

Thaís Pedruzzi do Nascimento

**Contribuições em Super-Resolução *Multi-Frame*  
Bayesiana Utilizando Informação Local e  
Registro Não Paramétrico**

Brasil

Maio de 2021



Thaís Pedruzzi do Nascimento

**Contribuições em Super-Resolução *Multi-Frame*  
Bayesiana Utilizando Informação Local e Registro Não  
Paramétrico**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Doutora em Engenharia Elétrica.

Universidade Federal do Espírito Santo  
Departamento de Engenharia Elétrica  
Programa de Pós-Graduação

Orientador: Evandro Ottoni Teatini Salles

Brasil  
Maio de 2021

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

---

P371c Pedruzzi do Nascimento, Thaís, 1989-  
Contribuições em Super-Resolução Multi-Frame Bayesiana Utilizando Informação Local e Registro Não Paramétrico / Thaís Pedruzzi do Nascimento. - 2021.  
130 f. : il.

Orientador: Evandro Ottoni Teatini Salles.  
Tese (Doutorado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Reconstrução de imagens. 2. Processamento de imagens. 3. Processamento de imagens - Técnicas digitais. 4. Imagens digitais. I. Ottoni Teatini Salles, Evandro. II. Universidade Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 621.3

---

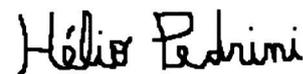
Thaís Pedruzzi do Nascimento

# **Contribuições em Super-Resolução *Multi-Frame* Bayesiana Utilizando Informação Local e Registro Não Paramétrico**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do título de Doutora em Engenharia Elétrica.

  
Evandro Ottoni Teatini Salles - UFES  
Orientador

  
Patrick Marques Ciarelli - UFES  
Membro Interno

  
Hélio Pedrini - UNICAMP  
Membro Externo

  
Jurgurta Rosa Montalvão Filho - UFS  
Membro Externo

  
Rodrigo Varejão Andreão - IFES  
Membro Externo

Brasil  
Maio de 2021

*Ao Vô Neno, que me ensinou a ter fé.*

*– Vô, sua Nininha agora é doutora!*

# Agradecimentos

Essa tese nasce de um processo longo de cinco anos e meio, que inundou todos os aspectos da minha vida. Eu tive apoio de muitas pessoas, que sempre acreditaram em mim e me acolheram de diversas maneiras. Temo esquecer de mencionar alguém e de escrever agradecimentos que não façam jus ao meu sentimento de gratidão. Mas segue, nessas páginas frias, a minha tentativa de agradecer a todo o calor que vem dessa minha rede de apoio infindável e incrível.

Aos meus pais, pelo suporte incondicional (financeiro, emocional e todos os outros tipos) e por não pouparem palavras para demonstrar o orgulho que sentem de mim. À minha irmã, Júlia, por ser minha parceira desde 1995. À toda a minha família, pelo acolhimento, torcida e orações pelo meu sucesso.

Ao Davi, Nanda e Larissa, por serem a primeira de muitas famílias que eu construí ao longo da vida. À Nathalia, por vibrar a todas as minhas conquistas e por me inspirar a me empoderar da profissão que eu escolhi. À Marcela e Carol, por estarem sempre presentes e torcerem por mim. À Érica e Luisa, por se manterem presentes, mesmo que a muitos quilômetros de distância. À Débora e Catherine, pela construção conjunta do nosso empoderamento profissional. Ao Bruno, pelo companheirismo diário, por vivermos juntos as nossas trajetórias acadêmicas, pela inesperada e maravilhosa transformação que a nossa relação sofreu nos últimos 3 anos. Ao Demuth, por me acompanhar nos meus 13 anos de UFES. Ao Léo, por manter o otimismo por todas as pessoas da Terra. Ao Fifi, por estar sempre disposto pra qualquer cilada. Ao Ricardo, Flávio, Clebeson, Gustavo, Daniel, Lucas, Matheus, Guilherme Butzke, Guilherme Cardoso, por todas as vezes que ajudamos uns aos outros a navegar as águas, por vezes tortuosas, da UFES. Ao Danilo, meu primo postiço, que está sempre a um WhatsApp de distância pra me dar conselhos de quem está a alguns degraus acadêmicos acima de mim. Aos amigos e amigas do MBV, por me inspirarem a lutar.

Ao Evandro, por ter topado me orientar quatro vezes, por não se assustar (tanto) com todas as lágrimas que eu derramei, por acolher as minhas inseguranças, por ser compreensivo, por acreditar no meu potencial. Por todas as conversas que ajudaram a moldar não só este trabalho, como a pesquisadora que sou.

Às professoras Raquel, Eliete, Márcia, Jussara e à Aline, por me inspirarem a ocupar, enquanto mulher, meu espaço nesse ambiente ocupado majoritariamente por homens.

Aos professores do PPGE, pelos ensinamentos (em aula e fora delas) e pelo apoio. À CAPES, pelas bolsas concedidas. À Universidade de Concórdia e ao prof. Nizar Bougila,

por me acolher enquanto pesquisadora *trainee* durante o doutorado sanduíche.

Aos professores doutores Patrick Marques Ciarelli, Hélio Pedrini, Jugurta Rosa Montalvão Filho e Rodrigo Varejão Andreão, por aceitarem o convite para compor a banca deste trabalho. Obrigada pelos valiosos apontamentos e sugestões.

A todas as professoras e professores que passaram pela minha formação, do jardim de infância ao doutorado, passando pelos cursos de inglês e piano, obrigada por despendarem esforço e dedicação pela educação de suas alunas e alunos. A todas as pessoas com quem eu troquei (e ainda troco) conhecimento que, hoje mais do que nunca, sei que vem de todos os lugares.

A todas as brasileiras e brasileiros que lutam por uma educação pública de qualidade.

# Resumo

A super-resolução *multi-frame* regularizada tem o objetivo de gerar uma imagem de alta resolução a partir de várias imagens de baixa resolução, pelo alinhamento delas e pela regularização, que ajusta a relação entre supressão de ruído e recuperação de informações de borda. Neste contexto, a partir do uso de informações locais e do registro *Demons*, três abordagens foram propostas nesta tese: a abordagem híbrida, a super-resolução com registro *Demons* e a abordagem baseada em *patches*. Para a abordagem híbrida, foram aplicadas múltiplas redes neurais de uma camada oculta, treinadas pelo algoritmo *Regularized Extreme Learning Machine*, em conjunto com os métodos de super-resolução *multi-frame Bilateral Total Variation* (BTV) e do *Iterative Re-Weighted Super-Resolution* (IRWSR), derivando nos métodos HyBTV e HyRWSR, respectivamente. Para a super-resolução com registro *Demons*, dois métodos foram propostos: o D-BTVIR e o D-IRWIR, que combinam o registro *Demons* com reconstrução baseada no BTV e IRWSR, respectivamente. Para a abordagem baseada em *patches* (PB, *patch-based*), que consiste em alinhar os *patches*, aplicar uma varredura e reconstruí-los individualmente, foram propostos três métodos: PB - sem classificação, PB - suavidade e PB - variância. No PB - sem classificação, todos os *patches* são super-resolvidos pelo método IRW, baseado no IRWSR. Para o PB - suavidade, os *patches* são classificados como homogêneos ou não, de acordo com a métrica “suavidade”, e super-resolvidos pela interpolação bicúbica, se homogêneos, e pelo IRW, caso contrário. Para o PB - variância, o mesmo é feito, mas considerando a variância como métrica de homogeneidade. Foram feitos experimentos utilizando deformação simulada em 119 imagens, dos conjuntos *Set5*, *Set14* e *B100*. As métricas PSNR, SSIM e o tempo de processamento foram analisados, empregando os testes de hipótese de Friedman, Nemenyi e Wilcoxon, além de análises por inspeção visual. O teste de Wilcoxon apontou melhor desempenho do HyBTV sobre o BTV com confiabilidade de 99% (valor- $p = 1,26 \times 10^{-11}$ ), e do HyRWSR sobre o IRWSR (valor- $p = 3,17 \times 10^{-6}$ ), além de melhor relação de tempo de processamento e qualidade do HyBTV sobre o IRWSR, que é, em média, 3,6 vezes mais lento. Para a abordagem baseada no *Demons*, o teste de Wilcoxon apontou melhor desempenho do D-BTVIR sobre o BTV com valor- $p = 3,52 \times 10^{-7}$ , e do D-IRWIR sobre o IRWSR com valor- $p = 2,95 \times 10^{-8}$ . O teste de Nemenyi apontou equivalência estatística entre o D-BTVIR e o IRWSR, mas o D-BTVIR se mostrou, em média, 7,2 vezes mais rápido. O teste de Nemenyi apontou melhor desempenho das três variações do método PB, comparadas ao BTV e IRWSR, além de equivalência estatísticas entre os três, com vantagem do PB - variância em relação ao tempo de processamento. Por fim, as análises por inspeção visual corroboraram os resultados dos testes de hipótese.

**Palavras-chave:** Super-resolução *multi-frame* regularizada. Registro *Demons*. *Extreme Learning Machine*. Super-resolução local.



# Abstract

Regularized multi-frame super-resolution aims to generate a high-resolution image from several low-resolution images, from the alignment of them and the use of regularization, which weighs noise suppression and edge information restoration. In this context, three approaches are proposed in this thesis, using local information and Demons registration: the hybrid approach, Demons super-resolution, and the patch-based approach. Regarding the hybrid approach, multiple single-layer neural networks were trained by the Regularized Extreme Learning Machine algorithm and implemented jointly with the multi-frame super-resolution methods Bilateral Total Variation (BTV) and Iterative Re-Weighted Super-Resolution (IRWSR), resulting in HyBTV and HyIRWSR, respectively. For Demons super-resolution two methods were proposed: D-BTVIR and D-IRWIR, which combine Demons registration with BTV-based an IRWSR-reconstruction, respectively. And for the patch-based approach (PB), which consists of aligning the patches, scan and super-resolve them individually, three methods were proposed: PB - no classification, PB - smoothness, and PB - variance. For PB - no classification all patches are super-resolved by IRW, which is based on IRWSR. For PB - smoothness, the patches are classified as homogeneous or not, by using smoothness as metrics, and super-resolved via bicubic interpolation, if homogeneous or via IRW, otherwise. PB - variance is based on the same procedure, but considering variance as the homogeneity metric. Experiments were conducted using simulated deformation in 119 images, from Set5, Se14 and B100. PSNR, SSIM and the execution time were analyzed by the use of Friedman, Nemenyi and Wilcoxon hypothesis tests, besides visual analysis. The Wilcoxon tests suggested a better performance of HyBTV over BTV with 99% reliability ( $p$ -value =  $1.26 \times 10^{-11}$ ), and of HyIRWSR over IRWSR ( $p$ -value =  $3.17 \times 10^{-6}$ ), besides better time-quality trade-off of HyBTV over IRWSR, which is, on average, 3.6 times slower than the first one. For the Demons-based approach, the Wilcoxon test suggested better performance of D-BTVIR over BTV with  $p$ -value =  $3.52 \times 10^{-7}$ , and the same thing was noted considering D-IRWIR over IRWSR, with  $p$ -value  $2.95 \times 10^{-8}$ . The Nemenyi test suggested statistical equivalence between D-BTVIR and IRWSR, however, D-BTVIR was 7.2 times faster. Nemenyi test also suggested better performance of the three variations of PB, when compared to BTV and IRWSR, besides statistical equivalence with each other, with PB - variance being the fastest one of them. Finally, the visual analysis supported the results from the hypothesis tests.

**Keywords:** Regularized Multi-frame Super-resolution. Demons Registration. Extreme Learning Machine. Local Super-resolution.



# Lista de ilustrações

Figura 1 – Modelo típico de um sistema de aquisição de imagens . . . . .	27
Figura 2 – Exemplo ilustrativo das etapas do processo de gerar uma imagem digital, (a) representação de uma imagem analógica, (b) seção transversal (ainda analógica) da região entre os pontos A e B da imagem, (c) amostragem e quantização e (d) versão digital da seção transversal A-B . . . . .	36
Figura 3 – Ilustração comparativa de diferentes resoluções. A região à esquerda da linha verde tem maior resolução e, a região à direita tem menor resolução. . . . .	37
Figura 4 – Esquema simplificado do modelo de aquisição. . . . .	38
Figura 5 – Exemplo de imagem super-resolvida utilizando o método BTV em que (a) é a imagem LR de entrada e (b) é a imagem HR estimada. . . . .	43
Figura 6 – Exemplo de imagem super-resolvida utilizando o método IRWSR em que (a) é a imagem LR de entrada e (b) é a imagem HR estimada. . . . .	45
Figura 7 – Esquema simplificado do funcionamento de um método de Super-resolução <i>single-frame</i> baseado em aprendizado, quanto à etapa de aprendizado com exemplos externos. . . . .	46
Figura 8 – Fluxograma das etapas de (a) treinamento e (b) reconstrução do método de SR <i>single-frame</i> baseado em exemplos com agrupamento usando informações do gradiente. . . . .	48
Figura 9 – Fluxograma do cálculo de um vetor de saída, redimensionado para dimensão $d_c \times d_c$ , colocado sobre o <i>grid</i> da imagem de alta frequência estimada, a partir de uma região $d_c \times d_c$ da imagem de baixa resolução interpolada. . . . .	50
Figura 10 – Fluxograma do cálculo de um <i>pixel</i> da imagem de alta frequência estimada. . . . .	51
Figura 11 – Representação gráfica de uma SLFN. . . . .	52
Figura 12 – Esquema simplificado do modelo de aquisição. . . . .	52
Figura 13 – Exemplo ilustrativo de um gráfico de distância crítica de Nemenyi. . . . .	56
Figura 14 – Espectro de Fourier das imagens residuais resultantes dos métodos (a) BTV, de Farsiu et al. (2004) e (b) IRWSR, de Köhler et al. (2016), quando aplicados às imagens <i>baby_GT</i> do <i>Set5</i> , e <i>barbara</i> e <i>coastguard</i> do <i>Set14</i> . . . . .	62
Figura 15 – Esquema ilustrativo do método híbrido proposto. No topo, em verde, a etapa de treino, e embaixo, em azul a etapa de teste. . . . .	63
Figura 16 – Imagens utilizadas para treinamento das redes neurais. . . . .	64
Figura 17 – Imagens utilizadas para o teste, provenientes do conjunto <i>Set5</i> . . . . .	64
Figura 18 – Imagens utilizadas para o teste, provenientes do conjunto <i>Set14</i> . . . . .	65
Figura 19 – Imagens utilizadas para o teste, provenientes do conjunto <i>B100</i> . . . . .	66

Figura 20 – Gráfico de diferença crítica (nível de significância de 1%) dos resultados dos métodos BTV, PSNR, HyBTV e HyIRWSR, considerando as imagens do <i>Set5</i> , <i>Set14</i> e <i>B100</i> e as métricas (a) PSNR e (b) SSIM. . . . .	69
Figura 21 – Tempo médio de processamento por <i>pixel</i> superresolvido de cada etapa dos métodos <i>BTV</i> , <i>HyBTV</i> , <i>IRWSR</i> e <i>HyIRWSR</i> aplicados às imagens do <i>Set14</i> . . . . .	71
Figura 22 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>156065</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	72
Figura 23 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>134035</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	73
Figura 24 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>37073</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	74
Figura 25 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>21077</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	75
Figura 26 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>145086</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	76

Figura 27 – Imagens super-resolvidas, referentes à imagem original de alta resolução <i>219090</i> , (a) <i>ground-truth</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) detaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do <i>ground truth</i> . . . . .	77
Figura 28 – Esquema simplificado do método “Super-resolução com registro <i>Demons</i> e reconstrução Bayesiana Regularizada” . . . . .	80
Figura 29 – Gráficos de distância crítica, referente ao teste de Nemenyi aplicado aos métodos BTV, IRWSR, D-BTVIR e D-IRWIR. Têm-se (a) o gráfico referente aos valores de PSNR, (b) o gráfico referente aos valores de SSIM e (b) o gráfico referente ao tempo de processamento. . . . .	86
Figura 30 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>flowers</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) detaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	87
Figura 31 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>156065</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) detaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	88
Figura 32 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>ppt3</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) detaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	89
Figura 33 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>zebra</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) detaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	90

Figura 34 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>monarch</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	91
Figura 35 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>37073</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) <i>ground truth</i> . . . . .	92
Figura 36 – Exemplo ilustrativo de uma mesma imagem armazenada em diferentes resoluções: (a) 50 dpi, (b) 100 dpi e (c) 500 dpi. . . . .	93
Figura 37 – Esquema simplificado do método “Super-Resolução Local com Registro <i>Demons</i> , Varredura em Patches e Classificação de Homogeneidade” . . . . .	94
Figura 38 – Exemplo de super-resolução com varredura de <i>patches</i> em que se observa artefatos nas regiões limítrofes entre os <i>patches</i> , sinalizados pelas setas vermelhas. . . . .	96
Figura 39 – Exemplo ilustrativo do <i>padding</i> realizado antes da varredura de <i>patches</i> , considerando $p_h = p_v = 2$ e uma imagem de dimensão (a) $4 \times 4$ antes de <i>padding</i> e dimensão (b) $8 \times 8$ depois do <i>padding</i> . . . . .	97
Figura 40 – Exemplos de <i>patches</i> homogêneos. . . . .	98
Figura 41 – Exemplos de <i>patches</i> não homogêneos. . . . .	99
Figura 42 – Distribuição dos valores das métricas variância e suavidade (a) de todos os 7.126 <i>patches</i> extraídos e (b) dos <i>patches</i> que apresentam valores das métrica dentro das faixas escolhidas para fazer a validação dos hiperparâmetros. . . . .	101
Figura 43 – Exemplo ilustrativo de imagem contaminada com ruído gaussiano branco aditivo interpolada por (a) interpolação bicúbica e (b) interpolação por vizinho mais próximo. . . . .	102
Figura 44 – <i>Ranks</i> médios das 88 variantes consideradas: em (a) o 1° ao 22° colocado, em (b) o 23° ao 44° colocado, em (c) o 45° ao 66° colocado e em (d) o 67° ao 88° colocado. Os círculos azuis referem-se aos resultados do Cenário 1, os quadrados laranjas referem-se ao Cenário 2, os losangos verdes referem-se ao Cenário 3 e, finalmente, os triângulos rosas referem-se ao Cenário 4. . . . .	105

Figura 45 – Gráficos de distância crítica referente ao teste de Nemenyi aplicado às cinco variantes melhores classificadas nos experimentos decorridos para a validação, em (a) o gráfico referente aos valores de qualidade de imagem, PSNR e SSIM e em (b) o gráfico referente ao tempo de processamento. . . . .	107
Figura 46 – Gráficos de distância crítica, referente ao teste de Nemenyi aplicado aos métodos BTV, IRWSR, PB - sem classificação, PB - variância e PB - <i>smoothness</i> . Têm-se, (a) o gráfico referente aos valores de PSNR, (b) o gráfico referente aos valores de SSIM e (b) o gráfico referente ao tempo de processamento. . . . .	110
Figura 47 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>156065</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	111
Figura 48 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>ppt3</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	112
Figura 49 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>flowers</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	113

Figura 50 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>zebra</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	114
Figura 51 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>monarch</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	115
Figura 52 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução <i>37073</i> , (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do <i>ground truth</i> . . . . .	116

# Lista de quadros

Quadro 1 – Nomes e descrições resumidas dos métodos propostos nesta tese e os capítulos que abordam cada um deles. . . . .	32
Quadro 2 – Parâmetros do modelo de aquisição, referente aos experimentos híbridos utilizando os métodos <i>multi-frame</i> BTV e IRWSR. . . . .	67
Quadro 3 – Parâmetros do agrupamento e R-ELM, referente aos experimentos híbridos utilizando os métodos <i>multi-frame</i> BTV e IRWSR. . . . .	67
Quadro 4 – Parâmetros do modelo de aquisição utilizado nos experimentos referentes à super-resolução com registro <i>Demons</i> . . . . .	83
Quadro 5 – Cenários considerados para a validação do método proposto e respectivos parâmetros. . . . .	104
Quadro 6 – Parâmetros do modelo de aquisição utilizado nos experimentos referentes à super-resolução local, com registro <i>Demons</i> , varredura de <i>patches</i> e classificação de homogeneidade proposta. . . . .	108



# Lista de tabelas

Tabela 1	– Dimensão da matrizes e vetores presentes no modelo de aquisição . . . .	41
Tabela 2	– Valores críticos para o teste de Nemenyi bicaudal. . . . .	55
Tabela 3	– Resultado do teste de Wilcoxon para os métodos HyBTV e HyIRWSR, quando comparados aos métodos BTV e IRWSR, respectivamente . . . .	68
Tabela 4	– Tempo de processamento médio, em segundos, de cada experimento, considerando o tempo médio de cada imagem dos conjuntos de teste <i>B100</i> , <i>Set14</i> e <i>Set5</i> . Os valores em negrito representam o melhor resultado para cada conjunto e os valores sublinhados representam o segundo melhor resultado. . . . .	70
Tabela 5	– Resultado do teste de Wilcoxon para os métodos D-BTVIR e D-IRWIR, quando comparados aos métodos BTV e IRWSR, respectivamente . . . .	85



# Lista de abreviaturas e siglas

AWGN	<i>Additive White Gaussian Noise</i>
BTV	<i>Bilateral Total Variation</i>
CCD	<i>Charge-Couple Device</i>
CD	<i>Critical Difference</i>
CMOS	<i>Complementary Metal-Oxide-Semiconductor</i>
CP	<i>Control points</i>
D-BTVIR	<i>Demons Bilateral Total Variation Image Reconstruction</i>
D-IRWIR	<i>Demons Iterative Re-Weighted Image Reconstruction</i>
ELM	<i>Extreme Learning Machine</i>
FMD	<i>Flickr Material Database</i>
HF	<i>High frequency</i>
HR	<i>High Resolution</i>
HVS	<i>Human Visual System</i>
IRWSR	<i>Iterative Re-Weighted Super-Resolution</i>
IRW	<i>Iterative Re-Weighted Image Reconstruction</i>
LR	<i>Low Resolution</i>
MSE	<i>Minimum Squared Error</i>
PB	<i>Patch-based</i>
PSF	<i>Point Spread Function</i>
PSNR	<i>Peak Signal-to-Noise Ratio</i>
R-ELM	<i>Regularized Extreme Learning Machine</i>
RGB	<i>Red Green Blue</i>
SCG	<i>Scaled Conjugate Gradient</i>

SLFN	<i>Single Hidden Layer Feed-foward Neutal Network</i>
SR	Super-Resolução
SSIM	<i>Structural Similarity Index</i>
TV	<i>Total Variation</i>

# Lista de símbolos

$\mathbf{x}$	Imagem de alta resolução original, em formato lexicográfico
$\hat{\mathbf{x}}$	Imagem de alta resolução estimada, , em formato lexicográfico
$K$	Quantidade de imagens de baixa resolução
$k$	Índice da imagem de baixa resolução, onde $1 \leq k \leq K$
$\mathbf{y}_k$	Imagem de baixa resolução de índice $k$ , em formato lexicográfico
$M$	Número de colunas das imagens de alta resolução
$N$	Número de linhas das imagens de alta resolução
$m$	Número de colunas das imagens de baixa resolução
$n$	Número de linhas das imagens de baixa resolução
$s$	Fator de magnificação ( $M = m \cdot s$ e $N = n \cdot s$ )
$\mathbf{W}_k$	Operador de deslocamento ( <i>warping</i> ) referente à imagem de baixa resolução de índice $k$
$h$	Horizontal
$v$	Vertical
$\mathbf{w}_h$	Vetor de translação horizontal
$\mathbf{w}_v$	Vetor de translação vertical
$\mathbf{w}_\theta$	Vetor de rotação
$\mathbf{B}$	Operador de borramento ( <i>blurring</i> )
$\mathbf{D}$	Operador de subamostragem ( <i>downsampling</i> )
$F(\cdot)$	Termo de fidelidade (ou norma do erro)
$R(\cdot)$	Termo regularizador (ou <i>priori</i> )
$\Gamma_{\text{BTV}}$	Termo regularizador para o método BTV
$\alpha_0$	Fator de escala de ponderação da <i>priori</i> , para o método BTV

$A$	Mapa de confiança, usado para ponderação da <i>priori</i> , para o método IRWSR
$\Gamma_{\text{IRWSR}}$	Termo regularizador usado pelo método IRWSR
$\mathbf{Y}_0$	Imagem estimada inicial, referente aos métodos baseados em aprendizado
$\mathbf{Z}$	Imagem residual, referente aos métodos baseados em aprendizado
$\ \cdot\ _i$	Norma- $\ell_i$
$\mathbf{g}$	Vetor gradiente
$\mathbf{G}$	Matriz de vetores de gradiente
$C$	Termo de regularização do R-ELM
$d_g \times d_g$	Dimensão do <i>patch</i> para calcular as estatísticas do gradiente
$d_c \times d_c$	Dimensão do <i>patch</i> para extrair o vetor de característica
$T$	Tempo de processamento
$I_{\text{dataset}}$	Quantidade de imagens de um dado <i>dataset</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>27</b>
<b>1.1</b>	<b>Hipóteses</b>	<b>31</b>
<b>1.2</b>	<b>Objetivos e Contribuições</b>	<b>31</b>
<b>1.3</b>	<b>Publicações</b>	<b>33</b>
<b>1.4</b>	<b>Estrutura da Tese</b>	<b>33</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>35</b>
<b>2.1</b>	<b>Conceitos</b>	<b>35</b>
2.1.1	Resolução Espacial	35
2.1.2	Modelo de Aquisição de Imagens	36
2.1.3	Registro	38
2.1.4	Registro Não Paramétrico <i>Demons</i>	39
2.1.5	Super-Resolução Bayesiana Regularizada	41
2.1.5.1	<i>Bilateral Total Variation - BTV</i>	42
2.1.5.2	<i>Iteratively Re-Weighted Super-Resolution - IRWSR</i>	43
2.1.6	Super-Resolução Baseada em Aprendizado de Máquinas	44
2.1.6.1	Estatísticas do Gradiente	46
2.1.6.2	Super-resolução Usando Múltiplos Modelos de Rede Neural	47
2.1.6.3	<i>Regularized Extreme Learning Machine</i>	49
2.1.7	Qualidade de Imagem	51
2.1.7.1	PSNR	53
2.1.7.2	SSIM	53
2.1.8	Testes de Hipótese	54
2.1.8.1	Teste de Friedman	54
2.1.8.2	Teste de Nemenyi	55
2.1.8.3	Teste de Wilcoxon Unicaudal	55
<b>2.2</b>	<b>Trabalhos Relacionados</b>	<b>57</b>
2.2.1	Super-Resolução <i>Multi-Frame</i> Regularizada	57
2.2.2	Super-Resolução Baseada em Redes Neurais	59
<b>3</b>	<b>SUPER-RESOLUÇÃO HÍBRIDA BAYESIANA REGULARIZADA COM MÚLTIPLAS REDES NEURAIAS</b>	<b>61</b>
<b>3.1</b>	<b>Método Proposto</b>	<b>62</b>
<b>3.2</b>	<b>Resultados Experimentais</b>	<b>67</b>
3.2.1	Comparações Estatísticas	67
3.2.1.1	Teste Pareado Unicaudal de Wilcoxon	67

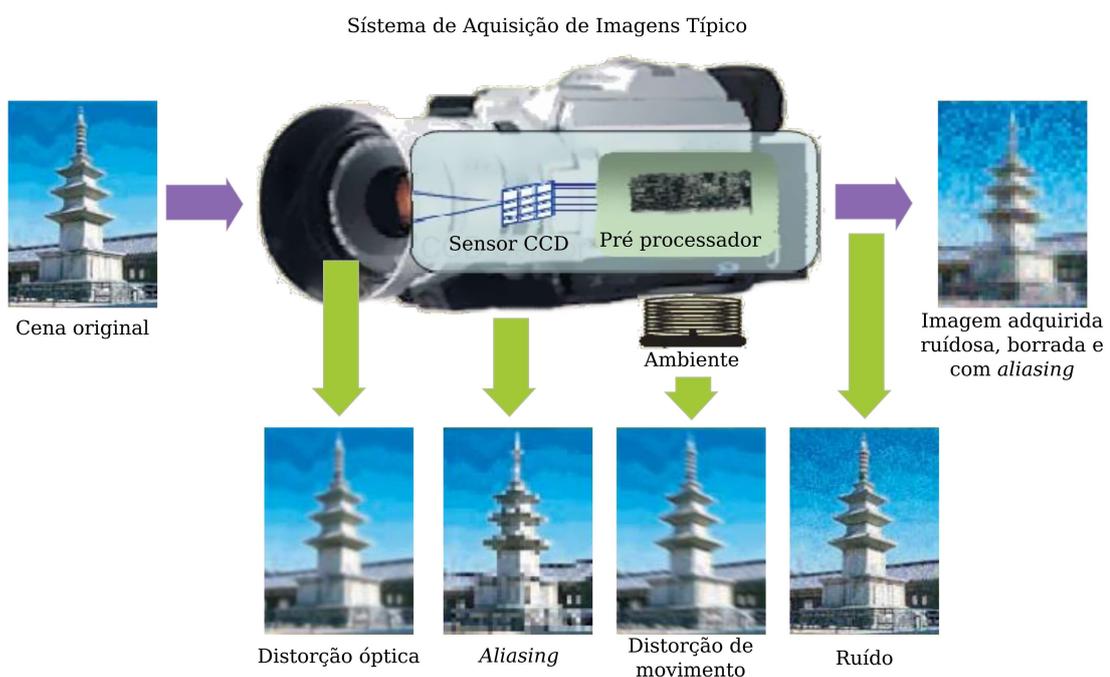
3.2.1.2	Testes de Friedman e Nemenyi . . . . .	69
3.2.2	Análise do Tempo de Processamento . . . . .	69
3.2.3	Análise por Inspeção Visual . . . . .	71
<b>4</b>	<b>SUPER-RESOLUÇÃO COM REGISTRO <i>DEMONS</i> E RECONSTRUÇÃO BAYESIANA REGULARIZADA . . . . .</b>	<b>79</b>
<b>4.1</b>	<b>Método Proposto . . . . .</b>	<b>80</b>
4.1.1	Registro <i>Demons</i> . . . . .	80
4.1.2	Operador de Deslocamento . . . . .	81
4.1.3	Reconstrução Bayesiana Regularizada . . . . .	82
4.1.3.1	BTVIR . . . . .	82
4.1.3.2	Reconstrução IRW . . . . .	82
<b>4.2</b>	<b>Experimentos e Resultados . . . . .</b>	<b>83</b>
4.2.1	Comparações Estatísticas . . . . .	83
4.2.1.1	Teste Pareado Unicaudal de Wilcoxon . . . . .	84
4.2.1.2	Teste de Friedman e Nemenyi . . . . .	84
4.2.2	Análise por Inspeção Visual . . . . .	86
<b>5</b>	<b>SUPER-RESOLUÇÃO LOCAL COM REGISTRO <i>DEMONS</i>, VARREDURA EM PATCHES E CLASSIFICAÇÃO DE HOMOGENEIDADE . . . . .</b>	<b>93</b>
<b>5.1</b>	<b>Método Proposto . . . . .</b>	<b>94</b>
5.1.1	Registro . . . . .	95
5.1.2	Varredura de <i>Patches</i> . . . . .	95
5.1.3	Classificação de Homogeneidade . . . . .	97
5.1.3.1	Variância . . . . .	99
5.1.3.2	Suavidade . . . . .	99
5.1.3.3	Distribuição dos Valores de Variância e Suavidade . . . . .	100
5.1.4	Interpolação Bicúbica . . . . .	101
5.1.5	Reconstrução IRW . . . . .	102
5.1.6	Remontagem da Imagem HR Estimada . . . . .	103
<b>5.2</b>	<b>Validação . . . . .</b>	<b>103</b>
<b>5.3</b>	<b>Experimentos e Resultados . . . . .</b>	<b>107</b>
5.3.1	Comparação Estatística . . . . .	108
5.3.2	Análise por Inspeção Visual . . . . .	109
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>117</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>121</b>

# 1 Introdução

A visão é o mais avançado dos sentidos, de forma que não é surpreendente que as imagens exerçam um papel importante na percepção humana. Uma imagem pode ser definida como uma função bidimensional cujo valor em cada par de coordenadas representa a intensidade de cor no ponto. Quando tal função é discreta, ou seja, composta por um valor finito de elementos, a imagem representada é do tipo digital e cada elemento é denominado *pixel* (GONZALEZ; WOODS, 2012).

O teorema de amostragem de Nyquist-Shannon garante a recuperação de um sinal de banda limitada após discretização, desde que a frequência de amostragem seja maior que o dobro da maior frequência presente no mesmo (OPPENHEIM; WILLSKY, 2010). Entretanto, ao se tratar de imagens, a banda de frequência é limitada por características construtivas do sensor óptico, cujo projeto envolve uma série de compromissos, de forma que sistemas de aquisição de imagem sejam tipicamente projetados com algum nível de *aliasing* (MILANFAR, 2011). Além do *aliasing*, outras distorções podem estar presentes durante a aquisição de uma imagem digital, como a distorção de movimento e o ruído, ilustradas na Figura 1. Em decorrência disso, as condições do teorema da amostragem não são satisfeitas, na prática, para imagens digitais, estando essas tipicamente sujeitas a distorções e perda de informação, que causam limitação no nível de detalhes discerníveis.

Figura 1 – Modelo típico de um sistema de aquisição de imagens



Fonte: Adaptada de Park, Park e Kang (2003)

Nesse contexto, define-se resolução espacial, que se refere à densidade de *pixel* de uma imagem digital, ou seja, a quantidade de *pixels* por unidade de área. Em outras palavras, quanto maior a resolução espacial, maior o número de detalhes perceptíveis. A resolução espacial é limitada pelos dispositivos de aquisição de imagem e sensores contidos neles que, desde a década de 70, são tipicamente do tipo CCD (do inglês, *Charge-Coupled Device*) ou CMOS (do inglês, *Complementary Metal-Oxide-Semiconductor*). As características construtivas dos sensores definem a dimensão da matriz de saída que representa a imagem e, conseqüentemente, a resolução espacial da mesma (MILANFAR, 2011). Todavia, alterar as características construtivas de um sensor pode ser oneroso ou, ainda, pode-se desejar aumentar a resolução da imagem após sua aquisição.

Considerando esse contexto, em que se deseja aumentar o nível de detalhes discerníveis em imagens digitais já adquiridas, conceitua-se como super-resolução *multi-frame* o processo de gerar uma imagem de saída de maior resolução espacial, a partir de um conjunto de imagens de menor resolução espacial. A imagem de entrada é denominada “imagem de baixa resolução” (LR, do inglês, *Low Resolution*) e a imagem de saída, por sua vez, é chamada de “imagem de alta resolução” (HR, do inglês, *High Resolution*) (FARSIU et al., 2004)(PELEG; KEREN; SCHWEITZER, 1987). A premissa da abordagem *multi-frame* é que a disponibilidade de múltiplas imagens da mesma cena provê informações complementares, desde que haja deslocamento *sub-pixel* entre tais imagens, ou seja, em frações de *pixels* (PARK; PARK; KANG, 2003). Uma vez que tal deslocamento é desconhecido, o cálculo dele é parte do problema e é denominado de registro. Além disso, é importante mencionar que o problema de super-resolução é tipicamente descrito como um problema inverso mal-posto <sup>1</sup> em que as imagens LR são as variáveis observadas, e a imagem HR é a variável desconhecida que se deseja estimar (KATSAGGELOS; MOLINA; MATEOS, 2007).

O paradigma Bayesiano, em particular, é amplamente utilizado na literatura (NASROLLAHI; MOESLUND, 2014), no qual todos os parâmetros e variáveis são considerados estocásticos, aos quais são atribuídas distribuições de probabilidade que, por sua vez, são baseadas em crenças subjetivas (KATSAGGELOS; MOLINA; MATEOS, 2007). Considerando a natureza mal-posta do problema, tais crenças subjetivas agem como regularização, contribuindo para restringir a busca por boas soluções, dentre as múltiplas possíveis (DAUBECHIES; DEFRISE; MOL, 2004). Em especial, a escolha do termo de regularização geralmente está atrelada à intenção de melhor reconstruir informações de alta frequência, como borda e textura. Nesse contexto, vale ressaltar o compromisso entre a reconstrução de bordas e a supressão de ruído, bem como a importante atuação do termo regularizador nesse sentido. Destacam-se o uso da regularização adaptativa (KÖHLER et al., 2016), em que a influência do termo regularizador é ponderada localmente.

<sup>1</sup> Um problema mal-posto, para esta tese, é aquele que não satisfaz pelo menos umas das condições de Hadamard (1923) para que seja bem-posto: (i) existência, (ii) unicidade e (iii) estabilidade da solução.

Quanto à implementação do paradigma Bayesiano e ao problema de registro, de acordo com [Katsaggelos, Molina e Mateos \(2007\)](#), as duas metodologias mais empregadas na literatura são: (i) a metodologia alternada, em que um procedimento cíclico descendente é aplicado para atualizar o cálculo da imagem HR e do deslocamento iterativamente; e (ii) a metodologia sequencial, em que assume-se que o deslocamento é conhecido ou calculado separadamente da imagem estimada HR e, portanto, considerado como entrada para o procedimento iterativo, que calcula, por sua vez, a imagem HR estimada. Nesta tese, aplica-se essas duas categorizações para qualquer implementação de super-resolução *multi-frame* e, além disso, refere-se à segunda etapa da metodologia sequencial de reconstrução *multi-frame*.

A supramencionada característica mal-posta do problema de super-resolução deve-se ao número de imagens LR e às distorções introduzidas pela câmera ([PARK; PARK; KANG, 2003](#)). Portanto, aplicar um método de registro não mitiga tal característica. Ou seja, sem perda de generalidade, assume-se que a etapa de reconstrução também é um problema mal-posto, tendo assim múltiplas soluções. A partir desse raciocínio, introduz-se, neste trabalho, a perspectiva de que a etapa de registro da metodologia sequencial é uma forma de mudar a condição inicial da etapa do problema iterativo de reconstrução *multi-frame*. Dessa maneira, apresenta-se aqui a primeira hipótese deste trabalho na qual a escolha de um método de registro eficaz pode contribuir na busca pela melhor solução do problema de reconstrução, conseqüentemente, aprimorando o resultado do método de super-resolução *multi-frame* sequencial.

O problema de registro pode ser categorizado em paramétrico ou não paramétrico ([ZITOVÁ; FLUSSER, 2003](#)). No caso paramétrico, assume-se uma função de mapeamento que relaciona a imagem a ser registrada e a imagem de referência, de forma que a solução do problema é dada ao encontrarem-se os parâmetros que melhor se ajustam a tal função. Entretanto, métodos paramétricos dependem de suposições acerca do deslocamento, que devem ser apropriadamente justificadas, o que é desafiador. Devido a essa dependência, tais métodos são limitados para garantir a factibilidade da solução não sendo, portanto, capazes de preservar singularidades tais como descontinuidades locais ([QIU; XING, 2013](#)). Por outro lado, métodos não paramétricos exploram as diferenças elemento-a-elemento entre as imagens deslocada e de referência, sendo focados em descrever o deslocamento espacial local ([VERCAUTEREN et al., 2007](#)), ([THIRION, 1998](#)). Além disso, segundo [Vandewalle, Sbaiz e Vetterli \(2011\)](#), o problema de registro é inerentemente não-linear quando da presença de translação. Uma vez que a literatura de super-resolução tipicamente considera translação e rotação para o deslocamento das imagens LR, o uso de um método de registro não-linear é justificado. Particularmente, dentre os métodos não-lineares de registro, o algoritmo *Demons*, proposto originalmente em ([THIRION, 1998](#)), é um método eficiente ([VERCAUTEREN et al., 2007](#)) capaz de resolver o problema de deslocamento *sub-pixel* de forma eficaz.

Em contrapartida, os métodos baseados em exemplos extraem novas informações a partir do mapeamento da relação entre as imagens LR e as imagens HR, utilizando um banco de dados externo e técnicas de aprendizado de máquina, como em (FREEMAN; JONES; PASZTOR, 2002) e (COSMO; SALLES, 2019). Nesse contexto, o problema de super-resolução é tido como um problema de regressão não-linear. Tipicamente, na etapa de treinamento, a imagem LR é transformada para o *grid* da imagem HR através de uma interpolação bicúbica e, então, subtraída da imagem original. Dessa subtração, resulta uma imagem com a predominância de componentes de alta frequência, chamada de “imagem HF” (do inglês, *High Frequency*) que, por sua vez, tem sua relação com a imagem LR mapeada. Esse mapeamento, junto à interpolação, é aplicado às imagens LR e, finalmente, gera-se a imagem HR estimada. Além disso, esse processo é, geralmente, executado considerando sub-regiões da imagem, ao invés da imagem inteira. Compreende-se, nesta tese, de que dessa forma é possível treinar regressores especializados localmente, dependendo das características das sub-regiões.

Outro trabalho em que se aplica a informação local presente nas imagens de entrada foi proposto por Su et al. (2012), em que uma varredura em *patches*<sup>2</sup> é feita, após o registro das imagens LR, seguida de uma classificação desses *patches*. As sub-regiões são super-resolvidas<sup>3</sup>, então, de acordo com a classe de cada uma delas, a partir da escolha de métodos que se ajustem melhor às características locais. A diferença significativa entre essa proposta e a proposta de Cosmo e Salles (2019), por exemplo, está no fato de que, no último os diferentes métodos são redes neurais com a mesma arquitetura e parâmetros diferentes, enquanto em (SU et al., 2012), os diferentes métodos (chamados de métodos candidatos, pelo autor) são de categorias completamente distintas, não tendo relação entre eles, de forma que tal proposta pode ser vista como mais adaptativa e flexível. Entretanto, vale ressaltar que ambos os métodos bem como (KÖHLER et al., 2016) se utilizam de informações geométricas da imagem, em especial, do que se chama, nesta tese, de homogeneidade. Considera-se, neste contexto, que a região de uma imagem é homogênea quando não há variação significativa de intensidade de *pixel* em nenhuma direção.

No que se diz respeito à análise de resultados, ao longo da literatura de super-resolução *multi-frame* dois tipos de análises estão presentes, sendo elas (i) uma análise visual, em que os resultados são comparados a partir de determinados aspectos subjetivos; e (ii) uma análise quantitativa baseada em métricas de qualidade, em que são apresentadas tabelas contendo os valores das métricas escolhidas considerando-se poucas imagens, muitas vezes selecionadas de forma arbitrária. Em ambos os casos observa-se uma lacuna tratando-se de análises estatísticas mais robustas no contexto de super-resolução *multi-frame*. Dessa

<sup>2</sup> Um *patch* refere-se a uma sub-região da imagem, constituída por pixels vizinhos. Utiliza-se o termo em inglês neste trabalho para manter coerência com a literatura.

<sup>3</sup> O verbo em inglês *to super-resolve* é utilizado na literatura para denominar o processo de estimar uma imagem por super-resolução. Neste trabalho, propõe-se a tradução livre do termo: super-resolver, suas conjugações e variantes (neste trecho, “super-resolvida”)

forma, introduz-se nesta tese o uso de testes de hipótese tais quais os apresentados em (DEMŠAR, 2006), que possibilitam a comparação de diferentes métodos em diversas imagens e banco de dados.

## 1.1 Hipóteses

O desenvolvimento deste trabalho decorre das hipóteses de que (i) o uso de informações locais pode trazer bons resultados quando aplicado a métodos de super-resolução *multi-frame*, sendo possível, a partir dele, reconstituir detalhes relevantes e fiéis à cena real, de onde a imagem foi observada e (ii) o uso do método de registro não paramétrico *Demons*, quando aplicado à uma abordagem de super-resolução *multi-frame* sequencial, proporciona um ponto de partida melhor para o processo iterativo de reconstrução, melhorando o resultado final da super-resolução em questão.

## 1.2 Objetivos e Contribuições

O objetivo geral deste trabalho é investigar o uso de características locais das imagens, sobretudo em relação à homogeneidade, bem como o uso do registro não paramétrico *Demons* em super-resolução *multi-frame* Bayesiana regularizada. Além disso, a fim de fornecer uma análise estatística robusta das abordagens propostas, apresenta-se a aplicação de testes de hipóteses aos resultados quantitativos dos métodos de super-resolução investigados. Dessa forma, esta tese apresenta as seguintes contribuições:

- Métodos de super-resolução híbrida, baseados na combinação de múltiplas redes neurais e super-resolução Bayesiana, chamados de HyIRWSR, quando do uso de super-resolução IRWSR e HyBTV, quando aplicada a super-resolução BTV;
- Métodos de super-resolução sequencial com registro *Demons* e reconstrução Bayesiana regularizada, chamados de D-IRWIR, quando do uso de reconstrução baseada em IRWSR e D-BTVIR, quando aplicada a reconstrução baseada em BTV;
- Abordagem sequencial com registro *Demons* e reconstrução em *patches*, de acordo com classificação local de homogeneidade, baseada em reconstrução Bayesiana IRW e interpolação simples;
- Análise estatística dos valores das métricas PSNR e SSIM dos resultados obtidos, baseada nos testes de hipótese de Friedman, Nemenyi e Wilcoxon.

O Quadro 1 mostra os nomes dos métodos propostos, relacionados com descrições resumidas e os capítulos que os abordam.

Quadro 1 – Nomes e descrições resumidas dos métodos propostos nesta tese e os capítulos que abordam cada um deles.

Nome do método	Descrição	Capítulo relacionado
HyBTV	Super-resolução híbrida baseada na combinação de múltiplas redes neurais e super-resolução regularizada BTV	Capítulo 3
HyIRWSR	Super-resolução híbrida baseada na combinação de múltiplas redes neurais e super-resolução regularizada IRWSR	Capítulo 3
D-BTVIR	Super-resolução sequencial com registro <i>Demons</i> e reconstrução regularizada baseada no método de super-resolução BTV	Capítulo 4
D-IRWIR	Super-resolução sequencial com registro <i>Demons</i> e reconstrução regularizada baseada no método de super-resolução IRWSR	Capítulo 4
PB - sem classificação	Super-resolução baseada em varredura de <i>patches</i> , sem classificação de homogeneidade, registro <i>Demons</i> e reconstrução IRW	Capítulo 5
PB - suavidade	Super-resolução baseada em varredura de <i>patches</i> , com classificação utilizando suavidade como métrica de homogeneidade, registro <i>Demons</i> , reconstrução IRW e interpolação simples	Capítulo 5
PB - variância	Super-resolução baseada em varredura de <i>patches</i> , com classificação utilizando variância como métrica de homogeneidade, registro <i>Demons</i> , reconstrução IRW e interpolação simples	Capítulo 5

Fonte: Elaborado pela autora

## 1.3 Publicações

Dois artigos de congresso e um artigo em periódico foram publicados como resultado desta pesquisa:

- NASCIMENTO, T. P.; Salles, E.O.T. A BAYESIAN MULTIFRAME SUPER-RESOLUTION ITERATIVE METHOD USING SSIM-BASED CONVERGENCE SUITABLE FOR REAL CASES. *In: XXI Congresso Brasileiro de Automática*, 2016, Vitória. CBA2016, 2016. p. 1512-1517
- NASCIMENTO, T. P.; SALLES, E. O. T A Hybrid Multiframe Superresolution Framework Using R-ELM Neural Networks and Maximum a Posteriori Estimation. *In: BRAZILIAN TECHNOLOGY SYMPOSIUM - BTSYM'18*, 2019, Campinas. BTSYM 2018 Proceedings, 2018. v. 1
- NASCIMENTO, T. P.; SALLES, E.O.T. Multi-Frame Super-Resolution Combining Demons Registration and Regularized Bayesian Reconstruction. **IEEE Signal Processing Letters**, v. 27, p. 2009-2013, 2020.

## 1.4 Estrutura da Tese

O Capítulo 1 contextualizou e descreveu o problema de super-resolução, introduzindo os aspectos revelantes para o desenvolvimento desta tese, além de tratar das hipóteses, objetivo e contribuições da mesma. O Capítulo 2 apresenta o referencial teórico deste trabalho, contendo os conceitos necessários para compreender o desenvolvimento do trabalho bem como um levantamento dos trabalhos relacionados. O Capítulo 3 explora os métodos de super-resolução híbrido, explicando a metodologia e apresentando os resultados obtidos. Da mesma forma, o Capítulo 4 trata dos métodos de super-resolução sequencial baseado em registro *Demons*, bem como o Capítulo 5 refere-se à terceira abordagem proposta nesta tese, baseada em reconstrução em *patches* e classificação de homogeneidade. Finalmente, o Capítulo 6 conclui esta pesquisa.



## 2 Referencial Teórico

Este capítulo apresenta os conceitos teóricos, relacionados ao desenvolvimento deste trabalho e os uma breve revisão da literatura, destacando-se os trabalhos relacionados a esta pesquisa.

### 2.1 Conceitos

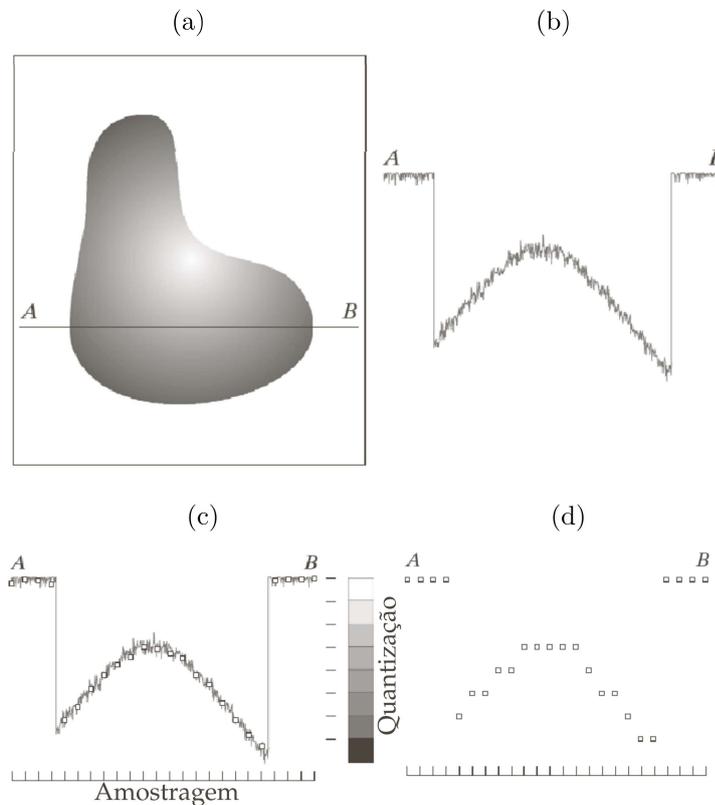
Primeiramente, apresenta-se o conceito de resolução espacial. Em seguida, o modelo de aquisição é explicado, uma vez que consiste na base da modelagem matemática utilizada na solução do problema de super-resolução *multi-frame*. Explana-se, logo após, a teoria de registro de imagens e o registro *Demons* utilizado em duas das três abordagens propostas neste trabalho. Na sequência, explicam-se os métodos de super-resolução *multi-frame* regularizados utilizados como base para todos os métodos proposto neste trabalho, além do método baseado em redes neurais utilizado como base para o método de super-resolução híbrido. As duas últimas seções do capítulo abordam as métricas PSNR e SSIM, tipicamente utilizadas para comparar resultados de métodos de super-resolução, e os testes de hipótese de Friedman e Nemenyi, introduzidos nesta tese como forma de analisar estatisticamente os resultados obtidos.

#### 2.1.1 Resolução Espacial

A criação de uma imagem digital envolve dois processos: quantização e amostragem (GONZALEZ; WOODS, 2012), como ilustrado na Figura 2. De forma geral, os parâmetros de tais processos definem a similaridade entre a imagem digital e sua versão analógica. A quantização se relaciona ao número de cores (ou tons de cinza, no caso de uma imagem preto e branca) contidos na imagem, o que se traduz numericamente em resolução de intensidade. A amostragem, por sua vez, se relaciona à quantidade de *pixels* contidos na imagem e implica na resolução espacial.

Unidades de medidas como “pares de linha por distância” e “pontos por unidade de distância” são alguns exemplos que definem, de forma numérica, a resolução espacial. Definições numéricas não são, entretanto, o foco deste trabalho. Interessa, aqui, bem como para a literatura de super-resolução em geral, a concepção intuitiva da resolução espacial, que diz respeito ao menor detalhe discernível em uma imagem. Nesse contexto, a resolução espacial é utilizada como um parâmetro comparativo entre diferentes versões de uma mesma imagem, como ilustra a Figura 3, em que o lado esquerdo da linha contínua verde apresenta maior resolução e o lado direito apresenta menor resolução.

Figura 2 – Exemplo ilustrativo das etapas do processo de gerar uma imagem digital, (a) representação de uma imagem analógica, (b) seção transversal (ainda analógica) da região entre os pontos A e B da imagem, (c) amostragem e quantização e (d) versão digital da seção transversal A-B .



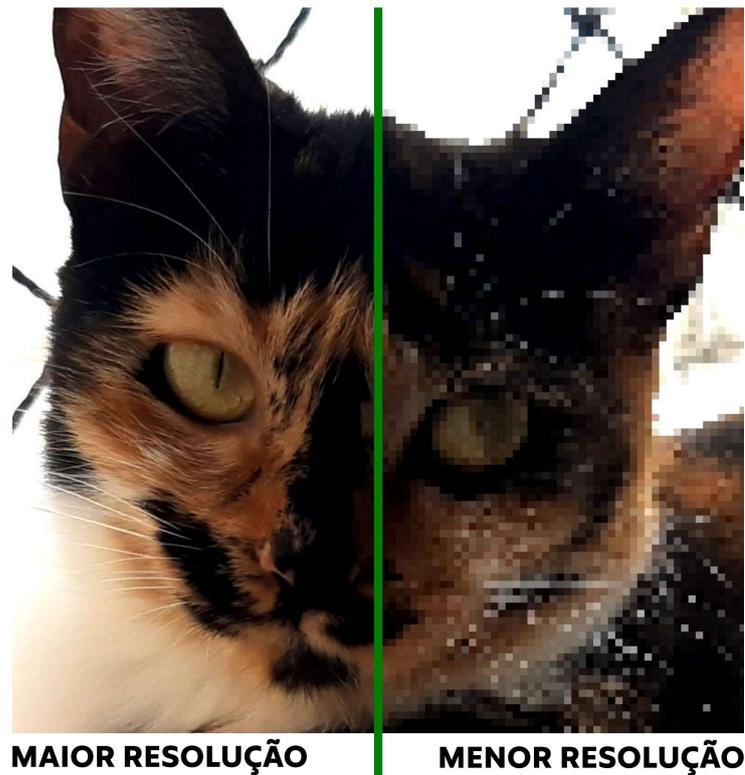
Fonte: Adaptada de [Gonzalez e Woods \(2012\)](#)

Neste caso, em que versões diferentes da mesma imagem são comparadas, as dimensões das matrizes que representam as imagens digital estão diretamente relacionadas à resolução das mesmas. Por este motivo, a versão de menor resolução será armazenada por uma matriz de menor dimensão do que a versão de maior resolução. Por fim, vale comentar que os termos “imagem de alta resolução” e “imagem de baixa resolução” são utilizados neste trabalho para se referir, respectivamente, às versões de maior e menor resolução de uma imagem digital.

### 2.1.2 Modelo de Aquisição de Imagens

O modelo de aquisição relaciona a imagem HR original às imagens de baixa resolução e é importante para definir uma formulação matemática conveniente, na qual se baseia a solução do problema de super-resolução. Segundo [Farsiu et al. \(2004\)](#), a estimação de uma imagem HR desconhecida não é baseada exclusivamente nas imagens LR observadas, mas também em suposições acerca do modelo matemático do ruído e do processo de deformação da imagem como um todo.

Figura 3 – Ilustração comparativa de diferentes resoluções. A região à esquerda da linha verde tem maior resolução e, a região à direita tem menor resolução.



Fonte: Elaborada pela autora.

Considerando a super-resolução *multi-frame*, o modelo de aquisição, ilustrado na Figura 4, é formulado como descreve a Equação 2.1. (ELAD; FEUER, 1997)

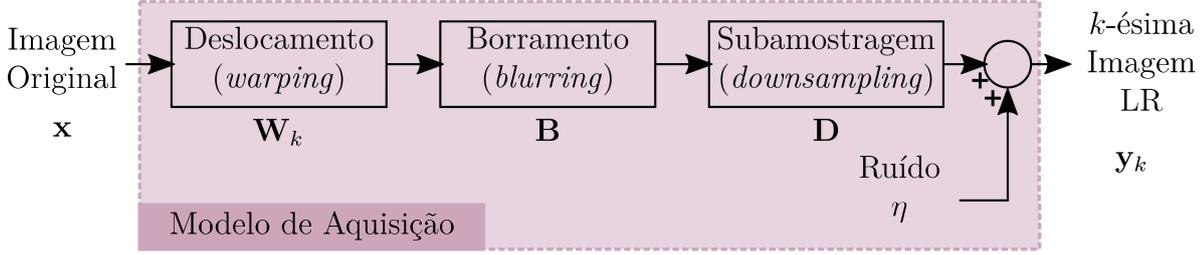
$$\mathbf{y}_k = \mathbf{DBW}_k \mathbf{x} + \eta = \mathbf{H}_k \mathbf{x} + \eta_k, \quad (2.1)$$

em que  $\mathbf{x}$  é a imagem original de alta resolução,  $\mathbf{y}_k$  é uma das  $K$  imagens de baixa resolução, ambas ordenadas lexicograficamente<sup>1</sup>, ou seja, todos os vetores envolvidos no processo são vetores-colunas.  $\mathbf{W}_k$  é o operador de deslocamento,  $\mathbf{B}$  é o operador de borramento,  $\mathbf{D}$  representa o operador de sub-amostragem e  $\eta_k$  é o ruído aditivo, também em formato lexicográfico.

O operador de borramento modela o efeito da função de espalhamento de ponto (PSF, do inglês, *point spread function*). A PSF descreve a forma como o ponto espalha sua energia ao redor dele mesmo, em um espaço bidimensional, dizendo respeito ao processo de formação da imagem dentro da câmera, mais especificamente ao fenômeno distorção óptica ilustrado na Figura 1. O operador de subamostragem atua de forma a produzir imagens LR de menor dimensão a partir da imagem HR original relacionada, de acordo com um fator de magnificação dado. Convém comentar que a ausência de tal operador transforma

<sup>1</sup> A ordenação lexicográfica consiste em transformar matrizes bidimensionais em vetores colunas, empilhando as colunas vizinhas.

Figura 4 – Esquema simplificado do modelo de aquisição.



Fonte: Elaborada pela autora.

o problema em um problema de restauração de imagens. O operador de deslocamento  $\mathbf{W}_k$ , por sua vez, descreve o deslocamento *sub-pixel* entre as imagens de baixa resolução, sendo definido pela transformação espacial das coordenadas  $\mathbf{s}_k$ , descrita na Equação 2.2.

$$\mathbf{s}_k = \begin{bmatrix} \cos \theta_k & \sin \theta_k & 0 \\ -\sin \theta_k & \cos \theta_k & 0 \\ t_{h_k} & t_{v_k} & 1 \end{bmatrix}, \quad (2.2)$$

em que  $\theta_k$  é o ângulo de rotação,  $t_{h_k}$  é a translação na direção horizontal e  $t_{v_k}$ , a translação na direção vertical da imagem LR  $k$ . Em relação ao ruído aditivo<sup>2</sup>, parte-se do pressuposto de que ele não está espacialmente conectado à imagem, ou seja, não há relação entre ele e os *pixels* (KHATTAB et al., 2020).

Considerando um modelo conjunto, no qual concatenam-se as  $K$  imagens LR (KÖHLER et al., 2016), tem-se a Equação 2.3.

$$\mathbf{y} = \mathbf{DBW}\mathbf{x} + \eta = \mathbf{H}\mathbf{x} + \eta, \quad (2.3)$$

em que  $\mathbf{y} = [\mathbf{y}_1^\top, \mathbf{y}_2^\top, \dots, \mathbf{y}_K^\top]^\top$ ,  $\mathbf{W} = [\mathbf{W}_1^\top, \mathbf{W}_2^\top, \dots, \mathbf{W}_K^\top]^\top$  e  $\eta = [\eta_1^\top, \eta_2^\top, \dots, \eta_K^\top]^\top$ .

Considerando os métodos de super-resolução *single-frame*<sup>3</sup>, o modelo de aquisição geralmente é formulado como descreve a Equação 2.4.

$$\mathbf{y} = \mathbf{DB}\mathbf{x} = \mathbf{H}\mathbf{x}, \quad (2.4)$$

em que  $\mathbf{H}$  é a matriz de deformação definida pelo modelo,  $\mathbf{x}$  e  $\mathbf{y}$  são, respectivamente, os vetores coluna obtidos através da organização lexicográfica da imagem de alta resolução  $\mathbf{X}$  e da imagem de baixa resolução  $\mathbf{Y}$  (YANG et al., 2010).

### 2.1.3 Registro

No contexto de processamento digital de imagens, registro é o processo de alinhar duas ou mais imagens de uma mesma cena, em que a transformação geométrica entre

<sup>2</sup> Alguns trabalhos de super-resolução *multi-frame* consideram outras categorias de ruído, como ruído Laplaciano aditivo ou ruído multiplicativo (NASROLLAHI; MOESLUND, 2014), entretanto, esses tipos de ruído estão fora do escopo desta tese.

<sup>3</sup> Apesar deste trabalho ser contextualizado nos métodos *multi-frame*, a abordagem híbrida proposta aqui, é baseada em um método *single-frame*, por isso essa modelagem também é apresentada neste capítulo.

as imagens é desconhecida, de forma que o problema se torna estimar tal transformação e usá-la, então, para alinhar as imagens em questão. Chamam-se imagens deslocadas aquelas que se deseja alinhar e, em contrapartida, chama-se imagem de referência a imagem em relação à qual as imagens deslocadas serão alinhadas (GONZALEZ; WOODS, 2012). A maioria dos métodos de registro são formados por quatro etapas: (i) detecção de características, (ii) correspondência de características, (iii) estimação do modelo de transformação e (iv) interpolação ou transformação das imagens. Na etapa (i) detecção de características, objetos distintos e relevantes são detectados, manual ou automaticamente, chamados pontos de controle (CP, do inglês *control points*). Vale ressaltar, entretanto, que tais pontos de controle podem ser os próprios *pixels* da imagem. A categorização dos métodos de registro, em paramétricos ou não paramétricos, depende de como as etapas (ii) e (iii) são executadas. Nos métodos paramétricos, a correspondência de característica é a etapa na qual se estabelece a relação entre as imagens deslocada e de referência, de forma que a etapa (iii) de estimação do modelo é reduzida a encontrar os parâmetros que melhor se ajustam ao modelo estabelecido. Por outro lado, nos métodos não paramétricos as etapas (ii) e (iii) são executadas simultaneamente e a correspondência entre as duas imagens pode ser estimada de forma local, baseando-se nas diferenças elemento-a-elemento entre elas (ZITOVÁ; FLUSSER, 2003). Finalmente, na etapa de interpolação, a imagem deslocada é transformada fazendo uso do deslocamento estimado pelas outras etapas, sendo alinhada à imagem de referência.

Os métodos com os quais as propostas desta tese são comparadas utilizam o método de registro paramétrico proposto por Evangelidis e Psarakis (2008). Entretanto, devido à sua eficácia tratando-se de deslocamentos *sub-pixels*, desempenho eficiente (VERCAUTEREN et al., 2007) e natureza não paramétrica, introduzem-se, nesta tese, métodos de super-resolução *multi-frame* que utilizam registro não paramétrico *Demons*.

#### 2.1.4 Registro Não Paramétrico *Demons*

O método de registro *Demons* foi proposto por Thirion (1998) a partir do conceito de modelos de difusão, análogo aos *demônios de Maxwell*, introduzidos no século 19 por James Clerk Maxwell como um experimento teórico acerca do problema de um gás composto por dois tipos de partículas, de diferentes temperaturas, separadas por uma membrana semi-permeável (KNOTT, 1911). Com o passar do tempo, as partículas tendem a se misturar até que a temperatura se equilibre. Entretanto, o pressuposto desse experimento é o de que haja um pequeno ser (o chamado demônio de Maxwell) que pode controlar uma pequena porta localizada na membrana e, sabendo a localização inicial de cada partícula, abrir e fechar a porta para que elas retornem às suas posições iniciais, separando-se novamente.

Análoga a tal conceito, a proposta de Thirion (1998) consiste em considerar as fronteiras dos objetos da imagem fixa como uma membrana semi-permeável e permitir

à imagem deslocada, pressuposta como uma grade deformável, difundir-se através de tal interface a partir da ação de agentes (tal qual os demônios de Maxwell) situados na membrana. Uma vez que algumas implementações diferentes são propostas por Thirion (1998), neste trabalho utiliza-se a implementação iterativa descrita no Algoritmo 1, feita em um esquema multi-escalar de  $L$  níveis (Linhas 3, 4 e 5). A operação descrita na Linha 7 consiste em uma interpolação da imagem deslocada que pode ser compreendida como o movimento das partículas em direção às suas posições iniciais, ou seja, os *pixels* da imagem, de forma que o agente, nesse caso, é a função interpoladora. A informação sobre a localização de cada *pixel* é dada pelo cálculo descrito pela Equação 2.5.

$$\Delta \mathbf{d}^{(i)}(u, v) = -\frac{(\mathbf{y}_k^{(i)}(u, v) - \mathbf{y}_r(u, v)) \nabla \mathbf{y}_r(u, v)}{(\nabla \mathbf{y}_r(u, v))^2 + (\mathbf{y}_k^{(i)}(u, v) - \mathbf{y}_r(u, v))^2}, \quad (2.5)$$

na qual  $\mathbf{y}_r$  é a imagem de referência,  $\mathbf{y}_k^{(i)}$  é imagem deslocada na iteração atual,  $\nabla \mathbf{y}_r$  é o gradiente da imagem de referência,  $\Delta \mathbf{d}^{(i)}$  é a atualização aditiva do campo denso de deslocamento e  $(u, v)$  são as coordenadas de um determinado *pixel*. Destaca-se, portanto, que o cálculo do campo de deslocamento é baseado na diferença elemento-a-elemento entre as imagens e no gradiente da imagem de referência, o que garante a natureza local do registro *Demons*. Após a atualização do campo de deslocamento, um filtro Gaussiano é aplicado ao mesmo (Linha 10) e assim sucessivamente até finalizar as iterações bem como os níveis do esquema multi-escala da implementação. Finalmente, a imagem deslocada é interpolada com o uso do campo de deslocamento, gerando a versão alinhada  $\tilde{\mathbf{y}}_k$ .

---

**Algoritmo 1:** Registro *Demons*


---

**Entrada:** Imagem de referência  $\mathbf{y}_r$ , imagem deslocada  $\mathbf{y}_k$ , parâmetro  $\sigma_{\text{diff}}$ , nível máximo de pirâmide  $L$  e número máximo de iterações

**Saída:** Versão alinhada da imagem deslocada,  $\tilde{\mathbf{y}}_k$

- 1 Inicializa  $\mathbf{d} = \mathbf{0}$  e  $j = 1$
  - 2 Calcula gradiente de  $\mathbf{y}_r$ ,  $\nabla \mathbf{y}_r$
  - 3 **enquanto**  $j \leq L$  **faça**
  - 4     Subamostra  $\mathbf{y}_r$  e  $\mathbf{y}_k$  usando a escala  $0.5^{L-j}$
  - 5     Calcula  $\mathbf{d}$  para que se ajuste às dimensões de  $\mathbf{y}_r$  e  $\mathbf{y}_k$
  - 6     **enquanto** não alcança número máximo de iterações **faça**
  - 7         Calcula  $\mathbf{y}_k^{(i)}$ , usando interpolação bilinear em  $\mathbf{y}_k^{(i-1)}$  aplicando  $\mathbf{d}$
  - 8         Calcula  $\Delta \mathbf{d}^{(i)}$  de acordo com a Equação 2.5
  - 9         Executa atualização aditiva  $\mathbf{d} = \mathbf{d} + \Delta \mathbf{d}^{(i)}$
  - 10        Aplica regularização no campo de deslocamento  $\mathbf{d}$  usando filtro Gaussiano definido por  $\sigma_{\text{diff}}$
  - 11     **fim**
  - 12     Atualiza nível da pirâmide  $j = j + 1$
  - 13 **fim**
  - 14 Interpola imagem deslocada usando  $\mathbf{d}$  gerando versão alinhada  $\tilde{\mathbf{y}}_k$
- 

O registro *Demons* é computacionalmente eficiente devido ao esquema multi-escalar adotado e robusto devido à regularização Gaussiana, o que é especialmente benéfico dado

que o modelo de aquisição explicado na Seção 2.1.2 considera ruído aditivo (VERCAUTEREN et al., 2007).

### 2.1.5 Super-Resolução Bayesiana Regularizada

As abordagens de super-resolução Bayesiana regularizada fazem uso de conhecimento *a priori*, que representam crenças subjetivas sobre características da formação das imagens, como uma tentativa de estabilizar o processo de inversão, compensar a falta de informações das imagens LR, além de contribuir para a remoção de artefatos e melhorar a taxa de convergência (FARSIU et al., 2004), (KATSAGGELOS; MOLINA; MATEOS, 2007), (YUE et al., 2016), (MOURABIT et al., 2017), (KHATTAB et al., 2020).

Considerando que imagem HR tem dimensão  $M \times N = sm \times sn$  e a imagem LR,  $m \times n$ , sendo  $s(s > 1)$  a escala de magnificação da super-resolução, as dimensões dos vetores e matrizes presentes no modelo de aquisição (Equação 2.3) são dadas na Tabela 1 em que se observa que a matriz  $\mathbf{H}$  é quadrada quando  $s^2 = K$ . Para que isso seja garantido, o número de imagens LR deve ser o mesmo que o quadrado da magnificação. Entretanto, isso não é suficiente para garantir a existência de uma única solução para o problema, que é mal posto (NASROLLAHI; MOESLUND, 2014).

Tabela 1 – Dimensão da matrizes e vetores presentes no modelo de aquisição

	$\mathbf{y}$	$\mathbf{x}$	$\mathbf{H}$	$\eta$
Dimensão	$Kmn \times 1$	$s^2mn \times 1$	$Kmn \times s^2mn$	$Kmn \times 1$

Fonte: Elaborada pela própria autora.

Sob a perspectiva Bayesiana, a sequência de imagens de baixa resolução, bem como a imagem de alta resolução, são modeladas como variáveis aleatórias (ELAD; FEUER, 1997). Então, segundo a Regra de Bayes, tem-se a Equação 2.6.

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}), \quad (2.6)$$

tomando o logaritmo de  $p(\mathbf{y}|\mathbf{x})$  e  $p(\mathbf{x})$ , a maximização da função  $p(\mathbf{x}|\mathbf{y})$  em relação a  $\mathbf{x}$  resulta na Equação 2.7

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \left[ \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}) \right], \quad (2.7)$$

assumindo ruído  $\eta$  da Equação 2.3 como sendo aditivo gaussiano e branco (ou seja, AWGN, do inglês *Additive White Gaussian Noise*) e reescrevendo a probabilidade *priori* ( $p(\mathbf{x})$ ) na forma de Gibbs

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \left[ \frac{1}{C_1} \exp\left(\frac{-\|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2}{2\sigma_x^2}\right) + \frac{1}{C_2} \exp(-R\mathbf{x}) \right], \quad (2.8)$$

em que  $C_1$  e  $C_2$  são constantes,  $\sigma_x^2$  é a variância do ruído e  $R(\mathbf{x})$  é a função energia da *priori* (NASROLLAHI; MOESLUND, 2014). A equação pode ser reescrita como mostra a Equação 2.9.

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left[ F(\mathbf{x}) + \lambda R(\mathbf{x}) \right] = \arg \min_{\mathbf{x}} \left[ \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R(\mathbf{x}) \right], \quad (2.9)$$

em que  $R(\mathbf{x})$  é o termo de regularização associado a uma *priori*, que diminui o espaço de possíveis soluções e a influência de distúrbios indesejados. A influência da regularização é ponderada pelo fator  $\lambda$ , e  $F(\mathbf{x}) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$  é denominada fidelidade. A solução para a Equação 2.9 pode ser encontrada de forma iterativa, aplicando métodos de otimização como o gradiente descendente.

A fidelidade age como função custo que mede o resíduo entre as imagens LR e a imagem HR estimada – por isso, deve ser minimizada (ZENG; YANG, 2013). O modelo matemático aplicado ao termo de fidelidade relaciona-se ao tipo de ruído considerado na imagem de baixa resolução. Por outro lado, a *priori* é escolhida de forma a acentuar uma dada característica, tal como elementos de borda ou, ainda, com o objetivo de remover ruído.

#### 2.1.5.1 Bilateral Total Variation - BTM

Uma das funções de regularização mais referenciadas é a função de custo Tikhonov (NASROLLAHI; MOESLUND, 2014)(ELAD; FEUER, 1997) que contém filtro passa-altas como o operador derivativo, laplaciano ou até mesmo uma matriz identidade. A ideia por trás de tal método de regularização é limitar a energia total da imagem (através da matriz identidade) ou forçar suavização espacial (através do operador derivativo ou Laplaciano). Uma vez que *pixels* contaminados por ruído e de borda contêm energia de alta frequência, ambas categorias de *pixel* serão removidas no processo de regularização e a imagem resultante não terá bordas pronunciadas.

Proposto originalmente por Farsiu et al. (2004), o método denominado BTM (do inglês, *Bilateral Total Variation*) apresenta a *priori* modelada de forma a ressaltar bordas na imagem HR e é baseado no método *Total Variation* (TV) (RUDIN; OSHER; FATEMI, 1992), que penaliza variações da intensidade dos *pixels* da imagem, medidas pela norma -  $\ell_1$  da magnitude do seu gradiente, sendo um método de regularização eficaz em aplicações de *denoising* e *deblurring* (FARSIU et al., 2004). A função de regularização BTM é derivada aplicando-se um filtro bilateral à regularização TV resultando em na Equação 2.10.

$$R(\mathbf{x}) = \Gamma_{\text{BTM}}(\mathbf{x}) = \sum_{m=-P}^P \sum_{n=-P}^P \alpha_0^{|m|+|n|} \|\mathbf{x} - \mathbf{S}_m^v \mathbf{S}_n^h \mathbf{x}\|_1, \quad (2.10)$$

para  $P \geq 1$  e  $\alpha_0 \in ]0, 1]$ .  $\mathbf{S}_m^v$  e  $\mathbf{S}_n^h$  são as derivadas parciais vertical e horizontal, respectivamente. Uma janela de dimensão  $(2P + 1) \times (2P + 1)$  define uma região sobre

a qual as derivadas vertical  $\mathbf{S}_m^v$  e horizontal  $\mathbf{S}_n^h$  são calculadas ( $m$  e  $n$  representam a localização do passo da varredura para o cálculo dessas derivadas) e  $\alpha_0$  é um fator de valor entre 0 e 1, tipicamente escolhido de forma empírica. A regularização descrita pelo termo  $\Gamma_{\text{BTV}}$  preserva bordas porque não penaliza de forma severa as bruscas mudanças locais de gradiente. A Figura 5 exemplifica o resultado da super-resolução quando utilizado o método BTV.

Para fins de conveniência, a Equação 2.10 pode ser reescrita como (KÖHLER et al., 2016) a Equação 2.11.

$$R(\mathbf{x}) = \Gamma_{\text{BTV}}(\mathbf{x}) = \sum_{m=-P}^P \sum_{n=-P}^P \|\mathbf{S}^{m,n} \mathbf{x}\|_1 = \|\mathbf{S} \mathbf{x}\|_1, \quad (2.11)$$

na qual a matriz  $\mathbf{S} \in \mathbb{R}^{N \times N'}$ , para  $N' = (2P + 1)^2 N$ , define uma transformação linear dada por

$$\mathbf{S} = (\mathbf{S}^{-P,-P} \mathbf{S}^{-P+1,-P} \dots \mathbf{S}^{P,P})^\top, \quad (2.12)$$

para  $\mathbf{S}^{m,n} = \alpha_0^{|m|+|n|} (\mathbf{I}_{N \times N} - \mathbf{S}_m^v \mathbf{S}_n^h)$ . A imagem super-resolvida  $\hat{\mathbf{x}}$ , então, é estimada resolvendo-se o seguinte problema de otimização:

$$\hat{\mathbf{x}} = \arg \min_x \left[ \|\mathbf{H} \mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{S} \mathbf{x}\|_1 \right]. \quad (2.13)$$

Figura 5 – Exemplo de imagem super-resolvida utilizando o método BTV em que (a) é a imagem LR de entrada e (b) é a imagem HR estimada.



Fonte: Elaborada pela autora.

### 2.1.5.2 Iteratively Re-Weighted Super-Resolution - IRWSR

Köhler et al. (2016) propuseram uma versão generalizada da regularização BTV – com o objetivo de aprimorar seu desempenho quanto à preservação de borda – e um termo de fidelidade adaptativa a fim de ponderar a influência de cada *pixel*. O método proposto estima, simultaneamente, os parâmetros do modelo e a imagem super-resolvida utilizando

um esquema de pirâmide *coarse-to-fine*. Assim, modifica-se o termo regularizador  $R(x)$  da Equação 2.9, de modo que a nova função de regularização seja formulada como

$$R(\mathbf{x}, \boldsymbol{\alpha}) = \Gamma_{\text{IRWSR}} = \|\mathbf{M}_{\boldsymbol{\alpha}} \mathbf{S} \mathbf{x}\|_1 \quad (2.14)$$

em que  $\mathbf{M}_{\boldsymbol{\alpha}} = \text{diag}(\alpha_1, \dots, \alpha_{N'})$  com  $N' = (2P + 1)^2(MN)$ , em que  $M \times N$  é a dimensão da imagem de alta resolução. A matriz  $\mathbf{M}_{\boldsymbol{\alpha}}$  é construída como uma matriz diagonal a partir dos valores de  $\alpha_i$ . O peso  $\alpha_i \in [0, 1]$  associado à *priori* possibilita que a regularização seja localmente adaptável de forma a aprimorar a reconstrução das descontinuidades (por exemplo, regiões de bordas). Tais pesos devem ser estimados de forma que  $\alpha_i = 0$  para descontinuidades diminui-se a influência da regularização e para  $\alpha_i = 1$  aumenta-se a influência da regularização, com o objetivo de suavizar regiões homogêneas. Além disso, o fator de regularização  $\lambda$  é selecionado automaticamente de forma a melhorar o resultado no que tange à presença de ruído na imagem final. Se o valor de  $\lambda$  for menor que o ideal, a super-resolução se torna mal condicionada e é afetada por ruído residual, caso contrário, se o valor for maior, a imagem super-resolvida torna-se borrada (KÖHLER et al., 2016).

A definição do termo de fidelidade, apresentado na Equação 2.9, geralmente parte do pressuposto de que o ruído é invariante no espaço, o que pode atingir resultados não ótimos para todos os *pixels*. O método IRWSR parte do pressuposto de que, em aplicações reais, o ruído segue distribuição mista. Nesse sentido, propõe-se uma distribuição gaussiana localmente ponderada definida por meio de um mapa de confiança  $\mathbf{M}_{\boldsymbol{\beta}} = \text{diag}(\beta_1, \dots, \beta_{Kmn})$ , no qual há um escalar  $\beta_j$  para cada *pixel* das imagens LR. Os pesos  $\beta_j \in [0, 1]$  modelam a confiança de cada *pixel*, controlando a influência de cada um deles na estimação final. Assim,  $\beta_j = 0$  indica que dado pixel não é bem descrito pela distribuição gaussiana sendo, portanto, considerado um *outlier*; Por outro lado,  $\beta_j = 1$  indica um pixel bem descrito pela distribuição. Associando o mapa de confiança  $\mathbf{M}_{\boldsymbol{\beta}}$  e o termo de regularização  $\Gamma_{\text{IRWSR}}$  à Equação 2.9, a imagem super-resolvida é dada por

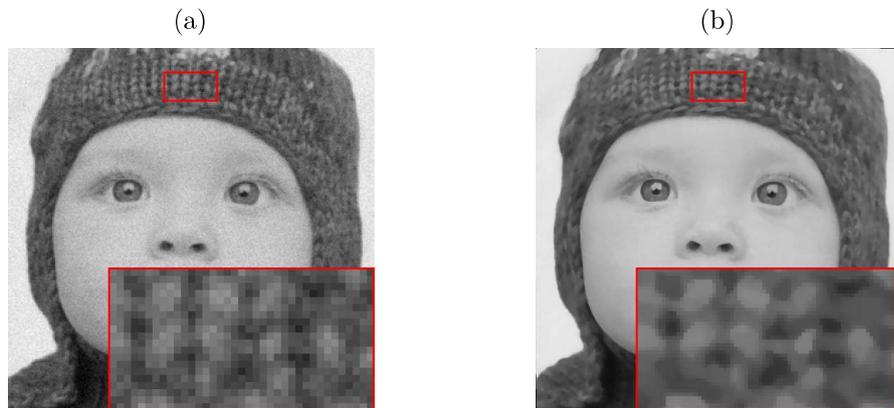
$$\hat{\mathbf{x}} = \arg \min_x \left[ (\mathbf{y} - \mathbf{H} \mathbf{x})^\top \mathbf{M}_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{H} \mathbf{x}) + \lambda \|\mathbf{M}_{\boldsymbol{\alpha}} \mathbf{S} \mathbf{x}\|_1 \right]. \quad (2.15)$$

As matrizes  $\mathbf{M}_{\boldsymbol{\alpha}}$  e  $\mathbf{M}_{\boldsymbol{\beta}}$ , bem como o hiperparâmetro  $\lambda$ , são tratados como variáveis latentes inferidas juntamente à imagem super-resolvida. A Figura 6 exemplifica o resultado da super-resolução quando utilizado do método IRWSR.

### 2.1.6 Super-Resolução Baseada em Aprendizado de Máquinas

Durante o processo de subamostragem de uma imagem, as informações de alta frequência, muitas vezes necessárias para a compreensão total da imagem, se perdem. A abordagem analítica e genérica, geralmente proposta nos trabalhos de super-resolução *multi-frame*, regulariza a solução mas não recupera, necessariamente, tais frequências perdidas (NASROLLAHI; MOESLUND, 2014). Dessa forma, os métodos baseados em aprendizado

Figura 6 – Exemplo de imagem super-resolvida utilizando o método IRWSR em que (a) é a imagem LR de entrada e (b) é a imagem HR estimada.

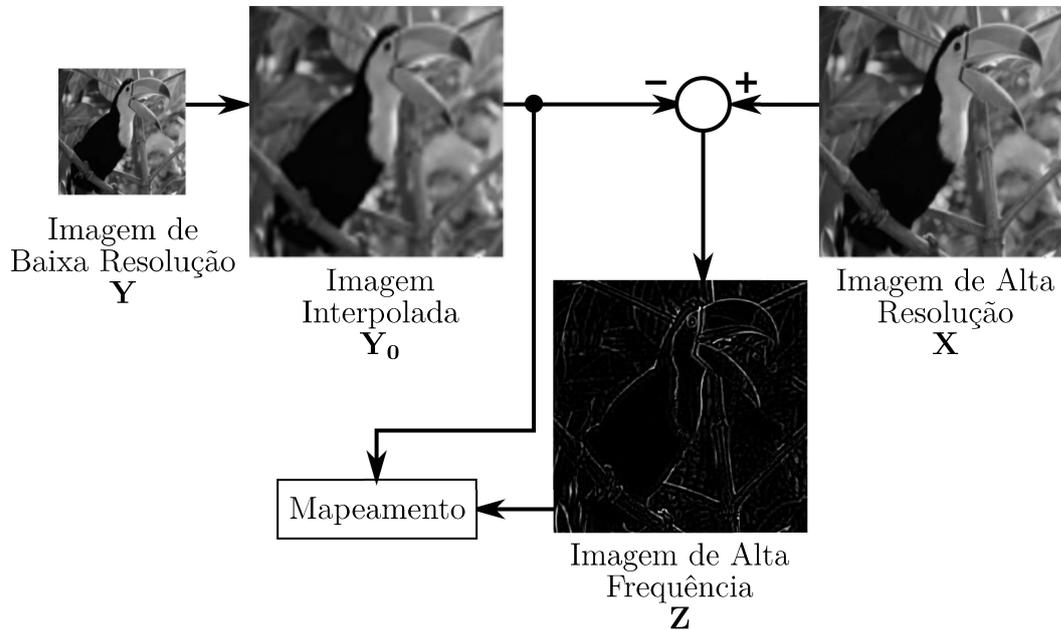


Fonte: Elaborada pela autora.

de máquinas utilizam bancos de dados externos para, a partir deles, aprender a relação entre  $\mathbf{Y}$  e  $\mathbf{X}$ . Esse aprendizado é aplicado, então, às imagens de baixa resolução, na etapa de teste. Entretanto, tal relação não é aprendida de forma direta, uma vez que uma primeira interpolação (por exemplo, a interpolação bicúbica) pode ser executada, resultando em uma imagem estimada inicial - denotada, aqui, como  $\mathbf{Y}_0$  (FREEMAN; JONES; PASZTOR, 2002). Tal imagem é subtraída da imagem original de alta resolução  $\mathbf{X}$ , resultando em uma nova imagem - denotada  $\mathbf{Z}$  - que representa o resíduo da interpolação executada. Portanto, o banco de dados externos é utilizado para mapear a relação entre a imagem estimada inicial ( $\mathbf{Y}_0$ ) e a imagem residual ( $\mathbf{Z}$ ). A Figura 7 mostra um esquema simplificado de tal procedimento. É importante ressaltar que o procedimento explicado neste parágrafo refere-se ao treinamento da arquitetura utilizada e, por isso, as imagens originais de alta resolução estão disponíveis. Para o teste, portanto, outro banco de dados, não usado para treinamento, é considerado.

Parte-se do pressuposto de que existem regularidades em imagens digitais, e tais regularidades podem ser extraídas de um banco de dados externo bem escolhido. A fim de melhorar a capacidade de generalização do conjunto externo, trabalha-se com pequenas regiões das imagens, ou *patches* (FREEMAN; JONES; PASZTOR, 2002). Cada imagem do banco de dados é deformada de acordo com o modelo de aquisição descrito pela Equação 2.4, formando assim, um conjunto de pares de *patches*. Cada par consiste em um *patch* de alta resolução e o *patch* de baixa resolução relacionado. Vale ressaltar que o problema de super-resolução é um problema de regressão, do ponto de vista de aprendizado de máquinas, em que objetiva-se estimar o valor da intensidade de cinza de cada *pixel* do *patch*. Dado um conjunto externo muito grande, entretanto, a busca pelo par correspondente ao *patch* de entrada pode ser muito custosa. Dessa forma, Li et al. (2009) sugerem uma classificação

Figura 7 – Esquema simplificado do funcionamento de um método de Super-resolução *single-frame* baseado em aprendizado, quanto à etapa de aprendizado com exemplos externos.



Fonte: Adaptada de Cosmo, Inaba e Salles (2017).

dos *patches*, sendo possível treinar regressores especializados para cada *cluster*<sup>4</sup>.

### 2.1.6.1 Estatísticas do Gradiente

A classificação dos *patches* da Imagem Interpolada  $Y_0$  é baseada em características dadas por análise do gradiente, obtida a partir do cálculo de autovalores e autovetores (COSMO et al., 2018). A análise do gradiente foi proposta por XiaoGuang Feng e Milanfar (2002) e retorna três características: a orientação predominante, a força média e a coerência dos gradientes em um *patch* de dimensão  $d_g \times d_g$ . A força mede a intensidade média dos gradientes e a coerência mede a exatidão da orientação predominante do gradiente no *patch*. Regiões com bordas marcantes e na mesma direção possuem alta força e coerência. Baixa coerência e força indicam regiões com textura fina ou falta de estrutura, por exemplo, regiões homogêneas. Alta força e baixa coerência são típicas de regiões com bordas curvas ou estruturas multidirecionais.

O cálculo da orientação pode ser formulado como a busca de um vetor unitário  $\hat{\mathbf{a}}$  que minimize a média dos ângulos entre  $\mathbf{a}$  e os vetores gradiente  $\mathbf{g}_i = \nabla f(h_i, v_i), i = 1, \dots, d^2$ .

<sup>4</sup> Um *cluster*, neste contexto, é um grupo de *patches* que tem informações similares, segundo a classificação aplicada. Como para o termo *patch*, utiliza-se o termo em inglês para manter coerência com a literatura.

O ângulo entre os dois vetores é dado pela Equação 2.16

$$\theta_i = \cos^{-1} \left( \frac{\langle \mathbf{a}, \mathbf{g}_i \rangle}{\|\mathbf{a}\|_2 \|\mathbf{g}_i\|_2} \right), \quad (2.16)$$

em que  $\langle, \rangle$  é o produto interno. Dessa forma,  $\hat{\mathbf{a}}$  pode ser calculado como

$$\hat{\mathbf{a}} = \arg \max_{\mathbf{a}} \sum_{i=1}^{d_g^2} \langle \mathbf{a} \cdot \mathbf{g}_i \rangle^2 = \arg \max_{\mathbf{a}} \mathbf{a}^T \mathbf{Q} \mathbf{a}, \quad (2.17)$$

em que,

$$\mathbf{Q} = \begin{bmatrix} \sum_{i=1}^{d_g^2} \mathbf{g}_h^{(i)} \mathbf{g}_h^{(i)} & \sum_{i=1}^{d_g^2} \mathbf{g}_h^{(i)} \mathbf{g}_v^{(i)} \\ \sum_{i=1}^{d_g^2} \mathbf{g}_v^{(i)} \mathbf{g}_h^{(i)} & \sum_{i=1}^{d_g^2} \mathbf{g}_v^{(i)} \mathbf{g}_v^{(i)} \end{bmatrix}.$$

O vetor  $\mathbf{a}$ , que maximiza 2.17, é o autovetor de  $\mathbf{Q}$  que corresponde ao seu maior autovalor ( $\lambda_1$ ). A direção de  $\mathbf{a}$  é a orientação predominante do gradiente naquela região.

Para calcular a matriz  $\mathbf{Q}$ , os vetores gradientes  $\mathbf{g}_i$  são concatenados, formando a matriz  $\mathbf{G}$

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_h^{(1)} & \mathbf{g}_v^{(1)} \\ \mathbf{g}_h^{(2)} & \mathbf{g}_v^{(2)} \\ \vdots & \vdots \\ \mathbf{g}_h^{(d^2)} & \mathbf{g}_v^{(d^2)} \end{bmatrix}, \quad (2.18)$$

e  $\mathbf{Q} = \mathbf{G}^T \mathbf{G}$ .

A força média é dada pela raiz quadrada do maior autovalor de  $\mathbf{Q}$ , ou seja,  $\sqrt{\lambda_1}$ . A coerência dos gradientes varia entre 0 e 1 e é dada por

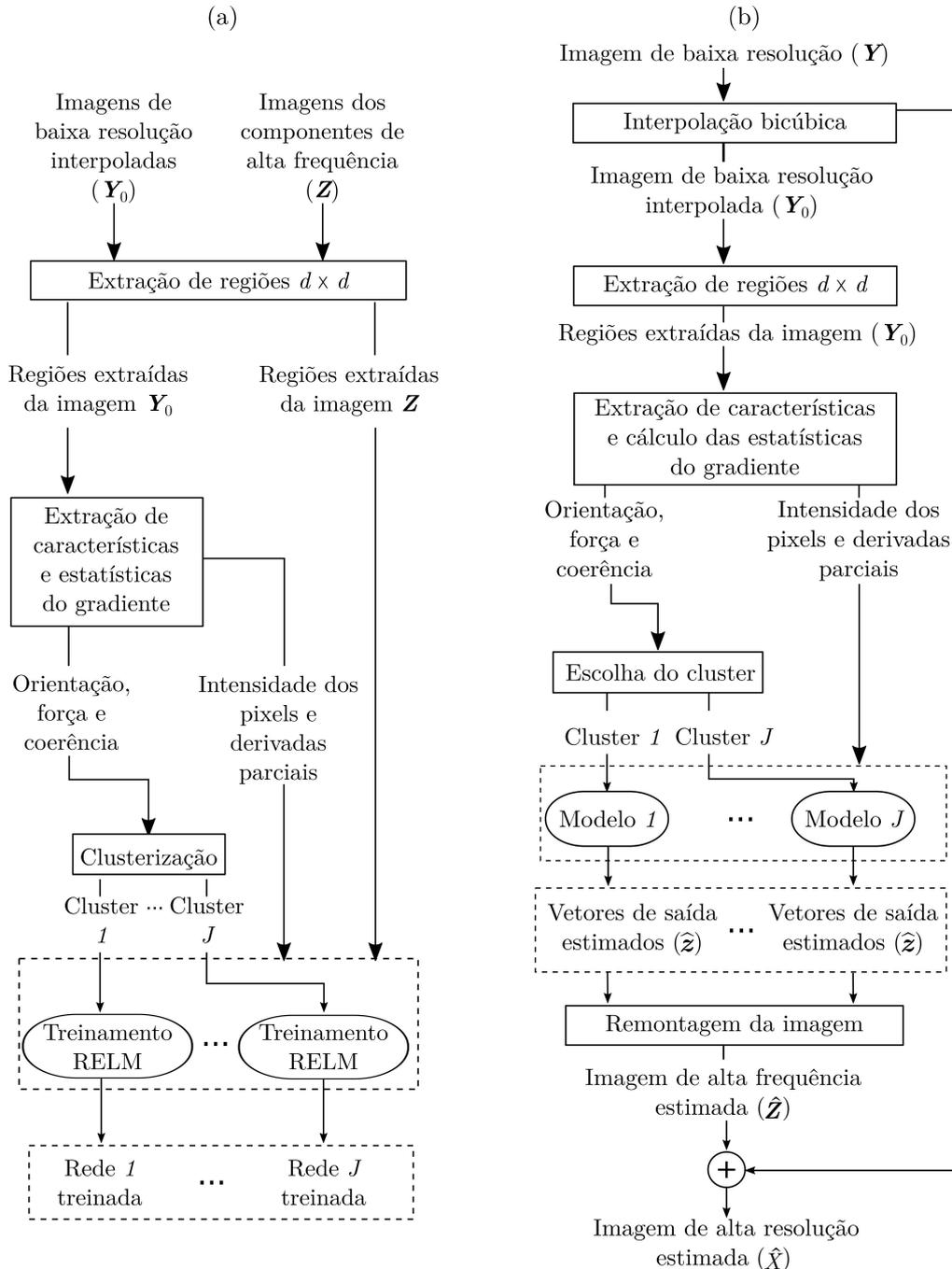
$$\mu = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1} + \sqrt{\lambda_2}}, \quad (2.19)$$

em que  $\lambda_1$  e  $\lambda_2$  são os autovalores de  $\mathbf{Q}$ .

### 2.1.6.2 Super-resolução Usando Múltiplos Modelos de Rede Neural

Cosmo, Inaba e Salles (2017) propõem a utilização de  $J$  clusters, definidos de acordo com a informação geométrica extraída dos pares de *patches* do banco de dados externo; e a aplicação de uma rede neural com uma camada oculta (do inglês, *Single-Layer FeedForward Network*, SLFN) para cada *cluster* definido. Da mesma forma como é feito por Romano, Isidoro e Milanfar (2017), as estatísticas do gradiente são utilizadas para determinar os *clusters*, distribuídos em um histograma 3D, em que cada *bin* é definido pelas respectivas coerência, força e orientação. Portanto, o número de *bins* também define o número de SLFNs treinadas. Uma vez que o agrupamento é executado, os múltiplos modelos de rede neural são treinados. Enquanto as estatísticas do gradiente são utilizadas para definir os grupos, as derivadas parciais de primeira e segunda ordens são dadas como entradas para as redes, juntamente com o valor das intensidades dos *pixels* da região. A

Figura 8 – Fluxograma das etapas de (a) treinamento e (b) reconstrução do método de SR *single-frame* baseado em exemplos com agrupamento usando informações do gradiente.



Fonte: Adaptada de Cosmo et al. (2018).

Figura 8a esquematiza esse processo de agrupamento e treinamento e a Figura 8b mostra a etapa de reconstrução de forma análoga à etapa de treinamento.

A Figura 9 mostra a etapa de reconstrução do ponto de vista de um *pixel* arbitrário. Para cada *pixel* são definidos dois *patches*, um de dimensão  $d_g \times d_g$  e outro de dimensão

$d_c \times d_c$ , e ambos são construídos a partir do *pixel* central e seus vizinhos. Os dois *patches* têm o mesmo *pixel* central, de forma que, se  $d_c = d_g$ , os *patches* são equivalentes. As estatísticas do gradiente, calculadas a partir da região  $d_g \times d_g$ , definem o modelo de rede neural a ser utilizado. No exemplo ilustrado na Figura 9, dos  $J$  possíveis modelos, o modelo 2 será utilizado, destacado pela cor laranja na figura. As derivadas parciais e intensidades dos *pixels* são extraídas da região  $d_c \times d_c$  e formam o vetor de características, que alimenta o modelo de rede neural escolhido. A saída dessa rede, por sua vez, é o vetor de saída estimado que contribui para a imagem de alta frequência estimada completa. Observa-se que o vetor de características tem dimensão  $6d_c^2$ ,  $\mathbf{y}_0 = [\mathbf{Y}_{0(d_c \times d_c)}, \frac{\partial \mathbf{Y}_0(i,j)}{\partial h}, \frac{\partial \mathbf{Y}_0(i,j)}{\partial v}, \frac{\partial^2 \mathbf{Y}_0(i,j)}{\partial h^2}, \frac{\partial^2 \mathbf{Y}_0(i,j)}{\partial v^2}, \frac{\partial^2 \mathbf{Y}_0(i,j)}{\partial hv}]$ , no qual  $(i, j)$  é a posição do *pixel* central do *patch*. Por outro lado, o vetor de saída  $\mathbf{z}_0$  estimado pela rede neural tem dimensão  $d_c^2$ . Tal vetor é redimensionado e posicionado sobre o *grid* de maior dimensão.

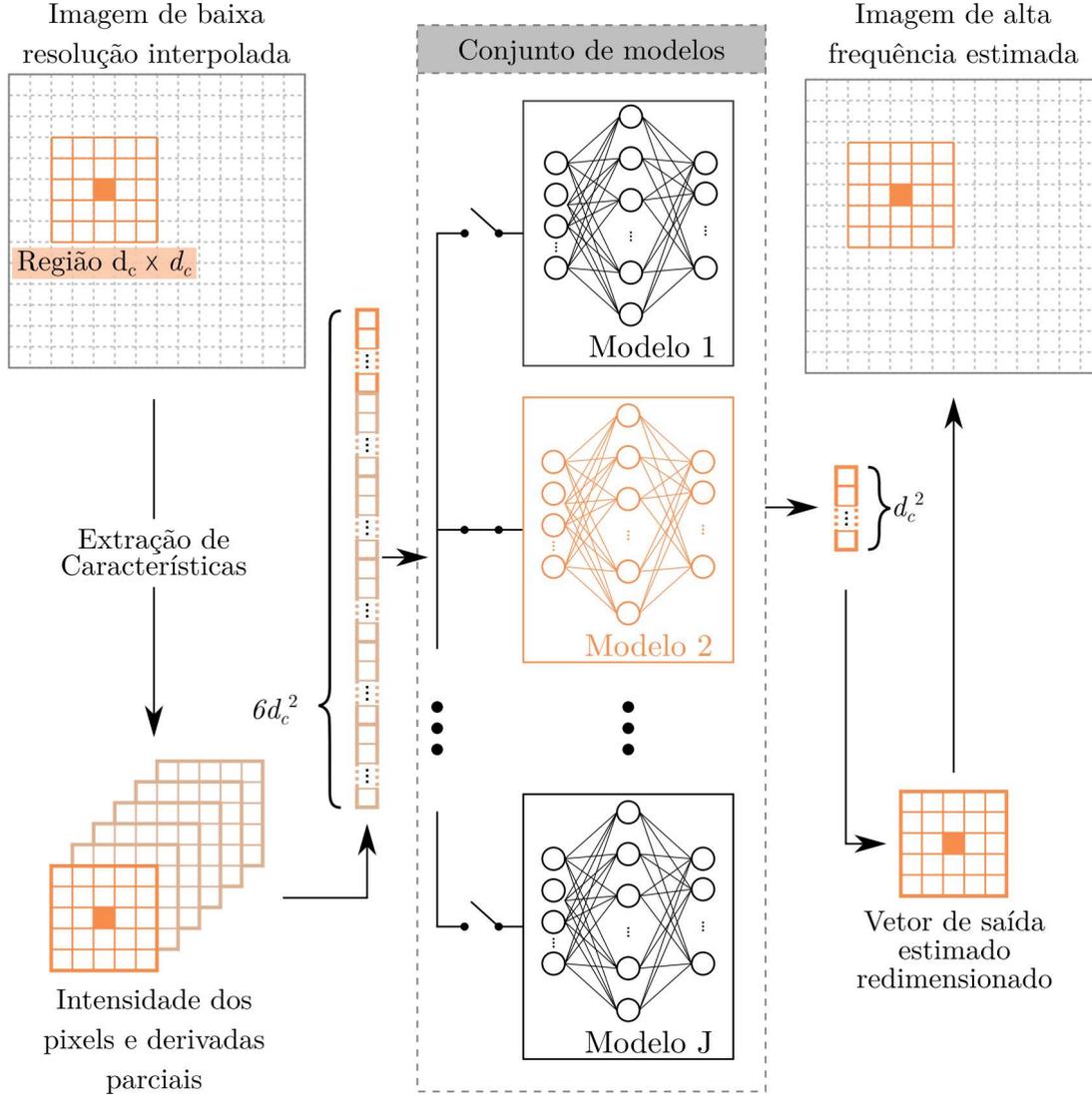
Entretanto, é importante destacar que os valores de intensidade de *pixel* desse vetor não são os valores finais da imagem estimada por esse processo. Uma vez que cada *pixel* é analisado como o centro de um *patch*, os *pixels* contribuem para o valor final dos seus vizinhos. Especificamente, cada *pixel* sofrerá a contribuição de, no máximo,  $d_c^2$  *pixels*, incluindo ele mesmo. Os *pixels* das bordas sofrem menos contribuição, pois têm menos vizinhos. Essa contribuição é calculada como a média simples do valor do *pixel* naquela posição, em todas as contribuições. Na Figura 10, é possível observar esse processo, de forma simplificada, em que o *pixel* analisado está destacado, em cor cinza, nas  $d_c^2$  contribuições e na imagem de alta frequência estimada.

### 2.1.6.3 Regularized Extreme Learning Machine

A Figura 11 mostra uma representação gráfica de uma SLFN, onde se observam a camada de entrada com  $e + 1$  neurônios, a camada oculta com  $L$  neurônios, a camada de saída, com  $s$  neurônios, bem como os pesos  $\mathbf{w}_i$  e vieses  $b_i$  entre a camada de entrada e a camada oculta e os pesos  $\beta_i$  entre a camada oculta e a camada de saída.

Métodos tradicionais, como *Backpropagation*, calculam os pesos e vieses da rede iterativamente, demandando um maior tempo de processamento. Nesse contexto de treinamento de SLFN, os autores de (DENG; ZHENG; CHEN, 2009) propuseram o método *Regularized Extreme Learning Machine* (R-ELM), uma modificação do método *Extreme Learning Machine* (ELM), proposto por Huang, Zhu e Siew (2004). O ELM é um método de treinamento de SLFN que tem como característica baixo tempo de treinamento. Tal técnica é baseada em uma solução fechada para o problema de calcular os pesos entre a camada oculta e de saída de uma SLFN, dado que os parâmetros da camada oculta são escolhidos de forma aleatória. O R-ELM, em especial, adiciona um termo regularizador que garante robustez a *overfitting*, o que garante uma melhor solução em comparação ao ELM e motiva a escolha do R-ELM (COSMO et al., 2018). Dessa forma, os pesos de saída

Figura 9 – Fluxograma do cálculo de um vetor de saída, redimensionado para dimensão  $d_c \times d_c$ , colocado sobre o *grid* da imagem de alta frequência estimada, a partir de uma região  $d_c \times d_c$  da imagem de baixa resolução interpolada.



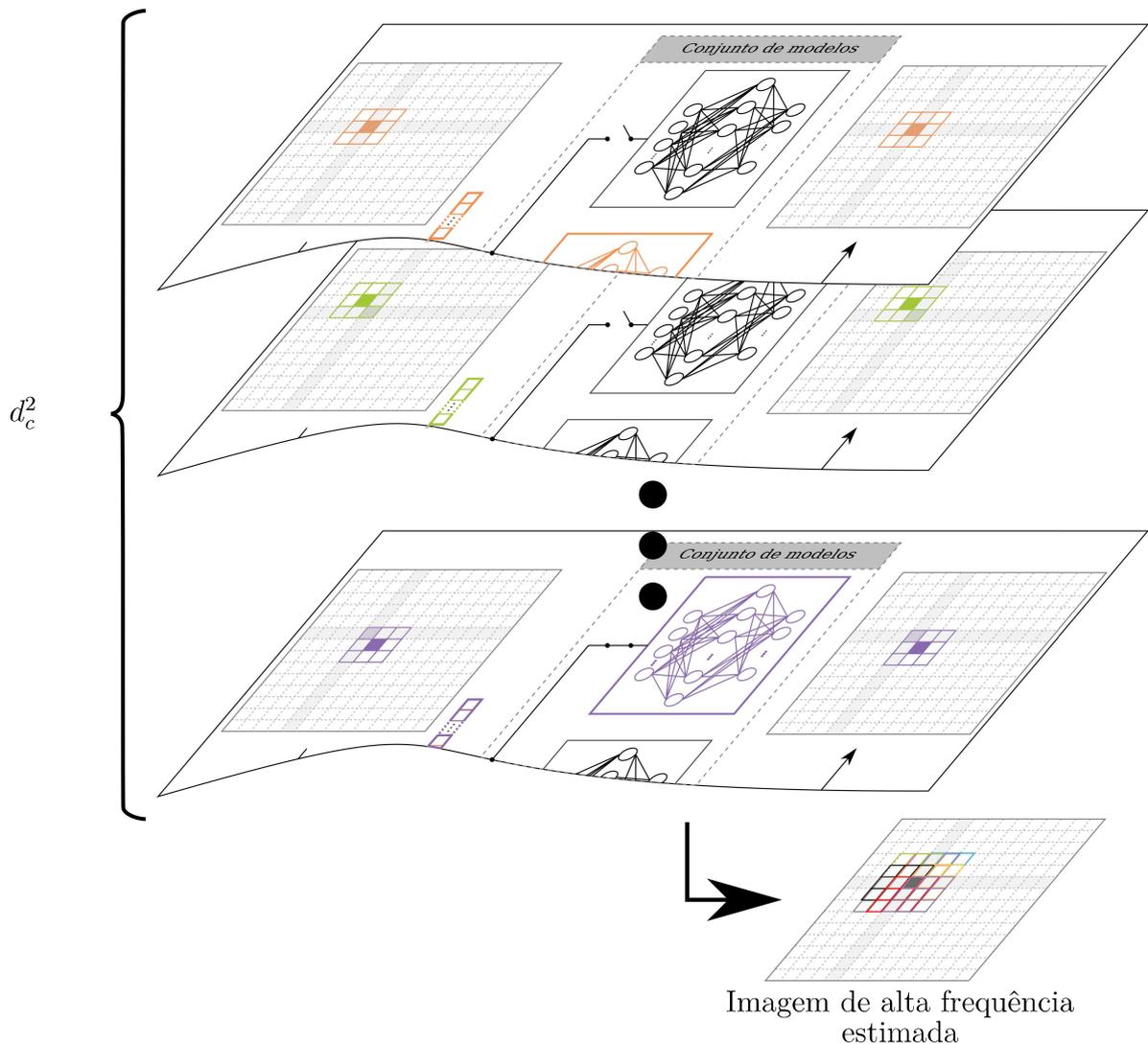
Fonte: Elaborada pela autora.

são dados pela Equação 2.20.

$$\mathbf{B}_{opt} = \left( \frac{\mathbf{I}}{C} + \mathbf{L}^T \mathbf{L} \right)^{-1} \mathbf{L}^T \mathbf{O}, \quad (2.20)$$

em que  $\mathbf{L}$  é a matriz de valores de saída da camada oculta,  $C$  é o parâmetro de regularização,  $\mathbf{B}_{opt}$  é a matriz de pesos de saída,  $\mathbf{I}$  é a matriz identidade e  $\mathbf{O}$  é a matriz de saída.

Cada *pixel* da imagem  $\mathbf{Y}_0$  gera uma amostra diferente, resultando em  $MNs^2$  amostras para uma imagem de tamanho  $Ms \times Ns$ . Entretanto, para o treinamento do R-ELM, é necessário que todas as amostras sejam carregadas no algoritmo ao mesmo tempo. Por conta de restrições de memória RAM, um subconjunto aleatório de amostras é escolhido, a partir do banco de dados de treinamento (COSMO; INABA; SALLES, 2017).

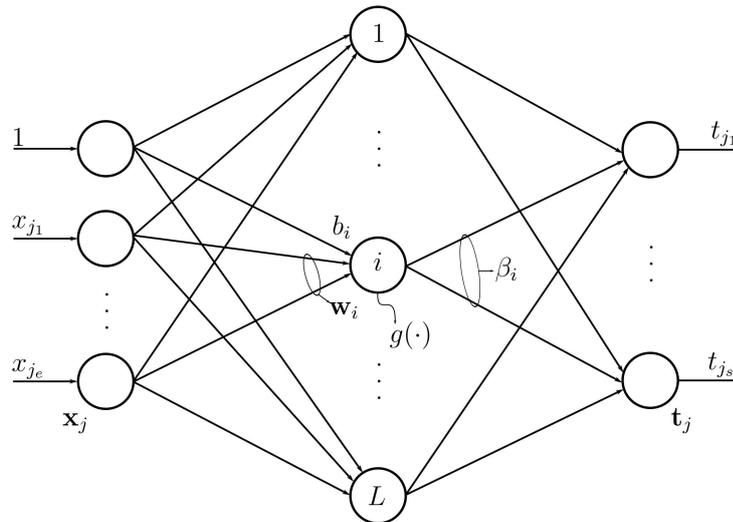
Figura 10 – Fluxograma do cálculo de um *pixel* da imagem de alta frequência estimada.

Fonte: Elaborada pela autora.

### 2.1.7 Qualidade de Imagem

Em uma aplicação real, o usuário tem uma determinada expectativa quanto à imagem melhorada, por exemplo, a leitura de uma placa de carro. Nesse caso, o *software* de super-resolução atinge seu objetivo se for possível ler (ou segmentar, em casos automáticos) a informação esperada com sucesso. Ou seja, a qualidade da super-resolução, em casos reais e isolados, é verificada por inspeção visual. Para fins de comparação de algoritmos, entretanto, além da inspeção visual faz-se necessário utilizar uma métrica de qualidade, capaz de analisar a eficiência do método de forma objetiva e quantificada. Dessa forma, simula-se o processo real, uma vez que não é possível comparar uma imagem digital – formada por uma matriz numérica – a uma cena observada pelo olho humano. Nessa simulação a imagem original é deformada pelo modelo de aquisição e, posteriormente, as

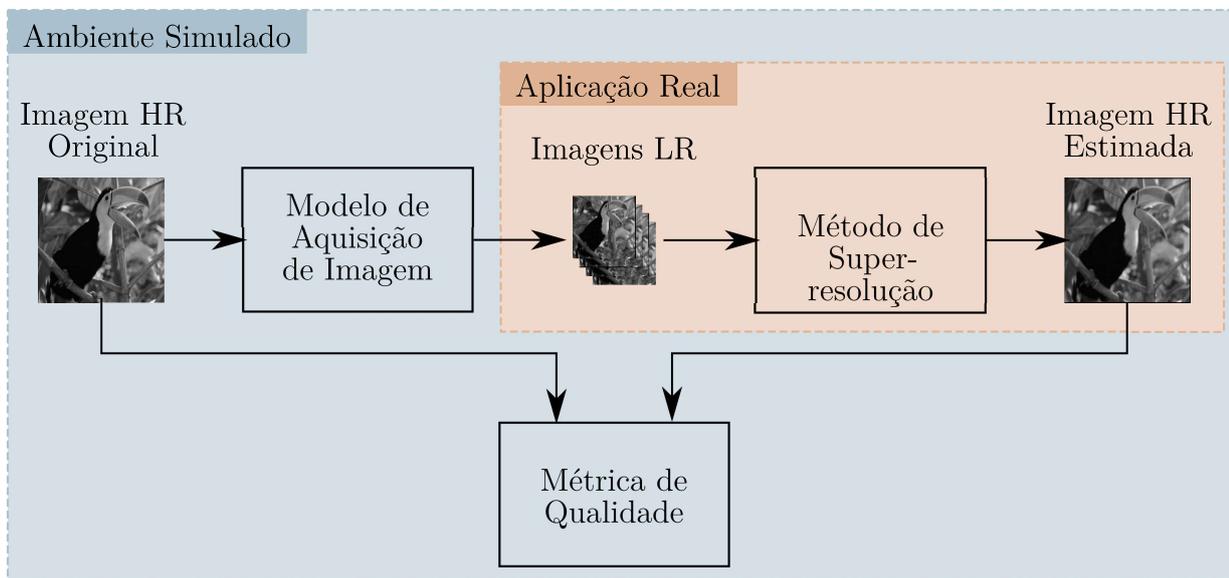
Figura 11 – Representação gráfica de uma SLFN.



Fonte: Adaptada de Inaba et al. (2018)

imagens LR são super-resolvidas, resultando em uma imagem HR estimada. Tal imagem resultante é, então, comparada com a imagem original, como esquematizado na Figura 12.

Figura 12 – Esquema simplificado do modelo de aquisição.



Fonte: Elaborada pela autora.

A escolha das métricas de qualidade objetiva se aproxima da avaliação humana. Em outras palavras, as métricas devem estimar a qualidade da imagem da mesma forma que uma pessoa o faria (WANG; BOVIK, 2006). A classificação das métricas de qualidade é dada pela disponibilidade da imagem original – que serve como referência. Quando a imagem original está disponível, as métricas são do tipo *Full Reference*. Nesses casos, compara-se a imagem estimada a ser avaliada com a imagem original. Neste trabalho, duas

métricas *Full Reference* são utilizadas: PSNR (Relação Sinal-Ruído de Pico, do inglês, *Peak Signal-to-Noise Ratio*) e SSIM (Índice de Similaridade Estrutural, do inglês, *Structural Similarity Index*).

### 2.1.7.1 PSNR

A PSNR é utilizada em muitos trabalhos nas áreas de compressão de sinal, imagem e vídeo. O cálculo dela utiliza o Erro Quadrático Médio (do inglês, *Mean Squared Error - MSE*), obtido entre a imagem avaliada e a referência, definido por

$$\text{MSE}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{1}{MN} \sum_{i=1}^{MN} [\mathbf{x}(i) - \hat{\mathbf{x}}(i)]^2, \quad (2.21)$$

em que  $\mathbf{x}$  é a imagem de referência e  $\hat{\mathbf{x}}$  é a imagem estimada a ser medida, ambas em suas formas lexicográficas e  $MN$  é o número de pixels contidos nas imagens.

O valor da PSNR é dado em decibéis e é calculado por,

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_{\mathbf{x}}^2}{\text{MSE}} \right), \quad (2.22)$$

em que  $\text{MAX}_{\mathbf{x}}$  é o valor de intensidade de cinza mais alto permitido na imagem, dado o número de bits de armazenamento de cada pixel. Para imagens com 8 bits de armazenamento,  $\text{MAX}_{\mathbf{x}} = 2^8 - 1 = 255$  (HORE; ZIOU, 2010). O valor da PSNR é dado em decibéis, devido à presença do logaritmo na Equação 2.22.

### 2.1.7.2 SSIM

Durante as três últimas décadas, muitos trabalhos estiveram focados em desenvolver uma forma de avaliação de qualidade de imagens que aproveite características conhecidas do sistema visual humano (HVS - *human visual system*) (WANG et al., 2004). A maioria dos modelos desenvolvidos tem como estratégia modificar a medida MSE de forma que os erros sejam penalizados de acordo com sua visibilidade. Wang et al. (2004) desenvolveram uma medida de similaridade estrutural – denominada SSIM (*Structural Similarity Index*) – comparando padrões locais de intensidade que são normalizados pela luminância e contraste, definida por

$$\text{SSIM}(x, \hat{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}, \quad (2.23)$$

em que  $\mu_x$  e  $\mu_{\hat{x}}$  são as médias de  $x$  e  $\hat{x}$ , respectivamente. Além disso,  $\sigma_x$  e  $\sigma_{\hat{x}}$  representam as variâncias de  $x$  e  $\hat{x}$ , respectivamente;  $\sigma_{\hat{x}x}$ , a covariância entre os dois sinais e  $c_1$  e  $c_2$  são constantes que estabilizam a divisão, no caso de um denominador próximo a zero.

O valor do SSIM varia entre -1 e 1. Assume-se 1 o valor resultante quando duas imagens idênticas são comparadas, e 0 quando duas imagens completamente diferentes são

dadas como entrada. Valores negativos correspondem a casos nos quais estruturais locais são invertidas (WANG; BOVIK; SIMONCELLI, 2005).

### 2.1.8 Testes de Hipótese

Demšar (2006) propõe o uso de testes estatísticos para comparar o desempenho de mais de dois algoritmos em múltiplos bancos de dados, no contexto de classificação, para verificar estatisticamente a melhora de desempenho dos algoritmos analisados. O autor recomenda o uso do teste de Friedman por ser não paramétrico, em conjunto com o teste de Nemenyi (este último quando necessário o uso de um teste *post-hoc*). Neste trabalho, introduz-se o uso desses testes para comparar estatisticamente os métodos de super-resolução analisados. Dessa forma, esta seção explica brevemente os procedimentos dos dois testes.

#### 2.1.8.1 Teste de Friedman

O teste de Friedman é um teste não paramétrico, no qual os algoritmos são ordenados para cada conjunto de dados separadamente, de forma que o algoritmo melhor classificado tem *rank* 1, o segundo melhor tem *rank* 2 e assim sucessivamente. Ou seja, para aplicar o teste de Friedman, os dados avaliados, que no caso da análise de métodos de super-resolução são as tabelas de valores PSNR e SSIM dos resultados, devem ser transformados em tabelas de *rank* dos algoritmos. Quando há empate, *ranks* médios são atribuídos. Em outras palavras, a hipótese avaliada pelo teste de Friedman no contexto de análise de algoritmos é dada pela questão: num conjunto de  $k$  algoritmos, pelos menos dois deles representam resultados com valores de medianas diferentes? Para responder essa pergunta, levantam-se (i) a hipótese nula, que declara que as medianas (ou, os *ranks*) de todos os algoritmos são iguais e (ii) a hipótese alternativa, que declara, por sua vez, que pelo menos dois algoritmos possuem *ranks* diferentes.

Considerando  $r_i^j \in \mathbb{R}$  o *rank* do  $j$ -ésimo algoritmo (dentre  $k$  algoritmos) para o  $i$ -ésimo conjunto de dados (dentre  $N$  conjuntos), o teste compara o *rank* médio dos algoritmos  $R_j = \frac{1}{N} \sum_i r_i^j$  por meio do cálculo da estatística de Friedman  $\chi_F^2$  :

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[ \sum_j R_j^2 - \frac{k(k+1)^2}{4} \right]. \quad (2.24)$$

Além disso, alguns autores recomendam corrigir  $\chi_F^2$  pelo número de empates que ocorrem nos dados (SHESKIN, 2011), utilizando o fator de correção de empates  $C_t$ , dado por

$$C_t = 1 - \frac{\sum_{i=1}^s (t_i^3 - t_i)}{n(k^3 - k)}, \quad (2.25)$$

em que  $s$  é o número de amostras contendo empates e  $t_i$  é o número de empates na  $i$ -ésima amostra. A estatística de Friedman corrigida é dada por

$$\chi_{F_c}^2 = \frac{\chi_F^2}{C_t}, \quad (2.26)$$

e o grau de liberdade  $k - 1$  fornecem o valor- $p$ . Define-se um valor de significância  $\alpha$ , tipicamente 5% ou 1%, e se o valor- $p$  for menor do que esse valor, diz-se que a hipótese nula pode ser rejeitada.

### 2.1.8.2 Teste de Nemenyi

Quando a hipótese nula é rejeitada, o próximo passo da análise estatística é a aplicação de um teste *post-hoc*, dentre os quais o teste de Nemenyi (NEMENYI, 1963) é usado para comparar todos os métodos entre si (DEMŠAR, 2006). Segundo o teste de Nemenyi, dois métodos são significantemente diferentes entre si se suas classificações médias diferirem em, pelo menos, a distância crítica, dada por

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \quad (2.27)$$

em que  $k$  é o número de métodos avaliados,  $N$  é o número de conjuntos que, no caso de super-resolução, é a quantidade de imagens avaliadas e  $q_\alpha$  é dado pela Tabela 2.

Tabela 2 – Valores críticos para o teste de Nemenyi bicaudal.

No. métodos	2	3	4	5	6	7	8	9	10
$q_{0,05}$	1,960	2,343	2,569	2,728	2,850	2,949	3,031	3,102	3,164
$q_{0,10}$	1,645	2,052	2,291	2,459	2,589	2,693	2,780	2,855	2,920

Fonte: Adaptada de Demšar (2006)

Além disso, os resultados do teste são tipicamente representados em um gráfico de distância crítica, como exemplificado na Figura 13, em que os algoritmos conectados pela linhas mais espessas não são significantemente diferentes entre si. Nesse exemplo ilustrativo, o método 1 é significantemente diferente do métodos 3 e 4 que, por sua vez, não são significantemente diferentes entre si. Vale ressaltar que não há transitividade nos gráficos de distância crítica, ou seja, no exemplo da Figura 13, embora tanto os métodos 1, 3 e 4 sejam equivalentes ao método 2, isso não significa que 1 seja equivalente a 3.

### 2.1.8.3 Teste de Wilcoxon Unicaudal

Os testes de Friedman e Nemenyi são utilizados, neste trabalho, nos casos em que mais de dois métodos são comparados. Por outro lado, quando comparam-se apenas dois métodos o teste de Wilcoxon Unicaudal é utilizado por empregar mais informações e, em



4. Cálculo da estatística  $z$ , de acordo com,

$$z = \frac{w - n(n+1)/4}{\sqrt{\frac{n(n+1)(2n+1)}{24} - C_t}}, \quad (2.31)$$

em que  $C_t$  é o fator de correção de empates, dado por

$$C_t = \frac{\sum_{i=1}^r t_i^3 - \sum_{i=1}^r t_i}{48}, \quad (2.32)$$

em que  $r$  é o número de posições para as quais ocorreu empate e  $t_i$  é o número de amostras empatadas na  $i$ -ésima posição;

5. Cálculo do valor- $p$ , nesse caso dado pela função cumulativa da distribuição  $\mathcal{N}(0, 1)$  em  $z$ , para cauda à esquerda e  $-z$ , para cauda à direita. Neste trabalho, em particular, considera-se a cauda à direita de forma que, o valor- $p$  é dado por

$$p = F(-z|0, 1) = 1 - F(z|0, 1) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt. \quad (2.33)$$

Dado o valor de significância  $\alpha$ , se  $p < \alpha$  a hipótese nula é rejeitada e, portanto, a hipótese alternativa é confirmada considerando-se, em outras palavras, que o método  $M1$  é estatisticamente melhor do que o método  $M2$ , de acordo com a métrica escolhida.

## 2.2 Trabalhos Relacionados

Tsai e Huang (1984) propuseram o primeiro método de super-resolução de imagem, utilizando múltiplas entradas e considerando o domínio da frequência de Fourier. Três anos depois, Peleg, Keren e Schweitzer (1987) propuseram o primeiro trabalho a utilizar imagens no domínio do espaço, ainda empregando múltiplas imagens de entrada. Desde então, muitos trabalhos foram desenvolvidos baseando-se no domínio da frequência (KIM; BOSE, 1990), (AIZAWA; KOMATSU; SAITO, 1991), (RHEE; KANG, 1999), (NGUYEN; MILANFAR, 2000), fazendo uso da relação entre a imagem HR e as observações LR a partir de uma base teórica simples e de alto desempenho computacional (YUE et al., 2016). Entretanto, tais métodos apresentam limitações aparentes, como sensibilidade a erros de modelo e dificuldade para tratar modelos de deslocamento mais complexos, o que inviabilizou a continuidade do uso do domínio de Fourier em super-resolução. Por essa razão, os métodos baseados no domínio do espaço passaram a ser mais frequentes (PARK; PARK; KANG, 2003).

### 2.2.1 Super-Resolução *Multi-Frame* Regularizada

A abordagem *multi-frame* parte do pressuposto de que a disponibilidade de múltiplas imagens da mesma cena fornece informações complementares, desde que haja deslocamento

*sub-pixel* entre tais imagens, ou seja, deslocamento em frações de *pixels* (YUE et al., 2016). Os métodos baseado em *Iterative Back Projection* (IBP) estão entre os primeiros métodos de super-resolução *multi-frame*, como (IRANI; PELEG, 1991) e (IRANI; PELEG, 1993), sendo estes trabalhos consagrados na literatura. Outras duas abordagens comuns na literatura são aquelas baseadas em Filtro Adaptativo Iterativo (ELAD; FEUER, 1999) e em *Projection Onto Convex Sets* (POCS) (STARK; OSKOU, 1989).

A abordagem Bayesiana regularizada pode ser encontrada em diversos trabalhos como (ELAD; FEUER, 1997), (HARDIE; BARNARD; ARMSTRONG, 1997), (SCHULTZ; MENG; STEVENSON, 1998), (NGUYEN; MILANFAR; GOLUB, 2001), (BAKER; KANADE, 2002), (FARSIU et al., 2004), (HE et al., 2007), (NG et al., 2007), (SHEN et al., 2007), (LAGHRIB et al., 2014), (LAGHRIB et al., 2016), (KÖHLER et al., 2016), (LAGHRIB et al., 2019), construídos a partir da suposição de que as imagens podem ser descritas como variáveis aleatórias. Esses métodos tratam a natureza mal posta do problema adicionando um termo de regularização, definido a partir de suposições acerca das características geométricas da imagem como, por exemplo, a presença de bordas. Molina et al. (2003) e Shen et al. (2007) propuseram modelos quadráticos, considerando ruído gaussiano e regularização quadrática, em especial Tikhonov para o segundo caso. Por outro lado, Farsiu et al. (2004) propuseram regularização não quadrática (baseada em norma- $\ell_1$ ) considerando a preservação de borda como prioridade. Farsiu et al. (2004) e Cho e Lee (2011) propuseram modelos de observação não quadráticos mais convenientes para ruídos não-gaussianos.

Os métodos supracitados são baseados em suposições acerca das imagens e, uma vez que tais suposições não são necessariamente ideais para todos os *pixels* da imagem, uma questão relevante diz respeito ao tratamento de *outliers* (são considerados como *outliers* os *pixels* que violam as suposições assumidas para construir a modelagem da solução). Nesse sentido, trabalhos como os propostos por Zeng e Yang (2013), Köhler et al. (2016) e Liu et al. (2018) apresentam modelos que são adaptados localmente a partir de parâmetros que podem ser estimados de forma automática. Esses trabalhos resolvem, assim, questões como diferença de homogeneidade em distintas partes da imagem e *outliers* em relação à suposição do ruído. Essas soluções locais e adaptativas apresentam bons resultados, em relação à qualidade das imagens estimadas, entretanto, demandam alto custo computacional, devido à estimação automática de hiperparâmetros.

O problema de super-resolução consiste em três fases: o registro, a fusão e a reconstrução (KHATTAB et al., 2020). O registro é usado para estimar o deslocamento entre as imagens de baixa resolução, a fusão é a etapa na qual as imagens alinhadas são combinadas e interpoladas em uma única *grid* de alta resolução, e a reconstrução é aplicada para garantir que a imagem final não tenha distorções. Uma vez que a premissa da super-resolução *multi-frame* é adquirir informações a partir do deslocamento *sub-pixel*,

o registro é um componente essencial do problema (EL-GAMAL; ELMOGY; ATWAN, 2016) que pode ser categorizado como paramétrico ou não paramétrico. No caso dos métodos paramétricos, considera-se uma função de mapeamento entre as imagens alinhada e deslocada, de forma que o problema se torna encontrar o melhor valor para os parâmetros dessa função. E para o registro não paramétrico, a diferença ponto-a-ponto entre as imagens é considerada, tendo como foco descrever as diferenças locais entre as imagens alinhada e deslocada. Essa ideia, na qual se baseia o registro não paramétrico, foi introduzida por Bajcsy e Kovačič (1989), e tal tipo de registro também é chamado de “registro elástico” (ZITOVÁ; FLUSSER, 2003).

Muitos métodos de super-resolução consideram registro paramétrico, como é o caso de (SCHULTZ; MENG; STEVENSON, 1998), (HE et al., 2007), (NG et al., 2007), (SHEN et al., 2007), (NGUYEN; MILANFAR; GOLUB, 2001), (KÖHLER et al., 2016) e (LIU et al., 2018). Entretanto, segundo (QIU; XING, 2013), os métodos paramétricos não são capazes de preservar singularidades locais, o que influencia negativamente na recuperação de informações como borda e textura. Nesse sentido, alguns métodos consideram registro não paramétrico, como (BAKER; KANADE, 2000), (LAGHRIB et al., 2014), (LAGHRIB et al., 2016), (WRONSKI et al., 2019), (LAGHRIB et al., 2019).

Dentre os métodos de registro não paramétricos, destaca-se o método Demons (THIRION, 1998), que combina iterações de fluxo óptico com regularização Gaussiana, resultando em um algoritmo computacionalmente eficiente, capaz de resolver deslocamento sub-pixel e robusto a ruído (VERCAUTEREN et al., 2007), (VERCAUTEREN et al., 2009). Até a finalização deste documento, foram encontrados, na literatura, três trabalhos de super-resolução *multi-frame* em que o Demons foi aplicado, sendo todos eles focados em aplicações médicas (RAHMAN; WESARG, 2010), (WU; XIAO; ZHANG, 2014), (SHI et al., 2015).

### 2.2.2 Super-Resolução Baseada em Redes Neurais

As abordagens *single-frame* baseadas em exemplos empregam métodos de aprendizado de máquina (em especial, os regressores não lineares) para recuperar informações de alta frequência que se perdem no processo de discretização e subamostragem da imagem (MILANFAR, 2011) (NASROLLAHI; MOESLUND, 2014). A abordagem baseada em exemplos consiste em usar um conjunto de treinamento, contendo pares de imagens de baixa e alta resolução, para aprender o relacionamento entre as mesmas (COSMO, 2019). Nesse contexto, Freeman, Jones e Pasztor (2002) propuseram um dos primeiros trabalhos proeminentes nessa área e, desde então, várias abordagens foram consideradas, utilizando, por exemplo SVR (*Support Vector Regression*) (JEBADURAI; PETER, 2017), treinamento de dicionário (YANG et al., 2010), árvores de decisão (GRABNER et al., 2017) e redes neurais (DONG et al., 2016) (COSMO; SALLES, 2019).

As imagens podem, ainda, ser divididas em regiões que são classificadas previamente ao treinamento e tratadas por regressores especializados, como proposto por [Li et al. \(2009\)](#). Nesse contexto, [Romano, Isidoro e Milanfar \(2017\)](#) propuseram uma classificação baseada em estatísticas do gradiente. Finalmente, com o objetivo de reduzir o tempo de treinamento, [Cosmo, Inaba e Salles \(2017\)](#) propuseram o uso do R-ELM, algoritmo baseado em uma solução analítica – portanto, rápida – e robusta a *outliers*.

A atual relevância das redes neurais profundas (*deep learning*) está presente nos trabalhos de super-resolução, sobretudo naqueles que abordam métodos *single-frame* ([DONG et al., 2014](#)), ([DONG; LOY; TANG, 2016](#)), ([LEDIG et al., 2017](#)), ([WANG; CHEN; HOI, 2020](#)) e ([ZHANG et al., 2020](#)). No caso dos métodos *multi-frame*, as redes profundas se encontram em trabalhos como ([DEUDON et al., 2020](#)) e ([KAWULOK et al., 2020](#)), os quais consideram imagens de sensoriamento remoto. Entretanto, as técnicas baseadas em redes profundas necessitam tipicamente de hardware especializado, principalmente na forma de GPUs (*Graphic Processor Units*), para que a rede possa ser treinada em um tempo viável (normalmente em dias) e as imagens de teste possam ser reconstruídas de forma rápida. Além disso, as técnicas que utilizam redes neurais profundas necessitam de uma grande quantidade de dados de treinamento para treinar um modelo eficaz ([COSMO; SALLES, 2019](#)), não estando, portanto, no escopo deste trabalho.

### 3 Super-resolução Híbrida Bayesiana Regularizada com Múltiplas Redes Neurais

Os métodos de super-resolução regularizados, explicados na Seção 2.1.5, partem de informação *a priori* para construir o termo regularizador e, assim, estabilizar o problema de inversão e compensar a ausência de informações nas imagens LR. Particularmente, a regularização contribui para ajustar o compromisso entre supressão de ruído e preservação de borda. Entretanto, segundo [Khattab et al. \(2020\)](#), os variados métodos que se encaixam nessa categoria ainda sofrem em relação à solução desse problema, de forma que ainda é uma questão desafiadora no contexto de super-resolução. Ou seja, ainda que a regularização contribua no sentido de contornar a natureza mal-posta do problema, diminuindo o espaço de busca pela solução ótima (que culmina na imagem HR estimada), essa abordagem é limitada, o que sugere, portanto, que as informações contidas na imagem HR original não são inteiramente recuperadas a partir do uso desse artifício.

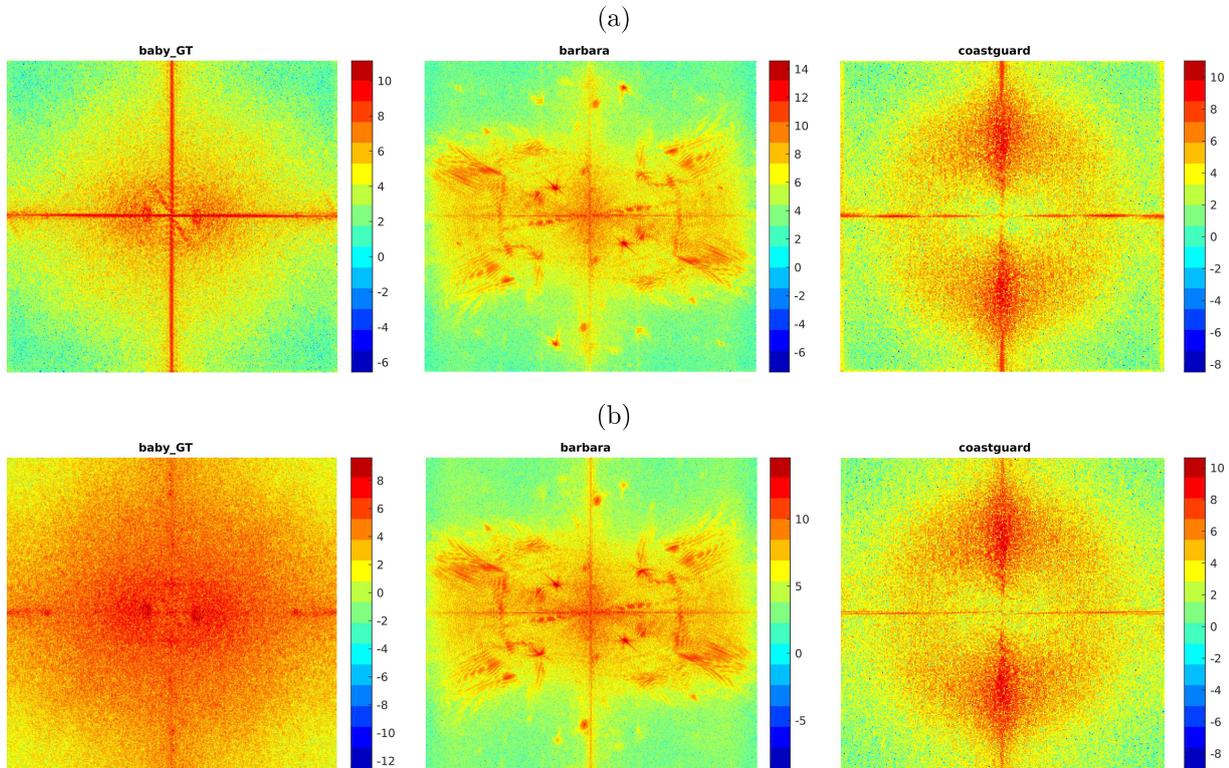
Por outro lado, retoma-se a premissa da super-resolução baseada em aprendizado de máquina (Seção 2.1.6), em que se utiliza a imagem de alta frequência (obtida pela subtração da imagem HR original por uma primeira versão interpolada da imagem LR) e a própria imagem interpolada como dados de entrada de um conjunto de regressores não lineares que tem, portanto, o objetivo de mapear a imagem interpolada à imagem de alta frequência, complementando, assim, a imagem HR estimada.

A Figura 14 mostra o espectro de Fourier das imagens residuais resultantes dos métodos de super-resolução regularizada BTV e IRWSR. Para se obter essas imagens, a imagem HR estimada pelos determinados métodos foi subtraída da imagem HR original (conhecida por se tratar de um contexto simulado) resultando na chamada imagem residual, da qual foi extraída o espectro de Fourier (i.e., a magnitude da transformada de Fourier) e, em seguida, aplicada a transformação logarítmica  $1 + \log |F(u, v)|$ <sup>1</sup>. Observam-se, na Figura 14, imagens residuais com informação não nula em várias regiões do espectro de frequência. Isso sugere a possibilidade de se extrair informações úteis dessas imagens residuais, assim como é feito nos métodos de super-resolução baseada em aprendizado de máquina a partir das imagens de alta frequência.

Nesse contexto, propõe-se um *framework* que combina super-resolução Bayesiana regularizada e super-resolução baseada em múltiplas redes neurais de forma de que o uso das redes neurais contribuam para recuperação das informações que não foram recuperadas pela etapa de super-resolução regularizada. Particularmente, destaca-se o uso das múltiplas

<sup>1</sup> A transformação logarítmica é feita para normalizar o intervalo de valores do espectro, originalmente dominado pelo termo DC ([GONZALEZ; WOODS, 2012](#)).

Figura 14 – Espectro de Fourier das imagens residuais resultantes dos métodos (a) BTV, de Farsiu et al. (2004) e (b) IRWSR, de Köhler et al. (2016), quando aplicados às imagens *baby\_GT* do *Set5*, e *barbara* e *coastguard* do *Set14*.



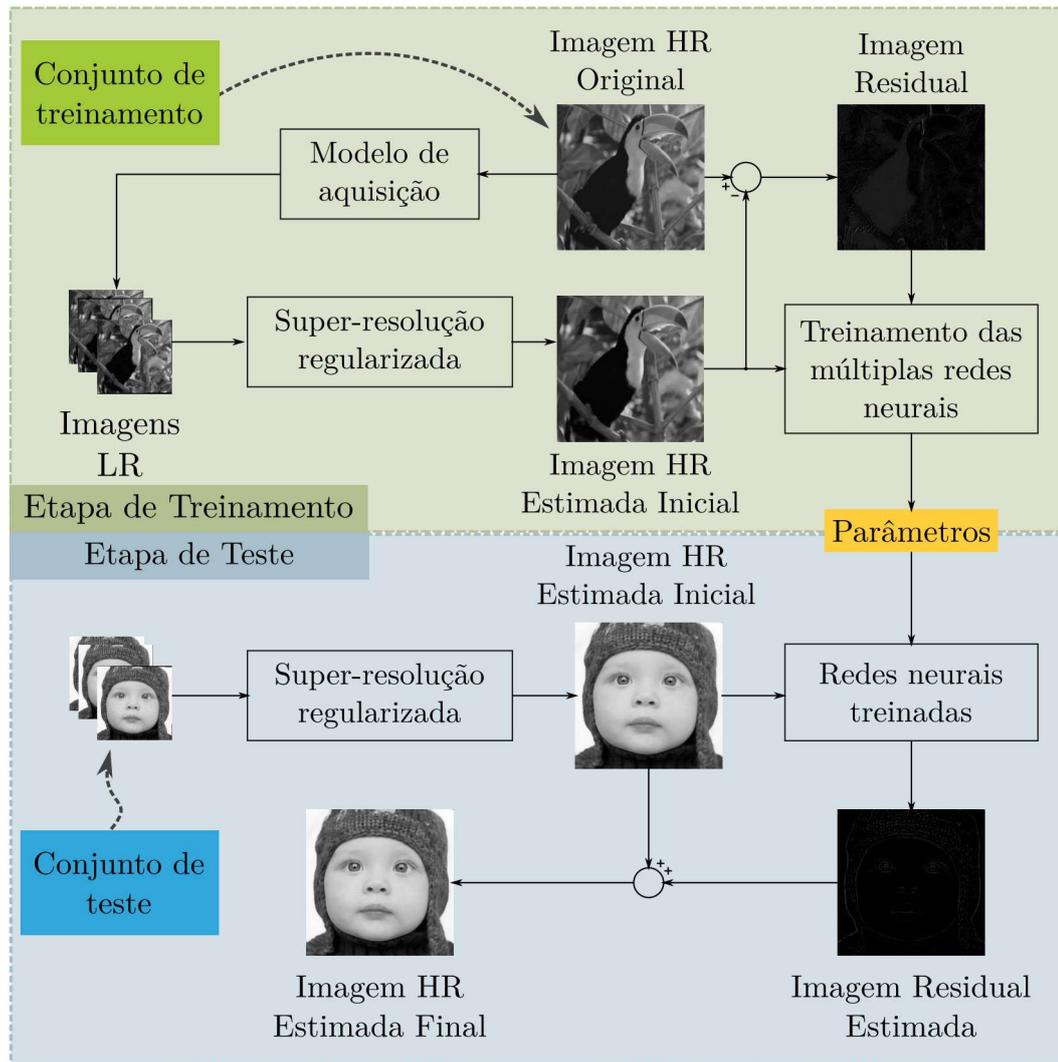
redes neurais nos métodos de super-resolução baseados em aprendizado da Seção 2.1.6 cujas entradas e escolha dos modelos dependem diretamente da extração de características de sub-regiões da imagem e de informações de vizinhança como as derivadas parciais e as estatísticas do gradiente. Em especial, o conjunto de modelos proposto por Cosmo et al. (2018) é escolhido por usar o algoritmo de treinamento R-ELM, que apresenta bom desempenho em termos de tempo de processamento. A proposta deste capítulo se relaciona à primeira hipótese apresentada no Capítulo 1, ou seja, a hipótese de que o uso de informações locais pode trazer bons resultados quando aplicado a métodos de super-resolução *multi-frame*, sendo possível, a partir dele, constituir detalhes relevantes e fiéis à cena real, de onde a imagem foi observada.

### 3.1 Método Proposto

A Figura 15 mostra um esquema ilustrativo e simplificado do método híbrido proposto. Na etapa de treinamento, um banco de imagens é escolhido, de onde as imagens originais de alta resolução são extraídas e deformadas por meio do modelo de aquisição, resultando em  $K$  imagens de baixa resolução (imagens LR) cada. Tais imagens são primeiramente reconstruídas por um método de super-resolução *multi-frame* regularizado

e subtraídas da imagem HR original, resultando na imagem residual<sup>2</sup>. Aplica-se uma varredura em *patches* na imagem residual extraíndo-se as estatísticas do gradiente (ou seja, os valores de força, coerência e orientação) dos *patches* que são, portanto, classificados de acordo com esses valores. As múltiplas redes neurais são alimentadas com os vetores de características extraídos dos *patches* e treinadas utilizando o algoritmo R-ELM.

Figura 15 – Esquema ilustrativo do método híbrido proposto. No topo, em verde, a etapa de treino, e embaixo, em azul a etapa de teste.

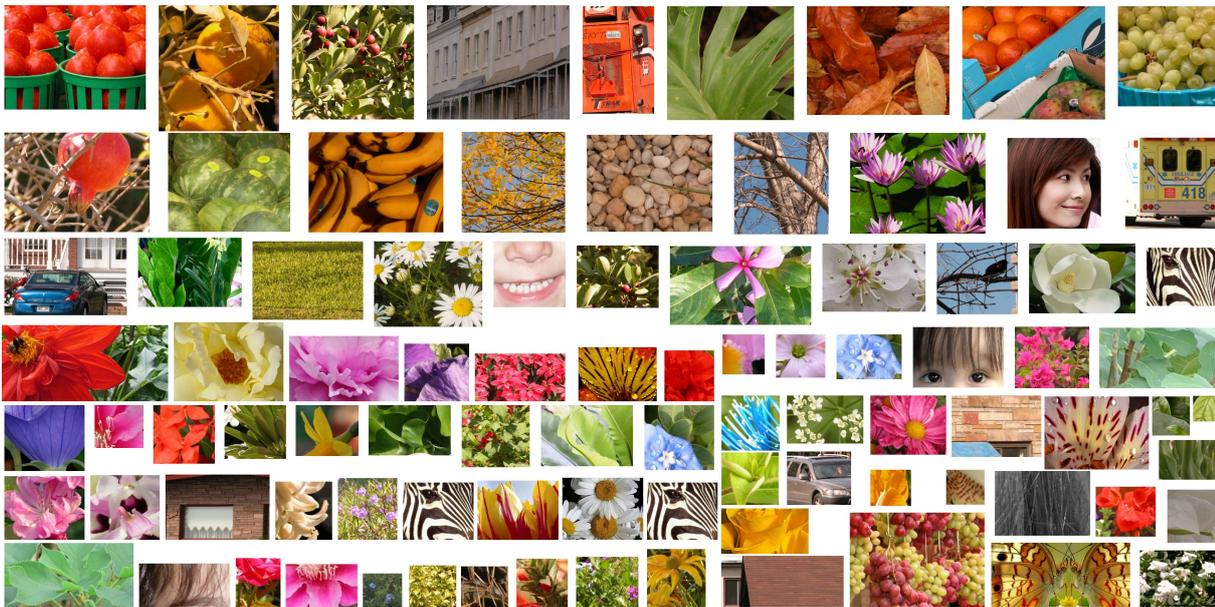


Fonte: Elaborada pela autora.

Para a etapa de testes, o modelo de aquisição e o método de super-resolução regularizado utilizados são os mesmos utilizados na etapa de treinamento. A imagem estimada pelo método regularizado, da qual os vetores de características são extraídos, como mostrado na Figura 9, é dada como entrada para as redes – já treinadas e carregadas com os parâmetros encontrados no treinamento – que resultam nos vetores de saída e,

<sup>2</sup> Cosmo et al. (2018) referem-se à essa imagem como “imagem de alta frequência”, entretanto, nesta tese se prefere chamar de “imagem residual” por ser um termo mais amplo.

Figura 16 – Imagens utilizadas para treinamento das redes neurais.



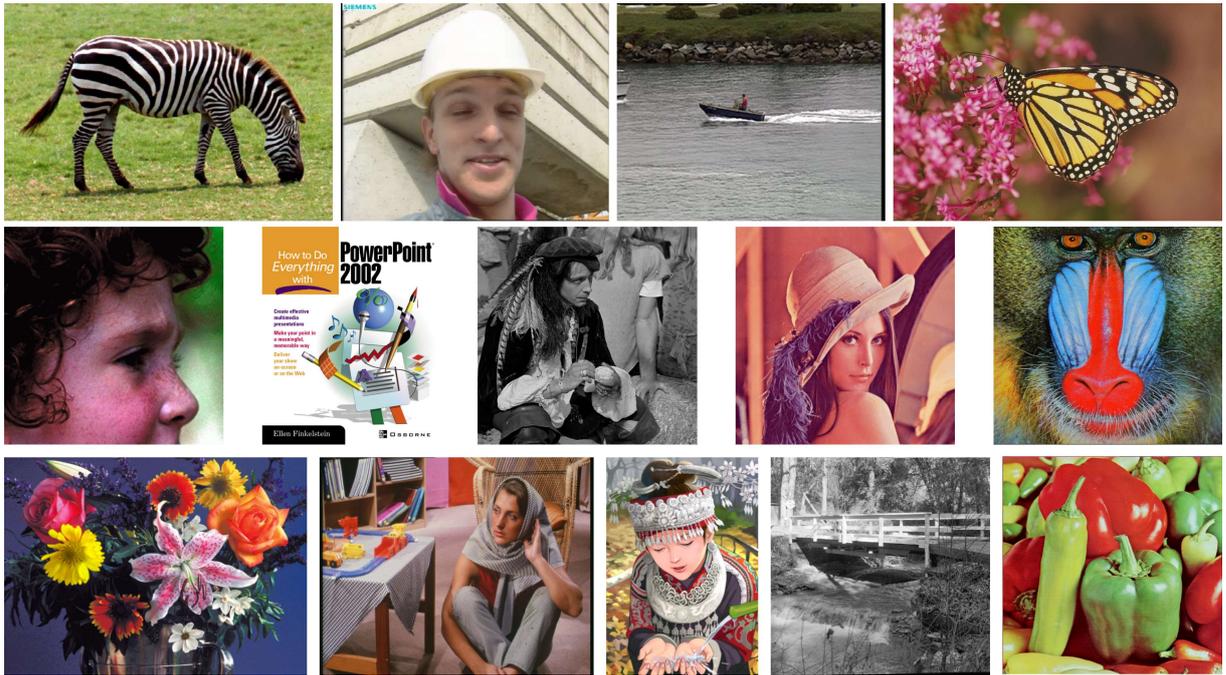
Fonte: Adaptada de [Dong et al. \(2016\)](#)

Figura 17 – Imagens utilizadas para o teste, provenientes do conjunto *Set5*.

Fonte: Adaptada de [Bevilacqua et al. \(2012\)](#)

posteriormente, na imagem residual estimada (remontada como mostra a Figura 10). Soma-se a imagem residual estimada à imagem estimada pelo método regularizado, obtendo-se, assim, a imagem estimada final. Para avaliar o desempenho do método proposto, a imagem estimada final é comparada à imagem original HR, utilizando as métricas SSIM e PSNR. Como métodos regularizados, foram escolhidos os métodos BTV, proposto por [Farsiu et al. \(2004\)](#) e o método proposto por [Köhler et al. \(2016\)](#), chamado, aqui, de IRWSR. Dessa forma, dois métodos híbridos são propostos: (i) HyBTV, baseado no BTV e (ii) HyIRWSR, baseado no IRWSR.

Utiliza-se o mesmo banco de dados de treinamento empregado por [Dong et al. \(2016\)](#), contendo 91 imagens, mostradas na Figura 16. Para o teste, foram utilizados os conjuntos *B100* ([MARTIN et al., 2001](#)), *Set14* ([ROMANO; ISIDORO; MILANFAR, 2017](#)) e *Set5* ([BEVILACQUA et al., 2012](#)), totalizando 119 imagens de tamanhos variados, que são mostradas nas Figuras 17, 18 e 19.

Figura 18 – Imagens utilizadas para o teste, provenientes do conjunto *Set14*.

Fonte: Adaptada de Romano, Isidoro e Milanfar (2017)

Cada imagem de teste – assumida como *ground truth* – foi deformada de acordo com o modelo de aquisição descrito pela Equação 2.3 e os parâmetros dados no Quadro 2 para gerar 8 imagens de baixa resolução. Os vetores de translação horizontal e vertical são definidos aleatoriamente dentro do limite de  $-2$  a  $2$  *pixels* e o vetor de rotação é definido aleatoriamente dentro de limite de  $-1^\circ$  a  $1^\circ$ . Para a *priori* BTV, o fator de regularização  $\lambda$  é definido como 0,0025, a constante  $P$  (que define o tamanho da janela do cálculo da derivada primeira no termo de regularização) é igual a 2 e o fator de escala  $\alpha_0 = 0,6$ . No caso do método IRWSR, o fator de escala é substituído pela matriz  $\mathbf{M}_\alpha$ , calculada automaticamente pelo método. Os métodos híbridos utilizam os mesmos parâmetros para a etapa de reconstrução e, para a etapa de treinamento, os parâmetros dados no Quadro 3 são utilizados. O fator de magnificação  $s$  utilizado é 2. Todos os experimentos foram implementados em MATLAB R2016b em um computador dotado de um processador Intel Core i7-8700k e 32 GB de memória RAM.

As imagem coloridas, originalmente armazenadas no formato RGB, são convertidas para nível de cinza a partir da relação, utilizada por padrão pelo MATLAB R2016b,

$$\mathbf{Y}_{\text{gray}} = 0,2989\mathbf{Y}_R + 0,5870\mathbf{Y}_G + 0,1140\mathbf{Y}_B, \quad (3.1)$$

em que  $\mathbf{Y}_{\text{gray}}$  é imagem em nível de cinza,  $\mathbf{Y}_R$  é o canal R,  $\mathbf{Y}_G$  é o canal G e  $\mathbf{Y}_B$  é o canal B da imagem colorida.



Quadro 2 – Parâmetros do modelo de aquisição, referente aos experimentos híbridos utilizando os métodos *multi-frame* BTV e IRWSR.

Parâmetro	Valor
Número de imagens LR	8
Largura da PSF	0, 4
Desvio padrão do ruído aditivo	0, 02
Limite da definição aleatória do vetor de translação	$[(-2, -2), (2, 2)]$
Limite da definição aleatória do ângulo de rotação	$[-1^\circ, 1^\circ]$

Fonte: Elaborada pela autora.

Quadro 3 – Parâmetros do agrupamento e R-ELM, referente aos experimentos híbridos utilizando os métodos *multi-frame* BTV e IRWSR.

Parâmetro	Valor
<i>Patch</i> para cálculo das estatísticas de gradiente ( $d_g \times d_g$ )	$7 \times 7$
<i>Patch</i> para extração do vetor de características ( $d_c \times d_c$ )	$5 \times 5$
Número de amostras aleatórias	200.000
Termo de regularização $C$ (Equação 2.20)	$2^8$
Número de neurônios na camada oculta	1.000
Função de ativação	Sigmóide
Número de <i>clusters</i>	180
<i>Bins</i> de Orientação	$[0, \frac{\pi}{20}, \frac{2\pi}{20}, \dots, \frac{19\pi}{20}, \pi]^T$
<i>Bins</i> de Força	$[0, \frac{1}{3}, \frac{2}{3}, 1]^T$
<i>Bins</i> de Coerência	$[0, \frac{1}{3}, \frac{2}{3}, 1]^T$

Fonte: Elaborada pela autora.

## 3.2 Resultados Experimentais

Esta seção apresenta as análises experimentais feitas para validar o método proposto, são elas (i) uma comparação estatística utilizando os testes de hipótese de Wilcoxon, Friedman e Nemenyi; (ii) uma comparação do tempo de processamento dos algoritmos e (iii) uma análise por inspeção visual. Para isso, além dos métodos propostos HyBTV e HyIRWSR, os métodos BTV de Farsiu et al. (2004) e IRWSR de Köhler et al. (2016) também foram testados.

### 3.2.1 Comparações Estatísticas

#### 3.2.1.1 Teste Pareado Unicaudal de Wilcoxon

A proposta geral apresentada neste capítulo trata-se de combinar o uso de múltiplas redes neurais de uma camada oculta treinadas com o algoritmo R-ELM com o uso de métodos de super-resolução regularizados, de forma que, do método BTV derivou-se o método HyBTV; e do IRWSR, derivou-se o HyIRWSR. Neste contexto, o teste de Wilcoxon

é aplicado para validar a hipótese de que fazer tal combinação de fato proporciona melhorias ao processo de super-resolução multi-frame, testando a diferença entre as classificações do método BTV e HyBTV (e, analogamente, do método IRWSR e HyIRWSR) considerando as métricas PSNR e SSIM, quando aplicados às imagens do *Set5*, *Set14* e *B100*, totalizando 119 imagens.

Assim sendo, define-se a hipótese nula  $H_0$  e a hipótese alternativa  $H_1$ :

$H_0$ : A mediana da diferença dos valores das métricas do método híbrido em relação ao método regularizado (ou seja, as classificações referentes ao BTV subtraídas do HyBTV ou as classificações referentes ao IRWSR subtraídas de HyIRWSR), considerando PSNR ou SSIM é igual a zero;

$H_1$ : A mediana da diferença dos valores das métricas do método híbrido em relação ao método regularizado (ou seja, as classificações referentes ao BTV subtraídas do HyBTV ou as classificações referentes ao IRWSR subtraídas de HyIRWSR), considerando PSNR ou SSIM diferente de zero;

em que a hipótese alternativa  $H_1$  (ou, de forma equivalente, a rejeição da hipótese nula  $H_0$ ) indica que há evidências de diferença maior que zero entre as medianas das amostras, sugerindo uma melhoria estatisticamente significativa decorrente do uso do método híbrido proposto.

Observa-se, pela Tabela 3, que tanto o método HyBTV quanto o método HyIRWSR apresentam melhores resultados, quando comparados aos métodos BTV e IRWSR, respectivamente, com valores- $p$  consideravelmente menores do que 1%, de forma que há evidências de diferenças entre as duas amostras com nível de significância de 1%, o que sugere a eficácia dos métodos propostos neste capítulo. Além disso, quando comparados o método HyBTV e o método IRWSR, com nível de significância de 5%, há evidências estatísticas que indicam um melhor desempenho do HyBTV em relação ao IRWSR.

Tabela 3 – Resultado do teste de Wilcoxon para os métodos HyBTV e HyIRWSR, quando comparados aos métodos BTV e IRWSR, respectivamente

Métodos comparados	valor- $p$	
	para PSNR	para SSIM
HyBTV - BTV	$1,26 \times 10^{-11}$	$1,40 \times 10^{-10}$
HyIRWSR - IRWSR	$3,17 \times 10^{-6}$	$9,52 \times 10^{-5}$
HyBTV - IRWSR	0,04	0,04

Fonte: Elaborada pela autora.

### 3.2.1.2 Testes de Friedman e Nemenyi

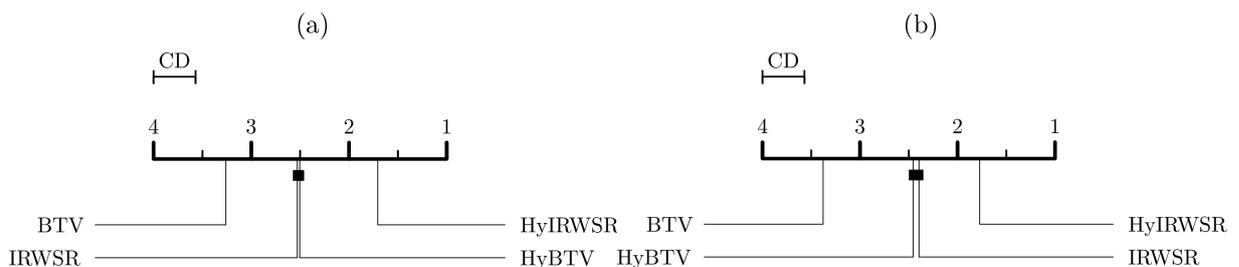
Os testes de Friedman e Nemenyi são utilizados com o objetivo de comparar os quatro métodos, BTV, HyBTV, IRWSR e HyIRWSR, de forma conjunta, definindo-se para o teste de Friedman, a hipótese nula  $H_0$  e a hipótese alternativa  $H_1$ :

$H_0$ : A mediana das classificações considerando a métrica escolhida (PSNR ou SSIM) é equivalente para os quatro métodos, BTV, HyBTV, IRWSR e HyIRWSR;

$H_1$ : Pelo menos dois dos quatro métodos tem medianas diferentes, considerando a métrica escolhida (PSNR ou SSIM).

O valor- $p$ , considerando as classificações a partir dos valores de PSNR, foi de  $1,31 \times 10^{-18}$  e, considerando o SSIM, foi de  $3,47 \times 10^{-20}$ . Ou seja, a hipótese nula  $H_0$  pode ser rejeitada em ambos os casos. Dessa forma, na sequência, aplica-se o teste de Nemenyi para gerar os gráficos de diferença crítica da Figura 20, em que se observa que o método HyIRWSR apresenta o melhor *rank* e o BTV, o pior, para as métricas PSNR e SSIM. Por outro lado, para ambas as métricas, os métodos HyBTV e IRWSR são estaticamente equivalentes, segundo o teste de Nemenyi, uma vez que estão conectados por uma faixa em negrito menor do que a distância crítica.

Figura 20 – Gráfico de diferença crítica (nível de significância de 1%) dos resultados dos métodos BTV, PSNR, HyBTV e HyIRWSR, considerando as imagens do *Set5*, *Set14* e *B100* e as métricas (a) PSNR e (b) SSIM.



Fonte: Elaborada pela autora.

### 3.2.2 Análise do Tempo de Processamento

A Tabela 4 mostra o tempo médio de processamento dos quatro métodos considerando os três bancos de dados utilizados, em que se observa que o IRWSR apresenta tempo de processamento, em média, 3,6 maior do que o HyBTV. O teste de Wilcoxon mostra um melhor resultado do HyBTV comparado ao IRWSR enquanto o teste de Nemenyi sugere um relativo empate entre os dois métodos. Ou seja, o método HyBTV apresenta melhor relação de compromisso entre a qualidade e o tempo de processamento.

Tabela 4 – Tempo de processamento médio, em segundos, de cada experimento, considerando o tempo médio de cada imagem dos conjuntos de teste *B100*, *Set14* e *Set5*. Os valores em negrito representam o melhor resultado para cada conjunto e os valores sublinhados representam o segundo melhor resultado.

	BTV	HyBTV	IRWSR	HyIRWSR
<i>B100</i>	<b>3,57</b>	<u>6,12</u>	28,25	32,42
<i>Set14</i>	<b>5,39</b>	<u>9,01</u>	45,93	49,90
<i>Set5</i>	<b>3,05</b>	<u>5,19</u>	21,69	24,48

Fonte: Elaborada pela autora.

Além disso, os métodos IRWSR e HyIRWSR mostraram-se, em média, cinco vezes mais lentos que os métodos BTV e HyBTV, respectivamente. Dessa forma, este segundo experimento tem o objetivo de analisar o tempo de processamento das diferentes etapas dos quatro métodos. A implementação do método BTV retorna, como imagem de saída, a solução do problema de otimização descrito pela Equação 2.13, etapa denominada “reconstrução multi-frame”. Para o método IRWSR, por outro lado, há, os cálculos das matrizes  $\mathbf{M}_\alpha$  e  $\mathbf{M}_\beta$ , além da solução da Equação 2.15 – etapa da reconstrução multi-frame referente a esse método. Os experimentos mostraram que o tempo de processamento do cálculo de  $\mathbf{M}_\beta$  representa 1,9% do tempo de processamento de  $\mathbf{M}_\alpha$ , por isso, esses valores foram somados. Finalmente, as versões híbridas dos respectivos métodos – HyBTV e HyIRWSR – apresentam, além das etapas supracitadas, a etapa denominada “pós-processamento R-ELM”, referente à execução do método baseado em aprendizado, conforme explicado na Seção 2.1.6.

Os experimentos foram executados sobre as imagens do *Set14* e foram aplicados os mesmos parâmetros apresentados na Seção 3.2. Uma vez que o *Set14* tem imagens de dimensões variadas, o tempo de processamento total de cada etapa foi dividido pelo número de *pixels* super-resolvidos, ou seja, a diferença entre o número de *pixels* das imagens de alta e baixa resolução. E, por fim, foi calculada a média das 14 imagens. Essa métrica denomina-se “tempo por *pixel* super-resolvido”, é dada em milissegundos de acordo com a Equação 3.2.

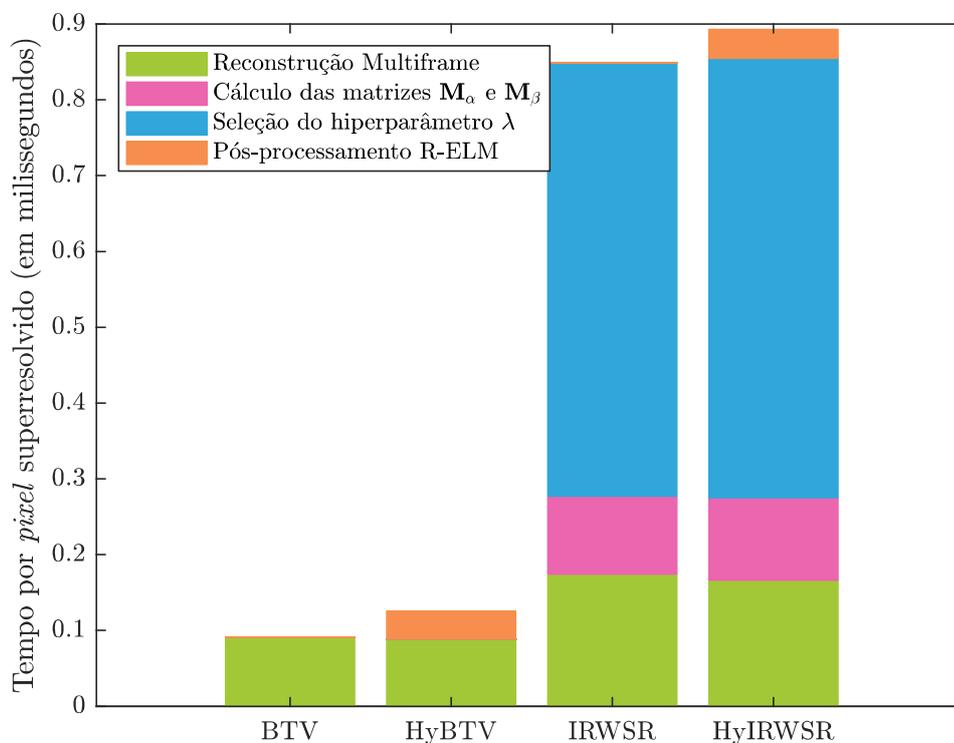
$$\text{Tempo por } \textit{pixel} \text{ super-resolvido} = \frac{T}{I_{dataset} \cdot mn(s^2 - 1)}, \quad (3.2)$$

em que  $T$  é o tempo de processamento total de cada etapa,  $I_{dataset}$  é a quantidade de imagens no *dataset* (neste caso,  $I_{dataset} = 14$ ),  $m \times n$  é a dimensão das imagens LR e  $s$  é o fator de magnificação.

A partir da Figura 21 é possível observar, visualmente, que o cálculos das matrizes  $\mathbf{M}_\alpha$  e  $\mathbf{M}_\beta$  – porção azul das barras – e a seleção de  $\lambda$  – em rosa – representam a maior parte do aumento no tempo de processamento, comparando os métodos baseados em IRWSR aos métodos baseados em BTV. Mais precisamente, a partir dos valores utilizados

para gerar os gráficos, infere-se que o pós-processamento R-ELM aumenta o tempo de processamento em 31% comparando o HyBTV ao BTV e em 5% comparando o HyIRWSR ao IRWSR. Além disso, observa-se um aumento no tempo de processamento de 750% do IRWSR em relação ao BTV e de 550% em relação ao HyBTV. Desse aumento, os cálculos das matrizes  $\mathbf{M}_\alpha$  e  $\mathbf{M}_\beta$  e a seleção de  $\lambda$  representam de 89% quando o IRWSR é comparado ao BTV e de 93% quando comparado ao HyBTV. Ou seja, a natureza adaptativa inerente ao método IRWSR aumenta o tempo de processamento de forma mais significativa do que o tempo adicionado pelo uso das múltiplas redes neurais, proposto neste capítulo.

Figura 21 – Tempo médio de processamento por *pixel* superresolvido de cada etapa dos métodos *BTV*, *HyBTV*, *IRWSR* e *HyIRWSR* aplicados às imagens do *Set14*.



Fonte: Elaborado pela autora.

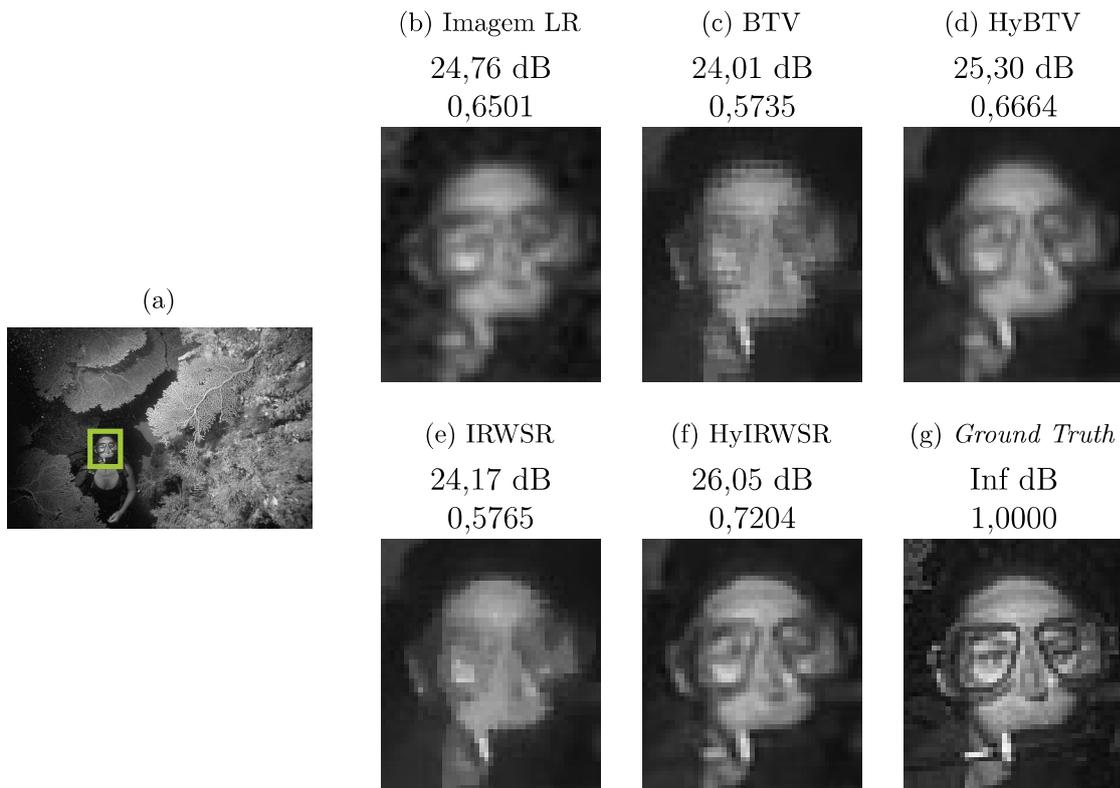
### 3.2.3 Análise por Inspeção Visual

Esta seção trata de análises qualitativas, feitas por inspeção visual com o objetivo de identificar características visuais presentes nas imagens super-resolvidas, sobretudo explorando as vantagens e limitações dos métodos propostos neste capítulo. Para isso, dentre as 119 imagens usadas nos experimentos, cinco foram escolhidas, em que as quatro primeiras (ilustradas nas Figuras 22, 23, 24 e 25) mostram as vantagens visualmente perceptíveis dos métodos HyBTV e HyIRWSR, e as últimas (ilustradas na Figura 26 e 27) referem-se a uma situação em que tais métodos obtêm piores resultados. Os resultados

quantitativos (em termos de PSNR e SSIM) também estão presentes, nas legendas das figuras.

A Figura 22 mostra os resultados referentes à imagem 156065, do B100, em que se observa, no detalhe, o rosto de uma mulher debaixo d'água usando óculos de mergulho e um aparato na boca. A vantagem do HyBTV e do HyIRWSR sobre os métodos BTV e IRWSR, presente nos valores das métricas, se mantém visualmente, uma vez que o rosto da mulher nas Figuras 22c e 22e está praticamente desfigurado, sobretudo na Figura 22c, em que se observa, além disso, significantes artefatos quadriculados nas regiões de borda, causando um aspecto *pixelado* à imagem. Por outro lado, nas Figuras 22d e 22f, o rosto está mais perceptível, em decorrência de uma melhor reconstrução das informações de alta frequência, além da ausência de artefatos quadriculados.

Figura 22 – Imagens super-resolvidas, referentes à imagem original de alta resolução 156065, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.

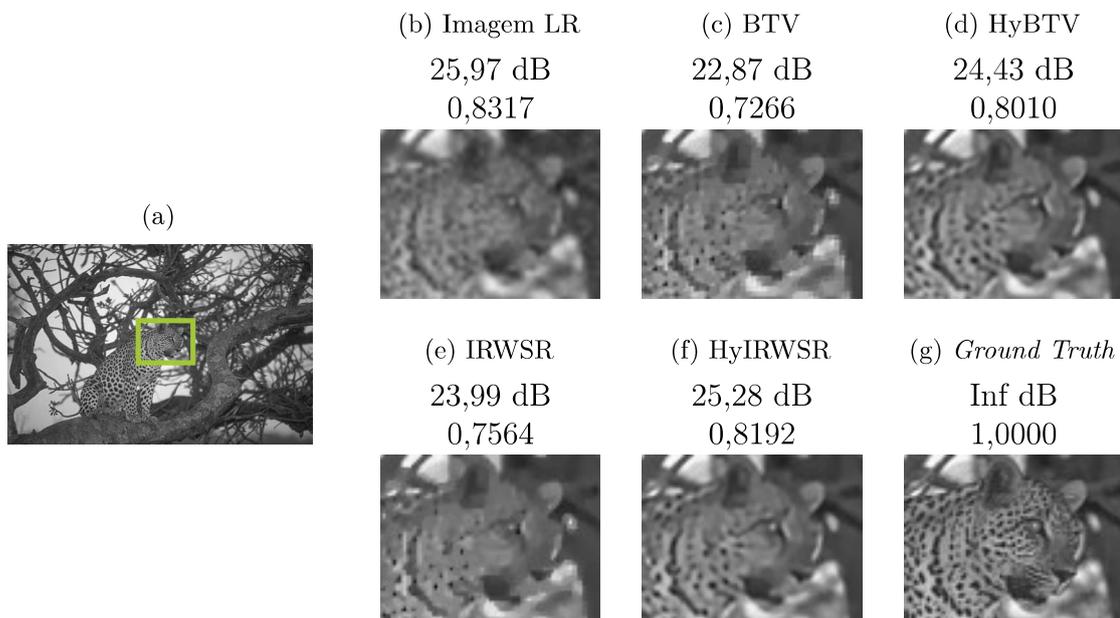


Fonte: Elaborada pela autora.

O aspecto *pixelado* se observa, também, no resultado do método BTV referente à imagem 134035 do B100, em que tem-se a vista lateral da região entre o rosto e pescoço de uma onça, como mostra a Figura 23c. Além disso, tanto no resultado do BTV quanto no

resultado do IRWSR, na Figura 23e, o padrão das pintas da onça sofreu uma deformação considerável, causando um aspecto cartunesco à imagem. As imagens resultantes dos métodos HyBTV e HyIRWSR, representadas nas Figuras 23d e 23f, respectivamente, são similares entre si e, apesar de apresentarem demasiada suavização no rosto e pescoço da onça, não manifestam aspecto cartunesco e apresentam uma melhor reconstrução das pintas do animal. As melhorias mencionadas, presentes nas imagens do HyBTV e HyIRWSR são refletidas nos resultados quantitativos.

Figura 23 – Imagens super-resolvidas, referentes à imagem original de alta resolução 134035, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.

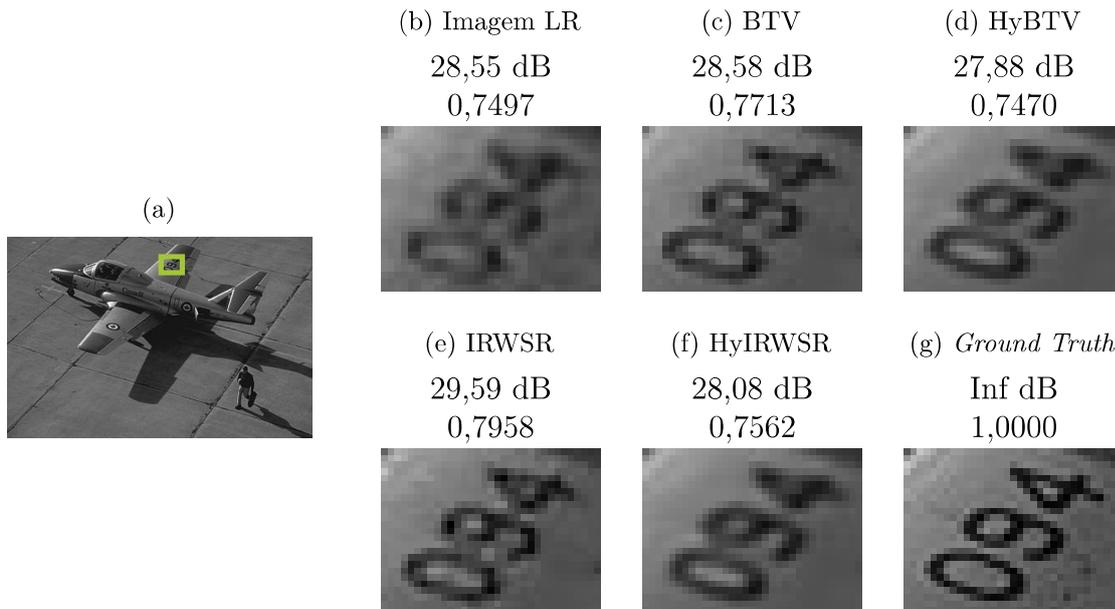


Fonte: Elaborada pela autora.

Ademais, destaca-se na Figura 24 parte da asa de um avião em que há os caracteres “094”, como mostra a Figura 24g, registrados equivocadamente como “034” na imagem LR mostrada na Figura 24b. Quanto à reconstrução de tais caracteres, na imagem referente ao IRWSR (na Figura 24e) o carácter “9” não se mostra inteiramente reconstruído, podendo ser confundido com um “3”, apesar desta região ter sido extraída da a imagem com melhor valores das métricas de qualidade (29,59 dB de PSNR e 0,7958 de SSIM); o mesmo se observa na imagem referente ao método BTV (Figura 24c com 28,58 dB de PSNR e 0,7713 de SSIM), além do supramencionado artefato quadriculado nas regiões de borda, também presente no resultado do IRWSR. Em contrapartida, as imagens super-resolvidas pelos métodos HyBTV (Figura 24d) e HyIRWSR (Figura 24f) não apresentam artefato

quadriculado e possuem melhor reconstrução dos caracteres, estando legível o número “094”.

Figura 24 – Imagens super-resolvidas, referentes à imagem original de alta resolução 37073, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.

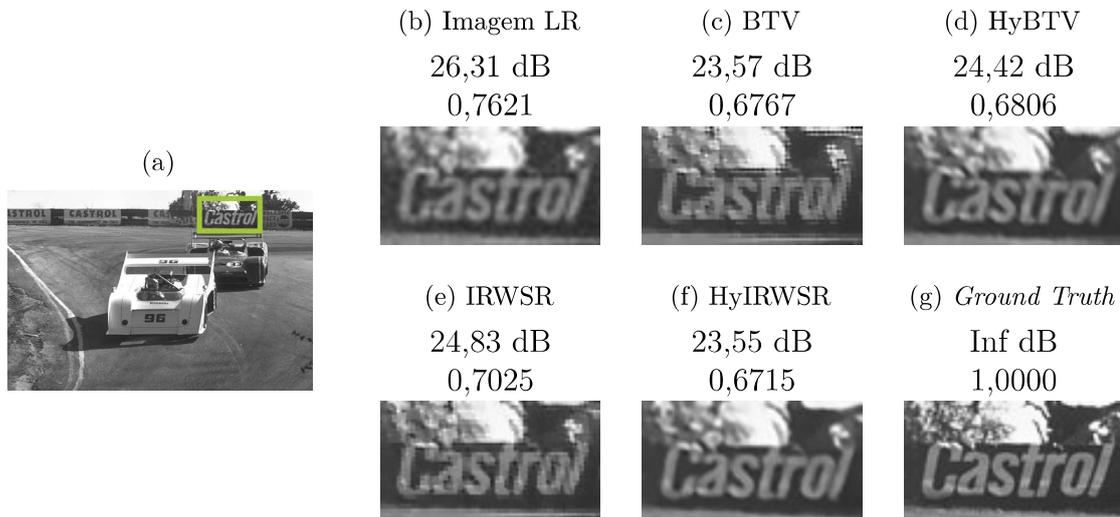


Fonte: Elaborada pela autora.

Outra situação na qual se observam caracteres está apresentada na Figura 25 em que se tem, no destaque, a palavra “Castrol” escrita em um muro e algumas copas de árvores ao fundo. Nesse caso, a palavra foi corretamente reconstruída em todas as imagens, o que era esperado, uma vez que isso também ocorre na imagem LR (Figura 25b). O já mencionado aspecto *pixelado* aparece na imagem referente ao BTV (Figura 25c) e, de forma menos perceptível, no resultado do IRWSR (Figura 25e), ainda que este último apresente os melhores valores de PSNR e SSIM. Como visto em situações anteriores, tal aspecto não aparece nas imagens referentes aos métodos HyBTV (Figura 25d) e HyIRWSR (Figura 25f), mas se nota demasiada suavização.

Considerando a imagem 145086, do B100, a diferença entre os resultados quantitativos dos métodos comparados é mais discrepante, de forma que os métodos propostos neste trabalho (HyBTV e HyIRWSR) obtêm resultados de 5% a 10% piores do que os métodos BTV e IRWSR. Entretanto, a Figura 26c, referente ao BTV, apresenta significativo artefato de movimento, que não está presente na Figura 26d (referente ao método HyBTV). Com relação às Figuras 26e (método IRWSR) e 26f (HyIRWSR), a primeira apresenta um leve

Figura 25 – Imagens super-resolvidas, referentes à imagem original de alta resolução 21077, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.



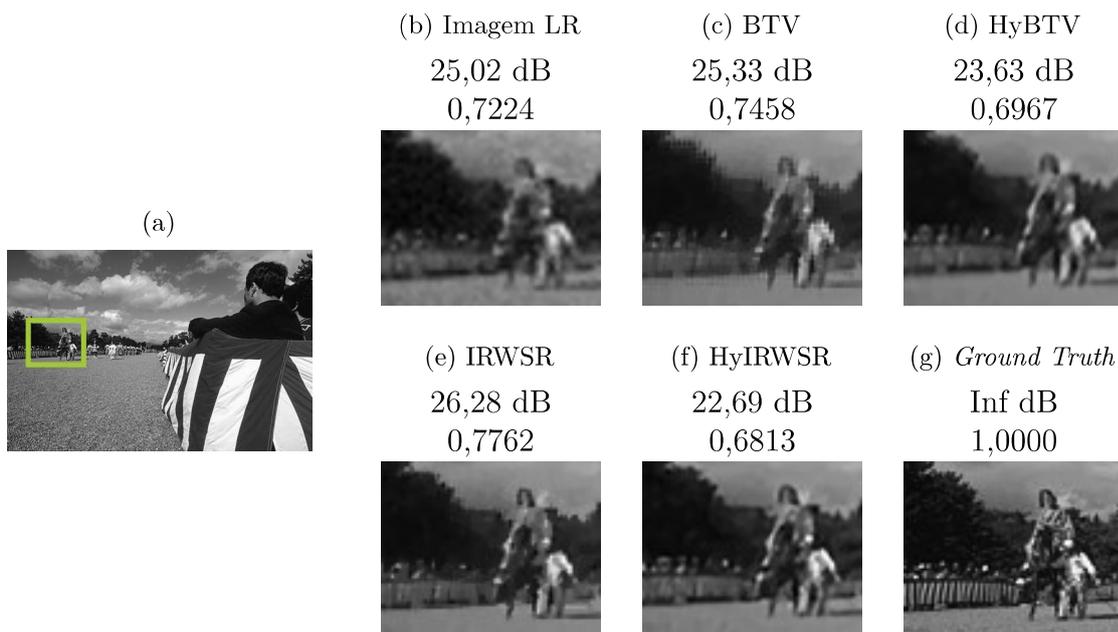
Fonte: Elaborada pela autora.

aspecto cartunescos e a segunda apresenta suavização demasiada.

O pior desempenho dos métodos HyBTV e HyIRWSR acontece quando considerada a imagem 219090, também do B100, cujos resultados são mostrados na Figura 27, em que se observa significativo borramento de movimento nas Figuras 27d e 27f, particularmente. Além disso, a imagem resultante do método IRWSR, na Figura 27e contém artefato quadriculado na borda, bem como o resultado do BTV (Figura 27c) apresenta leve artefato nas regiões de borda.

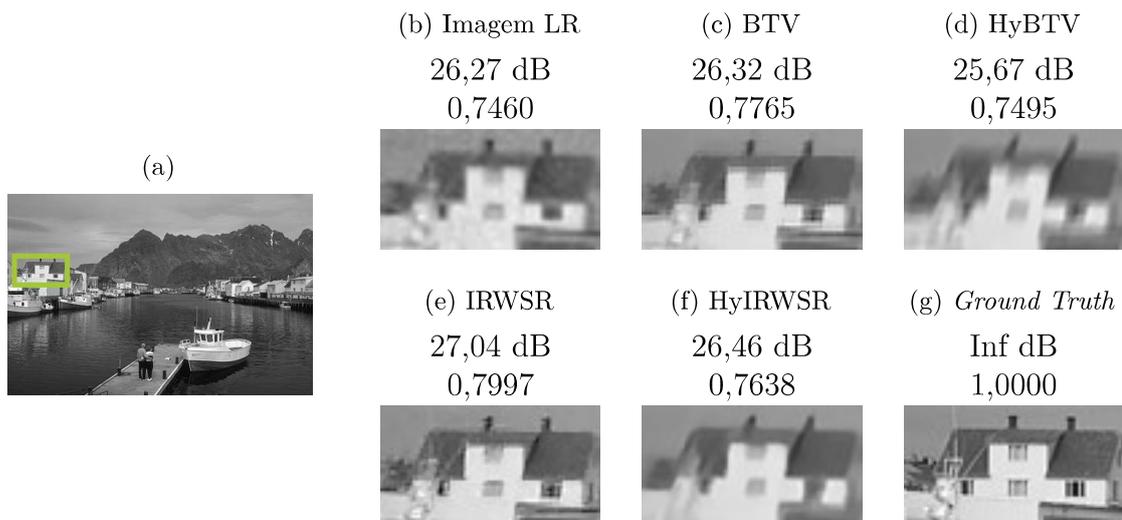
A análise por inspeção visual mostra, em suma, que os métodos híbridos propostos neste capítulo se destacam por não apresentarem (i) aspecto *pixelado* em decorrência de artefato quadriculado nas regiões de borda, (ii) aspecto cartunescos e (iii) deformação de estruturas como faces e caracteres. Entretanto, ainda se observa demasiada suavização nas imagens, perceptível sobretudo, em regiões de bordas e textura, sendo essa uma importante limitação dos métodos HyBTV e HyIRWSR. Ou seja, o uso das múltiplas redes neurais após a aplicação de um método regularizado de super-resolução *multi-frame* traz melhorias no que se diz respeito à reconstrução de informações de alta frequência, mas ainda não resolve a super-resolução por completo. A presença de artefatos em relação à deformação por movimento motiva uma investigação acerca do método de registro, que alinha as imagens LR, o que, em particular, fomenta os métodos propostos nos Capítulos 4 e 5 desta tese.

Figura 26 – Imagens super-resolvidas, referentes à imagem original de alta resolução 145086, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.



Fonte: Elaborada pela autora.

Figura 27 – Imagens super-resolvidas, referentes à imagem original de alta resolução 219090, (a) *ground-truth*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método HyBTV, (e) destaque da imagem super-resolvida pelo método IRWSR, (f) destaque da imagem super-resolvida pelo método HyIRWSR e (g) destaque do *ground truth*.



Fonte: Elaborada pela autora.



## 4 Super-resolução com Registro *Demons* e Reconstrução Bayesiana Regularizada

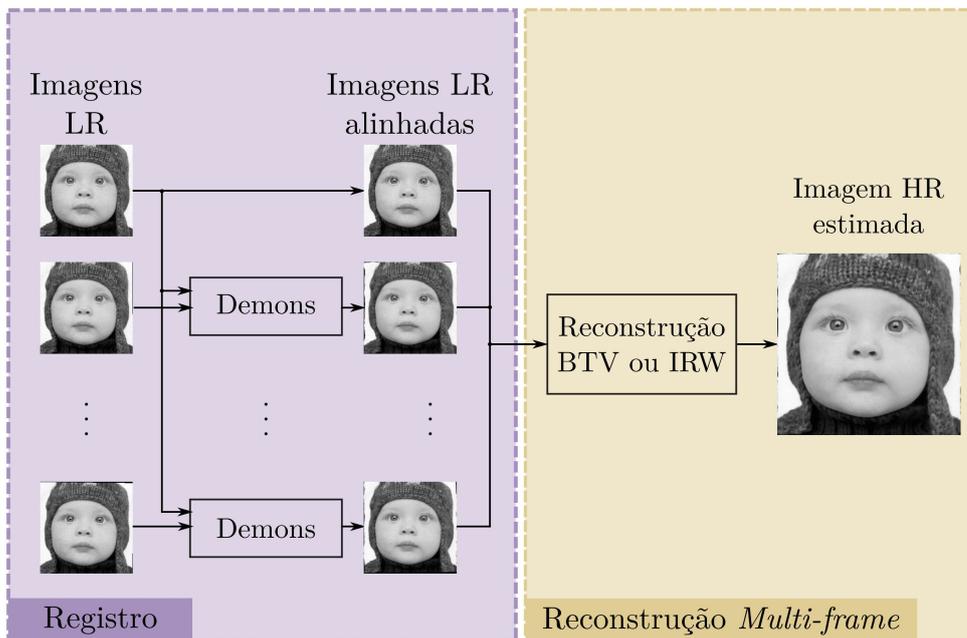
Os métodos de super-resolução *multi-frame* Bayesianos, como visto no Capítulo 1, podem ser categorizados como métodos alternados ou sequenciais (KATSAGGELOS; MOLINA; MATEOS, 2007), dependendo de como o alinhamento das imagens LR é feito. No caso dos métodos sequenciais, em particular, as imagens são previamente alinhadas, por um método de registro, de forma que a segunda etapa do problema (chamada, aqui, de *reconstrução*), toma como entrada versões alinhadas das imagens LR. O problema de super-resolução *multi-frame* é um problema mal-posto e, portanto, possui múltiplas soluções possíveis. Uma vez que tal natureza mal-posta decorre do número de imagens LR e do operador de borramento (PARK; PARK; KANG, 2003), o problema de reconstrução é, também, mal-posto, conseqüentemente resultando em múltiplas soluções. A implementação do paradigma Bayesiano para super-resolução *multi-frame* tipicamente considera soluções numéricas iterativas, cujos resultados dependem das condições iniciais definidas. Dessa forma, introduz-se, nesta tese, a perspectiva de que a escolha do método de registro muda a condição inicial do problema de reconstrução, influenciando no resultado final do mesmo, já que um registro mal feito pode ocasionar erros de estimação, propagados ao longo das iterações da etapa de reconstrução.

Os métodos de registro paramétricos dependem de suposições acerca do modelo de deslocamento, que nem sempre estão disponíveis, de forma que são limitados à factibilidade da solução, não sendo capazes de preservar singularidades locais (QIU; XING, 2013). Essa característica é particularmente desvantajosa no que diz respeito à recuperação de bordas e texturas, motivando a busca por métodos não paramétricos eficientes para serem aplicados às propostas desta tese. Retoma-se que o método de aquisição mostrado na Seção 2.1.2 considera deslocamento sub-*pixel* com rotação e translação, cuja presença torna o problema de registro inerentemente não-linear, segundo (VANDEWALLE; SBAIZ; VETTERLI, 2011). Nesse sentido, motiva-se o uso do método *Demons* (THIRION, 1998), um dos métodos mais eficientes para o caso não linear (VERCAUTEREN et al., 2007), além de ser adequado para problemas com deslocamento sub-*pixel*. Dessa forma, este capítulo tem o objetivo de validar a hipótese (ii) desta tese de que o uso do registro *Demons* aplicado a um método de super-resolução sequencial melhora o desempenho do mesmo.

## 4.1 Método Proposto

A Figura 28 e o Algoritmo 2 esquematizam o método de super-resolução *multi-frame* sequencial proposto, composto por duas etapas: registro e reconstrução. O registro *Demons* é aplicado às imagens LR de entrada, gerando versões alinhadas das mesmas, que são fornecidas, como entradas, ao método de reconstrução *multi-frame* que, por sua vez, gera a imagem HR estimada. A etapa de reconstrução, por sua vez, explorada na Seção 4.1.3, é baseada nos métodos BTV, de Farsiu et al. (2004), e no método IRWSR, de Köhler et al. (2016). Dessa forma, dois métodos são propostos: (i) o método D-BTVIR, formado pela combinação do registro *Demons* com a reconstrução baseada em BTV e (ii) D-IRWIR, formado pela combinação do registro *Demons* com a reconstrução baseada no IRWSR (KÖHLER et al., 2016). Como os métodos de reconstrução baseados no BTV e no IRWSR requerem, além das imagens LR, um operador de deslocamento, compreendido como hiperparâmetro, neste caso, considera-se a transformação identidade para o mesmo (Linha 7), uma vez que as imagens de entrada já estão alinhadas.

Figura 28 – Esquema simplificado do método “Super-resolução com registro *Demons* e reconstrução Bayesiana Regularizada”



Fonte: Elaborada pela autora.

### 4.1.1 Registro *Demons*

Os métodos de registro tipicamente tomam duas imagens como entradas, a imagem de referência e a imagem deslocada, com o objetivo de alinhar a imagem deslocada ao *grid* da imagem de referência (ZITOVÁ; FLUSSER, 2003). Portanto, para a etapa de registro em métodos de super-resolução *multi-frame*, é necessário escolher uma das imagens LR

**Algoritmo 2:** Super-resolução sequencial usando registro *Demons*


---

```

1 Entradas:  $\mathbf{y}_k, \mathbf{B}, \mathbf{D}, \sigma_{\text{diff}}, L, I, \rho, \phi$ 
2 Saída: Imagem HR estimada  $\hat{\mathbf{x}}$ 
3 Seleciona imagem LR de referência,  $\tilde{\mathbf{y}}_r = \mathbf{y}_r, r \in \{1, \dots, K\}$ 
4 for  $k \in \{1, \dots, K\} \setminus \{r\}$  do
5   | Gera versões alinhadas das imagens LR  $\tilde{\mathbf{y}}_k = \text{Demons}(\mathbf{y}_k, \tilde{\mathbf{y}}_r, \sigma_{\text{diff}}, L, I)$ 
6 end
7 Constrói  $\widetilde{\mathbf{W}}_k$  considerando  $\tilde{\mathbf{s}}_k = \mathbf{I}_{3 \times 3}, k = 1, \dots, K$ 
8 Combina operadores  $\widetilde{\mathbf{H}}_k = \mathbf{DB}\widetilde{\mathbf{W}}_k, k = 1, \dots, K$ 
9 Constrói estrutura  $\widetilde{\mathbf{H}}$  com  $\widetilde{\mathbf{H}}_k, k = 1, \dots, K$ 
10 Constrói estrutura  $\tilde{\mathbf{y}}$  com  $\tilde{\mathbf{y}}_k, k = 1, \dots, K$ 
11 Estima  $\hat{\mathbf{x}}$  executando  $\hat{\mathbf{x}} = \text{Reconstruction}(\rho, \phi, \tilde{\mathbf{y}}, s, \mathbf{B}, \widetilde{\mathbf{W}})$ 

```

---

como sendo a imagem de referência (ou seja, aquela que não sofreu deslocamento a partir da imagem HR original)  $\mathbf{y}_r$ , de forma que ela é tomada como sua própria versão alinhada, como descrito na Linha 3 do Algoritmo 2. Para as outras imagens LR, o registro *Demons* é calculado separadamente gerando versões alinhadas  $\tilde{\mathbf{y}}_k, k \in \{1, \dots, K\} \setminus \{r\}$  de cada uma delas, como descrito nas Linhas 4, 5 e 6 do Algoritmo 2.

### 4.1.2 Operador de Deslocamento

A etapa de reconstrução considera as versões alinhadas das imagens LR como variáveis observadas para resolver o problema inverso relacionado à transformação direta descrita por

$$\tilde{\mathbf{y}}_k = \mathbf{DB}\widetilde{\mathbf{W}}_k\mathbf{x} + \tilde{\boldsymbol{\eta}}_k = \widetilde{\mathbf{H}}_k\mathbf{x} + \tilde{\boldsymbol{\eta}}_k, \quad (4.1)$$

onde  $\tilde{\mathbf{y}}_k$  é a versão alinhada da  $k$ -ésima imagem LR,  $\widetilde{\mathbf{W}}_k$  é o operador de deslocamento que mapeia uma imagem de entrada arbitrária no *grid* dela mesma e  $\tilde{\boldsymbol{\eta}}_k$  é o ruído aditivo, que assume-se ser AWGN. O operador de deslocamento descreve uma transformação geométrica composta por duas etapas (i) aplicando uma transformação espacial  $\tilde{\mathbf{s}}_k$  e (ii) aplicando uma interpolação de intensidade que atribui valores de intensidade de cor para os *pixels* espacialmente transformados por  $\tilde{\mathbf{s}}_k$ . No contexto do método proposto, a interpolação é incorporada ao processo de reconstrução, de forma que a transformação espacial é, na verdade, um hiperparâmetro (representado pelo operador  $\widetilde{\mathbf{W}}_k$ ) para a reconstrução. Esta última, por sua vez, deriva dos métodos BTV, de (FARSIU et al., 2004) e IRWSR, de (KÖHLER et al., 2016), que consideram, na implementação utilizada, a transformação afim. Como as imagens de entrada da reconstrução  $\tilde{\mathbf{y}}_k$  são tidas como alinhadas (uma vez que são obtidas a partir do registro *Demons*), a transformação geométrica representada é a transformação identidade que, quando do modelo afim, é dada pela matriz identidade (GONZALEZ; WOODS, 2012), ou seja,  $\tilde{\mathbf{s}}_k = \mathbf{I}_{3 \times 3}, k = 1, \dots, K$ , em que  $K$  é o número de

imagens LR. A formação de  $\widetilde{\mathbf{W}}_k$  é feita a partir da versão lexicográfica de  $\tilde{\mathbf{s}}_k$  e pode ser lida com mais detalhes em (HE et al., 2007).

### 4.1.3 Reconstrução Bayesiana Regularizada

A etapa de reconstrução foi implementada baseada nas duas abordagens de super-resolução *multi-frame* BTV e IRW, considerando as versões alinhadas das imagens LR  $\tilde{\mathbf{y}}_k$  e o operador de deslocamento  $\widetilde{\mathbf{W}}_k$  como entradas.

#### 4.1.3.1 BTVIR

A partir do método de super-resolução *multi-frame* BTV (FARSIU et al., 2004), propõe-se o método de reconstrução *multi-frame* BTVIR, sendo a imagem HR estimada calculada de acordo com a Equação 4.2.

$$\begin{aligned} \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} & \sum_{k=1}^K \|\widetilde{\mathbf{H}}_k \mathbf{x} - \tilde{\mathbf{y}}_k\|_1 + \\ & \lambda \sum_{l=-P}^P \sum_{m=-P}^P \alpha^{|m|+|l|} \|\mathbf{x} - S_h^l S_v^m \mathbf{x}\|_1, \end{aligned} \quad (4.2)$$

em que o parâmetro de regularização  $\lambda$  é um escalar que pondera o segundo termo à direita (a regularização) em relação ao primeiro termo (também chamado de fidelidade),  $\alpha \in [0, 1]$ ,  $S_h^l$  e  $S_v^m$  são operadores que deslocam  $\mathbf{x}$  em  $l$  e  $m$  *pixels* nas direções horizontal e vertical, respectivamente, em uma região de dimensão  $(2P + 1) \times (2P + 1)$  com  $P \geq 1 \in \mathbb{Z}$ . No Algoritmo 2, quando D-BTVIR é considerado, o método de reconstrução  $\rho$  é BTVIR, e o conjunto de atributos  $\phi$  é  $\{\lambda, \alpha, P\}$ .

#### 4.1.3.2 Reconstrução IRW

O método de reconstrução *multi-frame* IRW é derivado a partir da super-resolução *multi-frame* robusta IRWSR (KÖHLER et al., 2016), sendo a imagem HR estimada calculada a partir da Equação 4.3.

$$\begin{aligned} \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} & (\tilde{\mathbf{y}} - \widetilde{\mathbf{H}}\mathbf{x})^\top \mathbf{M}_\beta (\tilde{\mathbf{y}} - \widetilde{\mathbf{H}}\mathbf{x}) + \\ & \lambda \sum_{l=-P}^P \sum_{m=-P}^P \|\mathbf{M}_\alpha (\mathbf{x} - S_h^l S_v^m \mathbf{x})\|_1, \end{aligned} \quad (4.3)$$

em que  $\mathbf{M}_\alpha = \text{diag}(\alpha_1, \dots, \alpha_{MN})$ ,  $\mathbf{M}_\beta = \text{diag}(\beta_1, \dots, \beta_{MN})$  e  $M \times N$  é a dimensão da imagem HR  $\mathbf{x}$ . IRWIR pondera o efeito do primeiro e do segundo termo localmente, uma vez que o método considera matrizes ao invés de escalares como parâmetros de ponderação. Além disso,  $\mathbf{M}_\alpha$ ,  $\mathbf{M}_\beta$  e  $\lambda$  também são estimados conjuntamente à imagem HR estimada  $\hat{\mathbf{x}}$ , o que confere uma natureza robusta ao método, de acordo com os autores de (KÖHLER et al., 2016). No Algoritmo 2, quando do uso do D-IRWIR, o método de reconstrução  $\rho$  é

IRWIR, e o conjunto de atributos  $\phi$  é  $\{P\}$ . As Equações 4.2 e 4.3 são solucionadas usando o algoritmo SCG (*Scaled Conjugate Gradient*) (NABNEY, 2002).

## 4.2 Experimentos e Resultados

Essa seção apresenta análises experimentais para validar o uso do registro *Demons* combinado aos métodos de BTV e IRW aplicados como reconstrução *multi-frame*, que compõe as duas propostas deste capítulo, o método D-BTVIR e o método D-IRWIR. Para tal fim, duas análises foram feitas (i) uma comparação estatística, usando os testes de hipótese de Friedman, Nemenyi e Wilcoxon Unicaudal e (ii) uma análise por inspeção visual, com o objetivo de encontrar características recorrentes das imagens super-resolvidas pelos dois métodos propostos. As imagens dos conjuntos *Set5* (BEVILACQUA et al., 2012), *Set14* (ROMANO; ISIDORO; MILANFAR, 2017) e *B100* (MARTIN et al., 2001) são utilizadas como *ground truth* e os métodos BTV (FARSIU et al., 2004) e IRWSR (KÖHLER et al., 2016) são submetidos à mesma bateria de testes para fins de comparação.

O Quadro 4 mostra os parâmetros referentes ao modelo de aquisição utilizado para gerar as imagens LR a partir das imagens consideradas como *ground truth*. Além disso, para os métodos IRWSR e BTV, e para a etapa de reconstrução dos métodos D-BTVIR e D-IRWIR, o parâmetro que define a janela do cálculo da derivada primeira no termo de regularização é  $P = 2$ ; para os métodos BTV e D-BTVIR, o fator de regularização é  $\lambda = 0,005$  e o fator de escala  $\alpha = 0,5$ ; e para o registro *Demons* (primeira etapa do D-BTVIR e D-IRWIR) o número máximo de iterações é  $I = 100$ , o número de níveis do esquema multi-escalar é  $L = 3$  e a regularização Bayesiana tem desvio padrão  $\sigma_{\text{diff}} = 1,0$ .

Quadro 4 – Parâmetros do modelo de aquisição utilizado nos experimentos referentes à super-resolução com registro *Demons*.

Parâmetro	Valor
Número de imagens LR	8
Largura da PSF	0,4
Desvio padrão do ruído aditivo	0,02
Limite da definição aleatória do vetor de translação	$[(-2, -2), (2, 2)]$
Limite da definição aleatória do ângulo de rotação	$[-1^\circ, 1^\circ]$

Fonte: Elaborada pela autora.

### 4.2.1 Comparações Estatísticas

Esta seção discorre sobre as comparações estatísticas, que foram feitas usando os testes de hipótese de Friedman, Nemenyi e Wilcoxon Unicaudal

#### 4.2.1.1 Teste Pareado Unicaudal de Wilcoxon

A proposta apresentada neste capítulo baseia-se no uso do registro não paramétrico *Demons* para super-resolução multi-frame Bayesiana regularizada, de forma que a combinação da reconstrução BTVIR com o registro *Demons* derivou o método D-BTVIR bem como a reconstrução IRWIR combinada a tal registro derivou o método D-IRWIR. Nesse contexto, como foi feito na Seção 3.2.1.1, o teste pareado unicaudal de Wilcoxon foi usado para comparar estatisticamente o método D-BTVIR ao BTV e, de forma análoga, o método D-IRWIR ao IRWSR, com a intenção de validar a hipótese de que o uso do *Demons* como registro melhora o resultado quantitativo da super-resolução, particularmente quando considerados os cenários apresentados neste trabalho.

As hipóteses nula  $H_0$  e alternativa  $H_1$  para o teste de Wilcoxon são definidas como se segue:

$H_0$ : A mediana da diferença dos valores das métricas (PSNR ou SSIM) do método com registro *Demons* em relação ao método com registro paramétrico (ou seja, os valores das métricas referentes ao BTV subtraídas do D-BTVIR ou os valores referentes ao IRWSR subtraídos dos valores referentes ao D-IRWIR) é igual a zero;

$H_1$ : A mediana da diferença dos valores das métricas (PSNR ou SSIM) do método com registro *Demons* em relação ao método com registro paramétrico (ou seja, os valores das métricas referentes ao BTV subtraídas do D-BTVIR ou os valores referentes ao IRWSR subtraídos dos valores referentes ao D-IRWIR) é maior do que zero;

em que a hipótese alternativa  $H_1$  (ou, de forma equivalente, a rejeição da hipótese nula  $H_0$ ) indica que há evidências de diferença maior que zero entre as medianas das amostras, sugerindo uma melhoria estatisticamente significativa decorrente do uso dos métodos propostos.

A Tabela 5 mostra os valores- $p$  dos dois cenários, ou seja, para D-BTVIR comparado ao BTV e para D-IRWIR comparado ao IRWSR, sendo todos os valores significativamente menores do que 0,01. Isso indica que, para as duas comparações, há evidências que respaldam a hipótese alternativa  $H_1$ , ou seja, entre outras palavras, há evidências que sugerem que tanto o D-BTVIR quanto o D-IRWIR apresentam resultados melhores, em termos de PSNR e SSIM, do que o BTV e o IRWSR, respectivamente, com nível de significância 1%, o que valida a eficácia dos dois métodos propostos neste capítulo.

#### 4.2.1.2 Teste de Friedman e Nemenyi

Os testes de Friedman e Nemenyi, em oposição ao teste pareado de Wilcoxon, possibilitam uma análise conjunta dos quatro métodos: BTV, IRWSR, D-BTVIR e D-IRWIR. Ademais, uma vez que para tais testes os *ranks* dos métodos são considerados,

Tabela 5 – Resultado do teste de Wilcoxon para os métodos D-BTVIR e D-IRWIR, quando comparados aos métodos BTV e IRWSR, respectivamente

Métodos comparados	valor- $p$	
	para PSNR	para SSIM
D-BTVIR - BTV	$3,52 \times 10^{-7}$	$6,37 \times 10^{-9}$
D-IRWIR - IRWSR	$2,95 \times 10^{-8}$	$3,53 \times 10^{-12}$

Fonte: Elaborada pela autora.

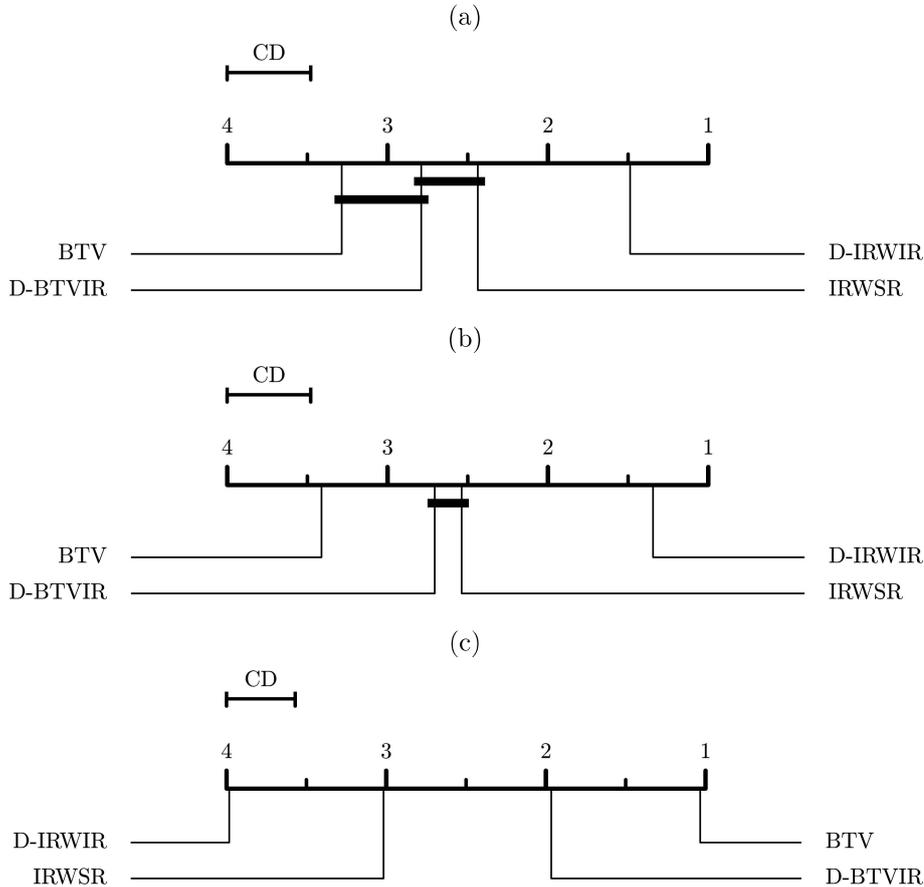
as diferentes escalas entre os valores das métricas não influenciam negativamente no resultado do teste, sendo possível concatenar os *rank* referentes aos valores de PSNR e SSIM na mesma tabela, gerando uma única análise em termos de qualidade das imagens super-resolvidas. Além desta análise, apresenta-se uma análise de tempo de processamento. Definem-se, para tal, a hipótese nula  $H_0$  e a hipótese alternativa  $H_1$ :

$H_0$ : A mediana dos *ranks*, em termos de qualidade ou tempo de processamento, é equivalente para os quatro métodos: BTV, IRWSR, D-BTVIR e D-IRWIR;

$H_1$ : Pelo menos dois dos quatro métodos tem medianas diferentes, em termos de qualidade ou tempo de processamento.

O valor- $p$ , quando consideradas os *ranks* de PSNR é igual a  $1,3110^{-26}$ , de SSIM é igual a  $5,44 \times 10^{-34}$  e, para o tempo de processamento, é de  $1,49 \times 10^{-75}$ , o que rejeita a hipótese nula  $H_0$  nos três casos. Aplica-se, dessa forma, o teste de Nemenyi com  $\alpha = 0,01$ , que resulta nos gráficos de distância crítica apresentados na Figura 29. Os gráficos da Figuras 29a e 29b mostram que o método D-IRWIR é estatisticamente superior aos outros três, estando classificado na melhor posição, para PSNR e SSIM, sem estar conectado a nenhum outro com a linha grossa (que indica distância entre *ranks* inferior à distância crítica). Considerando o SSIM (Figura 29b), o método BTV é o pior dos quatro, não sendo estatisticamente equivalente aos outros, e para o PSNR (Figura 29a) é estatisticamente equivalente ao D-BTVIR e ao IRWSR. Os métodos D-BTVIR e IRWSR estão conectados pela linha espessa nos dois gráficos supramencionados, indicando uma equivalência estatística entre eles, em termos de qualidade das imagens (PSNR e SSIM). Em relação ao tempo de processamento, por outro lado, o gráfico de distância crítica da Figura 29c não apresenta nenhuma linha espessa, sugerindo ausência de equivalência estatística entre os métodos para este quesito. Particularmente, considerando os valores de tempo de processamento do método IRWSR divididos pelos valores de tempo de processamento do método D-BTVIR, o método IRWSR é, em média, 7,2 vezes mais lento do que o método D-BTVIR o que, paralelo ao fato dos dois métodos serem estatisticamente equivalentes em termos de qualidade de imagem (PSNR e SSIM), indica um melhor compromisso entre tempo de processamento e qualidade do método D-BTVIR, proposto neste capítulo.

Figura 29 – Gráficos de distância crítica, referente ao teste de Nemenyi aplicado aos métodos BTV, IRWSR, D-BTVIR e D-IRWIR. Têm-se (a) o gráfico referente aos valores de PSNR, (b) o gráfico referente aos valores de SSIM e (c) o gráfico referente ao tempo de processamento.



Fonte: Elaborada pela autora.

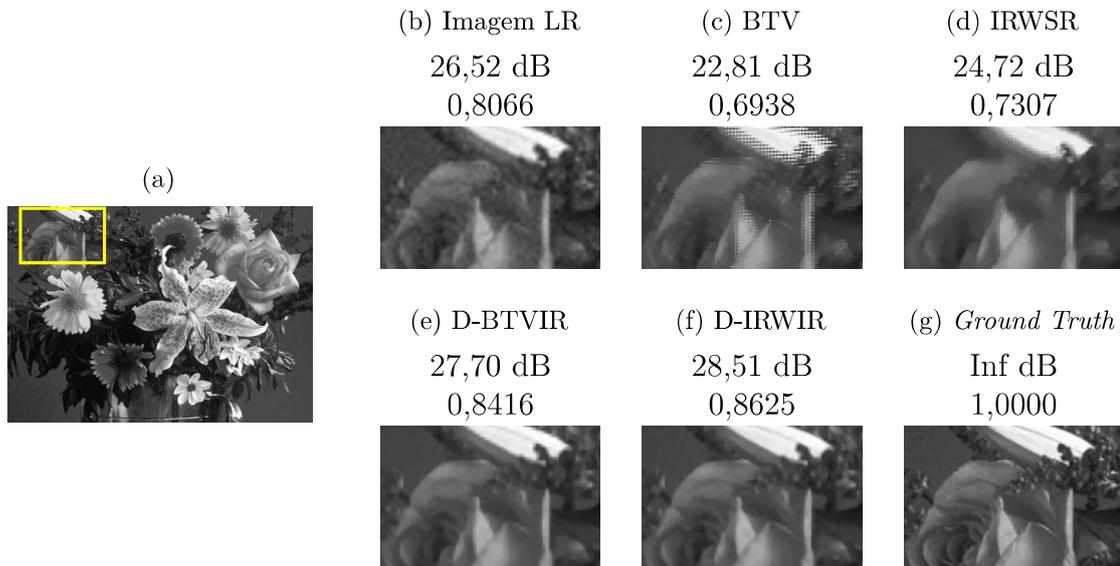
#### 4.2.2 Análise por Inspeção Visual

Apresenta-se, nesta seção, uma análise por inspeção visual, considerando alguns casos representativos dos testes executados. O objetivo é investigar as características visuais recorrentes nas imagens examinadas.

O resultado do BTV, na Figura 30c, apresenta presença intensa de artefatos, provavelmente devido ao erro de registro. Conforme mostrado na Figura 30d, o método IRWSR produz uma imagem com menos artefato na mesma região devido à robustez do método em relação ao erro de movimento, por outro lado, apresenta demasiada suavização. Ambos D-BTVIR (Figura 30e) e D-IRWIR (Figura 30f) produzem imagens sem artefatos, devido ao desempenho bem-sucedido do registro *Demons* e sem a demasiada suavização presente na imagem referente ao IRWSR.

A Figura 31 mostra as imagens referentes à figura 156065, do conjunto *B100*, em que se observa o rosto de uma mulher debaixo d'água com um par de óculos de mergulho e um

Figura 30 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *flowers*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.



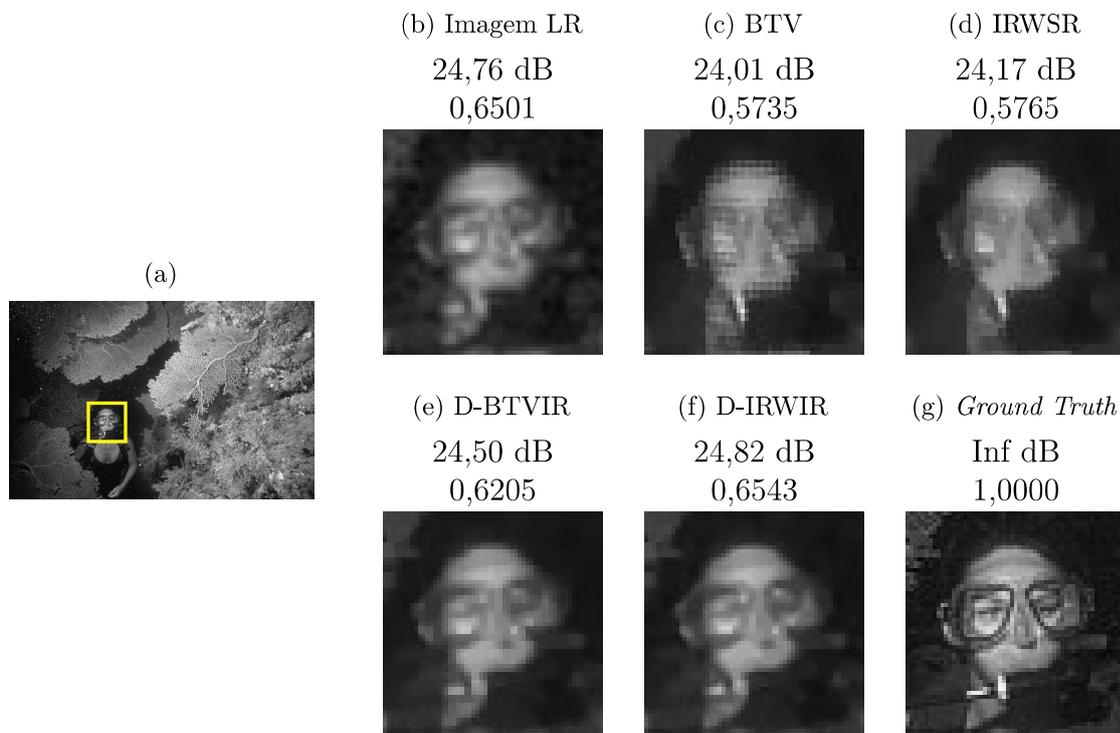
Fonte: Elaborada pela autora.

equipamento de mergulho na boca. Os resultados dos métodos BTV (Figura 31c) e IRWSR (Figura 31d) apresentam artefato expressivo nas bordas, que deformam significativamente o rosto da mulher presente na imagem. Por outro lado, os métodos D-BTVIR (Figura 31e) e D-IRWIR (Figura 31f) retornam imagens com melhor aspecto, em especial no caso do D-IRWIR.

A Figura 32 mostra os resultados referentes à imagem *ppt3* do *set14*, em que se observa na região destacada o texto “OSBORNE”. Neste caso, o texto é praticamente ilegível no resultado do método BTV (Figura 32c) bem como na imagem referente ao IRWSR (Figura 32d). Os resultados dos métodos D-BTVIR, na Figura 32e, e D-IRWIR, na Figura 32f, são similares entre si, nos quais a palavra se mostra legível. Entretanto, os resultados quantitativos não acompanham a análise visual, uma vez o IRWSR apresenta melhor SSIM do que o D-BTVIR, apesar do D-BTVIR ser visualmente melhor do que o IRWSR.

Para a região destacada na imagem *zebra*, do *Set14*, o resultado do método BTV (na Figura 33c) apresenta um artefato de difícil percepção visual, que remete a um possível erro de registro, enquanto a Figura 33d, referente ao método IRWSR, apresenta um aspecto cartunesco, apesar de não apresentar o mencionado artefato de deslocamento. Em contrapartida, os resultados dos métodos D-BTVIR (Figura 33e) e D-IRWIR (Figura

Figura 31 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução 156065, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.

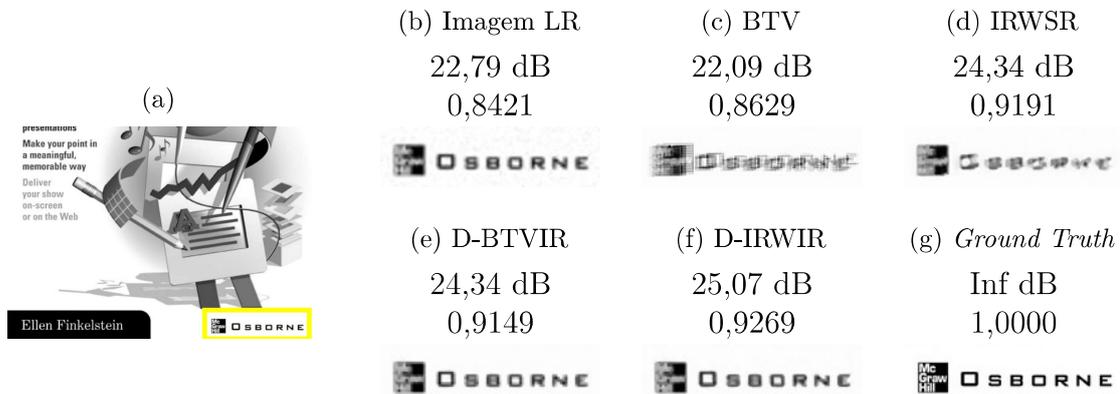


Fonte: Elaborada pela autora.

33f) são significativamente similares entre si, apresentando aspecto suavizado mas sem artefatos. Nota-se que o método IRWSR tem melhor PSNR do que o método D-BTVIR, mas tem pior SSIM, o que pode ser explicado pelas diferentes formas de cálculo das duas métricas (o SSIM se baseia em informações de vizinhança, diferentemente do PSNR, que é baseado na diferença média quadrática das imagens).

As duas últimas situações a serem analisadas, referentes às imagens *monarch*, do *Set14* e *37073*, do *B100*, são situações em que os métodos propostos neste capítulo, D-BTVIR e D-IRWIR, apresentam resultados visualmente inferiores ao método IRWSR. Para o primeiro caso, representado na Figura 34, o resultado do método BTV (Figura 34c) apresenta, como nas análises anteriores, artefato na região de borda, ocasionado por erro de registro. O método IRWSR (Figura 34d) apresenta o melhor resultado visual, sem artefatos e com a imagem de melhor aspecto. O resultado do D-BTVIR (Figura 34e) apresenta leve artefato na borda, enquanto o resultado do D-IRWIR apresenta aspecto demasiadamente suavizado, como observado na Figura 34f. Entretanto, o método IRWSR tem o segundo melhor resultado quantitativo.

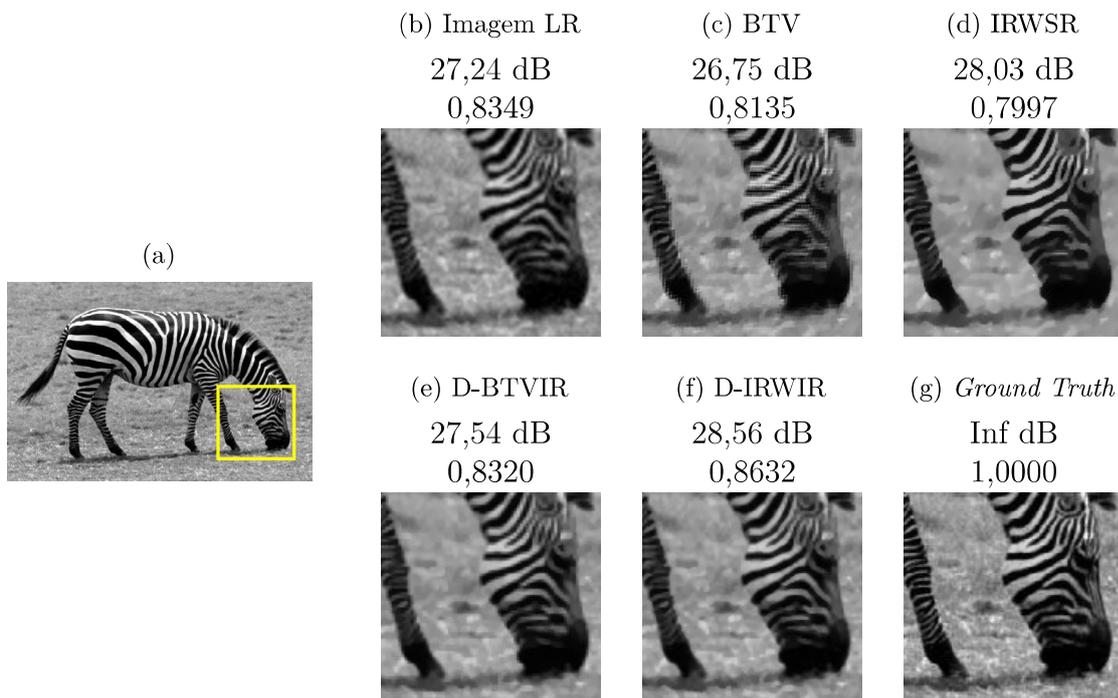
Figura 32 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *ppt3*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.



Fonte: Elaborada pela autora.

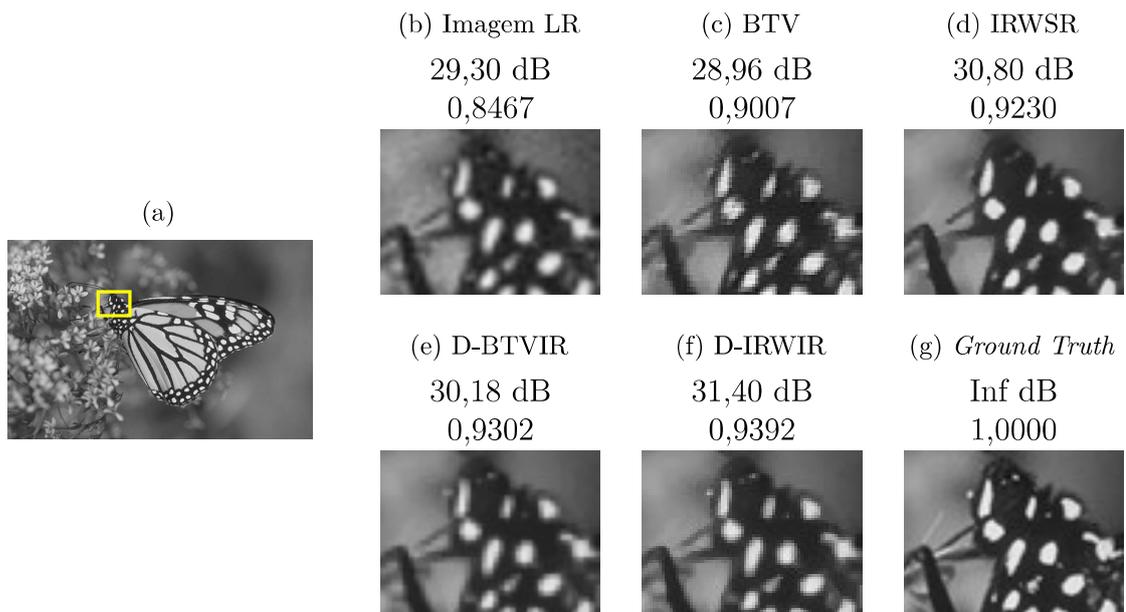
Finalmente, a Figura 35 mostra os resultados referentes à imagem *37073*, do *B100*, nos quais se observa um melhor desempenho do método IRWSR (Figura 35d), uma vez que o número “094” está mais legível para este caso, apesar dos artefatos de borda presentes na imagem. Em relação ao método BTV, a Figura 35c apresenta menos artefatos nas bordas e também é possível ler o número “094”. O método D-BTVIR (Figura 35e) retorna a imagem com pior aspecto visual, apresentando suavização demasiada e não sendo possível ler o número, podendo ser confundido com “034”. O mesmo ocorre na Figura 35f, do método D-IRWIR, que não contém artefato de borda mas, sim, um aspecto demasiadamente suavizado.

Figura 33 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *zebra*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.



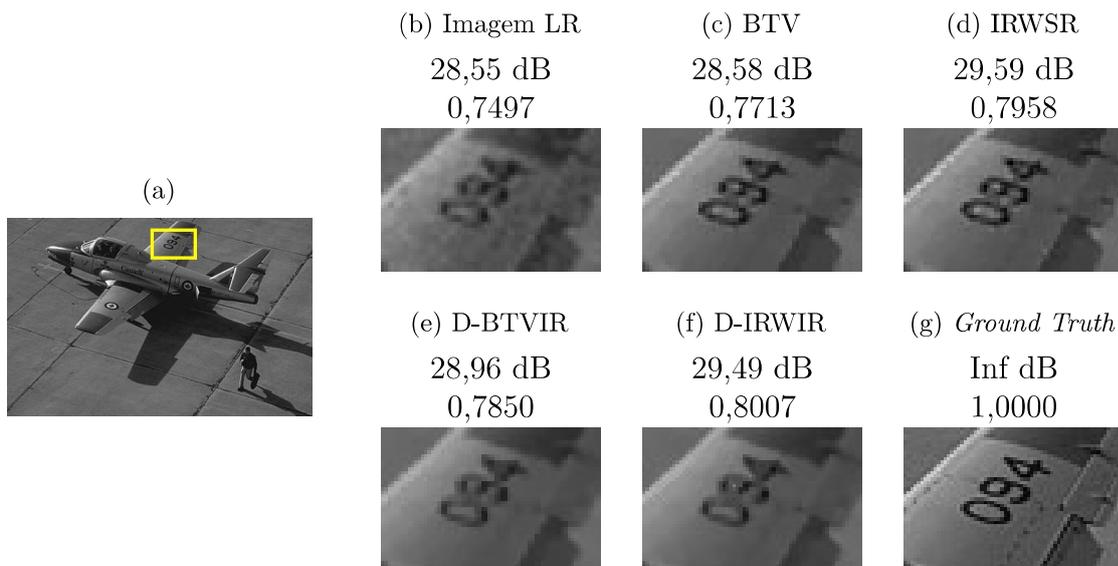
Fonte: Elaborada pela autora.

Figura 34 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *monarch*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.



Fonte: Elaborada pela autora.

Figura 35 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução 37073, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método D-BTVIR, (f) destaque da imagem super-resolvida pelo método D-IRWIR e (g) *ground truth*.

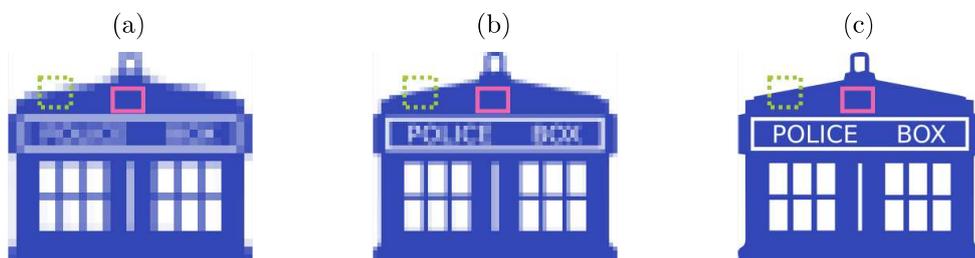


Fonte: Elaborada pela autora.

## 5 Super-Resolução Local com Registro *Demons*, Varredura em Patches e Classificação de Homogeneidade

A resolução espacial mede o menor detalhe discernível em uma imagem, quantitativamente representada em grandezas como pares de linhas por unidade de distância ou pontos (*pixels*) por polegada (*dpi*, do inglês, *dots per inch*) (GONZALEZ; WOODS, 2012). Nesse contexto, ela se relaciona com o quanto é possível expandir uma imagem, através do processo de interpolação, sem causar deformação na imagem interpolada. A Figura 36 mostra três versões de diferentes resoluções de uma mesma imagem, que foram redimensionadas através da interpolação pelo vizinho mais próximo para serem exibidas em uma mesma dimensão. Duas regiões foram destacadas nas imagens por um retângulo verde de borda pontilhada e um retângulo rosa de borda contínua. Observa-se que a região dentro do primeiro difere consideravelmente entre as imagens, de forma que há artefatos de borda nas Figuras 36b e 36a. Por outro lado, a região dentro do retângulo rosa de borda contínua é bastante similar nas três figuras. Nota-se que a região circundada pelo retângulo rosa contém, na realidade, menos informação (i.e., a cor azul) do que a região marcada pelo retângulo de borda pontilhada e, portanto, para aumentar a dimensão de tal região perfeitamente homogênea é suficiente copiar essa informação para os novos *pixels* que surgem no processo de sobre-amostragem.

Figura 36 – Exemplo ilustrativo de uma mesma imagem armazenada em diferentes resoluções: (a) 50 dpi, (b) 100 dpi e (c) 500 dpi.



Fonte: Elaborada pela autora

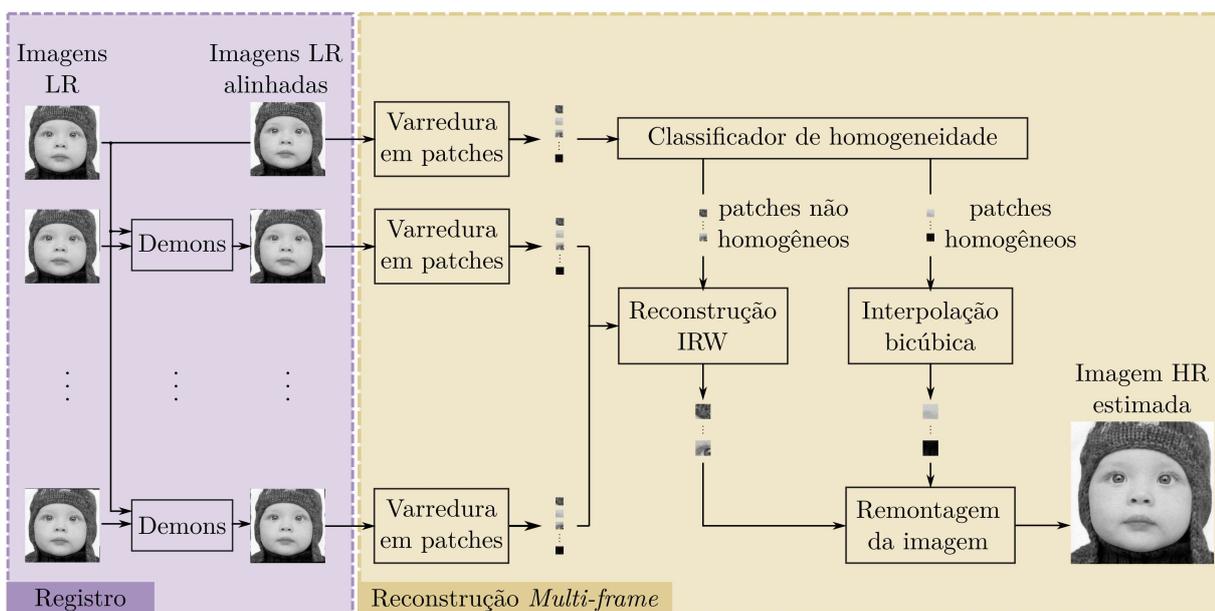
O exemplo apresentado no parágrafo anterior ilustra a intuição por trás da motivação da proposta a ser apresentada neste capítulo, que consiste em classificar os *patches* das imagens LR em homogêneos ou não e, a partir dessa classificação, estimar versões HR destes que, por sua vez, constituirão a imagem super-resolvida final. Propõe-se super-resolver os *patches* homogêneos utilizando interpolação, por esses conterem menos informação do que

aqueles classificados como não homogêneos que serão, então, super-resolvidos utilizando o método de super-resolução regularizado robusto IRW, proposto em (KÖHLER et al., 2016). A proposta deste capítulo se relaciona à hipótese (i) desta tese, ou seja, a hipótese de que o uso de informações locais pode favorecer o processo de super-resolução, particularmente, a informação em relação à homogeneidade em sub-regiões das imagens LR.

## 5.1 Método Proposto

A Figura 37 mostra o esquema simplificado do método proposto, denominado *Super-Resolução Local com Registro Demons, Varredura de Patches e Classificação de Homogeneidade*. As imagens LR são registradas, utilizando o método *Demons* (THIRION, 1998) gerando, assim, as imagens LR alinhadas. Em seguida, uma varredura de *patches* é aplicada nas imagens LR alinhadas, definida pela dimensão dos *patches* e pela sobreposição entre *patches* consecutivos. Os *patches* da imagem LR de referência são, então, classificados entre homogêneos e não homogêneos que são super-resolvidos por interpolação bicúbica e reconstrução IRW (KÖHLER et al., 2016), respectivamente. Essa classificação é feita a partir de métricas que descrevem a homogeneidade do *patch*, particularmente a variância e a suavidade, explicadas com detalhes na Seção 5.1.3. Finalmente, os *patches* HR são combinados para formarem a imagem HR estimada, de acordo com o esquema explicado na Seção 5.1.6.

Figura 37 – Esquema simplificado do método “Super-Resolução Local com Registro *Demons*, Varredura em Patches e Classificação de Homogeneidade”



Fonte: Elaborada pela autora.

### 5.1.1 Registro

O primeiro passo deste método é o registro, cuja eficiência é importante para garantir que o processo de varredura seja feito em todas as imagens LR de forma equivalente, sendo possível, assim, combinar os *patches* LR através do método de reconstrução *multi-frame* IRW, como será explicado na Seção 5.1.5. As vantagens do método *Demons*, exploradas no Capítulo 4, tais quais sua natureza não paramétrica e eficiência computacional, justificam o uso deste método de registro na proposta deste capítulo.

Uma das imagens LR (geralmente, a primeira delas) é definida como a imagem de referência, de forma que ela mesma já é a versão alinhada, como mostrada na Figura 37; as outras, uma a uma, são dadas como imagem deslocada, junto à imagem de referência, ao algoritmo *Demons* (explicado no Algoritmo 1) que, por sua vez, fornece como saída a versão alinhada de cada uma das imagens LR. Convém recordar que, além da imagem de referência e deslocada, têm-se como entrada o desvio padrão da suavização Gaussiana  $\sigma_{\text{diff}}$ , o número máximo de iterações  $I$  e o número de níveis do esquema multi-escalar  $L$ .

### 5.1.2 Varredura de *Patches*

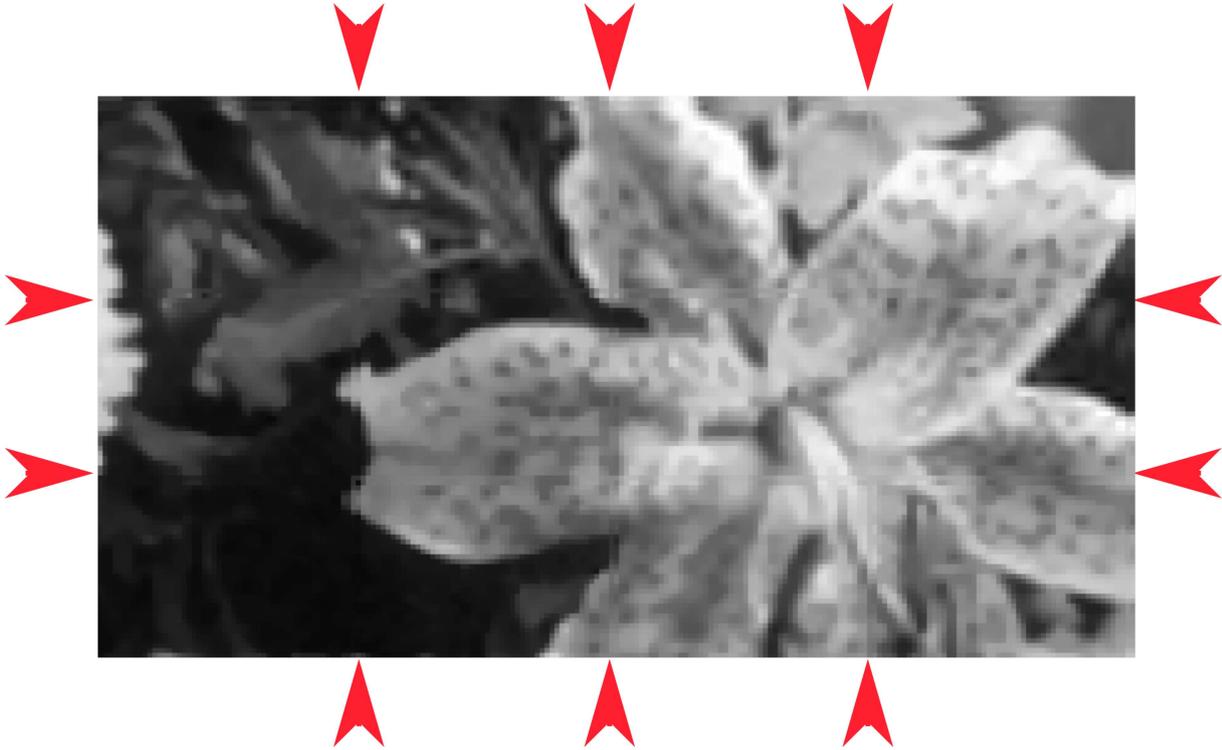
Após o registro, a varredura de *patches* é feita, da esquerda pra direita, de cima pra baixo, simultaneamente em todas as imagens LR alinhadas. A dimensão dos *patches* é definida dada a razão desejada entre o tamanho do *patch* e o tamanho da imagem. Como essa razão pode resultar em um número não inteiro de *pixels*, arredonda-se para baixo o resultado da multiplicação dela pela dimensão da imagem. Além disso, convencionou-se que caso a razão seja igual a 0 o *patch* será composto por um *pixel* e, em contrapartida, caso a razão seja igual a 1, o *patch* é composto pela imagem inteira, a partir do entendimento de que o usuário deseja o maior *patch* possível. Ou seja,

$$m_p = \begin{cases} \lfloor m \cdot r_p \rfloor, & \text{se } 0 < r_p < 1 \\ 1, & \text{se } r_p = 0 \\ m, & \text{se } r_p = 1 \end{cases} \quad e \quad n_p = \begin{cases} \lfloor n \cdot r_p \rfloor, & \text{se } 0 < r_p < 1 \\ 1, & \text{se } r_p = 0 \\ n, & \text{se } r_p = 1 \end{cases} \quad (5.1)$$

em que a dimensão dos *patches* é  $m_p \times n_p$ , a dimensão das imagens LR é  $m \times n$  e  $0 \leq r_p \leq 1$  é a razão aproximada desejada entre as dimensões dos *patches* e da imagem.

O problema de super-resolução, neste contexto, é dividido em sub-problemas que, por sua vez, tomam como entrada os *pixels* pertencentes às sub-regiões delimitadas pela varredura. Dessa forma, ainda que haja uma dependência entre *patches* consecutivos, dado que esses fazem parte de regiões locais próximas, os sub-problemas são resolvidos, na prática, de forma independente, sobretudo quando os *patches* em questão são classificados diferentemente. Em decorrência disso, pode haver uma discrepância nas regiões limítrofes entre *patches* causando um efeito indesejado na imagem estimada final, como pode ser observado na Figura 38, na qual se observam as linhas de encontro dos *patches*.

Figura 38 – Exemplo de super-resolução com varredura de *patches* em que se observa artefatos nas regiões limítrofes entre os *patches*, sinalizados pelas setas vermelhas.



Fonte: Elaborada pela autora.

Com o objetivo de evitar o artefato mencionado e observado na Figura 38, a varredura é executada considerando uma sobreposição entre *patches* consecutivos, de forma que a dependência existente entre eles esteja presente na implementação do método. Tal sobreposição  $r_s$  define, por sua vez, o valor do passo da varredura, ou seja, o número de *pixels* entre o início de dois *patches* consecutivos,

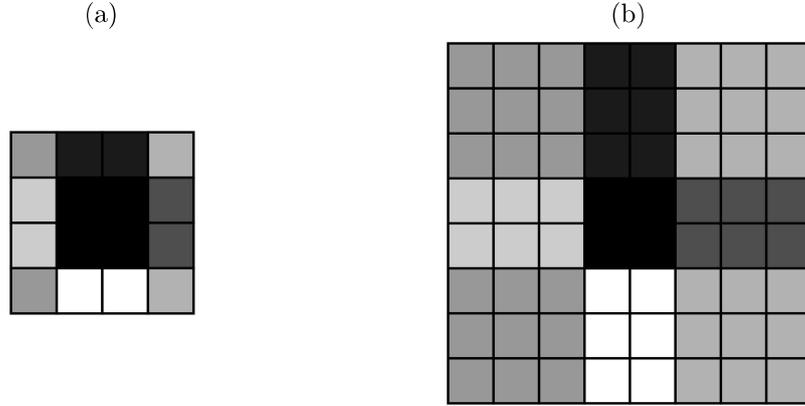
$$\beta_{s_h} = \lfloor (1 - r_s)m_p \rfloor \quad e \quad \beta_{s_v} = \lfloor (1 - r_s)n_p \rfloor, \quad (5.2)$$

em que  $\beta_{s_h}$  é o passo na direção horizontal,  $\beta_{s_v}$  é o passo na direção vertical e  $0 < r_s < 1$ . Faz-se esse cálculo (em oposição a se definir um valor inteiro para a sobreposição) para que o método possa lidar com imagens de tamanho diferentes, da mesma forma que o tamanho do *patch* também é dado como uma razão.

Em relação ao *padding*, processo realizado antes da varredura para garantir que não se perca informação nas regiões de borda da imagem, os *pixels* são replicados  $p_h$  vezes dos lados esquerdo e direito e  $p_v$  vezes em cima e embaixo das imagens LR alinhadas, como ilustrado na Figura 39. Os valores de  $p_h$  e  $p_v$  são definidos de forma que a dimensão total das imagens LR após *padding* menos a dimensão do *patch* seja múltipla do passo da varredura, para que esta seja efetuada nas imagens LR inteiras. Ou seja,  $p_h$  e  $p_v$  são os menores números inteiros tais que,

$$(m + 2p_h - m_p) \bmod \beta_{s_h} = 0 \quad e \quad (n + 2p_v - n_p) \bmod \beta_{s_v} = 0, \quad (5.3)$$

Figura 39 – Exemplo ilustrativo do *padding* realizado antes da varredura de *patches*, considerando  $p_h = p_v = 2$  e uma imagem de dimensão (a)  $4 \times 4$  antes de *padding* e dimensão (b)  $8 \times 8$  depois do *padding*.



Fonte: Elaborada pela autora.

em que mod é o operador resto da divisão inteira.

O *patch*  $l$  da  $k$ -ésima imagem LR alinhada é constituído pelos *pixels* da  $k$ -ésima imagem LR alinhada contidos na região definida pela varredura, ou seja,

$$\tilde{\mathbf{y}}_k^{(l)} = \{\tilde{\mathbf{y}}_k(u, v) \mid \mathbf{l}_h \leq u \leq \mathbf{l}_h + m_p - 1, \mathbf{l}_v \leq v \leq \mathbf{l}_v + n_p - 1\}, \quad (5.4)$$

em que  $\tilde{\mathbf{y}}_k$  é a  $k$ -ésima imagem LR alinhada,  $(\mathbf{l}_h, \mathbf{l}_v)$  é o *pixel* inicial do *patch*  $l$  e  $m_p \times n_p$  é a dimensão dos *patches*.

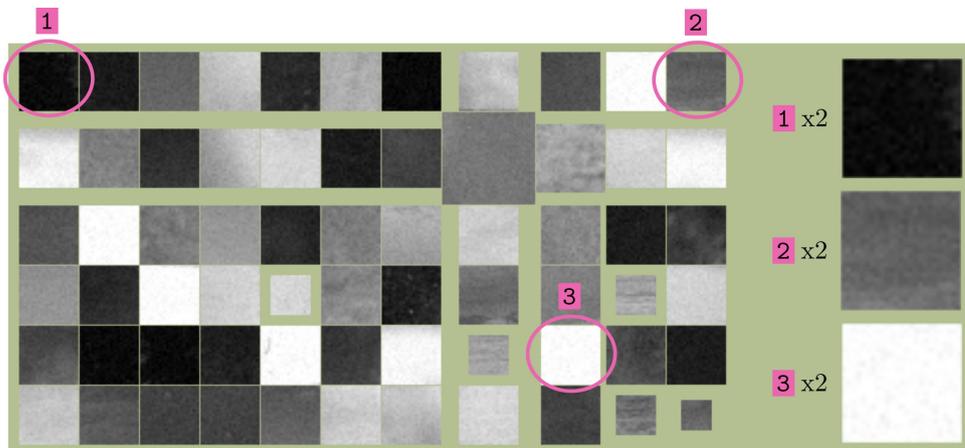
### 5.1.3 Classificação de Homogeneidade

A introdução deste capítulo apresenta um exemplo ilustrativo (Figura 36) em que se observa uma região homogênea satisfatoriamente representada em diferentes resoluções de uma mesma imagem. Tal exemplo trata de uma imagem sintética, em que a região em questão é perfeitamente homogênea, ou seja, os *pixels* presentes nela têm exatamente a mesma intensidade de cor. Essa não é, entretanto, uma característica comum em imagens naturais, adquiridas a partir de cenas reais, análogas a sinais reais analógicos. Em particular, as imagens de entrada do classificador de homogeneidade que se pretende construir são imagens sujeitas a uma série de deformações (como explicado na Seção 2.1.2) que incluem ruído aditivo, de forma que não seria coerente considerar *patches* idealmente homogêneos. Dessa forma, apresenta-se nesta seção o escopo no qual se compreende *patches* homogêneos e não homogêneos, bem como as duas métricas utilizadas para classificar tais sub-regiões. A classificação em si é feita a partir do valor da métrica escolhida que, sendo menor do que um dado limiar, caracteriza um *patch* homogêneo, caso contrário, tal *patch* é considerado não homogêneo.

Muitos pesquisadores da área de super-resolução destacam desafios relacionados à reconstrução de informações de borda, textura e demais estruturas que apresentam significativa mudança de intensidade de cor (NASROLLAHI; MOESLUND, 2014), (KHAT-TAB et al., 2020), (KATSAGGELOS; MOLINA; MATEOS, 2007), (MILANFAR, 2011). Nesse contexto, a intenção da classificação proposta aqui é considerar homogêneos aqueles *patches* menos desafiadores, ou seja, que possuem pouca ou nenhuma informação de tais estruturas, de forma que seja possível aumentar a resolução espacial dos mesmos a partir da aplicação de um método de interpolação mais simples do que os métodos de super-resolução conhecidos da literatura, uma vez que os últimos apresentam custo computacional significativo.

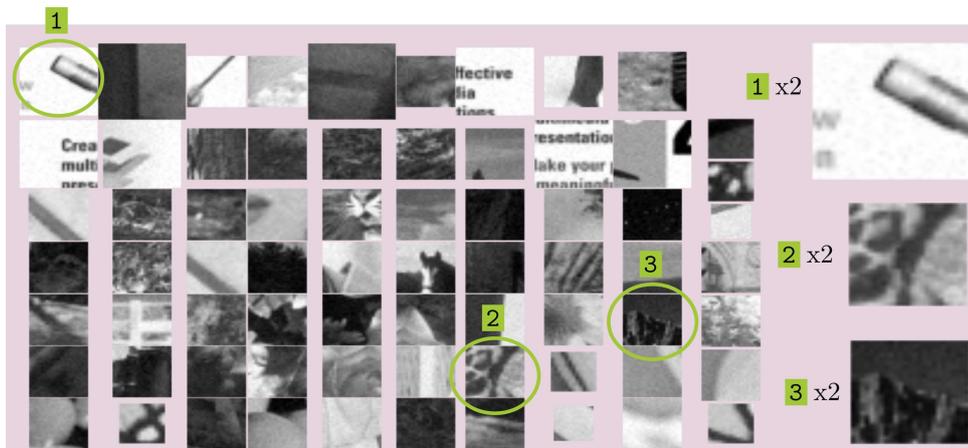
Apresenta-se uma metodologia para a classificação de homogeneidade em imagens. Foram extraídos e rotulados, a partir de análise visual, 679 *patches* homogêneos e 6.447 *patches* não homogêneos (totalizando 7126 *patches*) de imagens LR geradas a partir do modelo de aquisição utilizando os parâmetros apresentados no Capítulo 3 e alinhadas pelo algoritmo *Demons* utilizando os parâmetros apresentados no Capítulo 4. As Figuras 40 e 41 mostram alguns exemplos desses *patches*, além de três exemplos de versões interpoladas de cada tipo. Observa-se que os *patches* interpolados da primeira figura, ou seja, do tipo homogêneo, tem melhor aspecto do que os *patches* interpolados do tipo não homogêneo, o que ilustra o argumento discutido no parágrafo anterior.

Figura 40 – Exemplos de *patches* homogêneos.



Fonte: Elaborada pela autora.

Com a intenção de descrever quantitativamente a homogeneidade dos *patches* rotulados, foram escolhidas duas métricas, cujas distribuições dos valores foram analisadas para encontrar um valor de limiar que melhor separasse as duas classes de *patches* – homogêneos e não homogêneos. Destaca-se, porém, que a extração e rotulagem mencionada não pretende ser exaustiva nem definitiva, mas um meio de se encontrar uma faixa de valores para essas duas métricas, que serão utilizadas como hiperparâmetro do método,

Figura 41 – Exemplos de *patches* não homogêneos.

Fonte: Elaborada pela autora.

posteriormente submetido a um processo de validação explicado na Seção 5.2. Além disso, convém mencionar que, apesar dos métodos de aprendizado de máquina constituídos por uma etapa de treinamento serem bastante utilizados para tarefas de classificação no geral, acredita-se que um banco de dados mais volumoso seria necessário para um bom desempenho da classificação de homogeneidade pretendida.

### 5.1.3.1 Variância

Considerando o *patch* de índice  $l$  da  $k$ -ésima imagem LR alinhada  $\tilde{\mathbf{y}}_k$ , a variância amostral deste é dada pela Equação 5.5.

$$\sigma_l^2 = \frac{1}{m_p \cdot n_p - 1} \sum_{i=l_h}^{l_h+m_p-1} \sum_{j=l_v}^{l_v+n_p-1} (\tilde{\mathbf{y}}_k(i, j) - \mu_l)^2, \quad (5.5)$$

em que  $m_p \times n_p$  é a dimensão dos *patches*, o *pixel* inicial de tal *patch* está posicionado em  $(l_h, l_v)$  e  $\mu_l$  é a média dos valores dos *pixels* que constituem o *patch*  $l$ . A variância mede a dispersão dos valores das amostras de uma população em relação a média dessa população. Ou seja, quanto menor a variância das intensidades de cinza dos *pixels* presentes em um *patch*, mais homogêneo é esse *patch*.

### 5.1.3.2 Suavidade

Su et al. (2012) introduziu a métrica suavidade (do inglês, *smoothness*), calculada a partir do *structure tensor* que, por sua vez, indica as direções predominantes do gradiente de uma imagem na vizinhança de um dado ponto. A cada *pixel* das imagens LR alinhadas,

uma matriz  $\mathbf{Q}_{(u,v)}$  de dimensão  $2 \times 2$  é atribuída tal que

$$\begin{aligned} \mathbf{Q}_{(u,v)} &= \nabla \tilde{\mathbf{y}}_k(u, v) \cdot \nabla \tilde{\mathbf{y}}_k(u, v)^\top \\ &= \begin{bmatrix} \frac{\partial \tilde{\mathbf{y}}_k(u,v)}{\partial u} & \frac{\partial \tilde{\mathbf{y}}_k(u,v)}{\partial u} \frac{\partial \tilde{\mathbf{y}}_k(u,v)}{\partial v} \\ \frac{\partial \tilde{\mathbf{y}}_k(u,v)}{\partial v} & \frac{\partial \tilde{\mathbf{y}}_k(u,v)}{\partial v} \end{bmatrix}, \end{aligned} \quad (5.6)$$

em que  $(u, v)$  é a posição do *pixel* na imagem,  $\nabla \tilde{\mathbf{y}}_k$  é o gradiente da imagem LR alinhada  $\tilde{\mathbf{y}}_k$ ,  $\frac{\partial \tilde{\mathbf{y}}_k}{\partial u}$  é a derivada parcial no eixo  $u$  (direção horizontal) e  $\frac{\partial \tilde{\mathbf{y}}_k}{\partial v}$  é a derivada parcial no eixo  $v$  (direção vertical). O autor define, então, uma *structure matrix* para cada *patch*, a partir da média dos elementos das matrizes  $\mathbf{Q}_{(u,v)}$  dos *pixels* que constituem o *patch*, ou seja,

$$\mathbf{Q}_l = \frac{1}{m_p \cdot n_p} \sum_{u=l_h}^{l_h+m_p-1} \sum_{v=l_v}^{l_v+n_p-1} \mathbf{Q}_{(u,v)}, \quad (5.7)$$

em que  $\mathbf{Q}_l$  é a matriz  $2 \times 2$  que representa a estrutura geométrica do *patch*  $l$  que tem *pixel* inicial posicionado em  $(l_h, l_v)$ , cujos autovalores  $\lambda_1^{(l)}$  e  $\lambda_2^{(l)}$  medem a força do gradiente em duas direções perpendiculares. O maior autovalor corresponde à direção com gradiente mais forte e o menor corresponde à direção com o gradiente mais fraco. A suavidade  $\gamma_l$  do *patch*  $l$  é dada por,

$$\gamma_l = |\lambda_1^{(l)}| + |\lambda_2^{(l)}|, \quad (5.8)$$

em que quanto menor o valor de  $\gamma_l$ , mais homogêneo é o *patch*  $l$ .

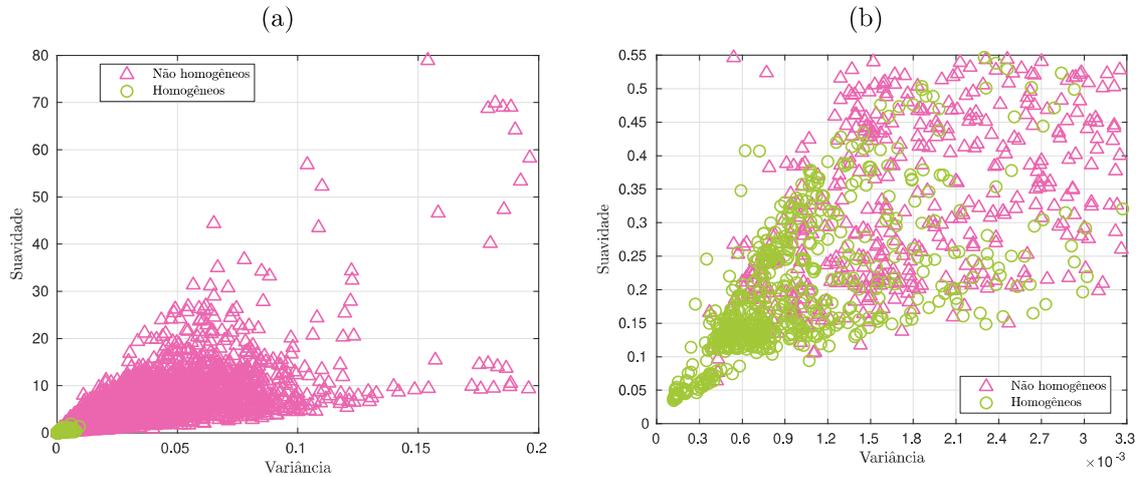
### 5.1.3.3 Distribuição dos Valores de Variância e Suavidade

O objetivo final do uso da variância e/ou da suavidade é definir um limiar que divida, da melhor forma possível, o grupo de *patches* homogêneos dos *patches* não homogêneos, de forma que a classificação seja feita a partir desse limiar. Assim, o objetivo desta seção é encontrar uma faixa de valores de variância e de suavidade para ser posteriormente validada. Isto é, as duas métricas são compreendidas como hiperparâmetros pelo processo de validação a ser apresentado na Seção 5.2.

A Figura 42 mostra a distribuição dos valores das métricas variância e suavidade de todos os 7.126 *patches* extraídos. Uma vez que deseja-se encontrar o melhor limiar entre os *patches* homogêneos e os não homogêneos, a região de maior interesse é aquela que compreende a maior parte dos *patches* homogêneos, sendo esta, a região mostrada na Figura 42b, que contém 89% deles. O domínio e contradomínio dessa região definem os candidatos a limiar dos valores de suavidade e variância a serem validados, sendo eles:

$$\begin{aligned} \gamma &= [0,05 \quad 0,10 \quad 0,15 \quad 0,20 \quad 0,25 \quad 0,30 \quad 0,35 \quad 0,40 \quad 0,45 \quad 0,5] \\ &\quad \text{e} \\ \sigma^2 &= [3 \quad 6 \quad 9 \quad 12 \quad 15 \quad 18 \quad 21 \quad 24 \quad 27 \quad 30] \times 10^{-4}. \end{aligned} \quad (5.9)$$

Figura 42 – Distribuição dos valores das métricas variância e suavidade (a) de todos os 7.126 *patches* extraídos e (b) dos *patches* que apresentam valores das métrica dentro das faixas escolhidas para fazer a validação dos hiperparâmetros.



Fonte: Elaborada pela autora.

#### 5.1.4 Interpolação Bicúbica

Os *patches* da imagem LR de referência classificados como homogêneos são super-resolvidos por interpolação utilizando o fator de magnificação dado como entrada do método de super-resolução. Isso porque, como explicado na introdução desta seção, esses *patches* tem menos informação e, por isso, parte-se da hipótese de que seja suficiente estimar a versão HR deles de forma mais simples com o objetivo principal de diminuir o tempo de processamento do método proposto. O exemplo da Figura 36 foi gerado utilizando a interpolação de vizinho mais próximo, entretanto, trata-se de uma imagem sintética, na qual as regiões sem bordas ou sem estruturas de alta frequência são perfeitamente homogêneas. No caso das imagens naturais, por outro lado, são raras as situações em que as regiões homogêneas são como esse caso ideal, de forma que considerar apenas essas situações não diminuiria consideravelmente o tempo de processamento do algoritmo, já que a proporção de classificações positivas em comparação às classificações negativas (ou seja, de regiões não homogêneas) seria baixa. No entanto, utilizar a interpolação pelo vizinho mais próximo em situações não idealmente homogêneas pode causar um efeito cartunesco na imagem, uma vez que essa interpolação mitiga as transições suaves de intensidade de cor, além dessa interpolação ser susceptível a ruído, como ilustrado na Figura 43.

Devido a essas desvantagens em se aplicar a interpolação por vizinho mais próximo, nas imagens naturais em que se considera ruído aditivo, utiliza-se a interpolação bicúbica para super-resolver os *patches* homogêneos. O procedimento é aplicado apenas à imagem LR de referência, ou seja, para os *patches* classificados como homogêneos, a informação das restantes das imagens LR é descartada. O cálculo de cada *pixel* da imagem interpolada

envolve a vizinhança  $4 \times 4$  ao redor dele (GONZALEZ; WOODS, 2012).

Figura 43 – Exemplo ilustrativo de imagem contaminada com ruído gaussiano branco aditivo interpolada por (a) interpolação bicúbica e (b) interpolação por vizinho mais próximo.



Fonte: Elaborada pela autora.

### 5.1.5 Reconstrução IRW

Os *patches* classificados como não homogêneos são super-resolvidos pela reconstrução IRW, baseada no método proposto em (KÖHLER et al., 2016), utilizada também no método proposto no Capítulo 4. Assim como naquele caso, o método IRW é escolhido aqui por ser robusto e por demandar a pré-seleção de apenas um hiperparâmetro (o fator  $P$ , que define a dimensão da janela em que a derivada primeira é calculada, no termo de regularização da função a ser otimizada), além de ser o estado-da-arte em super-resolução *multiframe* regularizada. Como as imagens de entrada para o método são previamente alinhadas pelo Registro *Demons*, é necessário passar como parâmetro de deslocamento a matriz identidade. Por isso, denomina-se *reconstrução* IRW, por não efetuar o alinhamento das imagens LR, mas sim a fusão, interpolação e reconstrução.

A estimação da versão em alta-resolução do *patch*  $l$  é dada pela Equação 5.10.

$$\begin{aligned} \hat{\mathbf{x}}^{(l)} = \arg \min_{\mathbf{x}^{(l)}} & (\tilde{\mathbf{y}}^{(l)} - \tilde{\mathbf{H}}^{(l)} \mathbf{x}^{(l)})^\top \mathbf{M}_\beta^{(l)} (\tilde{\mathbf{y}}^{(l)} - \tilde{\mathbf{H}}^{(l)} \mathbf{x}^{(l)}) + \\ & \lambda \sum_{j=-P}^P \sum_{k=-P}^P \|\mathbf{M}_\alpha^{(l)} (\mathbf{x}^{(l)} - S_h^j S_v^k \mathbf{x}^{(l)})\|_1, \end{aligned} \quad (5.10)$$

em que  $\tilde{\mathbf{y}}^{(l)}$  é o vetor em ordem lexicográfica obtido a partir da concatenação de  $\tilde{\mathbf{y}}_k^{(l)}$ ,  $k = 1, \dots, K$ , bem como  $\tilde{\mathbf{H}}^{(l)}$  é a matriz obtida a partir da concatenação de  $\tilde{\mathbf{H}}_k^{(l)}$ ,  $k = 1, \dots, K$ ,  $\mathbf{M}_\alpha^{(l)} = \text{diag}(\alpha_1, \dots, \alpha_{m_p n_p})$ ,  $\mathbf{M}_\beta^{(l)} = \text{diag}(\beta_1, \dots, \beta_{m_p n_p})$ ,  $m_p \times n_p$  é a dimensão de  $\hat{\mathbf{x}}^{(l)}$ , ou seja da versão super-resolvida do *patch*  $l$  e, finalmente,  $S_h^j$  e  $S_v^k$  são operadores que deslocam  $\mathbf{x}^{(l)}$  em  $l$  e  $m$  *pixels* nas direções horizontal e vertical, respectivamente, ao longo de uma janela de dimensão  $(2P + 1) \times (2P + 1)$ , com  $P \geq 1 \in \mathbb{Z}$ . Além disso,  $\tilde{\mathbf{H}}_k^{(l)} = \mathbf{D} \mathbf{B} \mathbf{W}_k^{(l)}$ , em que  $\mathbf{W}_k^{(l)}$  é formado considerando  $\tilde{\mathbf{s}}_k^{(l)} = \mathbf{I}_{3 \times 3}$ .

### 5.1.6 Remontagem da Imagem HR Estimada

A Figura 38 ilustra artefatos em decorrência da natureza local do método proposto, o que justifica a sobreposição entre *patches* adjacentes, como explicado na Seção 5.1.2. Neste contexto, a última etapa do método proposto tem o objetivo de remontar a imagem HR estimada final, a partir dos *patches* HR estimados na etapa anterior, considerando a mencionada sobreposição entre os *patches* adjacentes. Similarmente ao que é feito em (COSMO et al., 2018) e ilustrado na Figura 10, a intensidade de cada *pixel* da imagem HR estimada é influenciada pelos diferentes valores desse *pixel* nos *patches* em que faz parte, determinados pela varredura. Essa influência é definida por uma máscara Gaussiana de mesma dimensão dos *patches* ( $m_p \times n_p$ ) e com desvio padrão igual a  $0,5 \times m_p$  que pondera os *pixels* de tal forma que aqueles mais ao centro do *patch* tenham mais relevância. O valor final de cada *pixel*, então, é igual à combinação linear dos valores do *pixel* em todos os *patches* que faz parte multiplicados pelo fator definido pelas máscaras Gaussianas de tais *patches*.

## 5.2 Validação

A etapa de validação, amplamente utilizada no contexto de aprendizado de máquina, auxilia tanto na escolha do algoritmo mais apropriado para um determinado conjunto de dados quanto no ajuste de parâmetros (FACELI et al., 2011). Particularmente, neste trabalho, a validação é aplicada com os objetivos de (i) definir a métrica de homogeneidade, (ii) definir o valor do limiar de classificação, dentre os limiares candidatos (Equação 5.9), (iii) definir a razão entre a dimensão dos *patches* e da imagem LR e (iv) validar o uso da classificação de homogeneidade no método de super-resolução proposto. O conjunto *Train Set* (DONG et al., 2016), utilizado para o treinamento do método do Capítulo 3 e mostrado na Figura 16, é utilizado como conjunto de validação, enquanto os conjuntos *Set5*, *Set14* e *B100* são, posteriormente, utilizados nos experimentos de teste.

O procedimento de validação se inicia na aquisição das imagens LR, realizada de acordo com o mesmo modelo e parâmetros utilizados para o método apresentado na Seção 3.2. Utilizam-se 90 imagens do *Train Set* como *ground truth*. As imagens LR são, então, super-resolvidas pelo método proposto neste capítulo (esquematizado na Figura 37), inicializado com algumas variações dos hiperparâmetros, cujas escolhas fazem parte dos objetivos da validação, principalmente quanto à classificação de homogeneidade e à varredura de *patches*, culminando nos seguintes cenários:

1. classificador de homogeneidade considerando variância como métrica tomando o valor do limiar de homogeneidade  $\sigma_{th}^2$  variando nos candidatos  $\sigma^2$  da Equação 5.9;
2. classificador de homogeneidade considerando suavidade como métrica tomando o

- valor do limiar de homogeneidade  $\gamma_{th}$  variando nos candidatos  $\gamma$  da Equação 5.9;
3. sem classificação de homogeneidade, considerando todos os *patches* não homogêneos;
  4. sem classificação de homogeneidade, considerando todos os *patches* homogêneos.

Além disso, para todos os casos, a razão entre a dimensão do *patch* e da imagem LR ( $r_p$ ) varia dentro do conjunto de valores  $\{0, 1, 0, 2, 0, 5, 0, 8\}$ . Os parâmetros considerados para cada cenário estão resumidos no Quadro 5 e, a combinação de todos eles para os quatro cenários resulta no total de 88 variantes do método proposto.

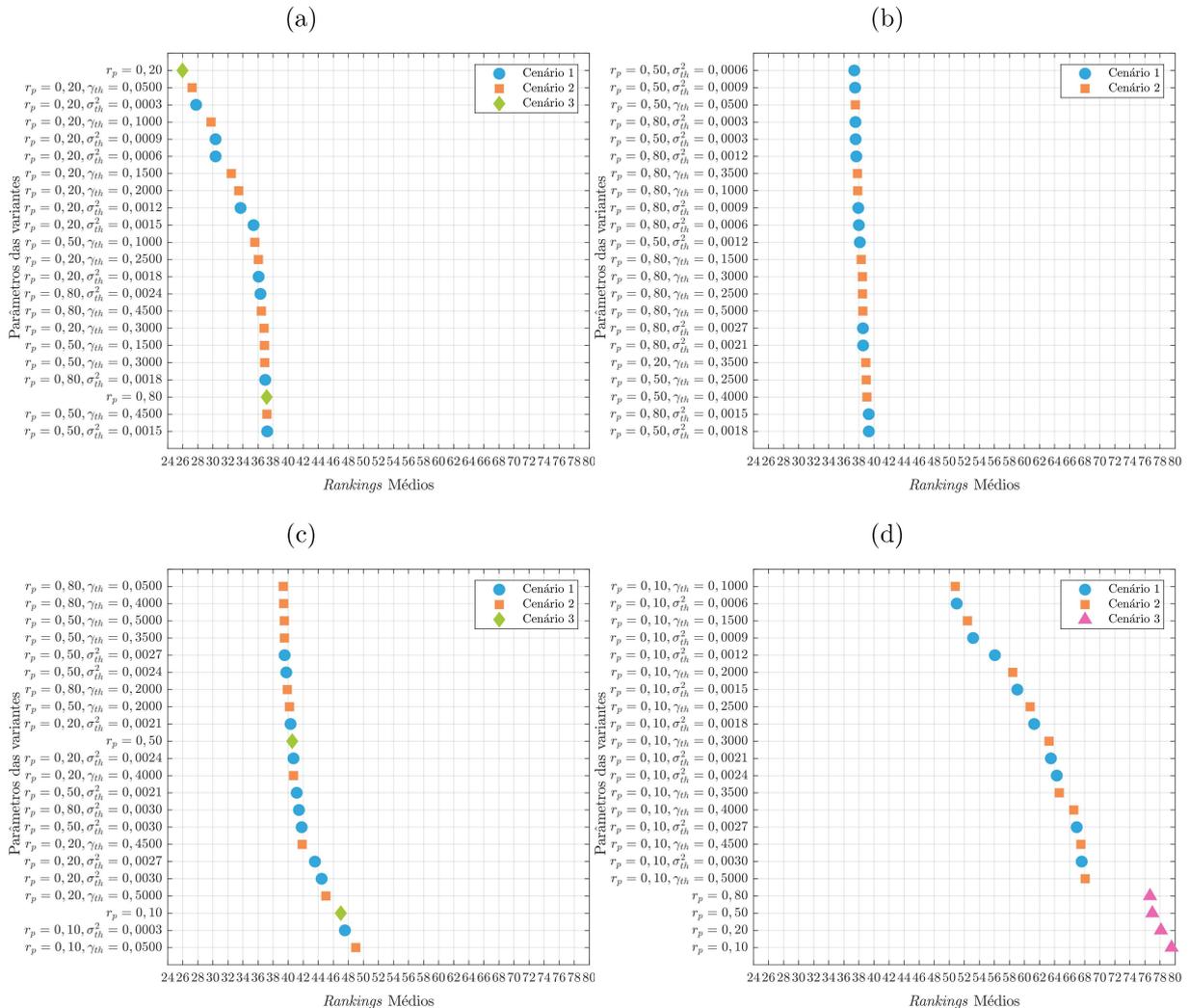
Quadro 5 – Cenários considerados para a validação do método proposto e respectivos parâmetros.

Cenário	Parâmetros
1 (40 variantes)	Métrica de homogeneidade: variância $\sigma^2 \in \{3, 6, 9, 12, 15, 18, 21, 24, 27, 30\} \times 10^{-4}$ $r_p \in \{0, 1, 0, 2, 0, 5, 0, 8\}$
2 (40 variantes)	Métrica de homogeneidade: suavidade $\gamma \in \{0, 05, 0, 10, 0, 15, 0, 20, 0, 25, 0, 30, 0, 35, 0, 40, 0, 45, 0, 50\}$ $r_p \in \{0, 1, 0, 2, 0, 5, 0, 8\}$
3 (4 variantes)	Todos <i>patches</i> considerados não homogêneos $r_p \in \{0, 1, 0, 2, 0, 5, 0, 8\}$
4 (4 variantes)	Todos <i>patches</i> considerados homogêneos $r_p \in \{0, 1, 0, 2, 0, 5, 0, 8\}$

A fim de alcançar os quatro objetivos da validação, mencionados no primeiro parágrafo desta seção, o desempenho médio das 88 variantes é analisado pela qualidade das imagens super-resolvidas, medida em PSNR e SSIM, e pelo tempo de processamento de cada variante para cada imagem. Uma vez que se trata de um número considerável de amostras (90 resultados para tempo de processamento e 180 para qualidade das imagens) e métodos comparados (as 88 variantes supramencionadas), faz-se uso do *rank* médio dos métodos, da mesma forma que é feito para os testes de Friedman e Nemenyi. Dessa forma, para cada imagem e métrica, as variantes são ordenadas, resultando em três tabelas de dimensão  $90 \times 88$ . As tabelas de PSNR e SSIM são concatenadas em uma só tabela de dimensão  $180 \times 88$  que representa o desempenho das variantes em termos de qualidade de imagem. O *rank* médio de cada variante é calculada, resultando em um vetor de 88 elementos, representado nos gráficos de dispersão da Figura 44.

Os gráficos da Figura 44 foram obtidos a partir do particionamento, em quatro partes, do vetor de *rank* médio de qualidade (PSNR e SSIM) para fins de melhor visualização, de forma que a Figura 44a mostra as variantes da 1ª à 22ª colocada, a Figura 44b mostra da 23ª ao 44ª colocada, a Figura 44c mostra da 45ª ao 66ª colocada e a Figura 44d

Figura 44 – *Ranks* médios das 88 variantes consideradas: em (a) o 1° ao 22° colocado, em (b) o 23° ao 44° colocado, em (c) o 45° ao 66° colocado e em (d) o 67° ao 88° colocado. Os círculos azuis referem-se aos resultados do Cenário 1, os quadrados laranjas referem-se ao Cenário 2, os losangos verdes referem-se ao Cenário 3 e, finalmente, os triângulos rosas referem-se ao Cenário 4.



Fonte: Elaborada pela autora.

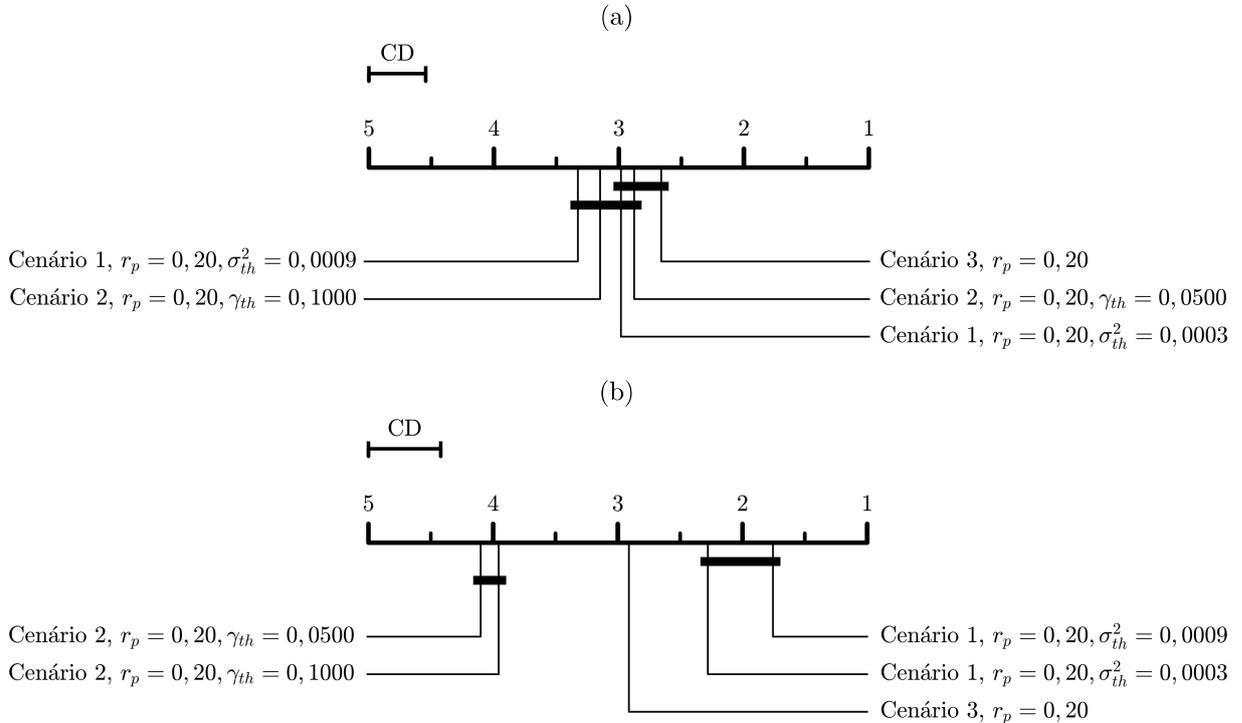
mostra, finalmente, da 67<sup>a</sup> ao 88<sup>a</sup> colocada. Os resultados referentes ao Cenário 1 (ou seja, variância como métrica de homogeneidade) são mostrados em círculos azuis, os resultados referentes ao Cenário 2 (suavidade como métrica de homogeneidade) são representados pelos quadrados laranja, os losangos verdes referem-se ao Cenário 3 (sem classificação, como todos os *patches* considerados não homogêneos) e os triângulos rosas referem-se ao Cenário 4 (sem classificação, com todos os *patches* considerados homogêneos). Além disso, no eixo *y* tem-se os valores de  $r_p$  e do limiar de classificação ( $\gamma_{tm}$  ou  $\sigma_{th}^2$ ), quando for o caso, enquanto o eixo *x* apresenta a colocação de cada variante – ou seja, quanto mais à esquerda o ponto estiver, melhor é o desempenho da variante referente.

Observa-se, no gráfico da Figura 44d, que as variantes em que todos os *patches* são considerados homogêneos apresentam os piores desempenhos, o que era esperado, uma vez que a interpolação bicúbica foi utilizada para super-resolver todos os *patches* e esta não reconstrói as informações de borda e textura, tendo pior desempenho do que a reconstrução IRW (usada para super-resolver os *patches* não homogêneos). Além disso, o domínio do eixo  $x$ , ou seja, dos *ranks* médios está contido entre 24 e 80. Particularmente, o 1º colocado (o losango verde no canto superior esquerdo da Figura 44a) tem *rank* médio igual a 25, 96, enquanto o 88º colocado (o triângulo rosa no canto inferior direito da Figura 44d) tem *rank* médio igual a 79, 54. Isso indica que, ao longo das 180 amostras (90 imagens com métricas PSNR e SSIM), existiram muitos empates e/ou os *ranks* das variantes divergiram entre si pois, caso contrário, o domínio das classificações médias teria uma distribuição mais ampla entre 1 e 88. Entretanto, o teste de Friedman aplicado às amostras retornou valor- $p$  aproximadamente nulo, indicando que a hipótese nula de que todas as 88 variantes são equivalentes pode ser rejeitada, o que respalda o uso das classificações médias para a validação pretendida. Para escolher os hiperparâmetros, observam-se as 10 variantes melhores classificadas na Figura 44a para as quais se nota que o valor de  $r_p$  é 0, 20, sendo este, então, o valor definido para tal hiperparâmetro – objetivo (iii) da validação, mencionado no primeiro parágrafo desta seção.

Por outro lado, para definir a métrica de homogeneidade, o limiar de classificação, bem como avaliar o uso da classificação segundo os experimentos efetuados, uma nova análise é feita considerando apenas as 5 melhores variantes da análise referente à Figura 44, na qual os valores de PSNR, SSIM e tempo de processamento são reordenados para cada imagem, gerando novas tabelas de *ranks* médios, uma referente ao desempenho em qualidade (PSNR e SSIM), com dimensão  $180 \times 5$  e outra referente ao desempenho em tempo de processamento, com dimensão  $90 \times 5$ . O teste de Friedman foi aplicado a essas duas tabelas, sendo que o valor- $p$  para os dados de qualidade é  $8,92 \times 10^{-4}$  e, para os dados de tempo de processamento, o valor- $p$  é igual a  $1,10 \times 10^{-31}$ , o que rejeita a hipótese nula de que as 5 variantes são estatisticamente equivalentes, sobretudo em relação ao tempo de processamento. Nesse caso, o teste de Nemenyi foi aplicado, gerando os gráficos de distância crítica apresentados na Figura 45.

Nota-se, na Figura 45a, que a variante com melhor desempenho em termos de qualidade é aquela referente ao Cenário 3, seguida da variante do Cenário 2 com limiar de classificação  $\gamma_{th} = 0,0500$  e, finalmente, a variante do Cenário 1 com limiar de classificação  $\sigma_{th}^2 = 0,0003$ . Entretanto, as três variantes mencionadas estão conectadas por uma linha mais espessa, que indica que a distância entre elas é inferior à distância crítica, de forma que, segundo o teste de Nemenyi, são consideradas estaticamente similares. Em contrapartida, ao observar a Figura 45b, observa-se que o Cenário 3 ocupa a terceira posição em tempo de processamento, e as variantes do Cenário 2 ocupam a quarta e quinta posição (com distância significativamente menor do que a crítica), o que indica uma pequena vantagem

Figura 45 – Gráficos de distância crítica referente ao teste de Nemenyi aplicado às cinco variantes melhores classificadas nos experimentos decorridos para a validação, em (a) o gráfico referente aos valores de qualidade de imagem, PSNR e SSIM e em (b) o gráfico referente ao tempo de processamento.



Fonte: Elaborada pela autora.

das variantes referentes ao Cenário 1, ou seja, da variância como métrica de homogeneidade. Em termos de limiar de classificação, define-se  $\gamma_{th} = 0,05$  para suavidade e  $\sigma_{th}^2 = 0,0003$  para variância em decorrência do melhor desempenho em termos de qualidade de imagem. Finalmente, uma vez que a Figura 45a indica um empate estatístico entre as três melhores variantes, os experimentos de testes foram executados considerando-as e seus respectivos parâmetros.

### 5.3 Experimentos e Resultados

Os experimentos efetuados para avaliar o desempenho do método proposto neste capítulo seguem a mesma metodologia dos experimentos dos Capítulos 3 e 4, ou seja, a bateria de experimentos foi aplicada às imagens dos conjuntos *Set5* (BEVILACQUA et al., 2012), *Set14* (ROMANO; ISIDORO; MILANFAR, 2017) e *B100* (MARTIN et al., 2001), utilizando, além do método proposto, os métodos BTV (FARSIU et al., 2004) e IRWSR (KÖHLER et al., 2016) como comparação. Os testes de Friedman e Nemenyi foram aplicados aos resultados das métricas de qualidade de imagem (PSNR e SSIM) bem como aos valores de tempo de processamento. Além disso, uma análise por inspeção visual

é apresentada, para serem analisados as vantagens e desvantagens do método proposto. A partir da validação apresentada na Seção 5.2 são testadas três variantes do método proposto, referido como PB (sigla para *patch based*) chamadas de (i) *PB - sem classificação*, referente ao método proposto feito sem classificação de homogeneidade, com razão entre a dimensão dos *patches* e das imagens LR de entrada  $r_p = 0,2$ ; (ii) *PB - suavidade*, considerando classificação de homogeneidade com a métrica suavidade com limiar de classificação  $\gamma_{th} = 0,05$  e  $r_p = 0,2$ ; e (iii) *PB - variância*, considerando classificação com variância como métrica de homogeneidade com limiar de classificação  $\sigma_{th}^2 = 0,0003$  e  $r_p = 0,2$ .

O Quadro 6 mostra os parâmetros referente ao modelo de aquisição utilizado para gerar as imagens LR a partir das imagens consideradas como *ground truth*. Além disso, para o método IRWSR, para a etapa IRW do método PB e para o método BTV, o parâmetro que define a janela do cálculo da derivada primeira no termo de regularização é  $P = 2$ ; para o método BTV, o fator de regularização é  $\lambda = 0,005$  e o fato de escala  $\alpha = 0,5$ ; e para o registro *Demons* usado nas variantes do método PB, o número máximo de iterações é  $I = 100$ , o número de níveis do esquema multi-escalar é  $L = 3$  e a regularização Bayesiana tem desvio padrão  $\sigma_{diff} = 1,0$ .

Quadro 6 – Parâmetros do modelo de aquisição utilizado nos experimentos referentes à super-resolução local, com registro *Demons*, varredura de *patches* e classificação de homogeneidade proposta.

Parâmetro	Valor
Número de imagens LR	8
Largura da PSF	0,4
Desvio padrão do ruído aditivo	0,02
Limite da definição aleatória do vetor de translação	$[(-2, -2), (2, 2)]$
Limite da definição aleatória do ângulo de rotação	$[-1^\circ, 1^\circ]$

Fonte: Elaborada pela autora.

### 5.3.1 Comparação Estatística

Os testes de Friedman e Nemenyi são usados, neste capítulo, para comparar os métodos BTV, IRWSR, PB - sem classificação, PB - variância e PB - suavidade. A partir dos valores de PSNR, SSIM e tempo de processamento, os cinco métodos são comparados, para cada imagem, resultando em três tabelas de *ranks* com 119 linhas e 5 colunas cada. A partir dessas tabelas, o teste de Friedman e, se for o caso, de Nemenyi são aplicados. Define-se, nesse sentido, as hipóteses nula  $H_0$  e alternativa  $H_1$ :

$H_0$ : A mediana dos *ranks* considerando PSNR, SSIM ou tempo de processamento é equivalente para os cinco métodos;

$H_1$ : Pelo menos dois dos cinco métodos tem medianas diferentes, considerando qualidade PSNR, SSIM ou tempo de processamento.

O valor- $p$  considerando PSNR resultou em  $5,11 \times 10^{-28}$ , considerando SSIM resultou em  $2 \times 10^{-37}$  e, considerando tempo de processamento,  $4,42 \times 10^{-92}$ . Ou seja, a hipótese nula  $H_0$  pode ser rejeitada nos três casos. Dessa forma, aplica-se o teste de Nemenyi, resultando nos gráficos de distância crítica mostrados na Figura 46. Observa-se, nos gráficos referentes ao PSNR e SSIM (Figuras 46a e 46b), que as três variantes do método PB tem melhor *rank* médio, sendo estatisticamente equivalentes entre si, uma vez que são ligadas pelas linhas mais espessa, indicando que diferem de uma distância menor do que CD (i.e, a distância crítica). O *rank* médio dessas variantes, que estão em torno de 2,5 demonstra melhor desempenho estatístico da abordagem proposta neste capítulo, em relação tanto ao IRWSR, quanto ao BTV. Em termos de processamento, o método BTV tem *rank* médio igual a 1 e o IRWSR tem *rank* médio igual a 2, o que está dentro do esperado uma vez que o método PB constitui, no fim das contas, na resolução de vários subproblemas, devido à varredura de *patches*. Por outro lado, entre as variantes do método proposto, a PB - variância demonstra melhor desempenho em termos de tempo de processamento, com *rank* médio igual a aproximadamente 3,2, indicando uma melhor relação de compromisso entre os métodos comparados.

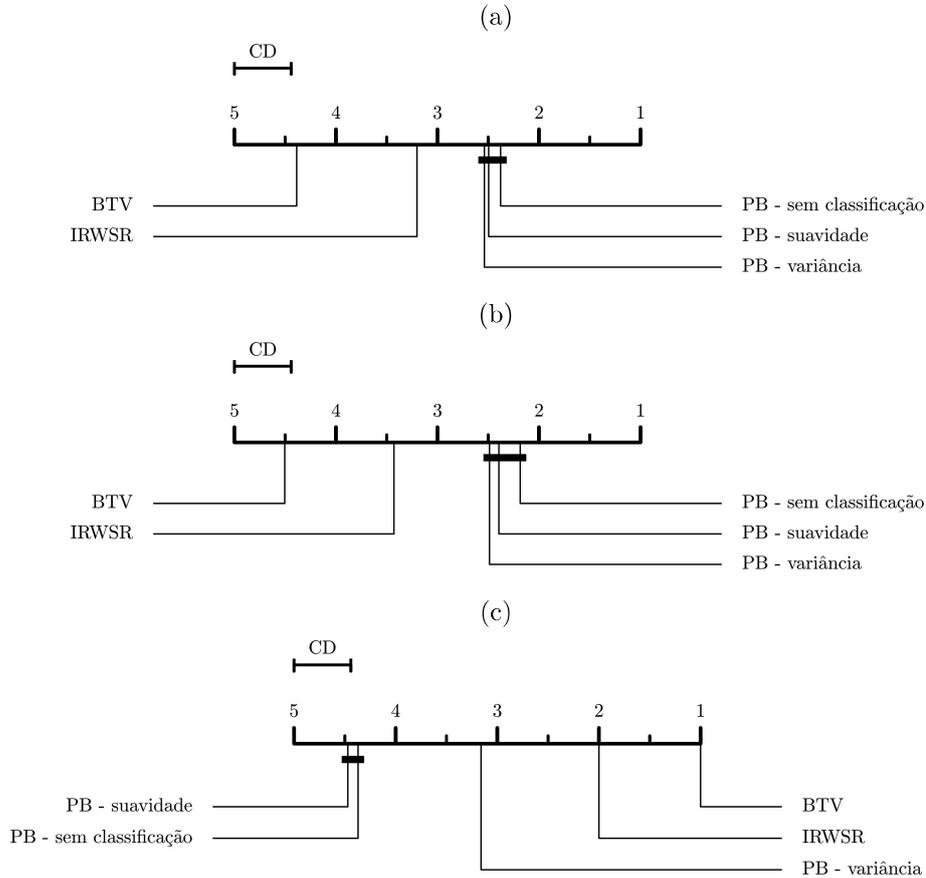
### 5.3.2 Análise por Inspeção Visual

A análise por inspeção visual apresentada nesta seção tem o objetivo de identificar características visuais recorrentes nas imagens resultantes dos métodos PB - sem classificação, PB - variância e PB - suavidade, propostos neste capítulo, a fim de identificar as vantagens e desvantagens dos mesmos.

A Figura 47 mostra as imagens super-resolvidas referentes à imagem *156065* do conjunto *B100*. Observa-se na região destacada pelo retângulo amarelo contínuo da Figura 47a, um rosto de uma mulher de óculos de mergulho e com um aparato na boca. Tanto a imagem resultante do método BTV (Figura 47c), quanto a imagem do método IRWSR (Figura 47d) apresentam deformações severas no rosto presente na imagem com, respectivamente, artefato e demasiada suavização nas bordas sendo, inclusive, difícil observar os óculos de mergulho no rosto da mulher. Por outro lado, as imagens estimadas pelas variações do método PB proposto neste capítulo não apresentam artefatos, além de apresentarem melhor reconstrução de bordas, apesar de ainda ser visível demasiada suavização. Tais características visuais respaldam os valores da avaliação quantitativa, uma vez que os métodos PB têm os melhores resultados de PSNR e SSIM.

A Figura 48 mostra os resultados referentes à imagem *ppt3*, do *Set14*, em que se observa o texto “OSBORNE”, ilegível na Figura 48d, resultante do método IRWSR e com

Figura 46 – Gráficos de distância crítica, referente ao teste de Nemenyi aplicado aos métodos BTV, IRWSR, PB - sem classificação, PB - variância e PB - suavidade. Têm-se, (a) o gráfico referente aos valores de PSNR, (b) o gráfico referente aos valores de SSIM e (c) o gráfico referente ao tempo de processamento.



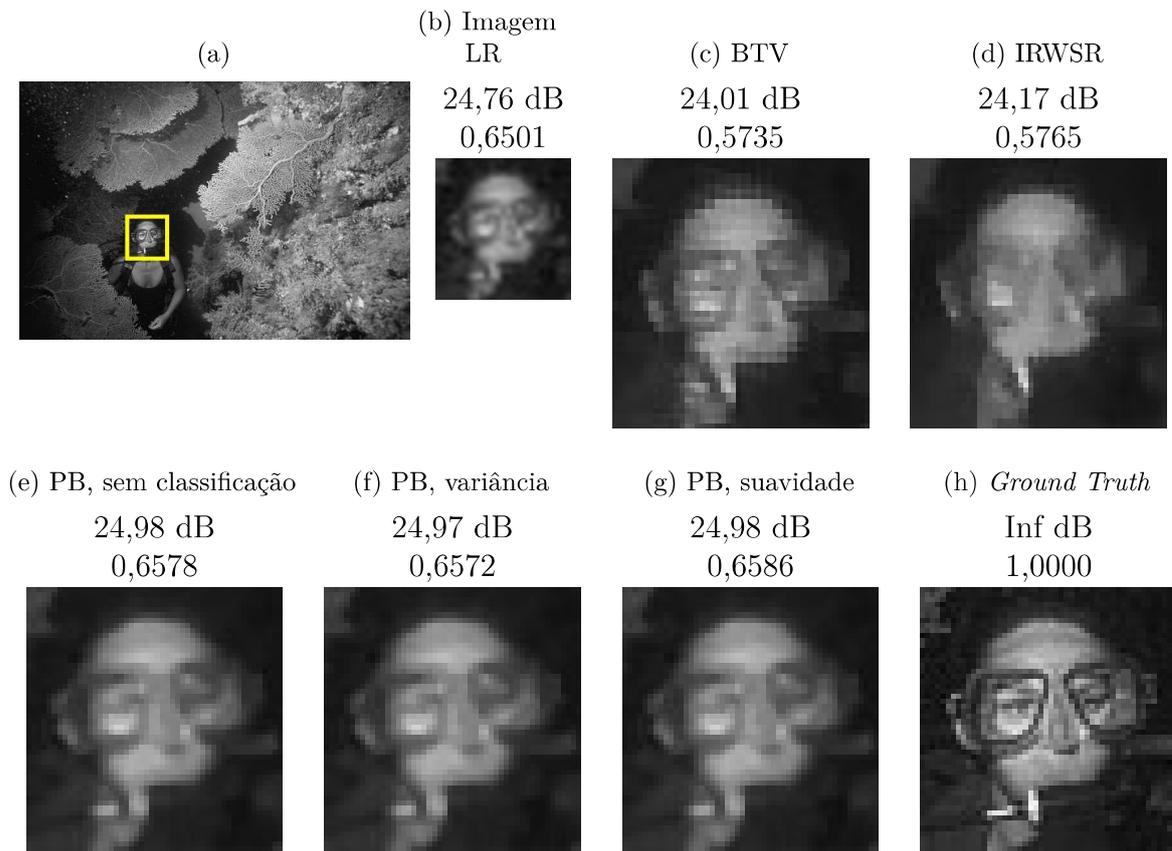
Fonte: Elaborada pela autora.

significativo artefato na Figura 48c, resultante do método BTV. Por outro lado, o texto está legível nas Figuras 48e, 48f e 48g, referentes aos métodos PB - sem classificação, PB - variância e PB - suavidade, respectivamente. Além disso, as três imagens dos métodos PB não apresentam artefatos ou deformações de movimento.

A Figura 49 mostra os resultados referentes à imagem *flowers* do *Set14*, em que análise visual respaldam os resultados qualitativos, uma vez que o resultado do BTV, na Figura 49c, apresenta significativo artefato nas bordas, aparentemente devido à deformação por movimento. O método IRWSR retorna uma imagem (na Figura 49d) sem tais artefatos, com aspecto borrado que também aparenta ser devido à deformação por movimento. Por outro lado, as três imagens referentes às variantes do método PB, ou seja, aquelas mostradas nas Figuras 49e, 49f e 49g, não apresentam artefato ou aspecto borrado. Nota-se, sobre as três últimas, resultados visualmente similares.

A figura referente à imagem *zebra* (Figura 50) do *Set14* apresenta resultados

Figura 47 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução 156065, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.

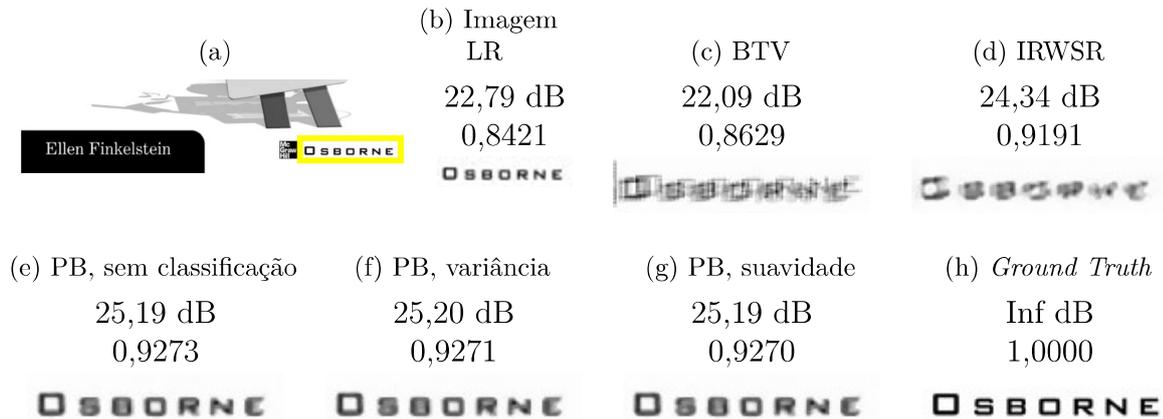


Fonte: Elaborada pela autora.

similares aos da imagem *flowers*. Nesse caso, o resultado do BTV (Figura 50c) apresenta significativo artefato nas bordas e o resultado do IRWSR (Figura 50d) apresenta efeito cartunesco e suavização demasiada na textura da grama. Por outro lado, os métodos PB - sem classificação (Figura 50e), PB - variância (Figura 50f) e PB - suavidade (Figura 50g) fornecem imagens com melhores aspectos visuais, similares entre si, com melhor reconstrução da textura da grama e sem artefatos nas bordas.

A Figura 51 mostra, por outro lado, uma situação em que, apesar do método PB - variância apresentar o melhor PSNR e o método PB - sem classificação apresentar o melhor SSIM, a análise visual aponta um melhor resultado do IRWSR. O resultado do método BTV (Figura 51c) apresenta artefato nas bordas, as três variantes do PB (Figuras 51e, 51f e 51g) apresentam suavização demasiada das estruturas de borda e o resultado do IRWSR (Figura 51d) apresenta efeito cartunesco, porém com uma imagem mais nítida.

Figura 48 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *ppt3*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.

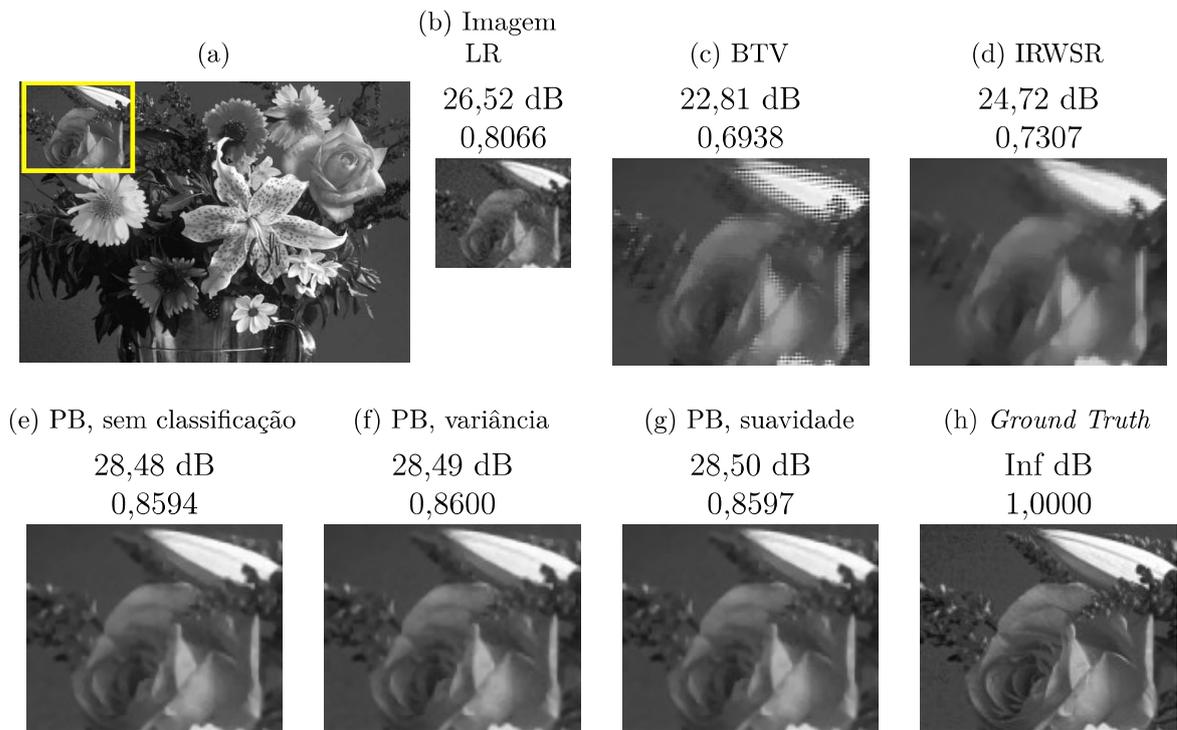


Fonte: Elaborada pela autora.

A Figura 52 mostra os referentes à imagem *37073*, cujo detalhe contém o número “094”, que foi reconstruído de forma legível nas Figuras 52c (relativa ao BTV) e 52d (relativa ao IRWSR). Nas variações do método PB (Figuras 52e, 52f e 52g) entretanto, o algarismo “9” pode ser confundido com um “3”, apesar do resultados de PSNR e SSIM dessas imagens serem melhores.

A partir da análise por inspeção visual apresentada nesta seção, nota-se, em suma, que as variantes do método PB, propostas nesta tese, não apresentaram artefatos e deformações, sobretudo nas estruturas de borda e textura, como apresentaram o BTV e, em algumas situações, o IRWSR. Os métodos PB também não apresentaram aspecto cartunesco, como ocorreu com o método IRWSR. Essa análise sugere que o uso da informação local a partir da varredura de *patches* proposta resulta em imagens com melhor reconstrução de borda e textura, quando comparada aos métodos BTV e IRWSR nas 119 imagens testadas. Por outro lado, nota-se que os métodos PB - sem classificação, PB - variância e PB - suavidade apresentam resultados visualmente similares, sendo imperceptíveis as diferenças visuais entre eles. Esse resultado respalda a análise feita a partir dos gráficos de diferenças críticas mostrados nas Figuras 46a e 46b, em que as três variantes do método PB são estatisticamente equivalentes. Ou seja, em outras palavras, apesar da melhoria do aspecto visual proporcionada pelas abordagens com varredura de patches (PB), pode-se dizer que a classificação de homogeneidade proposta não apresentou diferenças significativas nos experimentos feitos, no que diz respeito ao

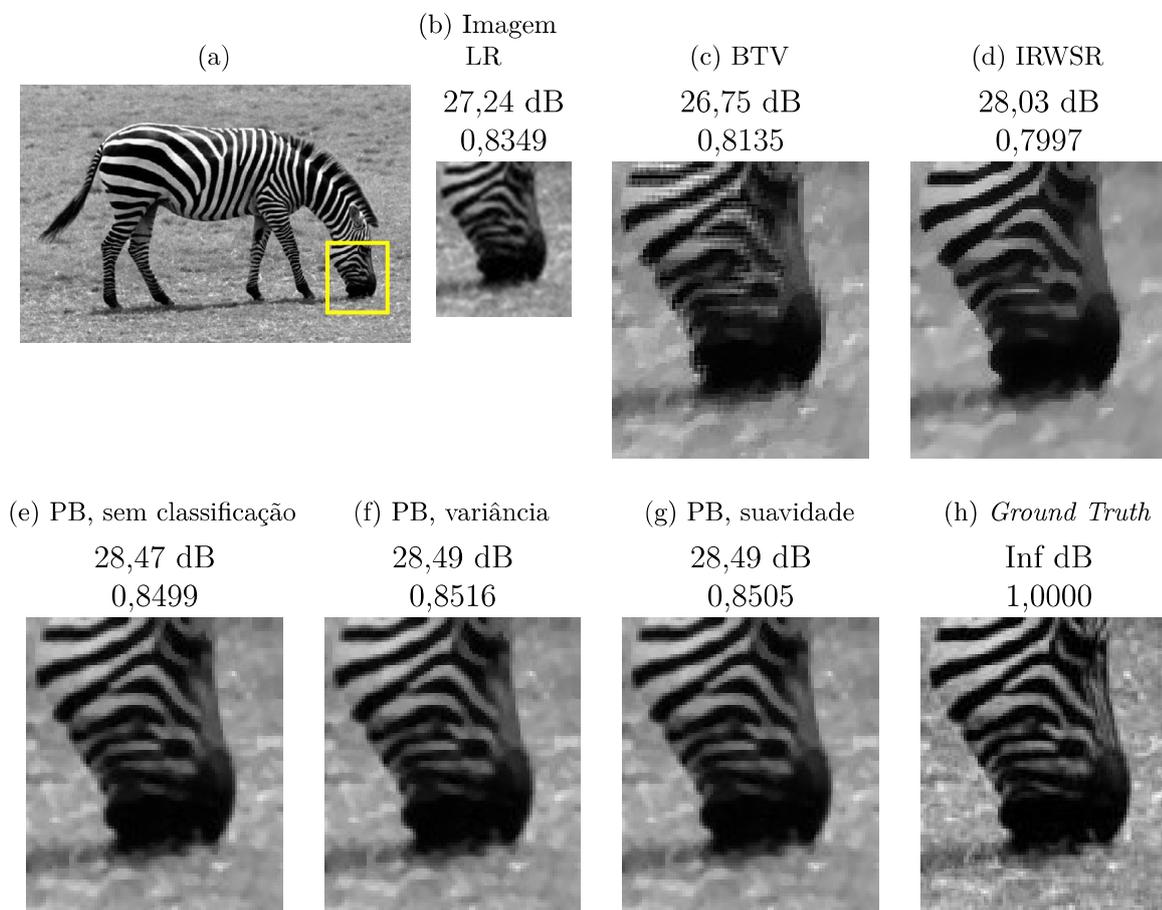
Figura 49 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *flowers*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.



Fonte: Elaborada pela autora.

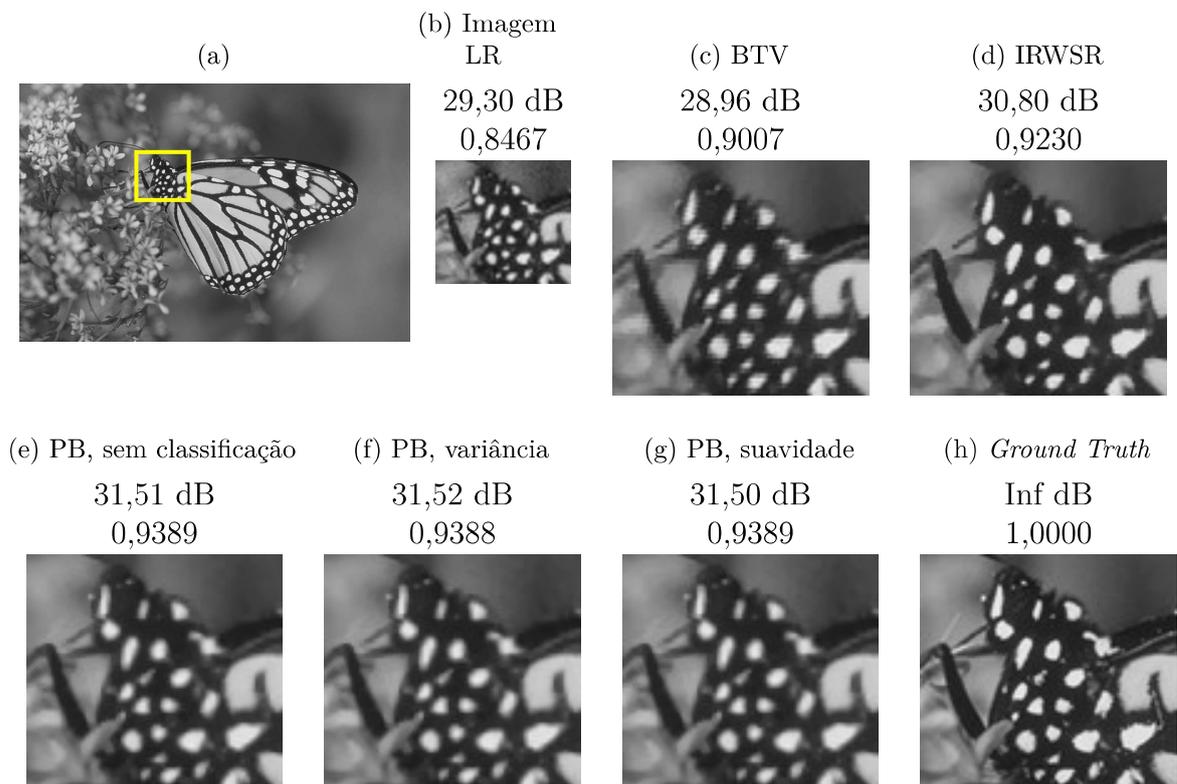
desempenho em termos de qualidade, porque o método PB - sem classificação apresentou resultado estatisticamente similar aos outros dois (PB - suavidade e PB - suavização), que contém classificação.

Figura 50 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *zebra*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.



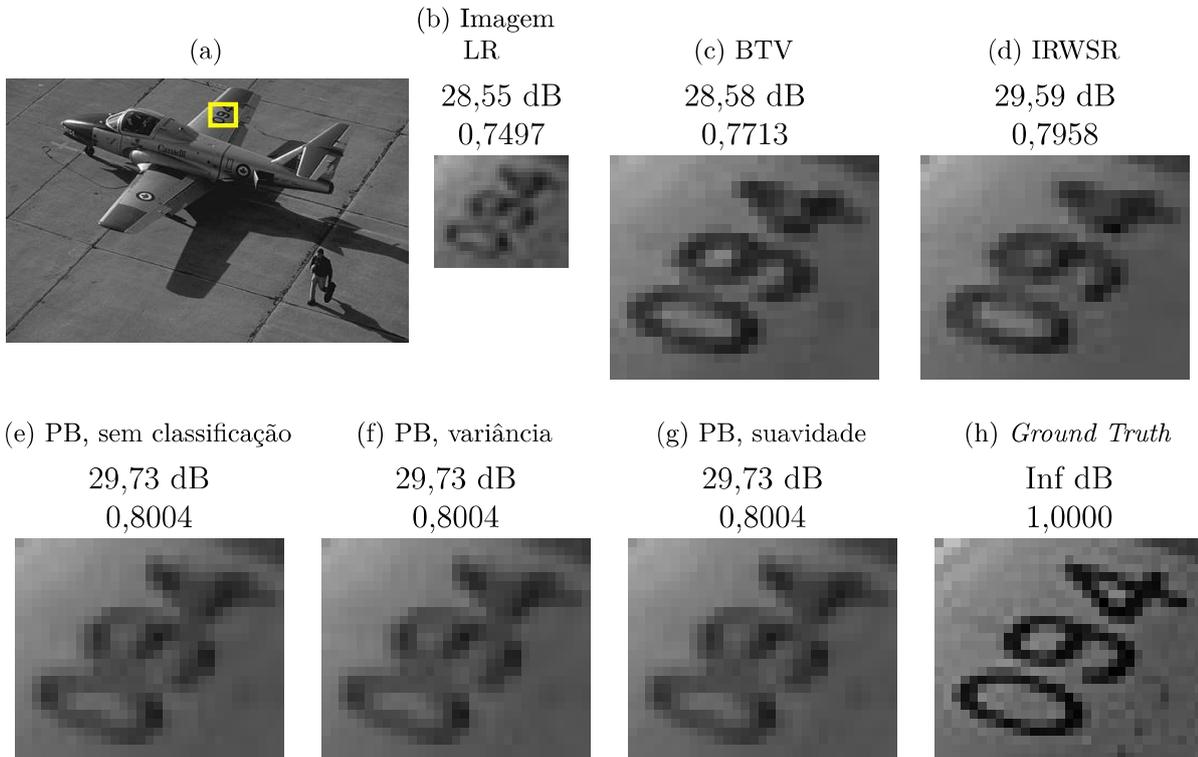
Fonte: Elaborada pela autora.

Figura 51 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução *monarch*, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.



Fonte: Elaborada pela autora.

Figura 52 – Imagens super-resolvidas, referentes à (a) imagem original de alta resolução 37073, (b) destaque de uma das imagens LR, (c) destaque da imagem super-resolvida pelo método BTV, (d) destaque da imagem super-resolvida pelo método IRWSR, (e) destaque da imagem super-resolvida pelo método PB - sem classificação, (f) destaque da imagem super-resolvida pelo método PB - variância, (g) destaque da imagem super-resolvida pelo método PB - suavidade e (h) destaque do *ground truth*.



Fonte: Elaborada pela autora.

## 6 Conclusão

Este trabalho se insere no contexto de super-resolução *multi-frame* de imagens, em especial, baseando-se no paradigma Bayesiano. As hipóteses que nortearam o desenvolvimento que se apresenta são (i) o uso de informações locais pode trazer bons resultados quando aplicado a métodos de super-resolução *multi-frame*, sendo possível, a partir dele, constituir detalhes relevantes e fiéis à cena real, de onde a imagem foi observada e (ii) o uso do método de registro não paramétrico *Demons*, quando aplicado à uma abordagem de super-resolução *multi-frame* sequencial, proporciona um ponto de partida melhor para o problema iterativo de reconstrução, melhorando o resultado final da super-resolução em questão.

Três abordagens, que culminaram em sete métodos, foram propostas para testar tais hipóteses: a abordagem híbrida (métodos HyBTV e HyIRWSR), a abordagem baseada em registro *Demons* (métodos D-BTVIR e D-IRWIR) e a abordagem baseada em *patches* (métodos PB - sem classificação, PB - suavidade e PB - variância). Os resultados da abordagem híbrida, apresentados no Capítulo 3, corroboram a hipótese (i), particularmente a partir do uso de múltiplas redes neurais, treinadas e selecionadas de acordo com a informação geométrica local das imagens, por meio das estatísticas do gradiente. Os resultados dos experimentos feitos com os métodos D-BTVIR e D-IRWIR, apresentados no Capítulo 4, corroboram a hipótese (ii), uma vez que se tratam diretamente da aplicação do registro *Demons* ao problema de super-resolução *multi-frame*, além de também respaldar indiretamente a hipótese (i), uma vez que o método *Demons* é um método de registro local. Finalmente, os resultados dos experimentos feitos com os métodos PB - sem classificação, PB - variância e PB - suavidade, apresentados no Capítulo 5, corroboram as duas hipóteses, a partir do uso do método *Demons* combinado com uma varredura e classificação de *patches* e posterior reconstrução *multi-frame* Bayesiana regularizada.

A discussão acerca dos espectros de Fourier, mostrada no início do Capítulo 3, sugere que o uso da regularização em super-resolução *multi-frame* não é o suficiente para ocasionar uma reconstrução completa das informações perdidas no processo de aquisição da imagem. Nesse sentido, a proposta do Capítulo 3 é utilizar o paradigma do aprendizado de máquinas, em especial as múltiplas redes neurais de uma camada (SLFN), para contribuir com tal reconstrução. O teste de Wilcoxon mostra uma melhoria significativa (nível de significância de 1%) do método HyBTV sobre o BTV e do HyIRWSR sobre o IRWSR, o que indica uma melhoria significativa em decorrência do uso das redes neurais após a aplicação dos métodos BTV e IRWSR. Além disso, o teste também mostra um melhor desempenho (com nível de significância de 5%) do método HyBTV sobre o IRWSR.

O teste de Friedman, por sua vez, mostra que os quatro métodos (BTV, HyBTV, IRWSR e HyIRWSR) não são equivalentes entre si. E o teste de Nemenyi, juntamente à análise de tempo de processamento, indica um melhor custo benefício do método HyBTV sobre o método IRWSR, uma vez que o IRWSR tem tempo de processamento, em média, 5,6 maior do que HyBTV. A complexidade do método IRWSR, pode-se dizer, está no cálculo das matrizes  $M_\alpha$  e  $M_\beta$ . Por outro lado, para os métodos HyBTV e HyIRWSR, há uma complexidade referente às redes neurais. A análise do tempo de processamento em etapas mostrou que a complexidade da fase de aprendizado de máquinas do HyBTV impactou menos significativamente no tempo do que o cálculo das matrizes  $M_\alpha$  e  $M_\beta$  do IRWSR. A partir desse resultado, conclui-se que a etapa de aprendizado de máquina ocasionou no ganho de qualidade de imagem com um menor tempo de processamento.

A análise por inspeção visual feita no Capítulo 3 mostrou que o HyBTV e o HyIRWSR se destacam por não apresentarem (i) artefato quadriculado nas regiões de borda, (ii) aspecto cartunesco e (iii) deformação de estruturas como faces e caracteres. Por outro lado, ainda se observa demasiada suavização nas imagens, perceptível sobretudo em regiões de bordas e textura, sendo essa uma importante limitação dos métodos HyBTV e HyIRWSR. A presença de artefatos em relação à deformação por movimento motiva uma investigação acerca do método de registro, que alinha as imagens LR, o que fomenta os métodos propostos nos Capítulos 4 e 5.

A investigação na literatura de super-resolução *multi-frame* focada em registro mostrou a predominância do uso de métodos paramétricos, que apresentam como desvantagem o fato de dependerem de suposições acerca do operador de deslocamento. Além disso, essa investigação trouxe à tona o fato do registro ser não linear quando da presença da translação, que é o caso do problema de super-resolução. Nesse contexto dos métodos de registro não paramétricos e não lineares, o método *Demons* se destaca e, por isso, foi o escolhido para compor a segunda hipótese apresentada nesta tese. O Capítulo 4 mostra os experimentos feitos neste sentido, a partir dos quais o resultado do teste de Wilcoxon sugere que há uma melhoria estatisticamente significativa em decorrência do uso do registro *Demons*. Além disso, os testes de hipótese juntamente à análise de tempo de processamento mostram um melhor custo benefício do método D-BTVIR sobre o IRWSR, já que IRWSR tem tempo de processamento, em média, 7,2 maior do que D-BTVIR.

Os resultados da inspeção visual mostram, em algumas situações, suavidade demasiada nas imagens super-resolvidas pelo D-BTVIR e D-IRWIR. Uma vez que essa suavidade demasiada não se apresenta nos resultados do BTV e IRWSR, nesses casos, é razoável considerar que o registro *Demons* é a causa delas. Destaca-se, nesse sentido, o uso da regularização Bayesiana no registro *Demons*, na prática se tratando de um filtro suavizador. Sugere-se, como trabalho futuro a partir desta tese, uma investigação dos efeitos dos parâmetros desse filtro na imagem final que pode ocasionar em uma melhora,

em especial considerando a suavidade demasiada. Em relação à modelagem matemática utilizada para o desenvolvimento dos métodos D-BTVIR e D-IRWIR, considerou-se o mesmo modelo de aquisição utilizado para os métodos BTV e IRWSR no qual o operador de deslocamento foi tido como uma matriz identidade. Entretanto, uma vez que supõe-se que as imagens de entrada já estão alinhadas, tal operador pode ser desconsiderado do modelo de aquisição o que pode fomentar uma investigação acerca desse novo modelo de aquisição, como trabalho futuro. Em especial, nesse sentido, sugere-se analisar a manutenção ou não do pressuposto de ruído AWGN após a alteração.

A motivação da proposta apresentada no Capítulo 5 surge da ideia de super-resolver diferentes regiões da imagem de entrada de acordo com características referentes à homogeneidade. Esse raciocínio se alinha com a proposta do Capítulo 3, no sentido de super-resolver diferentes regiões da imagem de forma especializada. Entretanto, enquanto para o HyBTV e HyIRWSR essa classificação é feita de forma não supervisionada, nas soluções PB (*patch based*), propõe-se que essa classificação seja feita de forma supervisionada. Outra diferença significativa entre as propostas do Capítulo 5, quando comparadas às propostas do Capítulo 3, é o uso do registro *Demons* e a ausência da etapa de aprendizado de máquina, de forma que a solução do Capítulo 5 dependa de menos suposições acerca do modelo de aquisição.

Foram propostas três variações do método PB, de acordo com a classificação de homogeneidade utilizada: PB - sem classificação, PB - suavidade e PB - variância. O teste de hipótese de Nemenyi retornou equivalência estatística entre as três técnicas, em relação às métricas de qualidade mas, por outro lado, retornou vantagem do PB - variância em termos de tempo de processamento. A análise por inspeção visual corrobora a equivalência entre os métodos, em termos de qualidade de imagem. As duas análises indicaram um melhor desempenho dos métodos PB em relação aos métodos BTV e IRWSR, o que valida os métodos propostos. Em contrapartida, a equivalência entre os três métodos PB sugere que as classificações de homogeneidade propostas não influenciaram no desempenho em termos de qualidade. Sugere-se, como trabalho futuro, a aplicação de outros métodos de classificação de homogeneidade aos métodos PB propostos. Ademais, no Capítulo 5 apresentam-se, ainda, um processo de validação utilizando os testes de hipótese de Friedman e Nemenyi.

Obversa-se comumente, nos trabalhos de super-resolução, análises visuais sobre um grupo cuidadosamente selecionado de imagens. Apesar de comum, este trabalho partiu do pressuposto de que não se trata de uma análise robusta, podendo levar a conclusões enviesadas sobre o desempenho dos métodos comparados. Por esse motivo, propôs-se o uso de testes de hipóteses bem estabelecidos pela literatura de estatística não paramétrica. As análises visuais foram mantidas, entretanto, para que fosse avaliada a convergência entre os dois tipos de análises, já que os métodos comparados não tinham referência de

testes de hipótese. O fato dos métodos propostos apresentarem melhorias em ambas as análises indica a eficácia e validade dos mesmos para a literatura de super-resolução. As análises de comparação estatística propostas, utilizando testes de Wilcoxon, Friedman e Nemenyi, podem ser utilizadas como metodologia de investigação dos resultados das métricas de qualidades, inclusive para outras métricas, além do PSNR, SSIM e tempo de processamento, em outros trabalhos de super-resolução e restauração de imagem.

## Referências

- AIZAWA, K.; KOMATSU, T.; SAITO, T. Acquisition of very high resolution images using stereo cameras. In: *Visual Communications and Image Processing '91: Visual Communication*. Boston: [s.n.], 1991. v. 1605, p. 318 – 328. Citado na página 57.
- BAJCSY, R.; KOVAČIČ, S. Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, Academic Press Professional, Inc., USA, v. 46, n. 1, p. 1–21, 1989. Citado na página 59.
- BAKER, S.; KANADE, T. Hallucinating faces. In: *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. [S.l.: s.n.], 2000. p. 83–88. Citado na página 59.
- BAKER, S.; KANADE, T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 24, n. 9, p. 1167–1183, 2002. Citado na página 58.
- BEVILACQUA, M.; ROUMY, A.; GUILLEMOT, C.; MOREL, M. line A. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: *Proceedings of the British Machine Vision Conference*. Surrey: BMVA Press, 2012. p. 135.1–135.10. Citado 3 vezes nas páginas 64, 83 e 107.
- CHO, S.; LEE, S. Handling Outliers in Non-Blind Image Deconvolution. In: *Proceedings of 2011 International Conference on Computer Vision*. Barcelona: [s.n.], 2011. p. 495–502. Citado na página 58.
- COSMO, D. L. *Super resolução single frame utilizando redes extreme learning machine*. Tese (Doutorado em Engenharia Elétrica) — Programa de Pós Graduação em Engenharia Elétrica, Universidade Federal do Espírito Santo, Vitória, 2019. Citado na página 59.
- COSMO, D. L.; INABA, F. K.; SALLES, E. O. T. Single Image Super-Resolution Using Multiple Extreme Learning Machine Regressors. In: *Proceedings of 2017 30th Conference on Graphics, Patterns and Images (SIBGRAPI)*. Niterói: [s.n.], 2017. p. 397–404. Citado 4 vezes nas páginas 46, 47, 50 e 60.
- COSMO, D. L.; NASCIMENTO, T. P. do; SALLES, E. O. T.; CIARELLI, P. M. Improving Super-Resolution Reconstruction with Regularized Extreme Learning Machine Networks. In: *Proceedings of 2018 International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro: [s.n.], 2018. v. 51, n. 1, p. 1–8. Citado 6 vezes nas páginas 46, 48, 49, 62, 63 e 103.
- COSMO, D. L.; SALLES, E. O. T. Multiple sequential regularized extreme learning machines for single image super resolution. *IEEE Signal Processing Letters*, IEEE, v. 26, n. 3, p. 440–444, 2019. Citado 3 vezes nas páginas 30, 59 e 60.
- DAUBECHIES, I.; DEFRISE, M.; MOL, C. D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, Wiley Online Library, v. 57, n. 11, p. 1413–1457, 2004. Citado na página 28.

- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, Microtome Publishing, v. 7, n. Jan, p. 1–30, 2006. Citado 4 vezes nas páginas 31, 54, 55 e 56.
- DENG, W.; ZHENG, Q.; CHEN, L. Regularized Extreme Learning Machine. In: *Proceedings of 2009 IEEE Symposium on Computational Intelligence and Data Mining*. Nashville: IEEE, 2009. v. 51, n. 1, p. 389–395. Citado na página 49.
- DEUDON, M.; KALAITZIS, A.; GOYTOM, I.; AREFIN, M. R.; LIN, Z.; SANKARAN, K.; MICHALSKI, V.; KAHOU, S. E.; CORNEBISE, J.; BENGIO, Y. Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery. *arXiv preprint arXiv:2002.06460*, 2020. Citado na página 60.
- DONG, C.; LOY, C. C.; HE, K.; TANG, X. Learning a deep convolutional network for image super-resolution. In: SPRINGER. *European conference on computer vision*. [S.l.], 2014. p. 184–199. Citado na página 60.
- DONG, C.; LOY, C. C.; HE, K.; TANG, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 38, n. 2, p. 295–307, 2016. Citado 3 vezes nas páginas 59, 64 e 103.
- DONG, C.; LOY, C. C.; TANG, X. Accelerating the super-resolution convolutional neural network. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 391–407. Citado na página 60.
- EL-GAMAL, F. E.-Z. A.; ELMOGY, M.; ATWAN, A. Current trends in medical image registration and fusion. *Egyptian Informatics Journal*, Elsevier, Cairo, v. 17, n. 1, p. 99–124, 2016. Citado na página 59.
- ELAD, M.; FEUER, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, IEEE, v. 6, n. 12, p. 1646–1658, 1997. Citado 4 vezes nas páginas 37, 41, 42 e 58.
- ELAD, M.; FEUER, A. Superresolution restoration of an image sequence: adaptive filtering approach. *IEEE Transactions on Image Processing*, IEEE, v. 8, n. 3, p. 387–395, 1999. Citado na página 58.
- EVANGELIDIS, G. D.; PSARAKIS, E. Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 30, n. 10, p. 1858–1865, 2008. Citado na página 39.
- FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. d. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. [S.l.]: Livros Técnicos e Científicos Editora, 2011. Citado na página 103.
- FARSIU, S.; ROBINSON, M. D.; ELAD, M.; MILANFAR, P. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, IEEE, v. 13, n. 10, p. 1327–1344, 2004. Citado 14 vezes nas páginas 11, 28, 36, 41, 42, 58, 62, 64, 67, 80, 81, 82, 83 e 107.
- FREEMAN, W. T.; JONES, T. R.; PASZTOR, E. C. Example-Based Super-Resolution. *IEEE Computer Graphics and Applications*, IEEE, v. 22, n. March, p. 56–65, 2002. Citado 3 vezes nas páginas 30, 45 e 59.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 3. ed. New Jersey: Pearson Prentice Hall, 2012. Citado 8 vezes nas páginas 27, 35, 36, 39, 61, 81, 93 e 102.

GRABNER, A.; POIER, G.; OPITZ, M.; SCHULTER, S.; ROTH, P. M. Loss-specific training of random forests for super-resolution. In: *Proceedings of the 22nd Computer Vision Winter Workshop*. Austria: TU Wien, Pattern Recognition and Image Processing Group, 2017. Citado na página 59.

HADAMARD, J. *Lectures on The Cauchy Problem in Linear Partial Differential Equations*. New Haven: Yale University Press, 1923. Citado na página 28.

HARDIE, R. C.; BARNARD, K. J.; ARMSTRONG, E. E. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE transactions on Image Processing*, IEEE, v. 6, n. 12, p. 1621–1633, 1997. Citado na página 58.

HE, Y.; YAP, K.-H.; CHEN, L.; CHAU, L.-P. A nonlinear least square technique for simultaneous image registration and super-resolution. *IEEE Transactions on Image Processing*, IEEE, v. 16, n. 11, p. 2830–2841, 2007. Citado 3 vezes nas páginas 58, 59 e 82.

HORE, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. In: *Proceedings of 2010 20th international conference on pattern recognition*. [S.l.]: IEEE, 2010. p. 2366–2369. Citado na página 53.

HUANG, G.-b.; ZHU, Q.-y.; SIEW, C.-k. Extreme Learning Machine : A New Learning Scheme of Feedforward Neural Networks. *IEEE International Joint Conference on Neural Networks*, IEEE, v. 2, p. 985–990, 2004. Citado na página 49.

INABA, F. K.; SALLES, E. O. T.; PERRON, S.; CAPOROSSI, G. Dgr-elm–distributed generalized regularized elm for classification. *Neurocomputing*, Elsevier, v. 275, p. 1522–1530, 2018. Citado na página 52.

IRANI, M.; PELEG, S. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, v. 53, n. 3, p. 231–239, 1991. Citado na página 58.

IRANI, M.; PELEG, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, v. 4, n. 4, p. 324–335, 1993. Citado na página 58.

JEBADURAI, J.; PETER, J. D. Sk-svr: Sigmoid kernel support vector regression based in-scale single image super-resolution. *Pattern Recognition Letters*, Elsevier, v. 94, p. 144–153, 2017. Citado na página 59.

KATSAGGELOS, A. K.; MOLINA, R.; MATEOS, J. *Super resolution of Images and Video*. 1. ed. California: Morgan & Claypool Publishers, 2007. 1–134 p. (Synthesis Lectures on Image, Video, and Multimedia Processing, 7). Citado 5 vezes nas páginas 28, 29, 41, 79 e 98.

KAWULOK, M.; BENECKI, P.; PIECHACZEK, S.; HRYNCZENKO, K.; KOSTRZEWA, D.; NALEPA, J. Deep learning for multiple-image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, v. 17, n. 6, p. 1062–1066, 2020. Citado na página 60.

- KHATTAB, M. M.; ZEKI, A. M.; ALWAN, A. A.; BADAWY, A. S. Regularization-based multi-frame super-resolution: A systematic review. *Journal of King Saud University - Computer and Information Sciences*, Elsevier, v. 21, n. 7, p. 755–762, 2020. Citado 5 vezes nas páginas 38, 41, 58, 61 e 98.
- KIM, S.; BOSE, N. K. Recursive Reconstruction of Image from Noisy and Blurred Multiframe. *IEEE Transactions on Acoustics*, IEEE, v. 38, n. 6, p. 1013–1027, 1990. Citado na página 57.
- KNOTT, C. G. Quote from Undated Letter from Maxwell to Tait. In: *Life and Scientific Work of Peter Guthrie Tait*. Cambridge: University Press, 1911. p. 215. Citado na página 39.
- KÖHLER, T.; HUANG, X.; SCHEBESCH, F.; AICHERT, A.; MAIER, A.; HORNEGGER, J. Robust Multiframe Super-Resolution Employing Iteratively Re-Weighted Minimization. *IEEE Transactions on Computational Imaging*, IEEE, v. 2, n. 1, p. 42–58, 2016. Citado 18 vezes nas páginas 11, 28, 30, 38, 43, 44, 58, 59, 62, 64, 67, 80, 81, 82, 83, 94, 102 e 107.
- LAGHRIB, A.; GHAZDALI, A.; HAKIM, A.; RAGHAY, S. A multi-frame super-resolution using diffusion registration and a nonlocal variational image restoration. *Computers & Mathematics with Applications*, Elsevier, v. 72, n. 9, p. 2535–2548, 2016. Citado 2 vezes nas páginas 58 e 59.
- LAGHRIB, A.; HADRI, A.; HAKIM, A.; RAGHAY, S. A new multiframe super-resolution based on nonlinear registration and a spatially weighted regularization. *Information Sciences*, Elsevier, v. 493, p. 34–56, 2019. Citado 2 vezes nas páginas 58 e 59.
- LAGHRIB, A.; HAKIM, A.; RAGHAY, S.; RHABI, M.; MARNE, F. Champs-sur. Robust super resolution of images with non-parametric deformations using an elastic registration. *Applied Mathematical Sciences*, Hikari Ltd., v. 8, n. 179, p. 8897–8907, 2014. Citado 2 vezes nas páginas 58 e 59.
- LEDIG, C.; THEIS, L.; HUSZÁR, F.; CABALLERO, J.; CUNNINGHAM, A.; ACOSTA, A.; AITKEN, A.; TEJANI, A.; TOTZ, J.; WANG, Z.; SHI, W. Photo-realistic single image super-resolution using a generative adversarial network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 105–114. Citado na página 60.
- LI, X.; LAM, K. M.; QIU, G.; SHEN, L.; WANG, S. Example-based image super-resolution with class-specific predictors. *Journal of Visual Communication and Image Representation*, Elsevier Inc., v. 20, n. 5, p. 312–322, 2009. Citado 2 vezes nas páginas 45 e 60.
- LIU, X.; CHEN, L.; WANG, W.; ZHAO, J. Robust multi-frame super-resolution based on spatially weighted half-quadratic estimation and adaptive btv regularization. *IEEE Transactions on Image Processing*, IEEE, v. 27, n. 10, p. 4971–4986, 2018. Citado 2 vezes nas páginas 58 e 59.
- MARTIN, D.; FOWLKES, C.; TAL, D.; MALIK, J. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In: . Vancouver: [s.n.], 2001. v. 2, n. July, p. 416–423. Citado 4 vezes nas páginas 64, 66, 83 e 107.

- MILANFAR, P. *Super-Resolution Imaging*. New York: CRC Press, 2011. Citado 4 vezes nas páginas 27, 28, 59 e 98.
- MOLINA, R.; VEGA, M.; ABAD, J.; KATSAGGELOS, A. K. Parameter estimation in bayesian high-resolution image reconstruction with multisensors. *IEEE Transactions on Image Processing*, IEEE, v. 12, n. 12, p. 1655–1667, 2003. Citado na página 58.
- MOURABIT, I. E.; RHABI, M. E.; HAKIM, A.; LAGHRIB, A.; MOREAU, E. A new denoising model for multi-frame super-resolution image reconstruction. *Signal Processing*, Elsevier, v. 132, p. 51 – 65, 2017. Citado na página 41.
- NABNEY, I. *NETLAB: Algorithms for Pattern Recognition*. 1. ed. UK: Springer Science & Business Media, 2002. Citado na página 83.
- NASROLLAHI, K.; MOESLUND, T. B. Super-resolution: A comprehensive survey. *Machine Vision and Applications*, Springer, v. 25, n. 6, p. 1423–1468, 2014. Citado 7 vezes nas páginas 28, 38, 41, 42, 44, 59 e 98.
- NEMENYI, P. *Distribution-free Multiple Comparisons*. Tese (Doutorado), New Jersey, 1963. Citado na página 55.
- NG, M. K.; SHEN, H.; LAM, E. Y.; ZHANG, L. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, Springer, v. 2007, n. 1, p. 1–16, 2007. Citado 2 vezes nas páginas 58 e 59.
- NGUYEN, N.; MILANFAR, P. A wavelet-based method for superresolution (Wavelet Superresolution). *Circuits System Signal Process*, Springer, v. 19, n. 4, p. 321–338, 2000. Citado na página 57.
- NGUYEN, N.; MILANFAR, P.; GOLUB, G. A computationally efficient superresolution image reconstruction algorithm. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, IEEE, v. 10, n. 4, p. 573—583, 2001. Citado 2 vezes nas páginas 58 e 59.
- OPPENHEIM, A. V.; WILLSKY, A. S. *Sinais e sistemas*. 2. ed. São Paulo: Pearson Education do Brasil Ltda., 2010. Citado na página 27.
- PARK, S. C.; PARK, M. K.; KANG, M. G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, IEEE, v. 20, n. 3, p. 21–36, 2003. Citado 5 vezes nas páginas 27, 28, 29, 57 e 79.
- PELEG, S.; KEREN, D.; SCHWEITZER, L. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, Elsevier Science Inc., v. 5, n. 3, p. 223–226, 1987. Citado 2 vezes nas páginas 28 e 57.
- QIU, P.; XING, C. On Nonparametric Image Registration. *Technometrics*, Taylor & Francis, v. 55, n. 2, p. 174–188, 2013. Citado 3 vezes nas páginas 29, 59 e 79.
- RAHMAN, S. u.; WESARG, S. Combining short-axis and long-axis cardiac MR images by applying a super-resolution reconstruction algorithm. In: DAWANT, B. M.; HAYNOR, D. R. (Ed.). *Proceedings of Medical Imaging 2010: Image Processing*. California: SPIE, 2010. v. 7623, p. 187 – 198. Citado na página 59.

- RHEE, S.; KANG, M. G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm. *Optical Engineering*, SPIE, v. 38, n. 8, p. 1348 – 1356, 1999. Citado na página 57.
- ROMANO, Y.; ISIDORO, J.; MILANFAR, P. RAISR: Rapid and Accurate Image Super Resolution. *IEEE Transactions on Computational Imaging*, IEEE, v. 3, n. 1, p. 110–125, 2017. Citado 6 vezes nas páginas 47, 60, 64, 65, 83 e 107.
- RUDIN, L. I.; OSHER, S.; FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, Elsevier, v. 60, n. 1–4, p. 259–268, 1992. Citado na página 42.
- SCHULTZ, R. R.; MENG, L.; STEVENSON, R. L. Subpixel motion estimation for super-resolution image sequence enhancement. *Journal of visual communication and image representation*, Elsevier, v. 9, n. 1, p. 38–50, 1998. Citado 2 vezes nas páginas 58 e 59.
- SHEN, H.; ZHANG, L.; HUANG, B.; LI, P. A MAP approach for joint motion estimation, segmentation, and super resolution. *IEEE Transactions on Image Processing*, IEEE, v. 16, n. 2, p. 479–490, 2007. Citado 2 vezes nas páginas 58 e 59.
- SHEKIN, D. J. *Handbook of Parametric and Nonparametric Statistical Procedures*. 5. ed. New York: CRC Press, 2011. Citado 2 vezes nas páginas 54 e 56.
- SHI, F.; CHENG, J.; WANG, L.; YAP, P.-T.; SHEN, D. Longitudinal Guided Super-Resolution Reconstruction of Neonatal Brain MR Images. In: *Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data. STIA 2014. Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2015. v. 8682, p. 67–76. Citado na página 59.
- STARK, H.; OSKOUI, P. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, OSA, v. 6, n. 11, p. 1715–1726, Nov 1989. Citado na página 58.
- SU, H.; TANG, L.; WU, Y.; TRETTER, D.; ZHOU, J. Spatially adaptive block-based super-resolution. *IEEE Transactions on Image Processing*, IEEE, v. 21, n. 3, p. 1031–1045, 2012. Citado 2 vezes nas páginas 30 e 99.
- THIRION, J.-P. Image Matching as a Siffusion Process: an Analogy with Maxwell’s demons. *Medical image analysis*, Elsevier, v. 2, n. 3, p. 243–260, 1998. Citado 6 vezes nas páginas 29, 39, 40, 59, 79 e 94.
- TSAI, R.; HUANG, T. Multiframe Image Restoration and Registration. *Advances in Computer Vision and Image Processing*, JAI Press, v. 1, p. 317–339, 1984. Citado na página 57.
- VANDEWALLE, P.; SBAIZ, L.; VETTERLI, M. Registration for super-resolution: Theory, algorithms, and applications in image and mobile video enhancement. In: MILANFAR, P. (Ed.). *Super-Resolution Imaging*. 1. ed. New York: CRC Press, 2011. Citado 2 vezes nas páginas 29 e 79.
- VERCAUTEREN, T.; PENNEC, X.; MALIS, E.; PERCHANT, A.; AYACHE, N. Insight into efficient image registration techniques and the demons algorithm. In: SPRINGER. *Proceedings of the Biennial International Conference on Information Processing in Medical Imaging*. Kerkrade, 2007. p. 495–506. Citado 5 vezes nas páginas 29, 39, 41, 59 e 79.

- VERCAUTEREN, T.; PENNEC, X.; PERCHANT, A.; AYACHE, N. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, Elsevier, v. 45, n. 1, p. S61–S72, 2009. Citado na página 59.
- WANG, Z.; BOVIK, A. C. Modern Image Quality Assessment. In: BOVIK, A. C. (Ed.). *Synthesis Lectures on Image, Video, and Multimedia Processing*. [S.l.]: Morgan & Claypool Publishers, 2006. v. 2, p. 1–156. Citado na página 52.
- WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, IEEE, v. 13, n. 4, p. 600–612, 2004. Citado na página 53.
- WANG, Z.; BOVIK, A. C.; SIMONCELLI, E. P. Structural approaches to image quality assessment. In: BOVIK, A. C. (Ed.). *Handbook of Image and Video Processing*. 2. ed. Austin: Academic, 2005. Citado na página 54.
- WANG, Z.; CHEN, J.; HOI, S. C. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 43, n. 10, p. 3365–3387, 2020. Citado na página 60.
- WRONSKI, B.; GARCIA-DORADO, I.; ERNST, M.; KELLY, D.; KRANIN, M.; LIANG, C.-K.; LEVOY, M.; MILANFAR, P. Handheld Multi-Frame Super-Resolution. *ACM Transactions on Graphics (TOG)*, ACM New York, v. 38, n. 4, p. 1–18, 2019. Citado na página 59.
- WU, X.; XIAO, S.; ZHANG, Y. Registration Based Super-Resolution Reconstruction for Lung 4D-CT. In: *Proceedings of 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago: IEEE, 2014. p. 2444–2447. Citado na página 59.
- XiaoGuang Feng; MILANFAR, P. Multiscale principal components analysis for image local orientation estimation. In: *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*. Pacific Grove: IEEE, 2002. v. 1, p. 478–482. Citado na página 46.
- YANG, J.; WRIGHT, J.; HUANG, T. S.; MA, Y. Image Super-Resolution via Sparse Representation. *IEEE Transactions on Image Processing*, IEEE, v. 19, n. 11, p. 2861–2873, 2010. Citado 2 vezes nas páginas 38 e 59.
- YUE, L.; SHEN, H.; LI, J.; YUAN, Q.; ZHANG, H.; ZHANG, L. Image super-resolution: The techniques, applications, and future. *Signal Processing*, Elsevier, v. 128, p. 389–408, 2016. Citado 3 vezes nas páginas 41, 57 e 58.
- ZENG, X.; YANG, L. A Robust Multiframe Super-Resolution Algorithm Based on Half-Quadratic Estimation with Modified BTV Regularization. *Digital Signal Processing: A Review Journal*, Academic Press, Inc., v. 23, n. 1, p. 98–109, 2013. Citado 2 vezes nas páginas 42 e 58.
- ZHANG, Y.; TIAN, Y.; KONG, Y.; ZHONG, B.; FU, Y. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 43, n. 7, p. 2480–2495, 2020. Citado na página 60.

---

ZITOVÁ, B.; FLUSSER, J. Image Registration Methods: A Survey. *Image and Vision Computing*, Elsevier, v. 21, n. 11, p. 977–1000, 2003. Citado 4 vezes nas páginas [29](#), [39](#), [59](#) e [80](#).