Gentil Auer Neto

Identificação de Contorno, Gênero e Massa Corporal por Meio de Redes Convolucionais: Uma Contribuição para a Estimativa Automática de Somatotipo

Vitória

2020

Gentil Auer Neto

Identificação de Contorno, Gênero e Massa Corporal por Meio de Redes Convolucionais: Uma Contribuição para a Estimativa Automática de Somatotipo

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Universidade Federal do Espírito Santo

Centro Tecnológico

Programa de Pós-Graduação em Engenharia Elétrica

Orientador: Prof. Dr. Patrick Marques Ciarelli

Vitória

2020

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

---

Auer Neto, Gentil, 1990-
A917i Identificação de contorno, gênero e massa corporal por meio de redes convolucionais: uma contribuição para a estimativa automática de somatotipo / Gentil Auer Neto. - 2020.
108 f. : il.

Orientador: Patrick Marques Ciarelli.
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Somatotipo. 2. Redes Neurais Profundas. 3. Segmentação do Corpo Humano. 4. Ambientes não controlados. 5. Classificação de Gênero. 6. Estimativa de massa corporal. I. Marques Ciarelli, Patrick. II. Universidade Federal do Espírito Santo. Centro Tecnológico. III. Título.

CDU: 621.3

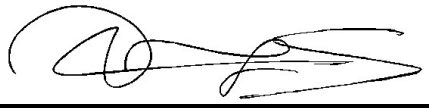
Gentil Auer Neto

Identificação de Contorno, Gênero e Massa Corporal por Meio de Redes Convolucionais: Uma Contribuição para a Estimativa Automática de Somatotipo

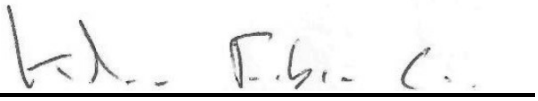
Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Engenharia Elétrica.

Trabalho aprovado. Vitória, 15 de junho de 2020:

Prof. Dr. Patrick Marques Ciarelli
Orientador



Prof. Dr. Jorge Leonid Aching Samatelo
Membro Externo



Prof. Dr. Klaus Fabian Côco
Membro Externo

Vitória
2020

À Sarah e Theo

Agradecimentos

Agradeço primeiramente a Deus, por conduzir meus passos e pelo refúgio seguro e inabalável em quaisquer circunstâncias.

Aos meus pais, Paulo e Maria Ângela, e à minha irmã, Lorena, por sempre terem me incentivado e nunca medido esforços para me dar o suporte que, sem dúvidas, foi fundamental para alicerçar a base que me sustentou até aqui. Obrigado pelos tão importantes conselhos, puxões de orelha e momentos carinhosos.

À Sarah, por todo amor e zelo, me impulsionando há mais de uma década para alcançar os meus objetivos. Obrigado por contribuir para que eu tivesse o foco necessário nos tempos de imersão, por me suportar pacientemente nos períodos de estresse, por me aguentar quando eu mostrava "telas em grego" seguidas de infinitas imagens de resultados, as quais felizmente ajudam a entender alguma coisa do que fiz aqui. Acima de tudo, obrigado por sempre construir valiosos conselhos para a minha tentativa de evoluir como ser humano.

Ao Theo, obrigado por simplesmente existir. É inexplicável a experiência maravilhosa que você me faz vivenciar todos os dias... você é um carinha legal.

Aos queridos amigos de Iúna, o meu muito obrigado! Jailda, João e Bruna, o apoio de vocês nunca será esquecido.

Ao Patrick, pelos inúmeros conselhos e direcionamentos na realização deste trabalho. Em especial, pela enorme paciência com meus períodos de menor disponibilidade, sem jamais desistir de me conduzir nesta pesquisa.

Ao Jorge e ao Klaus, por aceitarem o convite e disponibilizarem tempo compondo a banca de avaliação deste trabalho.

Aos colegas da ArcelorMittal Tubarão, agradeço pela parceria e colaboração de sempre, pela amizade valiosa que vem sendo construída ao longo do tempo e que, acreditem, faz um período de plantão virar moleza, tamanha a qualidade do ambiente que vocês proporcionam para quem está em volta.

Sorte é quando a preparação encontra a oportunidade.
— Autor Desconhecido.

Resumo

O corpo humano pode ser referenciado por uma métrica que mensura suas formas e proporções: o somatotipo. Estudos demonstram que a sua identificação é útil em diversas áreas da saúde e da educação física. Atualmente, o principal método para sua estimativa demanda bastante tempo, treinamento especializado e uso de equipamentos antropométricos para aferição de medidas corporais. Há pesquisas visando simplificar o teste, por exemplo, automatizando-o com processamento digital de imagens. Contudo, há limitações como a adequação de ambientes para a captura de fotografias com boa qualidade, um requisito importante para sua viabilidade. Dados os benefícios de obter um método robusto e rápido, tal que simplifique sua realização e reduza custos, a proposta principal deste trabalho é pesquisar e desenvolver métodos que possibilitem estimar de forma automática informações do corpo humano por meio de imagens obtidas em ambientes sem uma preparação prévia. Neste contexto, este trabalho propõe combinar técnicas de aprendizado de máquina, em especial aprendizado profundo, que alcancem a segmentação detalhada do contorno corporal de indivíduos em ambientes não controlados. Além disso, a identificação de gênero e estimativa de massa corporal são feitas, ambas por análise facial. Quatro diferentes bases de dados são utilizadas: face2bmi, Wiki, Fisio-Somatotipo e Internet. As duas primeiras com imagens de faces de pessoas em geral e as duas últimas com fotos de corpos inteiros de atletas. Os resultados mostram que o estudo é promissor, tendo alcançado um bom nível de segmentação em ambientes adversos e 91,52% de acurácia na classificação de gênero. Também é avaliado como o gênero influencia no modelo de estimativa de massa corporal, que alcançou um erro médio de 5,56 kg para o melhor caso.

Palavras-chave: Somatotipo. Redes Neurais Profundas. Segmentação do Corpo Humano. Ambientes não controlados. Classificação de Gênero. Estimativa de massa corporal.

Abstract

The human body can be referenced by a metric that measures up shapes and proportions: the somatotype. Studies shows that its identification is useful in several areas of health and physical education. Currently, the main method for its estimation requires a lot of time, specialized training and the use of anthropometric equipment for measuring body parts. Some researches are looking for simplifying the exam, for example, by automating it with digital image processing. However, there are limitations, such as the adaptation of environments for capturing good quality photographs, an important requirement for its viability. Considering the benefits of obtaining a robust and fast method, such that simplifies the exam realization and reduces its costs, the main purpose of this work is to research and develop methods that are able to automatically predict information from the human body through images obtained in environments without previous preparation. In this context, this work proposes to combine machine learning techniques, in particular deep learning, in order to achieve detailed segmentation of the body contour of individuals in uncontrolled environments. In addition, gender identification and weight estimation are also done, both by facial analysis. Four different datasets are used: face2bmi, Wiki, Fisio-Somatotipo and Internet. The two first containing face images of people in general and the last two with photos of full bodies of athletes. The results shows that the study is promising, having achieved a good level of segmentation in adverse environments and 91.52% of accuracy for gender classification. It is also discussed how gender can influency on the weight prediction model, which reached a mean error of 5.56 kg as best result.

Keywords: Somatotype. Deep Neural Network. Human body segmentation. Uncontrolled Environments. Gender Classification. Body Weight Estimation.

Lista de ilustrações

Figura 1 – Caracterização visual do somatotipo humano. 24
Figura 2 – Fluxo da metodologia proposta neste trabalho. 27
Figura 3 – Somatotipos: ectomorfo, mesomorfo e endomorfo. 29
Figura 4 – Exemplo de imagens capturadas em três posições para aplicação do método de Sheldon, Stevens e Tucker (1940). 30
Figura 5 – Instrumentos antropométricos utilizados na obtenção de medidas durante a estimativa de somatotipo. 33
Figura 6 – Resultados das competições de classificação de imagens ILSVRC com algoritmos clássicos de Visão Computacional (verde escuro), algoritmos de *Deep Learning* (verde claro) e rotulação realizada por humanos (cinza). 39
Figura 7 – Exemplo de arquitetura de uma Rede Neural Convolucional destinada à classificação de meios de transporte. 40
Figura 8 – Diferença entre um número escalar, um vetor, uma matriz e um tensor. 41
Figura 9 – Convolução entre tensor de entrada com tamanho 3 × 3 e um *kernel* 2 × 2, que opera como uma janela deslizante sobre o tensor de entrada para formar o mapa de características de tamanho 2 × 2 42
Figura 10 – Estrutura do Modelo R-CNN. 43
Figura 11 – Estrutura do Modelo Fast R-CNN. A imagem é convertida em mapa de características e dele são propostas as regiões de interesse via *selective search*. Para cada RoI, camadas totalmente conectadas (FC) convertem o mapa em um vetor de características, que por sua vez é enviado para as camadas de saída: uma camada faz regressão linear para encontrar as 4 coordenadas dos *bounding boxes* e outra camada os classifica usando um operador *softmax*. 44
Figura 12 – Estrutura do Modelo FCN. 46
Figura 13 – Estrutura do Modelo Faster-CNN com a *Region Proposal Network*. . . 47
Figura 14 – Funcionamento da *RoI pooling*, operação de janela deslizante que gera âncoras com informações obtidas da RPN e alimenta as camadas de classificação e regressão de coordenadas dos *bbox* dos objetos. 48
Figura 15 – Mask R-CNN: associação de modelo Faster R-CNN com camadas FCN na saída. 50
Figura 16 – Estrutura do Modelo Mask R-CNN. 51
Figura 17 – Arquiteturas simplificadas das camadas de saída dos modelos. 51
Figura 18 – Funcionamento do *RoI Align*. 52
Figura 19 – Comparação do funcionamento das técnicas *RoI Pooling* e *RoI Align*. . 53

Figura 20 – Bloco residual utilizado na arquitetura ResNet. 54
Figura 21 – Imagem original e seu *trimap* à direita. 54
Figura 22 – Imagem de entrada (a), seu *trimap* (b) e o resultado da aplicação das técnicas *Closed-Form Matting* (c) e *Deep Image Matting* (d). 55
Figura 23 – Exemplos dos dois conjuntos de imagens da base Fisio-Somatotipo. Ambiente com pouco controle de iluminação: (a) e (b). Ambiente com controle maior: (c) e (d). 58
Figura 24 – Imagens da base Wiki: (a) classe masculina; (b) classe feminina; (c) fundo vazio. 59
Figura 25 – Imagens da base Face-to-BMI e os respectivos valores de IMC de cada indivíduo. 60
Figura 26 – Imagens da base Internet para a classe masculina. 60
Figura 27 – Imagens da base Internet para a classe feminina. 61
Figura 28 – Fluxo da metodologia proposta neste trabalho. 61
Figura 29 – Segmentação dos corpos de indivíduos com o modelo Mask R-CNN: (a) e (c) com *backbones* ResNet-50, em azul escuro; (b) e (d) com ResNet-101 em azul claro. 64
Figura 30 – Geração automática do *trimap*. A imagem original (a), após segmentação via Mask R-CNN é binarizada (b). A dilatação de (b) resulta em (c). A erosão de (b) resulta em (d). A subtração entre (c) e (d) resulta em (e). O *trimap* resultante é a figura (f), em escala de cinza. 65
Figura 31 – Resultado da aplicação das técnicas de *Alpha Matting*. Em (a) a imagem original e em (b) o seu *trimap*. Os resultados da *Deep Image Matting* e *Closed-Form Matting* são mostrados em (c) e (d), respectivamente. . . . 66
Figura 32 – Variação de espessura da borda de incerteza para: (a) Trimap-2; (b) Trimap-4; (c) Trimap-6; (d) Trimap-8. 67
Figura 33 – Imagem de treinamento (a) para o modelo Mask R-CNN e máscaras para aprendizado da classe “cão” (b) e classe “livro” (c). 69
Figura 34 – Interface da ferramenta LabelMe com uma máscara preenchida em imagem de treinamento da base de dados Wiki. 70
Figura 35 – Imagens de treinamento com máscaras considerando a cabeça inteira (a) e apenas a face (b). 71
Figura 36 – Comparação de resultado da segmentação da Mask R-CNN com *backbone* ResNet-50 (a) e ResNet-101 (b) para os atletas de identificação número 65 e 60 da base Fisio-Somatotipo. 77
Figura 37 – Mask R-CNN com *backbone* ResNet-101 (b) nem sempre segmenta todas as regiões melhor que com ResNet-50 (a). 78

Figura 38 – Geração de *trimaps* utilizando o método Mask R-CNN e pós-processamento. Imagem (a) resultante do processamento com Mask R-CNN, binarizada em (b) e o resultado para Trimap-2 (c) e Trimap-8 (d). 79
Figura 39 – Resultados do processamento das técnicas *Deep Image Matting* (a) e *Closed-Form Matting* (b). 80
Figura 40 – Resultados do *Closed-Matting* para *Trimap*-2, *Trimap*-6 e *Trimap*-8. . . 81
Figura 41 – Detalhes da comparação entre os resultados obtidos com *Trimap*-2 (a) e *Trimap*-8 (b). 82
Figura 42 – Imagem original ampliada da base Internet e sua segmentação final. . . 83
Figura 43 – Segmentação corporal para atleta da base Fisio-Somatotipo. (a) atleta 54, (b) segmentação obtida por Gonçalves (2017), (c) segmentação obtida neste trabalho e (d) imagem resultante. 83
Figura 44 – Segmentação corporal para atleta da base Fisio-Somatotipo. (a) atleta 63, (b) segmentação obtida por Gonçalves (2017), (c) segmentação obtida neste trabalho e (d) imagem resultante. 84
Figura 45 – Matriz de confusão para a melhor classificação de gênero obtida. 85
Figura 46 – Imagens das bases Fisio-Somatotipo e internet: (a) amostras de entrada; (b) resultado da segmentação das faces utilizando Mask R-CNN para classificação de gênero. 86
Figura 47 – Estimativa de máscara facial com falhas. 87
Figura 48 – *Boxplots* dos erros de estimativa de massa corporal para cada gênero. . 89
Figura 49 – Exemplo 1 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 102
Figura 50 – Exemplo 2 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 103
Figura 51 – Exemplo 3 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 104
Figura 52 – Exemplo 1 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 105
Figura 53 – Exemplo 2 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 106
Figura 54 – Exemplo 3 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação. 107

Lista de tabelas

Tabela 1 – Distribuição de imagens por base de dados e por gênero. 56
Tabela 2 – Resultados de classificação de gênero para cada abordagem avaliada. Classe masculina considerada como classe positiva para as métricas usadas. 84
Tabela 3 – Métricas para o primeiro conjunto de imagens da base Fisio-Somatotipo. 88
Tabela 4 – Métricas para o segundo conjunto de imagens da base Fisio-Somatotipo. 88

Lista de abreviaturas e siglas

Ac	*Acurácia*
BBOX	*Bounding Box*
BMI	*Body Mass Index*
CNN	*Convolutional Neural Network*
COCO	*Common Objects in Context*
Es	*Especificidade*
FAIR	*Facebook Artificial Intelligence Research*
FC	*Fully Connected*
FCN	*Fully Connected Network*
FN	*Falso Negativo*
FP	*Falso Positivo*
GP	*Gaussian Process*
GPU	*Graphics Processing Unit*
IMC	*Índice de Massa Corporal*
LSE	*Least Squares Estimation*
MAE	*Mean Absolute Error*
ME	*Mean Error*
MAPE	*Mean Absolute Percentage Error*
MSE	*Mean Square Error*
PNG	*Portable Network Graphics*
R-CNN	*Regions with CNN features*
ResNet	*Residual Network*
RMSE	*Root Mean Square Error*

RNA	*Rede Neural Articial*
RoI	*Region of Interest*
RPN	*Region Proposal Network*
Se	*Sensibilidade*
SVM	*Support Vector Machine*
SVR	*Support Vector Regression*
VN	*Verdadeiro Negativo*
VP	*Verdadeiro Positivo*

Sumário

1 INTRODUÇÃO 23
1.1 Descrição do Problema 24
1.2 Motivação 25
1.3 Hipótese e Objetivos 26
1.4 Método Proposto 27
1.5 Estrutura da Dissertação 27

2 SOMATOTIPO E TRABALHOS RELACIONADOS 29
2.1 Somatotipo 29
2.1.1 Método Clássico de Sheldon, Stevens e Tucker 30
2.1.2 Método Heath-Carter 31
2.1.3 Problemas dos Métodos Tradicionais de Medição 32
2.2 Trabalhos Relacionados a Somatotipo 34
2.3 Trabalhos Relacionados a Identificação de Gênero 35
2.4 Trabalhos Relacionados a Estimativa de IMC e Massa Corporal . . . 35
2.5 Segmentação de Imagens em Ambientes Não Controlados 36
2.6 Considerações Finais do Capítulo 37

3 TÉCNICAS UTILIZADAS 38
3.1 Redes Neurais Convolucionais 38
3.1.1 Camada de convolução 40
3.1.2 Camada de subamostragem 41
3.1.3 Camadas totalmente conectadas 42
3.2 Redes Convolucionais Profundas Aplicadas à Detecção de Objetos em Imagens 43
3.2.1 R-CNN: *Region-based Convolutional Neural Network* 43
3.2.2 Fast R-CNN 44
3.2.3 FCN - Fully Convolutional Network 45
3.2.4 Faster R-CNN 46
3.2.5 Mask R-CNN 49
3.3 *Transfer Learning* 52
3.4 *Alpha Matting* 54

4 METODOLOGIA 56
4.1 Bases de Dados 56
4.1.1 Base de dados Fisio-Somatotipo 57

4.1.2 Base de dados Wiki . . . 57
4.1.3 Base de dados Face-to-BMI . . . 58
4.1.4 Base de dados Internet . . . 59
4.2 Estratégia de Execução do Estudo . . . 61
4.3 Segmentação dos Indivíduos Utilizando Modelo Mask R-CNN Pré-Treinado . . . 62
4.4 Refinamento de Contornos Corporais com *Alpha Matting* . . . 63
4.5 Classificação de Gênero Utilizando Modelo Mask R-CNN Adaptado 68
4.5.1 Preparação de Dados Para o Treinamento de um Modelo Mask R-CNN . . 68
4.5.2 Estratégias de Aplicação da Mask R-CNN Para a Classificação de Gênero . 69
4.6 Estimativa de Massa Corporal via IMC . . . 72
4.7 Métricas de Avaliação . . . 72
4.7.1 Avaliação da Classificação de Gênero . . . 73
4.7.2 Avaliação da Estimativa de IMC e Massa Corporal . . . 74

5 RESULTADOS E DISCUSSÕES . . . 75
5.1 Implementação e Recursos Computacionais . . . 75
5.2 Resultados da Segmentação do Contorno Corporal . . . 76
5.2.1 Mask R-CNN com backbone ResNet-50 vs ResNet-101 . . . 76
5.2.2 *Deep Image Matting* vs *Closed-Form Matting* . . . 78
5.2.3 Segmentação em ambientes não controlados . . . 79
5.3 Resultados da Classificação de Gênero . . . 84
5.4 Resultados das Estimativas de IMC e Massa Corporal . . . 87
5.5 Comentários Finais . . . 90

6 CONCLUSÕES E TRABALHOS FUTUROS . . . 91
6.1 Conclusão . . . 91
6.2 Trabalhos Futuros . . . 92

REFERÊNCIAS . . . 94

ANEXO A – RESULTADO DA SEGMENTAÇÃO DA BASE INTERNET . . . 102

ANEXO B – AUTORIZAÇÃO DO USO DE IMAGENS DA IFBB-ES 108

1 Introdução

O funcionamento do corpo humano e como as suas características o influenciam são temas que sempre despertaram o interesse comum. Ao longo do tempo, diversos povos deixaram registros visuais acerca dessa complexa máquina em muitos objetos e estruturas, o que mostra que a curiosidade remonta a tempos antigos. Por certo, é possível afirmar que entender e ampliar o conhecimento sobre como ocorrem os fenômenos no corpo humano são, desde sempre, questões vitais para a sobrevivência e a manutenção da espécie humana.

Sousa (1942) relata que a partir do século XVI foram iniciados os estudos de Anatomia Topográfica, um ramo de estudo direcionado à aplicação prática da Medicina, sobretudo acerca das manifestações físicas e constituição das proporções corpóreas da espécie humana. Entretanto, as primeiras pesquisas com rigor científico nessa área surgiram apenas a partir do século XX.

Buscando entender mais sobre o funcionamento específico da estrutura física do corpo humano, Sheldon, Stevens e Tucker (1940) fundamentaram a teoria que criou o termo batizado por eles como somatotipo. Trata-se da estrutura do corpo representada por uma combinação de folhetos embrionários, sendo caracterizada por três componentes: ectomorfismo, mesomorfismo e endomorfismo.

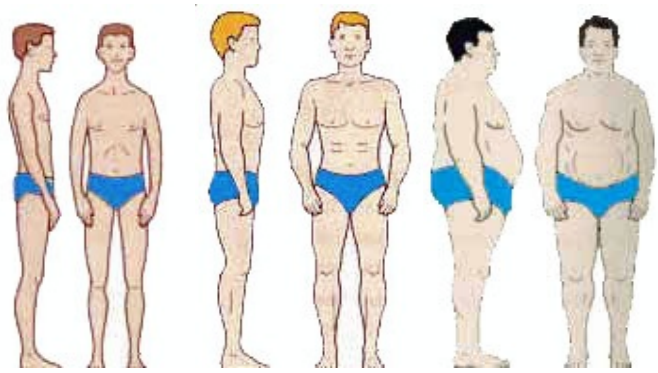
Segundo a teoria do somatotipo, em todas as pessoas há a presença de tais folhetos, mas a proporção entre as partes varia de indivíduo para indivíduo. Cada um dos três se refere a um aspecto principal, conforme segue: ectomorfismo, caracterizado pela linearidade corporal; mesomorfismo: marcado pela robustez músculo-esquelética; endomorfismo: reconhecido pela predominância de gordura corporal (SHELDON; STEVENS; TUCKER, 1940). A Figura 1 ilustra três indivíduos que tipificam as principais componentes do somatotipo humano.

As pesquisas que relacionam o somatotipo e a ocorrência de doenças tiveram início em meados do século XX (SHELDON; STEVENS; TUCKER, 1940; GARN, 1954; SHELDON, 1971). Alguns estudos buscaram mostrar como é viável direcionar a escolha de modalidades de atividades para pessoas buscando melhorias de saúde de maneira geral. Por exemplo, diversas áreas da educação física e da saúde utilizam a identificação do somatotipo de indivíduos, tornando-a uma ferramenta bastante relevante no meio esportivo e na prevenção de doenças (KOLEVA; NACHEVA; BOEV, 2000; KOLEVA; NACHEVA; BOEV, 2002; SINGH, 2007).

Sobre a análise do somatotipo, em relação ao aumento do desempenho de atletas de alto rendimento, alguns trabalhos mostraram que as características antropométricas, somatotípicas e morfológicas influenciam diretamente no sucesso do atleta de alto ren-

Figura 1 – Caracterização visual do somatotipo humano.

(a) Ectomorfo (b) Mesomorfo (c) Endomorfo



Fonte: (MACKENZIE, 2001)

dimento, por exemplo, em esportes como o futebol e o voleibol (BANDYOPADHYAY, 2007), o basquete (BAYIOS et al., 2006) e o judô (NOH; KIM; KIM, 2014).

Outros estudos demonstraram que em grupos de modalidades que exigem habilidades parecidas, é verificado que muitos atletas de categorias semelhantes possuem o somatotipo parecido (GUTNIK et al., 2015; PAVLOVIć, 2015).

Na tentativa de produzir mais conhecimento acerca do corpo humano, há trabalhos que modelam o seu formato a partir de descritores de Fourier (LEE; PARK, 1991), ou por meio de imagens para detectar padrões de movimentação (GREEN; GUAN, 2004), ou ainda para determinar o somatotipo (GONÇALVES, 2017).

1.1 Descrição do Problema

Com foco na realização do exame de identificação do somatotipo, Sheldon, Stevens e Tucker (1940) quantificaram as três componentes somatotípicas por intermédio de análises e medições realizadas diretamente em fotografias de indivíduos (CARTER, 2002). Considerando o avanço da tecnologia e a redução de custos no ramo da fotografia nas últimas décadas, é de se imaginar que o custo do método era elevado, pois era uma época em que haviam diversas limitações tecnológicas e de acesso aos equipamentos para trabalhos envolvendo essa área.

Para exemplificar algumas das dificuldades para uma análise assertiva dos resultados dos exames por imagem, podem ser citadas a baixa qualidade das câmeras disponíveis antigamente e a demora para revelar as fotos analógicas. Além disso, havia uma dedicação significativa em aspectos de preparação de ambientes, como iluminação apropriada, sempre com o objetivo de conseguir imagens que enriquecessem as características visuais necessárias para os exames.

Tempos depois, Heath e Carter (1967) propuseram um método que simplificava a infraestrutura tecnológica necessária para o exame do somatotipo, dessa vez passando a realizá-lo diretamente no corpo do indivíduo com instrumentos antropométricos, realizando medições em pontos definidos de forma padronizada e relacionando as medidas em equações cuidadosamente definidas em seu estudo.

Dobras cutâneas, perímetros e diâmetros corporais constituem parte do grupo de medições que eram aferidas em braços, pernas, troncos, entre outras partes, a fim de obter as informações necessárias para realizar a inferência do somatotipo. As relações matemáticas resultavam na proporção estimada das componentes de endomorfismo, mesomorfismo e ectomorfismo no corpo do indivíduo examinado.

O método de Heath e Carter (1967) carece de treinamento de profissionais habilitados para a realização do exame, uma vez que a utilização dos equipamentos de medição deve ser realizada de forma criteriosa para redução de erros. Além disso, nem todos os equipamentos para realizar as medições são tão portáteis.

Motivado para criar uma forma simplificada de identificação do somatotipo, Gonçalves (2017) propôs um método automático de realização do exame por meio de processamento digital de imagens, aproveitando os avanços e a facilidade atual de acesso aos equipamentos fotográficos para reduzir a necessidade de que a avaliação seja realizada por profissionais especialistas na tarefa, ampliando, portanto, o acesso à ela.

Embora Gonçalves (2017) tenha automatizado a obtenção de várias medidas corporais e proposto um método baseado em redes neurais para estimar o somatotipo, ainda assim o método precisa que certas informações sejam inseridas manualmente, sendo elas o gênero e a massa corporal das pessoas. Além disso, o método proposto tem a limitação de precisar de um ambiente controlado para estimar com precisão as medidas corporais.

1.2 Motivação

Simplificar a realização do exame de identificação do somatotipo fará com que uma maior parcela da população possa ter acesso a informações que contribuam para melhorar sua qualidade de vida, tanto em aspectos de saúde quanto para conhecer e tirar melhor proveito de suas próprias condições corporais na escolha de esportes.

A metodologia de exame proposta por Gonçalves (2017) mostrou a viabilidade de explorar a sua realização sem a necessidade de utilizar aparelhos antropométricos. Entretanto, há nela a dependência de preparação da iluminação do ambiente, por exemplo, com controle de saturação de intensidade de luz e redução de incidência de sombras no indivíduo registrado nas imagens. Além disso, ela não dispõe de inferência automática de

gênero e massa corporal dos indivíduos, duas características que influenciam no modelo proposto mas que, no momento, são informadas ao método manualmente.

Com os avanços recentes em tecnologias de visão computacional, desafios como a detecção e localização de humanos em imagens de ambientes com condições adversas de iluminação e variação de plano de fundo vêm sendo superados, ampliando consideravelmente o leque de aplicação das técnicas.

Há um campo muito rico que pode ser explorado em relação aos recursos de *hardware* e *software*, com técnicas de aprendizado de máquina que podem contribuir para um avanço no sentido de simplificar o exame de somatotipo. Eventualmente, pode ser atingido um nível de eficácia que permita automatizar a identificação do mesmo por meio de visão computacional, mesmo que as imagens sejam obtidas em ambientes adversos e sem controle específico a priori.

1.3 Hipótese e Objetivos

A hipótese deste trabalho é a de que é possível contribuir para a pesquisa de identificação automática de somatotipo por meio da identificação e segmentação detalhada do contorno corporal de indivíduos, por meio de análise de imagens digitais, mesmo em ambientes não controlados, visando obter informações úteis como silhueta, gênero e massa corporal.

Dessa forma, este trabalho objetiva aplicar técnicas de processamento de imagens e aprendizado de máquinas, mais especificamente aprendizado profundo, para expandir o contexto de aplicação da proposta formulada por Gonçalves (2017). O intuito é viabilizar sua aplicação em ambientes não controlados, ou seja, em locais em que não é necessária uma preparação a priori do fundo e da iluminação. Com isso, é desejado obter a segmentação corporal detalhada dos indivíduos, com técnicas que se apliquem tanto à base de imagens utilizada por Gonçalves (2017), quanto para novas imagens, considerando variedades de ambientes.

Além disso, é objetivo deste trabalho acrescentar a identificação automática de gênero dos indivíduos por meio de análise facial, dado que no método citado esta informação era fornecida externamente, de forma manual. Finalmente, acredita-se que é possível aliar ao estudo a estimativa de IMC (Índice de Massa Corporal) também por análise da face e, de forma indireta, inferir a massa corporal dos indivíduos, integrando à automação do método mais um dado que era fornecido de forma manual. Portanto, os objetivos específicos desta pesquisa são:

1. Realizar a segmentação de contorno corporal humano a nível de pixel, mesmo em ambientes não controlados;

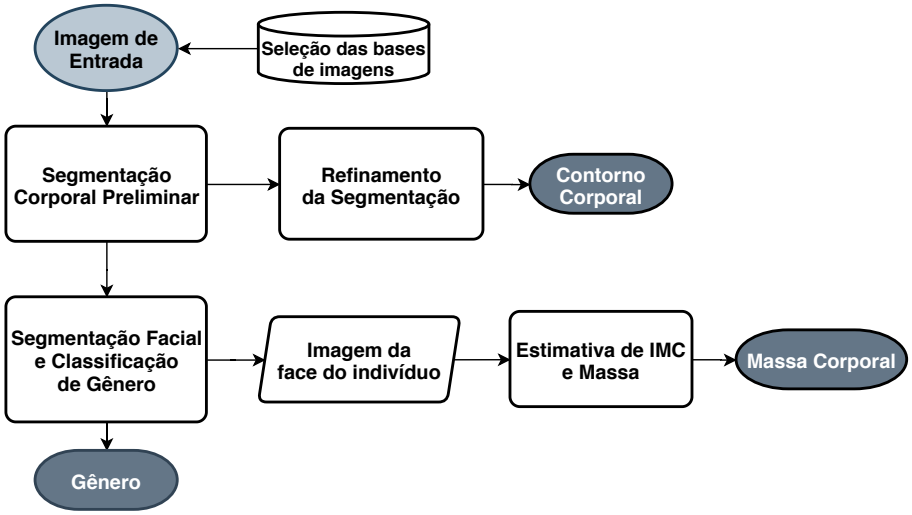
2. Realizar a identificação de gênero dos indivíduos via análise facial;

3. Realizar a estimativa de massa corporal dos indivíduos via análise facial.

1.4 Método Proposto

A Figura 2 mostra um diagrama simplificado da metodologia proposta neste trabalho. Dada uma imagem de entrada de uma base de dados, inicialmente é realizada a localização da pessoa na imagem e uma segmentação básica, que depois é refinada para obter o contorno corporal detalhado da mesma. Com a localização inicial, é encontrada a sua face, a qual é utilizada para a identificação do gênero e estimativas do IMC e massa corporal. Mais detalhes da abordagem proposta estão no Capítulo 4.

Figura 2 – Fluxo da metodologia proposta neste trabalho.



Fonte: o próprio autor.

1.5 Estrutura da Dissertação

Este trabalho possui a seguinte estrutura.

No Capítulo 2 é apresentada uma revisão teórica sobre o somatotipo, com um histórico acerca dos procedimentos clássicos e atuais utilizados na realização do exame. Também são apresentados alguns trabalhos relacionados com esta pesquisa.

No Capítulo 3 é realizada uma descrição de algumas técnicas de segmentação, que evoluíram até constituir o algoritmo mais utilizado ao longo desta dissertação. Além disso, também serão mostradas técnicas auxiliares que fazem parte do contexto de processamento digital de imagens e que foram inseridas neste trabalho para formação de uma metodologia híbrida de segmentação corporal.

No Capítulo 4 são apresentadas as características das imagens das bases de dados utilizadas e será detalhada a metodologia adotada para o alcance de cada um dos três objetivos específicos que são aqui propostos.

No Capítulo 5 são exibidos e discutidos os principais resultados obtidos para cada etapa de realização do trabalho.

Finalmente, o Capítulo 6 consiste em conclusões e caminhos que se abrem para trabalhos futuros.

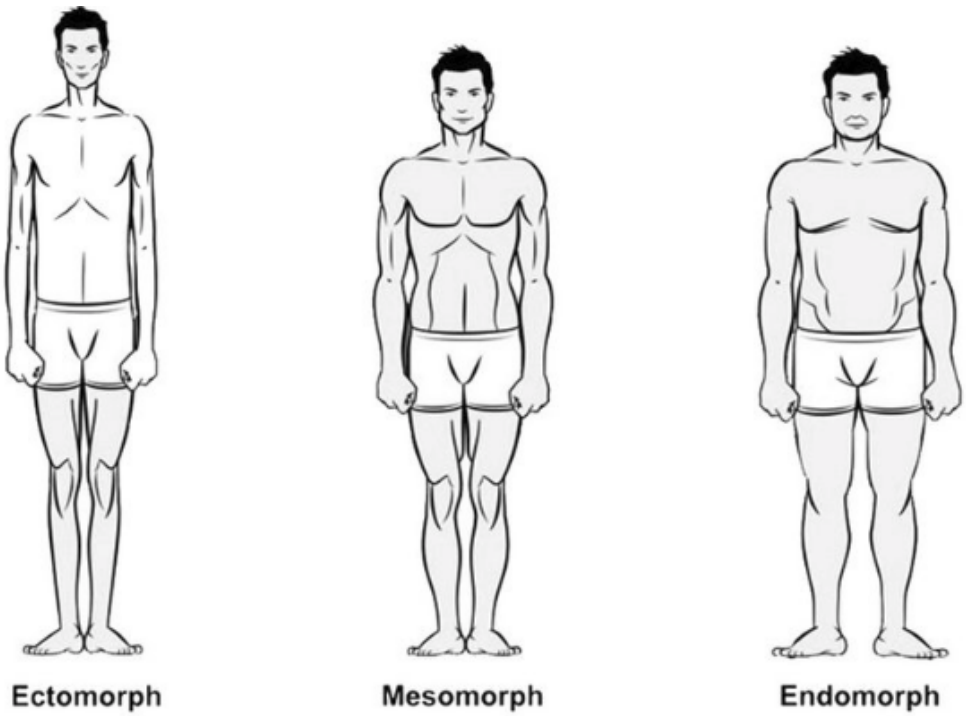
2 Somatotipo e Trabalhos Relacionados

Neste capítulo serão apresentados aspectos teóricos acerca de somatotipo. Serão abordados o seu significado e a sua importância. Em seguida, será apresentada uma breve revisão histórica acerca de alguns métodos de obtenção do somatotipo e algumas de suas limitações. Por fim, serão apresentados alguns trabalhos relacionados com o tema no contexto de estimativa do somatotipo e e suas principais limitações, sendo estas as motivações para a realização do presente trabalho.

2.1 Somatotipo

O termo somatotipo foi originalmente apresentado por Sheldon, Stevens e Tucker (1940) e é utilizado para classificar a estrutura física do corpo humano como uma combinação de três componentes: endomorfismo, mesomorfismo e ectomorfismo (SHELDON; STEVENS; TUCKER, 1940). A Figura 3 ilustra indivíduos que representam as características de cada uma das três classes que, combinadas, compõem o somatotipo de uma pessoa.

Figura 3 – Somatotipos: ectomorfo, mesomorfo e endomorfo.



Fonte: Adaptado de (STREUBER et al., 2016).

A componente ectomorfa é associada com a proporção entre altura e massa corporal,

a mesomorfa com a quantidade de massa magra e a endomorfa com a concentração de gordura corporal. Casos em que uma das componentes é predominantemente presente são chamados de "tipos puros", mas a maioria dos indivíduos possuem características de pelo menos duas classes (PAVLOVIć, 2015).

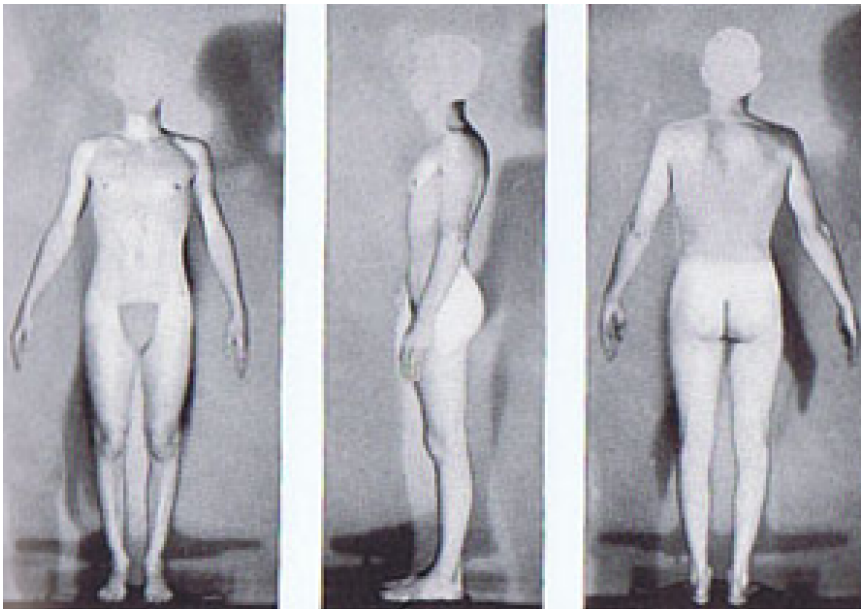
Estudos mostram que a componente predominante prevalece durante toda a vida da pessoa, sendo responsável por prover as características mais marcantes do corpo, ainda que a proporção entre as componentes possa mudar com o passar do tempo (FILHO, 2003).

Nas áreas de educação física e saúde, a identificação do somatotipo de cada indivíduo é uma métrica relevante na análise de pacientes e atletas. Diversos estudos mostram que é possível observar o efeito de determinadas doenças e características de nutrição através do somatotipo, o que contribui para a orientação de dietas e treinamentos, seja com foco em melhorias específicas de saúde, ou até mesmo para aprimoramento de desempenho de atletas competidores (SILVA; TRINDADE; ROSE, 2003).

2.1.1 Método Clássico de Sheldon, Stevens e Tucker

Para estudar e conceituar o somatotipo, Sheldon, Stevens e Tucker (1940) criaram um método que consiste em realizar medições em imagens obtidas de pessoas em três posições: perfil, frontal e dorsal. A posição do indivíduo em cada classe de imagem era padronizada: braços esticados e afastados do corpo, pernas esticadas e afastadas e as palmas das mãos voltadas para as coxas, conforme exibido na Figura 4.

Figura 4 – Exemplo de imagens capturadas em três posições para aplicação do método de Sheldon, Stevens e Tucker (1940).



Fonte: (SHELDON; STEVENS; TUCKER, 1940).

Para efetuar as medições, Sheldon, Stevens e Tucker (1940) dividiram as imagens em cinco regiões principais:

1. Cabeça, pescoço e face;
2. Tórax;
3. Ombros, membros superiores e mãos;
4. Abdômen;
5. Membros inferiores e pés.

Na região da cabeça, são medidas as larguras faciais na parte superior e também inferior da face, além da largura do pescoço de perfil e de frente. Na região do tórax, mede-se a largura superior do tronco, a espessura do tronco e o menor diâmetro perceptível da cintura. Na terceira região, mede-se a espessura superior do braço, a espessura superior do antebraço e a espessura inferior do mesmo. Na região abdominal é medida a espessura e a largura do tronco. Na quinta, e última região, são medidas a espessura superior da coxa, espessura superior da perna logo abaixo do joelho, espessura no ponto de maior volume da panturrilha e espessura no ponto mais estreito do tornozelo (SHELDON; STEVENS; TUCKER, 1940).

Portanto, dezesseis medições eram realizadas nas fotografias, e relacionadas de forma a caracterizar o somatotipo do indivíduo, considerando a combinação das três componentes principais.

O pioneirismo de Sheldon, Stevens e Tucker (1940) resultou em uma técnica que possibilitou o início dos estudos de somatotipo, porém a falta de tecnologia na época tornou a sua aplicação prática um tanto complexa. Fotografar indivíduos em posições padronizadas e em ambientes com iluminação controlada, revelar as fotografias e realizar medições diretamente nas imagens eram tarefas demoradas, principalmente considerando os equipamentos fotográficos que eram utilizados na época. Além disso, havia a necessidade de pessoas altamente capacitadas no assunto, uma vez que é fundamental tentar minimizar erros humanos nos exames (FILHO, 2003).

Com o objetivo de simplificar o método de Sheldon, Stevens e Tucker (1940) e não depender de medições em imagens, Heath e Carter (1967) propuseram uma nova forma e dexame, que possibilitou realizar as medições diretamente no corpo dos indivíduos.

2.1.2 Método Heath-Carter

Criado há cerca de 50 anos, este método utiliza equipamentos antropométricos para realizar dez medições distintas nos indivíduos de forma direta, e não por meio de

imagens. Para a realização dessa tarefa, são medidos: altura da pessoa, massa corporal, diâmetro ósseo bi-epicondilar de úmero e de fêmur, perímetro do braço flexionado em contração máxima, perímetro de panturrilha, dobras cutâneas de tríceps, supra espinhal, subescapular e panturrilha medial. As dez medidas são aplicadas em equações definidas e então é obtido um resultado quantitativo, o que permite classificar o somatotipo (HEATH; CARTER, 1967).

Sendo amplamente aplicada atualmente, a técnica de Heath e Carter (1967) utiliza cinco equipamentos, que podem ser vistos na Figura 5. Uma balança é utilizada para medir a massa corporal do indivíduo. A estatura é medida com um estadiômetro de cabeça móvel. As dobras cutâneas são medidas com o auxílio de um adipômetro. Com um paquímetro são medidos os diâmetros ósseos e, finalmente, uma trena antropométrica é usada para medir os perímetros (HEATH; CARTER, 1967).

Uma vez tomadas as medidas, há duas possibilidades para chegar ao resultado do somatotipo. A primeira é inserir os dados em um formulário de valores tabelados, marcando os valores mais próximos daqueles obtidos nas medições. Este é um método pouco confiável, devido à inserção de erros de aproximação e arredondamentos (TÓTH et al., 2014). A segunda possibilidade é utilizar os dados diretamente nas equações propostas por Heath e Carter (1967), que resultam em valores mais assertivos e aumentam a confiabilidade do exame. O resultado do somatotipo é representado por uma tripla ordenada com os valores de endomorfismo, mesomorfismo e ectomorfismo. As componentes devem ser arredondadas para múltiplos de 0,1 ou 0,5, a depender da aplicação, e não podem ser negativas (CARTER, 2002).

2.1.3 Problemas dos Métodos Tradicionais de Medição

Havia fatores complicadores da sistemática de avaliação de somatotipo utilizando o método clássico de Sheldon, Stevens e Tucker (1940), por exemplo, em relação à obtenção das fotografias. Era uma tarefa que demandava a preparação de ambientes com controle de iluminação e posicionamentos, tanto do paciente quanto do fotógrafo. Por este motivo, o método de Heath e Carter (1967) ganhou maior aplicação prática.

Por sua vez, o exame proposto por Heath e Carter (1967) é demorado, precisa de equipamentos que não são tão portáteis, e possui vários detalhes em sua forma de aplicação, o que agrega complexidade e demanda de profissionais treinados. Outro ponto importante é a precisão. Alguns aparelhos antropométricos possuem precisão de 0,1mm, como é o caso do adipômetro de Harpender, enquanto em outros, menos precisos, a imprecisão chega a 1mm, como é o caso das fitas métricas e estadiômetros (FILHO, 2003).

Devido à possibilidade de adição de erro humano ao erro intrínseco dos aparelhos antropométricos, é recomendável que os profissionais que realizam o exame sejam altamente

Figura 5 – Instrumentos antropométricos utilizados na obtenção de medidas durante a estimativa de somatotipo.

(a) Balança (b) Estadiômetro (c) Adipômetro

(d) Paquímetro (e) Trena

Fonte: (ONLINE, 2020).

capacitados para tal tarefa (SILVA et al., 2014). O objetivo da capacitação é reduzir o erro, ou ainda, fazer com que a imprecisão inserida seja próxima de um valor constante, considerando a individualidade de cada profissional (FILHO, 2003).

Em outras palavras, diferentes profissionais podem realizar exames que resultem, para um mesmo paciente, em diagnósticos quantitativos aproximados, porém diferentes. Considerando um mesmo examinador, é importante tentar minimizar a diferença de resultado para um paciente específico (FILHO, 2003). Uma forma de contornar a imprecisão é garantir que o exame seja feito duas, ou até mesmo três vezes para um mesmo par examinador-paciente e que o resultado final seja a média dos resultados parciais (FILHO, 2003).

2.2 Trabalhos Relacionados a Somatotipo

Algumas pesquisas recentes têm como objetivo principal reduzir algumas das dificuldades relacionadas aos métodos tradicionais de estimativa de somatotipo, sobretudo a velocidade e a acurácia do exame.

Com o intuito de acelerar o diagnóstico do exame por meio da computação, Houcine, Ahmed e Saddek (2014) propuseram um *software* que deriva do método de Heath e Carter (1967) para a obtenção do somatotipo, porém ainda depende de humanos realizando as medições corporais.

A pesquisa realizada por Houcine, Ahmed e Saddek (2014) apresenta uma outra proposta de método matemático para estimar o somatotipo, mas com o foco principal de automatizar a obtenção das medidas corporais, que são de fato a entrada do algoritmo. Para isso, foi proposta a utilização de equipamentos eletrônicos que, acoplados ao corpo humano, façam as medições necessárias e já as informem para o *software* integrado ao equipamento, gerando finalmente um diagnóstico automatizado. Porém, esta abordagem ainda depende de ação humana para posicionar corretamente os equipamentos, o que pode exigir algum treinamento específico.

O trabalho de Silva et al. (2014) faz uso de métodos meta-heurísticos para, dado um conjunto de medidas de determinado paciente ou atleta, não só determinar o seu somatotipo atual, mas também compará-lo com o somatotipo que é desejado atingir no futuro. Com base nessa relação, o método consegue auxiliar na sugestão de planos de treino e dietas, porém também depende de profissionais capacitados para a obtenção manual das medidas corporais.

Visando eliminar a necessidade de obtenção manual de medidas e fatores como imprecisão humana e lentidão para execução dos exames, foi proposto por Gonçalves et al. (2016) um método de avaliação do somatotipo utilizando modelos de aprendizado de máquina, e não mais a utilização das equações definidas por Heath e Carter (1967). Posteriormente, Gonçalves (2017) apresentou um método de avaliação automática de somatotipo, estimando medidas de atletas fisiculturistas por meio de imagens digitais e aplicando-as em modelos obtidos por técnicas de aprendizado de máquina.

Entretanto, o método proposto por Gonçalves (2017) possui algumas limitações, como a necessidade de ambientes controlados e previamente preparados, com iluminação uniforme, fundo homogêneo e distinto do corpo humano. Além disso, há a limitação de uma pessoa por imagem, além de ser fornecido por meio externo a massa corporal e o gênero do indivíduo.

Este trabalho tem como objetivo mitigar ou reduzir os efeitos dessas limitações, de forma a tornar mais prática a obtenção de somatotipo de indivíduos.

2.3 Trabalhos Relacionados a Identificação de Gênero

Os estudos de classificação de características humanas por meio de análise facial com imagens evoluiu substancialmente nas últimas décadas, e recentemente as redes neurais convolucionais têm tido destaque nesta tarefa. Lopes et al. (2017) propuseram um sistema de reconhecimento de expressões faciais que combina redes neurais convolucionais e técnicas de pré-processamento de imagem. Outro trabalho envolvendo a classificação de emoções por análise facial foi realizado por AlMarri (2019), o qual utilizou redes profundas baseadas na proposição automática de regiões de interesse. Estes trabalhos são apenas alguns, dentre inúmeros, que indicam a relevância de redes neurais profundas no campo dos estudos de análise facial computadorizada.

Outro tipo de aplicação de técnicas de aprendizado de máquina em análise facial, e que é um dos focos desta dissertação, é a estimativa de gênero. Há diversos estudos empregando modelos de aprendizado profundo visando realizar tal tarefa, como em (LEVI; HASSNCER, 2015) e (SENDIK; KELLER, 2019), onde ambos também realizam a estimativa de idade do indivíduo analisando a face. Há também avanços no sentido de possibilitar estas análises em tempo real: utilizando redes convolucionais com arquiteturas simplificadas, como em (LIEW et al., 2016), ou ainda com redes mais sofisticadas rodando em *hardwares* um pouco mais robustos (SUN, 2019).

2.4 Trabalhos Relacionados a Estimativa de IMC e Massa Corporal

Ainda sobre pesquisas relacionadas com análise facial por imagem, alguns estudos já exploraram a possibilidade de estimar o Índice de Massa Corporal (IMC) e a massa corporal de indivíduos por meio de processamento de dados inferidos de fotos de faces. No trabalho de Coetzee et al. (2010) foram definidas três medições a serem realizadas nas imagens faciais e foi testada a relação estatística entre as medidas selecionadas e o IMC, comprovando o potencial de viabilidade do método.

O IMC é uma métrica utilizada para definir características baseadas nas medidas de altura e massa corporal e classificar humanos dentre diferentes grupos. É uma métrica comumente interpretada sobretudo em relação à gordura corporal (NUTTALL, 2015), tendo sido criada para análise de populações em geral, sendo utilizada para acompanhamento de saúde de indivíduos principalmente durante a infância e, em diversas aplicações, ao longo da vida adulta. Também é utilizada para mensurar a qualidade de saúde populacional em programas governamentais.

Wen e Guo (2013) também exploraram a predição de IMC por meio de algoritmos de regressão aplicados às coordenadas de pontos específicos marcados nas faces. O problema foi modelado como uma aplicação de aprendizado estatístico. As técnicas de regressão

utilizadas na ocasião foram *Support Vector Regression* (SVR), *Least Squares Estimation* (LSE), e *Gaussian Process* (GP).

Uma vez estimado o IMC ($\widehat{IMC}$), a massa corporal estimada do indivíduo ($\widehat{Massa}$) pode ser calculado utilizando a altura real conhecida, conforme a Equação 2.1.

$$\widehat{Massa} = \widehat{IMC} \cdot (Altura)^2 \ [kg] \tag{2.1}$$

Outros trabalhos mais recentes têm explorado a aplicação de redes neurais profundas para estimar o IMC. Utilizando a imagem do rosto, Kocabey et al. (2017) criaram um modelo capaz de estimar para um par de faces qual dos indivíduos é o mais pesado. Ao comparar os resultados com as estimativas realizadas por humanos, em diversas faixas de IMC o resultado do modelo se aproxima muito ao da inferência humana.

Indo além, Dantcheva, Bremond e Bilinski (2018) propuseram a estimativa direta do IMC com uma rede neural profunda adaptada para o problema de regressão e, além disso, também mostraram alguns aspectos estatísticos sobre como o gênero pode influenciar no modelo. Portanto, as duas pesquisas citadas demonstraram o quão promissora é a utilização de redes profundas nesta linha de pesquisa.

2.5 Segmentação de Imagens em Ambientes Não Controlados

A segmentação de imagens consiste na tarefa de subdividi-las em regiões ou objetos que as compõem. O tipo de problema a ser resolvido define o nível de detalhe que a segmentação deve atingir. Segundo Gonzalez e Woods (2006), uma das tarefas mais complexas no processamento de imagens é a segmentação de imagens não triviais, sendo que o sucesso ou o fracasso da análise automatizada, independente da aplicação computacional, é diretamente influenciado pela precisão da segmentação.

Falhas de distribuição de iluminação, presença de ruídos, sobreposição de objetos, falta de padronização de posicionamento, diferenças de textura, variações de profundidade, sombras, além de grande variabilidade nos cenários e planos de fundo são barreiras que vêm sendo vencidas. Avanços recentes nas pesquisas passaram a possibilitar a extração de características e construção de conhecimento baseado em imagens obtidas nestes ambientes desafiadores (PAROLIN, 2011).

Nos últimos anos houve progresso significativo nos estudos de técnicas voltadas para detecção de objetos e segmentação semântica e de instâncias. Entretanto, a terceira tarefa ainda é especialmente desafiadora (HE et al., 2017).

Na segmentação semântica, além de detectar que determinada imagem possui um objeto conhecido, é necessário obter a região exata onde o objeto se encontra e também diferenciá-lo das outras classes de objetos disponíveis na imagem.

Por sua vez, na segmentação de instâncias, cada um dos objetos deve ser classificado e separado mesmo que pertençam à mesma classe, ainda que exista sobreposição. Inúmeros métodos são capazes de realizar segmentação de instâncias, porém o nível de detalhe e refinamento da segmentação ainda é um grande desafio. Muitas aplicações carecem de evolução para ampliar as possibilidades de utilização em larga escala, sobretudo em ambiente industrial ou em cenários de alta demanda por robustez e acurácia, como em áreas da saúde e da segurança.

2.6 Considerações Finais do Capítulo

O objetivo deste capítulo foi apresentar o histórico relacionado ao estudo do somatotipo, desde o início dos estudos de Sheldon, Stevens e Tucker (1940) até a criação da técnica de Heath e Carter (1967), que segue sendo utilizada atualmente.

Posteriormente, foram apresentadas algumas dificuldades relacionadas às técnicas clássicas de obtenção do somatotipo, sobretudo com relação à necessidade de profissionais com alta capacitação e a aplicação dispendiosa da técnica com o uso de equipamentos de medição antropométrica. Nesse contexto, foram citados alguns estudos propostos para simplificar o processo de estimativa de somatotipo.

Em especial, foi citado o trabalho de Gonçalves (2017), que consiste na estimativa do somatotipo de indivíduos por meio de análise automática de imagens em ambientes controlados. Tal pesquisa é o ponto de partida deste trabalho, que propõe a aplicação de técnicas atuais de visão computacional em imagens obtidas em ambientes não controlados. Aqui, o foco é automatizar a obtenção de informações importantes para a estimativa do somatotipo: contorno corporal do indivíduo, gênero e massa corporal. Todas essas informações são necessárias e suficientes para o modelo desenvolvido por Gonçalves (2017) ser automatizado[1].

[1] Considerando que as imagens das pessoas sejam obtidas nas posições estabelecidas pelo protocolo desenvolvido em Gonçalves (2017).

3 Técnicas Utilizadas

Neste capítulo serão apresentadas as principais técnicas de detecção e segmentação de objetos de imagens utilizadas neste trabalho, com destaque para as redes neurais convolucionais e foco em suas aplicações para ambientes não controlados. Serão explicados os funcionamentos de alguns modelos recentes de redes profundas e apresentados os avanços que resultaram na principal técnica utilizada neste trabalho. Também serão apresentados algoritmos auxiliares de processamento de imagens, além de aspectos relacionados às suas aplicações nas tarefas de proposição de regiões de objetos, segmentação e classificação de instâncias.

3.1 Redes Neurais Convolucionais

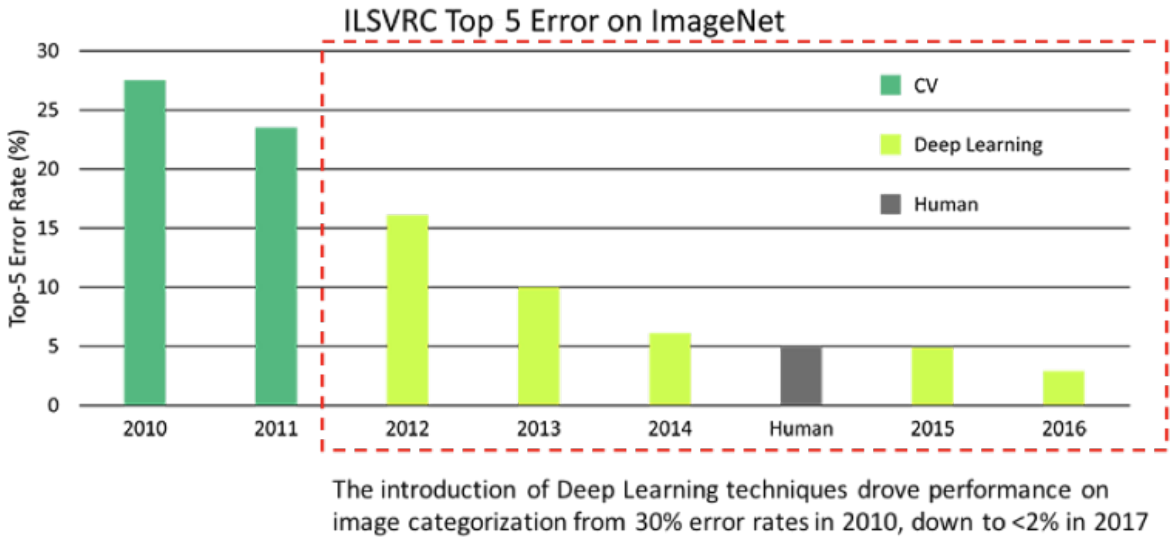
Na última década, avanços em recursos de *hardware* proporcionaram soluções computacionais cada vez mais robustas no campo científico. Com o avanço da aplicação de algoritmos de aprendizado profundo, está cada vez mais em evidência a aplicação das Redes Neurais Convolucionais, do inglês *Convolutional Neural Networks* (CNN), motivadas pela usabilidade, adaptabilidade e grande variedade de aplicações em diversas áreas de pesquisa (GOODFELLOW; BENGIO; COURVILLE, 2016).

Propostas originalmente por Lecun et al. (1998), as CNN vêm sendo amplamente utilizadas em processamento de imagens, especialmente a partir do início da última década. Sua utilização foi impulsionada pelo recorde alcançado por Krizhevsky, Sutskever e Hinton (2012) na competição ILSVRC-2012 (*ImageNet Large Scale Visual Recognition Challenge 2012*) (RUSSAKOVSKY et al., 2015). Na ocasião, foi treinada uma CNN que bateu em mais de 10% o resultado de todas as técnicas concorrentes para a tarefa de classificação de objetos em imagens para 1000 categorias.

Até o ano de 2011 as técnicas campeãs do ILSVRC não eram baseadas em CNN e sua utilização era desacreditada por parte da comunidade[1] (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). A Figura 6 mostra os resultados de taxa de erro dos modelos campeões das competições ILSVRC desde 2010. Vale destacar que, a partir de 2012, as técnicas campeãs são baseadas em CNN e, em 2015, a técnica ResNet (*Residual Neural Network*) (HE et al., 2016) entrou para a história ao superar os resultados de testes feitos por humanos.

[1] Segundo Krizhevsky, Sutskever e Hinton (2017), por volta de 2008 um artigo de Yann LeCun e seus colaboradores foi recusado em uma importante competição internacional de Visão Computacional por utilizar redes neurais ao invés de um sistema cuidadosamente *hand-designed*. Era considerado que para o treinamento eficaz de um modelo, jamais seria suficiente apenas utilizar imagens e rótulos de objetos contidos nelas.

Figura 6 – Resultados das competições de classificação de imagens ILSVRC com algoritmos clássicos de Visão Computacional (verde escuro), algoritmos de *Deep Learning* (verde claro) e rotulação realizada por humanos (cinza).



Fonte: (RUSSAKOVSKY et al., 2015).

O nome CNN tem referência ao fato de este tipo de rede utilizar o operador matemático de convolução, que consiste em um tipo de operação linear, conforme será descrito adiante. Explicando seu funcionamento de forma sucinta, segundo Goodfellow, Bengio e Courville (2016), as CNN são simplesmente redes neurais que utilizam o operador de convolução no lugar de multiplicações matriciais genéricas em ao menos uma de suas camadas do modelo.

Uma característica interessante das CNN é que são especialmente eficazes em tarefas de processamento de imagem pois, em baixo nível, se beneficiam da estrutura espacial dos dados, criando relacionamentos entre *pixels* de uma vizinhança específica, ao invés de tratar de forma equivalente regiões distantes (ACADEMY, 2019).

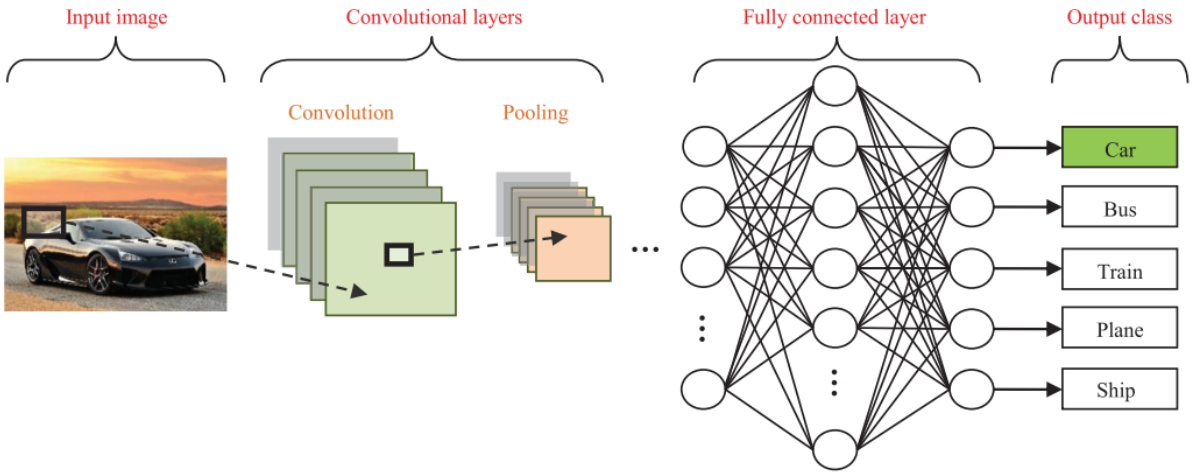
Além disso, são de fácil utilização e adaptação, propiciando testes em larga escala de dados com alta produtividade graças aos treinamentos e processamentos de modelos, especialmente utilizando *Graphics Processing Unit* (GPU), onde é possível paralelizar em larga escala os cálculos de multiplicações matriciais realizados. Treinamentos de CNN em GPU foram realizados pela primeira vez por Chellapilla, Puri e Simard (2006) e revolucionaram o tema.

Redes neurais convolucionais são compostas por várias camadas de processamento linear e não linear, onde a saída de cada camada anterior alimenta a entrada de sua camada posterior (DENG, 2014). Em geral, a CNN utiliza combinações de três tipos básicos de

camadas em sua arquitetura[2]: camada de convolução, camada de subamostragem (*pooling*) e camada totalmente conectada (KANG; WANG, 2014).

A Figura 7 mostra um exemplo simplificado de arquitetura de CNN para realização de classificação de tipos de meios de transporte, com destaque para as camadas convolucionais e totalmente conectadas.

Figura 7 – Exemplo de arquitetura de uma Rede Neural Convolucional destinada à classificação de meios de transporte.



Fonte: (RAWAT; WANG, 2017).

3.1.1 Camada de convolução

As camadas de convolução utilizam o operador linear mais presente nas CNN: a convolução, que é realizada deslizando uma máscara (*kernel*) de tamanhos e valores definidos e com um passo (*stride*) fixo sobre cada *pixel* da imagem de entrada. Aproveitando a condição de que *pixels* próximos estão diretamente relacionados, a camada convolucional extrai características que contribuem para a tarefa de identificação de padrões nas imagens. O resultado da operação de convolução é um mapa de características (*feature map*) da imagem de entrada.

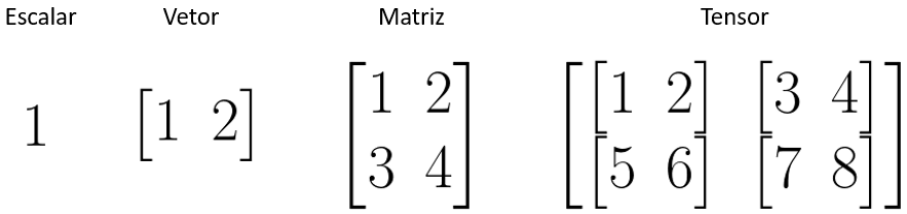
Segundo Gonzalez e Woods (2006), a convolução é uma técnica bastante presente em aplicações de processamento de imagens. A operação é definida pela Equação (3.1), onde $I(x, y)$ é a matriz correspondente à imagem, $K(s, t)$ é o *kernel* de dimensões $m \times n$, sendo $m = 2a + 1$ e $n = 2b + 1$, e $F(x, y)$ é o mapa de características resultante e a e b são as dimensões do *kernel*.

$$F(x, y) = K(s, t) * I(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} K(s, t)\, I(x - s, y - t) \tag{3.1}$$

[2] Estes são os tipos mais comuns. Entretanto, existem outros tipos de camadas que também podem ser utilizadas, mas que variam entre as diferentes arquiteturas.

Em aplicações de CNN, geralmente as entradas são entidades multidimensionais que representam os canais das imagens e os *kernels* são entidades multidimensionais de parâmetros adaptados pelo algoritmo de aprendizado. Para estas entidades multidimensionais é associada a nomenclatura tensor (GOODFELLOW; BENGIO; COURVILLE, 2016), cujo exemplo é ilustrado na Figura 8. Devido à nomenclatura, muitas das operações realizadas em uma CNN são denominadas operações tensoriais. Um escalar, um vetor e uma matriz são casos particulares de tensores, respectivamente de ordem zero, um e dois.

Figura 8 – Diferença entre um número escalar, um vetor, uma matriz e um tensor.



Fonte: o próprio autor.

A Figura 9 mostra uma operação de convolução entre dois tensores 2D. O *kernel* possui dimensão 2×2, o tensor de entrada possui dimensão 3×3 e o *stride* é igual a 1.

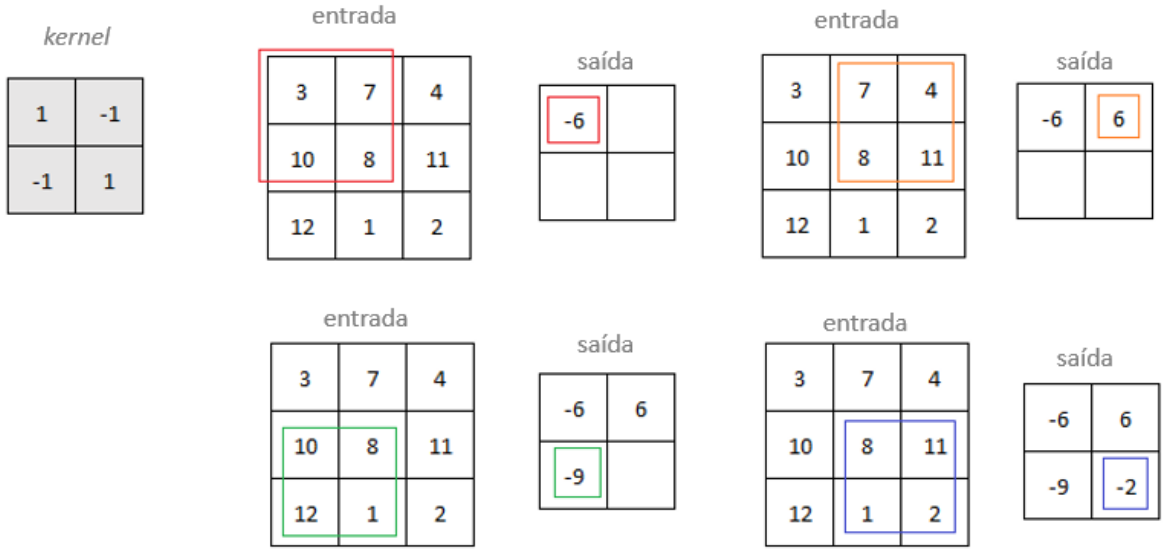
Após o processo de convolução, é comum passar o mapa de características por uma função de ativação não linear, o que adiciona uma maior capacidade de realização de tarefas complexas às redes. Uma das funções de transferência mais comumente usadas é a ReLU (*Rectified Linear Units*), devido à sua eficácia em relação ao treinamento aliada à velocidade com a qual a operação é realizada (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

3.1.2 Camada de subamostragem

No contexto de processamento de imagens, a subamostragem reduz a resolução de uma imagem e busca não alterar significativamente seu aspecto, evitando a perda de informação relevante. No contexto de CNN, o objetivo é reduzir a dimensão do mapa de características mantendo as informações principais do mapa original (YAMASHITA et al., 2018).

Algumas formas de operação de subamostragem são a seleção do valor máximo (*max pooling*), mínimo (*min pooling*), ou ainda os valores médios (*average pooling*), dentre outras. A redução das dimensões também reduz a quantidade de pesos a serem treinados no modelo, o que resulta em menor custo computacional. Devido à simplificação do modelo, as operações de subamostragem podem contribuir para sua generalização e reduzir o *overfitting* (KHAN et al., 2019).

Figura 9 – Convolução entre tensor de entrada com tamanho 3 × 3 e um *kernel* 2 × 2, que opera como uma janela deslizante sobre o tensor de entrada para formar o mapa de características de tamanho 2 × 2 .



Fonte: o próprio autor.

As camadas de convolução e subamostragem consistem na base da rede e fazem o processo de extração de características.

3.1.3 Camadas totalmente conectadas

As camadas totalmente conectadas das CNN, do inglês, *fully connected* (FC), funcionam de maneira similar às das clássicas RNA (Redes Neurais Artificiais). Localizadas no topo da CNN, as camadas FC são responsáveis pela classificação dos padrões de entrada, o que é feito conectando cada entrada a uma saída do modelo por um peso (YAMASHITA et al., 2018).

Ao contrário das operações realizadas nas camadas de convolução e subamostragem, a camada totalmente conectada realiza uma operação global. Esta camada recebe informações das etapas de extração de características e, por meio de operações não-lineares, gera valores de relação direta entre cada entrada associada a cada saída. A quantidade de saídas, em um modelo classificador, equivale ao número classes conhecidas pelo mesmo (KHAN et al., 2019).

Nas saídas das camadas FC, geralmente são inseridas funções de ativação específicas para a tarefa a ser realizada. Para a aplicação de classificação, uma função de ativação bastante presente é a *softmax*, que faz a normalização dos valores de saída da FC, de forma que eles representem a associação de determinada Região de Interesse (do inglês *Region of Interest* - RoI) com cada classe conhecida (YAMASHITA et al., 2018).

3.2 Redes Convolucionais Profundas Aplicadas à Detecção de Objetos em Imagens

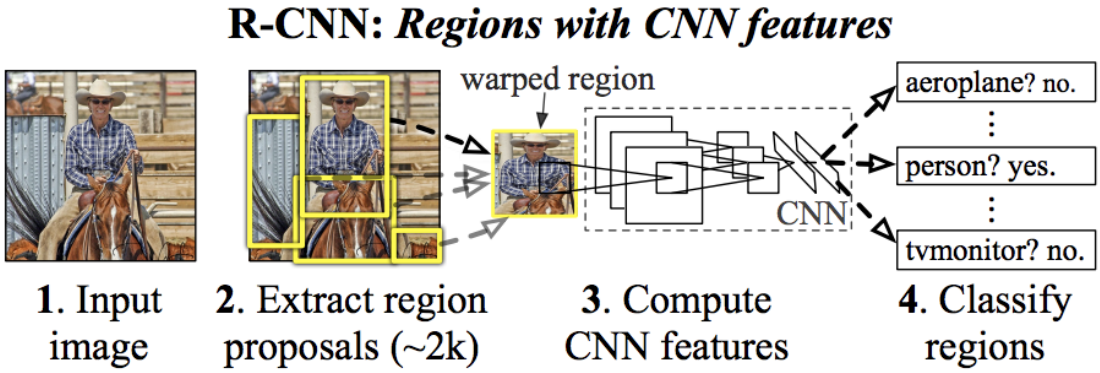
Um meta-algoritmo é um código que encapsula diferentes algoritmos para aumentar o desempenho da tarefa objetivada, tirando proveito de cada componente e mesclando suas execuções, de forma que sejam adaptadas automaticamente para utilizar o que melhor se aplicar a um determinado contexto ou conjunto de dados de entrada específico. Considerando um exemplo hipotético, se há diferentes métodos capazes de realizar a predição de valores de temperatura dada uma série histórica, mas cada um deles funciona melhor em uma faixa de temperatura diferente, é possível combinar diferentes algoritmos de forma a utilizar aquele que melhor se aplique a uma determinada faixa de temperatura de entrada do modelo em um instante específico.

Esta seção traz uma revisão técnica de métodos baseados em redes neurais profundas que, após pesquisas e evoluções, resultaram no método Mask R-CNN (HE et al., 2017), um meta-algoritmo dedicado não só à localização e classificação de objetos em imagens, mas também à segmentação de instâncias a nível de *pixel*.

3.2.1 R-CNN: *Region-based Convolutional Neural Network*

Projetado para detectar objetos em imagens, o método R-CNN utiliza a proposição de regiões de interesse ao invés da técnica de deslizamento de janelas, muito utilizada até então, como etapa preliminar para atingir a detecção (GIRSHICK et al., 2013). Após a apresentação de uma imagem de entrada, o algoritmo R-CNN é dividido em três módulos, conforme ilustra a Figura 10.

Figura 10 – Estrutura do Modelo R-CNN.



Fonte: (GIRSHICK et al., 2013).

Primeiro, um algoritmo de busca seletiva (UIJLINGS et al., 2013) extrai uma grande quantidade de regiões candidatas, que representam fronteiras de potenciais objetos.

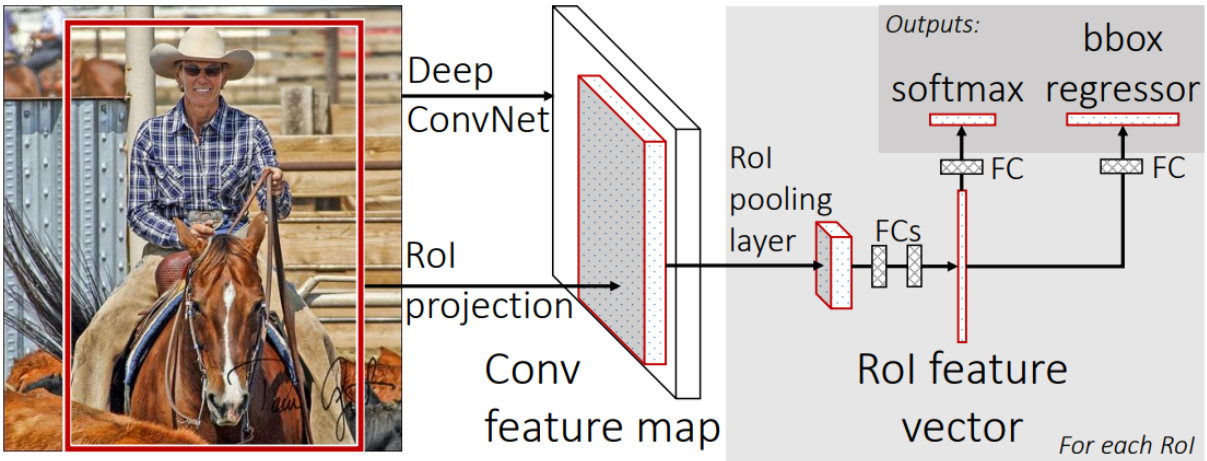
Geralmente são propostas 2000 regiões. Na segunda etapa, uma rede convolucional profunda realiza o mapeamento de características de cada uma das regiões propostas. Por fim, um supressor de não-máximos realiza o filtro de regiões repetidas e a classificação das regiões é feita por meio do classificador SVM (*Support Vector Machine*) (GIRSHICK et al., 2013).

Cada um dos três estágios é ajustado de forma independente, o que resulta em um sistema lento não só para a realização do treinamento dos modelos, mas também em testes de validação (GIRSHICK et al., 2013). Portanto, duas desvantagens deste modelo são relacionadas à sua lentidão. A primeira é o treinamento demorado, pois é realizado em mais de um estágio. A segunda é a execução lenta, pois para cada imagem de entrada são classificadas 2000 sub-imagens.

3.2.2 Fast R-CNN

Visando aumentar a velocidade de execução da rede apresentada na Seção 3.2.1, foi proposta por Girshick (2015) a arquitetura Fast R-CNN, cujo modelo é exibido na Figura 11. A nova arquitetura partiu de um princípio chave: compartilhar o processamento da camada convolucional com todas as RoI propostas, o que simplifica o método e, como principal consequência, reduz o tempo de processamento da rede. Além disso, na Fast R-CNN o resultado do processamento das camadas de saída da rede passa a ser reutilizado nas camadas da base do modelo (*backpropagation*).

Figura 11 – Estrutura do Modelo Fast R-CNN. A imagem é convertida em mapa de características e dele são propostas as regiões de interesse via *selective search*. Para cada RoI, camadas totalmente conectadas (FC) convertem o mapa em um vetor de características, que por sua vez é enviado para as camadas de saída: uma camada faz regressão linear para encontrar as 4 coordenadas dos *bounding boxes* e outra camada os classifica usando um operador *softmax*.



Fonte: (GIRSHICK, 2015).

No modelo Fast R-CNN, ao invés de várias RoI alimentarem uma rede neural, apenas a imagem original é alimentada à rede, que gera um mapa de características.

As coordenadas das RoI são propostas por um algoritmo de busca seletiva. O mapa de características é então usado para a avaliação das RoI, que são transformadas em imagens de tamanho fixo por meio do algoritmo batizado de *pooling* de região de interesse (*RoI pooling*).

O RoI *pooling* funciona a partir do mapa de características gerado pela rede convolucional da base do modelo. Para cada RoI proposta é feita a classificação utilizando uma camada *softmax*, que calcula a probabilidade do objeto pertencer a cada classe de treinamento. Além disso, é feita também a associação do rótulo ao *bounding box* da RoI avaliada (GIRSHICK, 2015).

Conforme o artigo original, considerando um mesmo *hardware*, os autores relatam que os tempos de processamento das imagens passaram da ordem de 50s com a R-CNN para algo em torno de 2s com o método Fast R-CNN, para cada imagem. (GIRSHICK, 2015). O ganho desse modelo foi fruto da aplicação de uma única rede convolucional, que gera o mapa de características geral a ser avaliado selecionando posteriormente as regiões de interesse, ao invés de definir primeiro as RoI para depois rodas 2000 vezes a rede convolucional (GIRSHICK, 2015).

3.2.3 FCN - Fully Convolutional Network

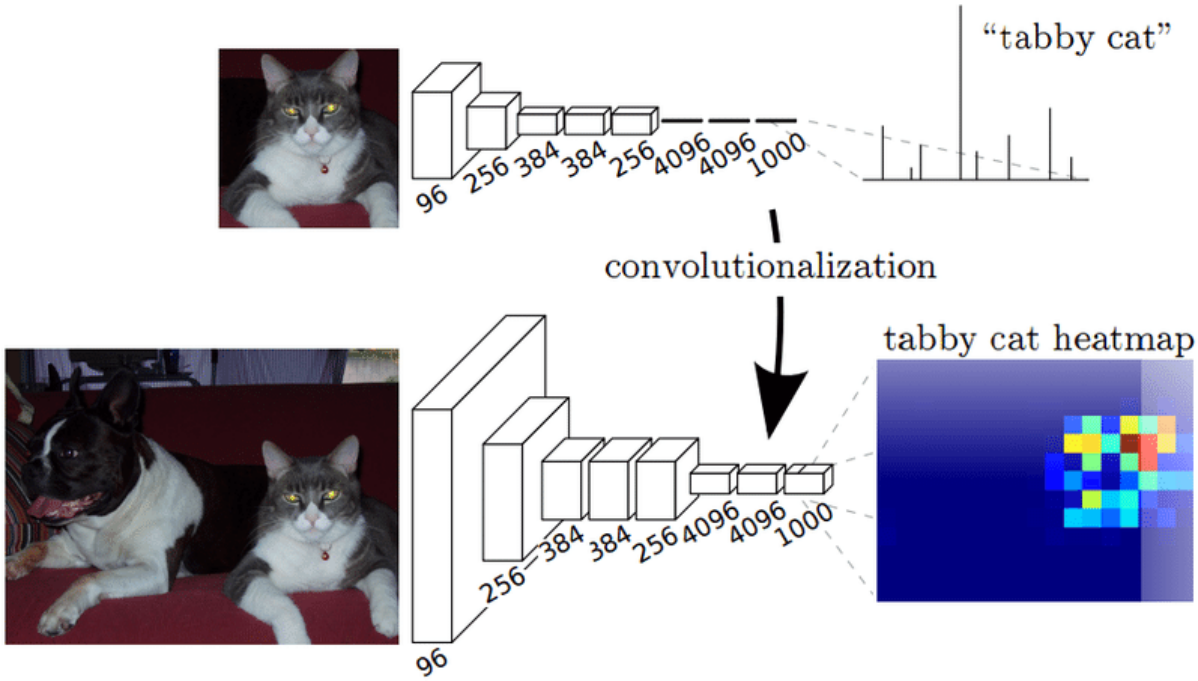
A rede totalmente convolucional, do inglês, *Fully Convolutional Network* (FCN), foi proposta por Long, Shelhamer e Darrell (2015) para a tarefa de segmentação semântica como uma continuidade de estudos promissores voltados para classificação de cenas (PINHEIRO; COLLOBERT, 2014) e estimativa de poses de humanos (TOMPSON et al., 2014). Os autores Long, Shelhamer e Darrell (2015) propuseram uma rede composta basicamente de camadas convolucionais, como o próprio nome FCN sugere.

Os componentes básicos das FCN são convolução, *pooling* e funções de ativação, que para operar sobre os dados de entrada dependem apenas da relação espacial das coordenadas.

Arquiteturas FCN funcionam no modelo *encoder-decoder*. Na etapa *encoder*, a rede realiza uma sequência de operações de convolução e *pooling*, criando mapas de características cada vez menores a cada passo dado. Essa etapa é destinada à classificação do objeto.

O resultado da última camada convolucional entra no fluxo *decoder*, onde são realizadas sequências de deconvolução (alguns autores preferem o termo convolução transposta) e *upsampling*, gerando no final do modelo um mapa com a mesma dimensão da imagem de entrada, que contém o *heatmap* a nível de *pixel* dos objetos, isto é, a probabilidade de cada pixel pertencer a cada classe conhecida (LONG; SHELHAMER; DARRELL, 2015). A Figura 12 ilustra de forma resumida o funcionamento da FCN.

Figura 12 – Estrutura do Modelo FCN.



Fonte: (LONG; SHELHAMER; DARRELL, 2015).

Uma vantagem das redes FCN é que elas trabalham com imagens de dimensão variável e produzem como saída um mapa de dimensões espaciais correspondentes às da imagem de entrada. A particularidade existente no caso da segmentação semântica é que o número de canais de saída corresponde à quantidade de classes do modelo (LONG; SHELHAMER; DARRELL, 2015).

Por exemplo, considerando imagens de entrada com dimensão $h \times w$ (sendo h a altura e w a largura), se o modelo da FCN for treinado para segmentar carros, bicicletas, pessoas e plano de fundo, totalizando quatro classes, a dimensão da imagem de saída será $h \times w \times 4$.

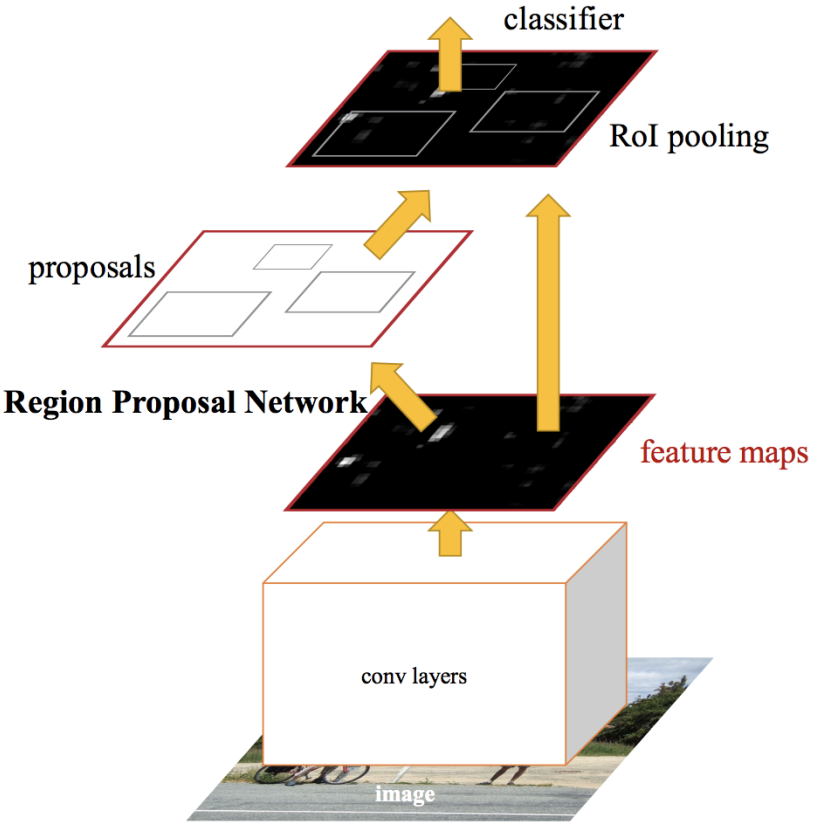
Para fins de exibição, os 4 canais passam por um operador *softmax*, resultando em uma imagem de canal único e com as 4 classes bem definidas, visualmente distintas por cores específicas para cada uma. Vale ressaltar que a classificação é realizada a nível de *pixel*, o que é outra vantagem desse tipo de modelo (LONG; SHELHAMER; DARRELL, 2015).

3.2.4 Faster R-CNN

A constante busca por melhorias levou Ren et al. (2017) à construção de uma solução para tornar a rede Fast R-CNN ainda mais rápida, resultando em um novo modelo chamado Faster R-CNN.

De forma similar ao que é realizado na Fast R-CNN, o primeiro passo desse algoritmo é gerar um único mapa de características com a imagem de entrada. Porém, nesse modelo o algoritmo de busca seletiva por RoI foi substituído por uma rede do tipo FCN (LONG; SHELHAMER; DARRELL, 2015), que faz a predição de regiões candidatas avaliando o mapa de características e foi batizada de RPN (*Region Proposal Network*) (REN et al., 2017).

Figura 13 – Estrutura do Modelo Faster-CNN com a *Region Proposal Network*.



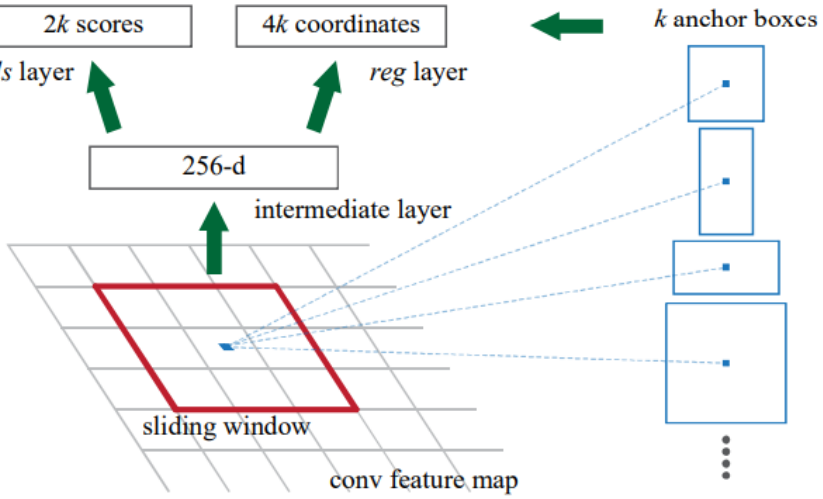
Fonte: (REN et al., 2017).

A Figura 13 mostra o modelo Faster R-CNN com a presença da RPN, cujo funcionamento consiste em uma rede completamente convolucional baseada em janela deslizante, que atua sobre o *feature map* obtido da imagem original, é invariante à translação (REN et al., 2017), e gera como resultado as proposições de coordenadas das RoI baseadas no mapa de características.

Após isso, é como se o modelo entregasse os dados para uma Fast R-CNN. Ou seja, é feito o RoI *pooling*: cada janela deslizante é mapeada em regiões de dimensões menores, denominadas âncoras (*anchors*), que podem ser visualizadas na Figura 14. As âncoras servem de entrada para duas camadas irmãs totalmente conectadas localizadas na saída da rede, uma dedicada à regressão do *bbox* (*bouding box)*, que refina na imagem original as coordenadas do *bbox* associado à âncora, e outra responsável pela *bbox-classification*

(REN et al., 2017).

Figura 14 – Funcionamento da *RoI pooling*, operação de janela deslizante que gera âncoras com informações obtidas da RPN e alimenta as camadas de classificação e regressão de coordenadas dos *bbox* dos objetos.



Fonte: (REN et al., 2017).

Para uma quantidade k de âncoras, são calculadas $4k$ coordenadas, que são as saídas da camada de regressão para os k *bounding boxes*. Também são determinadas $2k$ probabilidades para cada objeto das k regiões em relação a pertencerem à classe avaliada. Vale ressaltar que durante a etapa de busca por regiões, também é realizada a supressão de não-máximos para evitar o excesso de RoI candidatas (REN et al., 2017).

A RoI *pooling* atua com um objetivo importante, que é transformar as proposições em amostras (âncoras) de tamanho definido, passando-as para as camadas FCN de classificação e regressão da saída do modelo. Objetivo da RoI *pooling* é que o modelo funcione para imagens de objetos em diferentes escalas, sem perda de acurácia (REN et al., 2017).

De acordo com Ren et al. (2017), a Faster R-CNN pode ser executada em tempo real, pois durante as validações o tempo de avaliação de cada imagem levou em torno de 0,2s. Nos testes foi utilizada uma GPU com 12GB de memória e 2880 núcleos de processamento. Conforme explicado anteriormente, tal resultado é possível graças à substituição da busca seletiva na determinação das regiões de interesse por um método mais eficiente, a RPN, que reduz o número de iterações realizadas.

Resumindo de forma muito simplificada, a Faster R-CNN é uma Fast R-CNN que retirou a *selective search* e colocou uma RPN no lugar. Vale ressaltar que R-CNN, Fast R-CNN e Faster R-CNN trabalham apenas com a detecção de objetos, com regressão de *bounding box* e a classificação do objeto contido no mesmo. Tais redes não realizam segmentação a nível de pixel.

3.2.5 Mask R-CNN

Em abril de 2017 foi publicada a técnica Mask R-CNN (HE et al., 2017). Desenvolvida por um grupo de integrantes do FAIR (*Facebook Artificial Intelligence Research*), ela acrescentou relevantes avanços à Faster R-CNN (REN et al., 2017): levou a tarefa de segmentação de objetos para o nível de *pixel*, tratando inclusive casos de sobreposição de instâncias de uma mesma classe de objeto.

Considerando os avanços dos últimos anos, os métodos de segmentação semântica podem ser categorizados em duas famílias principais.

A primeira é impulsionada pela aplicação de arquiteturas do tipo R-CNN (GIRSHICK et al., 2013), onde a solução da tarefa parte de uma proposição de regiões de interesse a priori, que posteriormente são avaliadas por classificadores treinados para rotular essas RoI dentre diferentes categorias semânticas conhecidas pelo modelo.

A segunda é dirigida pelo uso de modelos do tipo FCN (LONG; SHELHAMER; DARRELL, 2015), que geralmente iniciam a tarefa utilizando o resultado de uma segmentação semântica de imagem completa, separando aquilo que pode ser objeto conhecido daquilo que estatisticamente é pouco provável que seja, e só depois são feitas as divisões desses resultados preliminares em instâncias individuais de objetos.

Conforme mostra a Figura 15, Mask R-CNN é a combinação de uma Faster R-CNN com uma FCN aplicada em cada RoI. Segundo He et al. (2017), a técnica foi desenvolvida para funcionar como um meta algoritmo, sob o ponto de vista do funcionamento independente que ocorre sobre cada região de interesse da imagem, tirando o melhor proveito de cada técnica interna e podendo ser estendida para algoritmos futuros (HE et al., 2017).

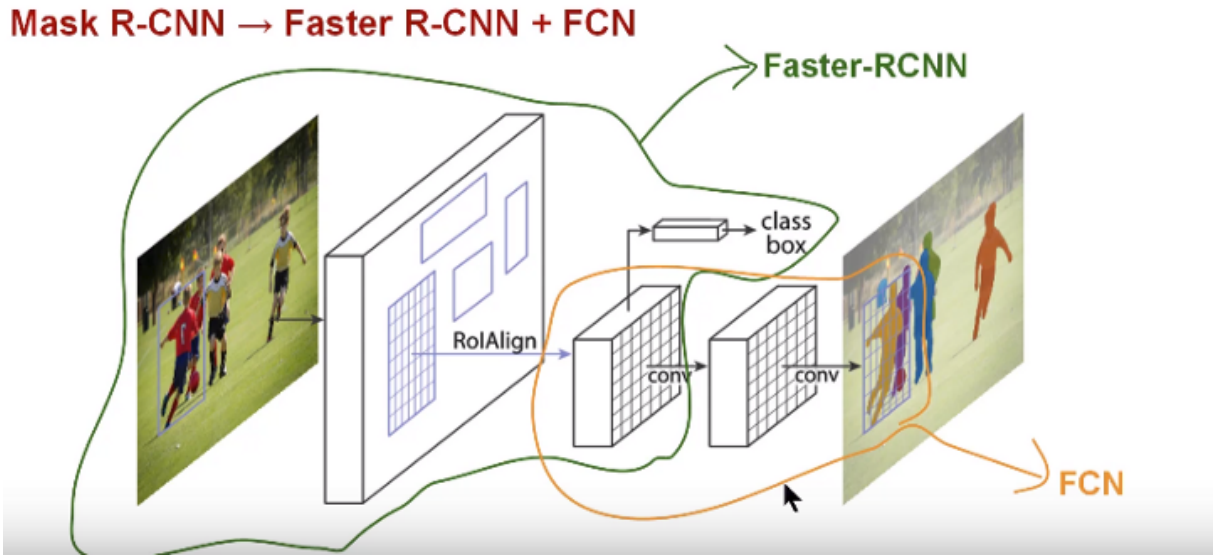
Dada uma imagem de entrada, a Mask R-CNN tem como saídas as classificações dos objetos identificados nela, seus *bounding boxes* e suas máscaras, que são a grande contribuição do método, pois representam a aproximação mais fiel possível encontrada para o objeto em nível de *pixel*.

Conforme ilustrado em detalhes na Figura 16, a Mask R-CNN é uma extensão da Faster R-CNN e adiciona uma nova ramificação paralela na saída do modelo, que é responsável por realizar a segmentação de instâncias e gerar como saída as suas máscaras, somando então essa funcionalidade à já presente regressão dos *bounding boxes* dos objetos, daí a escolha do termo *Mask* no nome do método (HE et al., 2017).

A Faster R-CNN funciona de forma similar à R-CNN, mas ao invés de tratar duas tarefas em passos subsequentes, realiza a separação de objetos e a regressão das *bounding boxes* de forma paralela, o que reduz significativamente o tempo de processamento do método.

A Figura 16 mostra o modelo completo da Mask R-CNN, e a Figura 17 dá ênfase

Figura 15 – Mask R-CNN: associação de modelo Faster R-CNN com camadas FCN na saída.



Fonte: (DWIVEDI, 2019).

nas camadas de saída da R-CNN, Faster R-CNN e o modelo em questão. Primeiro, está representado o modelo de processamento lento e sequencial da R-CNN (17a). Em seguida, na Faster R-CNN a introdução do paralelismo na saída do modelo para aumento de velocidade (17b) e, finalmente, a inclusão de uma terceira ramificação para a obtenção das máscaras na Mask R-CNN (17c).

Outra contribuição da Mask R-CNN foi a criação de um operador inédito, chamado *RoI align* (HE et al., 2017), que é uma evolução do operador *RoI pooling*.

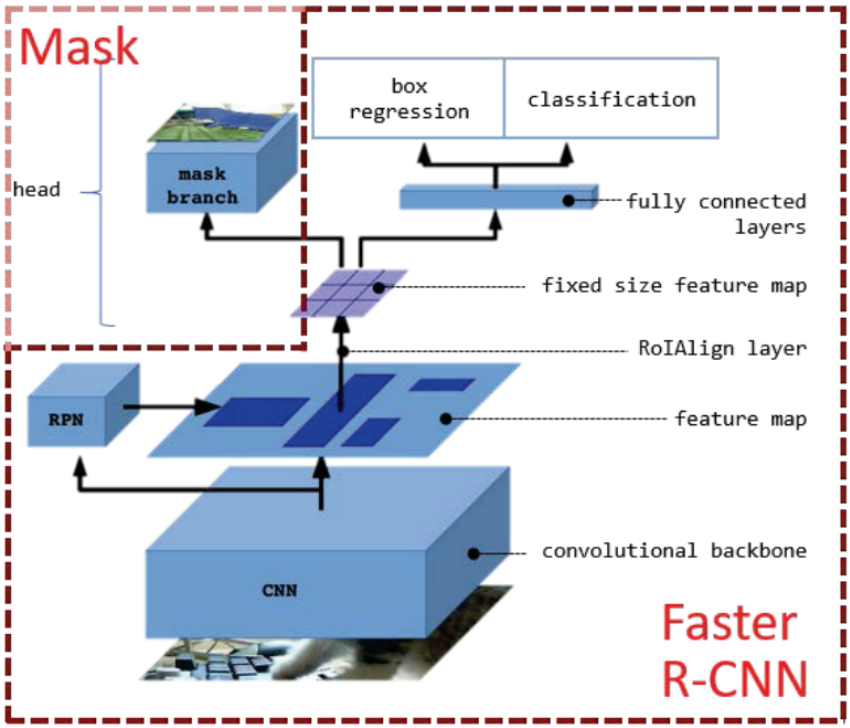
A Figura 18 exemplifica o *RoI Align* para um mapa de tamanho 5×5, em que quatro máscaras de tamanho 2×2 são aplicadas e, após a interpolação bilinear seguida de *max-pooling*, resulta em uma representação 2×2 para a RoI. O tamanho da máscara do *RoI Align* é um hiperparâmetro de treinamento do modelo.

O progresso consiste em não haver perda de informação no processamento das regiões de interesse. Isso é possível devido a não haver arredondamento no tratamento de coordenadas como no *RoI pooling*. Ao invés disso, após a conversão da RoI em mapa de características, é feita uma interpolação bilinear nas coordenadas desse mapa para gerar uma saída de dimensão fixa, porém sem que haja arredondamento das coordenadas dos *pixels* da região mapeada (HE et al., 2017), conforme ilustrada na Figura 19.

Na prática, a interpolação bilinear garante que as coordenadas sejam tratadas como ponto flutuante, trazendo uma condição de proporcionalidade à intensidade dos *pixels* da região interpolada, sem arredondamento e sem truncamento.

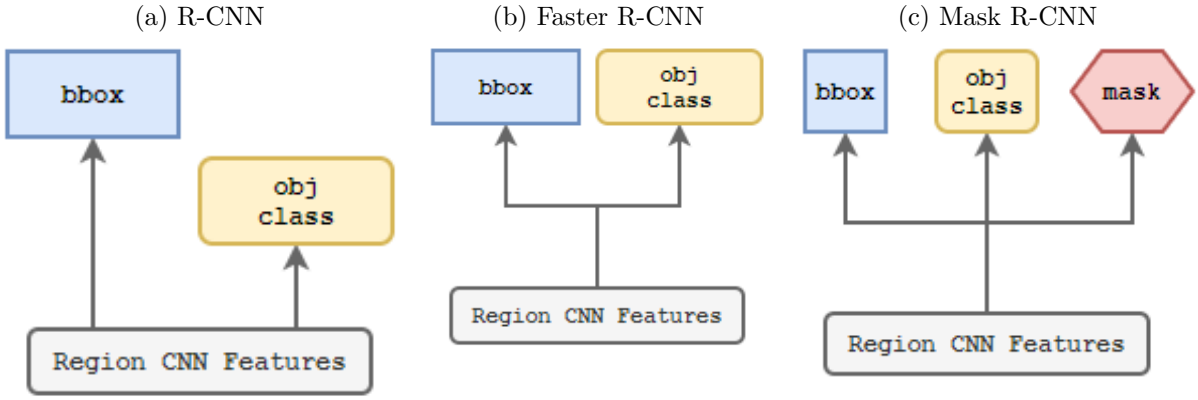
RoI pooling funciona bem para segmentação de objetos, quando é exigida apenas a

Figura 16 – Estrutura do Modelo Mask R-CNN.



Fonte: (BIENIAS et al., 2019).

Figura 17 – Arquiteturas simplificadas das camadas de saída dos modelos.

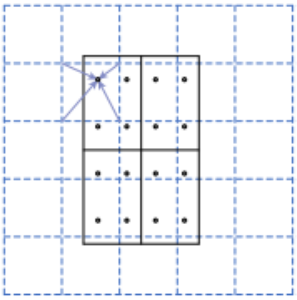


Fonte: o próprio autor.

predição do *bounding box* e o truncamento citado não reduz significativamente a qualidade do resultado, enquanto para o caso do *RoI align* há um refinamento no tratamento matemático, permitindo realizar a tarefa de encaixe da região em nível de *pixel*, abrindo caminho para aumentar a acurácia da predição da máscara do objeto (HE et al., 2017).

É possível utilizar dentro do modelo Mask R-CNN diferentes *backbones* para a rede convolucional que produz o *feature map* da imagem de entrada. Alguns bastante utilizados são baseados na arquitetura ResNet (*Residual Network*) (HE et al., 2016), que permite

Figura 18 – Funcionamento do *RoI Align*.



Fonte: (HE et al., 2016).

treinar redes com muitas camadas convolucionais, porém sem inserir complexidade a ponto de inviabilizar a execução da tarefa. Isso é feito por meio do uso de camadas que funcionam como blocos de aprendizado residual: elas não apenas repassam para a próxima camada o aprendizado obtido da camada atual, mas também o repassam para camadas 2 ou 3 níveis adiante, dependendo da arquitetura desejada. Um bloco residual simples é exibido na Figura 20.

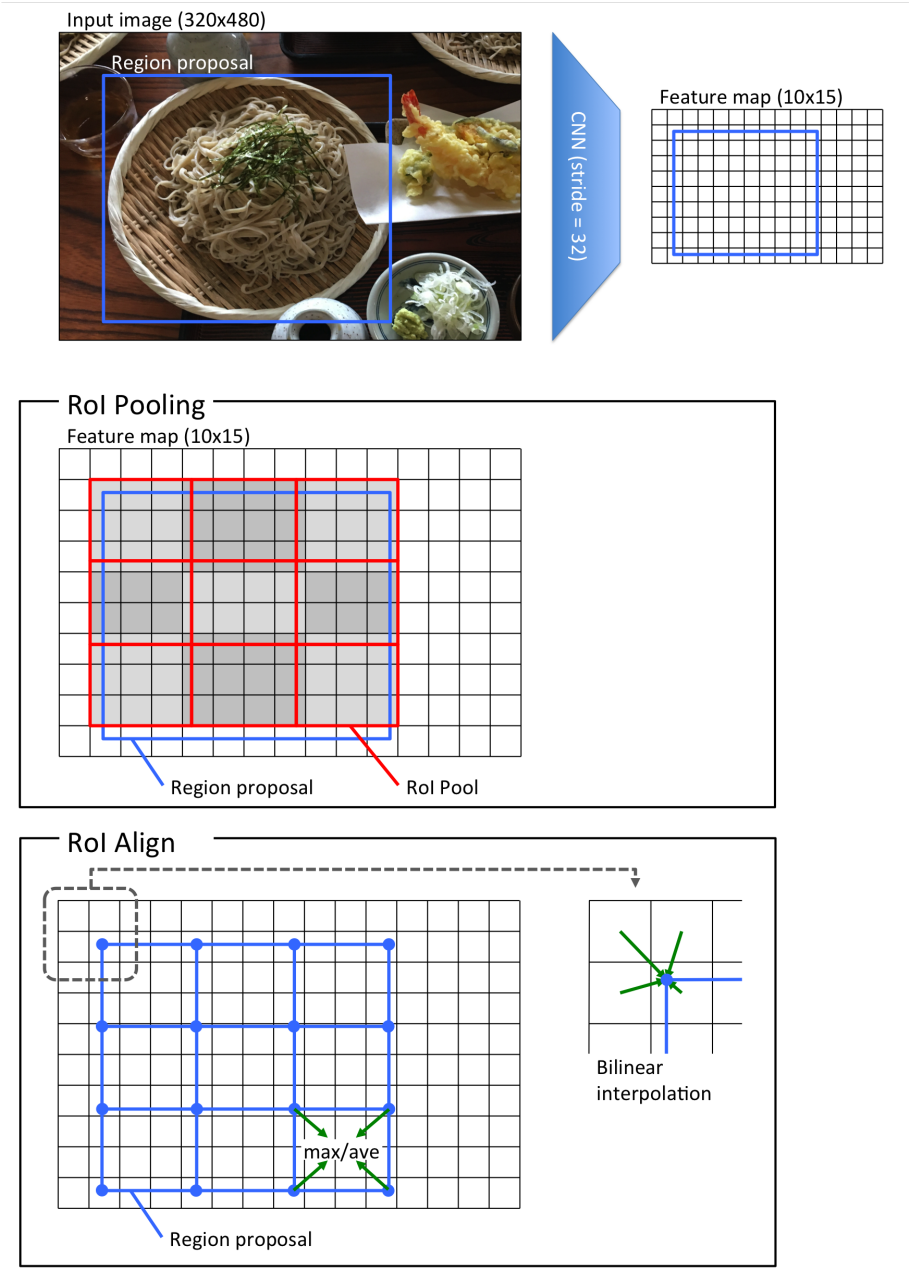
Os autores da Mask R-CNN disponibilizaram modelos pré-treinados ResNet-50 e ResNet-101, com 50 e 101 camadas convolucionais, respectivamente, que podem ser aplicados diretamente para a tarefa de classificação de objetos dentre 1000 classes pertencentes à base ImageNet, ou então podem ser adaptados por meio de *transfer learning* para a realização de outras tarefas.

3.3 *Transfer Learning*

A transferência de aprendizado, do inglês, *transfer learning*, é uma técnica de aprendizado de máquinas que consiste na reutilização de um modelo treinado para um problema de outro domínio. Esta técnica permite usar um modelo de rede neural treinado em um conjunto de dados e modificar apenas algumas camadas do modelo para resolver um outro problema semelhante, porém aproveitando parte dos pesos já treinados para solução do problema original (GOODFELLOW; BENGIO; COURVILLE, 2016). Quanto mais semelhante a tarefa original for do novo domínio, mais camadas pré-treinadas podem ser aproveitadas.

O treinamento de uma rede profunda utilizando conjuntos muito grandes de imagens é algo muito custoso computacionalmente. Dentre as vantagens do *transfer learning*, pode-se citar: o ganho de tempo em utilizar um modelo que foi pré-treinado em um conjunto de dados muito grande; a possibilidade de aplicar a rede em problemas que têm poucas amostras de treinamento. Muitas vezes, alterando apenas a última camada do modelo, por exemplo, já viabiliza ajustá-lo para a resolução de uma nova tarefa.

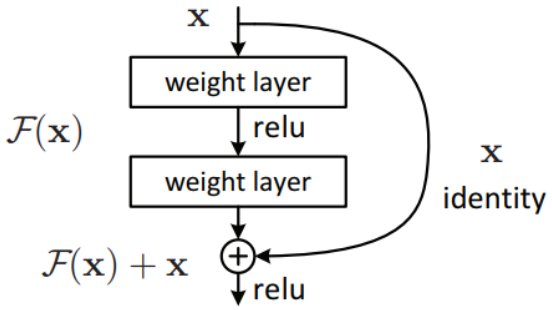
Figura 19 – Comparação do funcionamento das técnicas *RoI Pooling* e *RoI Align*.



Fonte: (UCHIDA, 2020)

Quando são utilizados modelos pré-treinados, a quantidade de parâmetros do novo aprendizado é reduzida, passando da ordem de centenas para dezenas de milhões, ou até centenas de milhares, dependendo da complexidade da estrutura da CNN (GOODFELLOW; BENGIO; COURVILLE, 2016).

Figura 20 – Bloco residual utilizado na arquitetura ResNet.



Fonte: (UCHIDA, 2020)

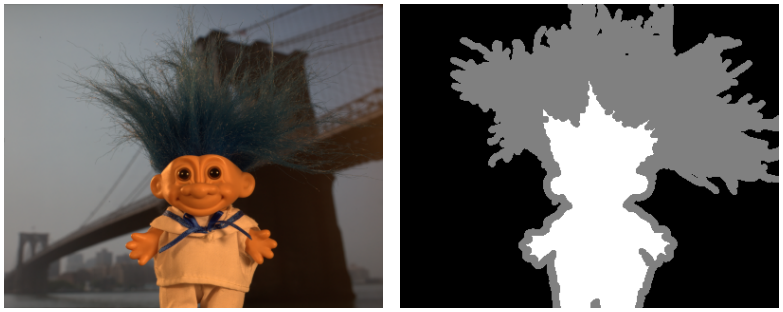
3.4 *Alpha Matting*

Alpha Matting é uma técnica que consiste em diferenciar os *pixels* do objeto no primeiro plano de uma imagem dos que estão no plano de fundo. Também pode ser definida como a tarefa de estimar o que é o objeto principal e o que não é, *pixel* a *pixel*. É comum que a técnica trabalhe a partir de duas informações de entrada: a imagem principal e um *trimap* (também conhecido como mapa de incerteza) (ZHENG; KAMBHAMETTU, 2009).

O *trimap*, cujo exemplo aparece na Figura 21b, é a imagem de entrada quantizada em apenas três níveis de cinza: 0, 128 e 255. O nível 0 é atribuído para todo *pixel* pertencente ao fundo, enquanto o 255 caracteriza o objeto principal em primeiro plano e, por fim, o nível 128 representa todo *pixel* que pertence a uma região de incerteza entre o objeto e o fundo.

Figura 21 – Imagem original e seu *trimap* à direita.

(a) Imagem original (b) *Trimap*



Fonte: (RHEMANN et al., 2009).

O modelo matemático comumente utilizado para o *Alpha Matting*, apresentado na Equação 3.2, define que cada *pixel* x de uma imagem I é a mistura entre o primeiro plano (*foreground*) I_F e o plano de fundo (*background*) I_B. A mistura, denominada *alpha matte*,

é quantificada por uma máscara *alpha*, em que $\alpha \in [0, 1]$.

$$I(x) = \alpha(x)I_F(x) + (1 - \alpha(x))I_B(x) \tag{3.2}$$

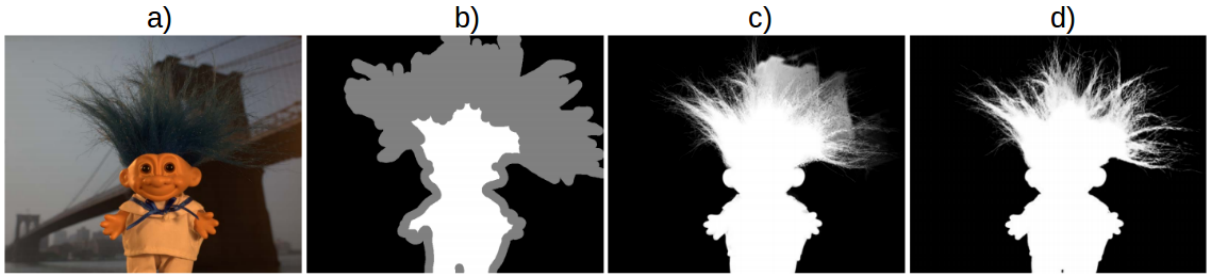
A tarefa-alvo dos algoritmos de *Alpha Matting* é recuperar I_F, I_B e α de I e há diversos métodos computacionais criados para realizá-la. Por exemplo, o *Closed-Form Matting* (LEVIN; LISCHINSKI; WEISS, 2008) considera que I_F e I_B são aproximadamente constantes dentro de pequenas janelas 3×3 ao redor de cada *pixel*, e pela sobreposição de janelas há a propagação de informação entre *pixels* vizinhos. O objetivo do método é encontrar o *alpha* que minimiza uma função custo pré-definida.

Utilizando CNN em dois estágios, o método *Deep Image Matting* (XU et al., 2017) utiliza primeiro uma rede para calcular de forma direta o *alpha matte* utilizando a imagem de entrada e o *trimap*. Ao invés de focar prioritariamente na informação de cor da imagem, a tarefa se baseia na estrutura natural presente em *alpha mattes* aprendidos pelo modelo.

Por exemplo, cabelo e pelos possuem estrutura de texturas muito dinâmica, bem como regiões de borrão em imagens capturadas de objetos em movimento. Estas características são complexas e usualmente menos diferenciadas por algoritmos menos sofisticados. No segundo estágio, há uma rede totalmente convolucional capaz de refinar as bordas utilizando como entrada a imagem original e o *alpha* estimado pela rede do primeiro estágio.

A Figura 22 ilustra o resultado obtido para uma imagem de exemplo com o *Closed-Form Matting* e o *Deep Image Matting*. As duas técnicas foram escolhidas para fazer parte dos testes desta pesquisa devido aos resultados de boa qualidade atingidos pelo segundo método e por este se comparar ao primeiro no trabalho original proposto em Xu et al. (2017).

Figura 22 – Imagem de entrada (a), seu *trimap* (b) e o resultado da aplicação das técnicas *Closed-Form Matting* (c) e *Deep Image Matting* (d).



Fonte: (XU et al., 2017)

4 Metodologia

Neste capítulo é descrito o método proposto para a realização do trabalho. São apresentadas algumas características das bases de dados utilizadas, além de todas as etapas da estratégia montada para atingir os objetivos específicos. Posteriormente, são apresentadas as métricas selecionadas para avaliar os resultados da pequisa.

4.1 Bases de Dados

Conforme citado no Capítulo 1, o intuito deste trabalho é dar continuidade à pesquisa iniciada por Gonçalves (2017), e consiste em extrair informações destinadas a contribuir para a automatização da estimativa de somatotipo. Assim, o principal foco deste trabalho é obter o contorno corporal de pessoas em imagens obtidas em ambientes não controlados. Além disso, faz parte dos objetivos ampliar o estudo no sentido de fazer a identificação automática de gênero e a estimativa de massa corporal dos indivíduos.

Devido à diversidade de tarefas a serem realizadas, houve dificuldade para encontrar uma base de dados rotulada que continha todas as informações necessárias e com um número adequado de amostras. Portanto, ao invés de usar uma única base de dados de imagens, foi selecionado um conjunto de bases de dados, cada uma com suas características próprias e com um uso específico dentro do tema deste trabalho. Algumas bases possuem imagens do corpo inteiro do indivíduo, outras apenas as faces. Ao todo, foram utilizadas quatro bases de dados.

A Tabela 1 mostra a distribuição[1] das imagens de acordo com o gênero para cada base, considerando apenas as imagens separadas para este trabalho. As próximas subseções detalham as informações acerca de cada uma das bases.

Tabela 1 – Distribuição de imagens por base de dados e por gênero.

Base	Masc	Fem	Fundo vazio	Total
Fisio-Somatotipo	78	76	0	154
Wiki	200	200	100	500
Face-to-BMI	2438	1768	0	4206
Internet	14	14	0	28

Em respeito à privacidade dos indivíduos que aparecem nas imagens contidas nesta dissertação, as faces dos mesmos receberam uma tarja. Entretanto, vale ressaltar que todos

[1] Imagens de fundo vazio foram utilizadas apenas na etapa de treinamento dos modelos de Classificação de Gênero, uma vez que na etapa de validação a imagem classificada já é de um indivíduo previamente identificado na etapa anterior de segmentação corporal.

os testes e validações foram realizados com as imagens originais.

4.1.1 Base de dados Fisio-Somatotipo

A primeira base de dados utilizada, aqui denominada Fisio-Somatotipo, foi formada por Gonçalves (2017) e é composta por imagens de corpo inteiro, em posição frontal e lateral, de atletas de fisiculturismo da IFBB-ES (*International Federation of Bodybuilding* - ES), Federação de Fisiculturismo do Estado do Espírito Santo. A autorização de uso concedida pela IFBB-ES está disponível no Anexo B. Além das fotos, há informações como altura e massa corporal dos atletas, dentre outras medidas corporais.

Esta base está dividida entre dois conjuntos: o primeiro foi obtido em um ambiente com pouco controle de iluminação, com presença de sombras e um fundo não homogêneo, além de haver objetos no fundo que não são de interesse; o segundo conjunto de imagens, além de ter sido obtido em um ambiente com maior controle de iluminação, também foi elaborado com uma padronização maior das poses dos atletas.

A utilização dessa base de imagens abrange todo o escopo de validação deste trabalho: para imagens frontais, cabe a aplicação dos algoritmos de segmentação de pessoas para obtenção dos contornos corporais, bem como a validação da classificação de gênero e, por fim, a estimativa de massa corporal; para imagens laterais, apenas a aplicação da segmentação. A Figura 23 mostra algumas imagens da base Fisio-Somatotipo.

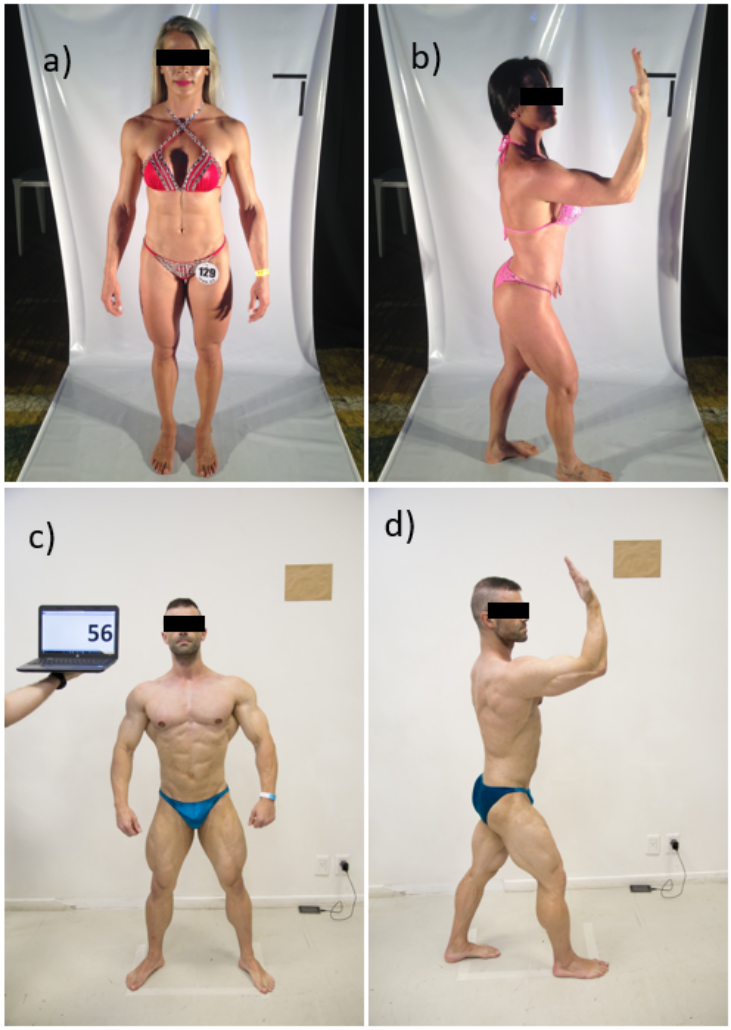
4.1.2 Base de dados Wiki

A segunda fonte para formar a base de dados deste trabalho foi a Wiki[2] (ROTHE; TIMOFTE; GOOL, 2015), que disponibiliza imagens de forma gratuita na *internet*. Esta base possui mais de 62 mil fotos de faces masculinas e femininas, tendo sido originalmente construída para a realização de estudos de estimativa de gênero e idade, informações estas que estão disponíveis na base junto com as fotos dos indivíduos.

Estas imagens possuem diversidade de iluminação de fundo, sombras e posicionamento das faces. Há também imagens utilizadas para a formação de um cenário de fundo vazio, isto é, imagens com ausência de pessoas nas fotos, onde não deve haver detecção de faces. O objetivo desta base de dados para este trabalho é explorar as técnicas de classificação de gênero com base nos rostos. A Figura 24 mostra algumas imagens da base Wiki.

[2] Wiki *dataset*, disponível em: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

Figura 23 – Exemplos dos dois conjuntos de imagens da base Fisio-Somatotipo. Ambiente com pouco controle de iluminação: (a) e (b). Ambiente com controle maior: (c) e (d).



Fonte: (GONÇALVES, 2017).

4.1.3 Base de dados Face-to-BMI

A terceira base de dados para esta pesquisa é a base Face-to-BMI[3] (KOCABEY et al., 2017), que possui 4206 imagens de faces humanas, além de anotações para cada imagem com o gênero e o índice de massa corporal (do inglês *body mass index* - BMI) dos indivíduos.

As fotografias da base Face-to-BMI contam com variedade de posicionamento dos rostos, enquadramento de câmera e variação de iluminação. O foco dessa fonte de imagens para este trabalho é o treinamento e a aplicação de métodos para estimativa de massa corporal. A Figura 25 exibe imagens da base Face-to-BMI e os respectivos valores reais de IMC dos indivíduos.

[3] Face-to-BMI *dataset*, disponível em: http://face2bmi.csail.mit.edu/

Figura 24 – Imagens da base Wiki: (a) classe masculina; (b) classe feminina; (c) fundo vazio.

Fonte: (ROTHE; TIMOFTE; GOOL, 2015)

4.1.4 Base de dados Internet

Por fim, algumas imagens para validação dos modelos de segmentação e classificação de gênero foram obtidas em diversos sites da *internet*, tendo suas fontes descritas neste trabalho de forma específica junto às figuras em que aparecem. São imagens cujo objetivo é explorar o potencial das técnicas aqui utilizadas nas aplicações em ambientes adversos, com diferentes características de fundos, padrões de iluminação e presença de sombras.

Algumas das principais características das imagens selecionadas são: falta de uniformidade da incidência de luzes, bem como fundos com diferentes texturas, cores e formas; diferentes variações de escala, ou proximidade entre os corpos e as câmeras ou enquadramentos das fotografias, e também o posicionamento dos corpos dos indivíduos, buscando variar e obter amplitude de cenas de teste.

Foram escolhidas imagens em que o indivíduo está posicionado com braços afastados do corpo, em poses semelhantes à pose frontal utilizada nos exames de somatotipo. Esta base foi denominada Internet e as Figuras 26 e 27 mostram algumas de suas amostras.

Figura 25 – Imagens da base Face-to-BMI e os respectivos valores de IMC de cada indivíduo.

Fonte: (KOCABEY et al., 2017)

Esta base será usada somente para validação das metodologias de segmentação de contorno corporal e classificação de gênero, uma vez que não são conhecidas as informações de massa corporal dos indivíduos.

Figura 26 – Imagens da base Internet para a classe masculina.

Fonte: (PROLEAGUENEWS, 2020; BRASIL, 2020)

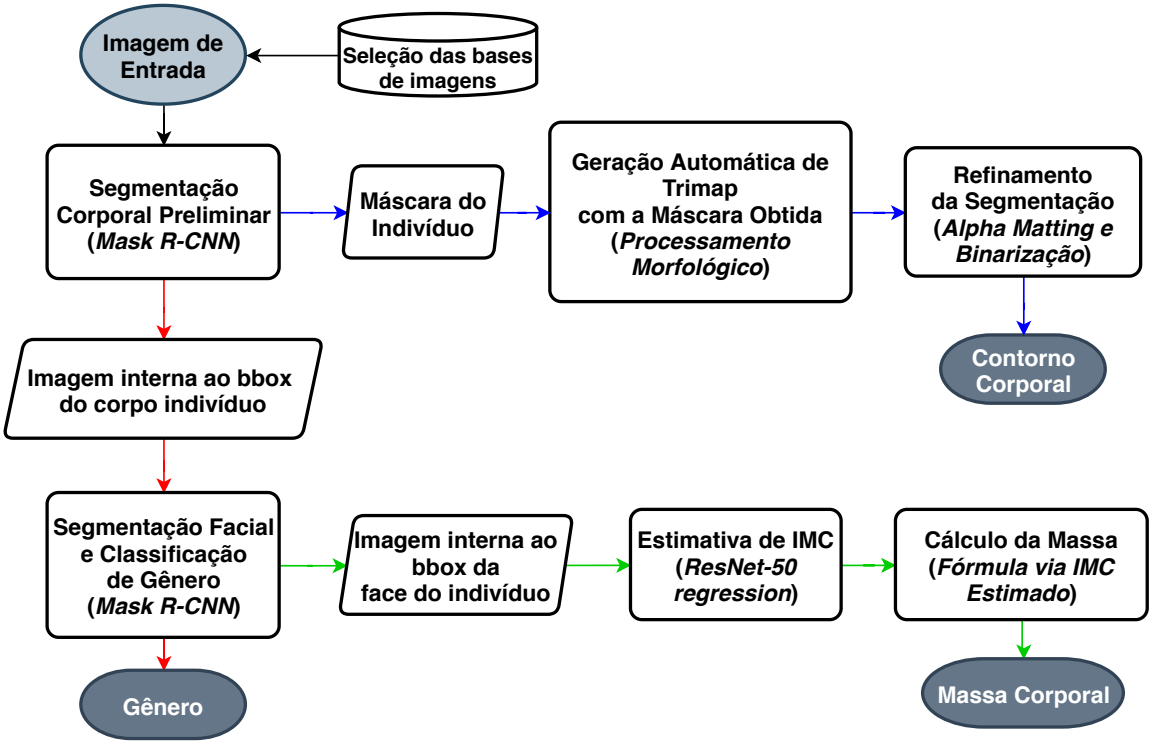
Figura 27 – Imagens da base Internet para a classe feminina.

Fonte: (CELEBZZ, 2020)

4.2 Estratégia de Execução do Estudo

A Figura 28 mostra um diagrama com o fluxo da metodologia proposta neste trabalho e as técnicas envolvidas nas principais etapas.

Figura 28 – Fluxo da metodologia proposta neste trabalho.



Fonte: o próprio autor.

Conforme ilustrado na Figura 28, uma vez definidas as bases de dados do estudo, foi estabelecida uma sequência de etapas para atingir os objetivos específicos deste trabalho. Em resumo, as tarefas ocorrem conforme segue:

1. Segmentação dos corpos dos indivíduos para obtenção do contorno corporal completo;
2. Segmentação do *bounding box* da face e definição da máscara contendo o contorno facial;
3. Classificação de gênero a partir da máscara da face;
4. Estimativa de IMC e massa corporal analisando a imagem de dentro do *bounding box* da face.

O objetivo principal é fornecer como saída do estudo três informações para cada imagem: o contorno mais detalhado possível do corpo da pessoa, a classificação do gênero em masculino ou feminino e, por fim, a estimativa de IMC e massa corporal.

Gonçalves (2017), ao estimar o somatotipo de uma pessoa, realizou também a estimativa da altura do indivíduo por meio das imagens, porém, isso demandou ajustes fixos de distância de captura da imagem não apenas entre a câmera e o indivíduo, mas também entre o indivíduo e o fundo. Aqui, pensando em uma aplicação prática futura, foi considerado que a altura é previamente conhecida, pois tem pouca variação ao longo da vida adulta e é comum que as pessoas saibam essa informação.

Neste trabalho, o desafio principal é obter a segmentação dos indivíduos em ambientes não controlados. Isto se deve ao fato de que esta tarefa deve ser realizada de forma a obter os contornos dos corpos dos indivíduos com o maior nível de detalhe possível, sobretudo em bordas e curvas.

As seções a seguir detalham a proposta de como realizar cada um dos objetivos propostos neste trabalho.

4.3 Segmentação dos Indivíduos Utilizando Modelo Mask R-CNN Pré-Treinado

A principal técnica aplicada para a segmentação de pessoas neste trabalho foi a Mask R-CNN. Funcionando como o principal pilar, sua aplicação foi escolhida por ter sido considerada como o estado da arte em relação à tarefa de segmentação semântica de instâncias na época em que este trabalho foi iniciado (HE et al., 2017). Outro motivo de sua escolha é que ela permite obter o contorno do objeto identificado em nível de pixel ao mesmo tempo em que é realizada a classificação do mesmo. Sua aplicação visa obter três informações em dois casos de aplicação complementares:

1. Obter a máscara relativa ao contorno corporal das pessoas;

2. Definir o *bounding box* do contorno corporal, posteriormente utilizado para busca da face na tarefa de classificação de gênero e estimativa de massa corporal, como explicado mais adiante, nas Seções 4.5.2 e 4.6.

Como modelos de redes convolucionais da Mask R-CNN, foram utilizados os *backbones* ResNet-50 e ResNet-101. Para a detecção e segmentação de pessoas foi aplicado primeiramente um modelo ResNet-50 e, posteriormente para efeito de comparação, um modelo ResNet-101. Ambos já possuíam pesos pré-treinados na base COCO (*Common Objects in Context* (LIN et al., 2014)) para a tarefa de segmentação de instâncias. Entretanto, foi realizada a restrição na saída do modelo para considerar apenas humanos como classe de interesse a ser segmentada nas imagens.

A Figura 29 mostra alguns exemplos de máscaras resultantes da segmentação corporal com o método Mask R-CNN utilizando backbones ResNet-50 e ResNet-101 para indivíduos da base Fisio-Somatotipo.

Posteriormente, com o objetivo de refinar os resultados obtido nos contornos corporais, foi decidido testar a técnica *Alpha Matting*, porém de forma automatizada.

4.4 Refinamento de Contornos Corporais com *Alpha Matting*

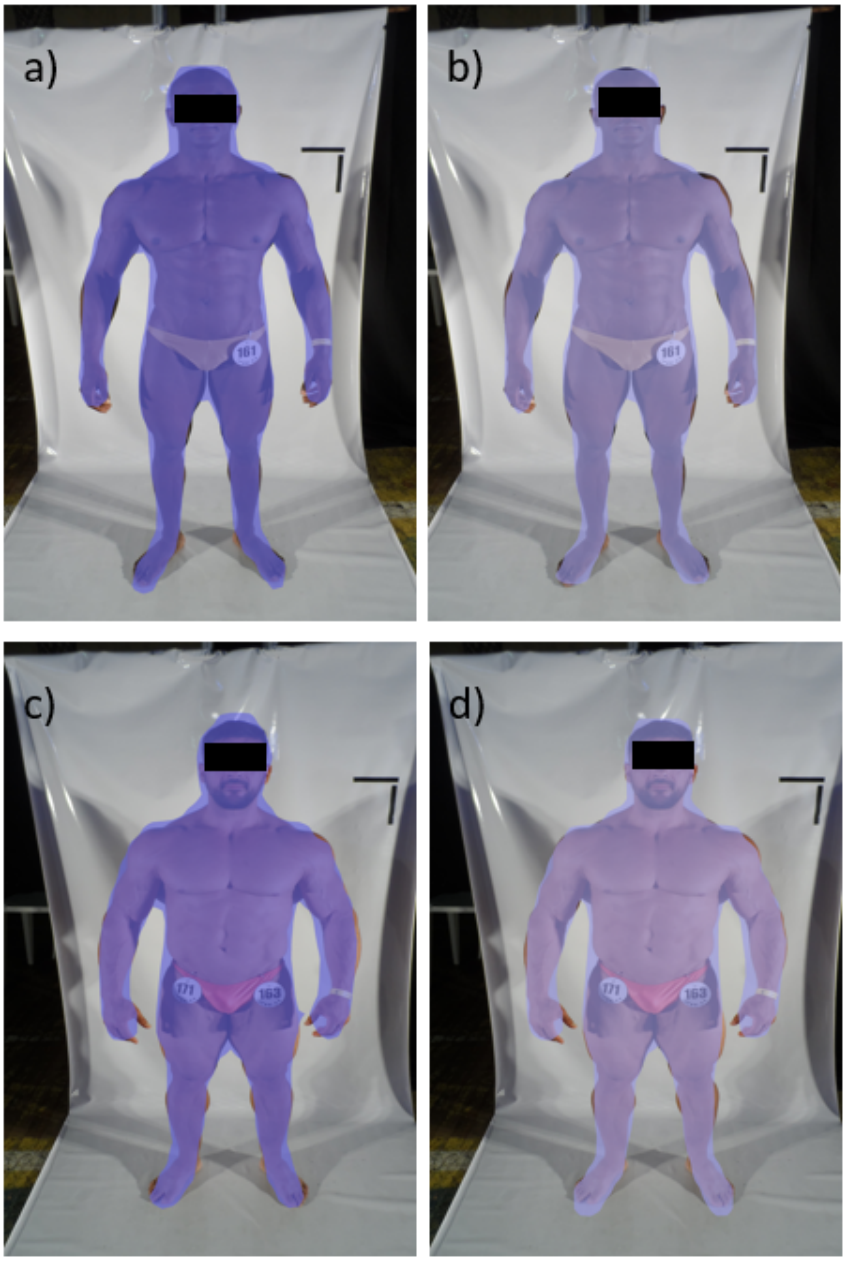
Após a segmentação utilizando a Mask R-CNN, duas implementações da técnica *Alpha Matting* foram aplicadas para tornar o resultado da segmentação do contorno corporal mais assertivo. A primeira implementação, baseada em CNN, foi a *Deep Image Matting* (XU et al., 2017). A segunda foi a *Closed-Form Matting* (LEVIN; LISCHINSKI; WEISS, 2008).

Como ambas implementações dependem que seja informado como entrada do algoritmo tanto a imagem original quanto o seu *trimap* com a região de incerteza, surgiu a ideia de buscar uma forma de gerar o *trimap* de forma automática. A maneira com que isso foi automatizado consiste em utilizar a Mask R-CNN não mais como um gerador de imagem final segmentada, mas como um gerador de *trimaps*.

Para isso, seja I_{mask} a imagem resultante da segmentação do contorno corporal com a Mask R-CNN, porém binarizada. Utilizando um *kernel* de tamanho 15×15 são geradas duas imagens auxiliares: I_{dil}, resultante da dilatação morfológica (GONZALEZ; WOODS, 2006) de I_{mask}, e I_{ero} resultante da erosão de I_{mask}.

Uma terceira imagem I_{borda} é obtida realizando a subtração entre a imagem dilatada I_{dil} e a imagem erodida I_{ero}, conforme a Equação 4.1. I_{borda} é utilizada para definir a região de incerteza do *trimap*. A quantidade de iterações realizadas de dilatação e de

Figura 29 – Segmentação dos corpos de indivíduos com o modelo Mask R-CNN: (a) e (c) com *backbones* ResNet-50, em azul escuro; (b) e (d) com ResNet-101 em azul claro.



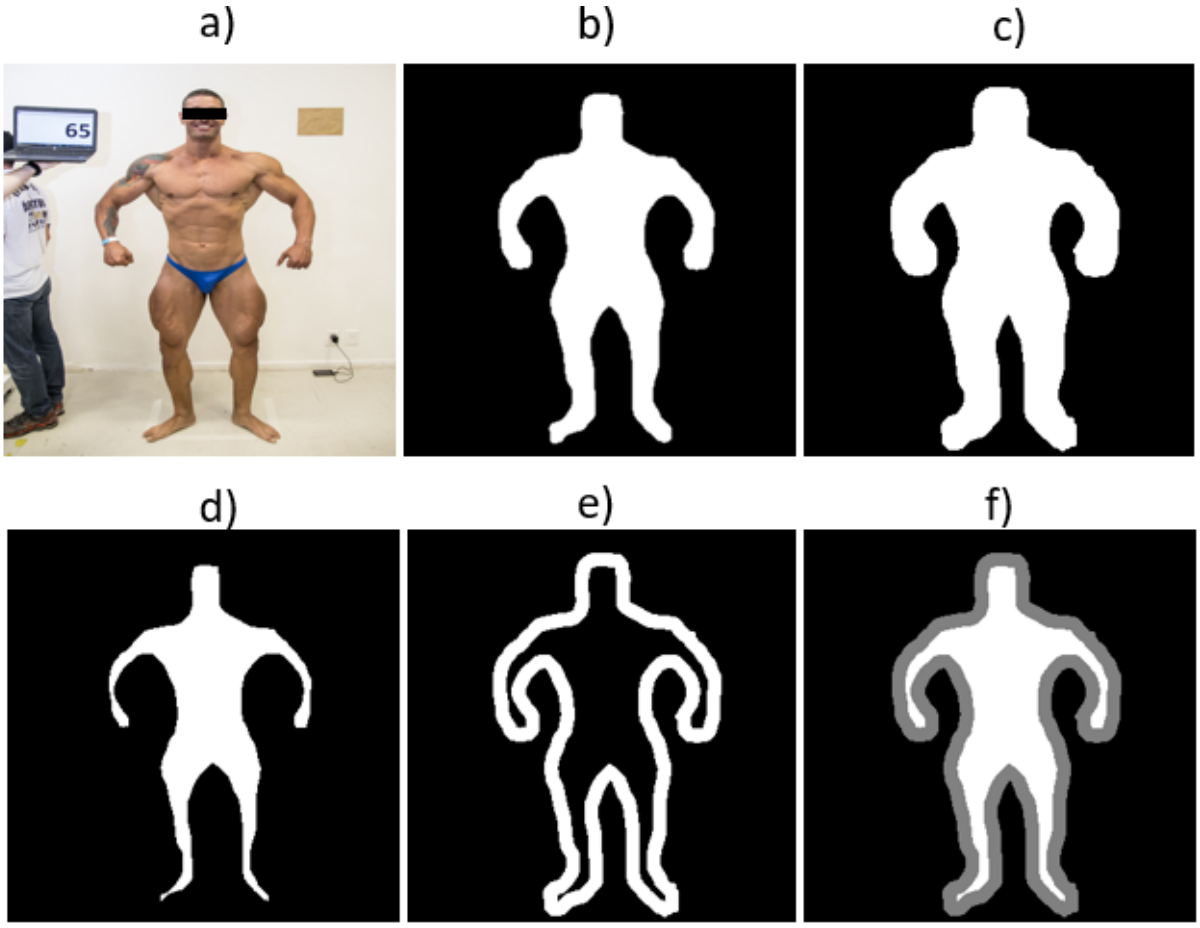
Fonte: o próprio autor.

erosão definem a espessura de I_{borda}. Finalmente, o *trimap* I_{trimap} em escala de cinza é obtido realizando a operação da Equação 4.2. A Figura 30 ilustra as etapas da geração automática do *trimap*. As operações de erosão e dilatação foram implementadas utilizando a biblioteca OpenCV do Python.

$$I_{borda} = I_{dil} - I_{ero} \tag{4.1}$$

$$I_{trimap} = 128 * I_{borda} + 255 * I_{ero} \tag{4.2}$$

Figura 30 – Geração automática do *trimap*. A imagem original (a), após segmentação via Mask R-CNN é binarizada (b). A dilatação de (b) resulta em (c). A erosão de (b) resulta em (d). A subtração entre (c) e (d) resulta em (e). O *trimap* resultante é a figura (f), em escala de cinza.



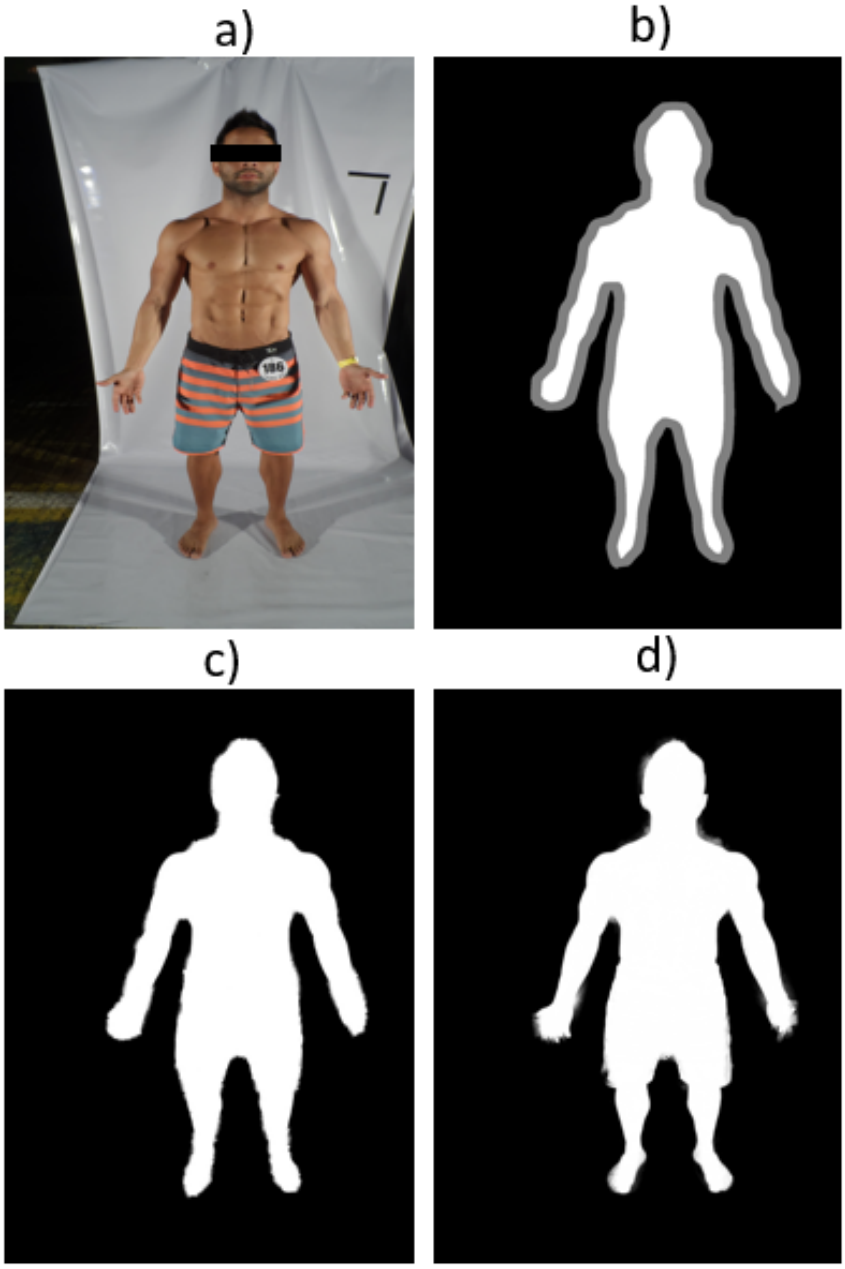
Fonte: o próprio autor.

Uma vez alcançada a geração automática dos *trimaps*, o próximo passo foi testá-los na aplicação das técnicas de *Alpha Matting*. A Figura 31 ilustra um exemplo obtido com tal processamento, tanto para o *Deep Image Matting*, quanto para o *Closed-Form Matting*, ambos descritos na Seção 3.4.

Para ampliar as possibilidades de testes, foram consideradas quatro espessuras distintas para compor a região de incerteza dos *trimaps*. A Figura 32 ilustra um exemplo de variação das bordas. A partir daí, os *trimaps* foram nomeados conforme a quantidade de iterações realizadas de dilatação e erosão. Um *trimap* fruto de duas dilatações e duas erosões recebeu o nome de Trimap-2, e assim sucessivamente.

Os resultados obtidos variando a espessura da região de incerteza dos *trimaps* estão detalhados no Capítulo 5.

Figura 31 – Resultado da aplicação das técnicas de *Alpha Matting*. Em (a) a imagem original e em (b) o seu *trimap*. Os resultados da *Deep Image Matting* e *Closed-Form Matting* são mostrados em (c) e (d), respectivamente.

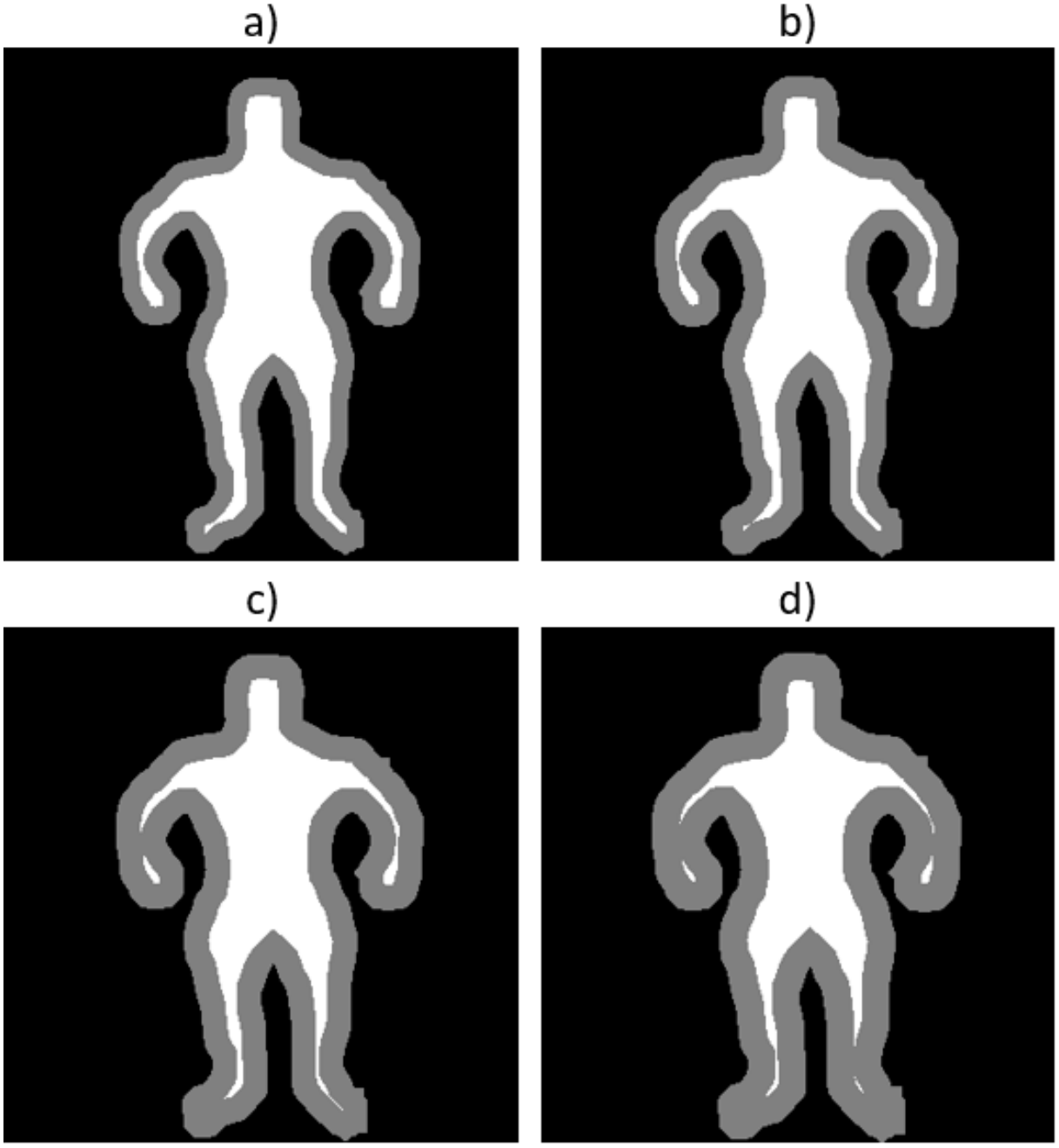


Fonte: o próprio autor.

Portanto, o objetivo buscado com a combinação dos algoritmos aqui utilizados é que, primeiramente, o Mask R-CNN encontre a localização aproximada do contorno da pessoa e que, no segundo momento, o *Alpha Matting* refine a segmentação.

Finalizadas as etapas citadas nesta seção, é alcançado o primeiro objetivo específico do trabalho: obter o contorno corporal do indivíduo com detalhes a nível de *pixel* em bordas e curvas.

Figura 32 – Variação de espessura da borda de incerteza para: (a) Trimap-2; (b) Trimap-4; (c) Trimap-6; (d) Trimap-8.



Fonte: o próprio autor.

A seguir é detalhado o procedimento para identificar o gênero da pessoa.

4.5 Classificação de Gênero Utilizando Modelo Mask R-CNN Adaptado

O próximo passo é classificar o gênero da pessoa. Esta tarefa pode ser realizada buscando a face no interior do *bounding box* que contém o indivíduo, o que economiza processamento, pois evita tratar a imagem original completa.

Para a classificação de gênero, foi decidido aplicar a Mask R-CNN com *backbone* ResNet-50, processando apenas a imagem restrita ao *bounding box* da pessoa. A ideia aplicada partiu de aproveitar o modelo pronto que serviu de base para a Seção 4.3, já com os pesos pré-treinados. Entretanto, para efeito de testes comparativos, foram realizadas quatro abordagens distintas: nas duas primeiras, foi feito o treinamento completo de todas as camadas da rede; nas outras duas propostas foram aplicados modelos pré-treinados e realizado apenas o ajuste fino da camada no topo da rede por meio de *transfer learning*. Os testes contemplaram apenas um modelo de *backbone* devido ao elevado custo computacional envolvido nas tarefas de treinamento.

Antes de explorar detalhadamente as estratégias adotadas, é necessário falar um pouco sobre a forma como é realizado um treinamento de modelo Mask R-CNN e o esforço manual envolvido.

4.5.1 Preparação de Dados Para o Treinamento de um Modelo Mask R-CNN

Na fase de treinamento de um modelo do método Mask R-CNN, é necessário apresentar para cada instância de objeto contido em uma imagem de treinamento quatro informações básicas: a imagem original que possui o objeto; um vetor com as coordenadas dos quatro vértices do *bounding box* do objeto; o rótulo da classe do objeto; um arquivo de imagem contendo apenas o desenho da máscara do objeto.

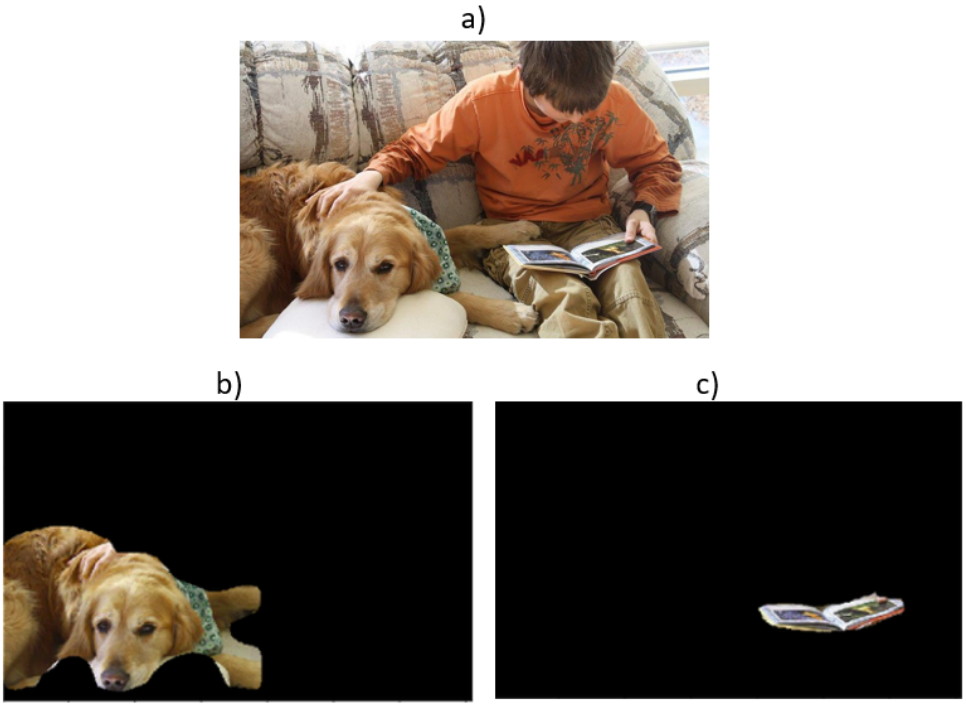
A Figura 33 ilustra uma imagem de treinamento e duas máscaras para classes distintas.

Há ferramentas que auxiliam a tarefa de anotação manual de imagens, por exemplo a LabelMe[4] (RUSSELL et al., 2007), que foi utilizada para a criação das máscaras de treinamento para classificação de gênero. A Figura 34 mostra a interface de uso da ferramenta, que é disponível de forma gratuita na *internet*, bastando estar cadastrado na plataforma. Uma vez finalizada a utilização, a ferramenta LabelMe disponibiliza para *download* arquivos texto com as coordenadas dos *bounding boxes* das máscaras e o rótulo do objeto, além dos arquivos das máscaras.

As máscaras são disponibilizadas na forma de imagens binárias em arquivos com extensão PNG (*Portable Network Graphic*) e podem ser modificadas manualmente em

[4] LabelMe, *the open annotation tool*, disponível em: http://labelme2.csail.mit.edu/

Figura 33 – Imagem de treinamento (a) para o modelo Mask R-CNN e máscaras para aprendizado da classe “cão” (b) e classe “livro” (c).



Fonte: adaptado de (SHARMA, 2020).

qualquer editor de imagens compatível com o formato, caso seja necessário algum ajuste posterior fora da ferramenta LabelMe. O motivo da utilização deste recurso é detalhado na Seção 4.5.2.

4.5.2 Estratégias de Aplicação da Mask R-CNN Para a Classificação de Gênero

As bases Wiki e Fisio-Somatotipo foram utilizadas como conjunto de dados para a classificação de gênero, totalizando 554 imagens com indivíduos, sendo:

1. 400 imagens da base Wiki: 200 homens e 200 mulheres;
2. 154 imagens da base Fisio-Somatotipo: 78 homens e 76 mulheres;

Das 554 imagens, 330 imagens (60%) foram separadas para treinamento e 224 (40%) para a validação do método.

Para as imagens de treinamento foram utilizadas somente as imagens da base Wiki, para as quais foi necessário criar de forma manual os arquivos das máscaras-alvo das faces humanas. Esta tarefa foi realizada utilizando a ferramenta LabelMe devido sua praticidade de utilização e a forma organizada como ela gera os dados para utilização posterior, conforme detalhado na Seção 4.5.1.

Figura 34 – Interface da ferramenta LabelMe com uma máscara preenchida em imagem de treinamento da base de dados Wiki.

Fonte: o próprio autor.

Uma vez apresentado o contexto do treinamento da rede Mask R-CNN para atingir o objetivo da classificação de gênero, cabe detalhar cada uma das quatro abordagens utilizadas nesse caminho.

O primeiro formato de teste consistiu em realizar o treinamento completo de todas as camadas da rede sem realizar *transfer learning* e sem utilizar imagens de fundo vazio (aquelas sem indivíduos nas fotos). Esse método é resumido como "sem *transfer learning* e sem fundo vazio" e foi abreviado como nTFnFV. O objetivo desse teste é avaliar a eficácia de um modelo completamente dedicado para as tarefas de segmentação facial e classificação de gênero.

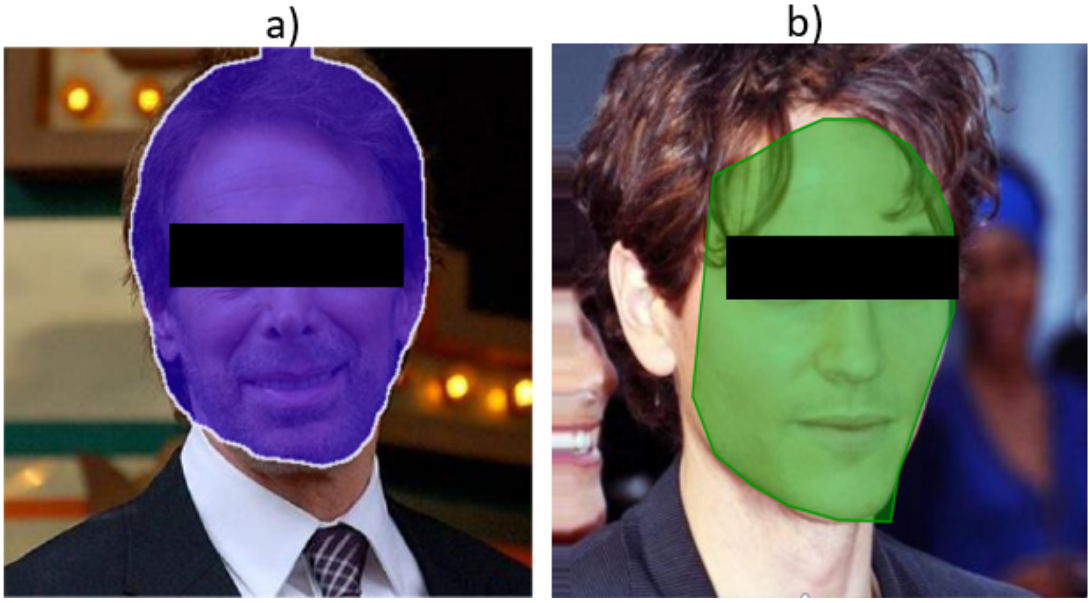
A segunda abordagem consistiu em ajustar apenas os pesos da última camada da rede. Foi aproveitado o conhecimento retido em um modelo ResNet-50 pré-treinado na base ImageNet (RUSSAKOVSKY et al., 2015), que aqui foi adaptado para a tarefa de classificação de gênero. O primeiro objetivo dessa abordagem foi tentar elevar o acerto

de classificação e ao mesmo tempo acelerar os testes, pois com menos parâmetros de aprendizado o custo computacional de treinamento da rede é substancialmente reduzido. Outro interesse foi avaliar se há alteração da acurácia aplicando um modelo pré-treinado em uma base com maior amplitude de cenas. Esse método é resumido como "*transfer learning* sem fundo vazio" e abreviado para TFnFV.

No terceiro momento, foi utilizada estratégia semelhante à da segunda abordagem, mas dessa vez foi inserido apenas ao conjunto de treinamento 100 imagens de fundo vazio, objetivando avaliar uma possível alteração na eficácia do modelo. Esse método é resumido como "*transfer learning* com fundo vazio" e foi abreviado como TFFV.

Por fim, além das alterações para o terceiro teste, na quarta abordagem foi realizada uma revisão de ajuste fino nas máscaras das imagens de treinamento: os preenchimentos das máscaras nas regiões dos contornos faciais foram cuidadosamente alinhados nas bordas de cada face. Nas três abordagens anteriores, as máscaras anotadas manualmente continham a região completa da cabeça do indivíduo. Aqui, a região foi reduzida para se restringir apenas à face, eliminando interferências devido a cortes de cabelo ou possíveis adereços, conforme ilustra a Figura 35. A quarta abordagem é resumida como "*transfer learning* com fundo vazio e máscaras reajustadas" e abreviada para TFFVm.

Figura 35 – Imagens de treinamento com máscaras considerando a cabeça inteira (a) e apenas a face (b).



Fonte: o próprio autor.

4.6 Estimativa de Massa Corporal via IMC

A proposta inicial deste trabalho previu a tentativa de estimar a massa corporal do indivíduo utilizando sua imagem. Contudo, devido à dificuldade em encontrar uma base de imagens gratuita com informações de massa corporal das pessoas, e somado ao fato de serem encontrados trabalhos que estimam o IMC com base em faces e que também disponibilizam suas bases de dados, foi optado por estimar o IMC e, indiretamente, a massa corporal do indivíduo.

Dando continuidade às aplicações envolvendo CNN, foi decidido testar uma rede baseada na arquitetura ResNet-50 com pesos previamente treinados para a estimativa de IMC, tarefa esta realizada analisando unicamente as imagens das faces das pessoas, semelhante ao realizado por Dantcheva, Bremond e Bilinski (2018). O objetivo foi utilizar o IMC estimado para calcular a massa corporal e avaliar a aplicação do método em uma base de dados diferente das já testadas pelos autores da estimativa de IMC. Nesse caso, foi escolhida a base de imagens de fisiculturistas formada por Gonçalves (2017), uma vez que é um grupo que possui características que destoam da população geral em relação aos aspectos físicos como massa corporal e gordura.

Na entrada do algoritmo de estimativa de IMC é informada apenas a imagem contida dentro do *bounding box* do indivíduo previamente encontrado na busca pelo contorno corporal. Em seguida, um detector de faces da biblioteca dlib[5] (KING, 2009) é usado para localizar a face da pessoa. A saída utilizada da dlib é selecionada de forma direta, sem distorção, com apenas uma sub-imagem relativa ao conteúdo interno ao *bounding box* da face, que então é repassada como entrada para o modelo de estimativa de IMC.

Considerando que para a base de dados Fisio-Somatotipo a altura real das pessoas está disponível, e esta é uma informação comumente conhecida pelas pessoas, a abordagem proposta pode ser promissora, pois foca o erro de estimativa apenas na parcela relativa à massa corporal, retirando a parcela relativa à altura.

4.7 Métricas de Avaliação

Conforme já mencionado, neste trabalho são geradas três informações: a segmentação do contorno corporal dos indivíduos, a classificação de gênero e a estimativa de IMC e massa corporal.

Para a segmentação corporal, devido à falta de base de dados com marcação e a complexidade de fazer uma adequada, a avaliação foi feita apenas visualmente e de forma qualitativa. Além disso, foi feita uma avaliação progressiva de acordo com a dificuldade do

[5] A biblioteca dlib possui API disponível em linguagens C++ e Python, e vem sendo atualizada desde sua criação em 2009, acrescentando implementações de técnicas recentes de aprendizado de máquina. O detector de faces utilizado implementa uma rede ResNet.

ambiente em que a imagem foi adquirida. Para isso, foram utilizadas algumas imagens da base Fisio-Somatotipo com algum controle do ambiente, outras com pouco controle do ambiente, e, por fim, imagens da base Internet, com ambientes ainda mais desafiadores.

A validação dos resultados da classificação de gênero e estimativa de IMC e massa corporal foi realizada para a base Fisio-Somatotipo, onde são conhecidos os gêneros e as medidas dos indivíduos.

4.7.1 Avaliação da Classificação de Gênero

A avaliação quantitativa da classificação de gênero foi realizada através das medidas de acurácia (Ac), sensibilidade (Se) e especificidade (Es). A acurácia calcula a taxa de acerto do algoritmo, sendo definida pelo total de acertos dividido pelo total de amostras avaliadas. É uma métrica bastante utilizada quando a base de dados está balanceada entre as classes, como é o caso da base Fisio-Somatotipo somada à base Wiki, em que foram selecionadas quantidades muito próximas de imagens para as classes masculina e feminina, conforme citado na Seção 4.5.2 .

Para a classificação binária de gênero, foi considerada como positiva a classe masculina. Seja M o conjunto de amostras rotuladas na classe masculina e F na feminina. São definidas como Verdadeiros Positivos (VP) as amostras de M classificadas corretamente como M, ao passo que os Verdadeiros Negativos (VN) são as amostras de F classificadas corretamente como F. Os Falsos Positivos (FP) são amostras de F classificadas como M e os Falsos Negativos são as de M classificadas incorretamente como F.

Na medida de acurácia, os resultados tidos como acertos contemplam a soma de VP e VN. A sensibilidade afere a capacidade que o classificador possui em identificar todos os casos positivos, ao passo que a especificidade o faz para os casos negativos.

As Equações 4.3 a 4.5 apresentam as métricas Acurácia (Ac), Sensibilidade (Se) e Especificidade (Es), onde: VP são os verdadeiros positivos, VN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos. Quanto mais próximas de 100%, melhor é o classificador.

$$Ac = \left(\frac{VP + VN}{VP + VN + FP + FN}\right) \times 100\% \tag{4.3}$$

$$Se = \left(\frac{VP}{VP + FN}\right) \times 100\% \tag{4.4}$$

$$Es = \left(\frac{VN}{VN + FP}\right) \times 100\% \tag{4.5}$$

4.7.2 Avaliação da Estimativa de IMC e Massa Corporal

Diferentes critérios podem ser utilizados para medir o desempenho de um modelo treinado, sendo alguns deles relacionados ao erro do estimador. Neste trabalho foram utilizadas cinco métricas de avaliação de erro: Erro Médio (*Mean Error - ME*), Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error — MAPE*), Erro Absoluto Médio (*Mean Absolute Error — MAE*), Erro Quadrático Médio (*Mean Square Error — MSE*) e Raiz do Erro Quadrático Médio (*Root Mean Square Error — RMSE*) (MYTTENAERE et al., 2016).

As Equações 4.6 a 4.10 mostram como calcular as cinco métricas citadas, onde: $\widehat{y_i}$ é a saída obtida pelo estimador, y_i é a saída desejada e n é o número de amostras avaliadas.

$$ME = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i) \tag{4.6}$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|\widehat{y_i} - y_i|}{y_i}\right) \times 100\% \tag{4.7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\widehat{y_i} - y_i| \tag{4.8}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2 \tag{4.9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2} \tag{4.10}$$

Para todas as métricas citadas, quanto maior o desempenho do modelo, mais próximo de 0 é o valor resultante, sendo que se todas atingirem simultaneamente 0 significa que houve a estimativa ideal do modelo. Destas, a única métrica que assume valores negativos é a *ME*.

Valores de *ME* positivos ou negativos indicam que o modelo, na média, estima a saída acima ou abaixo do valor real, respectivamente, o que indica se há enviesamento do modelo.

Enquanto o *MAPE* calcula o erro percentual, comparando os valores estimados com os valores reais, o *MAE* calcula o erro absoluto, sem causar distorção entre o valor real e o estimado. O *MSE* e o *RMSE* são métricas utilizadas para penalizar erros maiores, uma vez que as diferenças entre o real e o estimado são elevadas à segunda potência.

5 Resultados e Discussões

Neste capítulo serão apresentados e analisados os resultados obtidos para cada uma das metodologias propostas neste trabalho.

5.1 Implementação e Recursos Computacionais

Os códigos-fontes deste trabalho foram desenvolvidos na linguagem Python 3 e executados no computador pessoal do autor. Os modelos Mask R-CNN prontos que serão utilizados para os procedimentos apresentados na Seção 4.3 fazem parte da API Detectron[1] (GIRSHICK et al., 2018), e foram construídos pelo autores originais utilizando o *framework* PyTorch (PASZKE et al., 2019). Os modelos Mask R-CNN criados por seus autores são robustos e foram treinados para segmentação semântica de instâncias (incluindo seres humanos) usando milhões de imagens da base ImageNet e utilizando 8 GPU simultaneamente, o que restringe adaptações utilizando *hardwares* mais simples.

Os demais modelos Mask R-CNN, utilizados para o desenvolvimento da metodologia da Seção 4.5.2, foram construídos com o *framework* Keras (CHOLLET et al., 2015) sobre o *backend* TensorFlow (ABADI et al., 2015). O Keras é destinado à construção dos modelos neurais de forma mais amigável que o TensorFlow, que também pode ser usado para tal tarefa, mas que é principalmente o responsável por fazer interface com a GPU em mais baixo nível. O Keras é mais flexível que o PyTorch no quesito gerenciamento de memória de GPU, permitindo limitar a alocação da mesma, o que viabiliza implementações em placas de vídeo mais modestas, mesmo que mais lentamente.

Em termos de hiperparâmetros, para o treinamento das Mask R-CNN da classificação de gênero foi utilizada uma *learning rate* de 0,001 e realizadas 1000 iterações por 20 épocas. Foram utilizadas 200 RoI de treinamento por imagem e *score* mínimo de 0,5 por instância de objeto de treinamento. Conforme citado no Capítulo 4, o conjunto de treinamento possui 330 imagens. O tempo de treinamento foi aproximadamente de 45 minutos por época utilizando *batch size* igual a 16. O tempo de inferência de cada imagem em etapa de validação é de aproximadamente 0,6 segundo.

As configurações da máquina utilizada são:

1. GPU: Nvidia GTX 1070ti 8GB;
2. CPU: Intel Core i5-8400 2,8GHz;
3. RAM: 16GB DDR4 2400MHz;

[1] API Detectron, disponível em: https://github.com/facebookresearch/detectron

4. SSD: Kingston A400 480GB;

5. Placa mãe: Gigabyte B360M Aorus;

6. SO: Linux Ubuntu 16.04 LTS.

5.2 Resultados da Segmentação do Contorno Corporal

Esta seção é destinada aos resultados e análises da segmentação do contorno corporal. Primeiro, será feita uma comparação da aplicação isolada do modelo Mask R-CNN para dois *backbones* diferentes. Depois, serão discutidos os resultados obtidos após a inserção da técnica de *Alpha Matting* no contexto. Por fim, há uma seção dedicada para a análise dos resultados com foco em imagens em ambientes não controlados.

5.2.1 Mask R-CNN com backbone ResNet-50 vs ResNet-101

A Figura 36 mostra uma comparação de segmentação corporal obtida com o método Mask R-CNN utilizando *backbones* ResNet-50 e ResNet-101. Apesar de a imagem testada se tratar de um ambiente com certo controle de iluminação, com poucas sombras e fundo sem muita diversidade de texturas, há uma diferença bastante relevante entre os resultados obtidos. O modelo menos profundo (ResNet-50) obteve erros grosseiros de segmentação, sobretudo em curvas acentuadas, como na região dos ombros e axilas. Até mesmo em regiões menos curvilíneas, como na dorsal e no entorno dos tornozelos, o resultado da segmentação é melhor quando aplicado o modelo mais profundo, o ResNet-101.

No primeiro momento, ao analisar os resultados da Figura 36, a diferença entre ambos é tão notória que há uma impressão de que o modelo mais profundo possa ser potencialmente melhor. Porém, ao alterar o cenário das imagens analisadas, foram obtidos resultados que inserem uma ressalva nessa conclusão prévia, conforme segue abaixo.

A Figura 37 mostra uma nova comparação, mas para outra imagem da base Fisio-Somatotipo. Visivelmente, as discrepâncias entre os resultados foram menores e novamente o modelo mais profundo se saiu melhor. Porém, este mesmo modelo não foi o mais eficaz na totalidade das regiões possíveis.

Na Figura 37 foram destacadas duas regiões em que o resultado obtido para o *backbone* ResNet-101 atingiu uma melhor separação, sendo elas a região entre as pernas e também a região entre a cintura e o braço esquerdo da atleta. Contudo, o modelo com *backbone* ResNet-50 se comportou melhor no quesito de aderência da segmentação aos contornos em algumas partes, por exemplo no antebraço direito e na parte inferior da coxa esquerda.

Figura 36 – Comparação de resultado da segmentação da Mask R-CNN com *backbone* ResNet-50 (a) e ResNet-101 (b) para os atletas de identificação número 65 e 60 da base Fisio-Somatotipo.

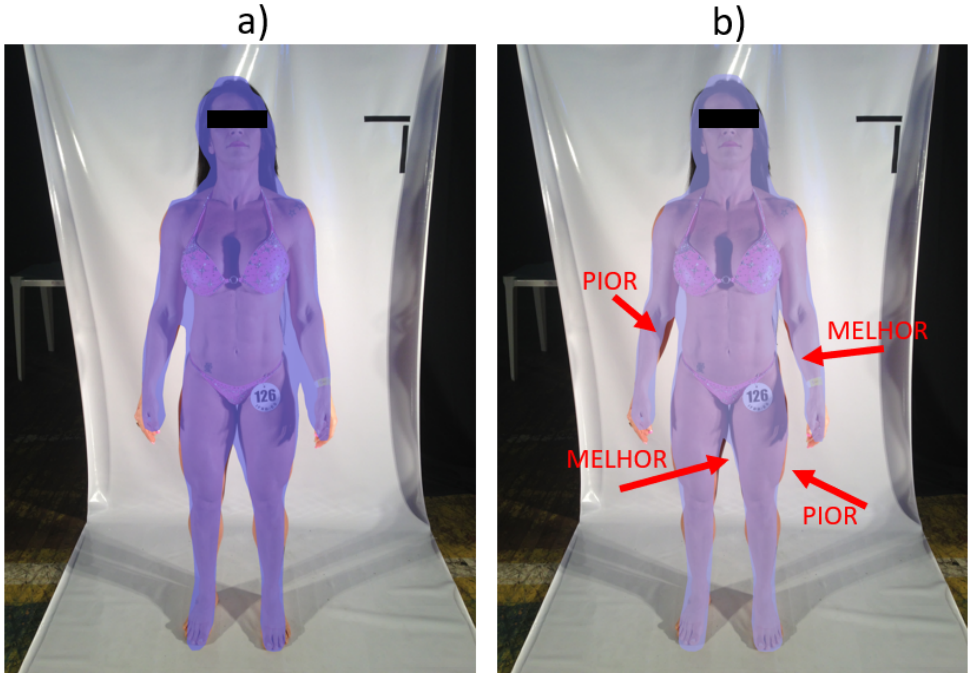
Fonte: o próprio autor.

O resultado apontou que o modelo mais simples falhou pela falta de detalhamento na segmentação em regiões estreitas, como entre o tronco e os braços, enquanto o mais profundo acabou retirando em excesso *pixels* pertencentes ao objeto-alvo, e isso causou perda de qualidade no ajuste fino dos contornos. Esta situação ocorreu especialmente quando o indivíduo estava com os membros próximos do corpo, como visto na Figura 37. Esta condição foi melhor contornada quando o ângulo de abertura de braços e pernas era maior, como visto na Figura 36.

Em resumo, ao aplicar o modelo mais profundo foi encontrada uma solução seguida de um problema. Em geral, ele se saiu melhor e resolveu de maneira mais controlada a separação dos membros, mas nem sempre foi o mais eficaz no ajuste fino ao longo do contorno da pele, sendo necessária a aplicação de uma técnica auxiliar posterior para resolver isso.

A partir daqui, todos os testes executados partiram da utilização da Mask R-CNN

Figura 37 – Mask R-CNN com *backbone* ResNet-101 (b) nem sempre segmenta todas as regiões melhor que com ResNet-50 (a).



Fonte: o próprio autor.

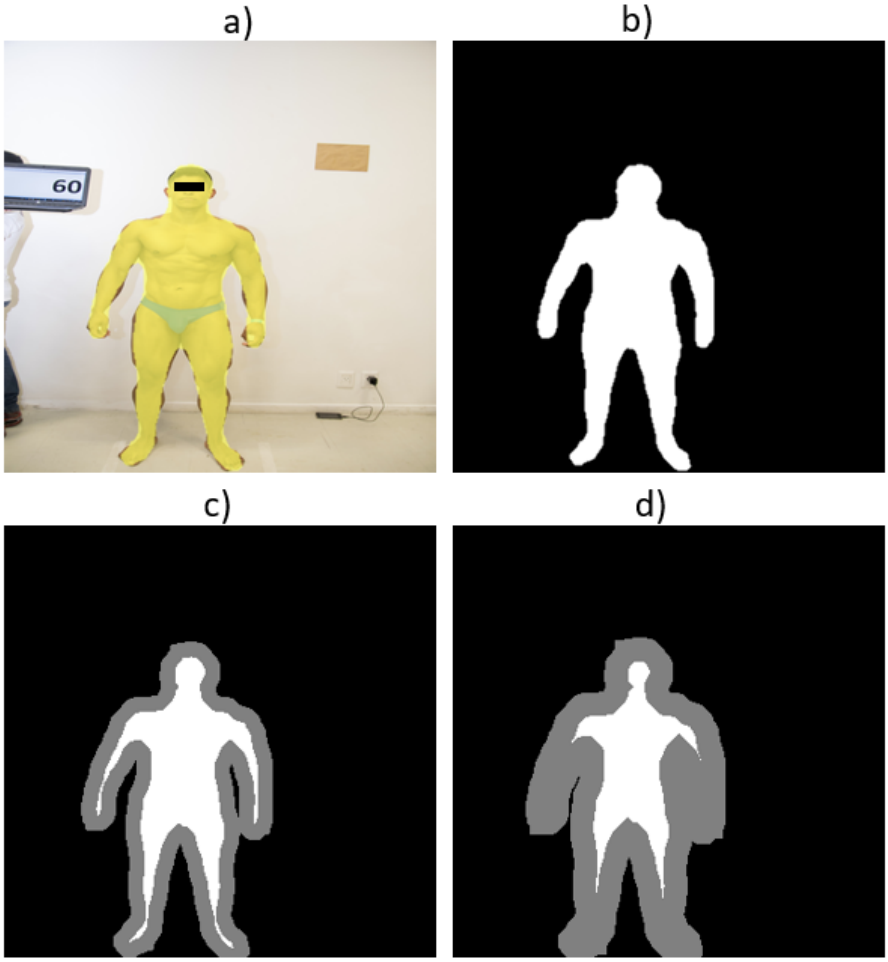
apenas com o *backbone* ResNet-101, pois ele se mostrou mais robusto para ser aproveitado no próximo passo da pesquisa, que foi utilizar o Mask R-CNN para gerar automaticamente os *trimaps* a serem aplicados no *Alpha Matting*. A Figura 38 mostra dois *trimaps* gerados com diferentes espessuras de borda de incerteza, utilizando o processo que foi detalhado na Seção 4.4.

A espessura das bordas dos *trimaps* dependem de dois quesitos: a quantidade de iterações de dilatação e erosão aplicadas; e a resolução da imagem. A próxima seção expõe os resultados obtidos com a utilização da técnica *Alpha Matting*, sendo que foram testadas duas de suas implementações com diferentes espessuras de borda de incerteza dos *trimaps*.

5.2.2 *Deep Image Matting* vs *Closed-Form Matting*

A Figura 39 mostra resultados obtidos para imagens da base Fisio-Somatotipo com a aplicação das duas técnicas de *Alpha Matting* abordadas neste trabalho. Como pode ser observado, o resultado para a *Closed-Form Matting* foi mais preciso na segmentação dos contornos de maneira geral, com destaque para as regiões dos contornos de ombros, orelhas, antebraços, mãos, joelhos e panturrilhas. Todas estas regiões foram melhor segmentadas pelo *Closed-Form Matting*. Apesar disso, este modelo ainda apresenta algumas limitações, sobretudo na separação entre o fundo da imagem e o indivíduo no caso em que os braços estão muito próximos do corpo, condição que eleva a complexidade do desafio.

Figura 38 – Geração de *trimaps* utilizando o método Mask R-CNN e pós-processamento. Imagem (a) resultante do processamento com Mask R-CNN, binarizada em (b) e o resultado para Trimap-2 (c) e Trimap-8 (d).



Fonte: o próprio autor.

O modelo utilizado para a aplicação do *Deep Image Matting* possui uma limitação[2] que exige o redimensionamento das imagens de entrada para a resolução de 256 × 256 *pixels*. Isso fez com que o nível de detalhe obtido nos contornos ficasse aquém do desejado, e por esse motivo foi decidido seguir adiante apenas com a técnica *Closed-Form Matting*, aliada à Mask R-CNN.

5.2.3 Segmentação em ambientes não controlados

Uma vez apresentados os resultados preliminares para ambientes não tão complexos, como os da base Fisio-Somatotipo, o foco da análise de segmentação corporal passa para

[2] Para transpor essa barreira, seria necessário retreinar completamente o modelo, condição inviável para a quantidade de memória de GPU disponível no *hardware* aqui utilizado, mesmo ajustando hiperparâmetros de treinamento como o *batch-size* reduzido, o que poderia causar excessiva lentidão na convergência do modelo, além de poder causar *overfitting*.

Figura 39 – Resultados do processamento das técnicas *Deep Image Matting* (a) e *Closed-Form Matting* (b).

Fonte: o próprio autor.

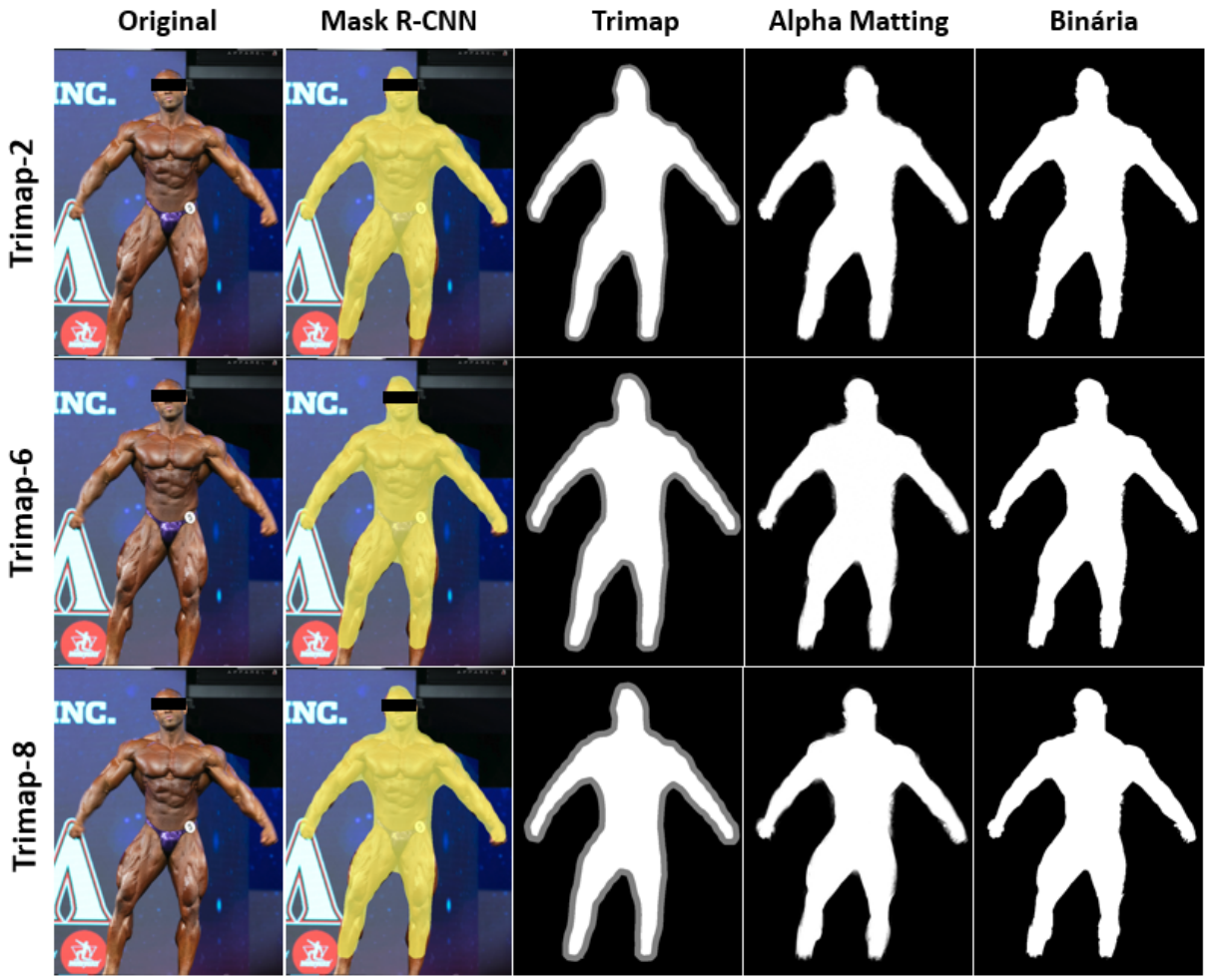
os ambientes mais desafiadores, com variações de texturas de fundo, iluminação, sombras, posicionamentos, como os da base Internet.

A Figura 40 mostra o resultado da segmentação para três *trimaps* diferentes, de espessuras 2, 6 e 8, respectivamente.

Com o aumento da espessura da borda do *trimap* de 2 para 6 há um salto na qualidade da segmentação mais significativo do que entre o *trimap* 6 e o 8. O último apresentou o melhor resultado, sobretudo na região entre o trapézio e o ombro, além da cintura e das laterais da região dorsal, joelhos e panturrilha. A Figura 41 detalha algumas das diferenças citadas.

A Figura 42 mostra a imagem original e o melhor resultado final obtido, que ocorreu para o Trimap-8. Vale destacar a variação de intensidade de iluminação ao longo do corpo do indivíduo, com regiões de saturação de luz, sobretudo nos membros superiores e nas coxas, além dos braços com a imagem borrada por efeito da movimentação que era realizada pela pessoa no momento de captura da foto.

Figura 40 – Resultados do *Closed-Matting* para *Trimap*-2, *Trimap*-6 e *Trimap*-8.

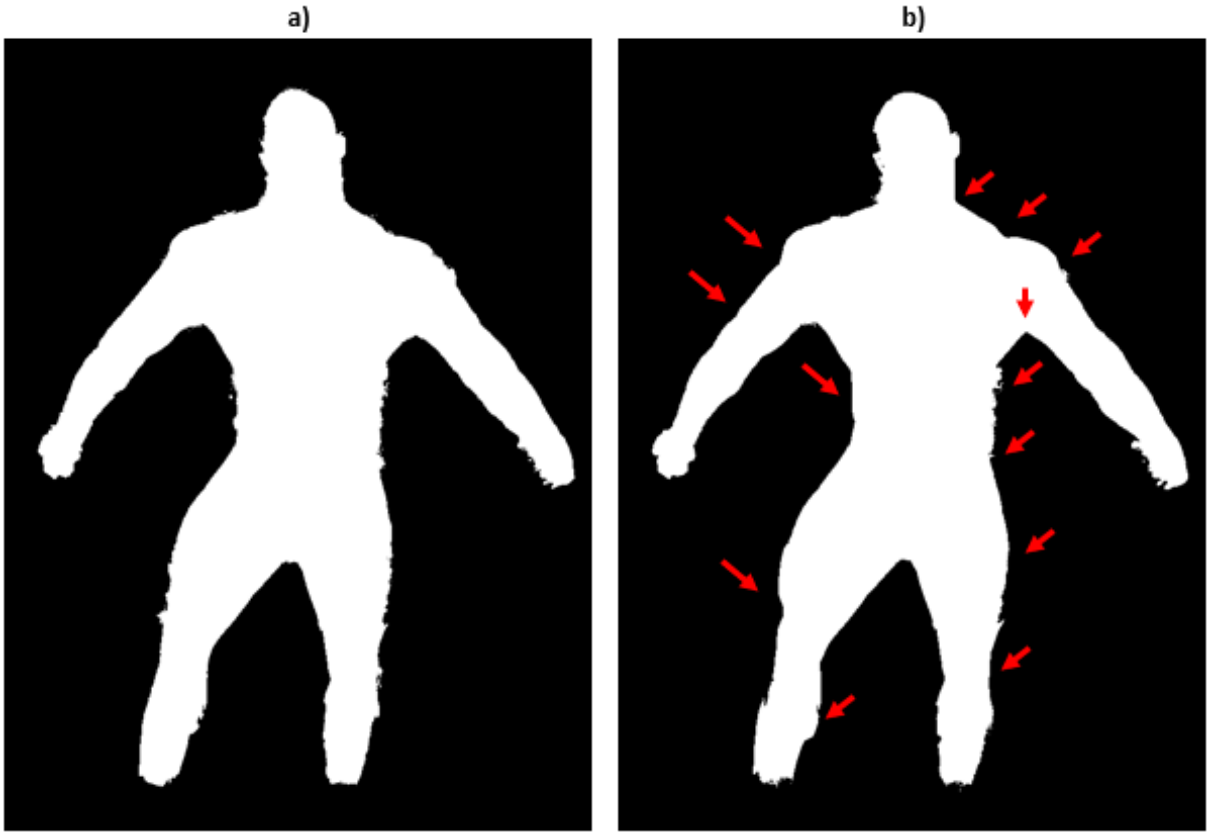


Fonte: o próprio autor.

No Anexo A são exibidas imagens dos resultados finais obtidos para as segmentações corporais das imagens da base Internet. São mostrados casos com os mais variados tipos de plano de fundo e poses dos indivíduos. De maneira geral, os resultados demonstraram que a segmentação atingiu um bom nível de detalhe, conforme objetivado para os contornos corporais mesmo em ambientes desafiadores. Uma ressalva em relação à aplicabilidade da técnica consiste nas poses dos indivíduos, que influenciam diretamente nos resultados quando há regiões fechadas no corpo, por exemplo com braços encostados no tronco ou mãos apoiadas na cintura. Nestas situações foi bem comum parte do fundo ser erroneamente associado ao corpo. Entretanto, conforme o protocolo desenvolvido por Gonçalves (2017) para aquisição de fotos para estimar o somatotipo, as poses sempre devem ser com braços e pernas afastadas, tal que não sejam formadas regiões fechadas. Então, nestas condições o método desenvolvido não teria o problema mencionado.

Ao comparar os resultados de segmentação corporal obtidos pelo trabalho de Gonçalves (2017) com aqueles obtidos por este trabalho, há alguns pontos a serem

Figura 41 – Detalhes da comparação entre os resultados obtidos com *Trimap*-2 (a) e *Trimap*-8 (b).



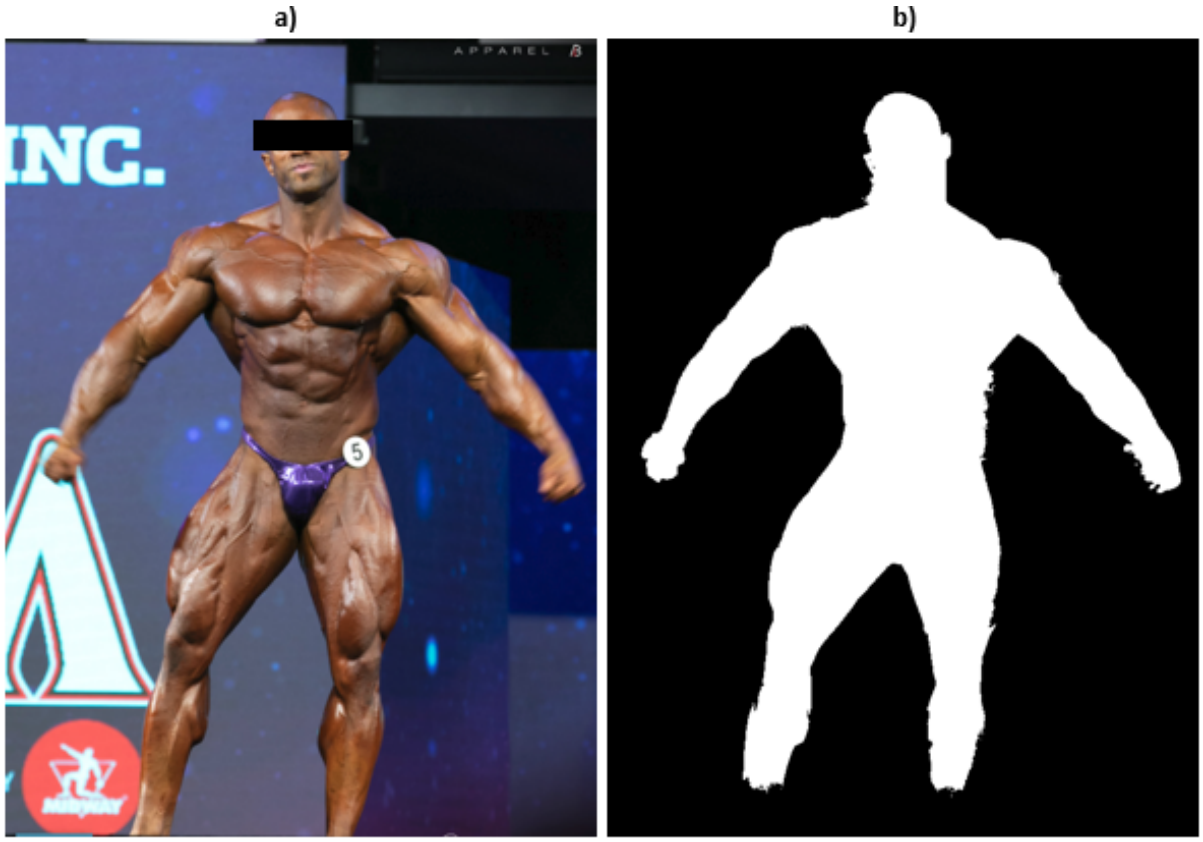
Fonte: o próprio autor.

destacados. A Figura 43 mostra um exemplo em que a segmentação realizada pelo primeiro trabalho não obteve defeitos significativos no resultado final, ao passo que neste trabalho o resultado gerou alguns artefatos no entorno da coxa direita do atleta, por exemplo, devido à proximidade entre a mão e a coxa, o que sugere um ponto fraco no algoritmo aqui desenvolvido.

Entretanto, ao analisar a Figura 44, em que no trabalho anterior foi obtido um resultado que gerava uma região desconexa entre o antebraço e mão esquerda do atleta, é possível notar que este problema foi resolvido pelo algoritmo aqui desenvolvido. Um ponto forte do algoritmo deste trabalho é mitigar a desconexão de regiões do corpo quando há objetos ofuscando a parte da imagem do indivíduo, ao passo que alguns artefatos indesejados podem surgir em alguns resultados, item que poderia ser evoluído com maior avanço em termos de pós processamento. Ademais, vale destacar que a principal evolução alcançada por este estudo é a liberdade de utilizar imagens de ambientes sem controle de iluminação e sem fundo homogêneo, item que gerava limitações em Gonçalves (2017).

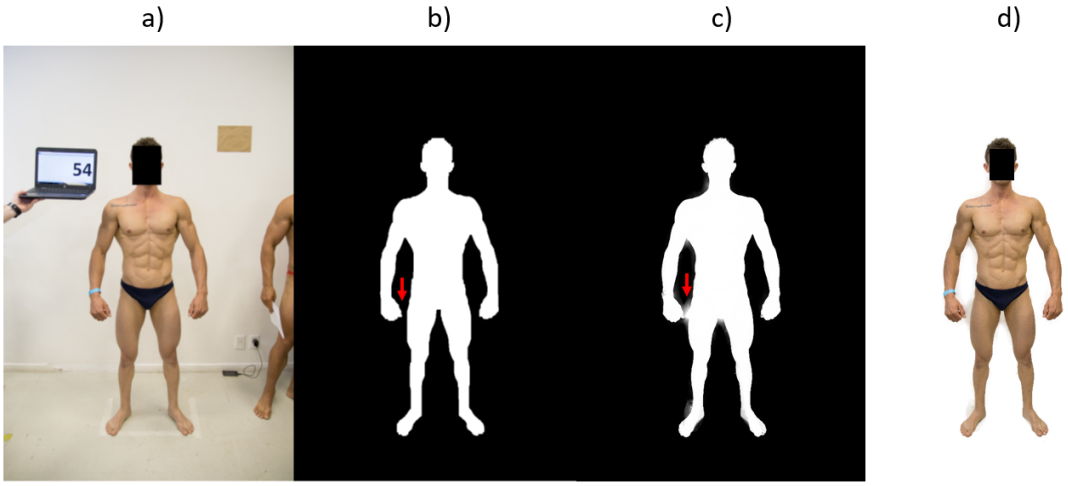
A próxima etapa do trabalho foi utilizar a imagem apenas do interior do *bounding box* do indivíduo, localizado pela Mask R-CNN, para encontrar a face e, baseado na

Figura 42 – Imagem original ampliada da base Internet e sua segmentação final.



Fonte: o próprio autor.

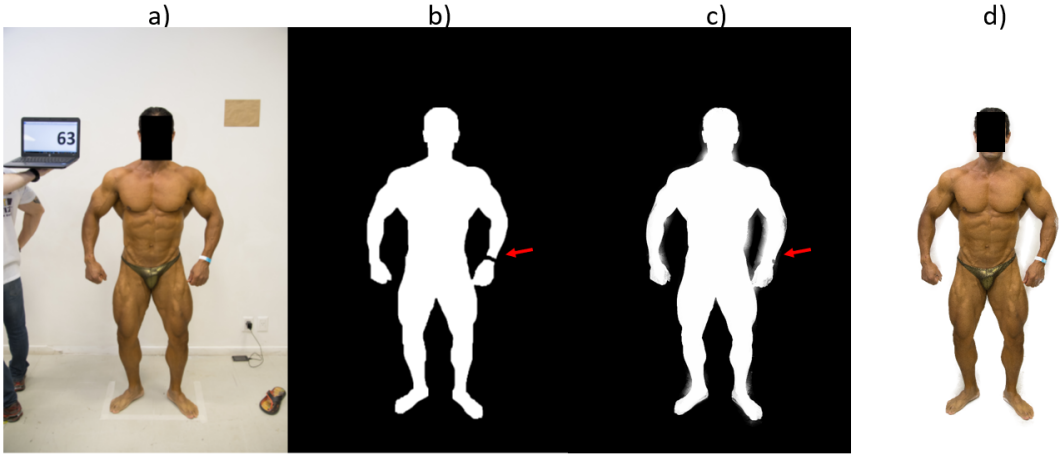
Figura 43 – Segmentação corporal para atleta da base Fisio-Somatotipo. (a) atleta 54, (b) segmentação obtida por Gonçalves (2017), (c) segmentação obtida neste trabalho e (d) imagem resultante.



Fonte: o próprio autor.

imagem, realizar a classificação de gênero da pessoa.

Figura 44 – Segmentação corporal para atleta da base Fisio-Somatotipo. (a) atleta 63, (b) segmentação obtida por Gonçalves (2017), (c) segmentação obtida neste trabalho e (d) imagem resultante.



Fonte: o próprio autor.

5.3 Resultados da Classificação de Gênero

Esta seção é destinada aos experimentos de classificação de gênero, sendo exibidos e analisados os resultados obtidos.

A classificação de gênero foi realizada utilizando as quatro diferentes abordagens detalhadas na Seção 4.5. Para essa tarefa, foram utilizadas 40% das imagens das bases Fisio-Somatotipo e Wiki previamente separadas para validação. A Tabela 2 exibe os resultados obtidos para as métricas de acurácia (Ac), sensibilidade (Se) e especificidade (Es). As métricas foram calculadas considerando o gênero masculino como classe positiva de um classificador binário.

Tabela 2 – Resultados de classificação de gênero para cada abordagem avaliada. Classe masculina considerada como classe positiva para as métricas usadas.

Abordagem	Amostras	*VP*	*VN*	*FP*	*FN*	*Ac*	*Se*	*Es*
nTFnFV	224	93	85	27	19	79,46%	83,04%	75,89%
TFnFV	224	94	89	23	18	81,70%	83,93%	79,46%
TFFV	224	95	89	23	17	82,14%	84,82%	79,46%
TFFVm	224	107	98	14	5	**91,52%**	**95,54%**	**87,50%**

A primeira abordagem, nTFnFV, consistiu em treinar todas as camadas do modelo da rede e não utilizar imagens de fundo vazio no conjunto de treinamento. O resultados obtido por essa abordagem foi pior em todas as métricas, quando comparado com as demais abordagens.

No segundo teste, TFnFV, foi aproveitado um modelo Mask R-CNN pré-treinado na base ImageNet e adaptado para predizer apenas duas classes de saída. A alteração em

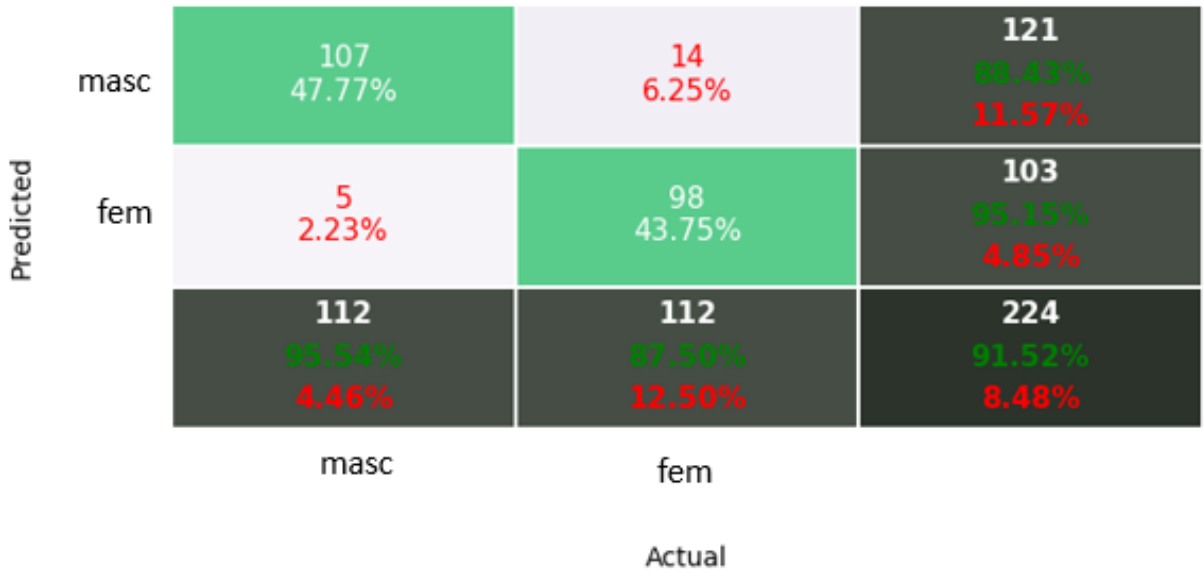
relação ao método anterior surtiu efeito positivo em todas as métricas avaliadas, o que mostra que o classificador passou a avaliar com melhor eficácia que o método anterior.

Na terceira abordagem, TFFV, foram acrescentadas ao grupo de treinamento as imagens de fundo vazio. A estratégia nesse caso é agregar ao modelo mais conhecimento sobre o que é uma imagem em que não há objeto algum de interesse, seja de classe positiva ou negativa. O resultado mostra que a alteração não surtiu efeito algum na especificidade do classificador, mas aumentou timidamente a acurácia, em aproximadamente 0,44%, e a sensibilidade em 0,89%, em relação a abordagem TFnFV.

Como as alterações anteriores surtiram pouco efeito no resultado, foi decidido alterar a forma de lidar com as imagens dos indivíduos. As imagens da base de treinamento possuíam máscaras que consideravam o rosto completo, então foi decidido restringir as regiões das máscaras de forma que possíveis adornos ou cortes de cabelo não influenciassem no modelo.

Após o ajuste fino das máscaras, realizado conforme explicado na Seção 4.5.1, a melhora no resultado foi bastante positiva. A Figura 45 mostra a matriz de confusão para o melhor resultado obtido. A acurácia saltou para 91,52%, a sensibilidade passou para 95,54% e a especificidade para 87,5%. Vale ressaltar que com a última abordagem o número de falsos negativos foi reduzido de 17 para 5, sendo que esses são os casos em que indivíduos do gênero masculino são classificados como gênero feminino. A melhora no acerto foi atribuída ao refinamento das regiões de interesse, fazendo com que brincos, por exemplo, não influenciassem no modelo.

Figura 45 – Matriz de confusão para a melhor classificação de gênero obtida.



Fonte: o próprio autor.

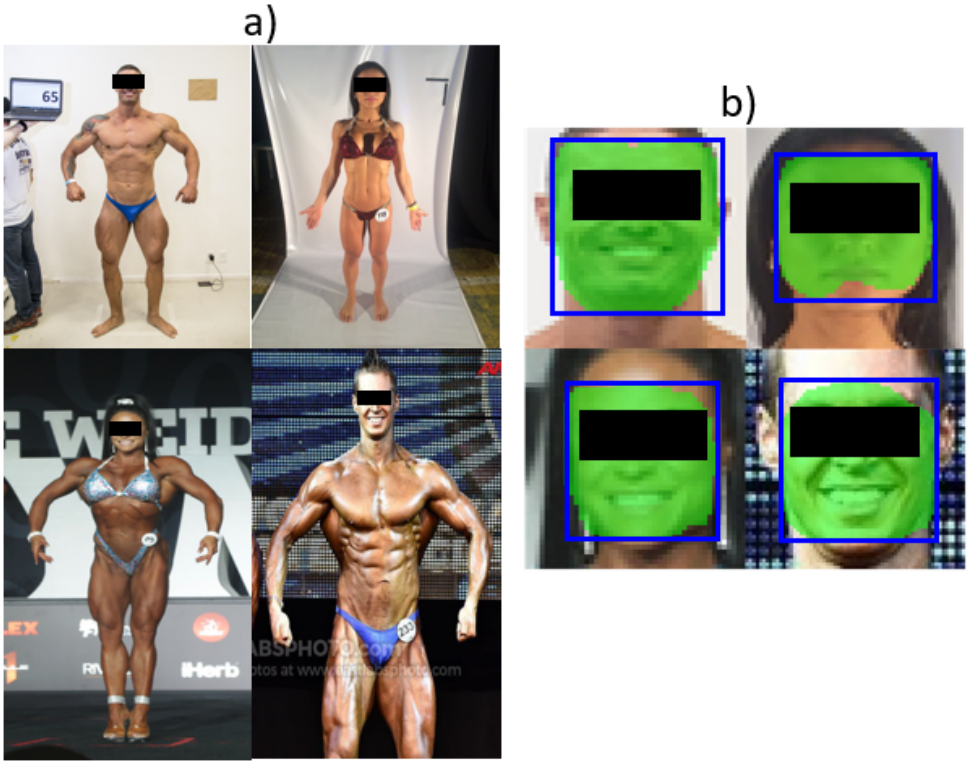
A Figura 46 exemplifica alguns resultados de segmentação facial obtidos para

imagens da base Fisio-Somatotipo e da base Internet. O método foi aplicado em 40% das imagens das duas bases citadas, conjunto este separado para validação do método. Como a identificação facial já parte de uma imagem pré-posicionada restrita ao *bounding box* do indivíduo, isso contribui para a localização da face.

Para todos os testes foi adotado um limiar de 50% na probabilidade associada à classe rotulada para ser considerada válida, ou seja, objetos classificados com *score* inferior a 0,5 são descartados. Para todas as imagens testadas em que há face humana, não houve caso em que a face não fosse encontrada. Além disso, ao testar imagens de fundo vazio, não foram encontradas regiões de face com *score* minimamente relevante.

No entanto, em alguns casos a máscara estimada para a face apresentou certa imprecisão. A Figura 47 exemplifica alguns casos em que o resultado não foi tão eficaz, se comparados aos da Figura 46.

Figura 46 – Imagens das bases Fisio-Somatotipo e internet: (a) amostras de entrada; (b) resultado da segmentação das faces utilizando Mask R-CNN para classificação de gênero.



Fonte: o próprio autor.

De maneira geral, da primeira para a última estratégia, a acurácia do modelo melhorou bastante, mas a principal observação é que, independente da mudança, o classificador sempre obteve resultados melhores para a classe masculina. Isto pode ser concluído pois esta classe é considerada como positiva para o classificador e a métrica de sensibilidade atingiu valores superiores aos de especificidade em todos os cenários de teste.

Figura 47 – Estimativa de máscara facial com falhas.

Fonte: o próprio autor.

É importante salientar que as bases de dados utilizadas para treinamento e validação da classificação estavam balanceadas entre as classes masculina e feminina.

Uma das possibilidades consideradas para explicar o pior desempenho para a classe feminina é que o conjunto de imagens que foi aplicado possui uma grande concentração de imagens de atletas fisiculturistas. Este é um nicho bastante específico, no qual, eventualmente, algumas mulheres podem possuir alguns traços faciais masculinizados, como os formatos do queixo e do contorno mandibular. Esta condição pode enganar o classificador, fazendo com que diminua a quantidade de verdadeiros negativos.

5.4 Resultados das Estimativas de IMC e Massa Corporal

Nesta seção são mostrados os resultados e as análises das estimativas de IMC e massa corporal de pessoas.

As estimativas de IMC e massa corporal foram realizadas para dois conjuntos de imagens da base Fisio-Somatotipo, sendo que ela foi dividida entre as imagens com ambiente menos controlado e ambiente com maior controle de iluminação, conforme abordado na explicação do levantamento dessa base de dados na Seção 4.1.1. Como o modelo de predição utilizado é pré-treinado em outra base, a Face-to-BMI, a validação deste método foi feita em 100% das imagens da base Fisio-Somatotipo.

As Tabelas 3 e 4 mostram os resultados das métricas *ME*, *MAPE*, *MAE*, *MSE* e *RMSE* para a etimativa de IMC e massa corporal por gênero para os dois conjuntos de imagens que formam a base Fisio-Somatotipo. Lembrando que o segundo conjunto de imagens da base Fisio-Somatotipo possui um maior controle de iluminação.

Olhando para a Tabela 3, o primeiro dado que chama atenção ao avaliar os resultados é que o erro do estimador para o gênero feminino foi absolutamente menor, tanto para o IMC quanto para a massa corporal. Há também para o segundo conjunto de imagens (Tabela 4) uma discrepância significativa entre os resultados obtidos por gênero, sendo

Tabela 3 – Métricas para o primeiro conjunto de imagens da base Fisio-Somatotipo.

	IMC			MASSA CORPORAL		
	Geral	Masc	Fem	Geral	Masc	Fem
ME	2,99	3,39	**2,60**	8,41	10,06	**6,77**
MAPE	11,46%	11,65%	**11,26%**	11,46%	11,65%	**11,26%**
MAE	3,19	3,41	**2,97**	8,98	10,12	**7,84**
MSE	14,52	15,73	**13,32**	112,36	135,44	**89,28**
RMSE	3,81	3,97	**3,65**	10,60	11,64	**9,45**

Tabela 4 – Métricas para o segundo conjunto de imagens da base Fisio-Somatotipo.

	IMC			MASSA CORPORAL		
	Geral	Masc	Fem	Geral	Masc	Fem
ME	2,87	3,27	**2,16**	8,39	10,01	**5,56**
MAPE	12,09%	13,31%	**9,94%**	12,09%	13,31%	**9,94%**
MAE	3,41	3,88	**2,60**	10,07	11,97	**6,76**
MSE	18,07	21,79	**11,55**	157,35	205,47	**73,14**
RMSE	4,25	4,67	**3,40**	12,54	14,33	**8,55**

que mais uma vez o resultado para o feminino foi melhor.

Analisando a Tabela 4, enquanto que para o gênero masculino o erro médio na estimativa de massa corporal chega a um valor de 10,01kg, o resultado da mesma métrica para as mulheres cai quase que pela metade, atingindo o valor de 5,56kg. Explorando um pouco mais os dados do segundo conjunto, o *MSE* da massa corporal do grupo feminino é quase um terço do masculino.

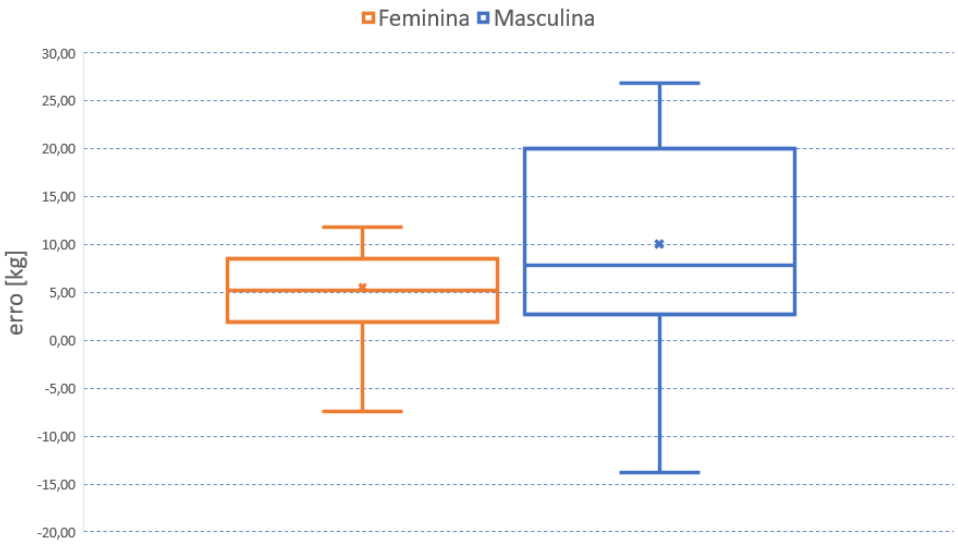
Portanto, observa-se que o erro na estimativa da massa corporal feminina se mostrou menor do que o da masculina e, além disso, a dispersão do erro na estimativa feminina é menor do que na masculina. Por se tratar de uma base de dados sobretudo com atletas fisiculturistas, uma possível explicação para o resultado obtido é a de que atletas desta modalidade e do gênero masculino tendem a atingir valores de massa corporal muito acima da média populacional. Fisiculturistas destoam da população em geral em termos de relação massa corporal vs percentual de gordura, e fisiculturistas homens destoam ainda mais. Vale destacar também que o IMC é uma métrica criada baseada em população geral, o que pode resultar em baixa aderência quando avaliado um conjunto específico como é o caso dos fisiculturistas.

Observando o *ME* sozinho é possível verificar que a média do erro é bem menor. Indo além, ao observar os valores das estimativas obtidas nos dados brutos, foi constatado somente um erro de estimativa maior que 20kg para uma mulher, enquanto que para os homens o estimador atingiu seis amostras com erro absoluto superior a esse, o que faz as métricas *MSE* e *RMSE* penalizarem muito mais o resultado para o masculino. Isto indica

que há uma maior dispersão do erro, conforme a segunda observação citada no parágrafo anterior.

As observações citadas podem também ser analisadas pelos gráficos de *boxplot* dos erros, exibidos na Figura 48, em que as linhas internas são as medianas dos erros, enquanto os X são os próprios *ME*. Para o caso feminino, mediana e média estão próximas, mas para o masculino a média é puxada para uma posição mais distante da mediana, além de serem maiores do que para o caso feminino. Também pode ser observado que há uma dispersão maior do erro para o caso masculino.

Figura 48 – *Boxplots* dos erros de estimativa de massa corporal para cada gênero.



Fonte: o próprio autor.

Para as estimativas de IMC, como é de se imaginar, pode-se observar que o gênero feminino também apresenta os melhores resultados, já que a diferença entre as estimativas do IMC e da massa corporal é dada por uma constante: a altura ao quadrado.

Devido às discrepâncias obtidas para os resultados considerando os diferentes gêneros, foi decidido realizar um teste estatístico sobre as bases de dados e avaliar a hipótese de que o gênero possa influenciar nas estimativas de IMC e massa corporal. Essa condição é sugerida em Dantcheva, Bremond e Bilinski (2018).

Ao realizar o teste estatístico de Mann-Whitney[3] para as amostras de IMC masculinas e femininas da base Fisio-Somatotipo, o p-valor obtido foi 0,00362, portanto, considerando um nível de significância de 5%, existe uma diferença estatística.

[3] O teste de Mann-Whitney é um teste não-paramétrico que testa a hipótese nula de que dois grupos de amostras pertençam à mesma distribuição, ou ainda, à mesma população (NACHAR, 2008). Há diferença significativa entre os grupos de amostras quando *p-value* é menor que a significância definida.

5.5 Comentários Finais

Neste capítulo foram exibidos e discutidos primeiramente os resultados obtidos para a segmentação corporal, depois os de classificação de gênero e, por fim, os da estimativa de IMC e massa corporal.

A segmentação corporal atingiu um nível de detalhe de qualidade elevada nos contornos para imagens da base Fisio-Somatotipo e resultados promissores para as imagens da base Internet. Com isso, o objetivo principal do trabalho foi atingido: elaborar uma técnica que permita segmentar pessoas para aproveitar o resultado no contexto do exame de somatotipo, ampliando o contexto para imagens em ambientes não controlados.

A classificação de gênero atingiu um nível de acurácia de 91,52% para um conjunto de imagens, porém já é um valor que mostra que é completamente viável a abordagem de automatizar a identificação do gênero para não depender de fornecer esse dado manualmente para o algoritmo de exame de somatotipo.

Por fim, em relação à estimativa de IMC e massa corporal, o resultado de erro médio de apenas 5,56kg para o gênero feminino demonstra que um modelo único de estimador pode ser um ponto de partida para a realização da tarefa. No entanto, o resultado para o masculino e o teste estatístico de Mann-Whitney para as amostras da base Fisio-Somatotipo mostram que há diferença estatística considerável entre a distribuição de valores para os dois gêneros, o que leva à conclusão de que não só era provável haver diferença nos resultados obtidos para os diferentes gêneros, como também pode ser recomendável treinar modelos separados por gênero para as estimativas de IMC e massa corporal, o que segue como sugestão de trabalho futuro.

6 Conclusões e Trabalhos Futuros

Motivado pela pesquisa de automatização da estimativa de somatotipo iniciada por Gonçalves et al. (2016), este trabalho testou a aplicação de técnicas conhecidas e apresentou metodologias para obter informações do corpo humano de forma automática utilizando imagens digitais. Para isso, foram utilizadas quatro bases de dados com características importantes para a realização da pesquisa.

6.1 Conclusão

O primeiro objetivo deste trabalho foi realizar a segmentação do contorno corporal dos indivíduos a nível de pixel. Para isso, foi aplicada uma técnica de redes convolucionais, a Mask R-CNN, que foi capaz de realizar uma segmentação preliminar dos contornos, permitindo a aplicação em imagens com diversidade de ambientes, sem necessidade de preparação específica ou pré-processamento das imagens. Posteriormente, a segmentação foi refinada por meio da aplicação automatizada de um algoritmo de *Alpha Matting*, sem a necessidade de gerar *trimaps* manualmente.

Foram testados dois *backbones* para a rede convolucional Mask R-CNN e duas implementações de *Alpha Matting*, sendo que o modelo mais profundo de *backbone* foi mais eficaz. Além disso, foi obtido um refinamento maior dos contornos corporais com a técnica *Closed-Form Matting*. Contudo, devido à falta de recursos computacionais mais robustos, foi inviabilizada a utilização do *Deep Image Matting* em imagens com resolução maior que 256×256 *pixels*, condição que o fez atingir um resultado de menor qualidade que a outra técnica de *Alpha Matting*.

Pelos resultados, foi possível verificar que o algoritmo híbrido (*Mask* + *Matting*) resolve de forma satisfatória o problema tanto em ambientes controlados, quanto em ambientes menos controlados, desde que a pose do indivíduo obedeça a algumas restrições, como o afastamento entre as duas pernas e entre cada braço e o tronco, condição esta que já havia sido recomendada por Gonçalves (2017).

O objetivo principal do trabalho, que era realizar a segmentação em ambientes desafiadores quanto à iluminação e texturas de fundo, foi atingido de maneira satisfatória. Entretanto, não há dúvidas de que podem ser exploradas melhorias em relação ao ajuste fino da segmentação de contorno corporal nas condições em que a pose do indivíduo não contribui positivamente para a realização desta tarefa.

O segundo objetivo foi montar um classificador de gênero que atuasse por meio de análise facial. Baseada em CNN, esta tarefa foi dividida em quatro diferentes estratégias,

abordando técnicas como o treinamento completo de uma rede neural, a adaptação de um modelo pré-treinado e a aplicação de transferência de aprendizado e ajustes na base de dados de treinamento. A tarefa resultou em um bom modelo, que atingiu 91,52% de acurácia, 95,54% de sensibilidade e 87,5% de especificidade.

Por fim, foi realizada a estimativa de IMC e massa corporal, também utilizando redes convolucionais aplicadas à análise facial. Os resultados foram melhores para os indivíduos femininos, tendo atingido valores de erro absoluto médio de $2,6kg/m^2$ e $6,76kg$, respectivamente, enquanto que para os masculinos os resultados médios foram $3,41kg/m^2$ e $10,12kg$.

A discrepância obtida nos resultados entre os gêneros masculinos e femininos foi reforçada pelo teste estatístico realizado, que demonstrou que os erros de estimativa para gênero feminino são estatisticamente menores do que para o masculino. Assim como o IMC, a massa corporal e a altura também são quesitos influenciados pelo gênero, uma vez que são métricas intimamente relacionadas com aquela (DANTCHEVA; BREMOND; BILINSKI, 2018).

Ampliando a análise, os resultados obtidos sugerem que devido ao IMC ser um índice criado para população em geral e, além disso, ser voltado para a determinação de diagnóstico relacionado à quantidade de gordura corporal, sua aplicação para análise de atletas pode não ser tão eficaz, ou nem mesmo aderente, dependendo da modalidade.

No caso do fisiculturismo, muitos atletas possuem massa corporal acima da média para um mesmo gênero, se comparado com a população geral, entretanto, para fotos obtidas em dias de campeonato, a gordura corporal é pouco presente. Sendo assim, a métrica IMC se mostra frágil quando analisados atletas cujo biotipo destoa muito em relação à população geral para a qual o índice foi criado. Esta conclusão é reforçada, por exemplo, comparando os resultados obtidos para a base face2bmi, disponíveis em Dantcheva, Bremond e Bilinski (2018), em que o erro médio de estimativa do IMC é menor do que o obtido neste trabalho para a base Fisio-Somatotipo.

Concluindo, os resultados aqui obtidos são considerados satisfatórios e reforçam o potencial a ser explorado nesta área de pesquisa. Há espaço para a especialização dos modelos de estimativa de IMC em relação ao gênero, tarefa que não foi realizada com maior profundidade devido ao prazo aproximado do término desta pesquisa.

6.2 Trabalhos Futuros

Como trabalho futuro, a primeira possibilidade sugerida é ampliar as amostras das bases de dados para cenários além do fisiculturismo, mas para isso é necessário um enorme trabalho de anotação manual em imagens para estabelecer uma base de treinamento. Ainda

nessa linha, também pode ser interessante realizar a tarefa de anotação de máscara-alvo no conjunto de dados de validação para poder mensurar resultados de forma quantitativa.

Em segundo lugar, podem ser testadas outras técnicas de segmentação de instâncias, sob o interesse de avaliar um maior suporte ao ajuste fino de identificação de contornos corporais, como a técnica *Point Rend* (KIRILLOV et al., 2019) ou a *Open Pose* (CAO et al., 2019)[1].

Uma terceira possibilidade a ser considerada é testar outras implementações da técnica *Alpha Matting*, visto que ela se mostrou bastante promissora para a realização do refinamento das bordas durante a etapa final de segmentação. Algumas que se destacam atualmente são a *Disentangled Image Matting* (CAI et al., 2019) e a SampleNet Matting (TANG et al., 2019).

Uma quarta sugestão é utilizar CNN para tentar estimar o somatotipo diretamente pelas imagens, sem realizar segmentação ou extração prévia de outros componentes, em uma tentativa de identificação direta dos pesos associados às classes de folhetos embrionários que compõem o somatotipo.

Outra proposta é implementar em equipamentos móveis, como celulares, as técnicas já testadas aqui e em (GONÇALVES, 2017). Há a expectativa de que a portabilidade permita ampliar o acesso ao exame de somatotipo para profissionais que não dispõem de todo treinamento necessário para sua execução, mas que podem tirar proveito do diagnóstico automático e ampliar seu leque de análises e orientações para os pacientes e atletas. Vale ressaltar que o *hardware* em si é um requisito custoso apenas no contexto de treinamento de modelos, porém suas execuções para inferência não demandam investimentos significativamente altos.

Para a avaliação de resultados de segmentação, uma proposta é utilizar variáveis qualitativas ordinais. Submetendo-os a questionários para uma população, seria possível qualificá-los dentre categorias, por exemplo: péssimo, ruim, regular, bom e ótimo. Dessa forma, seria possível avaliar os resultados obtidos com os algoritmos de segmentação sem ter, a priori, uma base de dados com máscaras de referência anotadas manualmente. As máscaras resultantes classificadas como ótimas, ou até mesmo as boas, poderiam ser utilizadas para compor uma base de dados para *benchmarks* futuros.

Finalmente, devido à influência do gênero no modelo de estimativa de IMC e massa corporal, uma sugestão é que sejam treinados modelos separados por classe. Outra possibilidade é treinar mais de uma combinação de modelos, separando por gênero e idade, ou por gênero e estatura, enfim, combinações que possam ser testadas de forma a criar modelos especialistas, tal que possam se beneficiar de informações que geralmente são conhecidas pelos usuários, de forma a serem úteis para aumentar a acurácia do sistema.

[1] Vale ressaltar que durante testes finais e fase de escrita deste trabalho ambas as técnicas já haviam sido desenvolvidas, mas por motivo de prazo não foi possível realizar testes com as mesmas.

Referências

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <http://tensorflow.org/>. Citado na página 75.

ACADEMY, D. S. *Deep Learning Book*. [s.n.], 2019. Acesso em: 04/04/2020. Disponível em: <http://deeplearningbook.com.br>. Citado na página 39.

ALMARRI, S. B. S. *Real-Time Facial Emotion Recognition Using Fast R-CNN*. Dissertação (Mestrado) — Rochester Institute of Technology, 2019. Disponível em: <https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=11364&context=theses>. Citado na página 35.

BANDYOPADHYAY, A. Anthropometry and body composition in soccer and volleyball players in west bengal, india. *Journal of physiological anthropology*, v. 26, p. 501–5, 06 2007. Citado na página 24.

BAYIOS, I.; BERGELES, N.; APOSTOLIDIS, N.; NOUTSOS, K.; KOSKOLOU, M. Anthropometric, body composition and somatotype differences of greek elite female basketball, volleyball and handball players. *The Journal of sports medicine and physical fitness*, v. 46, p. 271–80, 07 2006. Citado na página 24.

BIENIAS, L. T.; GUILLAMON, J. R.; NIELSEN, L. H.; ALSTROM, T. S. Insights into the behaviour of multi-task deep neural networks for medical image segmentation. In: *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109/mlsp.2019.8918753>. Citado na página 51.

BRASIL, I. *Classic Games*. 2020. Acesso em: 13/04/2020. Disponível em: <http://ifbbbrasil.com.br/wp-content/uploads/2019/02/Classic-Games.pdf>. Citado na página 60.

CAI, S.; ZHANG, X.; FAN, H.; HUANG, H.; LIU, J.; LIU, J.; LIU, J.; WANG, J.; SUN, J. Disentangled image matting. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2019. Disponível em: <https://doi.org/10.1109/iccv.2019.00891>. Citado na página 93.

CAO, Z.; MARTINEZ, G. H.; SIMON, T.; WEI, S.-E.; SHEIKH, Y. A. OpenPose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2019. Disponível em: <https://doi.org/10.1109/tpami.2019.2929257>. Citado na página 93.

CARTER, J. E. L. *The Heath-Carter anthropometric somatotype - Instruction Manual*. 2002. Citado 2 vezes nas páginas 24 e 32.

CELEBZZ. *Daily Celebrities Pictures*. 2020. Acesso em: 13/04/2020. Disponível em: <http://www.celebzz.com/>. Citado na página 61.

CHELLAPILLA, K.; PURI, S.; SIMARD, P. High Performance Convolutional Neural Networks for Document Processing. In: LORETTE, G. (Ed.). *Tenth International Workshop on Frontiers in Handwriting Recognition*. La Baule (France): Suvisoft, 2006. Http://www.suvisoft.com. Disponível em: <https://hal.inria.fr/inria-00112631>. Citado na página 39.

CHOLLET, F. et al. *Keras*. [S.l.]: GitHub, 2015. <https://github.com/fchollet/keras>. Citado na página 75.

COETZEE, V.; CHEN, J.; PERRETT, D. I.; STEPHEN, I. D. Deciphering faces: Quantifiable visual cues to weight. *Perception*, SAGE Publications, v. 39, n. 1, p. 51–61, jan. 2010. Disponível em: <https://doi.org/10.1068/p6560>. Citado na página 35.

DANTCHEVA, A.; BREMOND, F.; BILINSKI, P. Show me your face and i will tell you your height, weight and body mass index. In: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018. Disponível em: <https://doi.org/10.1109/icpr.2018.8546159>. Citado 4 vezes nas páginas 36, 72, 89 e 92.

DENG, L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, Cambridge University Press (CUP), v. 3, 2014. Disponível em: <https://doi.org/10.1017/atsip.2013.9>. Citado na página 39.

DWIVEDI, P. *Semantic Segmentation — Popular Architectures*. 2019. Acesso em: 16/04/2020. Disponível em: <https://towardsdatascience.com/semantic-segmentation-popular-architectures-dff0a75f39d0>. Citado na página 50.

FILHO, J. F. *A prática da avaliação física: testes, medidas, avaliação física em escolares, atletas e academias de ginástica*. [S.l.: s.n.], 2003. Citado 4 vezes nas páginas 30, 31, 32 e 33.

GARN, S. M. Atlas of men. a guide for somatotyping the adult male at all ages. william h. sheldon. harper, new york, 1954. xvi 357 pp. illus. $10. *Science*, American Association for the Advancement of Science (AAAS), v. 120, n. 3128, p. 980–981, dez. 1954. Disponível em: <https://doi.org/10.1126/science.120.3128.980>. Citado na página 23.

GIRSHICK, R. Fast r-CNN. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015. Disponível em: <https://doi.org/10.1109/iccv.2015.169>. Citado 2 vezes nas páginas 44 e 45.

GIRSHICK, R.; RADOSAVOVIC, I.; GKIOXARI, G.; DOLLáR, P.; HE, K. *Detectron*. 2018. <https://github.com/facebookresearch/detectron>. Citado na página 75.

GIRSHICK, R. B.; DONAHUE, J.; DARRELL, T.; MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2013. Disponível em: <http://arxiv.org/abs/1311.2524>. Citado 3 vezes nas páginas 43, 44 e 49.

GONÇALVES, T. de O. *Identificação do Somatotipo de Fisiculturistas Através de Imagens Digitais*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2017. Disponível em: <http://repositorio.ufes.br/handle/10/6916>. Citado 17 vezes nas páginas 17, 24, 25, 26, 34, 37, 56, 57, 58, 62, 72, 81, 82, 83, 84, 91 e 93.

GONÇALVES, T. de O.; CIARELLI, P. M.; COCO, K. F.; GONÇALVES, E. C.; CASTRO, J. W. B. Uma análise da viabilidade de estimar somatotipos de fisiculturistas através de medições em imagens. *XXI Congresso Brasileiro de Automática (CBA)*, p. 3416–3421, 2016. Disponível em: <https://plataforma.swge.com.br/PROCEEDINGS/PDF/CBA2016-0983.pdf>. Citado 2 vezes nas páginas 34 e 91.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. USA: Prentice-Hall, Inc., 2006. ISBN 013168728X. Citado 3 vezes nas páginas 36, 40 e 63.

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 5 vezes nas páginas 38, 39, 41, 52 e 53.

GREEN, R. D.; GUAN, L. Quantifying and recognizing human movement patterns from monocular video images-part ii: applications to biometrics. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 14, n. 2, p. 191–198, 2004. Citado na página 24.

GUTNIK, B.; ZUOZA, A.; ZUOZIENĖ, I.; ALEKRINSKIS, A.; NASH, D.; SCHERBINA, S. Body physique and dominant somatotype in elite and low-profile athletes with different specializations. *Medicina*, MDPI AG, v. 51, n. 4, p. 247–252, 2015. Disponível em: <https://doi.org/10.1016/j.medici.2015.07.003>. Citado na página 24.

HE, K.; GKIOXARI, G.; DOLLAR, P.; GIRSHICK, R. Mask R-CNN. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017. Disponível em: <https://doi.org/10.1109/iccv.2017.322>. Citado 6 vezes nas páginas 36, 43, 49, 50, 51 e 62.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016. Disponível em: <https://doi.org/10.1109/cvpr.2016.90>. Citado 3 vezes nas páginas 38, 51 e 52.

HEATH, B. H.; CARTER, J. E. L. A modified somatotype method. *American Journal of Physical Anthropology*, Wiley, v. 27, n. 1, p. 57–74, jul. 1967. Disponível em: <https://doi.org/10.1002/ajpa.1330270108>. Citado 5 vezes nas páginas 25, 31, 32, 34 e 37.

HOUCINE, A.; AHMED, A.; SADDEK, Z. Designing a software to count the body composition and somatotype and its role in pursing the morphological state of spotsmen. *AASRI Procedia*, Elsevier BV, v. 8, p. 38–43, 2014. Disponível em: <https://doi.org/10.1016/j.aasri.2014.08.007>. Citado na página 34.

KANG, K.; WANG, X. Fully convolutional neural networks for crowd segmentation. 11 2014. Citado na página 40.

KHAN, A.; SOHAIL, A.; ZAHOORA, U.; SAEED, A. A survey of the recent architectures of deep convolutional neural networks. 01 2019. Citado 2 vezes nas páginas 41 e 42.

KING, D. E. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, v. 10, p. 1755–1758, 2009. Citado na página 72.

KIRILLOV, A.; WU, Y.; HE, K.; GIRSHICK, R. PointRend: Image segmentation as rendering. In: . [S.l.: s.n.], 2019. Citado na página 93.

KOCABEY, E.; CAMURCU, M.; OFLI, F.; AYTAR, Y.; MARíN, J.; TORRALBA, A.; WEBER, I. Face-to-bmi: Using computer vision to infer body mass index on social media. 03 2017. Citado 3 vezes nas páginas 36, 58 e 60.

KOLEVA, M.; NACHEVA, A.; BOEV, M. Somatotype, nutrition, and obesity. *Reviews on Environmental Health*, Walter de Gruyter GmbH, v. 15, n. 4, jan. 2000. Disponível em: <https://doi.org/10.1515/reveh.2000.15.4.389>. Citado na página 23.

KOLEVA, M.; NACHEVA, A.; BOEV, M. Somatotype and disease prevalence in adults. *Reviews on Environmental Health*, Walter de Gruyter GmbH, v. 17, n. 1, jan. 2002. Disponível em: <https://doi.org/10.1515/reveh.2002.17.1.65>. Citado na página 23.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *NIPS*. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 38 e 41.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, Association for Computing Machinery (ACM), v. 60, n. 6, p. 84–90, maio 2017. Disponível em: <https://doi.org/10.1145/3065386>. Citado na página 38.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Institute of Electrical and Electronics Engineers (IEEE), v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <https://doi.org/10.1109/5.726791>. Citado na página 38.

LEE, N. S.; PARK, K. S. titleanthropomorphic classification using three-dimensional fourier descriptor/title. In: BOVIK, A. C.; HOWARD, V. (Ed.). *Biomedical Image Processing II*. SPIE, 1991. Disponível em: <https://doi.org/10.1117/12.44291>. Citado na página 24.

LEVI, G.; HASSNCER, T. Age and gender classification using convolutional neural networks. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2015. Disponível em: <https://doi.org/10.1109/cvprw.2015.7301352>. Citado na página 35.

LEVIN, A.; LISCHINSKI, D.; WEISS, Y. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 30, n. 2, p. 228–242, 2008. Citado 2 vezes nas páginas 55 e 63.

LIEW, S. S.; KHALIL-HANI, M.; RADZI, F.; BAKHTERI, R. Gender classification: A convolutional neural network approach. *Turkish Journal of Electrical Engineering and Computer Sciences*, v. 24, p. 1248–1264, 03 2016. Citado na página 35.

LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft coco: Common objects in context. In: FLEET, D.; PAJDLA, T.; SCHIELE, B.; TUYTELAARS, T. (Ed.). *Computer Vision – ECCV*

2014. Cham: Springer International Publishing, 2014. p. 740–755. ISBN 978-3-319-10602-1. Citado na página 63.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015. Disponível em: <https://doi.org/10.1109/cvpr.2015.7298965>. Citado 4 vezes nas páginas 45, 46, 47 e 49.

LOPES, A. T.; AGUIAR, E. de; SOUZA, A. F. D.; OLIVEIRA-SANTOS, T. Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. *Pattern Recognition*, Elsevier BV, v. 61, p. 610–628, jan. 2017. Disponível em: <https://doi.org/10.1016/j.patcog.2016.07.026>. Citado na página 35.

MACKENZIE, B. *Body Types and Build*. 2001. Acesso em: 10/05/2020. Disponível em: <https://www.brianmac.co.uk/bodytype.htm>. Citado na página 24.

MYTTENAERE, A. de; GOLDEN, B.; GRAND, B. L.; ROSSI, F. Mean absolute percentage error for regression models. *Neurocomputing*, Elsevier BV, v. 192, p. 38–48, jun. 2016. Disponível em: <https://doi.org/10.1016/j.neucom.2015.12.114>. Citado na página 74.

NACHAR, N. The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology*, TQMP, v. 4, n. 1, p. 13–20, 2008. Disponível em: <http://www.tqmp.org/RegularArticles/vol04-1/p013/p013.pdf>. Citado na página 89.

NOH, J.-W.; KIM, J.-H.; KIM, J. Somatotype analysis of elite judo athletes compared with nonathletes for health science research. *Toxicology and Environmental Health Sciences*, Springer Science and Business Media LLC, v. 6, n. 2, p. 99–105, jun. 2014. Disponível em: <https://doi.org/10.1007/s13530-014-0193-x>. Citado na página 24.

NUTTALL, F. Q. Body mass index. *Nutrition Today*, Ovid Technologies (Wolters Kluwer Health), v. 50, n. 3, p. 117–128, 2015. Disponível em: <https://doi.org/10.1097/nt.0000000000000092>. Citado na página 35.

ONLINE, P. H. *Equipamentos antropométricos*. 2020. Acesso em: 13/04/2020. Disponível em: <https://www.produtoshospitalaresonline.com.br/>. Citado na página 33.

PAROLIN, A. *Segmentação de imagens de pessoas em tempo real para videoconferências*. Dissertação (Mestrado) — Universidade do Vale do Rio dos Sinos, 2011. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/3005>. Citado na página 36.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KOPF, A.; YANG, E.; DEVITO, Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.; STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. Pytorch: An imperative style, high-performance deep learning library. In: WALLACH, H.; LAROCHELLE, H.; BEYGELZIMER, A.; ALCHé-BUC, F. d'; FOX, E.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019. p. 8024–8035. Disponível em: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>. Citado na página 75.

PAVLOVIć, R. *ANALYSIS OF THE ELITE ATHLETES' SOMATOTYPES*. 2015. Citado 2 vezes nas páginas 24 e 30.

PINHEIRO, P. O.; COLLOBERT, R. Recurrent convolutional neural networks for scene labeling. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. [S.l.]: JMLR.org, 2014. (ICML'14), p. I–82–I–90. Citado na página 45.

PROLEAGUENEWS. *Mr Olympia Contest Pictures.* 2020. Acesso em: 13/04/2020. Disponível em: <https://proleaguenews.com/category/contest-pictures/mr-olympia-contest-pictures/bodybuilding/>. Citado na página 60.

RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, MIT Press - Journals, v. 29, n. 9, p. 2352–2449, set. 2017. Disponível em: <https://doi.org/10.1162/neco_a_00990>. Citado na página 40.

REN, S.; HE, K.; GIRSHICK, R.; SUN, J. Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Institute of Electrical and Electronics Engineers (IEEE), v. 39, n. 6, p. 1137–1149, jun. 2017. Disponível em: <https://doi.org/10.1109/tpami.2016.2577031>. Citado 4 vezes nas páginas 46, 47, 48 e 49.

RHEMANN, C.; ROTHER, C.; WANG, J.; GELAUTZ, M.; KOHLI, P.; ROTT, P. A perceptually motivated online benchmark for image matting. In: *Proceddings of the IEEE Conference on Computer Vision and Pattern Recognition*. [s.n.], 2009. Posterpräsentation: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR '09, Miami, Florida, USA; 2009-06-20 – 2009-06-25. Disponível em: <http://publik.tuwien.ac.at/files/PubDat_180666.pdf>. Citado na página 54.

ROTHE, R.; TIMOFTE, R.; GOOL, L. V. DEX: Deep EXpectation of apparent age from a single image. In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2015. Disponível em: <https://doi.org/10.1109/iccvw.2015.41>. Citado 2 vezes nas páginas 57 e 59.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015. Citado 3 vezes nas páginas 38, 39 e 70.

RUSSELL, B. C.; TORRALBA, A.; MURPHY, K. P.; FREEMAN, W. T. LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, Springer Science and Business Media LLC, v. 77, n. 1-3, p. 157–173, out. 2007. Disponível em: <https://doi.org/10.1007/s11263-007-0090-8>. Citado na página 68.

SENDIK, O.; KELLER, Y. DeepAge: Deep learning of face-based age estimation. *Signal Processing: Image Communication*, Elsevier BV, v. 78, p. 368–375, out. 2019. Disponível em: <https://doi.org/10.1016/j.image.2019.08.003>. Citado na página 35.

SHARMA, P. *Computer Vision Tutorial: Implementing Mask R-CNN for Image Segmentation (with Python Code).* 2020. Acesso em: 22/04/2020. Disponível em: <https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/>. Citado na página 69.

SHELDON, W. The new york study of physical constitution and psychotic pattern. *Journal of the History of the Behavioral Sciences*, v. 7, n. 2, p. 115–126, 1971. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/1520-6696>. Citado na página 23.

SHELDON, W.; STEVENS, S.; TUCKER, W. *The varieties of human physique: an introduction to constitutional psychology*. Harper, 1940. (Human constitution series). Disponível em: <https://books.google.com.br/books?id=ZG0iAAAAMAAJ>. Citado 8 vezes nas páginas 15, 23, 24, 29, 30, 31, 32 e 37.

SILVA, C. A. D.; MENDES, D. dos S.; OLIVEIRA, E.; ALMEIDA, H. A.; ASCENSO, R. M. T. BodyShifter – software to determine and optimize an individual's somatotype. *Procedia Technology*, Elsevier BV, v. 16, p. 1456–1461, 2014. Disponível em: <https://doi.org/10.1016/j.protcy.2014.10.165>. Citado 2 vezes nas páginas 33 e 34.

SILVA, P. R. P. da; TRINDADE, R. de S.; ROSE, E. H. D. Composição corporal, somatotipo e proporcionalidade de culturistas de elite do brasil. *Revista Brasileira de Medicina do Esporte*, FapUNIFESP (SciELO), v. 9, n. 6, p. 403–407, nov. 2003. Disponível em: <https://doi.org/10.1590/s1517-86922003000600005>. Citado na página 30.

SINGH, S. Somatotype and disease – a review. *Anthropologist*, v. 3, 01 2007. Citado na página 23.

SOUSA, O. M. de. Anatomia Topográfica (Conceitos e Métodos de Estudo). *Revista de Medicina*, Departamento de Anatomia da Faculdade de Medicina da Universidade de São Paulo, 1942. Acesso em: 20/05/2020. Disponível em: <https://www.revistas.usp.br/revistadc/article/download/50129/54244/>. Citado na página 23.

STREUBER, S.; RAMIREZ, M. A. Q.; HILL, M.; HAHN, C.; ZUFFI, S.; O'TOOLE, A.; BLACK, M. Body talk: crowdshaping realistic 3d avatars with words. *ACM Transactions on Graphics*, v. 35, p. 1–14, 07 2016. Citado na página 29.

SUN, P. *Facial Expression Classification Using R-CNN Based Methods*. Dissertação (Mestrado) — Auckland University of Technology, 2019. Disponível em: <https://openrepository.aut.ac.nz/handle/10292/12161?> Citado na página 35.

TANG, J.; AKSOY, Y.; ÖZTIRELI, C.; GROSS, M.; AYDIN, T. O. Learning-based sampling for natural image matting. In: *Proc. CVPR*. [S.l.: s.n.], 2019. Citado na página 93.

TOMPSON, J.; JAIN, A.; LECUN, Y.; BREGLER, C. Joint training of a convolutional network and a graphical model for human pose estimation. 06 2014. Citado na página 45.

TÓTH, T.; MICHALÍKOVÁ, M.; BEDNARČÍKOVÁ, L.; ŽIVČÁK, J.; KNEPPO, P. Somatotypes in sport. *Acta Mechanica et Automatica*, Walter de Gruyter GmbH, v. 8, n. 1, p. 27–32, mar. 2014. Disponível em: <https://doi.org/10.2478/ama-2014-0005>. Citado na página 32.

UCHIDA, Y. *Entenda corretamente a diferença entre o RoI Align do último método de detecção de objeto Mask R-CNN e o RoI Pooling of Fast (er) R-CNN*. 2020. Acesso em: 24/04/2020. Disponível em: <https://qiita.com/yu4u/items/5cbe9db166a5d72f9eb8>. Citado 2 vezes nas páginas 53 e 54.

UIJLINGS, J. R. R.; SANDE, K. E. A. van de; GEVERS, T.; SMEULDERS, A. W. M. Selective search for object recognition. *International Journal of Computer Vision*, Springer Science and Business Media LLC, v. 104, n. 2, p. 154–171, abr. 2013. Disponível em: <https://doi.org/10.1007/s11263-013-0620-5>. Citado na página 43.

WEN, L.; GUO, G. A computational approach to body mass index prediction from face images. *Image and Vision Computing*, Elsevier BV, v. 31, n. 5, p. 392–400, maio 2013. Disponível em: <https://doi.org/10.1016/j.imavis.2013.03.001>. Citado na página 35.

XU, N.; PRICE, B.; COHEN, S.; HUANG, T. Deep image matting. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 311–320. Citado 2 vezes nas páginas 55 e 63.

YAMASHITA, R.; NISHIO, M.; DO, R. K. G.; TOGASHI, K. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, Springer Science and Business Media LLC, v. 9, n. 4, p. 611–629, jun. 2018. Disponível em: <https://doi.org/10.1007/s13244-018-0639-9>. Citado 2 vezes nas páginas 41 e 42.

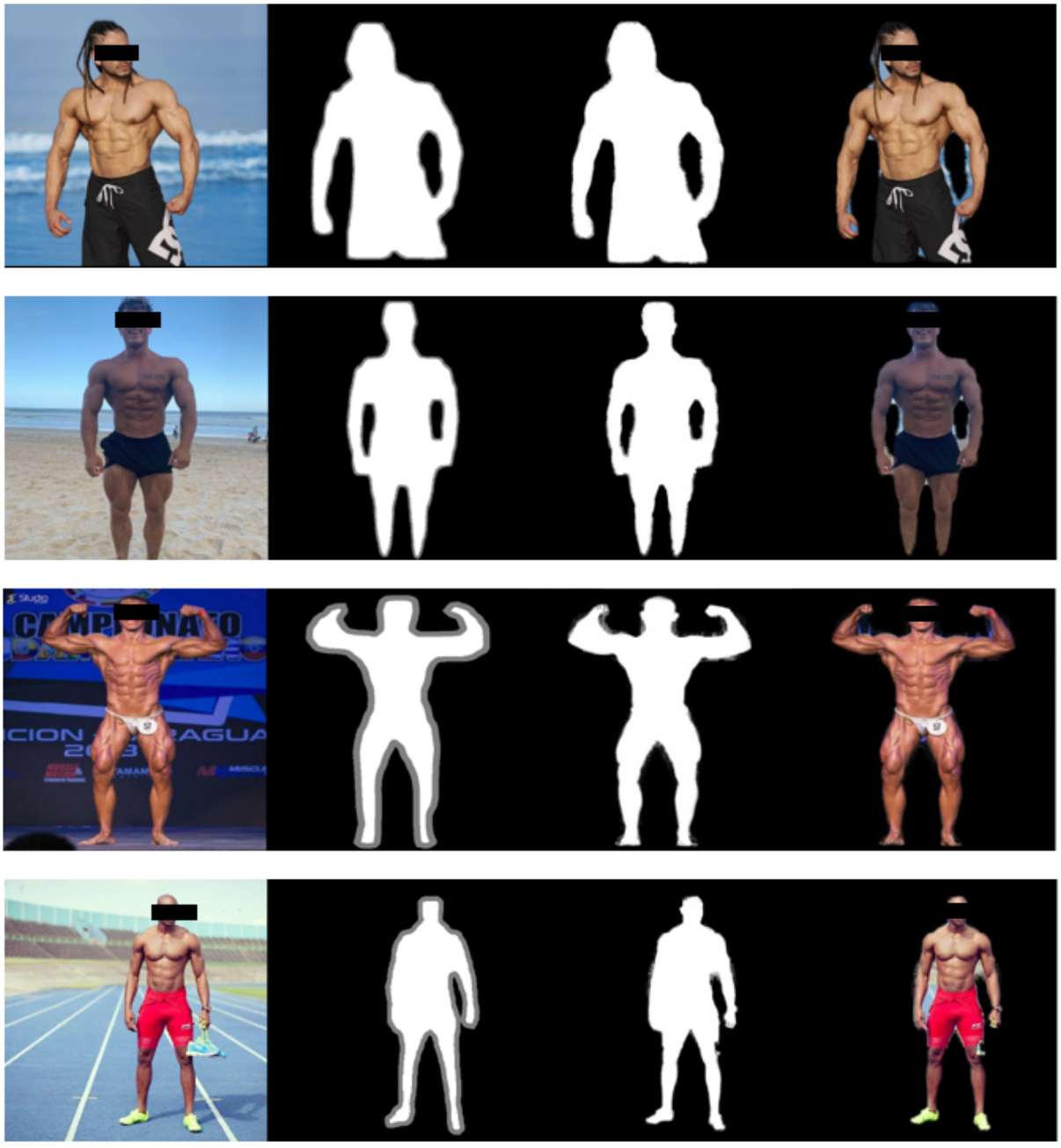
ZHENG, Y.; KAMBHAMETTU, C. Learning based digital matting. In: . [S.l.: s.n.], 2009. p. 889–896. Citado na página 54.

ANEXO A – Resultado da segmentação da base Internet

Figura 49 – Exemplo 1 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.

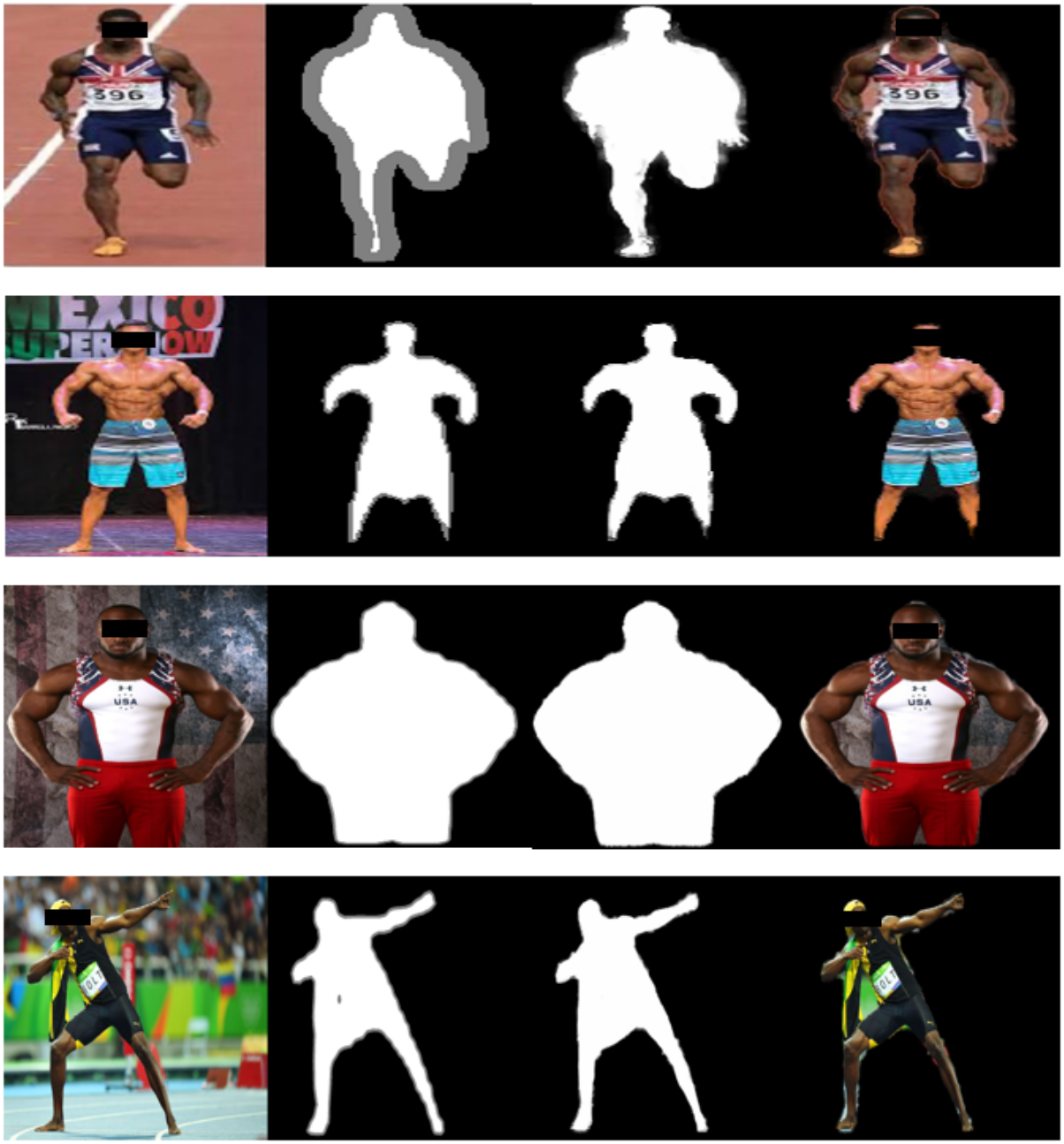
Fonte: o próprio autor.

Figura 50 – Exemplo 2 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.



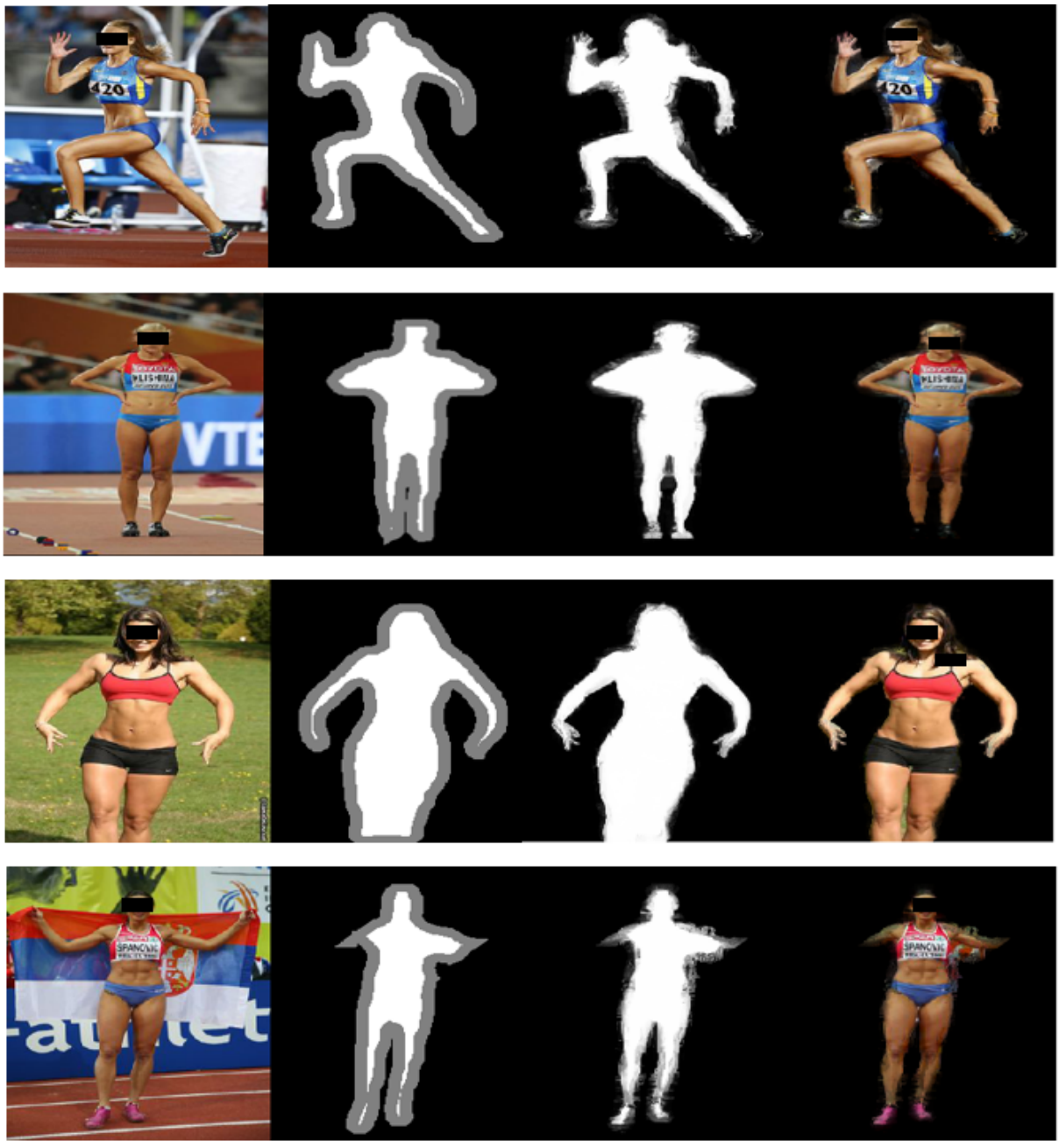
Fonte: o próprio autor.

Figura 51 – Exemplo 3 masculino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.



Fonte: o próprio autor.

Figura 52 – Exemplo 1 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.



Fonte: o próprio autor.

Figura 53 – Exemplo 2 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.

Fonte: o próprio autor.

Figura 54 – Exemplo 3 feminino: imagem original; *trimap*; *alpha matte*; resultado da segmentação.

Fonte: o próprio autor.

ANEXO B – Autorização do uso de imagens da IFBB-ES

FEDERAÇÃO CAPIXABA DE BODYBUILDING E FITNESS
IFBB-ES

AUTORIZAÇÃO

Vitória, 06 de março de 2016.

A IFBB-ES no uso de suas atribuições definidas no Estatuto da IFBB-ES e no que estabelece o Livro de Regras da IFBB Internacional, edição de 2014, na seção 1, artigo 7, autoriza o uso das imagens de seus atletas na pesquisa de mestrado do engenheiro **Thales de Oliveira Gonçalves** no Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Espírito Santo, como também permite que sejam feitas medições antropométricas nos competidores para a referida pesquisa durante a etapa de inscrição e pesagem em seus campeonatos.

A IFBB-ES ressalta que tal autorização se estabelece mediante a devida comunicação dos resultados do estudo a esta Federação bem como a menção desta nas publicações resultantes deste trabalho.

Atenciosamente,

Andreia Borges
PRESIDENTE DA IFBB-ES