

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ  
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

David Silva de Medeiros

**Conjunto de dados de imagens internas e externas rotuladas visando  
ao auxílio à navegação de deficientes visuais**

Fortaleza – CE

16 de dezembro de 2021



David Silva de Medeiros

**Conjunto de dados de imagens internas e externas  
rotuladas visando ao auxílio à navegação de deficientes  
visuais**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Sistemas Embarcados

Orientador: Prof. Dr. Elias Teodoro da Silva Júnior

Coorientador: Prof. Dr. Geraldo Luis Bezerra Ramalho

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

16 de dezembro de 2021



Dados Internacionais de Catalogação na Publicação  
Instituto Federal do Ceará - IFCE  
Sistema de Bibliotecas - SIBI  
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

---

- M488c    Medeiros, David.  
          Conjunto de dados de imagens internas e externas rotuladas visando ao auxílio à navegação de deficientes visuais / David Medeiros. - 2021.  
          77 f. : il.
- Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2021.  
          Orientação: Prof. Dr. Elias Teodoro da Silva Júnior.  
          Coorientação: Prof. Dr. Geraldo Luis Bezerra Ramalho.
1. Tecnologias Assistivas. 2. Navegação por imagens. 3. Conjuntos de dados. 4. Redes Neurais Convolucionais. I. Título.

CDD 004

---



Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)  
Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)  
Programa de Pós-Graduação em Ciência da Computação (PPGCC)

David Silva de Medeiros

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, sendo aprovada pela Coordenação do Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará e pela banca examinadora:

---

**Orientador: Prof. Dr. Elias Teodoro da Silva  
Júnior**

Instituto Federal de Educação, Ciência e Tecnologia do  
Ceará (IFCE)

---

**Coorientador: Prof. Dr. Geraldo Luis Bezerra  
Ramalho**

Instituto Federal de Educação, Ciência e Tecnologia do  
Ceará (IFCE)

---

**Prof. Dr. Romuere Rodrigues Veloso e Silva**  
Universidade Federal do Piauí (UFPI)

---

**Prof. Dr. Amauri Holanda de Souza Junior**  
Instituto Federal de Educação, Ciência e Tecnologia do  
Ceará (IFCE)

Fortaleza – CE  
16 de dezembro de 2021

*Este trabalho é dedicado a todos os alunos de escolas públicas,  
da periferia que mesmo com todas as dificuldades e injustiças sociais  
se esforçam em busca de seus sonhos.*



# Agradecimentos

Gostaria de deixar meus sinceros agradecimentos a todas as pessoas que, de alguma forma, contribuíram para que eu pudesse estar onde estou. A minha família. Meus pais, Maria das Graças e Francisco Alberto. Meus irmãos, Yasmin Silva e Lucas Silva.

Ao Professor Dr. Elias Teodoro da Silva Júnior pela paciência, ensinamentos, orientação e amizade durante o desenvolvimento deste trabalho.

Aos professores(as) Dr. Silviane da Silva Rocha, Dr. Eurípedes Carvalho da Silva e Marcos Monte Cruz pelo reconhecimento e por acreditarem na minha capacidade e esforço. Atitudes e palavras, por menores que sejam, podem mudar completamente o caminho de alguém. A professora MSc. Andréa Coutinho Pessoa de Oliveira por me ensinar a construir pensamento crítico e ser alguém independente na esfera do pensamento.

Gostaria de agradecer também, aos amigos bolsistas Francisco Gleyson Braga Vieira, Nathália Moura Cardoso e Thiago Henrique Araújo que ajudaram em alguma parte no desenvolvimento desse trabalho.

*“Quem é rico de sonhos não envelhece nunca. Pode até ser que morra de repente, mas  
morrerá em pleno voo”  
(Rubem Alves)*

# Resumo

Técnicas de aprendizado de máquina podem ser utilizadas para auxiliar a navegação de deficientes visuais e robôs terrestres autônomos. O aprendizado de máquina tem sido amplamente usado em diversas aplicações baseadas em imagem. As aquisições e construções de novos conjuntos de dados oferecem grande contribuição para a comunidade científica para que sejam avaliadas as técnicas de aprendizado de máquina existentes e explorações de novas. Esta pesquisa apresenta um novo conjunto de dados que associa dados inerciais a imagens de caminhada sob diferentes condições e trajetórias em ambientes internos e externos para classificação de movimentos laterais e frontais. O conjunto de dados apresentado se caracteriza como um problema de regressão, tendo sido agrupado em classes para as análises. CNNs clássicas foram usadas para avaliar o conjunto de dados produzido. Os resultados mostraram uma boa capacidade de representação do problema de evitar colisões e desviar de obstáculos somente a partir das imagens. Uma implementação da VGG16 consegue taxas de acerto superior a 90% quando as opções de direção são arrumadas em 5 classes.

**Palavras-chave:** Tecnologias Assistivas; Navegação por imagens; Conjuntos de dados; Redes Neurais Convolucionais.

# Abstract

Machine learning techniques can be used to aid visually impaired navigation as well as autonomous terrestrial robots. Machine learning has been widely used in many image based applications. The acquisitions and construction of new datasets offer a great contribution to the scientific community to evaluate existing machine learning techniques and explore new ones. This research presents a new dataset that associates inertial data with walking images under different conditions and trajectories in indoor and outdoor environments to classify lateral and frontal movements. The dataset presented is characterized as a regression problem, having been grouped into classes for analysis. Classic CNNs were used to evaluate the produced dataset. The results showed a good ability to represent the problem of avoiding collisions and obstacles only from the images. A VGG16 implementation achieves hit rates of over 90% when the direction options are arranged in 5 classes.

**Keywords:** Assistive technologies; Image based navigation; Datasets; Convolutional Neural Networks.

# Lista de ilustrações

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – A estrutura de uma CNN, consistindo de uma camada convolucional, de pooling e totalmente conectada Saleh et al. (2017) . . . . .                                                                                                                                                                                                                                                                                                                                                           | 27 |
| Figura 2 – Representação em blocos da arquitetura da VGG16 Simonyan e Zisserman (2015a). . . . .                                                                                                                                                                                                                                                                                                                                                                                                      | 29 |
| Figura 3 – Blocos residual da ResNet (HE et al., 2016). . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                       | 30 |
| Figura 4 – Arquitetura da CNN DroNet (a) e Blocos residuais utilizados (b) Loquercio et al. (2018) . . . . .                                                                                                                                                                                                                                                                                                                                                                                          | 31 |
| Figura 5 – Representação do mapa de características de transferência de aprendizagem profunda baseada em rede. Primeiro, a rede foi treinada no domínio de origem com um conjunto de dados de treinamento em grande escala. Em segundo lugar, parte da rede pré-treinada para o domínio de origem é transferida para fazer parte de uma nova rede projetada para o domínio de destino. Finalmente, a sub-rede transferida pode ser atualizada na estratégia de ajuste fino Tan et al. (2018). . . . . | 32 |
| Figura 6 – Diagrama do funcionamento da medição do SSIM Wang et al. (2004). .                                                                                                                                                                                                                                                                                                                                                                                                                         | 33 |
| Figura 7 – Representação da maneira como o dispositivo de captura foi acoplado ao corpo. Um Smartphone foi preso a um cinto de quatro pontos que foi vestido no corpo do pedestre. . . . .                                                                                                                                                                                                                                                                                                            | 42 |
| Figura 8 – Representação tridimensional de caminhada com as acelerações e ângulos de Euler correspondentes. . . . .                                                                                                                                                                                                                                                                                                                                                                                   | 43 |
| Figura 9 – Exemplo de amostras do conjunto das calçadas para velocidades laterais por classe. . . . .                                                                                                                                                                                                                                                                                                                                                                                                 | 45 |
| Figura 10 – Exemplo de amostras do conjunto de corredores para velocidades laterais por classe. . . . .                                                                                                                                                                                                                                                                                                                                                                                               | 46 |
| Figura 11 – Amostra de imagem da Figura 11a com seu respectivo espelhamento para a Classe 0 Figura 11b) . . . . .                                                                                                                                                                                                                                                                                                                                                                                     | 46 |
| Figura 12 – Exemplo de amostras do conjunto de calçadas para velocidades frontais por classe. . . . .                                                                                                                                                                                                                                                                                                                                                                                                 | 47 |
| Figura 13 – Exemplo de amostras do conjunto de corredores para velocidades frontais por classe. . . . .                                                                                                                                                                                                                                                                                                                                                                                               | 48 |
| Figura 14 – Fluxograma com as etapas do processo de construção do conjunto de dados com destaque em verde para os momentos de aplicação dos métodos de similaridade. . . . .                                                                                                                                                                                                                                                                                                                          | 48 |
| Figura 15 – Representação do fluxo de aplicação do método de remoção de imagens similares dentro da própria classe em 3 grupos de exemplo com tamanhos diferentes em uma iteração do método utilizado. . . . .                                                                                                                                                                                                                                                                                        | 50 |

|                                                                                                                                                                                                                                                                                                                                                                                  |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 16 – Boxplot com os resultados de similaridade entre classes na matriz de confusão. . . . .                                                                                                                                                                                                                                                                               | 50 |
| Figura 17 – Representação da aplicação do método de remoção de imagens similares entre classes na matriz de confusão com exemplo de execução na classe 2.                                                                                                                                                                                                                        | 51 |
| Figura 18 – Representação do método de remoção de imagens similares na linha temporal de captura do conjunto de dados. As amostras estão sequenciadas no tempo e a amostra 6 foi a pivô. Dentro do intervalo da janela escolhida para aplicação do SSIM, a amostra 5 e 6 obtiveram similaridade maior do que o limiar e a amostra 5 foi descartada do conjunto de dados. . . . . | 52 |
| Figura 19 – Erro médio quadrático pelo número de épocas para o treinamento da Dronet. Até a época 60 os pesos foram congelados e e as demais épocas até a época 300 os pesos foram descongelados. . . . .                                                                                                                                                                        | 53 |
| Figura 20 – Erro médio quadrático pelo número de épocas para o treinamento da ResNet50. Até a época 20 os pesos foram congelados e e as demais épocas até a época 100 os pesos foram descongelados. . . . .                                                                                                                                                                      | 54 |
| Figura 21 – Exemplos de amostra do conjunto das calçadas. 21a é uma amostra que foi classificada como classe "Acelere"(Classe 4) e 21b é uma amostra que pertence a classe "Acelere". . . . .                                                                                                                                                                                    | 58 |
| Figura 22 – Exemplos de amostra do conjunto das calçadas. 22a é uma amostra que foi classificada como classe 4 e 22b é uma amostra que pertence a classe 4.                                                                                                                                                                                                                      | 61 |
| Figura 23 – Exemplos de amostra do conjunto dos corredores. 23a é uma amostra que foi classificada como classe 4 e 23b é uma amostra que pertence a classe 4. . . . .                                                                                                                                                                                                            | 62 |
| Figura 24 – Boxplot com os resultados de similaridade entre classes na matriz de confusão no conjunto das velocidades frontais para corredores . . . . .                                                                                                                                                                                                                         | 63 |
| Figura 25 – Exemplos de amostra do conjunto dos corredores com SSIM onde houve confusão com a classe "Acelere"(Classe 4). . . . .                                                                                                                                                                                                                                                | 65 |

# Lista de tabelas

|           |                                                                                                                                                                             |    |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1  | – Descrição das classes de 0 a 4 no Conjunto das velocidades laterais e a quantidade de amostras por classe para os conjuntos de calçadas e corredores. . . . .             | 44 |
| Tabela 2  | – Descrição das classes de velocidades frontais e quantidade de amostras por classe para o conjunto das calçadas e corredores. . . . .                                      | 47 |
| Tabela 3  | – Métricas de avaliação das velocidades frontais para calçadas e corredores.                                                                                                | 56 |
| Tabela 4  | – Resultados de acurácia por classe para o conjunto das velocidades frontais.                                                                                               | 56 |
| Tabela 5  | – Matriz de confusão média da <b>VGG16</b> para velocidades frontais no conjunto das calçadas. . . . .                                                                      | 57 |
| Tabela 6  | – Matriz de confusão média da <b>ResNet50</b> para velocidades frontais no conjunto dos corredores. . . . .                                                                 | 57 |
| Tabela 7  | – Matriz de confusão média da <b>VGG16</b> e <b>ResNet50</b> para velocidades frontais em taxas proporcionais para a classe Reduza muito a velocidade.                      | 58 |
| Tabela 8  | – Métricas das velocidades laterais para os conjuntos de calçadas e corredores. . . . .                                                                                     | 59 |
| Tabela 9  | – Resultados de acurácia média por classe para o conjunto das velocidades laterais. . . . .                                                                                 | 60 |
| Tabela 10 | – Matriz de confusão média da <b>VGG16</b> para velocidades laterais no conjunto das calçadas. . . . .                                                                      | 60 |
| Tabela 11 | – Matriz de confusão média da <b>VGG16</b> para velocidades laterais no conjunto dos corredores. . . . .                                                                    | 60 |
| Tabela 12 | – Matriz de confusão média da <b>VGG16</b> para velocidades laterais em taxas proporcionais para as classes Desvie pouco para esquerda e Desvie pouco para direita. . . . . | 61 |
| Tabela 13 | – Matriz de similaridade média da matriz de confusão para as velocidades frontais no conjunto dos corredores. . . . .                                                       | 63 |
| Tabela 14 | – Métricas das velocidades frontais para os conjuntos dos corredores antes e após da aplicação do método de SSIM. . . . .                                                   | 64 |
| Tabela 15 | – Acurácia média por classe para o conjunto dos corredores com SSIM. .                                                                                                      | 65 |
| Tabela 16 | – Matriz de confusão da rede ResNet50 no conjunto de corredores com SSIM para as velocidades frontais. . . . .                                                              | 66 |
| Tabela 17 | – Matriz de confusão em porcentagem proporcional a quantidade de amostras dos conjuntos sem e com SSIM. . . . .                                                             | 66 |



# Lista de abreviaturas e siglas

|        |                                                                                   |
|--------|-----------------------------------------------------------------------------------|
| AT     | Assistive Technology (Tecnologia Assistiva)                                       |
| CNN    | Convolutional Neural Networks (Redes Neurais Convolucionais)                      |
| DVS    | Dynamic Vision Sensor (Sensores de Visão Dinâmica)                                |
| ELU    | Exponential Linear Unit (Unidade Linear Exponencial)                              |
| ENet   | Efficient Neural Network (Rede Neural Eficiente)                                  |
| ERFNet | Efficient Residual Factorized Network (Rede Fatorizada Residual Eficiente)        |
| EOA    | Electronic Orientation Aids (Ajuda Eletrônica para Orientações)                   |
| ETA    | Electronic Travel Aids (Ajuda Eletrônica para Viagens)                            |
| GMM    | Gaussian Mixture Models (Modelo de Mistura Gaussiana)                             |
| GPS    | Global Position System (Sistema de Posicionamento Global)                         |
| GPU    | Graphics Processing Unit (Unidade de Processamento Gráfico)                       |
| MLP    | Multilayer Perceptron (Perceptron Multicamadas)                                   |
| PLD    | Position Locator Devices (Dispositivos Localizadores de Posição)                  |
| OMS    | Organização Mundial da Saúde                                                      |
| ReLU   | Rectified Linear Unit (Unidade Linear Retificada)                                 |
| ResNet | Residual Neural Network (Rede Neural Residual)                                    |
| RGB-D  | Red, Green, Blue-Depth                                                            |
| VGG    | Visual Geometric Group (Grupo de Geometria Visual)                                |
| SSIM   | Structural Similarity Index Measure (Medida do Índice de Similaridade Estrutural) |
| RMSE   | Root Mean Square Error (Erro Médio Quadrático)                                    |



# Lista de símbolos

|               |                                   |
|---------------|-----------------------------------|
| $x$           | Imagem de entrada x               |
| $y$           | Imagem de entrada y               |
| $\mu_x$       | Média da entrada x                |
| $\mu_y$       | Média da entrada y                |
| $\sigma_x$    | Desvio padrão da entrada x        |
| $\sigma_y$    | Desvio padrão da entrada y        |
| $\sigma_{xy}$ | Covariância entre os sinais x e y |

# Sumário

|            |                                                                                                        |           |
|------------|--------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>                                                                                      | <b>21</b> |
| <b>1.1</b> | <b>Justificativa</b>                                                                                   | <b>22</b> |
| <b>1.2</b> | <b>Hipóteses</b>                                                                                       | <b>23</b> |
| <b>1.3</b> | <b>Objetivos Geral e Específicos</b>                                                                   | <b>23</b> |
| <b>1.4</b> | <b>Relevância</b>                                                                                      | <b>23</b> |
| <b>1.5</b> | <b>Problematização</b>                                                                                 | <b>23</b> |
| <b>1.6</b> | <b>Estrutura da Dissertação</b>                                                                        | <b>24</b> |
| <b>2</b>   | <b>FUNDAMENTAÇÃO TEÓRICA</b>                                                                           | <b>25</b> |
| <b>2.1</b> | <b>Tecnologias Assistivas</b>                                                                          | <b>25</b> |
| 2.1.0.1    | Navegação de deficientes visuais                                                                       | 25        |
| <b>2.2</b> | <b>Redes Neurais Convolucionais</b>                                                                    | <b>26</b> |
| 2.2.1      | Composição da CNN                                                                                      | 26        |
| 2.2.1.1    | Camada de Entrada                                                                                      | 27        |
| 2.2.1.2    | Camada Convolucional                                                                                   | 27        |
| 2.2.1.3    | Camadas de agrupamento ou subamostragem                                                                | 28        |
| 2.2.1.4    | Função de ativação                                                                                     | 28        |
| 2.2.1.5    | Lote de normalização                                                                                   | 28        |
| 2.2.1.6    | Camada completamente conectada                                                                         | 28        |
| <b>2.3</b> | <b>Arquiteturas de CNN</b>                                                                             | <b>29</b> |
| 2.3.1      | VGG                                                                                                    | 29        |
| 2.3.2      | ResNet                                                                                                 | 29        |
| 2.3.3      | DroNet                                                                                                 | 30        |
| <b>2.4</b> | <b>Transferência de aprendizado</b>                                                                    | <b>31</b> |
| 2.4.1      | Transferência de aprendizado profundo baseada em rede                                                  | 31        |
| <b>2.5</b> | <b>Medida do Índice de Similaridade Estrutural</b>                                                     | <b>32</b> |
| <b>3</b>   | <b>TRABALHOS RELACIONADOS</b>                                                                          | <b>35</b> |
| <b>3.1</b> | <b>Estado da arte para navegação dos deficientes visuais</b>                                           | <b>35</b> |
| <b>3.2</b> | <b>Estado da arte para CNN com aplicações em navegações com imagens para regressão e classificação</b> | <b>36</b> |
| <b>3.3</b> | <b>Estado da arte para transferência de aprendizado</b>                                                | <b>37</b> |
| <b>3.4</b> | <b>Estado da arte para conjuntos de dados de navegação</b>                                             | <b>38</b> |
| <b>3.5</b> | <b>Considerações</b>                                                                                   | <b>39</b> |
| <b>4</b>   | <b>METODOLOGIA</b>                                                                                     | <b>41</b> |

|            |                                                                          |           |
|------------|--------------------------------------------------------------------------|-----------|
| <b>4.1</b> | <b>Conjunto de dados</b>                                                 | <b>41</b> |
| 4.1.1      | Método de captura das imagens                                            | 41        |
| 4.1.2      | Rotulação das imagens                                                    | 41        |
| 4.1.3      | Pós-processamento do conjunto de dados                                   | 43        |
| 4.1.4      | Conjunto das velocidades laterais                                        | 44        |
| 4.1.5      | Conjunto das velocidades frontais                                        | 45        |
| <b>4.2</b> | <b>Método de remoção de imagens por similaridade</b>                     | <b>46</b> |
| 4.2.1      | Similaridade por classe                                                  | 48        |
| 4.2.2      | Similaridade entre classes na matriz de confusão                         | 49        |
| 4.2.3      | Similaridade entre classes na sequência temporal                         | 51        |
| <b>4.3</b> | <b>CNNs utilizadas no treinamento e avaliação dos conjuntos de dados</b> | <b>52</b> |
| 4.3.1      | Transferência de aprendizado                                             | 53        |
| 4.3.2      | Métricas de avaliação                                                    | 54        |
| <b>5</b>   | <b>RESULTADOS</b>                                                        | <b>55</b> |
| <b>5.1</b> | <b>Avaliação da metodologia proposta</b>                                 | <b>55</b> |
| 5.1.1      | Velocidades frontais                                                     | 55        |
| 5.1.2      | Velocidades laterais                                                     | 58        |
| <b>5.2</b> | <b>Propostas de aprimoramento do conjunto de dados</b>                   | <b>62</b> |
| <b>5.3</b> | <b>Discussão de resultados</b>                                           | <b>66</b> |
| <b>5.4</b> | <b>Reflexões sobre a aplicação embarcada</b>                             | <b>67</b> |
| <b>6</b>   | <b>CONCLUSÕES</b>                                                        | <b>69</b> |
|            | <b>REFERÊNCIAS</b>                                                       | <b>71</b> |



# 1 Introdução

A visão é o principal meio de orientação e navegação de pessoas, de modo que, perdê-la total ou parcialmente implica em ter dificuldades para realizar qualquer tarefa. Segundo a Organização Mundial da Saúde (OMS) é estimado que existam 252,6 milhões de pessoas com deficiência visual, dos quais, 36 milhões possuem deficiência total da visão e 216,6 milhões, deficiência leve ou moderada World Health Organization (2021).

Dentre os problemas que os deficientes visuais enfrentam o de mobilidade é um dos mais estudados pelos pesquisadores porque permite uma maior independência desses indivíduos para navegarem em ambientes internos e externos. As estratégias variam desde o uso de sensores na bengala guia, uso de câmeras ou uma combinação de ambos como dispositivos vestíveis para auxiliar os deficientes visuais a terem segurança na caminhada garantindo a manutenção de um trajeto seguro evitando perigos como colisões, como mostrado em Cheng et al. (2018), Aladren et al. (2016) e Lesecq et al. (2017).

O reconhecimento de padrões a partir dos dados de sensores tornam a capacidade de exploração espacial ainda melhor quando utilizados em algoritmos de aprendizagem de máquina como Redes Neurais Convolucionais (CNN) para problemas de classificação ou regressão. A maioria dos estudos realizados para auxiliar deficientes visuais utilizando CNN são para problemas de classificação como é mostrado em Barontini et al. (2020).

As Redes Neurais precisam ser alimentadas com conjuntos de dados rotulados para realizarem o treinamento dos modelos que serão utilizados para resolver problemas específicos. Apesar de já existirem diversos conjuntos de dados que podem ser utilizados para criar modelos, os dados podem não refletir a realidade do que está sendo investigado e, então, muitos pesquisadores precisam criar seus próprios conjuntos de dados para validarem seus experimentos como é avaliado em Roh, Heo e Whang (2019).

Nos últimos anos as abordagens de aprendizado por transferência começaram a atrair cada vez mais interesses de todas as áreas de pesquisa em reconhecimento de padrões e Aprendizado de Máquina. Quando as técnicas regulares de aprendizado de máquina atingem seus limites, o aprendizado por transferência permite novas possibilidades de um novo fluxo que pode mudar fundamentalmente a maneira como usamos para aprender as coisas e como costumávamos tratar as tarefas de classificação ou regressão (SHAO; ZHU; LI, 2015).

Neste trabalho apresentamos uma abordagem que utiliza imagens para evitar colisões e desvios de obstáculos em ambientes internos e externos. O método proposto consiste em: (1) Captura de imagens ambulantes; (2) Rotulagem de imagem com categorias de velocidade frontal e lateral; (3) Treinar uma CNN com o conjunto de dados gerado com

métodos de aprendizagem por transferência; (4) Avaliação do conjunto de dados proposto. A abordagem utiliza CNN para classificar as velocidades frontais e laterais durante uma caminhada em níveis de intensidade. Este método pode ser aplicado para navegação em robótica terrestre e acessibilidade para deficientes visuais com perda total ou parcial da visão. Além disso, pretende-se utilizar o conjunto de dados proposto, treinado com CNN, para ser aplicado em um sistema embarcado e ser utilizado como dispositivo vestível.

Visto que atualmente as CNN são utilizadas em diversas aplicações, tanto de regressão quanto de classificação, e problemas que possuem mais complexidade, as CNN conseguem extrair mais características dos conjuntos de imagens e consequentemente os modelos treinados com essas redes alcançam, na maioria dos casos, melhores resultados. Por conta disso, a abordagem com CNN foi utilizada neste trabalho. Apesar de utilizar CNN como modelo proposto a ser embarcado, modelos mais simples de classificadores podem ser utilizados visto as limitações de recursos computacionais das plataformas embarcadas.

Para o problema de orientação para deficientes visuais é fundamental que o conjunto de dados forneça informações sobre desvios e colisões no contexto da navegação. Não foram encontrados na literatura conjuntos de dados que relacionam os movimentos durante um caminhada com as imagens no instante do movimento. Portanto, o conjunto de dados proposto neste trabalho representa melhor uma das soluções desse problema.

Sendo assim, esta pesquisa apresenta um estudo sobre dados coletados através das ações de uma pessoa com visão saudável durante uma caminhada e utiliza essas ações como rótulos das imagens. Além disso, utiliza o conjunto de dados proposto em modelos de CNN treinados para tomadas de decisões de desvios e paradas. Portanto, este trabalho pretende investigar o comportamento do conjunto de dados em Redes Neurais Convolucionais (CNN), visando auxiliar a navegação de deficientes visuais em ambientes internos e externos com e sem obstruções de natureza estática ou dinâmica.

## 1.1 Justificativa

Existem vários trabalhos para auxiliar na navegação de deficientes visuais bem como diversos estudos com CNN para esse mesmo propósito Barontini et al. (2020). No entanto, não foram encontrados na literatura conjuntos de imagens rotuladas com dados de velocidades que caracterizam paradas e desvios, utilizando Visão Computacional com CNN para a finalidade de auxiliar a navegação de deficientes visuais capazes de inferir desvios de obstáculos e paradas considerando os sinais responsáveis pelas tomadas de decisões em função dos elementos contidos numa imagem.

## 1.2 Hipóteses

Este trabalho foi embasado nas seguintes hipóteses.

- As intenções de movimento de um pedestre podem ser associadas com imagens.
- Velocidades vetoriais de caminhada conseguem representar desvios e paradas a partir de modelos treinados com CNN.
- O modelo treinado com essas velocidades é capaz de auxiliar na navegação de deficientes visuais.

## 1.3 Objetivos Geral e Específicos

O objetivo geral deste trabalho é construir um conjunto de dados que pode ser treinado com CNN e possua taxas de acurácia que possibilite uma navegação segura para deficientes visuais.

Este trabalho possui como objetivos específicos:

- Construir um conjunto de dados contendo imagens e rótulos de velocidades vetoriais laterais e frontais.
- Propor um método de remoção de amostras que prejudicam a qualidade do conjunto de dados.
- Realizar treinamentos com esses conjuntos de dados em variadas CNN.
- Avaliar os resultados obtidos nessas CNN.
- Avaliar o modelo escolhido para uso em navegação de deficientes visuais.

## 1.4 Relevância

Esta dissertação de mestrado é relevante por ser um estudo que estabelece uma relação entre dados vetoriais de velocidades obtidas em caminhadas e imagens como entrada para treinamentos de Redes Neurais Convolucionais em problemas de classificação.

Em relação a aplicações, esta dissertação apresenta uma abordagem mais generalista de navegação para deficientes visuais que não foi encontrada na literatura e que pode ser aproveitada em formas de navegação diferentes das humanas.

## 1.5 Problematização

Esta dissertação busca avaliar se dados vetoriais obtidos a partir das capacidades motoras de uma pessoa que possui visão saudável conseguem generalizar situações de navegação através de modelos treinados com CNN e esse modelo pode ser utilizado para auxiliar na navegação de deficientes visuais.

## 1.6 Estrutura da Dissertação

No Capítulo 2, são apresentadas as principais fundamentações teóricas utilizadas nesta dissertação. No Capítulo 3, são descritos o estado da arte para a navegação de deficientes visuais apresentando as principais abordagens utilizadas. Em seguida, são apresentadas as principais técnicas utilizadas na literatura nas aplicações com uso de CNN voltadas para navegação utilizando imagens. Além disso, são mostrados trabalhos que utilizaram transferência de aprendizado para problemas de classificação. Por fim, são apresentados trabalhos que envolvem construção de conjuntos de dados voltados para a navegação de deficientes visuais.

O Capítulo 4 descreve as abordagens utilizadas nesta dissertação. A aplicação adotada bem como os conjuntos de dados utilizados são apresentados. Em seguida, são mostrados os processos de treinamento, classificação e avaliação das Redes Neurais Convolucionais utilizadas. São apresentadas as técnicas utilizadas para realizar a transferência de aprendizado. Finalmente, é apresentado um método que reduz as imagens similares que podem prejudicar a construção do conjunto de dados.

No Capítulo 5 são apresentados e discutidos os resultados alcançados nos modelos treinados com o conjunto de dados apresentado. No Capítulo 6, são mostradas as conclusões desta pesquisa bem como sua contribuição e trabalhos futuros.

## 2 Fundamentação teórica

Neste capítulo são apresentados os principais conceitos utilizados nesta pesquisa. Inicialmente será introduzido o conceito de Tecnologias Assistivas no contexto desenvolvido neste trabalho. Em seguida as CNN utilizadas são detalhadas. A transferência de aprendizado é explorada na sequência. Por fim, são mostrados os métodos de avaliação empregados.

### 2.1 Tecnologias Assistivas

Tecnologias Assistivas (AT) podem ser entendidas como dispositivos pessoais utilizados como suporte para que as pessoas com deficiência otimizem suas funções, independência e participações na sociedade Smith et al. (2018). Este trabalho foi focado em AT para portadores de deficiência visual.

#### 2.1.0.1 Navegação de deficientes visuais

A Substituição Visual pode ser explicada como a alternativa para cegos, onde uma imagem pode ser capturada de uma câmera, a informação é processada e a saída é fornecida ao usuário de forma não visual, como modo auditivo, modo vibratório ou uma combinação de ambos. Manjari, Verma e Singal (2020).

De acordo com Elmannai e Elleithy (2017) a Substituição Visual pode ser subdividida em Ajuda Eletrônica para Viagens (ETAs), Ajuda Eletrônica para Orientações (EOAs) e Dispositivos Localizadores de Posição (PLDs). Cada uma dessas subdivisões são responsáveis por resolver problemas específicos de navegação para deficientes visuais. As definições e características segundo Manjari, Verma e Singal (2020) são apresentadas a baixo.

Os ETAs recuperam dados do ambiente e os fornecem para o usuário através de sensores. As características que definem um ETA são:

- Detecção de obstáculos próximo ao corpo do usuário do chão a cabeça;
- Localizar itens ao redor dos obstáculos;
- Conscientizar o usuário com a distância entre ele e o obstáculo com orientações essenciais;
- Fornecer instruções a uma pessoa sobre as condições de uma superfície. Informações a cerca de buracos e texturas.

Os EOAs orientam a pessoa sobre seu caminho, fornecendo instruções ou através

de sinais de caminho, que indicam o que deve ser feito. Possuem como características:

- Seleção do melhor caminho através da definição de rotas;
- Calcular a localização aproximada de um usuário traçando o caminho;
- Orientar o usuário fornecendo sinais de caminho e instruções.

Os PLDs localizam o usuário encontrando sua posição. Isso ajuda as pessoas cegas a se localizarem enquanto viajam. O Sistema de Posicionamento Global (GPS) é um exemplo popular de Dispositivos Localizadores de Posição.

A partir dessas definições e características a solução apresentada nesta pesquisa é um híbrido de ETA e EOA porque detecta obstáculos próximos ao corpo do usuário, conscientiza o usuário com a distância entre ele e os obstáculos e orienta o usuário fornecendo sinais de caminho e instruções. Portanto, é uma solução muito completa para substituição visual.

## 2.2 Redes Neurais Convolucionais

CNN LeCun et al. (1989) são um dos melhores algoritmos de aprendizagem para compreender o conteúdo da imagem e têm mostrado desempenho exemplar em segmentação, classificação, detecção e recuperação de imagens Cireşan et al. (2012), Liu, Deng e Yang (2018). CNN é uma rede hierárquica de múltiplas camadas de alimentação direta, onde cada camada, usando um banco de kernels convolucionais, realiza várias transformações LeCun, Kavukcuoglu e Farabet (2010).

A topologia da CNN é dividida em vários estágios de aprendizagem compostos por uma combinação das camadas convolucionais, unidades de processamento não lineares e camadas de subamostragem Jarrett et al. (2009). A operação de convolução ajuda na extração de recursos úteis de pontos de dados correlacionados localmente. A saída dos kernels convolucionais é então atribuída à unidade de processamento não linear (função de ativação), que não apenas ajuda na aprendizagem de abstrações, mas também incorpora a não linearidade no espaço de recursos. Essa não linearidade gera diferentes padrões de ativações para diferentes respostas e, assim, facilita o aprendizado das diferenças semânticas nas imagens. A saída da função de ativação não linear é geralmente seguida por subamostragem, o que ajuda a resumir os resultados e também torna a entrada invariante para distorções geométricas Scherer, Müller e Behnke (2010), LeCun, Kavukcuoglu e Farabet (2010).

### 2.2.1 Composição da CNN

O modelo padrão da CNN tem uma estrutura que consiste na camada de entrada, camadas convolucionais alternadas com camadas de agrupamento ou subamostragem e

camadas não lineares. A última consiste em uma ou mais camadas totalmente conectadas Namatêvs (2017).

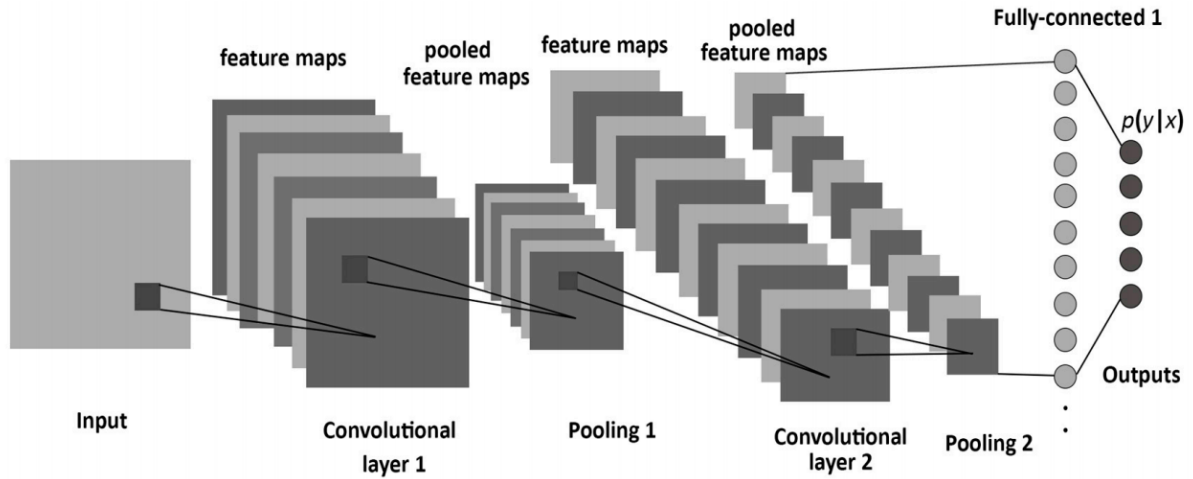


Figura 1 – A estrutura de uma CNN, consistindo de uma camada convolucional, de pooling e totalmente conectada Saleh et al. (2017)

### 2.2.1.1 Camada de Entrada

A entrada é geralmente uma matriz multidimensional de dados onde os dados são alimentados para a rede Ferreira e Giraldi (2017). Os dados de entrada podem ser pixels de imagem ou sua transformação, padrões, séries temporais ou sinais de vídeo.

### 2.2.1.2 Camada Convolucional

O objetivo principal da convolução é extrair características distintas da entrada. Krig (2016) descreve que essas camadas são compostas por uma série de filtros ou kernels aprendíveis que visam extrair características locais da entrada, e cada kernel é usado para calcular um mapa de características ou mapa do kernel.

A primeira camada convolucional extrai baixo nível significativo recursos como bordas, cantos, texturas e linhas. A próxima camada convolucional extrai recursos de nível superior, mas os recursos de nível mais alto são extraídos na última camada de convolução Ren et al. (2012). O tamanho do kernel refere-se ao tamanho do filtro, que gira em torno do mapa de recursos, enquanto o valor pelo qual o filtro desliza (processo de deslizamento) é o passo. Ele controla como o filtro se transforma em torno do mapa de feições. Portanto, o filtro convolve em torno das diferentes camadas do mapa de recursos de entrada, deslizando uma unidade a cada vez Goodfellow, Bengio e Courville (2016).

### 2.2.1.3 Camadas de agrupamento ou subamostragem

As características que resultam de uma saída da operação de convolução podem ocorrer em locais diferentes na imagem. Uma vez que as características são extraídas, sua localização exata se torna menos importante, desde que sua posição aproximada em relação as outras seja preservada. Agrupamento ou subamostragem é uma operação local que resume informações semelhantes na vizinhança do campo receptivo e produz a resposta dominante nesta região local Lee, Gallagher e Tu (2018).

O uso da operação de pooling ajuda a extrair uma combinação de características, que são invariáveis a mudanças translacionais e pequenas distorções Ranzato et al. (2007)(Huang et al. 2007; Scherer et al. 2010). A redução no tamanho do mapa de recursos para o conjunto de características invariáveis não apenas regula a complexidade da rede, mas também ajuda a aumentar a generalização, reduzindo o sobre ajuste. Diferentes tipos de formulações de agrupamento, como agrupamento máximo, a média de uma vizinhança retangular e o agrupamento por redução da resolução são usados na CNN (NAMATEVS, 2017).

### 2.2.1.4 Função de ativação

A função de ativação serve como uma função de decisão e ajuda no aprendizado de padrões intrincados. A seleção de uma função de ativação apropriada pode acelerar o processo de aprendizagem. Na literatura, diferentes funções de ativação, como sigmóide, tanh, maxout, SWISH, ReLU e variantes de ReLU, como ReLU e ELU são usadas para aprender uma combinação não linear de características LeCun et al. (2012).

### 2.2.1.5 Lote de normalização

A normalização em lote é usada para resolver os problemas relacionados a mudança de covariância interna nos mapas de características. A mudança de covariância interna é uma mudança na distribuição dos valores das unidades ocultas, o que retarda a convergência (forçando a taxa de aprendizagem a um valor pequeno) e requer inicialização cuidadosa de parâmetros.

### 2.2.1.6 Camada completamente conectada

A camada totalmente conectada é usada principalmente no final da rede para classificação. Ao contrário do agrupamento e da convolução, é uma operação global. Ele obtém a entrada dos estágios de extração de características e analisa globalmente a saída de todas as camadas anteriores Yoshioka et al. (2015). Conseqüentemente, ele faz uma combinação não linear das características selecionadas, que são usadas para a classificação dos dados Rawat e Wang (2017).

## 2.3 Arquiteturas de CNN

Nesta pesquisa foram utilizadas as arquiteturas de CNN VGG16, ResNet50 e DroNet.

### 2.3.1 VGG

A arquitetura VGG foi criada com o objetivo de investigar o efeito de adicionar profundidade a uma CNN, avaliando sua precisão quando aplicada ao reconhecimento de imagens em grande escala Simonyan e Zisserman (2015a). Especificamente, a arquitetura VGG16 envolve cerca de 138 milhões de parâmetros distribuídos ao longo de suas camadas. Inicialmente, a imagem de entrada é convertida para um tamanho predefinido de  $224 \times 224$ , depois é passada por uma pilha de camadas convolucionais com filtros  $3 \times 3$ , seguida por cinco camadas convolucionais  $2 \times 2$  que realizam um agrupamento espacial, além de três camadas totalmente conectadas, representando cerca de 89% de todos os parâmetros do modelo, e a camada de saída com a função softmax Gopalakrishnan et al. (2017). Uma das principais vantagens deste modelo CNN é que ele é uma arquitetura uniforme com seus pesos pré-treinados no banco de dados ImageNet Russakovsky et al. (2015) publicamente disponível e tem sido aplicado em muitas outras aplicações e desafios como um extrator de recursos eficiente.

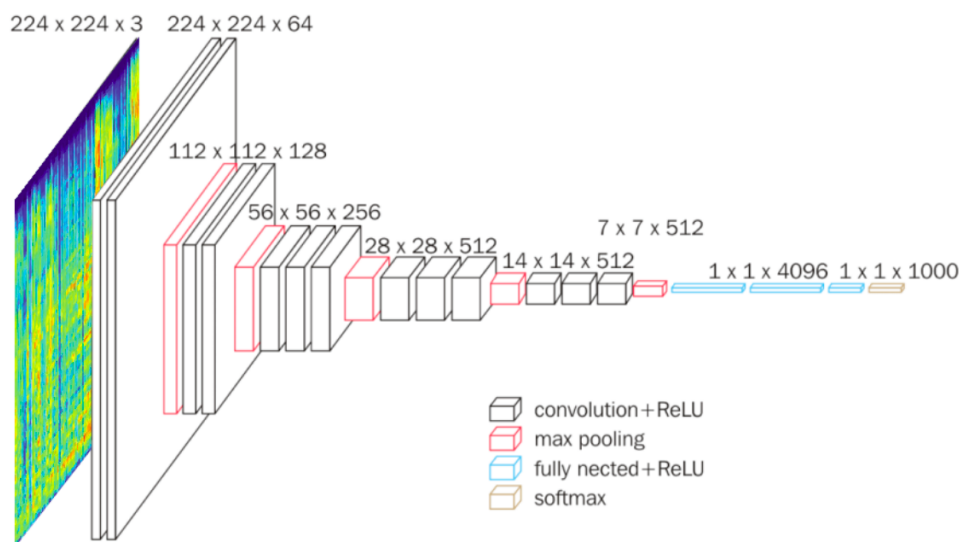


Figura 2 – Representação em blocos da arquitetura da VGG16 Simonyan e Zisserman (2015a).

### 2.3.2 ResNet

A ResNet introduziu o conceito de aprendizado residual nas CNNs e desenvolveu uma metodologia eficiente para o treinamento de redes profundas. Com a ResNet foi

proposto uma CNN com 152 camadas de profundidade, que venceu a competição 2015-ILSVRC. A arquitetura do bloco residual do ResNet é mostrada na Figura 3. A ResNet, que era 20 e 8 vezes mais profundo do que AlexNet e VGG, respectivamente, mostrou menos complexidade computacional do que as redes propostas anteriormente Krizhevsky, Sutskever e Hinton (2017), Simonyan e Zisserman (2015b). Foi mostrado empiricamente que a ResNet com camadas 50/101/152 tem menos erros na tarefa de classificação de imagens do que 34 camadas da Rede simples. Além disso, a ResNet obteve uma melhoria de 28% no conjunto de dados para reconhecimento de imagem denominado COCO Lin et al. (2014). O bom desempenho do ResNet em tarefas de reconhecimento e localização de imagens mostrou que a profundidade representacional é de importância central para muitas tarefas de reconhecimento visual. A ResNet sugeriu que as funções residuais são fáceis de otimizar e podem ganhar precisão para aumentar consideravelmente a profundidade.

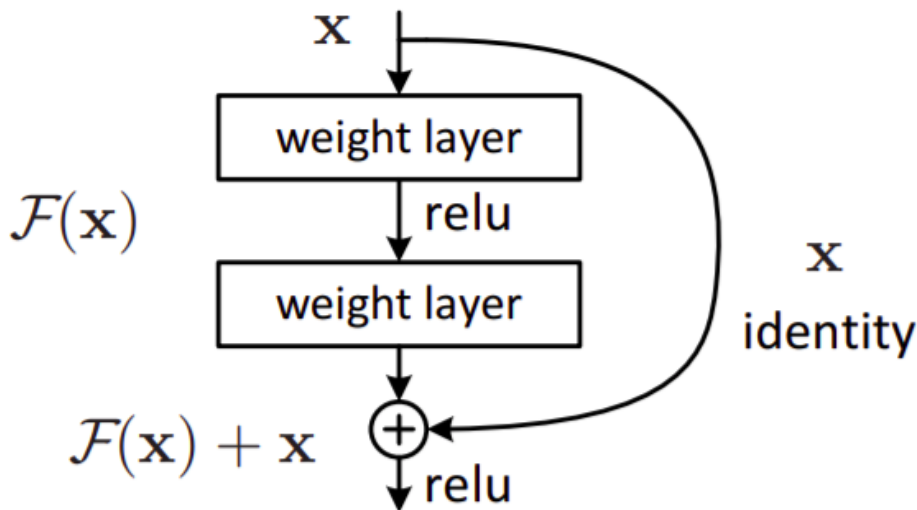


Figura 3 – Blocos residual da ResNet (HE et al., 2016).

### 2.3.3 DroNet

DroNet foi desenvolvida com o objetivo de conduzir de forma confiável um drone autônomo pelas ruas de uma cidade. A DroNet é uma rede neural convolucional bifurcada que prevê, a partir de uma única imagem  $200 \times 200$  em escala de cinza, um ângulo de direção e uma probabilidade de colisão. A parte compartilhada da arquitetura consiste em um ResNet-8 com 3 blocos residuais, seguido por dropout e função de ativação não linear ReLU. Posteriormente, a rede se ramifica em 2 camadas totalmente conectadas separadas, uma para realizar a previsão de direção e a outra para inferir a probabilidade de colisão Loquercio et al. (2018).

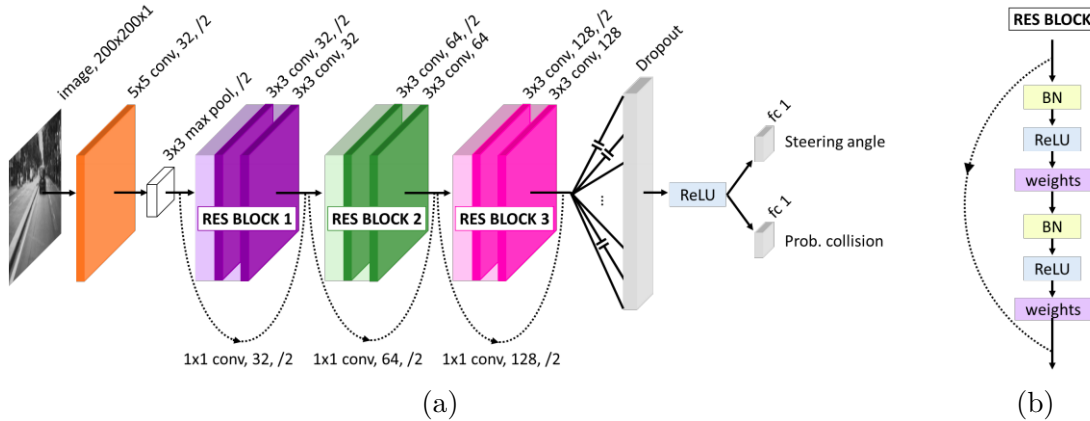


Figura 4 – Arquitetura da CNN DroNet (a) e Blocos residuais utilizados (b) Loquercio et al. (2018)

## 2.4 Transferência de aprendizado

O aprendizado por transferência é uma ferramenta importante no aprendizado de máquina para resolver o problema de dados de treinamento insuficientes. Ele tenta transferir o conhecimento do domínio de origem para o domínio de destino, considerando a suposição de que os dados de treinamento e os dados de teste devem ser independentes e distribuídos de forma idêntica. Isso levará a um grande efeito positivo em muitos domínios que são difíceis de melhorar devido à insuficiência de dados de treinamento Tan et al. (2018).

### 2.4.1 Transferência de aprendizado profundo baseada em rede

A Transferência de aprendizado profundo baseada em rede refere-se à reutilização da rede parcial que está pré-treinada no domínio de origem, incluindo sua estrutura de rede e parâmetros de conexão, transfere-a para ser uma parte da rede neural profunda que é usada no domínio de destino. Baseia-se no pressuposto de que “a rede neural é semelhante ao mecanismo de processamento do cérebro humano e é um processo de abstração iterativo e contínuo. As camadas frontais da rede podem ser tratadas como um extrator de recursos e os recursos extraídos são versáteis”. Uma representação do mapa de características do aprendizado de transferência profunda baseado em rede é mostrado na Figura 5 Tan et al. (2018).

Neste trabalho foi utilizado transferência de aprendizado nas redes VGG16 e ResNet50 pré-treinadas com a base de dados ImageNet Deng et al. (2009) e a rede DroNet com pesos pré-treinados pelos autores.

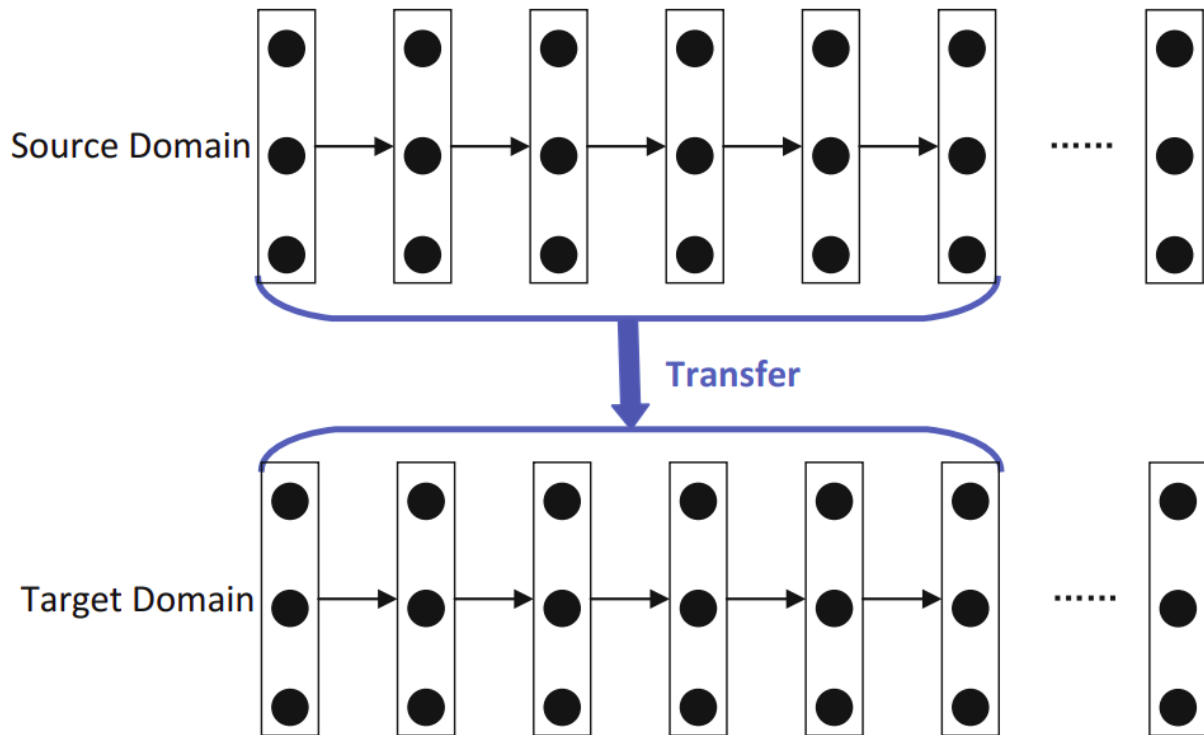


Figura 5 – Representação do mapa de características de transferência de aprendizagem profunda baseada em rede. Primeiro, a rede foi treinada no domínio de origem com um conjunto de dados de treinamento em grande escala. Em segundo lugar, parte da rede pré-treinada para o domínio de origem é transferida para fazer parte de uma nova rede projetada para o domínio de destino. Finalmente, a sub-rede transferida pode ser atualizada na estratégia de ajuste fino Tan et al. (2018).

## 2.5 Medida do Índice de Similaridade Estrutural

O Índice de Similaridade Estrutural ou Structural Similarity Index Measure (SSIM) é um método para medir a similaridade entre duas imagens e foi proposto por Wang et al. (2004). O índice SSIM pode ser visto como uma medida de qualidade de uma das imagens comparadas, desde que a outra imagem seja considerada de qualidade perfeita. O valor do SSIM varia entre -1 e 1. Assume-se 1 o valor resultante quando duas imagens idênticas são comparadas, e -1 quando duas imagens completamente diferentes são dadas como entrada. Na aplicações práticas, resultado de SSIM igual a 0 já representa uma comparação onde as imagens são completamente distintas.

A Figura 6 mostra o diagrama de como o SSIM é feito. Um dois sinais de entrada é considerado como a imagem de referência. A medida de similaridade da imagem de referência pode servir como uma medida quantitativa da qualidade da segunda imagem de entrada. O sistema separa a tarefa de medição de similaridade em três comparações: luminância, contraste e estrutura. Primeiro, a luminância de cada sinal é comparada. Depois, a intensidade média do sinal é removida. Finalmente, o sinal é normalizado

(dividido) por seu próprio desvio padrão, de modo que os dois sinais comparados tenham o desvio padrão da unidade.

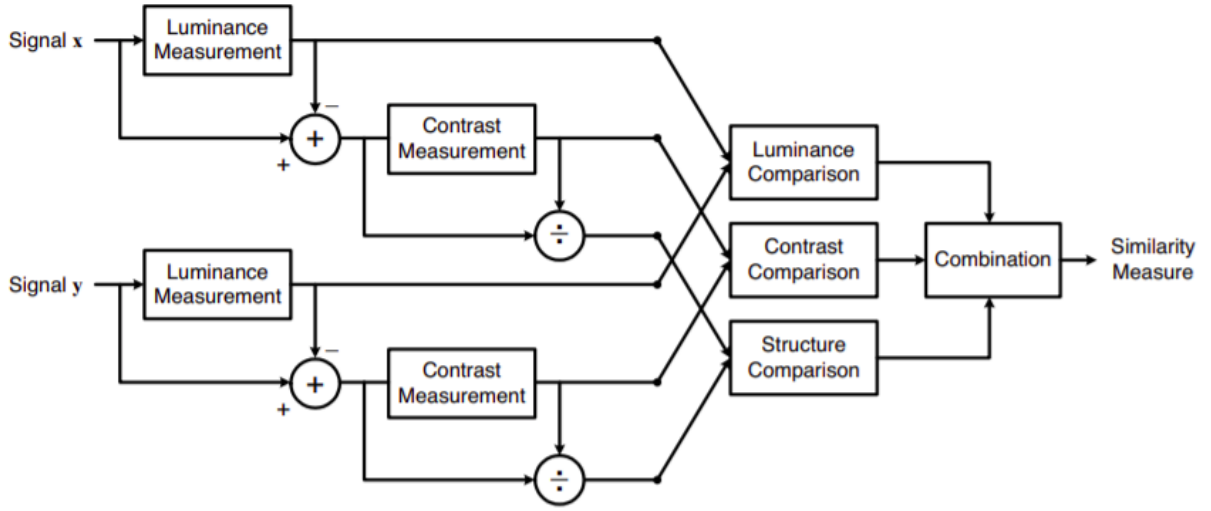


Figura 6 – Diagrama do funcionamento da medição do SSIM Wang et al. (2004).

O SSIM é definido de acordo com a Equação 2.1. Onde  $\mu_x$  e  $\mu_y$  são a média das entradas  $x$  e  $y$ ,  $\sigma_x$  e  $\sigma_y$  são os desvios padrão das entradas  $x$  e  $y$ ,  $\sigma_{xy}$  é a covariância entre os sinais de entrada  $x$  e  $y$  e  $C1$  e  $C2$  são constantes que estabilizam a divisão, no caso de um denominador próximo de zero.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.1)$$



## 3 Trabalhos relacionados

Neste capítulo são descritos os trabalhos relacionados que utilizaram abordagens para o problema de navegação dos deficientes visuais. Pesquisas que utilizaram Redes Neurais Convolucionais com imagens para problemas de navegação tanto para regressão quanto classificação e utilização de transferência de aprendizado. Além disso, são apresentados trabalhos que envolvem construção de conjuntos de dados voltados para aplicações de navegação.

### 3.1 Estado da arte para navegação dos deficientes visuais

Muitos estudos são realizados utilizando dispositivos vestíveis para que deficientes visuais possam desfrutar de uma melhor qualidade de vida. Para auxiliar a navegação de deficientes visuais durante uma caminhada foram realizados estudos como em Lesecq et al. (2017) que propõem uma arquitetura com um conjunto de sensores (LIDAR, ultrassom e radar UWB RF) e sensores inerciais (Acelerômetro e giroscópio) que realizam detecção espacial para ser acoplado em uma bengala guia como dispositivo vestível considerando baixo custo e consumo de energia. A escolha dos sensores foi baseada na capacidade do sistema em detectar grande variedade de obstáculos em diferentes condições ambientais e situações particulares. A arquitetura apresentada utiliza somente as informações dos sensores como meio de navegação e não utiliza aprendizado de máquina.

É utilizado em Wang et al. (2018) uma técnica de segmentação semântica com aprendizado profundo para detectar e classificar múltiplos obstáculos nas imagens RGB-D capturadas a partir de um dispositivo vestível com uma câmera. As imagens foram treinadas utilizando as arquiteturas SegNet, ENet e ERFNet. Os autores utilizam um notebook equipado com GPU como plataforma de processamento. A solução apresentada faz uma escala em função da distância de cada objeto a fonte detetora e informa ao usuário essa aproximação. A abordagem apresentada requer muito poder computacional fazendo com que uma plataforma de alto desempenho com grandes dimensões seja utilizada e poucos objetos são identificáveis.

Em Mancini, Frontoni e Zingaretti (2018) é proposto um dispositivo vestível para orientar deficientes visuais a realizarem caminhadas e corridas utilizando visão computacional. O dispositivo é equipado com uma câmera monocular. Técnicas de processamento digital de imagem foram utilizadas para estabilização de imagem e detecção das regiões de interesse. Detecção de linhas são utilizadas para orientar o deficiente visual a manter-se dentro dos limites da região de caminhada. No erro de distância entre o usuário e a linha detectada foi aplicado em uma lógica fuzzy para estimar a intensidade de giro que o usuário

deve fazer para realizar uma curva. Essa abordagem cria a restrição de haver linhas no trajeto para que funcione. O conjunto de dados utilizado não foi disponibilizado.

Os autores Cordeiro e Pedrino (2018) se propõem a detectar colisões em imagens com informações de profundidade obtidas do sensor Kinect. Algumas técnicas de PDI foram usadas para extrair informações de distância em colisões de objetos estáticos ou dinâmicas. Vários modelos de classificadores foram avaliados e o MLP obteve o melhor resultado obtendo 100% de acurácia. A abordagem foi testada em ambientes internos e com poucos objetos para colisões embora muitos tipos de colisões tenham sido considerados.

Alguns trabalhos utilizam reconhecimento de objetos para auxiliar na navegação de deficientes visuais como foi explorado em Ye e Qian (2018) que apresenta um método de reconhecimento de objetos 3D como paredes, escadas, mesas etc em um dispositivo vestível. O método utiliza sinais de um módulo de captura 3D para realizar treinamentos em Modelo de Mistura Gaussiana (GMM). Os autores não fazem nenhuma menção a utilização do método em ambientes externos.

Diversas abordagens para navegação foram exploradas como em Elmannai e Elleithy (2018) que utiliza logica de controle Fuzzy para indicar desvios de obstáculos. Os autores utilizaram imagens de ambientes internos (escritórios) e externos (parques) para detectar somente colisões. Trabalhos que utilizam câmeras nos óculos para detectar desvios são explorados em Jiang et al. (2019). Os autores utilizaram imagens de ambientes urbanos que tiveram objetos marcados para ajudar a navegação. Sensores de Visão Dinâmica (DVS) com abordagem semelhante e uso de computação em nuvem foi explorado em Everding et al. (2016).

## 3.2 Estado da arte para CNN com aplicações em navegações com imagens para regressão e classificação

Outros tipos de dispositivos vestíveis foram propostos a fim de melhorar a navegação como é mostrado em Mattoccia e Macrì (2015) que é proposta uma abordagem de dispositivo vestível utilizando um óculos 3D para auxiliar a navegação de deficientes visuais para desviar de obstáculos. O dispositivo utiliza imagens com mapas de profundidade para identificar as regiões de interesse e então notifica ao usuário a orientação de desvio. Foi utilizado a arquitetura de CNN LeNet conectada com uma rede MLP para classificar. Por limitações de tempo de classificação, foi adotada a estratégia de reconhecer e retornar apenas o obstáculo mais próximo primeiro e evitar que tenha excesso de informações para o usuário.

Em Loquercio et al. (2018) foi explorado uma arquitetura de CNN para navegação de drones em ambientes externos. Foi utilizado um conjunto de dados próprio com informações

de ângulo e colisões. A Dronet foi uma modificação da ResNet8 com modificação na última camada completamente conectada onde é treinado uma rede de regressão, para os desvios, e outra com classificação binária, para colisões. A partir das inferências de regressão e classificação realiza-se uma conversão para sinais de movimento para os lados onde o Drone faz desvios ou paradas. O conjunto de dados é utilizado apenas em ruas e sua classificação de parada possui apenas duas classes. O conjunto de dados é disponibilizado, porém não representa uma modelagem do problema que este trabalho visou solucionar.

Trabalhos para auxiliar a navegação de deficientes visuais em ambientes externos também são explorados como em Breve e Fischer (2020) que utiliza algumas arquiteturas de CNN para realizar treinamentos e avaliar os modelos para classificar diversos tipos de pisos. O conjunto de dados utilizado é composto por imagens de pisos com diferentes características. O classificador treinado foi binário (piso livre e piso com obstrução). Além disso, foi utilizada uma abordagem de transferência de aprendizado para melhorar a taxa de acurácia do modelo. Na pesquisa os autores não exploraram uma abordagem multiclases.

### 3.3 Estado da arte para transferência de aprendizado

Muitas pesquisas foram realizadas avaliando CNN com transferência de aprendizado. Parte desses trabalhos foram explorados o uso de modelos pré-treinados em conjunto de dados com poucas amostras para melhorar a generalização dos modelos como foi investigado em Schilling et al. (2017) que utiliza modelo pré-treinado com VGG16 para classificação de terrenos para robôs autônomos e em Ahn et al. (2019) que utiliza CNN com transferência de aprendizado para criar um extrator de atributos que transforma o conjunto de características da rede pré-treinada com ImageNet para uma representação no domínio de imagens médicas.

Para auxiliar a navegação de deficientes na identificação de caminhos foi utilizado em Saleh et al. (2017) uma Rede Totalmente Convolutiva baseada na arquitetura da VGG16. É utilizado uma abordagem de segmentação semântica em diferentes tipos de ambientes para localizar a área para caminhos disponíveis para navegação. Eles adicionaram camadas na VGG16 pré-treinadas específicas para segmentação. Os autores não mencionam como a detecção do caminho pode ser traduzida para orientar o deficiente visual durante uma caminhada.

Em Monteiro et al. (2017) os autores utilizaram uma abordagem para reconhecimento de atividades do cão guia como maneira de auxiliar o deficiente visual. São utilizadas CNN para avaliar os modelos de classificação dessa atividades. Foram explorados pesos pré-treinados da ImageNet nas redes AlexNet, GoogleLeNet e combinação delas.

### 3.4 Estado da arte para conjuntos de dados de navegação

Em Park et al. (2020) os autores propõem um conjunto de dados composto por imagens que auxiliam pessoas com deficiência a navegar por calçadas. Foi avaliada a detecção de objetos, segmentação semântica e visão estéreo. O dataset foi coletado a partir da perspectiva de cadeirantes. Foram utilizados um smartphone e uma câmera ZED para coletar as imagens. Redes de diferentes propósitos foram utilizadas como YOLO, RetinaNet e Faster R-CNN para detecção de objetos e segmentação semântica e PSMNet para visão estéreo. O dataset apresentado é apenas para detecção de objetos e não foi mencionado como isso ajuda a navegação do deficiente.

Em Ahmed e Yeasin (2017) os autores utilizam as redes VGG16, ResNet e GoogleNet em um dataset de calçadas para classificação em uma Raspberry Pi 3 avaliando parâmetros como consumo de energia e quantidade mínima de parâmetros para a rede continuar mantendo a acurácia em sua forma original. Foram definidas 5 classes em que cada classe contém características de construção, multidões, buracos em calçadas, pessoas em bicicletas e pessoas com animais de estimação.

Em Weiss et al. (2019) os autores propõem um sistema de navegação que utiliza aprendizado por reforço profundo que utiliza imagens de calçadas para guiar deficientes visuais. O conjunto de dados foi construído utilizando imagens capturadas em 360°. A câmera utilizada foi a Vuze+. Os dados foram rotulados em função do nome das ruas e número das casas. O método proposto utiliza um sistema de recompensa em função das ações realizadas pelo pedestre. O sistema produz uma saída em 5 categorias com desvios leves para a esquerda e direita, desvios fechados para a esquerda e direita e para frente. O estudo apresenta a limitação de oferecer apenas desvios para a esquerda e direita como graus liberdade de movimento durante a navegação.

Em Watanabe et al. (2021) foi proposto um sistema de navegação para pessoas que possuem deficiência motora e utilizam cadeira de rodas como meio de navegação. O método consistiu em utilizar um acelerômetro fixado na cadeira de rodas e um smartphone coletando as imagens das superfícies do chão a sua frente. As imagens foram utilizadas como guia para rotular os sinais de aceleração em 4 classes. Foi utilizado CNN como extrator de características desses sinais e utilizados como entrada para os classificadores MLP e SVM. Os autores apresentam uma abordagem com uma ideia semelhante com este trabalho, porém utilizam apenas o sinal de aceleração. Além disso, a informação de navegação fornecida diz respeito a identificação da superfície da navegação e não fornece informações para tomadas de decisão relacionadas diretamente a navegação.

Em Flores e Manduchi (2018) os autores propõem um conjunto de dados contendo informações inerciais de caminhadas de pessoas que possuem deficiência visual com diferentes modos de navegação dentro de ambientes internos. Foram acoplados dois sensores,

um em cada pé do participante, que monitoram as acelerações dos passos. Além de criar o conjunto de dados o trabalho buscou mostrar as variações de navegação entre diferentes maneiras de caminhadas como navegação com bengala e cão guia e como essas maneiras influenciam na velocidade e desvios. O conjunto de dados não foi avaliado em algoritmos de classificação e não foi apresentado nenhum modelo sobre como esses dados podem ser utilizados para auxiliar na navegação dos deficientes.

## 3.5 Considerações

A maioria dos trabalhos exploram uso de sensores de distância, segmentação de caminhos e reconhecimento de objetos. Diferente das pesquisas apresentadas, este trabalho explora o conjunto de dados em uma solução com mais graus de liberdade para navegação que inclui desvios e paradas utilizando modelos de classificação. É importante destacar que a maioria dos conjuntos de dados não foram disponibilizados pelos autores dos trabalhos apresentados ou se foram o conjunto de dados não consegue representar a tarefa que foi apresentada como solução para o problema de orientação dos deficientes visuais.



## 4 Metodologia

Neste capítulo estão descritos as técnicas e experimentos utilizados no desenvolvimento desta pesquisa. Inicialmente será apresentado a descrição do conjunto de dados. Em seguida, são apresentadas as particularidades dos conjuntos de dados de calçada e corredores. Posteriormente, é apresentado o método de extração das imagens similares por classe e entre classes no conjunto de dados. Por fim, o processo de classificação e as métricas de avaliação serão descritos.

### 4.1 Conjunto de dados

#### 4.1.1 Método de captura das imagens

O Conjunto de Dados para navegação com desvios de obstáculos é uma das contribuições deste trabalho. Todas as imagens possuem resolução de 2160x3840 pixels. As imagens foram capturadas a partir da câmera de um smartphone acoplado em um cinto de 4 pontos fixado na tórax do pedestre com gravação de vídeo em 30 quadros por segundo. Uma representação de como o smartphone estava preso no cinto de quatro pontos é mostrado na Figura 7. As imagens possuem a perspectiva de visão do que está a frente de quem caminha e foram rotuladas em função das decisões durante a caminhada. As decisões são baseadas em desvios de obstáculos e intenções de parada. Além disso, o modo de caminhar foi realizado com características de caminhadas lentas para que fosse parecido com o modo de caminhar do grupo que foi embasado o problema deste trabalho.

O conjunto de dados foi dividido em dois grupos. Um para ambientes externos, contendo imagens de calçadas, e outro para ambientes internos, contendo imagens de corredores. Os ambientes externos possuíam iluminação natural e os internos possuíam a iluminação artificial das construções. Todas as imagens foram capturadas durante o período diurno. Ambos os ambientes possuíam fluxo dinâmico de pessoas e objetos estáticos no trajeto. Após a aquisição, é aplicado redimensionamento em todas as imagens para as dimensões de entrada das respectivas CNN's utilizadas neste trabalho.

#### 4.1.2 Rotulação das imagens

Após a aquisição das imagens foram criados os conjuntos de imagens de calçadas e corredores. Eles foram rotulados em função das intensidades das velocidades frontais e laterais. As estimativas das velocidades foram realizadas a partir de estudo prévio da variação dessa grandeza durante caminhadas com a utilização de uma Unidade de Medida Inercial (UMI) do sensor MPU6050. A Fig. 8 mostra a representação dessas velocidades

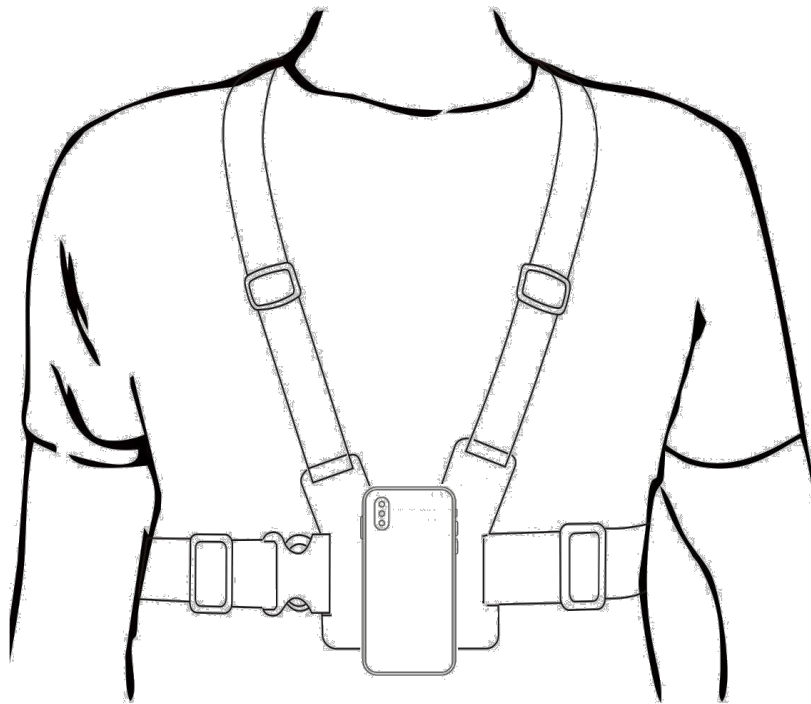


Figura 7 – Representação da maneira como o dispositivo de captura foi acoplado ao corpo. Um Smartphone foi preso a um cinto de quatro pontos que foi vestido no corpo do pedestre.

durante a caminhada. O eixo x representa as velocidades frontais e o eixo y representa as velocidades laterais. Existem outros movimentos relacionados a caminhada, mas neste trabalho foram utilizados somente as velocidades dos eixos x e y.

O processo de rotulação deu-se de modo automático. Cada imagem possui um rótulo de velocidade lateral e um rótulo de velocidade frontal. A velocidade foi utilizada em *ms*. Na sequência de captura, os quadros foram extraídos e as amostras selecionadas com espaçamento de 3 quadros entre elas para compor o conjunto. Para cada subconjunto da sequência de amostras que representava um determinado tipo de movimento foi realizado a rotulação automática a partir de parâmetros de entrada que representavam o movimento indicado no subconjunto selecionado. Esse procedimento foi utilizado em todas as sequências capturadas.

A partir da rotulagem, as velocidades foram discretizadas entre 0 e 1 em cada conjunto de dados. Os subconjuntos foram categorizados em 5 classes. A definição de cada classe deu-se em função da divisão das velocidades discretizadas em 5 partes iguais de modo que cada classe representasse um determinado tipo de movimento. A quantidade de classes pode ser aumentada ou reduzida em função da graduação do movimento que se deseja alcançar, ou seja, mais classes significa movimentos com maior precisão, menos classes, menor precisão. Para o problema estudado nesta pesquisa, uma graduação de

escala de 5 classes foi utilizada para representar os movimentos. O problema de navegação do deficiente visual com a utilização de muitas classes pode desviar muito a capacidade de atenção durante a caminhada e afetar os demais sentidos.

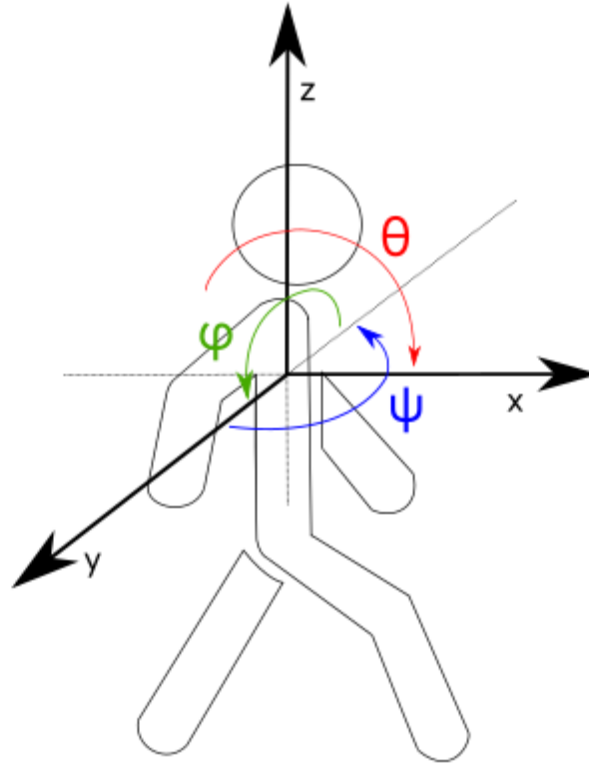


Figura 8 – Representação tridimensional de caminhada com as acelerações e ângulos de Euler correspondentes.

#### 4.1.3 Pós-processamento do conjunto de dados

A rotulação automática pode ocasionar erros durante seu processo, principalmente nas regiões onde existem mudanças de velocidades. Para minimizar esses erros foram propostos métodos de remoção das amostras por similaridade. Foi aplicado o método de remoção por classes. Na remoção por classe, as amostras que pertenciam a mesma classe foram comparadas entre si com a finalidade de remover amostras redundantes.

Com o conjunto de dados resultante da remoção por classe foram realizados treinamentos com as CNNs. A partir desses treinamentos foram obtidos as matrizes de confusão e nelas foi aplicado o SSIM nas colunas dos falsos positivos com os verdadeiros positivos com isso foi obtido um limiar que foi utilizado para realizar a remoção entre classes na sequência temporal. Após a realização desse último passo, foi obtido o conjunto de dados final com a minimização dos erros de rotulação originados pela rotação automática.

O conjunto inicial  $C1$  sofreu uma redução na quantidade de amostras, pois foi balanceado em função da classe que possuía menor quantidade de amostras. Após a execução de cada método de remoção por similaridade o conjunto de dados foi reduzido

ainda mais. A remoção de amostras por classes foi aplicado no conjunto  $C1$  e gerou um conjunto  $C2$ , onde  $C1 < C2$ . A remoção de amostras entre classes foi aplicada no conjunto  $C2$  e foi gerado um conjunto  $C3$ , onde  $C2 < C3$ .

#### 4.1.4 Conjunto das velocidades laterais

O subconjunto das velocidades laterais foram categorizados de modo que houvesse uma referência com variações de desvios para a esquerda e para a direita durante a caminhada. A Tabela 1 mostra a categorização deste conjunto e a quantidade de amostras por classe que foram utilizadas em cada conjunto. Para o conjunto das Calçadas foi obtido um total 8150 amostras divididas igualmente para cada classe, resultando em 1630 amostras por classe. Para o conjunto dos corredores o total de amostras foi de 1810 amostras com 362 amostras por classe. Este conjunto possui uma classe que representa um movimento de referência central. A classe sem desvio (Classe 2) foi a escolhida como referência. As demais classes determinam a intensidade de desvio a partir da classe de referência. Os desvios para a esquerda foram representados como classes 0 e 1, onde a classe 0 representa uma intensidade maior de desvio quando comparado a classe de referência e a classe 1, representa uma intensidade de desvio menor do que a classe 0 e maior do que a classe de referência. Os desvios para a direita foram representados pelas classes 3 e 4 com a mesma ideia de intensidade utilizada nos desvios para a esquerda, mas com sentido oposto. A Figura 9 apresenta de amostras para cada classe categorizada.

Tabela 1 – Descrição das classes de 0 a 4 no Conjunto das velocidades laterais e a quantidade de amostras por classe para os conjuntos de calçadas e corredores.

| Classe | Descrição                    | Quantidade de amostras |            |
|--------|------------------------------|------------------------|------------|
|        |                              | Calçadas               | Corredores |
| 0      | Desvie muito para a esquerda | 1630                   | 362        |
| 1      | Desvie pouco para a esquerda | 1630                   | 362        |
| 2      | Sem desvio                   | 1630                   | 362        |
| 3      | Desvio pouco para a direita  | 1630                   | 362        |
| 4      | Desvie muito para a direita  | 1630                   | 362        |

O subconjunto das velocidades laterais foi estendido utilizando o método de espelhamento das imagens. O objetivo do conjunto estendido é aumentar a quantidade de imagens nas classes que possuem as menores quantidades de imagens no conjunto. As classes que possuem mais deficiência de quantidade são as classes 0 e 4. Como representam desvios com maior intensidade eles acontecem com menor frequência uma vez que na maior parte do trajeto houve pouco fluxo de pessoas e obstáculos.

O método escolhido para estender o conjunto de imagens foi o espelhamento horizontal. Após esse processamento o rótulo da imagem foi invertido, dado que após o espelhamento a imagem torna-se da classe oposta a qual foi espelhada, devido a relação



Figura 9 – Exemplo de amostras do conjunto das calçadas para velocidades laterais por classe.

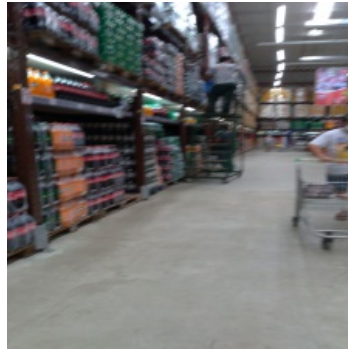
de simetria entre as classes. Com isso cada classe deste conjunto foi composta por 1630 imagens de calçadas e 362 imagens de corredores. A Figura 11 mostra uma exemplo da aplicação do espelhamento com uma amostra da Classe 4.

#### 4.1.5 Conjunto das velocidades frontais

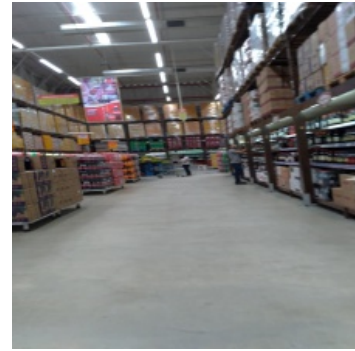
O subconjunto das velocidades frontais permite descrever os movimentos de modo a inferir variações entre paradas e acelerações da caminhada. O movimento varia desde pare (Classe 0) até acelere (Classe 4). Diferente da velocidade lateral, este conjunto não possui um movimento de referência. Além disso, o conjunto das velocidades frontais não possuem uma relação de simetria, logo o método do espelhamento não foi aplicado como maneira de ampliar a quantidade de amostras. A Tabela 2 mostra a categorização deste subconjunto e a quantidade de amostras por classe que foram utilizadas em cada conjunto. Para o conjunto das Calçadas foi obtido um total de 20115 amostras com 4023 amostras por classe. Para o conjunto dos Corredores foi obtido um total de 7070 amostras com 1414 amostras por classe. A Figura 8 mostra a orientação dessa velocidade representada pelo eixo x. A Figura 12 contém amostras de imagens para cada classe neste conjunto.



(a) Desvie muito para a esquerda.



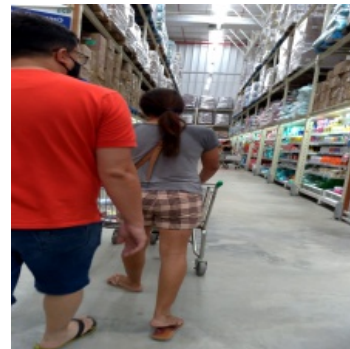
(b) Desvie pouco para a esquerda.



(c) Sem desvio

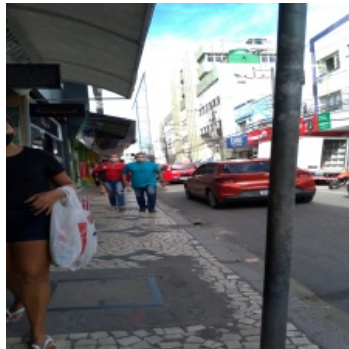


(d) Desvie pouco para a direita.



(e) Desvie muito para a esquerda.

Figura 10 – Exemplo de amostras do conjunto de corredores para velocidades laterais por classe.



(a) Amostra classe 4.



(b) Amostra espelhada da classe 4.

Figura 11 – Amostra de imagem da Figura 11a com seu respectivo espelhamento para a Classe 0 (Figura 11b)

## 4.2 Método de remoção de imagens por similaridade

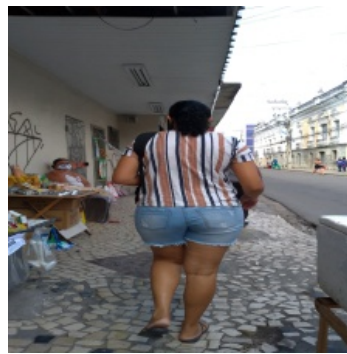
Após a construção do conjunto de dados e rotulação das amostras, foi aplicado o método SSIM para remoção de amostras por similaridade. Esse método foi aplicado em 3 momentos. No primeiro momento foi aplicado a remoção de similaridade por classe e foi obtido um novo conjunto de dados. Em um segundo momento, o modelo treinado com esse

Tabela 2 – Descrição das classes de velocidades frontais e quantidade de amostras por classe para o conjunto das calçadas e corredores.

| Classe | Descrição                 | Quantidade de amostras |            |
|--------|---------------------------|------------------------|------------|
|        |                           | Calçadas               | Corredores |
| 0      | Pare                      | 4023                   | 1414       |
| 1      | Reduza muito a velocidade | 4023                   | 1414       |
| 2      | Reduza pouco a velocidade | 4023                   | 1414       |
| 3      | Mantenha a velocidade     | 4023                   | 1414       |
| 4      | Acelere                   | 4023                   | 1414       |



(a) Pare.



(b) Reduza muito a velocidade.



(c) Reduza pouco a velocidade.



(d) Mantenha a velocidade.



(e) Acelere

Figura 12 – Exemplo de amostras do conjunto de calçadas para velocidades frontais por classe.

novo conjunto foi utilizado para obter a matriz de confusão onde o método de similaridade foi aplicado novamente utilizando uma abordagem entre classes, com isso foi obtido um limiar que foi utilizado na remoção por similaridade na linha temporal. Após a execução de todas essas etapas, um conjunto de dados mais equilibrado foi obtido. A Figura 14 mostra um fluxograma das etapas realizadas para a construção do conjunto de dados final.



Figura 13 – Exemplo de amostras do conjunto de corredores para velocidades frontais por classe.

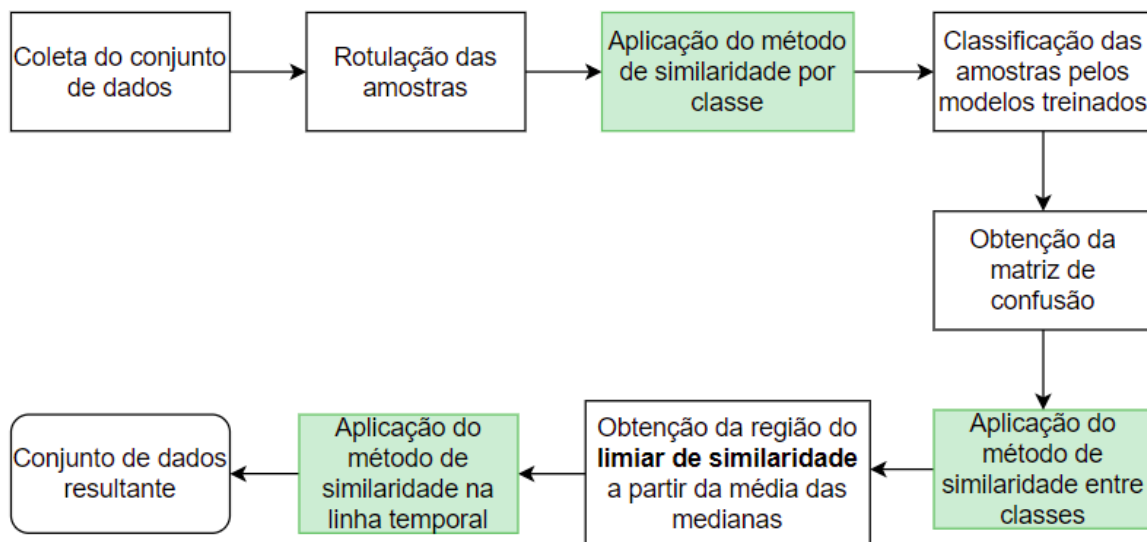


Figura 14 – Fluxograma com as etapas do processo de construção do conjunto de dados com destaque em verde para os momentos de aplicação dos métodos de similaridade.

#### 4.2.1 Similaridade por classe

Para eliminar as imagens redundantes do conjunto de dados foi aplicado um método de remoção de imagens similares dentro das classes. O método consiste em iterar sobre

grupos de amostras das classes, aplicar o algoritmo de SSIM, remover as amostras que são mais similares e criar um novo conjunto de dados  $C^2$  a partir das amostras que não foram removidas. A Figura 15 mostra um fluxograma de como essa remoção é realizada. A entrada do algoritmo é o tamanho do grupo que será analisado e quantas amostras serão removidas no critério de similaridade. O método permite escolher a quantidade de grupos que serão iterados por vez e a quantidade de amostras serão removidas. Os círculos em vermelho representam as amostras que foram removidas no processo por alcançarem os maiores índices de similaridade.

Todas as imagens do grupo são comparadas entre si a partir da combinatória sem repetição e é gerado um conjunto de pares de amostras. Em cada par de amostras o método de SSIM é aplicado e é gerado um resultado de similaridade entre essas amostras. Após a aplicação de SSIM em todos os pares os resultados de similaridade são comparados e é retornado um conjunto com os pares cujos índices de similaridade foram os menores desse grupo. Para cada par de amostras cujo índice de similaridade está entre os maiores, uma amostra é escolhida aleatoriamente para compor o conjunto das amostras que serão descartadas. Caso uma amostra da comparação dos pares já tenha sido descartada a imagem que restou será excluída.

Nos conjuntos utilizados neste trabalho foram realizados testes nos conjuntos resultantes após a aplicação do método com diferentes entradas. Para a velocidade  $x$ , tanto no conjunto das calçadas quanto no conjunto dos corredores, foi utilizada a regra de 1 remoção para grupos de 4 amostras.

#### 4.2.2 Similaridade entre classes na matriz de confusão

A partir da observação de estudos preliminares de classificação com modelos treinados com o conjunto de dado  $C^2$ , algumas amostras de uma classe confundem-se como amostras de classes vizinhas. Para minimizar esse efeito, o método de similaridade entre classes a partir da matriz de confusão foi utilizado. O método consiste em aplicar o SSIM nas amostras classificadas incorretamente com a respectiva classe correta na classe de entrada. Desse modo, tem-se uma relação de similaridade das amostras com a classe para as quais foram classificadas incorretamente. Esse procedimento permite inferir, em média, qual o índice de similaridade das amostras que está havendo dentro das predições incorretas. A Figura 17 mostra um diagrama dessa operação. As amostras incorretas, representadas com círculos vermelhos, foram classificações da classe 2 que não pertencem a esta classe.

Após o fim da execução do método de similaridade na matriz de confusão apenas os pontos da matriz de confusão que obtiveram resultados de similaridade, ou seja, houveram classificações incorretas, são considerados na etapa analítica. Esses resultados são plotados em boxplots para análise do índice de similaridade existente dentro do conjunto de dados.

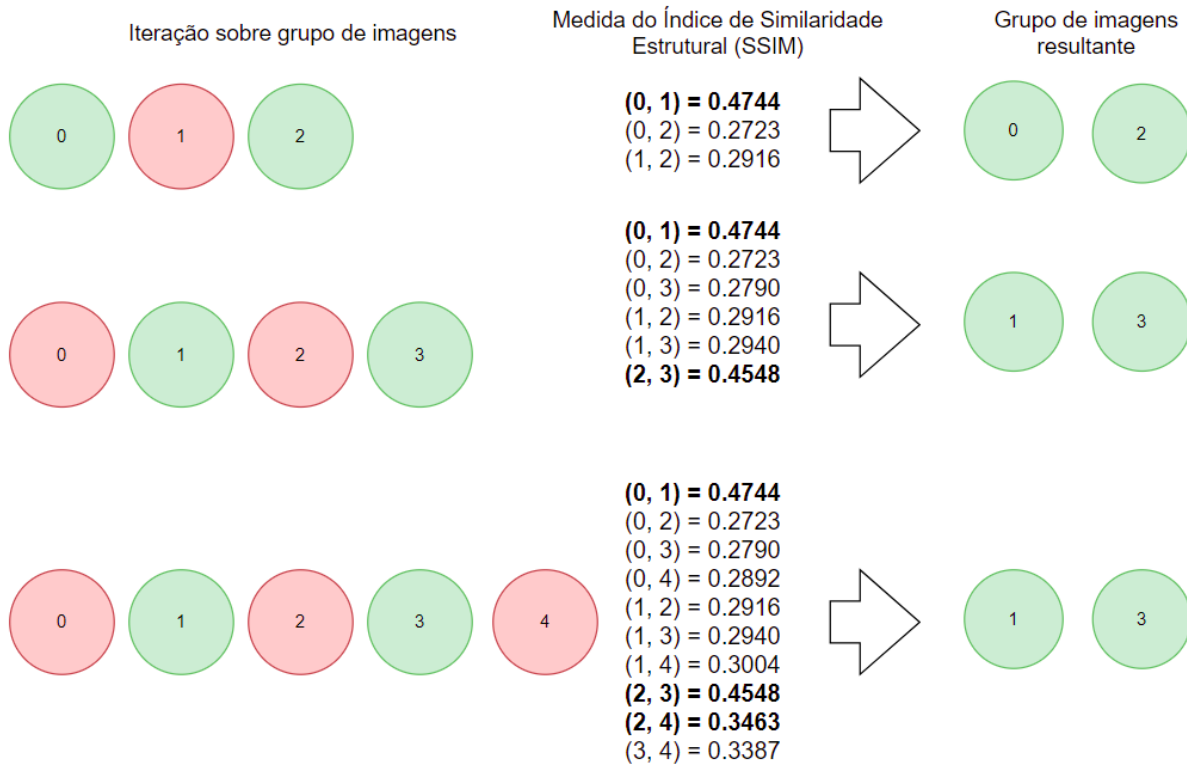


Figura 15 – Representação do fluxo de aplicação do método de remoção de imagens similares dentro da própria classe em 3 grupos de exemplo com tamanhos diferentes em uma iteração do método utilizado.

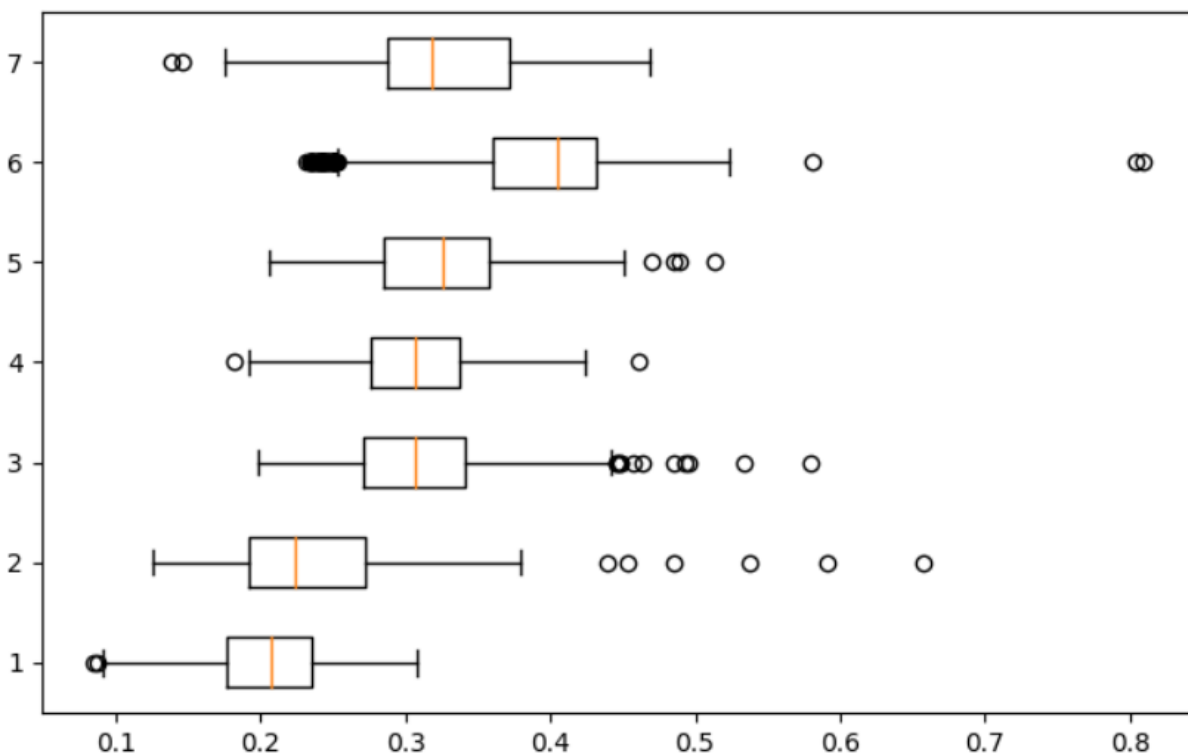


Figura 16 – Boxplot com os resultados de similaridade entre classes na matriz de confusão.

O boxsplot com resultados de similaridade entre classes é apresentado na Figura 16. Foi considerado a média das medianas como limiar para ser utilizado na etapa de busca e remoção por similaridade na linha temporal. A mediana é representada pela linha amarela nos boxsplots. Nesta etapa não é realizada nenhuma remoção de amostras, apenas a descoberta do limiar a ser utilizado na remoção das amostras na sequência temporal.

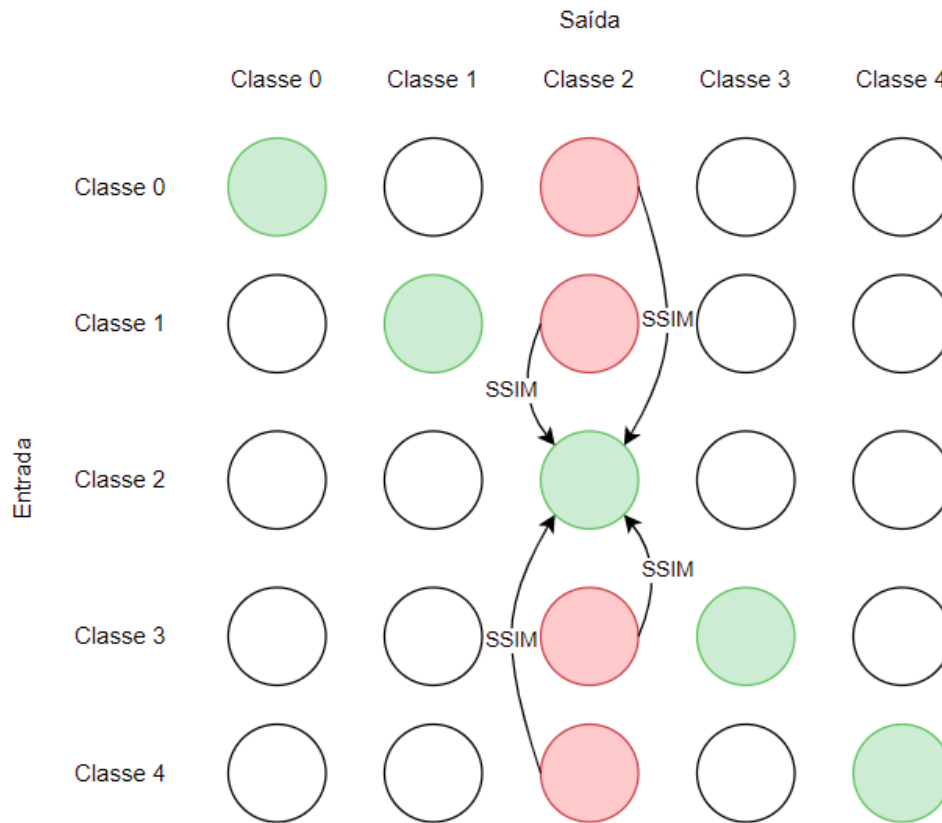


Figura 17 – Representação da aplicação do método de remoção de imagens similares entre classes na matriz de confusão com exemplo de execução na classe 2.

#### 4.2.3 Similaridade entre classes na sequência temporal

Para melhorar mais a qualidade do conjunto de dados foi proposto um método para remoção de similaridade entre classes na sequência temporal de aquisição de imagens. O método de remoção de similaridade foi introduzido na etapa onde os frames estão distribuídos na sequência temporal. Uma amostra central foi escolhida como pivô e nas 5 amostras anteriores e posteriores a essa amostra foi aplicado o método de SSIM. Esse procedimento é realizado em todo a extensão do conjunto de dados. As amostras que atingirem o limiar definido na etapa anterior são removidas do conjunto de dados pois pertencem ao conjunto das amostras que podem ser confundidas por estarem próximas as regiões de transição aplicadas pelo método de rotulação utilizado. A região de limiar para remoção por similaridade entre classes na linha temporal foi definida como a média das medianas das similaridades entre classes da matriz de confusão do modelo que alcançou a

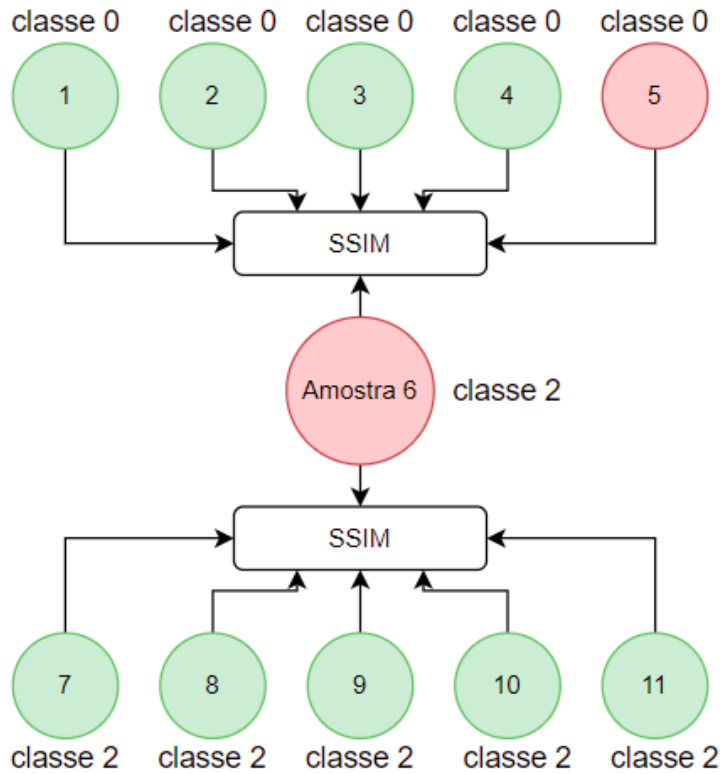


Figura 18 – Representação do método de remoção de imagens similares na linha temporal de captura do conjunto de dados. As amostras estão sequenciadas no tempo e a amostra 6 foi a pivô. Dentro do intervalo da janela escolhida para aplicação do SSIM, a amostra 5 e 6 obtiveram similaridade maior do que o limiar e a amostra 5 foi descartada do conjunto de dados.

melhor acurácia média no conjunto de testes. Uma região de transição é definida como a região do conjunto de dados onde há passagem de uma classe para outra dentro da sequência temporal de captura.

A Figura 18 ilustra uma região de transição e a aplicação do método de remoção por similaridade. A amostra 6 é a amostra pivô e pertence a classe 2. As 5 amostras anteriores pertencem a classe 0 e as 5 posteriores pertencem a classe 2. Com o método proposto a amostra 6 terá um grande índice de similaridade com a amostra 5, logo a amostra 5 será removida do conjunto.

### 4.3 CNNs utilizadas no treinamento e avaliação dos conjuntos de dados

As CNNs utilizadas para treinamento e avaliação do conjunto de dados foram VGG16, ResNet50, SqueezeNet e Dronet. A VGG16 e ResNet50 são redes bastante conhecidas e exploradas na literatura em problemas multiclases e por isso foram escolhidas para serem utilizadas neste trabalho. A Dronet foi utilizada por ser um rede que foi aplicada

em problemas de navegação em ruas e possui semelhanças com o problema de navegação estudado neste trabalho. Além disso, a Dronet possui tamanhos de modelos de treinamento pequenos sendo um forte candidato de modelo a ser embarcado em um computador vestível devido a baixa quantidade de processamento para ser executado e consequentemente consumo menor de energia.

Todas as CNNs foram treinadas utilizando otimização Adam com taxa de aprendizado  $10^{-4}$ , taxa de decaimento exponencial e com tamanho de lote 64. Os conjuntos de dados foram divididos em 70% das amostras utilizadas para treino, 15% para validação e 15% utilizadas para teste. Todas as amostras foram normalizadas utilizando min-max normalization. Todos os treinamentos foram realizados com instâncias g4dn da Amazon Web Service (AWS) e foi utilizado Keras (CHOLLET et al., 2015) como framework para uso das CNNs.

#### 4.3.1 Transferência de aprendizado

Todos os conjunto de dados foram treinados utilizando transferência de aprendizado. Foram utilizados os pesos pré-treinados da ImageNet (DENG et al., 2009) nas redes VGG16 e ResNet50 e para a Dronet foram utilizados os pesos pré-treinados dos próprios autores (LOQUERCIO et al., 2018). Na transferência de aprendizado os pesos das camadas convolucionais foram congelados durante 20 épocas para as redes VGG16 e ResNet50 e em seguida foram descongelados por 80 épocas. Na Dronet foram realizados 60 épocas com os pesos congelados e 240 épocas com os pesos descongelados.

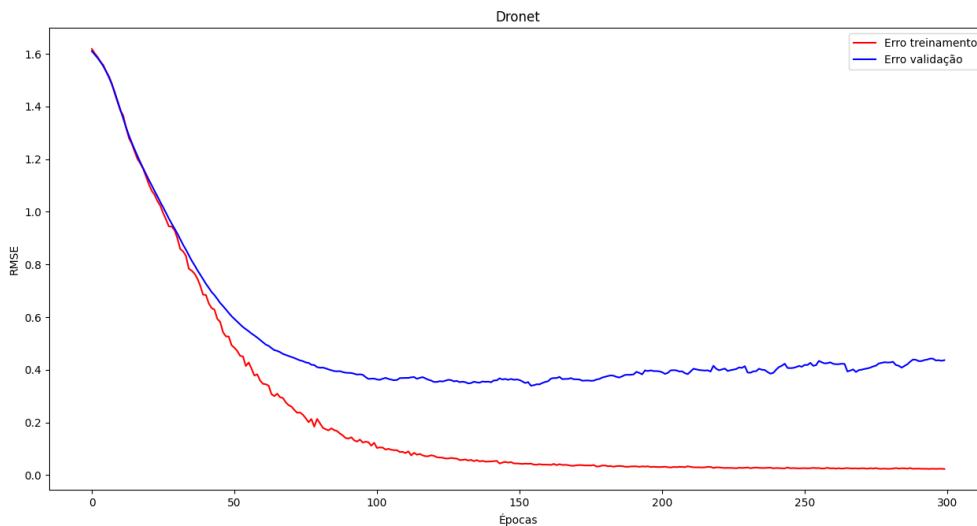


Figura 19 – Erro médio quadrático pelo número de épocas para o treinamento da Dronet. Até a época 60 os pesos foram congelados e e as demais épocas até a época 300 os pesos foram descongelados.

Para as CNNs utilizadas, o congelamento dos pesos se deu até que o RMSE

alcançasse em torno de 20% do total de épocas considerando a época em que foi alcançada a estabilização. Como pode-se observar nas Figuras 19 e 20 a estabilização da Dronet ocorreu na época 300 e na ResNet50 e VGG16 com 20 épocas.

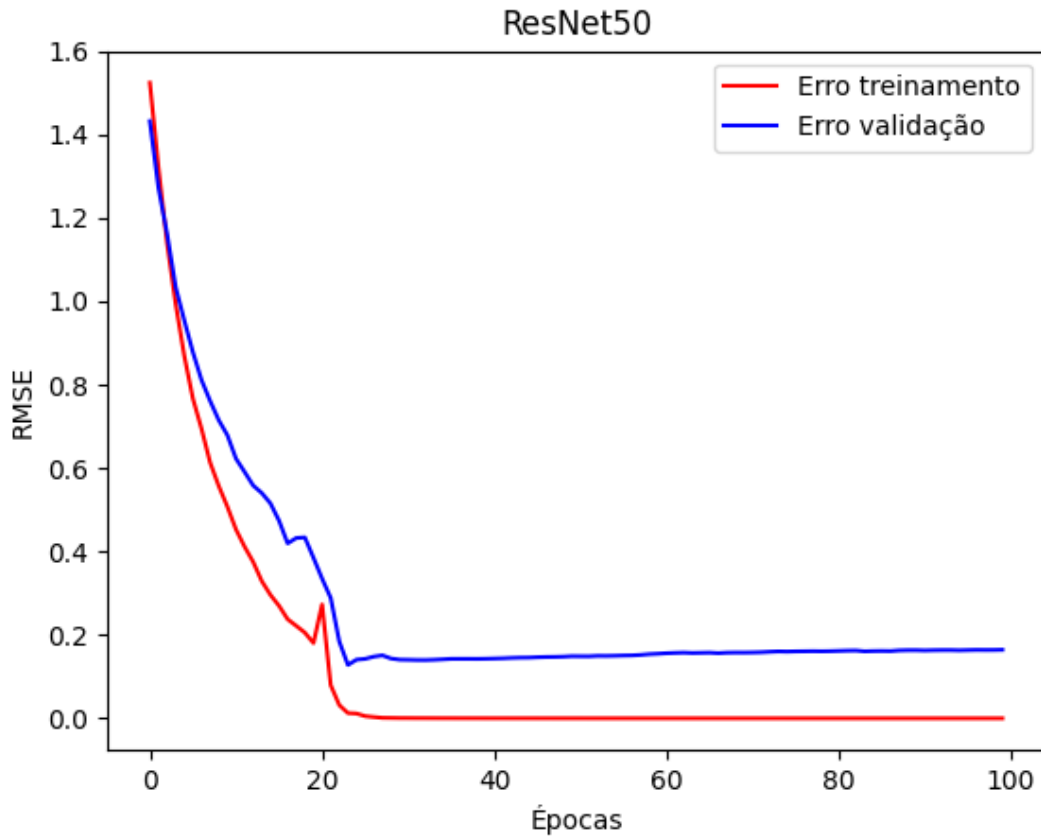


Figura 20 – Erro médio quadrático pelo número de épocas para o treinamento da ResNet50. Até a época 20 os pesos foram congelados e as demais épocas até a época 100 os pesos foram descongelados.

#### 4.3.2 Métricas de avaliação

Para avaliar o desempenho do classificador no problema apresentado foram utilizadas duas métricas. Acurácia e F1. A métrica F1 representa uma média harmônica da sensibilidade e especificidade. Todas essas medidas foram obtidas a partir da matriz de confusão. Para cada combinação de conjunto de dados e CNN foram feitas 10 realizações com conjuntos de dados feitos com distribuições aleatórias para treino, validação e teste. As medidas representam uma média dessas execuções.

## 5 Resultados

Este capítulo descreve os resultados alcançados nos treinamentos dos conjuntos de dados de velocidades frontais e laterais nos conjuntos de calçadas e corredores. Na Seção 5.1, são apresentados os resultados nos conjuntos finais de velocidades frontais e laterais com transferência de aprendizado. Na Seção 5.2 são apresentados os resultados de aprimoramento no conjunto de dados de corredores para velocidades frontais. Finalmente, na Seção 5.3 são apresentadas reflexões sobre o uso desses modelos em um sistema embarcado.

### 5.1 Avaliação da metodologia proposta

A partir do estudo inicial constata-se que a aplicação da técnica de aumento de dados combinada com a transferência de aprendizado melhora significativamente os resultados com as métricas utilizadas. Sendo assim, esse método foi aplicado em todos os experimentos apresentados nesta seção.

#### 5.1.1 Velocidades frontais

Os resultados da Tabela 3 foram obtidos a partir da média de 10 realizações nas CNNs VGG16, ResNet50 e Dronet e constata-se que a VGG16 alcançou uma média de 93,10% de acurácia no conjunto de dados de calçadas. Além disso, possui o menor desvio padrão entre as redes avaliadas. Para o conjunto de dados dos corredores as redes VGG16 e ResNet50 tiveram resultados semelhantes com acurácia média maior que 94% com vantagem de 0,39% na acurácia para a ResNet50. A Dronet obteve os piores resultados em ambos os conjuntos de dados. No conjunto de dados de calçada Dronet alcançou maior diferença quando comparado a rede que obteve os melhores resultados de acurácia no conjunto de dados de corredores.

Para avaliar o desempenho das redes foram destacados as acurácias por classe dos modelos que alcançaram a melhor acurácia dentre as realizações executadas em cada conjunto. Esses resultados são apresentados na Tabela 4. A classe "Acelere" alcançou os melhores resultados em ambos os conjuntos de dados. A Classe "Acelere" por ser uma classe que possui um padrão bem definido, ou seja, as imagens sempre representam caminhos sem obstáculos, possui melhores resultados. Para o conjunto das calçadas as redes VGG16 e Dronet obtiveram padrões semelhantes em termos de acurácia nas classes. Para essas duas redes a classe "Reduza muito a velocidade" alcançou o pior resultado. Esse padrão não se repetiu na rede ResNet50 diferindo nas classes "Reduza pouco a velocidade" e "Reduza

Tabela 3 – Métricas de avaliação das velocidades frontais para calçadas e corredores.

| Conjunto de dados calçadas   |              |             |
|------------------------------|--------------|-------------|
| CNN                          | Acurácia (%) | F1          |
| <b>VGG16</b>                 | 93,10±00,45  | 93,10±00,43 |
| ResNet50                     | 93,25±01,26  | 93,25±01,27 |
| Dronet                       | 89,46±00,57  | 89,46±00,58 |
| Conjunto de dados corredores |              |             |
| <b>VGG16</b>                 | 94,33±00,88  | 94,33±00,88 |
| ResNet50                     | 94,72±00,67  | 94,72±00,66 |
| Dronet                       | 93,77±00,73  | 93,76±00,73 |

muito a velocidade". A acurácia reduzida nessas duas classes ocorre devido a similaridade existente nas transições entre classes.

Para o conjunto dos corredores a classe "Reduza pouco a velocidade" atingiu os piores resultados em todas as redes avaliadas. Por ser uma classe que representa o movimento médio é mais difícil haver um padrão a ser detectado o que justifica a baixa acurácia nessa classe. No conjunto de dados de corredores as classes extremas, ou seja, "Pare" e "Acelere", alcançaram resultados melhores do que no conjunto das calçadas. Além disso, os piores resultados do conjunto dos corredores também atingiu melhores resultados do que o conjunto das calçadas. Esse resultado sugere que o conjunto das calçadas é um problema mais difícil para classificar do que o conjunto dos corredores.

Tabela 4 – Resultados de acurácia por classe para o conjunto das velocidades frontais.

| Classe                    | Acurácia (%) |             |             |
|---------------------------|--------------|-------------|-------------|
|                           | Calçadas     |             |             |
|                           | VGG16        | ResNet50    | Dronet      |
| Pare                      | 93,92±01,40  | 94,44±01,11 | 88,55±01,44 |
| Reduza muito velocidade   | 87,93±01,49  | 88,08±00,98 | 84,02±02,46 |
| Reduza pouco a velocidade | 90,30±01,39  | 89,92±05,01 | 85,70±01,51 |
| Manter velocidade         | 95,57±01,30  | 95,54±01,38 | 92,19±01,20 |
| Acelere                   | 97,79±00,54  | 98,26±00,35 | 96,85±00,58 |
| -                         | Corredores   |             |             |
| Pare                      | 94,58±00,89  | 93,73±00,84 | 94,35±01,00 |
| Reduza muito velocidade   | 93,54±03,14  | 95,66±01,29 | 93,54±01,99 |
| Reduza pouco a velocidade | 91,42±01,75  | 91,38±01,81 | 89,67±02,20 |
| Manter velocidade         | 92,92±02,47  | 93,20±01,94 | 92,79±02,82 |
| Acelere                   | 99,20±00,86  | 99,62±00,28 | 98,49±00,75 |

Foi extraída a matriz de confusão média das 10 realizações executadas com o modelo de CNN que alcançou o melhor resultado. Para o conjunto das calçadas a matriz de confusão média da VGG16 é mostrada na Tabela 5. Observa-se que a maior concentração de predições erradas ocorre nas regiões de vizinhança devido a aproximação temporal que essas regiões apresentam durante a fase de coleta das imagens. Além disso, há uma tendência de redução de erro na medida que ocorre distanciamento da classe correta. Para o conjunto dos corredores a matriz de confusão média é mostrada na Tabela 6. Existe o mesmo padrão de erros na vizinhança observado no conjunto das calçadas, assim como a redução do erro a partir do distanciamento da classe correta.

Tabela 5 – Matriz de confusão média da **VGG16** para velocidades frontais no conjunto das calçadas.

| Classe  |   | Saída (Quantidade de amostras) |              |              |              |              |
|---------|---|--------------------------------|--------------|--------------|--------------|--------------|
|         |   | 0                              | 1            | 2            | 3            | 4            |
| Entrada | 0 | 567,10±08,38                   | 029,90±08,16 | 004,60±02,61 | 001,70±02,10 | 000,50±00,50 |
|         | 1 | 027,80±04,85                   | 530,40±09,09 | 039,50±07,83 | 004,10±02,84 | 001,40±01,68 |
|         | 2 | 001,80±01,40                   | 036,20±05,17 | 545,10±08,52 | 018,30±05,00 | 002,20±01,77 |
|         | 3 | 000,80±01,24                   | 003,80±02,71 | 014,20±04,16 | 577,20±07,78 | 007,90±03,23 |
|         | 4 | 000,60±00,80                   | 002,30±01,67 | 001,70±01,10 | 008,70±03,28 | 590,20±03,18 |

As taxas de acurácia melhores no conjunto dos corredores do que no conjunto das calçadas ocorre porque a diversidade de padrões que compõe as imagens de calçada é maior do que no conjunto dos corredores, devido essa condição as CNN tem dificuldade em acertar uma classificação.

Tabela 6 – Matriz de confusão média da **ResNet50** para velocidades frontais no conjunto dos corredores.

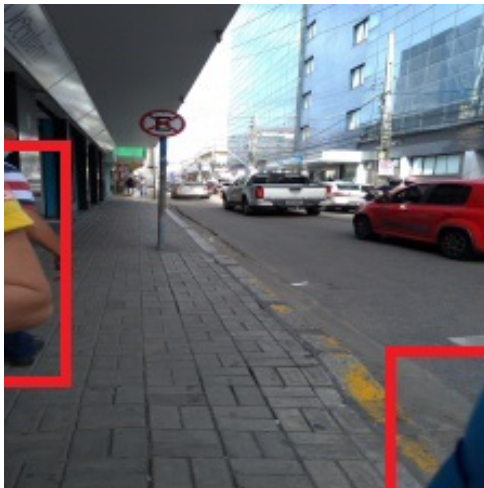
| Classe  |   | Saída (Quantidade de amostras) |              |              |              |              |
|---------|---|--------------------------------|--------------|--------------|--------------|--------------|
|         |   | 0                              | 1            | 2            | 3            | 4            |
| Entrada | 0 | 199,00±01,89                   | 013,00±02,04 | 000,10±00,30 | 000,00±00,00 | 000,20±00,40 |
|         | 1 | 004,40±02,33                   | 202,90±02,73 | 004,70±02,14 | 000,00±00,00 | 000,10±00,30 |
|         | 2 | 000,00±00,00                   | 003,40±02,24 | 194,00±03,82 | 014,90±02,98 | 000,00±00,00 |
|         | 3 | 000,00±00,00                   | 000,10±00,30 | 013,00±03,79 | 197,60±04,12 | 001,30±00,90 |
|         | 4 | 000,00±00,00                   | 000,10±00,30 | 000,00±00,00 | 000,70±00,64 | 211,50±00,80 |

Destacando a classe que alcançou a pior acurácia média em ambos os conjuntos, Temos "Reduza muito a velocidade" com VGG16 para o conjunto das calçadas e "Reduza pouco a velocidade" para o conjunto dos corredores com a ResNet50. Esse dado é mostrado na Tabela 7. Podemos observar que não existe erro nas classes extremas no conjunto dos corredores, já no conjunto das calçadas os erros são distribuídos ao longo das demais classes. Os erros nas classes vizinhas representam 11,14% para o conjunto das calçadas enquanto para o conjunto dos corredores essa taxa é de 8,61%, que é 2,53% menor.

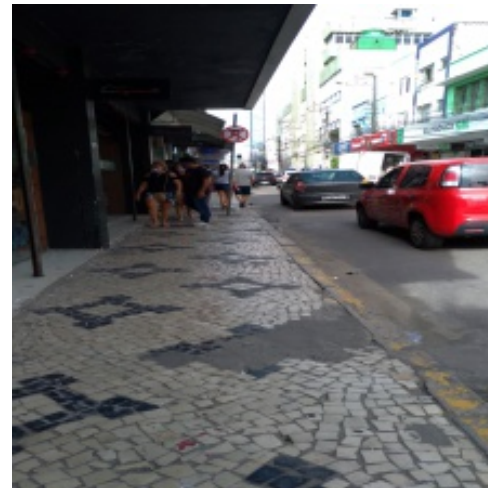
Tabela 7 – Matriz de confusão média da **VGG16** e **ResNet50** para velocidades frontais em taxas proporcionais para a classe Reduza muito a velocidade.

| Classe  |                         | Saída       |             |             |             |             |
|---------|-------------------------|-------------|-------------|-------------|-------------|-------------|
|         |                         | 0           | 1           | 2           | 3           | 4           |
| Entrada | Calçadas VGG16 (%)      |             |             |             |             |             |
|         | 1                       | 04,60±00,80 | 87,93±01,50 | 06,54±01,30 | 00,68±00,47 | 00,23±00,27 |
|         | Corredores ResNet50 (%) |             |             |             |             |             |
|         | 2                       | 00,00±00,00 | 01,60±01,05 | 91,38±01,80 | 07,01±01,40 | 00,00±00,00 |

Observa-se na Tabela 5 que existem erros extremos, ou seja, amostras da classe 0 (Pare) sendo classificadas como classe 4 (Acelere). A Figura 21a foi uma amostra classificada erroneamente como classe 4. A Figura 21a possui padrões semelhantes a uma amostra da classe 4 como é mostrado na Figura 21b. A Figura 21a foi rotulada como Pare (classe 0) devido a uma possíveis colisões marcadas com retângulos vermelhos. Esse mesmo efeito foi observado no conjunto dos corredores.



(a) Amostra da classe "Pare"(Classe 0)



(b) Amostra da classe "Acelere"(Classe 4)

Figura 21 – Exemplos de amostra do conjunto das calçadas. 21a é uma amostra que foi classificada como classe "Acelere"(Classe 4) e 21b é uma amostra que pertence a classe "Acelere".

### 5.1.2 Velocidades laterais

As métricas de acurácia e F1 foram extraídas a partir da matriz de confusão dos conjuntos de calçadas e corredores para as velocidades laterais. Foram aplicados 10 realizações de treinamento em cada CNN. Esses resultados são apresentados na Tabela 8 e representam a média dessas realizações. A VGG16 alcançou os melhores resultados de acurácia em ambas os conjuntos de dados. Para o conjunto das calçadas foi alcançado resultado médio de 90,66% de acurácia. As demais redes obtiveram resultados pelo menos 3% menores de acurácia média. Para o conjunto dos corredores as redes VGG16 e ResNet50

Tabela 8 – Métricas das velocidades laterais para os conjuntos de calçadas e corredores.

| Conjunto de dados calçadas   |              |             |
|------------------------------|--------------|-------------|
| CNN                          | Acurácia (%) | F1          |
| <b>VGG16</b>                 | 90,66±00,96  | 90,65±00,98 |
| ResNet50                     | 87,13±07,51  | 87,08±07,61 |
| Dronet                       | 83,94±00,89  | 83,93±00,89 |
| Conjunto de dados corredores |              |             |
| <b>VGG16</b>                 | 95,58±02,12  | 95,59±02,12 |
| ResNet50                     | 95,22±01,75  | 95,22±01,75 |
| Dronet                       | 89,85±01,68  | 89,85±01,67 |

alcançaram resultados próximos com uma vantagem de 0,36% para a VGG16. Em ambos os conjuntos a Dronet obteve os piores resultados com as métricas avaliadas. Entretanto, a Dronet alcançou os menores desvios padrão o que significa que foi a rede mais estável entre as avaliadas. Novamente, o conjunto dos corredores obteve melhores resultados, isso mostra que possui padrões mais fáceis de serem detectados pelas redes utilizadas.

As acurácias por classe foram extraídas de ambos os conjuntos e de todas as redes para avaliar os desempenhos. A Tabela 8 mostra esses resultados. Para o conjunto das calçadas a rede VGG16 foi a mais estável considerando as acurácias médias de todas as classes com uma diferença de 2,17% entre a menor acurácia que foi obtida na classe Desvie pouco para a esquerda e a maior que foi obtida pela classe Sem desvio. Em seguida, a ResNet50 com diferença de 5,42% e a Dronet com 6,06%. Todas as redes obtiveram os melhores resultados na classe Sem desvio. A ResNet50 obteve os maiores desvios padrão quando comparados as demais redes.

Para o conjunto dos corredores as classes mais distantes da Sem desvio alcançaram as melhores acurácias média em todas as redes avaliadas. A ResNet50 obteve o maior equilíbrio de acurácia média com uma diferença de 2,77%. As demais redes alcançaram 3,59% e 6% respectivamente. Todas as redes possuíram confusão para classificar as classes Sem desvio e Desvie pouco para a direita. Essas duas classes representam as menores taxas de acerto em todos os classificadores.

A Tabela 10 apresenta a matrix de confusão média para o conjunto das calçadas extraída da VGG16. Observa-se que a maior taxa de erros ocorre nas classes vizinhas e o erro é reduzido quanto mais distante da classe correta. Do mesmo modo ocorre no conjunto dos corredores, como é mostrado na Tabela 11.

A Tabela 12 destaca as classes que obtiveram as piores acurácias médias. Para o conjunto das calçadas foi a classe Desvie pouco para a esquerda e para o conjunto dos corredores foi a classe Desvie pouco para a Direita. Observa-se que os erros na vizinhança

Tabela 9 – Resultados de acurácia média por classe para o conjunto das velocidades laterais.

| Class                      | Acurácia (%) |             |             |
|----------------------------|--------------|-------------|-------------|
|                            | Calçadas     |             |             |
|                            | VGG16        | ResNet50    | Dronet      |
| Desvie muito para esquerda | 91,33±02,41  | 89,49±03,80 | 84,36±01,91 |
| Desvie pouco para esquerda | 89,40±04,11  | 86,32±01,50 | 84,90±01,90 |
| Sem desvio                 | 91,57±01,82  | 89,49±06,19 | 87,35±01,89 |
| Desvie pouco para direita  | 91,54±01,87  | 84,07±01,61 | 81,29±01,96 |
| Desvie muito para direita  | 89,44±01,71  | 86,25±05,42 | 81,78±03,37 |
| -                          | Corredores   |             |             |
| Desvie muito para esquerda | 96,86±02,19  | 96,49±02,54 | 94,82±03,18 |
| Desvie pouco para esquerda | 95,20±04,47  | 95,20±03,99 | 89,52±04,30 |
| Sem desvio                 | 94,27±03,14  | 93,72±03,11 | 88,82±03,72 |
| Desvie pouco para direita  | 93,97±04,02  | 94,15±04,93 | 86,18±04,08 |
| Desvie muito para direita  | 97,59±02,62  | 96,48±02,67 | 89,87±04,45 |

Tabela 10 – Matriz de confusão média da **VGG16** para velocidades laterais no conjunto das calçadas.

| Classe  |   | Saída (Quantidade de amostras) |              |              |              |              |
|---------|---|--------------------------------|--------------|--------------|--------------|--------------|
|         |   | 0                              | 1            | 2            | 3            | 4            |
| Entrada | 0 | 223,50±05,85                   | 012,50±03,32 | 002,20±01,46 | 001,70±01,18 | 004,80±02,92 |
|         | 1 | 013,40±04,22                   | 218,60±10,12 | 006,40±03,61 | 003,50±02,90 | 002,60±02,24 |
|         | 2 | 003,20±01,86                   | 006,30±01,67 | 224,00±04,51 | 006,60±02,10 | 004,50±02,01 |
|         | 3 | 001,10±00,70                   | 001,40±00,80 | 006,00±02,48 | 224,10±04,54 | 012,20±05,23 |
|         | 4 | 005,00±02,52                   | 002,10±01,51 | 005,20±02,22 | 013,50±02,87 | 218,60±04,27 |

Tabela 11 – Matriz de confusão média da **VGG16** para velocidades laterais no conjunto dos corredores.

| Classe  |   | Saída (Quantidade de amostras) |              |              |              |              |
|---------|---|--------------------------------|--------------|--------------|--------------|--------------|
|         |   | 0                              | 1            | 2            | 3            | 4            |
| Entrada | 0 | 052,60±01,42                   | 001,10±00,83 | 000,30±00,45 | 000,00±00,00 | 000,30±00,64 |
|         | 1 | 001,80±01,72                   | 052,00±02,79 | 000,10±00,30 | 000,70±00,78 | 000,00±00,00 |
|         | 2 | 000,60±00,66                   | 000,70±00,90 | 051,10±01,75 | 001,60±00,91 | 000,20±00,40 |
|         | 3 | 000,40±00,66                   | 001,00±00,89 | 000,40±00,66 | 051,50±02,24 | 001,50±01,74 |
|         | 4 | 000,30±00,45                   | 000,00±00,00 | 000,30±00,45 | 000,70±01,10 | 052,80±01,53 |

no conjunto dos corredores corresponde a 3,46% enquanto que no conjunto das calçadas essa proporção equivale a 8,09% que é 4,63% maior que no conjunto dos corredores.

Na matriz de confusão da Tabela 10 observa-se as mesmas questões encontradas nas velocidades frontais. As partes críticas são aquelas onde é indicado um desvio para

um lado e é predito para o lado oposto. Para a classe 0 (Desvie muito para a esquerda) obteve-se em média 4,80 amostras classificadas para o oposto. Outra região crítica é a classe Sem desvio e o classificador indicar desvios para os extremos.

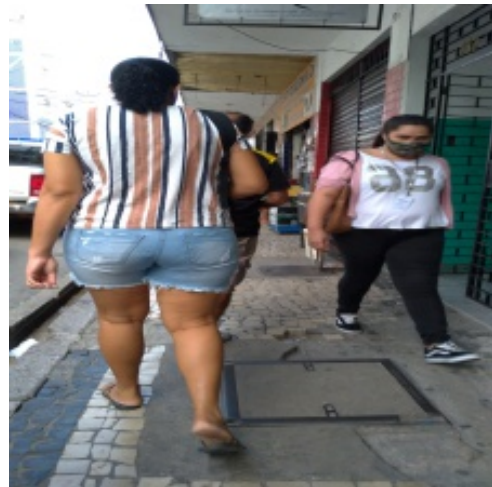
Tabela 12 – Matriz de confusão média da **VGG16** para velocidades laterais em taxas proporcionais para as classes Desvie pouco para esquerda e Desvie pouco para direita.

| Classe  |                      | Saída       |             |             |             |             |
|---------|----------------------|-------------|-------------|-------------|-------------|-------------|
|         |                      | 0           | 1           | 2           | 3           | 4           |
| Entrada | Calçadas VGG16 (%)   |             |             |             |             |             |
|         | 1                    | 05,48±01,72 | 89,40±04,14 | 02,61±01,47 | 01,43±01,18 | 01,06±00,91 |
|         | Corredores VGG16 (%) |             |             |             |             |             |
|         | 3                    | 00,73±01,21 | 01,82±01,63 | 00,73±01,21 | 93,97±04,10 | 02,73±03,18 |

Na Figura 22a é uma amostra que foi rotulada como Desvie muito para a esquerda, mas foi classificada como Desvie muito para a direita. Analisando os frames anteriores e posteriores observa-se que a amostra classificada errada está numa região onde foi realizado um desvio para a esquerda e correção desse desvio imediatamente para a direita. Como existe uma pessoa na frente o classificador entendeu que pertencia a classe de desvio para a direita.



(a) Amostra da classe "Desvie muito para a esquerda" (Classe 0)



(b) Amostra da classe "Desvie muito para a direita" (Classe 4)

Figura 22 – Exemplos de amostra do conjunto das calçadas. 22a é uma amostra que foi classificada como classe 4 e 22b é uma amostra que pertence a classe 4.

Na Tabela 11 também mostra casos de erros extremos no conjunto dos corredores. Extraíndo as amostras que caem nessa condição observa-se um padrão semelhante ao encontrado no conjunto das calçadas. A Figura 23a é de um movimento indicando desvio para a esquerda e a Figura 23b é o frame seguinte a esse desvio e que corresponde a correção do movimento para o centro do corredor. Essas imagens foram muito semelhantes apesar de pertencerem a classes opostas e por conta disso confundiu a classificação.



(a) Amostra da classe "Pare"(Classe 0)



(b) Amostra da classe "Acelere"(Classe 4)

Figura 23 – Exemplos de amostra do conjunto dos corredores. 23a é uma amostra que foi classificada como classe 4 e 23b é uma amostra que pertence a classe 4.

## 5.2 Propostas de aprimoramento do conjunto de dados

Esta seção apresenta os resultados obtidos com a aplicação do método da Medida do Índice de Similaridade Estrutural (SSIM) no conjunto dos corredores para as velocidades frontais.

A partir dos resultados obtidos para as velocidades frontais e laterais constata-se que as as velocidades frontais possuem maiores taxas de erros nas classes vizinhas, logo apresentam mais confusão para classificar do que as velocidades laterais. Desse modo, foi escolhido o conjunto dos corredores para aplicar o método de SSIM.

A modelo da VGG16 no conjunto dos corredores foi escolhido por possuir acurácia média muito próximo da arquitetura que alcançou o melhor resultado e por possuir maior variação de acurácia em suas classes. O modelo com maior acurácia entre as realizações para velocidades frontais foi extraído e aplicado o algoritmo de SSIM na matriz de confusão para avaliar a similaridade entre as amostras que foram classificadas erradas.

A similaridade média entre classes na matriz de confusão é apresentada na Tabela 13. Podemos observar que as amostras confundidas da classe "Reduza pouco a velocidade"(Classe 2) possuem 0,3910 de similaridade com a classe "Manter velocidade"(Classe 3) e 0,3082 com a classe "Reduza muito a velocidade"(Classe 1), o que elege a classe "Reduza pouco a velocidade"como a mais confundida. Observa-se também que para a classe "Manter devlocidade"possui a segunda maior confusão. De fato, ao observar a Tabela 4 constata-se que as classes "Reduza pouco a velocidade"e "Manter velocidade"obtiveram as menores acurácias média. Isso sugere que os erros dessas classes são consequência da semelhança de suas amostras com as classes de distância adjacente.

A partir da Tabela 13 verifica-se que existem 7 resultados de confusão com a

Tabela 13 – Matriz de similaridade média da matriz de confusão para as velocidades frontais no conjunto dos corredores.

| Classe     |   | SSIM Saída  |             |             |             |             |
|------------|---|-------------|-------------|-------------|-------------|-------------|
|            |   | 0           | 1           | 2           | 3           | 4           |
| SSIM Entra | 0 | -           | 0,234±0,053 | 0,000±0,000 | 0,000±0,000 | 0,000±0,000 |
|            | 1 | 0,204±0,041 | -           | 0,304±0,046 | 0,000±0,000 | 0,000±0,000 |
|            | 2 | 0,000±0,000 | 0,308±0,049 | -           | 0,391±0,058 | 0,000±0,000 |
|            | 3 | 0,000±0,000 | 0,000±0,000 | 0,323±0,049 | -           | 0,325±0,052 |
|            | 4 | 0,000±0,000 | 0,000±0,000 | 0,000±0,000 | 0,000±0,000 | -           |

similaridade extraída. Com esses resultados foi gerado um boxsplot que é mostrado na Figura 24. O boxsplot 6 corresponde ao resultado de similaridade 0,39 encontrado na confusão da classe 2 com a classe 3. Observa-se que existem amostras com alto índice de similaridade que se aproximam de 0,8 numa escala de 0 a 1, onde 1 significa imagem completamente similar. É importante notar que existem casos com vários outliers com similaridade superiores a 0,4.

A média da mediana das similaridades, que é destacada na Figura 24 com as linhas amarelas nos gráficos de boxSplot, foi utilizada como região de limiar. A média obtida foi de 0,2896. Após a aplicação do método de remoção das imagens semelhantes entre classes o novo conjunto de dados criado obteve um total de 1386 imagens por classe.

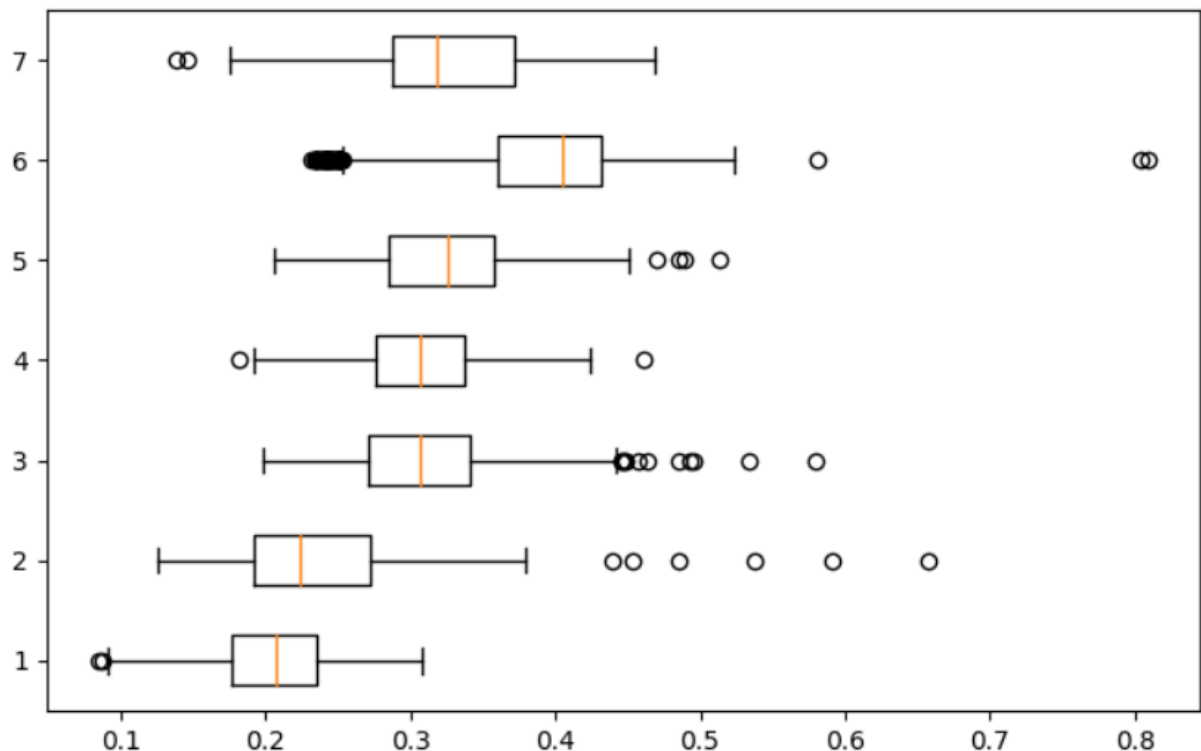


Figura 24 – Boxplot com os resultados de similaridade entre classes na matriz de confusão no conjunto das velocidades frontais para corredores

Tabela 14 – Métricas das velocidades frontais para os conjuntos dos corredores antes e após da aplicação do método de SSIM.

| Conjunto de corredores       |             |             |
|------------------------------|-------------|-------------|
| Sem remoção por similaridade |             |             |
| CNN                          | Acurácia(%) | F1          |
| VGG16                        | 94,33±00,88 | 94,33±00,88 |
| <b>ResNet50</b>              | 94,72±00,67 | 94,72±00,66 |
| <b>Dronet</b>                | 93,77±00,73 | 93,76±00,73 |
| Com remoção por similaridade |             |             |
| VGG16                        | 95,64±00,45 | 95,64±00,45 |
| <b>ResNet50</b>              | 95,85±00,38 | 95,85±00,38 |
| Dronet                       | 94,72±00,63 | 94,71±00,63 |

A Tabela 14 apresenta os resultados de acurácia por CNN para o conjunto dos corredores após a aplicação do método de extração de similaridade. Observa-se que a VGG16 e ResNet50 alcançaram resultados próximos com diferença de acurácia média de 0,21%. Comparando esses resultados com a Tabela 3 observa-se que o conjunto dos corredores sem método de SSIM manteve o padrão onde VGG16 e ResNet50 obtiveram resultados semelhantes. Comparando os resultados de acurácia média da CNN que alcançou os melhores resultados em ambos os conjuntos tem-se que o conjunto com SSIM foi 1,13% melhor do que o sem SSIM. Em ambos os conjuntos, com e sem SSIM, a Dronet foi a que alcançou os piores resultados com as métricas utilizadas.

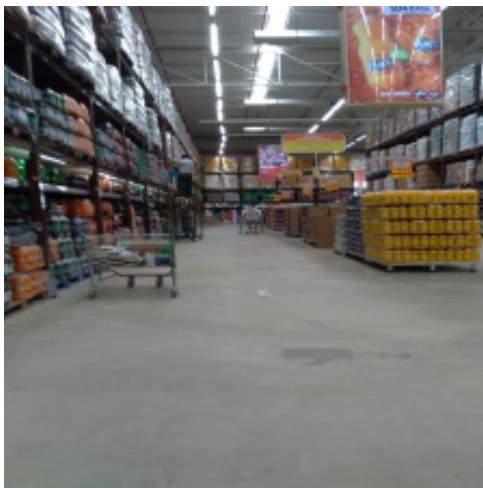
Extraíndo a acurácia média por classe para o conjunto com SSIM, observa-se na Tabela 15 que a classe Reduza pouco a velocidade obteve a pior acurácia média e que a classe Acelere alcançou o melhor resultado em todas as redes avaliadas. Esse resultado foi o mesmo observado no conjunto sem SSIM. Comparando os resultados de acurácia por classe da CNN que alcançaram o melhor resultado de acurácia média tem-se que o conjunto com SSIM possui médias maiores em todas as classes, exceto a classe Acelere onde o conjunto sem SSIM foi 0,06% melhor. No conjunto com SSIM as classes Pare e Manter Velocidade obtiveram uma acurácia média pelo menos 2% melhor do que a sem SSIM.

A Tabela 16 mostra a matriz de confusão média da ResNet50. Nota-se que os erros ocorrem sempre nas vizinhanças em todas as classes, exceto a classe "Reduza muito a velocidade"(Classe 1) onde houve confusão com a classe "Acelere"(Classe 4). A Figura 25a apresenta a amostra que foi classificada como classe "Acelere"(Classe 4). Ela possui as características de uma imagem que pertence a classe "Acelere"(Classe 4), porém foi rotulada como classe "Reduza muito a velocidade"(Classe 1) porque os frames anteriores continham uma possível colisão, como é mostrada na Figura 25b, quando a colisão deixou de estar no

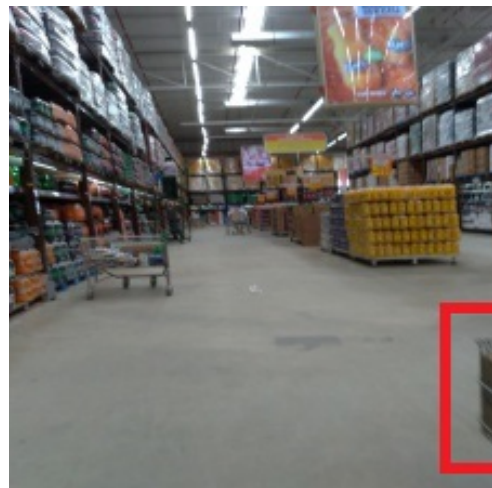
Tabela 15 – Acurácia média por classe para o conjunto dos corredores com SSIM.

| Class                     | Acurácia (%) |             |             |
|---------------------------|--------------|-------------|-------------|
|                           | Corredores   |             |             |
|                           | VGG16        | ResNet50    | Dronet      |
| Pare                      | 96,82±01,34  | 96,20±01,22 | 95,81±01,63 |
| Reduza muito a velocidade | 94,95±01,96  | 95,81±01,80 | 94,71±01,77 |
| Reduza pouco a velocidade | 91,57±01,82  | 92,45±01,72 | 89,66±02,10 |
| Manter velocidade         | 92,59±02,23  | 95,24±01,06 | 94,56±01,56 |
| Acelere                   | 99,32±00,57  | 99,56±00,45 | 98,84±00,57 |

campo de visão da câmera, mas ainda estava ao lado de quem estava caminhando, então o algoritmo a rotulou como classe "Reduza muito a velocidade"(Classe 1).



(a) Amostra da classe "Reduza muito a velocidade"(Classe 1) imediatamente após a passagem de uma possível colisão.



(b) Amostra da classe "Reduza muito a velocidade"(Classe 1) da possível colisão.

Figura 25 – Exemplos de amostra do conjunto dos corredores com SSIM onde houve confusão com a classe "Acelere"(Classe 4).

Comparando a classificações por classe proporcionais a quantidade contida nos conjuntos com e sem SSIM observa-se na Tabela 17 que o conjunto com SSIM apresenta menos confusão nas classes mais distantes em todas as classes. Além disso, os erros nas regiões de vizinhança são menores do que no conjunto sem SSIM. A classe "Reduza muito a velocidade"(Classe 1) apresentou uma redução de 2,33% na vizinhança no conjunto com SSIM. As demais classes reduziram em algum percentual a quantidade de erros nas vizinhanças esquerda e direita, exceto a classe "Reduza pouco a velocidade"(Classe 2) que apresentou aumento de 0,09% na vizinhança esquerda.

Tabela 16 – Matriz de confusão da rede ResNet50 no conjunto de corredores com SSIM para as velocidades frontais.

| Classe  |   | Saída        |              |              |              |              |
|---------|---|--------------|--------------|--------------|--------------|--------------|
|         |   | 0            | 1            | 2            | 3            | 4            |
| Entrada | 0 | 200,10±02,54 | 007,90±02,54 | 000,00±00,00 | 000,00±00,00 | 000,00±00,00 |
|         | 1 | 004,50±02,24 | 199,30±03,74 | 004,10±02,11 | 000,00±00,00 | 000,10±00,30 |
|         | 2 | 000,00±00,00 | 003,00±02,64 | 192,30±03,57 | 012,70±03,00 | 000,00±00,00 |
|         | 3 | 000,00±00,00 | 000,00±00,00 | 009,30±02,36 | 198,10±02,21 | 000,60±00,48 |
|         | 4 | 000,00±00,00 | 000,00±00,00 | 000,00±00,00 | 000,90±00,94 | 207,10±00,94 |

Tabela 17 – Matriz de confusão em porcentagem proporcional a quantidade de amostras dos conjuntos sem e com SSIM.

| Classe  | Saída                            |             |             |             |             |             |
|---------|----------------------------------|-------------|-------------|-------------|-------------|-------------|
|         | -                                | 0           | 1           | 2           | 3           | 4           |
| Entrada | Corredores ResNet50 sem SSIM (%) |             |             |             |             |             |
|         | 0                                | 93,73±00,89 | 06,12±00,96 | 00,04±00,14 | 00,00±00,00 | 00,09±00,18 |
|         | 1                                | 02,07±01,10 | 95,66±01,29 | 02,21±01,01 | 00,00±00,00 | 00,04±00,14 |
|         | 2                                | 00,00±00,00 | 01,60±01,05 | 91,38±01,80 | 07,01±01,40 | 00,00±00,00 |
|         | 3                                | 00,00±00,00 | 00,04±00,14 | 06,13±01,79 | 93,20±01,94 | 00,61±00,42 |
|         | 4                                | 00,00±00,00 | 00,04±00,14 | 00,00±00,00 | 00,33±00,30 | 99,62±00,38 |
|         | Corredores ResNet50 com SSIM (%) |             |             |             |             |             |
|         | 0                                | 96,20±01,22 | 03,79±01,22 | 00,00±00,00 | 00,00±00,00 | 00,00±00,00 |
|         | 1                                | 02,16±01,08 | 95,81±01,80 | 01,97±01,01 | 00,00±00,00 | 00,04±00,14 |
|         | 2                                | 00,00±00,00 | 01,44±01,27 | 92,45±01,72 | 06,10±01,44 | 00,00±00,00 |
|         | 3                                | 00,00±00,00 | 00,00±00,00 | 04,47±01,13 | 95,24±01,06 | 00,28±00,23 |
|         | 4                                | 00,00±00,00 | 00,00±00,00 | 00,00±00,00 | 00,43±00,45 | 99,56±00,45 |

### 5.3 Discussão de resultados

Para o conjunto das velocidades laterais o método de aumento do conjunto de dados conseguiu melhorar a taxa de acertos em 3,3% em relação ao conjunto sem a aplicação do método.

A VGG16 alcançou os melhores resultados de acurácia média em quase todos os conjuntos de dados, exceto no conjunto de corredores para velocidades frontais, onde a ResNet50 foi um pouco melhor em 0,39% em relação a VGG16. Para as velocidades frontais a VGG16 alcançou acurácia média de pelo menos 93% em ambos os conjuntos de dados. Para as velocidades laterais a VGG16 alcançou resultados de acurácia média de pelo menos 90%.

## 5.4 Reflexões sobre a aplicação embarcada

Pode-se observar na Tabela 14 que as arquiteturas VGG16 e ResNet50 são fortes candidatos de modelos para serem embarcados por terem alcançado os melhores resultados de acurácia. Quando é observado o resultado de acurácia por classe é possível perceber algumas diferenças nas taxas de acerto. Para o problema de navegação de deficientes visuais as classes, no conjunto de velocidades frontais, que possuem mais importância são as de paradas e reduções, porque essas decisões reduzem as chances de haver acidentes. Considerando essa premissa, a VGG16 possui resultados um pouco melhores do que as demais redes. Além disso, por ser uma rede que possui menor profundidade seu tempo de classificação é menor. No entanto, a VGG16 possui um arquivo de peso 524MB que deve ser alocado em memória para realizar as predições o que é uma restrição. Já a ResNet50, por ser uma rede mais profunda seu tempo de resposta será maior do que a VGG16, porém o arquivo de peso a ser alocado em memória é de 93MB, que é cerca de 5,6 vezes menor.



## 6 Conclusões

Este trabalho apresentou um estudo sobre o problema de navegação de deficientes visuais durante uma caminhada propondo dois novos conjuntos de dados que relacionam imagens com os movimentos do pedestre para evitar choque com obstáculos. Um dos conjuntos representa ambientes externos (calçadas urbanas) e outro representa ambientes internos (corredores de lojas). Cada conjunto explora dois domínios de movimentos. Velocidade frontais, que determinam acelerações e paradas, e velocidades laterais, que englobam movimentos de desvios para a esquerda e para a direita. Além disso, esses conjuntos foram utilizados para treino e teste nas arquiteturas CNN VGG16, ResNet50 e Dronet. O estudo mostrou que aplicar transferência de aprendizado em ambos os conjuntos aumentou suas taxas de acurácia.

O conjunto de dados dos corredores apresenta resultados de acurácia média melhores do que os conjuntos de calçadas. Tanto nas velocidades frontais quanto nas laterais. Além disso, as matrizes de confusão do conjunto dos corredores possuem menos erros extremos do que o de calçadas. Isso sugere que o conjunto dos corredores é um problema mais fácil para as redes classificarem do que o conjunto das calçadas. As classificações erradas que ocorrem dentro do sistema de navegação corrigem-se com os frames seguintes por ser um sistema dinâmico, logo é tolerável que eles ocorram.

A atribuição de rótulos às imagens foi obtida de forma automática, a partir das informações dos sensores inerciais no momento da captura das imagens. A fim de melhorar a marcação das amostras, foi proposto um método de remoção das imagens por similaridade, que foi aplicado nas velocidades frontais do conjunto de dados dos corredores. Tal método reduziu as classificações incorretas na matriz de confusão e aumentou a acurácia média em 1,13% quando comparado ao conjunto sem esse tratamento. Esse resultado se mostra promissor para melhorar conjuntos de dados que utilizam técnicas de capturas e rotulações seguindo uma sequência temporal.

Os conjuntos de dados propostos têm relevância em aplicações que visam a auxiliar sistemas de navegação para deficientes visuais e aplicações de robôs autônomos, pois permitem classificar, a partir de imagens, os principais movimentos relacionados a evitar colisão com obstáculos, como paradas, acelerações e desvios.

Visando à sua utilização em robôs ou em dispositivos vestíveis, as arquiteturas de CNN utilizadas devem sofrer otimizações no consumo de memória e tempo de classificação. A melhoria no tempo de classificação permite aumentar a taxa de classificações por segundo. Essa melhoria é necessária para que a navegação possa tomar decisões em pouco tempo, portanto, mais seguras.

Como trabalhos futuros sugere-se:

- Avaliar o conjunto de dados em outras outras redes neurais.
- Embarcar classificador em um dispositivo vestível.
- Avaliar custos computacionais na plataforma embarcada.
- Realizar testes de navegação em campo.

# Referências

- AHMED, F.; YEASIN, M. Optimization and evaluation of deep architectures for ambient awareness on a sidewalk. In: . IEEE, 2017. Disponível em: <<https://doi.org/10.1109/ijcnn.2017.7966186>>. Citado na página 38.
- AHN, E. et al. Unsupervised deep transfer feature learning for medical image classification. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019. Disponível em: <<https://doi.org/10.1109/isbi.2019.8759275>>. Citado na página 37.
- ALADREN, A. et al. Navigation assistance for the visually impaired using RGB-d sensor with range expansion. *IEEE Systems Journal*, Institute of Electrical and Electronics Engineers (IEEE), v. 10, n. 3, p. 922–932, set. 2016. Disponível em: <<https://doi.org/10.1109/jsyst.2014.2320639>>. Citado na página 21.
- BARONTINI, F. et al. Integrating wearable haptics and obstacle avoidance for the visually impaired in indoor navigation: A user-centered approach. *IEEE Transactions on Haptics*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2020. Disponível em: <<https://doi.org/10.1109/toh.2020.2996748>>. Citado 2 vezes nas páginas 21 e 22.
- BREVE, F.; FISCHER, C. N. Visually impaired aid using convolutional neural networks, transfer learning, and particle competition and cooperation. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020. Disponível em: <<https://doi.org/10.1109/ijcnn48605.2020.9207606>>. Citado na página 37.
- CHENG, R. et al. Visual localization of key positions for visually impaired people. In: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/icpr.2018.8545141>>. Citado na página 21.
- CHOLLET, F. et al. *Keras*. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>. Citado na página 53.
- CIREŞAN, D. et al. Multi-column deep neural network for traffic sign classification. *Neural Networks*, Elsevier BV, v. 32, p. 333–338, ago. 2012. Disponível em: <<https://doi.org/10.1016/j.neunet.2012.02.023>>. Citado na página 26.
- CORDEIRO, N. H.; PEDRINO, E. C. An architecture for collision risk prediction for visually impaired people. In: *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/sibgrapi.2018.00046>>. Citado na página 36.
- DENG, J. et al. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009. Disponível em: <<https://doi.org/10.1109/cvpr.2009.5206848>>. Citado 2 vezes nas páginas 31 e 53.
- ELMANNAI, W.; ELLEITHY, K. Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions. *Sensors*, MDPI AG, v. 17, n. 3, p. 565, mar. 2017. Disponível em: <<https://doi.org/10.3390/s17030565>>. Citado na página 25.

ELMANNAI, W. M.; ELLEITHY, K. M. A novel obstacle avoidance system for guiding the visually impaired through the use of fuzzy control logic. In: *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/ccnc.2018.8319310>>. Citado na página 36.

EVERDING, L. et al. A mobility device for the blind with improved vertical resolution using dynamic vision sensors. In: *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, 2016. Disponível em: <<https://doi.org/10.1109/healthcom.2016.7749459>>. Citado na página 36.

FERREIRA, A.; GIRALDI, G. Convolutional neural network approaches to granite tiles classification. *Expert Systems with Applications*, Elsevier BV, v. 84, p. 1–11, out. 2017. Disponível em: <<https://doi.org/10.1016/j.eswa.2017.04.053>>. Citado na página 27.

FLORES, G. H.; MANDUCHI, R. WeAllWalk. Association for Computing Machinery (ACM), v. 11, n. 1, p. 1–28, abr. 2018. Disponível em: <<https://doi.org/10.1145/3161711>>. Citado na página 38.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. (Adaptive computation and machine learning). ISBN 9780262035613. Disponível em: <<https://books.google.co.in/books?id=Np9SDQAAQBAJ>>. Citado na página 27.

GOPALAKRISHNAN, K. et al. Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. Elsevier BV, v. 157, p. 322–330, dez. 2017. Disponível em: <<https://doi.org/10.1016/j.conbuildmat.2017.09.110>>. Citado na página 29.

HE, K. et al. Deep residual learning for image recognition. In: . IEEE, 2016. Disponível em: <<https://doi.org/10.1109/cvpr.2016.90>>. Citado 2 vezes nas páginas 11 e 30.

JARRETT, K. et al. What is the best multi-stage architecture for object recognition? In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009. Disponível em: <<https://doi.org/10.1109/iccv.2009.5459469>>. Citado na página 26.

JIANG, B. et al. Wearable vision assistance system based on binocular sensors for visually impaired users. *IEEE Internet of Things Journal*, Institute of Electrical and Electronics Engineers (IEEE), v. 6, n. 2, p. 1375–1383, abr. 2019. Disponível em: <<https://doi.org/10.1109/jiot.2018.2842229>>. Citado na página 36.

KRIG, S. *Computer Vision Metrics*. Springer International Publishing, 2016. Disponível em: <<https://doi.org/10.1007/978-3-319-33762-3>>. Citado na página 27.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. 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 30.

LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, MIT Press - Journals, v. 1, n. 4, p. 541–551, dez. 1989. Disponível em: <<https://doi.org/10.1162/neco.1989.1.4.541>>. Citado na página 26.

- LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE, 2010. Disponível em: <<https://doi.org/10.1109/iscas.2010.5537907>>. Citado na página 26.
- LECUN, Y. A. et al. Efficient BackProp. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2012. p. 9–48. Disponível em: <[https://doi.org/10.1007/978-3-642-35289-8\\_3](https://doi.org/10.1007/978-3-642-35289-8_3)>. Citado na página 28.
- LEE, C.-Y.; GALLAGHER, P.; TU, Z. Generalizing pooling functions in CNNs: Mixed, gated, and tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Institute of Electrical and Electronics Engineers (IEEE), v. 40, n. 4, p. 863–875, abr. 2018. Disponível em: <<https://doi.org/10.1109/tpami.2017.2703082>>. Citado na página 28.
- LESECQ, S. et al. INSPEX: Design and integration of a portable/wearable smart spatial exploration system. In: *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*. IEEE, 2017. Disponível em: <<https://doi.org/10.23919/date.2017.7927089>>. Citado 2 vezes nas páginas 21 e 35.
- LIN, T.-Y. et al. Microsoft coco: Common objects in context. In: FLEET, D. et al. (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 740–755. Citado na página 30.
- LIU, X.; DENG, Z.; YANG, Y. Recent progress in semantic image segmentation. *Artificial Intelligence Review*, Springer Science and Business Media LLC, v. 52, n. 2, p. 1089–1106, jun. 2018. Disponível em: <<https://doi.org/10.1007/s10462-018-9641-3>>. Citado na página 26.
- LOQUERCIO, A. et al. DroNet: Learning to fly by driving. *IEEE Robotics and Automation Letters*, Institute of Electrical and Electronics Engineers (IEEE), v. 3, n. 2, p. 1088–1095, abr. 2018. Disponível em: <<https://doi.org/10.1109/lra.2018.2795643>>. Citado 5 vezes nas páginas 11, 30, 31, 36 e 53.
- MANCINI, A.; FRONTONI, E.; ZINGARETTI, P. Mechatronic system to help visually impaired users during walking and running. *IEEE Transactions on Intelligent Transportation Systems*, Institute of Electrical and Electronics Engineers (IEEE), v. 19, n. 2, p. 649–660, fev. 2018. Disponível em: <<https://doi.org/10.1109/tits.2017.2780621>>. Citado na página 35.
- MANJARI, K.; VERMA, M.; SINGAL, G. A survey on assistive technology for visually impaired. *Internet of Things*, Elsevier BV, v. 11, p. 100188, set. 2020. Disponível em: <<https://doi.org/10.1016/j.iot.2020.100188>>. Citado na página 25.
- MATTOCCIA, S.; MACRI, P. 3d glasses as mobility aid for visually impaired people. In: *Computer Vision - ECCV 2014 Workshops*. Springer International Publishing, 2015. p. 539–554. Disponível em: <[https://doi.org/10.1007/978-3-319-16199-0\\_38](https://doi.org/10.1007/978-3-319-16199-0_38)>. Citado na página 36.
- MONTEIRO, J. et al. Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017. Disponível em: <<https://doi.org/10.1109/ijcnn.2017.7966130>>. Citado na página 37.

NAMATĒVS, I. Deep convolutional neural networks: Structure, feature extraction and training. *Information Technology and Management Science*, Riga Technical University, v. 20, n. 1, jan. 2017. Disponível em: <<https://doi.org/10.1515/itms-2017-0007>>. Citado 2 vezes nas páginas 27 e 28.

PARK, K. et al. SideGuide: a large-scale sidewalk dataset for guiding impaired people. In: . IEEE, 2020. Disponível em: <<https://doi.org/10.1109/iros45743.2020.9340734>>. Citado na página 38.

RANZATO, M. et al. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007. Disponível em: <<https://doi.org/10.1109/cvpr.2007.383157>>. Citado na página 28.

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](https://doi.org/10.1162/neco_a_00990)>. Citado na página 28.

REN, J. S. et al. An unsupervised feature learning approach to improve automatic incident detection. In: *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2012. Disponível em: <<https://doi.org/10.1109/itsc.2012.6338621>>. Citado na página 27.

ROH, Y.; HEO, G.; WHANG, S. E. A survey on data collection for machine learning: A big data - AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, Institute of Electrical and Electronics Engineers (IEEE), p. 1–1, 2019. Disponível em: <<https://doi.org/10.1109/tkde.2019.2946162>>. Citado na página 21.

RUSSAKOVSKY, O. et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, Springer Science and Business Media LLC, v. 115, n. 3, p. 211–252, abr. 2015. Disponível em: <<https://doi.org/10.1007/s11263-015-0816-y>>. Citado na página 29.

SALEH, K. et al. Navigational path detection for the visually impaired using fully convolutional networks. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017. Disponível em: <<https://doi.org/10.1109/smc.2017.8122809>>. Citado 3 vezes nas páginas 11, 27 e 37.

SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: *Artificial Neural Networks – ICANN 2010*. Springer Berlin Heidelberg, 2010. p. 92–101. Disponível em: <[https://doi.org/10.1007/978-3-642-15825-4\\_10](https://doi.org/10.1007/978-3-642-15825-4_10)>. Citado na página 26.

SCHILLING, F. et al. Geometric and visual terrain classification for autonomous mobile navigation. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017. Disponível em: <<https://doi.org/10.1109/iros.2017.8206092>>. Citado na página 37.

SHAO, L.; ZHU, F.; LI, X. Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, Institute of Electrical and Electronics Engineers (IEEE), v. 26, n. 5, p. 1019–1034, maio 2015. Disponível em: <<https://doi.org/10.1109/tnnls.2014.2330900>>. Citado na página 21.

- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: BENGIO, Y.; LECUN, Y. (Ed.). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. [s.n.], 2015. Disponível em: <<http://arxiv.org/abs/1409.1556>>. Citado 2 vezes nas páginas 11 e 29.
- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Citado na página 30.
- SMITH, R. O. et al. Assistive technology products: a position paper from the first global research, innovation, and education on assistive technology (GREAT) summit. *Disability and Rehabilitation: Assistive Technology*, Informa UK Limited, v. 13, n. 5, p. 473–485, jun. 2018. Disponível em: <<https://doi.org/10.1080/17483107.2018.1473895>>. Citado na página 25.
- TAN, C. et al. A survey on deep transfer learning. In: *Artificial Neural Networks and Machine Learning – ICANN 2018*. Springer International Publishing, 2018. p. 270–279. Disponível em: <[https://doi.org/10.1007/978-3-030-01424-7\\_27](https://doi.org/10.1007/978-3-030-01424-7_27)>. Citado 3 vezes nas páginas 11, 31 e 32.
- WANG, J. et al. An environmental perception and navigational assistance system for visually impaired persons based on semantic stixels and sound interaction. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2018. Disponível em: <<https://doi.org/10.1109/smc.2018.00332>>. Citado na página 35.
- WANG, Z. et al. Image quality assessment: From error visibility to structural similarity. *Institute of Electrical and Electronics Engineers (IEEE)*, v. 13, n. 4, p. 600–612, abr. 2004. Disponível em: <<https://doi.org/10.1109/tip.2003.819861>>. Citado 3 vezes nas páginas 11, 32 e 33.
- WATANABE, T. et al. Wheelchair behavior recognition for visualizing sidewalk accessibility by deep neural networks. In: . Springer Singapore, 2021. p. 16–29. Disponível em: <[https://doi.org/10.1007/978-981-16-0575-8\\_2](https://doi.org/10.1007/978-981-16-0575-8_2)>. Citado na página 38.
- WEISS, M. et al. *Navigation Agents for the Visually Impaired: A Sidewalk Simulator and Experiments*. 2019. Citado na página 38.
- World Health Organization. *Global Cooperation on Assistive Technology (GATE)*. 2021. <<https://www.who.int/disabilities/technology/gate/en/>>, Last accessed on 2021-03-08. Citado na página 21.
- YE, C.; QIAN, X. 3-d object recognition of a robotic navigation aid for the visually impaired. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Institute of Electrical and Electronics Engineers (IEEE), v. 26, n. 2, p. 441–450, fev. 2018. Disponível em: <<https://doi.org/10.1109/tnsre.2017.2748419>>. Citado na página 36.
- YOSHIOKA, T. et al. The NTT CHiME-3 system: Advances in speech enhancement and recognition for mobile multi-microphone devices. In: *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2015. Disponível em: <<https://doi.org/10.1109/asru.2015.7404828>>. Citado na página 28.