



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Gabriel Nandi Astrissi

**Detecção de Distração e Sonolência em Condutores Veiculares Utilizando
Redes Neurais Convolucionais**

Araranguá
2023

Gabriel Nandi Astrissi

**Detecção de Distração e Sonolência em Condutores Veiculares Utilizando
Redes Neurais Convolucionais**

Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Computação submetido ao Centro de Ciências, Tecnologias e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Antonio Carlos Sobieranski, Dr.

Araranguá

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Astrissi, Gabriel Nandi

Detecção de Distração e Sonolência em Condutores
Veiculares Utilizando Redes Neurais Convolucionais /
Gabriel Nandi Astrissi ; orientador, Antonio Carlos
Sobieranski, 2023.

35 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Campus Araranguá,
Graduação em Engenharia de Computação, Araranguá, 2023.

Inclui referências.

1. Engenharia de Computação. 2. Detecção de Distração e
Sonolência. 3. Sistema de Apoio à Direção. 4. Visão
Computacional. 5. Redes Neurais Convolucionais. I.
Sobieranski, Antonio Carlos. II. Universidade Federal de
Santa Catarina. Graduação em Engenharia de Computação. III.
Título.

Gabriel Nandi Astrissi

**Detecção de Distração e Sonolência em Condutores Veiculares Utilizando
Redes Neurais Convolucionais**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia de Computação e aprovado em sua forma final pelo Curso de Graduação em Engenharia de Computação.

Araranguá, 29 de Novembro de 2023.

Prof. Jim Lau, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Antonio Carlos Sobieranski, Dr.
Orientador

Prof. Jim Lau, Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Rodrigo Vinícius Mendonça Pereira,
Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Fábio Rodrigues De La Rocha, Dr.
Avaliador Suplente
Universidade Federal de Santa Catarina

Detecção de Distração e Sonolência em Condutores Veiculares Utilizando Redes Neurais Convolucionais

Gabriel Nandi Astrissi

2023, NOVEMBRO

Resumo

É constatado que a principal causa de óbitos por acidentes de trânsito decorre da falta de atenção do condutor, associado com outros fatores tais como a falta do cinto, embriaguez, imprudência, dentre outros. Conforme levantamento realizado pela Organização Pan-Americana da Saúde, é apontado que mais de 1 milhão de mortes são causadas por acidentes de trânsito. Boa parte destes óbitos poderiam ser evitados com medidas educativas, campanhas conscientizadoras, inclusive nas séries iniciais nas escolas, e também nas auto-escolas. Sempre com o intuito de formar motoristas e pedestres mais conscientes. Neste sentido, o emprego de sistemas baseados em Visão Computacional pode ser uma solução complementar no auxílio da medição dos níveis de atenção dos condutores e na geração de alertas visuais e sonoros. Como resultado, espera-se que estes alertas aumentem a segurança do tráfego e diminuam o número de acidentes. O propósito deste trabalho é desenvolver um sistema com a capacidade de detectar distração e sonolência para auxiliar motoristas na direção. Para tal, será analisado o uso de Redes Neurais Convolucionais e como esta tecnologia pode ser empregada na detecção automatizada de padrões de distração e de sonolência de condutores. Resultados preliminares apontam a viabilidade da solução proposta, cujo resultado obtido foi de 95% e 84% de precisão em testes distintos de identificação de padrão de sonolência em condutores.

Palavras-chaves: Detecção de Distração e Sonolência, Sistema de Apoio à Direção, Visão Computacional, Redes Neurais Convolucionais, Sistema Assistivo de Direção.

Detection of Distraction and Drowsiness in Vehicle Drivers Using Convolutional Neural Networks

Gabriel Nandi Astrissi

2023, NOVEMBER

Abstract

It has been found that the main cause of deaths from traffic accidents is the lack of driver's attention, associated with other factors such as lack of seat belt, drunkenness, etc. A survey carried out by the Pan American Health Organization indicates that more than 1 million deaths are caused by traffic accidents. Majority of those deaths could be prevented by usage of educational measures, awareness raising campaigns, including in the initial grades in schools, and in driving schools. Always with the intention of raising more conscientious drivers and pedestrians. In this way, the employment of systems based on Computer Vision can be a complementary solution in helping the measurement of the drivers' attention levels and in generating visual and sound alerts. As result, it's expected that this alerts raise the traffic security and lower the number of accidents. The purpose of this work is to develop a system with the ability to detect distraction and drowsiness to assist drivers in driving. To this end, the usage of Convolutional Neural Networks and how this technology can be employed in the automated pattern detection of driver distraction and drowsiness shall be analysed. Preliminary results point towards the viability of the proposed solution, which the obtained result was of 95% and 84% accuracy in distinct tests in the identification of drowsiness patterns in conductors.

Key-words: Drowsiness and Distraction Detection, Steering Support System, Computer Vision, Convolutional Neural Networks, Steering Assist System.

1 Introdução

Acidentes de trânsito são uma grande causa de mortes no mundo, sendo responsáveis por 1,3 milhão de óbitos por ano segundo a Organização Mundial de Saúde – OMS ([Organização Pan-Americana da Saúde, 2021](#)). As formas comuns de acidente são ocasionadas por: excesso de velocidade, falta do uso de cinto de segurança, uso de bebida alcoólica, dentre outros. No Brasil são em média 40 mil mortes por ano ([Câmara dos Deputados, 2017](#)). Segundo a Associação Brasileira de Medicina de Tráfego (ABRAMET), a falta de atenção do condutor foi a principal causa de acidentes de trânsito e a sonolência foi a terceira maior causa de acidentes de trânsito no período de 2014 à 2020 ([ABRAMET, 2020](#)). Conforme a Tabela 1, é observado que existe uma significativa fração de acidentes causados pela falta de atenção, que poderiam ser evitados com o uso de políticas de segurança pública.

Figura 1 – Tabela de Acidentes no Período 2014-2020

| ACIDENTES DE TRÂNSITO NAS RODOVIAS FEDERAIS DO BRASIL - 2014 À 2020* | | | | | | | | |
|--|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Causa do Acidente | ACIDENTES | | | | | | | |
| | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020* | TOTAL |
| Falta de Atenção à Condução | 55.089 | 36.805 | 26.451 | 34.495 | 25.799 | 24.992 | 11.770 | 215.401 |
| Ingestão de Álcool | 7.391 | 6.745 | 6.332 | 6.448 | 5.196 | 5.419 | 2.737 | 40.268 |
| Condutor Dormindo | 4.794 | 4.067 | 3.763 | 3.798 | 2.632 | 2.483 | 1.146 | 22.683 |
| Mal Súbito | - | - | - | 969 | 677 | 715 | 341 | 2.702 |
| Restrição de Visibilidade | - | - | - | 884 | 539 | 493 | 289 | 2.205 |
| Ingestão de Substâncias Psicoativas | - | - | - | 78 | 73 | 94 | 47 | 292 |
| TOTAL | 67.274 | 47.617 | 36.546 | 46.672 | 34.916 | 34.196 | 16.330 | 283.551 |

Fonte: ABRAMET

No que diz respeito aos acidentes causados por sonolência, muitos poderiam ser evitados se o motorista que se encontra neste estado deixar a condução do veículo para outra pessoa, ou parar a atividade de direção para repouso conforme é previsto no art. 67-C da Lei nº 13.103 ([BRASIL, 2015](#)) que exige paradas para descanso do condutor veicular após certo período na atividade de condução.

Com o advento das tecnologias e sistemas computacionais para assistência à direção, alertas podem ser gerados em situações de periculosidade para contribuir com a condução veicular e reduzir acidentes. Exemplos de sistemas computacionais veiculares podem ser baseados em detecção das faixas de sinalização, detecção de proximidade, frenagem automatizada, detecção de padrões irregulares de direção, dentre outros. Tais sistemas são embarcados diretamente nos sistemas computacionais do veículo, realizando leituras de telemetria tal como *Driver Alert* da Ford ([NIEH, 2011](#)) e *Driving Assistant* da BMW ([BMW, 2013](#)).

Um outro tipo de sistema crucial para que o condutor se mantenha atento à direção são os sistemas que detectam padrões de atenção e sonolência por meio de visão computacional. Para estes sistemas, são utilizadas câmeras que operam no espectro de luz visível ou infravermelho, que realizam a leitura facial do condutor analisando decréscimos de atenção em relação aos seus sinais vitais de atenção. Com o incremento da capacidade computacional obtida recentemente, uma série de algoritmos de processamento digital de

imagens, reconhecimento de padrões e visão computacional tornaram-se viáveis a serem aplicados para o processamento em tempo real. Mais recentemente, com o ressurgimento da Inteligência Artificial, modelos de redes neurais obtiveram a sua evolução para incluir camadas profundas e convolutivas, tornando solucionadores gerais de problemas, que operam em função do aprendizado dos padrões que determinadas classes de problemas possuem. De fato, redes neurais convolucionais vem sendo aplicadas para uma série de problemas, o mesmo se aplica à área de assistência ao motorista.

Ao longo da literatura, sistemas computacionais para monitoramento de atenção de condutores baseados em modelos clássicos de visão computacional ou baseados em modelos convolutivos podem ser encontrados. Sistemas clássicos de visão computacional são baseados em detecção facial pelo método de Viola-Jones (VIOLA, 2001), HAAR Cascades (VIOLA, 2001), modelos probabilísticos e estatísticos (DAVIES, 2005a), detecção de bordas (DAVIES, 2005c) e segmentação de imagens (DAVIES, 2005d). Modelos baseados em Inteligência Artificial e técnicas de Machine Learning que utilizam redes neurais para tarefa de monitoramento de atenção do motorista usam das técnicas Redes Convolucionais (YUEN; MARTIN; TRIVEDI, 2016) e LSTM (LI et al., 2021).

O presente trabalho aborda o uso de Redes Neurais Convolutivas para a detecção dos padrões de decréscimo de atenção por parte de um condutor veicular. Com o uso de uma câmera convencional, foi desenvolvido um ambiente experimental onde leituras são realizadas do condutor em operação, e algoritmo híbrido de visão computacional e uso de redes neurais convolucionais a posteriori. São adquiridas imagens faciais e o algoritmo desenvolvido possui 2 etapas: (i) detecção facial com o Algoritmo de Viola-Jones e localização da região dos olhos, (ii) tomada de decisão em relação à leitura da região dos olhos e face para determinar os estágios de decréscimo de atenção e sonolência do usuário. Os resultados obtidos demonstram um grau de assertividade superior a 84% no ambiente experimental desenvolvido, indicando que a solução desenvolvida é viável de ser embarcada para propósitos de identificação de pontos de atenção em condutores veiculares.

O restante deste trabalho está organizado conforme segue: Seção 2 são levantados os trabalhos correlatos à presente pesquisa. Na Seção 3 a fundamentação teórica acerca do tema explorado é apresentada. A metodologia do presente trabalho é apresentada na Seção 4, enquanto que na Seção 5 são demonstrados os resultados experimentais e a validação da abordagem computacional desenvolvida. Por fim, a Seção 6 apresenta as Conclusão e Trabalhos Futuros em relação ao tema.

2 Trabalhos Correlatos

Nesta Seção serão abordados os trabalhos correlatos com o tema sistemas de detecção de falta de atenção do condutor por visão computacional. Foram considerados os trabalhos publicados entre 2005 a 2023, foi utilizado a Keyword “driver assistant system”, sendo descritos abaixo os que mais se aproximam do que foi realizado neste trabalho:

2.1 Métodos Clássicos

Em V. Sanjay Kumar (Sanjay Kumar et al., 2020) o autor usa detecção de fadiga, detector de álcool e sensor piezoelétrico para detecção de batidas para monitorar o estado fisiológico do motorista. Utiliza-se de classificador de cascatas HAAR para detectar atributos da face e dos olhos utilizando de *bounding boxes* para classificar os olhos, gerando um alerta sonoro no caso destes ficarem fechados por muito tempo. Também é utilizado um

sensor de gás MQ3 para analisar a presença de álcool no ar, e de um sensor piezoelétrico para detectar colisões. Os sensores são conectados a um *minicomputador* Raspberry Pi via USB. Nos testes não foram gerados alarmes falsos e a taxa de detecção foi de 85%.

Em Jesús Nuevo (NUEVO et al., 2006) é usado um rastreador AAM (Active Appearance Model) para monitorar o nível de alerta do motorista utilizando de lógica fuzzy para analisar múltiplos parâmetros. Para o sistema de captura de imagens foi usada uma câmera de espectro infravermelho próximo posicionada no meio de dois emissores de infravermelho próximo. Em conjunto com o AAM para o rastreamento de rosto foram utilizados dois filtros Kalman, um para cada pupila.

Em Zhencheng Hu (HU et al., 2015) foi utilizado um sensor Kinect para capturar imagens infravermelho normais e com padrão 3D, as imagens com padrão 3D são usadas para detectar a região do rosto através de histograma de profundidade, enquanto as imagens infravermelho sem o padrão de profundidade são utilizadas com detector Haar para rastreamento da orientação do rosto.

Em Mahdi Rezaei (REZAEI; KLETTE, 2011) foi buscado os parâmetros ideais para o treino de um detector de cascatas Haar para detecção em tempo real do status dos olhos. Foi assumido dois estados: olhando para a frente e olhos abertos. Foram assumidos 5 cenários para a realização da detecção dos quais é derivado o grau de distração do motorista: olhando para frente, rosto inclinado, a frente do rosto não é detectado mas um dos olhos é, olhos fechados e nem rosto nem olhos são detectados.

Em Parakram Singh (SINGH; MAHIM; PRAKASH, 2022) foi detectado sonolência através da proporção dos olhos utilizando o algoritmo de detecção Dlib Facial Feature sendo obtida uma precisão de 85.2% no dataset de teste.

Em Nikhil Tyagi (TYAGI; GOYAL; RAUTELA, 2022) foi utilizado do cálculo da proporção dos olhos para a detecção de sonolência sendo que para a detecção de face e olhos foi utilizado método Viola-Jones enquanto para encontrar a localização destes foi através do método de cascatas Haar.

2.2 Métodos Contemporâneos (Inteligência Artificial e Redes Neurais Convolucionais)

Em Jing Huang (HUANG; LIU; PENG, 2022) o autor usou 4 métodos diferentes (XGBoost, CNN, ConvLSTM and uma combinação de CNN e LSTM) para analisar 6 canais diferentes de sinais biológicos (single channel of ECG, EDA, RSP, R-R interval and two channels of EEG) para detectar sonolência do motorista no volante. Para o classificador XGBoost, 61 features dos domínios do tempo, espectral e *wavelet* foram extraídos dos sinais de 6 canais. Para a CNN, uma versão com redução de amostragem de 500 Hz para 100 Hz do sinal foi usado na forma de uma matriz 6 por 100 como entrada em uma CNN com 6 camadas convolucionais que a cada duas camadas convolucionais possui uma *camada de dropout*, então os atributos passam por uma *camada para flattening* conectada em uma *camada softlayer* para completar a tarefa de classificação. Para a LSTM de convolução, ao invés de usar o padrão, foi criada uma para processar os sinais dos 6 canais de forma síncrona, usando 4 camadas ConvLSTM, uma camada totalmente conectada e uma *camada softlayer* para a classificação final. A precisão obtida nos testes com um sinal de 1s: XGBoost 63.6%, CNN 95.3%, ConvLSTM 79.3% e CNN + LSTM 96.2%.

Em Kevan Yuen (YUEN; MARTIN; TRIVEDI, 2016) foi utilizado da CNN AlexNet em combinação com Stacked Hourglass modificado para incluir informação de oclusão no mapa de temperatura com o intuito de detectar a localização da face e a pose da cabeça. A

pose da cabeça é estimada usando o algoritmo POSIT. A validação foi feita no VIVA-Face dataset, atingindo uma taxa de detecção de 86,1%.

Em E. Mary Bearly (BEARLY; CHITRA, 2023) foi criado um sistema para a detecção de sonolência que é composto por quatro partes: um detector facial para detectar o rosto em vídeos, rede DGAN (Dependent Generative Adversarial Network) para detectar sonolência a partir de características faciais indentificadas, três níveis de atenção e uma técnica de suavização para decréscimo de predições incorretas. A precisão obtida foi de 93% no dataset NTHU-DDD.

Em Goutam Kumar Sahoo (SAHOO; DAS; SINGH, 2023) foi usado um sistema que utiliza da rede DCNN leve SqueezeNet 1.1 para detectar dez posturas de direção distraída para detecção de distração em dispositivos com poder computacional restrito. A precisão de detecção obtida foi de 99.93%.

3 Fundamentação Teórica

3.1 Processamento Digital de Imagens

Segundo D'Antuono (D'ANTUONO, 2022) imagens são representações da realidade e o poder computacional representar todos os aspectos nem sempre está a disposição, para obter estas representações da realidade são utilizados detectores que convertem a radiação da região do espaço de interesse nestas representações. Para o caso do imageamento digital são geradas pelos detectores informações digitais para a representação da faixa de espectro de luz visível o qual é a faixa de interesse, tipicamente sendo dividida a magnitude de intensidade em 256 valores (representação em 8-bits) em três canais para a representação da intensidade das cores vermelho, verde e azul ou 256 valores em um canal de luminosidade para imagens em escala cinza.

3.1.1 Imagem Digital e Formas de Aquisição

A aquisição de imagens de alta qualidade é de extrema importância para atividades de visão computacional. Segundo Davies (DAVIES, 2005b) qualquer deficiência na aquisição das imagens pode resultar que os propósitos da aplicação de visão computacional não serem cumpridos, tendo em vista que os objetos não são reconhecíveis devido às distorções que estes poderiam vir a sofrer.

Também segundo Davies (DAVIES, 2005b) à aquisição de imagens digitais é feita através de câmaras CCD que gera tensões analógicas que são digitalizadas em informações em escala cinza contidas entre 6 a 8 bits o qual é feita por conversores analógicos para digital (ADC converter).

Em Bigas (BIGAS et al., 2006) os sensores CMOS são descritos como tendo um baixo consumo de energia, baixo custo, alta velocidade de imageamento porém possuem menor qualidade de imagem que o CCD. Tipicamente um gerador de imagens CMOS composto por um vetor de pixels, processador de sinais analógicos, seletor de linha e coluna e um componente denominado de Temporização e Controle (Timing and Control).

3.1.2 Algoritmos de Filtragem e Redução de Ruídos

De acordo com Pulfer (PULFER, 2019), filtros de Blur são filtros Passa-Baixa e tem o objetivo de fazer com que a imagem fique menos pixelada assim reduzindo o ruído

através do ajuste do valor de cada pixel na direção de um valor médio que leva em conta todos os pixels da imagem.

O filtro Gaussiano é descrito por Pulfer (PULFER, 2019) como sendo um filtro Passa-Baixa que realiza a convolução da imagem com um kernel Gaussiano que é calculado por uma função Gaussiana tipicamente sendo aplicado primeiramente no sentido horizontal e depois no sentido vertical pois a implementação desta forma requer um número inferior de cálculos comparado a percorrer ambas direções simultaneamente.

Conforme Sobieranski (SOBIERANSKI et al., 2008) o filtro Perona-Malik se trata de um filtro isotrópico que possui detecção de bordas mais precisa do que o filtro Canny apesar da imprevisibilidade por somente atingir estabilidade nas bordas após um número alto de iterações. O modelo discreto do filtro é descrito pela Equação 1

$$I(s, t + 1) = I(s, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla I_{s,p}(t)| \nabla I_{s,p}(t)) \quad (1)$$

Onde $I(s,t)$ é a imagem discreta em termos espaciais e temporais, s é a posição do pixel em uma grade bidimensional discreta, t é o passo temporal discreto e tem de ser maior ou igual a 0, λ é a constante de velocidade de difusão e η_s representa um conjunto de vizinhos espaciais do pixel s .

Também de acordo com Sobieranski (SOBIERANSKI et al., 2008) o filtro Weickert é um filtro anisotrópico que devido a sua complexidade computacional tem de utilizar uma quantidade predefinida de iterações para controlar parâmetro responsável pela intensidade da difusão λ . Este é um filtro que utiliza de processos iterativos sendo descrito pelo seguinte modelo discreto na Equação 2.

$$u_{i,j}^{k+1} = u_{i,j}^k + \tau \sum_{(m,n) \in N(i,j)} g_{m,n}^k \frac{u_{m,n}^k - u_{i,j}^k}{|(m,n) - (i,j)|^2} \quad (2)$$

3.1.3 Algoritmos de Detecção de Bordas

Segundo Ziou (ZIOU; TABBONE, 1998) é difícil fazer um inventário de todos os algoritmos disponíveis para a operação de detecção de bordas devido ao aumento contínuo que ocorre desde o início da criação do processamento de imagens.

Segundo Ansari (ANSARI; KURCHANIYA; DIXIT, 2017), o método de detecção de bordas Sobel é um método que utiliza um par de máscaras convolucionais 3x3 o qual dá maior peso ao pixel central quando tirada a média, já o método de detecção de bordas Roberts é um método computacionalmente eficiente e de baixa complexidade que mensura o gradiente espacial da imagem, recebendo como entrada uma imagem em escala cinza como entrada a qual é aplicada um par de máscaras convolucionais 2x2 e gerando uma imagem de saída a qual é composta pelos valores estimados da magnitude do gradiente espacial para cada pixel. Segundo Naidu (NAIDU P. LAKSHMAN RAO, 2012), o método de detecção de bordas Robison utiliza de 4 filtros 3x3 e a multiplicação destes por -1 para obter uma aproximação da primeira derivada da imagem para obter as bordas.

De acordo com Ansari (ANSARI; KURCHANIYA; DIXIT, 2017) o método Canny para extração de bordas é o método mais utilizado por ser superior aos outros tendo a capacidade de controlar a quantidade de detalhes na imagem de saída que contém as bordas e realiza a extração do ruído da imagem antes de extrair as bordas tendo uma remoção de ruído eficiente. Utiliza uma máscara de convolução 3x3. O método é constituído pelos seguintes quatro passos:

1. Blur Gaussiano para remover ruídos.
2. Cálculo dos gradientes locais e direções das bordas.
3. Processo de supressão de não-máximos.
4. Processo de conexão de bordas.

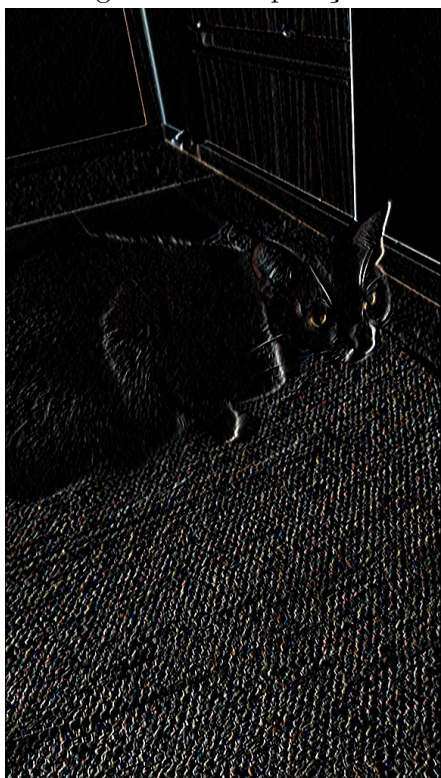
Para a comparação dos métodos de detecção de bordas citados acima, na Figura 2 temos uma imagem a qual a saída após ter sido ser processada por estes métodos de detecção de bordas é demonstrada na Figura 3.

Figura 2 – Imagem original

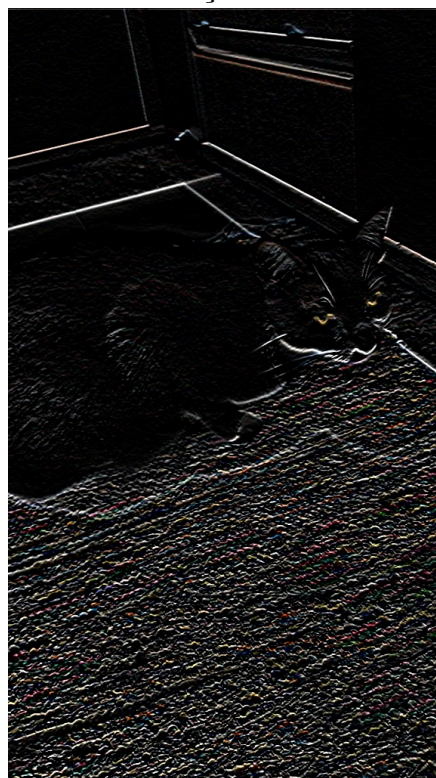


Fonte: Próprio autor

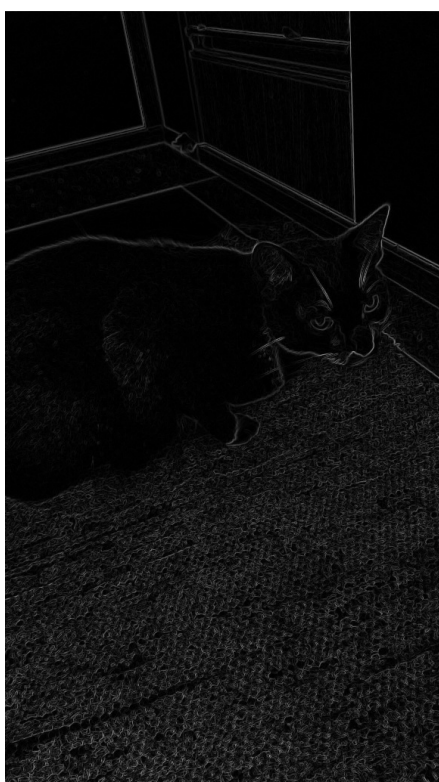
Figura 3 – Comparação entre os algoritmos de detecção de bordas



(a) Imagem gerada pelo algoritmo Sobel com dx ordem 1



(b) Imagem gerada pelo algoritmo Sobel com dy ordem 1



(c) Imagem gerada pelo algoritmo Roberts



(d) Imagem gerada pelo algoritmo Canny

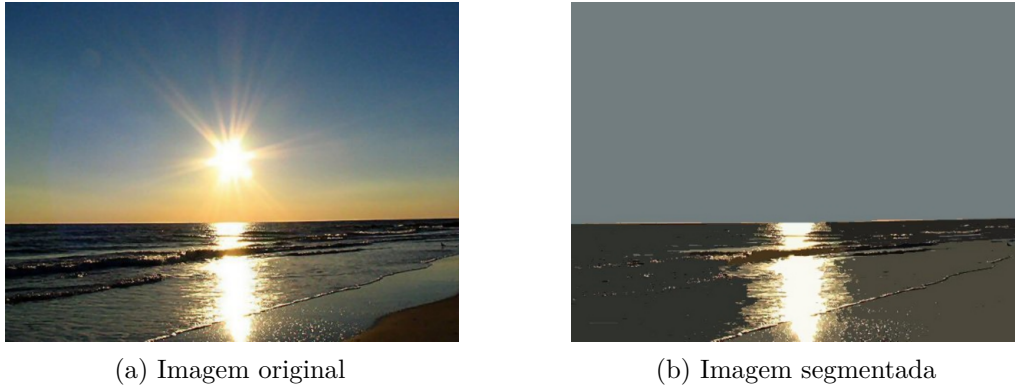
Fonte: Próprio autor

3.1.4 Algoritmos de Segmentação de Imagens

Segundo Sobieranski (SOBIERANSKI et al., 2008) a segmentação busca dividir uma imagem em segmentos que são compostos por pixels agrupados por similaridade, a qual é definida por pixels que obtiveram um mesmo label atribuído pertencerem a uma mesma região. Estas regiões devem possuir algum grau de homogeneidade das características que definem estes pixels tal como a cor ou a luminosidade. Inicialmente os algoritmos de segmentação trabalhavam com imagens em escala cinza, sendo posteriormente desenvolvidos algoritmos com a capacidade de trabalharem com imagens coloridas e com maiores resoluções após a expansão de recursos computacionais disponíveis permitir que a implementação destes fosse viável.

Conforme Sobieranski (SOBIERANSKI et al., 2008) Componentes Conexos é uma extensão da *teoria de grafos* que funciona buscando em uma imagem de entrada por pixels adjacentes (conexos) que possuam intensidade próxima (contínuos) rotulando estes em regiões pelo processo conhecido como *labeling*. Há duas formas diferentes de considerar quais pixels são vizinhos conexos, somente utilizando os pixels vizinhos nos eixos horizontais e verticais e outra considerando os pixels nas diagonais também como vizinhos conexos juntamente com os pixels vizinhos pertencentes aos eixos horizontais e verticais. Geralmente há duas formas para calcular qual valor utilizar para os pontos de referência, de forma estática a qual o valor de referência é fixo e a comparação é feita diretamente se o valor do pixel está dentro do *threshold*, e de forma dinâmica a qual o valor do ponto de referencial vai sendo ajustado conforme a posição atual para evitar mudanças abruptas de região entre pixels que localmente são similares.

Figura 4 – Resultado da segmentação por Componentes Conexos



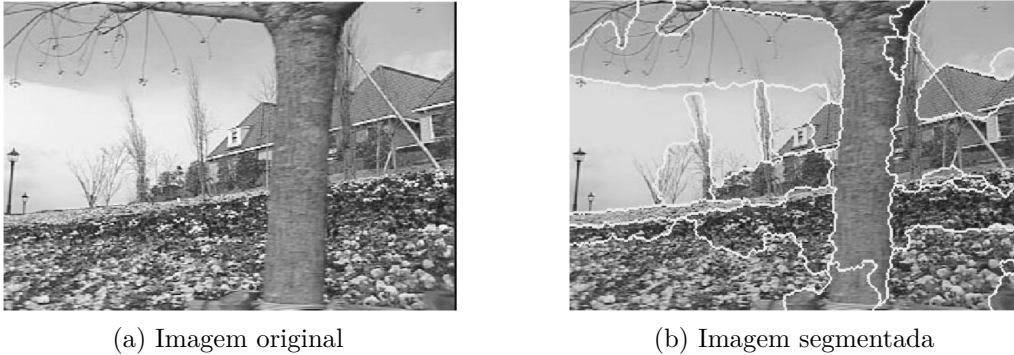
Fonte: (SOBIERANSKI et al., 2008)

Em Sobieranski (SOBIERANSKI et al., 2008) o algoritmo JSEG é descrito como um algoritmo que trabalha no espaço de cores CIELUV e preza pela identificação de texturas o qual utiliza da quantização de cores para então, sobre os agrupamentos obtidos aplicar um algoritmo de crescimento de regiões efetuando o cálculo de similaridade utilizando do valor de erro acumulado da quantização para definir a homogeneidade do segmento da imagem, quanto menor o erro, mais homogêneo este segmento é. A fraqueza deste algoritmo é que ele necessita de grandes variações de luminosidade na cena para conseguir segmentar esta.

Em Deng (DENG; MANJUNATH, 2001) é mencionado que a separação em duas etapas é benéfica pois é muito complicado analisar a distribuição e a similaridade das

cores ao mesmo tempo, sendo que a etapa em que é aplicado o algoritmo de crescimento de regiões é chamada de etapa de segmentação espacial e utiliza de uma heurística para encontrar os pontos iniciais para o crescimento das regiões e após isto estes pontos iniciais vão crescendo pixel a pixel até ser obtida as várias regiões que compõem a imagem.

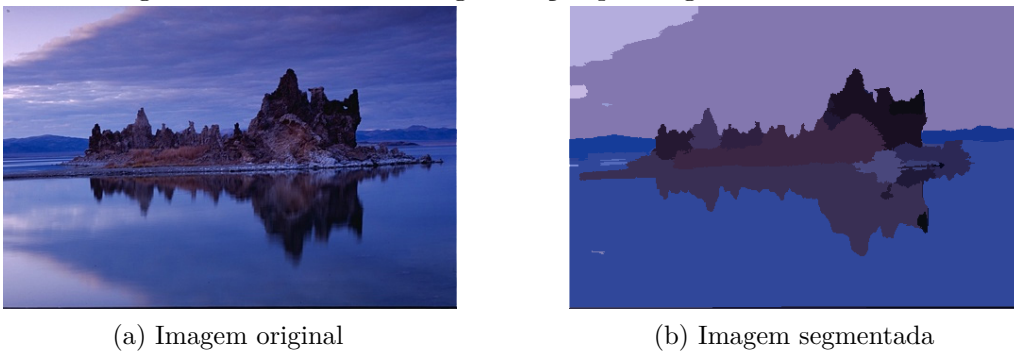
Figura 5 – Resultado da segmentação pelo algoritmo JSEG



Fonte: (DENG; MANJUNATH, 2001)

Conforme Sobieranski (SOBIERANSKI et al., 2008) algoritmo Edge Detection and Image Segmentation System ou de forma abreviada EDISON é um sistema para a extração de características em baixo nível que possibilita a segmentação de variações suaves dos gradientes na imagem o qual é composto pela combinação dos algoritmos de detecção de borda baseado em confiança que tem a utilidade de atuar em como identificar e agrupar as regiões para penalizar a função de densidade do algoritmo de segmentação Meanshift que é o outro algoritmo que compõe este sistema.

Figura 6 – Resultado da segmentação pelo algoritmo EDISION

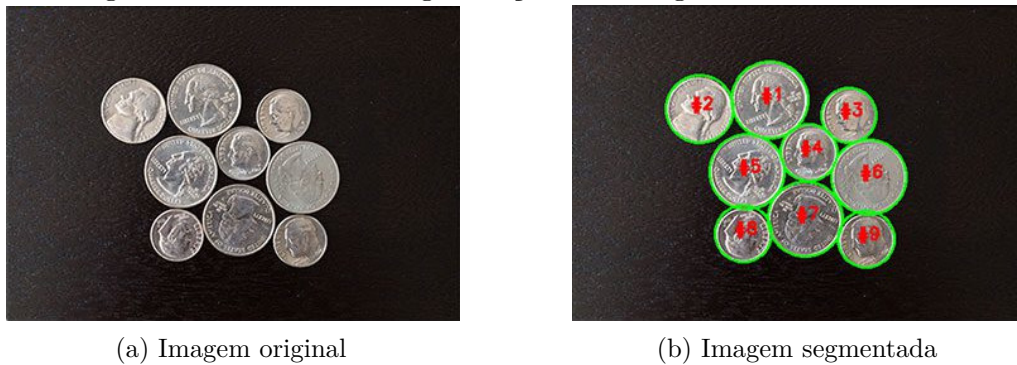


Fonte: (Aldo von Wangenheim,)

Em Zhao (ZHAO et al., 2008) é relatado que o algoritmo de Watershed foi inicialmente proposto por Beucher e Vincent, sendo um algoritmo baseado em morfologia matemática que simula inundações em um mapa topológico em que pixels de alta clareza seriam picos de montanhas em um oceano de pixels de baixa clareza. Também é mencionado que ruídos e irregularidades na imagem causam sobre-segmentação neste algoritmo de *threshold* interativo-adaptativo.

Em Sobieranski (SOBIERANSKI et al., 2008) o algoritmo de Watershed é descrito como sendo tipicamente aplicado ao gradiente da imagem e cujo o conceito básico é que a entrada é interpretada como sendo uma bacia hidrográfica. A posição de cada pixel são as coordenadas espaciais bidimensionais e a altitude do relevo é calculada da intensidade do valor da cor do pixel sendo localizado os mínimos locais de cada região para servir como ponto referencial de onde o escoamento de água começará para inundar de forma uniforme todas as bacias hidrográficas. Sendo assim localizadas as linhas de *watershed* que são as linhas divisórias entre as bacias. A partir de determinação destas, para impedir que com o aumento gradual do nível da água haja surgimento das diferentes represas, são inseridos vários diques que são localizados através de operações como a dilatação e erosão sendo repetido até que somente os diques permaneçam visíveis sobre as linhas de *watershed* então obtendo-se as fronteiras entre as regiões da imagem.

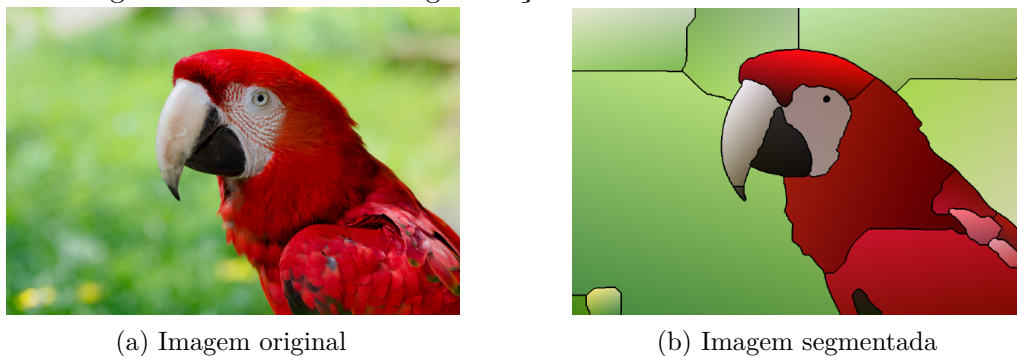
Figura 7 – Resultado da segmentação com o algoritmo de Watershed



Fonte: (Adrian Rosebrock, 2015)

De acordo com Sobieranski (SOBIERANSKI et al., 2008) o modelo Mumford-Shah é um modelo matemático cuja as aplicações vão além da segmentação de imagens podendo ser utilizado para várias outras aplicações desde a estimativa de oclusão de objetos a até a movimentação em vídeos dentre outras, possuindo a vantagem de ter a base variacional em equações diferenciais parciais assim podendo descrever qualquer fenômeno de forma generalizada porém o desconhecimento da função de aproximação da imagem e a possibilidade de existir mínimos locais dificulta a utilização desta.

Figura 8 – Resultado da segmentação com o modelo Mumford-Shah



Fonte: (KIEFER; STORATH; WEINMANN, 2020)

3.2 Algoritmo de Detecção de Objetos Viola-Jones

O método Viola-Jones segundo Ferrassini (FERRASSINI, 2014) utiliza da extração de características de uma imagem para a detecção da região em que se encontra um objeto por uma questão de velocidade de processamento pois por se tratarem de regiões de pixels, o volume de características é significativamente inferior a o de pixels. O algoritmo utiliza 3 formatos de características: a de dois retângulos a qual é a diferença da soma entre os pixels pertencentes ao retângulos, três retângulos que é o resultado da diferença da soma dos pixels de um retângulo interno e a soma dos pixels pertencentes aos dois retângulos restantes e quatro retângulos que é calculada de pares diagonais de retângulos obtendo a diferença entre estes.

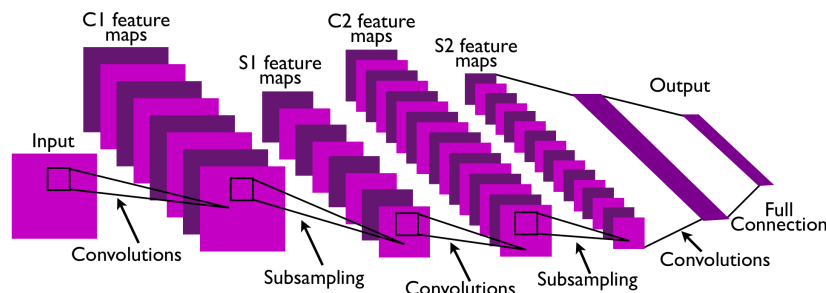
3.3 Machine Learning

Segundo Vaz (VAZ, 2021) as técnicas de aprendizado dentro do *machine learning* podem se dividir em *aprendizado supervisionado* o qual utiliza de grupos de entrada-saída (os dados de treino são pré-classificados) predefinidos para aprender uma função de mapeamento, *aprendizado não supervisionado* que utiliza de um conjunto de dados não classificados para aprender padrões que compõem este conjunto de dados, e *aprendizado por reforço* que utiliza de agentes cujo o objetivo é obter máxima recompensa cumulativa realizando ações dentro de um ambiente simulado.

3.4 Redes Neurais Convolucionais

De acordo com LeCun (LECUN; KAVUKVUOGLU; FARABET, 2010) redes neurais convolucionais são arquiteturas multiestagio treináveis em que cada estágio é composto por um vetor de entrada e outro de saída denominados de *mapas de características* ou *feature maps* os quais a dimensionalidade depende do tipo de dado o qual a rede trabalhará com. Para processamento de imagens os *mapas de características* são vetores bi-dimensionais. A Figura 9 demonstra a organização de uma rede neural convolucional de dois estágios.

Figura 9 – Ilustração de uma rede neural convolucional com dois estágios



Fonte: (LECUN; KAVUKVUOGLU; FARABET, 2010)

Para a tarefa de reconhecimento de imagens, segundo LeCun (LECUN; KAVUKVUOGLU; FARABET, 2010) cada estágio é composto pelas seguintes camadas:

Camada de Banco de Filtros(Filter Bank Layer): a qual é composta pela operação de convolução discreta do *mapa de características* com um kernel. A Equação 3 descreve a operação realizada, aonde x é o vetor de entrada o qual é composto de n_1 cada *mapa de características* de dimensão $n_2 \times n_3$ sendo que x_i denota um *mapa de características* e x_{ijk} denota um componente individual, y é um vetor de saída composto por m_1 *mapas de características* de dimensão $m_2 \times m_3$, k_{ij} é o kernel e possui dimensão de $l_1 \times l_2$, b_j é um valor de bias treinável e $*$ é o operador de convolução discreta.

$$y_j = b_j + \sum_i k_{ij} * x_i \quad (3)$$

Camada de Não-Linearidade(Non-Linearity Layer): que é a camada na qual os neurônios passam por uma função de ativação que tem o propósito de causar competição entre as características adjacentes no *mapa de características* sendo tradicionalmente uma função sigmoid más que em implementações mais recentes estão sendo utilizadas outras funções de não-linearidade como a sigmoid retificada.

Camada de Pooling de Características(Feature Pooling Layer): que computa o valor médio na vizinhança de cada *mapa de características* assim causando uma redução na resolução do *mapa de características* de saída.

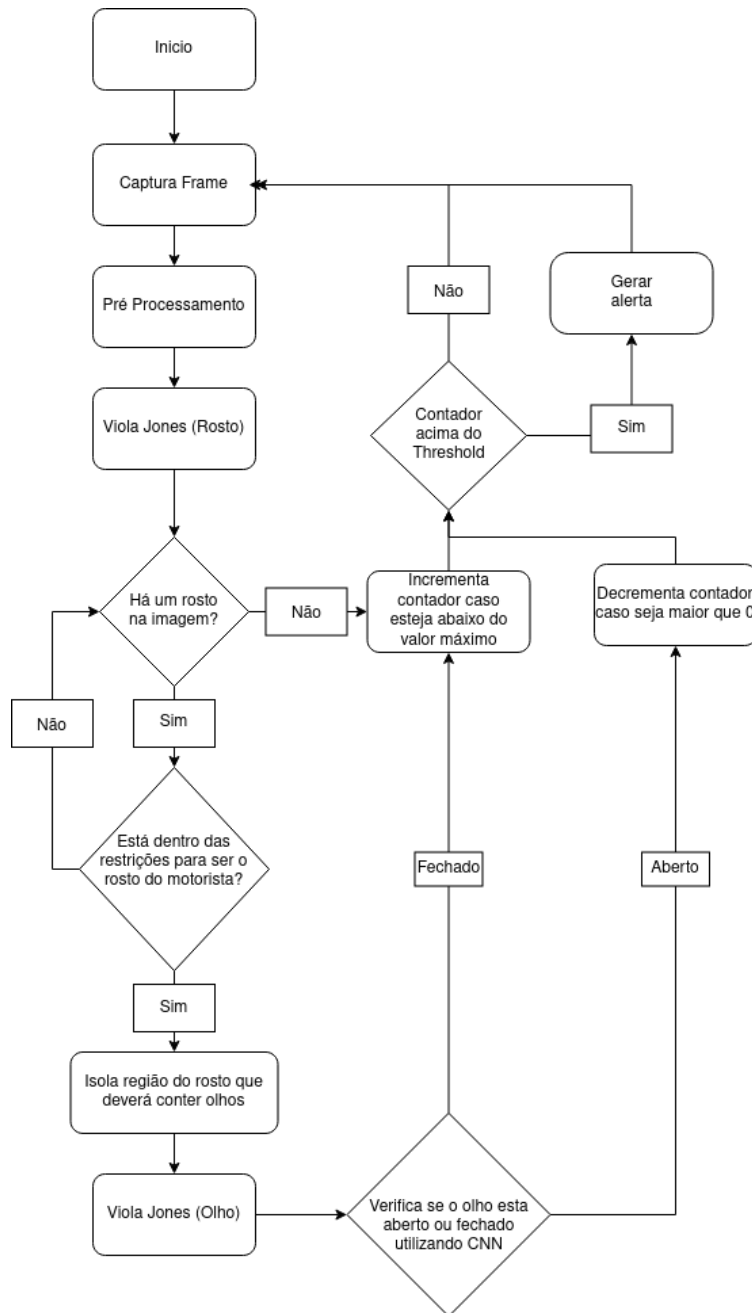
Por fim, segundo Canal (CANAL, 2021) a informação passa por uma *camada densa* à qual é responsável por tomar a decisão final do processo de classificação por ser a camada responsável por operações lógicas de alto nível, porém as operações realizadas nesta camada são custosas por causa de sua topologia em que todos os neurônios estão conectados a todos os neurônios da camada anterior o que resulta em um alto volume de cálculos. Na Figura 9 é possível visualizar está camada no final do processo, sendo chamadada de Full Connection.

4 Metodologia

A solução computacional composta é sumarizada conforme o diagrama da Figura 10 e é composta de cinco etapas principais:

1. Pré-processamento;
2. Detecção do rosto através do algoritmo de Viola-Jones;
3. Detecção de regiões candidatas a conter um olho através do algoritmo de Viola-Jones;
4. Processamento por redes neurais convolucionais para identificar se o olho está aberto ou fechado;
5. Contagem de frames para calcular se há distração ou sonolência.

Figura 10 – Fluxograma da lógica do algoritmo



Fonte: Próprio autor

4.1 Pré-processamento

A primeira etapa realizada é a de pré-processamento que tem a finalidade de preparar o frame para as próximas etapas. O pré-processamento realizado em cada frame trata-se da conversão do espaço de cores para grayscale e em seguida utilizar um filtro Gaussiano no frame já em grayscale para remover ruídos e melhorar a qualidade da imagem que servirá de entrada a ser processada pelo algoritmo de Viola-Jones. A conversão em grayscale é feita com o objetivo de permitir que o frame seja processado pelo algoritmo

Viola-Jones do OpenCV pois este requer entradas neste espaço de cores específico para ser utilizado.

4.2 Detecção Inicial da Face

A segunda etapa realizada é a de detecção do rosto, esta é uma etapa crucial para a funcionalidade do sistema ocorrer de forma correta e poder ser realizada a detecção do estado de distração ou de sonolência no motorista. Esta etapa é realizada através do algoritmo de Viola-Jones utilizando de entrada a saída da etapa de pré-processamento. O objetivo principal desta etapa é isolar o rosto do motorista dos demais objetos e outros rostos que possam estar presentes na cena, para isto é utilizado o algoritmo de Viola-Jones.

Primeiramente feita a assumpção de que o posicionamento da câmera foi feito de forma que o rosto do motorista será o rosto de maior tamanho presente na imagem portanto sendo a região de maior tamanho que será detectada pelo algoritmo de Viola-Jones. A segunda assumpção realizada é de que o rosto do motorista está a uma distância mínima da câmera e portanto regiões com um tamanho menor que um tamanho mínimo devem ser ignoradas por não se tratarem de regiões que possam conter o rosto do motorista.

Com base nestas assumpções, somente a maior região detectada e que tenha um tamanho que seja pelo menos um valor mínimo pré-determinado é válida para conter o rosto do motorista. Se não é encontrada uma região que atenda a estas demandas, deve-se assumir que não há um motorista presente ou que o motorista está com comportamento distraído.

4.3 Detecção de Regiões com Olhos

A terceira etapa realizada é a de detecção de regiões que contém olhos. Esta é uma etapa que utiliza do algoritmo de Viola-Jones para extrair as regiões que serão processadas pela rede neural convolucional para a determinação de distração e sonolência do motorista na direção veicular.

A etapa é iniciada com o isolamento de uma subregião do rosto a qual é calculada removendo-se aproximadamente 41% da parte inferior da região original para ser processada pelo algoritmo de Viola-Jones que retorna as regiões candidatas a ter olhos. Isto é realizado com o objetivo de reduzir a quantidade de operações a serem realizadas e de diminuir a quantidade de regiões não relevantes detectadas pois o algoritmo de Viola-Jones para detecção de olhos não é tão preciso quanto o para detecção de rostos. Então é criada uma lista a qual é inserida estas regiões recortadas do frame original após o redimensionamento para a resolução $64px$ por $64px$ e o espelho no eixo horizontal deste o que resulta em um vetor de $[2X, 64, 64, 3]$ no qual X é a quantidade de regiões candidatas que foram obtidas.

O motivo de também ser analisado o espelho no eixo horizontal da região do olho é que diminui-se a quantidade de treino a ser realizado na rede pois só se faz necessário treinar para um dos lados devido a simetria do dado analisado. Também há a possibilidade de nenhuma região ser detectada, o que resulta em uma lista vazia.

4.4 Identificação do Estado dos Olhos

A quarta etapa realizada é a de identificação do estado dos olhos a qual é realizada pela rede neural convolucional. É utilizada como entrada para a rede neural a lista gerada na etapa anterior e para cada região na lista é verificado a qual classe ela pertence, sendo

que quando há uma região pertencente a classe de olho aberto uma flag é limpa para indicar para a etapa de geração de alertas que este frame não é um frame de distração ou sonolência.

O valor de predição (grau em que entrada pertence a determinada classe) que foi usado para a comparação foi de 0.9, então caso haja uma região em que ela ou o espelho no eixo horizontal possua o valor de predição para a classe de olho aberto maior que 0.9 esta região é tratada como um olho aberto.

4.5 Algoritmo de Geração de Alertas

A quinta etapa realizada é a de geração de alertas, esta etapa é centrada em torno de um contador inicializado no início do sistema o qual é denominado contador de distração e sonolência. Este contador de distração e sonolência é um contador cumulativo que é responsável pela contagem de frames que há distração e sonolência presentes.

O funcionamento do contador ocorre da seguinte forma: é verificado um flag que é setado no início do loop principal e é limpo na etapa anterior caso o evento esperado ocorra (se é encontrado um olho aberto), se o flag está setado então o contador é incrementado, caso contrário, isto é o flag está limpo é verificado se o contador está com um valor maior que 0 e se sim este é decrementado. Após isto é verificado se o contador está com o valor maior que um valor determinado (foi utilizado o valor 30), se sim é gerado um alerta e então é verificado se o contador está acima do valor máximo (foi usado o valor 40) e caso estiver é atribuído o valor máximo ao contador.

O motivo de ser feito desta forma é que assim obtém-se um período de graça antes de ser gerado o alerta o que permite movimentações naturais de direção como olhar aos lados por curtos instantes e de ações como piscar os olhos sem ser gerado um alerta pois estes tipos de comportamento não são comportamentos típicos de distração e de sonolência, porém caso haja uma extensão muito longa então pode-se considerar uma situação válida para a geração de alerta.

O valor limite máximo é para que o sistema pare de gerar o alerta após o motorista voltar a ter foco na pista, que também é esperado que ocorra ao longo de múltiplos frames portanto o motivo de ter sido utilizado uma faixa para gerar o alerta ao invés de somente o valor máximo.

4.6 Arquitetura da Rede Neural Convolutacional

Foi utilizada uma rede neural convolutacional (CNN ou Convolutional Neural Network) no framework Tensorflow de seguinte configuração para fazer a detecção do estado do olho: um layer convolutacional 2D (filtros: 64; kernel: 3; strides: 3) conectado a um layer convolutacional 2D (filtros: 8; kernel: 3; strides: 3) conectado a uma camada de max pooling (2 x 2) conectado a um layer denso de dimensão 128 conectado a outro layer denso de dimensão 2 o qual é a saída que indica em qual estado está o olho do motorista.

5 Resultados Experimentais e Validação

5.1 Dataset

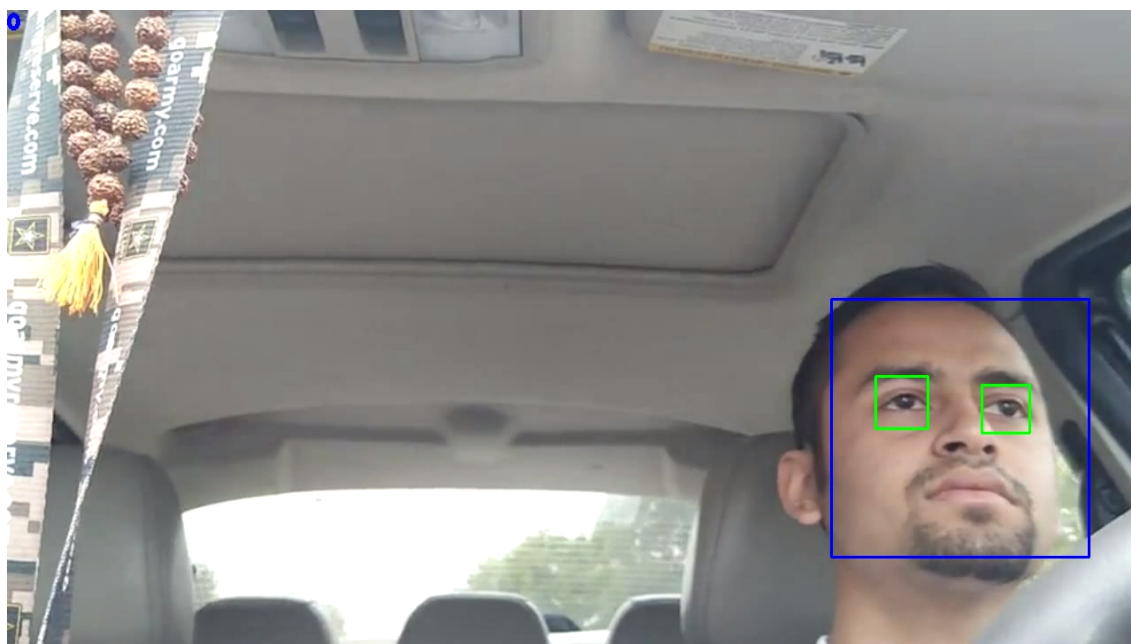
O dataset utilizado no experimento foi o UTA Real-Life Drowsiness Dataset (GHODDOOSIAN; GALIB; ATHITSOS, 2019) o qual contém 60 participantes e para cada participante possui 3 vídeos mas somente 36 dos 60 participantes forneceram permissão para mostrar seus rostos em publicações.

5.2 Resultados Experimentais

Nesta subsecção são apresentados algumas saídas da rede, para os *bounding boxes* dos olhos verde representa que a rede detecta o olho como estando aberto e vermelho como o olho estando fechado e o *bounding box* azul representa a região detectada como a região em que está presente o rosto. Os resultados experimentais obtidos pela abordagem proposta são também publicizados através do link: <<https://youtu.be/oaQ2YoqAt0g>>.

Abaixo na Figura 11 temos o motorista com foco na pista e detecção ocorre de forma correta, sendo detectado ambos os olhos pelo algoritmo de Viola-Jones (etapa 3) e sendo obtida a saída correta pela rede neural convolucional (etapa 4) de que o motorista está focado na pista.

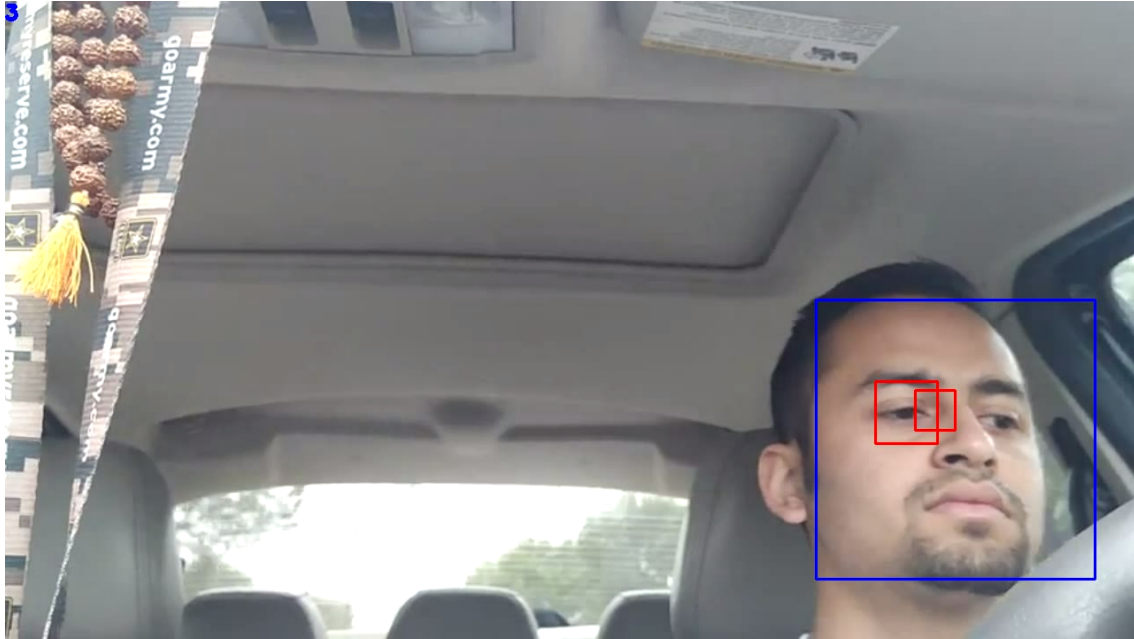
Figura 11 – Motorista dirigindo corretamente



Fonte: Próprio autor, saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 4 vídeo 00

Já na Figura 12 abaixo temos o motorista com um olhar não focado na pista não sendo gerado um alerta pois o evento ainda está sendo de curta duração com o contador de distração contendo somente o valor de 3. Podemos observar que há um olhar que aparenta estar com foco fora da pista, provavelmente o motorista está olhando para o velocímetro neste instante. Apesar de somente o olho do lado direito do rosto ter sido detectado pelo algoritmo de Viola-Jones, o sistema julga de forma correta qual estado classificar o motorista neste frame.

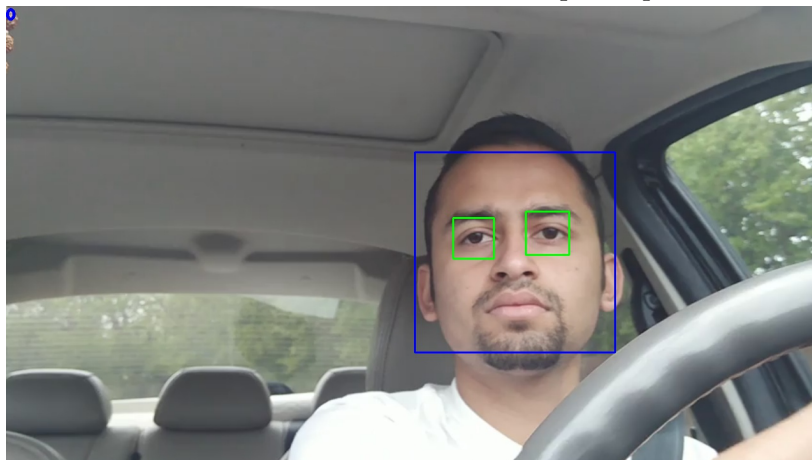
Figura 12 – Motorista com o foco fora da estrada



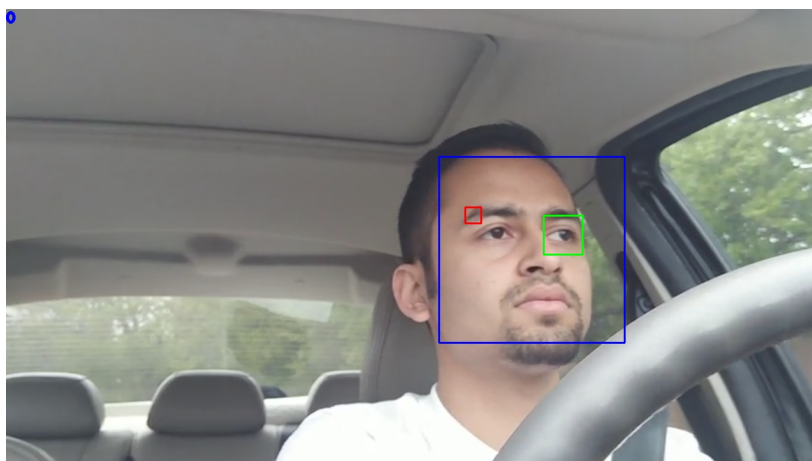
Fonte: Próprio autor, saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 4 vídeo 00

Abaixo, as Figuras 13 e 14 mostram uma sequência de frames da saída do sistema do vídeo em que o motorista está dirigindo. Através da progressão da sequência de frames de número 100 a 200 pode ser observado que apesar de nem todos os olhos terem sido detectados em todos os frames pelo algoritmo de Viola-Jones com alguns dos frames contendo somente um dos olhos, o sistema ainda obteve a saída correta para o estado de alerta do motorista. Isto ocorre devido a forma de funcionamento do contador de frames de distração e sonolência e ao fato de o algoritmo requerer somente um olho com estado de aberto para não classificar o frame como um de distração ou sonolência.

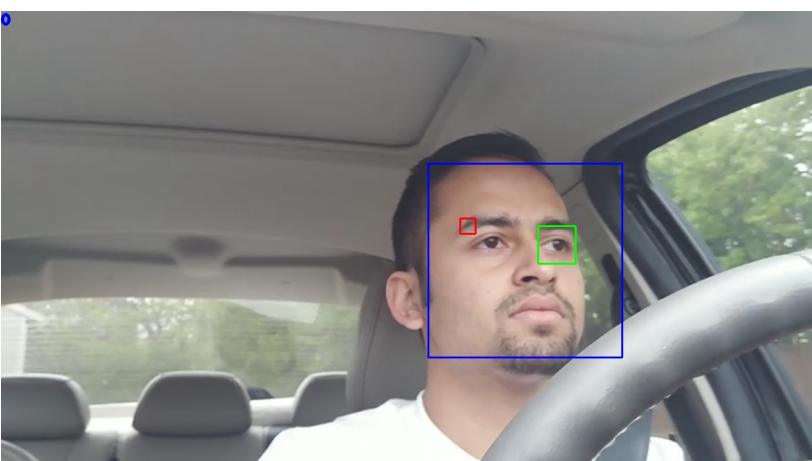
Figura 13 – Sequência de frames de saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 4 vídeo 00



(a) Imagem frame número 100, é observado que ambos os olhos foram detectados sem problemas neste frame.



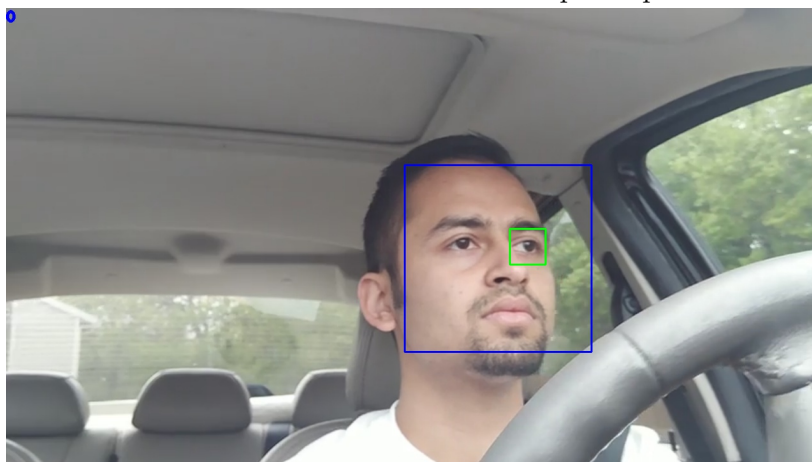
(b) Imagem frame número 120, é observado que o olho do lado direito do rosto não foi detectado pelo algoritmo de Viola-Jones na terceira etapa.



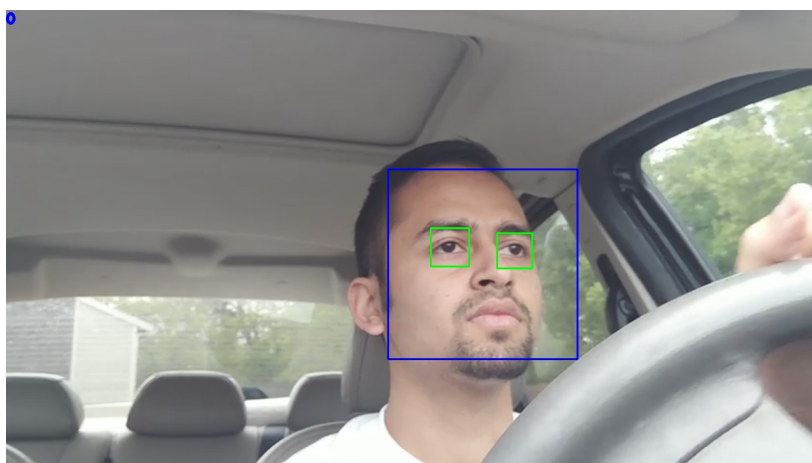
(c) Imagem frame número 140, a situação em relação a detecção do olho do lado direito do rosto continua a mesma a da imagem do frame de número 120.

Fonte: Próprio autor

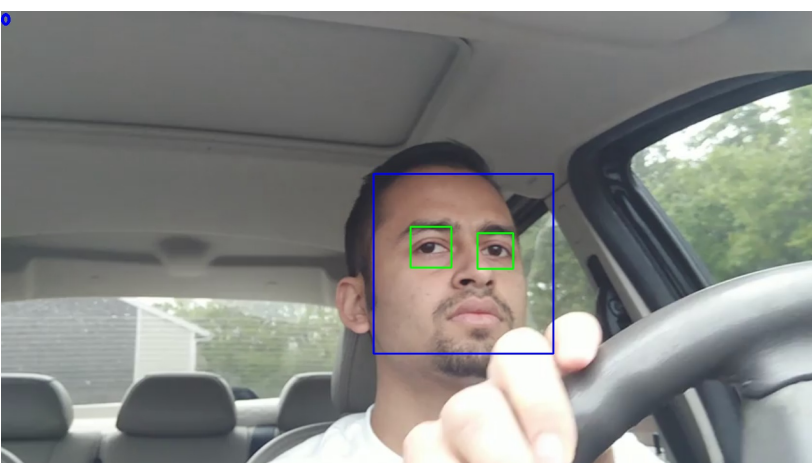
Figura 14 – Sequência de frames de saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 4 vídeo 00



(a) Imagem frame número 160, igualmente a o que ocorreu no frame de número 120 e 140, o olho do lado direito do rosto não foi detectado pelo algoritmo de Viola-Jones na terceira etapa.



(b) Imagem frame número 180, a partir deste frame ambos os olhos foram detectados de forma correta pelo algoritmo de Viola-Jones.

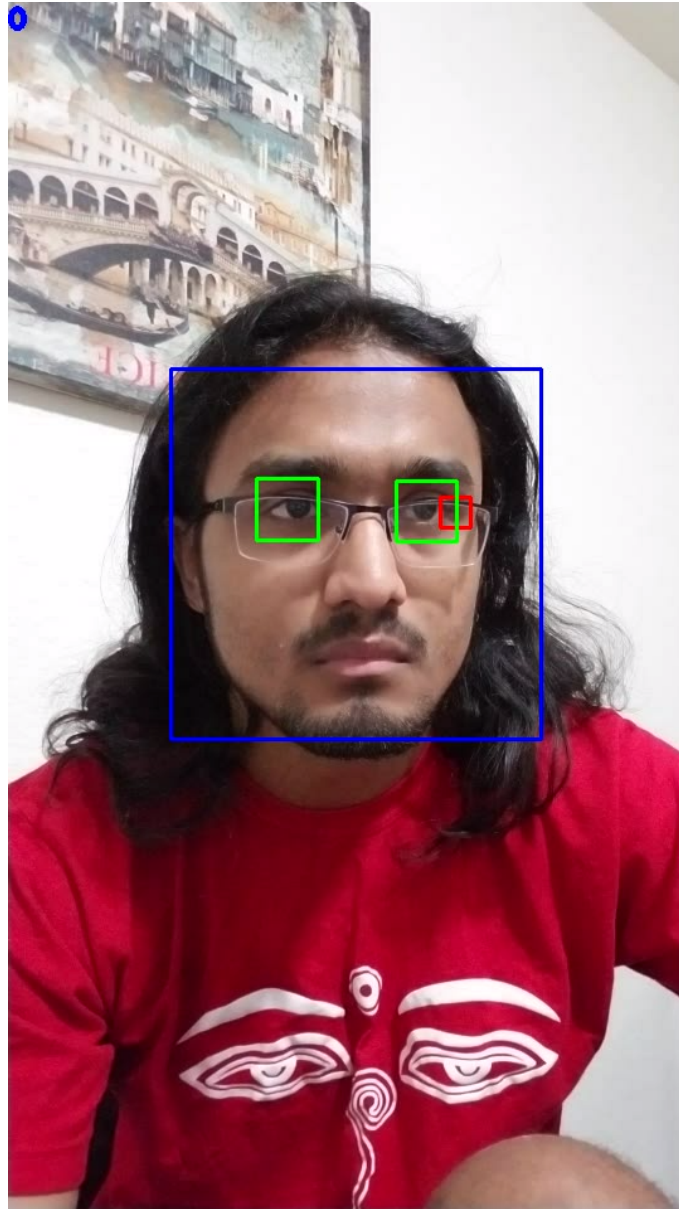


(c) Imagem frame número 200, ambos os olhos continuaram a ser detectados de forma correta pelo algoritmo de Viola-Jones assim como no frame de número 180.

Fonte: Próprio autor

Abaixo na Figura 15 temos a situação em que o motorista está de óculos. O sistema continua mantendo sua capacidade de detecção da posição e situação dos olhos.

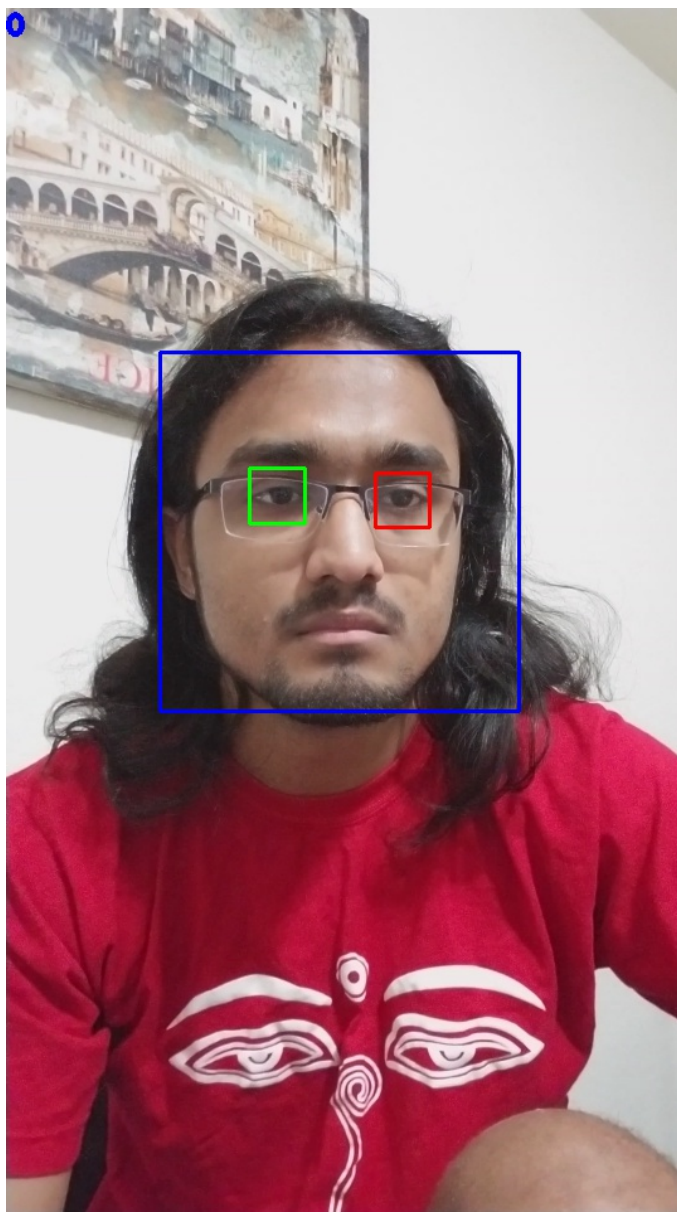
Figura 15 – Teste com usuário de óculos



Fonte: Próprio autor, saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 9 vídeo 10

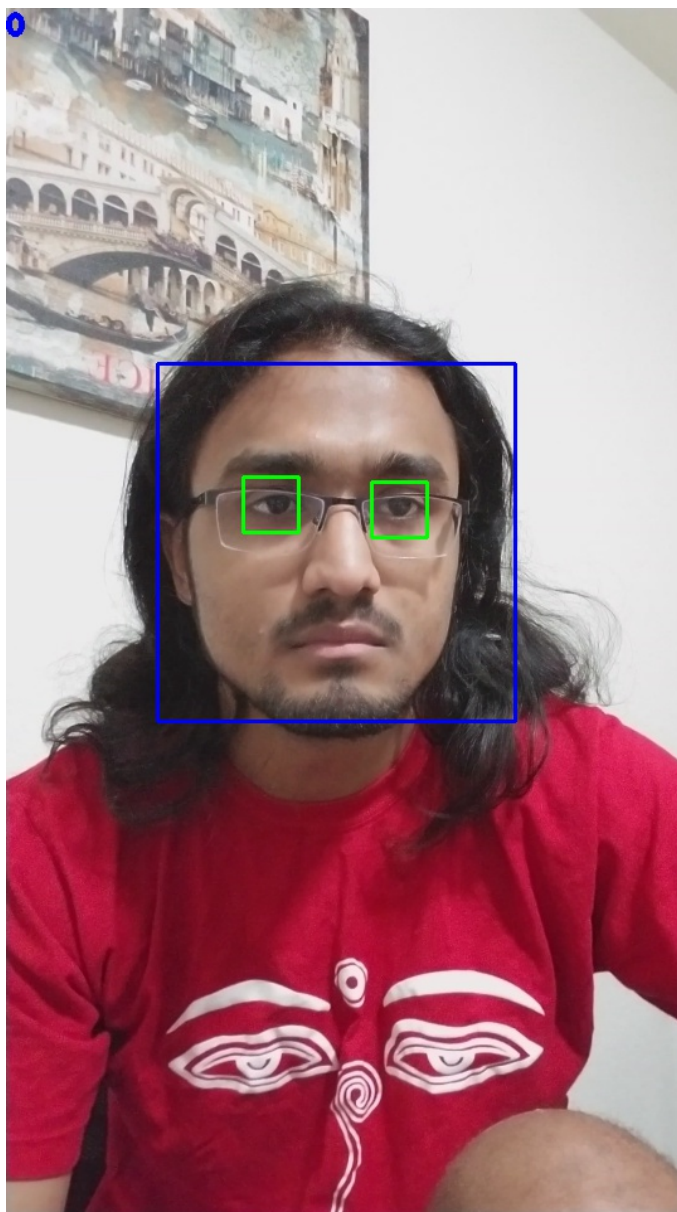
As Figuras 16, 17 e 18 mostram uma sequência de frames do vídeo em que é testado o funcionamento do sistema na situação em que o motorista esteja utilizando óculos.

Figura 16 – Sequência de frames de saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 9 vídeo 10, imagem frame número 100



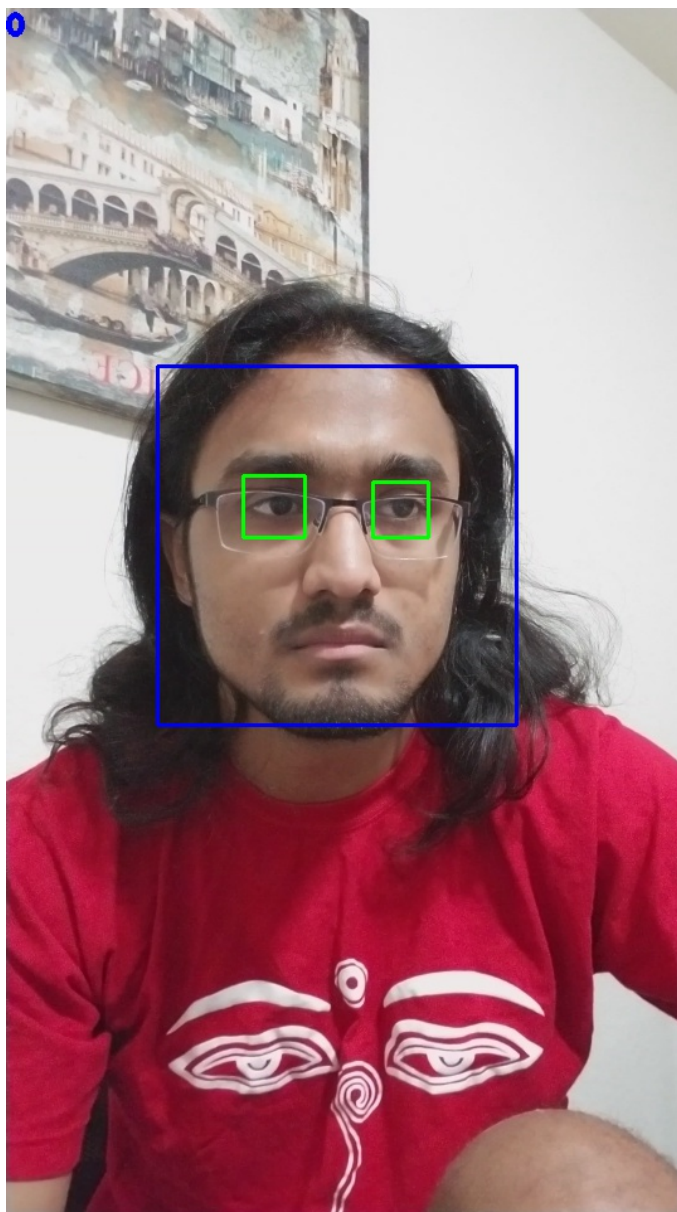
Fonte: Próprio autor

Figura 17 – Sequência de frames de saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 9 vídeo 10, imagem frame número 120



Fonte: Próprio autor

Figura 18 – Sequência de frames de saída do sistema utilizando como entrada o vídeo do dataset UTA Real-Life Drowsiness Dataset participante 9 vídeo 10, imagem frame número 140



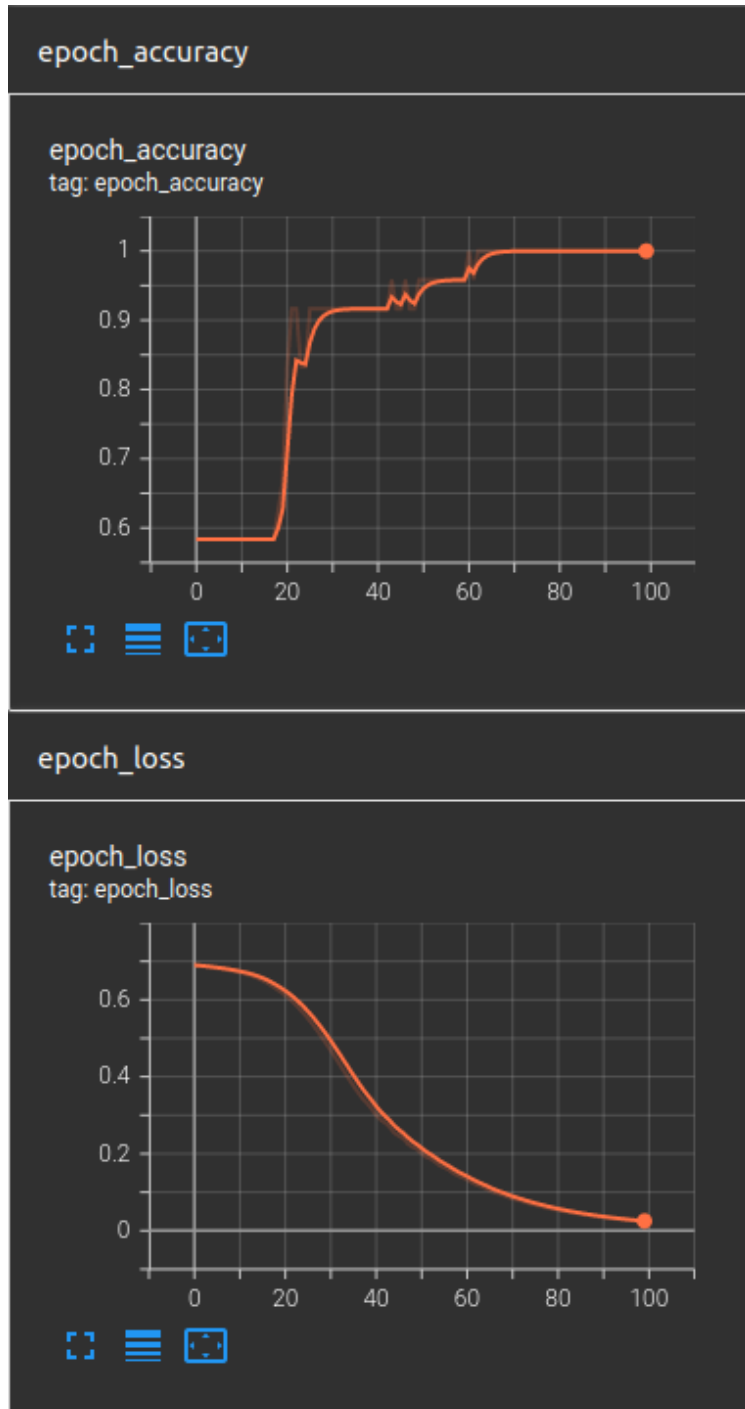
Fonte: Próprio autor

Como pode ser observado em um ambiente no qual não há glare no óculos, o sistema tem a capacidade de detecção mesmo que o motorista esteja utilizando óculos. Caso haja glare não é esperado que o sistema funcione corretamente pois as distorções causadas na imagem podem causar inclusive humanos de não conseguir distinguir o estado do olho, simplesmente por não haver visibilidade do olho em si devido a reflexão.

5.3 Validação

Segundo os gráficos gerados pelo Tensorboard a rede convergiu para precisão máxima no set de treino.

Figura 19 – Gráficos gerados pelo Tensorboard da precisão de cada época e perda em cada época durante o treino



Fonte: Próprio autor

Para testar a precisão do sistema foram marcados em dois vídeos os períodos em que há frames que devem de gerar o resultado de distração baseado em como o sistema foi construído e após isto foi comparada a saída do sistema com estes períodos com o qual foi obtida uma precisão de 95,41% em um vídeo e de 84,59% no outro vídeo.

5.4 Performance em Tempo de Execução

Durante os testes em uma máquina com as seguintes especificações: Intel® Core™ i7-10700F CPU @ 2.90GHz × 16, 32GB de memória RAM, GPU GeForce GTX 1660 (6GB RAM) em sistema operacional Linux Ubuntu 22.04.02 LTS utilizando o ambiente Python Anaconda com o framework Tensorflow e a biblioteca OpenCV foi obtido o tempo médio de 80068 microsegundos (aproximadamente 80 milissegundos) para a execução de cada frame pelo algoritmo, analisando a utilização de recursos pela máquina em que a utilização da CPU não ultrapassou 75% em nenhum dos núcleos e que o máximo do poder de processamento usado da GPU foi de 30% pode-se inferir que o gargalo se tratava de algum delay mínimo de comunicação na parte de envio dos frames para a GPU processar com a rede neural o qual é inerente da própria arquitetura computacional para executar processamento em GPUs.

Figura 20 – Uso da CPU e memória RAM durante a execução algorítmica



Fonte: Próprio autor

6 Conclusão e Trabalhos Futuros

Em grande parte o sistema se comportou de forma desejada nos testes realizados obtendo as precisões de 95% e 84%, o que é bem promissor porém também demonstra as limitações dos métodos clássicos os quais estão entrando em obsolescência, principalmente da utilização do algoritmo de Viola-Jones para a detecção da posição dos olhos o qual pode ocasionar problemas para a tarefa de assistência a um motorista que tenha somente um dos olhos pois houve a ocorrência de não detecção da posição de um dos olhos ou detecção de regiões que não continham olhos como contendo um olho o que causa desperdício de recursos de processamento. Já para a detecção da localização do rosto o algoritmo de Viola-Jones se mostrou bem adequado e resiliente para a tarefa após a calibração dos parâmetros para a distância do motorista da câmera.

As Redes Neurais Convolucionais se mostraram bem eficientes na detecção do estado do olho e realizaram um bom papel na classificação dos olhos mesmo quando o algoritmo de Viola-Jones retornou regiões sem um olho como contendo um olho. Entretanto, se uma região que contém o olho não é detectada pelo algoritmo de Viola-Jones então ela não será processada pela CNN, o que significa que para melhorar o desempenho do sistema em trabalhos futuros terá de ser modificado o método para a localização da posição dos olhos.

A utilização de um temporizador cumulativo para a tarefa de contagem de frames em que há distração e sonolência se provou como sendo uma excelente estratégia para detectar distrações e situações de sonolência sem punir o motorista por realizar tarefas normais de condução como olhar aos lados para manobragem, verificar a velocidade marcada no velocímetro dentre outras situações normais a quais não deve-se causar um alarme por ocorrerem.

Idealmente para trabalhos futuros as modificações realizadas seriam de substituir a parte de detecção de rosto e de posição dos olhos por soluções totalmente baseadas em redes neurais artificiais as quais neste momento se mostram promissoras em conseguir realizar esta tarefa, mas devido a complexidade para realizar estas tarefas no ambiente de execução proposto requerem muitos recursos humanos e computacionais para desenvolver.

Referências

ABRAMET. *Problemas na saúde de motoristas causaram mais de 280 mil acidentes nas rodovias desde 2014, aponta Abramet*. 2020. Acesso em: 07 Jun. 2023. Disponível em: <<https://www.abramet.com.br/noticias/problemas-na-saude-de-motoristas-causaram-mais-de-280-mil-acidentes-nas-rodovias-desde-2014-aponta-abramet>>. Citado na página [7].

Adrian Rosebrock. *Watershed OpenCV*. 2015. Acesso em: 3 Out. 2023. Disponível em: <<https://pyimagesearch.com/2015/11/02/watershed-opencv/>>. Citado na página [16].

Aldo von Wangenheim. *Combinando Deslocamento para a Média com Detectores de Bordas (EDISON)*. Acesso em: 3 Out. 2023. Disponível em: <<https://lapix.ufsc.br/ensino/visao/combinando-medias-moveis-com-detectores-de-bordas-edison/>>. Citado na página [15].

ANSARI, M.; KURCHANIYA, D.; DIXIT, M. A comprehensive analysis of image edge detection techniques. *International Journal of Multimedia and Ubiquitous Engineering*, v. 12, p. 1–12, 11 2017. Citado na página [11].

BEARLY, E. M.; CHITRA, R. Automatic drowsiness detection for preventing road accidents via 3dgan and three-level attention. *Multimedia Tools and Applications*, Nov 2023. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-023-17272-y>>. Citado na página [10].

BIGAS, M. et al. Review of cmos image sensors. *Microelectronics Journal*, v. 37, n. 5, p. 433–451, 2006. ISSN 0026-2692. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0026269205002764>>. Citado na página [10].

BMW. *BMW model upgrade measures taking effect from the summer of 2013*. 2013. Acesso em: 08 Nov. 2022. Disponível em: <<https://www.press.bmwgroup.com/global/article/detail/T0141144EN/bmw-model-upgrade-measures-taking-effect-from-the-summer-of-2013>>. Citado na página [7].

BRASIL. *LEI Nº 13.103, DE 2 DE MARÇO DE 2015*. Brasília, DF: Presidente da República, 2015. Dispõe sobre o exercício da profissão de motorista; altera a Consolidação das Leis do Trabalho - CLT[...]. Acesso em: 10 Dec. 2023. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/113103.htm>. Citado na página [7].

CANAL, F. Z. *Reconhecimento de Expressão Faciais Baseado em Redes Neurais Convolucionais para Aplicação no Sistema Tutor Inteligente MAZK*. 2021. TCC(graduação) - Universidade Federal de Santa Catarina. Campus Araranguá. Engenharia da Computação. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/223660>>. Citado na página [18].

Câmara dos Deputados. *MAIO Amarelo alerta: 40 mil morrem por ano no Brasil em acidentes de trânsito*. Câmara dos Deputados. 2017. Acesso em: 20 Fev. 2023. Disponível em: <<https://www.camara.leg.br/radio/programas/512920-maio-amarelo-alerta-40-mil-morrem-por-ano-no-brasil-em-acidentes-de-transito/>>. Citado na página [7].

D'ANTUONO, R. Basic digital image acquisition, design, processing, analysis, management, and presentation. In: NECHYPORUK-ZLOY, V. (Ed.). *Principles of Light Microscopy: From Basic to Advanced*. Cham: Springer International Publishing, 2022. p. 77–104. ISBN 978-3-031-04477-9. Disponível em: <https://doi.org/10.1007/978-3-031-04477-9_4>. Citado na página [10].

DAVIES, E. Chapter 24 - statistical pattern recognition. In: DAVIES, E. (Ed.). *Machine Vision (Third Edition)*. Third edition. Burlington: Morgan Kaufmann, 2005, (Signal Processing and its Applications). p. 687–723. ISBN 978-0-12-206093-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780122060939500277>>. Citado na página [8].

DAVIES, E. Chapter 27 - image acquisition. In: DAVIES, E. (Ed.). *Machine Vision (Third Edition)*. Third edition. Burlington: Morgan Kaufmann, 2005, (Signal Processing and its Applications). p. 781–803. ISBN 978-0-12-206093-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780122060939500307>>. Citado na página [10].

DAVIES, E. Chapter 5 - edge detection. In: DAVIES, E. (Ed.). *Machine Vision (Third Edition)*. Third edition. Burlington: Morgan Kaufmann, 2005, (Signal Processing and its Applications). p. 131–157. ISBN 978-0-12-206093-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780122060939500083>>. Citado na página [8].

DAVIES, E. Chapter 6 - binary shape analysis. In: DAVIES, E. (Ed.). *Machine Vision (Third Edition)*. Third edition. Burlington: Morgan Kaufmann, 2005, (Signal Processing and its Applications). p. 159–205. ISBN 978-0-12-206093-9. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780122060939500095>>. Citado na página [8].

DENG, Y.; MANJUNATH, B. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 8, p. 800–810, 2001. Citado (2) vezes nas páginas [14 e 15].

FERRASSINI, R. B. *Sistema de detecção de sonolência, por meio de visão computacional*. 2014. Trabalho de Conclusão de Curso - Escola de Engenharia de São Carlos - USP. Disponível em: <<http://www.tcc.sc.usp.br/tce/disponiveis/18/180450/tce-29012015-154320/?&lang=br>>. Citado na página [17].

GHODDOOSIAN, R.; GALIB, M.; ATHITSOS, V. A realistic dataset and baseline temporal model for early drowsiness detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2019. p. 0–0. Citado na página [22].

HU, Z. et al. Face orientation estimation for driver monitoring with a single depth camera. In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. [S.l.: s.n.], 2015. p. 958–963. Citado na página [9].

HUANG, J.; LIU, Y.; PENG, X. Recognition of driver’s mental workload based on physiological signals, a comparative study. *Biomedical Signal Processing and Control*, v. 71, p. 103094, 2022. ISSN 1746-8094. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1746809421006911>>. Citado na página [9].

KIEFER, L.; STORATH, M.; WEINMANN, A. PALMS Image Partitioning - A New Parallel Algorithm for the Piecewise Affine-Linear Mumford-Shah Model. *Image Processing On Line*, v. 10, p. 124–149, 2020. <<https://doi.org/10.5201/ipol.2020.295>>. Citado na página [16].

LECUN, Y.; KAVUKVUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: *Proc. International Symposium on Circuits and Systems (ISCAS’10)*. [S.l.]: IEEE, 2010. Citado na página [17].

LI, W. et al. A survey on vision-based driver distraction analysis. *Journal of Systems Architecture*, v. 121, p. 102319, 2021. ISSN 1383-7621. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1383762121002174>>. Citado na página [8].

NAIDU P. LAKSHMAN RAO, M. S. P. B. K. V. L. B. B. R. Framework for efficient edge detection techniques—comparison among robert, rewitt, sobel, robinson, kirsch and canny. *International Journal of Computer Science And Technology*, v. 3, n. 2, 2012. Citado na página [11].

NIEH, M. *Ford’s Wake-Up Call for Europe’s Sleepy Drivers*. 2011. Acesso em: 08 Nov. 2022. Disponível em: <https://web.archive.org/web/20110513232258/http://media.ford.com/article_print.cfm?article_id=34562>. Citado na página [7].

NUEVO, J. et al. Real-time robust face tracking for driver monitoring. In: *2006 IEEE Intelligent Transportation Systems Conference*. [S.l.: s.n.], 2006. p. 1346–1351. Citado na página [9].

Organização Pan-Americana da Saúde. *OMS lança Década de Ação pela Segurança no Trânsito 2021-2030*. 2021. Acesso em: 20 Fev. 2023. Disponível em: <<https://www.paho.org/pt/noticias/28-10-2021-oms-lanca-decada-acao-pela-seguranca-no-transito-2021-2030>>. Citado na página [7].

PULFER, E.-M. *Different Approaches to Blurring Digital Images and Their Effect on Facial Detection*. 2019. Computer Science and Computer Engineering Undergraduate Honors Theses Disponível em: <<https://scholarworks.uark.edu/csceuht/66>>. Citado (2) vezes nas páginas [10 e 11].

REZAEI, M.; KLETTE, R. 3d cascade of classifiers for open and closed eye detection in driver distraction monitoring. In: REAL, P. et al. (Ed.). *Computer Analysis of Images and Patterns*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. p. 171–179. ISBN 978-3-642-23678-5. Citado na página [9].

SAHOO, G. K.; DAS, S. K.; SINGH, P. A deep learning-based distracted driving detection solution implemented on embedded system. *Multimedia Tools and Applications*, v. 82, n. 8, p. 11697–11720, Mar 2023. ISSN 1573-7721. Disponível em: <<https://doi.org/10.1007/s11042-022-13450-6>>. Citado na página [10].

Sanjay Kumar, V. et al. Smart driver assistance system using raspberry pi and sensor networks. *Microprocessors and Microsystems*, v. 79, p. 103275, 2020. ISSN 0141-9331. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0141933120304348>>. Citado na página [8].

SINGH, P.; MAHIM, S.; PRAKASH, S. Real-time ear based drowsiness detection model for driver assistant system. In: *2022 International Conference on Signal and Information Processing (IconSIP)*. [S.l.: s.n.], 2022. p. 1–4. Citado na página [9].

SOBIERANSKI, A. C. et al. An anisotropic diffusion filtering implementation to execute in parallel distributed systems. In: *2008 11th IEEE International Conference on Computational Science and Engineering - Workshops*. [S.l.: s.n.], 2008. p. 182–187. Citado (4) vezes nas páginas [11, 14, 15 e 16].

TYAGI, N.; GOYAL, R.; RAUTELA, R. Real time drowsiness detection system. In: *2022 International Conference on Cyber Resilience (ICCR)*. [S.l.: s.n.], 2022. p. 1–3. Citado na página [9].

VAZ, V. C. *Obtenção de Batimetria em Estuários Através da Aplicação de Métodos de Aprendizado de Máquina em Imagens Multiespectrais de Satélites*. 2021. TCC(graduação) - Universidade Federal de Santa Catarina. Campus Araranguá. Engenharia da Computação. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/223658>>. Citado na página [17].

VIOLA, M. J. P. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2001. p. 1–511–1–518. ISBN 0-7695-1272-0. Citado na página [8].

YUEN, K.; MARTIN, S.; TRIVEDI, M. M. Looking at faces in a vehicle: A deep cnn based approach and evaluation. In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. [S.l.: s.n.], 2016. p. 649–654. Citado (2) vezes nas páginas [8 e 9].

ZHAO, Y. et al. Improved watershed algorithm for dowels image segmentation. In: *2008 7th World Congress on Intelligent Control and Automation*. [S.l.: s.n.], 2008. p. 7644–7648. Citado na página [15].

ZIOU, D.; TABBONE, S. 'edge detection techniques: An overview'. *International Journal of Pattern Recognition and Image Analysis*, v. 4, p. 537–559, 01 1998. Citado na página [11].