

Classificação dos Olhos utilizando Rede Neural Convolucional para Detecção de Sonolência

Cézar Augusto Gobbo Passamani
Departamento de Informática
(UFES) Universidade Federal do Espírito Santo
Vitória, Brasil
gobbo.cezar@lcad.inf.ufes.br

Alberto Ferreira de Souza
Departamento de Informática
(UFES) Universidade Federal do Espírito Santo
Vitória, Brazil
alberto@lcad.inf.ufes.br

Abstract—As condições de sonolência no corpo humano podem afetar a mudança de algumas partes do comportamento do corpo, como a mudança de comportamento de olhos, os quais estarão fechados ao dormir. Essa mudança de comportamento dos olhos trás o benefício de detectar o estado de dormindo. Nesta pesquisa, coletamos imagens de olhos para construir um conjunto de dados e separar esse conjunto de dados em quatro classes com base em nossas classificações, essas classes são "open", "closed", "forced" e "na". Assim, existem 4 conjuntos de dados diferentes nesta pesquisa, o primeiro conjunto de dados para a classe open contém 139152 imagens, o segundo conjunto de dados para a classe closed contém 157396 imagens, o terceiro conjunto de dados para a classe forced contém 124080 imagens, e o quarto conjunto de dados para a classe na contém 155000 imagens, esse conjunto de dados foi coletado de aproximadamente 20 pessoas, em diferentes iluminações e distâncias da camera. O algoritmo de detecção foi executado no Raspberry Pi 3.

Keywords—convolutional neural network, sleeping detection, raspberry pi, lenet, eyes

I. INTRODUÇÃO

Os acidentes de trânsito são os problemas comuns que todos os países enfrentam, uma pessoa que morre a cada 25 segundos em acidentes de trânsito [1]. Em 2015, a OMS divulgou que a cada ano em todo o mundo morriam mais de 1,25 milhão de pessoas e 50 milhões de feridos graves em acidentes de trânsito [2]. A maioria dos sobreviventes de acidentes de trânsito é pobre e precisa de reabilitação mental para curar seus traumas [3]. O governo sugere descansar a cada 4 horas durante uma longa viagem para refrescar o corpo e a mente, mas se os motoristas não obedecerem a essas sugestões, infelizmente isso pode transformar suas condições mental e física em fadiga e sonolência. Nesta pesquisa, propomos a detecção de sonolência usando rede neural convolucional (CNN) a partir de um conjunto de dados de imagens.

II. MÉTODO

O diagrama baixo mostra o passo a passo da pesquisa.

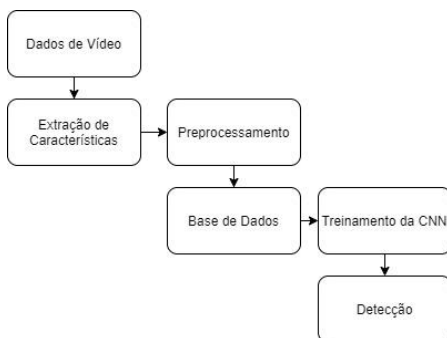


Figura 1. Diagrama de passo a passo da pesquisa.

A. Dados de Vídeo

Os dados de vídeo neste estudo, são de muitos vídeos de olhos em tempo real, e variações desses vídeos são necessárias para construir um bom conjunto de dados. Variações desses vídeos são iluminação (brilho e escuridão), distância entre o rosto e a câmera em relação aos eixos x, y e z. Essas informações foram usados para construir um conjunto de dados usando o método detecção de landmarks do rosto da Dlib [4] para extração de características. A Dlib utiliza o método haar cascade [5] para fazer a detecção do rosto.

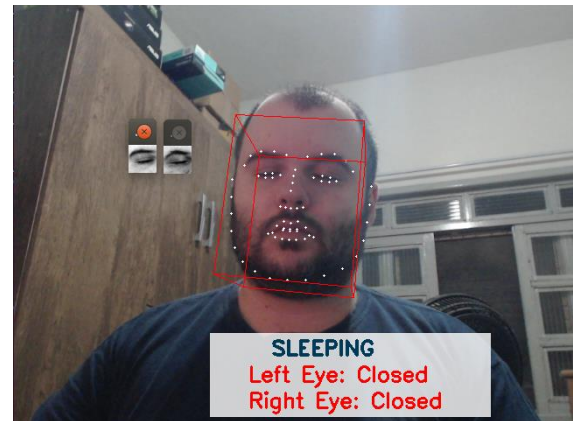


Figura 2. Figura mostrando a detecção feita pela Dlib utilizando haar cascade, e a classificação dos olhos no estado closed feita pela Lenet.

B. Extração de Características e préprocessamento.

Nesta seção, enquanto a Dlib faz o reconhecimento dos landmarks da face, definimos o ROI a partir dos pontos do rosto que estão perto dos olhos, como mostra a figura 2, recortamos a imagem desse ROI e aplicamos uma equalização de histograma para aumentar o contraste dos olhos. A partir disso, a cada frame, redimensionamos a imagem para 32x32 e gravamos em uma pasta. Isto para os 2 olhos. Para o olho da esquerda, após aplicar o histograma, é feito um espelhamento na horizontal, assim podemos gravar o olho da direita na mesma pasta que o da esquerda, gerando assim a base de dados de imagens prontas para treino. Esse espelhamento é mostrado na figura 2, na imagem pequena recordada ao lado do rosto, onde o olho esquerdo foi flipado na horizontal.

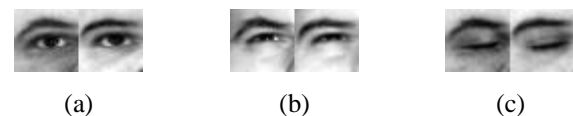


Figura 3. Exemplos de imagens da base de dados de 3 classes: (a) open, (b) forced e (c) closed.

C. Base de Dados

A base de dados contém 4 classes referentes a 4 pastas de imagens: a classe “open” para os olhos no estado aberto, a classe “closed” para os olhos no estado fechado, a classe “forced” para os olhos na classe “forçado”, e a classe “na” para as imagens negativas que não contém olhos. Com os experimentos do dia a dia, percebemos que quando tem sol sobre o rosto da pessoa, ou alguma fonte de luz forte, ela contrai os olhos de modo que eles fiquem quase fechados, mas de maneira “forçada”, exemplo na figura 4. Inicialmente o treino foi feito utilizando apenas 2 classes, open e closed, porém quando o sol, ou alguma fonte de luz forte, pairava sobre o rosto da pessoa e ela entrava no estado forçado, a rede classificava como closed, o que gerava falso-positivo. Assim com a nova classe “forced”, isso foi resolvido setando o estado para open. Alguns exemplos são mostrados na figura 3.

Atualmente, a base de dados contém 575628 imagens no total e pode ser encontrada no link: [base de olhos](#).

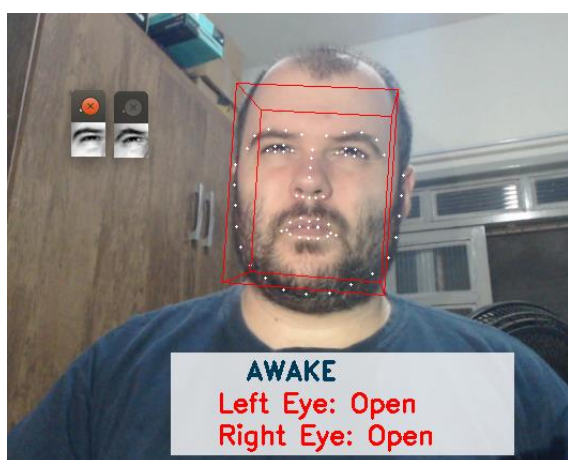


Figura 4. Classificação dos olhos no estado forced feita pela Lenet, devido à fonte de luz forte pairando sobre o rosto.

D. Treinamento da CNN

Com a base de dados pronta, escolhemos a rede neural convolucional projetada com a arquitetura LeNet [6], que é mostrada na figura 3, usando o framework Caffe para treinar a rede. Para isso, 80% (460502) das imagens do dataset foram usadas para treino e 20% (115126) para validação. O uso da LeNet é devido ao algoritmo ser executado no Raspberry Pi 3, que não possui um poder de processamento high-end, e como a LeNet é considerada uma rede light, não tivemos problema com falta de desempenho.

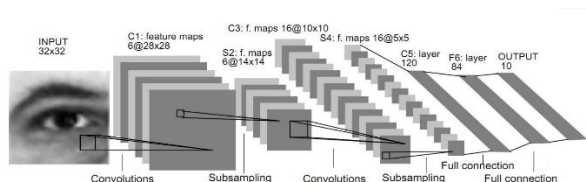


Figura 3. Exemplo de uma imagem do olho recortada pelo algoritmo e sendo enviada para a rede neural classificar.

O treino da rede neural foi realizado em um notebook portando um Core i7 e uma placa de video RTX 2060 (6GB). O treino teve uma acurácia de 0.984281 e uma perda de

0.049167, o que são resultados muito bons para o treino. Os gráficos da acurácia e da perda mostrados a seguir.

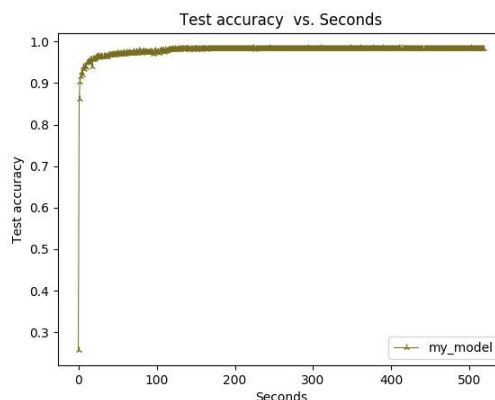


Figura 5. Evolução da acurácia do treino da rede, com um resultado de 0.984281.

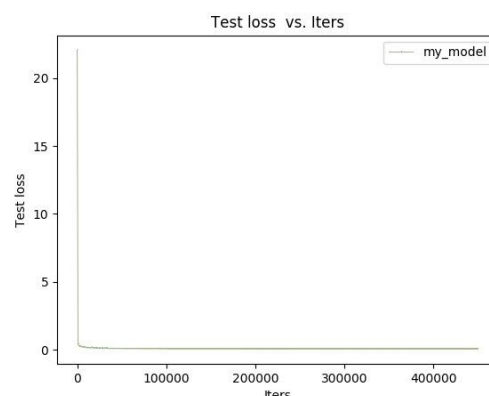


Figura 6. Evolução da perda do treino da rede, com um resultado de 0.049167.

Na tabela abaixo são mostrados alguns Hiperparâmetros da rede neural, e os resultados da acurácia e da perda.

Learning Rate	0.0001
Max Iterations	450000
N images train	460502
N images test	115126
Accuracy	98.42%
Loss	4.91%

Tabela 1. Hiperparâmetros da rede neural.

E. Algoritmo, Detecção e Experimento

Depois de treinar a rede, o algoritmo de detecção de sonolência irá carregar ela memória e realizar inferências, seguindo o fluxograma na figura 1. O algoritmo funciona da seguinte forma: o algoritmo inicia dizendo que os olhos estão abertos e a pessoa está acordada; para cada imagem que vem da câmera, os olhos são recortados; a equalização de histograma é aplicada nestes recortes para aumentar o contraste; os recortes são enviados para a rede neural que está carregada na memória; a rede classifica os olhos retornando uma das quatro classes (open, closed, forced ou na) para cada olho; se os 2 olhos forem classificados como closed simultaneamente, um contador de tempo é iniciado a partir de zero em uma variável estática, se durante 3 segundos ou mais

a classificação da rede continuar retornando closed para os 2 olhos, o algoritmo irá dizer que a pessoa está dormindo, como mostra a figura 2, caso a pessoa abra um ou mais olhos, a contagem irá voltar para 0 e o algoritmo irá dizer que a pessoa está acordada; por fim, o contador de tempo apenas voltará a iniciar depois que a rede classificar como closed os 2 olhos novamente. As figuras 2, 4 e 7 mostram o resultados das detecções feita pelo algoritmo utilizando a rede treinada.

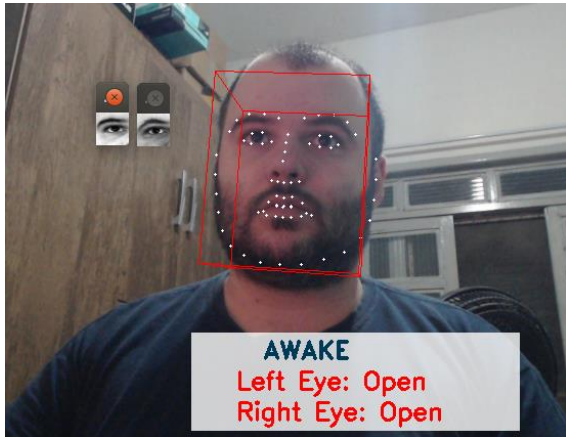


Figura 7. Classificação dos olhos no estado open feita pela Lenet.

III. REFERÊNCIAS

- [1] L. B. K. Donaldson, "Orthopaedic trauma from road crashes: is enough being done?," [Online]. Available: <http://scopus.com/record>.
- [2] "List of countries by traffic-related death rate," 2019. [Online]. Available: https://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate.
- [3] WHO, "Global Status Report on Road Safety," em *WHO Press*, 2015.
- [4] "Dlib C++ Library," Boost Software, [Online]. Available: <http://dlib.net/>.
- [5] M. J. Paul Viola, "Rapid Object Detection using a Boosted Cascade of Simple Features," em *Computer Vision and Pattern Recognition*, 2001.
- [6] L. B. Y. B. a. P. H. Yann LeCun, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, 1998.