UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

GABRIEL MORAES DA CRUZ

Relatório Final de Atividades de Iniciação Científica: Captura, processamento e análise de grandes massas de dados usando Algoritmos de Deep Learning.

São Paulo
29 de Abril de 2021
GABRIEL MORAES DA CRUZ

Relatório Final de Atividades de Iniciação Científica: Captura, processamento e análise de grandes massas de dados usando Algoritmos de Deep Learning.

Versão Original

Relatório final de atividades apresentado à Fundação da Universidade de São Paulo no projeto de iniciação científica oferecido através do Laboratório de Arquitetura e Redes de Computadores da Escola Politécnica da Universidade de São Paulo.

Coordenador do Projeto:
Prof. Dr. Wilson Vicente Ruggiero

Docente Responsável:
Profª. Drª. Graça Bressan

Mentor Responsável:
M. José Carlos Gutiérrez Menéndez

São Paulo
2021

FOLHA DE APROVAÇÃO

GABRIEL MORAES DA CRUZ

Relatório Final de Atividades de Iniciação Científica: Captura, processamento e análise de grandes massas de dados usando Algoritmos de Deep Learning.

Relatório final de atividades apresentado à Fundação da Universidade de São Paulo no projeto de iniciação científica oferecido através do Laboratório de Arquitetura e Redes de Computadores da Escola Politécnica da Universidade de São Paulo.

Aprovada em: __/__/2021

Examinador:

_____

RESUMO

A pesquisa foi conduzida de forma a entender o comportamento de estruturas de redes neurais profundas aplicada a reconhecimento facial em diversos tipos de etnias humanas. Para melhor entender esse comportamento foi utilizado seis tipos de bases de dados e dois métodos de aprendizado para comparar não só o desempenho entre os dados mas entre os métodos. Os resultados obtidos mostram que utilizando uma base pré-treinada somada a um treinamento em um banco de dados suficientemente grande é possível eliminar grandes variações de desempenho, diferentemente de estruturas neurais não inicializadas que não são previsíveis tampouco eficientes em ambientes onde há grande diversidade étnica.

Palavras-chave: Rede Neural; Reconhecimento Facial; Etnias; Base Pré-treinada.

Sumário

Introdução

Com a evolução dos métodos de Machine Learning um número maior de problemas vêm sendo resolvidos por meio dessa ferramenta que se mostra tão eficiente para lidar com desafios que apresentam um número exorbitante de dados.

O reconhecimento facial é um dos problemas que é mais atacado pelo método de Machine Learning pois uma imagem digitalizada é um grande amontoado de dados que precisa de um método bastante robusto para poder se adaptar aos mais variados cenários como quantidade de luz, ângulo de fotografia, apetrechos (tais como bonés, óculos, chapéus, etc.), tentativas de fraude, distância, qualidade de imagem, etc. São esses alguns dos problemas que devem ser resolvidos para conseguir-se alguma eficiência nesse tipo de problema.

O problema que é proposto neste trabalho é entender como a variedade de etnias influencia na eficiência de um algoritmo de Machine Learning que pretende classificar indivíduos e reconhecê-los posteriormente com uma taxa operacional, isto é, possível de implementação em um ambiente real para fins de identificação. O trabalho propõe não apenas entender o problema mas também apontar direções para que ele possa ser resolvido, seja por tratamento dos dados seja por estrutura neural ou algum outro caminho.

Utilizar algoritmos Deep Learning para classificar grupos étnicos é um desafio grande, como mostra o trabalho *Ahmed et al, N. 2020 Race estimation with deep networks. Journal of King Saud University - Computer and Information Science*. Por algum motivo essas estruturas têm grande dificuldade em lidar com o conceito de raça, que para um ser humano é de tão fácil reconhecimento. Entretanto a abordagem desse trabalho não é a classificação racial em si mas o comportamento de uma mesma estrutura dentro de diversos grupos raciais.

Metodologia

O banco de dados

O primeiro passo para resolver um problema de Machine Learning é saber que tipo de dado o algoritmo irá trabalhar. Para objetivo deste trabalho será necessário possuir um grande número de fotografias faciais de pessoas de diferentes sexos e etnias para que o algoritmo seja capaz de classificá-las conforme sua identidade.

O treinamento será feito com uma base de dados que possui cinco etnias: Afro Americanos, asiáticos, hispânicos, caucasianos e mestiços; ainda foi feito um sexto tipo em que misturou-se todas essas etnias em uma outra base que será chamada multiracial. Na *Figura 1* temos alguns exemplos de como são a qualidade e a diferença entre os dados de cada uma das bases. Na *Tabela 1* temos a quantidade de imagens e pessoas de cada uma das etnias.

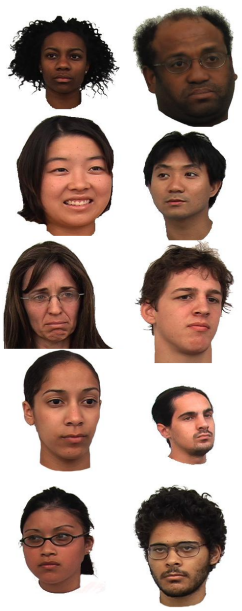

Figura 1 - Amostra de cada etnia. De cima para baixo temos: afro americanos, asiáticos, caucasianos, hispâncos, mestiços. Na coluna da esquerda temos uma mulher e na da direita um homem de cada etnia.

Etnia	**Número de pessoas**	**Imagens totais**
Afro americanos	34	273
Asiáticos	53	487
Caucasianos	105	1522
Hispânicos	19	163
Mestiços	18	180
Total:	227	2625

O tratamento

As imagens apresentadas não possuem grande padronização. Elas apresentam ângulos de fotografia que variam de 0 a 45 graus à esquerda ou à direita, expressões faciais diferentes, apetrechos na cabeça ou nos olhos e rotações em relação ao eixo horizontal.

Faz-se necessário, portanto, padronizar essas imagens conforme um critério comum, e o padrão escolhido foi o alinhamento dos olhos com o eixo horizontal e o recorte da região facial. Esses critérios visam eliminar características alheias ao indivíduo, como cor da roupa, corte de cabelo, etc. O objetivo é capturar a face em uma mesma direção conservando apenas os dados essenciais de cada pessoa para que diminua-se o erro.

Ambos os procedimentos de tratamento utilizam os olhos como referência para finalizar a operação. O primeiro procedimento é a rotação. Identificando a posição dos olhos é feita uma rotação na imagem até que os olhos fiquem alinhados com o eixo horizontal. O procedimento está ilustrado na *Figura 2*. É notável a perda de qualidade da imagem mas a qualidade ainda é suficiente para os objetivos deste trabalho.

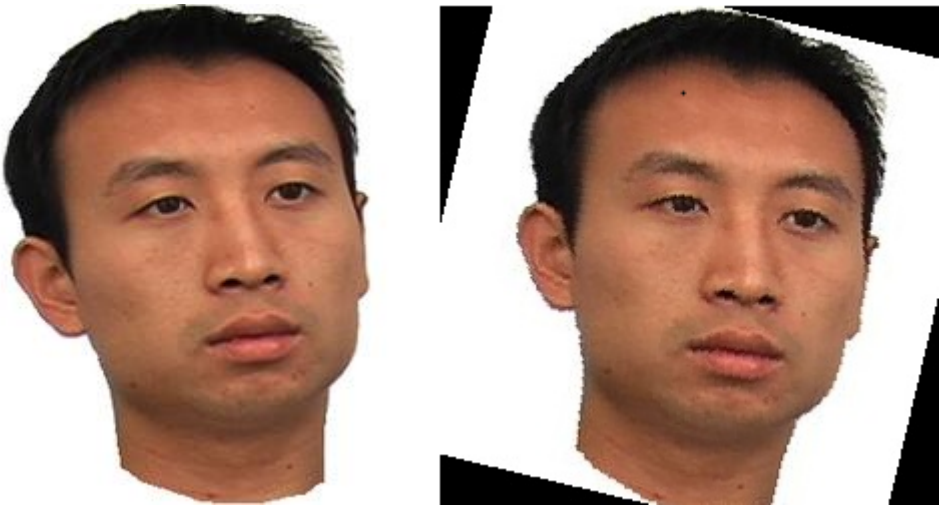

Figura 2 - Antes de depois de uma imagem rotacionada.

Após o alinhamento é identificada a região facial e é feito um recorte e, por fim, um redimensionamento da imagem para o valor de 150x150. O procedimento completo está ilustrado na *Figura 3*. É importante notar que a ordem dos procedimentos é essencial. Quando a função de recorte é aplicada às imagens não alinhadas, o recorte não é tão bem realizado como vê-se abaixo.

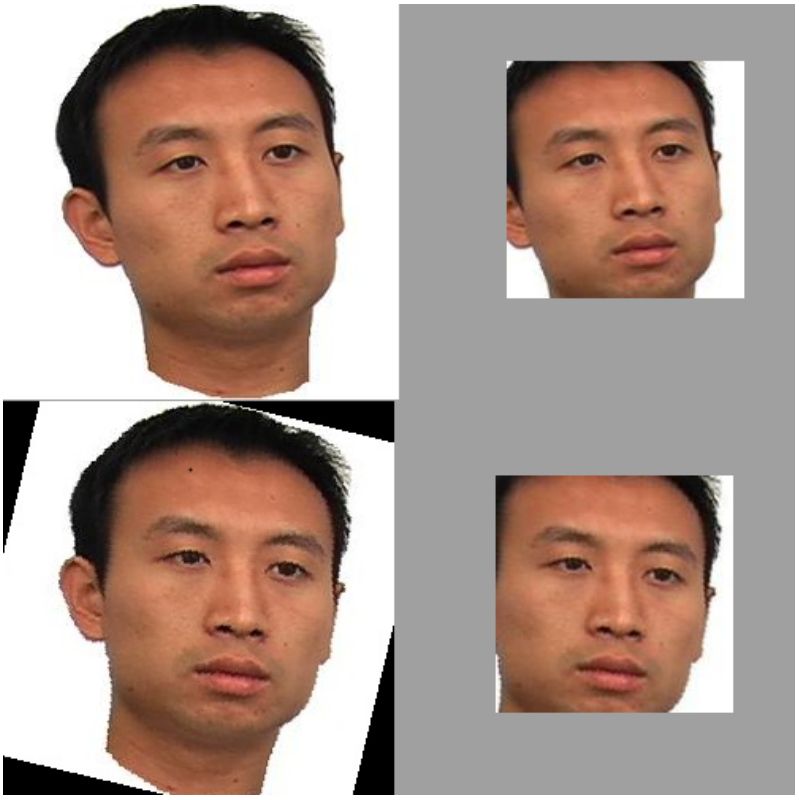

Figura 3 - Comparação dos recortes com rotação e sem rotação.

Todos esses procedimentos foram aplicados por meio de algoritmos escritos em Python e utilizou-se as bibliotecas NumPy, PIL, cv2 e autocrop. Esses procedimentos foram realizados de modo a gerar os arquivos intermediários que estão expostos neste relatório, entretanto isso foi feito para fins de recolhimento de dados e não fins práticos. Em um ambiente operacional é preferível que a imagem seja capturada completamente e processada para a forma matricial final e introduzido à rede neural diretamente, sem intermediários. Isso pode e deve ser feito.

O Pré-processamento

Uma imagem em computação é descrita no padrão RGB por uma matriz que possui as dimensões de sua quantidade de pixels horizontais, sua quantidade de pixels verticais e mais três dimensões de espessura em que cada uma delas corresponde a um valor de 0 a 255 correspondentes, respectivamente, ao grau de vermelho, verde e azul. Por exemplo, se temos uma imagem 127x255, então teremos uma matriz (127, 255, 3). Este conceito está ilustrado na *Figura 4*.

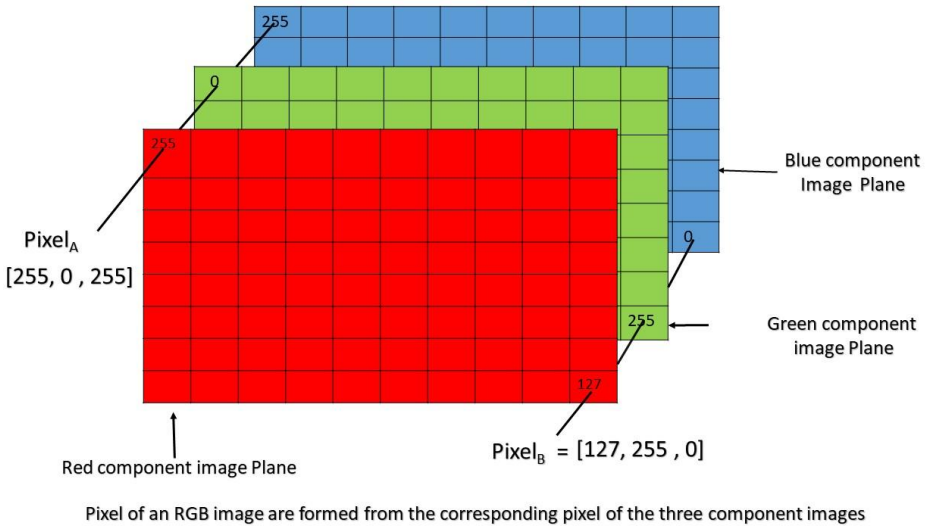

Figura 4 - Matriz RGB de uma imagem 127x255.

Para adaptar os dados para o algoritmo será feita uma normalização dos valores, dividindo-os todos por 255 (Valor máximo do padrão RGB), e, por fim, a matriz será verticalizada completamente. No exemplo teríamos, por fim, uma matriz (97155, 1), em que 97155 é o produto 127 vezes 255 vezes 3.

Eis a forma final dos dados que serão computadorizados pelos métodos de Machine Learning no procedimento de Back Propagation, o procedimento de tratamento. Por fim, os dados serão colocados em sua forma final utilizando a

biblioteca de pré-processamento ImageDataGenerator que faz parte do ferramental da biblioteca Keras, famosa biblioteca de Aprendizado Profundo em Python que funciona baseada na biblioteca TensorFlow. Utilizaremos 20% das imagens para validação dos resultados e 80% para o treinamento, dessa forma a quantidade de imagens para cada fim está explicitado na *Tabela 2*.

Etnia	**Número de imagens usadas no treinamento**	**Número de Imagens usadas na validação**
Afro americanos	234	39
Asiáticos	418	69
Caucasianos	1263	259
Hispânicos	141	22
Mestiços	158	22
Multiracial	2214	411

Tabela 2 - Relação de quantidade de imagens para treino e para validação.

A estrutura Deep Learning

Os dois modelos

A partir das bases construídas é necessário estruturar algoritmos de Machine Learning para o treinamento e reconhecimento dessas imagens para a obtenção de resultados operacionais.

Foram construídas duas estruturas para os treinamentos. A primeira, exposta na Figura 5, é uma rede profunda inicializada aleatoriamente e treinada desde o seu início com o banco de dados já apresentado. A biblioteca Keras do Python foi utilizada para sua construção, treinamento e validação.

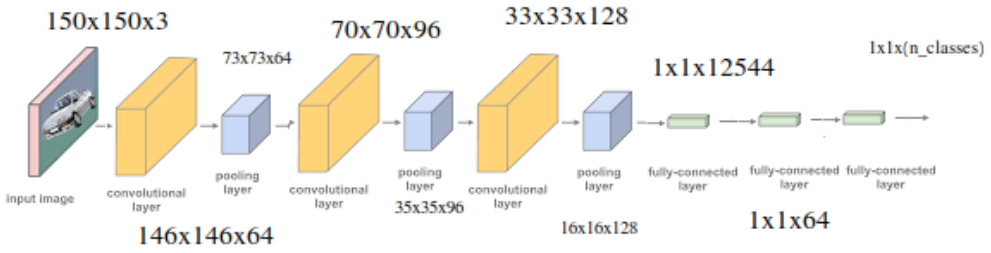

Figura 5 - Estrutura da rede inicializada aleatoriamente.

A segunda estrutura é mais robusta. Ela baseia-se no conceito de Transfer Learning, isto é, na utilização de uma base de dados pré-treinada para obtenção de uma eficiência maior.

Entre tantas alternativas na academia, foi escolhida a base VGG16 proposta no artigo *Simonyan & Zisserman, N. Very Deep Convolutional Networks for scale image recognition. Apresentado no ICLR 2015, 10 de Abril, 2015, 14. Disponível em https://arxiv.org/abs/1409.1556.* Trata-se de uma rede profundíssima como mostra a arquitetura da Figura 6, treinada com uma base de dados de mais de 1.2 milhões de imagens e 1000 classes diferentes, obtendo uma taxa de acerto de 92,7% na ImageNet, uma base de dados famosa na academia para disputar resultados.

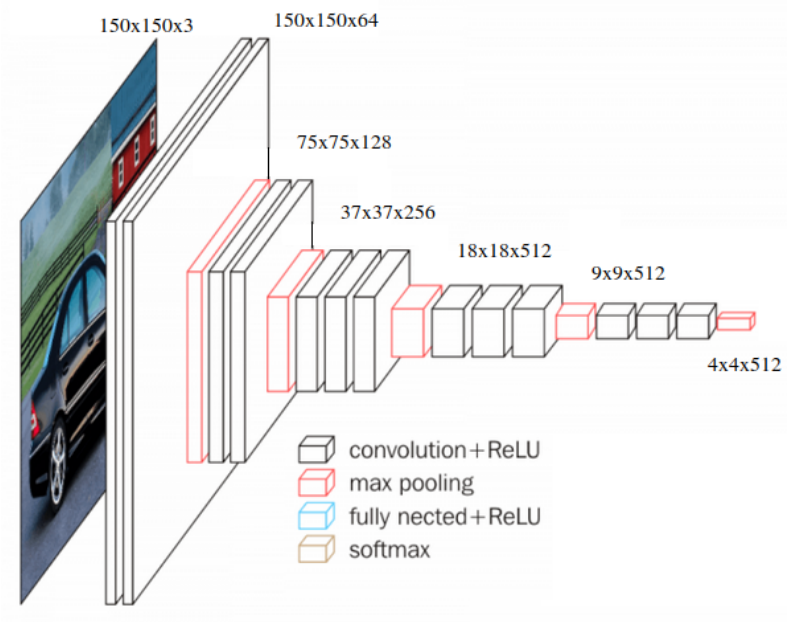

Figura 6 - Estrutura VGG16.

A preparação para sua utilização foi adaptar a entrada da rede para dados no formato que a base aqui apresentada necessita (150, 150, 3), congelar todas as camadas pré-treinadas para que os parâmetros obtidos pelos autores não sejam alterados e, por fim, acoplar 3 camadas finais treináveis para que a base seja aplicável ao problema aqui proposto. O resultado final está exposto graficamente na Figura 7.

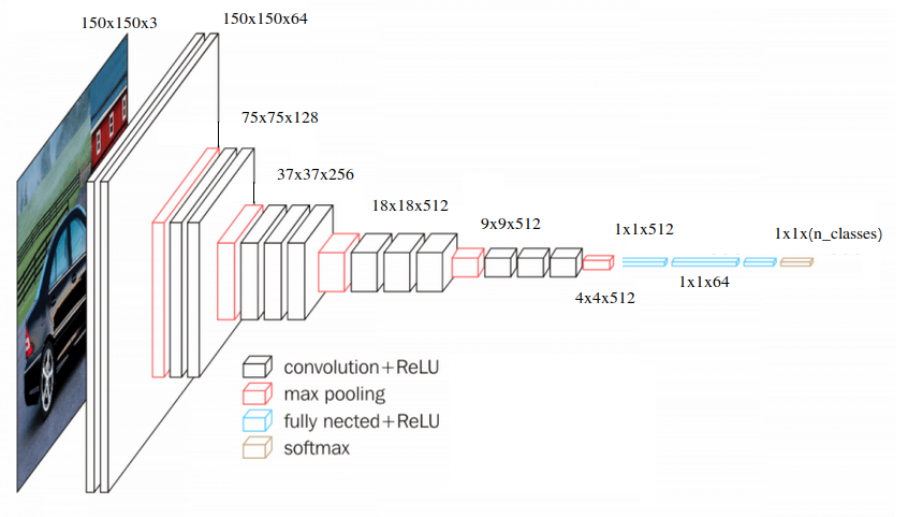

Figura 7 - Modelo com Transfer Learning.

A quantidade de parâmetros treináveis e não treináveis de cada um dos modelos encontra-se na *Tabela 3*.

Modelo	**Parâmetros treináveis**	**Parâmetros não-treináveis**	**Total de Parâmetros**
Sem Transfer Learning	1,328,395	1,088	1,329,483
Com VGG16	310,763	14,714,688	15,025,451

Tabela 3 - Quantidade de parâmetros de cada modelo.

O treinamento

O treinamento consistiu em treinar cada uma das 6 bases de dados em cada uma das 2 estruturas e os dados recolhidos de cada um dos treinamentos foram: a perda de treinamento, taxa de acerto de treinamento, perda de validação e taxa de acerto de validação.

O ambiente utilizado para treinamento foi o Google Colab. Os motivos que levaram a essa escolha foram a capacidade de integração com o Google Drive, facilitando a versatilidade e praticidade da pesquisa; e o hardware fornecido, uma GPU NVIDIA P100s, K80s ou P4s sendo impossível saber com qual delas foi feito cada treinamento. Mas todas elas são suficientes para o tipo de processamento exigido neste trabalho.

Tratamento dos dados

Para a visualização dos dados foi utilizada a biblioteca MatPlotLib e por meio dela serão plotados os dados de acordo com as epochs. As epochs são a repetição do processo iterativo de aprendizado.

Foi utilizado a média móvel para análise dos dados. A escolha desse filtro foi feita por sua capacidade de suavizar curvas e analisar tendências, isso condiz com o objetivo deste trabalho.

Devido aos resultados obtidos foi escolhido ignorar os dados de perdas por não serem capazes de transmitir alguma informação conclusiva. Pelo contrário, os gráficos de accuracy(Taxa de acerto, em inglês) conseguem mostrar o desempenho dos algoritmos e dos bancos de dados.

Resultados

Todos os dados foram recolhidos e encontram-se na Figura 8, onde estão listados na ordem já enunciada nas tabelas acima.

É visível o sucesso dos modelos. Todas as parcelas de treinamento, à exceção da multiracial que será discutida a seguir, foram levadas à exaustão ao ponto de possuírem não só um alto desempenho mas uma baixa variação, isto é uma alta estabilidade.

Entretanto, a grande diferença entre os algoritmos se dá na validação. A rede profunda inicializada aleatoriamente obteve grandes dificuldades na estabilidade de seus resultados. As variações chegam a 15% de acerto. Já a rede que utilizou o transfer learning não mostrou esse tipo de problema, sua validação varia tão pouco quanto no treinamento.

As bases que obtiveram maior estabilidade foram a de asiáticos e a de caucasianos treinados com o algoritmo VGG16. A base que apresentou maior dificuldades para treinamento foi a base de mestiços, isso ocorreu com os dois

algoritmos e por incrível que pareça a base inicializada aleatoriamente saiu-se melhor neste caso embora possua grande variação na validação. Entretanto nenhum dos dois casos é operacionável.

A base com todos os dados apresentou a maior diferença de resultados entre os modelos. É impensável operar essa banco sem a base pré-treinada VGG16. A eficiência com a inicialização aleatória não alcançou 15% no treinamento tampouco na validação.

Considerando, por fim, as dificuldades dos bancos de dados, o tamanho das bases de dados e a diferença dos tipos de face a serem reconhecidas os resultados foram satisfatórios para indicar direções por meio da qual o problema de reconhecimento multi-étnico possa vir a ser resolvida com alto desempenho em todos os casos.

Conclusão

Fica claro com os resultados obtidos a necessidade de uma rede neural profundíssima para lidar com o problema. A rede inicializada aleatoriamente não alcançou resultados satisfatórios na validação em nenhum dos casos, seja pelo valor final de acerto ou pela instabilidade desse mesmo valor.

Por outro lado, a utilização da base pré-treinada trouxe avanços significativos ao nosso problema. A rede neural constituída dessa ferramenta conseguiu resultados bons mas que não condizem com o estado de arte do reconhecimento facial atual, que está acima de 90%.

O trabalho mostrou que a utilização de uma mesma estrutura em bases diferentes adapta-se bem a qualquer rede obtendo aproximadamente 67% de acerto final de validação em todas elas. Mas ainda é necessário elevar todos esses resultados.

Um modelo ainda mais profundo pode ser um caminho interessante para resolver isso. No modelo VGG16 acoplamos apenas mais 3 camadas não-lineares, talvez acoplar algumas de convolução antes dessas pode ser uma boa alternativa.

O problema étnico que o trabalho se propõe a resolver não causou grandes impactos exceto na étnica de mestiços. Não pode-se concluir ao certo o motivo dessa etnia mostrar-se tão alheia aos resultados das outras. Outras bases de dados desse tipo devem ser utilizadas para entender se o problema é local (refere-se ao banco de dados) ou global (refere-se ao algoritmo).

Por fim, crê-se que o trabalho contribui para o entender o comportamento de Redes Neurais Profundas para identificação facial em diversos cenários diferentes. Foi mostrado as vantagens de utilizar bases pré-treinadas para acelerar o treinamento e obter melhores resultados e também foram indicados problemas que devem ser superados.

Referências

Valiente, Rodolfo; Gutiérrez, José C; Sadaike, Marcelo T; Bressan, Graça; "Automatic Text Recognition in Web Images", Proceedings of the 23rd Brazilian Symposium on Multimedia and the Web, 241-244,2017.

Gutiérrez, José C; Valiente, Rodolfo; Sadaike, Marcelo T; Soriano, Daniel F; Bressan, Graça; Ruggiero, Wilson V; "Mechanism for Structuring the Data from a Generic Identity Document Image using Semantic Analysis",Proceedings of the 23rd Brazilian Symposium on Multimedia and the Web,,,213-216,2017.

Gutiérrez, Armando M; Pacheco, Patricia A; Gutiérrez, José C; Bressan, Graça; "Development of a naive bayes classifier for image quality assessment in biometric face images", Proceedings of the 25th Brazilian Symposium on Multimedia and the Web,,,177-180,2019.

R. Szeliski, Computer Vision: algorithms and applications. London; New York: Springer, 2011

O. Kahm and N. Damer, "2D face liveness detection: An overview," BIOSIG-Proceedings IEEE Int. Conf. the. Biometrics Spec. Interes. Gr. (BIOSIG)., pp. 171–182, 2012.

M. Saini and C. Kant, "Liveness Detection for Face Recognition in Biometrics : A Review," pp. 31–36.

M. Singh and A. S. Arora, "A Novel Face Liveness Detection Algorithm with Multiple Liveness Indicators," Wirel. Pers. Commun., vol. 100, no. 4, pp. 1677–1687, 2018.

Simonyan & Zisserman, N. Very Deep Convolutional Networks for scale image recognition. Apresentado no ICLR 2015, 10 de Abril, 2015, 14. Disponível em https://arxiv.org/abs/1409.1556.