

Desenvolvimento Preliminar de um Algoritmo de Reconhecimento Facial em Tempo Real através de Redes Neurais Convolucionais

Diana Rocha, João Pereira, Maria Melo, Bruno C. Bispo e Pedro M. Rodrigues

Resumo— Este trabalho apresenta os resultados preliminares do desenvolvimento de um algoritmo de reconhecimento facial em tempo real através de redes neurais convolucionais (CNNs). Elementos faciais, nomeadamente, rosto, boca, nariz e olhos, são detectados pelo algoritmo de Viola-Jones. Cada elemento facial é utilizado para treinar uma CNN. Os resultados de treinamento mostram uma acurácia de identificação de 100%. Testes em tempo real demonstram necessidade de aprimoramento. A base de imagens será futuramente ampliada para realização de um rigoroso procedimento de treinamento e teste do algoritmo.

Palavras-Chave— Reconhecimento facial, tempo real, rede neural convolucional, algoritmo de Viola-Jones.

Abstract— This work presents the preliminary results of the development of a real-time facial recognition algorithm through convolutional neural networks (CNNs). Facial elements, namely, face, mouth, nose and eyes, are detected by the Viola-Jones algorithm. Each facial element is used to train a CNN. The training results show 100% identification accuracy. Real-time tests demonstrate a need for improvement. The image base will be expanded in the future to carry out a rigorous procedure for training and testing the algorithm.

Keywords— Face recognition, real time, convolutional neural network, Viola-Jones algorithm.

I. INTRODUÇÃO

A face é o principal elemento para a identificação de uma pessoa. O reconhecimento facial é crucial em muitas aplicações como, por exemplo, ciber segurança e monitorização de pessoas em tempo real [1]. Este trabalho apresenta os resultados preliminares do desenvolvimento de um algoritmo para identificar, em tempo real, indivíduos através de uma rede neural convolucional (CNN) e elementos faciais, nomeadamente, rosto, boca, nariz e olhos.

II. METODOLOGIA

O método aplicado pode ser dividido em 3 etapas: aquisição digital de imagem e pré-processamento; segmentação e extração de elementos faciais; reconhecimento em tempo real. A Figura 1 resume a metodologia utilizada.

A. Aquisição e pré-processamento de imagem

A primeira etapa constituiu na aquisição de 300 imagens do rosto de cada um dos 3 indivíduos, que participaram deste estudo inicial, através de uma *webcam*. As imagens na escala RGB foram convertidas para a escala de cinza. Imagens do nariz, boca e olhos foram extraídas de cada imagem do rosto.

Diana Rocha, João Pereira, Maria Melo, e Pedro M. Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal, e-mail: prodrigues@porto.ucp.pt; Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, e-mail: bruno.bispo@ufsc.br.

B. Segmentação e extração de elementos faciais

Para detectar os elementos faciais em tempo real, nomeadamente, rosto, boca, nariz e olhos, utilizou-se o algoritmo Viola-Jones. Este algoritmo tem a capacidade de processar imagens a uma alta velocidade e de apresentar altas taxas de detecção [2], [3]. Apesar de amplamente utilizado para detectar rostos, ele pode ser utilizado para detectar qualquer tipo de objeto. O algoritmo é composto por 4 etapas [2], [3]: características Haar; imagens integrais; treinamento *adaboost*; e classificador em cascata. Estas etapas são brevemente descritas abaixo. Maiores detalhes podem ser encontrados em [2], [3].

1) *Características Haar*: Características, ou *kernels*, constituídas de regiões retangulares integralmente formadas por pixels brancos ou pretos, onde as regiões possuem o mesmo tamanho e são adjacentes horizontal ou verticalmente. Possuem esse nome por lembrarem as funções base da Haar. Características com até 4 regiões são usadas. Aplicadas a uma sub-janela da imagem, o valor resultante é a diferença entre a soma dos pixels dentro das regiões claras e a soma dos pixels dentro das regiões escuras. Valores elevados indicam semelhança entre a sub-janela e a característica Haar.

2) *Imagens integrais*: O valor das características Haar pode ser calculado rapidamente utilizando uma representação da imagem denominada imagem integral. O valor de um pixel na imagem integral é definido como a soma dos valores dos pixels à esquerda e acima, inclusive, na imagem original.

3) *Treinamento AdaBoost*: O processo de treinamento utiliza AdaBoost para selecionar um subconjunto significativo de características. Classificadores fracos são treinados, através da otimização de um limiar de classificação, utilizando apenas uma característica de forma a minimizar falsos negativos. Ou seja, um classificador por característica. Classificadores fortes são então construídos ao combinar classificadores fracos.

4) *Classificadores em cascata*: A detecção é realizada através de classificadores em cascata. No primeiro estágio, todas as sub-janelas da imagem são aplicadas ao melhor classificador forte construído com um apenas um classificador fraco (ou característica). As sub-janelas não detectadas corretamente são definitivamente descartadas. As outras sub-janelas são aplicadas ao classificador do segundo estágio, o qual foi construído combinando alguns classificadores fracos. O número de classificadores fracos combinados para construir o classificador forte aumenta conforme o avanço nos estágios. O processo se repete até que a meta geral de falsos positivos e taxa de detecção é atendida. O conjunto de sub-janelas detectadas no último estágio forma a imagem do objeto.

O algoritmo de Viola-Jones foi aplicado às 300 imagens de cada indivíduo. Isso resultou em 300 imagens de cada um dos elementos faciais (rosto, boca, nariz e olhos) por indivíduo.

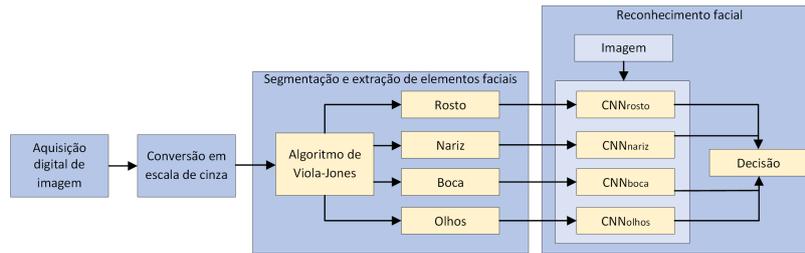


Fig. 1. Diagrama de fluxo da metodologia do algoritmo desenvolvido.

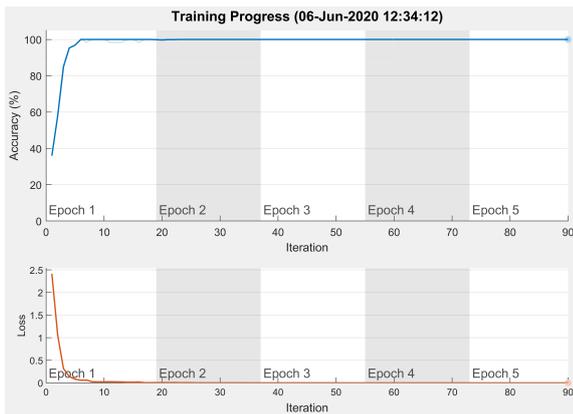


Fig. 2. Treinamento da CNN relativa ao rosto.

C. Reconhecimento em tempo real

A identificação dos indivíduos é realizada através de CNNs. As CNNs são utilizadas nos principais métodos do estado-da-arte de reconhecimento facial [4]. Neste trabalho, optou-se por uma CNN para cada elemento facial, totalizando 4 CNNs, com o intuito de tornar o reconhecimento mais robusto.

Para cada elemento facial, as 300 imagens de cada indivíduo foram utilizadas como pares entrada/saída para treinar a correspondente CNN com validação cruzada. Utilizou-se uma validação cruzada estratificada em 5 vezes, onde os dados de treino são divididos em 5 grupos. Por época, 20% dos dados de treino são alocados para validação cruzada de forma a validar o modelo. O algoritmo utilizado foi o SGDM (*stochastic gradient descent with momentum*), com taxa inicial de aprendizagem de 0,01 e funções de ativação ReLU (*rectified linear unit*) para todas as camadas, exceto na última onde foi utilizada a função de ativação softmax. O tamanho do *kernel* e o número de filtros são obtidos pela técnica de força bruta.

A decisão final do reconhecimento é baseada nas saídas das CNNs. Caso um indivíduo seja identificado por, pelo menos, 3 CNNs, ele é reconhecido pelo algoritmo. Caso contrário, assume-se que não houve reconhecimento do indivíduo.

III. RESULTADOS E DISCUSSÃO

Os resultados de treinamento da CNN relativa ao rosto são exibidos na Figura 2. Resultados muito similares foram obtidos para as outras CNNs. Observa-se que uma acurácia de 100% é atingida após as primeiras iterações, indicando que as CNNs se ajustam completa e rapidamente aos dados.

Testes do reconhecimento em tempo real foram realizados. A Figura 3 ilustra o reconhecimento correto de Diana e Maria, autoras do trabalho. Note que o nome dos indivíduos é exibido no retângulo amarelo acima do rosto.

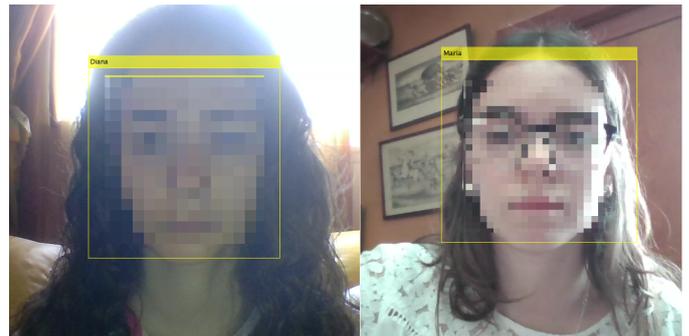


Fig. 3. Exemplos do algoritmo funcionando em tempo real.

Exemplos do algoritmo funcionando em tempo real estão disponíveis em <https://drive.google.com/file/d/1Hp6m0PcuMKc6PQmWJyVOxLtTqycvn2xN>.

Algumas limitações do algoritmo desenvolvido foram notadas. O algoritmo de Viola-Jones necessita de otimização para diminuir os falsos positivos porque, às vezes, detecta mais elementos faciais que os corretos. Há necessidade de proximidade entre a câmera e o indivíduo para identificação. Movimentos bruscos dos indivíduos provocam erros de identificação.

IV. CONCLUSÃO

Este trabalho apresentou os resultados preliminares do desenvolvimento de um algoritmo de reconhecimento facial em tempo real. Elementos faciais, nomeadamente, rosto, boca, nariz e olhos, são obtidos pelo algoritmo de Viola-Jones. Cada elemento facial é utilizado para treinar uma CNN. Apesar de resultados preliminares promissores, onde a prova de conceito foi conseguida, é necessário aumentar a base de imagens para realizar um rigoroso procedimento de treinamento e teste do algoritmo. Além disso, algumas melhorias no algoritmo são necessárias para efetiva aplicação em tempo real.

AGRADECIMENTOS

Esse trabalho foi apoiado pelos Fundos Nacionais da Fundação para a Ciência e a Tecnologia (FCT), Portugal, através do projeto UIDB/50016/2020.

REFERÊNCIAS

- [1] J. B. Alam, M. M. Islam, T. Jabid, and S. Ahmed, "System development using face recognition," in *Proceedings of the International Conference on Automation, Computational and Technology Management*, April 2019.
- [2] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.
- [3] —, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December 2001, pp. 511–518.
- [4] S. Balaban, "Deep learning and face recognition: the state of the art," in *Proceeding of XII Conference on Biometric and Surveillance Technology for Human and Activity Identification*, 2015.