



Aprendizado de máquina para predição de idade e gênero em tempo real

Machine learning for real-time age and gender prediction

João Vitor Pizzoni Paulino¹, Gustavo Henrique Paetzold²

RESUMO

A estimativa de idade de uma pessoa pode ser definida de forma automática por outras pessoas, porém com o avanço da tecnologia surgiu a ideia de uma máquina predeterminar a idade de uma pessoa. Contudo, devido a particularidade e complexidade de cada rosto torna-se um problema para a computação abstrair cada característica. O problema de predeterminar as idades tornou-se importante devido a sua aplicação, como de exemplo na arte forense, Gestão Eletrônica de Relacionamento com o cliente e controle de segurança. Neste trabalho foram utilizados modelos de máquina do tipo CNN's (Rede Neural Convolucional) para identificação de idade e de gênero em tempo real. Os resultados mostram que o sistema é capaz de ter uma boa precisão. Contudo, foi realizado um teste com três pessoas, dois são homens de 20 anos e uma mulher de 22 anos, o teste foi individualmente por pessoa e em seguida com os três simultaneamente.

PALAVRAS-CHAVE: Aprendizado de Máquina; Predição de Idade e Gênero; Visão Computacional.

ABSTRACT

A person's age estimate can be defined automatically by other people, but with the advancement of technology, the idea of a machine predetermining a person's age has emerged. However, due to the particularity and complexity of each face, it becomes a problem for computing to abstract each characteristic. The problem of predetermining ages has become important due to its application, for example in forensic art, Electronic Customer Relationship Management and security control. In this work, CNN machine models (Convolutional Neural Network) were used to identify age and gender in real time. The results show that the system is capable of good accuracy. However, a test was carried out with three people, two were 20-year-old men and one was a 22-year-old woman, the test was carried out individually by person and then with all three simultaneously.

KEYWORDS: Machine Learning; Age and Gender Prediction; Computer Vision.

INTRODUÇÃO

O avanço da tecnologia proporcionou diversas ferramentas, desde o celular que pode ser usado tanto para o lazer quanto para agilizar a vida cotidiana. A partir disso, um problema proposto pela visão computacional seria uma máquina determinar a idade e o gênero de uma pessoa apenas o seu rosto. O desafio apresenta pontos importantes, visto que até para o próprio ser humano dificilmente acerta a idade de outro. A aplicação surge em várias do conhecimento, tornando – se importante o desenvolvimento de tal sistema.

A construção do código para a detecção em tempo real de idade e gênero, utiliza de modelos de aprendizagem de máquina. O sistema construído foi a partir de modelos

¹ Voluntário. Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil. E-mail: jpaulino@alunos.utfpr.edu.br. ID Lattes: 2201841350680822.

² Docente do curso Engenharia da Computação. Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil. E-mail: ghpaetzold@utfpr.edu.br. ID Lattes: 3576463426605379.

como Rede Neural Convolucional (CNN), também uma complexa rede de imagens classificadas, além de um reconhecimento de facial.

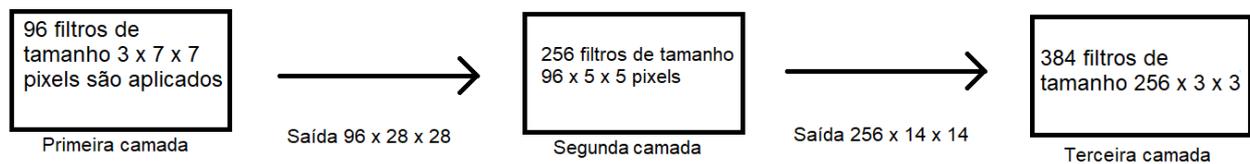
MATERIAIS E MÉTODOS

Para realizar o projeto foram utilizados modelos de redes neurais convolucionais a um sistema de visão que opera em tempo real com a ajuda de uma webcam. A rede LeNet-5 foi uma das primeiras para o reconhecimento óptico de caracteres, mesmo sendo uma rede menos profunda comparada as outras.

A rede neural baseia-se em uma espécie de rede neural artificial, tende-se com o intuito de diminuir o overfitting, nome que se dá quando o sistema não aprendeu o suficientemente as relações existentes para o teste [5], foi utilizado no código exemplo uma rede composta por três camadas convolucionais e duas camadas conectadas com um pequeno número de neurônio [3].

As imagens para serem utilizadas pelo código passam por um processamento onde primeiramente são redimensionadas para 256 x 256 de tamanho e um corte de 227 x 227 para alimentar a rede. Logo, as camadas convolucionais são definidas como:

Figura 1 – Camadas Convolucionais



Fonte: Demonstração de como as camadas estão relacionadas e os filtros aplicados. Autoria própria.

Todas as camadas empregam o Rectified Linear Units (ReLU), uma função que desempenha o papel de aperfeiçoar os resultados em cada camada do código por meio de atribuição de pesos aos parâmetros, sendo demonstrada pela equação (1) [1].

$$o = \sum_i^{n-1} \theta x_i \quad (1)$$

As camadas estão conectadas entre si, logo a primeira camada recebe a saída da terceira e contém 512 neurônios seguido por um ReLU e uma camada dropout, a segunda recebe da primeira e realiza o mesmo processo, já a terceira mapeia todas as classes de idade e de gênero. Depois de todo esse processo é aplicada a uma camada softmax que calcula a probabilidade de acerto na imagem teste [3].

O treinamento do código foi realizado com um banco de dados disponibilizado na internet. Um conjunto de pessoas organizaram várias fotos, onde todas estão classificadas com idades e gêneros. O banco de dados utilizado é o “Age/Gender benchmark of unfiltered face”, fotos que apresentam imagens como essa como exemplo.



Figura 2 - Age and Gender Estimation of Unfiltered Faces



Fonte: Eran Eidinger, Roeen Enbar, Tal Hassner. Age and Gender Estimation of Unfiltered Faces. Transactions on Information Forensics and Security (IEEE-TIFS), special issue on Facial Biometrics in the Wild, Volume 9, Issue 12, pages 2170 - 2179, 2014.

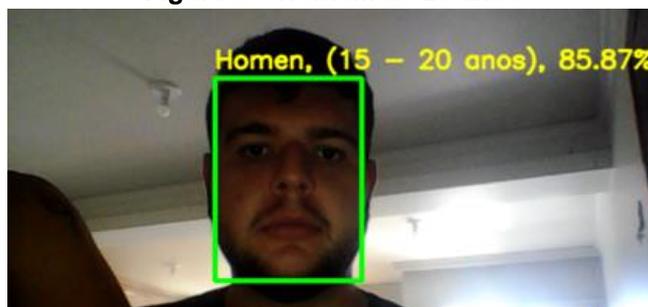
Além de usar uma rede convolucional mais enxuta para diminuir overfitting, foi utilizado mais dois métodos, o primeiro é aplicação de aprendizado dropout que é uma marca oculta nas camadas com aleatoriedade de 0,5, uma forma muito eficiente de realizar a média do modelo com redes neurais [2]. O segundo método utilizado foi arranjar e rearranjar as os pixels nas medidas de 227 x 227 das fotos de 256 x 256, aleatoriamente e passando pelas camadas para frente e para trás [3].

O código exemplo utilizado tem duas repartições principais, a idade e o gênero, cada uma tem suas próprias características. A idade ela é dividida em oito faixa etárias, sendo [0-2], [4 - 6], [8 - 13], [15 - 20], [25 - 32], [38 - 43], [60 - +]. O gênero é dividido em dois, em apenas homem e mulher.

RESULTADOS

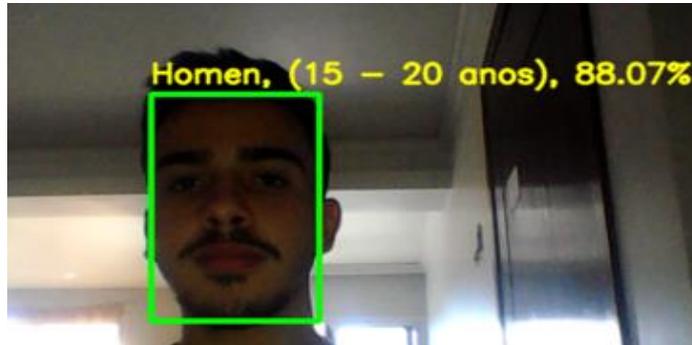
Para realizar um teste foi colocado em prática com três pessoas, dois são homens de 20 anos e uma mulher de 22 anos.

Figura 3 – Homem de 20 anos



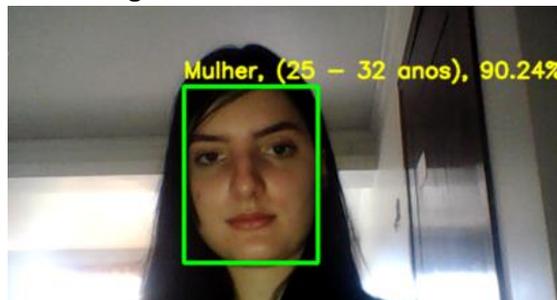
Fonte: Autoria própria.

Figura 4 – Homem de 20 anos



Fonte: Autoria própria.

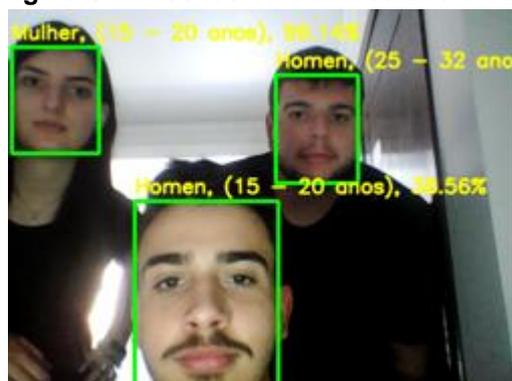
Figura 5 – Mulher de 22 anos



Fonte: Autoria própria.

Além da predição de idade e gênero, o código também retornar sua porcentagem de assertividade. Contudo, além dessas características, o sistema consegue captar além de um rosto na mesma imagem.

Figura 6 – Pessoas com idades diferentes



Fonte: Autoria própria.

CONCLUSÃO

O programa desenvolvido apresenta uma complexidade enxuta comparado a outros aprendizados de máquinas mais sofisticados, porém apresenta uma assertividade



relativamente alta. Devido a suas relações entre as camadas convolucionais, apresenta uma forma simples de aprender como funciona o treinamento, a parte de abstrair os rostos e prever as idades e gêneros.

Contudo, os resultados demonstraram uma boa porcentagem em situações controladas, com a iluminação artificial, pouco movimento e uma câmera com resolução média. Entretanto, apresentou dificuldades quando as pessoas se mexem e demonstrando faixas etárias semelhantes a que tinha apresentado antes.

Portanto, para um aperfeiçoamento da predição de idades e gêneros, há diversas formas, como a criação de mais camadas convolucionais, mais imagens para a realização de treinamento e melhor enquadramento dos rostos. Trabalhos mais complexos utilizam uma gama de imagens testes e uma aprimoração na capacidade do aprendizado de máquina, a partir desse trabalho cria-se a oportunidade de várias formas de aprimoramento futuro.

Agradecimentos

Agradeço pelos recursos fornecidos pela UTFPR – Toledo para a realização deste trabalho.

Conflito de interesse

Não há conflitos de interesse.

REFERÊNCIAS

AGARAP, Abien Fred. Deep learning using rectified linear units (relu). **arXiv preprint arXiv:1803.08375**, 2018.

HINTON, Geoffrey E. et al. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.

LEVI, Gil; HASSNER, Tal. Age and gender classification using convolutional neural networks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition workshops**. 2015. p. 34-42.

RUSSAKOVSKY, Olga et al. Imagenet large scale visual recognition challenge. **International journal of computer vision**, v. 115, p. 211-252, 2015.

YING, Xue. An overview of overfitting and its solutions. In: **Journal of physics: Conference series**. IOP Publishing, 2019. p. 022022.