

## << Energia Visual Emocional (E.V.E. - Versão 2.0) - Plataforma de Acessibilidade Comunicativa que Utiliza Reconhecimento Facial e Mensagens em Áudio Para Inclusão de Pessoas com Deficiência Visual ou Cegueira>>

Nathalia Assunção das Chagas, Thiago Pinto Martins  
Robson Ferreira Lopes, Gema Galgani Bezerra  
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo  
Câmpus Guarulhos

### Resumo

Na contemporaneidade, a comunicação cinésica é considerada um aspecto facilitador nas relações interpessoais, isso porque transmite informações que muitas vezes passam despercebidas no discurso verbal. Desse modo, pessoas com deficiência visual ou cegueira acabam perdendo uma parte das informações transmitidas no decorrer de uma conversa, o que pode resultar em má compreensão durante as interações sociais e na reclusão desses indivíduos. Em virtude disso, a plataforma Energia Visual Emocional (E.V.E. versão 2.0) tem como objetivo fomentar a inclusão social das pessoas com deficiência visual a partir de um protótipo de software capaz de reconhecer as faces de pessoas presentes em uma imagem e de traduzir em mensagens de áudio sua identificação e traços visuais que indiquem suas emoções. Utilizando o método de engenharia a partir da prototipação, a plataforma utiliza as técnicas de Aprendizagem Profunda (*Deep Learning*) em conjunto com a biblioteca *Face Recognition* para realizar o reconhecimento facial. Os próximos passos consistem em aprimorar a eficiência do reconhecimento de emoções, além de aprofundar os estudos para adaptar o projeto para o dispositivo móvel. Os processos de desenvolvimento, bem como os fatores levados em consideração na etapa inicial do projeto são detalhados do decorrer dessa pesquisa.

**Palavras-chave:** Inclusão. Deficiência visual ou cegueira. Inteligência artificial. Reconhecimento de faces.

## 1. Introdução

Na contemporaneidade, a sociedade torna-se cada vez mais orientada visualmente. Segundo Viana (2014) as expressões faciais são um dos meios mais importantes nas relações interpessoais, isso porque funcionam como uma confirmação de ideias, reforçando o discurso, além de serem a forma mais comum de expressar as emoções dos indivíduos. Ainda segundo Viana (2014), que analisa o estudo do antropólogo Ray Birdwhistell (1971)<sup>1</sup>, a comunicação cinésica, como gestos e expressões faciais, transmitem entre 65% e 70% das informações durante uma interação social, mostrando que apenas cerca de 30% das informações são transmitidas pela comunicação verbal. Desse modo, pessoas com deficiência visual ou cegueira (pessoas doravante mencionadas neste trabalho como deficientes visuais ou cegas) acabam perdendo uma parte das informações transmitidas durante uma conversa, o que pode resultar em má compreensão durante as interações sociais.

De acordo com a psicóloga Maria C.C. Barczinski (2001), que comenta o estudo do psiquiatra Luis Cholden (1984), a falta de visão se torna um grande gatilho para a perda da identidade, falta de vontade de viver e depressão, reações semelhantes às do luto. Em vista disso, contornar esse obstáculo na comunicação é um grande passo para reduzir as lacunas de uma vida sem visão, incluindo os deficientes visuais na sociedade e diminuindo os casos de isolamento e depressão.

Ademais, o projeto procura responder com essa pesquisa: e se for possível uma pessoa cega identificar outras pessoas em uma imagem e saber quem está nela? E se for possível construir uma ferramenta ou dispositivo capaz de reconhecer pessoas e suas respectivas emoções em uma imagem e identificá-las para a pessoa cega por meio de mensagens de áudio? Como garantir maior autonomia aos deficientes visuais para saberem quais pessoas estão em uma fotografia, por exemplo? Em virtude do que foi mencionado, o projeto tem como objetivo a construção de uma plataforma capaz de identificar as faces e emoções humanas, primeiramente em imagens, e, depois, em interações reais, utilizando a inteligência artificial, e descrevê-las por meio de mensagens de áudio, melhorando a qualidade de vida das pessoas com deficiência visual, facilitando seu convívio e interações sociais. Nesse trabalho serão descritos os processos de levantamento de necessidades e material teórico, bem como o processo de prototipação da plataforma que utiliza elementos do *Deep Learning* com a ciência social aplicada.

<sup>1</sup> BIRDWHISTELL, Ray L. *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia: **University of Pennsylvania Press**. 1971

## 2. Materiais e Métodos

### 2.a Fundamentação teórica

Essa seção deste trabalho tem como objetivo apresentar os conceitos utilizados para a fomentação dessa pesquisa.

#### 2.a.a *Deep Learning*

Segundo Goodfellow, Bengio e Courville (2016), *Deep Learning* ou aprendizagem profunda, tecnologia que se encontra como uma subcategoria da inteligência artificial, é um ramo da ciência da computação que propõe modelos de dispositivos que simulem a capacidade de raciocínio humano em termos de hierarquia de conceitos, produzindo decisões e resultados confiáveis, com a menor intervenção humana possível.

Dessa tecnologia resulta uma série de camadas de processamento: a máquina realiza etapas para tentar descobrir determinada informação. Apresentando a imagem de um objeto para um computador – um *input* (dados fornecidos) –, ele tentaria entendê-lo dividindo a análise desse objeto em diversas camadas de processamento chamadas de *Hidden Layers*. Usando várias camadas de processamento de dados não lineares, obtém-se uma representação complexa e abstrata dos dados de forma hierárquica, sem a intervenção humana no processo.

Dados sensoriais (pixels de imagens, por exemplo) são alimentados em uma primeira camada, sendo que a saída de cada camada se torna a entrada da camada seguinte. O aprendizado profundo é baseado no conceito de redes neurais artificiais, ou sistemas computacionais que imitam a maneira como o cérebro humano funciona.

#### 2.a.b *Python e as bibliotecas*

Segundo a [python.org](http://python.org), Python é uma linguagem fácil de aprender e poderosa, pois possui estruturas de dados de alto nível, o que a torna mais eficiente com uma abordagem simples e efetiva de programação orientada a objetos.

Neste projeto os módulos usados mais importantes são: A *Face Recognition*, a biblioteca Dlib, Pyttx4, um sintetizador de voz compacto de software de código aberto para inglês e outros idiomas, e a API Keras em conjunto com o dataset FER2013 (Facial Expression Recognition 2013 Dataset) para identificar as emoções.

A Face Recognition foi criada por Adam Geitgey, uma API (*Application Programming Interface*) que possibilita o reconhecimento facial, baseada no estado da

arte da biblioteca Dlib. É desenvolvida na linguagem Python e disponibilizada gratuitamente no GitHub do próprio autor. ([https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition)).

O Dlib é uma biblioteca gratuita, desenvolvida em C ++, projetada para ser utilizada como uma coleção de componentes de framework que podem ser utilizados em qualquer tipo de aplicação, seja comercial ou acadêmica. No presente, ele fornece módulos que suportam álgebra linear, processamento de imagens, otimização e outras funções, e todos fornecem uma documentação rica e de simples acesso.

O Keras é um framework, biblioteca Python voltada a aprendizado de máquina profundo. Esta biblioteca implementa funcionalidades de frameworks de deep learning, já consolidadas, entre elas o TensorFlow.

Para desenvolvimento do trabalho foi utilizado a base de dados Facial Expression Recognition 2013 (FER-2013). Esta base de dados foi desenvolvida por Pierre Luc Carrier e Aaron Courville, e possui um total 35887 imagens faciais RGB de diferentes expressões, com dimensões de 48x48 pixels (GOODFELLOW et al., 2013). Os dados possuem as seis expressões faciais de Ekman, adicionadas à expressão neutra. As amostras desta base de dados estão distribuídas em: 7215 imagens de alegria; 436 de desgosto; 4097 de medo; 4965 de neutra; 3995 de raiva; 3171 de surpresa; e 4830 de tristeza.

### **2.a.c Rede Neural**

De acordo com Vargas, Carvalho e Vasconcelos (2016), uma CNN - *Convolutional Neural Network*, cuja tradução literal é “Rede Neural Convolutacional”, é um algoritmo do *Deep Learning* inspirado no “processo biológico de processamentos de dados visuais”.

Enquanto a abordagem tradicional de computação utiliza uma série de blocos lógicos para executar uma determinada tarefa, Redes Neurais utilizam redes compostas por nós (os quais atuam como neurônios) e arestas (que por sua vez atuam como sinapses) para processar dados. As entradas então percorrem o sistema e uma série de saídas é gerada. Essas saídas são então comparadas com dados conhecidos.

Segundo Goodfellow, Bengio e Courville (2016) a CNN, na sua execução, utiliza uma construção matemática que é tipicamente composta por três tipos de camadas (ou blocos de construção): convolução, agrupamento e camadas totalmente conectados. As duas primeiras, camadas de convolução e agrupamento, realizam extração de recursos,

enquanto a terceira, uma camada totalmente conectada, mapeia as características extraídas para a saída final, como classificação.

## 2.b Metodologia

Para programar o protótipo, foi necessária uma pesquisa que resultou na utilização de uma programação em python em conjunto com a biblioteca *Face Recognition* (escrita em linguagem python), que identifica características do rosto e é uma das ferramentas fundamentais usadas para o reconhecimento facial.

Para viabilizar o uso da biblioteca, foram instalados os pacotes python3-numpy==1.16.4, python3-scipy, python3-matplotlib ipython3, python3-pandas, python3-sympy, python3-nose, python3-keras, python3-opencv, python3-Tkinter, python3-pyinstaller, python3-pytxs3, python3-imutils em uma máquina com sistema operacional Windows e, depois, com o pip (ferramenta utilizada para baixar pacotes do repositório Python) foi feita a instalação da *face\_recognition*. Além da biblioteca, foram utilizados módulos como dlib, que é um moderno conjunto de algoritmos escritos em C++ para aprendizado de máquina, PIL.image e Pytx4 (sintetizador de voz compacto de software de código aberto para inglês e outros idiomas).

Vários testes da biblioteca foram realizados até chegar ao protótipo usado neste projeto. Na primeira versão foi desenvolvido o código *facesample01.py* e os resultados foram as coordenadas de cada rosto em qualquer foto indicada no programa. No segundo teste foi possível identificar as faces e fazer o software indicar a quantidade de pessoas na foto.

Na versão beta, em primeira instância, é importada a biblioteca *face\_recognition* e as bibliotecas adicionais. Em seguida, são carregadas as imagens pré-registradas no banco de dados, o que permite ao programa comparar as imagens recebidas com as que ele já conhece (Anexo 01).█

O programa utiliza uma rede neural para aprender sobre as imagens, identificando, respectivamente, as faces presentes na imagem e a matriz de dados geradas por ela. Para iniciar a execução do reconhecimento um arquivo é carregado para teste de reconhecimento (Anexo 02). Tendo a imagem carregada, a biblioteca PIL gera uma nova imagem de modo a não alterar a imagem original, sendo criado o objeto retângulo que será impresso no novo arquivo com as faces identificadas. Identificada a quantidade de faces existentes na imagem de teste, o programa compara essas faces com as imagens que o software conhece. As faces identificadas recebem um retângulo e,

abaixo, um label com o nome da pessoa. Caso o programa não reconheça a face, será registrada a palavra “Desconhecido” no label da face. No final, temos a principal função do sistema, que é a reprodução em áudio dos nomes das pessoas encontradas e a indicação das pessoas desconhecidas (Anexo 03)

Na versão gama, assim como na versão beta, o programa recebe uma imagem de entrada e compara com as imagens pré-registradas no banco de dados, criando uma caixa de texto ao redor da face com o nome do indivíduo identificado. Em seguida, é realizada uma extração dos pontos faciais das imagens de treino. Esses pontos são arquivados em um arquivo .npy e lançados na rede neural, que compara os descritores faciais da nova imagem de entrada com os descritores faciais já salvos no treinamento (Anexo 04).

Com a comparação finalizada e os pontos identificados, a imagem passa pela biblioteca Keras, que cria uma matriz com os descritores faciais, reduzindo a imagem em pixels de 48x48px, realizando uma comparação com os pontos salvos no dataset FER2013.

O programa então pega os descritores faciais da imagem de teste e compara com as imagens presentes no dataset, classificando com qual emoção a imagem mais se encaixa (Anexo 05).

Para validar o protótipo, o programa foi executado novamente com outro arquivo de teste (Anexo 06).

Com a utilização da biblioteca Tkinter montamos uma interface simples (anexo 07). O programa se inicia com um pequeno widget com uma breve caixa de texto e um botão. O botão recebe uma função e todo código implementado está dentro do escopo da função, ao clicar, o botão inicia uma caixa de busca de arquivos direcionada para img, essa imagem é armazenada em uma variável e esta mesma variável entra com entrada no nosso código de reconhecimento com a raiz da img selecionada

### **3. Resultados e Discussão**

Até o presente momento, o programa conseguiu identificar a quantidade de pessoas na foto e exibi-las separadamente, identificando cada uma delas por meio de áudio (Anexo 01). No entanto, o protótipo consegue identificar apenas uma emoção por vez caso haja mais de um indivíduo presente na imagem.

Para sua validação, foram realizados testes indiretos com um indivíduo surdo-cego. Porém, devido sua realização de maneira virtual, não foi possível uma comprovação científica de seus resultados.

O protótipo mostrou-se eficiente quanto às técnicas de programação e uso da biblioteca, mas cabe salientar que a sua validação ainda não foi realizada devido à pandemia de Covid-19. Nas fases seguintes, tem-se o objetivo de realizar novas simulações para que seja possível a identificação das emoções em todos os indivíduos presentes na imagem, além de aprofundar as pesquisas do reconhecimento de emoções a partir da utilização das ações fundamentais dos músculos individuais (*Action Units*), o que permitirá uma eficiência acima de 70% nos resultados do programa. Vale ressaltar que o projeto tem como prioridade testar o programa com pessoas deficientes visuais ou cegas, validando-o devidamente.

#### **4. Considerações Finais**

O objetivo desta pesquisa foi realizar um estudo que demonstre a importância da consolidação da autonomia comunicativa da pessoa com deficiência visual ou cegueira por meio da criação de novos meios de comunicação, o que pode aumentar a qualidade de vida desses indivíduos para uma melhor socialização no mundo contemporâneo. Dada a importância do assunto, o projeto E.V.E tem como principais objetivos o reconhecimento facial em imagens, tendo como seu diferencial a utilização da tecnologia *Deep Learning*, em conjunto com o reconhecimento facial, como uma ferramenta de inclusão. O desenvolvimento do presente estudo possibilitou o aprendizado de conceitos como *Deep Learning* e redes neurais, que são conhecimentos avançados na área de inteligência artificial, além de permitir um melhor conhecimento sobre programação utilizando a linguagem python.

A versão 2.0 do projeto é capaz de identificar a quantidade de pessoas na foto e exibi-las separadamente com apoio da biblioteca *face recognition*, identificando cada uma das faces presentes por meio da reprodução de áudio. Além de identificar as emoções do indivíduo presente na foto.

Para o futuro, a intenção é aprimorar o código do protótipo, melhorando a eficiência na identificação de emoção. Além disso, o projeto pretende, ainda, melhorar a

interface gráfica do projeto, tornando-a acessível, permitindo autonomia para a pessoa com deficiência visual ao utilizar a plataforma.

## 5. Referências

BARCZINSKI, Maria Cristina de Castro. Reações psicológicas à perda da visão. **Benjamin Constant**, n. 18, 2001. Disponível em: <<http://revista.ibc.gov.br/index.php/BC/article/view/585>> Acesso em: 13 abril. 2021.

DLIB C++ Library. **Documentation**. Disponível em: <http://dlib.net/>. Acesso em: 17 nov. 2020.

GEITGEY, Adam. **Face Recognition documentation**. 2017. Disponível em: <<https://face-recognition.readthedocs.io/en/latest/index.html>>. Acesso em: 17 nov. 2020.

GOODFELLOW, Ian *et al.* **Deep Learning**. Cambridge: MIT press, 2016. Disponível em: <[http://www.academia.edu/38223830/Adaptive\\_Computation\\_and\\_Machine\\_Learning\\_series\\_Deep\\_learning\\_The\\_MIT\\_Press\\_2016\\_pdf](http://www.academia.edu/38223830/Adaptive_Computation_and_Machine_Learning_series_Deep_learning_The_MIT_Press_2016_pdf)> Acesso em: 17 nov. 2020

PYTHON Software Foundation. **Documentação Python 3.8.2** Disponível em: <<https://docs.python.org/pt-br/3/index.html>>. Acesso em: 17 nov. 2020.

VARGAS, Ana Caroline Gomes; PAES, Aline; VASCONCELOS, Cristina Nader. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: **Proceedings of the XXIX Conference on Graphics, Patterns and Images**. 2016. Disponível em: <<http://gibis.unifesp.br/sibgrapi16/e proceedings/wuw/7.pdf>>. Acesso em: 19 nov. 2020

VIANA, Isabel. Comunicação não verbal e expressões faciais das emoções básicas. **Revista de Letras**, v. 13, p. 165-181, 2014. Disponível em: <[https://www.researchgate.net/profile/Bel-Viana/publication/330729356\\_Comunicacao\\_nao\\_verbal\\_e\\_expressoes\\_faciais\\_das\\_emocoes\\_basicas/links/5c518ffca6fdccd6b5d4112b/Comunicacao-nao-verbal-e-expressoes-faciais-das-emocoes-basicas.pdf](https://www.researchgate.net/profile/Bel-Viana/publication/330729356_Comunicacao_nao_verbal_e_expressoes_faciais_das_emocoes_basicas/links/5c518ffca6fdccd6b5d4112b/Comunicacao-nao-verbal-e-expressoes-faciais-das-emocoes-basicas.pdf)>. Acesso em: 29 abril. 2021.

## ANEXOS



Anexo 01 - Arquivos de imagem usados como base para comparação pelo programa do projeto E.V.E.




		
Chris Evans	Chris Hemsworth	Robert Downey Jr.

Foto: (Reprodução/Cinema com Rapadura;

Reprodução/AdoroCinema; Reprodução/Legião dos Heróis).

Anexo 02 - Imagem com três rostos dos atores usada como

teste no projeto



Foto: Perfil do ator Chris Hemsworth no Instagram - Disponível

em <<https://www.instagram.com/p/Bv7x8fFjnGA/>>

Anexo 03 - Imagem com os três rostos identificados pelo programa EVE versão beta.



Foto: Perfil do ator Chris Hemsworth no Instagram - Disponível em <<https://www.instagram.com/p/Bv7x8FjnGA/>>

Anexo 04 - Imagem utilizada para validação da do programa E.V.E. versão gama.

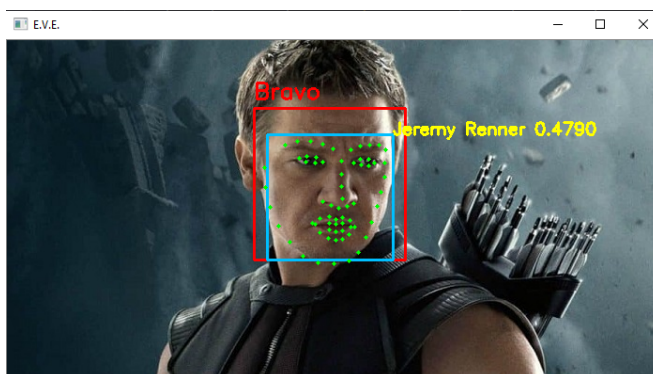


Foto: Reprodução/Geek Blog.

Anexo 05 - Imagem com o rosto e emoção identificados pelo programa EVE versão gama.

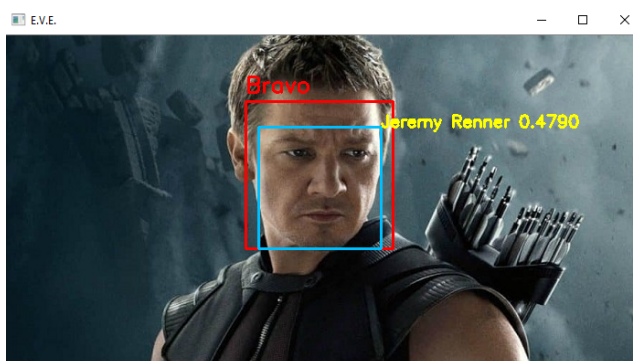


Foto: Reprodução/Geek Blog.

Anexo 06 - Imagem utilizada para validação da do programa E.V.E. versão gama.

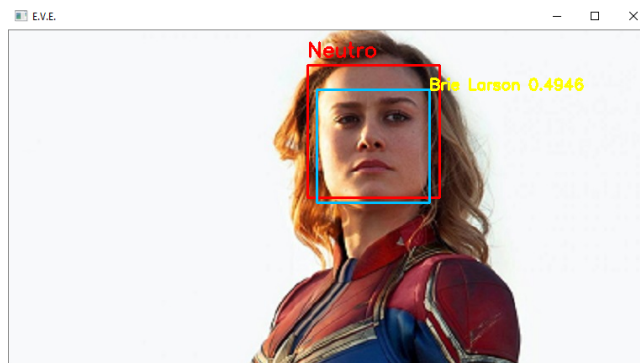


Foto: Reprodução/Ei Nerd. Disponível em:

<<https://www.einerd.com.br/vingadores-ultimato-estrela-2-anos/>>

Anexo 07 - Primeira implementação gráfica com o Tkinter.

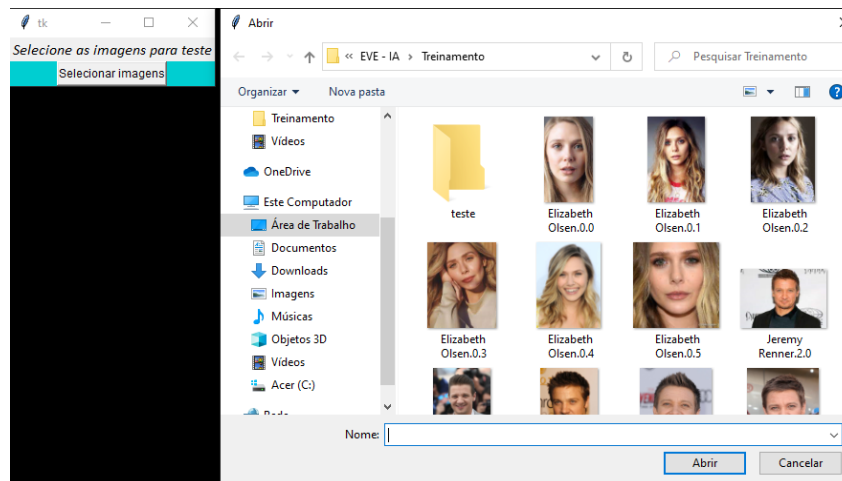


Foto: Arquivo pessoal do projeto