

DEEP-EMOTIVE

PROTÓTIPO DE SISTEMA PARA RECONHECIMENTO DE
EXPRESSÕES FACIAIS UTILIZANDO APRENDIZADO
PROFUNDO

DIÓGENES ADEMIR DOMINGOS

ANDREZA SARTORI - ORIENTADORA

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Roteiro

- **Introdução**
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Introdução

- Cisco Systems (2016)
 - 73,1 Exabytes (EB) de tráfego de dados
- Inteligência artificial
- Computação Afetiva
 - Picard (1997)
- Emoções
 - Alegria, Tristeza, Medo, Desgosto, Raiva, Surpresa - Ekman et al. (1987)

Roteiro

- Introdução
- **Objetivos**
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Objetivo geral

Desenvolver um protótipo capaz de reconhecer as emoções universais de Ekman através da identificação das expressões faciais obtidas em imagens.

Objetivos específicos

- Reconhecer a face em imagens utilizando técnicas de visão computacional;
- Extrair as características faciais relevantes para identificar as expressões faciais utilizando técnicas de visão computacional;
- Implementar um classificador de emoções utilizando redes neurais artificiais e aprendizado profundo
- Reconhecer as seis emoções universais definidas por Ekman (1987).

Roteiro

- Introdução
- Objetivos
- **Fundamentação teórica**
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Fundamentação teórica

- Conceito de emoção e as expressões faciais.
- Conceito de computação afetiva
- Conceitos de visão computacional.
- Características de redes neurais
- Aprendizado Profundo ou Deep Learning

Emoções

- Padrões formados por reações químicas e neurais (DAMÁSIO, 2000).
- As emoções desempenham um papel fundamental (REEVE, 2006).
- Reações apropriadas e adaptativas (SCHULTZ et al. 2001).



Fonte: extraídas da base de dados CK+ (LUCEY et al. 2010).

Expressões faciais e Computação Afetiva

- Sistema de Codificação da Ação Facial – FACS (EKMAN; FRIESEN, 1976).
- Inflexão de voz, Comportamentos observáveis, Expressões faciais, Linguagem corporal (PICARD, 1997)

AU 1	AU 2	AU 4
		
Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together
AU 5	AU 6	AU 7
		
Upper eyelids are raised.	Cheeks are raised.	Lower eyelids are raised.
AU 1+4	AU 4+5	AU 1+2
		
Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.	Inner and outer portions of the brows are raised.
AU 1+2+4	AU 1+2+5+6+7	AU 0 (neutral)
		
Brows are pulled together and upward.	Brow, eyelids, and cheek are raised.	Eyes, brow, and cheek are relaxed.

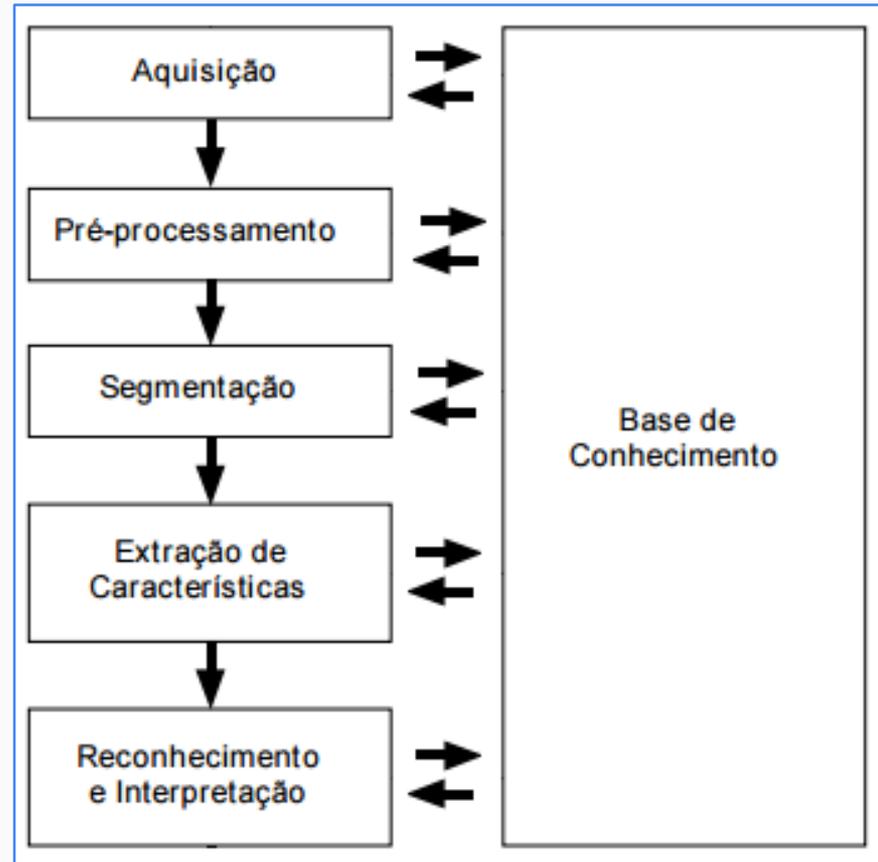
Fonte: Adaptado de Tian, Kanade e Cohn (2001, p. 3).

Visão computacional

- Visão humana é capaz forma, iluminação e distribuição de cor, e ainda segmentar objetos na cena observada (SZELISKI, 2011).
- Visão Computacional tenta realizar o inverso, transformar essas propriedades em dados que o computador consiga interpretar (SZELISKI, 2011).

Processamento Digital de Imagens

- Pré-processamento
- Segmentação
- Extração de características
- Reconhecimento de padrões



Fonte: adaptado de Marques Filho e Vieira Neto (1999, p. 9).

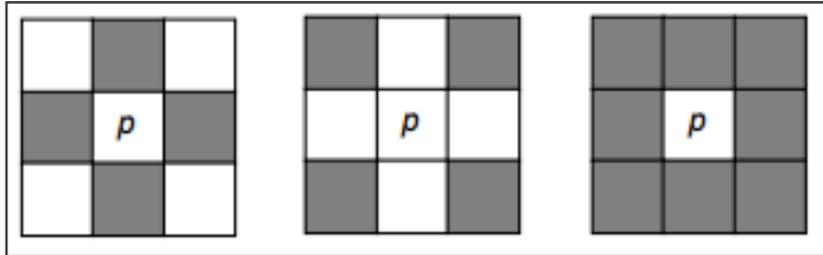
Imagem Digital

$$f(x, y) \approx \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, M-1) \\ f(1,0) & f(1,1) & \dots & f(1, M-1) \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1, M-1) \end{bmatrix}$$

Fonte: Gonzalez e Woods (2010).

Pré-processamento

Núcleos

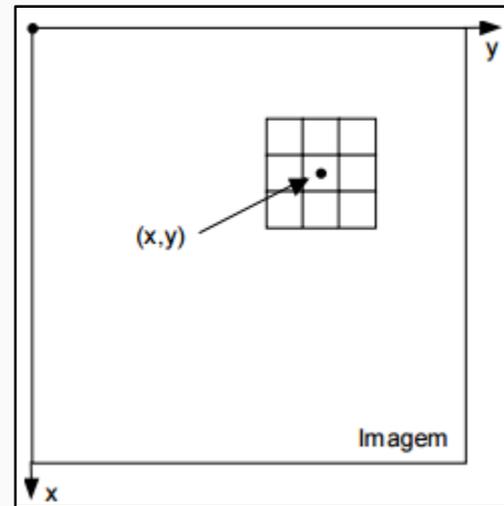


Fonte: Marques Filho e Vieira Neto (1999, p. 26).

Convolução discreta

$$S(x,y) = (K * I)(x,y) = \sum_M \sum_N I(x,y)K(x-m,y-n)$$

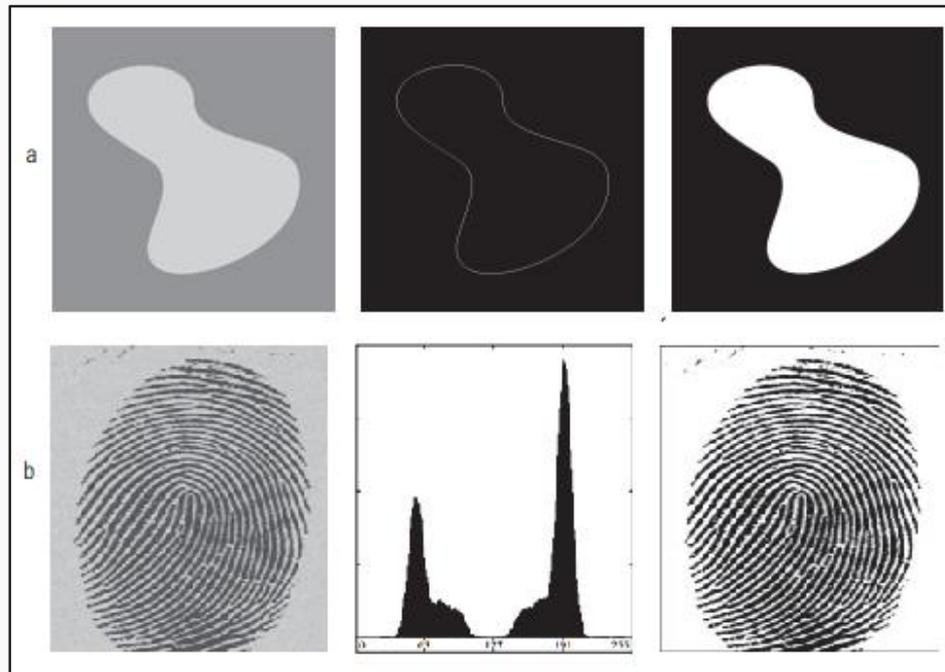
Deslocamento do núcleo sobre a imagem



Fonte: Gonzalez e Woods (2010, p. 84).

Segmentação

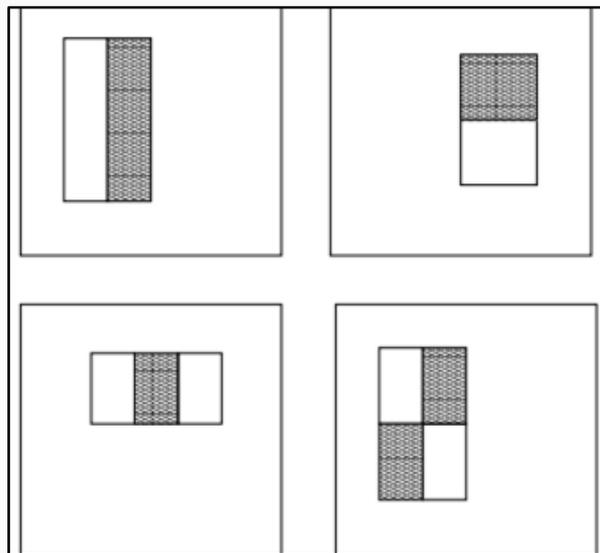
- Subdivide uma imagem em partes ou regiões de interesse (GONZALEZ; WOODS, 2010).
- Descontinuidade (a)
- Similaridade (b)



Fonte: Gonzalez e Woods (2010, p.455 e 489).

Extração de características

- Reconhecimento da face em uma imagem (VIOLA; JONES, 2001).



Fonte: Viola e Jones (2001, p. 512).



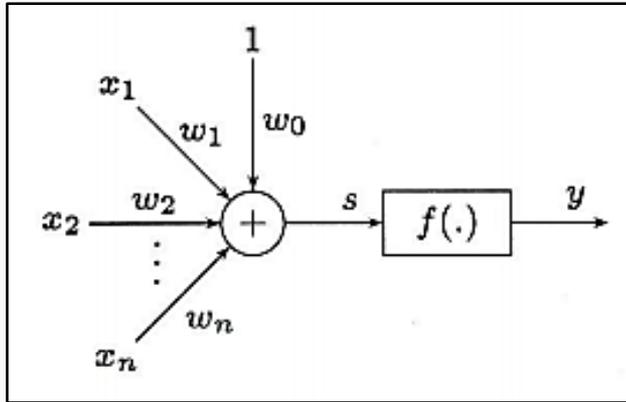
Fonte: Viola e Jones (2001, p. 514).

Reconhecimento de padrões

- Processo que atribui um rótulo a um objeto (GONZALEZ; WOODS, 2010)
- Matching
- Redes Neurais Artificiais

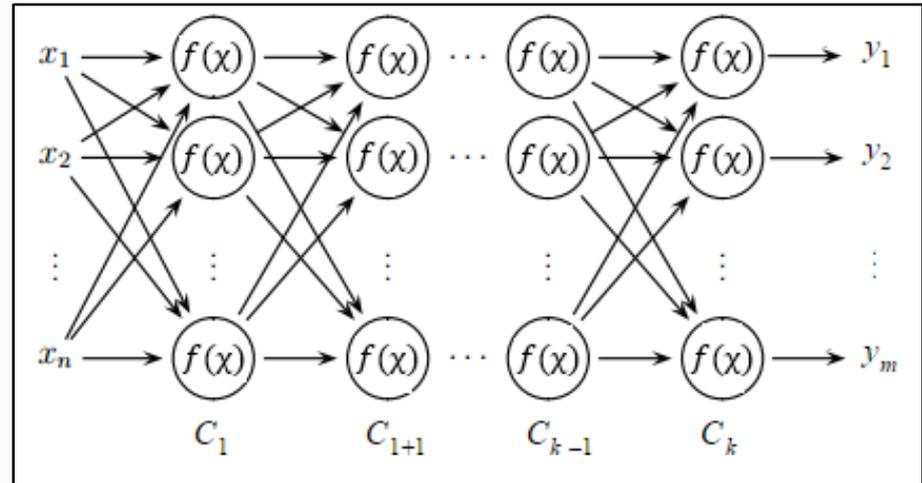
Redes Neurais Artificiais

Mcculloch e Pitts (1943)



Fonte: Marques (2005, p. 164).

Perceptron Multi camadas



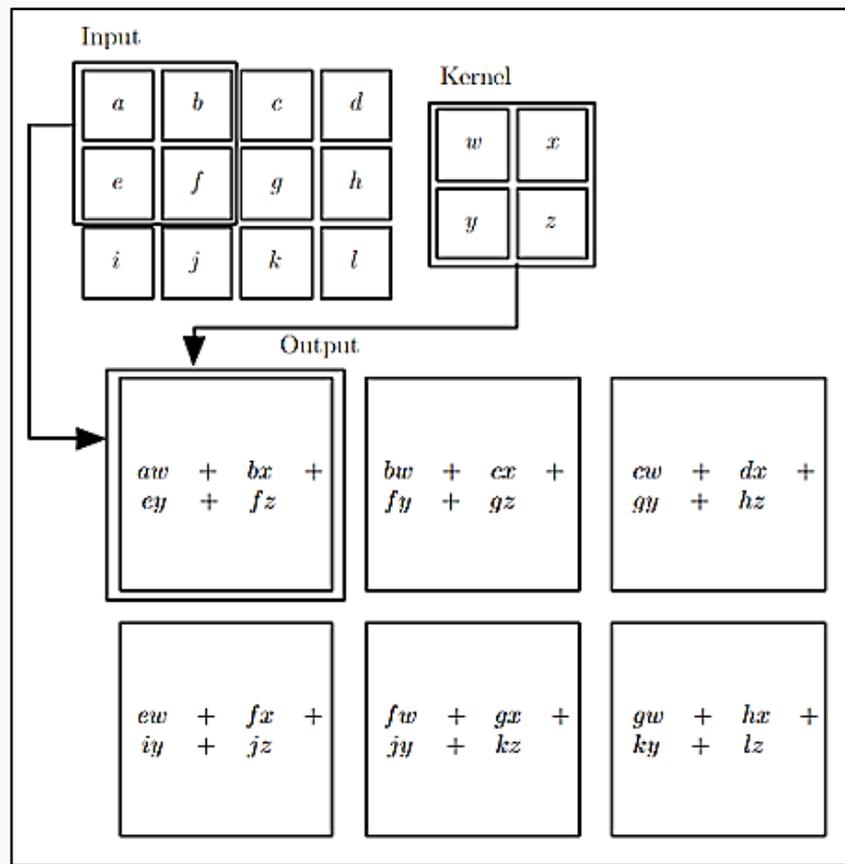
Fonte: baseado em Gonzalez e Woods (2010, p. 585).

Aprendizado profundo – Deep Learning

- Camadas dispostas hierarquicamente
- Camadas de baixo e alto nível
- Análogo ao sistema biológico (DICARLO; ZOCCOLAN; RUST, 2012)
- Aprendizagem de representação (LECUN; BENGIO; HINTON, 2015)

Redes neurais convolucionais

- Reconhecimento de padrões em imagens (GOODFELLOW; BENGIO; COURVILLE, 2016).
- Convolução
- Núcleos
- Mapas de características
- Função de ativação



Fonte: Goodfellow, Bengio e Courville (2016, p. 330).

Subamostragem

Max Pooling

1	2	2	4
4	5	6	3
3	2	8	7
1	2	4	5

→

5	6
3	8

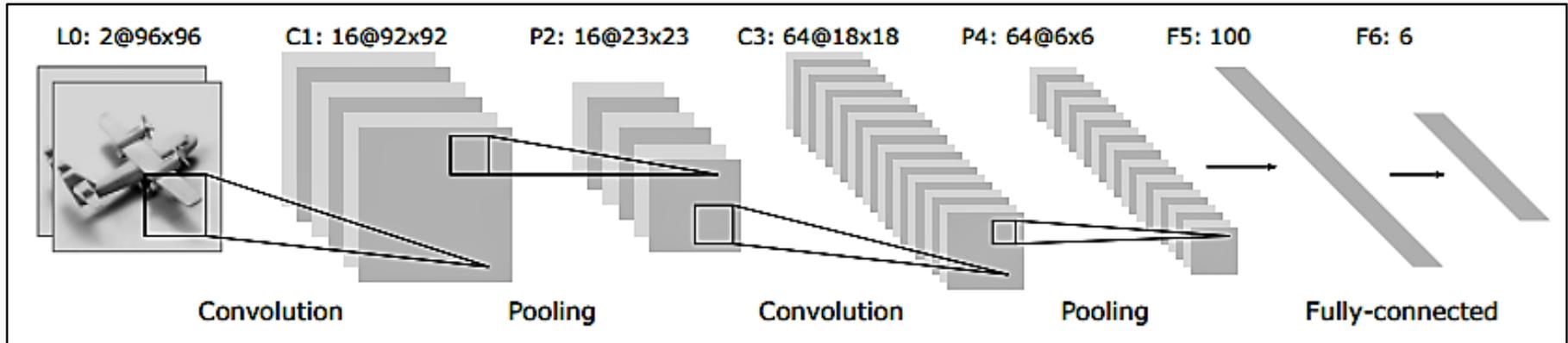
Average Pooling

1	2	2	4
4	5	6	3
3	2	8	7
1	2	4	5

→

3	3,75
2	6

Arquitetura de uma Rede Neural Convolutacional



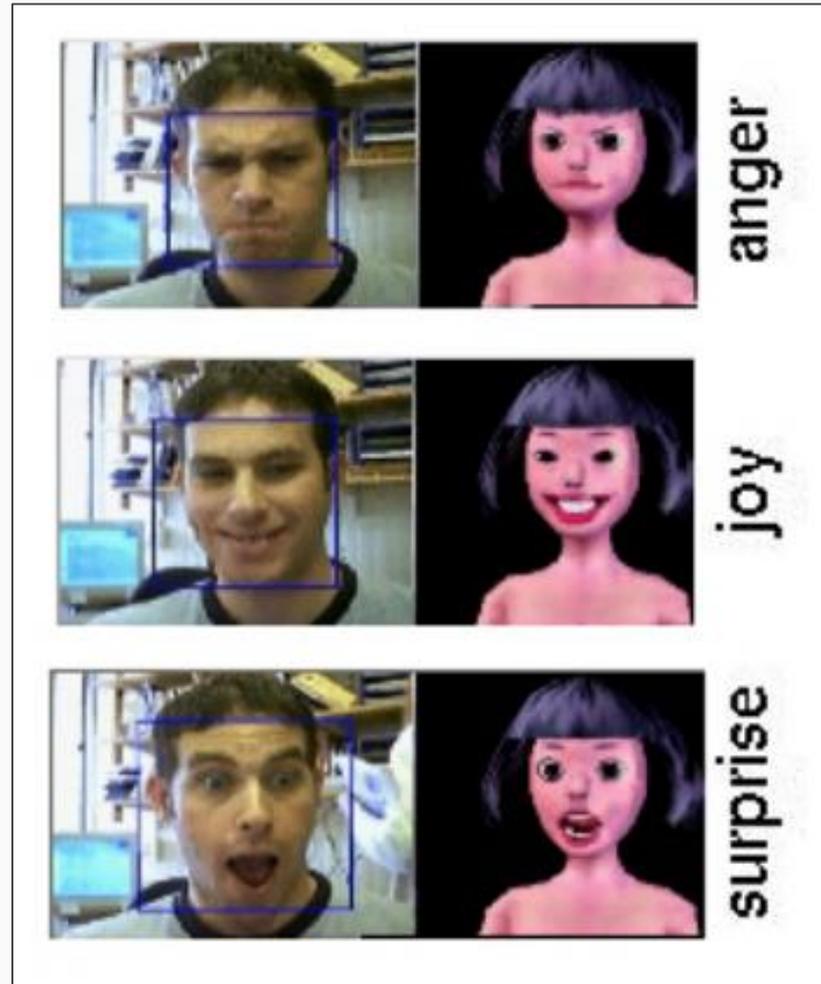
Fonte: Scherer, Müller e Behnke (2010, p 3).

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- **Trabalhos correlatos**
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

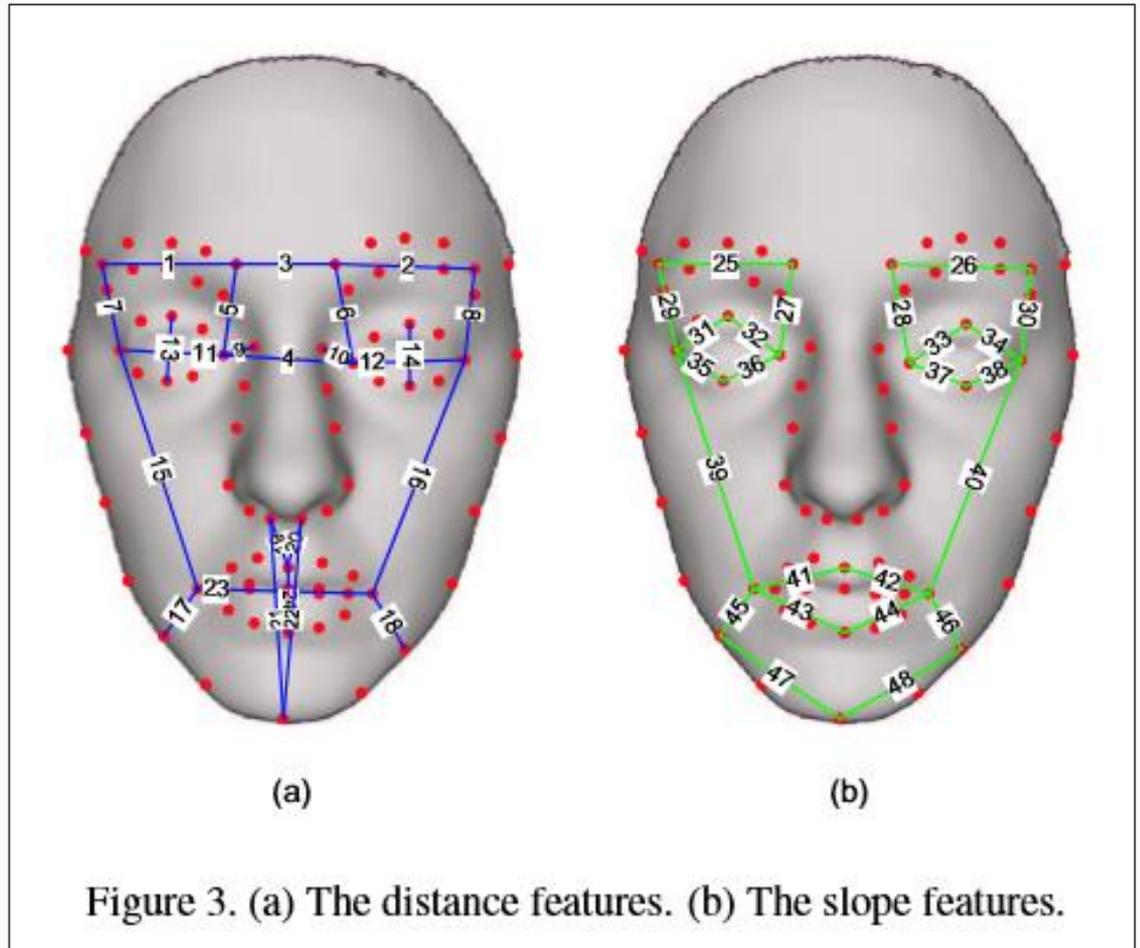
Barlett et al. (2003)

- Tempo real.
- Reconhece sete expressões faciais (Neutra)
- Viola-Jones
- Classificador SVM
- Precisão de 93%



Tang e Huang (2008)

- Expressões faciais em 3D
- Reconhece seis expressões faciais
- Distância entre pontos
- Classificador SVM
- Precisão de 87,1%



Candra et al. (2016)

- Expressões faciais para aplicação em psicoterapias.
- Reconhece sete expressões faciais (Desprezo).
- Viola-Jones e E-HOG
- Classificador SVM
- Precisão de 96,4%

Detected Face using Viola-Jones' cascaded face detector



Detected Face Image



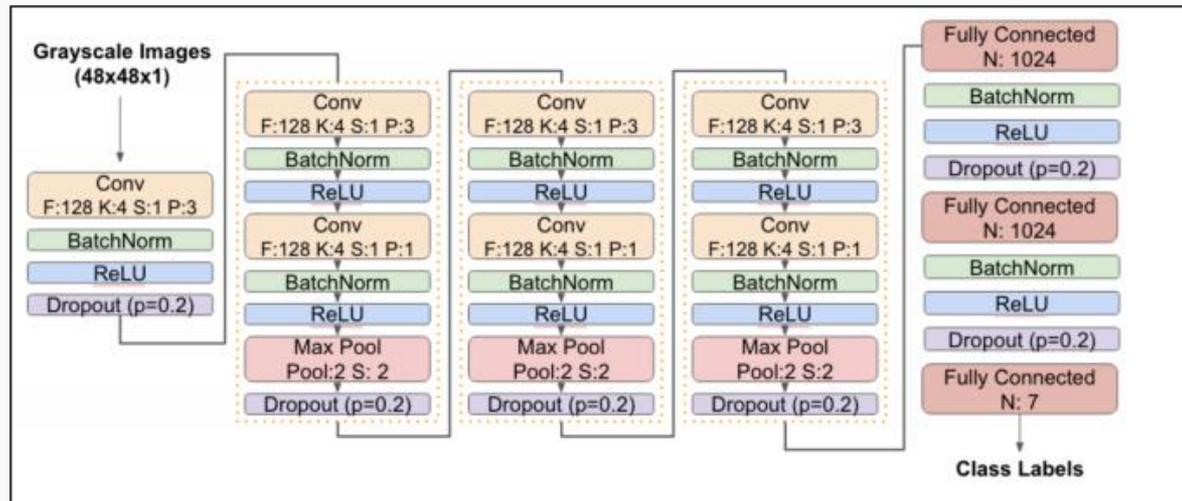
Detected eyes using Viola-Jones' cascaded eyes detector on the upper part of the face



Detected mouth using Viola-Jones' cascaded mouth detector on the lower part of the face

Amin, Chase e Sinha (2017)

- Reconhecimento de expressões faciais
- Aprendizado Profundo
- Sete emoções
- Precisão de 61,05%



Fonte: Amin, Chase e Sinha (2017, p. 4).

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- **Requisitos**
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Requisitos

O protótipo tem como requisitos funcionais:

- Realizar a detecção da face em um imagem
- Criar uma rede neural profunda
- Utilizar a técnica de Aprendizado Profundo para o aprendizado de máquina
- Permitir que o usuário insira imagens no sistema para reconhecimento
- Classificar a emoção através das expressões faciais.

Requisitos

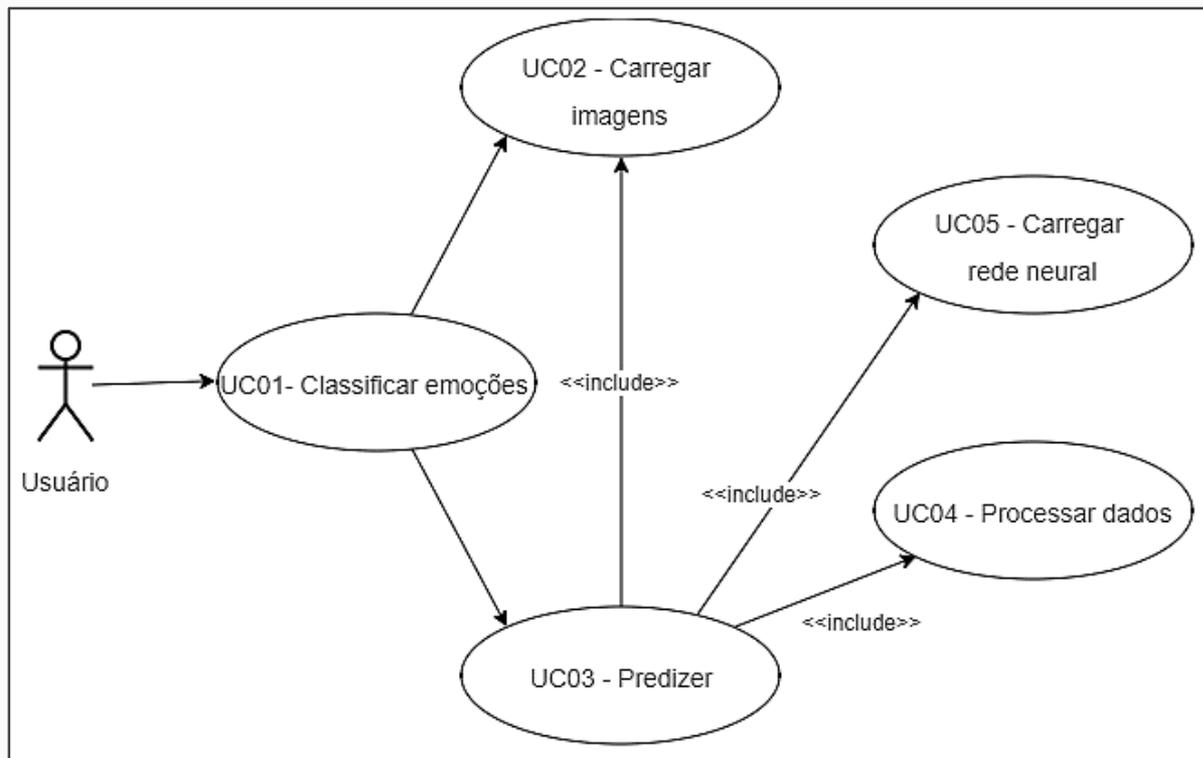
O protótipo tem como requisitos não funcionais:

- Ser desenvolvido utilizando a linguagem Python
- Utilizar Jupyter Notebook como ambiente de desenvolvimento
- Utilizar o framework Keras para aplicação da Aprendizado Profundo
- Utilizar frameworks OpenCV, Scikit Learn, Numpy, Matplotlib para computação científica e visualização dos dados
- Reconhecer imagens nos formatos JPG e PNG (RNF).

Roteiro

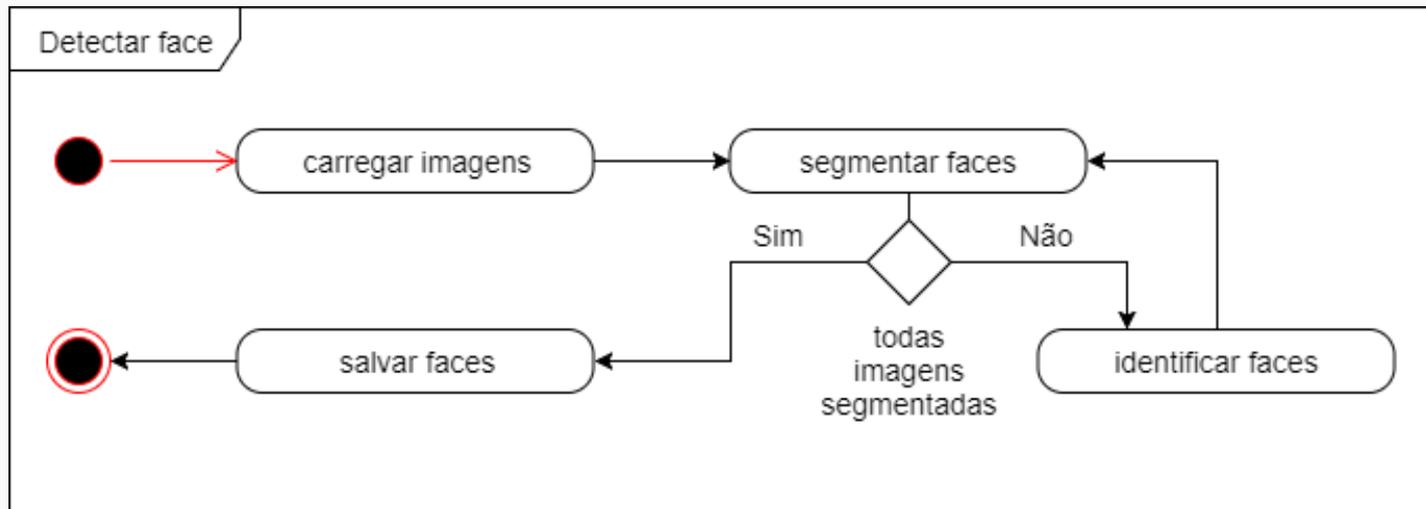
- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- **Especificação**
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Especificação – Diagrama de caso de uso



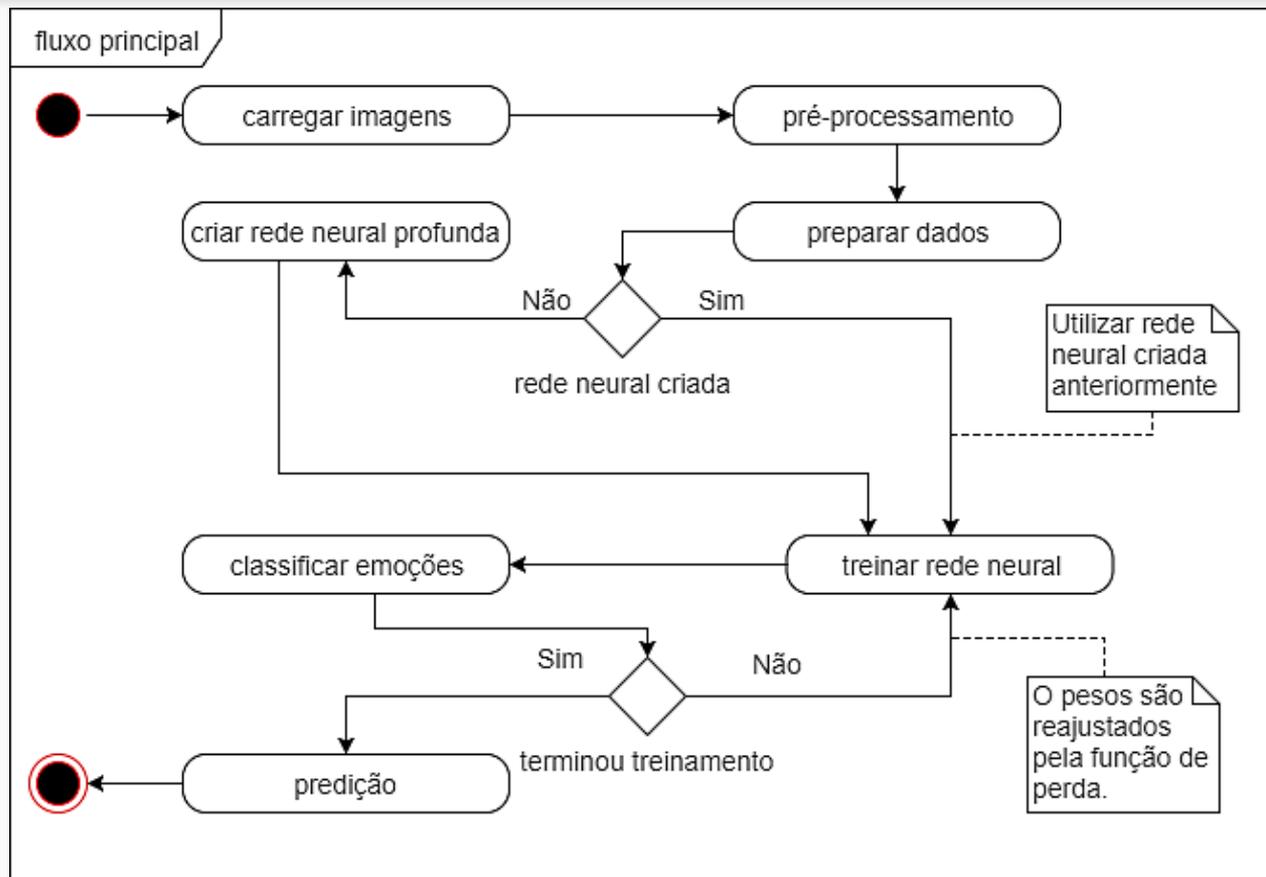
Fonte: elaborado pelo autor.

Especificação – Diagrama de atividades



Fonte: elaborado pelo autor.

Especificação – Diagrama de atividades



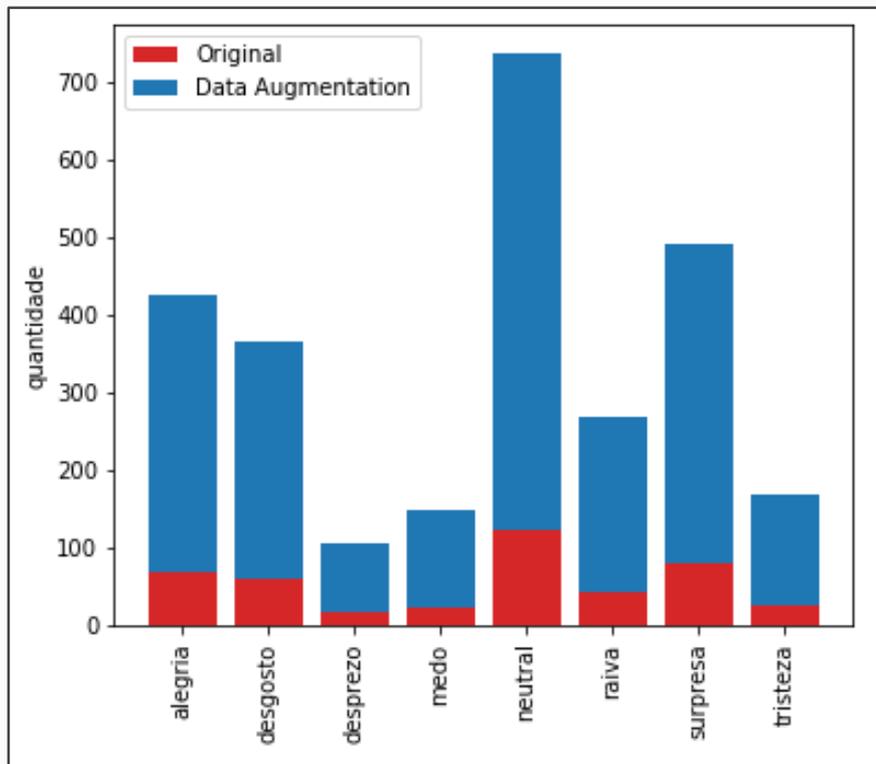
Fonte: elaborado pelo autor.

Roteiro

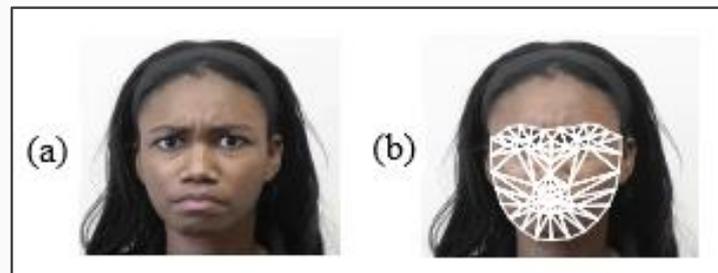
- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- **Implementação**
- Operacionalidade da implementação
- Resultado e discussões

Implementação

- Data Augmentation 2718

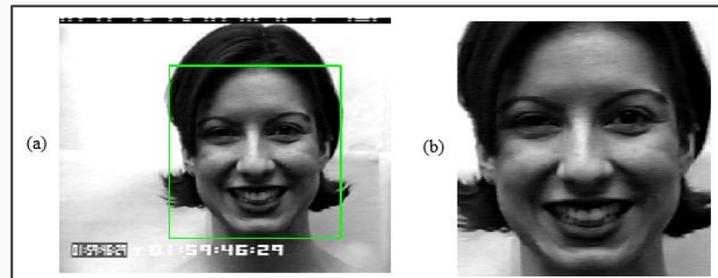


- Refinamento das imagens 593/453



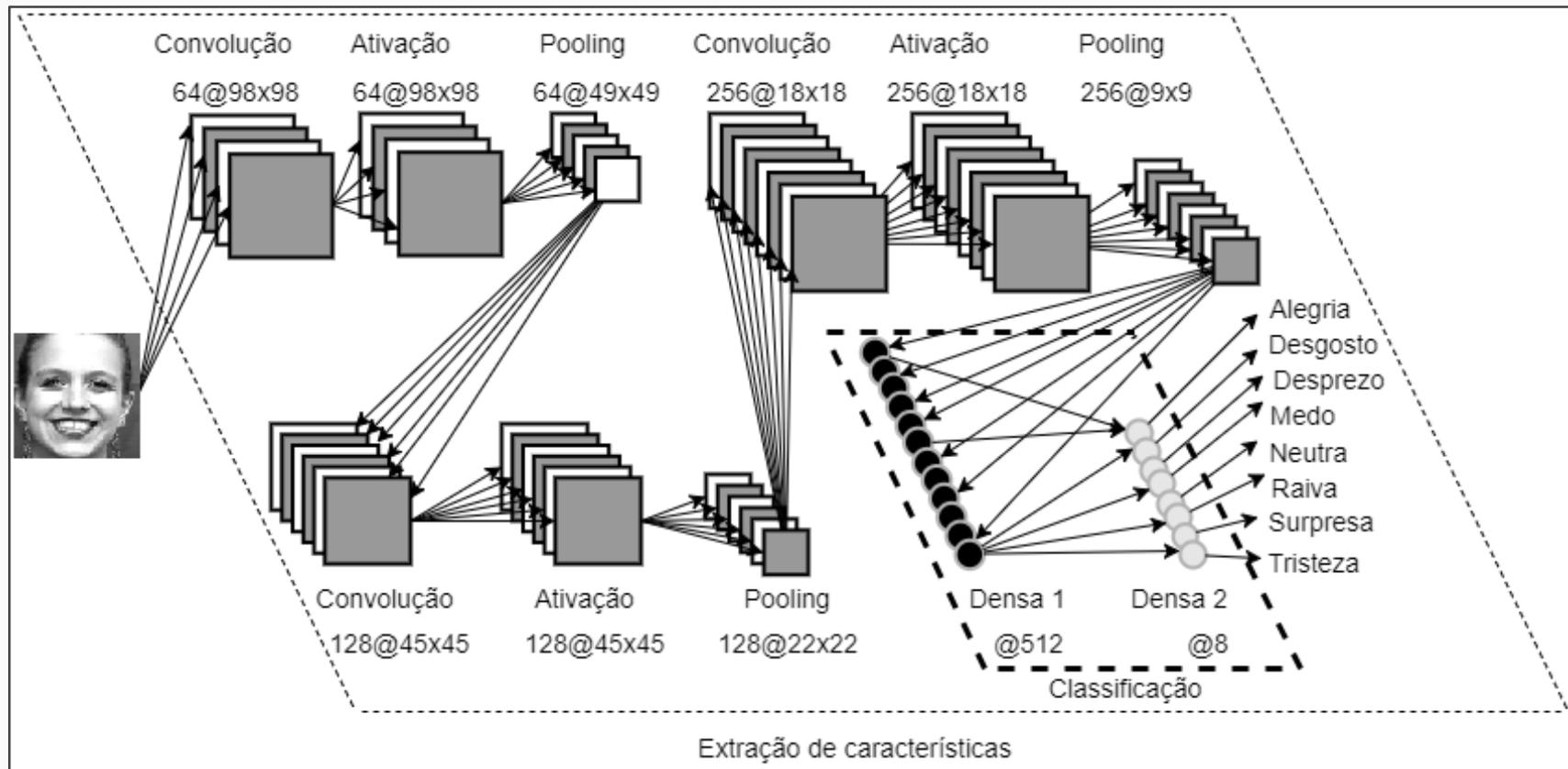
Fonte: Lucey et al. (2010).

- Detecção da face



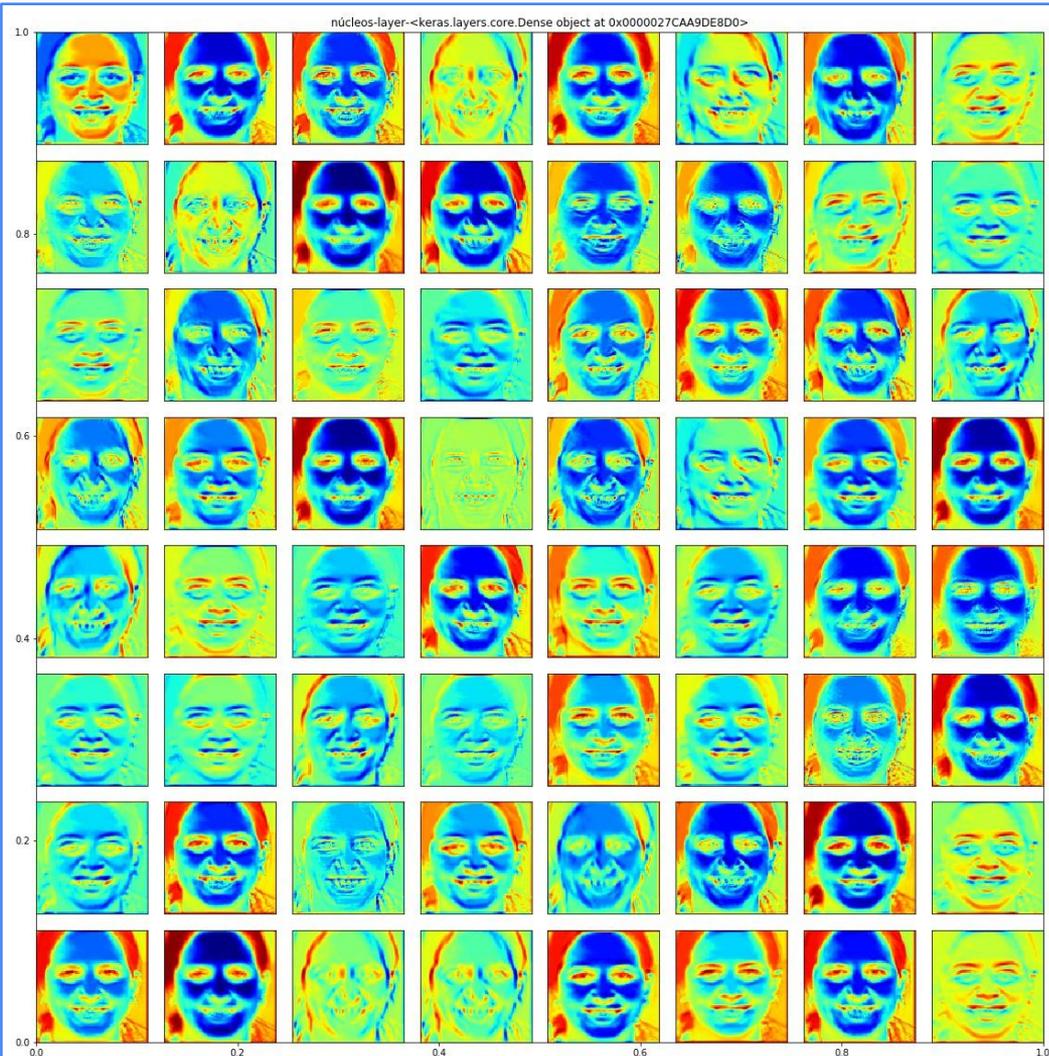
Implementação

- Rede Convolutacional profunda – Deep-Emotive



Implementação

- 64 mapas de características
- Saída da primeira camada de convolução
- 128 Segunda camada
- 256 Terceira camada



Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- Resultado e discussões

Operacionalidade

- Classificar a emoção
- Uma imagem
- Várias imagens
- Modelo treinado

```
$ python classificar_emocao.py -m "teste_usuario/imagem_usuario.png"

*****
Prototipo DeepEmotive
Modulo de classificacao de emocao de uma imagem.

*****
*****
Importando imagem: teste_usuario/imagem_usuario.png
Imagem importada com sucesso!

*****
Formato dos pixels da imagens: (98, 98)

Configurando o array de pixels para procesamento da rede...

Trabalhando com tensorflow: (98, 98, 1)
Novo formato dos pixles da imagem: (1, 98, 98, 1)
Preparacao da imagem completa!

*****
Carregando o modelo...
Modelo carregado com sucesso!

Carregando os melhores pesos...
Carregamento dos melhores pesos completo!

*****
*****
1/1 [=====] - 0sUsing TensorFlow backend.

Classificando a imagem...

probabilidade da emocao: alegria:          0.80%
probabilidade da emocao: desgosto:        0.00%
probabilidade da emocao: desprezo:        0.00%
probabilidade da emocao: medo:            0.20%
probabilidade da emocao: neutral:         0.00%
probabilidade da emocao: raiva:           0.00%
probabilidade da emocao: surpresa:       0.00%
probabilidade da emocao: tristeza:        0.00%

Emocao classificada pela rede: alegria
```

Roteiro

- Introdução
- Objetivos
- Fundamentação teórica
- Trabalhos correlatos
- Requisitos
- Especificação
- Implementação
- Operacionalidade da implementação
- **Resultado e discussões**

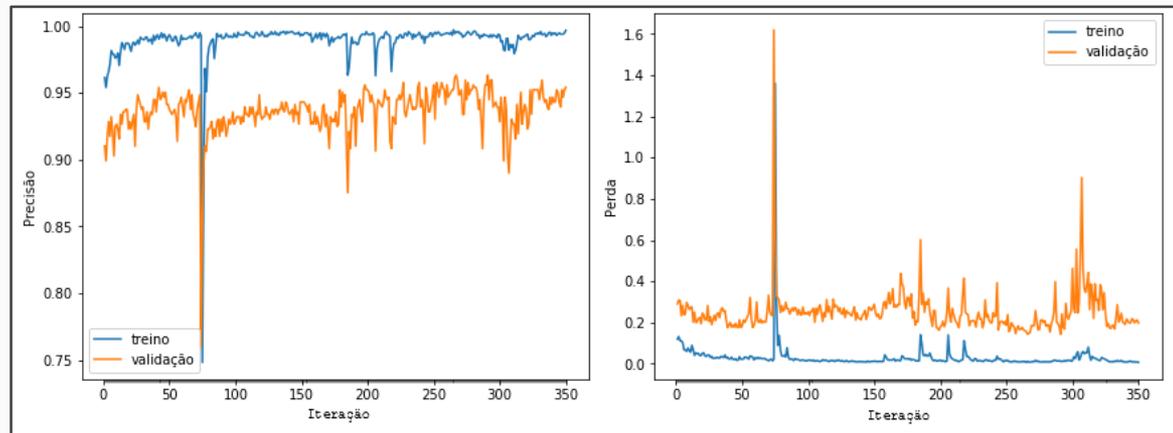
Resultado e discussões

- 16 modelos
- Hiperparâmetros
- Precisão de 96,33%

Experimento	Pixels	Campo receptivo	Núcleos	Função Ativação	Alteração	Precisão %
Modelo 1	640x490	3x3	32,64,128,128	ReLU	---	---
Modelo 2	150x150	3x3	32, 64, 128	ReLU	Redimensionar	31,08
Modelo 3	150x150	3x3	32, 64, 128	ReLU	Dropout	37,36
Modelo 4	150x150	3x3	32, 64, 128	ReLU	Normalização	24,18
Modelo 5	224x224	5x5	64, 128, 256	PReLU	Sem normalização	52,75
Modelo 6.1	224x224	5x5	64, 128, 256	PReLU	Otimizador - Sgd	27,89
Modelo 6.2	224x224	5x5	64, 128, 256	PReLU	Otimizador -Nadam	28,96
Modelo 6.3	224x224	5x5	64, 128, 256	PReLU	Otimizador - Adagrad	35,96
Modelo 6.4	224x224	5x5	64, 128, 256	PReLU	Otimizador - Adamax	64,85
Modelo 6.5	224x224	5x5	64, 128, 256	PReLU	Otimizador - Adadelta	65,87
Modelo 6.6	224x224	5x5	64, 128, 256	PReLU	Otimizador - Rmsprop	81,65
Modelo 6.7	224x224	5x5	64, 128, 256	PReLU	Otimizador - Adam	85,69
Modelo 7	98x98	5x5	64, 128, 256	PReLU	Segmentação	90,00
Modelo 8	98x98	5x5	64, 128, 256	PReLU	Paralelismo	47,05
Modelo 9	98x98	5x5	64, 128, 256	PReLU	GlobalMaxPooling	86,97
Modelo 10	98x98	5x5	64, 128, 256	PReLU	AveragePooling	96,33

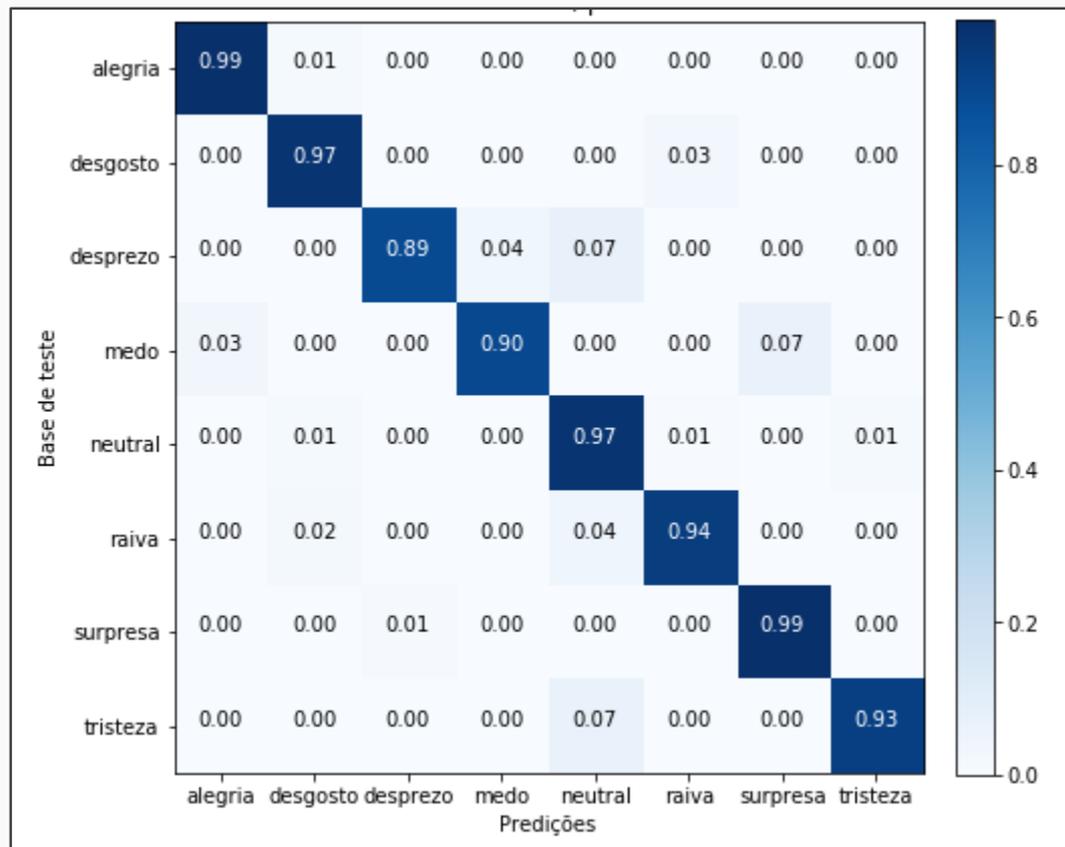
Resultado e discussões

- Gráfico de aprendizado
- Emoções previstas



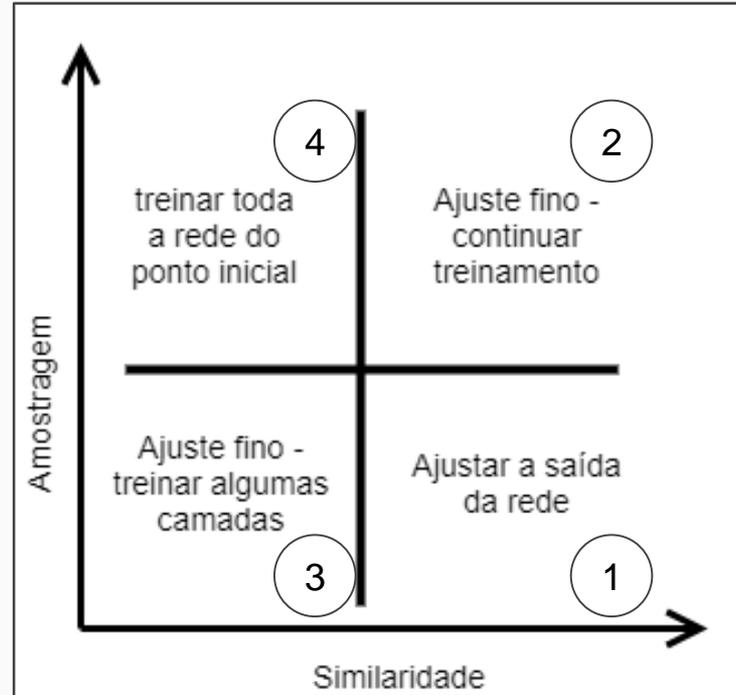
Resultado e discussões

- Matriz de confusão
- Desprezo, 108 imagens
- Medo, 150 imagens



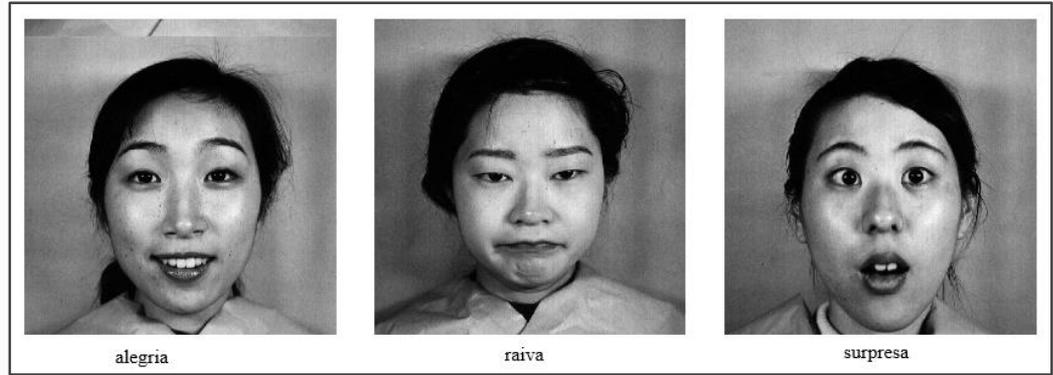
Transferência de aprendizado

- Utilizar o conhecimento adquirido na resolução de um problema e aplicado para um novo problema relacionado (PAN; YANG, 2010).



Transferência de aprendizado

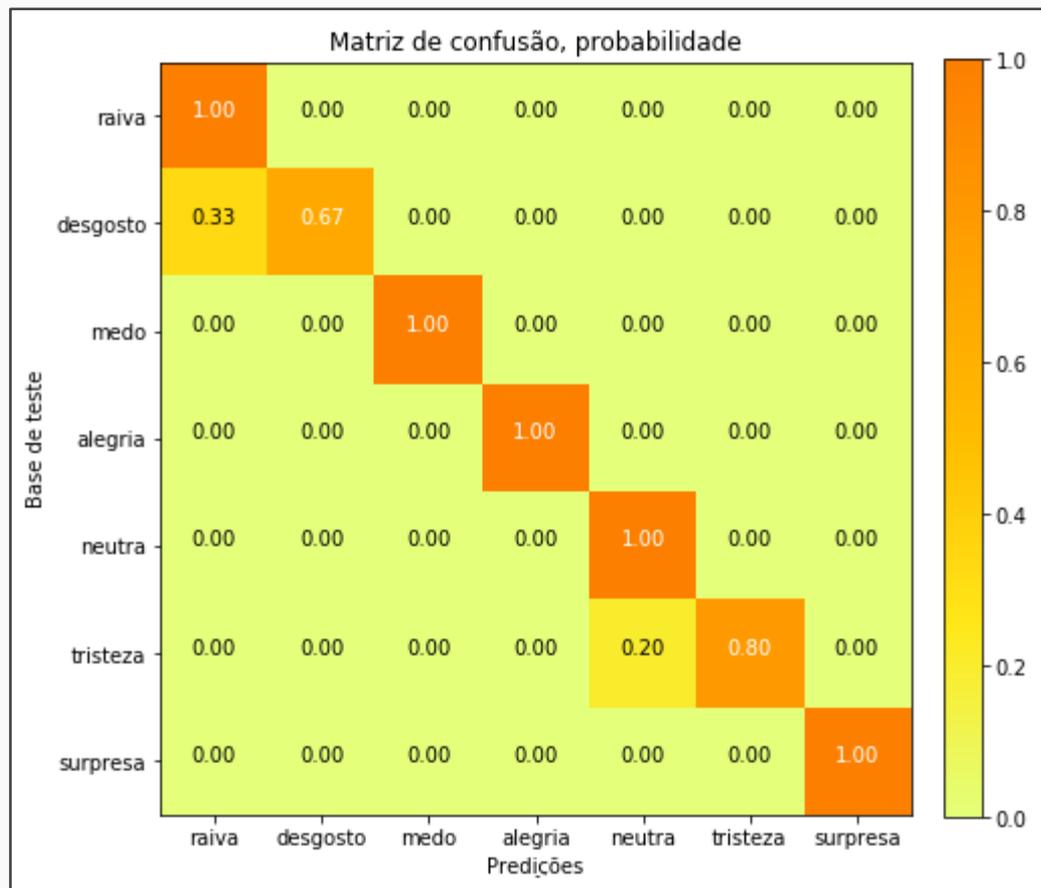
- Base de dados Japanese Female Facial Expression (JAFFE).
- Similaridade
- Bom desempenho
- 213 imagens



Trabalho	Lyons et al. (1998)	Ducan; Shine; English (2016)	DeepEmotive (2017)
Técnica	Filtros Garbor	Aprendizado Profundo	Aprendizado Profundo
Precisão	67%	57,1%	93,02%

Transferência de aprendizado

- Base de dados Japanese Female Facial Expression (JAFFE).
- Matriz de confusão



Transferência de aprendizado

- Base de dados Facial Expression Recognition 2013 (FER-2013)

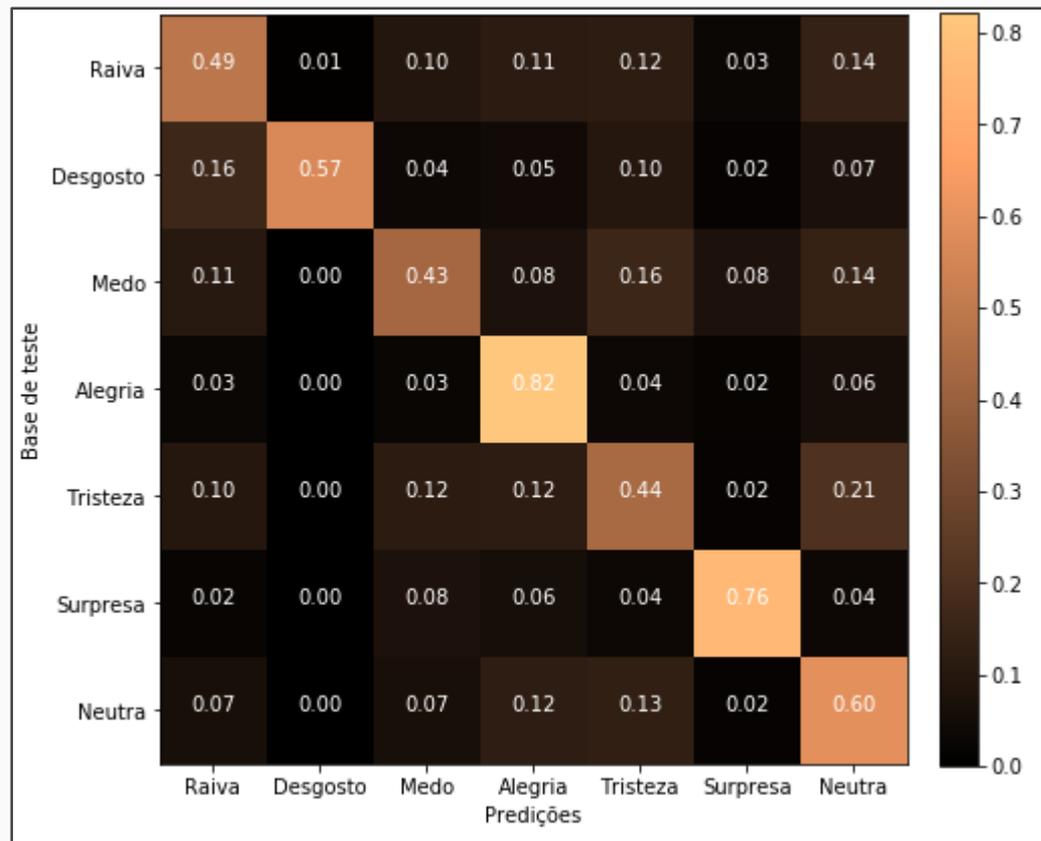
- Dissimilaridade
- Problemas de visão computacional
- The ICML 2013
- 28709 imagens



Trabalho	Tang (2013)	DeepEmotive (2017)
Técnica	Aprendizado profundo	Aprendizado profundo
Classificador	L2-SVM	Softmax
Precisão	71,16%	60,62%

Transferência de aprendizado

- Base de dados Facial Expression Recognition 2013 (FER-2013)
- Matriz de confusão



Comparação dos resultados

Características / Trabalho	Bartlett et al. (2003)	Tang; Huang, (2008)	Candra et al. (2016)	Amin; Chase; Sinha (2017)	DeepEmotive (2017)
Dimensão das imagens	2D	3D	2D	2D	2D
Extrator de características	GBoost, Viola-Jones	Distância entre Pontos	Viola-Jones, E-HOG	Aprendizado profundo	Viola-Jones, Aprendizado profundo
Classificador	SVM	SVM	SVM	Softmax	Softmax
Emoções reconhecidas	7 (Ekman et al. 1987) + Neutra	6 (Ekman et al. 1987)	7 (Ekman et al. 1987) + Desprezo	7 (Ekman et al. 1987) + Neutra	8 (Ekman et al. 1987) + Desprezo e Neutra
Tempo real	Sim	Não	Não	Não	Não
Precisão	93%	87,1%	96,4%	61,05%	96,33%

Conclusões

- O trabalho alcançou o objetivo proposto de reconhecer seis emoções
- O trabalho superou o objetivo reconhecendo oito emoções (desprezo e neutra)
- Transferência de aprendizado validou o algoritmo
- Transferência de aprendizado apresentou limitações do protótipo
- Tempo de processamento
- Novos algoritmos
- Interação homem máquina, marketing, e no monitoramento de pacientes.

Extensões

- Aprimorar o conhecimento da rede neural, através da transferência de aprendizado, aumentando a generalização do conhecimento da rede
- Implementar a estrutura proposta por este trabalho, em um ambiente tecnológico com maiores recursos, a fim de fornecer mais características a serem aprendidas pela rede
- Treinar a rede neural em uma nova base de dados, que possua uma distribuição de amostras mais homogênea, para que seja possível aprimorar a precisão do reconhecimento das emoções de desprezo e medo
- Desenvolver uma API para distribuição do conhecimento adquirido pela rede, tornando possível sua utilização por outras áreas de estudo

Extensões

- Fazer experimentos utilizando outro tipo de classificador e verificar a precisão em uma base de dados dissimilar como a FER-2013;
- Complementar o protótipo com a funcionalidade de reconhecimento em tempo real;
- Utilizar o modelo preditivo, criado por este trabalho, em conjunto com outras abordagens para análise de sentimento, como análise de emoção em frames de vídeos.

DEMO