Descrição de Abordagem de Reconhecimento Facial utilizando Arquitetura *Convolution Neural Network*

Rodolfo Labiapari Mansur Guimarães
Departamento de Computação
Universidade Federal de Ouro Preto - UFOP
Ouro Preto - MG – Brasil
rodolfolabiapari@decom.ufop.br

***Resumo*—Current approaches to human face detection problems is following neural network algorithms. This process is necessary for big difficulty of computational models to have on recognition of objects in images where there are different faces with different backgrounds. Techniques that use neural convolutional network architecture have a great prominence in the academic environment, because it is a procedure that obtains good quality accuracy results and hasn't large losses with trade-off and can even be processed in personal environments and with low processing power.**

***Keywords*-Face Detection; Convolutional Neural Network.**

***Resumo*—Abordagens atuais para problemas de detecção de face humana utilizando algoritmos é baseada em redes neurais. Tal processo é necessário pela grande dificuldade de modelos computacionais terem sucesso no reconhecimento de objetos em imagens onde existam diferentes faces com diferentes e complexos fundos. Técnicas que utilizam arquitetura tipo *convolutional neural network* estão tendo bastante destaque no ambiente acadêmico, por ser um procedimento que obtêm resultados com boa acurácia e sem ter grandes perdas com *trade-off* podendo até ser processado em ambientes pessoais e de baixo poder de processamento.**

***Keywords*-Deteção de Face; Rede Neural Convolucional.**

I. Introdução

A. Apresentação

Detecção de face é um problema pertencente à visão computacional. Métodos modernos podem facilmente detectar faces em ambientes controlados, ou seja, com fundo que respeite um padrão pré-estabelecido, sem obstrução facial como óculos, barba, etc.

Uma criança de três anos de idade consegue reconhecer a foto de um pássaro. Parece algo fácil, mas cientistas da computação tiveram como enigma este problema por cerca de 50 anos.

Como o estado da arte é concreto em detecção de face em ambientes controlados, pesquisas mas recentemente partem a buscar métodos que tivessem foco principalmente em detecção de face em ambientes onde existe uma maior complexidade, sem perder a eficiência já obtida até então. As pesquisas em detecção de face hoje procuram métodos eficientes e rápidos para detecção onde elas aparecem não só posicionada em perfil mas sim, com distâncias variadas, oclusões, iluminações extremas entre outras variações que poderia representar mais situações que representassem o mundo real na captura de imagens [1].

As maiores dificuldades do tema atualmente são a larga variações de visualização de faces humanas em fundos não-padronizados e também a procura espacial onde cada face pode estar posicionada em diferentes posições e tamanhos.

Porém, utiliza-se nos dias atuais uma abordagem de reconhecimento de objetos usando arquitetura *deep convolutional neural network* também conhecida como *convolutional neural network* ou só CNN e ela será descrita neste trabalho.

O artigo está disposto em três seções principais onde Seção I-B define justificativas para a realização deste trabalho, além de um Referencial Teórico abordando o tema na Seção II. Em Seção III é descrito em detalhes o processo de reconhecimento de face utilizando arquitetura *convolutional neural network*.

B. Justificativa

Sistemas computacionais anteriores à redes neurais possuíam características bastante ineficientes para imagens que tinham como propriedade fundos não-padronizados.

Considerando o sucesso de algoritmos que utilizam *convolutional neural network* em reconhecimento de padrão em tarefas como áudio e imagens, como será descrito na Seção II, é possível ter bons resultados utilizando tal arquitetura para reconhecer faces humanas em imagens não padronizadas. Como o algoritmo não é de porte específico, é possível utilizá-lo em vários conjuntos de problemas criando assim um algoritmo robusto para deteção de padrões [1].

II. Referencial Teórico

A. Reconhecimento de Objetos utilizando Neural Networks

Proposta inicialmente por McCulloch e Pitts [2] as redes neurais computacionais, exibida na Figura 1, estão presentes a mais de 50 anos na computação e evoluíram drasticamente com o passar o tempo [3].

Inicialmente utilizava algoritmos de *backpropagation*, onde cada neurônio propagava seu resultado à todos os neurônios da camada adiante, como exibido na Figura 2. Sabendo que a ativação de determinado neurônio dar-se pelo processo de somatório de pesos, esta arquitetura de *feed-forward neural network* (FFNN) não é prática para vários tipos de problemas reais da computação. Um exemplo é a baixa habilidade de reconhecimento de objetos em visualizações. Isso deve-se ao

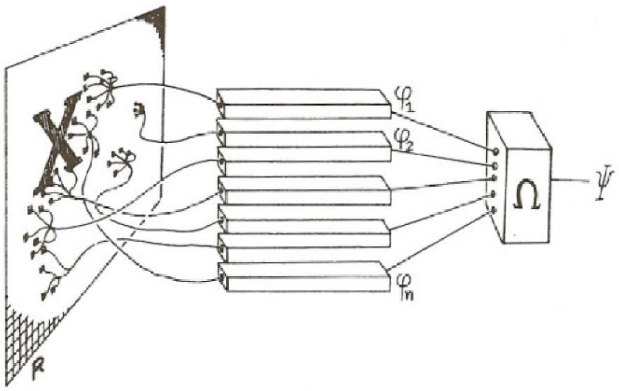

Figura 1. Rede neural de arquitetura tipo *feed-forward neural network*.

fato de que cada unidade da camada $n-1$ é conectada em todas as outras da camada n na sua propagação. Dessa forma, os números de neurônios e seus cálculos podem crescer exponencialmente com facilidade em relação à sua entrada [4].

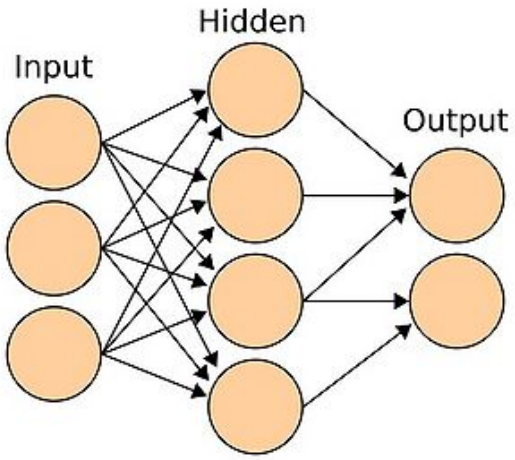

Figura 2. Rede neural FFNN simples.

Para contornar este problema foi necessário uma nova arquitetura que explorasse as restrições espaciais bidimensionais impostas pela sua modalidade de entrada onde reduz simultaneamente a quantidade de parâmetros na sua formação. O nome desta é *convolutional neural network*.

B. Proposta da Arquitetura Convolutional Neural Network

Como proposta ao problema do *feed-forward neural network* tem-se a arquitetura *Convolutional Neural Network*. É biologicamente inspirada em variantes de *Multilayer Perceptron* Convolutional Neural Networks (CNN) are biologically-inspired variants of MLPs.

Ela explora a correlação espacial reforçando um padrão de conectividade local entre neurônios nas camadas adjacentes, ou seja, as entradas da camada m são de um conjunto da camada anterior $m-1$ relacionados de uma forma contígua como é exibido na Figura 3 [5].

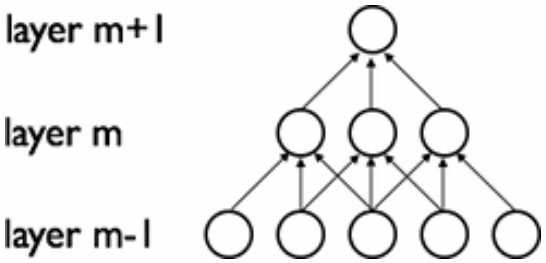

Figura 3. Exemplo das camadas de uma rede *convolutional neural network*.

O algoritmo baseia-se em vários pequenos processos. São eles a *Convolution*, *Subsampling* e a mescla dos dois processos formando o maior.

O processo de *Convolução* trata-se da prática de aplicar repetidas vezes a saída de determinada função como entrada de outra. Um termo comumente utilizado para tal processo é também *filtro*, onde a entrada passa por avaliações antes de entrar na segunda e consecutivamente. Já o processo de *Subsampling* utiliza o conhecido algoritmo de *max_pooling* para reduzir o tamanho de amostras a ser processada pela rede que fará a avaliação booleana da objeto a ser reconhecido [6].

Tais processos serão praticados na Seção III-A.

III. Desenvolvimento

A. Procedimentos de Operação de uma Convolutional Neural Network

Neural Networks podem resolver problemas computacionais complexos utilizando um conjunto simples de neurônios como por exemplo a estimação de preço de uma casa como e exibido na Figura 4.

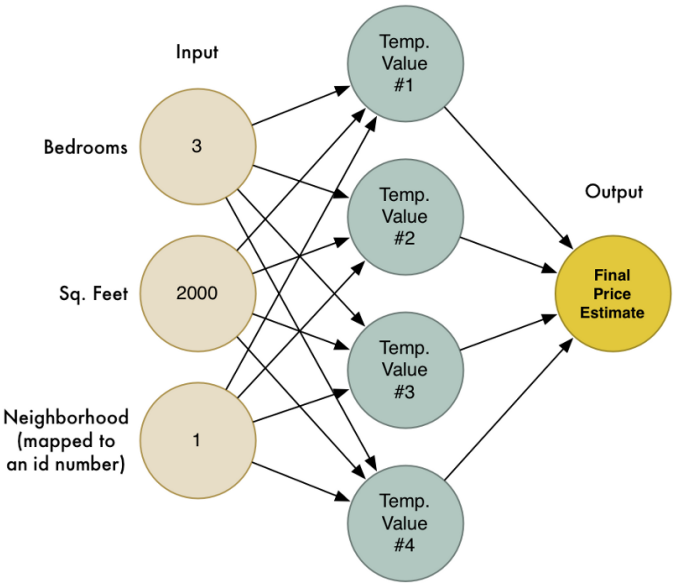

Figura 4. Rede neural simples para um problema computacional complexo. Exemplo de um estimador de preço de casa.

São algoritmos genéricos, então alterado este acima, é possível que reconheçamos o número 8 à mão.

Para que possamos reconhecer determinado item, temos que ter uma prévia de dados para ensinar o algoritmo. Quanto mais dados sobre determinado item, como é mostrado na Figura 5, que no caso é o 8 à mão, melhor será sua acurácia.

Diferente do exemplo do preço que possui somente três entradas, agora temos inúmeras possíveis entradas/escritas do valor 8. As imagens da escrita são digitalizadas e então viram uma entrada para o nosso algoritmo. Usando um padrão de entrada sendo uma imagem 18x18, temos um total de 324 pixels e dessa forma, nosso algoritmo passa de 3 para 324 entradas. Ao contrário da primeira, agora tem-se duas saídas onde cada uma representa a validação do processamento realizado. Um visão geral é exibida na Figura 6.

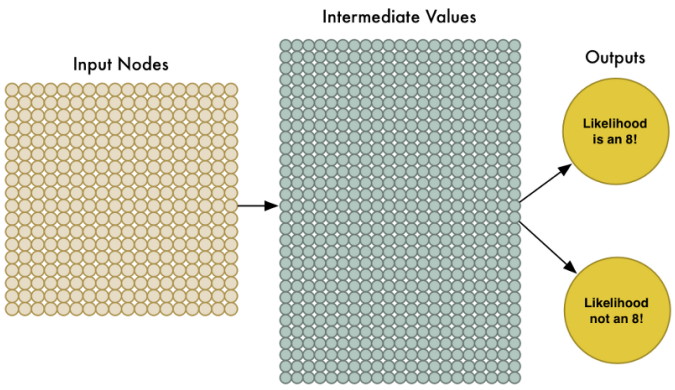

Figura 6. Visão geral do sistema descrito.

Por mais que agora tenhamos centenas de neurônios em nossa rede, computadores modernos ainda conseguem trabalhar e com bom desempenho. Isso é importante pois significa que pode-se executar tais procedimentos em sistemas embutidos onde têm como requisito a economia de energia e processamento mais brandos.

Após estes passos, deve-se treinar o algoritmo tanto para imagens que representam o valor 8 tanto para as que não representam. Com o algoritmo treinado, temos então uma rede neural capaz de reconhecer valores de 8 escritos à mão com alto valor de acurácia.

1) Enquadramento da Imagem: Antes do algoritmo de executar, é necessário enquadrar o que se quer pesquisar para enviar-se como entrada ao algoritmo. Para isso existe algumas técnicas:

a) Busca utilizando Janela Deslizante: Neste algoritmo, existe uma máscara que percorre toda a imagem. A cada iteração, a máscara seleciona parte da imagem e realiza a pesquisa na rede neural. Realizada a operação em determinado ponto, a máscara realiza o deslizamento para um outro ponto próximo predefinido para a realização de uma nova pesquisa;

b) Utilização de dados mais variantes: Sabendo-se que a imagem a ser testada pode estar com com uma posição diferente, tamanho ou até mesmo rotação diferente da treinada no algoritmo, permite-se então a criação de outra técnica que procura mais dados com posições variantes para que a rede também possa captar tais tipos de entradas em momentos futuros.

Entretanto, esta técnica faz com que o algoritmo tenha mais camadas e consequentemente mais complexo[1].

Ambos, *a)* e *b)* possuem um procedimento custoso, pois utiliza-se técnica de força bruta para a sua pesquisa.

c) Convolução: Ao invés de usar uma imagem inteira como alimento para a rede neural, utiliza técnicas mais elaboradas para extrair uma vantagem maior.

Primeiro quebra-se a imagem em vários quadros[2] sendo o nome deste é *Convolução*. Feito isso, utiliza-se cada quadro em uma pequena rede neural. Os resultados serão salvos numa matriz que segue a indexação da imagem original. A matriz de resultados final será avaliada com algoritmo de `max_pooling` onde selecionará o melhor resultado em determinado quadrante resultando em uma matriz só de soluções válidas e este processo chama-se *Redução de Amostragem*.

A função `max_pooling` é uma forma de sub-amostragem não-linear. Ela particiona a imagem de entrada em um conjunto não sobreposto de retângulos e para cada tipo de sub-região, retorna o máximo valor encontrado. Seu propósito é a eliminação de valores não maximais reduzindo o processamento computacional nas camadas posteriores. Ela também provê uma forma de transação invariante no processo realizado. Isso provê uma robustez adicional na posição, ela também é uma forma inteligente de reduzir a dimensionalidade de representações intermediárias do processo.

Com esse processo, elimina-se os itens irrelevantes para o processamento. O processo é demonstrado na Figura 7.

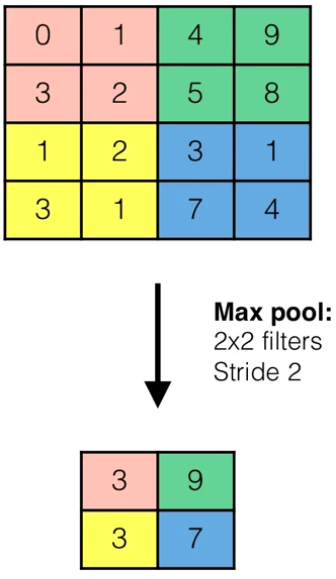

Figura 7. O processo de seleção de melhores resultados. Procedimento executado pela função `max_pooling`.

[1]Técnica ideal para utilização de GPUs.

[2]Semelhante a ideia de Janelas Deslizantes.

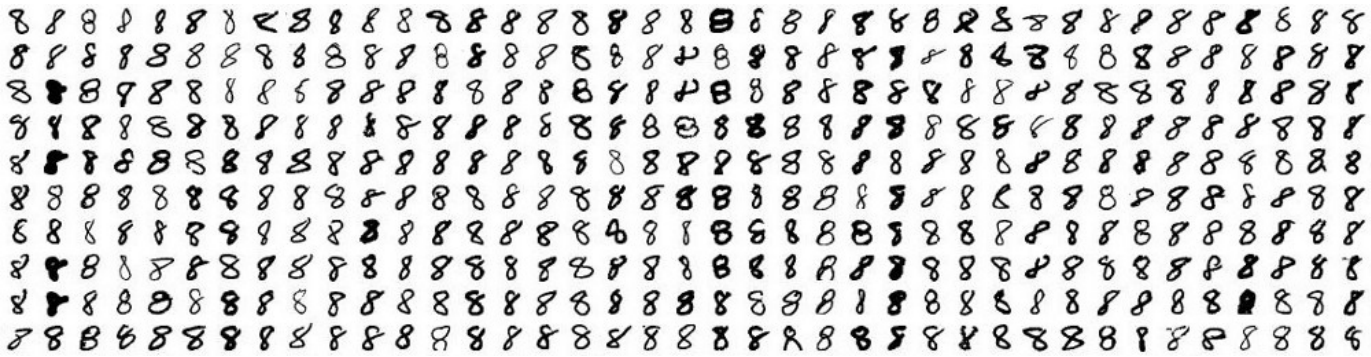

Figura 5. Exemplo de treinamento do algoritmo utilizando vários formatos de 8 escritos à mão livre.

Os itens resultantes serão entrada para outra rede neural que fará de fato a decisão final de casamento. Tal teoria é exibida na Figura 8.

Com tais procedimentos, conclui-se o processo de avaliação utilizando *convolutional neural network*.

B. Data Set

Para testes do algoritmo a ser implementado, será utilizado o conjunto de dados e *benchmark* disponibilizado gratuitamente pela Universidade de Massachusets Amherst [7].

Tal *data-set* possui um conjunto de anotações para 5171 faces identificadas num conjunto de 2845 imagens retiras do conjunto nomeado *Faces in the Wild*, disponível em http://tamaraberg.com/faceDataset/index.html. O conjunto inteiro possui um total de 30281 imagens com faces coletadas da fonte *News Photographs*. Todas foram automaticamente etiquetadas por meio do sistema descritivo *Who's in the Picture*, no qual possui 80% de acurácia [8].

Os dados estão em formato compactado `jpg`. Junto com o pacote de imagens, existe um arquivo com o seguinte formato:

```
...
<image name i>
<number of faces in this image =im>
<face i1>
<face i2>
...
<face im>
...
```

onde cada face é denotada por

```
<major_axis_radius minor_axis_radius angle
 center_x center_y 1>
```

no qual tais valores representam a posição da face como é exibido na Figura 9.

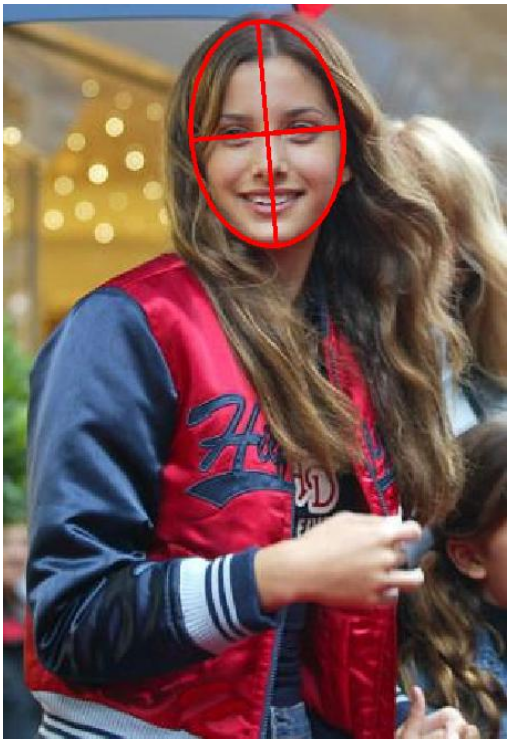

Figura 9. Exemplo de identificação de rosto especificado pelo *data-set* disponibilizado pela Universidade de Massachusets Amherst.

REFERÊNCIAS

[1] L. Haoxiang and Lin, "A Convolutional Neural Network Approach for Face Identification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5325–5334, 2015. [Online]. Available: http://www.cv-foundation.org/openaccess/content{_}cvpr{_}2015/html/Li{_}A{_}Convolutional{_}Neural{_}2015{_}CVPR{_}paper.htmlhttp://www.cs.nyu.edu/{~}xd283/face2012.pdf

[2] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[3] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, also published as a book. Now Publishers, 2009.

[4] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 9, pp. 249–256, 2010. [Online]. Available: http://www.jmlr.org/proceedings/papers/v9/

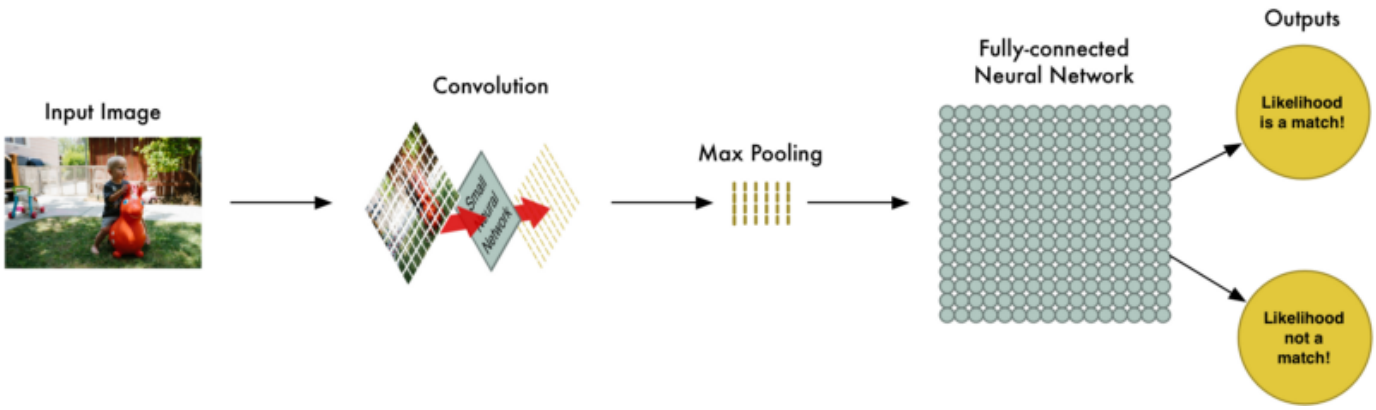

Figura 8. O processo simples de uma arquitetura de reconhecimento de padrão *convolutional neural network* final após todos os processos.

glorot10a/glorot10a.pdf?hc{_}location=ufihttp://machinelearning.wustl.edu/mlpapers/paper{_}files/AISTATS2010{_}GlorotB10.pdf

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[6] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *2013 IEEE International Conference on Image Processing, ICIP 2013 - Proceedings*, 2013, pp. 4034–4038. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/6738831/

[7] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2010-009, 2010.

[8] T. L. Berg, A. C. Berg, J. Edwards, and D. A. Forsyth, "Who's in the picture," *Advances in neural information processing systems*, vol. 17, pp. 137–144, 2004.