



Reconhecimento de Gestos

Raphael Ramos
André Luis Souto
Rafaela Sinhoroto

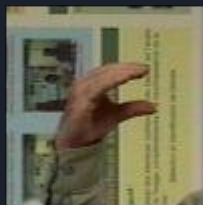
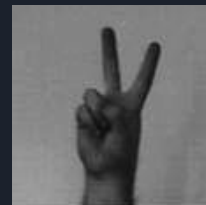
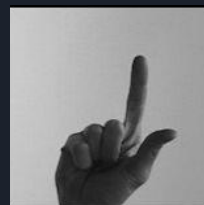
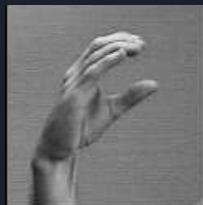
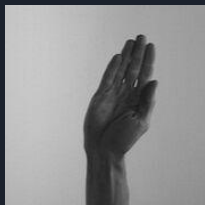


Introdução

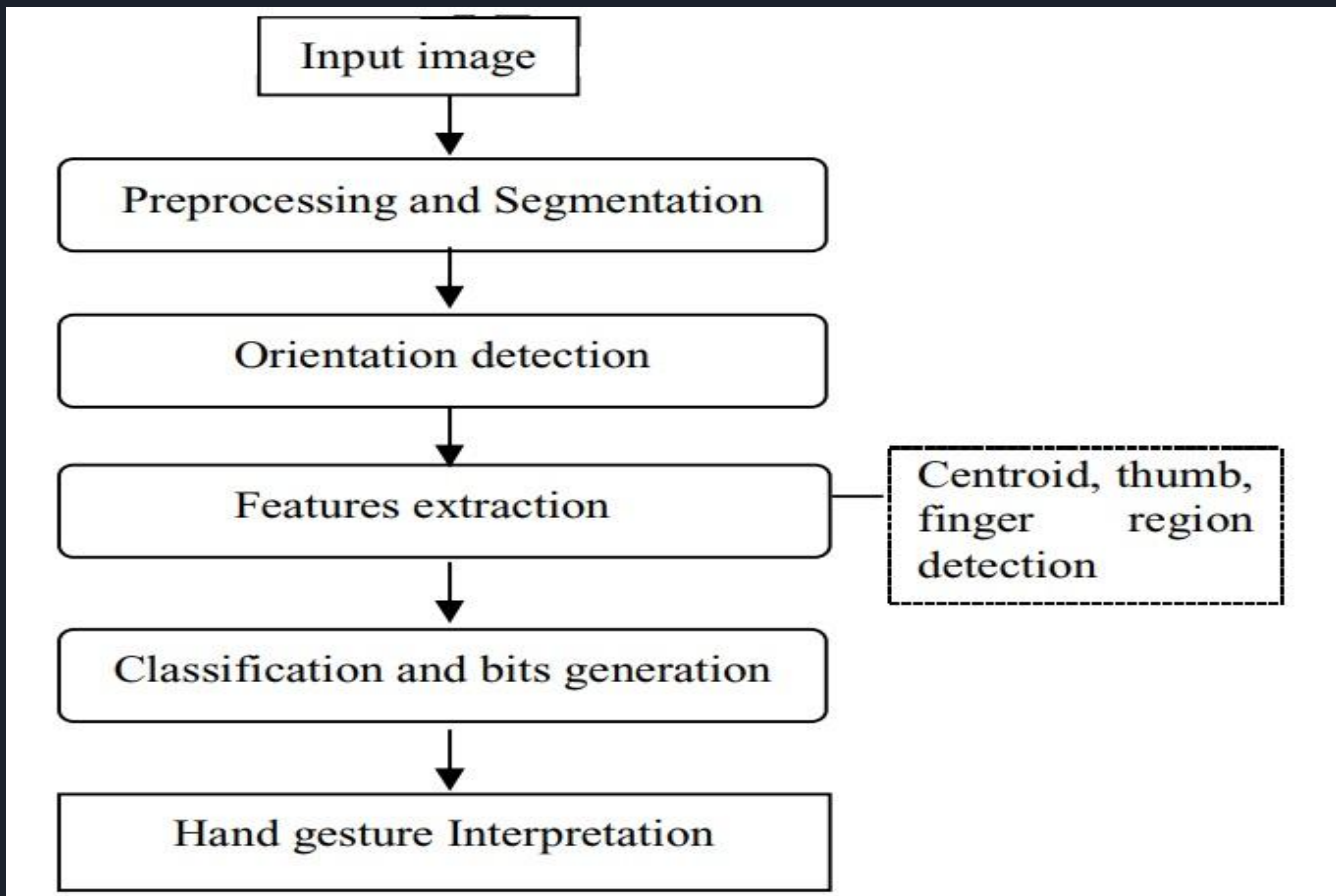
Neste trabalho foram comparados dois modelos diferentes para realizar reconhecimento de gestos: redes neurais convolucionais (CNN) e reconhecimento baseado em *shape parameters*.

Foi usado o *Marcel dataset* que consiste em 6 sinais de mão (A, B, C, FIVE, POINT, V) executados por 24 pessoas em três tipos diferentes de *background*.

Marcel Dataset

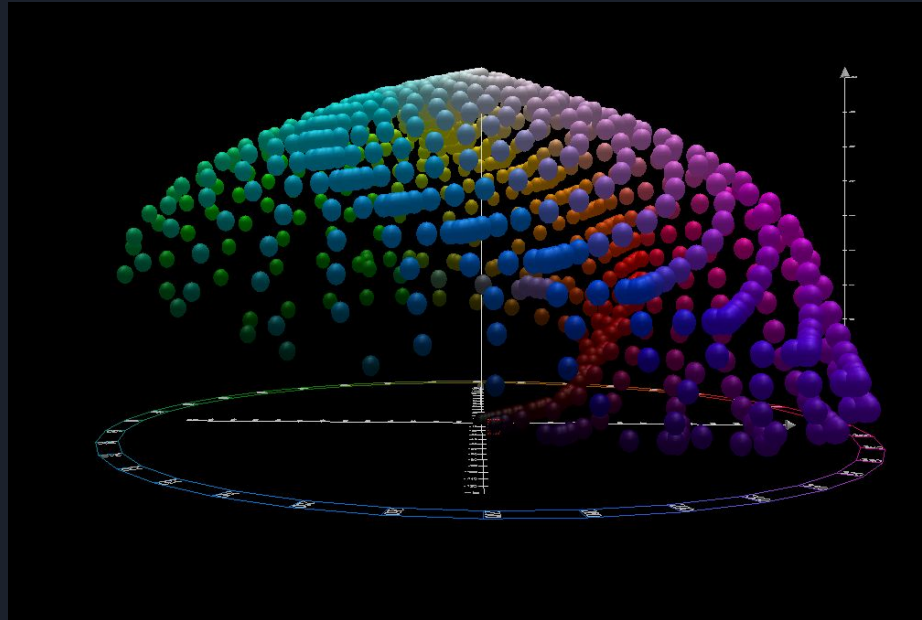
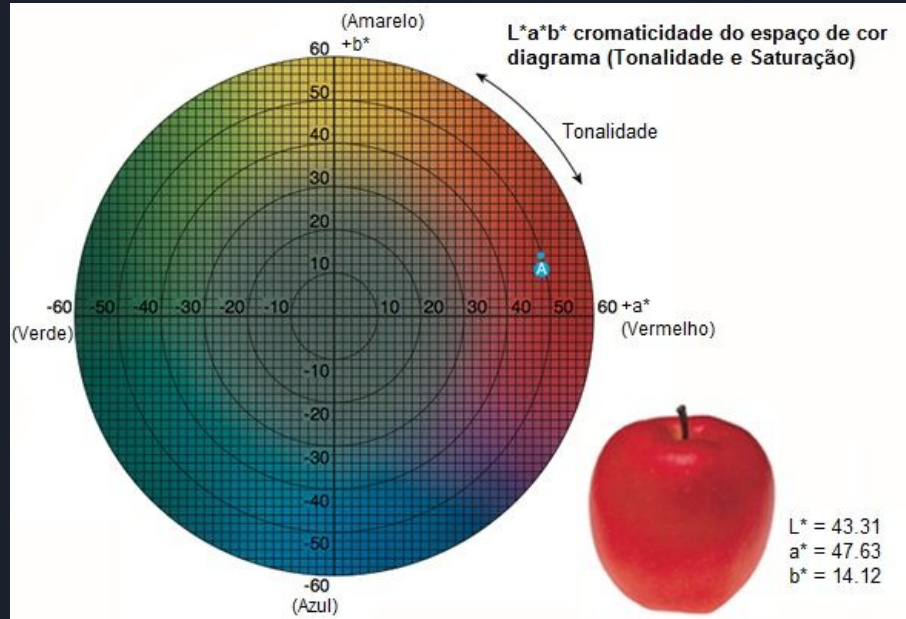


Primeiro Método - *Shape Parameters*

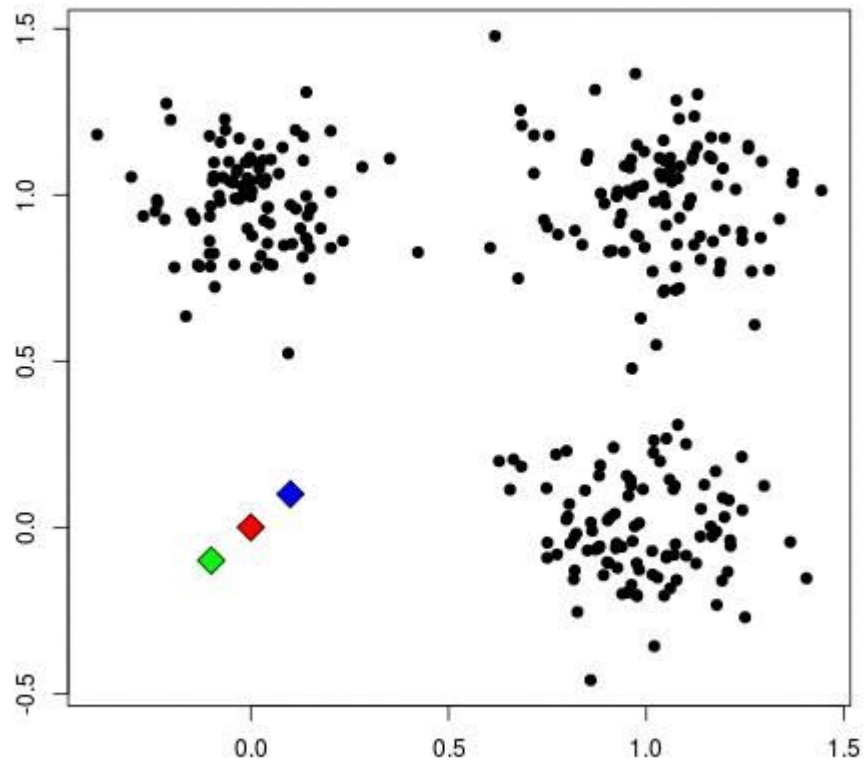


Segmentação da mão

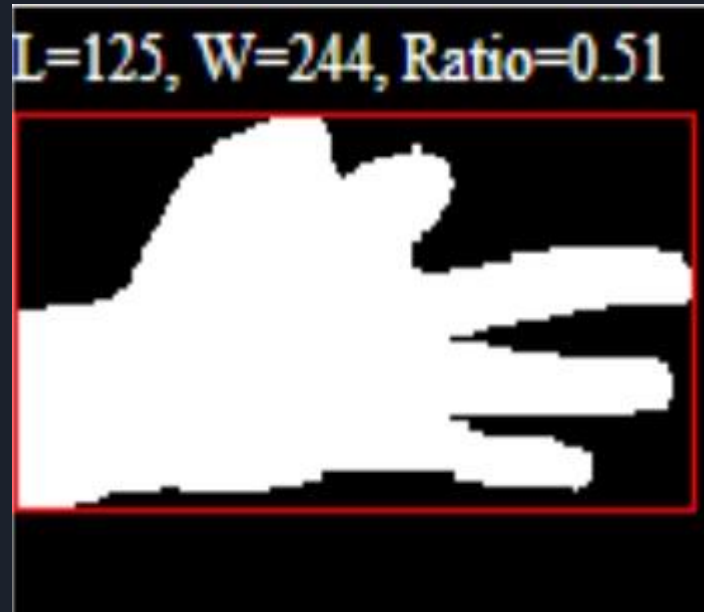
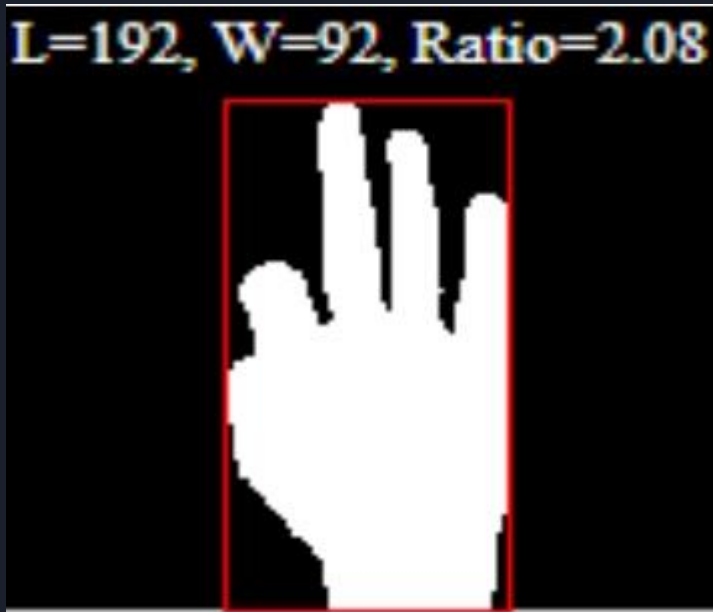
Utilizando a técnica de *k-means clustering* e agrupando os pixels de acordo com a cor no espaço $L^*a^*b^*$



Start!

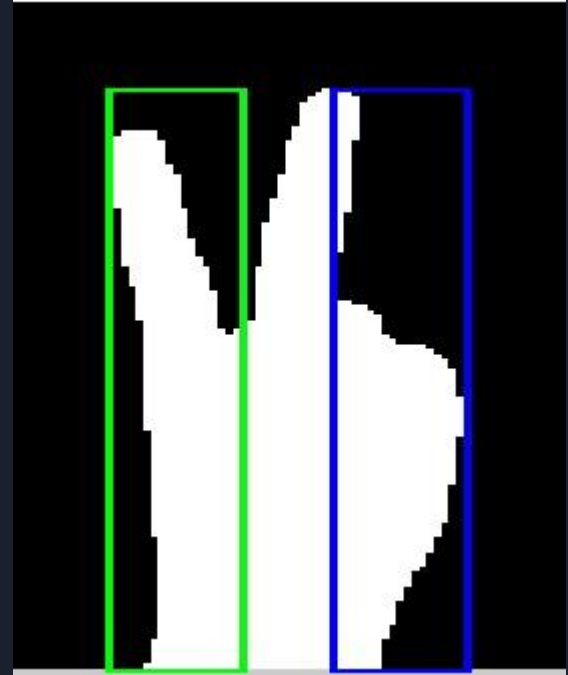


Detecção de orientação



Extração de *features* relevantes

- Centro de Massa
- Detecção de polegar
- Identificação dos dedos





Classificação

- Para classificação, é gerado um vetor de 5 bits que representa o estado dos dedos e dessa forma, representa cada um dos gestos. Cada bit corresponde a um dedo, e tem o valor 0 se dobrado ou 1 se levantado.

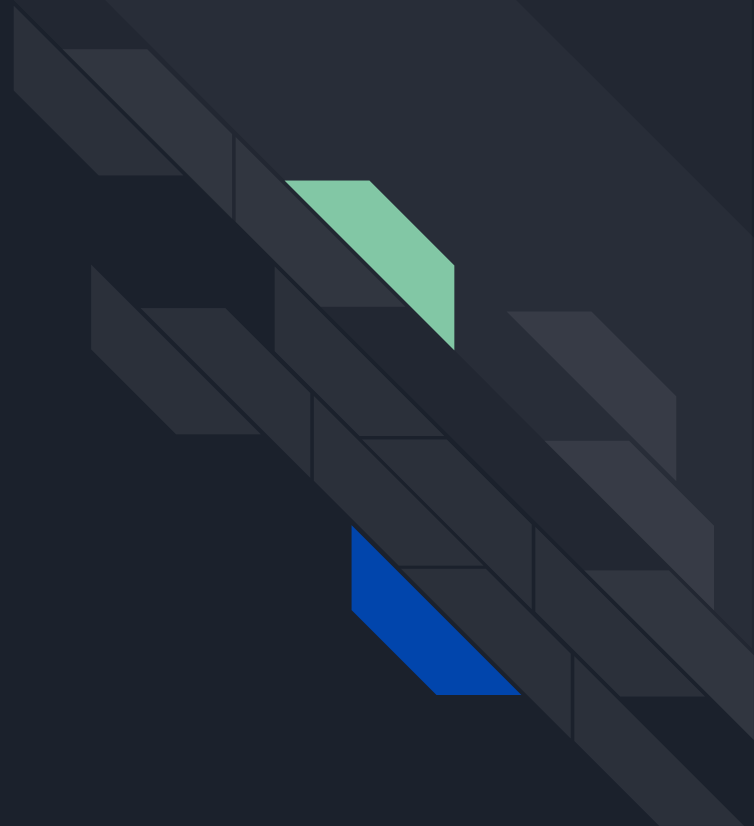
Classificação

- Exemplo: Gesto com os 5 dedos levantados é representado pelos bits [1 1 1 1 1]




~17%

É a acurácia obtida ao utilizar o método dos parâmetros de forma para classificar gestos variados em situações não triviais (ex: plano de fundo não uniforme, pessoas utilizando camisas de manga comprida, dor de fundo próxima de cor de pele)



Confusion Matrix

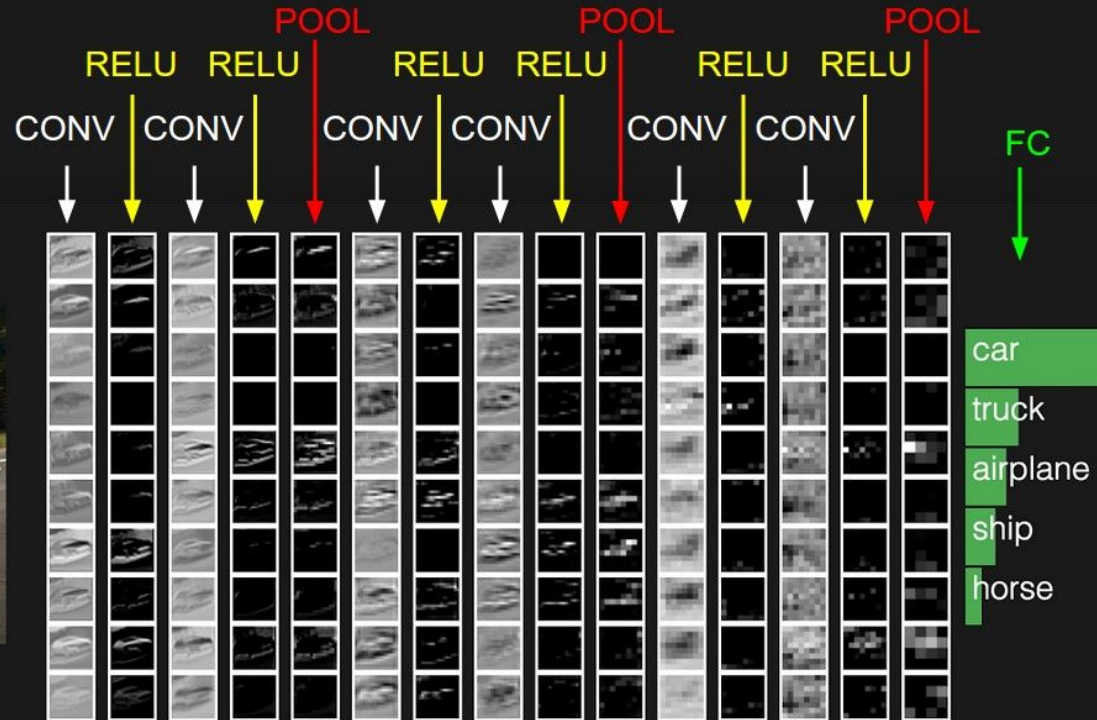
Output Class	1	2	3	4	5	6	7	
	16 2.4%	17 2.6%	6 0.9%	27 4.1%	23 3.5%	12 1.8%	0 0.0%	15.8% 84.2%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	0 0.0%	0 0.0%	1 0.2%	14 2.1%	0 0.0%	0 0.0%	0 0.0%	93.3% 6.7%
	31 4.7%	41 6.2%	9 1.4%	7 1.1%	44 6.7%	13 2.0%	0 0.0%	30.3% 69.7%
	1 0.2%	0 0.0%	11 1.7%	4 0.6%	1 0.2%	37 5.6%	0 0.0%	68.5% 31.5%
	48 7.3%	44 6.7%	85 12.9%	82 12.5%	51 7.8%	33 5.0%	0 0.0%	0.0% 100%
Target Class	1	2	3	4	5	6	7	
	16.7% 83.3%	0.0% 100%	0.0% 100%	10.4% 89.6%	37.0% 63.0%	38.8% 61.1%	NaN% NaN%	16.9% 83.1%



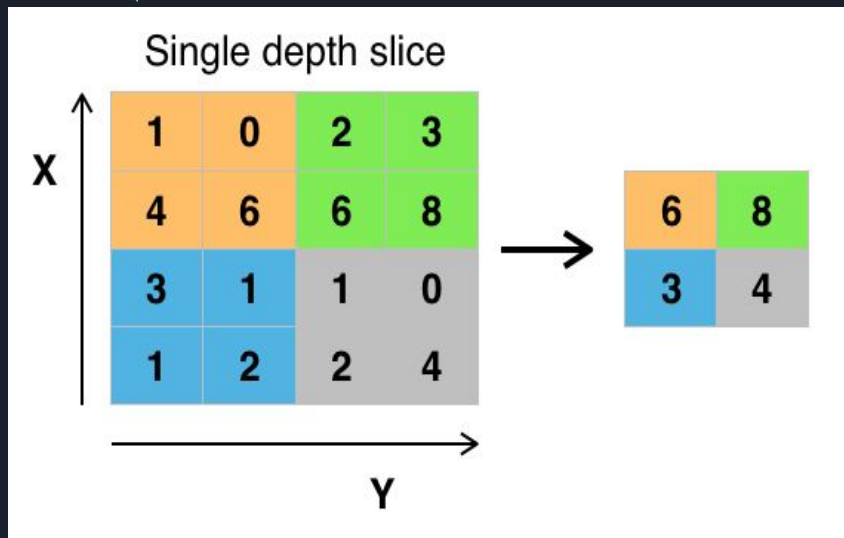
Segundo Método - Redes Neurais Convolucionais (CNN)

- Redes neurais são capazes de fazer previsões aprendendo o relacionamento entre características dos seus dados e algumas respostas observadas.
- Em uma CNN cada camada age como um filtro para a detecção de features específicas ou padrões presentes nos dados.
- As primeiras camadas em uma CNN detectam características que podem ser reconhecidas e interpretadas relativamente fácil. Camadas posteriores detectam features mais abstratas e usualmente presentes em muitas das features detectadas por camadas anteriores. A última camada realiza a classificação combinando todas as features detectadas pelas camadas anteriores no dados de input.
- Redes neurais convolucionais profundas alcançaram alta performance e são a base dos resultados do estado da arte para reconhecimento de imagens, detecção de objetos, reconhecimento de faces, reconstrução tridimensional de objetos, reconhecimento de faces, reconhecimento de discurso, entre outros.
- Desnecessidade de engenharia de características!

Primeiro Método - Redes Neurais Convolucionais (CNN)



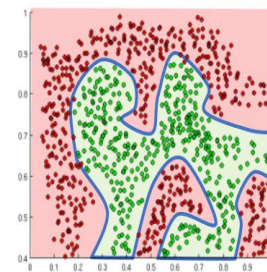
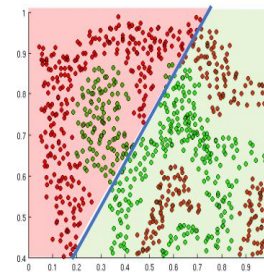
Primeiro Método - Redes Neurais Convolucionais (CNN)



Pooling Operator - Reduz quantidade de parâmetros e cálculos na rede

Importance of Activation Functions

The purpose of activation functions is to introduce non-linearities into the network



Funções de ativação lineares são geralmente usadas em output neurons quando o objetivo é regressão em vez de classificação



Redes Neurais Convolucionais (CNN) - Inception

- Importante marco no desenvolvimento de CNN.
- Antes da sua origem (hehe), a maioria das CNNs apenas empilhavam camadas de convolução cada vez mais profundas, em busca de uma melhor performance.
- Em contrapartida, as redes Inception são complexas. Usam vários truques para melhorar performance, tanto em termos de velocidade quanto acurácia.
- Hoje já existem várias versões: Inception v1, Inception v2, Inception v3, Inception v4, Inception-ResNet v1, Inception-ResNet v2 e Xception.

Redes Neurais Convolucionais (CNN)



Meme referenciado no primeiro paper da Inception: Going deeper with convolutions.

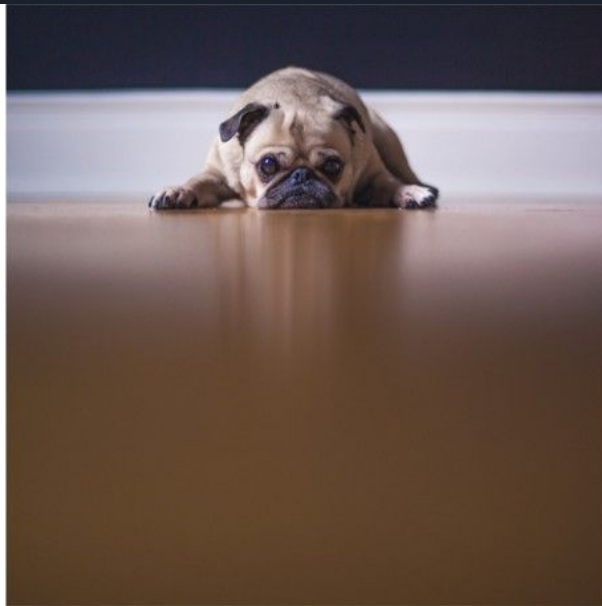


Redes Neurais Convolucionais (CNN)

Problemas:

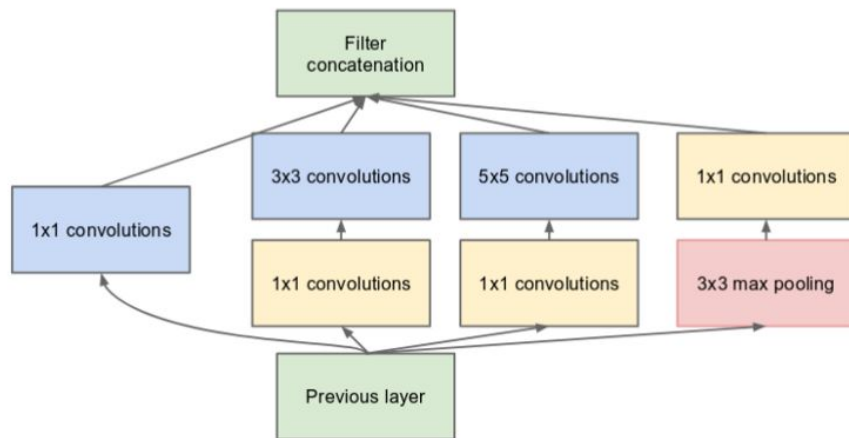
1. Partes salientes da imagem podem ter variação no tamanho.
2. Redes neurais muito profundas tendem a sobreajuste dos dados. Também é difícil passar o gradiente por toda a rede.
3. Operação de convolução é cara!

Redes Neurais Convolucionais (CNN)

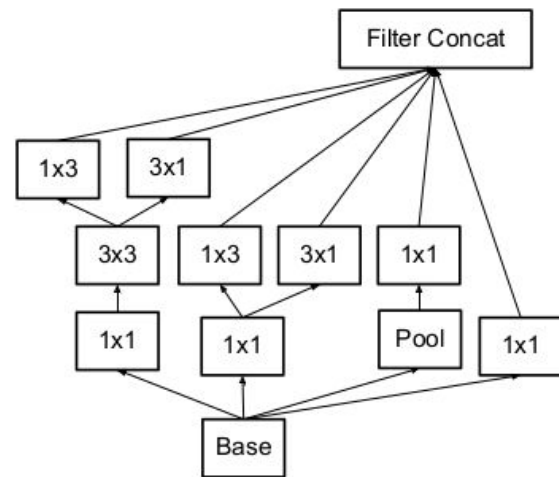


Fotos de cachorros ocupando áreas diferentes da imagem.

Redes Neurais Convolucionais (CNN) - Inception



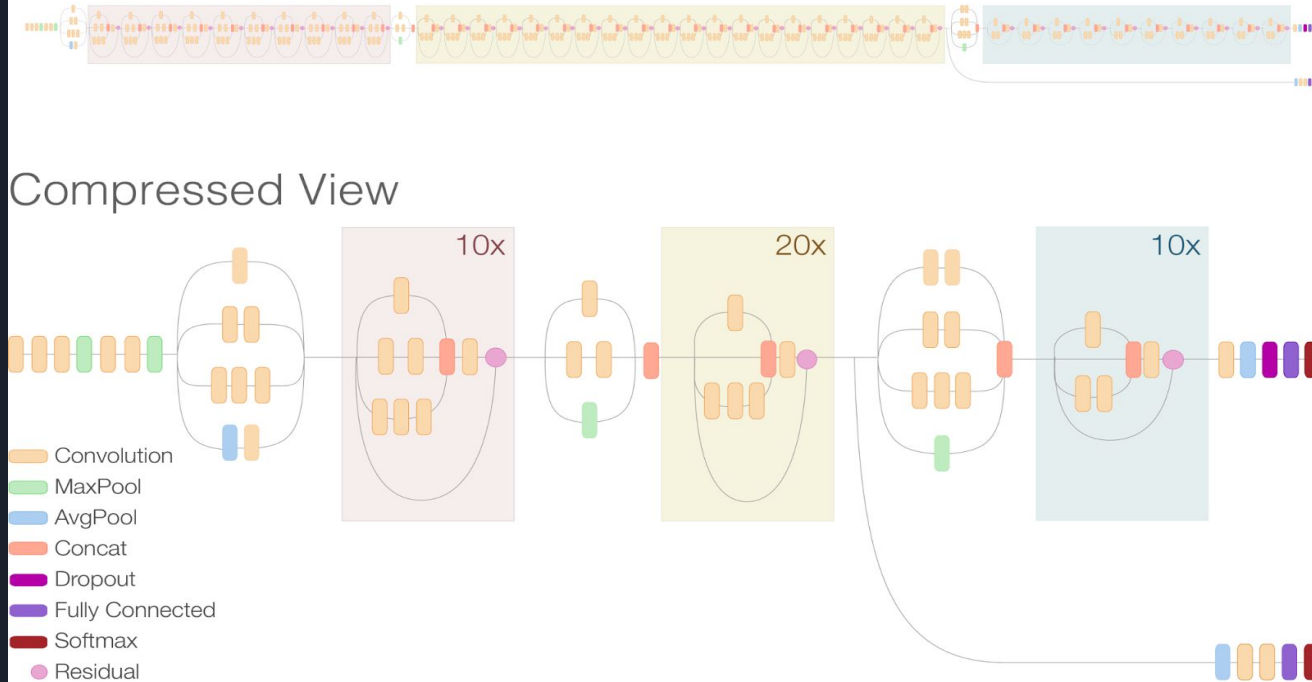
(b) Inception module with dimension reductions



Módulo Inception original à esquerda e fatorado a direita. Os autores notaram que uma convolução com kernel de tamanho $n \times n$ é 33% mais cara que uma combinação de uma convolução $1 \times n$ seguida por uma $n \times 1$.

Redes Neurais Convolucionais (CNN) - Inception ResNet V2

Inception Resnet V2 Network





Redes Neurais Convolucionais - Resultados

	Dropout	Acurácia (Top-1)
Xception-1.0	60%	89.82%
Xception-2.0	80%	91.45%
Inception ResNet V2	80%	91.45%
GoogleNet (artigo)		78.22%
VGG (artigo)		64.19%

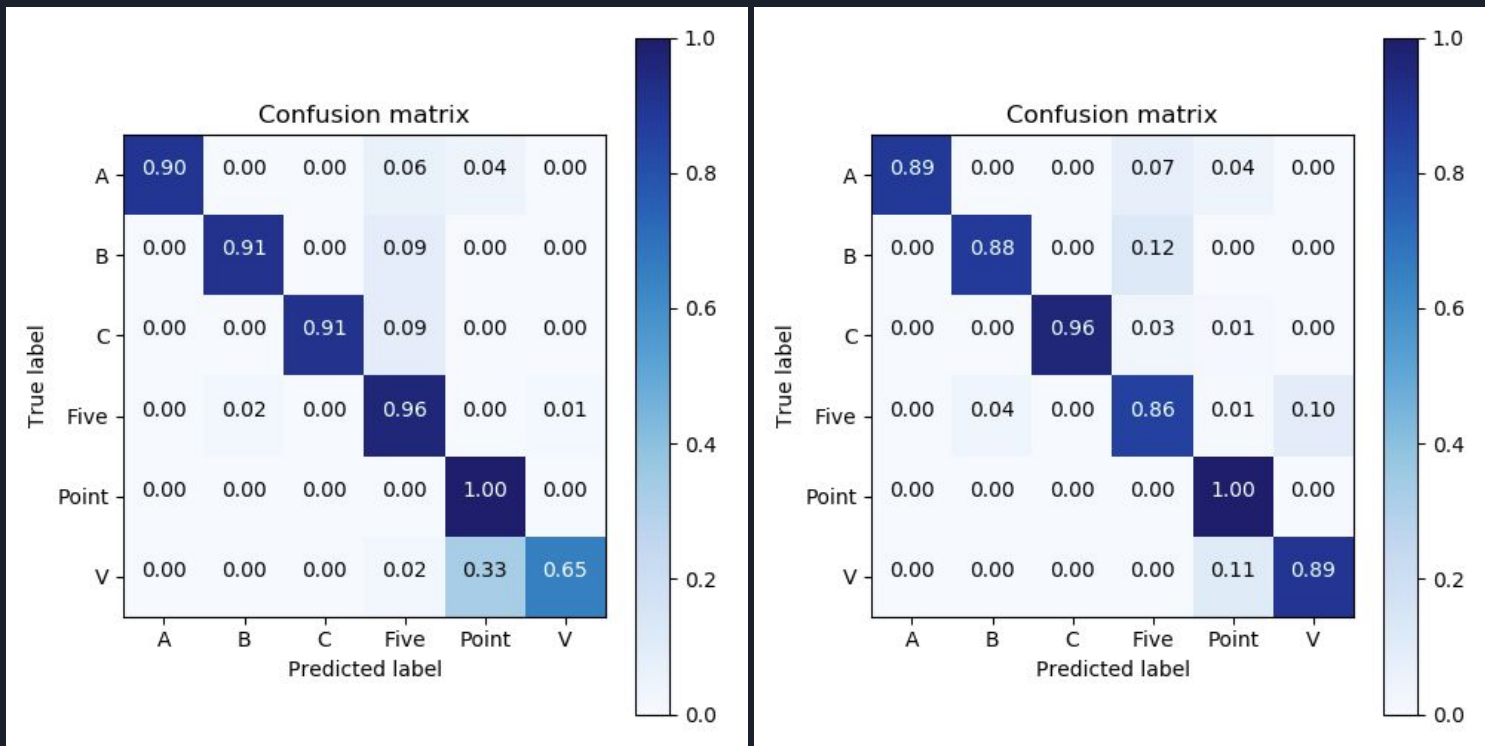
Resultados obtidos por mim e pelos autores do artigo Hand Gesture Recognition using Deep Convolutional Neural Networks.



Redes Neurais Convolucionais - Resultados

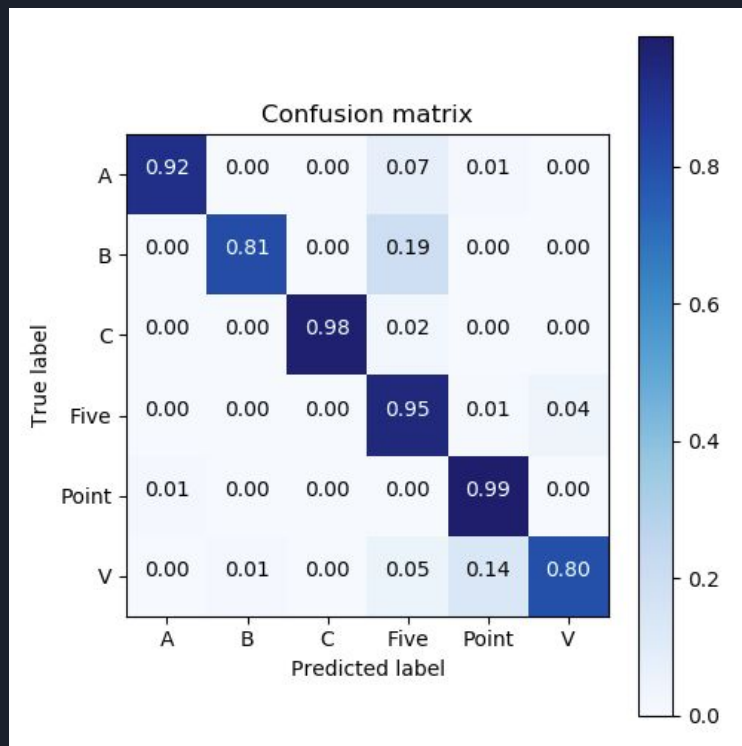
	precision	recall	f1-score	support
A	1.00	0.89	0.94	96
B	0.95	0.88	0.91	102
C	1.00	0.96	0.98	112
Five	0.84	0.86	0.85	134
Point	0.88	1.00	0.94	119
V	0.87	0.89	0.88	95
avg / total	0.92	0.91	0.92	658

Redes Neurais Convolucionais - Resultados



Matriz de confusão Xception-1.0 e Xception-2.0, respectivamente.

Redes Neurais Convolucionais - Resultados



Matriz de confusão Inception ResNet V2.

Redes Neurais Convolucionais - Demonstração

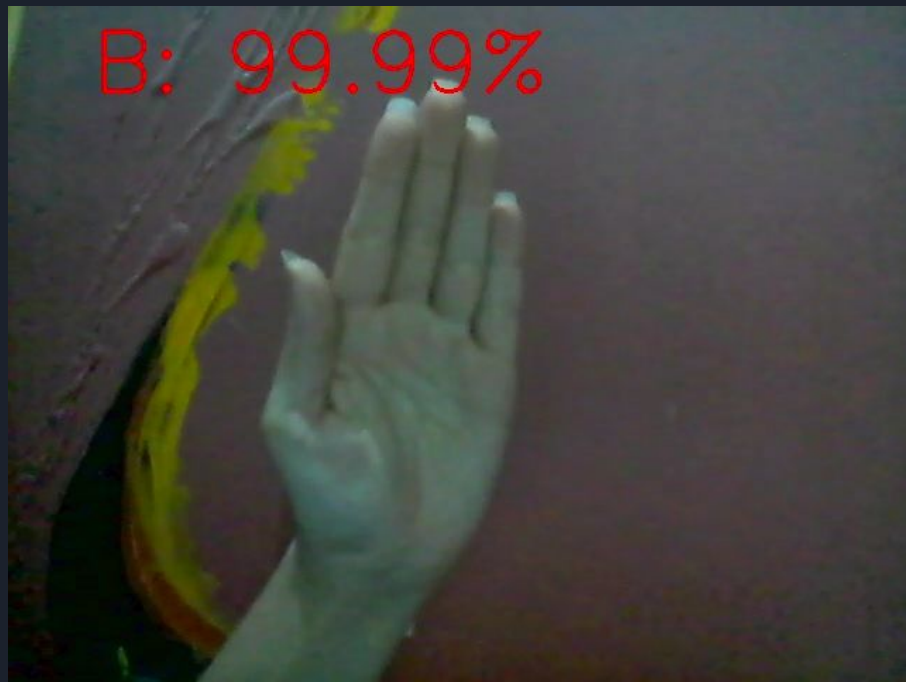
A: 99.68%



A: 95.13%



Redes Neurais Convolucionais - Demonstração



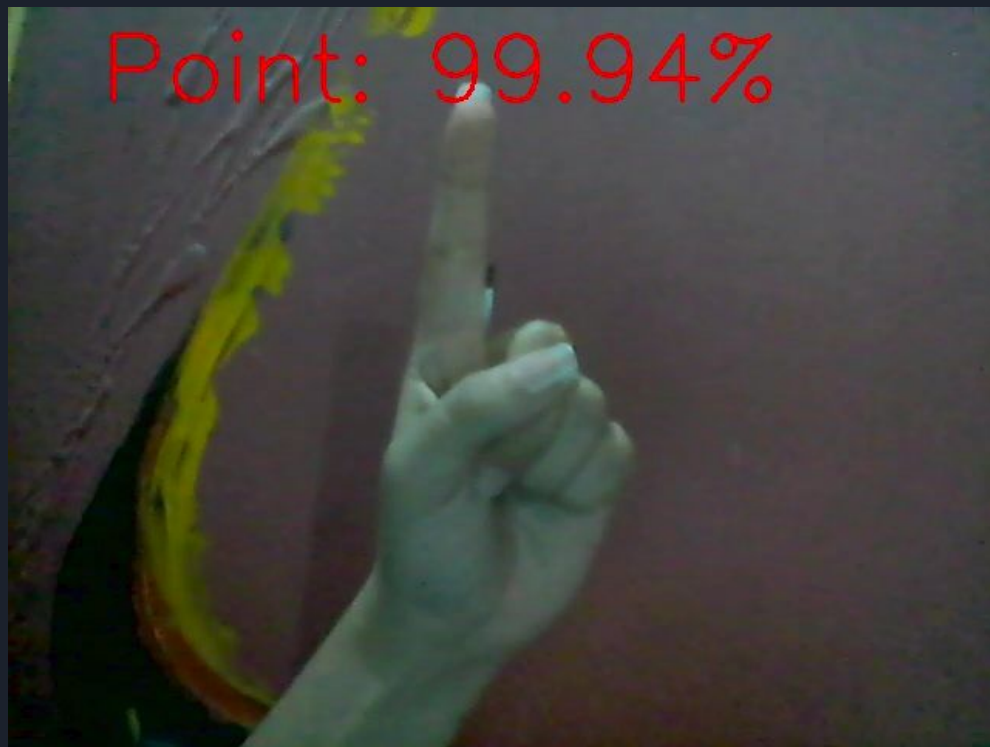
Redes Neurais Convolucionais - Demonstração



Redes Neurais Convolucionais - Demonstração



Redes Neurais Convolucionais - Demonstração

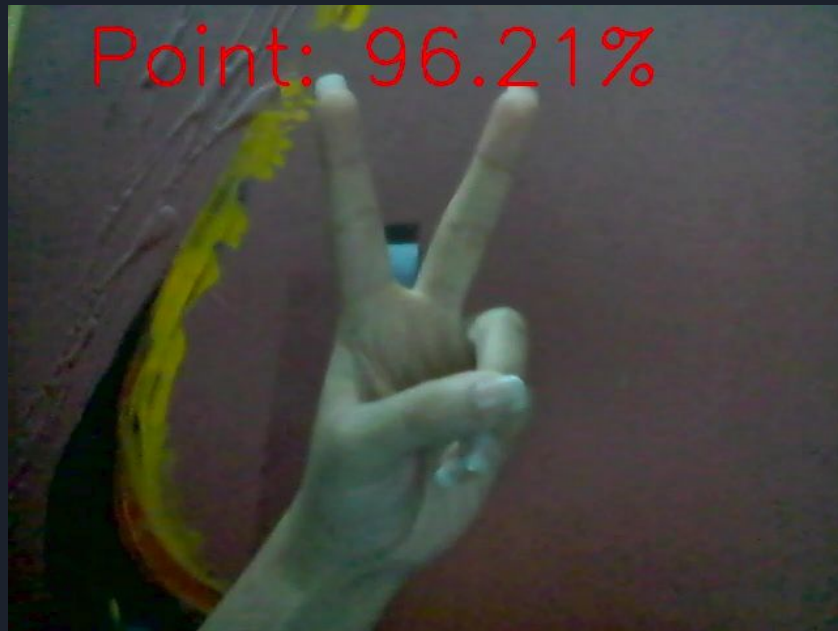


Redes Neurais Convolucionais - Demonstração

V: 85.35%



Point: 96.21%





Conclusão

- De acordo com os resultados obtidos, conclui-se que o método Shape Parameters não é adequado para reconhecimento de gestos quando as imagens não são feitas em ambiente controlado, ou seja, quando o fundo não é uniforme ou a posição da mão varia com relação a câmera.
- Já o método de Redes Neurais Convolucionais, mostrou resultados satisfatórios.