

# Detecção e Segmentação

## Redes Neurais e Aprendizado Profundo

Moacir A. Ponti

[www.icmc.usp.br/~moacir](http://www.icmc.usp.br/~moacir) — [moacir@icmc.usp.br](mailto:moacir@icmc.usp.br)

São Carlos-SP/Brasil

# Agenda

Detecção de Objetos, regressão+classificação e pixel-to-pixel

# Agenda

Detecção de Objetos, regressão+classificação e pixel-to-pixel

# Classificação + regressão

Objetivo: classificar e localizar



## Saída da rede

- ▶ Classes
- ▶ Valores de uma caixa (*bounding box*)

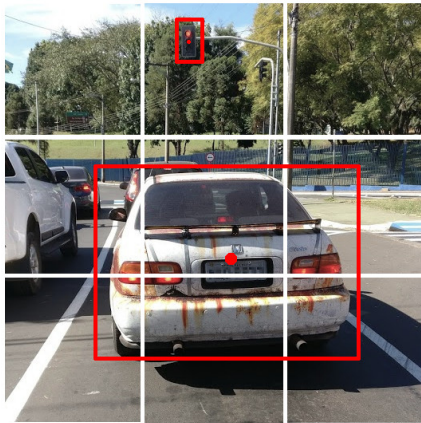
# Classificação + regressão

Formato da predição (saída da rede): presença do objeto, bounding box e classes.

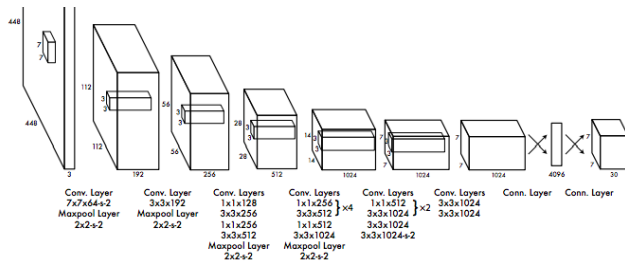


# Classificação + regressão: em um grid

Treinamento considera grid  $S \times S$  (comumente  $19 \times 19$ ) e  $B$  caixas em formatos pré-definidos, chamados de âncoras.



# YOLO: You Only Look Once



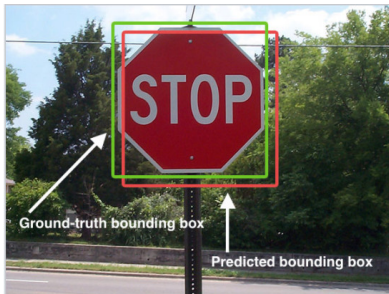
**Figure 3: The Architecture.** Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating  $1 \times 1$  convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution ( $224 \times 224$  input image) and then double the resolution for detection.


Confiança é calculada com:  $P(classe) \cdot IoU$

Saída é de tamanho  $S \times S \times (5B + C)$

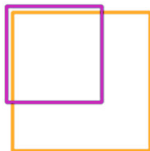
# YOLO: You Only Look Once + IoU

## Intersecção sobre União



$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


IoU: 0.4034



IoU: 0.7330



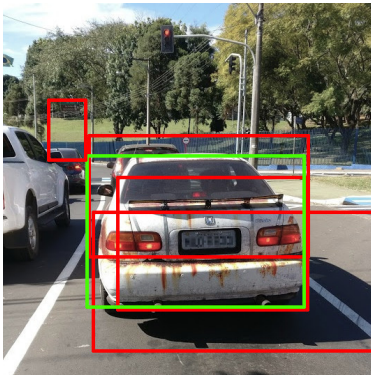
IoU: 0.9264





# YOLO: You Only Look Once + Non-Max Supression

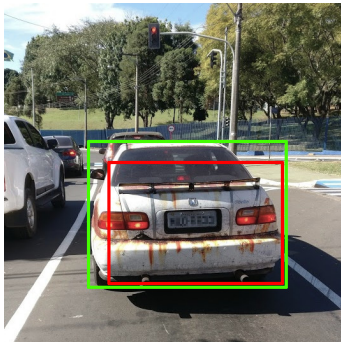
Supressão de não máximos



- ▶ descartar  $p_c \leq 0.6$
- ▶ seleccionar maior  $p_c$
- ▶ descartar caixas com  $IoU \geq 0.5$  da anterior

# YOLO: You Only Look Once + Non-Max Supression

Supressão de não máximos



- ▶ descartar  $p_c \leq 0.6$
- ▶ seleccionar maior  $p_c$
- ▶ descartar caixas com  $IoU \geq 0.5$  da anterior

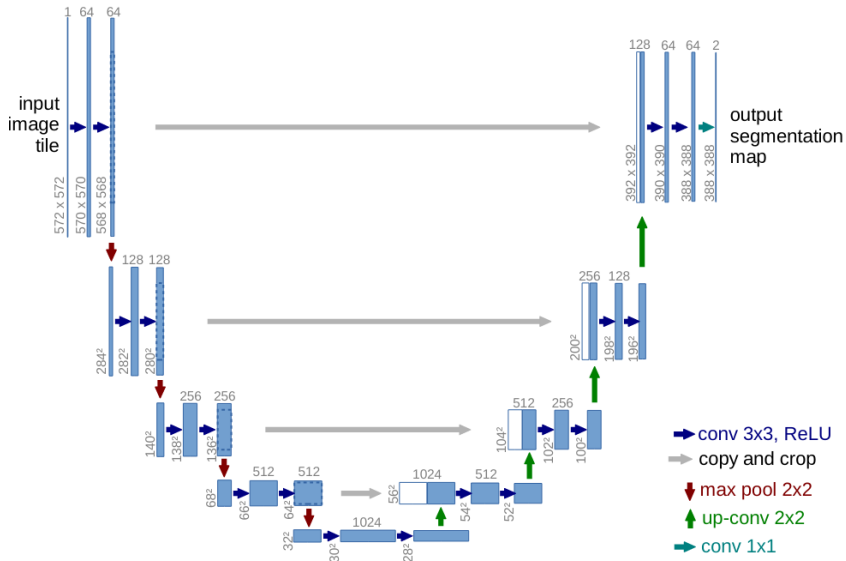
# Detecção de pontos de referência (landmark)

Exemplo: encontrar pontos de uma face

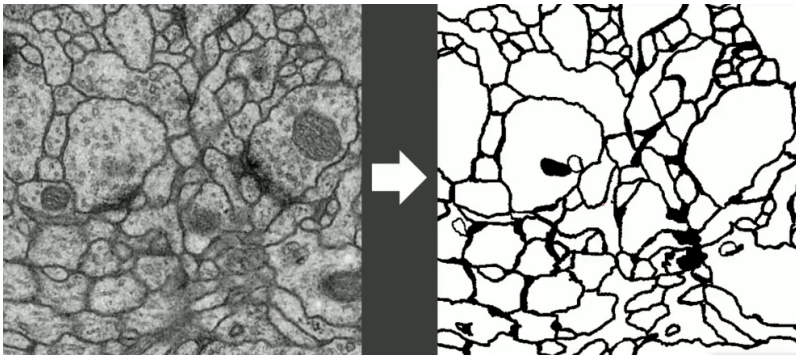
Formato da predição (saída da rede): presença do objeto de interesse, coordenadas para cada landmark



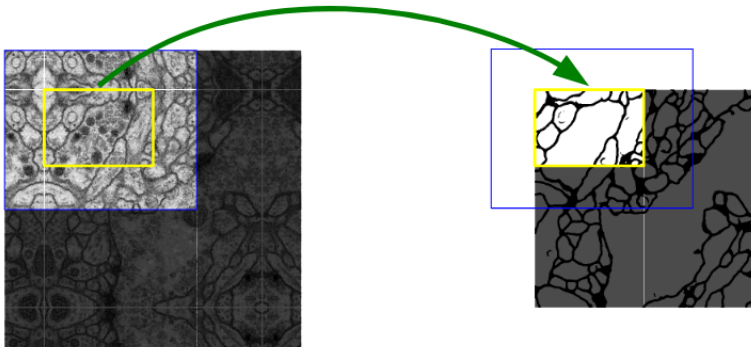
# U-Net



# U-Net



# U-Net



## Função de custo

- ▶ Softmax pixel-a-pixel (ao longo dos canais) + Entropia Cruzada
- ▶ Pesos computados para:
  1. compensar desbalanceamento (fundo é comumente mais proeminente do que os alvos)
  2. dar mais peso às bordas das regiões a serem segmentadas

## Treinamento:

- ▶ Data augmentation: utilizando deformação suave das imagens
- ▶ Inicialização dos pesos por camada
  1. baseada na camada anterior
  2. distribuição Gaussiana/normal  $\sigma = \sqrt{2/N}$ ,  $N$  sendo o número de nós de entrada
    - exemplo: camada anterior com 64 filtros  $3 \times 3$ ,  $\sigma = \sqrt{2/(9 \cdot 64)} = \sqrt{2/576}$



## Referências: Classificação e regressão para detecção de objetos

- ▶ Material em Português sobre YOLO:  
<https://iaexpert.academy/2020/10/13/deteccao-de-objetos-com-yolo-uma-abordagem-moderna/>
- ▶ Blog com implementação básica em Keras: <https://machinelearningmastery.com/how-to-perform-object-detection-with-yolov3-in-keras/>
- ▶ Artigo sobre detecção de objetos: [https://www.researchgate.net/profile/Li-Liu-76/publication/327550187\\_Deep\\_Learning\\_for\\_Generic\\_Object\\_Detection\\_A\\_Survey/links/5ddf876aa6fdcc2837f083ea/Deep-Learning-for-Generic-Object-Detection-A-Survey.pdf](https://www.researchgate.net/profile/Li-Liu-76/publication/327550187_Deep_Learning_for_Generic_Object_Detection_A_Survey/links/5ddf876aa6fdcc2837f083ea/Deep-Learning-for-Generic-Object-Detection-A-Survey.pdf)