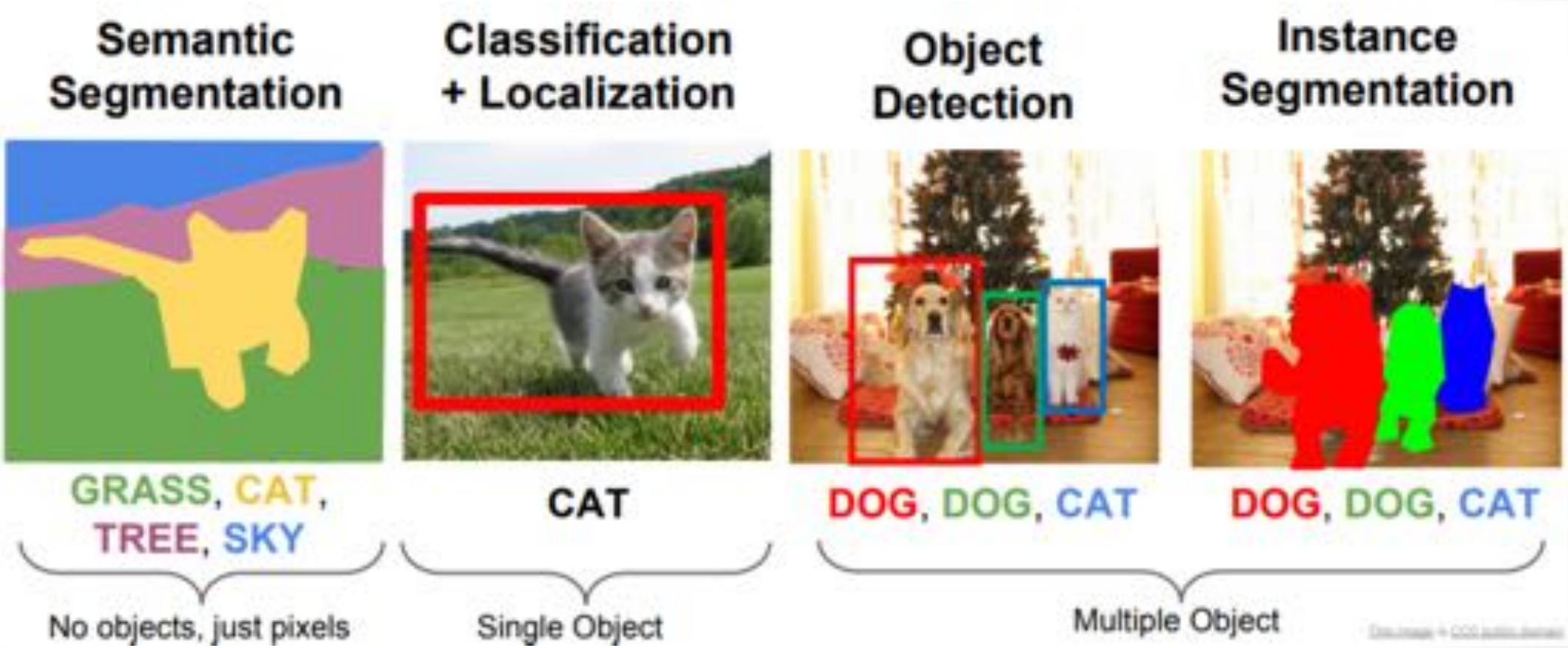
Segmentação Semântica: U-Net

Visão Computacional

Programa de Pós-Graduação em Ciência da Computação – UFMA
Prof. Geraldo Braz Junior

Notas de aula baseada em cs231n

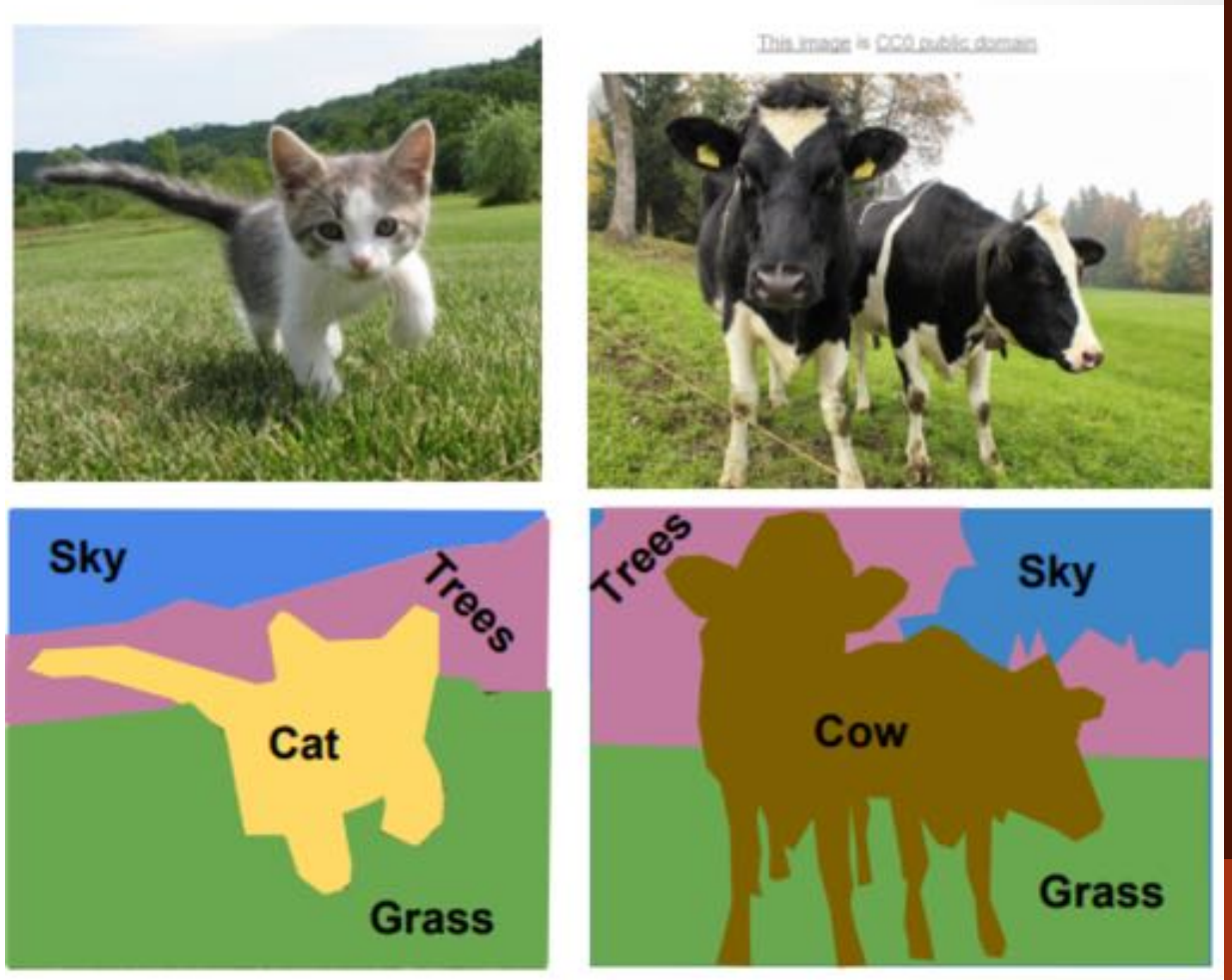
Tipos


Semantic Segmentation
Classification + Localization
Object Detection
Instance Segmentation
GRASS, CAT, TREE, SKY
CAT
DOG, DOG, CAT
DOG, DOG, CAT
No objects, just pixels
Single Object
Multiple Object

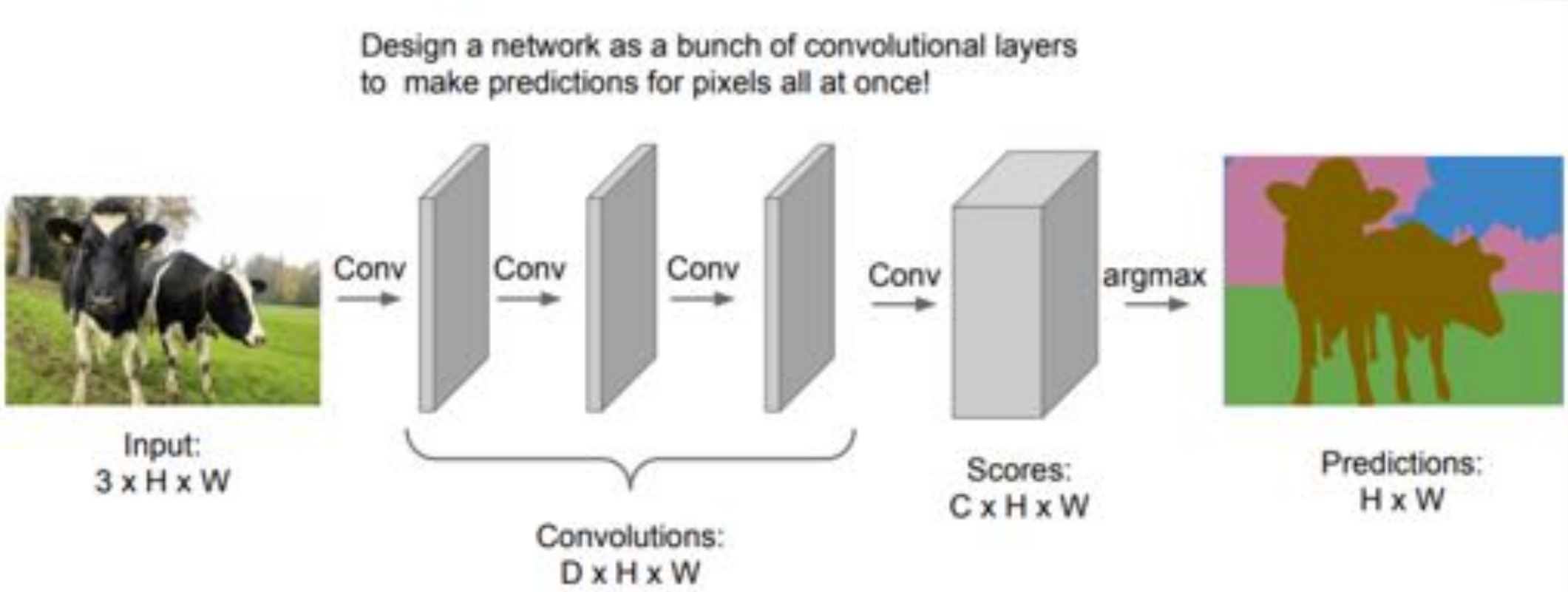
Segmentação semântica

- Classifica cada pixel
- Não diferencia objetos, apenas pixels

- Algumas soluções
  - Janela deslizante: muito custoso
  - Uma MLP ou CNN geral com vetores de características final: tb custoso

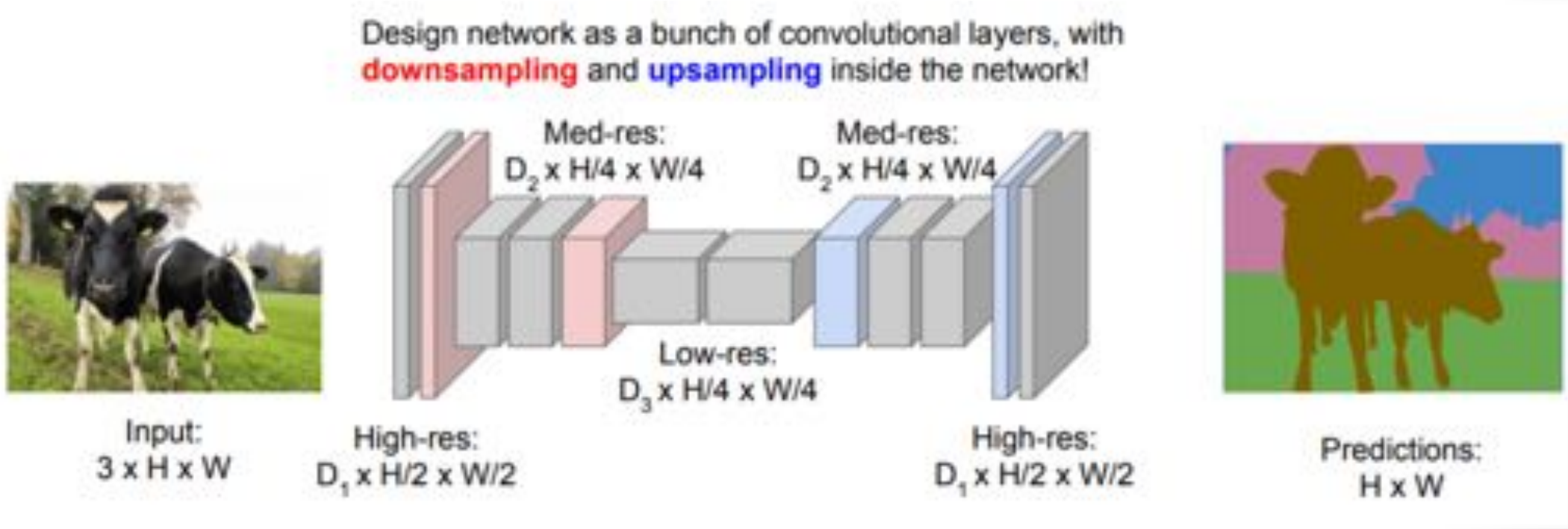


Segmentação semântica com CNN direto



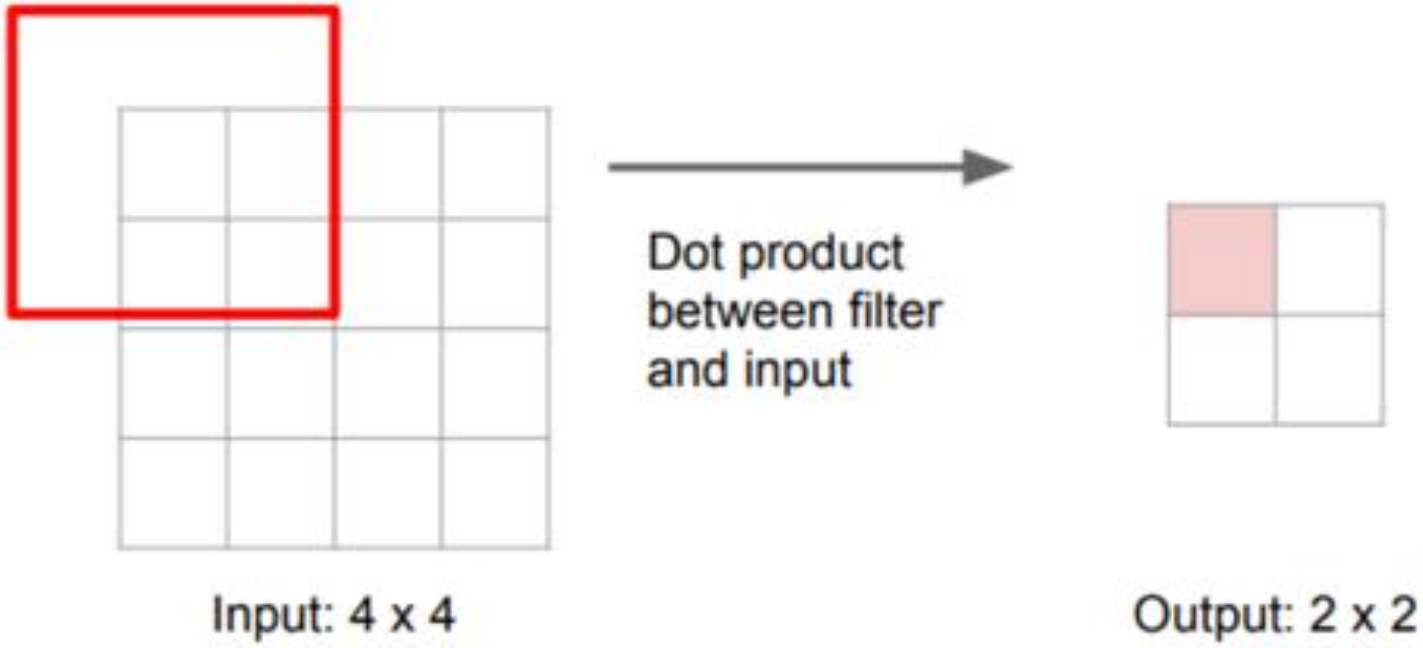
Como seria rotular essa base?
O quanto custa manter a proporção?

Abordagens mais recentes



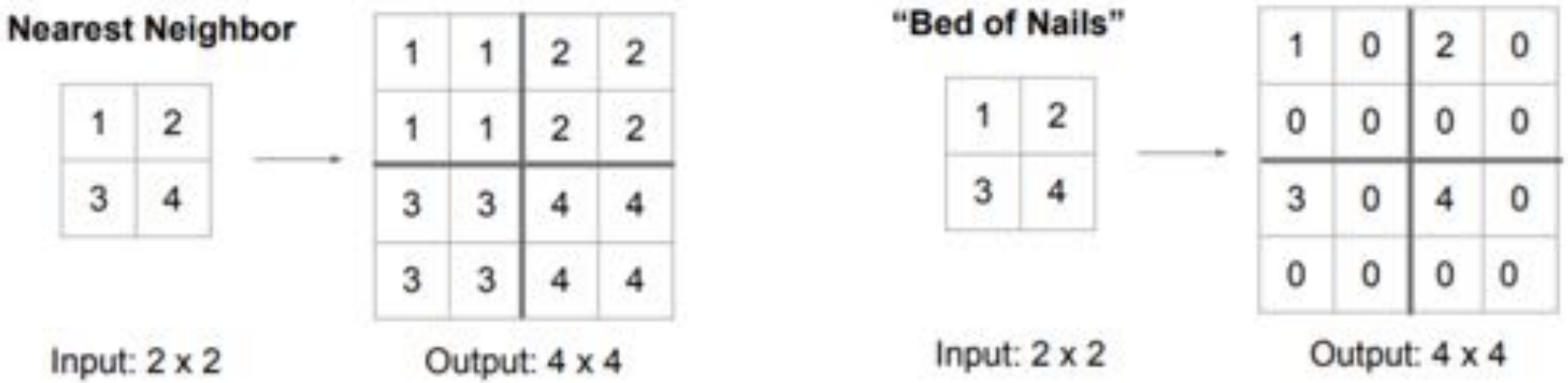
Downsampling

- Basicamente pooling ou convolução com stride > 1

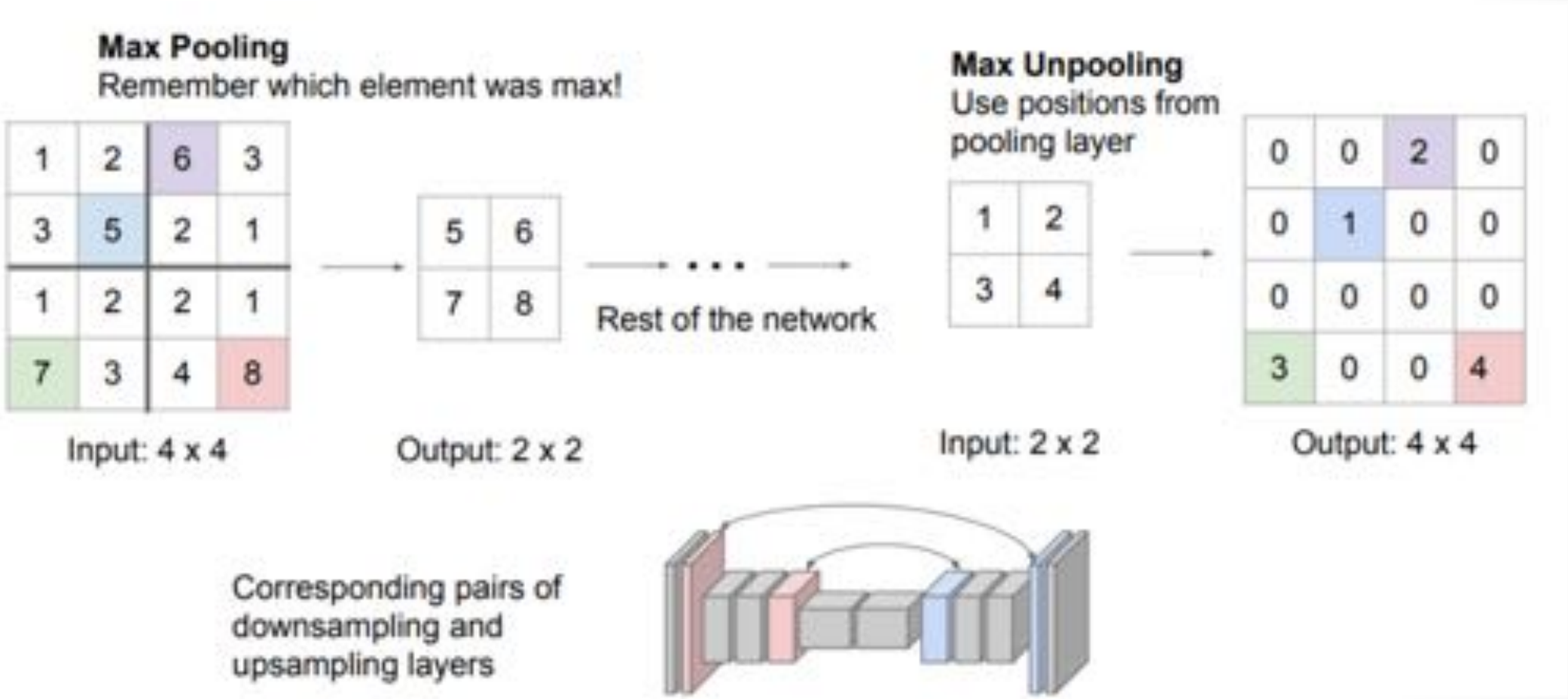


- O filtro será aprendido com backpropagation

Upsampling: abordagens

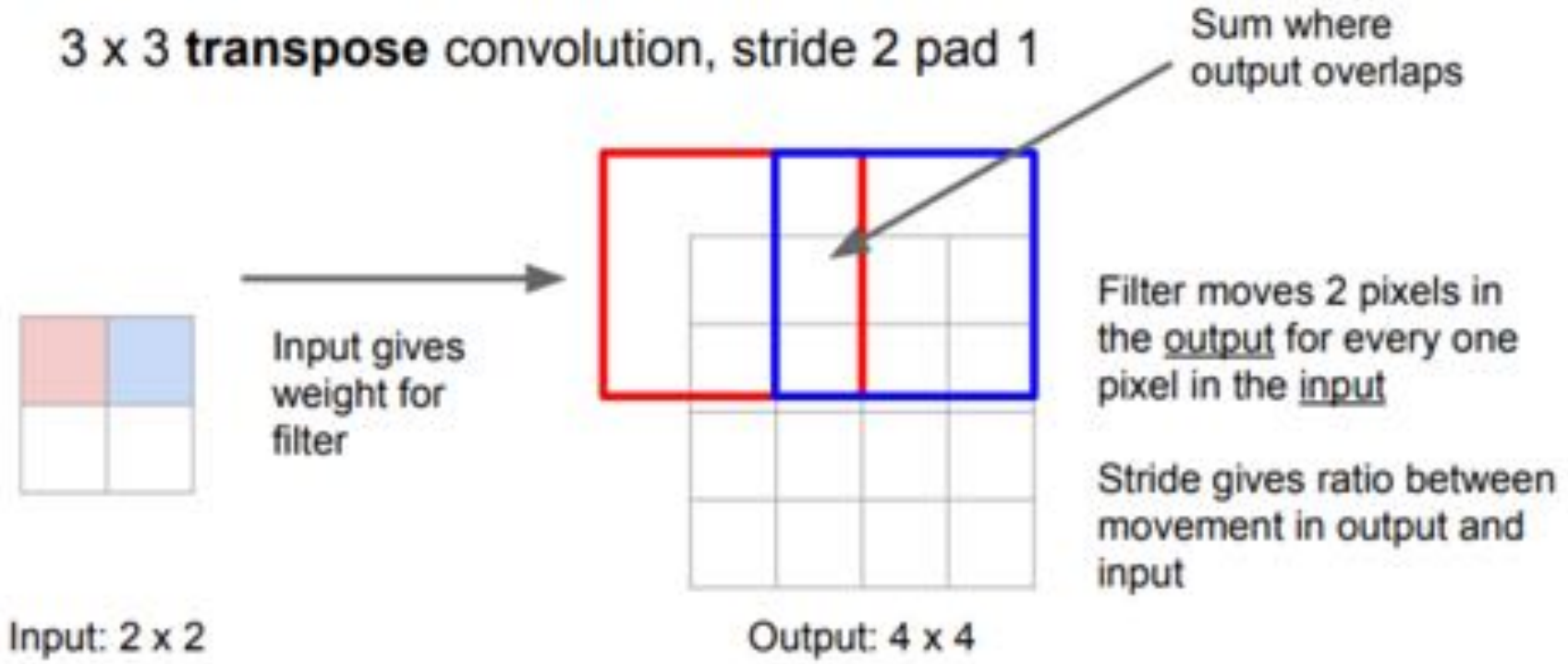


Upsampling: abordagens

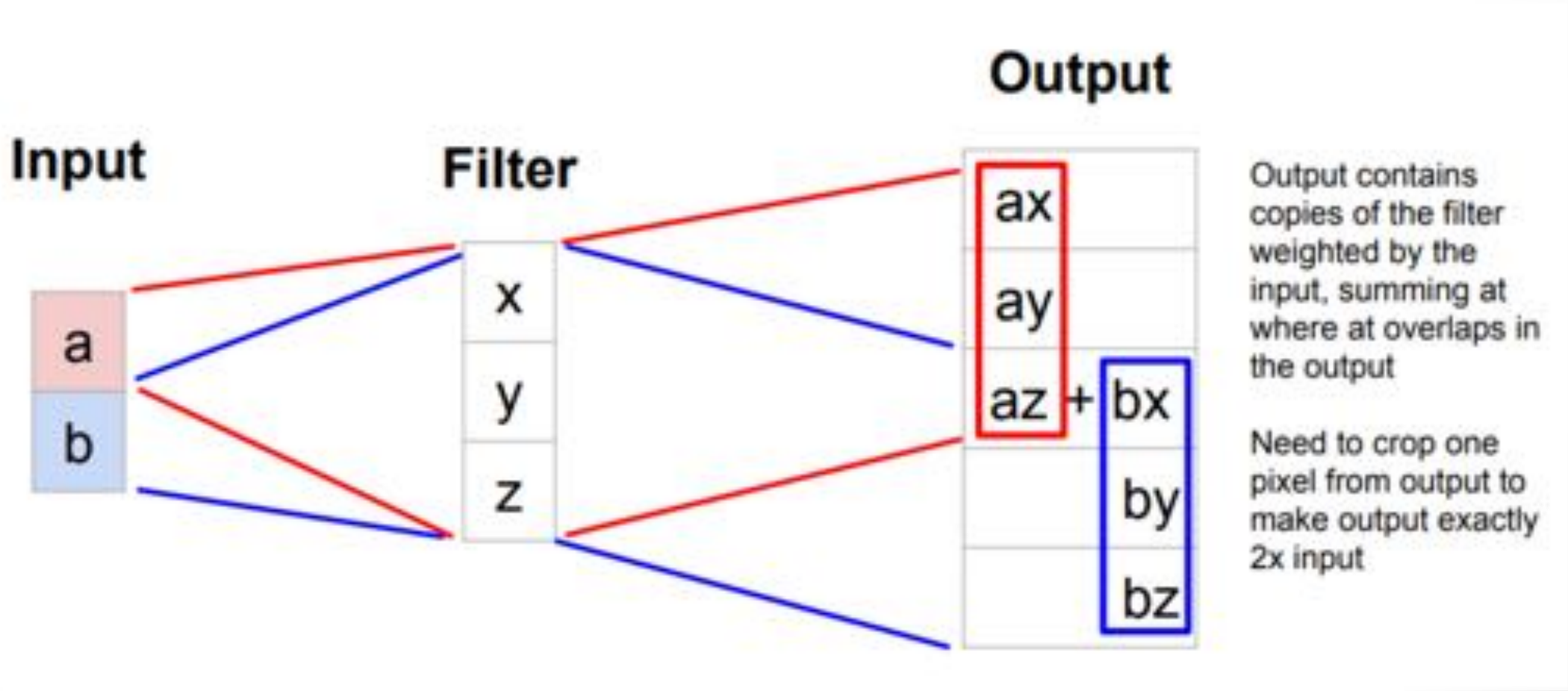


Upsampling: o melhor, transpose convolution

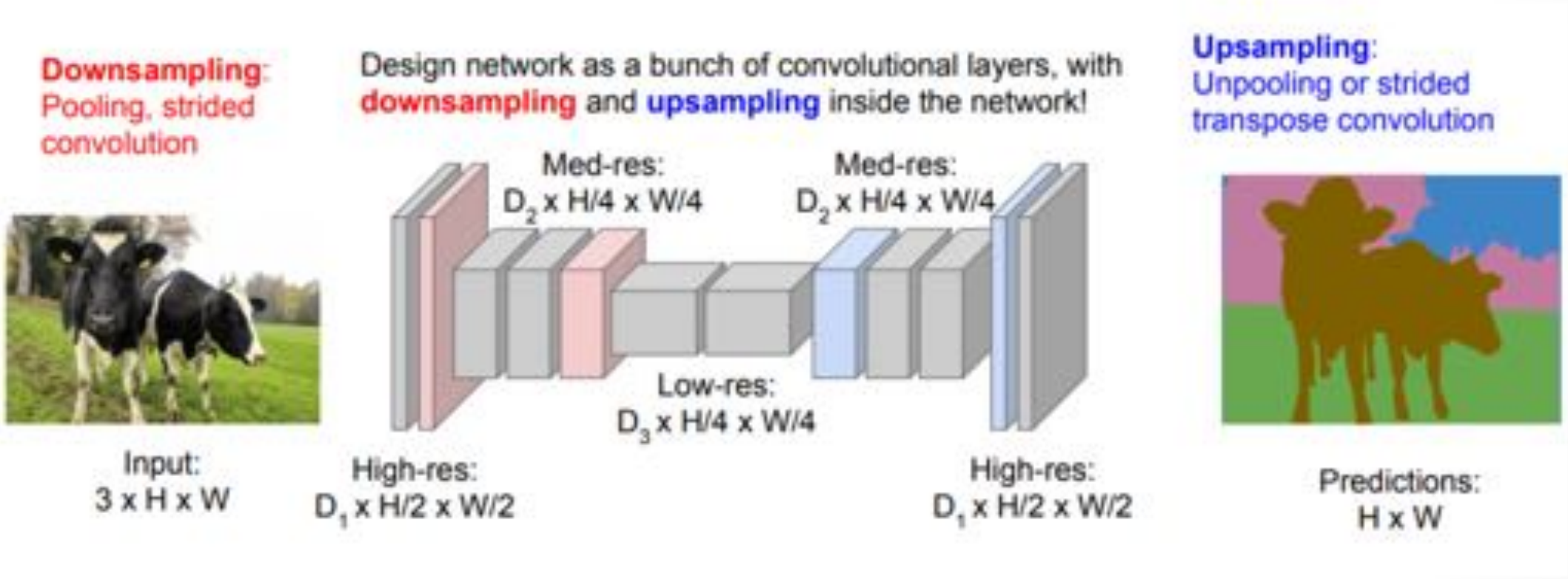
- Lembre-se que é uma convolução, e quem tem filtros, e que serão aprendidos no backpropagation



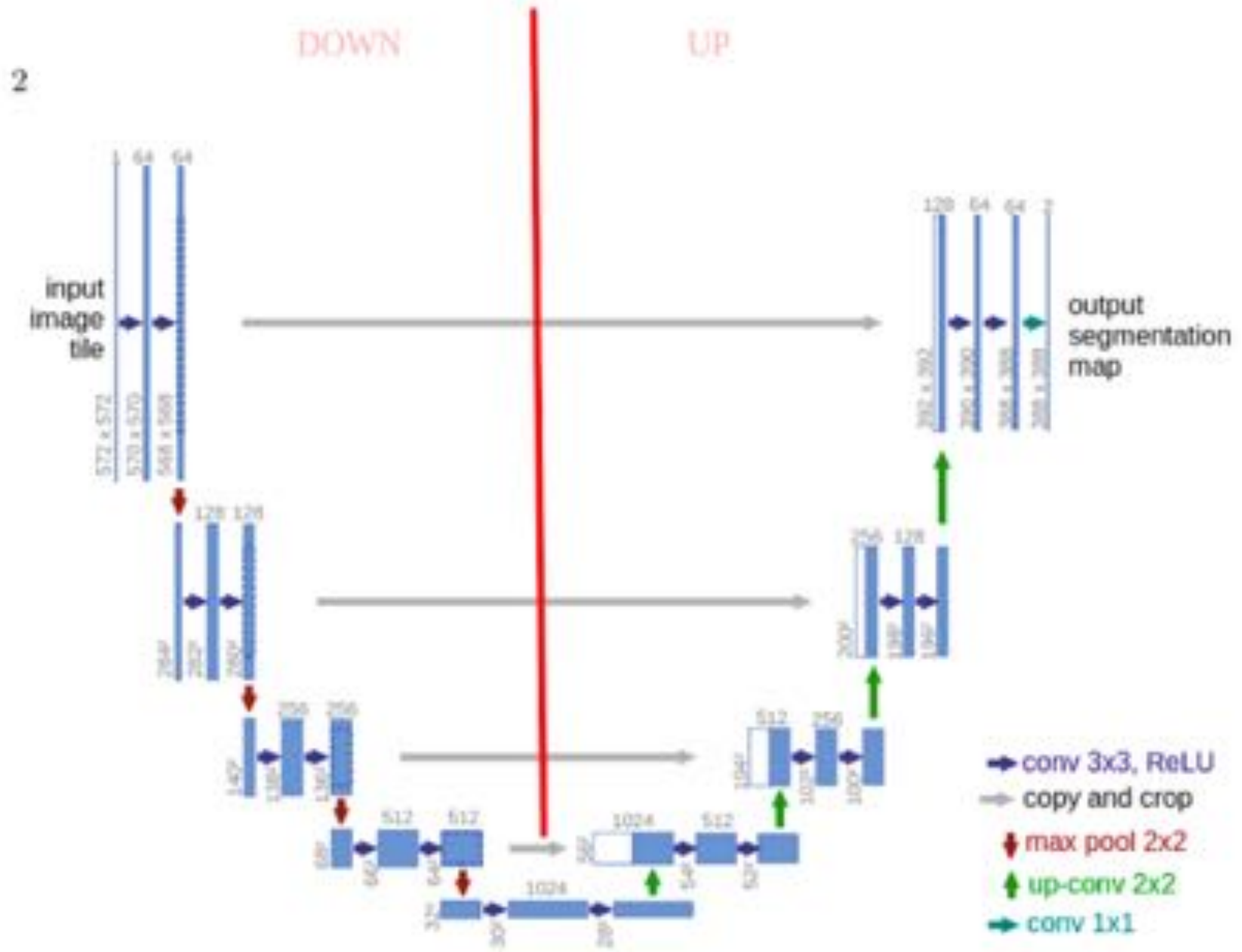
Upsampling: o melhor, transpose convolution



Abordagens mais recentes



U-Net


DOWN
UP
input image tile
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

Ideia por trás

"The main contribution of U-Net in this sense compared to other fully convolutional segmentation networks is that while ***upsampling*** *and going* ***deeper*** *in the network we are* ***concatenating the higher resolution features*** *from* ***down*** *part* ***with the upsampled features*** *in order to* ***better localize*** *and* ***learn*** *representations with following convolutions."*

U-net: Convolutional Networks for Biomedical Image Segmentation: *https://arxiv.org/abs/1505.04597*

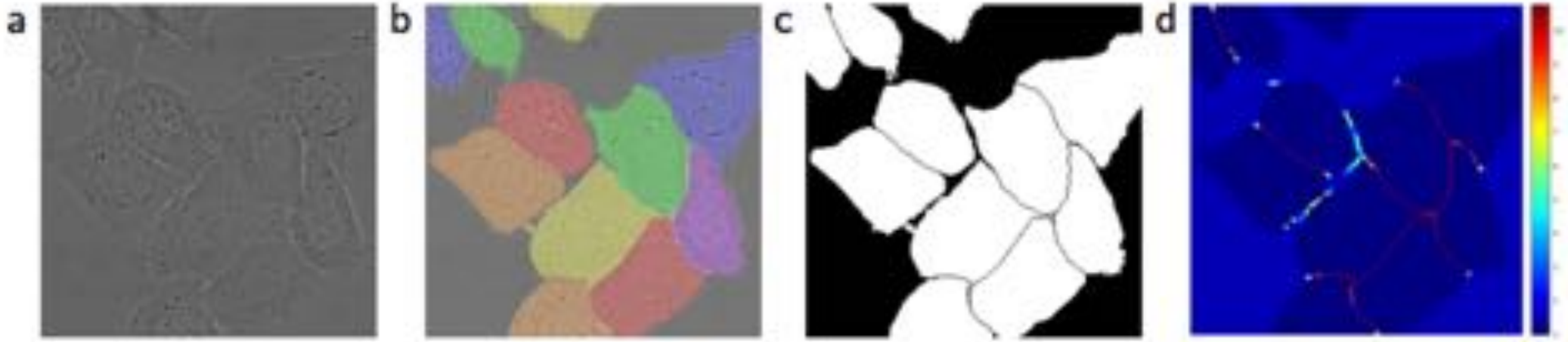


Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background). (**d**) map with a pixel-wise loss weight to force the network to learn the border pixels.

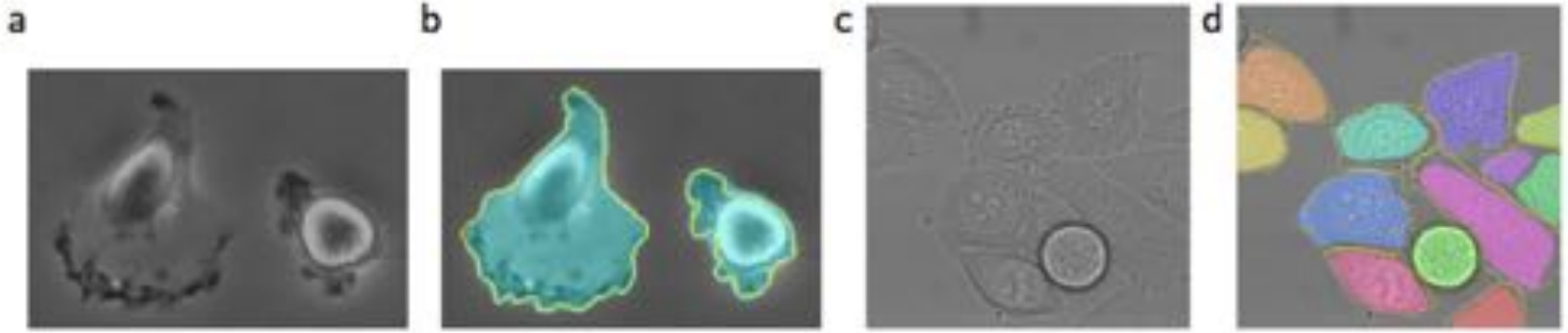
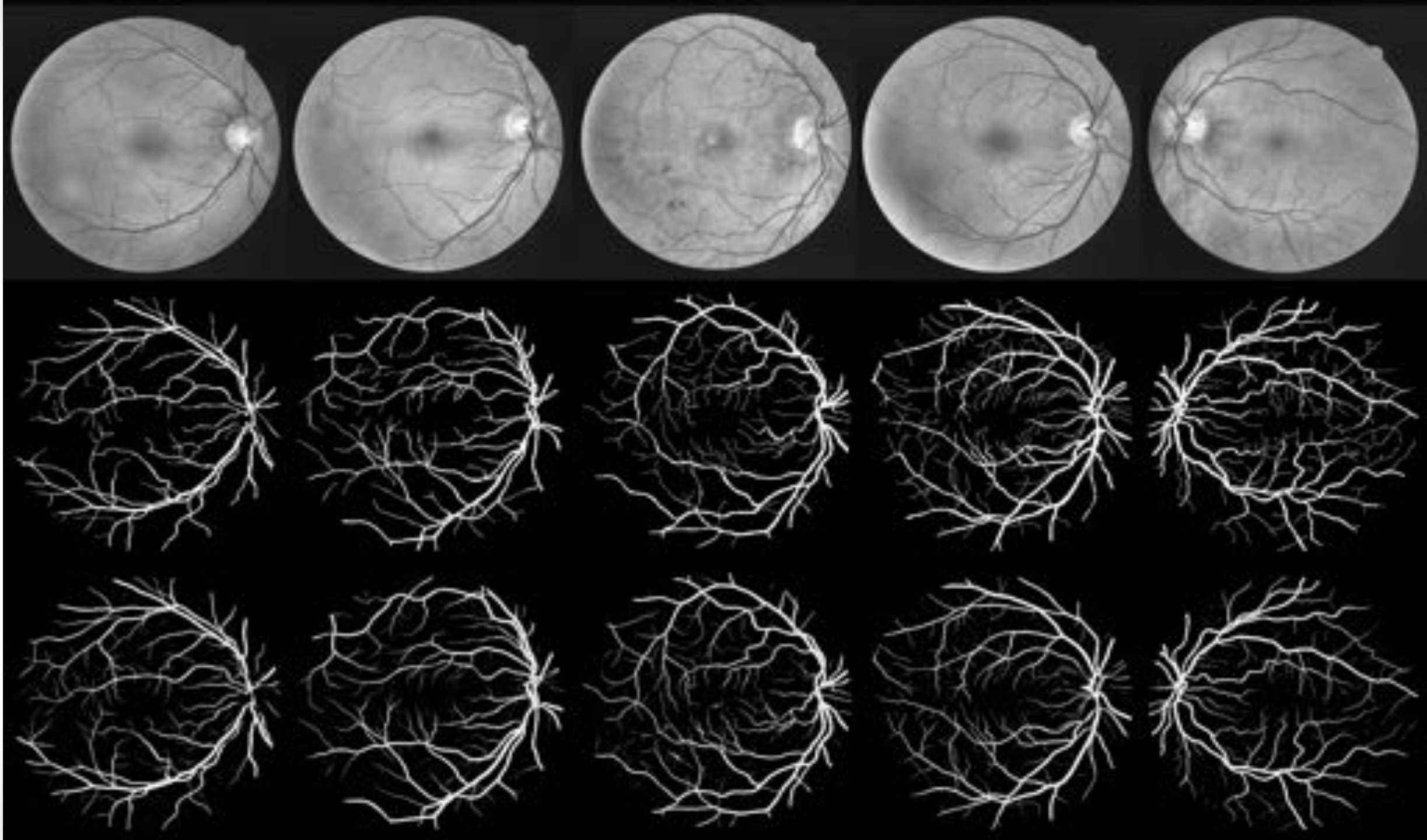
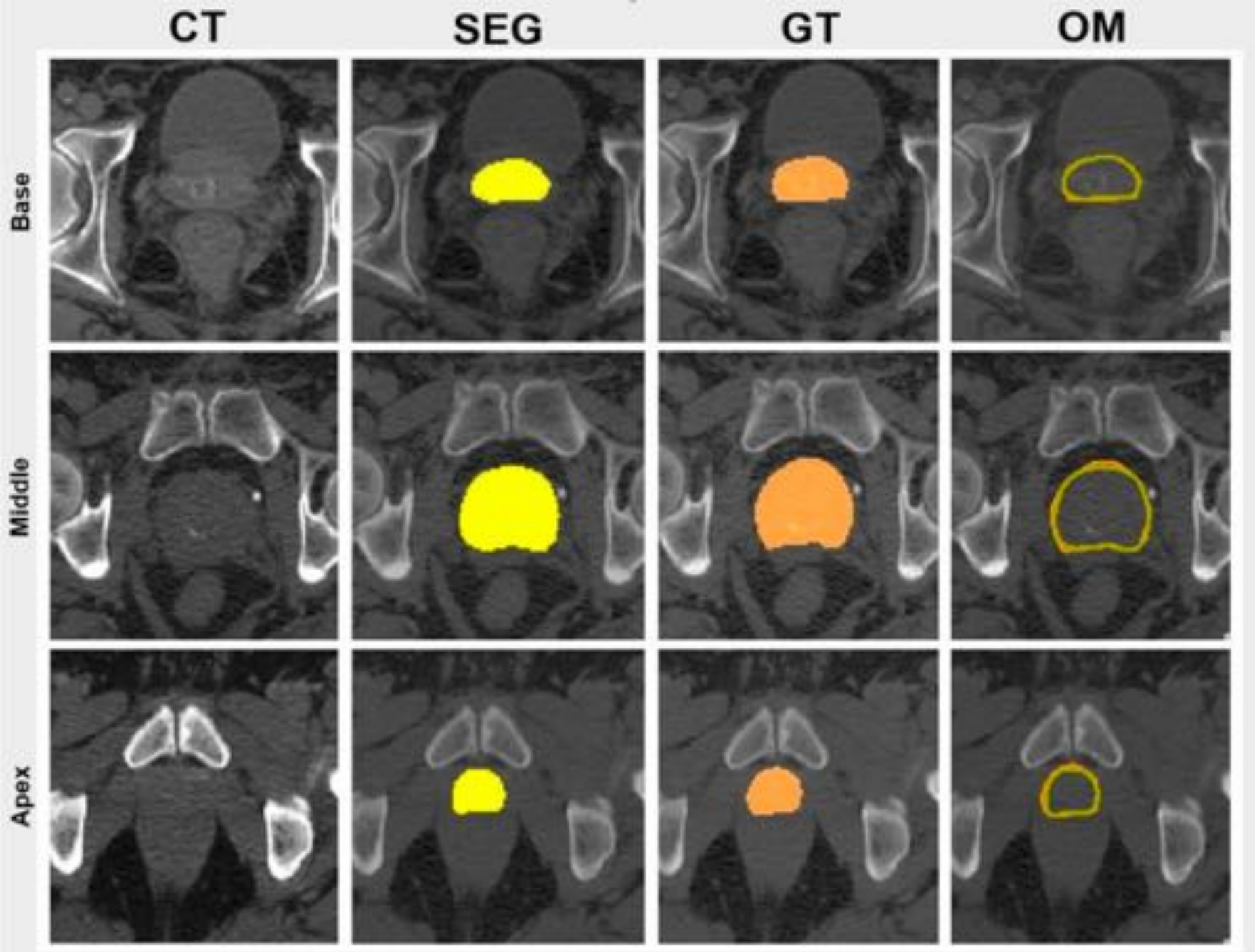


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).



- Segmentação de prostata



Deep Features, pq não? TernausNet

- Transfer Learning com VGG11
  - Conceito de **Encode**: pré-treinado
  - Treine o **Decoder**

- TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation: *https://arxiv.org/abs/1801.05746*