

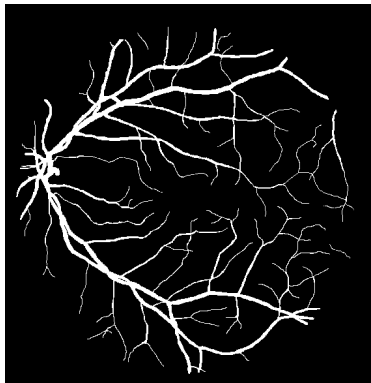
MAC5921 – Deep Learning

Aula 17 – 24/10/2023

**Discussão sobre Task 2, segmentação de imagens
aspectos práticos, image-to-image, etc**

Nina S. T. Hirata

Task 2



U-Net (2015)

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, Thomas Brox

<https://arxiv.org/abs/1505.04597>

U-Net

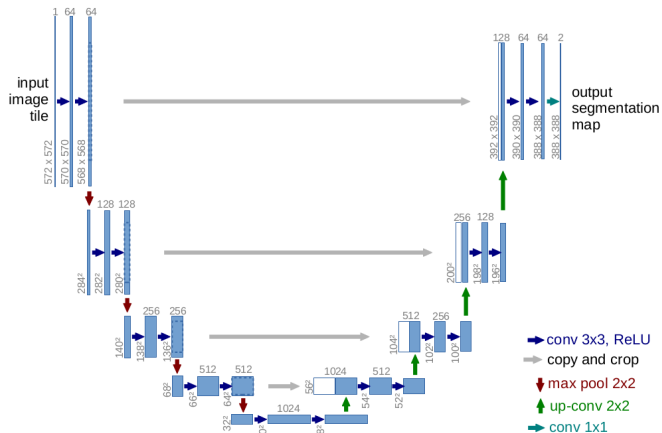
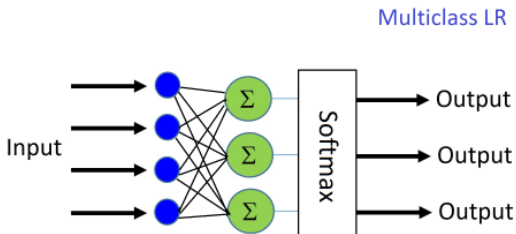
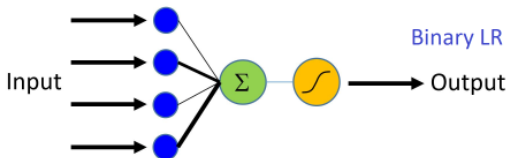


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Classificação binária × multi-classes



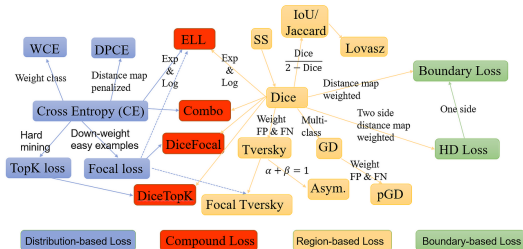
Fonte: https://www.cntk.ai/pythondocs/CNTK_103B_MNIST_LogisticRegression.html

Funções de perda

Segmentation Loss Odyssey

Jun Ma

<https://arxiv.org/abs/2005.13449>



Funções de perda

A survey of loss functions for semantic segmentation

Shruti Jadon

<https://arxiv.org/abs/2006.14822>

pix2pix

Image-to-Image Translation with Conditional Adversarial Networks

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros

<https://arxiv.org/abs/1611.07004>

Alguns pontos discutidos

- Dificuldades: o que está no paper não é exatamente o mesmo que está nas implementações disponíveis
- Diferentes ajustes no tamanho da imagem (crop, preenchimento, resize) para se enquadrar na estrutura da U-Net
- Normalização; inclusive alguns ground-truth tem valor 0-225 e outros 0-1
- Data-augmentation: ajuda? Quais fazem sentido? Dependendo do tipo de transformação, deve-se fazer o mesmo na imagem ground-truth
- Desbalanceamento (pesos para as classes 0 e 1; imagem máscara, máscara usando transformada distância como no artigo da U-Net, funções de perda espertas, ...)

Alguns pontos discutidos

- Segmentação binária (sigmoid/softmax na saída?), segmentação multi-classes (múltiplos canais na saída; um para cada classe)
- Função de perda (Cross-entropy com peso, dice loss, focal loss, combinação de perdas de tipos distintos...)
- Pós-processamento: valor de threshold sobre $p(y = 1|x)$, curva ROC para escolher threshold, faz sentido escolher threshold de forma dinâmica (levando em conta características locais, por exemplo)?
- Modelos pré-treinado? Arquiteturas de tamanhos distintos? Atenção no módulo decoder?