

U-Net: Convolutional Networks for Biomedical Image Segmentation

Dado o cenário de redes neurais profundas a qual a U-Net foi desenvolvida e proposta, observou-se que redes com uma grande quantidade de dados tinham melhores desempenhos, visto que cada vez mais tem-se dados anotados, apresentando redes com até milhões de entradas, cenário diferente ao visto em contextos de saúde. Nesta situação propôs-se uma arquitetura que utiliza de forma mais eficiente os dados disponíveis e também uma boa abordagem de aumento de dados, contraindo o caminho de forma a expandir as experiências visualizadas anteriormente. Possibilitando assim um treinamento que exige uma baixa quantidade de dados anotados.

O estado da arte anterior visava a predição através de uma janela deslizante provendo as classes do pixel e seu derredor, porém apresentavam resultados que eram lentos, se comparados com a arquitetura proposta e também a existência de uma troca entre acurácia local e sobre o contexto geral, enquanto propostas recentes consideram os atributos de múltiplas camadas.

Para alcançar bons resultados, esta rede baseou-se em arquiteturas convolucionais e estendeu-se de forma a necessitar de poucas amostras e uma segmentação mais precisa, para isto, a ideia foi complementar uma rede usual de camadas sucessivas, onde os operadores de pooling são substituídos por operadores de upsampling, onde existe um caminho de contração, dado pelo lado esquerdo de entrada, combinado com uma saída, lado direito da rede, de upsampling, resultando em imagens com boas resoluções de saída.

Para as áreas de upsampling, obtém-se um grande número de atributos o que coopera para a rede ter um alto poder de propagação de contexto, sem utilizar nenhuma camada totalmente conectada. Para obter mais valores de entrada e aumentar a variação dos exemplos vistos, utilizou-se a aumento de dados, através de deslocamentos, rotações e algumas deformações usando vetores de deslocamento aleatório, além de utilizar drop-out como uma aumento implícita. Também têm-se como desafio a segmentação de objetos da mesma classe, para isto neste trabalho, propõe-se uma função de perda ponderada que é capaz de separar os rótulos de fundo e as células.

No treinamento da rede, utilizou-se altos valores de momentum para que a rede consiga lembrar valores previstos a mais tempos. A função de energia é calculada por um softmax em termos de pixel sobre o mapa final combinado com a função de perda de entropia cruzada. O softmax é dado por uma função que relaciona a ativação em um canal na posição de um pixel em relação a somatória das ativações e a Entropia Cruzada então penaliza as posições que desviam do valor desejado. Cada um dos mapas de pedo foram pré calculados para compensar a diferença das frequências de pixel, utilizando a frequência das classes e as distâncias entre as bordas de proximidade. Também foram estudadas e propostas inicializações ótimas de pesos, neste trabalho idealmente iniciados através de uma função de distribuição gaussiana.

Os testes foram realizados utilizando principalmente duas diferentes bases de dados. Uma para segmentação de dados, disponibilizados por ISBI 2012, onde a U-Net apresentou os melhores, até então, e baixos valores de erros, warping error de 0.0003529 e um rand-error de 0.0382. Também aplicou-se a rede para a tarefa de segmentação imagens microscópicas de luz com resultados de IOU ("intersecção sobre união") de 92%. o segundo conjunto de dados foi o DIC-HeLa com células HeLa em um vidro plano registrado por microscopia de contraste de interferência diferencial (DIC), apresentando IOU de 77.5%, resultados estes bem melhores que o estado da arte no momento.