Robust and generalizable nuclei segmentation using deep learning

Christian Eschen¹, MSc student in Medicine and Technology
Supervisors: Jeppe Thaagarj¹, ²Thomas Ebsstrup³, Ole Winther¹ and Rasmus R. Paulsen¹
¹Technical University of Denmark, ²Vsiopharm A/S

Abstract
Identification of cell nuclei in biomedical images is of great importance for research, drug discovery and diagnosis of disease. Examination of cancer suspicious tissue on the microscopic level is the golden standard for diagnosis of almost any cancer type. One of the greatest difficulties in segmentation of cell nuclei in histopathology is separation of adjoining cell nuclei. Separation of adjoining cell nuclei in deep learning is difficult due to the fact that convolutional neural networks (CNN) typically use a category cross entropy, which only work on a pixel-wise level and therefore lacks more global information context. Therefore, there is an increasing demand for robust segmentation algorithms which are designed for separation of cell nuclei.

We present improved segmentation results using a feature engineered weighted loss and an adversarial loss. The best results were obtained using a U-Net architecture and an adversarial loss yielding an AJI equal to 0.452 on the validation set on the MoNuSeg grand challenge. We denot this architecture Pix2pix U-Net.

Data
The MoNuSeg [1] data consists of 30 H&E stained images of cell nuclei from multiple organs with 40x magnification and associated ground truth annotations of cell nuclei. The images are separated into 16 training images and 14 images for validation. A third class is introduced describing boundary of the cell nuclei.

Figure 1: From left to right: H&E stained image of colon tissue. Ground truth image with three classes. Weight map used for training UNET SAW. The weight map is normalized for visualization only.

Methods

U-Net:
We employ a U-Net architecture [2] with a VGG encoder and a decoder with feature concatenation and bilinear upsampling, see Figure 2. The loss function used for training U-Net is:

\[ L = \sum_{i=1}^{K} \sum_{x \in C_i} W(x; i) \cdot g(x; k) \cdot \log(p(y|x)) \]

where \( g(x; k) \) describes the ground truth image with pixels \( X \) and classes \( k \); and the softmax activation is described as \( p(y|x) \). For the U-Net, we set the weight map to \( W(x) = 1 \)

UNET SAW:
As described in [3], the separation border between cells can be enforced using a weight map in the cost function, which takes class frequencies into account and penalizes predictions close to the border of the nuclei:

\[ W(x; i) = \frac{1}{\max(\text{hist}(x; i))} \cdot \frac{1}{\text{class frequency}(i)} + \text{inverse distance} \]

The weight map \( W(x) \) now has a contribution from the class areas \( \text{hist}(x; i) \) and the inverse distance transform \( \text{inverse distance} \cdot \frac{\max(\text{hist}(x; i))}{\text{class frequency}(i)} \) smoothed with a Gaussian kernel \( F_x \). The inverse distance transform is defined by the following equation with \( r \) describing normalization, \( \text{hist}(x; i) \) is the euclidean distance transform of the foreground \( T_F \).

\[ u_{\text{hist}}(x; i) = \begin{cases} \frac{1}{\max(\text{hist}(x; i))} \cdot \frac{1}{\text{class frequency}(i)} + \frac{\max(\text{hist}(x; i))}{\text{class frequency}(i)} \cdot \text{inverse distance} \cdot \frac{1}{F_X(r)} \cdot \frac{1}{\text{class frequency}(i)} \, \quad x \in T_F \\ 0 \, \quad \text{otherwise} \end{cases} \]

The weight map is presented in Figure 1.

Pix2pix:
Instead of feature engineering the loss, one could define an objective “be indistinguishable from the ground truth”. This is the goal in generative adversarial networks (GANs). We utilize the GAN proposed in [4] for semantic segmentation. This is achieved by a so-called min-max game in which a generator \( G \) tries to fool a discriminator \( D \):

\[ G^* = \arg \min G \\arg \max D \mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_{L2}(G) \]

The \( \mathcal{L}_{L2}(G) \) term acts as a regularizer forcing the segmentations to be as close as possible to the ground truth. The Pix2pix U-Net framework is presented in Figure 3.

Figure 2: VGG U-Net

Figure 3: Pix2pix segmentation framework

Results
The segmentation performance is measured using the Dice score, Hausdorff distance and the AJI index [1]. The validation metrics for the models are presented in Figure 4 and the segmentation maps are presented in Figure 5.

Discussion
We present improved discriminative properties of the CNN regarding adjoining cell nuclei separation using U-Net SAW. The best segmentation performance was achieved using the Pix2pix U-Net yielding a mean AJI equal to 0.452. However, the Pix2pix U-Net has not learned the boundary class sufficiently. Future work will investigate methods for successfully learning the boundary class.

References

Contact information
Christian.eschen@hotmail.com
Ph. 20652915

Figure 4: Comparison of the performance of the models

Figure 5: a) H&E stained image and ground truth image of breast tissue. b) Softmax activation and segmentations using U-Net. c) Softmax activation and segmentations using U-Net SAW. d) Softmax activation and segmentations using Pix2pix U-Net.

Figure 6: a) H&E stained image and ground truth image of breast tissue. b) Softmax activation and segmentations using U-Net. c) Softmax activation and segmentations using U-Net SAW. d) Softmax activation and segmentations using Pix2pix U-Net.