ISBI 2018 SUB-CHALLENGE 1: LESION SEGMENTATION

Jaemin Son, Woong Bae

VUNO Inc.
{woalsdnd,iorism}@vuno.co
6F, 507, Gangnam-daero, Seocho-gu,
Seoul, Republic of Korea

1. APPROACHES

We first compared U-Net with atrous CNN for segmentation [1] in the original resolution without down-scaling. We replaced max-pooling in U-Net with the atrous convolution that maintains resolution of feature maps. We experimentally observed that atrous CNN qualitatively segments regions more smoothly than U-Net, however, it yields lower AU-PR on the validation dataset (reserved dataset from the given training dataset). Therefore, we chose U-Net architecture as our baseline.

Then, we reduced image resolution by about 8 times and compared the results with those of original resolution on the validation dataset. We empirically saw that U-Net with lower resolutional input generated qualitatively better segmentation with less false positive and high sensitivity. We suspect that this is because there exists noise in the original image such as jpeg artifact and human annotators would not perceive pixel level granularity but find boundary from lower resolution that can capture the context.

After fixing the model and data input size, we optimized loss function to achieve the best AU-PR on the validation dataset. We modified weighted binary cross entropy to include hyperparameter that can be optimized and find the best solution given limited computing resources.

Our source code is available at `https://bitbucket.org/woalsdnd/isbi-2018-fundus-challenge`.

2. METHODS

2.1. Image Preprocessing

Original image (4288×2848) is cropped to (3500×2848) with offset in x direction to minimize black areas in fundus images. Then, images are padded to (3500×3500) and resized to (640×640) through bicubic interpolation for the network input. Each image is divided 255 to bound values onto [0,1]. In case of HE and MA, each image is divided by 255 and then subtracted with the mean. We did not divide an image by std in order to maintain color contrast which would be important in discriminating subtle changes around lesions. We observed that performance worsens with division by std.

During training, an image is perturbed with affine transformation (flip, scaling, rotation, translation, shear) and random intensity re-scaling.

2.2. Model Architecture

We modified U-Net [2] so that upsampling layers have the same number of feature maps with the corresponding initial layers that are concatenated. This choice bases on the motivation that features in initial layers and upsampled layers have the same importance for segmentation, thus should have the same number of feature maps. Additionally, we adjusted the number of max-pooling in a way that the largest lesion spans a pixel in the most coarse layer. For hard exudates and hemorrhage, our modified U-Net is max-pooled 6 times, while the network for soft exudates and microaneurysm max-pools 4 and 2 times. Details of networks are given in Table. 1, Table. 2, Table. 3. When it comes to MA, we used inverse pixel shuffling to convert $1280 \times 1280 \times 3$ fundus image to $640 \times 640 \times 12$ for network input and pixel shuffling [3] to convert $640 \times 640 \times 4$ segmentation map into $1280 \times 1280 \times 1$.

2.3. Training Details

Pairs of normalized fundus images and lesion segmentations are fed into the network as input and segmentation result is output with range of [0,1]. We used weighted binary cross entropy as loss function which is given by

$$L = \frac{1}{k}\sum_{i=1}^{k}\left[-\alpha \mathbf{y}_{true}^{i} \log \mathbf{y}_{pred}^{i} - (1-\mathbf{y}_{true}^{i}) \log(1-\mathbf{y}_{pred}^{i})\right] \tag{1}$$

where k denotes the number of images in a batch, $\mathbf{y}_{true}^{i}$ and $\mathbf{y}_{pred}^{i}$ represent true segmentation and predicted segmentation for ith image. α is given by

$$\alpha = \frac{N_0^i}{\gamma N_1^i} \tag{2}$$

Table 1: Details of model architecture for segmentation of Hard Exudates and Hemorrhage. Concatenation is described with brackets. *up* means up-scaling of feature maps by scale of 2.

Block	U-Net Operation	U-Net Output size
Input	fundus	(640,640,3)
conv 1	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
pool 1	2 × 2 maxpool	(320, 320, 32)
conv 2	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
pool 2	2 × 2 maxpool	(160, 160, 64)
conv 3	{3 × 3 conv, batch-norm, ReLU} × 2	(160, 160, 128)
pool 3	2 × 2 maxpool	(80, 80, 128)
conv 4	{3 × 3 conv, batch-norm, ReLU} × 2	(80, 80, 256)
pool 4	2 × 2 maxpool	(40, 40, 256)
conv 5	{3 × 3 conv, batch-norm, ReLU} × 2	(40, 40, 512)
pool 5	2 × 2 maxpool	(20, 20, 512)
conv 6	{3 × 3 conv, batch-norm, ReLU} × 2	(20, 20, 512)
pool 6	2 × 2 maxpool	(10, 10, 512)
conv 7	{3 × 3 conv, batch-norm, ReLU} × 2	(10, 10, 512)
up 1	up(conv7)	(20,20,512)
conv 8	{3 × 3 conv, batch-norm, ReLU} × 1	(20, 20, 512)
concat 1	[conv8, conv6]	(20, 20, 1024)
conv 9	{3 × 3 conv, batch-norm, ReLU} × 2	(20, 20, 512)
up 2	up(conv9)	(40,40,512)
conv 10	{3 × 3 conv, batch-norm, ReLU} × 1	(40, 40, 512)
concat 2	[conv10, conv5]	(40, 40, 1024)
conv 11	{3 × 3 conv, batch-norm, ReLU} × 2	(40, 40, 512)
up 3	up(conv11)	(80,80,512)
conv 12	{3 × 3 conv, batch-norm, ReLU} × 1	(80, 80, 256)
concat 3	[conv12, conv4]	(80, 80, 512)
conv 13	{3 × 3 conv, batch-norm, ReLU} × 2	(80, 80, 256)
up 4	up(conv13)	(160,160,256)
conv 14	{3 × 3 conv, batch-norm, ReLU} × 1	(160, 160, 128)
concat 4	[conv14, conv3]	(160, 160, 256)
conv 15	{3 × 3 conv, batch-norm, ReLU} × 2	(160,160, 128)
up 5	up(conv15)	(320,320,128)
conv 16	{3 × 3 conv, batch-norm, ReLU} × 1	(320, 320, 64)
concat 5	[conv16, conv2]	(320, 320, 128)
conv 17	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
up 6	up(conv17)	(640,640,64)
conv 18	{3 × 3 conv, batch-norm, ReLU} × 1	(640, 640, 32)
concat 6	[conv18, conv1]	(640,640,64)
conv 19	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
conv 20	{1 × 1 conv, sigmoid} × 1	(640, 640, 1)

Table 2: Details of model architecture for segmentation of Soft Exudates. Concatenation is described with brackets. *up* means up-scaling of feature maps by scale of 2.

Block	U-Net Operation	U-Net Output size
Input	fundus	(640,640,3)
conv 1	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
pool 1	2 × 2 maxpool	(320, 320, 32)
conv 2	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
pool 2	2 × 2 maxpool	(160, 160, 64)
conv 3	{3 × 3 conv, batch-norm, ReLU} × 2	(160, 160, 128)
pool 3	2 × 2 maxpool	(80, 80, 128)
conv 4	{3 × 3 conv, batch-norm, ReLU} × 2	(80, 80, 256)
pool 4	2 × 2 maxpool	(40, 40, 256)
conv 5	{3 × 3 conv, batch-norm, ReLU} × 2	(40, 40, 512)
up 1	up(conv7)	(80,80,512)
conv 6	{3 × 3 conv, batch-norm, ReLU} × 1	(80, 80, 256)
concat 1	[conv6, conv4]	(80, 80, 512)
conv 7	{3 × 3 conv, batch-norm, ReLU} × 2	(80, 80, 256)
up 2	up(conv7)	(160,160,256)
conv 8	{3 × 3 conv, batch-norm, ReLU} × 1	(160, 160, 128)
concat 2	[conv14, conv3]	(160, 160, 256)
conv 9	{3 × 3 conv, batch-norm, ReLU} × 2	(160,160, 128)
up 3	up(conv9)	(320,320,128)
conv 10	{3 × 3 conv, batch-norm, ReLU} × 1	(320, 320, 64)
concat 3	[conv10, conv2]	(320, 320, 128)
conv 11	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
up 4	up(conv11)	(640,640,64)
conv 12	{3 × 3 conv, batch-norm, ReLU} × 1	(640, 640, 32)
concat 4	[conv12, conv1]	(640,640,64)
conv 13	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
conv 14	{1 × 1 conv, sigmoid} × 1	(640, 640, 1)

Table 3: Details of model architecture for segmentation of Microaneurysm. Concatenation is described with brackets. *up* means up-scaling of feature maps by scale of 2.

Block	U-Net Operation	U-Net Output size
Input	fundus	(1280,1280,3)
IPS	Inverse Pixel Shuffling	(640,640,12)
conv 1	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
pool 1	2 × 2 maxpool	(320, 320, 32)
conv 2	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
pool 2	2 × 2 maxpool	(160, 160, 64)
conv 3	{3 × 3 conv, batch-norm, ReLU} × 2	(160, 160, 128)
up 1	up(conv3)	(320,320,128)
conv 4	{3 × 3 conv, batch-norm, ReLU} × 1	(320, 320, 64)
concat 1	[conv4, conv2]	(320, 320, 128)
conv 5	{3 × 3 conv, batch-norm, ReLU} × 2	(320, 320, 64)
up 2	up(conv5)	(640,640,64)
conv 6	{3 × 3 conv, batch-norm, ReLU} × 1	(640, 640, 32)
concat 2	[conv6, conv1]	(640,640,64)
conv 7	{3 × 3 conv, batch-norm, ReLU} × 2	(640, 640, 32)
conv 8	{3 × 3 conv, sigmoid} × 1	(640, 640, 4)
PS	Pixel Shuffling	(1280,1280,1)
conv 9	{3 × 3 conv} × 1	(1280, 1280, 1)
conv 10	{3 × 3 conv, sigmoid} × 1	(1280, 1280, 1)

where N_0^i and N_1^i denote the number of background and foreground in the ith image. Since background overwhelms foreground in the lesion segmentation, this loss function that penalizes false negative would boost sensitivity which is important in detecting lesions. γ is left as a hyperparameter and chosen out of $\{0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 256, 512\}$ to yield the highest AU-PR on the validation set. γ values are summarized in Table. 4.

Table 4: γ values in Eq. 2.

Hard Exudates	Soft Exudates	Hemorrhage	Microaneurysm
64	512	8	32

Networks are trained over total 300 epochs. We used Adam optimizer [4] with hyper-parameters of $\beta_1 = 0.5, \beta_2 = 0.999$ and learning rate of $2e^{-4}$ until 250 epochs and $2e^{-5}$ until the end.

We added images labeled with no-DME in sub-challenge 2 as negative samples for hard exudate since no-DME means there exists no hard exudates. Also, images labeled as no-DR are included as negative samples for hemorrhage, microaneurysm. For soft exudate, we only used images with apparent retinopathy since it is easy to be under-segmented.

Entire dataset is split into training dataset and validation dataset. Validation dataset consists of the first 10% images after sorting image files by name. For each lesion, testset is processed with a model that yielded the best AU-PR on the validation dataset.

All implementation is done by Keras 2.0.8 with tensorflow backend 1.4.0.

2.4. Results

We measured Area Under the Precision and Recall curve (AU-PR) and Area Under the Receiver Operating Curve (AU-ROC) for the entire dataset, and training and validation datasets. All segmented images are up-scaled to the original resolution and concatenated to compute the metrics to measure the aggregate performance on the dataset. Metrics module in sklearn package (v.0.18.1) to compute AU-ROC and AU-PR. AU-PR is calculated after retrieving a PR curve and measured the area under the curve with the trapezoidal rule.

Results are shown in Table. 5 and Table. 6.

Table 5: AU-PR and AU-ROC for lesions.

Lesion	AU-PR (all)	AU-PR (training)	AU-PR (validation)
Hard Exudates	0.8668	0.8716	0.8680
Soft Exudates	0.6370	0.6263	0.7160
Hemorrhage	0.8162	0.8246	0.6171
Microaneurysm	0.4840	0.4889	0.4084

Table 6: AU-PR and AU-ROC for lesions.

Lesion	AU-ROC (all)	AU-ROC (training)	AU-ROC (validation)
Hard Exudates	0.9992	0.9994	0.9984
Soft Exudates	0.9984	0.9983	0.9990
Hemorrhage	0.9966	0.9987	0.9503
Microaneurysm	0.9965	0.9968	0.9904

3. REFERENCES

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[3] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[4] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.