MICCAI 2020 REFUGE2 TECHNICAL REPORT

Jaemin Son

VUNO Inc.
{woalsdnd}@vuno.co
5F, 507, Gangnam-daero, Seocho-gu,
Seoul, Republic of Korea

All implementation is done by Keras 2.1.6 with TensorFlow backend 1.13.1 in Python3. Our source code is available at `https://bitbucket.org/woalsdnd/refuge2/src/` (codes will be available after the final competition).

1. AUGMENTATION

In all tasks, we used the same augmentation strategies, namely, a combination of rotation, flip, affine transformation, image perturbation (color, contrast, brightness, sharpness, RGB shift, Gamma), noise (blur, ISO noise, Gaussian noise, JPEG compression, sunflare), resize, elastic transform, and down-sampling. For detailed hyperparameters of each augmentation, readers can refer to the code repository.

2. GLAUCOMA CLASSIFICATION

2.1. Preprocessing

Two preprocessing methods were applied. One is normalization, which divides the pixel values by 255. The other is what we call zero-local-mean, which subtract local mean computed by a gaussian kernel. The exact code for zero-local-mean can be implemented by cv2 library (addWeighted and GaussianBlur) in Python.

2.2. Model Architecture

We used the network that was used in ISBI 2020 ADAM challenge for AMD classification. Some descriptions may overlap with the technical report in ISBI 2020 ADAM challenge.

We used an architecture of a commercial system that outputs multiple ophthalmologic findings, as introduced in [1] with an architectural modification reflecting the tenet of EfficientNet [2] (Fig1).

Input size of the network is 1024×1024 and the output feature maps are fifteen findings (hemorrhage, hard exudate, cotton wool patch, drusen, retinal pigmentary change, vascular abnormality, membrane, fluid accumulation, chorioretinal atrophy, choroidal lesion, myelinated nerve fiber, retinal nerve fiber layer defect, glaucomatous disc change, non-

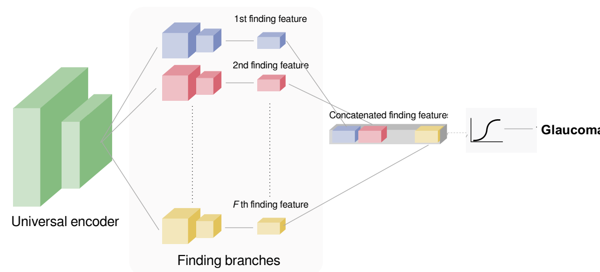

Fig. 1: Architecture of classification network

glaucomatous disc change, and macular hole). All feature maps of findings are concatenated and a fuuly connected layer follows with a sigmoid activation to output the score of glaucoma in the range of 0 to 1. The last fully connected layer is the only part that is trainable.

2.3. Training Details

We trained the network using REFUGE 1 (1200 images) dataset and validation dataset of REFUGE 2 (400 images). The validation dataset of REFUGE 2 was split into 4 folds with the equal positive-negative ratio and used used as a validation set. REFUGE 1 data were used for training throughout 4 folds. For each of 4 folds, 1500 images were used for training, and 100 images were used as a validation set.

For each fold, a model that achieved the best validation auroc was chosen. For each of two preprocessing methods, we obtained 4 models. The total of 8 models were ensembled by averaging the outputs.

Batch size was set to 64.

2.4. Learning Rate

SGD was used as an optimizer and a learning rate was initially set to 10e-3. If a validation auroc does not improve over 10 epochs, the learning rate was reduced by dividing by 10. We set the minimum threshold value below which the learning rate could not be set. The learning rate was set to the

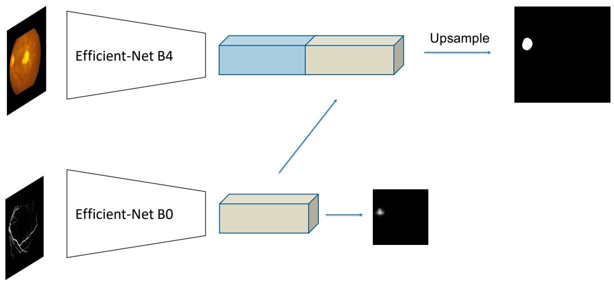

Fig. 2: Architecture of disc, cup segmentation network

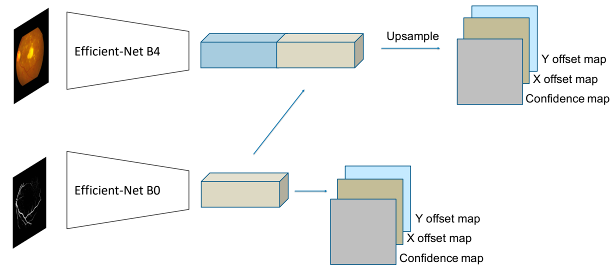

Fig. 3: Architecture of fovea localization network

initial value in case the learning is reduced to a value below the minimum threshold.

3. CUP, DISC SEGMENTATION

3.1. Preprocessing

A given fundus image was standardized channel-wise, or formally described as

$$\hat{I}_i = \frac{I_i - \mu_i}{\sigma_i} \tag{1}$$

where $I_i \in \mathbb{R}^2$ one of RGB channels and μ_i, σ_i represents the mean and standard deviation of values at the channel i.

3.2. Model Architecture

We used the network that was used in ISBI 2020 ADAM challenge for disc segmentation. Some descriptions may overlap with the technical report in ISBI 2020 ADAM challenge. The same architecture was used for disc and cup segmentation (Fig2).

The input size is 640×640 for both a standardized fundus image (whole image) and a vessel segmentation.

The network consists of two branches. One branch, which is Efficient-Net [2] B4, processes a fundus image, and the other branch, which is Efficient-Net B0, operates on a vessel segmentation image. The penultimate feature maps of the fundus branch, with the resolution of 10×10, are concatenated to those of the vessel branch. Then, decoding layers are appended to the concatenated feature maps using depth-wise separable convolutions [3], swish activation functions, and depth-wise feature concatenation with the same-sized features of the fundus branch, and upscale by a factor of 2. Batch normalization was not used. The decoder upscaled up to 160×160. The 160×160 feature maps were upscaled into 640×640 feature maps and the final segmentation layer is generated using a 1×1 convolution followed by a sigmoid function.

Batch size was set to 4.

3.3. Training Details

We finetuned the network that was presented in ISBI 2020 ADAM challenge. That network used several external datasets for training - RIGA [4], IDRiD [5], REFUGE [6], PALM [7]. We used a GAN-based vessel segmentation model to segment retinal vasculatures [8].

We post-processed a segmentation map from the network by filling holes inside the largest blob found after the thresholding for both disc and cup segmentation. The optimal threshold was found through measures on the leader board. When merging binary masks for disc and cup segmentations to generate the submission format, any foreground pixels for the cup outside the disc are considered as background.

Top models on the leader board at threshold 0.5 were ensembled for final submission. Four models were selected for disc segmentation and two models were chosen for cup segmentation.

4. FOVEA LOCALIZATION

4.1. Preprocessing

The same standardization method was used as in the disc, cup segmentation.

4.2. Model Architecture

We used the network that was used in ISBI 2020 ADAM challenge for disc segmentation. Some descriptions may overlap with the technical report in ISBI 2020 ADAM challenge.

The architecture (Fig3) is similar to that was used in disc, cup segmentation except for the last part - the decoding layers upscale up to 80×80 and 3 output maps are generated consequently, namely, a confidence map, a map for x-offset, and a map for y-offset. Elements in every output map take values in the range of 0 to 1, as processed by a sigmoid function. Maps for x-offset and y-offsets are maps for bounding box offsets in detection networks.

The network takes a standardized fundus image (whole image) and a vessel segmentation map as input.

4.3. Training Details

We used training dataset of REFUGE 1 as the annotations in the validation and test sets in REFUGE 1 seem inaccurate.

In fact, the results were worse on the leader board when the validation and test sets were used for training.

We finetune the network that was used for ISBI 2020 competition. That network was trained using several external datasets - IDRiD [5], REFUGE [6], PALM [7]. We used a GAN-based vessel segmentation model to segment retinal vasculatures [8].

Learning rate was set to 1e-5 uniformly for training. Ten Models that has around L2 error of 8 pixels in the original resolution were ensembled by taking the average for the final submission.

Batch size was set to 4.

4.4. Postprocessing

A cell with the highest confidence, or the highest value in the confidence map, was searched and the exact coordinates were computed using x and y offsets at the corresponding cell. No other postprocessing was applied.

5. REFERENCES

[1] Jaemin Son, Joo Young Shin, Hoon Dong Kim, Kyu-Hwan Jung, Kyu Hyung Park, and Sang Jun Park, "Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images," *Ophthalmology*, vol. 127, no. 1, pp. 85–94, 2020.

[2] Mingxing Tan and Quoc V Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[3] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[4] Ahmed Almazroa, Sami Alodhayb, Essameldin Osman, Eslam Ramadan, Mohammed Hummadi, Mohammed Dlaim, Muhannad Alkatee, Kaamran Raahemifar, and Vasudevan Lakshminarayanan, "Retinal fundus images for glaucoma analysis: the riga dataset," in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*. International Society for Optics and Photonics, 2018, vol. 10579, p. 105790B.

[5] Prasanna Porwal, Samiksha Pachade, Manesh Kokare, Girish Deshmukh, Jaemin Son, Woong Bae, Lihong Liu, Jianzong Wang, Xinhui Liu, Liangxin Gao, et al., "Idrid: Diabetic retinopathy–segmentation and grading challenge," *Medical image analysis*, vol. 59, pp. 101561, 2020.

[6] José Ignacio Orlando, Huazhu Fu, João Barbossa Breda, Karel van Keer, Deepti R Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, et al., "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical image analysis*, vol. 59, pp. 101570, 2020.

[7] Huazhu Fu; Fei Li; José Ignacio Orlando; Hrvoje Bogunović; Xu Sun; Jingan Liao; Yanwu Xu; Shaochong Zhang; Xiulan Zhang, "Palm: Pathologic myopia challenge," 2019.

[8] Jaemin Son, Sang Jun Park, and Kyu-Hwan Jung, "Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks," *Journal of digital imaging*, vol. 32, no. 3, pp. 499–512, 2019.