



Published in final edited form as:

Mach Learn Med Imaging. 2022 September ; 13583: 436–445. doi:10.1007/978-3-031-21014-3_45.

Multi-scale Multi-structure Siamese Network (MMSNet) for Primary Open-Angle Glaucoma Prediction

Mingquan Lin¹, Lei Liu², Mae Gorden², Michael Kass², Sarah Van Tassel¹, Fei Wang¹, Yifan Peng¹

¹Weill Cornell Medicine, New York, NY, USA

²Washington University School of Medicine in St. Louis, St. Louis, MO, USA

Abstract

Primary open-angle glaucoma (POAG) is one of the leading causes of irreversible blindness in the United States and worldwide. POAG prediction before onset plays an important role in early treatment. Although deep learning methods have been proposed to predict POAG, these methods mainly focus on current status prediction. In addition, all these methods used a single image as input. On the other hand, glaucoma specialists determine a glaucomatous eye by comparing the follow-up optic nerve image with the baseline along with supplementary clinical data. To simulate this process, we proposed a Multi-scale Multi-structure Siamese Network (MMSNet) to predict future POAG event from fundus photographs. The MMSNet consists of two side-outputs for deep supervision and 2D blocks to utilize two-dimensional features to assist classification. The MMSNet network was trained and evaluated on a large dataset: 37,339 fundus photographs from 1,636 Ocular Hypertension Treatment Study (OHTS) participants. Extensive experiments show that MMSNet outperforms the state-of-the-art on two “POAG prediction before onset” tasks. Our AUC are 0.9312 and 0.9507, which are 0.2204 and 0.1490 higher than the state-of-the-art, respectively. In addition, an ablation study is performed to check the contribution of different components. These results highlight the potential of deep learning to assist and enhance the prediction of future POAG event. The proposed network will be publicly available on <https://github.com/bionlplab/MMSNet>.

Keywords

Deep learning; Primary open-angle glaucoma (POAG); Fundus photographs; Siamese network

1 Introduction

Primary open-angle glaucoma (POAG) is one of the leading causes of blindness worldwide [1]. In the United States, POAG is the most common form of glaucoma and is the leading cause of blindness among African-Americans [22] and Hispanics [11]. POAG can be asymptomatic until very advanced stages. Fortunately, most blindness caused by POAG can be avoided by early identification and treatment [23]. Therefore, early prediction of eyes that

will develop POAG plays an important role in patient monitoring and medical and surgical treatments [5,21].

Fundus photography provides a convenient and inexpensive way to record the optic nerve head structure, which is the gold standard for showing a classic glaucomatous appearance. Unfortunately, the low prevalence of glaucoma, the limited number of trained physicians, and the complex logistics of traditional screening programs are obstacles to timely screening for many patients [14]. Therefore, it is crucial to develop automatic models with high accuracy to assist clinicians in predicting future glaucoma event from fundus photographs, which can help many patients avoid blindness.

Recently, deep learning models have been successfully applied to biology and medicine [3,6–9,18,20,25,26]. In the ophthalmology domain, several models have been proposed to detect POAG from fundus photographs [2,4,16,17]. All of these studies predict the current glaucomatous status of a patient.

In this study, we seek to predict the probability of POAG onset from fundus photos. Such prediction may identify patients appropriate for early treatment. The Ocular Hypertension Treatment Study cohort (OHTS) [12], a large-scale clinical trial from 22 centers in the United States, includes longitudinal fundus photographs and disease assessment over a period of approximately 16 years. It provides an unprecedented opportunity to investigate POAG onset prediction using the dynamic fundus images.

To the best of our knowledge, only two previous works have focused on prediction of future POAG event [15,24]. However, the input to all these approaches is a single image, which might affect the performance of the model. In clinical practice, patients usually have follow-up visits to screen for glaucoma progression. Therefore, the glaucoma specialists compare the follow-up with the baseline image (the image taken at the first visit of a study) to trace the relevant feature, as shown in Fig. 1.

In this study, we used fundus images to predict an eye's progress to POAG (which may never occur) within specific inquired durations from the current visit. The inquired duration was selected in advance (2-year or 5-year), and it was relative to the time when the image was taken, not to the time of the baseline visit. Unlike prior studies, for one eye, the inputs included one fundus image taken at the baseline (first visit), and one image was taken at the current visit (follow-up image). Therefore, our proposed method is suitable for screening patients during follow-up visits. We never need "future images" to screen people. The output was the probability that the time to POAG onset exceeds the inquired duration. To handle the pair of images, we proposed a novel Siamese network model with side output and additional convolution, called multi-scale multi-structure Siamese network (MMSNet), by comparing the differences between two input images. Different from previous Siamese work, MMSNet used the additional 2D features by convolution operation together with cosine similarity, instead of the cosine similarity only to measure the similarity between two outputs of the network instead of the cosine similarity between the two outputs. In addition, the MMSNet also consists of side output [19] to ease the vanishing gradient problems in training deep models and to drive the hidden layers for favoring discriminative features. To the best of our

knowledge, it is the first time in the ophthalmology domain that two fundus images have been utilized and compared for automated glaucoma prediction before onset by Siamese networks.

Our work has the following contributions: (1) We proposed a model to compare the similarity between baseline and follow-up images. The model simulates the glaucoma screening process in clinical practice. To the best of our knowledge, no previous studies tackle the glaucoma prediction before onset problem this way. (2) We use the 2D features by convolution operation together with cosine similarity, instead of the cosine similarity only to measure the similarity between the outputs of two networks. (3) We incorporate the side-output to improve the performance further. The side-output can ease the vanishing gradient problems in training deep models and drive the hidden layers to favor discriminative features. (4) Our approach achieves superior POAG prediction 2-year and 5-year before onset results (93.12% and 95.07% in AUC) against several competitive baselines on the large, multi-institutional benchmarks.

2 Methods

2.1 MMSNet Architecture

MMSNet comprises two convolutional blocks that share the weight and two prediction blocks (Fig. 2). In the beginning, two fundus images x_1 and x_2 are passed through the convolutional neural network, DenseNet-201 [10], respectively. We used the output of last (F_{d4} and F_{d4n}) and second to last Dense Blocks (F_{d3} and F_{d3n}). For each output, we feed the outputs into the prediction block. Each prediction block has two paths to generate prediction results. For the first path, we got the feature embedding by average pooling, followed by cosine similarity to measure the similarity of feature embedding, which is a traditional Siamese network operation. We got the prediction result by sigmoid activation. Another path is the 2D block. We concatenated two outputs, followed by a 1×1 convolution, a batch normalization (BN), and rectified linear units (ReLU). In the end, a global average pooling and a fully connected layer with sigmoid activation are attached. The average of the predicted values of the two paths is used as the final output of each prediction block, regarded as the side output.

2.2 Loss Function

In this study, we use binary cross-entropy as the loss function in the MMSNet. In addition, to overcome the severe class imbalance for the POAG classification, we apply the weighted cross-entropy, a commonly used loss function in classification. The adopted weighted cross-entropy was as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N [\beta y_n \log(\hat{y}_n(x_n, \theta_s)) + (1 - \beta)(1 - y_n) \log(1 - \hat{y}_n(x_n, \theta_s))] \quad (1)$$

N is the number of training examples. β is the balancing factor between positive and negative samples. Here, we used inversely proportional to POAG frequency in the training

data. y_n is the observe value, \hat{y} is the probability predicted by the classifier, and θ_s represents the parameters of the neural network.

The overall loss function is the average of the losses associated with the prediction from the two prediction blocks:

$$\mathcal{L}_s = \alpha \mathcal{L}^{(1)} + (1 - \alpha) \mathcal{L}^{(2)}, \quad (2)$$

2.3 Data Augmentation

In this work, we sequentially apply the following augmentation techniques on the fly during training: (1) random rotation between 0° and 10° , (2) random translation: an image was translated randomly along the x- and y-axes by distances ranging from 0 to 10% of width or height of the image, and (3) random flipping.

3 Results

3.1 Datasets

In this study, we include one independent dataset (Table 1). This database is a large-scale, cross-sectional, longitudinal, and population-based study.

The dataset is obtained from the Ocular Hypertension Treatment Study (OHTS). OHTS is one of the largest longitudinal clinical trials in POAG (1,636 participants and 37,399 images) from 22 centers in the United States. This study does not need Institutional Review Board approval because it does not constitute human subjects research.

The participants in this dataset were selected according to both eligibility and exclusion criteria. The gold standard POAG labels were graded at the Optic Disc Reading Center. In brief, two masked certified readers were arranged to independently detect the optic disc deterioration. If there was a disagreement between two readers, a senior reader reviewed it in a masked fashion. The POAG diagnosis in a quality control sample of 86 eyes (50 normal eyes and 36 with progression) showed test-retest agreement at $\kappa = 0.70$ (95% confidence interval [CI], 0.55–0.85). More details of the reading center workflow have been described in [12]. For the OHTS dataset, we split the entire dataset randomly at the patient level. We take one group (20% of total subjects) as the hold-out test set and the remaining as the training set.

We compare the baseline image with each follow-up image and they compromise pairs separately. In each pair, the follow-up image and the baseline image come from the same eye. In this study, all eligible subjects are non-POAG at baseline. For 2-year POAG prediction before onset, in POAG pair, x_1 refers to the baseline image and x_2 refers to the follow-up image that converted to POAG within two years. In normal pair, x_1 refers to the baseline image and x_2 refers to the follow-up image that did not convert to POAG within two years no matter whether it converts to POAG eventually. The definition of the POAG pair and the normal pair in the 5-year POAG prediction before onset can be derived similarly.

3.2 Evaluation Metrics

To evaluate the performance of POAG within a certain duration, we compute accuracy, sensitivity, specificity, and AUC (Area Under the ROC curve).

3.3 Experimental Settings

We first trained DenseNet-201 on POAG prediction using a single image as input. Then we initialized the subnets in the MMSNet using the DenseNet-201 and fine-tuned the entire network in an end-to-end manner.

All images are resized to $224 \times 224 \times 3$ as input of the proposed model. The models were implemented by Keras with a backend of Tensorflow. The proposed network was optimized using the Adam optimizer method [13]. The learning rate is 5×10^{-5} and α is 0.8. The experiments were performed on Intel Core i9-9960 X 16 cores processor and NVIDIA Quadro RTX 6000 GPU. The training time was 103 mins, and the testing time was 12 mins.

3.4 Results and Discussion

We compare our method with six models on POAG prediction on the OHTS dataset, including the DenseNet-201 with a single image as input, MobileV2 with a single image as input [24], the traditional Siamese network with an absolute difference, the traditional Siamese network with cosine similarity, MMSNet using the last DenseNet Block (MMSNet w/o side output), and MMSNet without 2D Block (MMSNet w/o 2D block).

2-Years POAG Prediction Before the Onset.—Table 2 shows the performance comparison for 2-year POAG prediction before onset. Our model achieved the best results, with an accuracy of 0.9337, a sensitivity of 0.6485, a specificity of 0.9414, and an AUC of 0.9312. Compared to DenseNet-201, which is the best model among those baseline models with a single image as input, the proposed MMSNet possesses 27.90% higher accuracy, 28.70% higher specificity, and 22.04% higher AUC, let alone MobileV2 which also uses a single image as input that is used in [24]. The result obtained by the Siamese network with cosine similarity is better than the results obtained by the DenseNet-201, indicating that the Siamese network, which imitates the clinical process (with multiple inputs), is more precise for 2-year POAG prediction before onset. As the results listed in row 3 and row 4, the Siamese network with cosine similarity works better than the Siamese network with an absolute difference.

5-Year POAG Prediction Before the Onset.—Table 3 compares the results of MMSNet with 6 state-of-the-art models on the OHTS dataset for POAG prediction 5-years before onset. Our model obtained the best results, with an accuracy of 0.9414, a sensitivity of 0.7530, a specificity of 0.9520, and an AUC of 0.9507. Compared to the baseline (DenseNet-201), MMSNet has higher accuracy (12.48% improvement), sensitivity (12.80% improvement), specificity (12.46% improvement), and AUC (14.90% improvement). Analogously, in this 5-year task, the result obtained by the Siamese network with cosine similarity is better than the results obtained by the DenseNet-201, indicating that the imitation of the clinical process (with multiple inputs) is superior to single-visit input. As

the results listed in row 3 and row 4 of Table 3, the Siamese network with cosine similarity works better than the Siamese network with an absolute difference.

3.5 Ablation Studies

In this section, we conduct the ablation study to analyze the effect of the two components (1) using multi-scale features by side output; (2) using multi-structure features with the 2D block integrated into the proposed network. Note that MMSNet will reduce to the traditional Siamese network with cosine similarity by removing the side output and the 2D block. The results obtained by the Siamese network with cosine similarity are listed in the fourth row of Table 2 and Table 3. The 5 and 6 rows list the performance obtained by MMSNet without side output and MMSNet without 2D block, respectively. The results demonstrated that both the side output mechanism and 2D block utilize convolution to measure the similarity between two outputs could boost the performance of MMSNet. The last row lists the results obtained by the proposed method that consists of both side output and 2D block, and the results improve further.

4 Conclusions

In conclusion, this study proposed a new end-to-end deep learning network that simulates the process for automatic POAG prediction within a certain duration from fundus photographs. It is a first attempt to predict the POAG before onset by simulating the glaucoma screening process. The proposed network consists of a 2D block and side output. The 2D block via convolution operation utilizes the 2D features and cosine similarity to measure the similarity between two outputs. The side output drives the hidden layer for favoring discriminative features. One large dataset from multi-institutions was used to evaluate the proposed model. The results demonstrated that the proposed network performs well on POAG prediction before onset.

Acknowledgment.

This work was also supported by awards from the National Eye Institute, the National Center on Minority Health and Health Disparities, National Institutes of Health (grants EY09341, EY09307), Horncrest Foundation, awards to the Department of Ophthalmology and Visual Sciences at Washington University, the NIH Vision Core Grant P30 EY 02687, Merck Research Laboratories, Pfizer, Inc., White House Station, New Jersey, and unrestricted grants from Research to Prevent Blindness, Inc., New York, NY.

References

1. Bourne RR, et al. : Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Glob. Health* 1(6), e339–e349 (2013) [PubMed: 25104599]
2. Chen X, Xu Y, Wong DWK, Wong TY, Liu J: Glaucoma detection based on deep convolutional neural network. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 715–718. IEEE (2015)
3. Ching T, et al. : Opportunities and obstacles for deep learning in biology and medicine. *J. Roy. Soc. Interface* 15(141) (2018). 10.1098/rsif.2017.0387
4. Christopher M, et al. : Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. *Sci Rep.* 8(1), 1–13 (2018) [PubMed: 29311619]
5. Doshi V, Ying-Lai M, Azen SP, Varma R, Los Angeles Latino Eye Study Group, et al. : Sociodemographic, family history, and lifestyle risk factors for open-angle glaucoma and ocular

- hypertension: the Los Angeles Latino Eye Study. *Ophthalmology* 115(4), 639–647 (2008) [PubMed: 17900693]
6. Ehteshami Bejnordi B, et al. : Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318(22), 2199–2210 (2017). 10.1001/jama.2017.14585 [PubMed: 29234806]
 7. Esteva A, et al. : Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639), 115–118 (2017). 10.1038/nature21056 [PubMed: 28117445]
 8. Ghahramani GC, et al. : Multi-task deep learning-based survival analysis on the prognosis of late AMD using the longitudinal data in AREDS. medRxiv (2021)
 9. Han Y, et al.: Using radiomics as prior knowledge for thorax disease classification and localization in chest X-rays. In: AMIA Annual Symposium Proceedings, vol. 2021, p. 546. American Medical Informatics Association (2021) [PubMed: 35308939]
 10. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
 11. Jiang X, Torres M, Varma R, Los Angeles Latino Eye Study Group, et al. : Variation in intraocular pressure and the risk of developing open-angle glaucoma: the Los Angeles Latino Eye Study. *Am. J. Ophthalmol* 188, 51–59 (2018) [PubMed: 29360458]
 12. Kass MA, et al. : The ocular hypertension treatment study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch. Ophthalmol* 120(6), 701–713 (2002) [PubMed: 12049574]
 13. Kingma DP, Ba J: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
 14. Kolomeyer NN, et al. : Lessons learned from 2 large community-based glaucoma screening studies. *J. Glaucoma* 30(10), 875–877 (2021) [PubMed: 34334703]
 15. Li L, Wang X, Xu M, Liu H, Chen X: DeepGF: glaucoma forecast using the sequential fundus images. In: Martel AL, et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 626–635. Springer, Cham (2020). 10.1007/978-3-030-59722-1_60
 16. Li L, et al. : A large-scale database and a CNN model for attention-based glaucoma detection. *IEEE Trans. Med. Imaging* 39(2), 413–424 (2019) [PubMed: 31283476]
 17. Li Z, He Y, Keel S, Meng W, Chang RT, He M: Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 125(8), 1199–1206 (2018) [PubMed: 29506863]
 18. Lin M, Jiang M, Zhao M, Ukwatta E, White JA, Chiu B: Cascaded triplanar autoencoder m-net for fully automatic segmentation of left ventricle myocardial scar from three-dimensional late gadolinium-enhanced MR images. *IEEE J. Biomed. Health Inform* 26(6), 2582–2593 (2022) [PubMed: 35077377]
 19. Lin M, et al. : Fully automated segmentation of brain tumor from multiparametric MRI using 3D context deep supervised U-Net. *Med. Phys* 48, 4365–4374 (2021) [PubMed: 34101845]
 20. Lin M, et al. : Artificial intelligence in tumor subregion analysis based on medical imaging: a review. *J. Appl. Clin. Med. Phys* 22(7), 10–26 (2021) [PubMed: 34164913]
 21. Quigley HA, Katz J, Derick RJ, Gilbert D, Sommer A: An evaluation of optic disc and nerve fiber layer examinations in monitoring progression of early glaucoma damage. *Ophthalmology* 99(1), 19–28 (1992) [PubMed: 1741133]
 22. Sommer A, et al. : Racial differences in the cause-specific prevalence of blindness in east Baltimore. *New Engl. J. Med* 325(20), 1412–1417 (1991) [PubMed: 1922252]
 23. Tatham AJ, Medeiros FA, Zangwill LM, Weinreb RN: Strategies to improve early diagnosis in glaucoma. *Prog. Brain Res* 221, 103–133 (2015) [PubMed: 26518075]
 24. Thakur A, Goldbaum M, Yousefi S: Predicting glaucoma before onset using deep learning. *Ophthalmol. Glaucoma* 3(4), 262–268 (2020) [PubMed: 33012331]
 25. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM: ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462–3471. IEEE (2017). 10.1109/CVPR.2017.369

26. Wanyan T, et al. : Supervised pretraining through contrastive categorical positive samplings to improve COVID-19 mortality prediction. In: Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, pp. 1–9 (2022)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

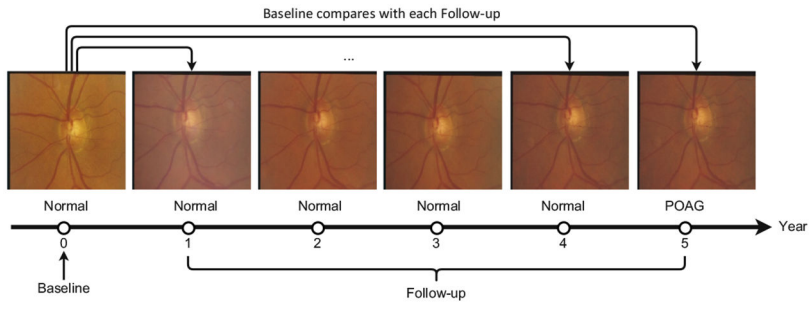


Fig. 1. Longitudinal fundus images of a patient.

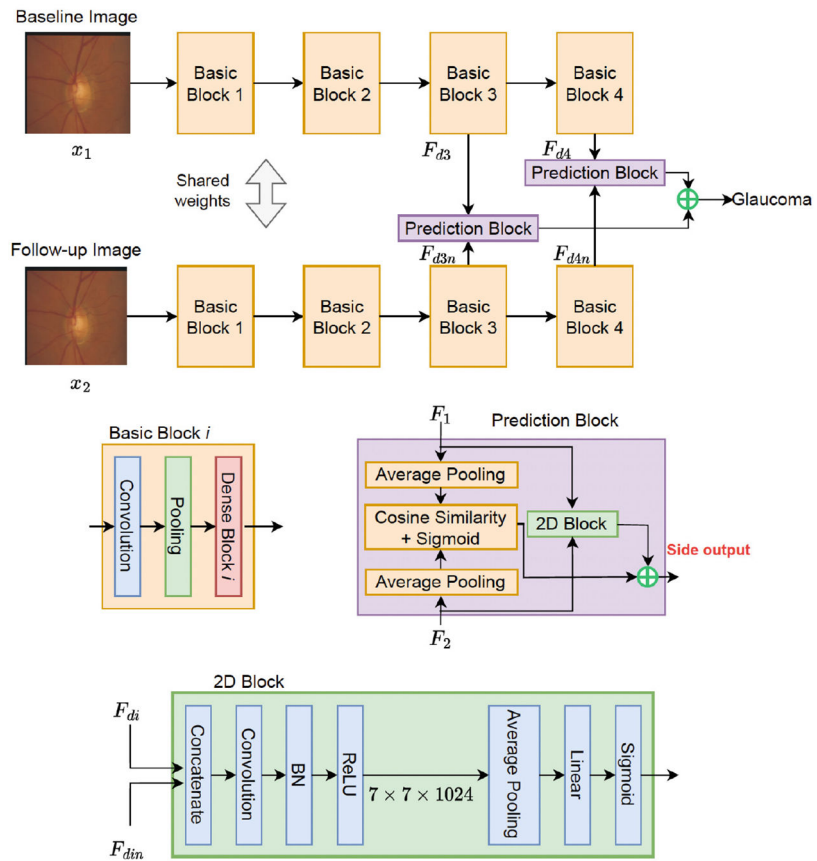


Fig. 2.
The architecture of the proposed MMSNet.

Table 1.

Characteristics of the OHTS.

	Train	Val	Test
Participants	2,503	115	654
2-year prediction			
POAG	463	133	163
Normal	23,315	1,557	6,113
5-year prediction			
POAG	961	284	336
Normal	22,817	1,406	5,940

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Comparisons on the test set of the OHTS dataset for 2-years POAG prediction before onset.

Method	Accuracy	Sensitivity	Specificity	AUC
DenseNet-201	0.6547	0.6667	0.6544	0.7108
MobileV2	0.8368	0.1697	0.8549	0.5114
Siamese network (absolute difference)	0.8634	0.4424	0.8748	0.8089
Siamese network (cosine similarity)	0.9423	0.4121	0.9566	0.8798
MMSNet w/o side output	0.9047	0.7515	0.9089	0.9085
MMSNet w/o 2D block	0.9132	0.6848	0.9193	0.8987
MMSNet	0.9337	0.6485	0.9414	0.9312

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Comparisons on the test set of the OHTS dataset for 5-years POAG prediction before onset.

Method	Accuracy	Sensitivity	Specificity	AUC
DenseNet-201	0.8166	0.625	0.8274	0.8017
MobileV2	0.5915	0.6667	0.5872	0.6799
Siamese network (absolute difference)	0.9312	0.5536	0.9525	0.9109
Siamese network (cosine similarity)	0.9433	0.6250	0.5412	0.9210
MMSNet w/o side output	0.9111	0.7292	0.9219	0.9238
MMSNet w/o 2D block	0.9348	0.6012	0.9535	0.9363
MMSNet	0.9414	0.7530	0.9520	0.9507

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript