Structurally Guided Channel Attention Networks: SGCA-Net

Veysi Yildiz
Cognitive Systems Laboratory, ECE Department, Northeastern University
Boston, MA, USA

Jennifer Dy
ECE Department, Northeastern University
Boston, MA, USA

J. Peter Campbell
Department of Ophthalmology, Oregon Health & Science University
Portland, Oregon, USA

Susan Ostmo
Department of Ophthalmology, Oregon Health & Science University
Portland, Oregon, USA

Michael F. Chiang
Department of Ophthalmology, Oregon Health & Science University
Portland, Oregon, USA

Stratis Ioannidis
ECE Department, Northeastern University
Boston, MA, USA

Deniz Erdoğmuş
Cognitive Systems Laboratory, ECE Department, Northeastern University
Boston, MA, USA

ABSTRACT

In this paper, we propose Structurally Guided Channel Attention Networks (SGCA-Net), a principled way to guide the channel attention of CNNs. Convolution operator constructs features maps by using both channel and spatial information within the receptive fields of its filters. Prior research has investigated the impact of strengthening the representational power of CNNs using channel attention modules. In this work, we guide the channel attention of networks using feature vectors that contain clinically relevant information. We do so by attaching guided attention modules into a state-of-the-art network architecture, and guiding these attention modules with domain knowledge using feature vectors. Experiments on a dataset of 5512 posterior retinal images, taken using a wide angle fundus camera, show that SGCA-Net achieves 0.983 and 0.948 AUC to predict plus and normal categories, respectively. SGCA-Net achieves higher performance than CNNs without attention modules and CNNs with unguided attention modules.

KEYWORDS

neural networks, attention, CNNs, channel attention, ROP

ACM Reference Format:
Veysi Yildiz, Jennifer Dy, J. Peter Campbell, Susan Ostmo, Michael F. Chiang, Stratis Ioannidis, and Deniz Erdoğmuş. 2021. Structurally Guided Channel Attention Networks: SGCA-Net. In *The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021), June 29-July 2, 2021, Corfu, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3453892.3454006

PETRA 2021, June 29-July 2, 2021, Corfu, Greece

ACM ISBN 978-1-4503-8792-7/21/06…$15.00
https://doi.org/10.1145/3453892.3454006

1 INTRODUCTION

Retinopathy of prematurity (ROP) is a disease that affects premature infants and is a leading cause of childhood blindness [8]. Around 14,000-16,000 premature born infants are affected by of ROP in the U.S. each year [13]. An international classification system to standardize ROP diagnosis was developed in 2005 [6]. Plus disease is the most important disease feature in determining the need for treatment, and it is defined as abnormal tortuosity and dilation of the posterior retinal blood vessels. An important factor for treatment planing is the correct classification of three levels of plus disease, *normal*, *pre-plus*, and *plus*. Early detection and treatment of plus disease plays an important role to avoid impaired vision or blindness [13]. As the survival rate of prematurely born babies increases, the number of infants at risk of ROP also increases [7]. Lack of access to ROP experts remains a challenge. These factors bring about a need for an accurate ROP detection system.

Neural networks have made high impact on many medical tasks [5, 11, 16, 17], including detection of ROP from fundus images [3, 24]. State-of-the-art ROP detection systems employ convolutional neural networks (CNNs) [3] and achieve up to 0.947 and 0.982 area under the ROC curve (AUC) in the discrimination of *normal* and *plus* levels of ROP. Recent studies show that incorporating attention in network architecture improves CNNs classification performance, including ROP detection [22, 27].

Attention in neural networks is a set of additional operations that generate a soft mask to weigh the outputs of intermediate layers. It selectively emphasizes informative features and suppress less useful ones. Attention masks in CNNs can be learned to weigh the spatial information [9, 22, 27] or the channels of layer outputs [12, 26]. When used for channel attention, network generates attention vectors to suppress channels of the layer outputs that are irrelevant for target classification [12, 23]. Similarly, in an actual diagnosis process, clinicians would only focus on relevant features while diagnosing a disease. For example, in ROP diagnosis, ophthalmologists focus on abnormal tortuosity and dilation of the posterior retinal blood vessels [6].

Clinicians' focus on relevant features is based on their knowledge of the disease. We believe that such structural domain knowledge can be leveraged to enhance CNN performance. In standard attention architectures, network tries to learn attention masks while only supervised with the class labels. Although attention in CNNs has been used for mimicing what clinicians do while diagnosing, most of the current attention architectures do not incorporate structural domain knowledge in training. Several studies [14, 19, 25] use spatial attention to guide the network attention to clinically important regions using manually annotated segmentation of classification target, however, they do not suggest a principal way of incorporating domain knowledge in guidance of network attention.

Yildiz et al. [27] show that incorporating structural domain knowledge in guiding spatial attention increases ROP detection performance. They attach spatial attention modules into the state-of-the-art ROP detection network [3], and improve its performance up to 3% [27]. They guide the attention of the network to the regions of an image that contains clinically relevant information. Our work extends Yildiz et al. [27] by using guidance in channel attention. Specifically, we depart from Yildiz et al. by guiding network attention for weighing the importance of feature map channels at the output of convolutional layers instead of the spatial information within feature maps. We do so by using domain knowledge-based feature vectors during network training. Several studies extract structural features dedicated to diagnosis of ROP [1, 15, 21, 28]. Closest to us, Yildiz et al. [28] achieve 0.94 AUC in detection of ROP by extracting features related to tortuosity and dilation of posterior blood vessels. We leverage features extracted by Yildiz et al. [28] in guiding channel attention of ROP detection network.

In this paper, we propose structurally guided channel attention networks (SGCA-Net). We present a novel way to incorporate domain knowledge in guiding channel attention of CNNs. We do so by guiding the network using features dedicated to classification task. By achieving 0.948 and 0.983 AUC in detection of *normal* and *plus* ROP levels, we achieve higher performance than baseline ROP classification network and baseline channel attention network.

2 MOTIVATION

Accurate classification of three levels of plus disease, which is characterized by abnormal tortuosity and dilation in the posterior retinal blood vessels plays an important role for treatment planing. We present two sample fundus images from *normal* (top) and *plus* (bottom) classes in the left column of Fig. 1. Also, the right column of Fig. 1 presents images in which blood vessels are segmented via the procedure proposed by Brown et al. [2]. In addition to methods which detect ROP by quantifying structural featuers such as dilation and tortuosity [1, 28], Brown et al. [3] employ CNNs. Our motivation in this study is to improve the classification performance of ROP detection CNN by incorporating structural features in the training process of CNNs using attention mechanisms.

3 METHOD

3.1 Channel Attention

A convolutional layer in a CNN convolves its entire input with multiple filters. It generates feature maps at the output of the layer after summation of convolution channels. This process uses channel and spatial information only within the receptive field of the filters. While channel dependencies are implicitly embedded in convolutional layers, they are entangled with the spatial correlations.

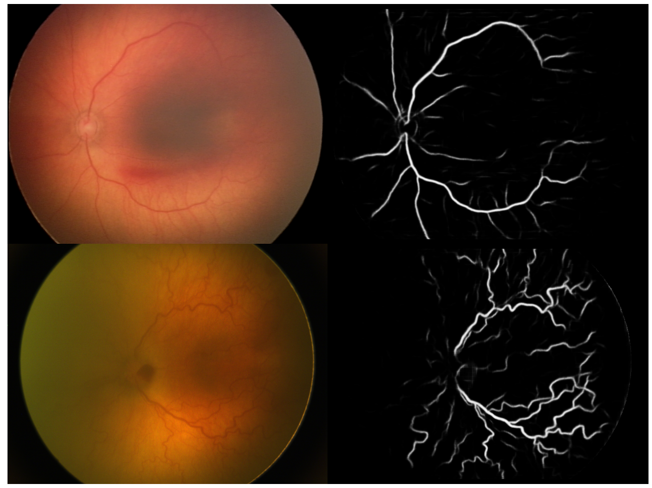

Figure 1: Sample fundus images (left) from *normal* (top) and *plus* (bottom) classes with vessel segmentations (right).

Channel attention strengthens the representational power of a CNN by learning to use global information in feature maps to selectively emphasise informative feature map channels and suppress less useful ones [12]. Hu et al. [12] use Squeeze-Excitation blocks in channel attention networks and, improve networks classification performance. Yildiz et al. [27] show that guiding the spatial attention of network with features based on domain knowledge during the attention process improves the network classification performance. We believe that exploiting the impact of guiding network attention with domain knowledge will also increase the performance of channel attention networks.

3.2 Problem Definition

Given a dataset containing N images, indexed by $i \in \{1, 2, \ldots, N\}$, every image i is represented as $\mathbf{X}_i \in \mathbb{R}^{H \times W}$ where H and W are the height and width of the image, respectively. For each image $\mathbf{X}_i$, a labeler generates a label $y_i \in \{$*normal, pre-plus, plus*$\}$, which indicates the ROP level. Let feature vector $\mathbf{m}_i^f \in [0, 1]^d$ be the domain knowledge based features of $\mathbf{X}_i$. **Our goal is to learn CNNs that perform classification well, and simultaneously learn to emphasise informative feature map channels as guided by domain knowledge based features $\mathbf{m}_i^f$.**

As shown in Fig. 2, attention modules are effectively additional layers generating attention vectors. Let $j \in \{1, .., n\}$ be the index of the j-th attention module parameterized by θ_j, and $\mathbf{F}_{j,i}, \mathbf{F}_{j,i}^M \in \mathbb{R}^{C \times H_j \times W_j}$ be the feature maps of $\mathbf{X}_i$ before and after j-th attention mask is applied, respectively. $\mathbf{F}_{j-1,i}^M$ and attention vector $\mathbf{m}_i^{\theta_j} \in [0, 1]^C$ are the input-output pair of the j-th attention module.

3.3 Squeeze-Excitation Networks

Hu et al. [12] use squeeze-excitation blocks for applying channel attention in CNNs. Squeeze-excitation blocks use global information in feature maps and alter their importance in three steps: (1) squeeze, (2) excitation and (3) scaling.

Squeeze: Squeeze operation finds global information embedding of feature maps in convolutional layers. It squeezes the spatial information in each feature map channel into a channel descriptor $\mathbf{z} \in \mathbb{R}^C$.

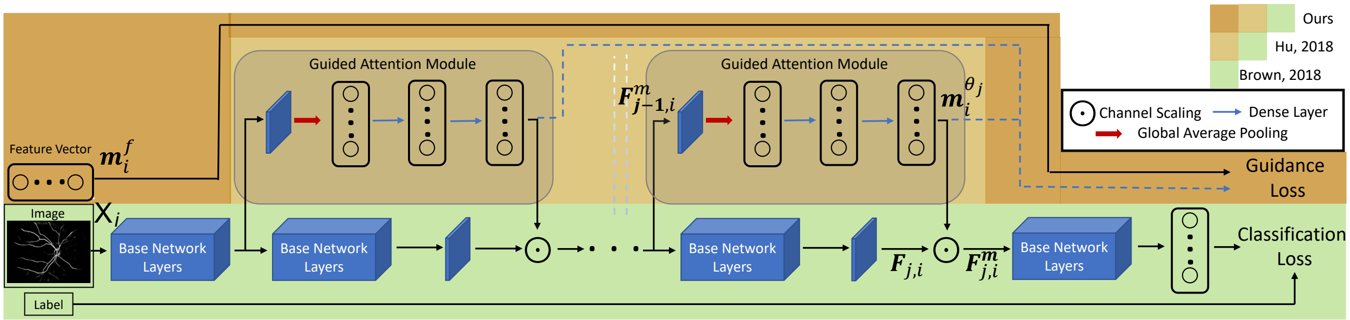

Figure 2: Proposed architecture of Structurally Guided Channel Attention Networks, SGCA-Net

They obtain channel descriptors by applying global average pooling to feature maps. Let $\mathbf{F}_i \in \mathbb{R}^{C\times H\times W}$ be the feature maps of image $\mathbf{X}_i$, channel descriptor $\mathbf{z}_i$ is find by $\mathbf{z}_{i,c} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\mathbf{F}_{i,c}(h,w)$, where $c \in \{1,..,C\}$ is the index of the feature channel.

Excitation: After global information in feature map channels are squeezed into channel descriptors, excitation operation aims to capture the channel-wise dependencies.

Hu et al. [12] uses two fully connected layers as excitation operation. The first layer has C/r nodes where r is defined as a reduction ratio. This layer is followed by a RELU activation function. The second layer has C nodes followed by a sigmoid function. The dense layers at jth attention block are parametrized with θ_j. Excitation function generates an attention vector $\mathbf{m}_i^{\theta_j} \in \mathbb{R}^C$.

Scaling: The attention vector $\mathbf{m}_i^{\theta_j}$ is used for scaling the feature map channels following $\mathbf{F}_{j,i,c}^m = \mathbf{m}_{i,c}^{\theta_j}\mathbf{F}_{j,i,c}$, where $\mathbf{m}_{i,c}^{\theta_j}$ is the cth element of attention vector $\mathbf{m}_i^{\theta_j}$, $\mathbf{F}_{i,j,c}$ is the cth channel of feature maps $\mathbf{F}_{i,j}$, and $\mathbf{F}_{i,j}^m$ are the attention applied channel maps.

Traditionally, attention module and base network layers are trained jointly by minimizing categorical cross entropy loss $\mathcal{L}_C$:

$$\mathcal{L}(\mathbf{X}_i, y_i) = \mathcal{L}_C(\mathbf{X}_i, y_i) = -\log\left(e^{s_{p_i}(\mathbf{X}_i,\Theta)}/\sum_l e^{s_l(\mathbf{X}_i,\Theta)}\right), \quad (1)$$

where p_i is the index of the true class for image i, $s_l(\mathbf{X}_i, \Theta)$ is the score of class $l \in \{1, 2, ..., L\}$ produced by the model, parameterized by Θ, prior to the soft-max layer that is explicitly indicated above, and L is the number of classes.

3.4 Guiding Channel Attention

We incorporate domain knowledge in guiding channel attention networks by introducing an additive term in loss function. Our loss function consists of two terms: (a) classification loss $\mathcal{L}_C(\mathbf{X}_i, y_i)$ and (b) guidance loss $\mathcal{L}_G(\mathbf{m}_i, \mathbf{m}_i^f)$. We use categorical cross entropy loss in Eq. (1) for $\mathcal{L}_C(\mathbf{X}_i, y_i)$. Given image $\mathbf{X}_i$, we generate a feature vector that contains important features $\mathbf{m}_i^f \in [0,1]^d$, as described in Section 3.5. We use these to guide the attention vectors generated by the network $\mathbf{m}_i^{\theta_j} \in [0,1]^C$, via the guidance loss, we define as:

$$\mathcal{L}_G(\mathbf{m}_i^{\theta}, \mathbf{m}_i^f) = \sum_j \frac{1}{d}\sum_{c=1}^{d}(\mathbf{m}_{i,c}^f - \mathbf{m}_{i,c}^{\theta_j})^2, \quad (2)$$

where the mean squared error (MSE) is used. Note that when calculating MSE, we use the first d elements of $\mathbf{m}_i^{\theta_j}$. Our proposed SGCA-Net loss function is defined as:

$$\mathcal{L}(\mathbf{X}_i, y_i, \mathbf{m}_i^{\theta}, \mathbf{m}_i^f) = \mathcal{L}_C(\mathbf{X}_i, y_i) + \lambda\mathcal{L}_G(\mathbf{m}_i^{\theta}, \mathbf{m}_i^f), \quad (3)$$

where λ is a trade-off control parameter between classification and guidance loss.

3.5 Extracting ROP Features

Vessel dilation and tortuosity are commonly used for the definition of ROP [6]. Following the pipeline from Yildiz et al. [28], we compute 143 features related to vessel tortuosity and dilation. For every image $\mathbf{X}_i$, we generate a feature vector, $\mathbf{m}_i^f \in [0,1]^{143}$.

4 EXPERIMENTAL SETUP

Dataset. Our dataset contains 5512 retinal fundus images. According to its disease level, clinicians assign a label (as plus, pre-plus or normal) to each image following a reference standard diagnosis [18]. The dataset contains 163 plus, 802 pre-plus, and 4547 normal images. We use images in which vessels are segmented via the procedure proposed by Brown et al. [3].

4.1 Evaluation Metrics

We binarize labels as (a) plus vs. other classes (PvO), and (b) normal vs. other classes (NvO). We calculate the Area Under the ROC Curve (AUC), accuracy (ACC), F1 score (F1) and Area Under the Precision-Recall Curve (PRAUC) scores with five fold cross-validation. We present the mean of five folds and calculate the 95% confidence intervals as $1.96 \times \sigma_A$, where σ_A is the standard deviation, we calculate following Hanley et al. [10].

4.2 Base CNN Architecture and Training

As the base network to attach attention modules, we employ Inception v.1 architecture [20], which has shown great performance in many classification tasks including ROP [3]. We initialize the network weights with pretrained weights on ImageNet [4]. We employ stochastic gradient descent with learning rate 0.0001 for 100 epochs to optimize the network weights.

4.3 Competing Methods

We explore the effects of guiding the channel attention in classification of ROP in 3 different setups. When used, we attach an attention module to *inception 5a* block.

No Attention [3]: We train the base CNN architecture without any attention as in Brown et al. [3]'s ROP classification model.

Unguided Attention [12]: We train the baseline channel attention network as explained in Section 3.3.

Guided Attention (SGCA-Net): We use the feature vectors for guiding the network attention as explained in the Section 3.4. Also, we study the effect of trade-off control parameter λ in Eq. (3) by repeating the same experiment with λ values ranging in $[0, 50]$.

5 RESULTS AND DISCUSSION

Table 1: Cross validation results for competing methods

	Task	No Attention [3]	Unguided Attention [12]	Guided with Feature Vector
AUC	NvsO	0.947(0.024)	0.944(0.025)	**0.948(0.024)**
	PvsO	0.982(0.006)	**0.983(0.006)**	**0.983(0.006)**
ACC	NvsO	0.872(0.001)	0.872(0.001)	**0.876(0.001)**
	PvsO	0.916(0.0)	**0.945(0.0)**	0.943(0.0)
F1	NvsO	0.917(0.001)	0.917(0.001)	**0.92(0.001)**
	PvsO	0.954(0.0)	**0.971(0.0)**	**0.971(0.0)**
PR-AUC	NvsO	0.832(0.039)	**0.838(0.039)**	0.835(0.039)
	PvsO	0.726(0.019)	0.699(0.019)	**0.729(0.019)**

We present the cross validation results of predicting normal vs other and plus vs other classes in Table 1. SGCA-Net achieves 0.948 and 0.983 AUC in predicting normal vs other classes and plus vs other classes, respectively. It achieves higher AUC than unguided channel attention [12] in predicting normal vs other classes. Also, SGCA-Net achieves higher scores than baseline ROP detection network [3] in all metrics. In predicting normal vs other classes, unguided network achieves lower AUC than baseline network, whereas, SGCA-Net avoids the performance drop by incorporating guidance in the attention. The results suggest that incorporating guidance in channel attention increases the network performance.

6 CONCLUSION

In this paper, we propose a novel way to incorporate domain knowledge in guiding channel attention in CNNs. We attach attention modules into state-of-the-art network architecture and guide the network by using feature vectors created with domain knowledge. We do so with an additional guidance loss in network training. Experiments show that guiding the network attention using domain knowledge increases performance of ROP detection networks. We believe that performance of SGCA-Net can be further improved with different architectures in attention module. Also, SGCA-Net guides the network attention in a predefined set of feature channels, for example in this paper, we guided first 143 channels of *inception 5a* block. It would be interesting to select which feature channels to guide based on their activations.

ACKNOWLEDGMENTS

Our work is supported by NIH (R01EY019474), NSF (SCH-1622542 at MGH, SCH-1622536 at Northeastern, SCH-1622679 at OHSU), a Facebook Research Award, and by unrestricted departmental funding from Research to Prevent Blindness (OHSU).

REFERENCES

[1] E Ataer-Cansizoglu, V Bolon-Canedo, J P Campbell, A Bozkurt, D Erdogmus, J Kalpathy-Cramer, S Patel, K Jonas, RV P Chan, S Ostmo, et al. 2015. Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the "i-ROP" system and image features associated with expert diagnosis. *TVST* 4, 6 (2015).

[2] J M Brown, J P Campbell, A Beers, K Chang, K Donohue, S Ostmo, RV P Chan, J Dy, D Erdogmus, S Ioannidis, et al. 2018. Fully automated disease severity assessment and treatment monitoring in retinopathy of prematurity using deep learning. In *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, Vol. 10579. International Society for Optics and Photonics.

[3] J M Brown, J P Campbell, A Beers, K Chang, S Ostmo, RV P Chan, J Dy, D Erdogmus, S Ioannidis, J Kalpathy-Cramer, et al. 2018. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA ophthalmology* 136, 7 (2018).

[4] J Deng, W Dong, R Socher, L Li, K Li, and L Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *CVPR*. IEEE.

[5] A Esteva, B Kuprel, R A Novoa, J Ko, S M Swetter, H M Blau, and S Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017).

[6] International Committee for the Classification of Retinopathy of Prematurity et al. 2005. The international classification of retinopathy of prematurity revisited. *Archives of ophthalmology (Chicago, Ill.: 1960)* 123, 7 (2005).

[7] C Gilbert, A Fielder, L Gordillo, G Quinn, R Semiglia, P Visintin, A Zin, et al. 2005. Characteristics of infants with severe retinopathy of prematurity in countries with low, moderate, and high levels of development: implications for screening programs. *Pediatrics* 115, 5 (2005).

[8] C Gilbert and A Foster. 2001. Childhood blindness in the context of VISION 2020: the right to sight. *Bulletin of the WHO* 79 (2001).

[9] Q Guan, Y Huang, Z Zhong, Z Zheng, L Zheng, and Y Yang. 2018. Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. *arXiv preprint arXiv:1801.09927* (2018).

[10] J A Hanley and B J McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 1 (1982).

[11] L Hou, D Samaras, T M Kurc, Y Gao, J E Davis, and J H Saltz. 2016. Patch-based convolutional neural network for whole slide tissue image classification. In *CVPR*.

[12] J Hu, L Shen, and G Sun. 2018. Squeeze-and-excitation networks. In *CVPR*.

[13] National Eye Institute. 2019. Retinopathy of Prematurity. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/retinopathy-prematurity.

[14] K Li, Z Wu, K Peng, J Ernst, and Y Fu. 2018. Tell me where to look: Guided attention inference network. In *CVPR*.

[15] Elias Khalili Pour, Hamidreza Pourreza, Kambiz Ameli Zamani, Alireza Mahmoudi, Arash Mir Mohammad Sadeghi, Mahla Shadravan, Reza Karkhaneh, Ramak Rouhi Pour, and Mohammad Riazi Esfahani. 2017. Retinopathy of prematurity-assist: novel software for detecting plus disease. *Korean Journal of Ophthalmology* 31, 6 (2017), 524–532.

[16] P Rajpurkar, J Irvin, K Zhu, B Yang, H Mehta, T Duan, D Ding, A Bagul, C Langlotz, K Shpanskaya, et al. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[17] O Ronneberger, P Fischer, and T Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*. Springer.

[18] M C Ryan, S Ostmo, K Jonas, A Berrocal, K Drenser, J Horowitz, T C Lee, C Simmons, M Martinez-Castellanos, RV P Chan, et al. 2014. Development and evaluation of reference standards for image-based telemedicine diagnosis and clinical research studies in ophthalmology. In *AMIA annual symposium*, Vol. 2014. American Medical Informatics Association.

[19] J Son, W Bae, S Kim, S Park, and K Jung. 2018. Classification of findings with localized lesions in fundoscopic images using a regionally guided cnn. In *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer.

[20] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich. 2015. Going deeper with convolutions. In *CVPR*.

[21] D K Wallace, J Jomier, S R Aylward, and M Landers III. 2003. Computer-automated quantification of plus disease in retinopathy of prematurity. *Journal of American Association for Pediatric Ophthalmology and Strabismus* 7, 2 (2003).

[22] F Wang, M Jiang, C Qian, S Yang, C Li, H Zhang, X Wang, and X Tang. 2017. Residual attention network for image classification. In *CVPR*.

[23] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.

[24] D E Worrall, C M Wilson, and G J Brostow. 2016. Automated retinopathy of prematurity case detection with convolutional neural networks. In *Deep Learning and Data Labeling for Medical Applications*. Springer.

[25] H Yang, J Kim, H Kim, and S P Adhikari. 2019. Guided soft attention network for classification of breast cancer histopathology images. *IEEE TMI* (2019).

[26] X Yao, D She, S Zhao, J Liang, Y Lai, and J Yang. 2019. Attention-aware polarity sensitive embedding for affective image retrieval. In *ICCV*.

[27] V Yildiz, S Ioannidis, I Yildiz, P Tian, P Campbell, S Ostmo, J Kalpathy-Cramer, M Chiang, , D Erdogmus, and J Dy. forthcoming. Structural Visual Guidance Attention Networks in Retinopathy of Prematurity. In *ISBI*.

[28] V M Yildiz, P Tian, I Yildiz, J M Brown, J Kalpathy-Cramer, J Dy, S Ioannidis, D Erdogmus, S Ostmo, S J Kim, et al. 2020. Plus Disease in Retinopathy of Prematurity: Convolutional Neural Network Performance Using a Combined Neural Network and Feature Extraction Approach. *TVST* 9, 2 (2020).