BIRA-NET: BILINEAR ATTENTION NET FOR DIABETIC RETINOPATHY GRADING

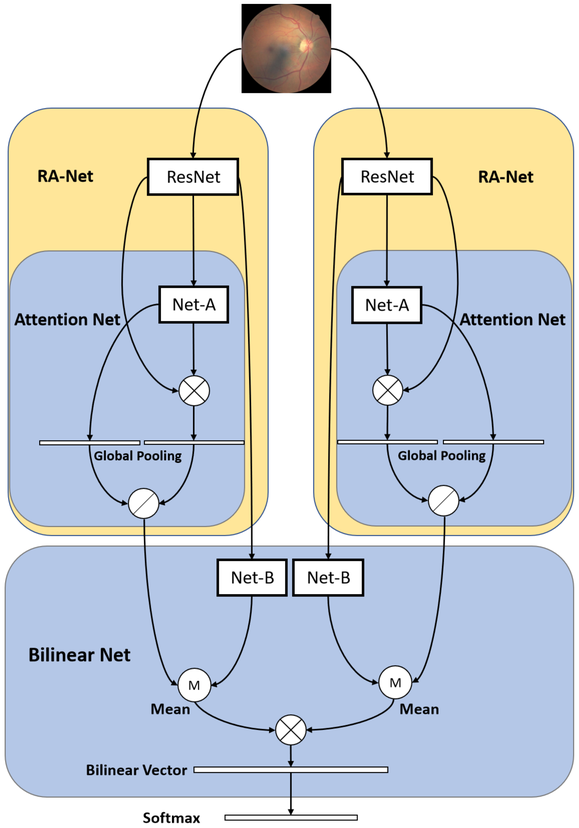
Ziyuan Zhao, Kerui Zhang*, Xuejie Hao, Jing Tian, Matthew Chin Heng Chua, Li Chen, Xin Xu*

*(*Equal Contribution)*

NUS & UESTC & WUST

Diabetic retinopathy (DR) is a common retinal disease that leads to blindness. For diagnosis purposes, DR image grading aims to provide automatic DR grade classification, which is not addressed in conventional research methods of binary DR image classification. Small objects in the eye images, like lesions and microaneurysms, are essential to DR grading in medical imaging, but they could easily be influenced by other objects.

To address these challenges, we propose a new deep learning architecture, called **BiRA-Net**, which combines the attention model for feature extraction and bilinear model for fine-grained classification.

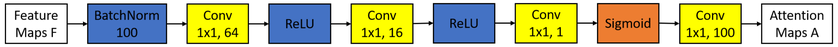


ResNet

- Shortcut connection lets some input skip the layer indiscriminately, which would avoid adding on new parameters and having too much calculation on the network.
- Avoid the loss of information and degradation problem.
- Saliently increase the training speed and effects

Attention Net

- Input: $F \in \mathbb{R}^{100\times20\times20}$ (feature maps from ResNet)
- Net-A: $A \in \mathbb{R}^{100\times20\times20}$ (attention maps from Net-A)



- Output: $Output = GAP(A^l) \oslash GAP(A^l \otimes F^l)$

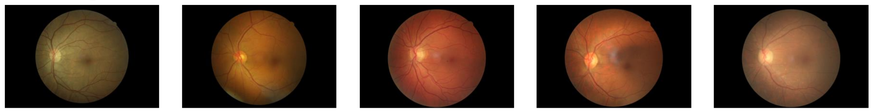
A^l and F^l are l-th attention map and l-th feature map.
$\otimes$ and $\oslash$ denote element-wise multiplication and element-wise division.

Bilinear Net

- Two same streams of **RA-Net** are trained simultaneously
- Take the output of **Attention Net** and the output of the **ResNet** as inputs.
- The resulting bilinear vector $B(B = ZZ^T)$ is passed through signed square-root step ($Y \longleftarrow sign(B)\sqrt{|B|}$) and l_2 normalization ($Z \longleftarrow Y/\|Y\|_2$) to improve the performance.

Dataset and implementation

- The retinal images are provided by EyePACS consisting of 35126 images.



- **Cropped** to keep the whole retina regions in the square areas.
- **Resized** to 610 × 610 pixels.
- **Standardized** by subtracting mean and dividing by standard deviation that is computed over all pixels in all training images.
- The **histogram equalization** is used for contrast enhancement.

Baseline Methods

- **Bi-ResNet:** A pre-trained ResNet-50 using the bilinear strategy.
- **RA-Net**: Only one single stream of the proposed **BiRA-Net** is used.
- **BiRA-Net**: The proposed architecture.

Results

- BiRA-Net outperforms all other methods in ACA, Marco F1 and Micro F1.

	ACA	Marco F1	Micro F1
Bravo *et al.* [16]	0.5051	0.5081	0.5052
ResNet-50 [19]	0.4689	0.4753	0.4689
Bi-ResNet	0.4889	0.5503	0.4897
RA-Net	0.4717	0.5268	0.4724
BiRA-Net	**0.5431**	**0.5725**	**0.5436**

- In the confusion matrix, each class is most likely to be predicted into the right class, except class 1, which is mostly classified into class 0. It is clear that class 1 is the most difficult to differentiate and normal (class 0) is the easiest to detect.