GANet: Group attention network for diabetic retinopathy image segmentation

Ye, Lei, Zhu, Weifang, Feng, Shuanglang, Chen, Xinjian

Lei Ye, Weifang Zhu, Shuanglang Feng, Xinjian Chen, "GANet: Group attention network for diabetic retinopathy image segmentation," Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 1131307 (10 March 2020); doi: 10.1117/12.2548310

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

GANet: Group Attention Network for Diabetic Retinopathy Image Segmentation

Lei Ye [1,#], Weifang Zhu [1#], Shuanglang Feng [1], Xinjian Chen [1,2*]

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, 215006, China
[2]State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, 215123, China

ABSTRACT

The assistance of deep learning techniques for clinic doctors in disease analysis, diagnosis and treatment is becoming popular and popular. In this paper, we propose a U-shape architecture based Group Attention network (named as GANet) for symptom segmentation in fundus images with diabetic retinopathy, in which Channel Group Attention(CGA) module and Spatial Group Attention Upsampling (SGAU) module are designed. The CGA module can adaptively allocate resources based on the importance of the feature channels, which can enhance the flexibility of the network to handle different types of information. The original U-Net directly merges the high-level features and low-level features in decoder stage for semantic segmentation, and achieves good results. To increase the nonlinearity of the U-shape network and pay more attention to the lesion area, we propose a Spatial Group Attention Upsampling (SGAU) module. In summary, our main contributions include two aspects: (1) Based on the U-shape network, the CGA module and SGAU module are designed and applied, which can adaptively allocate the weight of channels and pay more attention to the lesion area, respectively. (2) Compared with the original U-Net, the Dice coefficients of the proposed network improves by nearly 2.96% for hard exudates segmentation and 2.89% for hemorrhage segmentation, respectively.

KEYWORDS: Deep Learning, Diabetic Retinopathy, Segmentation, Group Attention Network

1. INTRODUCTION

Diabetic retinopathy(DR) is a chronic complication characterized by retinal ischemia, which is irreversible and has become one of the four major blinding diseases. According to statistics, there are about 425 million diabetics worldwide, and this number is still soaring. It is expected that this number will reach 592 million by 2035[1]. The symptoms of diabetic retinopathy include microaneurysms, hard exudation, hemorrhage, etc. Hemorrhage is caused by retinal vascular obstruction, which is shown as a red spot in fundus image. The hard exudate in the fundus image presents as yellowish speckles[2]. Surveys show that about one-third of diabetics will develop DR. Grading of DR is important for the diagnosis and treatment of DR. Retinal photograph are widely used in the early screening and diagnosis of several eye diseases, such as DR, glaucoma, etc. Hard exudates present in the early stage of DR, while retinal hemorrhage is a relatively late symptom of DR. Accurate segmentation of hard exudates and hemorrhage is a key step for DR grading.

In recent years, we have witnessed the development of deep learning in medical image processing, FCN[3] is the first proposed end-to-end image segmentation technology, from which the U-shape network structure received more and more concerns. PAN[4] used feature pyramid attention model and channel attention model to improve segmentation performance. Xiang Li et al [5] proposed the SGE (Spatial Group-wise Enhance) module to highlight the features in the correct semantic region and suppress the ones in the unrelated region. Receptive field plays an important role in image segmentation. Large receptive field can make segmentation more accurate. Instead of stacking convolution layers to enlarge the receptive field, Wang X et al [6] proposed the non-local module to capture the long-range dependencies, which computed attention map on the whole graph. Unlike non-local mechanism, EMANet[7] obtained bases through the expectation-maximization (EM) algorithm, and ran the attention mechanism on this bases, which greatly reduced the computational complexity.

*Corresponding author: E-mail: xjchen@suda.edu.cn, # indicates these authors contributed equally to this work

Medical Imaging 2020: Image Processing, edited by Ivana Išgum, Bennett A. Landman,
Proc. of SPIE Vol. 11313, 1131307 · © 2020 SPIE · CCC
code: 1605-7422/20/$21 · doi: 10.1117/12.2548310

In this paper, we purposefully improve the U-shape network with Channel Group Attention(CGA) module and Spatial Group Attention Upsampling (SGAU) module for accurate segmentation of hard exudates and hemorrhage in DR fundus images.

2. METHODS

In this section the proposed method will be introduced including the three parts: overall architecture, channel group attention module(CGA), and spatial group attention upsampling (SGAU) module.

2.1 Overall Architecture

Inspired by previous work, based on U-shape network, we propose a Group Attention network (named as GANet) for symptom segmentation in DR fundus images, in which CGA module and SGAU module are designed and inserted. The CGA module focuses on feature channel selection, and the SGAU module focuses on enhancing the expression of lesion features. The overall structure of GANet is shown in Fig 1.

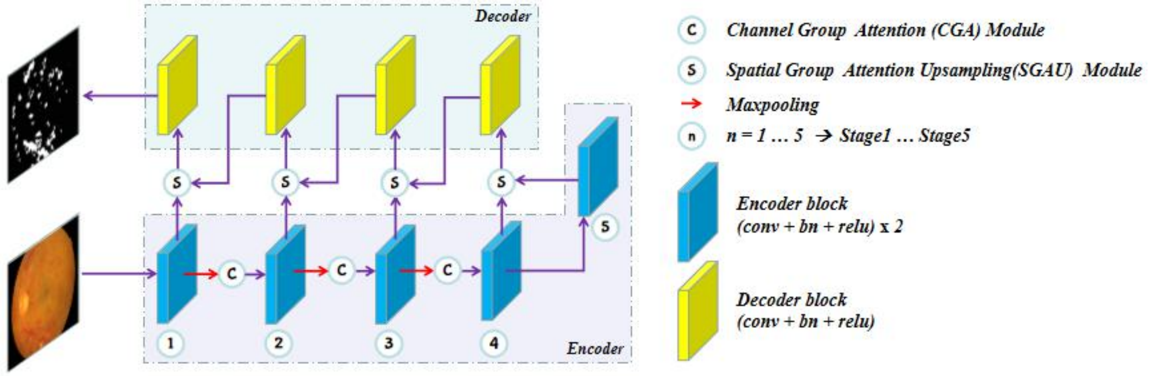
Decoder
Encoder
S
C
1
2
3
4
5
C Channel Group Attention (CGA) Module
S Spatial Group Attention Upsampling(SGAU) Module
Maxpooling
n n = 1 … 5 → Stage1 … Stage5
Encoder block
(conv + bn + relu) x 2
Decoder block
(conv + bn + relu)

Fig 1. Overview of proposed GANet for semantic segmentation. The number of output channels in the first stage is 32, and the number of channels in each subsequent stage is twice that of the previous stage.

2.2 Channel Group Attention(CGA) Module

It has been first proved in SENet [8] that the segmentation performance of the network can be improved significantly by considering the dependencies between feature channels. Inspired by SENet, we propose the CGA module, as shown in Fig 2. First, C channels are divided into m equal groups. Then (C/m) feature maps in each group are summed in the channel direction and normalized, and m feature maps(m*H*W) are generated. Second, we normalize and concatenate the obtained m feature maps and get the channel weight after global pooling and full-connected operation.

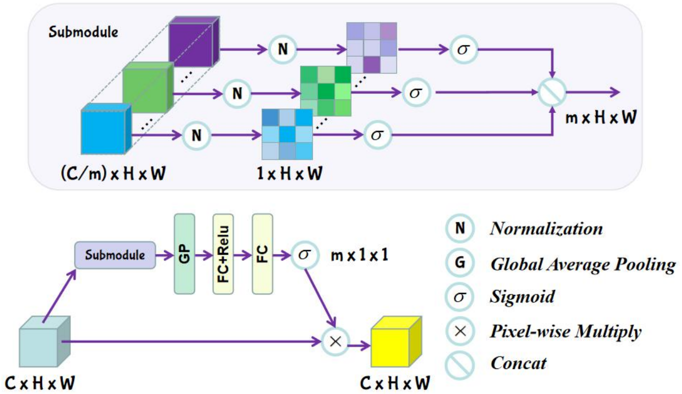

Fig 2. Components of the Channel Group Attention(CGA) Module. After passing through the CGA module, the features which contains the correct semantics will be automatically weighted.

The CGA module structure can be expressed as:

$$\text{Submodule} = \overset{j=32}{\underset{j=1}{Cat}}(\sigma(Norm(Div(F^{(n)})))) \tag{1}$$

$$\text{Y} = \sigma(Conv(GP(\text{Submodule};\theta))) \otimes F^{(n)} \tag{2}$$

Where **Submodule** denotes the feature grouping stage in CGA module and **Y** represents the module output. F^n means features in stage **n**, Div represents feature grouping. **Cat** represents the concatenation operation, and θ represents the model parameter.

Compared with SENet, the CGA module proposed by this paper has better performance in Dice, because the semantics expressed by feature map after channel fusion are more accurate than those expressed by single feature map. The experiment results are shown in Table 1.

2.3 Spatial Group Attention Upsampling (SGAU) Module

To increase the nonlinearity and pay more attention to the lesion region, the Spatial Group Attention Upsampling (SGAU) module is designed and inserted as the skip-connections (shown in Fig.3), which is inspired by Attention U-Net[6]. Take stage 4 and stage 5 as an example. First, like the CGA module, we divide the feature maps of stage 4 (Low-level features) and stage 5 (High-level features) into **m** groups respectively, as shown in Fig 3(a). To enhance the differences between foreground and background, a self-attention operation is done on each group of features in high-level. Second, the self-attention feature map is summed in the channel direction and normalized to obtain a feature map, which is the input of the sigmoid layer. Third, upsample and multiply the resulted spatial attention map with a group of features in stage 4. We also introduce a residual structure to help the network optimize. The output feature of SGAU modules is shown in Fig. 4(d), from which it can be clearly seen that the foreground is enhanced and the background is suppressed.

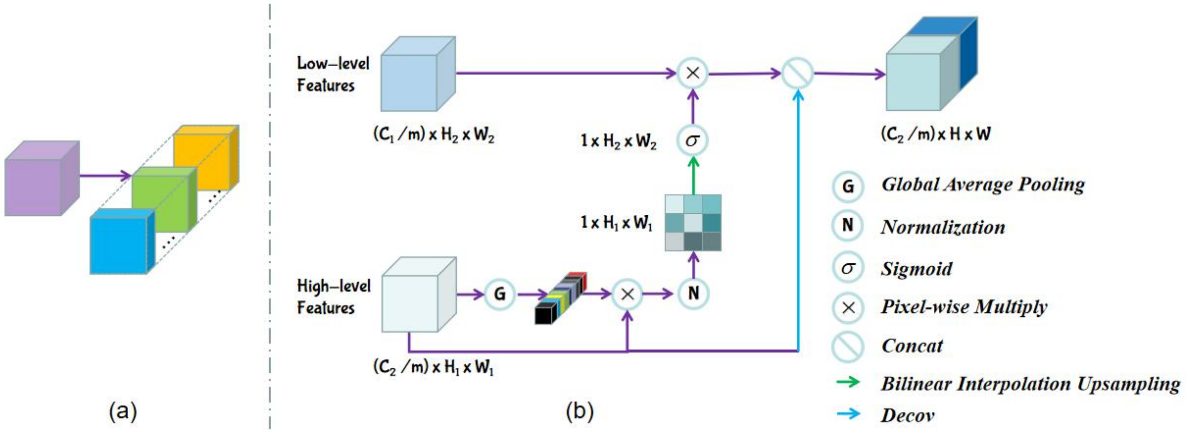
Low–level Features
$(C_1/m) \times H_2 \times W_2$
$1 \times H_2 \times W_2$
$1 \times H_1 \times W_1$
$(C_2/m) \times H \times W$
High–level Features
$(C_2/m) \times H_1 \times W_1$
G
N
σ
G Global Average Pooling
N Normalization
σ Sigmoid
× Pixel-wise Multiply
⃠ Concat
Bilinear Interpolation Upsampling
Decov
(a)
(b)

Fig 3. The illustrations of Spatial Group Attention Upsampling (SGAU) module. We use the spatial attention map generated by the high-level features to weight the low-level features, which is beneficial for the network to pay more attention to the lesion region. C2=2C1, H2=2H1, and W2=2W1.

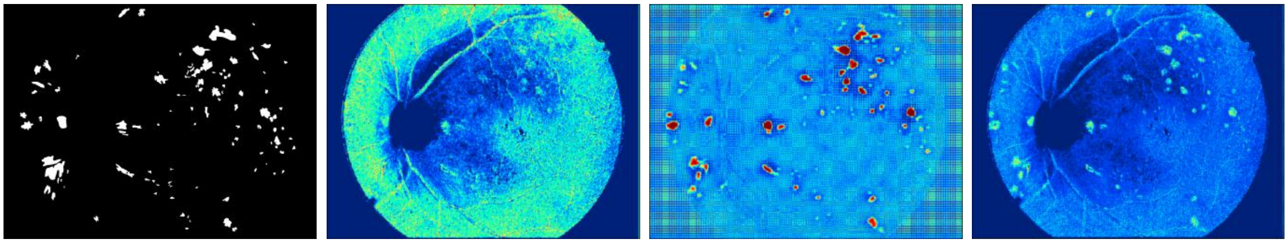

Fig 4. Example of feature map weighted by the SGAU module. (a) Ground truth. (b) Low-level feature map generated by the encoder. (c) Attention map generated by high-level feature. (d) Pixel-wise multiplication of (b) and (c).

The SGAU module process can be summarized as follows:

$$\mathrm{Y} = Cat(\sigma(N(F^{n+1} \otimes GP(F^{n+1}))) \otimes F^{n}; F^{n+1}) \tag{3}$$

3. RESULTS

3.1 Implementation Details

In the encoder, we use U-Net as our backbone. The generalized Dice loss [9] is applied as our loss function. The implement of the network is based on Pytorch and NVIDIA Tesla K40 GPU with 12GB memory. In the training stage, the Adam optimizer is used and the initial learning rate is set as 1e-4. The number of epochs is set as 300 and the batch size is set to 8. The group number **m** is set to 32. The original image size is 4428 × 2848. A part of background is cropped and the image is resized to 512 × 384. Data augmentation technology is adopted, which contains randomly horizontal flipping, randomly vertical flipping and random rotation. 3-fold cross validation strategy is adopted. The symptoms of DR including hard exudates and hemorrhages are segmented respectively.

3.2 Datasets

The performance of the propose GANet is evaluated based on a public dataset IDRiD, which contains 81 fundus images with diabetic retinopathy. All the subjects underwent mydriasis[10]. Retinal fundus images of diabetic patients were captured with 39 mm distance between lenses and examined eye using non-invasive fundus camera having xenon flash lamp (kowa vx-10 α , FOV 50°). The pixel-wised annotations were accomplished using special software developed by ADCIS [11].

3.3 Evaluation metrics

In this paper we use four metrics to evaluate our approach, Dice similarity coefficient, accuracy, sensitivity, and specificity. G denotes ground truth and P represents the model predict. TP , TN, FP, and FN represent true positive, true negative false positive, and false negative, respectively. Dice similarity coefficient is used to measure the overlay ratio between the prediction result and the ground truth, which is the main evaluate metrics. Accuracy represents the pixel-wise classification accuracy. Sensitivity represents the proportion of positive predictions in all real positives. Specificity represents the percentage of negative predictions in all real negatives.

$$Dice = \frac{2|G \cap P|}{|G|+|P|} \tag{4}$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{6}$$

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

3.4 Results

We remove a convolution from each layer of the decoder path of U-Net and take it as our baseline. Table 1 shows the results of comparison experiments and ablation experiments in the form of mean ± standard deviation. It can be seen from Table 1 that compared with the baseline, our SGAU module improved by 2.12% and 0.82% respectively in hard exudates and hemorrhage segmentation in Dice, while the CGA module improved by 2.01% and 0.58% respectively. The final results after adding two modules improved by 2.39% and 1.52% compared with the baseline. Fig. 5 shows some examples of hard exudate and hemorrhage segmentation results. As it can be seen in Fig. 5, the group attention network proposed by this paper can effectively suppress the false positives and true negatives, because SGAU and CGA modules can effectively enhance the feature expression of the network from space and channel respectively.

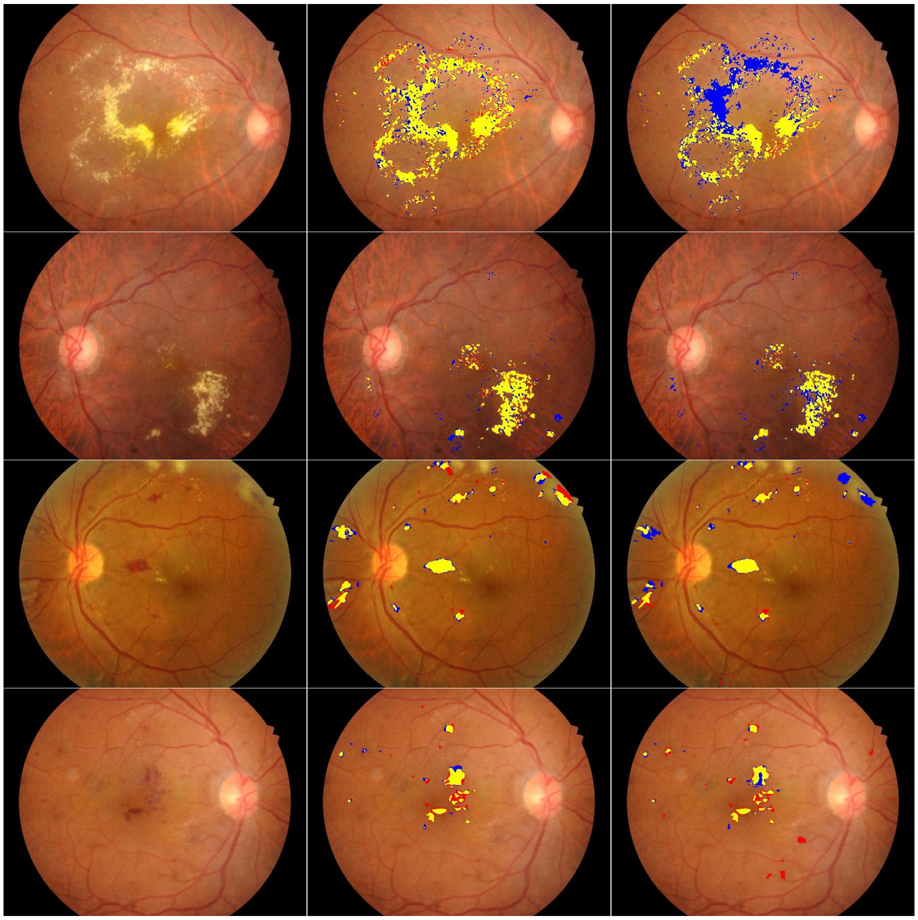

Fig 5. Examples of segmentation results. Blue, red, and yellow represent the false negative, false positive and true positive, respectively. From left to right: original image, results of the proposed GANet, and the results of U-Net. First and second row: hard exudate segmentation results, and third and fourth row: hemorrhage segmentation results.

Table 1. Performance comparison for hard exudate(EX) and hemorrhage(HE) segmentation.

Methods		Dice(%)	Acc(%)	Sensitivity(%)	Specificity(%)	# para
Unet	HE	41.52 ± 2.49	**99.11 ± 0.14**	37.37 ± 3.11	**99.79 ± 0.029**	7.764M
	EX	55.41 ± 5.77	99.42 ± 0.23	50.44 ± 6.62	**99.87 ± 0.045**	
Baseline	HE	42.02 ± 2.57	99.11 ± 0.16	39.01 ± 3.64	99.77 ± 0.034	6.981M
	EX	56.85 ± 4.53	99.40 ± 0.23	55.97 ± 3.58	99.8 ± 0.061	
Baseline + SGAU	HE	44.14 ± 3.48	99.09 ± 0.15	43.64 ± 5.44	99.76 ± 0.057	6.981M
	EX	57.67 ± 4.24	99.39 ± 0.23	57.11 ± 4.37	99.79 ± 0.069	
Baseline + SE	HE	43.50 ± 2.46	99.09 ± 0.15	44.87 ± 7.03	99.73 ± 0.12	7.025M
	EX	56.72 ± 5.70	99.39 ± 0.22	58.48 ± 7.51	99.74 ± 0.08	
Baseline + CGA	HE	44.03 ± 2.62	99.05 ± 0.19	44.70 ± 7.22	99.66 ± 0.158	6.987M
	EX	57.43 ± 5.01	99.40 ± 0.23	57.27 ± 7.03	99.78 ± 0.074	
Baseline + SGAU+CGA	HE	**44.41 ± 1.71**	99.08 ± 0.17	**44.96 ± 6.26**	99.71 ± 0.168	6.987M
	EX	**58.37 ± 4.52**	**99.42 ± 0.21**	**59.41 ± 4.82**	99.77 ± 0.073	

4. CONCLUSIONS

In this paper, we have purposefully improved the U-shape network. We propose two attention modules including SGAU and CGA, in which CGA module focuses on channel weight distribution and SGAU module focuses on spatial weight distribution. The generalized Dice loss is adopted to overcome the problem of unbalanced distribution. The experimental results show that the performance of the proposed GANet has been significantly improved, comparing with the original U-Net. The proposed method has the potential of assistance of DR diseases diagnosis.

5. REFERENCE

1. Dhoot, Dilsher S., et al. "Baseline factors affecting changes in diabetic retinopathy severity scale score after intravitreal aflibercept or laser for diabetic macular edema: post hoc analyses from VISTA and VIVID." Ophthalmology 125.1 (2018): 51-56.
2. Davidson, Jaime A., et al. "How the diabetic eye loses vision." Endocrine 32.1 (2007): 107-116.
3. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition.2015.
4. Li, Hanchao, et al. "Pyramid attention network for semantic segmentation." arXiv preprint arXiv:1805.10180 (2018).
5. Li, Xiang, Xiaolin Hu, and Jian Yang. "Spatial Group-wise Enhance: Improving Semantic Feature Learning in Convolutional Networks." arXiv preprint arXiv:1905.09646(2019).
6. Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 7794-7803).
7. Li, X., Zhong, Z., Wu, J., Yang, Y., Lin, Z., & Liu, H. (2019). Expectation-maximization attention networks for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (pp. 9167-9176).
8. Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
9. Sudre, Carole H., et al. "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations." Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, Cham, 2017. 240-248
10. Porwal, Prasanna, et al. "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research." Data 3.3 (2018): 25.
11. Advanced Concepts in Imaging Software (ADCIS). Aphelion Image Annotator. ADCIS France. 2018. Available online: http://www.adcis.net/en/Image-Processing-And-Analysis-Software-And-CustomEngineering-Developments.html (accessed on 2 July 2018).