Triple ANet: Adaptive Abnormal-aware Attention Network for WCE Image Classification

Xiaoqing Guo and Yixuan Yuan

Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China

Motivation & background

●Background

Digestive system cancer is the most common cancer with 328,030 new cases being diagnosed in the United States in 2019 [10]. GI tract cancers dominate the overwhelming responsibility for digestive system cancers, and about 79,830 people may die from the GI tract cancers [10].

●Challenges

WCE image classification still remains challenging due to the relatively low contrasts and ambiguous boundary between abnormalities and normal regions. Additionally, the huge intra-class variations, alone with the high degree of visual similarities shared by inter-class lesions prevent the network from robustly classification.

●Our contribution

1) An Adaptive Dense Block (ADB) is developed to adaptively assign one attention score for each dense connected layer in dense blocks, and the score reveals importance of feature maps in different depth. 2) We propose an Abnormal-aware Attention Module (AAM) that can gradually adjust the respective field according to the abnormal regions. This AAM aims to combine local information with context and help network pay attention to the abnormal region. 3) A novel angular contrastive loss (AC Loss) is proposed to reduce the intra-class variations and enlarge the inter-class differences effectively.

Methodology

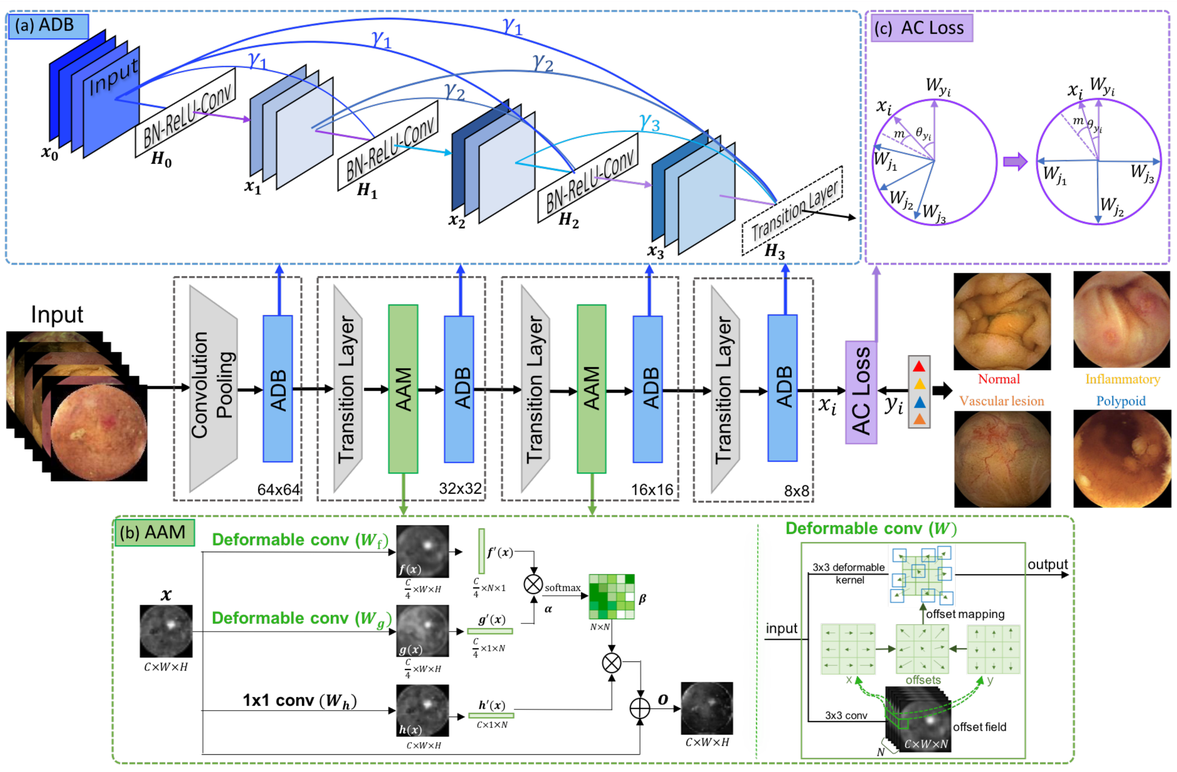


Fig. 1. Illustration of the proposed Triple ANet.

The source code for Triple ANet is available at https://github.com/Guo-Xiaoqing/Triple-ANet.

●Adaptive Dense Block

ADB is proposed to introduce direct and suitable connections from any layer to all subsequent layers as in Fig. 1(a). Let x_l denote the output of l^{th} layer and γ_l is the corresponding weight scalar, adaptive dense connectivity is formulated as $x_l = H_l([\gamma_0 x_0, \gamma_1 x_1, \dots, \gamma_{l-1} x_{l-1}])$, where $H_l(\cdot)$ is a composite function with Batch Normalization, ReLU and Convolution, and each $H_l(\cdot)$ produces k = 12 feature maps. All the weights γ are initialized as 1s and optimized with iteration, which makes the useful convolutional signals gradually enhanced. Specifically, Triple ANet includes four ADBs as shown in Fig. 1, and they are comprised of 6, 12, 24, 16 densely connected layers, respectively.

●Abnormal-aware Attention Module

Considering that features in the neighborhood of abnormalities also make contribution to the identification, AAM (Fig.1(b)) is proposed to combine context with local information, including three branches. The 1st and 2nd branches are deformable convolutions, while the 3rd branch is 1x1 convolution. Especially, the deformable convolution enables the receptive fields to gradually expand around the abnormalities. Thus, the AAA module can calculate region-to-region correlation and focus training on important region as well as its neighbors.

● Angular Contrastive Loss

The softmax cross entropy loss is widely used and can be formulated as: $L_{softmax} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^{\top}x_i+b_{y_i}}}{\sum_{j=1}^{n}e^{W_j^{\top}x_i+b_j}}$ However, softmax loss function does not explicitly optimize the feature to ensure higher differences for interclass features and similarity for intraclass features.

For simplification, we firstly discard bias item and rewrite the formula of FC layer as $F(W_j, x_i) = \|W_j\| \cdot \|x_i\| \cos\theta_j$, where θ_j is the angle between the features x_i and the weight W_j. Then x_i and W_j are respectively normalized by L2 normalization. Our AC Loss takes a general form of FC layer and introduces an angle margin penalty m and a regularization term, which can be formulated as:

$$L_{AC} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{sA(\theta_{y_i}+m)}}{e^{sA(\theta_{y_i}+m)}+\sum_{j=1,j\neq y_i}^{n}e^{sA(\theta_j)}} + \sum_{y_i=1}^{n}\sum_{j=1,j\neq y_i}^{n}W_{y_i}^{\top}W_j$$

where the angular activation function is $A(\theta_j) = \frac{1+e^{(-\frac{\pi}{2k})}}{1-e^{(-\frac{\pi}{2k})}}\cdot\frac{1-e^{(\frac{\theta_j}{k}-\frac{\pi}{2k})}}{1+e^{(\frac{\theta_j}{k}-\frac{\pi}{2k})}}$. k is a hyper parameter to control the gradient of angular activation function and the performance of loss function directly.

As shown is Fig. 1 (c), converged L_{AC} maximizes the separability of inter-class features and enables intra-class features to cluster toward the weight of their corresponding class.

Experiment Results

●Dataset

We evaluated Tripe ANet model on a combined WCE dataset from CAD-CAP[2], KID [6] and our collected polyp dataset. It consists of 2846 WCE images, including 771 normal frames, 728 inflammatory ones, 762 vascular lesion ones and 585 polyps.

●ADB Analysis

Fig. 2 shows the change of γ values in the 4th ADB with 16 connected layers, and γ_i indicates the learned attention score for the ith layer. The assigned attention values for different layers varies significantly. It can be seen that the deeper the layer is, the higher the value of γ is, indicating the deep feature will make more contributions for the WCE image classification.

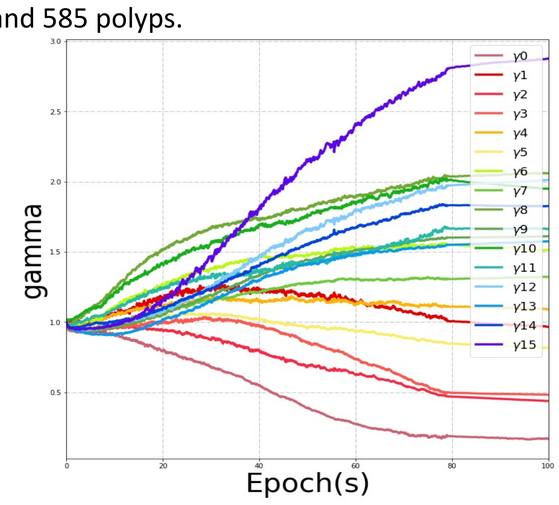
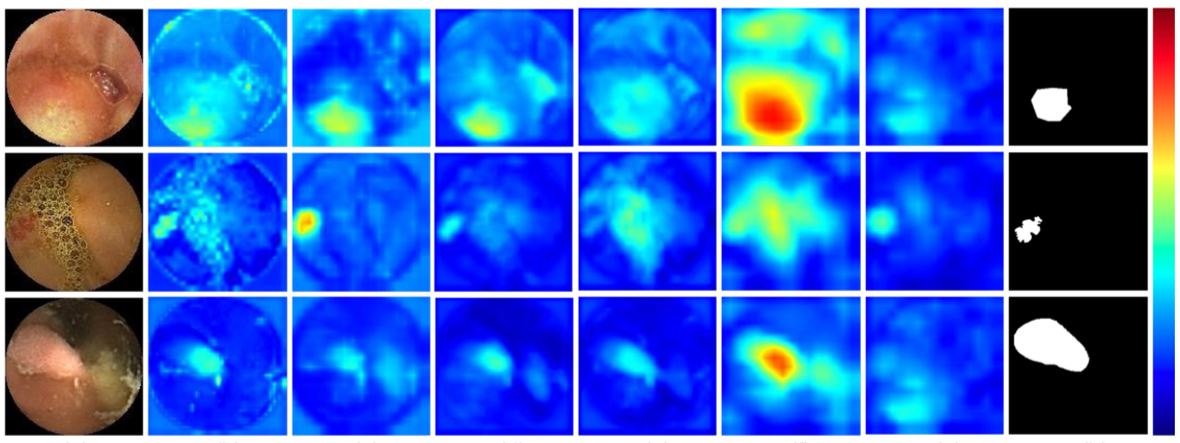


Fig. 2. γ in the fourth ADB for different epochs.

●AAM Analysis

Fig. 3 (b-c) show the feature maps extracted from the two AAMs, and we could find that AAMs make feature maps highlight the abnormal regions successfully. Fig. 3 (d-g) show the offset fields obtained in these AAMs, and red regions indicate the large receptive fields. It can be seen that the respective fields are tend to be larger at abnormalities compared with normal regions. Therefore, context information of abnormalities are incorporated with local information to lead better classification performance.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. From top to bottom, they are respectively inflammatory, vascular lesion and polyp samples.

●Comparison with State-of-the-art Methods

To individually demonstrate the effectiveness of the proposed ADB, AAM and AC Loss, we conducted several comparison experiments and the results were shown in Table 1 row 1–4. In general, it is clear that the proposed ADB, AAM and AC Loss make contribution to the promotion of performance, because involving any of them leads to relatively better performance compared with traditional DenseNet [4].

The proposed method shows superior performance with an increment of 12.31%, 11.58%, 2.74%, 1.48% in OA, 16.52%, 15.51%, 3.67%, 1.98% in Cohen's Kappa compared with state-of-the-art methods [3,5,7,9].

Table 1. Comparison results for WCE image classification. w/ADB, w/AAM and w/AC Loss denote DenseNet with ADB (instead of original dense block), DenseNet with AAM and DenseNet minimized by AC Loss (instead of original softmax cross entropy loss), respectively.

Methods	Normal ACC (%)	Inflammatory ACC (%)	Vascular lesion ACC (%)	Polyp ACC (%)	OA	Cohen's Kappa
DenseNet [4]	92.69 ± 0.49	90.12 ± 0.54	93.71 ± 0.52	97.59 ± 0.39	87.05 ± 0.45	82.66 ± 0.60
w/ADB	93.03 ± 0.40	89.94 ± 0.18	93.61 ± 0.15	97.71 ± 0.27	87.14 ± 0.21	82.78 ± 0.28
w/AAM	**94.06 ± 0.17**	91.49 ± 0.81	94.47 ± 0.77	97.78 ± 0.42	88.89 ± 0.52	85.13 ± 0.69
w/AC Loss	93.59 ± 0.30	91.20 ± 0.33	94.94 ± 0.34	97.69 ± 0.46	88.70 ± 0.20	84.87 ± 0.27
Triple ANet	94.03 ± 0.09	**91.73 ± 0.29**	**95.26 ± 0.33**	**97.81 ± 0.20**	**89.41 ± 0.23**	**85.82 ± 0.31**
Fan et al. [3]	85.44 ± 1.43	83.09 ± 0.79	90.19 ± 0.96	95.47 ± 0.89	77.10 ± 1.14	69.30 ± 1.58
Jia et al. [5]	86.16 ± 1.07	83.37 ± 0.71	90.32 ± 0.88	95.81 ± 0.59	77.83 ± 1.28	70.31 ± 1.74
Seguí et al. [7]	92.11 ± 0.60	89.71 ± 0.48	94.21 ± 0.57	97.31 ± 0.12	86.67 ± 0.84	82.15 ± 1.12
Yuan et al. [9]	93.44 ± 0.30	90.79 ± 0.26	93.91 ± 0.17	97.73 ± 0.35	87.93 ± 0.07	83.84 ± 0.08

Conclusion

Automatic abnormality classification is a challenge task due to the diverse characteristics rendered on WCE images. Our method is fundamentally different from the previous works with traditional convolutional network applications. Instead, we proposed a novel Triple ANet with Adaptive Dense Block (ADB), Abnormal-aware Attention Module (AAM) and Angular Contrastive loss (AC Loss). Our methods can be flexibly transferred to a wide range of medical image classification tasks to extract discriminative features and boost the classification performance.

Reference

[1] Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: CVPR, pp. 4690–4699 (2019)
[2] Dray, X., et al.: Cad-cap: une base de donnees francaise a vocation internationale, pour le developpement et la validation d'outils de diagnostic assiste par ordinateur en videocapsule endoscopique du grele. Endoscopy 50(03), 000441 (2018)
[3] Fan, S., et al.: Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. Phys. Med. Biol. 63(16), 165001 (2018)
[4] Huang, G., et al.: Densely connected convolutional networks. In: CVPR. pp. 4700{4708 (2017)
[5] Jia, X., Meng, M.Q.H.: A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: EMBC, pp. 639–642 (2016)
[6] Koulaouzidis, et al.: Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. Endoscopy international open 5(06), (2017)
[7] Segu´ı, S., et al.: Generic feature learning for wireless capsule endoscopy analysis. Comput. Biol. Med. 79, 163–172 (2016)
[8] Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR, pp. 7794–7803 (2018)
[9] Yuan, Y., et al.: Riis-densenet: Rotation-invariant and image similarity constrained densely connected convolutional network for polyp detection. In: MICCAI. pp. 620{628. Springer (2018)
[10] R. L. Siegel, et al.: "Cancer statistics, 2019." Ca A Cancer Journal for Clinicians, vol. 69, no. 1, pp. 7–34, 2019.