Feature Patch Based Attention Model for Dental Caries Classification

Genqiang Ren[1], Yufei Chen[1(✉)], Shuai Qi[2], Yujie Fu[2], and Qi Zhang[2]

[1] College of Electronics and Information Engineering, Tongji University, Shanghai, China
yufeichen@tongji.edu.cn

[2] Department of Endodontics, School and Hospital of Stomatology, Tongji University, Shanghai Engineering Research Center of Tooth Restoration and Regeneration, Shanghai, China

Abstract. Dental caries is a common dental disease. According to statistics, about 90% of adults suffer from dental caries. Therefore, early detection and treatment of dental caries are crucial to dental health. According to the depth of carious lesions, dental caries can be classified into shallow, moderate, and deep caries. Among them, the accurate classification of moderate caries and deep caries is important to making the subsequent treatment plan. Clinically, doctors can make the diagnosis with the help of CBCT images. However, due to the spatial complexity of the 3D volume, the difficulty of labeling the carious lesion, and the insignificant difference between moderate and deep caries, there is still a great challenge to accurately identifying moderate and deep caries. And to the best of our knowledge, there is no study on automatic dental caries classification based on CBCT images. In this paper, we propose a feature patch based attention model to improve the classification accuracy of dental caries in CBCT images. We extract overlapping patches from the 3D feature maps and assign every patch with a corresponding weight computed by adaptive learning to achieve automatic screening of regions that are critical for classification. We collect a real-world dental dataset which includes 167 CBCT scans with moderate caries and 157 CBCT scans with deep caries. A series of experiments demonstrate that our algorithm achieves 92% accuracy on caries classification, which outperforms state-of-the-art methods by a large margin.

Keywords: Dental caries diagnosis · Attention mechanism · Dental cbct

1 Introduction

Dental caries is a multifactorial, dynamic disease that results in net mineral loss of dental hard tissues and a carious lesion [1]. The teeth are composed of three main parts, from the outside to the inside: enamel, dentin, and pulp. Caries start on the enamel surface and progress to the dentin until it affects the whole dentin

Y. Chen et al. (Eds.): CLIP 2022, LNCS 13746, pp. 62–71, 2023.
https://doi.org/10.1007/978-3-031-23179-7_7

layer and finally lead to the inflammation of the pulp. The inflammation of the dental pulp will result in great pain for the patient. As one of the most common chronic dental diseases, caries can occur throughout life, affecting between 60% and 90% of school children and most adults [2,3]. Therefore, early detection and treatment of caries are crucial to dental health.

Corresponding to the depth of the carious lesion, caries has several stages: shallow caries, moderate caries, and deep caries. Different carious lesion depths correspond to different treatment strategies and prognoses. By examining the content of those slices in all three views (Axial, Sagittal, Coronal), clinicians may figure out the carious lesions' depth and the specific relationship between the carious lesions and the pulp in CBCT images, thus formulate further treatment plans. However, this evaluation process is sometimes rather time-consuming and challenging. Clinicians must check hundreds of slices in all three views to acquire comprehensive information about the carious lesions. In addition, CBCT is not as prevalent as periapical radiographs or bitewings in our routine use, leading to the capacity variance among clinicians regarding radiograph diagnosis. Hence, it is of great significance to develop an automated computer-aided system based on CBCT, boosting caries evaluating efficiency on CBCT images and improving the accuracy of diagnosing carious lesions. So far, in the field of dentistry, some deep learning-based of computer-aided diagnosis works have been investigated [4–12]. But, to the best of our knowledge, there is no existing work to classify dental caries in CBCT images. The main technical challenges are as follows: (1) inter-slice variance. The slices of the same CBCT image correspond to different caries categories, for example, some slices belong to moderate caries, and others to deep caries. Labeling all slices in CBCT images is time-consuming and laborious. This limits the use of most 2D CNN models, such as ResNet [13]. (2) low signal-to-noise ratio. The percentage of the carious lesion in the dental CBCT images is tiny, The percentage of the carious lesion in the dental CBCT images is tiny, resulting in the convolution mechanism being unable to capture distinguishing features. Previous studies have shown that caries diagnosis is related to the severity of caries intrusion into the teeth, especially in the dentin region. Hence, directly feeding the whole CBCT image into a deep model usually results in an unsatisfactory diagnosis, as the irrelevant dental region would harm, instead of help, the model's discriminatory power.

The attention mechanism is widely used in deep learning-based models to enhance the influence of task-relevant information while suppressing irrelevant information. Meanwhile, compared to 2D models, 3D models are more popular since they can not only focus on disease-specific volumes but also explore the inter-slice context of those regions, which might be critical to this diagnosis task. In this work, we propose a patch-based attention mechanism that extracts overlapping patches from the 3D feature maps and assigns every patch with a corresponding weight computed by adaptive learning. To enhance the ability to distinguish features of the carious lesion and ignore the irrelevant features, when a patch is a key to the image-level decision, the weight is high; on the contrary, the weight is lower. Inspired by the fact that in diagnosing dental caries, doctors should pay attention to the local information of the lesion area

and the global information, which is the relative depth of carious lesions into the dentin, compared to extracting non-overlapping patches strategy, we extract overlapping patches to capture the global information. Meanwhile, compared to the method of extracting patches from the raw image, which will decompose the target into multiple patches, our approach that extracts patches from the 3D feature maps is better for classification by retaining the integrity of the target. The contributions of this paper can be summarized as follows:

- A one-stage dental caries classification framework is proposed. It takes the classification task directly on the source image without additional pre-localiz ation steps.
- A patch attention model is designed based on the 3D feature maps. It adaptively calculates the contribution of each patch and then assigns the corresponding weight so that the lesion region can get more attention for better classification.
- We evaluate the proposed algorithm on a dental dataset containing 324 sets of CBCT images of moderate and deep caries, which are two easily confused diseases in the clinic. Our method has obtained superior results in the task.

2 Related Work

This section briefly introduces previous studies on computer-aided diagnosis methods with 3D medical image data and review attention mechanism related works in medical imaging analysis.

Medical Image Classification. The existing automatic diagnosis in 3D images data are based on segmentation or detection tasks requiring detailed annotations [14,15]. With the auxiliary task, it is possible to first localize the lesion area, which in turn only needs to be fed into a classifier. For example, Xu et. al [14] leveraged the pixel-wise annotations to detect lesions and constructed a classifier based on the segmentation results. To reduce the time and labor cost of manual annotation, 3D medical diagnosis using merely patient-level labels has become a promising alternative. However, due to the low caries signal-to-noise ratio, it is difficult for the conventional CNN to extract more discriminative features when guided by only patient-level labels. Inspired by ViT's [16] idea of cutting the original image into smaller patches to reduce computation, we propose to cut the whole CBCT into multiple patches and filter out the critical patches for classification through an attention mechanism to improve the efficiency of the model. But unlike ViT, which performs non-overlapping cropping on the original image in a way that destroys the integrity of the target (the same object will be divided into multiple patches), we extract patches on the feature space to ensure the integrity of the target.

Attention Mechanism. Motivated by the success of attention in NLP, the attention mechanism has been widely used in the medical imaging analysis

domain to enhance the influence of task-relevant information while suppressing irrelevant information [17–19]. Different task-oriented attention modules have been proposed to enhance the features of disease-related regions in images, thus improving the accuracy of classification or segmentation. For instance, Schlemper et. al [17] propose a novel attention gate (AG) model for medical image analysis that automatically learns to focus on target structures of varying shapes and sizes.

Although deep learning-based approaches have performed well in the field of medical image analysis, due to the spatial complexity of the 3D volume, the difficulty of labeling the carious lesion, and the insignificant difference between moderate and deep caries, there is still a great challenge to accurately identifying moderate and deep caries. Considering that attentional mechanisms can enhance task-related features, we explore a patch-based attention mechanism to improve the classification accuracy of dental caries in CBCT images.

3 Methodology

According to the above directions, we construct a framework to perform an automated diagnosis of dental caries using merely patient-level labels in CBCT images. The proposed framework is illustrated in Fig. 1. Compared to ViT [16], instead of extracting patches from raw input images, the CBCT images are fed into a 3D CNN and the patch-level features are obtained through this CNN. Then we assigns every patch with a corresponding weight computed by attention module so that the lesion region can get more attention for better classification.

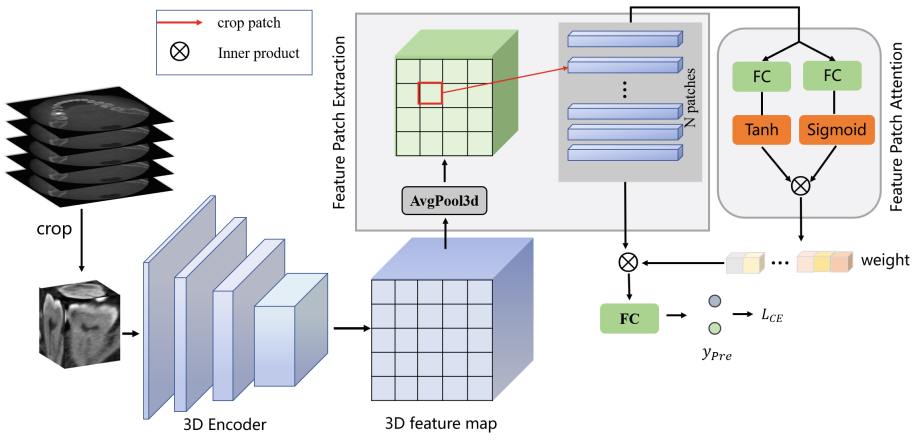

Fig. 1. Illustration of the pipeline of the proposed framework.

3.1 Feature Patch Extraction

Existing popular approaches [20] generally divide the CT/MR images into multiple cubic patches with a fixed size without overlapping. However, such a manner would split the target into different patches, neglecting the spatial and global information. Inspired by the patch extraction in [21], we propose a simple but effective module that extracts patches from feature maps instead of the input image. In addition, to capture the local and global information, we propose extracting overlapping patches. In this manner, each patch has relatively complete semantic information which is essential for the classification task. Generally speaking, given a 3D dental CBCT scan X_i with the shape of $D*H*W$, we can get a 3D feature map $\tilde{X}_i$ using a 3D Encoder. Then, the feature map is fed into an average pooling layer and a new feature map $\hat{X}_i$ is obtained. The shape of $\hat{X}_i$ is $C*D*H*W$, where C, D, H, W represent the channel, depth, high, width, respectively. We view each point of the feature map as a patch with the shape of $C*1$. Hence, in this way, for the given CBCT scan X_i, we extract $N = D*H*W$ patches in total. We define the patch generator operation as ψ, and the patches are as follow:

$$\mathcal{H}_i = \{h_1, h_2, \ldots, h_N\} \tag{1}$$

where N denotes the quantity of patches, $h_i \in \mathbb{R}^{N*C}$. Note that the raw location of corresponding patches on the 3D CBCT scan can be easily derived according to the location of patches on the feature map. Formally, this step can be formulated into:

$$\mathcal{H}_i = \psi\left(X_i\right) \tag{2}$$

In conclusion, the patch extraction module in our work not only transforms patches into embedding space but generates patches that are not defined before(such as ViT). Viewing each point in the feature maps as a patch is a straightforward routine to extract 3D patches that consider the spatial relations between patches. The main difference with the existing method [21] is that our generator can apply on 3D data.

3.2 Feature Patch Attention

After obtaining a set of 3D patches $\mathcal{H}_i$, inspired by [22], we propose an attention-based patches selection module. The attention-based patches selection is defined by

$$z = \sum_{n=1}^{N} a_n \boldsymbol{h}_n \tag{3}$$

where,

$$a_n = \frac{\exp\left\{\boldsymbol{w}^\top \left(\tanh\left(\boldsymbol{V}\boldsymbol{h}_n^\top\right) \odot \operatorname{sigm}\left(\boldsymbol{U}\boldsymbol{h}_n^\top\right)\right)\right\}}{\sum_{j=1}^{N} \exp\left\{\boldsymbol{w}^\top \left(\tanh\left(\boldsymbol{V}\boldsymbol{h}_j^\top\right) \odot \operatorname{sigm}\left(\boldsymbol{U}\boldsymbol{h}_j^\top\right)\right)\right\}} \tag{4}$$

where $\boldsymbol{w} \in \mathbb{R}^{N*1}$, $\boldsymbol{V} \in \mathbb{R}^{N*D}$ and $\boldsymbol{U} \in \mathbb{R}^{N*D}$ are trainable parameters. $\odot$ is an element-wise multiplication and sigm(.) is the sigmoid non-linearity. We use the

hyperbolic tangent $\tanh(\cdot)$ element-wise non-linearity for proper gradient flow. In simple terms, if a 3D patch has a more significant weight value calculated by the attention module, it means that the semantic features in this patch are essential for the classification task. From another perspective, the weight values calculated by the attention module directly reflect the contribution of each patch extracted in the previous step to the patient-level diagnosis. Therefore, the patch-based attention mechanism gives strong interpretability to the predictions. In summary, let σ_a with parameters θ_{σ_a} represent the attention-based patches selection, this step can be formulated into:

$$z_i = \sigma_a\left(\mathcal{H}_i\right) \tag{5}$$

On the feature maps extracted by a 3D Encoder, the patch-based attention is able to filter out task-relevant semantic features. Because of the higher resolution of CBCT images, the number of patches is also quite large, resulting in a surge in weight calculation.

3.3 Loss Function

We train the proposed caries classification model using a patient-level label. The loss function we use in model training based on the cross entropy loss is described as:

$$\mathcal{L}(\mathbf{W}) = -\frac{1}{N}\sum_{n=1}^{N}\log\left(P\left(Y_n \mid \boldsymbol{X}_n; \mathbf{W}\right)\right) \tag{6}$$

where N is the number of images, $P\left(Y_n \mid \boldsymbol{X}_n; \mathbf{W}\right)$ is the probability of correct prediction for $\boldsymbol{X}_n$, $(\boldsymbol{X}_n, Y_n)$ is the training sample, $\mathbf{W}$ is the parameter of the model.

4 Experiments and Analysis

4.1 Dataset and Settings

Dataset. In this study, we collected a real-world dental dataset comprised of 324 CBCT scans between 2020 and 2022 from Stomatology Hospital. The dataset includes 167 CBCT scans with moderate caries and 157 CBCT scans with deep caries. The random selected dental CBCT image are illustrated in Fig. 2. All images are resampled to 0.1 * 0.1 * 0.1 mm isotropic resolution and labeled by experienced dentists. The splitting of the training set and testing set is according to the patient level. This study and all research were approved and conducted following relevant guidelines/regulations.

Implementation and Details. The framework is implemented using Pytorch 1.9 and trained on a workstation equipped with a NVIDIA GTX 3090 graphics card. During the training process, the model is optimized by Adaptive Moment Estimation(Adam) algorithm with 500 epochs, and the weight decay is $1 * 10^{-4}$. The initial learning rate is 0.0001. For the tooth dataset, 70% of the samples and

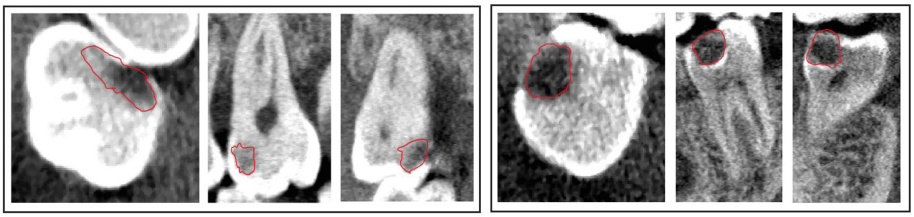

Fig. 2. The visualization of dental CBCT slices from the collected dataset. (a)middle caries (b)deep caries

20% of the samples are respectively selected as the training set and validation set to supervise the training of the model, and 10% of the samples are as the testing set to evaluate the performance.

Evaluation Metrics. In order to verify the effectiveness of the framework, 5-fold cross-validation is adopted, and some evaluation metrics, including $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, $F1\ score = \frac{2*Precision*Recall}{Precision+Recall}$, $Recall = \frac{TP}{TP+FN}$, and $Precision = \frac{TP}{TP+FP}$ are used to evaluate the classification performance, where TP, TN, FP, and FN are the number of true positive samples, true negative samples, false positive samples, and false negative samples, respectively.

4.2 Experimental Results

Table 1 shows the result of caries classification on CBCT images. All the used algorithms are achieving promising performance. Among them, our method significantly outperforms the ResNet-18 [13] 3D, Resnet-50 [13] 3D, and GCNet [23] 3D models on almost all metrics.

Table 1. Comparison of classification results of different methods.

	Accuracy	F1-score	Precision	Recall
Resnet-18 [13] 3D	0.7815	0.7746	0.7856	0.7915
Resnet-50 [13] 3D	0.8127	0.7997	0.8100	0.7940
GCNet [23] 3D	0.8235	0.8126	0.8094	0.8035
Proposed	0.9218	0.9193	0.9200	0.9189

It reveals that compared to existing methods, our model is the best in several metrics and can obtain a more interpretable result, as illustrated in Fig. 3. We employed the salience mask of the final trained model to identify the critical

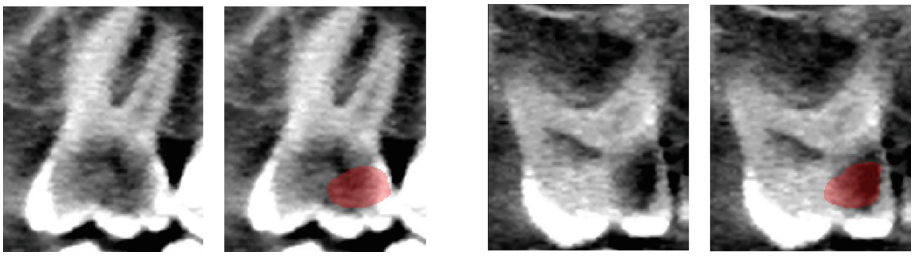

Fig. 3. The visualization of key patches(in red) in dental CBCT image. (Color figure online)

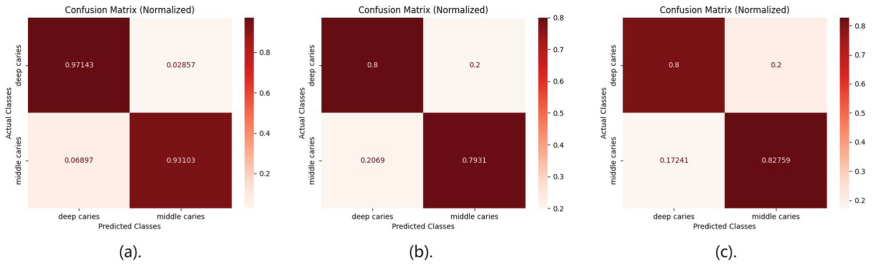

Fig. 4. The confusion matrix of two classes classification: moderate caries and deep caries. (a).Proposed (b). ResNet-50 [13] 3D (c). GCNet [23] 3D

regions for moderate caries vs. deep caries classification. Fig. 4 shows the confusion matrixes of our method, Resnet-50 3D, and GCNet 3D. And our method obtains a balance performance. Comparing ResNet-18 and ResNet-50, two more commonly used backbone models in deep learning, it can be found that simply increasing the model complexity does not significantly improve the accuracy of the model. Comparing GCNet with our model, although both use the attention mechanism, our approach tends to extract features more efficiently, as demonstrated by visualizing the critical patches.

5 Conclusion

Compared with traditional convolution-based mechanisms, our method of patch-based attention model can capture more discriminative features and achieve automatic identification and classification of lesion regions without the aid of auxiliary tasks to localize lesion areas such as segmentation and detection, and the results are interpretable, which can assist clinicians in diagnosis. However, some samples are difficult to distinguish between medium and deep caries solely based on imaging information, and further reference to other clinical information of the patient is needed. In addition, for better clinical application, the current binary classification needs to be extended to four classifications, i.e. normal, superficial, medium and deep caries, which is our future work to be carried out.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (No. 62173252, 61976134), the Clinical Research Plan of Shanghai Hospital Development Center (grant no. SHDC2020CR3058B), the Shanghai Municipal Health Commission (grant no. 202040282).

References

1. Machiulskiene, V., et al.: Terminology of dental caries and dental caries management: consensus report of a workshop organized by ORCA and cariology research group of IADR. Caries Res. **54**(1), 7–14 (2020)
2. Pitts, N.B., et al.: Dental caries. Nat. Rev. Dis. Primers. **3**(1), 1–16 (2017)
3. Munteanu, A., Holban, A.M., Păuna, M.R., Imre, M., Farcaiu, A.T., Farcaiu, C.: Review of professionally applied fluorides for preventing dental caries in children and adolescents. Appl. Sci. **12**(3), 1054 (2022)
4. Duan, W., Chen, Y., Zhang, Q., Lin, X., Yang, X.: Refined tooth and pulp segmentation using U-Net in CBCT image. Dentomaxillofacial Radiol. **50**(6), 20200251 (2021)
5. Lin, X., et al.: Micro-computed tomography-guided artificial intelligence for pulp cavity and tooth segmentation on cone-beam computed tomography. J. Endodontics **47**(12), 1933–1941 (2021)
6. Yang, X., Chen, Y., Yue, X., Lin, X., Zhang, Q.: Variational synthesis network for generating micro computed tomography from cone beam computed tomography. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1611–1614. IEEE (2021)
7. Lee, J.H., Kim, D.H., Jeong, S.N., Choi, S.H.: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J. Dent. **77**, 106–111 (2018)
8. Cantu, A.G., et al.: Detecting caries lesions of different radiographic extension on bitewings using deep learning. J. Dent. **100**, 103425 (2020)
9. Casalegno, F., et al.: Caries detection with near-infrared transillumination using deep learning. J. Dent. Res. **98**(11), 1227–1233 (2019)
10. Schwendicke, F., Elhennawy, K., Paris, S., Friebertshäuser, P., Krois, J.: Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study. J. Dent. **92**, 103260 (2020)
11. Moutselos, K., Berdouses, E., Oulis, C., Maglogiannis, I.: Recognizing occlusal caries in dental intraoral images using deep learning. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1617–1620. IEEE (2019)
12. Liu, L., Xu, J., Huan, Y., Zou, Z., Yeh, S.C., Zheng, L.R.: A smart dental health-IoT platform based on intelligent hardware, deep learning, and mobile terminal. IEEE J. Biomed. Health Inf. **24**(3), 898–906 (2019)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
14. Xu, X., et al.: A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering **6**(10), 1122–1129 (2020)
15. Roth, H.R., et al.: Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE Trans. Med. Imaging **35**(5), 1170–1181 (2015)

16. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
17. Schlemper, J., et al.: Attention gated networks: learning to leverage salient regions in medical images. Med. Image Anal. **53**, 197–207 (2019)
18. Wang, S., Li, L., Zhuang, X.: AttU-Net: attention U-Net for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) International MICCAI Brainlesion Workshop. BrainLes 2021, vol. 12963, pp. 302–311. Springer, Cham (2022)
19. Shen, C., et al.: Attention-guided pancreatic duct segmentation from abdominal CT volumes. In: Oyarzun Laura, C., et al. (eds.) DCL/PPML/LL-COVID19/CLIP -2021. LNCS, vol. 12969, pp. 46–55. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-90874-4_5
20. Gao, X., Qian, Y., Gao, A.: COVID-VIT: Classification of COVID-19 from CT chest images based on vision transformer models. arXiv preprint arXiv:2107.01682 (2021)
21. Yuan, K., Guo, S., Liu, Z., Zhou, A., Yu, F., Wu, W.: Incorporating convolution designs into visual transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 579–588 (2021)
22. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning, pp. 2127–2136. PMLR (2018)
23. Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H.: GCNet: non-local networks meet squeeze-excitation networks and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (2019)