Inter-slice Correlation Weighted Fusion for Universal Lesion Detection

Muwei Jian[1,2*]
School of Computer Science and Technology
Shandong University of Finance and Economics
Jinan, China
jianmuweihk@163.com

Yue Jin[1]
School of Computer Science and Technology
Shandong University of Finance and Economics
Jinan, China
jinyue@mail.sdufe.edu.cn

Rui Wang[1]
School of Management Science and Engineering
Shandong University of Finance and Economics
Jinan, China
rachelwang_cs@163.com

Xiaoguang Li[3]
Faculty of Information Technology,
Beijing University of Technology,
Beijing, China
lxg@bjut.edu.cn

Hui Yu[2*]
School of Creative Technologies
University of Portsmouth
Portsmouth, UK
hui.yu@port.ac.uk

***Abstract*—Universal lesion detection using computerised tomography (CT) scans is a critical computer-aided diagnosis measure in clinical diagnosis. One of the key issues during the diagnosis is the correlations between sequential slices to improve the feature representation of CT scans. In the process of fusing slice features containing temporal correlations, the correlation between the contextual slices in the channel dimension and the target slices is closely related to the spatial distance in practice. However, convolutional fusion approaches commonly ignore that features of different distances have unequal weights. To tackle this issue, we present a temporal correlation weighted fusion lesion detection network, called TCW-Net. Specifically, for the slices in the channel dimension, we develop a weighted feature fusion module to adjust the more discriminative features using learned weights. Then, we adapt a spatial offset attention mechanism that allows the detection network to pay more attention to the lesion's slight spatial offset and thus improve the model's capacity for distinguishing between different lesion features. Extensive experiments carried out on the DeepLesion dataset show that the proposed algorithm has superior performance over the state-of-the-art methods.**

Keywords—Universal lesion detection, CT, Temporal correlation, Weighted fusion

I. INTRODUCTION

Computerized Tomography (CT) scanning is a widely used medical imaging technique that utilizes X-rays to scan the body layer by layer, producing digital images that accurately depict the internal structure of the human body through computer-aided processing [1][35]. This technique provides high-resolution images that can effectively display the tissues and organs of the body. Additionally, it is relatively safe and efficient, making it an essential tool in disease diagnosis, particularly in the differentiation between benign and malignant lesions. The prevalence of CT examinations requires more and more radiologists for analysis, which results in an imbalance between the overload of the doctors and the speed limitation of the CT reports.

There is thus an urgent need for computer technologies to assist the diagnosis process. Intelligent computer-aided diagnosis has gradually attracted the attention [37]. According to the surveys on clinical demands, radiologists usually need to observe and diagnose the lesion areas of multiple organs at the same time. For example, when cancer cells spread in the body of the patients, a universal lesion detection method is needed to assist in diagnosis.

The lesions in CT images are usually characterized by small-sized and blurred edges. Therefore, how to extract more efficient features information from CT images has become a challenge in universal lesion detection. Recently, a series of findings have been achieved for enhancing feature representation [1-7][36]. The key ideas of these method are: 1) introducing attention mechanisms [2,5-6], such as channel or spatial attention mechanisms, which focus the information of task concerns on prominent regions and thus improve the richness of feature representation; 2) adding deformable convolution [4] in the network, which adaptively adjust the size of perceptual field; 3) using pseudo-3D [3,7] techniques for 3D CT slices. In addition, improvement of universal lesion detection has been achieved by optimizing the anchor [8-10] or anchor-free mechanism[11-12], and by embedding domain knowledge [13].

Inspired by the camouflaged object detection in the video task [13], we observed that correlations between frames is helpful for the detection of camouflaged objects. We thus assume that enhancing temporal correlations between slices of CT images can help to improve the feature representation of CT images which leads to better lesion detection. Furthermore, the correlation between the channel dimension context slices and target slices depends on spatial distance, which most existing methods ignore[38].

We propose a new method for universal lesion detection called TCW-Net, which addresses the issue of unequal weighting of features in existing methods by including a weighted feature fusion module that gives weights to each slice feature in the channel dimension. This allows for more adaptive highlighting of discriminative feature information during temporal correlation fusion of individual slices. In addition, we also introduce a spatial offset attention mechanism that mimics the visual receivers generated by the human eyes when observing small target objects, making the detection network more focused on the lesion regions. The contributions can be summarized as follows.

i) A novel TCW-Net with temporal correlation of CT slices is proposed for universal lesion detection.

ii) A weighted feature fusion module is designed, which allocates weights to each slice feature within the fusion process for recalibrating feature responses to target slice.

iii) A multi-branch spatial offset attention mechanism is created to further promote the detection network's ability to discriminate lesions.

iv) The experiments on the widely used datasets, the DeepLesion benchmarks indicate that the proposed method is superior to existing methods.

The rest of this paper is structured as follows. We outline relevant research in Section 2. The TCW-Net for universal lesion detection is thoroughly described in Section 3. The full experimental results are then provided in Section 4. In Section 5, we finally give a summary of this work and offer a few concepts for the direction of future research.

II. RELATED WORK

Universal lesion detection (ULD) is in increasing demand in clinical applications and has a great potential for development in computer-aided diagnosis systems. In recent years, it has been gaining attention from researchers and its performance has been continuously improved [39][40]. Current works typically use 2D networks to extract meaningful information from 2D CT slices, as annotation of medical images is labor-intensive and highly specialized requirement, especially for 3D CT images. Yan et al. [20] proposed a 3D context enhanced regional convolutional neural network, called 3DCE. To exploit details in multiple slices, it incorporated 3D context into a 2D CNN network to extract and aggregate features, and finally used a 2D detection network for prediction. In [6], to extend the feature representation and enhance the discriminability, an inter-slice contextual attention was introduced, which selectively aggregated features of different slices, and focused feature learning on the most salient regions within slices using spatial attention. Shao et al. [2] proposed a multiscale enhancer with improved feature pyramid network using both channel and spatial attention mechanism, where lesion responses from each pyramid feature map were recorded using a distinct expansion rate. Similarly, the MVP-Net proposed by Li et al. [23] also chose to extract multi-scale features to detect small lesions, where the difference was that the MVP-Net derives three window width values suitable for multi-organ lesion detection, which effectively improved the detection metrics. In addition, pseudo 3D has also been used for lesion detection, such as the MP3D FPN proposed by Zhang et al. [3], in which depthwise separable convolutional filters and a group transform module were used to effectively extract the 3D context of CT slices for universal lesion detection. And a new pre-training method for 3D networks was also introduced to accelerate the convergence. Apart from the above-mentioned methods, there are also methods based on domain knowledge enhancement, such as Sheoran et al. [13]. Meanwhile, it is also a common way to refine the anchor mechanism. Zlocha et al. [8] used an evolution searching algorithm to optimize the ratios and scales of the anchor to boost the effectiveness of lesion detection model based on a single stage detector. Sheoran et al. [11] argued that the predefined anchor limited the execution of the detection network, and thus presented a robust single stage anchor-free damage detection network.

One issue that existing methods typically overlook is the temporal correlation between sequential slices, despite its importance in enhancing feature representation. To address this issue, we propose a novel network, named TCW-Net in this study. TCW-Net introduces a Weighted Feature Fusion (WFF) module and a Spatial Offset Attention (SOA) mechanism focusing on lesions and performing weighted fusion, respectively. The WFF module assigns weights to each slice feature to highlight more discriminative features during correlation fusion, and the SOA mechanism simulates the visual receivers generated by the human eyes when observing target objects of smaller sizes. It can significantly improve the detection network's accuracy on lesion regions. The details are described in Section 3 and Section 4.

III. THE PROPOSED METHOD

To improve the sensitivity of the network to lesion information in the TCW-Net, we design a spatial offset attention (SOA) that includes a multi-branch structure and multiple dilated convolutions to simulate the centrally generated receivers of the human retina. We also utilized weight redirection of the weighted feature fusion (WFF) module to obtain the importance of features of different slices to the target slice, allowing the feature fusion process to retain a more favorable representation of the target slice in detection task. Fig. 1 illustrates the framework of the proposed method.

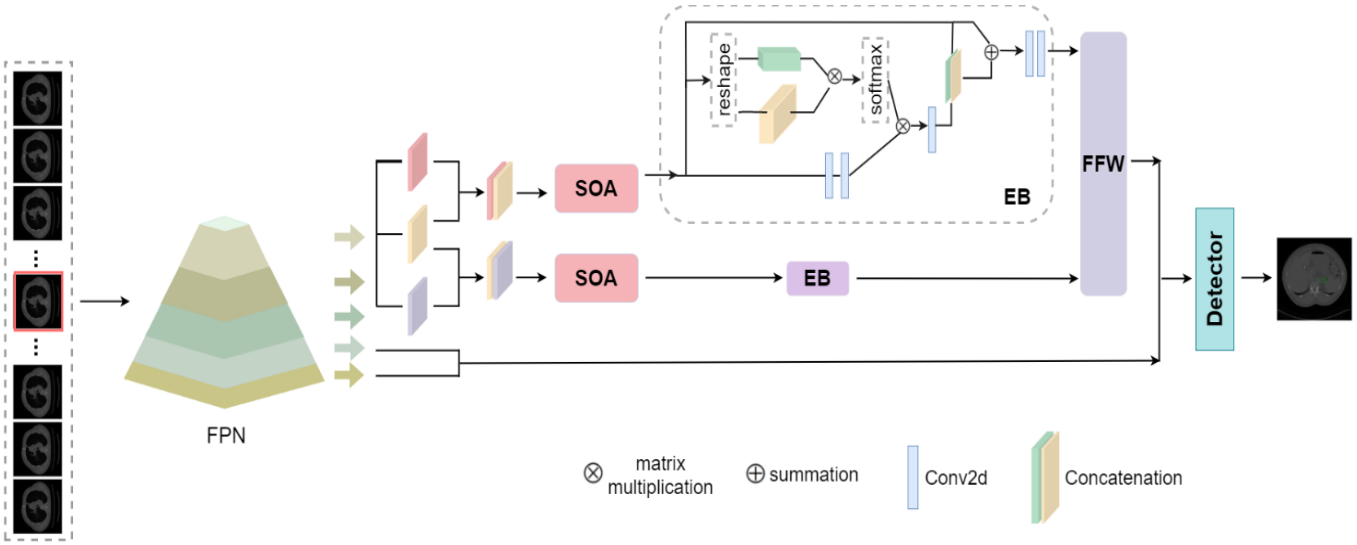

Fig. 1. The framework of the proposed TCW-Net. SOA (Spatial Offset Attention) brings network focus to lesion features, WFF (Weighted Feature Fusion) weighted for feature fusion process

A. Preprocessing

A certain number of sequential CT slices are selected as input, which are arranged successively and cascaded together in the channel dimension. The target slice that needs to be predicted is located in the center of the input sequence.

In universal lesion detection, the lesion size and location are considered as critical factors that affect the detection performance. To deal with this, we employ a pre-trained FPN [25] with ResNet-50 [26]. on ImageNet [24] to extract multi-scale feature maps that capture rich detailed information and abstract semantic information while preparing for the detection of lesions in different scales. Specifically, feature maps of a specific resolution can be obtained at each layer of the FPN, with low-resolution feature maps at the higher layers containing semantic information in CT slices and high-resolution feature maps at the lower layers including the location information of lesions. The feature maps output by the FPN are sequentially defined from the higher to lower levels denoted as F_i , $i = 0,1,...,4$. Then, the higher level feature maps (F_0 to F_2) are uniformly divided into N groups (N is set to 3 in the experiment) in the channel dimension and labelled as $\{G_0, G_1, ..., G_{N-1}\}$, with each pair of adjacent feature groups being cascaded together for presentation.

B. SOA

The human visual system is capable of generating receivers of different sizes to adaptively capture the target and observe the motion of minor objects. To simulate this visual structure, Liu et al. [16] proposed a feature extraction module called Receptive Field Block (RFB). This module allows the detection model to be more sensitive to small movement changes of lesions in CT images, focusing on the potential lesion to be detected. Inspired by this idea, we propose a spatial offset attention (SOA) module adapting to the small size and blurred edges of lesions in CT images. The SOA module aims to further improve the sensitivity of the network to lesion information in the TCW-Net.

We utilize SOA to expand the receptive fields of features obtained by cascading G_0 to G_{N-1} . The process can be expressed mathematically as

$$Q_1 = S(cat(G_0, G_1), ..., cat(G_{N-2}, G_{N-1})), \quad (1)$$

where $S(\cdot)$ is the SOA operation, cat represents cascade and Q_1 denotes the output features of SOA.

The internal structure of the SOA module is shown in Fig. 2. The module adopts a multi-branch pathway design. We use the b_0 branch with a 1×1 convolution kernel to maximize the conservation of the original features. To simulate information processing of different sizes by human eye receivers, we design the b_1 and b_2 branches using convolutional layers with different sizes of kernels k, which extract different semantic and detailed information, respectively. Furthermore, we decompose the $k \times k$ convolution kernels in each branch into two convolutions of $k \times 1$ and $1 \times k$ to reduce the number of parameters. Due to the limited information of the lesion features, we design three dilated convolutions with different dilation rates ($r = 1,2,5$) for expanding the range of the receptive field. This dilation rate can provide continuity between the convolution kernels of the dilated convolution, thus obtaining continuous feature information and avoiding the gridding effect. Then, we use cascade and convolution to fuse the different feature information obtained from the three branches and combine the fused feature maps with the original feature maps input to the spatial offset attention module for residual operations to further extract features that is missed in the earlier stage. This process can ensure capturing the characteristic of the lesion area and improve its discrimination ability accordingly.

C. WFF

The effectiveness of feature information from neighboring slices for lesion detection varies, which requires a weighting operation during feature fusion. To address this issue, we incorporate a weight redirection operation in the WFF module, inspired by the SENet proposed by Hu et al. [16]. This operation allows for adaptive learning of the weight assignment of the feature map, resulting in increased sensitivity of the target slice to informative features from neighboring slices. Obtaining the importance levels of each neighboring slice to the target slice can further improve the network's accuracy in lesion detection.To be specific, the WFF first compresses the features along the dimension in

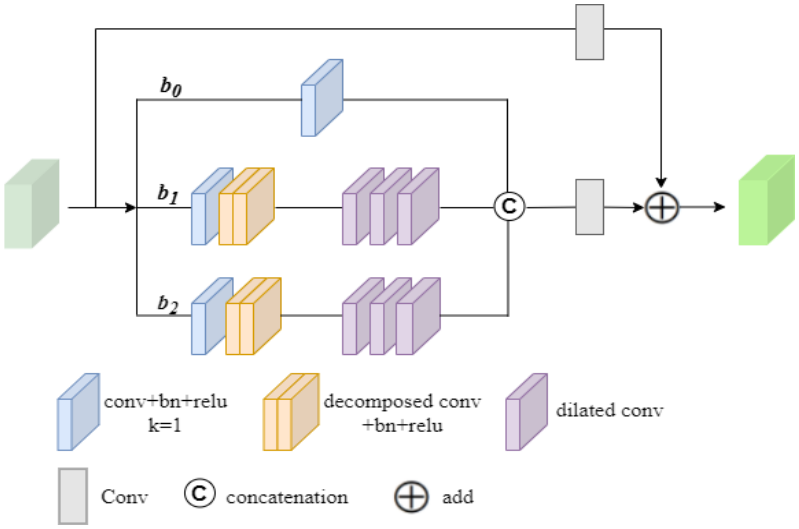

Fig. 2. The pipeline of the SOA module.

which the channel is located, so that each 2D feature becomes a real number, keeping the number of channels constant at this time, while containing certain global receptive fields in each dimension of the output result, which represent the global distribution of the feature map in the channel direction.

Before the above process, we first use enhanced blocks (EBs) to cascade each pair of adjacent feature groups to extract correlation information between sliced features and generate feature groups containing correlation information. Then, to maximize and exploit feature representations that are more favorable to target slices, we design the WFF module to apply weights to each feature channel and fuse the generated $N-1$ feature groups. This process can be formulated as follows:

$$Q_2 = W(E(Q_1)), \tag{2}$$

where $E(\cdot)$ denotes the enhanced block, $W(\cdot)$ is the weighting operation of the WFF module, and Q_2 is the feature map after final processing.

Fig. 3 illustrates the process of the WFF module. It first compresses the features along the channel dimension to generate a 2D feature map with a constant number of channels and certain global receptive fields that represent the global distribution of the feature map in the channel direction. Then, the parameters W following the work by Hu et al. [16] are learned to generate the weights of each channel in the feature map and then calculate correlations between channels. To redirect the weights from the input feature map channels to the feature channels of neighboring slices with respect to the target slice features, a weight redirection operation is performed. This operation converts the individual weights obtained from the input feature map channels to the weights of the feature channels of neighboring slices. The particular conversion operation is formulated as follows.

$$D = \frac{w_i}{w^*}, i = 0,1,...,N. \tag{3}$$

where $\{w_0, w_1, ..., w^*, ..., w_{N-1}, w_N\}$ is the obtained channel weights, and w^* is the average of the channel weights of the feature map corresponding to the target slice, and N is a natural number.

We have derived the feature map distribution that is significant for the target slices so far. Next, we use the output weights as the importance scores of each neighboring slice feature channel and multiply it with the input feature. Finally, we design a residual structure and a convolution layer to retain the original feature information.

D. Total Loss

Finally, we cascade the processed lower-level detailed features , and the higher level semantic features , which are input into the detector to perform the lesion detection. We adopt Faster R-CNN[15] as the detector. We utilize the same bounding-box regression loss and lesion classification loss in Faster R-CNN to train the TCW-Net.

$$L(\{p_i\},\{t_i\}) = \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) + \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*), \tag{4}$$

where λ is weight balance parameter, N_{cls} denotes the size of mini-batch, N_{reg} represents the number of anchor locations. When anchor [i] is a positive sample, parameter $p_i^* = 1$; otherwise, it takes 0. The parameterized coordinates of the predicted bounding box are denoted by t_i , while t_i^* and p_i stand for the ground truth and the predicted classification probability of anchor [i] respectively. Loss L_{reg} uses smooth L_1 , and L_{cls} adopts cross entropy. All these parameters are optimized jointly in the process.

IV. Experiments

A. Datasets and Evaluation Metrics

DeepLesion [17] is a dataset of clinical medical CT images with lesion-level annotations, which is the largest of its kind that the NIH Clinical Center has ever published. The dataset includes a total of 10,594 CT scans from 4,427 anonymous patients, with 32,735 labelled lesions across multiple categories. Three sets of the data are created: 70% are used for training, 15% are used for validation, and 15% are used for testing. By calculating the average sensitivity for various false positive rates on the training set, we assess how well it performed on the test set. This evaluation metric is employed in order to compare our model's performance with that of existing methods.

In our framework, we utilize FPN with pre-trained ResNet-50 on ImageNet [24] and train the model on the NVIDIA GeForce RTX 3090 GPU. Our RPN has five anchor scales (16, 32, 64, 128, 256) and three anchor ratios (0.5, 1.0, 2.0), with an IOU threshold value being set to 0.7. Additionally, the input slice is resized to a resolution of 800×800 pixels.

Our model is trained using the stochastic gradient descent (SGD) optimizer with 16 epochs. The initial learning rate is

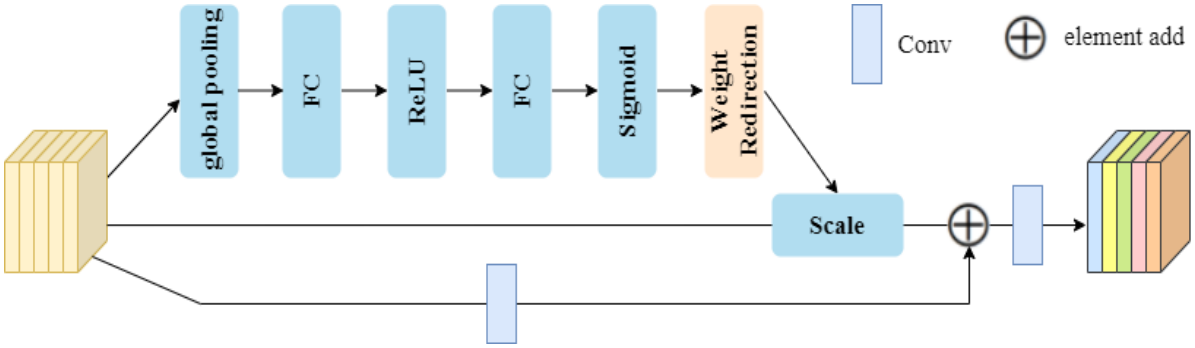

Fig. 3. The architecture of the proposed WFF module.

set to 0.002, and then it is gradually reduced after the 11th and 13th iterations by a factor of 10.

B. Comparison with Other Methods

We compare the proposed TCW-Net with contemporary techniques, including GPN [19], 3DCE [20], ULDOR [21], LENS [22], Deformable Dilated Faster R-CNN [4] and MVP-Net [23]. As shown in Table 1, we perform the evaluation by computing the sensitivity values at various false positives rates.

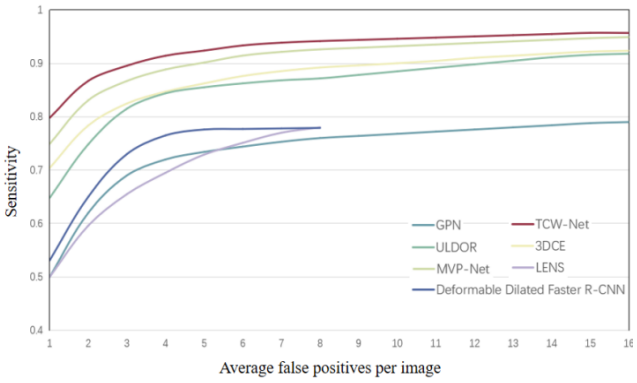

Fig. 4. FROC curves of various methods on the DeepLesion.

a) Quantitative results: Table 1 presents the comparison results of TCW-Net's prediction performance with other ULD networks on the DeepLesion test set. The results indicate that TCW-Net has clear advantages over other networks in lesion detection sensitivity regardless of false positive rates, which demonstrates the effectiveness of the spatial offset attention module and the weighted feature fusion module in the fusion process. Specifically, when nine slices are used as input, TCW-Net outperforms other networks, with an improvement of 5.78% and 4.88% in FPs@0.5 and FPs@1.0 indicators respectively, which leads to a lower rate of leak check of lesion. Moreover, the sensitivity values at various false positive rates are converted into the FROC curve, as shown in Fig. 4. The area under the TCW-Net curve is larger than that of other methods, which further confirms the improvement of the model's performance indicators.

b) Qualitative results: The prediction results of TCW-Net have been compared with other ULD networks, and the visualization of lesion prediction is demonstrated using CT images of various organs in Fig. 5. The results indicate that TCW-Net outperforms other detection models in predicting lesion sizes of different scales across different tissue organs with higher accuracy. TCW-Net also shows a good performance in lesion localization.

TABLE I. ON THE DEEPLESION TEST SET, THE PROPOSED METHOD IS COMPARED WITH THE STATE-OF-THE-ART METHODS. LESION DETECTION SENSITIVITY VALUES ARE NOW REPORTED AT DIFFERENT FALSE POSITIVE (FP) RATES. THE NUMBER AFTER "," DENOTES THE NUMBER OF SLICES.

Methods	**FP@**						
	0.5	**1**	**2**	**3**	**4**	**8**	**16**
GPN[19]	0.36	0.5	0.62	-	0.72	0.76	0.79
3DCE[20], 3	0.522	0.6297	0.7351	0.7881	0.8263	0.8919	0.9294
3DCE, 9	0.5482	0.6547	0.7459	0.7897	0.8166	0.8719	0.9151
3DCE, 27	0.5967	0.704	0.7836	0.8247	0.8469	0.8921	0.9235
ULDOR[21]	0.5286	0.648	0.7484	-	0.8438	0.8717	0.918
LENS[22]	0.403	0.5	0.596	-	0.695	0.78	
Deformable Dilated Faster R-CNN[4]	0.427	0.531	0.65	-	0.775	0.775	-
MVP-Net[23], 3	0.5704	0.6893	0.796	0.8394	0.8634	0.9161	0.9442
MVP-Net, 9	0.6224	0.7492	0.8308	0.8669	0.8884	0.9261	0.9491
TCW-Net, 9	**0.6802**	**0.798**	**0.8671**	**0.8955**	**0.9142**	**0.9416**	**0.9568**

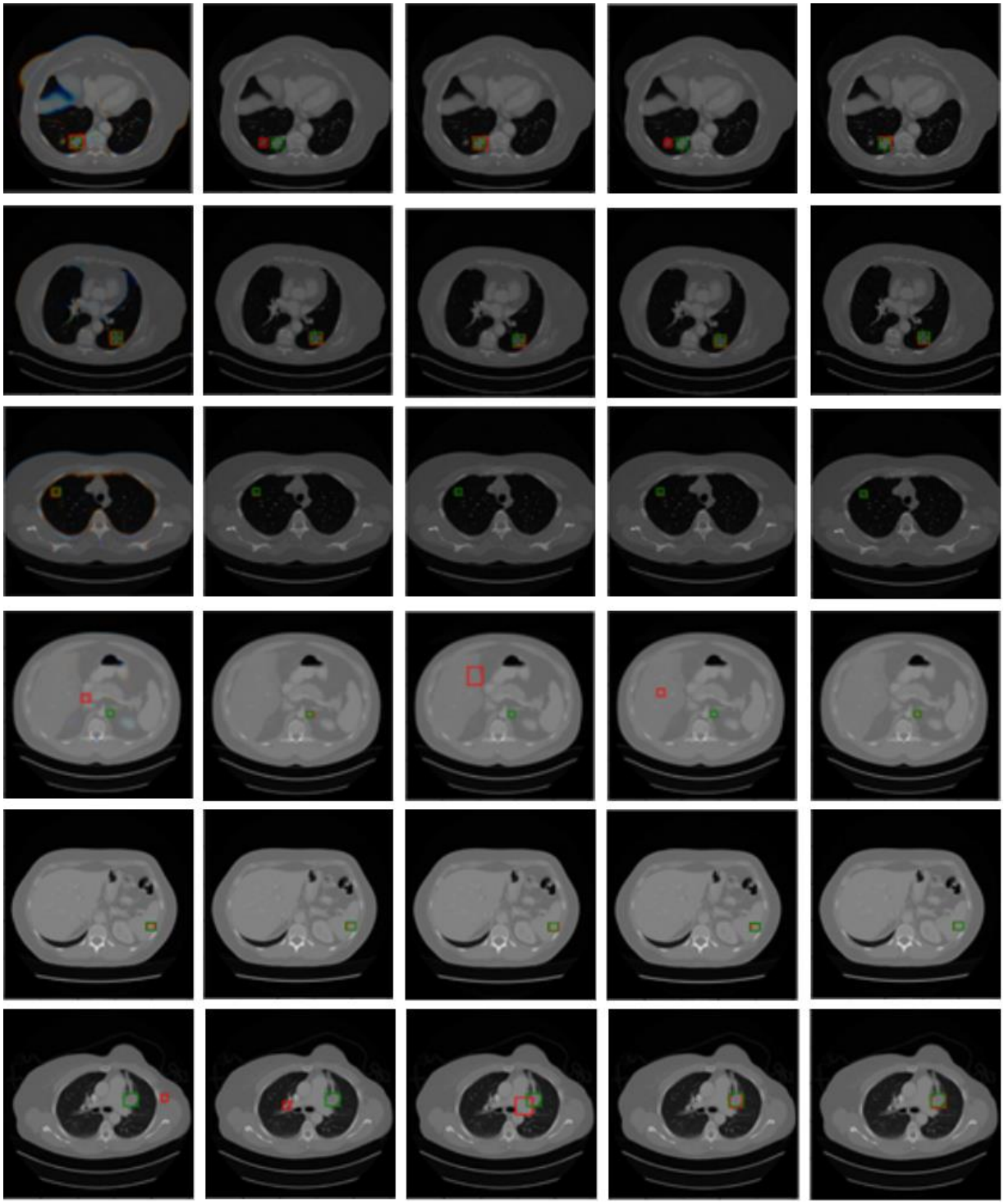

Fig. 5. Qualitative comparison of several existing ULD models and ours on the testing set of DeepLesion. Green and red boxes correspond to ground-truths in the test set and predicted true positives. The number after "," denotes the number of slices.

TABLE II PERFORMANCE OF THE PROPOSED METHOD FOR DOING ABLATION EXPERIMENTS ON THE DEEPLESION DATASET.

Methods	**FP@**						
	0.5	**1**	**2**	**3**	**4**	**8**	**16**
FPN	0.5943	0.7136	0.8101	0.8498	0.8748	0.9194	0.9497
FPN+SOA	0.6722	0.7889	0.8658	0.8919	0.9127	0.9217	0.956
FPN+WFF	0.6769	0.7914	0.8666	0.8927	0.9135	0.9225	0.9566
TCW-Net	**0.6802**	**0.798**	**0.8671**	**0.8955**	**0.9142**	**0.9416**	**0.9568**

C. Ablation Experiment

We have conducted ablation experiments on TCW-Net using the DeepLesion test set to evaluate the impact of incorporating the SOA mechanism and WFF module. Specifically, we compare the performance of adding SOA on the basis of FPN [25], adding WFF, and the final TCW-Net. The results, presented in Table 2, demonstrate that the TCW-Net significantly improves performance at different false positive rates. These findings suggest that paying attention to the subtle offset of lesions and assigning fusion weights to features can enhance the contextual and content information

of slice features, improve the expression of features, and optimize the prediction results of the detection model.

V. Conclusion

In this study, we propose a temporal correlation weighted fusion lesion detection network (TCW-Net) for ULD tasks in CT images. The WFF module is introduced to better emphasize the feature representations that are more important for the target slice by weighting multiple feature channels containing temporal correlations. Moreover, a SOA mechanism is designed to enhance the capture of lesion information. The findings from the experiments validate the effectiveness of the proposed TCW-Net for lesion detection tasks. However, the detection of lesions is a challenging task that involves multiple organs, and the dataset should have a balanced amount of CT images of diverse organs to boost the network's generalization capability. In the future, we would increase CT images of organs to the dataset to further improve the performance of TCW-Net in future research.

Acknowledgment

This work was supported by National Natural Science Foundation of China (NSFC) (61976123, 61601427); Taishan Young Scholars Program of Shandong Province; and Key Development Program for Basic Research of Shandong Province (ZR2020ZD44).

References

[1] J. Hsieh, "Computed tomography: principles, design, artifacts, and recent advances," 2003.

[2] Q. Shao, L. Gong, K. Ma, H. Liu, Y. Zheng, "Attentive CT Lesion Detection Using Deep Pyramid Inference with Multi-scale Booster," Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11769, pp. 301-309, 2019.

[3] S. Zhang, J. C. Xu, Y. C. Chen, J. C. Ma, Y. Z. Yu, "Revisiting 3D Context Modeling with Supervised Pre-training for Universal Lesion Detection in CT Slices," 2020.

[4] F. Hellmann, Z. Ren, E. André, B. W. Schuller, "Deformable Dilated Faster R-CNN for Universal Lesion Detection in CT Images," 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, pp. 2896-2902, 2021.

[5] Z. Liu, K. Han, K. F. Xue, Y. Q. Song, L. Liu, Y. Y. Tang, Y. Zhu, "Improving CT-image universal lesion detection with comprehensive data and feature enhancements," Multimedia Systems, vol. 28, pp. 1741-1752, 2022.

[6] Q. Y. Tao, Z. Y. Ge, J. F. Cai, J. X. Yin, S. See, "Improving Deep Lesion Detection Using 3D Contextual and Spatial Attention," Medical Image Computing and Computer Assisted Intervention – MICCAI, vol. 11769, pp. 185-193, 2019.

[7] Z. F. Qiu, T. Yao, T. Mei, "Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks," IEEE International Conference on Computer Vision (ICCV), pp. 5534-5542, 2017.

[8] M. Zlocha, Q. Dou, B. Glocker, "Improving RetinaNet for CT Lesion Detection with Dense Masks from Weak RECIST Labels," Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11769, pp. 402–410, 2019.

[9] H. Li, H. Han, S. K. Zhou, "Bounding Maps for Universal Lesion Detection," Medical Image Computing and Computer Assisted Intervention – MICCAI, vol. 12264, pp. 417–428, 2020.

[10] H. Li, L. Chen, H. Han, Y. Chi, S. K. Zhou, "Conditional Training with Bounding Map for Universal Lesion Detection," Medical Image Computing and Computer Assisted Intervention – MICCAI, vol. 12905, pp. 141–152, 2021.

[11] M. Sheoran, M. Dani, M. Sharma, L. Vig, "An Efficient Anchor-Free Universal Lesion Detection in Ct-Scans," IEEE 19th International Symposium on Biomedical Imaging (ISBI), pp. 1-4, 2022.

[12] J. Z. Cai, K. Yan, C. T. Cheng, J. Xiao, C. H. Liao, L. Lu, A. P. Harrison, "Deep Volumetric Universal Lesion Detection Using Light-Weight Pseudo 3D Convolution and Surface Point Regression," Medical Image Computing and Computer Assisted Intervention – MICCAI, vol. 12264, pp. 3-13, 2020.

[13] M. Sheoran, M. Dani, M. Sharma, L. Vig, "DKMA-ULD: Domain Knowledge augmented Multi-head Attention based Robust Universal Lesion Detection," arXiv preprint arXiv:2203.06886, 2022.

[14] X. L. Cheng, H. Xiong, D. P. Fan, Y. R. Zhong, M. Harandi, T. Drummond, Z. Y. Ge, "Implicit Motion Handling for Video Camouflaged Object Detection," Conference on Computer Vision and Pattern Recognition, pp. 13854-13863, 2022.

[15] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 39(6), pp. 1137-1149, 2016.

[16] S. Liu, D. Huang, "Receptive field block net for accurate and fast object detection," Proceedings of the European conference on computer vision (ECCV), vol. 11215, pp. 385-400, 2018.

[17] J. Hu, L. Shen, G. Sun, "Squeeze-and-excitation networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132-7141, 2018.

[18] K. Yan, X. S. Wang, L. Lu, L Zhang, A. P. Harrison, M. Bagheri, R. M. Summers, "Deep Lesion Graphs in the Wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-Scale Lesion Database," Computer Vision and Pattern Recognition (CVPR), pp. 9261-9270, 2018.

[19] Y. Li, "Detecting Lesion Bounding Ellipses with Gaussian Proposal Networks," Machine Learning in Medical Imaging, vol. 11861, pp. 337-344, 2019.

[20] K. Yan, M. Bagheri, R. M. Summers, "3D Context Enhanced Region-Based Convolutional Neural Network for End-to-End Lesion Detection," Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, vol. 11070, pp. 511-519, 2018.

[21] Y. B. Tang, K. Yan, Y. X. Tang, J.M. Liu, J. Xiao, R. M. Summers, "Uldor: A Universal Lesion Detector For Ct Scans With Pseudo Masks And Hard Negative Example Mining," IEEE 16th International Symposium on Biomedical Imaging, pp. 833-836, 2019.

[22] K. Yan, J. Z. Cai, Y. J. Zheng, A. P. Harrison, D. K. Jin, Y. B. Tang, Y. X. Tang, L. Y. Huang, J. Xiao, L. Lu, "Learning From Multiple Datasets With Heterogeneous and Partial Labels for Universal Lesion Detection in CT," IEEE Transactions on Medical Imaging, vol. 40, pp. 2759-2770, 2021.

[23] Z. H. Li, S. Zhang, J. G. Zhang, K. Q. Huang, Y. Z. Wang, Y. Z. Yu, "MVP-Net: Multi-view FPN with Position-Aware Attention for Deep Universal Lesion Detection," Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, vol. 11769, pp. 13-21, 2019.

[24] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. F. Li, "ImageNet: A large-scale hierarchical image database," IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255, 2009.

[25] T. Y. Lin, P. Dollár, R. Girshick, K. M. He, B. Hariharan, S. Belongie, "Feature Pyramid Networks for Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936-944, 2017.

[26] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

[27] Á. S. Hervella, J. Rouco, J. Novo, M. Ortega, "Retinal microaneurysms detection using adversarial pre-training with unlabeled multimodal images", Information Fusion, vol. 79, pp. 146-161, 2022.

[28] D. Zhang, M. Ye, Y. G. Liu, L. Xiong, L.H. Zhou, "Multi-source unsupervised domain adaptation for object detection", Information Fusion, vol. 78, pp. 138-148, 2022.

[29] M. R. Hassan, S. Huda, M. M. Hassan, J. Abawajy, A. Alsanad, G. Fortino, "Early detection of cardiovascular autonomic neuropathy: A multi-class classification model based on feature selection and deep learning feature fusion", Information Fusion, vol. 77, pp. 70-80, 2022.

[30] G. Li, J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," Information Fusion, vol. 91, pp. 93-102, 2023.

[31] X. Li, M. L. Li, P. F. Yan, G. Y. Li, Y. C. Jiang, H. Luo, S. Yin, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," IJNDI, vol. 2, pp. 93–116, 2023.

[32] B. Rahi, M. Z. Li, M. Qi, “A Review of Techniques on Gait-Based Person Re-Identification,” IJNDI, vol. 2, pp. 66–92, 2023.

[33] F. M. Shakiba, M. Shojaee, S. M. Azizi, M. Zhou, “Real-Time Sensing and Fault Diagnosis for Transmission Lines,” IJNDI, vol. 1, pp. 36–47, 2022.

[34] B. Rahi, M. Li, M. Qi, “A Review of Techniques on Gait-Based Person Re-Identification,” IJNDI, vol. 2, pp. 66–92, 2023.

[35] M. W. Jian, H. Y. Chen, C. Tao, X. G. Li, G. G. Wang, “Triple-DRNet: A triple-cascade convolution neural network for diabetic retinopathy grading using fundus images,” Comput. Biol. Medicine, vol. 155, pp. 106631-106640, 2023.

[36] Y. C. Yin, Z. M. Han, M. W. Jian, G. G. Wang, L. Y. Chen, R. Wang, “AMSUnet: A neural network using atrous multi-scale convolution for medical image segmentation,” Comput. Biol. Medicine, vol. 162, pp. 107120, 2023.

[37] Z. M. Han, M. W. Jian, G. G. Wang, “ConvUNeXt: An efficient convolution neural network for medical image segmentation,” Knowl. Based Syst, vol. 253, pp. 109512-109517, 2022.

[38] M. W. Jian, L. S. Zhang, H. D. Jin, X. G. Li, “3DAGNet: 3D Deep Attention and Global Search Network for Pulmonary Nodule Detection,” Electronics, vol. 12, pp. 2333-2347, 2023.

[39] M. W. Jian, R. H. Wu, H. Y. Chen, L. Q. Fu, C. D. Yang, “Dual-Branch-UNet: A Dual-Branch Convolutional Neural Network for Medical Image Segmentation,” CMES-Computer Modeling in Engineering & Sciences, vol. 137, 2023.

[40] S. Yu, J. Y. Cong, K. X. Zhang, M. W. Jian, B. Z. Wei, “Unsupervised medical image feature learning by using de-melting reduction auto-encoder,” Neurocomputing, vol. 523, pp. 145-156, 2023.