MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231095

MSPA-DLA++: A Multi-Scale Phase Attention Deep Layer Aggregation for Lesion Detection in Multi-Phase CT Images

Titinunt KITRUNGROTSAKUL[a], Yingying XU[a], Qingqing CHEN[b], Jing LIU[a], Yinhao LI[c], Lanfen LIN[d], Hongjie HU[b], Ruofeng TONG[d], Jingsong LI[a] and Yen-Wei CHEN[c1]

[a]*Research Center for Healthcare Data Science, Zhejiang Lab, China*
[b]*Department of Radiology, Sir Run Run Shaw Hospital, China*
[c]*Graduate School of Information Science and Engineering, Ritsumeikan Univ., Japan*
[d]*College of Computer Science and Technology, Zhejiang Univ., China*

Abstract. Object detection using convolutional neural networks (CNNs) has achieved high performance and achieved state-of-the-art results with natural images. Compared to natural images, medical images present several challenges for lesion detection. First, the sizes of lesions vary tremendously, from several millimeters to several centimeters. Scale variations significantly affect lesion detection accuracy, especially for the detection of small lesions. Moreover, the effective extraction of temporal and spatial features from multi-phase CT images is also an important issue. In this paper, we propose a group-based deep layer aggregation method with multiphase attention for liver lesion detection in multi-phase CT images. The method, which is called MSPA-DLA++, is a backbone feature extraction network for anchor-free liver lesion detection in multi-phase CT images that addresses scale variations and extracts hidden features from such images. The effectiveness of the proposed method is demonstrated on public datasets (LiTS2017) and our private multiphase dataset. The results of the experiments show that MSPA-DLA++ can improve upon the performance of state-of-the-art networks by approximately 3.7%.

Keywords. CT Imaging, detection, mullti-phase, attention

1. Introduction

Object detection using convolutional neural networks has recently achieved high-performance results. We can categorize object detection networks based on their detection procedures into two categories: anchor-based [1-7] and anchor-free [8-10] detection methods. Compared to natural images, medical images present several challenges for lesion detection. First, the sizes of lesions vary tremendously, from several millimeters to several centimeters. Scale variations significantly affect lesion detection accuracy, especially for the detection of small lesions. Moreover, the effective extraction of spatial and temporal features from multiphase CT images is also an important issue.

[1] Corresponding Author: Yen-Wei CHEN, email: chen@is.ritsumei.ac.jp

2. Methods

Figure 1 shows the proposed network's overview. There are two parts in our proposed detection network: a feature extraction network and a detection parts. In this research, we focus our research on the feature extraction network, which is called MSPA-DLA++. MSPA-DLA++ is proposed to solve the scale variation problem by extracting phase attention features with group-based multiscale features from multi-phase CT images. The MSPA-DLA++ can be used or combined with any detection head.

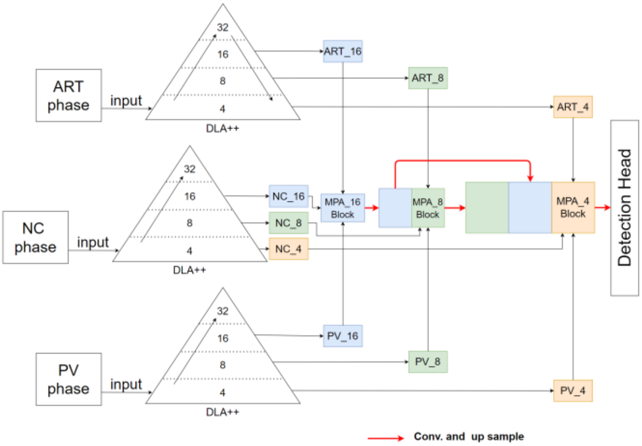

Figure 1. Overview of the architecture of MSPA-DLA++.

2.1. A Full-Scale Connected Deep Layer Aggregation (DLA++) Network

The original DLA and CenterNet's DLA used plain connections or deformable convolution. The purpose of the DLA network was to address the problems faced by other skip connection networks, which use linear skip connections to pass features with the same scale from the lower layer to the upper layer. However, after the feature passes each network layer, these networks still lose some information. To maintain and reuse full-scale features, we follow our originally proposed baseline network of DLA++ [11] which is use the concept derived from DenseNet [12].

2.2. Phase Attention (PA)

To utilize the features from each phase, a multiphase combination technique is applied in our work. In this work, we propose phase attention (PA). PA is used to extract enhancement patterns and inter-phase features. In our PA module, a squeeze and excitation network (SENet) [13] is applied and used to extract channel attention features for each phase except the NC phase.

$$F^{C,s}_{x_1,x_2} = F^{s}_{x_1} \otimes \sigma\left(MLP\left(AvgPool\left(F^{s}_{x_2}\right)\right)\right) \qquad (1)$$

where $F^{C,s}_{x}$ denotes a feature at scale *s* in the *x* phase, *MPL* denotes a multilayer perceptron, *AvgPool* denotes the average pooling operation, and σ denotes the sigmoid function. $F^{C}_{x_1,x_2}$denotes the channel attention feature of phase x_2 elementwise multiplied (⊗) with the features of phase x_1. In our work, elementwise multiplication is applied to the NC phase features after extracting the channel attention features of each phase. NC-ART channel features and NC-PV channel features are generated.

We extract spatial attention features from the NC-ART channel and NC-PV channel features based on the concept of the CBAM, in which average-pooling, max-pooling, and convolution layers are used to generate spatial attention F^S.

$$\mathrm{F}^{\mathrm{S,s}}_{\mathrm{x_1,x_2}} = \mathrm{f}^{7\times7}\left(\left[\mathrm{AvgPool}\left(\mathrm{F}^{\mathrm{C,s}}_{\mathrm{x_1,x_2}}\right), \mathrm{MaxPool}\left(\mathrm{F}^{\mathrm{C,s}}_{\mathrm{x_1,x_2}}\right)\right]\right) \tag{2}$$

where $f^{7\times7}$refers to a convolution operation with a 7×7 filter size, *AvgPool,* and *MaxPool* are the average-pooling operation and max-pooling operation, respectively, and the concatenation operation is denoted as [·]. After the NC-ART and NC-PV spatial features are generated, we merge both sets of features using elementwise summation. The sigmoid function (σ) and elementwise multiplication (⊗) are applied to merge the NC phase features into the final features $\hat{F}^{s}_{NC}$, as shown in Figure 2.

$$\hat{\mathrm{F}}^{\mathrm{s}}_{\mathrm{NC}} = \mathrm{F}^{\mathrm{s}}_{\mathrm{NC}} \otimes \sigma\left(\mathrm{F}^{\mathrm{S,s}}_{\mathrm{NC,ART}} + \mathrm{F}^{\mathrm{S,s}}_{\mathrm{NC,PV}}\right) \tag{3}$$

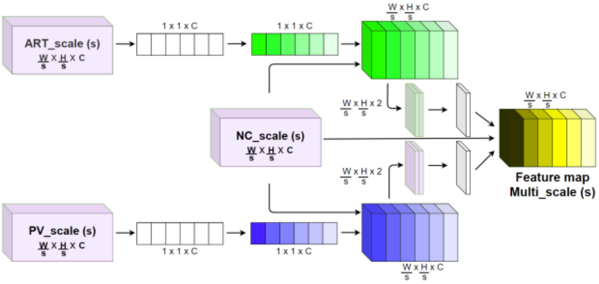

Figure 2. The structure of the phase attention module.

2.3. MSPA-DLA++: Multi-Scale Phase Attention Deep Layer Aggregation

Figure 1 illiterates the MSPA-DLA++ method. PA is used to extract phase features ($\hat{F}^{s}_{NC}$) from multi-phase CT images at each scale s in the DLA++ network. The 1×1 convolution and upsampling processes are used to transform the features. We concatenate the small features and large-scale features with the concept of residual connections to retain the lower features in the final decision. The CT data are abdominal CT scans from 118 patients. Each data point from the patients comprises 3 phases (NC, PV and ART phases).

3. Results

In this paper we evaluate our method on our private multi-phase CT dataset from Sir Run Run Shaw Hospital collected from 2015 through 2017. In addition to our private dataset, we evaluate the detection methods on LiTS2017, a public dataset from MICCAI that is used to evaluate segmentation performance with respect to liver lesions. We use the segmentation labels to generate detection ROIs for the lesion detection task.

We compare our MSPA-DLA++ with several state-of-the-art methods: SSD [1], Faster RCNN [2], GSSD [5], DeepLung [14], 3DCE [15], CornerNet [8], ExtremeNet [9], and CenterNet [10]. In Table 1, the LiTS2017 and our private dataset columns show the evaluation performance of the tested detection methods. The results show that the detection performances on LiTS2017 are similar to those on our private dataset; the anchor-based methods perform worse than the anchor-free methods. However, the performance of the DeepLung and 3DCE methods is better than that of other methods, including both anchor-based and anchor-free methods. We investigate the differences between the two datasets. We find that some lesions in the LiTS2017 dataset have curved shapes, which makes them look like two lesions instead of one. The lesion shape problem makes detection methods detect such lesions as two lesion objects. The detection methods can detect some parts of the lesion (two fragment lesions), but when we calculate the performance using the AP@0.5 and 0.75 metrics, the sizes of the detection boxes are insufficient to count as true positives. Regarding the 3DCE and our proposed method, additional spatial information can assist these detection methods in calculating correct detection results. As shown in Table 1, the AP@0.75 results of DeepLung, the 3CDE, and our PV MSPA-DLA++ are approximately 45–48%, while those of the anchor-free methods are 30–35%, and those of the anchor-based methods are 42–45%.

Table 1. Comparison of detection result with state-of-the-art methods.

Method	**Sir Ran Ran Shaw Hospital**		**LiTS201**	
	AP@0.5	**AP@0.7**	**AP@0.5**	**AP@0.7**
SSD[1]	51.48± 1.78	51.48± 1.78	51.48± 1.78	51.48± 1.78
Faster RCNN[2]	66.38±4.10	41.82±10.97	58.24±3.91	35.74±9.93
GSSD[5]	57.81±1.99	46.25±4.88	50.21±4.81	34.52±10.04
DeepLung[14]	68.01±9.91	47.44±9.19	70.29±4.79	45.41±7.89
3DCE, 27 slices [15]	68.09±1.70	47.60±4.78	69.22±4.90	38.72±8.91
3DCE multiphase [15]	83.68±3.78	59.48±5.35	-	-
CornerNet [8]	72.48±4.76	49.54±7.33	64.56±5.01	42.73±9.13
ExtremeNet [9]	75.52±3.30	56.91±5.93	66.87±5.77	45.98±6.90
CenterNet [10]	81.07±3.89	53.29±6.01	73.55±4.21	45.31±7.27
PV MSPA-DLA++	-	-	**74.00±4.23**	**48.25±6.32**
MSPA-DLA++	**84.26±3.02**	**60.08±5.16**	-	-

4. Discussion

To utilize the medical image features, multi-phase features, and improve the performance of detection network, we propose a multi-scale phase attention deep layer aggregation (MSPA-DLA++) for lesion detection in multi-phase CT images. Our proposed network is designed and developed based on our baseline network, DLA++. To extract channel information from the PV and ART phases to capture the enhancement patterns of the lesion, we proposed phase attention(PA).

5. Conclusions

By combining multi-phase attentions under the concept of multiscale phase attention, which is try to concatenate all scale features and multi-phase attention features, we can obtain more accurate results than a result from single phase attention.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (62202435, 82071988), in part by China Postdoctoral Science Foundation (2020TQ0293), in part by Key Research Project of Zhejiang Lab (2022ND0AC01).

References

[1] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single Shot MultiBox Detector. In: Leibe B, Matas J, Sebe N, Welling M, editors. European Conference on Computer Vision (ECCV); Lecture Notes in Computer Science, vol 9905. Cham: Springer; 2016, doi: 10.1007/978-3-319-46448-0_2.

[2] Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal. In: Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS'15); Cambridge, MA, USA: MIT Press; 2015. p. 91-9, doi: 10.5555/2969239.2969250.

[3] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA; 2016. p. 779-88, doi: 10.1109/CVPR.2016.91.

[4] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: IEEE International Conference on Computer Vision (ICCV); Venice, Italy; 2017. p. 2980-8, doi: 10.1109/ICCV.2017.322.

[5] Lee S, Bae JS, Kim H, Kim JH, Yoon S. Liver lesion detection from weakly-labeled multi-phase CT volumes with a grouped single shot multibox detector. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G, editors. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI); MICCAI 2018. Lecture Notes in Computer Science, vol 11071. Cham: Springer; 2018, doi:10.1007/978-3-030-00934-2_77.

[6] Liang D, Lin1 L, Chen X, Hu H, Zhang Q, Chen Q, Iwamoto Y, Han X, Chen YW, Tong R, Wu J. Multi-stream scale-insensitive convolutional and recurrent neural networks for liver tumor detection in dynamic ct images. In: IEEE International Conference on Image Processing (ICIP); Taipei, Taiwan; 2019. p. 794-8, doi: 10.1109/ICIP.2019.8803730.

[7] Deng D, Liu H, Li X, Cai D. PixelLink: detecting scene text via instance segmentation. AAAI [Internet]. 2018 Apr 27 [cited 2023 Apr 4];32(1). Available from: https://ojs.aaai.org/index.php/AAAI/article/view/12269, doi: 10.1609/aaai.v32i1.12269.

[8] Law H, Deng J. CornerNet: detecting objects as paired keypoints. Int J Comput Vis. 2019;128:642-56, doi:10.1007/s11263-019-01204-1.

[9] Zhou X, Zhuo J, Krahenbuhl P. Bottom-up object detection by grouping extreme and center points. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Long Beach, CA, USA; 2019. p. 850-9, doi: 10.1109/CVPR.2019.00094.

[10] Zhou X, Wang D, Krahenbuhl P. Objects as points. arXiv preprint arXiv:1904.07850. 2019; doi: 10.48550/arXiv.1904.07850.

[11] Kitrungrotsakul T, Iwamoto Y, Takemoto S, Yokota H, Ipponjima S, Nemoto T, Lin L, Tong R, Li J, Chen YW. Accurate and fast mitotic detection using an anchor-free method based on full-scale connection with recurrent deep layer aggregation in 4D microscopy images. BMC Bioinformatics. 2021 Feb;22(1):91, doi: 10.1186/s12859-021-04014-w.

[12] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI, USA; 2017. p. 2261-9, doi: 10.1109/CVPR.2017.243.

[13] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Salt Lake City, UT, USA; 2018. p. 7132-41, doi: 10.1109/CVPR.2018.00745.

[14] Gu Y, Lu XQ, Yang LD, Zhang BH, Yu DH, Zhao Y, Gao L, Wu L, Zhou T. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. Comput Biol Med. 2018 Dec;103:220-31, doi: 10.1016/j.compbiomed.2018.10.011.

[15] Yan K, Bagheri M, Summers RM. 3D context enhanced region based convolutional neural network for end-to-end lesion detection. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention (MICCAI); Lecture Notes in Computer Science, vol 11070. Cham: Springer, MICCAI 2018; 2018. p. 511–9, doi:10.1007/978-3-030-00928-1_58.