2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Dual Skip Connections Minimize the False Positive Rate of Lung Nodule Detection in CT images

Jiahua Xu*, Philipp Ernst*, Tung Lung Liu*, Andreas Nürnberger
Data & Knowledge Engineering Group, Faculty of Computer Science
Otto von Guericke University Magdeburg, Germany
{Jiahua.Xu, Philipp.Ernst, Andreas.Nuernberger}@ovgu.de
tung-lung.liu@st.ovgu.de

***Abstract*— Pulmonary cancer is one of the most commonly diagnosed and fatal cancers and is often diagnosed by incidental findings on computed tomography. Automated pulmonary nodule detection is an essential part of computer-aided diagnosis, which is still facing great challenges and difficulties to quickly and accurately locate the exact nodules' positions. This paper proposes a dual skip connection upsampling strategy based on Dual Path network in a U-Net structure generating multiscale feature maps, which aims to minimize the ratio of false positives and maximize the sensitivity for lesion detection of nodules. The results show that our new upsampling strategy improves the performance by having 85.3% sensitivity at 4 FROC per image compared to 84.2% for the regular upsampling strategy or 81.2% for VGG16-based Faster-R-CNN.**

Pulmonary nodule detection, Dual skip connections, Dual Path U-Net, Region Proposal Network.

I. Introduction

Pulmonary cancer is one of the most commonly diagnosed and fatal cancers among other cancers in medical research [1]. Other pulmonary diseases, such as COVID-19 or pulmonary infection, may also cause serious damage to the lung. The most common problem during diagnostic in radiology is solitary pulmonary nodules, i.e. single, round or oval nodules generally smaller than $3\,\mathrm{cm}$ [2].

Computed tomography (CT) is one of the most common non-invasive screening approaches for diagnosing pulmonary diseases [3]. Pulmonary nodules that appear on the images have a high variability in terms of size, shape and location in the pulmonary regions [4]. Small nodules are very difficult to observe because there are many other tissues in the thorax (e.g., blood vessels, airways, lymph nodes) with morphological features similar to nodules [5]. It is challenging to reduce misdiagnoses and false positives (FPs) in early-stage lung cancer diagnosis [6]. The main issue which leads to a high ratio of FPs, particularly for pulmonary nodule detection, comes from the variability of nodules in terms of size, shape and location [4] and, compared to regular RGB images, gray-scale medical images provide less information in terms of edges and textures to distinguish different tissues [7].

In recent researches, deep learning has greatly improved performance and efficiency in nodule detection. Variations based on U-Net [8], Faster R-CNN [9], 3D-CNN [10], and 3D-Dual Path Network [11] have been reported. One study [12] utilized neighboring slices to extract the volumetric and contextual information around the nodules as well as keeping the computational effort of the method low compared to 3D models that have a larger number of network parameters.

In this paper, we propose a dual skip connection upsampling strategy using Faster R-CNN [13] with Dual Path Network (DPN) [14] as the backbone in a U-Net [15] structure.

*Contributed equally

II. Method

A. Dataset: DeepLesion

DeepLesion [16] was released by the National Institutes of Health Clinical Center. It consists of 32,120 axial CT slices from 10,594 CT scans (studies) of 4,427 unique patients. We have extracted CT images that are annotated with pulmonary nodules, which resulted in 2,394 CT images for our dataset in our case. 1,916 CT images are used for training and 478 CT images for validation.

B. Data Preprocessing

The CT slices of DeepLesion are normalized by converting Hounsfield units to mass attenuation coefficients and dividing by the 99th percentile of the entire dataset. Each slice has $1\,\mathrm{mm}$ to $5\,\mathrm{mm}$ thickness in most cases while some of the images are $0.625\,\mathrm{mm}$ or $2\,\mathrm{mm}$. For each lesion, there is one key slice with $30\,\mathrm{mm}$ of extra slices in front of and behind the key slice. However, only the key slice has the annotation data including lesion types, coordinates of 2D bounding-boxes and RECIST diameters for the lesions. The CT slices were resized to $512\,\mathrm{px} \times 512\,\mathrm{px}$. For training, we adapted data augmentation where images are flipped horizontally and vertically and rotated with a probability of 50% respectively to enrich the variability of the CT images. During augmentation, the corresponding coordinates of the bounding boxes are updated as well. To enhance the spatial information, we concatenate one more slice in front of and behind the key slice to get a 2.5D model. This approach increases the information for the model while being a lot more lightweight than a 3D model.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

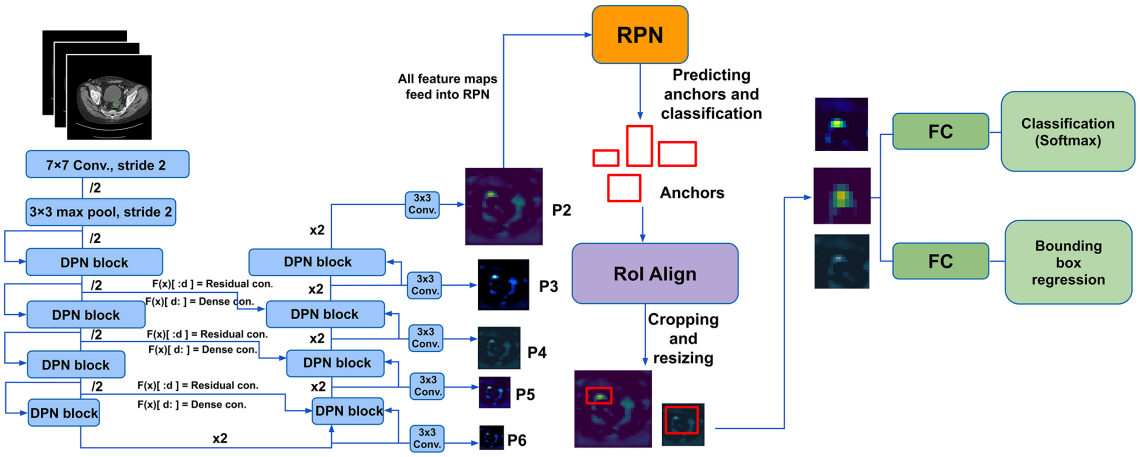

Fig. 1. Our model generates five feature maps for RPN. ROI Align crops the feature maps according to the sizes of the anchors on the corresponding feature map.

C. Proposed Model

We adopt ideas from different algorithms to tackle the issue of variability of pulmonary nodules. The two-stage object detection model Faster R-CNN is used, which is hoped to reduce the ratio of FPs by providing more spatial and contextual information. Fig. 1 shows the outline of the architecture of the proposed model.

The backbone architecture of the first stage is a combination of U-Net and DPN with skip connections between the encoding and decoding path to provide more spatial information and to improve the flow of gradients. Five feature maps of different resolutions are derived during the decoding process that are fed into a Region Proposal Network (RPN) [13] for the first part of classification and anchor box regression.

ROI Align from Mask R-CNN [17] crops and resizes the regions of interest of RPN to a fixed resolution. Eventually, two sets of fully connected layers, which classify nodule or non-nodule, are attached at the end of our network for the second part of classification and anchor box regression, respectively.

1) Encoder: The encoder is a modified DPN architecture that performs DenseNet and ResNet in parallel. The model starts with having 64 kernels with a 7x7 convolutional layer and a stride of 2 followed by a 3x3 max pooling layer with a stride of 2 subsequently as an initial block. Afterwards, there are four more bottleneck blocks [14] having 1x1, 3x3 and 1x1 convolutional layers where each convolutional layer is followed by batch normalization and ReLU activation. Grouped convolutions on all channels are performed within the 3x3 convolutional layer like in ResNet [18].

2) Decoder: The decoder stages consist of scaling up the feature maps and concatenating or adding the skip connections in the same encoder stage, where the upsampling starts after scaling up the feature map, as shown in Fig. 2 (Type I). This approach is relatively straightforward and easy to implement since ResNet-like, DenseNet-like or FCNs have only a single type of operation to combine the different

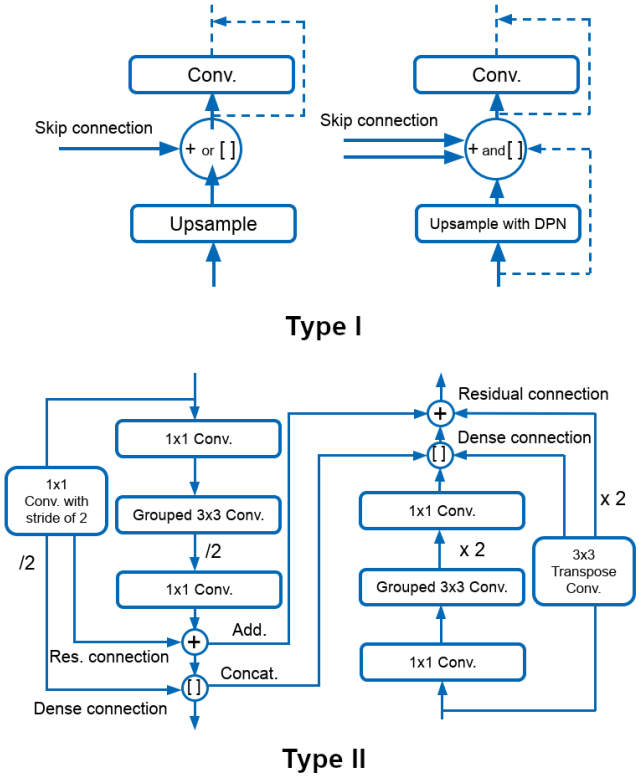

Fig. 2. Upsampling Type I and Type II

connections, yet, DPN has both of these operations during encoding. The implementation of starting the shortcut before the upsampling layer is shown in Fig. 2 (Type II). Hence, it provides us a discussion space to observe the performance if we start the shortcut before upsampling and implement both addition and concatenation to connect with the skip connection as a regular DPN block during decoding. Finally, an extra 3x3 convolutional layer is attached after concatenating or adding the skip connections.

We have implemented two different types of upsampling

with DPN blocks as the backbone to observe the performance on pulmonary nodule detection. Type I is a regular upsampling approach and starts the DPN block after upsampling as shown in Fig. 2. For Type II, we extract a part of the data stream before upsampling, where the main data stream is upscaled in resolution by a DPN block. This is done by the 3x3 convolutional layer within the bottleneck in a DPN block with a stride of 2. The detailed visualization of our proposed upsampling strategy is depicted in Fig. 2, which shows how the two operations of concatenation and addition interact during encoding and decoding.

D. Generating Ground Truth Labels

Considering each scale of an anchor with three different ratios is performed on one single level of the feature map, the concatenation for training labels is not applicable because the height and width are different on each level. For this reason, we vectorize all labels and concatenate them. However, this approach requires a lot more attention to making sure the order of the labels is valid in the order of how convolutional filters are doing windowing on feature maps from different levels.

For RPN, we mark an anchor as positive if the IoU is more than 0.7 with the ground truth bounding box and as negative if the IoU is lower than 0.3. Those anchors having IoU between these maximum and minimum threshold are considered as neutral, and are not involved for training.

The sub-network Classifier relies on the prediction of RPN to provide information on the location of the target. We consider locations as background if the IoU is between 0.3 and 0.5 and foreground if the IoU is greater than 0.5. Only positive samples calculate the regression parameters in this stage as well.

E. Loss Function

Our loss function follows the definition of the multi-task loss function of Faster R-CNN (cf. Eq. (1-2) in [13]), i.e.

$$L(p,t) = \frac{1}{N_{cls}} \sum L_{cls}(p, p^*) + \lambda \frac{1}{N_{reg}} \sum p^* L_{reg}(t, t^*),$$

consisting of:

- The classification loss L_{cls}, the log-loss over two classes (object vs. not object) for RPN and multiple classes for Classifier.
- The regression loss L_{reg}, that we set to the smooth L1 loss [19].

The term $p^* L_{reg}$ calculates the regression loss only for positive anchors ($p^* = 1$). The weighting parameter λ is set to 1, since it was proven insensitive in [13]. For bounding box regression, we adopt the regression parameters of the 4 coordinates following R-CNN's definition (cf. Eq. (6-9) in [20]).

F. Metric

The free-response receiver operating characteristic (FROC) is one of the standard metrics in lesion detection [21]. Its evaluation is performed by measuring the sensitivities (%) with respect to their corresponding average FP rate per scan. TPs and FPs are determined by thresholding a confidence measure of the predictions. For our evaluations, we calculate the IoU of the predicted bounding boxes with the ground truth bounding boxes. If it is larger than 0.5, it represents a TP, or an FP otherwise.

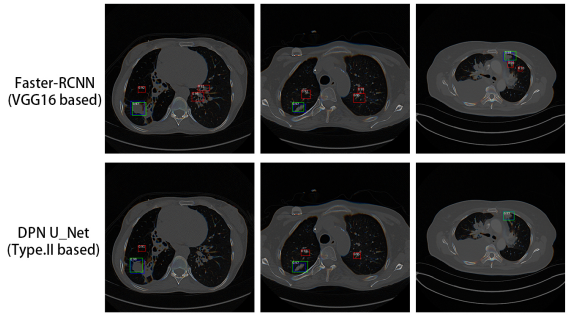

Fig. 3. Exemplary predictions on the CT slices using VGG16 based Faster-RCNN and DPN U-Net (Type II).

G. Experimental Setup

For training, we use Adam as optimizer with its default parameters (learning rate=1×10^{-3}, betas=(0.9, 0.999), eps=1×10^{-8}) and a weight decay of 1×10^{-4}. The initial learning rate is reduced to 1×10^{-4} after 5 epochs. Due to time limitations, all models are trained within 15 epochs on the training dataset. We use data augmentation, dropout and normalization to prevent overfitting. The initialization of the layers' parameters is also kept at default, i.e. Xavier uniform for kernels and zeros for biases. All trainings and tests are performed on Google Colab Pro (NVIDIA Tesla K80) utilizing Keras 2.3.1 with tensorflow as backend.

III. Results

We implemented three different models. Our baseline model is based on Faster R-CNN with VGG16 as backbone where the prediction was only performed on the last feature map. The other two models are our proposed models with the regular upsampling strategy (Type I) and the proposed upsampling approach (Type II).

Tab. I shows the the sensitivities at 1/2, 1, 2, 4, 8 and 16 average FPs per scan to compare our models. At 4 average FPs per scan, e.g., the sensitivity of our model using Type II upsampling is increased by $1.1\,\%$ compared to Type I upsampling, and is even $4\,\%$ above the Faster R-CNN baseline.

The first row of Fig. 3 visualizes the detection results in the official test dataset of DeepLesion for the baseline model while the second row of Fig. 3 shows the results of the proposed model DPN U-Net Type II. Blue, green and red boxes represent the ground truth, TP and FP boxes respectively, and the number on the top-left corner of the boxes represents the confidence.

IV. Discussion

From Tab. I, we see that the average FROC of DPN U-Net Type II yields 80.5%, surpassing the DPN U-Net Type I

TABLE I

SENSITIVITIES AT DIFFERENT FPS PER IMAGE (MULTIPLE LESIONS).

	Sensitivity (%) at FPs					
Model	0.5	1	2	4	8	16
Faster R-CNN (VGG16)	55.8	66.3	74.7	81.2	84.8	87.5
DPN U-Net, Type I	60.9	71.3	77.7	84.2	87.6	88.9
DPN U-Net, Type II	64.6	74.1	80.7	**85.3**	88.3	89.8

with 78.4%, and the baseline model with 75.1%. The baseline model is a single scale model where the RPN only looks for targets on the same resolution of the feature map. By looking at Fig. 3, we can see that the baseline model generates more FP predictions than DPN U-Net Type II on the same CT images in general. Our results show that multiscale feature maps can help to improve the performance.

The traditional upsampling approach in Type I might lose some information when upsampling, although it has the skip-connection in the same stage. DPN U-Net Type II intends further to provide more contextual information upon the original structure. The results show that Type II is an efficient approach to reuse the DPN block and can provide more contextual information by adding a shortcut connection before upsampling. Furthermore, during the experiment, DPN U-Net Type I and II require a similar computational time per batch, while the baseline model requires only 50% of the time for computation. This behavior is expected since DPN consists of more complex operations and structure.

We also compared our proposed models with the state of art, such as a 3DCE_CS_Att network [22] with 21 slices for pulmonary nodule detection achieving a sensitivity of 92% at 4 FPs while having average FROC of 83.9% for all lesions from DeepLesion. Yet, the sensitivity of DPN U-Net Type II at 4 FPs is at 85.3%, which is already quite close to 3DCE with 89%. By theory, 3D input provides more contextual information for a deep learning model, yet, it also requires more computational resources.

V. CONCLUSION

This paper proposed a dual skip connection upsampling strategy to locate pulmonary nodules in various shapes and sizes compared with two baseline networks. Our work shows that the proposed model DPN U-Net Type II surpasses the results performed by the single skip connections model (Type I) and single-scale feature map model (Faster R-CNN). The proposed model DPN U-Net Type II reuses the DPN block throughout the whole network, which is an efficient way to explore new potential features and prevent vanishing gradients by having both operations from ResNet and DenseNet. Overall, our proposed upsampling strategy has successfully reduced the false positives in the evaluation of nodule detection. We assume that the performance of our proposed model might still be improved by fine tuning the hyperparameters.

ACKNOWLEDGEMENT

This work was conducted within the International Graduate School MEMoRIAL at OVGU Magdeburg, supported by the ESF (project no. ZS/2016/08/80646).

REFERENCES

[1] J. Ferlay, I. Soerjomataram, R. Dikshit *et al.*, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International Journal of Cancer*, pp. E359–E386, 2015.

[2] A. Toghiani, A. Adibi, and A. Taghavi, "Significance of pulmonary nodules in multi-detector computed tomography scan of noncancerous patients," *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences*, pp. 460–464, 2015.

[3] D. E. Midthun, "Early detection of lung cancer [version 1; peer review: 3 approved]," *F1000Research*, 2016.

[4] G. D. Rubin, "Lung nodule and cancer detection in computed tomography screening," *Journal of thoracic imaging*, pp. 130–138, 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25658477

[5] L. A. Torre, R. L. Siegel, and A. Jemal, "Lung cancer statistics," in *Lung cancer and personalized medicine*. Springer, 2016, pp. 1–19.

[6] Q. Dou, H. Chen, L. Yu *et al.*, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, pp. 1558–1567, 2017.

[7] M. M. Escobar, "An interactive color pre-processing method to improve tumor segmentation in digital medical images," Master's thesis, Iowa State University, 2008.

[8] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung nodule detection via 3d u-net and contextual convolutional neural network," in *2018 International Conference on Networking and Network Applications (NaNA)*, 2018, pp. 356–361.

[9] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, 2017, pp. 559–567.

[10] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018, pp. 673–681.

[11] H. Jiang, F. Gao, X. Xu *et al.*, "Attentive and ensemble 3d dual path networks for pulmonary nodules classification," *Neurocomputing*, pp. 422–430, 2020.

[12] K. Yan, M. Bagheri, and R. M. Summers, "3d context enhanced region-based convolutional neural network for end-to-end lesion detection," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, 2018, pp. 511–519.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1137–1149, 2017. [Online]. Available: https://doi.org/10.1109/TPAMI.2016.2577031

[14] Y. Chen, J. Li, H. Xiao *et al.*, "Dual path networks," *arXiv preprint arXiv:1707.01629*, 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.

[16] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of Medical Imaging*, pp. 1 – 11, 2018.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[18] S. Xie, R. Girshick, P. Dollár *et al.*, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987–5995.

[19] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[20] R. Girshick, Donahue, and et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[21] A. A. A. Setio, A. Traverso, T. de Bel *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge," *Medical Image Analysis*, pp. 1–13, 2017.

[22] Q. Tao, Z. Ge, J. Cai *et al.*, "Improving deep lesion detection using 3d contextual and spatial attention," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, 2019, pp. 185–193.