EasyChair Preprint

№ 5897

Small Cell Lung Cancer Semantic Segmentation in Pathological Section Images Based on Multi-Scale Fusion Network

Qincheng Zhang, Junhai Xu, Ke Zheng, Zhiwen Zhang and Jiangyong Yu

June 23, 2021

Small Cell Lung Cancer Semantic Segmentation in Pathological Section Images Based on Multi-scale Fusion Network

Qincheng Zhang[1], Junhai Xu[1], Ke Zheng[1], Zhiwen Zhang[2(✉)], and Jiangyong Yu[3(✉)]

[1] College of Intelligence and Computing, Tianjin University(TJU), Tianjin, China.
[2] Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China.
zhangzhw109@126.com
[3] Department of Medical Oncology, Beijing Hospital, National Center of Gerontology; Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China.
yujiangyong2011@163.com

Abstract. Lung cancer is a serious threat to human health, accounting for 26% of the total number of cancer deaths. In order to develop treatment plans for patients with small cell lung cancer, doctors are required to evaluate their pathological materials, which requires a high level of experience and knowledge and is time-consuming and labor. In this paper, we propose an automatic segmentation model of small cell lung cancer tumor from pathological section images. A multi-scale convolutional network is trained on pathological images using ground truth ROIs that were manually delineated by pathologists for 31 patients. We use high-resolution to low-resolution parallel convolution to extract multi-scale features in the first three down-sampling stages of the network's encoding path. Use mIOU(mean intersection over union) as the evaluation index to compare the training results of this network with U-Net, and the results showed that the segmentation results of this network were significantly improved compared with U-Net. Code and models are available at https://github.com/QinchengZhang/PathologySegmentation.

Keywords: Medical image segmentation· Convolutional neural networks· Multiscale.

Table of Contents

1 Introduction

Lung cancer is a serious threat to human health, accounting for 26% of the total number of cancer deaths[1], and ranks first in the incidence and mortality of malignant tumors in China. Small cell lung cancer (SCLC) accounts for about 13% of lung cancer, with high aggressiveness, high mortality and 5-year overall survival (OS) of only 6.3% [2,3]. Two-thirds of patients with SCLC have distant metastases, and the most common targets are liver, bone, brain, and lung[4]. For all patients with malignant metastases, brain-metastases (BM) are frequent in SCLC patients with a high fatality rate.

In recent years, with the increase of detection methods such as CT, PET/CT and MR and the increase of treatment methods for lung cancer[5,6,7], the survival rate of lung cancer is increasing year by year. They have facilitated the development of automated analytical algorithms to reduce the burden and improve the performance of pathologists. Therefore, automatic medical image segmentation has been widely studied in the field of image analysis.

In recent years, a large number of deep learning (DL) methods have been proposed for automatic image analysis from the cell level to the image level[8].

With the advent of convolutional neural networks (CNNs), near-radiologist-level performance can be achieved in automated medical image analysis tasks[9,10], including pulmonary pathology section image segmentation[11]. High expression

ability, fast reasoning and filter sharing make CNN a practical standard for image segmentation. Fully convolutional networks (FCNs) [12] and the U-Net [13] are two commonly used architectures.

High-resolution representation learning plays an essential role in semantic segmentation, especially with high-resolution medical images. Therefore, how to retain high resolution features in feature extraction has become a problem we need to solve. Ke Sun et al proposed HRNet [14]. The HRNet maintains high-resolution representations by connecting high-to-low resolution convolutions in parallel and strengthens high-resolution representations by repeatedly performing multi-scale fusion across parallel convolutions.

In this paper, we introduce a novel network based on Encoder-Decoder to perform the segmentation of the tumor area in pulmonary pathology section images. This network uses a multi-scale encoder like HRNet to extract high resolution features. The article is structured as follows. In the first part, we will introduce the related work. Then we will introduce the method we proposed, including the dataset we used, the network structure, the evaluation and so on.

2 Related Work

2.1 U-Net

In 2015, Ronneberger et al.[13] proposed a U-shaped fully convolutional network for medical image segmentation called U-Net, which has a typical symmetrical codec network structure. U-Net has an obvious advantage that it can make good use of GPU memory. This advantage is mainly related to extraction of image features on multiple image scales. U-Net transfers the feature maps obtained in the encoding stage to the corresponding decoding stage, and merges the feature maps of different stages through "skip connection" to merge coarse and fine-level dense predictions. Since then, U-Net has become the baseline for most medical image segmentation tasks, and has inspired more researchers to think about the U-shaped structure.

2.2 HRNet

In 2019, Ke Sun et al. published a paper named Deep High-Resolution Representation Learning for Human Pose Estimation[14], which proposed a high-resolution network HRNet. The high-resolution representation is maintained by parallelizing high-resolution to low-resolution convolution, and the high-resolution representation is enhanced by performing multi-scale fusion repeatedly across parallel convolution. The effectiveness of the method is proved in many image tasks.

3 Methods

U-Net realized multi-scale feature fusion by skipping connection of decoding path, but its encoding path did not well in multi-scale feature fusion, while

HRNet realized multi-scale feature extraction. Therefore, we tried to combine the advantages of the two to achieve an improved method. In this section, we describe the dataset and our proposed model for segmentation of pulmonary pathological sections.

3.1 Dataset

In this study, we performed H&E staining of lung tissue sections from 31 patients using a digital slide scanner(Hamamatsu Nano Zoomer Digital Pathology) with an objective magnification of 20 times. The SCLC was morphologically diagnosed according to the criteria of World Health Organization(WHO) classification of Tumors of the Lung, Pleura, Thymus and Heart (the fourth Edition)[15]. The area containing tumor cells was circled by one thoracic pathologist.

3.2 Preprocessing

Due to the influence of staining ratio, staining platform, imaging platform and even different operators, different pathological sections have great differences in color, which are not suitable for direct training. Therefore, we first use an improved color normalization method based on Vahadane method[16,17] to process the data. The close look of different tissues in the slides can be seen in Figure 1.

Since the original data annotation is saved in XML format files, we then generated a binary mask according to the annotation, and we used ASAP to read the images and annotations in NDPI format. Since the image is too huge, it is impossible to input the original image into the network during segmentation, so it must be cut into patches one by one. We intercepted each annotation according to the annotation information and resized it to 1024×1024 size. Then I got 134 patches. The segmented image with the corresponding mask is shown in the Figure 2.

Next, we use data enhancement methods such as mirroring, flipping, rotating and offsetting to expand the dataset. We finally got 670 patches.

3.3 Network architecture

Our proposed network structure is shown in the Figure 3.

Network description Our proposed network is based on U-Net, and we use high-resolution to low-resolution parallel convolution in the first three down-sampling stages of the encoding path, high-resolution to low-resolution parallel convolution is like a fully connected structure, in which the high-order features are connected to the low-order features of the next stage through up-sampling, and the low-order features are connected to the high-order features of the next stage through stride convolution, and features of the same order are connected through convolution with a kernel size of 3 and stride length of 1. so as to ensure the fusion of multi-scale features and the preservation of high-resolution features.

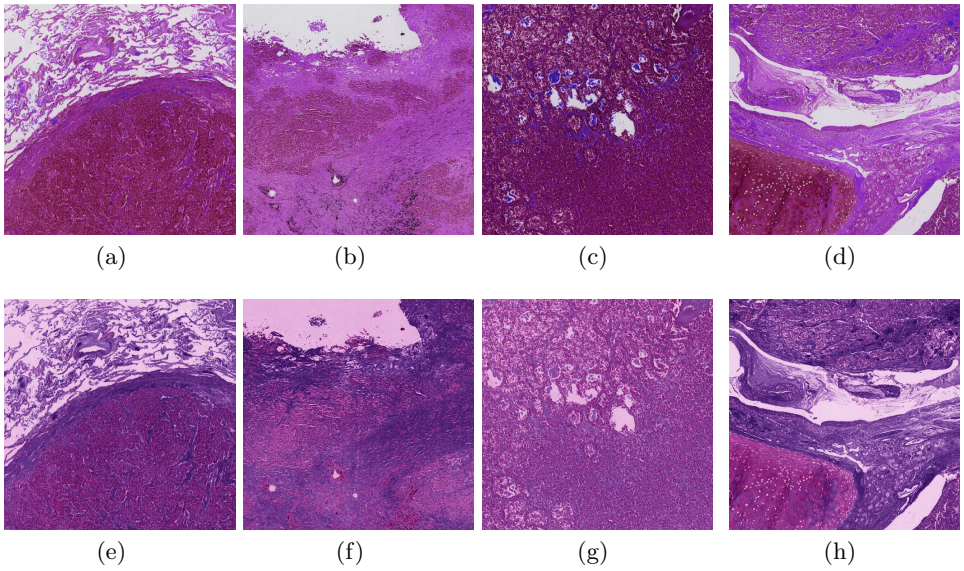

Fig. 1. Example of original images and color normalized images, where (a), (b), (c) and (d) are original images, and (e), (f), (g) and (h) are color normalized images

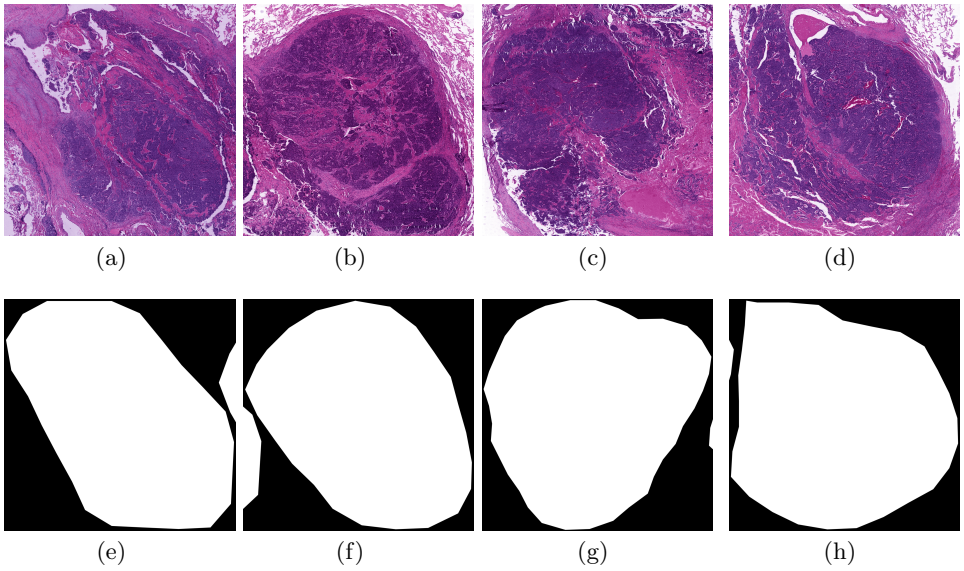

Fig. 2. (e), (f), (g) and (h) are the binary masks of (a), (b), (c) and (d) respectively

In order to reduce the number of parameters, we do not use high-resolution to low-resolution parallel convolution in the fourth down-sampling stage. Instead, the highest order features of the previous stage are down-sampled to get the highest order features of this stage, and the rest features are convolved to get the features of each order of this stage. No changes have been made to the decoding path. In our experiments, the input and output sizes are $256 \times 256 \times 3$ and $256 \times 256 \times 1$ respectively.

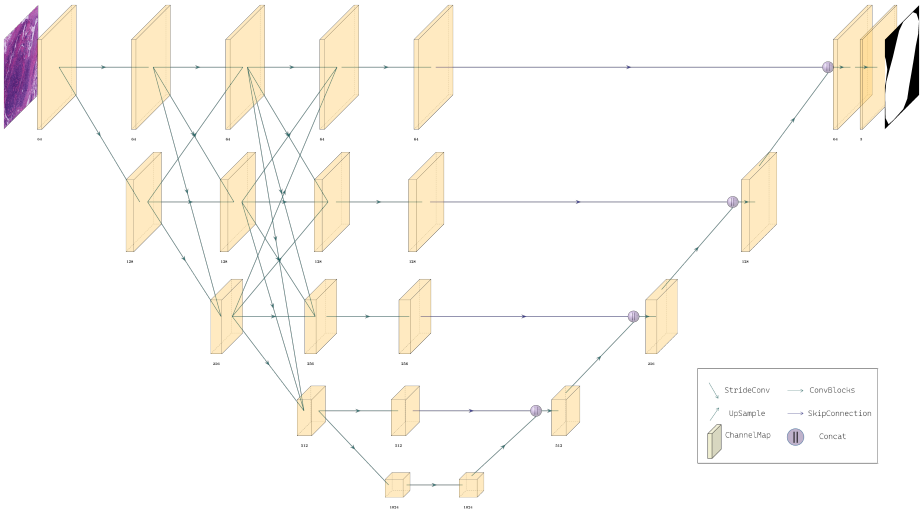

Fig. 3. The network we proposed, where StrideConv is two convolution operations with a stride size of 2 and a convolution kernel of 3. Convblocks is several convolution + BatchNorm + ReLU operations with stride size of 1 and convolution kernel of 3.

3.4 Training scheme

The loss used for computing the gradients is a combination of binary Dice Similarity Coefficient(DSC) and Cross-Entropy (CE) losses defined by Isensee et al.[18] and computed as $L_{DSC} + L_{CE}$. This dual loss benefits from the smooth and bounded cross-entropy loss gradients and the explicitly optimized dice scores used for evaluation and their robustness against class imbalances. The Dice loss is computed as

$$L_{DSC} = 1 - 2\frac{\sum_i \hat{y}_i y_i}{\sum_i \hat{y}_i + \sum_i y_i} \tag{1}$$

where $\hat{y} \in [0, 1]$ is the softmax output,$y \in \{0, 1\}$ is the ground truth mask and the sum is computed over all voxels. The cross-entropy loss is computed as

$$L_{CE} = \sum_i y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \tag{2}$$

We train the networks with a SGD optimizer with a batch size of 12 and use Poly Learning Rate Policy to adjust the learning rate, the learning rate decline formula is shown as

$$lr = lr_{initial} * (\frac{1 - iter}{iter_{max}})^{power} \tag{3}$$

3.5 Evaluation

To assess the impact of the attention module, our model is compared with other methods. To enable a fair comparison, data pre-processing, validation and hyper-parameter search were performed the same way as for our model. We divided the dataset into 30:1 training set and validation set. The training set and validation set contain 649 and 21 images respectively.

To evaluate the agreement between the ground truth and the 2-classes prediction, we used the mIOU and kappa coefficient. The mIOU is computed as

$$mIOU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{4}$$

Where p_{ij} represents the real value of i, the predicted number of j, and k+1 is the number of categories.p_{ii} is the actual quantity.p_{ij} and p_{ji} represent false positive and false negative respectively.

The kappa coefficient is computed as

$$kappa = \frac{p_o - p_e}{1 - p_e} \tag{5}$$

Where p_o is the sum of the number of samples for each correct classification divided by the total number of samples, which is the total classification accuracy. p_e is the sum of the products of the actual and predicted numbers for each category, divided by the square of the total number of samples.

4 Experimental settings and results

4.1 Experimental settings

The 670 patches obtained were divided into a training set and a test set, in which the training set contained 650 patches and the test set 20 patches. During training, resize each patch to 256×256.

We set the number of convolution blocks at each down-sampling stage as 2 and 4 respectively, conducted two training sessions and compared them with other models.

The network was implemented in python with the PaddlePaddle library [19] and is based on the PaddleSeg project [20,21]. Training and testing were conducted using the mentioned setup and hyper-parameters from scratch on single NVIDIA Titan RTX of 24GB. 100000 iterations were fixed as upper bound limit for training each single network, but in all the conducted trials, the best model always reached around 10000 ± 2000 iterations.

4.2 Experimental results

With the above configurations we obtained the quantitative results expressed in Table 1. This shows that the network we proposed has better robustness. The segmentation accuracy of U-Net and Attention U-Net is 0.8952 and 0.9383 respectively. And the segmentation accuracy of ours with 2 ConvBlocks and 4 ConvBlocks is 0.9403 and 0.9409 respectively. Mean intersection over union(mIOU) is the main evaluation criterion, the mIOU of U-Net, Attention U-Net, our method with 2 ConvBlocks and our method with 4 ConvBlocks is 0.8682, 0.8823, 0.8860, 0.8871 respectively. And the kappa of U-Net, Attention U-Net, our method with 2 ConvBlocks and our method with 4 ConvBlocks is 0.8588, 0.8749, 0.8790, 0.8803 respectively. The above three evaluation criterion of our method with 4 ConvBlocks is the best, but it also has a much larger number of parameters, reaching 122.15M. While our method with 2 ConvBlocks reduces the number of parameters by 41.7% at the expense of very little accuracy.

	Number of Params	mIOU	Accuracy	Kappa
U-Net	13.40M	0.8682	0.8952	0.8588
Attention U-Net	34.89M	0.8823	0.9383	0.8749
Ours(2 ConvBlocks)	71.21M	0.8860	0.9403	0.8790
Ours(4 ConvBlocks)	122.15M	0.8871	0.9409	0.8803

Table 1. Performance comparison of different networks.

4.3 Segmentation results

Figure 4 shows the segmentation result of the network we proposed. The top row is the original image, the middle is ground truth, The proposed model(4 ConvBlocks) predictions are on the bottom row. The black areas represent normal areas and the white areas represent lesion areas. According to the segmentation result, we can observe intuitively that the generated image is very similar to the mask image. In addition, in the unlabeled part, we found some small white areas, which means that the unlabeled pathological cells have also been isolated. Therefore, we think that the features learned by neural network through marker information have good generalization, and can segment the scattered pathological cells in the unlabeled area. The results show that the proposed network is better than expected in learning multi-scale feature information, which proves the powerful application of neural network in the field of medical imaging.

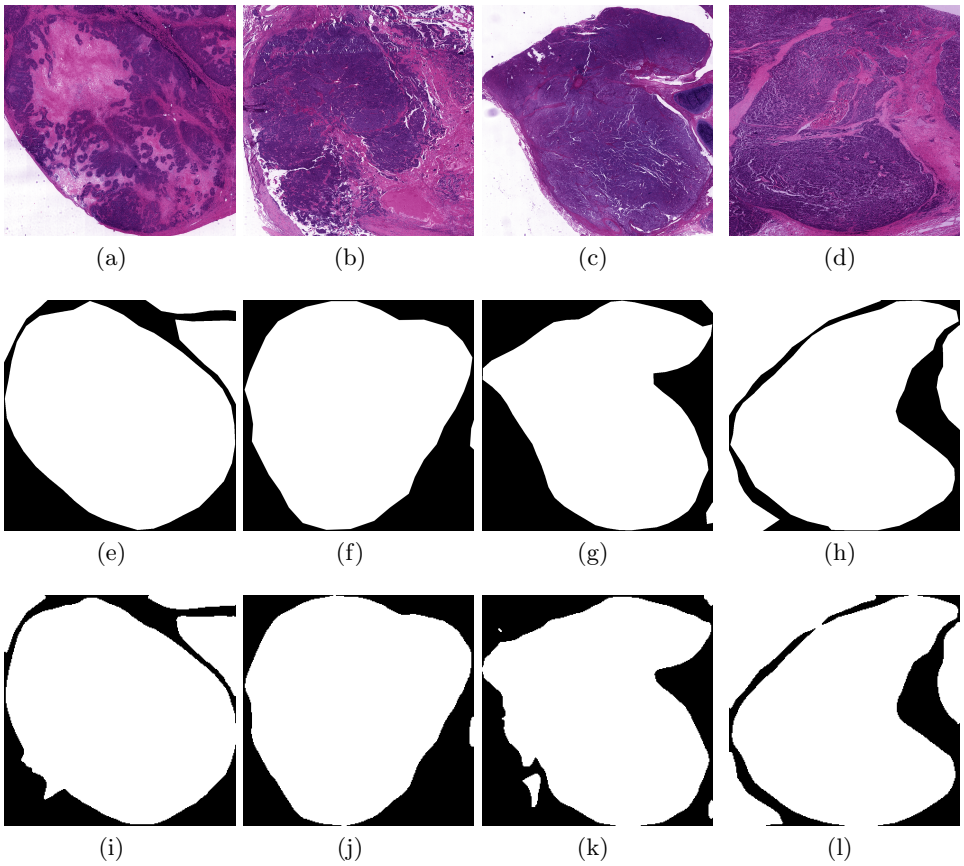

Fig. 4. Segmentation examples from four patches. The top row is the original image, the middle is ground truth, The proposed model(4 ConvBlocks) predictions are on the bottom row.

5 Conclusion

In this paper, we developed an automatic segmentation model of small cell lung cancer tumor from pathological section images based on multi-scale fusion network. This is an important problem because the development of U-shaped networks is closely related to the development of medical image segmentation, and the extraction of multi-scale features is conducive to the development of high-resolution medical image segmentation algorithms. We propose a U-shaped network based on multi-scale feature fusion and compare it with two kinds of networks. The experimental data we get are even more convincing. We use mIOU as the main evaluation criterion of segmentation effect. The results show that the mIOU of our proposed network is 0.0189 higher than that of U-Net, and 0.0048 higher than that of Attention U-Net. We can conclude that our proposed network is superior to the other two in the segmentation of small cell lung cancer based on pathological section images of the lung.

Acknowledgement

The work was supported by the Project funded by China Postdoctoral Science Foundation(2020M680905).

References

1. R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1), 2019.
2. G. P. Kalemkerian, W. Akerley, R. J. Downey, D. S. Ettinger, and C. C. Williams. Small cell lung cancer. *J Natl Compr Canc Netw*, 37(3):783–796, 2016.
3. A. Amini, L. A. Byers, J. W. Welsh, and R. U. Komaki. Progress in the management of limited-stage small cell lung cancer. *Cancer*, 120(6), 2014.
4. M. Fruh, D De Ruysscher, S. Popat, L. Crino, S. Peters, and E. Felip. Small-cell lung cancer (sclc): Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 24 Suppl 6(suppl 6):vi99, 2013.
5. B. T. Kung, T. Auyong, and C. Tong. Prevalence of detecting unknown cerebral metastases in fluorodeoxyglucose positron emission tomography/computed tomography and its potential clinical impact. *World Journal of Nuclear Medicine,13,2(2014-08-19)*, 13(2):108–111, 2014.
6. P. C. Davis, P. A. Hudgins, S. B. Peterman, and J. C. Hoffman. Diagnosis of cerebral metastases: Double-dose delayed ct vs contrast-enhanced mr imaging. *American Journal of Neuroradiology*, 12(2):293–300, 1990.
7. Julio Sanchez De Cos Escuin, D. M. Menna, MAS González, J. Z. Quirantes, and Mcp Calvo. [silent brain metastasis in the initial staging of lung cancer: evaluation by computed tomography and magnetic resonance imaging]. *Archivos De Bronconeumología*, 43(7):386–391, 2007.
8. Chetan L. Srinidhi, Ozan Ciga, and Anne L. Martel. Deep neural network models for computational histopathology: A survey. 2019.

9. Mitko Veta A, Yujing J. Heng B, Nikolas Stathonikos C, Babak Ehteshami Bejnordi D, Francisco Beca E, Thomas Wollmann F, Karl Rohr F, Manan A. Shah G, Dayong Wang B, and Mikael Rousson and H. Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. *Medical Image Analysis*, 56:111–121, 2019.
10. Zizhao Zhang, Pingjun Chen, Mason Mcgough, Fuyong Xing, and Lin Yang. Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. *Nature Machine Intelligence*.
11. Zhang Li, Jiehua Zhang, Tao Tan, et al. Deep learning methods for lung cancer segmentation in whole-slide histopathology images – the acdc@lunghp challenge 2019, 2020.
12. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2015.
13. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, page 234–241, 2015.
14. Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, 2019.
15. Who classification of tumours of the lung, pleura, thymus and heart. fourth edition. 2015.
16. Goutham Ramakrishnan, Deepak Anand, and Amit Sethi. Fast GPU-Enabled Color Normalization for Digital Pathology. *arXiv:1901.03088 [cs]*, January 2019. arXiv: 1901.03088.
17. Abhishek Vahadane, Tingying Peng, Shadi Albarqouni, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Amit Sethi, Irene Esposito, and Nassir Navab. Structure-preserved color normalization for histological images. In *IEEE International Symposium on Biomedical Imaging*, 2015.
18. Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018.
19. Yanjun Ma, Dianhai Yu, Tian Wu, and Haifeng Wang. Paddlepaddle: An open-source deep learning platform from industrial practice. *Frontiers of Data and Computing*, 1(1):105, 2019.
20. Yi Liu, Lutao Chu, Guowei Chen, Zewu Wu, Zeyu Chen, Baohua Lai, and Yuying Hao. Paddleseg: A high-efficient development toolkit for image segmentation, 2021.
21. PaddlePaddle Authors. Paddleseg, end-to-end image segmentation kit based on paddlepaddle. https://github.com/PaddlePaddle/PaddleSeg, 2019.