UDC 004.8
IRSTI 28.23.15

https://doi.org/10.55452/1998-6688-2024-21-2-42-53

[1]*Nam D.,
Master of Tech. Sciences, PhD Student, ORCID ID: 0000-0002-9356-3114, e-mail: d.nam@kbtu.kz
[1]Pak A.,
Candidate of Tech. Sciences, Professor, ORCID ID: 0000-0002-8685-9355, e-mail: a.pak@kbtu.kz

[1]Kazakh-British Technical University, 050000, Almaty, Kazakhstan

COMPARATIVE ANALYSIS OF U-NET, U-NET++, TRANSUNET AND SWIN-UNET FOR LUNG X-RAY SEGMENTATION

Abstract

Medical image segmentation is a widely used task in medical image processing. It allows us to receive the location and size of the required instance. Several critical factors should be considered. First, the model should provide an accurate prediction of the mask. Second, the model should not require a lot of computational resources. Finally, the distribution between the false positive and false negative predictions should be considered. We provide the comparative analysis between four deep learning models, base U-Net and its extension U-Net++, TranUNet, and Swin-UNet for lung X-ray segmentation based on trainable parameters, DICE, IoU, Hausdorff Distance, Precision and Recall. CNN models with the smallest number of parameters show the highest DICE and IoU scores than their descendants on the limited-size dataset. Based on the experiment results provided in the article U-Nethas maximum DICE, IoU, and precision. It makes the model the most appropriate for medical image segmentation. SwinU-Net is the model with minimum Hausdorff Distance. U-Net++ has the maximum Recall.

Key words: CNN, segmentation, transformers, medical image processing

Introduction

Medical image segmentation is the process of automatic extraction of masks from the images containing the area of relevant parts. The mask could contain the damaged part of the organ or the organ by itself. An analysis of the mask allows to get information about the current state of the organ. In opposition to classification and detection image segmentation allows to understand the class, location, and size of the relevant image at the same time.

However, medical image processing is a complicated and expensive process because it requires a deficit of data for training. According to the authors of the original U-Net architecture [1], it was proposed specifically for medical image segmentation and could be trained on a limited-size dataset. U-Net was applied for various medical image segmentation subtasks [2–4] and shows good performance. This was a motivation to use its key features, such as mirror encoder and decoder application, and skip-connections were used in different other architectures. U-Net architecture could be combined with residual [5], recurrent [6], or dense [7] blocks forming a new convolutional neural network. U-Net architecture also could be adopted for transformer-based architecture [8, 9, 10].

Although the application of additional components increases segmentation quality, the number of trainable parameters increases as well. The idea of automated medical image segmentation is the reduce the burden on the medical sector through pre-medical diagnostics and patient management. There are several criteria for methods used for the segmentation process. It should not only provide the exact area of the required instance but also work fast without requiring large computing power. So in the current comparative analysis, we take into account the number of trainable parameters of each observed model because, for deep learning models, this is a key characteristic responsible for the speed and required resources for the training process.

Another important component in the process of evaluating neural networks for medical image segmentation is the distribution between false positive and false negative predictions. Image segmentation could be described as a pixel-wise classification, when each pixel belongs to some class, for example, a cancer or non-cancer area.

We compared U-Net with three U-Net-based algorithms: U-Net++ [11], TransUNet [8], and Swin-UNet [10] among the number of trainable parameters and segmentation quality. We used DICE, IoU, and Hausdorff Distance to evaluate the distance and similarity between the ground true and predicted image. Also, we calculated precision and recall for predicted and ground true masks to evaluate the distribution between false positive and false negative predictions. Based on our experiments Base U-Net with the smallest number of parameters shows the highest DICE and IoU scores than its descendants on the limited size dataset. However, it lost in Hausdorff Distance, Precision, and Recall to its descendants.

Background

Medical images are an extensive source of information about the human body, its current state, and possible illness. Together with the disease history and related visual data, the doctor could draw up a complete picture of the disease and appropriate treatment. There are a lot of different formats of medical image data that are appropriate depending on the case: computed tomography images, X-rays, biopsy images, electrocardiograms, and others. In the current article, we stopped on X-ray images because this format has some benefits among others, and it is the widely spread type of equipment to collect medical image data.

X-rays are one of the cheapest ways to collect data about the human body. It is widely available in various medical centers, including small rural ones which are usually not equipped with more expensive types of equipment like CT scanners or MRI machines. Also taking X-ray images are quick and non-invasive, requiring minimal preparation. X-rays provide valuable insights into the internal structures of the body, making them versatile for a range of diagnostic purposes. They are commonly used to examine bones, detect fractures, identify abnormalities in the chest, and assist in various other diagnostic evaluations. Also, X-ray involves a relatively low radiation dose. The X-ray was created in the 19th century and still is the most popular way of screening.

The use of X-ray images allows for gathering information about various organs in the human body. In this article, we focus on X-ray images of the lungs, as this organ is particularly susceptible to various diseases. Currently, lung cancer stands as the most lethal type of cancer in Kazakhstan and worldwide. Lung cancer can be classified into two main types: non-small cell lung cancer and small-cell lung cancer. The most spread and the least dangerous type of non-small cell lung cancer is adenocarcinoma. It grows more slowly than other types and is often treatable due to early detection. Usually, it could be found in the outer part of the lung. Another type of non-small cell lung cancer is squamous cell carcinoma. Typically found in the central part of the lung. Squamous cell carcinoma is often linked to smoking. The last subtype of non-small cell lung cancer is large cell carcinoma. This type of cancer is characterized by rapid growth and non-predicted location in every part of the lung. The second type of cancer is Small cell lung cancer. It is generally the more aggressive type of lung cancer with a higher potential for rapid spread compared. This type of cancer has higher mortality and a more negative prognosis of treatment because of several factors. First is the rapid growth. It faster spreads inside the lung and to other human organs. It leads to the second issue the second issue with this type of cancer. It gives early metastasis that spreads to distant organs, particularly the brain. This can complicate treatment and affect the overall prognosis. Last but not least, the difficulties in treating this type of cancer are twofold. On one hand, frequent late detection of small cell lung cancer often makes surgical intervention impossible. As the cancerous area attains a considerable size, it becomes impractical to remove it surgically without affecting other organs and important blood

vessels. On the other hand, the traditional method of cancer treatment, chemotherapy, often yields positive results at the beginning of treatment, but there is a high risk of recurrence. This prevents the ability to control the cancer in the long term and maintain a stable outcome.

There are a lot of factors that cause lung cancer. Some of them are possible to control, such as smoking, and harmful production. However, the main part of them is difficult to control, like passive smoking, radon gas exposure, age over sixty fixe, and genetic factors.

Another widely spread type of lung disease is tuberculosis. Tuberculosis is widely spread in Kazakhstan. Tuberculosis is a contagious bacterial infection caused by Mycobacterium tuberculosis. This airborne disease primarily affects the lungs but can also target other organs, leading to a range of symptoms and complications. Tuberculosis remains a global health concern, with a significant impact on morbidity and mortality.

Tuberculosis is a highly contagious type of illness. Tuberculosis can be transmitted when an infected person coughs or sneezes, releasing microscopic droplets containing the bacteria into the air. The primary modes of tuberculosis transmission include inhaling air containing tiny particles of the microbes and contact with the infection through the skin or mucous membranes. Additionally, transmission from mother to child during pregnancy or childbirth is also possible. Due to the necessity of contact with an infected person, tuberculosis is often transmitted in prisons and other places of confinement. This is because many individuals are housed in poorly ventilated environments.

Tuberculosis could be in two forms: latent in active. In active form, there are common symptoms of tuberculosis including persistent cough, chest pain, weight loss, fatigue, night sweats, and fever. A latent form of tuberculosis occurs when the bacteria are present in the body but not causing symptoms. However, a patient with a latent form of tuberculosis also is dangerous because of the risk of spread. TB is treatable with a combination of antibiotics, usually taken for several months. Adherence to the full course of treatment is crucial to prevent the development of drug-resistant strains.

One more lung disease that was the reason of the pandemic and creatine in 2019 is Covid 19. It was caused by a novel coronavirus SARS-CoV-2 and afflicted economic, public health, and geo-political situations over the world. COVID-19 primarily spreads through respiratory droplets when an infected person coughs, sneezes, or talks. Common symptoms include fever, cough, and shortness of breath, but the virus can also cause a range of respiratory, gastrointestinal, and neurological issues. Certain individuals, particularly the elderly and those with underlying health conditions, are at a higher risk of severe complications. Despite the mortality rate of Covid 19 significantly decreasing and the initiation of vaccination, various new strains of the virus have begun to emerge with differing symptomatology, transmission rates, and mortality.

Although the 2019 pandemic dealt a significant blow to humanity, it undoubtedly exposed healthcare challenges linked to a shortage of qualified medical personnel. Overall, the medical field faces a significant challenge with timely disease diagnosis and subsequent patient management due to a shortage of healthcare professionals and necessary equipment. In this context, the use of automated tools in preclinical medical diagnostics has the potential to significantly improve the quality of services provided and the overall survival rates of patients with various illnesses. Hand segmentation of the disease or organ is an expensive and long process. Moreover, due to the significant burden on healthcare professionals and the fact that images are not pre-ranked based on the severity of the patient's condition, individuals with more severe forms of illness may not receive timely treatment. It is also worth noting that the use of automated tools cannot fully replace a doctor since there is a significant challenge in predicting false-positive results. In the context of pulmonary diseases, a false-positive prediction may lead to a lung puncture, which is an invasive procedure.

We have focused on segmenting the region of the lung itself, as this step allows for the initial assessment of the organ's condition. It involves defining boundaries and size, as well as the potential identification of entities with radiological characteristics deviating from the standard values of the lung area.

Main provision

We provide a comparative analysis of base U-Net with three U-Net-based algorithms: U-Net++, TransUNet, and Swin-UNet. The main contribution of the article could be described as follows:

- The authors of Swin-UNet provide the comparison of TransUNet and Swin-UNet for the aorta, gallbladder, spleen, left kidney, right kidney, liver, pancreas, spleen, and stomach. We checked the algorithms for the task of lung segmentation.
- The comparison of TransUNet and Swin-UNet was provided in the original article Swin-UNet based on DICE and Hausdorff Distance. We complete the comparative analysis by adding IoU. Also, for medical processing tasks, it is critical to consider the distribution between false positive and false negative predictions. In the lung segmentation task, we took pixels belonging to the lung as positive, and all other areas of the image as negative. We also calculated precision and recall characterizing the distribution between segmentation errors.
- The comparison of TransUNet and Swin-UNet in the original article was done with the use of hardware with different computation power. For the current research, we used the same equipment for training all four models.
- We provide a comparison between two convolutional (U-Net, U-Net++) and two transformer-based (TransUNet, Swin-Unet) models to check which type of model shows better results for medical image segmentation on the dataset with a limited size. Also, we used a backbone model with a smaller number of trainable parameters for TransUNet to bring the number of trainable parameters of transformers and convolutional neural networks closer to each other.

Materials and methods

Image segmentation models

We compared four deep-learning algorithms for the lung segmentation task: U-Net, U-Net++, TransUnet, Swin-UNet. U-Net++ is an extension of the U-Net architecture for semantic segmentation, incorporating skip connections and deep supervision to enhance the ability of the model to capture hierarchical features in medical image analysis. TransUNet is the combination of a transformer encoder and a convolutional decoder. Swin-UNet is a fully transformer-based model. In contrast to convolutional-based models, transformer-based architectures utilize an attention mechanism. This mechanism can be conceptualized as a sequence of linear layers incorporating key, query, and value components. By doing so, the attention mechanism adeptly highlights relevant features within the original input sequence, allowing transformers to capture intricate dependencies and patterns across the entire sequence for various tasks. The combination of the attention mechanism in the encoder and convolutional decoder in TransUNet is the reason for additional calculation inside skip connections because the convolutional part operates with matrixes while the transformer part works with sequences. This architectural singularity and the number of layers are the reasons why Transnet used the biggest number of parameters among all observed models.

TransUNet and Swin-UNet models were compared by the authors of Swin-UNet architecture on the Synapse multi-organ computer tomography dataset [12] and ACDC dataset [13]. The synapse dataset consists of computed tomography images of the aorta, gallbladder, kidney, liver, pancreas, spleen, and stomach and corresponding masks. ACDC dataset contains MRI images with labeled ventricles and myocardium.

Data

For the current comparative analysis, we used X-ray images of the lung, because the results of the original article do not provide information about the quality of segmentation based on X-rays, and the lung was not used in the original articles as well. We used an open-source dataset of chest X-rays with corresponding lung masks [14, 15]. We used 704 images from the dataset, which were randomly split into train and test sets. So the train set contains 563 images and the test set consists

of 141 images. All X-ray images were provided in PNG format. An example of an image from the dataset and corresponding mask is provided in Figure 1.

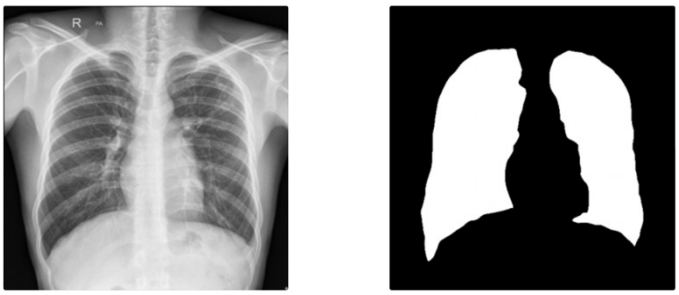

Figure 1 – X-ray image (left) and corresponding binary lung mask (right)

We used 704 images from the dataset, which were randomly split into train and test sets. We used the same distribution between the train and test for honest comparison. The train set contains 563 images, and the test set consists of 141 images. We used the image with the size 224*224 because this image size was applied in SwinU-Net architecture. So, we applied the same for all of the others for honest comparison.

The dataset also has additional annotations that include information about age, gender, sex, and diagnosis. We used only a part of the dataset. So, the distribution between genders is 442 to 262 for males and females correspondingly. The age distribution could be described as it is shown in Table 1. As it is shown in Table 1 the main part of the dataset is the X-ray images of young adults with age from 19 to 35.

Table 1 – Age distribution

Age span	Number of examples
0–12	27
13–18	14
19–35	321
36–60	278
60+	64

The dataset includes a variety of pulmonary conditions diagnosed through chest X-rays. There are descriptions of the most popular cases that appeared in the dataset. Also, the dataset includes some other cases like Smear positive, active TB, LUL apex; Bilateral secondary PTB, left encapsulated intrathoracic fluid; Left PTB, left pleural thickening, and others which appeared in the dataset no more than five times.

Normal (358 cases) indicate that there are no apparent signs of pulmonary abnormalities. These images serve as a baseline for comparison with cases exhibiting pathology.

Bilateral PTB (48 cases) signifies the presence of pulmonary tuberculosis affecting both lungs. This condition necessitates careful monitoring and treatment due to the bilateral involvement, as tuberculosis can significantly impact respiratory function.

Right PTB (42 cases) indicates the presence of pulmonary tuberculosis specifically affecting the right lung. The focus on the right lung allows for targeted diagnosis and treatment planning.

PTB in the Right Upper Field (26 cases) specifies tuberculosis localized to the upper region of the right lung. This precise characterization assists in identifying the affected area for accurate diagnosis and treatment.

Left PTB (18 cases) highlights cases where pulmonary tuberculosis is isolated to the left lung. Like right-sided cases, specific localization aids in tailored medical intervention.

These classifications play a crucial role in medical diagnostics, guiding healthcare professionals in identifying and addressing pulmonary pathologies through the analysis of chest X-ray images. Understanding the nature and location of abnormalities is vital for effective treatment strategies and patient care.

Loss function

For training all of the selected models we used the same hyper-parameters and loss function. We applied DICE Loss for the training process because it is widely used for image segmentation. Dice loss is a metric used in image segmentation tasks that quantifies the difference between predicted and ground truth masks. We used dice loss with square values for the predicted mask to increase the distance between the incorrect prediction and the ground true mask (Equation 1).

$$\text{DICE Loss} = 1 - \text{DICE} = 1 - \frac{2 * |X^2 \cap Y^2|}{|X^2| + |Y^2|} \quad (1)$$

Model evaluation

As medical image segmentation is used as a premedical step of image diagnostics, the quality of segmentation is critical. However, the similarity between the real and predicted mask is not only one important characteristic. To evaluate the similarity and distance between ground true and predicted mask we used DICE (Equation 2), IoU (Equation 3) and Hausdorff Distance (Equation 4).

$$\text{DICE} = \frac{2 * |X \cap Y|}{|X| + |Y|} \quad (2)$$

$$\text{IoU} = \frac{|X \cap Y|}{|X| \cup |Y|} \quad (3)$$

$$\text{HD} = \max\left(\max_{a \in A} \min_{b \in B} d(a, b), \max_{b \in B} \min_{a \in A} d(a, b)\right) \quad (4)$$

It is important to have an opportunity to use the model on different hardware with different computation resources. For deep learning models, the number of trainable parameters is the characteristic responsible for the speed and required hardware for training, because it has a direct relation with the number of layers of the model.

Image segmentation task could be described as pixel-wise classification. The distribution between false positive and false negative predictions is also important for medical image processing tasks. We used precision (Equation 5) and recall (Equation 6) to evaluate if the model better classified positive or negative examples. Although precision and recall both characterize the quality of classification, one of them could be significantly higher than the other. Precision is more important if the number of false positive predictions has more crucial consequences than false negative. Otherwise, when false negative predictions are more critical, it is better to increase recall.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (5)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (6)$$

Hardware

All experiments were done on the same hardware: NVIDIA A100 80GB GPU with CUDA version 11.7, paired with an AMD EPYC 7663 56-Core Processor for CPU tasks., RAM 1,5 Ti. Computational resources have been temporarily provided by the Institute of Information and Computational Technologies (IICT).

Results and discussion

We calculated the number of parameters for all the models. The cheapest model we used was Base U-Net. We used five metrics for the evaluation of segmentation. DICE, IoU, Precision, and Recall should be maximized, and Hausdorff Distance should be minimized for better segmentation results. We used a constant number of epochs for training. The results with input size 224*224 are provided in Table 2.

Table 2 – Comparative results for U-Net, U-Net++, Swin-UNet, TransUNet with the input size 224 * 224

Model	N param	DICE	IoU	HD	Prec	Rec
Base U-Net	3.25E+07	0.9546	0.9147	16.73	0.9643	0.9474
Swin-UNet	4.14E+07	0.9543	0.9141	13.6969	0.9623	0.9487
U-Net++	4.89E+0.7	0.9537	0.9128	23.05	0.9476	0.9618
TransUNet	9.23E+07	0.9442	0.8965	14.3628	0.9580	0.9340

As we can see in Table 1, DICE and IoU decreased when the number of trainable parameters increased. It could be caused by the limited size of the data. The model with the minimum Hausdorff Distance is SwinU-Net, which means that it provides maximum similarity in the form of a predicted mask. U-Net also has maximum Precision, this makes U-Net a preferred model in situations where it is important to minimize the number of false positive predictions. Usually, the problem is false positive predictions from the point of view of the patient and the doctor. The model with the highest Recall is U-Net++, which means that it makes fewer false negative predictions.

Literature review

The U-Net architecture was originally constructed for the main challenges of computer vision: it could be trained on the set with limited size, considering the fact the relevant part of the image could be significantly smaller, like borders between cells on a biopsy image, than all other parts, and does not require big computational resources because its architecture distinctive features. Base U-Net consists of main parts: mirror encoder-decoder, skip connection layers, and bottleneck. U-Net shows high performance for medical [16, 17, 18] and non-medical image segmentation [19]. The high performance of U-Net has been a motivation for its extension for various U-Net-based algorithms [8, 9, 10, 11, 20]. U-Net architecture could be combined with residual [5], recurrent [6], or dense [7] blocks forming a new convolutional neural network. U-Net also could be adopted for 3D image segmentation [21, 22] which is also a widely used format of visual representation of the human body.

The author of article [23] provides a comprehensive theoretical review of more than eight U-Net-based and original U-Net models. The authors provided a list of their applications for different human organs and illnesses segmentation based on the information given in other articles.

The article [24] provides a comparative analysis of 2D U-Net and 3D U-Net for brain tissue segmentation. The authors used the Open Access Series of Imaging Studies (OASIS) [25] for training and testing. The article [26] provides a comparative analysis of U-Net-based algorithms for liver segmentation. 3D U-Net was better in both cases. The article [27] provides a comparison of 2D and

3D U-Net for pulmonary nodule segmentation based on the LIDS-IDRI Dataset [28]. Two Level U-Net shows the best scores for image segmentation than 2D and 3D U-Net.

U-Net architecture could be extended not only as a part of the new convolutional neural network but also as a new transformer architecture. The first application of transformer-based architectures in computer vision was described in the article [29]. The authors described an application of a visual transformer for the image classification task. The architecture was extended for the image segmentation task. Examples of combination U-Net with transformers are TransUNet and Swin-UNet. TransUNet and Swin-UNet were compared for caries segmentation in the article [30]. The article [31] proposed new transformer-based architecture which was tested with TranUNet and SwinU-Net for ACDC Dataset. In both cases, TransUNet showed a higher DICE score, which is the opposite result provided by the authors of the original SwinU-Net article.

Although transformer-based architectures are powerful tools for image processing, they require much more computation resources than convolutional neural networks. Sometimes their application is not worth the resources spent. The article [33] provides an adaptation of U-Net called LargeKernelU-Net (LKU-Net) which shows higher DICE on OASIS dataset than TransMorph [34].

Although U-Net-based algorithms show stable and high performance for the segmentation of instances with different types, they are not the only type of architecture, used in medical image processing. Example Mask R-CNN [35] originally described for non-medical object detection, also could be adopted for medical image segmentation task [36, 37] by itself. The article [38] provides a comparison between U-Net, Mask RCNN, and their ensembled for nuclear segmentation.

Conclusion

Medical image processing could be applied as a pre-medical diagnostic to increase the quality of patient management. Image segmentation is one of the computer vision tasks, that is suitable for medical image diagnostics because it allows to find not only the class of required instance, but the location, size, and form as well. However, there are a lot of changes and limitations for computer vision medicine. The first one is the computational and data resources. The second is the quality of segmentation because the model must find the accurate size and form of the required instance. Finally, the distribution between false positive and false negative predictions influences the choice of the model. The limitation of medical image processing was the motivation for the creation of U-Net architecture for medical image segmentation. U-Net shows high performance on various medical image tasks and was extended by combination with other architectures. We compared U-Net with three of its extensions TranUNet, Swin-UNet and UNet++. Based on our results Base U-Net still shows comparative results with its descendants with a bigger number of trainable parameters.

Future work

In the current article, we provide a comparative analysis of four deep learning models: U-Net, U-Net++, TransUNet and SwinUNet for lung segmentation tasks. We trained all models on similar circumstances and compared results with each other. We used open-source data for comparison. All of the images were provided in PNG format. We used the dataset with a small number of images. So, this research has some limitations, which could be extended in future work.

First, the images that we used are provided in PNG format. One pixel in PNG could be described as an integer value from 0 to 255. However, the usual format of X-ray images is an application of DICOM format. Dicom format is also used for computed tomography images. This format could contain much more value of each pixel which allows to separate instances of different types by the grey value on the image by itself. So the X-ray image in Dicom contains much more radionic features than the PNG image that we used.

Another limitation of the experiment that we provide is the size of the dataset. Although it was one of the purposes of the research to check the behavior of selected models on small-size datasets,

this data could not be enough for training transformer-based models. So to extend current research the dataset with a bigger size could be used.

The distribution between the lung area and another part of the X-ray also should be considered. In case we describe the segmentation of the lung on X-ray as pixel-ways classification, the distribution between the lung and other instances is quite balanced because the area of the lung is big. However, we used metrics like DICE and IoU which are partly sensitive to the size of the set. It should be mentioned that for disease segmentation, we usually work with examples where the relevant region is significantly smaller than other parts of the body. So, the behavior of the model could not be predicted before real tests were done for another problem domain. Also, these experiments could be done several types for different types of disease because all of them have different radiological features.

We also use image size 224*224 while the size of the original X-ray image could be different depending on the equipment used for taking the data. So some images were compressed or extended to have similar sizes. Moreover, the Dicom format contains some additional information like the scale of the image, which was lost because of the use of PNG. However, this information could be deleted because in medicine all the information should be anonymized as it is possible. However, for private usage, the quality of segmentation and image processing additional information about disease history, age, sex, and smoking could increase the quality of the model work.

Another possible way to extend current research is not in comparison of the models but in increasing segmentation quality. We do not apply any types of data augmentation to increase the quality of segmentation for honest comparison. Also, we had a goal to compare models on the small-size dataset, so we to not increase the size of the training set in any possible way. So an application of affine transformation could make the segmentation of the models better.

The dataset also was annotated for classification purposes. For current experiments, we used only a part of the data with X-ray images and corresponding lung masks. However, the original dataset also provides information that was contained from two data sources: Shenzhen and Montgomery. The dataset also contains information about gender, age, pathology existing, diagnosis, and disease. There are nine possible types of pathology in the dataset: tuberculosis, pleurisy, atelectasis, pleural thickening, pneumothorax, pneumonia, fibrosis, granuloma (or cavitary), and infiltrate. The main part of the X-ray images does not contain any types of pathologies. Also, some X-ray images do not have labeled pathology and are indicated as an unknown class. Currently, the open-source data is labeled for lung segmentation only, however it cloud be extended by additional labeling for mentioned diseases. However, the fact that medical labeling is expensive and long process should be taken into account as well.

Also, we used only U-Net-based architectures for comparison reasons. However, as it is mentioned in the Literature review, there are a lot of other ways for medical image segmentation. Currently, we cannot effectively evaluate the reason why the neural network decides on any type of image-processing task. However the X-ray images, especially in Dicom format, could be segmented by non-neural network methods. So the lung could be segmented based on morphological features of the X-ray image. Subsequently, the research could be extended by the comparison of deep learning methods with classical computer vision approaches.

Information about funding

This work was supported by the Ministry of Education and Sciences of the Republic of Kazakhstan under the following grant #AP14871214. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1 Ronneberger O., Fischer P. and Brox T. (2015) U-Net: Convolutional networks for biomedical image segmentation, in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., pp. 234–241.

2 Shaziya H., Shyamala K. and Zaheer R. (2018) Automatic Lung Segmentation on Thoracic CT Scans Using U-Net Convolutional Network, 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0643–0647. https:// doi.org/10.1109/ICCSP.2018.8524484.

3 Robin M., John J. and Ravikumar A. (2021) Breast Tumor Segmentation using U-NET, 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 1164–1167. https:// doi.org/ 10.1109/ICCMC51019.2021.9418447.

4 Frid-Adar M., Ben-Cohen A., Amer R. and Greenspan H. (2018) Improving the segmentation of anatomical structures in chest radiographs using U-Net with an ImageNet pre-trained encoder, in Image Analysis for Moving Organ, Breast, and Thoracic Images. Cham, Switzerland: Springer, pp. 159–168.

5 He K., Zhang X., Ren S. and Sun J. (2016) Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778. https:// doi.org/10.1109/CVPR.2016.90.

6 Liang, Ming, and Xiaolin Hu (2015) Recurrent convolutional neural network for object recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

7 Huang G., Liu Z., Van Der Maaten L. and Weinberger K.Q. (2017) Densely Connected Convolutional Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 2261–2269. https://doi.org/ 10.1109/CVPR.2017.243.

8 Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y. ... and Zhou Y. (2021) Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

9 Pan S., Liu X., Xie N. and Chong Y. (2023) EG-TransUNet: a transformer-based U-Net with enhanced and guided models for biomedical image segmentation. BMC bioinformatics, 24(1), 85.

10 Cao H., Wang Y., Chen J., Jiang D., Zhang X., Tian Q. and Wang M. (2023, February). Swin-unet: Unetlike pure transformer for medical image segmentation. In Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III, pp. 205–218, Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25066-8_9.

11 Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N. and Liang J. (2018) Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, pp. 3–11, Springer International Publishing.

12 Multi-Atlas Labeling Beyond the Cranial Vault – Workshop and Challenge. Synapse multi-organ computer tomography dataset. [Data set]. Synapse.org. https://repo-prod.prod.sagebase.org/repo/v1/doi/loca.te?id=syn3193805type=ENTITY. https://doi.org/10.7303/SYN3193805

13 Bernard O., Lalande A., Zotti C. et al. (2018) Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Trans Med Imaging, vol.37, no.11, pp. 2514–2525. https://doi.org/10.1109/TMI.2018.2837502.

14 Jaeger S., Karargyris A., Candemir S. et al. (2014) Automatic tuberculosis screening using chest radiographs. IEEE Trans Med Imaging, vol.33, no. 2, pp. 233–245. https://doi.org/10.1109/TMI.2013.2284099. PMID: 24108713

15 Candemir S., Jaeger S., Palaniappan K., Musco JP, Singh RK, Xue Z., Karargyris A., Antani S., Thoma G. and McDonald CJ. (2014) Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans Med Imaging, vol. 33, no. 2, pp. 577–590. https://doi.org/10.1109/TMI.2013.2290491. PMID: 24239990

16 Dong H., Yang G., Liu F., Mo Y. and Guo Y. (2017) Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, Proceedings 21, pp. 506–517, Springer International Publishing.

17 Roja Ramani D. and Siva Ranjani S. (2019) U-Net based segmentation and multiple feature extraction of dermascopic images for efficient diagnosis of melanoma. In Computer Aided Intervention and Diagnostics in Clinical and Medical Images, pp. 81–101, Springer International Publishing. https:// doi.org/10.1007/978-3-030-04061-1_9.

18 Song L.I., Geoffrey K.F. and Kaijian H.E. (2020) Bottleneck feature supervised U-Net for pixel-wise liver and tumor segmentation. Expert Systems with Applications, no. 145, p. 113131. https://doi.org/10.1016/j.eswa.2019.113131.

19 Abderrahim N. Y. Q., Abderrahim S. and Rida A. (2020) Road Segmentation using U-Net architecture, 2020 IEEE International conference of Moroccan Geomatics (Morgeo), Casablanca, Morocco, pp. 1–4, https://doi.org/10.1109/Morgeo49228.2020.9121887.

20 Zhuang, J. (2018) LadderNet: Multi-path networks based on U-Net for medical image segmentation. arXiv preprint arXiv:1810.07810.

21 Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T. & Ronneberger O. (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, Proceedings, Part II 19, pp. 424–432. Springer International Publishing.

22 Chen W., Liu B., Peng S., Sun J. & Qiao X. (2019) S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation. Lecture Notes in Computer Science, pp. 358–368. https://doi.org/10.1007/978-3-030-11726-9_32.

23 Siddique N., Paheding S., Elkin C.P. and Devabhaktuni V. (2021) U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications, in IEEE Access, vol. 9, pp. 82031–82057. https://doi.org/10.1109/ACCESS.2021.3086020.

24 Woo B. and Lee M. (2021) Comparison of tissue segmentation performance between 2D U-Net and 3D U-Net on brain MR Images, 2021 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Korea (South), pp. 1–4. https:// doi.org/10.1109/ICEIC51217.2021.9369797.

25 Marcus D.S., Wang T.H., Parker J., Csernansky J.G., Morris J.G. and Buckner R.L. (2007) Open Access Series of Imaging Studies (OASIS): Crosssectional MRI data in young, middle aged, nondemented, and demented older adults, J. Cogn. Neurosci., vol. 19, pp. 1498–1507.

26 Meine H., Chlebus G., Ghafoorian M., Endo I. and Schenk A. (2018) Comparison of u-net-based convolutional neural networks for liver segmentation in ct. arXiv preprint arXiv:1810.04017.

27 Song G., Nie Y., Zhang J. and Chen G. (2020) Research on the fusion method of 2D and 3D UNet in pulmonary nodules segmentation task, 2020 International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China, pp. 44–47. https:// doi.org/10.1109/ICCSMT51754.2020.00016.

28 Armato S.G. 3rd, McLennan G., Bidaut L. et al. (2011) The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. Medical Physics, no.38, pp. 915–931. https://doi.org/10.1118/1.3528204

29 Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T. ... and Houlsby N. (2020) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

30 Ying S., Wang B., Zhu H., Liu W. and Huang, F. (2022) Caries segmentation on tooth X-ray images with a deep network, Journal of Dentistry, no. 119, p. 104076. https://doi.org/10.1016/j.jdent.2022.104076.

31 Wang H., Xie S., Lin L., Iwamoto Y., Han X.H., Chen Y. W. and Tong R. (2022, May). Mixed transformer u-net for medical image segmentation. In ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2390–2394 IEEE. https:// doi.org/10.1109/ICASSP43922.2022.9746172.

32 Jia X., Bartlett J., Zhang T., Lu W., Qiu Z. and Duan J. (2022, September) U-net vs transformer: Is u-net outdated in medical image registration? In International Workshop on Machine Learning in Medical Imaging, pp. 151–160. Cham: Springer Nature Switzerland.

33 Chen J., Frey E.C., He Y., Segars W.P., Li Y. & Du Y. (2022) Transmorph: Transformer for unsupervised medical image registration. Medical image analysis, no. 82, p. 102615. https://doi.org/10.1016/j.media.2022.102615.

34 He K., Gkioxari G., Dollár P. and Girshick R. (2017) Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

35 Liu M., Dong J., Dong X., Yu H. and Qi L. (2018, September). Segmentation of lung nodule in CT images based on mask R-CNN. In 2018 9th International Conference on Awareness Science and Technology (iCAST), pp. 1–6, IEEE.

36 Zhang Y., Chan S., Park V.Y., Chang, K.T., Mehta S., Kim M. J. ... and Su M. Y. (2022). Automatic detection and segmentation of breast cancer on MRI using mask R-CNN trained on non–fat-sat images and tested on fat-sat images. Academic Radiology, 29, pp. 135–144.

37 Vuola A.O., Akram S.U. and Kannala J. (2019) Mask-RCNN and U-Net Ensembled for Nuclei Segmentation, 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, pp. 208–212. https:// doi.org/10.1109/ISBI.2019.8759574.

[1]*Нам Д.,
техника магистрі, PhD студенті, ORCID ID: 0000-0002-9356-3114,
e-mail: d.nam@kbtu.kz
[1]Пак А.,
т.ғ.к., ORCID ID: 0000-0002-8685-9355, e-mail: a.pak@kbtu.kz

[1]Қазақстан-Британ техникалық университеті, 050000, Алматы қ., Қазақстан

ӨКПЕНІҢ РЕНТГЕНДІК КЕСКІНДЕРІН СЕГМЕНТТЕУ МӘСЕЛЕСІНДЕ U-NET, U-NET++, TRANSUNET ЖӘНЕ SWIN-UNET МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ

Аңдатпа

Медициналық кескіндерді сегменттеу медициналық кескіндерді өңдеуде кеңінен қолданылатын мәселе. Медицинада сегменттеуді қолдану қажетті нысанның орналасуы мен мөлшерін анықтауға мүм- кіндік береді. Модельді таңдауда бірнеше маңызды фактор ескерілді. Біріншіден, модель масканы дәл болжауды қамтамасыз етуі керек. Екіншіден, модель есептеу ресурстарының үлкен көлемін қажет етпеуі керек. Ақырында, жалған оң және жалған теріс болжамдар арасындағы бөліністі ескеру қажет. Біз DICE, IoU, дәлдік пен толықтық, Хаусдорф қашықтығы сияқты параметрлер негізінде өкпенің рентген кескіндерін сегменттеу үшін қолданылатын терең оқытудың төрт моделіне салыстырмалы талдау жасаймыз, олар: негізгі U-Net және оның U-Net ++ кеңейтімі, TranUNet және Swin-UNet. Ең аз параметрлі CNN модельдері деректер жинағы шектелген көп параметрлі модельдермен салыстырғанда DICE және IoU ең жоғары көрсеткішті көрсетеді. Мақалада ұсынылған эксперимент нәтижелеріне сәйкес U-Net максималды DICE, IoU мен дәлдікке ие. Ал бұл оны медициналық кескінді сегменттеуде пайдаланылатын ең қолайлы модельге айналдырады. SwinU-Net – Хаусдорф қашықтығы ең аз модель. U-Net++ максималды толық модель.

Тірек сөздер: CNN, сегменттеу, трансформерлер, медициналық кескінді өңдеу.

[1]*Нам Д.,
магистр техн. наук, PhD студент, ORCID ID: 0000-0002-9356-3114, e-mail: d.nam@kbtu.kz
[1]Пак А.,
канд. техн. наук, профессор, ORCID ID: 0000-0002-8685-9355, e-mail: a.pak@kbtu.kz

[1]Казахстанско-Британский технический университет, 050000, г. Алматы, Казахстан

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ U-NET, U-NET++, TRANSUNET AND SWIN-UNET В ЗАДАЧЕ СЕГМЕНТАЦИИ РЕНТГЕН-СНИМКОВ ЛЕГКОГО

Аннотация

Сегментация медицинских изображений является широко используемой задачей в обработке медицинских изображений. Использование сегментации в медицине позволяет получить местоположение и размер необходимой сущности. Существует несколько важных факторов при выборе модели. Во-первых, модель должна обеспечивать точное предсказание маски. Во-вторых, модель не должна требовать большого объема вычислительных ресурсов. Наконец, следует учесть распределение между ложноположительными и ложноотрицательными предсказаниями. Мы предоставляем сравнительный анализ четырех моделей глубокого обучения: базовой U-Net и ее расширений U-Net++, TranUNet и Swin-UNet для сегментации легких по рентгеновским снимкам на основе обучаемых параметров, DICE, IoU, расстояния Хаусдорфа, точности и полноты. Модели CNN с наименьшим количеством параметров показывают самые высокие показатели DICE и IoU по сравнению с моделями с большим количеством параметров на ограниченном по размеру наборе данных. Согласно результатам эксперимента, представленным в статье, U-Net имеет максимальные DICE, IoU и точность. Это делает модель наиболее подходящей для сегментации медицинских изображений. SwinU-Net – модель с минимальным расстоянием Хаусдорфа. U-Net++ имеет максимальную полноту.

Ключевые слова: CNN, сегментация, трансформеры, обработка медицинских изображений.