Tumor Segmentation in Intraoperative Fluorescence Images Based on Transfer Learning and Convolutional Neural Networks

Weijia Hou, MSc[1] , **Liwen Zou, PhD[2], and Dong Wang, PhD[3]**

Abstract

Objective: To propose a transfer learning based method of tumor segmentation in intraoperative fluorescence images, which will assist surgeons to efficiently and accurately identify the boundary of tumors of interest.
Methods: We employed transfer learning and deep convolutional neural networks (DCNNs) for tumor segmentation. Specifically, we first pre-trained four networks on the ImageNet dataset to extract low-level features. Subsequently, we fine-tuned these networks on two fluorescence image datasets (ABFM and DTHP) separately to enhance the segmentation performance of fluorescence images. Finally, we tested the trained models on the DTHL dataset. The performance of this approach was compared and evaluated against DCNNs trained end-to-end and the traditional level-set method.
Results: The transfer learning-based UNet++ model achieved high segmentation accuracies of 82.17% on the ABFM dataset, 95.61% on the DTHP dataset, and 85.49% on the DTHL test set. For the DTHP dataset, the pre-trained Deeplab v3 + network performed exceptionally well, with a segmentation accuracy of 96.48%. Furthermore, all models achieved segmentation accuracies of over 90% when dealing with the DTHP dataset.
Conclusion: To the best of our knowledge, this study explores tumor segmentation on intraoperative fluorescent images for the first time. The results show that compared to traditional methods, deep learning has significant advantages in improving segmentation performance. Transfer learning enables deep learning models to perform better on small-sample fluorescence image data compared to end-to-end training. This discovery provides strong support for surgeons to obtain more reliable and accurate image segmentation results during surgery.

Keywords

image-guided surgery, intraoperative fluorescent image segmentation, convolutional neural networks, deep learning, transfer learning

Introduction

The primary approach for treating solid tumors involves surgical resection. In tumor surgery, it is crucial for surgeons to rapidly and accurately identify the precise boundaries of the tumor to ensure its complete removal. Inaccurate resection can lead to increased local recurrence rates, reducing patients' chances of survival.[1] Fluorescence-guided surgery, as a modern medical technique, has found extensive application in surgical procedures. It provides real-time fluorescence images to assist surgeons in more accurately identifying tumors, localizing abnormal tissues or lesions, while minimizing damage to healthy tissues.[2-5] This technology offers significant support to surgical procedures and has the potential to enhance precision and success rates while improving patient outcomes. In intraoperative fluorescent imaging, a molecular fluorescent dye is injected into the patient before surgery and illuminated using near-infrared light. When the fluorescent dye enters tumor cells, it emits near-infrared light when excited by an external light source. The fluorescence navigation system utilizes near-infrared light in combination with white light to generate real-time images of the tumor. Then, surgeons can accurately locate the tumor in real-time on a display screen. This technology provides valuable guidance, making

[1]College of Science, Nanjing Forestry University, Nanjing, China
[2]Department of Mathematics, Nanjing University, Nanjing, China
[3]Group A: Large-Scale Scientific Computing and Media Imaging, Nanjing Center for Applied Mathematics, Nanjing, China

Corresponding Author:
Dong Wang, Nanjing Center for Applied Mathematics, No.26, Zhishi Road, Nanjing 211135, China.
Email: wangdong@njcam.org.cn

surgery more precise and efficient. Intraoperative fluorescence imaging not only plays a critical role in tumor surgery but has also found extensive applications in various medical fields. In neurosurgery, for instance, fluorescent dyes help surgeons track the distribution of nerve tissue in real-time, thereby preventing inadvertent damage to nerve structures and increasing the safety and success of surgery.[6] In cardiac and vascular surgery, it aids in the precise identification of vascular structures, including coronary arteries, veins, and heart valves, contributing to more accurate decision-making and procedures.[7,8]

Currently, in the realm of intraoperative fluorescence image processing, several research teams have undertaken preliminary investigations. In the domain of image reconstruction, Gao et al[9] introduced a method for time-domain fluorescence diffuse optical tomography. The research proposes a linear image reconstruction algorithm in the context of time-domain fluorescence molecular tomography (FMT). This algorithm allows the simultaneous reconstruction of both yield and lifetime images for multiple fluorescent substances and has the potential to advance the field of fluorescence molecular tomography. In the field of image fusion, the team led by Tian et al[10] proposed an adaptive brightness fusion method (ABFM) for near-infrared fluorescence imaging based on an attention network. Experiments showed that this method exhibited excellent fusion performance in perception and quantification, enhancing the usability of fluorescence imaging in surgical procedures. Zhang et al[11] introduced a novel laparoscopic hepatectomy navigation system (LHNS) that integrates preoperative three-dimensional models with Indocyanine Green (ICG) fluorescence imaging to achieve real-time surgical navigation.

In previous research, the focus on intraoperative fluorescence image processing has predominantly centered around aspects such as image fusion, image reconstruction, and intraoperative navigation systems, with limited exploration into the segmentation of intraoperative fluorescence images. This suggests a certain limitation in the comprehensive analysis of fluorescence images in previous works. Therefore, we conducted an in-depth investigation into the task of segmenting intraoperative fluorescence images, providing an innovative contribution that fills a gap in prior research. Simultaneously, addressing challenges related to the segmentation of these novel image types and the issues associated with small-sample training is a primary focus of this work. The lack of extensive research in this domain is primarily attributed to the following challenges.

One of the challenges in image segmentation work stems from the unique characteristics of intraoperative fluorescent images. During the image acquisition process, factors such as variations in illumination, instrument noise, and other sources of interference can introduce various types of noise and artifacts. Simultaneously, these images exhibit a low contrast feature, further complicating the differentiation between different regions. The boundaries of lesions or tumors are often unclear, posing difficulties in precise localization. Traditional image segmentation methods often struggle to address these issues, as they find it challenging to accurately capture the distinctive features of intraoperative fluorescent images, thereby impacting segmentation performance.

As is widely acknowledged, effective training of deep convolutional neural networks for image segmentation typically demands substantial data and corresponding annotations. Nevertheless, another challenge we confront lies in the relative scarcity of intraoperative fluorescence image data, coupled with the intricate medical knowledge and time-intensive effort required for data annotation. This scarcity of data is a prevalent issue in the realm of medical imaging. To surmount this constraint, we contemplate the application of transfer learning. Through transfer learning, our aspiration is to enhance performance in the task of intraoperative fluorescence image segmentation, ultimately providing a more precise and dependable adjunct tool for medical practice.

The innovation of this study lies in addressing the dual challenges of segmentation for novel images and limited data volume. To tackle these issues, we employed a transfer learning strategy, combined with a multimodal and multi-dataset approach, through a multicenter design. To the best of our knowledge, it is the first attempt to perform tumor segmentation in intraoperative fluorescent images, which fills a critical gap in the field of image segmentation and is of great clinical importance. The automatic identification of the boundaries of fluorescent tumor regions can be employed to minimize harm to healthy tissues during surgery. The remaining sections of this article are structured as follows: Related Works describes the relevant work; Methodology provides an in-depth description of the primary methods and datasets employed in this study; Experiments presents the experimental results; Discussion conducts a discussion of the research findings; and Conclusion offers the conclusion and closing remarks of this article.

Related Works

Deep Learning Techniques for Image Segmentation

Deep convolutional neural networks show powerful feature learning capabilities, enabling automatic extraction of feature from images. This approach better adapts to the specific characteristics of fluorescence images, thereby enhancing segmentation accuracy and robustness. In image segmentation tasks, two commonly used tools are the encoder-decoder structure and the Atrous Spatial Pyramid Pooling (ASPP). The encoder-decoder structure,

by gradually reducing the spatial dimensions of feature maps while capturing higher-level semantic information, provides powerful feature extraction capabilities for image segmentation tasks. Ronneberger et al[12] introduced a deep convolutional neural network architecture known as U-Net. This structure employs an encoder-decoder framework, with the encoder responsible for feature extraction from input images, and the decoder responsible for mapping these features. To effectively handle features at different scales and resolutions, U-Net also incorporates skip connections, enabling the network to simultaneously focus on features at various hierarchical levels. Due to its outstanding performance, U-Net has rapidly expanded across various medical imaging domains, including medical images, cellular images, and neuroimaging, among others. Kermi et al[13] proposed a fully automatic brain tumor segmentation method based on a two-dimensional deep convolutional neural network, which effectively extracts the entire tumor and internal regions from multi-modal 3D-MRI, including enhanced tumor, edema, and necrosis. The network structure is based on U-Net and addresses the issue of class imbalance in brain tumor data by introducing improvements such as Weighted Cross Entropy and Generalized Dice Loss. Testing on the BraTS'2018 dataset shows average Dice coefficients of .783, .868, and .805 for the tumor, whole tumor, and tumor core, respectively. Gao et al[14] achieved automatic segmentation of retinal vessels using the U-Net network. They treated the vessel segmentation task as a multi-label problem, combining pre-processed Gaussian matched filtering with the U-Net network to form a vessel segmentation framework. The method was tested on the DRIVE public dataset, achieving an average classification accuracy of .9636.

However, U-Net also has some limitations, such as relatively weak modeling of global contextual information, loss of features at specific scales, and sensitivity to class imbalances. In order to overcome these limitations and meet the demand for high-precision medical image segmentation, Zhou et al[15] proposed the UNet++ architecture in 2018, which represents an improvement over the classical U-Net structure. This enhanced structure introduces a multi-level nested U-Net design to improve feature extraction and image segmentation. The multi-level U-Net nesting structure effectively captures features at different scales, contributing to enhanced segmentation performance. Additionally, UNet++ incorporates an Adaptive Gating Unit (AGU) to enhance the information transfer process. Research findings indicate that UNet++ achieves superior segmentation performance compared to traditional U-Net and other deep learning architectures on public datasets. Wang et al[16] utilized the UNet++ network for deep learning to automatically segment liver tumors in magnetic resonance imaging (MRI). The experimental results demonstrated

Dice Similarity Coefficient (DSC) values exceeding .9 for the liver and over .6 for liver tumors, indicating that UNet++ can automatically segment normal liver tissue and liver tumors on MR images. This methodological foundation enhances delineation efficiency and meets the subsequent demands for radiological and deep learning analyses in liver tumor segmentation.

Although UNet++ has made significant progress in capturing multi-scale features and facilitating information propagation, it also has some potential drawbacks. For instance, the increased complexity of the network leads to higher computational and memory requirements, demanding a substantial amount of training data. In the same year, Chen et al[17] proposed the Deeplab v3+ network architecture. The design of Deeplab v3+ combines the Atrous Spatial Pyramid Pooling (ASPP) module with the encoder-decoder structure and introduces an innovative convolution operation called "Dilated Separable Convolution." The ASPP module in Deeplab v3+ is utilized to capture multi-scale contextual information. The encoder-decoder structure is employed to gradually reduce the spatial dimensions of feature maps and restore spatial information, maintaining a clear perception of target boundaries. The uniqueness of Dilated Separable Convolution lies in its variable dilation rates, achieved by introducing different dilation rates for convolutional operations to increase the receptive field of the convolutional kernel. This helps the model better capture contextual information when dealing with objects at different scales, simultaneously reducing computational burden and the number of parameters, thereby enhancing the performance of image segmentation. Furthermore, Deeplab v3+ can be pretrained on large datasets to better learn universal features, improving the model's generalization ability and performance on various tasks and datasets. Khodadadi Shoushtari et al[18] employed the Deeplab v3+ architecture with pre-trained Resnet18 weights for semantic segmentation of glioblastoma images. The study utilized BraTS 2020 data for training, and experimental results demonstrated the high accuracy of Deep-Net in segmenting Glioblastoma Multiforme (GBM) tumors. Sun et al[19] utilized the Deeplab v3+ semantic segmentation model based on the TensorFlow framework to segment liver CT images and locate lesions. This model, combining deep convolutional neural networks (DCNNs) and probabilistic graphical models (DenseCRFs), exhibited excellent performance across various computer vision tasks. Roy Choudhury et al[20] performed brain tumor segmentation using the Deeplab v3+ network, constructing 18 models with different combinations of T1CE (Contrast-Enhanced T1), FLAIR (Fluid Attenuated Inversion Recovery), T1, and T2 images for region identification. Leveraging training data from MICCAI BraTS (Multimodal Brain Tumor Segmentation Challenge), their testing results showed average Dice coefficients of .7086

for enhanced tumors, .7897 for tumor cores, and .8755 for the entire tumor region, indicating strong performance.

The Swin UNETR (Swin Transformer U-Net with Rectified Attention) network was introduced by the Hatamizadeh research team[21] in 2021. Built upon the Swin Transformer architecture, this network is specifically designed for addressing medical image segmentation tasks, with a particular focus on semantic segmentation of brain tumors in multi-modal MRI images. Swin UNETR adopts a U-shaped network structure, utilizing the Swin Transformer as the encoder and a CNN-based decoder. The network employs skip connections at different resolutions to effectively link the encoder and decoder, enhancing its capability to capture semantic information in medical images. The architecture aims to improve the accurate segmentation of structures such as organs and lesions in medical images. Kakavand et al[22] conducted a study on enhancing individual-specific knee joint finite element (FE) modeling, introducing a semi-automated segmentation algorithm. Initially, they employed 3D Swin UNETR for the initial segmentation of the femur and tibia, followed by adjustments using a statistical shape model (SSM) to improve surface roughness and continuity. Compared to traditional manual segmentation methods, the semi-automated segmentation achieved outstanding results, with Dice similarity coefficients exceeding 98%.

In summary, current deep learning techniques for image segmentation have made significant advancements. However, there is a lack of research on the segmentation of novel images, such as intraoperative fluorescent images. This paper aims to explore segmentation methods in the specific context of intraoperative fluorescent images, utilizing mature image segmentation techniques. By applying state-of-the-art deep learning models to this new domain, we aspire to provide a more comprehensive and accurate solution for medical image analysis. This endeavor seeks to bridge existing research gaps and propel the development of intraoperative fluorescent image segmentation.

Transfer Learning Methods for Image Segmentation

Transfer learning and fine-tuning methods serve as solutions to the issues of data scarcity and expense prevalent in the field of medical image segmentation. By utilizing neural network models that have undergone pretraining on large-scale datasets, the weights of these models are fine-tuned for the target domain, resulting in improved performance in image segmentation tasks. Given the relatively small size of the target dataset, data augmentation is typically applied during the fine-tuning stage to prevent overfitting, involving subtle alterations to the images. This aids the model in learning essential features pertinent to the segmentation task rather than image-specific characteristics.[23] Data augmentation enhances the model's ability to generalize to unknown domains. The primary advantage of transfer learning and fine-tuning methods lies in their ability to leverage knowledge acquired from large-scale pretrained datasets, thereby achieving exceptional image segmentation performance even in cases of limited target data. This approach reduces training time, data sample requirements, and is applicable to various fields, including new medical image segmentation tasks and remote sensing image segmentation. Sanford et al[24] demonstrated that combining transfer learning, data augmentation, and fine-tuning can enhance the performance of a prostate segmentation model, enabling precise segmentation of the entire prostate and transition zone. They presented a method that employs data augmentation and fine-tuning to extend a centrally trained model to multiple independently trained centers. Amiri et al[25] investigated the impact of fine-tuning different layers of pretrained UNet models on image segmentation performance. Results indicated that the choice of which layers to fine-tune affects segmentation outcomes. Specifically, performance was better when fine-tuning shallow layers of the pretrained network, possibly because these layers encompass specific low-level feature patterns pertinent to medical images. Kaur et al[26] explored how to leverage pretrained segmentation models on a large-scale dataset of patient images to successfully train deep learning pathology segmentation models for different diseases. By training a UNet model on a large, proprietary, multi-center, multi-scanner clinical trial dataset for Multiple Sclerosis (MS) with over 3500 multi-modal MRI samples, they subsequently transferred the model to the BraTS 2018 dataset for segmentation. The research results indicate that fine-tuning the encoder and decoder of the network trained on the larger MS dataset improves brain tumor segmentation when only a small subset of samples is available. Lu et al[27] proposed a transfer learning strategy based on existing knowledge of white matter tract segmentation to segment new white matter tracts with fewer samples. They also developed a data augmentation strategy called TraceMix. Results demonstrated significant improvements in white matter tract segmentation performance with this method when working with a limited number of samples.

In conclusion, the current integration of transfer learning techniques with deep learning methods for image segmentation has been widely adopted. However, there has been no prior research on intraoperative fluorescence images. The main objective of this paper is to address the limited data availability in intraoperative fluorescence

images and enhance the performance of segmentation models using transfer learning.

Methodology

Datasets

This experiment aims to perform tumor segmentation on intraoperative fluorescence images, and we selected the intraoperative fluorescence image dataset as the experimental data. The team led by Tian[10] proposed an adaptive brightness fusion method (ABFM) for image fusion in intraoperative fluorescence images. The dataset used in their paper is a publicly available dataset, and we chose 61 images from it, with a resolution of 1024 × 1024, including images of breast cancer lymphatic vessels and breast tumors. We named this dataset the ABFM dataset. Additionally, our team collected tumor data from Nanjing Drum Tower Hospital, which was captured using the NOVADAQ device—a widely used equipment for capturing fluorescence images during surgery. Surgeons at Drum Tower Hospital collected five intraoperative video data using the NOVADAQ device. We selected frames with clear tumor visibility from these video data and further preprocessed and cropped them to obtain 106 images with a resolution of 1024 × 1024 for training and analysis. These images include 69 images of Pancreatic Neuroendocrine Tumors (PNETs), 26 images of Solid Pseudopapillary Tumors (SPTs), and 11 images of Hepatic Perivascular Epithelioid Cells (HPCs). We named this dataset as DTHP (Drum Tower Hospital Pancreas) dataset. Figures 1 and 2 display some samples from these two datasets.

Following that, we also collected surgical videos for the fluorescence-guided treatment of liver cancer at Gulou Hospital. After injecting ICG, there was no drainage observed in the left half of the liver and the right anterior region, while ICG stayed in these areas. Therefore, doctors clearly marked the boundary lines between the left and right halves of the liver and between the right anterior and right posterior regions. The ICG retention line was then used as the cutting line for liver resection at the diaphragmatic surface. We selected frames from the video data where tumors were clearly visible, subjected them to the same preprocessing and cropping procedures, resulting in a total of 24 images with a resolution of 1024 × 1024, named the DTHL (Drum Tower Hospital Liver) dataset. Figure 3 shows some samples of this dataset.

Ground Truth

The annotation process for the intraoperative fluorescence image dataset in this study consists of two stages. The basic strategy involves initial annotation by our team, followed by submission of the preliminary results to medical professionals for correction, resulting in the final annotations used for network training. The specific procedures are as follows: For the ABFM dataset, preliminary annotation is based on the intraoperative near-infrared fluorescence images, visible light images, and corresponding fused images for each image. These preliminary annotations are subsequently reviewed and corrected by medical professionals, forming the ultimate annotations for network training. For the DTHP and DTHL datasets, we utilized intraoperative fluorescence navigation system videos from Nanjing Drum Tower Hospital for preliminary annotation. Through careful observation of the videos, we generated corresponding ground truth data, which was then submitted to medical professionals for annotation correction, yielding the final annotations for network training. During the annotation process, ITK-SNAP 3.4.0 software was employed. ITK-SNAP, short for Insight Segmentation and Registration Toolkit for Snap, is a free, open-source, cross-platform software application designed for medical image analysis and three-dimensional visualization, serving as a robust tool for medical image annotation.[28]

Data Preprocessing

To effectively prevent overfitting, we employed various data augmentation techniques[29] to enhance the diversity of experimental data. The original images underwent rotation operations, including rotations of 90, 180, and 270 degrees, as well as horizontal and vertical flips. These operations were performed to generate more diverse images, making the training model more robust. After augmentation, the ABFM dataset consisted of 366 images, and the DTHP dataset contained 636 augmented images. It is worth noting that in the intraoperative fluorescence image dataset, lesions often exhibit a small yet high-density distribution, posing a significant challenge to the image segmentation task. To better adapt to this characteristic, we processed the original images with a resolution of 1024 × 1024, focusing on labeled regions of interest with a size of 512 × 512 for network training. This strategy aims to concentrate on the lesion area, enhancing the model's sensitivity to critical regions and further improving segmentation performance. Table 1 provides an overview of the dataset division and data augmentation.

Models

Leveraging knowledge from the domain of natural images and transferring it to the domain of medical image segmentation is an approach. Transfer learning is a technique that enhances the flexibility of training convolutional neural networks.[30-35] Its core idea is to transfer knowledge from one domain (the source domain) to another domain
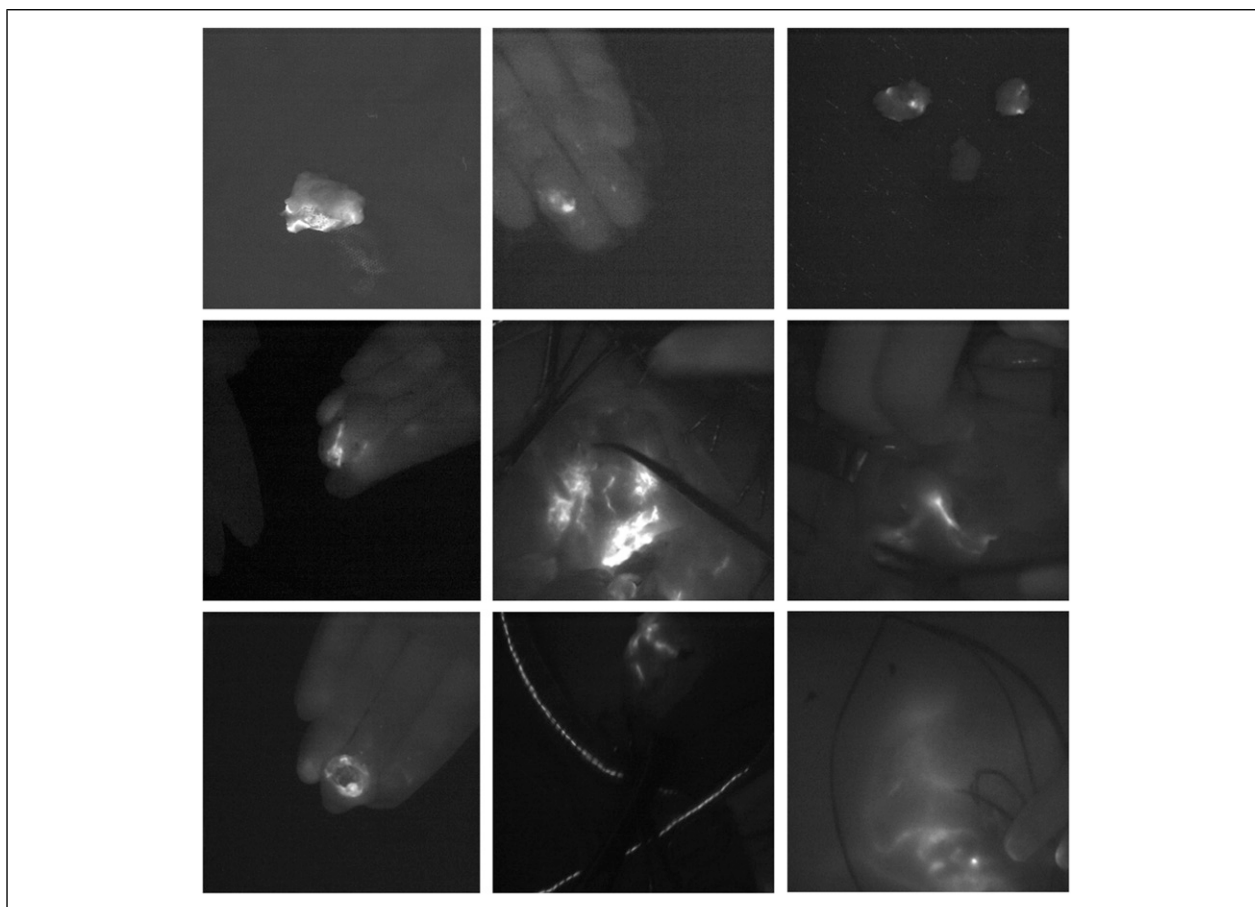
Figure 1. ABFM dataset sample images.

(the target domain) to improve the performance of models in the target domain. This involves transferring the features and knowledge learned by deep learning models pretrained on related tasks to the target task. In transfer learning, the weights of the lower-level feature extractor are typically kept unchanged, and only the top layers of the model are fine-tuned to adapt to the new task. This approach effectively utilizes generic features learned in the source task, reducing training time and data requirements for the target task. Research has shown that transferring features and fine-tuning them to different datasets can lead to better generalization and help avoid overfitting of neural network models.[36]

In this study, we employed a transfer learning strategy, integrating multimodal and multi-dataset approaches, and designed a system involving multiple centers and queues. Throughout the transfer learning process, we chose three distinct network architectures: UNet, Deeplab v3+, UNet++ and Swin UNETR. ImageNet, a vast dataset comprising over 14 million images, is widely utilized for general image classification across 1000 categories.[37] Although ImageNet itself lacks labels for image segmentation,

applying the pre-trained convolutional neural network's lower-level features to segmentation tasks proves effective.[38] Specifically, we trained the encoder parts of the backbones on the ImageNet dataset and freezing the parameters of the encoders. Subsequently, we trained the decoders on our task-specific dataset to achieve model fine-tuning.

In addition, we conducted comparative experiments using end-to-end training of neural networks and traditional level set method DRLSE (Distance Regularized Level Set Evolution).[39] To evaluate the model's performance and reduce the randomness in the evaluation results, we employed a five-fold cross-validation approach. The workflow is illustrated in Figure 4. In this workflow, this neural network framework represents the utilization of four specific deep learning architectures in our approach: UNet, Deeplab v3+, UNet++ and Swin UNETR. The detailed configurations of each backbone are omitted for simplicity in the workflow. Please refer to the corresponding literature for in-depth architecture details. To verify the accuracy and performance of the proposed method, we also tested the trained model on the DTHL dataset.
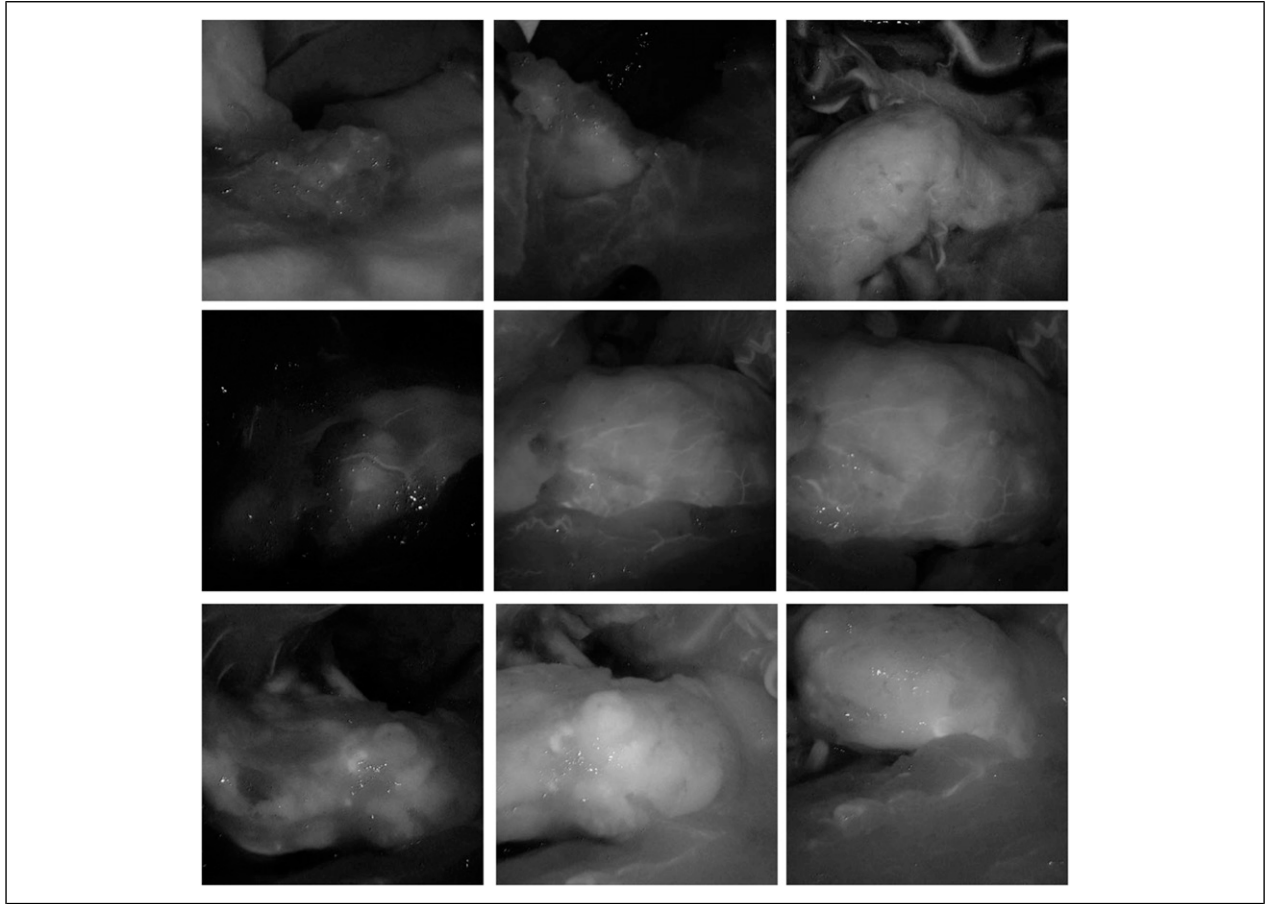
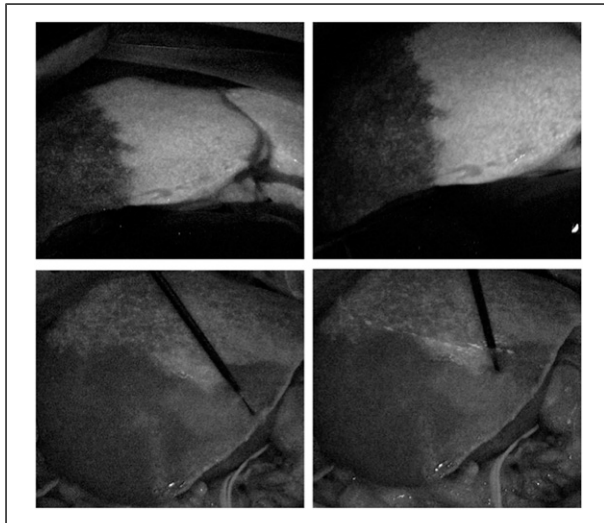**Figure 2.** DTHP dataset sample images.

Figure 3. DTHL dataset sample images.

Table 1. Summary of Fluorescence Image Training Dataset.

Type		Training	Validation	Total
ABFM dataset	Original	51	10	61
	Augmented	287	79	366
DTHP dataset	Original	85	21	106
	Augmented	508	128	636

Experiments

Metrics

We employed two common evaluation metrics to assess the model's performance. The Dice coefficient is used to measure the degree of overlap between the segmentation results and the ground truth labels. A higher Dice coefficient indicates that the model's predictions are closer to the actual labels. Additionally, we utilized the Hausdorff
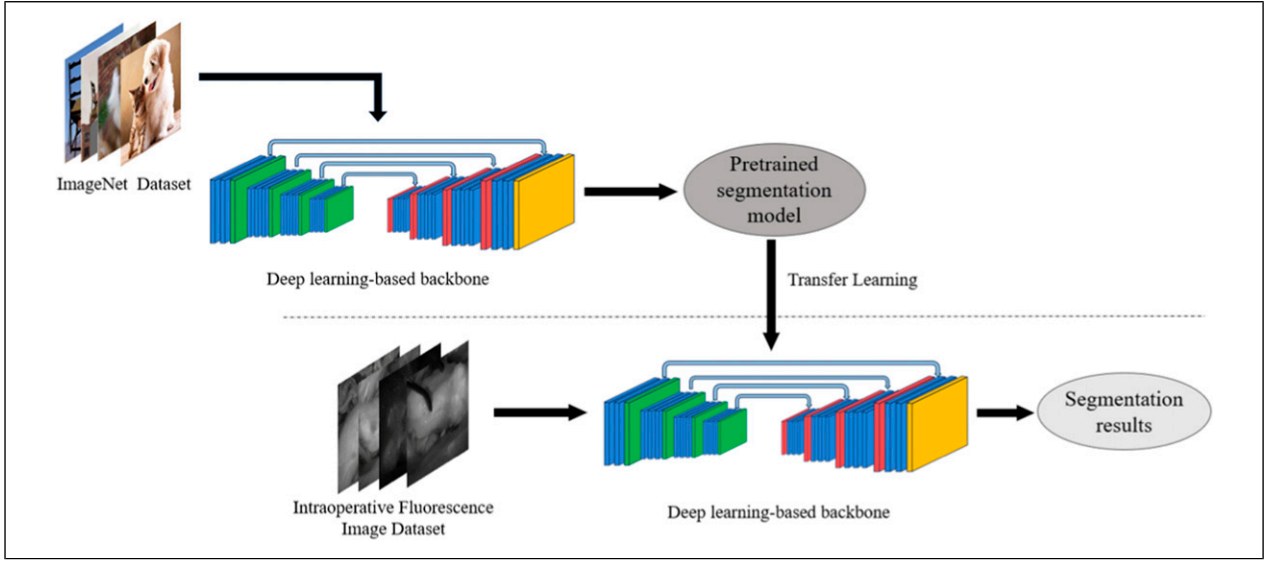
Figure 4. Outline of the proposed method for the segmentation model. The backbone represents the network structures of UNet, UNet++, Deeplab v3+, and Swin UNETR.

distance (HD) to gauge the accuracy of boundary segmentation. The computation method is as follows:

$$Dice = \frac{2 \cdot |X \cap Y|}{|X| + |Y|}. \tag{1}$$

$$HD(X, Y) = \max(h(X, Y), h(Y, X)). \tag{2}$$

$$h(X, Y) = \max_{x \in X} \min_{y \in Y} d(x, y). \tag{3}$$

Among them, X represents the prediction mask of the model. Y represents the ground truth mask. $|X \cap Y|$ represents the correct number of pixels for model segmentation. $|X|$ represents the total number of pixels in the model segmentation. $|Y|$ represents the total number of pixels in the real label. Furthermore, we introduced the evaluation metrics proposed by Zou et al,[40] including False Positive Segmentation Rate (FPSR) and False Negative Segmentation Rate (FNSR). These metrics offer robust support for evaluating the model's performance and segmentation accuracy. The calculation formulas are as follows:

$$FPSR(X, Y) = \frac{|X - Y|}{|X| + |Y|}. \tag{4}$$

$$FNSR(X, Y) = \frac{|Y - X|}{|X| + |Y|}. \tag{5}$$

Implementation Details

The proposed scheme is developed using Python 3.7.8, with PyTorch 1.9.0. The machine on which the experiments were conducted has an NVIDIA TITAN XP GPU with 12 GB memory. We used the Adam optimizer to calculate the optimal weights during back-propagation. All the learning-based models were trained for 200 epochs with a batch size of 4.

In this study, to enhance transparency and reproducibility, we have established an online repository containing the source code and data information related to our research. We sincerely invite you to visit the following webpage link for more detailed information at https://github.com/ZouLiwen-1999/TSIFI. This repository will be regularly updated to ensure that you can access the latest information at any time.

Experimental Results

Segmentation Results on the ABFM Dataset. Table 2 presents the segmentation results of the ABFM dataset. Clearly, the segmentation accuracy of deep learning is over 35.34% higher than traditional segmentation methods, and the segmentation performance with transfer learning consistently outperforms end-to-end training. The DRLSE method achieved a Dice coefficient of 34.06% on the ABFM dataset, indicating suboptimal segmentation performance. Therefore, further calculation of other performance metrics for this approach was

Table 2. ABFM Dataset Segmentation Results. Bold Indicates the Best Performance.

Methods	Backbones	Dice ↑	HD ↓	FPSR ↓	FNSR ↓
Traditional segmentation	DRLSE	34.06	–	–	–
End-to-End Learning	UNet	80.41	35.83	3.97	15.61
	Deeplab v3+	69.40	60.20	5.89	24.71
	UNet++	80.71	**33.34**	5.08	14.20
	Swin UNETR	74.42	48.95	5.69	19.89
Transfer Learning	UNet	80.76	34.76	**3.88**	15.36
	Deeplab v3+	71.71	58.94	7.14	21.15
	UNet++	**82.17**	37.49	4.04	**13.79**
	Swin UNETR	75.21	39.27	5.27	19.52

decided against. Specifically, the transferred Deeplabv3+ network achieved an improvement in the Dice coefficient by 2.31%, the UNet++ network improved by 2%, and the Swin UNETR network improved by .79%. Among the nine models, the UNet++ network demonstrated outstanding performance after transfer learning, achieving an accuracy of 82.17%. The transferred UNet network followed closely with an accuracy of 80.76%. In the realm of deep convolutional neural networks, the worst-performing model was the end-to-end trained Deeplab v3+, with an accuracy of only 69.4%. The HD metric was at its lowest for the end-to-end trained UNet++ network, indicating a high degree of similarity between its segmentation results and the reference standard. The transferred UNet network achieved the lowest false positive segmentation rate, while the transferred UNet++ network achieved the lowest false negative segmentation rate. Partial segmentation results of the ABFM dataset are illustrated in Figure 5.

Segmentation Results on the DTHP Dataset. Table 3 presents the segmentation results for the DTHP dataset. It is evident that the overall segmentation performance of the DTHP dataset surpasses that of the ABFM dataset. Similarly, the segmentation accuracy of deep learning is over 35.91% higher than traditional segmentation methods, and the segmentation performance of transfer learning is better than that of end-to-end trained models. The DRLSE method achieves a Dice coefficient of 54.12% on the DTHP dataset, indicating suboptimal segmentation performance, leading to the decision not to further calculate other performance metrics for this approach. Specifically, the transferred Deeplab v3+ network shows the best performance, with an accuracy of 96.48%, followed by the transferred UNet++ network with an accuracy of 95.61%. All networks achieve segmentation accuracies exceeding 90%. The HD metric is lowest for the end-to-end trained Deeplab v3+ network, indicating a high degree of global similarity between its segmentation results and the reference standard. Among the models pre-trained through transfer learning, the Deeplab v3+ network has the lowest false positive and false

negative rates, followed by the UNet++ network. Figure 6 shows partial segmentation results of the DTHP dataset.

We compiled the segmentation results for different tumor types in the DTHP dataset, as shown in Table 4. It is observed that the model excelled in segmenting Solid Pseudopapillary Tumor (SPT), achieving an outstanding Dice coefficient of 98.02%, with the lowest false-negative rate. For Hepatic Perivascular Epithelioid Cell tumor (HPC), the model demonstrated optimal performance in terms of the HD metric and false-positive rate. However, the segmentation results for Pancreatic Neuroendocrine Tumor (PNET) were relatively poorer compared to the other two tumor types.

Segmentation Results on the DTHL Dataset. Table 5 presents the segmentation results for the DTHL dataset. It can be seen that the segmentation accuracy of deep learning is higher than that of traditional segmentation methods, and the segmentation effect of transfer learning is better than that of end-to-end training. The DRLSE method achieved a Dice coefficient of 64.26% on the DTHL dataset, indicating suboptimal segmentation performance, leading us to decide against further computing other performance metrics. Specifically, the transfer-learned UNet++ network demonstrated the best performance with a Dice coefficient of 85.49%, along with the lowest HD metric and false-negative rate. Figure 7 displays partial segmentation results for the DTHL dataset.

Discussion

Analysis of Experimental Results

Considering the segmentation results of nine models on three datasets, it can be observed that in the task of segmenting intraoperative fluorescence images, the segmentation performance of deep learning methods is superior to traditional methods, and the accuracy after transfer learning is better than direct end-to-end training. According to the experimental results, the UNet++ network with transfer learning performs the best overall in segmenting tumor fluorescence images.
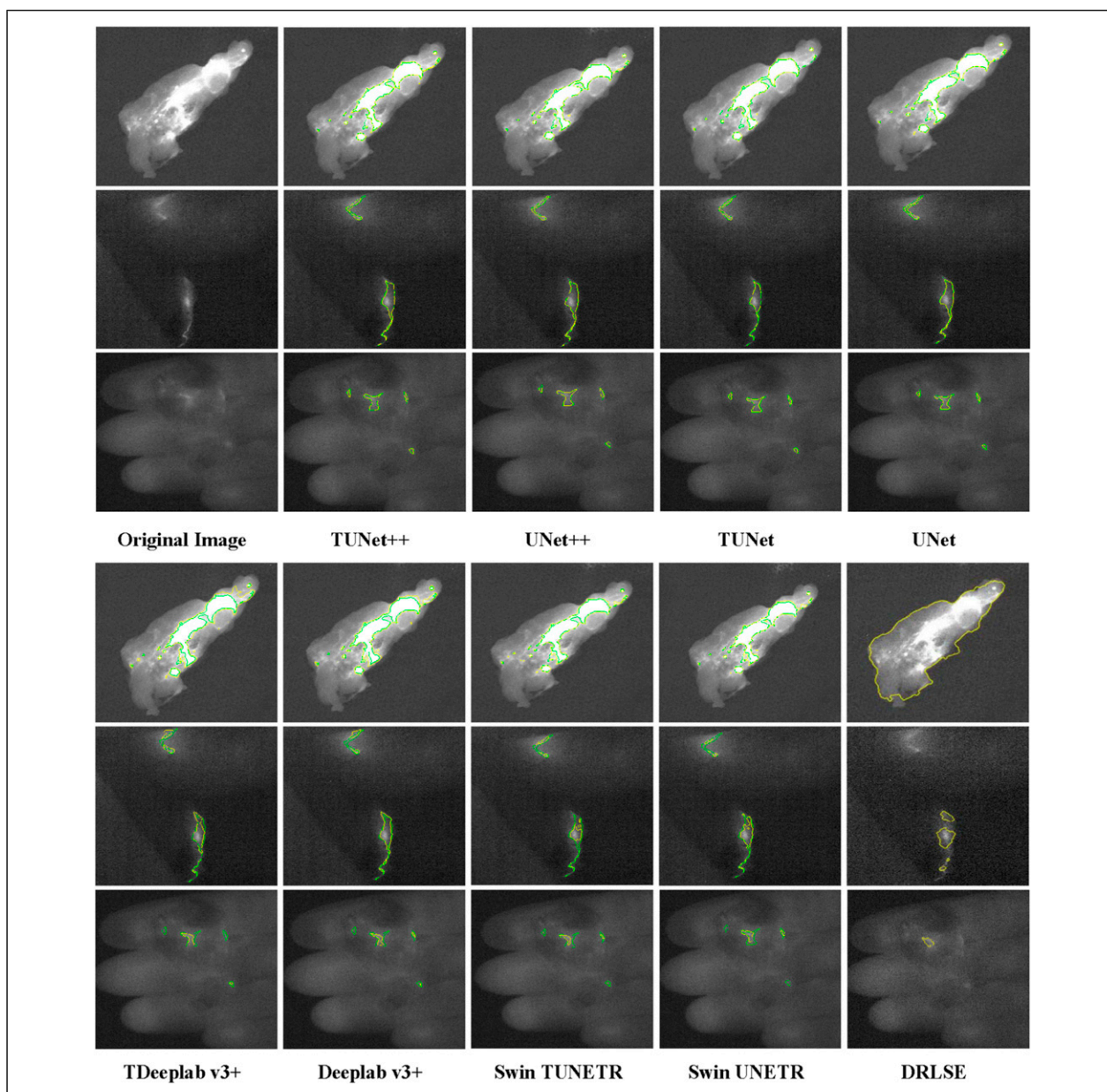
Figure 5. Display of ABFM dataset segmentation results. The green and yellow contours represent the ground truth and the predicted segmentation, respectively. Among them, TUNet++, TUNet, TDeeplab v3+, and Swin TUNETR represent transfer learning models.

Figures 8 and 9 depict detailed renderings of the ABFM dataset and DTHP dataset, respectively, showcasing the effects of different segmentation network models. From the figures, it is evident that the transferred Deeplab v3+ network excels in segmenting large lesions; however, its segmentation accuracy is relatively lower for small and dense lesions. Several factors contribute to this:

Pixel-level segmentation difficulty: Segmenting small and densely distributed lesions at the pixel level is generally more challenging due to the limited number of

pixels, making it harder for the network to capture local features.

Receptive field issue: The backbone network used in Deeplab v3+ may have a larger receptive field, focusing more on capturing global contextual information. While effective for segmenting large lesions, it may not be as suitable for smaller lesions, as a larger receptive field might make it difficult for the network to accurately locate the positions and boundaries of small lesions.

Training data issue: The training dataset may contain relatively few samples of small lesions, hindering the

Table 3. DTHP Dataset Segmentation Results. Bold Indicates the Best Performance.

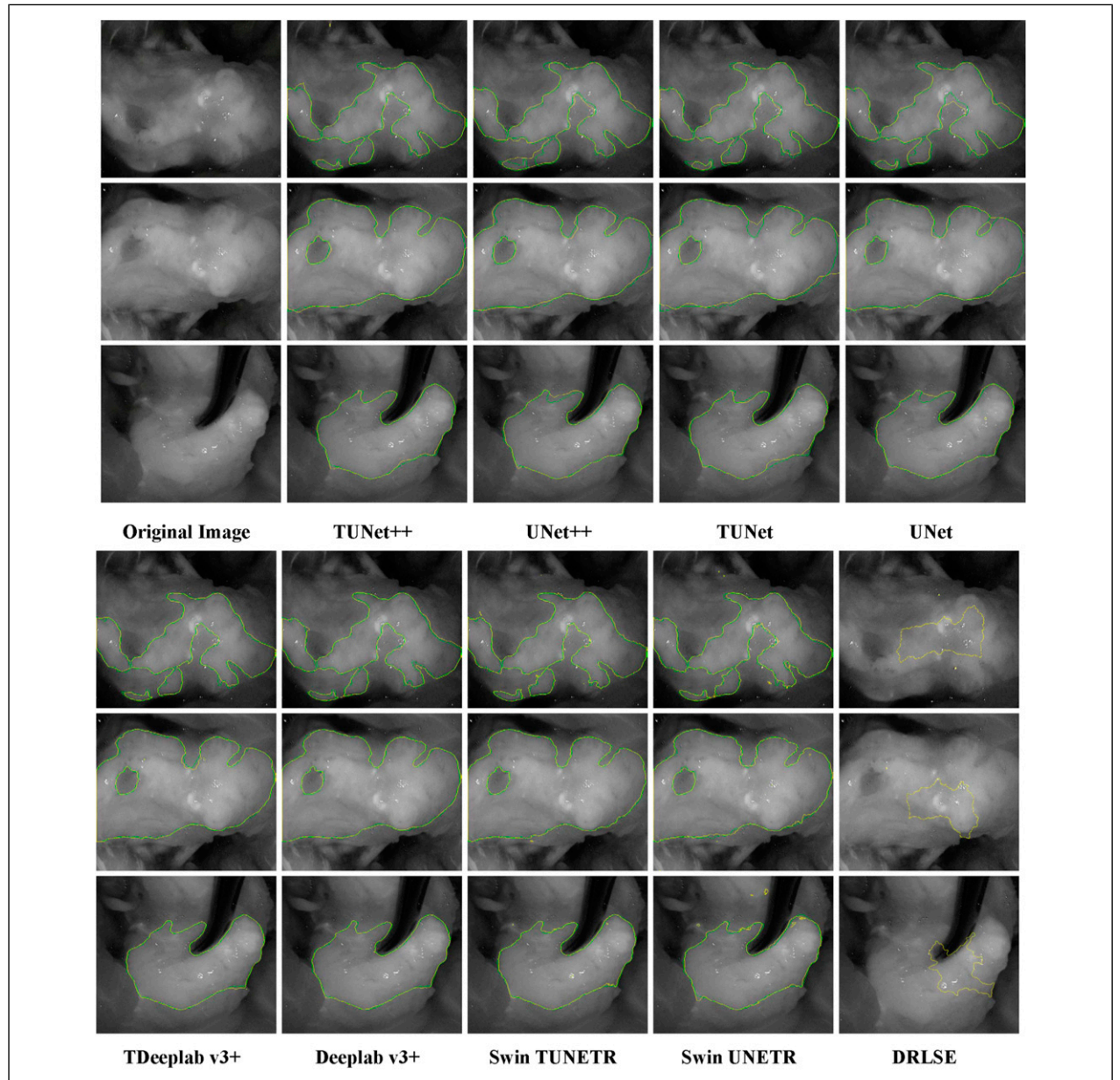| Methods	Backbones	Dice ↑	HD ↓	FPSR ↓	FNSR ↓
Traditional segmentation	DRLSE	54.12	–	–	–
End-to-End Learning	UNet	94.25	32.96	2.77	2.98
	Deeplab v3+	96.35	**20.03**	1.83	**1.82**
	UNet++	94.90	29.16	2.61	2.49
	Swin UNETR	90.03	101.32	4.19	5.78
Transfer Learning	UNet	94.33	32.11	2.84	2.82
	Deeplab v3+	**96.48**	20.25	**1.70**	**1.82**
	UNet++	95.61	25.77	2.03	2.37
	Swin UNETR	91.82	82.14	3.27	4.91

Figure 6. Display of DTHP dataset segmentation results. The green and yellow contours represent the ground truth and the predicted segmentation, respectively. Among them, TUNet++, TUNet, TDeeplab v3+, and Swin TUNETR represent transfer learning models.

Table 4. Comparison of Tumor Segmentation Performance by the Transferred Deeplab v3+ for Patients With Different Type Tumors on the DTHP Dataset in Terms of Four Metrics (Arrows Indicate Which Direction is Better). Bold Indicates the Best Performance.

Tumor type	Dice ↑	HD ↓	FPSR ↓	FNSR ↓
PNET	95.73	22.76	1.94	2.33
SPT	**98.02**	18.53	1.29	**.68**
HPC	97.54	**8.62**	**1.15**	1.31

network's ability to learn how to accurately segment these lesions. UNet networks typically have a smaller receptive field, making them better suited for segmenting small lesions by specializing in extracting local features from high-resolution images.

Moreover, compared to deep neural network methods, the traditional level set DRLSE method demonstrates lower accuracy in segmenting intra-operative fluorescence images, possibly due to frequent noise interference and variations in lighting.

Table 5. DTHL Dataset Segmentation Results. Bold Indicates the Best Performance.

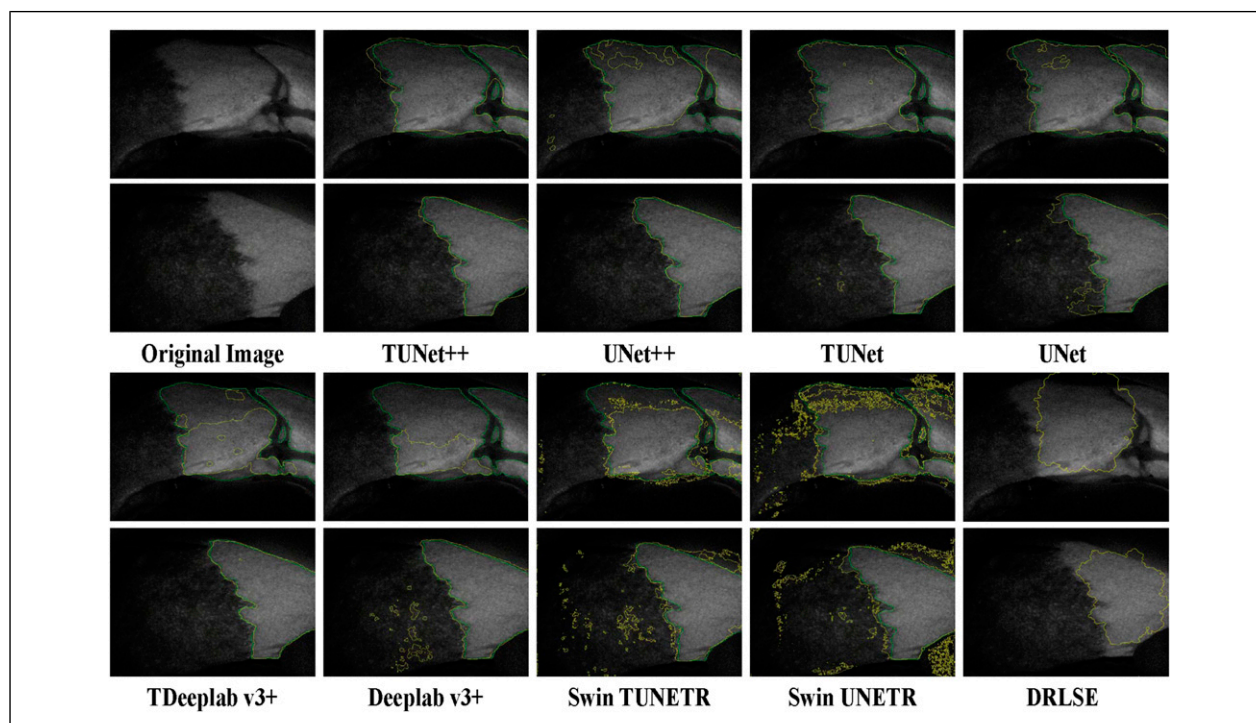| Methods	Backbones	Dice ↑	HD ↓	FPSR ↓	FNSR ↓
Traditional segmentation	DRLSE	64.26	–	–	–
End-to-End Learning	UNet	81.48	303.21	5.63	12.89
	Deeplab v3+	80.97	381.06	8.87	10.16
	UNet++	83.68	335.40	**5.03**	11.29
	Swin UNETR	75.59	505.02	8.84	15.57
Transfer Learning	UNet	83.59	342.25	10.73	5.67
	Deeplab v3+	82.76	321.95	8.29	8.95
	UNet++	**85.49**	**293.47**	12.49	**2.02**
	Swin UNETR	77.66	495.03	6.81	15.53

Figure 7. Display of DTHL dataset segmentation results. The green and yellow contours represent the ground truth and the predicted segmentation, respectively. Among them, TUNet++, TUNet, TDeeplab v3+, and Swin TUNETR represent transfer learning models.
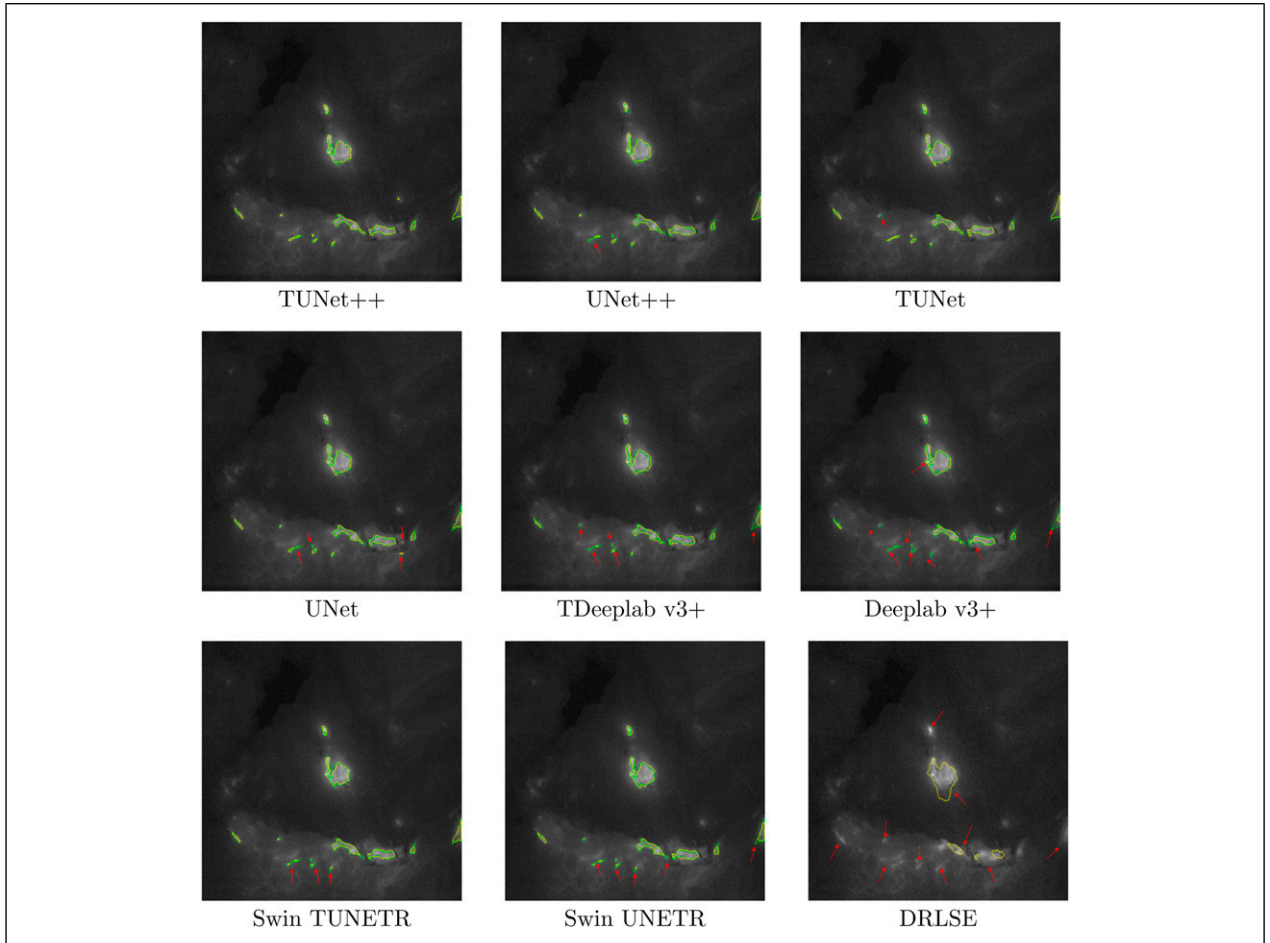
Figure 8. Comparison of segmentation results of different network models on the ABFM dataset. The green line represents the ground truth segmentation, the yellow line represents the predicted segmentation, and the red arrows point to areas with poor segmentation performance.

DRLSE's lower sensitivity to noise and lighting changes may affect its effectiveness in addressing these issues. Additionally, in certain scenarios, the boundaries of targets in intraoperative fluorescence images may lack clarity, posing challenges for level set methods to precisely locate segmentation boundaries.

Limitations

This study acknowledges certain limitations in the following aspects, and we discuss these limitations. Firstly, in terms of data, we recognize the relative limitation of the current dataset. To address this issue, we employed a transfer learning strategy to tackle the challenges of small-sample training. By utilizing models trained on existing datasets and fine-tuning them on relatively smaller new datasets, we could more effectively leverage the limited intraoperative fluorescence images. Compared to end-to-end training, the transfer learning strategy improved the

performance of the segmentation model, a viewpoint strongly supported by our experimental results.

We acknowledge that the field of intraoperative fluorescence imaging is continuously evolving, with the introduction of new technologies and devices potentially leading to inconsistencies in the generated images or videos. This inconsistency further results in a lack of standardization in datasets. It is crucial to fully consider this inconsistency in the data processing and analysis process. Therefore, in this experiment, we selected frames from the collected intraoperative fluorescence imaging surgical videos where tumors were clearly visible. Subsequently, we applied preprocessing, cropping, and data augmentation to obtain images with a resolution of 1024×1024. We established this relatively standardized processing workflow, which can be applied similarly to new data.

Regarding image annotation, we adopted a two-stage annotation approach. Initially, we employed a preliminary annotation method based on relevant
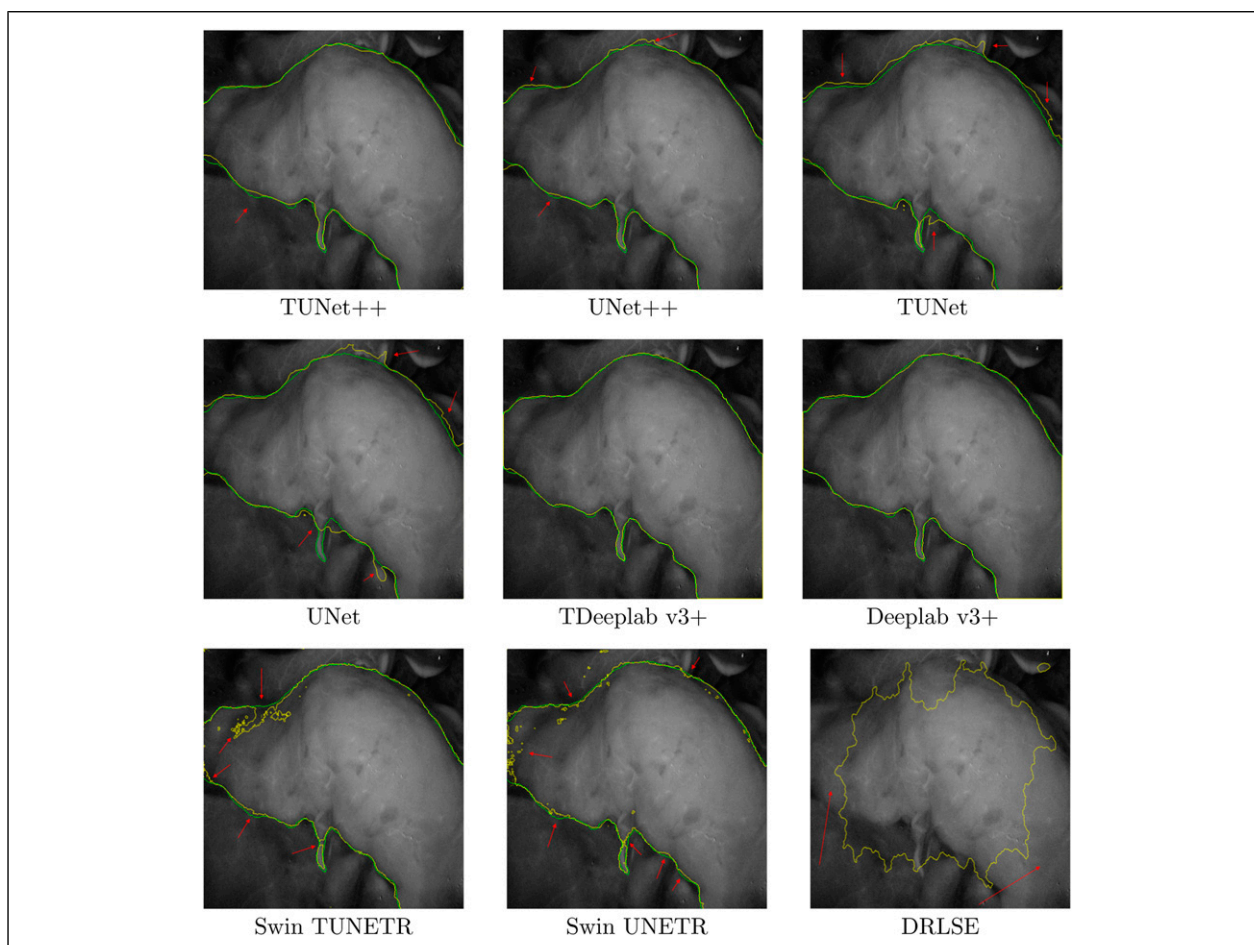
Figure 9. Comparison of segmentation results of different network models on the DTHP dataset. The green line represents the ground truth segmentation, the yellow line represents the predicted segmentation, and the red arrows point to areas with poor segmentation performance.

references rather than relying solely on direct annotations by medical professionals. Direct annotation by doctors is time-consuming and involves substantial workload, especially when dealing with large medical image datasets. To enhance annotation efficiency, our strategy involves preliminary annotation based on video references and personal experience. Subsequently, these preliminary annotations are submitted to medical professionals for correction, ensuring the accuracy and professionalism of the annotations. This two-stage annotation process strikes a balance between annotation speed and accuracy, making the annotation process more efficient.

We actively collect new data to continually expand the dataset. Currently, our research team focuses primarily on pancreatic cancer, emphasizing the collection of intraoperative fluorescence navigation data related to pancreatic cancer. In future work, we aim to establish a large-scale dataset covering multiple organs and diseases. By integrating diverse data, we strive to enhance the model's versatility, enabling it to perform well in tasks related to different tumor types and organs. The preliminary research on intraoperative fluorescence images lays the foundation for more in-depth future work.

This study focuses on the segmentation of intraoperative fluorescence images, initially exploring the model's training limitations due to data volume constraints. Therefore, we employed transfer learning, achieving certain results. Although we initially applied this segmentation method to intraoperative fluorescence navigation images, it is essential to emphasize that the applicability of this method is not limited to specific types of images. In fact, the combination of transfer learning and deep neural networks for image segmentation has been successfully applied in other medical imaging domains. We view the application of this method to intraoperative fluorescence images as pioneering work, laying the foundation for future applications in a broader range of medical image fields.

Conclusion

This paper addresses the task of tumor segmentation for intraoperative fluorescence images for the first time. In network training, we employ techniques such as transfer learning, data augmentation, and fine-tuning due to the limited size of intraoperative fluorescence image datasets. Specifically, we pre-trained four networks (UNet++, UNet, Deeplab v3+, and Swin UNETR) on the ImageNet dataset. Subsequently, we fine-tuned these networks using two sets of different intraoperative fluorescence image data (ABFM and DTHP), conducting model training and validation. Finally, we tested the trained models on another set of intraoperative fluorescence images (DTHL). As comparative experiments, we performed end-to-end training of the same networks and employed the traditional level-set method. The results indicate that deep learning segmentation methods outperform traditional methods in intraoperative fluorescence image segmentation, and the accuracy improves further with transfer learning compared to direct end-to-end training.

Our work represents a novel exploration in the field of medical imaging, aiming to enhance the automatic segmentation performance of intraoperative fluorescence images. By leveraging the advantages of deep learning and a profound understanding of fluorescence images, we aim to provide surgeons with more reliable and accurate image segmentation results during surgery. This endeavor holds promise in improving the accuracy and success rates of tumor resection procedures, thereby reducing unnecessary harm to patients.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

Weijia Hou ⬤ https://orcid.org/0009-0001-4016-9420

References

1. Bertrand TE, Cruz A, Binitie O, Cheong D, Letson GD. Do surgical margins affect local recurrence and survival in extremity, nonmetastatic, high-grade osteosarcoma? *Clin Orthop Relat Res*. 2016;474:677-683. doi:10.1007/s11999-015-4359-x.
2. Hernot S, van Manen L, Debie P, Mieog JSD, Vahrmeijer AL. Latest developments in molecular tracers for fluorescence image-guided cancer surgery. *Lancet Oncol*. 2019;20: e354-e367. doi:10.1016/S1470-2045(19)30317-1.
3. Barabino G, Porcheron J, Cottier M, et al. Improving surgical resection of metastatic liver tumors with near-infrared optical-guided fluorescence imaging. *Surg Innov*. 2016;23: 354-359. doi:10.1177/1553350615618287.
4. Wang X, Teh CSC, Ishizawa T, et al. Consensus guidelines for the use of fluorescence imaging in hepatobiliary surgery. *Ann Surg*. 2021;274:97-106. doi:10.1097/SLA.0000000000004718.
5. Alander JT, Kaartinen I, Laakso A, et al. A review of indocyanine green fluorescent imaging in surgery. *J Biomed Imaging*. 2012;2012:940585-940587. doi:10.1155/2012/940585.
6. Whitney MA, Crisp JL, Nguyen LT, et al. Fluorescent peptides highlight peripheral nerves during surgery in mice. *Nat Biotechnol*. 2011;29:352-356. doi:10.1038/nbt.1764.
7. Yamamoto M, Sasaguri S, Sato T. Assessing intraoperative blood flow in cardiovascular surgery. *Surg Today*. 2011;41: 1467-1474. doi:10.1007/s00595-010-4553-0.
8. Detter C, Wipper S, Russ D, et al. Fluorescent cardiac imaging: a novel intraoperative method for quantitative assessment of myocardial perfusion during graded coronary artery stenosis. *Circulation*. 2007;116:1007-1014. doi:10.1161/CIRCULATIONAHA.106.655936.
9. Gao F, Zhao H, Tanikawa Y, Yamada Y. A linear, featured-data scheme for image reconstruction in time-domain fluorescence molecular tomography. *Opt Express*. 2006; 14:7109-7124. doi:10.1364/oe.14.007109.
10. Zhang C, Wang K, Tian J. Adaptive brightness fusion method for intraoperative near-infrared fluorescence and visible images. *Biomed Opt Express*. 2022;13:1243-1260. doi:10.1364/BOE.446176.
11. Zhang P, Luo H, Zhu W, et al. Real-time navigation for laparoscopic hepatectomy using image fusion of pre-operative 3D surgical plan and intraoperative indocyanine green fluorescence imaging. *Surg Endosc*. 2020;34: 3449-3459. doi:10.1007/s00464-019-07121-1.
12. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Paper presented at: MICCAI 2015: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention; October 5-9, 2015; Munich, Germany. Accessed June 20, 2023.
13. Kermi A, Mahmoudi I, Khadir MT. Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. Paper presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018; September 16, 2018; Granada, Spain. Accessed June 20, 2023.
14. Gao X, Cai Y, Qiu C, Cui Y. Retinal blood vessel segmentation based on the Gaussian matched filter and U-Net. Paper presented at: 10th International Congress *on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); October 14-16*, 2017; Shanghai, China. Accessed June 21, 2023.
15. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested U-Net architecture for medical image segmentation. Paper presented at: 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; September 20, 2018; Granada, Spain. Accessed June 21, 2023.

16. Wang J, Peng Y, Jing S, Han L, Li T, Luo J. A deep-learning approach for segmentation of liver tumors in magnetic resonance imaging using UNet+. *BMC Cancer*. 2023;23(1): 1060. doi:10.1186/s12885-023-11432-x.

17. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Paper presented at: Proceedings of the European Conference on Computer Vision (ECCV); September 8-14, 2018; Munich, Germany. Accessed June 21, 2023.

18. Khodadadi Shoushtari F, Sina S, Dehkordi ANV. Automatic segmentation of glioblastoma multiform brain tumor in MRI images: using Deeplabv3+ with pre-trained Resnet18 weights. *Phys Med*. 2022;100:51-63. doi:10.1016/j.ejmp.2022.06.007.

19. Sun Y, Shi C. Liver tumor segmentation and subsequent risk prediction based on Deeplabv3+. Paper presented at: *IOP Conference Series: Materials Science and Engineering*; October, 2019; United Kingdom. Accessed June 21, 2023.

20. Roy Choudhury A, Vanguri R, Jambawalikar SR, Kumar P. Segmentation of brain tumors using DeepLabv3+. Paper presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018; September 16, 2018; Granada, Spain. Accessed June 21, 2023.

21. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNet: Swin transformers for semantic segmentation of brain tumors in MRI images. Paper presented at: International MICCAI Brainlesion Workshop; September 27, 2021; Online. Accessed June 23, 2023.

22. Kakavand R, Palizi M, Tahghighi P, et al. Integration of Swin UNETR and statistical shape modeling for a semi-automated segmentation of the knee and biomechanical modeling of articular cartilage. *Sci Rep*. 2024;14(1):2748.

23. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. Paper presented at: *AMIA Annual Symposium Proceedings*; November, 2017; American. Accessed June 23, 2023.

24. Sanford TH, Zhang L, Harmon SA, et al. Data augmentation and transfer learning to improve generalizability of an automated prostate segmentation model. *AJR Am J Roentgenol*. 2020;215(6):1403-1410. doi:10.2214/AJR.19.22347.

25. Amiri M, Brooks R, Rivaz H. Fine-tuning U-Net for ultrasound image segmentation: different layers, different outcomes. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2020;67(12): 2510-2518. doi:10.1109/TUFFC.2020.3015081.

26. Kaur B, Lemaître P, Mehta R. Improving pathological structure segmentation via transfer learning across diseases. Paper presented at: First MICCAI Workshop, DART 2019, and First International Workshop. MIL3ID 2019; October 13 and 17, 2019; Shenzhen, China. Accessed June 23, 2023.

27. Lu Q, Liu W, Zhuo Z, et al. A transfer learning approach to few-shot segmentation of novel white matter tracts. *Med Image Anal*. 2022;79:102454. doi:10.1016/j.media.2022.102454.

28. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-1128. doi:10.1016/j.neuroimage.2006.01.015.

29. Zeiser FA, da Costa CA, Zonta T, et al. Segmentation of masses on mammograms using data augmentation and deep learning. *J Digit Imaging*. 2020;33:858-868. doi:10.1007/s10278-020-00330-4.

30. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 23-28, 2014; Columbus, OH, USA. Accessed June 23, 2023.

31. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35(5):1285-1298. doi:10.1109/TMI.2016.2528162.

32. Samala RK, Chan HP, Hadjiiski LM, Helvie MA, Cha KH, Richter CD. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. *Phys Med Biol*. 2017;62(23):8894-8908. doi:10.1088/1361-6560/aa93d4.

33. Pang S, Yu Z, Orgun MA. A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images. *Comput Methods Programs Biomed*. 2017; 140:283-293. doi:10.1016/j.cmpb.2016.12.019.

34. Yu Y, Hu S, Yang Y, et al. Successive monitoring surveys of selected banned and restricted pesticide residues in vegetables from the northwest region of China from 2011 to 2013. *Information*. 2017;18(3):91. doi:10.3390/info8030091.

35. Karri SPK, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed Opt Express*. 2017;8(2): 579-592. doi:10.1364/BOE.8.000579.

36. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Adv Neural Inf Process Syst* 2014;27:3320-3328. doi:10.5555/2969033.2969197

37. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L.Imagenet: A large-scale hierarchical image database. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 20-25, 2009; Miami, FL, USA. Accessed June 23, 2023.

38. Dai J, He K, Sun J. Instance-aware semantic segmentation via multi-task network cascades. Paper presented at: *IEEE conference on computer vision and pattern recognition; June 27-30*, 2016; Las Vegas, Nevada, USA. Accessed June 23, 2023.

39. Li C, Xu C, Gui C, Fox MD. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans Image Process*. 2010;19(12):3243-3254. doi:10.1109/TIP2010.2069690.

40. Zou L, Cai Z, Qiu Y, Gui L, Mao L, Yang X. CTG-Net: an efficient cascaded framework driven by terminal guidance mechanism for dilated pancreatic duct segmentation. arXiv preprint arXiv:2303.02944; 2023. 10.1088/1361-6560/acf110.