




Article

Image Navigation System for Thoracoscopic Surgeries Driven by Nuclear Medicine Utilizing Channel R-CNN

Chuanwang Zhang ^{1,2}, Yueyuan Chen ^{1,2}, Dongyao Jia ^{3,*}  and Bo Zhang ^{1,2,4}

¹ China Nuclear Power Engineering Co., Ltd., Beijing 100840, China; zhangcwb@cnpe.cc (C.Z.); chenyyd@cnpe.cc (Y.C.); zhangbob@cnpe.cc (B.Z.)

² CNNC Engineering Research Center for Fuel Reprocessing, Beijing 100840, China

³ School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China

⁴ School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China

* Correspondence: dongyaojia1974@163.com; Tel.: +86-188-1149-2178

Abstract: Breast cancer, a prevalent and significant cause of cancer-related mortality in women, often necessitates precise detection through nuclear medicine techniques. Despite the utility of computer-aided navigation in thoracoscopic surgeries like mastectomy, challenges persist in accurately locating and tracking target tissues amidst intricate surgical scenarios. This study introduces a novel system employing a channel R-CNN model to automatically segment target regions in thoracoscopic images and provide precise cutting curve indications for surgeons. By integrating a Detection Network Head and Thorax Network Head, this multi-channel framework outperforms existing single-task models, marking a pioneering effort in cutting curve indication for thoracoscopic procedures. Utilizing a specialized dataset, the model achieves a notable region segmentation *mIOU* of 79.4% and *OPA* of 83.2%. In cutting path planning, it attains an *mIOU* of 68.6% and *OPA* of 77.5%. The system operates at an average speed of 23.6 frames per second in videos, meeting the real-time response needs of surgical navigation systems. This research underscores the potential of advanced imaging and AI-driven solutions in enhancing precision and efficacy in thoracoscopic surgeries.

Keywords: image-guided surgery; semantic segmentation; surgical navigation; precision surgery



Academic Editor: Rossella Bedini

Received: 17 December 2024

Revised: 24 January 2025

Accepted: 25 January 2025

Published: 30 January 2025

Citation: Zhang, C.; Chen, Y.; Jia, D.; Zhang, B. Image Navigation System for Thoracoscopic Surgeries Driven by Nuclear Medicine Utilizing Channel R-CNN. *Appl. Sci.* **2025**, *15*, 1443. <https://doi.org/10.3390/app15031443>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last decades, breast cancer has ranked as the most commonly diagnosed cancer among US women (excluding skin cancers) and is the second leading cause of cancer death among women after lung cancer [1]. There were 2.26 million cases of female breast cancer in the year 2020, which constitutes the most commonly diagnosed cancer worldwide [2]. Due to the reluctance to undergo radiation therapy and the fear of recurrence, breast-conserving-surgery-eligible patients are increasingly electing mastectomy [3]. Thoracoscope surgery is a technique for breast-related operation in the treatment of thorax illness [4], which prolongs progression-free survival rates and improves the prognostic effect for metastatic diseases. For thoracoscope surgery, the entire breast may be removed, including all of the breast tissue and sometimes other nearby tissues. Both breasts of severe patients could be removed in a double mastectomy [5,6]. Therefore, thoracoscope surgery requires the experiences and skills of surgeons [7]. Errors in operations, including target area miscuts and incorrect cutting curves, may lead to blood vessel bleeding, spasm, and even paralysis and surgical failure [8]. Moreover, reoperation of the surgery

could also increase patient stress and the risk of failure. Aiming at addressing such problems, computer-aided surgery systems can provide effective indications by giving accurate guidance regarding target issues to promote the surgeon's perception in the operation environment and the control of surgical tools [9–11]. Recent studies about thoracoscope surgery navigation mainly focus on the resection of small and deep-seated pulmonary nodules [12], the assessment of the micro-coil position relative to the lesions [13], and the localization of the tumor [14]. To our best knowledge, there are no related systems for thoracoscope surgery to guide the surgeons' cutting paths.

For the cutting curve indication in thoracoscopic surgery, given in Figure 1a, specific category segmentation masks (e.g., electrotome, fibrous tissue, and pectoralis major) are depicted in Figure 1b. Our proposed thoracoscopic surgery indication system based on the channel R-CNN is applied to segment, classify, and track the excised contours of target areas and to generate the cutting curves for clinical operation. However, detecting and tracking the target areas of thoracoscopic tissues under fickle and complex surgery situations, which usually includes the surgery instrument occlusion and blurred blood stains, could be a challenging task for clinical applications. Furthermore, operation errors like the cutting of target areas of the pectoralis major and blood vessels may influence the effects of mastectomy and can even cause surgery failure. Therefore, traditional approaches including machine learning segmentation models are not suitable for this task under certain situations.

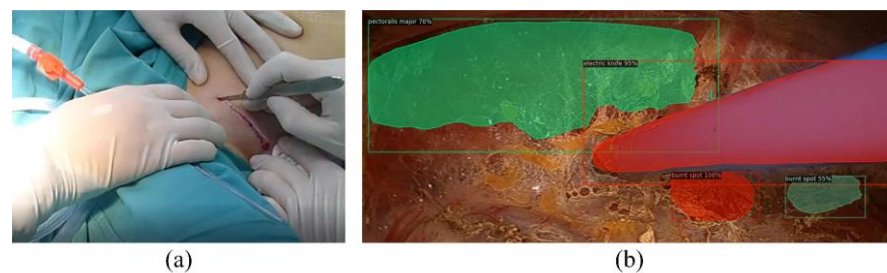


Figure 1. System interface. (a) Clinical shooting of thoracoscopic surgery system. (b) Target region indication of a specific frame.

This paper proposes a novel system that can automatically and accurately segment target regions from thoracoscopic surgery images and provide cutting curve indications for surgeons. Mask R-CNN [15] is utilized to build the proposed two-stage structure network, the channel R-CNN, where the feature maps of surgery frames are extracted by DenseNet first. Then, the features are input into the parallel channels to generate the instance segmentation masks (channel 1#) and the rough cutting curve (channel 2#). Our previously proposed leapfrog algorithm [16] is applied for path planning in the prediction of refined cutting curves. We also design an improved objective function to evaluate the segmentation performance of the rough cutting path; moreover, the replanning strategy based on the leapfrog algorithm chooses the optimized refined cutting curve for further generation.

The proposed system is related to image segmentation, cutting indication, and computer-aided surgery navigation. These works are briefly reviewed in this section.

Image segmentation. Recent advancements in neural networks have significantly transformed the field of image segmentation [17–21]. Liu et al. [22] proposed an efficient medical image segmentation network based on an alternating mixture of a CNN and Transformer tandem, which is called Eff-CTNet, which achieved better performance with less computation sources. Alam et al. [23] proposed a graph model initialized by a fully convolutional network (FCN) named Graph-FCN for image semantic segmentation. H-DenseUNet [24] extracted the intra-slice features efficiently and used the 3-D counterpart

with hierarchically aggregating volumetric contexts for liver and tumor segmentation. Chen et al. [25] highlighted the convolution with upsampled filters, atrous spatial pyramid pooling (ASPP), and combining methods from DCNNs as well as probabilistic graphical models for better segmentation performance. Mask R-CNN [15], which served as the base model for this paper, has been widely applied for medical image segmentation and achieves accurate and fast performance by adding a branch for predicting the object mask. Mask R-CNN is still one of the most representative segmentation models and is widely applied in clinical contexts; its variations [26–29] provide stable foundations for medical image segmentation. However, the proposed channel R-CNN distinguishes itself by employing a novel multi-channel architecture; by integrating DenseNet as a base model, it processes feature maps through two parallel channels: one for instance segmentation using Mask R-CNN and another for generating surgical cutting curves. This dual-channel approach not only enhances segmentation accuracy but also provides real-time surgical guidance, setting it apart from traditional single-task models.

Cutting indication. An improper cutting curve not only makes the surgery difficult but may also lead to the injury of other tissues; therefore, cutting curve indication is a difficult and critical challenge in medical applications [30]. Chrysovergis et al. [31] applied the CNN model for the assessment of unconstrained surgical cuttings in VR; different cutting trajectories were distinguished and selected for optimization. Fast and efficient fluid dynamic visualizations for the heart surgery simulation were proposed by Sugeng et al. [32]; comparison of the frame rates for the surgery simulation proved the effectiveness of their approaches. Jin et al. [33] presented the meshless total Lagrangian adaptive dynamic relaxation (MTLADR) algorithm to address the shortcomings, including high computational cost and the need for re-meshing in clinical surgical cutting. Tang et al. [34] designed a hybrid CNN–Transformer network to capture both the local and global information and performed experiments on two datasets to demonstrate its superior capability. A periacetabular tumor resection was simulated using a pelvic bone model by Cartiuax et al. [35], by which the location of the cut planes with respect to the target planes was significantly improved. However, the existing methods still face limitations with real-time requirements and are not able to handle the cutting curve indication in thoracoscopic surgery; high computation costs and a complex operation background add difficulties to their clinical applications. Despite these advancements, existing methods face limitations in real-time applications and complex surgical environments. The channel R-CNN overcomes these challenges by incorporating a Thorax Network Head (TNH) that generates rough cutting curves, specifically designed for thoracoscopic surgery. This integration allows for real-time navigation and precise cutting path planning, addressing the high computational demands and operational complexities of current methods.

Computer-aided surgery navigation system. Image-guided surgery navigation techniques have been widely applied in various clinical scenes [36–38]. The main purpose of image-guided surgery (IGS) is to provide help to surgeons in order to perform safer and less invasive procedures while removing tissue tumors, resulting in surgeries that are conducted more efficiently and less riskily. As for thoracoscopic surgery, Lee et al. [14] developed a thoracoscopic surgical navigation system with real-time augmented image guidance to assess the potential benefits for minimally invasive resection of chest tumors, which improved the accuracy of tumor localization. Hanna et al. [39] described the evolution of thoracoscopic spine surgery from basic endoscopic procedures using fluoroscopy and anatomical localization through developmental iterations. Moreover, a virtual reality simulation system [40,41], three-dimensional navigation [42,43], intraoperative fluorescence visualization [44], CT imaging [45], integrated models [46], and holographic laser projection [47] models have been deployed for thoracoscopic surgeries. However, no existing

studies focus on the cutting path planning process of chest surgery. To our best knowledge, the indication difficulties for the clinical operations caused by complex backgrounds and huge computation costs have not yet been addressed. The channel R-CNN contributes to this field by offering a comprehensive solution that combines instance segmentation and surgical navigation. Its dual-channel architecture not only segments surgical areas with high accuracy but also guides cutting paths, providing a level of functionality and integration that surpasses existing multi-task models. By leveraging the strengths of DenseNet and the innovative use of parallel channels, the channel R-CNN sets a new standard in both segmentation and navigation capabilities, making it a valuable tool in the realm of computer-aided surgery.

Moreover, the proposed framework based on the channel R-CNN is an effective attempt at thoracoscopic surgical navigation, which has the advantages of high accuracy, light weight, and fast real-time response. Moreover, influences from occlusion regions by surgical instruments are also reduced, and the robustness of our system is strongly proved by experiments. A flowchart of the proposed framework is given in the next section. The contributions of our system are listed as follows:

- (1) A real-time image-guided thoracoscopic surgery navigation system for surgeon operation with target region segmentation and cutting path indication is proposed. To our best knowledge, this is the first trial in the cutting operation indication of thoracoscopic surgery;
- (2) The channel R-CNN network is innovatively designed, the Detection Network Head (DNH) and Thorax Network Head (TNH) operate in parallel to process the surgery navigation task;
- (3) The improved leapfrog algorithm is applied to refine the cutting curve generation based on rough segmentation results, and the combination with the region detection results ensures the accuracy of the cutting curve and reduces the surgery risk.

This paper is organized as follows. Section 2 presents the Materials and Methods. Results and Discussions are given in Section 3. Section 4 presents the Conclusions.

2. Materials and Methods

2.1. Dataset Preparation

A high-quality dataset carries important prior knowledge for network training. Our proposed dataset is provided by the Peking Union Medical College Hospital comprising six videos captured from real thoracoscopic surgeries. This study was approved by the ethics committee of Research Center for Big Data and Intelligent Measurement and Control from Beijing Jiaotong University; all subjects provided informed consent for enrolling in this study.

2.1.1. Thoracoscopic Surgery Image Dataset

The proposed dataset for segmentation contains 7320 image frames from the thoracoscopic surgery video. All the patients' sensitive information is removed and checked by the hospital's ethics committee. The Nottingham grading system presented by the World Health Organization (WHO) serves as the reference for histological labeling of sample categories.

The 7320 images are from 8 patients who had thoracoscopic surgeries based on the needs of breast beauty. Each frame consists of different imaging components (histopathological and non-histopathological parts), including an electrotome (5190), fibrous tissue (3420), pectoralis major (4510), burnt spots (2030), rib periosteum (2800), fat (3070), blood vessel (1050), and nervous tissue (910), which are presented in Figure 2. In the clinical practice of thoracoscopic surgery, cutting into tissues such as muscles and nerves could

cause potential bleeding, muscle weakness, and paralysis. Therefore, accurate and timely surgery navigation reduces the risk and improves the cutting efficiency.

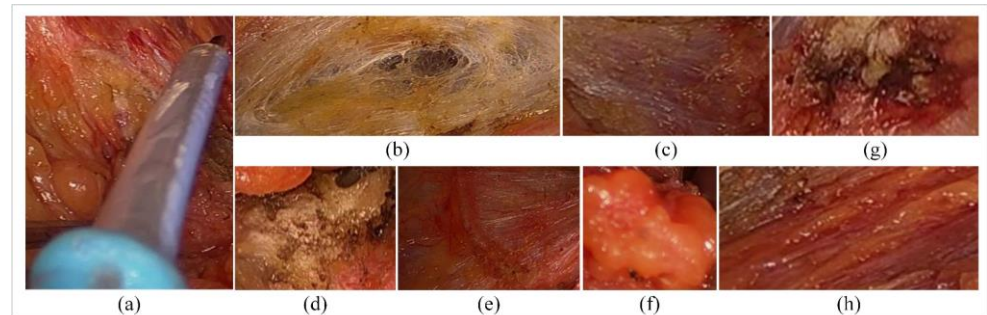


Figure 2. Typical imaging components from video frames. (a) Electrotome. (b) Fibrous tissue. (c) Pectoralis major. (d) Burnt spots. (e) Rib periosteum. (f) Fat. (g) Blood vessel. (h) Nervous tissue.

Slides from the proposed dataset are digitized with a semi-rigid electronic thoracic video endoscope (Olympus, LTF-160, Olympus UK, Southend-on-Sea, UK), which has a flexible tip and incorporates a 2.8 mm working channel. The video output applies the raw format of the thoracoscope; then, the output is divided into the RGB frames with 8-bit color depth for each channel. Typical samples of the dataset are given in Figure 3, by which the visual details are provided.

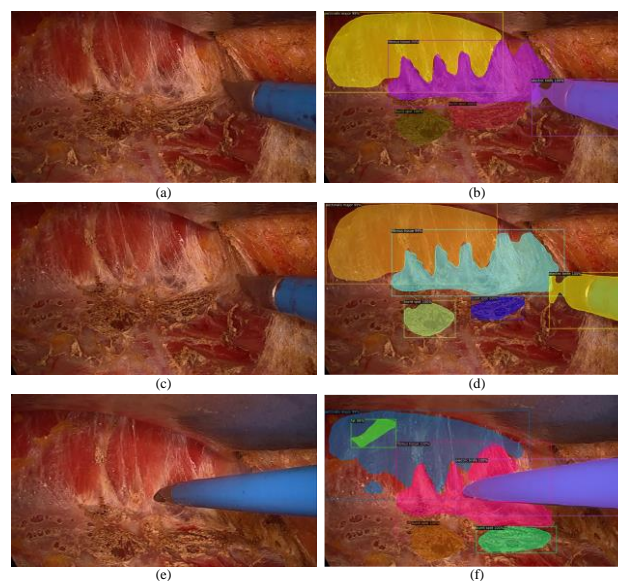


Figure 3. Three samples of the segmentation dataset. (a,c,e) are the original images, and (b,d,f) are the results of the manual annotation and network output. Class labels are provided on the top left corner of the bounding box; segmented masks are distinguished with different color attributes.

Representative samples are extracted and labeled by four pathologists with decades of clinical experience. The content of each frame is hand-annotated with the web-based tool Labellmg [48], the frame size is saved to 1280×720 , and the ground truth is also provided. Aiming at providing effective labels, all the annotations are made in consideration of the consistency principle. The original dataset is preprocessed by the combinations of data augmentation approaches, which include cropping, flipping, mirroring, rotation. Then the preprocessed dataset is input into the network for training. The proposed dataset is divided into training, validation, and testing sets according to the ratio of 6:2:2; 5-fold cross-validation is applied to ensure robustness.

2.1.2. Cutting Curve Dataset

In the video frames, we choose the essential images I_{ess} where cutting indications are completely visible and crucial for thoracoscopic surgeries. The dataset for training the TNH model is generated by manual labeling of the essential images I_{ess} to produce the frames I_{cc} with clear cutting curves. We select 4 frequently faced surgical cutting lines and generate 5158 images in total from the 8 video sequences.

Feature maps generated from DenseNet are applied for the training of the TNH with a resolution of 1280×720 . As given in Figure 4, examples from the cutting curve dataset with corresponding splitting indications labeled by our pathologists are demonstrated. To our best knowledge, there is no available dataset for thoracic cutting indication at present.

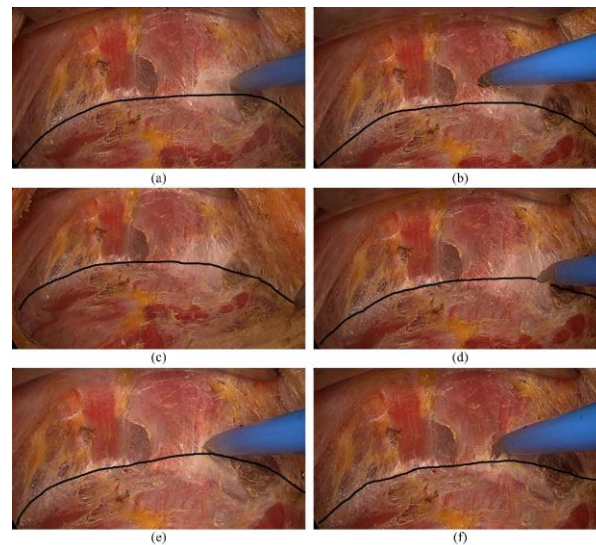


Figure 4. Six samples of the cutting curve dataset. Labeled cutting curves are annotated with the black pixels for illustration. Subfigures (a)–(f) are the six selected samples for illustration.

2.2. Methods

The main content of this section is as follows: First, the navigation problem in thoracoscopic surgery is specifically defined. After that, the proposed framework used for surgery navigation is given. Finally, evaluation metrics which indicate the model navigation performance are provided.

2.2.1. Problem Definition

The proposed model is applied to address the navigation issues of thoracoscopic surgery, which include the cutting indication of the electrotome in operation, as given in Figure 1a. It consists of instance segmentation and trajectory planning. We assign the stromal parts of the original frames as the background and assign the key regions annotated by pathologists as the RoIs. They are the instance segmentation objects including the fibrous tissue, pectoralis major, burnt spots, etc. Thoracoscopic surgery is a step-by-step incision process working towards the muscle gaps using the electrotome; our goal is to provide the cutting indications while eliminating interference from the RoIs. In this case, the surgery navigation system is designed to be a two-stage structure; a rough cutting curve is generated in the first stage; then, it is refined using the restrictive conditions extracted from the RoIs in the second stage, by which the final cutting path is produced. The proposed framework could be achieved by object location (put bounding boxes around the RoIs), classification (determine the specific categories of the RoIs: fibrous tissue, pectoralis major, burnt spots, etc.), instance segmentation (draw accurate masks for the RoIs), and trajectory planning (refinement of the initial cutting curve). The outputs of the instance segmentation from the

proposed channel R-CNN are provided (Figure 1b), and different RoIs are indicated by uniquely colored masks; corresponding class labels are also generated.

2.2.2. Model Definition

(1) Network Structure: As given in Figure 5, details of the proposed channel R-CNN are provided. DenseNet is applied as the base model, which outputs the feature maps of the surgery frames. Then, the features are fed into two parallel channels to generate the instance segmentation masks and the rough cutting curve. In channel 1#, the Mask R-CNN structure is applied. Feature maps from the DenseNet are input into the region proposal network (RPN) [49] to obtain the alternative proposals, which are the RoIs. Then, the proposed Detection Network Head (DNH) is applied to output the specific category, accurate bounding box, and mask graph. As for channel 2#, the proposed Thorax Network Head (TNH) produces a rough cutting curve to navigate the operation of the thoracoscopic surgery, depending on the feature maps from DenseNet. Finally, the refined cutting curve is generated based on the outputs of the DNH and TNH, and an improved leapfrog algorithm [16] provides the path planning for the prediction. Due to the multi-channel architecture of the proposed model, we name it the channel R-CNN.

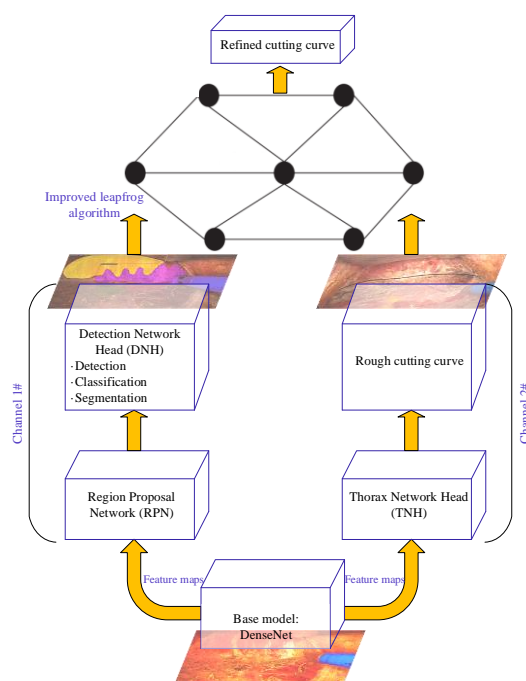


Figure 5. Overview of the proposed channel R-CNN structure. DenseNet is applied as the base model, and feature maps are extracted and fed into the two parallel channels (channel 1# and 2#). The Detection Network Head (DNH) is applied to output the specific category, accurate bounding box, and mask graph; the Thorax Network Head (TNH) produces a rough cutting curve. Finally, the refined cutting curve is generated based on the outputs of the DNH and TNH using the improved leapfrog algorithm.

(2) Objective Function: the proposed system aims to navigate the thoracoscopic surgery by outputting the refined cutting curve. The DNH and TNH models accomplish this task. Specific loss functions are designed: for channel 1#, L_{det} evaluates whether the RoIs are located accurately, L_{cat} is applied to measure the classification accuracy for different thoracic components, and L_{mask} determines whether the bounding boxes regress to true ROI edges.

For channel 2#, a cutting prediction loss L_{cut} is proposed for the TNH, which evaluates the segmentation performance of the cutting path, as follows:

$$L_{cut} = \sum_{k=1}^N (-q_k \log(p_k) - (1 - q_k) \log(1 - p_k)) \quad (1)$$

where N represents the total number of applied frames in the model training; $p_k \in (0, 1)$, $p_k = n_{in}/n_{total}$, and n_{in} represents the pixel number in the two-dimensional confidence interval of the annotation path; n_{total} represents the total pixel number of the cutting curve generated by the TNH. p_k is output by the sigmoid function of the proposed model. $q_k = 1$ when the created curve has an intersection with the ground truth; if not, $q_k = 0$. Therefore, the loss function L of the proposed channel R-CNN consists of two parts (channel 1# and 2#), by which the rough cutting curve is generated.

$$L = \underbrace{L_{det} + L_{cat} + L_{mask}}_{DNH} + \underbrace{L_{cut}}_{TNH} \quad (2)$$

(3) Cutting path replanning: after generating the segmentation mask and the rough cutting curve from the DNH and TNH, we refined the initial cutting path based on the previously published work, using the improved leapfrog algorithm [16]. The mask image is effectively combined with the initial cutting indication information by the refined process, in order to avoid the key areas that would be likely to affect the patients' bodily function if they are cut up accidentally, as given in Figure 6. The image which combines the segmentation mask and the rough cutting line is marked as the directed graph $G = (V, E)$, V represents the pixel nodes, and E stands for the detected edges (the mask E_M and the rough cutting lines E_R). Then, the frogs are randomly generated to construct the initial population, which depicts the potential path among pixel nodes, and they are recorded as $U = (U_1, U_2, \dots, U_d)$; here, d represents the dimension of path space solution. Then, the source node is set as V_S , which iterates over the U to find the optimal path E_O comparing with E_R . Then, the forwarding-satisfaction rate of each option is applied to select a more effective path, which is given as follows:

$$P_S(s, e) = \frac{\sum_{j=1}^n x_{i,j} \delta_{i,j}}{\varepsilon} \quad (3)$$

where $x_{i,j}$ represents the pixel value at the position (i, j) , and $P_S(s, e)$ stands for the forwarding-satisfaction rate from the source node to the end node. $\delta_{i,j}$ is the degree of information correlation between the pixel $x_{i,j}$ and its neighborhood pixels. ε represents the routing hops of the path to be calculated. Then, the optimal path E_O is selected according to the presented routing satisfaction rate.

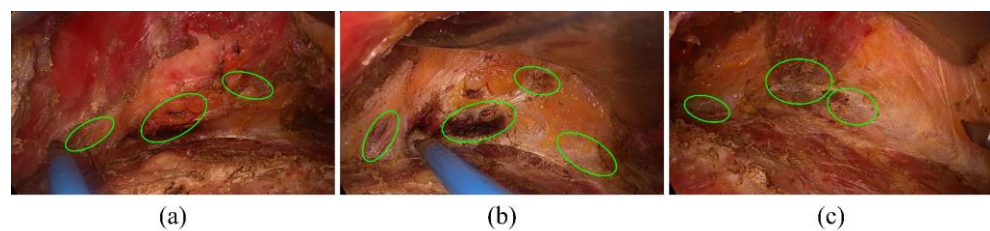


Figure 6. Key areas for the cutting curve replanning. Regions within the green ellipse are vital targets that the refined cutting lines try to avoid. Subfigures (a)–(c) are the three selected samples for illustration.

Moreover, the optimal path E_O is further refined using the detected edges (segmentation mask E_M); we define the positive factor f_{PO} , negative factor f_{NE} , and the penalty factor f_{PE} to modulate the refined path dynamically.

$$E_R = (\alpha \times f_{PO} + (1 - \alpha) \times f_{NE} + f_{PE} \times E_M^* \cap E_O(E)) \rightarrow E_O(E) \quad (4)$$

where $\alpha \in (0, 1)$, which is the adjusting item of the factors. When α increases, the force that constrains the optimal cutting line E_O close to the segmentation mask will increase. f_{PO} and f_{NE} adjust the refined cutting path jointly, which closes to the segmentation mask border while maintaining the rough boundary. E_M^* refers to the alarming parts in the segmentation masks, including the pectoralis major, rib periosteum, and nervous tissue. f_{PE} keeps the refined cutting line away from the sensitive areas by testing the intersection of E_M^* and E_O . Therefore, the initial cutting curve is reoptimized, the critical areas of the thoracoscopic surgery are avoided, the operational efficiency is improved, and the potential risk is reduced. The generation process of the refined cutting curve using the improved leapfrog algorithm is given in Algorithm 1:

Algorithm 1. Refined Cutting Curve Generation

Input: ~ directed graph $G = (V, E)$

Output: ~ fine-tuned cutting path E_F

- 1: Treat pixels of graph G as individuals in frog population and record its size as P , set the population number as M , and set the first and maximum iteration number as 1 and I ;
 - 2: Define the frogs as the vector $\{X_1, X_2, \dots, X_i, \dots, X_p\}$, where $X_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in})$. When $x_{in} = 0$, the node is removed; otherwise, the node is selected;
 - 3: Calculate the forwarding-satisfaction rate $P_S(s, e)$ of each path using Equation (3); then, adjust the frog order;
 - 4: Divide P frogs into M populations, and set the sub-population as $\{F_1, F_2, \dots, F_n, \dots, F_M\}$. Iterate over the potential paths $U = (U_1, U_2, \dots, U_d)$, and set X_h as the path with highest forwarding-satisfaction rate and X_l as the path with the lowest rate. Update the value of X_h in sub-population F_n ;
 - 5: Re-adjust the frogs of sub-population F_n according to $P_S(s, e)$, and generate the optimal subpopulation. Obtain the globally optimal path E_O by comparing with $U = (U_1, U_2, \dots, U_d)$;
 - 6: Re-arrange the frog order in sub-population F_n , and apply the positive factor f_{PO} , the negative factor f_{NE} , and the penalty factor f_{PE} for the calculation of $P_S(s, e)$. Iterate over the frog population on the segmentation mask E_M ;
 - 7: Update the refined cutting path E_F in circulation, loop step 3, and compare with the local optimal cutting path E_L ;
 - 8: **end if**
 - 9: Determine whether the E_F outperforms the E_L ;
 - 10: Update the refined cutting line: $E_F = E_F \cap E_L$;
 - 11: **Return** E_F
-

(4) Implementation and Transfer learning: the proposed channel R-CNN model is constructed based on the Visual Studio 2018, Opencv 3.4.1, and the Tensorflow1.12.0. The Win 10 (64-bit) system with 64 GB of RAM and an Intel Core (TM) i7-8700 are applied for training. Due to the limited hardware configuration, the original input data (1280×720) are divided into 32 patches with overlap, and each patch is resized to 512×512 pixels. The labeled information provided with the resized patches are input into the model for

training; as for the testing stage, the image patches are reassociated together and output as the whole files.

Moreover, the originality of the thoracoscopic surgery navigation system results in insufficient medical image data. The proposed framework is limited by accurate annotation information and the balanced data category. Therefore, transfer learning [50] is applied to address the problem of lacking data. Shallower layers in the deep learning model obtain more general parameters, which map the data from the source feature space to the target feature space [51]. Pre-training preserves the shared parameters and saves the training procedure, which also reduces the overfitting on the limited labeled dataset. Considering the diversity across object categories and scenes, the benchmark status in computer vision tasks, the benefits of transfer learning, and the widespread adoption within the research community, the MS COCO dataset [52] was applied for the pre-training of the channel R-CNN, which contains 91 target categories, 328,000 images, and 2.5 million labels. Parameters of the pre-trained model are initialized with the MS COCO dataset; then, the DNH and TNH channels are fine-tuned with the proposed thoracoscopic sugary image dataset. The channel R-CNN framework is trained with the multi-stage strategy, which is divided into three steps.

Stage 1: pre-train the higher layers (Dense Block and Transition Layer) of the DenseNet backbone [53] with the MS COCO dataset. The DNH is trained using the feature maps from DenseNet. Stochastic Gradient Descent (SGD) with backpropagation is applied for model training. Based on the principles of transfer learning, the DNH is firstly trained for 50 epochs; then, the DenseNet with the DNH is fine-tuned using the thoracoscopic sugary image dataset;

Stage 2: according to the proposed cutting prediction loss function L_{cut} , the fixed parameters in Stage 1 are transferred to the TNH; then, it is fine-tuned with the thoracoscopic sugary cutting dataset, by which the rough cutting line is generated. In the training process, Stochastic Gradient Descent (SGD) is applied. As given in Figure 7, the fined-tuned model is observed to converge after 210 epochs; we found the transfer learning works well in practice;

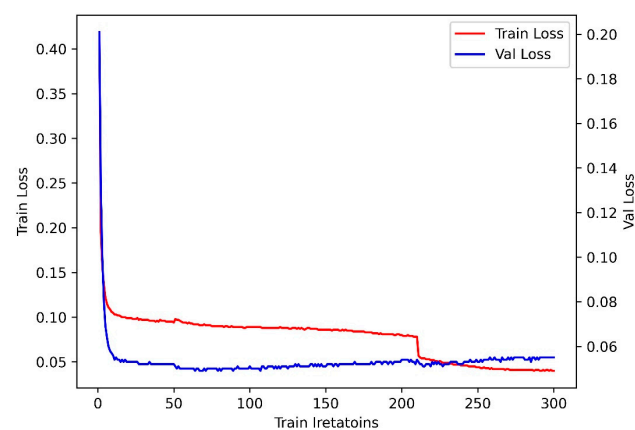


Figure 7. Training process of the proposed model.

Stage 3: take the segmentation mask and the rough cutting curve from the above two stages, initialize and augment the frog population P , and conduct the first and second generation to obtain the globally optimized cutting curve. Set the number of nodes as 1000, the total number of frogs as 2400, and the number of iterations I as 95; then, operate the generation process of the refined cutting curve using the improved leapfrog algorithm, and calculate the forwarding-satisfaction rate for potential paths $U = (U_1, U_2, \dots, U_d)$. Use the

adjusting factors f_{PO}, f_{NE}, f_{PE} to update the forwarding-satisfaction rate $P_S(s, e)$; iterate all the alternative options until the refined cutting path E_F is generated.

2.2.3. Evaluation Metrics

As given in Table 1, P' & N' refer to the prediction categories generated by the proposed channel R-CNN; P & N are the actual categories provided by the annotation. Aiming at illustrating the indication performance of the proposed channel R-CNN framework, the mean Intersection Over Union ($mIOU$), Overall Pixel Accuracy (OPA), Standard Mean Accuracy (SMA), and Dice coefficient (DC) are used to evaluate the segmentation performance. The DNH channel achieves the classification of segmented masks based on the confusion matrix in Table 1.

Table 1. Confusion matrix for the evaluation of detection performance.

	Prediction Category (P')	Prediction Category (N')
Actual category (P)	True positive (TP)	False negative (FN)
Actual category (N)	False positive (FP)	True negative (TN)

As for the segmentation performance, record the segmentation results as r and the annotation as a . Define the pixel confusion matrix $C(C_{1,1}, C_{1,2}, \dots, C_{i,j}, \dots, C_{n,n})$; n is the total number of statistical pixels, and $C_{i,j}$ is the number of pixels annotated as a_i while segmented as r_j . The Jaccard coefficients J_i for classes a_i are calculated as follows:

$$J_i = \frac{TP}{TP + FP + FN} = \frac{C_{i,i}}{T_i + P_j - C_{i,i}} \tag{5}$$

where T_i represents the total number of pixels annotated with a_i , and P_j refers to the number of pixels segmented as r_j . Thus, $mIOU$ is presented as follows:

$$mIOU = \frac{1}{N} \sum J_i \tag{6}$$

N is the number of categories. OPA is given as follows:

$$OPA = \frac{\sum_i C_{i,i}}{\sum_i \sum_j C_{i,j}} \tag{7}$$

SMA is calculated as follows:

$$SMA = \frac{1}{N} \sum_i \frac{C_{i,i}}{\sum_j C_{i,j}} \tag{8}$$

DC is presented as follows:

$$DC = \frac{2TP}{2TP + FP + FN} \tag{9}$$

3. Results and Discussions

3.1. Evaluations on Thoracic Tissue Contour Segmentation

This section discusses the quantitative results for thoracic tissue contour segmentation, which are shown in Table 2. Four methods are applied for performance comparison. FCN-8s achieves competitive results due to its ability to capture spatial information at multiple scales; DeepLab utilizes atrous convolution (also known as dilated convolution) to effectively capture object boundaries and fine details in images; U-Net is commonly used

for biomedical image segmentation tasks and is designed to efficiently capture both local and global features through a U-shaped architecture with skip connections; manual traits refers to a manual feature extraction approach where handcrafted features are used for segmentation, and it involves extracting and selecting features based on domain knowledge and manual analysis. The average performance (indicated by $mIOU$, OPA , SMA , and DC) of the proposed DNH model is compared with these four baseline approaches using our dataset.

Table 2. Performance comparison of segmenting thoracic histological images as “Electrotome” (EL), “Fibrous tissue” (FT), “Pectoralis major” (PM), “Burnt spots” (BS), “Rib periosteum” (RP), “Fat” (FA), “Blood vessel” (BV), and “Nervous tissue” (NT). Time represents the processing time spent on one frame; the unit is ms. Bold results refer to the best performance.

	J_{PM}	J_{RP}	J_{BV}	J_{NT}	$mIOU$	OPA	SMA	Time
FCN-8s [54]	78.5%	75.4%	80.6%	82.2%	76.8%	83.1%	82.8%	24.4
DeepLab [25]	64.4%	51.5%	63.9%	61.1%	59.5%	74.4%	57.9%	27.2
U-Net [55]	80.7%	72.3%	63.7%	75.4%	73.6%	82.0%	85.1%	26.5
Manual traits [56]	61.2%	46.7%	52.3%	N/A	49.5%	76.7%	53.6%	29.1
Channel R-CNN	83.1%	83.6%	73.5%	83.9%	79.4%	83.2%	88.4%	23.3

Table 2 (Row 5) shows the segmentation performance of our model, which achieves 79.4% $mIOU$, 83.2% OPA , and 88.4% SMA among the proposed dataset. The proposed approach outperformed the other baseline models with slightly better results. In the subclasses, the channel R-CNN reaches relatively good performance in PM, RP, and NT classification. Meanwhile, it only achieves a value of 73.5% for blood vessels. This is due to the fickle appearance of vessels, which adds to the model detection performance. For the comparison of processing efficiency, the channel R-CNN achieves an efficiency of 23.3 ms on one frame, which outperforms the other models with the leading edge of 1.1 ms, 3.9 ms, 3.2 ms, and 5.8 ms. We credit the segmentation improvement to the following two differences between our model and the baseline models: First, a two-stage structure was applied in the left channel; the RPN module (1st stage) transfers the focusing region information to the DNH module (2st stage) using the “attention” mechanisms. Second, a larger neural network, DenseNet, was adopted for the segmentation avoiding failing into the overfitting dilemma with a mass of parameters. As for the efficiency advantage, the channel R-CNN outperforms others due to its ability to integrate region proposal networks (RPNs) and achieve instance segmentation within a single framework. This integration allows for shared computation between the two tasks, leading to improved speed and performance. Additionally, the use of features shared across tasks helps reduce redundant computations.

Segmentation outcomes of the thoracic tissue contour and surgical instruments are as given in Figure 8. Clear and complete borders of different regions are produced by the DNH channel.

As given in Figure 9, we compare the segmentation performance of FCN-8s, U-Net, and the channel R-CNN for specific thoracic tissue categories. Examining the results presented in Figure 8a–c, it is concluded that the proposed system outperforms the existing approaches with the increasing rates of metrics varying from 1% to 11%. Feature maps from the DenseNet provide more explicit semantic characteristics for segmentation and propose obvious flags for boundary division. All the models gain better results for the pectoralis major (PM) and burnt spots (BS); this is due to their unique texture features and richer datasets. As for the “Blood vessel” (BV) and “Nervous tissue” (NT), there still exist much room for improvement; smaller morphological areas and chromaticity similar

to the background increase the difficulties of segmentation and detection. However, the segmentation accuracy of the proposed channel R-CNN is still close to that of clinicians, which provides a technical basis for further promotion.

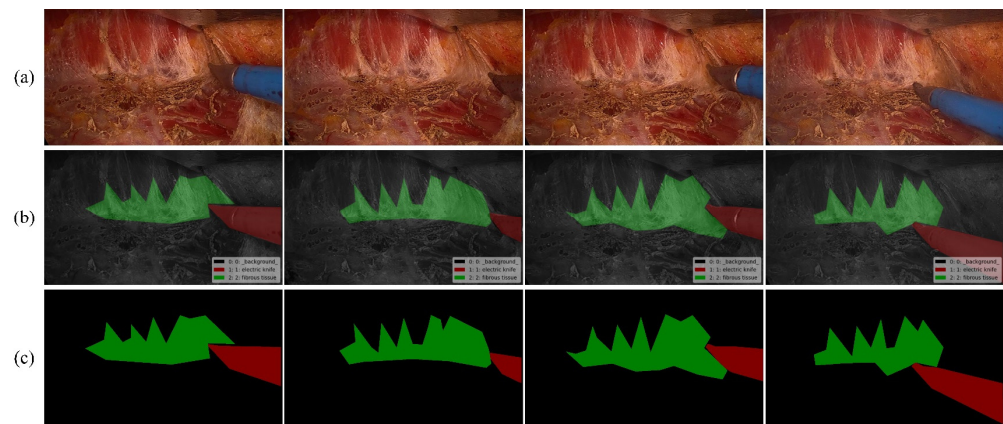


Figure 8. Visual segmentation results of the DNH. (a) Original image. (b) Labeled image. (c) Image mask.

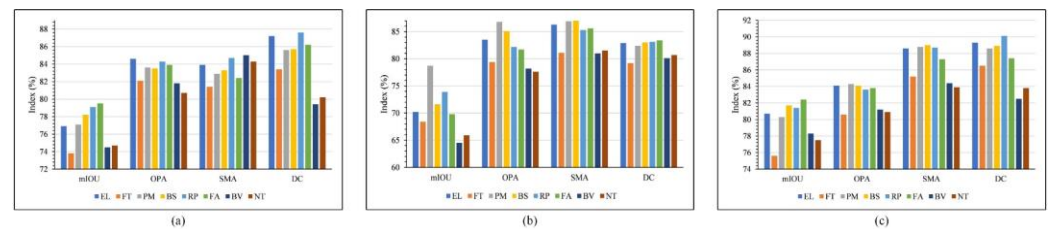


Figure 9. Performance comparison of thoracic tissue contour segmentation on specific categories. (a) FCN-8s. (b) U-Net. (c) Channel R-CNN.

3.2. Evaluations of Cutting Curve Segmentation

DenseNet [57] is applied as the baseline model for the cutting curve. Considering the predictable relationship between waves at different pixels, the l_1 and l_2 error of the cutting curve with the segmentation mask along with the peak signal-to-noise ratio (PSNR) are taken to evaluate the cutting indication performance. Ablation studies are conducted to further illustrate the improvement produced by channel 2# and the leapfrog algorithm, as given in Table 3. It is concluded that the proposed TNH module with the optimization process performs the best on all the metrics. The addition of the TNH and leapfrog algorithm increases the segmentation PSNR and reduces the mis-division error, and more target pixels are included in the mask, which is depicted by the higher segmentation metrics. The proposed channel R-CNN outperforms the Mask R-CNN method with improvements of 1.9%, 0.6%, 0.5%, and 1.1% for the *mIOU*, *OPA*, *SMA*, and *DC* metrics, respectively. For the U-Net model, its segmentation performance is close to the proposed network for the metrics of *OPA* and *SMA*, while its efficiency (24.7 ms) is relatively poor in comparison to all models, which may be due to the deep architecture and extensive skip connections. The numerous layers and complex connections in U-Net lead to a higher computational load during both training and inference. As for the comparison of efficiency, the baseline model DenseNet obtains the fastest processing time of 20.4 ms, the baseline model combining with the TNH and leapfrog algorithm exhibit a decrease in processing efficiency (0.8 ms and 1.4 ms, respectively), but considering the improvement in detection accuracy, the processing time of 21.8 ms is acceptable on the premise of meeting clinical real-time requirements. The mask segmentation performance of the proposed network structure is proved by the model performance comparison.

Table 3. l_1 and l_2 error and commonly applied evaluation metrics for the cutting curve segmentation using thoracic histological images. Time represents the processing time spent on one frame; the unit is ms. Bold results refer to the best performance.

	l_1	l_2	PSNR	mIOU	OPA	SMA	DC	Time
DenseNet [57]	2.1	0.2	35.62	61.2%	73.9%	72.7%	69.6%	20.4
+TNH	1.8	0.2	35.74	63.4%	76.3%	75.1%	73.7%	21.2
+Leapfrog	1.6	0.1	35.97	70.5%	78.1%	80.2%	81.3%	21.8
Mask R-CNN [15]	1.7	0.1	35.83	68.6%	77.5%	79.7%	80.2%	22.1
U-Net [55]	1.7	0.3	34.48	67.5%	77.4%	78.2%	80.5%	24.7

3.3. Robustness Analysis

In order to evaluate the robustness of the proposed DNH module for thoracoscopic surgery region segmentation, we conducted comparison experiments using clinical surgery videos with a resolution of 1280×720 at 24 fps. Surgical images under different complex backgrounds were input into the DNH module, and the region segmentation results of three test videos captured in one surgery are as shown in Figure 10, which effectively handles various states of surgical instrument occlusion. There rarely exists an abrupt direction change in the thoracoscopy angle during surgery; therefore, the segmentation of target regions is conducted based on the frames ordered in a small time period. Electrosurgery focus is indicated by the segmented mask, by which the chief surgeons obtain clearer cutting views. The proposed DNH can achieve robust segmentation of critical areas in clinical video sequences.

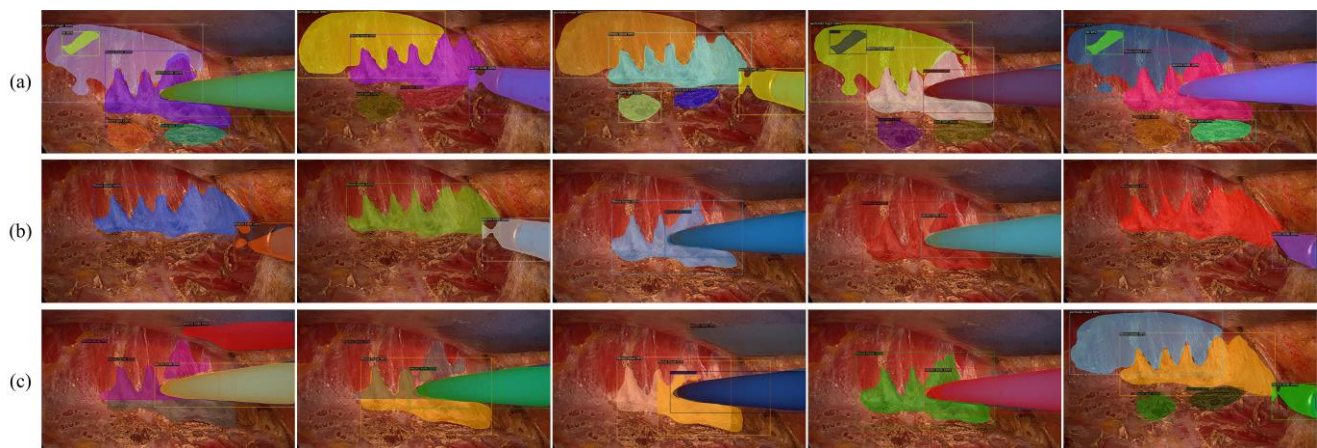


Figure 10. Target region segmentation by the DNH in different clinical videos. (a–c) Three individual test videos from one surgery.

As given in Figure 11, four main states with frequent occurrence including appliance occlusion (AO), tissue occlusion (TO), partial occlusion (PO), and complete visible (CV) are presented. It is concluded from Figure 11 that the proposed channel R-CNN effectively produces clear cutting curves in complex situations, which reduces the interferences from background pixels.

To further analyze the cutting indication performance of the proposed method for thoracoscopic surgery, visual segmentation on extra real surgery videos from cooperative units is carried out, using images with a resolution of 1280×720 at 24 fps from the cutting curve dataset. Table 4 illustrates the total number of frames and the percentage of frames with different states for each video sequence on cutting curve segmentation. The table outlines the channel R-CNN model's segmentation efficiency across eight video segments, showcasing the distinct percentages of events like AO, TO, PO, and CV. Variations in event

distribution are evident, with average AO event percentages spanning from approximately 16.54% in v8 to 35.09% in v1. The TO event percentages fluctuate notably, ranging from 18.17% in v4 to 48.48% in v8. Meanwhile, PO event percentages average around 16.88%, with v3 exhibiting the highest at 19.49% and v8 the lowest at 9.27%. CV event percentages show a range from 5.47% in v2 to 32.13% in v4. The average number of frames across segments is 669, with v5 recording the highest at 1320 frames and v8 the lowest at 284 frames. The average processing for each frame remains stable at 41.6 ms. These quantitative metrics highlight the diverse performance characteristics of the model across different video segments, indicating the influence of event composition, frame count, and duration on segmentation efficiency, which also indirectly illustrates the stronger segmentation robustness of the proposed model.

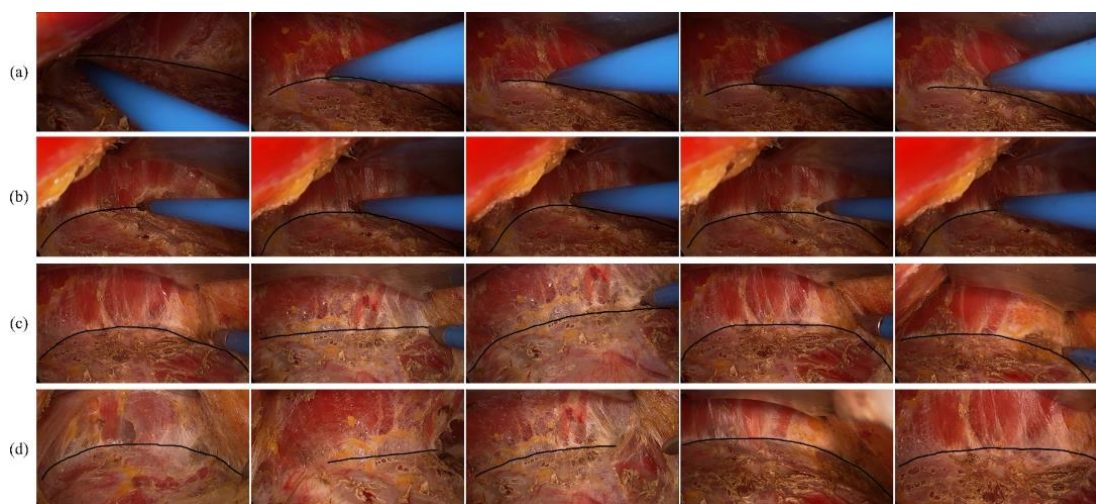


Figure 11. Visual cutting curves of the TNH under different conditions from a real surgery video. (a) Appliance occlusion (AO). (b) Tissue occlusion (TO). (c) Partial occlusion (PO). (d) Complete visible (CV).

Table 4. Descriptions of the duration, frame count, and the proportion of frames showcasing primary states within the cutting curve dataset for video sequences.

	Events [%]				Frames	Time [s]	t/Frame (ms)
	AO	TO	PO	CV			
v1	35.09	23.11	13.28	28.52	634	26.41	41.6
v2	25.04	42.93	26.56	5.47	515	21.45	41.7
v3	31.89	24.23	19.49	24.39	798	33.25	41.7
v4	32.78	18.17	16.92	32.13	606	25.25	41.7
v5	19.09	39.45	21.73	19.73	1320	55.00	41.7
v6	26.37	42.76	14.61	16.26	493	20.54	41.6
v7	22.29	38.06	15.57	24.08	508	21.16	41.7
v8	16.54	48.48	9.27	25.71	284	11.83	41.7

Table 5 shows the segmentation performance including the *mIOU*, *OPA*, *SMA*, *DC*, and time cost per frame (ms) for each video sequence to evaluate the efficiency and robustness of the proposed model. The average speed of our system is 23.6 fps in videos. It is concluded from Table 5 that the channel R-CNN achieves stable and consistent distribution for segmentation in each scoring index, the results of video sequence v3 and v4 on *SMA* and *DC* reach over 90%, reaching 91.2% and 90.9%, respectively. Therefore, the generalization ability of our system is proved by performance evaluation using the video sequences, by

which the thoracoscopic surgery is indicated using the clinical information from target regions and cutting curves.

Table 5. Segmentation performance and time costs of one-frame processing for video sequences (CON: thoracic tissue contour segmentation; CUR: cutting curve segmentation). Bold results refer to the best performance.

		v1	v2	v3	v4	v5	v6	v7	v8
Time (ms)	CON	27	26	24	24	23	26	24	24
	CUR	23	22	23	21	22	23	22	23
mIOU	CON	78.3%	82.5%	79.6%	76.1%	73.8%	80.7%	83.4%	80.9%
	CUR	71.8%	70.3%	68.6%	72.1%	73.4%	72.2%	70.6%	69.2%
OPA	CON	77.4%	83.7%	84.5%	83.2%	81.1%	82.6%	83.8%	81.0%
	CUR	78.5%	77.0%	76.1%	78.4%	80.6%	77.3%	77.5%	79.1%
SMA	CON	90.1%	86.8%	85.7%	91.2%	88.5%	86.2%	89.5%	88.7%
	CUR	78.2%	79.8%	80.7%	78.4%	78.6%	79.0%	78.3%	80.6%
DC	CON	86.9%	89.2%	90.9%	84.6%	88.1%	89.3%	90.8%	88.8%
	CUR	80.4%	76.5%	80.8%	79.1%	81.6%	80.7%	81.2%	80.4%

3.4. User Study

Five chest doctors were invited to experience and evaluate our indication system. After the test, all the experimenters were asked to complete the questionnaire as the raw material for the system evaluation, in which there were five key questions: (1) the accuracy of thoracic tissue contour segmentation; (2) the accuracy of cutting curve indication; (3) system smoothness; (4) the cutting guidance under obstruction of appliances and tissues; (5) the utility of online guiding. All the questions are scored between 0 and 10, where the value is positively correlated with satisfaction degree. Table 6 records the average scores for the five questions, and the limits to be improved are also illustrated. As depicted in Table 6, the accuracy of the thoracic tissue contour segmentation and the correctness of the cutting curve indication have received better feedback (8.5 higher on average). The cutting guidance under view occlusion is also acceptable. For the system smoothness, corresponding scores are rather general. This is due to the current system's computational efficiency; clinical surgery indication requires timely responses, and changes in patient states caused by the surgical procedure could range between a few seconds. Therefore, we need to concentrate more on real-time performance, by optimizing the algorithm efficiency with TensorRT or reducing network parameters. Moreover, it is also a feasible choice to process surgery data parallelly using multi-thread processing.

Table 6. Survey results by questionnaire.

Questions	D1	D2	D3	D4	D5	Mean	Standard Deviation
1	8.5	9.8	9.4	9.9	8.4	9.2	0.64
2	8.3	8.9	8.4	8.6	8.8	8.6	0.23
3	8.1	7.7	7.5	7.9	7.3	7.7	0.29
4	8.0	8.5	8.9	8.5	8.0	8.4	0.37
5	8.5	7.5	8.5	7.9	8.6	8.2	0.43

3.5. Evaluation of Time and Computation Complexity

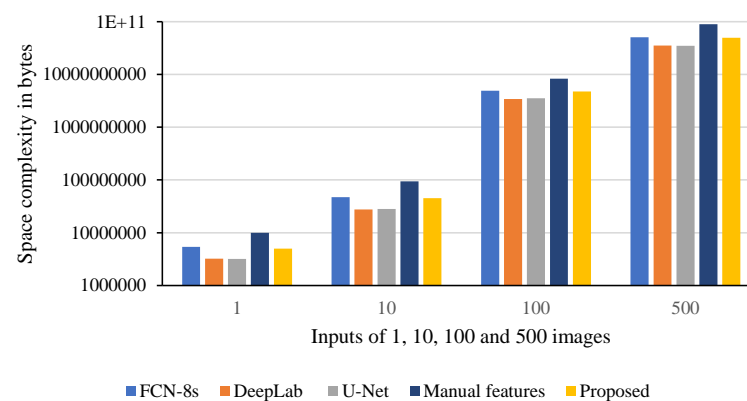
The hardware of the channel R-CNN is given in the section "Implementation and Transfer learning", and a comparison of the processing efficiency between the proposed system and the other similar methods is depicted in Table 7.

Table 7. Comparison of processing efficiency. Bold results refer to the best performance.

Method	Size	Processing Time (ms/Image)
FCN-8s [54]	1280 × 720	54
DeepLab [25]	1280 × 720	46
U-Net [55]	1340 × 1030	48
Manual features [56]	1280 × 720	62
Proposed	1280 × 720	53

As given in Table 7, the processing time for one image is listed. DeepLab [25] achieves the highest efficiency among all the models of 46 ms/image; atrous convolution and probabilistic graphical models contribute to the improvement. U-Net [55] reaches a similar running time as DeepLab of 48 ms/image, which is superior to other existing methods. Due to the proposed loss function and multi-channel network structure, the channel R-CNN achieves a lower running efficiency of 53 ms/image compared to U-Net, while outperforming the FCN-8s [54] and manual features [56] methods with an improvement of 1 ms and 9 ms, respectively. Considering the segmentation performance of target regions and cutting curves, the accuracy of the proposed system compensates for the response efficiency. However, in view of the real-time and safety requirements of clinical applications, frame rates above 30/s are common, and the efficiency of the proposed model needs to be further optimized.

Figure 12 shows the computation complexity in bytes for FCN-8s, DeepLab, U-Net, manual features, and the proposed channel R-CNN. It is concluded that DeepLab and U-Net achieve relatively lower complexity when facing different scales of input data. The proposed system and FCN-8s gain similar space complexity, and the channel R-CNN runs faster than FCN-8s, whose segmentation performance is also higher. Although the efficiency and computation complexity of the proposed model is slightly worse than DeepLab and U-Net, its segmentation results are obviously better. The channel R-CNN has taken the segmentation accuracy and efficiency into account concurrently, which has the potential for clinical application in thoracoscopic surgery.

**Figure 12.** Comparison of computation complexity on system inputs of 1, 10, 100, and 500 images from our proposed dataset. Values are generated on a virtual machine and Python 3.8.6, which will be only applied for comparison experiments.

To further evaluate the practical utility of the proposed system, we sought the expertise of chest specialists to provide insights on the processing time in the context of real-world applications. The feedback from these medical professionals offers a valuable perspective on the system's readiness for use in clinical settings. According to Dr. Zichen Wang from Qilu Hospital of Shandong University, the processing time of the proposed system appears to align well with practical requirements for timely diagnosis and decision-making in

chest imaging tasks. Dr. Lin Wei from CNPC Central Hospital noted that the system's performance is promising and could significantly enhance workflow efficiency in the clinical environment. Their expert assessment adds depth to our evaluation of the system's processing time and reinforces its potential application in real-world scenarios.

4. Conclusions

A novel channel R-CNN network for thoracoscopic surgery navigation to assist surgeons in cutting manipulation is proposed. The excised contours of target regions are tracked through semantic segmentation by a Detection Network Head (DNH) in channel 1#. Then, the cutting curve indication is conducted using the Thorax Network Head (TNH) in channel 2#, which involves rough and refined cutting operations. Results from the indicators exhibit an *mIOU* of 79.4% and 70.5% and *OPA* of 83.2% and 78.1% for the target contour and cutting curve segmentation, respectively. The contribution of the framework modules to the final results is verified by an ablation study. The comparative experimental results illustrate that our framework could achieve accurate and robust navigation under complex surgery situations in a clinical environment. It is feasible to implement semantic segmentation and navigation for thoracoscopic surgery using the RNN network.

Our proposed technique, while promising, has room for enhancement, particularly in clinical applications such as segmentation accuracy and real-time responsiveness. In future studies, we aim to optimize the system's efficiency by reducing network configuration parameters and implementing an inference acceleration refiner. This will be crucial for meeting the stringent demands of real-time surgical navigation. To further improve segmentation accuracy, we plan to refine the loss function to minimize surgical risks, especially in critical tissue areas. This enhancement is vital for ensuring patient safety and surgical precision. Considering the framework structure, we will explore the application of more natural scenes to adapt the real-time network, maintaining accuracy while significantly boosting efficiency. This adaptation is essential for the model's practical deployment in dynamic surgical environments.

Additionally, to meet real-time surgical navigation requirements, this future study will further optimize the model architecture to improve real-time performance and response speed; explore the use of specialized hardware (such as GPU, TPU, etc.) for acceleration to improve the computational efficiency; introduce online learning mechanisms to enable models that learn and adjust in real-time from new data to adapt to constantly changing surgical environments; and design end-to-end systems. On the other hand, the proposed model is also subjected to potential limitations: under real-time requirements, the model may require more computing resources and storage space, which may limit its deployment in resource-constrained environments; real time surgical navigation involves sensitive medical data; therefore, strict data privacy protection and security measures are required; the model needs to have good robustness in real-time surgical environments and be able to handle data noise, uncertainty, and unexpected situations, which may be challenging and a key direction for our future study. By expanding the application of the proposed model and addressing current limitations, the channel R-CNN could become an invaluable tool in enhancing surgical outcomes and training interns.

Author Contributions: Conceptualization, C.Z. and D.J.; Data curation, Y.C. and B.Z.; Methodology, Y.C.; Resources, D.J.; Software, C.Z.; Supervision, D.J.; Validation, B.Z.; Writing—original draft, C.Z.; Writing—review and editing, C.Z. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding; the APC was funded by China Nuclear Power Engineering Co., Ltd.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Research Center for Big Data and Intelligent Measurement and Control from Beijing Jiaotong University (protocol code EA2024090101 and date of approval 2024.9.1).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: For any inquiries, requests, or intentions for further collaboration regarding this study or the dataset, please contact the corresponding author.

Acknowledgments: We thank Zhigang Guo of Weichai Power (Weifang) Co., Ltd., Hairui Ge of Hollysys (Xian) Co., Ltd., and Xingyi Du of Peking Union Medical College Hospital for comments and help on data processing.

Conflicts of Interest: The authors declare that they work in China Nuclear Power Engineering Co., Ltd. The funders participated in the design of the study, as well as in the analyses and interpretation of data, the writing of this article or the decision to submit it for publication. Authors Chuanwang Zhang, Yueyuan Chen and Bo Zhang were employed by the company China Nuclear Power Engineering Co., Ltd. The remaining authors declare that the re-search was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- DeSantis, C.E.; Ma, J.; Gaudet, M.M.; Newman, L.A.; Miller, K.D.; Goding Sauer, A.; Jemal, A.; Siegel, R.L. Breast cancer statistics, 2019. *CA A Cancer J. Clin.* **2019**, *69*, 438–451. [[CrossRef](#)] [[PubMed](#)]
- Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **2021**, *149*, 778–789. [[CrossRef](#)] [[PubMed](#)]
- Kunkler, I.H.; Williams, L.J.; Jack, W.J.; Cameron, D.A.; Dixon, J.M. Breast-conserving surgery with or without irradiation in early breast cancer. *N. Engl. J. Med.* **2023**, *388*, 585–594. [[CrossRef](#)]
- Aujayeb, A.; Jackson, K. A review of the outcomes of rigid medical thoracoscopy in a large UK district general hospital. *Pleura Peritoneum* **2020**, *5*, 20200131. [[CrossRef](#)] [[PubMed](#)]
- Giannakeas, V.; Lim, D.W.; Narod, S.A. Bilateral mastectomy and breast cancer mortality. *JAMA Oncol.* **2024**, *10*, 1228–1236. [[CrossRef](#)]
- Safran, T.; Vorstenbosch, J.; Viezel-Mathieu, A.; Davison, P.; Dionisopoulos, T.J.P.; Surgery, R. Topical tranexamic acid in breast reconstruction: A double-blind randomized controlled trial. *Plast. Reconstr. Surg.* **2023**, *152*, 699–706. [[CrossRef](#)] [[PubMed](#)]
- Haidari, T.A.; Nayahangan, L.J.; Bjerrum, F.; Hansen, H.J.; Konge, L.; Massard, G.; Batirel, H.F.; Novoa, N.M.; Milton, R.S.; Petersen, R.H.; et al. Consensus on technical procedures for simulation-based training in thoracic surgery: An international needs assessment. *Eur. J. Cardio-Thoracic Surg.* **2023**, *63*, ezad058. [[CrossRef](#)] [[PubMed](#)]
- Mustajab, M.; Shuja, M.I. Evaluation of Factors in Women’s Decisions Regarding Breast Surgery. *Evaluation* **2023**, *8*, 11.
- Fan, S.; Sáenz-Ravello, G.; Diaz, L.; Wu, Y.; Davó, R.; Wang, F.; Magic, M.; Al-Nawas, B.; Kämmerer, P.W. The accuracy of zygomatic implant placement assisted by dynamic computer-aided surgery: A systematic review and meta-analysis. *J. Clin. Med.* **2023**, *12*, 5418. [[CrossRef](#)]
- Westbrook, K.; Rollor, C.; Aldahmash, S.A.; Fay, G.G.; Rivera, E.; Price, J.B.; Griffin, I.; Tordik, P.A.; Martinho, F.C. Comparison of a novel static computer-aided surgical and freehand techniques for osteotomy and root-end resection. *J. Endod.* **2023**, *49*, 528–535.e521. [[CrossRef](#)]
- Bhuskute, K.P.; Jadhav, V.; Sharma, M.; Reche, A. Three-dimensional Computer-aided Design System used in Orthodontics and Orthognathic Surgery for Diagnosis and Treatment Planning-A Narrative Review. *J. Clin. Diagn. Res.* **2023**, *17*, ZE7–ZE10. [[CrossRef](#)]
- Kuo, S.-W.; Tseng, Y.-F.; Dai, K.-Y.; Chang, Y.-C.; Chen, K.-C.; Lee, J.-M. Electromagnetic navigation bronchoscopy localization versus percutaneous CT-guided localization for lung resection via video-assisted thoracoscopic surgery: A propensity-matched study. *J. Clin. Med.* **2019**, *8*, 379. [[CrossRef](#)]
- Chen, J.; Pan, X.; Gu, C.; Zheng, X.; Yuan, H.; Yang, J.; Sun, J. The feasibility of navigation bronchoscopy-guided pulmonary microcoil localization of small pulmonary nodules prior to thoracoscopic surgery. *Transl. Lung Cancer Res.* **2020**, *9*, 2380. [[CrossRef](#)] [[PubMed](#)]
- Lee, C.Y.; Chan, H.; Ujiie, H.; Fujino, K.; Kinoshita, T.; Irish, J.C.; Yasufuku, K. Novel thoracoscopic navigation system with augmented real-time image guidance for chest wall tumors. *Ann. Thorac. Surg.* **2018**, *106*, 1468–1475. [[CrossRef](#)]

15. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
16. Jia, D.; Zou, S.; Li, M.; Zhu, H. Adaptive multi-path routing based on an improved leapfrog algorithm. *Inf. Sci.* **2016**, *367*, 615–629.
17. Ni, Z.-L.; Bian, G.-B.; Li, Z.; Zhou, X.-H.; Li, R.-Q.; Hou, Z.-G. Space squeeze reasoning and low-rank bilinear feature fusion for surgical image segmentation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 3209–3217. [[CrossRef](#)]
18. Xu, J.; Zeng, B.; Egger, J.; Wang, C.; Smedby, Ö.; Jiang, X.; Chen, X. A review on AI-based medical image computing in head and neck surgery. *Phys. Med. Biol.* **2022**, *67*, 17TR01. [[CrossRef](#)] [[PubMed](#)]
19. Tsai, A.Y.; Carter, S.R.; Greene, A.C. Artificial intelligence in pediatric surgery. In *Seminars in Pediatric Surgery*; Elsevier: Amsterdam, The Netherlands, 2024; p. 151390.
20. Yue, W.; Zhang, J.; Hu, K.; Xia, Y.; Luo, J.; Wang, Z. Surgicsam: Efficient class promptable surgical instrument segmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 26–27 February 2024; pp. 6890–6898.
21. Yellu, R.R.; Kukalakunta, Y.; Thunki, P. Medical Image Analysis-Challenges and Innovations: Studying challenges and innovations in medical image analysis for applications such as diagnosis, treatment planning, and image-guided surgery. *J. Artif. Intell. Res. Appl.* **2024**, *4*, 93–100.
22. Liu, S.; Yue, W.; Guo, Z.; Wang, L. Multi-branch CNN and grouping cascade attention for medical image classification. *Sci. Rep.* **2024**, *14*, 15013. [[CrossRef](#)] [[PubMed](#)]
23. Alam, M.; Zhao, E.J.; Lam, C.K.; Rubin, D.L. Segmentation-assisted fully convolutional neural network enhances deep learning performance to identify proliferative diabetic retinopathy. *J. Clin. Med.* **2023**, *12*, 385. [[CrossRef](#)] [[PubMed](#)]
24. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.-W.; Heng, P.-A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [[CrossRef](#)]
25. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)]
26. Huang, S.-T.; Chu, Y.-C.; Liu, L.-R.; Yao, W.-T.; Chen, Y.-F.; Yu, C.-M.; Yu, C.-M.; Tung, K.-Y.; Chiu, H.-W.; Tsai, M.-F.; et al. Deep learning-based clinical wound image analysis using a mask R-CNN architecture. *J. Med. Biol. Eng.* **2023**, *43*, 417–426. [[CrossRef](#)]
27. Xia, G.; Ran, T.; Wu, H.; Wang, M.; Pan, J. The development of mask R-CNN to detect osteosarcoma and oste-ochondroma in X-ray radiographs. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2023**, *11*, 1869–1875. [[CrossRef](#)]
28. Salh, C.H.; Ali, A.M. Automatic detection of breast cancer for mastectomy based on MRI images using Mask R-CNN and Detectron2 models. *Neural Comput. Appl.* **2024**, *36*, 3017–3035. [[CrossRef](#)]
29. Li, X.; Young, A.S.; Raman, S.S.; Lu, D.S.; Lee, Y.-H.; Tsao, T.-C.; Wu, H.H. Automatic needle tracking using Mask R-CNN for MRI-guided percutaneous interventions. *Int. J. Comput. Assist. Radiol. Surg.* **2020**, *15*, 1673–1684. [[CrossRef](#)] [[PubMed](#)]
30. Brower, V. The cutting edge in surgery. *EMBO Rep.* **2002**, *3*, 300–301. [[CrossRef](#)]
31. Chrysovergis, I.; Kamarianakis, M.; Kentros, M.; Angelis, D.; Protopsaltis, A.; Papagiannakis, G. Assessing unconstrained surgical cuttings in VR using CNNs. *arXiv* **2022**, arXiv:2205.00934.
32. Rianto, S.; Li, L. Fluid dynamic visualisations of cuttings-bleeding for virtual reality heart beating surgery simulation. In Proceedings of the Thirty-Third Australasian Conferenc on Computer Science, Brisbane, Australia, 18–22 January 2010; Volume 102, pp. 53–60.
33. Jin, X.; Joldes, G.R.; Miller, K.; Yang, K.H.; Wittek, A. Meshless algorithm for soft tissue cutting in surgical simulation. *Comput. Methods Biomech. Biomed. Eng.* **2014**, *17*, 800–811. [[CrossRef](#)]
34. Tang, H.; Chen, Y.; Wang, T.; Zhou, Y.; Zhao, L.; Gao, Q.; Du, M.; Tan, T.; Zhang, X.; Tong, T.J.; et al. HTC-Net: A hybrid CNN-transformer framework for medical image segmentation. *Biomed. Signal Process. Control.* **2024**, *88*, 105605. [[CrossRef](#)]
35. Cartiaux, O.; Banse, X.; Paul, L.; Francq, B.G.; Aubin, C.-É.; Docquier, P.-L. Computer-assisted planning and navigation improves cutting accuracy during simulated bone tumor surgery of the pelvis. *Comput. Aided Surg.* **2013**, *18*, 19–26. [[CrossRef](#)] [[PubMed](#)]
36. Lin, Q.; Xiongbo, G.; Zhang, W.; Cai, L.; Yang, R.; Chen, H.; Cai, K. A novel approach of surface texture mapping for cone-beam computed tomography in image-guided surgical navigation. *IEEE J. Biomed. Health Inform.* **2023**, *28*, 8. [[CrossRef](#)] [[PubMed](#)]
37. Wilson, J.P., Jr.; Fontenot, L.; Stewart, C.; Kumbhare, D.; Guthikonda, B.; Hoang, S. Image-Guided Navigation in Spine Surgery: From Historical Developments to Future Perspectives. *J. Clin. Med.* **2024**, *13*, 2036. [[CrossRef](#)]
38. Zeng, J.; Fu, Q. A review: Artificial intelligence in image-guided spinal surgery. *Expert Rev. Med. Devices* **2024**, *21*, 689–700. [[CrossRef](#)] [[PubMed](#)]
39. Hanna, G.; Kim, T.T.; Uddin, S.-A.; Ross, L.; Johnson, J.P. Video-assisted thoracoscopic image-guided spine surgery: Evolution of 19 years of experience, from endoscopy to fully integrated 3D navigation. *Neurosurg. Focus* **2021**, *50*, E8. [[CrossRef](#)]
40. Ujiiie, H.; Yamaguchi, A.; Gregor, A.; Chan, H.; Kato, T.; Hida, Y.; Kaga, K.; Wakasa, S.; Eitel, C.; Clapp, T.R. Developing a virtual reality simulation system for preoperative planning of thoracoscopic thoracic surgery. *J. Thorac. Dis.* **2021**, *13*, 778. [[CrossRef](#)] [[PubMed](#)]

41. Nakamura, S.; Hayashi, Y.; Kawaguchi, K.; Fukui, T.; Hakiri, S.; Ozeki, N.; Mori, S.; Goto, M.; Mori, K.; Yokoi, K. Clinical application of a surgical navigation system based on virtual thoracoscopy for lung cancer patients: Real time visualization of area of lung cancer before induction therapy and optimal resection line for obtaining a safe surgical margin during surgery. *J. Thorac. Dis.* **2020**, *12*, 672. [[PubMed](#)]
42. Lopez-Lopez, V.; Sánchez-Esquer, I.; Crespo, M.J.; Navarro, M.Á.; Brusadin, R.; Conesa, A.L.; Barrios, A.N.; Miura, K.; Robles-Campos, R. Development and validation of advanced three-dimensional navigation device integrated in da Vinci Xi[®] surgical robot for hepatobiliary surgery: Pilot study. *Br. J. Surg.* **2023**, *110*, 108–110. [[CrossRef](#)] [[PubMed](#)]
43. Tzelnick, S.; Rampinelli, V.; Sahovaler, A.; Franz, L.; Chan, H.H.; Daly, M.J.; Irish, J.C. Skull-base surgery—A narrative review on current approaches and future developments in surgical navigation. *J. Clin. Med.* **2023**, *12*, 2706. [[CrossRef](#)] [[PubMed](#)]
44. Huang, W.; Wang, K.; Chen, F.; Li, G.; Chen, X.; Yang, Q.; Li, N.; He, K.; Chen, F.; Tian, J. Intraoperative Fluorescence Visualization in Thoracoscopic Surgery. *Ann. Thorac. Surg.* **2022**, *115*, e79–e81. [[CrossRef](#)]
45. Yoshida, R.; Yoshizako, T.; Tanaka, S.; Ando, S.; Nakamura, M.; Kishimoto, K.; Kitagaki, H. CT-guided color marking of impalpable pulmonary nodules prior to video-assisted thoracoscopic surgery. *Clin. Imaging* **2021**, *74*, 84–88. [[CrossRef](#)] [[PubMed](#)]
46. Liu, Z.; Zheng, L.; Gu, L.; Yang, S.; Zhong, Z.; Zhang, G. InstrumentNet: An integrated model for real-time segmentation of intracranial surgical instruments. *Comput. Biol. Med.* **2023**, *166*, 107565. [[CrossRef](#)] [[PubMed](#)]
47. Sato, Y.; Sugimoto, M.; Tanaka, Y.; Suetsugu, T.; Imai, T.; Hatanaka, Y.; Matsuhashi, N.; Takahashi, T.; Yamaguchi, K.; Yoshida, K. Holographic image-guided thoracoscopic surgery: Possibility of usefulness for esophageal cancer patients with abnormal artery. *Esophagus* **2020**, *17*, 508–511. [[CrossRef](#)] [[PubMed](#)]
48. Wang, F.; Su, J. Based on the improved YOLOV3 small target detection algorithm. In Proceedings of 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 18–20 June 2021; pp. 2155–2159.
49. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)]
50. Salehi, A.W.; Khan, S.; Gupta, G.; Alabdullah, B.I.; Almjally, A.; Alsolai, H.; Siddiqui, T.; Mellit, A. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability* **2023**, *15*, 5930. [[CrossRef](#)]
51. Menghani, G. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *ACM Comput. Surv.* **2023**, *55*, 1–37. [[CrossRef](#)]
52. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich Switzerland, 6–12 September 2014; pp. 740–755.
53. Liao, T.; Li, L.; Ouyang, R.; Lin, X.; Lai, X.; Cheng, G.; Ma, J. Classification of asymmetry in mammography via the DenseNet convolutional neural network. *Eur. J. Radiol. Open* **2023**, *11*, 100502. [[CrossRef](#)]
54. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
55. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
56. Gertych, A.; Ing, N.; Ma, Z.; Fuchs, T.J.; Salman, S.; Mohanty, S.; Bhele, S.; Velásquez-Vacca, A.; Amin, M.B.; Knudsen, B.S. Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Comput. Med. Imaging Graph.* **2015**, *46*, 197–208. [[CrossRef](#)] [[PubMed](#)]
57. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.