Artificial Intelligence In Medicine (2023)

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Multi-Task Learning Framework to Predict the Status of Central Venous Catheter Based on Radiographs

Yuhan Wang[a], Hak Keung Lam[a,*], Yujia Xu[a], Faliang Yin[a], Kun Qian[b]

[a]*Department of Engineering, King's College London, Strand, London, WC2R 2LS, United Kingdom*
[b]*Center for the Developing Brain, School of Biomedical Engineering and Imaging Sciences, King's College London, St Thomas' Campus, St Thomas' Hospital, Westminster Bridge Road, London, SE1 7EH, United Kingdom*

ARTICLE INFO

Keywords:
X-ray Image
Catheter Status Recognition
Central Venous Catheter (CVC)
Convolutional Neural Network (CNN)
Multi-Task Learning

ABSTRACT

Hospital patients can have catheters and lines inserted during the course of their admission to give medicines for the treatment of medical issues, especially the central venous catheter (CVC). However, malposition of CVC will lead to many complications, even death. Clinicians always detect the status of the catheter to avoid the above issues via X-ray images. To reduce the workload of clinicians and improve the efficiency of CVC status detection, a multi-task learning framework for catheter status classification based on the convolutional neural network (CNN) is proposed. The proposed framework contains three significant components which are modified HRNet, multi-task supervision including segmentation supervision and heatmap regression supervision as well as classification branch. The modified HRNet maintaining high-resolution features from the start to the end can ensure to generation of high-quality assisted information for classification. The multi-task supervision can assist in alleviating the presence of other line-like structures such as other tubes and anatomical structures shown in the X-ray image. Furthermore, during the inference, this module is also considered as an interpretation interface to show where the framework pays attention to. Eventually, the classification branch is proposed to predict the class of the status of the catheter. A public CVC dataset is utilized to evaluate the performance of the proposed method, which gains 0.823 AUC (Area under the ROC curve) and 82.6 % accuracy in the test dataset. Compared with two state-of-the-art methods (ATCM method and EDMC method), the proposed method can perform best.

1. Introduction

During the course of admission, hospital patients require different kinds of catheters including central venous catheter (CVC) to give medications and do medical tests quickly [29]. CVC which is regarded as one of the most essential indwelling devices is peripherally inserted into a large central vein and is terminally at the inferior vena cava, superior vena cava, or right atrium. After the insertion procedure, the position of CVC is commonly verified based on radiographs [30] to avoid malposition. More specifically, landmark determination of CVC terminal tip location is regarded as the general way to detect the malposition of the CVC based on X-ray images belonging to radiographs [18]. However, if malposition of the CVC cannot be immediately detected, it will lead to many complications even death [20]. In previous research [1], machine learning methods were proposed to avoid complications for ICU patients, showing good performance. Therefore, inspired by human prediction, an automatic multi-task learning framework

*Corresponding author: hak-keung.lam@kcl.ac.uk

including catheter tip detection, catheter course segmentation, and catheter status classification is proposed to assist clinicians in detecting the status of the catheter effectively to avoid complications for patients.

Similar to the challenges faced by [25] [12], automatic classification of the status of CVC based on X-ray images meets four main challenges: 1) As shown in Fig. 1, the low-contrast X-ray images are in poor quality, making it difficult to distinguish the CVC from the background. 2) There are different kinds of tubes and anatomical structures that have similar appearances with CVC simultaneously appearing in one X-ray image, making it difficult to distinguish the CVC from other structures. 3) The course of the CVC which is regarded as one of the most significant information for catheter status classification is thin in X-ray images, which means that the number of pixels containing the catheter is much less than the background, making it difficult to classify the status of the catheter. 4) Because of the impact of patient rotation or incomplete inspiration, the location of the tip which is the most significant information for catheter status classification [10] is controversial in some images, making it difficult to find the precise tip location. To deal with the above challenges, plenty of methods were proposed for catheter classification based on radiographs.

To deal with the first to third challenges mentioned above, early approaches of catheter status classification are mainly based on hand-engineered rules, significant image preprocessing, and feature extraction [23]. However, these methods based on human-created features have lower detection accuracy, even if these features utilized to detect and classify the catheters are explainable. Recently, deep learning frameworks that can automatically extract features have been widely implemented in the classification of the type of medical instruments as well as the classification of the status of the instruments. One-stage deep learning frameworks based on Inception V3 [25] automatically extracting features from X-ray images can directly give the status of the catheter, improving the classification performance. However, these end-to-end deep learning frameworks only can output the status of the catheter. However, the position of the whole body of the catheter as well as the tip of the catheter cannot be generated simultaneously, making the results untrustworthy. To make the framework more trustworthy, in [12], the multi-stage framework which can give the mask of the catheter and the tip of the catheter was proposed, based on a high-resolution recovery process to detect the whole body of the catheter as well as the tip of the catheter. However, during the whole process, high-resolution features cannot be maintained from the start to the end, leading to poor segmentation performance of the whole body of the catheter and the catheter tip. Then, the imprecise segmentation masks of the whole body of the catheter and the catheter tip make the catheter status misclassified by the classification network. Therefore, an automatic framework that can generate the precise mask of the whole body of the catheter, the location of the tip of the catheter as well as the status of the catheter simultaneously is required, to make the clinicians know the status of the catheter [33].

To evaluate the performance of the proposed framework, we use X-ray images from CLiP, catheter, and line position dataset published in [29]. In the CVC dataset, there are two kinds of statuses of CVC which are dangerous and safe. When the status of the catheter is classified as safe, it means that the CVC is appropriately positioned and does not require repositioning. When the status is classified as dangerous, it means that the catheter requires repositioning.

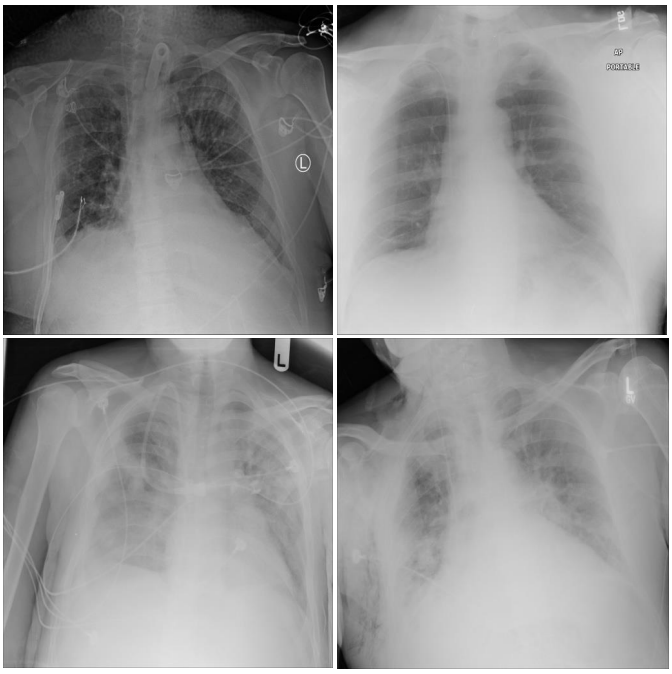


Fig. 1. Several X-ray images from the CVC Dataset [29].

The contribution of this work includes: 1) An automatic catheter status classification framework based on X-ray images is proposed in this paper to assist clinicians in detecting the status of the CVC to avoid complications for patients. 2) To deal with the third challenge, a high-resolution network structure [32] is adopted to maintain high-resolution features from start to end. 3) To tackle the second challenge and make the framework more explainable, the framework not only can output the status of the catheter but also can generate the mask of the whole body of the catheter as well as the tip position of the catheter to assist clinicians to check the place of the catheter, making the classification framework trustworthy. 4) We compared the performance of the proposed framework with two methods: ATCM Method [25] which is a one-stage framework and EDMC Method [12] which is a multi-stage framework. The proposed framework can get the highest AUC, which is 0.823.

2. Related Work

2.1. Computer-aided Assessment of Catheters on Radiographs

Traditional computer vision methods based on manual feature extraction were adopted to classify the type of the catheters and the status of the catheters [23] [22]. In [22], based on the hypotheses assuming two different kinds of catheters (NGTs and ETTs) have different intensity profiles, template matching was utilized to classify the type of catheter. However, this method based on the appearance features is limited, when

catheters have the same appearance and only differ in their course relative to anatomic landmarks. Therefore, in [23], geometric features are considered along the appearance features to classify the types of catheters. The regions containing relevant anatomical structures, such as the lung, mediastinum, and stomach, were identified and used to determine the relative positions of tubes in the image. These regions served as landmarks for tube detection and classification. For the above methods, even if these features utilized to detect and classify the catheters are explainable, these methods cannot achieve high detection and classification performance. In addition, to get the final classification result, the images should be fed into a framework comprising multiple stages, which is complex. To improve the performance, machine learning method such as SVM [11] was utilized as a classifier based on the human-created features. However, the performance is not high enough to meet the requirement and the framework also contains different stages to extract different features and give the final classification result.

In recent years, data-driven methods based on deep learning have been widely implemented in surgical scenarios [27] [24] [9], including surgical instrument segmentation [19], detection [16] and classification. Especially, convolutional neural network (CNN) has shown the leading technique for classification task [6] [28], including surgical instrument classification [25] [12] [26]. There are two main kinds of deep learning frameworks for catheter classification, which are the one-stage framework and the multi-stage framework respectively. The one-stage framework [25] can also be regarded as an end-to-end framework, which combines the feature extraction process and classifier into a unique network to give the final class. For multi-stage framework [12], the features such as the mask of the catheters or anatomical structures are extracted from the original image, then combined with the original image to feed into the classifier to get the final class.

In [25], the one-stage framework was proposed based on pre-trained deep convolutional neural networks (DCNNs) to classify the tube placement which is the non-critical insertion or critical insertion. Compared with different kinds of DCNNs such as ResNet-50 and DenseNet-121, Inception V3 was adopted to classify the tube placement. In [7], to classify multiple types of catheters on the neonatal chest and abdominal radiographs, a kind of convolutional neural network based on pretrained ResNet-50 was proposed. The performance achieved by the one-stage deep learning framework is much higher than that of the traditional method. In addition, the process of a one-stage framework is much simpler, which means that the feature extraction and classification are combined into a unique network. However, because the output of the framework is only the final class of the catheter without any intermediate output such as the course of the catheter or the place of the catheter tip, the features learned from the framework are unknown, making the classification result more untrustworthy.

To improve the classification performance and make the framework more explainable, which can explain which features of the image are utilized to predict the class, different kinds of multi-stage deep learning frameworks [12] [26] [5] were proposed. In [26], the multi-stage framework was proposed to detect and classify different types of catheters. Firstly, the course of the catheter and anatomical structures are segmented based on U-Net. Secondly, the features are extracted based on the course of the catheters as well as the anatomical structures, combined with high-precision custom features such as the size and shape of the catheters. Finally, these features are considered as the input for the classification model based on the neural network utilized to extract features and random forest used to give the final classification result. In [12] [5], two-stage frameworks were proposed to classify the placement of the catheter. In [5], the tip and course of the catheter as well as the interested region where the catheter tip might be segmented based on U-Net. Then, these segmentation results are combined to feed into the classification network (EfficientNet) to give the classification result. Similar to [5], in [12], firstly the two endpoints of the catheter and the course of the catheter are segmented based on U-Net. The 2-channel segmentation results are combined with the original image to feed into EfficientNet to classify whether the catheter is malpositioned or not. However, the performance of the classification highly depends on the performance of segmentation results. To classify the placement of CVC, based on the similar structure of [12], in [5], the anatomical structure (trachea) was segmented by the segmentation network (UNet) to further promote the classification performance. If the course of the catheter, tip of the catheter, interested regions or anatomical structures such as trachea are failed to be segmented, the frameworks will achieve low classification performance.

2.2. Multi-Task Learning

In the proposed framework, multi-task learning is adopted to tackle the second challenge illustrated in Section 1 and to reduce the risk of overfitting as well as to make the classification result more trustworthy. The idea of multi-task learning is first investigated in [4], showing that learning multiple tasks with a model can get a better result than learning them independently. There are two main categories of multi-task learning framework, which are hard-parameters sharing framework and soft-parameters sharing framework [31]. For the hard-parameters sharing framework, the hidden layers of the network are shared among all tasks, while keeping several task-specific output layers. In the soft parameters sharing framework, the model is defined for each task with its parameters. Then, the distance between the parameters of the model is regularized to encourage the parameters to be similar. The proposed method is inspired by the hard-parameters sharing framework.

In recent years, multi-task learning has been widely implemented in surgical instrument detection and tracking. In addition, most of the methods are based on a hard-parameters sharing framework, which means that there is a shared encoder with task-oriented decoders in the framework. In [16], multi-task learning was adopted to improve the performance of the keypoint detection of the surgical instruments with the task of surgical instrument segmentation. In [21], to reconstruct the 3D geometry of the surgical site and detect instruments within it, the tool segmentation, as well as the disparity estimation, are joined into a unique network. Additionally, in [8], task-oriented decoders are utilized to predict scanpath while tracking instruments based on the shared encoder. Therefore, in the proposed

framework, to get better performance of catheter status classification as well as make the classification result more convincing, the catheter segmentation task and catheter tip detection task are combined with the classification task into a unique network.

3. Methods

The main goal of our work is to classify whether the status of the CVC is safe or not. In order to achieve the main goal, the location of the catheter tip as well as the course of the catheter are considered as the auxiliary outputs of the framework to promote the classification performance. Therefore, the input of the framework is the X-ray image and the outputs of the framework combine three parts which are the position of the catheter tip, the course, and the class of the catheter, as shown in Fig. 2.

3.1. Network Architecture

In this part, the design of each component of the proposed catheter status classification network shown in Fig. 3 is presented. There are three main components which are modified HRNet, multi-task supervision and classification branch. In addition, the proposed classification network shown in Fig. 3 is regarded as a component of the whole framework shown in Fig. 2, starting from the X-ray image and ending with three outputs which are the heatmap of the catheter tip, segmentation mask of the catheter, and class of catheter in Fig. 3.

3.1.1. Modified HRNet

The classification network shown in Fig. 3 is proposed based on HRNet. Compared with the encoder-decoder structure such as U-Net, HRNet can maintain high-resolution features from the start to the end, avoiding the loss of significant information from high-resolution features, i.e. HRNet begins with the highest-resolution convolution stream, gradually combining with the lower-resolution convolution streams step by step in parallel [32].

The modified HRNet in the proposed classification network shown in Fig. 3 begins with the X-ray image represented as a matrix with the size of $512 \times 512 \times 3$. In the first stage, two stride-2 3×3 convolution operations are implemented to decrease the resolution of the features to 128×128 which is 1/4 of the input, and increase the number of the channel to 64. Then, the feature map with the size of $128 \times 128 \times 64$ is fed into 3 Bottleneck Residual units shown in Fig. 4, followed by two different streams. In the first stream, one stride-1 3×3 convolution operation is implemented to get features with the size of $128 \times 128 \times 32$ which is regarded as one part of the output for the 1st stage. Another stream includes one stride-2 convolution operation to increase the channel of the features to 64 and decrease the resolution of the features to 1/8 of the input which is regarded as another part of the output for the 1st stage. Then, these two outputs of the 1st stage are regarded as the input for the 2nd stage.

As demonstrated in Fig.3, the 2nd stage begins with two streams and ends up with three outputs. There are four stages in the modified HRNet. Consequently, the number of streams for the first, second, third, and fourth stages is 1, 2, 3, and 4 respectively. The number of channels of the output for the first, second, third, and fourth streams is 32, 64, 128, and 256 respectively. Additionally, the resolution of the feature map is 1/4, 1/8, 1/16, and 1/32 of the input features for the first, the second, the third, and the fourth stream. Except for the first stage, the inputs with different sizes of features are fed into 3 basic Residual units (each of them consists of two 3×3 convolutions followed by BatchNorm and ReLu) shown in Fig. 4 of different streams, in each stage. At the end of each stage, there is a feature fusion operation where the higher-solution features are downsampled based on the convolution operation to add the lower-resolution features generated by other streams. Moreover, the lower-resolution features are upsampled based on convolution operation to add with the higher-resolution features. Eventually, these features with high to low resolutions are regarded as the input for different streams of the next stage. As demonstrated in Fig. 3, there are 1, 4, and 3 modularized blocks in the 2nd, 3rd, and 4th stages, respectively. The modularized block consists of multi-resolution parallel convolutions and multi-resolution fusion.

3.1.2. Multi-Task Supervision

In the previous studies [16] [15], multi-task learning was adopted to alleviate the interruption of similar appearance structures in X-ray images. To deal with the same challenge, besides the class of the catheter status, the proposed framework also outputs the segmentation mask of the catheter and heatmap of the catheter tip, which also can improve the classification accuracy and make the classification result more trustworthy. These two components are regarded as two components of the multi-task supervision module.

The course of the catheter is one of the most essential pieces of information for catheter status detection. In the proposed framework, the course of the catheter is represented as the segmentation mask of the catheter. The segmentation mask contains two classes which are background and catheter. More specifically, the matrix is utilized to represent the ground truth segmentation mask of the catheter. In the matrix, the value of the pixels belonging to the catheter is assigned as 1 and the value of the pixels belonging to the background is assigned as 0. In the proposed network shown in Fig. 3, to generate the segmentation mask, the highest resolution feature maps of HRNet are considered as the input. After an extra 1×1 convolution operation reducing the channel from 64 to 1, the mask of the catheter at a lower resolution (1/4 of the initial input image) is generated, supervised by sigmoid cross-entropy loss.

The placement of the catheter tip is also regarded as significant information when radiologists detect the status of the catheter based on the X-ray image. In the proposed framework, the placement of the catheter tip is represented as the heatmap. More specifically, a 2D Gaussian function is utilized to represent the heatmap of the catheter tip, which is:

$$heatmap_{x,y} = \exp\Big(-\Big(\frac{(x-x_0)^2}{2\delta_x^2} + \frac{(y-y_0)^2}{2\delta_y^2}\Big)\Big) \tag{1}$$

where (x, y) denotes the coordinates of the pixel in the heatmap. The Gaussian kernel is used to represent the coordinate of the

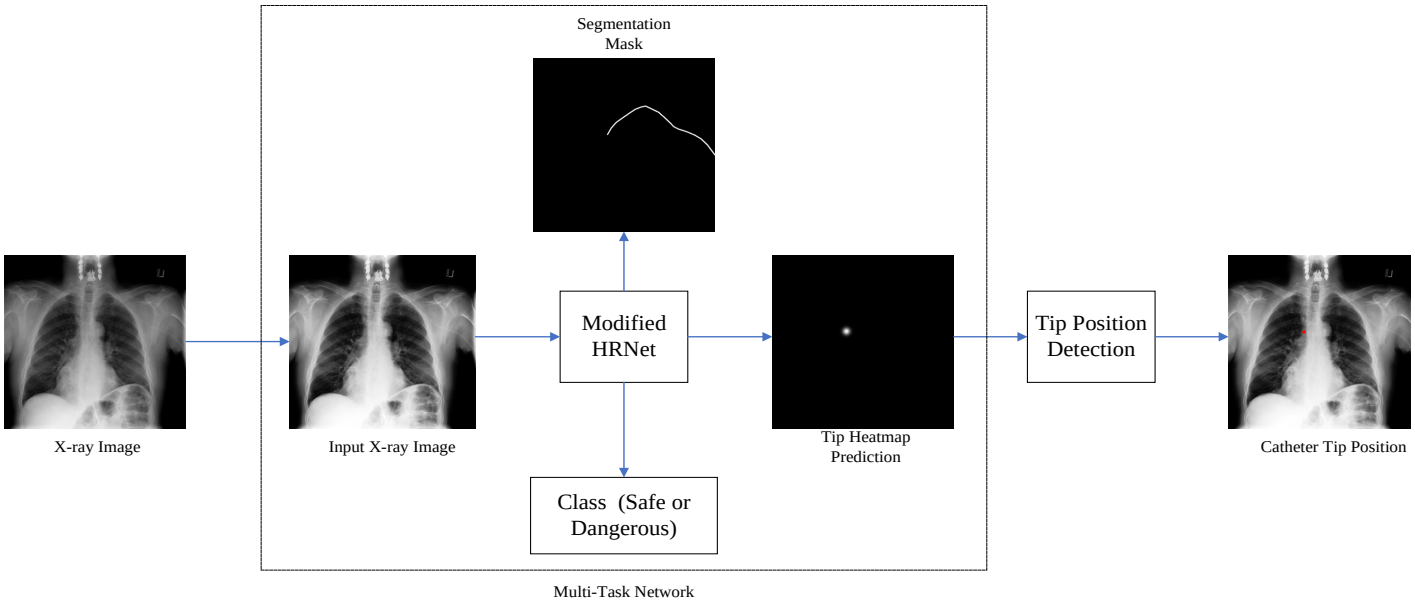


Fig. 2. The overall framework for catheter status classification. The input of the framework is the X-ray image which is regarded as the input of the status classification network. The segmentation mask of the catheter, the heatmap of the catheter tip, and the status of the catheter are the outputs of the classification network shown in Fig. 3. The detected coordinate of the catheter tip is highlighted as a red dot.

catheter tip in the heatmap. The horizontal and vertical coordinate of the tip is represented as x_0 and y_0. The variance of the x-axis and y-axis is indicated by δ_x and δ_y, determining the region where the catheter tip can be in. Additionally, in the proposed framework, the values of δ_x and δ_y are the same. In the proposed classification network shown in Fig. 3, to generate the heatmap of the catheter tip, the highest resolution feature maps of the modified HRNet are regarded as the input. After a 1×1 convolution operation reducing the channel from 64 to 1, the heatmap of the catheter tip at a lower resolution (1/4 of the initial input image) is generated, supervised by MSE (mean squared error) loss.

3.1.3. Classification Branch

The classification branch is proposed to classify the status of the catheter. There are two classes which are "Safe" and "Dangerous" respectively. "Safe" means that the catheter is correctly placed, which is represented as 1 in the proposed network. "Dangerous" means that the place of the catheter should be corrected, which is represented as 0 in the proposed network. After the final stage of the modified HRNet, the feature maps in different resolutions are downsampled to feature maps with the size of $16 \times 16 \times 256$ by convolution operations. Then these downsampled feature maps are summed up to produce a new feature map with the size of $16 \times 16 \times 256$, which is regarded as the input of the catheter status classification branch. Then a max pooling layer is added to the network, followed by a fully connected layer generating the class of the catheter status. Eventually, sigmoid cross entropy loss is utilized to supervise the classification branch.

3.1.4. Loss Function

In the loss function of catheter status classification, there are three main parts: catheter status classification loss; catheter segmentation loss; and heatmap regression loss. More specifically, binary cross-entropy loss is regarded as the loss function for the classification task in the proposed framework. There are two main reasons why binary cross-entropy loss is selected. Firstly, the cross-entropy loss is normally utilized in classification tasks [12]. Additionally, there are only two classes (safe and dangerous) in our task. Consequently, compared with the softmax cross-entropy loss which is implemented on the multi-class classification task, the binary cross-entropy loss is more suitable for our task. Similar to the status classification branch, the binary cross-entropy loss is also considered as the loss function for the segmentation task in the proposed framework. The reason why the binary cross-entropy loss is chosen for the segmentation task is similar to the loss function selection for the classification task. Firstly, the cross-entropy loss is widely implemented in segmentation tasks [2], which can perform well. Secondly, in our application scenario, only two classes which are catheter and background should be segmented. Therefore compared with the softmax cross-entropy loss, the binary cross-entropy loss is more appropriate for the segmentation task. Same as [16] where MSE loss was implemented to do heatmap regression of catheter tip, MSE loss is regarded as the loss function for the heatmap regression of the CVC tip in the proposed framework.

The binary cross-entropy loss [3] for CVC status classification can be represented as:

$$L_c = -(Y \log(C) + (1 - Y) \log(1 - C)) \tag{2}$$

where Y denotes the ground truth class of the catheter status and C represents the predicted class of the status. Moreover, the binary cross-entropy loss [17] for catheter segmentation can be

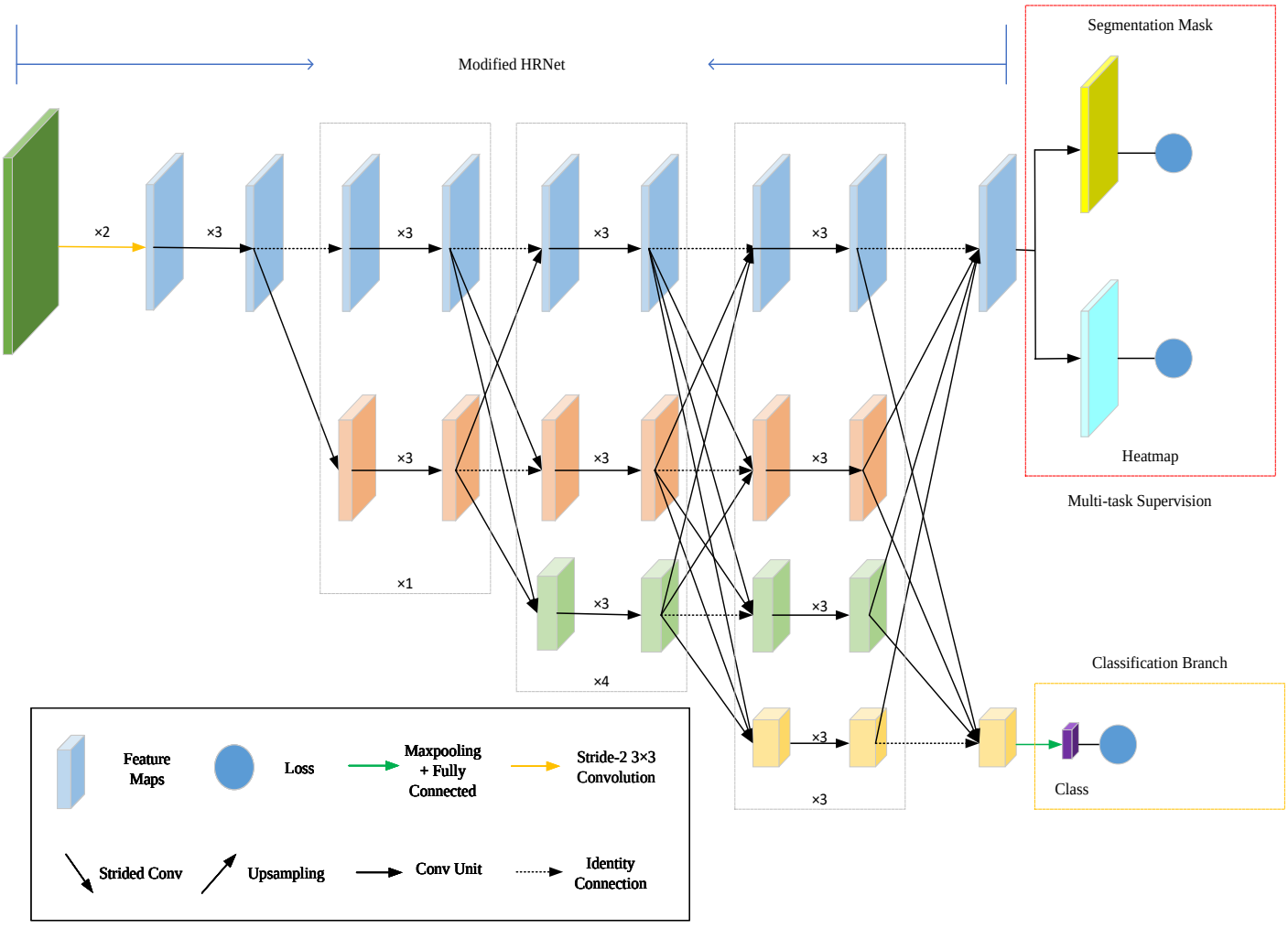


Fig. 3. The network structure for catheter status classification. The multi-task supervision is in the red box and the classification branch is in the yellow box. The input image for the classification network is the X-ray image shown in Fig. 2. Segmentation mask and heatmap indicate the segmentation mask of the catheter and the heatmap of the catheter tip in Fig. 2 respectively.

represented as:

$$L_s = \frac{1}{w \times h} \sum_{i \le w; j \le h} M_{ij} \log(S_{ij}) + (1 - M_{ij}) \log(1 - S_{ij}) \quad (3)$$

where the ground truth mask of the catheter is indicated by M and M_{ij} denotes the pixel value of the ground truth mask at the coordinates (i, j). S denotes the predicted segmentation mask of the catheter and S_{ij} is the pixel value of the predicted segmentation mask at the coordinates (i, j). Furthermore, the loss function for the heatmap regression can be represented as:

$$L_h = \frac{1}{w \times h} \sum_{i \le w; j \le h} (G_{ij} - P_{ij})^2 \quad (4)$$

where w and h denote the width and the height of the heatmap respectively. G is the ground truth heatmap of the tip and the pixel value at the coordinates (i, j) is represented as G_{ij}. The predicted heatmap is denoted by P and P_{ij} indicates the pixel value at the coordinates (i, j).

The loss function for the classification framework is:

$$L = \lambda L_h + L_s + \eta L_c \quad (5)$$

where η and λ are non-zero constant scalars to be determined which are used to balance the segmentation result, tip detection result, and classification result.

3.2. CVC Dataset Description

The dataset utilized in this paper is from CLiP, catheter, and line position dataset published in [29]. The public dataset contains 30,083 chest radiographs with image-level labels. Four different tubes including CVC are manually given the annotation of the category. There are three categories of the CVC given in the public dataset which are normal, borderline, and abnormal. The normal category includes CVCs which are not required to be reposed. The borderline category included CVCs that would ideally be reposed. Additionally, the abnormal category contained the CVCs which should be immediately reposed. In the CVC dataset utilized in this paper, the CVCs labeled as abnormal and borderline are considered in the same category labeled as dangerous, which means that the CVCs require repositioning. Moreover, the CVCs labeled as normal in the public dataset are classified as safe, which means that CVCs do not require repositioning.

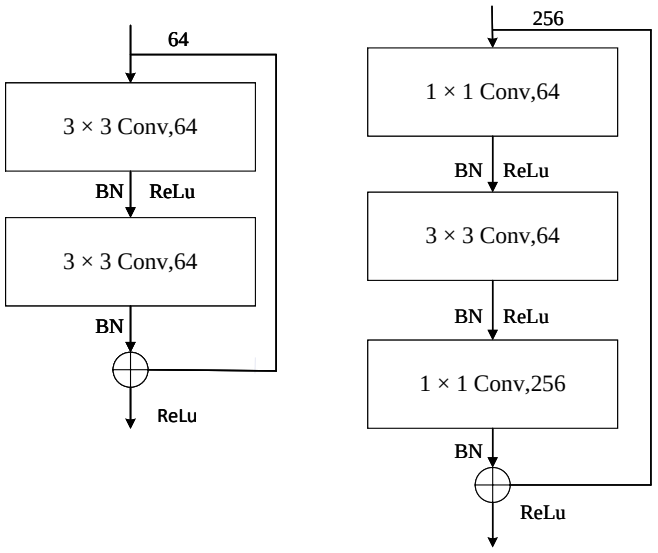


Fig. 4. The Bottleneck Residual unit is on the right-hand side and the basic Residual unit is on the left-hand side. BN indicates batch normalization and ⊕ indicates feature addition. The Bottleneck Residual unit is the basic convolution unit for the 1st stage in the module of modified HRNet in Fig. 3. The basic Residual unit is the basic convolution unit for the rest of the stages in the module of modified HRNet in Fig. 3.

Moreover, we only focus on status classification for a single CVC in our research. Therefore, the X-ray images which only contain a single CVC are selected from the public dataset. Additionally, only a subset of the chest radiographs in the public dataset are given the related segmented labels. However, the proposed framework requires segmentation annotation, which means that the X-ray images with related segmentation labels are selected for the CVC dataset in this paper. After image selection, we found that the number of samples labeled as safe is larger than the number of samples labeled as dangerous. Therefore, offline data augmentation is implemented on the CVC dataset to enlarge the number of samples for catheters classified as dangerous, to make the number of samples in different classes more balanced. More details of data augmentation are illustrated in Section 3.3. After data selection and augmentation, the CVC dataset contains 5089 training samples and 1095 test samples.

3.3. Image Rescaling and Augmentation

The size of each image in the dataset is ranging from 1024×1024 to 1536×1536. However, the proposed framework requires a fixed-size image to be input into. Therefore, ahead of inputting the X-ray image into the proposed framework, all the samples in the CVC dataset are downsampled to 512×512. In data augmentation, image cropping is implemented with a scale of 0.75. Because CVCs might be cropped in some of the samples, not all the samples in the CVC dataset can be cropped. Therefore, to avoid this issue, the annotation mask of the catheter is utilized to determine whether this image can be cropped or not when implementing image cropping.

4. Experiment and Results

4.1. Evaluation Metrics

Evaluation metrics are required to select the hyperparameters and model weights, as well as compare the performance of the proposed framework with other methods. Therefore, we investigate the performance by two different metrics which are AUC and accuracy.

AUC (Area under the ROC Curve) is defined based on the ROC Curve (receiver operating characteristic curve) which is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: true positive rate and false positive rate. That is, AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1). Moreover, the higher the value the model achieves, the better performance it is. Accuracy which is typically expressed as a percentage is the count of predictions where the predicted value is equal to the true value. It is binary (true/false) for a particular sample. Same as AUC, the higher the value the model achieves, the better performance it is.

4.2. Implementation Details

The size of the input image is 512×512 for the proposed framework. Additionally, the size of the segmentation mask and the regression heatmap is 128×128. The *sigma* of the 2D Gaussian function is set as 5 in the ground truth of the catheter tip heatmap. λ and η for the loss function are set as 10 and 0.1 respectively. In addition, the values of λ and η are achieved by trial and error. In each stage, the number of blocks is set as 3, chosen by experiments in Section 4.3.1. Additionally, the dropout rate of the dropout layer for the classification branch is set as 0.3, chosen by experiments shown in Section 4.3.2. The Adam optimizer [13] is adopted and the batch size is set as 4. Additionally, the learning rate is initially set as 0.001, adjusted to 0.0001 at epoch 6 and to 0.00001 at epoch 60. The modified HRNet of the proposed framework shown in Fig. 3 is initialized with the ImageNet pre-trained model provided by [14]. The learning epoch is set as 100 when training the proposed framework. The framework is implemented on PyTorch 1.8 and Python 3.8. All experiments are conducted with Nvidia Tesla V100.

4.3. Hyperparameter Selection

For the proposed framework, there are several hyperparameters required to be selected. To select them, 5-fold cross-validation is implemented on the training dataset to select the dropout probability of the classification branch and the number of basic blocks for the modified HRNet of the classification network. AUC and accuracy are utilized as the evaluation metrics to show the performance of each model with different values of each hyperparameter. More details are shown below.

4.3.1. Number of Basic Blocks

In each stage of the modified HRNet, the proposed network can contain more than one convolution block in different resolutions. The more blocks it contains, the more parameters the network will have and the larger computation complexity it will

Table 1. Results of AUC, accuracy, number of parameters (Params), computation complexity (GFLOPs) for block number selection and inference time (Time) for each image based on 5-fold cross-validation

Number of Block	Avg AUC×100	Avg Accuracy (%)	Params(M)	GFLOPs	Times (ms)
1	82.6 ± 1.4	83.3 ± 1.4	**10.3**	**17.9**	**8.8**
2	82.7 ± 0.7	83.7 ± 0.8	16.6	25.8	9.0
3	84.3 ± 1.2	**84.7 ± 1.6**	22.9	33.7	11.2
4	**84.4 ± 1.6**	84.6 ± 1.6	29.3	41.7	15.0

Table 2. Results of mean Accuracy and mean AUC for dropout probability selection for classification branch based on 5-fold cross-validation

Dropout Probability	Avg AUC×100	Avg Accuracy (%)	Params (M)
0.1	83.2 ± 0.8	84 ± 1.4	22.9
0.2	83.4 ± 0.6	83.9 ± 0.6	22.9
0.3	**84.3 ± 1.2**	**84.7 ± 1.6**	22.9
0.4	83.3 ± 1.1	83.9 ± 1.2	22.9

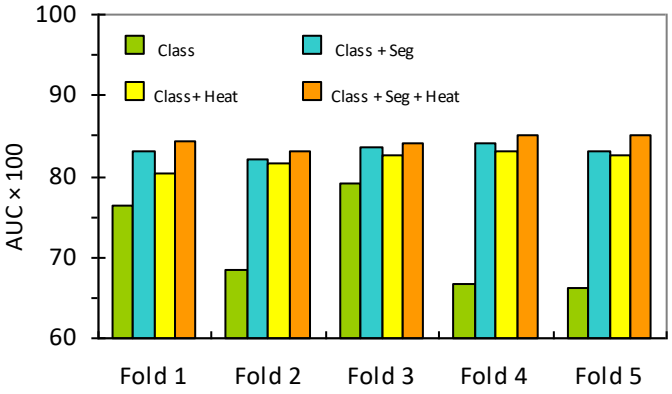


Fig. 5. The bar chart of the AUC results for four catheter status classification methods based on a 5-fold validation dataset. Notation: Class denotes classification branch; Heat denotes heatmap regression supervision; Seg denotes segmentation supervision.

have. Therefore, to find the number of blocks that are the most suitable for the proposed network, several experiments are carried out. To alleviate the effect of other hyperparameters, the dropout probability of the classification branch is set as 0.3 for all experiments. We set the number of blocks as 1, 2, 3, and 4 in the 5-fold cross-validation experiments. The average results can be found in Table 1. Average AUC and accuracy for five folds are considered when choosing the most suitable settings for both hyperparameters. However, the number of blocks not only affects the value of AUC and accuracy but also the inference time. In addition, inference time is one of the most important factors. Because, the faster the inference time it has, the faster the status of the catheter could be detected to avoid complications. Therefore, the inference time for the model with different numbers of blocks is given in Table 1. It is demonstrated that the AUC is the highest, but the inference time is highest when the block is set as 4. On the contrary, the inference time is the lowest, but the average AUC is lowest when setting the block as 1. Compared to the performance gained by the framework with 3 blocks and 4 blocks, it is illustrated that the average AUC of the framework with 4 blocks is only 0.01 higher than that of the framework with 3 blocks. However, the average accuracy for the model with 3 blocks is 0.8 % higher than the method with 4 blocks. Moreover, the inference time for each image of the method with 3 blocks is much less than that of the method with 4 blocks. Therefore, to make a trade-off between the classification accuracy, the AUC, and the inference time, the number of blocks is set as 3 in the proposed framework.

4.3.2. Dropout Probability

To overcome the overfitting problem and make the framework more generalized, the dropout layer is added at the end of the classification branch. To find the optimal value of dropout probability, several simulations are done and the results are illustrated in Table 2. The optimal value of the model is searched based on 5-fold cross-validation. When doing simulation, to reduce the effect of other hyperparameters, the number of the convolution block for each resolution is set as 3 in each stage. Therefore, the parameters of each model are the same. To show the performance for different values of the dropout probability, the average AUC and accuracy for the model with different values of dropout probability based on 5-fold cross-validation are shown in Table 2. It is illustrated that the framework can get the highest average AUC and accuracy for five folds when the dropout probability is set as 0.3. Therefore, the value of the dropout probability is set as 0.3, in our proposed framework.

4.4. Ablation Study

To verify the proposed framework and integration of multi-task supervision including heatmap regression supervision and segmentation supervision, performance is evaluated with different modules based on 5-fold cross-validation. The average AUC for five-folds, the average accuracy for five-folds, the number of the parameters, and the computation complexity of each model are demonstrated in Table 3. From the performance shown in Table 3, it is shown that the model only with the classification branch (case 1) achieves the lowest average AUC and lowest accuracy among all methods. On the contrary, the model with segmentation supervision, the heatmap regression supervision, and the classification branch (case 4) can gain the highest average AUC which is 0.843 and the highest average accuracy which is 84.7 %. Moreover, compared with case 1, case 4 can increase 0.129 AUC and 9.7 % accuracy, without an increase in the number of parameters.

4.4.1. Evaluation with multi-task Supervision

The multi-task supervision combines two parts which are heatmap regression supervision as well as segmentation super-

Table 3. The average AUC and the average accuracy of 5-fold cross-validation, the computation complexity (GFLOPs) as well as the number of parameters (Params) for four catheter status classification methods. The Class indicates the classification branch, the Seg represents segmentation supervision and the Heat indicates heatmap regression supervision.

Case	Model	Class	Heatmap	Seg	Average AUC × 100	Average Accuracy	Params (M)	GFLOPs
1	HRNet	✓	×	×	71.4 ± 6.0	75.0 ± 4.9	22.9	**33.741**
2	HRNet	✓	×	✓	83.1 ± 1.0	83.9 ± 1.4	22.9	33.742
3	HRNet	✓	✓	×	82.0 ± 1.1	82.5 ± 1.1	22.9	33.742
4	HRNet	✓	✓	✓	**84.3 ± 1.2**	**84.7± 1.6**	22.9	33.743

vision. To verify the contribution of this module, a set of experiments is carried out. After 5-fold cross-validation, the performance of each model is shown in Table 3. It is illustrated that the model with multi-task supervision (case 4) gains 0.129 average AUC increase and 9.7% accuracy increase, compared with the model only including the classification branch (case 1). To demonstrate the performance of the model with and without the multi-task supervision module for each fold, the AUC for each model is shown in Fig. 5. It shows that, for each fold, the AUC of case 1 is lower than that of case 4. Especially for fold 4 and fold 5, the difference in the AUC between these two models is much larger. Therefore, after 5-fold cross-validation on the training dataset, it can be inferred that multi-task supervision is useful for the proposed catheter status classification framework. More details of the contribution of each component (segmentation supervision and heatmap regression supervision) of multi-task supervision are given in Section 4.4.2 and Section 4.4.3.

4.4.2. Evaluation with Heatmap Regression Supervision

To test the contribution of heatmap regression supervision, two sets of experiments are carried out based on 5-fold cross-validation. Without the effect of segmentation supervision, the performance of the models with heatmap (case 3) and the model without it (case 1) is shown in Table 3. It is demonstrated that case 3 can bring 0.106 average AUC growth and 7.5% average accuracy increase, compared with case 1. In addition, from Fig. 5, it can be found that the AUC of case 3 for each fold is larger than it of case 1. Furthermore, to verify the contribution of heatmap regression supervision when segmentation supervision is considered, the performance of the model with both heatmap and mask (case 4) and the model with only the mask (case 2) is demonstrated in Table 3. It is illustrated that case 4 can bring 0.023 average AUC increase and 2.2% average accuracy growth, compared with case 2. From Fig. 5, it is shown that the model with both heatmap and mask can perform better than the model with two outputs for each fold.

4.4.3. Evaluation with Segmentation Supervision

To investigate the contribution of the mask segmentation supervision, two sets of experiments are carried out, based on 5-fold cross-validation. The average AUC of five folds, the average accuracy of five folds, the number of parameters, and the computation complexity for the model only with the class of the catheter status (case 1), as well as the model with segmentation supervision and the class of the catheter status (case 2), are shown in Table 3. Compared with case 1, case 2 can bring a 0.117 average AUC growth and 7.5 % average accuracy growth, without the increase in the number of parameters. In Fig. 5, the AUC for each fold of case 1 and case 2 is demonstrated. It can be found that the AUC for the model with two outputs can perform better than the model without the segmentation supervision for each fold. Besides, when the effect of the heatmap regression supervision for the proposed framework is considered, the performance of the model combining the class of the status of the catheter, the heatmap regression supervision, and the segmentation supervision (case 4), as well as the model only including the heatmap regression supervision and the class of the catheter status (case 3), is compared based on experiments. From Table 3, it is shown that the average AUC of case 4 is 0.023 higher and the average accuracy is 2.2% larger than case 3. From Fig. 5, it is demonstrated that the AUC of each fold for the model including the segmentation supervision is higher than the model without it.

Table 4. AUC, Accuracy, number of parameters (Params) and inference time (Time) for three comparison frameworks on test dataset

Method	AUC × 100	Accuracy(%)	Params(M)	Time (ms)
ATCM[25]	68.2	73.2	**21.8**	**5.5**
EDMC[12]	77.0	76.0	28.3	9.0
Proposed	**82.3**	**82.6**	22.9	11.2

4.5. Comparison Study

Several state-of-the-art models for catheter status classification are carried out for comparison study. To ensure a fair comparison, all the experiments are conducted, using the same rescaling and augmentation techniques on 5 cross-validation datasets. For each method, the best performance model that achieves the highest AUC for each dataset is selected based on the model achieving the best validation AUC. Comparing among all the best-performing models gained by 5 datasets, the one that gains the highest validation AUC is selected as the best-performance model of each method. Two existing methods which are ACTM [25] method and EDMC [12] method are selected as the comparison methods. ACTM method is a one-stage framework only outputting the predicted class of the catheter. EDMC method is a two-stage framework dividing the catheter status classification into two stages. In the first stage, a network is implemented to segment the mask of the catheter as well as two endpoints of the catheter. In the second stage, a classification network is implemented to generate the class of the catheter based on the concatenation of the X-ray image and the segmented masks generated by the network in the first stage.

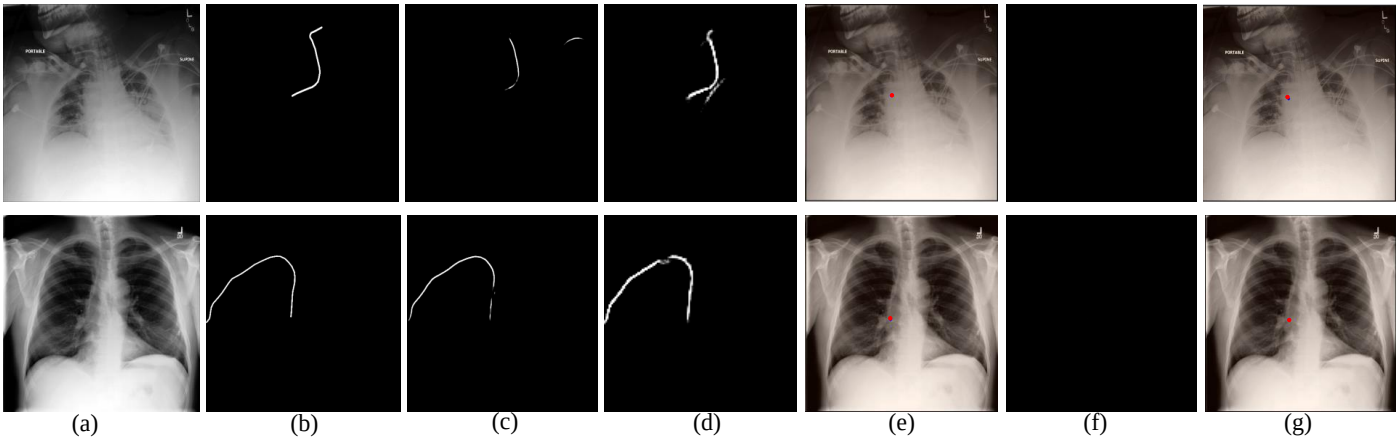


Fig. 6. (a)The X-ray image (b) The ground truth mask of the catheter (c) The catheter mask generated by the segmentation network of the EDMC method (d) The mask of the catheter produced by the proposed method (e) The ground truth position of catheter represented as a red dot (f) The mask of the catheter endpoint generated by EDMC method (g) The position of the catheter tip generated by the proposed method, which is represented as a red dot

The performance of the proposed method and the two comparison methods based on the test dataset is shown in Table 4. It can be found that the ACTM method [25] achieves the lowest AUC and the lowest accuracy, which are 0.141 and 9.4%, respectively, lower than the proposed method with a similar number of parameters (The number of parameters for ACTM method is only 1.1 M smaller than the proposed method). Even though, it is shown that the inference time of the proposed framework is longer than the inference time of the ACTM method. However, the ACTM gains the shortest inference time, sacrificing the accuracy and AUC. Additionally, although both the ACTM method and the proposed method belong to a one-stage framework, the ACTM method is only trained to learn the class of the catheter. However, the proposed method can learn the tip position of the catheter and the course of the catheter. From the above results, it can be inferred that the segmentation supervision as well as the heatmap supervision can help to increase classification accuracy with less increase in the number of parameters. Comparing the ACTM method with the EDMC method, it is shown that the EDMC method which is the two-stage framework can gain higher AUC and accuracy than the ACTM method which is the one-stage framework without being assisted by other auxiliary information such as the course of the catheter. Therefore, it can be concluded that, compared with the one-stage framework without being assisted by any other auxiliary information, the two-stage framework can gain higher AUC and accuracy.

Compared with the EDMC method, the AUC and the accuracy of the proposed method are 0.053 and 6.6% higher. Although the inference time of the classification network for the EDMC method is shorter than that of the proposed method. However, the two-stage framework does not only include the classification network but also contains the segmentation network, which means that the inference time for the segmentation network of the two-stage framework also should be considered. Additionally, the inference time for the segmentation network is 16 ms for each X-ray image. Therefore, for each X-ray image, to generate the class of the catheter status, the whole inference time for the two-stage framework is 25 ms. To justify the reason why the proposed method performs better than the two-stage framework, the mask of the whole body of the catheter and the tip position generated by the EDMC method and these of the proposed method shown in Fig. 6 are investigated. It is shown that the mask of the whole body of the catheter and the tip position generated by the proposed method are much better than those generated by the EDMC method. In the first row of Fig. 6, it is demonstrated that the course of the catheter generated by the EDMC method does not contain several significant parts such as the area near the catheter tip. Moreover, it is shown that the tip position cannot be segmented by the EDMC method. Therefore, the low-quality segmentation results of the EDMC method lead the network to produce wrong classification results. In the second row of Fig. 6, it can be found that there is a huge gap between the two parts of the catheter mask generated by the EDMC method. Additionally, the two endpoints of the CVC are not able to be detected. However, the mask of the catheter is considered a part of the input of the EDMC method, which means that the performance of the classification network is highly dependent on the quality of the generated mask. If the quality of the mask is not high enough, then it will have a negative effect on the performance of the catheter classification. However, it is demonstrated that the tip position and the course of the CVC generated by the proposed method are much more similar to the ground truth. Because the input of the proposed method is only the X-ray image, the proposed method does not highly rely on the segmentation mask of the catheter as well as the tip position. The placement of the catheter is regarded as assisted information to help interpret the framework and assist in increasing the classification performance. Moreover, the proposed method is based on the network maintaining high-resolution features from the start to the end. However, the segmentation network of the EDMC method is based on an encoder-decoder structure, which means that the high-resolution features cannot be considered from the start to

the end.

4.6. Error Analysis of Results

When checking the classification accuracy and AUC for each image, it can be found that the performance of most of the X-ray images with a small scale performs worse than the X-ray images with a large scale. As the catheter status classification by clinicians, when enlarging the scale of the X-ray image, the details such as the location of the catheter tip as well as the essential points of other anatomical structures will become much clearer. Then it will be easier to recognize the location of the catheter tip and other anatomical structures that are utilized to determine the status of the catheter. Additionally, from Fig. 7, two examples are used to illustrate the classification error made by the proposed method. The label of the sample shown in the first row of Fig. 7 should be 'safe', but it is classified as 'dangerous'. It can be found that the sample in the first row of Fig. 7 includes different kinds of tubes and catheters and the position of the CVC is difficult to distinguish from the background. From the course of the catheter and the tip of the catheter generated by the proposed method, it is shown that a part of the other tube is considered a part of the CVC. Additionally, the tip position of the CVC is also unsuccessfully detected. A point along another tube is considered the tip. Therefore, it can be inferred that the model does not find the place where the CVC is located, leading to the classification error. The label of the sample demonstrated in the second row of Fig. 7 should be 'dangerous', however, it is classified as 'safe'. The class of the catheter is wrongly predicted, although the course of the catheter, as well as the tip position, are both successfully detected. The main reason why the status of the catheter is misclassified is that the position of the catheter is inserted close to the safe place where the CVC should be inserted.

5. Discussion

5.1. Discussion on Multi-task Supervision

From the results shown in Fig. 5, it is shown that the proposed method with multi-task supervision including segmentation supervision and heatmap regression supervision can gain better performance than the method only with the classification branch. Therefore, it can be inferred that based on the same backbone network (HRNet) which can maintain high-resolution features from the start to the end, the framework with multi-task supervision can perform better. The multi-task supervision can supply the network with more assisted information when generating the class of the catheter. To investigate the contribution of segmentation supervision and heatmap regression supervision separately, the ablation experiments are also carried out and the results are shown in Section 4.4. It is demonstrated that the accuracy and AUC achieved by the framework with only segmentation supervision or heatmap regression supervision are lower than those gained by the method with both of them. Therefore, it can be inferred that the more assisted information is given, the higher performance will be achieved. Furthermore, comparing the method with heatmap regression supervision with the method with segmentation supervision, it is shown that the method with segmentation supervision can get better AUC and accuracy than the method with heatmap regression supervision. Consequently, it can be inferred that the course of the catheter is much more significant when generating the class of the status of the catheter. Additionally, as illustrated in Section 4.5, compared with the proposed method with the ATCM method which only generates the class of the catheter, the accuracy of the proposed method is much higher than that of the ATCM method, with a similar number of parameters. Therefore, based on the above experiments, it can be concluded that multi-task supervision is helpful for the classification task.

5.2. Discussion on Explainability

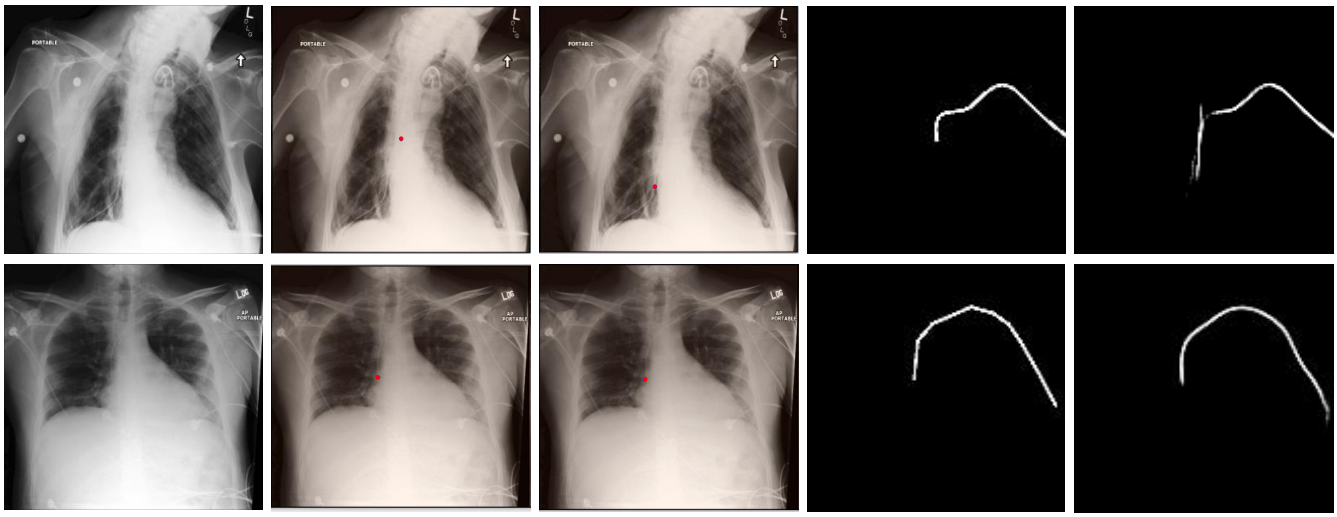
The proposed method is not only able to generate the class of the status of the catheter but also produce the course of the catheter and the tip position of the catheter, shown in Fig. 3. During inference, the course of the catheter and the tip position of the catheter can be regarded as a piece of evidence generated by the framework to show the reason why the class of the status of the catheter belongs to safe or dangerous. Therefore, it makes the framework more understandable to the human. Furthermore, when the status of the catheter fails to be classified, the segmentation mask of the catheter as well as the catheter tip position can assist in finding the reason why the mistake is made, which makes the classification prediction more explainable.

5.3. Discussion on Framework Limitation

When more than one CVC appears in an X-ray image, the proposed framework cannot classify the status of each individual CVC separately. The model which can detect multiple CVCs in the same image should be considered in future work. Additionally, even though, the AUC and accuracy for the proposed framework are much better than those of the one-stage framework only generating the class of the status of the catheter. However, the one-stage framework only outputting the class of the catheter status has a shorter inference time for each test image than the proposed method. Consequently, a more efficient multi-task learning framework can be investigated in the future.

6. Conclusion

To increase the efficiency of CVC status classification, a multi-task framework that is not only able to generate the class of the catheter but also the tip position of the catheter and the mask of the whole body of the catheter is proposed. There are three main innovations in our work: 1) The proposed method adopts HRNet as the backbone, which can maintain high-resolution features from the start to the end to generate more precise assisted information such as the mask of the whole body of the catheter and the tip position of the catheter. 2) The proposed multi-task framework combines three tasks (mask segmentation, heatmap regression, and classification) into a unique network to further increase the classification performance. 3) During the inference, the generated mask of the catheter and the tip position of the catheter can make the classification result more understandable to a human. After the ablation study,



(a)X-ray Image (b)Ground Truth Tip Position (c)Predicted Tip Position (d)Ground Truth Mask (e)Predicted Mask

Fig. 7. Two misclassified samples. The label of the sample shown in the first row is safe. The label of the sample demonstrated in the second row is dangerous. (a) The X-ray image (b) The ground truth tip position of the catheter represented as a red dot (c) The predicted tip position of the catheter represented by a red dot (d) The ground truth of the mask of the catheter (e) The predicted mask of the catheter

it is shown that each component of the proposed framework is helpful for classification performance. Additionally, after comparing the proposed method with two state-of-the-art methods for catheter status classification, it is shown that the proposed method performs the best.

Acknowledgments

This work was partly supported by King's College London and the China Scholarship Council and has been performed using resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service funded by EPSRC Tier-2 capital grant EP/P020259/1, Cirrus UK National Tier-2 HPC Service at EPCC funded by the University of Edinburgh and EPSRC grant EP/P020267/1, and Jade2 National Tier-2 HPC Service funded by EPSRC grant EP/T022205/1.

Declaration of Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Amador, T., Saturnino, S., Veloso, A., Ziviani, N., 2022. Early identification of icu patients at risk of complications: Regularization based on robustness and stability of explanations. Artificial Intelligence in Medicine. 128, 102283.

[2] Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 39, 2481–2495.

[3] Buja, A., Stuetzle, W., Shen, Y., 2005. Loss functions for binary class probability estimation and classification: Structure and applications. Working draft. 3, 13.

[4] Caruana, R., 1998. Multitask learning.

[5] Hansen, L., Sieren, M., Hobe, M., Saalbach, A., Schulz, H., Barkhausen, J., Heinrich, M.P., 2021. Radiographic assessment of cvc malpositioning: How can ai best support clinicians?, in: Medical Imaging with Deep Learning.

[6] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[7] Henderson, R.D., Yi, X., Adams, S.J., Babyn, P., 2021. Automatic detection and classification of multiple catheters in neonatal radiographs with deep learning. Journal of digital imaging. 34, 888–897.

[8] Islam, M., Vibashan, V., Lim, C.M., Ren, H., 2021. St-mtl: Spatio-temporal multitask learning model to predict scanpath while tracking instruments in robotic surgery. Medical Image Analysis. 67, 101837.

[9] Itzkovich, D., Sharon, Y., Jarc, A., Refaely, Y., Nisky, I., 2021. Generalization of deep learning gesture classification in robotic-assisted surgical data: From dry lab to clinical-like data. IEEE Journal of Biomedical and Health Informatics. 26, 1329–1340.

[10] Kang, M., Bae, J., Moon, S., Chung, T.N., 2021. Chest radiography for simplified evaluation of central venous catheter tip positioning for safe and accurate haemodynamic monitoring: a retrospective observational study. BMJ open 11, e041101.

[11] Kao, E.F., Jaw, T.S., Li, C.W., Chou, M.C., Liu, G.C., 2015. Automated detection of endotracheal tubes in paediatric chest radiographs. Computer methods and programs in biomedicine. 118, 1–10.

[12] Khan, A.B.M., Ali, S.M.A., 2021. Early detection of malpositioned catheters and lines on chest x-rays using deep learning, in: International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), pp. 51–55.

[13] Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

[14] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 60, 84–90.

[15] Laina, I., Rieke, N., Rupprecht, C., Vizcaíno, J.P., Eslami, A., Tombari, F., Navab, N., 2017. Concurrent segmentation and localization for tracking of surgical instruments, in: Medical Image Computing and Computer-Assisted Intervention(MICCAI), pp. 664–672.

[16] Ma, H., Smal, I., Daemen, J., van Walsum, T., 2020. Dynamic coronary roadmapping via catheter tip tracking in x-ray fluoroscopy with deep learning based bayesian filtering. Medical image analysis. 61, 101634.

[17] Ma, Y.d., Liu, Q., Qian, Z.B., 2004. Automated image segmentation using improved pcnn model based on cross-entropy, in: Proceedings of 2004

International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004., pp. 743–746.

[18] Moureau, N.L., 2019. Vessel health and preservation: the right approach for vascular access. Springer Nature.

[19] Ni, Z.L., Bian, G.B., Li, Z., Zhou, X.H., Li, R.Q., Hou, Z.G., 2022. Space squeeze reasoning and low-rank bilinear feature fusion for surgical image segmentation. IEEE Journal of Biomedical and Health Informatics. 26, 3209–3217.

[20] Polderman, K.H., Girbes, A.R., 2002. Central venous catheter use: Part 1: Mechanical complications. Intensive Care Medicine. 28, 1–17.

[21] Psychogyios, D., Mazomenos, E., Vasconcelos, F., Stoyanov, D., 2022. Msdesis: Multitask stereo disparity estimation and surgical instrument segmentation. IEEE transactions on medical imaging. 41, 3218–3230.

[22] Ramakrishna, B., Brown, M., Goldin, J., Cagnon, C., Enzmann, D., 2011. Catheter detection and classification on chest radiographs: an automated prototype computer-aided detection (cad) system for radiologists, in: Medical Imaging 2011: Computer-Aided Diagnosis, pp. 892–897.

[23] Ramakrishna, B., Brown, M., Goldin, J., Cagnon, C., Enzmann, D., 2012. An improved automatic computer aided tube detection and labeling system on chest radiographs, in: Medical Imaging 2012: Computer-Aided Diagnosis., pp. 231–237.

[24] Sharan, L., Romano, G., Koehler, S., Kelm, H., Karck, M., De Simone, R., Engelhardt, S., 2021. Mutually improved endoscopic image synthesis and landmark detection in unpaired image-to-image translation. IEEE Journal of Biomedical and Health Informatics. 26, 127–138.

[25] Singh, V., Danda, V., Gorniak, R., Flanders, A., Lakhani, P., 2019. Assessment of critical feeding tube malpositions on radiographs using deep learning. Journal of Digital Imaging. 32, 651–655.

[26] Subramanian, V., Wang, H., Wu, J.T., Wong, K.C., Sharma, A., Syeda-Mahmood, T., 2019. Automated detection and type classification of central venous catheters in chest x-rays, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 522–530.

[27] Sun, Y., Wang, L., Jiang, Z., Li, B., Hu, Y., Tian, W., 2020. State recognition of decompressive laminectomy with multiple information in robot-assisted surgery. Artificial Intelligence in Medicine. 102, 101763.

[28] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.

[29] Tang, J.S., Seah, J.C., Zia, A., Gajera, J., Schlegel, R.N., Wong, A.J., Gai, D., Su, S., Bose, T., Kok, M.L., et al., 2021. Clip, catheter and line position dataset. Scientific Data. 8, 1–7.

[30] Turgay, A.S., Khorshid, L., 2010. Effectiveness of the auscultatory and ph methods in predicting feeding tube placement. Journal of Clinical Nursing. 19, 1553–1559.

[31] Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., Van Gool, L., 2021. Multi-task learning for dense prediction tasks: A survey. IEEE transactions on pattern analysis and machine intelligence. 44, 3614–3633.

[32] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al., 2020. Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence. 43, 3349–3364.

[33] Yi, X., Adams, S.J., Henderson, R.D., Babyn, P., 2020. Computer-aided assessment of catheters and tubes on radiographs: how good is artificial intelligence for assessment? Radiology: Artificial Intelligence. 2.