Semi-supervised Learning with Contrastive and Topology Loss for Catheter Segmentation and Misplacement Prediction

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

Chest X-ray is commonly used to assess the appropriateness of catheter placement since misplacement could lead to serious complications. With the emergence of deep learning, computer aided detection methods were developed to assist radiologists by detecting catheter misplacement and highlighting the course of the catheter. However, large pixel-wise labeled datasets are often unavailable since such annotation is laborious. To confront this problem, we proposed a novel semi-supervised learning method combining contrastive loss and topology loss that exploits prior knowledge about the segmentation target. This takes advantages of the topology property of catheter which does not require labeling. We collected 5767 chest X-rays from ***** which are labeled for misplacement of nasogastric tube and endotracheal tube catheters and possess a pixel-wise annotation of each catheter and relevant anatomical landmarks. To enhance the robustness and generalizability of our model, we use a public CLiP dataset with catheters as unlabeled data for semi-supervised learning. We use U-Net with classification head to perform segmentation of catheter and misplacement classification simultaneously. We reached AUC of 0.96 for classification and 0.59 for segmentation Dice score.

Keywords: Semi-supervised Learning, Contrastive Learning, Topology Loss, Catheter Misplacement, Catheter Segmentation

1. Introduction

Chest X-ray is commonly used to assess the appropriateness of catheter placement. Misplacement of catheter can lead to serious complications and thus catheter placement needs to be evaluated quickly and precisely. For example, incorrect placement of an endotracheal tube (ETT) might leads to complications such as hypoxemia, pneumothorax and pneumonia (**?**), while a misplaced nasogastric tube (NGT) might causes respiratory failure and even death (**?**). Studies have shown that the estimated misplacement rate of ETT and NGT can be up to 28% and 15% respectively (**?**). With the advancement of computer vision and deep learning, computer aided detection (CAD) methods were developed to assist radiologists by detecting catheter misplacement (**?????**). Nonetheless, for CAD to be clinically useful, model explainability is an important factor to consider. Thus semantic segmentation methods of catheter on chest X-ray were also developed (**??**).

However, large datasets with pixel-wise labels are often unavailable since pixel-level annotation is laborious and requires expert knowledge. There is also a class imbalance problem since misplaced catheters occur far less than correctly placed catheters. Not to mention that the course of the catheter can vary significantly, especially for misplaced

© 2023 Anonymous.

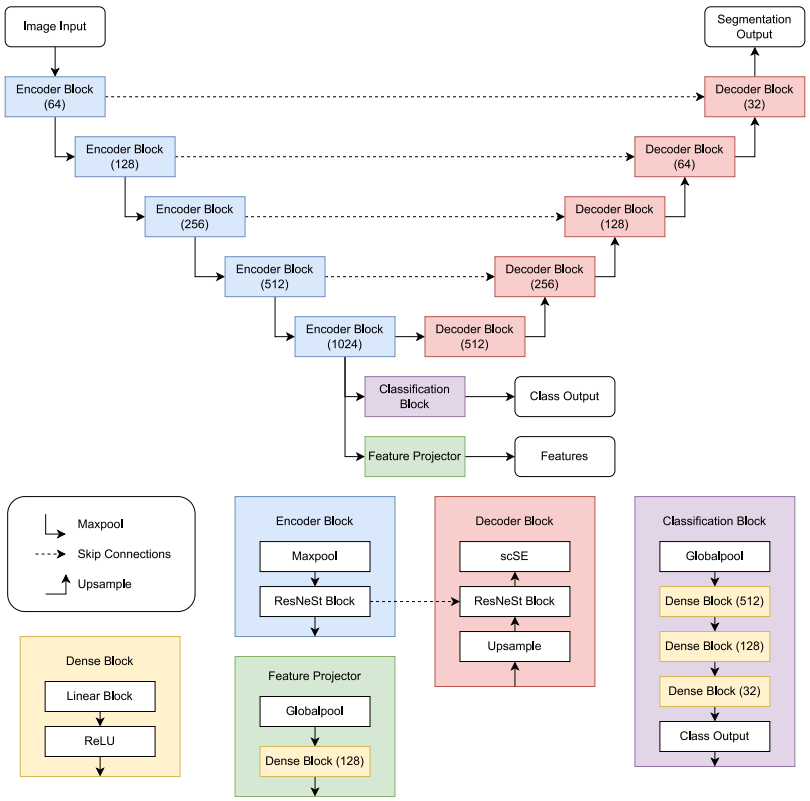

Figure 1: Modified U-Net

catheters. It is strenuous for a model to segment misplaced catheters as well as predicting if catheter is not present or misplaced. Semi-supervised contrastive learning techniques have been used to confront this unbalance problem by utilizing large unannotated dataset (**?**). Realizing that the catheter is a physical connected object, we can use its expected topology as prior to compute a topology loss. Note that topology loss also does not require labeling as the topology is assumed to be the same across all images. In this work, we utilize both methods and propose semi-supervised learning with contrastive loss and topology loss.

We developed an algorithm to do both catheter segmentation and misplacement prediction simultaneously using U-Net (**?**) with a classification head. The model will predict misplacement of ETT or NGT, as well as segment the catheters and relevant anatomical landmarks.

2. Methods

2.1. Model

The model we use is based on U-Net (**?**), a fully convolutional neural network popular for biomedical image segmentation. U-Net contains an encoder that extracts features of different spatial resolutions and uses skin connections to pass feature maps to a decoder that reconstruct the mask by pixel-wise classification. The backbone of the encoder we

use is ResNeSt-50 (**?**), which implements channel-wise attention mechanism to capture cross-feature interactions and learn diverse representations. We also apply Spatial and Channel Squeeze and Excitation (scSE) (**?**) to the decoder blocks, which aggregates spatial information and feature information to enhance segmentation performance.

In addition to the U-Net architecture for segmentation, we add a classification head and a feature projector after the encoder. Figure 1 shows the full network architecture of our model.

2.2. Self-Supervised Contrastive Learning

The goal of self-supervised contrastive learning is to learn visual representation from unlabeled images. We use the SimSCL framework proposed by (**?**). For each input batch $B = \{x_1, x_2, ..., x_b\}$, where x_i represents each input image, every image is performed augmentations twice to obtain $B' = \{x_1, x_2, ..., x_{2b}\}$. Let $\sigma(i)$ be the index of image that is derived from the same image as index i. Then denote the encoder as $E(\cdot)$ and the feature projector as $g(\cdot)$. The contrastive loss is defined as

$$L = -\frac{1}{2b}\sum_{i=1}^{2b}\log\frac{\exp(\mathrm{sim}(z_i, z_{\sigma(i)})/\tau)}{\sum_{k=1}^{2N}\mathbb{1}_{[k\neq i]}\exp(\mathrm{sim}(z_i, z_k)/\tau)} \tag{1}$$

where $z_i = g(E(x_i))$ and $\mathrm{sim}(u, v) = \frac{u^T v}{||u||\cdot||v||}$. The augmentations we using is the composition of contrast adjust, Gaussian noise, affine, 2D-elastic, grid distortion and horizontal flip.

2.3. Topology Loss

Topology loss is a connectivity loss for segmentation that is based on persistent homology proposed in (**?**). It requires topology prior regarding the object to segment, which are the Betti numbers. For example, zero-dimensional Betti number, β_0, is the number of connected components and a one-dimensional Betti number, β_1, is the number of holes. This loss is computed by first constructing a barcode diagram. This is based on changes of topology features with respect to different pixel-wise thresholds on the output probability map of segmentation. When the threshold is increasing from zero to one, topological structures on the output are created and destroyed. The birth and death of each topological structure is recorded in the diagram as a barcode. We can then put constraints on this diagram based on the topology prior. The loss is defined as

$$Topo(\hat{y}) = \sum_k\sum_{l=1}^{\beta_k}(1 - |B_{l,k}|)^2 + \sum_{\beta_k+1}^{l_{max}}|B_{l,k}|^2 \tag{2}$$

where $|B_{l,k}|$ is the length of l-th longest barcode of dimension k. Since the input of our task is a two-dimensional image, the Betti numbers for two-dimensional and above is always zero. Moreover, due to the nature of catheters, we can restrict the output to be one connected component but not on the number of holes since it is possible for catheter to curl and form holes. As for the value of l_{max}, we found 20 to work well in our setting since most barcodes will be zero.

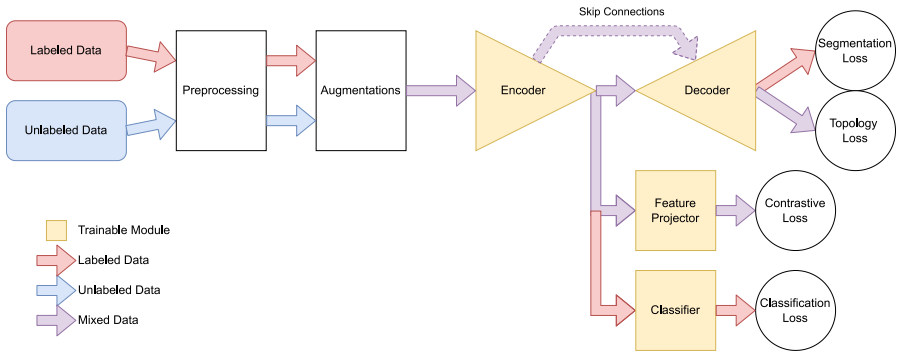

Figure 2: The complete training pipeline.

One thing to note is that computing this topological loss is very computationally expensive, and it can only be computed on CPU. In practice, we before we compute the loss we first downscale the output to 16×16 by max-pooling since the time consumption scales linearly with the number of pixels. Downsampling to such resolution still makes sense since it gives a constraint on the global structure of the segmentation prediction.

For image with no catheter present,

2.4. Semi-supervised Learning

Figure 2 shows the semi-supervised training pipeline that we proposed. Noticing that both contrastive learning and topology loss do not require labels to be trained, we combine both method for semi-supervised training. Each input batch is mixed with labeled and unlabeled data, where only topology loss and contrastive loss are computed for unlabeled data and all losses are computed on labeled data.

Topology loss can help the model to learn predicts segmentation with given topology prior, acting as a constrain to the parameter space. As for the contrastive loss, it helps the encoder to learn useful visual representations even for unlabeled data, further increasing the robustness of the model.

3. Experiments and Results

3.1. Data

3.1.1. CLiP

This Catheter and Line Position (CLiP) dataset is provided by (**?**). It contains 30,083 chest X-rays that includes data with misplaced catheter. It contains misplacement label of NGT and ETT, as well as their segmentation mask. In our task this dataset is used as unlabeled data.

3.1.2. *****

We collected 5767 chest X-rays from ***** (*****) which are labeled with misplacement state of nasogastric tube (NGT) and endotracheal tube (ETT), as well as segmentation

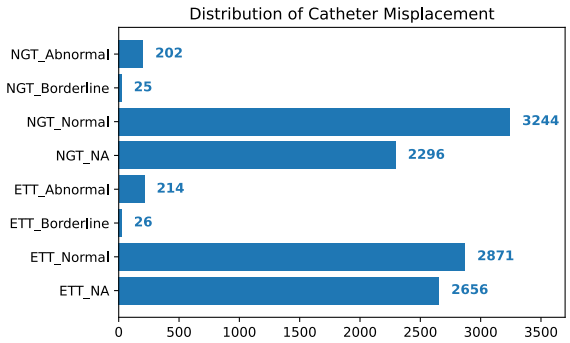

Figure 3: The distribution of catheter misplacement of ***** dataset.

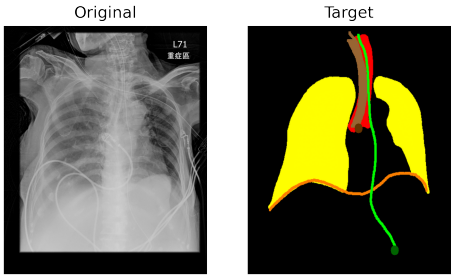

Figure 4: Sample of chest X-ray and its segmentation mask.

annotation of aforementioned catheters and relevant anatomical landmarks. Each chest X-ray is given a one-hot label of length three, where the index represents Not Applicable (NA), Normal and Abnormal respectively. NA means the catheter is not present in the chest X-ray. The label has exception for borderline cases, where instead of treating it as another class, we give a soft label of $[0, 0.5, 0.5]$ for such cases. Note that this is still a three-class classification problem for each catheter. From figure 3 we can see that there is a serious class imbalance problem.

Segmentation annotation is given in form of points, so it has to be interpolated into a line then dilated to become a segmentation mask. We also define the last point of the annotation as the tip of catheter, treating it as another channel for segmentation mask. In total there will be seven channels of segmentation mask, namely, ETT, ETT tips, trachea, NGT, NGT tips, lung and diaphragm. Since this dataset is private, a sample annotation is shown in figure 4 with image from CLiP dataset that is inferenced with our trained model.

3.2. Preprocessing

Each input image is resized to 512×512. Then each image is applied with contrast-limited adaptive histogram equalization (CLAHE) (**?**) to enhance the image. For the parameters, we took 0.02, 0.05, 0.1 of the image size as the kernel size and 0.02, 0.05, 0.1 for clip limit,

which makes nine different parameters pair. The image is applied with CLAHE nine times and stacked into a nine-channel array.

3.3. Training

Data Splitting We split the data with ratio $train : valid : test = 6 : 1 : 3$. Since abnormal cases is very rare, we split more data to testing set.

Data Sampling We use weighted random sampler as out sampler. Since this is a multi-class multi-label problem giving weight is not trivial. We give a weight of three to input that contain at least one misplaced catheter and a weight of one otherwise. We avoided giving weight to a input by state of each catheter and multiply them together since the abnormality of each catheter is independent.

Postprocessing To have a smoother segmentation loss and metric, the output and target segmentation mask will be applied with dilation with a 3×3 square.

Hyperparameters We use AdamW as our optimizer and learning rate scheduler with cosine annealing. The initial learning rate is $1e^{-4}$ and weight decay is $1e^{-5}$. We use a batch size of 16 with contain eight labeled data and eight unlabeled data,

Losses The losses we use is an aggregation of six losses. Since the shape of catheter is a line, it is considered a highly imbalanced pixel-wise. To account this problem, we use exponential logarithm loss proposed by (**?**) to combine Dice loss and our proposed RW-BCE. The four other losses are the contrastive loss, topology loss, pixel-wise focal loss and categorical cross entropy loss, which are combined linearly.

3.4. Results

We evaluate the performance of our proposed methods using the following metrics on testing set ($n = 1730$). We compute 95% confidence interval (CI) and p-value over baseline using bootstrapping ($n = 1000$).

Dice The Dice coefficient is used to measure the agreement between predicted and targeted segmentation masks. Predicted mask is binarized with threshold 0.5. A smoothing constant of $1e^{-4}$ is added to numerator and denominator to avoid zero division. The result is a mean Dice over all seven channels.

AUC This is area under receiver operating characteristic curve (AUC) to evaluate classification performance. It is computed with one-vs-rest for NA, Normal and Abnormal and then take the macro average. The result is a mean AUC over all catheters.

AP This is the average precision (AP), which is equivalent to area under precision recall curve. Since we have a heavily imbalanced data, AP is a more suitable metric to evaluate the classification performance (**?**). Is is computed for abnormal-vs-rest for each catheter and takes the average.

Table 1: Performance of different methods.

Previous Studies	Dice	AUC (95% CI)	AP (95% CI)
Ting Chen et al. (**?**)	-	0.80	-
Lakhani P. (**?**)	-	0.81	-
V. Singh et al. (**?**)	-	0.87 (0.80,0.94)	-
Milan Aryal et al. (**?**)	-	0.96	-
Marta Gherardini et al. (**?**)	0.55	-	-
Ryan Sullivan et al. (**?**)	0.74	-	-
Ours			
Baseline	0.449	0.882 (0.864,0.898)	0.137 (0.096,0.200)
TP	0.500	0.928* (0.911,0.944)	0.245 (0.170,0.352)
TP + Self-SCL	0.520	0.944* (0.927,0.961)	0.414* (0.314,0.529)
Semi-SCL	0.508	0.958* (0.943,0.971)	0.469* (0.361,0.579)
TP + Semi-SCL	**0.594**	**0.961*** (0.946,0.974)	**0.544*** (0.432,0.654)
Inference on CLiP Dataset			
Baseline	-	0.841 (0.831,0.850)	0.011 (0.009,0.017)
TP + Semi-SCL	-	**0.898*** (0.885,0.910)	**0.132*** (0.094,0.193)

TP = Topology Loss, Self-SCL = Self-supervised Contrastive Learning, Semi-SCL = Semi-supervised Contrastive Learning
* Statistically significant over baseline with p-value < 0.002

4. Discussion & Conclusion

As seen in Table 1, combining semi-supervised contrastive learning and topology loss can achieve Dice score of 0.596, AUC of 0.961 and AP of 0.544 on ***** dataset. This result is statistically significant over baseline with p-value < 0.002. Our results are comparable with metrics reported in previous studies (**??????**), but note that the dataset and task are different and therefore a direct comparison to prior work is not possible. By ablation study, each components of semi-supervised learning, contrastive learning and topology loss was shown to contribute positively to the performance. Lacking annotated data is always obstacle for deep learning development, especially for medical image. We proposed semi-supervised contrastive learning with topology loss to confront this problem for object segmentation with a known topological prior. Our method allows us to achieve state-of-the-art level performance on our publicly available CLiP dataset even without training with its label.