Real-time Duidewire Endpoint Localization in Continuous Fluoroscopy Images

Haoyun Wang:36920221153119AI,[1] Qiying Sun:36920221153113AI, [1] Sihan Liu:36920221153102AI, [1] Xiaowei Lu:36920221153105AI [1]

[1] Xiamen University
Xiamen, China

Abstract

In hepatic arterial chemoembolization (TACE: transcatheter arterial chemoembolization), real-time localization of guidewire endpoints on intraoperative images is a prerequisite for computer-assisted interventional surgery, which can reduce radiation dose, contrast dose, and operative time. At the same time, since the guide wire is a non-rigid body with a slender structure, it increases the difficulty of the endpoint localization task in noisy X-ray fluoroscopy. However, current research methods mainly focus on guidewire endpoint localization in a single fluoroscopic image, and there are still few methods for guidewire endpoint localization in serial X-ray images. Therefore, we combined the characteristics of the guide wire process in TACE surgery, replaced the branches of the Association LSTM (Long Short-Term Memory) algorithm and modified some settings, finally proposed a new framework for guide wire endpoint and state analysis in continuous X-ray images. Quantitative and qualitative analytical evaluations on a dataset consisting of 389 X-ray sequences from 48 patients show that the proposed framework significantly outperforms other existing frameworks, and the experimental results outperform previously published state-of-the-art results for this task. The proposed framework addresses the problem of real-time positioning of guidewires in continuous images, achieving state-of-the-art performance. At the same time, it is also expected to be extended to other interventional procedures, so as to realize robot-assisted interventional procedures.

Introduction

Advanced liver cancer greatly threatens human health and is the biggest killer of humans(Organization et al. 2022). However, the surgery has gradually exposed problems in safety and flexibility (Andreassi et al. 2016; Klein et al. 2015; Klein and Campos 2017; Yin et al. 2015). The keypoint localization of surgical instruments plays a weighty role in computer-assisted interventions. From the keypoint localization results, the pose of the instrument can be estimated and the use status of the instrument can also be inferred. The most important surgical instrument used in transcatheter arterial chemoembolization (TACE) is the guidewire. During the intervention, the guidewire is observed and navigated under the real-time fluoroscopy images, as shown in 1. The endpoint is the most meaningful keypoint of the guidewire, and it is of great significance to localize the guidewire endpoint in real-time.It could be applied in computer-assisted interventions to help computers understand the surgical process of TACE in real-time. To deliver the guidewire to the desired coronary artery, real-time fluoroscopy images are used for observation and navigation. As shown in 1,The two ends of the visible part of the guide wire correspond to the two end points respectively,one is the real endpoint of the guidewire and the other is the endpoint of the radiopaque material. These two points are very similar in appearance in fluoroscopy images, so how to correctly identify the real guidewire endpoint from these two points is one of the difficulties of the guidewire endpoint localization task.

Compared with the keypoint localization tasks of other surgical instruments, the challenges of guidewire endpoint localization in fluoroscopy images lie in:

- Simple appearance of the guidewire: Simple appearance means there will be more similar objects (such as rib outlines) in fluoroscopy images.
- The low contrast of the slender guidewire structure.
- Small size of visible part: Only a 3cm tip of the guidewire is with the radiopaque material coating and visible. Other parts of the guidewire are almost invisible.

Current surgical instrument key point localization algorithms mainly focus on instrument localization in a single image. However, during interventional procedures, all that is seen is continuous images. Therefore, inspired by (Lu, Lu, and Tang 2017a) and taking full advantage of the rich temporal information inherent in video data, we propose a framework for guidewire endpoint localization in continuous images, a straightforward solution is to introduce Association LSTM, which can fully learn Compared with ordinary RNN, the temporal information in video sequences can perform better in longer sequences. At the same time, we replace the SSD algorithm with the YOLOv5 algorithm. Therefore, our method is mainly composed of YOLOv5 static target detection algorithm and LSTM. Our Method is elaborately compared with other comparable keypoint localization methods on the TAGL dataset. The experimental results show that our method achieves state-of-the-art performance for guidewire endpoint localization.

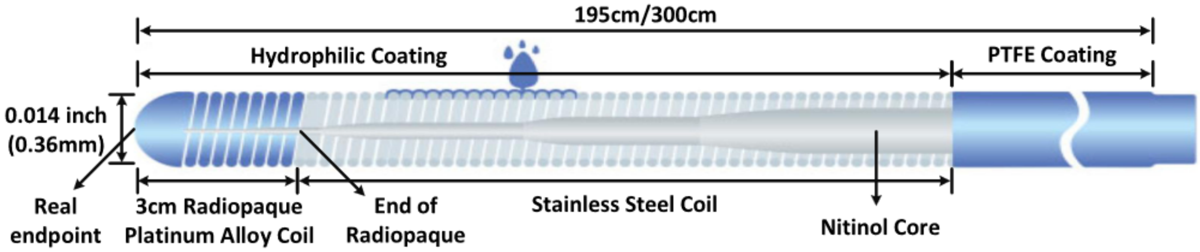

Figure 1: The structure diagram of the guidewire.

.

Related Work

Duidewire Endpoint Lolization

In recent years, there have been many studies on surgical instruments, including segmentation (Garcia-Peraza-Herrera et al. 2017; Zhou et al. 2021a), detection (Alvar and Bajić 2018a; Sarikaya, Corso, and Guru 2017; Sznitman et al. 2011; Richa et al. 2011), and also keypoint localization (Li et al. 2021; Sznitman et al. 2012; Laina et al. 2017; Rieke et al. 2016; Li et al. 2022; Kurmann et al. 2017).The guidewire endpoint localization task can be regarded as a keypoint localization task of the surgical instrument. The existing keypoint localization methods fall into two categories. The first category is the traditional computer vision algorithms (Reiter, Allen, and Zhao 2012; Sznitman, Becker, and Fua 2014; Du et al. 2018). The methods of this category extract hand-crafted features from keypoints to learn the appearance templates, and detect or track the instrument's parts using the learned templates. In recent years, with the extensive application of deep learning methods in medical images, some methods of using deep CNNs to localize the keypoints of the surgical instrument have emerged (Li et al. 2022, 2021; Zhou et al. 2021b; Du et al. 2018). Since the deep learning methods can extract the high-level semantic information in images, their localization accuracy is greatly improved compared with the traditional methods. However, there are some problems in the direct use of the existing keypoint localization methods for multi-guidewire endpoint localization: In these studies, surgeries are basically laparoscopic surgery or retinal surgery. Their surgical instruments are rigid bodies, so there is a relatively fixed positional relationship between the keypoints, which also makes some keypoint localization methods unsuitable for the guidewire endpoint localization(Du et al. 2018; Li et al. 2021).

Guidewire Detectionin in Continuous Images

Accomplishing real-time multi-guidewire endpoint localization means the localization is carried out on consecutive fluoroscopy images. Associating multiple detection results in successive frames is the mainstream way for multiple object tracking (MOT) called tracking-by-detection.In recent years, some methods have emerged to solve the object tracking task by combining the context information contained in consecutive frames with object detectors. Recurrent YOLO (Alvar and Bajić 2018b) proposed to send the output of the YOLO detector (Redmon and Farhadi 2018) to a Long Short-Term Memory (LSTM), which is used to improve the detection results by learning the context information. For Recurrent YOLO, it is necessary to prepare additional continuous frame data to train the LSTM. MV-YOLO (Alvar and Bajić 2018a) is designed for compressed-domain data, which contains Motion Vector (MV) information.The final tracking result can be obtained by combining the above results. These methods are flawed, we not only have to locate and classify the guidewire endpoints, but also associate features to represent each output object.

Method

In this project, we utilize the Association LSTM (ALSTM) model (Lu, Lu, and Tang 2017b) as the baseline for addressing sequential association tasks. The ALSTM is a variant of the Long Short-Term Memory (LSTM) network that introduces an associative memory module to store and retrieve information from past time steps. This allows the ALSTM to capture long-term dependencies in the data, improving its ability to learn complex patterns. In addition, the ALSTM incorporates a self-attention mechanism that allows the network to weight the importance of different input features at each time step, further enhancing its performance. Overall, the ALSTM offers a powerful and flexible approach for tackling sequential association tasks.

In summary, the main focus of ALSTM is to develop a long short-term memory (LSTM) network that can jointly perform object regression and object association in video object detection. To achieve this, the LSTM receives spatial information from input frames and applies the Single Shot Detector (SSD) to extract objects in the frames. The SSD produces a location-score vector for each detected object, as well as a fixed-size descriptor using region of interest (RoI) pooling. These are concatenated and stacked with past frames to form a frame-level tensor, which is fed into the LSTM network. The LSTM outputs improved predictions for the current frame with respect to the ground truth, including object locations, category scores, and association features. The network is designed to solve the object regression and object association tasks jointly, using carefully designed loss functions that consider the accuracy of object locations and class scores, as well as smoothness constraints across consecutive frames and association error based on dot products between normalized association features.

The final objective function combines the regression and

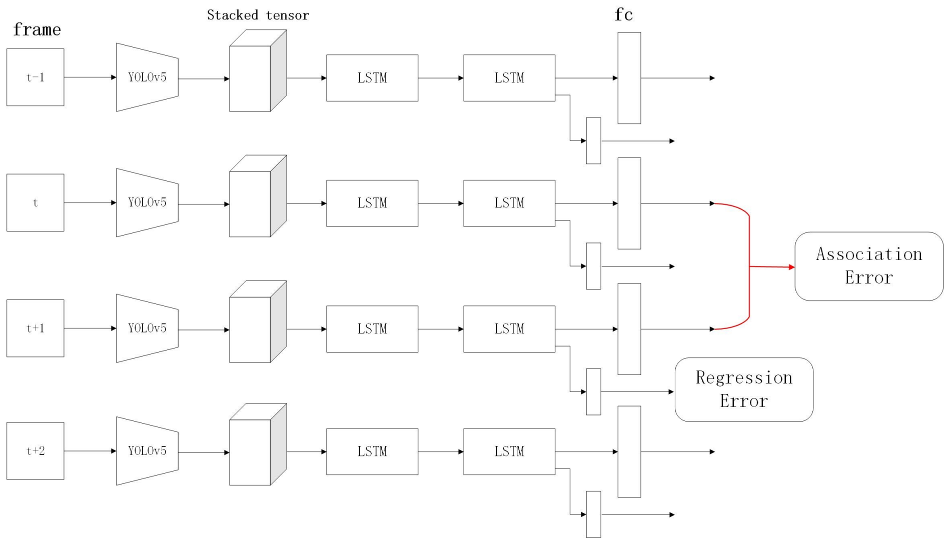

Figure 2: Our LSTM architecture in detail.

association loss functions, weighted by a hyperparameter ξ. The LSTM is trained using both fully labeled and weakly labeled datasets, and the batch normalization technique is used to accelerate the training process and normalize output features. The resulting network is able to accurately regress object locations and scores, and associate objects across multiple frames. Regression loss function consists of three components:

$$\mathcal{L}_{reg} = \sum \left(L_{conf}(c, c^*) + \lambda L_{loc}(l, g)\right) + \alpha \cdot \mathcal{L}_{smooth} \quad (1)$$

The first two terms are the object regression error, where the localization loss Llocis a smooth L1 loss between the predicted box (l) and ground truth box (g). The confidence loss Lconf is the softmax loss over multiple classes confidences c toward ground truth score vector c. The third terms is to apply the smoothness constraint across consecutive time-steps to regularize the LSTM model.

The margin contrastive loss function:

$$\mathcal{L}_{asso} = \sum_{t} \sum_{i,j} \theta_{ji} |\phi_{t-1}^{i} \cdot \phi_{t}^{j}| \quad (2)$$

where $\theta_{jk} \in 0$ 1 is an indicator, $\theta_{ji} = 1$if and only if object i in frame $t-1$ is associated with j^{th} object in frame t, they have the smallest distance among all pairs. $\cdot$ is a dot product operation.

In this paper, we improved upon the original Association LSTM (ALSTM) network by replacing the SSD detector with YOLOv5 (Yan et al. 2021)for extracting objects from the input frames. Our LSTM architecture in Figure2. YOLOv5 is a state-of-the-art object detection method that combines the strengths of multiple detection approaches to achieve high accuracy and speed performance. In our experiments, we found that using YOLOv5 for object detection in ALSTM significantly improved the performance of the network for detecting guidewire endpoints. Our results show that our modified ALSTM network outperformed the original ALSTM network in detecting guidewire endpoints.Overall, the use of YOLOv5 in ALSTM demonstrates the flexibility of the network to incorporate different object detection methods, and highlights the potential for further improvements in the performance of ALSTM for various object detection tasks.

Experiments

Datasets

To evaluate the performance of our proposed method for interventional guidewire localization in TACE (Transarterial Chemoembolization) procedures(Li et al. 2021), we have collected a dataset of fluoroscopy images taken during these procedures. The dataset, called the TACE Intervention Guidewire Localization (TIGL) dataset, contains 35 sequences from 11 subjects. All the video were collected from the same clinical center and were captured using the Siemens Artis zee III ceiling system with a flat panel detector. The frame rate of the images is approximately 8 FPS, and each image has a resolution of 512 × 512 pixels. guidewire in video were manually labeled with bounding boxes and coordinates for their two endpoints. The images were divided into a training set (18 sequences) and a testing set (17 sequences). The TIGL dataset includes images from all stages of TACE procedures, including the angiography phase, guidewire delivery phase, and balloon/stent placement phase. These images may contain various sources of interference, such as contrast agents and stents, which makes the dataset more representative of real-world conditions.

Implementation Detail

In the TACE interventional procedure, we utilized the YOLOv5 technology to detect the endpoints of the guidewire. To extract features, we utilized four feature

Table 1: Comparison Results of Three Dection Methods

Method	$\text{PCK}_5(\%)$		$\text{PCK}_7(\%)$		$\text{PCK}_9(\%)$	
	Real end	RP end	Real end	RP end	Real end	RP end
Ours	81.45	**90.26**	87.05	**92.51**	**88.53**	**92.87**
FGFA	64.89	78.14	79.68	86.48	84.08	89.13
STMM	78.12	86.73	**87.36**	88.36	85.68	91.04
MANet	**83.62**	88.32	84.97	88.73	87.59	91.28

maps.These feature maps were chosen based on their activation to high confidence objects in test sets using a pretrained YOLOv5 model. The training process involved sampling video snippets and training a two-layer LSTM model using back-propagation through time and RMSProp with a learning rate of 0.0003 and a decay rate of 0.85 for 200 epochs. The LSTM models used two-layer stateful LSTMs with 150 hidden units for state estimation and 300 hidden units for data association. The network was implemented using Pytorch and optimization was performed using the SGD optimizer. Data augmentation techniques, including random grayscale adjustment, random contrast ratio, and random scaling, were applied to the input image. The training process took about 12 hours on an NVIDIA Titan XP for 100 epochs and used all the training samples, with 200 samples randomly selected from the testing set as the validation set and the remaining 841 samples used for testing.

Evaluation Metric

In this paper, the evaluation index of the experiment is based on the OKS(Li et al. 2022) evaluation index.The evaluation metric used in this study for multi-guidewire endpoint detection is the Average Precision Percentage of Correct Keypoints (APPCK).The OKS is simplified by setting the standard deviation of each keypoint and the scale of the object as constants, and the localization is considered successful when the distance between the predicted and ground-truth keypoints is less than a threshold. Three thresholds are used in this study: 5 pixels, 7 pixels, and 9 pixels, resulting in APPCK 5, APPCK 7, and APPCK 9, respectively. This evaluation metric allows for the analysis of the localization performance on each keypoint separately and provides an intuitive way to evaluate the overall performance.

$$\mathcal{OKS}_{guidewire} = \sum_{i}^{m} \exp(-d_i^2/C), C = 2S^2K^2 \quad (3)$$

Therefore, for guidewires, the k_i in OKS can be set as a constant k_i. Besides, we assume that the scale of the guidewire is similar in different images so that s in OKS can also be set as a constant S. Since all keypoints in the dataset are visible, the δ(vi >0) can be set as 1. After the above simplifications, the OKS for guidewires will degenerate into a function inversely proportional to distance d_i.

After a series of simplifications, a threshold is set for each d_i, and when d_i is less than this threshold, the localization is considered successful, making the OKS similar to the metric named Percentage of Correct Keypoints (PCK). PCK also sets a threshold for each keypoint and reports the percentage of localization errors that less than the threshold.

Overall Results

By comparing our method with other dectetion methods on the test set, we demonstrate that our method can achieve favorable dectectino performance, as shown in Table 1. The comparison experiments are carried out in two experiment modes. Three methods are applied in comparison.

In Table1, Our method and MANet obtain similar detection results, while FGFA has the worst detection result. These detection performances are expected. However, since the guidewire detection task is not complicated, all methods achieve good detection results, and their gap is not obvious. As for localization results, Our Method achieve best detection results.

References

Alvar, S. R.; and Bajić, I. V. 2018a. MV-YOLO: Motion vector-aided tracking by semantic object detection. In *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, 1–5. IEEE.

Alvar, S. R.; and Bajić, I. V. 2018b. MV-YOLO: Motion vector-aided tracking by semantic object detection. In *2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, 1–5. IEEE.

Andreassi, M. G.; Piccaluga, E.; Guagliumi, G.; Del Greco, M.; Gaita, F.; and Picano, E. 2016. Occupational health risks in cardiac catheterization laboratory workers. *Circulation: Cardiovascular Interventions*, 9(4): e003273.

Du, X.; Kurmann, T.; Chang, P.-L.; Allan, M.; Ourselin, S.; Sznitman, R.; Kelly, J. D.; and Stoyanov, D. 2018. Articulated multi-instrument 2-D pose estimation using fully convolutional networks. *IEEE transactions on medical imaging*, 37(5): 1276–1287.

Garcia-Peraza-Herrera, L. C.; Li, W.; Fidon, L.; Gruijthuijsen, C.; Devreker, A.; Attilakos, G.; Deprest, J.; Vander Poorten, E.; Stoyanov, D.; Vercauteren, T.; et al. 2017. Toolnet: holistically-nested real-time segmentation of robotic surgical tools. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5717–5722. IEEE.

Klein, L. W.; and Campos, E. P. 2017. Occupational Hazards in the Cath Lab—Physician, Protect Thyself! *Journal of Invasive Cardiology*, 30(2).

Klein, L. W.; Tra, Y.; Garratt, K. N.; Powell, W.; Lopez-Cruz, G.; Chambers, C.; Goldstein, J. A.; for Cardiovascular Angiography, S.; and Interventions. 2015. Occupational health hazards of interventional cardiologists in the current decade: results of the 2014 SCAI membership survey. *Catheterization and Cardiovascular Interventions*, 86(5): 913–924.

Kurmann, T.; Marquez Neila, P.; Du, X.; Fua, P.; Stoyanov, D.; Wolf, S.; and Sznitman, R. 2017. Simultaneous recognition and pose estimation of instruments in minimally invasive surgery. In *International conference on medical image computing and computer-assisted intervention*, 505–513. Springer.

Laina, I.; Rieke, N.; Rupprecht, C.; Vizcaíno, J. P.; Eslami, A.; Tombari, F.; and Navab, N. 2017. Concurrent segmentation and localization for tracking of surgical instruments. In *International conference on medical image computing and computer-assisted intervention*, 664–672. Springer.

Li, R.-Q.; Xie, X.-L.; Zhou, X.-H.; Liu, S.-Q.; Ni, Z.-L.; Zhou, Y.-J.; Bian, G.-B.; and Hou, Z.-G. 2021. Real-Time Multi-Guidewire Endpoint Localization in Fluoroscopy Images. *IEEE Transactions on Medical Imaging*, PP: 1–1.

Li, R.-Q.; Xie, X.-L.; Zhou, X.-H.; Liu, S.-Q.; Ni, Z.-L.; Zhou, Y.-J.; Bian, G.-B.; and Hou, Z.-G. 2022. A Unified Framework for Multi-Guidewire Endpoint Localization in Fluoroscopy Images. *IEEE Transactions on Biomedical Engineering*, 69(4): 1406–1416.

Lu, Y.; Lu, C.; and Tang, C.-K. 2017a. Online video object detection using association LSTM. In *Proceedings of the IEEE International Conference on Computer Vision*, 2344–2352.

Lu, Y.; Lu, C.; and Tang, C.-K. 2017b. Online Video Object Detection Using Association LSTM. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2363–2371.

Organization, W. H.; et al. 2022. World health statistics 2022: monitoring health for the SDGs, sustainable development goals.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Reiter, A.; Allen, P. K.; and Zhao, T. 2012. Feature classification for tracking articulated surgical tools. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 592–600. Springer.

Richa, R.; Balicki, M.; Meisner, E.; Sznitman, R.; Taylor, R.; and Hager, G. 2011. Visual tracking of surgical tools for proximity detection in retinal surgery. In *International Conference on Information Processing in Computer-Assisted Interventions*, 55–66. Springer.

Rieke, N.; Tan, D. J.; Tombari, F.; Vizcaíno, J. P.; San Filippo, C. A. d.; Eslami, A.; and Navab, N. 2016. Real-time online adaption for robust instrument tracking and pose estimation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 422–430. Springer.

Sarikaya, D.; Corso, J. J.; and Guru, K. A. 2017. Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE transactions on medical imaging*, 36(7): 1542–1549.

Sznitman, R.; Ali, K.; Richa, R.; Taylor, R. H.; Hager, G. D.; and Fua, P. 2012. Data-driven visual tracking in retinal microsurgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 568–575. Springer.

Sznitman, R.; Basu, A.; Richa, R.; Handa, J.; Gehlbach, P.; Taylor, R. H.; Jedynak, B.; and Hager, G. D. 2011. Unified detection and tracking in retinal microsurgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1–8. Springer.

Sznitman, R.; Becker, C.; and Fua, P. 2014. Fast part-based classification for instrument detection in minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 692–699. Springer.

Yan, B.; Pan, F.; Lei, X.; Liu, Z.; and Yang, F. 2021. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing*, 13: 1619.

Yin, X.; Guo, S.; Xiao, N.; Tamiya, T.; Hirata, H.; and Ishihara, H. 2015. Safety operation consciousness realization of a MR fluids-based novel haptic interface for teleoperated catheter minimally invasive neurosurgery. *IEEE/ASME Transactions On Mechatronics*, 21(2): 1043–1054.

Zhou, Y.-J.; Liu, S.-Q.; Xie, X.-L.; Zhou, X.-H.; Wang, G.-A.; Hou, Z.-G.; Li, R.-Q.; Ni, Z.-L.; and Fan, C.-C. 2021a. A Real-Time Multi-Task Framework for Guidewire Segmentation and Endpoint Localization in Endovascular Interventions. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 13784–13790.

Zhou, Y.-J.; Xie, X.-L.; Zhou, X.-H.; Liu, S.-Q.; Bian, G.-B.; and Hou, Z.-G. 2021b. A Real-Time Multifunctional Framework for Guidewire Morphological and Positional Analysis in Interventional X-Ray Fluoroscopy. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3): 657–667.