Vein Localization on Ultrasound Images for Robotic Intravenous Insertion

Jing Zhang, Guangsheng Ma, Daniel Buckland

Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708, USA

Abstract

Robotic Intravenous (IV) Insertion system has been proposed to make catheter insertion more reliable and efficient in delivering medical fluids. Vein visual locating helps much in positioning guidance for catheter insertion, and thus is required to behave fast and accurately enough. Two deep learning models - UNet and YOLOv3 were applied to identity vein locations on ultrasound images, and both demonstrated high accuracy and efficient processing. UNet displayed more precise centroid positions of veins, whereas YOLOv3 showed more stable and stringent predictions. A combination of the two models is likely to benefit the real-time operation for robotic intravenous insertion.

Introduction

The veins appear as darker and nearly circular regions in the ultrasound images. With evident background noises, it does require humans a decision time to confirm where the veins. Traditional methods, like template matching, cannot deal with unexpected cases well and process slowly. However, convolutional neural networks have been publicly proven to behave well in recognizing objects visually. Locating veins in images could be achieved through image segmentation, like UNet or object detection, like YOLOv3.

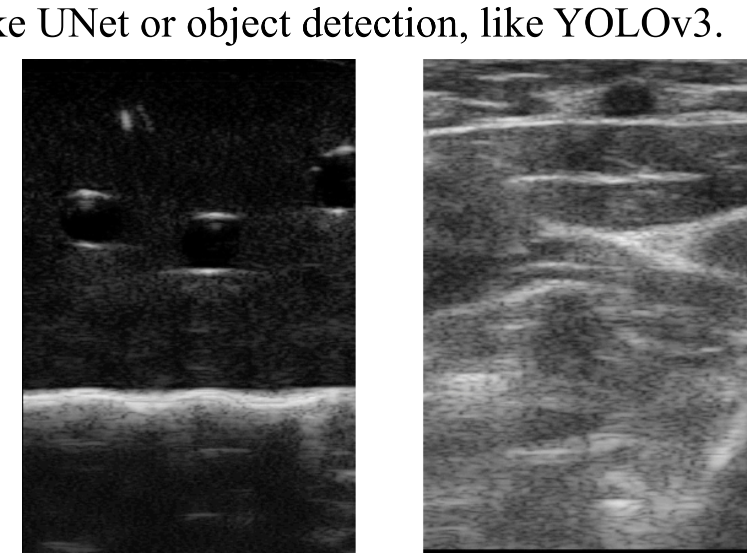


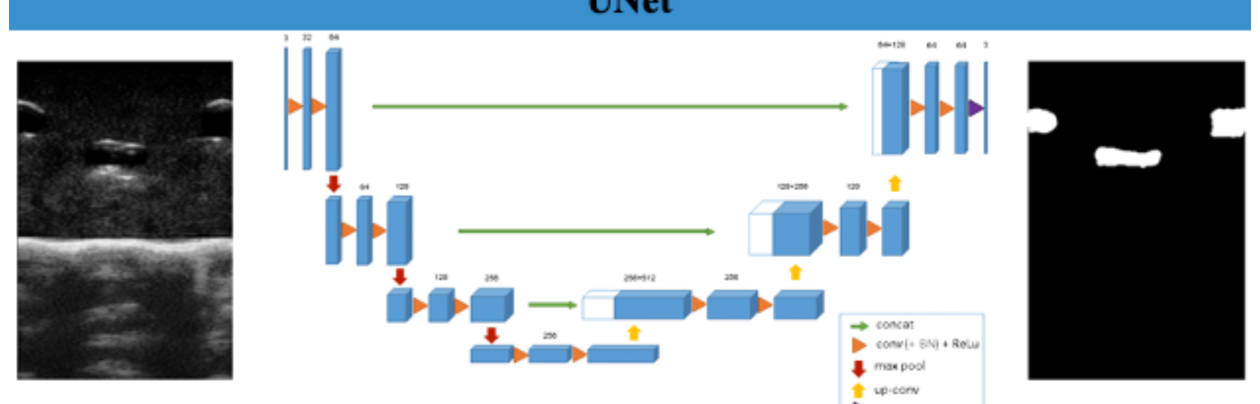
Fig1. Typical Ultrasound Images

Methods

Setup: We collected and labeled two ultrasound image (grayscale) datasets – Phantom (20 samples of 504 x 747) and Invivo (90 samples of 494 x 754) from a phantom arm and a real human arm.

UNet symmetrically combines the location information from the left downsampling path with the contextual information in the right upsampling path to predict a good segmentation map.

UNet



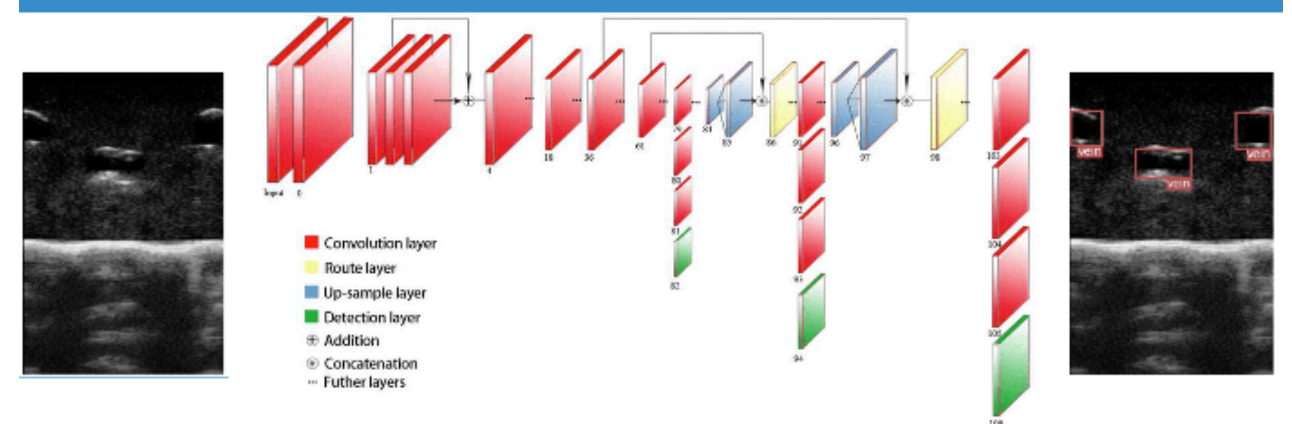
Pixel-wise Binary Cross Entropy Loss

$$\sum_{x\in\Omega} -\big(y\log(p(x)) + (1-y)\,log(1-p(x))\big)$$

Fig2. Model Structure and Training Loss Formula of UNet

YOLO (You Only Look Once) intuitively regresses from pre-defined grid cells to bounding box coordinates and class probabilities. Using Darknet-53 as the backbone convolutional neural net to extract features, YOLOv3 makes predictions at three different scales. It predicts faster, accurately, and highly generalizes to new domains.

YOLOv3



Bounding Box Coordinates MSE + Binary Cross Entropy Loss for Objectness and Classness

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2+(w_i-\hat{w}_i)^2+\left(h_i-\hat{h}_i\right)^2\right]$$
$$+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\left[\lambda_{obj}1_{ij}^{obj}\left(-log\left(p(O_i)\right)\right)+\lambda_{noobj}1_{ij}^{noobj}\left(-log(1-O_i)\right)\right]$$
$$+\sum_{i=0}^{S^2} 1_i^{obj}\,binarycrossentropy(C_i)$$

Fig3. Model Structure and Training Loss Formula of YOLOv3

Results

By evaluating the trained UNet on the Invivo dataset, we found that UNet predicts like the Normal for around 77% samples with the Worst still working well, and it also possibly takes Non-vein regions as positive. We can view these false-positive predictions as suspected veins that were not confirmed while labeling.

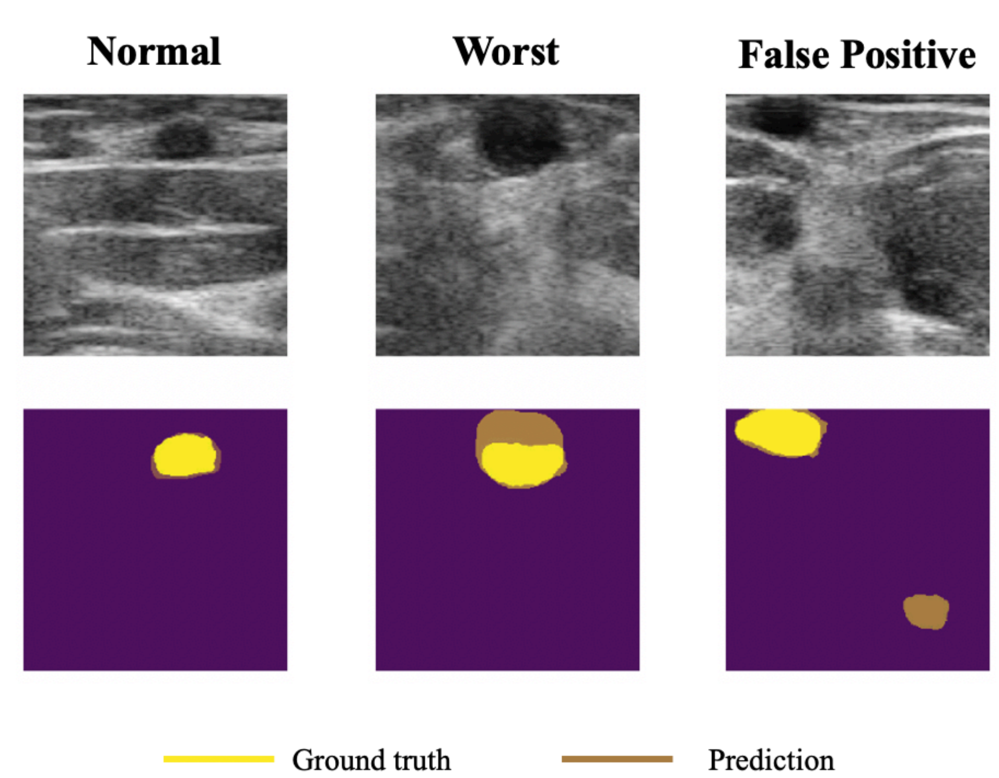


Fig4. UNet Predictions on Invivo

Note: The above samples are cropped patches around the veins for clear comparison.

Similarly, we applied the trained YOLOv3 on the Invivo dataset and found that YOLOv3 predicts exceptionally well no matter the number of veins in the image. Moreover, it achieved zero false positives and a very high AP (Average Precision) on condition that we adjust the objectness-confidence threshold and non-maximum-suppression threshold to proper values.

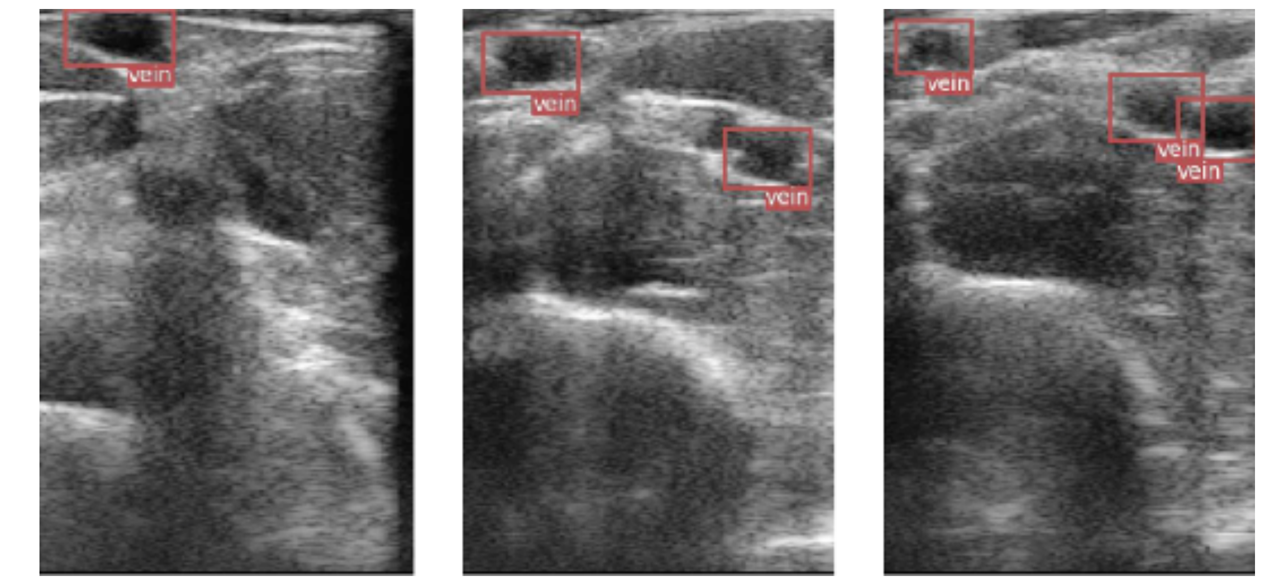


Fig5. YOLOv3 Predictions on Invivo

Evaluation Performance Statistics

Model	UNet		YOLOv3	
Dataset	Phantom	Invivo	Phantom	Invivo
Validation Score	0.975	0.969	0.780	0.793
Avg. Centroid Distance	1.995	4.218	6.736	8.132
Outliers Ratio	7.02%	5.80%	0.00%	4.26%
Avg. Prediction Speed (seconds per sample)	0.081 (GPU) 2.694 (CPU)	0.022 (GPU) 2.530 (CPU)	0.025 (GPU) 0.437 (CPU)	0.023 (GPU) 0.453 (CPU)

Fig6. Performance Comparison between UNet and YOLOv3

Note: Validation score for UNet is dice coefficient and for YOLOv3 is IOU; Centroid distance is the Euclidean distance between prediction and ground truth; Outliers ratio is out of the amount of effective predictions; the gpu mentioned above is GeForce RTX 2060.

The performance comparison on two datasets uniformly showed that the centroid prediction error of UNet is much smaller than YOLOv3 but less reliable. Their prediction speeds match up real-time processing and are quite close while using the GPU. However, when the CPU provided only, UNet needs 2.53 seconds to process a single image while YOLOv3 still can process two frames per second.

Summary

UNet and YOLOv3 both have pros and cons. UNet suggests more accurate centroid positions of veins on ultrasound images and points out all possible vein-like regions. YOLOv3 strictly predicts less false positives on the background and is a better option when we have no access to the GPU power. It is probably better to combine them into real-time operations through comprehensive coordination rather than choose one from them as the vision module of the robotic IV insertion system.

References

[1] Brewer, R.D., & Salisbury, J.K. (2010). Visual vein-finding for robotic IV insertion. *2010 IEEE International Conference on Robotics and Automation*, 4597-4602.

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv, abs/1505.04597.*

[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[4] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.