EasyChair Preprint

№ 8167

Inexact Multi-Task Learning for Fetal Anastomoses Detection

Alessandro Casella, Gaia Romana De Paolis, Elena De Momi, Dario Paladini, Sara Moccia and Leonardo Mattos

June 1, 2022

Inexact Multi-Task Learning for Fetal Anastomoses Detection

Alessandro Casella[1,2], Gaia Romana De Paolis[2], Elena De Momi[2], Dario Paladini[3], Sara Moccia[4], and Leonardo S. Mattos[1]

[1]*Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy*
[2]*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
[3]*Department of Fetal and Perinatal Medicine, Istituto "Giannina Gaslini", Genoa, Italy*
[4]*The BioRobotics Institute and Department of Excellence in Robotics & AI, Scuola Superiore Sant'Anna, Pisa, Italy*

INTRODUCTION

Twin-to-Twin Transfusion Syndrome (TTTS) is a rare pathology that may affect monochorionic twin pregnancies. TTTS depends on the unbalanced blood transfer from one twin (the donor) to the other (the recipient) through abnormal placental vascular anastomoses. Currently, the treatment for TTTS consists of the photo-ablation of abnormal anastomoses in fetoscopic laser surgery [1]. Residual anastomoses still represent a major complication [2] and their identification is not a trivial task. Visual challenges such as small field of view, amniotic fluid turbidity, low-resolution imaging, and unfavourable views are due to the position of the insertion site for the tools. To support surgeons, researchers are working on vessel and placenta segmentation [3], [4]. Recently, [5] presented the first multi-centre large-scale dataset to improve the current state-of-the-art in segmentation and registration in fetoscopy. However, to date, there is no work in the literature on anastomosis detection. There is also no available datasets for this task.

This work aims to develop a deep-learning-based framework for anastomosis detection in intra-operative fetoscopic videos from inexact labels. Considering the challenges of labelling anastomoses, we propose a weakly-supervised strategy by training a multi-task convolutional neural network (CNN) for (i) segmenting vessels in the fetoscopy frame and (ii) classifying frames as containing anastomoses or not. Relying on class activation mapping (CAM), anastomosis detection is then accomplished.

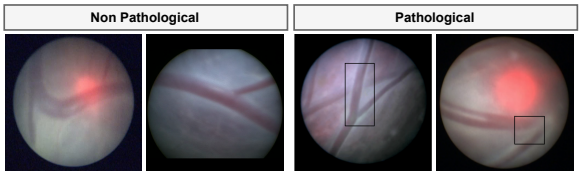

Fig. 1: A sample of the anastomoses dataset. Black bounding boxes highlight the pathological anastomosis.

MATERIALS AND METHODS

In this work, we want to exploit the representation learning capability of CNNs for inexact supervised anastomoses detection. CNNs can extract meaningful features from data, and CAM can be used for investigating the CNN visual encoding process and localising anastomoses. As shown in Fig. 2, the proposed framework is composed by a backbone and three branches. The backbone comprises (i) a dense-feature encoder and (ii) a Feature Pyramid Network (FPN) decoder. The decoder is connected to a Y-shape head consisting of two branches, for vessel segmentation and frame classification. The localisation branch extracts activation maps from the last three segmentation blocks of the encoder to perform anastomoses detection with CAM.

The encoder is based on the DenseNet121 architecture. Dense connectivity in DenseNet is implemented to improve gradients' flow among layers, avoiding the problem of gradient vanishing in deeper layers and improve network training efficiency. The decoder is designed as FPN to leverage feature hierarchy and semantics learning at different scales. All the feature maps from each pyramid level in the FPN are processed by a segmentation block to produce rough segmentation maps, one at each pyramid scale. The segmentation block contains a 3×3 convolution followed by *Group Normalisation* and *ReLU* activation. The partial segmentation maps are summed up and processed by the Y-shape head. The segmentation branch in the Y-shape head consists of 1×1 convolution for channel reduction, bilinear upsampling to recover the original size, and sigmoid activation. The classification branch is made of a Global Average Pooling (GAP) layer followed by three fully connected layers. We add one dropout layer to each of the first two layers to regularise the training process [6]. In the localisation branch, we compute the weighted sum of all Gradient-weighted Class Activation Map (Grad-CAM) for the segmentation blocks in the decoder, which represents the feature map response at each scale. The overall activation map is then rescaled using MinMax scaling, and a threshold is applied to extract regions of interest that may contain anastomoses. We experimentally found that 0.75 is a reasonable threshold value.

A. Experimental protocol

To develop and test our framework, we built a dataset by merging a publicly available dataset[1] presented in [4] that we extended with additional data collected at

[1]https://www.ucl.ac.uk/interventional-surgical-sciences/fetoscopy-placenta-data

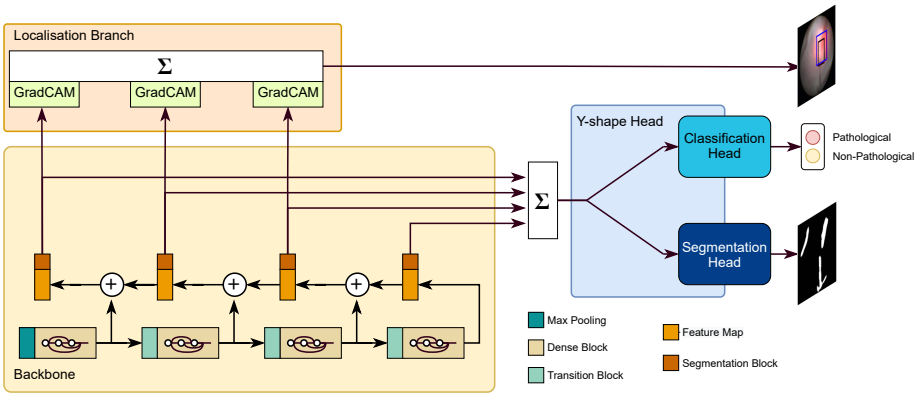

Fig. 2: Overview of the proposed WAYNet framework.

TABLE I: Results of the ablation studies for the backbone network. In the baseline the Segmentation Head is missing.

		Classification			Segmentation	Detection
		Acc	*Rec*	*Prec*	*IoU*	*IoU*
Baseline	ResNet50	0.6200	0.0570	0.7500	-	0.0424
Fine-Tuning	ResNet50	0.6827	0.5377	0.6064	**0.5770**	0.0062
	Dense	**0.7380**	**0.5849**	**0.6966**	0.5426	0.0217
Multi-Task	ResNet50	**0.9889**	0.9811	**0.9905**	**0.5976**	0.0570
	Proposed	0.9668	**0.9906**	0.9292	0.5314	**0.2206**

Istituto Giannina Gaslini, Italy for a total of 18 TTTS procedures. We manually extracted 1476 frames and asked an expert clinician to annotate vessels and the presence of pathological anastomoses in each frame. The 27.6% of frames includes pathological anastomosis. A sample of our dataset is shown in Fig. 1. For testing the detection performance, we further annotated 83 frames from 3 additional patients with bounding boxes. The overall framework was trained end-to-end using a combination (L_{Overall}) of two loss functions. L_{Overall} is defined as: $L_{\text{Overall}} = L_S + L_C$, where L_S is the binary cross entropy used for the segmentation task and L_C is the weighted cross entropy used for the classification task. We used a weighted cross entropy to account for class imbalance.

We evaluated the performance of the proposed framework in terms of accuracy (Acc), recall (Rec), and precision ($Prec$) for the classification task. As for the segmentation and detection task, intersection over union (IoU) was utilized. A Wilcoxon test assessed the statistical significance of the results.

In the ablation study, we investigated several configurations and training strategies for our framework. The **Baseline** consists of a classification network based on ResNet50. In **fine-tuning** configurations (ResNet50 and Dense-FPN), we trained the backbone on vessel segmentation and then on the classification task, while in **Multi-Task** tests (ResNet50 and Proposed), we trained the backbone on both tasks end-to-end.

RESULTS

As shown in Table I, the proposed framework achieved good performance in classification ($Acc = 0.9668$, $Rec = 0.9906$, $Prec = 0.9292$) and segmentation ($IoU = 0.5314$). For anastomoses detection we achieved an $IoU = 0.2206$. Our framework performed comparably with the other tested approaches for the classification and segmentation task, while it showed by far the best performance for the anastomosis detection task.

CONCLUSIONS AND DISCUSSION

This paper presented a first step towards the automatic localisation of anastomoses in TTTS surgical videos. Our framework showed promising results and outperformed tested approaches for weakly-supervised anastomoses detection from inexact labels. The vessel segmentation and anastomoses classification performance of our framework is reliable, with few misclassifications. Sometimes pathological and non-pathological anastomoses may look very similar, affecting classification performance, especially with such small datasets. Contrastive learning approaches might tackle this challenge. However, the high imbalance between classes could be an issue for those techniques. Collecting more data and data stratification are reasonable improvements to our experimental protocol. Despite achieving promising results, there is still work to do to tackle the complexity of the task.

REFERENCES

[1] Roberts et al. Interventions for the treatment of twin-twin transfusion syndrome. *Cochrane Database of Systematic Reviews*, 2014(1), 2014.

[2] Lopriore, et al. Residual Anastomoses After Fetoscopic Laser Surgery in Twin-to-Twin Transfusion Syndrome: Frequency, Associated Risks and Outcome. *Placenta*, 28(2-3):204–208, 2 2007.

[3] Casella, et al. A shape-constraint adversarial framework with instance-normalized spatio-temporal features for inter-fetal membrane segmentation. *Medical Image Analysis*, 70:102008, 2021.

[4] Bano, et al. Deep placental vessel segmentation for fetoscopic mosaicking. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 763–773. Springer, 2020.

[5] Bano et al. FetReg: Placental Vessel Segmentation and Registration in Fetoscopy Challenge Dataset. *ArXiv*, 2021.

[6] Lee, et al. Robust tumor localization with pyramid grad-cam. *ArXiv*, abs/1805.11393, 2018.