Medical Image Analysis 99 (2025) 103330

Contents lists available at ScienceDirect

Medical Image Analysis

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

Real-time placental vessel segmentation in fetoscopic laser surgery for Twin-to-Twin Transfusion Syndrome

Szymon Płotka [a,b,c], Tomasz Szczepański [a], Paula Szenejko [d], Przemysław Korzeniowski [a], Jesús Rodriguez Calvo [e], Asma Khalil [f], Alireza Shamshirsaz [g,h], Robert Brawura-Biskupski-Samaha [i], Ivana Išgum [b,c,j], Clara I. Sánchez [b,c], Arkadiusz Sitek [h,k,*]

[a] Sano Centre for Computational Medicine, Cracow, Poland
[b] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
[c] Department of Biomedical Engineering and Physics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[d] First Department of Obstetrics and Gynecology, The University Center for Women and Newborn Health, Medical University of Warsaw, Warsaw, Poland
[e] Fetal Medicine Unit, Obstetrics and Gynecology Division, Complutense University of Madrid, Madrid, Spain
[f] Fetal Medicine Unit, Saint George's Hospital, University of London, London, United Kingdom
[g] Maternal Fetal Care Center, Boston Children's Hospital, Boston, MA, United States of America
[h] Harvard Medical School, Boston, MA, United States of America
[i] Department of Obstetrics, Perinatology and Neonatology, The Medical Centre of Postgraduate Education, Warsaw, Poland
[j] Department of Radiology and Nuclear Medicine, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[k] Center for Advanced Medical Computing and Simulation, Massachusetts General Hospital, Boston, MA, United States of America

ARTICLE INFO

Dataset link: FetReg2021 dataset, TTTSNet data set

MSC:
41A05
41A10
65D05
65D17

Keywords:
Deep learning
Semantic segmentation
Twin-to-Twin Transfusion Syndrome (TTTS)
Fetoscopic Laser Surgery

ABSTRACT

Twin-to-Twin Transfusion Syndrome (TTTS) is a rare condition that affects about 15% of monochorionic pregnancies, in which identical twins share a single placenta. Fetoscopic laser photocoagulation (FLP) is the standard treatment for TTTS, which significantly improves the survival of fetuses. The aim of FLP is to identify abnormal connections between blood vessels and to laser ablate them in order to equalize blood supply to both fetuses. However, performing fetoscopic surgery is challenging due to limited visibility, a narrow field of view, and significant variability among patients and domains. In order to enhance the visualization of placental vessels during surgery, we propose TTTSNet, a network architecture designed for real-time and accurate placental vessel segmentation. Our network architecture incorporates a novel channel attention module and multi-scale feature fusion module to precisely segment tiny placental vessels. To address the challenges posed by FLP-specific fiberscope and amniotic sac-based artifacts, we employed novel data augmentation techniques. These techniques simulate various artifacts, including laser pointer, amniotic sac particles, and structural and optical fiber artifacts. By incorporating these simulated artifacts during training, our network architecture demonstrated robust generalizability. We trained TTTSNet on a publicly available dataset of 2060 video frames from 18 independent fetoscopic procedures and evaluated it on a multi-center external dataset of 24 in-vivo procedures with a total of 2348 video frames. Our method achieved significant performance improvements compared to state-of-the-art methods, with a mean Intersection over Union of 78.26% for all placental vessels and 73.35% for a subset of tiny placental vessels. Moreover, our method achieved 172 and 152 frames per second on an A100 GPU, and Clara AGX, respectively. This potentially opens the door to real-time application during surgical procedures. The code is publicly available at https://github.com/SanoScience/TTTSNet.

1. Introduction

Twin-to-Twin Transfusion Syndrome (TTTS) is an infrequent yet severe complication impacting around 10%–15% of monochorionic twin pregnancies (Lewi et al., 2008). It occurs when there is an imbalance in the blood flow between the twins, which can lead to serious complications and even death for both fetuses, if left untreated (Haverkamp et al., 2001). The condition is caused by the presence of abnormal blood vessel connections in the placenta, called arteriovenous anastomoses, which connect the blood circulations of both twins. These anastomoses are not present in normal monochorionic twin pregnancies but are almost always present in TTTS. The pathological blood flow imbalance results in one twin (the "recipient") receiving too much

* Corresponding author.
E-mail address: asitek@mgh.harvard.edu (A. Sitek).

https://doi.org/10.1016/j.media.2024.103330
Received 4 December 2023; Received in revised form 7 June 2024; Accepted 27 August 2024
Available online 30 August 2024

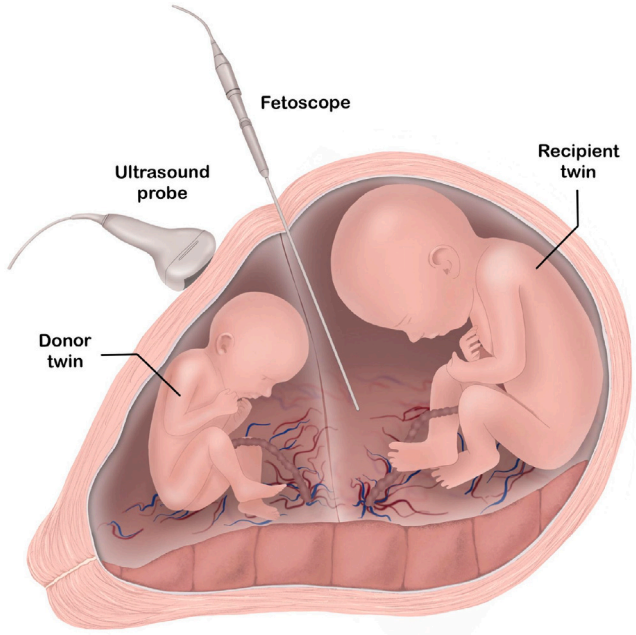

Fig. 1. An overview of FLP for TTTS. Twin fetuses, each within their own amniotic sac, are shown. The monochorionic twin pregnancy is characterized by a single shared placenta, typically with vascular connections that allow an exchange of blood between twins. A fetoscope is used to inspect the placental vessels and find pathological connections which cause an imbalance in blood exchange. When such connections are identified, they are coagulated using laser light. An ultrasound probe is typically used to guide the insertion of the fetoscope.

blood while the other twin (the "donor") receives too little (Umur et al., 2002).

Fetoscopic laser photocoagulation (FLP) is a surgical procedure that is used to treat TTTS. The procedure involves the use of ultrasound-guided insertion of a fetoscope into the amniotic sac, where the fetal surgeon identifies and ablates abnormal blood vessels in the placenta (arteriovenous anastomoses) using a laser, as shown in Fig. 1. This procedure has been shown to achieve a 70% survival rate of both twins and a survival rate of at least one fetus in more than 90% of cases (Bamberg and Hecher, 2019). The surgeon also coagulates a narrow area along the placental equator, which divides the placenta into territories that supply each twin, called the Solomon technique (Ruano et al., 2013). This minimizes the risk of a serious post-surgery complication, Twin Anemia Polycythemia Sequence (TAPS) (Bamberg and Hecher, 2019), which can lead to serious complications such as heart failure and brain injury in the affected twins.

Accurately detecting and identifying placental vessels during FLP surgery for TTTS is essential for successful outcomes and reducing the risk of surgical complications like TAPS or preterm birth (Chalouhi et al., 2011, Baschat et al., 2013). However, the visibility of the placental vessels can be hindered by the turbid environment inside the amniotic sac, poor texture visibility, low image resolution, non-planar view, particularly with anterior placenta, occlusions due to the fetus and ablation tool, and striking highlights making it difficult for the surgeon to target the abnormal blood vessels correctly. Therefore, improving the detection and identification of placental vessels would likely increase the chances of success of the treatment.

Prior studies have addressed the segmentation of placental vessels; however, these investigations relied on ex-vivo images and classical computer vision algorithms might not meet the necessary standards for clinical applications (Almoussa et al., 2011, Chang et al., 2013). In recent years, deep learning algorithms have emerged as a promising approach for the segmentation of placental vessels. Sadda et al. (2019) introduced the first deep learning-based solution using 345 in-vivo video frames from 10 TTTS surgeries, utilizing a U-Net-based network architecture. Bano et al. (2020) further explored this approach by testing various variants of U-Net and with different backbones (VGG-16, ResNet-50, and ResNet-101). They improved the segmentation by implementing a larger backbone (ResNet-101) compared to the basic U-Net. The performance and robustness of the segmentation algorithm were evaluated on 483 in-vivo video frames from 6 independent TTTS procedures.

To foster research in the field, the *FetReg2021* (Bano et al., 2021, Bano et al., 2023) dataset was released as part of the Endoscopic Vision Challenge.[1] This two-site dataset includes 2717 video frames from 24 in-vivo TTTS procedures, providing a valuable resource for developing robust and generalized models. Therefore, we utilize the *FetReg2021* dataset, maintaining its original distribution, for both training and as part of the test set to develop and evaluate our method. However, the aforementioned methods are computationally demanding, which makes them impractical for real-time application. Additionally, we noticed inconsistencies in annotations in the publicly available dataset (Bano et al., 2021, Bano et al., 2023). These inconsistencies include inaccurately delineated placental vessels at the edge of the field of view or omitted tiny placental vessels. In this work, expert fetal surgeons manually corrected them, and we investigated the impact of these corrections on the segmentation performance of placental vessels.

The research in this field is still constrained by the limited availability of comprehensive expert-annotated datasets collected from various surgical settings, essential for capturing such variability. This limitation primarily arises due to the infrequent occurrence of TTTS, making systematic data collection challenging, coupled with a shortage of annotators possessing sufficient domain expertise to ensure clinically accurate ground truth. It follows that methods mentioned in the previous paragraph were evaluated on relatively small size datasets drawn from two clinical centers. This lack of diversity may hinder their generalizability and robustness in segmentation performance on new, unseen data. To address these limitations, we summarize our contributions as follows:

1. Motivated by the need for real-time analysis and the optimal trade-off between computational efficiency and segmentation performance, we adapt state-of-the-art lightweight segmentation neural networks — DABNet and LMFFNet. Here, we modify multi-scale feature fusion and an attention mechanism to enhance placental vessel visualization during FLP for TTTS. The multi-scale approach allows the network to effectively capture both fine and coarse-grained details of the vessels while being computationally efficient and enabling real-time analysis. Additionally, the channel-attention mechanism provides the network with the ability to focus on the most important regions of the image, thereby further improving segmentation performance,
2. To address challenges associated with poor visibility within the amniotic sac environment and other artifacts, we introduce novel data augmentation approaches, mimicking laser pointer effects, amniotic sac particles, camera structural defects, and fiber artifacts,
3. To foster research in the field, we introduce and release, a novel, comprehensive, expert-annotated dataset, which includes data from four fetal medicine centers across Europe. The dataset consists of 1690 video frames from 18 in-vivo TTTS procedures. To the best of our knowledge, this is the most diverse dataset of intraoperative video frames during FLP for TTTS treatment to date,
4. We develop our method using the publicly available *FetReg2021* dataset (Bano et al., 2021, Bano et al., 2023). We identify inconsistencies in the annotations of this dataset, which are

[1] https://weiss-develop.cs.ucl.ac.uk/fetreg/

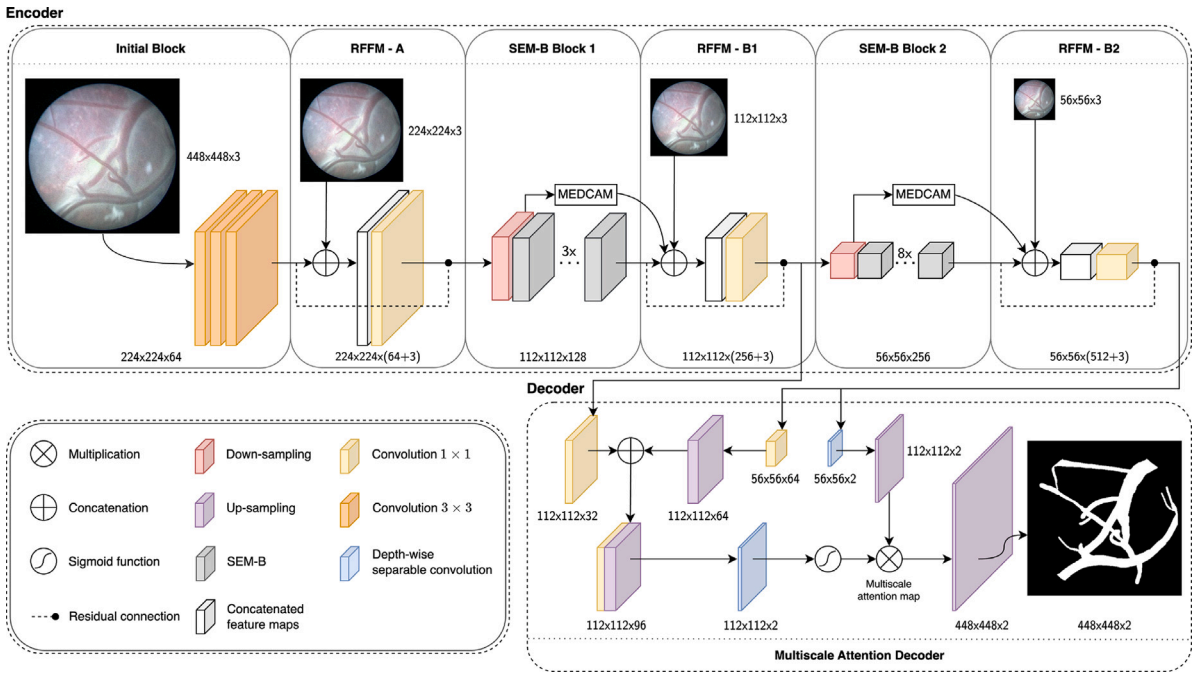

Fig. 2. An overview of the TTTSNet network architecture for real-time placental vessel segmentation during FLP for TTTS. The TTTSNet is designed as an asymmetric encoder–decoder neural network, taking a three-channel RGB input image and producing binary segmentation maps as output. In the encoder part, TTTSNet consists of the Initial Block, Residual Feature Fusion Module (RFFM) blocks, and Split-Extract-Merge Bottleneck (SEM-B) blocks, including a channel-attention mechanism called Max Pooled Channel-Attention Mechanism (MEDCAM). The encoder part allows the extraction of contextual features with low computational complexity, allowing efficient and fast processing with a few model parameters. In RFFM modules, a • preceded with the dashed line denotes residual connections, which aid the model in learning complex features without increasing the number of model parameters In the decoder part, we use the lightweight Multi-scale Attention Decoder (MAD). The MAD, with its multi-scale attention mechanism, allows the decoder to effectively recover spatial feature representation by using a minimal number of parameters.

subsequently corrected by expert fetal surgeons. We then demonstrate the enhanced generalization ability of models trained on the corrected dataset. Finally, we release these improved annotations to the community.

The rest of the paper is organized as follows. We present the method and implementation details in Section 2. The data is described in Section 3. Experimental design, and results are presented in Sections 4 and 5, respectively. We discuss the results in Section 6, and conclude the paper in Section 7.

2. Methods

This section presents our network architecture, TTTSNet, for real-time placental vessel segmentation for TTTS surgery. Our approach includes an asymmetric encoder–decoder neural network, feature fusion module, and channel-attention mechanism. Additionally, we introduce novel data augmentation approaches to increase the robustness and generalizability of the trained model against artifacts.

2.1. TTTSNet

TTTSNet is inspired by the efficient encoder design in DABNet (Li and Kim, 2019), and the lightweight Multi-scale Attention Decoder (MAD) in LMFFNet (Shi et al., 2022). Our solution adapts these methods for placental vessel segmentation through proposed Residual Feature Fusion Module (RFFM) blocks and Max Pooled Channel-Attention Mechanism (MEDCAM), designed to improve the precision for tiny vessel segmentation and facilitate model learning of complex features without increasing the number of parameters. Motivated by the lightweight design and real-time application capabilities of both DABNet and LMFFNet, our approach balances computational complexity during inference with high-quality segmentation. DABNet's encoder efficiently generates a sufficient receptive field and densely incorporates contextual information. At the same time LMFFNet's decoder precisely recovers multi-scale details of the input images through its attention mechanism. The MAD is particularly advantageous for segmenting vessels with significant size variability. During FLP, the distance from the camera to the placenta can fluctuate considerably, impacting the apparent size of the vessels within the field of view. In such cases, the fusion of features at different scales becomes highly desirable, enhancing the precision and effectiveness of the segmentation process. To segment placental vessels, we use the asymmetric encoder–decoder strategy, as it has shown promising results and a good balance between providing high segmentation accuracy and fast inference (Li et al., 2019, Li and Kim, 2019, Zhuang et al., 2021, Gao et al., 2021). An overview of TTTSNet is illustrated in Fig. 2.

2.1.1. Encoder

The encoder of TTTSNet consists of the Initial Block, Residual Feature Fusion Module (RFFM) blocks, and Split-Extract-Merge Bottleneck (SEM-B) blocks, as shown in Fig. 2.

The main objective of a feature extractor is the efficient extraction of rich and multi-scale features. It is achieved through a rapidly progressive reduction in the dimensions of feature maps coupled with multiple connections between the different levels of the architecture. This strategy is implemented using three components. Firstly, the Initial Block consisting of convolutional layers removes redundant information, reducing the original input image size by half. Secondly, RFFM blocks capture multi-scale features and context information between adjacent layers' feature map representations and combine multi-scale semantic information from different depths. They leverage also the proposed MEDCAM with max pooling operation to ensure sensitivity to small vessels. Finally, the SEM-B Blocks play a role in increasing the feature extraction efficiency by enlarging the receptive field, which allows the network to capture more context information from the input image.

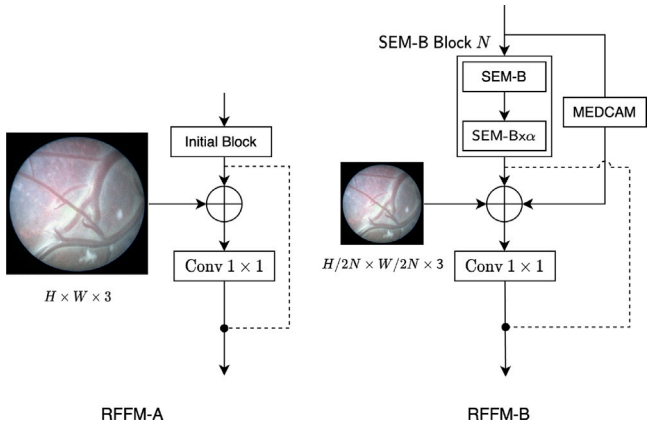

Fig. 3. Proposed RFFM preserves the identity function to aid the model in learning complex features without increasing the number of model parameters. In RFFM-A, we concatenate an identity path of input block features and process with convolution 1×1 concatenated features of the raw image and Initial Block. In RFFM-B, processed input feature maps are concatenated to the output of the SEM-Bs, down-sampled raw image, MEDCAM's output, and residual connection of input feature maps. The SEM-B Block N bounded in the dashed box comprises (α+ 1) SEM-Bs, where N corresponds to block 1 or 2.

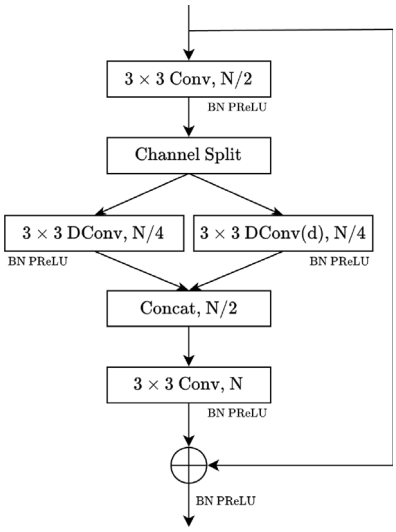

Fig. 4. Architecture of the SEM-B structure. The SEM-B starts with a 3×3 convolution (Conv) that is applied to extract feature maps and reduce the number of input channels by half. The output of this convolution is then split into two branches, consisting of a depth-wise convolution (DConv) and a depth-wise dilated convolution (DConv(d)). To fuse multi-scale feature maps, a 3×3 convolution is utilized. Batch Normalization (BN) and PReLU activation are applied after every convolutional operation. The module's output concatenates the last convolutional layer's output and the identity of the input feature map. N, and $\oplus$ denote the number of feature channels and concatenation, respectively.

The Initial Block serves as the starting point and performs key operations to lower computational complexity, reducing the resolution of feature maps. Simultaneously, removing redundant information ensures that the network focuses on the most relevant visual features for accurate placental vessel segmentation. A convolutional layer with a stride of 2 reduces the size of the input image while creating a 64-channel deep feature map

The encoder of TTTSNet includes two types of proposed RFFM presented in Fig. 3. We adapt the original FFM module with residual connection, thus creating RFFM to improve training efficiency without increasing the number of model parameters. RFFM-A block integrates the down-sampled image I_{ds_1} with the convolutional result of the Initial Block F_{IB} from the input stage and feds it through a point-wise product of 1×1 convolutional filter, expressed as:

$$F_{\text{RFFM-A}} = f_{conv_{1\times1}}(F_{\text{IB}} \oplus I_{\text{ds}_1}) + F_{\text{IB}}, \tag{1}$$

where $\oplus$ denotes concatenation operation along channels dimension and F corresponds to feature map. The second one, RFFM-B, establishes a long-range skip connection, reaching the down-sampled input image and concatenating it with the feature map, and a short-range skip connection fusing the output of the SEM-B Block N and processed with the proposed MEDCAM down-sampled output of the RFFM from the previous network stage, given by equation:

$$\begin{aligned} F_{\text{RFFM-B}_N} = f_{conv_{1\times1}}(I_{\text{ds}_{1+N}} \oplus F_{\text{SEM-B}_{B_N}} \oplus F_{\text{MEDCAM}}) \\ + F_{\text{SEM-B}_{B_N}}, \end{aligned} \tag{2}$$

where N corresponds to Blocks 1 or 2. The proposed adaptation with residual connection joins the output of one earlier layer to the input of another via summation. At the same time, intermediate operations are skipped, which, through the identity function, aids the model in learning more complex functions and reduces a vanishing gradient problem. We sum original SEM-B Block N feature maps from before concatenation with the output of the convolution operation.

The SEM-B is used to improve the feature extraction efficiency. It operates by enlarging the model's receptive field, allowing it to extract features more efficiently with fewer parameters. SEM-B Block consists of a variable number of SEM-Bs in different layers of the architecture. SEM-B Block 1 consists of three SEM-Bs and SEM-B Block 2 of eight SEM-Bs. Fig. 4 depicts an overview of the SEM-B. The 3×3 convolutions are used on both ends of SEM-B, which enlarge the receptive field of the model. Then, after the channel split, depth-wise dilated convolutions are performed. In this module, the activation function is essential to the convolution operation. Parametric Rectified Linear Unit (PReLU) (He et al., 2015) and Batch Normalization (BN) are applied in the SEM-B, which is important because the PReLU activation function typically performs better than the ReLU in lightweight networks. Moreover, the BN helps to increase the convergence speed. The module's output concatenates the last convolutional layer's output and the identity of the input feature map F_{in}, which is finally followed by the PReLU activation function. We omit PReLU activation and BN for brevity and express the result of SEM-B as:

$$F_{\text{SEM-B}_{B_N}} = f_{conv_{3\times3}}(\oplus^2 f_{Dconv_{3\times3}}(\updownarrow^2 (f_{conv_{3\times3}}(F_{\text{in}})))) + F_{\text{in}}, \tag{3}$$

where $\updownarrow^2$ denotes halving channel split operation, and $\oplus^2$ concatenation of halved in previous step channels. A novel channel-attention mechanism module called Max Pooled Channel-Attention Mechanism (MEDCAM) is proposed for improving the feature fusion ability of the RFFMs. We adapt the pooling operation with max pooling instead of average pooling in LMFFNet's PMCA module. This adjustment aims to improve precision, especially when segmenting the finest vessels. Our module prioritizes the features within significant channels while it links the down-sampled previous stage RFFM result and SEM-B Block output. Fig. 5 shows an overview of the MEDCAM module architecture. In the proposed attention module, the input feature maps F_{in} are processed with Adaptive Max Pooling $\text{AMP}_{2\times2}$, where the feature maps are partitioned into four equally sized regions, and global max pooling is applied to each. A weighted sum is applied to the partition-pooled feature vectors through learned weights of 2×2 depth-wise convolutional filters, focusing more on specific spatial partitions among channels. Finally, Squeeze Excitation (SE) Block (Hu et al., 2020) is used for dynamic channel-wise feature re-calibration of concatenated pooled-partitions with globally pooled $\text{AMP}_{1\times1}$ input features F_{in} to calculate a channel-attention vector used to multiply with the input features. The resulting modified with channel attention vector att output feature map is given as $F_{\text{out}} = F_{\text{in}} \times att(F_{\text{in}})$, where the MEDCAM att is expressed as:

$$att(x) = \text{SE}(f_{Dconv_{2\times2}}(\text{AMP}_{2\times2}(x)) + \text{AMP}_{1\times1}(x)) \tag{4}$$

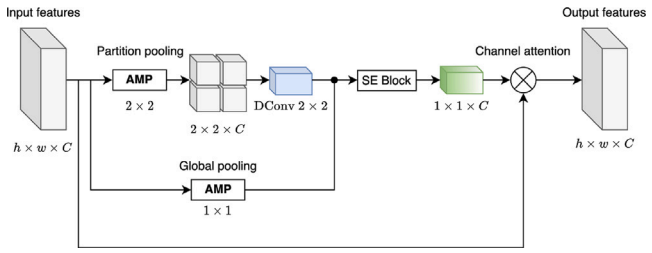

Fig. 5. The proposed MEDCAM module architecture. We utilize Adaptive Max Pooling (AMP) globally and on partitioned feature maps to leverage multi-scale features while preserving vessel details. The weighted sum is applied to the partition-pooled feature vector through learned depth-wise convolutional filters, focusing more on specific spatial partitions among channels. Squeeze Excitation (SE) Block allows for dynamic channel-wise feature re-calibration, resulting in the meaningful channel attention vector finally being applied to input features. The MEDCAM utilizes a channel attention mechanism to focus on specific feature channels and capture important information about tiny placenta vessels. h, w, C, $\otimes$ denote feature map height, and width, number of channels, and multiplication, respectively.

The MEDCAM module allows the model to pay attention to the narrow placental vessels, which are often visible in only a small area of the image. The primary objective of utilizing max pooling is to mitigate the influence of background features on segmentation outcomes, as demonstrated in the study by Nirthika et al. (2022), while simultaneously augmenting the information on the most delicate vessel fragments. In contrast, using average pooling often results in losing intricate details and smaller fragments. Although average pooling proves advantageous for generalizing natural images (Khosravan and Bagci, 2018), it is less preferred for TTTS segmentation, where the extraction of fine vessels holds greater importance. The MEDCAM utilizes an attention mechanism to focus on specific feature channels and capture important information about small vessels.

2.1.2. Decoder

The decoder in the TTTSNet consists of the Lightweight Multi-scale Attention Decoder (MAD). It leverages an attention mechanism to process multi-scale features and efficiently recover spatial details. The employed MAD combines two-scale features in one stage to refine and generate more accurate attention maps corresponding to the two RFFM-Bs in the encoder. To achieve fast inference speed, MAD aims to recover the input information by gathering low-level and high-level features with less computational complexity. The lightweight decoder with a multi-scale attention mechanism has the potential to recover the feature map's spatial details with only 0.35M parameters.

2.2. Data augmentation

We propose novel data augmentations to improve model generalization and avoid specific challenges in TTTS segmentation (Bano et al., 2021). These challenges mainly correspond to artifacts such as laser pointers, amniotic sac particles, fiber comb pattern artifacts, and fiber structural defects, as shown in Fig. 6. The following subsections outline the origin of the artifacts and the segmentation difficulties they cause. We also provide details on how augmentations mimic the artifacts.

The proposed data augmentations share the general principle of the image generation process unless otherwise stated. In the first part of this process, we build an augmentation prototype and then combine it with the input image to create the final image. Let $I : \Omega \subseteq \mathbb{Z}^2 \rightarrow \{0, 1, \ldots, 255\}$ be a single-channel gray-scale image, where Ω denotes the domain of the image, assumed to be a convex set, specifically $\Omega = \{1, \ldots, n\} \times \{1, \ldots, n\}$. The domain Ω represents discrete pixel coordinates in the image plane, with values greater than zero. Firstly, we initialize a prototype matrix $I_p : \Omega \rightarrow \mathbb{Z}$, where each element $I_p(x, y)$ represents the intensity value at pixel (x, y), initially filled with zeros. This prototype undergoes a series of operations and is finally

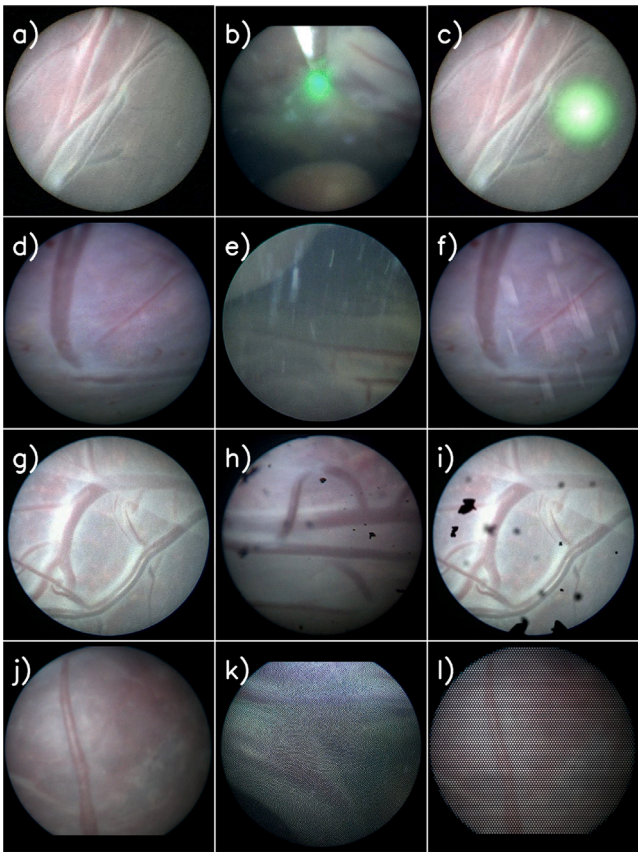

Fig. 6. A summary of the custom data augmentations is presented. Four examples of different data augmentations are shown in each row, including laser pointer, amniotic sac particles, structural defects, and optical fiber artifacts. The images depict the input image on the left, the real artifact in the middle, and the artificial artifact on the right.

merged with the input image I_{in}, yielding the augmented image I_{aug}, defined as:

$$I_{\text{aug}}(x, y) = I_{\text{in}}(x, y) + I_{\text{p}}(x, y). \tag{5}$$

Throughout the description of each of the following augmentation methods, we operate on the intensities of pixels within the field of view (FoV) of the camera, which we refer to as I_{FoV}. This set consists of pixels satisfying the relation:

$$(x - x_c)^2 + (y - y_c)^2 \leq r_{FoV}^2, \tag{6}$$

where (x_c, y_c) are the coordinates of the center of the image, (x, y) are the integer coordinates of a points in 2D space, and r_{FoV} is the radius of the FoV. Proposed augmentations to ensure fidelity to the actual artifacts and the variety of images generated include operations leveraging randomness, where random numbers come from a uniform distribution.

2.2.1. Laser pointer

During the ablation procedure, a laser is used to project a light spot onto the placenta's surface for targeting (Su et al., 2015), which can lead to local invisibility of vessels or the presence of unrealistic colors. To address this, we create augmentation that generates a realistic laser pointer (lp) effect with varying color C_{lp}, size S_{lp} and intensity I_{lp}. This augmentation is only performed after we ensure that there are no existing laser spots on the original input image. We present the final result in Fig. 6(c). Our initial step involves confirming that there are no existing laser spots in the image before applying the enhancement. We then build a prototype of the blurred laser spot fused with a randomized

overexposure effect, defined as:

$$I_{\mathrm{p}}(x,y) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} B(i,j) I_{\mathrm{lp}}(x+i, y+j), \tag{7}$$

where $I_{\mathrm{lp}}(x,y)$ represents the normalized intensity value of the laser spot image, $B(i,j)$ denotes the value of the box filter at position (i,j), and n is the size of the box filter kernel B, an odd integer. To generate a laser spot we select its center point $p_{\mathrm{c}}(x,y) \in I_{\mathrm{FOV}}$, where $I_{\mathrm{FOV}} \subseteq \Omega$ denotes pixels within the field of view of the camera. We draw a circle of size S_{l}, where $S_{\mathrm{l}} = \alpha r_{\mathrm{FOV}}$ and $\alpha \in [0.1, 0.3]$ is a random real number. The laser spot takes one out of three available colors C_{lp} by assigning a random intensity in the range of $I \in [50, 150]$ to a randomly selected channel. The intensity diminishes from the spot center, mimicking real-world behavior (Ritt, 2019; Račiukaitis et al., 2011). We determine the laser's intensity distribution based on the Distance Transform (DT), which generates a map $D : \Omega \to \mathbb{R}$ where each pixel p represents the smallest distance from the laser's spot edge pixels. The intensity map of the laser spot is defined as:

$$D(p) = \min\left\{ d(p,q) \mid q \in \mathcal{O}^c \right\}, \tag{8}$$

where $\mathcal{O}^c$ is the complement set of pixels belonging to the laser spot, i.e. pixels on its edge, and $d(p,q)$ is the Euclidean distance. Additionally, overexposure is applied to the central part of the laser spot, as commonly observed during the procedure. This effect is obtained by superimposing with probability p_{oe} of 0.3 s smaller laser spot of size $S_{l\prime} = \beta r_{FoV}$, where $\beta \in [0.7, 0.9]$ at the exact point p_{c}. Let $I : \Omega \to \mathbb{Z}$ be the laser spot image. Given a clipping parameter $c \in \mathbb{N}^+$ to achieve an overexposure effect, the normalization operation is defined as:

$$I_{\mathrm{lp}}(x,y) = \min\left(\max\left(I(x,y), 0\right), c\right). \tag{9}$$

2.2.2. Amniotic sac particles

The amniotic fluid is turbid and contains many particles, such as vernix caseosa coating the fetus's skin (Narendran et al., 2000; Akinbi et al., 2004). These particles can move freely around the amniotic sac and reflect light. Additionally, the light transmission through amniotic fluid decreases throughout gestation (Steigman et al., 2010). These factors limit visibility, causing difficulties for surgeons and the segmentation model. To simulate particles obscuring the vessels of the placenta, we implement data augmentation that mimics particles and their behavior within the field of view. The augmentation parameters, including particle shape, size, and motion blur strength, are randomly sampled from experimentally determined intervals to ensure fidelity to the original data. Fig. 6(f) illustrates the augmentation of the amniotic sac particle artifact.

The geometric shape of particles is defined based on polygons interpolated with Bezier curves. Let $P : \mathbb{R}^2 \to \mathbb{R}^2$ denote a Bezier curve function with control points P_i, where $i = 0, 1, \ldots, n$. The particle's shape S is defined as the interpolation of these curves:

$$S(t) = \sum_{i=0}^{n} B_{i,n}(t) P_i, \tag{10}$$

where $B_{i,n}(t)$ denotes the Bernstein polynomial of degree n at parameter t, given as:

$$B_{n,k}(t) = \begin{pmatrix} n \\ k \end{pmatrix} t^k (1-t)^{n-k}. \tag{11}$$

We define control points positions $P_i = \left(x_i, y_i\right)$ with $i \in \{5, \ldots, 20\}$ in the space of the prototype as:

$$P_i(t) : \begin{cases} \sum N_{i=0} x_i B_{i,N}(t), \\ \sum N_{i=0} y_i B_{i,N}(t). \end{cases} \tag{12}$$

Motion blur applied to particles is accomplished using a 2D linear filter. The filter is applied to the particle shapes in a randomly selected direction to generate streak-like effects. Let K be the 2D filter kernel of odd size n initialized with zeros. We denote the center row of K as k_{c} and create a filter that will smooth the image in the horizontal direction $k_{\mathrm{c}} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$. To achieve variability in motion blur through smoothing strength and direction, we apply an affine transformation to K with a random rotation angle α and scaling factor s. Let T denote the 2D affine transformation matrix, combining rotation and scaling:

$$T = \begin{pmatrix} \cos(\alpha)s & -\sin(\alpha)s \\ \sin(\alpha)s & \cos(\alpha)s \end{pmatrix} \tag{13}$$

Then, we apply the affine transformation T to the kernel matrix K:

$$K' = T \times K \tag{14}$$

To apply the 2D filter K' to the particles prototype I_{p}, we perform operation $I_{\mathrm{p}} = K' \circledast I$, where $\circledast$ represents convolution.

2.2.3. Structural defects

The fibers within the fiber bundle of the fiberscope are fragile and, during exploitation, they are prone to various fiber structure defects (Perperidis et al., 2020). These defects range in size from single optical fibers to groups connecting a dozen or more. The damaged fibers are unable to transmit light and appear as black spots in the image (Olivas et al., 2015) which can affect segmentation performance. An example of the effect of structural defects is depicted in Fig. 6(i).

For this augmentation, instead of adding the prototype I_{p} we subtract it from the input image I_{in} to imitate the real defects' significant drop in intensity. We simulate structural defects within the fiber bundle with dark gray or black spots based on polygons interpolated with Bezier curves. The polygons have a random number of vertices, which serve as control points for the curves. The parameter r controls the distance of the control points P_{c} from the curve's endpoints P_{e}. Specifically, it steers the distance of the control points relative to the length of the line segment between the endpoints, given as:

$$P_{\mathrm{c}} = P_{\mathrm{e}} + r(\cos(\theta), \sin(\theta)), \tag{15}$$

where r is the distance from the endpoint to the control point and θ is the angle formed between the horizontal axis and the line segment connecting the endpoint to the control point. We parameterize curves also with *edginess* parameter e, which modifies the angle θ between consecutive pairs of curve segments, defined as:

$$\theta_{\mathrm{new}} = \begin{cases} e\theta_1 + (1-e)\theta_2 + \pi & \text{if } |\theta_2 - \theta_1| > \pi, \\ e\theta_1 + (1-e)\theta_2 & \text{if } |\theta_2 - \theta_1| \leq \pi. \end{cases} \tag{16}$$

Increasing the value of e leads to sharper changes in direction along the curve, resulting in a more angular appearance. The number of defects is randomized, and their positions are randomly selected within the field of view. Gaussian blur with varying kernel sizes is applied to blur the spots and mimic out-of-focus blur. The likelihood of a strongly blurred defect is the same as that of a slightly blurred one.

2.2.4. Optical fiber artifacts

Flexible endoscopes, or fiberscopes, are utilized in modern laparoscopic surgery (Elter et al., 2006). They consist of a semi-rigid bending section that guides light through optical fibers, forming coherent bundles of flexible cores surrounded by opaque cladding (Waterhouse et al., 2018). This core-cladding relationship creates a distinctive comb structure in the image, with bright transmission points and dark borders (Winter et al., 2006), which degrades image quality (Kim et al., 2021). To mimic the image-degrading structure, our augmentation creates a hexagonal pattern. Fig. 6(l) illustrates the optical fiber artifact augmentation.

We employ Euclidean tiling with convex regular polygons to replicate the way optic fibers are packed in fiberscopes. Hexagons are used for regular tiling. Let $I_{\mathrm{in}}(x,y)$ be the intensity of the original image pixel located at coordinates (x,y) and let $\mathrm{Hex}(x_h, y_h)$ represent a regular

hexagon centered at coordinates (x_h, y_h). We define set of k vertices $V_{\text{hex}}(x, y)$ of a hexagon, as:

$$V_{\text{hex}}(x, y) : \begin{cases} \cos\left(\frac{2\pi k}{6} + \frac{\pi}{2}\right) \\ \sin\left(\frac{2\pi k}{6} + \frac{\pi}{2}\right) \end{cases} \tag{17}$$

The function $f(I_{\text{in}}, \text{Hex})$ assigns to hexagon the same color as the pixel of original input image located at its center (x_h, y_h), defined as:

$$f(I_{\text{in}}, \text{Hex}) \rightarrow I_{\text{in}}\left(x_h, y_h\right) \text{ for } (x, y) \in \text{Hex}\left(x_h, y_h\right). \tag{18}$$

The augmentation generates hexagons of various sizes based on the parameter of fiber density fd $\in [0.1, 0.3]$, which indicates the number of fibers used for image propagation to the sensor. Furthermore, a dark gradient generated with *DT* transform is applied to all the fiber edges or inside their centers to imitate the cladding comb structure.

3. Data

To develop and evaluate our method, we used a fetoscopic video dataset which consists of 4,408 frames (2,060 for training and validation and 2,348 for testing) obtained from 42 independent patients with TTTS fetoscopic procedures done in six European fetal surgery centers, namely:

1. Center A: Fetal Medicine Unit, University College London Hospital, London, United Kingdom,
2. Center B: Department of Fetal and Perinatal Medicine, Instituto "Giannina Gaslini", Genoa, Italy,
3. Center C: Department of Obstetrics, Perinatology and Neonatology, The Medical Center of Postgraduate Education, Bielański Hospital, Warsaw, Poland,
4. Center D: First Department of Obstetrics and Gynecology, The University Center for Women and Newborn Health, Medical University of Warsaw, Warsaw, Poland,
5. Center E: Fetal Medicine Unit, Obstetrics and Gynecology Division, Hospital Universitario 12 de Octubre, Complutense University of Madrid, Madrid, Spain,
6. Center F: Fetal Medicine Unit, Saint George's Hospital, University of London, London, United Kingdom.

Data from Centers A and B are publicly available within *FetReg2021* dataset (Bano et al., 2021, Bano et al., 2023), while data from Centers C-F is our in-house acquired dataset. The data were acquired from both anterior and posterior placental cases, providing a comprehensive representation of the procedure. An anterior placenta is attached to the abdominal ceiling, while a posterior placenta is attached to the back of the uterus. Overall, the dataset consists of 17 anterior placenta and 25 posterior placenta cases. Prior to usage, the dataset was anonymized in accordance with the ethical standards listed in the Helsinki Declaration. All patients provided written informed consent to use TTTS videos for research purposes.

3.1. Data acquisition

The publicly available data from Centers A and B were acquired with an original image size varying from 470 × 470 to 720 × 720 pixels (Bano et al., 2023). The data from Centers C through F were acquired with an original image size varying from 384 × 288 to 1430 × 1080 pixels. It should be noted that the frames have different sizes because the fetoscopic videos were captured at different centers with various models of fetoscope devices or light scopes. The details about the different sets are given in Table A1, and Table A2 in the Supplementary material, respectively. The data from Centers A through F were acquired by various Karl Storz GmbH (Tuttlingen, Germany) fetoscopes with an acquisition frame rate of 25 frames per second (FPS). The curved rigid 11508 AAK and straight rigid 11506 AAK fetoscopes were used to obtain data from Center A and Center B respectively. Center C and Center D employed the 11510 A fetoscopes, while Center E and Center F used the 11508 AAK, and 11530 KA fetoscopes, respectively. The data from Centers C and D were acquired by the same fetal surgeon but with different clinical settings such as support staff with fetal medicine expertise. For data from Centers C through F, we extract consecutive video frames among those in which the placental vessel is detected by expert fetal surgeons. Each video frame differs in quality, brightness, contrast, and percentage of vessels on the placental surface. We decompose the videos into frames, and the excess area outside the field of view is cropped to obtain squared images of the field of view. The dataset shows high variability in terms of image quality, noise, motion blur, the field of view size, brightness and contrast, placenta position, camera view, and stage of TTTS, as shown in Fig. 8, and Fig. 9. Table 1 shows the quantity of the dataset divided into training and test sets.

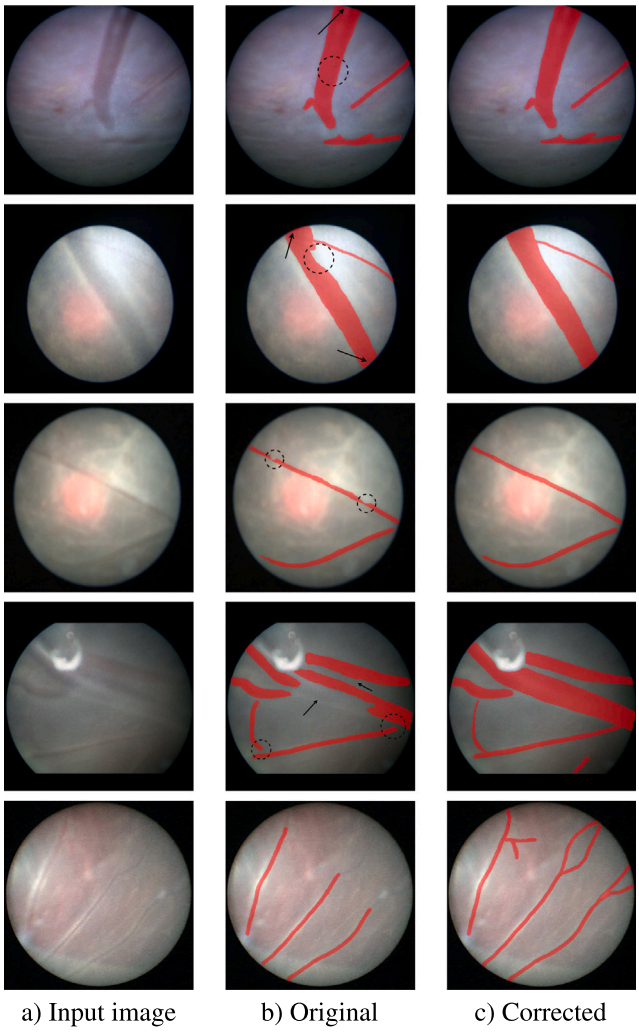

Fig. 7. Examples of corrected annotations: The input (left), original annotation (middle), and corrected annotation (right). Dotted circles emphasize inaccurate annotations, and arrows pinpoint labeling inconsistencies such as annotations beyond the field of view or not adhering to the edge. The first row illustrates an annotation that failed to fill in the gaps. In the second row, inaccurately delineated edges of the placental vessel are emphasized. The third row demonstrates the discontinuous annotation of vessels resulting from amniotic sac particle artifacts. The fourth row shows omitted big placental vessels. Lastly, the final row exhibits omitted small placental vessels.

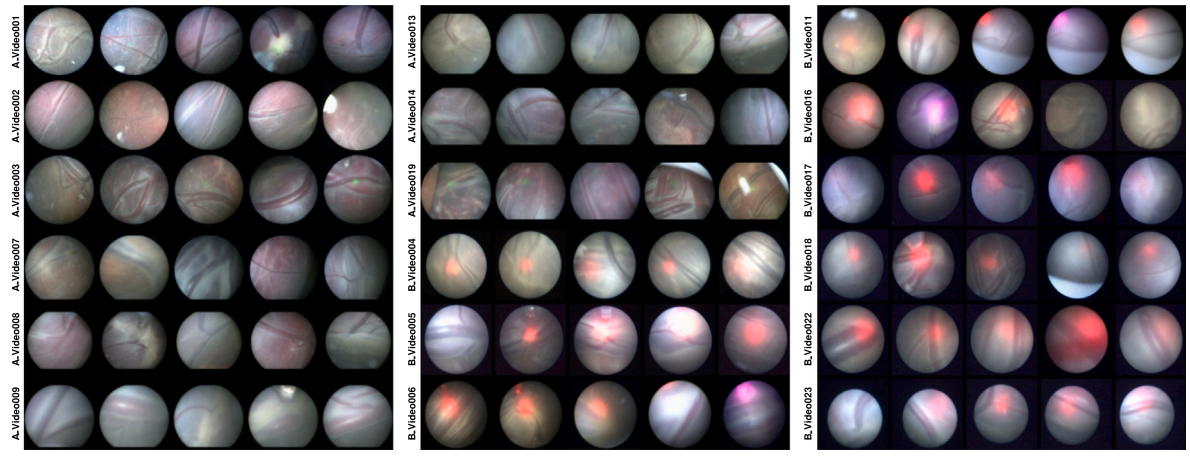

Fig. 8. Representative video frames from the training set from Center A and Center B. Each row illustrates five consecutive data samples extracted from a single video. In total 90 video frames from 18 independent in-vivo TTTS procedures are presented.

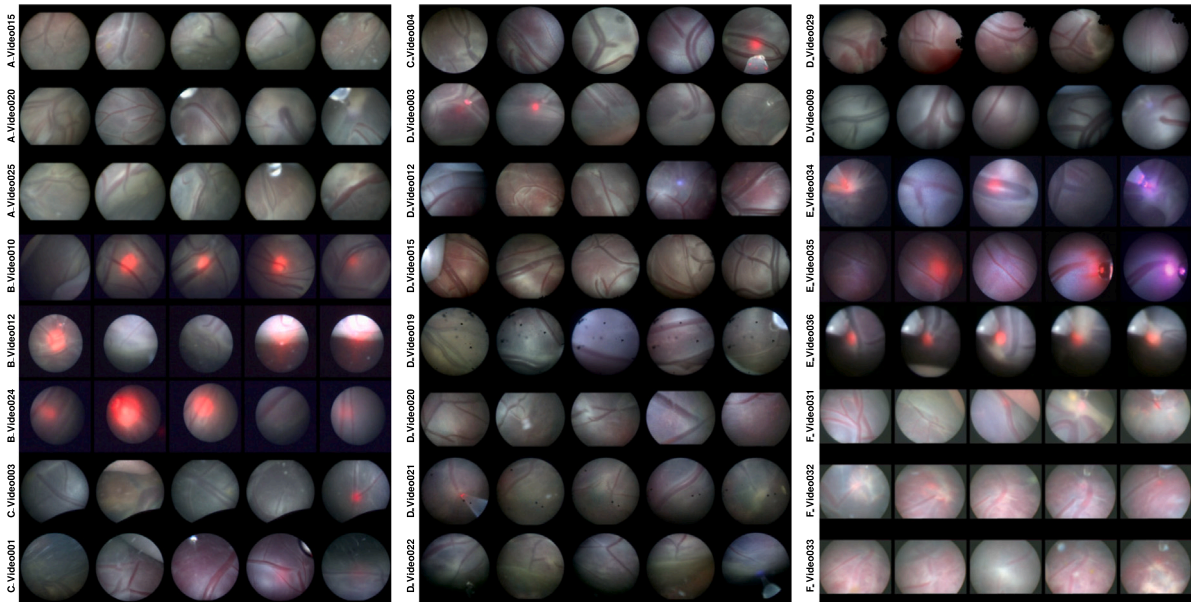

Fig. 9. Representative video frames from the test set from four centers – Center C, through Center F. Each row illustrates five consecutive data samples extracted from a single video. In total 120 video frames from 24 independent in-vivo TTTS procedures are presented.

Table 1
The total number of videos and video frames from each of the six centers used for training and testing.

No.	Center	Videos (frames)	Training videos (frames)	Test videos (frames)
1.	A	12 (1,518)	9 (1,160)	3 (358)
2.	B	12 (1,200)	9 (900)	3 (300)
3.	C	3 (300)	–	3 (300)
4.	D	9 (450)	–	9 (450)
5.	E	3 (340)	–	3 (340)
6.	F	3 (600)	–	3 (600)
	Total	42 (4,408)	18 (2,060)	24 (2,348)

3.2. Annotation protocol

The annotators segmented placental vessels adhering to the following guidelines: (1) The binary map should closely match the outline of the vessel, (2) should the object be obscured by artifacts or shadows, the vessel needs to be filled in by the annotators, (3) in cases where the placental vessel goes underneath another vessel, only the visible vessel should be annotated, (4) the vessel must remain within the field of view.

The *FetReg2021* dataset (from Centers A and B), currently publicly accessible, contains pixel-wise annotations that exclude small placental vessel segmentation and feature incomplete labels for larger vessels. Within the context of this paper, these segmentations were revised to align with our segmentation protocol. This process was executed by six junior clinicians and later reviewed by two clinical experts. To reduce inter-observer variability, each image was examined and, if necessary, rectified by one junior clinician and then reviewed by two clinical experts. Two field experts conducted quality control, and video frames were re-annotated as needed. Fig. 7 provides a selection of examples of the corrected pixel-wise annotations. We release the corrected annotations to the community to foster research in this domain. In total, 520 intraoperative video frames underwent correction. For the test set (from Centers C through F), annotations were made by a consensus of two field experts. Video frames were annotated using Supervisely, a publicly available web-based platform.[2]

[2] Online annotation tool https://supervise.ly/

3.3. Training and validation sets

We utilized a publicly available multi-center training set with corrected annotations (see Table 1). This dataset comprises 2060 pixel-wise annotated video frames obtained from 18 independent fetoscopic TTTS procedures performed in-vivo. For the training and validation sets, video frames were sourced from two distinct centers. Center A contributed 1160 video frames derived from 9 videos, while Center B provided 900 video frames obtained from 9 videos.

Although the dataset originally included four labels (background, placenta vessels, ablation tool, and fetus), we were only using the background and placenta vessel classes as our focus was on vessel segmentation. Fig. 8 shows representative examples of the video frames in the training and validation sets.

3.4. Test set

We used the multi-center test set from six European fetal medical centers, which consisted of 2348 pixel-wise annotated video frames from 24 in-vivo independent fetoscopic TTTS procedures (see Table 1). Of these, 658 video frames from 6 in-vivo procedures were sourced from FetReg2021 (Bano et al., 2023), while 1690 video frames from 18 in-vivo procedures were our in-house dataset. Fig. 9 shows representative frames from the test set.

To test our method for placental vessel segmentation, clinicians selected a diverse set of images that represent the range of variations in placental vessel appearance and structure that are typically encountered in clinical practice. The test images were selected such that the set is representative of real-world scenarios. The following criteria for selection are observed: (1) the video frame contains a placental vessel hereafter referred to as *object*, (2) the video frame (preferably) contains occluded parts, (i.e. by artifacts related to the laser pointer, optical fiber, amniotic sac particles, or structural defects), (3) the target object might be presented with different objects like *ablation tool*, and *fetus*, (4) the video frames that differ in the size of object are selected, (5) the target objects are represented in different positions, angles, and light, (6) the consecutive video frames do not represent the same *object*.

4. Experimental design

This section presents implementation details for reproducing our work, the evaluation metrics used to compare both speed and segmentation performance, and an explanation of the experimental setup that was followed.

4.1. Implementation details

We resized all video frames to 448 × 448 pixels and used them as input to the neural network. We implemented our model with PyTorch 1.11.0 (Paszke et al., 2019) on an Ubuntu workstation with 24 cores of 2.20 GHz and trained it using 2 × NVIDIA A100 80 GB GPUs and CUDA 11.3 with a mini-batch size of 16 and an initial learning rate of 1×10^{-4} with a cosine annealing learning rate scheduler (Loshchilov and Hutter, 2017) which is defined as:

$$\eta_t = \eta_{min} + \frac{1}{2}\left(\eta_{max} - \eta_{min}\right)\left(1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right), \tag{19}$$

where η_{min}, η_{max} define the range for the learning rate, T_{cur} accounts for the number of epochs performed since the last restart, and T_{max} is the max number of epochs, which is set to 300.

The neural network was initialized with default PyTorch weights, including Xavier initialization for linear, and Kaiming initialization for convolutional layers, respectively. We trained and optimized the neural network hyperparameters using six-fold cross-validation on the training set. For each fold, we used video frames from 15 videos for training and video frames from 3 videos for validation.

As a loss function we use a weighted sum of $\mathcal{L}_{dice}$ and $\mathcal{L}_{CE}$ losses, which is defined as:

$$\mathcal{L} = \mathcal{L}_{dice} + \Lambda\mathcal{L}_{CE}, \tag{20}$$

A limited grid search for optimal Λ was carried out by changing the value of Λ in the range between 0.5 and 1. We found that the network performed best if the value was 1.

To minimize the loss function $\mathcal{L}$, we employed an ADAM optimizer (Kingma and Ba, 2015) with L1 regularization of 1×10^{-5}. To handle class imbalanced data, we only counted loss for the placenta vessel class. In addition to custom augmentations, we applied default Albumentations-based (Buslaev et al., 2020) data augmentation as follows:

- *HorizontalFlip* and *VerticalFlip* with $p = 0.5$,
- *OneOf*, which select one of transform to apply, including *Blur*, limit $\in [3, 7]$, with $p = 0.25$ and *BlurMotion*, limit $\in [3, 7]$, with $p = 0.75$. *OneOf* is applied with $p = 0.2$,
- *ShiftScaleRotate* with shift limit 0.025, rotate limit = 40, scale limit = 0.2, and constant border mode with $p = 0.5$,
- *ColorJitter* with saturation = 0.2, hue = 0.15, with $p = 0.4$,
- *RandomBrightnessContrast* with brightness limit $\in [-0.15, 0.05]$, contrast limit $\in [-0.1, 0.2]$ with $p = 0.5$,
- *CLAHE* with clip limit = 1.0, tile grid size $\in [16, 16]$ with $p = 0.15$,
- *PiecewiseAffine* with scale $\in [0.004, 0.007]$, number of rows and columns = 12 with $p = 0.3$,
- *ChannelShuffle* with $p = 0.05$,
- and our four custom augmentations are as follows. *OneOf Laser pointer*, *Optical fiber artifacts*, *Amniotic sac particles*, and *Structural defects*, each applied with $p = 0.15$. *OneOf* is applied with $p = 0.5$.

After identifying the appropriate hyperparameters through six-fold cross-validation, we trained a single model using an 80% random data split for training and 20% for validation. Subsequently, we evaluated our method on an external test set. We used CometML[3] as the Machine Learning Operation (MLOps) platform to track our experiments.

4.2. Evaluation metrics

The performance of the proposed TTTSNet, along with two different configurations based on TTTSNet, was evaluated and compared to other state-of-the-art methods following the guidelines defined by the *FetReg2021* challenge, using the Intersection-over-Union (IoU) as the segmentation metric

For the analysis of statistical differences, one-way analysis of variance (ANOVA) were calculated, with $p < 0.05$ indicating statistically significant differences.

We performed an inference speed test on a 60-second video sequence. To ensure a fair comparison of inference time, synchronization between the host and device (i.e., the CPU and GPU) is utilized. This means that the time recording was only initiated after the process running on the GPU has been completed. Additionally, a GPU warm-up of 300 iterations is performed to stabilize the final results. The inference was tested on an NVIDIA Clara AGX equipped with RTX 6000 Quadro 24 GB GPU and a single NVIDIA A100 GPU to investigate the ability to deploy our method on portable devices.

4.3. Ablation studies

We conducted two ablation studies. Specifically, we investigate the following:

Impact of each key component. We validate the effectiveness of different key components in the proposed network in four configurations. First, we trained a baseline neural network. To establish a

[3] https://www.comet.com/site/

Table 2
Experimental results of ablation study with different approaches to each of key components of TTTSNet. The results of the test set are presented. The first row is the result of the baseline neural network as a part of TTTSNet, and the rest three rows refer to additional components added to the baseline.

	Baseline	RFFM	MEDCAM	mIoU (%)	p-value
Configuration 1	✓	–	–	67.54	< 0.05
Configuration 2	✓	–	✓	69.95	< 0.05
Configuration 3	✓	✓	–	70.22	< 0.05
Configuration 4	✓	✓	✓	73.08	< 0.05

baseline for comparison, we adopt DABNet. This serves as our reference method against which we evaluate our proposed approach. Second, we added Max Pooled Channel-Attention Mechanism (MEDCAM) module to the baseline. Then, we added Residual Feature Fusion Module (RFFM) to the baseline and both RFFM and MEDCAM module, respectively.

Impact of custom data augmentation. We trained a baseline neural network without using any custom data augmentations (configuration 1). We used the baseline method called TTTSNet, from the previous ablation study (i.e., Baseline + RFFM + MEDCAM). We added *Laser pointer* data augmentation to the baseline (configuration 2). In configuration 3, we added both *Laser pointer* and *Optical fiber* data augmentation to the baseline and so on (see Table 3).

4.4. Placental vessel segmentation experiments

The proposed TTTSNet network architecture was compared with eleven state-of-the-art segmentation models, including UNet (Ronneberger et al., 2015), ESNet (Wang et al., 2019), FBSNet (Gao et al., 2022), CFPNet (Lou and Loew, 2021), DABNet (baseline) (Li and Kim, 2019), UNet++ (Zhou et al., 2019), LMFFNet (Shi et al., 2022), 2D Swin UNETR (Tang et al., 2022), SwinPA-Net (Du et al., 2022), FetReg2021 top 1 performing method – Baseline (Bano et al., 2021), FetReg2021 top 2 performing method – RREB (Bhattarai et al., 2023) and two TTTSNet-based configurations: TTTSNet-S and TTTSNet*. Small TTTSNet (TTTSNet-S) is a variant of our proposed TTTSNet, trained with a reduced number of feature map scaling (32 for TTTSNet-S compared to 64 for TTTSNet). On the other hand, TTTSNet* is trained using the same parameters as the proposed TTTSNet, but it is trained on the original pixel-wise annotations provided by the *FetReg2021* challenge. For a fair comparison, we implement all methods in the same programming environment and computational settings. For training the networks, we use the same dataset (with corrected annotations) and employ the same data augmentations. The segmentation performance of the TTTSNet and other methods were compared between all centers (see Table 4).

Furthermore, to demonstrate the robustness and generalization of our method in segmenting tiny placental vessels, we conducted an additional experiment using video frames that exclusively featured tiny placental vessels. Fetal surgeons were tasked with selecting video frames containing only tiny placental vessels from data collected across each center. Clinicians selected 28, 35, 92, 68, 89, and 87 video frames from the test sets of centers A through F, respectively. In total, these selections amounted to 398 video frames, representing nearly 17% of the test set. The segmentation performance of the TTTSNet and other methods were compared between all centers (see Table 5).

Additionally, to show variability between each video sample, we performed a quantitative segmentation performance evaluation of the top 5 performing methods across each of the 24 test video samples from six centers. Moreover, we indicate the type of placenta (anterior or posterior) for each video sample to demonstrate differences in segmentation performance for both placental types (see Table 6).

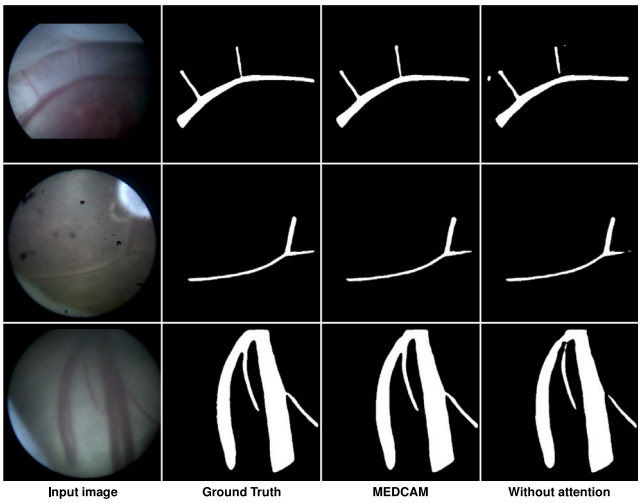

Fig. 10. A qualitative comparison of the impact of the MEDCAM module on the segmentation of placental vessels. Each row shows an example from the test set. Input image, ground truth, MEDCAM, and without attention module are presented from left to right, respectively.

4.5. Comparison with FetReg2021 challenge solutions

We compared TTTSNet with the methods from Task 1 of the FetReg2021 challenge in the following ways. Firstly, we conducted six-fold cross-validation between TTTSNet and the Baseline (Bano et al., 2020) on the FetReg2021 training dataset, maintaining the same data distribution as outlined in Bano et al. (2021). By adopting a patient-centric approach to dataset composition, we maintained consistency in the size of validation and training datasets, thereby ensuring diverse data representation in each fold. Due to the multicentric nature of the *FetReg2021* dataset, patients from all centers were included in each fold to ensure comprehensive representation. As a Baseline method, we treat UNet (Ronneberger et al., 2015) with ResNet50 (He et al., 2016) backbone with the same settings described in Bano et al. (2021). We compared both methods using original and corrected annotations. Secondly, we compared TTTSNet with the next top 5 performing methods, which include Baseline, RREB, GRECHID, SANO, and AQ-ENIB solutions. For methods lacking publicly available code, we re-implemented them to the best of our abilities, following the reproducibility descriptions provided by Bano et al. (2021).

5. Results

This section presents the results of two ablation studies conducted to demonstrate the impact of each key component in the proposed TTTSNet, along with the custom data augmentations, the results of placental vessel segmentation, and comparison with FetReg2021 challenge solutions.

5.1. Ablation studies

Table 2 presents the segmentation performance of each key component of TTTSNet in different ablated configurations. The baseline method (configuration 1) achieved an mIoU of 67.54%. The incorporation of the MEDCAM module (configuration 2) resulted in a 2.41% increase in mIoU, while the addition of the RFFM (configuration 3) to the baseline yielded a 2.68% improvement. Notably, when both modules were incorporated (configuration 4) into the baseline, they significantly outperformed the previous configurations, achieving an mIoU of 73.08%. The one-way ANOVA revealed a statistically significant difference between *configurations 1–3* and the proposed method

Table 3
Experimental results of ablation study with different approaches to custom data augmentation methods used for TTTSNet training. The results of the test set are presented. We listed five different approaches. The first row was the result of TTTSNet trained without any custom data augmentations as the baseline, and the other four rows refer to progressively adding each type of data augmentation.

	Laser pointer	Optical fiber	Amniotic sac particles	Structural defects	mIoU (%)	*p*-value
Configuration 1	–	–	–	–	73.08	< 0.05
Configuration 2	✓	–	–	–	73.87	< 0.05
Configuration 3	✓	✓	–	–	75.16	< 0.05
Configuration 4	✓	✓	✓	–	76.76	< 0.05
Configuration 5	✓	✓	✓	✓	78.26	< 0.05

Table 4
A summary of the number of parameters in millions, inference speed on both A100 GPU (GPU) and Clara AGX (Clara) hardware in FPS, and values of mIoU (%) for placental vessel segmentation obtained with different state-of-the-art methods computed using the test set. Each column shows the method, results per Center, as well as overall results. All methods were compared with the same image size of 448 × 448 pixels. The *p*-value indicates the pairwise comparison of the significance between TTTSNet and each method. The results are in order of the segmentation performance. The best results are bolded.

Method	Parameters (M)	GPU (FPS)	Clara (FPS)	Center A	Center B	Center C	Center D	Center E	Center F	Overall ↑	*p*-value
							mIoU (%)				
CFPNet	**0.55**	125.81	110.52	62.54	57.98	76.75	78.41	56.24	68.51	66.74 ± 9.42	< 0.05
FBSNet	0.61	67.52	48.51	63.95	59.54	78.56	79.15	55.25	71.42	67.98 ± 9.98	< 0.05
DABNet	0.75	205.61	186.21	65.43	60.21	78.84	81.25	59.45	69.56	69.12 ± 9.26	< 0.05
ESNet	1.66	**209.46**	**190.51**	66.06	61.25	79.43	81.25	58.43	72.44	69.81 ± 9.46	< 0.05
TTTSNet*	5.30	171.59	154.11	67.21	63.45	81.24	83.15	62.67	74.55	72.05 ± 8.93	< 0.05
LMFFNet	1.40	190.93	160.21	69.45	64.78	82.45	82.21	64.98	73.25	72.85 ± 7.98	< 0.05
UNet++	9.16	142.81	125.27	70.25	64.54	84.21	82.11	67.09	71.54	73.29 ± 8.05	< 0.05
TTTSNet-S	1.35	173.04	155.12	70.44	64.25	84.15	85.15	67.31	77.56	74.81 ± 8.81	< 0.05
UNet	31.07	87.57	65.15	72.32	67.20	83.19	83.59	68.92	75.42	75.11 ± 7.01	< 0.05
Swin UNETR	25.14	27.48	18.98	72.68	67.78	84.12	84.33	70.22	75.10	75.71 ± 7.04	< 0.05
FetReg_top2	44.01	50.78	46.32	73.18	67.56	84.22	84.15	71.16	75.62	75.98 ± 6.88	< 0.05
SwinPA-Net	117.0	18.14	13.89	73.45	67.95	84.26	85.98	71.76	76.15	76.59 ± 7.14	< 0.05
FetReg_top1	82.45	29.05	20.98	74.25	68.16	85.56	84.01	70.27	77.86	76.68 ± 7.12	< 0.05
TTTSNet	5.30	171.59	154.11	**75.01**	**70.11**	**86.15**	**86.09**	**72.08**	**80.12**	**78.26 ± 6.96**	–

Table 5
A summary of values of mIoU (%) for tiny placental vessel segmentation obtained with different state-of-the-art methods computed using the test set. Each column shows the method, results per Center, as well as overall results. All methods were compared with the same image size of 448 × 448 pixels. The *p*-value indicates the pairwise comparison of the significance between TTTSNet and each method. The results are in order of the segmentation performance. The best results are bolded.

Method	Center A	Center B	Center C	Center D	Center E	Center F	Overall ↑	*p*-value
				mIoU (%)				
CFPNet	48.75	45.47	71.75	70.23	56.84	65.78	59.81 ± 11.17	< 0.05
FBSNet	48.81	49.73	73.19	71.00	55.12	68.21	61.01 ± 11.05	< 0.05
DABNet	50.02	52.00	74.44	71.30	59.56	69.08	62.73 ± 10.37	< 0.05
ESNet	51.76	53.50	75.76	74.85	58.77	70.22	64.14 ± 10.79	< 0.05
TTTSNet*	50.22	55.49	81.24	78.78	61.59	73.67	66.83 ± 12.88	< 0.05
LMFFNet	50.75	56.10	82.15	80.20	63.73	75.70	68.11 ± 13.16	< 0.05
UNet++	55.00	57.36	77.83	77.52	70.15	72.10	68.33 ± 9.90	< 0.05
UNet	56.69	56.35	79.45	76.40	71.80	73.54	69.04 ± 10.04	< 0.05
Swin UNETR	56.28	57.04	79.49	80.16	70.15	71.78	69.15 ± 10.47	< 0.05
SwinPA-Net	56.85	56.59	79.55	80.42	71.60	70.68	69.28 ± 10.51	< 0.05
TTTSNet-S	56.37	55.47	81.92	80.59	68.50	74.08	69.49 ± 11.56	< 0.05
FetReg_top2	57.45	59.44	80.64	82.87	72.36	74.62	71.15 ± 10.53	< 0.05
FetReg_top1	58.07	59.89	81.04	83.34	72.87	75.60	71.80 ± 10.63	< 0.05
TTTSNet	**59.90**	**60.40**	**82.70**	**84.26**	**74.35**	**78.49**	**73.35 ± 10.79**	–

(configuration 4), which significantly outperformed the other evaluated configurations.

The segmentation results of the five evaluated configurations, which are associated with different approaches to custom data augmentations, are listed in Table 3. The baseline method (TTTSNet) without any custom data augmentations (configuration 1) achieved an mIoU of 73.08%. The incorporation of the *Laser pointer* augmentation (configuration 2) into the baseline improved segmentation performance by 0.79%. Furthermore, the addition of the *Optical fiber* augmentation (configuration 3) increased the mIoU by 1.29% compared to the previous configuration. The introduction of *Amniotic sac particles* and *Structural defects* as augmentations (configurations 4 and 5) further improved the mIoU by 1.6% and 3.1%, respectively. Overall, the TTTSNet trained with all augmentations outperforms the TTTSNet without any augmentations by 5.18% using mIoU as an evaluation metric.

To further analyze the effect of the attention module in highlighting tiny placental vessels, we visualize the probability maps generated by TTTSNet with and without the MEDCAM module, as shown in Fig. 10. When MEDCAM is not applied, the probability map exhibits inaccuracies and is susceptible to artifacts. Conversely, when MEDCAM is applied, the probability map effectively emphasizes both thick and thin placental vessels, remains robust in the presence of artifacts, consistently aligns with the ground truth, and closely resembles it.

5.2. Placental vessel segmentation

Table 4 presents a quantitative comparison of results among our proposed TTTSNet, eleven state-of-the-art methods, and two TTTSNet-based configurations. This comparison includes the number of model parameters, speed on two different devices, and segmentation performance across all evaluated centers. The proposed TTTSNet has a

Table 6
A summary of values of mIoU (%) ± standard deviation for TTTSNet and the next top 5 performing methods per each video from the test set. Each column shows the video name, Center, type of placenta and method. All methods were compared with the same image size of 448 × 448 pixels and the same training settings, i.e. data augmentations. The results are in order of the segmentation performance (the best on the left). The best results are bolded.

No.	Video name	Center	Type of placenta	TTTSNet	FetReg_top1	SwinPA-Net	FetReg_top2	Swin UNETR	UNet
1.	B_Video010	B	anterior	**75.57 ± 8.00**	72.25 ± 9.10	71.98 ± 9.06	71.66 ± 9.05	71.81 ± 9.20	71.73 ± 8.51
2.	B_Video012	B	anterior	**61.98 ± 13.11**	61.75 ± 12.77	61.53 ± 12.88	61.13 ± 12.64	61.37 ± 12.66	60.95 ± 12.67
3.	A_Video015	A	anterior	**75.14 ± 8.35**	74.73 ± 11.16	74.13 ± 11.57	71.44 ± 11.26	74.45 ± 9.28	71.92 ± 9.40
4.	A_Video020	A	posterior	**74.25 ± 8.10**	73.18 ± 9.08	72.26 ± 9.42	71.85 ± 9.38	71.63 ± 9.34	71.56 ± 8.91
5.	B_Video024	B	posterior	**72.77 ± 10.12**	70.49 ± 8.69	70.33 ± 8.81	69.89 ± 8.91	70.13 ± 9.06	68.93 ± 8.60
6.	A_Video025	A	posterior	**75.69 ± 7.71**	75.39 ± 7.29	74.02 ± 7.32	74.37 ± 7.27	73.37 ± 7.38	73.92 ± 7.150
7.	C_Video001	C	posterior	**83.07 ± 9.85**	82.44 ± 9.83	80.98 ± 9.94	80.62 ± 9.76	80.53 ± 9.76	80.52 ± 8.81
8.	C_Video003	C	anterior	**87.62 ± 5.32**	87.18 ± 5.34	85.82 ± 5.50	86.11 ± 5.52	86.03 ± 5.52	83.90 ± 4.52
9.	C_Video004	C	posterior	**87.75 ± 8.49**	87.02 ± 8.13	85.97 ± 8.21	85.92 ± 8.08	85.81 ± 8.06	85.14 ± 8.38
10.	D_Video006	D	anterior	**89.70 ± 12.98**	87.38 ± 13.11	88.32 ± 11.78	86.74 ± 11.62	87.02 ± 11.77	85.04 ± 11.59
11.	D_Video009	D	anterior	82.18 ± 13.75	80.70 ± 13.56	**82.84 ± 12.75**	80.10 ± 12.68	81.38 ± 12.51	79.06 ± 13.85
12.	D_Video012	D	posterior	**82.78 ± 9.09**	80.84 ± 9.00	82.52 ± 8.30	79.76 ± 8.34	81.70 ± 9.19	79.38 ± 7.65
13.	D_Video015	D	posterior	**83.84 ± 10.26**	81.16 ± 10.16	82.82 ± 9.78	81.94 ± 10.02	81.50 ± 10.05	81.08 ± 10.59
14.	D_Video019	D	posterior	**87.52 ± 4.28**	82.22 ± 6.27	85.10 ± 5.97	84.14 ± 6.13	83.70 ± 5.65	85.70 ± 3.45
15.	D_Video020	D	anterior	**85.16 ± 7.57**	82.02 ± 8.41	84.40 ± 6.86	82.76 ± 7.10	83.96 ± 7.49	83.18 ± 6.89
16.	D_Video021	D	posterior	85.06 ± 8.00	83.96 ± 7.47	**86.70 ± 6.60**	85.14 ± 6.44	84.16 ± 6.52	82.80 ± 8.25
17.	D_Video022	D	anterior	**91.76 ± 7.97**	91.50 ± 4.29	91.20 ± 5.31	90.62 ± 5.29	88.90 ± 4.95	90.68 ± 8.01
18.	D_Video029	D	posterior	86.90 ± 10.24	86.32 ± 8.59	88.92 ± 7.58	86.14 ± 8.03	86.64 ± 8.10	85.38 ± 9.92
19.	E_Video034	E	posterior	69.50 ± 11.15	68.91 ± 12.31	**70.26 ± 12.26**	69.84 ± 12.21	69.09 ± 12.55	67.24 ± 12.27
20.	E_Video035	E	posterior	**71.28 ± 10.89**	68.62 ± 11.87	69.75 ± 11.55	69.31 ± 11.69	67.74 ± 12.35	67.43 ± 12.32
21.	E_Video036	E	anterior	75.78 ± 5.09	73.93 ± 6.12	**76.06 ± 5.99**	75.08 ± 5.64	74.83 ± 6.30	72.69 ± 6.39
22.	F_Video031	F	posterior	**72.37 ± 5.80**	70.11 ± 7.05	67.71 ± 7.09	67.31 ± 7.14	67.04 ± 7.37	69.35 ± 7.08
23.	F_Video032	F	posterior	**82.74 ± 9.80**	81.04 ± 9.02	79.72 ± 9.30	79.30 ± 9.42	78.48 ± 9.27	77.40 ± 9.65
24.	F_Video033	F	anterior	**85.23 ± 8.90**	82.45 ± 8.65	81.06 ± 9.03	80.26 ± 9.35	79.79 ± 9.53	79.52 ± 10.01

Table 7
Results of six-fold cross-validation for both baseline and TTTSNet methods on original and corrected annotations. Vessel Original and Vessel Corrected classes are abbreviated as VO and VC, respectively.

Video	Center	$VO_{Baseline}$	$VC_{Baseline}$	$VO_{TTTSNet}$	$VC_{TTTSNet}$	Fold	Images per fold	Type of placenta
Video001	A	86.1	89.5	87.9	90.2			posterior
Video006	B	69.4	74.8	73.2	75.8	1	352	posterior
Video016	B	85.2	88.9	87.3	89.8			posterior
Video002	A	81.0	85.2	83.2	87.6			posterior
Video011	B	73.2	77.1	74.9	78.9	2	353	posterior
Video018	B	80.9	84.4	83.1	86.4			anterior
Video004	B	81.1	85.2	83.3	86.8			posterior
Video019	A	78.8	82.4	80.8	83.6	3	349	posterior
Video023	B	82.9	86.8	84.4	87.7			anterior
Video003	A	82.1	86.9	83.7	89.5			posterior
Video005	B	77.7	82.1	79.8	83.2	4	327	anterior
Video014	A	84.3	88.9	85.2	87.9			anterior
Video007	A	78.2	82.4	79.1	83.2			anterior
Video008	A	76.4	80.6	77.2	81.8	5	350	anterior
Video022	B	81.8	87.4	83.3	87.6			posterior
Video009	A	80.7	85.3	83.0	86.2			anterior
Video013	A	78.3	82.8	79.9	84.4	6	329	anterior
Video017	B	67.9	70.6	70.5	74.1			posterior
		79.22 ± 4.98	83.41 ± 5.08	81.10 ± 4.71	84.71 ± 4.63			

Table 8
A summary of the values of mIoU (%) for TTTSNet and the top 5 performing methods from the FetReg2021 challenge. We provide placental vessel segmentation performance for each Center, as well as overall results for both original and corrected annotations. All methods were compared using the same image size of 448 × 448 pixels and the same training settings, i.e., data augmentations. The results are ordered by segmentation performance, with the best results bolded.

Method	Original annotations							Corrected annotations						
	A	B	C	D	E	F	Overall ↑	A	B	C	D	E	F	Overall ↑
AQ-ENIB	52.63	49.42	68.72	68.72	51.07	56.44	57.83 ± 8.75	60.00	59.86	74.12	76.12	62.56	67.27	66.65 ± 7.11
SANO	58.81	52.14	72.53	73.18	55.13	61.95	62.29 ± 8.83	67.10	65.40	78.62	81.45	68.67	72.08	72.22 ± 6.50
GRECHID	65.92	60.98	79.06	79.45	60.82	67.15	68.90 ± 8.42	72.58	67.60	83.90	83.75	71.20	75.95	75.83 ± 6.75
RREB	66.04	61.84	79.68	80.12	58.11	70.50	69.38 ± 9.14	73.18	67.56	84.22	84.15	71.16	75.62	75.98 ± 6.88
Baseline	66.48	62.21	80.45	81.51	59.48	71.15	70.21 ± 9.24	74.25	68.16	85.56	84.01	70.27	77.86	76.68 ± 7.12
TTTSNet	**67.21**	**63.45**	**81.24**	**83.15**	**62.67**	**74.55**	**72.05 ± 8.93**	**75.01**	**70.11**	**86.15**	**86.09**	**72.08**	**80.12**	**78.26 ± 6.96**

total of 5.3 million parameters and operates at speeds of 154.11 and 171.11 FPS on the two hardware configurations. For the test data from each of the six participating centers, our proposed TTTSNet achieves an overall mean IoU of 78.26%. The IoU values range from 70.11% to 86.15% across all centers. Notably, our TTTSNet demonstrates the best segmentation performance among the compared methods, with an average IoU score improvement of 1.58% over the second-best FetReg top 1 performing solution–Baseline (Bano et al., 2020). Particularly noteworthy is that our method outperforms the other segmentation

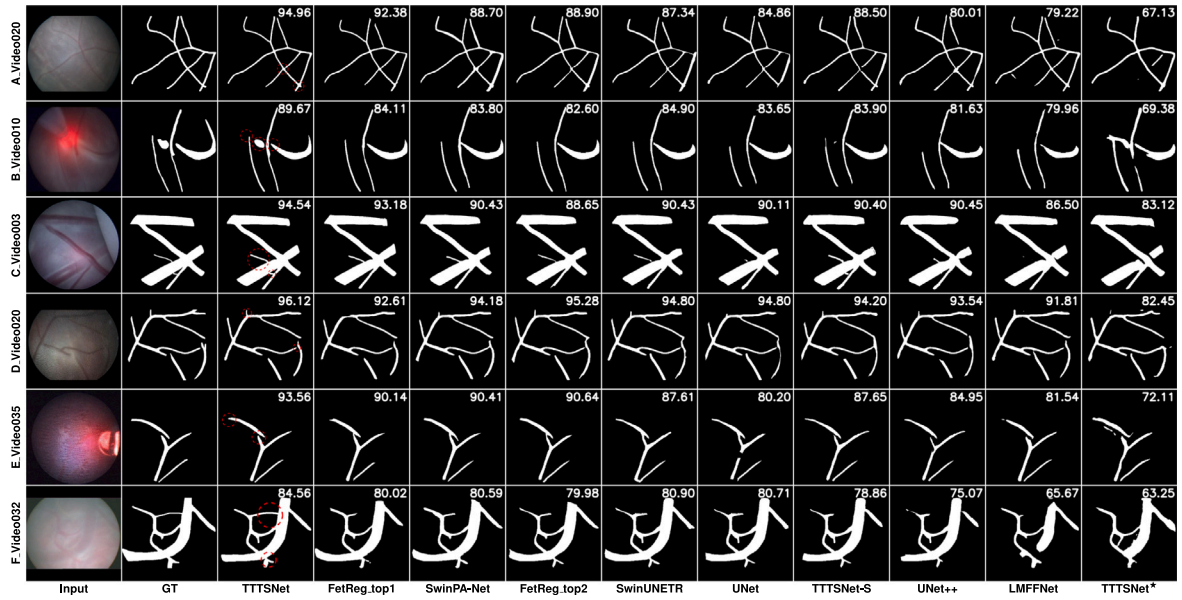

Fig. 11. Examples of segmentation results obtained on the test set by our proposed TTTSNet model, compared with several state-of-the-art methods and two TTTSNet-based configurations. Ground truth is abbreviated as GT. TTTSNet* denotes TTTSNet trained on original pixel-wise annotations provided by the *FetReg2021* challenge. The images are arranged in order of the best overall score, with the best results on the left. Each row corresponds to a different video, and each column shows the input image, ground truth, results of TTTSNet, and results of other state-of-the-art methods.

methods by a statistically significant margin across all evaluated centers (with $p < 0.05$).

Table 5 provides a quantitative evaluation of our proposed TTTSNet, eleven state-of-the-art methods, and two TTTSNet-based configurations. This assessment focuses specifically on the segmentation performance within a subset of the test dataset, which exclusively contains video frames featuring tiny placental vessels. Within this subset, our TTTSNet demonstrates strong segmentation performance, yielding an overall mean IoU score of 73.35% across all six participating centers. The IoU values range from 58.43% to 84.26% across these centers. Notably, our TTTSNet stands out as the top-performing method in this context, achieving an average IoU score improvement of 1.55% over the second-best FetReg top 1 performing solution – Baseline. Importantly, our method consistently outperforms several segmentation approaches across all assessed centers, with statistical significance (with $p < 0.005$).

Table 6 presents a quantitative comparison of the results among our proposed TTTS and the next top 5 performing methods for each video in the test set along six centers. Our TTTSNet achieves an mIoU for the anterior placenta ranging from 61.98 to 91.76%, with a mean of 81.01 ± 8.97%. For the posterior placenta, the mIoU ranges from 71.28 to 87.52%, with a mean of 79.68 ± 6.65%.

Fig. 11 presents a qualitative comparison of segmentation results among ground truth, our proposed TTTSNet, several state-of-the-art methods, and two TTTSNet-based configurations. This figure illustrates that TTTSNet consistently delivers accurate segmentation for both thick and thin placental vessels. In contrast, other methods encounter difficulties with thin placental vessels and regions affected by artifacts. Our approach successfully achieves accurate segmentation in these challenging regions, as evident in the qualitative comparison.

5.3. Comparison with FetReg2021 challenge solutions

Table 7 presents a comparison of results between TTTSNet and the Baseline (Bano et al., 2020), utilizing six-fold cross-validation on the training FetReg2021 dataset with both original and corrected annotations. TTTSNet achieved mIoU of 81.10 ± 4.71 and 84.71 ± 4.63 compared to 79.22 ± 4.98 and 83.41 ± 5.08 for the Baseline, for original and corrected annotations, respectively. In both scenarios, our proposed TTTSNet outperformed the Baseline method. Overall, both compared methods demonstrated stable performance across all folds, although limitations were observed in certain challenging scenarios.

Table 8 shows a comparison of results between TTTSNet and the top 5 performing methods of the FetReg2021 challenge. For comparison, we used both original and corrected annotations. In both scenarios, our TTTSNet outperformed the next best method—RREB by 2.67% and 1.28% for original and corrected annotations, respectively.

6. Discussion

Our proposed network architecture, TTTSNet, is designed to address the challenging task of segmenting placental vessels within frames extracted from video feeds obtained during FLP surgery, a procedure aimed at treating TTTS. Specifically, we adapt both state-of-the-art lightweight segmentation networks, DABNet and LMFFNet, by modifying the multi-scale feature fusion module and the attention mechanism to enhance placental vessel segmentation. Our method demonstrated superior performance compared to eleven state-of-the-art methods, achieving an mIoU of 78.26% in placental vessel segmentation. Notably, TTTSNet achieves an inference speed of 150 FPS, enabling real-time interactive usage and potentially facilitating prompt decision-making by surgeons.

We assessed the generalizability of our method by testing it on data from multiple centers. This multi-center placental vessel segmentation study is by far the most extensive to date, involving 24 in-vivo videos (comprising 2,348 meticulously annotated frames) from six different European institutions. 18 of these video samples are novel in-house test set from four centers, and remaining 6 coming from two centers are already publicly available within *FetReg2021*. We observed that the TTTSNet segmentation performance differs across data from different centers. These differences are likely due to several reasons. The multi-center test set was acquired by fetal surgeons with varying levels of clinical experience, each employing different techniques for performing the TTTS procedure. In addition, videos vary in the acquisition quality, which is influenced by different equipment used during surgery (see Fig. 9) and different sizes of fields of view (Table A1 and Table A2 in the Supplementary material). Moreover, we have used the publicly available data from Center A, Center B, and our in-house acquired

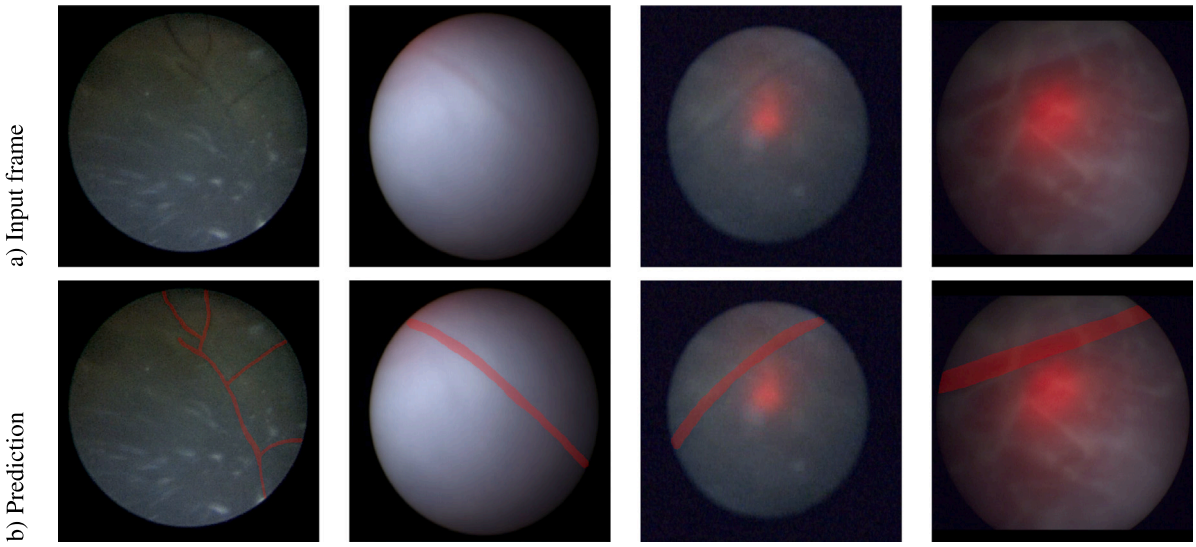

Fig. 12. Examples of poor visibility in video frames and their corresponding overlay prediction mask of placental vessels. The first row shows input video frames, and the second row shows the overlay prediction mask of placental vessels. Here, we demonstrate how a deep learning-based model may improve the visibility of placental vessels to assist fetal surgeons during TTTS fetoscopic surgery.

Center E using fetoscopes of lower optical quality compared to those used in the other centers.

The *FetReg2021* challenge was released in 2021 (Bano et al., 2021, Bano et al., 2023) for the evaluation of various segmentation methods using a two-center dataset comprising 24 in-vivo procedures. In the challenge, the best-performing method achieved an mIoU of 67.03% (Bano et al., 2023). We extended the publicly available *FetReg2021* dataset with additional data from four European fetal medicine centers. In the *FetReg2021* dataset, we identified areas for potential improvement in the pixel-wise annotations, which we aimed to enhance in our research. The annotations were carefully corrected and agreed upon by two domain experts. Our findings indicate that training our approach on improved annotations resulted in superior performance compared to using the publicly available annotations prior to correction. To foster research in this field, we are releasing the corrected annotations and making them available to the scientific community.

To address challenges associated with poor visibility within the amniotic sac environment, we introduced a novel data augmentation approach mimicking laser pointer effect, amniotic sac particles, camera structural defects, and fiber artifacts. This approach helps in building a robust and generalizable method. Specifically, TTTSNet trained with these custom data augmentation methods achieved superior quantitative segmentation performance compared to results without these methods, with an mIoU of 78.26% vs. 73.08%, respectively (see Table 3). Moreover, the proposed data augmentations are not only limited to TTTS but can also be used in different fetoscopy-based surgeries such as Amniotic Band Syndrome (ABS) and Fetoscopic Endoluminal Tracheal Occlusion (FETO) (Nassr et al., 2018).

Specifically, we obtained an average mIoU of 78.26%, whereas the mIoU using the uncorrected annotations was 72.05%, as shown in Table 4. Our quantitative outcomes suggest a noticeable correlation between the segmentation performance and the specific fetoscope hardware utilized in the data collection process, as substantiated by mIoU values of 75.01%, 70.11%, and 72.08%, for Center A, B, and E, respectively. Furthermore, it is important to note that the data gathered from both Center C and Center D were obtained using fetoscopes of higher optical quality compared to those used in the other centers, resulting in the most accurate segmentation outcomes. We achieved the highest mIoU of 86.15% and 86.09% for Centers C and D, which are quite similar. This may be correlated, as the same fetal surgeon acquired these data. However, they were obtained from two different centers with different hardware and clinical settings.

To demonstrate the robustness and sensitivity of our method in segmenting tiny placental vessels, we conducted an additional experiment on a subset of the test set, which consisted of video frames containing only tiny placental vessels. Overall, clinical experts selected nearly 17% of the test set, and our method, TTTSNet, outperformed eleven compared methods, and two TTTSNet-based configurations achieving a mean mIoU of 73.35 ± 10.79%, as shown in Table 5. It is worth noting that the second-best method, TTTSNet-S, demonstrates that our proposed neural network architecture can excel in the challenging task of tiny placental vessel segmentation. Both configurations achieved accurate segmentation of tiny placental vessels, despite the challenges posed by artifacts and a turbid environment (see the last three rows of Fig. 11).

Fig. 11 demonstrated that our method consistently provided accurate and robust segmentation covering the entire map of the placental vessel, including challenging regions like reduced contrast tiny vessels. Furthermore, TTTSNet produced smoother segmentation maps when compared to other state-of-the-art methods. Our proposed TTTSNet performs well in both types of placenta: anterior and posterior (see Table 6). Moreover, it is worth noting that video frames from anterior placenta cases exhibit enhanced visibility, with vessels appearing more detailed in the field of view due to their proximity to the camera (see Fig. 13). While performing FLP (Bamberg and Hecher, 2019) on the anterior placenta is a more challenging task than on the posterior placenta, the segmentation performance revealed that posterior placenta cases were much harder (mIoU of 81.01 ± 8.97% vs. 79.68 ± 6.65%). Notably, the worst-performing case (B_Video012) with an mIoU of 61.98 ± 13.11% came from a posterior placenta. This low segmentation performance is associated with the lowest resolution of the video frames (320 × 320 pixels). Moreover, the next two worst-performing cases (E_Video035 and F_Video031) with mIoUs of 71.28 ± 10.89% and 72.37 ± 5.80%, respectively, also came from posterior placenta cases. In contrast, the two best-performing cases (D_Video022 and D_Video006) with mIoUs of 91.76 ± 7.97% and 89.70 ± 12.98%, respectively, came from anterior placenta cases (see Table 6).

Two ablation studies demonstrated the effectiveness of various components developed in this work. Both neural network innovations and data augmentations significantly improved and contributed to achieving state-of-the-art segmentation performance (see Table 2). We showed that our proposed module, MEDCAM, enhances segmentation

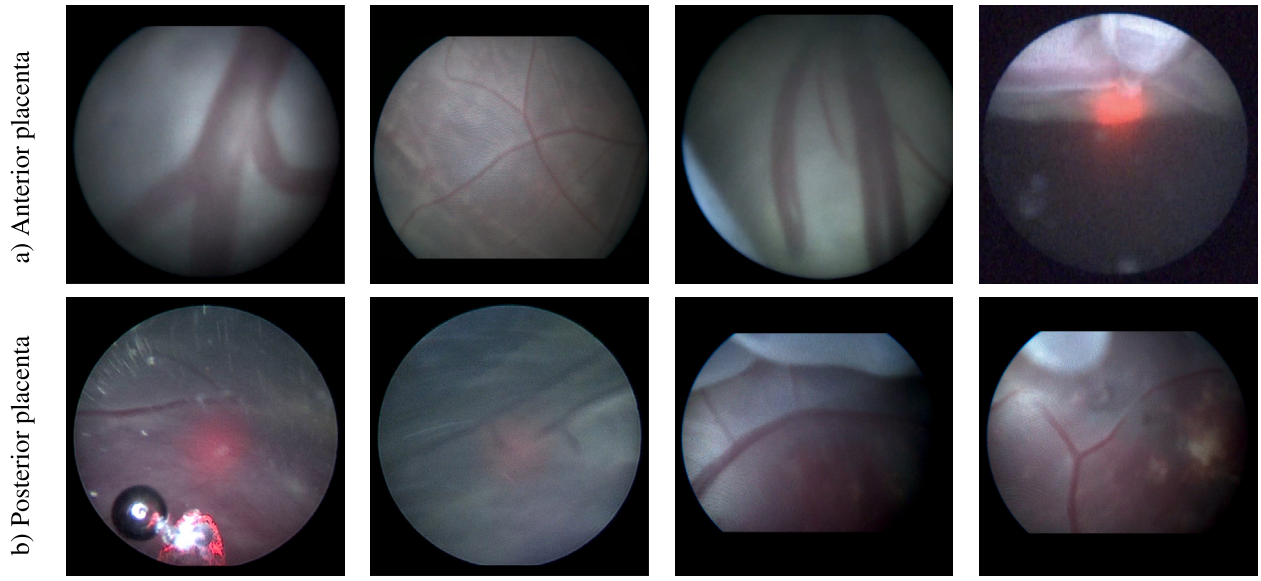

Fig. 13. Examples of video frames from two types of the placenta: (a) anterior, and (b) posterior placenta. We demonstrate that anterior placenta cases exhibit better visibility of placental vessels within the field of view compared to posterior cases, which impacts the segmentation performance of both types.

quality and effectively addresses limitations such as detecting tiny placental vessels within a small region of the field of view (see Fig. 10). This capability is particularly crucial during TTTS surgery to avoid postoperative complications. Any residual connections left unnoticed and uncoagulated may further enhance TTTS stages. In cases of fetal demise, such connections may also be responsible for the death of the second twin, or if the second twin survives, for brain hemorrhage leading to brain injury. Another postoperative complication is Twin Anemia Polycythemia Sequence (TAPS), which is one of the most prevalent side effects of FLP and is mostly due to a surgeon missing small placental intertwin vessels.

Fig. 12 illustrates an example of poor visibility in video frames alongside their corresponding overlay prediction masks. This visual representation underscores the challenge of identifying crucial anatomical features in such scenarios. Through our research, we establish that employing a deep learning-based method can significantly aid fetal surgeons during real-time TTTS surgery. By utilizing deep learning, we can effectively highlight regions containing placental vessels that would otherwise be challenging for the human eye to detect. This augmentation of visual perception may offer support to surgeons enhancing their ability to navigate during the procedure.

We compared TTTSNet with the Baseline method (Bano et al., 2020) of the FetReg2021 challenge using six-fold cross-validation on the training set, detailing an mIoU for individual videos based on both original and corrected annotations. We demonstrate that TTTSNet outperformed the Baseline method in both scenarios. Furthermore, we observed consistent results across all folds for both methods. However, in some cases such as Video006, Video011, and Video017, issues like intense laser glare or shadows and posterior placenta type coupled with low image resolution led to inaccurate vessel segmentation (see Table 7).

Furthermore, to demonstrate the robustness of TTTSNet and ensure a fair comparison, we evaluated TTTSNet against the top 5 performing methods from Task 1 of the FetReg2021 challenge. This comparison was conducted using both original and corrected annotations across the entire test set encompassing all six centers. In both scenarios, TTTSNet outperformed the next-best method, Baseline, by mIoU margins of 1.84% and 1.58%, respectively. Interestingly, it is evident that all compared methods achieved their highest results at Centers C, D, and F, while their performance was comparatively lower at Centers A, B, and E (see Table 8).

Although our method achieved overall success in qualitative outcomes, it also encountered some errors in quantitative results, but we did not observe recognizable patterns in the predictions. The high absolute error, devoid of any consistent trends or patterns, can be attributed to the inherent challenges of dealing with highly variable and competitive data encountered during TTTS surgery. The observed discrepancy in the performance of the mIoU metric for placental vessel segmentation in TTTS surgery can be attributed to various sources of error. Challenges arise due to poor image quality, characterized by low resolution, noise, and motion blur, which have a detrimental impact on the accuracy of vessel segmentation. These factors create difficulties for the model in detecting and outlining intricate vessel structures. Additionally, artifacts such as shadows, reflections, or occlusions caused by medical instruments further complicate the segmentation process, leading to errors in vessel boundaries. Moreover, irregular vessel structures commonly associated with placental pathologies, type of placenta and TTTS abnormalities introduce additional complexity. These irregularities include vessel tortuosity, abnormal branching patterns, and vessel dilation, deviating from the typical appearance of vessels and posing challenges for achieving accurate segmentation.

Our work is subject to limitations, which we intend to address through future research. Firstly, while our network architecture is designed for the segmentation of placental vessels, it would be clinically beneficial to differentiate vessels into arteries and veins. This is a challenging task due to the high similarity between the two. Additionally, we have not included segmentation of other structures such as the fetus or ablation tools as they are not considered clinically relevant. We found that including more classes in the training process can negatively impact the performance of vessel segmentation, and thus, we have chosen to focus solely on vessel segmentation. Another limitation is that the proposed network architecture relies on a single frame for segmentation. While this approach is efficient and fast, it does not take into account the temporal context. We plan to explore $2D + t$ spatio-temporal feature representation learning as a means of incorporating temporal context into the segmentation process in the future. Additionally, as with any deep learning project, we would like to utilize more data for training, validation, and testing to increase data diversity using different clinical centers, obtained on different equipment, and performed by other surgeons to test our solution. Lastly, our goal is to create a clinically useful system, and as such, it should be evaluated using clinical utility measures such as improvement in FLP success rate. This will be an important aspect in future evaluations.

7. Conclusions

We have proposed a network architecture for real-time placental vessel segmentation in videos obtained during FLP for TTTS. To improve performance, we have developed custom network and data augmentations specifically tailored for this task. Our experiments on a large and diverse test set have shown that TTTSNet is not only accurate in terms of segmentation metric but also robust in terms of generalizability to datasets from different institutions. Furthermore, our method demonstrates superior performance compared to current state-of-the-art methods. In the future, the use of TTTSNet may aid surgeons during real-time fetoscopic fetal surgery to accurately identify critical structures and ultimately improve outcomes of TTTS treatments.

CRediT authorship contribution statement

Szymon Płotka: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tomasz Szczepański:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Paula Szenejko:** Resources, Data curation. **Przemysław Korzeniowski:** Data curation. **Jesús Rodriguez Calvo:** Resources, Data curation. **Asma Khalil:** Resources, Data curation. **Alireza Shamshirsaz:** Resources, Data curation. **Robert Brawura-Biskupski-Samaha:** Resources, Data curation. **Ivana Išgum:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Clara I. Sánchez:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Arkadiusz Sitek:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Szymon Plotka reports financial support was provided by Horizon Europe 2020. Szymon Płotka reports financial support was provided by Foundation for Polish Science. Przemyslaw Korzeniowski reports equipment, drugs, or supplies was provided by NVIDIA Corp. Tomasz Szczepanski reports financial support was provided by Horizon 2020. Tomasz Szczepanski reports financial support was provided by Foundation for Polish Science. Arkadiusz Sitek reports financial support was provided by National Institutes of Health. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The FetReg2021 dataset is publicly available. Prior to usage, the TTTSNet dataset was anonymized in accordance with the ethical standards listed in the Helsinki Declaration. All patients provided written informed consent for the use of video from the TTTS procedure for research purposes. Both corrected annotations and external test set of TTTSNet dataset are publicly available.

Acknowledgments

This work is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 857533 (Sano) and the International Research Agendas programme of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund. This work is supported in part by National Institutes of Health (NIH) grant number HL159183. We would like to thank Patrycja Kaczmarczyk for the graphic design. We would like to thank NVIDIA Corporation for the in-kind donation of AGX Clara hardware.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2024.103330.

References

Akinbi, H.T., Narendran, V., Pass, A.K., Markart, P., Hoath, S.B., 2004. Host defense proteins in vernix caseosa and amniotic fluid. Am. J. Obstet. Gynecol. 191 (6), 2090–2096.

Almoussa, N., Dutra, B., Lampe, B., Getreuer, P., Wittman, T., Salafia, C., Vese, L., 2011. Automated vasculature extraction from placenta images. In: Medical Imaging 2011: Image Processing, vol. 7962. SPIE, pp. 501–510.

Bamberg, C., Hecher, K., 2019. Update on twin-to-twin transfusion syndrome. Best Pract. Res. Clin. Obstet. Gynaecol. 58, 55–65.

Bano, S., Casella, A., Vasconcelos, F., Moccia, S., Attilakos, G., Wimalasundera, R., David, A.L., Paladini, D., Deprest, J., De Momi, E., et al., 2021. FetReg: placental vessel segmentation and registration in fetoscopy challenge dataset. ArXiv preprint arXiv:2106.05923.

Bano, S., Casella, A., Vasconcelos, F., Qayyum, A., Benzinou, A., Mazher, M., Meriaudeau, F., Lena, C., Cintorrino, I.A., De Paolis, G.R., et al., 2023. Placental vessel segmentation and registration in fetoscopy: literature review and MICCAI FetReg2021 challenge findings. Med. Image Anal. 103066.

Bano, S., Vasconcelos, F., Tella-Amo, M., Dwyer, G., Gruijthuijsen, C., Van-der Poorten, E., Vercauteren, T., Ourselin, S., Deprest, J., Stoyanov, D., 2020. Deep learning-based fetoscopic mosaicking for field-of-view expansion. Int. J. Comput. Assist. Radiol. Surg. 15 (11), 1807–1816.

Baschat, A.A., Barber, J., Pedersen, N., Turan, O.M., Harman, C.R., 2013. Outcome after fetoscopic selective laser ablation of placental anastomoses vs equatorial laser dichorionization for the treatment of twin-to-twin transfusion syndrome. Am. J. Obstet. Gynecol. 209 (3), 234–e1.

Bhattarai, B., Subedi, R., Gaire, R.R., Vazquez, E., Stoyanov, D., 2023. Histogram of oriented gradients meet deep learning: A novel multi-task deep network for 2D surgical image semantic segmentation. Med. Image Anal. 85, 102747.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: fast and flexible image augmentations. Information 11 (2), 125.

Chalouhi, G., Essaoui, M., Stirnemann, J., Quibel, T., Deloison, B., Salomon, L., Ville, Y., 2011. Laser therapy for twin-to-twin transfusion syndrome (TTTS). Prenat. Diagn. 31 (7), 637–646.

Chang, J.-M., Huynh, N., Vazquez, M., Salafia, C., 2013. Vessel enhancement with multiscale and curvilinear filter matching for placenta images. In: 2013 20th International Conference on Systems, Signals and Image Processing. IWSSIP, IEEE, pp. 125–128.

Du, H., Wang, J., Liu, M., Wang, Y., Meijering, E., 2022. SwinPA-Net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation. IEEE Trans. Neural Netw. Learn. Syst. 35 (4), 5355–5366.

Elter, M., Rupp, S., Winter, C., 2006. Physically motivated reconstruction of fiberscopic images. In: 18th International Conference on Pattern Recognition, vol. 3. ICPR'06, IEEE, pp. 599–602.

Gao, G., Xu, G., Li, J., Yu, Y., Lu, H., Yang, J., 2022. FBSNet: A fast bilateral symmetrical network for real-time semantic segmentation. IEEE Trans. Multimed.

Gao, G., Xu, G., Yu, Y., Xie, J., Yang, J., Yue, D., 2021. Mscfnet: a lightweight network with multi-scale context fusion for real-time semantic segmentation. IEEE Trans. Intell. Transp. Syst. 23 (12), 25489–25499.

Haverkamp, F., Lex, C., Hanisch, C., Fahnenstich, H., Zerres, K., 2001. Neurodevelopmental risks in twin-to-twin transfusion syndrome: preliminary findings. Eur. J. Paediatr. Neurol. 5 (1), 21–27.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026–1034.

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.

Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023.

Khosravan, N., Bagci, U., 2018. S4ND: Single-shot single-scale lung nodule detection. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11. Springer, pp. 794–802.

Kim, M.-k., Yeo, Y.-s., Shin, H.-j., 2021. Binarization for eliminating calibration in fiberscope image processing. Opt. Commun. 497, 127198.
Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: International Conference on Learning Representations.
Lewi, L., Jani, J., Blickstein, I., Huber, A., Gucciardo, L., Van Mieghem, T., Doné, E., Boes, A.-S., Hecher, K., Gratacós, E., et al., 2008. The outcome of monochorionic diamniotic twin gestations in the era of invasive fetal therapy: a prospective cohort study. Am. J. Obstet. Gynecol. 199 (5), 514–e1.
Li, G., Kim, J., 2019. Dabnet: Depth-wise asymmetric bottleneck for real-time semantic segmentation. In: British Machine Vision Conference.
Li, H., Xiong, P., Fan, H., Sun, J., 2019. Dfanet: Deep feature aggregation for real-time semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9522–9531.
Loshchilov, I., Hutter, F., 2017. Sgdr: stochastic gradient descent with warm restarts. In: International Conference on Learning Representations.
Lou, A., Loew, M., 2021. Cfpnet: channel-wise feature pyramid for real-time semantic segmentation. In: 2021 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1894–1898.
Narendran, V., Wickett, R.R., Pickens, W.L., Hoath, S.B., 2000. Interaction between pulmonary surfactant and vernix: a potential mechanism for induction of amniotic fluid turbidity. Pediatr. Res. 48 (1), 120–124.
Nassr, A.A., Erfani, H., Fisher, J.E., Ogunleye, O.K., Espinoza, J., Belfort, M.A., Shamshirsaz, A.A., 2018. Fetal interventional procedures and surgeries: a practical approach. J. Perinat. Med. 46 (7), 701–715.
Nirthika, R., Manivannan, S., Ramanan, A., Wang, R., 2022. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. Neural Comput. Appl. 1–27.
Olivas, S.J., Arianpour, A., Stamenov, I., Morrison, R., Stack, R.A., Johnson, A.R., Agurok, I.P., Ford, J.E., 2015. Image processing for cameras with fiber bundle image relay. Appl. Opt. 54 (5), 1124–1137.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.
Perperidis, A., Dhaliwal, K., McLaughlin, S., Vercauteren, T., 2020. Image computing for fibre-bundle endomicroscopy: A review. Med. Image Anal. 62, 101620.
Račiukaitis, G., Stankevičius, E., Gečys, P., Gedvilas, M., Bischoff, C., Jäger, E., Umhofer, U., Völklein, F., 2011. Laser processing by using diffractive optical laser beam shaping technique.. J. Laser Micro/Nanoeng. 6 (1).
Ritt, G., 2019. Laser safety calculations for imaging sensors. Sensors 19 (17), 3765.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.
Ruano, R., Rodo, C., Peiro, J., Shamshirsaz, A., Haeri, S., Nomura, M., Salustiano, E., De Andrade, K., Sangi-Haghpeykar, H., Carreras, E., et al., 2013. Fetoscopic laser ablation of placental anastomoses in twin–twin transfusion syndrome using 'Solomon technique'. Ultrasound Obstet. Gynecol. 42 (4), 434–439.
Sadda, P., Imamoglu, M., Dombrowski, M., Papademetris, X., Bahtiyar, M.O., Onofrey, J., 2019. Deep-learned placental vessel segmentation for intraoperative video enhancement in fetoscopic surgery. Int. J. Comput. Assist. Radiol. Surg. 14 (2), 227–235.
Shi, M., Shen, J., Yi, Q., Weng, J., Huang, Z., Luo, A., Zhou, Y., 2022. LMFFNet: A well-balanced lightweight network for fast and accurate semantic segmentation. IEEE Trans. Neural Netw. Learn. Syst..
Steigman, S.A., Kunisaki, S.M., Wilkins-Haug, L., Takoudes, T.C., Fauza, D.O., 2010. Optical properties of human amniotic fluid: implications for videofetoscopic surgery. Fetal Diagn. Ther. 27 (2), 87–90.
Su, B., Tang, J., Liao, H., 2015. Automatic laser ablation control algorithm for an novel endoscopic laser ablation end effector for precision neurosurgery. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 4362–4367.
Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A., 2022. Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20730–20740.
Umur, A., Van Gemert, M., Nikkels, P., Ross, M., 2002. Monochorionic twins and twin–twin transfusion syndrome: the protective role of arterio-arterial anastomoses. Placenta 23 (2–3), 201–209.
Wang, Y., Zhou, Q., Xiong, J., Wu, X., Jin, X., 2019. ESNet: An efficient symmetric network for real-time semantic segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision. PRCV, Springer, pp. 41–52.
Waterhouse, D.J., Luthman, A.S., Yoon, J., Gordon, G.S., Bohndiek, S.E., 2018. Quantitative evaluation of comb-structure correction methods for multispectral fibrescopic imaging. Sci. Rep. 8 (1), 1–14.
Winter, C., Rupp, S., Elter, M., Munzenmayer, C., Gerhauser, H., Wittenberg, T., 2006. Automatic adaptive enhancement for images obtained with fiberscopic endoscopes. IEEE Trans. Biomed. Eng. 53 (10), 2035–2046.
Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39 (6), 1856–1867.
Zhuang, M., Zhong, X., Gu, D., Feng, L., Zhong, X., Hu, H., 2021. LRDNet: A lightweight and efficient network with refined dual attention decorder for real-time semantic segmentation. Neurocomputing 459, 349–360.