Medical Image Analysis (2022)

Contents lists available at ScienceDirect

Medical Image Analysis

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

FetReg2021: A Challenge on Placental Vessel Segmentation and Registration in Fetoscopy

Sophia Bano[a,*], Alessandro Casella[b,c], Francisco Vasconcelos[a], Abdul Qayyum[e], Abdesslam Benzinou[e], Moona Mazher[f], Fabrice Meriaudeau[g], Chiara Lena[c], Jessica Biagioli[c], Gaia Romana[c], Ilaria Anita Cintorrino[c], Daria Grechishnikova[h], Jing Jiao[i], Bizhe Bai[j], Yanyan Qiao[k], Binod Bhattarai[a], Rebati Raman Gaire[l], Ronast Subedi[l], Eduard Vazquez[m], Szymon Płotka[n], Aneta Lisowska[n], Arkadiusz Sitek[n], George Attilakos[o,p], Ruwan Wimalasundera[o,p], Anna L. David[o,p,q], Dario Paladini[r], Jan Deprest[p,q], Elena De Momi[c], Leonardo S. Mattos[b], Sara Moccia[e], Danail Stoyanov[a]

[a] *Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) and Department of Computer Science, University College London, London, UK*
[b] *Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy*
[c] *Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
[d] *The BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Italy*
[e] *ENIB, UMR CNRS 6285 LabSTICC, France*
[f] *Department of Computer Engineering and Mathematics, University Rovira i Virgili, Spain*
[g] *ImViA Laboratory, University of Bourgogne Franche-Comté, France*
[h] *Department of Physics, Moscow State University, Russia*
[i] *Fudan University, China*
[j] *Medical Computer Vision and Robotics Group, Department of Mathematical and Computational Sciences, University of Toronto, Canada*
[k] *MicroPort Robotics*
[l] *NepAL Applied Mathematics and Informatics Institute for Research (NAAMII), Nepal*
[m] *Redev Technology, UK*
[n] *Warsaw University of Technology, Poland*
[o] *Fetal Medicine Unit, Elizabeth Garrett Anderson Wing, University College London Hospital, London, UK*
[p] *EGA Institute for Women's Health, Faculty of Population Health Sciences, University College London, UK*
[q] *Department of Development and Regeneration, University Hospital Leuven, Leuven, Belgium*
[r] *Department of Fetal and Perinatal Medicine, Istituto "Giannina Gaslini", Genoa, Italy*

ARTICLE INFO

Article history:
Received Day Month 2022
Received in final form Day Month 2022
Accepted Day Month 2022
Available online Day Month 2022

Keywords: Fetoscopic videos,
Placental scene segmentation,
Video mosaicking,
Twin-to-twin transfusion syndrome
Surgical data science

ABSTRACT

Fetoscopy laser photocoagulation is a widely adopted procedure for treating Twin-to-Twin Transfusion Syndrome (TTTS). The procedure involves photocoagulation pathological anastomoses to restore a physiological blood exchange among twins. The procedure is particularly challenging, from the surgeon's side, due to the limited field of view, poor manoeuvrability of the fetoscope, poor visibility due to amniotic fluid turbidity, and variability in illumination. These challenges may lead to increased surgery time and incomplete ablation of pathological anastomoses, resulting in persistent TTTS. Computer-assisted intervention (CAI) can provide TTTS surgeons with decision support and context awareness by identifying key structures in the scene and expanding the fetoscopic field of view through video mosaicking. Research in this domain has been hampered by the lack of high-quality data to design, develop and test CAI algorithms. Through the *Fetoscopic Placental Vessel Segmentation and Registration (FetReg2021)* challenge, which was organized as part of the MICCAI2021 Endoscopic Vision (EndoVis) challenge, we released the first large-scale multi-centre TTTS dataset for the development of generalized and robust semantic segmentation and video mosaicking algorithms with a focus on creating drift-free mosaics from long duration fetoscopy videos. For this challenge, we released a dataset of 2060 images, pixel-annotated for vessels, tool, fetus and background classes, from 18 in vivo TTTS fetoscopy procedures

*Corresponding author.
e-mail: sophia.bano@ucl.ac.uk (Sophia Bano)

and 18 short video clips of an average length of 411 frames for developing placental scene segmentation and frame registration for mosaicking techniques. Seven teams participated in this challenge and their model performance was assessed on an unseen test dataset of 658 pixel-annotated images from 6 fetoscopic procedures and 6 short clips. The challenge provided an opportunity for creating generalized solutions for fetoscopic scene understanding and mosaicking. In this paper, we present the findings of the FetReg2021 challenge alongside reporting a detailed literature review for CAI in TTTS fetoscopy. Through this challenge, its analysis and the release of multi-centre fetoscopic data, we provide a benchmark for future research in this field.

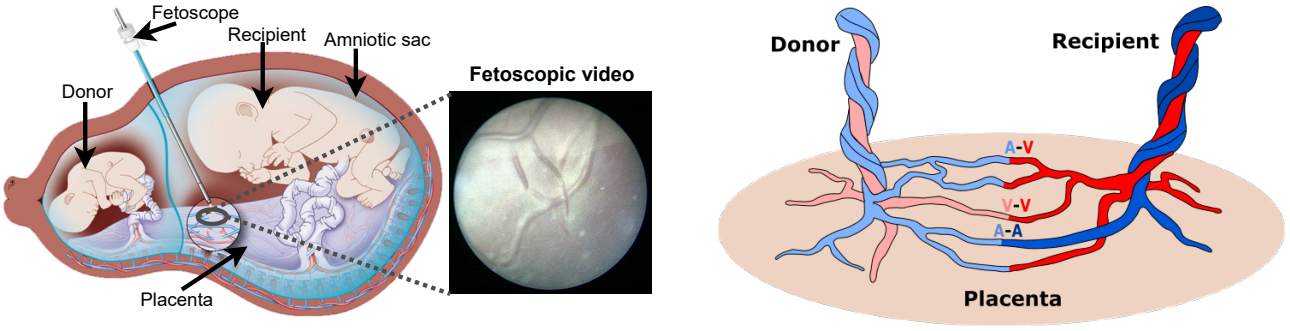
1. Introduction

Twin-to-twin transfusion syndrome (TTTS) is a severe complication of monochorionic twin pregnancies. TTTS is characterized by the development of an unbalanced and chronic blood transfer from one twin (the donor twin) to the other (the recipient twin) through placental anastomoses (Baschat et al., 2011). This shared circulation is responsible for serious complications, which may lead to profound fetal hemodynamic and cardiovascular disturbances (Lewi et al., 2013). In 2004, a randomized, controlled trial demonstrated that fetoscopic laser ablation of placental anastomoses in TTTS had a higher survival rate for at least one twin than other treatments, such as serial amnioreduction. Laser ablation further showed a lower incidence of complications, such as cystic periventricular leukomalacia and neurologic complications (Senat et al., 2004). The trial included pregnancy at 16-26 weeks' gestation. Such results were confirmed for pregnancy before 17 and after 26 weeks' gestation (Baud et al., 2013). A description of all the steps that brought laser surgery for coagulation of placental anastomoses to be the elective treatment for TTTS can be found in Deprest et al. (2010).

Fetoscopic laser photocoagulation involves the ultrasound-guided insertion of a fetoscope into the amniotic sac. Through the fetoscopic camera, the surgeon identifies abnormal anastomoses and laser ablate them to regulate the blood flow between the two fetuses (as illustrated in Fig. 1(a)). First attempts at laser coagulation included laser ablating all vessels that looked like anastomoses (a non-reproducible and operator-dependent technique), and laser ablating all vessels crossing the inter-fetus membrane (an approach that relies on the assumption that all vessels crossing the dividing membrane are pathological anastomoses) (Quintero et al., 2007). Today, the recognized elective treatment is the selective laser photocoagulation, which consists of the precise identification and lasering of placental pathological anastomoses. The selective treatment relies on the identification of the anastomoses (shown in Fig. 1(b)) and their classification into arterio-venous (from donor to recipient, AVDR, or from the recipient to donor, AVRD), arterio-arterial (AA) or veno-venous (VV) anastomoses. The identified AVDR anastomoses are laser ablated to regulate the blood flow between the two fetuses.

Despite all the advancements in instrumentation and imaging for TTTS (Cincotta and Kumar, 2016; Maselli and Badillo, 2016), residual anastomoses after monochorionic placentas treated with fetoscopic laser surgery still represent an issue (Lopriore et al., 2007). This may be explained considering the challenges in fetoscopic laser surgery, such as limited field of view (FOV), low fetoscopic image quality, poor visibility and high inter-subject variability. In this complex scenario, computer-assisted intervention (CAI) and surgical data science (SDS) methodologies may be exploited to provide surgeons with context awareness and decision support. However, the research in this field is still in its infancy, and several challenges still have to be tackled (Pratt et al., 2015). This includes dynamically changing views with poor texture visibility, low resolution, non-planar view, especially in the case of the anterior placenta, occlusions due to the fetus or working challenge port, fluid turbidity and specular highlights. Moreover, inter and intra-patient variabilities in the fetoscopic scenes are high. All these challenges hinder the designing of robust fetoscopic mosaicking methods for supporting navigation during in vivo fetoscopy.

In the context of TTTS fetoscopy, approaches for anatomical landmark segmentation (inter-fetus membrane, vessel), event detection, and mosaicking exist (see Sec. 2. Even though fetoscopic videos have large inter and intra procedure variabilities, the majority of the segmentation and event detection approaches are validated on a small subset of in vivo TTTS videos. Existing mosaicking approaches are validated only on a small subset of ex vivo Tella-Amo et al. (2019), in vivo Peter et al. (2018); Bano et al. (2020a) or underwater phantom sequences Gaisser et al. (2018). Recent intensity-based image registration Bano et al. (2020a); Li et al. (2021) methods relied on placental vessel segmentation maps for registration which facilitated in overcoming some of the visibility challenges (e.g. floating particles, poor illumination), however, such method fails when the predicted segmentation map is inaccurate, or the vessels are inconsistent across frames or are absent from the view. Deep learning-based flow-field matching for mosaicking Alabi et al. (2022) has also been proposed, which results in accurate registration even in regions with poor or weak vessels but such an approach fails when the fetoscopic scene is homogenous having poor texture. Besides developing placental vessel segmentation and mosaicking algorithms, a major effort is needed to collect large, high-quality, multi-centre datasets that can capture the variabilities of fetoscopic video.



(a) (b)

Fig. 1: Illustrations of Twin-to-Twin Transfusion Syndrome. (a) shows the fetoscopic laser photocoagulation procedure where the field-of-view of the fetoscope is extremely narrow. (b) shows the types of anastomoses (i) A-V: arterio-venous, (ii) V-V: veno-venous, and (iii) A-A: arterio-arterial.

1.1. Our Contributions

Placental Vessel Segmentation and Registration for Mosaicking (FetReg2021) [1] challenge is a crowdsourcing initiative to address key problems in fetoscopy towards developing CAI techniques for providing TTTS with decision support and context awareness. With FetReg2021, we collected a large multi-centre dataset to better capture not only inter- and intra-procedure variabilities but also inter-domain (data captured in two different clinical sites) variability. The FetReg2021 dataset can support developing robust and generalized models, paving the way for the translation of deep-learning methodologies in the actual surgical practice. The dataset is available to the research community [2] to foster research in the field. FetReg2021 was organized as part of the MICCAI 2021 Endoscopic Vision (EndoVis) [3] challenge, and aimed at solving two tasks: placental scene segmentation and frame registration for mosaicking.

In this paper, we present the results and findings of the FetReg2021 challenge, in which 7 teams participated. We further provide a detailed review of the relevant literature on CAI for fetoscopy. To conclude, we benchmark FetReg2021 participants methods against the existing state of the art in fetoscopic scene segmentation and mosaicking method.

2. Related work

This section surveys the most relevant current CAI methods developed in the field of TTTS surgery. This includes anatomical structure segmentation (Sec. 2.1), mosaicking and navigation (Sec. 2.2), and surgical event recognition (Sec. 2.3) methods for fetoscopic videos.

2.1. Anatomical structure segmentation

Segmentation algorithms in CAI for TTTS partition mainly focuses on vessel (Sec. 2.1.1) and placenta (Sec. 2.1.2) segmentation, as reference anatomical structures to provide surgeons with context awareness.

2.1.1. Placental vessel segmentation

Since the abnormal distribution of the anastomoses on the chorionic plate is responsible for the TTTS, exploration of the placenta vascular network is crucial during the photocoagulation procedure. The work presented by Almoussa et al. (2011) is among the first in the field of automatic placenta vasculature segmentation. The work, developed and tested with ex vivo images, combines the segmentation from Hessian-based filtering and a custom neural network trained on handcrafted features. The approach was improved by Chang et al. (2013), which introduced a vessel enhancement filter that combines multi-scale and curvilinear filter matching. The multi-scale filter extends the Hessian filter introducing two scaling parameters to tune vesselness sensitivity. The curvilinear filter matching refines vessel segmentation, preserving all the structures that fit in the vessel shape template defined by a curvilinear function.

The main limitation of both methods lies in the initial design that focuses on using ex vivo images due to the unavailability of intraoperative images. These methods were tested and optimised for segmenting vessels from ex-vivo placenta images. Ex-vivo placentas, immediately after removal, are placed on the operating table. After cleansing, the expert pathologist takes high resolution photos, in optimal light conditions and a static environment, unlike the gestational cavity. Therefore, acquired images will have visual characteristics that are highly different from intraoperative images. These methods were also proposed when the available computing power was considerably lower. Thus, neural networks require a manual feature engineering process to achieve acceptable performance. More importantly, Hessian-based methods have been proven to perform poorly in the case of tortuous and irregular vessels (Moccia et al., 2018).

More recently, researchers have focused their attention on convolutional neural networks (CNNs) to tackle the variability of intra-operative TTTS frames. Sadda et al. (2019) uses U-Net, achieving segmentation performance in terms of Dice Similarity Coefficient (DSC) on a dataset of 345 in-vivo fetoscopic frames of 0.55 ± 0.22.

U-Net is further explored in Bano et al. (2020a), as segmented vessels are used as a prior for fetoscopic mosaicking (Sec. 2.2.3). The authors tested several versions of U-Net, including the original version by Ronneberger et al. (2015), and U-Net with different backbones (i.e. VGG16, ResNet50

[1] FetReg challenge: `https://www.synapse.org/#!Synapse:syn25313156/`

[2] FetReg dataset: `https://www.ucl.ac.uk/interventional-surgical-sciences/weiss-open-research/weiss-open-data-server/fetreg-dataset`

[3] EndoVis Challenges: `https://endovis.grand-challenge.org/`

Table 1: Overview of the existing segmentation (Sec. 2.1-2.1.2, fetoscopic event detection (Sec. 2.3) and video mosaicking methods (Sec. 2.2). The type of dataset used in each method is also reported. Key: IFM - inter-fetus membrane, GMS - grid-based motion statistics, EMT - electromagnetic tracker.

Reference	Task	Methodology	Imaging type
Almoussa et al. (2011)	Vessel segmentation	Hessian filter and Neural Network trained on handcrafted features	Ex vivo
Chang et al. (2013)	Vessel segmentation	Combined Enhancement Filters	Ex vivo (150 images)
Sadda et al. (2019)	Vessel segmentation	Convolutional Neural Network (U-Net)	In vivo (345 frames from 10 TTTS surgeries)
Bano et al. (2019)	Vessel segmentation	Convolutional Neural Network	In vivo (483 frames from 6 TTTS surgeries)
Casella et al. (2020)	IFM segmentation	Adversarial Neural Network (ResNet)	In vivo (900 frames from 6 TTTS surgeries)
Casella et al. (2021)	IFM segmentation	Spatio-temporal Adversarial Neural Network (3D DenseNet)	In vivo (2000 frames from 20 TTTS surgeries)
Reeff et al. (2006)	Mosaicking	Hybrid feature and intensity-based	In water ex vivo placenta
Daga et al. (2016)	Mosaicking	Feature-based with GPU for real time computation	Ex vivo, Phantom placenta
Tella et al. (2016)	Mosaicking	Combined EM and visual tracking probablistic model	Ex vivo w/laparoscope& EMT
Gaisser et al. (2016)	Mosaicking	Deep-learned features through contrastive loss	Ex vivo and Phantom placenta video frames
Yang et al. (2016)	Mosaicking	SURF features matching and RANSAC for transformation estimation	Ex vivo and monkey placentas w/laparoscope
Gaisser et al. (2017)	Mosaicking	Handcrafted features and LMedS for tranformation estimation	Ex vivo, In water placenta phantom
Tella-Amo et al. (2018)	Mosaicking	Combined EM and visual tracking with bundle adjustment	Ex vivo placenta w/laparoscope & EMT
Gaisser et al. (2018)	Mosaicking	Extended Gaisser et al. (2016) to detect stable vessel regions	In water placenta phantom
Sadda et al. (2018)	Mosaicking	AGAST detector with SIFT followed by GMS matching	In vivo (# frames/clips)
Peter et al. (2018)	Mosaicking	Direct pixel-wise alignment of image gradient orientations	In vivo (# frames/clips)
Tella-Amo et al. (2019)	Mosaicking	Pruning through EM and super frame generation	Ex vivo placenta w/laparoscope & EMT
Bano et al. (2019, 2020a)	Mosaicking	Deep learning-based four point registration in consecutive images	Synthetic, Ex vivo, Phantom, In vivo phantom)
Bano et al. (2020a)	Mosaicking	Direct aligment of predicted vessel maps	In vivo fetoscopy placenta (6 procedures) 4
Alabi et al. (2022)	Mosaicking	FlowNet 2.0 with robust estimation for direct registration	Extended in vivo fetoscopy placenta data (6 procedures) 4
Vasconcelos et al. (2018)	Ablation detection	Binary classification using ResNet	In vivo fetoscopy videos (5 procedures)
Bano et al. (2020c)	Event detection	Spatio-temporal model for multi-label classification	In vivo fetoscopy videos (7 procedures)

and ResNet101). Segmentation performance is evaluated on a dataset of 483 in-vivo images from six TTTS surgery, the first publicly available [4].

Despite the advances introduced by CNN-based methods, the state of the art methods cannot tackle the high variability of intraoperative images. From one side, the datasets used to train these algorithms are small in size and the challenges of intraoperative images, as listed in Sec. 1, are not always represented. At the same time, encoder-decoder architectures trained to minimize cross-entropy and DSC loss fail in segmenting poor contrasted vessels and vessels with uneven margins.

The research in the field is strongly limited by the low availability of comprehensive expert-annotated datasets collected in different surgical settings that could encode such variability. This is mainly due to the low incidence of TTTS, which make difficult systematic data collection, and the lack of annotators with sufficient domain expertise to ensure clinically correct groundtruth.

2.1.2. Inter-fetus membrane segmentation

At the beginning of the surgical treatment, due to the very limited field of view and poor image quality, the surgeon has to find references to keep oriented within the amniotic cavity. The structure identified for this purpose is the inter-fetus membrane. The visibility of this membrane can be very variable, depending on the chorios characteristics, in addition to the challenges described so far in fetoscopic images. Once located, the surgeon refers to the inter-fetus membrane as a reference during placental vascular network exploration.

Automatic inter-fetus membrane segmentation has been introduced by Casella et al. (2020). In this work, an adversarial segmentation network based on ResNet is proposed to enforce placenta-shape constraining. The method was tested on a dataset of 900 intraoperative frames from 6 TTTS patients with an average DSC of 91.91%. Despite the promising results, this method suffers when illumination is too bright or dark, so the membrane is barely visible.

The work in Casella et al. (2020) is extended in Casella et al. (2021) by exploiting dense connectivity and spatio-temporal information to improve membrane segmentation accuracy and tackle high illumination variability. The segmentation accuracy was tested on the first publicly available dataset of 2000 in-vivo images from 20 TTTS surgeries outperforming the method previously proposed.

Despite the promising results achieved in the literature, the task of inter-fetus membrane segmentation is still poorly explored, and more research needs to be performed.

2.2. Fetoscopic Mosaicking and Navigation

Video mosaicking aims at generating an expanded FOV image of the scene by registering and stitching overlapping video frames. Video mosaicking of high-resolution images has been extensively used as navigation guidance in the context of aerial, underwater, and street view imaging and also in consumer photography to build panorama shots. However, the outputs from off-the-shelf mosaicking methods have significantly poorer quality or fail completely when applied to fetoscopy videos due to the added visibility challenges of intra-operative images. Nevertheless, fetoscopy video mosaicking remains an active research topic within the context of computer-assisted intervention. Such a technique can facilitate the surgeon during the procedure in better localization of the anastomotic sites, which can improve the procedural outcomes.

Mosaicking for FOV expansion in fetoscopy has been explored using handcrafted feature-based and hybrid methods (Sec. 2.2.1), intensity-based (Sec. 2.2.2), and deep learning-based (Sec. 2.2.3) methods. These methods are either devised

[4]Fetoscopy placenta dataset: `https://www.ucl.ac.uk/interventional-surgical-sciences/fetoscopy-placenta-data`

for synthetic placental images, ex vivo placental images/videos or in vivo videos.

2.2.1. Handcrafted feature-based and hybrid methods

Feature-based methods involve detecting and matching features across adjacent or overlapping frames, followed by estimating the transformation between the image pairs. On the other hand, hybrid methods utilize multimodal data (combination of image and electromagnetic tracking data) or a combination of feature-based and intensity-based methods.

Early approaches focused on accomplishing fetoscopic mosaicking from videos or overlapping a pair of images only for image registration and mosaicking. Reeff et al. (2006) proposed a hybrid method that used classical feature detection and matching approach for first estimating the transformation of each image with respect to a reference frame, followed by global optimization by minimizing the sum of the squared differences of pixel intensities between two images. Multi-band blending was applied for seamless stitching. For testing the hybrid method, the authors recorded one ex vivo placenta fixed in a hemispherical receptacle submerged in water to mimic an in vivo imaging scenario. Such an experiment also allowed capturing camera calibration to remove lens distortion. A short sequence of 40 frames sampled at 3 frames per second was used for the evaluation. The matched feature correspondences were visually analyzed to mark them as correct or incorrect, which is a labour-intensive task. The generated mosaic with and without global optimization was shown for qualitative comparison.

Handcrafted feature-based methods, similar to what is commonly used in high-resolution image stitching in computer vision, were also explored for fetoscopic mosaicking. Daga et al. (2016) presented the first approach toward generating real-time mosaics. The approach considered using SIFT for feature detection and matching. For real-time computation, texture memory was used on GPU for computing extremes of the difference of Gaussian (DoG) that describes SIFT features. Planar images of ex vivo phantom placenta recorded by mounting a fetoscope to a KUKA robotic arm were used for validating the approach. The robot was programmed to follow a spiral path that facilitated qualitative evaluation. Yang et al. (2016) proposed a SURF feature detection and matching based approach for generating mosaics from 100 frames long sequences that captured ex vivo phantom and monkey placentas. Additionally, pair of images correspondence failure approach was proposed based on the statistical attributes of the feature distribution and an adaptive updating mechanism for parameter tuning to recover registration failures. Gaisser et al. (2017) used different keypoint descriptors (SIFT, SURF, ORB) along with Least Median of Squares (LMedS) for estimating the transformation between overlapping pairs of images.

Through experiments on both ex vivo and in-water phantom sequences, the authors showed that handcrafted features returns either no features or low confidence features due to texture paucity and dynamically changing visual conditions. This leads to inaccurate or poor transformation estimation.

Sadda et al. (2018) proposed a feature-based method that relied on extracting AGAST corner detector Mair et al. (2010), SIFT as descriptor and grid-based motion statistics (GMS) Bian et al. (2017) for refining feature matching for homography estimation. The validation was performed on 22 in vivo fetoscopic image pairs. Additionally, in a hybrid approach by Sadda et al. (2019), vessel segmentation masks were also used for selecting AGAST features only around the vessel regions. However, the reported error was large mainly because of linear and single vessels in the 22 image pairs under analysis. Using handcrafted feature descriptors such as SIFT shows poor performance in the case of in vivo placental videos due to the added challenges introduced by poor visibility, texture paucity and low resolution imaging.

A few approaches used an additional electromagnetic tracker in an ex vivo setting to design a feature-based method for improved mosaicking.

Tella et al. (2016); Tella-Amo et al. (2018) assumed the placenta to be planar and static and used a combination of visual and electromagnetic tracker information for generating robust and drift-free mosaics. Mosaicking performance was increased in Tella-Amo et al. (2019), where the pruning of overlapping frames and generation of a super frame for reducing computational time was proposed. An Aurora electromagnetic tracker (EMT) was mounted on the tip of a laparoscope to obtain camera pose measurements. Using this setup, a data sequence of 701 frames was captured from a phantom (i.e., a printed image of a placenta). Additionally, a synthetic sequence of 273 frames following only planar motion was also generated for quantitative evaluation. The camera pose measurements from the EMT were incorporated with frame-based visual information using a probabilistic model to obtain globally consistent sequential mosaics. It is worth mentioning that laparoscopic cameras used are considerably better than fetoscopic cameras.

However, current clinical regulations and the limited form factor of the fetoscope hinder the use of such a tracker in intra-operative settings.

2.2.2. Intensity-based methods

Intensity-based image registration is an iterative process that uses raw pixel values for direct registration through first selecting features, such as edges, contours, followed by a metric, such as mutual information, cross-correlation, the sum of squared difference, absolute difference, for describing how similar two overlapping input images are and an optimizer for obtaining the best alignment through fitting a spatial transformation model.

The use of direct pixel-wise alignment of oriented image gradients for creating a mosaic was proposed by Peter et al. (2018) that was validated on only one in vivo fetoscopic sequence of 600 frames. An offline bag of words was used to improve the global consistency of the generated mosaic.

Bano et al. Bano et al. (2020a) proposed a placental vessel-based direct registration approach. A U-Net model was trained on a dataset of 483 vessel annotated images from 6 in vivo fetoscopy for segmenting vessels. The vessel maps from consecutive frames were registered, estimating the affine transformation between the frames. Testing was performed on 6 additional in vivo fetoscopy video clips. The approach facilitated overcoming visibility challenges, such as floating particles and varying

illumination.

However, the method failed when the predicted segmentation map is inaccurate or in views with thin or no vessels.

2.2.3. Deep learning-based methods

Existing deep learning-based methods for fetoscopic mosaicking mainly focused on training a CNN network Bano et al. (2019, 2020b) for directly estimating homography between adjacent frames, extracting stable regionsGaisser et al. (2016) in a view, or learning robust key points Alabi et al. (2022) for consecutive pairs of images registration.

A deep learning-based feature extractor was proposed by Gaisser et al. (2016) that used similarity learning using contrastive loss when training a Siamese convolutional neural network (CNN) architecture between pairs of similar and dissimilar small patches extracted from ex vivo placental images. The learned feature extractor was used for extracting features from pairs of overlapping images, followed by using LMedS for the transformation estimation. Due to motion blur and texture paucity that affected the feature extractor performance, the method was validated only on a short sequence (26 frames) that captured an ex vivo phantom placenta. Gaisser et al. (2018) extended their similarity learning approach (Gaisser et al., 2016) for detecting stable regions on the vessels of the placenta. These stable regions' representation is used as features for placental image registration in an in-water phantom setting. The obtained homography estimation did not result in highly accurate registration, as the learned regions were not robust to visual variabilities in underwater placental scenes.

Methods for estimating 4-point homography using direct registration with deep learning exist in computer vision literature DeTone et al. (2016); Nguyen et al. (2018). Bano et al. (2019, 2020b) extended the work of DeTone et al. (2016) to propose one of the first homography-based methods for fetoscopic mosaicking, which was tested on 5 diverse placental sequences, namely, synthetic sequence of 811 frames, ex vivo placenta planar sequence of 404 frames, ex vivo phantom placenta sequence of 681 frames, in vivo phantom placenta sequence of 350 frames and in vivo TTTS fetoscopic video of 150 frames. In (Bano et al., 2019, 2020b), a VGG-like model was trained to estimate 4-point homography between two patches extracted from the same image with known transformation. Controlled data augmentation was applied to the two patches for network training. Filtering is then applied during testing to obtain the most consistent homography estimation. The proposed approach led to advancing the literature on fetoscopic mosaicking. However, the proposed network mainly focused on estimating rigid transformation (rotation and translation) between adjacent frames due to controlled data augmentation. As a result, the generated mosaics in non-planar sequences accumulated drift over time.

More recently, deep learning-based optical flow combined with inconsistent motion filtering for robust fetoscopy mosaicking has been proposed Alabi et al. (2022). Their method relied on FlowNet-v2 Ilg et al. (2017) for obtaining dense correspondence between adjacent frames, robust estimation using RANSAC and local refinement for removing the effect of floating particles and specularities for improved registration. Unlike Bano et al. (2020a) which used placental vessel prediction to drive mosaicking, Alabi et al. (2022) did not rely on vessels, as a result, it managed to generate robust and consistent mosaic for longer duration of fetoscopic videos. Their approach was tested on the extended fetoscopy placenta dataset from Bano et al. (2020a). While this approach significantly improved fetoscopic mosaicking, further analysis is needed to investigate its performance in low-textured placental regions.

2.3. Surgical event recognition

TTTS laser therapy has a relatively simple workflow with an initial inspection of the placenta's surface and vasculature to identify and visualise photocoagulation targets. Fetoscopic laser therapy is conducted by photocoagulation of each identified target in sequence. Automatic identification of these surgical phases and surgical events is an essential step towards general scene understanding and tracking of the photocoagulation targets. This identification can provide temporal context for tasks such as segmentation and mosaicking. It could provide prior to finding the most reliable images for registration (before ablation) or identify changes in the scene appearance (after ablation).

The CAI literature has hardly explored event detection or workflow analysis methods. Vasconcelos et al. (2018) used a ResNet encoder to detect ablation in TTTS procedures, additionally also indicating when the surgeon is ready for ablating the target vessel. The method was validated on 5 in vivo fetoscopic videos. Bano et al. (2020c) combined CNNs and recurrent networks for the spatio-temporal identification of fetoscopic events, including clear view, occlusion (i.e., fetus or working channel port in the field-of-view), laser tool presence, and ablating laser tool present. The method was effective in identifying clear view segments Bano et al. (2020c) suitable for mosaicking and was validated on 7 in vivo fetoscopic videos.

3. The FetReg Challenge: Dataset, Submission, Evaluation

In this section, we present the dataset of the *EndoVis FetReg 2021* challenge and its tasks (Sec. 3.1), the evaluation protocol designed to assess the performance of the participating methods (Sec. 3.2) and an overview of the challenge setup and submission protocol(Sec. 3.3).

3.1. Dataset and Challenge Tasks

The *EndoVis FetReg 2021* challenge aimed at advancing the current state-of-the-art in placental vessel segmentation and mosaicking (Bano et al., 2020a) by providing a benchmark multicentre large-scale dataset that captured variability across different patients and different clinical institutions. We also aimed to perform out-of-sample testing to validate the generalization capabilities of trained models. The participants were required to complete two sub-tasks which are critical in fetoscopy, namely:

- **Task 1: Placental semantic segmentation**: The participants were required to segment four classes, namely, background, vessels, tool (ablation instrument, i.e. the tip of

Table 2: Summary of the *EndoVis FetReg 2021* training and testing dataset. For each video, image resolution, the number of annotated frames (for the segmentation task), the occurrence of each class per frame and the average number of pixels per class per frame are presented. For the registration task, the number of unlabelled frames in each video clip is provided. Key: BG - background.

TRAINING DATASET											
Sr.	**Video name**	**Image Resolution (pixels)**	**No. of Labelled frames**	**Occurrence (frame)**				**Occurrence (Avg. pixels)**			**Unlabel--led clips # frames**
				Vessel	**Tool**	**Fetus**	**BG**	**Vessel**	**Tool**	**Fetus**	
1.	Video001	470×470	152	152	21	11	196463	21493	1462	1482	346
2.	Video002	540×540	153	153	35	1	271564	16989	3019	27	259
3.	Video003	550×550	117	117	52	32	260909	27962	3912	9716	541
4.	Video004	480×480	100	100	21	18	212542	14988	1063	1806	388
5.	Video005	500×500	100	100	35	30	203372	34350	2244	10034	722
6.	Video006	450×450	100	100	49	4	171684	28384	1779	653	452
7.	Video007	640×640	140	140	30	3	366177	37703	4669	1052	316
8.	Video008	720×720	110	105	80	34	465524	28049	13098	11729	295
9.	Video009	660×660	105	104	40	14	353721	68621	7762	5496	265
10.	Video011	380×380	100	100	7	37	128636	8959	184	6621	424
11.	Video013	680×680	124	124	54	21	411713	36907	8085	5695	247
12.	Video014	720×720	110	110	54	14	464115	42714	6223	5348	469
13.	Video016	380×380	100	100	16	20	129888	11331	448	2734	593
14.	Video017	400×400	100	97	20	3	151143	7625	753	479	490
15.	Video018	400×400	100	100	26	11	139530	15935	1503	3032	352
16.	Video019	720×720	149	149	15	31	470209	38513	1676	8002	265
17.	Video022	400×400	100	100	12	1	138097	21000	650	253	348
18.	Video023	320×320	100	92	14	8	94942	6256	375	828	639
	All training videos		**2060**	2043	581	293	4630229	467779	58905	74987	**7411**
TESTING DATASET											
19.	Video010	622×622	100	92	7	28	341927	40554	1726	19410	320
20.	Video012	320×320	100	100	54	0	95845	5132	1422	0	507
21.	Video015	720×720	125	124	83	28	452552	47221	12082	6545	530
22.	Video020	720×720	123	100	15	1	436842	59884	15259	6415	307
23.	Video024	320×320	100	110	72	13	203372	34350	2244	10034	269
24.	Video025	720×720	110	648	320	83	459947	43189	9801	5464	272
	All testing videos		**658**	2043	581	293	1880090	205009	40638	37879	**2205**

the laser probe) and fetus, on the provided dataset. Fetoscopic frames from 24 TTTS procedures collected in different centres were annotated for the four classes that commonly occur during the procedure. This task was evaluated on unseen test data (6 videos) independent of the training data (18 videos). The segmentation task aimed to assess the generalization capability of segmentation models on unseen fetoscopic video frames.

- **Task 2: Registration for Mosaicking**: The participants were required to perform the registration of consecutive frames to create an expanded FOV image of the fetoscopic environment. Fetoscopic video clips from 18 multicentre fetoscopic procedures were provided as the training data. No registration annotations were provided as it is not possible to get the groundtruth registration during the in vivo clinical fetoscopy. The task was evaluated on 6 unseen video clips extracted from fetoscopic procedure videos, which were not part of the training data. The registration task aimed to assess the robustness and performance of registration methods for creating a drift-free mosaic from unseen data.

The *EndoVis FetReg 2021* dataset is unique as it is the first large-scale fetoscopic video dataset of 24 different TTTS fetoscopic procedures. The videos contained in this dataset are collected from two fetal surgery centres across Europe, namely,

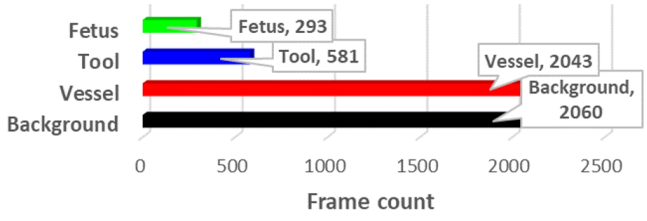
- Centre 1: Fetal Medicine Unit, University College London Hospital (UCLH), London, UK,
- Centre 2: Department of Fetal and Perinatal Medicine, Istituto "Giannina Gaslini", Genoa, Italy,

Both centres contributed with 12 TTTS fetoscopic laser photocoagulation videos each. A total of 9 videos from each centre (18 videos in total) form the train set, while 3 videos from each centre (6 videos in total) form the test set. Alongside capturing the intra-case and inter-case variability, the multi-centre data collection allowed capturing the variability that arises due to different clinical settings and imaging equipment at different clinical sites. At UCLH, the data collection was carried out as part of the GIFT-Surg[5] project. The requirement for formal ethical approval was waived, as the data were fully anonymized in the corresponding clinical centres before being transferred to the organisers of the *EndoVis FetReg 2021* challenge.

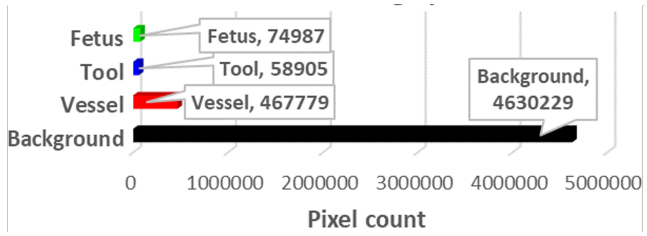
Table 2 summarises *EndoVis FetReg 2021* dataset characteristics. Specifically, video number 001, 002, 003, 007, 008, 009, 013, 014, 015, 019, 020, 025 were obtained from the first centre, and video number 004, 005, 006, 010, 011, 012, 016, 017,

[5]GIFT-Surg project: `https://www.gift-surg.ac.uk/`

018, 022, 023, 024 were obtained from the second centre. Further details about the segmentation and registration datasets are provided in following sections.

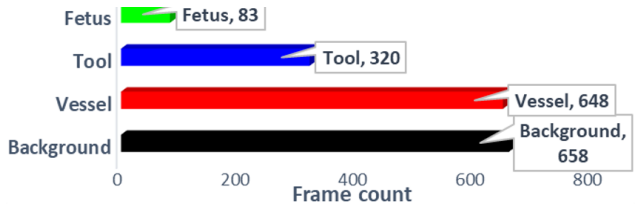


(a) Occurrence per frame

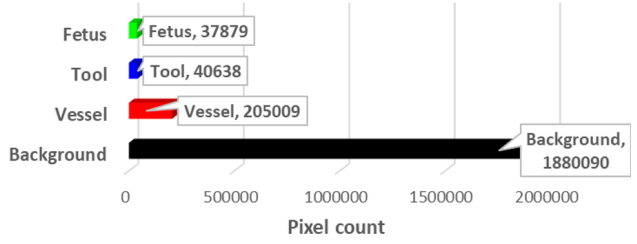


(b) Occurrence average pixel

Fig. 2: Training dataset distribution: (a) and (b) segmentation classes and their overall distribution in the segmentation data.



(a) Occurrence per frame



(b) Occurrence average pixel

Fig. 3: Testing dataset distribution: (a) and (b) segmentation classes and their overall distribution n the segmentation data.

3.1.1. Dataset for placental semantic segmentation

Fetoscopy videos acquired from the two different fetal medicine centres were first decomposed into frames and excess black background was cropped to obtain squared images capturing mainly the fetoscope FOV. From each video, a subset of non-overlapping informative frames (in range 100-150) is selected and manually annotated. All pixels in each image are labelled with background (class 0), placental vessel (1), ablation tool (2) or fetus class (3). Labels are mutually exclusive.

Annotation is performed by four academic researchers and staff members with a solid background in fetoscopic imaging. Additionally, annotation services are obtained from Humans in the Loop (HITL)[6] for a subset of videos. All the annotations are verified by two fetal medicine specialists to confirm the correctness and consistency of the labels. The publicly available Supervisely[7] platform is used for annotating the dataset.

The *FetReg* train and test dataset for the segmentation task contains 2060 and 658 annotated images from 18 and 6 different in-vivo TTTS fetoscopic procedures, respectively. Figure 2(a) and Fig. 2(b) show the overall class occurrence per frame and class occurrence in average pixels per frame on the training dataset. The same for test dataset is shown in Figure. 3(a) and Fig. 3(b). Note that the frames present different resolutions as the fetoscopic videos are captured at different centres with different facilities (e.g., device, light scope). The dataset is highly unbalanced: the *Vessel* is the most frequent class while *Tool* and *Fetus* are presented only in a small subset of images corresponding to 28% and 14%, respectively of the training dataset and 48% and 13% of the test dataset. When observing the class occurrence in average pixels per image, the *Background* class is the most dominant, with *Vessel*, *Tool* and *Fetus* occur 10%, 0.13% and 0.16% in train dataset and 11%, 0.22%, and 0.20% in test dataset, respectively.

Figure. 4 shows some representative annotated frames from each video. Note that the frame appearance and quality changes in each video due to the large variation in intra-operative environment among different cases. Amniotic fluid turbidity resulting in poor visibility, artefacts introduced due to spotlight light source and reddish reflection introduced by the laser tool, low resolution, texture paucity, non-planar views due to anterior placenta imaging, are some of the major factors that contribute to increase the variability in the data. Large intra-case variation can also be observed from these representative images. All these factors contribute towards limiting the performance of the existing placental image segmentation and registration methods (Bano et al., 2020a, 2019, 2020b). The *EndoVis FetReg 2021* challenge provided an opportunity to make advancements in the current literature by designing and contributing novel segmentation and registration methods that are robust even in the presence of the above-mentioned challenges.

3.1.2. Dataset for registration for mosaicking

A typical TTTS fetoscopy surgery takes approximately 30 minutes. Only a sub-set of fetoscopic frames are suitable for frame registration and mosaicking because of the presence of fetuses, laser ablation fibre and working channel port which can occlude the field-of-view of the fetoscope. Mosaicking is mainly required in occlusion-free video segments that capture the surface of the placenta Bano et al. (2020c) as these are the segments in which the surgeon is exploring the intraoperative environment to identify abnormal vascular connections. Expanding the field-of-view through mosaicking in these video segments can facilitate the procedure by providing better visualization of the environment.

For the registration for mosaicking task, we have provided one video clip per video for all 18 procedures in the training

[6]Humans in the Loop: `https://humansintheloop.org/`

[7]Supervisely: a web-based annotation tool: `https://supervise.ly/`

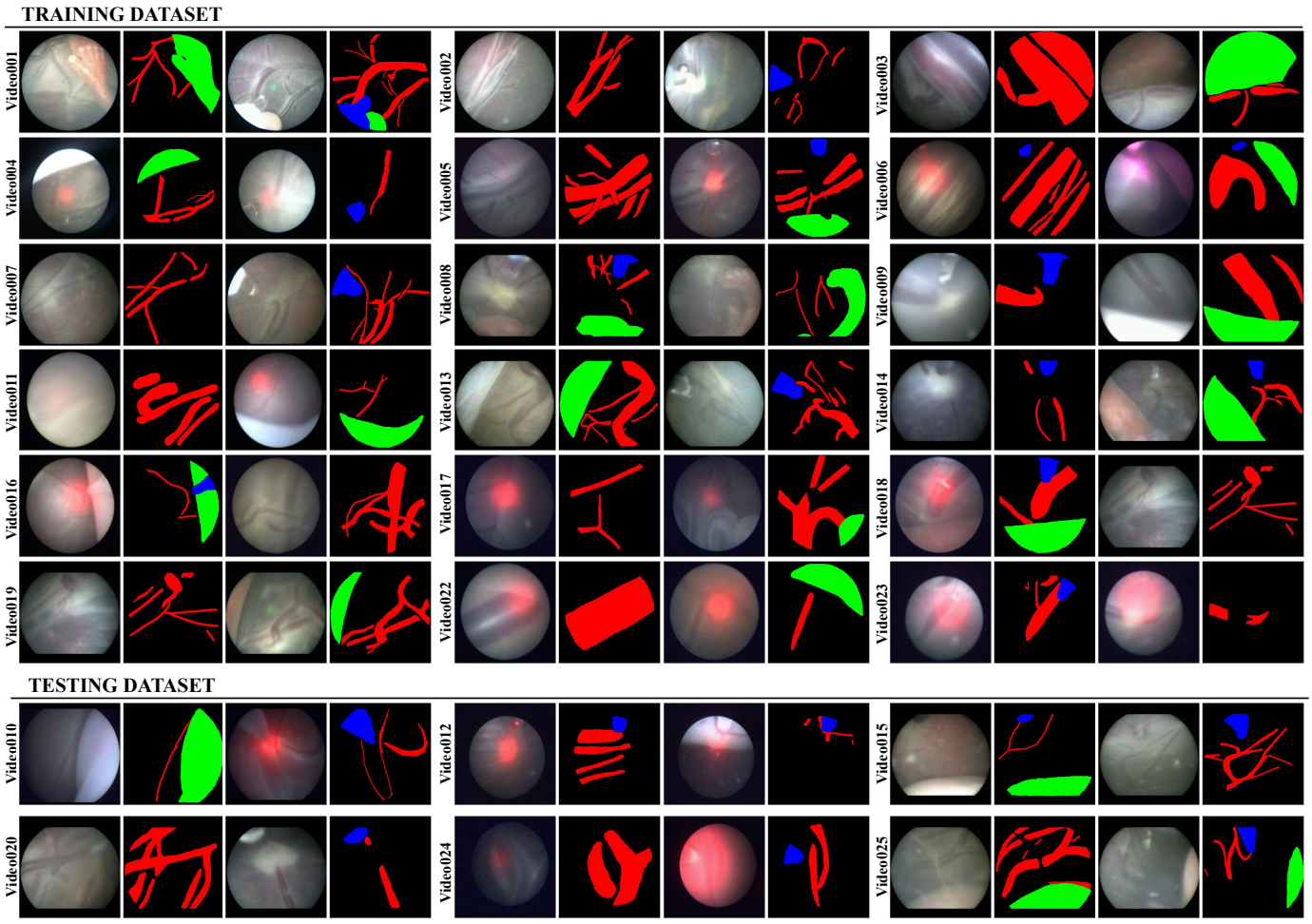


Fig. 4: Representative images from training and test datasets along with the segmentation annotations (groundtruth).

dataset. Likewise, one clip per video from all 6 procedures in the test dataset is selected for testing and validation. These frames are neither annotated with segmentation labels nor have registration groundtruth. The number of frames in each video clip is reported in Table 2 for training and test dataset. Representative frames from each clip are shown in Fig. 6. Representative frames at every 2 seconds from the some video clips are shown in Fig. 5. Observe the variability in the appearance, lighting conditions and image quality in all video clips. Even though there is no noticeable deformation in fetoscopic videos, which is usually thought to occur due to breathing motion, the views can be non-planar as the placenta can be anterior or posterior. Moreover, there is no groundtruth camera motion and scene geometry that can be used to evaluate video registration approaches for in-vivo fetoscopy. In Section 3.2.2, we detail how this challenge is addressed with an evaluation metric that is correlated with good quality, consistent, and complete mosaics Bano et al. (2020a).

3.2. Evaluation protocol

3.2.1. Segmentation Evaluation

For evaluating the performance of segmentation models (Task 1), we compute for each frame provided in the test set the mean Intersection over Union (*mIoU*) per class between the prediction and the manually annotated segmentation masks. Intersection-over-Union IoU is another most commonly used metric for evaluating segmentation algorithms which measures the spatial overlap between the predicted and groundtruth segmentation masks as:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{1}$$

where *TP* are the correctly classified pixels belonging to a class, *FP* are the pixels incorrectly predicted in a specific class and *FN* are the pixels in a class incorrectly classified as not belonging to it. The mean *mIoU* for each class and video samples is computed. Overall mean *mIoU* over all classes and all test samples is also computed and used for ranking different methods under comparison.

3.2.2. Frame Registration and Mosaicking Evaluation

For evaluating homographies and mosaics (Task 2), we use the evaluation metric presented in Bano et al. (2020a) in the absence of groundtruth. The metric, that we referred as N-frame SSIM, aims as evaluating the consistency in the adjacent frames. A visual illustration of the N-frame SSIM metric is presented in Fig. 6. Given N consecutive frames and a set of $N-1$ homographies $\{H_1, H_2, ..., H_{N-1}\}$, we evaluate

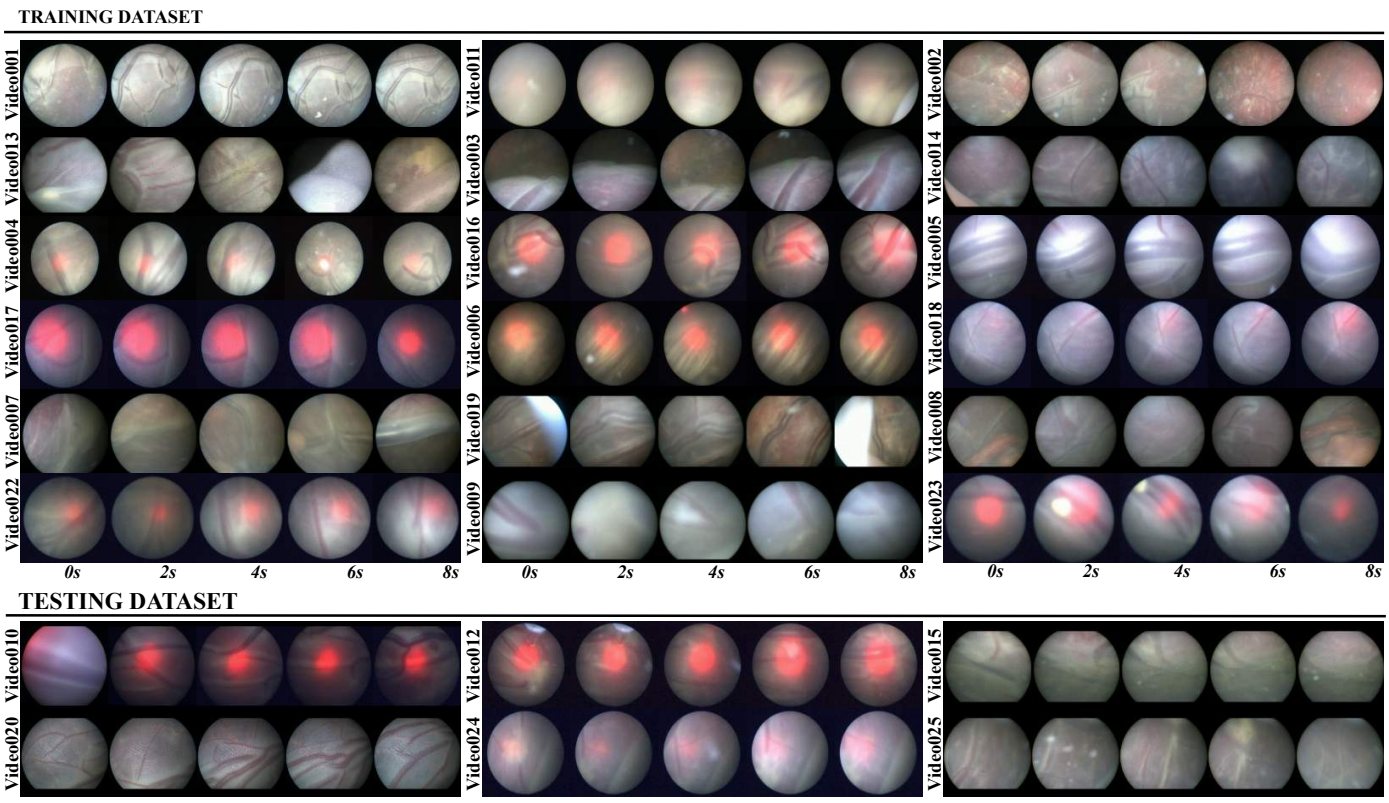


Fig. 5: Representative frames from training and test datasets at every 2 *seconds*. These clips are unannotated and the length of each clip mentioned in Table 2.

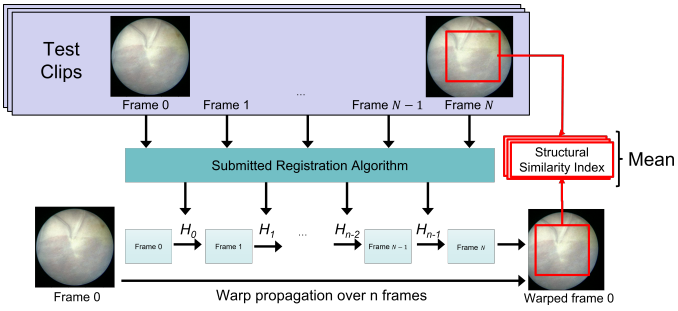


Fig. 6: Illustration of the N-frame SSIM evaluation metric from Bano et al. (2020a)

the consistency between them. The ultimate clinical goal of fetoscopic registration is to generate consistent, comprehensible and complete mosaics that map the placental surface and guide the surgeon. Considering adjacent frames will have large overlap along them, we evaluate the registration consistency between pairs of non-consecutive frames N frames apart that have a large overlap in field of view and present a clear view of the placental surface. Consider a source image I_i, a target image I_{i+n}, and a homography transformation $H_{i\to i+n}$ between them, we define the consistency s between these two images as:

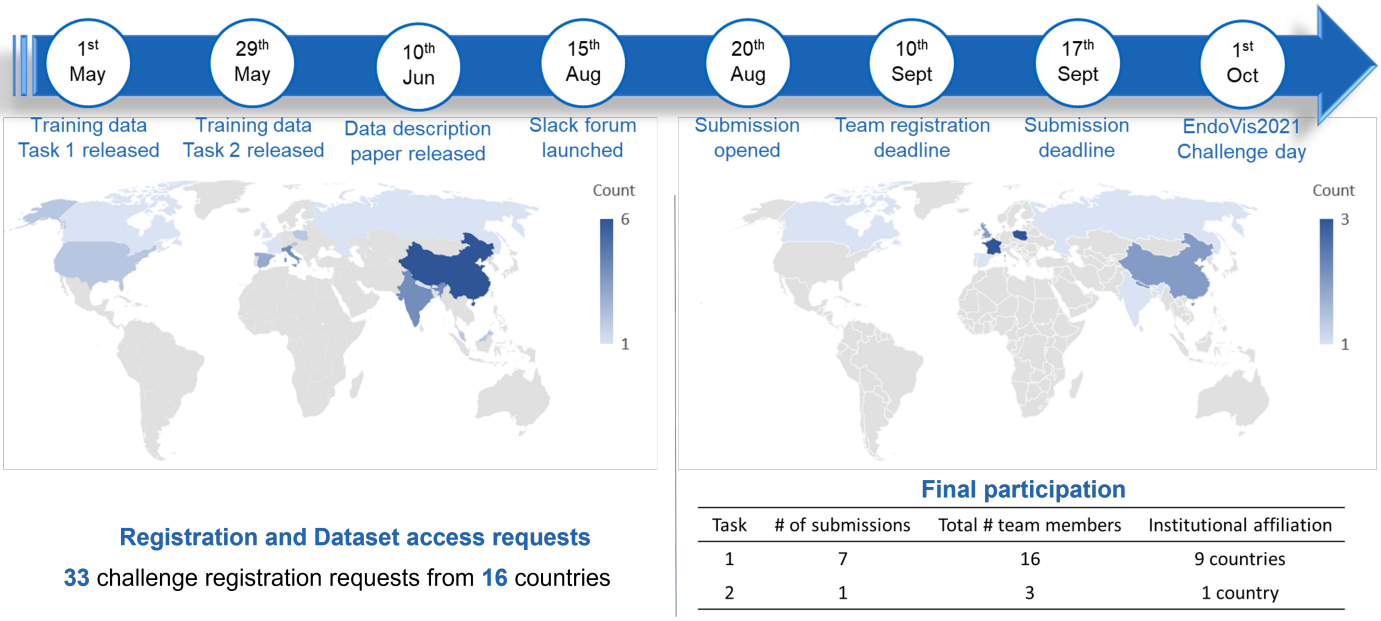
$$s_{i\to i+n} = \text{sim}(w(\tilde{I}_i, H_{i\to i+n}), \tilde{I}_{i+n}) \quad (2)$$

where sim is an image similarity metric that is computed based on target image and warped source image, and $\tilde{I}$ is a smoothed version of the image I. *Smoothing* $\tilde{I}$ is obtained by applying a 9×9 Gaussian filter with standard deviation of 2 to the original image I. This is fundamental to make the similarity metric robust to small outlier (e.g. particles) and image discretization artefacts. For computing the *similarity*, we start by determining the overlap region between the target $\tilde{I}$ and the warped source $w(\tilde{I}_i, H_{i\to i+n})$, taking into account their circular edges. If the overlap contains less than 25% of $\tilde{I}$, we consider that the registration failed as there will be no such cases in the evaluation pool. A rectangular crop is fit to the overlap, and the structural similarity index metric (SSIM) is calculated between the image pairs after having been smoothed, warped, and cropped.

3.3. Challenge Organization and Timeline

The challenge timeline and submission statistics are presented in Fig. 7. The challenge was announced on April, 1st 2021 through the FetReg2021 synapse 1 website. The training dataset for task 1 and task 2 were released on May, 1st and May, 29th, respectively. A challenge description paper Bano et al. (2021) that also included baseline method evaluation was also published. Additionally, a slack support forum was launched for faster communication with the participants. Docker submission was opened on August, 20th 2021 followed by team registration deadline of September, 10th and final submission deadline was set to September, the 17th.

The test dataset was not made available to the challenge participants to keep the comparison fair and avoid any misuse of the test data during training. Each participating team was required to make submissions as a docker container. The teams could submit multiple docker dockers during the submission time (from August 20th to September 17th 2021) to check the



Task	# of submissions	Total # team members	Institutional affiliation
1	7	16	9 countries
2	1	3	1 country

Fig. 7: FetReg2021 timeline and challenge participation statistics.

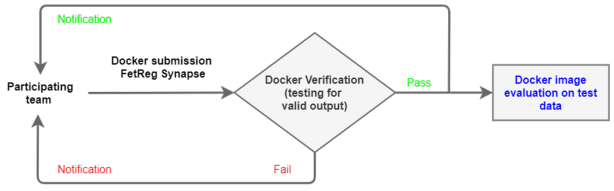


Fig. 8: FetReg2021 submission protocol illustrating the docker image verification protocol.

validity of the submitted docker. We provided the participants with docker examples for both tasks along with detailed submission guidelines through FetReg2021 GitHub repository [8]. The docker submission protocol is illustrated in Fig. 8. Each participating team submitted their docker through the Synapse platform. The submitted docker was verified for the validity of their output structure, i.e., they follow the same output format as requested and needed for the evaluation. Each participating team was then informed whether their submission passed the validity test. Each team was allowed to submit multiple dockers. However, only the last valid docker submission was used in the final evaluation.

We received 33 challenge registration requests from 16 different countries. A total of 13 team registration requests with a total number of 22 team members were received. For task 1, final submissions were received from 7 teams having 16 participants. For task 2, one single submission was received, probably because of the challenging nature of this task.

[8]FetReg2021 GitHub: `https://github.com/sophiabano/EndoVis-FetReg2021`

4. Summary of methods proposed by participating teams

In total 7 teams participated in the challenge. Out of these, one team did not qualify to be included in this article as the achieved performance was extremely low with a *mIoU* of 0.060. In this section, we summarize the methodology proposed by each participating team.

4.1. AQ-ENIB

Team AQ-ENIB are Abdul Qayyum, Abdesslam Benzinou, Moona Mazher and Fabrice Meriaudeau from ENIB (France), University Rovira i Virgili (Spain) and University of Bourgogne (France). The method proposed by AQ-ENIB implemented a model made by a dense encoder followed by a non-dense decoder. Dense encoder is chosen to enable the flow of information and gradients throughout the network, facilitating training convergence. The dense encoder consists of 5 dense blocks, each consisting of 6 dense layers followed by a transition layer. Each dense layer consists of 2 convolutional layers with batch normalization (BN) and ReLU activation functions. The first convolutional layer uses 1×1 kernels, while the second uses 3×3 kernels. The transition layers consists of a BN layer, a 1×1 convolutional layer, and a 2×2 average pooling layer. The transition layer helps to reduce feature-map size. The dense blocks in the encoder have an increasing number of feature maps at each encoder stage. The model is trained using 5-fold cross-validation. To compute the final prediction, test time augmentation is performed. This means that the model is fed with raw images and their augmented versions (using flipping and rotation with different angles). The model predicts, for each input, a segmentation mask. All the segmentation masks are ensembled using maximum majority voting.

4.2. BioPolimi

The team BioPolmini from Politecnico di Milano (Italy) are Chiara Lena, Jessica Biagioli, Gaia Romana and Ilaria

Anita Cintorrino. The model proposed by BioPolimi has a ResNet50 He et al. (2016) backbone followed by the U-Net Ronneberger et al. (2015) decoder for segmentation. The model is trained for 700 epochs with 6-fold cross-validation, using learning rate and batch size of 10^{-3} and 32, respectively. To be consistent with the FetReg Challenge baseline, training images are resized to 448×448 pixels. Data augmentation, consisting of random crop with size 256×256 pixels, random rotation (in range $(-45°, +45°)$), horizontal and vertical flip and random variation in brightness (in range $(-20\%, +20\%)$), is applied to the training data. During inference, testing images are cropped in patches of dimension 256x256 pixels. The final prediction is obtained by overlapping the prediction obtained for each patch with a stride equal to 8.

4.3. GRECHID

Team GRECHID is Daria Grechishnikova from Moscow State University (Russia). The method proposed by GRECHID consists of a U-Net with SERsNeXt50 backbone Hu et al. (2018) trained separately for each class (i.e., vessels, fetus and surgical tools). The *SEResNeXt50* backbone contains *Squeeze-and-Excitation* (SE) blocks, which allow the model to weigh adaptively each channel of SE blocks.

Before training, exact and near-duplicates were removed using an online software[9], obtaining 783 unique images from the original training dataset. Multi-label stratification split is performed to allocate images into train, test and validation sets. All the images are resized to 224×244 pixels. To improve model generalization, data augmentation is performed using horizontal and vertical flip, random rotation and flipping. The model is trained using Adam optimizer and cosine annealing with restart as learning rate scheduler, with a loss that combines Dice and modified cross-entropy losses. The modified cross-entropy loss has additional parameters to penalize either false positives and false negatives. Training is carried out in two stages. During the first stage, the model is trained for 30 epochs with a higher learning rate of 10^{-3}, then the learning rate is lowered to 10^{-5}. Cosine annealing with restart scheduling is used until best convergence.

A threshold-based post-processing is applied on the model output to remove spurious pixels.

4.4. OOF - Overoverfitting

Team OOF are Jing Jiao, Bizhe Bai and Yanyan Qiao from Fudan University (China), University of Toronto (Canada) and MicroPort Robotics. Team OOF used UNET++ Zhou et al. (2018) as the segmentation model. EfficientNetb-0 Tan and Le (2019) pre-trained on the Imagenet dataset is used as UNET++ encoder. To tackle illumination variability, median blur preprocessing and contrast limited Adaptive histogram equalization are applied to the images before feeding them to the model. Data augmentation, including random rotation, flip, and elastic transform, is applied during training. Adam optimizer with an initial learning rate of 10^{-4} is used. The learning rate increases exponentially with 5 warm-up epochs.

[9] https://github.com/idealo/imagededup

4.5. RREB

Team RREB are Binod Bhattarai, Rebati Raman Gaire, Ronast Subedi and Eduard Vazquez from University College London (UK), NepAL Applied Mathematics and Informatics Institute for Research (Nepal) and Redev Technology (UK). The model proposed by RREB uses U2-Net Qin et al. (2020) as segmentation network. A regressor branch is added on top of each decoder layer to learn the Histogram of Oriented Gradients (HoG) at different scales. The loss L minimised during the training is defined as:

$$L = \alpha \mathrm{CE}_{\mathrm{seg}} + \beta \mathrm{MSE}_{\mathrm{HoG}} \tag{3}$$

where $\alpha = 1$, $\mathrm{CE}_{\mathrm{seg}}$ is the cross-entropy loss for semantic segmentation, $\beta = 1$ and $\mathrm{MSE}_{\mathrm{HoG}}$ is the mean squared error of the HoG regressor.

All the images are resized to 448×448 pixels, and random crops of 256×256 are extracted. Random rotation between $(-45°, +45°)$, cropping at different corners and centres, and flipping are applied as data augmentation. The entire model is trained for 200000 iterations using Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and a batch size of 16. The initial learning rate is set to 0.0002 and then is halved at 75000, 125000, 175000 iterations. The proposed model is validated through cross-validation.

4.6. SANO

Team SANO from Warsaw University of Technology (Poland) are Szymon Płotka, Aneta Lisowska and Arkadiusz Sitek. This is the only team that participated in both tasks.

Segmentation. The model proposed by SANO is a Feature Pyramid Network (FPN) Lin et al. (2017) that uses ResNet-152 He et al. (2016) with pre-trained weights as backbone. The first convolutional layer has a 3-input channel, $n = 64$ feature maps, 7×7 kernel with stride = 2, and padding = 3. The following three convolutional blocks have $2n$,$4n$ and $32n$ feature maps. Our bottleneck consists of three convolutional blocks with BN. During training, the images are resized to 448×448 pixels and following augmentations are applied:

- color jitter (brightness = [0.8, 1.2], contrast = [0.8, 1.2], saturation = [0.8, 1.2], and hue = [-0.1, 0.1])
- random affine transformation (rotation = [-90, 90], translation = [0.2, 0.2], scale = [1, 2], shear = [-10, 10])
- horizontal and vertical flip.

The overall framework is trained with cross-entropy loss using a batch size of 4, Adam as optimizer with an initial learning rate of 10^{-4}, weight decay and step learning rate by 0.1, and cross-entropy loss. Validation is performed with 6-fold cross-validation.

Registration. The algorithm uses the channel corresponding to the placental vessel (PV) from the segmentation network and the original RGB images. The algorithm only models translation with the precision of 1 pixel. If frames are indexed by $i = 1, \ldots, t, \ldots, T$, the algorithm finds $T - 1$ translations between neighbouring frames. To compute the placenta vasculature (PV) image, softmax is applied to the raw output of the segmentation. The PV channel is extracted and multiplied by 255. A mask of non-zero pixels is computed from the raw image and applied to the PV image. The homography is then computed in two steps: The shift between PV images t and $t+1$ is computed using masked Fast Fourier Transform. Then, the rotation matrix between t and the shifted $t+1$ image $T+1_s$ is computed by minimizing the mean square error.

4.7. Baseline

As the baseline model, we trained a U-Net Ronneberger et al. (2015) with ResNet50 He et al. (2016) backbone as described in Bano et al. (2020a). Softmax activation is used at the final layer. Cross-entropy loss is computed and back-propagated during training. Before training, the images are first resized to 448×448 pixels. To perform data augmentation, at each iteration step, a patch of 256×256 pixels is extracted at a random position in the image. Each of the extracted patches is augmented by applying a random rotation in range $(-45°, +45°)$, horizontal and vertical flip, scaling with a factor in the range of $(-20\%, +20\%)$ and random variation in brightness $(-20\%, +20\%)$ and contrast $(-10\%, +10\%)$. Segmentation results are obtained by inference using 448×446 pixels resized input image. The baseline model is trained for 300 epochs on the training dataset. We create 6 folds, where each fold contains 3 procedures, to preserve as much variability as possible while keeping the number of samples in each fold approximately balanced. The final model is trained on the entire dataset, splitting videos in 80% for training and 20% for validation. The data is distributed to represent the same amount of variability in both subsets.

5. Quantitative and Qualitative Evaluation Results

5.1. Placental Scene Segmentation Task

We perform both quantitative and qualitative comparison to evaluate the performance of the submitted placental scene segmentation methods. Table 3 shows the *mIoU* for each team individually over each of the 6 test video samples and the overall *mIoU* over all videos. To test the rank stability, total number of times a team is ranked 1st on a video is also reported. Figure 9(a) shows the qualitative comparison of each team on each video and Fig. 9(b) shows the comparison of each team on individual segmentation classes.

The qualitative results for the placental scene segmentation task are presented in Fig. 10. Among the challenge participants, the best performing approach is that of RREB, which achieved an overall *mIoU* of 0.6411. RREB obtained the best performance for all videos, but Video010 and Video012, where AQ-ENIB and GRECHID were the best, respectively. RREB performed the best among participants for all the three classes, with median IoU for vessel and surgical tools that overcome 60%. However, RREB obtained poor results for fetus segmentation, with a median IoU lower than 40% with a large dispersion among images. As shown in Fig. 10(c) and (d), RREB meets challenges in presence of fetus and surgical tools. In the first case, RREB does not segment the fetus, while in the second the tool is segmented as fetus.

GRECHID scored second among all the participants, with a *mIoU* of 0.5865. As for RREB, GRECHID grants the best and lowest performance for surgical tools and vessels, respectively. Figure 10(b) and (f) show that GRECHID wrongly identifies and segments the fetus when it is not present in the field of view, while in Fig. 10(c), where the fetus is present, GRECHID does not segment it.

With an overall *mIoU* of 0.5741, SANO scored third, with the best performance achieved for vessels. SANO shows high variability in the IoU computed among frames for both fetus and surgical tools. Despite the generalised good visual performance among videos, SANO tends to underestimate the areas.

AQ-ENIB obtaines an overall *mIoU* of 0.5503 with the worst performance obtained with fetus segmentation. Despite the good performance for vessel segmentation, vessel area is often underestimated as shown in Fig. 10(b), (e) and (f).

BioPolimi and OOF show the worst performance with an *mIoU* of 0.3443 and 0.2526, respectively. OOF faces challenges also in images where one single vessel is present in the field of view, as shown in Fig. 10(a). Despite the low overall performance of BioPolimi, especially in tool and fetus segmentation, vessels are correctly segmented when visible and continuous (i.e., particles or specularities does not interrupt vessels surface), as shown in Fig. 10(d).

The baseline method is the best performing method achieving an overall *mIoU* of 0.6763, overcoming the performance of the challenge participants for all videos but Video024 and Video025 where RREB is the best performing method.

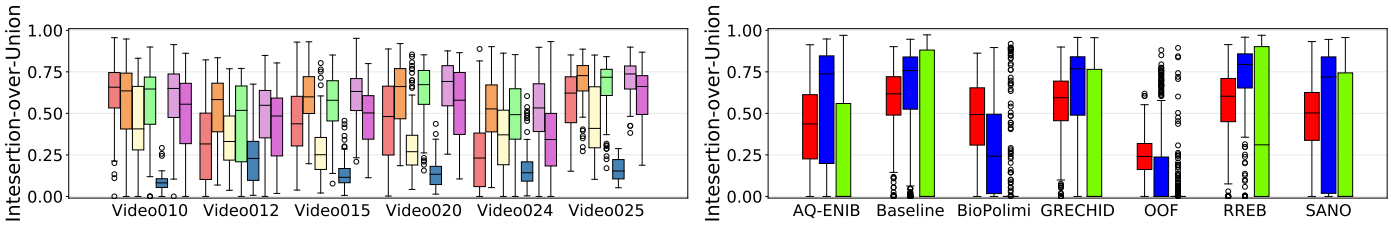
5.2. Registration for Mosaicking Task

Quantitative and qualitative results for the mosaicking task are presented in Table 4, Fig. 12 and Fig. 11.

The mosaics from the baseline and SANO methods and their 5-frame SSIM metric for every pair of images 5 frames apart in a sequence are shown Fig. 11 for all 6 test video clips. Both methods utilized placental vessel maps for estimating the transformation between adjacent frames. From the mosaic of Video010, we observe that both methods follow different strategies for registration. SANO utilized translation registration having fewer degrees of freedom, while baseline performs affine registration onf vessel having more degrees of freedom. Therefore, baseline is able to deal with perspective warpings while SANO's approach is unable to deal with perspective changes and overestimates translation to compensate such changes. As a result the 5-frame SSIM for SANO is lower compressed to the baseline in Video001. On Video012, both methods struggled to generate a meaningful mosaic but overall the baseline resulted in upper 5-frame SSIM metric compared to SANO (see Table 4). Video015 is an anterior placenta case in which the placental surface is not fronto-parallel to the camera.

Table 3: Results of segmentation for the Task 1 using test dataset.

Team name	Video010	Video012	Video015	Video020	Video024	Video025	Overall mIoU	# Video won
AQ-ENIB	0.5611	0.2745	0.4855	0.4848	0.3342	0.6414	0.5503	0
Baseline Bano et al. (2020a)	**0.5750**	**0.4122**	**0.6923**	**0.6757**	0.5514	0.7045	**0.6763**	4
BioPolimi	0.3891	0.2806	0.2718	0.2606	0.3666	0.3943	0.3443	0
GRECHID	0.4768	0.3792	0.5884	0.5744	0.3097	0.6534	0.5865	0
OOF	0.1874	0.1547	0.2745	0.2074	0.0872	0.3724	0.2526	0
RREB	0.5449	0.3765	0.6823	0.6191	**0.6443**	**0.7585**	0.6411	2
SANO	0.4682	0.3277	0.5201	0.5863	0.4132	0.6609	0.5741	0



(a) (b)

Fig. 9: Qualitative comparison showing (a) *mIoU* for each team on each video, and (b) overall *mIoU* for each team per segmentation class.

As a result, there is large perspective warping across different frames. SANO's approach failed here as it attempts to estimate only translation transformation. On the other hand, the baseline successfully estimate the warping through affine transformation resulting is better 5-frame SSIM metric. Qualitatively, SANO performs better on Video020 compared to the baseline, especialy in regions where vessels are visible and the mosaic remains bounded due to only translation transformation estimation. However, the error between 5 frames is particularly large for SANO as the warping are not accurate. Video024 and Video025 show interesting cases where in some frames there are no distinguishable structures like vessels (frame 90 in Video024 and frame 148 in Video025), hence both methods loose tracking intermediantly. Quantitatively, SANO's performance is slightly better than the baseline on Video024. Through rank stability test, we found that baseline performance was better in 5 out of 6 videos (se. Table 4).

Figure. 11 shows the qualitative comparison using 1 to 5 frame SSIM metric. We observe that with increasing frame distance, the error becomes large. In the case of SANO, video010 and video015 results in large drift even at 2-frame distance. As SANO used a translation transformation estimation, its error becomes quite large in all videos when oberving from 1 to 5 frames SSIM. The basline followed an affine transformation estimation, as a result, its errors appear to be relatively smaller than SANO, which mainly occurred when no visible vessels were present in the scene.

6. Discussion

An accurate placental semantic segmentation is necessary for better understanding and visualization of the fetoscopic environment; as a result this may facilitate surgeons in improved localization of the anastomoses and better surgical outcome. However, the high intra and inter-procedure variability remain a key challenge as only a small subset of images from each procedure were manually annotated for model training. Additionally, datasets captured from different clinical centres varies in terms of the resolution, imaging device and light source, making model generalization even more challenging. From the segmentation model results on individual 6 test videos, we observed large variability in the *mIoU* values of all methods (see Table 3). Note that Video 010, 012 and 024 are from Centre 2 and the remaining were acquired from Centre 1. The performance of RREB, i.e. the winning team, may be explained considering that RREB uses a multi-task approach to segment the vascular structures while regressing the HoG. We hypothesise that training a CNN to regress HoG may help the CNN to enhance poorly contrasted vessels. The poor performance of RREB for fetus segmentation may be attributed to the low number (293, i.e. 14.22% of training frames) and high variability of fetus frames used for training. Baseline method, which is also the top performing method, was relatively low on Centre 2 videos (average *mIoU* of 0.5130) compared to Centre 1 videos (average *mIoU* of 0.6910). Team RREB, the second best method, also performed poorly on Video012 (mIoU of 0.3765). There was no single method that outperformed on all 6 test video samples. This suggests that the proposed methods did not fully generalized to the dataset distributions from the two centres. To better model the variabilities in the dataset, either more annotated images would be needed for supervised learning. Limited annotation problems can also be addressed through pseudo labelling using semi-supervised learning techniques.

A reliable and consistent mosaic is needed for visualizing an increased field-of-view image of the placental environment. The two methods under comparison relied on accurate placental vessel segmentation for mosaicking. However, during fetoscopy, the placenta regions might appear either with very thin and weak vessels or no vessels at all. A segmentation algo-

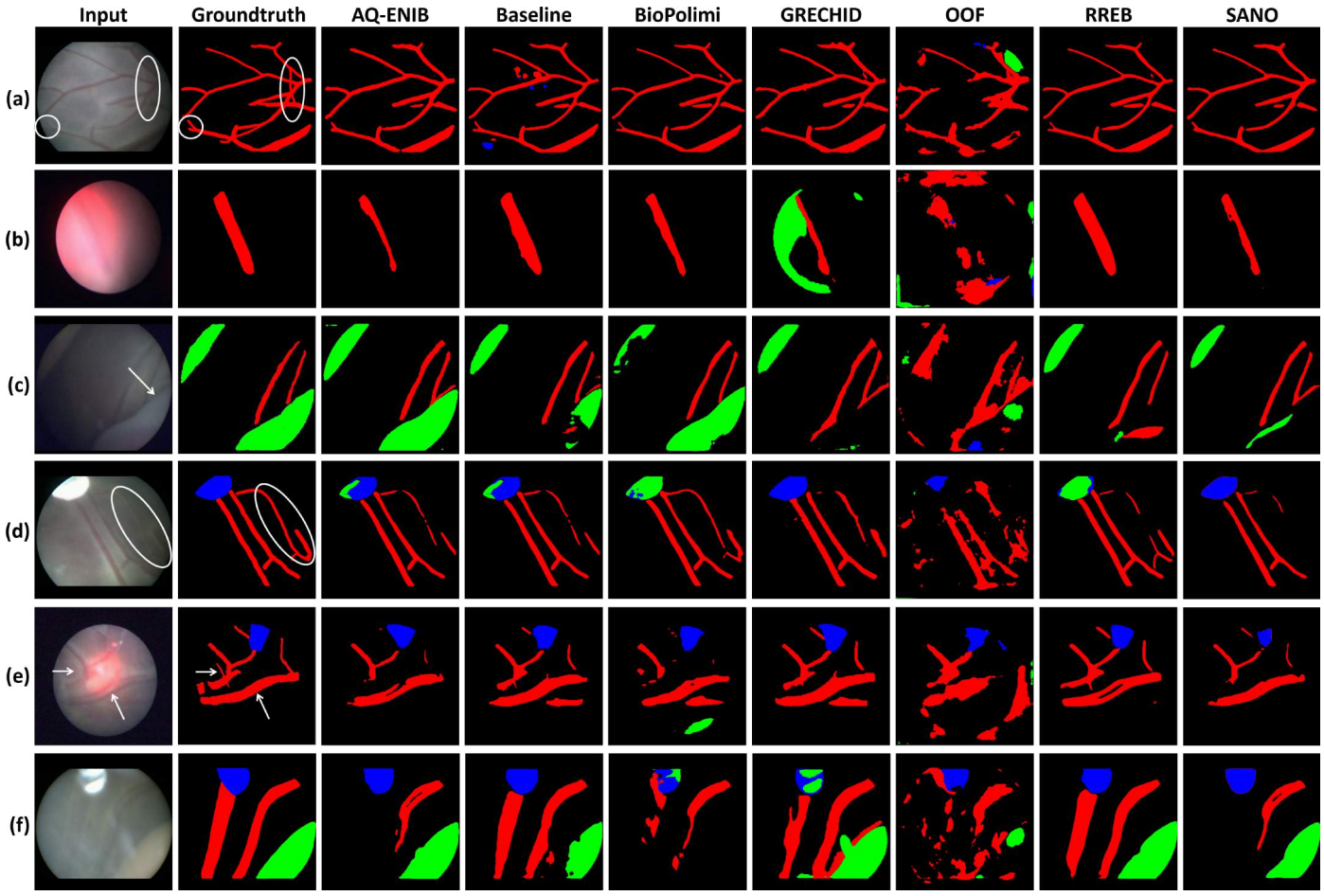


Fig. 10: Qualitative comparison of the 7 methods under analysis. Both baseline and RREB better generalize over the placental scene dataset. Baseline achieved better segmentation than RREB in (c), (d) and (e). OOF is the worst as it failed to generalize, wrongly segmented vessels and missed segmenting the fetus class.

Table 4: Results of Registration for the Task 2 using test video clips. Mean and Median of 5-frame-SSIM metric over individual video clips is reported.

Team name		Video010	Video012	Video015	Video020	Video024	Video025	Overall	# Video won
Baseline Bano et al. (2020a)	**Mean**	0.9048	0.9204	0.9695	0.9169	0.9336	0.9558	0.9348	5
	Median	0.9303	0.9330	0.9767	0.9301	0.9478	0.9712	0.9524	
SANO	**Mean**	0.8231	0.9164	0.9588	0.8276	0.9420	0.9234	0.9019	1
	Median	0.8837	0.9289	0.9746	0.8825	0.9563	0.9608	0.9434	

rithm may fail in these scenarios, especially when no vessels are visible, leading to failure in consecutive frames registration for mosaicking. This suggests that a registration algorithm should not solely rely on vessel segmentation predictions. More recent deep learning-based keypoint and matching approaches DeTone et al. (2018); Sarlin et al. (2020); Sun et al. (2021) could be useful in improving placental frame registration for mosaicking.

7. Conclusion

Surgical data science has the potential to enhance intraoperative imaging by providing better visualization of the surgical environment with increased field-of-view to support surgeon's decision during the procedure. Deep learning-based semantic segmentation algorithms can help in better understanding the fetoscopic placental scene during fetoscopy. However, large labelled datasets are required for training robust segmentation models. Through thr FetReg2021 challenge, which was part of the MICCAI2021 Endoscopic vision challenge, we contributed a large scale multicentre fetoscopy dataset containing data from 18 fetoscopy procedures for training and 6 fetoscopy procedures for testing. The test data was hidden from the challenge participants but followed similar distribution to the training dataset. The challenge focused on solving the task of placental semantic segmentation and fetoscopy video frame registration for mosaicking. The segmentation solutions presented by the participating teams achieved promising results though they were unable to beat the baseline method. Achieving generalizability remained an open question, and none of the methods outperformed in all test video samples. The contributed mosaicking approaches relied on accurate vessel segmentation and the presence of vessels in the fetoscopic placental view.

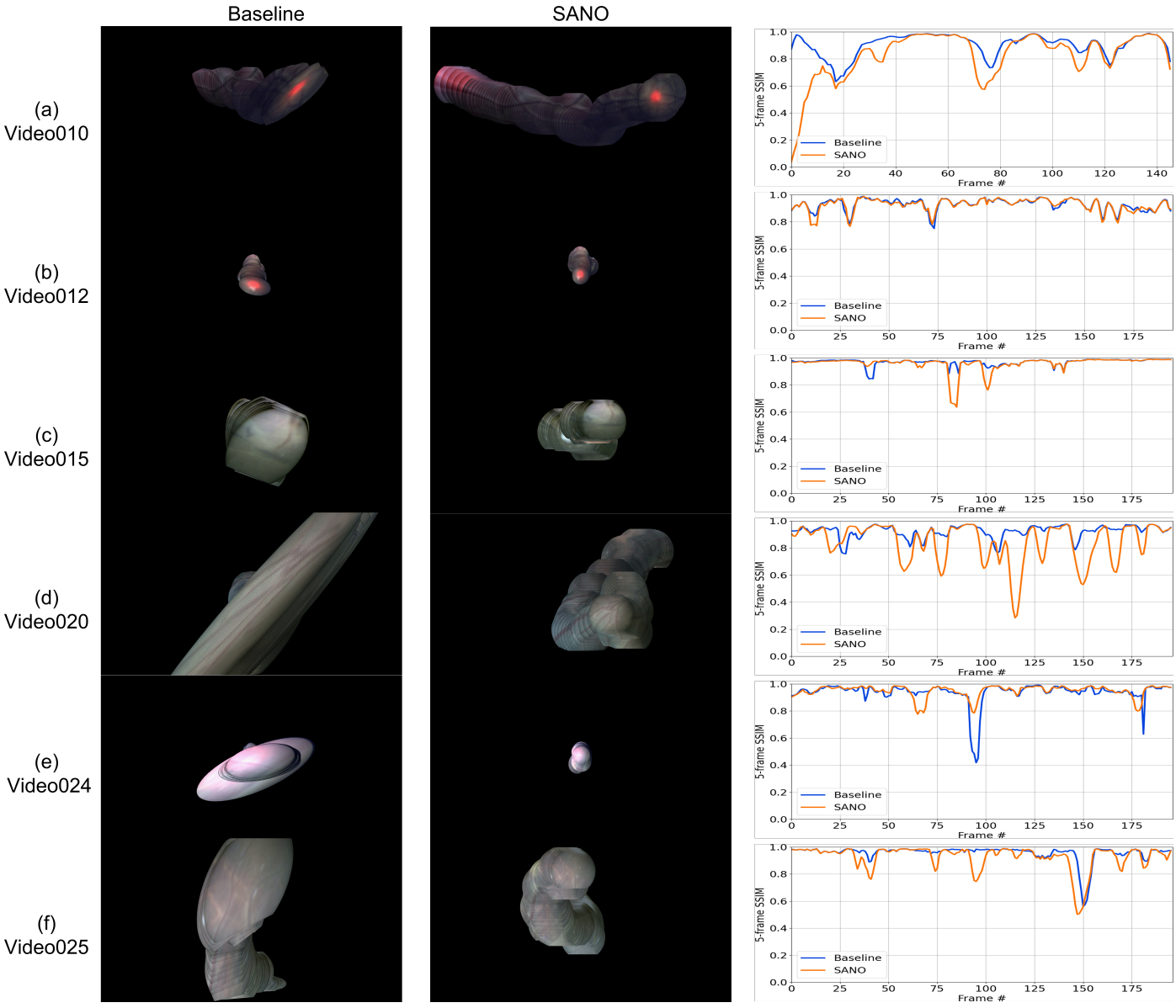


Fig. 11: Qualitative comparison of the Baseline Bano et al. (2020a) and SANO methods showing (first column) generated mosaics from the Baseline method, (2nd column) generated mosaics from the SANO method, and (3rd column) 5-frame SSIM per frame for both methods. Baseline performance is better in all videos except Video020.

Through the FetReg2021 challenge, we contributed a benchmark dataset for advancing the research in fetoscopic mosaicking.

Acknowledgement

We are grateful to NVIDIA¡ Medtronic and E4 Computing for sponsoring the FetReg2021 challenge. This work was partly supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) at UCL (203145Z/16/Z), EPSRC(EP/P027938/1, EP/R004080/1,NS/A000027/1), the H2020 FET (GA863146)and Wellcome [WT101957]. For the purpose of open access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission.

References

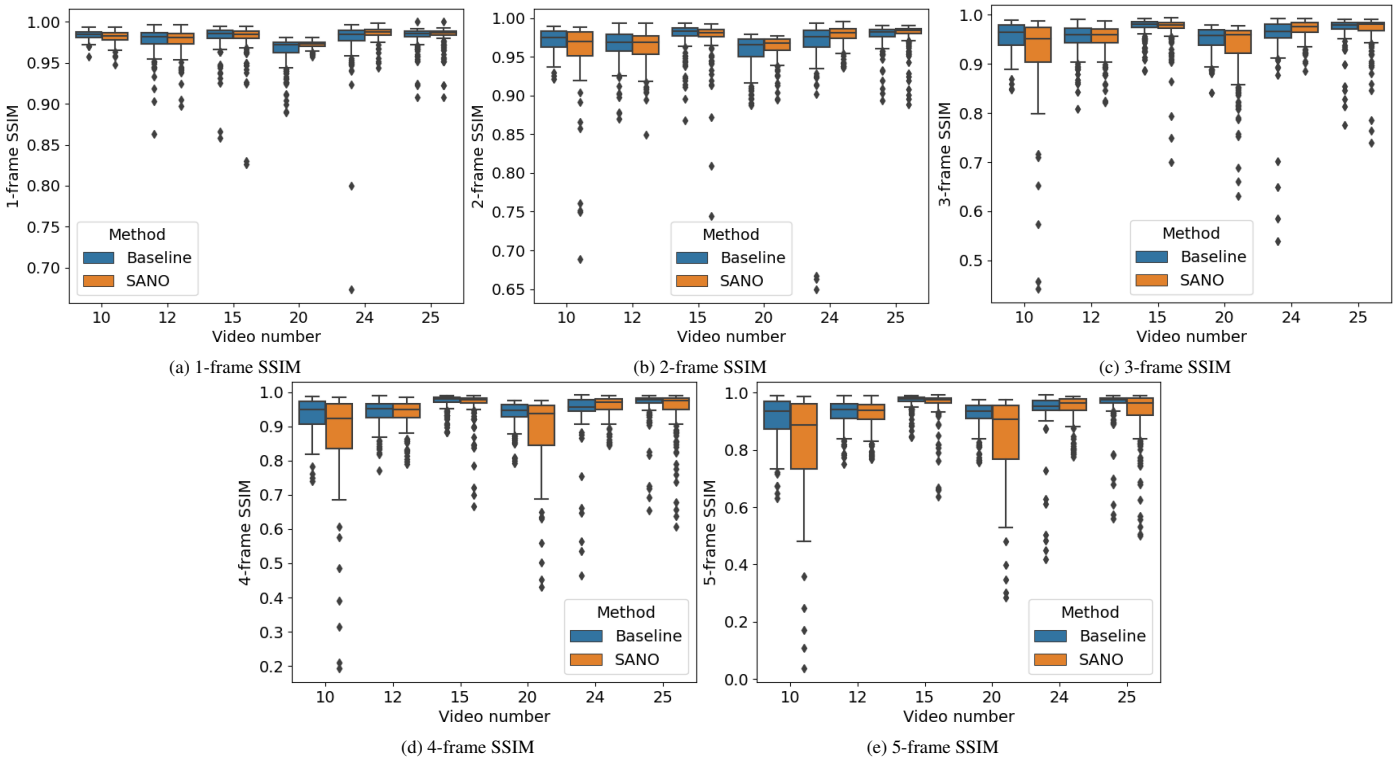
Alabi, Oluwatosin andBano, S., Vasconcelos, F., L. David, A., Deprest, J., Stoyanov, D., 2022. Robust Fetoscopic Mosaicking from Deep Learned Flow Fields. International Journal of Computer Assisted Radiology and Surgery .

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, SPIE. p. 79621L. doi:10.1117/12.878343.

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., Vasconcelos, F., Shepherd, L.M., Vander Poorten, E., Vercauteren, T., Ourselin, S., David, A.L., Deprest, J., Stoyanov, D., 2020a. Deep placental vessel segmentation for fetoscopic mosaicking, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 763–773.

Bano, S., Vasconcelos, F., Tella Amo, M., Dwyer, G., Gruijthuijsen, C., Deprest, J., Ourselin, S., Vander Poorten, E., Vercauteren, T., Stoyanov,



(a) 1-frame SSIM (b) 2-frame SSIM (c) 3-frame SSIM

(d) 4-frame SSIM (e) 5-frame SSIM

Fig. 12: Quantitative comparison of the Baseline Bano et al. (2020a) and SANO methods using the N-frame SSIM metric.

D., 2019. Deep sequential mosaicking of fetoscopic videos, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 311–319. URL: http://arxiv.org/abs/1907.06543http://dx.doi.org/10.1007/978-3-030-32239-7_35, doi:10.1007/978-3-030-32239-7{_}35.

Bano, S., Vasconcelos, F., Tella-Amo, M., Dwyer, G., Gruijthuijsen, C., Vander Poorten, E., Vercauteren, T., Ourselin, S., Deprest, J., Stoyanov, D., 2020b. Deep learning-based fetoscopic mosaicking for field-of-view expansion. International Journal of Computer Assisted Radiology and Surgery doi:10.1007/s11548-020-02242-8.

Bano, S., Vasconcelos, F., Vander Poorten, E., Vercauteren, T., Ourselin, S., Deprest, J., Stoyanov, D., 2020c. FetNet: a recurrent convolutional network for occlusion identification in fetoscopic videos. International Journal of Computer Assisted Radiology and Surgery 15, 791–801. doi:10.1007/s11548-020-02169-0.

Baschat, A., Chmait, R.H., Deprest, J., Gratacós, E., Hecher, K., Kontopoulos, E., Quintero, R., Skupski, D.W., Valsky, D.V., Ville, Y., 2011. Twin-to-twin transfusion syndrome (TTTS). Journal of Perinatal Medicine 39, 107–112. doi:10.1515/JPM.2010.147.

Baud, D., Windrim, R., Keunen, J., Kelly, E.N., Shah, P., Van Mieghem, T., Seaward, P.G.R., Ryan, G., 2013. Fetoscopic laser therapy for twin-twin transfusion syndrome before 17 and after 26 weeks' gestation. American Journal of Obstetrics and Gynecology 208, 1–197. doi:10.1016/j.ajog.2012.11.027.

Bian, J., Lin, W.Y., Matsushita, Y., Yeung, S.K., Nguyen, T.D., Cheng, M.M., 2017. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4181–4190.

Casella, A., Moccia, S., Frontoni, E., Paladini, D., De Momi, E., Mattos, L.S., 2020. Inter-foetus Membrane Segmentation for TTTS Using Adversarial Networks. Annals of Biomedical Engineering 48, 848–859. doi:10.1007/s10439-019-02424-9.

Casella, A., Moccia, S., Paladini, D., Frontoni, E., De Momi, E., Mattos, L., 2021. A shape-constraint adversarial framework with instance-normalized spatio-temporal features for inter-fetal membrane segmentation. Medical Image Analysis 70, 102008.

Chang, J.M., Huynh, N., Vazquez, M., Salafia, C., 2013. Vessel enhancement with multiscale and curvilinear filter matching for placenta images, in: International Conference on Systems, Signals, and Image Processing, IEEE Computer Society. pp. 125–128. doi:10.1109/IWSSIP.2013.6623469.

Cincotta, R., Kumar, S., 2016. Future Directions in the Management of Twin-to-Twin Transfusion Syndrome. Twin Research and Human Genetics 19, 285–291. doi:10.1017/thg.2016.32.

Daga, P., Chadebecq, F., Shakir, D.I., Herrera, L.C.G., Tella, M., Dwyer, G., David, A.L., Deprest, J., Stoyanov, D., Vercauteren, T., Ourselin, S., 2016. Real-time mosaicing of fetoscopic videos using SIFT, in: Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling, SPIE. p. 97861R. doi:10.1117/12.2217172.

Deprest, J.A., Flake, A.W., Gratacos, E., Ville, Y., Hecher, K., Nicolaides, K., Johnson, M.P., Luks, F.I., Adzick, N.S., Harrison, M.R., 2010. The making of fetal surgery. doi:10.1002/pd.2571.

DeTone, D., Malisiewicz, T., Rabinovich, A., 2016. Deep image homography estimation. arXiv preprint arXiv:1606.03798 .

DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 224–236.

Gaisser, F., Jonker, P.P., Chiba, T., 2016. Image Registration for Placenta Reconstruction, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society. pp. 473–480. doi:10.1109/CVPRW.2016.66.

Gaisser, F., Peeters, S.H., Lenseigne, B., Jonker, P.P., Oepkes, D., 2017. Fetoscopic panorama reconstruction: Moving from ex-vivo to in-vivo. volume 723 of *Communications in Computer and Information Science*. Springer International Publishing, Cham. URL: http://link.springer.com/10.1007/978-3-319-60964-5, doi:10.1007/978-3-319-60964-5{_}51.

Gaisser, F., Peeters, S.H., Lenseigne, B.A., Jonker, P.P., Oepkes, D., 2018. Stable image registration for in-vivo fetoscopic panorama reconstruction. Journal of Imaging 4. doi:10.3390/jimaging4010024.

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., Sun, G., 2018. Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. doi:10.1109/CVPR.2018.00745.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2017.

Flownet 2.0: Evolution of optical flow estimation with deep networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2462–2470.

Lewi, L., Deprest, J., Hecher, K., 2013. The vascular anastomoses in monochorionic twin pregnancies and their clinical consequences. doi:10.1016/j.ajog.2012.09.025.

Li, L., Bano, S., Deprest, J., David, A.L., Stoyanov, D., Vasconcelos, F., 2021. Globally optimal fetoscopic mosaicking based on pose graph optimisation with affine constraints. IEEE Robotics and Automation Letters 6, 7831–7838.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.

Lopriore, E., Middeldorp, J.M., Oepkes, D., Klumper, F.J., Walther, F.J., Vandenbussche, F.P., 2007. Residual Anastomoses After Fetoscopic Laser Surgery in Twin-to-Twin Transfusion Syndrome: Frequency, Associated Risks and Outcome. Placenta 28, 204–208. doi:10.1016/j.placenta.2006.03.005.

Mair, E., Hager, G.D., Burschka, D., Suppa, M., Hirzinger, G., 2010. Adaptive and generic corner detection based on the accelerated segment test, in: European conference on Computer vision, Springer. pp. 183–196.

Maselli, K.M., Badillo, A., 2016. Advances in fetal surgery. doi:10.21037/atm.2016.10.34.

Moccia, S., De Momi, E., El Hadji, S., Mattos, L.S., 2018. Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics. Computer Methods and Programs in Biomedicine 158, 71–91.

Nguyen, T., Chen, S.W., Shivakumar, S.S., Taylor, C.J., Kumar, V., 2018. Unsupervised deep homography: A fast and robust homography estimation model. IEEE Robotics and Automation Letters 3, 2346–2353.

Peter, L., Tella-Amo, M., Shakir, D.I., Attilakos, G., Wimalasundera, R., Deprest, J., Ourselin, S., Vercauteren, T., 2018. Retrieval and registration of long-range overlapping frames for scalable mosaicking of in vivo fetoscopy. International Journal of Computer Assisted Radiology and Surgery 13, 713–720. doi:10.1007/s11548-018-1728-4.

Pratt, R., Deprest, J., Vercauteren, T., Ourselin, S., David, A.L., 2015. Computer-assisted surgical planning and intraoperative guidance in fetal surgery: A systematic review. doi:10.1002/pd.4660.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O.R., Jagersand, M., 2020. U2-net: Going deeper with nested u-structure for salient object detection. Pattern Recognition 106, 107404.

Quintero, R.A., Ishii, K., Chmait, R.H., Bornick, P.W., Allen, M.H., Kontopoulos, E.V., 2007. Sequential selective laser photocoagulation of communicating vessels in twin-twin transfusion syndrome. Journal of Maternal-Fetal and Neonatal Medicine 20, 763–768. doi:10.1080/14767050701591827.

Reeff, M., Gerhard, F., Cattin, P., Székely, G., 2006. Mosaicing of endoscopic placenta images, in: INFORMATIK 2006 - Informatik fur Menschen, Beitrage der 36. Jahrestagung der Gesellschaft fur Informatik e.V. (GI), pp. 467–474. URL: https://www.researchgate.net/publication/221384260.

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.

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. doi:10.1007/s11548-018-1886-4.

Sadda, P., Onofrey, J.A., Bahtiyar, M.O., Papademetris, X., 2018. Better feature matching for placental panorama construction, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag. pp. 128–137. doi:10.1007/978-3-030-00807-9{_}13.

Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4938–4947.

Senat, M.V., Deprest, J., Boulvain, M., Paupe, A., Winer, N., Ville, Y., 2004. Endoscopic Laser Surgery versus Serial Amnioreduction for Severe Twin-to-Twin Transfusion Syndrome. New England Journal of Medicine 351, 136–144. doi:10.1056/nejmoa032597.

Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X., 2021. LoFTR: Detector-free local feature matching with transformers, in: Conference on Computer Vision and Pattern Recognition, pp. 8922–8931.

Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International conference on machine learning, PMLR. pp. 6105–6114.

Tella, M., Daga, P., Chadebecq, F., Thompson, S., Shakir, D.I., Dwyer, G., Wimalasundera, R., Deprest, J., Stoyanov, D., Vercauteren, T., Ourselin, S., 2016. A Combined em and Visual Tracking Probabilistic Model for Robust Mosaicking: Application to Fetoscopy, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society. pp. 524–532. doi:10.1109/CVPRW.2016.72.

Tella-Amo, M., Peter, L., Shakir, D.I., Deprest, J., Stoyanov, D., Iglesias, J.E., Vercauteren, T., Ourselin, S., 2018. Probabilistic visual and electromagnetic data fusion for robust drift-free sequential mosaicking: application to fetoscopy. Journal of Medical Imaging 5, 1. doi:10.1117/1.jmi.5.2.021217.

Tella-Amo, M., Peter, L., Shakir, D.I., Deprest, J., Stoyanov, D., Vercauteren, T., Ourselin, S., 2019. Pruning strategies for efficient online globally consistent mosaicking in fetoscopy. Journal of Medical Imaging 6, 1. doi:10.1117/1.jmi.6.3.035001.

Vasconcelos, F., Brandão, P., Vercauteren, T., Ourselin, S., Deprest, J., Peebles, D., Stoyanov, D., 2018. Towards computer-assisted TTTS: Laser ablation detection for workflow segmentation from fetoscopic video. International Journal of Computer Assisted Radiology and Surgery 13, 1661–1670. doi:10.1007/s11548-018-1813-8.

Yang, L., Wang, J., Ando, T., Kubota, A., Yamashita, H., Sakuma, I., Chiba, T., Kobayashi, E., 2016. Towards scene adaptive image correspondence for placental vasculature mosaic in computer assisted fetoscopic procedures. The International Journal of Medical Robotics and Computer Assisted Surgery 12, 375–386. doi:10.1002/rcs.1700.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation, in: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, pp. 3–11.