FetNet: A Recurrent Convolutional Network for Occlusion Identification in Fetoscopic Videos

Sophia Bano, Francisco Vasconcelos, Jan Deprest, Sebastien Ourselin, Emmanuel Vander Poorten, Tom Vercauteren, and Danail Stoyanov

Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS),
University College London, London, UK

✉sophia.bano@ucl.ac.uk

Twin-to-Twin Transfusion Syndrome (TTTS)

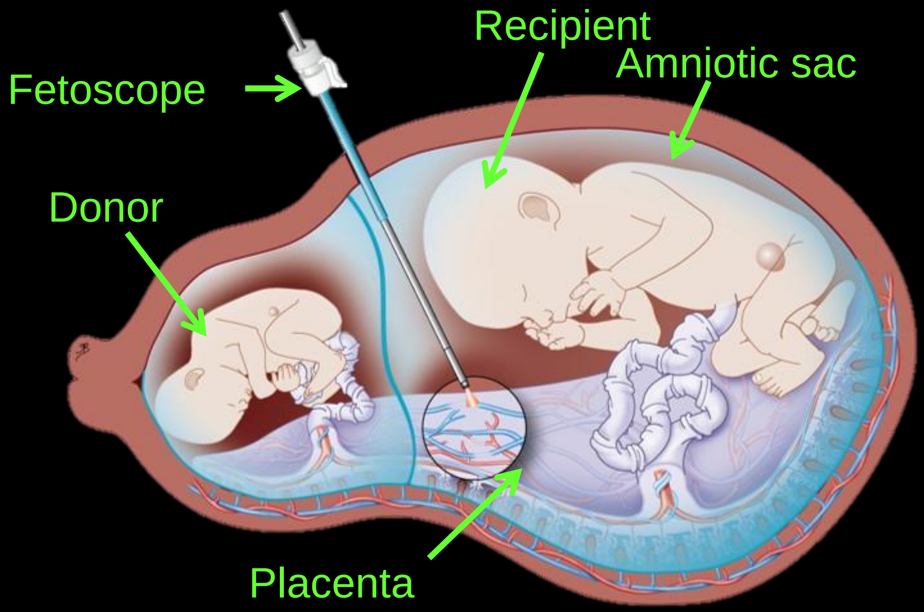
TTTS is a fetal anomaly affecting identical twins sharing a monochorionic placenta

Consequence

- Unbalanced flow of blood
- Donor may experience much slower growth
- Recipient at risk of heart failure

Common treatment

- Fetoscopic laser photocoagulation



Fetoscopic Laser Photocoagulation

Fetoscopic laser photocoagulation is a minimally invasive surgery

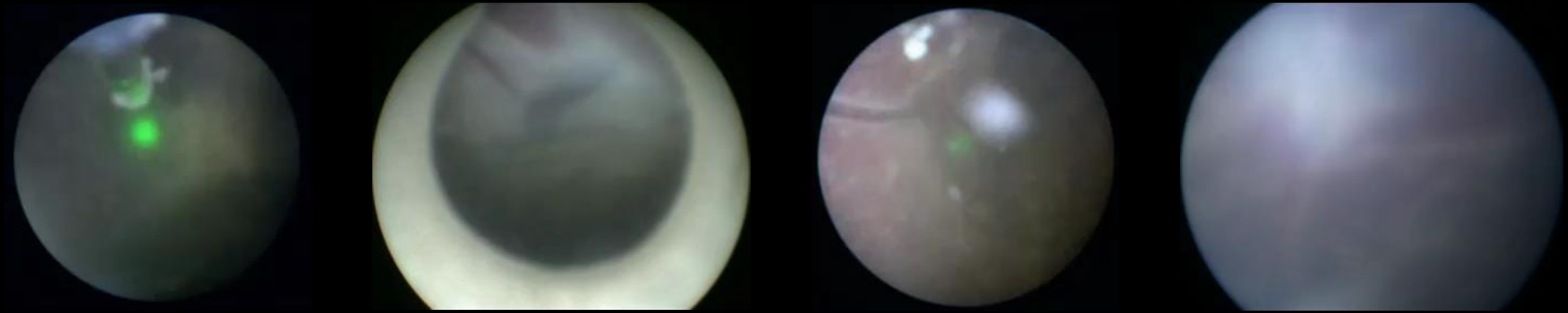
- Visually explore the placenta using fetoscopic camera to identify vascular anastomose

- Localize the target vessels and use the laser to ablate them

Safe procedure requires

- Clear view of the placenta
- Clear path between the ablation tool and the target vessels

Fetoscopic Video Analysis: Common Challenges



Difficult visual conditions

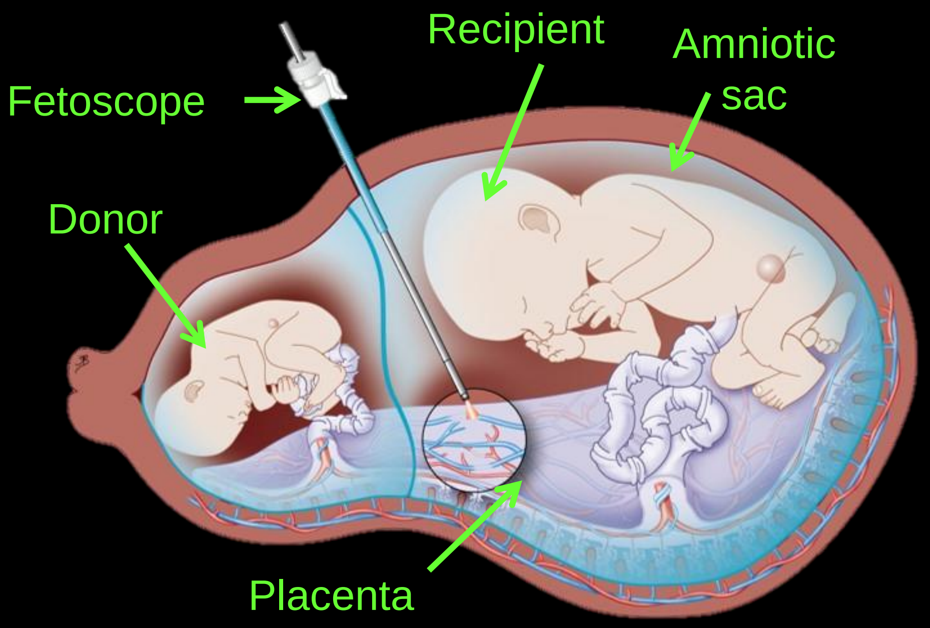
- Poor visibility (low resolution, low illumination, amniotic fluid turbidity)
- Occlusion due to the fetus and working channel port
- Specular highlights resulting in glare and reflection

Fetoscopic Event Identification - Motivation

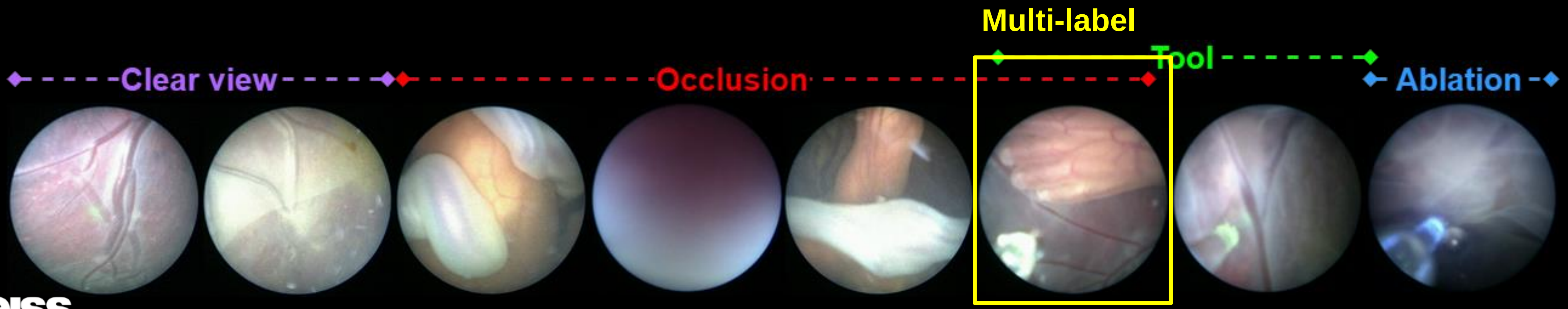
UCL

Identifying fetoscopic events can

- Assist surgeons during the TTTS procedure
- Provide context for navigation and mapping



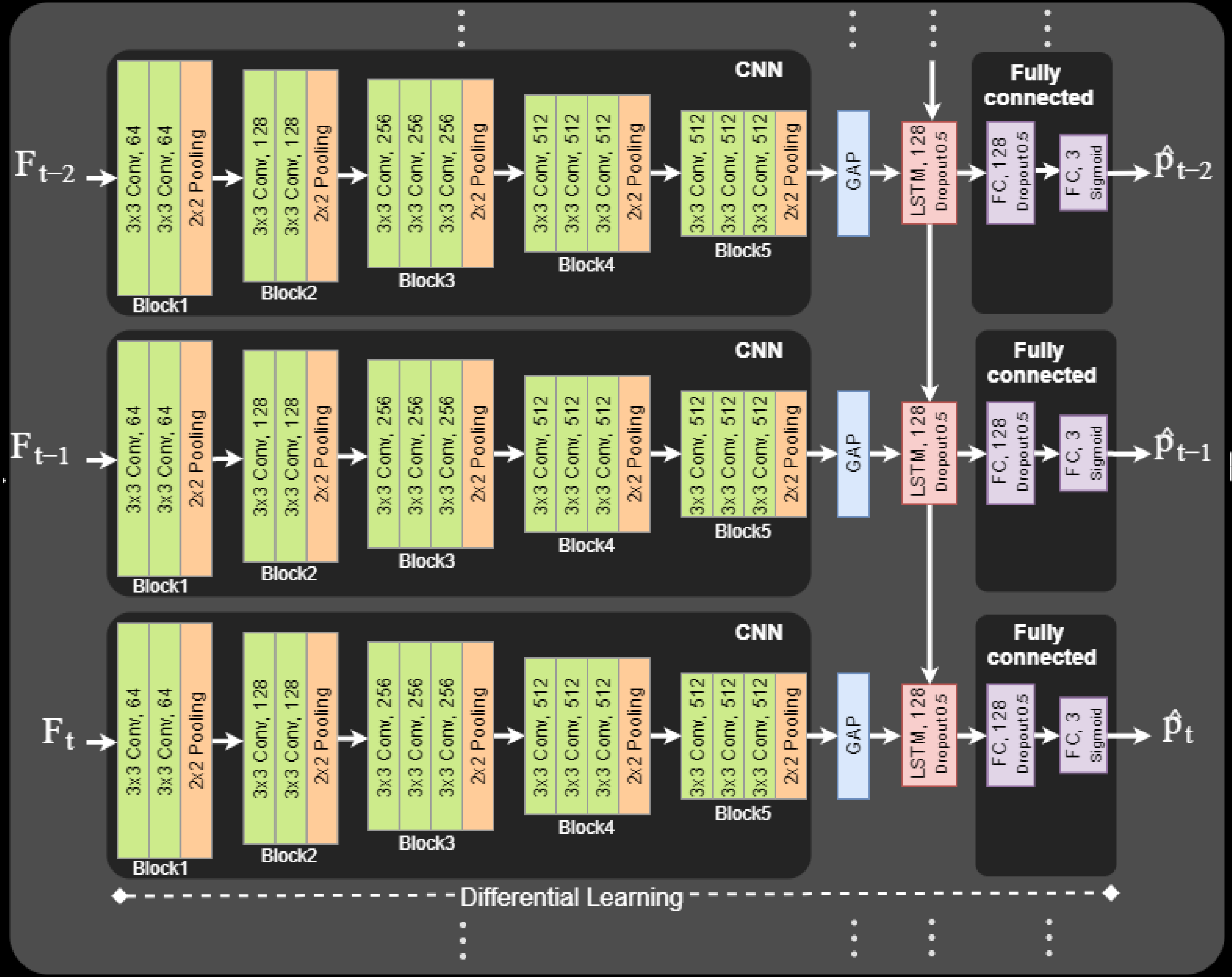
Four event labels created are



weiss

Proposed FetNet Architecture

A recurrent convolutional network for spatio-temporal identification of fetoscopic events



weiss

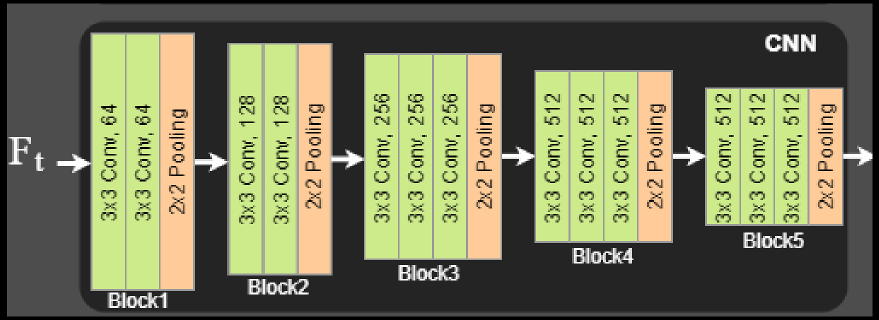
Proposed FetNet Architecture

A recurrent convolutional network for spatio-temporal identification of fetoscopic events

- Integrates:
  - Convolutional Neutral Network (CNN) for encoding spatial cues

CNN

Input frame



CNN features

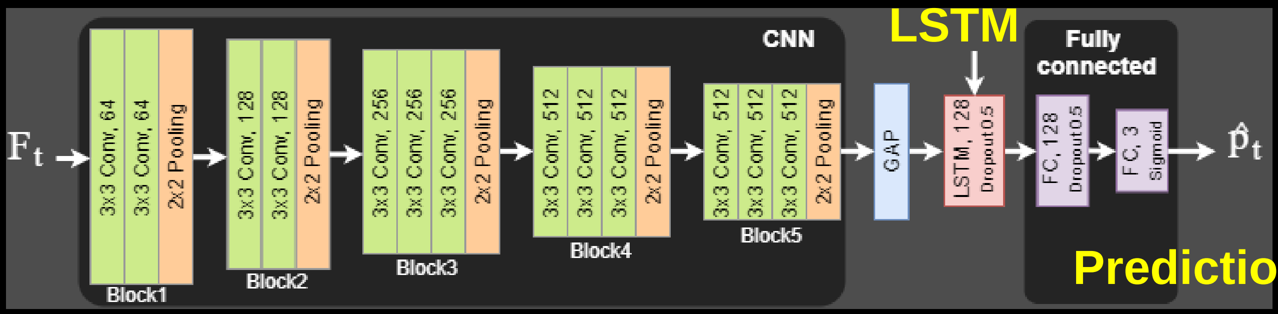
Proposed FetNet Architecture

A recurrent convolutional network for spatio-temporal identification of fetoscopic events

- Integrates:
  - Convolutional Neutral Network (CNN) for encoding spatial cues
  - Long Short-Term Memory (LSTM) for encoding temporal cues

- Multi-labels handled using sigmoid activation
  - Independent prediction probabilities

Input frame



Prediction

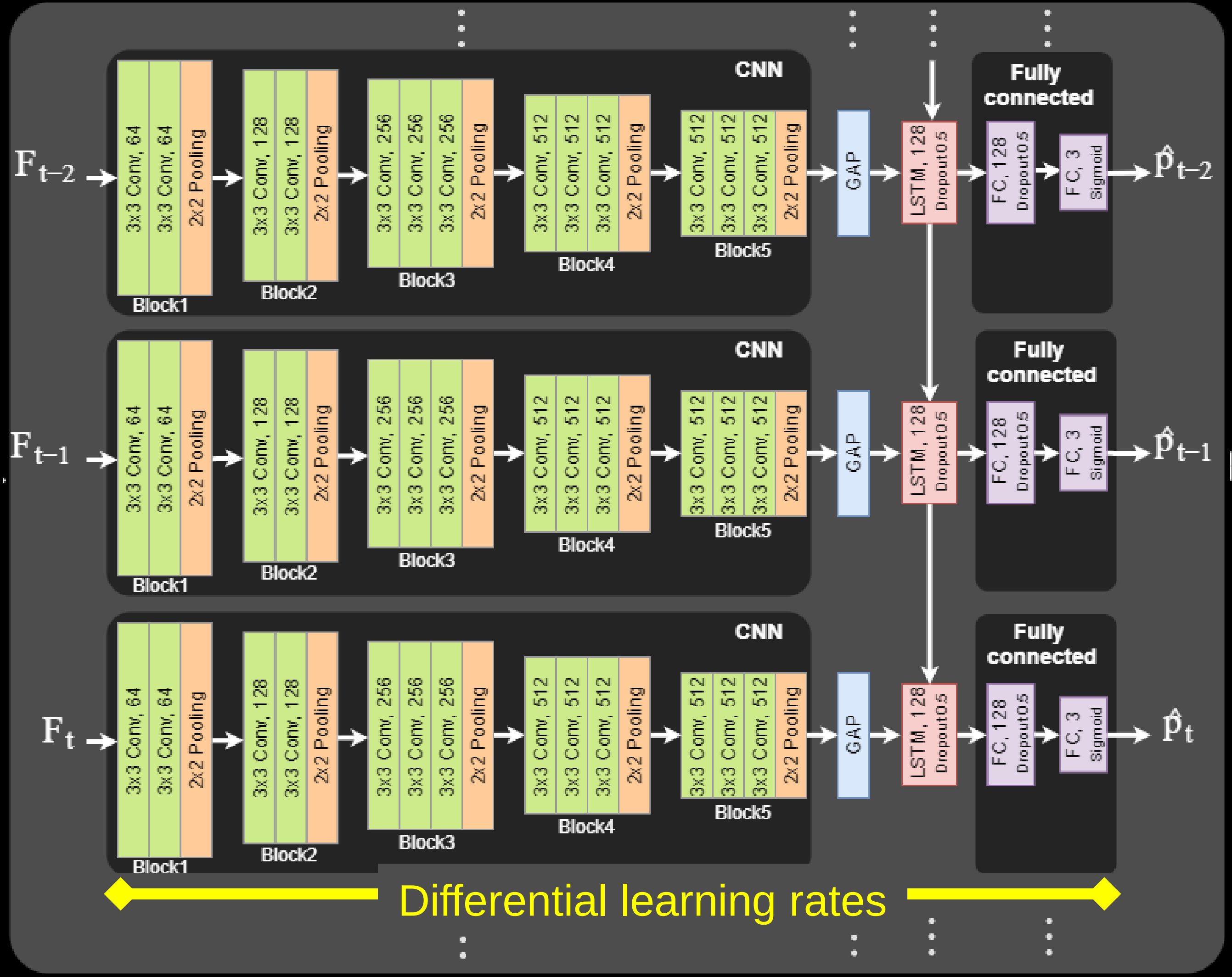
Proposed FetNet Architecture

A recurrent convolutional network for spatio-temporal identification of fetoscopic events

- Integrates:
  - Convolutional Neutral Network (CNN) for encoding spatial cues
  - Long Short-Term Memory (LSTM) for encoding temporal cues

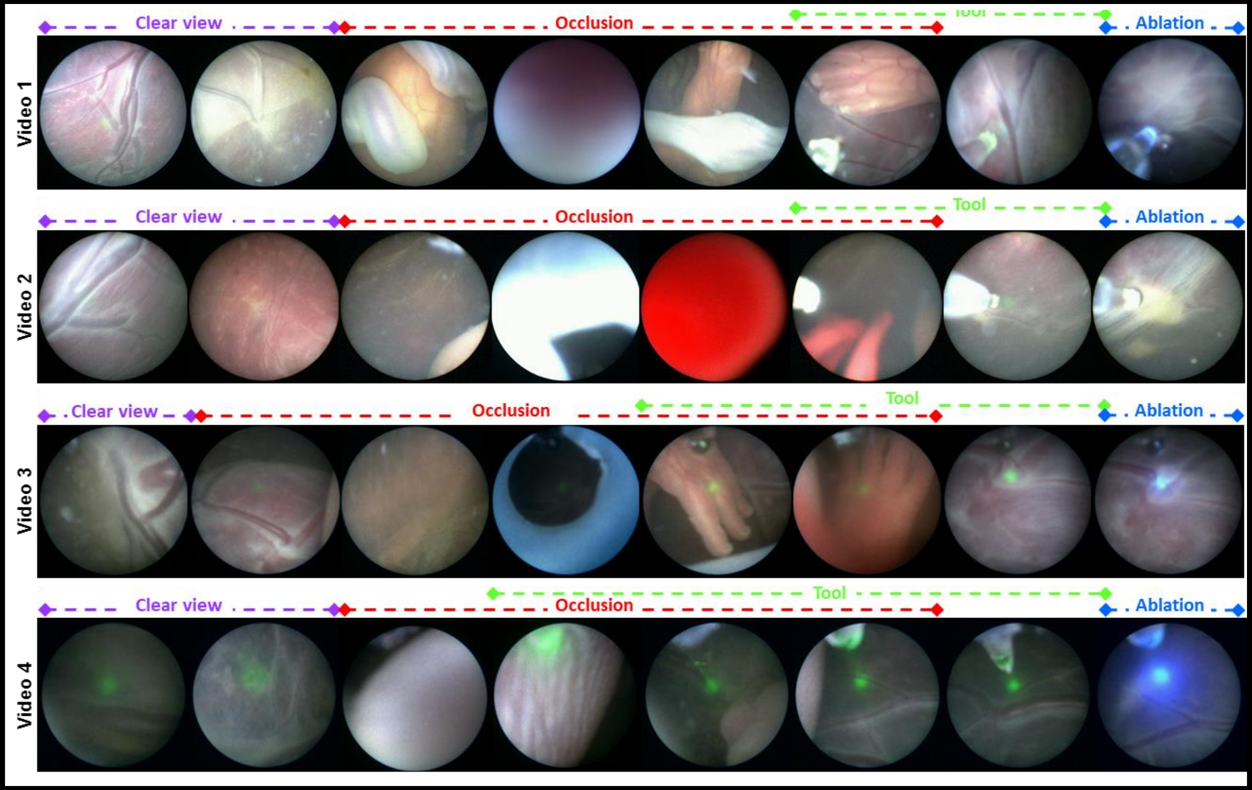
- Multi-labels handled using sigmoid activation
  - Independent prediction probability

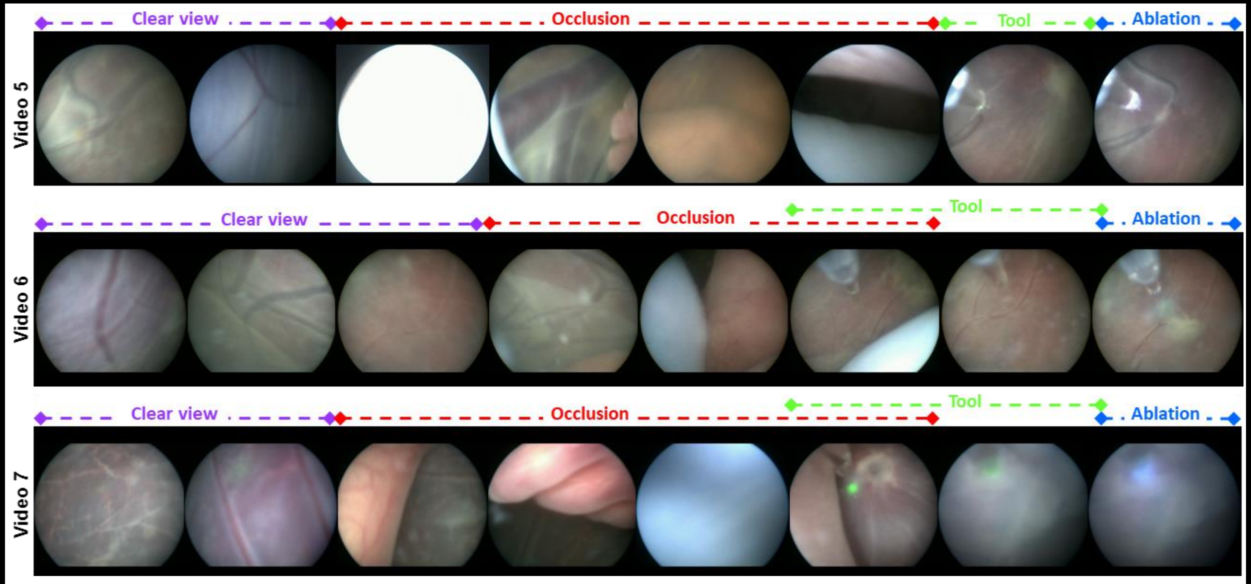
- Differential learning rates
  - Pretrained CNN weights



Dataset collection and annotation

- Seven fetoscopic videos from different patients
- Average duration of each video is *800s*
- Frame-level manual annotation for events





Quantitative Analysis and Comparison

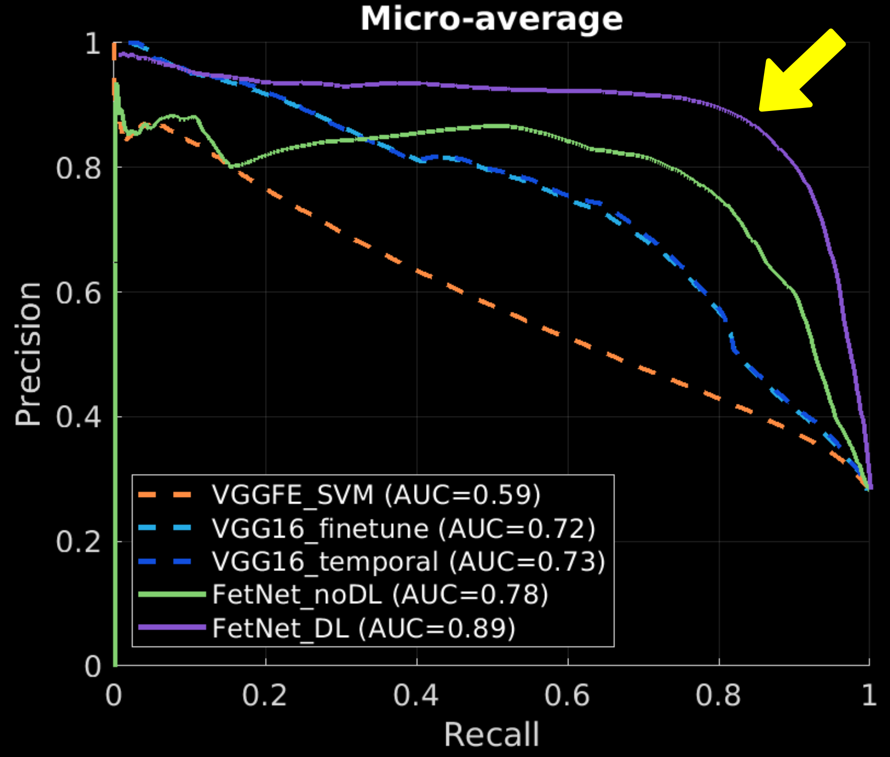
- 7-fold cross-validation

Comparison Methods	
Ablation_detect	Fetoscopic ablation detection method [Vasconcelos_IJCAR2018]
VGGFE_SVM	CNN features with SVM classifier [Cadene_arXiv2016]
VGG16_fine	Fine-tuning of VGG16 [Simonyan_ICLR2015]
VGG16_temporal	Fine-tuning and temporal smoothing of VGG16 [Cadene_arXiv2016]
FetNet_noDL	Proposed without differential learning
FetNet_DL	Proposed with differential learning

Quantitative Analysis and Comparison

- 7-fold cross-validation

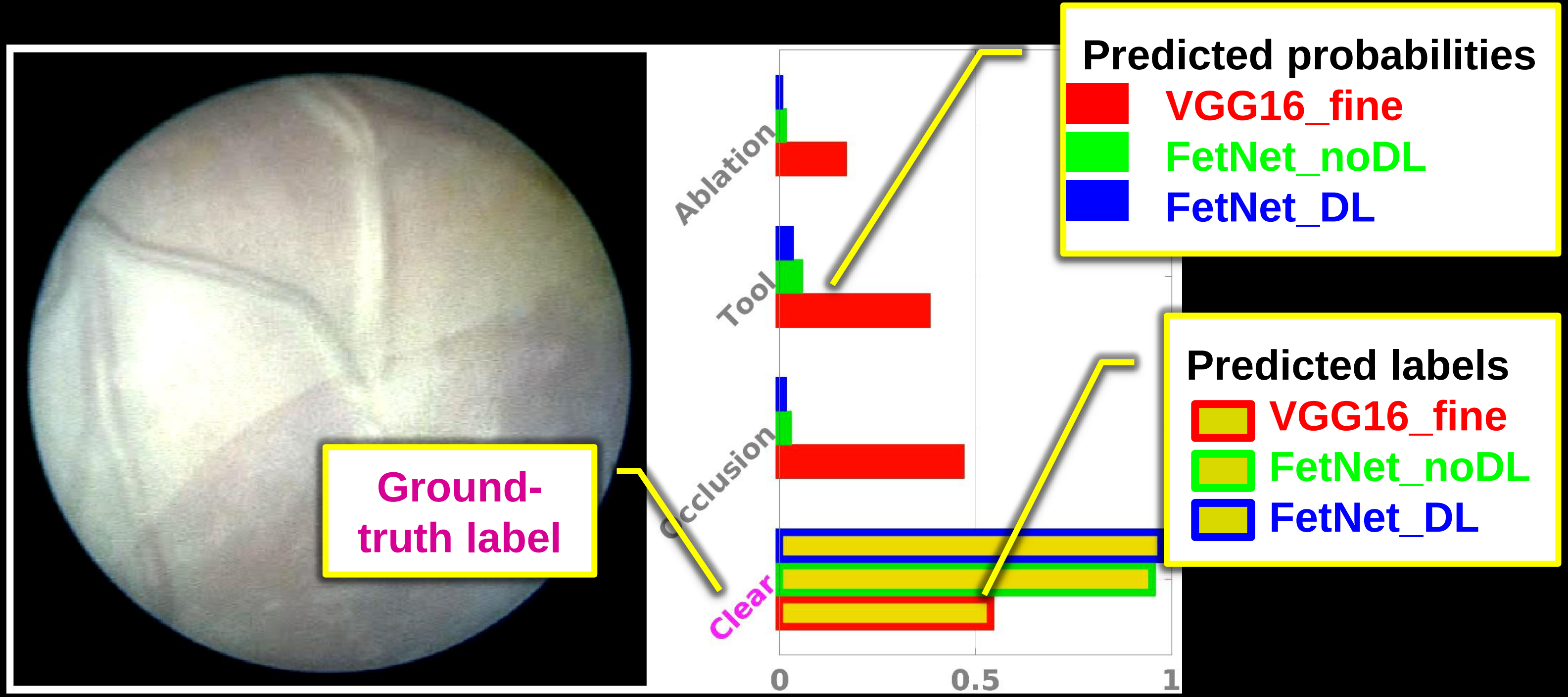
Method \ Class		Clear	Occlusion	Tool	Ablation	Average
Ablation_detect	Precision	-	-	-	0.81	0.81
	Recall	-	-	-	0.71	0.71
	F1-score	-	-	-	0.76	0.76
VGGFE_SVM	Precision	0.52	0.55	0.68	0.32	0.52
	Recall	0.42	0.70	0.50	0.19	0.45
	F1-score	0.46	0.62	0.58	0.24	0.47
VGG16_fine	Precision	0.66	0.69	0.76	0.96	0.77
	Recall	0.47	0.69	0.73	0.61	0.63
	F1-score	0.55	0.69	0.74	0.75	0.68
VGG16_temporal	Precision	0.72	0.70	0.76	0.96	0.79
	Recall	0.46	0.68	0.73	0.56	0.61
	F1-score	0.56	0.69	0.74	0.71	0.68
FetNet_noDL	Precision	0.72	0.70	0.86	0.95	0.81
	Recall	0.78	0.60	0.90	0.69	0.74
	F1-score	0.74	0.65	0.88	0.80	0.77
FetNet_DL	Precision	0.86	0.69	0.92	0.96	0.86
	Recall	0.84	0.79	0.94	0.95	0.88
	F1-score	**0.85**	**0.74**	**0.93**	**0.95**	**0.87**



Qualitative Analysis

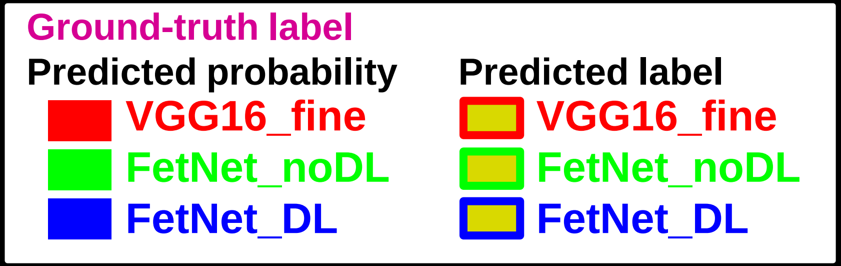
Ground-truth label

Predicted probability	Predicted label
VGG16_fine	VGG16_fine
FetNet_noDL	FetNet_noDL
FetNet_DL	FetNet_DL

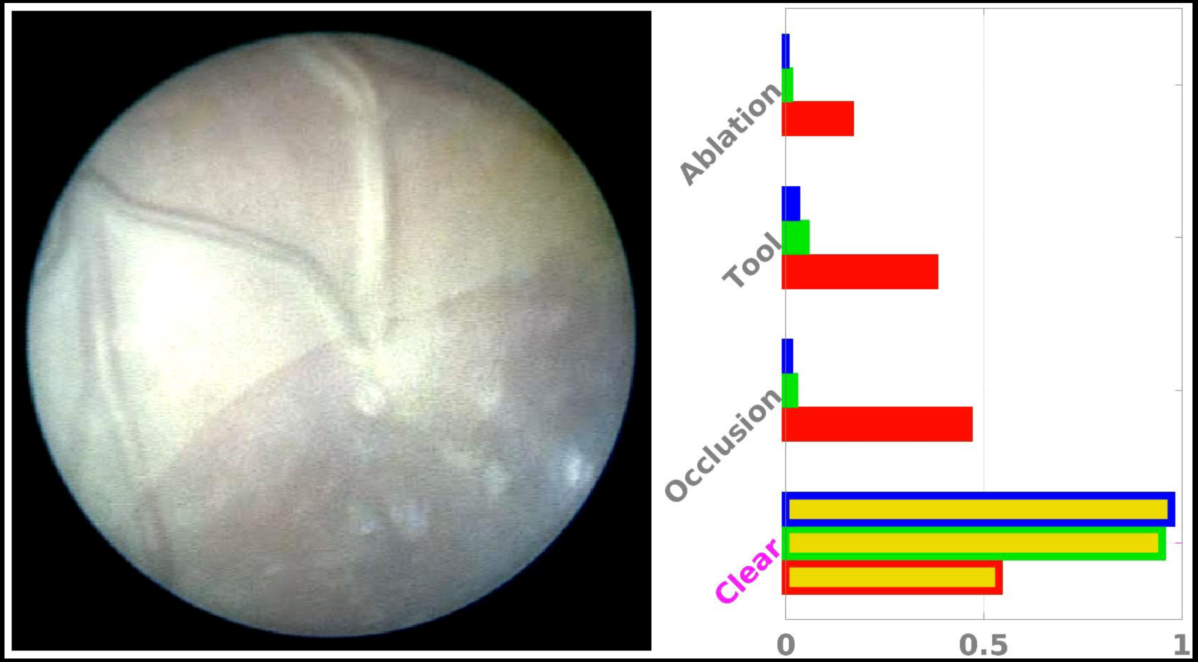


Qualitative Analysis

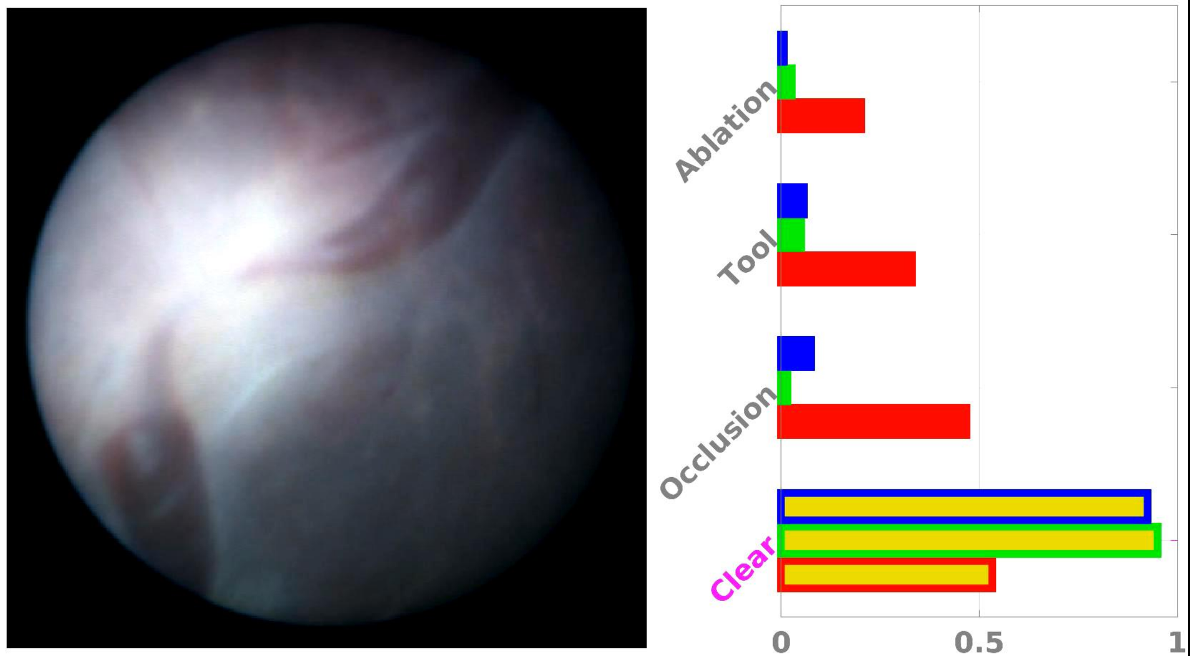
Single label per frame



Clip 1: Clear view or occlusion



Clip 2: Tool or ablation



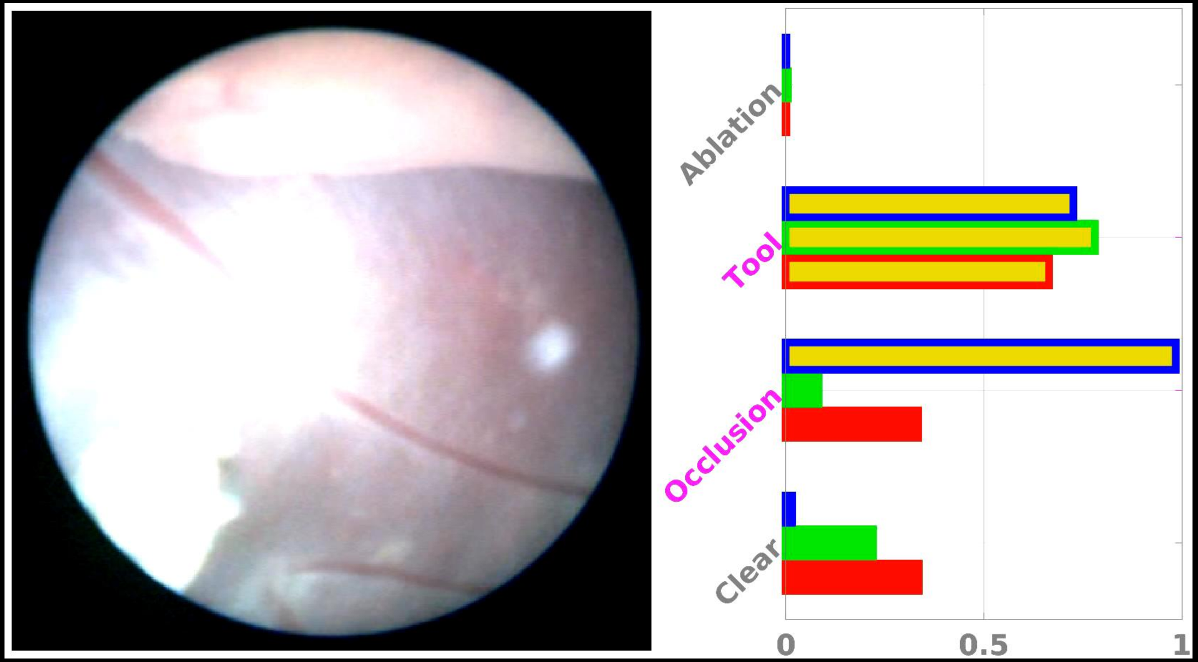
Qualitative Analysis

Multi-labels per frame

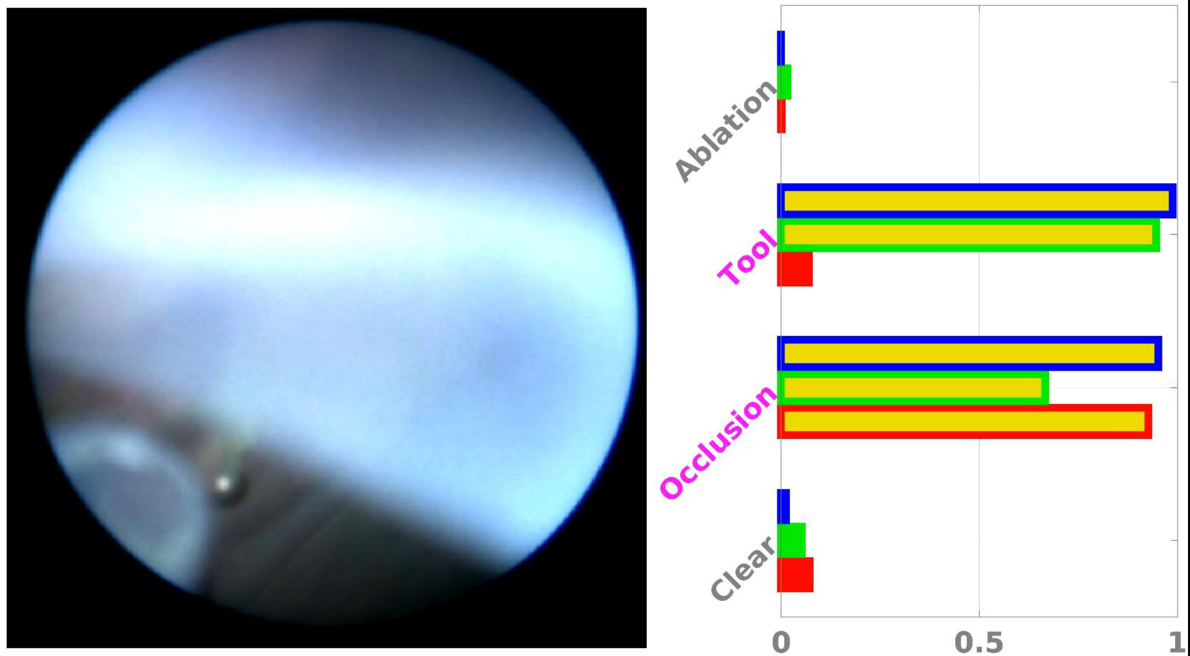
Ground-truth label

Predicted probability	Predicted label
VGG16_fine	VGG16_fine
FetNet_noDL	FetNet_noDL
FetNet_DL	FetNet_DL

Clip 3: Occlusion and tool



Clip 4: Occlusion, tool and ablation



Conclusion

Proposed FetNet architecture

- Occlusion Identification in Fetoscopic Videos
- Obtained an overall F1-score of 87%
- Outperformed existing methods
- Online testing returned a frame rate of 114 *fps*

Future work

- Possible integration in real-world systems
- Clear view segmentations are suitable for the field-of-view expansion

Thank you

sophia.bano@ucl.ac.uk

EPSRC
Engineering and Physical Sciences Research Council
wellcome trust
UCL
weiss
GIFT Surg
KU LEUVEN
UZ LEUVEN
KING'S College LONDON