MICCAI 2021

Endoscopic Vision Challenge 2021

HeiChole Surgical Workflow Analysis and Full Scene Segmentation (HeiSurF)

1st of October

Comitee

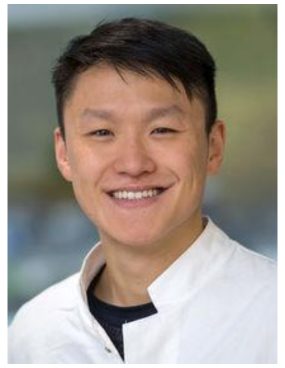
National Center for Tumor Diseases (NCT) Dresden

Sebastian Bodenstedt

Stefanie Speidel

Heidelberg University Hospital

Martin Wagner



Jonathan Chen

Anna Kisilenko

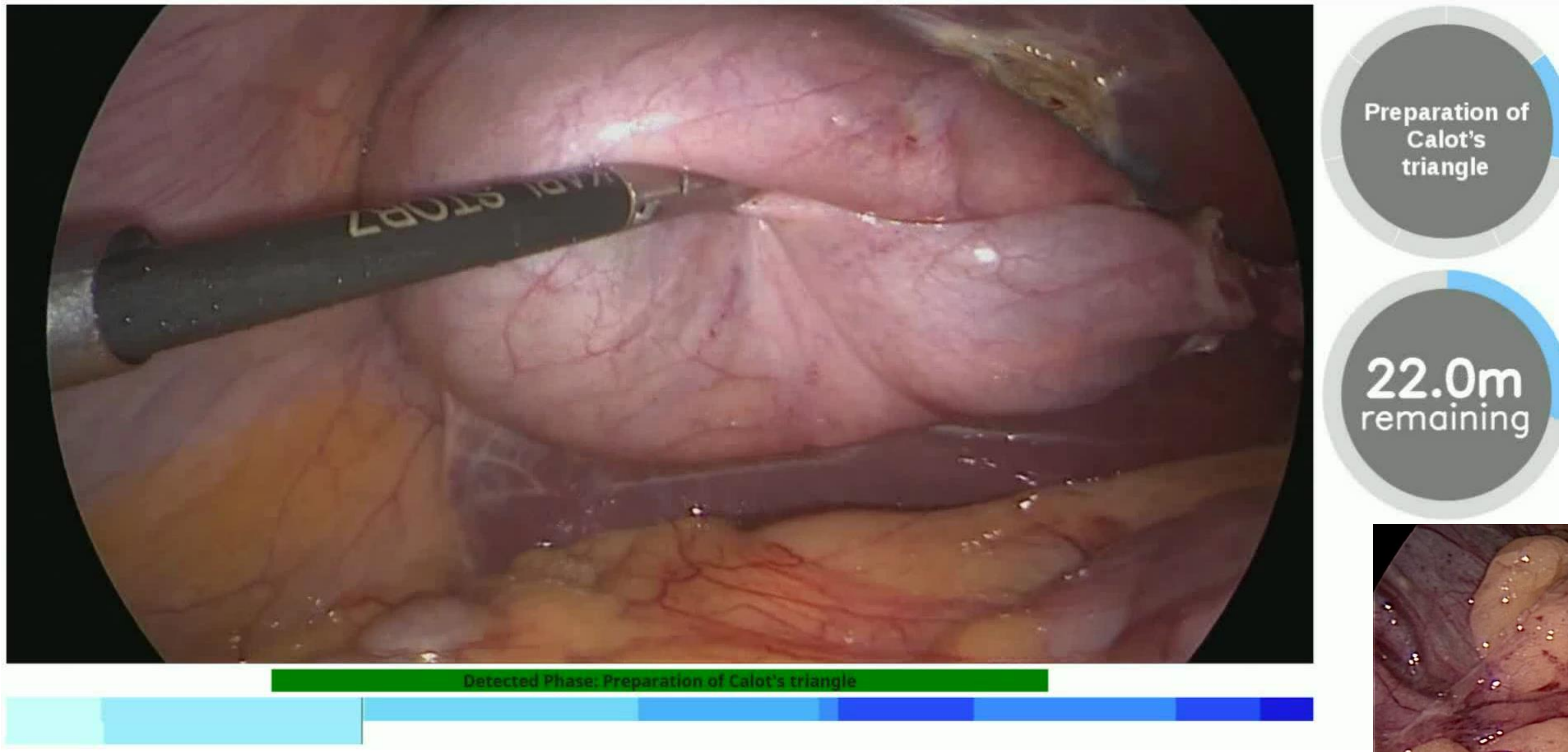
Beat Müller

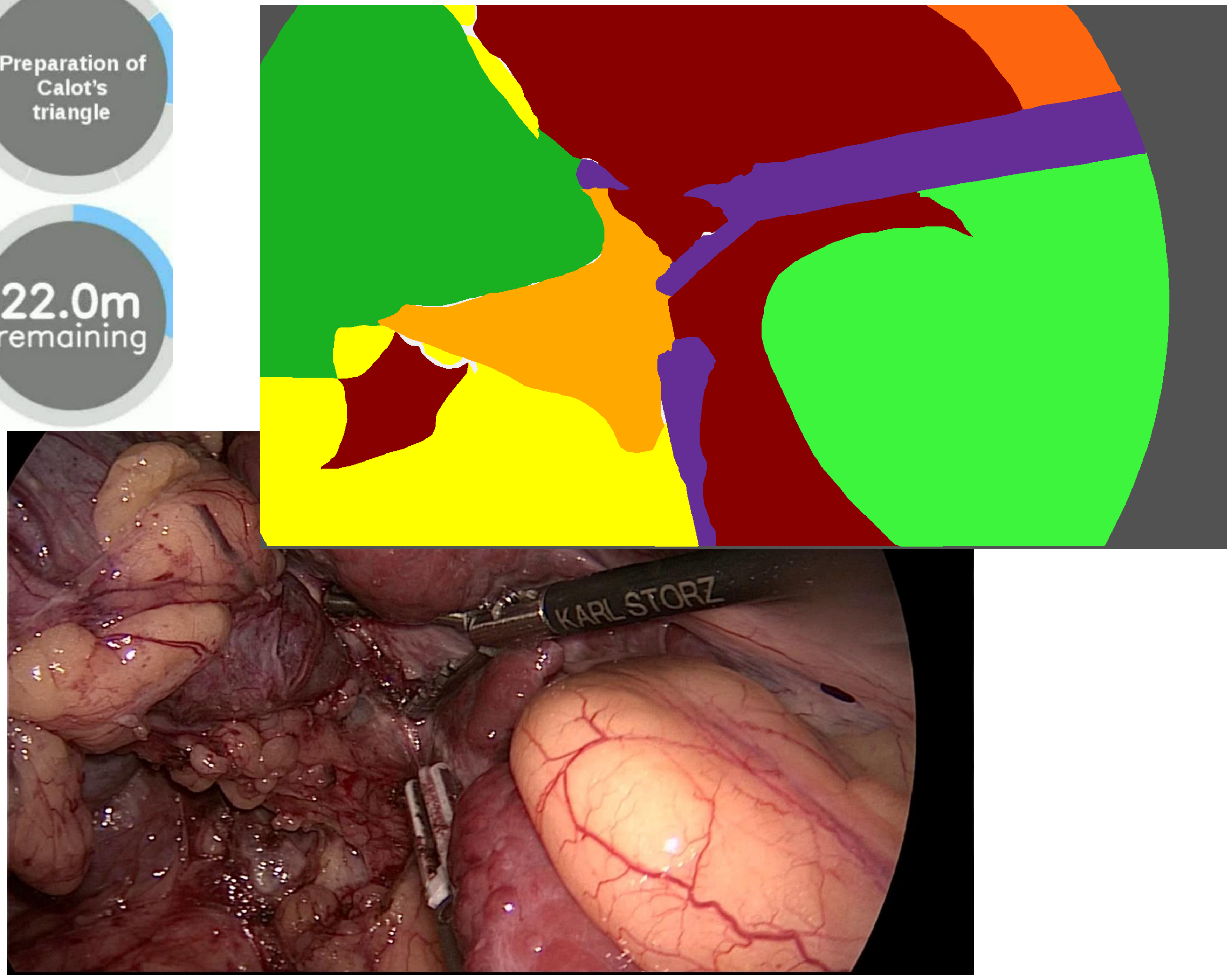
German Cancer Research Center

Lena Maier-Hein

Clinical Context

Surgical Workflow Analysis and Image Segmentation for context-aware CAS



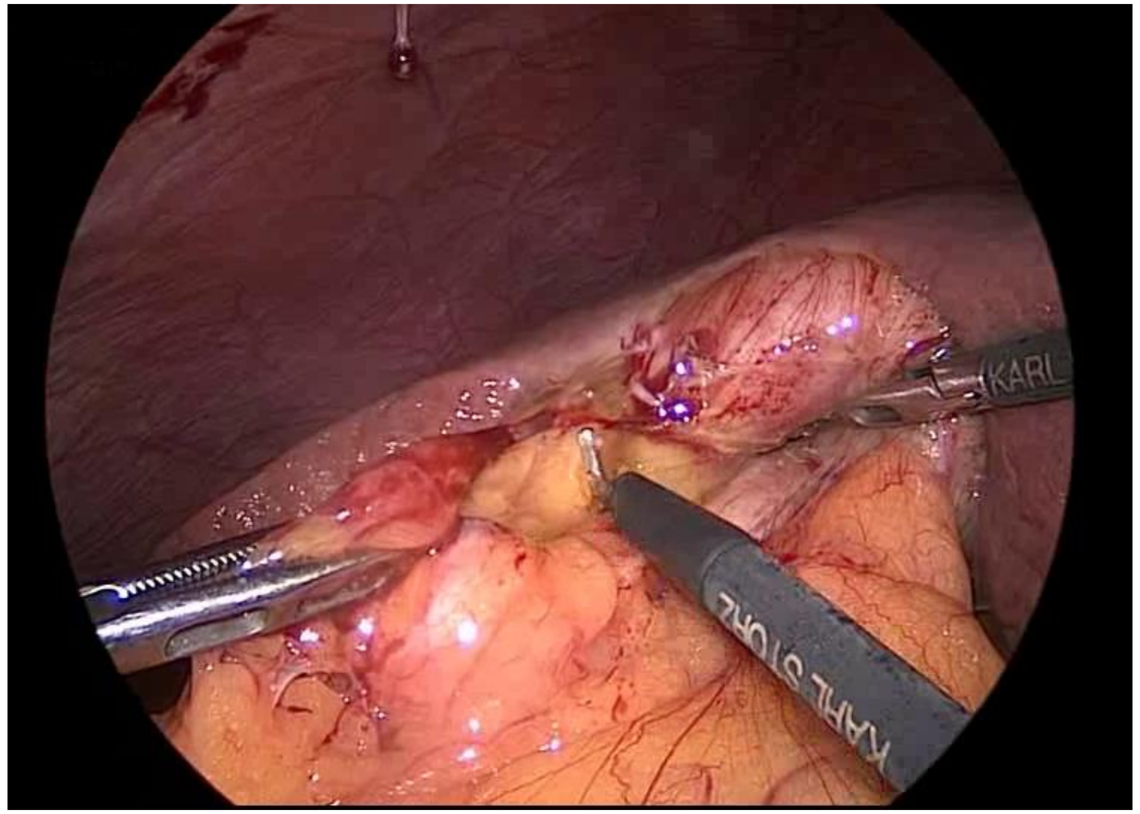


Dataset

- 33 laparoscopic cholecystectomies (gallbladder removals): Hei-Chole
  - 2019 challenge (http://arxiv.org/abs/2109.14956)
  - 3 centers
    - 15 cases from University Hospital Heidelberg (complex cases)
    - 15 cases from Salem Hospital (standard cases)
    - 3 cases from Sinsheim Hospital (standard cases)
  - 23.3h duration (42.4 minutes average procedure duration)
  - Anonymized videos (outside frames filtered out)
  - Cases representative of actual day-to-day procedures (no cherry picking)

- Four tasks
  - Full scene segmentation
  - Phase segmentation
  - Instrument presence detection
  - Action recognition

Task: Full Scence Segmentation



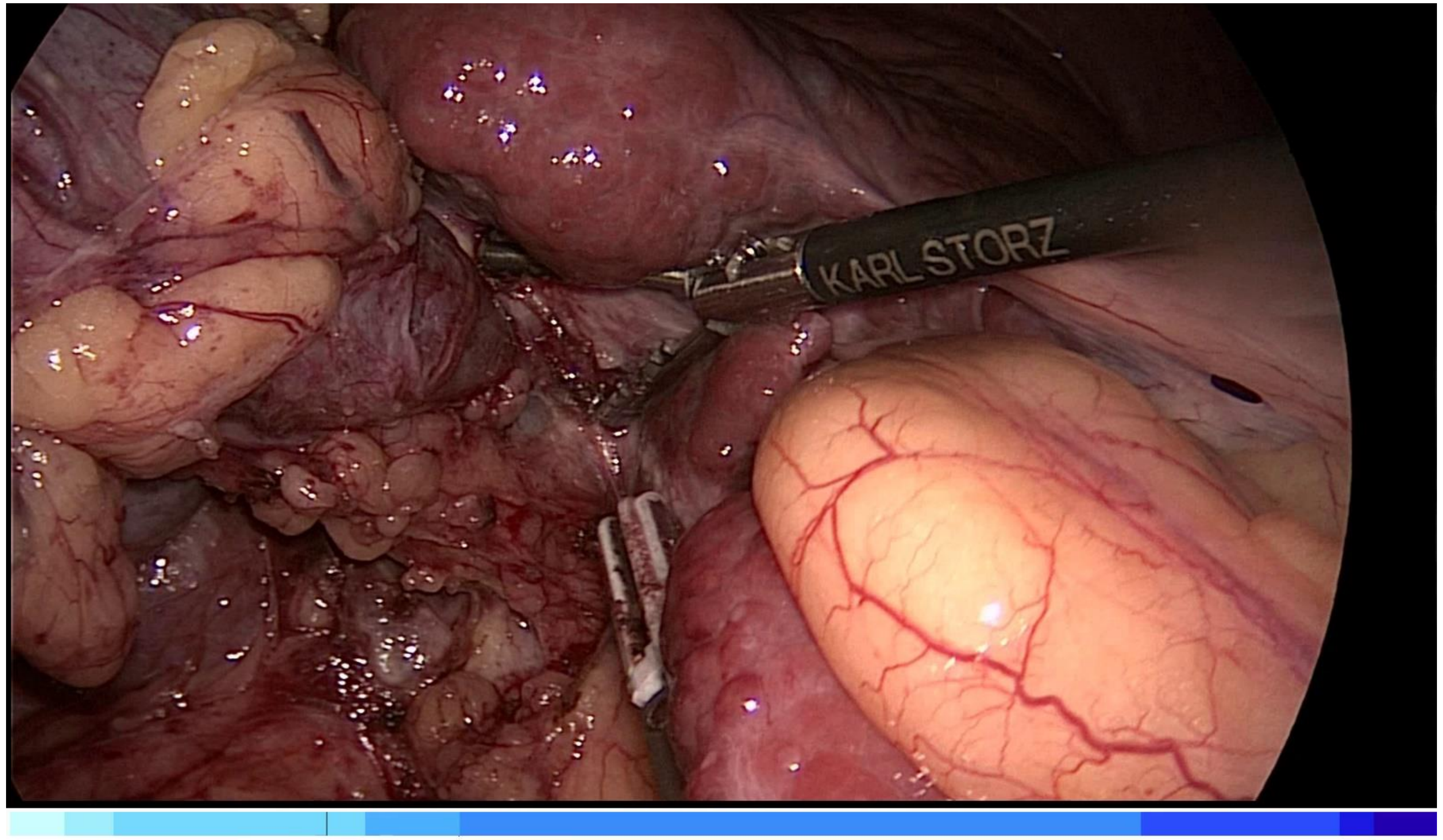
- **Annotation**: 3 expert annotators
  - Annotated independently
  - Discrepancies solved via discussion
  - Quality control by surgeon
  - 827 annotated images
- **Evaluation metrics**:
  - F1-score
  - Hausdorff distance
  - Computed per class
  - Averaged over all classes

$$F_1 = \frac{T_p}{2T_p + F_n + F_p} = \frac{2 \cdot precession \cdot recall}{precession + recall}$$

$$d_{\mathrm{H}}(X,Y) = \max\left\{ \sup_{x \in X} d(x,Y),\ \sup_{y \in Y} d(X,y) \right\},$$

Classes
Abdominal Wall and Diaphragm
Blood Pool
Censored out of Body
Clip
Cystic Artery
Cystic Duct
Drainage
Fatty Tissue
Gallbladder
Gallbladder Resection Bed
Gastrointestinal Tract
Gauze
Hilum of Liver / Hepatoduodenal Ligament
Inside of Trocar
Instrument
Liver
Other
Out of Image
Round and Falciform Ligament of the Liver
Specimenbag
Trocar

Task: Phase Segmentation



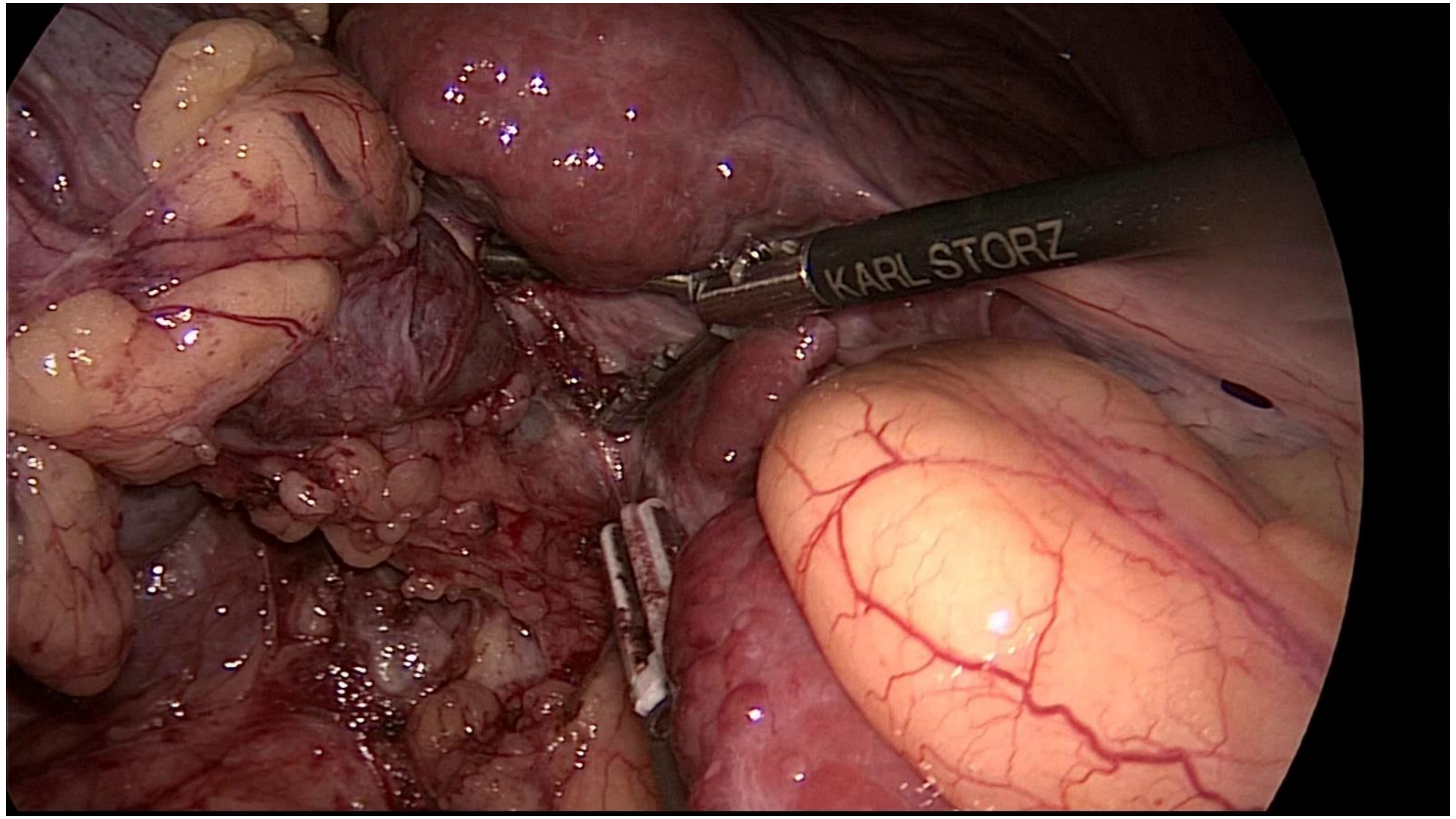
Phase name	Phase ID
Preparation	0
Calot triangle dissection	1
Clipping and cutting	2
Gallbladder dissection	3
Gallbladder packaging	4
Cleaning and coagulation	5
Gallbladder retraction	6

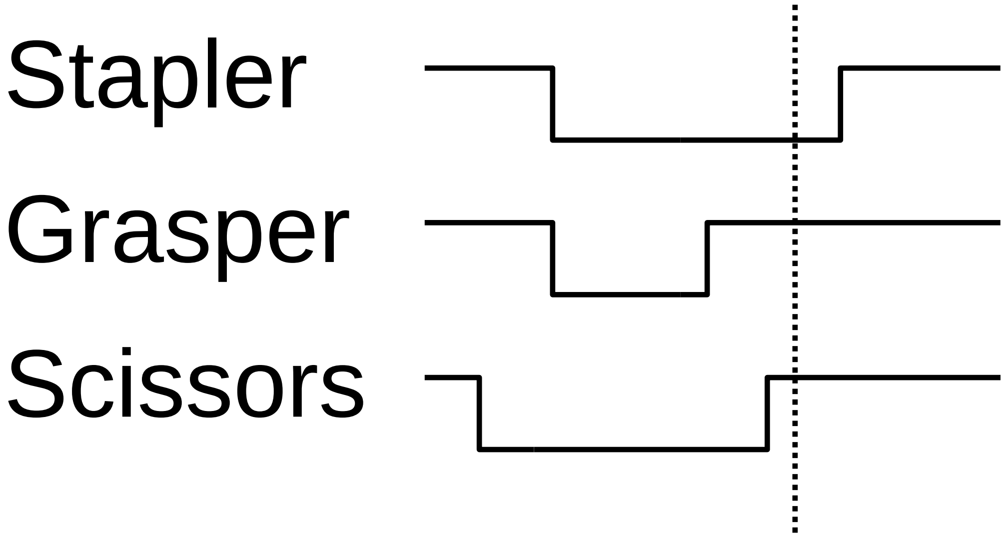
- **Evaluation metric**: F1-score
  - Computed per phase
  - Averaged over all phases

$$F_1 = \frac{T_p}{2T_p + F_n + F_p} = \frac{2 \cdot precesion \cdot recall}{precesion + recall}$$

- **Annotation**: 3 surgical experts
  - Annotated independently
  - Discrepancies solved via discussion

Task: Instrument presence detection





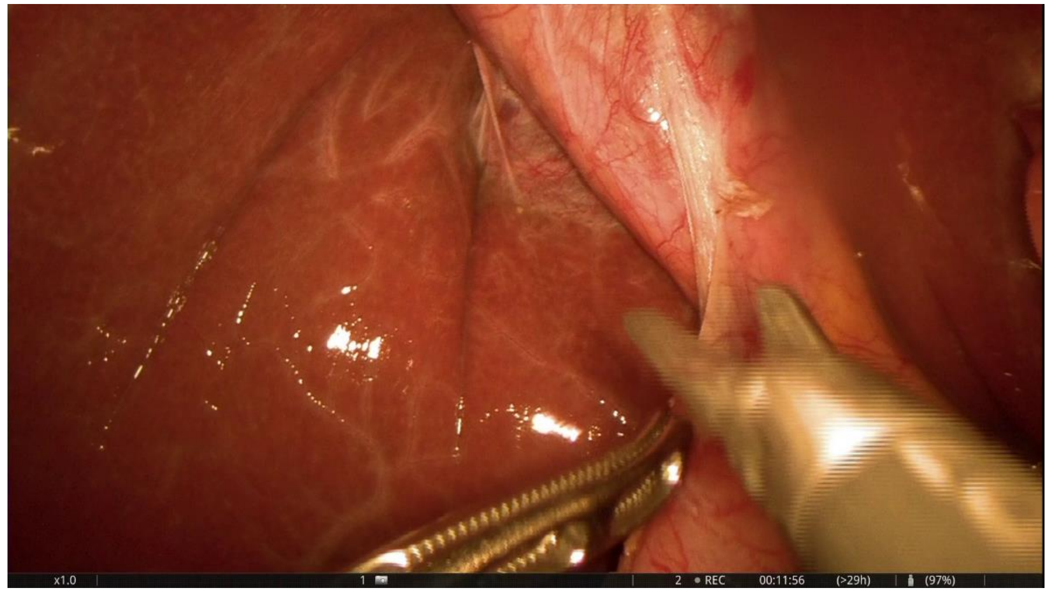
Tool category name	Tool category ID
Grasper	0
Clipper	1
Coagulation instrument	2
Scissors	3
Suction/irrigation	4
Specimen bag	5
Stapler	6

- **Evaluation metric**: F1-score
  - Computed per instrument category
  - Averaged over all categories

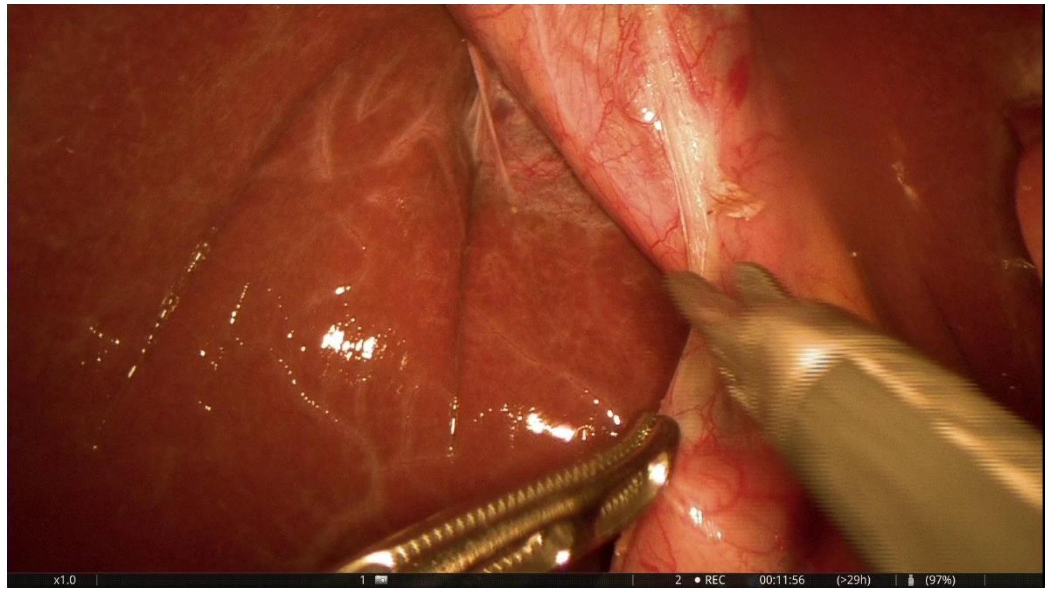
$$F_1 = \frac{T_p}{2T_p + F_n + F_p} = \frac{2 \cdot precesion \cdot recall}{precesion + recall}$$

- **Annotation**: 1 surgical expert
  - 21 classes
  - Merged into 7 overall categories

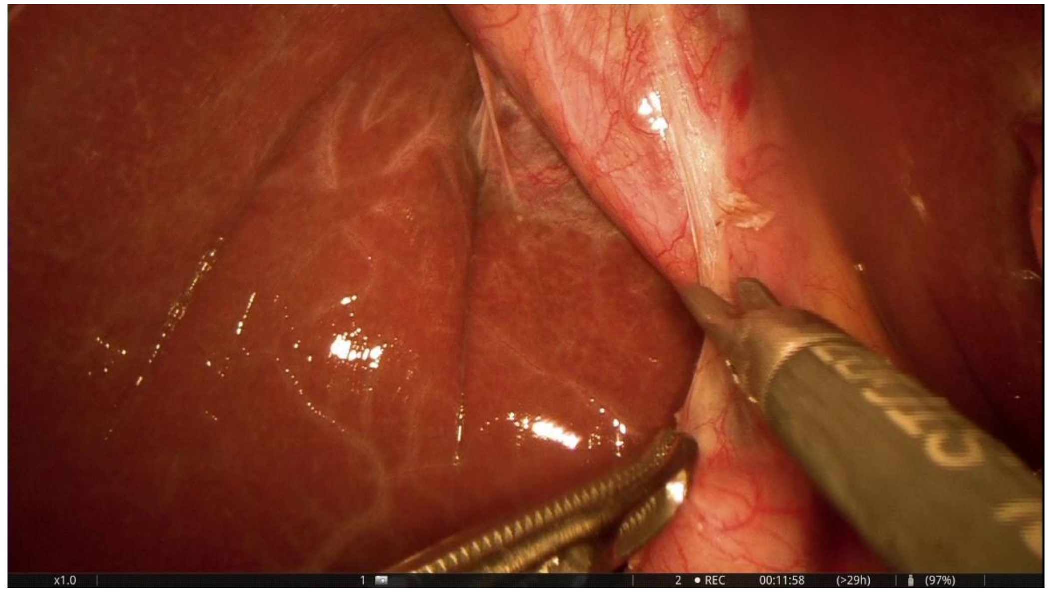
Task: Action recognition



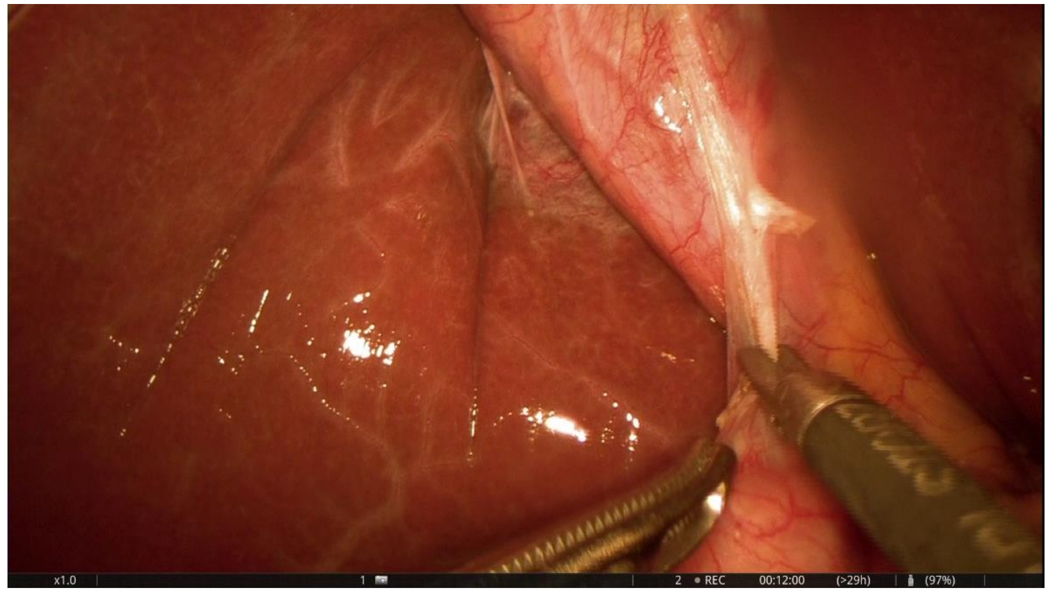
…



…



…



Action name	Action ID
Grasp	0
Hold	1
Cut	2
Clip	3

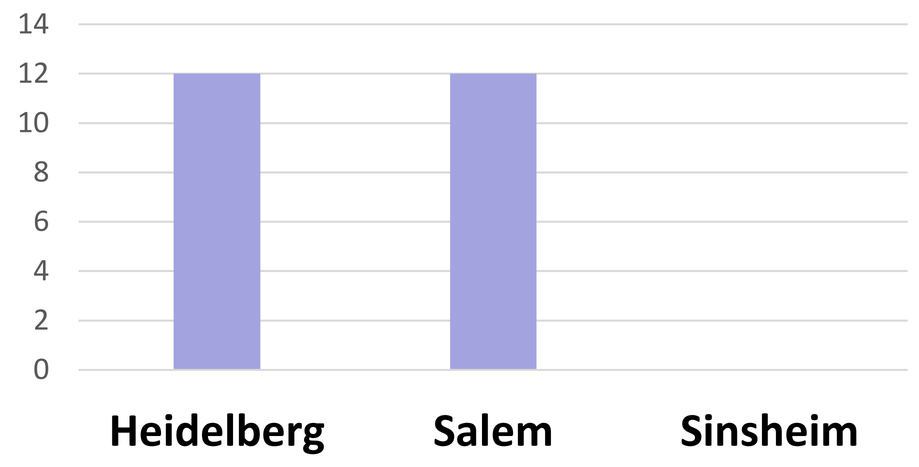
- **Evaluation metric**: F1-score
  - Computed per action class
  - Averaged over all actions

$$F_1 = \frac{T_p}{2T_p + F_n + F_p} = \frac{2 \cdot precesion \cdot recall}{precesion + recall}$$

- **Annotation**: 1 surgical expert

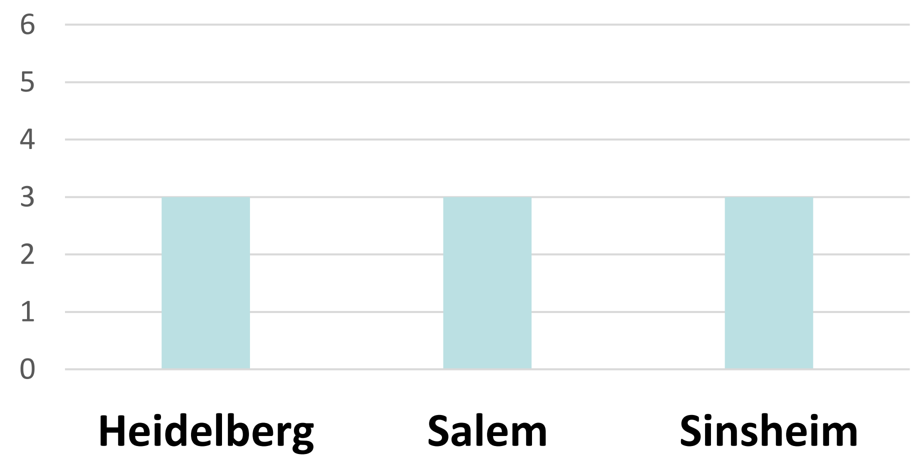
Training & testing data

Training set: 24 annotated laparoscopies



- 581 images with full scene segmentation
- Every frame annotated for phase and instrument usage
- 4 actions annotated

Test set: 9 annotated laparoscopies



- 248 images with full scene segmentation
- Every frame annotated for phase and instrument usage
- 4 actions annotated

- **Test data** was not released to participants
  - Participants were asked to submit Docker containers based on training data
  - Dockers were then evaluated on the cluster at NCT

Participating teams

Team 2AI
- Bruno Oliveira
- Helena Torres
- Pedro Morais
- Jaime Fonseca
- João Vilca

Team VisionAI Hutom
- Seungbum Hong
- SeulGi Hong
- Minkook Choi

Team ARCSeg
- Jonathan Zamora-Anaya
- Shunkai Yu
- Yuheng Zhi
- Jingpei Lu
- Shan Lin
- Michael Yip

Team UCL
- Daniel Corvesor

Team Digital Surgery
- Rahim Mohammadi
- Maria Grammatikopoulou
- David Owen
- Imanol Luengo
- Danail Stoyanov

Team SIAT-CAMI
- Tong Xia
- Fucang Jia

Team Muroran-IT
- Satoshi Kondo

Team GMCAO
- Sylvain Guy
- Sandrine Voros

Team Wintegral
- Wolfgang Reiter
- Jonas Hajek

Team CMEMS
- Nuno Freitas
- Ismael Vaz
- Estêvão Lima
- Carlos Lima

Team UESTC
- Baosheng Zou
- Guotai Wang
- Lu Wang

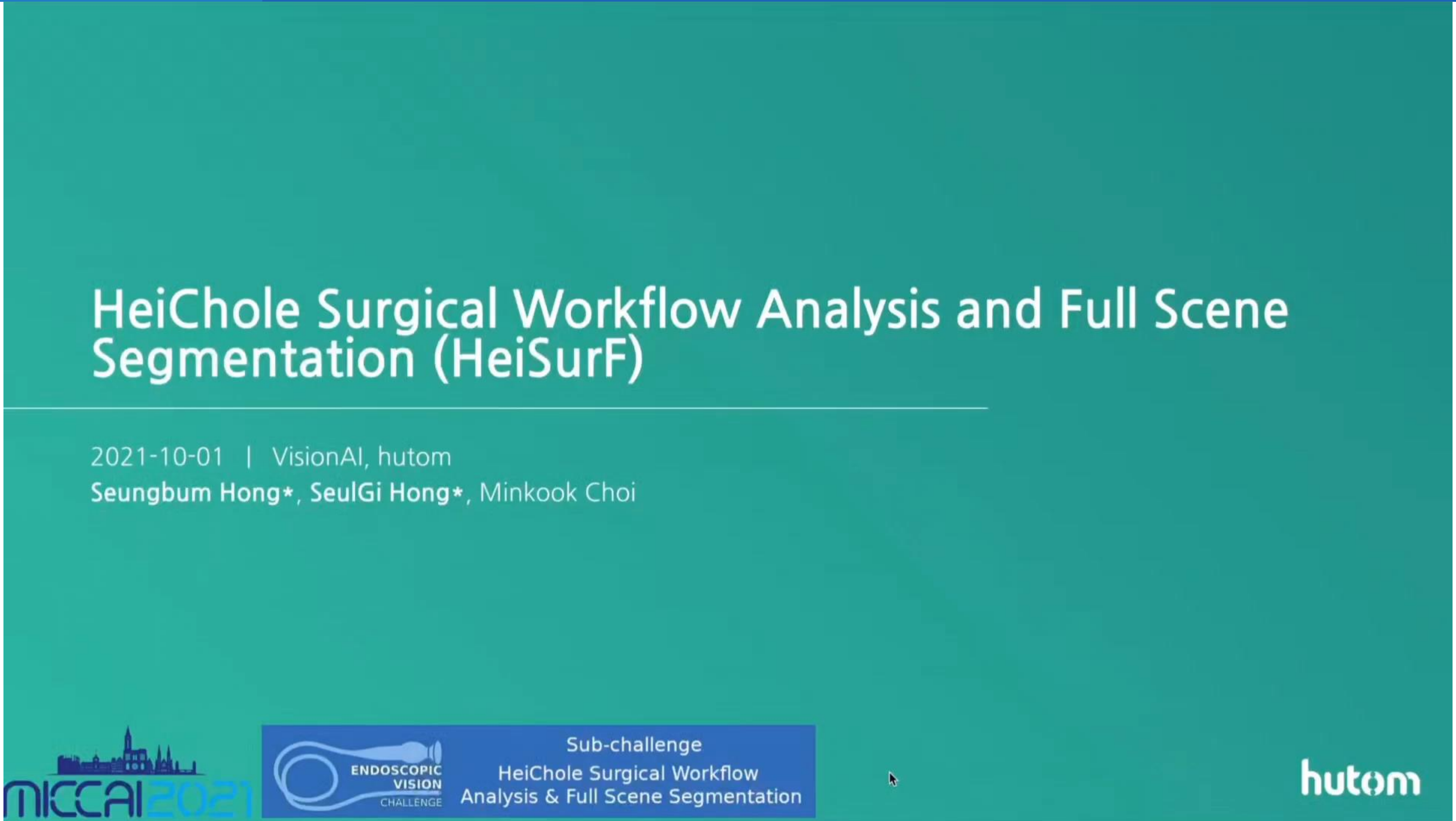
Team 2AI

Surgical Video Analysis using an Ensemble of Multi-task Recurrent Convolutional Neural Networks

Bruno Oliveira [a,b,c], Helena R. Torres [a,b,c], Pedro Morais [c], Jaime C. Fonseca [b], João L. Vilaça [c]

[a] Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal;
[b] Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal
[c] 2Ai – Polytechnic Institute of Cávado and Ave, Barcelos, Portugal

Team VisionAI Hutom



Team ARCSeg

docs.google.com

ARCSeg Team Presentation for HeiSurF Challenge 2021 - Google Slides

ARCSeg Team

Jonathan Zamora-Anaya[1], Shunkai Yu[1], Yuheng Zhi[1], Jingpei Lu[1], Shan Lin[1], Michael Yip[1]

[1] University of California San Diego, Department of Electrical and Computer Engineering

{jzamoraa, shyu, yzhi, jil360, shl102, yip}@ucsd.edu

Surgical Phase recognition for HeiSurf 2021 competition

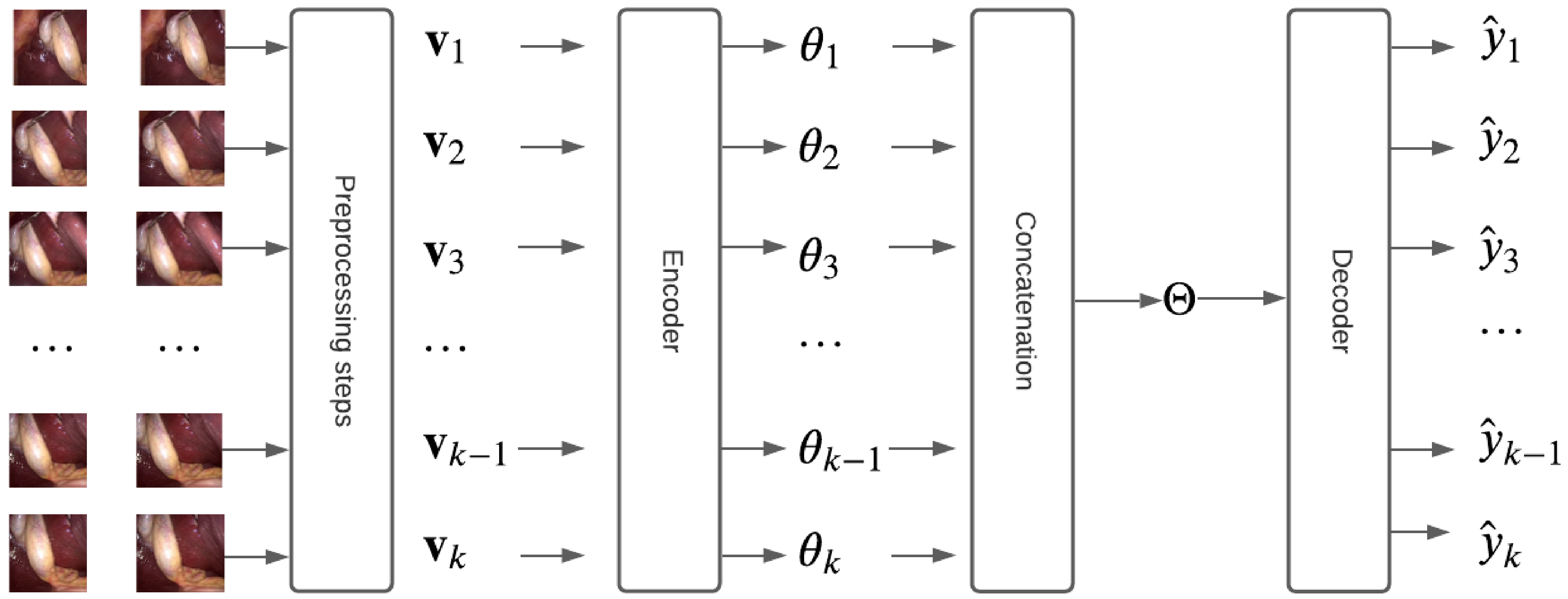
Daniel Corvesor (UCL)

Motivation

- Completed analysis and experimentation as part of my MSc Thesis titled ‘Cross Domain adaptation for surgical workflow recognition’
- I was attempting to use a labelled dataset (Cholec80) to improve on the unsupervised performance on HeiSurf2021.
- In doing these experiments, I submit to this competition one preliminary baseline approach I used which does in fact use the HeiSurf2021 labels for supervised learning
- Idea is to transfer the strong visual features from a larger dataset and then learn the different annotation and temporal features from the HeiSurf2021 dataset (particularly the repeated phases not present in Cholec80).

Method

- We use the classic encoder-decoder model with our encoder selected to be Resnet50 (using the Imagenet weights) and then fine tuned with the Cholec80 dataset.
- We then use an MSTCN decoder model which we train with the HeiSurf2021 dataset.
- This approach is a simple transfer learning technique.



Architecture Details

- The MSTCN I used had 4 stages each with causal dilated convolutions (for online inference) and 32 feature maps.
- We trained using two Nvidia GPUs and trained the encoder model with a Step LR scheduler 0.001 starting lr and the decoder model with a 0.0001 and a cyclic lr scheduler. The encoder model had early stopping based on accuracy and the decoder model based on f1 score.

Team Digital Surgery

Workflow Recognition

Used a two-stage model

- Encoder with resnet backbone trained in a multitask setup
  - ResNet features fed into different head for surgical phase, action and instrument detection
  - Representations adapted for each task using 2 fully connected layers
- Decoder, a 4 stage TCN model to incorporate temporal information

Trained the models using binary cross entropy

Weighted positive examples to tackle unbalanced dataset

Segmentation

DeepLab-ResNet50

Dataset

- Pretrain on CholecSeg8k+HeiSurF, map to nearest equivalent classes
- Finetune on HeiSurF alone

Class imbalance

- train/val/test splits based on class distributions
- Balanced train sampling by class presence (not area)

Focal cross-entropy loss, Adam lr=1e-4

ENDOSCOPIC VISION CHALLENGE

Team SIAT CAMI

MICCAI 2021

Multi-task Mutual Channel Recurrent Net for Fine-grained Surgical Workflow Recognition

Tong Xia[1,2], Fucang Jia[1]

1 Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
2 University of Chinese Academy of Sciences

Motivation

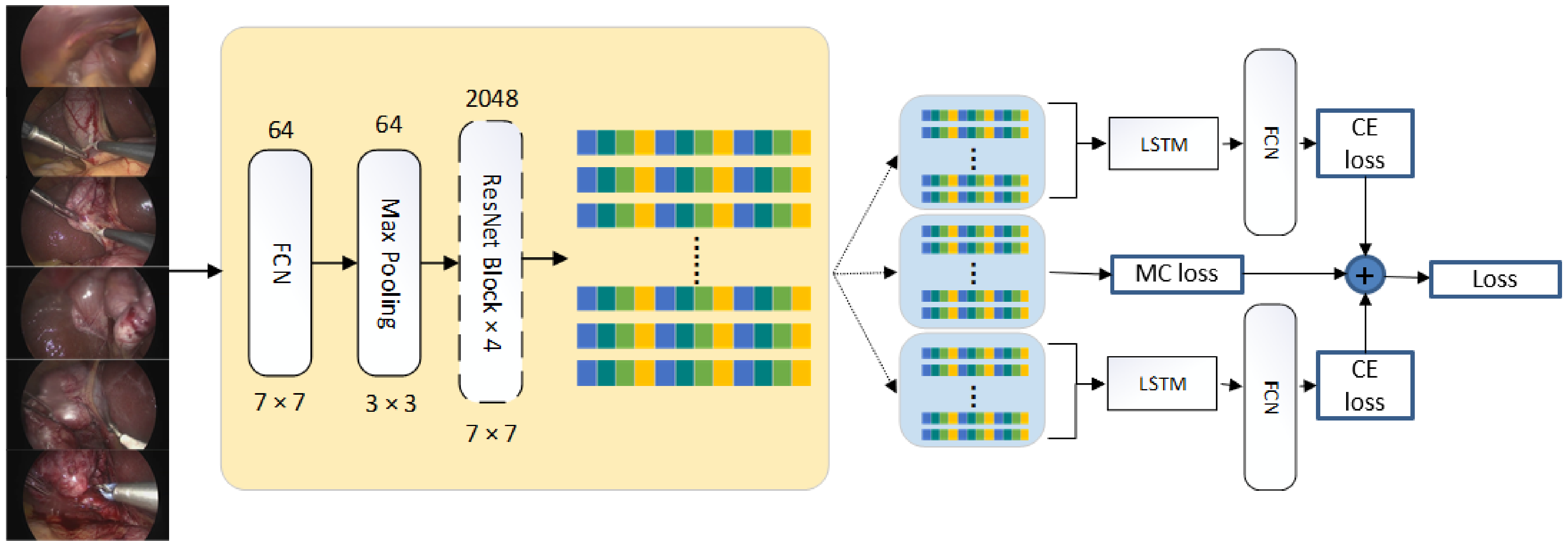
- Surgical multi-task consists of instrument, action, and workflow.

- Instrument is the most obvious object in surgical scene
- Action and workflow is difficult to recognize for its complexity in temporal information and fine-grained local features.

- The key to extract these fine-grained features is to find discriminative local regions corresponding to specific class.

Team SIAT CAMI

Method



- MCLNet, a multi-task network with mutual channel loss.
- The mutual channel bases on channel-wise attention.
- Extended LSTM modules capture temporal motion information.

$$L_{MC}(F) = \lambda_1 L_{\mathrm{dis}}(F) + \lambda_2 L_{\mathrm{div}}(F)$$

$$L_{\mathrm{dis}}(F) = L_{CE}(y, Softmax(e^{g(F_i)}))$$

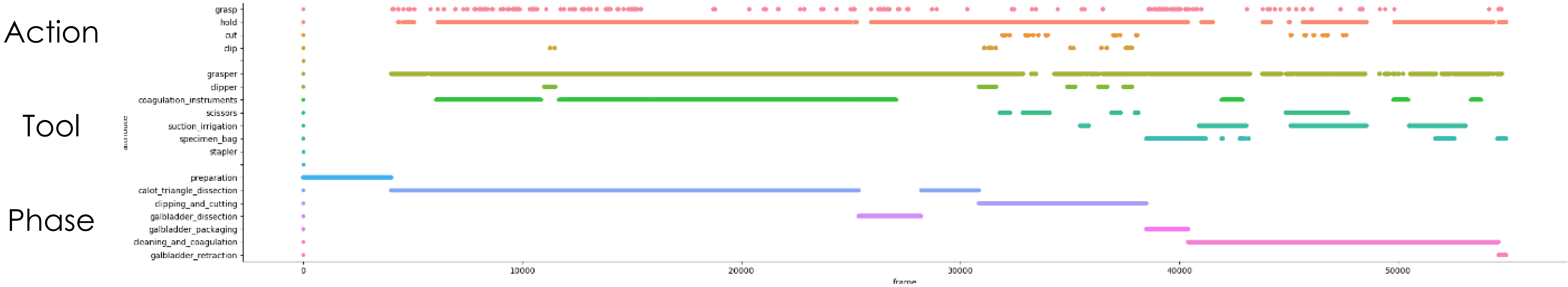
$$L_{\mathrm{div}}(F) = \frac{1}{15}\sum_{i=0}^{14} h(F_i)$$

Workflow Analysis in HeiSurF Challenge

Satoshi Kondo
Muroran Institute of Technology
室蘭工業大学
MURORAN INSTITUTE OF TECHNOLOGY

Basic Idea

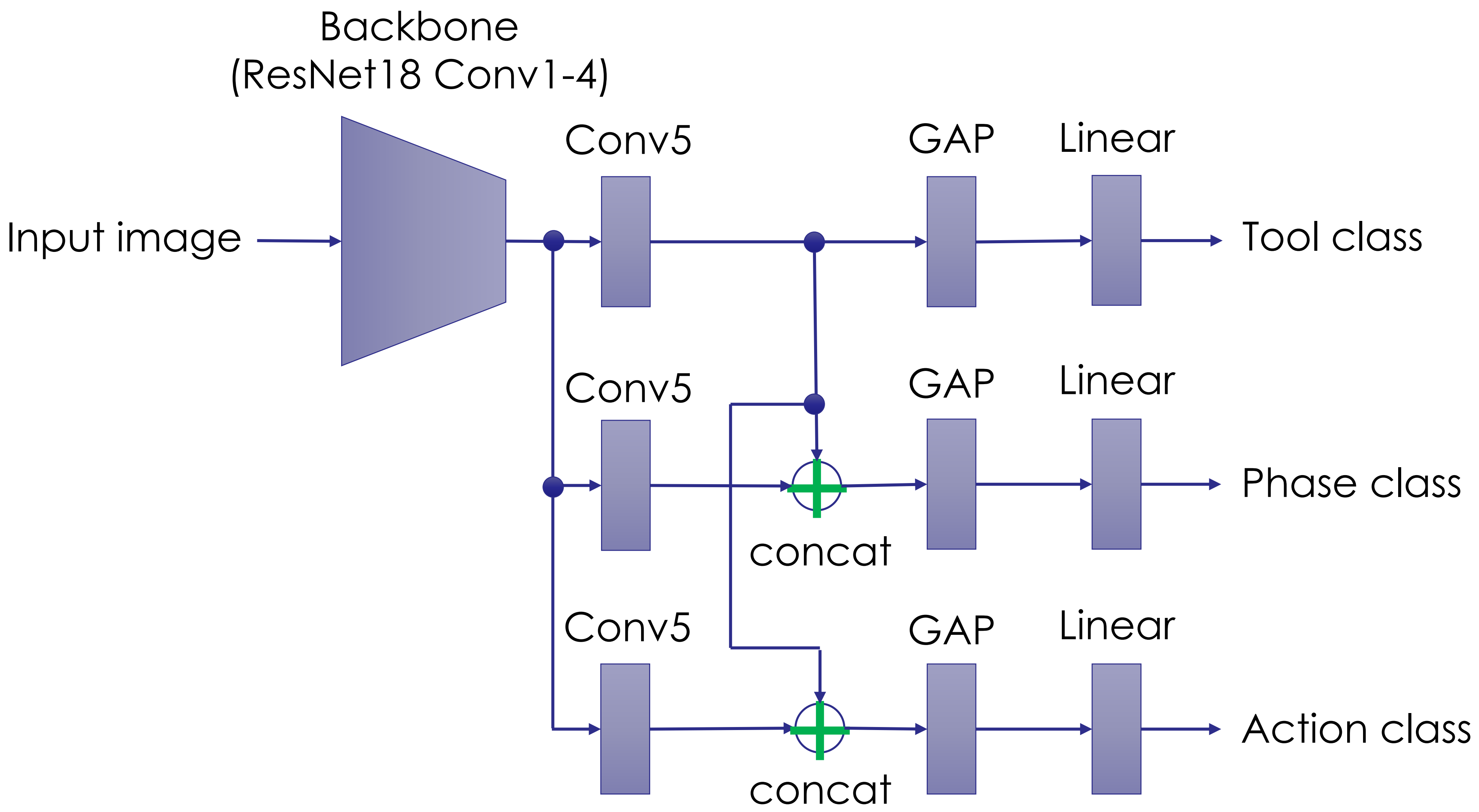
- Usage of some tools correlate to some actions and phases.
  - Ex.1) Tool ‘clipper’ is used in action ‘clip’
  - Ex.2) Tool ‘specimen_bag’ appears in phase ‘gallbladder_packaging’



Example of time chart of phase, action and tool.

Network Architecture

- Backbone + 3 head networks
  - Head networks are for tool / phase / action recognition
  - Feature maps in 'tool' head network are concatenated to feature maps in 'phase' and 'action' networks.



Configuration of Training

- Dataset: 18 videos for training, 3 videos for validation
- Preprocessing: Subsampled to 320 x 320 pixels / 5 fps
- Augmentation: Shift, Scale, Rotate, Color Jitter, Blur
- Optimizer: Adam, Initial learning rate = 2.3e-5 (optimized by Optuna)

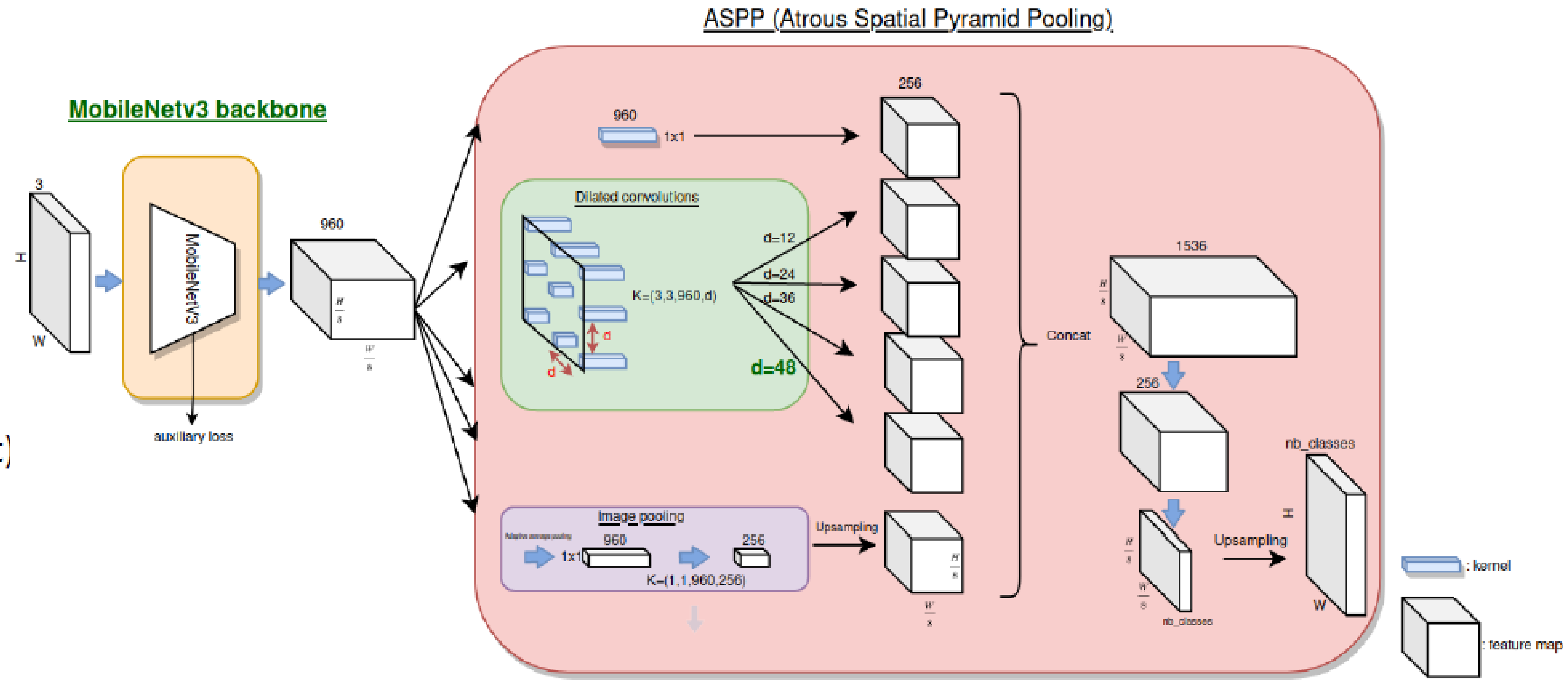
- Performance for validation dataset (F1 value)
  - Phase: 0.559, Action: 0.865, Tool: 0.790

Multi-class semantic segmentation on Heichole dataset

S. Guy[1], S. Voros[1,2]

[1] TIMC Laboratory, Univ. Grenoble Alpes, CNRS, UMR 5525, Grenoble, France ----- [2] INSERM

- **Architecture:**
  - **Deeplabv3 [1] with MobileNetv3 [2] backbone**:
    - modified kernel atrous rates: 11% gain of performance compared to default rates
    - Comparison to 20 models (incl. Unet++ and Attention Unet, and several ResNet/MobileNet backbones)
- **Data augmentation:**
  - random flip, crop, rotate
  - Tried without improvement:
    - vertical flip, random shuffle of image grid cells,
      random brightness contrast,
      random gamma correction, gaussian blur,
      artificial "smoke"
- **Pretraining:**
  - On Coco
  - (Endovis2018 enables faster training but no improvement)
- **Loss:**
  - Dice + Cross entropy
  - Comparison to:
    - cross entropy
    - Log(dice) + cross entropy
- **Optimizer:**
  - Adam (betas=(0.9,0.999), weight decay=5e-4)
  - Reduce on Plateau learning rate policy



Deeplabv3 [1] with MobileNetv3 [2] backbone and modified kernel atrous rates d=(12,24,36,48)

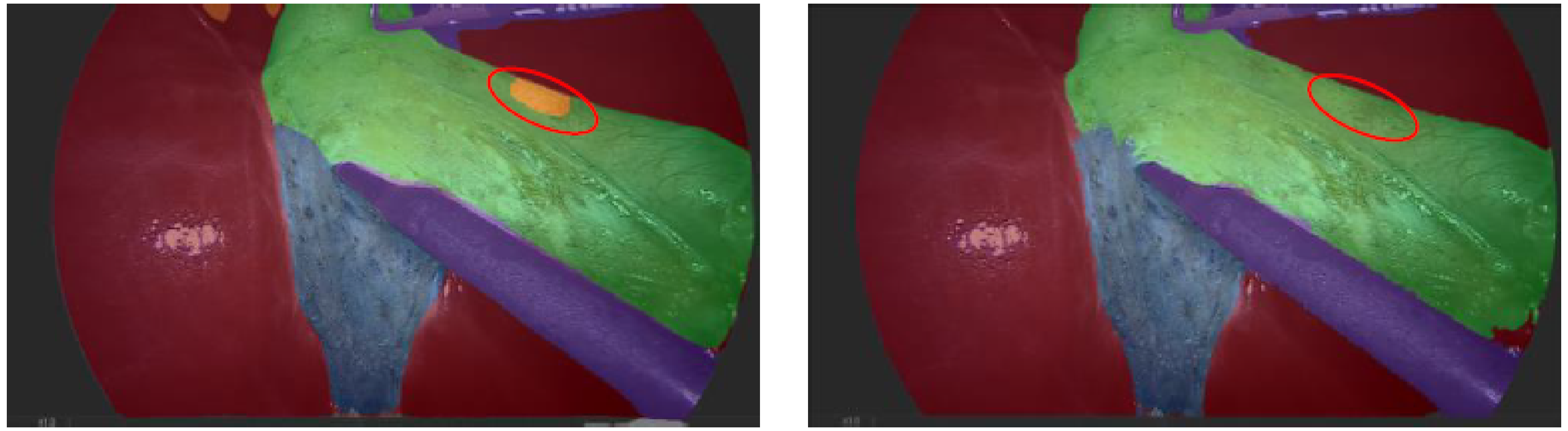
[1] Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
[2] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1314-1324).

Inference:

- Test-time augmentation: Average network probability predictions over [0.75, 1., 1.25] scales and the horizontally flipped version of the image

Post-processing:

- Remove small inconsistencies in the final image : DenseCRF
  - Fully connected Conditional Random Field model [3]
  - Grid search on DenseCRF parameters: w1=6, alpha=50, beta=5
- Small area removal
  - removing pixels of a class that has very few occurrences in an image
- censored class: white images detected outside of the neural network



Before (left) and after (right) applying DenseCRF[3] postprocessing

Results:

- Test set: videos 13-11-22-5-4 - Train set: the other videos
- Metric : IoU averaged per image

Blood pool	Drain.	Inside of trocar	Cystic artery	Clip	Trocar	Cystic duct	Specimen bag	Gb. resection bed	Hilum	Fat	Gastro. tract	Ligament	Abdo.	Gauze	Gallbladder	Instr.	Liver	Out of image	Cens.	Other	Averaged
0.00	0.00	0	0.03	0.06	0.13	0.17	0.25	0.27	0.37	0.51	0.51	0.52	0.55	0.63	0.63	0.72	0.80	0.87	1.00	nan	0.40

Poor results due to:
- Few occurrences of classes
- Difficult detection even for human eye without context (temporal)

Conclusion:

- **Deeplabv3** with **MobileNetv3 backbone**, slightly modified to give **higher spatial context**
- Important: **data augmentation, DenseCRF and test-time data augmentation**
- Performance highly variable depending on the class
- important **drop in performance between train/validations and test sets**: low generalization from one video to another
- Didn't address the problem of reflections on organs and instruments

[3] Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems, 24, 109-11

Team wintegral

- Wolfgang Reiter
- Jonas Hajek

Phase

- Strategy:
  - Fine tune resnet50-i3d with non-local blocks pretrained on Kinetics400 on 32-frame clips
  - Map output features to embedding space
  - Train LSTM on embeddings
- Preprocessing:
  - Downsample + Center crop 256x256
  - Clip wise random augmentations: zoom center, flip, colour (hue, saturation, contrast)
- Results:
  - Short term temporal information(i3d) gives larger improvements than long term (LSTM)
  - resnet50-i3d is too large for this dataset: freeze first part of model

Tool

- Strategy:
  - Concatenate max-pooling and fc-head from Cholec80 trained ResNet-50
  - Remove low-frequency class stapler in labels for first training
  - Use trained weights for additional binary model to predict stapler
- Preprocessing:
  - Resize to 240x427x3
  - Random augmentations: zoom center, rotate, colour (hue, saturation, contrast)
- Results:
  - ResNet-50 performs better than Resnet50-i3d
  - Performance measure on stapler requires more data

Segmentation

- Strategy:
  - U-Net shaped architectures with and without Inception blocks
  - Compare Inception U-Nets with and without skip connections
- Preprocessing:
  - Resize to 544x960x3
  - Connect RGB and HSV colour space as input
  - random augmentations: zoom center, translate, rotate, colour (hue, saturation, contrast)
- Results:
  - Traditional U-Net without Inception blocks performs slightly better
  - Close results in Inception U-Nets with and without skip connections

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

UMINHO CMEMS
CENTER FOR MICROELECTROMECHANICAL SYSTEMS

Universidade do Minho
Escola de Engenharia

HeiChole Surgical Workflow Analysis and Full Scene Segmentation (HeiSurF)

MICCAI 2021

Surgical Full Scene Segmentation by Using Boundary Multi-Mask Scoring R-CNN Trained with the 2nd Momentum

Nuno R. Freitas, Ismael Vaz, Estevão Lima, Carlos S. Lima

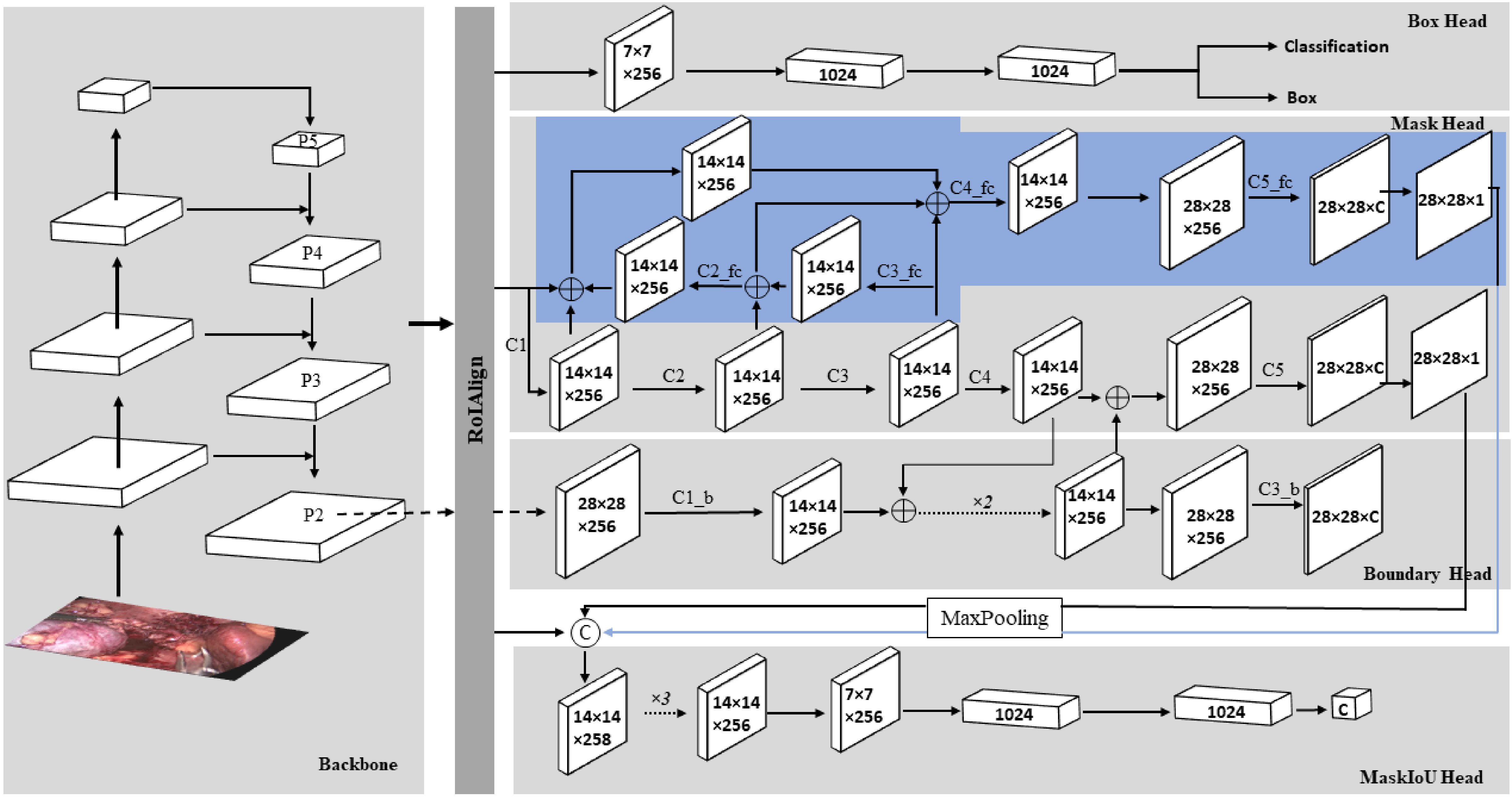
Proposed B2MS-RCNN

Task: Segmentation

Main purpose: Improving the quality of predicted masks.

HOW?

By generating an additional mask (blue), richer in low-level spatial information.



New added module to the baseline BMask-RCNN with Mask-scoring.

Second Momentum Training Algorithm

- Main Purpose: Perform a more efficient training and more robust to non-tuned hyperparameters with the second-order momentum optimizer.

- SGD + Momentum updating rule: $w^{k+1} = w^k + \beta^k(w^k - w^{k-1}) - \eta^k \nabla w^k f(w^k)$

As the n^{th} order derivative is just the derivative of the $(n\text{-}1)^{th}$ derivative then the n^{th} moment is the momentum of the $(n\text{-}1)^{th}$ momentum.
Therefore, the second momentum, which is the acceleration of the weight coefficients, can be obtained by the momentum of the momentum:

- SGD + 2nd Momentum updating rule:

$$w^{k+1} = w^k + (\beta^k + \boldsymbol{\gamma^k})(w^k - w^{k-1}) - \boldsymbol{\beta^k \gamma^k}(\boldsymbol{w^{k-1}} - \boldsymbol{w^{k-2}}) - \eta^k \nabla w^k f(w^k)$$

Discussion/Conclusion

- The proposed algorithm was trained using **subjects 1 to 19** and validated on the remaining **subjects 20 to 24**.
- **SGD with momentum** and the **SGD with second-order momentum** were the used optimizers. In both cases we used $\boldsymbol{\beta = 0.9}$ and for the second momentum was used $\boldsymbol{\gamma = 0.7}$, and the learning rate was set to 0.001. The algorithm was implemented in Pytorch by using the detectron2 object detection library.

- Model structure:

  - An improvement of around 1.5% was verified in terms of both mAP and Dice score, which shows the effectiveness of mixing up high and low-level mask information.

- Optimizer:

  - The training with the second-order momentum outperformed the classical momentum up to 2% in both metrics.

Team

Nuno R. Freitas: (PhD Candidate) (nrafeb@gmail.com) -CMEMS-UMinho Research Unit, Universidade do Minho, Guimarães, Portugal

Ismael Vaz: (PhD) -Algorithmi Research Unit, Universidade do Minho, Guimarães, Portugal

Estêvão Lima: (MD, FEBU, PhD) -Life and Health Sciences Research Institute, University of Minho, Campus Gualtar, 4710-057, Braga, Portugal;

-ICVS/3Bs - PT Government Associate Laboratory, Braga/Guimarães, Portugal;

-Department of Urology, CUF Hospitals, 4100-180 Oporto, Portugal

Carlos S. Lima: (PhD) **-**CMEMS-UMinho Research Unit, Universidade do Minho, Guimarães, Portugal

Acknowledgments

NRF gratefully acknowledges Fundação para a Ciência e a Tecnologia for the doctoral grant (SFRH/BD/139061/2018).

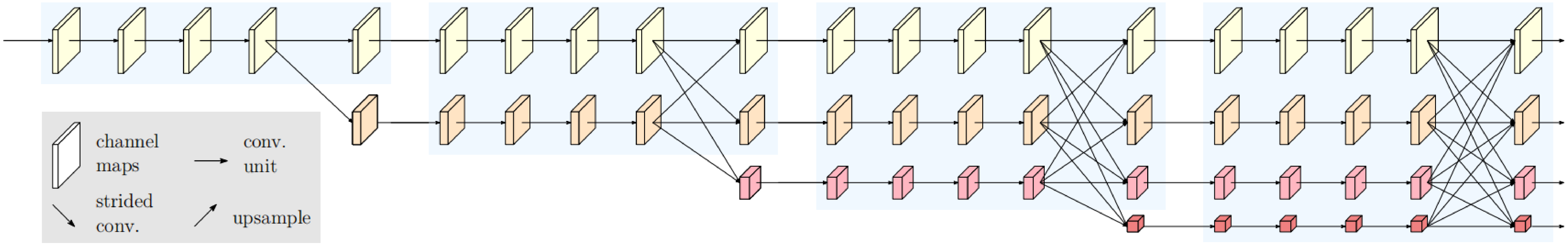
Team UESTC

Endoscope image segmentation based on HRNet

BaoSheng Zou

School of Mechanical and Electrical Engineering,
University of Electronic Science and Technology of China, Sichuan, China

backbone



HRNet

Team UESTC

Data transforms

- Rescale
- Pad
- Crop
- Flip
- Normalize

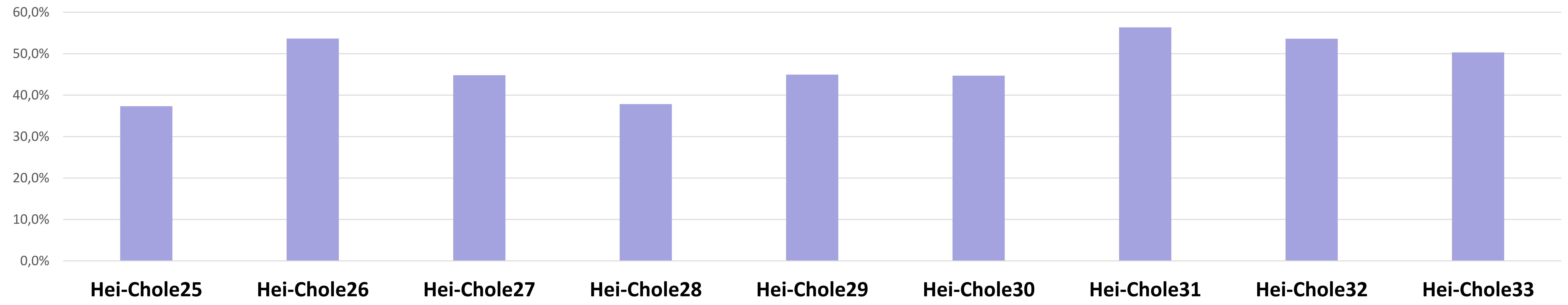
tta

- Ensemble

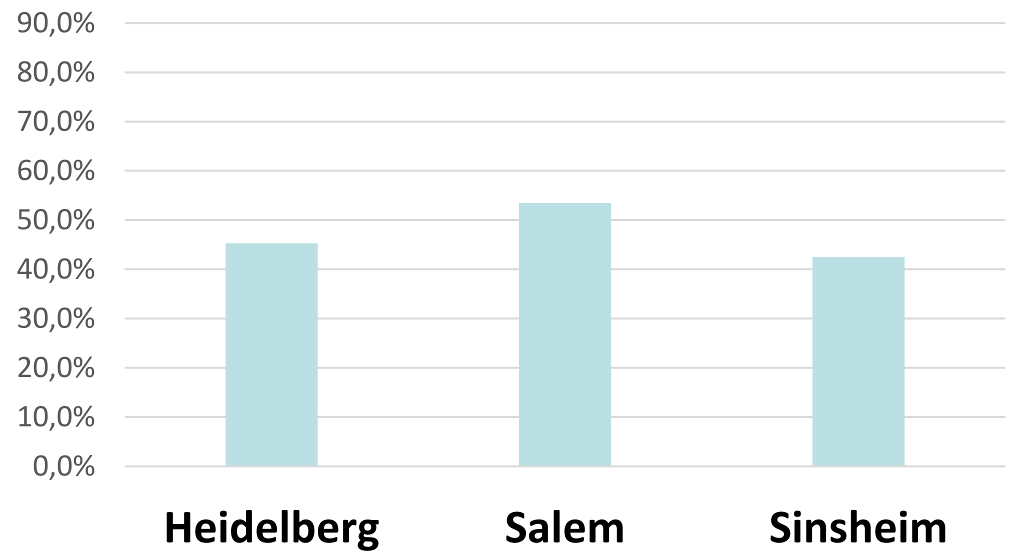
Results

Results: Full Scene Segmentation

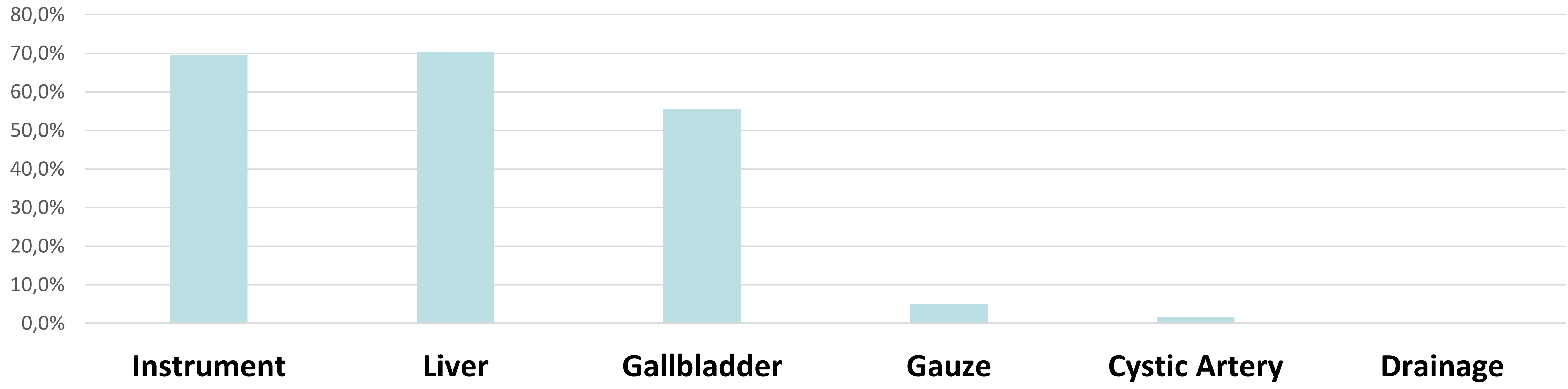
- Submissions
  - 8 teams submitted
  - Average F1-scores between 27% and 57.5%
  - Average NHD between 0.57 and 0.3
- Average F1-score by video



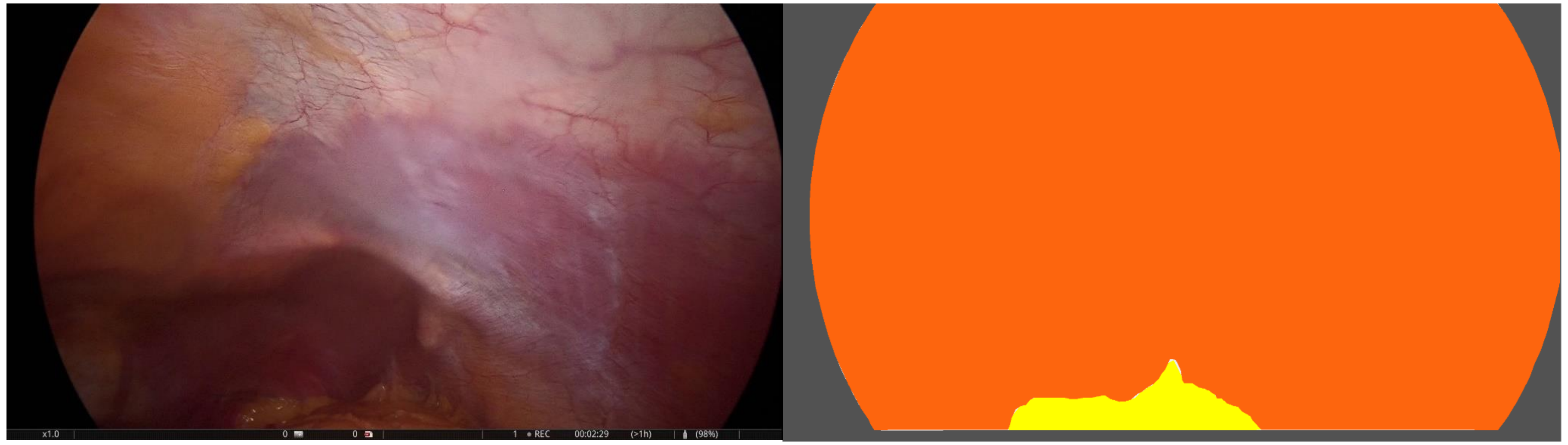
- Average F1-score by center

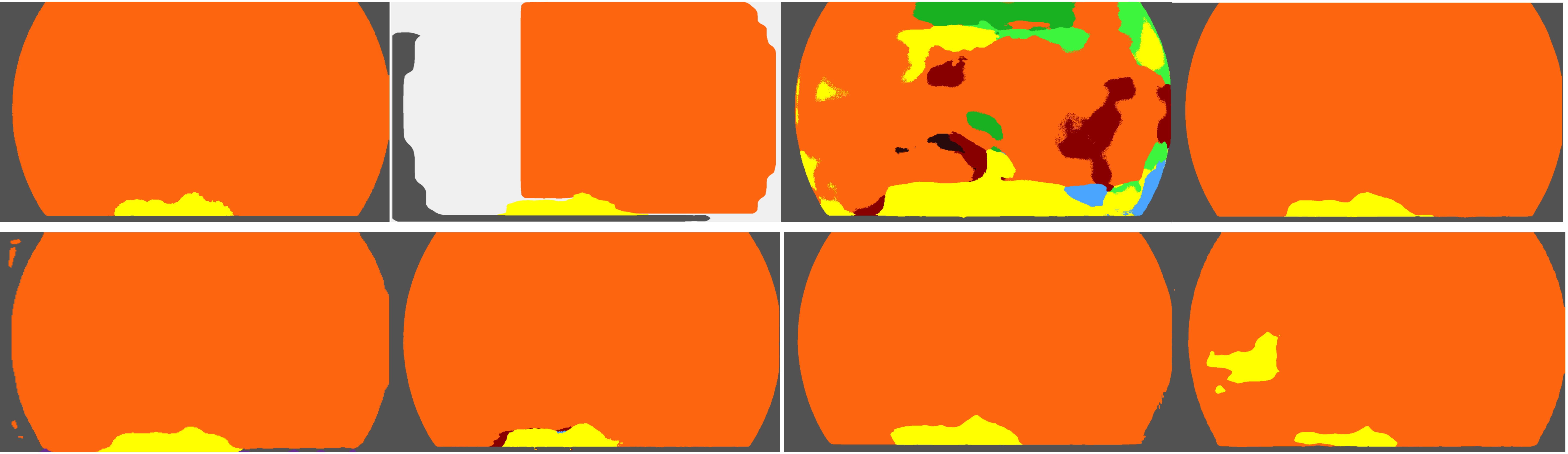


- Average F1-score by class (selection)

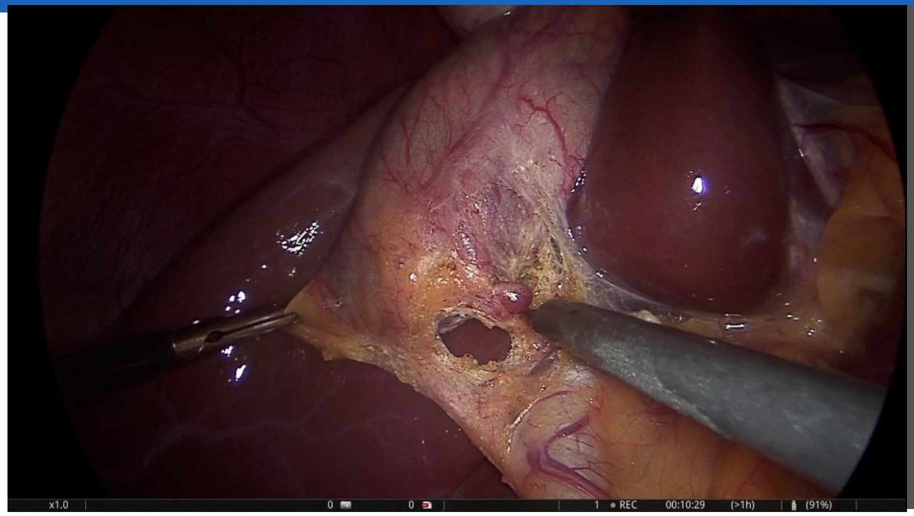


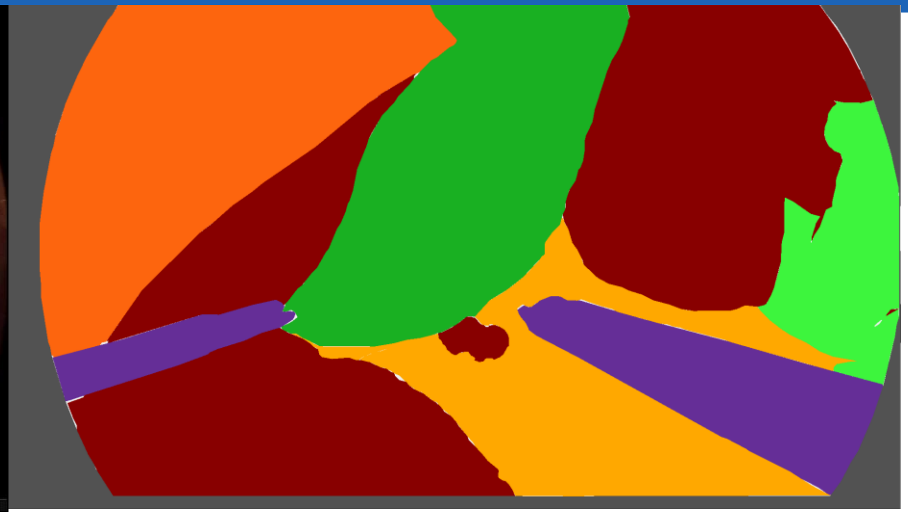
Best examples (according to F1)

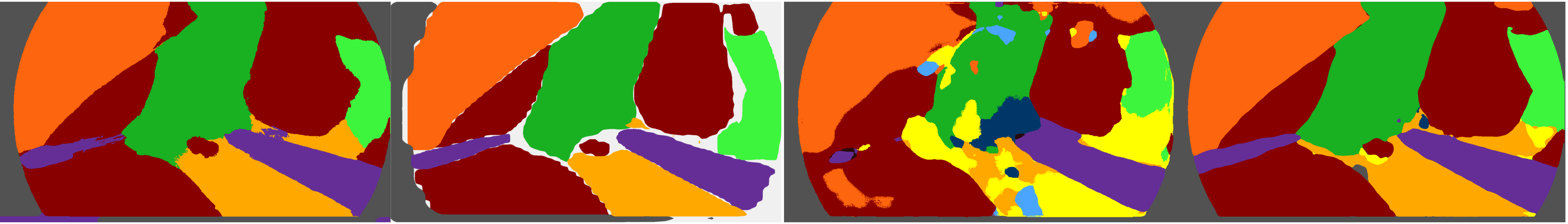


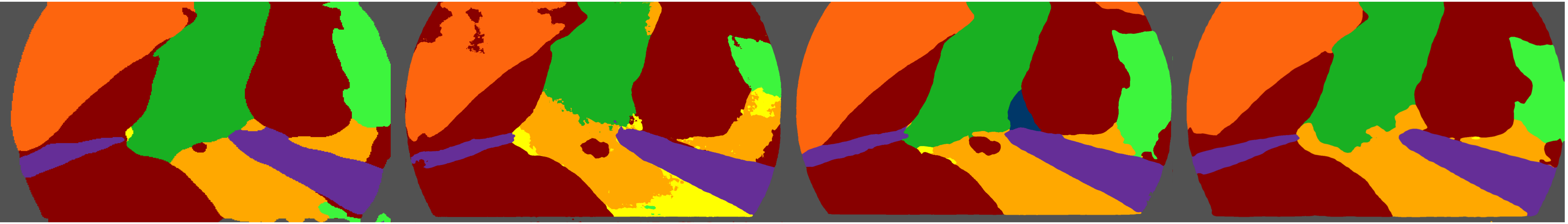


Best examples (according to F1)

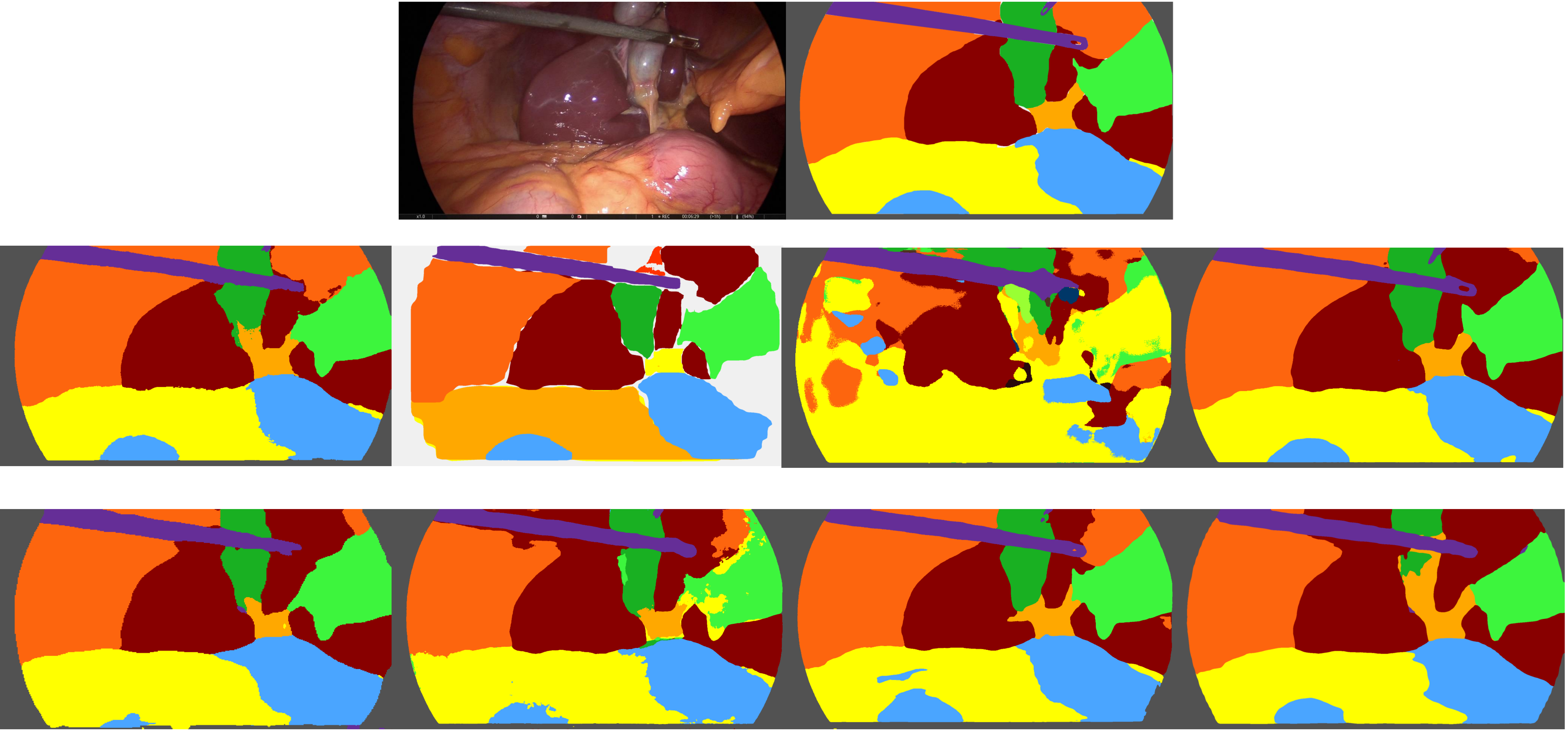




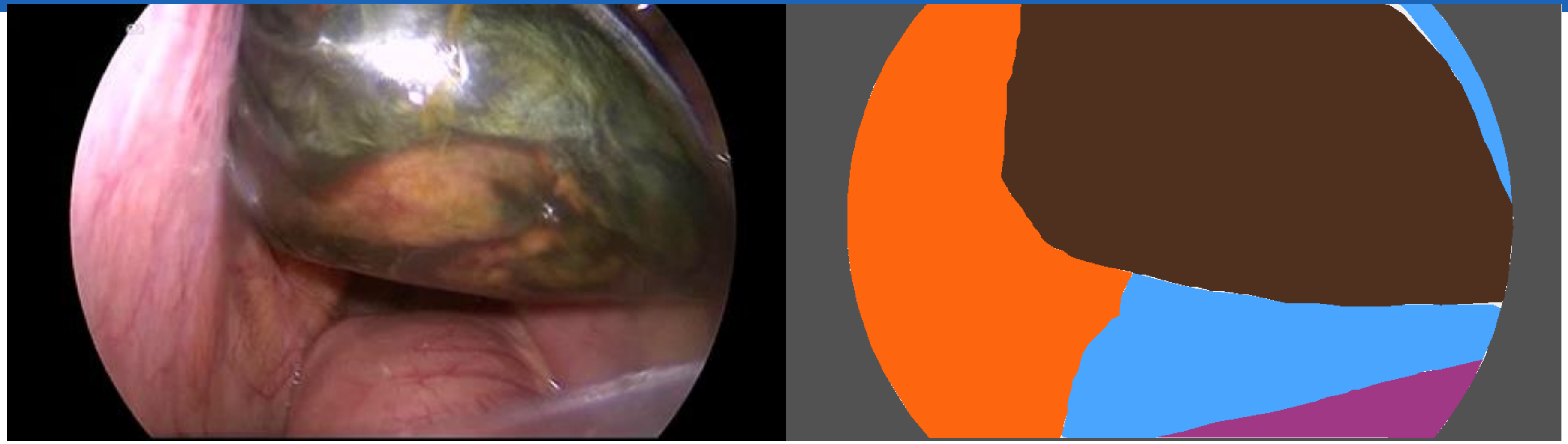


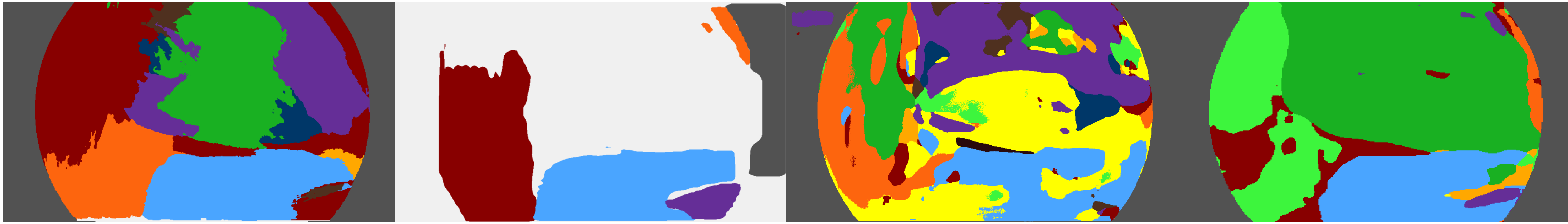


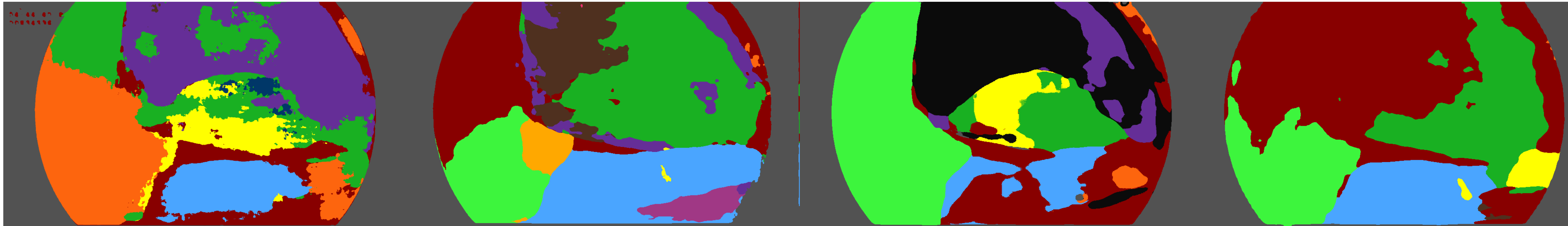
Best examples (according to F1)



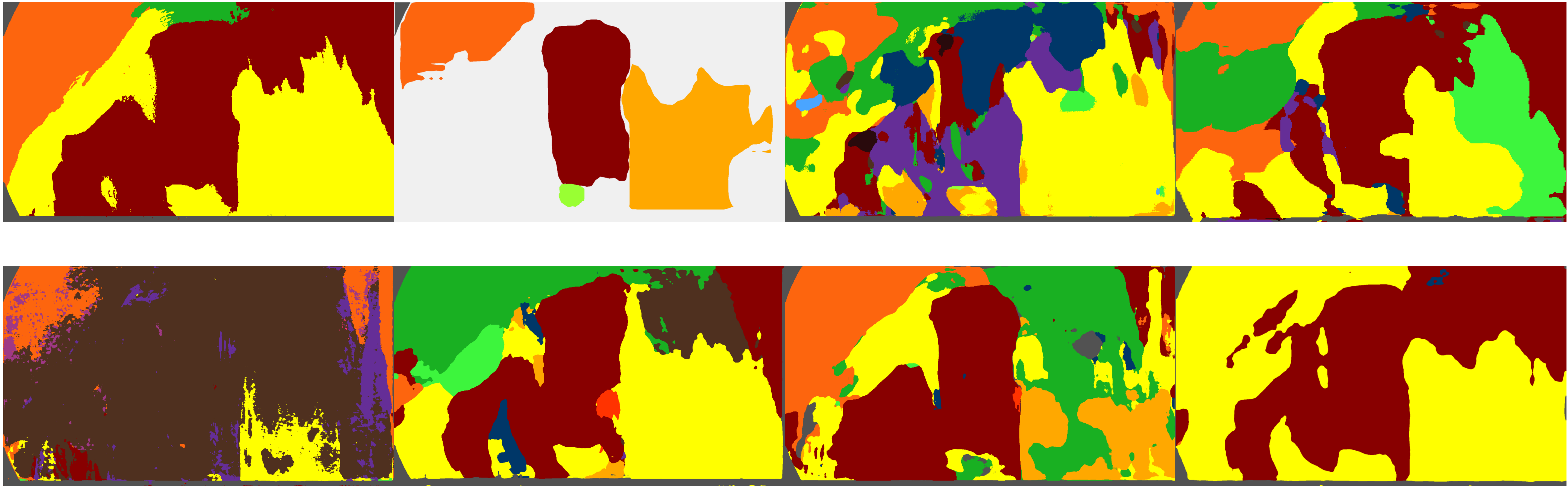
Worst examples (according to F1)





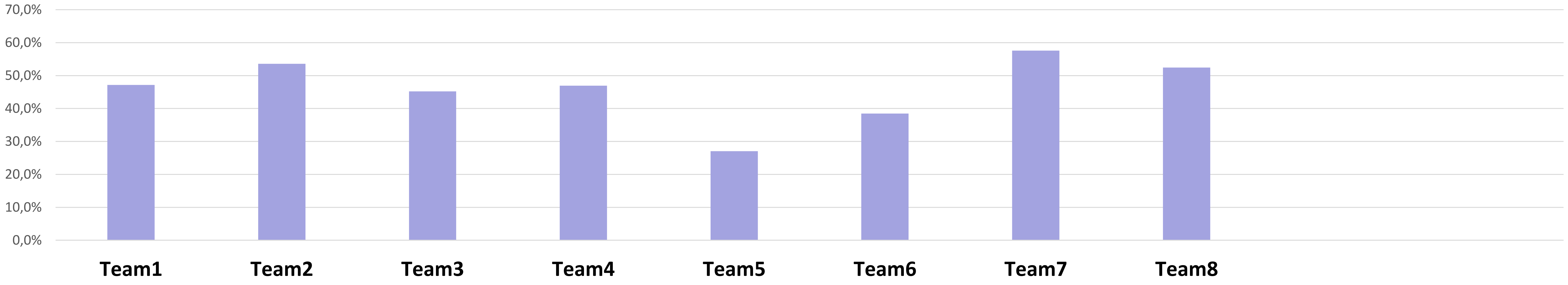


Worst examples (according to F1)

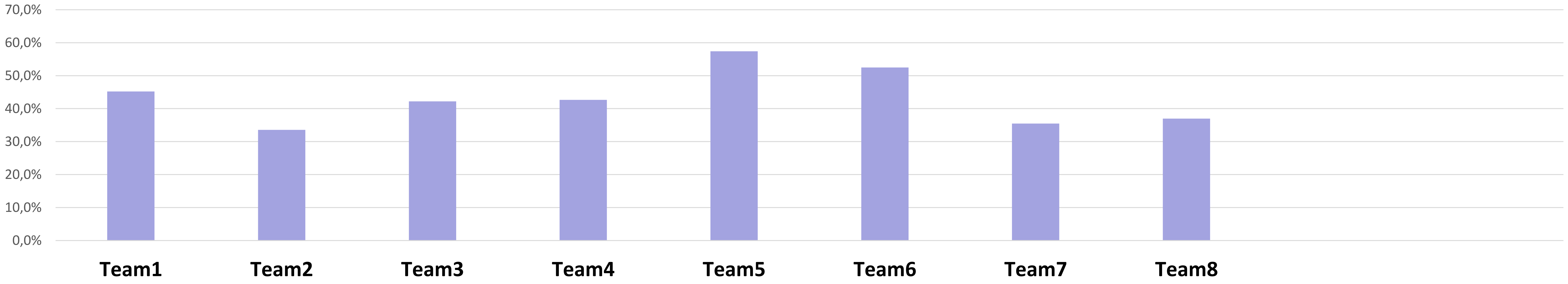


Results: Full Scene Segmentation

- Submissions
  - 8 teams submitted
  - Average F1-scores between 27% and 57.5%
  - Average NHD between 0.57 and 0.3
- Average F1-score by team



- Average NHD-score by team



Ranks

TeamName	Average F1 (# highest)	Average NHD (# lowest)
Team 1	47.2% (24)	0.45 (9)
Team 2	53.6% (54)	0.34 (87)
Team 3	45.2% (10)	0.42 (15)
Team 4	46.9% (14)	0.43 (19)
Team 5	27.0% (0)	0.57 (1)
Team 6	38.5% (4)	0.52 (6)
Team 7	57.6% (99)	0.35 (59)
Team 8	52.4% (43)	0.37 (52)

Ranks

TeamName	Rank F1	Rank NHD
Team 1	4	6
Team 2	2	1
Team 3	6	4
Team 4	5	5
Team 5	8	8
Team 6	7	7
Team 7	1	2
Team 8	3	3

Ranks

TeamName	Rank F1	Rank NHD	Average Rank
Team 1	4	6	5
Team 2	2	1	1.5
Team 3	6	4	5
Team 4	5	5	5
Team 5	8	8	8
Team 6	7	7	7
Team 7	1	2	1.5
Team 8	3	3	3

Ranks

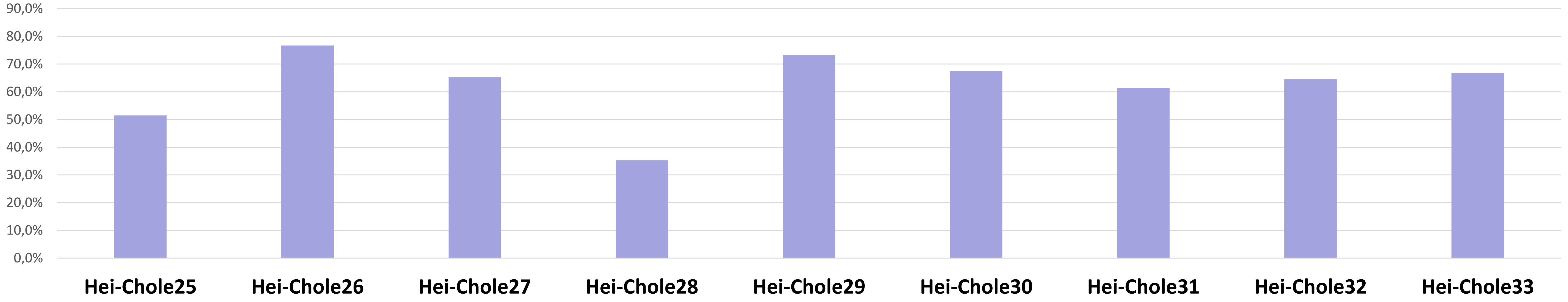
TeamName	Rank F1	Rank NHD	Average Rank
2AI	4	6	5
VisionAI Hutom	2	1	1.5
ArcSeg	6	4	5
Digital Surgery	5	5	5
Wintegral	8	8	8
CMEMS	7	7	7
GMCAO	1	2	1.5
UESTC	3	3	3

KARL STORZ EndoVis Segmentation Award

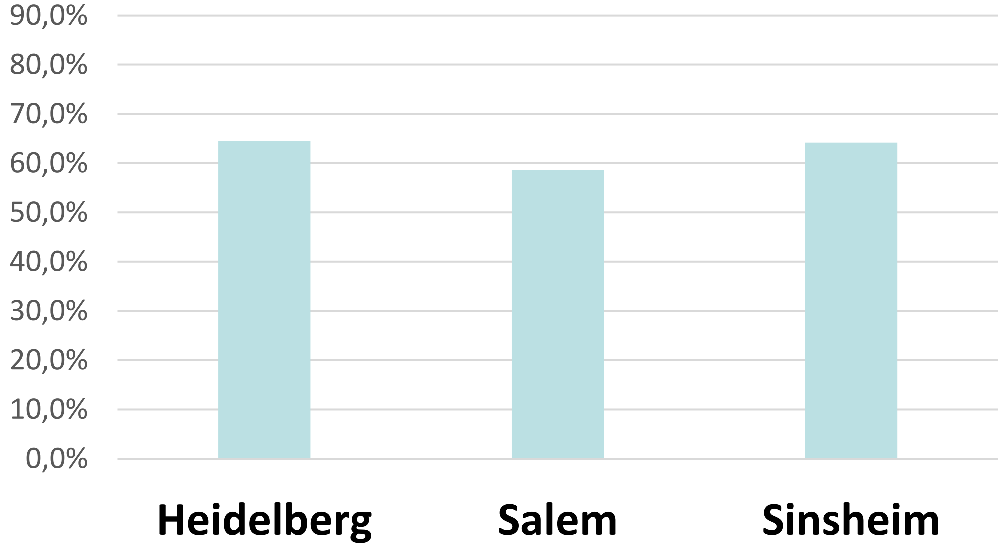
Team VisionAI Hutom
Team TIMIC

Results: Phase segmentation

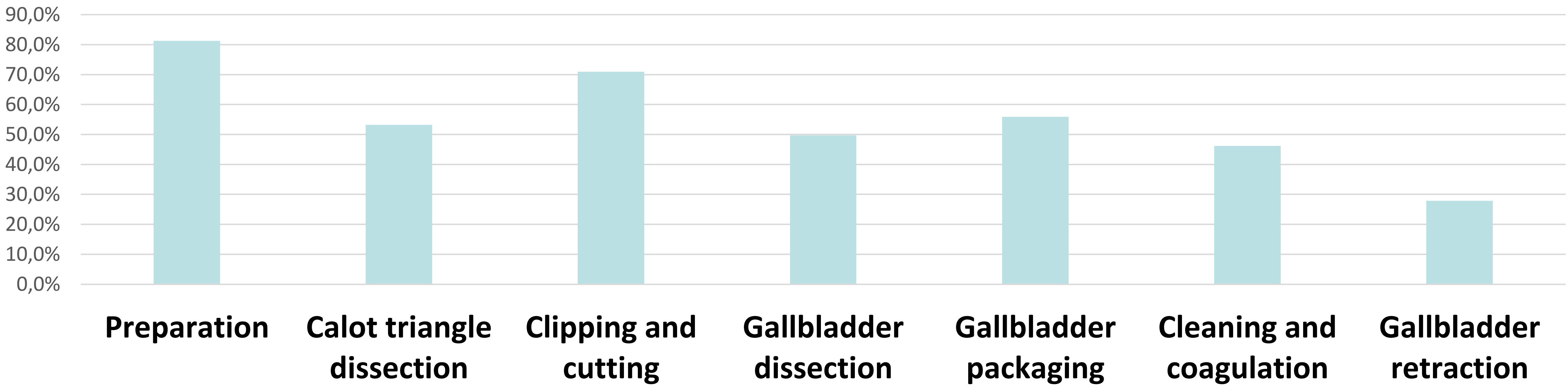
- Submissions
  - 7 teams submitted
  - Average F1-scores between 52.9% and 70.3%
- Average F1-score by video



- Average F1-score by center



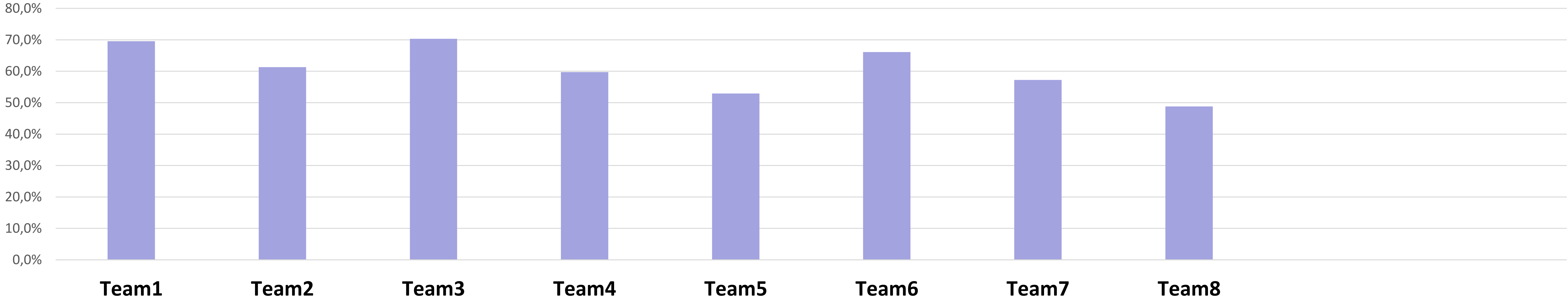
- Average F1-score by phase



Results: Phase segmentation

- Submissions
  - 7 teams submitted
  - Average F1-scores between 52.9% and 70.3%

- Average F1-score by team



Ranks

TeamName	Average F1	# Videos Best
Team 3	70.3%	3
Team 2	69.6%	2
Team 7	66.1%	3
Team 8	61.3%	1
Team 1	59.7%	
Team 4	57.2%	
Team 5	52.9%	

Last years winner: 65.4%

Ranks

TeamName	Average F1	# Videos Best
Digital Surgery	70.3%	3
2AI	69.6%	2
UCL	66.1%	3
VisionAI Hutom	61.3%	1
SIAT-CAMI	59.7%	
Muroran-IT	57.2%	
Wintegral	52.9%	

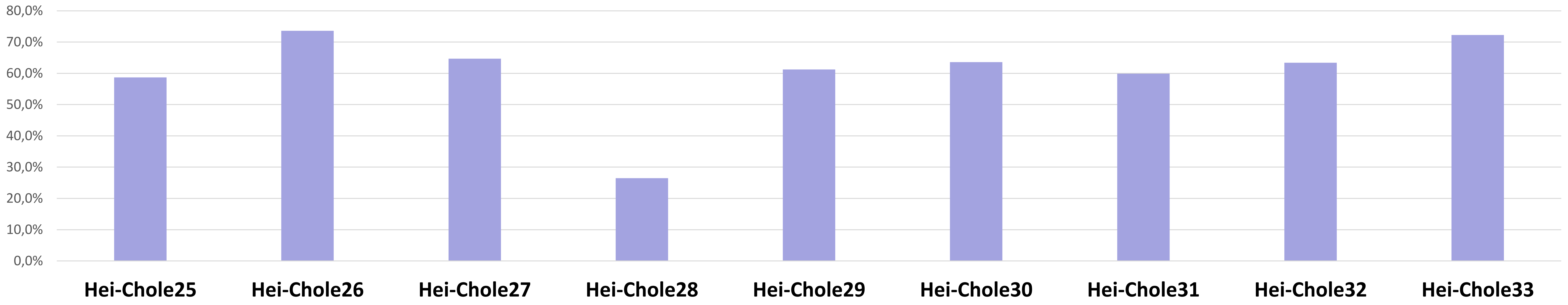
KARL STORZ EndoVis Workflow Award
Category Phase Segmentation

STORZ
KARL STORZ — ENDOSKOPE

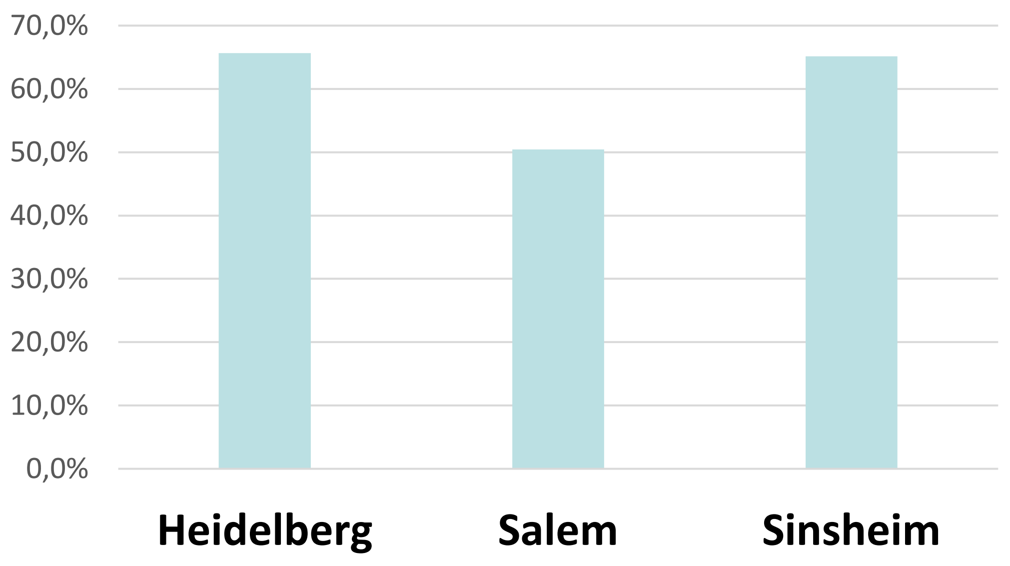
Team Digital Surgery

Results: Instrument presence detection

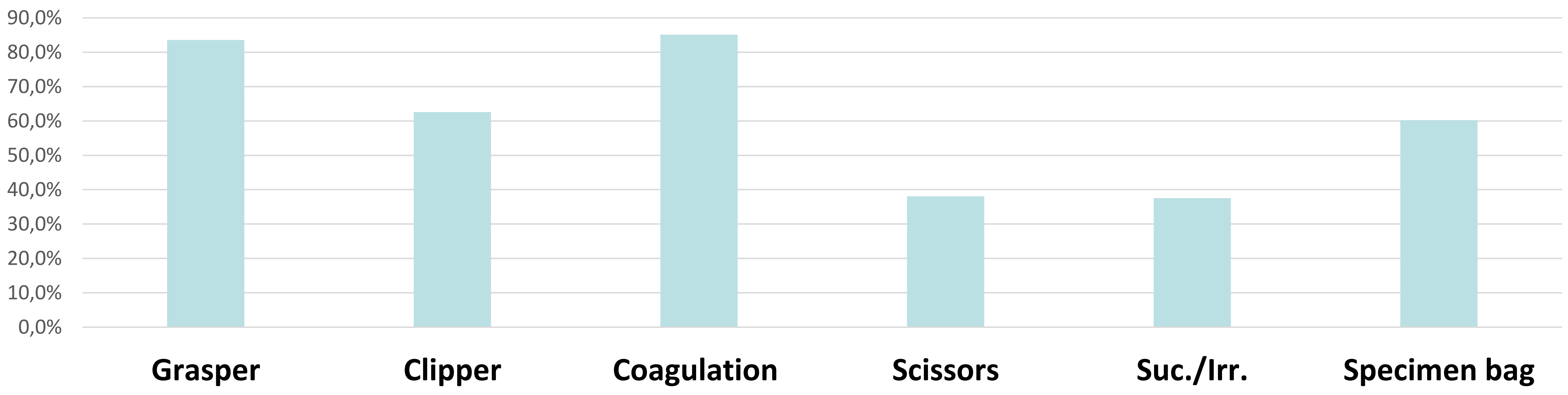
- Submissions
  - 6 teams submitted
  - Average F1-scores between 45.6% and 72.9%
- Average F1-score by video



- Average F1-score by center



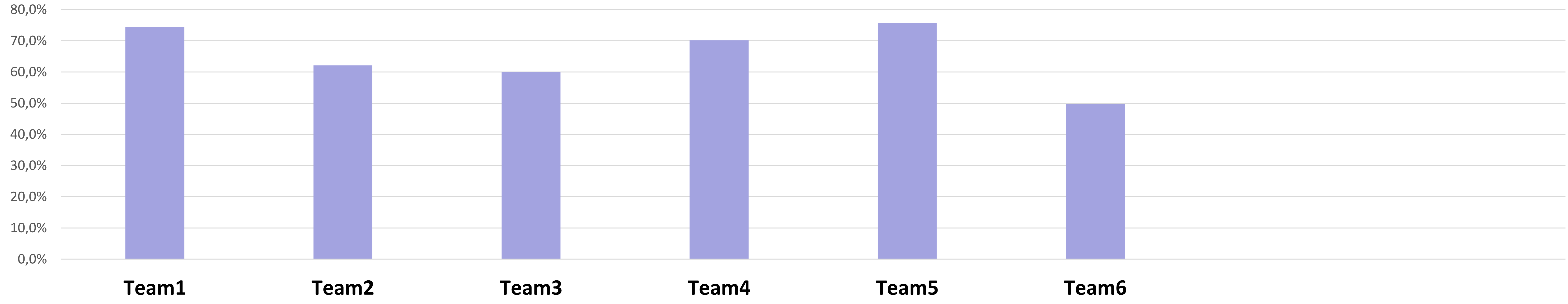
- Average F1-score by instrument type



Results: Instrument presence detection

- Submissions
  - 6 teams submitted
  - Average F1-scores between 45.6% and 72.9%

- Average F1-score by team



Ranks

TeamName	Average F1	# Videos Best
Team 1	72.9%	8
Team 4	65.3%	1
Team 5	65.3%	
Team 3	60.1%	
Team 6	53.3%	
Team 2	45.7%	

Last years winner: 63.8%

Ranks

TeamName	Average F1	# Videos Best
2AI	72.9%	8
SIAT-CAMI	65.3%	1
Wintegral	65.3%	
Digital Surgery	60.1%	
Muroran-IT	53.3%	
VisionAI Hutom	45.7%	

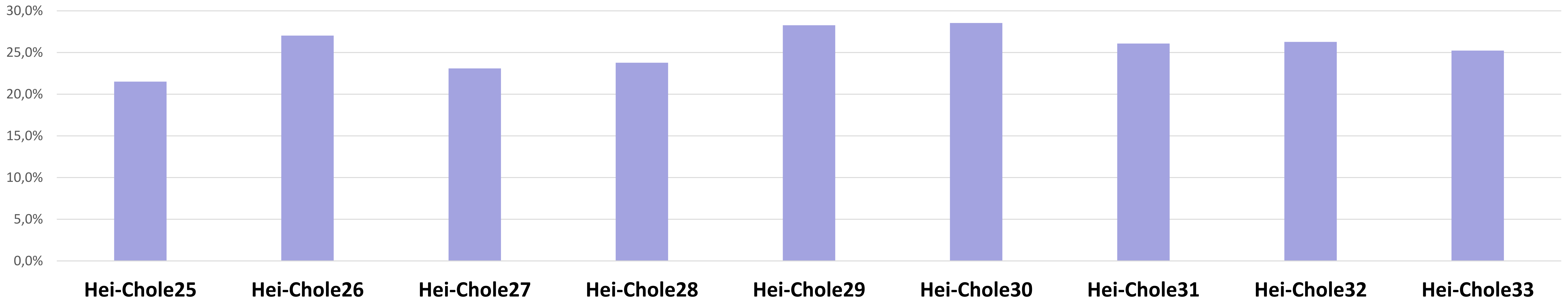
KARL STORZ EndoVis Workflow Award
Category Instrument Detection

STORZ
KARL STORZ — ENDOSKOPE

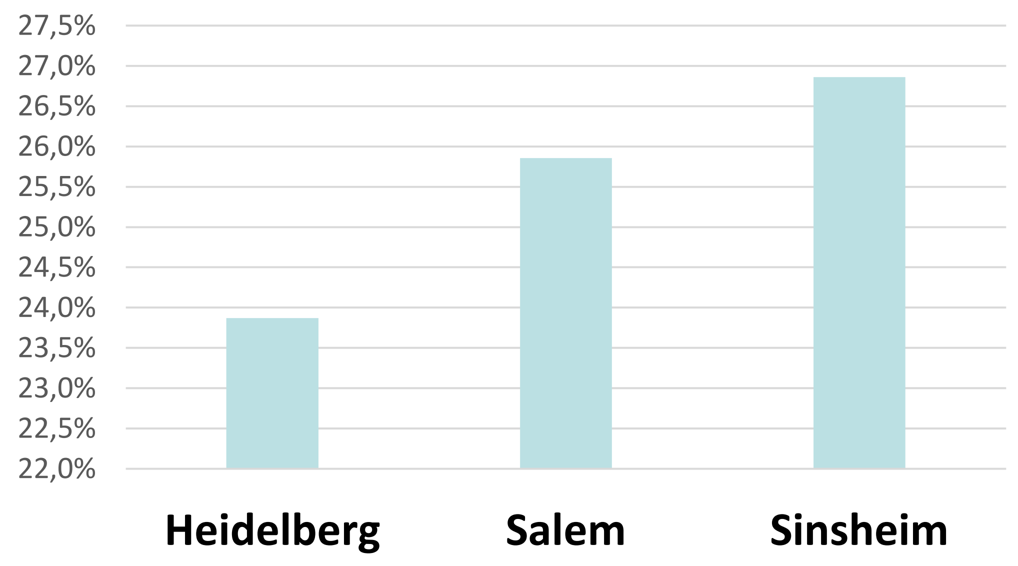
Team 2AI

Results: Action recognition

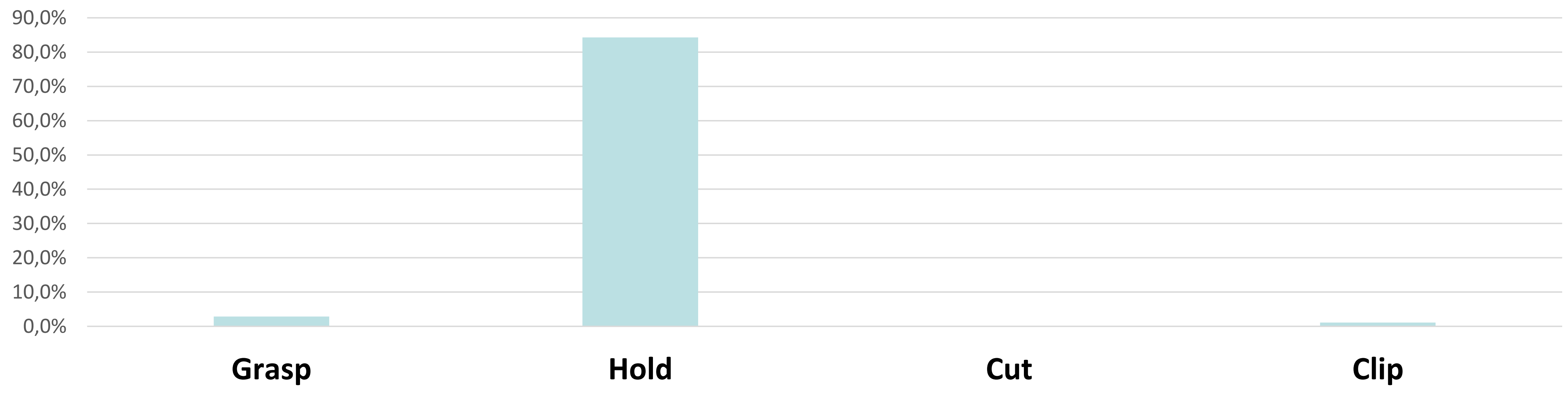
- Submissions
  - 5 teams submitted
  - Average F1-scores between 22% and 30.7%
- Average F1-score by video



- Average F1-score by center



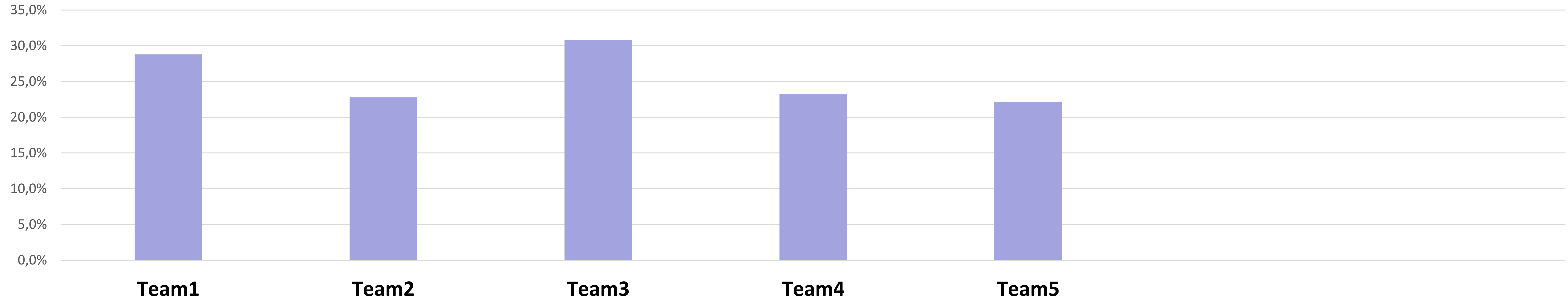
- Average F1-score by action



Results: Action recognition

- Submissions
  - 5 teams submitted
  - Average F1-scores between 22% and 30.7%

- Average F1-score by team



Ranks

TeamName	Average F1	# Videos Best
Team 3	30.8%	5
Team 1	28.8%	4
Team 4	23.2%	
Team 2	22.8%	
Team 5	22.1%	

Last years winner: 23.3%

Ranks

TeamName	Average F1	# Videos Best
Digital Surgery	30.8%	5
2AI	28.8%	4
SIAT-CAMI	23.2%	
VisionAI Hutom	22.8%	
Muroran-IT	22.1%	

Award: Action detection

KARL STORZ EndoVis Workflow Award
Category Action Detection

KARL STORZ — ENDOSKOPE

Team Digital Surgery

Conclusion

- Four tasks
  - Full scene segmentation: F1-scores 27% - 57.5%, NHD 0.57 - 0.3
  - Phase segmentation: F1-scores 52.9% - 70.3%
    - Compare to EndoVis 19[1] winner: 65.4%
  - Instrument presence detection: F1-scores 45.6% - 72.9%
    - Compare to EndoVis 19[1] winner: 63.8%
  - Action recognition: F1-scores 22% - 30.7%
    - Compare to EndoVis 19[1] winner: 23.3%

- 11 teams and 12 submissions
  - 8 submissions segmentation
  - 7 submissions workflow
    - 7 phase segmentation
    - 6 instrument detection
    - 5 action detection

1. https://arxiv.org/abs/2109.14956

Questions?

STORZ
KARL STORZ — ENDOSKOPE

NCT
NATIONAL CENTER FOR TUMOR DISEASES
PARTNER SITE DRESDEN
UNIVERSITY CANCER CENTER UCC

Heidelberg University Hospital

dkfz.