Not End-to-End: Explore Multi-Stage Architecture for Online Surgical Phase Recognition

Fangqiu Yi*, Yanfeng Yang*, and Tingting Jiang✉

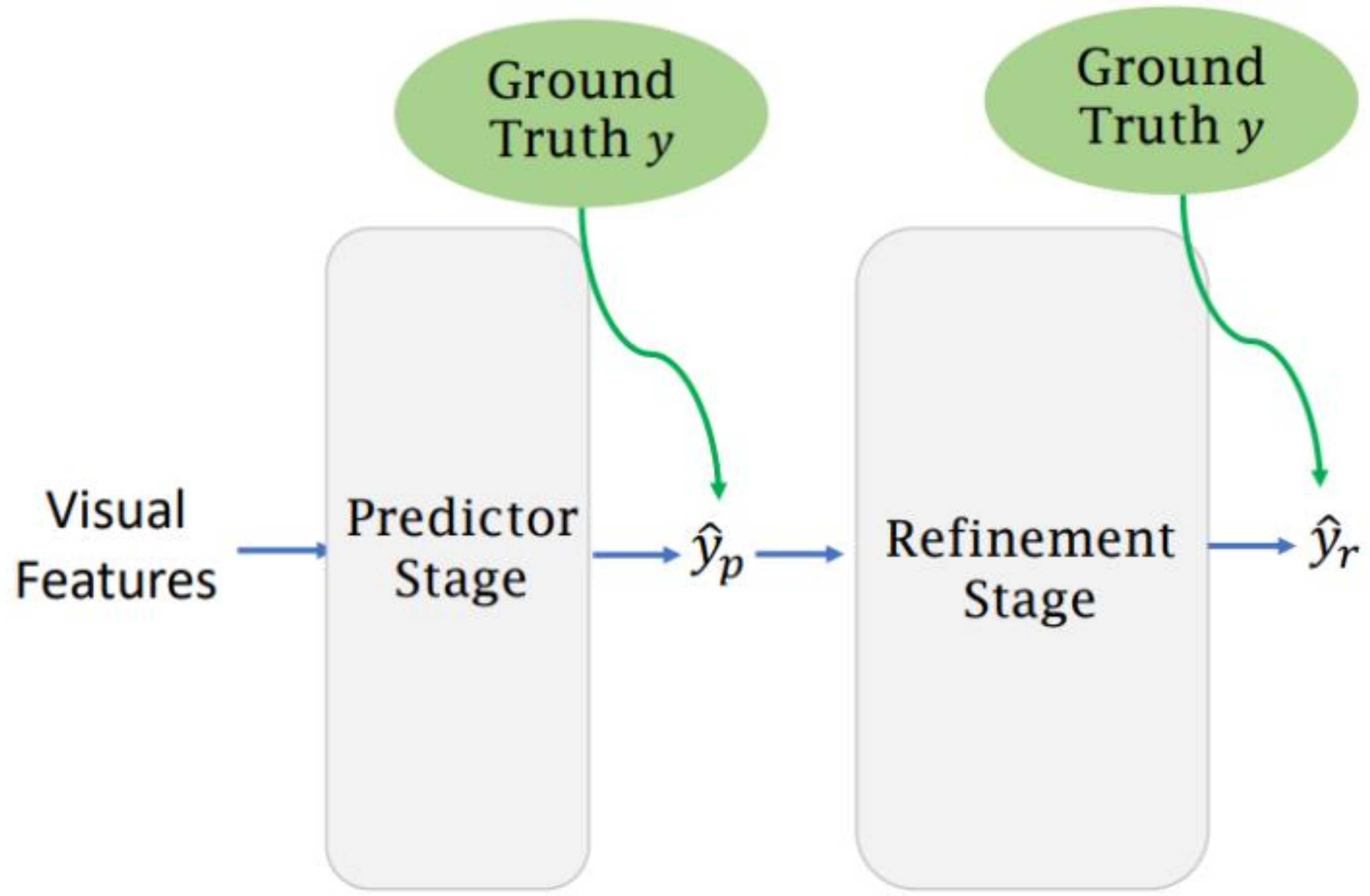
National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China

What is Surgical Phase Recognition?

Predict what surgical phase is occurring at each frame in the surgical videos.

Why Multi-Stage Architecture?

The imperfect predictions can be further refined.
Surgical video contents contain rich temporal patterns.

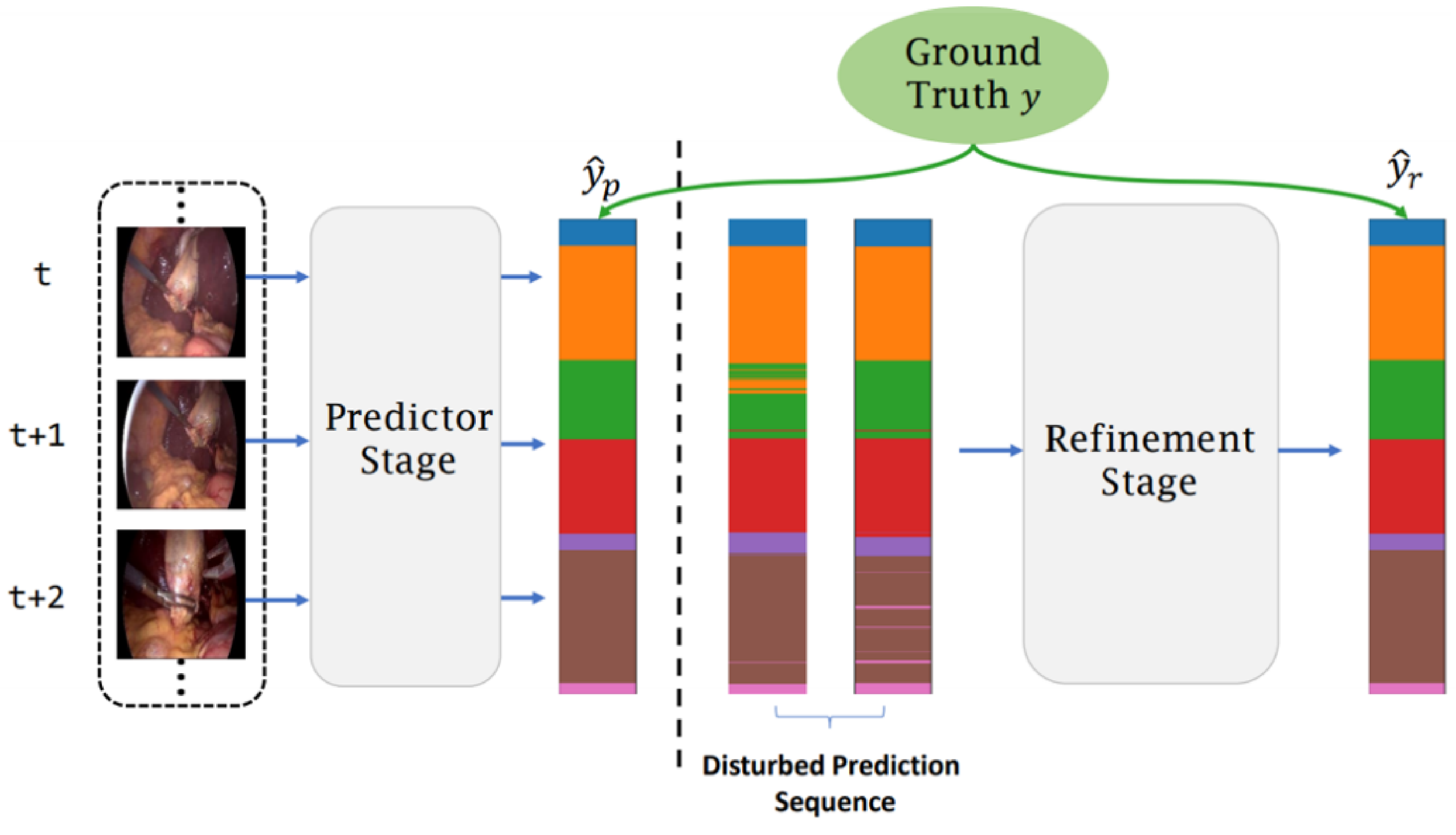


End-to-End not Work in Multi-Stage

1. The inputs of the refinement stage during training and inference are different.
2. The limited size of current datasets.

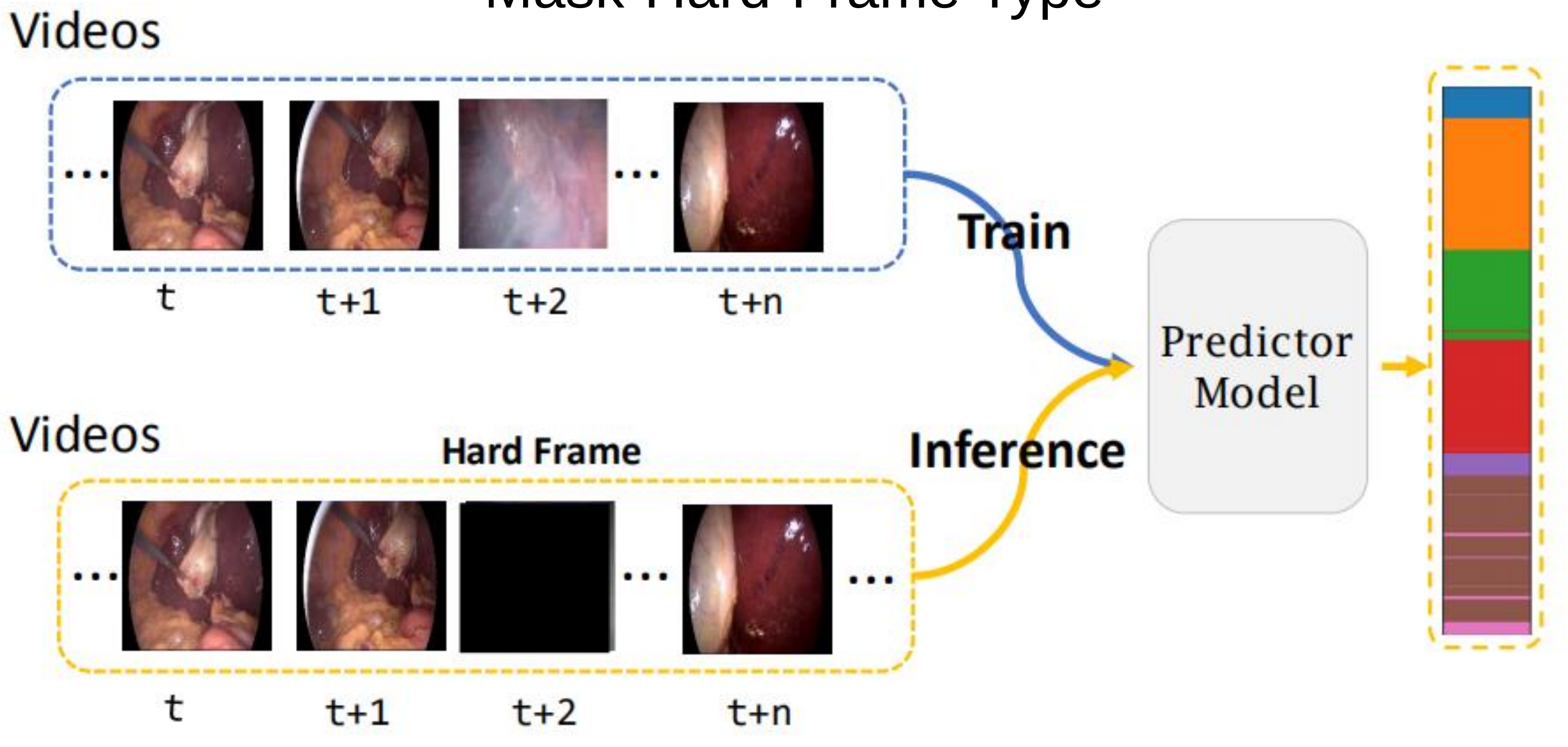
Our Solution

Train predictor stage and refinement stage separately.
Design two types of training sequences to simulate the real output of the predictor during inference.

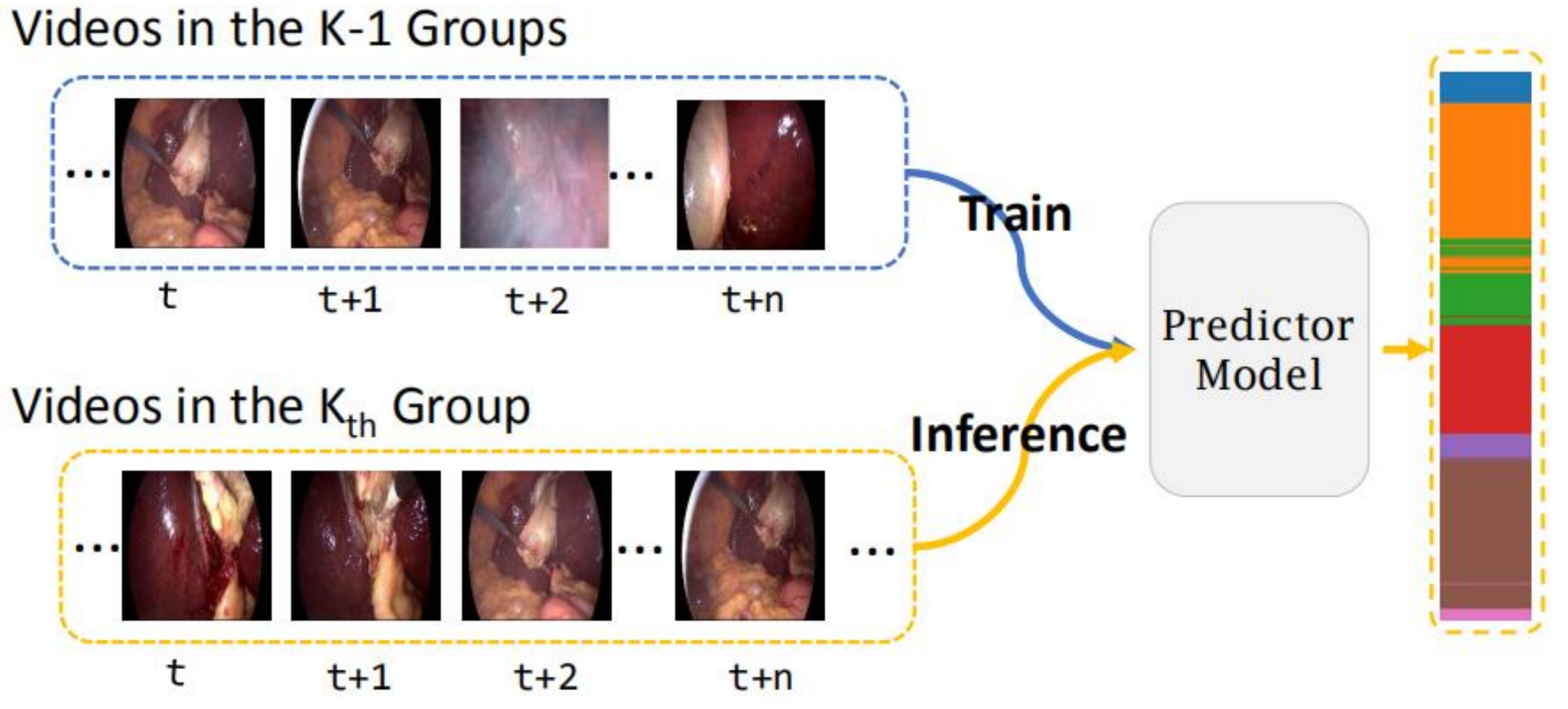


Disturbed Prediction Sequence Generation

Mask-Hard-Frame Type



Cross-Validate Type



Experiments

Comparison with SOTA on Cholec80 dataset

Method	Acc	JACC	Rec
ResNet [8]	78.3±7.7	52.2±15.0	-
PhaseLSTM [25]	80.7±12.9	64.4±10.0	-
PhaseHMM [25]	71.1±20.3	62.4±10.4	-
EndoNet [13]	81.7±4.2	-	79.6±7.9
EndoNet-GTbin [13]	81.9±4.4	-	80.0±6.7
SV-RCNet [7]	85.3±7.3	-	83.5±7.5
OHFM [8]	87.0±6.3	66.7±12.8	-
TeCNO [9]	88.6±2.7	-	85.2±10.6
OperA [20]	85.8±1.0	-	87.7±0.7
Trans-SVNet [21]	90.3±7.1	**79.3±6.6**	**88.8±7.4**
causal TCN	88.8±6.3	73.2±9.8	84.9±7.2
Ours	**92.0±5.3**	77.1±11.5	87.0±7.3

Comparison with End-to-End on Cholec80 dataset

Method	Acc	JACC	Rec
Predictor	88.8±6.3	73.2±9.8	84.9±7.2
End-to-End+GRU	87.1±7.8	69.7±12.6	83.2±9.4
End-to-End+causal TCN	87.7±6.3	77.7±11.2	84.3±6.3
End-to-End+TCN	89.8±6.6	75.8±8.4	87.4±7.5
Ours+GRU	90.8±7.0	75.5±11.1	85.6±10.0
Ours+causal TCN	91.0±5.2	74.2±11.8	84.1±9.6
Ours+TCN	**92.8±5.0**	**78.7±9.4**	**87.5±8.3**

Conclusion

A new non end-to-end training strategy to minimize the distribution gap between the training and inference.

References:

[7] Jin, Y., Dou, Q., Chen, H., Yu, L., Qin, J., Fu, C.W., Heng, P.A.: SV-RCNet: Workflow recognition from surgical videos using recurrent convolutional network. IEEE Transactions on Medical Imaging 37 (2018) 1114–1126

[8] Yi, F., Jiang, T.: Hard frame detection and online mapping for surgical phase recognition. In: Medical Image Computing and Computer Assisted Intervention. (2019)

[13] Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M., Padoy, N.: EndoNet: A deep architecture for recognition tasks on laparoscopic videos. IEEE Transactions on Medical Imaging 36 (2017) 86–97