Method of Team VIE-PKU

EndoVis 2019 Surgical Workflow and Skill Analysis Subchallenge

Mengqi Guo, Beihang University, China, 836918658@qq.com
Chen Gong, Peking University, China, 1700013027@pku.edu.cn
Daochang Liu, Peking University, China, finspire13@gmail.com
(Mengqi Guo and Gong Chen have equal contribution)

2019/10/07

1. Tasks:

- Surgical phase segmentation
- Action recognition
- Instrument presence detection

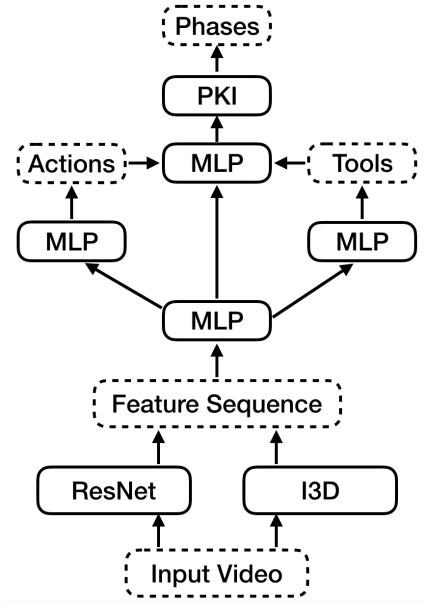
2. Model:

Our model is illustrated on the right. We formulate the three tasks as a multi-task learning problem.

Features are first extracted from the input video using ResNet and I3D. ResNet is pre-trained on ImageNet. I3D is pre-trained on ImageNet and Kinetics. Both ResNet and I3D are fixed during training. The two types of features are mixed by late-fusion.

Features are then sent into several frame-wise multi-layer perceptrons to get the predictions of the three tasks. Since phase segmentation is more high-level, the predictions of actions and instruments are re-wired into the phase MLP. The phase results are post-processed by PKI at last.

To train our model, cross-entropy loss is used for phase task. And weighted binary cross-entropy loss is used for action and instrument tasks to handle class imbalance problem.

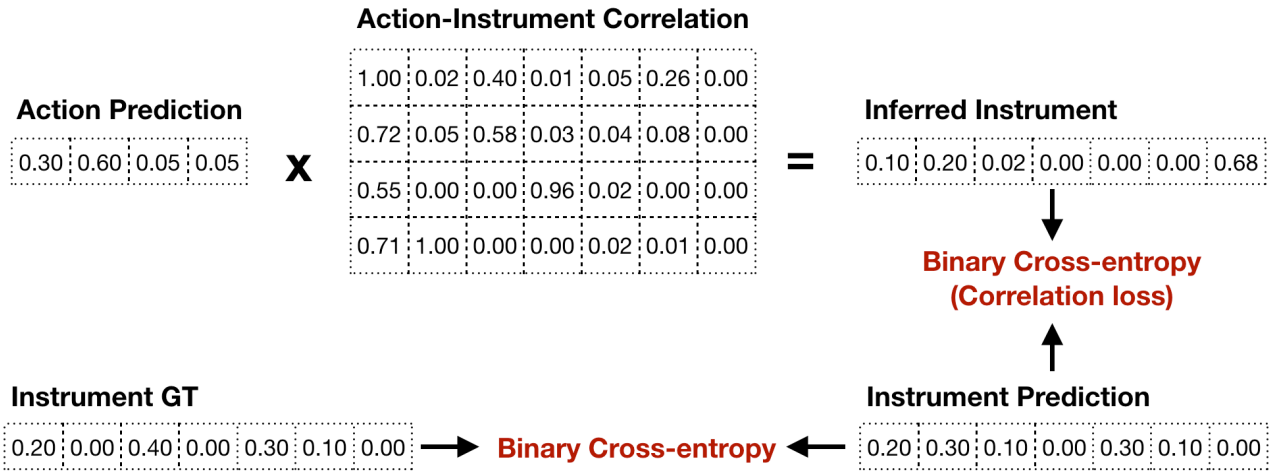


3. Correlation Constraints:

In a surgery, the occurrences of phases, actions and instruments are highly correlated. For example, some instrument only appear in a certain phase. Thus we compute correlation statistics from the training set and use them as loss constraints during training.

3.1 Inter-task correlation:

For each two tasks, we can compute the correlation matrix from the training set. As the action-instrument example shown below. The correlation matrix is the probabilities of occurrence of each instrument given the occurrence of each action, i.e. P(instrument|action). Then during training, we can use the action prediction to infer instrument probabilities, i.e., P(instrument|action) * P(action). Then we compute the binary cross-entropy between the inferred instrument probabilities and the predicted instrument probabilities as a correlation loss.

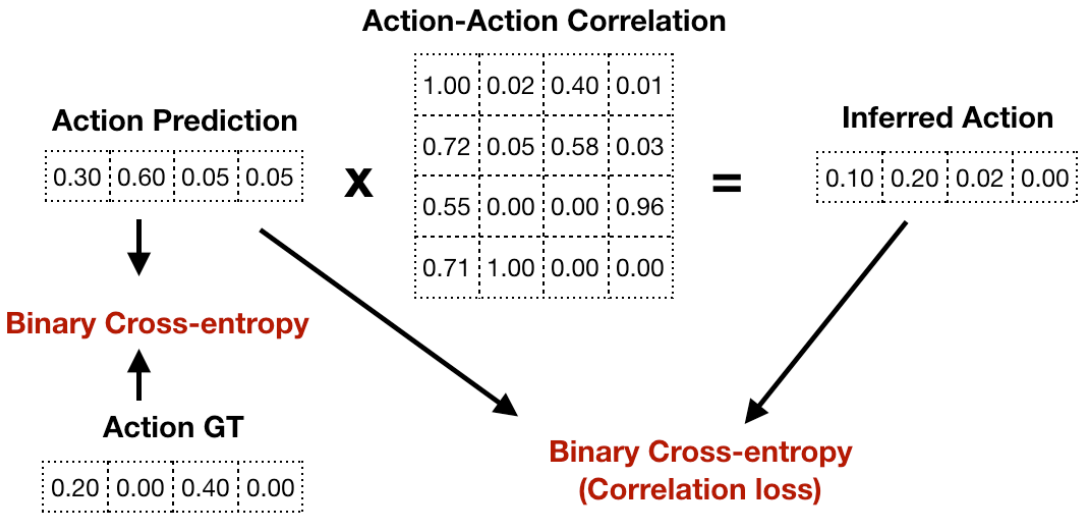


The values are only for demonstration.

In this manner, there will be 6 inter-task correlation losses in total (Action-Phase, Action-Instrument, Phase-Action, Phase-Instrument, Instrument-Action, Instrument-Phase).

3.2 Intra-task correlation:

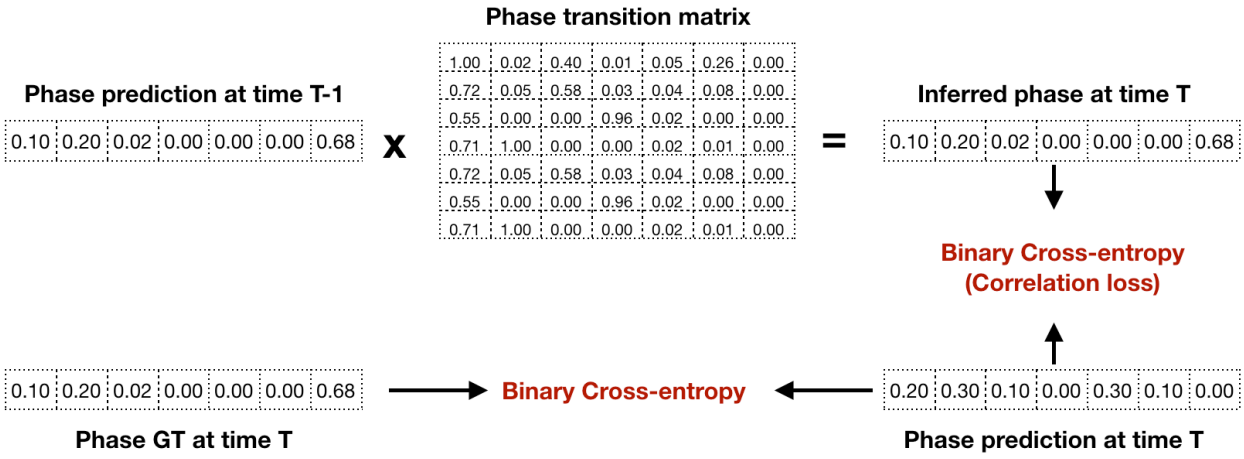
For a single task, there are also correlations between different classes. For example, some action always co-occurs with another action. Therefore, we also design losses for intra-task correlations, as the action example shown below. In this manner, there will be 2 intra-task correlation losses in total (Action, Instrument).



The values are only for demonstration.

3.3 Phase-transition correlation:

Another type of correlation is for phase-transition. Some phase are more likely to transit to certain phase than others. Therefore, as shown below, we also design correlation loss to constrain phase transitions between neighboring video frames. The phase transition matrix is from dataset statistics.



The values are only for demonstration.