

Supplementary Experiment of SKiT: a Fast Key Information Video Transformer for Online Surgical Phase Recognition

Yang Liu, Jiayu Huo, Jingjing Peng, Rachel Sparks, Prokar Dasgupta, Alejandro Granados,
Sebastien Ourselin*

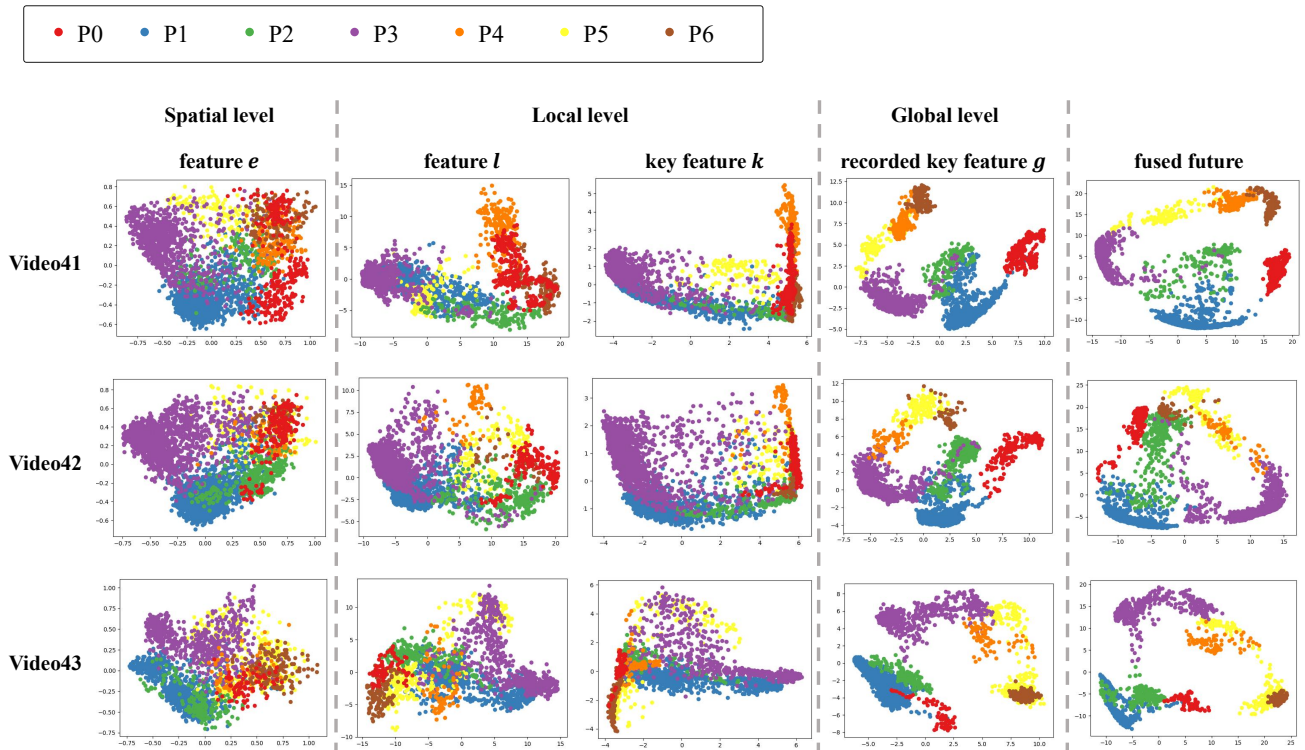
King's College London

{yang.9.liu, firstname.lastname}@kcl.ac.uk

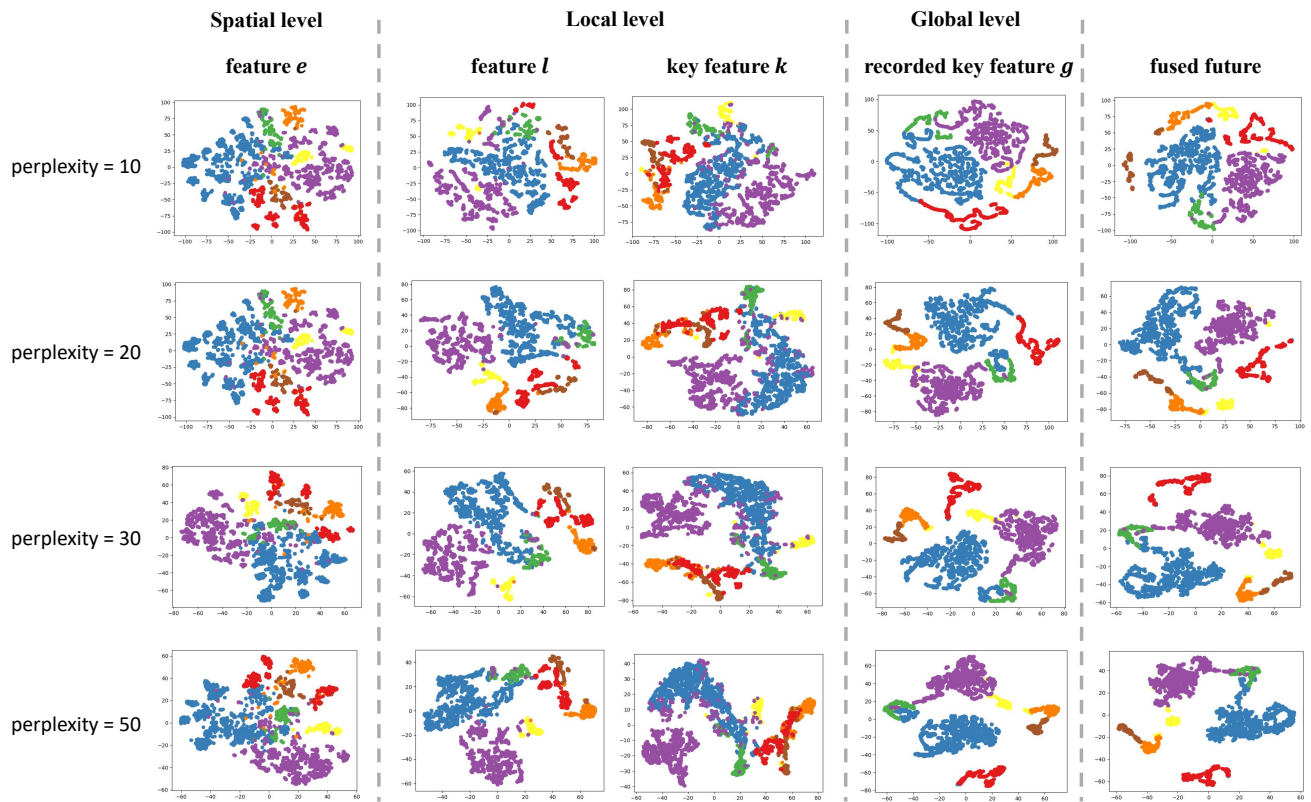
1. Visualization of Different-level Features

Visual examples of different-level features are presented in Figure [SM1](#), obtained through PCA and t-SNE methods. The recorded key feature g and fused feature were used to demonstrate the effectiveness of the proposed Key-recorder in improving phase recognition performance. These findings provide strong evidence of the benefits of the key-recorder.

*Corresponding author



(a) Examples of visualized comparison between different-level features using PCA.



(b) Example of visualized comparison between different-level features using t-SNE with varying perplexity parameters (Video41 on test set of Cholec80).

Figure SM1: Illustrative examples of the different-level features. PCA and t-SNE techniques are employed to visualize various level features, such as spatial feature e , local temporal feature l , local key feature k , global recorded key feature g , and fused feature, using video examples from the Cholec80 test dataset. The phase labels are indicated by different colors, while different points correspond to distinct frame features of the video.