

T-DEED: Temporal-Discriminability Enhancer Encoder-Decoder for Precise Event Spotting in Sports Videos

Supplementary Material

A. Implementation Details for T-DEED

Here we outline the configuration used for each T-DEED model in the SOTA comparison from Table 1 in the main paper.

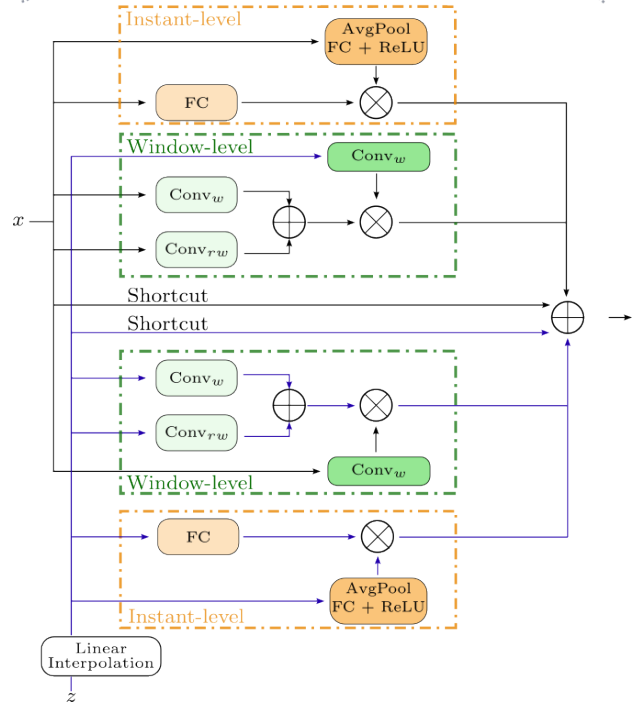
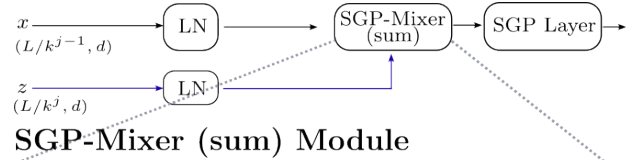
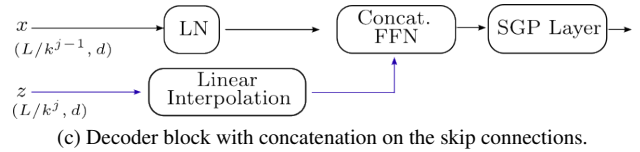
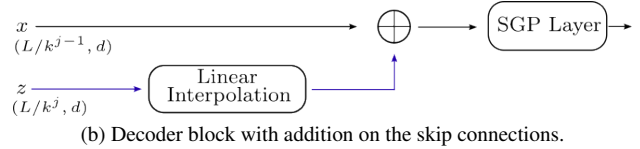
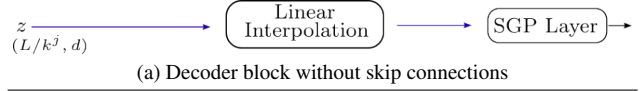
All models utilize data augmentations, including mixup with $\alpha = \beta = 0.2$, color jitter with probability 0.25, and Gaussian blur with probability 0.25. For FigureSkating, frames of size 398×224 are randomly cropped to 224×224 , while for FineDiving and FineGym, frames are resized to 224×224 . The detection radius r_E is set to 2 frames for FineDiving, and 1 frame for FigureSkating. Additionally, we weight the positive classes with $w = 5$ within the cross-entropy loss.

Among model-specific hyperparameters, we have the number of blocks (B), the kernel size (ks), and the scale factor within the SGP module (r). These are chosen independently for FineDiving, FS-Comp, and FS-Perf:

- FineDiving (T-DEED w/ 200MF): $B = 2, ks = 7, r = 4$.
- FineDiving (T-DEED w/ 800MF): $B = 2, ks = 9, r = 4$.
- FS-Comps(T-DEED w/ 200MF): $B = 3, ks = 5, r = 2$.
- FS-Comp (T-DEED w/ 800MF): $B = 2, ks = 9, r = 4$.
- FS-Perf (T-DEED w/ 200MF): $B = 3, ks = 9, r = 2$.
- FS-Perf (T-DEED w/ 800MF): $B = 2, ks = 9, r = 4$.

B. Decoder blocks alternative approaches

In Figure 1, we illustrate different approaches to the decoder block design and the information aggregation within skip connections. Figure 1a depicts the case without skip connections, where information from the previous layer undergoes linear interpolation and an SGP Layer processes it. Figure 1b illustrates aggregation through addition, where upsampled features from the previous layer are added to skip connection features before further processing with an SGP Layer. Similarly, Figure 1c shows concatenation-based aggregation, where concatenated features undergo linear projection before SGP Layer processing. Finally, Figure 1d showcases an approach akin to our SGP-Mixer layer but using addition instead of concatenation and linear projection for feature aggregation.



(d) Decoder block with SGP-Mixer merging branches information with addition.

Figure 1. Decoder block alternative approaches.