Supplemental Materials on Video Object Segmentation Using Global and Instance Embedding Learning

Wenbin Ge\textsuperscript{1,}\textsuperscript{*}, Xiankai Lu\textsuperscript{2,}\textsuperscript{†}, Jianbing Shen\textsuperscript{3,1}

\textsuperscript{1} Beijing Institute of Technology \quad \textsuperscript{2} School of Software, Shandong University \quad \textsuperscript{3} Inception Institute of Artificial Intelligence

gewenbin292@bit.edu.cn, carrierlxk@gmail.com

1. Results on YouTube-VOS 2019 validation dataset

We report complete results of our proposed method on YouTube-VOS 2019 validation dataset \cite{4}. Table 1 shows results of STM \cite{2}, LWL \cite{1} and our proposed method.

\begin{table}[h]
\centering
\begin{tabular}{c|cccc}
 & STM \cite{2} & LWL \cite{1} & Ours_{fast} & Ours \\
\hline
Overall & 79.2 & 81.0 & 79.1 & 80.5 \\
Mean J \uparrow & 79.6 & \textbf{79.6} & 78.2 & 79.5 \\
Mean F \uparrow & 83.6 & \textbf{83.8} & 82.3 & 83.6 \\
Mean J \uparrow & 73.0 & \textbf{76.4} & 73.8 & 76.0 \\
Mean F \uparrow & 80.6 & \textbf{84.2} & 82.1 & 82.9 \\
\end{tabular}
\caption{Evaluation of O-VOS on Youtube-VOS\textsubscript{19} val set, with region similarity J and boundary accuracy F. “Overall”: averaged over the four metrics.}
\end{table}

2. Visualization

We present a qualitative evaluation of our proposed method on each sequence of DAVIS\textsubscript{17} val dataset \cite{3}, DAVIS\textsubscript{17} test-dev dataset \cite{3} and Youtube-VOS val dataset \cite{4} including Youtube-VOS\textsubscript{18} and Youtube-VOS\textsubscript{19}. Specifically, Figure. 1 shows the visualization results of Ours on DAVIS\textsubscript{17} val dataset \cite{3}. Figure. 2 shows the visualization results of Ours_{fast} on DAVIS\textsubscript{17} val dataset \cite{3}. Figure. 3 shows the visualization results of Ours on DAVIS\textsubscript{17} test-dev dataset \cite{3}. Figure. 4 shows the visualization results of Ours_{fast} on DAVIS\textsubscript{17} test-dev dataset \cite{3}. Figure. 5 shows the visualization results of Ours on Youtube-VOS val dataset \cite{4}. Figure. 6 shows the visualization results of Ours_{fast} on Youtube-VOS val dataset \cite{4}.

References


\cite{4} Ning Xu, Linjie Yang, Yuchen Fan, Jianchao Yang, Dingcheng Yue, Yuchen Liang, Brian Price, Scott Cohen, and Thomas Huang. Youtube-vos: Sequence-to-sequence video object segmentation. In ECCV, 2018.
Figure 1: Qualitative results of Ours on DAVIS17 val dataset [3]. From top to bottom are bike-packing, blackswan, bmx-trees, dogs-jump, horsejump-high and judo.

Figure 2: Qualitative results of Ours\textunderscore fast on DAVIS17 val dataset [3]. From top to bottom are bike-packing, blackswan, bmx-trees, dogs-jump, horsejump-high and judo.
Figure 3: Qualitative results of Ours on DAVIS$_{17}$ test-dev dataset [3]. From top to bottom are horsejump-stick, hoverboard, mtb-race, skate-jump, tandem and tractor.

Figure 4: Qualitative results of Ours$_{fast}$ on DAVIS$_{17}$ test-dev dataset [3]. From top to bottom are horsejump-stick, hoverboard, mtb-race, skate-jump, tandem and tractor.
Figure 5: Qualitative results of Ours on Youtube-VOS val dataset [4]. From top to bottom are 3dd327ab4e, 45d898acc4, 63ca8970d9, 83a5056a16, 193aa74f36 and 3674b2c70a.

Figure 6: Qualitative results of Ours_fast on Youtube-VOS val dataset [4]. From top to bottom are 3dd327ab4e, 45d898acc4, 63ca8970d9, 83a5056a16, 193aa74f36 and 3674b2c70a.