method	venue	high-level summary
CtxSyn [5]	CVPR 2018	estimate $F_{0\to1}$ and $F_{1\to0}$, extract features and splat them to time t, synthesize the output I_t using a neural network
Super SloMo [3]	CVPR 2018	estimate $F_{0\to1}$ and $F_{1\to0}$, obtain $F_{t\to0}$ and $F_{t\to1}$ and visibilities using a network, backwarp and merge I_0 and I_1
DAIN [1]	CVPR 2019	estimate $F_{0\to1}$ and $F_{1\to0}$, splat flows subject to Z, extract and backwarp features to time t, obtain I_t using a network
SoftSplat [6]	CVPR 2020	estimate $F_{0\to1}$ and $F_{1\to0}$, extract feature pyramids and splat them to time t, synthesize I_t using a neural network
EDSC [2]	arXiv 2020	estimate kernels as well as masks and offsets at t using a network, apply deformable convolutions and merge the images
BMBC [7]	ECCV 2020	estimate $F_{t\to0}$ and $F_{t\to1}$ as well as $F_{0\to1}$ and $F_{1\to0}$, extract features and splat them to t, obtain I_t using a network
AnimeInterp [4]	CVPR 2021	estimate coarse $F_{0\to1}$ and $F_{1\to0}$, refine flows recurrently, extract feature and splat them, synthesize I_t using a network
XVFI [8]	ICCV 2021	extract features, estimate $F_{0\to1}$ and $F_{1\to0}$, splat flows and refine them, backwarp features and obtain I_t using a network
Ours	N/A	estimate $F_{0\to1}$ and $F_{1\to0}$, optionally upsample them, splat flows to get $F_{t\to0}$ and $F_{t\to1}$, backwarp and merge I_0 and I_1

Table 1: Overview of recent video frame interpolation approaches that take two images as input and yield an interpolation result at an arbitrary time. For brevity we use the term features in the high-level summary to refer to various feature representations such as the input images themselves, contextual features, feature pyramids, depth maps, edge maps, occlusion masks, and visibility maps.

References

[1] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-Aware Video Frame Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[2] Xianhang Cheng and Zhenzhong Chen. Multiple Video Frame Interpolation via Enhanced Deformable Separable Convolution. *arXiv/2006.08070*, 2020.

[3] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik G. Learned-Miller, and Jan Kautz. Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[4] Siyao Li, Shiyu Zhao, Weijiang Yu, Wenxiu Sun, Dimitris N. Metaxas, Chen Change Loy, and Ziwei Liu. Deep Animation Video Interpolation in the Wild. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.

[5] Simon Niklaus and Feng Liu. Context-Aware Synthesis for Video Frame Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[6] Simon Niklaus and Feng Liu. Softmax Splatting for Video Frame Interpolation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[7] Junheum Park, Keunsoo Ko, Chul Lee, and Chang-Su Kim. BMBC: Bilateral Motion Estimation With Bilateral Cost Volume for Video Interpolation. In *European Conference on Computer Vision*, 2020.

[8] Hyeonjun Sim, Jihyong Oh, and Munchurl Kim. XVFI: eXtreme Video Frame Interpolation. In *IEEE International Conference on Computer Vision*, 2021.