Towards Holistic Modeling for Video Frame Interpolation with Auto-regressive Diffusion Transformers

Supplementary Material

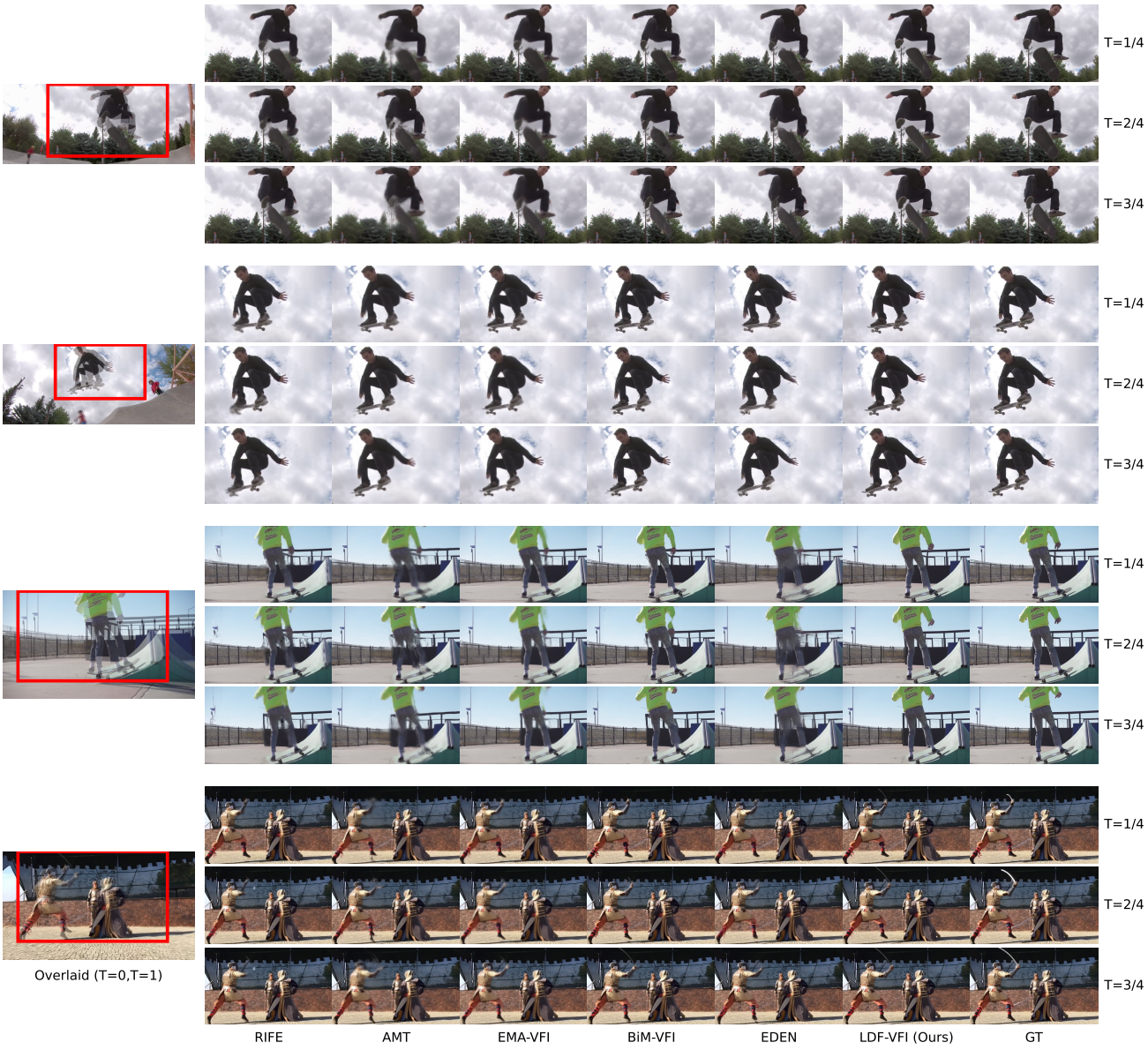

Figure A1. Qualitative comparisons of 16× interpolation on SNU-FILM-entire benchmark.

More visualization We provide additional visual comparisons in Figure A1.

Impact of temporal chunk length Table A1 compares chunk sizes under similar training computation (32K steps for length 5 vs. 8K steps for length 20). A longer chunk (20) improves temporal continuity, e.g., reducing X4K FVD from 510.63 to 121.29.

Table A1. Impact of temporal chunk length under similar total computation.

Temporal chunk	SNU-FILM-8×		X4K-16×	
	LPIPS↓	FVD↓	LPIPS↓	FVD↓
Length 5	**0.059**	42.15	0.198	510.63
Length 20	0.066	**24.87**	**0.126**	**121.29**