

Compressing Video Calls using Synthetic Talking Heads

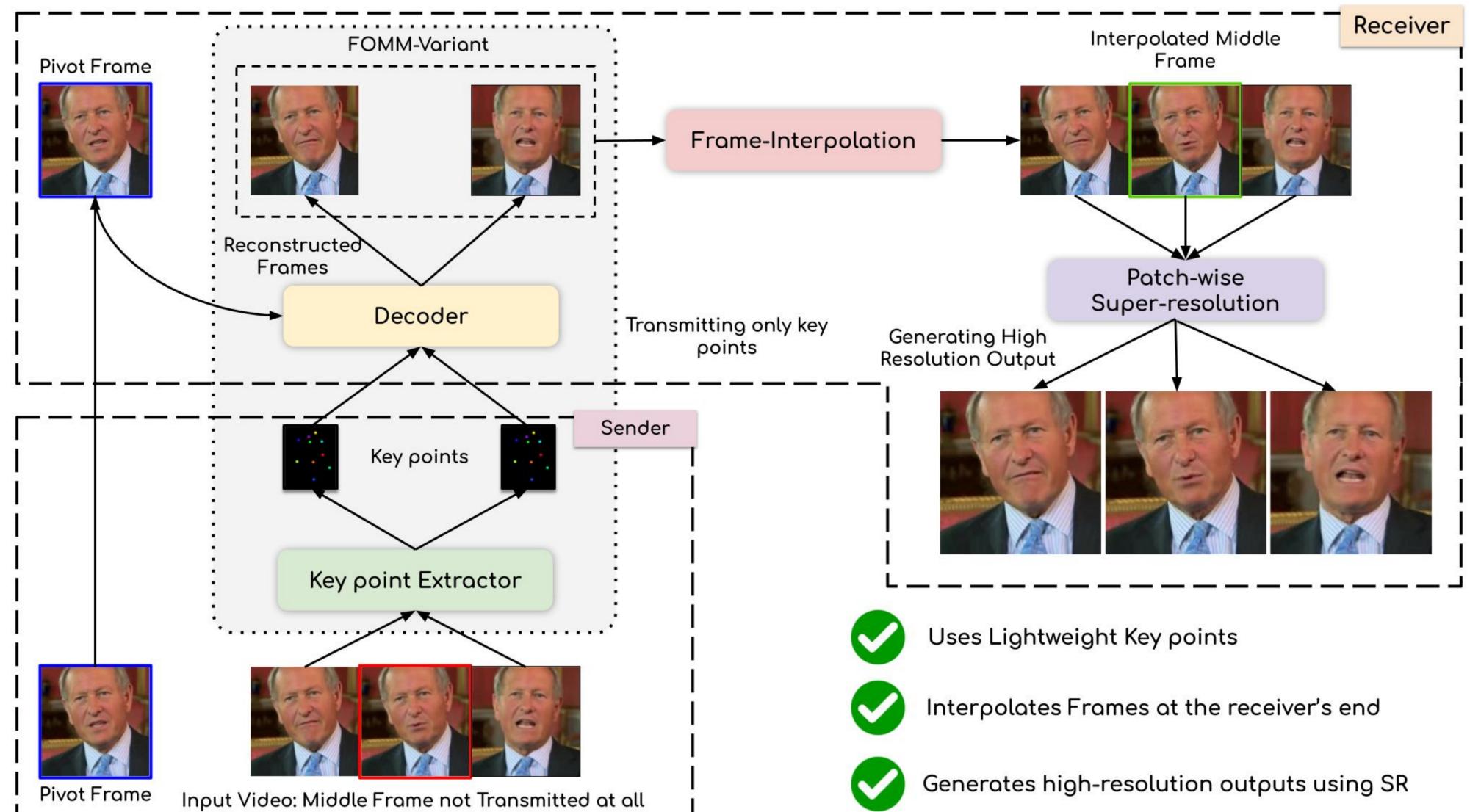


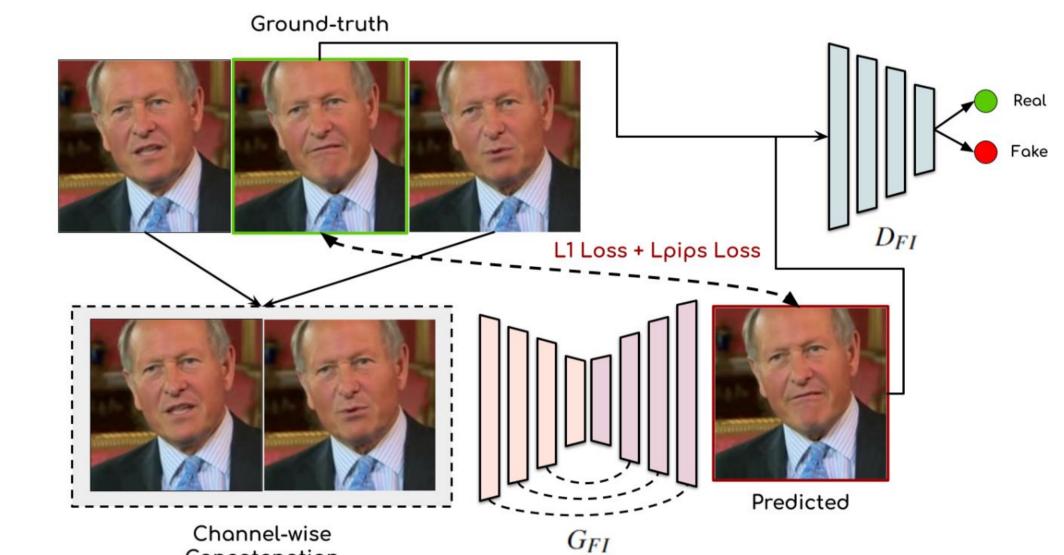
HYDERABAD

Madhav Agarwal, Anchit Gupta , Rudrabha Mukhopadhyay, Vinay Namboodiri, C.V. Jawahar

Overall Pipeline

Frame Interpolation





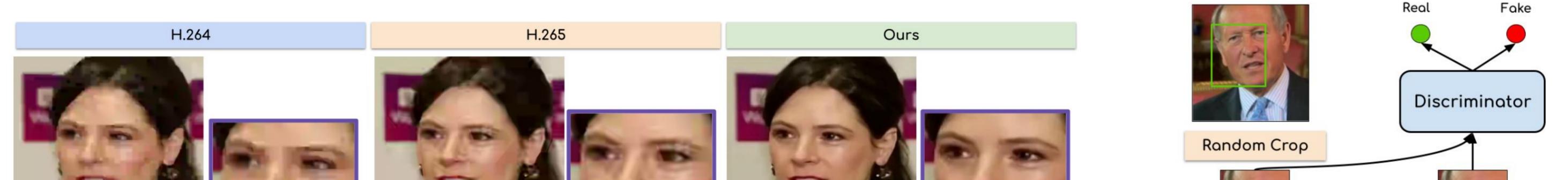
Concatenation

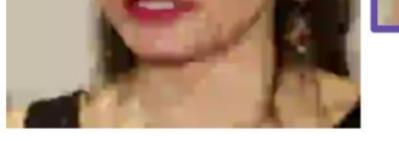
Frame-Interpolater Network

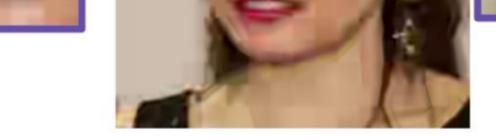
#Int. frames	BPP↓	PSNR ↑	SSIM [↑]	FID↓	
0	0.0024	29.28	0.83	9.10	
1	0.0012	28.49	0.82	12.42	
2	0.0008	28.23	0.81	12.77	
3	0.0006	27.73	0.79	12.92	

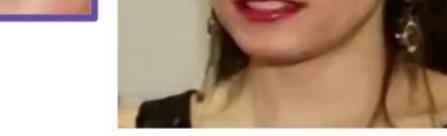
Comparison with H.264 and H.265

Super-Resolution



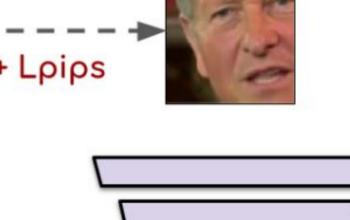


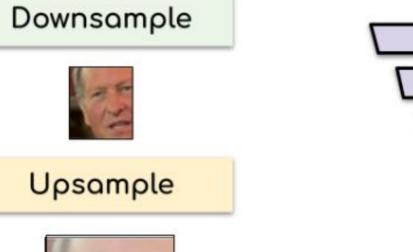


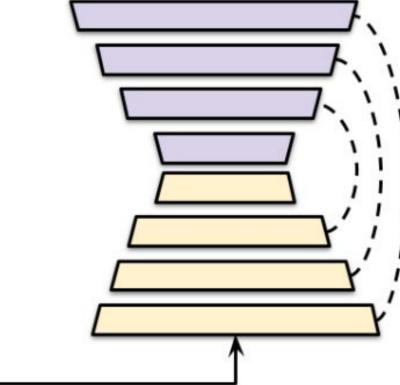






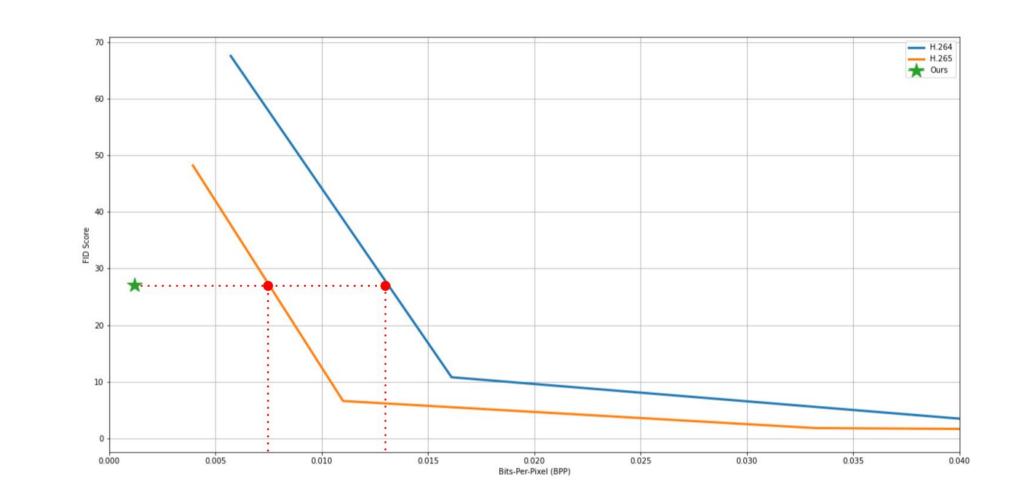






Patch-wise Super-resolution Network

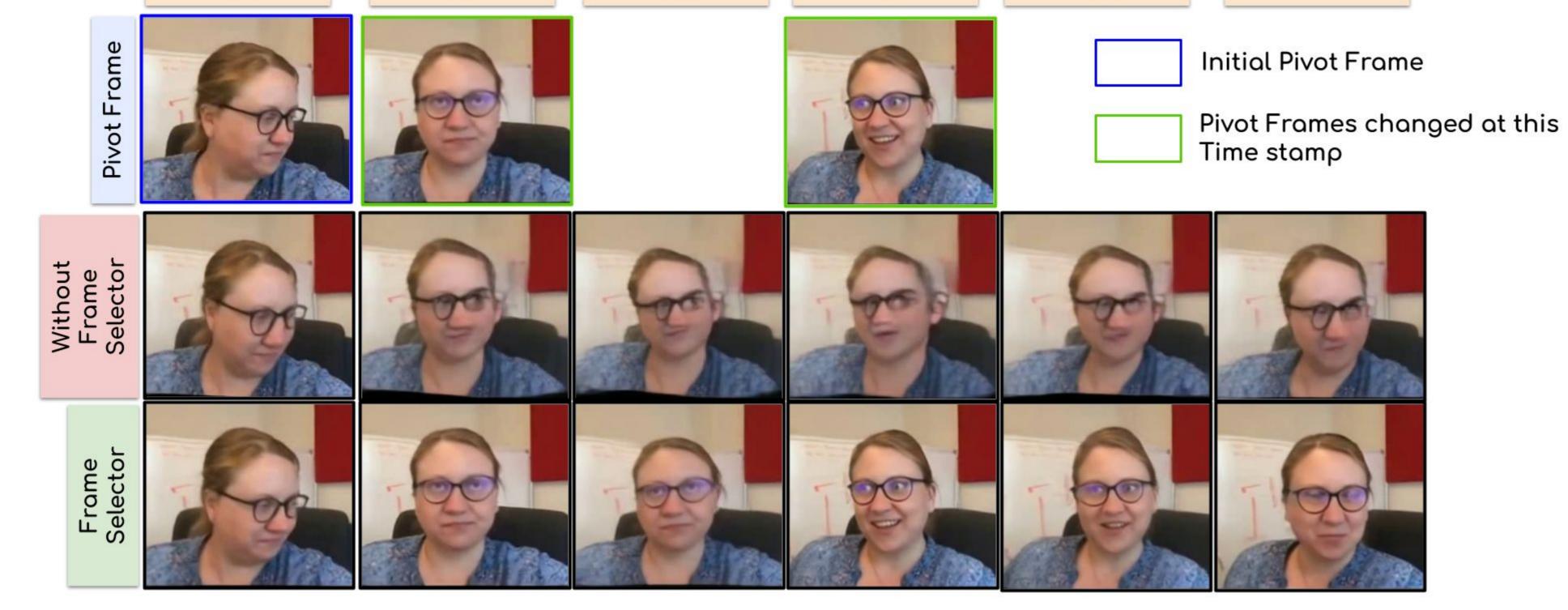
Patch Size	PSNR ↑	SSIM ↑	FID↓	
128×128	27.37	0.81	19.12	
64×64	27.47	0.80	20.16	
32×32	26.18	0.79	20.89	
16×16	25.34	0.78	21.17	



Method	BPP ↓	PSNR ↑	SSIM [↑]	FID↓
FOMM [20]	0.029	26.81	0.79	26.81
Face-vid2vid [25]	0.016	24.37	0.80	25.19
H.264 [26]	0.0057	30.25	0.78	67.54
H.265 [1]	0.0039	30.74	0.80	48.20
Key point Only (Ours)	0.0097	24.48	0.78	20.58
Key point + Frame Interpolation (Ours)	0.0048	24.21	0.78	23.03
Key point + Frame Interpolation + SR (Ours)	0.0012	26.73	0.81	27.81

T15 T50 Τ0 T10 T30

T70





Contact: madhav.agarwal@research.iiit.ac.in