ICLR 2024

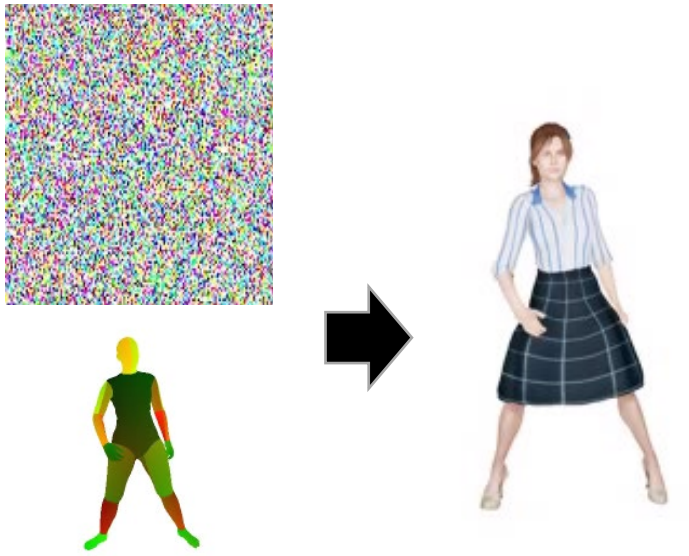
Korea Institute of Science and Technology

Adobe

Bidirectional Temporal Diffusion Model for Temporally Consistent Human Animation

Tserendorj Adiya[1,3] Jae Shin Yoon[2] Jungeun Lee[1] Sanghun Kim[1] Hwasup Lim[1]

[1]Korea Institute of Science and Technology [2]Adobe [3]CJ AI Center, CJ Corporation



Noise-to-Animation

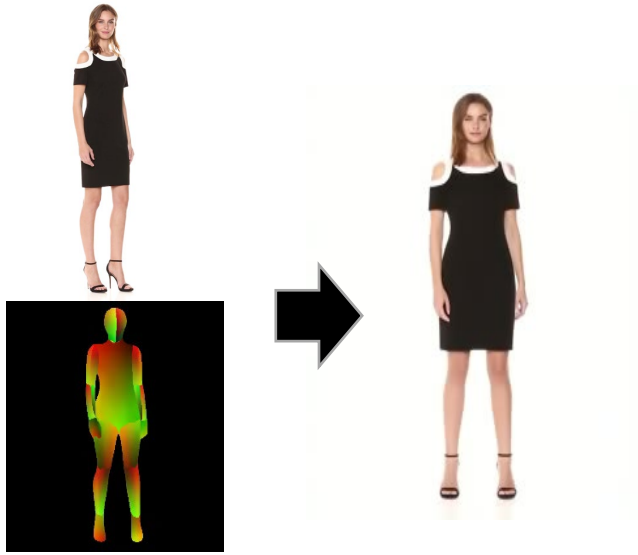
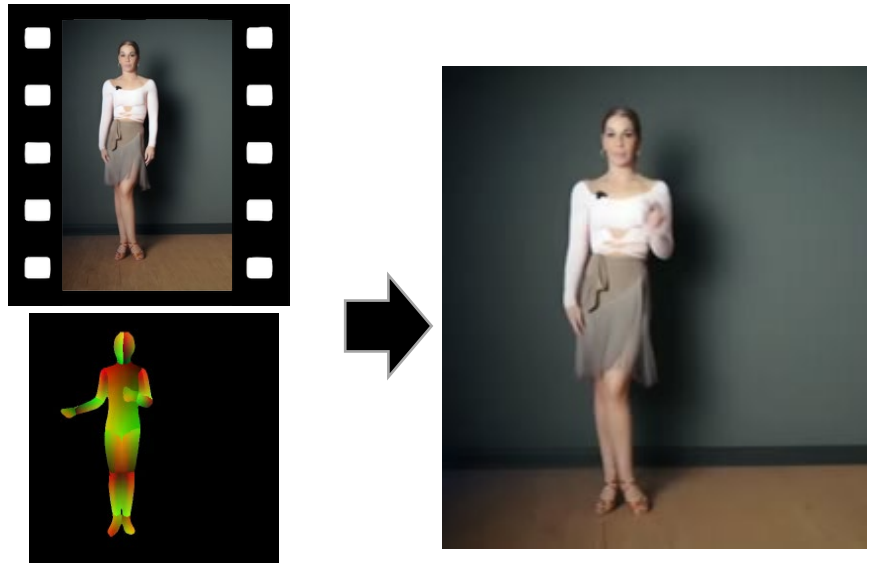


Image-to-Animation

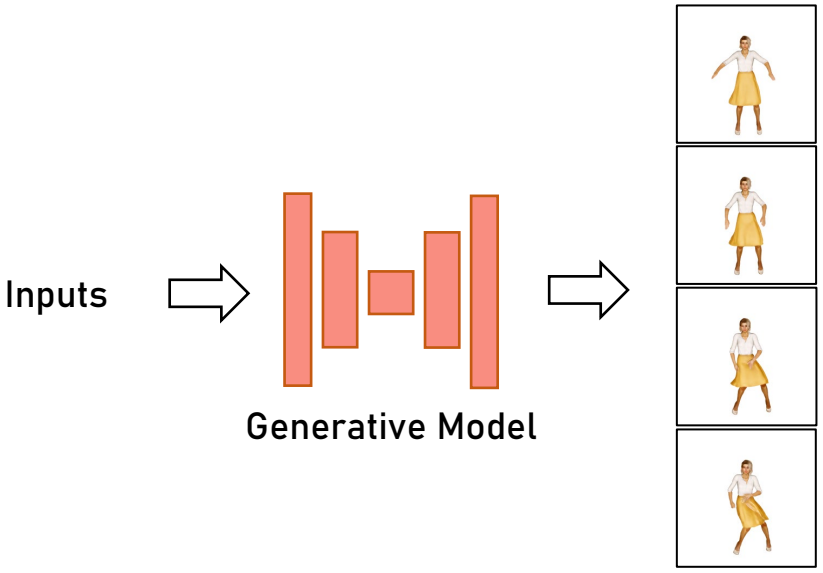


Video-to-Animation

*All the results in this video show the testing scenes

Motivation

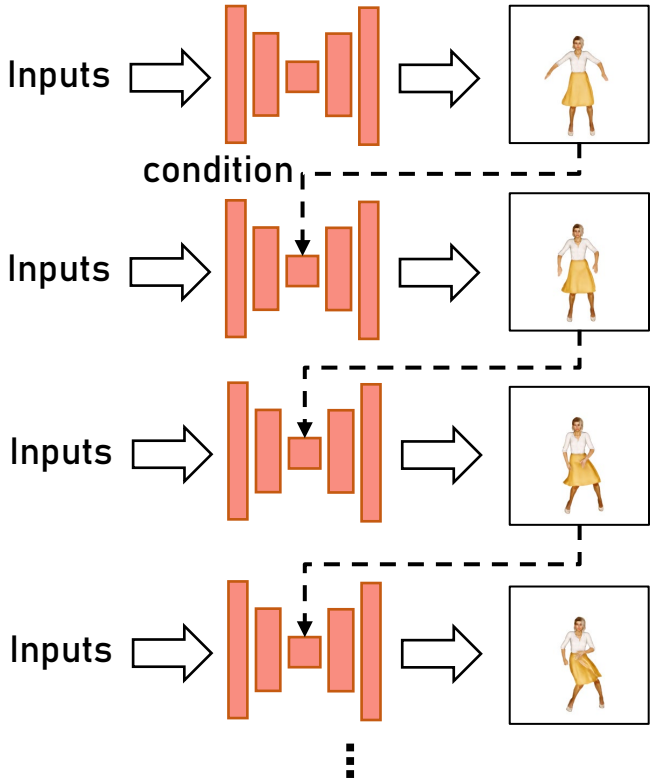
Parallel approach



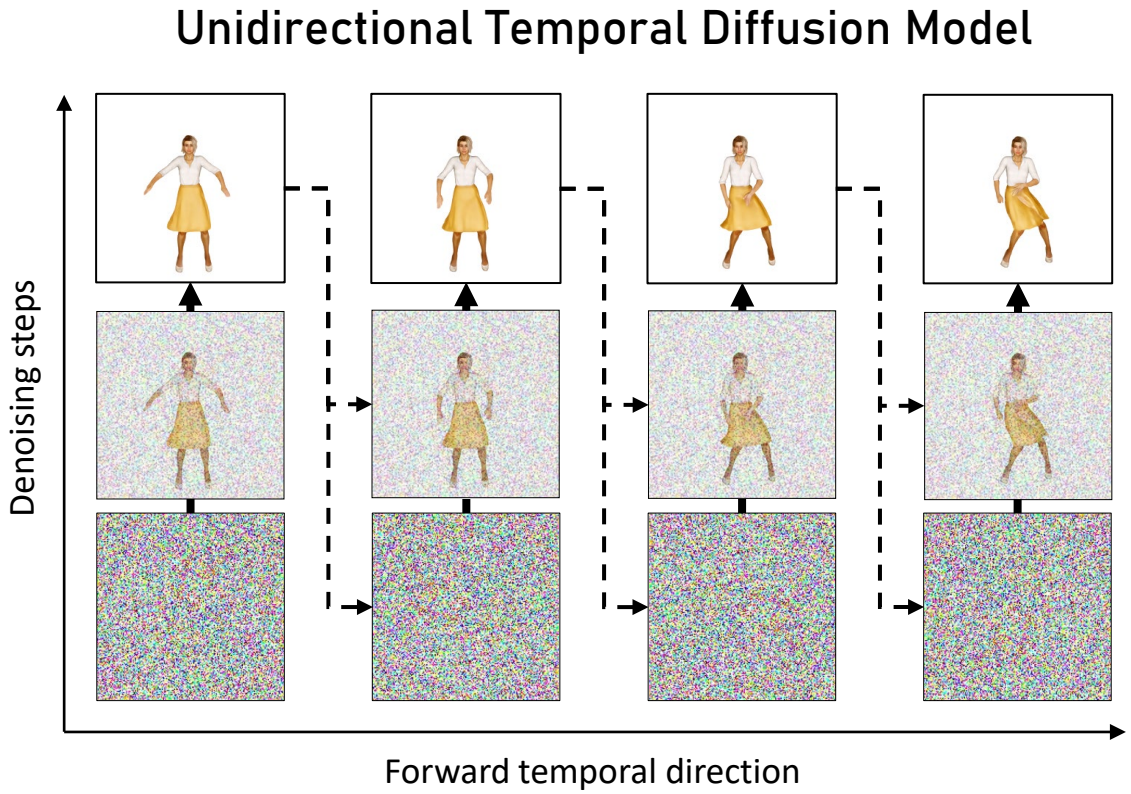
❌ Memory Efficient

❌ Generating Long Video

Auto-regressive approach

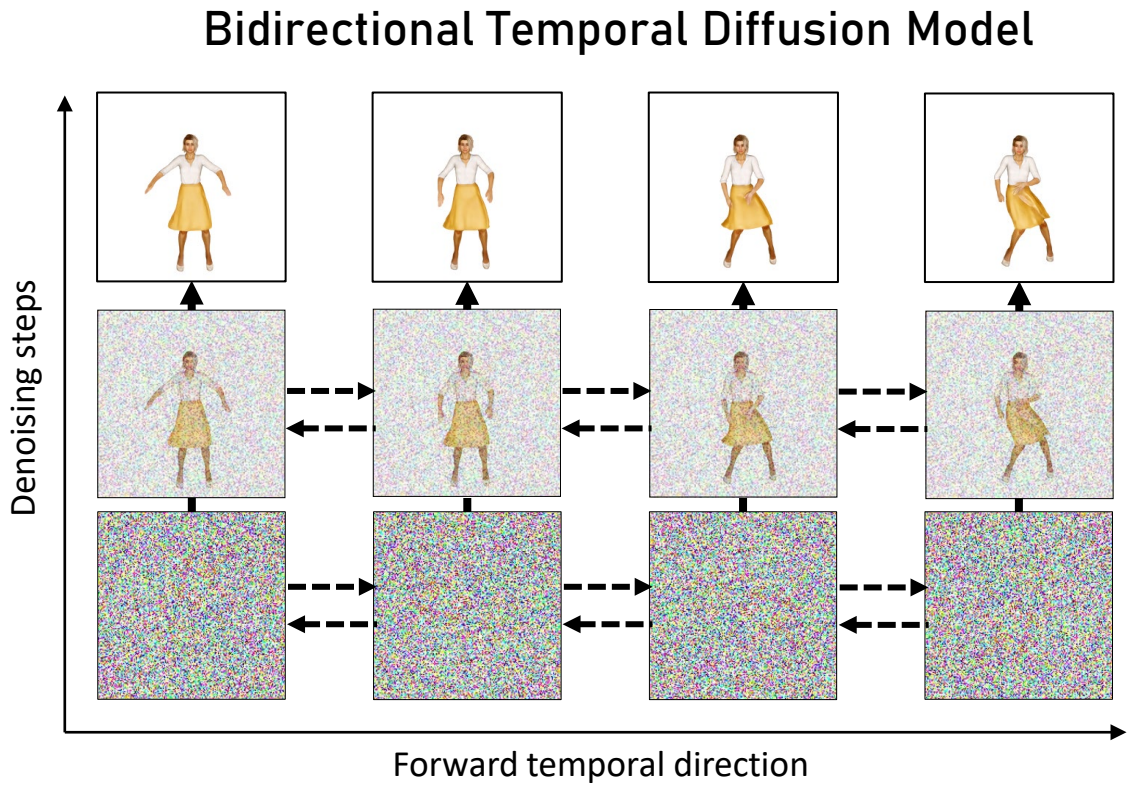


Motivation



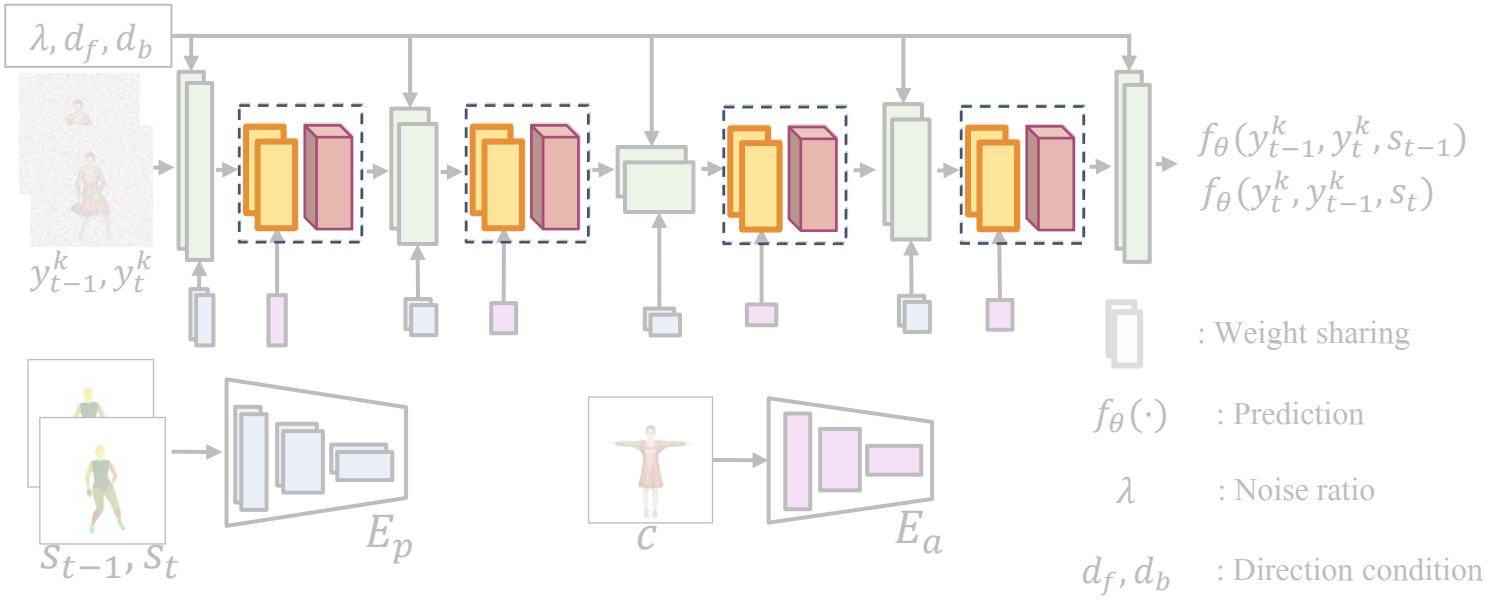
Errors increase gradually as they occur.

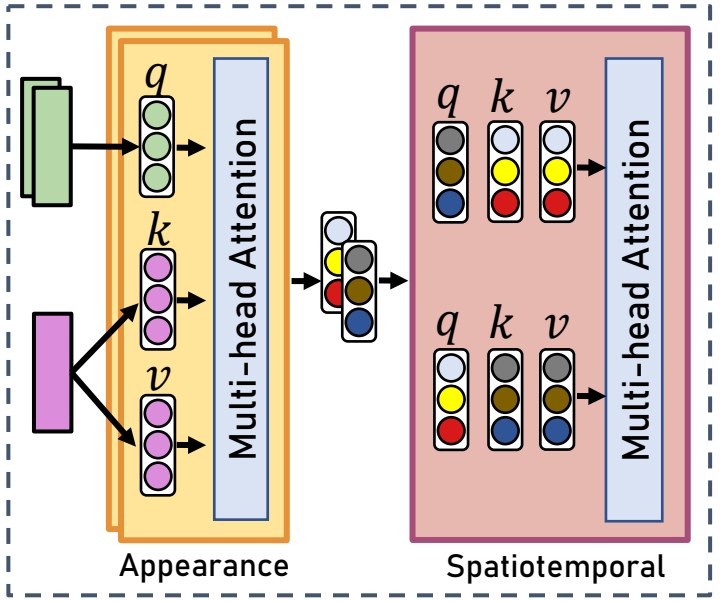
Proposed Method



No artifacts

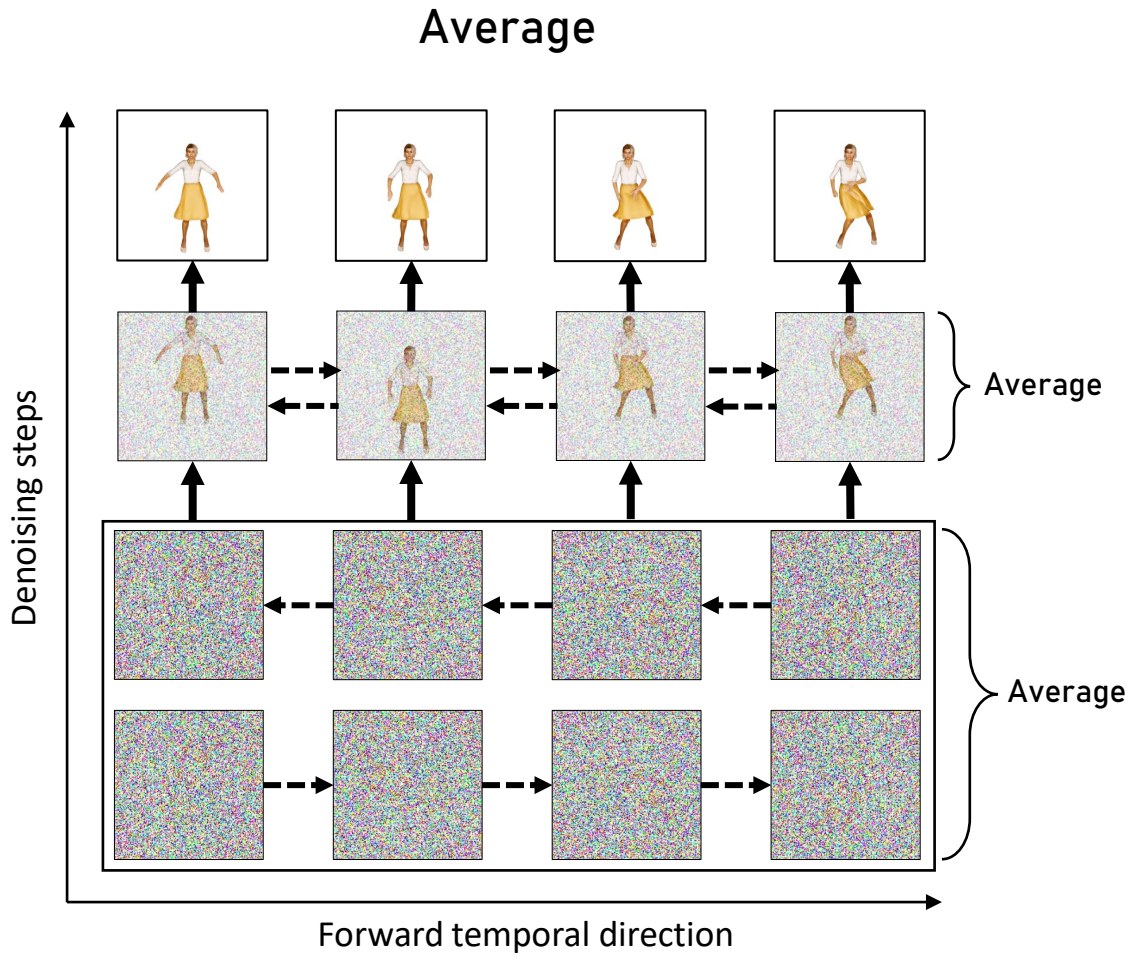
Overview: Bidirectional Temporal U-Net (BTU-Net)





Bidirectional Attention Block

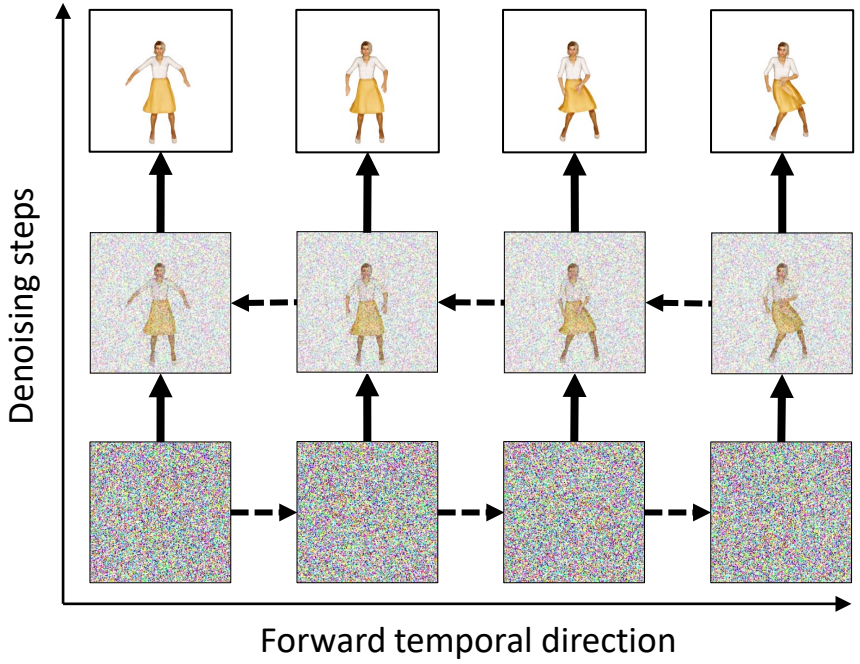
Bidirectional Recursive Sampling



The ghost effect occurs

Bidirectional Recursive Sampling

Alternative



Experiment results

Quantitative results of single image animation (image-to-animation).

Methods	SSIM↑	LPIPS↓	tLPIPS↓	tOF	FID↓
MRAA Siarohin et al. (2021)	0.894	0.140	0.011	13.58	67.68
TPS Zhao & Zhang (2022)	0.915	0.077	0.005	11.92	48.76
Ours	**0.958**	**0.036**	**0.003**	**8.93**	**11.14**

Quantitative results of person specific animation (video-to-animation).

Methods	Data 1 (6K)	Data 2 (10K)	Data 3 (10K)	Data 4 (15K)	Data 5 (15K)	Average
V2V	1.84/2.95/9.69	3.03/3.83/9.60	11.51/3.80/9.05	3.06/2.98/9.40	4.01/4.04/9.49	4.69/3.52/9.45
EDN	2.74/3.86/9.57	3.98/5.40/9.46	13.12/4.52/8.96	4.90/5.09/9.22	5.00/4.82/9.34	5.95/4.74/9.31
HFMT	3.68/4.41/9.48	6.39/8.54/9.26	13.27/4.62/8.91	6.08/3.22/9.10	6.86/4.53/9.21	7.26/5.06/9.19
DIW	1.83/2.88/9.68	2.70/4.11/9.61	11.89/4.09/9.03	2.83/4.66/9.45	4.14/5.20/9.48	4.68/4.19/9.45
MDMT	**1.76/2.58/9.73**	**2.68**/3.77/9.65	10.48/3.12/9.11	2.81/2.86/9.45	**3.81/4.12**/9.50	**4.31**/3.29/9.49
Ours	1.90/2.83/9.72	2.91/**3.76/9.65**	**10.32/2.52/9.28**	**2.78/2.85/9.57**	4.07/4.26/**9.51**	4.39/**3.22/9.55**

Comparison: Single Image Animation (Image-to-Animation)










Source Image | Driving Motion | MRAA | TPSMM | Ours

Comparison: Single Image Animation (Image-to-Animation)

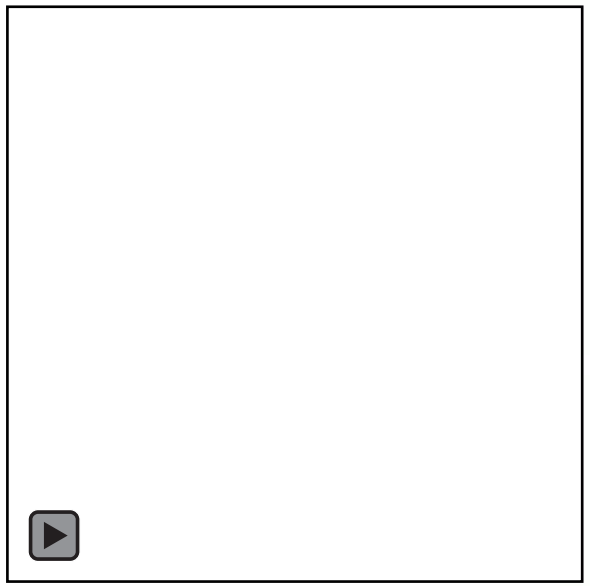
Source Image

MRAA

TPSMM

Ours

Comparison: Person Specific Animation (Video-to-animation)



Ground Truth

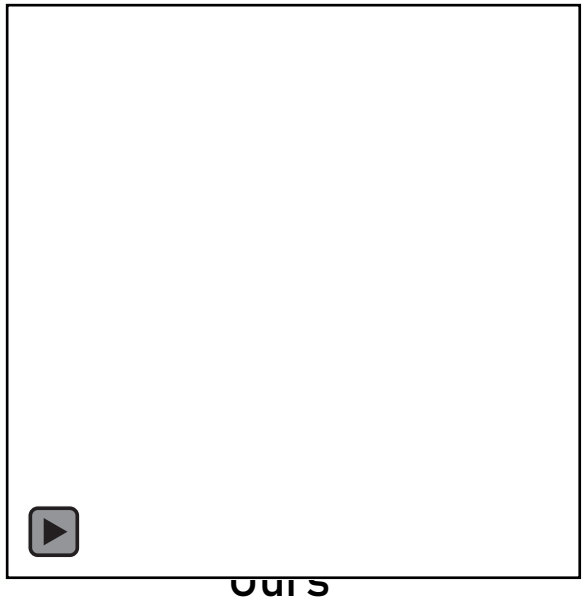
V2V

EDN

HFMT

DIW

MDMT



Our's

Comparison: Person Specific Animation (Video-to-animation)

V2V

EDN

HFMT

Ground Truth

DIW

MDMT

Ours

Ablation Study: Bidirectional vs Unidirectional

Source Image

Unidirectional

Bidirectional

Quantitative results of bidirectional vs unidirectional approaches.

Methods	SSIM↑	LPIPS↓	tLPIPS↓
Ours w/ Unidirectional Model	0.937	0.052	0.005
Ours	**0.958**	**0.036**	**0.003**

Unconditional Generation (Noise-to-Animation)








Thanks for your attention