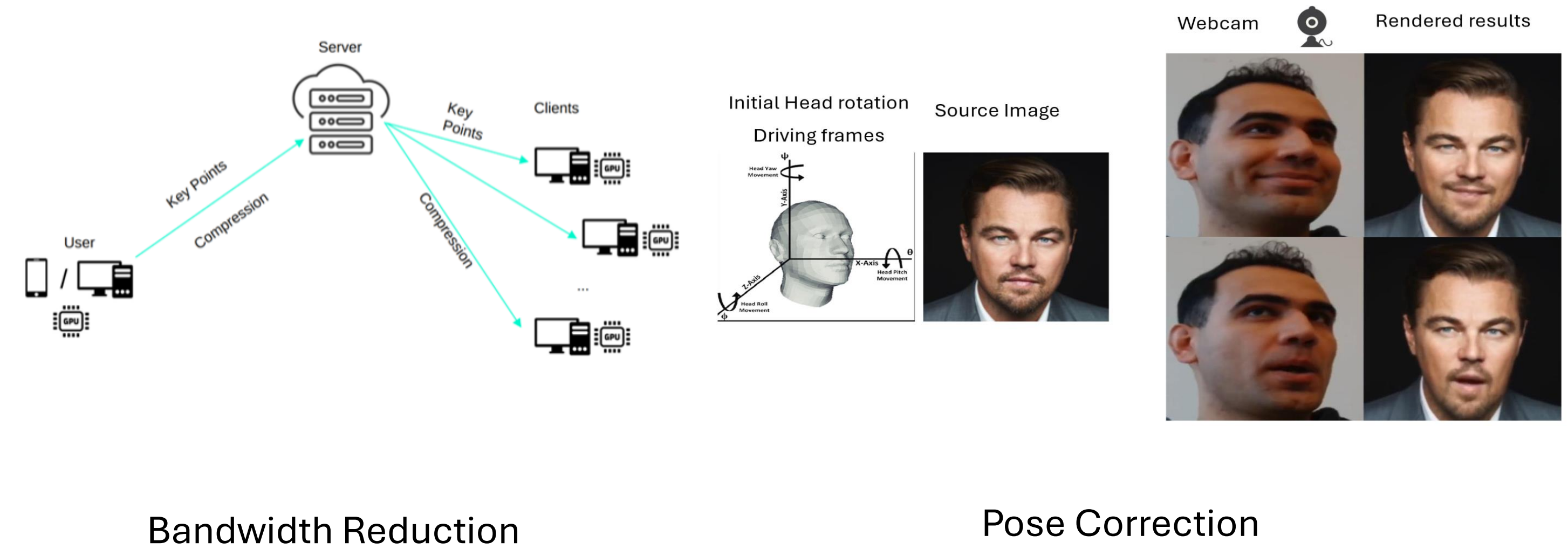


## Introduction

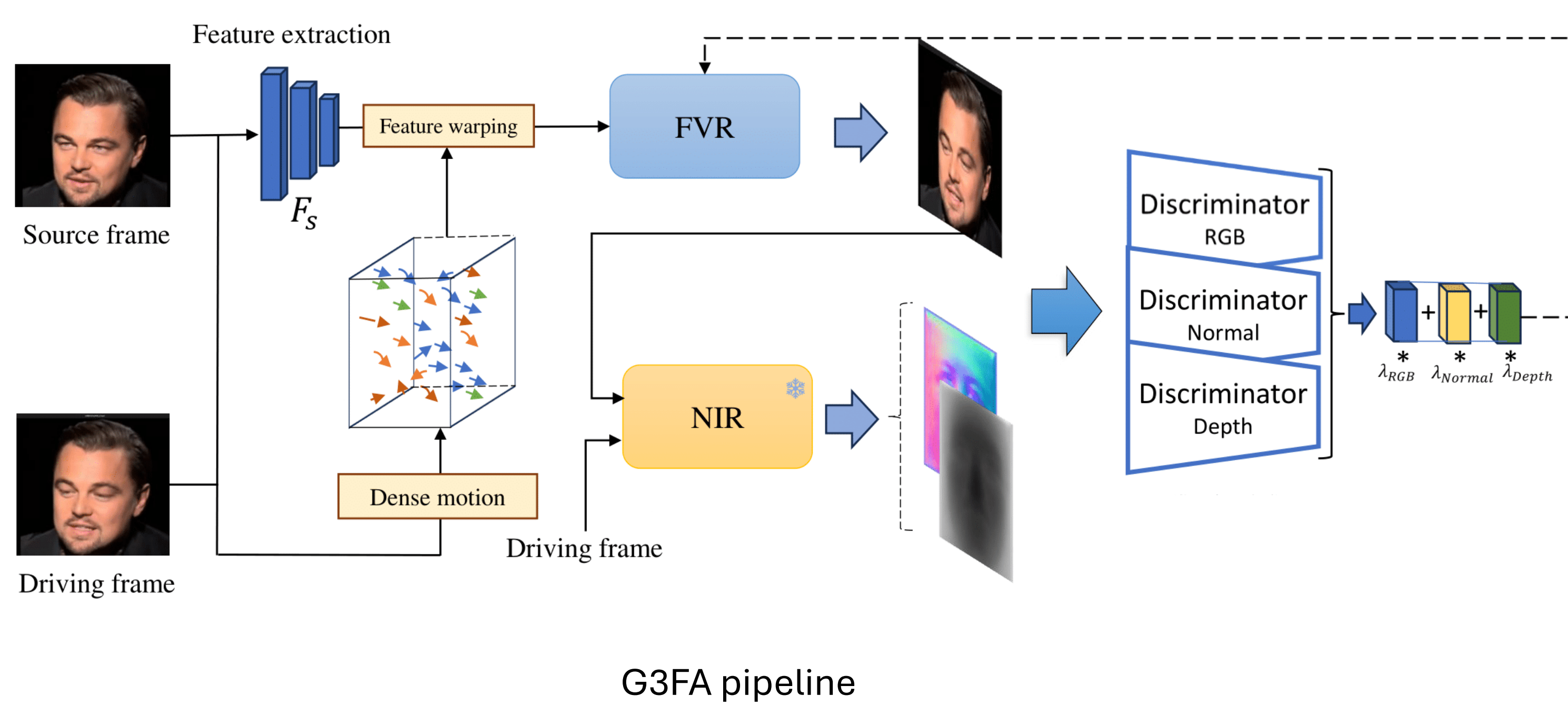
Animating human face images aims to synthesize a desired source identity in a natural-looking way mimicking a driving video's facial movements.

- **Bandwidth reduction:** Achieved by transmitting keypoints instead of full video frames in video conferencing.
- **Alternative appearance representation:** Allows displaying a different appearance of a participant.
- **Head pose correction:** Ensures proper eye contact and improves communication by adjusting head poses during video calls



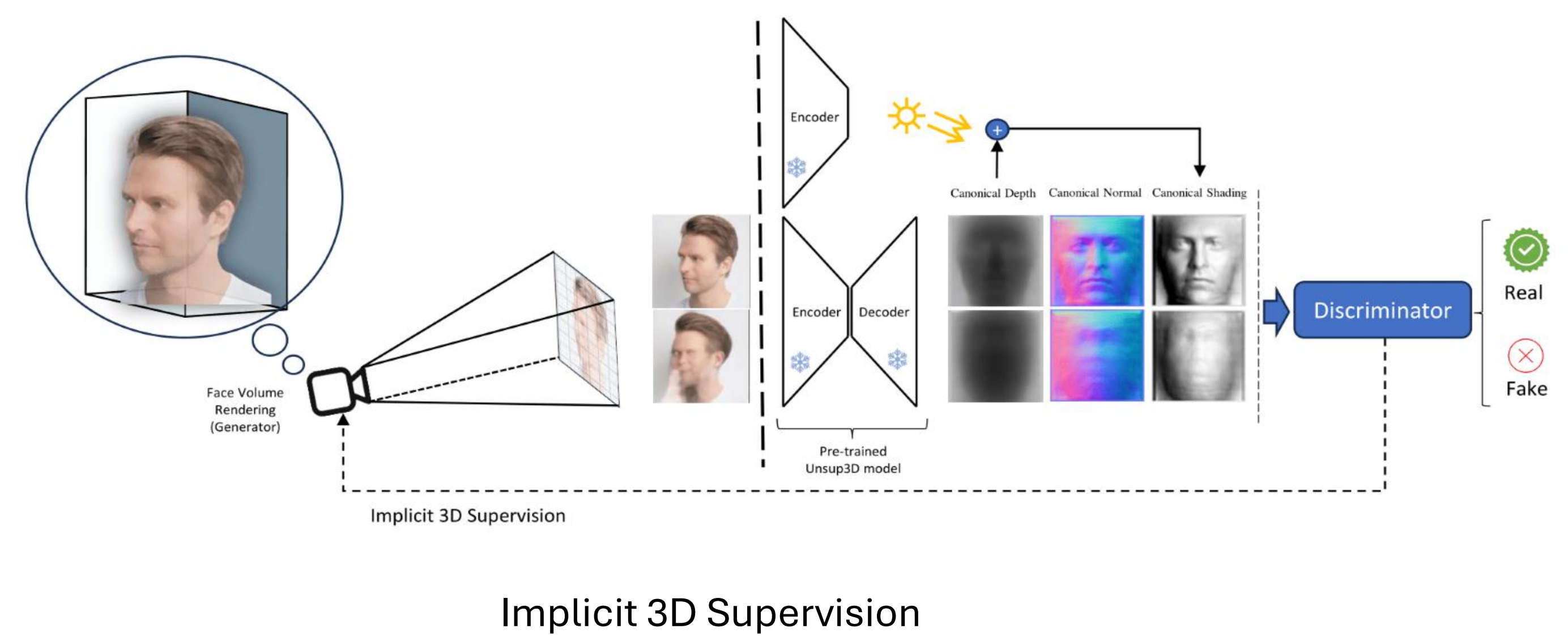
## Method

Generative Adversarial Networks (GANs) [1] excel in real-time face reenactment from a single image but often lack geometric consistency. In this paper, we introduce Geometry-guided GAN for Face Animation (G3FA) to overcome this challenge. Our method incorporates 3D information from 2D images, enhancing the image generation capabilities of talking head synthesis models. By using Unsup3D [2] as an inverse rendering model to extract 3D facial geometry and refining the generator with a weighted ensemble of discriminators, we improve visual consistency. Additionally, we employ 2D motion warping, orthogonal ray sampling, and volume rendering to produce high-quality, realistic outputs.



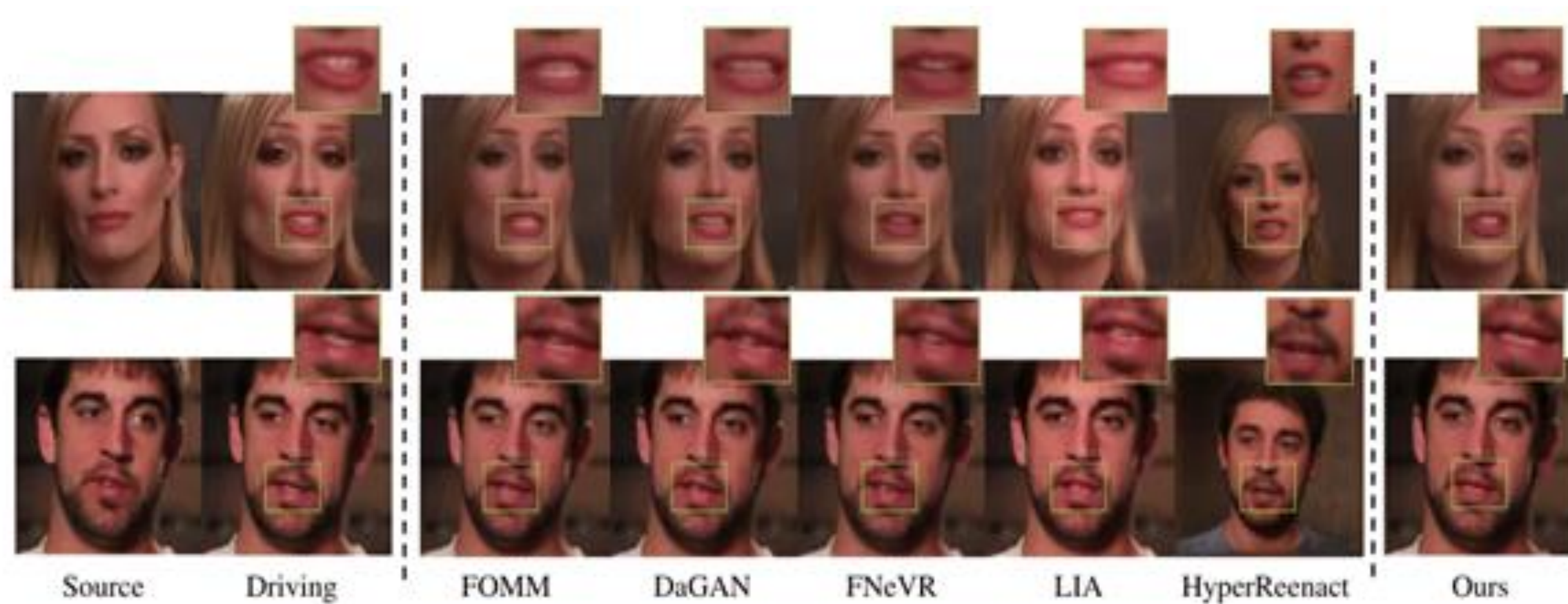
Our Contributions:

- **Implicit 3D supervision:** Leverages prior human head information to improve geometry consistency in face reenactment models without affecting inference time.
- **3D integration into GAN framework:** Introduces 3D properties into GAN-based face animation models, ensuring better geometry while keeping training computationally efficient.



## Evaluation

### Same identity Re-enactment



Method	L1↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	AKD↓
FOMM	12.31	0.109	23.52	0.71	24.59	1.89
Face vid2vid	11.13	0.125	24.15	0.84	21.78	1.72
DaGAN	11.22	0.117	<b>25.64</b>	0.88	22.83	0.91
FNeVR	11.16	0.094	24.18	0.77	21.11	0.95
LIA	12.02	0.106	25.57	0.84	<b>16.47</b>	1.12
HyperReenact	13.42	0.111	22.51	0.69	28.87	1.28
G3FA(Ours)	<b>10.87</b>	<b>0.081</b>	24.51	<b>0.91</b>	18.79	<b>0.80</b>

### Cross identity Re-enactment



Method	VoxCeleb2		TK	
	FID↓	CSIM↑	FID↓	CSIM↑
FOMM	142.18	0.5219	130.78	0.5402
Face vid2vid	139.74	0.5971	121.44	0.6114
DaGAN	130.77	0.6021	120.94	0.6264
FNeVR	132.36	0.5408	122.47	0.6021
LIA	<b>122.26</b>	0.6078	<b>118.59</b>	0.6338
HyperReenact	152.94	0.5144	147.63	0.4923
G3FA(Ours)	127.12	<b>0.6274</b>	122.83	<b>0.6455</b>

## References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NIPS), vol. 27, 2014.

[2] S. Wu, C. Rupprecht, and A. Vedaldi, "Unsupervised learning of probably symmetric deformable 3D objects from images in the wild," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-10, 2020.



Paper



Code