

FreeUV: Ground-Truth-Free Realistic Facial UV Texture Recovery via Cross-Assembly Inference Strategy

CVPR 2025

Xingchao Yang, Takafumi Taketomi, Yuki Endo, Yoshihiro Kanamori



Motivation

Recover high-quality 3D facial UV textures from single-view 2D images

Input



Recovered
UV texture



Rendered
3D face



Challenges

1. Reliance on costly ground-truth UV data or synthetic UV datasets
2. Robustness across diverse poses and in-the-wild conditions
3. Complex facial details: wrinkles, makeup, facial hair, occlusions

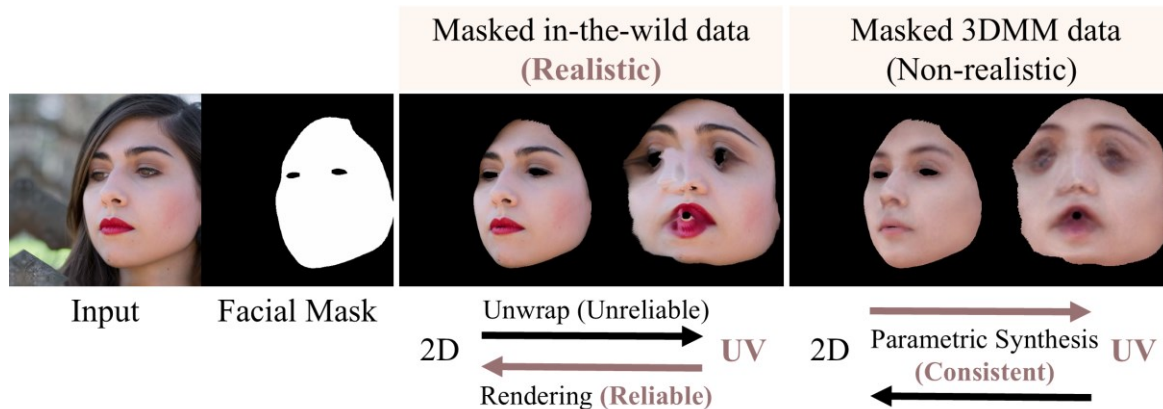
FreeUV: A scalable, data-efficient solution without ground-truth UV supervision

1. Eliminates need for annotated or synthetic UV data
2. Leverages pre-trained Stable Diffusion with Cross-Assembly Inference
3. Captures intricate details (e.g., wrinkles, makeup) robustly

Key idea

Cross-Assembly Inference Strategy

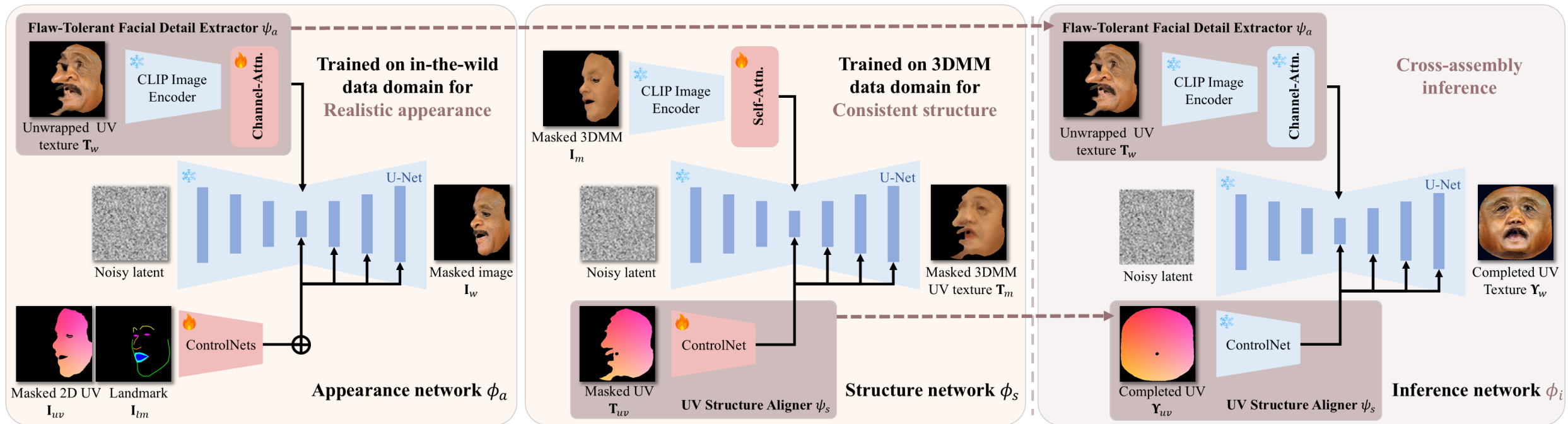
- Selectively combine **realistic appearance** (in-the-wild) and **structural consistency** (3DMM)
- **Training**: Two networks trained separately for Appearance and Structure
- **Inference**: Integrate UV-specific modules for **UV-to-UV** mapping



Mapping	Domain	Appearance	Structure	Ours
UV-to-2D	3DMM	Non-realistic	Consistent	✗
	In-the-wild	Realistic	Reliable	✓
2D-to-UV	3DMM	Non-realistic	Consistent	✓
	In-the-wild	Realistic	Unreliable	✗

Overview

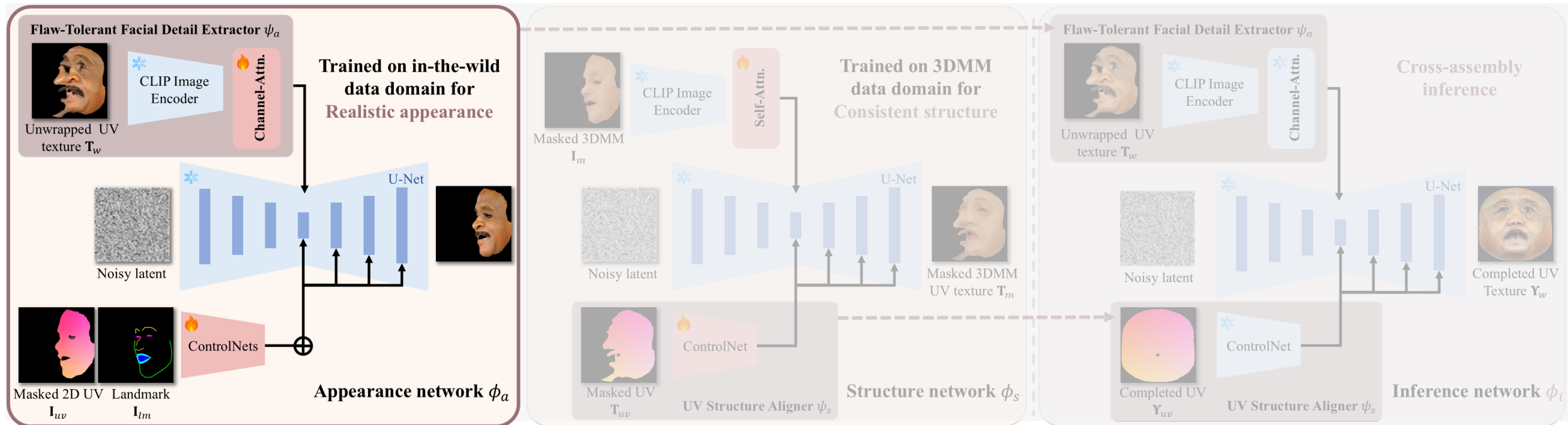
- Appearance network (**UV** to 2D) for feature extraction
- Structure Network (2D to **UV**) for structural reconstruction
- Inference Network (**UV** to **UV**) to generate textures align with the UV layout



Appearance network

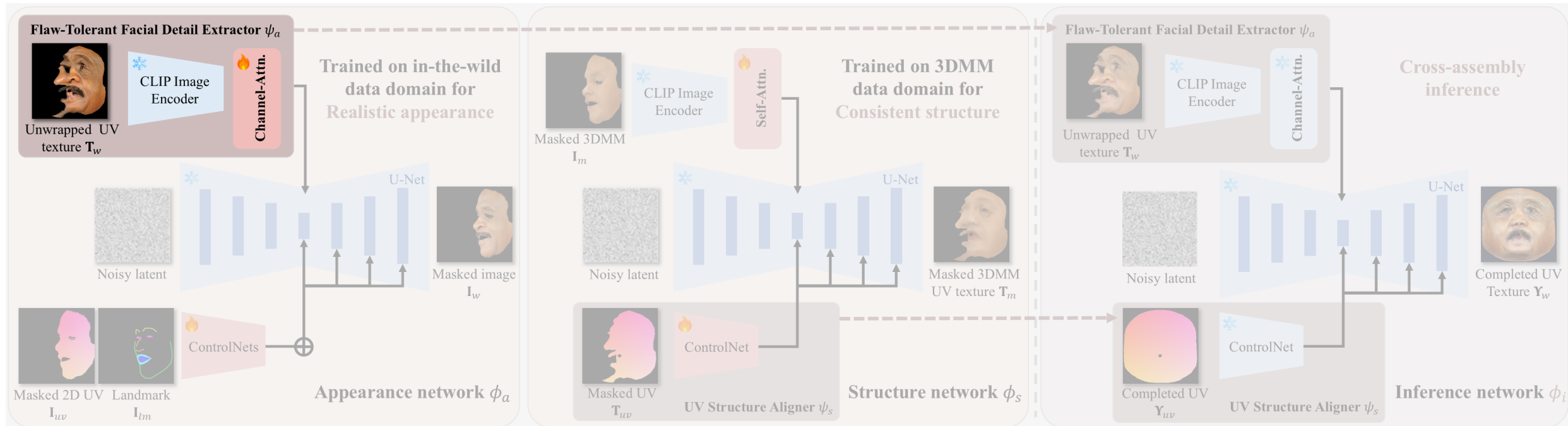
Training on in-the-wild realistic data domain

- **Process:** Maps masked unwrapped UV texture to 2D face image (**UV** to 2D)
- **Outcome:** High-fidelity 2D images (e.g., makeup, wrinkles, facial hair)



Flaw-Tolerant Facial Detail Extractor

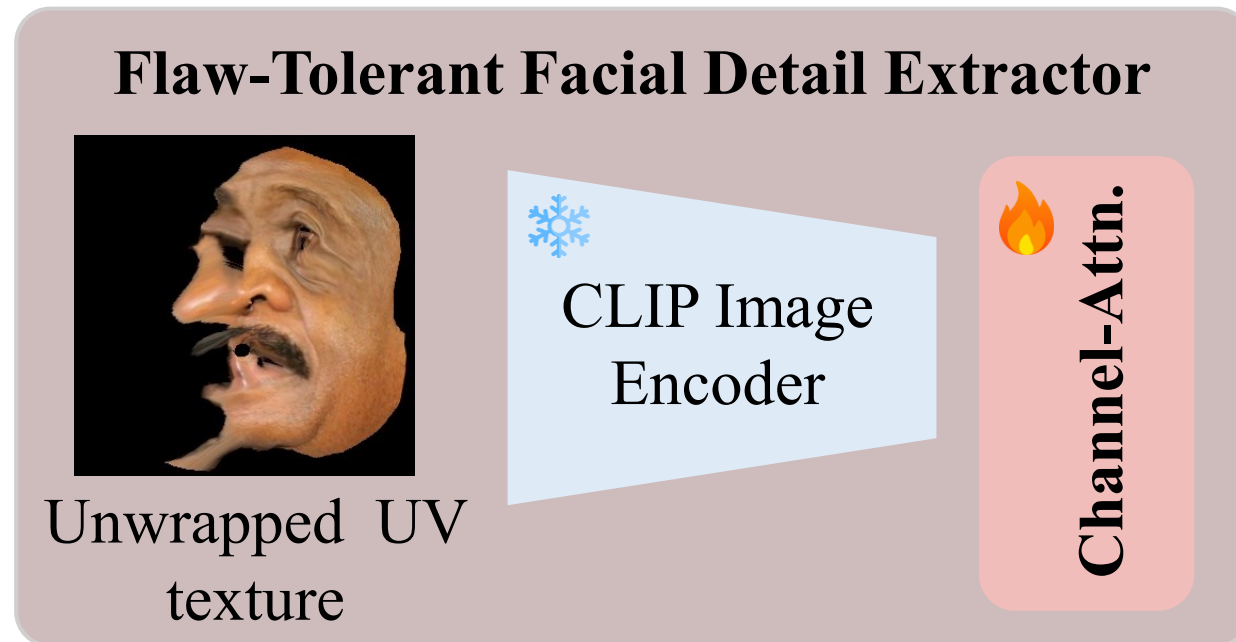
Captures realistic facial details despite UV unwrapping distortions



Flaw-Tolerant Facial Detail Extractor

Captures realistic facial details despite UV unwrapping distortions

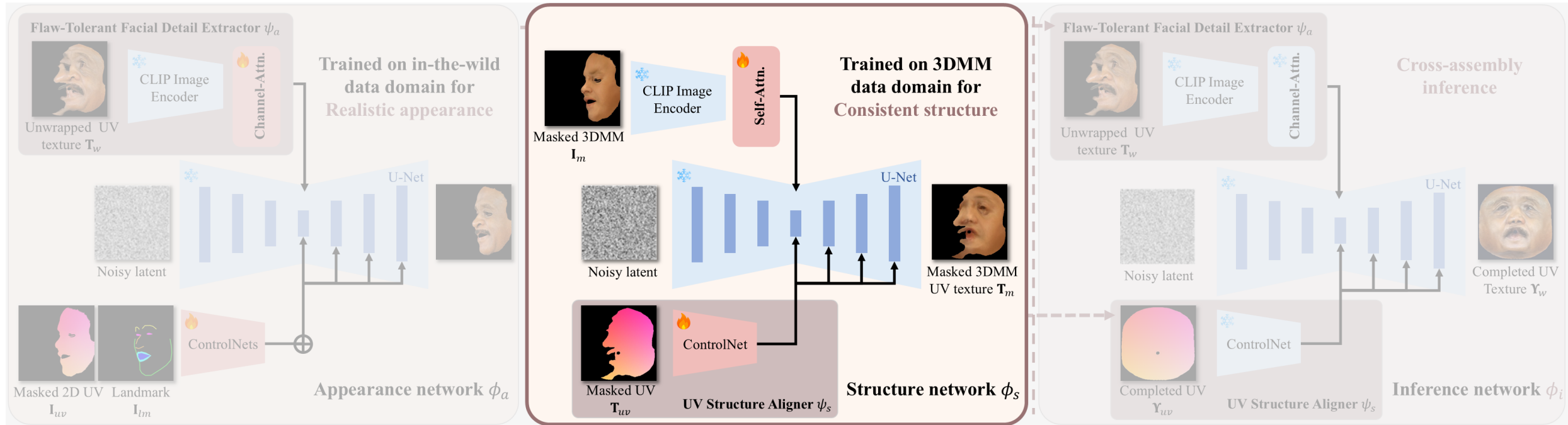
- Use **CLIP visual backbone layers** to capture realistic details from in-the-wild UV textures
- Use **Channel attention** to selectively extract relevant features and mitigate flaws in unwrapped UV textures



Structure network

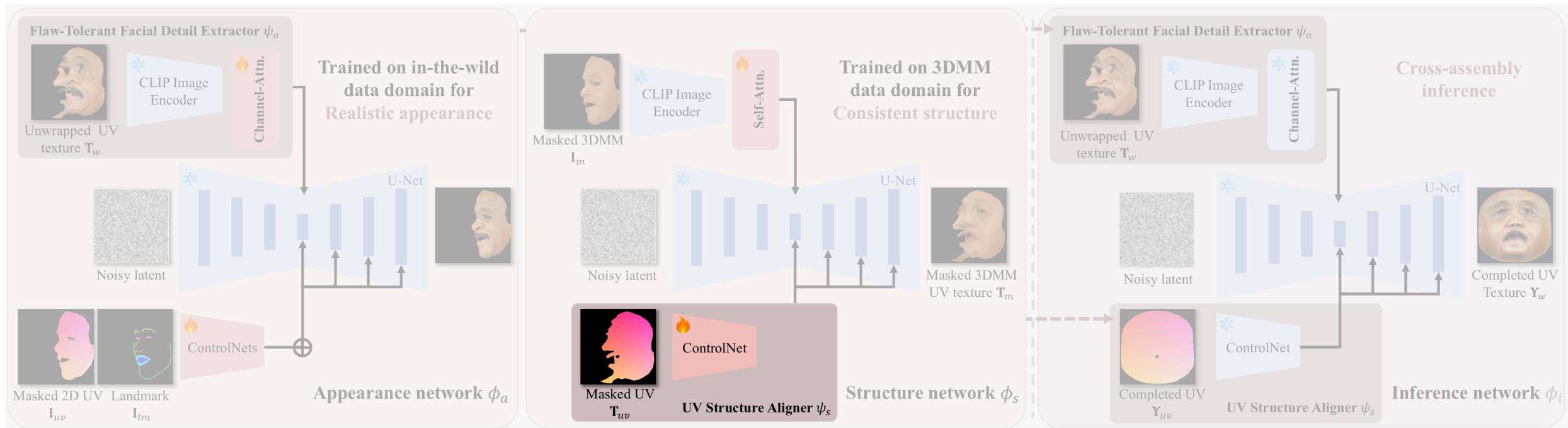
Training on 3DMM consistent data domain

- **Process:** Maps rendered 2D 3DMM face image to UV texture (2D to **UV**)
- **Outcome:** Structurally consistent UV textures



UV Structure Aligner

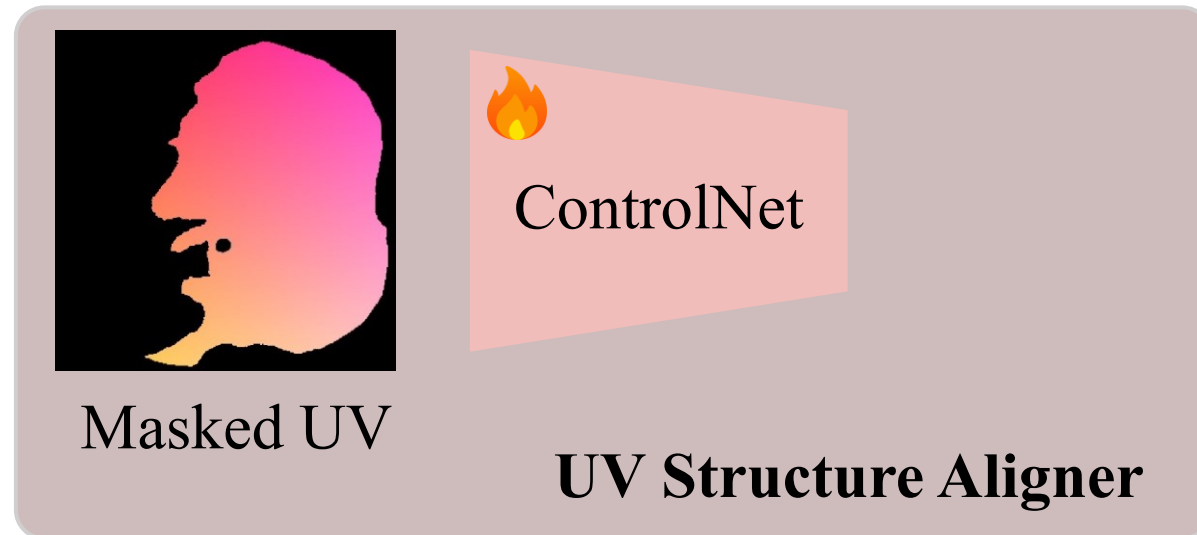
Ensures structural consistency aligned with 3DMM UV layout



UV Structure Aligner

Ensures structural consistency aligned with 3DMM UV layout

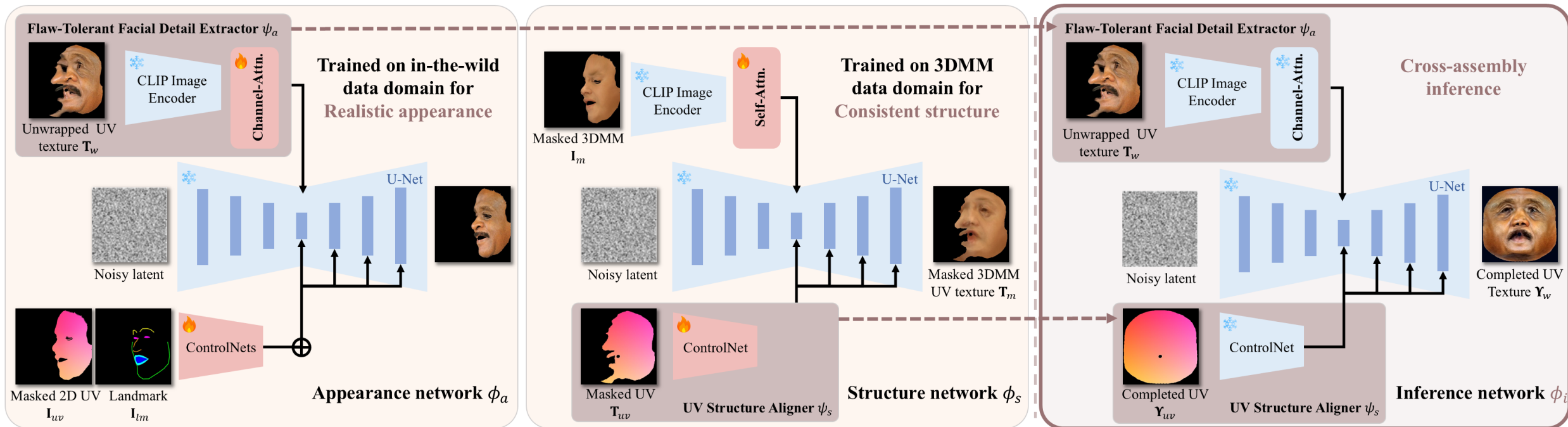
- A **masked UV map** guides the structure through ControlNet



Cross-Assembly Inference Strategy

Combines appearance and structure components for UV-to-UV mapping

- **Process:** Maps masked unwrapped UV texture to complete UV texture (**UV** to **UV**)
- **Outcome:** High-fidelity UV textures (e.g., makeup, wrinkles, facial hair)



Facial UV texture recovery

Robustly produces realistic textures despite challenging inputs

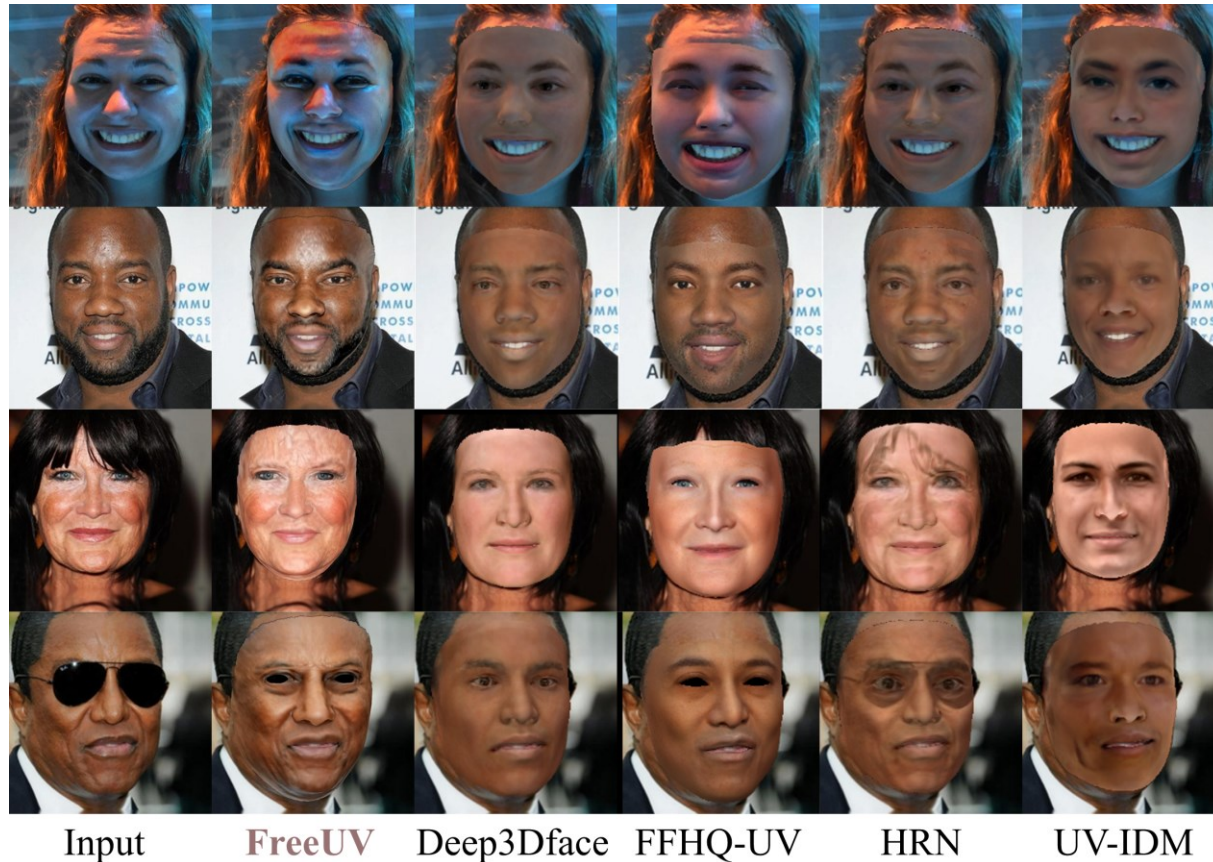
- Even with significant distortions, occlusions, and makeup



3D face reconstruction

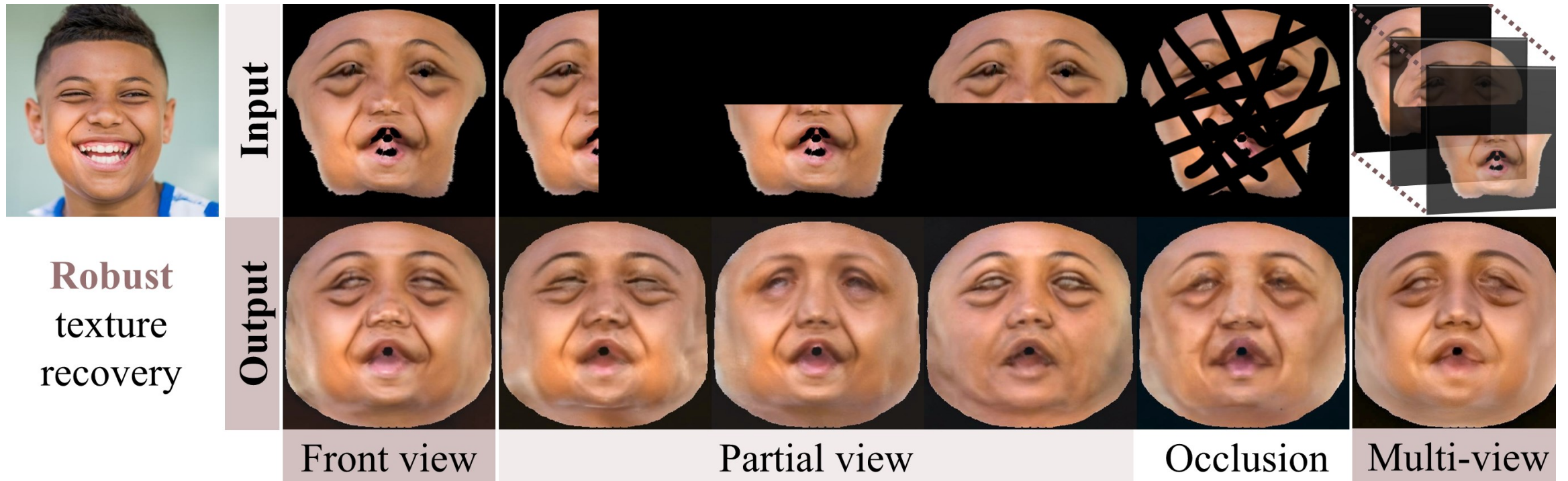
Achieves the closest match to the original input

- Even with extreme lighting, facial hair, and wrinkles



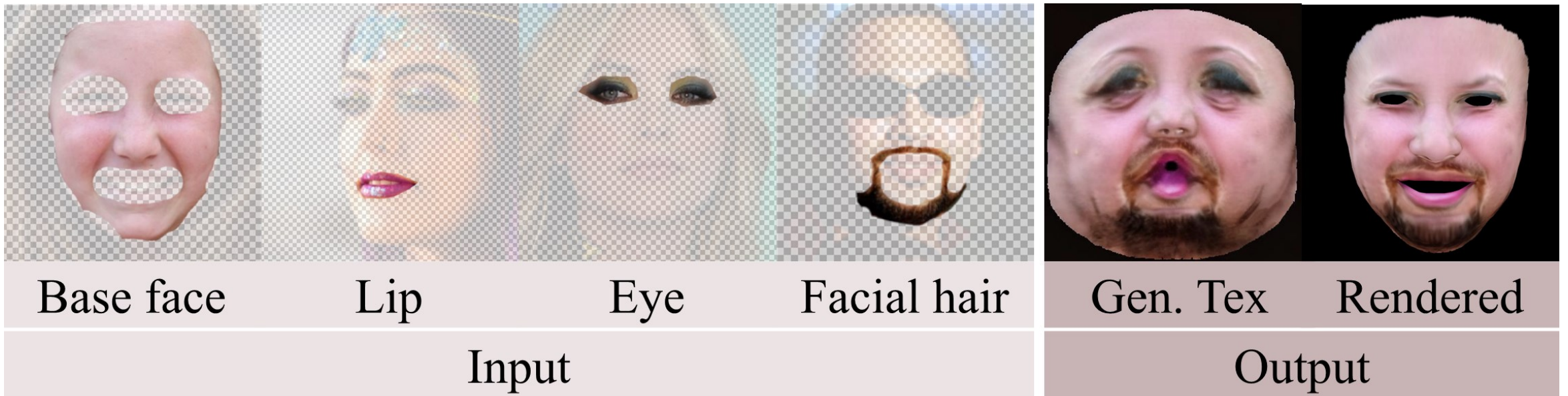
Robustness evaluation

Effectively handles partial views, seamlessly completing missing regions



Customized local editing

Enables transfer of specific facial features from different source, creating a coherent UV texture that combines multiple attributes



Quantitative Evaluation

FreeUV surpasses state-of-the-art methods

- Measured between original 2D face images and the recovered UV textures
- Capability to capture semantically meaningful and visually coherent features

Bold: best results; Underlined: second-best

Method	FFHQ			CelebAMask-HQ			LPFF		
	CLIP-I↑	DINO-I↑	FID↓	CLIP-I↑	DINO-I↑	FID↓	CLIP-I↑	DINO-I↑	FID↓
HRN [46]	<u>0.8327</u>	<u>0.7389</u>	166.19	<u>0.8259</u>	0.7382	189.74	0.7368	0.5951	142.82
UV-IDM [48]	0.7986	0.5836	228.74	0.7458	0.5690	258.34	0.7440	0.5345	239.10
FLAME-based [90]	0.8218	0.7269	<u>158.06</u>	0.8016	<u>0.7640</u>	<u>164.98</u>	<u>0.7822</u>	<u>0.6724</u>	166.31
FreeUV	0.8490	0.7559	142.39	0.8272	0.7948	153.43	0.7997	0.6835	<u>158.55</u>

Conclusion

- **FreeUV** provides a scalable, ground-truth-free solution for high-quality UV texture recovery
- Cross-Assembly Inference strategy integrates realistic appearance with structural consistency

Thank you!

