# *CraftGraffiti*: Exploring Human Identity with Custom Graffiti Art via Facial-Preserving Diffusion Models

*Ayan Banerjee*

CVC Centre de Visió per Computador

*Fernando Vilariño*

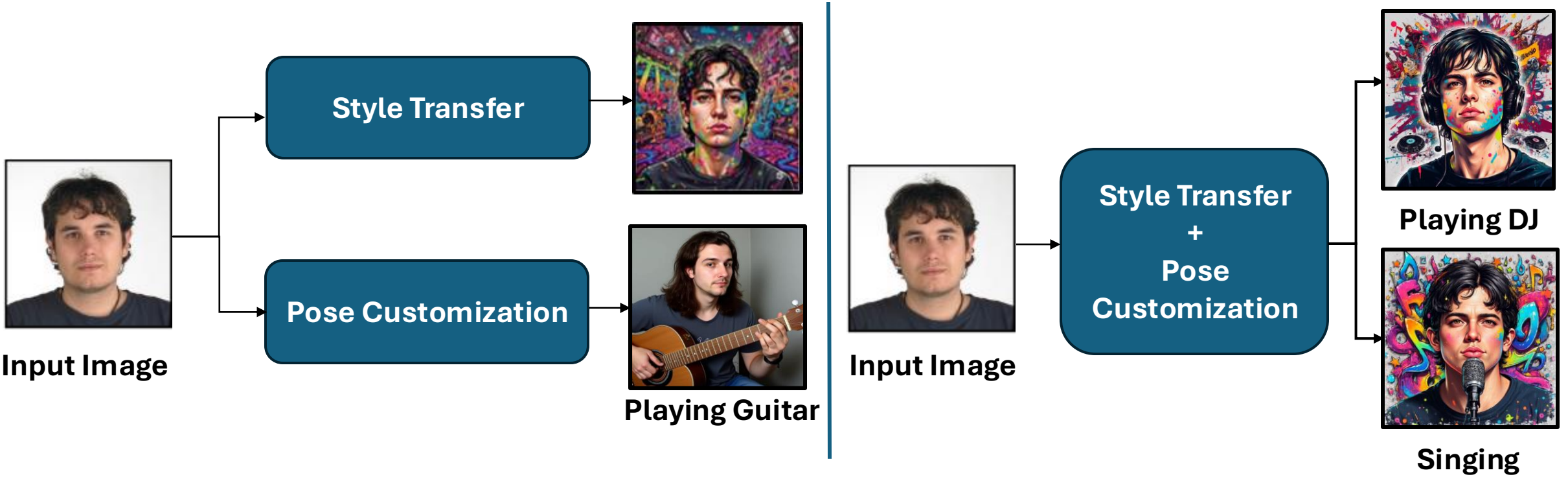
UAB Universitat Autònoma de Barcelona

*Josep Lladós*

NEURAL INFORMATION PROCESSING SYSTEMS 2025

# Problem Statement

## Style Transfer + Pose Customization




We define **style transfer with pose customization** as the problem of generating an image that preserves the **facial identity of a source image** adopts the **artistic or photometric style** and conforms to a **target pose representation.**

# Challenges of Style Transfer

## Over weightage on styles cause appearance drift

- Style transfer operates in feature-statistic space, not in identity space.
- It introduces appearance drift — modifications to color, lighting, or even local geometry.

# Solution

- Low-rank adaptation based style transfer.

# Why is this effective?

- Control the weight pruning in order to mitigate the appearance drift.

Input Image

Reference Style

Final Output

# Challenges of Pose Customization

## Affect face consistency

- For large pose shifts, the generator must hallucinate unseen regions (e.g., ear or cheek), which can break identity structure.
- Keypoint-based pose constraints ensure alignment but provide no supervision on facial texture leading to distorted or identity-inconsistent renderings.

# Solution

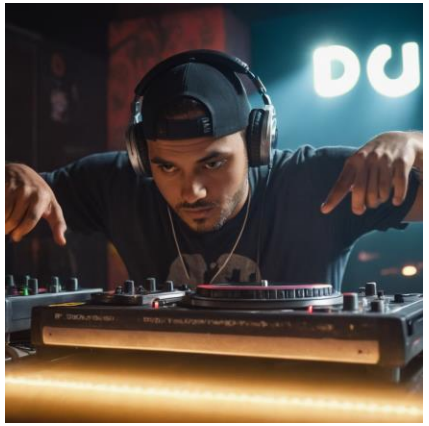
- CLIP- based pose customization (limits no. of pose)
- Face-consistent self-attention.

# Why is this effective?

- Preserve the facial attribute of the character through the extra dimension of identity embedding.

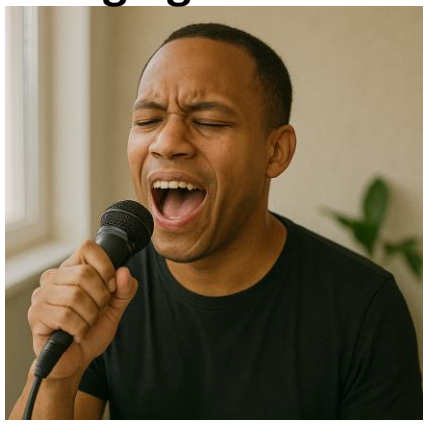
Input Image

Playing DJ



Playing Guitar

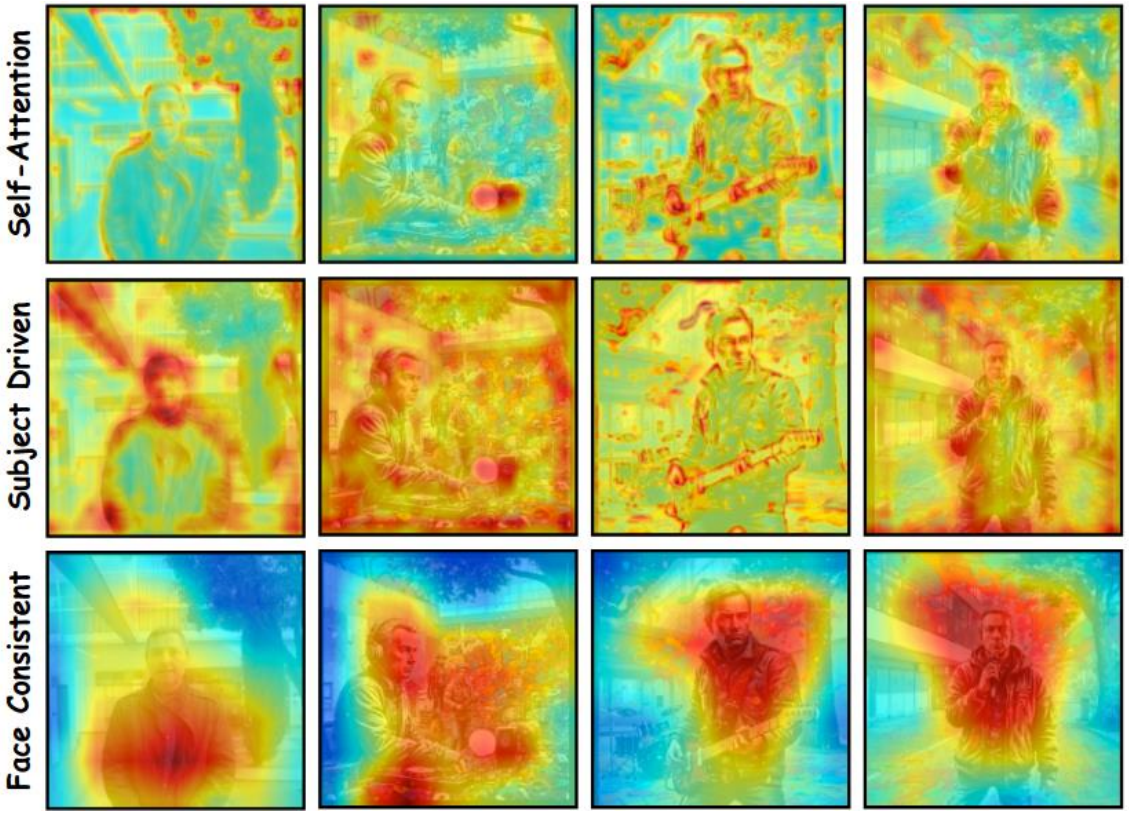
Singing in a mic

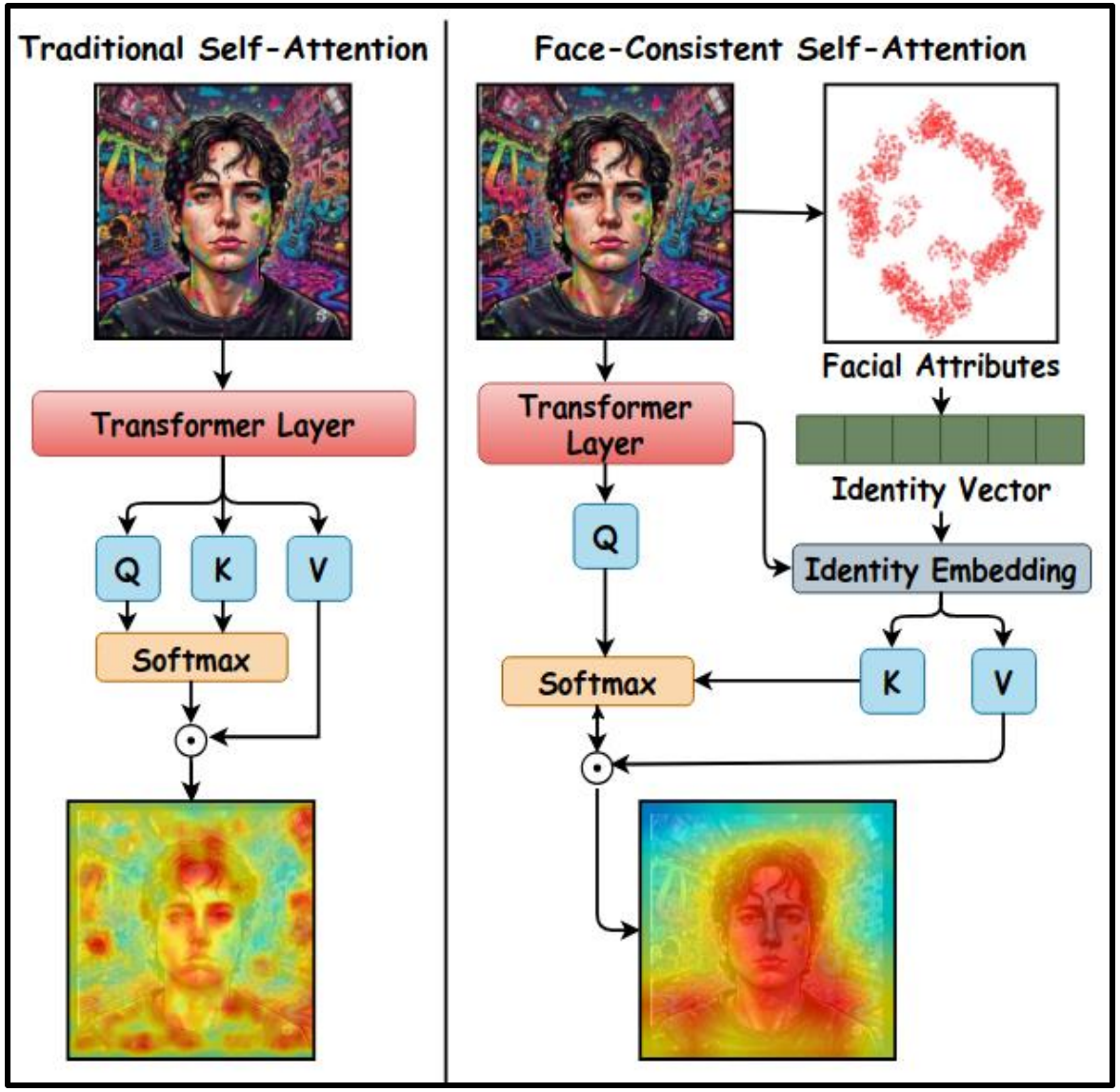


# Our main contribution

## Face-consistent self-attention

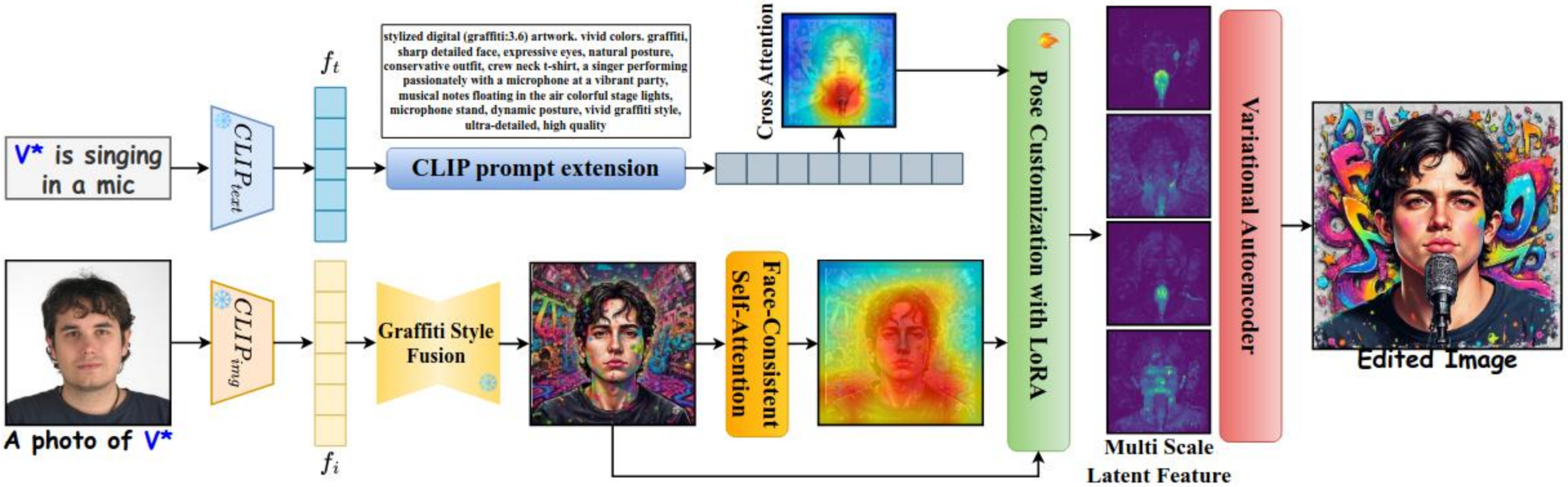
- **Preserve facial features in an identity vector and transfer it into the latent space through identity embedding.**



# *CraftGraffiti* Pipeline

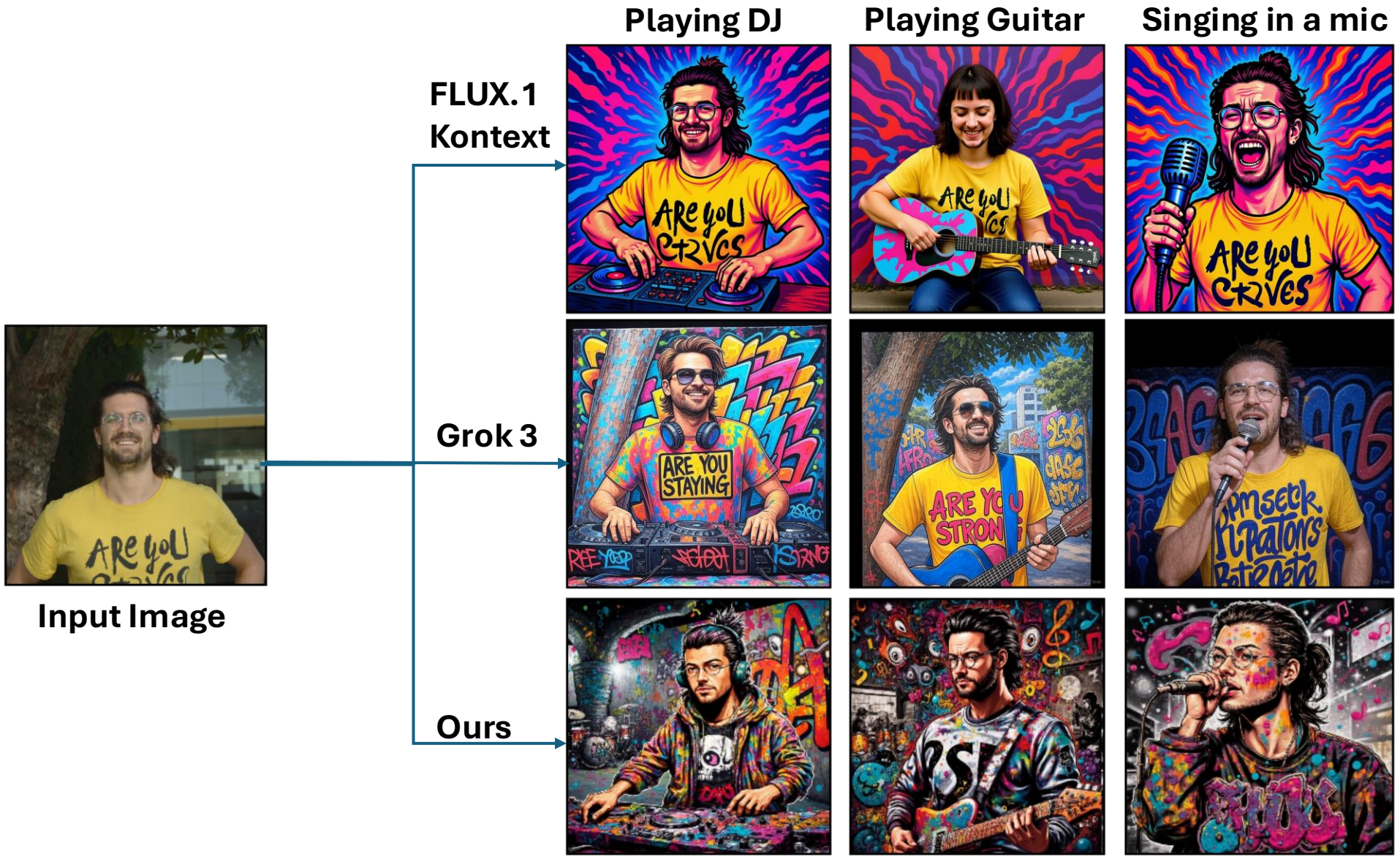
## Simple yet effective




It injects **graffiti style** using a **LoRA-tuned diffusion model**, then applies a **face-consistent** self-attention to preserve facial features. A **pose customization** module guided by **CLIP prompts** enables flexible pose control without retraining. Finally, **multi-scale latent processing** in a VAE captures both global structure and fine details, producing high-quality, identity-consistent graffiti portraits.

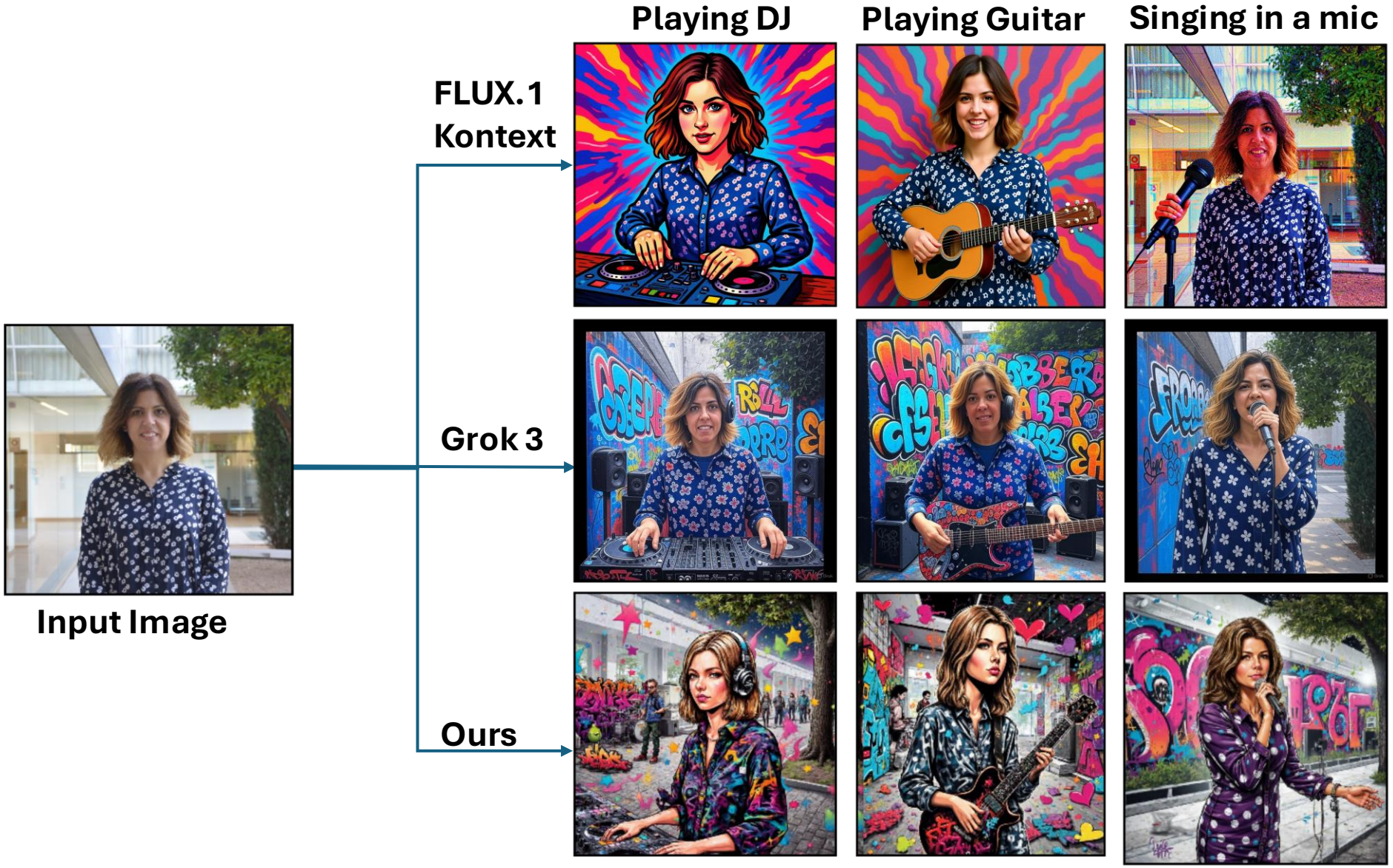
# Experimental Analysis & Comparative Study

## Works irrespective of gender bias



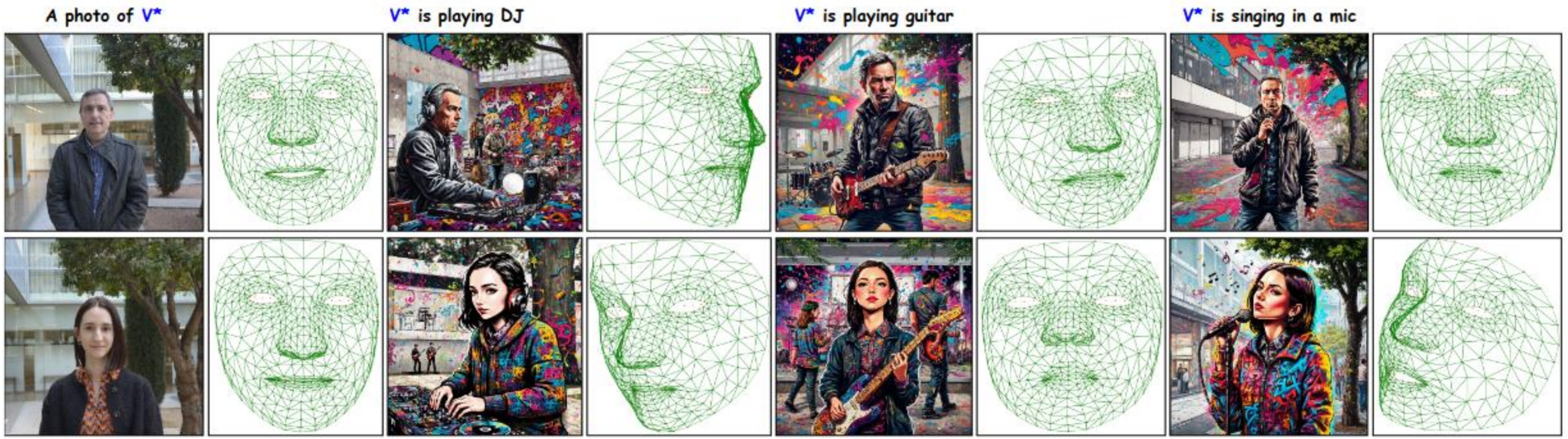
# Experimental Analysis & Comparative Study

## Works irrespective of gender bias



# Face Consistency

## Works irrespective of gender bias




As we can generate similar face mesh during different pose customization in graffiti style, ***Craftgraffiti*** preserve the facial features. (Note: this face meshes are extracted after the generation, this step doesn't involve in training)

# Quantitative Analysis

## Metric and Human Evaluation

NEURAL INFORMATION PROCESSING SYSTEMS 2025

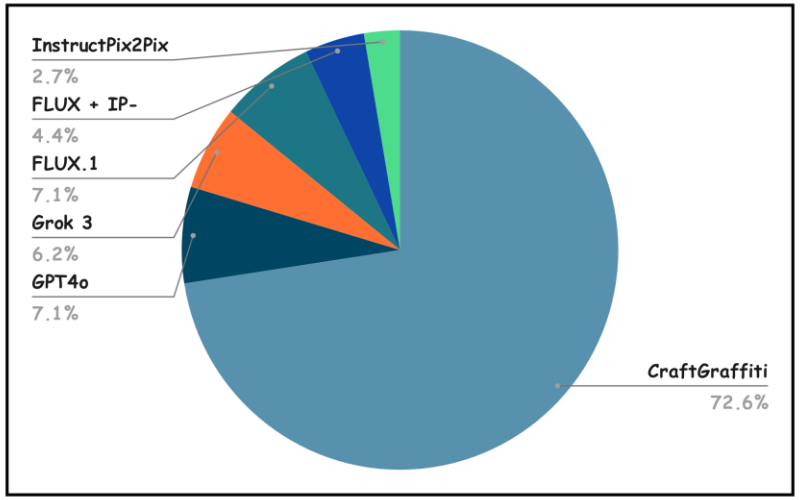
Table 1: Quantitative Evaluation of CraftGraffiti

| Method/Metric | FFC ↑ | Aes ↑ | HPS ↑ | Inf. Time (sec) ↓ |
|---|---|---|---|---|
| FLUX + IP-Adapter [56] | 0.8324 | 3.2414 | 0.3012 | 8.2 |
| FLUX.1 Kontext [27] | 0.6741 | 2.1749 | 0.2911 | 6.7 |
| InstructPix2Pix [6] | 0.7112 | 2.7272 | 0.1918 | 3.1 |
| GPT4o [22] | **0.8761** | 4.5193 | 0.3412 | 13.4 |
| Grok 3 [16] | 0.8513 | 4.1652 | 0.3102 | 11.2 |
| Ours (Baseline) | 0.7618 | 3.6913 | 0.2911 | **2.9** |
| Ours (+ Style Fusion) | 0.6814 | 4.7195 | 0.3001 | 5.3 |
| Ours (+ Face Consistent self attention) | 0.7713 | 5.1376 | 0.3176 | 8.7 |
| CraftGraffiti | 0.7713 | **5.2271** | **0.3536** | 10.1 |

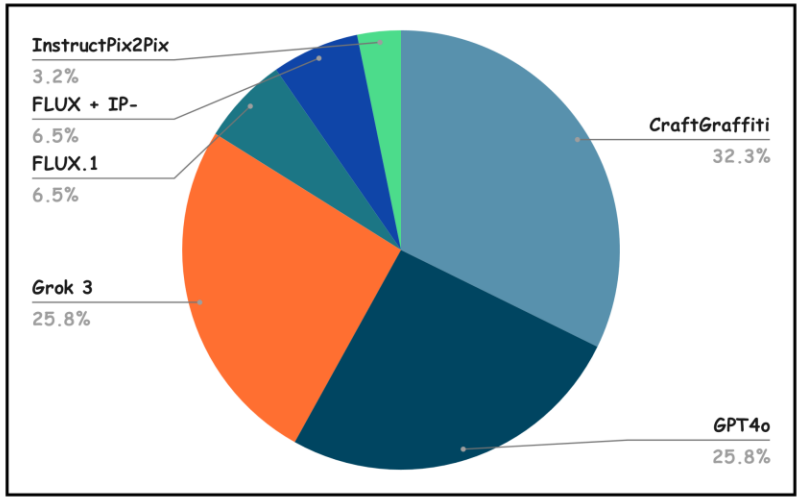
VLMs only focus on face consistency not on style and pose but CraftGraffiti attend every aspects

# Quantitative Analysis

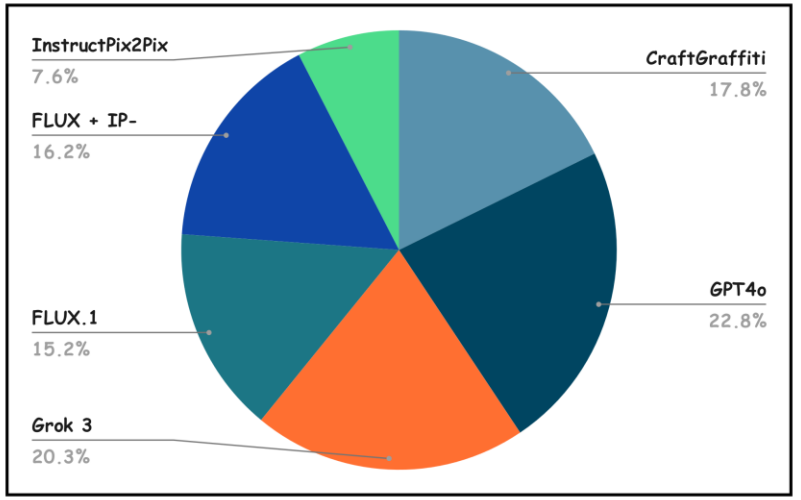
## Metric and Human Evaluation




**Style Blending**




**Aesthetics**




**Recognizability**

CraftGraffiti not only achieve superior performance also address the cultural diversity as the users from 11 different countries participate in human evaluation.

# Ablation Studies

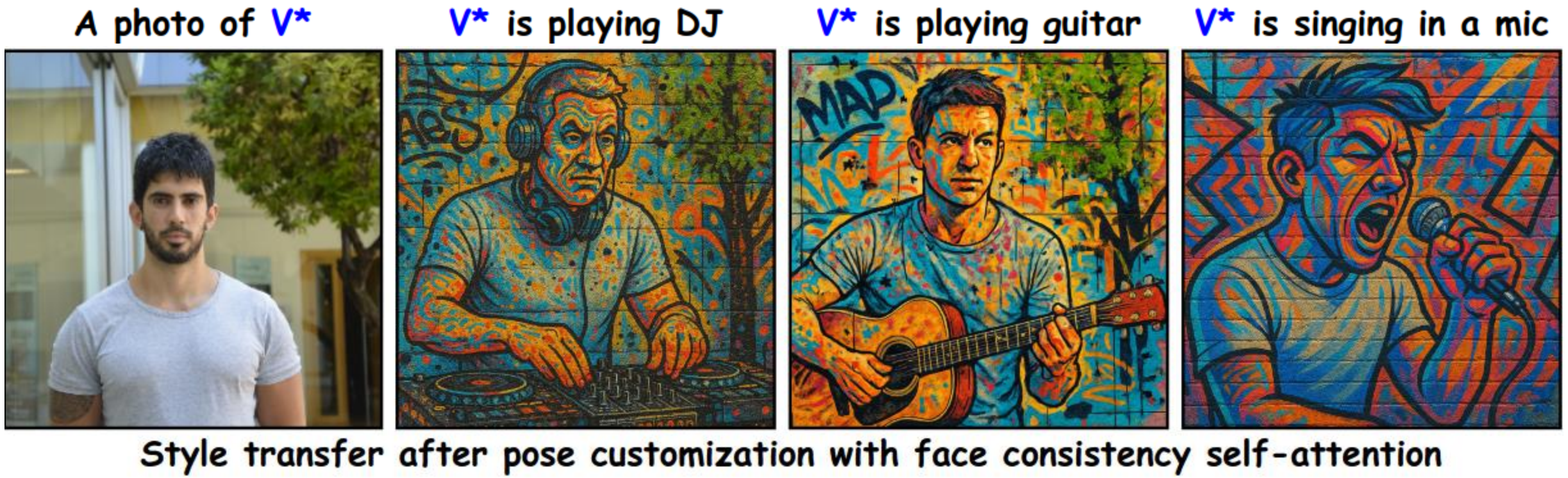
## Why each of the components are effective

# Ablation Studies

## Why each of the components are effective

# Ablation Studies

## Why each of the components are effective



# Ablation Studies

## Why each of the components are effective

# Qualitative Analysis

NEURAL INFORMATION PROCESSING SYSTEMS 2025

A photo of V*

V* is playing DJ

V* is playing guitar

V* is singing in a mic

**Blend perfect style and pose**

# Qualitative Analysis

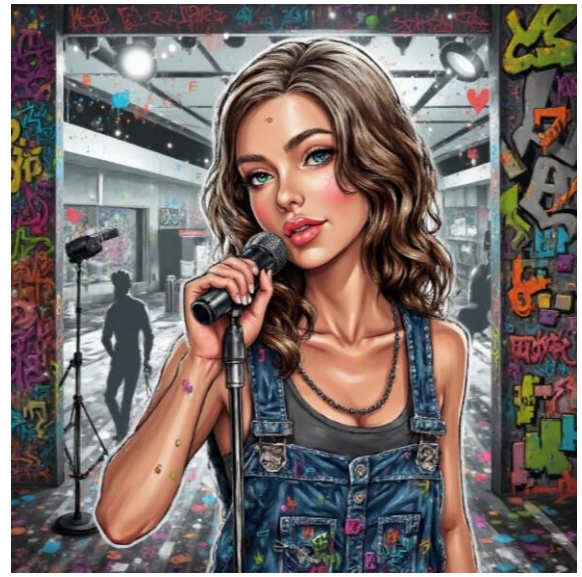
NEURAL INFORMATION PROCESSING SYSTEMS 2025

A photo of V*

V* is playing DJ

V* is playing guitar

V* is singing in a mic



**Mitigate gender bias**

# Qualitative Analysis

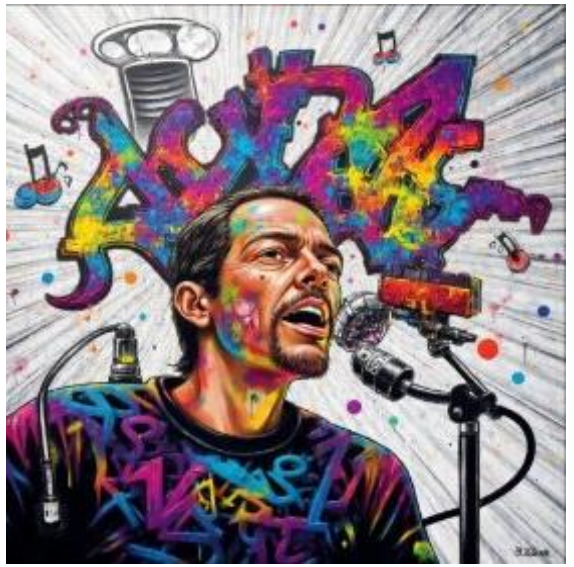
NEURAL INFORMATION PROCESSING SYSTEMS 2025

A photo of V*

V* is playing DJ

V* is playing guitar

V* is singing in a mic



**Works irrespective of the background**

# Qualitative Analysis

NEURAL INFORMATION PROCESSING SYSTEMS 2025

A photo of V*

V* is playing DJ

V* is playing guitar

V* is singing in a mic

Try to maintain hair color as well

# Contribution Bucket

- Face Consistency during style and pose customization.
- Mitigating gender bias.
- Achieve cultural diversity.

# Limitations

- Always generate younger face.
- Put extra make up on women's face.

# Future Scope

- Extending CraftGraffiti to handle a broader range of cultural art forms beyond graffiti.
- Integrating real-time bias detection and mitigation pipelines.


Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL

Plan de Recuperación, Transformación y Resiliencia

ENIA
ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL