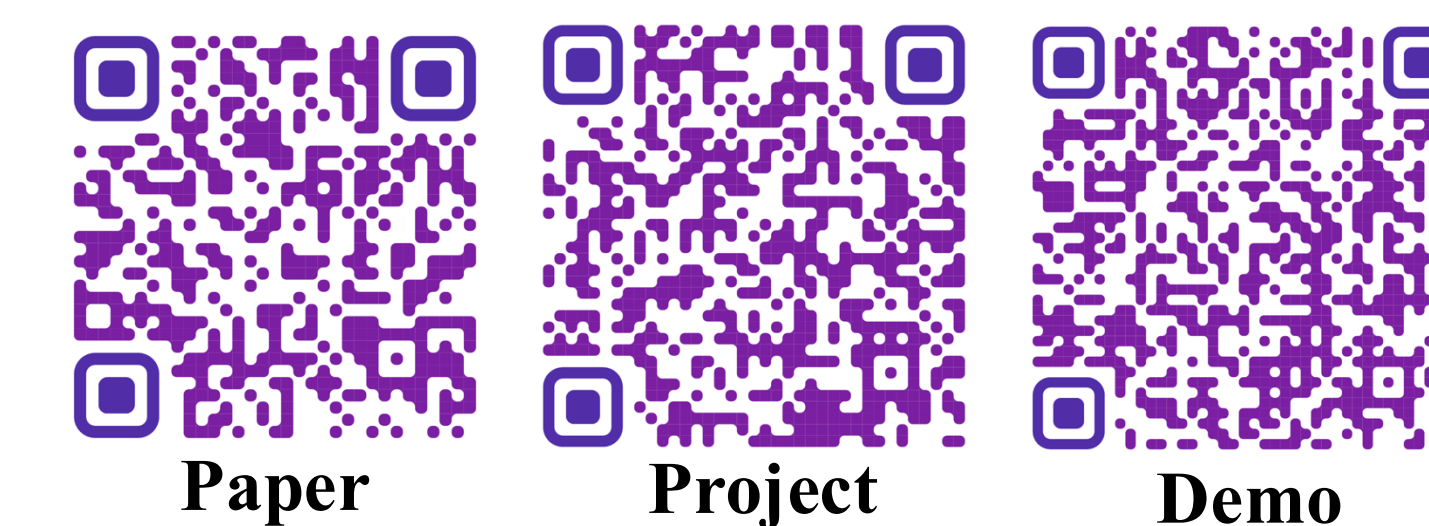


# ID-Patch: Robust ID Association for Group Photo Personalization

Yimeng Zhang<sup>1,2</sup>, Tiancheng Zhi<sup>1</sup>, Jing Liu<sup>1</sup>, Shen Sang<sup>1</sup>, Liming Jiang<sup>1</sup>, Qing Yan<sup>1</sup>, Sijia Liu<sup>2</sup>, Linjie Luo<sup>1</sup>

<sup>1</sup> ByteDance Inc. <sup>2</sup> OPTML Lab, Michigan State University



## Introduction

### Task

Synthesize personalized group photos and specify the positions of each identity offers immense creative potential.

### Challenges

OMG [1] uses a two-stage process: generate without IDs, then inject identities via segmentation.

- Relies on accurate segmentation
- Slow: separate denoising per person

InstantFamily [2] applies identity-aware attention masks in one pass.

- Suffers from mask inaccuracies and overlapping faces
- Information leakage through attention and convolution

### Contributions

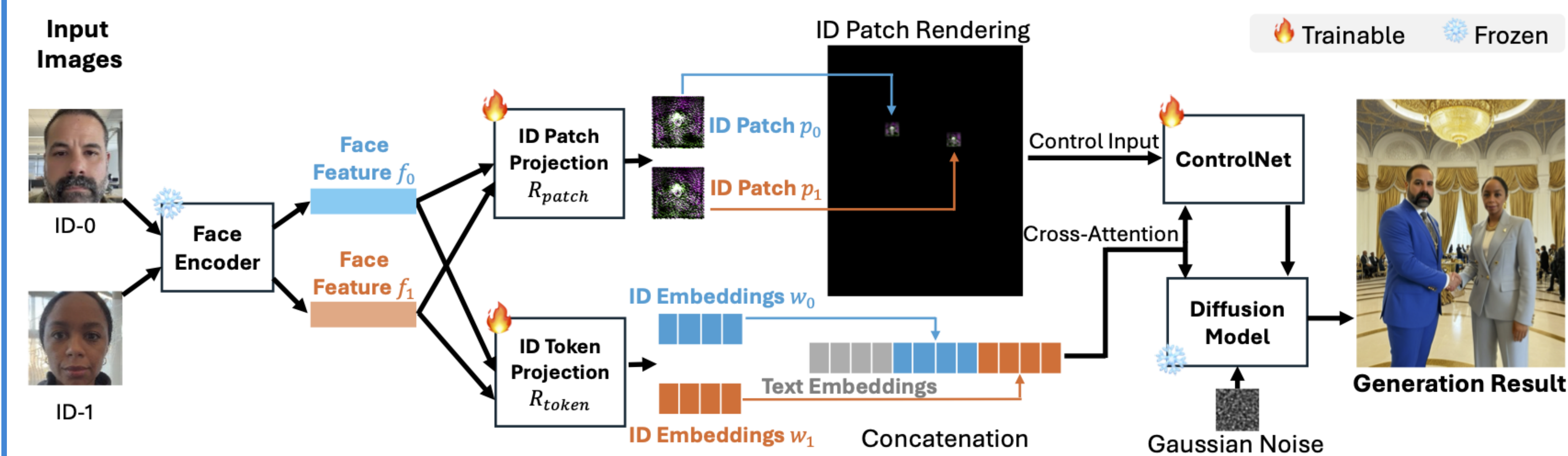
- ID-Patch** links identity features with spatial positions via visual patches for precise control.
- Efficient**: Simplifies multi-ID generation with minimal computation.
- Robust**: Avoids segmentation models; uses single-point control.
- Accurate**: Delivers superior identity fidelity and spatial placement, especially in complex scenes.

## Reference

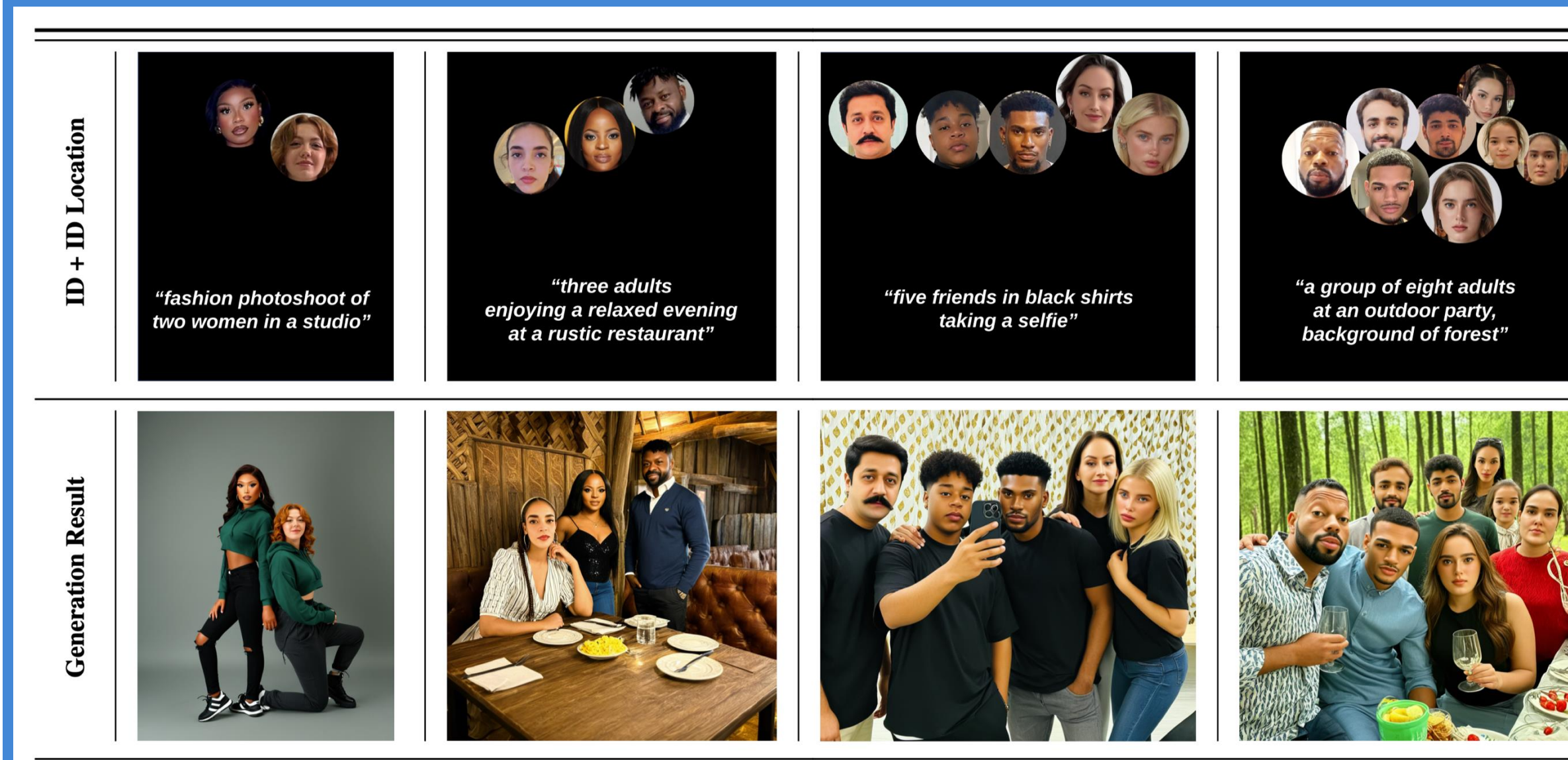
[1] Kong, Zhe, et al. "Omg: Occlusion-friendly personalized multi-concept generation in diffusion models." ECCV 2024.

[2] Kim, Chanran, et al. "Instantfamily: Masked attention for zero-shot multi-id image generation." arXiv preprint arXiv:2404.19427 (2024).

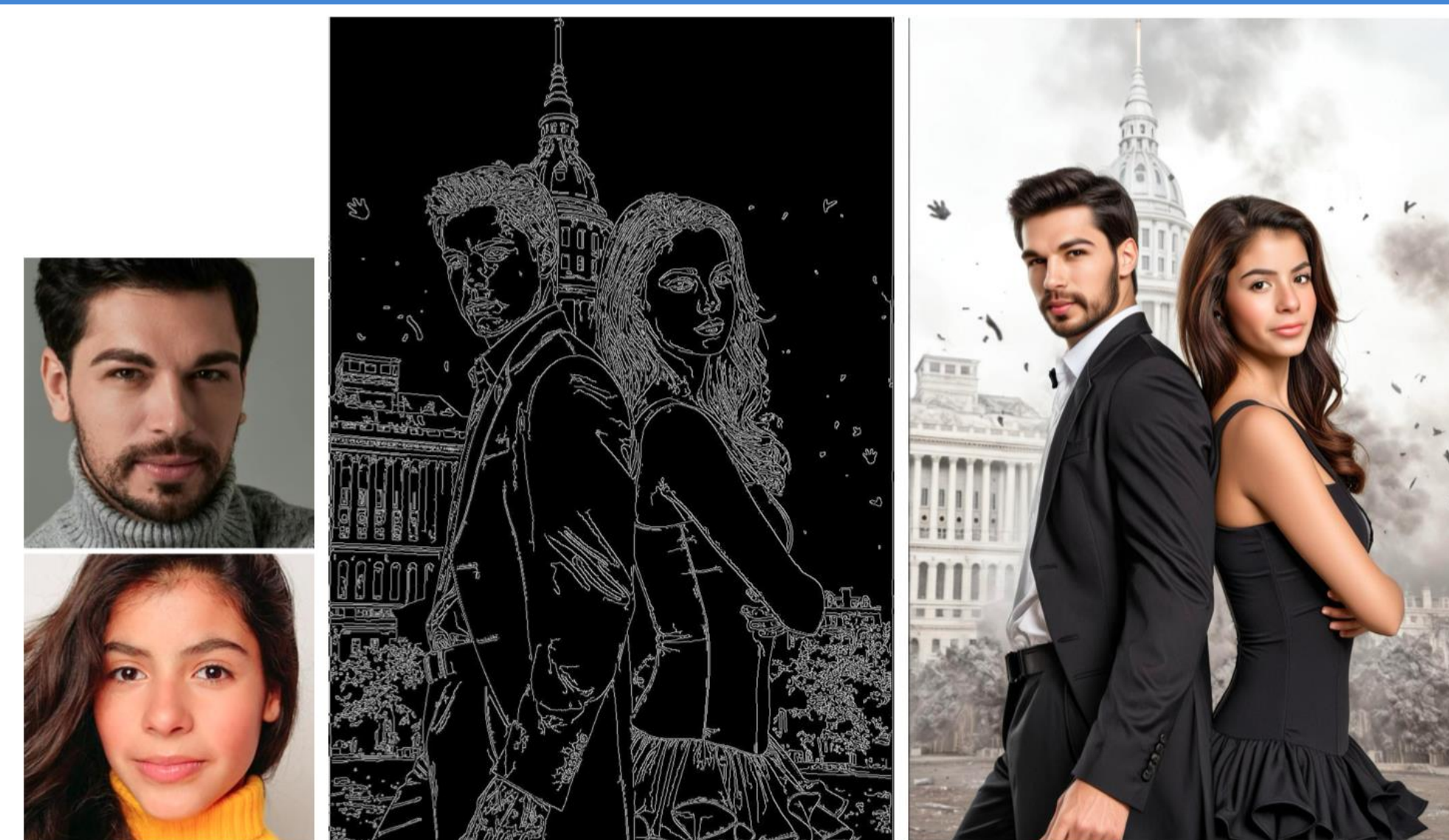
## ID-Patch: Build Identity-to-Position Association



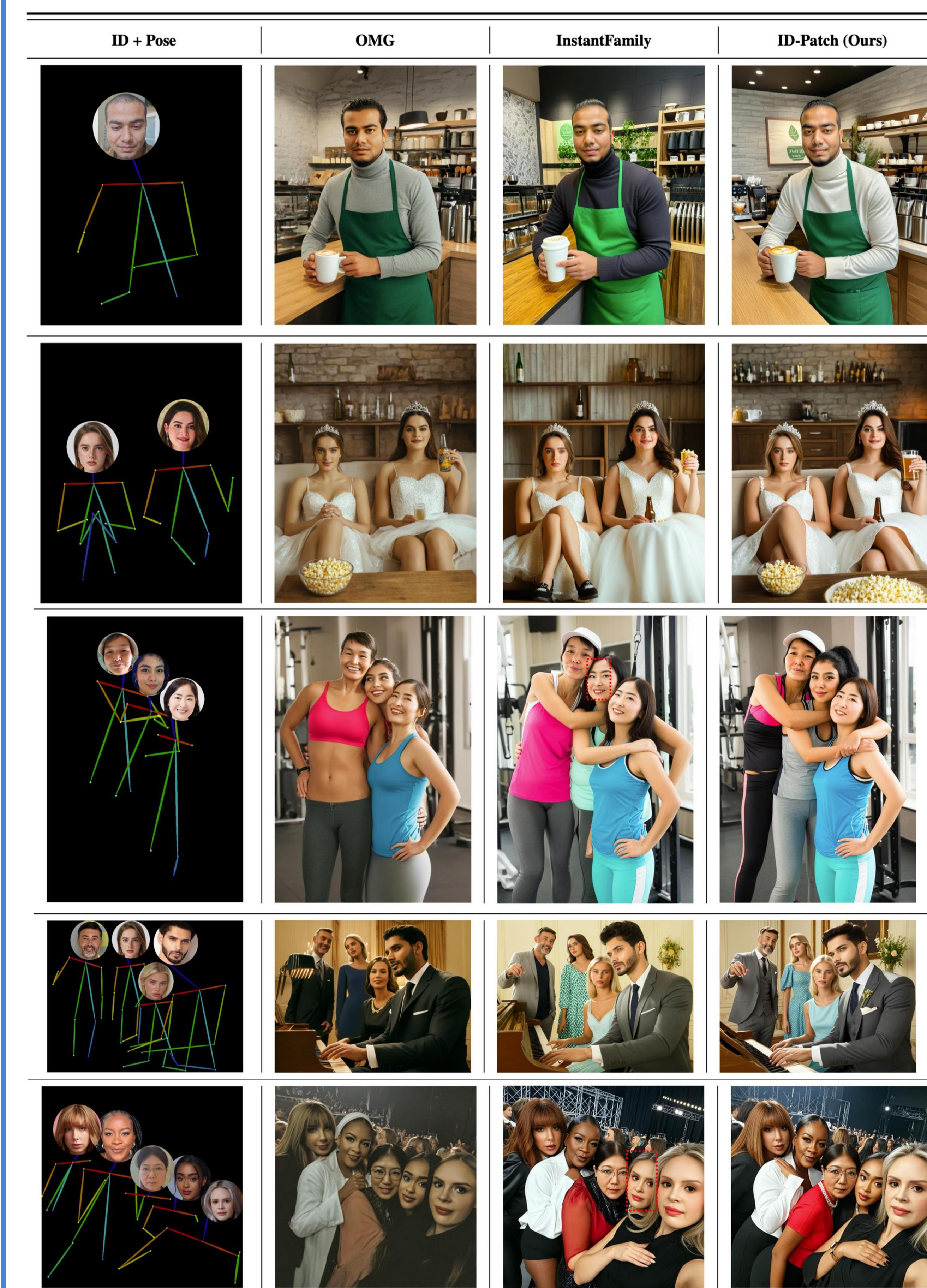
## Pose-Free Generation



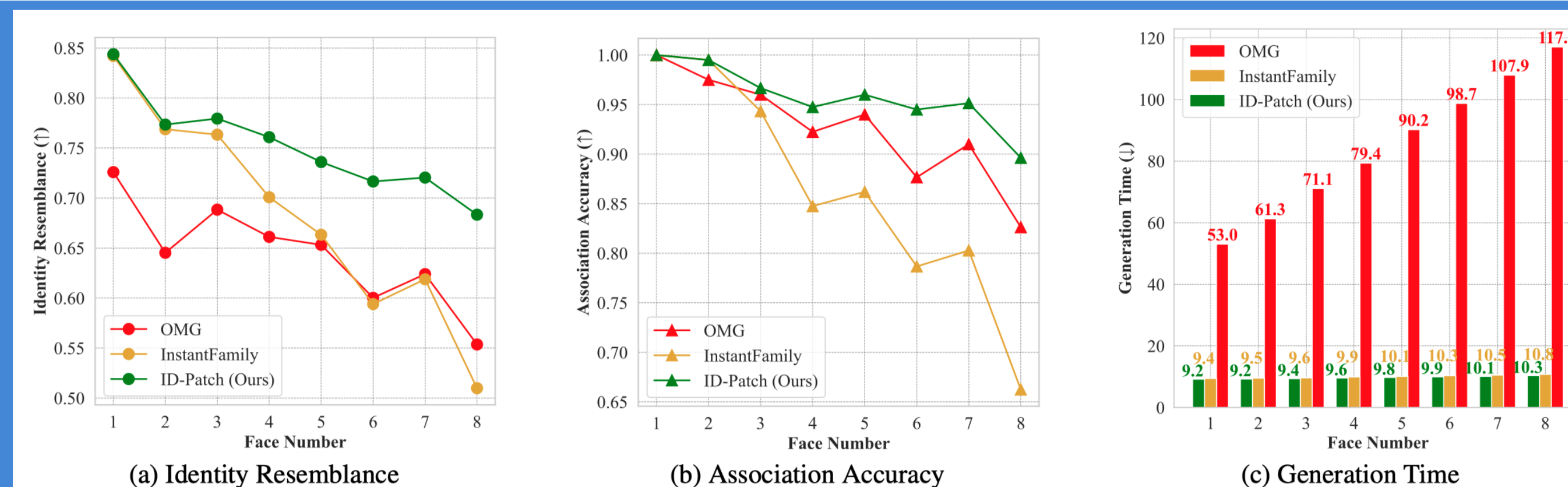
## Plug-and-Play: Canny Edge



## Visualizations



## Evaluations



## Limitations

