

Qualcomm

Summary: Resolving the Identity Crisis in Text-to-Image Generation

Shubhankar Borse, Farzad Farhadzadeh, Munawar Hayat,
Fatih Porikli

<https://arxiv.org/abs/2510.01399>

Presenter: Shubhankar Borse
Staff AI Reseacher
Qualcomm AI Reserach

@qualcomm

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries.
Qualcomm patents are licensed by Qualcomm Incorporated.



MOTIVATION



Our motivation is to Generate rich synthetic data generation for MultiHuman Task

Why do we need to do T2I?

- To do good Personalized T2I, we need rich data of portrait images of multiple people, which we don't have.
- If we can get that using T2I would solve the problem

What are we looking for in these images:

- They should not be overlapping faces
- Images should look realistic
- The prompt and Image should match

BACKGROUND: THE IDENTITY CRISIS

Gemini-NanoBanana (September 2025)



Flux-Krea (August 2025)



HiDream-Full (April 2025)



OmniGen2 (June 2025)



- To train our SOTA reference-based multi-human generation model, we need high quality data.
- We use SOTA T2I generation models for this purpose.
- SOTA models (including proprietary methods) struggle with the **Identity Crisis**.
- Apart from duplicate identities in the same image, they produce the same faces across the data generation process.
- Additionally, they are inaccurate at generating the correct number of humans.

Proposed Method



PROPOSED METHOD

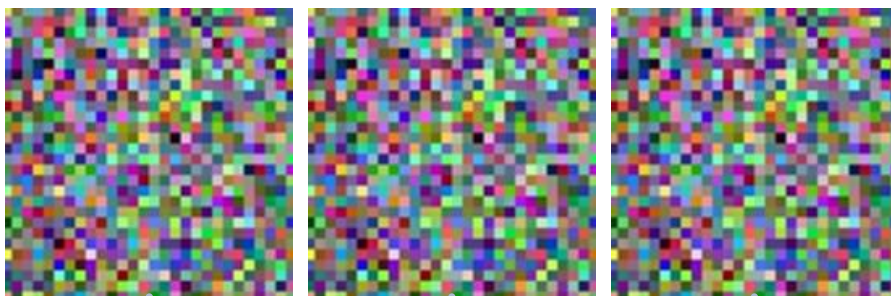
Prompt
"Seven people in a soccer field"

 Text-to-Image Generator

Policy

GRPO
Policy
Update

Intermediate Timesteps



Intra-Image
Diversity

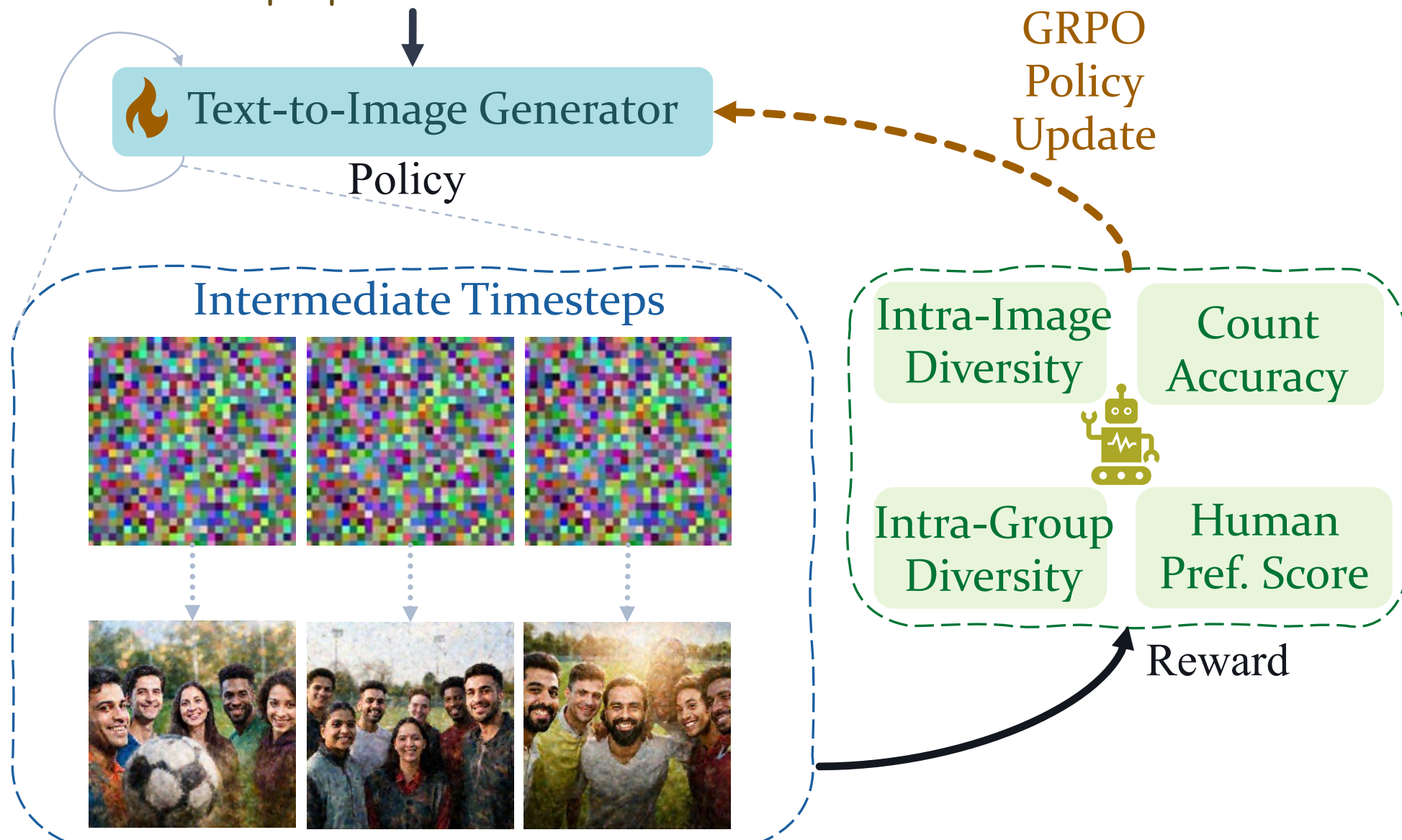
Count
Accuracy



Intra-Group
Diversity

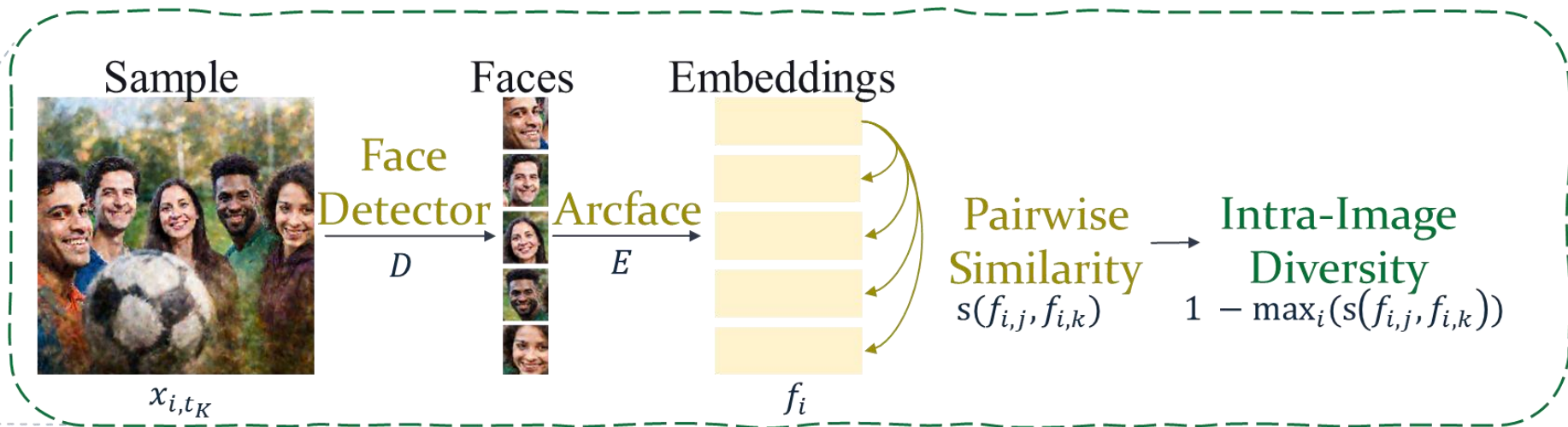
Human
Pref. Score

Reward

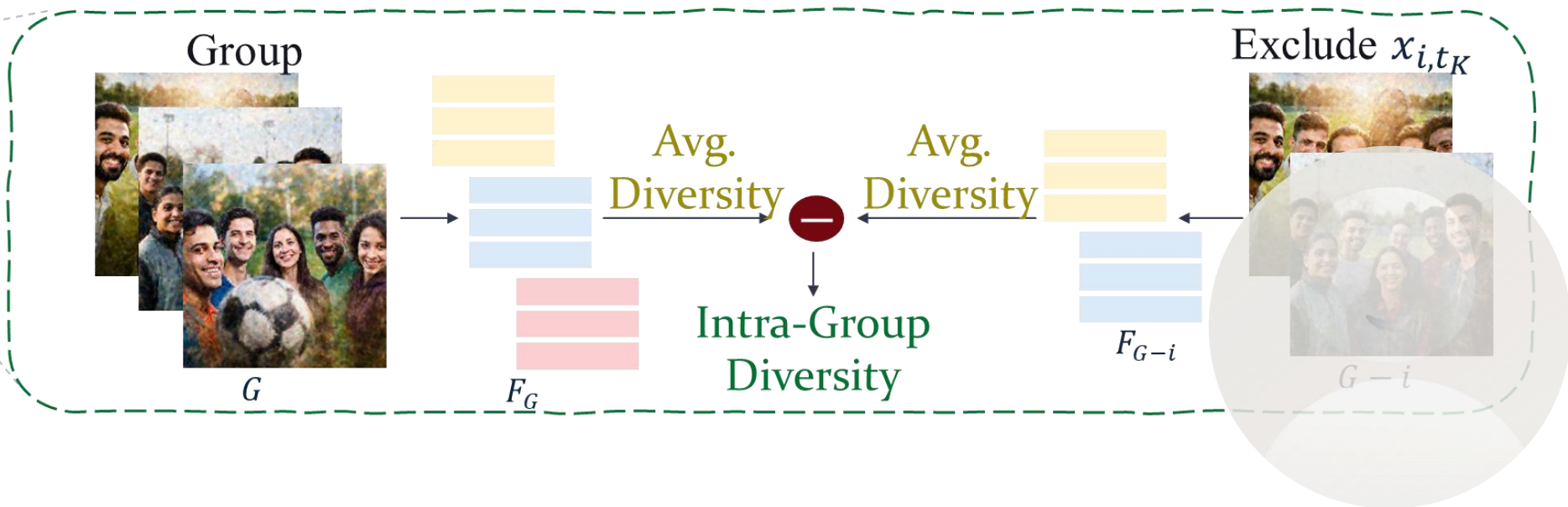


PROPOSED METHOD

Intra-Image Diversity



Intra-Group Diversity



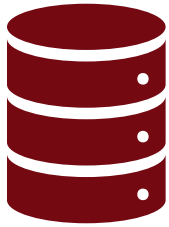
Experiments and Results



TEST DATASETS

Multi-Human Testbench*

- 1800 samples
- 1-5 People



Diverse-Humans

- 1200 samples
- 2-7 People



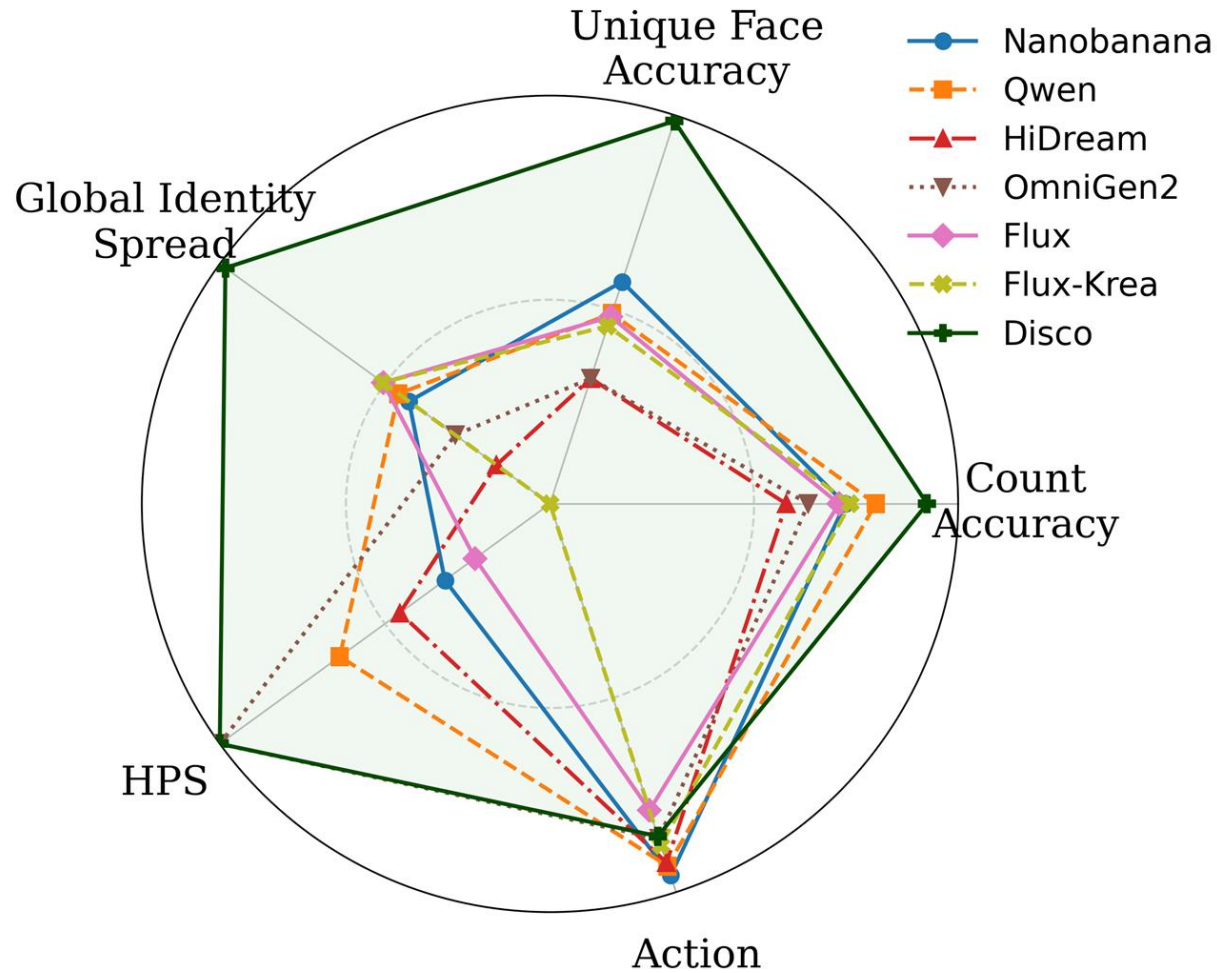
25% No tag: Five people on an island cove beach, High dynamic range, Group harmony, Professional portrait, Natural lighting, Smiling expressions

25% Diverse Tag: Five people in an antique arcade, High dynamic range, Sharp focus, Group harmony, Clear faces, Smiling expressions, **Diverse faces among people**

25% Single Ethnicity: Five people in a sidewalk cafe, Sharp focus, Bokeh background, Well lit, Clear faces, Group harmony, **Indian ethnicity**

25% Individual Assignments: Five people in a coastal market, Bokeh background, Sharp focus, Professional portrait, Portrait photography, One person is White, One person is Middle-eastern, One person is Asian, One person is Black, One person is Hispanic




































QUANTITATIVE RESULTS



Metrics:

- **Unique Face Accuracy:** Percentage of images without ANY ID overlap
- **Global Identity Spread:** Percentage of new faces across all the generated faces

QUALITATIVE RESULTS (DISCO-FLUX)

	GPT	Nanobanana	Flux-Krea	HiDream-Full	OmniGen2	Flux-Dev	DisCo(Flux-Dev)
"Six people in a castle's ramparts"	 ✓ ✗	 ✓ ✗	 ✗ ✗	 ✗ ✗	 ✗ ✗	 ✓ ✗	 ✓ ✓
"Seven people in a glacier lookout"	 ✓ ✓	 ✗ ✗	 ✓ ✗	 ✗ ✗	 ✗ ✗	 ✓ ✗	 ✓ ✓
"Seven people inside a desert cave"	 ✓ ✓	 ✓ ✓	 ✗ ✗	 ✗ ✗	 ✓ ✗	 ✗ ✗	 ✓ ✓
"Six people on a canopy bridge"	 ✓ ✗	 ✓ ✗	 ✗ ✗	 ✓ ✗	 ✗ ✗	 ✗ ✗	 ✓ ✓
"Six people inside a village plaza"	 ✗ ✗	 ✓ ✗	 ✓ ✗	 ✗ ✗	 ✗ ✗	 ✓ ✗	 ✓ ✓

Scoresheet

Count Accuracy:

- Win: ✓
- Loss: ✗

Unique Faces:

- Win: ✓
- Loss: ✗

QUALITATIVE RESULTS (DIVERSITY ACROSS DATASET)

Prompt

Flux-Dev

+DisCo

"Four people in a park, portrait photography, smiling expressions, **Indian** ethnicity"



"Four chefs in a kitchen, portrait photography, smiling expressions, **Asian** ethnicity"



"Three people in the snow, portrait photography, smiling expressions, **White** ethnicity"



Thank you

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

© Qualcomm Technologies, Inc. and/or its affiliated companies. All Rights Reserved.

Qualcomm and Snapdragon are trademarks or registered trademarks of Qualcomm Incorporated. Other products and brand names may be trademarks or registered trademarks of their respective owners.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes our licensing business, QTL, and the vast majority of our patent portfolio. Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of our engineering, research and development functions, and substantially all of our products and services businesses, including our QCT semiconductor business.

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries. Qualcomm patents are licensed by Qualcomm Incorporated.

Follow us on: [in](#) [X](#) [@](#) [v](#) [f](#)

For more information, visit us at qualcomm.com & qualcomm.com/blog

