Customizing Realistic Human Photos via Stacked ID Embedding

What is the Problem?

- **Generating realistic human photos based on text descriptions** is challenging, especially when maintaining the person's identity (ID) across different settings, expressions, and attributes.
- Existing methods like **DreamBooth are either slow** or require extensive fine-tuning, making them impractical for personalized content generation.

What Has Been Done Earlier?

- **Previous Work**:
  - **DreamBooth**:
    - Fine-tunes models on specific identities to generate personalized images.
    - High ID fidelity but time-consuming and requires a large dataset for each identity.
  - **Textual Inversion**:
    - Encodes a single image into a vector (embedding) for generating new images.
    - Faster but struggles with maintaining high ID fidelity and diversity.
  - **FastComposer/IPAdapter**:
    - Offers some improvements in speed and text controllability but still lacks the ability to balance efficiency, fidelity, and diversity

Suman Sahoo, b421056

Proposed Solution and Challenges

What Are the Remaining Challenges?

- **Challenges**:
  - **Balancing ID fidelity** (keeping the person recognizable) with diversity in generated images.
  - **Reducing the time** and computational resources needed for personalized image generation.
  - Ensuring the model does not memorize irrelevant details, which can reduce the quality and variety of outputs.

What Novel Solution is Proposed by the Authors?

- **Proposed Solution: PhotoMaker with Stacked ID Embedding**:
  - **Stacked ID Embedding**:
    - Combines multiple images of the same person into a single, rich embedding that better preserves identity.
    - This unified representation allows for more realistic and diverse image generation while keeping the person recognizable.
  - **Efficiency**:
    - Achieves high-speed image generation (130x faster than DreamBooth) with no need for test-time fine-tuning.
  - **Versatility**:
    - Supports a wide range of applications like attribute modification, identity mixing, and converting artworks into realistic photos.