香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Image Synthesis & Background Replacement

— Graph-to-Graph Triple U-net Generator & Patch-GAN Discriminator

GROUP 36:
Haixun ZHENG 122020291
Xinyuan QUAN 122090442
Penggan XU 122090625
Chenyun MO 122090406
Yang ZENG 122090714

Content

Part I. Introduction

Introduction

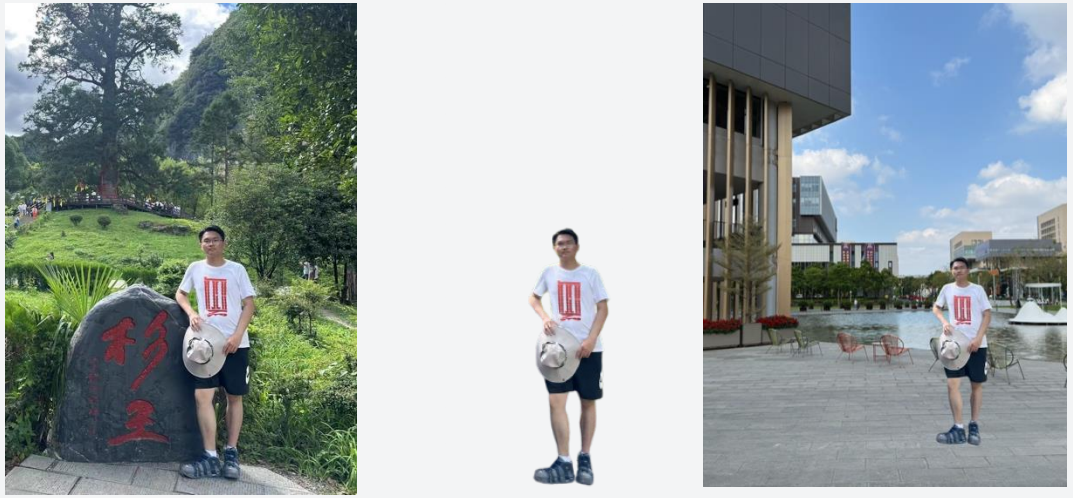
First Significance

Most of Existing Tools (baseline):

- Simply change the background
- Do not consider ***Size & Position*** of the person

Our Advancement:

- Can achieve the baseline performance



- ***Learn & Adjust*** the ***Reasonable Size & Position***

Introduction

Second Significance

Existing Research Methods

Three models (object placement, background harmonization, and shadow generation) are ***separate*** and are done ***step by step***.

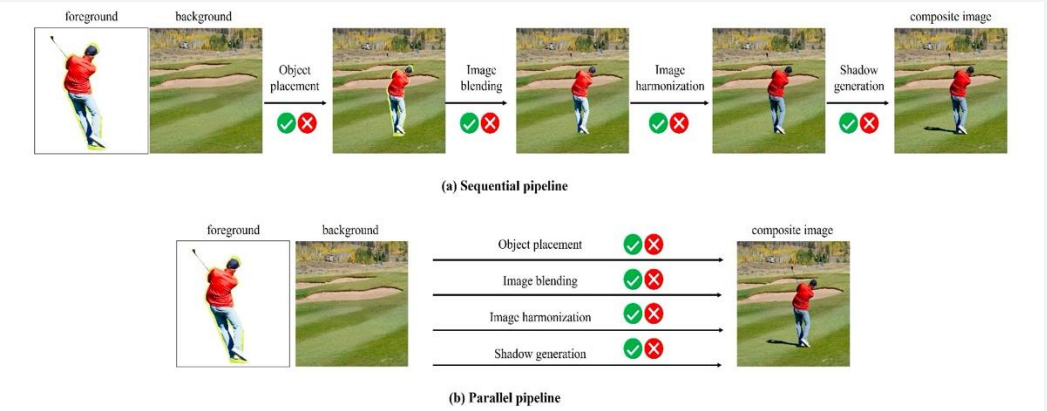


Fig. 3. Previous works perform multiple sub-tasks (*e.g.*, object placement, image blending, image harmonization, shadow generation) sequentially or parallelly to achieve the goal of image composition.

Our Advancement

- We tried to ***integrate these three*** functions ***into one model***

Benefits:

- **Enhance Efficiency:** reduce the number of ***operational steps & time***
- **Improved Consistency:** the generated results are more ***uniform & coordinated***

Part II.
Data Collection & Preprocessing

Data Collection

Image Synthesis & Background Replacement

5661 Images of Single-person

https://github.com/csjliang/PPR10K

Single-character Images in CUHKSZ and their background

Pre-train Base Model

Fine-tune Base Model

Data Preprocessing

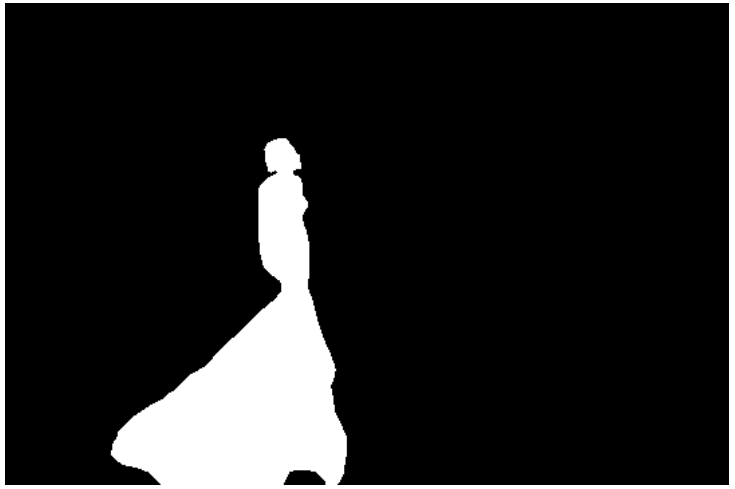
Step 1: Graph Matting

Model: **birefnet-portrait**
Package: **rembg** in python
https://github.com/danielgatis/rembg

Separated Person

Incomplete Background

Mask Image



Data Preprocessing

Step 2: Graph Inpainting

https://github.com/advimman/lama

Why Inpainting?

- **Preventing Indication of Position**

Data Preprocessing

Step 3: Augmentation

Enhancing:

- **model robustness**
- **generalization ability**

Enabling to Learn:

- **optimal size & position** through GAN

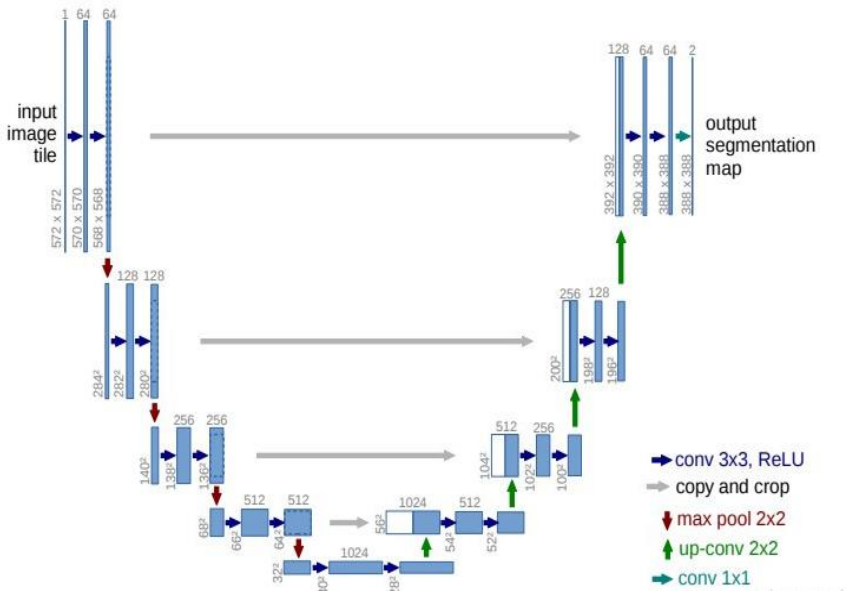
Resizing & Rotating

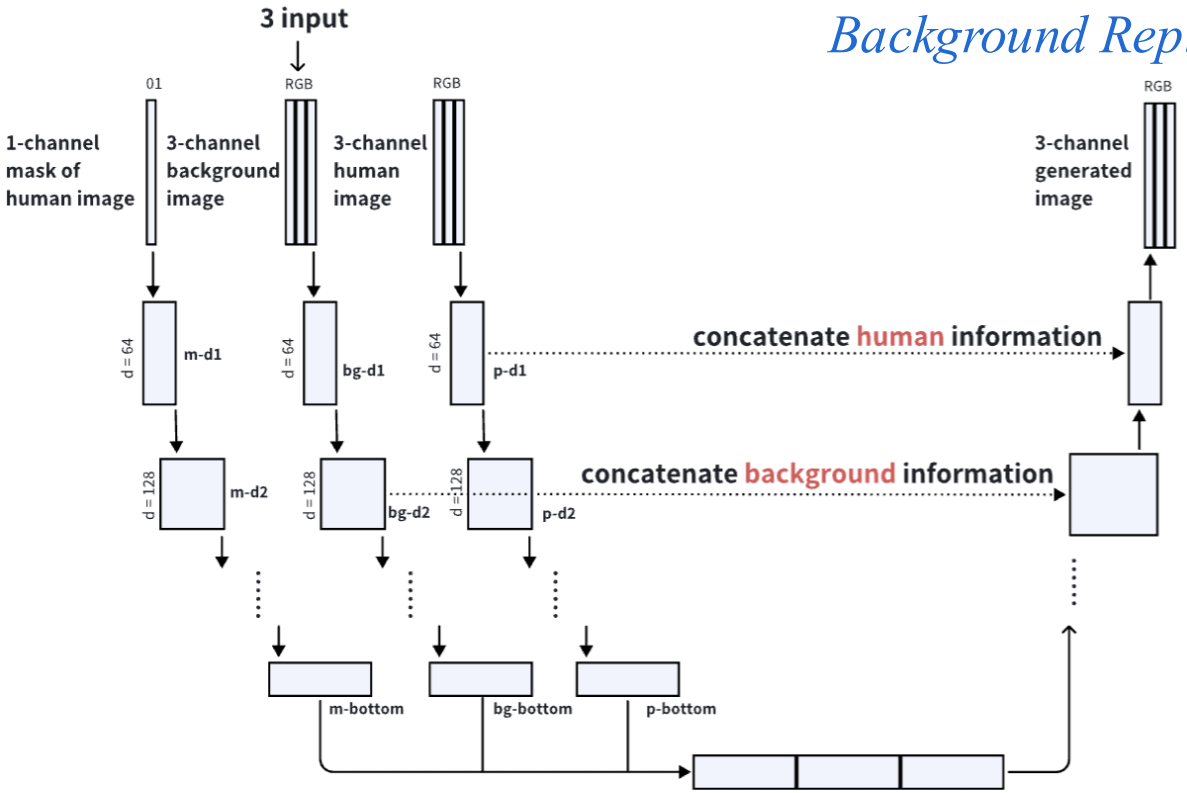
Shifting & Flipping

Shifting & Rotating

Part III. Methodology

Triple U-net Generator





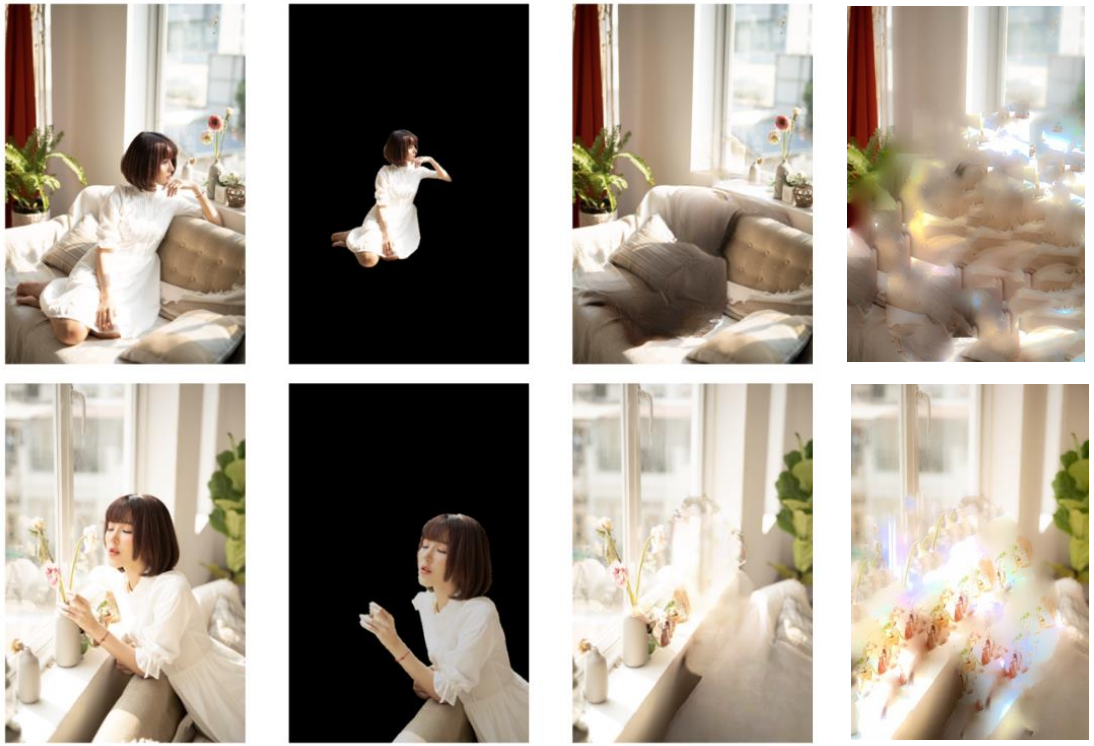
Triple U-net Generator

Improvements:

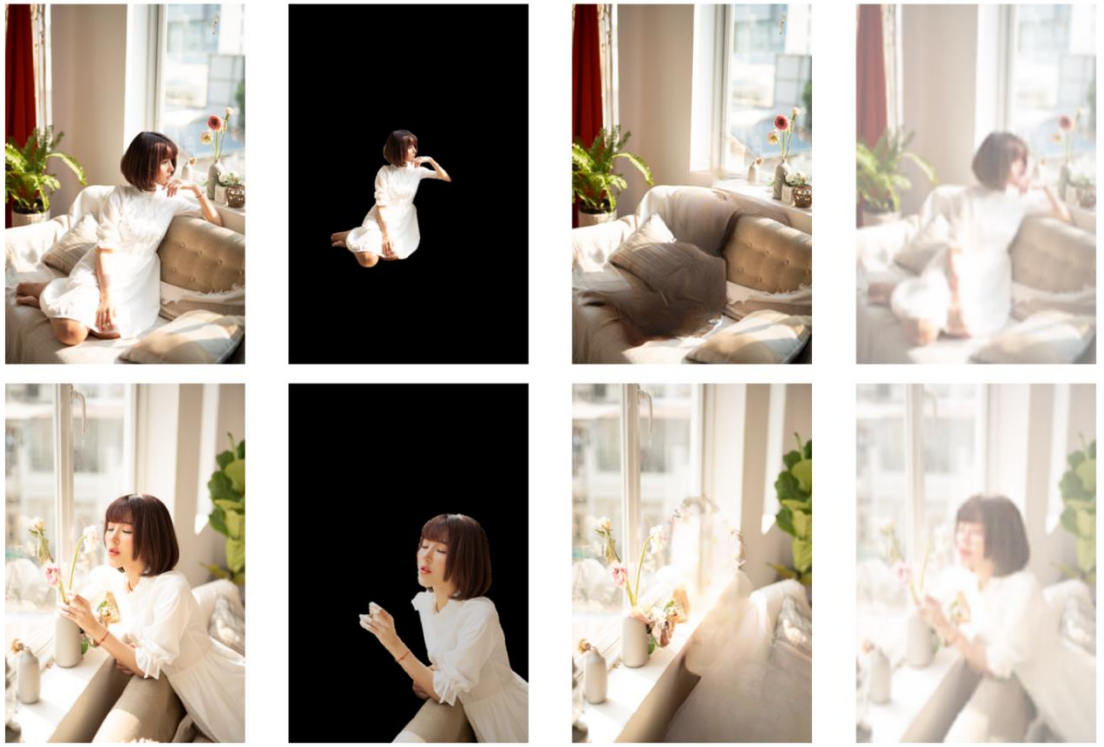
- Single input → **Multiple inputs (mask / people / background)**
- Single generator → **Multiple generators (one input ~ one generator)**
- Simple concatenate → **Dynamic concatenate (shift different information at each layer)**

Triple U-net Generator

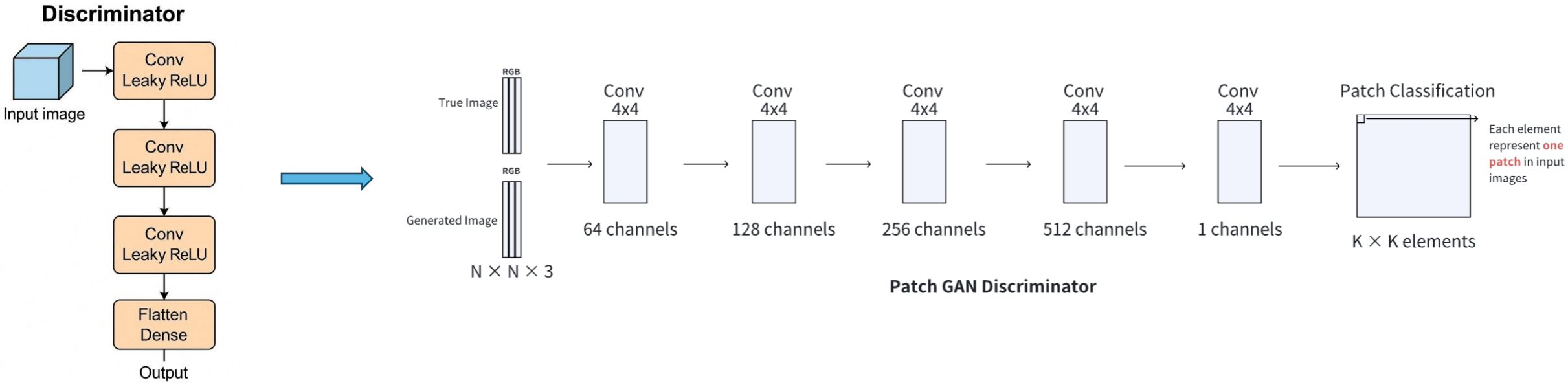
Only 1 Generator: 3 Inputs Stacking



3 Generators: One Input, One Generator



Patch GAN Discriminator



Improvements:

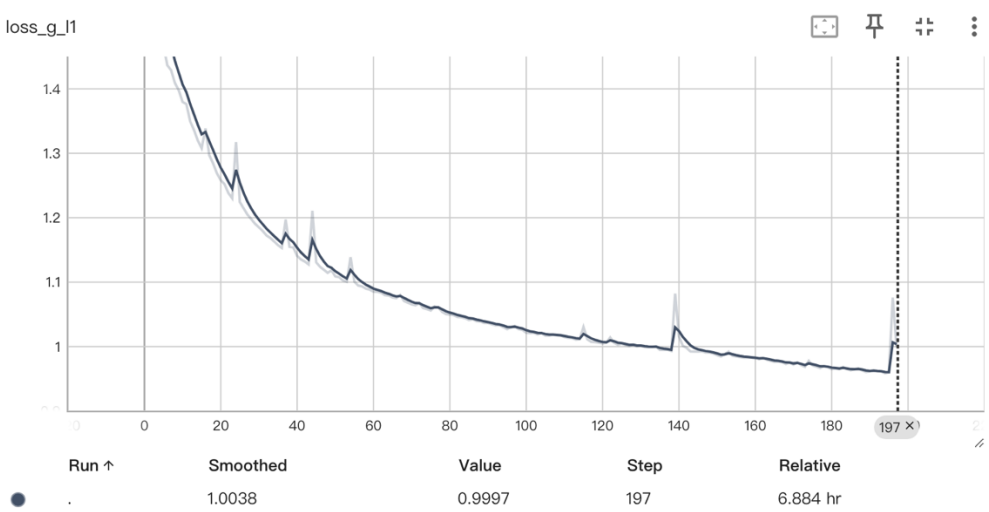
- **Whole-image classification → Patch-level probability**
- **Preserve the spatial structure information of the input image and local details**
- **More efficient calculations, only needs to process local regions rather than global features.**

Part IV.
Numerical & Experimental Results

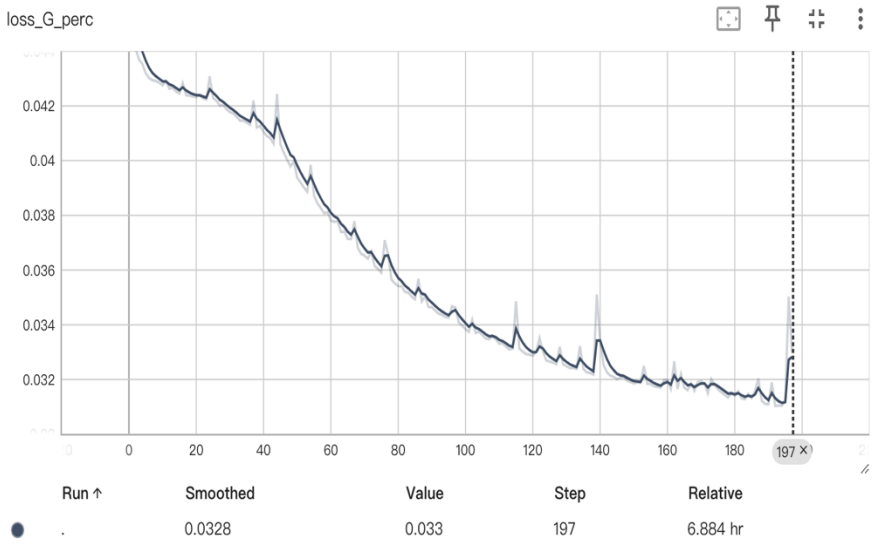
Numerical & Experimental Results

Numerical Results:

- 1



- 2



L1 Loss of Generators: Pix-to-Pix Difference

Decreasing from greater than 1.4 to smaller than 1

Decreased by 28.5%

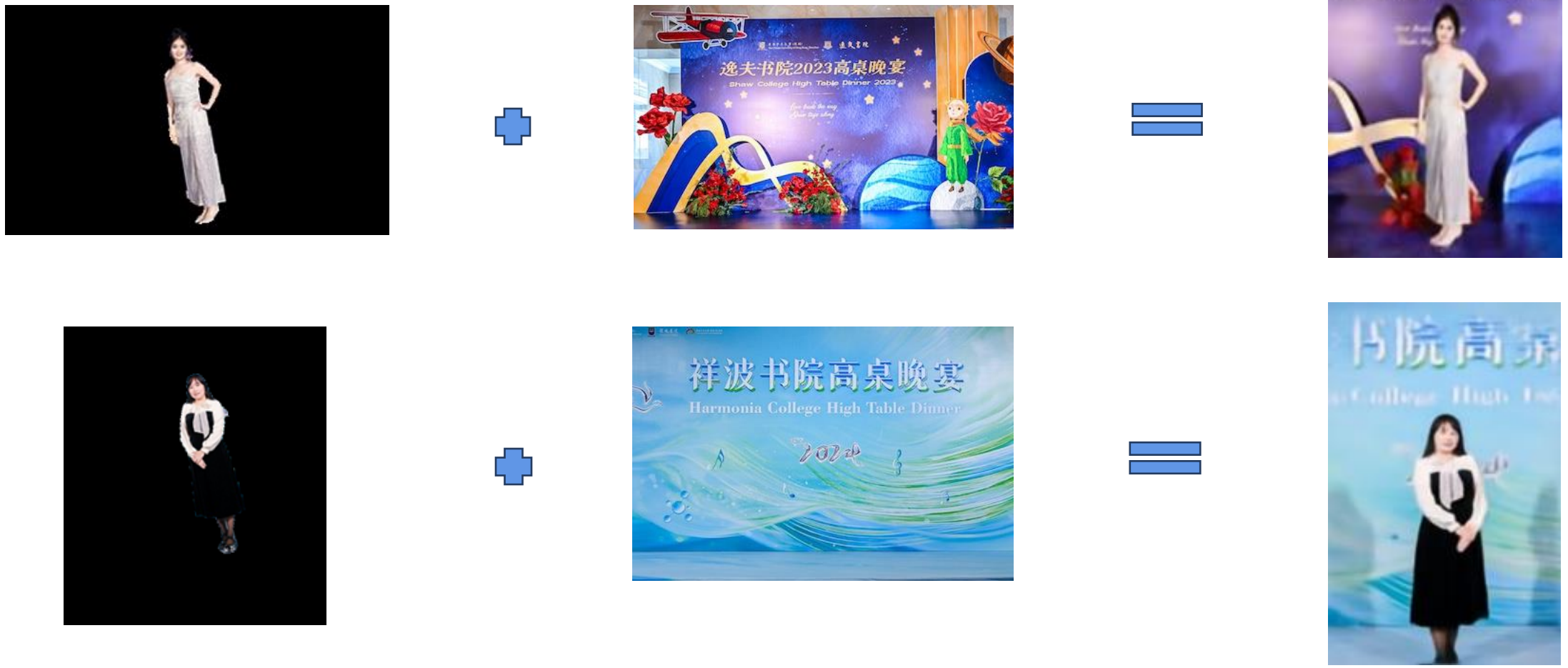
Perceptual Loss of Generators: High Level Structure difference

Decreasing from greater than 0.044 to smaller than 0.032

Decreased by 27.2%

Numerical & Experimental Results

Experimental Results:

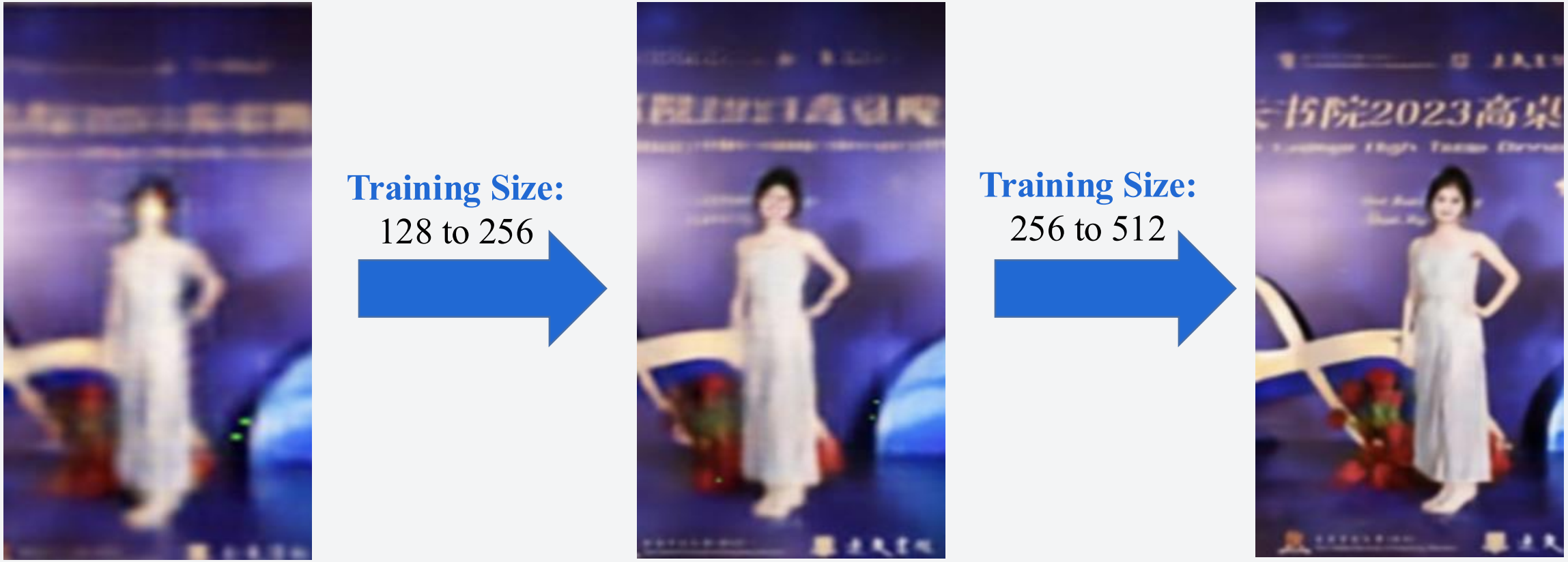


Part V. Limitations

Limitations

The Blurriness of the Image

Increasing the image input size enables the ***extraction of more features***, making the image ***clearer***



Limitations

Generalization Issue

- ***Size, Orientation, & Brightness*** of the person need to be intelligently adjusted
- Result may not be satisfactory for some ***individual special images***

Limited Dataset for CUHKSZ

- Limited to the specific situations and scenarios set by CUHKSZ
- ***OOP Problem:*** May not be able to accurately handle ***new situations*** that have ***not appeared*** in the training data

Limitations

OOP Performance Issue

- Due to **Limited Datasets**:
- Model could not extract the **casual outfit feature** (which has a big gap from formal outfits in our training dataset) very well

Limitations

Future Improvements

Optimize the Model Architecture

Further Reduce Image Blurriness

Enhance the Size & Position Adjustment structure

Handle more complex image scenarios.

Expand the dataset

Improve the model's generalization ability

Part VI. Conclusions

Conclusions

Progress

Our model:

- ***Learn & Adjust*** the reasonable ***Size & Position***
- Attempt to ***Integrate 3 Phases***:
  - Object placement,
  - Background harmonization,
  - Shadow generation

Application

- **Campus Event Promotion**
- **Student Personal Creation**

Thank You for Your Listening

GROUP 36:
Haixun ZHENG 122020291
Xinyuan QUAN 122090442
Penggan XU 122090625
Chenyun MO 122090406
Yang ZENG 122090714

References

- Azadi, S., Pathak, D., Ebrahimi, S., & Darrell, T. (2020). Compositional GAN: Learning image-conditional binary composition. *International Journal of Computer Vision, 128*(10), 2570-2585.
- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
- Niu, L., Cong, W., Liu, L., Hong, Y., Zhang, B., Liang, J., & Zhang, L. (2021). Making images real again: A comprehensive survey on deep image composition. *arXiv preprint arXiv:2106.14490*.
- *https://github.com/csjliang/PPR10K*
- *https://github.com/danielgatis/rembg*
- *https://github.com/advimman/lama*

Q & A