Virtual-Try-On

Rohit Kumar

https://rokmr.github.io/

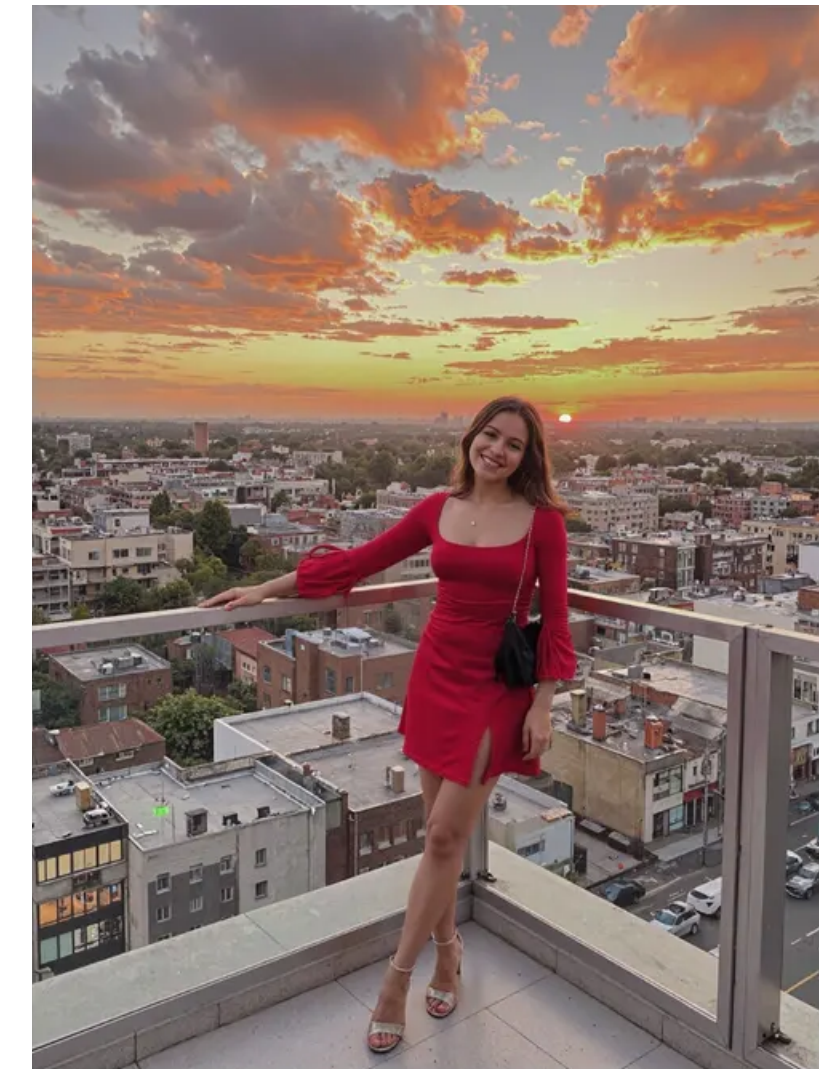# Problem Statement

Target Person

Required Garment

Virtual Try ON

VTON Output

Solving VTON

- It can be seen as a two-step process:

  - Extract the garment from the product image:

    - This can be done using segmentation approaches.

    - Recent text-based segmentation models to extract the required garment image.

  - Inpaint the garment onto the human image through a generative process.

- Attempted solving this task in different ways, referred to as version v0, v1 and v2, each addressing the drawbacks of the previous version.

v0: A simple 2 Stage Framework

- **Goal :** Build a simple pipeline to solve the given VTON task.

- **Process :**

  - **Stage 1**: Extract the **garment** from the **product image**.

    - Used Florence2[1], a conditional segmentation model to segment out the product.

  - **Stage 2:**

    - Explore most recent VTON models: **IDM-VTON**[2], **StableVITON**[3]

    - **IDM-VTON: Takes** Human and garment image as input and gives the desired VTON output.

[1] https://arxiv.org/pdf/2311.06242
[2] https://arxiv.org/pdf/2403.05139
[3] https://arxiv.org/pdf/2312.01725

Target Person

VTON Output

Upper-Body

IDM-VTON

Required Garment

Segmented garment

Florence2

T-shirt

vO: Test Sample 1

Target Person
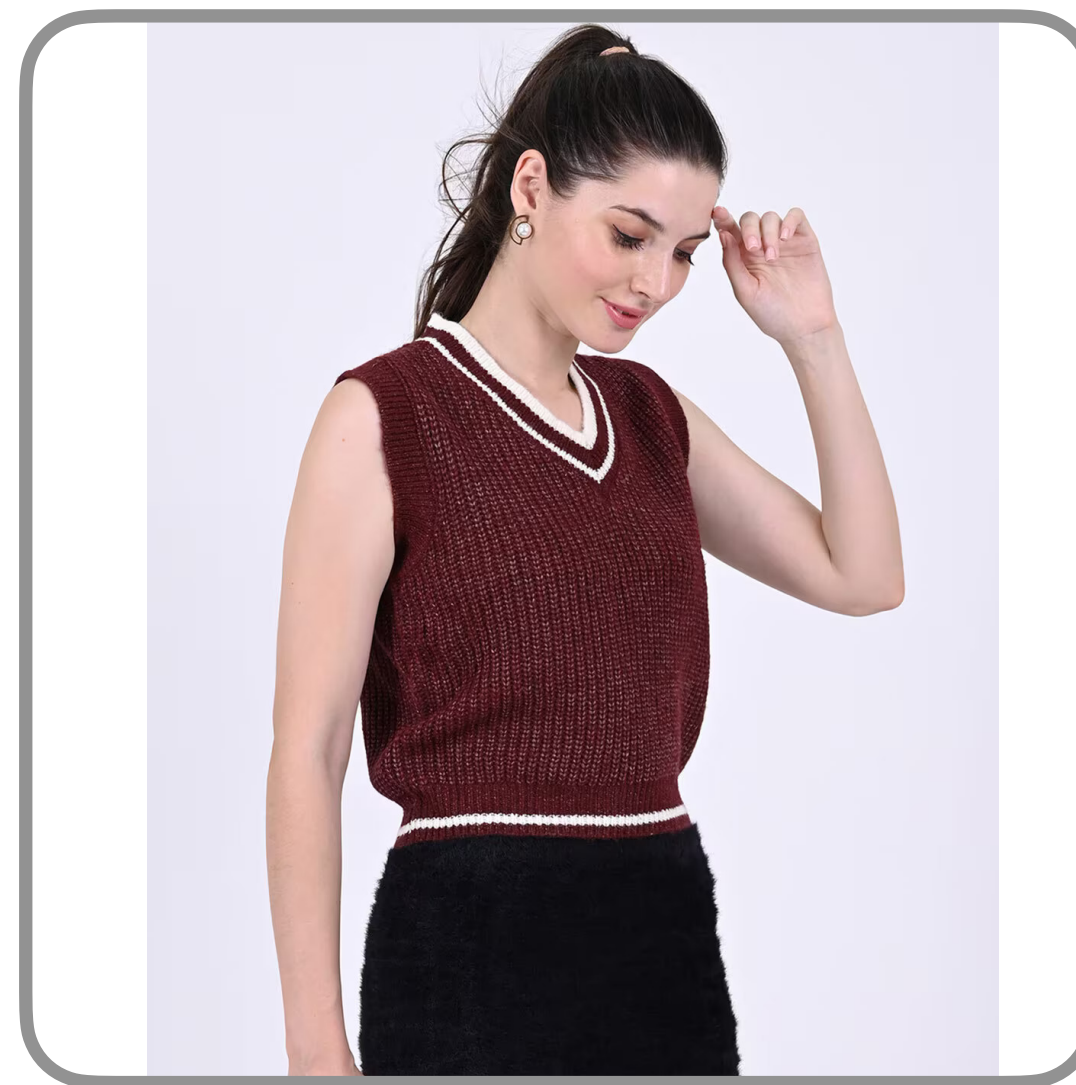
Required Garment

VTON Output

vO: Test Sample 2

Target Person

Required Garment

VTON Output
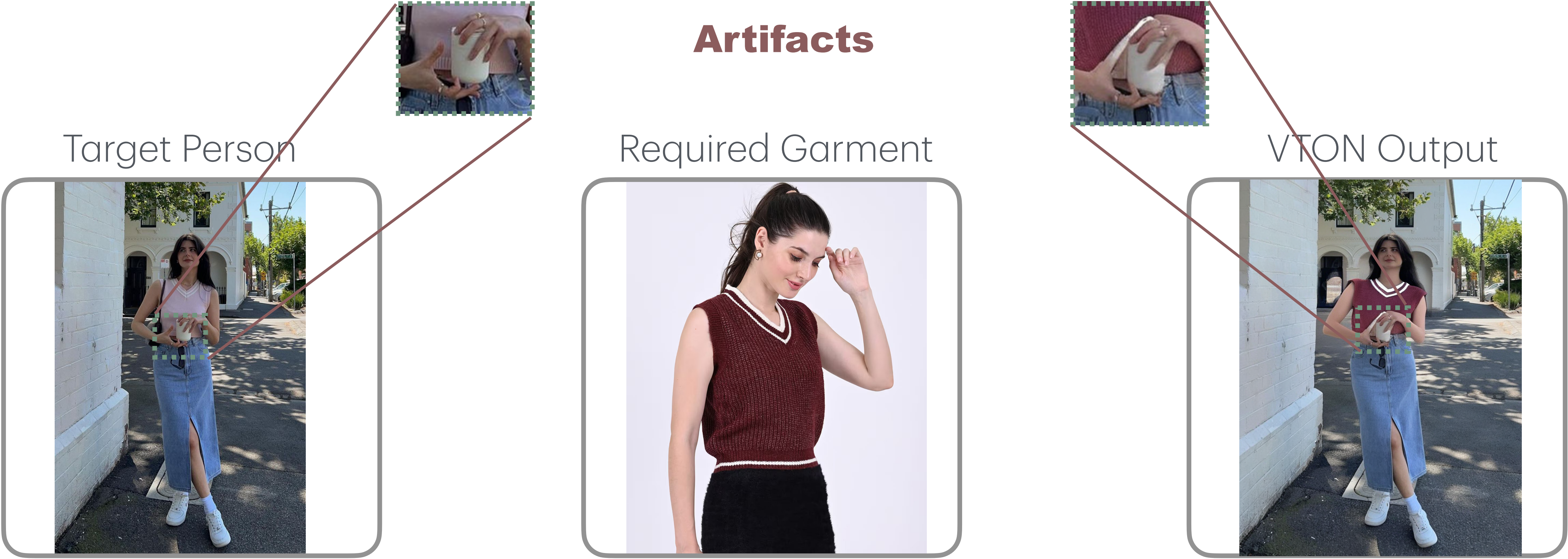
vO: Test Sample 3

Target Person

Required Garment

VTON Output

Failure Cases for v0

1. **Color mismatch** between garment and VTON Output.

2. **Generative Artifacts** introduced when there is occlusion in Person image.

3. **Inconsistent Patterns** between Garment image and the VTON Output.

Generative Artifacts

Artifacts

Target Person

Required Garment

VTON Output

Color Mismatch

Color Mismatch

Target Person

Target Person

VTON Output

Inconsistent Patterns

Color Mismatch

Inconsistent Pattern

Target Person

Required Garment

VTON Output

Addressing Failure Cases: Towards v1 !

- Here, the major issue is the inconsistency in the color and pattern of the dress.

- IDM-VTON is trained on the dataset of garment images which are front facing.

- Garment image extracted from the human image is not always front facing.

- IDM-VTON can fail in such scenarios.

- Can be addressed by transforming the extracted garment image to front view through a geometric transform or a generative model.

- Geometric transform cannot not handle occlusions in the garment image.

- Used CAT-TryOff from [1] model to transform garment image to front view always and then perform VTON using IDM-VTON.

[1] https://github.com/Zheng-Chong/CatVTON

v1: A 3-Stage Framework

Target Person

VTON Output

Upper-Body

IDM-VTON

Required Garment

Segmented

Segmented

Florence2

CAT-TRYOFF

T-shirt

v1: A 3-Stage Framework

Preprocess

Required Garment

Segmented

Output

Florence2

CAT-TRYOFF

T-shirt
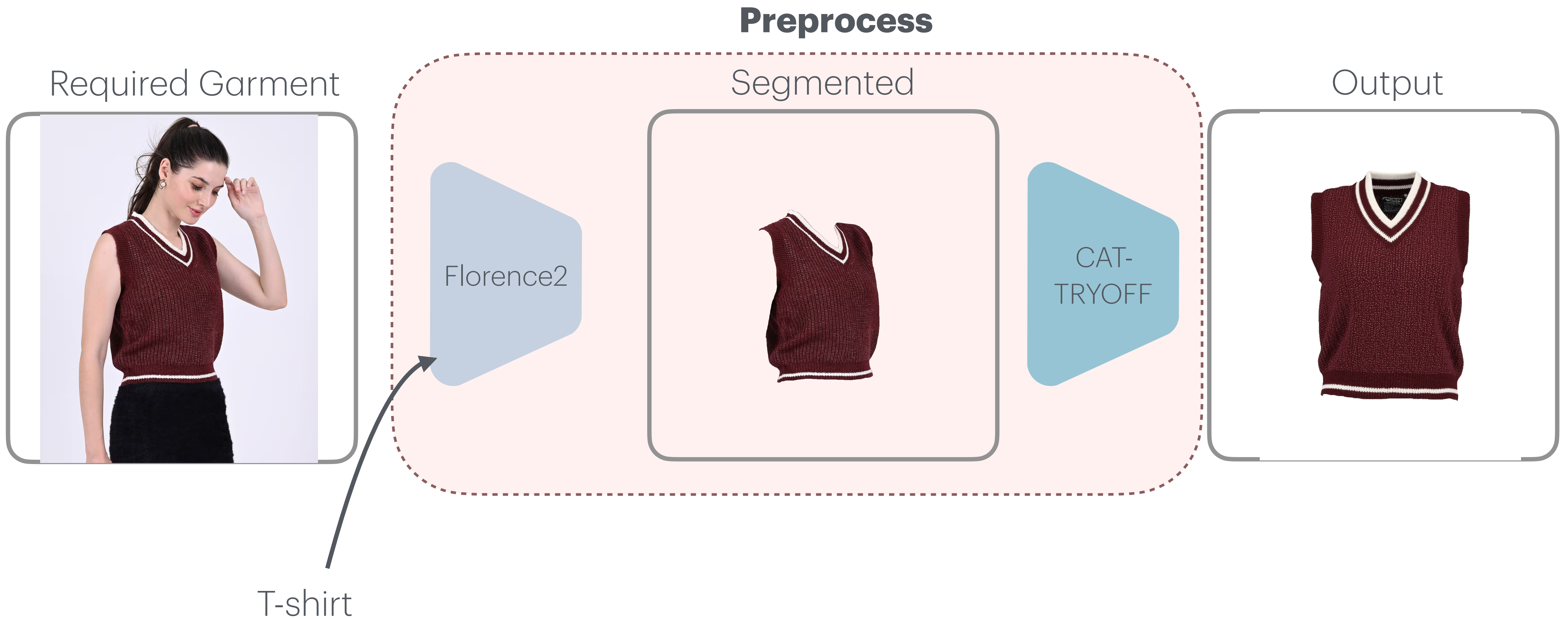
v1: Test Sample 1

Target Person

Required Garment

VTON Output

v1: Test Sample 2

Target Person

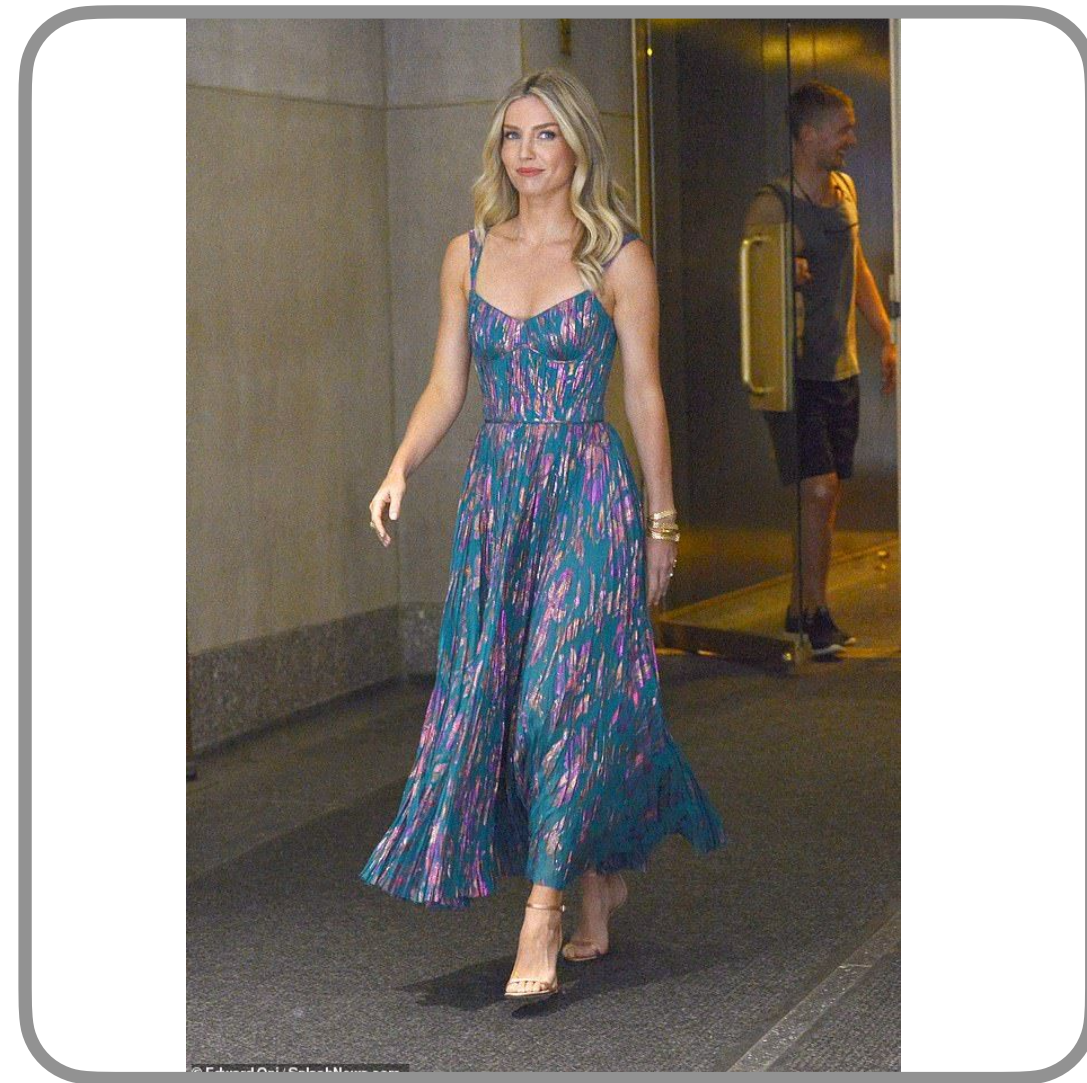### Required Garment

VTON Output

v1: Test Sample 3

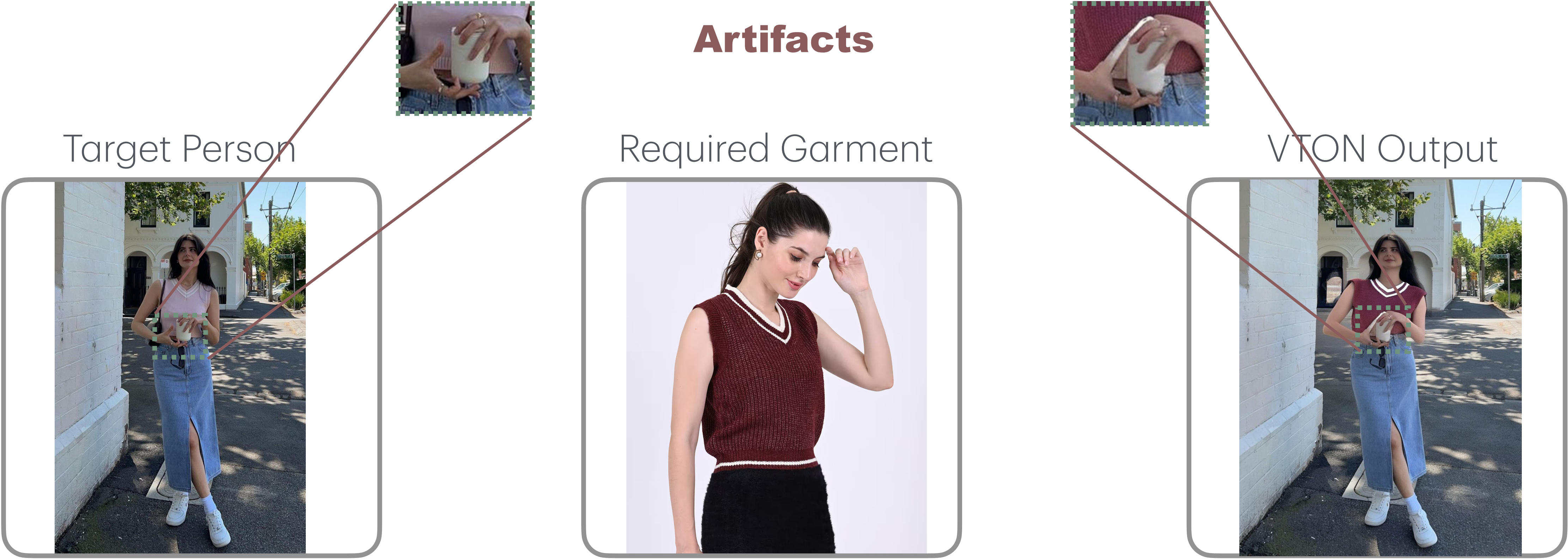Target Person

Required Garment

VTON Output

Better Color & Texture Consistency

Failure Cases for v1

1. **Latency:** Since it is 3-Stage process, it requires multiple models and hence increasing the inference time.

2. **Generative artifacts remain:** Due to incorrect initial segmentation of the person, generative artifacts from occlusion in the target human image remain.

3. **Garment fitting** onto the target person is not always right due to the difference in the size of extracted garment image extracted or variations in human physique as well.
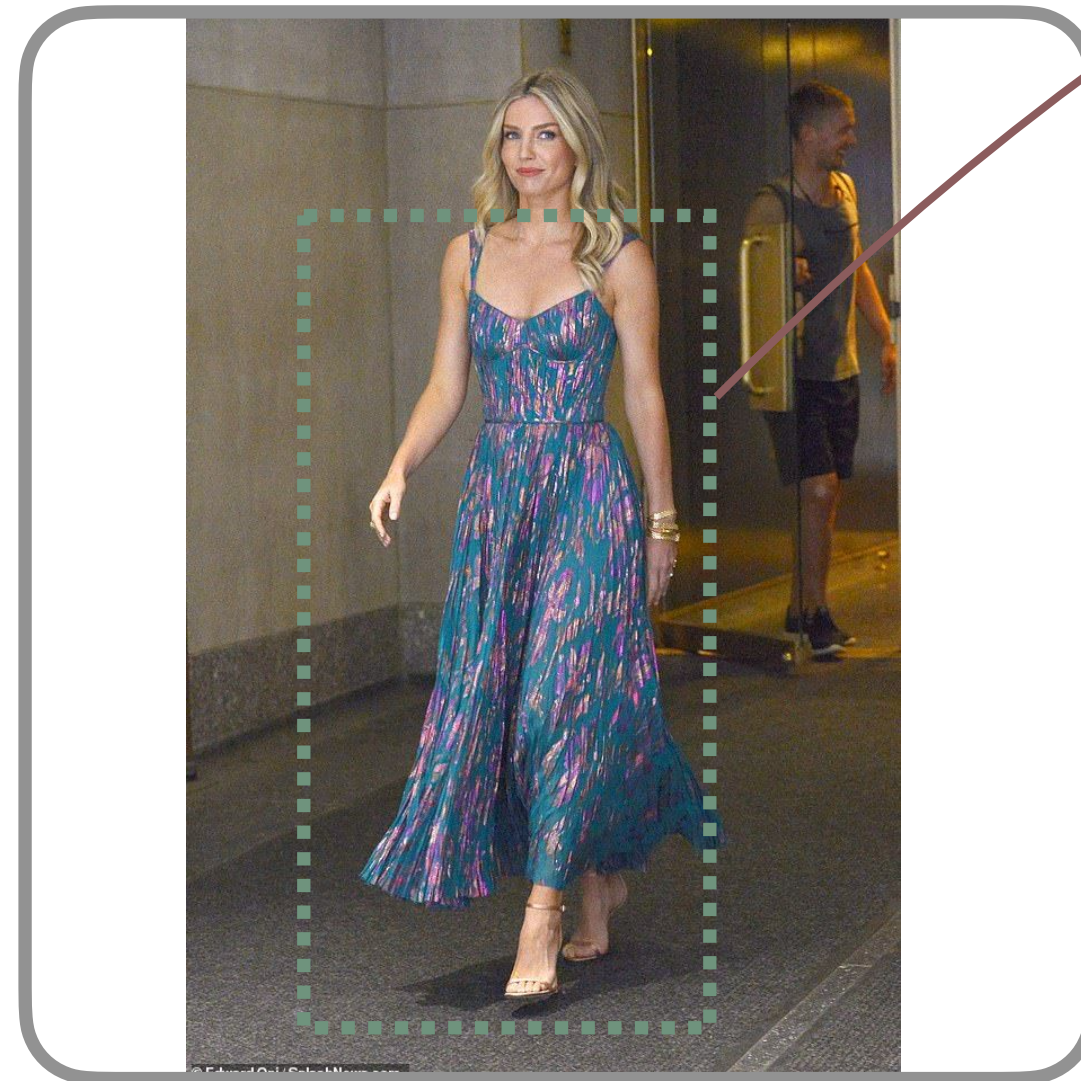
Generative Artifacts remain

Artifacts

Target Person

Required Garment

VTON Output

Garment fitting

Fitting Issue

Target Person

Required Garment

VTON Output

What are the drawbacks? A step towards v2 !

- A primary drawback is latency, which can be resolved if the whole process can be done in a **single stage.**

- Need a method that takes the garment human image and the target human image and gives the final VTON output with the required garment on the target human image.

- MFP-VTON[1] and Any2Any-Tryon[2] solves similar problem.

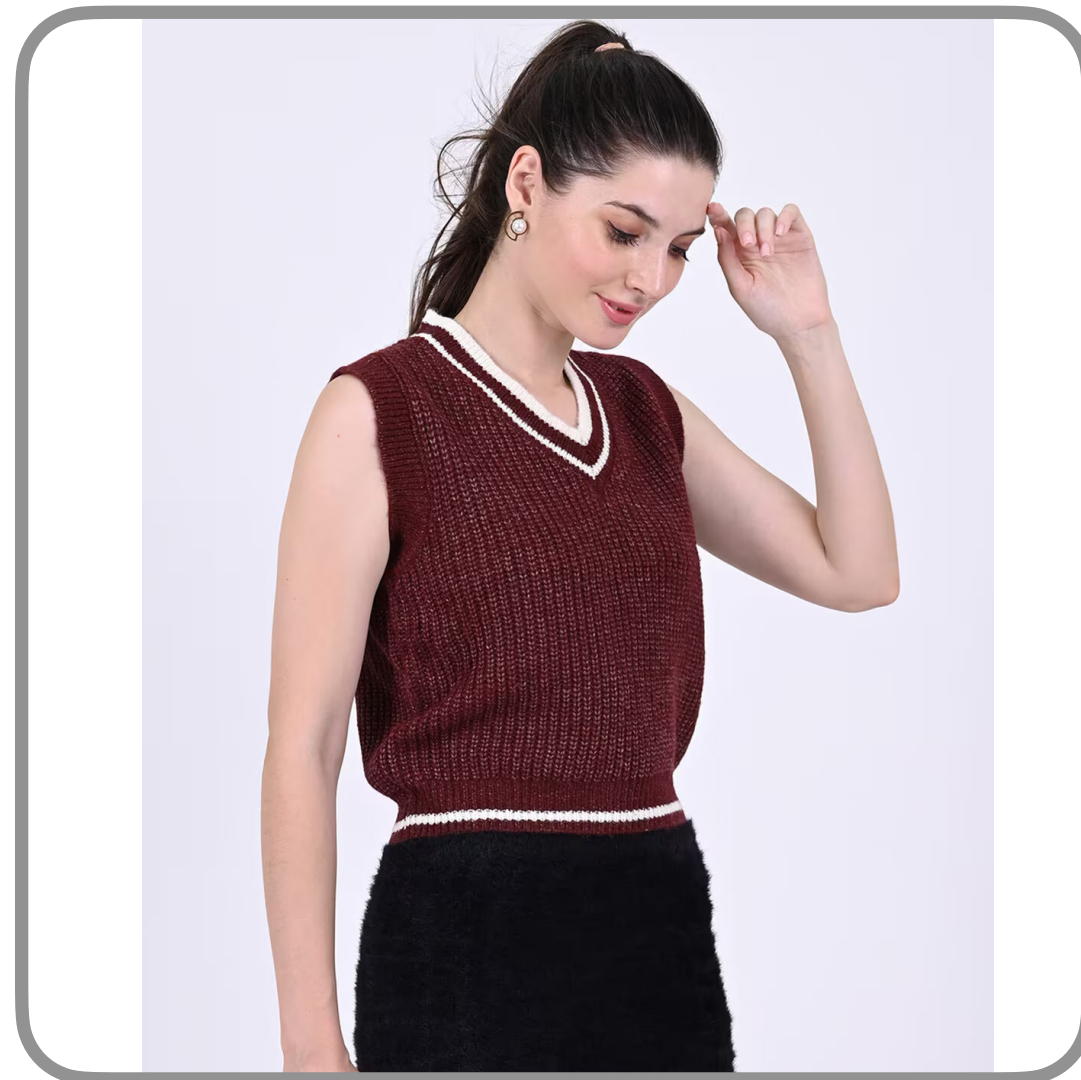- Any2Any-Tryon[2] is FlUX- based method which uses LoRA for the problem.

[1] https://arxiv.org/pdf/2502.01626

[2] https://logn-2024.github.io/Any2anyTryon/
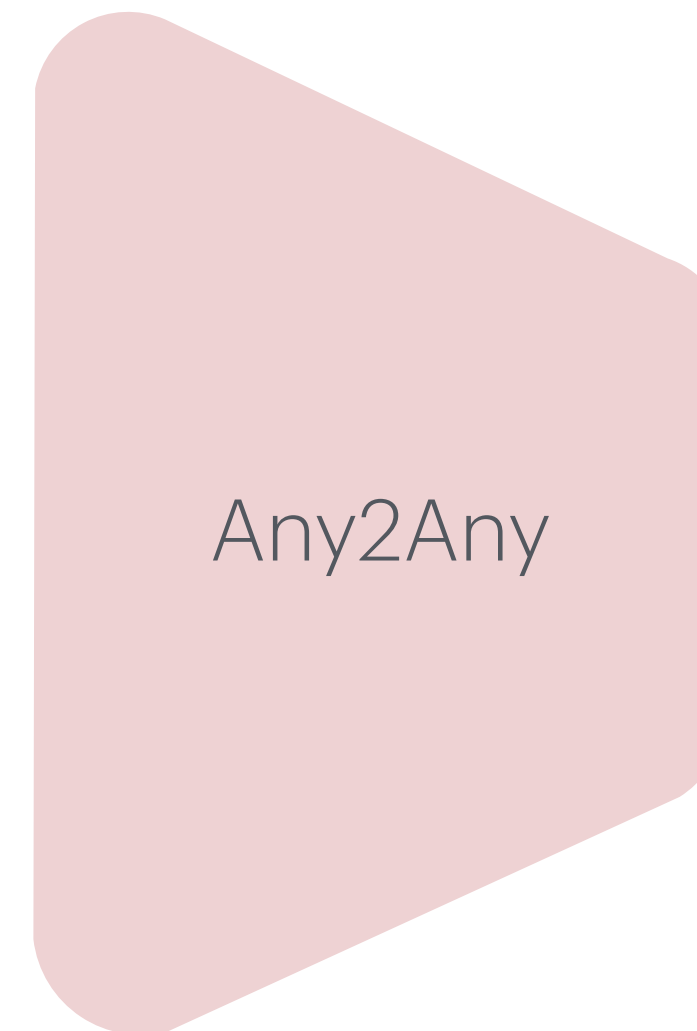
v2: A single Stage Framework

Target Person

Required Garment

T-shirt

Any2Any

VTON Output

v2: Test Sample 1

Target Person

Required Garment

VTON Output

v2: Test Sample 2

Target Person

Required Garment
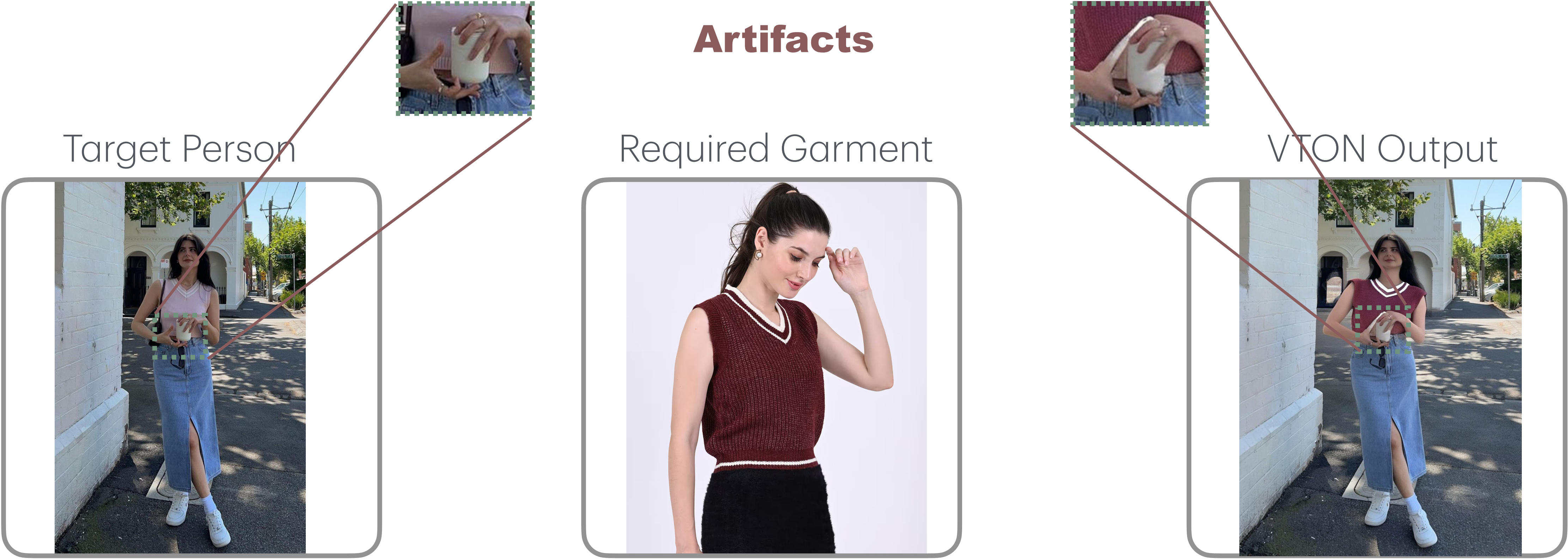
VTON Output

v2: Test Sample 3

Target Person

Required Garment

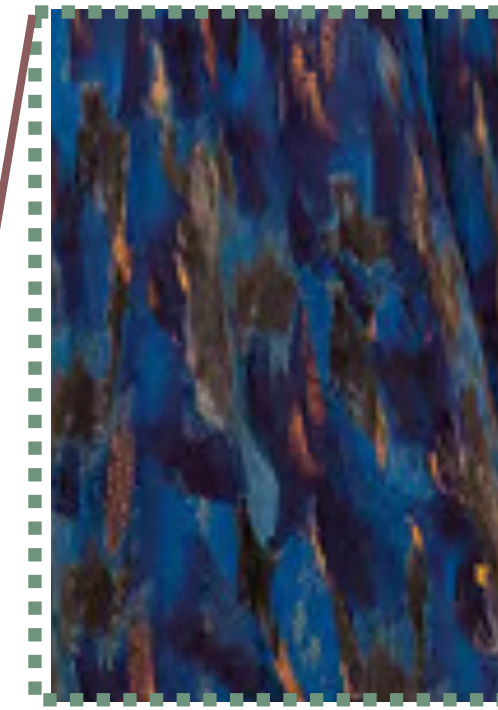VTON Output

Generative Artifacts still remain

Target Person

Artifacts

Required Garment

VTON Output

Inconsistent Patterns still remain

Inconsistent Patterns

Target Person

Required Garment

VTON Output

Conclusion

1. Existing models like IDM-VTON provide very impressive results.

2. While most simple cases can be handled by these SOTA methods, there is still scope for

   Improvement, particularly in avoiding any hallucination due to the generative process.

Future Work

1. Explore ways to improve consistency in textures or patterns while doing VTON.

2. Improving the inference time for real-time execution.