GFA23AI04 - FA23AI34

Virtual Shirt Fitting Model

based on Deep Learning and Computer Vision

@mentoredbyVuHA

AIP490 - FALL 2023

Our Team

Duong Vien Thach
SE150534

Tran Ngoc Xuan Tin
SE150422

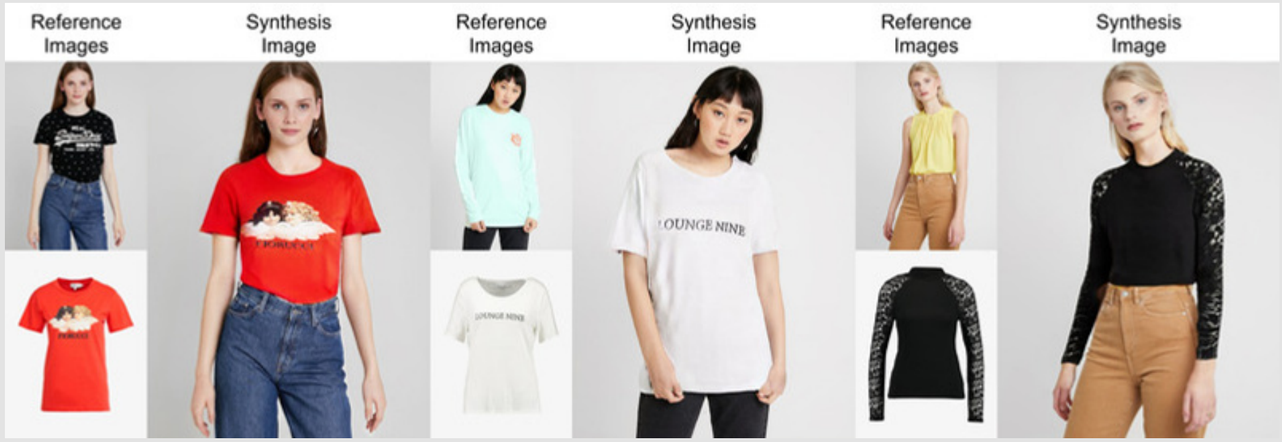
Nguyen Thanh Anh
Duy - SE150090

Our Goals

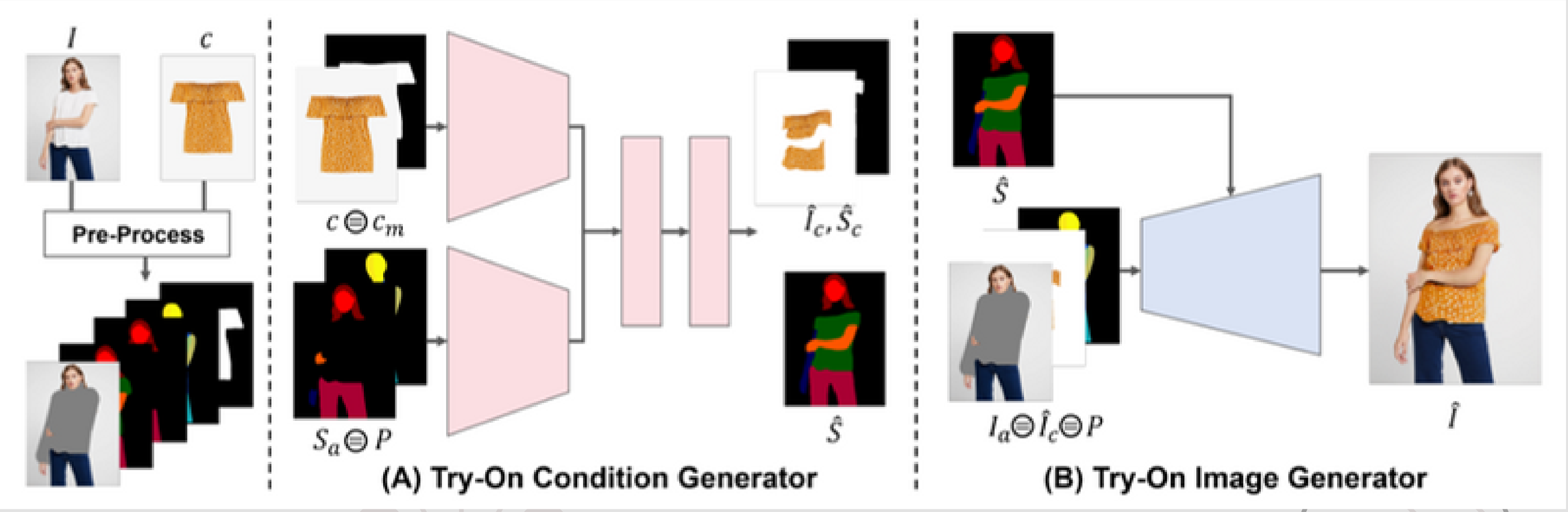
- To develop a model that can accurately identify and track the body's keypoints in different body shapes and sizes. This is important for ensuring that the shirt is rendered in a way that fits the body accurately
- To develop a model that can render the shirt in a realistic way in different poses. This is important for ensuring that the shirt looks realistic in different body positions
- To develop a model that is easy to use and accessible to a wide range of users. This is important for ensuring that the model is widely adopted by consumers

Related Works

After researching the virtual try-on task and consulting relevant works, we have decided to choose HR-VITON as our base model



Overview of the authors' framework

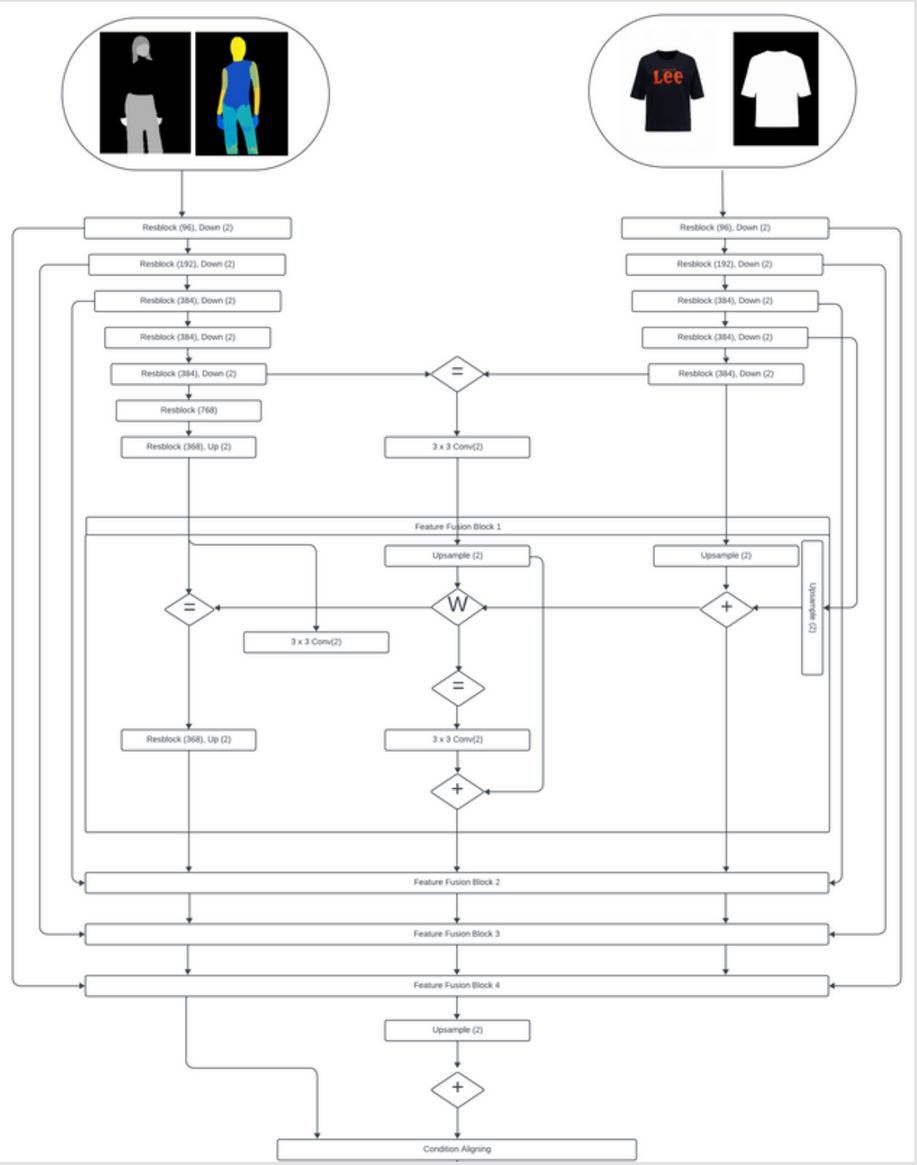


04

HR-VITON

Try-On Condition Generator

A novel component that addresses the limitations of previous virtual try-on methods. It effectively handles misalignments and occlusions between the clothing item and the person's body, leading to more realistic and accurate try-on results



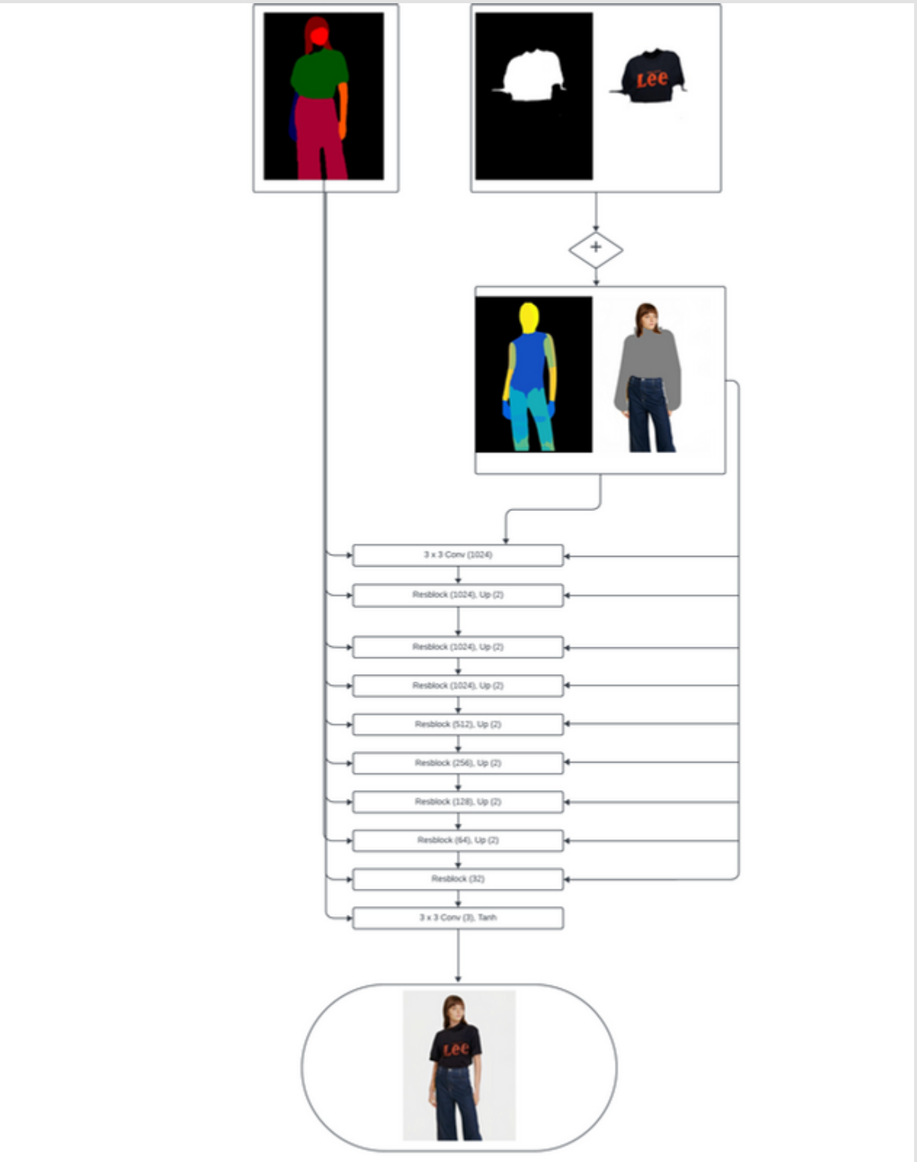
HR-VITON

Try-On Condition Generator

A novel component that addresses the limitations of previous virtual try-on methods. It effectively handles misalignments and occlusions between the clothing item and the person's body, leading to more realistic and accurate try-on results

Try-On Image Generator

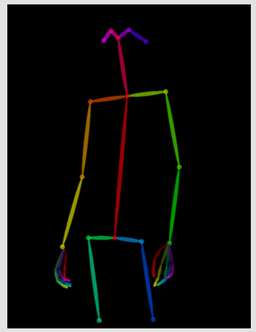
A powerful tool for synthesizing high-quality images of people wearing different clothing items. It is based on the SPADE architecture and uses a spatial attention module to focus on the relevant parts of the image. This allows the generator to produce more realistic images and to better handle misalignments and occlusions



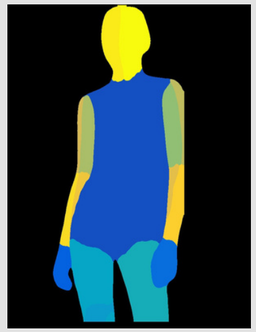
Data Preprocessing

Cloth Mask

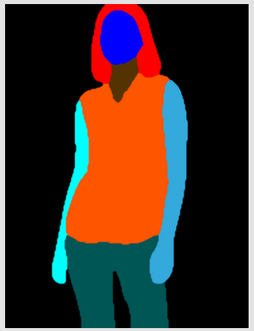
OpenPose



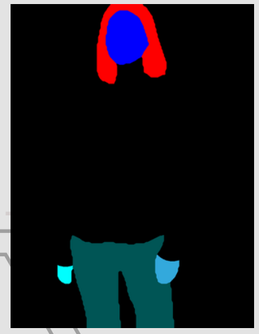
DensePose



Human Parse



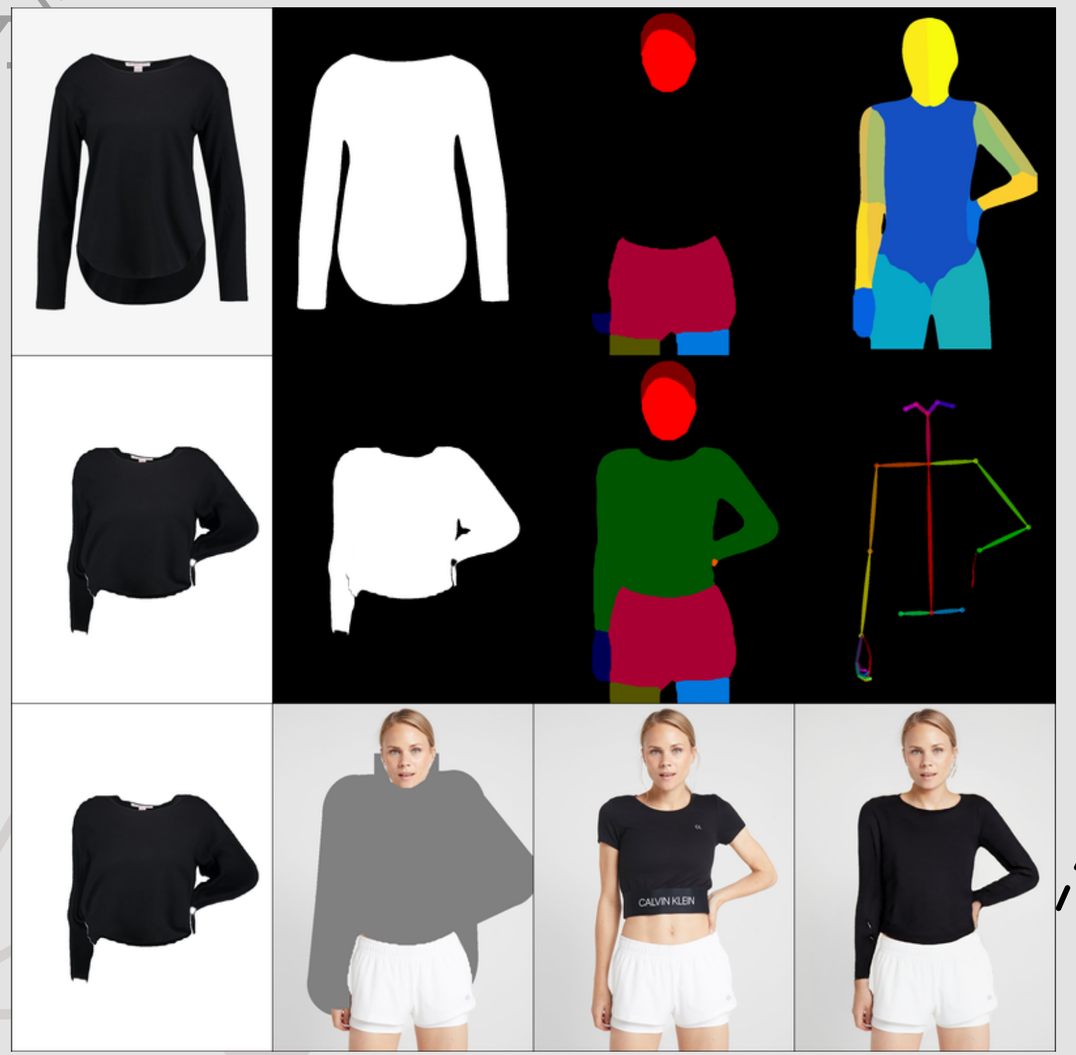
Parse Agnostic



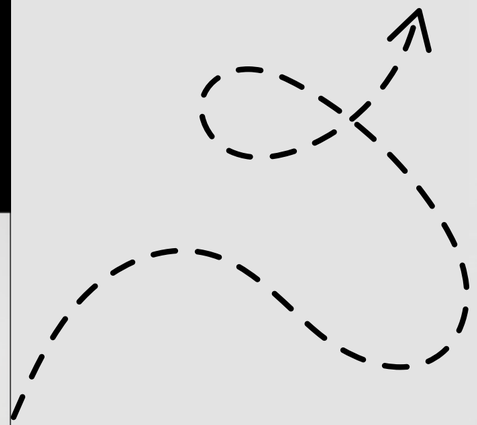
Human Agnostic

Results run on what authors given

Results run on what authors given

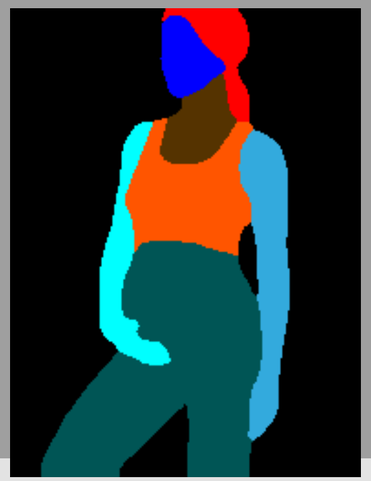


Results run on what authors given



Trying to reproduce data preprocessing

Having bad results at the Human Parse step on 768x1024 images



Details on 192x256 images

768 x 1024

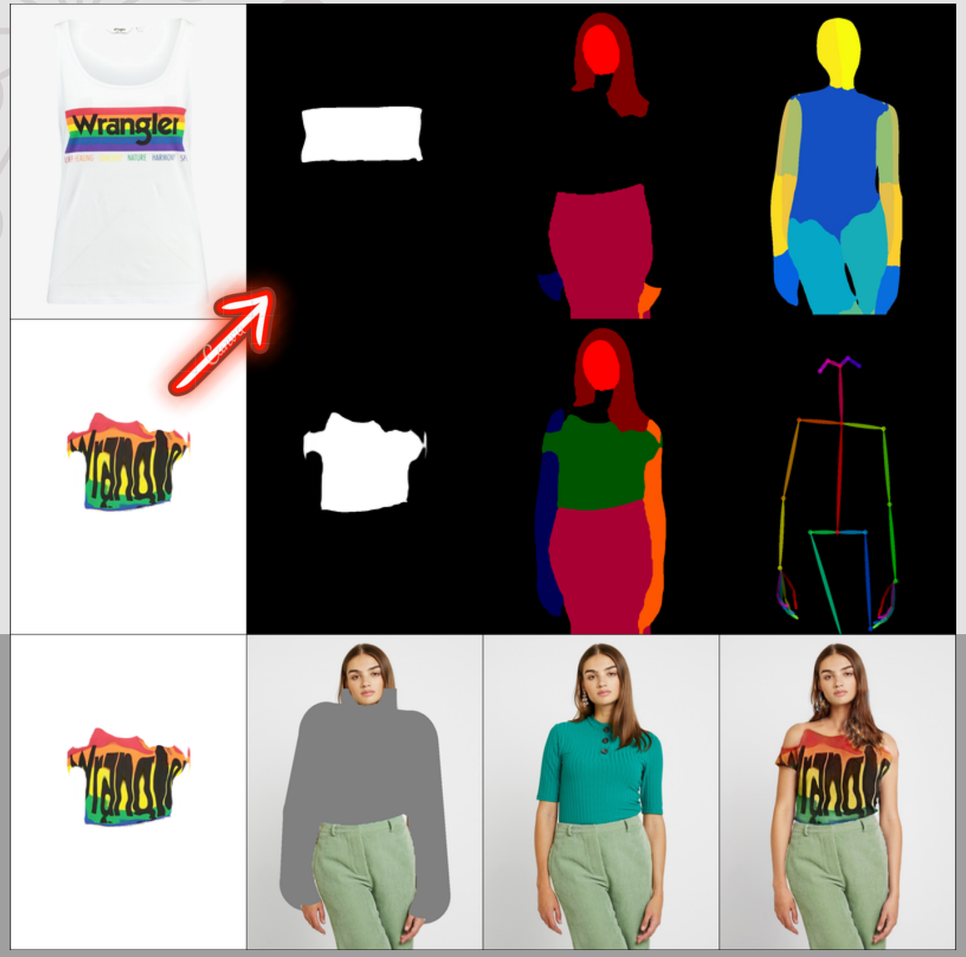
192 x 256

Upscaling

Normal resize methods

Interpolate method

Cloth Mask problem



HR-VITON

Our proposed method

HR-VITON

Our proposed method

HR-VITON

Our proposed method

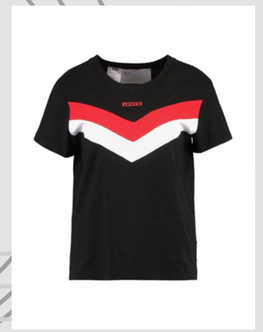
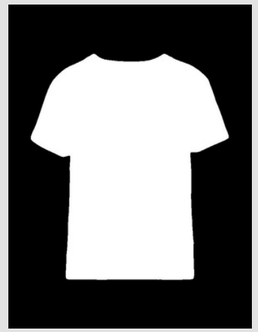
HR-VITON

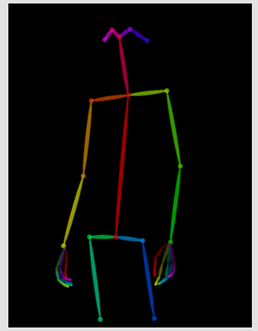
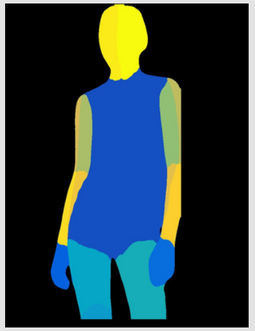
Our proposed method

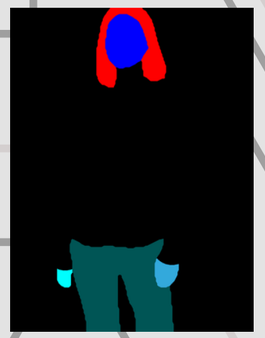
Proposed data preprocessing

According to the explanation from the authors, at least 6 steps are needed for getting all of the required inputs of the model

We’ve added one more preprocessing step and modified other steps in order to get better try-on results

Comparison

HR-VITON

SSIM: 0.825291 / **MSE**: 0.050458 / **LPIPS**: 0.230576
IS_mean: 3.189770 / **IS_std**: 0.000000

Our proposed data preprocessing

SSIM: 0.821920 / **MSE**: 0.051742 / **LPIPS**: 0.229801
IS_mean: 3.227256 / **IS_std**: 0.000000

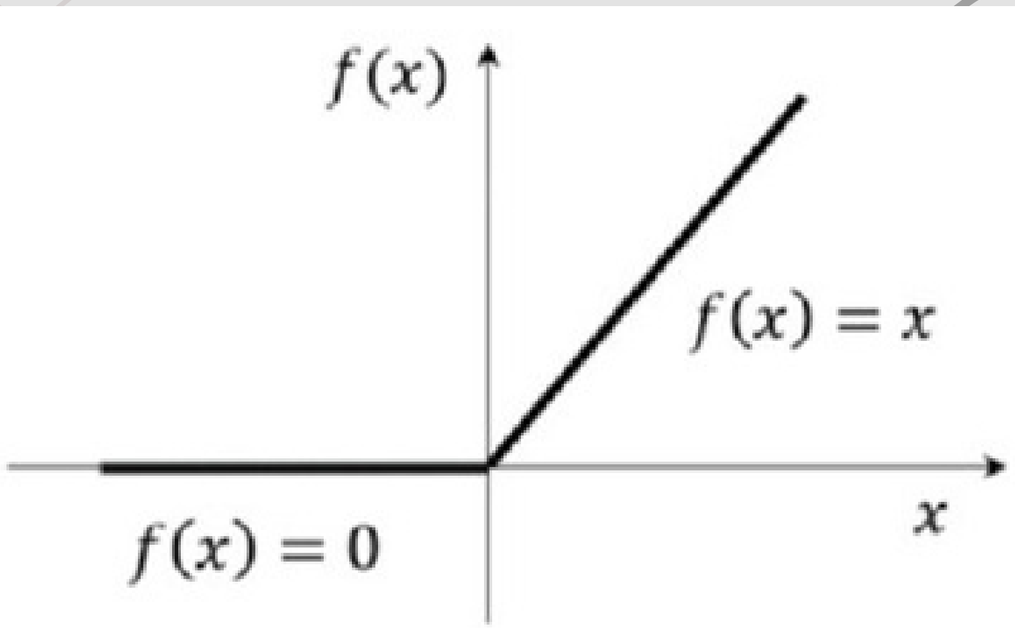
Optimization

We trained with 1000/300000 steps to check which one is the most optimal for the final results

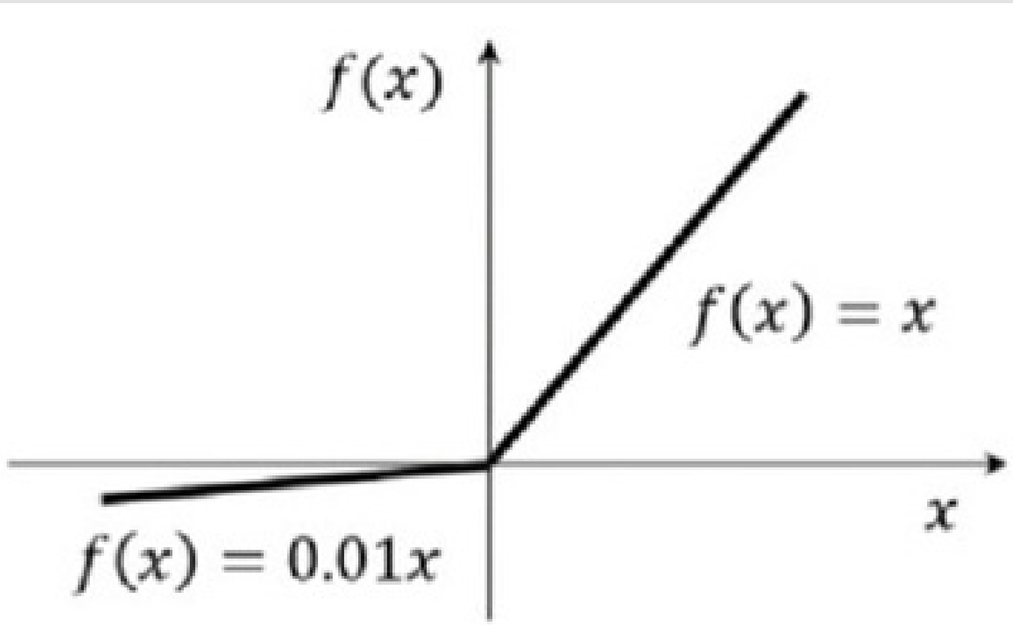
	SSIM	MSE	LPIPS	IS_mean
Base 1000 step	0.815446	0.052806	0.233563	3.195333
conv2d_norm_Relu_conv2d_Relu_norm	0.738314	0.081379	0.341204	3.525081
conv2d_norm_LeakyRelu_conv2d_norm	0.814557	0.051797	0.23412	3.335743
conv2d_norm_LeakyRelu_conv2d_norm_LeakyRelu	0.815976	0.052283	0.231399	3.041033
Tranposconv2d_norm_LeakyRelu_Tranposconv2d_norm	0.77468	0.056094	0.286347	2.35643
Tranposconv2d_norm_LeakyRelu_Tranposconv2d_LeakyRelu_norm	0.742297	0.070948	0.353888	3.369364
Tranposconv2d_norm_LeakyRelu(0.01)_Tranposconv2d_norm	0.775967	0.056168	0.286394	2.361891
Tranposconv2d_norm_LeakyRelu(0.01)_Tranposconv2d_LeakyRelu(0.01)_norm	0.775027	0.056589	0.286465	2.375879
conv2d_norm_LeakyRelu_conv2d_LeakyRelu(0.01)_norm	0.815831	0.051446	0.231517	3.159622

Optimization

Why did we add another leakyrelu layer?



ReLU activation function



LeakyReLU activation function

Optimization

We also re-ran the evaluation metrics and got results approximately the same as the author's

STEP	SSIM	MSE	LPIPS	IS_mean
3000	0.812811	0.054265	0.241237	3.108629
6000	0.813482	0.053982	0.239779	3.124469
9000	0.813805	0.053811	0.2389	3.128264
12000	0.812789	0.05404	0.240045	3.265919
15000	0.822069	0.052957	0.229872	3.265919
18000	0.82195	0.053811	0.226971	3.293342
21000	0.815857	0.051024	0.237209	3.294366
24000	0.816077	0.052957	0.236621	3.305738
27000	0.822359	0.053132	0.236827	3.310482
30000	0.822001	0.053142	0.235957	3.311385

Optimization Results

Original

Optimized

Optimization Results

Original

Optimized

Optimization Results

Original

Optimized

Optimization Results

Original

Optimized

Optimization Results

Original

Optimized

Conclusion

HR-VITON

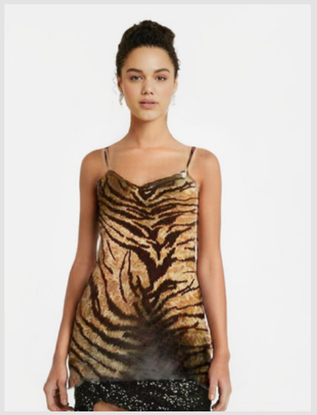
- No specific data preprocessing instructions are provided
- Some cases get bad results on Cloth Mask preprocessing step

Our researches

- Figured out how to preprocess data step by step
- Added background removal for cloth and image
- Fixed Cloth Mask problem by using another method

Model's restrictions

- Image of person does not show both arms cause bad generated result
- Cannot handle the length of the try-on shirt's shape (base on person's shirt)
- Only focus on upper clothes so can't try-on dress, jacket,...



AIP490 - FALL 2023

GFA23AI04 - FA23AI34

THANK YOU

@mentoredbyVuHA