Deep Fashion

Group 11

Motivation

The most important question while shopping for clothes is- "How it looks on me?". While shopping in a store, we can just put the clothes on, but this is not possible when shopping online.

We can only see how the model looks wearing the dress of our choice, but still the question remains unanswered- "How it looks on me?". To see that we have to order the dress & then try it on, it takes time & it’s not efficient.

We address this problem with our project- **Deep Fashion**

Problem Statement

We solve the problem- To see how we look in clothes of our choice while shopping online so that shopper can make much more informed decision.

Computer Vision Task-

Given an image of a model wearing the garment & the image of the shopper, generate a video of the shopper wearing the garment doing a catwalk.

Literature Review

This topic has been approached in several different ways, some of which are discussed below :

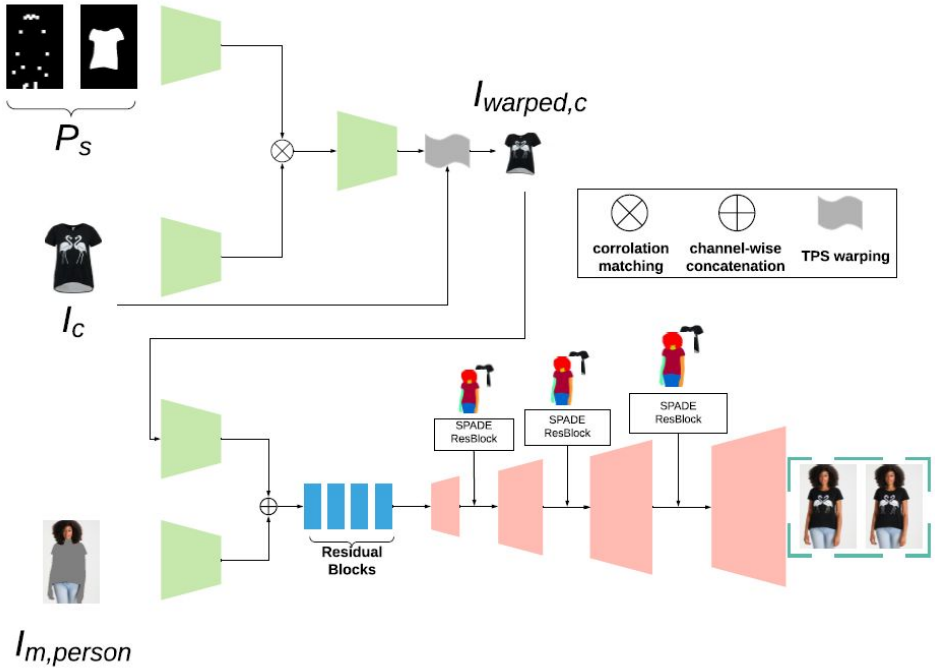
Neural Re-Rendering of Humans from a Single Image [link]

The novel algorithm represents body pose and shape as a parametric mesh which can be reconstructed from a single image and easily reposed. Instead of a colour-based UV texture map, this approach further employs a learned high-dimensional UV feature map to encode appearance. This rich implicit representation captures detailed appearance variation across poses, viewpoints, person identities and clothing styles better than learned colour texture maps. The body model with the rendered feature maps is fed through a neural image-translation net-work that creates the final rendered colour image. The above components are combined in an end-to-end-trained neural network architecture that takes as input a source person image, and images of the parametric body model in the source pose and desired target pose.

Literature Review

GarmentGan: Photo-realistic Adversarial Fashion Transfer [link]

GarmentGan breaks the problem in two parts, separating person's body part from their clothing, and the second part is to generate images with new clothes on. This paper uses GANs to perform image-base garment transfer. It requires two input images, target fashion item an an image of customer in the pose they want.

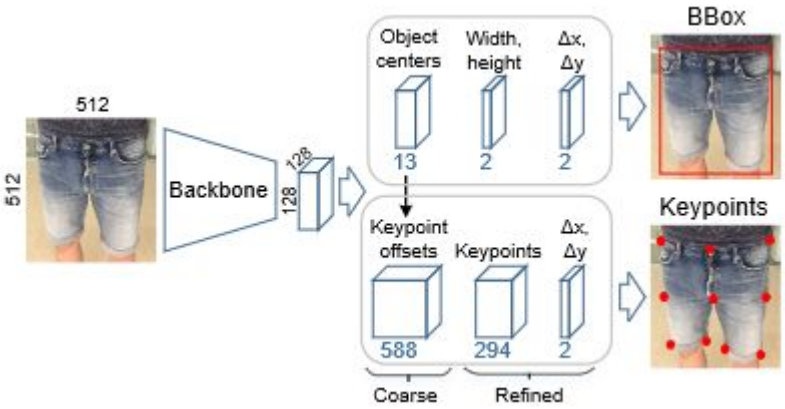


Literature Review

DeepMark: One-Shot Clothing Detection [link]

The one-shot approach, DeepMark, for fast clothing detection as a modification of a multi-target network, Center-Net , is proposed in the paper.

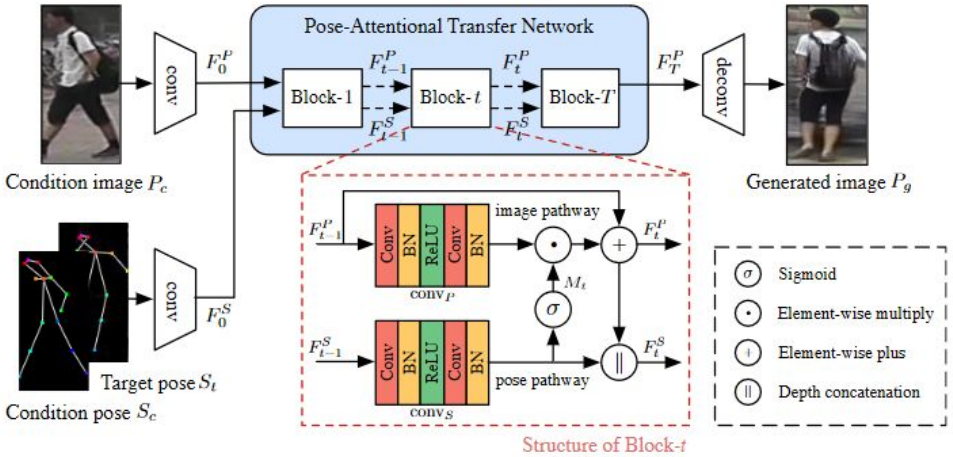
This paper uses CenterNet as a backbone and the deepfashion2 dataset includes 13 categories each of which contains a different number of keypoints (294 key-points total). This approach is based on CenterNet topology but handles two above mentioned restrictions with ease.



Literature Review

Progressive Pose Attention Transfer for Person Image Generation [link]

This paper proposes a new generative adversarial net-work for pose transfer, i.e., transferring the pose of a given person to a target pose. The generator of the net-work comprises a sequence of Pose-Attentional Transfer Blocks that each transfers certain regions it attends to, generating the person image progressively.



Dataset Description

Dataset that will be used for this project is DeepFashion2. It was compiled by the research team from the Chinese University of hong Kong(CUHK).

It consist of over 492k fashion images and each image is labeled into 13 popular clothing categories. The dataset is split into a training set (391K images), a validation set (34k images), and a test set (67k images). The dataset is well annotated. A full spectrum of task like clothes detection, landmark, segmentation are defined.

Dataset Description

Category_name : a string which tells about the category of the item like short sleeve top, skirt etc.

Category_id : a number corresponding to the category name

Bounding_box : separate bounding box for each clothing item

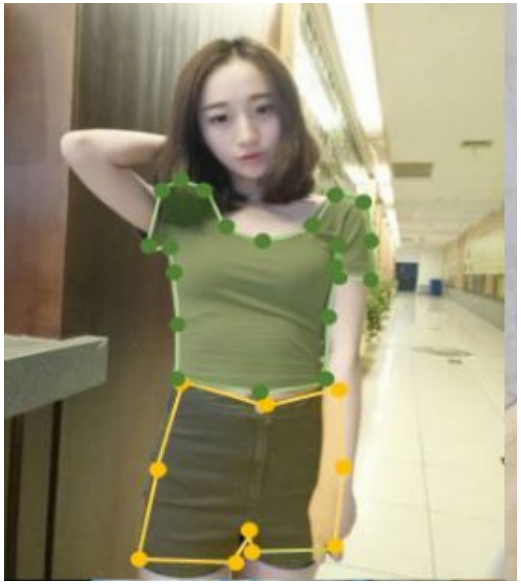
Landmarks : landmarks of clothing items with its corresponding visibility

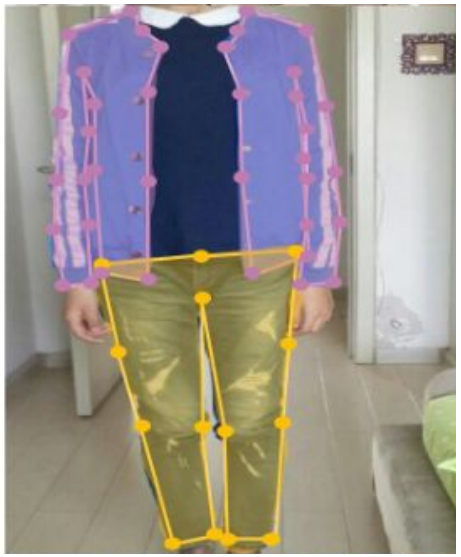
Segmentation : It describes a polygon which represents a single clothing item

Occlusion : A value from 1 to 3 is allocated based on occlusion of the clothing items

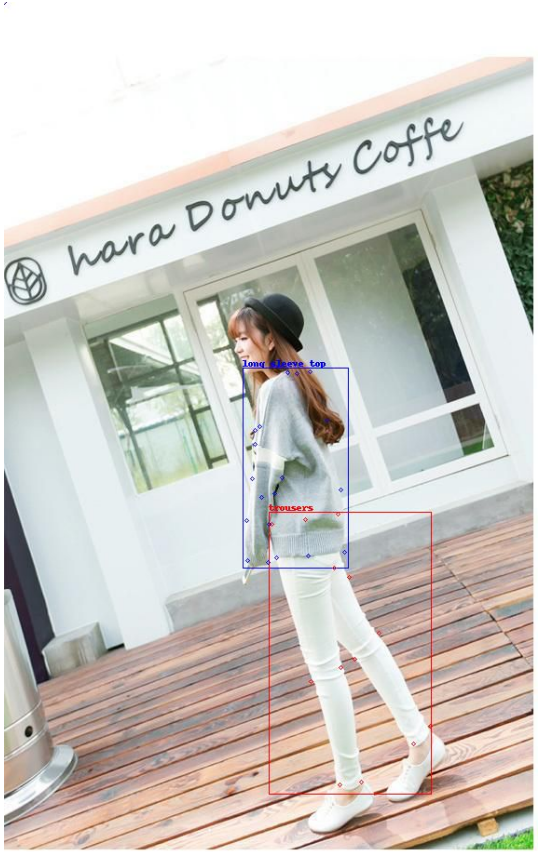
Viewpoint : A number, where 1 represents no wear, 2 represents frontal viewpoint and 3 represents side or back viewpoint.

Dataset Description





Dataset Description



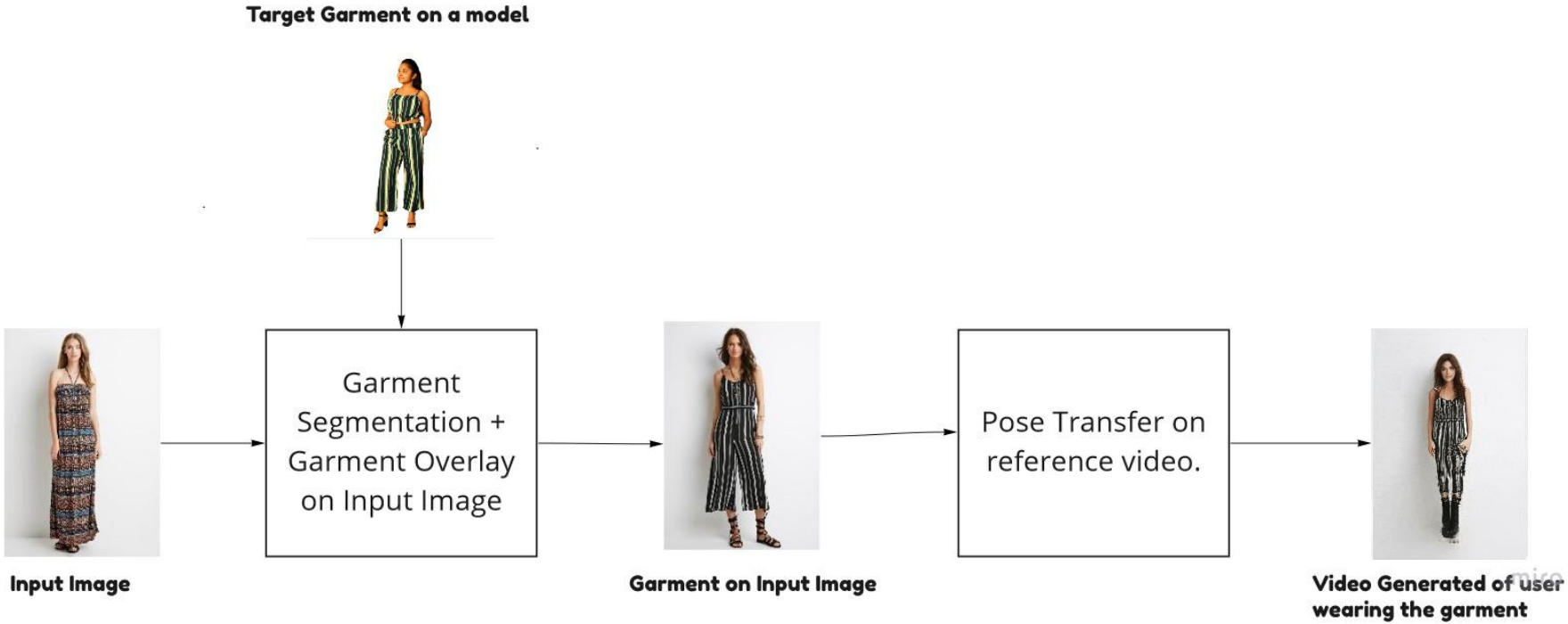
Methodologies

The problem is divided into 4 stages to make a end to end system.

1. Garment extraction from Target Model.
2. Garment Transfer from Target garment to Input Image.
3. Pose Transfer from Input Pose to Target Pose
4. Video Generation

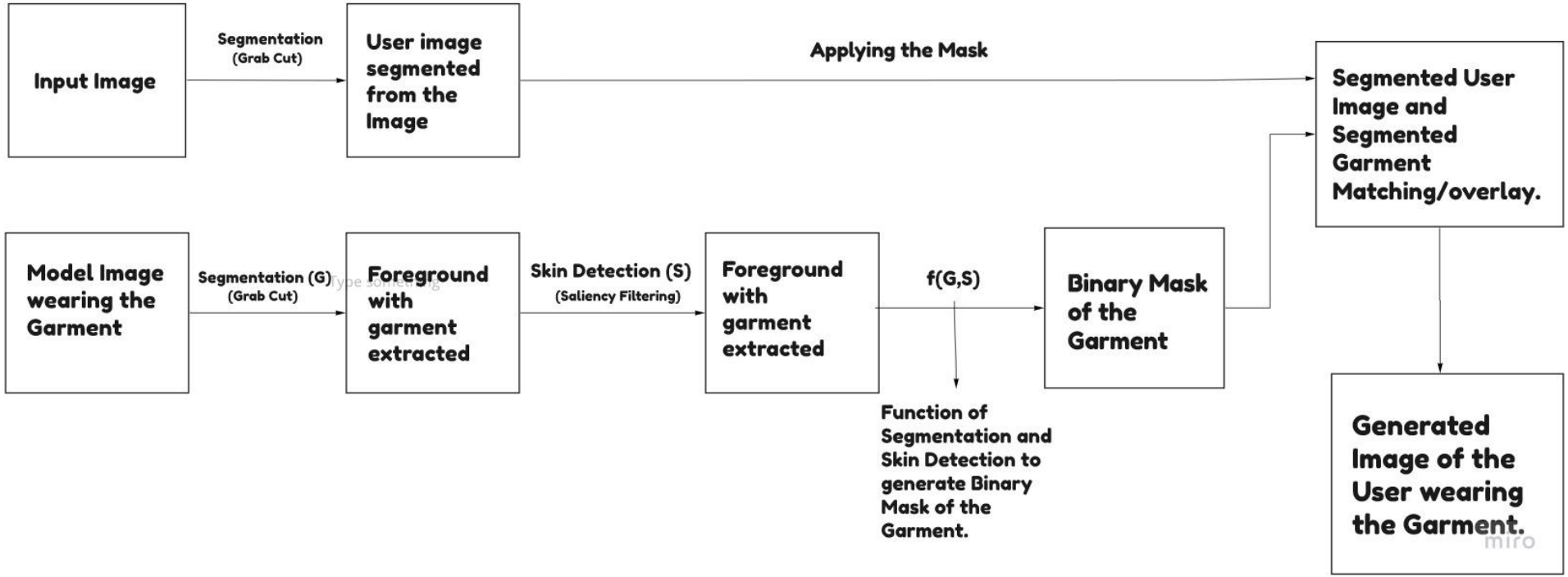
Methodologies

- Model Pipeline



Methodologies

- Garment Segmentation and Garment Transfer from Target garment to Input Image.



Methodologies

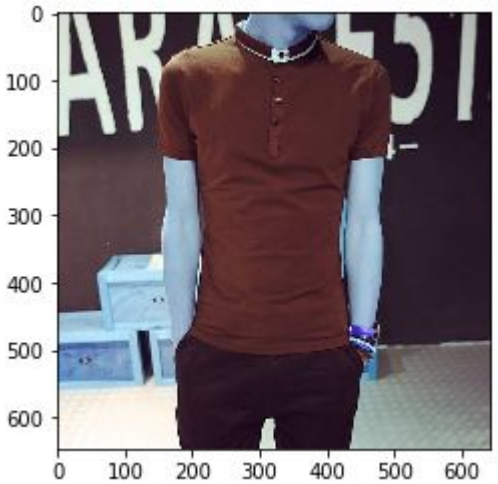
- Pose Transfer and Video Generation.

1. After the Image of user wearing the Target Garment is generated.

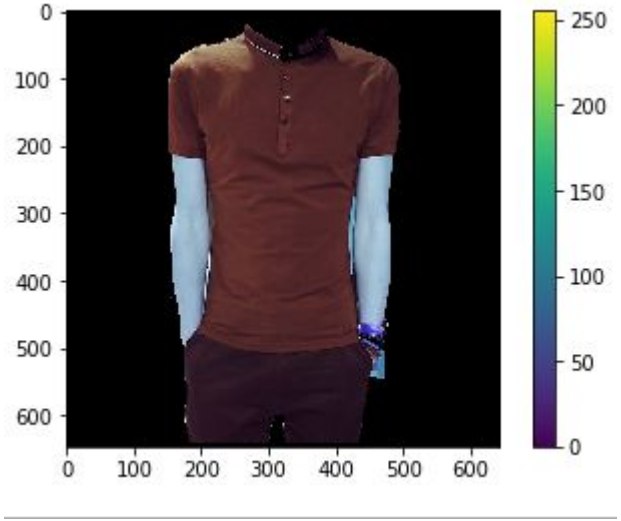
2. The Image is passed on to the Progressive Pose Attention Transfer for Person Image Generation Model to generate the Video of the User.

Preliminary Results

Currently, we have been able to segment the Garment from the Model Image using the Grab Cut Algorithm.



Grab Cut (Segmentation)



Work Left

1. Using the Deep Mark Model to get the individual bounding boxes of the clothing items then applying segmentation on them.
2. Applying Saliency filters for Skin Detection to remove them from the Segmented Garment as shown above.
3. Generation of a binary mask using both Saliency and Segmentation to get the entire garment.
4. Pose-Transfer and Video Generation of the User wearing the garment.

Thank You

Group 11
Arka Sarkar , Adarsh Kumar, Nitin Gupta, Utsav Singla