Deep Learning induced Background Matting

Overview:

The goal of the project was to remove the background (mannequin) from the clothing display images, generated from the client's shoot-data. While many tools now provide background replacement functionality, they also yield artifacts at boundaries, particularly in areas where there are finer details. In contrast, traditional image matting methods provide much higher quality results, but do not run at high resolution, and frequently require manual input (Trimaps).

Hence, in order to achieve this ghost mannequin effect on the clothing, we used a technique based on background matting. In this approach, an additional frame of the background is captured and used in recovering the alpha matte and the foreground layer. The objective is to compute a high-quality alpha matte, preserving the edge details, while processing high-resolution images specified by the client.

Client details:

Name: Confidential | **Industry**: Publishing | **Location**: USA

Technologies:

Deep Learning, PyTorch, Image Matting, Morphological Operations (OpenCV), Pandas, NumPy, Matplotlib, CUDA, Streamlit (frontend), Python (Uvicorn)

Deep Learning induced Background Matting

Project Description:

The technique used in the project is the first fully-automated, high-resolution matting technique, that produces state-of-the-art results at 4K (3840×2160) at 30fps and HD (1920×1080) at 60fps. The method relies on capturing an extra background image to compute the alpha matte and the foreground layer, an approach known as background matting.

To achieve this feat, we employ two neural networks:

- a base network that computes a low-resolution result
- a second network which refines the result by operating at high-resolution on selective patches.

In order to design a network that can handle high-resolution images, fine-grained refinement is targeted at relatively few regions in the image. Therefore a combination is used of a base network, which predicts the alpha matte and foreground layer at lower resolution, along with an error prediction map which specifies areas that may need high-resolution refinement. A refinement network then takes over the low-resolution result and the original image.

There are two large-scale video and image matting datasets: VideoMatte240K and PhotoMatte13K/85. This yields higher quality results, while simultaneously yielding a dramatic boost in both speed and resolution.

Input:

Background Images (4480 x 6720)

Deep Learning induced Background Matting

Output:

In addition to this, an alpha channel is also outputted for the cutout.

Dataset used for training the Background Matting Model:

Since it is extremely difficult to obtain a large-scale, high-resolution, high-quality matting dataset where the alpha mattes are cleaned by human artists, publicly available datasets were used.

Publicly Available Datasets: (we have to contact/mail the authors to get access though)

- The Adobe Image Matting (AIM) dataset provides 269 human training samples and 11 test samples, averaging around 1000×1000 resolution.
- A humans-only subset of Distinctions 646 consisting of 362 training and 11 test samples, averaging around 1700×2000 resolution.

The mattes were created manually and are thus high-quality.
However 631 training images are not enough to learn large variations in human poses and finer details at high resolution, so 2 more additional datasets were introduced

VideoMatte240K (publicly available)**:** 484 high-resolution green screen videos and generate a total of 240,709 unique frames of alpha mattes and foregrounds with chroma-key software Adobe After Effects.

Deep Learning induced Background Matting

PhotoMatte13K (not available publicly due to privacy and licensing issues)**:** A collection of 13,665 images shot with studio-quality lighting and cameras in front of a green-screen, along with mattes extracted via chroma-key algorithms with manual tuning and error repair.

Custom Training Timeline:

- We used the pre-trained checkpoint that was trained using the above datasets from generating the alpha masks and foregrounds (3D clothing).
- To further improve the output quality, we fine-tuned by training the network further using client foreground - background pairs for about 2-3 months till we got better results.
- We also added post processing techniques using OpenCV (morphological techniques) to further improve the outputs.
- Hosted a production ready Web Application on client's local network with simple UI where high resolution foreground and background images are pointed to the backend, thus generating the clothing , removing the mannequin.
- Updated the Web Application to cater batch processing.
- Added more post processing techniques to the generated outputs using Photoshop automation.

Network Architecture

It is a deep learning based model with 2 neural networks in it.
The model is trained on 3 datasets containing high resolution images and videos
The model basically takes in 2 inputs for inference -

- Image with foreground and the background
- Image with only background aka casually captured background

They call it

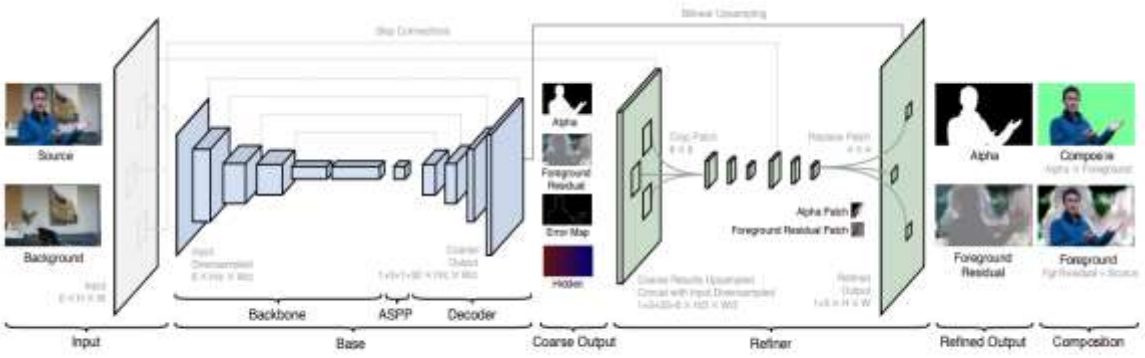

Figure 3: The base network G_{base} (blue) operates on the downsampled input to produce coarse-grained results and an error prediction map. The refinement network G_{refine} (green) selects error-prone patches and refines them to the full resolution.

casually captured background as an estimation of true background because it can contain slight movements, color differences, slight shadows and similar colors as the foreground (not same though)

Reference : https://arxiv.org/pdf/2012.07810.pdf