FASHION OUTFIT GENERATOR

Submitted By:

Tanishq Kakkar (2020UCP1778)
Rahul Guroo (2020UCP1124)

Under the Guidance of
Dr. Dinesh Gopalani, Associate Professor

Dept. of Computer Science and Engineering
Malaviya National Institute of Technology, Jaipur

Overview

Introduction

Problems Faced

1. Limited number of Outfits
2. Unable to customize according to our need
3. Adjust with the selected outfit

Solution

1. Ask user for the details of the outfit.
2. Prepare outfit according to user specifications (Customized outfit)

- A text-to-image model which generates an outfit according to user demand on the basis of details provided.

Literature Review

Text to Image Models

1. Generative Adversarial Networks (GANs)
2. Encoder-Decoder Architectures
3. Transformer Based Models

Fashion Development

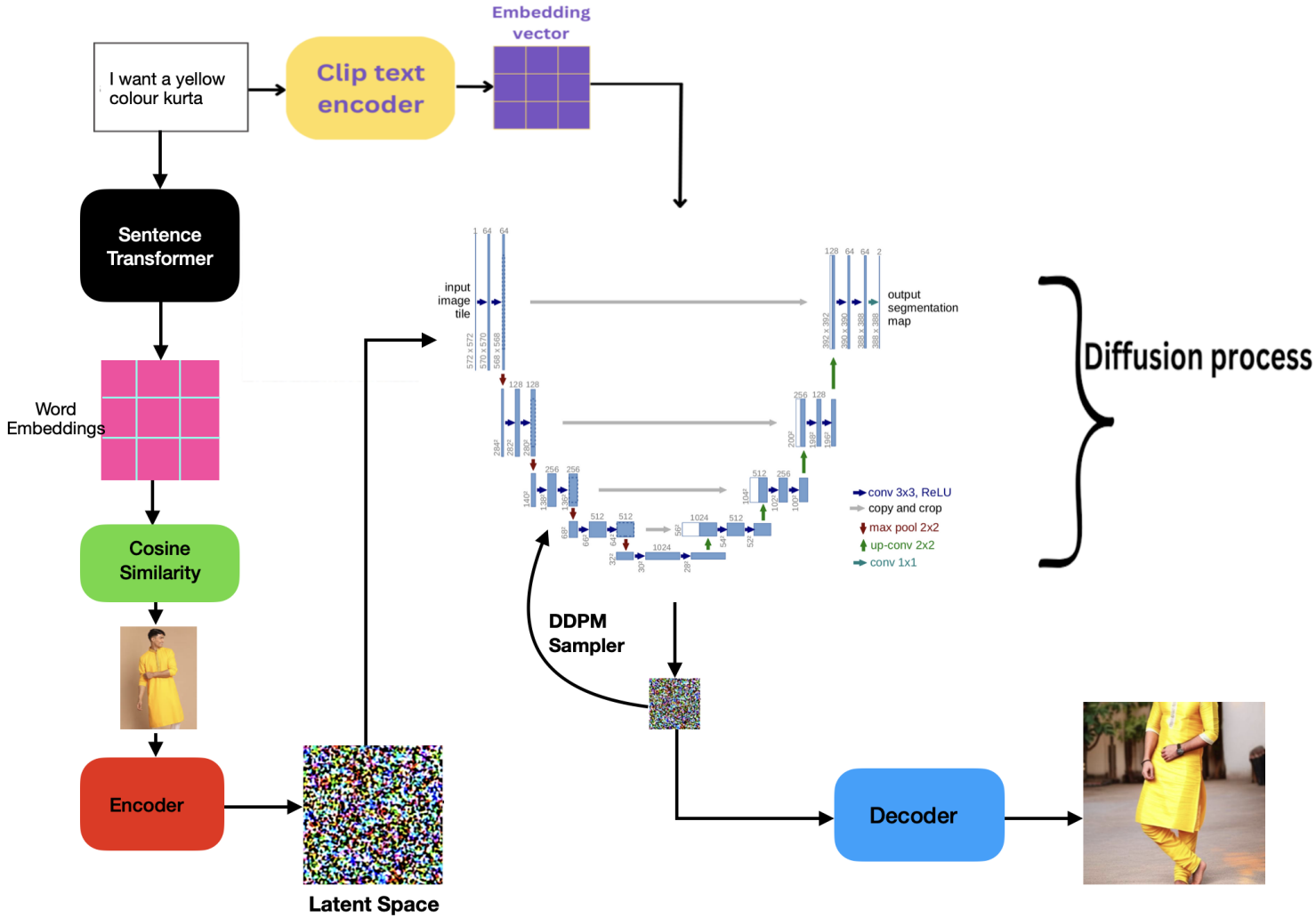
1. Fashion++: Minimal Edits for Outfit Improvement
2. FashionGAN: Fashion-Forward Image Synthesis with Generative Adversarial Networks
3. Text2Image: Generative Adversarial Synthesis from Language

Methodology

- Take input prompt from user as details of the outfit
- Using SentenceTransformer and CLIP Tokenizer convert the prompt to the word embeddings for inventory search and diffusion model respectively.
- Find the most similar outfit from the inventory according to the prompt using cosine similarity between word embedding of input prompt and description of each outfit.
- Now using the outfit collected pass it to encoder to convert the image to latent space and use this as the starting point of outfit generation.
- The embedding vector generated from CLIP Tokenizer and the latent space vector is now passed into the diffusion model for generation of outfit.
- The UNET model now takes this as input and generates an outfit as noise.
- DDPM Sampler removes the noise and add it back to UNET to improve the quality of outfit generated.
- After certain number of iterations, the image is then passed to Decoder to obtain actual dimensions, i.e. convert from latent space to original dimensions.

Pipeline

Here is the pipelined model of the complete project to explain the working.



Results

- Since we trained the model on latest outfits with newest trends. Hence we observed latest trends on the sample outfits too.
- Also model was trained on outfits for both men and women and hence it is able to generate outfits for everyone.
- We also observe variety in outfit, for example traditional and western wear were able to generate

Results

We tried our model with various input prompts some of them are:

Input Prompt: **I want a yellow colour kurta with white pyjama for diwali occasion for a man.**

Input Prompt: **Bel air silver glamour dress for women.**

Conclusion

- The project shows that using text-to-image creation for individualised fashion suggestions based on user descriptions is feasible.
- Cosine similarity is a useful tool for efficiently searching a huge clothing inventory for photographs of outfits that are most pertinent to the user's prompt.
- The deep learning pipeline enables image alteration to match the details specified in the prompt by utilising a UNET with an encoder-decoder architecture.
- When it comes to producing altered outfit photos that are semantically connected to the user's description, the model produces encouraging results.

Q & A

Thank You