

# Fashion Is All You Need

Diffusion-Based Deep Learning Architecture for Generating Textured Clothing in Multiple Poses

# Motivation

# Motivation

Why "All You Need Is Fashion?"

- **AI Generated Fashion Clothing**

The incorporation of Artificial Intelligence into fashion design can facilitate the production of new clothing designs that can bolster the creativity of clothing designers.

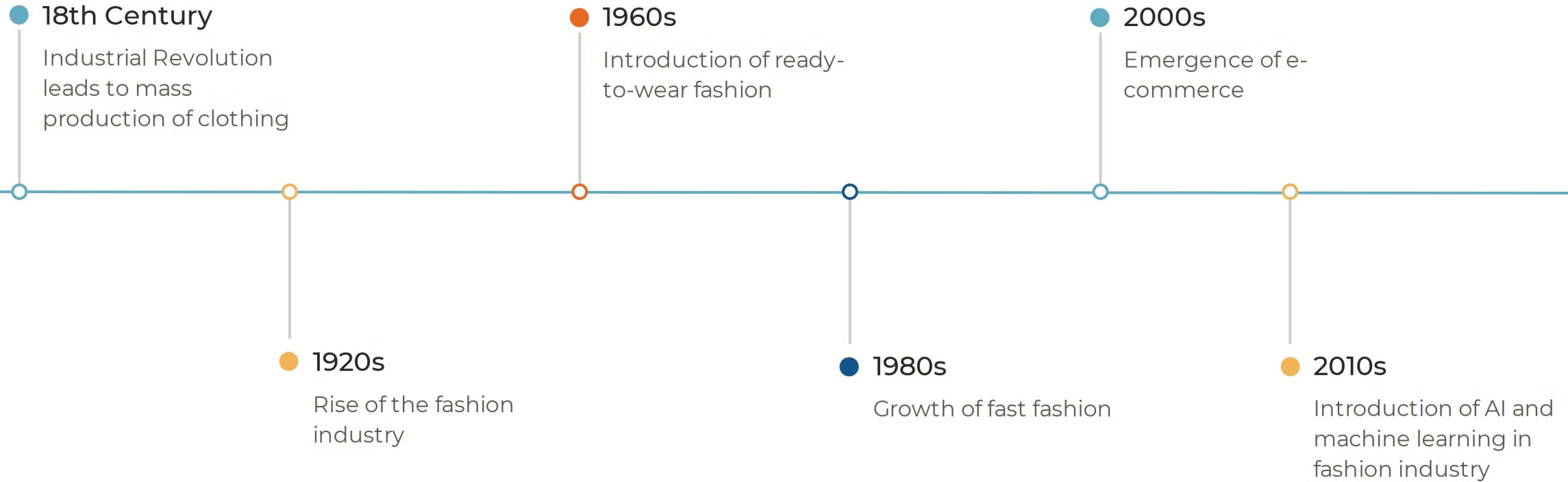
- **AI Pose Models**

Instead of spending a lot of money on traditional photoshoots for clothing, artificial intelligence models can be utilized to generate synthetic images of the desired clothing with predetermined poses.

- **Sustainability**

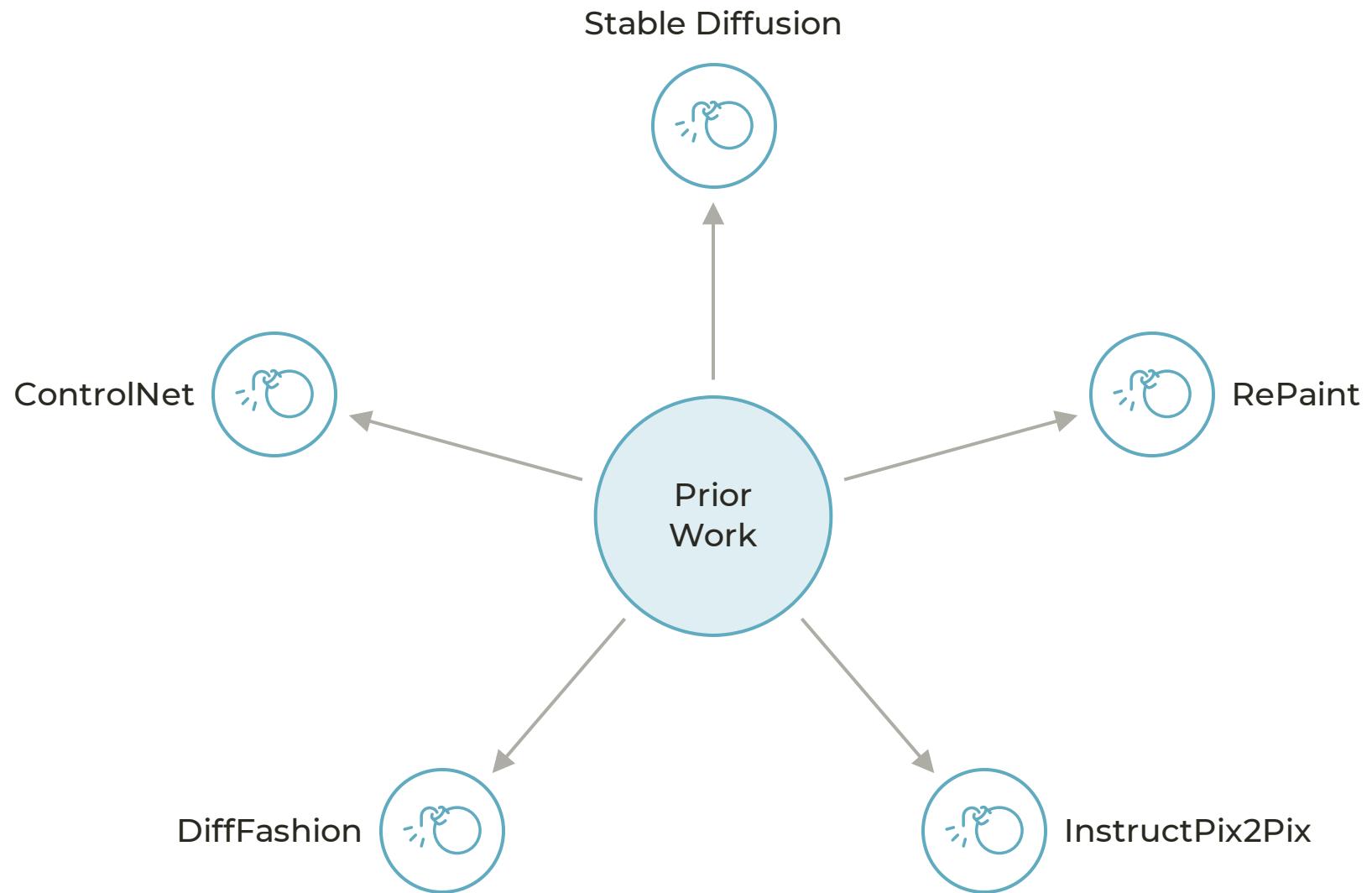
AI can enable us to reduce wastage, encourage inclusivity and diversity, and enable people to experiment with their style, ultimately altering our attitude and approach to fashion.

# Fashion Industry and Challenges



# Prior Work

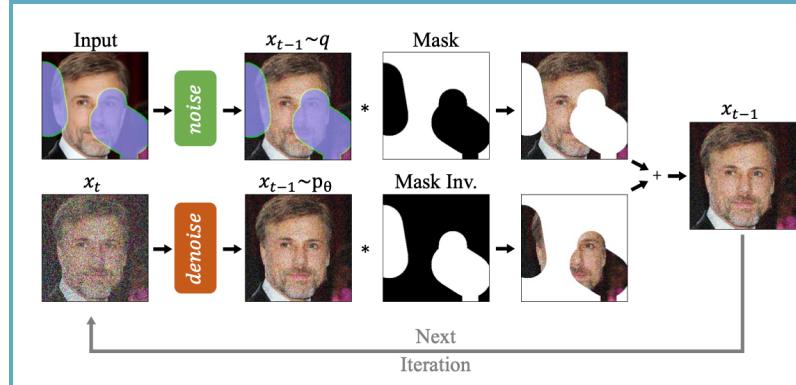
Acknowledging existing research that paved the way for our approach. We studied image synthesis and fashion, incorporating their techniques into our project.



# RePaint

## Introduction

A Denoising Diffusion Probabilistic Model (DDPM) based inpainting approach.



## Contribution

1

Synthesizes high-quality images with fine details by **inpainting** missing parts.

## Relevance

By leveraging RePaint's techniques for generating high-quality images and filling in missing details, we hoped to create more realistic and visually appealing fashion designs.

# Experiments with RePaint



Input Image



Segmentation Mask



Input + Mask

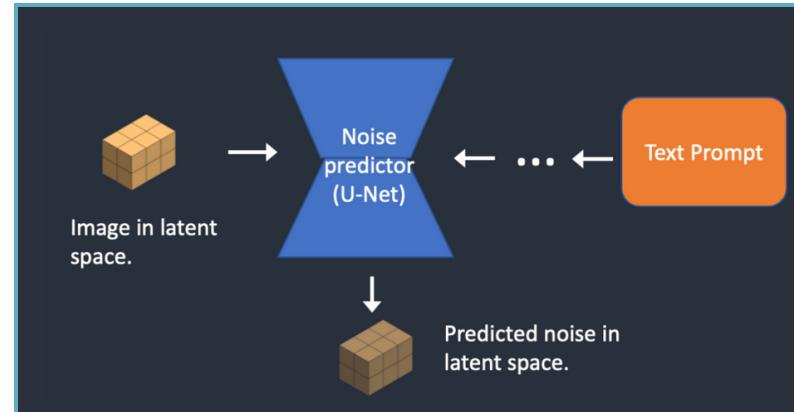


Output Image

# Stable Diffusion

## Introduction

Stable Diffusion is a **latent text-to-image diffusion model** capable of generating photo-realistic images given any text input.



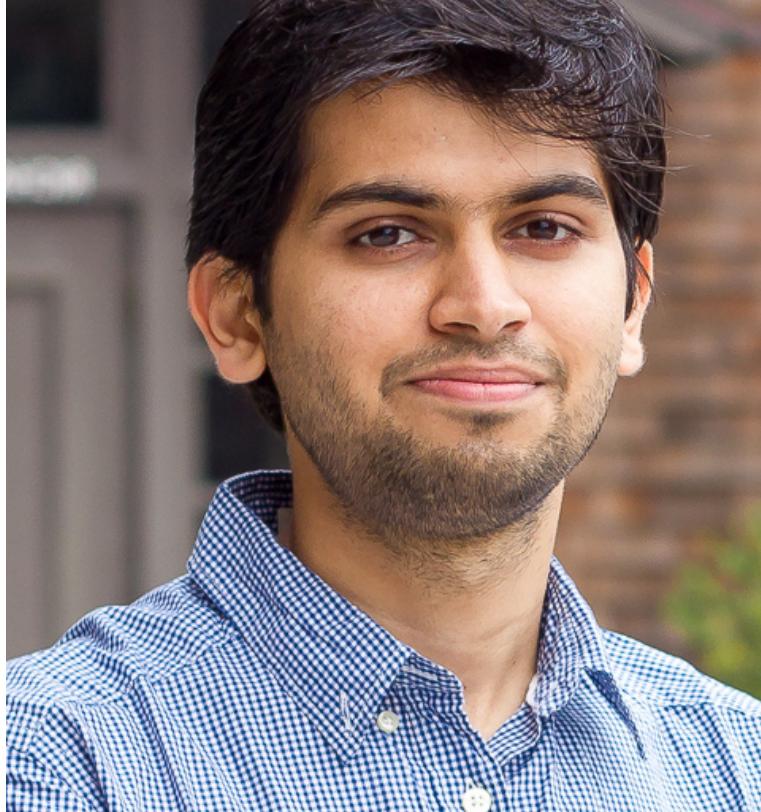
## Contributions

The model can be guided using **textual prompts** to generate images with specific attributes or features.

## Relevance

Stable Diffusion's use of textual prompts to guide image generation with specific attributes helps us generate images with desired features.

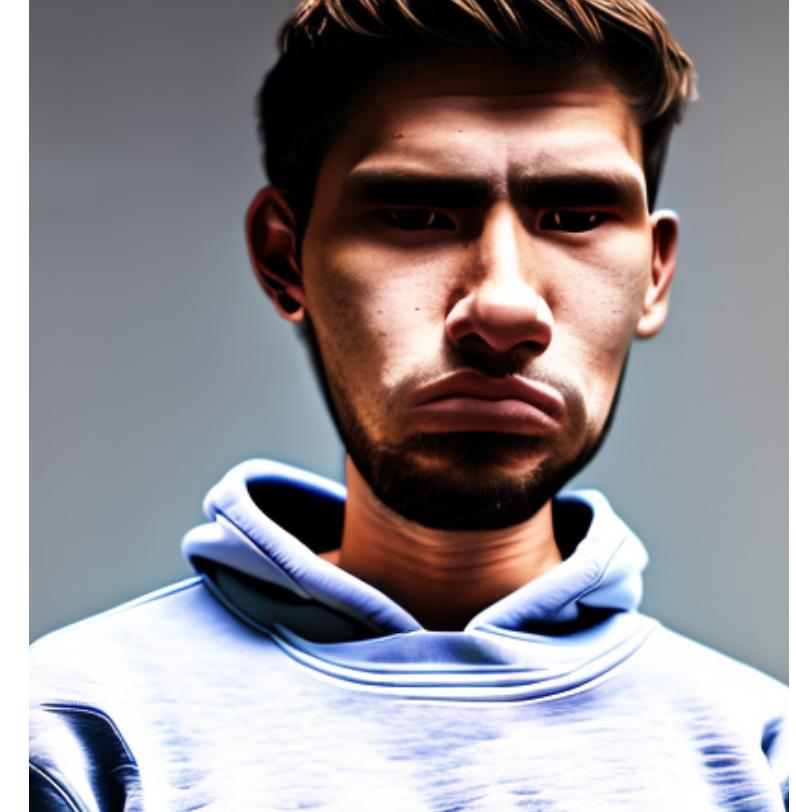
# Experiments with Stable Diffusion - 1.5



Input Image



Prompt: Change the shirt of the man  
to a colourful tee or T-shirt



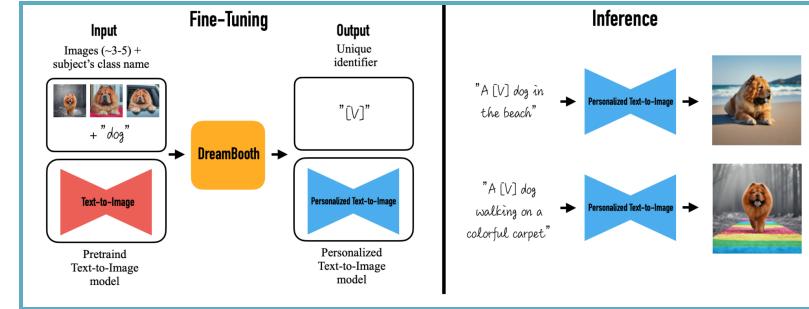
Prompt: Change the shirt of the man  
to a sweatshirt without changing his  
face

# DreamBooth

## Introduction

Another approach for "personalization" of text-to-image diffusion models.

## Methodology



## Contributions

ControlNet has many possibilities that allow us to control stable diffusion using object borders, lines, scribbles, pose skeletons, segmentation maps, depth maps, and more.

## Relevance

Can showcase the clothing in different poses and angles to give people a better idea of how it will look on them.

# Experiments with Fine-Tuned DreamBooth



Prompt: A photo of Adeesh wearing a red sweater.



Prompt: A photo of Adeesh wearing a pink shirt.

# ControlNet

## Introduction

ControlNet is a neural network structure to control diffusion models by adding extra conditions.

## Methodology

- ControlNet involves locking the original AI model and creating a separate external network to understand new inputs.
- The external network is trained using a stable diffusion training routine with image-caption pairs and depth maps.
- The trained external network is integrated with the main model to improve performance on the subtask.

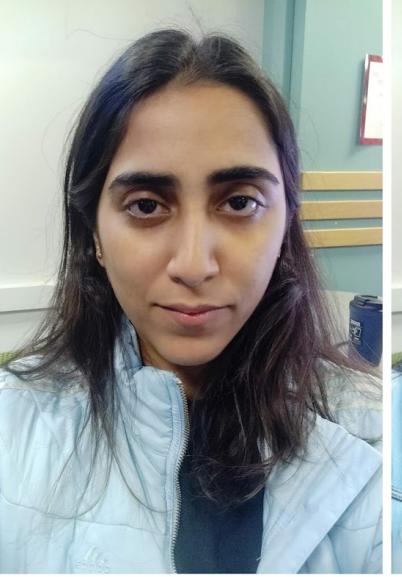
## Contributions

ControlNet has many possibilities that allow us to control stable diffusion using object borders, lines, scribbles, pose skeletons, segmentation maps, depth maps, and more.

## Relevance

Can showcase the clothing in different poses and angles to give people a better idea of how it will look on them.

# Our Work



Goal: Generate Pose Aware Photo-Realistic Clothes For A Target Image, using an input image of the cloth texture, and a text prompt

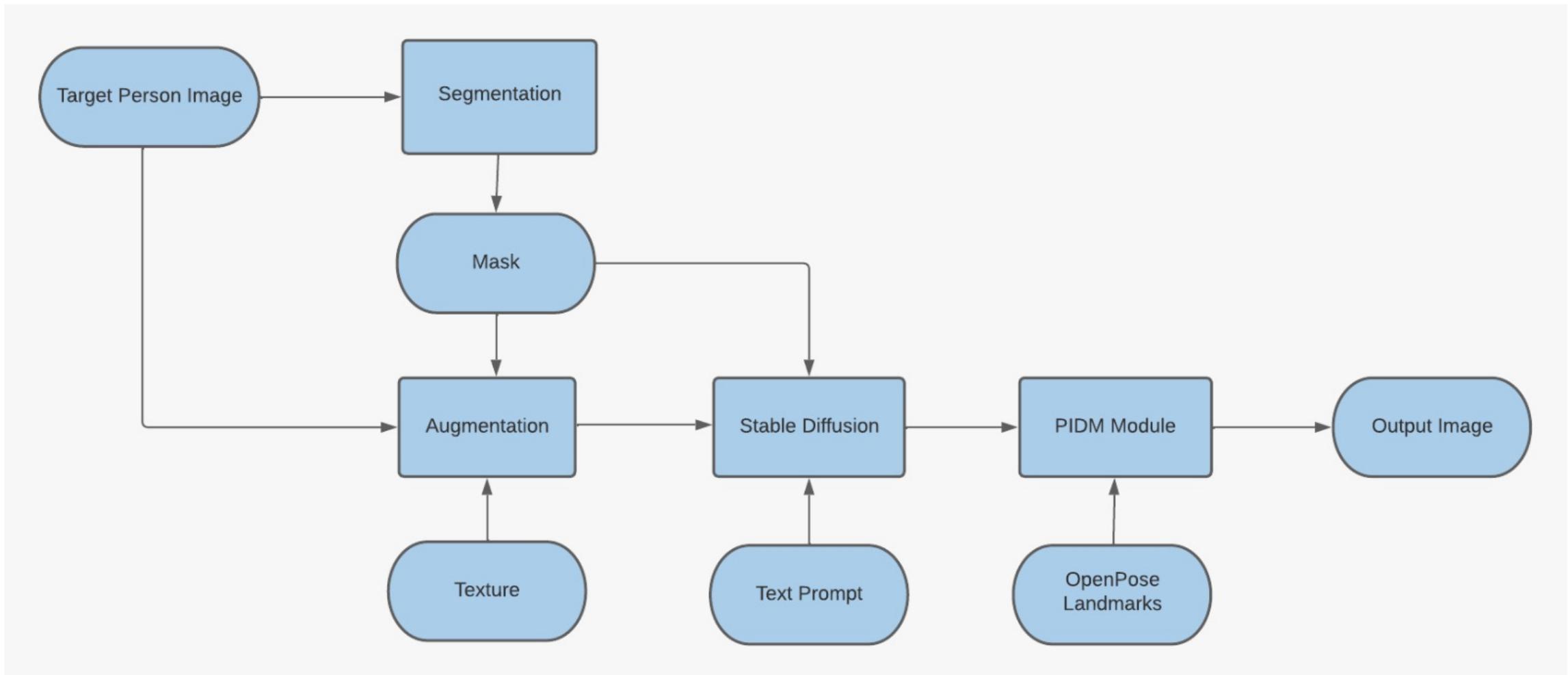


# Why Does Our Idea Work Intuitively?



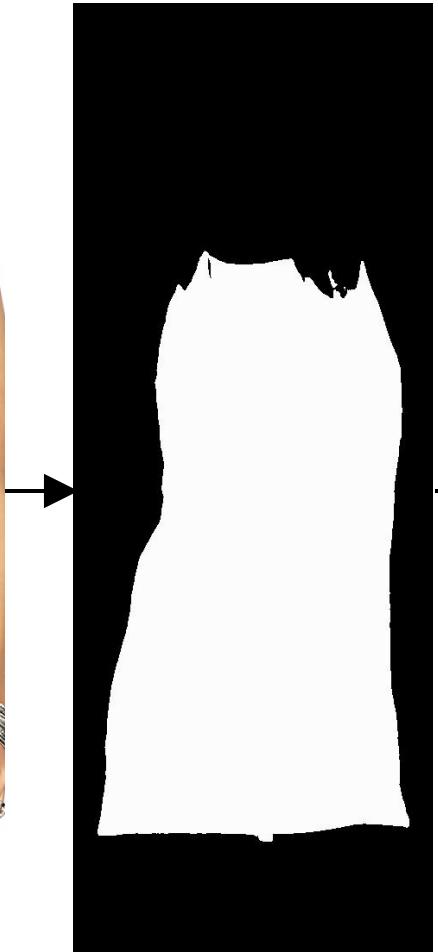
- Information such as cloth texture cannot be explained in text and visual context is needed
  - Cloth Segmentation helps provide image prior to the diffusion model and control regions in-painted
  - Augmentation is created such that the cloth texture is super imposed onto segmentation mask
    - Prompt: "Complete Woman's Dress in High Quality" provides context to the stable diffusion model that the inpainting is a "Woman's dress"
  - The information of the body shape embedded in the dress is lost when the texture image is superimposed and hence PIDM is used for preserving and generating new body postures ad shapes.

# Flow Chart



# Pipeline

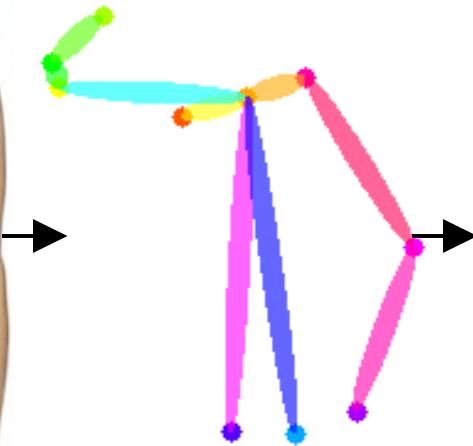
## Our Pipeline



New Texture  
Transfer



Stable Diffusion

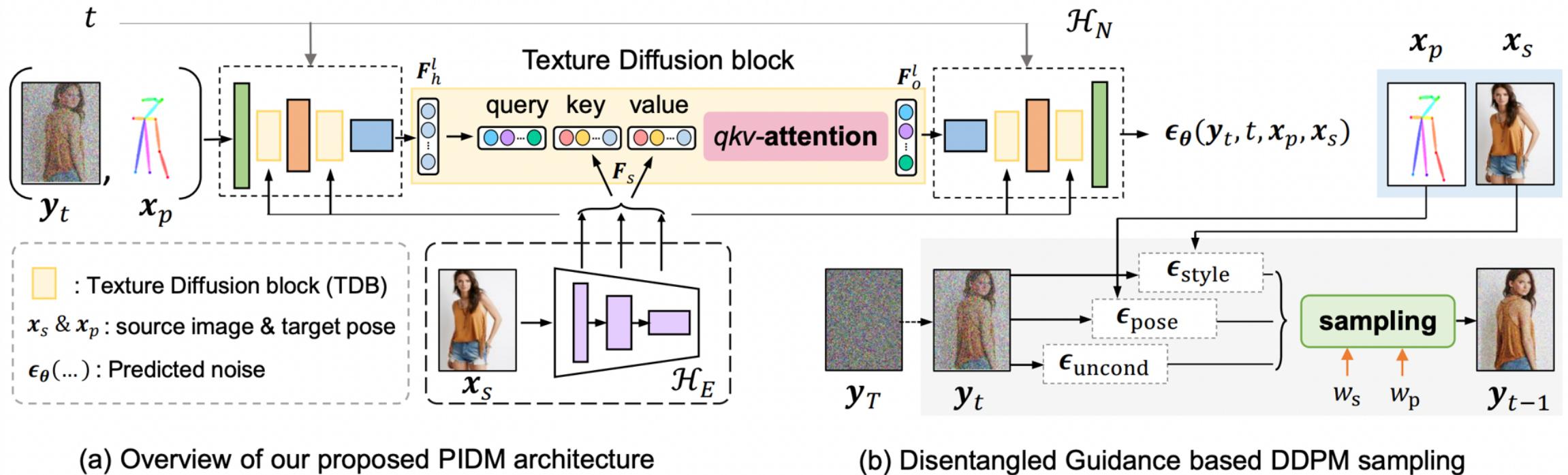


Pose Input



Pose Generation

# Pose Generation Architecture: PIDM



# Pipeline

## Our Pipeline



# RESULTS

# Results



Input Image



Segmentation



New Texture  
Transfer



Stable Diffusion  
Result

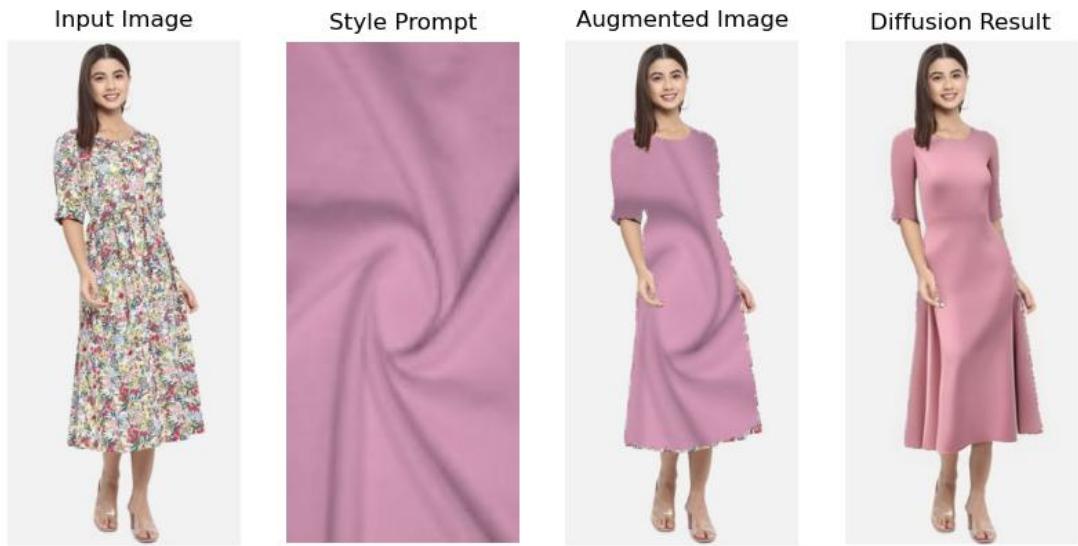


New Texture  
Transfer  
Augmentation



Stable Diffusion  
Result

Fashion Is All You Need: Stable Diffusion Results



Fashion Is All You Need: Stable Diffusion Results



Fashion Is All You Need: Stable Diffusion Results



Fashion Is All You Need: Stable Diffusion Results



# Pose Generation With Open-Pose Prompts



Reference image (On the left) is input to PIDM model to generate various poses

# Pose Generation With Open-Pose Prompt



Reference image (On the left) is input to PIDM model to generate various poses

## Limitations of the Model

- 1 | Depends on accurate segmentation of the clothes
- 2 | Requires heavy computational resources for stable diffusion inference
- 3 | Faces tend to get distorted in the pose generation phase



## References

- 1 | Stable Diffusion: <https://stable-diffusion-art.com/how-stable-diffusion-work/>
- 2 | High-Resolution Image Synthesis with Latent Diffusion Models, Robin Rombach et al, 2021
- 3 | DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation, Nataniel Ruiz et al, 2022
- 4 | Person Image Synthesis via Denoising Diffusion Model, Ankan Kumar Bhunia et al, 2022
- 5 | Adding Conditional Control to Text-to-Image Diffusion Models, Lvmin Zhang et al, 2023
- 6 | Person Image Synthesis via Denoising Diffusion Model, Ankan Kumar Bhunia et al, 2022
- 7 | PISE: Person Image Synthesis and Editing with Decoupled GAN, Jinsong Zhang et al, 2021
- 8 | InstructPix2Pix: Learning to Follow Image Editing Instructions, Tim Brooks et al, 2022
- 9 | Image-to-Image Translation with Conditional Adversarial Networks, Phillip Isola et al, 2016
- 10 | DiffFashion: Reference-based Fashion Design with Structure-aware Transfer by Diffusion Models, Shidong Cao et al, 2023

# Future Work