

FINE-TUNING STABLE DIFFUSION FOR STYLE TRANSFER

PROBLEM DEFINITION

The task of fine-tuning stable diffusion for style transfer involves optimizing diffusion methods to effectively imbue target images with desired artistic styles. This entails refining diffusion parameters to ensure stable and faithful style transfer results across various input images and artistic styles.



“An Image of a Doctor”



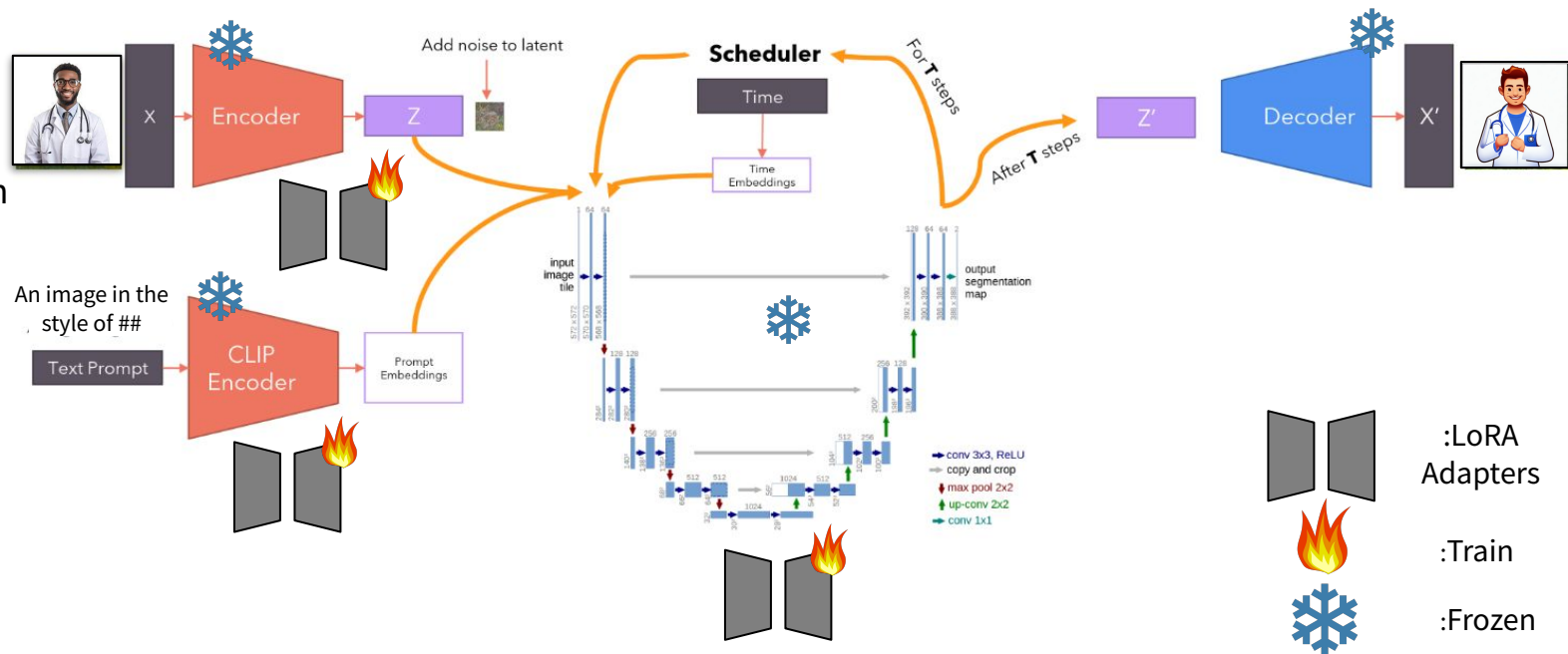
Style : Clipart

METHODOLOGY

Architecture (Image-To-Image)

1) LoRA Adapter on CLIP Encoder, Conditional-Unet and C-VAE

2) Plan to take a few shot approach with style transfer using only 8 images of the paradigm under consideration in that particular style

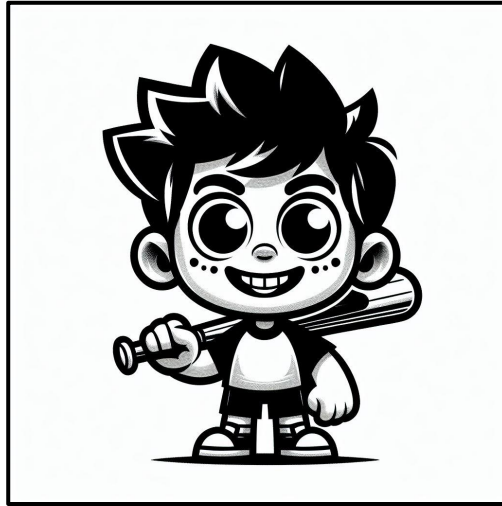


**RESULTS FOR
IMAGE GENERATION and
STYLE TRANSFER BY
CHANGING THE TRAINABLE
LAYERS**

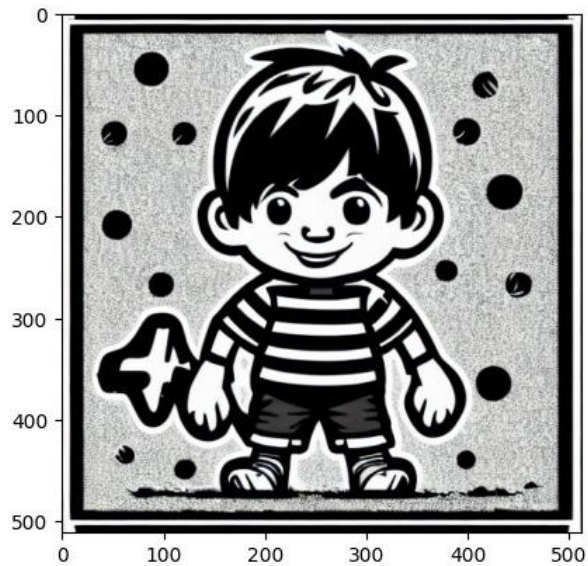
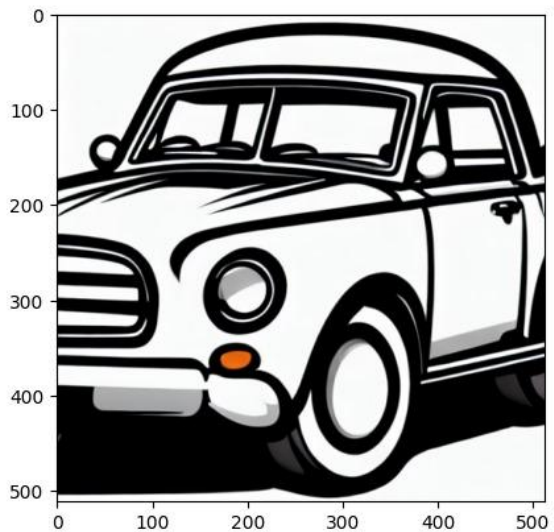
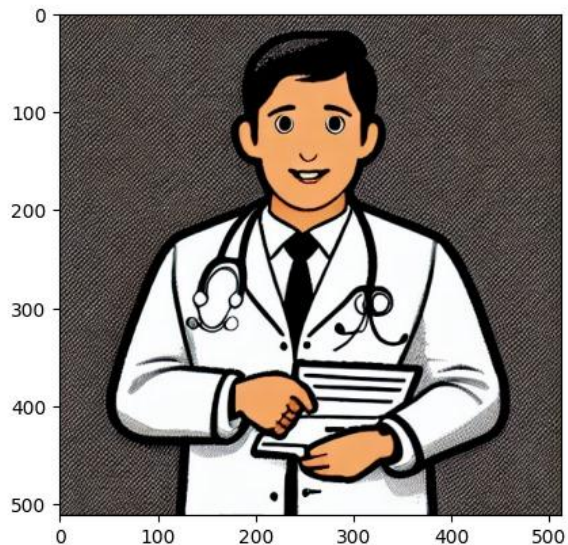
IMAGES USED FOR THE ANALYSIS OF STYLE TRANSFER:



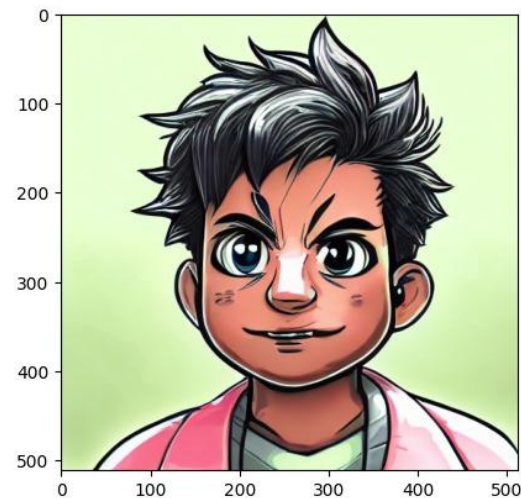
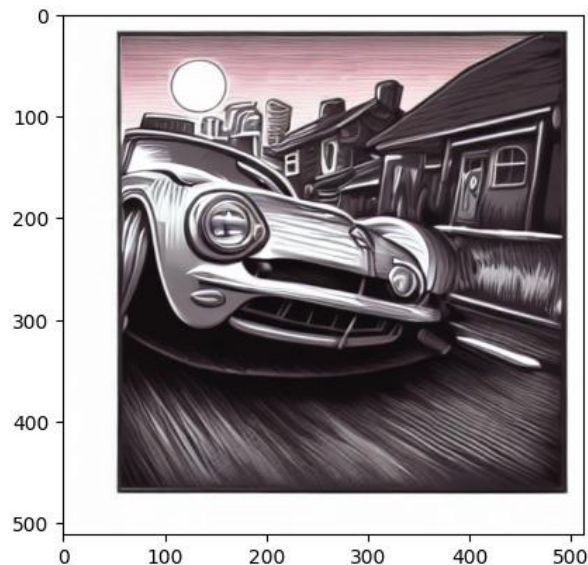
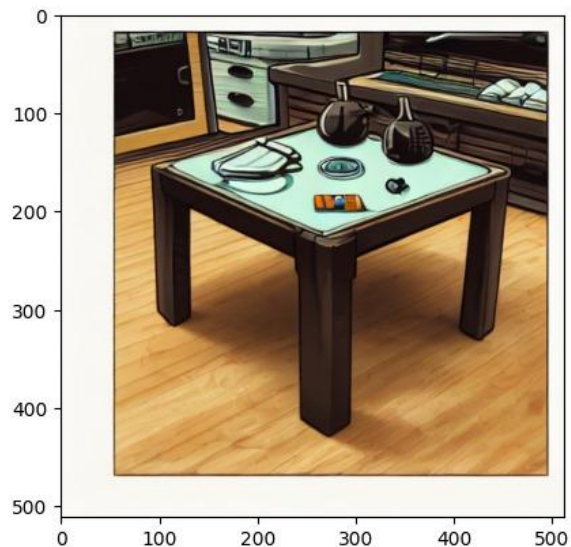
STYLE: CLIPART



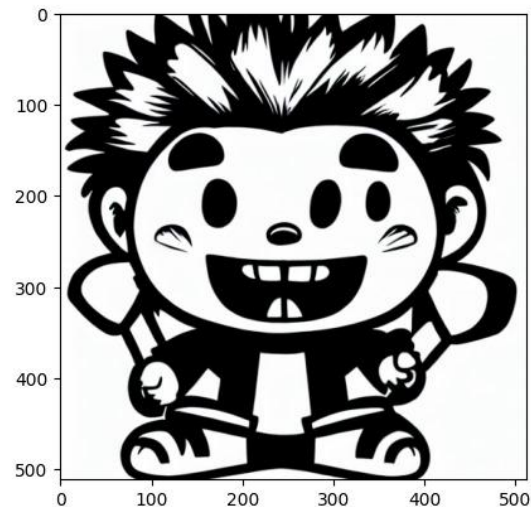
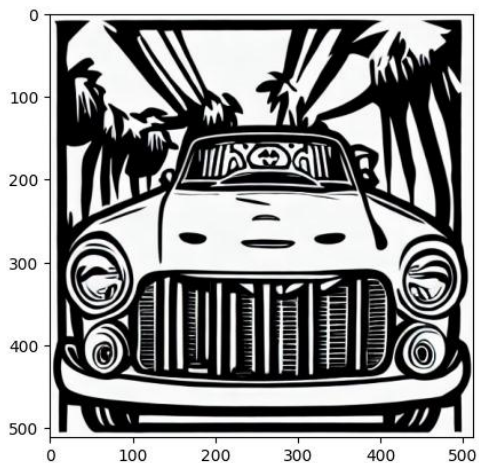
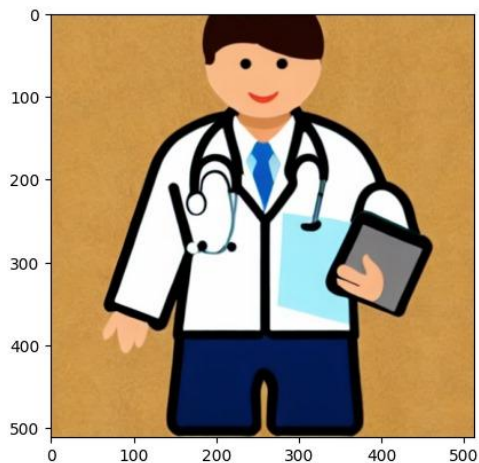
WITH FT. ALL (image generation)



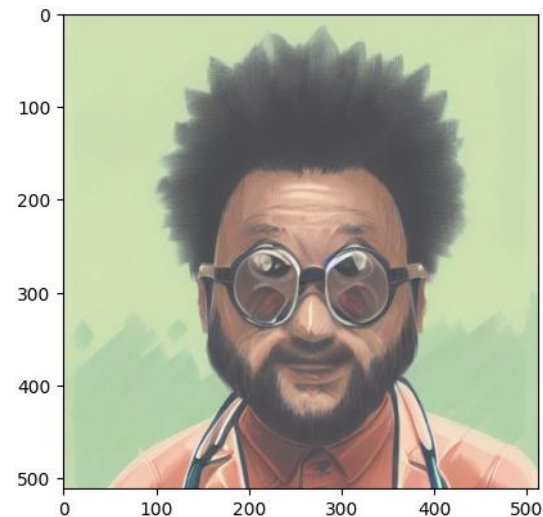
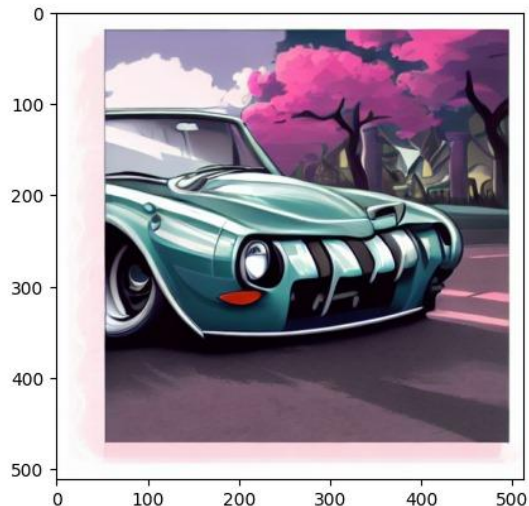
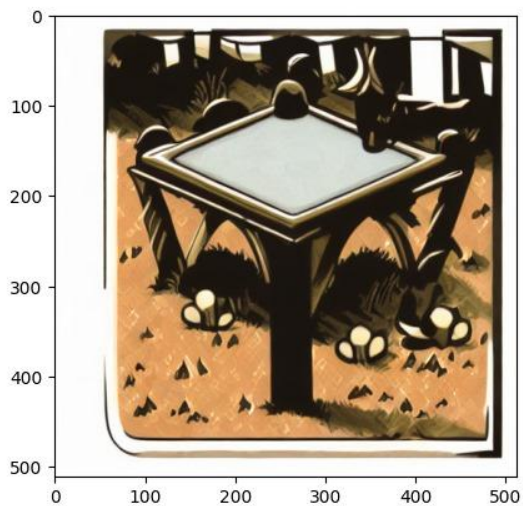
WITH FT. ALL (style transfer)



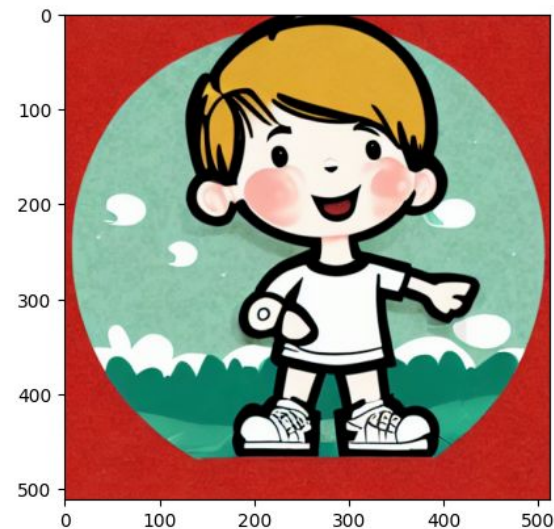
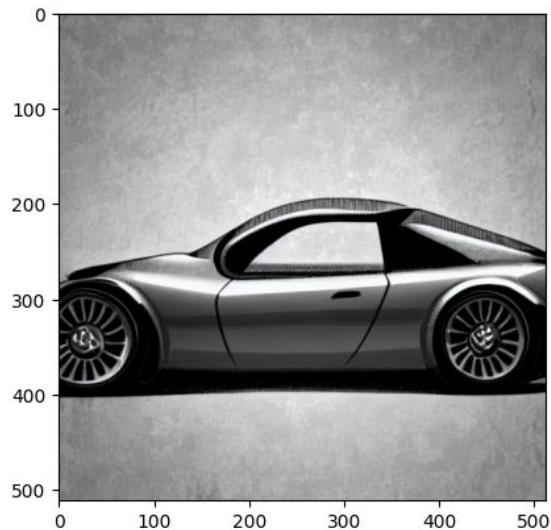
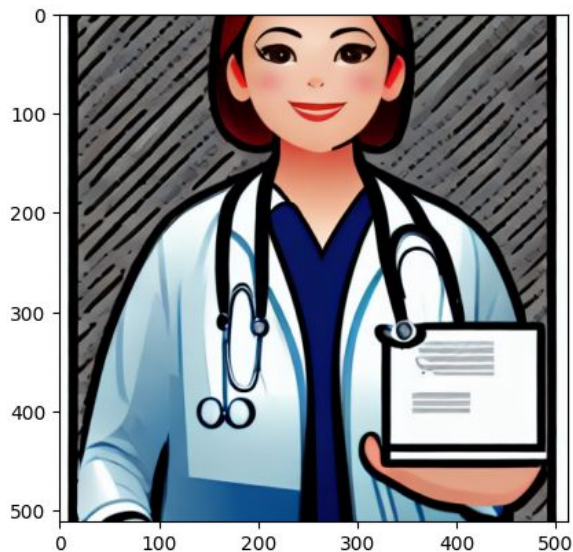
WITHOUT FT. VAE (image generation)



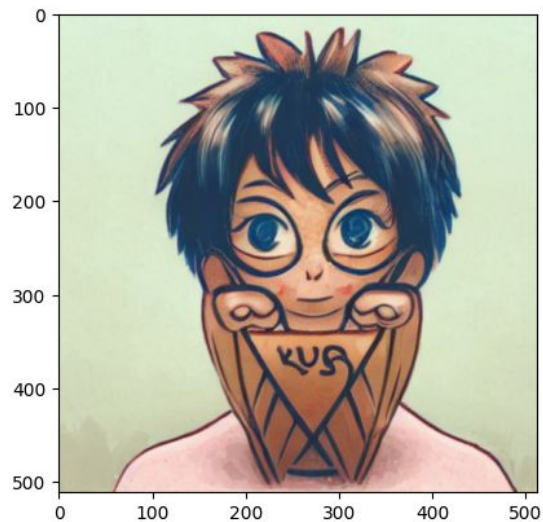
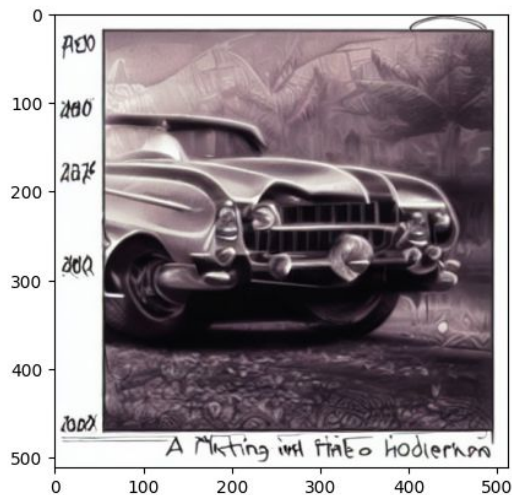
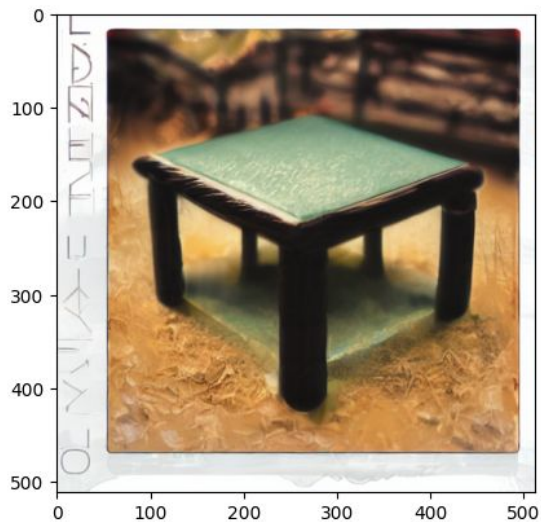
WITHOUT FT. VAE (style transfer)



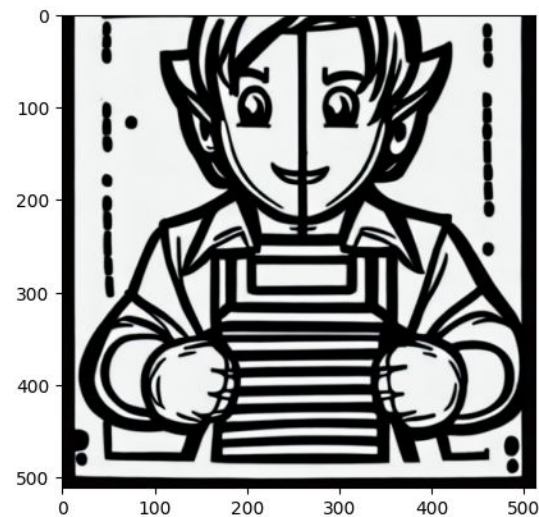
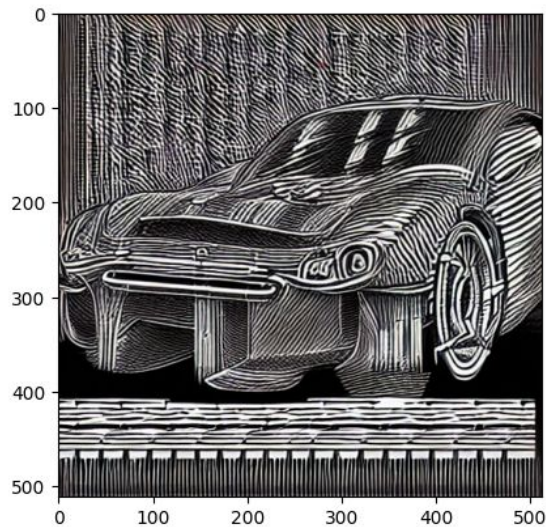
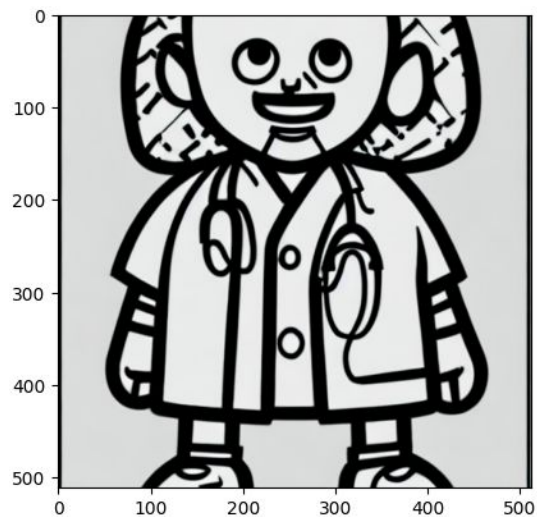
WITHOUT FT. TEXT (image generation)



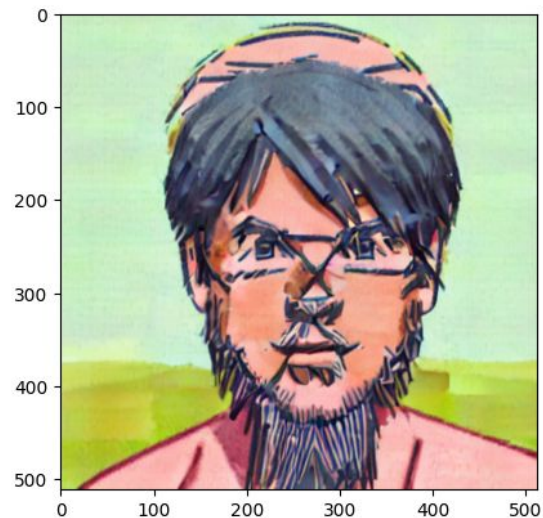
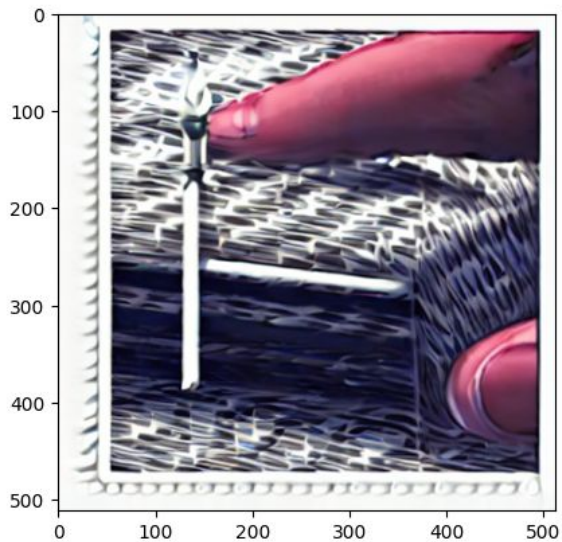
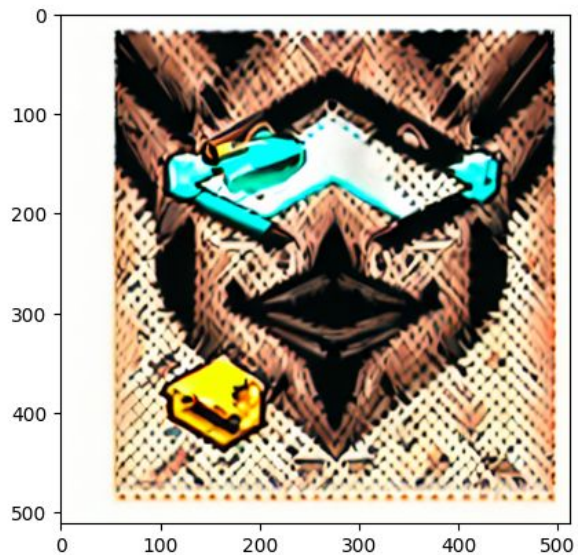
WITHOUT FT. TEXT (style transfer)



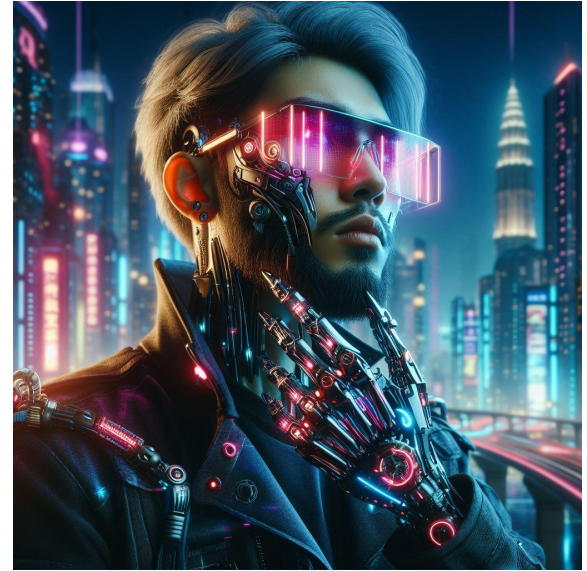
WITHOUT FT. UNET (image generation)



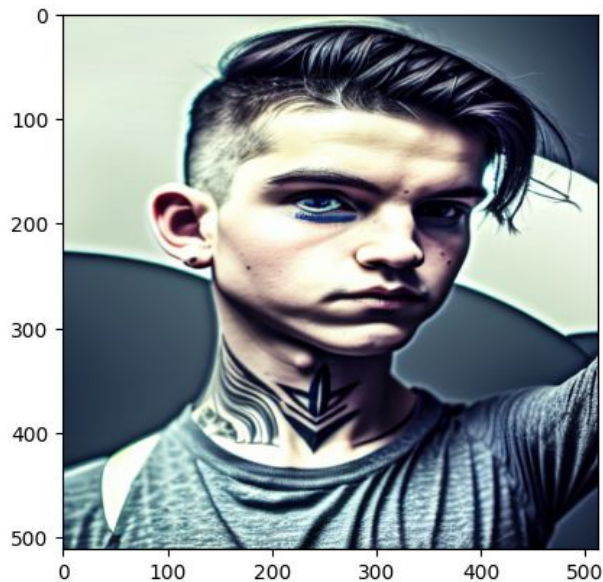
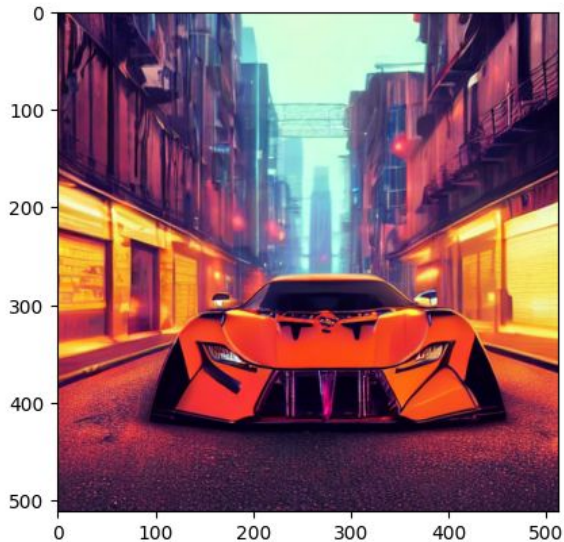
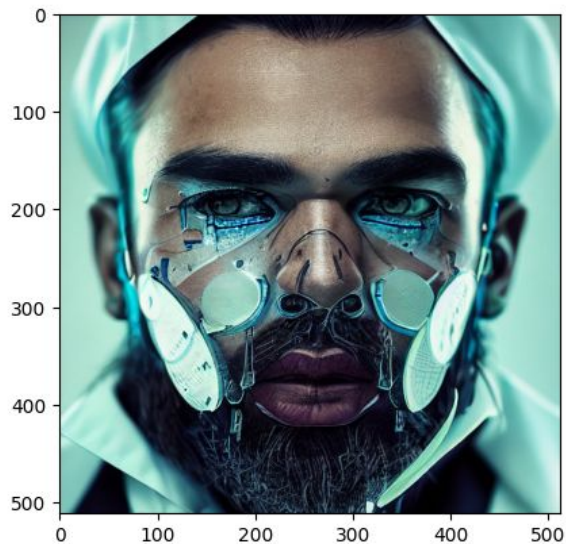
WITHOUT FT. UNET (style transfer)



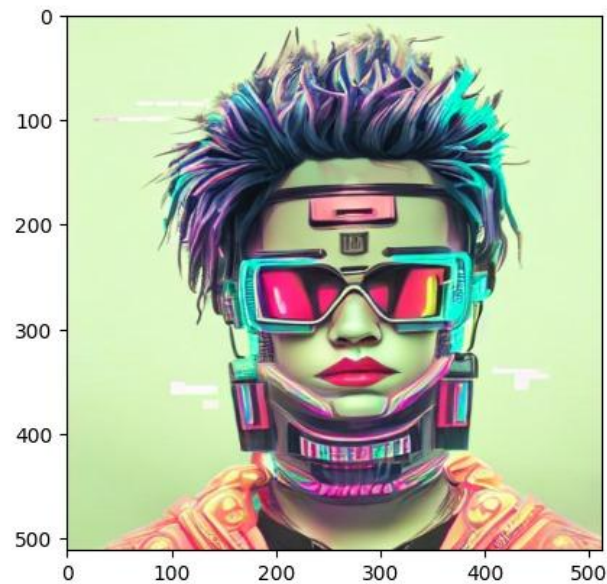
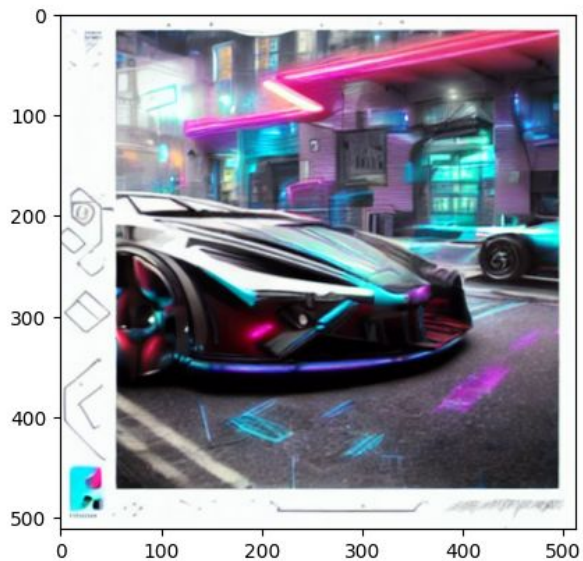
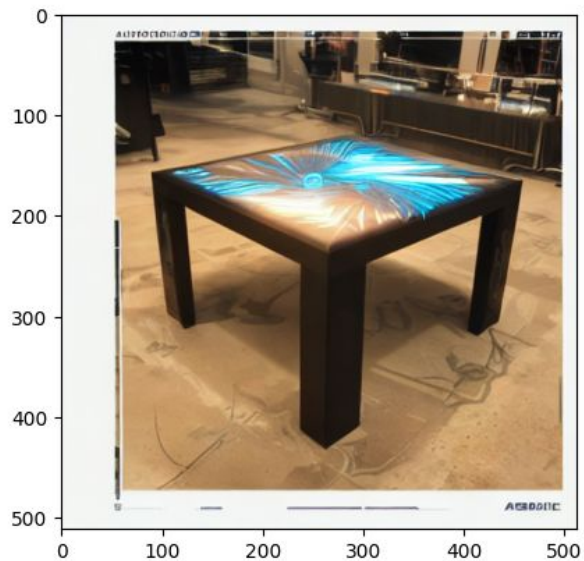
STYLE CYBERPUNK



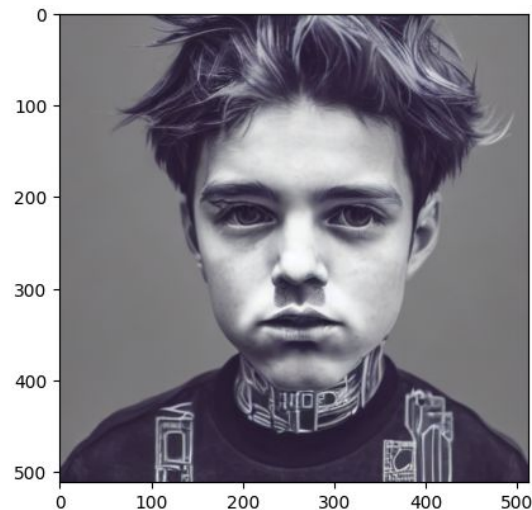
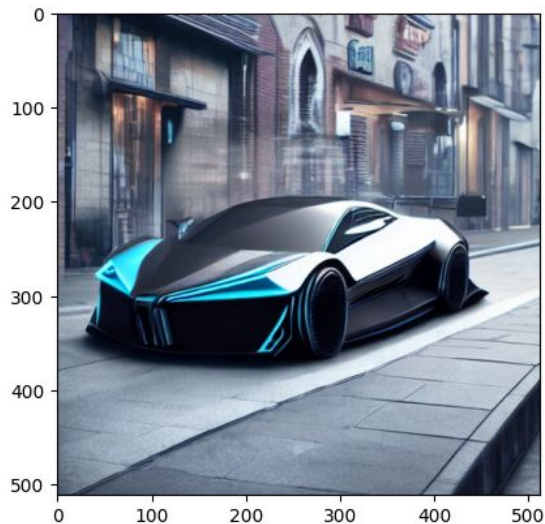
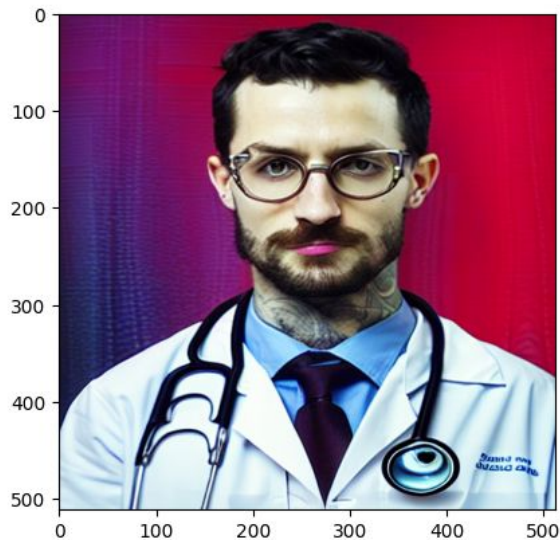
WITH FT. ALL (image generation)



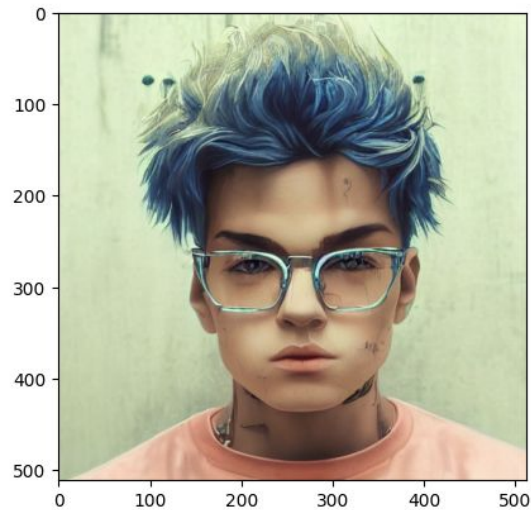
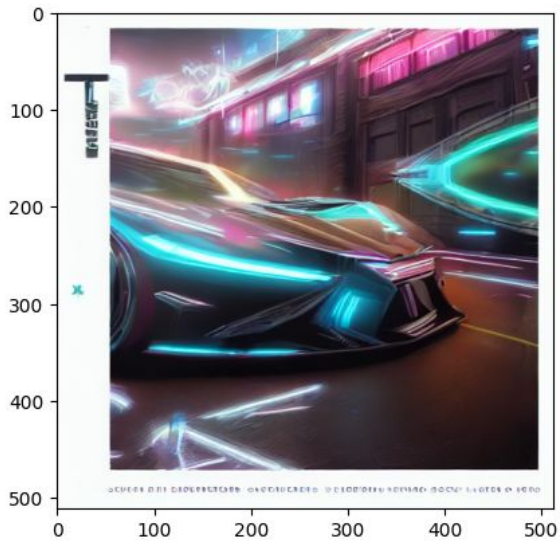
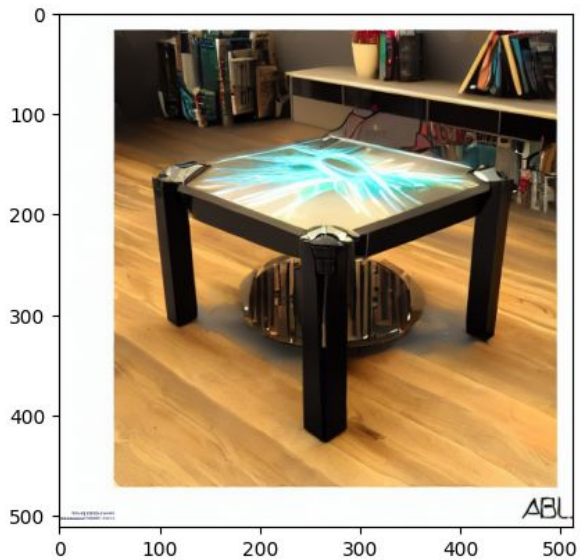
WITH FT. ALL (style transfer)



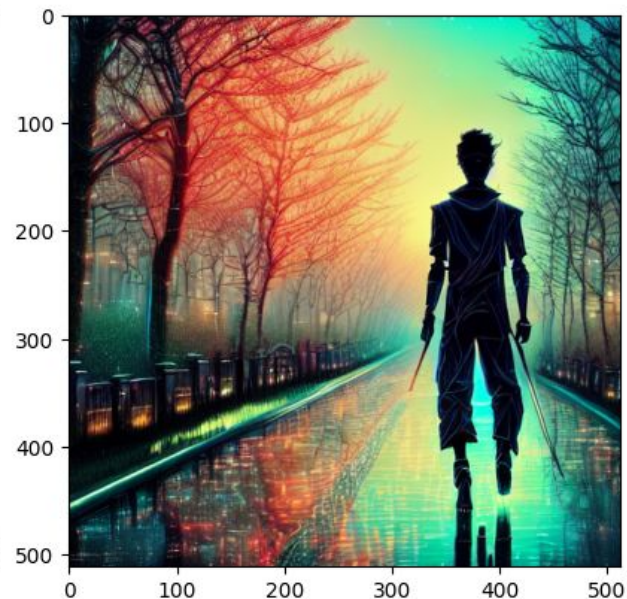
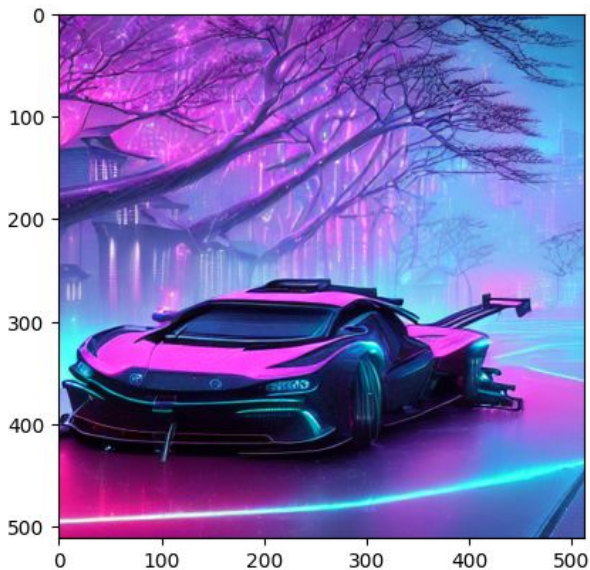
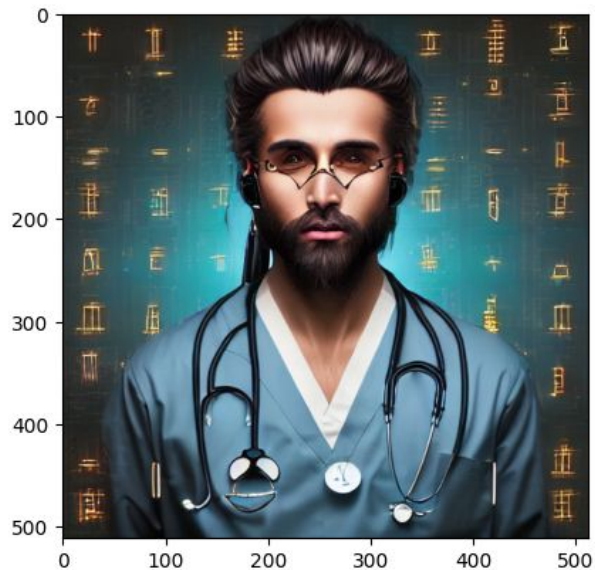
WITHOUT FT. VAE (image generation)



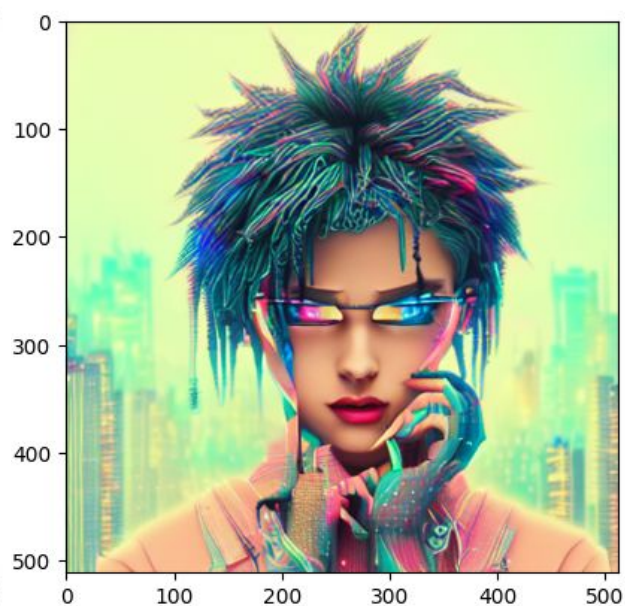
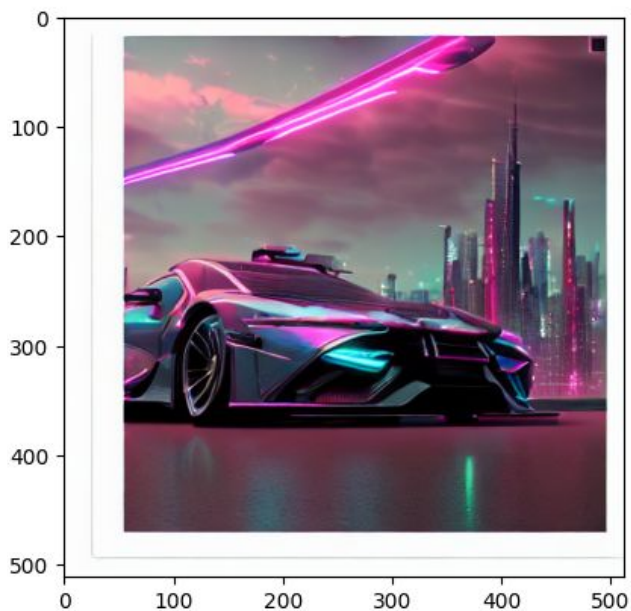
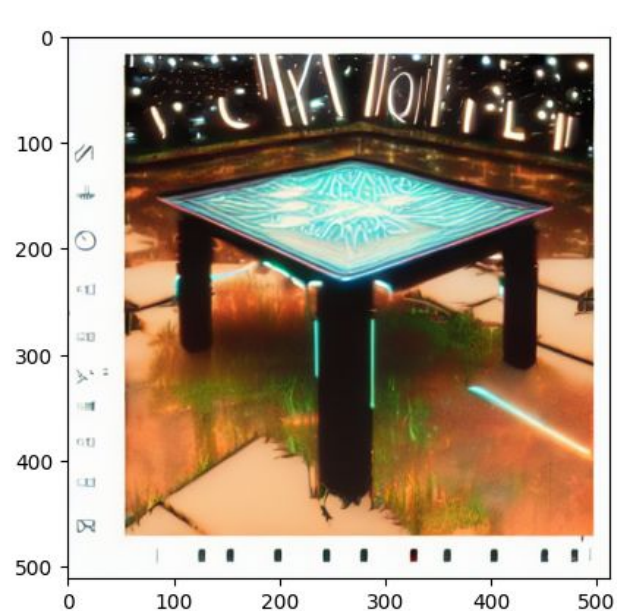
WITHOUT FT. VAE (style transfer)



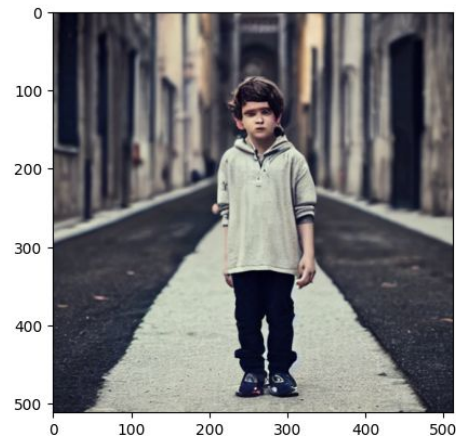
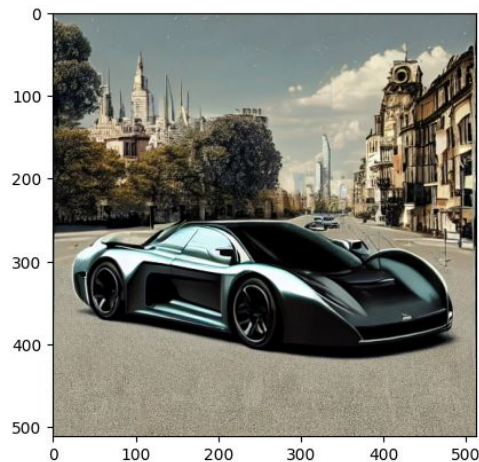
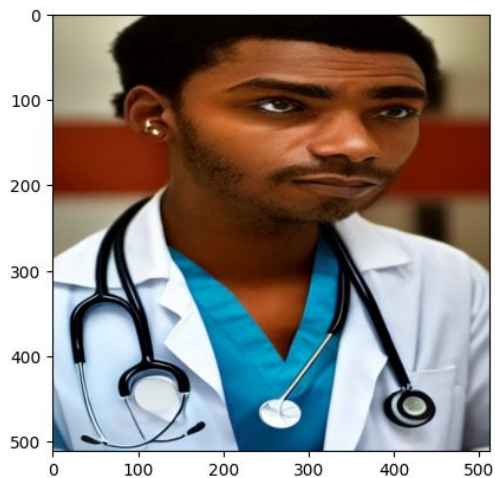
WITHOUT FT. UNET (image generation)



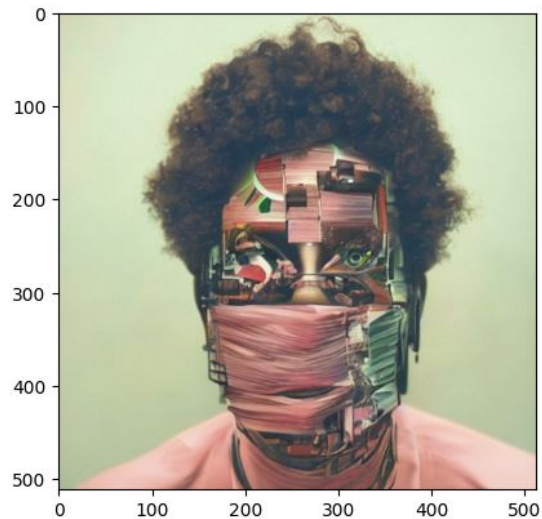
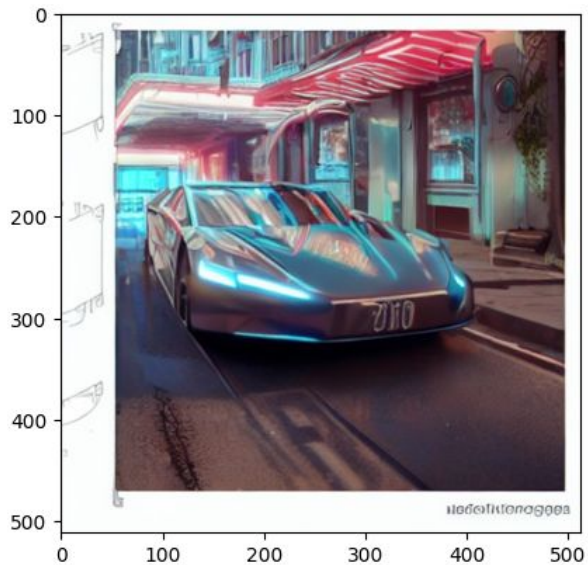
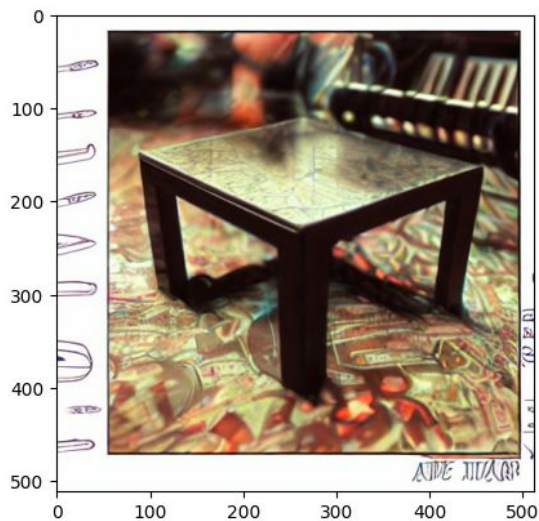
WITHOUT FT. UNET (style transfer)



WITHOUT FT. TEXT (image generation)



WITHOUT FT. TEXT (style transfer)



Conclusions

- Fine-tuning the text encoder plays a crucial role specially for the task of image generation but its effect is felt less for the task of style transfer from one image to another.
- Fine-tuning the UNet plays the most crucial role for a consistent output for both the tasks.
- Fine-tuning of the VAE is necessary for a cleaner and more stable output.

Thanks!

Arkadeep Acharya
2101A41



STABLE DIFFUSION 1.5
PROMPT: "STYLE TRANSFER"