Virtual Art- ADL

Florian Zogaj

January 16, 2024

Introduction

The main goal for this project is to create a system that not only recognizes and segments objects within images but also artistically alters them using neural style transfer. Traditional image processing methods often handle these tasks in isolation and therefore lack the comfort using these techniques together. While onventional neural style transfer alters the whole image, this project focuses on giving users the ability to stylize only certain regions. For example: I upload an image of myself and a cat and I want to turn the style of the cat into a Starry Night Van Gogh Style. Reading [Minaee et al., 2021] I learned about what the most well known techniques and datasets are in the context of image segmentation tasks.

Type of project:"Bring your own method" - Enhancing conventional neural style transfer with region-specific styling functionality (using SAM).

The primary objectives of this project are:

- **Develop an Integrated Pipeline:** Combining object detection (YOLOV6), segmentation (SAM), and neural style transfer (VGG-19) into a cohesive workflow
- **From an artistic perspective:** Achieve style transfer on objects from which the art can be recognized
- **Creating an application** that makes the system usable in real time

Methods

Object Detection - YOLOV6

Object detection is a crucial task in computer vision, involving the identification and localization of objects within images. We use this to detect objects and the localization to then recieve the masks. YOLO (You Only Look Once) with YOLOV6 being one of the latest versions. This model is capable of detecting objects in real-time and therefore fits our goal of building an application. Unlike other models that sequentially perform region proposal and classification, YOLO analyzes an entire image in a single evaluation and predicts bounding boxes and class probabilities simultaneously. These probabilities could further be used in our project to let the user decide if he wants to only see objects that are recognized with at least a certain confidence. [Li et al., 2022]

Figure 1: Detected objects with > 0.5 confidence

Mask Creation with Segment Anything Model (SAM)

In the project proposal I still was not sure if I'll be able to use this model as the paper I have read is very new and I have just heard about it. [Kirillov et al., 2023] In the end it fit really well in the application pipeline. Mask creation or image segmentation involves classifying each pixel of an image into a relevant category. The Segment Anything Model (SAM) offers an approach to segmenting various objects in an image. Some other models I have read about are trained on specific categories, while SAM generalizes well across different objects. This capability is used for our application and worked very accurate as further described in the error metrics.

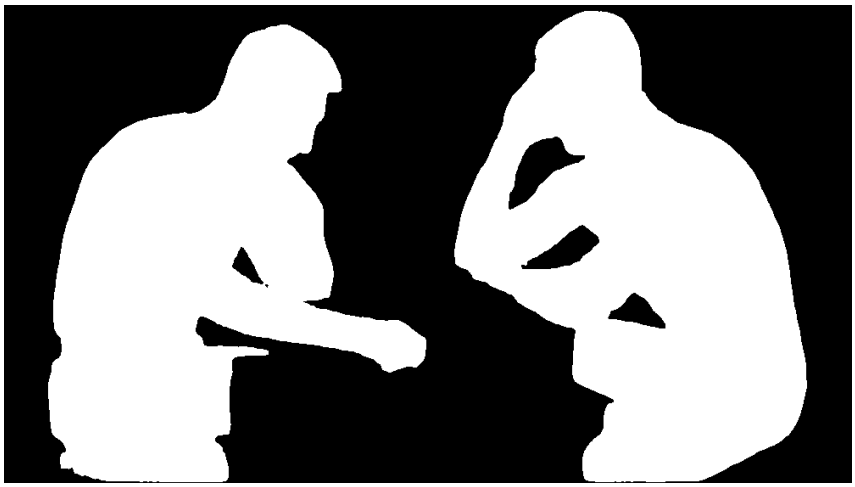

Figure 2: Masks of detected persons - SAM

Neural Style Transfer with VGG-19

Neural Style Transfer uses Convolutional Neural Networks (CNNs) to apply the artistic style of one image to the content of another. As mentioned before, this process is usually done to style the whole image. For this task I used the VGG-19 model. The goal here was to apply the style but not lose to too much of the structure of the object. This was done by experementing with different emphasis on content and style. As the objects seem to be less recognisable in contrast to their surroundings without style, I have reduced the style score and increased the content score in the calculation of loss. This resulted in better looking images. Further experimentation can be done by adjusting the weights. Also, a user study can be conducted to measure how well the style of the image is translated into the objects.

Figure 3: Applied style on the detected persons

Error Metric

I manually created masks of objects in my images and compared them to the masks created by SAM. In this case the manually created masks are the "ground truth" that is being compared to by IoU. If an object is detected and I want to transfer a style to that object, I would want the mask to be as close as possible to the real object. How good the style is transferred may be more subjective and can be explored via user studies. However, by looking at the loss of style, I was still able to compare how the changes in style weight affected the image. The goal was to achieve IoU scores of over 75% to ensure that good results are possible when transferring styles. In a test file on my github-repository, the IoU values of 2 images are shown. Here I detected the 2 persons on the image and created their masks. Both instances have shown IoU values of over 93%, which is more than enough to get a reasonable output style. Creating own masks manually is very time-consuming, but I wanted to go through this process myself to test the output.

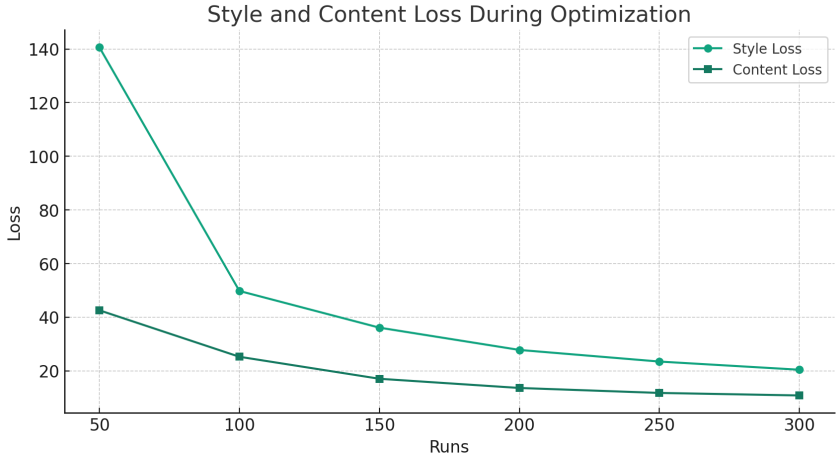

Figure 4: Style and content loss while optimizing

Work Breakdown Structure

Task	Estimated Time	Actual Time
Project setup	7	12
Design and build Pipeline	20	20
Fine Tuning and Tests	17	20
Application Development	6	9
Report	10	10
Final Presentation	3	3

In the project setup I had to first find out how I can make use of the models I used. Checkpoints were used for this task, which was quite time-consuming as this was the first time

I implemented a project like this. I also invested a lot of time in thinking of suitable error metrics and tried out different techniques before coming up with the ones I mentioned. The application development took longer than expected as I thought gradio would give me more freedom when designing the application.

Future Work

While working on the project I thought about many ways how this could be improved or applied to real world scenarios. I want to build a better application where the generated masks are displayed in the image and as the user hovers over them, they can select the objects to apply the style to. Further experementing with style weights can be done to adjust them according to the content image, as I think there are certain features in objects that would need more style weight, while others need less. As the models required image compression when applying the style, the results show lower resolution on the selected objects. This could be improved by dividing the image in multiple patches. However this would need more experimentation due to the irregularities the patches would have to each other when joining them back together. The combination of object detection and style transfer also brought me to the idea of how this pipeline can be used for blurring faces or number plates on pictures.

More Examples:

(a) The Great Wave (b) Starry Night (c) Kandinsky (d) Great Wave

References

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021.