Object Extract by YOLOv5 and Style Transfer for Masks in Images

### Wei Gao

### 23015920

### 15/03/2024

## Introduction

This project combines the powerful object detection capabilities of YOLOv5 with cutting-edge style transfer techniques and mask applications, creating a unique pipeline for processing images. By applying artistic style transfer to the objects in an image and accurately masking detected objects, we achieve a blend of artistic expression and object detection.

## Background

The project extracts and segments these elements in a single image by extracting key elements related to the epidemic (e.g., masks) from a dataset of realistic images and applying style migration techniques, using style transfer to give a unique visual language to a specific part of the image that has a representation of the Covid-19, and artistic style migration to the entire image to generate a unique visual representation to recall and document this special period. Yolov5 successfully trains excellent object recognition model using coco dataset, Gayts has also achieved very successful image style conversion work.

## Method

**Buld dataset**

* The required images were collected manually, partly from screenshots of images from official Chinese news and partly from personal photography.
* Poor results were found in the model training phase, and a specialised mask dataset was searched from Roboflow.
* Data labelling of the raw image in Roboflow, where the main job is to frame out the mouthpiece part of the image. The dataset is then divided into training set (70%), testing set (10%), and validation set (20%).

电脑视频的屏幕截图

描述已自动生成

Data label process

* Roboflow was able to download the dataset in yolov5 format, and after labelling was completed the dataset snippet was downloaded directly into the training phase via the API settings. (For your convenience,I have downloaded the data file in the project submission.)

电脑萤幕的截图

描述已自动生成

**Object detect - Yolov5 model training**

The main steps are training and inference using the open source yolov5 model, replacing the dataset with the mask dataset. Using the pre-trained model, the model training process was refined by adjusting the hyperparameters and supplementing the dataset, and here we thank the tutorial for the section on visualising the training effect metrics, which reduces some of the unnecessary workload.

* There were only 62 images in total in the first dataset, and the training results were very poor, with a very low confidence level (0.17) and a recall curve that showed very abnormal results, and inference using other images did not accurately reflect the results.

图表

中度可信度描述已自动生成

* By expanding the dataset to 200 images, the model label was increased from a single "mask" to "no mask", and the yaml file was adjusted accordingly. After that, we adjusted the values of batch and epoch several times, and finally, it is obvious that the confidence level has entered a reasonable interval, and there is also a reasonable recall curve.

图表, 折线图

描述已自动生成 图表, 折线图

描述已自动生成

At the same time, training weights deposited in "best.pt", and inference using the new images yielded more accurate results, although misidentifications still occurred. However, due to running out of colab points and time constraints, refinement of the yolo model was not pursued.

穿着蓝色衣服的人在厨房里

中度可信度描述已自动生成图片包含 文本

描述已自动生成

**Border coordinate extraction**

## This one part is generated using LLM.

## However, this step frequently reports errors “ValueError: tile cannot extend outside image” So I defined the border coordinates manually in order to proceed smoothly with the following section.

**Style Transfer**

Extract the border coordinates and then style convert the coordinate area. This is done using the official Pytorch style conversion tutorial code.

Two ideas are used here. Firstly, an attempt was made to use just the boxed region of coordinates as an input to the style migration model. Firstly, we tried to use only the boxed area as input for the style migration model, but after several attempts, we got several errors about dimension mismatch. The reason for this is that the border area coordinates and the style image are not one size fits all.

I then chose the other method of resizing the whole image to the same size as the styled image, which was much easier than changing the last method, and then styling the whole image, using a mask to just make the part inside the border show through, and finally superimposing it over the original image.

For the style image choose the Picasso style from the tutorial.

## Results

## Model training results

## 图表, 折线图 描述已自动生成图示 描述已自动生成图表 描述已自动生成图表, 折线图 描述已自动生成图形用户界面, 图表, 折线图 描述已自动生成图表, 折线图 描述已自动生成图形用户界面, 图表 描述已自动生成许多照片放在一起 描述已自动生成日程表 描述已自动生成日程表 描述已自动生成

## Style Transfer result

图形用户界面, 应用程序

描述已自动生成

穿蓝色衣服的人在厨房里

低可信度描述已自动生成

## Discussion

The project spent more time than expected on yolov5 training, and by constantly adjusting the batch, epoch, learning rate, expanding the dataset and other methods, we finally achieved more satisfactory training results.

For a better artistic effect, I was looking for some more perfect implementations, this is a ComfyUI integration for Yolov8, which I can refer to later to improve my personal project.

[Comfyui-Yolov8](file:///Users/galway/Downloads/For%20a%20better%20artistic%20effect,%20I%20was%20looking%20for%20some%20more%20perfect%20implementations,%20this%20is%20a%20ComfyUI%20integration%20for%20Yolov8,%20which%20I%20can%20refer%20to%20later%20to%20improve%20my%20personal%20project.)

## Conclusion

The project eventually achieved object recognition and style conversion for specific areas technically, but the artistic performance of the border extraction part and the final style conversion was not as good as it could have been. The project has been more successful in the application of the Yolo model training and the style transformation model, where the model has a reasonable level of confidence and the style transformations can be run successfully and with a reasonable loss.

## Ethical considerations

The data were derived from relevant photography in official Chinese news reports and from my own personal photography during the epidemic, and there was no tracing of photography in the news to its original, original source, and these may have led to problems.

## LLM disclaimer

LLM-Generated Features: Border Extraction and Blending

The methodologies for extracting object borders and blending them seamlessly with stylized backgrounds were developed with the assistance of AI prompts provided to OpenAI's GPT models. These processes leverage advanced algorithms to ensure the integrity of object detection while integrating artistic elements.

I used LLM for the English translation of part of the content paper content.

## Credits

Inspiration: <https://www.noufaljowaysir.com/#/adp/>

YOLOv5 GitHub Repository: <https://github.com/ultralytics/yolov5>

PyTorch Style Transfer Documentation:

<https://pytorch.org/tutorials/advanced/neural_style_tutorial.html?highlight=style%20transfer>

Roboflow: [https://roboflow.com](https://roboflow.com/)

Data supplementation: https://public.roboflow.com/object-detection/mask-wearing/4