Project Update #1 for Custom Trainable YOLO Based Real-Time Object Detection in Animations

JiHo Han
Department of Engineering and Computer Science
University of Michigan
Michigan, USA
jihohan@umich.edu

***Abstract*—In this paper, I would like to show the current progress and health of previously proposed project - an object-detection in animations using You Only Look Once (YOLO) - a scheme for object detection in videos [1] [2].**

***Index Terms*—Real-Time Object Detection, Object Detection, Object Detection in Videos, Animation, YOLO**

I. Implementations and its Process

The proposed project and its deliverables are prepared through following:

1) Read the documentation of YOLO model [3] and clone the repository to use it
2) Choose the animation to work with. As this is a prototyping stage, animation which seems to have
3) Create new, original dataset using Computer Vision Annotation Tool (CVAT) [4]
4) Run a program to randomly divide the dataset into 2 groups - training group and testing group using simple Python program [5].
5) Train the YOLO model, with varying settings. In the best trained model, following settings were used:
   - Types of weight to use for model - YOLOv5-S
   - The size of the input image - 480×640
   - Number of images per batch - 16 images per batch
   - Number of epochs for training - 100 epochs
6) Check the numerical analysis proivded, and if the train seems promising, test with the video which is used to create the dataset
7) Run another test with videos which are new, i.e., videos which were not involved when creating the training dataset
8) Create an analysis and create project update based on the analysis

II. Accomplishments So Far

Training through 100 epochs took approximately 3 hours, and the program was able to successfully detect objects in videos. Each frame took about 40ms in 360p video, i.e., 25 frames per second, which is higher than 23.976 frames per second, average frame rate of animation feed. So far, the result proved to be extremely effective in differentiating 5 different characters in animation *Pororo the Little Penguin* [6].

Fig. 1: Left: 16 images batched as a single image with randomized sections and saturation for training.
Right: Statistics of the inputted dataset.

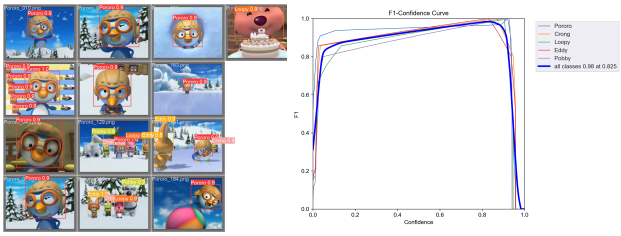

Fig. 2: Prediction made by the model and its F1 Curve

III. Critical Issues and Next Steps

There are no critical issues that needs to be addressed as the prototype showed promising values overall. However, there is a chance that in 3D-animation, the program may be underfitting for the variance in characters' faces aren't widespread. Also, attempting to use this program in 2D animations - which will be my next step - can prove to be difficult compare to what has been currently done so far, and additional action rather than simply training YOLOv5 model using dataset may be needed.

References

[1] Joseph Redmon, Ali Farhadi "*YOLOv3: An Incremental Improvement*" 2018.
[2] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao "*YOLOv4: Optimal Speed and Accuracy of Object Detection*" 2020.
[3] ultralytics. You Only Look Once (YOLO) v5 (Version 6.2) [Computer software]. https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data. 2022.
[4] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT) (Version 2.2.0) [Computer software]. https://github.com/opencv/cvat. 2022.
[5] Luiz Doleron, "*Training YOLOv5 custom dataset with ease* " 2022. https://medium.com/mlearning-ai/training-yolov5-custom-dataset-with-ease-e4f6272148ad
[6] Iconix Entertainment Co., Ltd, "*Pororo the Little Penguin*" 2003∼2021.