Project Update #2 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 add onto 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 2nd part of the proposed project and its deliverables are prepared through following:

1) Choose the animation to work with. This time, 2D animation where same character having varying features should be used to test the program more rigorously. Genres such as fantasy, action, or fictional will suit such requirements.
   - I figured the movie "Howl's Moving Castle" to be perfect for this, as the hair color or age drastically changes throughout the movie, changing the appearance by a large margin. Also, some characters are non-human with opacity, making it challenging for object detection.
2) Create new, original dataset using Computer Vision Annotation Tool (CVAT) [4]
3) Run a program to randomly divide the dataset into 2 groups - training group and testing group using simple Python program [5].
4) 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
5) Check the numerical analysis proivded, and if the train seems promising, test with the video which is used to create the dataset
6) Create an analysis and create project update based on the analysis

Note that the process is quite similar in comparison to prior project update - this is because the input is heavily related to what I am trying to test, rather than its process.
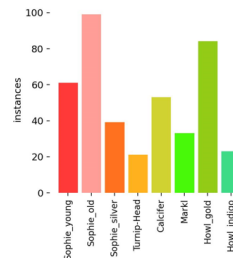
II. Accomplishments So Far

Fig. 1: Distribution of labels in training set

Achieving character detection when the appearance of characters are constantly changing was challenging. My approach to handle such problem was to divide same character into multiple categories with similar appearance.

Training process took 1.5 hours, and the program was able to successfully detect objects in videos. Each frame took about 30ms in 360p video, which is, again, sufficiently high enough in comparison to average frame rate of animation feed. So far, the result proved to be quite effective in differentiating main characters in animation *Howl's Moving Castle* [6], and showed some interesting aspects.

III. Critical Issues and Next Steps

Fig. 2: Example of images where program had hard time recognizing / differentiating the characters

One of the main problems were with the character "Calcifer" - the fire in the aforementioned image. It is noticable that the character is rather transparent and fluid, making the program difficult to recognize the character in certain frames. The main approach of fixing such manners will be through adding additional frames for certain part of animation where character changes drastically.

Another problem was where part of the background character or object have similar characteristics of main character (e.g., similar color, same hat, etc.). I propose this as the limitation of YOLO method in 2D animation, as it is trained to detect the part of characteristics if it's determined to be similar enough, and most of the time, it was.

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] Studio Ghibli, "*Howl's Moving Castle*" 2004.