Complex Character Retrieval from Comics using Deep Learning

Ravi Jain　　Hiroshi Watanabe

Graduate School of Fundamental Science and Engineering, Waseda University

1. Introduction

The field of computer vision has had exponential growth in the past couple of years. Meanwhile, Digital comic books have been replacing printed books. In this paper, we are going to discuss the state-of-the-art deep learning techniques applied for object detection within comics, that can subsequently lead to advanced character-based search within such digital books.

We discuss the application of multi-object detection with You Only Look Once (YOLO) algorithm to retrieve information from comic books in the form of text, frame face, and body. Figure 1 depicts certain issues that the models used to train real-life objects face while trying to detect comic, animated characters.

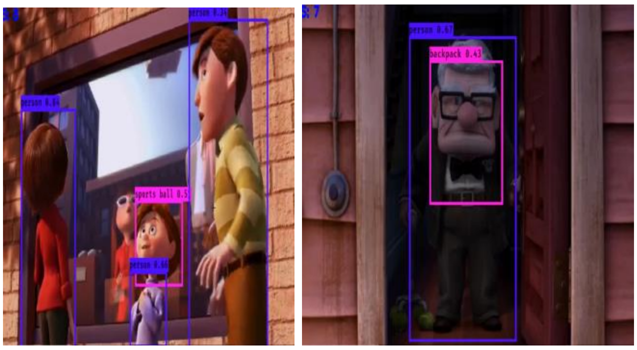


Figure 1: Classification of complex animated character faces by a model trained on real-life objects

The comic characters are not real life like, in these images, the face of one character looks like a backpack, while the other like a sports ball.

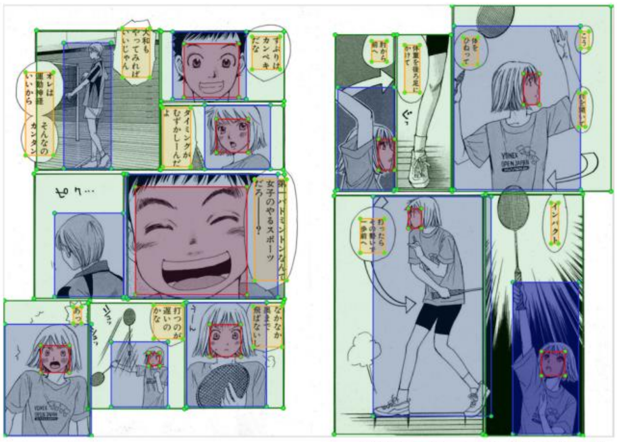
Such issues make object detection within comics a relatively difficult task as compared to real life object detection.

2. Experiment

We train our model on Manga109 dataset consisting of 10,619 images (10,130 with annotations) with labels for text, frame, face, and body of the character.

Figure 2 shows an example of the images in this dataset.

We feed in these images along with their corresponding annotations in the format required by YOLOv3, i.e. <object class> <x_center> <y_center> <width> <height> with float values relative to the width and height of the image. A python script was used to convert the original annotations into the form required by YOLOv3.



YamatoNoHane ©Saki Kaori

Figure 2: Sample Image in the dataset along with corresponding bounding boxes

The Manga109 dataset is then split into 9663 images for training, 956 images for validation (10,130 total, as the cover pages do not have annotations), we provide our model the pre-trained YOLOv3 weights, and training is initialized.

We take some images from each kind of comic and put it in the validation set so that the results are not biased towards one particular comic book.

The model which leads to the minimum validation loss is saved.

3. Evaluation

Table 1 shows the evaluation result using our model after 34,300 iterations.

	Frame	Text	Body	Face
AP (%)	90.45	89.37	85.66	85.21
Thresh	Precision	Recall	F1-Score	
0.25	0.93	0.85	0.89	
Thresh	TP	FP	FN	Avg. IoU
0.25	39748	2949	6747	77.13%
mAP	**87.67%**			

Table 1: Evaluation on Manga109 validation set

The issue with this evaluation is that Manga109 dataset consists of images that include mostly human-like comic characters. So, we attempted to test our model on a more general set of comics rather than manga comics, for this purpose, we use eBDtheque dataset, which has 100 images, and each of them has complex comic characters, with structure not in resemblance to any human. Figure 3 shows one result of our trained model on the eBDTheque and Manga109 dataset. Table 2 shows the average precision values for frames, and body after carrying out evaluation on eBDtheque dataset (we do not provide AP values for face, text, as labels are different for eBDtheque dataset).

	Frame	Body
AP (%)	73.69	43.63

Table 2: Evaluation on eBDtheque Dataset

4. YOLOv3

We use YOLOv3 for training as it is relatively faster than any other object detection techniques, and it has significantly improved compared to previous versions i.e. YOLOv2. Other techniques include detection using deep reinforcement learning of region proposal networks, i.e. drl-rpn, or methods that involve the use of convolutional neural networks to propose regions, such as Faster R-CNN. We might get a higher accuracy with these but the detection speed would be a tradeoff.

YOLOv3 works by convolutional implementation of sliding windows. The accuracy is measured by IoU, which is the measure of the overlap between two bounding boxes. To avoid overlapping objects problem, each object in training image is assigned to a grid cell that contains an object's midpoint and anchor box for the grid cell with highest IoU. To avoid multiple detections of the same object, non-max suppression (NMS) is used, independently carrying out NMS on each of the output classes. For each grid cell, we get a vector which corresponds to as many predicted bounding boxes as the number of anchor boxes, for example, if we have two anchor boxes, then we get two predicted boxes for each grid cell, and then using NMS select the box with the highest IoU.

K-means clustering is used on the training set bounding boxes to automatically find good anchor boxes.

5. Applications

If we develop a robust system, that can accurately retrieve required information from images, then it can lead to advanced search based on character-related input. Figure 4 shows one such example. Such a system would allow users to not go through long lists to find desired images while reading comics.

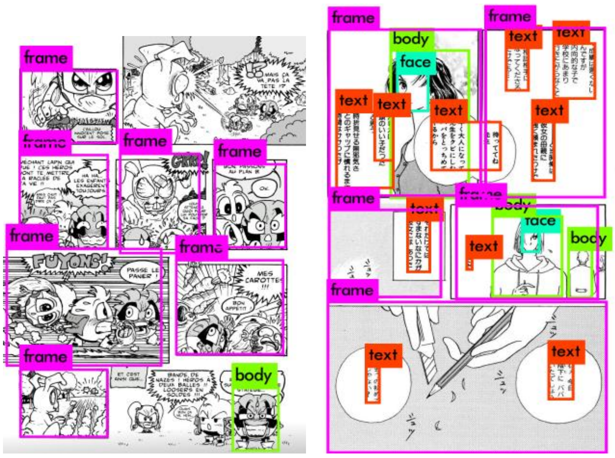
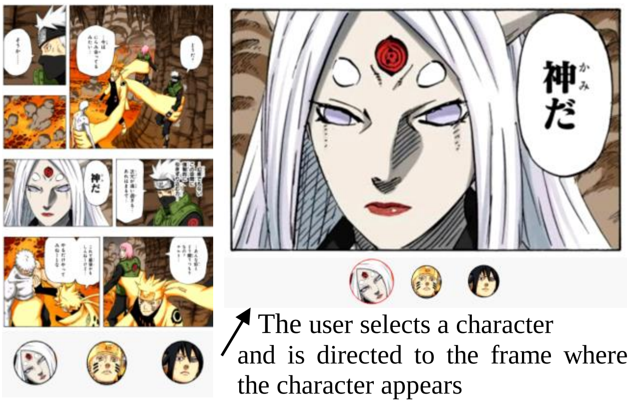


Figure 3: Example of a prediction made by trained model; eBDtheque on the left, Manga109 (one side) on the right



The user selects a character and is directed to the frame where the character appears

Figure 4: Example of an application of a robust character detection system

6. Conclusion

The detection accuracy on human-like comic characters is relatively good compared to the accuracy on object like comic characters.

The eBDtheque dataset is a relatively small dataset, so we do not use it for training. With state-of-the-art object detection techniques, we are able to detect human-like characters with the help of Manga109 dataset, but considering the huge number of complex characters in comic books, one large dataset that includes human-like, object like, animal-like and other complex characters would be ideal for developing a robust system of information retrieval from comics.

7. References

[1] Redmon, Joseph and Farhadi, Ali, "YOLOv3: An Incremental Improvement", arXiv, 2018

[2] Nguyen, N.-V.; Rigaud, C.; Burie, J.-C. "Digital Comics Image Indexing Based on Deep Learning.", *J. Imaging*, 2018, 4, 89

[3] T.Ogawa, A.Otsubo, R.Narita, Y.Matsui, T.Yamasaki, K.Aizawa、"Object Detection for Comics using Manga109 Annotations", arXiv:1803.08670, 2018