Wildlife Detection and Recognition in Digital Images Using YOLOv3

Extended Abstract

Mina Gabriel, Sangwhan Cha, Nushwan Yousif B. Al-Nakash, and Daqing Yun
Harrisburg University of Science and Technology
{mgabriel, scha, nal-nakash, dyun}@harrisburgu.edu

***Abstract*—Recent advances in hardware capability and machine learning techniques enable convenient monitoring of wildlife and their living environments. In this work, we apply Deep Learning (DL) methods to detect and recognize wildlife in digital images and report the experimental results conducted in a commodity workstation. Specifically, YOLOv3 and YOLOv3-Tiny are used to detect and classify several classes of animals based on 9051 digital images and they achieve 75.2% and 68.4% mean average precision, respectively.**

***Index Terms*—Deep learning, YOLOv3, wildlife detection and classification, digital images.**

I. Introduction

Nowadays, digital image datasets overflow from various sources and technologies such as Social Networking Service (SNS) and the Internet of Things (IoTs). The capability of detecting objects in digital images is vitally required by many to enable the developments of their unique applications for being more active and effective in their strategies towards their domain-specific goals. Feature extraction and object detection/recognition are two main steps of such data processing to identify and further analyze the detected objects in digital images. In contrast to the conventional computer vision techniques within the area of machine learning, Deep Learning (DL) techniques can effectively build accurate models for feature extraction, object detection, and image classification, and typically do not require significant human intervention on the problem-specific modeling that is oftentimes prone to oversimplification thus inaccurate in practice.

In this work, we employ the YOLOv3 [1] method with TensorFlow framework to detect and recognize/classify several classes of animals in digital images. Our solution consists of a series of steps that carry out data preprocessing, feature selection, object detection, and wildlife recognition/classification. We first introduce the data collection of the wildlife images used in model training and present representative samples, and then conduct experiments to evaluate the detection/recognition accuracy in a commodity workstation. The experimental results demonstrate the effectiveness of our solution as well as the potential benefits of using learning-based techniques to solve "traditional" problems in a more effective way.

The rest of this paper is organized as follows. In Section II, we describe the dataset of wildlife images and review the algorithms for object detections. In Section III, we conduct experiments and report evaluation results to demonstrate the effectiveness of the proposed DL-based wildlife detection solution. Finally, in Section IV, we conclude our work.

II. A Wildlife Detection and Recognition Solution Based on YOLOv3

A. Wildlife Detection

The importance of wildlife detection and recognition in digital data has been well-recognized in both academia and industry due to its social, economical, and technical significance [2]. For example, timely location and behavior information of wild animals detected from digital images could bring huge benefits to ecosystems conservation [3]. Detecting wildlife from digital images is challenging due to the complexities existed in the dataset caused by various factors such as types of snapshots, weather changes, seasonal changes, and other physical issues regarding data acquisition placements and maintenance.

Conventional object detection algorithms rely on finding features in images to aid in the discrimination of objects that belong or otherwise to particular classes. Such algorithms depend on handcrafted features from images relative to a particular class. As deep learning gets evolved, the Convolutional Neural Network (CNN) has been a dominant solution for object detection. The "You Only Look Once (YOLO)" method is another mainstream approach to provide real-time object detection capability. The processing speed of YOLOv3 [1] shows several times faster than other existing methods for roughly the same accuracy of object detection. In this work, we develop our wildlife detection and recognition solution based on YOLOv3, with Darknet53 [1] additionally incorporated for parameter tuning.

III. Experiments and Results

A. Overview

Fig. 3 highlights the process of our approach, i.e., using YOLOv3 to detect and recognize wildlife from digital images.

B. Data Preparation

The dataset used for training consists of 9051 digital images of several classes of animals, including cat, bird, mouse, and iguana. These digital images are taken by Hyperfire 2 high-definition (HD) cameras [4]. The cameras have an exceptionally quick detection circuitry with daytime color photos and nighttime imaging capabilities and an image resolution of 1080 pixels wide screen and 720 pixel HD video with audio. Fig. 1 shows several representative images of different classes of animals in the dataset. These 9000+ digital images are manually labelled in our experiments.

978-1-7281-8266-7/20/$31.00 ©2020 IEEE
DOI 10.1109/IEEECloudSummit48914.2020.00033

(a) Cat (b) Bird (c) Mouse (d) Iguana

Fig. 1. Example images of four classes of animals in the dataset.

(a) Cat (mAP: 0.8226) (b) Bird (mAP: 0.8502) (c) Mouse (mAP: 0.7462) (d) Iguana (mAP: 0.492)

Fig. 2. Examples of detected and recognized animals corresponding to Fig. 1.

C. Model Training

We merge our dataset together with VOC-2012 [5] to classify 26 classes of objects, 4 of which are for our study (cat, bird, mouse, and iguana). The dataset is then split in the ratio of 80/20 for training/testing. The neural network consists of 106 layers, the input size of the images used is 416×416, the batch size is 8, and the learning rate is 0.001. The model is initialized with the Darknet53.conv.74 [1] training weights.

D. Testbed Configurations

Our experiments are conducted using a commodity workstation equipped with two Nvidia GeForce GTX 1060 GPUs with 3 GB RAM and installed with Ubuntu 20.04 LTS OS. The test dataset includes 1065 digital images, and 4 classes of animals are detected and recognized.

E. Results

We observe a mean Average Precision (mAP) of 75.2% in the detection and recognition of the 4 classes of animals using YOLOv3; and a mAP of 68.4% using YOLOv3-Tiny. The average processing time of using YOLOv3-Tiny is approximately 50% of using YOLOv3. The detected and recognized animals corresponding to Fig. 1 are shown in Fig. 2.

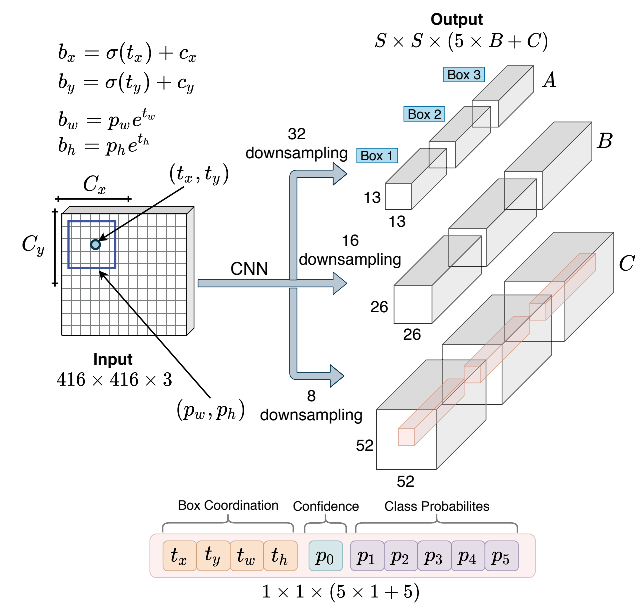

Fig. 3. Illustration of the process of using YOLOv3 [1] to detect and recognize wildlife from digital images.

IV. Conclusion

We applied YOLOv3 to detect and recognize wildlife based on a dataset consists of 9000+ digital images taken by HD cameras. We conducted experiments in a commodity workstation and presented results to show that the applied method achieved mean average precision of 75.2% and 68.4%, for YOLOv3 and YOLOv3-Tiny, respectively.

Acknowledgments

This research is sponsored by the Presidential Research Grant (PRG) of Harrisburg University under Grant No. PRG-2020-05.

References

[1] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," https://bit.ly/2y7UuXj.
[2] Z. Zhao *et al*, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019.
[3] M. Norouzzadeh *et al.*, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proc. of the National Academy of Sciences*, vol. 115, no. 25, pp. E5716–E5725, 2018.
[4] "Hyperfire 2 user manual," https://bit.ly/3g4LthK.
[5] "VOC2012 Development Kit," https://bit.ly/3i6Dwe3.