2022 5th International Conference on Artificial Intelligence for Industries (AI4I) | 978-1-6654-5961-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/AI4I54798.2022.00025

Utilization of Data Augmentation Techniques to Enhance Learning with Sparse Datasets

Richard Yarnell*
Electrical and Computer Engineering
University of Central Florida
Orlando, FL
richardyarnell@knights.ucf.edu

Daniel Brignac*
Center for Research in Computer Vision
University of Central Florida
Orlando, FL
dbrign9@knights.ucf.edu

Yanjie Fu
Department of Computer Science
University of Central Florida
Orlando, FL
yanjie.fu@ucf.edu

Ronald F. DeMara
Electrical and Computer Engineering
University of Central Florida
Orlando, FL
ronald.demara@ucf.edu

***Abstract* — Neural network-based object detection has many important applications but requires a vast amount of training data. In applications where training data may be scarce, data augmentation techniques can be used to expand the training set. This paper explores the performance of such techniques on You Only Look Once Version 5 (YOLOv5).**

Keywords — Machine Learning, You Only Look Once (YOLO)

I. Introduction

Data augmentation expands a training set by creating new and more diverse images based upon the transformation of images in the original training set. This paper explores the ability of various standard image data augmentation techniques to improve the performance of the You Only Look Once Version 5 (YOLOv5) object detection algorithm. All testing is performed using the Commonwealth Scientific and Industrial Research Organization (CSIRO) Crown-of-Thorn Starfish (COTS) Detection Dataset [1], which contains underwater footage of the Great Barrier Reef.

II. Image Augmentation

We evaluate three different image data augmentation techniques: mirroring (reflecting across the central vertical axis), flipping (reflecting across the central horizontal axis), and random cropping while retaining a minimum of 95% of the original image. Each image in the training set is mirrored with probability p_{mirror}, flipped with probability p_{flip}, and cropped with probability p_{crop}. If an image is selected for mirroring and/or flipping, a duplicate image is created, modified accordingly, and added to the training set. When an image is transformed, it is also necessary to update the coordinates of any objects contained therein. During the cropping process, deleted regions are filled with zeros (i.e., filled with black) to preserve the original image size for consistency. If cropping results in more than $^1/_3$ of any bounding box being removed, that box is deleted.

III. Experiments and Results

We use YOLOv5 pretrained on Common Objects in Context (COCO) as our detection model and evaluate on the COTS Detection Dataset. The COTS dataset contains three videos with 6,708, 8,232, and 8,561 extracted frames respectively. We choose to use videos 0 and 1 for training and reserve video 2 for testing. This maximizes the number of object instances that a model can learn from as videos 0 and 1 contain the greatest number of objects. We remove training images that do not contain any object instances reducing the total number of frames in each training video to 2,143 and 2,099 respectively. For testing, we seek to measure the effects of all data augmentations described above (flip, mirror, crop, and combination of all three) versus a training scenario without data augmentations. We evaluate using standard evaluation metrics: precision, recall, mAP, and F1 score.

We perform five experiments regarding data augmentation. The first experiment involves no augmentation, and we train the model only on the data provided. In the second experiment, we perform image flips with $p_{flip} = 0.5$ and expand the training set as described previously. In the third experiment, we mirror each image with $p_{mirror} = 0.5$ and similarly expand the training set. In the fourth experiment, we randomly crop each image (i.e., $p_{crop} = 1$) as discussed above. In the final experiment, we simultaneously perform all the above augmentations on the training set. The goal of these data augmentations is to increase the training data so the model has more examples to learn from and to encourage the model to be position invariant. Results are listed in Table I.

IV. Conclusion

Our results indicate that any individual augmentation technique when implemented alone does not yield substantial improvements. However, when performing an ensemble of augmentation techniques, each image undergoes a different random set of alterations, which leads to greater diversification of the training set, a more position invariant model, and superior performance. This amalgamation may provide a pathway to deploying a working model even when provided with what might otherwise be considered an insufficient training set.

TABLE I. Results

Augmentation	Precision	Recall	mAP	F1 Score
None	0.9082	0.5516	0.6253	0.69
Flip	0.8858	0.5468	0.6356	0.68
Mirror	0.9062	0.5472	0.6179	0.68
Crop	0.9015	0.5092	0.5978	0.65
All	0.9185	0.5655	0.6530	0.70

Acknowledgment

This work was supported in part by the Army Research Office under Grant Number W911NF-20-1-0174. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

References

[1] Jiajun Liu et al., "The CSIRO crown-of-thorn starfish detection dataset," [Online]. Available: https://doi.org/10.48550/arXiv.2111.14311.

* Equal contribution

978-1-6654-5961-7/22/$31.00 ©2022 IEEE
DOI 10.1109/AI4I54798.2022.00025