Object Instance Balancing Through Extended Mosaic Augmentation

Zafar Aziz, Odilbek Urmonov, Ali Haroon, HyungWon Kim*
ChungBuk National Univ.

Abstract

Augmentation strategies are key in training a well-generalized object detection model with scarce datasets. The available annotated object detection datasets are not only scarce but are also naturally unbalanced. The unbalanced nature of these datasets leads to poor or incorrect learning of minority class samples in the data. Hence, the resulting model suffers from a confirmation bias problem and performs poorly on test data. We propose an improved mosaic by combining a simple copy-paste technique with the existing mosaic augmentation strategy. Our proposed mosaic strategy alleviates the number of minority class samples in the data. In this method, we copy-paste available augmented ground truth minority samples into the mosaic with severe unbalance. The resultant training data now have more minority class samples to learn a much better and more accurate representation of these categories.

Ⅰ. Introduction

Object detection is a crucial computer vision task, providing foundational visual cues for scene understanding [1], and object recognition [2]. State-of-the-art models for object detection are trained using supervised learning. Hence, their effectiveness is influenced by the volume of labeled training data.

Additionally, real-world detection datasets often exhibit class imbalance. Hence, leading the models trained on such datasets to potentially overfit and exhibit poor generalization to test data examples belonging to minority classes [3]. Existing augmentation techniques fail to address the imbalance of minority class samples, a straightforward solution to the confirmation bias problem [4].

In this study, we extend existing mosaic augmentation [5] by combining it with a simple copy-paste strategy [3].

Ⅱ. Proposed Method

We propose a simple extension to mosaic augmentation that ensures at least one sample from each object category in every mosaic image. By training a model on this enhanced dataset, we expect to achieve a significantly improved representation learning of all the classes.

The proposed extended 4-Mosaic method is shown in Fig. 1. The details are provided in the sections below.

A. Sample Memory Bank Creation

We propose to augment the existing mosaic with transformed ground truth (GT) samples to address class imbalance. To achieve this, we create a memory bank by cropping GT objects from the original dataset and storing them for later use. By incorporating these diverse GT samples with adequate transformations into the mosaic augmentation process, we aim to improve the model's ability to detect the objects from underrepresented classes.

B. 4-mosaic to 6-mosaic

Generally, the Yolo family of models uses mosaic with 4 grids. In the first stage, we propose to extend 4 grids to 6 grid mosaic. Hence our method makes 6 grids instead of 4 as in the conventional 4-mosaic.

C. Conventional Mosaic Step

In the next step, we propose a slightly altered conventional 4-mosaic step. We perform 4-mosaic step but contrary to default setting during paste operation we select a random grid in the mosaic. Hence, we fill out random four grids out of six with selected images (four images). Two grids are left as it is for class-balancing.

D. Class Balancing

In the final stage, we count the samples of each category to select minority classes in the current mosaic. Finally, we fill the remaining two grids by copy-pasting minority class samples from the memory bank with adequate transformations.

Ⅲ. Experimental Results

To evaluate the effectiveness of proposed extended mosaic method, we perform experiments using a subset of BDD dataset [6] containing 15000 training images. We used full validation set for evaluation. We also consolidated the BDD object classes to only three important categories (vehicle, person and rider class). We opted Yolov5 [7] detector for our experiments. The mosaic augmentation in Yolov5 training is replaced with our extended mosaic method. The initial experimental results are provided in the Table I.

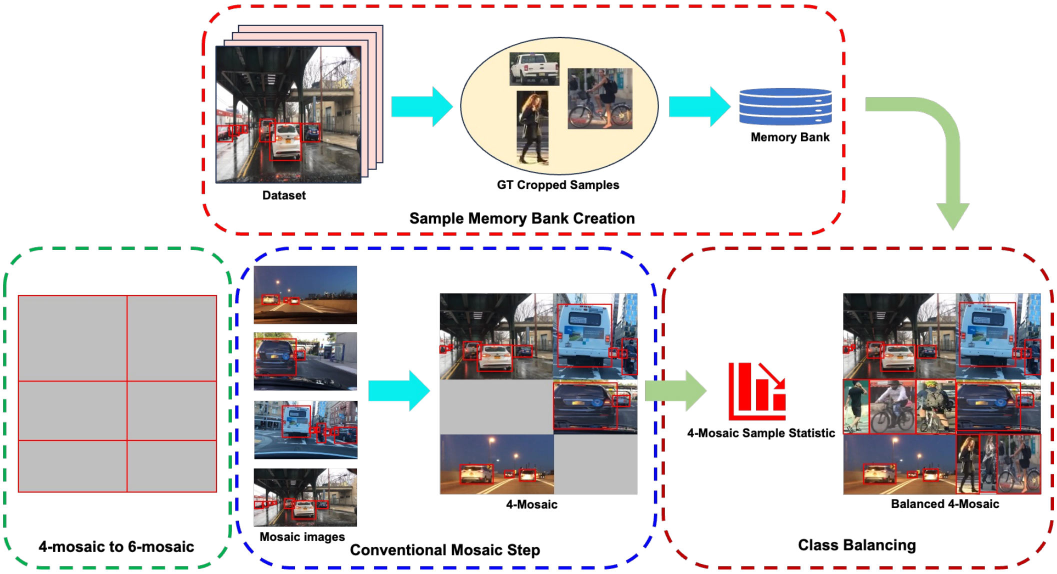

Fig. 1: Proposed extended 4-mosaic with minority class sample copy-paste strategy for class-balancing

TABLE I
STATISTICS COMPARISON BETWEEN CONVENTIONAL AND EXTENDED MOSAIC METHODS

Class-wise performance (mAP@50) and sample statistics comparison of conventional and extended mosaic				
Classes	Conventional mosaic (4-mosiac)		Extended mosaic (6-mosaic)*	
	samples	mAP@50	samples	mAP@50
Vehicle	12256	64.9	12951	65.1
Person	2716	44.9	7792	49.2
Rider	295	18.4	6636	34.7

IV. Conclusion

In this study, we propose an extended mosaic augmentation method with copy-paste strategy for alleviating the number of minority class samples. In our method, we copy-paste the minority class samples from existing ground truth data with adequate transformation into the mosaic image. The resultant data have boosted minority class samples helpful in better training the detector.

ACKNOWLEDGMENT

This Work was supported by Regional Leading Research Center (RLRC) of the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No.2022R1A5A8026986), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01304, development of Self-Learnable Mobile Recursive Neural Network Processor Technology), and the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Communication Technology Research Center support program (IITP-2024-2020-0-01462) supervised by the IITP.

References

[1] R. Surendran and D. Jude Hemanth, “Scene Understanding Using Deep Neural Networks—Objects, Actions, and Events: A Review,” in International Conference on Innovative Computing and Communications, vol. 1087, A. Khanna, D. Gupta, S. Bhattacharyya, V. Snasel, J. Platos, and A. E. Hassanien, Eds., in Advances in Intelligent Systems and Computing, vol. 1087. , Singapore: Springer Singapore, 2020, pp. 223–231. doi: 10.1007/978-981-15-1286-5_19.

[2] D. Balamurugan, S. S. Aravinth, P. C. S. Reddy, A. Rupani, and A. Manikandan, “Multiview Objects Recognition Using Deep Learning-Based Wrap-CNN with Voting Scheme,” Neural Process Lett, vol. 54, no. 3, pp. 1495–1521, Jun. 2022, doi: 10.1007/s11063-021-10679-4.

[3] Z. Aziz and H. Kim, “Copy-paste Augmentation for better saliency learning of underrepresented Object Classes,” in 2023 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Busan, Korea, Republic of: IEEE, Oct. 2023, pp. 1–4. doi: 10.1109/ICCE-Asia59966.2023.10326350.

[4] J. Yang et al., “Divide to Adapt: Mitigating Confirmation Bias for Domain Adaptation of Black-Box Predictors,” 2022, doi: 10.48550/ARXIV.2205.14467.

[5] P. Kaur, B. S. Khehra, and Er. B. S. Mavi, “Data Augmentation for Object Detection: A Review,” in 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA: IEEE, Aug. 2021, pp. 537–543. doi: 10.1109/MWSCAS47672.2021.9531849.

[6] F. Yu et al., “BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning,” 2018, doi: 10.48550/ARXIV.1805.04687.

[7] G. Jocher et al., ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. (Nov. 22, 2022). Zenodo. doi: 10.5281/ZENODO.3908559.