On synthetic and real images as training data for object detection - A brief review

Martin Georg Ljungqvist
Axis Communications AB / Lund, Sweden
martin.ljungqvist@axis.com

Abstract

To train neural networks, sufficiently large and diverse datasets are needed. To address this, the use of synthetic data has become popular because it is inherently scalable and can be automatically annotated. A brief overview of recent work on using synthetic and real images as training data for object detection is presented in this paper, with a focus on mixing real and synthetic training data. The trend is that having real data and adding some amount of synthetic data helps the performance in many studies. It was concluded that there appears to be no consensus on the ratio of real and synthetic image data.

1 Introduction

Presented here is a brief overview of selected articles on using both real and synthetic data for object detection, the papers reviewed here cited the article by Ljungqvist et al. (2023). The overview focused on mixing real and synthetic training data.

Synthetic data here refer to images produced entirely in a computer, while real data here refer to images capturing a natural scene by a camera. Data augmentation can be applied to both synthetic and real data.

Since synthetic images and real images come from different distributions due to sensor noise, texture and many other factors, there is a domain gap and the results of training on only synthetic data and testing on only real data can produce much lower results than if training on real data (Ljungqvist et al., 2023). However, real and annotated data might not always be available in large volumes and therefore mixing real and synthetic data can gain performance.

These are the papers covered:

- Searching for the Ideal Recipe for Preparing Synthetic Data in the Multi-Object Detection Problem (Staniszewski et al., 2025)
- In the Search for the Balance Between Real and Synthetic Images in Multi-Class Detection Systems (Cecchetti et al., 2024)
- Ai-generated images as data source: The dawn of synthetic era (Yang et al., 2023)
- Bridging the gap: Active learning for efficient domain adaptation in object detection (Menke et al., 2024)

2 Mixing real and synthetic data

2.1 Searching for the Ideal Recipe for Preparing Synthetic Data in the Multi-Object Detection Problem

In Staniszewski et al. (2025) the authors generated images of 3D synthetic objects in 6 categories aimed at public transport scenarios: bicycles, trolleys, wheelchairs, boxes, suitcases, and bags.

The synthetic objects were positioned on a background collage of randomized real images from news, sports, etc.

For real data a mix of COCO and other images was used; 3 of the classes were in the COCO dataset - the data were henceforth complemented with images from a dataset for public transportation and the Internet.

The synthetic images were only added to the training set. Data augmentations were used.

They used a YOLOv7 object detector pretrained on COCO and trained it with the backbone frozen. Freezing the backbone while training on synthetic data is motivated by the study of Hinterstoisser et al. (2018). However, Tremblay et al. (2018) showed promising results for unfrozen backbone using domain randomized synthetic data. Ljungqvist et al. (2023) reported that

From Electrophoresis to Wikidata: Festschrift in honor of Pierre Nugues, pages 16–19.
Published by University of Tartu Library, May 2025.

no particular difference could be seen in performance for frozen or unfrozen backbone. It seems that there is not a consensus whether freezing the backbone or not is preferred for training on synthetic data.

It should be noted that this study claimed that synthetic training data "significantly improves results", however there was no mention of multiple trainings, t-test, or p-values which are needed to test statistically significant differences. However, they used cross-validation for model selection.

They trained the detector on a mix of real and synthetic data using different ratios between real and synthetic training data.

Using only synthetic and no real data gives very low performance. Adding only 5% real data gives a tremendous boost.

The results showed a trend that adding synthetic data clearly improved the results. They observed a boost in performance when adding 25% or more synthetic data in relation to the amount of real data. Soon after 25% added synthetic data the performance flattened out.

Having real data and adding any amount of synthetic data increased the performance of all metrics in this study. Adding synthetic data did not decrease the performance except for some cases of the objectness metric. The objectness and classification metrics had more variance compared to the other metrics.

The best results without transfer learning were when having slightly more synthetic than real data, the best mean average precision at 0.50 intersection of union (mAP50) was for 1 real and 1.75 synthetic data, but it varied for different metrics between 1:1 and 1:2.

With transfer learning, the best mAP50 was for the ratio of 1 real and 0.75 synthetic data, but it varied for different metrics between 1:0.25 and 1:1.25.

In this study, the influence of the amount of synthetic data relative to real data was concluded to be minimal. However, it is worth noting that synthetic data can be a good replacement and complement to real data when real data is scarce and there is a need to balance data from multiple classes.

This study reported interesting results on the topic of mixing real and synthetic image data. However, the study had limitations such that only one synthetic dataset was used, and there was no statistical evaluation of differences in the results, such as error bars or t-test. Furthermore, balancing multiple classes may not always be strictly needed.

2.2 In the Search for the Balance Between Real and Synthetic Images in Multi-Class Detection Systems

A literature review on mixing real and synthetic images in 27 articles from the past 5 years was performed in Cecchetti et al. (2024).

Only three of the articles in the review presented a specific ratio; however, they did not present a comparison between the different ratios and their impact on object detection performance.

This literature review observed that most of the articles did not establish a relationship between the use of real and synthetic images. They conclude that there is a lack of in-depth analysis of the ratio relationship of mixing real and synthetic data for multi-class detection tasks. Hence, there is currently a scientific gap in terms of the number of synthetic images used to train object detection models.

It should be noted that the Staniszewski et al. (2025) study was published after this literature review was performed.

2.3 AI-generated images as data source: The dawn of synthetic era

In the review paper Yang et al. (2023) results were included for object detection results from Ge et al. (2022), and extended in Ge et al. (2023), for Pascal VOC and COCO detection reporting mAP50 and mAP. They used Faster RCNN with ResNet-50 backbone. Combining real and synthetic images (copy paste and foreground) resulted in higher mAP50 and mAP for both Pascal VOC and COCO.

They combine real and synthetic data in various ways using either synthetic foreground or background, so there is no specified ratio of distinct real and synthetic images.

On Pascal VOC, a model trained solely on synthetic data in the absence of any real images achieves comparable performance to a model trained on 1,464 real images. Furthermore, adding synthetic backgrounds when blending both real and synthetic led to the best performance.

For the COCO dataset, mixing both synthetic and real gave the greatest performance boost.

Hence, the conclusion is that mixing real and synthetic data gives a boost.

2.4 Bridging the gap: Active learning for efficient domain adaptation in object detection

Menke et al. (2024) combined real and synthetic data for training using active learning for domain adaptation. They used the synthetic sim10k dataset as the source domain and the natural images in the Cityscapes dataset as the target domain, in a synthetic-to-real domain adaptation setting. The aim was to adapt the object detection models trained on sim10k to perform well on Cityscapes. The objective was to select a portion of the target domain for labeling (aiming to outperform unsupervised domain adaptation methods that rely solely on unlabeled target data).

Sim10k contains 10,000 images from the computer game Grand Theft Auto 5, the paper did not specify a split for these data, so it is assumed that they used all for training. Cityscapes contains 2975 labeled images in the training set, and they used up to 25% of this for training, which results in approximately 743 images, which is about 7% of 10,000.

The mix of synthetic and real training data is performed using an active learning approach; real images from Cityscapes are selected for training with the aim of complementing the source domain (sim10k). This was performed by scoring the images using a scoring function to decide if the image should be used for training. In this study, three different scoring functions were proposed and evaluated. They used 25% of the target domain training images.

They show mAP performance on the percentage of labeled images (from 5% to 25%) and the number of sampled boxes (from about 1,000 to 12,000). The absolute mAP increased with increasing number of labeled target images and with the number of labeled boxes. This is interpreted in such a way that they used a fixed amount of synthetic training data, and for an increasing amount of real images the performance increased. The amount of real data was in all ratios smaller than the synthetic amount. Using 25% of the target data, the real data consists of about 7% of the amount of synthetic data.

Using 25% labeled target domain images, there was a 2.43 mAP improvement in object detection performance over the random baseline.

3 Discussion and Conclusion

From this brief literature review, it was concluded that there seems to be no consensus on the ratio of real and synthetic image data.

None of the studies had a statistical evaluation of differences in the results, such as standard deviation, error bars, or t-test.

The trend is that having real data and adding some amount of synthetic data helps the performance in many studies.

There is probably no ratio that works for everything; it could depend on factors such as:

- Data quality
- The class types
- Variations within the classes
- Domain gap factors
- Amount of real data available
- How the synthetic data is generated

And other factors may likely come into account here. There is no one-size-fits-all for mixing synthetic and real images as training data.

Acknowledgments

The author wishes to thank colleagues for valuable discussions on the topic.

References

Vitória Biz Cecchetti, Marcelo Rudek, and Roberto Z. Freire. 2024. https://doi.org/10.1109/ICIEA61579.2024.10664977 In the search for the balance between real and synthetic images in multi-class detection systems. In *2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA)*, pages 1–6.

Yunhao Ge, Jiashu Xu, Brian Nlong Zhao, Neel Joshi, Laurent Itti, and Vibhav Vineet. 2022. http://arxiv.org/abs/2206.09592 Dall-e for detection: Language-driven compositional image synthesis for object detection.

Yunhao Ge, Jiashu Xu, Brian Nlong Zhao, Neel Joshi, Laurent Itti, and Vibhav Vineet. 2023. http://arxiv.org/abs/2309.05956 Beyond generation: Harnessing text to image models for object detection and segmentation.

Stefan Hinterstoisser, Vincent Lepetit, Paul Wohlhart, and Kurt Konolige. 2018. http://arxiv.org/abs/1710.10710 On pre-trained image features and synthetic images for deep learning. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Martin Georg Ljungqvist, Otto Nordander, Markus Skans, Arvid Mildner, Tony Liu, and Pierre Nugues. 2023. Object detector differences when using synthetic and real training data. *SN computer science*, 4(3):302.

Maximilian Menke, Thomas Wenzel, and Andreas Schwung. 2024. https://doi.org/https://doi.org/10.1016/j.eswa.2024.124403 Bridging the gap: Active learning for efficient domain adaptation in object detection. *Expert Systems with Applications*, 254:124403.

Michał Staniszewski, Aleksander Kempski, Michał Marczyk, Marek Socha, Paweł Foszner, Mateusz Cebula, Agnieszka Labus, Michał Cogiel, and Dominik Golba. 2025. https://doi.org/10.3390/app15010354 Searching for the ideal recipe for preparing synthetic data in the multi-object detection problem. *Applied Sciences*, 15(1).

Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. 2018. http://arxiv.org/abs/1804.06516 Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop on Autonomous Driving*.

Zuhao Yang, Fangneng Zhan, Kunhao Liu, Muyu Xu, and Shijian Lu. 2023. http://arxiv.org/abs/2310.01830 Ai-generated images as data source: The dawn of synthetic era.