Investigating the Performance of Computer Vision Models in Terms of Object Counting and Localization When Trained on Synthetically Generated Data

As training data becomes increasingly valuable to firms racing to build ever potent neural networks, it continues to be expensive to manually collect and annotate large datasets, especially those consisting of visual media. One solution to this is to generate synthetic data, a process claimed to be "one of the most promising general techniques on the rise in modern deep learning" (Gerard).

In short, synthetic data is data that has been artificially produced in an attempt to mimic its real-world counterpart, with the aim to extend and fill gaps in naturally acquired datasets. This paper seeks to extend the research on synthetic data, particularly in relation to object counting and localization in computer vision, with more diverse datasets. To do so, three unrelated, natural datasets with different sizes and complexities were chosen, as shown in Figure 1. Next, an algorithm to produce synthetic replicas was developed based on the parameters of real samples, also visualized in Figure 1. Finally, a U-Net neural network was trained on different proportions of natural to synthetic images, and its ability to quantify the number of objects (object counting) and identify their positions (object localization) was evaluated.
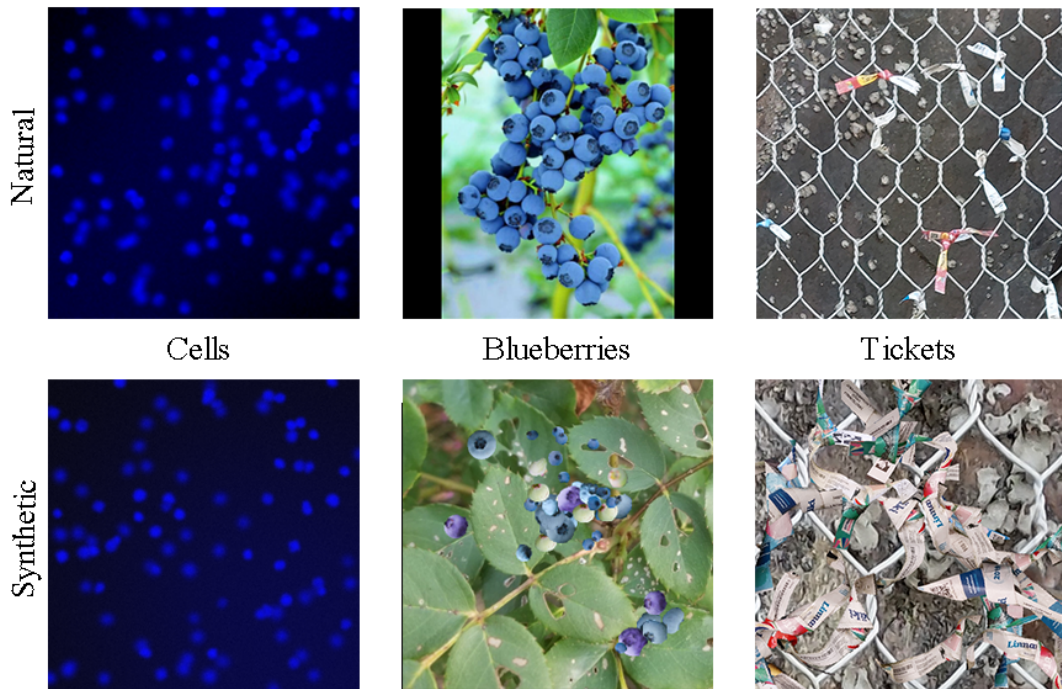

Figure 1: An example of what the natural and synthetically generated images from each dataset look like.

Overall, it was found that in cases where natural data was incomplete, synthetic data could provide substantial performance improvements. However, otherwise or in high concentrations, it could also hinder the model by overly 'diluting' the training set. Thus, the findings converge on the fact that artificial data is a viable method of improving the performance of a computer vision model in terms of object counting and localization, provided that the initial dataset is scarce, and the generated data closely corresponds to the one used for validation.