EE417 - Assignment #2

Custom Data Creation and Object Detection

April 2025

1 Introduction

Object detection is one of the most active research areas in computer vision, with a wide range of real-world applications such as autonomous driving, surveillance, healthcare, and industrial automation. The fundamental goal is to identify and localize objects of interest within an image. Despite the impressive advances in recent years, the performance of object detection models still heavily depends on the quality and size of the dataset used during training.

In many real-world scenarios, ready-to-use, annotated datasets are either unavailable or do not match the specific requirements of the task at hand. This creates a need for custom dataset creation, where users collect their own data, annotate it, and apply various data augmentation techniques to increase its diversity and robustness. Although this process can be time-consuming and labor-intensive, it allows for fine-tuned models that are tailored to the exact characteristics of the problem being solved. Although offline augmentation used to be popular in the early days of deep learning, nowadays online augmentation is the norm.

The objective of this assignment is to simulate such a real-world scenario. *You are asked to create a small custom image dataset*, annotate the objects of interest, and apply various online data augmentation methods, and train an object detection algorithm using the resulting dataset both from scratch and through fine-tuning. This hands-on experience will help students understand the challenges of dataset creation and give them insights into the complete pipeline of an object detection system — from raw data to model evaluation.

2 Task

2.1 Stage 1 - Dataset Creation (25 points)

Choose two everyday objects from your dorm room or elsewhere and take **at least 25 photos** of each object. Make sure to take these photos from different angles and distances. Do not use photographs other than your own!

2.2 Stage 2 - Labeling the dataset (5 points)

You can label the photographs with a labeling tool of your choice (Labelimg [1], LabelBox[2], CVAT[3], etc.). You need to be precise during labeling.

2.3 Stage 3 - Augment the dataset (20 points)

You must apply the online augmentations specified below.

2.3.1 Flip and rotate

You need to apply vertical and horizontal flips and rotate the original image ± 45 degrees for each side.

2.3.2 Sharpening and blurring

You need to sharpen and blur the images

2.3.3 Brighten and Darken

You need to brighten and darken the original images.

2.3.4 Hue augmentation

Adjust the hue of the images.

2.3.5 Saturation augmentation

Adjust the saturation of the images.

2.4 Stage 4 - Train models (30 points)

You are expected to train 2 object detection architectures of your choice — one must be a **single-stage detector** and the other must be a **two-stage detector**. Each of them will be trained in 2 modes: from scratch and with pretrained weights. Split your data into 60% training, 20% validation, and 20% testing.

2.4.1 Random-init

This mode is from scratch, meaning that you'll start with random weights. Be careful how you initialize them. Remember our discussion in class.

2.4.2 Fine-tune

This one means that you'll use pretrained models, and keep on training them with your dataset.

2.4.3 Evaluation Metrics

You need to provide precision, F1-score, mAP, and a confusion matrix for each of the 4 training sessions.

2.5 Stage 5 - Infer the models (10 points)

After training the object detection models, you are required to evaluate their performance on a separate set of unseen (i.e., test) images. For each image, the model should generate predictions in the form of bounding boxes and corresponding class labels. These predicted bounding boxes must be visually overlaid on the images, with the class name displayed on or near each box to indicate the identified object. This step not only facilitates qualitative analysis of the model's performance but also provides a clearer understanding of its strengths and potential failure cases.

2.6 Stage 6 - Reporting (10 points)

During the dataset preparation phase, you are expected to document the criteria and considerations that guided your data collection process. This includes explaining any challenges encountered (e.g., class imbalance, poor lighting conditions, occlusions, etc.) and the strategies adopted to address them. Additionally, you should report the resolution of the images used in your dataset, as it can significantly influence model performance and computational requirements.

Furthermore, you are required to clearly present the methods and models you experimented with, including the training hyperparameters (e.g., learning rate, batch size, number of epochs, optimizer type). A comparative analysis of the performance of different models should also be provided, highlighting both quantitative metrics. (from scratch VS fine-tune)

Please ensure that the report includes visual examples of the outputs generated through your data augmentation process. In addition, you are required to provide a Google Drive link to your prepared dataset. Both your dataset and the applied augmentation techniques will be subject to evaluation. Make sure that all cited references are accurate, relevant, and properly formatted.

3 What to submit

You are expected to submit

- the notebook with your code (filled version of the given notebook),
- a PDF report prepared in LaTeX of at most 3 pages (IEEE conference format),
- each submitted file must be named Name_Surname_000StudentNumber.

explaining your pipeline and the results that you have obtained.

4 Rules and penalties

In addition to the assignment and plagiarism rules outlined in the course syllabus:

- the notebook must be ready to run, with no errors. The teaching assistant will not solve your compilation/syntax/... errors. If it doesn't run, it will get a grade of ZERO.
- in order to get the grades of each stage, you need to provide both the working code and the respective explanations of how you solved it in your report.
- incorrect file naming will lead to point deductions.
- submissions that do not adhere to the specified reporting format (LaTeX using the IEEE style) will be subject to point deductions.
- notebook must contain a Google Drive link to your dataset.

5 Important notes

- Please use the notebook in SuCourse to fill in your solutions.
- You are free to utilize any libraries or frameworks of your choice, as long as they are appropriately documented and their usage is clearly justified within your implementation.

References

[1] HumanSignal, "Labelimg," https://github.com/HumanSignal/labelImg, [Accessed 11-04-2025].

[2] LabelBox, "Labelbox," https://labelbox.com/, [Accessed 11-04-2025].

[3] I. OPENVino, "Leading data annotation platform," https://www.cvat.ai/, [Accessed 11-04-2025].