PyTorch Project Proposal

Wassim Mecheri

March 2025

1 Project Description

My project idea is a cash detection assistant. Using a smartphone's camera, the application detects money, counts it accurately, and outputs the result on the screen, or using a voice system upon user request.

2 Use Cases

- This assistant would be of great help for visually impaired or blind people, who often use techniques such as feeling the size of the money or sorting methods to count it.
- It could also be used for elderly people who often experience deteriorating vision due to age and sometimes need a little help to be autonomous in their everyday lives.
- Another use case is for people visiting a new country. They may not be familiar with the local currency, and if the model is trained on multiple currencies, the application could output the value in the local currency along with a conversion to their home currency for reference.
- In general, this application could be used in any country where the banking system is not well developed and cash transactions remain the primary mode of payment, such as Algeria, for example.

3 Approach Idea

Taking inspiration from a project that detects the calories on a plate of food:

- I would first need a dataset of images of money. An interesting option would be a dataset featuring different currencies, with pictures of money in real-life situations (as opposed to images of perfectly flat money, which rarely occur in real life). Alternatively, I could create my own dataset by taking a few bills and coins and arranging them in various configurations to take pictures.
- We store all these images in a database.
- For object detection, I would use YOLO initially, as it is a common choice for this type of task. If time allows, I could try to build my own version of YOLO from scratch as an experiment. The results would likely be worse, but that does not really matter.
- We would then pass the results from YOLO to an LLM to process the information and output the desired result: the current cash value and, if needed, a conversion to the user's preferred currency.
- Finally, we could convert the LLM output into sound using a text-to-speech model.