I recently worked on a computer vision project for a client who wanted to automate product identification in a retail environment. The objective was to build an AI-powered system that could seamlessly identify and bill products, eliminating the need for manual cashier operations.

The project started with data collection, where we were tasked with capturing images continuously for a week. We developed custom code to capture frames every few minutes, resulting in a large dataset. This raw data was then prepared for the next stage, which involved annotation and labelling.

Given the massive amount of data, with over a lakh images to label, we decided to outsource this task to a third-party data labelling company. They used tools like LabelImg and LabelMe to annotate the images, marking each product with bounding boxes. Once we received the labelled data, we conducted a thorough quality check to ensure that the annotations were accurate.

With the labelled dataset in hand, we moved on to data preprocessing. The goal was to standardize and enhance the dataset to ensure consistency. We applied various preprocessing techniques, such as resizing images, normalizing them, and enhancing the contrast. We utilized tools like cloDSA and Augmentations for these tasks, ensuring that the dataset was clean and ready for the model training phase.

To further improve the dataset, we performed data augmentation, which involved generating additional images through various transformations. Starting with 5,000 images, we used augmentation techniques like rotation, scaling, flipping, and brightness adjustments to expand our dataset significantly. This resulted in a robust dataset of up to 80,000 images, which was essential for training a deep learning model.

For model selection, we chose YOLOv8 due to its balance of speed and accuracy, making it ideal for real-time product identification. We initialized the model with pre-trained weights from Ultralytics, which were trained on the COCO dataset, a collection of common objects. We then fine-tuned the model on our custom data, training it for 3,000 to 4,000 epochs. This training process took an entire day, but the results were promising, with the model accurately identifying products in the images.

Once the training was complete, we evaluated the model's performance using metrics like mean Average Precision (mAP). This allowed us to gauge how well the model was performing in terms of accuracy and detection. If the results were not up to the mark, we performed fine-tuning by adjusting hyperparameters, experimenting with different data augmentation techniques, and even retraining the model if necessary.

Finally, we deployed the trained model on an AWS EC2 instance. The deployment setup allowed the system to process images in real-time, identifying products and providing outputs such as the image ID, bounding box coordinates, and the S3 bucket URL where the image was stored.

This project was a significant step toward revolutionizing the retail checkout experience. By leveraging advanced computer vision techniques and deep learning, we were able to build a system that could automatically identify and bill products, greatly improving efficiency and customer satisfaction in retail environments.