Image classification and Object Detection – A Big Data Approach

Ali Mir

12/17/20

Project Business Case:

Image classification and object detection

Purpose of the pipeline

Image classification has been of an interest to different industries because of numerous valuable insights that a company can acquire by automatically labeling different images based on their filtering criteria.

A few examples would be:

- Harmful object detection of scanned luggage at the airport security
- Automatic flagging of medical images for potential diseases
- Automation of quality control on a production level at the assembly line of a particular product
- Rapid automatic sourcing of different goods

Benefit of the pipeline

Image classification and object detection may look like a very challenging concept to different companies; and it might very well be. However, the development of Big Data ingestion and processing tools is providing new and unimaginable opportunities to developers and companies to be benefited from the “goods” of using an image classifier while avoid most of the “bads”.

The purpose of this pipeline is to employ variety of Big Data tools and provide a scalable, understandable, and custom designed image classification and object detect platforms for clients’ needs.

While maintaining the standards of a reliable object detection platform, this pipeline’s aim is to make is as accessible to clients without in-depth knowledge of machine learning or data manipulation as possible.

Data Source

Zenserp – Google images

Tools

Kafka – Spark – MongoDB/Hive – AWS S3 – AWS SageMaker (For model training purposes, depending on the pricing and available resources, the last step of the pipeline could be replaced by local-training)

Project Timeline

Monday – 12/7: Data Ingestion

Wednesday – 12/9: Data Manipulation and storage in MongoDB/Hive

Saturday – 12/12: Data export to S3

Tuesday – 12/15: Model Training, Model preparation

Wednesday – 12/16: Cleanup – Model optimization – Hyperparameter tuning