Case Study: Video Stream Processing For Traffic Sign Classification

We selected a representative traffic management system case study following road safety inspection concerns [1]. Detecting and recognizing different traffic signs and anomalies in nearly real-time requires fast detection of objects in video frames and embedding the information on the detected objects in video streams at different encoding resolutions and bit rates [2]. Typical examples are broken, covered, worn-out or stolen traffic signs, or incorrectly painted road surface markings [3]. We represent this application as a DAG of eight stages depicted in the figure. Each independent stage contains a data store and communicates with other stages through a lightweight HTTP interface and a message queuing system [4, 5].

1) Encoding stage: receives and encodes the raw video stream in high resolution and bitrate near to the vehicles equipped with multiview-cameras. We use for this purpose the ffmpeg software suite [6] with the H.264 video codec for encoding, transcoding and packaging of the video streams.

2) Framing stage: utilizes OpenCV to produce still frames from different video scenes [7].

3) Dataset storage stage: provides the stored records such as 50,000 video frames of 43 different traffic sign classes [8] for other stages.

4) Low-accuracy inference stage: identifies features in the video stream, such as traffic signs on the road. This stage aims for a low classification accuracy of 70%.

5) High-accuracy inference stage: uses a machine learning model [9] capable of accurate inference when the low-accuracy stage has a poor confidence. We use a trained model of convolutional neural network with nine layers to classify the signs in the same video frames as for the low-accuracy inference.

6) Low-accuracy training stage: updates and retrains the multi-class classification model with a low accuracy until reaching a 70% accuracy. The stage uses Tensor Flow core version 2.3.0 for Python v3.7.4 to train a convolutional neural network with nine layers on localized signs from 50,000 video frames of various traffic sign classes. Every frame contains a traffic sign used for training and testing the neural network.

7) High-accuracy training stage: updates and retrains the multi-class classification model to learn from newly collected data [10]. We use the same convolutional neural network with nine layers to classify the signs. This stage is the upstage of the high-accuracy inference stage.

8) Packaging and delivery stage: provides the transcoded video stream together with the detected signs in the format required by the drivers. This stage is the downstage of multiple upstage services, and receives the data from inference and training stages.

REFERENCES

[1] Narges Mehran, Dragi Kimovski, and Radu Prodan. A two-sided matching model for data stream processing in the cloud – fog continuum. In 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud, and Internet Computing (CCGrid), pages 514–524. IEEE, 2021.

[2] Siniša Šegvić, et al. Exploiting temporal and spatial constraints in traffic sign detection from a moving vehicle. Machine vision and applications, 25(3):649–665, 2014.

[3] Siniša Šegvić, et al. A computer vision assisted geoinformation inventory for traffic infrastructure. In 13th International IEEE Conference on Intelligent Transportation Systems, pages 66–73. IEEE, 2010.

[4] Samodha Pallewatta, Vassilis Kostakos, and Rajkumar Buyya. Microservices-based iot application placement within heterogeneous and resource constrained fog computing environments. In Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing, pages 71–81, 2019.

[5] Nikolay Nikolov, Yared Dejene Dessalk, Akif Quddus Khan, Ahmet Soylu, Mihhail Matskin, Amir H Payberah, and Dumitru Roman. Conceptualization and scalable execution of big data workflows using domain-specific languages and software containers. Internet of Things, page 100440, 2021.

[6] Anatoliy Zabrovskiy, Christian Feldmann, and Christian Timmerer. Multi-codec dash dataset. In Proceedings of the 9th ACM Multimedia Systems Conference, pages 438–443. ACM, 2018.

[7] Framing a video. https://gist.github.com/SiNa88/c85d8cfac641918c6de8b4f31d8cdc22

[8] GTSRB - German Traffic Sign Recognition Benchmark; Multi-class, single-image classification challenge: https://www.kaggle.com/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign

[9] Ganesh Ananthanarayanan, Victor Bahl, Landon Cox, Alex Crown, Shadi Nogbahi, and Yuanchao Shu. Video analytics-killer app for edge computing. In Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, pages 695–696, 2019.

[10] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural networks, 32:323–332, 2012.