Knowing eye gaze through Webcam

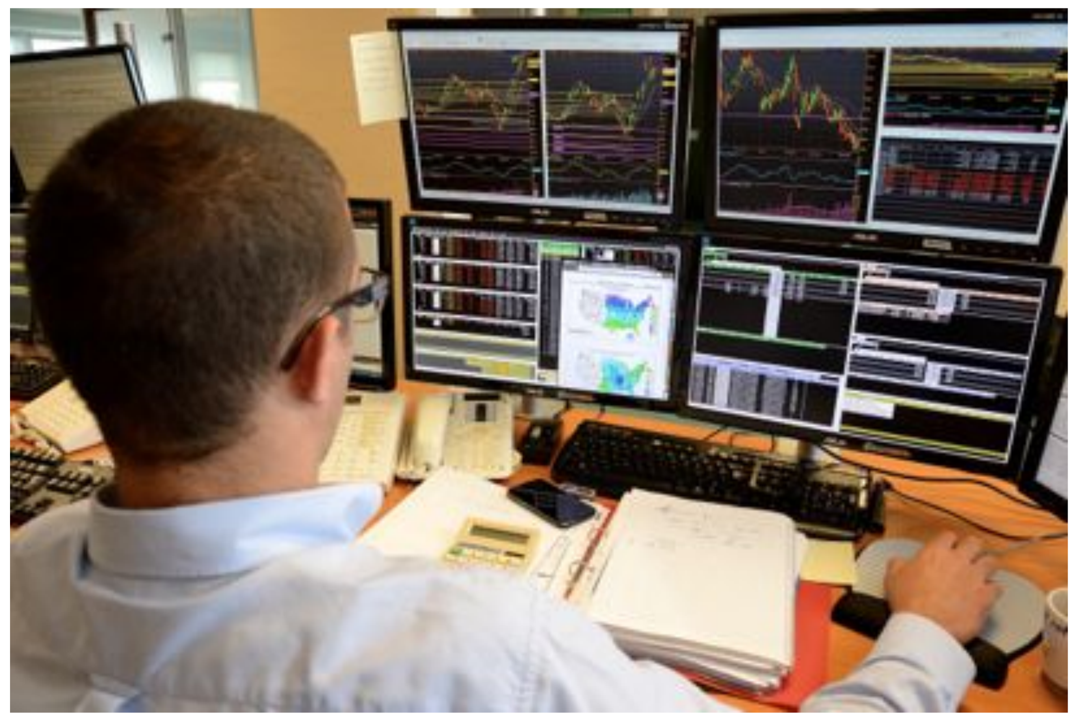
Data Science Institute
Columbia University

Minghao Chen, Feng Fan, Lehui Liu, Shaofei Wang, Zhida Zhang
Industry Mentor: Naftali Cohen
Faculty Mentor: Sining Chen

Data Science Capstone Project with J.P. Morgan AI Research

Introduction

One way to model the sequence of a trader to help the decision making process in financial company is to use people's gaze information on multi-screens, which could be recorded by using some specialized device such as GazePoint. However, it's expensive and not scalable. Our project provided a convenient and scalable solution by directly using webcam to capture the face images and make real-time prediction. To simplify the problem, we built nine classification models to predict which part of the screen people is looking at. The prediction results are basically different time series sequence, which could be further used for decision making process.



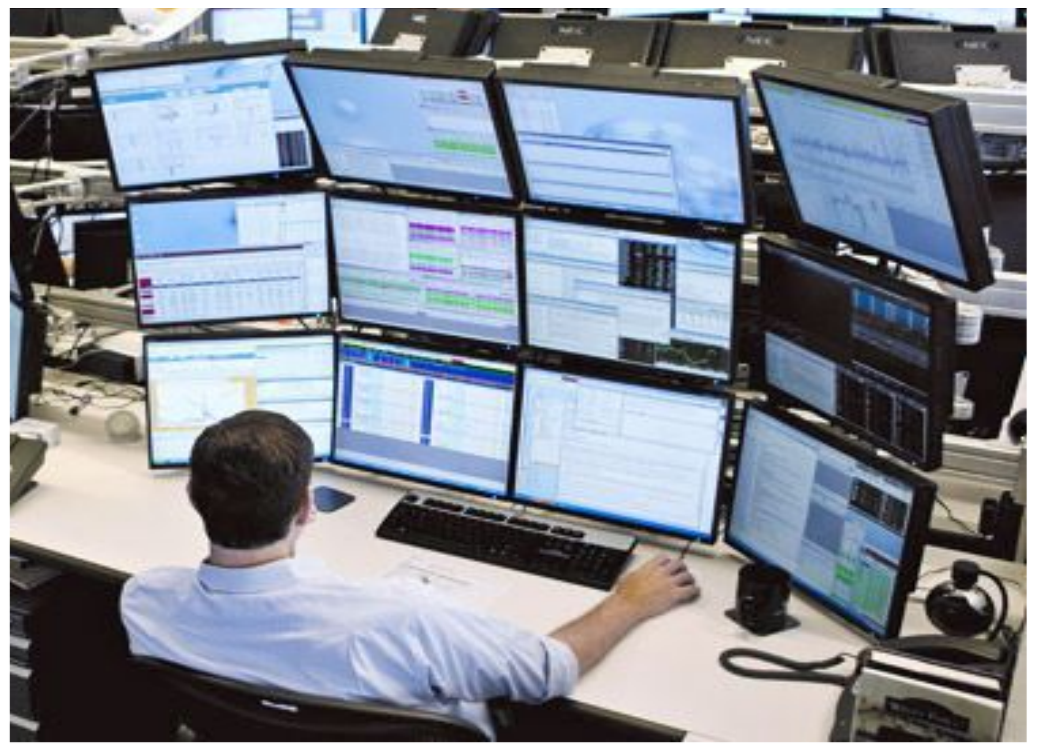


Figure 1. Two traders daily work

Architecture

The whole project pipeline consists of three parts: (1) Observer, it controls webcam to capture face images; (2) Eyegaze comparator, it uses gaze point device to provide labels; (3) Model, make prediction in real time

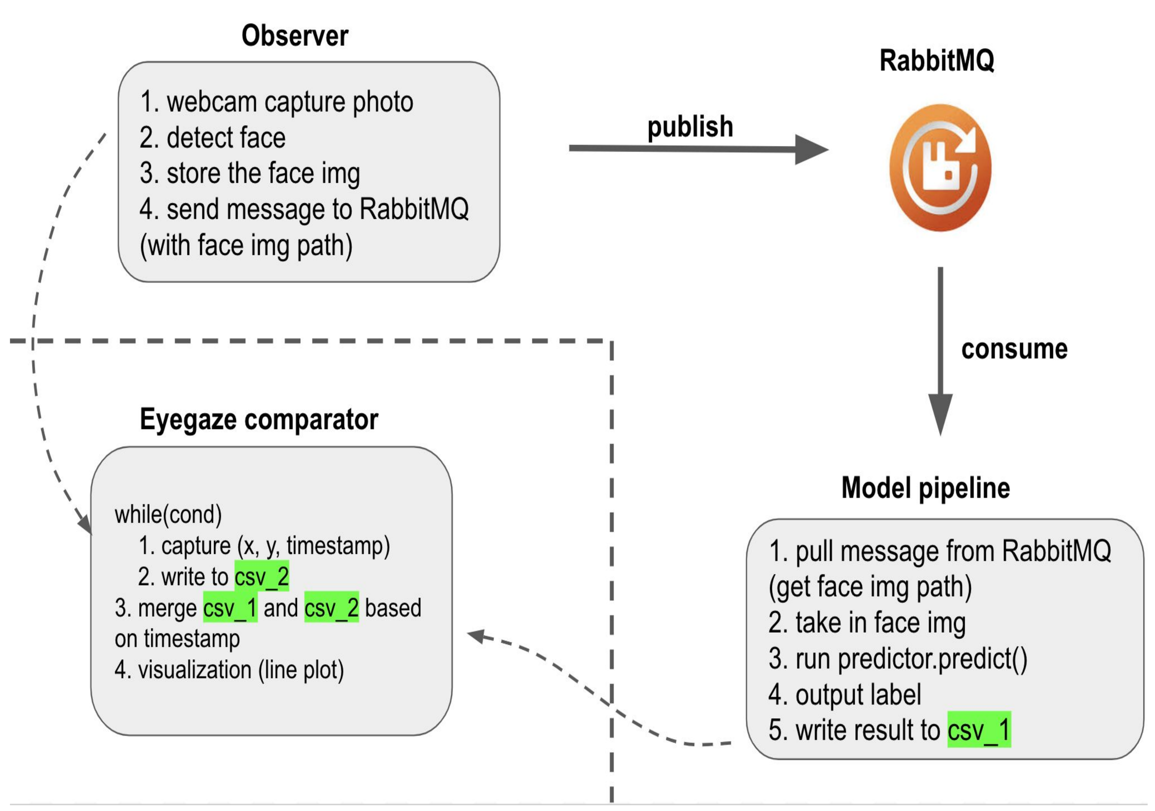


Figure 2. Project Pipeline

Model:

We implement pre-trained neural nets, such as ResNet[1], MobileNet[2], FaceNet[3] with additional dense layers to do classification.

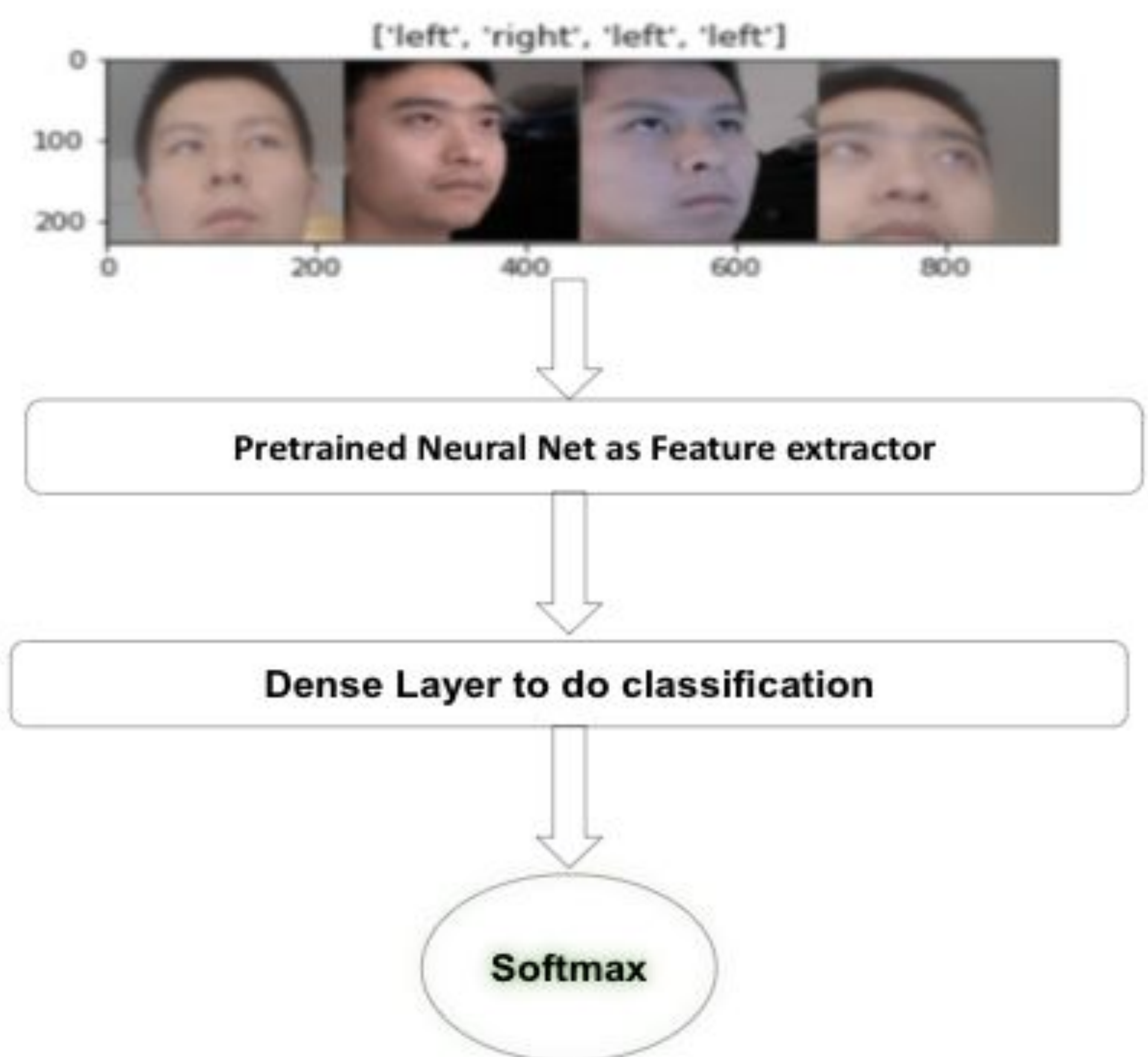


Figure 3. Model Pipeline

Evaluation

Model Performance:

Table 1. Experiment Result on our dataset

Model	LR+TruncatedSVD	ResNet	MobileNet	FaceNet
Inference Time	almost 0	1s per img	0.1s per img	1.5s per img
Accuracy	25.64%	53%	49%	45%

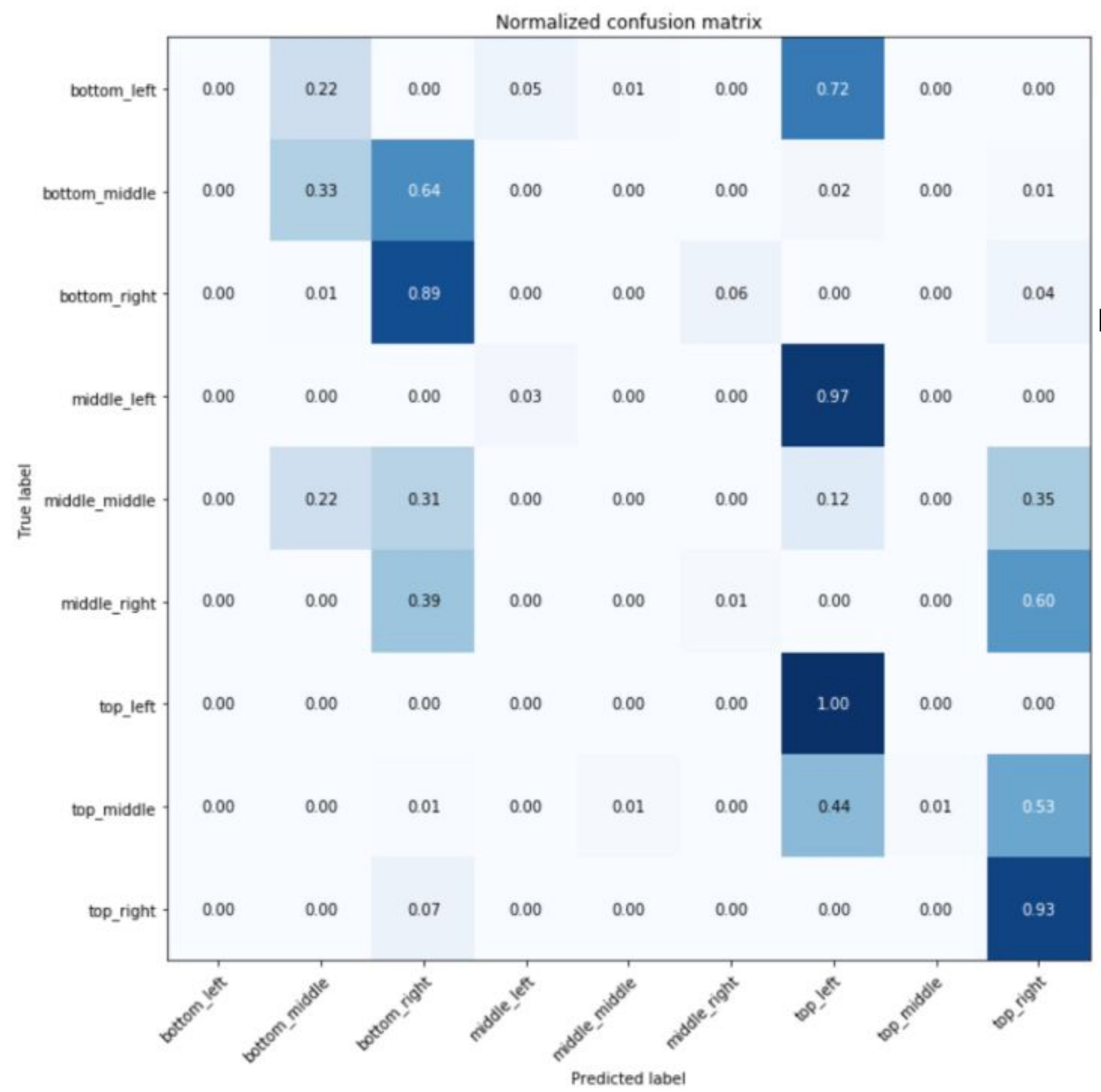


Figure 4. Confusion matrix on MobileNet Model

Real-time evaluation:

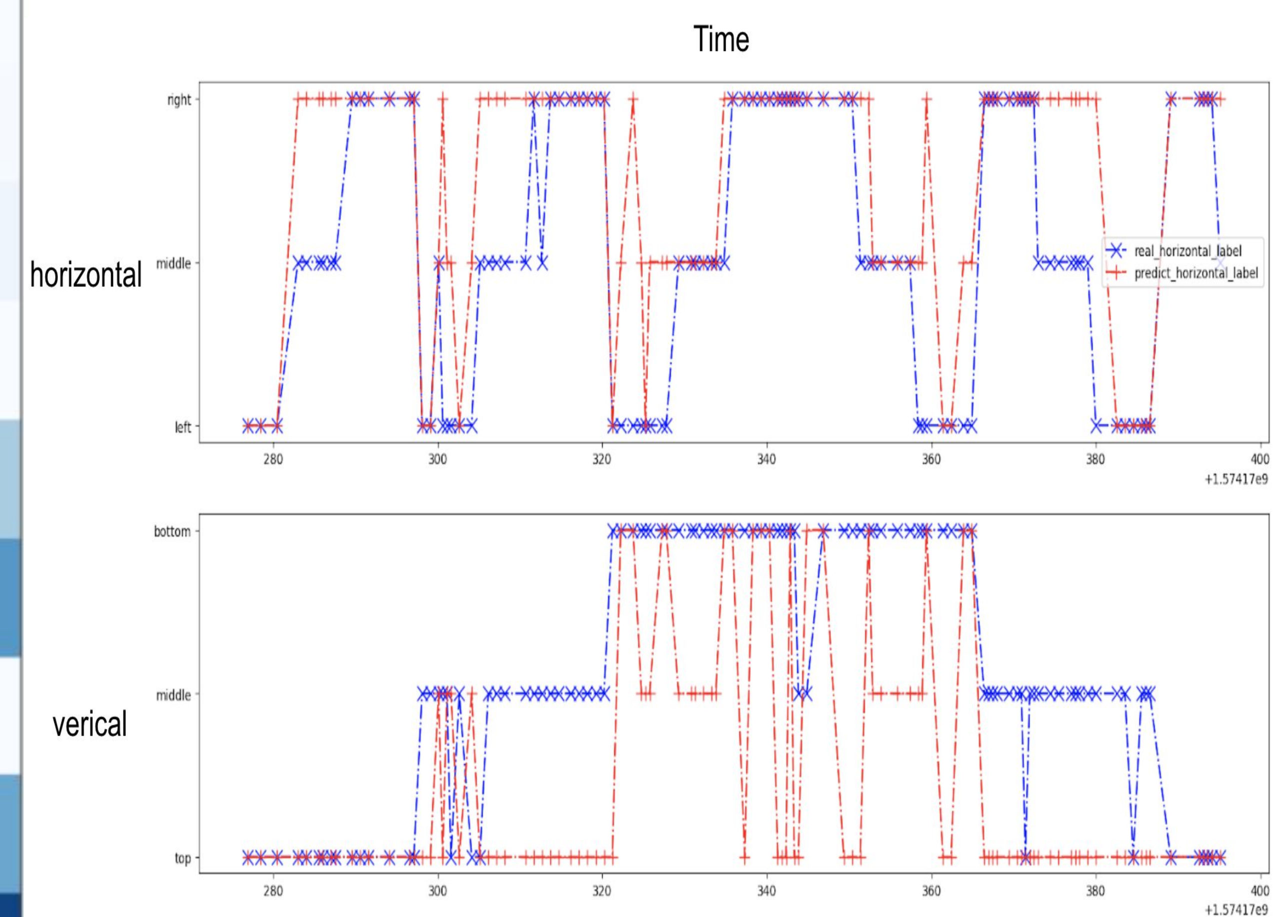


Figure 5. sample real-time evaluation result

Conclusion

This project provides insights on capturing gazing point with webcams. Using transfer learning, which was proven useful given similar contexts and users, we model traders' eye gazing areas when looking at the screen. While our experiment tries different models, we decide to use MobileNet under the trade off between accuracy and inference time.

Acknowledgments

This research was supported by J.P Morgan and Data Science Institute at Columbia University. We thank Dr. Naftali and his team for great comments and equipments support on our work.

References

[1] He et at. *"Deep Residual Learning for Image Recognition"*,2015
[2] Sandler et at.*"MobileNetV2: Inverted Residuals and Linear Bottlenecks"*, 2018
[3] Schroff et at.*" FaceNet: A unified embedding for face recognition and clustering"*, 2015