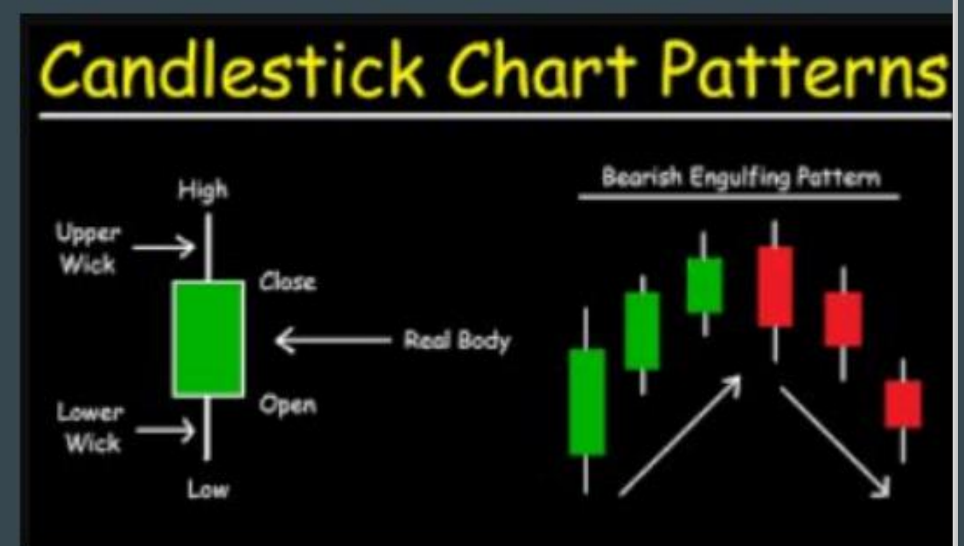
Fast Candlestick Patterns Detection with Limited Training Samples Using RGB Gramian Angular Field and YOLO-LITE-V1

Background and Problems

- Candlestick chart is a visualizable tokenization of price time series.
- Potential in high-frequency trading, so automation is very important.
- Many traders come up with new candlestick pattern indicators and need automatic pipeline to test their performance.
- Insufficient data
- Tedious labeling
- Numeric value based rather than graph based



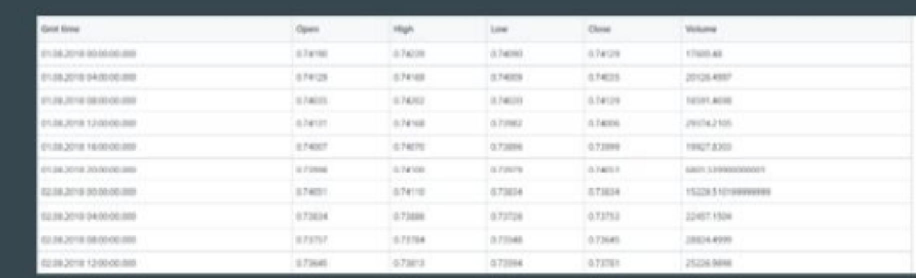
Datasets

- Input data:
  1. Time series price data

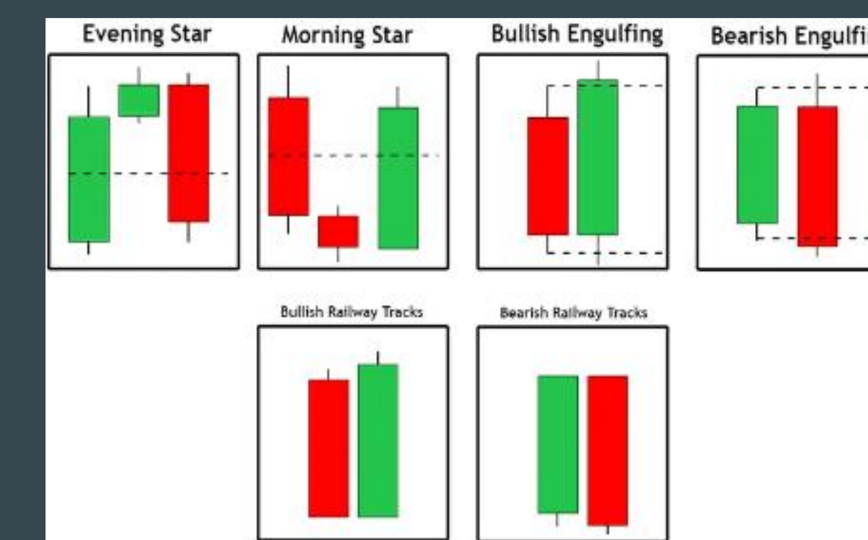
     WRDS with 4h, 1h, 30min, and 5min intervals on currency and SPY futures. It is roughly 800M in total
  2. Candlestick pattern label

     200-300 for each pattern
  3. Real candlestick charts(background changed)

     Yahoo Finance and TD_Ameritrade



Time series price data



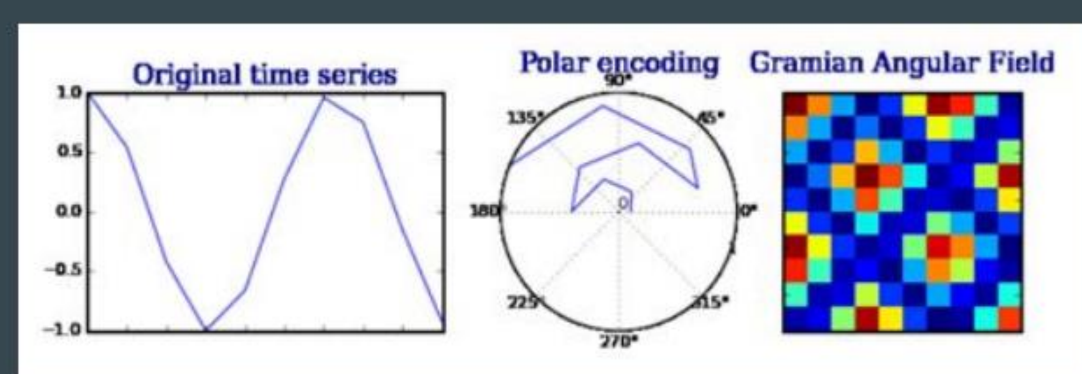
real candlestick charts(background changed)

Start_t,	End_t,	Label
4,	7,	2
27,	29,	3
28,	30,	3

Candlestick pattern label

Problem Statement

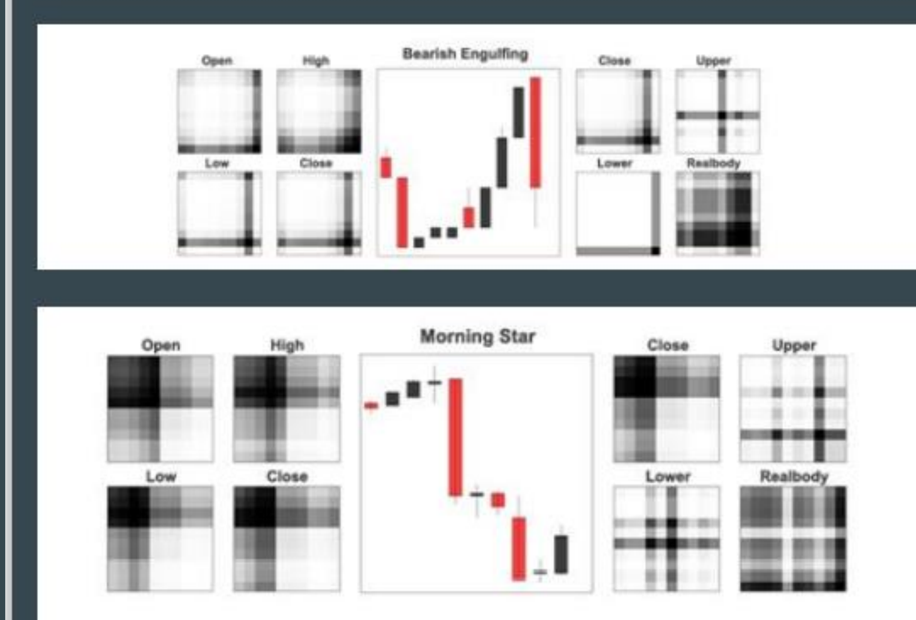
- Goal:
  A pipeline that can train on a small training set and yield an object detection model to detect candlestick patterns graphically
- Input data:
  1. Time series price data
  2. Candlestick pattern label
  3. pictures of real candlestick charts(background changed)
- Core Model
  YOLO-LITE for object detection
  Simple CNN with dropout/pooling for data augmentation
- Metrics
  mAP>0.4 for object detection
  ROC_AUC for simple CNN



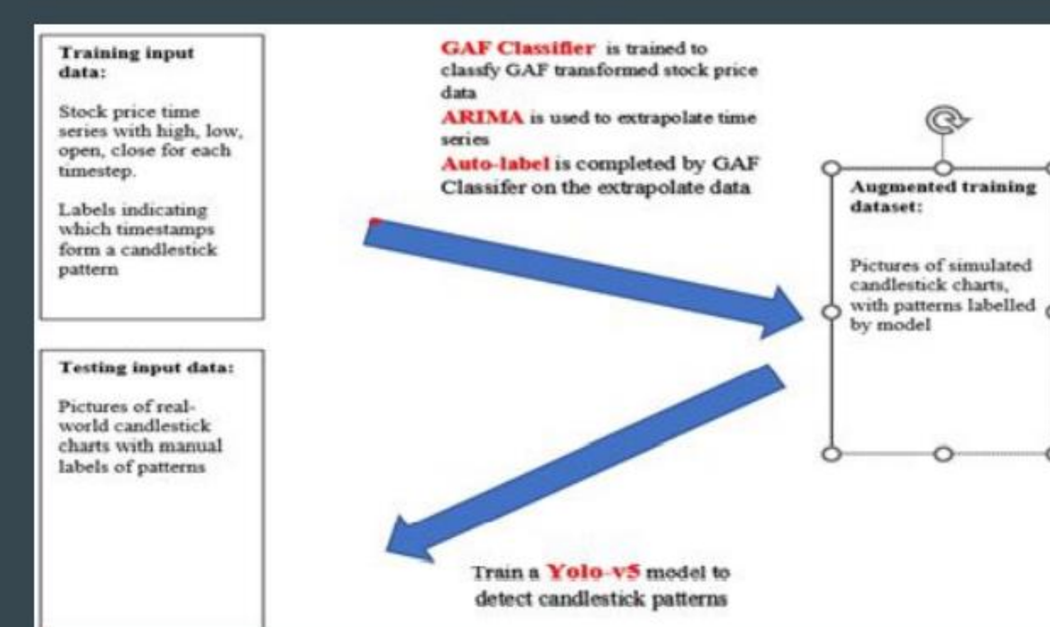
Time Series —> polar coordinate -> Gramian Angular Field

YOLO-LITE ←Augmented timeseries with labels <- SimpleCNN

Method



Chen's GAF without RGB

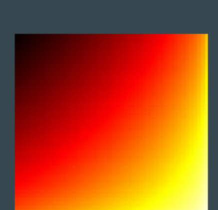


Model Structure

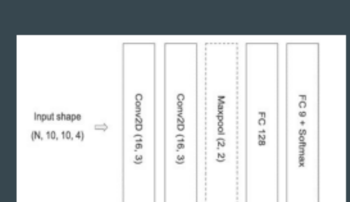
mAP>0.4 = 0.481

ROC_AUC = 0.71

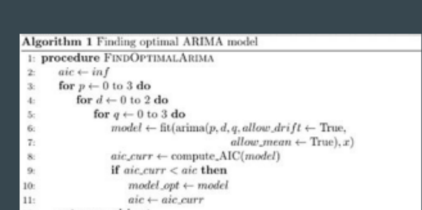
Method



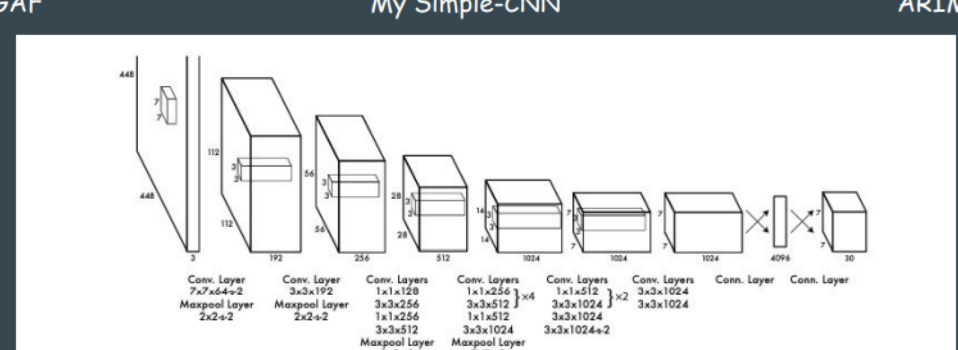
My RGB-GAF



My Simple-CNN



ARIMA



YOLO_LITE (# of parameters are adjusted)

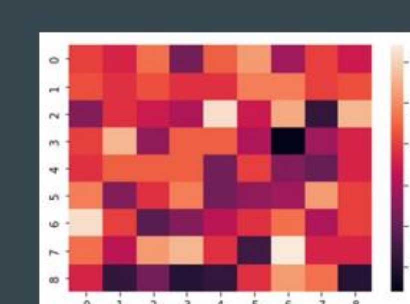
Conclusions

While there still exists cases that YOLO_LITE model labels far-off pictures for reasons that I cannot understand, the overall performance of both bounding box precision and classification precision is already reasonably good with a mAP>0.4 at 0.481.

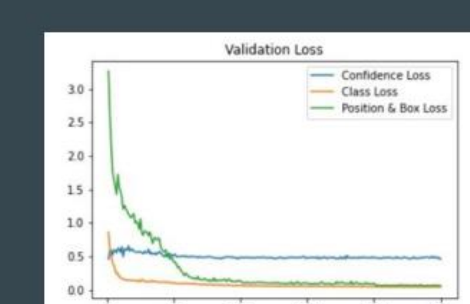
Part of the error is due to imperfectness in manual labelled test sets. My design of RGB Granian Angular Field, data ARIMA-CNN-based data augmentation, and model choice of YOLO-LITE contributes to the performance to a great extent.

If more time, computational resource, and people are available in future, we will increase the number of precisions of testing labels and explore more completed setup for object detection model.

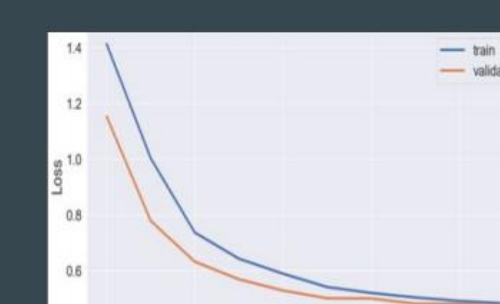
Results



Confusion Matrix for YOLO-LITE



Validation loss for YOLO-LITE



Training loss for Simple-CNN

Top examples by YOLO-LITE

Top examples by YOLO-LITE

Failed examples by YOLO-LITE