Paper Replication: (Re-)Imag(in)ing Price Trends

Final Project of MSBD 5013

Group 8

CEN Xinxin, MA Xiaoran and WU Xiang

{xcenab, xmabi, xwucb}@connect.ust.hk

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Presentation Video Link: https://www.bilibili.com/video/BV1TT4y1k7Jx/

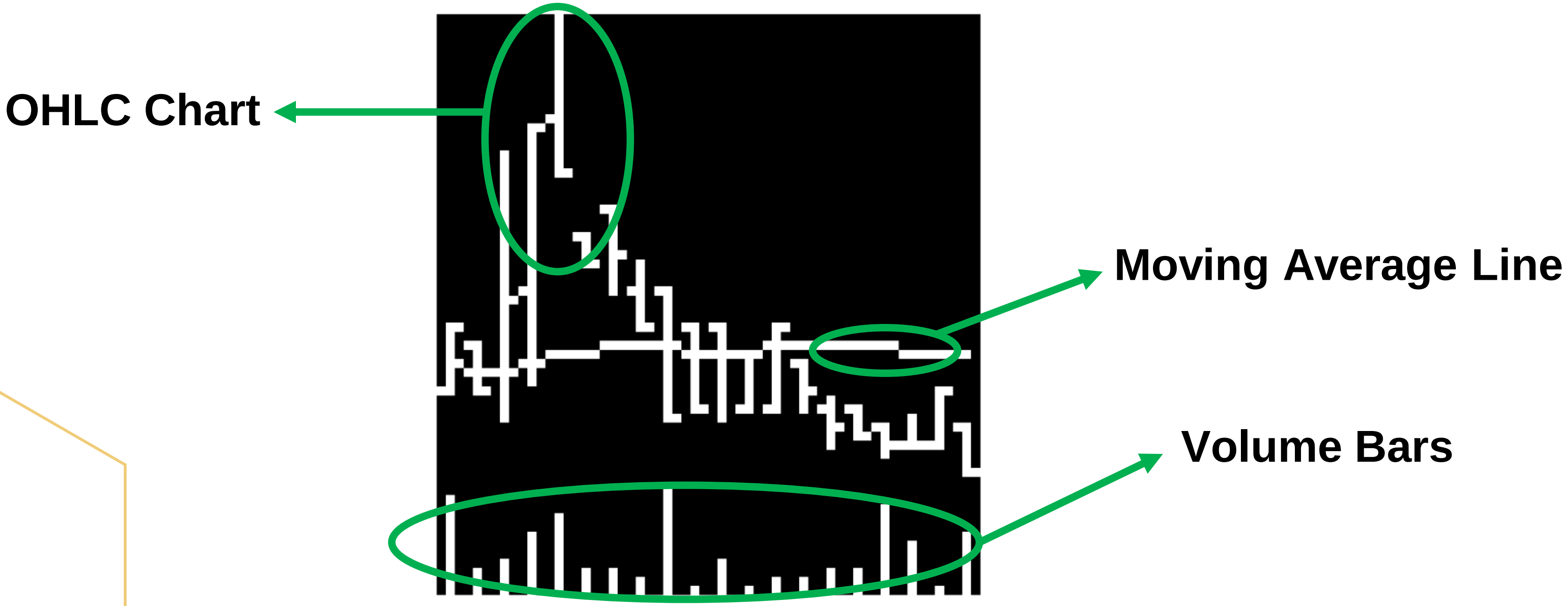
Contents

Introduction

- In our project, we implemented different image classification models to predict future returns of stocks. We used images of stock-level price charts as raw predictor data.
- We followed the paper *(Re-)Imag(in)ing Price Trends* to decide the data split strategy, model structure, evaluation metrics and robustness checking method.
- Other than CNN, we also investigate a traditional method based on HOG feature and Logistic Regression model. Finally, we compared all the models and analyzed the results in detail.

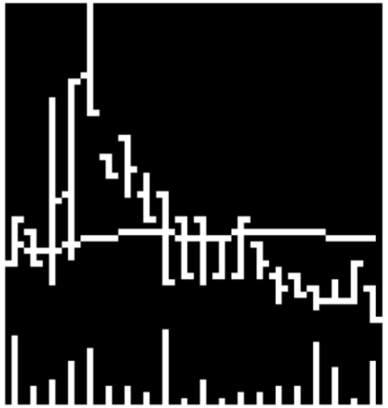
About Dataset

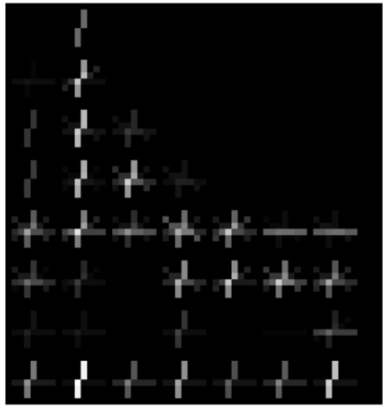
- Samples:
  - Input Images consists of three components: OHLC Chart, moving average lines, and volume bars of consecutive 20 days. The images of 20-days version are of size 64×60.
  - Labels takes value 1 for positive returns and 0 for non-positive returns of the subsequent 20-days.
- Sample period: from 1993 to 2019.



HOG-Based Method

- HOG (Histogram of Oriented Gradients) is a feature descriptor used in image processing, mainly for object detection.
- Getting the HOG feature descriptor consists of several steps.
  - The first is dividing the image into cells of 8×8 pixels, and calculating the gradient for each pixel of cell.
  - The second step is to group the gradients into nine bins using the direction of the gradient. If the direction is between two bins, then the vote is divided proportionally to each bin.
  - The third step is to applied L2-norm in each block consisting of 2×2 cells, since we want the histogram values to be insensitive to lighting.
  - The result feature vector has dimension of 1512, which is much smaller than the original image.
- The images below shows an example: the left one is the original image, and the right one is the visualization of HOG descriptors of each cell.
- Using Logistic Regression as the classifier. The performance is recorded in the resulting table, and was compared with other CNN models.

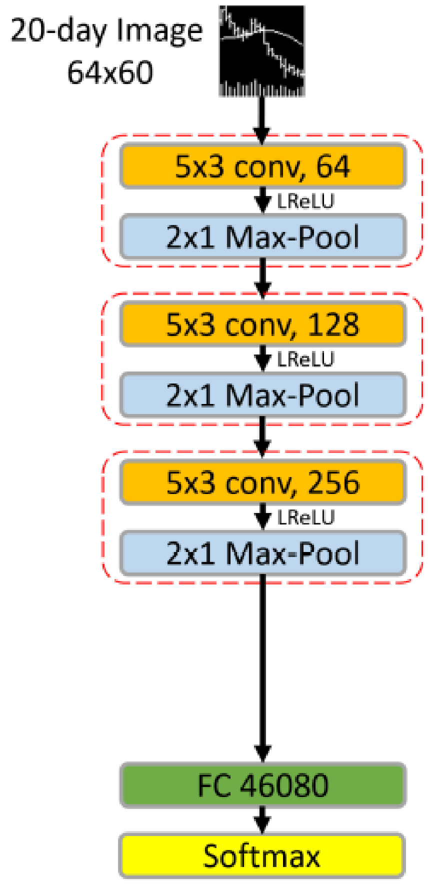




CNN-Based Method

- Data Preparation
  - Non-positive returns (‘down’) - 0
  - Positive returns (‘up’) - 1
  - NaN - 2 (remove)

- Model Design
  - Three convolution blocks
  - A fully connected layer
  - A softmax layer



CNN-Based Method

- Workflow Design
  - Data split
    - Eight years for training (70%) and validation (30%)
    - Nineteen years for testing
  - Loss and evaluation
    - Cross entropy loss
    - Accuracy (other evaluation metrics: Spearman、Pearson、AUC)
  - Train process
    - Xavier initialization
    - Dropout (0.5)
    - Batch normalization
    - Early stopping (epoch=2)

CNN-Based Method

➢Performance Evaluation

- ➢Ablation studies
- ➢Replicate Table 18
- ➢Results comparison

Our Replication

		Loss		Acc.		Correlation		ROC
		V	T	V	T	Spearman	Pearson	Auc
Baseline		**0.689**	**0.695**	**0.541**	**0.525**	**0.064**	**0.040**	**0.535**
Filters (64)	32	0.689	0.694	0.538	0.519	0.060	0.039	0.532
Layers (3)	2	0.690	0.698	0.541	0.521	0.057	0.037	0.531
Dropout (0.50)	0.00	0.693	0.697	0.531	0.520	0.047	0.032	0.526
BN (yes)	no	0.688	0.695	0.542	0.520	0.056	0.037	0.531
Xavier (yes)	no	0.688	0.693	0.542	0.523	0.059	0.039	0.532
Activation (LReLU)	ReLU	0.696	0.704	0.533	0.513	0.049	0.031	0.526
Max Pool Size (2×1)	2×2	0.689	0.694	0.538	0.520	0.059	0.038	0.532
Filter Size (5×3)	3×3	0.696	0.703	0.529	0.508	0.045	0.029	0.524
Dilation/Stride (2,1)/(3,1)	(1,1)/(3,1)	0.689	0.694	0.539	0.520	0.055	0.037	0.544
HOG Feature + Logistic Regression		0.691	0.693	0.525	0.511	0.038	0.022	0.521

Table 18

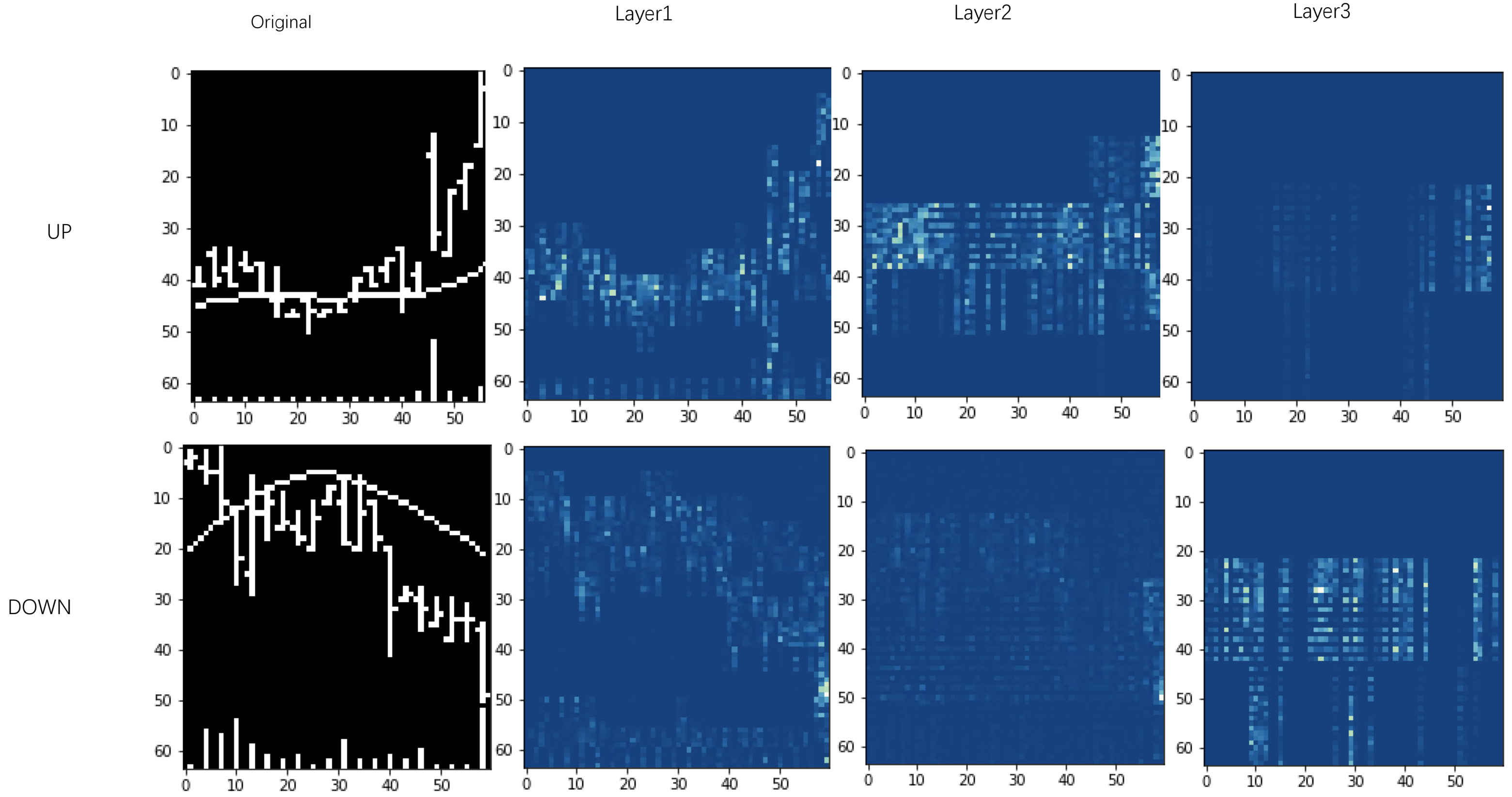
		Loss		Acc.		Correlation		Sharpe Ratio	
		V	T	V	T	Spearman	Pearson	EW	VW
Baseline		**0.687**	**0.690**	**0.542**	**0.533**	**0.059**	**0.034**	**2.16**	**0.49**
Filters (64)	32	0.687	0.690	0.543	0.534	0.058	0.033	2.00	0.28
	128	0.689	0.691	0.538	0.530	0.054	0.031	1.85	0.40
Layers (3)	2	0.688	0.690	0.541	0.534	0.054	0.030	1.77	0.33
	4	0.688	0.691	0.541	0.531	0.052	0.031	2.14	0.22
Dropout (0.50)	0.00	0.697	0.693	0.532	0.528	0.047	0.027	2.14	0.59
	0.25	0.689	0.692	0.541	0.528	0.055	0.032	2.31	0.51
	0.75	0.688	0.691	0.540	0.530	0.053	0.030	1.47	0.16
BN (yes)	no	0.685	0.691	0.550	0.532	0.062	0.037	2.33	0.51
Xavier (yes)	no	0.688	0.692	0.541	0.527	0.056	0.032	2.08	0.44
Activation (LReLU)	ReLU	0.688	0.691	0.540	0.531	0.053	0.029	1.49	0.23
Max-pool Size (2×1)	(2×2)	0.689	0.691	0.536	0.531	0.053	0.029	1.62	0.32
FilterSize (5×3)	(3×3)	0.690	0.691	0.536	0.530	0.051	0.027	1.53	0.16
	(7×3)	0.689	0.691	0.537	0.531	0.053	0.030	1.84	0.09
Dilation/Stride (2,1)/(3,1)	(2,1)/(1,1)	0.692	0.692	0.533	0.523	0.047	0.028	2.20	0.26
	(1,1)/(3,1)	0.689	0.691	0.540	0.532	0.057	0.032	2.00	0.30
	(1,1)/(1,1)	0.693	0.693	0.534	0.521	0.051	0.029	1.80	0.25

Interpretability of the CNN Model

- In order to aid interpretation of deep learning models, we use Grad-CAM method, which is known as gradient-weighted class activation mapping, to illustrate the regions of the most important for predicting a class.
- Grad-CAM can produce activation heat maps for each layer of the CNN model, and these heat maps help us to observe which parts of the image are important to final prediction.
- The main idea of Grad-CAM is to use backward to get the gradient corresponding to each pixel on each feature map and average each gradient map to get the weight.

Interpretability of the CNN Model

➢ We illustrate our CNN baseline model and draw both “up” and “down” images in 2019 respectively. The following figure shows the original images and Grad-CAM activation heat map at each of the CNN’s three layers.



Conclusions

- We implement two methods to solve this problem. The results show that CNN-Based method performs better than HOG-Based method.
- We perform sensitivity analysis of the CNN prediction model. We use loss, accuracy, correlation and AUC as evaluation metrics. The results we got is similar with the original paper shows in Table 18.
- We use a visualization method “Grad-CAM” to interpret CNN model. The results show that CNN model has ability to extract non-linear associations between price, high-low range and volume to obtain successful return prediction.

Thank you for listening!

Q & A