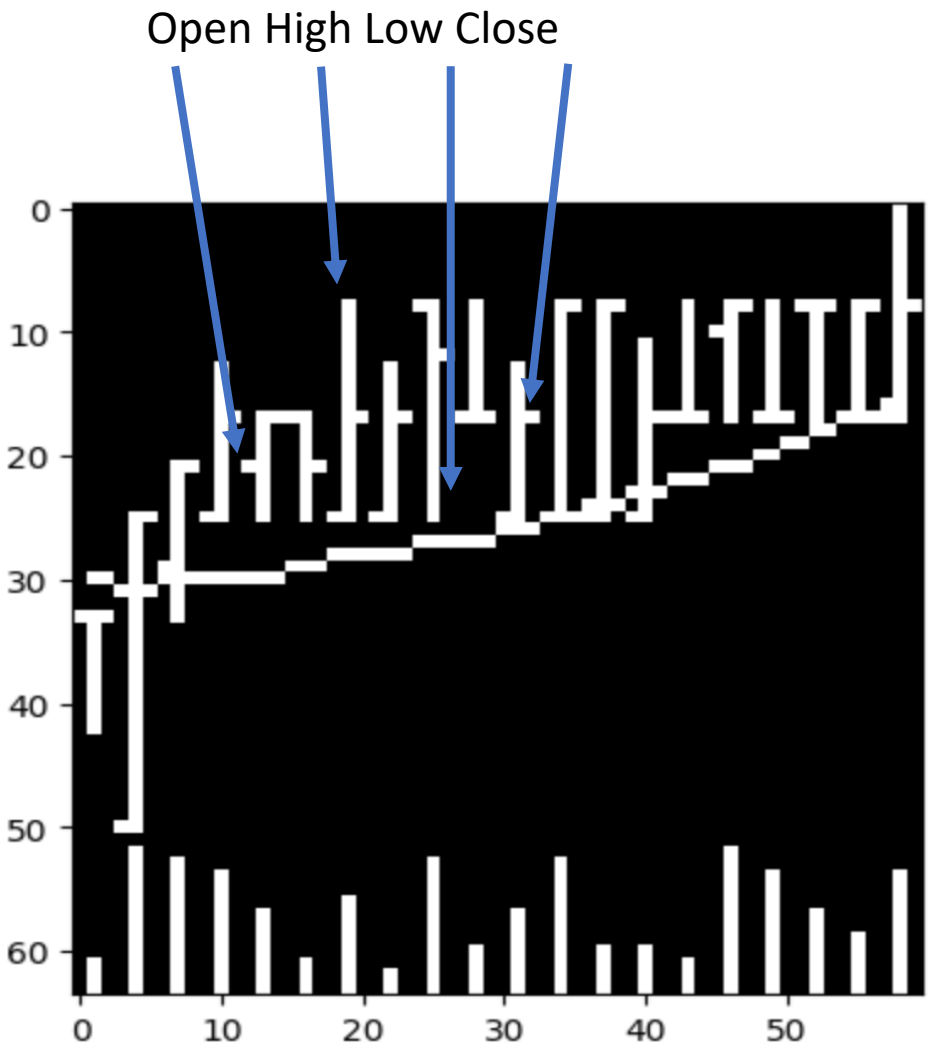
Market Data

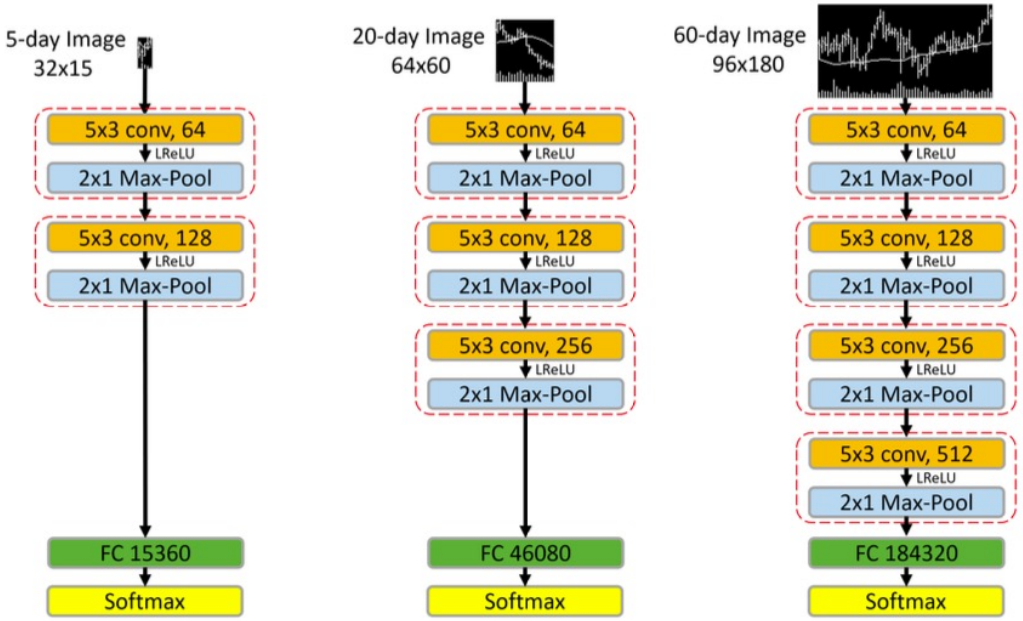


- Each OHLC chart is 5/20/60 business days of data
  - Each day 3pixels wide
  - Eg: Total image for 20 days == 60x64 pixels wide
- Returns scaled so the largest price is the top of the image
- Three types of information in image
  - Direction stock movements
  - Volatility data
  - SMA (5/20/60)

CNN Architecture

- Goal for CNN is to balance tradeoff between complex pattern recognition and interpretability
- Why CNN?
  - Alternative to the time series format that can uncover complex patterns
  - Technical traders have long relied on geometry and SMA's (cognition) to make trading decisions
  - CNN designed to solve positionality, scale etc… and may require less parameterization than NNs

Figure 3: Diagram of CNN Models

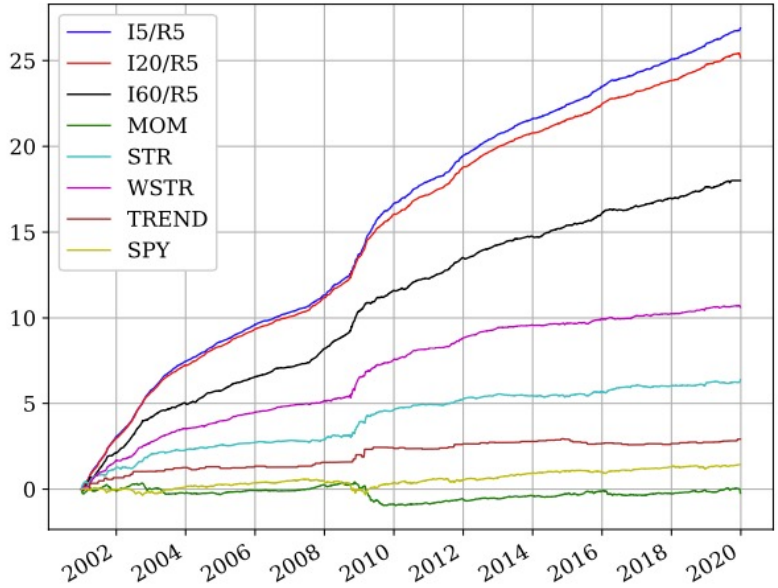


- **Architecture Highlights**
  - Loss Funct:
    - Adam, .0001
  - 50% dropout to fully connected layer
  - Stopping after 2 consecutive epochs of higher error
  - 70% train 30% validation from 1993-2000 and testing from 2001-2019
  - 9 separate models (combinations of I5, I20, I60)

Evaluation

- Prediction is most potent (accurate) in the 5 day period after the image is observed
  - Sharpe ratios exceed 5 in a High Low portfolio sorted by decile (High-Low)(10-1)
- Outperforms Momentum and Short Term Reversal with no parameterization needed
- Predictions are context independent
  - Work well in international markets as well as varying time periods
- Volatility surprisingly low

I-Images / R-Prediction

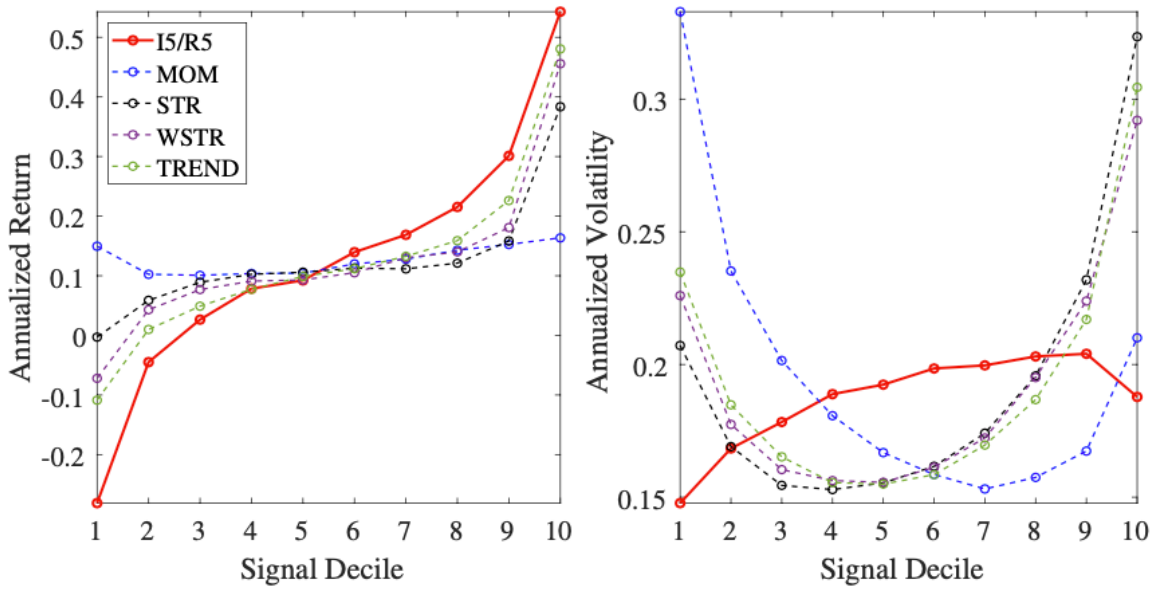


	Equal Weight													
	I5/R5		I20/R5		I60/R5		MOM/R5		STR/R5		WSTR/R5		TREND/R5	
	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR	Ret	SR
Low	-0.28	-1.92	-0.32	-1.94	-0.21	-1.10	0.15	0.44	-0.01	-0.03	-0.08	-0.34	-0.11	-0.46
2	-0.04	-0.27	-0.04	-0.21	0.02	0.12	0.10	0.44	0.06	0.35	0.04	0.24	0.01	0.05
3	0.03	0.15	0.04	0.20	0.07	0.35	0.10	0.50	0.09	0.58	0.08	0.48	0.05	0.30
4	0.08	0.41	0.08	0.43	0.11	0.58	0.10	0.57	0.10	0.67	0.09	0.58	0.08	0.50
5	0.09	0.48	0.12	0.65	0.14	0.75	0.10	0.63	0.11	0.68	0.09	0.60	0.10	0.64
6	0.14	0.70	0.15	0.80	0.16	0.88	0.12	0.76	0.11	0.70	0.11	0.65	0.11	0.71
7	0.17	0.84	0.19	0.97	0.17	0.93	0.13	0.83	0.11	0.64	0.13	0.75	0.13	0.78
8	0.22	1.06	0.23	1.19	0.20	1.08	0.14	0.90	0.12	0.62	0.14	0.72	0.16	0.85
9	0.30	1.48	0.27	1.40	0.22	1.23	0.15	0.91	0.16	0.68	0.18	0.81	0.23	1.04
High	0.54	2.89	0.52	2.76	0.33	1.85	0.16	0.78	0.38	1.19	0.46	1.56	0.48	1.58
H-L	0.83***	7.15	0.84***	6.75	0.54***	4.89	0.02	0.07	0.39***	1.76	0.53***	2.84	0.59***	2.92
Turnover	690%		667%		619%		123%		341%		660%		499%	

Evaluation (Cont.)

- Return becomes less significant when we ask the model to predict monthly or quarterly returns (20,60), but still above the benchmarks (MOM, WSTR) due to lower volatility
- When restricted to largest 500 stocks, returns are less potent
  - Is the CNN just learning to pick volatility?
- Highest correlated feature with the CNN is Short Term Reversal (-34%)
- CNN performs better than ~96% of all (7,846) technical trading strategies

Figure 6: Prediction Accuracy By Decile



Sensitivity

- Data Augmentation
  - Volume and MA bar seem to contribute to noise for I5, but very useful for I20, I60
  - Removing smaller stocks (due to volatility/swings) leaves model unchanged
  - Changing pixel representation (centered, top) results in similar scores (sharpe ratio & return)
- Model selection
  - Changing number of convolution layers, dropout rate, dilation, degrades model performance or leaves results largely unchanged
- Computer Vision Models
  - CNN v. HOG v. HAAR
    - CNN outperforms for I5
    - CNN mostly outperforms for I20, I60
    - HAAR feature selection using Adaboost as too many features (100Million+)

Table IA11: Sensitivity to Model Structure and Estimation, I20/R20

		Sharpe Ratio EW	Sharpe Ratio VW
Baseline		**2.16**	**0.49**
Filters (64)	32	2.00	0.28
	128	1.85	0.40
Layers (3)	2	1.77	0.33
	4	2.14	0.22
Dropout (0.50)	0.00	2.14	0.59
	0.25	2.31	0.51
	0.75	1.47	0.16
BN (yes)	no	2.33	0.51
Xavier (yes)	no	2.08	0.44
Activation (LReLU)	ReLU	1.49	0.23
Max-pool Size (2×1)	(2×2)	1.62	0.32
FilterSize (5×3)	(3×3)	1.53	0.16
	(7×3)	1.84	0.09
Dilation/Stride (2,1)/(3,1)	(2,1)/(1,1)	2.20	0.26
	(1,1)/(3,1)	2.00	0.30
	(1,1)/(1,1)	1.80	0.25

Technical Patterns/Transfer Learning

- Technical Patterns
  - Model extended to predict probability of different technical patterns producing positive 20-day return based on 10K trained images
- Transfer Learning
  - The paper also explores transfer learning with other markets, both on trained and untrained (transfer) models
    - Transferring the trained model produces high Sharpe in most other markets
    - When data is insufficient in other markets (stock count) this could be beneficial → Arbitrages

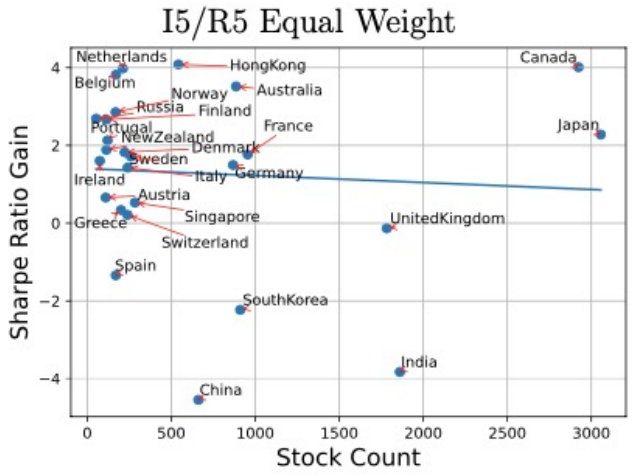


Table IA19: CNN Predictions on Simulated Charts

	Sign	20-Day Images Mean	20-Day Images Std	60-Day Images Mean	60-Day Images Std		Sign	20-Day Images Mean	20-Day Images Std	60-Day Images Mean	60-Day Images Std
Noise (Brownian motion)		48.6	(4.0)	50.7	(4.2)	Descending Triangle	–	42.1	(2.3)	46.5	(3.3)
Cup And Handle	+	46.7	(2.5)	51.3	(2.8)	Ascending Triangle	+	54.9	(2.0)	57.5	(2.8)
Head And Shoulders Top	–	56.3	(2.3)	55.5	(3.6)	Rounding Top	–	54.6	(1.9)	51.4	(2.5)
Head And Shoulders Bottom	+	49.3	(2.0)	54.8	(3.6)	Rounding Bottom	+	36.8	(1.8)	43.4	(2.8)
Broadening Top	–	45.7	(2.1)	52.4	(2.8)	Triple Top	–	62.0	(2.1)	60.3	(3.5)
Broadening Bottom	+	52.0	(2.6)	47.1	(3.0)	Triple Bottom	+	53.9	(2.0)	55.9	(3.7)
Triangle Top	–	44.6	(2.3)	48.7	(3.4)	Bearish Flag	–	55.8	(2.4)	56.4	(3.8)
Triangle Bottom	+	51.4	(2.2)	55.3	(3.0)	Bullish Flag	+	53.6	(2.1)	58.7	(3.4)
Double Top	–	55.6	(2.1)	56.0	(3.4)	Bearish Pennant	–	51.2	(2.5)	53.5	(3.7)
Double Bottom	+	50.1	(2.1)	54.3	(3.7)	Bullish Pennant	+	50.3	(2.1)	54.4	(3.2)
Rising Wedge	–	49.1	(1.9)	48.2	(1.9)	Diamond Top	–	59.3	(2.1)	62.4	(3.5)
Falling Wedge	+	40.0	(2.1)	41.6	(2.2)	Diamond Bottom	+	56.2	(2.0)	57.5	(3.6)

Thoughts

- This paper implies credibility to financial technical analysis, and takes the human out of the equation
- I am unsure of train/test though process
  - train and validation on 1993-2000 data. Markets may have fundamentally changed since then?
  - Why not sample different time periods from data?
- Paper made good effort to experiment with sensitivity of model parameters, possibly more can be experimented with…
- Sharpe ratios degraded slightly when stocks are limited to 500 largest cap, implying maybe more volatile picks when unrestricted (smaller and more illiquid stocks)..