SurfaceEdge

Predicting Next-Day Options Price Changes from Options Surface Images

Caleb Bettcher & Troy Arthur

CSCI 5922 — Neural Networks and Deep Learning | University of Colorado Boulder | Spring 2026

Motivation

Accurately pricing, and predicting the future price of options contracts is a major goal of retail and institutional traders alike.

A single percentage point advantage can result in hundreds of millions of dollars of additional profit for top institutions.

Millions of options contracts trade daily across hundreds of underlying equities, forming a rich two-dimensional structure.

This options surface encodes the market's real-time consensus on future uncertainty across all strikes and expiry dates.

The Gap

Existing deep learning approaches to options pricing treat each contract in isolation using scalar features, or apply LSTMs to capture time-series dynamics.

Importantly, neither approach leverages the full spatial structure of the options surface, nor does so through visual representation.

Additionally, many rely on explicit statistical denoising steps to handle market noise, rather than architectures with intrinsic noise immunity.

Our Edge

SurfaceEdge is novel in its approach by leveraging the inherent noise reduction capabilities of Convolution models through image representation of options surfaces.

We hypothesize that past performance and future expectation are already priced into current conditions, and we expand prior work by amending convolutional extracted features with contract-level scalar inputs.

Dataset

Raw Data

104	17 yrs	~9.4GB	200M+
U.S. Equity Tickers	Historical Coverage	Raw Parquet Data	Labeled Contracts

Surface Construction

1. **Split Adjustment**
   Adjust for historical stock splitting events using split-coefficients
2. **Grid Binning**
   Each day's options chain is discretized into a fixed 60×30 grid
   Y: log-moneyness (−1.0 to +1.0)
   X: days to expiry (1–61 days)
3. **RGB Encoding**
   = implied volatility,
   = log(open interest)
   = log(volume)
4. **Contract Specific Scalars**
   spot, strike, T, log-moneyness, mark, Greeks, spread, dividends, momentum
5. **Information Density**
   Drop days with < 100 valid contracts

Data Processing

Label: Next-day percentage mark price change

$\ell = (\text{mark}_{t+1} - \text{mark}_t) / \text{mark}_t$

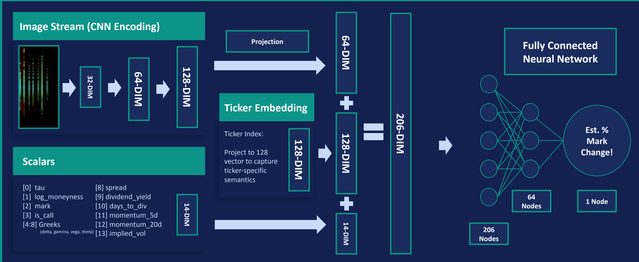
Naive Baseline: $\ell = 0$

	Contracts	Naive MAE	St Dev.
Original	36,174,983 (100%)	105.4%	1344.0%
Filtered	30,546,224 (84.4%)	13.2%	23.0%

1. Remove $|\ell| > 2.0$ (>200% overnight moves)
2. Remove spread > 0.5 (illiquid contracts)
   spread = (ask − bid) / mark
3. Select top four tickers: (MSFT, AAPL, GOOGL, AMZN)
   1.1 million training, 347k testing contracts

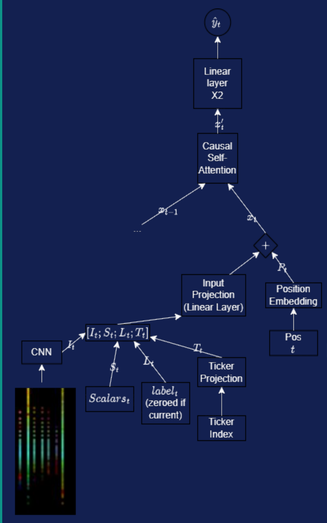
Model Architectures

Baseline Model



- CNN: Processes input images through a three-stage CNN backbone (3x3 kernels), using global average pooling to compress
- Multi-modal feature fusion: integrates visual imagery, ticker embeddings, and financial scalars into a unified representation
- Regularized MLP head: processes the fused features through a dense feed-forward network, employing layer normalization, ReLU, and dropout
- Uses same hyperparameters as Deep Residual Protection Head Model

Time-Series Causal Self-Attention



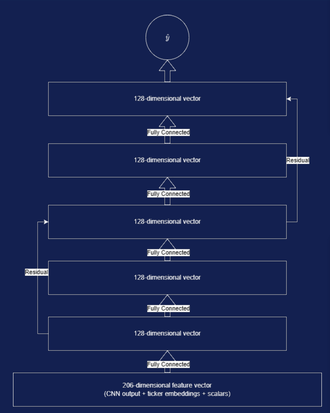
- Temporal Contextualization via self-attention: introduces a multi-head self-attention mechanism to capture long-range dependencies
- Positional encoding and masking: employs learnable positional embeddings to preserve chronological order
- Sequence length of 25, with 4 transformer heads per block and 2 blocks

Hyperparameters
- Batch size: 16
- sequence length: 25
- num heads: 4
- num layers: 2

Deep NN

- Deep residual prediction head: replaces baseline MLP with deep, dense neural network and twin residual blocks.
- Utilizes skip connections in the deep head to mitigate vanishing gradients
- Leverages the baseline model, focusing increased parameter capacity on allowing for rich feature interactions

Hyperparameters:
- LR: 1e-5
- Batch Size: 256
- Dropout: 0.2
- Hidden dim:128

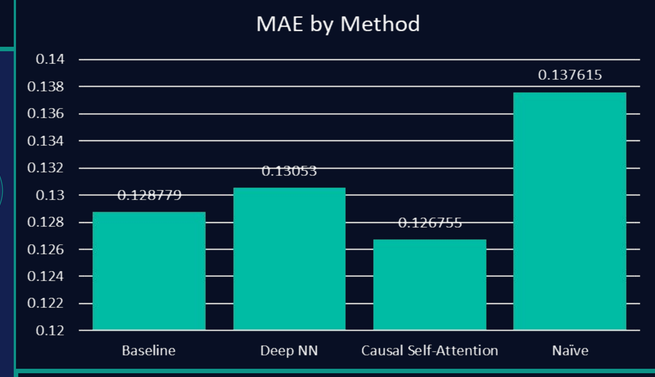


Train/Test Split

To avoid data contamination due to the time series based nature of financial data, train and test data was split based on a specific date, with the most recent data contributing to the test set.

Split On: 11/21/2024

Results



Improvement Over Naive

Baseline	Deep NN	Self-Attention
+6.421%	+5.148%	+7.892%

Discussion

- Causal self-attention performed the best, with ~.127 MAE compared to ~.138 naive MAE
  - Historical data for each contract provide key insights into "running patterns"
- Deep neural network performed the worst
  - Deep head, while supposed to increase complexity of feature extraction, may lead to overfitting

Key Takeaways

- Surface images encode predictive signal about next-day contract price movements, but cannot anticipate larger macroeconomic events which can render historical surface patterns obsolete overnight
- Time-series causal decoder modeling allows model to learn the rhythms of a contract over time, surpassing single-day snapshots
- Deeper prediction heads may not always allow for better accuracy when input data are not sufficiently rich
- Future experimentation will include bluesky text data relevant to tickers under consideration

Related Work

[1] Dubach (2024) — Source of our 104-ticker historical options chain dataset; GitHub repository accessed March 30, 2026

[2] Radfar (2025) — CNNs outperform LSTMs on financial time series; identifies temporal data leakage as a pervasive flaw in published financial DL literature

[3] Hutchinson, Lo & Poggio (1994) — First to show neural networks can recover Black-Scholes pricing from market data alone, establishing data-driven derivatives modeling without parametric assumptions

[4] Ruf & Wang (2020) — Survey of 150+ papers on neural networks for options pricing; identifies cross-contract spatial structure as an unexplored gap in the literature

[5] Bao et al. (2017) — Explicit multi-stage denoising via wavelets and autoencoders; we rely on CNN intrinsic spatial averaging instead, better suited to chaotic market dynamics