SurfaceEdge

Predicting Next-Day Options Price Changes from Options Surface Images

Caleb Bettcher

CSCI 4530 — Generative AI

The Problem

The Options Market

Options contracts give buyers the right to buy or sell an asset at a fixed price before expiry.

Millions of contracts trade daily across hundreds of underlying equities, collectively forming a rich two-dimensional structure encoding the market's real-time consensus on future uncertainty.

The Edge

A competitive edge in options pricing, even a fraction of a percent, can compound into several percentage points of alpha annually.

For institutional and retail traders alike, consistently pricing contracts ahead of the market is the difference between outperforming the S&P 500 and not.

Shortcomings

Famous models like Black-Scholes and the Binomial pricing model treat each contract through fixed scalar parameters, failing to account for the highly unpredictable nature of financial markets. No prior deep learning work has treated the full surface as a spatial visual input.

Each day, the options surface encodes collective market sentiment, but existing models largely ignore its spatial structure.

Existing Approaches & Their Gaps

Scalar-Only Models

Most options pricing models treat each contract in complete isolation; that is, each contract independently defined by its own parameters. Cross-contract spatial structure is ignored.

No surface context

LSTM / Time-Series Models

Sequential models capture temporal dynamics but require long historical windows and still fail to leverage the 2D spatial relationships across the full surface on any given day.

No spatial encoding

Temporal Data Leakage

Radfar (2025) and Ruf & Wang (2020) independently identify that a large fraction of published results are inflated by random train/test splits that violate time-series structure.

Overstated Performance

The Idea

Key Insight: Treat the options surface as a visual object. Encode it as an RGB image and apply convolutional feature extraction to capture spatial context across strikes and expiries.

Surface ⟶ Image

60×30 RGB grid
R = Implied Volatility
G = log(Open Interest)
B = log(Volume)

Encoded per trading day

CNN Feature Extraction

Convolution Network

Intrinsic noise immunity

Spatial averaging across strikes & expiries

Global average pooling

Hybrid Fusion

CNN embedding

+

Contract scalars

(τ, moneyness, greeks, etc.)

Dataset — Raw Data

104
U.S. Equity Tickers
Large-cap equities, ETFs, indices

17 yrs
Historical Coverage
2008 – 2025

~9.4 GB
Raw Parquet Data
options.parquet + underlying.parquet per ticker

200M+
Labeled Contracts
raw samples, before dataset construction

Per-Ticker Files

```
options.parquet
  -   strike
  -   expiry
  -   bid
  -   ask
  -   Implied Volatility
  -   Open Interest
  -   Volume
  -   Greeks

underlying.parquet
  -   daily OHLCV
      (Open, High, Low, Close, Volume)
  -   split coefficients
  -   dividend records
```

Dataset — Surface Construction

01 Split Adjustment

Strike prices are multiplied by the product of all future split coefficients, placing all historical contracts in consistent post-split terms.

02 Grid Binning

Each day's options chain is discretized into a fixed 60×30 grid.

Y-axis: log-moneyness (−1.0 to +1.0)
X-axis: days to expiry (1–61 days)

03 RGB Encoding

Three channels are normalized per image:

R = implied volatility,
G = log(open interest)
B = log(volume)

04 Contract Specific Scalars

Alongside each PNG, a 60×30×16 float16 array stores per-cell features:

spot, strike, τ, log-moneyness, mark, Greeks, spread, dividends, momentum

Dataset — Labels & Filtering

Label: Next-day percentage mark price change

$$\ell = (mark_{t+1} - mark_t) / mark_t$$

Contracts with no next-day quote receive $\ell = 0$ and are excluded from all evaluation.

Spread

$$spread = (ask - bid) / mark$$

Spread Filtering

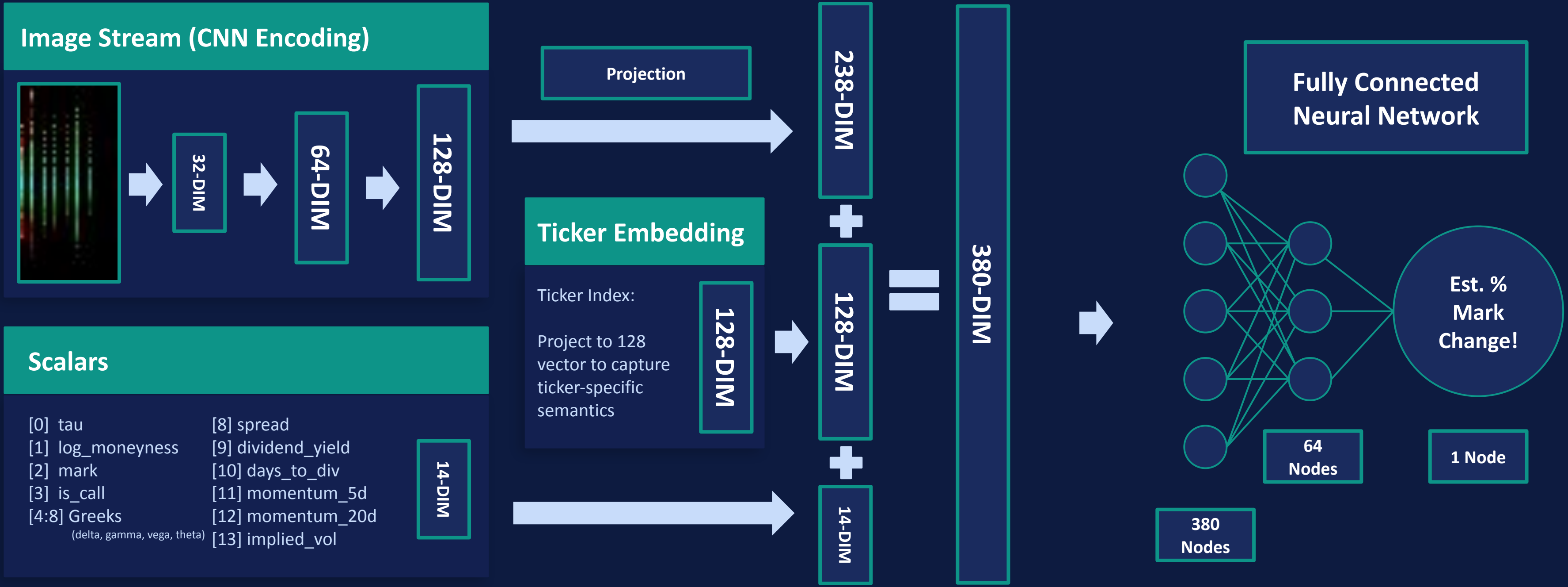
Spread Threshold	Contracts	Retained	Naive MAE	Standard Deviation
None (unfiltered)	36,174,983	100.0%	1.054 = 105.4%	13.44
≤ 0.5	30,546,224	84.4%	0.132 = 13.2%	0.230

** On average, any given contract in our training data will move 13.2% daily*

Filter Criteria

- Remove $|\ell| > 2.0$ (>200% overnight moves)
- Remove spread > 0.5 (illiquid contracts)
- Drop days with < 100 valid contracts
- Result: 92.2% MAE reduction
  Cleaner training signal, eliminates far OTM contracts

Model Architecture



Fusion: Concatenate (238 + 128 + 14 = 380 dim) → FC(380 → 64) → LayerNorm → FC(64 → 1)

Resulting Model has about 161K parameters

Model — Baseline & Measuring Performance

Naive Baseline: Predict $\ell = 0.0$ (no price change) for every contract

Baseline MAE = $\mathbb{E}[|\ell|]$ — the minimum bar the model must beat to show surface images carry predictive signal

Experiment: SurfaceEdge vs. Naive Baseline

Purpose:

Does the CNN extract predictive information from the options surface beyond the trivial zero-prediction?

Metric:

Test MAE vs. Naive MAE

The CNN shows it's able to extract information regarding current market conditions from the options surface if the test set MAE is significantly better than the Naive Baseline

	Baseline	Surface Edge
MAE	13.2%	TBD
SE Beat?	TBD	

Prevent Data Leakage:

Strict time series train/test split. Choose a day (20% of contracts), and any data after that date is test data.

Results

Filtered Dataset (4 Epochs, All 104 Tickers)

Epoch	Train MAE	Test MAE	Beat Naive?	Baseline - Test MAE	Improvement
1	0.124566 = 12.46%	0.117323 = 11.73%	YES ✓	1.47%	+11.14%
2	0.119084 = 11.91%	0.119970 = 11.99%	YES ✓	1.21%	+9.17%
3	0.111499 = 11.14%	0.125093 = 12.51%	YES ✓	0.69%	+5.23%
4	0.104604 = 10.46%	0.130104 = 13.01%	YES ✓	0.19%	+1.44%

Results (key takeaways)

- All runs outperformed naive assumption that option mark prices don’t change day-to-day
- Best run was 11.14% better than baseline
- Model training demonstrated unstable characteristics

Conclusions

Surface images carry signal - SurfaceEdge beats the naive baseline across all epochs by up to 11%. This means, on average, SurfaceEdge is within 11.73% of an options true next day price.

Chronological splits matter - enforcing strict temporal separation is essential to avoid unintentional data leakage, and for accurate test set results.

Filtering produces dramatic results - removing illiquid contracts reduced naive MAE by 92.2%!

The framing is novel - to my knowledge, no prior work treats the IV surface as an RGB image for price change prediction. This approach shows potential for groundbreaking research.

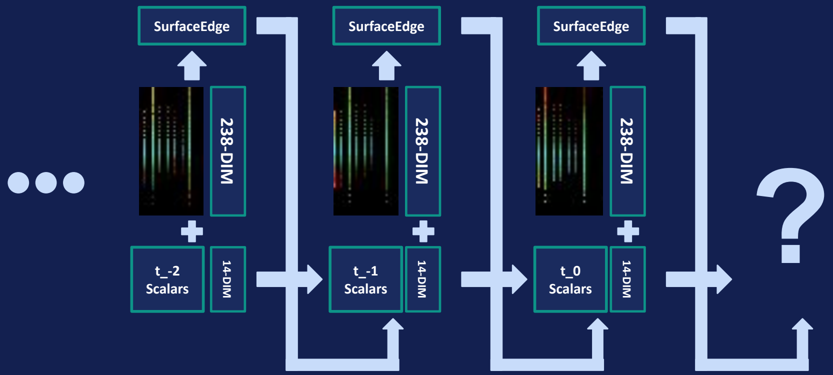
Model was unstable/imperfect - This initial model architecture demonstrated potential, but was highly unstable as demonstrated by reduced test set performance with more epochs.

Future Work

Temporal Surface Transformer

Using a similar methodology to how LLMs generate tokens in a sequence based on prior tokens.

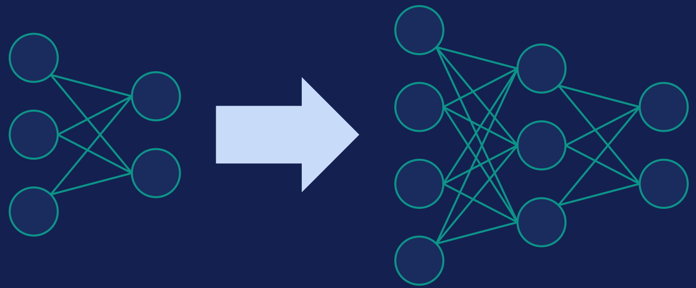
Combine the contextual power of transformers with Surface Edge!



Experiment with Deeper Models

It's well studied that, given the proper relative sizing of training data, deeper models can perform better due to their allocation.

Due to the complex nature of financial markets, SurfaceEdge may benefit from this expanded size.



I hope to expand this project beyond the current scope of this class. Perhaps, with a promising enough model, backtest against real financial data, and deploy as a form of algorithm trading!