# Time Series Forecasting with Transformers

Iram Liu, Jeffrey Xiang, Wen Chen, Ajay Sachdev, Nick Hsieh
Cornell Engineering CS 4782 Deep Learning Final Project

## Background

Transformers are well-suited for time series forecasting, as the attention mechanism can learn relationships between points in time.

However, simple linear models can **outperform** sophisticated Transformer-based models. [1]

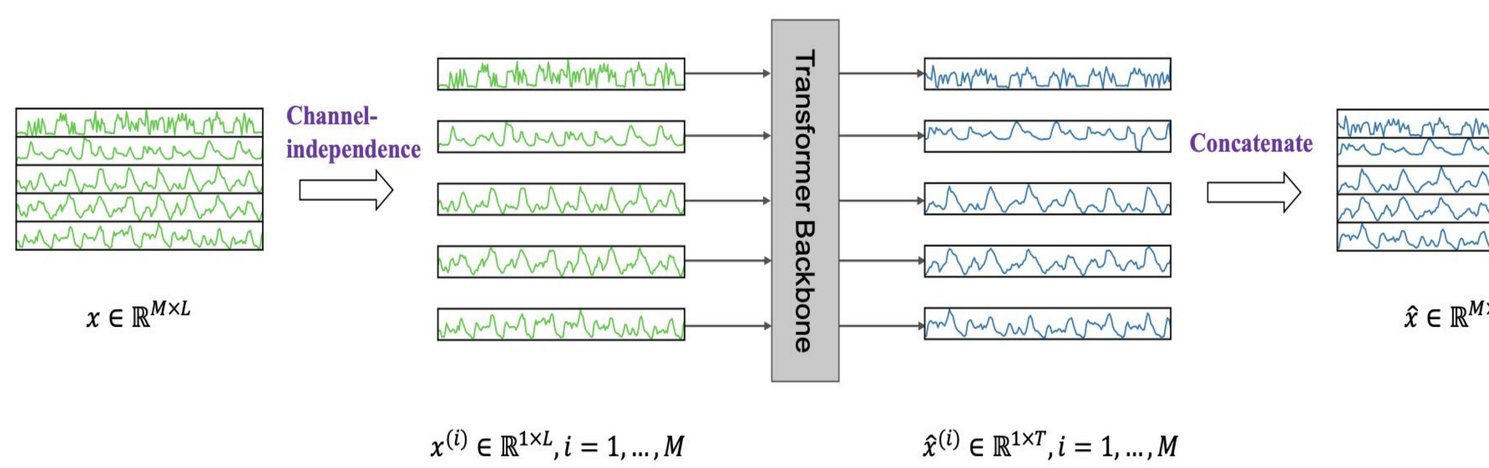
| Methods | | IMP. | Linear* | | NLinear* | | DLinear* | | FEDformer | | Autoformer | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric | | MSE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| Electricity | 96 | 27.40% | **0.140** | **0.237** | 0.141 | **0.237** | **0.140** | **0.237** | 0.193 | 0.308 | 0.201 | 0.317 |
| | 192 | 23.88% | **0.153** | 0.250 | 0.154 | **0.248** | **0.153** | 0.249 | 0.201 | 0.315 | 0.222 | 0.334 |
| | 336 | 21.02% | **0.169** | 0.268 | 0.171 | **0.265** | **0.169** | 0.267 | 0.214 | 0.329 | 0.231 | 0.338 |
| | 720 | 17.47% | **0.203** | 0.301 | 0.210 | **0.297** | **0.203** | 0.301 | 0.246 | 0.355 | 0.254 | 0.361 |

**How can we improve the performance of Transformer-based forecasting models?**

## New Design Features

Two key features are proposed in PatchTST:

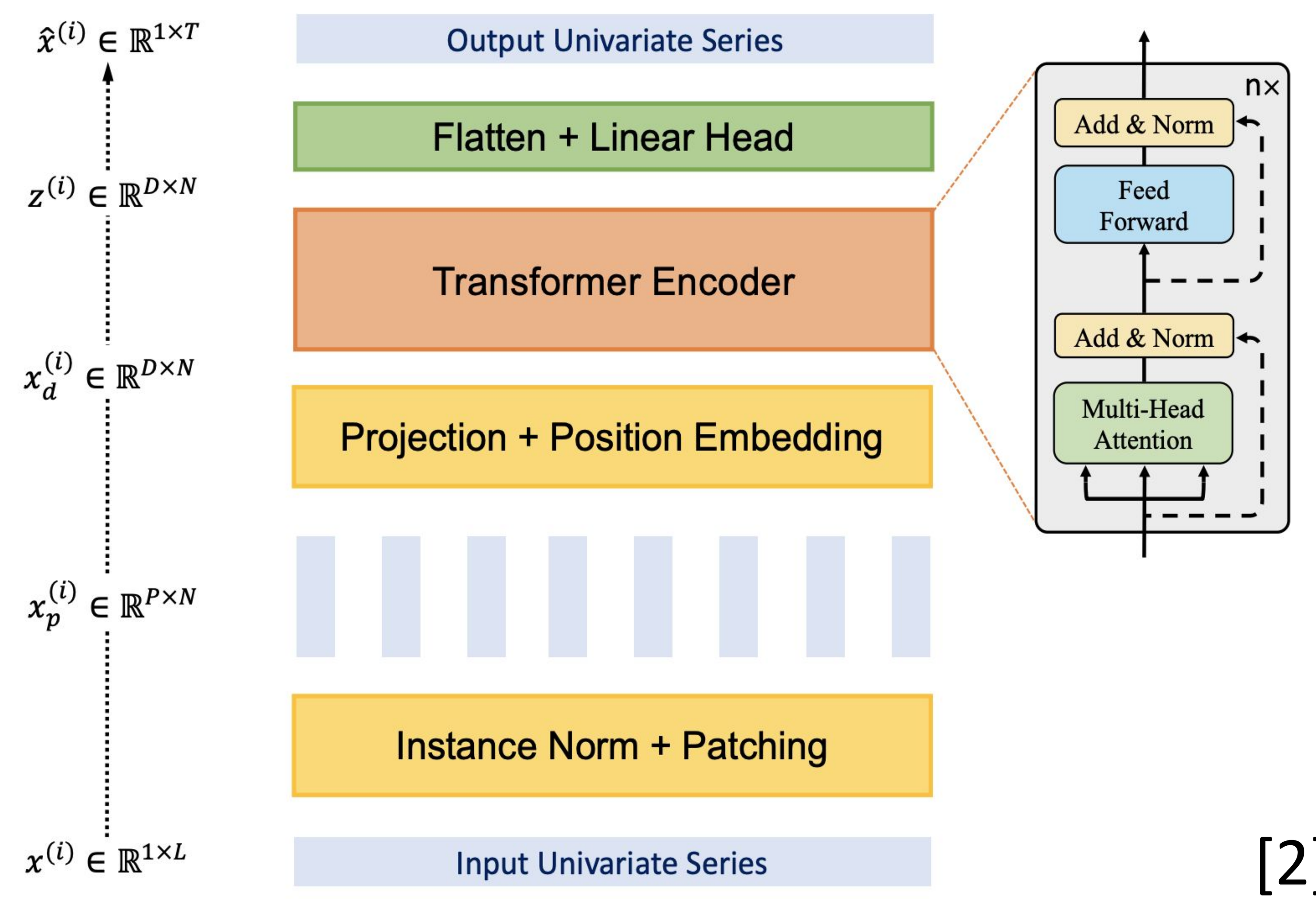
1) **Patching**
2) **Channel-independence**




[2]

**Impact:** local semantic info is better retained, reduced computation allowing for longer look back periods, better generalization

## Architecture




[2]

Apply same transformer backbone to all channels

## Datasets

We tested PatchTST on five datasets:

1) **Electricity**: electricity use for 321 households
2) **Illness**: number of patients and ILI ratio
3) **Traffic**: road occupancy rates from 860 sensors on San Francisco freeways
4) **ETTh1**: hourly powerload data from China
5) **ETTm1**: 15 minute powerload data from China

**Goal:** for each channel, we lookback 336 time steps and predict 96 time steps into the future

## Results

| Dataset | Our MSE | Paper MSE | Our MAE | Paper MAE |
|---|---|---|---|---|
| **Electricity** | 0.156 | 0.129 | 0.258 | 0.222 |
| **Illness*** | 2.071 | 1.319 | 0.974 | 0.754 |
| **Traffic** | 0.421 | 0.360 | 0.295 | 0.249 |
| **ETTh1** | 0.427 | 0.370 | 0.452 | 0.400 |
| **ETTm1** | 0.326 | 0.293 | 0.372 | 0.346 |

*model results unreliable due to dataset size and splitting

We got our metrics from the test set, using a 70/10/20 split for training, validation, and testing.

## Conclusions and Next Steps

The paper found significant improvements compared to other transformer models + basic linear models and **we were able to see the same type of improvement in our reimplementation.**

For our reimplementation, we focused on **supervised learning**. Next steps involve implementing **self-supervised learning** and testing hyperparameters.

## References

[1] Are transformers effective for time series forecasting? [Zeng et al., 2022]
[2] Long-term Forecasting with Transformers [Nie et al., 2022]