TECHNISCHE HOCHSCHULE DEGGENDORF

# Enhancing Spatiotemporal Networks with xLSTM

## A Scalar LSTM Approach for 5G Traffic Forecasting

*GI/ITG KuVS Fachgespräch "Network Softwarization"*
03-04 April 2025

Z. Bettouche, 2. April 2025

# Introduction

- Zineddine Bettouche
- Applied Computer Science Faculty in Deggendorf Institute of Technology
- Currently doing research in the application of language models beyond NLP

*zineddine.bettouche@th-deg.de*

# Contribution

# Enhancing Spatiotemporal Networks with xLSTM: A Scalar LSTM Approach for 5G Traffic Forecasting

Zineddine Bettouche, Khalid Ali, Andreas Fischer, Andreas Kassler
Deggendorf Institute of Technology
Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany
{zineddine.bettouche, khalid.ali, andreas.fischer, andreas.kassler}@th-deg.de

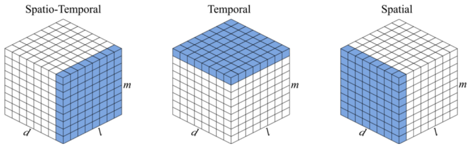
***Abstract*—Accurate spatiotemporal traffic forecasting is vital for optimizing 5G networks. Traditional LSTM models struggle with capturing complex spatiotemporal dependencies, limiting predictive performance. To address this, we propose an enhanced Spatiotemporal Network (STN) integrating Scalar LSTM (sLSTM), a more efficient variant designed to improve temporal modeling while reducing computational complexity. Our dual-path STN processes the input through an sLSTM for sequential feature extraction and a three-layer Conv3D path for spatial feature learning, with both outputs fused in a dedicated fusion layer for enhanced spatiotemporal representation. By incorporating sLSTM, our model stabilizes gradients, accelerates convergence,**

(sLSTM), introduced within the Extended-LSTM (xLSTM) framework. Unlike traditional LSTMs, sLSTM simplifies the gating mechanism using scalar operations, significantly reducing parameter complexity while improving convergence stability and temporal dependency modeling.

In this work, we enhance spatiotemporal traffic forecasting for 5G networks by integrating sLSTM into the STN framework. Our key contributions are:

- sLSTM-Enhanced STN: We incorporate sLSTM into the STN framework, reducing computational overhead

# Spatiotemporal (ST) Forecasting




## Challenge

**Multidimensional data**: time, location, features

## Approach

Enhance ST Network with **xLSTM**

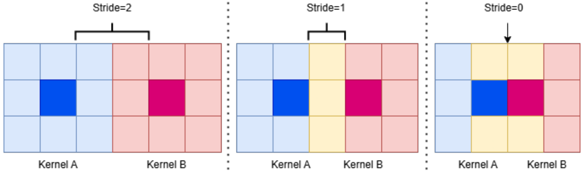
## Goal

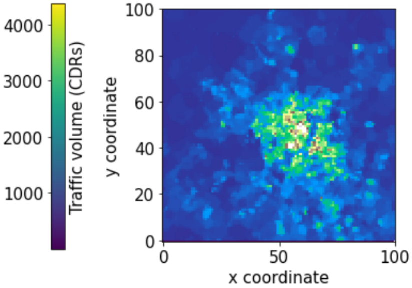
**Accurate forecasting of future frames (grids, video-frames, etc)**.

# Data and Setup: Milan Dataset

# Data and Setup: Milan Dataset

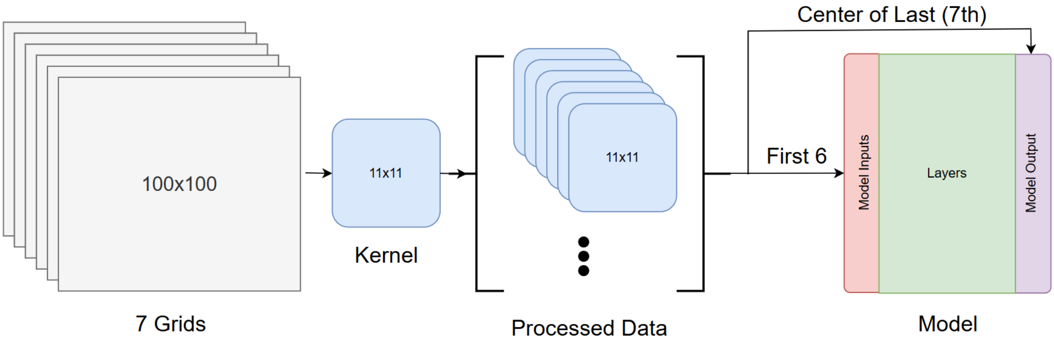
- **Training Data:** Milan dataset (70% of data)
  - 1 million samples with a stride of 6
  - 40 epochs of training
- **Testing Data:**
  - 7 million samples from 15% of the data with a stride of 1



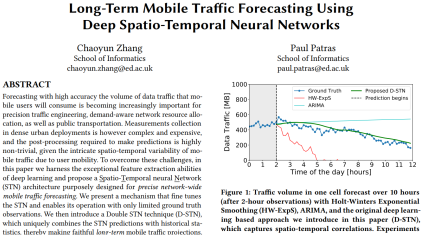
# Data Pre-Processing

# Data Pre-Processing



# Baseline: Spatiotemporal Network (STN)

# Baseline: Spatiotemporal Network (STN)



**Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks**

Chaoyun Zhang
School of Informatics
chaoyun.zhang@ed.ac.uk

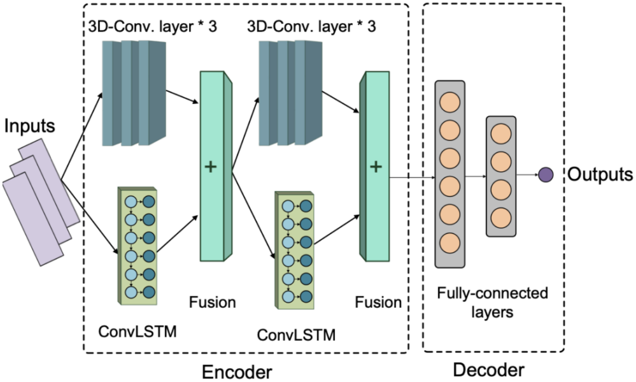
Paul Patras
School of Informatics
paul.patras@ed.ac.uk

**ABSTRACT**

Forecasting with high accuracy the volume of data traffic that mobile users will consume is becoming increasingly important for precision traffic engineering, demand-aware network resource allocation, as well as public transportation. Measurements collection in dense urban deployments is however complex and expensive, and the post-processing required to make predictions is highly non-trivial, given the intricate spatio-temporal variability of mobile traffic due to user mobility. To overcome these challenges, in this paper we harness the exceptional feature extraction abilities of deep learning and propose a Spatio-Temporal neural Network (STN) architecture purposely designed for *precise network-wide mobile traffic forecasting*. We present a mechanism that fine tunes the STN and enables its operation with only limited ground truth observations. We then introduce a Double STN technique (D-STN), which uniquely combines the STN predictions with historical statistics, thereby making faithful *long-term* mobile traffic projections.



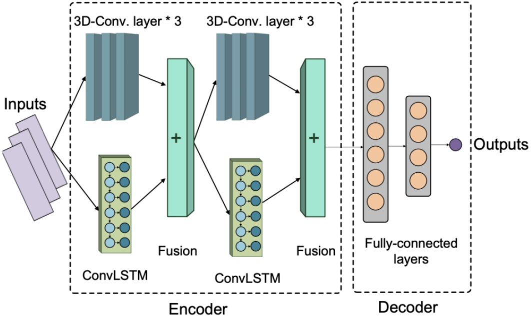
**Figure 1: Traffic volume in one cell forecast over 10 hours (after 2-hour observations) with Holt-Winters Exponential Smoothing (HW-ExpS), ARIMA, and the original deep learning based approach we introduce in this paper (D-STN), which captures spatio-temporal correlations. Experiments**



# Modernizing the STN

# Modernizing the STN

- Attention-based fusion layer: Cross-attention
- A better LSTM: the extended LSTM



# A Better LSTM: xLSTM

# A Better LSTM: xLSTM

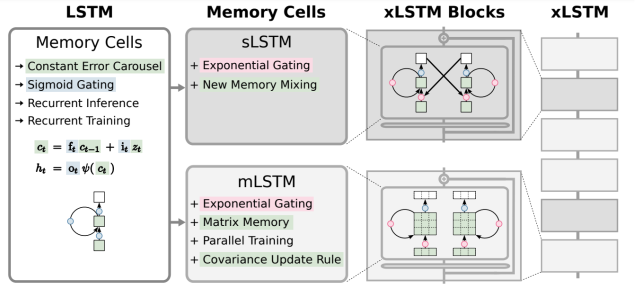
## xLSTM: Extended Long Short-Term Memory


Maximilian Beck* [1,2,3] Korbinian Pöppel* [1,2,3] Markus Spanring [1]
Andreas Auer [1,2] Oleksandra Prudnikova [1] Michael Kopp
Günter Klambauer [1,2,3] Johannes Brandstetter [1,2,3] Sepp Hochreiter [1,2,3]
*Equal contribution
[1]ELLIS Unit, LIT AI Lab, Institute for Machine Learning, JKU Linz, Austria
[2]NXAI Lab, Linz, Austria, [3]NXAI GmbH, Linz, Austria



### Abstract

In the 1990s, the constant error carousel and gating were introduced as the central ideas of the Long Short-Term Memory (LSTM). Since then, LSTMs have stood the test of time and contributed to numerous deep learning success stories, in particular they constituted the first Large Language Models (LLMs). However, the advent of the Transformer technology with parallelizable self-attention at its core marked the dawn of a new era, outpacing LSTMs at scale. We now raise a simple question: How far do we get in language modeling when scaling LSTMs to

# xLSTM: Performance on PALOMA

# xLSTM: Performance on PALOMA

| | Model | #Params M | C4 | MC4 EN | Wikitext 103 | Penn Treebank | Red Pajama | Refined Web | Dolma | M2D2 S2ORC | M2D2 Wikipedia | C4 Domains | Dolma Subreddits | Dolma Coding | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 125M | RWKV-4 | 169.4 | 26.25 | 22.33 | 29.18 | 38.45 | 8.99 | 32.47 | 17.04 | 23.86 | 21.42 | 22.68 | 37.08 | 5.12 | 23.74 |
| | Llama | 162.2 | 24.64 | 17.23 | 23.16 | 31.56 | 8.26 | 29.15 | 15.10 | 19.71 | 20.41 | 21.45 | 36.73 | 3.61 | 20.92 |
| | Mamba | 167.8 | 23.12 | 17.04 | 22.49 | 30.63 | 7.96 | 27.73 | 14.60 | 19.38 | 19.36 | 20.14 | 34.32 | 3.77 | 20.05 |
| | xLSTM[1:0] | 163.8 | 22.54 | 16.32 | 21.98 | 30.47 | 7.80 | 27.21 | 14.35 | 19.02 | 19.04 | 19.65 | 34.15 | 3.64 | 19.68 |
| | xLSTM[7:1] | 163.7 | **22.39** | **16.13** | **21.47** | **30.01** | **7.75** | **26.91** | **14.13** | **18.6** | **18.84** | **19.52** | **33.9** | **3.59** | **19.44** |
| 350M | RWKV-4 | 430.5 | 19.55 | 15.82 | 19.64 | 27.58 | 6.97 | 24.28 | 12.94 | 17.59 | 15.96 | 16.98 | 29.40 | 3.90 | 17.55 |
| | Llama | 406.6 | 18.38 | 13.28 | 16.41 | **21.82** | 6.56 | 22.09 | 11.76 | 15.05 | 15.25 | 15.99 | 28.30 | 3.12 | 15.67 |
| | Mamba | 423.1 | 17.33 | 13.05 | 16.11 | 22.24 | 6.34 | 21.04 | 11.42 | 14.83 | 14.53 | 15.16 | 27.02 | 3.20 | 15.19 |
| | xLSTM[1:0] | 409.3 | 17.01 | **12.55** | **15.17** | 22.51 | **6.20** | **20.66** | **11.16** | **14.44** | **14.27** | **14.85** | 26.70 | **3.08** | **14.88** |
| | xLSTM[7:1] | 408.4 | **16.98** | 12.68 | 15.43 | 21.86 | 6.23 | 20.70 | 11.22 | 14.62 | 14.30 | **14.85** | **26.61** | 3.11 | **14.88** |
| 760M | RWKV-4 | 891.0 | 15.51 | 12.76 | 14.84 | 21.39 | 5.91 | 19.28 | 10.70 | 14.27 | 13.04 | 13.68 | 24.22 | 3.32 | 14.08 |
| | Llama | 834.1 | 15.75 | 11.59 | 13.47 | 18.33 | 5.82 | 19.04 | 10.33 | 13.00 | 13.05 | 13.76 | 24.80 | 2.90 | 13.49 |
| | Mamba | 870.5 | 15.08 | 11.54 | 13.47 | 19.34 | 5.69 | 18.43 | 10.15 | 13.05 | 12.62 | 13.25 | 23.94 | 2.99 | 13.30 |
| | xLSTM[1:0] | 840.4 | **14.60** | **11.03** | **12.61** | 17.74 | **5.52** | **17.87** | **9.85** | **12.50** | **12.20** | **12.81** | 23.46 | **2.87** | **12.76** |
| | xLSTM[7:1] | 839.7 | 14.72 | 11.11 | 12.68 | **17.61** | 5.55 | 18.01 | 9.87 | 12.59 | 12.25 | 12.89 | **23.43** | 2.88 | 12.80 |
| 1.3B | RWKV-4 | 1515.2 | 14.51 | 12.04 | 13.73 | 19.37 | 5.62 | 18.25 | 10.11 | 13.46 | 12.10 | 12.87 | 22.85 | 3.25 | 13.18 |
| | Llama | 1420.4 | 13.93 | 10.44 | 11.74 | **15.92** | 5.29 | 17.03 | 9.35 | 11.61 | 11.53 | 12.24 | 22.63 | 2.74 | 12.04 |
| | Mamba | 1475.3 | 13.35 | 10.40 | 11.76 | 16.65 | 5.21 | 16.50 | 9.17 | 11.73 | 11.18 | 11.83 | 21.43 | 2.83 | 11.84 |
| | xLSTM[1:0] | 1422.6 | **13.13** | **10.09** | 11.41 | **15.92** | **5.10** | **16.25** | **9.01** | **11.43** | **10.95** | **11.60** | **21.29** | **2.73** | **11.58** |
| | xLSTM[7:1] | 1420.1 | 13.31 | 10.21 | **11.32** | 16.00 | 5.16 | 16.48 | 9.11 | 11.61 | 11.10 | 11.76 | 21.50 | 2.75 | 11.69 |

# Evaluation: STN-sLSTM v. STN

# Evaluation: STN-sLSTM v. STN

| Model | MAE | RMSE | $R^2$ Score | SSIM |
|---|---|---|---|---|
| STN | 7.3917 | 16.8849 | 0.9546 | 0.9853 |
| STN + Cross-att. | 7.2104 | 15.5359 | 0.9616 | 0.9858 |
| STN-sLSTM-AverageFusion | **5.6704** | **12.3191** | **0.9758** | **0.9912** |

*trained for 40 epochs on 1 million data-points with stride=6,*
*tested on 7 million unseen points (6, 11x11)*

# Evaluation on Autoregression: Visualization

**Setup:** 6 kernel frames in, 6 kernel frames out. I.e., first predicted frame becomes "new"last frame for the next step.

# Evaluation on Autoregression: Visualization

**Setup:** 6 kernel frames in, 6 kernel frames out. I.e., first predicted frame becomes "new"last frame for the next step.

# Evaluation on Autoregression: Results

**Setup:** 6 kernel frames in, 6 kernel frames out. I.e., first predicted frame becomes "new"last frame for the next step.

# Evaluation on Autoregression: Results

**Setup:** 6 kernel frames in, 6 kernel frames out. I.e., first predicted frame becomes "new"last frame for the next step.

*STN (original paper)*

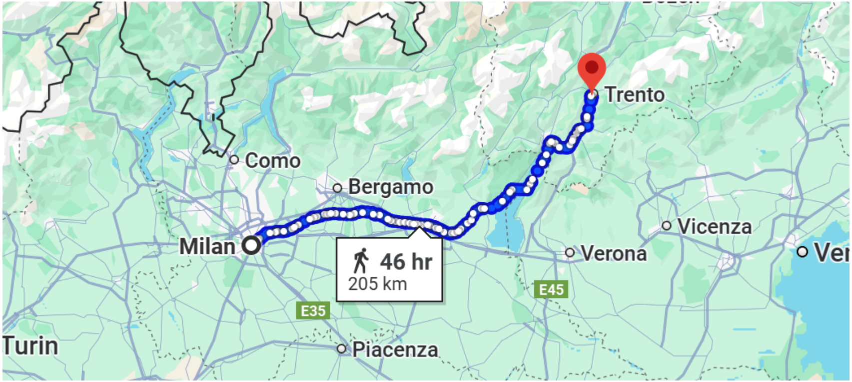
| Step | MAE | RMSE | SSIM |
|---|---|---|---|
| 1.0 | 8.92385 | 33.0079 | 0.940624 |
| 2.0 | 11.1397 | 37.4584 | 0.902465 |
| 3.0 | 12.9716 | 40.4277 | 0.878164 |
| 4.0 | 14.4332 | 44.2043 | 0.844633 |
| 5.0 | 16.2411 | 45.4263 | 0.806056 |
| 6.0 | 17.6542 | 47.3949 | 0.765161 |

*STN-sLSTM-AverageFusion*

| Step | MAE | RMSE | SSIM |
|---|---|---|---|
| 1.0 | 6.44511 | 23.4955 | 0.963972 |
| 2.0 | 7.70051 | 24.6218 | 0.942716 |
| 3.0 | 8.86864 | 26.0003 | 0.937315 |
| 4.0 | 9.79182 | 29.6569 | 0.92417 |
| 5.0 | 10.866 | 32.0609 | 0.907771 |
| 6.0 | 11.7697 | 34.5118 | 0.891746 |

# Dealing with Unseen Data: Trentino Dataset

# Dealing with Unseen Data: Trentino Dataset



# Dealing with Unseen Data: Trentino Dataset

# Dealing with Unseen Data: Trentino Dataset

| Model (6x11x11 in, 1 out) | MAE | RMSE | $R^2$ Score | SSIM |
|---|---|---|---|---|
| STN | 2.6344 | 7.6370 | 0.9116 | 0.9744 |
| STN-sLSTM-AverageFusion | **1.6915** | **5.2119** | **0.9588** | **0.9895** |

*STN (Autoregression)*

| Step | MAE | RMSE | SSIM |
|---|---|---|---|
| 1 | 2.65936 | 22.9203 | 0.98781 |
| 3 | 3.63895 | 28.0328 | 0.974009 |
| 5 | 4.56257 | 31.135 | 0.956859 |
| 6 | 4.89946 | 32.0503 | 0.949903 |

*STN-sLSTM-AvgFusion (Autoregression)*

| Step | MAE | RMSE | SSIM |
|---|---|---|---|
| 1 | 1.79022 | 18.4738 | 0.99405 |
| 3 | 2.35933 | 20.288 | 0.990245 |
| 5 | 3.00794 | 21.6852 | 0.984203 |
| 6 | 3.27395 | 22.2906 | 0.981255 |

# Conclusion

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations
- **Better Long-term Forecasting**: In autoregression tasks, the STN-sLSTM showed lower error accumulation over time

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations
- **Better Long-term Forecasting**: In autoregression tasks, the STN-sLSTM showed lower error accumulation over time

## Outlook

- More modern (attention-based/SS-based) baselines (e.g., STTRE and Mamba).

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations
- **Better Long-term Forecasting**: In autoregression tasks, the STN-sLSTM showed lower error accumulation over time

## Outlook

- More modern (attention-based/SS-based) baselines (e.g., STTRE and Mamba).
- Ablation studies.

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations
- **Better Long-term Forecasting**: In autoregression tasks, the STN-sLSTM showed lower error accumulation over time

## Outlook

- More modern (attention-based/SS-based) baselines (e.g., STTRE and Mamba).
- Ablation studies.
- Integrating the xLSTM into PredRNN++ architecture.

# Conclusion

## Summary

- **Improved Performance**: STN-sLSTM-AverageFusion achieved better metrics across all evaluations
- **Better Long-term Forecasting**: In autoregression tasks, the STN-sLSTM showed lower error accumulation over time

## Outlook

- More modern (attention-based/SS-based) baselines (e.g., STTRE and Mamba).
- Ablation studies.
- Integrating the xLSTM into PredRNN++ architecture.
- Video-based Forecasting (baselines: ViT, ViM...)

# Now For Your Questions