

TEMPTPI: INFORMER-BASED TRAJECTORY PREDICTION FOR MARITIME VESSELS

Kevin Ferneding[✉], Veronika Lietavcova[✉], Aleksandra M. Blachowiak[✉], Peder Heiselberg[✉]

Technical University of Denmark

Anker Engelunds Vej 101 2800 Kongens Lyngby, Denmark

ph@space.dtu.dk

Abstract—Accurate long-term trajectory prediction for maritime vessels is essential for safety and logistical efficiency. While deep learning models, particularly Transformers, have shown promise in processing Automatic Identification System (AIS) data, they often struggle with the quadratic computational complexity of self-attention and the loss of accuracy over extended forecasting horizons. This study proposes TempTPI, a novel prediction framework that integrates an Informer-based encoder with a multi-channel temporal encoding mechanism. The Informer architecture leverages a ProbSparse self-attention mechanism to reduce computational overhead and focus on the most significant dependencies, while the temporal encoder utilizes Fourier-like frequency expansions to capture cyclic patterns (hourly, daily, and seasonal) in vessel behavior. We evaluate our model against the state-of-the-art TPTrans architecture using AIS data from Danish waters. Experimental results demonstrate that TempTPI consistently outperforms existing methods across prediction windows of 1 to 5 hours. Notably, at a 5-hour horizon, the proposed model achieves a 55% improvement in Mean Squared Error (MSE), offering a robust solution for long-range maritime situational awareness.

Index Terms—Trajectory prediction, Automatic Identification System (AIS), Informer, Deep learning, Spatio-temporal modeling, Maritime safety, Long-sequence forecasting.

I. INTRODUCTION

Maritime cargo vessels make up a large portion of global trade [1] and require precise orchestration to run flawlessly. Large numbers of simultaneous maritime vessels in limited space logically lead to an increase in complexity for securing safe and efficient transportation [2]. Thus, the potential commercial benefits of a high-accuracy trajectory prediction for commercial vessels in order to enable a reliable vessel coordination are very high [3, 4].

A widely used communication system in the maritime domain is the *Automatic Identification System* (AIS), which aims to transmit real-time information about the current status of the ship [5] for collision avoidance. Through the wide distribution of this system, a significant amount of data has been collected over time, which provides a valuable foundation for deep learning research. The Danish Maritime Authority *Søfartsstyrelsen* publicly provides a collection of historic AIS data from the region of Denmark [6]. The system transmits two main types of data: static information (e.g., vessel name, type, and MMSI) and dynamic information (e.g., timestamp, longitude, Latitude, Speed Over Ground (SOG), and Course Over Ground (COG)). While AIS data is invaluable, its quality

can be heterogeneous. It is frequently affected by irregularities, including shared or spoofed MMSI identifiers, signal dropouts, and physically implausible transmissions (e.g., sudden jumps in location or speed). These inconsistencies necessitate rigorous preprocessing. Furthermore, maritime vessel traffic exhibits seasonal patterns, with private and recreational ships dominating during summer months. To ensure the relevance of our predictive model to global trade and commercial logistics, we focus exclusively on sampled data from the 1st to the 14th of Oct. 2025, selecting a time frame that emphasizes stable and high-volume movements of commercial cargo vessels.

Recent work increasingly applies deep learning to AIS-based vessel trajectory prediction, with Transformer architectures receiving particular attention due to their ability to capture long-range temporal dependencies. In addition to architectural innovations, several complementary strategies have been proposed to improve long-term predictions. SEMINT [7] incorporates large language models and historical maritime knowledge. Multi-modal knowledge-enhanced frameworks [8] combine AIS data with environmental and domain information, and probabilistic deep learning approaches [9] provide uncertainty-aware forecasts. These methods highlight challenges in long-term trajectory prediction that extend beyond standard sequence modeling. TPTrans [10] addresses some of these by retaining the standard Transformer encoder-decoder structure and augmenting it with convolutional layers that extract local spatiotemporal features before global attention is applied. While this CNN-Transformer hybrid improves prediction accuracy and captures complex spatial and temporal relationships in AIS data, it inherits several limitations of existing deep learning approaches: the model does not incorporate explicit temporal context features, predictions are generated in absolute rather than relative coordinates, and confidence decreases as the prediction horizon grows. Similar limitations are observed in TrAISformer [11], which employs a spatio-temporal Transformer and demonstrates improvements over recurrent architectures, yet still exhibits increasing error for extended prediction horizons.

These limitations are compounded by the computational constraints of the vanilla transformer, whose quadratic L^2 self-attention cost becomes prohibitive for long input sequences. This motivates the exploration of architectures specifically designed for long-horizon forecasting. The Informer [12] architecture addresses these challenges through probabilistic

sparse attention and a generative decoder, enabling efficient processing of long sequences while avoiding iterative autoregressive decoding. Across four benchmark datasets, Informer demonstrates stable long-range behavior, with errors increasing smoothly as the horizon widens, and consistently outperforms related Transformer variants such as Reformer [13] and LogTrans [14] as well as LSTMa [15] networks.

Taken together, TPTrans and informers provide complementary perspectives on trajectory prediction: TPTrans highlights the value of combining local convolutional features with attention, while informers show how architectural efficiency and sparse attention can support long-range forecasting. These developments form the basis of our benchmarking and clarify the methodological gaps our work aims to address.

II. METHODOLOGY

A. Data Processing

For the preprocessing of the AIS trajectory data, the following steps were undertaken: AIS messages outside the defined geographical study region were removed, and only vessels with valid MMSI identifiers and acceptable mobile classes (A and B) were retained. For each vessel, the trajectory was ordered by timestamp, duplicate records were removed, and physically implausible tracks were excluded based on minimum duration, maximum allowable SOG, and overall track length. Trajectories were segmented whenever the temporal gap between consecutive AIS messages exceeded 15 minutes, and segments failing the same quality criteria (minimum duration and motion) were discarded. All remaining trajectories were resampled to a fixed 6-minute interval and projected from geographic coordinates into Web Mercator (x, y) coordinates to enable faster distance calculations for filtering. Physically inconsistent points were filtered out by enforcing limits on step distance (≤ 50 km), time gaps (≤ 4 h), and implied speed (≤ 100 km/h), while points below a minimum speed threshold of 6 km/h were also removed to exclude near-stationary behaviour. Cleaned trajectories were divided into continuous blocks, and a sliding window technique was applied: a 60-minute window was moved with fixed stride, with the first 30 minutes forming the model input and the following 30 minutes serving as prediction targets.

B. Model Architecture

The proposed architecture TempTPI is closely related to the TPTrans model [10], but introduces two novel improvements in the form of a temporal input encoding and an Informer encoder. The input is processed in two separate encoder channels and combined before using a positional encoder [16]. The encoded input goes through a 1d-convolutional layer, followed by an informer encoder layer. The encoded content is mean-pooled and combined with a positional decoder output, before entering separate decoders in the form of fully connected layers for latitude and longitude prediction. The following paragraphs go into more detail on the temporal encoding and the informer layer.

The temporal information transmitted in the AIS data serves as a valuable resource to detect seasonal pattern variations in the vessel behavior. To allow our model to capture these variations, we propose a temporal encoder in the first layer of the model.

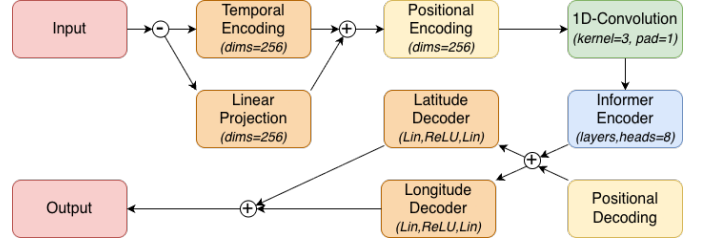


Fig. 1. Model Architecture of TempTPI

The dataset was extended to contain sine and cosine cyclic transformations to capture the *hour of day*, *day of week* and *month of year* cycles. The temporal encoder uses a linear projection for each cyclic feature, as well as one bias parameter for each projection. To process a richer representation of the cyclic features [17], we introduce a Fourier-like frequency expansion across a frequency range K such that the temporal encoding TE for each cyclic feature $c \in \{h, d, m\}$ is given as:

$$TE_c(x) \begin{cases} \sin(k * x) \\ \cos(k * x), \text{ for } k = 1, \dots, K. \end{cases}$$

These expanded features are fed into the corresponding projection layer. The encoder then combines all projections and their respective biases. The output of the temporal encoder is combined with the output of the normal data projection layer and fed into the positional encoder.

TempTPI builds on the **Informer** architecture, which relies on three main components: ProbSparse self-attention, self-attention distilling, and a generative decoder. ProbSparse Self-attention that is defined as:

$$\mathcal{A}(Q, K, V) = \text{Softmax}\left(\frac{\overline{Q}K^\top}{\sqrt{d}}\right)V, \quad (1)$$

where $\overline{Q}$ is a sparse matrix of the same size as Q , containing only the top- u queries under the sparsity measurement $M(q, K)$. This allows the model to evaluate only $O(\log L_Q)$ dot-products for each query, instead of the full $O(L_Q)$ required in the standard Transformer. The intuition behind this mechanism can be seen on Figure 2.

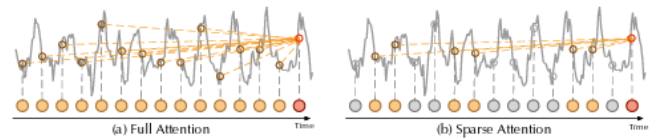


Fig. 2. Full vs Sparse Attention [18]

The Informer also modifies the architecture through self-attention distilling and a generative decoder. The encoder

compresses the sequence after each ProbSparse attention block using convolution and pooling, reducing memory from quadratic to roughly linear $O((2 - \varepsilon)L \log L)$ and enabling efficient processing of very long inputs. The decoder replaces autoregressive decoding with a single-step generative formulation, taking a start token and predicting the full horizon at once, which removes cumulative error and avoids the $O(L_y)$ iterative cost of the standard Transformer [12].

C. Model Training

The proposed architecture, TempTPI, was evaluated alongside TPInform (TPTrans, with an Informer encoder instead of a transformer) and TPTrans. The primary goal of the training phase was to establish the effectiveness of the temporal encoder and to quantify the contribution of the ProbSparse self-attention mechanism via an ablation study. Since TempTPI showed consistent improvement to TPInform, TPInform is omitted from the results.

All models were developed and trained using PyTorch 2.4.1 and executed on GPU with an available 4GB per core. The optimization strategy employed the Adam optimizer [19] with a learning rate $\mu = 1e-4$ and a batch size of 32. Training was executed for a maximum of 200 epochs. Additionally, a standard train-validation-test split (70%, 20%, and 10%, respectively) was applied to the total dataset. To compare the proposed trajectory prediction model with the aforementioned TPTrans model [10], we used *Mean Squared Error* (MSE) as the loss function. The loss was calculated with respect to the predicted and actual latitude and longitude as

$$MSE = \frac{1}{2N} \sum_{i \in N} [(lon_{true}^{(i)} - lon_{pred}^{(i)})^2 + (lat_{true}^{(i)} - lat_{pred}^{(i)})^2]$$

A systematic ablation study was designed to investigate the model's sensitivity to the data sampling parameters and its impact on the performance quality. The experiments were split into two blocks, targeting the goal of 21 tests for each model. The first block of experiments focused on the impact of the temporal dimensions of the input and prediction sequences. These tests used a full dataset ($k = 1000$ unique MMSIs of vessels) and a fixed sliding window stride of 30 minutes. The varied parameters were the input window size, providing a historical context across 240, 360, 480 minutes, and the prediction window size, targeting the forecast length varying across 60, 120, 180, 240, 300 minutes. The output of this experiment can be seen in Figure 3. While TempTPI shows an improvement, especially for longer predictions, it can also be seen that the performance dropped for longer input lengths. Since the dataset size decreases with increasing input length, we take from this that TempTPI requires more input to achieve good results.

To complete the study and evaluate model prediction quality under varying data volume and stride, a second block of 6 experiments was conducted. These tests used the fixed value of the input window size of 480 minutes and the prediction window size of 300 minutes. The varied parameters were the data volume k sampled across 100, 500, 1000 unique MMSIs

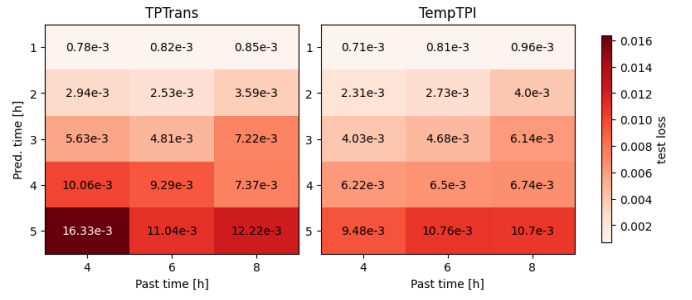


Fig. 3. Ablation study results for varying historical context and prediction time

simulating scenarios with high, moderate, and abundant data. The other tested parameter was sliding window stride that varied across 15, 30 minutes. The impact of this experiment can be seen in Figure 4. Since TempTPI performs significantly worse on the smallest sample size, it confirms the previous assumption that the required amount of data is larger than for TPTrans.

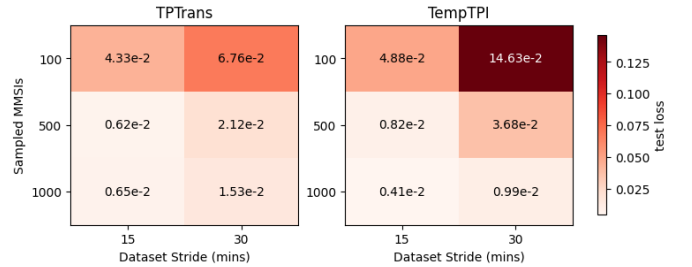


Fig. 4. Ablation study results for varying data volume and stride

III. RESULTS

To evaluate the performance of the model, we used the same evaluation metrics as explained in the previous section. The training for these models was conducted on identical training and validation datasets, with an increasing sample number for increasing prediction length, to prevent the underfitting mentioned in the ablation study by keeping a consistent dataset length of approx. 40000 windows. The tests were performed on a separate test dataset, consisting of approximately 8000 trajectory windows with a stride of 15 minutes, which was not part of the model training procedure. For all prediction lengths, a fixed past time window of 5 hours was used. The curve was fitted to a quadratic function $y(x) = ax^2 + bx + c$ using *curve_fit* [20], to get a generalized estimate of the loss over prediction length.

A comparison of the loss for different prediction lengths can be seen in Fig. 5. It can be seen that switching to an informer-based encoder (blue curve) significantly reduces the loss in comparison to the original TPTrans model (green curve). Additionally, adding the temporal encoder to this architecture gives a consistent improvement across all prediction lengths. Most importantly, at the prediction length of 5 hours, TempTPI shows a roughly 55% improvement over TPTrans.

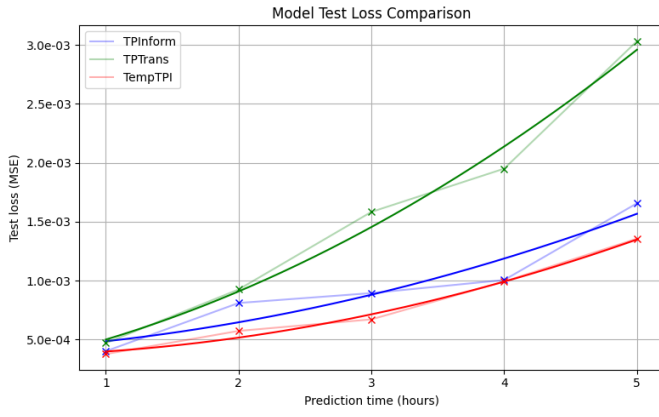


Fig. 5. Comparison of loss (MSE) over prediction length

In Fig. 6 we can see the development of the training and validation loss values during the training for the prediction length of 5 hours. The used learning rate was $\mu = 1e-4$ for both models. While TempTPI initially has a higher loss, after around 50 epochs it performs better than TPTrans. It additionally shows a longer and more stable learning curve.

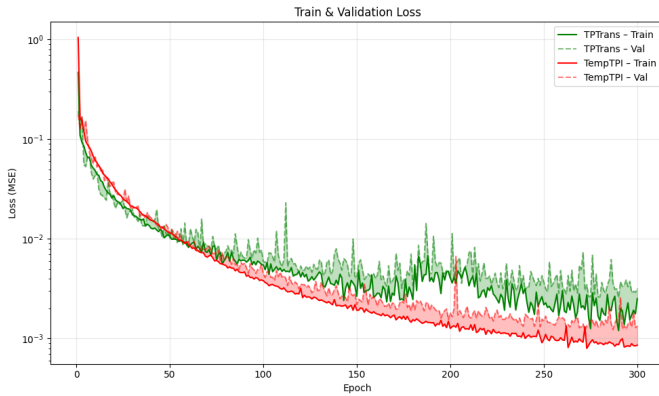


Fig. 6. Comparison of train and val. loss (MSE) over epochs

Looking at some samples of the predicted trajectories for long predictions, shown in Fig. 7, one can see that, with the given actual trajectory (blue), the predicted paths of TempTPI (red) are generally closer to the actual paths (light blue) than predictions from TPTrans (green).

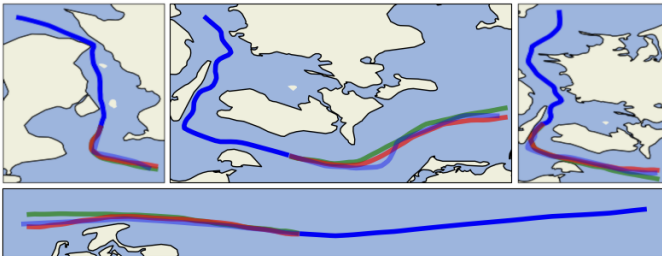


Fig. 7. Comparative prediction samples for 3-5 hours. TempTPI (red), actual (light blue), TPTrans (green)

We can see a tendency for curves to be underestimated in some and overestimated in other predictions. The predicted trajectories successfully navigate narrow pathways, e.g., in the Øresund and Great Belt straits, relatively well, but sometimes violate water boundaries.

In general the proposed model is very effective, even compared to a state-of-the-art model such as TPTrans. The significant reduction in test loss for very long predictions suggests greater stability and a higher learning potential, especially regarding cyclic fluctuations in vessel behavior captured with the temporal encoding.

IV. DISCUSSION

This study shows the potential of using an Informer-based encoder in combination with a temporal encoding, which has been compared to a transformer-based model and yielded significantly better results. The higher accuracy, especially for longer prediction windows, combined with capturing cyclic variances, suggests high potential for usage in the field of maritime logistics. The focus of this paper has been on the model architecture, and the layer parameters were kept relatively consistent; experimenting with different parameters could further increase the efficiency.

However, it has to be mentioned that the used dataset only captured traffic for two weeks in a single month. Since the temporal encoder aims to capture hourly, daily, and monthly variations, the model could benefit from further training on a larger dataset, ensuring that all seasons, such as spring and summer, are well represented and enabling further prediction windows. As of right now, only predictions in the area of Denmark have been investigated. To account for varying maritime domains, this work could potentially be extended to other areas of the world, focusing on high-traffic areas such as the Suez Canal or Panama Canal.

The results have partially shown trajectories being predicted to enter land boundaries. In a novel experiment, we introduced a land punishment during training, but found that it destabilized the training and reduced overall performance, thus it was left out of this paper. Future research could investigate options for penalized training or other land avoidance techniques in more detail.

To summarize, the presented architecture demonstrated a promising foundation for long-term prediction, but its full potential of capturing seasonal variations remains to be investigated in the future. These findings highlight a pathway toward more accurate, reliable, and operationally valuable forecasting models within the global maritime domain.

REFERENCES

- [1] United Nations Conference on Trade and Development (UNCTAD), "Shipping data: seaborne trade statistics," <https://unctad.org/news/shipping-data-unctad-releases-new-seaborne-trade-statistics>, 2025, accessed: 2025-11-23.
- [2] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Challenges in video based object detection in maritime scenario using computer vision,"

- CoRR*, vol. abs/1608.01079, 2016. [Online]. Available: <http://arxiv.org/abs/1608.01079>
- [3] X. Li, C. Liu, J. Li, L. Zhao, and Z. Du, "Advancing ship trajectory prediction: Integrating deep learning with enhanced reference trajectory correction techniques," *Ocean Engineering*, vol. 311, p. 118880, 2024. [Online]. Available: <https://doi.org/10.1016/j.oceaneng.2024.118880>
 - [4] Z. L. Szpak and J. R. Tapamo, "Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6669–6680, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417410013060>
 - [5] A. Felski, K. Jaskólski, and P. Banyś, "Comprehensive assessment of automatic identification system (AIS) data in regard to vessel movement prediction," *The Journal of Navigation*, vol. 67, no. 5, pp. 791–809, 2015. [Online]. Available: <https://doi.org/10.1017/S0373463314000253>
 - [6] D. M. A. Søfartsstyrelsen, "AIS-data," <https://www.soefartsstyrelsen.dk/sikkerhed-til-soes/sejladsinformation/aix-data>, accessed: 2025-11-23.
 - [7] N. Chen, A. Yang, H. Wu, L. Chen, W. Xiong, and N. Jing, "SEMINT: an LLM-empowered long-term vessel trajectory prediction framework," *International Journal of Geographical Information Science*, pp. 1–35, 2025.
 - [8] H. Yu, T. Li, K. Torp, and C. S. Jensen, "A multi-modal knowledge-enhanced framework for vessel trajectory prediction," in *Proceedings of the 19th International Symposium on Spatial and Temporal Data*, 2025, pp. 44–54.
 - [9] K. A. Sørensen, P. Heiselberg, and H. Heiselberg, "Probabilistic maritime trajectory prediction in complex scenarios using deep learning," *Sensors*, vol. 22, no. 5, p. 2058, 2022.
 - [10] W. Wang, W. Xiong, X. Ouyang, and L. Chen, "TPTrans: Vessel trajectory prediction model based on transformer using AIS data," in *ISPRS International Journal of Geo-Information*, vol. 13, no. 11, 2024. [Online]. Available: <https://doi.org/10.3390/ijgi13110400>
 - [11] Y. Li, J. Wang, T. Li, and Z. Fu, "TrAISformer: Spatio-temporal ship trajectory prediction based on transformer," in *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*. IEEE, 2024, pp. 1099–1104.
 - [12] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115. [Online]. Available: <https://doi.org/10.1609/aaai.v35i12.17325>
 - [13] N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," *arXiv preprint arXiv:2001.04451*, 2020.
 - [14] X. Nie, X. Zhou, Z. Li, L. Wang, X. Lin, and T. Tong, "Log-trans: Providing efficient local-global fusion with transformer and cnn parallel network for biomedical image segmentation," in *2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys)*. IEEE, 2022, pp. 769–776.
 - [15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
 - [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
 - [17] M. T. et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 7537–7547. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/55053683268957697aa39fba6f231c68-Paper.pdf
 - [18] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *CoRR*, vol. abs/2106.13008, 2021. [Online]. Available: <https://arxiv.org/abs/2106.13008>
 - [19] I. Loshchilov, F. Hutter et al., "Fixing weight decay regularization in adam," *arXiv preprint arXiv:1711.05101*, vol. 5, no. 5, p. 5, 2017.
 - [20] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright et al., "SciPy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.