

Article

D3AT-LSTM: An Efficient Model for Spatiotemporal Temperature Prediction Based on Attention Mechanisms

Ting Tian ¹, Huijing Wu ¹, Xianhua Liu ² and Qiao Hu ^{3,*}¹ School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China; 15639932467@163.com (T.T.); wuhuijingzzu@163.com (H.W.)² National Supercomputing Center in Zhengzhou, Zhengzhou 450001, China; z10619@163.com³ School of Geo-Technology, Zhengzhou University, Zhengzhou 450001, China

* Correspondence: qiao@zzu.edu.cn

Abstract: Accurate temperature prediction is essential for economic production and human society's daily life. However, most current methods only focus on time-series temperature modeling and prediction, ignoring the complex interplay of meteorological variables in the spatial domain. In this paper, a novel temperature prediction model (D3AT-LSTM) is proposed by combining the three-dimensional convolutional neural network (3DCNN) and the attention-based gated cyclic network. Firstly, the historical meteorological series of eight surrounding pixels are combined to construct a multi-dimensional feature tensor that integrates variables from the temporal domain as the input data. Convolutional units are used to model and analyze the spatiotemporal patterns of the local sequence in CNN modules by combining them with parallel attention mechanisms. The fully connected layer finally makes the final temperature prediction. This method is subsequently compared with both classical and state-of-art prediction models such as ARIMA (AR), long short-term memory network (LSTM), and Transformer using three indices: the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2). The results indicate that the D3AT-LSTM model can achieve good prediction accuracy compared to AR, LSTMs, and Transformer.

Keywords: temperature prediction; temporal and spatial characteristics; convolutional neural network; attention mechanism; LSTM



Citation: Tian, T.; Wu, H.; Liu, X.; Hu, Q. D3AT-LSTM: An Efficient Model for Spatiotemporal Temperature Prediction Based on Attention Mechanisms. *Electronics* **2024**, *13*, 4089. <https://doi.org/10.3390/electronics13204089>

Academic Editor: Domenico Rosaci

Received: 2 August 2024

Revised: 30 September 2024

Accepted: 15 October 2024

Published: 17 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Climate change is closely linked to the socioeconomic activities of human society, such as agriculture, transportation, and military operations. The increased frequency of natural disasters caused by extreme weather events has been observed worldwide in recent years. Thus, it is essential to establish an effective meteorological prediction scheme to avoid or mitigate the socioeconomic damage of extreme weather on human society.

Time-series meteorological data consists of various weather measurements arranged chronologically. This massive dataset often contains complex patterns and hidden information that can be used to infer future weather conditions.

There are two main approaches for temperature prediction: physical models [1] and data-driven models [2]. Physical models rely on the fundamental laws of atmospheric science to simulate atmospheric fluctuations by considering factors like atmospheric dynamics, thermodynamics, and radiative transfer. However, these models often require significant computational resources and time to run simulations. In contrast, data-driven methods use historical data to learn the relationship between past observations and future outcomes. These methods can capture internal temporal correlations within the data for a specific period and use this information for predictions. Data-driven approaches rely on large amounts of historical data and employ statistical models and machine learning techniques to establish the connection between input (past observations) and output (future

predictions). However, this method can only be analyzed from the perspective of time, ignoring the complex role of meteorological variables.

With the increasing availability of observation data, the development of artificial intelligence (AI) approaches in machine learning is enriching and promoting the application of data-driven methods in meteorological prediction. Recurrent neural networks (RNN) [3] are a powerful framework for spatiotemporal learning, making them well-suited for modeling long-sequence correlations. Consequently, they have become a standard tool for meteorological series prediction. Long short-term memory (LSTM) networks [4], a special type of RNN, are particularly popular due to their ability to address the gradient vanishing/exploding problem and the limited dependency modeling capability of traditional RNNs. This is achieved by introducing a gating control mechanism. Y. Liu et al. [5] successfully used LSTM model to predict exhaust gas temperature, achieving higher accuracy. Similarly, Liu X. et al. [6] used LSTM to predict future temperature by feeding it multiple long-time series meteorological elements (e.g., temperature) and achieved higher accuracy compared to the RNN model.

Additionally, elements in the time series domain often interact with the spatial domain, and this interaction should be considered when designing prediction models. CNNs (convolutional neural networks) use convolution operations to extract local spatial features and scan the entire image to obtain global contexts. Researchers have begun to combine CNNs and LSTMs to create hybrid models that can handle time series applications such as weather prediction. Ham et al. [7] used a CNN network to predict the El Niño phenomenon 1.5 years in advance. Their experiments demonstrate that CNNs are significantly better at capturing global correlations compared to the most advanced dynamic prediction systems currently available. Shi et al. [8] proposed a ConvLSTM network for radar echo extrapolation. ConvLSTM replaces the multiplication operation in the LSTM unit with convolution operations, allowing it to extract the high-dimensional spatial-temporal relationship of meteorological data and improve prediction accuracy. Kim et al. [9] proposed a method that combines a denoising autoencoder with ConvLSTM. The denoising autoencoder is used to extract important features from marine weather data and reduce noise. ConvLSTM is used in this method to capture the spatiotemporal dependence of the data. Their results show that this method performs well in predicting key meteorological parameters such as ocean surface temperature and wind speed. In conclusion, models that combine convolutional neural networks and LSTMs can effectively extract deeper information from the data's spatiotemporal domain, leading to significantly improved prediction accuracy.

Furthermore, 3DCNNs learn information differences in spatial and temporal domains simultaneously, which can mine deeper spatiotemporal dependency compared to traditional 2DCNNs. Researchers have started to improve 3DCNN by integrating LSTM networks for spatiotemporal temperature prediction. Wang et al. [10] introduced E3D-LSTM to effectively capture the spatiotemporal dynamics in video data, by using 3DCNN to capture spatial information between video frames and using LSTM to extract and memorize important temporal information in video sequences. This model can not only effectively predict the next video frame but also capture the detailed video changes, thereby producing high-quality video prediction.

Based on the analyses above, we can conclude that deep learning algorithms have achieved good performance in temperature prediction tasks due to their unique structures in modeling spatial and temporal dependency. Building upon traditional CNNs, 3DCNNs can model the local spatiotemporal range, thereby perceiving the deep evolution patterns of multiple continuous local spatiotemporal information. LSTMs are particularly effective at capturing long-term changes in time series data like temperature, especially when provided with appropriate meteorological variables. Combining CNNs and LSTMs in a single, hybrid model can fully explore the evolution laws of meteorological features in the spatial-temporal domains. However, this hybrid model cannot reflect the attention to the key time steps, so it cannot make a more comprehensive decision. Therefore, by introducing the attention mechanism, the paper proposes improvements to 3DCNN

and LSTM architectures to comprehensively model the spatiotemporal factors that affect the temperature sequence evolution, ultimately aiming to improve the accuracy of temperature prediction.

Numerous temperature prediction methods have been proposed, but many suffer from two primary limitations: their focus on a single time dimension and their neglect of internal interactions between meteorological sequences. Additionally, these methods often lack prediction schemes grounded in the internal evolutionary law of temperature. To address these shortcomings, this thesis introduces a multidimensional super-vegetal structure that incorporates time, space, and multiple meteorological variables as input features. This structure serves as the foundation for a composite model combining 3DCNN and LSTM networks. This model is designed to delve into the temporal and spatial relationships within the input features, enabling accurate short-term single-day temperature predictions.

2. Related Theories and Methods

This study focuses on obtaining real-time traffic volume on each road section. We achieve this by recording the number of vehicles entering and exiting specific road segments at key road nodes like intersections with traffic lights and entrances to residential or work areas. By establishing vehicle braking distance models, we can analyze the collected data to provide accurate information on traffic flow. This information serves as a foundation for developing effective recommendations for traffic dispersion and management.

2.1. The Definition of the Temperature Prediction Problem

Temperature prediction can be formulated as forecasting future temperature conditions based on historical time-series data as input. Specifically, meteorological characteristics at a given time t are represented by a vector F^t , and a historical observation sequence spanning t_i time steps into the future can be denoted as $X = \{F^t, F^{t+1}, \dots, F^{t+t_i}\}$. Given X as input, the goal is to predict the future temperature $Y = \{F^{t+t_i+1}, T^{t+t_i+2}, \dots, T^{t+t_i+t_o}\}$ of a specific grid point. Here, t_i and t_o represent the lengths of the input and output sequences, respectively. Since temperature T is a single variable, the output data is a one-dimensional temperature sequence value.

In our approach, the input feature X is a four-dimensional tensor encompassing both spatial and temporal information. The paper aims to enhance the accuracy of the output sequence Y by developing a deep learning model [11].

$$Y = NN(X; Weight) \quad (1)$$

where NN represents a neural network, which predicts the temperature series value Y for the future t_o time steps given the historical input data X , and $Weight$ is the parameter vector of the neural network.

2.2. Overview of the Spatiotemporal Temperature Prediction Models

2.2.1. 3DCNN

CNNs are one of the most prominent algorithms in deep learning. Hubel et al. [12] pioneered the concept of convolutional operations in 1959 based on their observations of how neurons in a cat's visual cortex respond to various stimuli. The core principle behind CNNs lies in progressively extracting high-level features from input data through a stacked architecture of convolutional, pooling, and fully connected layers. Both 2DCNNs and 3DCNNs represent extensions of CNNs designed for spatiotemporal prediction tasks.

Traditional 2D convolutions are primarily suited for extracting spatial features from two-dimensional data, such as images. However, in many real-world scenarios with temporal context, like video frames, it's crucial to consider both spatial and temporal correlations within the data. Since traditional 2DCNNs cannot capture correlations in the time dimension, 3DCNNs were specifically designed to address this limitation. 3DCNNs consist of convolutional layers, activation layers, pooling layers, and output layers [13]. They employ a 3D convolution kernel that performs a sliding window operation across three

dimensions: width, height, and depth (representing time). This process allows 3DCNNs to capture the significant patterns embedded within the 3D data space. Consequently, 3DCNNs have found widespread applications in computer vision tasks such as image recognition and motion capture. Similarly, in the context of spatiotemporal prediction tasks, 3DCNNs can be leveraged to simultaneously extract features from both the spatial and temporal domains, ultimately leading to improved prediction accuracy.

Equation (2) describes the 3DCNN convolution operation:

$$X_{i,j,k} = f_{conv} \left(\sum_{n=0}^{s_1} \sum_{m=0}^{s_2} \sum_{o=0}^{s_3} w_{n,m,o} x_{i+n,j+m,k+o} + b_{m,n,o} \right) \quad (2)$$

where $x_{i+n,j+m,k+o}$ represents the input matrix at the time step $k + o$, with row $i + n$ and column $j + m$; f_{conv} represents the Relu activation function; $w_{n,m,o}$ represents the value of the convolution kernel at row n and column m at depth o ; $b_{m,n,o}$ is the bias; and s_1, s_2, s_3 represent the size of the convolution kernel in height, width, and depth, respectively.

2.2.2. Attention Mechanism

The attention mechanism [14] is a valuable tool in deep learning, allowing models to prioritize informative elements within a dataset while disregarding less relevant data during input processing. This approach draws inspiration from the human visual system's ability to identify critical information within its field of view, filtering out hidden information automatically.

Attention mechanisms can be categorized into self-attention, spatial attention, and channel attention based on their internal structure. Squeeze-and-excitation networks (SE-Nets) [15] are a prime example of channel attention. SE-Nets allow each channel to incorporate information from the entire feature map by performing global pooling along the channel dimension. This process generates channel weights that are then used to re-weight the feature map, emphasizing important feature channels. This technique is widely adopted [16–18] because it overcomes the limitation of convolutional layers, whose receptive field can only capture local information. The core operation of SE-Net involves evaluating channel importance through a two-step process: squeeze and excitation. The squeeze operation utilizes global average pooling on the multi-dimensional feature map for each time step, compressing it into a single real number. Equation (3) describes the process, taking the 2D matrix as an example:

$$c_z = \frac{1}{hw} \sum_{i=1}^w \sum_{j=1}^h m_z(i, j) \quad (3)$$

where m_z is the feature matrix of the channel z and c_z is the value after global average pooling.

The excitation operation comprises two stages: feature fusion and dimension recovery, typically implemented using two fully connected layers:

$$n = \text{sigmoid}(f_2(\text{Swish}(f_1 c_z))) \quad (4)$$

The input c_z is passed through the first full connected layer f_1 . The obtained result is activated by the Swish function, followed by a second fully connected layer that normalizes the output to a vector between 0 and 1. This process increases the dimension of output weights are equal to the number of input feature channels, enabling the network to dynamically learn channel-wise importance weights.

The scale operation performs element-wise multiplication between the normalized weight vector n and the input feature map m_z :

$$q = m_z * n \quad (5)$$

where q represents the output feature map after incorporating channel-wise importance and n represents the normalized weight value, representing channel importance. Specifically,

the higher the importance of m_z , the closer the value of n is to 1 and vice versa; the lower the importance, the closer the value n is to 0.

SE-Nets are frequently integrated with various network architectures, such as CNNs and Transformers, to improve the model's expression ability. By dynamically adjusting channel weights, SE-Nets can emphasize informative feature channels while suppressing less relevant ones.

2.2.3. LSTM Model

LSTMs [19] add the mechanism of memory unit and gate to the traditional RNNs. The LSTM controls the information flow through three gating units: the forgetting gate determines the extent to which the LSTM forgets the previous state information, while the input gate determines the extent to which new information will be added to the memory unit. The output gate generates updated output information according to the current input and previous status. The whole process can be expressed as follows:

$$f_t = \text{sigmoid}(w_f[h_{t-1}, x_t] + b_f) \quad (6)$$

$$i_t = \text{sigmoid}(w_i[h_{t-1}, x_t] + b_i) \quad (7)$$

$$c_t = f_t c_{t-1} + i_t (\tanh(w_c[h_{t-1}, x_t] + b_c)) \quad (8)$$

$$o_t = \text{sigmoid}(w_o[h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t \tanh c_t \quad (10)$$

where x_t represents the input at time t , and c_{t-1} represents the memory state at the previous time $t - 1$; h_{t-1} represents the hidden state at time $t - 1$ and c_t is the memory state at time t ; h_t represents the hidden state at time t ; f_t , i_t , and o_t represent the forget gate, input gate, and output gate, respectively, at the current time step t ; w_f , w_i , w_o and w_c represent weight matrices corresponding to each gate (forget, input, output, and cell, respectively); b_f , b_i , b_o and b_c represent bias for each gate and the cell state; and sigmoid and $\tanh$ represent the activation functions used in the gating mechanism.

3. D3AT-LSTM Network

This paper proposes a new temperature prediction model (D3AT-LSTM) by combining the 3DCNN and the gated cyclic network (AT) attention mechanism. Firstly, the historical meteorological sequence of eight adjacent grid points of the target point is combined to construct a multi-dimensional feature tensor as the input data, integrating time, space, and multi-variable aspects. Secondly, the CNN modules combined with parallel attention branches are constructed, in which the convolution unit analyzes the spatiotemporal information of the local sequence, and the attention branch evaluates the importance of the time channel. Then, LSTM is used to capture the long-term dependencies among different sequences. Finally, the temperature prediction results are produced by the fully connected layer.

The steps of constructing the D3AT-LSTM model are as follows:

1. We use the supervoxel reconstruction to create meteorological features X as input.
2. The meteorological data X is fed to the 3DCNN module combined with the parallel attention branch, in which the 3DCNN module handles the feature abstraction in short terms. The standard 3DCNN model is challenging to handle issues that the characteristics of abnormal fluctuations and the time characteristics adjacent to the prediction stage are far more critical than that of the stationary stage in a continuous time series domain.
3. We introduce the channel attention mechanism to deal with the feature importance within the time series, in which all input features automatically learn the feature weight according to the loss and multiply it element by element with the output of the CNN module. By doing so, the model can focus on a specific time channel and

extract high-dimension spatiotemporal features. This structure reduces information redundancy and transforms the extracted fine-grained data into feature vectors with multiple time steps.

4. The data of each time step is input into the LSTM, which captures the coarse-grained dependencies of each time node in the long time series. Finally, the high-dimensional vector integrating the characteristics of different dimensions was passed into the full connection layer to produce the predicted temperature.

Figure 1 shows the structure of the D3AT-LSTM model, which is mainly composed of input layers, 3DCNN layers combined with a parallel attention mechanism, and an LSTM cycle neural unit. The framework of D3AT-LSTM is listed as Table 1.

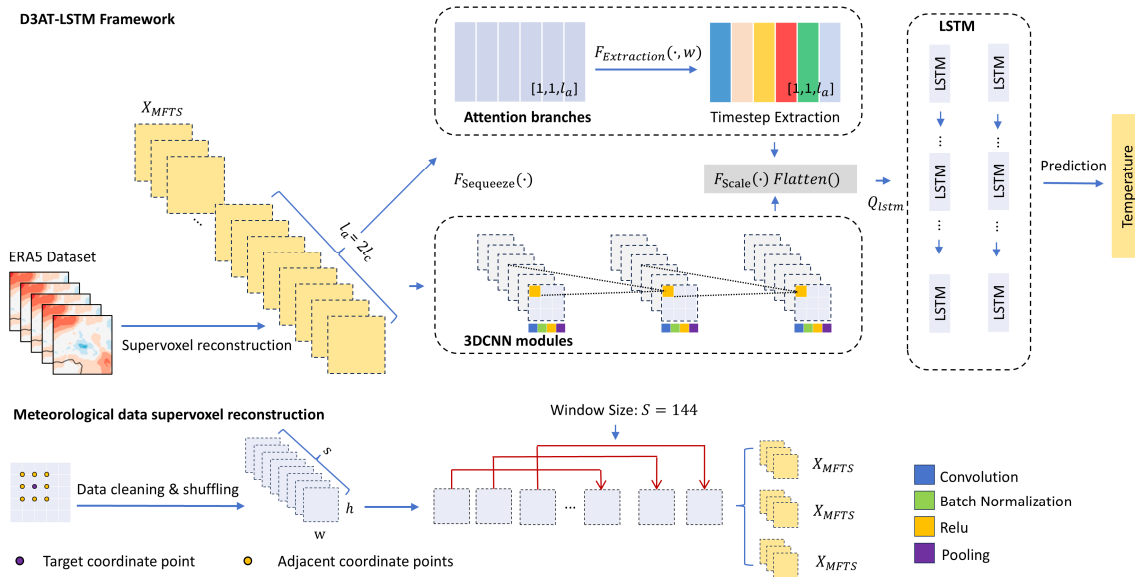


Figure 1. D3AT-LSTM network.

Table 1. D3AT-LSTM framework.

3DCNN Input Size: $S \times W \times H$					Attention Branch Input Size: $2S \times W \times H^*$		
Layers	Kernel Size	Stride	Input Channel	Output Channel	Operations	Parameters	Output Size
Conv1	$3 \times 3 \times 3$	1	1	32	Repeat	64	$2S \times W \times H \times 64$
Pool1	$2 \times 2 \times 2$	2	32	32	Average Pooling	Global	$2S \times 1 \times 1 \times 64$
Conv2	$3 \times 3 \times 3$	1	32	64	Squeeze	0.25	$\frac{S}{2} \times 1 \times 1 \times 64$
Pool2	$1 \times 2 \times 2$	$1 \times 2 \times 2$	64	64	Excitation	4	$2S \times 1 \times 1 \times 64$
Output_Size1 Q_1 : $\frac{S}{2} \times \frac{W}{4} \times \frac{H}{4} \times 64$					Cut	0.25	$\frac{S}{2} \times 1 \times 1 \times 64$
					Output_Size2 Q_2 : $\frac{S}{2} \times 1 \times 1 \times 64$		
$Q_{lstm} = Q_1 \times Q_2$							
Flatten				$Q_{lstm} \rightarrow 1D \ Q_{lstm}$			
LSTM Module				$Output = LSTM(Q_{lstm})$			
FC				$temperature = FC(Output)$			

* $2S$ is equal to l_a in Figure 1. S is equal to l_c in Table 1.

3.1. Local Spatiotemporal Information Extraction Based on 3DCNN

CNNs are good at extracting high-level features from data through their use of convolutional operations. This capability has led to their remarkable performance in image processing tasks. In this work, we leverage supervoxel reconstruction to create meteorological features X_{MFTS} with a consistent pixel shape $S \times H \times W$ across all time slices T_{Tot} , where

S is represented as time steps. This design captures both spatial and temporal correlations within the data, which is crucial for accurate temperature prediction. Traditional 2DCNNs are limited in their ability to analyze local time series dependencies within a specific time slice T_{part} ($T_{part} < T_{Tot}$) due to the fact that the depth of their convolution kernel is restricted to the number of channels in the input feature. In contrast, 3D convolutions can simultaneously capture spatial and temporal (depth) characteristics, overcoming this limitation. Therefore, we employ a 3DCNN architecture to perform spatiotemporal convolutions on the tensors of the 3D feature cube X_{MFTS} . This cube is constructed by stacking multiple consecutive feature matrices (Figure 2). This approach enables the network to capture the inherent differences within the spatial and temporal dimensions concurrently.

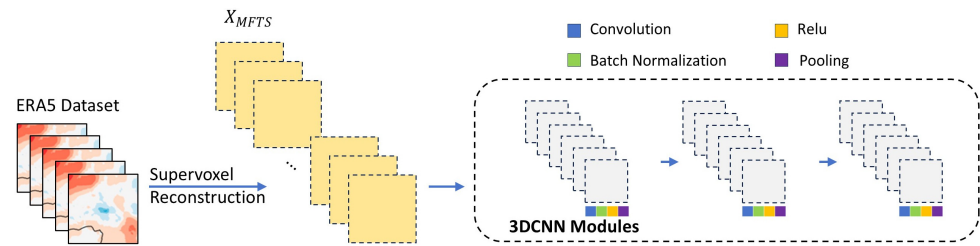


Figure 2. 3DCNN blocks.

This network progressively extracts spatiotemporal features from the input data, resulting in increasingly refined representations. As depicted in Figure 2, the convolutional subnet comprises two primary components. The first component utilizes a $3 \times 3 \times 3$ convolutional kernel with a stride of 1 and a $2 \times 2 \times 2$ pooling layer with a stride of 2 to reduce dimensionality and downsample the data. The second component employs a $3 \times 3 \times 3$ convolutional kernel with a stride of 1, followed by a $1 \times 2 \times 2$ pooling layer with a stride of $1 \times 2 \times 2$ to filter out redundant information and downsample along specific dimensions. Both components include a batch normalization layer and a ReLU activation function to accelerate convergence and mitigate gradient issues. In the 3DCNN architecture, we utilize $3 \times 3 \times 3$ kernels for the convolutional layers. The network includes two convolutional layers and two pooling layers. The hidden units in the two convolutional layers are set to 32 and 64, respectively. Each channel within the mechanism utilizes a single attention head. Table 1 offers a detailed illustration of these details.

3.2. Importance Evaluation of Combining Parallel Attention Branches

In time series temperature prediction, increasing the time step typically necessitates the inclusion of more historical data. Directly feeding this data into the network significantly increases prediction difficulty. However, in the continuous time series domain, information regarding abnormal fluctuations and adjacent moments holds greater importance compared to stationary periods. Standard 3DCNN models struggle to capture this characteristic effectively. We introduce a parallel channel attention branch based on the SE-Net within the 3DCNN module to address this limitation. This allows the network to assess the importance of different time steps. Furthermore, to ensure a comprehensive analysis of importance differences across time steps by the attention branch, we expand the sub-domain range of the attention module. Based on the input sequence length l_c of the convolution module, the feature domain is expanded by $\frac{1}{2}l_c$ units on both the left and right sides, resulting in a total expansion of $l_a = 2l_c$ (Figure 3). This expansion enables the attention module to make comprehensive decisions regarding feature importance for different time steps. SE-Net primarily evaluates the significance of the time domain through squeeze and excitation operations. For the input supervoxel information, the squeeze operation $F_{Squeeze}$ performs global average pooling across the spatial domain, compressing the features on each time slice into a single real number. The excitation operation $F_{Excitation}$ then utilizes a gating mechanism to adaptively learn the importance of each channel in the time domain. This process is applied in parallel with each convolution operation. The output weight sequence

of SE-Net is selected based on the same time range as the input sequence of the convolution. This weight sequence is then multiplied element-wise with the CNN output to obtain weighted high-order spatiotemporal information at each time slice. It allows for a more accurate perception of feature importance at the current stage compared to traditional methods that dynamically construct the weight matrix based on input-output data.

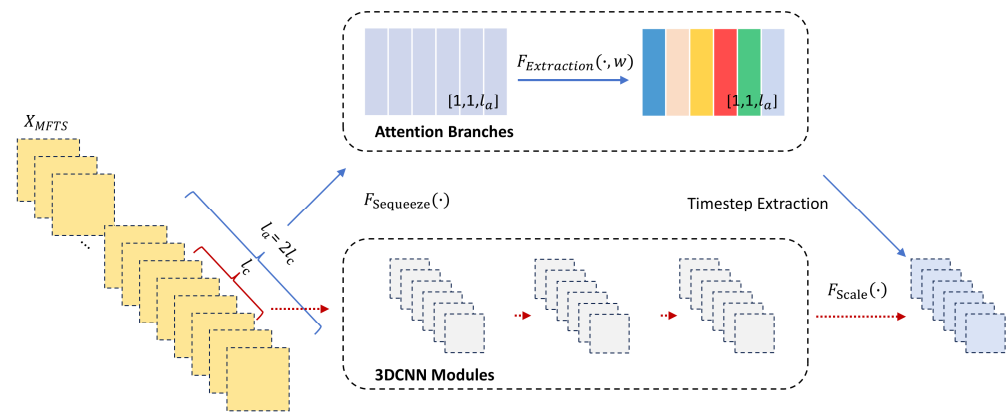


Figure 3. Convolution modules combined with parallel attention branch.

Each channel within the mechanism utilizes a single attention head. The weight matrices associated with this attention are defined as $\frac{S}{2} * 1 * 1$. Table 1 offers a detailed illustration of these details.

3.3. Long-Term Dependence Analysis with LSTM

CNNs equipped with attention mechanisms are good at extracting local spatiotemporal patterns from meteorological sequences. However, capturing the long-term global dependencies within these sequences is crucial for temperature prediction. While RNNs are adept at capturing the long-term dependencies of input features by modeling contextual information, traditional RNNs suffer from vanishing and exploding gradients. This makes it challenging for them to capture long-term dependencies in time-series meteorological data effectively. Long Short-Term Memory (LSTM) networks are employed to address this limitation. By introducing gating mechanisms, LSTMs prevent memory loss and gradient dispersion caused by longer step sizes. This enables them to effectively handle the dependencies present in long-sequence data. In our proposed approach, the fine-grained spatiotemporal features extracted by the CNN modules are fed forward to the LSTM unit. This step allows the LSTM to further extract the global, generalized relationships within the time series sequence. By fusing both coarse-grained and fine-grained features, the LSTM comprehensively describes the temporal relationships in the data.

In our approach, the high-dimensional features extracted from multiple time steps by the sub-CNNs are reshaped into one-dimensional vectors for each time step. These vectors are then fed as input to the corresponding time steps of the LSTM unit. The LSTM unit utilizes forget and input gates to manage its internal state. The forget gate determines the extent to which information from the previous sequence state is retained, while the input gate controls the degree to which new feature information is incorporated into the memory unit. By combining these gates, the LSTM continuously updates its hidden state. This allows the LSTM to dynamically adjust its internal structure to better capture the inherent structure of the meteorological time series. Consequently, the model can exploit long-term dependencies within the data, leading to improved prediction accuracy. The LSTM network incorporates two hidden layers, with each layer having a hidden size of 32 and 64, respectively.

4. Experiments and Results

4.1. Datasets

The ERA5 [20] global meteorological dataset, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), was employed in this study. Data from 1979 to the present were extracted at a spatial resolution of $0.25^\circ \times 0.25^\circ$ and a temporal resolution of 1 h. This paper focuses on the Yellow River Basin (defined by $95^\circ \text{ E} \sim 112^\circ \text{ E}$, $32^\circ \text{ N} \sim 42^\circ \text{ N}$) and utilizes the dataset from 2018 to May 2023 at the standard air pressure of 1000 hPa. Meteorological variables with strong correlation with temperature can well reflect the change of temperature [21]. In this paper, the size of Spearman correlation coefficient is used as the screening basis. Six meteorological elements [22–27] (evapotranspiration e_t , zonal wind speed sp_t , meridional wind speed v_t , atmospheric pressure sp_t , total precipitation tp , and relative humidity hr) commonly used in the field of temperature prediction and air temperature are used as input variables to observe the Spearman coefficient of air temperature variables. As shown in Figure 4, the correlation coefficients for meridional wind speed v_t and evapotranspiration e_t are 0.399 and 0.289, respectively, indicating a strong positive correlation with air temperature. Meanwhile, zonal wind speed sp_t and air pressure u_t also exhibit a certain influence on air temperature, with correlation coefficients of 0.132 and 0.116, respectively. The correlation coefficients of total precipitation tp and relative humidity hr are lower than 0.05, so these two variables were not considered. Therefore, meridional wind speed v_t , zonal wind speed sp_t , evapotranspiration e_t , and air pressure u_t were selected as the key variables for predicting air temperature T_t in this study.

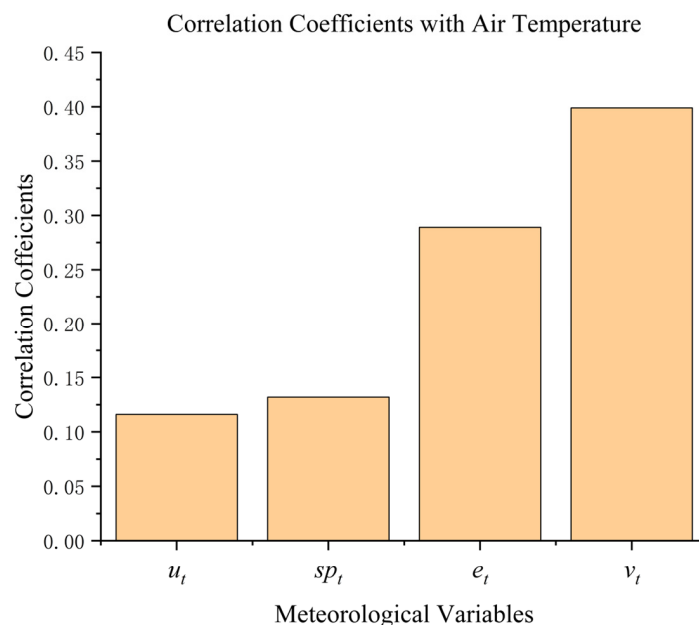


Figure 4. Correlation coefficient between meteorological characteristics and air temperature.

First, we clean up the obtained meteorological data and fill in the missing data with data from the last time step. Since the meteorological characteristics have different numerical regions, the network normalizes the feature of the meteorological dataset by using the linear Min-Max scaling function before feeding it into the network to accelerate the network convergence for obtaining the optimal solution.

The ERA5 dataset used in this paper is a network unit arranged according to a certain spatial resolution, and each unit has a corresponding unique meteorological value. In order to make the model fully consider the complex influence characteristics in time and space, the model incorporates its eight neighboring spatial points to predict the temperature at a specific point (Figure 5), effectively fusing the meteorological characteristics within the spatial-temporal domain as the subsequent input data. First, the meteorological elements

at a given time t for the station are represented as temperature T_t , zonal wind speed u_t , meridional wind speed v_t , air pressure sp_t , and evapotranspiration e_t . Through data reconstruction, a 2D matrix is formed in each time step. The size is $H * W$, where $W = 9$ represents the number of points and $H = 5$ represents the input features. To ensure compatibility with the 3DCNN data input format, the model first expands the channel dimension $C = 1$ in the second dimension. This transforms the original meteorological data into a multi-dimensional tensor $C * H * W$ for each time step. Subsequently, the network employs a sliding window with a size of $S = 144$ (the time series length of the input), dividing the data into multiple features with overlapping sequences. Finally, the input features are expressed as a four-dimensional tensor X_{MFTS} with the shape of $C * S * H * W$. X_{MFTS} is a supervoxel [18] structure that integrates time space multi meteorological variables. In this paper, it is used as the input of the follow-up network, so that the follow-up model can pay more attention to the internal and external complex spatiotemporal elements acting on the temperature. The process of supervoxel reconstruction of meteorological features is shown in Figure 5.

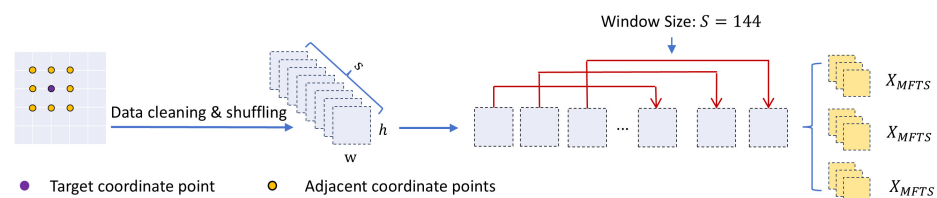


Figure 5. Supervoxel reconstruction process of meteorological characteristics.

4.2. Experiment Setup and Validation

4.2.1. Experiment Environment

Table 2 illustrates the experimental environment for this study.

Table 2. Experiment environment.

Name	Version
OS	Windows 10
Language	Python 3.7
GPU	NVIDIA GeForce GTX 1080 GPU
Framework	PyTorch 1.4
Graphic Card	3.2 G

4.2.2. Model Validation

To evaluate the effectiveness of the D3AT-LSTM model, we compared it with several state-of-the-art and classical temperature prediction methods, including ARIMA [28], LSTM-AE [29], Transformer [30], ConvLSTM [31], TCN-LSTM [32]. ARIMA serves as an example of a traditional machine-learning algorithm. LSTM-AE and Transformer represent purely deep learning approaches. ConvLSTM and TCN-LSTM are hybrid networks, allowing us to compare network performance across various architectures.

Before prediction, we preprocessed the input data to match the specific requirements of each model's architecture. ARIMA determines the model order and other parameter information through multiple iterations of historical data to fit the trend of temperature values. LSTM-AE integrates an Auto-Encoder (AE) with LSTM to extract features and reconstruct the input sequences. Transformer, equipped with a self-attention mechanism, allows the model to focus on important parts of the sequence. The input data for Transformer is a two-dimensional sequential structure defined by time and variables. ConvLSTM applies convolution operations in the spatial domain and LSTM operations in the temporal domain. The input for ConvLSTM is a two-dimensional temperature matrix arranged by time sequence. The TCN structure introduces dilated convolution and residual connections

to capture long-term dependencies within the 1D convolutional neural network. In temperature prediction tasks, the data dimension at each time step corresponds to the number of variables included at that specific time.

4.2.3. Parameter Setting and Evaluation Index

Meteorological data from 2018 to 2022 was used as the training set, while data from 2023 was employed as the test set. The model utilized a meteorological sequence from the past 144 h (six days) as input to predict the temperature for the next 24 h. During the training process, the model underwent 1000 iterations to arrive at the final network architecture. Feeding the input data through the first two layers (the convolutional layer and the pooling layer) resulted in channel numbers of 32 and 64, respectively. To expedite network convergence, a BatchNorm3D normalization layer and a Relu activation function were added after each convolution layer. The network was optimized using the Adam optimizer. After adjusting parameters, training parameters included 100 epochs and a batch size of 128. In order to learn the features more deeply, the hidden layer in LSTM is 2 and hidden size is set as 32 and 64 in both hidden layers. The initial learning rate was set to 0.01, with a decay mechanism that reduces the learning rate as the number of training rounds increases. Additionally, dropout regularization with a rate of 0.3 was employed. Finally, a fully connected layer with the output size is 24 to predict temperatures and then compared with the ground truth. The root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were used to evaluate the prediction performance of each model. Mean absolute error (MAE) represents the mean absolute error between the ground truth and the predicted value. The root mean square error (RMSE) is used to represent the square root after the square average of the difference between the ground truth value and the predicted value to measure the standard deviation of the prediction error. The coefficient of determination R^2 is used to measure the degree of fit of the model to the input data. Generally, the closer its value is to 1, the better the fitting effect of the model. In addition, we created index I to comprehensively evaluate the performance of the model. The MAE, RMSE, R^2 , and I calculation formula are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (12)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y} - \bar{y})^2}{\sum_{i=1}^n (y - \bar{y})^2} \quad (13)$$

$$I = \frac{R^2}{MAE \times RMSE} \quad (14)$$

where n represents the sample numbers (in the test set, $n = 24$); y_i represents the ground truth and $\hat{y}$ is the predicted value; and $\bar{y}$ represents the mean of ground truth.

4.3. Experiment Results

4.3.1. Accuracy Evaluation of Comparative Experiment

To increase the robustness of the experimental results and observe the prediction difference over different spaces, we selected five coordinates representing different spatial positions on behalf of the different locations in research area. Each point is combined with the adjacent eight points to roughly assess the model prediction over the whole study area, recorded as station1 (34.5° N, 103° E), station2 (34.5° N, 111° E), station3 (37° N, 107° E), station4 (39.5° N, 103° E), and station5 (39.5° N, 111° E), respectively (Figure 6). We obtained the hourly temperature of five stations by averaging the prediction of different methods, and acquire the 24-h prediction accuracy (Table 3 and Figure 7).

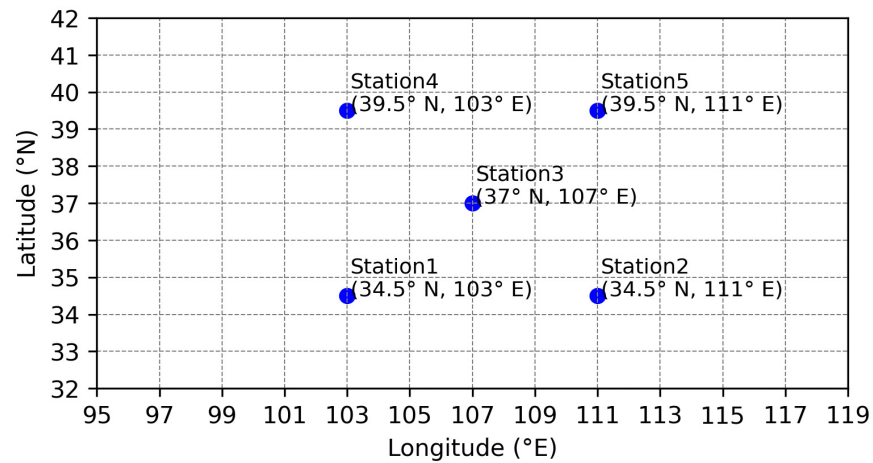


Figure 6. Search area of Station1–5.

Table 3. The 24-h average accuracy of the comparative experiment.

Models	Accuracy Index	Station					Avg.
		1	2	3	4	5	
AR	MAE	1.7127	1.7068	1.6980	1.7029	1.7103	1.7061
	RMSE	2.2438	2.2187	2.0452	2.0547	2.2389	2.1602
	R ²	0.9468	0.9475	0.9484	0.9477	0.9471	0.9475
	I	0.2463	0.2502	0.2730	0.2695	0.2473	0.2572
LSTM-AE	MAE	1.6327	1.6182	1.6225	1.6275	1.6323	1.6266
	RMSE	2.1207	2.0648	2.0780	2.0789	2.0961	2.0877
	R ²	0.9522	0.9530	0.9527	0.9524	0.9520	0.9524
	I	0.2750	0.2852	0.2825	0.2814	0.2782	0.2804
Transformer	MAE	1.5711	1.5661	1.5674	1.5692	1.5729	1.5693
	RMSE	2.1237	2.0804	2.0175	2.0207	2.1068	2.0698
	R ²	0.9583	0.9588	0.9593	0.9596	0.9585	0.9589
	I	0.2871	0.2942	0.3033	0.3026	0.2892	0.2952
TCN-LSTM	MAE	1.5361	1.5347	1.5294	1.5209	1.5320	1.5306
	RMSE	2.0637	2.0320	2.0163	1.9332	2.0964	2.0283
	R ²	0.9640	0.9650	0.9652	0.9656	0.9642	0.9648
	I	0.3040	0.3098	0.3129	0.3284	0.3002	0.3110
ConvLSTM	MAE	1.4841	1.4736	1.4513	1.4524	1.4913	1.4705
	RMSE	1.9837	1.9780	1.8361	1.9358	1.9841	1.9435
	R ²	0.9660	0.9669	0.9678	0.9673	0.9659	0.9667
	I	0.3281	0.3317	0.3631	0.3440	0.3264	0.3386
D3AT-LSTM	MAE	1.3543	1.3268	1.2685	1.3199	1.3263	1.3191 *
	RMSE	1.8126	1.7420	1.6838	1.7447	1.7420	1.7450
	R ²	0.9670	0.9697	0.9717	0.9698	0.9698	0.9696
	I	0.3939	0.4195	0.4549	0.4211	0.4197	0.4212
Avg.	MAE	1.5485	1.5377	1.5308	1.5321	1.5410	---
	RMSE	2.0513	2.0176	1.9512	1.9712	2.0440	---
	R ²	0.9590	0.9601	0.9608	0.9604	0.9596	---
	I	0.3019	0.3094	0.3216	0.3180	0.3046	---

* Bold indicates that our algorithm has the best performance compared with other algorithms.

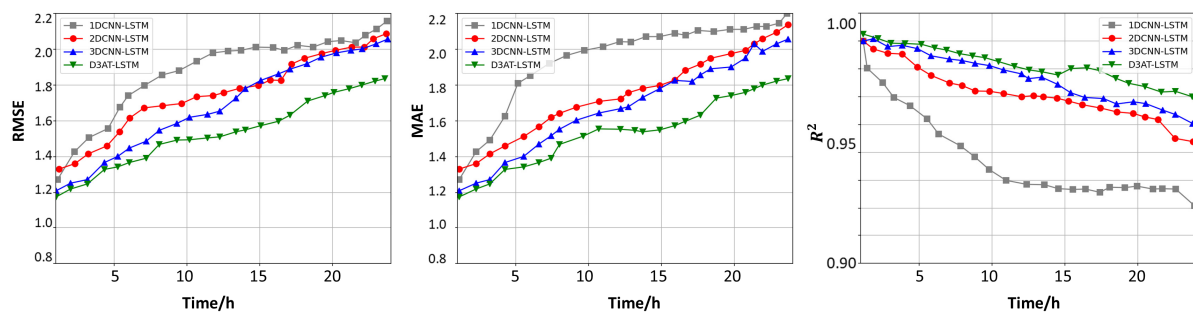


Figure 7. The accuracy of comparison experiment per hour.

The results demonstrate stable and generalizable predictions across different stations, which is likely due to the model's incorporation of the entire spatial domain (Figure 7 and Table 3). Station3, located in the center of the study area with moderate temperatures, exhibits strong model generalizability. This highlights the benefit of incorporating information from neighboring locations for improved prediction accuracy. Among the different methods, TCN-LSTM and ConvLSTM outperform single-model approaches like AR and LSTM-AE. The relatively strong performance of ConvLSTM suggests that this hybrid network effectively combines the strengths of CNNs and RNNs to capture both spatial and long-term temporal dependencies, leading to improved prediction accuracy. Furthermore, considering the entire time sequence, the ConvLSTM model surpasses other methods in terms of MAE, RMSE, and R^2 at the first and second time steps. This may be due to the small difference between local sequences in short-term prediction, ConvLSTM uses convolution operation in the LSTM structure to mine the overall spatiotemporal information and improve the prediction accuracy. Compared to the traditional AR, the D3AT-LSTM model achieves a reduction of approximately 0.4 in both MAE and RMSE, while R^2 is improved by more than 2% across the entire time series. Additionally, compared to LSTM-AE and Transformer sequence networks, the ConvLSTM model reduces MAE and RMSE by over 0.25 and 0.3, respectively, and increases R^2 by more than 1.3%. Moreover, compared to ConvLSTM and TCN-LSTM, the D3AT-LSTM model demonstrates significantly improved accuracy in the remaining time steps. When the time step increases, the error remains low and stable, indicating that the D3AT-LSTM method proposed in this study exhibits superior performance in 24-h short-term temperature prediction compared to other models.

To further illustrate the effectiveness of the D3AT-LSTM method, we present a case study focusing on 1 March 2023. A heatmap visualization is used to (Figure 8) compare the predicted 24-h average temperature distribution with the average actual temperatures observed on that day. The results reveal that our proposed method offers a more detailed representation compared to other approaches. This translates to greater sensitivity in capturing local temperature variations. This suggests that the D3AT-LSTM method effectively predicts future temperature distributions and provides robust long-term forecasts. However, a limitation of the method lies in its ability to predict extreme local values. This can likely be attributed to the inherent uncertainties associated with long-term temperature predictions.

In this paper, all experiments were completed using a 64-b Intel Core CPU i7-6900 k @ 3.20 GHz and NVIDIA GeForce GTX 1080 GPU. Table 4 shows the comparison of the training and testing time of the proposed D3AT-LSTM and the other five algorithms. It can be seen from the table that Transformer has the longest training time, which may be caused by its more complex model structure and higher computational complexity. As usual, neural networks are more time-consuming than traditional algorithms, because there are many parameters for network design optimization. Although the proposed D3AT-LSTM takes several hours, these training time can fully mine the temporal and spatial information and provide better feature representation for temperature prediction. In addition, the time in the test phase is short because there is no back propagation of the model in this phase.

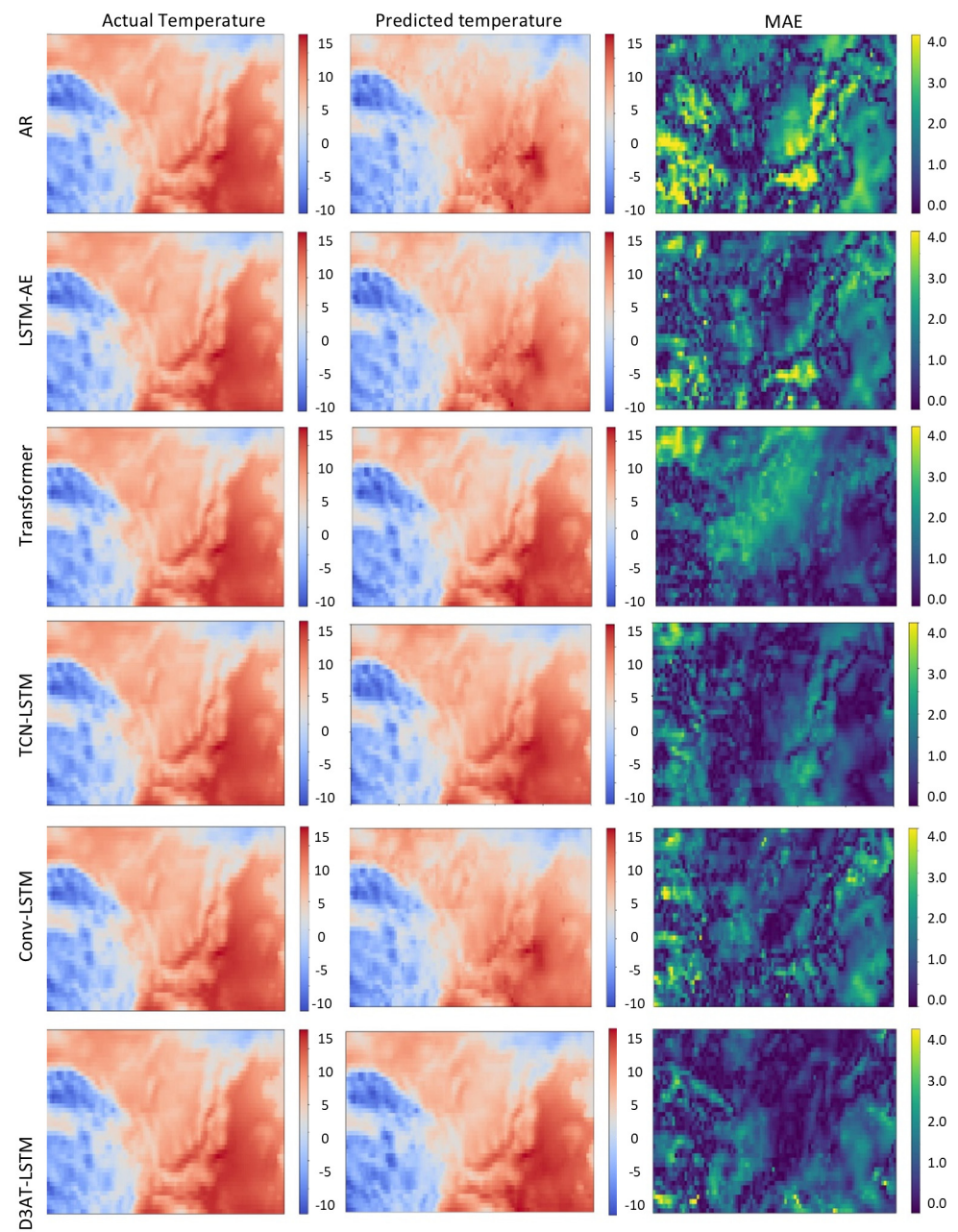


Figure 8. Thermal distribution of temperature predicted by each model.

Table 4. The running time comparison.

Models	Training (s)	Test (s)
AR	3540	1.3
LSTM-AE	6953	9.3
Transformer	16,861	24.2
TCN-LSTM	8930	10.9
ConvLSTM	10,908	13.5
D3AT-LSTM	11,244	16.8

4.3.2. Accuracy Evaluation of the Ablation Experiment

Ablation experiments were conducted to verify the effectiveness of each module in D3AT-LSTM by controlling specific variables in Table 5 and Figure 9. In the 1DCNN-LSTM model, only the target site's temperature variable at each time step was used as input. The 2DCNN-LSTM model incorporated the historical meteorological sequence of both the target site and the eight surrounding sites to assess the impact of supervoxel inclusion on prediction accuracy. The 3DCNN-LSTM model was used to evaluate the effectiveness of local convolution in the time dimension. Finally, the D3AT-LSTM model was used to verify the effectiveness of attention mechanism.

Table 5. Accuracy of 24-h mean temperature prediction in ablation experiment.

Models	Accuracy Index	Station1	Station2	Station3	Station4	Station5	Avg.
1DCNN-LSTM	MAE	1.6571	1.6493	1.6283	1.6432	1.6511	1.6458
	RMSE	2.2129	2.1356	2.0845	2.1037	2.1844	2.1442
	R^2	0.9525	0.9529	0.9538	0.9532	0.9527	0.9530
	I	0.2597	0.2705	0.2810	0.2757	0.2641	0.2702
2DCNN-LSTM	MAE	1.5519	1.5472	1.5449	1.5462	1.5495	1.5479
	RMSE	2.0988	2.1071	2.0961	2.1001	2.1121	2.1028
	R^2	0.9580	0.9588	0.9595	0.9590	0.9582	0.9587
	I	0.2941	0.2941	0.2963	0.2953	0.2928	0.2945
3DCNN-LSTM	MAE	1.4815	1.4783	1.4743	1.4757	1.4775	1.4774
	RMSE	2.0192	2.0157	1.9881	2.0116	2.0123	2.0093
	R^2	0.9654	0.9678	0.9695	0.9690	0.9670	0.9677
	I	0.3227	0.3248	0.3308	0.3264	0.3252	0.3260
D3AT-LSTM	MAE	1.3543	1.3268	1.2685	1.3199	1.3263	1.3191 *
	RMSE	1.8126	1.7420	1.6838	1.7447	1.7420	1.7450
	R^2	0.967	0.9697	0.9717	0.9698	0.9698	0.9696
	I	0.3939	0.4196	0.4549	0.4211	0.4198	0.4219
Avg.	MAE	1.5112	1.4937	1.4791	1.4962	1.5011	---
	RMSE	2.0358	2.0001	1.9631	1.9900	2.0127	---
	R^2	0.9607	0.9623	0.9636	0.9627	0.9619	---
	I	0.3123	0.3221	0.3319	0.3233	0.3184	---

* Bold indicates that our algorithm has the best performance compared with other algorithms.

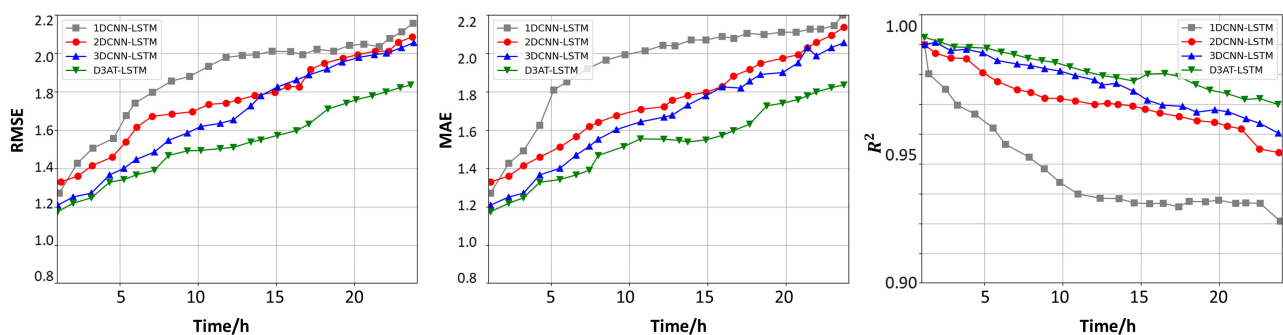


Figure 9. Accuracy curve of predicted accuracy at each time step of ablation experiment.

In the ablation experiment, the effectiveness of the D3AT-LSTM design idea and implementation method has been proven by controlling a single variable.

5. Summary

5.1. Conclusions

In this paper, we proposed a D3AT-LSTM method to predict daily temperature in the Yellow River Basin. This method utilized the ERA5 reanalysis meteorological dataset

as the benchmark. To predict the temperature at target sites, we combined the historical meteorological sequence of the surrounding area. This combined data, containing both spatial and temporal information, is reconstructed as input data. The D3AT-LSTM method then captures the spatiotemporal features of this input data. A 3DCNN extracts deeper spatiotemporal features between local sequences, while a parallel attention mechanism assesses the importance of the time dimension. The features at each time step are then fed into the LSTM to capture long-term dependencies within sequences, ultimately predicting the temperature for the next 24 h. The results demonstrate that this method achieves higher prediction accuracy than other methods. Finally, the proposed D3AT-LSTM framework proves effective for studying the effect of a single variable.

5.2. Further Perspective

In terms of temperature prediction in a short time range, the amount of data is relatively small, and D3AT-LSTM can greatly improve the accuracy of prediction. However, it should be noted that this paper discusses the problem of short-term temperature prediction. With the passage of time, the temperature presents a complex evolution law. Therefore, in the future research, it is necessary to pay more attention to the internal characteristics of the temperature, such as seasonal changes, so as to improve the robustness of the model in the long-term temperature prediction.

Author Contributions: Conceptualization: T.T.; Methodology, T.T.; Writing: T.T.; Review: H.W. and X.L.; Supervision: Q.H. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research on the key technologies of the construction and service of the Yellow River simulator for supercomputing, Henan major science and technology project (National Supercomputing Zhengzhou Center Innovation Ecosystem Construction Science and technology project) (201400210900).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, J.; Zhou, Y.S.; Shen, X.Y. Numerical Simulation Analysis of the Impact of Urbanization on an Extreme Precipitation Event over Beijing–Tianjin–Hebei, China. *Atmosphere* **2020**, *11*, 945. [\[CrossRef\]](#)
2. Cheng, A.R.; Lee, T.H.; Ku, H.; Chen, Y. Quality Control Program for Real-Time Hourly Temperature Observation in Taiwan. *J. Atmos. Ocean. Technol.* **2016**, *33*, 953–976. [\[CrossRef\]](#)
3. Ma, Z.; Zhang, H.; Liu, J. DB-RNN: An RNN for Precipitation Nowcasting Deblurring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5026–5041. [\[CrossRef\]](#)
4. Guo, R.; Li, H.; Huang, C. Operation stage division and RUL prediction of bearings based on 1DCNN-ON-LSTM. *Meas. Sci. Technol.* **2024**, *35*, 025035. [\[CrossRef\]](#)
5. Liu, Y.; Gan, H.; Cong, Y.; Hu, G. Research on fault prediction of marine diesel engine based on attention-LSTM. *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.* **2023**, *237*, 508–519. [\[CrossRef\]](#)
6. Liu, X.; Guo, H. Air quality indicators and AQI prediction coupling long-short term memory (LSTM) and sparrow search algorithm (SSA): A case study of Shanghai. *Atmos. Pollut. Res.* **2022**, *13*, 101551. [\[CrossRef\]](#)
7. Ham, Y.G.; Kim, J.H.; Luo, J.J. Deep learning for multi-year ENSO forecasts. *Nature* **2019**, *573*, 568–572. [\[CrossRef\]](#)
8. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9. [\[CrossRef\]](#)
9. Kim, K.-S.; Lee, J.-B.; Roh, M.-I.; Han, K.-M.; Lee, G.-H. Prediction of ocean weather based on denoising autoencoder and convolutional LSTM. *J. Mar. Sci. Eng.* **2020**, *8*, 805. [\[CrossRef\]](#)
10. Wang, Y.; Jiang, L.; Yang, M.H.; Li, L.J.; Long, M.; Li, F.F. Eidetic 3D LSTM: A model for video prediction and beyond. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 9 May 2019.
11. Jin, Q.; Tian, Y.; Sang, Q.; Liu, S.; Yu, J.; Wang, X. A deep learning model for joint prediction of three-dimensional ocean temperature, salinity and flow fields. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 573–577. [\[CrossRef\]](#)
12. Hubel, D.H.; Wiesel, T.N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **1959**, *148*, 574–583. [\[CrossRef\]](#)
13. Wu, K.; Shen, Y.; Wang, S. 3D convolutional neural network for regional precipitation nowcasting. *J. Image Signal Process.* **2018**, *7*, 200–212. [\[CrossRef\]](#)

14. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)
15. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [\[CrossRef\]](#)
16. Santavas, N.; Kansizoglou, I.; Bampis, L.; Karakasis, E.; Gasteratos, A. Attention! A Lightweight 2D Hand Pose Estimation Approach. *IEEE Sens. J.* **2021**, *21*, 11488–11496. [\[CrossRef\]](#)
17. Yang, H.; Yuan, C.; Zhang, L.; Sun, Y.; Hu, W.; Maybank, S.J. STA-CNN: Convolutional Spatial-Temporal Attention Learning for Action Recognition. *IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc.* **2020**, *29*, 5783–5793. [\[CrossRef\]](#)
18. Zhang, Y.; Yi, P.; Zhou, D.; Yang, X.; Yang, D.; Zhang, Q.; Wei, X. CSANet: Channel and Spatial Mixed Attention CNN for Pedestrian Detection. *IEEE Access* **2020**, *8*, 76243–76252. [\[CrossRef\]](#)
19. Cheng, J.; Dong, L.; Lapata, M. Long short-term memory-networks for machine reading. *arXiv* **2016**, arXiv:1601.06733. [\[CrossRef\]](#)
20. Harsbech, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [\[CrossRef\]](#)
21. Demeritt, D.; Nobert, S.; Cloke, H.; Pappenberger, F. Challenges in communicating and using ensembles in operational flood forecasting. *Meteorol. Appl.* **2010**, *17*, 209–222. [\[CrossRef\]](#)
22. Chen, P.-Y.; Pang, C.-C.; Sung, G.-Y. Developing an integrated model relating substrate water content to indoor temperature reduction for irrigation-decision support of a green roof. *Heliyon* **2024**, *10*, 33983. [\[CrossRef\]](#)
23. Lv, Y.; Guo, J.; Li, J.; Han, Y.; Xu, H.; Guo, X.; Cao, L.; Gao, W. Increased Turbulence in the Eurasian Upper-Level Jet Stream in Winter: Past and Future. *Earth Space Sci.* **2021**, *8*, e2020EA001556. [\[CrossRef\]](#)
24. Zhang, L.; Liu, D.; Yan, X.; Zhou, D.; Zheng, H.; Su, L. Analyzing and forecasting climate change in Harbin City, Northeast China. *Chin. Geogr. Sci.* **2011**, *21*, 65–73. [\[CrossRef\]](#)
25. Perevedentsev, Y.P.; Vasil'ev, A.A.; Shantalinskii, K.M.; Gur'yanov, V.V. Long-term variations in surface air pressure and surface air temperature in the Northern Hemisphere mid-latitudes. *Russ. Meteorol. Hydrol.* **2017**, *42*, 461–470. [\[CrossRef\]](#)
26. Pan, F.; Nagaoka, L.; Wolverson, S.; Atkinson, S.F.; Kohler, T.A.; O'Neill, M. A Constrained Stochastic Weather Generator for Daily Mean Air Temperature and Precipitation. *Atmosphere* **2021**, *12*, 135. [\[CrossRef\]](#)
27. Hao, J.; Lu, E. Variation of Relative Humidity as Seen through Linking Water Vapor to Air Temperature: An Assessment of Interannual Variations in the Near-Surface Atmosphere. *Atmosphere* **2022**, *13*, 1171. [\[CrossRef\]](#)
28. Polasek, W. Time Series Analysis and Its Applications: With R Examples, Third Edition by Robert H. Shumway, David S. Stoffer. *Int. Stat. Rev.* **2013**, *81*, 323–325. [\[CrossRef\]](#)
29. Luo, X.J.; Oyedele, L.O. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. *Adv. Eng. Inform.* **2021**, *50*, 101357. [\[CrossRef\]](#)
30. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021. [\[CrossRef\]](#)
31. Huang, F.; Zhang, Y.; Zhang, Y.; Shangguan, W.; Li, Q.; Li, L.; Jiang, S. Interpreting Conv-LSTM for Spatio-Temporal Soil Moisture Prediction in China. *Agriculture* **2023**, *13*, 971. [\[CrossRef\]](#)
32. Wang, X.; Liu, Y.; Liang, X.; Zhang, C.; Yang, C.; Gui, W. Learning an Enhanced TCN-LSTM Network for Temperature Process Modeling in Rotary Kilns. *IEEE Trans. Autom. Sci. Eng.* **2024**, 1–12. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.