EasyChair Preprint

№ 10785

STA-Net: Reconstruct Missing Temperature Data of Meteorological Stations Using a Spatiotemporal Attention Neural Network

Tianrui Hou, Li Wu, Xinzhong Zhang, Xiaoying Wang and Jianqiang Huang

August 27, 2023

STA-Net: Reconstruct Missing Temperature Data of Meteorological Stations Using a Spatiotemporal Attention Neural Network

Tianrui Hou[1][0009−0000−6474−957X], Li Wu[2][0000−0001−5242−3169], Xinzhong Zhang[3], Xiaoying Wang[4], and Jianqiang Huang[5]

[1] Department of Computer Technology and Applications, Qinghai University, No.251, Ningda Road, Xining, 810016, Qinghai, China
15651653492@163.com
[2] Department of Computer Technology and Applications, Qinghai University, No.251, Ningda Road, Xining, 810016, Qinghai, China
wuli777@qhu.edu.cn
[3] Beijing PRESKY Co.,Ltd., Research and development department, China
mail.zxz@163.com
[4] Department of Computer Technology and Applications, Qinghai University, No.251, Ningda Road, Xining, 810016, Qinghai, China
wxy_cta@qhu.edu.cn
[5] Department of Computer Technology and Applications, Qinghai University, No.251, Ningda Road, Xining, 810016, Qinghai, China
hjq@qhu.edu.cn
Corresponding author(s). E-mail(s): wuli777@qhu.edu.cn

Abstract. Reconstructing the missing meteorological site temperature data is of great significance for analyzing climate change and predicting related natural disasters, but is a trickily and urgently solved problem. In the past, various interpolation methods were used to solve this problem, but these methods basically ignored the temporal correlation of the site itself. Recently, the methods based on machine learning have been widely studied to solve this problem. However, these methods tend to handle the missing value situation of single site, neglecting spatial correlation between sites. Hence, we put forward a new spatiotemporal attention neural network (STA-Net) for reconstructing missing data in multiple meteorological sites. The STA-Net utilizes the currently state-of-the-art encoder-decoder deep learning architecture and is composed of two subnetworks which include local spatial attention mechanism (LSAM) and multidimensional temporal self-attention mechanism (MTSAM), respectively. Moreover, a multiple-meteorological-site data processing method is developed to generate matrix datasets containing spatiotemporal information so the STA-Net can be trained and tested. To evaluate the STA-Net, a large number of experiments on real Tibet and Qamdo datasets with the missing rates of 25%, 50% and 75%, respectively, are conducted, meanwhile compared with U-Net, PConvU-Net and BiLSTM. Experimental results have showed that our data processing method is effective and meantime and our STA-Net achieves greater reconstruction effect. In the case with the missing rate of 25% on Tibet test datasets and com-

pared to the other three methods, the MAE declines by 60.21%, 36.42% and 12.70%; the RMSE declines by 56.28%, 32.03% and 14.17%; the R2 increases by 0.75%, 0.20% and 0.07%.

Keywords: Attention mechanism · Deep learning · Neural networks · Missing data imputation · Meteorological station data · Time series.

1 Introduction

In meteorological field, temperature data are mostly obtained from corresponding meteorological stations. Complete temperature data are an important basis for meteorologists to conduct the relevant weather forecast and climate analysis; meanwhile, it also plays a critical role in agriculture and ecological research, and is also an important data source for relevant scientific research [1–4]. But the data obtained by meteorological sites are not always complete due to various issues such as electromagnetic interference, equipment failure, harsh environmental conditions or manual operation errors, etc [2], [5], [6]. Thus, reconstructing the missing temperature data is an essential previous work when ones conduct relevant scientific studies, and is also a trickily and urgently solved problem.

In order to impute missing meteorological site temperature data, related researchers have undertaken a lot of work, and proposed various kinds of methods. Generally, these methods can be divided into two categories: traditional space-based interpolation method and data-driven based machine learning method. The methods based on traditional interpolation often use the mathematical or physical properties in space to restore missing data [7]. For example, researchers have already reconstructed the missing meteorological station temperature data using interpolation based methods such as inverse-distance weighting (IDW) [8], kriging [9, 10], thin-plate splines [11] and multiple regressions [12]. However, these methods are inevitably affected by the spatial distance and the number of meteorological stations [2], [5]. Besides, they seldom consider the specific geographical status of station position, being prone to causing that the theoretical models used by such methods may not match the complexity of the actual data. Such limitations would restrict the performance of the methods based on interpolation.

On the other hand, the data-driven methods represented by machine learning can dig out the complex relationship or potential distribution that the data given exist and thus reconstruct the missing meteorological station data. These methods can also be divided into two categories: traditional machine learning and the most currently popular deep learning. Examples based on traditional machine learning to reconstruct various missing meteorological station data have k-nearest neighbor [13], logistic regression [14] and autoregressive integrated moving average [6], [15]. Although these methods have a relatively stronger imputation ability in datasets with a low missing rate, they may show a poor performance on datasets with a relatively high missing rate [16].

Meantime, due to meteorological station data are typically time-series data, in recent years, people have used deep learning models based sequence-to-

sequence to impute various missing meteorological station data. For instance, Cao et al. [17] applied bidirectional recurrent neural network (BiRNN) to reconstruct multiple types of time series data, such as air quality data collected by stations, etc. Zhang et al. [18, 19] and Dagtekin et al. [20] used long short-term memory (LSTM) [21] or gated recurrent units (GRUs) to structure network architectures to recover missing water quality data collected by stations. Xie et al. [22] applied LSTM based model to impute missing temperature station data. These models flexibly designed the corresponding network architecture based on the data used and achieved better reconstruction effects. However, they only considered the problem that the missing value existed in long-term time series data collected by single station, neglecting spatial correlation between multiple local sits.

To address against the faced problem, in this paper, we firstly propose a weather station point data processing method. The datasets produced by this method is a format of the numerical matrix, which can be input to the convolutional neural network (CNN). That is, we solve a tricky problem that previously people often use traditional interpolation method but not CNN-based network to impute missing data in space. Secondly, we design a completely new spatiotemporal attention neural network (STA-Net) which can effectively handle spatiotemporal information through two attention mechanism: local spatial attention mechanism (LSAM) and multidimensional temporal self-attention mechanism (MTSAM). In addition, this paper also discusses reconstruction effects of the proposed model on Tibet and Qamdo datasets with different missing rate 25%, 50%, and 75%, which provides a new technical means for other similar missing data imputation work.

The main contributions of this work include:(1) We design a simple multiple-meteorological-station data processing method. The method can generate matrix datasets containing spatiotemporal information. These generated datasets can be input into the CNN. Hence, we now can dig out spatial correlation between multiple local stations by taking advantage of CNN's characteristics;(2) Propose a novel reconstruction neural network (STA-Net) with a new spatiotemporal attention based on encoder-decoder network architecture. Our model is able to recover missing data based on the information from both the spatial and temporal relation;(3) Design two new attention mechanism: local spatial attention mechanism (LSAM) and multidimensional temporal self-attention mechanism (MTSAM), respectively. The LSAM mainly handles spatial correlation between multiple local meteorological site data. The MTSAM mainly solves the time sequence relationship of single meteorological site data itself.

2 Related works

2.1 Missing data reconstruction

In many research areas, missing data reconstruction is an important and necessarily previous beforehand work when conducting relevant scientific research.

Especially in recent years, with the continuous significant progress in deep learning, people are more and more inclined to use deep learning knowledge to solve this problem. For example, in the field of image repair, people have applied based on CNN network structure for image repair [23–26]. Similarly, in geography and meteorology, related researchers have begun to use deep learning networks to repair a variety of missing data. For example, relevant researchers used CNN based network model to reconstruct the global average temperature [27], recover the land water storage data [28], repair the remote sensing sea surface temperature image [29], and reconstruct the missing seismic data [7], [30]. In addition, for the missing weather station data, in the past, people had used the interpolation method for recovery. However, at present, based on meteorological station data is a typical timing data, meteorologists often use the LSTM based network model to repair the missing various meteorological data [17–20],[22]. However, so far, no one has used CNN based networks to reconstruct missing meteorological station data.

2.2 Attention

At present, much research has been done on attention. Here, we have selected several representative models for a brief description. On spatial attention, Yu et al. [31] first proposed contextual attention (CA) model which is used in image repair tasks. However, the model is faced with a regular missing region data reconstruction problem and is therefore not suitable for irregular ones. Although Shin et al. [25] and Peng et al. [26] made some improvements to CA, having not addressed the problem of irregular missing region data repair. On temporal attention, Bahdanau et al. [32] first proposed the concept of attention, having been applied to related temporal tasks and natural language translation tasks [33]. Meanwhile, this attention explicitly models global dependencies, which has been successfully applied in multiple models [34]. These tasks motivate us to try using attention to improve network performance.

3 Methods and network architecture

3.1 A data processing method of meteorological stations

The original used meteorological datasets in this paper are provided by China Meteorological Administration (http://data.cma.cn/) which is publicly accessible. They are Tibet regional meteorological datasets and Qamdo regional meteorological datasets, respectively. Both two datasets hourly store the observed meteorological data in text format.

According to the actual situation, any meteorological station data have its own idiosyncratic time sequence. Hence, we can apply LSTM-based neural network to impute missing data. However, these network models hardly deal with the spatial connection of multiple meteorological stations. Although the interpolation methods can reconstruct missing values between multiple meteorological

stations in space, these methods are usually driven by a mathematical or physical model [7], which is prone to cause incompatible between data and model. In addition, these methods ignore the temporal relationship of the weather station itself.

To overcome the above plights, we design a simple data processing method. It generate a numerical matrix data set including spatiotemporal relationship. The proposed method first selects s meteorological stations as the research object, and then integrates the meteorological data observed within 24 hours in chronological order (from 0:00 to 23:00). Finally, we will get a matrix data set with dimensions of $(24, s)$, where s is determined by the specific data set used. In addition, the ranking of s meteorological sites is based on the order in which they appear in the original data set.

Obviously, the new matrix data set can be used as the input of CNN, which effectively alleviates the deficiencies of the interpolation methods, better coping with spatial correlation between multiple local sites. At the same time, since each column of data is still a time series data, we can still use the LSTM-based network models. More importantly, this data set allows us to design a new network architecture to reconstruct the missing value from the space and time dimensions.

3.2 Attention mechanism and network architecture

Local spatial attention mechanism According to the characteristics of CNN, convolution operation only simply aggregates local information [30], neglecting the spatial correlation or similarity between adjacent local areas. Hence, when we apply the neural network model only containing CNN to reconstruct the missing data, a large missing rate will bring a great impact on the reconstruction effect. In addition, our input data is two-dimensional numerical matrix mathematically and basically can be equivalent to single channel image data. Meanwhile, previous studies showed that the nearby pixels between adjacent local areas in images could keep consecutive and consistent in most cases [35–38]. Particularly, Yu et al. [31] proposed contextual attention (CA) to repair the images including missing pixels by utilizing the principle. Therefore, based on the foregoing discussed theories and methods, we design a local spatial attention mechanism, drawing and exploiting the correlation in neighboring local areas to resolve the problem encountered at the beginning of this paragraph.

In the original paper, the CA model only processed the situation that missing pixels or part was regular. And it did not utilize all attention scores, either [26]. In contrast, the proposed local spatial attention model makes use of all the corresponding masks to judge whether the filled missing values are valid.

Fig. 1 shows how the local spatial attention mechanism works. The input feature map which concurrently can be regarded the as foreground feature map and the background feature map. The foreground feature map matches with the patches $b_{x',y'}$ extracted from the background feature map by the way of sliding window $b_{x',y'}$ with the step acquiescently set to 1, which is similar to the convolution operation. That is, the patches $b_{x',y'}$ can be treated as convolutional

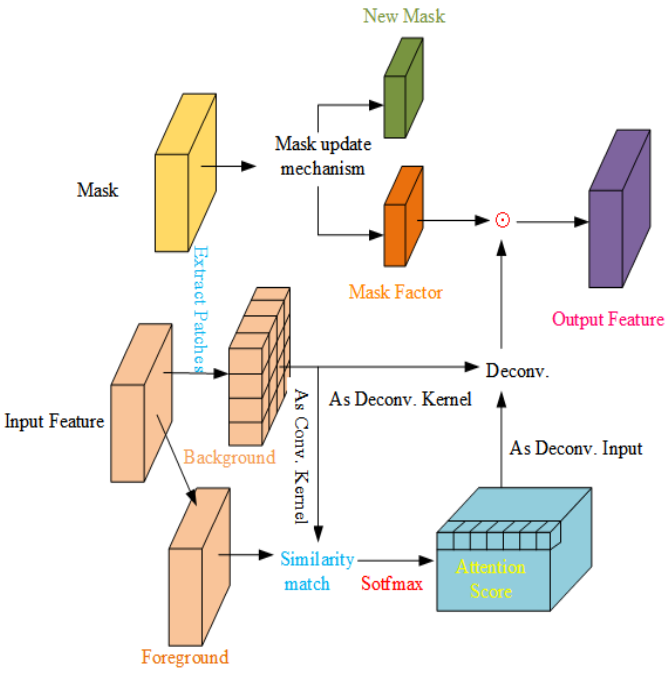

Fig. 1. Illustration of the local spatial attention mechanism. Firstly after the patches extracted from background feature and foreground feature go through the similarity matching calculation and $softmax$, we will get the attention scores. Meanwhile we attain mask factor by utilizing mask update mechanism with corresponding mask. Next, the result that deconvoluting attention score outputs and the mask factor execute hadamard product (element-wise multiplication). Finally, we get the output feature map.

kernel whose size is set to $(3, 3)$ in this paper, where $\left(x^{'}, y^{'}\right)$ denotes the central coordinate of the patches in the background feature map. Let $f_{x,y}$ denotes the sub feature map corresponding to the current sliding (convolution) window. Then, the attention score can be computed as:

$$s_{(x,y),(x',y')} = \langle \frac{f_{x,y}}{\|f_{x,y}\|}, \frac{b_{x',y'}}{\left\|b_{x',y'}\right\|} \rangle \tag{1}$$

$$a_{(x,y),(x',y')} = softmax(\alpha s_{(x,y),(x',y')}) \tag{2}$$

wherein $s_{(x,y),(x',y')}$ denotes cosine similarity between $f_{x,y}$ and $b_{x',y'}$, $a_{(x,y),(x',y')}$ denotes the attention scores for $f_{x,y}$ corresponding to $b_{x',y'}$, and α is a hyper-parameter that avoids smaller values in performing $softmax$ operations and is set to 10.

By processing mask M corresponding to input feature map using the mask update mechanism, we get the new mask and mask factor f_m whose element values can only be 1 or 0 which indicates whether the reconstructed missing values are valid. The f_m is computed as:

$$f_m = G(F(m, w^1)) \tag{3}$$

wherein w^1 is a matrix with elements of all 1 whose size is same as $b_{x',y'}$, and F and G denote convolutional function and cropping function that limits the value to $[0, 1]$, respectively.

Eventually, through the result output by the deconvolution operation $Deconv$ [39, 40] whose input is $a_{(x,y),(x',y')}$ and $b_{x',y'}$, and combines with f_m, we can get final outcome o, which is computed as:

$$o = f_m \odot Deconv(a_{(x,y),(x',y')}, b_{x',y'}) \tag{4}$$

wherein $\odot$ denotes element-wise multiplication.

Multidimensional temporal self-attention mechanism According to the data processing method proposed, per column in input data has its own temporal relationship which can be acquired by making use of LSTM based network models. In this paper, We used bidirectional LSTM (BiLSTM), an extended version of BiRNN [17], to cope with the temporal relationship.

However, existing BiLSTM simply concatenates forward hidden $\overrightarrow{y_i}$ and backward hidden state $\overleftarrow{y_i}$, not taking into account that the elements corresponding to sub dimensions between $\overrightarrow{y_i}$ and $\overleftarrow{y_i}$ have different importance to the s_i, where $\overrightarrow{y_i}$ and $\overleftarrow{y_i}$ denote high-dimensional vectors, as shown in Fig. 2. In order to weigh the significance of different sub elements values, we propose a multidimensional temporal self-attention mechanism (MTSAM).

The MTSAM proposed works as Fig. 3 shows. For the input time series data $X = (x_1, x_2, \cdots, x_i, \cdots, x_{n-1}, x_n)$, when they are input into the BiLSTM, we will obtain vector sequences of two states: forward hidden

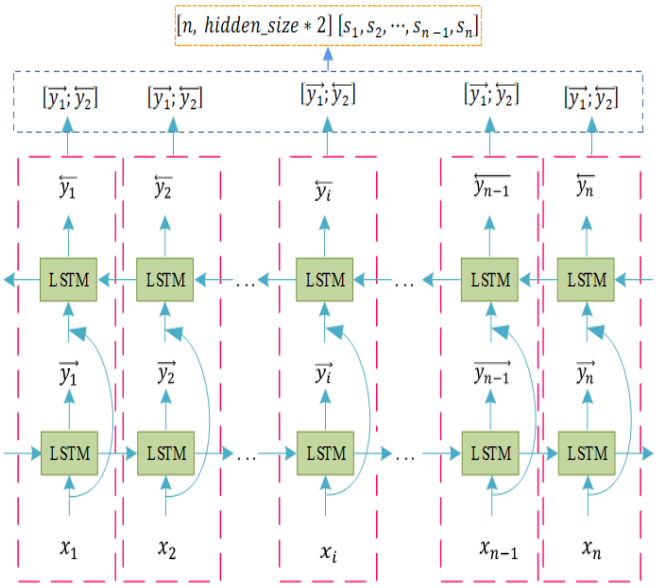

Fig. 2. Illustration of the existing BiLSTM model. The after input sequence data $X = (x_1, x_2, \cdots, x_i, \cdots, x_{n-1}, x_n)$ go through BiLSTM, we get a new sequence $S = (s_1, s_2, \cdots, s_i, \cdots, s_{n-1}, s_n)$. For each s_i, it is attained by concatenating forward hidden states $\overrightarrow{y_i}$ and backward hidden states $\overleftarrow{y_i}$, namely which can be expressed as $[\overrightarrow{y_i}; \overleftarrow{y_i}]$.

states $\overrightarrow{y} = (\overrightarrow{y_1}, \overrightarrow{y_2}, \cdots, \overrightarrow{y_i}, \cdots, \overrightarrow{y_n - 1}, \overrightarrow{y_n})$ and backward hidden states $\overleftarrow{y} = (\overleftarrow{y_1}, \overleftarrow{y_2}, \cdots, \overleftarrow{y_i}, \cdots, \overleftarrow{y_n - 1}, \overleftarrow{y_n})$. Here, we stack the states corresponding subscript between $\overrightarrow{y}$ and $\overleftarrow{y}$, obtaining the output $S = (s_1, s_2, \cdots, s_i, \cdots, s_{n-1}, s_n)$ whose size is $[n, 2, hidden\_size]$ contrary to the output of existing BiLSTM.

To figure out the influence degree between the element values in each sub dimensions of the status s_i in S on itself, we first transpose S and get the result $S^T = (s_1^T, s_2^T, \cdots, s_i^T, \cdots, s_{n-1}^T, s_n^T)$ whose size is $[n, hidden\_size, 2]$, computed as follows:

$$S^T = T(S) \tag{5}$$

wherein T denotes transpose function. Next S^T enters a single layer perceptron [41] and $softmax$ operation layer, in turn. After this step, getting attention scores s_a, which is computed as follows:

$$s_w = L(S^T; \theta) \tag{6}$$

$$s_a = softmax(s_w) \tag{7}$$

wherein L is a linear neural network, θ denotes learnable parameters and s_w shows importance of the element values which are in each sub dimensions of each sub status s_i in S.

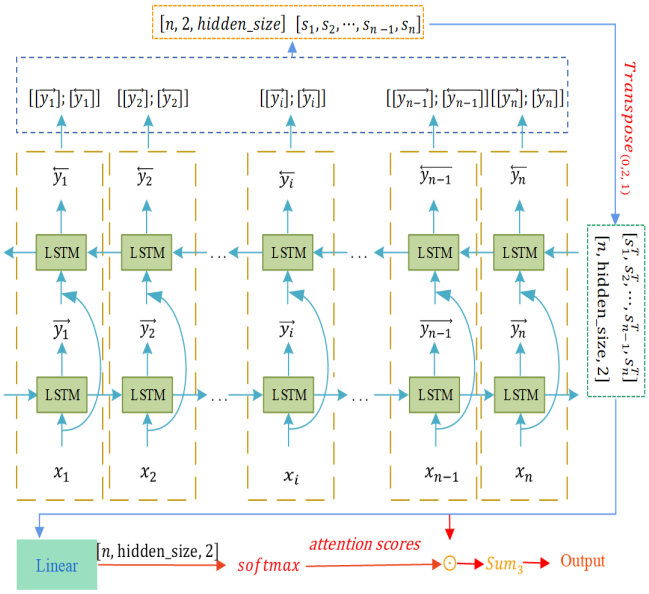

Fig. 3. Illustration of the multidimensional temporal self-attention mechanism. After input sequence data $X = (x_1, x_2, \cdots, x_i, \cdots, x_{n-1}, x_n)$ go through BiLSTM, the output sequence data $S = (s_1, s_2, \cdots, s_i, \cdots, s_{n-1}, s_n)$ is obtained by stacking forward hidden states $\overrightarrow{y_i}$ and backward hidden states $\overleftarrow{y_i}$ not concatenating them. Then the output S^T after transposing S executes the linear and $softmax$ operations to get self-attention scores s^*. Finally after summing the hadamard product of S^T and s^*, getting ultimate result.

Finally, we calculate the weighted sum S_w^T of the attention scores s_a and S^T. Now each dimension of S_w^T fuses the forward and backward information, which makes the amount of messages implied in S_w^T unchanged, but reduces data size by half compared with the previous BiLSTM model output. Hence, the cost of subsequent operations can be reduced. The S_w^T is computed as:

$$S_w^T = Sum(s_a \odot S^T) \tag{8}$$

wherein $\odot$ denotes element-wise multiplication and Sum represents summarizing function.

Variable factor analysis Intuitively, the temperature of a specific geographical environment is affected by a variety of factors. These factors usually include solar altitude (SA), latitude (Lat), longitude (Lon), cloud thickness (CT), cloud range (CR), wind speed (WS), and other related factors. Apparently, the influence focus of these factors on temperature varies in different regions and in different time periods. Here we express this relation by applying a function, which is defined as:

$$Tem = F_c(SA, Lat, Lon, CT, CR, WS, \cdots) \tag{9}$$

wherein Tem is multivariate function value, F_c denotes a multivariate function and SA et al. is variable.

When performing temporal regularity analysis of a certain meteorological site, the temperature data need to be decomposed and had better to be expressed by many factors, which is defined as:

$$F_m = NN_d(Tem; \theta) \tag{10}$$

wherein NN_d denotes a neural network layer with learnable parameter θ, and F_m is a multivariable representation symbol.

Specifically, by studying changes in the number of variables impacted on Tem, finding that the the size of hyperparameter $hidden\_size$ has an extremely similar meaning as F_m, and NN_d can be understood as the function consisting of the BiLSTM with learnable parameters θ. We will further elaborate this in detail in ablation experiments to clarify their relationship.

Network architecture In order to extract the spatiotemporal information of used data as much as possible to make the filled missing data more accurate, we carefully design a new network architecture containing spatiotemporal attention mechanism. It is named as spatiotemporal attention network (STA-Net), as shown in Fig. 4. STA-Net is made up of two encoder-decoder based sub network frameworks: the PConvU-Net [24] based on local spatial attention model (LSAPConvU-Net) and the PConvU-Net based on multidimensional temporal self-attention model (MTSAPConvU-Net). And both two sub-networks are based on encoder-decoder architecture [23].

The encoder layers and decoder layers of both LSAPConvU-Net and MTSAPConvU-Net adopt partial convolution (PConv) [24] block as shown in Fig. 7. By using the PConv in encoder layers, more abstract and higher representation or feature map [23] could be obtained. In decoder layers, utilize such features and the output of encoder layers to reconstruct the missing information [24]. Meanwhile, in both LSAPConvU-Net and MTSAPConvU-Net, we use the Hardswish activation function [42] but not the ReLU activation function [43] in encoder layers and the LeakyReLu activation function [44, 45] in decoder layers to accelerate network training. Following the activation function, the batch normalization [46] is also applied to accelerate training and improve the network performance. Besides, the architectures based on PConvU-Net apply the skip connections to the two corresponding feature maps in both encoder and decoder layers, which not only promotes the network performance and the training speed, but also alleviates the phenomenon of gradient disappearance or explosion, which is similar as the residual net [47]. In skip connections, the upsampling keeps the size of the two corresponding feature maps same. In the whole network architecture, the convolutional kernel size in both encoder and decoder layers is set to . The stride is set to 2 in encoder layers and set to 1 in decoder layers. In the last layer of network, kernel size, out-channels and stride are set to , 1 and 1, separately. The detailed parameters could be found in Table 1.

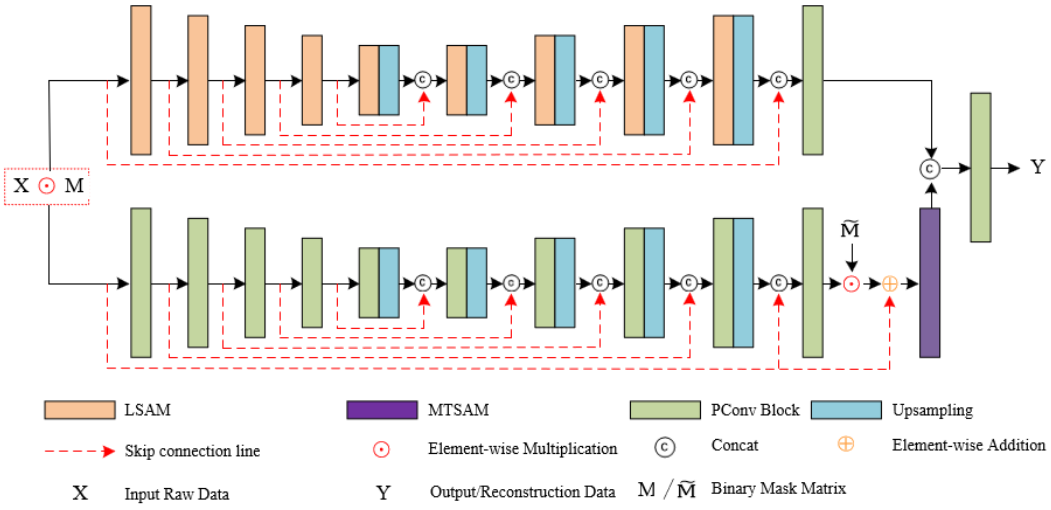

Fig. 4. The architecture of STA-Net which is consisting of two sub-networks: the PConvU-Net based on local spatial attention model on top (LSAPConvU-Net) and the PConvU-Net based on multidimensional temporal self-attention model beneath (MTSAPConvU-Net). First, we input the mask data ($X \odot M$) into both sub-networks, in turns. In MTSAPConvU-Net, the result obtained through PConvU-Net will be first multiplied by $\widetilde{M}$ (which equals to $1 - M$), secondly added by the mask data and input into multidimensional temporal self-attention. Next we concatenate the output of the two sub-networks. Finally, after the concatenated result goes through the last layer, we can get the final reconstruction data Y. The local spatial attention model (LSAM), multidimensional temporal self-attention model (MTSAM) and PConv Block are shown as Fig. 5, Fig. 6 and Fig. 7, respectively.

3.3 Loss function

When training a neural network, we need a suitable loss function to guide the network to convergence [30]. In meteorological field, root mean squared error (RMSE) is mostly used as the evaluation index for the reconstruction algorithm of missing data. Concurrently, due to the data input to proposed network is a numerical matrix which can be regarded as image data, we can also use the loss function in the field of image repair. According to the above reason, here we chose the L_2 function with a mask [23] as our loss function. For each ground truth input data X with corresponding mask M, our reconstruction network R can output corresponding imputation result $R(X \odot M)$. M is a binary mask with the same size as X, where 0 corresponds to the data is discarded while 1 means the opposite. Then the loss function is expressed as:

$$Loss_{mse} = ||(X - R(X \odot M)) \odot (1 - M)||_2^2 \tag{11}$$

wherein $\odot$ is the element-wise product operation. In our experiments, mimicking the preprocessing method of the image data in the computer vision (CV) field

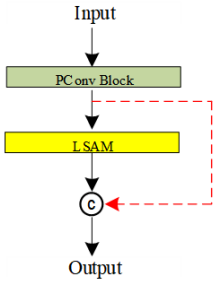

Fig. 5. The local spatial attention model. Taking the output from the upper layer of neural network as input, then passing through PConv Block and the local spatial attention mechanism, concatenating the output of two networks by channels to obtain result with local spatial information.

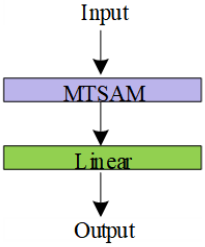

Fig. 6. The multidimensional temporal self-attention model. After the output of the previous layer goes through multidimensional temporal self-attention mechanism, expression containing timing information could be obtained. Since multidimensional temporal self-attention will change the size of input data, we apply the linear network to deal with this issue and get the output with temporal correlation.

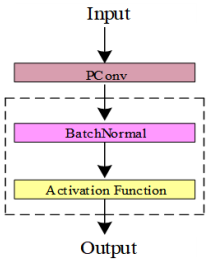

Fig. 7. The partial convolution block.

Table 1. The details of STA-Net

subnetwork_name	net_name	kernel_size	in_channels	out_channels	stride	activation function
LSAPConvU-Net	Net1	3	1	32	2	Hardswish
	Net2	3	64	128	2	Hardswish
	Net3	3	256	512	2	Hardswish
	Net4	3	1024	1024	2	Hardswish
	Net5	3	2048	2048	2	Hardswish
	Net6	3	6144	1536	1	LeakyReLU(0.3)
	Net7	3	4096	1024	1	LeakyReLU(0.3)
	Net8	3	2304	576	1	LeakyReLU(0.3)
	Net9	3	1216	304	1	LeakyReLU(0.3)
	Net10	3	609	1	1	LeakyReLU(1.)
MTSAPConvU-Net	Net1	3	1	32	2	Hardswish
	Net2	3	32	64	2	Hardswish
	Net3	3	64	128	2	Hardswish
	Net4	3	128	256	2	Hardswish
	Net5	3	256	512	2	Hardswish
	Net6	3	768	384	1	LeakyReLU(0.3)
	Net7	3	512	256	1	LeakyReLU(0.3)
	Net8	3	320	160	1	LeakyReLU(0.3)
	Net9	3	192	96	1	LeakyReLU(0.3)
	Net10	3	97	1	1	LeakyReLU(1.)
last layer	Net	3	2	1	1	LeakyReLU(1.)

before being input to the network [23–26], the data used was normalized to minimum-maximum normalization between $[-1, 1]$.

4 Experiments

To evaluate the imputation effectiveness of the spatiotemporal attention network (STA-Net) in the meteorological missing data reconstruction task, in three different missing rates: 25%, 50% and 75%, we performed a large number of experiments on two meteorological station datasets: Tibet and Qamdo. Meanwhile, STA-net was compared with other three classical models: U-Net [48], BiLSTM [17] and PConv-Net [24]. Besides, to examine the effectiveness of each part and design attention, we also performed ablation experiments on the Tibet dataset with three missing rates: 25%, 50% and 75%.

4.1 Experimental details and evaluation indicators

In training networks, the Adam algorithm [49] was used as the optimizer with an initial learning rate $1e^{-3}$. When the test rmse remained unchanged for 10 consecutive times, learning rate was reduced to one-tenth of the last time until $1e^{-8}$. The stochastic gradient descent (SGD) [50] was used to train our models where the batch size was set to 32, and training epoch was set to 300, 400 or 600

as required. Our all models have been trained and tested upon a server configured with Nvidia Tesla T4 GPUs, intel(R) Xeon(R) Gold 6242 CPU @ 2.80GHz and 386G memory. Meantime, in order to measure the reconstruction performance of four models, we selected three different indicators, including mean absolute error (MAE) and root mean squared error (RMSE)[2, 5, 17, 18], coefficient of determination (R^2) [51], computed as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{14}$$

wherein x_i and y_i are true and imputation values, respectively. $\overline{x}$ is mean value.

4.2 Datasets and mask

TibetDataset The original Tibet meteorological data set recorded the meteorological data observed hourly from 2009 to 2019. Used this data set, s was set to 61. After it went through the proposed data processing method, missing data was eliminated to avoid the impact of missing values contained in the data on the experimental accuracy. The non-missing data used in the experiment was obtained, named as TND. The TND contains 2918 numeric matrices and is divided into two parts: training set (80% of the total) and test set (20% of the total). The size of each numeric matrix is $(24, 61)$.

QamdoDataset The original Qamdo meteorological data set records the meteorological data observed hourly from 1978 to 2020. We chose 37 meteorological observation stations in Qamdo area to use in the following experiment, that is, s is set to 37. The relevant data processing process is same as TND. Finally, we acquired the the non-missing data used in the experiment, named as QND, in which 80% was used for training and 20% for testing and validation. It contained 4980 numeric matrices whose size was $(24, 37)$.

Mask To simulate the actual situation, we also analyzed the change rule of missing values of meteorological stations and found that the missing values were random in time. Based on the findings, by randomly masking at time, we made binary mask matrix with different missing ratios whose dimension was the same as used data.

4.3 Experimental results

Analysis of TND experimental results To compare the performance of four models in meteorological missing data reconstruction, we conducted experiments on the TND dataset when missing rates were 25%, 50% and 75%, respectively. We use three evaluation metrics for comparison: MAE, RMSE, and R^2. Table 2 lists the performance comparison of STA-Net with other methods on the TND dataset, and the best results are shown in bold. Moreover, by filling the missing data of a single site within different missing rates, we try to show the performance difference between models from the side. Here, we select Lhasa weather station as a representative example, the experimental results are shown in Fig. 8, Fig. 9 and Fig. 10.

Table 2. Performance comparison results of different models with 25%, 50% and 75% missing rate on TND dataset.

Missing rate	Method	MAE(↓)	RMSE(↓)	R^2(↑)
25%	U-Net[48]	0.3594	0.9362	0.9980
	PConv-Net[24]	0.2249	0.6022	0.9962
	BiLSTM[17]	0.1638	0.4769	0.9975
	STA-Net(ours)	**0.1430**	**0.4093**	**0.9982**
50%	U-Net[48]	0.8113	1.4799	0.9770
	PConv-Net[24]	0.5648	1.0530	0.9883
	BiLSTM[17]	0.4248	0.8599	0.9920
	STA-Net(ours)	**0.3874**	**0.7536**	**0.9939**
75%	U-Net[48]	1.2252	1.8140	0.9654
	PConv-Net[24]	1.1098	1.6656	0.9707
	BiLSTM[17]	0.8915	1.4296	0.9778
	STA-Net(ours)	**0.8819**	**1.3496**	**0.9794**

As it can be observed from the results in Table 2 and figures, the performance of STA-Net is the best compared to other methods. For example, in the case with the missing rate of 25% on test datasets and compared to the other three methods, the MAE declines by 60.21%, 36.42% and 12.70%; the RMSE declines by 56.28%, 32.03% and 14.17%; the R2 increases by 0.75%, 0.20% and 0.07%. At the same time, the experimental results also show that the multiple-meteorological-station data processing method proposed is effective and workable. Furthermore, the results also reflect that PConvU-Net performs better than U-Net, since partial convolution is proposed to solve the problem that ordinary convolutional neural network does not take into account the effect of the missing part on the extracted spatial features [24] in image repair field. Meanwhile, the performance of BiLSTM is much better than PConvU-Net and U-Net, since temporal dependence of site data self is stronger than spatial correlation reflected by matrix data generated by our data processing method. However, these models do not perform as well as STA-Net, since it makes full use of space-time correlation through the two attention mechanisms designed. The experimental results demonstrate that

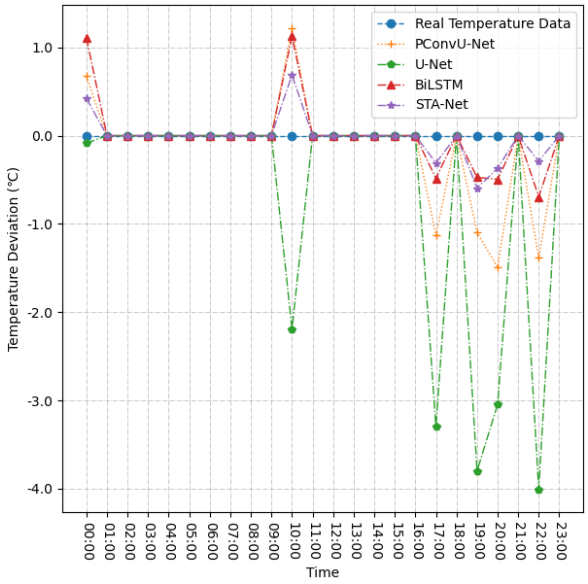

Fig. 8. The differences between reconstruction values Y and real values X and when the missing rate is 25% on TND test datasets, which are obtained by different models in Lhasa meteorological station.

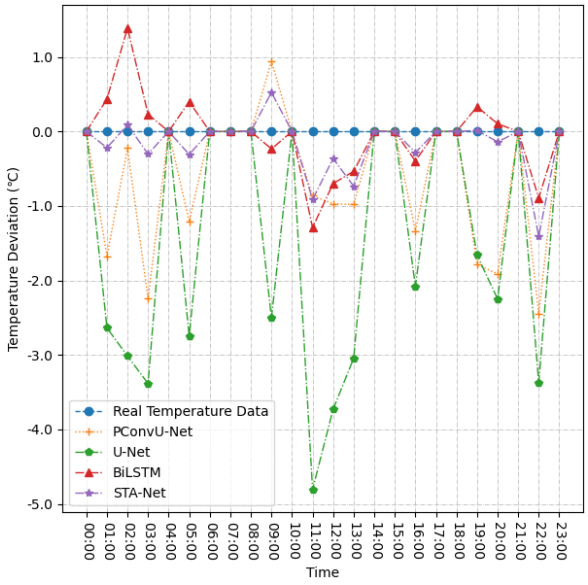

Fig. 9. The differences between reconstruction values Y and real values X and when the missing rate is 50% on TND test datasets, which are obtained by different models in Lhasa meteorological station.

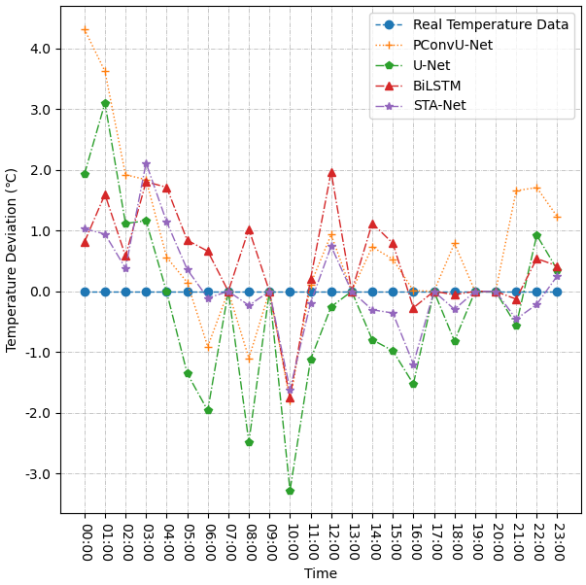

Fig. 10. The differences between reconstruction values Y and real values X and when the missing rate is 75% on TND test datasets, which are obtained by different models in Lhasa meteorological station.

our constructed spatiotemporal attention neural network is effective in meteorological missing data reconstruction tasks.

Analysis of QND experimental results To further evaluate the performance of our method in meteorological missing data reconstruction, we conducted experiments with 25%, 50% and 75% missing rate on QND dataset. Table 3 lists the performance comparison of STA-Net with other methods on the QND dataset, and the best results are shown in bold. It can be again observed that the designed data preprocessing method is effective and feasible, and generated datasets not only utilizes CNN-based networks to reconstruct missing data in space but also makes use of LSTM-based models to fill missing data in time sequence.

Besides, due to the meteorological stations in Qamdo area are relatively closer geographically compared with TND, the QND has stronger spatial correlation. Hence the performance gaps between PConvU-Net and BiLSTM are smaller. However, since these two models only consider space or time correlation, their reconstruction effects are not as good as STA-Net. As the same as previous experiments, here, Qamdo meteorological station is selected to side-on show analogous conclusions, the corresponding experimental results are shown in Fig. 11, Fig. 12 and Fig. 13. To sum up, the experimental results again show that our constructed spatiotemporal attention neural network is effective in meteorological missing data reconstruction tasks.

Table 3. Performance comparison results of different models with 25%, 50% and 75% missing rate on QND dataset

Missing rate	Method	MAE(↓)	RMSE(↓)	R^2(↑)
25%	U-Net[48]	0.2252	0.6088	0.9954
	PConv-Net[24]	0.1670	0.4642	0.9973
	BiLSTM[17]	0.1509	0.4400	0.9976
	STA-Net(ours)	**0.1427**	**0.4054**	**0.9980**
50%	U-Net[48]	0.5420	1.0274	0.9869
	PConv-Net[24]	0.4232	0.8135	0.9918
	BiLSTM[17]	0.3937	0.7945	0.9923
	STA-Net(ours)	**0.3303**	**0.6574**	**0.9946**
75%	U-Net[48]	0.9463	1.4563	0.9738
	PConv-Net[24]	0.9508	1.4454	0.9743
	BiLSTM[17]	0.8543	1.3749	0.9769
	STA-Net(ours)	**0.6480**	**1.0326**	**0.9869**

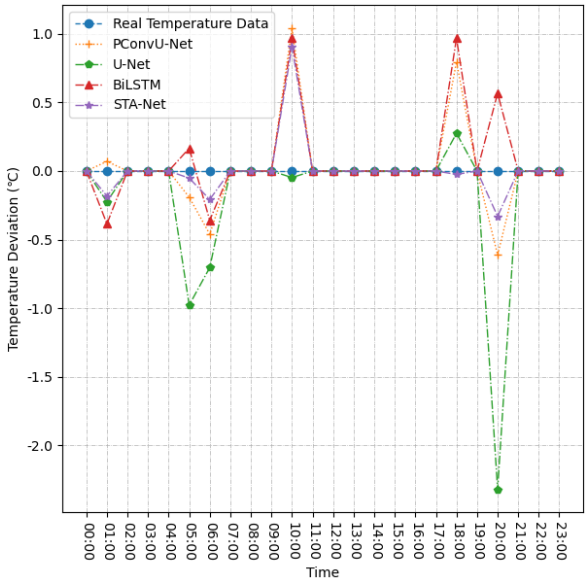

Fig. 11. The differences between reconstruction values Y and real values X and when the missing rate is 25% on QND test datasets, which are obtained by different models in Qamdo meteorological station.

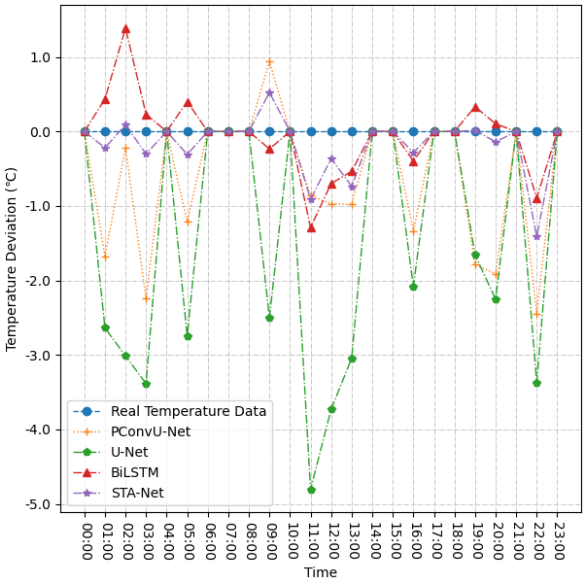

Fig. 12. The differences between reconstruction values Y and real values X and when the missing rate is 50% on QND test datasets, which are obtained by different models in Qamdo meteorological station.

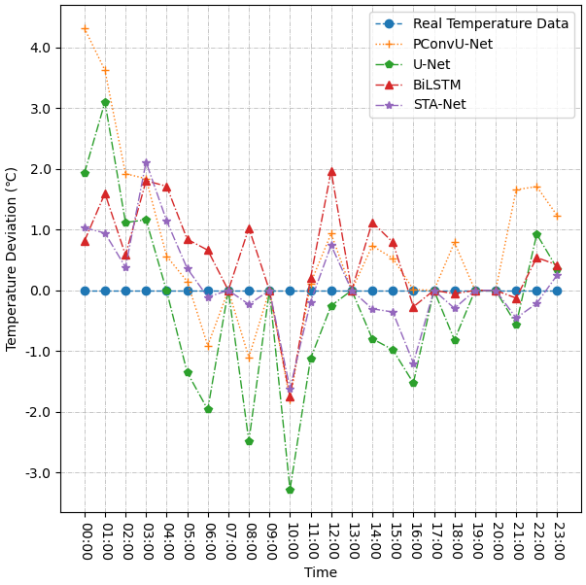

Fig. 13. The differences between reconstruction values Y and real values X and when the missing rate is 75% on QND test datasets, which are obtained by different models in Qamdo meteorological station.

4.4 Ablation experiments

We conduct the ablation experiments on the TND dataset with the 25%, 50% and 75% missing rate to evaluate that two different factors have an impact on network performance.

Effectiveness of the spatiotemporal attention model To validate the effectiveness and superiority of the local spatial attention model (LSAM) and multi-dimensional temporal self-attention model (MTSAM), we take the PConvU-Net as baseline network architecture. Then conduct the corresponding ablation experiments with the 25%, 50% and 75% missing rate, respectively. The results of ablation experiment are shown Table 4.

Table 4. The results of ablation experiments on STA-Net

Missing rate	Method	MAE(↓)	RMSE(↓)	R^2(↑)
25%	Baseline	0.2249	0.6022	0.9962
	Baseline+LSAM	0.1897	0.5117	0.9972
	Baseline+MTSAM	0.1487	0.4216	0.9981
	STA-Net(ours)	**0.1430**	**0.4093**	**0.9982**
50%	Baseline	0.5648	1.0530	0.9883
	Baseline+LSAM	0.5201	0.9703	0.9899
	Baseline+MTSAM	0.4154	0.7989	0.9931
	STA-Net(ours)	**0.3874**	**0.7536**	**0.9939**
75%	Baseline	1.1098	1.6656	0.9707
	Baseline+LSAM	1.0011	1.5165	0.9753
	Baseline+MTSAM	0.9363	1.4252	0.9781
	STA-Net(ours)	**0.8819**	**1.3496**	**0.9794**

Obviously, since PConvU-Net does not consider local spatial and temporal relations, its performance is worst. When LSAM or MTSAM is combined, due to the using attention mechanisms, spatial or temporal information is obtained so that the effect of reconstruction can get improved. When both LSAM and MTSAM are integrated, the performance will be the best because of obtaining spatiotemporal information.

Effectiveness of the hyperparameter hidden_size The hyperparameter hidden_size in LSTM is closely related to the types of factors affecting the temperature of meteorological stations. Typically, the performance of LSTM based models is influenced by it. Here, for simplicity, we conduct the ablation experiments with the 25%, 50% and 75% missing rate on MTSAPConvU-Net. The hidden size is set to 1, 32, 64, 128, and 256, respectively, to observe the impact on MTSAPConvU-Net. The results of ablation experiment are shown in Table 5, which show the impact from hidden size into MTSAPConvU-Net.

Table 5. The results of ablation experiments on MTSAPConvU-Net

Missing rate	Hidden_size	MAE(↓)	RMSE(↓)	R^2(↑)
25%	1	0.1991	0.5523	0.9968
	32	0.1774	0.4883	0.9975
	64	0.1773	0.4886	0.9975
	128	**0.1487**	**0.4216**	**0.9981**
	256	0.1731	0.4836	0.9975
50%	1	0.5208	0.9826	0.9898
	32	0.4549	0.8655	0.9921
	64	0.4750	0.8990	0.9913
	128	**0.4154**	**0.7989**	**0.9931**
	256	0.4186	0.8024	0.9929
75%	1	1.0223	1.5736	0.9744
	32	**0.0.9363**	**1.4252**	**0.9781**
	64	0.9928	1.4497	0.9762
	128	0.9535	1.4491	0.9774
	256	0.9438	1.4340	0.9779

It is observed from the experimental results in Table 5 that the performance comparison in different settings of the hidden size. Roughly, as the hidden size becomes larger, the performance of the model increases first and then drop. We infer that when the hidden size is small, multiple factors decomposed cannot fully express the data, which suppresses the performance of the model. Especially when it is set to 1 as an extreme setting, only one constraint has an impact on the data and the model performance theoretically is the worst in this case. However, if the size is too big, a huge number of factors may lead to over expression of the data, which will also degrade the performance of the model due to resource waste and reduce the time efficiency. Hence, the hidden size should be cautiously set in a reasonable range to achieve the best performance.

5 Conclusion and future work

In this research, we mainly designed a now network architecture STA-Net to fill missing data in multiple meteorological station data. We first proposed a simply novel a multiple-meteorological-station data processing method, making generated data able to utilize currently the most popular CNN to fill the missing values. At the same time, in order to exploit the spatio-temporal correlation in data as much as possible, we designed two attention mechanisms: local spatial attention mechanism and multidimensional temporal self-attention mechanism. Through a large number of experiments, it has been proved that the proposed data processing method is feasible. By conducting the ablation study, we further confirmed that the two attention models could greatly improve the effect of missing value reconstruction. In this way, our work can provide a meaningful reference for the repair of missing values of multiple meteorological station data, even other kinds of data. In the future, we will further extend our current

framework to be applicable to support other types of meteorological data, such as wind speed, relative humidity and so on. In addition, better network models could be further exploited to improve the accuracy and time.

Acknowledgements This work is funded by the National Natural Science Foundation of China (No.42265010, No.62162053, No.62062059, No.62166032), Natural Science Foundation of Qinghai Province (No.2023-ZJ-906M), Youth Scientific Research Foundation of Qinghai University (No.2022-QGY-6) and the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (No.2020-ZZ-03).

References

1. Peterson, T.C., Vose, R.S.: An overview of the Global Historical Climatology Network temperature database. Bulletin of the American Meteorological Society vol. 78(12), 2837–2850 (1997)
2. Lompar, M., Lalić, B., Dekić, L., Petrić, M.: Filling gaps in hourly air temperature data using debiased ERA5 data. Atmosphere vol. 10(1), pp. 13. Publisher, MDPI (2019)
3. Lara-Estrada, L., Rasche, L., Sucar, L.E., Schneider, U.A.: Inferring missing climate data for agricultural planning using Bayesian networks. Land vol. 7(1), pp. 4. Publisher, MDPI (2018)
4. Huang, M., Piao, S., Ciais, P., Peñuelas, J., Wang, X., Keenan, T.F., Peng, S., Berry, J.A., Wang, K., Mao, J., et al.: Air temperature optima of vegetation productivity across global biomes. Nature ecology & evolution vol. 3(5), pp. 772–779. Publisher, Nature Publishing Group UK London (2019)
5. Henn, B., Raleigh, M.S., Fisher, A., Lundquist, J.D.: A comparison of methods for filling gaps in hourly near-surface air temperature data. Journal of Hydrometeorology vol. 14(3), pp. 929–945. Publisher, American Meteorological Society (2013)
6. Afrifa-Yamoah, E., Mueller, U.A., Taylor, S., Fisher, A.: Missing data imputation of high-resolution temporal climate time series data. Meteorological Applications vol. 27(1), pp. 1873. Publisher, Wiley Online Library (2020)
7. Park, J., Yoon, D., Seol, S.J., Byun, J.: Reconstruction of seismic field data with convolutional U-Net considering the optimal training input data. In: SEG International Exposition and Annual Meeting. Publisher, OnePetro (2019)
8. Daly, S., Davis, R., Ochs, E., Pangburn, T.: An approach to spatially distributed snow modelling of the sacramento and san joaquin basins, california. Hydrological Processes vol. 14(18), pp. 3257–3271. Publisher, Wiley Online Library (2000)
9. Tobin, C., Nicotina, L., Parlange, M.B., Berne, A., Rinaldo, A.: Improved interpolation of meteorological forcings for hydrologic applications in a swiss alpine region. Journal of Hydrology vol. 401(1-2), pp. 77–89. Publisher, Elsevier (2011)
10. Garen, D.C., Johnson, G.L., Hanson, C.L.: Mean areal precipitation for daily hydrologic modeling in mountainous regions 1. JAWRA Journal of the American Water Resources Association vol. 30(3), pp. 481–491. Publisher, Wiley Online Library (1994)
11. Pape, R., Wundram, D., L¨offler, J.: Modelling near-surface temperature conditions in high mountain environments: an appraisal. Climate Research vol. 39(2), pp. 99–109 (2009)

12. Stahl, K., Moore, R., Floyer, J., Asplin, M., McKendry, I.: Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. Agricultural and forest meteorology vol. 139(3-4), pp. 224–236. Publisher, Elsevier (2006)
13. Belachsen, I., Broday, D.M.: Imputation of missing pm2. 5 observations in a network of air quality monitoring stations by a new k nn method. Atmosphere vol. 13(11), pp. 1934. Publisher, MDPI (2022)
14. Chen, M., Zhu, H., Chen, Y., Wang, Y.: A novel missing data imputation approach for time series air quality data based on logistic regression. Atmosphere vol. 13(7), pp. 1044. Publisher, MDPI (2022)
15. Kihoro, J., Athiany, K., et al.: Imputation of incomplete nonstationary seasonal time series data. Mathematical Theory and Modeling vol. 3(12), pp. 142—154 (2013)
16. Wang, H., Yuan, Z., Chen, Y., Shen, B., Wu, A.: An industrial missing values processing method based on generating model. Computer Networks vol. 158, pp. 61—68. Publisher, Elsevier (2019)
17. Cao, W., Wang, D., Li, J., Zhou, H., Li, L., Li, Y.: Brits: Bidirectional recurrent imputation for time series. Advances in neural information processing systems vol. 31 (2018)
18. Zhang, Y.F., Thorburn, P.J., Xiang, W., Fitch, P.: SSIM—A deep learning approach for recovering missing time series sensor data. IEEE Internet of Things Journal vol. 6(4), pp. 6618—6628. Publisher, IEEE (2019)
19. Zhang, Y., Thorburn, P.J.: A dual-head attention model for time series data imputation. Computers and Electronics in Agriculture vol. 189, pp. 106377. Publisher, Elsevier (2021)
20. Dagtekin, O., Dethlefs, N.: Imputation of Partially Observed Water Quality Data Using Self-Attention LSTM. In: 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1—8. Publisher, IEEE (2022)
21. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation vol. 9(8), pp. 1735—1780. Publisher, MIT press (1997)
22. Xie, C., Huang, C., Zhang, D., He, W.: BiLSTM-I: A deep learning-based long interval gap-filling method for meteorological observation data. International Journal of Environmental Research and Public Health vol. 18(19), pp. 10321. Publisher, MDPI (2021)
23. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536—2544 (2016)
24. Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 85—100 (2018)
25. Shin, Y.G., Sagong, M.C., Yeo, Y.J., Kim, S.W., Ko, S.J.: Pepsi++: Fast and lightweight network for image inpainting. IEEE transactions on neural networks and learning systems vol. 32(1), pp. 252—265. Publisher, IEEE (2020)
26. Peng, J., Liu, D., Xu, S., Li, H.: Generating diverse structure for image inpainting with hierarchical VQ-VAE. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10775—10784 (2021)
27. Kadow, C., Hall, D.M., Ulbrich, U.: Artificial intelligence reconstructs missing climate information. Nature Geoscience vol. 13(6), pp. 408—413. Publisher, Nature Publishing Group UK London (2020)
28. Irrgang, C., Saynisch-Wagner, J., Dill, R., Boergens, E., Thomas, M.: Self-Validating Deep Learning for Recovering Terrestrial Water Storage From Gravity

and Altimetry Measurements. Geophysical Research Letters vol. 47(17), pp. 2020—089258. Publisher, Wiley Online Library (2020)
29. Dong, J., Yin, R., Sun, X., Li, Q., Yang, Y., Qin, X.: Inpainting of remote sensing SST images with deep convolutional generative adversarial network. IEEE geoscience and remote sensing letters vol. 16(2), pp. 173—177. Publisher, IEEE (2018)
30. Yu, J., Wu, B.: Attention and hybrid loss guided deep learning for consecutively missing seismic data reconstruction. IEEE Transactions on Geoscience and Remote Sensing vol. 60, pp. 1-–8. Publisher, IEEE (2021)
31. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5505-–5514 (2018)
32. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems vol. 30, pp. 5998-–6008 (2017)
34. Chaudhari, S., Mithal, V., Polatkan, G., Ramanath, R.: An attentive survey of attention models. ACM Transactions on Intelligent Systems and Technology (TIST) vol. 12(5), pp. 1-–32. Publisher, AcM New York, NY, USA (2021)
35. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8628—8638 (2021)
36. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. Advances in Neural Information Processing Systems vol. 34, pp. 852-–863 (2021)
37. Liu, H., Ruan, Z., Zhao, P., Dong, C., Shang, F., Liu, Y., Yang, L., Timofte, R.: Video super-resolution based on deep learning: a comprehensive survey. Artificial Intelligence Review vol. 55(8), pp. 5981–6035. Publisher, Springer (2022)
38. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Transactions on Graphics (ToG) vol. 26(3), pp. 4. Publisher, AcM New York, NY, USA (2007)
39. Zeiler, M.D., Taylor, G.W., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. In: 2011 International Conference on Computer Vision, pp. 2018—2025. Publisher, IEEE (2011).
40. Shi, W., Caballero, J., Theis, L., Huszar, F., Aitken, A., Ledig, C., Wang, Z.: Is the deconvolution layer the same as a convolutional layer? arXiv preprint arXiv:1609.07009 (2016)
41. Rosenbaltt, F.: The perceptron–a perciving and recognizing automation. Cornell Aeronautical Laboratory (1957)
42. Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., et al.: Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314—1324 (2019)
43. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM vol. 60(6), pp. 84–90. Publisher, AcM New York, NY, USA (2017)
44. Maas, A.L., Hannun, A.Y., Ng, A.Y., et al.: Rectifier nonlinearities improve neural network acoustic models. In: Proc. Icml, vol. vol. 30, pp. 3 (2013). Atlanta, Georgia, USA
45. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853 (2015)

46. Ioffe, S., Szegedy, C.:Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448—456. Publisher, PMLR (2015)
47. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770—778. Publisher, IEEE (2016)
48. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234—241. Publisher, Springer (2015)
49. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
50. Bottou, L.: Stochastic gradient descent tricks. Neural Networks: Tricks of the Trade: Second Edition, pp. 421-–436. Publisher, Springer (2012)
51. Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M.: Methods for imputation of missing values in air quality data sets. Atmospheric environment vol. 38(18), pp. 2895-–2907. Publisher, Elsevier (2004)