Deep Learning for Near-Surface Air Temperature Estimation from FY-4A Satellite Data

Shanmin Yang, Qing Ren, Ningfang Zhou, Yan Zhang*, Xi Wu*

***Abstract*—Near-surface air temperature is a crucial weather parameter that significantly impacts human health and is widely utilized in numerical weather forecasting and climate prediction studies. However, the most common ground-based meteorological station observation and radar observation are often limited by geographic and natural constraints. With the advantages of global coverage and high spatiotemporal resolution, satellite remote sensing has become a valuable support in overcoming data scarcity issues related to ground-based station and radar observations in complex geographic and natural conditions. Although remote sensing indirectly reflects atmosphere variables (e.g., near-surface air temperature), accurately estimating the atmosphere variables through satellite remote sensing remains a significant challenge. This paper introduces a deep learning Transformer-based neural network (TaNet) for near-surface air temperature estimation. TaNet automatically extracts information from imageries captured by China's new-generation geostationary meteorological satellite FengYun-4A and generates grid near-surface air temperature data in near real-time. Extensive experiments conducted using the state-of-the-art operational reanalysis product ERA5 and meteorological station observations as benchmark standards demonstrate the effectiveness and superiority of TaNet. It achieves an impressive Pearson's correlation coefficient (CC) of 0.990 with ERA5 and 0.959 with station observations, outperforming the other products, such as CFSv2, CRA, and U-Net, on root mean square error (RMSE) and CC metrics. TaNet reduces the RMSE of CFSv2, CRA, and U-Net by a margin of 10.551% (2.594°C vs. 2.900°C), 2.261% (2.594°C vs. 2.654°C), and 5.535% (2.594°C vs. 2.746°C), respectively, using station observations as the benchmark.**

***Index Terms*—Near-Surface Air Temperature; Deep learning; FengYun 4A Satellite**

I. Introduction

NEAR-SURFACE air temperature, also known as the 2m temperature, is one of the most essential weather parameters for human health. Accurate and high-resolution near-surface air temperature is crucial for reducing loss caused by natural disasters such as drought, and various scientific studies such as numerical weather forecasting and climate prediction [1], [2]. Out of the three most commonly used meteorological measurement methods, ground-based station observations and radar observations may be hindered by geographical and natural barriers [1], [3], resulting in sparse and unevenly distributed data that can not accurately reflect the true spatial and temporal distribution of meteorological conditions. However, satellite remote sensing offers a solution to these challenges by providing global coverage and high temporal resolution data, making it an invaluable tool for overcoming the limitations of ground-based meteorological station observations and radar observations, especially in regions with complex terrain.

In recent decades, significant advancements in satellite remote sensing and deep learning technologies have opened up new possibilities to produce more accurate and real-time meteorological products based on massive satellite data and deep learning techniques. For instance, Tao et al. [4] propose to improve the accuracy of precipitation estimated from the Geostationary Operational Environmental Satellite infrared imagery with a stacked denoising autoencoder network. Similarly, Sadeghi et al. [5] improve the accuracy of near real-time precipitation estimation by using a U-Net convolutional neural network (CNN), the geographical and infrared (IR) information. Gao et al. [3] propose an attention-based U-Net framework for precipitation estimation from FengYun 4A (FY-4A) satellite data. Hu et al. [6] present a deep learning architecture consisting of four U-Net sub-networks for precipitation bias correction. Jiang et al. [7] propose a deep learning-based method to supplement the missing radar echo data by reconstruction from satellite data. By combining model-data-knowledge-driven and deep learning techniques, Wang et al. [8] suggest a method to overcome the ill-conditioned problem of land surface temperature retrieval by combining model-data-knowledge-driven and deep learning. Guo et al. [9] put forward a deep learning-based method for classifying rainfall levels in Synthetic Aperture Radar (SAR) images, with the aim of improving sea-surface wind speed retrieval in conjunction with existing rainfall correction models. Experiments conducted in these studies showed the potential and advantages of utilizing deep learning methods in meteorology. However, improvements are still needed for practical applications.

For near-surface air temperature estimating, Liu et al. [10] employ a statistical method on Landsat 8 imagery data and auxiliary data. The estimation accuracy of several machine-learning approaches over regions of China is evaluated by [1], [11]. Zhou et al. [12] develop a two-stage machine learning method to estimate near-surface air temperature from multi-source skin temperature. Shen et al. [13] propose a deep belief network (DBN) that estimates near-surface air temperature

Shanmin Yang is with Chengdu University of Information Technology, Chengdu, China (e-mail: yangsm@cuit.edu.cn).

Qing Ren is with Chengdu University of Information Technology, Chengdu, China.

Ningfang Zhou is with National Meteorological Center of CMA, Beijing, China.

Yan Zhang is with Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, National Satellite Meteorological Center (National Center for Space Weather), China Meteorological Administration, Innovation Center for FengYun Meteorological Satellite (FYSIC), Beijing, China.

Xi Wu is with Chengdu University of Information Technology, Chengdu, China (e-mail: xi.wu@cuit.edu.cn).

Corresponding authors: Yan Zhang, Xi Wu

using remote sensing, socioeconomic, and assimilation data. However, there is a lack of relevant deep learning-based studies on near-surface air temperature estimation based on FY-4A [1], which is still a challenging issue.

This paper presents a deep learning-based method named as TaNet to estimate the near-surface air temperature T_a. Based on FY-4A satellite data, TaNet can reconstruct T_a with an advantage of near-real-time over the reanalysis products such as ERA5, CRA, etc. Besides, experimental results demonstrate the excellent accuracy of TaNet. It outperforms CFSv2, CRA, and U-Net products on RMSE and Pearson's correlation coefficient metrics. TaNet provides a promising solution to address the shortage problem of ground-based meteorological observation data.

The remainder of this paper is structured as follows: Section II describes the study area and data preparation. The proposed method, loss function, and evaluation metrics are presented in section III. In section IV, we conduct extensive experiments to compare our approach with operational reanalysis products and meteorological station observation data. We discuss the experimental results and future research directions in section V. Finally, we summarize this paper in section VI.

II. Study Area and Data

A. Study Area

For our study, we selected the region between 50°S-50°N and 40°E-140°E, according to the coverage of the FY-4A remote sensing satellite. This region is highly diverse in topography, climate, and vegetation. It is also characterized by various weather phenomena, including tropical cyclones, monsoons, and droughts, which makes it an ideal location for evaluating the effectiveness of our proposed method across different environments.

B. Data

1) Fengyun-4A satellite data: Fengyun-4A satellite is the second generation of geostationary meteorological satellites developed and launched by the China Meteorological Administration (CMA). It was put into meteorological operation on May 1, 2018. The Fengyun-4A satellite is characterized by its onboard Advanced Geostationary Radiation Imager (AGRI), which captures satellite brightness temperature imageries with spatial resolutions of 4km and temporal intervals of $15-60$ minutes for full-disk image observations [14], [15]. Each imagery captured has a total of 14 spectral channels [16]–[18], covering visible/near-infrared ($0.45\sim0.90\mu m$), short-wave infrared ($1.36\sim2.35\mu m$), mid-wave infrared ($3.5\sim4.0\mu m$), long-wave infrared ($8.0\sim13.8\mu m$), and water vapor ($5.8\sim7.3\mu m$) bands. As only the 12 and 13 channels contain surface temperature-related information, we selected channels 12 and 13 as input in our paper, liking the existing work [1]. For more details and (or) download of FY4A/AGRI data, please refer to the China National Satellite Meteorological Center (http://satellite.nsmc.org.cn).

TABLE I
List of datasets used in this study. All datasets are selected in the region between 50°S-50°N and 40°E-140°E, with a spatial resolution of 0.25°; the training and testing datasets are disjoined.

Datasets	Variables	Spatial*/ Time Resolution	Usage
FY-4A/AGRI	bands 12	4km hourly	training, testing
	bands 13	4km hourly	testing
ERA5	2m temperature	0.25° hourly	training, testing
CRA	2m temperature	0.25° daily	testing
CFSv2	2m temperature	0.25° hourly	testing
Station Observation	2m temperature	- daily	testing

* The approximate conversion between degree and km is: $1° \approx 111$ km

2) Reanalysis Data: ERA5 is the fifth generation global reanalysis data product provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) [19]. It has become one of the most accurate global reanalysis data products, providing a wide range of climate factors such as temperature, pressure, humidity, wind speed, and precipitation. The ERA5 dataset is publicly available and widely used in climate research, weather forecasting, etc. For our study, ERA5 data in the region between 50°S-50°N and 40°E-140°E is selected, with a spatial resolution of $0.25°\times$ $0.25°$ and a temporal resolution of one hour. For more details about the ERA5 data, please refer to https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5.

CRA is China's first-generation global atmospheric reanalysis product, which provides a comprehensive view of the Earth's atmosphere from 1979 to the present. The CRA dataset includes a wide range of atmospheric variables such as temperature, pressure, precipitation, etc. These variables are available from the ground to an altitude of 55 kilometers [20], at time resolutions of 3 hours, 6 hours, daily, and monthly. Specifically, the daily CRA dataset is selected for testing in this paper. For more details about and download the CRA product, please refer to China Meteorological Data Service Centre (http://data.cma.cn/).

In addition, the NCEP Climate Forecast System Reanalysis (CFSR) product is introduced and analyzed in this paper. CFSR is a global climate reanalysis dataset produced by the National Centers for Environmental Prediction (NCEP), providing a comprehensive view of the Earth's climate from 1979 to the present. It was initially completed between 1979 and 2009 and extended to March 2011 [21]. On March 30, 2011, the CFSR dataset was upgraded to a new version (CFSv2) [22]. CFSv2 dataset with hourly temporal resolution and $0.25° \times 0.25°$ spatial resolution is selected in this study. For more details about CFSv2, please refer to https://rda.ucar.edu/.

3) Meteorological Observation Station Data: Meteorological observation station data is used as a benchmark in this study at the testing stage. In detail, meteorological station observation data from 2231 stations (between 50°S-50°N and 40°E-140°E), in January, April, July, and October of 2020 year, are used to measure the performance of the proposed method in this paper. And these data are provided by the China Meteorological Data Service Centre.

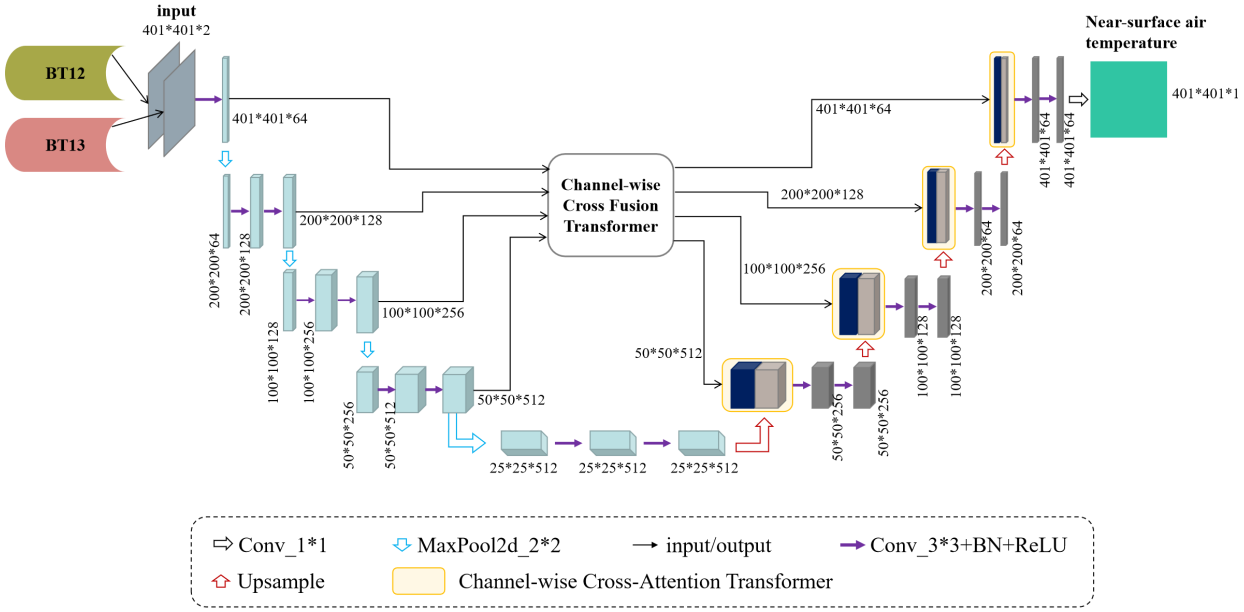

Fig. 1. Framework of TaNet model. BT12 and BT13 are the 12 and 13 spectral bands of FY-4A/AGRI satellite imagery, respectively; the light blue and gray boxes are the output in each layer at the encoding and decoding stages, respectively; the deep blue boxes are the output of the Channel-wise Cross Fusion Transformer; and the numbers are the input/output shape size (height, width, and channels respectively).

C. Data Preprocessing

All data used in this study are listed in Table I. Before sending to the proposed framework, imageries of the twelve and thirteen bands of FY-4A/AGRI are converted to equal latitude and longitude projection and aligned with the ERA5 data (with $0.25° \times 0.25°$ resolution, and 401 $\times$ 401 pixels) through spatial interpolation using the nearest-neighbor interpolation method. For matching with the ERA5 data in temporal, the latest 15-minute observations within one hour are selected. That is, one frame of FY-4A/AGRI data in every hour is chosen in this study.

Overall, the input data consists of 12 and 13 channels of FY-4A/AGRI data, with a spatial resolution of $0.25° \times 0.25°$ and a temporal resolution of one hour. Data from April 2021 to May 2022 (training set) are used for deep model training; and data from January, April, July, and October 2020 (test set) are used for performance verification. In addition, the ERA5 data and station observations are introduced as reference benchmarks for the grid and site values, respectively.

III. Method

A. Problem Formulation

Our method takes the twelve and thirteen bands of FY-4A/AGRI satellite imageries as input data. Let $BT12$ and $BT13$ represent the above-mentioned twelve and thirteen bands, the near-surface air temperature estimation process can be formulated as:

$$\hat{T_a} = G_f(x; \phi_f) \tag{1}$$

$$x = Concat(BT12, BT13) \tag{2}$$

where $Concat(.,.)$ means combing the multiple input sources along the channel dimension; $x \in R^{c \times h \times w}$ (c, h, w are the number of channels, height, and width) represents the merged input data; G_f is the near-surface air temperature estimation deep neural network parameterized by ϕ_f; $\hat{T_a}$ is the reconstructed near-surface air temperature results.

B. Model Structure

Deep learning technologies have attracted increasing attention from the meteorological community in recent years. As a classical encoder-decoder deep neural framework, U-Net [23] is known for combining low-level details with high-level semantic features and has been widely studied and demonstrated potential in many meteorological tasks, such as precipitation nowcasting [24] and quantitative precipitation estimation [3]. Recently, Wang et al. [25] demonstrate that not all skip connections in U-Net framework are beneficial because of the potential semantic inconsistency between features of the encoder and decoder paths. To overcome this issue, Wang et al. [25] propose a framework named UCTransNet. UCTransNet replaces the skip connections of the vanilla U-Net framework for concatenating encoder and decoder features with two modules. The first is a Channel-wise Cross Fusion Transformer (CCT) module designed to adaptively model dependencies between multi-scale encoder features from the channel-wise perspective for feature refining. The second is a Channel-wise Cross Attention (CCA) module, presented to fuse these refined encoder features with features learned during decoding and eliminate potential semantic inconsistency between them. Consistent performance improvements on three benchmark medical image segmentation tasks suggest the superiority of UCTransNet.

Inspired by [25], we adopt UCTransNet as the backbone of our TaNet framework, as shown in Fig. 1. In this framework, conv_1$*$1 represents a convolutional layer with a kernel size

of 1×1; Maxpool2d_2∗2 represents a 2D max pooling layer with a square window size of 2×2; BN denotes a batch normalization layer; ReLU indicates a nonlinear layer with the $ReLU$ activation function; and Upsample represents a upsample layer. Details of the Channel-wise Cross Fusion Transformer and Channel-wise Cross Attention modules are the same as in [25].

TaNet automatically extracts near-surface air temperature-related features from the input data through five encoding stages (the left path of the TaNet framework). The embedding features learned from the last encoding stage are used as input to the decoder (the right path of the TaNet framework), and the multi-scale features extracted at the first four encoding stages are directed into the CCT module. The CCT module refines these features through multi-scale feature fusion and multi-head channel-wise attention, which captures relationships between each input and the fused feature. Subsequently, the refined features are connected to the corresponding level of decoding feature with the help of the CCA module, which focuses on eliminating potential semantic inconsistencies.

C. loss function

The loss function used in this paper is a combination of mean square error (MSE) and mean absolute error (MAE), defined as follows:

$$L = \lambda_1 \times \frac{1}{n}\sum_{i=1}^{n}\left|T_i - \hat{T}_i\right| + \lambda_2 \times \frac{1}{n}\sum_{i=1}^{n}\left(T_i - \hat{T}_i\right)^2 \quad (3)$$

Where, T_i represents the ground-truth near-surface air temperature; $\hat{T}_i$ is the network estimation result; n is the number of samples; λ_1 and λ_2 are two parameters designed to balance the role of MSE (the second item) and MAE (the first item). By optimizing the overall loss function shown in Equation (3), the network aims to minimize the discrepancies between its predictions and the ground-truth values.

D. Evaluation Metrics

Similar to previous works such as [3], [5], [13], root mean square error (RMSE) and Pearson's correlation coefficient (CC) are adopted in this study as evaluation metrics. These metrics are defined as follows:

$$CC = \frac{\sum_{i=1}^{n}(\hat{T}_i - \bar{\hat{T}})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^{n}(\hat{T}_i - \bar{\hat{T}})^2}\sqrt{\sum_{i=1}^{n}(T_i - \bar{T})^2}} \quad (4)$$

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(\hat{T}_i - T_i)^2} \quad (5)$$

where, $\bar{T}$ is the average of all target value T_i; $\bar{\hat{T}}$ is the average of all predict vaule $\hat{T}_i$. By comparing the values of RMSE and CC against state-of-the-art products methods, we can get insights into the effectiveness of our approach in accurately estimating the near-surface air temperature.

IV. Results

A. Experimental Setting

The region between 50°S-50°N and 40°E-140°E with a resolution of $0.25° \times 0.25°$ is selected in our experiments. All products are spatiotemporally aligned before performance comparison. We implement all experiments with PyTorch 1.8.0 and Python 3.7. The framework is trained with SGD optimizer [26] with a momentum of 0.9 and a weight decay of $5e^{-4}$. The batch size is set to 8, and the initial learning rate is set to 0.001. The parameters λ_1 and λ_2 in Equation (3) are set to 1.0. All deep learning methods are trained and tested in the same way.

B. Comparisons based on ERA5

The proposed method is firstly verified on the test dataset, using ERA5 data as the benchmark. The traditional operational product CFSv2 and the deep learning framework U-Net are adopted as baselines for comparison. Results of our proposed TaNet method against the baselines are summarized in Table II. As can be seen, the two deep learning methods (U-Net and our TaNet) perform better than the CFSv2 product on both metrics. The performance gain of U-Net is 4.840% in average RMSE (1.809°C vs. 1.901°C) and 0.102% in Pearson's correlation coefficient (0.986 vs. 0.985) relative to CFSv2. While our method reduces the average RMSE of CFSv2 by a margin of 20.936% (from 1.901°C to 1.503°C), and reduces that of U-Net by 16.915% (from 1.809°C to 1.503°C). At the same time, it improves Pearson's correlation coefficient (CC) with ERA5 by 0.406% (0.990 vs. 0.986) and 0.508% (0.990 vs. 0.985) compared to U-Net and CFSv2, respectively. These performance improvements suggest the effectiveness and superiority of our proposed method.

TABLE II
COMPARSION RESULTS ON RMSE AND PEARSON'S CORRELATION COEFFICIENT (CC) USING STATION OBSERVATIONS AS THE BENCHMARK (THE UP ARROW INDICATES THAT A LARGER VALUE IS BETTER, WHILE THE DOWN ARROW SIGNIFIES THAT A SMALLER VALUE IS BETTER; BOLD VALUES INDICATE THE BEST PERFORMANCE).

Methods/Products	RMSE (°C)		CC	
	Value ↓	Gain	Value ↑	Gain
CFSv2	1.901	-	0.985	-
U-Net	1.809	4.840%	0.986	0.102%
TaNet (Ours)	**1.503**	**20.936%**	**0.990**	**0.508%**

Figure 2 presents the hourly evolutions of RMSE for products CFSv2, U-Net, and our TaNet. The lower curve indicates the smaller RMSE (the smaller, the better). It can be seen that all these products exhibit relatively more stable and smaller RMSE values in July 2020 (as shown in Figure 2c) than in January, April, and October. We speculate that this observation can be attributed to the influence of hot weather conditions in July, which results in higher temperatures and greater atmospheric stability. Still, CFSv2 (represented by the blue curves) achieves the largest RMSE in July, with values around 2.2°C. In addition, CFSv2 also shows the highest RMSE in almost all periods of January and July 2020 compared with the other products. In contrast, our proposed

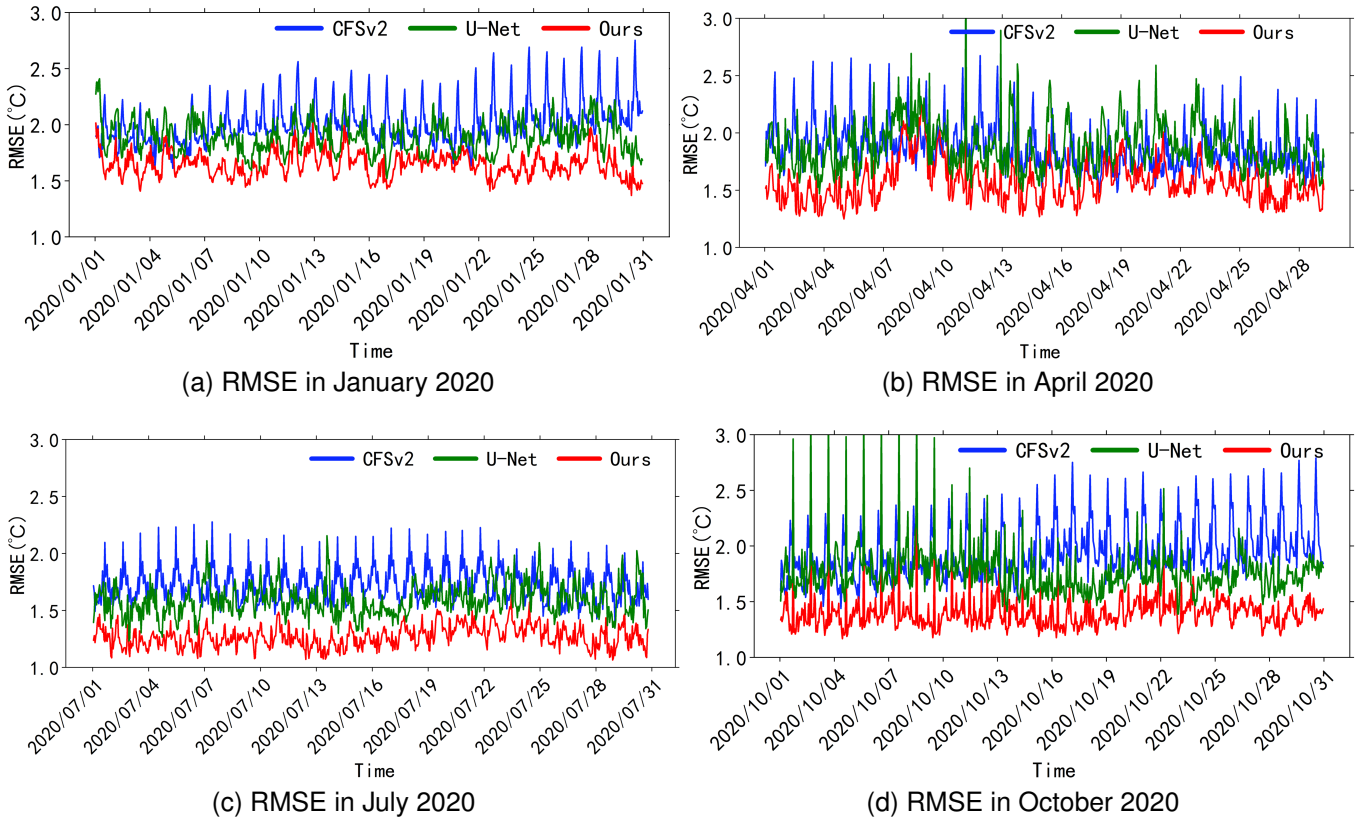

Fig. 2. Hourly evolutions of RMSE of CFSv2, U-Net, and TaNet (Ours) during the verification period (compared to the ERA5 benchmark).

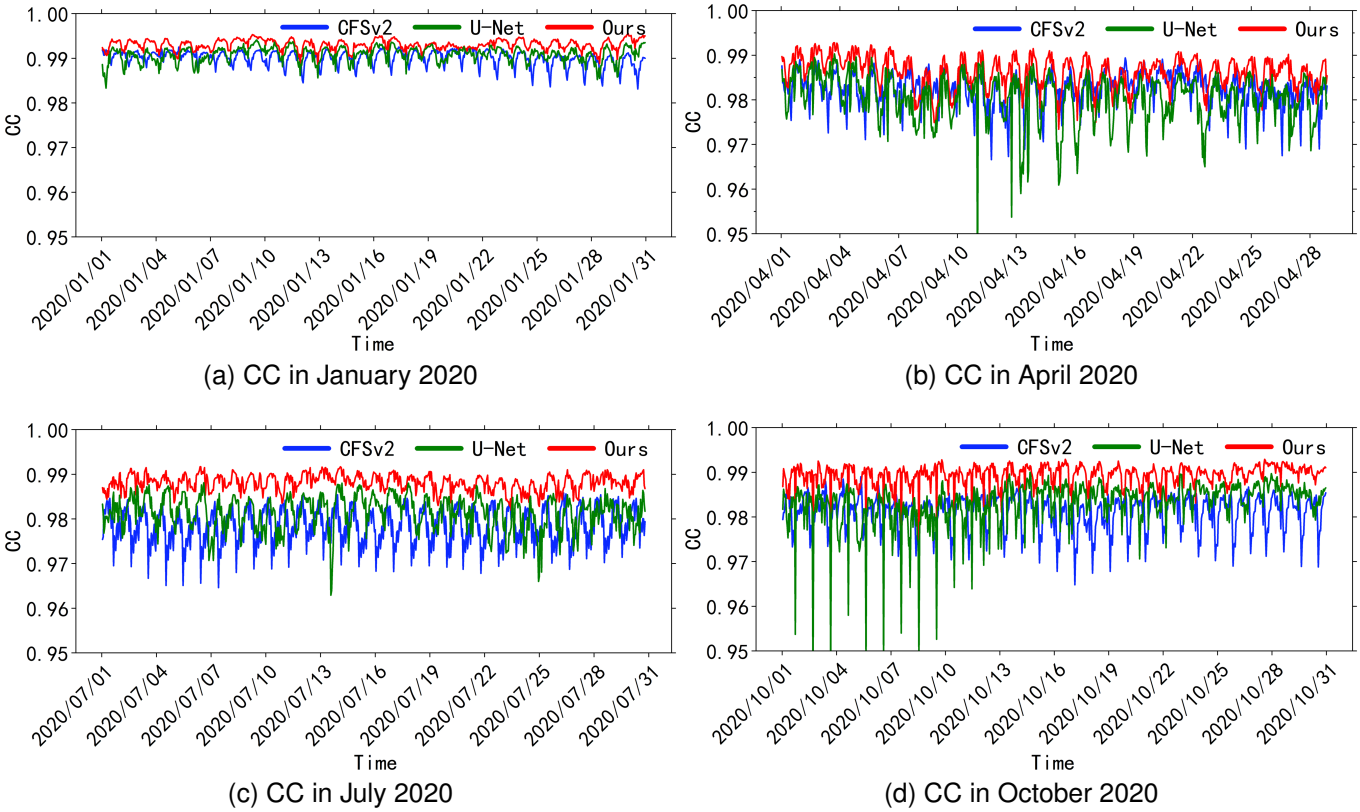

Fig. 3. Hourly evolutions of Pearson's correlation coefficient for CFSv2, U-Net, and Our TaNet during the verification period (with the ERA5 data).

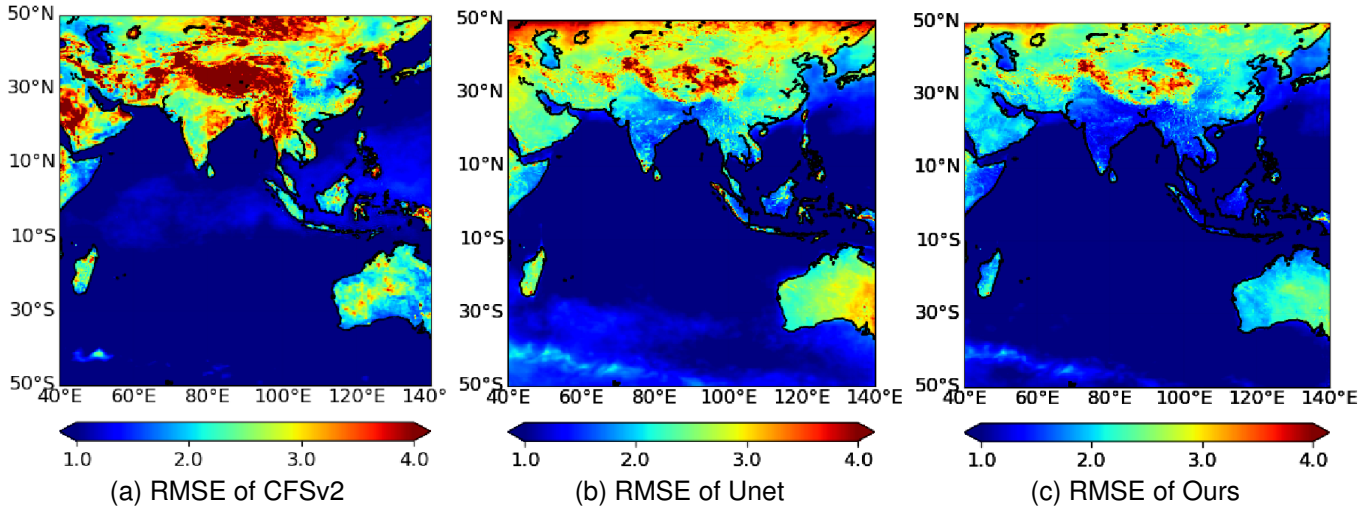

Fig. 4. Spatial distribution of RMSE for products CFSv2, U-Net, and our TaNet during the verification period (compared to the ERA5 benchmark).

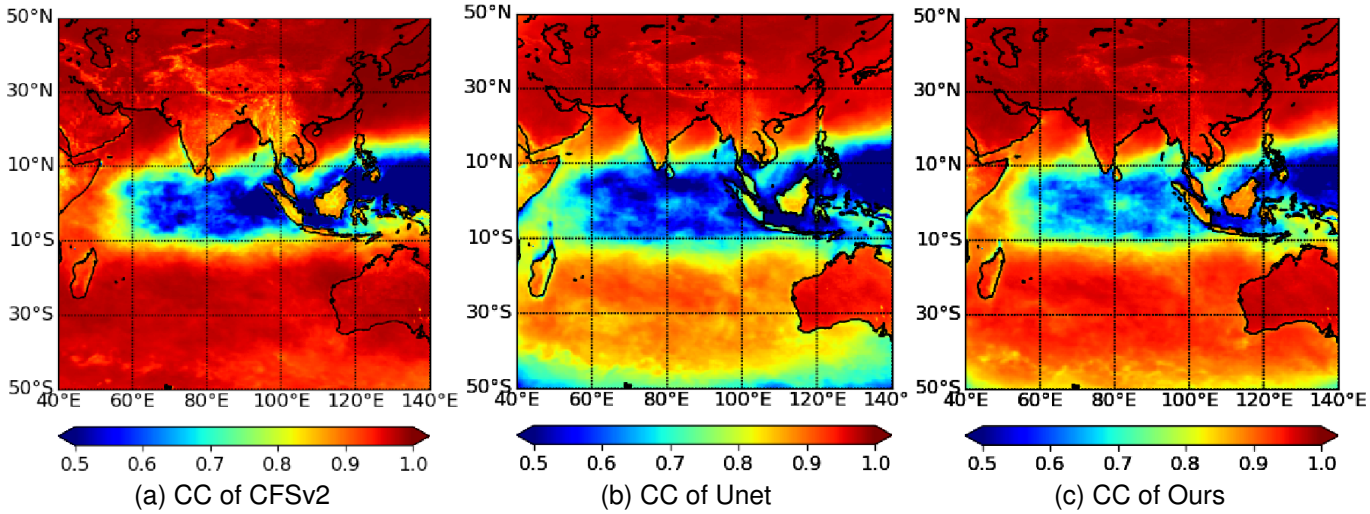

Fig. 5. Spatial distribution of Pearson's correlation coefficient (CC) for CFSv2, U-Net, and TaNet (ours) during the verification period (with the ERA5 benchmark).

TaNet (represented by the red curves) exhibits relatively lowest RMSE values throughout the entire verification period, with a significant performance gain observed in July. These results demonstrate the effectiveness and superiority of our proposed TaNet method.

Furthermore, Figure 3 presents the hourly evolutions of Pearson's correlation coefficient for products CFSv2, U-Net, and our method TaNet. The higher curve reflects the more significant correlation (the larger, the better) with ERA5 data. It is evident that the curves in Figure 3a are more stable and closer to 1.00 than curves in Figure 3b, 3c, and 3d. This observation suggests that meteorological conditions and seasonal fluctuations may have some influence on the performance of all these methods, to a certain extent. Nevertheless, our proposed TaNet method (represented by the red curves) consistently outperforms both CFSv2 (represented by the blue curves) and U-Net (represented by the green curves) throughout the verification period. These results demonstrate the remarkable stability and exceptional performance of our approach.

Figure 4 shows the spatial distribution of RMSE for CFSv2, U-Net, and our TaNet in the study area. The warm color indicates a large RMSE, while the cold color indicates a small RMSE. We can see that the colormap displayed in Figure 4a are overall warmer than that in Figure 4b and Figure 4c, which reveals the relatively larger and more widely distributed RMSE of the CFSv2 product compared to the other two methods. Furthermore, Figure 4b shows that the RMSE in the mid-high latitude region (e.g., 50°N) is higher than in other areas. In contrast, the colormap generated by our TaNet framework, as shown in Figure 4c, is relatively the coolest color. That is especially true for regions with challenging climate conditions, such as the mid-high latitude region and the Qinghai-Tibet Plateau. These results suggest the promising performance of our TaNet method in reconstructing near-surface air temperature.

Figure 5 illustrates the spatial distribution of Pearson's

TABLE III
COMPARISON RESULTS ON RMSE AND PEARSON'S CORRELATION COEFFICIENT (CC) USING STATION OBSERVATIONS AS THE BENCHMARK (THE UP ARROW INDICATES THAT A LARGER VALUE IS BETTER, WHILE THE DOWN ARROW SIGNIFIES THAT A SMALLER VALUE IS BETTER; BOLD VALUES INDICATE THE BEST PERFORMANCE, AND VALUES WITH UNDERLINES INDICATE THE SECOND-BEST PERFORMANCE).

Methods/Products	RMSE (°C) ↓	CC ↑
CFSv2	2.900	0.957
U-Net	2.746	0.954
CRA	2.654	0.955
ERA5	**2.249**	**0.969**
TaNet (ours)	<u>2.594</u>	<u>0.959</u>

correlation coefficient for CFSv2, U-Net, and our TaNet within the study area. The warmer colors indicate a higher correlation with the ERA5 benchmark data. It is worth noting that in the equatorial ocean region between 10°S and 10°N, all three products exhibit lower correlation coefficients compared to other regions. Still, our method (as shown in Figure 5c) shows an advantage over the CFSv2 and U-Net products.

C. Comparisons based on Station Observations

This section further evaluates the proposed method using station observations as the benchmark. We choose four products as the baselines for comparison: ERA5, CRA, CFSv2, and U-Net. As presented in Table III, ERA5 obtains the results with an RMSE of 2.249°C and a correlation coefficient of 0.969, outperforming all other products on both metrics. Furthermore, among the other products (except for ERA5), our proposed method TaNet achieves the closest performance to ERA5. TaNet outperforms CRA, CFSv2, and U-Net, reducing the RMSE of CFSv2 (2.900°C), CRA(2.654°C), and U-Net(2.746°C) by 10.551% (CFSv2), 2.261% (CRA), and 5.535% (U-Net), respectively. Meanwhile, it improves the correlation coefficient of CFSv2, CRA, and U-Net by 0.209%, 0.419%, and 0.524%, respectively. These results validate the effectiveness and promising performance of our proposed TaNet method.

Figure 6 shows the hourly evolution of the average RMSE of all compared methods during the verification period. We can see that the purple line is always the lowest, which indicates the relatively best performance of ERA5. The RMSE performance of our method TaNet (represented by the red lines) is lower than that of U-Net (denoted by the light blues) almost all the time, demonstrating the improvement of our TaNet network over the U-Net baseline network. Especially in July, our method continuously outperforms CFSv2, CRA, and U-Net methods. These results indicate the superiority of our proposed method.

Figure 7 shows the hourly evolution of Pearson's correlation coefficient (CC) during the verification period. In detail, the curves presented in Figure 7a are significantly closer to the upper bound line (maximum cc value of 1.00) than the corresponding ones in Figure 7b-7d, revealing the relatively better performance of these products in January 2020 than that in April, July, and October. Remarkably, each product exhibits a correlation coefficient with the station observations larger than 0.975. However, these products' correlation coefficients were relatively low in July 2020 (as shown to be furthest from the upper bound). Still, our method TaNet outperforms CFSv2, CRA, and U-Net noticeably in July 2020.

Figure 8 illustrates the spatial distribution of RMSE for ERA5, CRA, CFSv2, U-Net, and our method TaNet in the study area. The color represents RMSE with station observations, and the colder the color, the smaller the RMSE. It shows in Figure 8 that the color of each product in the mid-high latitude area is warmer than that in the other areas, which indicates the larger RMSE of all products in the mid-high latitude area. Still, our method TaNet shows promising performance. The RMSE of TaNet is lower (colder color) than that of CFSv2, CRA, and U-Net.

Figure 9 summarizes the spatial distribution of Pearson's correlation coefficient of ERA5, CRA, CFSv2, U-Net, and our method TaNet in the study area. Each color map denotes the correlation coefficient with station observation in the study area, with warmer colors indicating stronger correlations. It shows that the correlation coefficient of CFSv2, CRA, and U-Net in India and Myanmar are relatively lower (indicated by cooler colors) than other areas. In contrast, our method (as shown in Figure 9e) exhibits relatively better than CFSv2, CRA, and U-Net.

V. DISCUSSION

In this work, we propose a deep-learning neural network TaNet. It estimates near-surface air temperature in near real-time using FY-4A satellite data, providing timely and accurate near-surface air temperature data for regions lacking station observations. Two kinds of experiments are conducted to verify the performance of TaNet, using the operational product ERA5 and station observations as the benchmark standard, respectively. Based on the experimental results, our discussions are as follows:

Estimating near-surface air temperature from FY-4A satellite imagery based on deep learning is effective. As the three most common meteorological measurements, satellite remote sensing, ground-based radar, and station observations can reflect the same weather condition from different perspectives (corresponding to different data sources). So, there must be certain relationships between the three meteorological data sources. It is possible to model the relationships between the three meteorological data sources with deep learning networks and thus estimate anyone source data using the data from another or two sources, as deep learning networks are famous for approximating arbitrary functions with arbitrary accuracy. Extensive experiments conducted in this paper demonstrate the promising performance of the two deep learning models, U-Net and our proposed TaNet. Especially, TaNet achieves 0.990 and 0.959 on Pearson's correlation coefficient with the operational product ERA5 and station observations, respectively, outperforming the operational products CFSv2 and CRA.

Furthermore, choosing an appropriate deep neural network helps to improve performance. The backbone of our framework TaNet is an improvement of U-Net, by replacing the skip connections in U-Net with a channel-wise cross-fusion transformer (CCT) module and a channel-wise cross-attention (CCA) module. The CCT module captures global

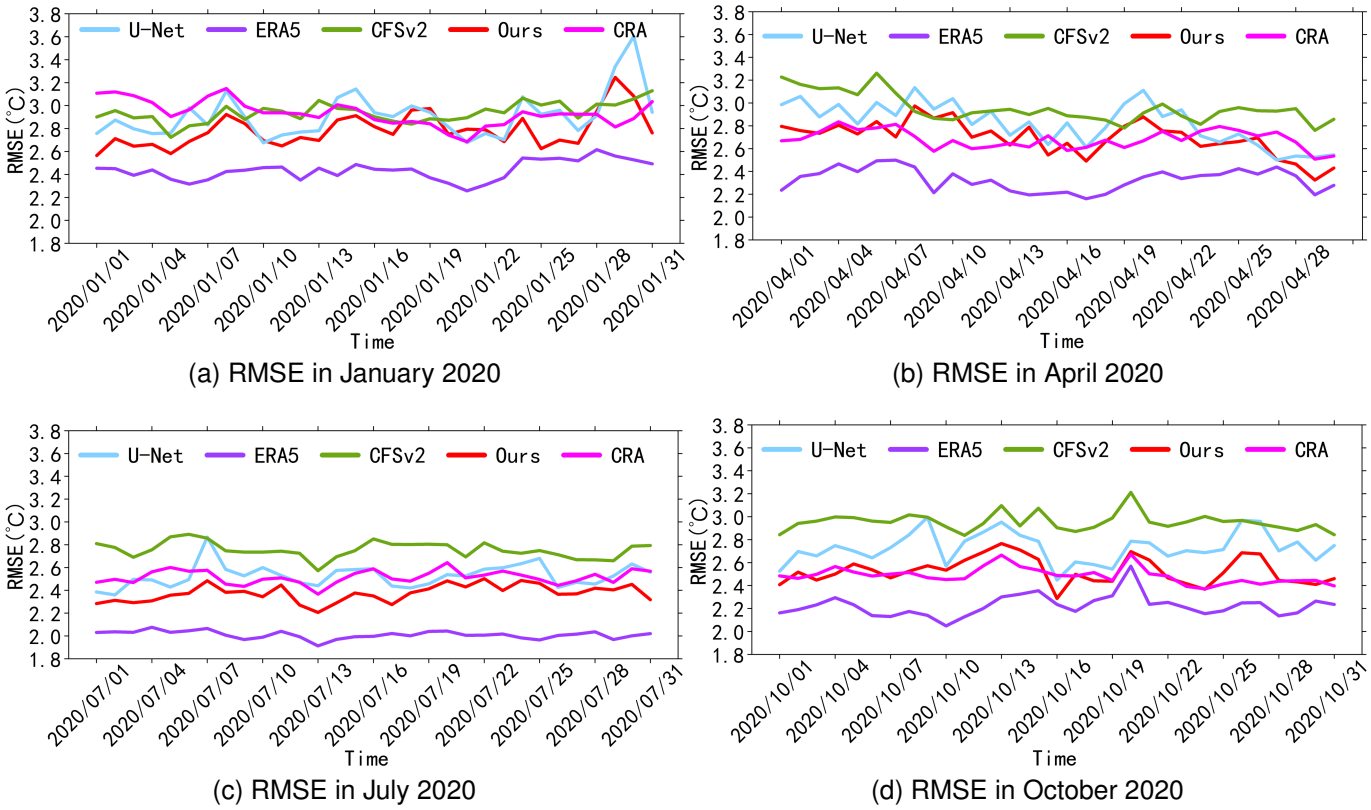

Fig. 6. Evolution of the average RMSE for products U-Net, ERA5, CFSv2, CRA, and our TaNet, compared to the station observations during the verification period.

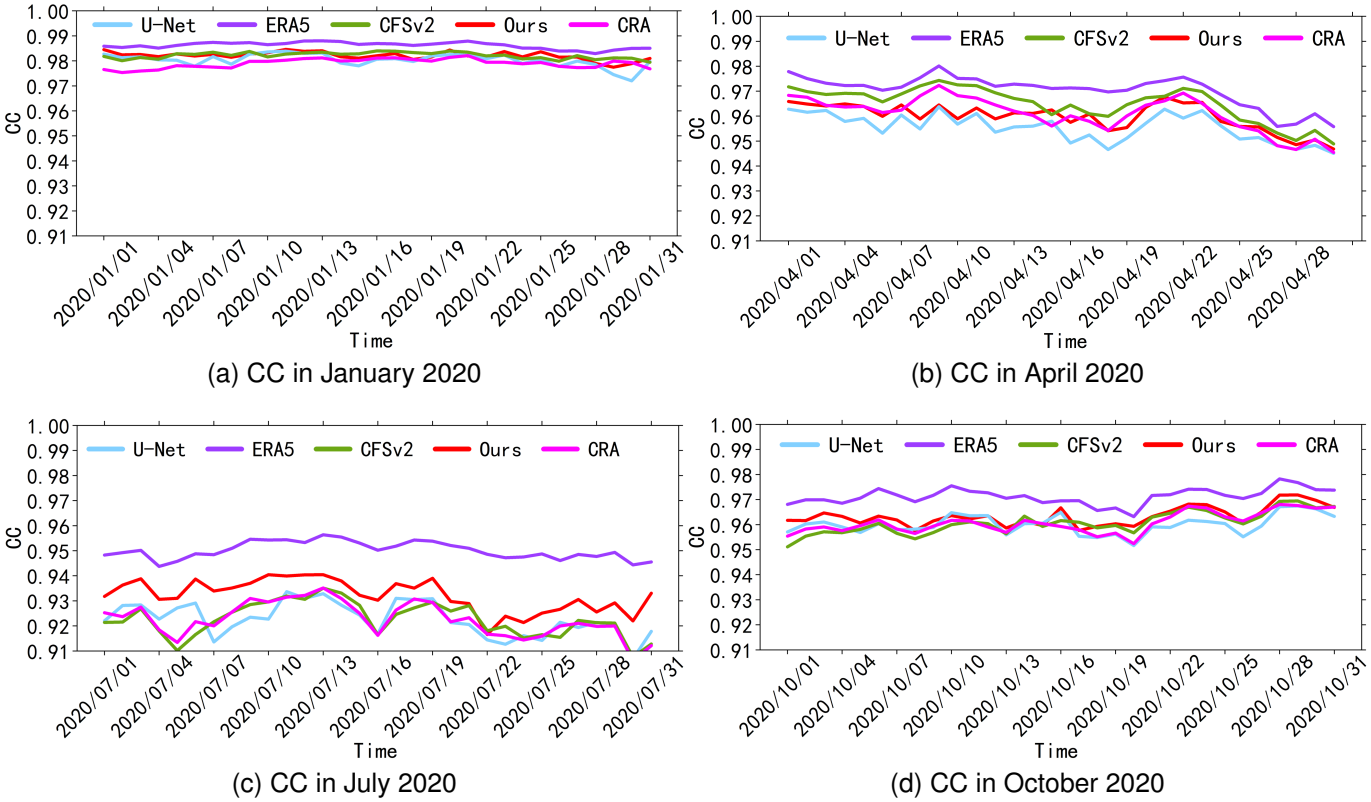

Fig. 7. Evolution of Pearson's correlation coefficient (CC) for products U-Net, ERA5, CFSv2, CRA, and our TaNet, with the station observations during the verification period.

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2023.3322343

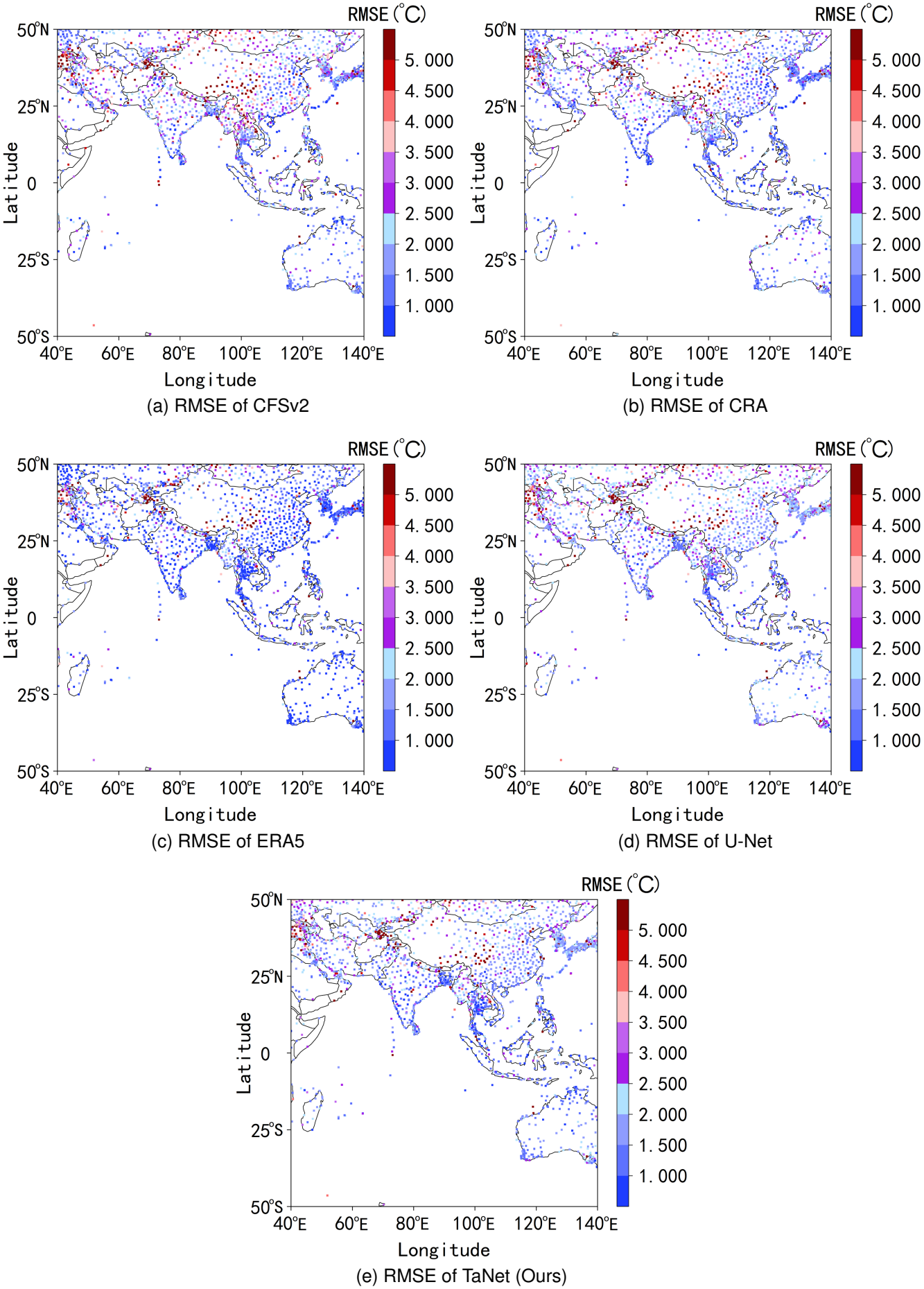

Fig. 8. Spatial distribution of RMSE of ERA5, CRA, CFSv2, U-Net, and our method TaNet, compared to station observations in the study area.

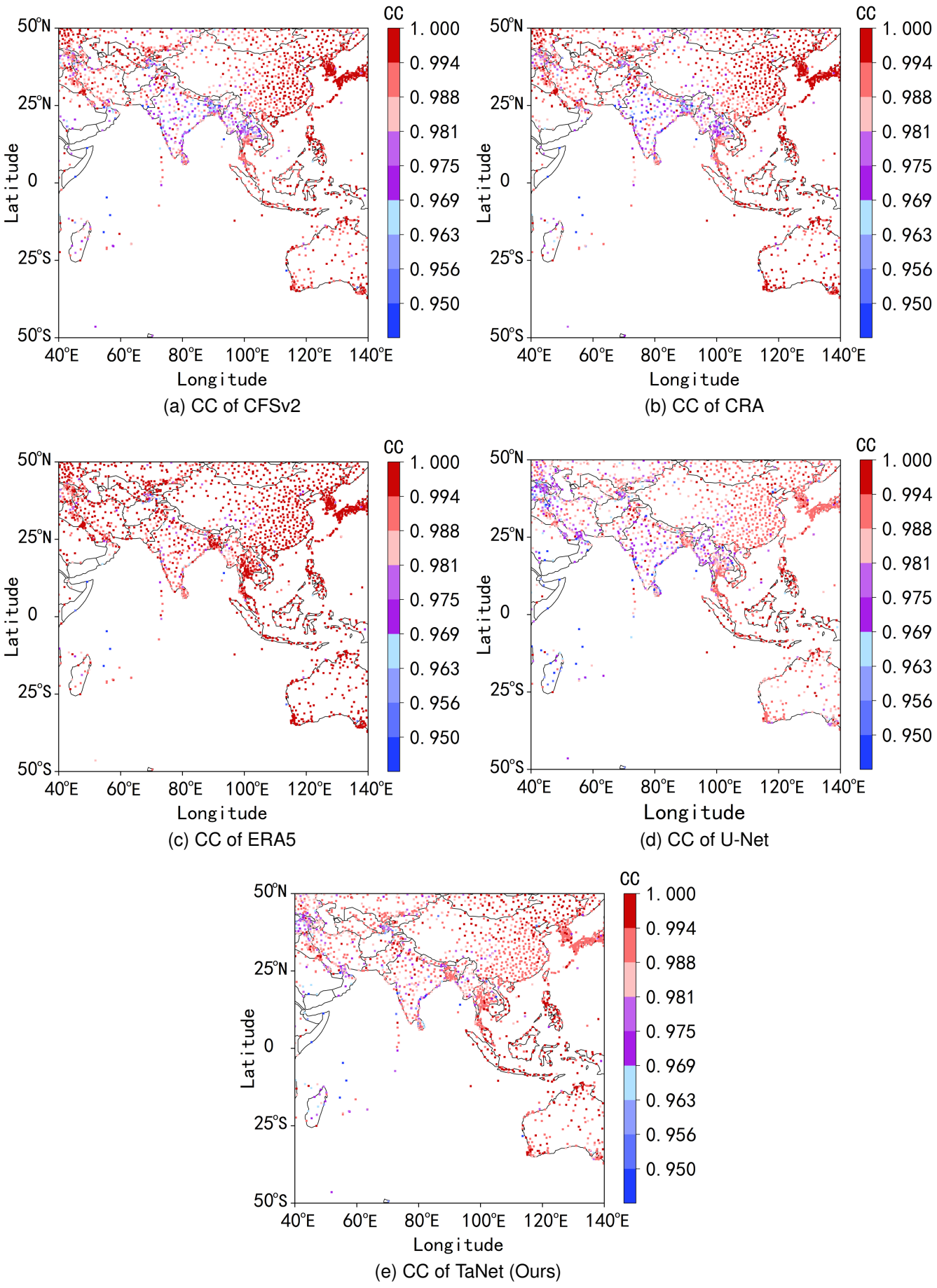

Fig. 9. Spatial distribution of Pearson's correlation coefficient of ERA5, CRA, CFSv2, U-Net, and our method TaNet, with station observations in the study area.

and local spatial relationships and helps refine multi-scale features learned at the encoding stages. The CCA module connects these refined features with features learned at the decoding stages, eliminating the potential semantic inconsistency between them. Jointly, the two modules contribute to the performance improvement of TaNet over U-Net. Trained and tested in the same way, the performance gain of TaNet is 6.915% (1.503°C vs. 1.809°C) in RMSE and 0.406% (0.990 vs. 0.986) in Pearson's correlation coefficient relative to U-Net, with ERA5 as the benchmark standard. These results demonstrate the effectiveness and superiority of our method TaNet.

Lastly, it is important to note that meteorological conditions, seasonal fluctuations, and geographic variations may influence the performance of near-surface air temperature estimation. Through extensive comparative experimental analysis, we discover that RMSE values and Pearson's correlation coefficient of all methods vary across different months/seasons and regions. For instance, the RMSE values in July 2020 are generally smaller and more stable (as exhibited in Figures 2 and 6), while the CC values in January 2020 are notably more excellent and more stable (as shown in Figures 3 and 7). Moreover, RMSE and CC values of mid to high-latitude regions often exhibit more diverse and unstable variations (as shown in Figures 4, 5, 8, and 9). Considering these findings, it is crucial for future work to incorporate factors such as latitude, longitude, topography, and climate into the estimation process, to further improve the accuracy and reliability of near-surface air temperature estimation.

VI. Conclusion

Near-surface air temperature estimation from satellite data provides a valuable way to fill the data shortage problem of ground-based meteorological station observation and radar observation under complex geographical and natural conditions. In this paper, we propose a deep learning framework TaNet. It automatically extracts information from FY-4A satellite imagery (the twelve and thirteen bands) and estimates the near-surface air temperature in near real-time. Specifically, TaNet is a U-shaped encoder-decoder framework. It adaptively refines multi-scale features learned at encoding stages using a transformer module, then fuses these refined features with those learned at decoding stages with an attention mechanism. Extensive experiments and visualizations demonstrate the effectiveness and superiority of the proposed TaNet method. It helps address the station observation lacking problem. In future work, we will introduce more meteorological data and incorporate additional factors such as latitude, longitude, topography, and climate to improve the accuracy and robustness of TaNet and test it over more prolonged periods.

VII. Acknowledgments

This research was funded by the China Meteorological Administration through the FengYun Application Pioneering Project (No. FY-APP-2022.0609), the Scientific Research Fund of Chengdu University of Information Technology (No. KYTZ202157), the National Key Research and Development Program of China (No. 2020YFA0608001), the National Science Foundation of China (No. 42075142), the Sichuan Science and Technology Program (No.2022YFG0042), the Opening Foundation of Agile and Intelligent Computing Key Laboratory of Sichuan Province.

References

[1] K. Zhou, H. Liu, X. Deng, H. Wang, and S. Zhang, "Comparison of machine-learning algorithms for near-surface air-temperature estimation from fy-4a agri data," *Advances in Meteorology*, vol. 2020, pp. 1–14, 2020.

[2] B. Du, K. Mao, S. M. Bateni, F. Meng, X.-M. Wang, Z. Guo, C. Jun, and G. Du, "A novel fully coupled physical statistical deep learning method for retrieving near-surface air temperature from multisource data," *Remote Sensing*, vol. 14, no. 22, p. 5812, 2022.

[3] Y. Gao, J. Guan, F. Zhang, X. Wang, and Z. Long, "Attention-unet-based near-real-time precipitation estimation from fengyun-4a satellite imageries," *Remote Sensing*, vol. 14, no. 12, p. 2925, 2022.

[4] Y. Tao, X. Gao, K. Hsu, S. Sorooshian, and A. Ihler, "A deep neural network modeling framework to reduce bias in satellite precipitation products," *Journal of Hydrometeorology*, vol. 17, no. 3, pp. 931–945, 2016.

[5] M. Sadeghi, P. Nguyen, K. Hsu, and S. Sorooshian, "Improving near real-time precipitation estimation using a u-net convolutional neural network and geographical information," *Environmental Modelling & Software*, vol. 134, p. 104856, 2020.

[6] Y. F. Hu, F. K. Yin, and W. M. Zhang, "Deep learning-based precipitation bias correction approach for yin–he global spectral model," *Meteorological Applications*, vol. 28, no. 5, 2021.

[7] S. Jiang, Y. Huang, and F. Zhang, "Deep learning reconstruction method of meteorological radar echo data based on satellite data," in *Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies*, 2020, pp. 70–74.

[8] H. Wang, K. Mao, Z. Yuan, J. Shi, M. Cao, Z. Qin, S. Duan, and B. Tang, "A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning," *Remote Sensing of Environment*, vol. 265, p. 112665, 2021.

[9] C. Guo, W. Ai, X. Zhang, Y. Guan, Y. Liu, S. Hu, and X. Zhao, "Correction of sea surface wind speed based on sar rainfall grade classification using convolutional neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 321–328, 2022.

[10] Y. Liu, S. Ortega-Farías, F. Tian, S. Wang, and S. Li, "Estimation of surface and near-surface air temperatures in arid northwest china using landsat satellite images," *Frontiers in Environmental Science*, vol. 9, p. 609, 2021.

[11] C. Wang, X. Bi, Q. Luan, and Z. Li, "Estimation of daily and instantaneous near-surface air temperature from modis data using machine learning methods in the jingjinji area of china," *Remote Sensing*, vol. 14, no. 8, p. 1916, 2022.

[12] B. Zhou, E. Erell, I. Hough, J. Rosenblatt, A. C. Just, V. Novack, and I. Kloog, "Estimating near-surface air temperature across israel using a machine learning based hybrid approach," *International Journal of Climatology*, vol. 40, no. 14, pp. 6106–6121, 2020.

[13] H. Shen, Y. Jiang, T. Li, Q. Cheng, C. Zeng, and L. Zhang, "Deep learning-based air temperature mapping by fusing remote sensing, station, simulation and socioeconomic data," *Remote Sensing of Environment*, vol. 240, p. 111692, 2020.

[14] H. Ding, L. Zhao, S. Liu, X. Chen, G. de Leeuw, F. Wang, F. Zheng, Y. Zhang, J. Liu, J. Li, L. She, Y. Si, and X. Gu, "Fy-4a/agri aerosol optical depth retrieval capability test and validation based on nnaerog," *Remote Sensing*, vol. 14, no. 21, p. 5591, 2022.

[15] H. Qiao, P. Zhang, Z. Li, and C. Liu, "A new geostationary satellite-based snow cover recognition method for fy-4a agri," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 372–11 385, 2021.

[16] J. Yang, Z. Zhang, C. Wei, F. Lu, and Q. Guo, "Introducing the new generation of chinese geostationary weather satellites, fengyun-4," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1637–1658, 2017.

[17] Y. Jiang, W. Cheng, F. Gao, S. Zhang, S. Wang, C. Liu, and J. Liu, "A cloud classification method based on a convolutional neural network for fy-4a satellites," *Remote Sensing*, vol. 14, no. 10, p. 2314, 2022.

[18] B. Zhong, Y. Ma, A. Yang, and J. Wu, "Radiometric performance evaluation of fy-4a/agri based on aqua/modis," *Sensors*, vol. 21, no. 5, p. 1859, 2021.
[19] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, "The era5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.
[20] "Integration, quality assurance, and usage of global aircraft observations in cra," *Journal of Meteorological Research*, vol. 35, no. 1, pp. 1–16, 2021.
[21] S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer *et al.*, "The ncep climate forecast system reanalysis," *Bulletin of the American Meteorological Society*, vol. 91, no. 8, pp. 1015–1058, 2010.
[22] S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H.-y. Chuang, M. Iredell *et al.*, "The ncep climate forecast system version 2," *Journal of climate*, vol. 27, no. 6, pp. 2185–2208, 2014.
[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.
[24] K. Trebing, T. Stanczyk, and S. Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," *Pattern Recognition Letters*, vol. 145, pp. 178–186, 2021.
[25] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, 2022, pp. 2441–2449.
[26] H. Robbins and S. Monro, "A stochastic approximation method," *The annals of mathematical statistics*, pp. 400–407, 1951.