Deriving Accurate Surface Meteorological States at Arbitrary Locations via Observation-Guided Continous Neural Field Modeling

Zili Liu* *Graduate Student Member, IEEE*, Hao Chen†, Lei Bai, Wenyuan Li, Keyan Chen *Graduate Student Member, IEEE*, Zhengyi Wang, Wanli Ouyang, Zhengxia Zou *Senior Member, IEEE* and Zhenwei Shi† *Senior Member, IEEE*

***Abstract*—Accurately retrieving surface meteorological states at arbitrary locations is of great application significance in weather forecasting and climate modeling. Since meteorological variables are typically provided as coarse-resolution gridded fields, common methods obtain the states at a specific location directly through spatial interpolation can lead to significant accuracy deviations compared to actual observations. Traditional downscaling, the process of obtaining fixed-scale high-resolution meteorological fields from low-resolution inputs, has been proposed as a way to indirectly improve the accuracy of retrieving states at arbitrary locations by providing more detailed subgrid-scale information. However, for arbitrary locations at the station scale, their states are influenced by sub-grid information, resulting in systematic biases between the downscaled results after interpolation and the actual observations at specific station locations. To address this issue, in this paper, we propose a new task called *Station-Scale Downscaling*, which aims to directly derive accurate meteorological states at any given station location from a coarse-resolution meteorological field. To achieve this, we propose a new downscaling model based on hypernetwork architecture, namely *HyperDS*, which efficiently integrates the multi-scale observational information to guide the continuous neural field modeling of the meteorological variables, enabling accurate sampling of the states at any target location. Through extensive experiments, our proposed method outperforms other specially designed baseline models on multiple surface variables. Notably, the mean squared error (MSE) for wind speed and surface pressure improved by 67% and 19.5% compared to other methods.**

***Index Terms*—Meteorological field downscaling, remote sensing, hypernetworks, earth observation.**

The work was supported by the National Natural Science Foundation of China under Grants 62125102, the National Key Research and Development Program of China (Grant No. 2022ZD0160401), the Beijing Natural Science Foundation under Grant JL23005, and the Fundamental Research Funds for the Central Universities, the National Key Research and Development Program of China(Grant No.2022ZD0160101). †Corresponding author: Zhenwei Shi and Hao Chen (e-mail:shizhenwei@buaa.edu.cn, chenhao1@pjlab.org.cn)

Zili Liu, Keyan Chen and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China, and also with the Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Hao Chen, Lei Bai, and Wanli Ouyang are with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Wenyuan Li is with the Department of Geography, University of Hong Kong, Hong Kong, China.

Zhengyi Wang is with the School of Oceanography, Shanghai Jiao Tong University, Shanghai 200030, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Zhengxia Zou is with the Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

*This work was done during his internship at Shanghai Artificial Intelligence Laboratory.

I. Introduction

In recent years, the application of artificial intelligence technologies such as deep learning in weather forecasting has garnered significant attention [1]–[3]. These efforts utilize large-scale gridded historical meteorological field data, combined with advanced models from computer vision, and have demonstrated powerful performance in many forecasting tasks [4]–[7], even surpassing the long-developed numerical weather prediction systems [8]. In addition to weather forecasting, these advanced models have also begun to be extensively studied in other important meteorological tasks such as downscaling [9], data assimilation [10], and satellite data retrieval [11]. These methods have achieved remarkable research results and have started to be applied in practice. Nevertheless, the reliance on methodologies from the field of computer vision has led to the acquisition of image-like meteorological field data in the form of fixed and relatively coarse resolution. This approach is at odds with the intrinsically multi-scale nature of meteorological variables. Furthermore, in practical applications, we typically aim to obtain the meteorological state at any given scattered location rather than over a broad regional area. Therefore, common methods usually use spatial interpolation methods (bilinear, etc.) to obtain the meteorological state of the target location from the grid meteorological field data. This process leads to significant systematic bias, and since the entire process is unlearnable, it is difficult to model the recovery of complex subgrid information. Consequently, as an indirect solution, *downscaling* has become an indispensable post-processing task within operational forecasting to obtain meteorological variables at finer scales and resolutions [3], [12].

The objective of downscaling in weather forecasting is typically to map coarse-resolution global-scale meteorological fields to high-resolution regional-scale fields [9]. This setup appears to be highly analogous to the task of image super-resolution in computer vision [13]. The traditional dynamic downscaling methods [14] are akin to numerical weather prediction techniques, involving the numerical solution of

atmospheric differential equations at a regional scale. This process is highly computationally demanding, particularly when the grid resolution is very high. As a result, statistic downscaling methods, primarily deep learning-based ones, have recently received increased attention as a parallel approach [3], [9], [12], [15]–[32]. Most previous deep learning-based downscaling works directly employ models and methods from image super-resolution tasks due to the similarity between the two tasks. Meteorological field data is treated as an image to achieve super-resolution at a fixed resolution and fixed upscaling factor, which is direct and efficient.

Unlike natural images that are directly captured through camera sensors, gridded meteorological field data are typically obtained by fusing and assimilating multi-source, multi-scale, and multi-modal observational and forecast data, commonly known as *analysis* data or *reanalysis* data. The observational information employed generally includes satellite remote sensing images, ground observation stations, radiosondes, and so on. A specific meteorological variable at a particular pixel can be considered as the average of all observed and predicted values within that pixel area. For instance, the widely used ERA-5 reanalysis data [33] have a temporal resolution of 1 hour and a spatial resolution of $0.25°$. The state value of any given pixel can be regarded as the average of all observations and forecast results within the $0.25° \times 0.25°$ grid over 1 hour. The same applies to forecast fields derived from analysis and reanalysis data. As a result, for meteorological fields with fixed resolution, although each grid can be considered as the average of all observations, many sub-grid observations cannot be effectively represented.

However, in practical applications, we aim to obtain the meteorological state at a specific precise location through the given gridded meteorological field, rather than merely obtaining the high-resolution gridded meteorological fields. The absence of sub-grid information results in a significant deviation between the state values of the meteorological field and the scattered stations, when interpolating directly from the grid meteorological field to the stations [34]. Therefore, for the downscaling of meteorological variables, it is crucial to recover information at the sub-grid scale. A straightforward method of recovering sub-grid information is to use high-resolution real-time observational data with multi-scale resolution to guide the downscaling task. Most of the existing downscaling frameworks based on deep learning are inspired by super-resolution tasks. They achieve downscaling solely by learning the mapping from low-resolution to high-resolution images, which does not allow for the integration of multi-scale observations into the model's training and inference processes. With its fixed super-resolution factor, the resulting downscaled output is also gridded. Therefore, to obtain the meteorological state at an arbitrary station location, further non-learnable interpolation is required, which can result in a significant loss of precision (as shown in Fig 1(a)). Thus, establishing a reasonable downscaling task specifically tailored to the meteorological field is of critical importance, especially with the ability to obtain accurate meteorological conditions at any location.

To address the aforementioned issues, our paper first constructs a new task for downscaling meteorological fields at station scale, guided by multi-source, multi-scale, and multi-modal observational data. Our goal is to downscale low-resolution meteorological fields to the scale of arbitrary scattered points. Specifically, the paper selects ERA5 reanalysis data [33] as the meteorological field data to be downscaled. For observational data, we utilize remote sensing images from the new generation geostationary meteorological satellite Himawari-8 (H8) [35] at L1-level as high-resolution gridded-scale indirect observational data. Additionally, we employ meteorological observation station data obtained from the Weather2K dataset [36] as direct observational data at the scattered station scale. This task setting is crucial for the downscaling of meteorological variables because it allows for the utilization of multi-scale observational information, and it can yield downscaling results that are adaptable to any given location by sampling from the learned continuous neural field (as shown in Fig. 1(b)).

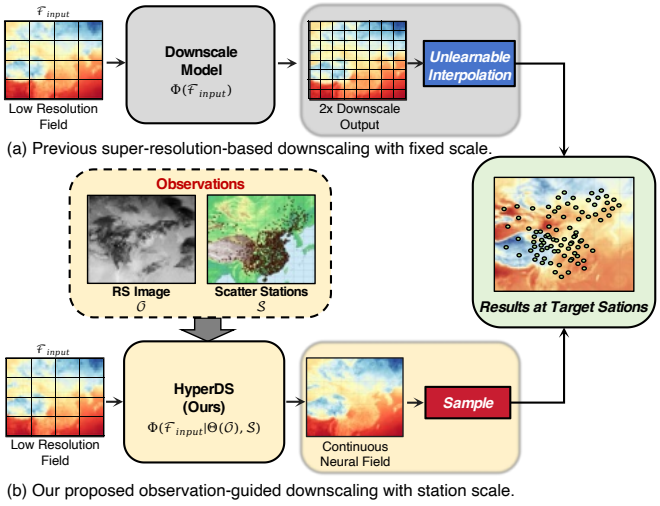


Fig. 1. The difference between the previous SR (super-resolution)-based downscaling pipeline with fixed grid-level scale [9] (a), and the proposed observation-guided downscaling pipeline with arbitrary scattered station-level scale (b). $\Phi(\cdot)$ is the downscale model, and $\Theta(\cdot)$ is the observational operator that maps the remote sensing images into the meteorological state domain. To obtain the meteorological status at target stations, the former method requires interpolation of the downscaled meteorological field, which is non-learnable. The latter, on the other hand, directly learns a continuous neural representation of the meteorological field, thereby directly obtaining the meteorological states of the target stations through sampling.

In response to the established novel task, inspired by the ability of implicit neural representation and neural field methods [37], [38] in computer vision to continuously model two-dimensional and three-dimensional data, and combined with a data-conditioned hypernetworks structure [39], we propose a novel model for the continuous resolution downscaling of meteorological fields, named *HyperDS*. *HyperDS* uses H8 observations as the additional conditional input and takes Weather2K station data as supervision at the scattered station scale. The overall architecture of *HyperDS* can be divided into a *dual-branch hypernetwork* and a *target network*. The former consists of two encoders based on convolutional neural networks that are used to extract semantic features from the low-resolution meteorological field and the relatively high-

resolution H8 data, respectively. Following these, an *implicit retrieval model* employs a cross-attention mechanism to implicitly learn the retrieval process from satellite imagery to meteorological fields, thereby efficiently integrating H8 data into the downscaling process. This results in the generation of high-level feature vectors that contain fused information. The target network is based on a multilayer perceptron (MLP), whose weights are generated from the fused features output by the hypernetwork. It achieves downscaling at arbitrary locations by sampling from the learned continuous neural field which inputs the coordinate values of the target location and obtains the corresponding meteorological state variables. We have also devised a training technique utilizing sub-grid sampling, and in combination with supervisory data from observational stations, it effectively reconstructs accurate state values for meteorological variables at the sub-grid level.

We established several baseline methods and compared them with the proposed method under the condition of identical input and supervision data. *HyperDS* shows superior downscaling performance at the scattered station scale. Additionally, through more extensive analysis and ablation experiments, we also verified the importance of incorporating observational data for the task of station-scale downscaling. We hope that more researchers in the field will engage in further studies on this new task, aiming to achieve more efficient continuous-resolution modeling of meteorological fields and more effective integration of observational data. To summarize, the main innovative contributions of this paper include the following three points:

- Given the systematic biases between surface gridded meteorological fields and the actual observed meteorological states at station locations, we propose a new task called *Station-scale Downscaling*, which aims to directly derive accurate meteorological states at arbitrary locations from coarse-resolution meteorological fields.
- Based on this new task, we propose a novel model structure called *HyperDS*, based on a data-conditioned hypernetwork model. We designed a dual-branch feature encoder and an implicit retrieval network tailored to the task characteristics and developed a multi-scale loss function, which achieves continuous modeling and station-scale downscaling of meteorological fields by efficiently integrating the multi-scale observational information from remote sensing images and in-situ observations.
- Through the design of fair baseline models and extensive experiments, we have validated the effectiveness of the proposed new model, which significantly outperforms comparative methods at the scale of scattered stations.

II. Related Work

In this section, we will briefly introduce and review the work related to the proposed task and model in this paper.

A. Meteorological Field Downscaling

The objective of downscaling meteorological fields is to obtain accurate weather forecast results with fine granularity and high resolution as required [3], [9]. Typically, meteorological forecast data generated by operational global forecasting systems are produced on a relatively coarse-resolution grid. Currently, the highest resolution global forecast and reanalysis data are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), with their operational Integrated Forecasting System (IFS) and ERA-5 analysis [33] with a spatial resolution of 0.25° and temporal resolution of 1 hour. To obtain higher-resolution regional-scale weather states, there have been numerous downscaling efforts in the past. Here, we focus our research on methods based on deep learning.

Due to the similarity between the task of downscaling and the task of super-resolution in the field of computer vision, the vast majority of previous work has been inspired by related efforts in the field of super-resolution. One of the most mainstream approaches is the use of super-resolution networks, such as UNet [40], with an encoder-decoder structure based on CNN [15], [18]–[27] and Transformer [28]. Furthermore, due to the successful application of generative modeling techniques in the field of super-resolution, many previous studies have also applied Generative Adversarial Networks (GANs) and Diffusion models to the task of downscaling meteorological fields, to obtain results with richer texture information [16], [29]–[31]. These studies solely utilize high-resolution meteorological field data for supervision, learning the mapping process from low-resolution meteorological fields to high-resolution ones [12]. However, the aforementioned methods are all direct applications of super-resolution models and do not incorporate special designs tailored to the characteristics of meteorological variables. They merely achieve downscaling results at specific magnifications following high-resolution supervisory data. Some recent works on simulating fluid field data [41] have attempted to achieve grid-independent continuous-resolution downscaling and have integrated physical information as prior [42]. However, the relevant physical information and data are difficult to apply in the context of real-world data. Our recent work DeepPhysiNet [43] bridges physical laws and deep learning for continuous weather modeling on real-world weather data. However, the aforementioned methods do not use sub-grid observational data as an auxiliary, which intuitively could be a very direct way to enhance downscaling information [44]. Incorporating observational information at multiple scales during the training process will also enhance the model's adaptability to various scales, allowing for more accurate recovery of sub-grid meteorological states [45]. Moreover, most studies focus solely on downscaling a single type of meteorological variable, such as temperature or precipitation, and are unable to simultaneously process multiple meteorological variables.

Addressing the issues present in previous downscaling efforts, we have specifically designed a station-scale downscaling task, combined with multi-scale observational data, tailored to the characteristics of meteorological variables. Our new model effectively reconstructs sub-grid states within coarse-resolution meteorological fields. Moreover, we have performed downscaling on five surface variables, which helps in obtaining more comprehensive meteorological state information at the scale of scattered stations.

B. Image Super Resolution

The widespread application of deep learning in meteorological downscaling is inseparable from the rapid development of image super-resolution tasks in the field of computer vision. The following is a brief introduction to the related work in image super-resolution (SR).

The field of image super-resolution (SR) has witnessed a significant transformation with the advent of deep learning techniques [13]. Pioneering work, Super-Resolution Convolutional Neural Network (SRCNN) [46], demonstrated the effectiveness of deep learning for this task. Building on this foundation, many studies have proposed various super-resolution networks based on CNN-based encoder-decoder structures, enhancing the performance of super-resolution tasks [47]–[49]. With the development of foundational models in vision, several super-resolution models based on GANs [50], [51] and Transformers [52], [53] have also been proposed. However, the aforementioned super-resolution models are only capable of achieving super-resolution at fixed magnifications. Inspired by implicit neural representations [37], recent works have begun to explore super-resolution tasks with continuous resolutions [54], [55]. The aim is to achieve super-resolution at any arbitrary position by learning the mapping from coordinates to RGB values. However, super-resolution methods based on implicit neural representations suffer from a lack of sub-grid supervision, resulting in what is called continuous-resolution being merely a more sophisticated form of smoothing.

Besides enhancing the resolution of natural images, there is also an urgent need to improve resolution in the field of remote sensing imagery [56]. Due to their overhead perspective and broader geographic coverage, remote sensing images bear a closer resemblance to meteorological downscaling tasks. Influenced by the super-resolution techniques applied to natural images, the development of remote sensing imagery has also incorporated unique features such as geographic information. Related approaches include methods CNN-based [57], GAN-based [58], and recent implicit representation-based [38], [59], [60] methods.

However, there is scarcely any existing work that has introduced super-resolution methods based on implicit neural representations into the realm of meteorological downscaling. Furthermore, unlike image super-resolution, meteorological variables often include a wealth of sub-grid station observational information, which can better assist models in learning information at continuous locations. Therefore, we incorporate scattered grid observations as auxiliary information into our task, combined with high-resolution remote sensing observations, in the hope of effectively integrating multi-scale observational data to recover sub-grid meteorological states and achieve meteorological downscaling at scattered station scale.

C. Hypernetworks

Hypernetworks [61] are the type of model architecture that utilizes one network (commonly referred to as the hypernetwork) to predict the weights of another network (typically called the target network). Compared to traditional network architectures, hypernetworks offer more flexibility in their structure and input/output modalities. They have been widely applied across various fields such as computer vision, solving differential equations, and uncertainty quantification [39]. Leveraging the hypernetwork structure, the traditional per-sample optimization approach of implicit neural representations can be transformed into a data-conditioned hypernetwork learning architecture. This structure allows for the learning of the target network's parameters based on different input samples, and the design of the target network's input and output according to the requirements of the task. Regarding the application of hypernetworks to downscaling in meteorological fields, to the best of our knowledge, there are currently no similar efforts. Given the characteristics of hypernetwork structures, they are particularly suitable for meteorological data, which often comprises multi-modal data types. Therefore, the *HyperDS* we propose utilizes the hypernetwork model structure and has been specifically designed to cater to the characteristics of the downscaling task.

III. Problem Setting

In this section, we will introduce the proposed novel *Station-Scale Downscaling* task guided by observations specifically designed based on our understanding of meteorological downscaling tasks. This includes a description of the task, the datasets we used, and the evaluation metrics.

A. Observation-Guided Station-Scale Downscaling

1) Task Description: Unlike the objectives of previous downscaling or super-resolution tasks to obtain high-resolution grid data, we focus on capturing the meteorological state at any given scattered station location, which has significant practical value [34]. The results produced by most current meteorological tasks are structured gridded data, which need to be further processed through methods such as interpolation to obtain the state values at the scattered point locations of interest. As a result, simple interpolation without any learnable process creates an inherent bias between the gridded data and the scattered stations (as shown in Fig.1(a)). Therefore, it is crucial to design specialized methods to effectively downscale gridded meteorological fields to scattered points and minimize the inherent bias between them.

To address this issue, inspired by data assimilation [10], [62]–[64], we realize that gridded meteorological field data are obtained through the integration of multi-source, multi-scale observational data, and gridded model forecast result. Therefore, using observational data to guide the downscaling of meteorological fields is a very direct and reasonable approach. This allows us to recover sub-grid information of low-resolution gridded meteorological fields through multi-scale, multi-resolution observational data, thereby achieving the purpose of downscaling at scattered station scale.

Specifically, we classify the observational data into two categories: one is the gridded high-resolution indirect observational data (such as satellite observations), and the other is the scattered sub-grid direct observational data (such as weather observation stations). This also conforms to the observation

TABLE I
METEOROLOGICAL VARIABLES USED FOR STATION-SCALE DOWNSCALING.

Long Name	Short Name	Description	Unit
10m u-component of wind	u_{10}	Eastward component of the wind speed, at the height of 10 meters above the surface of the Earth.	m/s
10m v-component of wind	v_{10}	Northward component of the wind speed, at the height of 10 meters above the surface of the Earth.	m/s
2m temperature	t_{2m}	Temperature of air at 2m above the surface of the land, sea, or inland waters.	K
surface pressure	sp	Pressure of the atmosphere at the surface of land, sea, and inland water.	hPa
total precipitation in 1 hour	tp_{1h}	Accumulated liquid and frozen water, comprising rain and snow, that falls to the Earth's surface in 1 hour.	mm

TABLE II
DATASET USED IN THIS PAPER, PERIOD: 2017-01-01 TO 2021-08-31.

Data Name	Data Type	Resolution	Descriptions
ERA-5 Reanalysis [33]	Meteorological Field	$0.25°$, 1 hour	Meteorological fields of 5 surface variables in Tab. I.
Himawari-8 L1 Gridded data [35]	High-resolution Gridded Observations	$5km$, 10 min	Rreflectance of channel 'albedo_03', 'albedo_05', 'tbb_08' and 'tbb_15'
Weather2K [36]	Scattered Station Observations	sub-grid, 1 hour	Observation state of air pressure, temperature, wind speed, and total precipitation in 1 hour.

type settings used in the operational assimilation forecasting in actual meteorological services [10]. Given multiple low spatial resolution meteorological fields data $\mathcal{F}_{input} \in \mathbb{R}^{1\times V\times LH\times LH}$ at a certain time step, gridded high-resolution indirect observational data sequence $\mathcal{O} \in \mathbb{R}^{T\times C\times TH\times TW}$ and scattered sub-grid direct observational data $\mathcal{S} \in \mathbb{R}^{1\times V\times\ N}$, where V is the number of meteorological variables, T and C is the number of gridded observation frames and channels, N is the number of scattered observations. our goal is to obtain the meteorological state values $\mathcal{F}_{output} \in \mathbb{R}^{1\times V\times\ M}$ at any M scatter point locations by:

$$\mathcal{F}_{output} = \Phi(\mathcal{F}_{input}|\Theta(\mathcal{O}), \mathcal{S}) \tag{1}$$

where $\Theta(\cdot)$ is a function that maps the indirect observational data into the meteorological variable domain. $\Phi(\cdot)$ is a downscaling model that is used for downscaling to the scattered station scale. It should be noted that, in order to verify the generalization ability of the downscaling process for different scattered point locations, we require that the N points in $\mathcal{S}$ and M points in $\mathcal{F}_{output}$ are disjoint. Under such a setting, downscaling to multiple grid scales or random scattered point scales can be achieved by altering the positions of the target points.

2) Meteorological Variables Selection: In order to select significant meteorological variables to verify the effectiveness of the downscaling method, we analyzed the characteristics of the observational data and specially selected five surface variables as the focus of our task: u-component wind (u_{10}), v-component wind (v_{10}), 2-meters temperature t_{2m}, surface pressure (sp) and total precipitation in 1 hour (tp_{1h}). For details please refer to Tab. I. We chose these five variables primarily because the observational data includes the aforementioned variables, or the values of related variables can be roughly inferred through indirect observations, or there is an implicit correlation between the observational information and the variables.

B. Dataset

This subsection will introduce the actual data used for the proposed task, with the main data information available in Table II. Based on the observational data and meteorological field data we used, we selected the research area with a boundary of $80°$E to $136°$E and $18°$N to $54°$N, as shown in Fig. 2. This area is an intersection of all the regions covered by all kinds of data we used. The selected study area covers most of mainland China in East Asia. It is important to note that since the Himawari-8 satellite cannot cover parts of western China, we excluded the corresponding regions and the observational station information contained therein.

1) Meteorological Field Data: We select the widely recognized ERA5 reanalysis data [33] as the meteorological field data. Its original resolution is $0.25°$, covering the globe, with a temporal resolution of 1 hour. The ERA5 reanalysis data has been widely used in the field of meteorological forecasting based on deep learning [65], [66], and it is employed as both the initial field and supervision data for models [4]–[7]. For our task, we have extracted data for the study area and, through the operation of average pooling, downsampled the data to a

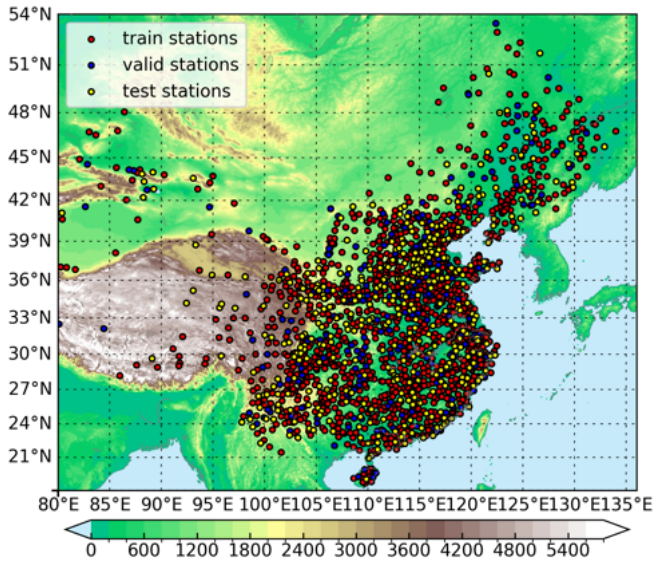


Fig. 2. The study area and scattered stations used in our paper. The red dots represent the training stations, the blue dots represent the validation stations and the yellow dots represent the test stations.

spatial resolution of 1° to serve as input for the model. This setup enables us to provide high-resolution grid supervision.

It should be noted that previous downscaling work was mostly based on forecasting tasks, which primarily involved downscaling coarse-resolution forecast fields and utilizing high-resolution analysis data for supervision [3], [28]. This setting is typically taken from the perspective of practical operational applications. Different from its starting point, our task aims to study more effective downscaling methods at the scale of scattered stations based on deep learning. Hence, we hope to utilize readily available public data to provide as many complete samples as possible for model training. However, most model forecast fields and high-resolution analysis data are often difficult to obtain, or they have a low temporal resolution [67], making it challenging to meet the requirements for a large sample size. Therefore, we have chosen the most commonly used ERA5 reanalysis data, which ensures the fulfillment of our task requirements. The methods developed on it can also be well extended to situations where the forecast fields are used as inputs.

2) Observation Data: Various observation data are crucial to forming a structured grid of meteorological fields. In the field of meteorology, data assimilation tasks [10], [62]–[64] specifically study how to integrate different observation data into forecast fields, improving the performance of forecast models based on real-time observations. Integrating observational data effectively into meteorological fields to obtain more accurate meteorological states at different locations is an important research direction. We hope to enhance the accuracy of downscaling by incorporating observational information into downscaling tasks. On the one hand, using observational information to improve the accuracy of downscaling, and on the other hand, treating downscaling as a fundamental task to explore effective methods for integrating observational data into meteorological fields.

Based on the task description provided earlier, we selected the L1 gridded data from the next-generation geostationary meteorological satellite Himawari-8 [35] as the gridded high-resolution indirect observational data, and we chose the station dataset provided by the Weather2K [36] as the scattered sub-grid direct observational data. Below is a brief introduction to the two types of observational data:

Himiwari-8 L1 Gridded Data [35] is generated by JAXA/EORC from the Himawari Standard Data with re-sampling to equal latitude-longitude grids. It includes 16 spectral bands, with a spatial resolution of up to 2 km and a temporal resolution of 10 minutes, obtained from the Advanced Himawari Imager (AHI) onboard the Himawari-8 satellite. Due to storage constraints, we primarily downloaded the version with a 5 km spatial resolution and cropped the data to the research area of our interest. Furthermore, to reduce the memory footprint of the input data, we empirically selected four representative bands from the visible, near-infrared, and far-infrared spectral ranges. It should be noted that, as we are using Level 1 observational data rather than the results of the satellite retrieval of meteorological variables, the designed model is required to learn the state values of meteorological variables from the indirect satellite radiance values. This also implies that the retrieval process is inherently included in our task.

Weather2K dataset [36] was originally designed for mesoscale weather forecasting tasks. It comprises hourly observations of 20 meteorological variables from 1866 ground observation stations across China, spanning from January 2017 to August 2021. We have selected four surface meteorological variables as the focus of our research, which are: air temperature, air pressure, wind speed, and precipitation in 1 hour. For the downscaling task at the scattered station scale, the primary objective is to verify the model's generalization performance at various scattered locations. Therefore, as shown in Fig. 2, we randomly split the 1866 stations into three non-overlapping parts, with 1266 stations used for training, 200 stations for validation, and the remaining 400 stations used for testing.

C. Evaluation Metrics

To evaluate the downscaling effects of different methods at the scale of scattered observation stations, referencing prior work [34], we have chosen the mean squared error (MSE) and mean absolute error (MAE) averaged over both stations and time as our assessment metrics. The calculation methods are as follows:

$$\begin{aligned} MSE &= \frac{1}{M*T}\sum_{i=1}^{M}\sum_{t=1}^{T}(Y_i^t - \hat{Y}_i^t)^2 \\ MAE &= \frac{1}{M*T}\sum_{i=1}^{M}\sum_{t=1}^{T}|Y_i^t - \hat{Y}_i^t| \end{aligned} \tag{2}$$

where M is the total number of test observations, Y_i^t is the ground truth value for a given variable state at i-th station and t-th time point, $\hat{Y}_i^t$ is the predicted downscaled value. Both metrics are commonly used in regression tasks to measure the

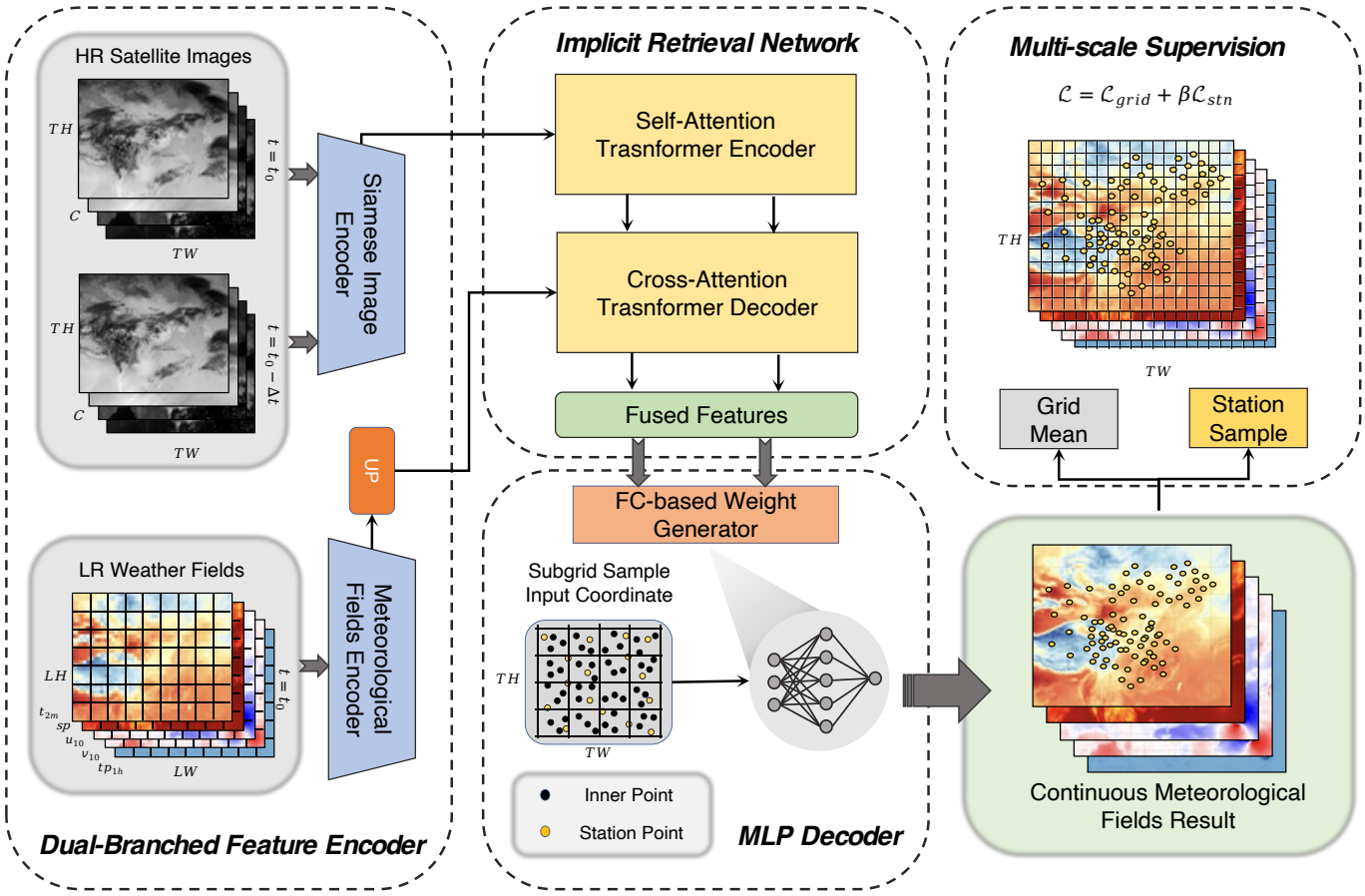


Fig. 3. The proposed *HyperDS* architecture. It mainly consists of three parts: a dual-branch feature encoder is used to extract semantic features from the input low-resolution meteorological field and high-resolution remote sensing images respectively; subsequently, the implicit retrieval network utilizes a cross-attention mechanism to implicitly fuse different feature information and align the remote sensing image features with meteorological field variables; and finally, the FC (fully connected)-based weight generator predicts the weight vector for the target network. The MLP (Multi-Layer Perceptron) decoder, the target network, learns the mapping from the sampled subgrid coordinates to the corresponding location state values. It is supervised at both the observation station scale and the high-resolution grid scale, allowing for the continuous modeling of the meteorological fields.

accuracy of the predicted values. Lower values of MSE and MAE indicate better model performance, with the MAE being particularly useful for understanding the error magnitude on an average per-observation basis.

IV. Method

In response to the *Station-Scale Downscaling* task described above, we developed a novel method, namely *HyperDS*, that effectively integrates high-resolution Himawari-8 satellite observations and scattered station observations to recover subgrid-scale meteorological states from low-resolution atmospheric fields, achieving continuous-resolution modeling of the meteorological field. This section will provide a detailed introduction to the structure and training strategy of our proposed method.

A. Overall Structure

The overall structure of *HyperDS* can be viewed as a data-conditional hypernetwork architecture [39]. Considering the type of observational data, we use the indirectly observed high-resolution Himawari-8 satellite images as auxiliary input to the model, and the direct scattered observation station data as the model's station-scale supervision. The reason for this setup is to enable the model to learn the implicit meteorological field information from indirect remote sensing observations through operations such as encoding and feature extraction of the former. At the same time, supervision from the observation stations is used to correct the inherent biases that occur when downscaling from grid scale to station scale.

As illustrated in Fig. 3, our model is composed of three sub-network structures: a dual-branch feature encoder based on CNN and an implicit inversion network based on a Transformer with cross-attention from the hypernetwork, which generates the fused features used to determine the weights of the target network; the target network, in turn, comprises an MLP-based decoder that learns the mapping from specific coordinates to meteorological states, using the weight parameters generated by the hypernetwork.

Additionally, in the input portion of the MLP decoder, we design a coordinate selection method based on subgrid sampling that more naturally and reasonably adapts to the capability of implicit neural representation for continuous state modeling. In such a case, by averaging the subgrid samples

pixel by pixel, we can utilize high-resolution grid scale data for supervision and learn the deviation loss between the predicted values and the site-scale observations by sampling specific scattered station locations. Averaging and sampling of this form are also better aligned with the processing methods used to integrate multiscale observational data in meteorological fields.

B. Dual-Branched Feature Encoder

For processing meteorological field data and satellite imagery simultaneously, we implemented a simple dual-branch encoding structure based on ResNet-18 [68] as the backbone for both the meteorological field encoder and the satellite image encoder. For the satellite image encoding branch specifically, we chose to input two adjacent Himawari-8 images in a Siamese configuration into the encoder, which strikes a balance between memory consumption and the amount of input information. To balance the semantic information and spatial information of the features, we selected the features from the intermediate layers of the two encoders as the outputs of the model. Specifically: given the low-resolution meteorological $\mathcal{F}_{\text{input}} \in \mathbb{R}^{1\times 5\times LH\times LH}$ and two frames high-resolution Himawari-8 remote sensing images $\mathcal{O} \in \mathbb{R}^{2\times 4\times TH\times TW}$, the extracted features from each encoder can be computed by:

$$\begin{aligned} F_{\text{field}} &= \text{Conv2d}_{\text{field}}(\text{MeteoEncoder}(\mathcal{F}_{\text{input}})) \\ F_{\text{field}} &= \text{Conv2d}_{\text{field}}(F_{\text{field}}) \\ F_{\text{h8}} &= \text{Concat}(\text{ImgEncoder}(\mathcal{O}_0), \text{ImgEncoder}(\mathcal{O}_1) \\ F_{\text{h8}} &= \text{Conv2d}_{\text{field}}(F_{\text{h8}}) \end{aligned} \tag{3}$$

The two features are extracted from the same stage in ResNet-18 and aligned across all dimensions into $C \times h \times w$ through upsampling and convolution operations.

C. Implicit Retrieval Network

To further integrate the two types of extracted features, we adopted a Transformer encoder-decoder network based on a cross-attention mechanism. This network implicitly retrieves indirect observation information into the meteorological field domain and effectively integrates it. We first flatten F_{field} and F_{h8} into shape $C \times hw$, and add the learnable positional encoding vector. Then, the features from high-resolution remote sensing images are fed to the self-attention Transformer encoder to further learn the relationships between different tokens. Following a cross-attention Transformer decoder receives features from both remote sensing images and meteorological fields to implicitly learn the relationships between different feature domains. Therefore, the fused features generated by the implicit retrieval network can be computed by:

$$\begin{aligned} F_{\text{h8}} &= \text{SelfAttnEncoder}(F_{\text{h8}}) \\ F_{\text{fused}} &= \text{CrossAttnDecoder}(F_{\text{h8}}, F_{\text{field}}) \end{aligned} \tag{4}$$

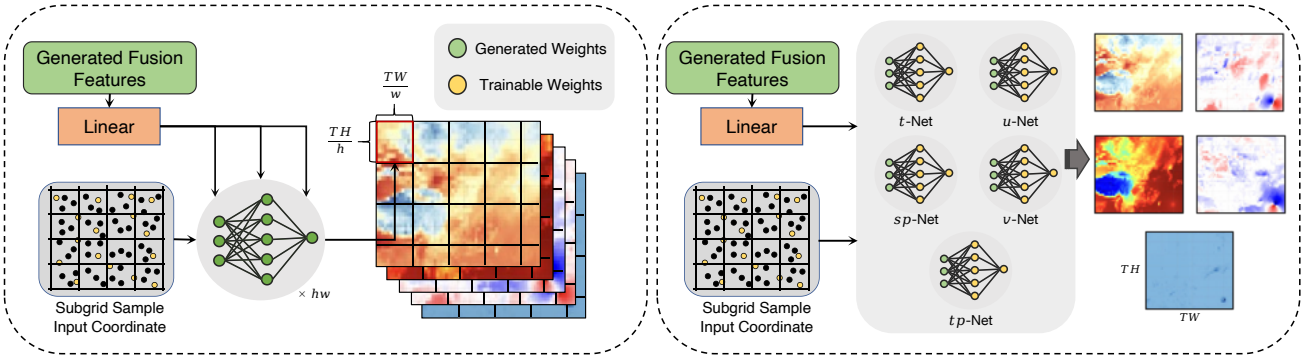
Through the above calculations, the generated fusion features contain high semantic features of both low-resolution meteorological fields and high-resolution satellite observations, laying the foundation for subsequent continuous-resolution modeling.

D. MLP Decoder with Subgrid-sampling

Based on the fusion features generated by the hypernetwork structure and combined with the latent neural representation method [37], we designed a decoder module based on the Multilayer Perceptron (MLP). By learning the mapping from coordinate positions to meteorological states, we realized continuous modeling of the meteorological field.

Unlike previous modeling methods based on latent neural representations that directly use the coordinates of the grid center to represent grid values [42], [43], [54], we designed a specialized sub-grid sampling method specifically for meteorological fields. This allows us to construct a continuous representation of the meteorological field more naturally and reasonably. Specifically, for a given pixel p in a high-resolution grid $\text{Grid}(\text{TH}, \text{TW})$, we randomly sample multiple inner points $I_p = \{(x_i, y_i) \mid i = 1, 2, \cdots, P\}$ within this pixel as the coordinate values to be input in this pixel. Then, we can obtain the state value for the corresponding resolution grid by calculating the average of the meteorological state values associated with all the inner points under that pixel, which can then be supervised using the corresponding high-resolution grid labels. Specifically, when a pixel includes the location of observation stations, we will also sample the positions of these observation stations. By using proposed the subgrid sampling method described above, we can average the data from observation stations into the mean value of the grid, thereby mitigating the conflict between the scattered point stations and the grid values. It should be noted additionally that, apart from the coordinate values, we also input the grid interpolation results of the state values corresponding to the coordinate points as auxiliary information into the MLP. For the sake of simplifying the expression, this part of the information will not be explicitly reflected in the following text.

As for the MLP modeling method used to represent the meteorological field, as shown in Fig. 4, we referenced previous work [43], [69] and designed two different modeling structures. The first type, the multi-block-based MLP decoder, divides the target meteorological field into several blocks, each of which is continuously modeled by a separate MLP that takes input coordinates and simultaneously outputs the state values of all target variables. The weights of the network are obtained entirely through the linear mapping of fused features. Different from the first one, the second type is a multivariate MLP decoder that uses different MLPs to model the entire meteorological field of specific variables separately. The shallow parameters of each MLP are obtained through linear mapping of the fused features, whereas the deeper parameters are randomly initialized and constitute learnable weights. The reason for this setup is that the former, multi-block-based modeling approach, although more conducive to modeling high-frequency information, is more memory-intensive and increases model complexity with the number of sub-blocks. The latter has a relatively fixed computational complexity and memory usage, but it makes the modeling task more challenging for a single MLP. Therefore, each method has its advantages and disadvantages, and we hope to provide



(a) Multi-block-based MLP decoder (b) Multi-variable-based MLP decoder

Fig. 4. Two variants of MLP decoders based on implicit neural representations with subgrid sampling.

more flexible options for our approach.

E. Loss Function

The supervision label data proposed for *HyperDS* includes two scales: grid-scale and station-scale. By inputting different sampling coordinates into the MLP decoder, predictions for the corresponding scales can be generated. Detailed introductions follow.

1) Grid-scale Loss: Referencing previous downscaling work based on Super-Resolution (SR), we use high-resolution meteorological fields as supervision at the grid scale. However, unlike previous work, we obtain the prediction result for a target pixel by calculating the average of the sub-grid inner points within that pixel. To be specific, given the high-resolution grid label field $\mathcal{F}_{\text{label}}^{\text{grid}} \in \mathbb{R}^{1\times 5\times TH\times TW}$, we sample P inner points $I_p = \{(x_i, y_i) \mid i = 1, 2, \cdots, P\}$ at each pixel p in $\text{Grid}(\text{TH}, \text{TW})$ and the predicted field can be computed by:

$$\mathcal{F}_{\text{output}}^{\text{grid}} = \frac{1}{P}\sum_{i=1}^{P} \text{MLPDecoder}(I, F_{\text{fused}}) \tag{5}$$

then, the grid-scale loss can be computed by:

$$\mathcal{L}_{\text{grid}} = \|\mathcal{F}_{\text{label}}^{\text{grid}} - \mathcal{F}_{\text{output}}^{\text{grid}}\|^2 \tag{6}$$

However, while high-resolution supervision can provide accurate mean supervision on fine grids, in actual applications, high-resolution grid supervision is often difficult to obtain. To account for station-scale downscaling in such situations, we referenced previous work on modeling dynamic systems [70] and designed a grid loss function for when high-resolution gridded supervision is not available. In such cases, still benefiting from the sub-grid sampling coordinate input form, we could also compute the mean values covered by the low-resolution pixel in $\text{Grid}(\text{LH}, \text{LW})$, at the same time interpolate the input low-resolution fields into a fine-grained one. Thereby simultaneously obtaining the interpolated high-resolution supervision as well as the low-resolution mean supervision:

$$\begin{aligned} \mathcal{L}_{\text{HR}} &= \|\text{Interp}(\mathcal{F}_{\text{input}}) - \mathcal{F}_{\text{output}}^{\text{grid}}\|^2 \\ \mathcal{L}_{\text{LR}} &= \|\mathcal{F}_{\text{input}} - \text{Avgpool}(\mathcal{F}_{\text{output}}^{\text{grid}})\|^2 \\ \mathcal{L}_{\text{grid}} &= \mathcal{L}_{\text{HR}} + \mathcal{L}_{\text{LR}} \end{aligned} \tag{7}$$

Through such an approach, in the absence of high-resolution grid supervision information, it is possible to provide as much grid supervision information as possible to the maximum extent.

2) Station-scale Loss: To integrate the observational information from scattered stations into the downscaling process and mitigate the inherent bias between the meteorological field and grid observations, the training process incorporates station-scale supervision to learn the cross-scale mapping from the meteorological field to the stations. To be specific, given the station-scale label $\mathcal{F}_{\text{label}}^{\text{station}} \in \mathbb{R}^{1\times 5\times M}$, where M is the number of stations, we could sample the corresponding coordinates and computed the station scale output by:

$$\mathcal{F}_{\text{output}}^{\text{station}} = \text{MLPDecoder}(I', F_{\text{fused}}) \tag{8}$$

and the station-scale loss is

$$\mathcal{L}_{\text{stn}} = \|\mathcal{F}_{\text{label}}^{\text{station}} - \mathcal{F}_{\text{output}}^{\text{station}}\|^2 \tag{9}$$

It should be noted that since only the wind speed (ws) variable, rather than its components, is provided in the Weather2K data [36], we base our calculation of wind speed loss on the following formula:

$$ws = \sqrt{u_{10}^2 + v_{10}^2}. \tag{10}$$

Combining the above loss functions, we can derive the final loss function as:

$$\mathcal{L} = \mathcal{L}_{\text{grid}} + \beta\mathcal{L}_{\text{stn}} \tag{11}$$

where β are the loss coefficients that need to be manually set.

V. Results

In this section, we design experiments and baseline methods tailored to the station-scale downscaling task we have constructed, which will be compared with our proposed *HyperDS* model.

A. Experiment Details

To validate the downscaling performance from low-resolution meteorological fields to arbitrarily scattered stations, we downsampled the original ERA5 reanalysis data [33] to a spatial resolution of 1° using the method of average pooling,

which served as the input meteorological field data. Moreover, we used the original ERA5 data at 0.25° resolution as high-resolution grid supervision. Additionally, we conducted experiments without high-resolution grid supervision by setting the loss according to Eq. 7. The dataset was divided into three parts in chronological order: data from January 1, 2017, to August 31, 2020, was used for training; data from September 1, 2020, to December 31, 2020, was used for validation; and data from January 1, 2021, to August 31, 2021, was utilized for testing. It should be noted that for the observational station data from Weather2K [36], we partitioned the dataset based on both time and station locations. This division requires that the model exhibit strong generalization performance both temporally and spatially when validated at the observational station scale, which presents a significant challenge. Therefore, the experiments we set up aim to recover the meteorological state variables at a scattered scale from a 1° spatial resolution meteorological field and to validate the performance at 400 randomly sampled test stations (as shown in Fig. 2) within the testing period. The hyperparameter number of samples in the MLP decoder is set to 10, and the loss coefficient $\beta = 0.05$.

To ensure a fair comparison of different methods, we have established the same training procedures and hyperparameters for all. We optimized the model using the Adam optimization method [71], employing a cyclical learning rate with cosine annealing [72], starting with an initial learning rate of 0.0001, for a total of 50 epochs of training. We choose the checkpoint with the lowest station-level loss in the validation sets for testing. We trained our proposed model using 4x NVIDIA A100 GPUs, setting the batch size to 4 per GPU.

B. Baselines

For the task we proposed, we designed two basic baseline methods for comparison to validate the effectiveness of our method. The following subsections provide a detailed introduction to these methods.

1) Interpolation of Weather Field into Station Scale: One of the simplest and most direct methods to obtain meteorological state variables at the scale of scattered stations from a meteorological field is through interpolation. We use the $\mathrm{DataArray.interp}$ function with the default setting from the open-source xarray library to perform interpolation on the meteorological field, based directly on the absolute positions of latitude and longitude, with each grid cell's state value corresponding to its center point coordinates. We mainly performed interpolation on meteorological fields with resolutions of 1° and 0.25°, corresponding to the interpolated results from the input low-resolution meteorological fields and the high-resolution supervision. Since the interpolation process does not incorporate any available observational information, this method can serve as our most basic baseline result. Moreover, it can reflect to some extent the inherent bias that exists between the meteorological fields and observations.

2) Super-Resolution-based Downscaling with Observations: Given the widespread application of super-resolution models in downscaling tasks, we specifically modified traditional super-resolution models for our proposed task, integrating multi-scale observational information into them. As shown in Fig. 1(a), traditional SR-based downscaling methods mainly learn the mapping between low-resolution input fields and the target high-resolution fields within an encoder-decoder architecture. We adopted a straightforward approach to incorporate high-resolution Himawari-8 (H8) satellite imagery observations and Weather2K station observations into the model. Specifically, given the high-resolution H8 images we encode them with a single convolutional layer, then align the dimension with the input meteorological fields by average pooling operation and concatenated them on top of the input low-resolution meteorological field before feeding them into the super-resolution model. After obtaining the meteorological field at the target resolution, the meteorological state of the scattered stations can be acquired through interpolation. Subsequently, both the grid supervision and the corresponding station supervision data are utilized to compute the loss function, which is then used for backpropagation. In terms of the choice of super-resolution models, we selected the classic UNet-based super-resolution model with ResNet-18 backbone [73] and the EDSR [48] and made modifications to their structures and implementations.

C. Comparison with Baselines

We compared the performance of our proposed *HyperDS* method with other baseline methods for station-scale downscaling with high-resolution grid supervision. Due to the trade-off between grid-scale supervision and station-scale supervision in the model optimization process, but as our current task mainly focuses on the performance at the station scale, we chose the checkpoint with the smallest station-scale loss on the validation set during the training process as the model for testing. The loss coefficients can also greatly affect site performance, so considering the need to balance the losses at both scales, we empirically set the loss coefficients in Eq. 11 as $\beta = 0.05$. Subsequent sections will further discuss the trade-off problem between the losses at the two scales. Since the Weather2K dataset [36] used as labels do not include the wind speed component, we compared the downscaling results of 4 surface variables, which are: wind speed (ws), surface pressure (sp), 2m temperature (t_{2m}) and total precipitation in 1 hour (tp_{1h}).

1) Test Results by Variables: Tab. III displays the test metrics for different variables using different methods at the 400 testing observation stations. From the results, it can be seen that our proposed *HyperDS* method outperforms the compared baseline methods on all variables. Particularly for wind speed and surface pressure, our method significantly exceeds the others, with the MSE for wind speed improving by 67% and for surface pressure by 19.5% compared to the best baseline results.

It should be noted that in the results of direct interpolation of meteorological fields, the ERA5 1° interpolation results are superior to the 0.25° interpolation results in metrics such as wind speed and precipitation, which seems counterintuitive. However, similar results have been reported in recent related work [34]. We believe this is due to the limited assimilation of observational station data in the ECMWF reanalysis data for the China region.

TABLE III
STATION-LEVEL DOWNSCALING RESULTS FOR WIND SPEED (ws), SURFACE PRESSURE (sp), 2M TEMPERATURE (t_{2m}) AND TOTAL PRECIPITATION IN 1 HOUR (tp_{1h}) OF VARIOUS METHODS.

Method	ws		sp		t_{2m}		tp_{1h}	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ERA5 1°	5.5642	1.7842	1048.3886	21.1125	7.7483	1.9235	1.1396	0.1896
ERA5 0.25°	6.2164	1.9118	801.3915	17.2155	6.7855	1.7770	1.2018	0.1893
UNet [73]	5.4575	1.7757	967.8221	20.1538	7.3537	1.8806	1.1426	0.1955
EDSR [48]	6.1547	1.8905	896.9313	19.2898	7.1336	1.8386	1.1572	0.1982
HyperDS (Ours, multi-var)	**1.7995**	**0.9568**	716.0126	15.4001	**6.3588**	**1.7656**	1.1278	0.1887
HyperDS (Ours, multi-block)	1.9671	1.0126	**645.0722**	**14.6524**	6.5747	1.8400	**1.1260**	**0.1859**

Upon further analysis of the test results for different variables, it is evident that our method shows the most significant improvement in wind speed. This is primarily because wind speed exhibits the most notable sub-grid variability, with local wind speeds often being influenced by a variety of small-scale meteorological processes such as turbulence, making it difficult to capture at the relatively coarse resolution of grid scales. On the contrary, for variables such as 2m temperature and precipitation, the improvement from our method is relatively small. This is because the variability of local temperature is relatively gradual, and as for precipitation, due to its sparse and long-tailed distribution [74], acceptable downscaling results can be obtained by simply interpolating the grid data.

For the two super-resolution-based comparison methods, we can see that although observational information has been incorporated, the improvement in overall station downscaling performance is quite limited. We believe that this is because traditional super-resolution methods, which are based on fixed-resolution grid supervision, place more emphasis on the regression task for each grid pixel, and the learning process remains a discrete mapping from coarse grid scales to fine grid scales. In contrast, the optimization goal of our proposed *HyperDS* method, which is based on hypernetworks and implicit neural representations, is to learn a mapping from arbitrary coordinates to meteorological states, constructing a continuous representation of the meteorological field. This endows the model with stronger capabilities for expressing sub-grid-scale information. This also explains why the EDSR model, which has stronger grid super-resolution capabilities, performs worse in station downscaling than the simpler UNet model. The reason is that EDSR has a stronger ability to fuse and extract features at the grid scale, which makes it difficult for the model to generalize well to the station scale, even with the inclusion of station-scale observations as supervisory labels.

For the two different MLP decoder structures we proposed, it can be seen that different decoder outcomes have certain performance differences for different variables. The multivariate-based HyperDS performs relatively better on wind speed and 2m temperature, while the multi-block-based HyperDS performs better on surface pressure and precipitation. We believe this is due to the different statistical distributions of various variables. Although we have normalized all variables by their mean and variance, making them roughly follow a Gaussian distribution with zero mean, there are still significant differences in the value ranges of each variable after normalization. For example, the variation range of surface air pressure is relatively larger, whereas the temperature variation range is comparatively smaller. The decoder based on multiple blocks, with each MLP representing a local region, can model the data for a specific area more effectively, thus avoiding the issue of too great a range of variable changes caused by global modeling. The decoder based on multiple variables uses a single MLP to model the entire meteorological field of a region, which is more effective for variables with smaller ranges of variation and can also save more computational memory consumption.

2) Result Visualization: Fig. 5 illustrates the results of downscaling at different test stations using various methods, where the color of each station represents the magnitude of the normalized mean square error at that station, with lighter colors indicating larger errors. The base map of each visual image is also the result of downscaling at the grid scale with $0.25°$ spatial resolution of each model. The results in the figure show that our proposed *HyperSR* method performs significantly better at the site scale compared to other methods. In particular, for the wind speed variable, the baseline methods exhibit a large error in the northeastern area of the study region (more white dots), but our method can effectively correct the downscaling bias in this area. Combining this with Fig. 2, we can see that for areas with sparser training stations (such as the northeastern and western regions), the downscaling performance of all methods tends to decline to different extents compared to the densely observed southeastern region. However, our method still exhibits better generalization performance; for instance, for the 2m temperature variable, our method has relatively fewer white dots in the northeastern region compared to other baseline methods. It is noteworthy that the downscaling method based on super-resolution (SR) achieves better downscaling results at the grid scale (i.e., base map) compared to our method. This is a limitation of our method and a direction for further improvement in the future.

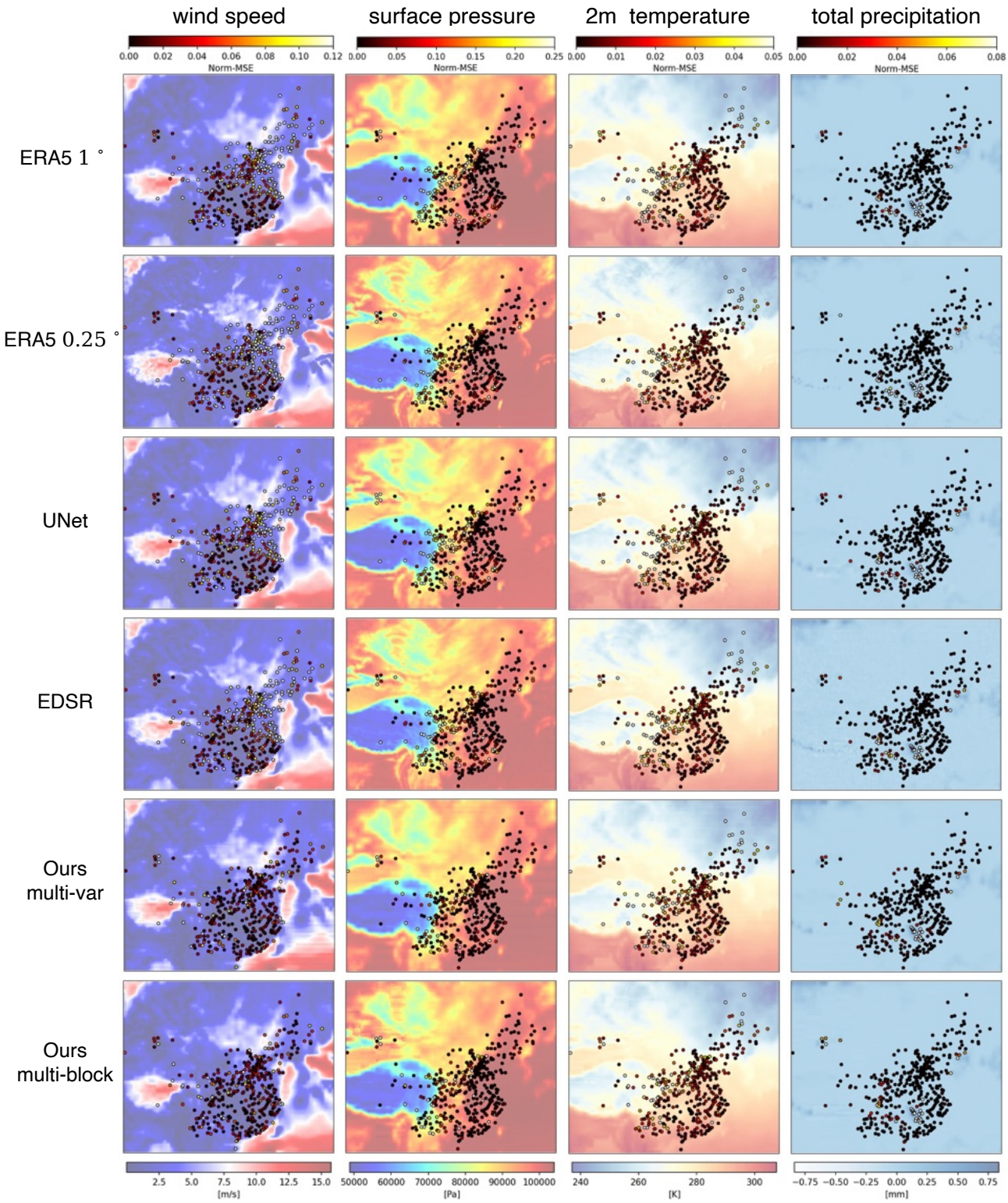


Fig. 5. Visualization comparison of downscaling to station-scale using different methods, where the color of each station represents the magnitude of the normalized mean square error at that station, with lighter colors indicating larger errors, i.e. the darker the color of the site, the better the performance of the downscaling. The base map of each visual image is also the result of downscaling at the grid scale with 0.25° spatial resolution of each model.

D. Downscaling performance across different subregions

From Fig. 2, it is evident that the density of stations exhibits significant variation according to different regions. Specifically, the distribution of stations is relatively dense in the southeastern coastal regions, whereas in the northwest and southwest regions (such as the Xinjiang area and the Tibetan Plateau), the distribution of sites is comparatively sparse. Consequently, to further analyze the impact of station density on the results of downscaling, we have established four distinct subregions to compare the downscaling results at the test stations within our proposed HyperDS and ERA5 interpolation. The division of the four subregions is as follows:

- Region 1 covers the southeast of the Chinese mainland with a boundary of $105°$E to $125°$E and $20°$N to $40°$N. A total of 270 test stations are included within this subregion.
- Region 2 covers the northeast of the Chinese mainland with a boundary of $120°$E to $135°$E and $42°$N to $54°$N. A total of 28 test stations are included within this subregion.
- Region 3 covers the northwest of the Chinese mainland with a boundary of $80°$E to $100°$E and $40°$N to $54°$N. A total of 7 test stations are included within this subregion.
- Region 4 covers the Tibetan Platea area of the Chinese mainland with a boundary of $80°$E to $100°$E and $25°$N to $40°$N. A total of 16 test stations are included within this subregion.

Tab. IV presents a comparison of the downscaling results for stations across different subregions. The results from the table clearly show significant differences in the downscaling outcomes across different subregions. Region 1, having the highest number of observational stations during the training stage, also exhibits the best downscaling results. It is noteworthy that the wind speed variable in Region 1 demonstrated poor performance with the ERA5 interpolation method. However, by leveraging extensive observational station data for training, our proposed HyperDS method significantly enhanced the downscaling results for wind speed, achieving the best outcomes. This outcome is also consistent with the visual results presented in Fig. 5. For the other three regions, which have relatively fewer sites, particularly Regions 3 and 4, the downscaling results for wind speed, temperature, and other meteorological variables exhibit varying degrees of decline. This is largely due to the discrepancies between the low-resolution meteorological fields used as inputs and the observations made at the stations in these regions. Nevertheless, HyperDS still enhanced the downscaling performance even with fewer observations.

In summary, the comparative results across different regions fully demonstrate the effectiveness of the HyperDS method. Additionally, it also reveals the uneven assimilation capabilities of the original meteorological field data in various regions. For areas where observation is particularly challenging, such as the Tibetan Plateau, future efforts will require more specialized methods to further enhance downscaling performance.

E. Downscaling without High-resolution Gridded Supervision

In practical applications, high-resolution gridded meteorological fields are often difficult to obtain, whereas direct observations at observation stations are relatively easier to access and the data quality is more stable. Obtaining station-scale meteorological states directly from low-resolution meteorological fields is a very meaningful task. Hence, we conducted station-scale downscaling experiments without high-resolution grid supervision based on the designed loss function Eq. 7.

Tab. V shows the downscaling results at the station level for different methods. The results from the table indicate that, even without high-resolution grid supervision, our proposed HyperDS outperforms other methods on the majority of the evaluation metrics for most variables, particularly for the wind speed variable. However, for 2m-temperature and surface pressure variables, compared to the results supervised with high-resolution grid data, there is a noticeable performance decrease (for surface pressure, the MSE decreased from 645.0722 to 805.9112, and for 2m-temperature, it decreased from 6.3588 to 7.1487). This performance degradation is because high-resolution grid supervision provides a significant improvement over coarse-resolution grid inputs at the station level for these two variables. Consequently, the absence of high-resolution grid data supervision leads to a decline in performance. In contrast, for the wind speed variable, the downscaling performance at the station level is slightly improved (MSE decreased from 1.7995 to 1.7815) because the interpolation results from low-resolution inputs are less dependent on high-resolution supervision.

Although there is a certain degree of performance decline, our proposed HyperDS method still outperforms the comparison methods even without high-resolution grid supervision. Even when the comparison methods incorporate high-resolution supervision (as shown in the results of Tab. III), our method remains superior on most metrics compared to those based on grid super-resolution networks.

F. Ablation Study

We further compared the performance of the *HyperDS* method under different settings to verify the impact of the inclusion of observational data and the number of samples on the method's performance.

Tab. VI shows the test performance of *HyperDS* under different experimental settings. The results indicate that the inclusion of station observation supervision is the most critical factor affecting the model's performance. This is intuitive, as previous work has also shown that there is an inherent bias between the meteorology itself and scattered station observations [34], which cannot be recovered solely through high-resolution grid supervision. Based on this result, coupled with the fact that we use station observations as our supervision labels, it means that no real-time station observations are needed during the model inference stage. The model itself can adaptively generalize the meteorological field to any station location, which also implies that the model has learned the inherent bias from the meteorological field to the station and has effectively

TABLE IV
STATION-LEVEL DOWNSCALING RESULTS FOR WIND SPEED (ws), SURFACE PRESSURE (sp), 2M TEMPERATURE (t_{2m}) AND TOTAL PRECIPITATION IN 1 HOUR (tp_{1h}) ACROSS DIFFERENT SUBREGIONS.

Subregions	#Stations	Methods	ws		sp		t_{2m}		tp_{1h}	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Region 1	270	ERA5 1°	5.5319	1.8160	721.5415	18.1236	4.6385	1.6041	1.3967	**0.2135**
		HyperDS (Ours, multi-var)	**1.7674**	**0.9430**	478.7216	13.3340	**4.6018**	**1.5668**	1.3752	0.2137
		HyperDS (Ours, multi-block)	1.9406	0.9742	**470.5051**	**12.1737**	4.6337	1.6130	**1.3746**	0.2147
Region 2	28	ERA5 1°	7.6884	2.2915	190.7777	10.4336	7.8760	1.8637	0.6024	**0.1195**
		HyperDS (Ours, multi-var)	**1.9086**	**1.05156**	111.2481	8.2178	**6.1165**	**1.7686**	0.5844	0.1242
		HyperDS (Ours, multi-block)	2.5825	1.2313	**82.6108**	**6.8603**	7.8511	1.8394	**0.5801**	0.1211
Region 3	7	ERA5 1°	3.1861	1.3679	1433.3450	29.4225	14.3588	3.0458	0.0703	**0.0421**
		HyperDS (Ours, multi-var)	**2.8637**	**1.2714**	1394.6950	27.1029	**10.7129**	**2.8938**	0.0700	0.0677
		HyperDS (Ours, multi-block)	3.0135	1.3189	**1061.6501**	**26.0353**	15.8731	3.1928	**0.0690**	0.0568
Region 4	16	ERA5 1°	3.9376	1.5139	2713.4776	42.2931	27.4980	3.9007	0.2903	0.1418
		HyperDS (Ours, multi-var)	**3.0041**	**1.3109**	1218.1292	26.3006	**15.4011**	**3.1412**	0.2753	0.1430
		HyperDS (Ours, multi-block)	3.3948	1.3837	**990.8851**	**24.5738**	23.3481	3.4003	**0.2682**	**0.1243**

TABLE V
STATION-LEVEL DOWNSCALING RESULTS FOR WIND SPEED (ws), SURFACE PRESSURE (sp), 2M TEMPERATURE (t_{2m}) AND TOTAL PRECIPITATION IN 1 HOUR (tp_{1h}) OF VARIOUS METHODS WITHOUT HIGH-RESOLUTION GRIDDED SUPERVISION.

Method	ws		sp		t_{2m}		tp_{1h}	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
UNet [73]	5.3051	1.7508	1078.9179	21.6993	7.6797	1.9225	1.1426	0.1992
EDSR [48]	5.2994	1.7495	1082.5246	21.7704	7.6539	1.9183	1.1426	**0.1902**
HyperDS (Ours, multi-var)	**1.7815**	**0.9613**	901.4435	17.9062	**7.1487**	**1.8239**	1.1319	0.1940
HyperDS (Ours, multi-block)	2.0379	1.0335	**805.9112**	**16.0210**	7.4909	1.9281	**1.1274**	0.1943

TABLE VI
ABLATION STUDY RESULTS OF HYPERDS WITH MULTI-BLOCK-BASED MLP DECODER. THE SETTINGS 'STATION' AND 'H8' REPRESENT THE STATION-LEVEL SUPERVISION AND H8 SATELLITE IMAGE INPUT. THE SETTING 'SAMPLE' REPRESENTS THE SUBGRID-SAMPLING STRATEGY IN MLP DECODER.

Method	Settings			ws		sp		t_{2m}		tp_{1h}	
	station	h8	sample	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
HyperDS	✗	✓	✓	5.4509	1.7769	1088.7840	23.3697	9.2645	2.2367	1.1427	0.1983
	✓	✗	✓	**1.8876**	**0.9893**	721.9860	15.2776	6.8517	1.8417	1.1262	0.1910
	✓	✓	✗	2.3029	1.1133	729.5865	15.7978	7.1172	1.8925	1.1322	0.1936
	✓	✓	✓	1.9671	1.0126	**645.0722**	**14.6524**	**6.5747**	**1.8400**	**1.1260**	**0.1859**

reduced it. Therefore, our method can be regarded as a general interpolation model from the meteorological field to stations.

Regarding the input of H8 remote sensing satellite images, although it is not a direct representation of meteorological conditions, through our designed feature extraction and implicit retrieval network, the model can learn useful information from indirect observations. However, the results indicate that the input from H8 did not show a positive impact on all variables. This is because the satellite observations we input are Level 1 radiance data, in which the meteorological state information is implicit and incomplete. From the types of Level 2 (L2) inversion products provided by the Himawari-8 satellite, it is evident that the primary meteorological variables related to it are surface temperature, humidity, and high-altitude wind speed (obtained indirectly based on cloud movement). Therefore, in the experimental results, the incorporation of H8 data has a more pronounced improvement in 2m temperature and surface pressure (which are strongly correlated with humidity). However, for the surface wind speed, since it has a significant deviation from the high-altitude wind speed, the results do not

show a direct enhancement.

Sampling at the subgrid coordinates is also one of the innovative aspects of our method. By using subgrid sampling, the traditional implicit neural representation methods can be better aligned with the continuous distribution characteristics of the meteorological field, and the scattered station observations are treated as samples to be averaged with other samples within the same pixel. We set subgrid sampling numbers to 2 for comparison, and the experimental results also indicate that more samples typically yield better experimental outcomes and faster convergence rates. However, excessive sampling usually means greater GPU memory usage; therefore, we only set the maximum number of samples to 10 during the experimental process.

G. Analysis of the effectiveness of Himawari-8 obervations

To further explain the negative impact of Himawari-8 observations on wind speed variables as shown in Tab. VI, we have designed an additional experiment for interoperability. Since the retrieval process in the HyperDS model proposed in this paper is implicitly represented through the cross-attention mechanism in the Transformer model, it is not possible to intuitively visualize the effects of the inversion. Therefore, we designed a vanilla CNN-based encoder-decoder network, taking H8 observations as input and using ERA5 for supervision, to simulate the retrieval process. As illustrated in Fig. 6, we visualized the albedo03 band of the input H8 observations, as well as the retrieved meteorological fields and the corresponding ground truth. By observing the texture information in the retrieved results, it can be seen that the surface pressure variables perform the best, followed by temperature and precipitation. However, the retrieved results for wind speed variables are inferior. This visualization results to some extent illustrate the negative impact of H8 observations on wind speed as shown in Tab. VI.

Further explanation for this can be analyzed from the imaging characteristics of the H8 satellite. The H8 bands we input include visible light, near-infrared, and infrared water vapor channels. The reflectance of these channels contains information related to surface reflectance, surface water vapor, and clouds. The surface pressure variable is most significantly influenced by air temperature and humidity, and they change relatively steadily over time. Therefore, rich texture information can be inverted from H8 observations. However, for wind speed, although our method inputs data from two adjacent frames of H8, hoping to indirectly obtain wind speed information through the movement of clouds, the high altitude of the clouds does not adequately represent surface wind speed. This leads to the inadequacy of H8 observations in capturing wind speed information.

H. Optimization process analysis

Since our method requires the simultaneous optimization of two losses at the grid scale and the station scale, we further discuss the trade-off between these two types of losses to analyze the impact of different losses and labels on the model optimization process. Both previous work [34] and the experimental results discussed above have adequately illustrated a substantial systematic bias between meteorological conditions at the station scale and the grid scale. Our approach uses the meteorological fields at the grid scale as input and, by modeling a continuous representation of the meteorological field, aims to obtain high-precision meteorological states at the station scale to alleviate this systematic bias. However, during the training process, as mentioned in Eq. 11, we introduced supervisory information from both the station scale and the grid scale and combined them in a weighted sum to serve as the objective function. This also creates a trade-off between the two different losses.

Fig. 7 displays the changes in the normalized MSE loss for the validation set at both the station scale and grid-scale during the training process. It should be noted that because we have incorporated strong prior information (such as interpolation results, etc.) as inputs into our model, the model is capable of achieving satisfactory convergence within just one epoch. As a result, the overall loss function appears relatively smooth. The curves in the figure indicate that for all models, the grid-scale loss decreases steadily with the progression of training. However, the station-scale loss is relatively more volatile and tends to first decrease and then increase as the number of training epochs increases (this is particularly evident for the models with $\beta = 0.05$). As β increases from 0.05 to 0.1, this phenomenon is somewhat mitigated, but there is a corresponding decline in performance at the grid scale. We believe that such results are primarily due to a certain level of discrepancy between the two types of losses, and the fact that there are fewer station-scale samples compared to grid-scale samples, leading to a degree of sample imbalance. Furthermore, the results in the table also reveal that high-resolution grid supervision has a significant impact on grid-scale performance, but the effect is relatively tolerable for station-scale predictions. Upon further analysis of cases with high-resolution station supervision, models with multi-block decoders exhibit a significant performance improvement at the grid scale compared to multi-variable decoders; however, this improvement is not as pronounced at the station scale. Additionally, the convergence speed of the former is noticeably superior to that of the latter.

From the analysis above, it is evident that for the novel task we proposed, it is unreasonable to focus solely on improving station-scale performance without considering the grid scale. Therefore, devising more rational model structures to further enhance the downscaling performance at both scales is an important research direction for the future.

VI. Discussion

The purpose of this paper is to downscale grid-scale meteorological field data to the scale of discrete scattered stations. This allows us to obtain the meteorological state at any location based on widely used coarse-resolution meteorological field data, which has significant practical importance. The new task proposed in this paper, along with the novel method HyperDS, integrates multi-scale observational information into the downscaling task, effectively achieving station-scale downscaling. However, there are still many aspects of this task

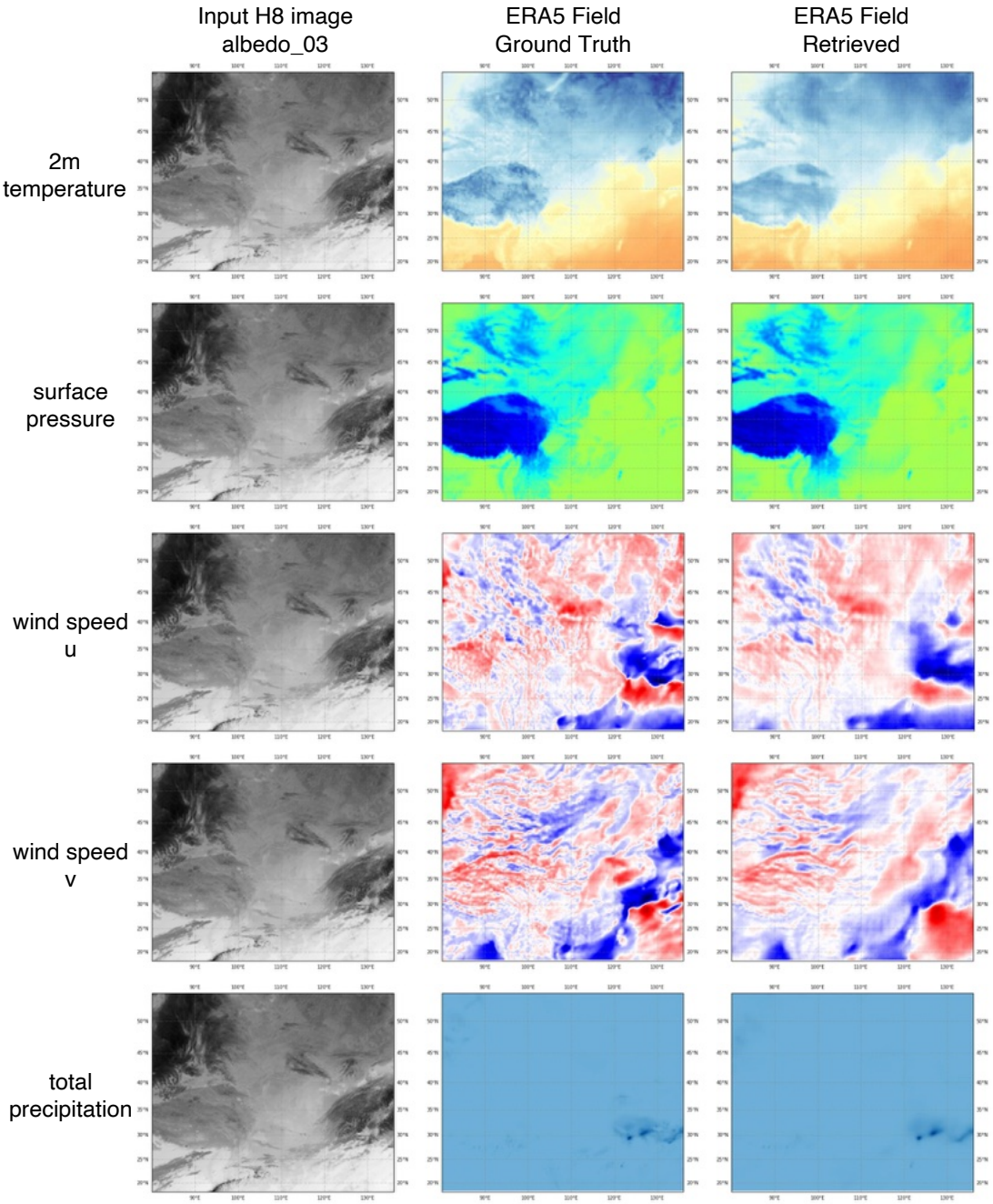
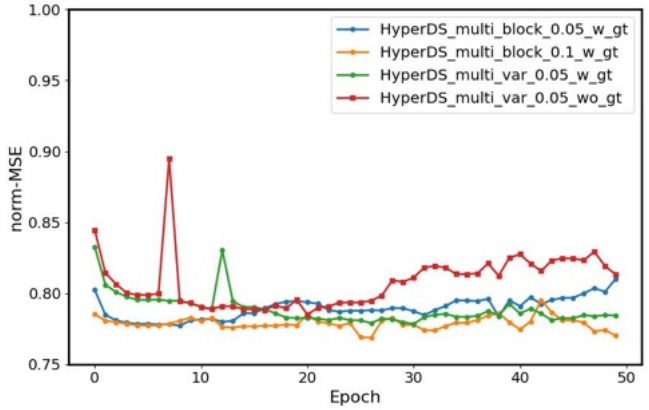


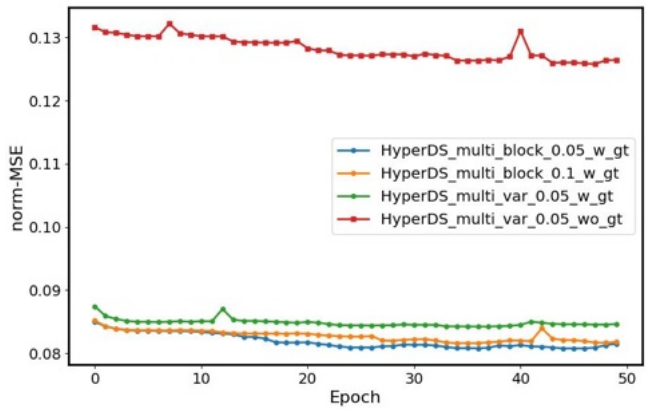
Fig. 6. Illustration of meteorological field retrieved based on Himawari-8 observations using a vanilla CNN-based encoder-decoder network. It can be seen that the retrieval results for temperature, pressure, and precipitation variables exhibit good detail in texture; however, it fails to recover high-frequency information for wind speed variables.

that can be further explored and researched. We will discuss the current issues and potential future directions from two perspectives: the data and the methodology.

1) Observation Data: Based on the experiments and analyses conducted, it is evident that incorporating multi-scale observational data plays a crucial role in the downscaling task. The dataset we currently propose includes only two types of observational information: station observations and geostationary satellite radiance values. Especially with regard to satellite observations, the data types contained within a single data source are quite limited. In the operational forecasting data assimilation process [62], multisource and multisensor remote sensing data are used to obtain different meteorological variables. Therefore, integrating more types of remote sensing observation data into the dataset is important for constructing a continuous-scale multivariate meteorological field. However, this task is also very challenging, involving specialized knowledge related to the sensors and corresponding meteorological variables, as well as complex data preprocessing procedures. In addition, the number of station data points we utilized is still relatively small. Previous work has used data from tens of thousands of stations to achieve high-accuracy station-



(a) Normalized MSE validation loss on station-level scale.



(b) Normalized MSE validation loss on grid-level scale.

Fig. 7. Illustration of the changes in the normalized MSE loss on the validation set under various hyperparameter settings during the training process, with 0.05 and 0.1 indicating the β values of the station loss in Eq. 11.

scale weather forecasts [34]. Incorporating more station data is also crucial for continuous-scale modeling. We also welcome more researchers to integrate more types of data into our task to enhance the capability of continuous-scale modeling for more meteorological variables, and we hope that related work can assist with the practical applications of meteorological forecasting.

2) Models and Methodology: Designing model architectures and methods suitable for the current task is also key to improving the performance of the task. Unlike traditional downscaling works based on super-resolution, downscaling to the scale of individual stations places higher demands on the model's ability to represent resolution continuously. The *HyperDS* proposed in this paper starts from this perspective and has achieved good performance in downscaling to the station scale. However, the current method sacrifices the performance of grid-scale downscaling to some extent in order to enhance station-scale downscaling, which is not a trade-off we desire to see. This is also a common problem with many super-resolution methods based on implicit neural representations. Therefore, how to design a more effective model that ensures multi-scale modeling accuracy while supporting continuous-resolution representation is an urgent issue to be addressed in the future.

VII. Conclusion

In this paper, we extend the traditional fixed-resolution grid-based downscaling task to the scale of scattered station scale to derive accurate surface meteorological states at arbitrary locations. To restore the meteorological state of the subgrid, we integrate multi-scale observational data into the downscaling process and propose a new model based on a hypernetwork structure called *HyperDS*. It uses high-resolution remote sensing images as prior input and scattered observation station data as station-scale labels. By continuously modeling the meteorological field with neural fields, it effectively integrates multi-scale observational information and achieves high-precision meteorological field downscaling at the station scale. Through extensive experimental comparisons with specially designed baseline methods, we have verified the effectiveness of our proposed approach, particularly in terms of performance on wind speed and surface pressure variables, where it significantly outperforms other methods. This paper represents the first exploration into observation-driven downscaling of meteorological fields to station scales. We hope that in the future, more researchers will build on this foundation to study more effective methods, enhancing the accuracy and capability of continuous meteorological field modeling.

References

[1] X. Ren, X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang, "Deep learning-based weather prediction: a survey," *Big Data Research*, vol. 23, p. 100178, 2021.

[2] S. K. Mukkavilli, D. S. Civitarese, J. Schmude, J. Jakubik, A. Jones, N. Nguyen, C. Phillips, S. Roy, S. Singh, C. Watson *et al.*, "Ai foundation models for weather and climate: Applications, design, and implementation," *arXiv preprint arXiv:2309.10808*, 2023.

[3] S. Chen, G. Long, J. Jiang, D. Liu, and C. Zhang, "Foundation models for weather and climate data understanding: A comprehensive survey," *arXiv preprint arXiv:2312.03014*, 2023.

[4] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3d neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, 2023.

[5] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu *et al.*, "Learning skillful medium-range global weather forecasting," *Science*, p. eadi2336, 2023.

[6] K. Chen, T. Han, J. Gong, L. Bai, F. Ling, J.-J. Luo, X. Chen, L. Ma, T. Zhang, R. Su *et al.*, "Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead," *arXiv preprint arXiv:2304.02948*, 2023.

[7] L. Chen, X. Zhong, F. Zhang, Y. Cheng, Y. Xu, Y. Qi, and H. Li, "Fuxi: A cascade machine learning forecasting system for 15-day global weather forecast," *arXiv preprint arXiv:2306.12873*, 2023.

[8] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.

[9] Y. Sun, K. Deng, K. Ren, J. Liu, C. Deng, and Y. Jin, "Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 208, pp. 14–38, 2024.

[10] A. J. Geer, "Learning earth system models from observations: machine learning or data assimilation?" *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200089, 2021.

[11] 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.

[12] P. Ren, N. B. Erichson, S. Subramanian, O. San, Z. Lukic, and M. W. Mahoney, "Superbench: A super-resolution benchmark dataset for scientific machine learning," *arXiv preprint arXiv:2306.14070*, 2023.

[13] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.

[14] Z. Xu, Y. Han, and Z. Yang, "Dynamical downscaling of regional climate: A review of methods and limitations," *Science China Earth Sciences*, vol. 62, pp. 365–375, 2019.

[15] N. Rampal, P. B. Gibson, A. Sood, S. Stuart, N. C. Fauchereau, C. Brandolino, B. Noll, and T. Meyers, "High-resolution downscaling with interpretable deep learning: Rainfall extremes over new zealand," *Weather and Climate Extremes*, vol. 38, p. 100525, 2022.

[16] L. Harris, A. T. McRae, M. Chantry, P. D. Dueben, and T. N. Palmer, "A generative deep learning approach to stochastic downscaling of precipitation forecasts," *Journal of Advances in Modeling Earth Systems*, vol. 14, no. 10, p. e2022MS003120, 2022.

[17] B. Kumar, K. Atey, B. B. Singh, R. Chattopadhyay, N. Acharya, M. Singh, R. S. Nanjundiah, and S. A. Rao, "On the modern deep learning approaches for precipitation downscaling," *Earth Science Informatics*, vol. 16, no. 2, pp. 1459–1472, 2023.

[18] F. Wang, D. Tian, L. Lowe, L. Kalin, and J. Lehrter, "Deep learning for daily precipitation and temperature downscaling," *Water Resources Research*, vol. 57, no. 4, p. e2020WR029308, 2021.

[19] Y. Sha, D. J. Gagne II, G. West, and R. Stull, "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part i: Daily maximum and minimum 2-m temperature," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 12, pp. 2057–2073, 2020.

[20] ——, "Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. part ii: Daily precipitation," *Journal of Applied Meteorology and Climatology*, vol. 59, no. 12, pp. 2075–2092, 2020.

[21] G. Liu, R. Zhang, R. Hang, L. Ge, C. Shi, and Q. Liu, "Statistical downscaling of temperature distributions in southwest china by using terrain-guided attention network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 1678–1690, 2023.

[22] K. Höhlein, M. Kern, T. Hewson, and R. Westermann, "A comparative study of convolutional neural network models for wind field downscaling," *Meteorological Applications*, vol. 27, no. 6, p. e1961, 2020.

[23] Y. Liu, A. R. Ganguly, and J. Dy, "Climate downscaling using ynet: A deep convolutional network with skip connections and fusion," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3145–3153.

[24] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, 2017, pp. 1663–1672.

[25] T. Vandal, E. Kodra, and A. R. Ganguly, "Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation," *Theoretical and Applied Climatology*, vol. 137, pp. 557–570, 2019.

[26] J. Wu, L. Xia, T. O. Chan, J. Awange, and B. Zhong, "Downscaling land surface temperature: A framework based on geographically and temporally neural network weighted autoregressive model with spatio-temporal fused scaling factors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 187, pp. 259–272, 2022.

[27] M. van der Meer, S. de Roda Husman, and S. Lhermitte, "Deep learning regional climate model emulators: A comparison of two downscaling training frameworks," *Journal of Advances in Modeling Earth Systems*, vol. 15, no. 6, p. e2022MS003593, 2023.

[28] X. Zhong, F. Du, L. Chen, Z. Wang, and H. Li, "Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind speed forecast," *Quarterly Journal of the Royal Meteorological Society*.

[29] M. Mardani, N. Brenowitz, Y. Cohen, J. Pathak, C.-Y. Chen, C.-C. Liu, A. Vahdat, K. Kashinath, J. Kautz, and M. Pritchard, "Generative residual diffusion modeling for km-scale atmospheric downscaling," *arXiv preprint arXiv:2309.15214*, 2023.

[30] J. Leinonen, D. Nerini, and A. Berne, "Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 9, pp. 7211–7223, 2020.

[31] N. J. Annau, A. J. Cannon, and A. H. Monahan, "Algorithmic hallucinations of near-surface winds: Statistical downscaling with generative adversarial networks to convection-permitting scales," *Artificial Intelligence for the Earth Systems*, vol. 2, no. 4, p. e230015, 2023.

[32] Y. Li, H. Wu, H. Chen, and X. Zhu, "A robust framework for resolution enhancement of land surface temperature by combining spatial downscaling and spatiotemporal fusion methods," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[33] 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.

[34] H. Wu, H. Zhou, M. Long, and J. Wang, "Interpretable weather forecasting for worldwide stations with a unified deep model," *Nature Machine Intelligence*, pp. 1–10, 2023.

[35] K. Bessho, K. Date, M. Hayashi, A. Ikeda, T. Imai, H. Inoue, Y. Kumagai, T. Miyakawa, H. Murata, T. Ohno *et al.*, "An introduction to himawari-8/9—japan's new-generation geostationary meteorological satellites," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 94, no. 2, pp. 151–183, 2016.

[36] X. Zhu, Y. Xiong, M. Wu, G. Nie, B. Zhang, and Z. Yang, "Weather2k: A multivariate spatio-temporal benchmark dataset for meteorological forecasting based on real-time observation data from ground weather stations," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2023, pp. 2704–2722.

[37] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar, "Neural fields in visual computing and beyond," in *Computer Graphics Forum*, vol. 41, no. 2. Wiley Online Library, 2022, pp. 641–676.

[38] K. Chen, W. Li, S. Lei, J. Chen, X. Jiang, Z. Zou, and Z. Shi, "Continuous remote sensing image super-resolution based on context interaction in implicit function space," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[39] V. K. Chauhan, J. Zhou, P. Lu, S. Molaei, and D. A. Clifton, "A brief review of hypernetworks in deep learning," *arXiv preprint arXiv:2306.06955*, 2023.

[40] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *Ieee Access*, vol. 9, pp. 82 031–82 057, 2021.

[41] K. Fukami, K. Fukagata, and K. Taira, "Super-resolution analysis via machine learning: a survey for fluid flows," *Theoretical and Computational Fluid Dynamics*, pp. 1–24, 2023.

[42] S. Esmaeilzadeh, K. Azizzadenesheli, K. Kashinath, M. Mustafa, H. A. Tchelepi, P. Marcus, M. Prabhat, A. Anandkumar *et al.*, "Meshfreeflownet: A physics-constrained deep continuous space-time super-resolution framework," in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, 2020, pp. 1–15.

[43] W. Li, Z. Liu, K. Chen, H. Chen, S. Liang, Z. Zou, and Z. Shi, "Deephysinet: Bridging deep learning and atmospheric physics for accurate and continuous weather modeling," *arXiv preprint arXiv:2401.04125*, 2024.

[44] S. Hagemann and L. D. Gates, "Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations," *Climate Dynamics*, vol. 21, pp. 349–359, 2003.

[45] S. Gao and X. Zhuang, "Multi-scale deep neural networks for real image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[46] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[47] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[48] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[49] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[50] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[51] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.
[52] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, "Transformer for single image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 457–466.
[53] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5791–5800.
[54] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8628–8638.
[55] Z. Chen, Y. Chen, J. Liu, X. Xu, V. Goel, Z. Wang, H. Shi, and X. Wang, "Videoinr: Learning video implicit neural representation for continuous space-time super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 2047–2057.
[56] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Science Reviews*, vol. 232, p. 104110, 2022.
[57] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local–global combined network," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1243–1247, 2017.
[58] ——, "Coupled adversarial training for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3633–3643, 2019.
[59] K. Zhang, D. Zhu, X. Min, and G. Zhai, "Implicit neural representation learning for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2022.
[60] H. Wu, N. Ni, and L. Zhang, "Learning dynamic scale awareness and global implicit functions for continuous-scale super-resolution of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.
[61] D. Ha, A. Dai, and Q. V. Le, "Hypernetworks," *arXiv preprint arXiv:1609.09106*, 2016.
[62] J. Eyre, W. Bell, J. Cotton, S. English, M. Forsythe, S. Healy, and E. Pavelin, "Assimilation of satellite data in numerical weather prediction. part ii: Recent years," *Quarterly Journal of the Royal Meteorological Society*, vol. 148, no. 743, pp. 521–556, 2022.
[63] S. Cheng, C. Quilodrán-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandeo, R. Fablet, D. Lucor, B. Iooss, J. Brajard *et al.*, "Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 6, pp. 1361–1387, 2023.
[64] K. Chen, L. Bai, F. Ling, P. Ye, T. Chen, K. Chen, T. Han, and W. Ouyang, "Towards an end-to-end artificial intelligence driven global weather forecasting system," *arXiv preprint arXiv:2312.12462*, 2023.
[65] S. Rasp, P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid, and N. Thuerey, "Weatherbench: a benchmark data set for data-driven weather forecasting," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 11, p. e2020MS002203, 2020.
[66] S. Rasp, S. Hoyer, A. Merose, I. Langmore, P. Battaglia, T. Russel, A. Sanchez-Gonzalez, V. Yang, R. Carver, S. Agrawal *et al.*, "Weatherbench 2: A benchmark for the next generation of data-driven global weather models," *arXiv preprint arXiv:2308.15560*, 2023.
[67] R. Swinbank, M. Kyouda, P. Buchanan, L. Froude, T. M. Hamill, T. D. Hewson, J. H. Keller, M. Matsueda, J. Methven, F. Pappenberger *et al.*, "The tigge project and its achievements," *Bulletin of the American Meteorological Society*, vol. 97, no. 1, pp. 49–67, 2016.
[68] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[69] T. R. Shaham, M. Gharbi, R. Zhang, E. Shechtman, and T. Michaeli, "Spatially-adaptive pixelwise networks for fast image translation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 882–14 891.
[70] C. Rao, P. Ren, Q. Wang, O. Buyukozturk, H. Sun, and Y. Liu, "Encoding physics to learn reaction–diffusion processes," *Nature Machine Intelligence*, vol. 5, no. 7, pp. 765–779, 2023.
[71] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[72] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
[73] O. Ronneberger, P. Fischer, and T. Brox, "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*. Springer, 2015, pp. 234–241.
[74] P. Hess, M. Drüke, S. Petri, F. M. Strnad, and N. Boers, "Physically constrained generative adversarial networks for improving precipitation fields from earth system models," *Nature Machine Intelligence*, vol. 4, no. 10, pp. 828–839, 2022.

Zili Liu (Graduate Student Member, IEEE) received the B.S. and M.S. degree from the School of Science, China University of Mining and Technology, Beijing in 2017 and 2021. He is currently working toward the Phd degree in the Image Processing Center, School of Astronautics, Beihang University.

His research interests include deep learning, AI for meteorology, remote sensing image processing, and efficient deep learning.

Hao Chen received the B.S. and Ph.D. degrees from the Image Processing Center, School of Astronautics, Beihang University, Beijing, China, in 2017 and 2023, respectively.

He is currently a Researcher at Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. His research interests include geospatial machine learning, remote sensing, Earth monitoring, and prediction

Lei Bai received the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2021.

He was a Post-Doctoral Researcher with the University of Sydney, Camperdown, NSW, Australia. He is currently a Research Scientist with Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. He has authored or co-authored a set of peer-reviewed papers in top AI conferences and journals, such as Neural Information Processing Systems (NeurIPS), Conference on Computer Vision and Pattern Recognition (CVPR), International Joint Conference on Artificial Intelligence (IJCAI), Knowledge Discovery and Data Mining (KDD), International Conference on Computer Vision (ICCV), International Conference on Ubiquitous Computing (Ubicomp), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (TITS). His research interests include machine learning, spatial-temporal learning, and their applications (e.g., Earth System Science and Smart City).

Dr. Bai is or was a Program Committee Member or Reviewer for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, NeurIPS, International Conference on Machine Learning (ICML), International Conference on Learning Representations (ICLR), CVPR, ICCV, Association for the Advancement of Artificial Intelligence (AAAI), IJCAI, KDD, European Conference on Computer Vision (ECCV), IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, and ACM Transactions on Sensor Networks. He was a recipient of the 2020 Google Ph.D. Fellowship, the 2020 UNSW Engineering Excellence Award, and the 2021 Dean's Award for Outstanding Ph.D. Theses.

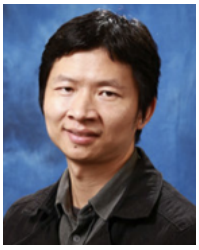
Wenyuan Li received the B.S. degree from North China Electric Power University, Beijing, China, in 2017, and the Ph.D. degree from Beihang University, Beijing, in 2023.

He is currently a Post-Doctoral Fellow with the Department of Geography, The University of Hong Kong, Hong Kong. His research interests encompass remote sensing image processing and agricultural crop mapping.

Keyan Chen (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2019 and 2022, respectively, where he is currently pursuing the Ph.D. degree with the Image Processing Center.

His research interests include remote sensing image processing, deep learning, pattern recognition, and multimodal. His personal website is https://kyanchen.github.io/.

Zhengyi Wang Zhengyi Wang received the B.S. and M.S. degree from China University of Petroleum (Beijing) and Ocean University of China in 2017 and 2020. He is currently working toward the Phd degree in the School of Oceanography, Shanghai Jiao Tong University. His research interests include remote sensing, artificial intelligence and marine science.



Wanli Ouyang received the Ph.D. degree from the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, in 2010.

He was an Associate Professor with The University of Sydney, Camperdown, NSW, Australia. He is now a Professor with Shanghai Artificial Intelligence (AI) Laboratory, Shanghai, China. His research interests include pattern recognition, machine learning, and AI for Science.

Dr. Ouyang served as an Associate Editor for International Journal of Computer Vision (IJCV) and Pattern Recognition (PR), the Senior Area Chair for Conference on Computer Vision and Pattern Recognition (CVPR), and the Guest Editor for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI).

Zhengxia Zou (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Beihang University, Beijing, China, in 2013 and 2018, respectively. He is currently a Professor with the School of Astronautics, Beihang University. From 2018 to 2021, he was a Postdoctoral Research Fellow at the University of Michigan, Ann Arbor, MI, USA. His research interests include remote sensing image processing and computer vision. He has published more than 60 peer-reviewed papers in top-tier journals and conferences, including Nature Communications, Proceedings of the IEEE, IEEE Transactions on Image Processing, IEEE Transactions on Geoscience and Remote Sensing, and IEEE/CVF Computer Vision and Pattern Recognition. Dr. Zou serves as an Associate Editor for IEEE Transactions on Image Processing (TIP). His personal website is https://zhengxiazou.github.io/.

Zhenwei Shi (Senior Member, IEEE) is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing, China. He has authored or co-authored over 200 scientific articles in refereed journals and proceedings. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Prof. Shi serves as an Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, Pattern Recognition, ISPRS Journal of Photogrammetry and Remote Sensing, Infrared Physics and Technology, and so on. His personal website is http://levir.buaa.edu.cn/.