

Research Article

A Climate Downscaling Deep Learning Model considering the Multiscale Spatial Correlations and Chaos of Meteorological Events

Bin Mu , Bo Qin , Shijin Yuan , and Xiaoyun Qin 

School of Software Engineering, Tongji University, Shanghai 100052, China

Correspondence should be addressed to Shijin Yuan; yuanshijin2003@163.com

Received 30 July 2020; Revised 27 September 2020; Accepted 20 October 2020; Published 16 November 2020

Academic Editor: Rafal Zdunek

Copyright © 2020 Bin Mu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Climate downscaling is a way to provide finer resolution data at local scales, which has been widely used in meteorological research. The two main approaches for climate downscaling are dynamical and statistical. The traditional dynamical downscaling methods are quite time- and resource-consuming based on general circulation models (GCMs). Recently, more and more researchers construct a statistical deep learning model for climate downscaling motivated by the single-image superresolution (SISR) process in computer vision (CV). This is an approach that uses historical climate observations to learn a low-resolution to high-resolution mapping and produces great enhancements in terms of efficiency and effectiveness. Therefore, it has provided an appreciable new insight and successful downscaling solution to multiple climate phenomena. However, most existing models only make a simple analogy between climate downscaling and SISR and ignore the underlying dynamical mechanisms, which leads to the overaveraged downscaling results lacking crucial physical details. In this paper, we incorporate the a priori meteorological knowledge into a deep learning formalization for climate downscaling. More specifically, we consider the multiscale spatial correlations and the chaos in multiple climate events. Depending on two characteristics, we build up a two-stage deep learning model containing a stepwise reconstruction process and ensemble inference, which is named climate downscaling network (CDN). It can extract more local/remote spatial dependencies and provide more comprehensive captures of extreme conditions. We evaluate our model based on two datasets: climate science dataset (CSD) and benchmark image dataset (BID). The results demonstrate that our model shows the effectiveness and superiority in downscaling daily precipitation data from 2.5 degrees to 0.5 degrees over Asia and Europe. In addition, our model exhibits better performance than the other traditional approaches and state-of-the-art deep learning models.

1. Introduction

Climate variations are influencing society's well-being all over the world, such as global warming, extreme storm, precipitation, and sea-level rising [1–3]. A lot of industrial and agricultural products have been severely affected by these climate changes with increasing intensity, duration, and frequency [4]. In recent years, climate researchers are looking for more reliable climate simulation methods, which can provide a more detailed evolution process of meteorological events, for more effective disaster mitigation and decision-making [5].

General circulation models (GCMs) are the traditional tools for simulating and investigating the various climate events [6, 7]. They are usually developed on the basis of

complicated physical dynamical principles and executed on large-scale supercomputers. These GCMs usually take a large number of physical variables into account and model the entire climate evolution processes into specific differential equations. These variables are treated as initial fields and have specific spatial resolutions, respectively, so that the meteorological researchers can investigate and forecast the potential trends that rely on these physical patterns with various resolutions under model integration. However, when simulating and studying the meteorological evolutions in small-scale areas, one may encounter the issues of too coarse spatial resolutions in GCMs, which limit revealing critical detailed physical processes, such as precipitations in key river basins and extreme typhoons. This severely hinders the assessment of the impact of climate events on these relevant key regions [8].

In this situation, climate downscaling has always been the hotspot in meteorological research, which aims to reconstruct more detailed and finer resolution meteorological data for small/medium scale regions. Technically, climate downscaling is used to enlarge the resolutions and elaborate the detailed contents for physical variables in climate events. There are two classic approaches to solve this problem: dynamical downscaling and statistical downscaling. Dynamical downscaling accounts for developing high-resolution GCMs or regional climate numerical models (RCMs) to simulate local meteorological processes. It usually includes embedding subgrid of key regions into GCM or using the output of GCM as the boundary conditions in RCMs for high-resolution projections. This is an advantage that dynamical downscaling has clear physical causalities and usually covers any regions even the entire Earth. However, limited to construction principles of GCMs or RCMs, developing and integrating the dynamical downscaling may need a large amount of calculation and time resources; in addition, modifying and configuring GCMs into a certain resolution are not convenient.

Statistical downscaling is a data-driven method and learns the projection relationship from low-resolution (LR) to high-resolution (HR) climate patterns via historical meteorological big data. A large number of existing studies have suggested that statistical downscaling has more efficient computation and almost equally or even more accurate downscaling results [9, 10], which is a reliable alternative. Therefore, climate researchers are paying more attention to it recently.

Statistical downscaling has been implemented by a variety of numerical calculation methods in wide scope; one of the most important ideas is to use regression to simulate the mapping projection including linear and nonlinear ways. Traditional linear regression makes preliminary progress in terms of downscaling of sea temperature and precipitation in early stage [11]. Subsequently, many studies focus on the nonlinear evolution in climate events and extend linear to nonlinear (such as artificial neural network (ANN)) approach based on machine learning to reduce downscaling errors. These nonlinear methods, including quantile regression neural networks [12], Bayesian model averaging [13], and expanded downscaling [14], can capture a more general relationship between LR and HR, so they achieve better results.

The above-mentioned statistical methods are efficient in climate downscaling. However, limited to simplified formalization in mathematics, the existing statistical downscaling can only learn the most superficial correlations contained in meteorological data but overlooks potential physical details, which causes the unreasonable downscaling results, such as overaveraging results or even describing the extreme events' failures [15]. In meteorological research, the extreme behaviors in the Earth system should be paid more attention to than normal events [1, 8], such as typhoons, unusual floods, and strong El Niño events [16], because extreme events may cause more severe damages. Therefore, there is a special requirement for climate downscaling:

reconstructing more accurate climate data for reflecting the truly detailed scenarios, instead of simple interpolation.

With the development of computing equipment, deep learning methods, as the extension of ANN, provide an appreciable new insight in many frontier areas, which inspire climate researchers to explore suitable alternatives in the field of deep learning computer vision (CV) and effectively transfer these ideas to climate downscaling. More specifically, single-image superresolution (SISR) problems in CV are to raise the image resolution for providing clearer and more comfortable pictures for human eyes. In terms of such purpose, SISR problem is quite similar to climate downscaling: they both learn a projection relationship between low-resolution (LR) and high-resolution (HR) data. SR is based on image data; climate downscaling relies on climate observed or simulated data. But the meteorological data has a natural property; that is, they are usually available on grid points. Therefore, we can make an analogy between SR and climate downscaling: treating the grid points in a climate data as pixels in an image and then modeling climate downscaling as enlarging resolution in SR via deep learning CV techniques [15].

Recently, a large number of established benchmark image databases and SR competitions have significantly promoted the development of SR problem in CV, so that we can absorb these successful experiences and apply them to climate downscaling. The first successful deep learning model was established by Dong et al. [17], who proposed a three-layer SRCNN and achieved superior performance than conventional methods. Then the SISR models are constructed larger scale, deeper and deeper by introducing residual learning [18] and dense connection [19], such as 20 layers in VDSR [20] and DCRN [21]. In addition, more and more CNN-based models improved the pioneering works and achieved better results, such as applying subpixel convolution [22], generative adversarial network [23], and Charbonnier penalty function [24] in the model structures to enhance model expression and ease model training. In general, deep learning methods are relatively easier to construct than dynamical downscaling and they can make up the shortcomings of execution time and computing resources. Furthermore, many latest deep learning technologies are also convenient to use in SR, which are more and more conducive to improving performance, such as dense connection and residual learning. These studies imply that deep learning can be a reliable tool in climate investigation, and have great potential in climate downscaling.

However, the relationship between climate downscaling and SR is not exactly the simple analogy. SR requires the modulated images to improve viewing comfort, including higher resolutions, adaptive contrast, and edges. On the contrary, climate downscaling requires absolutely true physical details and dynamic features that cannot be described well at low resolution, which may even generate "extra" spatial information. To tackle this challenge, we can absorb some general guidelines in the SR model and then combine them with specific concepts of meteorological knowledge to build up a more comprehensive formalization

for climate downscaling problem, which can increase the model interpretability and improve downscaling skills.

In this paper, we incorporate the a priori meteorological knowledge into the deep learning modeling process for climate downscaling. The multiscale spatial correlations and the chaos are the two key concepts in multiple climate events. We take both into account and novelly formulate climate downscaling as a two-stage process. In the first stage, the same LR climate data will be reconstructed to multiple same-size HR data considering multiple spatial scale correlations, which will generate many HR candidates. In the second stage, we will evaluate the reconstruction uncertainties of these HR candidates and then infer the optimal probability distribution of the results, which can provide the optimal mean and corresponding variance at the same time. Based on such formalization, we further implement a more detailed model named climate downscaling network (CDN), which implements the above-mentioned two stages.

To the best of our knowledge, it is the first time to integrate meteorological knowledge into deep learning in climate downscaling. We train and evaluate our proposed model on the real-world climate science dataset (CSD) and benchmark image dataset (BID), respectively, on the metrics of SSIM and PSNR. We carry out comparative experiments to determine a better model structure which can make a trade-off between performance and speed, and then we design several different magnification experiments with our two datasets and conduct extensive experiments to evaluate the performance of our model. From CSD experiments, our model shows the effectiveness and superiority in downscaling daily precipitation data from 2.5 degrees to 0.5 degrees over Asia and Europe. In addition, we also discover that what our model learns is not a fixed resolution mapping from LR to HR but the enlarged size, which means that our model is not limited by the size of input or output and has good extensibility for transfer learning [25]. As for BID experiments, our model has excellent performance, which outperforms than the other traditional approaches and state-of-the-art models, especially at the small magnifications.

2. Problem Formalization and Context

2.1. Formalization. Climate events can be treated as very complicated hydrodynamic phenomena, where different coupled physical variables develop under deterministic dynamic mechanisms. Climate downscaling is to enhance the resolution of a certain physical variable in the same area and provide more detailed forecast results. In other words, climate downscaling is to supplement the data on the finer grid points by extracting and learning the underlying physical laws of the original data.

Spatial correlation plays an important role in climate downscaling, which is an important internal characteristic. Increasing evidence suggests the remote correlations produce climate variabilities. Rong et al. [26] found that the positive sea surface temperature anomalies (SSTA) in the northern tropical Atlantic (NTA) can trigger the eastward Kelvin wave which induces low-level anticyclone over the western North Pacific during boreal summer. Rueda et al.

[27] found that the waves have great contributions to extreme sea-level fluctuations, for both local and remote windstorms. In general, climate events in the Earth system are not isolated; because multiples atmospheric circulations or ocean waves building up a bridge span the ocean (or land) and atmosphere, the coupled interactions connect all climate variables spatially together regardless of adjacency correlations or remote correlations.

However, such spatial correlations are not fully utilized in the climate downscaling problem. Traditional regression approaches and CNN filters usually extract local information in continuous spaces to interpolate extra grid points, which is confined to adjacency neighborhood, and tend to average the downscaling results; the remote meteorological spatial correlations are usually ignored, which makes the results lacking credible dynamic causality and interpretation.

From the perspective of climate dynamics, the interpolated details are mainly determined by local scales information or the so-called adjacency correlations; the large-scale information may also contain the potential trends, such as remote waves or wind forcing. So if multi-scale correlations can be taken into account in the downscaling model, the performance will be greatly improved. In addition, the chaos of the dynamic system is inevitable; meteorological researchers are paying more attention to describing the uncertainties in climate forecasting, leading to a turning from determinant single-value predictions to probabilistic predictions, such as ensemble prediction [28, 29]. Therefore, we will take good advantage of probabilistic predictions and provide more comprehensive downscaling.

On the basis of these two pieces of a priori meteorological knowledge, we formulate a 2-stage process for climate downscaling. The first stage (stage 1) is to reconstruct HR climate data considering different multiscale spatial correlations and obtain several HR candidates. The second stage (stage 2) is to infer the optimal probabilistic distributions for downscaling results via the obtained candidates from stage 1, providing the optimal mean and corresponding variance at the same time (see Figure 1).

Stage 1. In general, small-scale spatial correlation can reflect more adjacency information; large-scale spatial correlation contains more remote interactions. Therefore, to cover as many different spatial correlations as possible, we design a structure containing N path to reconstruct HR data, where every path extracts and learns different scale spatial correlations. Finally, we can obtain N HR candidates.

In every reconstruction process of these N paths, we propose the stepwise reconstruction process to gradually enlarge the climate data. Classic methods usually use the upsampling or bicubic interpolation to enlarge the data by integer multiples directly, but these operations are done in one step and always limited in integer multiples, such as from an original size 40×60 to a required size 80×120 ($2 \times$ magnification), which may fill too much content at one time and cause too large space skips. In fact, climate downscaling may have a more flexible magnification size according to requirement, which is usually not an integer-multiple

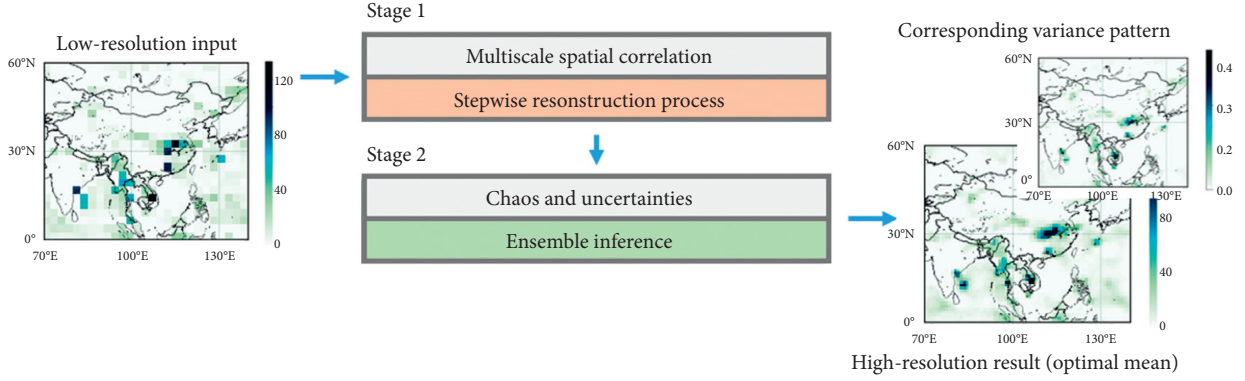


FIGURE 1: Our proposed two-stage formalization, considering the multiscale spatial correlations and chaos in multiple climate events.

amplification, such as from an original size 40×60 to a required size 50×80 , in which aspect traditional methods have limitations. Therefore, in every reconstruction path, we divide the entire process into gradually enlarged subprocesses, which can fully extract the spatial correlations of different scales.

More specifically, suppose the original climate data size is $i \times j$ and the required data size is $i' \times j'$ in a 2D downscaling scenario (channel size is 1 for expression simplification):

$$\begin{cases} \Delta h = i' - i, \\ \Delta w = j' - j, \end{cases} \quad (1)$$

where Δh and Δw are the resolution increments, respectively, for each edge of original climate data. According to equation (1), we divide Δh and Δw into M steps and enlarge the data size with $(n_{(m)}^h, n_{(m)}^w)$ in each step as depicted in

$$\begin{cases} \sum n_{(m)}^h = \Delta h \\ \sum n_{(m)}^w = \Delta w \\ y_{(m+1)} = g_{(n_{(m)}^h, n_{(m)}^w)}(y_{(m)}) \end{cases}, \quad (m = 1, 2, \dots, M), \quad (2)$$

where $y_{(m)}$ and $y_{(m+1)}$ are the input and output data of each step, respectively, and g is the magnification operator related to enlarged sizes $(n_{(m)}^h, n_{(m)}^w)$. For example, the input size for $y_{(m)}$ is $i_{(m)} \times j_{(m)}$ and the output size for $y_{(m+1)}$ is $(i_{(m)} + n_{(m)}^h) \times (j_{(m)} + n_{(m)}^w)$. This process is described in Figure 2, where (a) and (e), (b) and (f), and (c) and (g) show the processes for $M = 1, 2, 3$, respectively; (d) represents the process of dividing M into infinite magnifications, which is similar to calculus. We design a rule to deduce the enlarged sizes $(n_{(m)}^h, n_{(m)}^w)$ to guarantee them as equal as possible during these M steps as in equation (3). In such a process, $\lfloor \Delta h/M \rfloor$ or $\lceil \Delta h/M \rceil$ represents every enlarged size $n_{(m)}^h$ and the same for $n_{(m)}^w$.

$$\begin{cases} \Delta h = s_h \times \lfloor \Delta h/M \rfloor + t_h \times \lceil \Delta h/M \rceil, \\ \Delta w = s_w \times \lfloor \Delta w/M \rfloor + t_w \times \lceil \Delta w/M \rceil, \\ M = s_h + t_h = s_w + t_w. \end{cases} \quad (3)$$

For a certain climate downscaling problem, Δh and Δw are fixed values; the magnification factors $(n_{(m)}^h, n_{(m)}^w)$ and step size M are negatively related to each other; for example, when M is very small, $(n_{(m)}^h, n_{(m)}^w)$ are relatively large, but when M is very large, $(n_{(m)}^h, n_{(m)}^w)$ are relatively small. In addition, the smaller $(n_{(m)}^h, n_{(m)}^w)$ means the size change of climate data is not obvious, which leads the climate data to tend to project more on to the adjacency spatial correlations and learn more local information to fill the grid point of expanded resolution and vice versa. As we use N different paths to reconstruct the HR data and set as many different decomposed steps M as possible in every path where $(n_{(m)}^h, n_{(m)}^w)$ are totally different, we can fully demonstrate the adjacency and remote spatial interactions in the whole process. Figure 3 displays the detailed stepwise reconstruction structure in each path.

Stage 2. In meteorological research, the forecasts are more willing to be expressed in the form of a probability distribution. On the one hand, this is a more rigorous choice to measure the uncertainties of forecast results. On the other hand, climate events often imply potential extreme conditions, which cannot be ignored. Therefore, a large number of studies use ensemble forecasts to describe climate evolutions.

Taking the ensemble idea into account, we build up a probabilistic distribution for obtained downscaling candidates from stage 1 as

$$[\hat{y}, \sigma^2] = \mathcal{F}([y_1, y_2, \dots, y_N]), \quad (4)$$

where \hat{y} and σ^2 are the optimal mean and variance for the downscaling results, $[y_1, y_2, \dots, y_N]$ represents the N HR candidates from stage 1, and \mathcal{F} represents the downscaling ensemble system.

Data-driven methods, especially deep learning, show excellent performance in discovering the intricate nonlinear characteristics. Therefore, we construct a detailed deep learning architecture to approximate g and \mathcal{F} depicted in the following sections. This novel architecture is composed of two corresponding stages. The first stage is to reconstruct downscaling data from the perspective of multiscale spatial

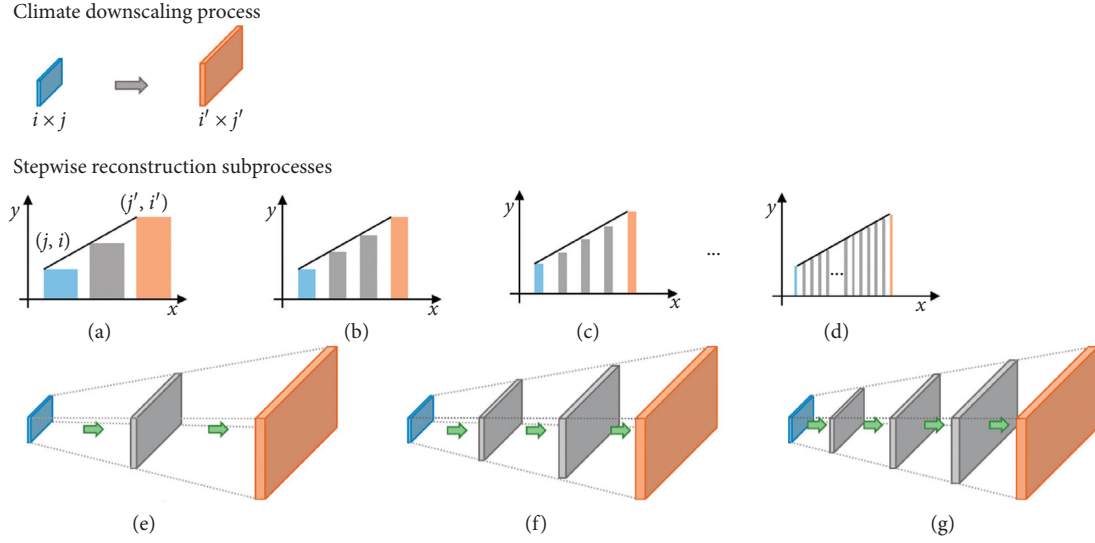


FIGURE 2: Some examples of the stepwise reconstruction process. The upper row means the original task: enlarging the data from $i \times j$ to $i' \times j'$. The lower row means dividing the entire process into M steps. $M = 2$ for (a) and (e); $M = 3$ for (b) and (f); $M = 4$ for (c) and (g). (d) represents the process of dividing M into infinite magnifications.

Original size	Stepwise reconstruction subprocesses						Downscaling size
$i \times j$	$(i \times \Delta h/2) \times (j \times \Delta w/2)$						$i' \times j'$
	$(i + \Delta h/3) \times (j + \Delta w/3)$			$(i + 2\Delta h/3) \times (j + 2\Delta w/3)$			$i' \times j'$
	$(i + \Delta h/4) \times (j + \Delta w/4)$		$(i + 2\Delta h/4) \times (j + 2\Delta w/4)$		$(i + 3\Delta h/4) \times (j + 3\Delta w/4)$		$i' \times j'$
	⋮						⋮
	$(i + \Delta h/n)$ × $(j + \Delta w/n)$	$(i + 2\Delta h/n)$ × $(j + 2\Delta w/n)$	$(i + 3\Delta h/n)$ × $(j + 3\Delta w/n)$	$(i + 4\Delta h/n)$ × $(j + 4\Delta w/n)$...	$(i + ((n-1)\Delta h)/n)$ × $(j + ((n-1)\Delta w)/n)$	$i' \times j'$

FIGURE 3: The detailed process for an N path reconstruction based on the stepwise reconstruction process. Every row in the green block represents a path, where each box means the enlarged size in this step. After reconstructing in N paths, we can collect N HR candidates shown in red block.

correlation, and the second stage is to approximate the mean and variance for the obtained candidates and provide the optimal result.

2.2. Data Collection and Application Scenario. We use a deep learning network to implement the magnification operator g and downscaling ensemble system \mathcal{F} . When the deep learning network structure is determined, the training dataset quality directly affects the downscaling performance. From the perspective of meteorology, multiresolution datasets are easily available. One of the ways is to collect real-world observation data on the same area and time period with different resolutions as a training set; another is to use simulation data by adjusting the resolution setting of a certain numerical climate model. No matter in which way, the LR climate data can be treated as the input of the deep learning model, and the corresponding HR climate data can be applied to supervised learning.

However, one may also encounter the situation that there is a single source for a certain physical variable, which only provides one fixed resolution climate data. It is difficult to obtain adequate data. In this situation, we can draw on the idea of SISR problem to augment the dataset in three ways. (1) Scaling: randomly downscale the original climate data between $[0.5, 1]$. (2) Rotation: randomly rotate image by 90° , 180° , or 270° . (3) Flipping: flip images horizontally or vertically with a probability of 0.5. The expended dataset can be used as HR data, then generating the LR training set using the bicubic downsampling method. It is worth noting that what the model actually learns via such LR-HR data pair is more similar to the inverse process of the downsampling kernel (such as the inverse of the bicubic kernel) [30]. Therefore, it is a better choice to use the real-world or simulated LR-HR climate data to train the network.

Furthermore, our model is very expandable. What the model learns is not a fixed resolution mapping from LR to HR, but the enlarged size, which can be applied to any

resolution enlargement. For example, the model is trained for enlarging (30×40) to (50×70) with downscaling size $(+20, +30)$; then it can be reused for enlarging (20×30) to (40×60) by transfer learning. This approach greatly reduces the difficulty of climate downscaling in similar situations.

3. Methodology

Based on the proposed architecture, we design a more detailed network for providing improved climate downscaling, named climate downscaling network (CDN). The structures and intentions of the two-stage structure are interpreted in the following sections.

3.1. stage 1: Stepwise Reconstruction Process. According to our proposed formalization, we build up a deep learning model which contains N different paths, and every path enlarges the LR data in M steps gradually. It will reconstruct and generate N HR candidates. The network structure is designed as in Figure 4.

The input is the original climate data or climate data after some preprocess operations, such as min-max scale, standardization, and normalization. After the climate data passed into network, it is assigned to N different paths, respectively. In each path, the original climate data will be expended in M steps gradually to the required size, which is called the stepwise reconstruction process as stair-like green blocks in Figure 4.

In this structure, the path number N and corresponding step number M are crucial for model structure and downscaling performance, which will influence the change sizes $(n_{(m)}^h, n_{(m)}^w)$ directly. Suppose the climate data sizes for HR and LR are $i' \times j'$ and $i \times j$, respectively; we apply three effective strategies for determining the path number N and step number M in every path. ① Nature-strategy (NS): the step size in the first path is 1, and the step size between adjacent paths is increased by 1 from 1 to $\min((i' - i), (j' - j))$. ② Even (Odd)-strategy (ES or OS): the step size in the first path is 1 (2), and the step size between adjacent levels is increased by 2 from 1 (2) to $\min((i' - i), (j' - j))$. ③ Prime-strategy (PS): choose all prime numbers between 3 and $\min((i' - i), (j' - j))$ as the step size for every path. An example is shown in Figure 5. In the following experiments, we will verify the pros and cons of these three strategies.

The most import module in the stepwise reconstruction process is the green block in Figure 4, which is used to enlarge the feature maps by $(n_{(m)}^h, n_{(m)}^w)$. As our proposed formalization, change size $(n_{(m)}^h, n_{(m)}^w)$ may be small. Therefore, we propose a special dense block to configurably enlarge the size of feature map by $(n_{(m)}^h, n_{(m)}^w)$ described in Figure 6. In this block, the input feature map is firstly expended by c convolution layers as χ for feature extraction; then the channel-expended feature map χ is connected to a k -stair-like structure, which has the following characteristics:

- (1) Every stair contains one convolution layer and one final dilated transformed convolution layer.
- (2) The first stair uses feature map χ as input. The other stairs concatenate the last stair's input and the output of the convolutional layer and then use it as the input of this layer.

After all stairs enlarge the feature maps' size by the dilated transformed convolution layer, we concatenate the output from every stair together and shrink the channel size to the original input channel size by continuous d convolution layers. The example shown in Figure 6 exhibits one structure with $c = 1, d = 3$, and $k = 3$ stairs.

Such dense block fully consists of convolutional layers and can be considered to contain 3 different effects: feature extraction, mapping, and shrinking. The first c convolution layers represent the feature extraction. The k -stair-like structure represents the mapping, which can discover the mapper between coarse resolution (LR) and fine resolution (HR). The final d convolution layers represent the shrinking, which can compress and filter the redundant features after multiple mappings. We design such structure by dense connection, where the feature representations at lower levels are shared and connected directly with higher levels and thus can increase the nonlinearity of the network to learn complex mappings at the finer levels.

In this dense block, the key point is the dilated transformed convolution layer, which is used to expand the size of feature maps efficiently in each stair. We improve the original dilated transformed convolution layer according to $(n_{(m)}^h, n_{(m)}^w)$. For example, when expanding data size $i_{(m)} \times j_{(m)}$ to $(i_{(m)} + n_{(m)}^h) \times (j_{(m)} + n_{(m)}^w)$, the contributions of $(n_{(m)}^h, n_{(m)}^w)$ are twofold: firstly, $(n_{(m)}^h, n_{(m)}^w)$ are used to define the padding size around original feature maps, which means padding zeros around the feature maps to make the size expand as required. The left-up-right-bottom padding sizes are $\lceil \lceil n_{(m)}^h/2 \rceil, \lceil n_{(m)}^w/2 \rceil, \lfloor \lfloor n_{(m)}^h/2 \rfloor, \lfloor \lfloor n_{(m)}^w/2 \rfloor \rfloor$, respectively, shown as dotted rectangles around the light blue rectangles in Figure 7; secondly, $(n_{(m)}^h, n_{(m)}^w)$ control the dilated rate (dr) for the kernel of dilated transformed convolution layer, where the relationship between dr and $(n_{(m)}^h, n_{(m)}^w)$ is described in the following equation:

$$\begin{cases} \text{dr}^h = \left\lceil \frac{n_{(m)}^h}{2} \right\rceil, \\ \text{dr}^w = \left\lceil \frac{n_{(m)}^w}{2} \right\rceil. \end{cases} \quad (5)$$

The difference between dilated and common convolution lies in that the dilated rate dr means inserting $(\text{dr} - 1)$ zeros between two elements in convolution kernel, as grey rectangles inserted into the dark blue kernel in Figure 7(b). Because the convolution kernel is a skipped structure, the dilated kernels are concerned not only with continuous spatial information (when $\text{dr} = 2$) but also with remote grid points with discontinuous remote spatial correlations (when $\text{dr} > 2$). This is also related to the receptive field in deep

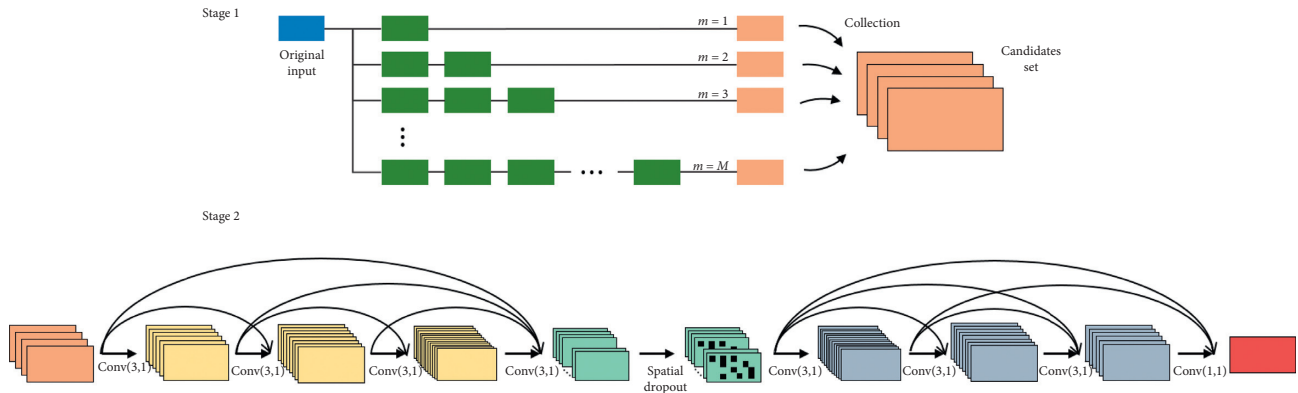


FIGURE 4: Our proposed two-stage deep learning model. The upper row represents the stepwise reconstruction model, the input is the original data, and the output is the HR candidates' collection set. The lower row represents the ensemble inference model, which is a symmetrical encoder-decoder structure, the input is the HR candidates set, and the output is the final downscaling result.

Strategy	Path number	Step size	Data size change
Nature-strategy	20	1	$(30 \times 30) \rightarrow (50 \times 50)$
		2	$(30 \times 30) \rightarrow (40 \times 40) \rightarrow (50 \times 50)$
		3	$(30 \times 30) \rightarrow (36 \times 36) \rightarrow (43 \times 43) \rightarrow (50 \times 50)$
		⋮	⋮
		20	$(30 \times 30) \rightarrow (31 \times 31) \rightarrow (32 \times 32) \rightarrow \dots \rightarrow (50 \times 50)$
Odd-strategy	10	1	$(30 \times 30) \rightarrow (50 \times 50)$
		3	$(30 \times 30) \rightarrow (36 \times 36) \rightarrow (43 \times 43) \rightarrow (50 \times 50)$
		5	$(30 \times 30) \rightarrow (34 \times 34) \rightarrow (38 \times 38) \rightarrow (42 \times 42) \rightarrow (46 \times 46) \rightarrow (50 \times 50)$
		⋮	⋮
		19	$(30 \times 30) \rightarrow (31 \times 31) \rightarrow (32 \times 32) \rightarrow \dots \rightarrow (50 \times 50)$
Even-strategy	10	2	$(30 \times 30) \rightarrow (40 \times 40) \rightarrow (50 \times 50)$
		4	$(30 \times 30) \rightarrow (35 \times 35) \rightarrow (40 \times 40) \rightarrow (45 \times 45) \rightarrow (50 \times 50)$
		6	$(30 \times 30) \rightarrow (33 \times 33) \rightarrow (36 \times 36) \rightarrow (39 \times 39) \rightarrow (42 \times 42) \rightarrow (46 \times 46) \rightarrow (50 \times 50)$
		⋮	⋮
		20	$(30 \times 30) \rightarrow (31 \times 31) \rightarrow (32 \times 32) \rightarrow \dots \rightarrow (50 \times 50)$
Prime-strategy	7	3	$(30 \times 30) \rightarrow (36 \times 36) \rightarrow (43 \times 43) \rightarrow (50 \times 50)$
		5	$(30 \times 30) \rightarrow (34 \times 34) \rightarrow (38 \times 38) \rightarrow (42 \times 42) \rightarrow (46 \times 46) \rightarrow (50 \times 50)$
		7	$(30 \times 30) \rightarrow (33 \times 33) \rightarrow (35 \times 35) \rightarrow (38 \times 38) \rightarrow (41 \times 41) \rightarrow (44 \times 44) \rightarrow (47 \times 47) \rightarrow (50 \times 50)$
		⋮	⋮
		19	$(30 \times 30) \rightarrow (31 \times 31) \rightarrow (32 \times 32) \rightarrow \dots \rightarrow (50 \times 50)$

FIGURE 5: The detailed subprocesses for our proposed 3 path strategies with corresponding step size and detailed size changes. Suppose we enlarge the climate data from 30×30 to 50×50 . The first block represents the NS, where the path number is 20. The second (third) block represents the OS (ES), where the path number is 10. The last block represents the PS, where the path number is 7.

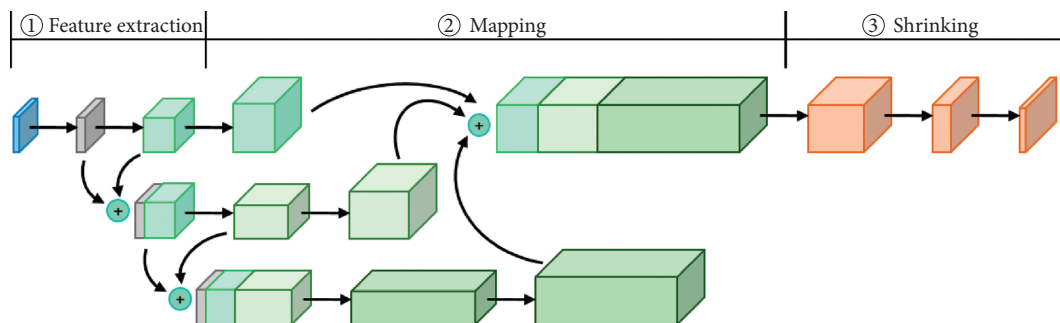


FIGURE 6: Our proposed dense block, which is a stair-like full convolution structure. It can be considered to contain 3 different effects: feature extraction, mapping, and shrinking.

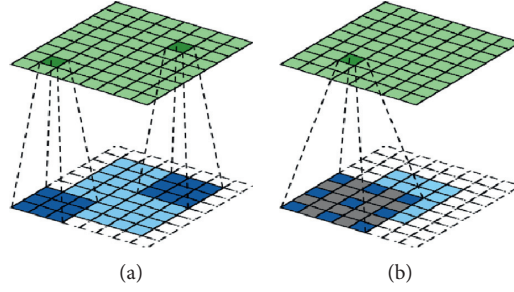


FIGURE 7: Two examples of improved dilated transformed convolution layer, the contributions of which are twofold: enlarging the feature maps and enlarging the receptive fields to capture discontinuous remote spatial correlations. (a) $n_{(m)}^h = n_{(m)}^w = 2, dr^h = dr^w = 1$; (b) $n_{(m)}^h = n_{(m)}^w = 4, dr^h = dr^w = 2$.

learning, which is determined by the real size of convolution kernel. The smaller dr is, the smaller the convolution kernel is, the more the network pays attention to the adjacency information, and vice versa. In our network architecture, we reconstruct the HR data into multiple paths, and we divide the entire enlarge process into various steps, so that the network covers as many multiscale spatial correlations as possible.

3.2. stage 2: Ensemble Inference for Uncertainty Estimation. After obtaining N reconstruction candidates, we will estimate the optimal probabilistic distribution. Mathematically, when solving the optimal downscaling result y^* for a new observed sample x^* , it allows us to infer the probabilistic distribution $p(y^* | x^*)$ via Bayesian formula as

$$p(y^* | x^*, D) = \int p(y^* | x^*, \theta) p(\theta | D) d\theta, \quad (6)$$

where D is the original dataset and θ is the optimal trainable parameters in the downscaling system. However, the posterior distribution $p(\theta | D)$ usually cannot be evaluated analytically. Variational inference is often used to approximate it, which uses a relatively simple distribution $q(\theta)$ to simulate $p(\theta | D)$ by minimizing the Kullback-Leibler (KL) divergence in equation (7). This process is shown in Figure 8.

$$\theta = \arg \min_{\theta} \text{KL}[q(\theta) \| p(\theta | D)]. \quad (7)$$

Dropout variational inference is one of the useful techniques to approximate the posterior distribution. As proved by Gal et al. [31], the posterior distribution $p(\theta | D)$ is approximately equivalent to the dropout mask distribution $q'(\theta)$, which can be viewed as a Bernoulli distribution related to dropout rate z . At the meanwhile, the results can be described by sampling T times in testing, which is a stochastic process. Therefore, equation (6) can be regarded as follows:

$$p(y^* | x^*, D) \approx \int p(y^* | x^*, \theta) q'(\theta) d\theta \approx \frac{1}{T} \sum_{t=1}^T p(y^* | x^*, \theta). \quad (8)$$

Via equation (8), we can estimate the approximate mean and variance for the downscaling results. As for stage 2, we design an encoder-decoder structure to implement it as shown in Figure 4 (lower row). This structure is also fully convolutional. The input of this network is the concatenation of N candidates obtained from stage 1. During the feature maps propagation forward in this network, the width and height are never changed; we only use the convolution filters to expand and reduce the channel size. The encoder and decoder are designed to be symmetrical. In the encoder, the channel size is gradually increased, such that we use [16, 32, 64] filters in Figure 4 (lower row). On the contrary, the channel size is gradually decreased in the decoder, such as [64, 32, 16]. In encoder and decoder, all layers are adopted dense connection, where the input of every layer is the outputs' concatenation of the previous layers. Such structure reuses the extracted features and strengthens the feature propagation.

Furthermore, a trainable spatial dropout layer [32] is applied to connect encoder and decoder, which is to drop out some redundant information points spatially. This layer provides uncertainties for the whole network and allows the approximation of $p(y^* | x^*, D)$ from the stochastic dropout sampling. On the other hand, we can take advantage of all our training dataset, because the vast majority of samples are neural events which contain little extreme information for the regression; determinant single-value traditional methods tend to be overaveraged downscaling naturally (neutralization), but the probability distribution is convenient to describe the true situation.

3.3. Loss Function and Metrics. Our goal is to approximate the stepwise magnification operator g and the downscaling ensemble system \mathcal{F} for generating an HR data \hat{y} , which is close to the ground truth y . According to our formalization, we design two different loss functions for these two different stages, respectively, and combine them together, which can ensure the spatial consistency of reconstructed HR data.

For stage 1, we use N paths to reconstruct N different HR candidates by stepwise subprocesses. Therefore, we

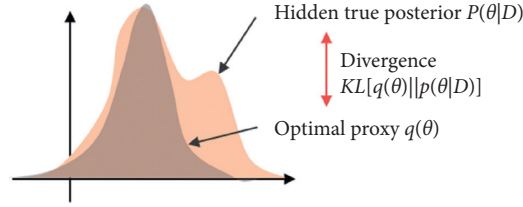


FIGURE 8: The variational inference for analyzing the optimal posterior distribution $p(\theta|D)$ by estimating a relatively simpler distribution $q(\theta)$.

optimize every candidate y_n in every path by ℓ_2 norm, which is described in equation (9):

$$\mathcal{L}_{S1} = \frac{1}{N} \sum_{n=1}^N y - y_{n2}. \quad (9)$$

For stage 2, we combine ℓ_2 norm and ℓ_1 norm together, where ℓ_2 norm is to measure the smoothness of the reconstructed HR data and ℓ_1 norm is to measure the peak distribution of the reconstructed HR data, described in equation

$$\mathcal{L}_{S2} = y - \hat{y}_2 + y - \hat{y}_1. \quad (10)$$

The total loss is as follows:

$$\mathcal{L} = \mathcal{L}_{S1} + \mathcal{L}_{S2}. \quad (11)$$

After training the model, we evaluate the reconstructed HR data in a validation set with two commonly used quality metrics: PSNR and SSIM [33], shown as equations

$$\left\{ \begin{array}{l} \text{MSE} = \frac{1}{i'j'} \sum \sum (y - \hat{y})^2, \\ \text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right), \end{array} \right. \quad (12)$$

$$\left\{ \begin{array}{l} l = \frac{2\mu_{\hat{y}}^2 \mu_y^2 + c_1}{\mu_{\hat{y}}^2 + \mu_y^2 + c_1}, \\ c = \frac{2\sigma_{\hat{y}y} + c_2}{\sigma_{\hat{y}}^2 + \sigma_y^2 + c_2}, \\ s = \frac{\sigma_{\hat{y}y} + c_3}{\sigma_{\hat{y}} \sigma_y + c_3}, \\ \text{SSIM} = l^\alpha \cdot c^\beta \cdot s^\gamma. \end{array} \right. \quad (13)$$

In PSNR, MAX is the maximum value among grid points or pixels, such as MAX = 255 in image evaluation (the pixel range is [0, 255] in an image). In SSIM, l , c , and s represent bright, contrast, and structure comparison, respectively. $\mu_{\hat{y}}$ and μ_y are the averages of all grid point in \hat{y} and y , $\sigma_{\hat{y}}$ and σ_y are the standard deviations of all grid point in \hat{y} and y , and $\sigma_{\hat{y}y}$ is the covariance between \hat{y} and y . We set $\alpha = \beta = \gamma = 1$, and c_1 , c_2 , and c_3 are all small constants which are to prevent denominator from 0.

4. Experiment Schema and Results

4.1. Climate Science Dataset and Benchmark Image Dataset.

For fully measuring the performance of our proposed model, we collect two types of datasets and carry out extensive experiments on them, respectively: ① Climate science dataset (CSD): this dataset is composed of real-world observed daily precipitation data with different 3 resolutions covering the globe [34, 35], which are 0.5°, 1°, and 2.5°, respectively (0.5° only for Europe). ② Benchmark image dataset (BID): this dataset is composed of several different datasets. The training and validation set is from DIV2K [36], containing 800 images for training and 100 images for validation, respectively. The testing set is from 5 datasets: SET5 [37], SET14 [38], BSDS100 [39], URBAN100 [40], and MANGA109 [41]. Among these datasets, SET5, SET14, and BSDS100 consist of natural scenes; URBAN100 contains challenging urban scenes images with details in different frequency bands; and MANGA109 is a dataset of Japanese manga.

For CSD, we crop the daily precipitation data covering Europe and Asia from original data, respectively; the resolution and size are shown in Table 1. Europe covers (150°W to 65°E, 30°N to 80°N), and Asia covers (70°E to 140°E, 0° to 60°N). We make pairs of these data at the same area under different resolutions and treat such pairs as an individual subdataset for training and testing so that we obtain 4 different subsets: 2.5° → 1° Europe, 2.5° → 0.5° Europe, 1° → 0.5° Europe, and 2.5° → 1° Asia. In every subset, we select 85% randomly as the training set, 5% as the validation set, and 10% as the testing set. This dataset is with one channel and preprocessed to [0, 1] by max-min scale as

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (14)$$

where x is the original data, x_{norm} is the scaled data, and x_{\max} and x_{\min} are the maximum and minimum value in original data x . In addition, in the following experiments, we choose lower resolution data as input and higher resolution data in the same area as output of the model. The PSNR and SSIM indices are calculated for the reconstructed HR data directly, where MAX = 1 for PSNR because the data is scaled to [0, 1].

For BID, we design experiments with 2× and 3× magnifications. Due to the characteristics of climate down-scaling, meteorological researchers tend to choose climate data with a resolution close to the requirement for down-scaling, so the magnifications of climate data are often not

TABLE 1: Climate science dataset (CSD).

Region	HOAPS/GPCC	
	Resolution	Size
Europe	0.5°	100 × 190
	1°	50 × 95
	2.5°	20 × 38
Asia	1°	60 × 70
	2.5°	24 × 28

too large, which can basically meet the needs within 3×. The original size of every dataset is not the same; we randomly sample small patches with the size of 64×64 (120×120) as the HR training set and generate the corresponding LR training set with 32×32 (40×40) by bicubic downscaling with 2 (3) scale factor. The input and output are both in RGB color space (original channel=3). The PSNR and SSIM indices are calculated on the luminance channel (a.k.a. Y channel) of YCbCr color space, and $MAX = 255$ for PSNR.

We run our model on a GPU server (Intel i7 8700k and dual NVIDIA GTX 1080Tis). The experiments are divided into two parts. Firstly, we investigate the optimal model structure by determining the hyperparameters. Secondly, we measure and investigate the performance by conducting comparison experiments on the above-mentioned two datasets, respectively.

4.2. Model Analysis. The structure of a deep learning model affects the downscaling performance directly. Our proposed model has many adjustable modules, and we design a lot of comparative experiments to find the optimal combination.

4.2.1. Network Structure. To determine the path number N and corresponding step size M in each path, we evaluate our model with three proposed strategies. We train and test our model on BID with 2× magnifications, where we set $c = 1, d = 3$, and $k = 3$ in the dense block of stage 1 and set encoder (decoder) depth of stage 2 at 3. Under such settings, we dynamically adjust the strategies to evaluate the model performance. The results are summarized in Table 2. From the results, the PS has obvious better performance than others. Specifically, when building the whole model with NS, the results will obtain the largest downscaling errors. It may be due to the fact that the enlargement process is too finely divided, even containing overmuch continuous (+1, +1) in many paths. This will lead to an overly dense network structure, so that too many redundant dense blocks may exist in this process, which leads to overfitting. In addition, the OS or ES also exhibits a little bit overfitting. This demonstrates that the subprocesses should not be divided too much, and it will also lead to an increase in the number of trainable parameters and extend the training time.

On the other hand, the division of subprocesses can be designed adaptively, such as Tolerance2-Strategy (T2S) or Tolerance3-Strategy (T3S), where the difference of step size between adjacent paths is 2 or 3. The results are also described in Table 2. The T2S has downscaling results similar to PS, which outperforms than T3S. T3S divides sparser paths

in the entire model, which may lead to some big skips in the dilated rate and discontinuous spatial correlations. This implies that the model is unable to learn enough detailed features with sparse division of stepwise reconstruction subprocesses.

4.2.2. Network Depth. In our proposed dense block, by reducing the stair depth k to 1 in our proposed model, the block falls back to a network structure similar to FSRCNN [42]. On the contrary, when increasing the depth k of dense block, the block is similar to the structure of LapSRN [24] with the dense connections attached. Increasing the depth k will definitely bring more trainable parameters of the model, causing the considerable extension of training time and downscaling speed, but also improves the model performance. To find the trade-offs between performance and speed, we train the model with different block structure combinations, which represents the different total network depth. We use PS in stage 1 and set encoder (decoder) depth of Stage 2 at 3 and then dynamically adjust the depth of the dense block. The parameter combinations of dense block and trade-offs between performance and speed are shown in Table 3. From the results, we can conclude that deeper model will perform better than shallow. As the network gradually deepens, the performance improvement is not very obvious. Therefore, we choose $k = c = d = 3$ in each dense block of stage 1 for the following experiments to balance the efficiency and performance.

For stage 2, we also train the proposed model by changing the depth of symmetrical encoder and decoder (with PS and $k = c = d = 3$ in the dense block). The results are shown in Table 4. It is the same as stage 1, when the network deepens, the performance will increase. We choose encoder (decoder) depth at 5 because we do not observe significant performance gains by using more convolutional layers.

4.3. Climate Downscaling Performance on CSD. We apply our model to CSD dataset. The CSD dataset contains 5 subdatasets as different regions and resolutions; we divide the entire dataset into 4 groups of LR-HR training pairs. The downscaling of CSD can be considered as 2× (1 to 0.5), 2.5× (2.5 to 1), and 5× (2.5 to 0.5) according to the groups. The LR and HR climate are treated as the input and output of model, respectively. The results are described in Table 5.

From Table 5, we can detect that the restored HR is very close to the real observation in general, where SSIM and PSNR are both maintained at a high level. As the magnification increases, SSIM and PSNR will decrease slightly. This is because the LR data (20×38) especially in 5× experiments is too sparse for the required size (100×190) and cannot provide enough spatial information for more accurate results. In addition, at the same magnification such as in 2.5× experiments, although the input data of the model are different, such that the input sizes are 20×38 and 24×28 for Europe and Asia, respectively, the downscaling results are almost the same. This implies that our model is almost not limited by the input size under the same magnification. In addition, the sizes of the input and output have significant

TABLE 2: The results on the proposed 3 different strategies with 2 other strategies to determine the path number.

Strategies	Set5	Set14
	PSNR/SSIM	PSNR/SSIM
NS	36.53/0.957	31.76/0.905
OS (ES)	37.74/0.959	32.48/0.914
PS	38.27/0.961	34.10/0.921
T2S	38.12/0.961	33.87/0.920
T3S	37.85/0.960	33.88/0.918

TABLE 3: Trade-off between network depth and performance on the depth of the dense block in stage 1. We dynamically adjust the parameters k , c , and d in the dense block, where we set $c = d$ for simplification.

k	c and d	Set5	Set14
		PSNR/SSIM	PSNR/SSIM
1	2	37.54/0.936	32.85/0.890
	3	37.63/0.939	32.87/0.892
	5	37.69/0.942	32.94/0.894
2	2	37.84/0.946	33.84/0.909
	3	37.89/0.949	33.86/0.911
	5	38.03/0.951	33.91/0.912
3	2	38.26/0.961	34.09/0.919
	3	38.27/0.961	34.10/0.921
	5	38.29/0.962	34.12/0.921
5	2	38.30/0.963	34.13/0.920
	3	38.30/0.964	34.13/0.919
	5	38.31/0.964	34.13/0.922

TABLE 4: Trade-off between network depth and performance on the depth of encoder and decoder in stage 2.

Encoder (Decoder) Depth	Set5	Set14
	PSNR/SSIM	PSNR/SSIM
2	38.18/0.956	34.03/0.910
3	38.21/0.959	34.07/0.911
5	38.27/0.961	34.10/0.921
8	38.28/0.963	34.14/0.921
10	38.31/0.964	34.17/0.922

TABLE 5: Quantitative evaluation of our proposed model on CSD: averaged PSNR/SSIM for 4 groups of subdatasets, containing $2\times$, $2.5\times$, and $5\times$ magnifications.

Region	Magnification	Downscaling	Loss Value	
			SSIM	PSNR
Europe	$2\times$	$1 \rightarrow 0.5$	0.966	37.97
	$2.5\times$	$2.5 \rightarrow 1$	0.943	36.81
	$5\times$	$2.5 \rightarrow 0.5$	0.896	31.74
Asia	$2.5\times$	$2.5 \rightarrow 1$	0.967	37.95

impacts on the training time. The larger the size is, the more the time it takes to train the model. Therefore, under the premise of ensuring model performance, it is important to choose a suitable input and output size.

Furthermore, we choose real-world observation data in the validation set to exhibit the downscaling results as shown in Figures 9–12.

In these figures, in the left three columns: the second column (b) describes the original HR climate data, the first column (a) shows the results of bicubic interpolation, and the third column (c) shows the output of our proposed model. The rightmost column (d) in these figures describes the downscaling variance of our model. By comparing the downscaling results, we can see that the bicubic interpolation tends to underestimate and especially overestimate the downscaling results, such as excessively expanding the region of the large value area (red areas in the figures) compared to ground truth; this phenomenon is particularly common. Technically, climate downscaling is to insert more grid points into original data. Traditional approaches are similar to heat diffusion process, which tends to smoothly transport energy between grid points. This makes the results more average, ignoring the real physical mechanism. However, our model considers the different multiscale spatial correlations at the same time in the modeling process; it pays more attention to extract potential features in the discontinuous remote spaces. Therefore, our model can achieve better performance.

From the perspective of variance ((d) in these figures), the largest variance always appears at the large value area. This indicates that, in areas where the events are relatively strong, the uncertainties are high. As the magnification increases, the uncertainty gradually increases and becomes more and more obvious. It may be that the LR data provides too little information, which leads to increased uncertainties. Overall, it still maintains at a very low level. From these real-world downscaling results, although the climate downscaling has become more and more accurate in recent years, it is still unavoidable that uncertainties will interfere with results. Therefore, quantitative evaluation of the uncertainties is very important in meteorological research.

4.4. Comparison with the State-of-the-Art on BID. We compare our proposed model with 8 state-of-the-art SR methods and models: FSRCNN [42], VDSR [20], EDSR [43], LapSRN [24], RDN [44], RCAN [45], AdaRCAN [46], and RFANet [47]. These models are all based on deep learning. All the models are trained and validated on the same dataset, DIV2K, and tested on 5 datasets: SET5, SET14, BSDS100, URBAN100, and MANGA109. In addition, the learning rate is initialized as 10^{-5} and decays to half after every 50 epochs to total convergence.

We measure the performance of the above-mentioned models with PNSR and SSIM. We train these models with $2\times$ and $3\times$ to detect the differences on performance. The results are shown in Table 6. Note that all results of other methods given in this section are generated by the officially released models or papers.

As shown in Table 6, in general, our proposed model is roughly comparable with or even outperforms other models especially in $2\times$ experiments, which indicates the superiority of the stepwise reconstruction and ensemble inference. At

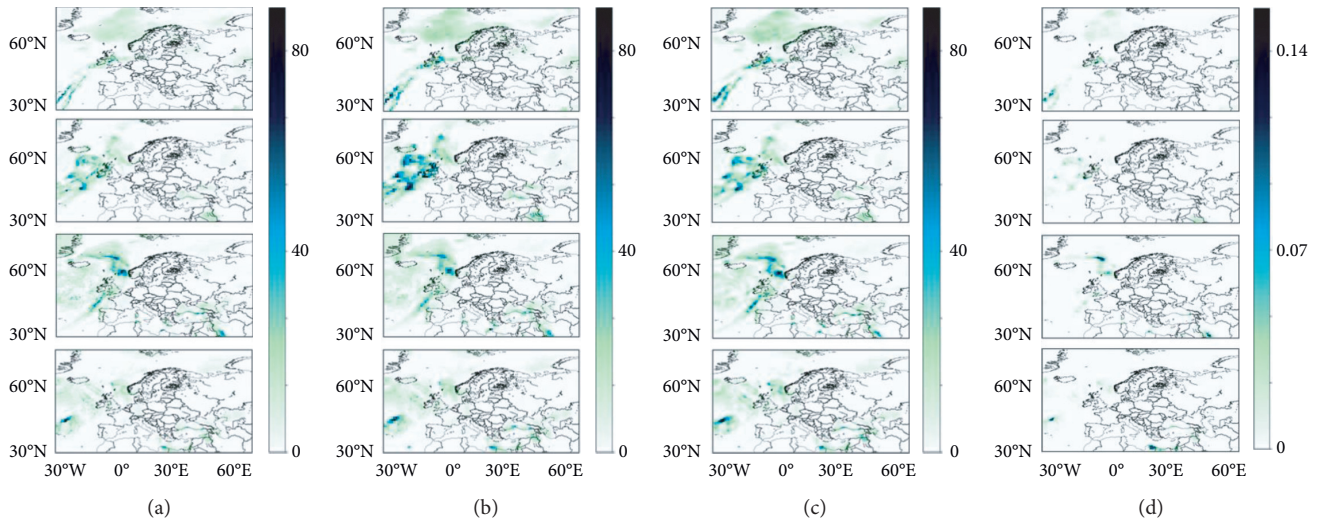


FIGURE 9: The comparison results for our model and bicubic results in $2\times$ experiments ($1 \rightarrow 0.5$) on real-world observation from 2019.12.1 to 2019.12.4 over Europe. The ground truth is in the second column (b). The first column (a) represents the bicubic results, the third column (c) represents our model results, and the forth column (d) represents the downscaling variance of our model.

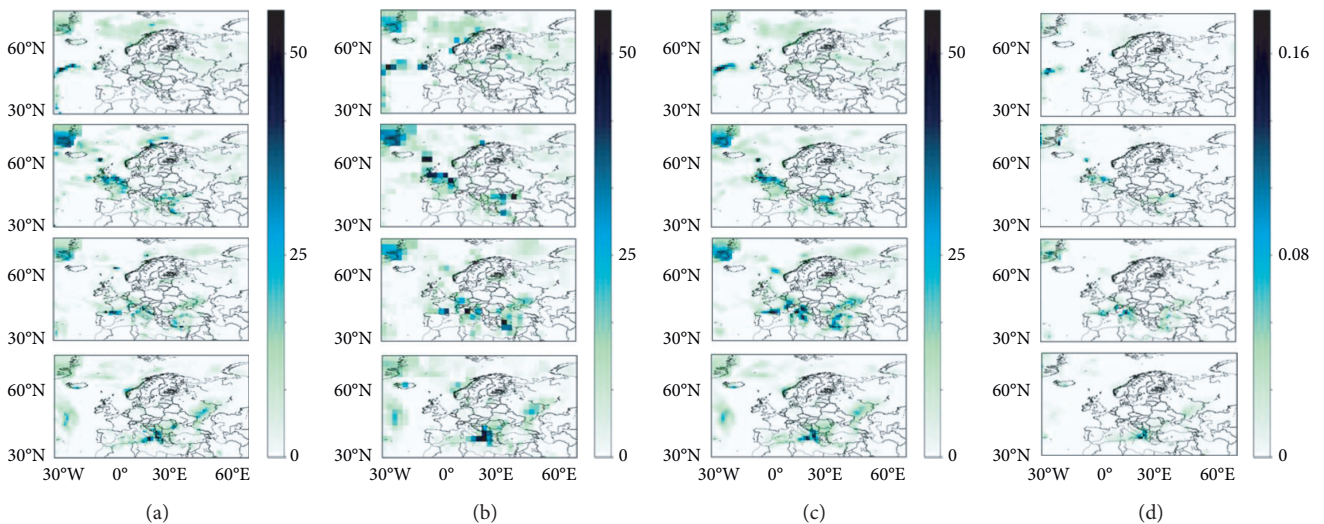


FIGURE 10: The same with Figure 9, but for $2.5\times$ experiments ($2.5 \rightarrow 1$) on real-world observation from 2018.12.1 to 2018.12.4 over Europe. (a) Bicubic results, (b) ground truth HR, (c) our results, and (d) downscaling variance.

present, the existing deep learning models are designed deeper and deeper, and they make full use of dense connection and residual learning in the model structures. However, these models are usually one step to enlarge the feature maps directly (such as from 30×30 to 60×60 directly), which may cause too large feature size changes and incomplete feature extractions. On the other hand, these models only contain the integer-multiple magnification block, so sometimes they cannot adjust any enlargement size such as $2.5\times$ magnification. Due to the stepwise reconstruction process, our model tends to use as many magnification blocks as possible, which enlarge the data gradually and make full use of spatial correlations. Therefore, it can extract more spatial dependencies and achieve better results.

In addition, our model performs better on $2\times$ experiments than $3\times$ experiments in general. This may be because the LR images have closer resolutions with HR images in $2\times$ experiments.

It is worth noting that the SSIM indicator obtained by our model is better, which usually ranks higher than PSNR. According to the mathematical definitions of PSNR and SSIM, PSNR tends to describe the ratio of the noise information (MSE) between two images, while SSIM thoroughly measures the structures of the two images under multiple aspects, such as luminance and contrast, many values of which are calculated by correlation coefficients of these two images based on equation (13). It shows that SSIM pays more attention to the trend and spatial consistency of

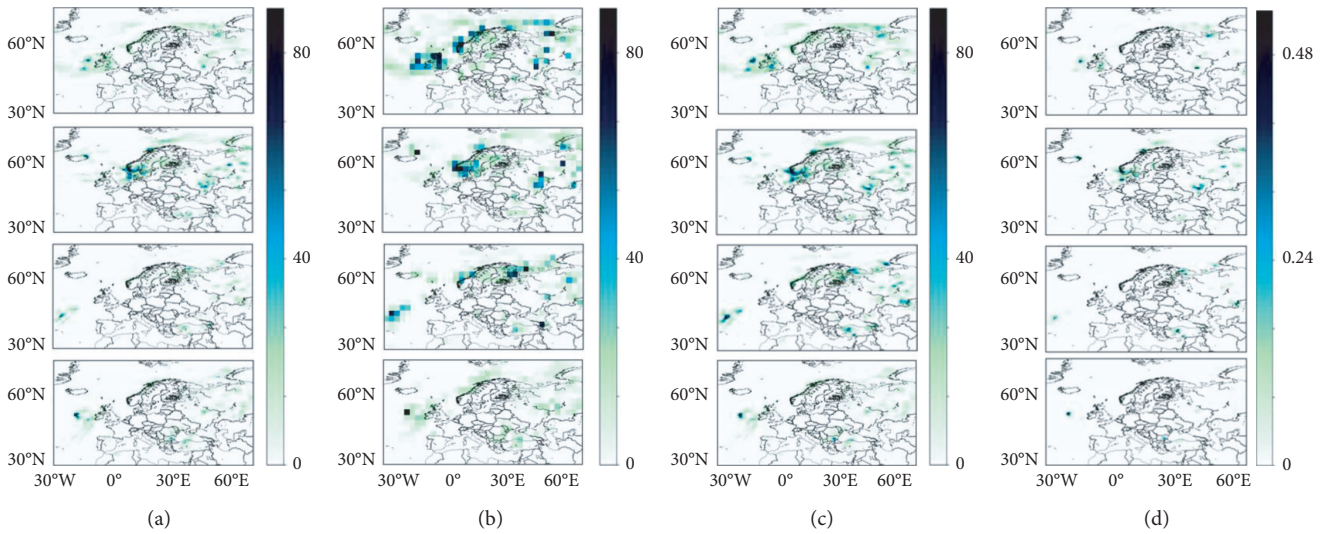


FIGURE 11: The same with Figure 9, but for $5\times$ experiments ($2.5 \rightarrow 0.5$) on real-world observation from 2019.6.1 to 2019.6.4 over Europe. (a) Bicubic results, (b) ground truth HR, (c) our results, and (d) downscaling variance.

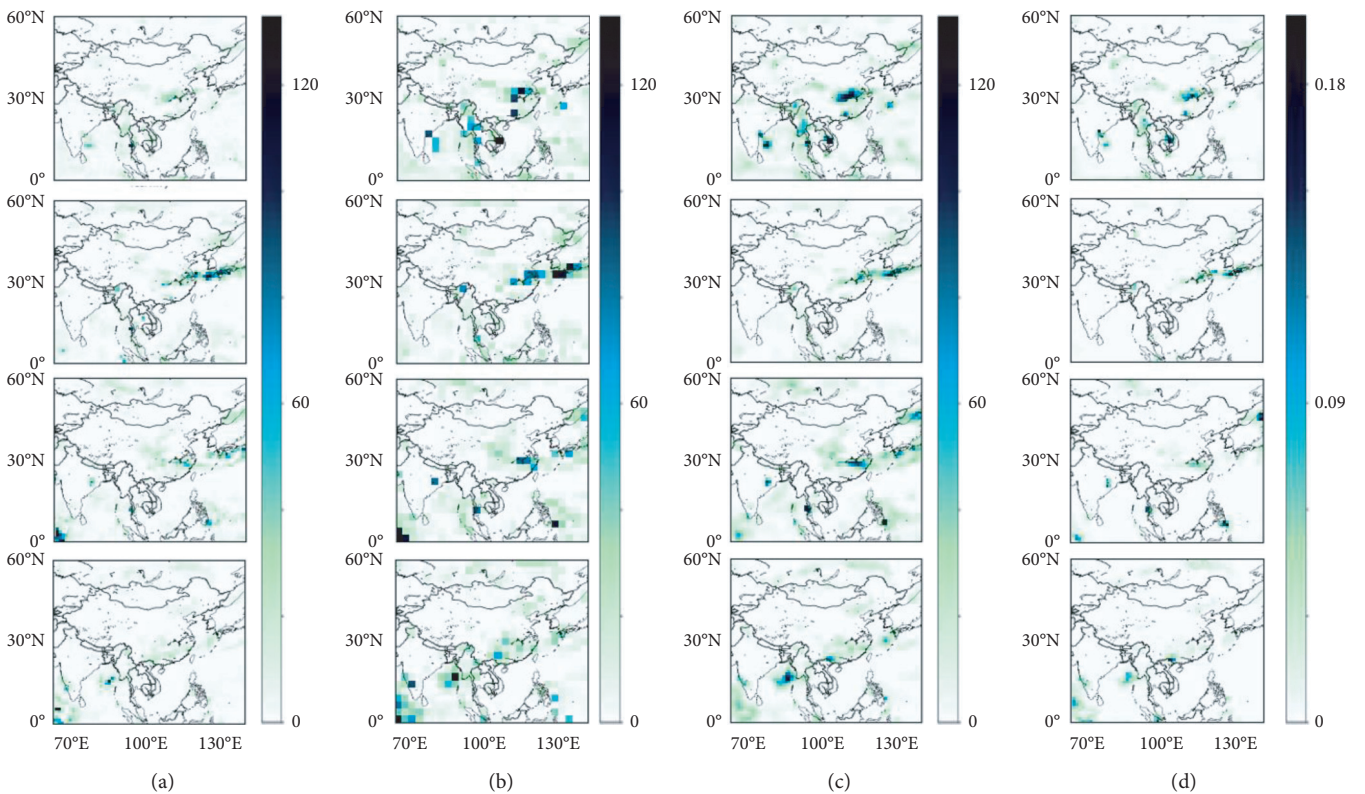


FIGURE 12: The same with Figure 9, but for $2.5\times$ experiments ($2.5 \rightarrow 1$) on real-world observation from 2019.6.1 to 2019.6.4 over Asia. (a) Bicubic results, (b) ground truth HR, (c) our results, and (d) downscaling variance.

the downscaling results. Therefore, this implies that the downscaling results of our model will slightly overestimate or underestimate the real situation due to the interference of uncertainties, but it can exhibit more comprehensive spatial characteristics, which is more important in meteorological research.

5. Discussions

In which aspect does our proposed model consider the multiple spatial remote correlations of climate events? In order to take good advantage of this intrinsic characteristic of climate phenomena, we associate the concept of receptive

TABLE 6: Quantitative evaluation with other 8 state-of-the-art models: averaged PSNR/SSIM for 2× and 3× magnifications. Red text indicates the best and blue text indicates the second best performance.

Scale	Algorithm	SET5	SET14	BSDS100	URBAN100	MANGA109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
×2	Bicubic	33.65/0.930	30.34/0.870	29.56/0.844	26.88/0.841	30.84/0.935
	FSRCNN	36.99/0.955	32.73/0.909	31.51/0.891	29.87/0.901	36.62/0.971
	VDSR	37.53/0.958	32.97/0.913	31.90/0.896	30.77/0.914	37.16/0.974
	EDSR	38.20/0.961	34.02/0.920	32.37/0.902	33.10/0.936	39.10/0.977
	LapSRN	37.52/0.959	33.08/0.913	31.80/0.895	30.41/0.910	37.27/0.974
	RDN	38.24/0.961	34.01/0.921	32.34/0.902	32.89/0.935	39.18/0.978
	RCAN	38.27/0.961	34.12/0.922	32.41/0.903	33.34/0.938	39.44/0.979
	AdaRCAN	38.28/0.962	34.12/0.922	32.41/0.903	33.29/0.938	39.44/0.979
	RFANet	38.26/0.961	34.16/0.922	32.41/0.903	33.33/0.939	39.44/0.978
	Ours	38.27/0.961	34.10/0.921	32.41/0.902	33.33/0.939	39.43/0.979
×3	Bicubic	30.39/0.868	27.55/0.774	27.21/0.739	24.46/0.735	26.95/0.856
	FSRCNN	33.18/0.914	29.37/0.824	28.53/0.791	26.43/0.808	31.10/0.921
	VDSR	33.66/0.921	29.77/0.831	28.82/0.798	27.14/0.828	32.01/0.9340
	EDSR	34.76/0.929	30.66/0.848	29.32/0.810	29.02/0.869	34.17/0.948
	LapSRN	—	—	—	—	—
	RDN	34.71/0.930	30.57/0.847	29.26/0.809	28.80/0.865	34.13/0.948
	RCAN	34.74/0.930	30.65/0.848	29.32/0.811	29.09/0.870	34.44/0.950
	AdaRCAN	34.79/0.930	30.65/0.848	29.33/0.811	29.03/0.869	34.49/0.950
	RFANet	34.79/0.930	30.67/0.849	29.34/0.816	29.15/0.872	34.59/0.951
	Ours	34.75/0.928	30.64/0.848	29.32/0.814	29.09/0.871	34.59/0.952

Dilated rate	Real kernel size
3	7
5	11
7	15
9	19

FIGURE 13: Some examples of the relationships between the real kernel size and dilated rate when the original kernel size is 3.

field in deep learning with spatial remote correlation and design a stepwise reconstruction process to enlarge climate data gradually. More specifically, in the proposed dense blocks, we adaptively design different dilated rate in dilated transformed convolution layers to capture multiscale spatial correlations. In detail, the real kernel size in dilated transformed convolution layers $\text{kernel}_{\text{real}} = \text{kernel} + (\text{kernel} - 1)(\text{dr} - 1)$, which means inserting $(\text{dr} - 1)$ zeros between every two elements of the original kernel. So when $\text{dr} \geq 2$, the kernel size is enlarged, and the receptive fields are also expanded. The space filtered by convolution kernels is not continuous, and there are lots of skips in space, so that broader remote spatial correlations are considered. Figure 13 shows some examples when $\text{kernel} = 3$.

In addition, in the stepwise subprocesses, we divide the entire enlargement process into many small processes that gradually enlarge the feature maps. The stepwise reconstruction process can cover as many different dilated rate dr as possible. Therefore, in the entire enlarging process with a specific strategy (such as Prime-strategy), the model usually contains a large dilated rate and small dilated rate at the same time, which can make every grid point participate in the reconstruction process with multiscale spatial correlations.

Such operation ensures that our proposed model can extract more comprehensive multiscale spatial information and achieve better results.

In the implementation detail, there are no pooling layers existing in the proposed model. Although pooling layers are the most common structure to enlarge the receptive field, it may reduce the size of the data, which is contrary to the downscaling purpose and loss information. Therefore, we design special dilated transformed convolution layers to dynamically adjust the receptive field while increasing the climate data resolution.

The chaos of the dynamic system is always a severe obstacle for climate forecasting. This is the direction for further climate research to try to capture the uncertainties. We design a key structure of a trainable spatial dropout layer, which can increase the randomness of the model, so that our model can infer the probability distribution of downscaling results and describe the optimal mean and variance at the same time.

6. Conclusions

In this paper, we propose a novel formalization for climate downscaling considering the multiscale spatial correlations and chaos in climate events and incorporate these pieces of a priori meteorological knowledge into the deep learning modeling process. Our formalization contains 2 stages. In the first stage, we creatively formulate climate downscaling into a parallel N -path process, and each path contains a stepwise reconstruction process to gradually enlarge the data considering the different scale spatial correlations. The same LR climate data will be reconstructed to N same-sized HR data. stage 1 will generate N HR candidates. In the second stage, we absorb the thought of ensemble prediction and

evaluate the reconstruction uncertainties of these HR candidates by inferencing the optimal probability distribution of the results, which can provide the optimal mean and corresponding variance at the same time.

During the implementation, we take advantage of deep learning technologies on SISR problem and build up a more detailed model named climate downscaling network (CDN). According to our proposed stepwise reconstruction and ensemble inference, we build up a model with two corresponding stages. stage 1 consists of an N -path process, and each path divides the entire process into different steps to gradually enlarge the data. These paths will dynamically design the enlargement step size. Then every path will generate an HR candidate, respectively. In stage 1, the most basic module is the dense block proposed by us. We design such blocks by making an analogy between the concept of receptive field in deep learning and multiscale spatial correlations of climate events. This block is a stair-like structure with dense connections. stage 2 is composed of a fully convolution network, which is used to infer the optimal probability distributions. A classic symmetrical encoder-decoder structure is used in stage 2 with dense connections, where the uncertainty of the network is controlled by a trainable spatial dropout layer between encoder and decoder. According to the proposed model, we design loss functions for these two stages, respectively, and combine them together as the final loss. SSIM and PSNR are set as the evaluation metrics.

We collect and make two different datasets to train and evaluate our proposed model: climate science dataset (CSD) and benchmark image dataset (BID). CSD contains the real-world observed daily precipitation data with different 3 resolutions covering the globe, which are 0.5° (Europe only), 1° , and 2.5° , respectively. BID is composed of 6 different image datasets, where DIV2K is for training and validation, and SET5, SET14, BSDS100, URBAN100, and MANGA109 are for testing.

After building up the deep learning model and collecting the dataset, we design two series of comparison experiments to determine the optimal network structure and depth. Firstly, we confirm and evaluate our proposed 3 different path strategies with 2 other adaptive strategies and find that the paths and step sizes should not be set too dense, which leads to heavy overfitting. The Prime-strategy (PS) is a better structure for the model. Then, to find the trade-off between the speed and performance, we gradually change the depth of the dense block in stage 1 and the depth of encoder-decoder in stage 2. In terms of the SSIM and PSNR, we determine the optimal model structure.

Then, we measure and investigate the performance of our model on the above-mentioned two datasets, respectively. In CSD, we use a real-world observed daily precipitation dataset covering Europe and Asia to train our model. We use different LR-HR pairs' dataset to conduct climate downscaling. Our proposed model effectively learns the mapper projection from LR to HR. The results demonstrate the efficiency and superiority of our model in downscaling daily precipitation data from 2.5 degrees to 0.5 degrees over Asia and Europe. Although as the magnification increases,

the model performance will decrease slightly, it still maintains at a relatively high level. In addition, the sizes of input and output will not affect the performance of the model. We further carry out downscaling simulations on several real-world observation data; the results show that the large uncertainties are always located in the areas where the events are relatively strong, which indicates that the chaos of the dynamic system is always a severe obstacle for climate forecasting. Despite this, our model progressively down-scales the original climate data with low variances.

In BID, we compare our model with 8 other state-of-the-art models in SSIM and PSNR. Sufficient experiments demonstrate that it is reasonable to use our proposed formalization which provides a reliable forecast up to at least 2 and 3 magnifications. Furthermore, our model is not limited in the magnification size compared to other models, tapping its great potentials in exploring detailed dynamic mechanisms for various climate events. What the model learns is not a fixed resolution mapping from LR to HR but the enlarged size, which can be applied to any resolution enlargement and convenient for transfer learning.

Accurate climate forecasting not only depends on models and experience but also requires an overall grasp of the laws of nature. Thus, we will use this approach to investigate the climate predictability. In the future, we will explore dynamical mechanism by downscaling more different physical variables (such as sea surface level and thermocline depth). Furthermore, through absorbing more effective meteorological knowledge, we will build a more targeted model structure to improve the model performance and reduce the computational complexity.

Data Availability

The original daily precipitation dataset can be accessed online (https://opendata.dwd.de/climate_environment/GPCC/html/HOGP_V002.html). The image datasets are also available online (<https://data.vision.ee.ethz.ch/cvl/DIV2K/>; <https://www.kaggle.com/brunoqin/super-resolution>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported in part by the Fundamental Research Funds for the Central Universities under Grant 22120190207, the Key Project Fund of Shanghai 2020 "Science and Technology Innovation Action Plan" for social development under Grant 20dz1200702, and National Natural Science Foundation of China under Grant 42075141.

References

- [1] R. J. Nicholls and A. Cazenave, "Sea-level rise and its impact on coastal zones," *Science*, vol. 328, no. 5985, pp. 1517–1520, 2010.

- [2] S. Villén-Peréz, J. Heikkinen, M. Salemaa, and R. Mkip, "Global warming will affect the maximum potential abundance of boreal plant species," *Ecography*, vol. 43, no. 6, pp. 801–811, 2020.
- [3] L. Wang, J. Chen, and L. Wu, "Impact of winter sst anomaly in tropical eastern pacific on following summer tropical cyclone precipitation in China," *International Journal of Climatology*, vol. 40, no. 2, 2019.
- [4] K. E. Trenberth, "Framing the way to relate climate extremes to climate change," *Climatic Change*, vol. 115, no. 2, 2012.
- [5] Q. Schiermeier, "The real holes in climate science," *Nature*, vol. 463, no. 7279, pp. 284–287, 2010.
- [6] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of cmip5 and the experiment design," *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 485–498, 2012.
- [7] R.-H. Zhang, Y. Yu, Z. Song et al., "A review of progress in coupled ocean-atmosphere model developments for ENSO studies in China," *Journal of Oceanology and Limnology*, vol. 38, no. 4, pp. 930–961, 2020.
- [8] G. Bürger, T. Q. Murdock, A. T. Werner, S. R. Sobie, and A. J. Cannon, "Downscaling extremes-an intercomparison of multiple statistical methods for present climate," *Journal of Climate*, vol. 25, no. 12, pp. 4366–4388, 2012.
- [9] E. P. Maurer and H. G. Hidalgo, "Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods," *Hydrology and Earth System Sciences*, vol. 12, no. 2, pp. 551–563, 2008.
- [10] 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, no. 1-2, pp. 557–570, 2017.
- [11] M. Hessami, P. Gachon, T. B. M. J. Ouarda, and A. St-Hilaire, "Automated regression-based statistical downscaling tool," *Environmental Modelling & Software*, vol. 23, no. 6, pp. 813–834, 2008.
- [12] A. J. Cannon, "Quantile regression neural networks: implementation in r and application to precipitation downscaling," *Computers & Geosciences*, vol. 37, no. 9, pp. 1277–1284, 2011.
- [13] X. Zhang and X. Yan, "A new statistical precipitation downscaling method with bayesian model averaging: a case study in China," *Climate Dynamics*, vol. 45, no. 9-10, pp. 2541–2555, 2015.
- [14] G. Bürger, "Expanded downscaling for generating local weather scenarios," *Climate Research*, vol. 7, no. 2, pp. 111–128, 1996.
- [15] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "Generating high resolution climate change projections through single image super-resolution," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, July 2017.
- [16] R.-H. Zhang and C. Gao, "The IOCAS intermediate coupled model (IOCAS ICM) and its real-time predictions of the 2015-2016 El Niño event," *Science Bulletin*, vol. 61, no. 13, pp. 1061–1070, 2016.
- [17] C. Dong, C. C. Loy, K. He, and X. Tang, *Learning a Deep Convolutional Network for Image Super-resolution*, Springer, Cham, Switzerland, 2014.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2015.
- [19] G. Huang, Z. Liu, V. D. M. Laurens, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2016.
- [20] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 2016.
- [21] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1637–1645, Las Vegas, NV, USA, June 2016.
- [22] W. Shi, J. Caballero, F. Huszár, J. Totz, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.
- [23] C. Ledig, L. Theis, F. Huszar et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
- [24] W. S. Lai, J. B. Huang, N. Ahuja, and M. H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5835–5843, Honolulu, HI, USA, July 2017.
- [25] Pan and S. Jialin, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [26] X. Rong, R. Zhang, and T. Li, "Impacts of Atlantic sea surface temperature anomalies on Indo-East Asian summer monsoon-ENSO relationship," *Chinese Science Bulletin*, vol. 55, no. 22, pp. 2458–2468, 2010.
- [27] A. Rueda, S. Vitousek, P. Camus et al., "A global classification of coastal flood hazard climates associated with large-scale oceanographic forcing," *Entific Reports*, vol. 7, no. 1, p. 5038, 2017.
- [28] K. R. Mylne, R. E. Evans, and R. T. Clark, "Multi-model multi-analysis ensembles in quasi-operational medium-range forecasting," *Quarterly Journal of the Royal Meteorological Society*, vol. 128, no. 579, pp. 361–384, 2002.
- [29] D. J. Stensrud, J.-W. Bao, and T. T. Warner, "Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems," *Monthly Weather Review*, vol. 128, no. 7, pp. 2077–2107, 2000.
- [30] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, "Toward real-world single image super-resolution: a new benchmark and a new model," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3086–3095, Seoul, South Korea, November 2019.
- [31] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, University of Cambridge, Cambridge, UK, 2015.
- [32] J. Tompson, R. Goroshin, A. Jain, Y. Lecun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2014.
- [33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [34] M. Schröder, A. Becker, F. Dietzsch et al., "HOAPS/GPCC European daily precipitation data record with uncertainty

- estimates using satellite and gauge based observations at 0.5° ,” 2019.
- [35] M. Schröder, A. Becker, F. Dietzsch et al., “HOAPS/GPCC Global Daily Precipitation Data Record with Uncertainty Estimates Using Satellite and Gauge Based Observations at 1.0° ,” 2019.
- [36] R. Timofte, E. Agustsson, L. Van Gool et al., “NTIRE 2017 challenge on single image super-resolution: methods and results,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, July 2017.
- [37] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” in *Proceedings of the British Machine Vision Conference 2012*, Oxford, UK, September 2012.
- [38] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” *Curves and Surfaces, Lecture Notes in Computer Science*, Springer, Berlin, Germany, pp. 711–730, 2010.
- [39] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.
- [40] J. B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Boston, MA, USA, June 2015.
- [41] Y. Matsui, K. Ito, Y. Aramaki et al., “Sketch-based manga retrieval using manga109 dataset,” *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21811–21838, 2017.
- [42] C. Dong, C. C. Loy, and X. Tang, *Accelerating the Super-resolution Convolutional Neural Network. European Conference on Computer Vision*, Springer, Cham, Switzerland, 2016.
- [43] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced Deep Residual Networks for Single Image Super-resolution,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, July 2017.
- [44] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.
- [45] 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)*, pp. 286–301, Springer, Berlin, Germany, 2018.
- [46] E. C. Mannshardt-Shamseldin, R. L. Smith, S. R. Sain, L. O. Mearns, and D. Cooley, “Downscaling extremes: a comparison of extreme value distributions in point-source and gridded precipitation data,” *The Annals of Applied Statistics*, vol. 4, no. 1, pp. 484–502, 2010.
- [47] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, “Residual feature aggregation network for image super-resolution,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2359–2368, Seattle, WA, USA, June 2020.