Self-Supervised Vision for Climate Downscaling

Karandeep Singh[1] , **Chaeyoon Jeong**[2,1] , **Naufal Shidqi**[2,1] , **Sungwon Park**[2,1] ,
Arjun Nellikkattil[3,4] , **Elke Zeller**[3,4] and **Meeyoung Cha**[1,2,5*]

[1]Data Science Group, IBS, South Korea
[2]School of Computing, KAIST, South Korea
[3]Center for Climate Physics, IBS, South Korea
[4]Department of Climate System, PNU, South Korea
[5]MPI-SP, Germany
{karandeepbrar, joungchaeyoon, hi.nshidqi, deu30303, arjunbabun11, iccp.elke, meeyoung.cha}@gmail.com

Abstract

Climate change is one of the most critical challenges that our planet is facing today. Rising global temperatures are already affecting Earth's weather and climate patterns with an increased frequency of unpredictable and extreme events. Future projections for climate change research are based on computer models like Earth System Models (ESMs). Climate simulations typically run on a coarser grid due to the high computational resources required, and then undergo a lighter downscaling process to obtain data on a finer grid. This work presents a self-supervised deep learning model that does not require high resolution ground truth data for downscaling. This is realized by leveraging salient distribution patterns and the hidden dependencies between weather variables for an individual data point at runtime. We propose three climate-specific components that well represent the patterns of underlying weather variables and learn intricate inter-variable dependencies. Extensive evaluation with 2x, 3x, and 4x scaling factors demonstrates that our model obtains 8% to 47% performance gain over existing baselines while greatly reducing the overall runtime. The improved performance and no dependence on high resolution ground truth data make our method a valuable tool for future climate research.

1 Introduction

The increase in global temperatures is resulting in unprecedented changes in Earth's weather and climate patterns, characterized by an increase in the frequency of extreme weather events [Seneviratne *et al.*, 2021] as well as other changes like melting ice caps, rising sea levels, lowering crop yields, and impacts on oceans and the fishing industry [H.-O. Pörtner, 2022]. Th climate change impact has critical consequences for Earth's ecology and human societies. Future climate projections are largely driven by a suite of computer-based physics models that simulate the Earth's climate system, called ESMs. These physics models are fundamentally important tools for climate research and the design of future climate policies and strategies. High-resolution (HR) climate data offer detailed predictions at a regional level and can help plan local mitigation strategies. However, finer-scale models are constrained by the enormous computational resources required to run them. Hence, simulations are generally run at coarser scales, and HR data on a finer scale is obtained by downscaling. We limit the scope of this work to statistical downscaling.

Climate downscaling resembles the super-resolution (SR) task in computer vision. Supervised SR models assume that paired or unpaired low-resolution (LR) and HR data are available. Moreover, ground truth data are assumed to be accessible in sufficient quantity for the desired model fit and generalizability. These models further assume that the test data distribution is similar to training. Under these assumptions, several advanced machine learning-driven downscaling models have been proposed that are fully or partially supervised [Groenke *et al.*, 2020; Vandal *et al.*, 2017; Park *et al.*, 2022].

However, access to ground truth data is not guaranteed in real-world climate research. In addition, the inherent stochastic nature of climate simulations makes it infeasible to obtain LR-HR data pairs. Minor changes in different simulation runs can lead to vastly different outcomes, a phenomenon also known as the *butterfly effect* [Palmer *et al.*, 2014]. Time scale brings another challenge, as climate simulations can span thousands or even millions of years, making it computationally infeasible to generate sufficient HR simulation data. For instance, an HR simulation with the Community Earth System Model (CESM) uses 250K core hours per simulated year, generating approximately one terabyte of data per compute day [Small *et al.*, 2014].

This research presents an interdisciplinary collaboration between AI and climate physics, introducing a deep learning based climate downscaling model that does not require HR ground truth data for training. Our model uses self-supervised learning to train an instance-specific model on a single data point and adapts to its data distribution at runtime. We designed three components to reflect underlying weather

*Corresponding author

variables and improve transferability. First is self-supervised pretraining, which enables knowledge transfer across the input data instances. Second is channel segregation, which enhances the learning representations of different weather variables that have vastly different characteristics (as shown in the precipitation and temperature in Figure 1. Third is topoclimatic attention, which forces the model to learn intricate inter-channel dependencies. The latter two components help extract improved features and understanding of physical relationship between climate variables.

The model is evaluated with multiple scaling factors by downscaling the surface temperature (TS) and total precipitation (PRECT) data obtained from CESM (v. 1.2.2) and utilizing the gradient of topography (dPHIS) as an additional helping variable. The evaluation results illustrate that our model achieves performance superiority ranging from 8% to 47% over existing baselines and greatly reduces the overall runtime computational complexity against that of self-supervised model baseline. In addition to quantitative performance gain, the qualitative analysis section demonstrates the enhanced downscaling quality of the model output. The model code and data is available at: https://github.com/k-s-b/climate_sd

Without reliance on ground truth HR data, and reduced computational complexity, our downscaling method has great potential for climate downscaling research. Our work contributes towards the ongoing efforts to enable (1) "temporal downscaling" by running coarser models with higher temporal resolution, (2) larger ensemble runs which are necessary for understanding climate variability, (3) democratizing climate research and allowing organizations with limited access to computational power to run HR simulation models, (4) downscaling already conducted simulations and increasing the resolution without rerunning the model, and (5) promoting a greener climate with a lower energy budget.

As research plays a crucial role in understanding, predicting, and developing mitigation and adaptation strategies for climate change, this work contributes to several aspects of the United Nations Sustainable Development Goals (UN SDGs), with SDG 13 (climate action) in particular. By increasing downscaling quality and reducing computational complexity, our work enables a wider adoption of climate data, empowering informed decisions by policymakers and stakeholders. Our contributions to democratizing climate predictions also align with SDG 9 (industry, innovation, and infrastructure). Further, the interdisciplinary research presented in this work is conducted with an active collaboration of computer science (AI) and climate researchers, which resonates well with SDG 17 (partnerships for the goals).

2 Related Work

Single Image Super-Resolution. This task aims to enhance image resolution, typically using convolutional neural networks (CNN) and generative adversarial networks (GAN) [Dong *et al.*, 2014; Ledig *et al.*, 2017; Kim *et al.*, 2016; Lim *et al.*, 2017; Haris *et al.*, 2018; Zhang *et al.*, 2018c; Ledig *et al.*, 2017; Zhang *et al.*, 2018b; Lu *et al.*, 2022]. These methods involve training CNN models or vision transformers to directly map the LR images to HR images. While

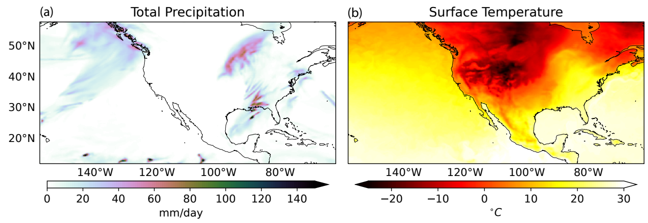

Figure 1: A snapshot representing total precipitation (PRECT) and surface temperature (TS) with different dynamics.

most methods rely on supervised learning with predefined blur kernels, some tackle unpaired or blind SR, addressing the absence of direct LR-HR pairs [Wang *et al.*, 2021; Ji *et al.*, 2020; Cornillère *et al.*, 2019; Zhang *et al.*, 2018a; Assaf Shocher, 2018].

Deep Learning for Climate Downscaling. Deep learning has been widely explored for for climate downscaling by applying SR techniques to various climate features. Previous research utilized deep neural networks such as CNN [Vandal *et al.*, 2017; Ducournau and Fablet, 2016; Kumar *et al.*, 2021; Baño-Medina *et al.*, 2020; Vandal *et al.*, 2019; Stengel *et al.*, 2020; Misra *et al.*, 2018]. Those works demonstrate the potential of deep learning methods over traditional approaches. Recent approaches point out that applying SR methods directly to climate data presents challenges due to the distinct nature of climate data compared to images. A branch of work has addressed this by incorporating climate-specific attributes or physics into deep learning models, yielding promising outcomes [Chou *et al.*, 2021; Rampal *et al.*, 2022].

3 Problem Setting

Problem: Given an LR climate data instance X and no availability of paired HR data, the goal is to obtain a HR version $X\uparrow$ of X such that visual quality of the downscaled data is improved in terms of the root mean squared error (RMSE), while minimizing the inference time.

Climate simulation data is represented in grids, where the height and width of the data represent the number of grid points over a spatial area. The larger the number of grid points for a given area, the higher the resolution of the data and the better the representation of regional climatic conditions. If a LR data instance for a given area has dimensions of $\mathbb{R}^{C\times W\times H}$, we downscale it to $\mathbb{R}^{C\times W'\times H'}$ such that $W' = s\times W$ and $h' = s\times H$, where s is the scaling factor ($s > 1$), and C denotes the number of variables in the data. TS and PRECT are critical variables that determine weather and climate conditions and impact climate adaptation and mitigation policies. These variables are depicted in Figure 1. Thus, we consider $C = 3$ with temperature (TS), total precipitation (PRECT), and topography gradient (dPHIS) as the input weather variables. We only downscale TS and PRECT; dPHIS remains constant throughout the simulation years. Nonetheless, dPHIS carries valuable information to predict the two target variables; for example, temperature decreases with height. Moreover, using dPHIS instead of PHIS reduces opposing biases on the

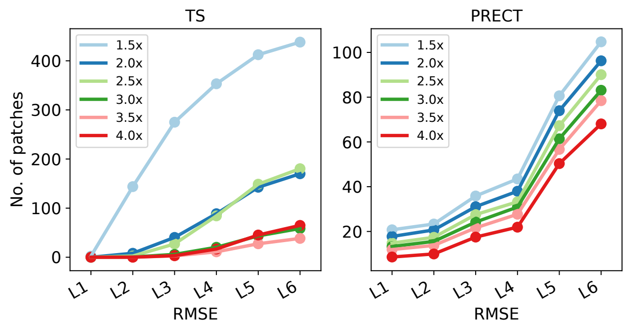

Figure 2: Climate data reveals information repetition across scales. The similarity between the 10×10 patches in the source image is compared across scales (1.5x–4.0x). The y-axis represents levels L1 to L6 or the increasing RMSE values for the cutoff threshold that establishes the degree of similarity. The cutoff range is 1E-2 to 1E-1 for TS and 0.5E-14 to 1E-12 for PRECT. The number of patches with RMSE values below the cutoff at that level is shown on the x-axis.

windward and leeward sides of mountains. Our approach is flexible, agnostic to the type and number of input weather variables, and can downscale multiple variables at once.

3.1 Data

The data utilized in this work is obtained from CESM 1.2.2 [Hurrell *et al.*, 2013; Small *et al.*, 2014], which logs the daily means of present-day climate conditions for 140 simulation years on a spatial resolution of approximately 0.25° in the atmosphere component and 0.10° in the ocean components [Chu *et al.*, 2020]. The last 20 (corresponding to 7,300 data points) years of data are used for this work, which are considered from the periods that have received an equilibrium state.

4 Model

Our model is a self-supervised climate downscaling model that operates without any HR ground truth dataset. Figure 3 depicts the core architecture of the model.

4.1 Information Redundancy in Climate Data

Natural images have been shown to possess recurring patches of information across their original images and their coarser versions [Glasner *et al.*, 2009]. The internal information of an image is known to have stronger predictive power than the information from external databases [Zontak and Irani, 2011]. Inspired by previous SR works [Assaf Shocher, 2018], we analyze information redundancy in the raw climate data. Figure 2 presents the statistics of this analysis from our dataset. It can be seen that climate data have information redundancy for both the TS and PRECT channels across multiple scales of the input data. This forms the basis for a SR model that learns a generalized mapping from the downgraded LR data point to its higher resolution.

4.2 Training Process

Self-supervised training using pseudo data pair. The main research question we address in this work is downscaling LR climate data when HR data is not available. Similar to [Assaf Shocher, 2018], we generate the "pseudo" LR-HR pair by degrading the original HR data (X^{HR}). We regard this degraded data as the LR climate data (X) to be downscaled, and apply degradation to generate its downgraded version $X{\downarrow_d}$ by a factor d ($d < 1$). The data pair $X{\downarrow_d}$ - X now becomes a pseudo-LR ($X{\downarrow_d}$) and pseudo-HR (X) pair and is used as training data of an instance-specific CNN model f_θ. The downscaled data $X{\uparrow_s}$ is obtained by applying the trained model on X, *i.e.*, $X{\uparrow_s} = f_\theta(X)$. Here, s indicates the target downscaling factor and d is the downgrading factor, with $d{=}1/s$ typically.

Before passing as an input to the model, X and $X{\downarrow_d}$ are rescaled to the target downscaled size. The optimization objective of the model becomes minimizing element-wise loss such as mean squared error (MSE) or mean absolute error (MAE) between $f_\theta(X{\downarrow_d})$ and X. The original HR data is only used for quantitative analysis of the final output and is never used for model training.

4.3 Self-supervised Pretraining

Training on just one data point poses two critical challenges: long convergence times and limited knowledge utilization from the rest of the data. Deep learning models trained on larger datasets for one task can be adapted to similar downstream tasks, known as transfer learning [Finn *et al.*, 2017; Sun *et al.*, 2019; Soh *et al.*, 2020]. Inspired by this, we perform a self-supervised pretraining step where a single model learns a mapping for the entire downgraded input LR data ($X{\downarrow_d}$) and the input LR (X) data itself. Exposing the model to the entire input data enables knowledge transfer across the dataset. Additionally, we finetune the pretrained model on randomly sampled data points augmented by uniform random noise, optimizing it for diverse scenarios and enhancing its generalization. This two-step approach combines pretraining with regularized data-specific adaptation, resulting in a faster, more knowledgeable, and ultimately more robust model. The effect of this component on performance and runtime is detailed in Section 5. The final downscaling task is realized by loading this pretrained model and optimizing it for a single data instance on the go (individual $X{\downarrow_d}$ - X pairs).

4.4 Channel Segregation

Figure 1 presents a snapshot of the TS and PRECT channels and highlights their characteristic differences. The TS channel has a smoother gradient and spans the entire spatial extent. In contrast, PRECT is highly non-linear, chaotic, localized and has a high degree of variance within localized regions. This characteristic contrasts with natural images, where different channels, albeit with different values, represent measurements of the same snapshot. The standard convolutional neural network (CNN) models operate under this assumption since a weighting filter convolves over the entire input volume to output a unified representation as another volume.

The model may be forced to simultaneously optimize for low-frequency TS features and high-frequency PRECT features at the same spatial location. This perplexity is unwanted and can lead to the model producing undesirable artifacts in

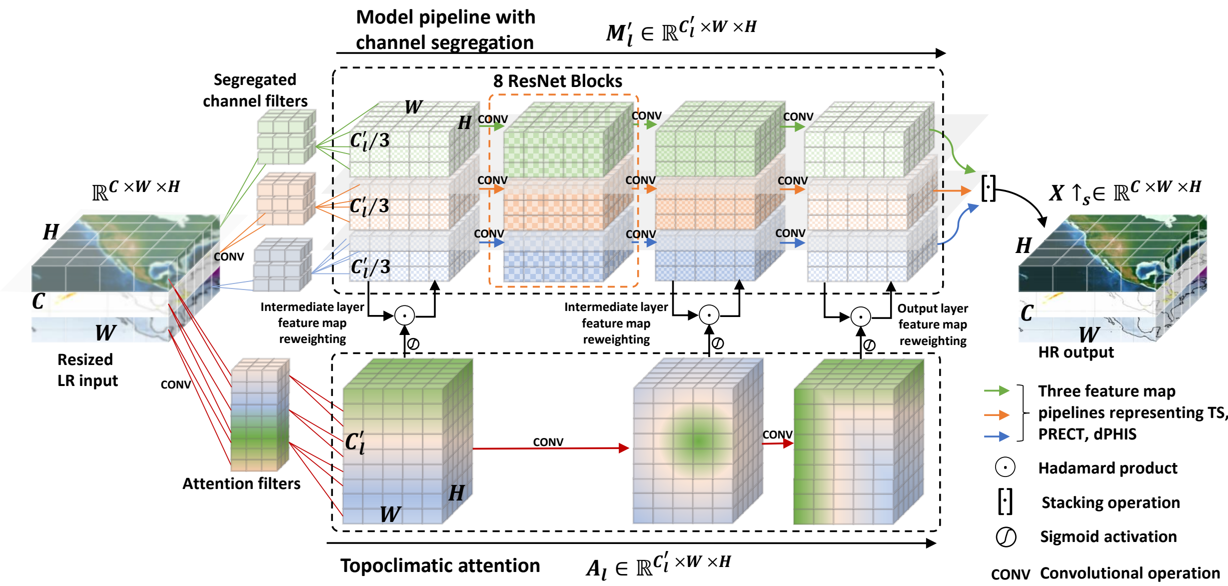

Figure 3: Core architecture of the proposed model. LR input X is downscaled to $X\uparrow_s$ ($s > 1$) and $(X, X)\uparrow_s \in \mathbb{R}^{C\times W\times H}$ where $C = 3$, represent three weather variables: surface temperature (TS), total precipitation (PRECT), and gradient of topography (dPHIS). X is resized to target size before feeding into the model. The upper half of the image shows the channel segregation with three pipelines, with respective filters for each input variable. The topoclimatic attention mechanism is the lower half parallel to the main model flow. The attention mechanism reweights the intermediate feature maps at multiple stages and enforces learning of topoclimatic relationships.

downscaled output. However, CNN should adapt to a particular weather variable's low- and high-level characteristics and leverage them for the SR task. A mechanism is needed to handle each channel's vastly diverse properties effectively. We enable it by constraining the CNN model to learn separate weights for each input weather variable. Our model consists of C different pipelines with several feature maps (evenly divided among the input channels) running in parallel, which are stacked at the last layer and optimized simultaneously. If C' represents the number of channels in the intermediate feature map, then each channel is convolved by a filter of size $\frac{C'}{C}$ where $C' = I \times C$ and I is an integer. This operation can be represented as:

$$\mathrm{Out}(N, C') = \left[\sum_{c} \mathrm{Filter}(C', c) \cdot \mathrm{Input}(N, c) + \mathrm{b}_{out}\right]_{\forall c \,\in\, C}$$

where N is the batch size, C the number of weather variables, C' the number of output channels, and $[\cdot]$ represents the stacking operation. This architecture can learn features for each channel by reducing the undesired artifacts in the model output.

4.5 Topoclimatic Attention

Separating the filters for individual channels can improve model performance but has the drawback of missing out on inter-channel dependencies. This is particularly true for weather data, where intricate relationships between temperature, precipitation, and topography can exist. To account for such relations, we propose a topoclimatic attention mechanism that learns the inter-channel dependencies and enforces them by reweighting low, intermediate, and high-level feature maps in the main model pipeline. We define low to high-level feature maps as the intermediate output of the model layers from the first through the last. The attention mechanism learns weights A_l ($l \in L$, where L are the model layers) that encompass these dependencies with the following steps.

Convolution over volume. If $\mathrm{M}'_l \in \mathbb{R}^{C'_l \times W\times H}$ is the feature map at the layer l, then $\mathrm{A}_l \in \mathbb{R}^{C'_l \times W\times H}$ is the attention weight at layer l, computed by a convolution over the entire input volume. A separate set of weight filters are trained for computing A_l. Spanning the entire volume layer-wise enables learning of dependencies between the weather variables at a given level l of the feature map.

Generation of attention weights. At given level l the attention weights A_l can be represented as:

$$\mathrm{A}_l = \sigma\Big(\phi\Big(\mathrm{Conv}^l\Big(\phi\Big(\ldots\Big(\phi(\mathrm{Conv}^1\Big(\mathrm{X}\Big)\Big)\Big)\ldots\Big)\Big)\Big)\Big)$$

where Conv^l is a 2D covolutional operator at layer l, ϕ and σ are LeakyReLU and Sigmoid activation functions respectively. If A'_l represents the raw output of the convolution operation at layer l (*i.e.* without the Sigmoid activation) then A_l ($l > 1$) is computed by passing A'_{l-1} through Conv^l. At $l = 1$, the A_l is computed over the input. To enable this operation to learn complex non-linear correlations, the convolved output is transformed by the LeakyReLU non-linear activation function. Finally, the entire output is normalized in the range [0, 1] by passing it through the sigmoid activation function.

Scale	2x		3x		4x	
Channel	TS	PRECT	TS	PRECT	TS	PRECT
Bilinear	0.6655	2.2302	1.1922	3.5510	0.8662	3.7215
Bicubic	0.6829	2.0246	1.2491	3.6584	0.8647	3.5651
PT	0.4879	1.5046	0.9343	3.0128	0.7534	3.3845
PT+FT	0.4607	**1.3552**	0.9109	2.8722	0.7435	3.4147
Ours	**0.3585**	1.3795	**0.7959**	**2.8703**	**0.7012**	**3.3837**

Table 1: Performance comparison (RMSE) against standard approaches and deep learning models for 2x, 3x, and 4x scaling factors for temperature (TS) and precipitation (PRECT).

Feature map reweighting. The intermediate feature map M'_l is updated to M_l by attention weight A_l as:

$$M_l = A_l \odot M'_l$$

where $\odot$ is the Hadamard product, and $M'_l, M_l, A_l \in \mathbb{R}^{C'_l \times W \times H}$. The attention mechanism reweights the feature maps to highlight the importance of inter-variable dependencies and the regions of the highest importance.

5 Experiments and Results

The model training and evaluation are performed with the dataset described in Section 3.1. The results are reported over 2x, 3x, and 4x factors, representing downscaling from resolutions of 50km, 75km, and 100km, respectively, to 25km. We use a scale factor of 2 when generating pseudo-LR and HR pairs in the pretraining phase. The model performance is measured by the RMSE between the downscaled model output and the ground truth HR data. Additional analyses are presented on the project's GitHub page.

Evaluation. The model evaluation is performed with the *entire* CESM dataset (7,300 data points of size $213 \times 321 \times 3$), as the ground truth data is not exposed to the model at any point during the training. The results are reported over 2x, 3x, and 4x scaling factors, representing downscaling from resolutions of 50km, 75km, and 100km, respectively, to 25km. For all tables, the RMSE units are in 1e-08 for PRECT. Smaller values are better for RMSE.

	Scale	2x		4x	
	Channel	TS	PRECT	TS	PRECT
Interpolation	Bicubic	0.6804	2.0618	0.8616	3.6487
Supervised	$\text{GINE}_{5\%}$	1.0712	1.7408	1.0859	3.5636
	$\text{GINE}_{10\%}$	0.8112	**1.2454**	0.9388	3.5069
Self-supervised	Base	0.5119	1.3367	0.7844	3.5329
	PT+FT	0.4580	1.3835	0.7399	3.4985
	Ours	**0.3568**	1.4607	**0.6980**	**3.4749**

Table 2: Performance comparison (RMSE) of the statistical method, supervised and self-supervised models based on 1,000 data points.

5.1 Performance Evaluation

Table 1 shows the model performance over 2x, 3x, and 4x downscaling factors for the entire dataset of 7,300 data points.

Model	Training time	Testing time	Total time	1 datapoint time (sec)
Base	-	38.8000	283.2394	139.68
PT+FT	3.1667	0.1693	1.2360	0.61
Ours	15.4167	0.2821	2.0592	1.02

Table 3: Time efficiency of the Base model, the pretrained model without channel segregation and topoclimatic attention, and our model. The experiment is performed over 1,000 data points and estimated for the entire 7,300 data points. All units are in hours, except the runtime for one datapoint, which is in seconds.

Linear and bicubic interpolation are commonly used baseline methods in climate downscaling research. We also compare the model performance against that of strong baselines such as self-supervised pretraining (PT), which is a generalized SRResNet model without any finetuning with all $(X \downarrow_d, X)$ pairs. Additionally, we evaluate the model against a more sophisticated baseline (see 'PT+FT' in the table) where the pretrained model is finetuned with individual data points but does not contain channel segregation and topoclimatic attention components.

For the 2x scaling factor, our model outperforms all baselines, with a 47.50% (TS) and 31.86% (PRECT) reduction in the RMSE over that of bicubic interpolation and a reduction of 26.52% (TS) and 8.32% (PRECT) over PT. The overall runtime is reduced greatly, as discussed in Section 5.2. Except for the 2x scaling factor of PRECT, the full model consistently outperforms the baselines for the 3x and 4x scaling factors.

We finally compare our model performance against GINE [Park *et al.*, 2022], a supervised model for climate downscaling. We train this model in a supervised fashion by using 5% and 10% of the HR ground truth data (X-X^{HR} pairs). As this model uses HR data for training, we compare its performance against 1,000 randomly chosen data points from the held out set. The results in Table 2 illustrate that our model outperforms the baseline, and even GINE model trained with some HR ground truth data, although it performs better for PRECT when using 10% labels. Base refers to a model such as [Assaf Shocher, 2018], with architecture as in Fig. 3, but without the proposed components. Given the limited compute and runtime constraints, we do not include vision transformer based baselines, although they should be explored for the future versions of this work.

The TS and PRECT channels do not show the same patterns of performance improvement because the two climate variables have distinct dynamics (as shown in Figure 1). The improvement in one channel results in a minor performance trade-off in the other channel.

5.2 Runtime

As climate simulations usually cover many time steps, the downscaling method should have a low overall runtime for practical use. We assess the overall computation times for a model trained on a single data instance from scratch (Base), a pretrained and finetuned model without channel segregation and topoclimatic attention components (PT+FT), and a full model with these components (Ours). The base model

Channel	TS	PRECT
Full model	**0.3585**	1.3795
(-) Channel Segregation (Sec. 4.4)	0.4055	**1.3245**
(-) Attention (Sec. 4.5)	0.3656	1.4539
(-) FT	0.4132	1.4106

Table 4: Ablation study: removing channel segregation, topoclimatic attention and noisy finetuning at the 2x scaling factor.

requires thousands of epochs for convergence. Due to the computational bottleneck induced by the base model, this particular assessment is only performed with randomly selected (seeded) 1,000 data points (only the analyses presented in Tables 2, 3, and 5 is performed over (the same set of) randomly chosen 1,000 datapoints, whereas other results are presented for the full dataset). A single NVIDIA Tesla A100 GPU is used for experiments. The results demonstrate that our model substantially reduces overall runtime while reducing the RMSE values.

5.3 Ablation Study

We remove each of the proposed components, i.e., channel segregation, topoclimatic attention, and finetuning on augmented data in the pretraining and finetuning on individual data points (represented by (-) FT), and measure the model performance for all data points. Table 4 confirms that every model component is essential for performance. Removing channel segregation results in some performance gain for PRECT over the full model. At the same time, it also causes a performance drop in the TS channel. This implies that after an initial reduction in the RMSE, the TS and PRECT performance exhibit orthogonality to some extent. Nonetheless, the performance is better than the most commonly used baselines.

The proposed components improve the quality of the downscaled output substantially. The zoomed-in regions in the Figure 4, we demonstrate that the channel segregation component effectively eliminates the undesirable artifacts present in the unsegregated model, resulting in output that closely resembles the ground truth.

5.4 Model Robustness

Non-ideal blur kernel. We test the model's ability to downscale climate data when the blur kernel is non-ideal (*e.g.*, not bicubic). To realize this kernel, we inject *freshly* generated uniform random noise into X and $X\downarrow_d$, *independently* for each data point. A noisy setting implies that the relationship between X and $X\downarrow_d$ can be different from $X\downarrow_d$ and X^{HR}. This step is also included in the pretraining and instance-wise finetuning phases. The results are shown in Table 5. Our model is robust against non-ideal blur settings and still outperforms the baselines by a substantial margin.

5.5 Supervised Learning

As most existing downscaling models are supervised, we extend our investigation to ascertain whether the integration of channel segregation and topoclimatic attention components can improve performance in scenarios where supervised climate downscaling is feasible. We present preliminary results

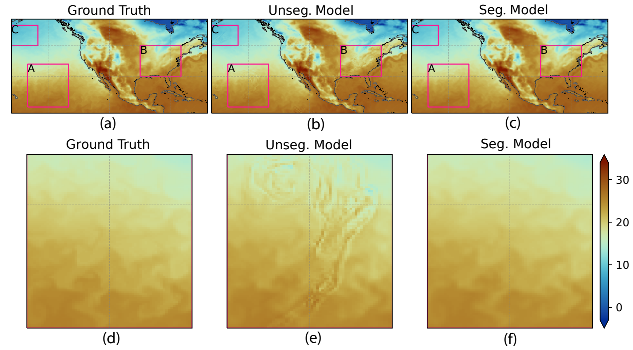

Figure 4: Ground truth HR data (a), compared to downscaled output without channel segregation (b), and with channel segregation (c). Regions with noticeable artifacts are marked with pink rectangles. Region A is zoomed-in Figures (d), (e), and (f). RMSE values for (b) and (c) are smaller than bicubic interpolation, but model output without channel segregation suffers from undesirable artifacts.

Model	TS	PRECT
Bilinear	3.4363	2.1406
Bicubic	4.3781	2.0650
Base	3.6050	**1.3096**
Ours	**3.3013**	1.3908

Table 5: RMSE of standard approaches and deep learning models, trained and tested on data with random noise by 2x scaling factor. The experiment is performed over 1,000 data points.

for this analysis on the CMIP6-ERA5 (LR-HR, respectively) data pairs. The LR data in these experiments is not a degraded version of the HR data but a separate dataset that emulates the real data. While CMIP6 is a collection of various coupled climate simulation models, ERA5 is the reanalysis data, that combines observational data and numerical predictions to create a continuous record of the past weather. For these datasets as well, we downscale TS and PRECT, while HR topography (PHIS) is the third variable that aids the model learning.

To reduce computational complexity, we regrid ERA5 data with CDO tool to 1° from the original resolution of 0.25°. Similarly, CMIP6 data is regridded from 1° to 3.75°. Thus, the LR CMIP6 data is downscaled with a factor of 3.75x (48x96 → 128x256). Both datasets are reconciled for spatial grid and temporal scopes, and the final dataset covers daily averages from the year 1979 to 2014. Moreover, to limit the extraneous perplexity, we also clip the polar regions from the data and utilize data in the region of 70° North and South of the equator. We use SRResNet as the baseline and the backbone network for all experiments [Ledig *et al.*, 2017]. The model was trained on $X - X^{HR}$ pairs, where $X \in CMIP6$ and $X^{HR} \in ERA5$. The data is split into train, validation, and test data as 7:1:2. The model was trained for 50 epochs, and the model checkpoint with minimum validation error was chosen for evaluation. Table 6 shows that using channel segregation and topoclimatic attention improves performance in both the channels. This preliminary analysis highlights that the proposed components are effective in general for deep learning

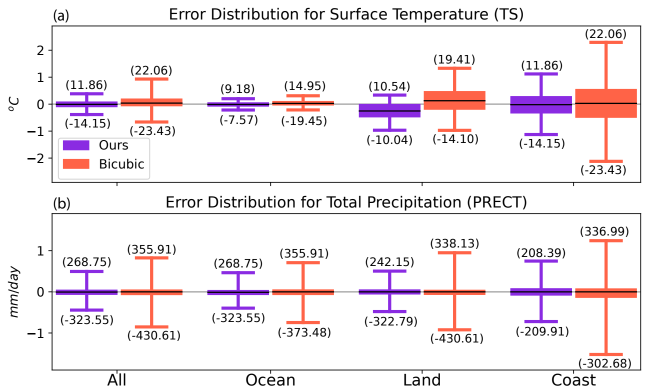

Figure 5: Error distribution from our model (purple) and bicubic interpolation (red) for TS (a) and PRECT (b). The vertical edges of solid boxes represent Q1 and Q3 of the error distribution.

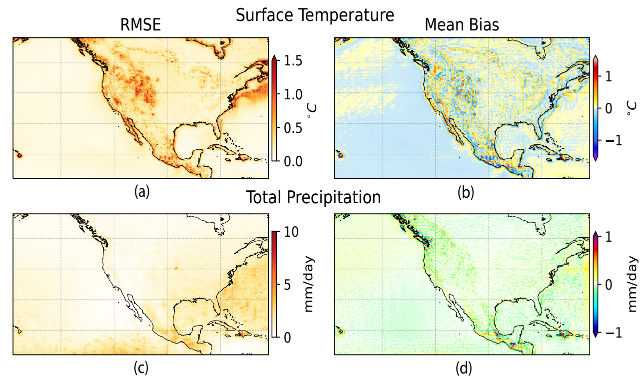

Figure 6: The spatial distribution of RMSE and mean bias across surface temperature (TS) and total precipitation (PRECT) for downscaling by the proposed model for North America.

Model	TS	PRECT
Bilinear	5.1150	3.0349
Bicubic	5.2944	3.0454
SRResNet	2.8484	2.7114
SRResNet+CS	2.7469	2.7156
SRResNet+Att	2.6632	2.7147
SRResNet+CS+Att	**2.6365**	**2.7103**

Table 6: Supervised experiment on CMIP6 (LR) and ERA5 (HR) data. The components in the table are abbreviations of Channel Segregation (CS) and Topoclimatic Attention (Att).

based climate downscaling models.

6 Discussion

We now compare our model with the bicubic interpolation baseline (standard approach in climate downscaling) and discuss its strengths and weaknesses in different domains within the region of interest. The metrics used for this analysis are the RMSE and mean bias (Figure 6). We look at the error distribution within domains such as oceans, land, and coastal areas (Figure 5). Analyses are performed for 2x scaling factor.

The two variables under consideration show considerable differences in their spatiotemporal characteristics. As the noisier variable, the PRECT has a larger error distribution spread than TS. Our model's error distribution (purple plots) shows improvement in all error metrics. The values corresponding to the 5^{th}, 25^{th}, 75^{th}, and 95^{th} percentiles are lower in our model than in the bicubic interpolation, extending the confidence in our model to perform better downscaling.

Even the maximum and minimum (above and below whiskers, 95^{th} and 5^{th} percentile) values of errors are smaller in our model (Figure 5 a,b). We emphasize that our inferences are valid for the entire region and separately in the different domains (ocean, land, and coast) within it. Both the bicubic interpolation baseline and our model demonstrate lower performance in TS over the coastal region (Figure 5a). PRECT also shows similar behavior, albeit with smaller differences between different domains (Figure 5b).

Figure 6 shows the spatial distribution of RMSE and mean bias in North America. We find that TS exhibits relatively large RMSE values over the mountainous regions. The errors are also high along the warm western boundary current along the east coast (Figure 6a). The former might stem from the model's inability to resolve mountainous topography well, while the latter is due to the unresolved small-scale eddies formed along the boundaries of the current system. Our model underestimates the TS over the Pacific and Atlantic Oceans (Figure 6b). The RMSE spatial distribution for PRECT indicates higher values at lower latitudes and along the continent's eastern coast (see Figure 6c). The errors are introduced on either side of the elevated terrain (Figure 6d). Therefore, error hot spots include coastlines, mountainous terrains, and eddy-rich oceanic regions. The above suggests that our model consistently outperforms baseline interpolation across all standard error distribution metrics.

7 Conclusion

This work proposed a deep learning-driven climate downscaling model that does not require HR ground truth data for training. Our model incorporates three components into the standard CNN model to make it adaptable to climate data: self-supervised pretraining that enables knowledge transfer from the input data and improves runtimes, channel segregation that enables learning better representations for individual weather variables, and topoclimatic attention that enforces the learning of the intricate relationship between the weather variables and topography.

Our work opens up possibilities such as effective downscaling of historical data, running larger climate ensembles and democratizing climate research with a reduced need for computational power. In the future, we would like to better understand trade-offs between weather variables of different natures, incorporating temporal dimension, extending the analysis in the supervised setting and newer methodologies like vision transformers, diffusion models, and improving the explainability of the deep learning downscaling model.

Acknowledgments

The authors would like to thank Axel Timmermann for his insightful feedback and Sun-Seon Lee for sharing the High-resolution CESM Climate simulation data. The work was supported by the Institute for Basic Sciences (IBS), South Korea, under IBS-R029-C2 (Singh, Shidqi, Park, and Cha) and IBS-R028-D1 (Nellikkattil and Zeller). This work was also partly supported by the National Research Foundation of Korea grant RS-2022-00165347 (Shidqi, Park, and Cha).

References

[Assaf Shocher, 2018] Michal Irani Assaf Shocher, Nadav Cohen. "zero-shot" super-resolution using deep internal learning. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[Baño-Medina *et al.*, 2020] Jorge Baño-Medina, Rodrigo Manzanas, and José Manuel Gutiérrez. Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4):2109–2124, 2020.

[Chou *et al.*, 2021] Christopher Chou, Junho Park, and Eric Chou. Generating high-resolution climate change projections using super-resolution convolutional LSTM neural networks. In *proc. of the International Conference on Advanced Computational Intelligence (ICACI)*, pages 293–298. IEEE, 2021.

[Chu *et al.*, 2020] Jung-Eun Chu, Sun-Seon Lee, Axel Timmermann, Christian Wengel, Malte F Stuecker, and Ryohei Yamaguchi. Reduced tropical cyclone densities and ocean effects due to anthropogenic greenhouse warming. *Science Advances*, 6(51):eabd5109, 2020.

[Cornillère *et al.*, 2019] Victor Cornillère, Abdelaziz Djelouah, Yifan Wang, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics (TOG)*, 38:1 – 13, 2019.

[Dong *et al.*, 2014] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing.

[Ducournau and Fablet, 2016] Aurelien Ducournau and Ronan Fablet. Deep learning for ocean remote sensing: an application of convolutional neural networks for super-resolution on satellite-derived sst data. In *proc. of the IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS)*, pages 1–6, 2016.

[Finn *et al.*, 2017] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *proc. of the International Conference on Machine Learning (ICML)*, pages 1126–1135. PMLR, 2017.

[Glasner *et al.*, 2009] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *proc. of the IEEE International Conference on Computer Vision (ICCV)*, pages 349–356, 2009.

[Groenke *et al.*, 2020] Brian Groenke, Luke Madaus, and Claire Monteleoni. Climalign: Unsupervised statistical downscaling of climate variables via normalizing flows. In *proc. of the International Conference on Climate Informatics*, pages 60–66, 2020.

[H.-O. Pörtner, 2022] E.S. Poloczanska K. Mintenbeck M. Tignor A. Alegría M. Craig S. Langsdorf S. Löschke V. Möller A. Okem H.-O. Pörtner, D.C. Roberts. *Summary for Policymakers*, pages 3–33. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2022.

[Haris *et al.*, 2018] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

[Hurrell *et al.*, 2013] James W Hurrell, Marika M Holland, Peter R Gent, Steven Ghan, Jennifer E Kay, Paul J Kushner, J-F Lamarque, William G Large, D Lawrence, Keith Lindsay, et al. The community earth system model: a framework for collaborative research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360, 2013.

[Ji *et al.*, 2020] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, pages 1914–1923, 2020.

[Kim *et al.*, 2016] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.

[Kumar *et al.*, 2021] Bipin Kumar, Rajib Chattopadhyay, Manmeet Singh, Niraj Chaudhari, Karthik Kodari, and Amit Barve. Deep learning–based downscaling of summer monsoon rainfall data over indian region. *Theoretical and Applied Climatology*, 143:1145–1156, 2021.

[Ledig *et al.*, 2017] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[Lim *et al.*, 2017] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, July 2017.

[Lu *et al.*, 2022] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022.

[Misra *et al.*, 2018] Saptarshi Misra, Sudeshna Sarkar, and Pabitra Mitra. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and applied climatology*, 134:1179–1196, 2018.

[Palmer *et al.*, 2014] TN Palmer, A Döring, and G Seregin. The real butterfly effect. *Nonlinearity*, 27(9):R123, 2014.

[Park *et al.*, 2022] Sungwon Park, Karandeep Singh, Arjun Nellikkattil, Elke Zeller, Tung Duong Mai, and Meeyoung Cha. Downscaling earth system models with deep learning. In *proc. of the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, aug 2022.

[Rampal *et al.*, 2022] Neelesh Rampal, Peter B Gibson, Abha Sood, Stephen Stuart, Nicolas C Fauchereau, Chris Brandolino, Ben Noll, and Tristan Meyers. High-resolution downscaling with interpretable deep learning: Rainfall extremes over new zealand. *Weather and Climate Extremes*, 38:100525, 2022.

[Seneviratne *et al.*, 2021] S.I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou. *Weather and Climate Extreme Events in a Changing Climate*, page 1513–1766. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

[Small *et al.*, 2014] R Justin Small, Julio Bacmeister, David Bailey, Allison Baker, Stuart Bishop, Frank Bryan, Julie Caron, John Dennis, Peter Gent, Hsiao-ming Hsu, et al. A new synoptic scale resolving global climate simulation using the community earth system model. *Journal of Advances in Modeling Earth Systems*, 6(4):1065–1094, 2014.

[Soh *et al.*, 2020] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3516–3525, 2020.

[Stengel *et al.*, 2020] Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020.

[Sun *et al.*, 2019] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 403–412, 2019.

[Vandal *et al.*, 2017] Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R. Ganguly. DeepSD. In *proc. of the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, aug 2017.

[Vandal *et al.*, 2019] Thomas Vandal, Evan Kodra, and Auroop R Ganguly. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology*, 137:557–570, 2019.

[Wang *et al.*, 2021] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10581–10590, 2021.

[Zhang *et al.*, 2018a] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[Zhang *et al.*, 2018b] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[Zhang *et al.*, 2018c] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.

[Zontak and Irani, 2011] Maria Zontak and Michal Irani. Internal statistics of a single natural image. In *proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 977–984, 2011.