시간 정보를 활용한 자기 지도 기후 초해상화: 북아메리카 지역을 중심으로

Naufal Shidqi[1,2], 정채윤[1,2], 박성원[1,2], Karandeep Singh[2], Arjun Babu Nellikkattil[3,4], Elke Zeller[3,4], 차미영[2,1]
[1]한국과학기술원 전산학부, [2]기초과학연구원 데이터 사이언스 그룹, [3]부산대학교 기후시스템전공, [4]기초과학연구원 기후물리 연구단
{naufal.shidqi, lily9991, psw0416}@kaist.ac.kr, ksingh@ibs.re.kr,
{arjunbabun, elkezeller}@pusan.ac.kr, mcha@ibs.re.kr

Leveraging Temporal Information for Self-Supervised Climate Downscaling in North America Region

Naufal Shidqi[1,2], Chaeyoon Jeong[1,2], Sungwon Park[1,2], Karandeep Singh[2], Arjun Babu Nellikkattil[3,4], Elke Zeller[3,4], Meeyoung Cha[2,1]
[1]School of Computing, KAIST, [2]Data Science Group, IBS, [3]Department of Climate System, PNU, [4]Center for Climate Physics, IBS

요 약

Climate downscaling is an important technique in climate research that projects low-resolution (LR) climate data to high-resolution (HR). Prior research has shown deep learning to be an effective technique for climate downscaling. However, most of the deep learning models for climate downscaling are fully supervised and require large amounts of high-resolution (HR) data for training, which is rarely available in practice. Additionally, the focus of these models is mostly on learning the spatial dependencies between the weather variables, whereas temporal information is generally ignored. In order to tackle these problems, this paper proposes a self-supervised convolutional neural network (CNN) model for downscaling climate data. We use a self-supervised learning setting that has no dependency on high-resolution (HR) data and incorporates the temporal climate variability into the model using a long short-term memory (LSTM) network. The proposed model is evaluated using precipitation and surface temperature data from the Community Earth System Model (CESM) v1.2.2 simulation. Results show significant improvements over existing baselines, demonstrating the effectiveness of capturing spatio-temporal dependencies in downscaling climate data.

1. Introduction

The Earth's climate is changing rapidly, and its impact on the planet is becoming increasingly apparent. To mitigate the effects of climate change, researchers must understand the mechanisms behind climate variability and predict future climate patterns. One of the major challenges in climate research is the ability to accurately project global climate data to a regional or local scale, which is known as "climate downscaling." There are two kinds of downscaling methods: dynamical downscaling and statistical downscaling. Dynamical downscaling employs high-resolution (HR) regional climate models to mimic the interactions between the atmosphere, land, seas, and other climate variables. This technique provides a more comprehensive depiction of local climate conditions, but it requires a substantial amount of computational power [1]. On the other hand, statistical downscaling utilizes climate data to find statistical relationships between HR climate models and observed local climate states. However, conventional statistical methods fail to capture the complex relationships between global and local climate patterns and interactions between different climate variables. This statistical downscaling can be considered a super-resolution (SR) task in computer vision research. While the SR task enhances the resolution of low-resolution *images*, statistical downscaling improves the resolution of coarse *climate data*.

Previous studies have shown deep learning to be a promising approach for downscaling climate models. Vandal *et al*. used a deep SR convolutional neural network (SRCNN) method to downscale climate variables [2]. Other studies have also used other deep learning models such as GAN, LSTM, and other deep neural network-based methods to generate climate data on a local scale by incorporating climate-physics properties into their models [3, 4, 5]. However, all such models are fully supervised and typically require large amounts of HR ground truth data, which is computationally expensive and difficult to obtain. Furthermore, only a few of the prior studies have incorporated the temporal aspect of the climate data, which dynamically evolves over time, as illustrated in **Figure 1**.

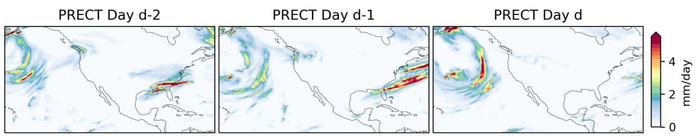

Figure 1. A three-day snapshot of total precipitation (PRECT) over the North America region. The temporal evolution can be clearly observed.

This paper presents a self-supervised CNN model for downscaling climate variables, given that no HR climate data is available for model training. The proposed model uses a self-supervised learning setting that downscales specific LR data points by learning their characteristics at the runtime. Moreover, the proposed model also incorporates temporal variability into our model by employing an LSTM network to capture the time-varying dynamics of the weather variables. We evaluate our model by downscaling total precipitation (PRECT) and surface temperature (TS) from the Community Earth System Model (CESM) v1.2.2 into 2x and 4x scale factors. To the best of our knowledge, this is the first study to investigate temporal information for a self-supervised climate downscaling model. The results obtained show substantial improvements in the performance compared to existing baselines, indicating that capturing the spatio-temporal dependencies is a promising approach for downscaling the climate data.

2. Methods

As the HR, or LR-HR paired climate data, is challenging to obtain due to the enormous computational complexity involved in the HR simulations, we employ a self-supervised setting for model training that has no dependency on the HR data. Given a model input X_t at specific time t, we generate pseudo-LR data $X_t{\downarrow}_d$ (d = downgrading factor) by

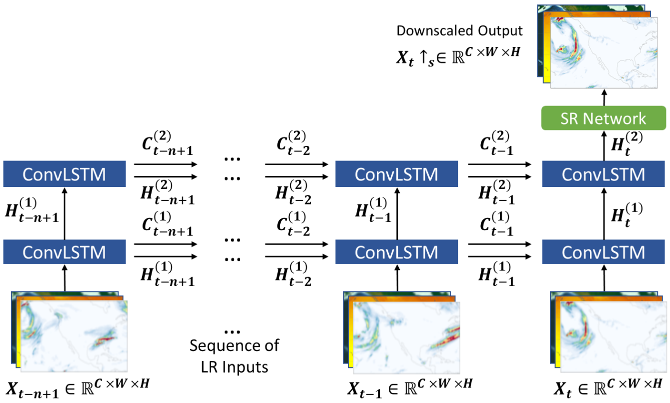

Figure 2. The overall architecture of the proposed model. A sequence of LR data points (X_{t-n+1}, ..., X_{t-1}, X_t) is passed to the model, yielding a downscaled output $X_t\uparrow_s$. Surface temperature (TS), total precipitation (PRECT), and topography gradient (dPHIS) are the model's input channels.

downgrading X_t, then train the model with the $X_t\downarrow_d - X_t$ pair as the pseudo LR - pseudo HR pair. Our aim is to obtain downscaled output $X_t\uparrow_s$(s = scaling factor, s = 1/d) by performing model inference on X_t, as depicted in **Figure 2**.

2.1 Convolutional LSTM

We employ a convolutional LSTM (ConvLSTM) network to handle n number of spatial data X_{t-n+1}, ..., X_{t-1}, X_t, recorded up to time t periodically [4, 6]. The LSTM can maintain temporal context in its current memory state C_t. Then, the memory state will be propagated through controlling gates to its hidden state H_t. The formula for a ConvLSTM cell at time t is defined as follows:

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f) \\
C_t &= f_t \odot C_{t-1} + i_t \odot tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \\
H_t &= o_t \odot tanh(C_t)
\end{aligned} \tag{1}
$$

where i_t, f_t, o_t are input, forget, and output gates; b are bias vectors; W are weight matrices; σ represents sigmoid function; $*$ denotes the convolution operation; and $\odot$ denotes Hadamard product. We then stack two ConvLSTM cells to encode more complex dynamical patterns in the spatio-temporal space of the climate data [6].

2.2 Super-resolution (SR) Network

The intermediate hidden state from the last ConvLSTM layer is then passed to our super-resolution (SR) network. We adopt SRResNet [7] as the network's backbone, with 32 channels and 8 ResNet blocks, and two CNN layers subsequently. The CNN layers in our downscaling network are designed to maintain the size of input data. The downscaling network operations are described as follows:

$$
\begin{aligned}
F_1 &= ResNetBlocks(H_t^{(2)}) \\
F_2 &= ReLU(IN(CONV(F_1))) \\
X_t\uparrow_s &= CONV(F_2 + H_t^{(2)})
\end{aligned} \tag{2}
$$

where IN is an instance normalization layer, $CONV$ is a CNN layer, and $X_t\uparrow_s$ is the output of the downscaling network with s scale factor.

2.3 Self-Supervised Learning

As discussed before, obtaining real LR-HR data is challenging in practice. Motivated by [8] and [9], we generate pseudo-LR and pseudo-HR for each instance, train a fresh model on the pseudo pair, and finally, perform model inference on LR data to generate the downscaled HR data. The intention behind this is to make the model learn the internal structure of the specific instance [8]. We synthesize the LR data at timestep t, X_t, from the original HR data at timestep t, X_t^{HR}, by bicubic-interpolation to a coarser resolution. This downgrading process is irreversible, implying that it is impossible to restore the original X_t^{HR} by merely interpolating back to its original resolution. The input X_t is then downgraded again to generate $X_t\downarrow_d$. We regard the $X_t\downarrow_d - X_t$ pair as a pseudo-LR and pseudo-HR data pair for our model training, i.e., $X_t = f_\theta(X_t\downarrow_d)$, where f_θ represents our model. Note that the original HR data (X_t^{HR}) is not used in model training. Then, the objective of the model training is to minimize the loss function L:

$$
L = \frac{1}{C \times W \times H} \sum_{C,W,H} MSE(f_\theta(X_t\downarrow_d),\ X_t) \tag{3}
$$

where MSE is an element-wise mean squared error. The inference of the model can be performed by passing the LR input data (pseudo-HR), X_t, to the trained model, which produces the final output, $X_t\uparrow_s = f_\theta(X_t)$. The performance of the model is calculated by evaluating the root mean squared error (RMSE) as follows:

$$
Error(X_t, \theta) = \frac{1}{C \times W \times H} \sum_{C,W,H} RMSE(f_\theta(X_t),\ X_t^{HR}) \tag{4}
$$

3. Experiment and Evaluation

3.1 Data and Temporal Settings

We used CESM v1.2.2 HR climate simulation data ($X_t^{HR} \in \mathbb{R}^{C \times H \times W}$) for the present-day condition of daily means over the North America region for 20 years (N = 20 x 365 = 7300). The dataset is configured sequentially in time. Each data point has a resolution of 213×321, with each pixel representing ~25 km grid resolution. Surface temperature (TS), total precipitation (PRECT), and gradient of topography (dPHIS) are the three climate variables that are used as the model's input data channels. The LR data ($X_t \in \mathbb{R}^{C \times H' \times W'}$) is synthesized by bicubic interpolation from the HR data into a coarser resolution ($H' \times W'$), and rescaled to the target output size ($H \times W$) by applying bicubic interpolation before proceeding to the model's input. We downscale PRECT and TS, as dPHIS remains constant during the simulation time.

3.2 Experiment Setup

As described in Section 2.3, we train a new model for every data instance at runtime. We use LeakyReLU with a negative slope of 0.2, instance normalization, the Adam optimizer, uniform convolution kernels of size 3, and padding and stride of 1 throughout the downscaling network. The model is run for 3000 epochs at a learning rate of 1E-4. The same ResNet backbone and self-supervised setting are used for all models and baselines. We evaluate the model by downscaling 500 data points from the CESM v1.2.2 climate simulation data (500 data points with the shape 3 × 213 × 321). For each target LR data point, two prior data points are utilized to extract the temporal information. The model's performance is measured by calculating the root mean squared error (RMSE) by comparing the model output and the original HR data over 2x and 4x scale factors.

3.3 Results and Discussion

Table 1 shows the model performance of downscaling PRECT and TS data for 2x and 4x scale factors. We evaluated the model against bilinear and bicubic interpolation, SRResNet with 2D and 3D convolution, and ours in a supervised setting. Bilinear and bicubic interpolation are widely used as conventional downscaling methods. SRResNet is a ResNet-based deep learning model for the SR task. Moreover, to incorporate temporal information, we additionally implemented SRResNet with 3D convolution. Bilinear, bicubic, and

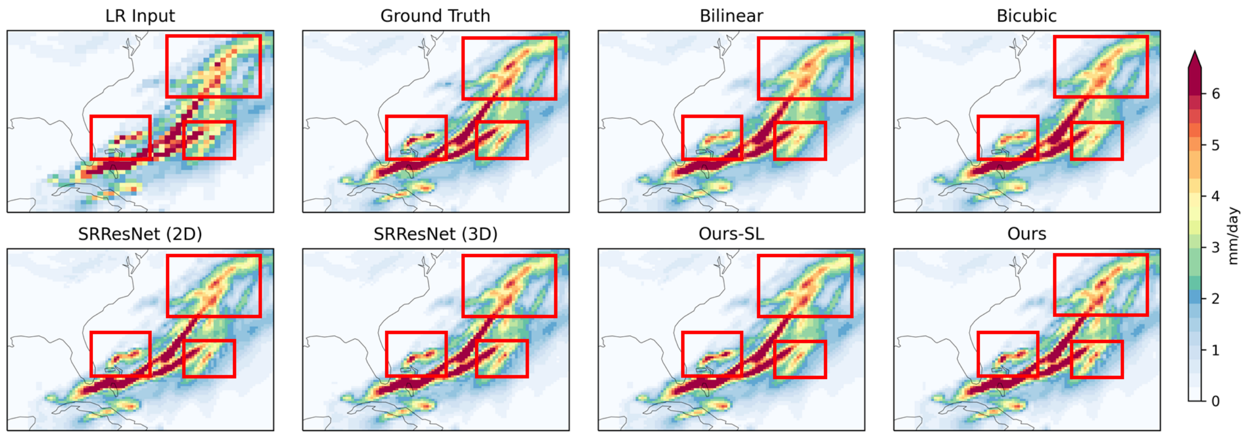

Figure 3. Visual examples of total precipitation (PRECT) from LR input, ground truth, bilinear, bicubic, SRResNet (2D), SRResNet (3D), Ours-SL (supervised learning), and Ours (self-supervised learning) for the same data point. The differences are indicated in the red squares.

Table 1. The RMSE (lower is better) comparison of conventional methods and deep learning models for 2x and 4x (units are 1e-8 for PRECT).

	Model	2x		4x	
		TS	PRECT	TS	PRECT
Spatial	Bilinear	0.6741	1.9102	0.8791	3.1666
	Bicubic	0.6936	1.8549	0.8803	3.1916
	SRResNet (2D)	0.5178	1.1697	0.8119	3.0708
Spatio-temporal	SRResNet (3D)	0.5433	1.1560	0.8240	3.0676
	Ours-SL	0.4583	1.1437	0.8391	**2.7912**
	Ours	**0.4389**	**1.0653**	**0.7777**	3.0417

SRResNet (2D) represent spatial models, while SRResNet (3D) and our models constitute spatio-temporal models. It is evident that our model outperforms baselines for both scale factors. In the 2x scale factor, our model improves the performance by reducing the RMSE with 29.85% (TS) and 46.87% (PRECT) decreases over the bicubic, and 20.32% (TS) and 8.96% (PRECT) decreases over the SRResNet (3D). Moreover, our model has better performance than Ours-SL (trained with HR data in a supervised setting) and the SRResNet (2D). The superior performance of our model highlights the importance of utilizing temporal information for the downscaling task. Similar performance improvements are also observed in the 4x scale factor, albeit our model in supervised setting is able to learn more for PRECT.

We further explored the number of prior data points used as the temporal information for the target data point, as presented in **Figure 4**. Due to the computational bottleneck, the results presented here are for 200 data points. The figure illustrates that our model consistently outperforms the SRResNet (2D). Although the performance improves when previous temporal points are provided (compare when $N > 0$ and $N = 0$), an increase in the number of prior data points led to an increase in the RMSE value. This indicates that while the immediate past data may be informative for downscaling the target data point, the distant ones may possess weak correlations with the target data point.

4. Conclusion and Future Work

In this paper, we propose a self-supervised spatio-temporal deep learning model that can effectively downscale climate variables using only low-resolution (LR) climate data. Our model incorporates temporal dynamics by employing an LSTM network, and we demonstrate its effectiveness by extensive evaluations on total precipitation (PRECT) and surface temperature (TS) variables obtained from the Community Earth System Model (CESM) v1.2.2. The proposed method offers a promising approach for downscaling climate data. It also indicates the efficacy of incorporating spatio-temporal dependencies in downscaling climate data. Future work could explore the use of other deep learning backbones and more datasets to provide further insights and improvements to our proposed model. This study stimulates further climate research explorations, such as predicting future climate patterns and mitigating extreme future climate events.

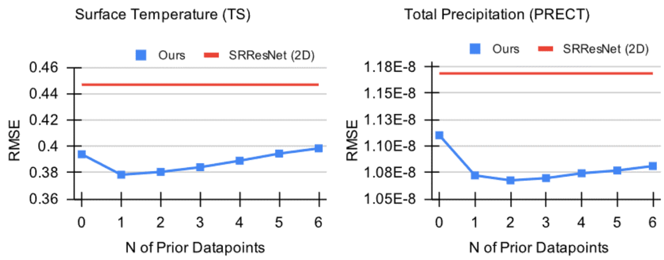

Figure 4. Performance comparison of Ours and SRResNet (2D) for TS and PRECT in the 2x scale factor over a number of prior data points.

Acknowledgments

The climate simulations were conducted on the IBS/ICCP supercomputer "Aleph", a 1.43 petaflops high-performance Cray XC50-LC Skylake computing system. This work was supported by the Institute for Basic Sciences (IBS), Republic of Korea, under IBS-R029-C2 and IBS-R028-D1.

References

[1] R. J. Small et al., "A new synoptic scale resolving global climate simulation using the Community Earth System Model," Journal of Advances in Modeling Earth Systems, vol. 6, no. 4, pp. 1065–1094, 2014.
[2] T. Vandal, E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, "DeepSD," Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.
[3] J. Cheng, J. Liu, Z. Xu, C. Shen, and Q. Kuang, "Generating High-Resolution Climate Prediction through Generative Adversarial Network," Procedia Computer Science, vol. 174, pp. 123–127, 2020.
[4] C. Chou, J. Park, and E. Chou, "Generating High-Resolution Climate Change Projections Using Super-Resolution Convolutional LSTM Neural Networks," Proceedings of the 13th International Conference on Advanced Computational Intelligence, 2021.
[5] S. Park, K. Singh, A. Nellikkattil, E. Zeller, T. D. Mai, and M. Cha, "Downscaling Earth System Models with Deep Learning," Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022.
[6] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," Neural Information Processing Systems, 2015.
[7] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[8] A. Shocher, N. Cohen, and M. Irani, "'Zero-Shot' Super-Resolution using Deep Internal Learning." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[9] K. Singh, C. Jeong, S. Park, A. N. Babur, E. Zeller and M. Cha, "Self-supervised learning for climate downscaling," IEEE International Conference on Big Data and Smart Computing (BigComp), 2023.