

Downscaling climate projections using single-image super-resolution

Andrew McDonald, Drew Hayward, Tyler Lovell, Sanjeev Thenkarai Lakshmi Narasimhan
Michigan State University
East Lansing, MI

{mcdon499, haywar40, lovellty, thenkara}@msu.edu

Abstract

Climate change is one of the most pressing problems facing modern society, and understanding its effects across spatial scales is essential for risk prevention and mitigation. To advance this understanding, we propose to study the problem of downscaling from climate science through the lens of single-image super-resolution in computer vision, mapping climate projections from coarse spatial scales to fine spatial scales using deep learning. We adapt three common super-resolution network architectures from the computer vision literature—namely, SRCNN [5], VDSR [14], and SRResNet [18]—to downscale temperature and precipitation data from the ERA5 Reanalysis Project [11] over the continental United States by a factor of $4\times$, conduct extensive experimental evaluation, discuss salient strengths and weaknesses of each model, and conclude with a discussion on future directions for research at the intersection of climate science and computer vision. All code associated with the project is freely available on GitHub.¹

1. Introduction

According to the IPCC Sixth Assessment Report [13], it is “virtually certain” that anthropogenic greenhouse gas emissions have significantly altered the modern climate with respect to the pre-industrial era, and “virtually certain” that the global mean temperature has already risen by approximately 1°C . Barring immediate and radical reductions in carbon emissions, extreme weather will continue to become more frequent and more severe; polar land and sea ice will continue to recede, accelerating with the albedo effect; mean sea level will continue to rise with melting ice and thermal expansion; biodiversity will fall with habitat destruction and displacement; and perhaps most concerningly, global systems such as the Atlantic Meridional Overturning Circulation and El Niño/Southern Oscillation may destabilize, bringing an era of unprecedented climate tipping points

and cascading second-order effects. Understanding the consequences of climate change across spatial scales—from the global level down to the hyperlocal level—is absolutely essential if we hope to enact policies preventing the worst possible outcomes under climate change. An important subset of these consequences may be characterized and analyzed using the data-driven methods of machine learning [23], and in particular, using the methods of deep learning (DL) [7] in the context of computer vision (CV) [27]. Towards this end, we propose to study the problem of *downscaling*—a notoriously difficult, under-constrained problem, central to climate change and event impact analyses in climate science [22]—by framing it as a problem of single-image super-resolution (SISR), a well-studied problem in CV [32].

2. Problem Statement

Downscaling involves the mapping of coarse-scale measurements to fine-scale measurements, and naturally aligns with the task of SISR when viewing spatially-gridded climate data as multi-channel images: given measurements of one or more variable(s) (e.g., temperature and/or precipitation) at a low spatial resolution, we aim to infer the values of the same variable at a higher spatial resolution. Used to refine the coarse-scale output of global climate models (GCMs)—which must be run at 100km-by-6hr (or lower) resolutions due to their computational complexity—downscaling allows for more interpretable climate change projections by localizing impacts such as heatwaves and severe storms to the city-level.

More formally, given low-resolution climate measurements $\mathbf{X}_L \in \mathbb{R}^{c \times h_L \times w_L}$ from a grid of size $h_L \times w_L$ and corresponding high-resolution measurements $\mathbf{X}_H \in \mathbb{R}^{c \times h_H \times w_H}$ from a grid of size $h_H \times w_H$, we aim to learn a mapping f_θ for which $f_\theta(\mathbf{X}_L) = \mathbf{X}_H$. Here, c denotes the number of physical climate variables we consider, akin to the number of color channels in an image, and θ denotes the set of learnable parameters associated to f_θ , typically implemented as a deep neural network (DNN). To make this mathematical formulation concrete, suppose we map temperature and precipitation from a spatial scale of 1° lati-

¹<https://github.com/andrewmcdonald27/CSE803FinalProject>

tude/longitude to a finer scale of 0.5° latitude/longitude using a ResNet—then, recalling that there are 180° latitude and 360° longitude on a sphere, we would have $h_L \times w_L = 180 \times 360$, $h_H \times w_H = 360 \times 720$, $c = 2$, and f_θ a ResNet with θ being the parameters of the ResNet.

If the mapping f_θ is a DNN, we may optimize its parameters θ by defining a loss $\mathcal{L}(f_\theta(\mathbf{X}_L), \mathbf{X}_H)$ for which $\lim_{\mathbf{X}_L \rightarrow \mathbf{X}_H} \mathcal{L} = 0$ and apply backpropagation [24] to minimize this loss. Typical loss functions used to train a downscaling network include pixel-wise mean squared error (MSE) and mean absolute error (MAE), though more complicated adversarial and content losses have been proposed in the SISR literature to encourage synthesis of images favorable to human perception [18]. Performance of a particular downscaling approach is generally evaluated in terms of pixel-wise mean squared error, pixel-wise mean absolute error, or pixel-wise correlation due to the physical, climatological significance of such metrics, though more general SISR techniques are often evaluated on the basis of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [30], or mean opinion score collected by human judges (MOS) [18], to emphasize image properties more important to human perception.

3. Related Work

SISR has gained popularity as a subfield in CV over the last 5-10 years, experiencing rapid advances in the wake of the deep learning revolution initiated by the breakthrough performance of AlexNet [15] on ImageNet [4]. In particular, the inductive biases of convolutional neural networks (CNNs) [17] promoting local feature extraction, hierarchical feature extraction, parameter sharing, and translational equivariance have proven essential to the incorporation of deep learning in SISR. Foundational deep learning-based SISR advances include the super-resolution CNN (SRCNN) architecture proposed in [5], the very deep super-resolution (VDSR) architecture proposed in [14], the super-resolution residual network (SRResNet) and super-resolution generative adversarial network (SRGAN) architectures proposed in [18], along with architectures such as efficient subpixel CNN (ESPCN) proposed in [25] and enhanced deep super-resolution (EDSR) proposed by [19]. We refer the reader to [31, 32] for comprehensive reviews of the state of modern SISR literature.

More recently, SISR has been adopted as an approach to downscaling within the climate science community. Vandal et al. [28] proposed DeepSD, downscaling daily precipitation from the PRISM dataset [2] from 1° to $1/8^\circ$ resolution. Subsequently, Liu et al. [20] proposed YNet, downscaling precipitation and temperature projections from an ensemble of 35 GCMs hosted by NASA. While these domain-specific approaches often outperform SISR architectures designed to operate on natural image data, domain-agnostic SISR

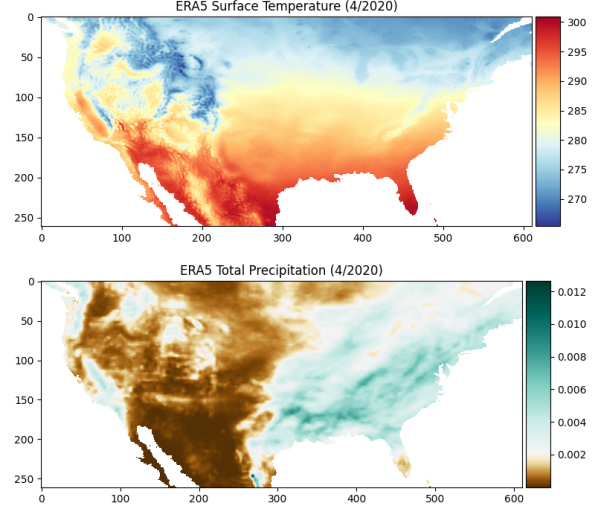


Figure 1. Average 2m temperature (top) and total precipitation (bottom) across the continental US in April 2020, sourced from ERA5. Diverging colormaps are centered at the mean value across all pixels to account for skewed distributions.

networks applied to climate data [1, 3, 16, 29] remain impressively performant and have gained community attention through events such as the Climate Change AI Workshops² at ICLR, ICML, and NeurIPS, and the EarthVision Workshop³ at CVPR.

4. Approach

4.1. Data

We source gridded climate data collected at a 0.1° resolution over the United States from 1950-2020 from the ERA5 Reanalysis Project [11] through the EU Copernicus Climate Data Store.⁴ Specifically, we obtain monthly averages of t_{2m} , temperature measured at 2m above ground level, and t_p , total precipitation, on a 0.1° grid ranging from longitudes 127°W to 66°W and latitudes 24°N to 50°N , amounting to a dataset of 852 images (12 months per year \times 71 years) with each image consisting of 2 channels (t_{2m} and t_p) having height 261 (10 pixels per degree \times 26° latitude) and width 611 (10 pixels per degree \times 61° longitude). Although ERA5 data is *not* GCM data—ERA5 consists of historical observations and not future projections—downscaling ERA5 observations defines a problem virtually equivalent to downscaling GCM projections while being a more convenient dataset to work with. Sample t_{2m} and t_p channels extracted from a single image in the dataset are presented in Figure 1.

ERA5 data is provided in .nc (NetCDF) files by EU

²<https://www.climatechange.ai/papers?>

³<http://www.classic.grss-ieee.org/earthvision2021/index.html>

⁴<https://cds.climate.copernicus.eu/cdsapp>

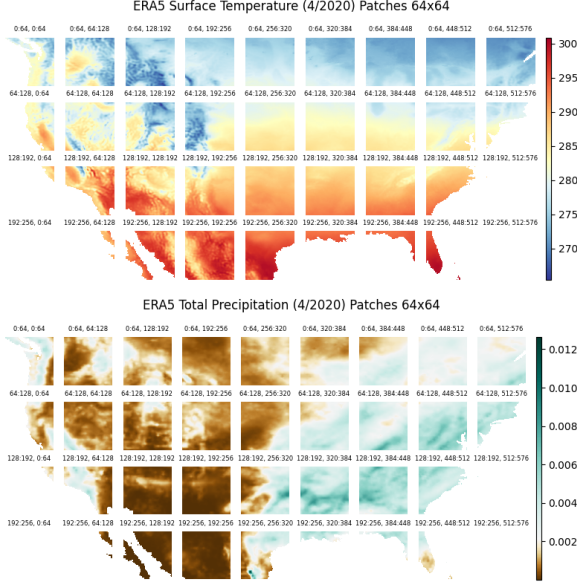


Figure 2. Patches of t_{2m} (top) and t_p (bottom) of shape 64×64 across the continental US in April 2020, sourced from ERA5.

Copernicus, which we extract into Numpy and PyTorch with shape $n \times c \times h \times w = 852 \times 2 \times 261 \times 611$. We subsequently split each 261×611 image of monthly-average t_{2m} and t_p over the US into a set of 64×64 patches for two reasons: (i) splitting images into patches expands the size of our dataset far beyond 852 images, effectively functioning as a dataset augmentation technique, and (ii) training an SISR model on small 64×64 patches is less computationally burdensome than training on entire 261×611 pixel images. In addition, a patch-based model will still be able to perform SISR on the original 261×611 pixel images through the use of a splitting, processing, and stitching pipeline at inference-time. Sample 64×64 patches from the dataset are presented in Figure 2.

To develop low-resolution input images \mathbf{X}_L and target high-resolution images \mathbf{X}_H , we upscale the original 0.1° data to 0.4° with average pooling—thus, the original 0.1° image serves as the target \mathbf{X}_H , and the artificially-pooled \mathbf{X}_L serves as model input. When applying average pooling to construct \mathbf{X}_L , we maintain patch spatial dimensionality at 64×64 , such that both \mathbf{X}_L and \mathbf{X}_H are of size 64×64 despite \mathbf{X}_L having a lower effective resolution. In other words, the 16 pixels in each 4×4 pool of \mathbf{X}_H remain 16 pixels in the corresponding 4×4 pool in \mathbf{X}_L , with all 16 pixels taking the same (average) value. Sample high-resolution patches at 0.1° and their corresponding low-resolution counterparts pooled $4 \times$ to 0.4° are presented in Figure 3.

Data is only valid over land surfaces in ERA5, and is set to NaN over oceans and lakes, as is evident in the white pix-

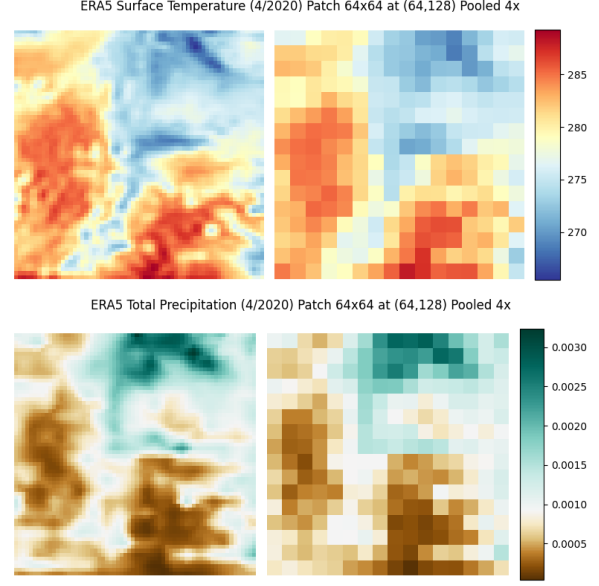


Figure 3. High-resolution (\mathbf{X}_H) and low-resolution (\mathbf{X}_L) patch of 2m temperature (top) and total precipitation (bottom) from April 2020, sourced from ERA5. Patch corresponds to that at row 64, column 128 in Figure 2 centered over the Rocky Mountains.

els in Figures 1-3. However, one may wish to downscale patches partially over land and partially over water, hence it is imperative to handle patches with partially-missing data. Towards this end, we implement a function which replaces NaN values in patches with the mean value of valid pixels in the patch, such that partial-NaN patches can be processed by downstream SISR models. For stability in downstream SISR model training, we discard patches having $>50\%$ NaN values and apply this mean-replacement approach on patches having $<50\%$ NaN values. An example patch to which this mean-replacement approach was applied is presented in Figure 4.

As a final preprocessing step, we normalize all images by subtracting the channel-wise mean and dividing by the channel-wise standard deviation. This is essential, as t_{2m} is measured in degrees Kelvin and takes values on the order of 10^2 , while t_p is measured in meters and takes values on the order of 10^{-2} .

4.2. Models

Here, we introduce and characterize the three SISR architectures considered in this work. For a comprehensive history of SISR and overview of the state of modern literature—which we believe may provide useful context to the remainder of this section—we refer the reader to [31, 32].

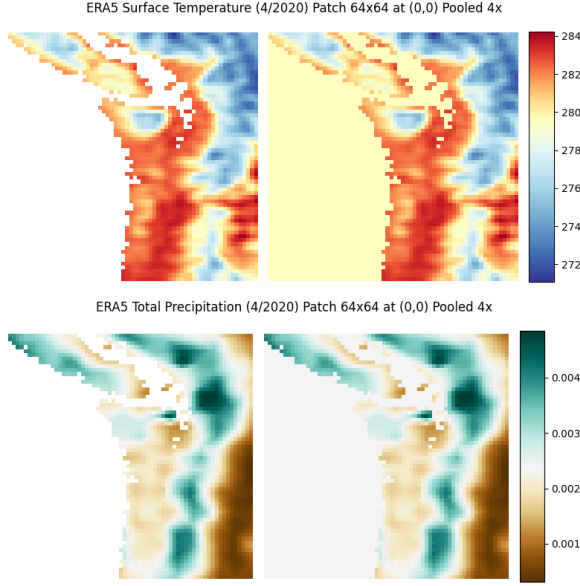


Figure 4. To deal with partial-NaN patches (left), we replace missing t_{2m} (top) and t_p (bottom) values with their in-patch mean (right). Patch corresponds to that at row 0, column 0 in Figure 2 centered over Washington and British Columbia.

4.2.1 SRCNN

Introduced in [5], the super-resolution CNN (SRCNN) was the first deep convolutional architecture proposed for single-image super-resolution, setting the baseline for state-of-the-art performance for years to come. While shallow by 2021 standards, SRCNN represented a significant step forward at the time of its introduction in both size and performance. SRCNN is fully-convolutional, consisting of three layers: the first layer extracts n_1 feature maps with a kernel of size f_1 on the input image of shape $c \times h \times w$; the second layer then extracts n_2 feature maps with a kernel of size f_2 ; and the final layer reconstructs an output image of shape $c \times h \times w$ with a kernel of size f_3 . In [5] and in our implementation, the architectural parameters are taken to be $(n_1, n_2) = (64, 32)$ and $(f_1, f_2, f_3) = (9, 1, 5)$. Rectified linear unit (ReLU) activation functions are applied after the first and second layer. The authors do not use padding and thus must contend with smaller output images when evaluating their approach; we use `padding="same"`, `padding_mode="replicate"` to maintain dimensionality. A schematic of SRCNN is presented in panel (a) of Figure 5.

4.2.2 VDSR

Introduced in [14], the very deep super-resolution (VDSR) architecture was inspired by the success of VGGNet [26] and the empirical observation that deeper networks outper-

form shallower networks when given sufficient computational resources, data, and time. The authors emphasize that a deeper architecture brings two main advantages to the task of SISR over SRCNN: first, deeper convolutional architectures capture a wider receptive field through kernel-stacking; and second, deeper neural architectures in general facilitate the representation of more complex functions through iterated (nonlinear) functional composition. The authors also employ a skip (residual) connection just prior to the output of the network, reasoning that this allows the network to simply learn a “refinement” of the input image.

VDSR consists of a single $k \times k$ convolutional input layer mapping images of shape $c \times h \times w$ to feature maps of shape $d \times h \times w$ with ReLU activation, followed by $\ell - 2$ blocks of $k \times k$ convolution and ReLU which preserve feature depth d . A $k \times k$ convolutional layer follows this sequence of blocks, mapping the image back to $c \times h \times w$, then added to the original input image through a skip connection just prior to the output of the network. We slightly modify this original formulation by moving the residual connection after the first convolution layer and prior to the final convolution layer, such that d -channel feature maps are residually-connected instead of c -channel images. This change facilitated more stable training. In [14], the architectural parameters are taken to be $(\ell, k, d) = (20, 3, 64)$; we use these parameters in our implementation as well, though we set the kernel size of the first and final convolution layers (before and after the modified residual connection) to 9×9 . We make one additional modification: in contrast to the original implementation’s use of zero-padding to maintain dimensionality throughout the network, we use `padding="same"`, `padding_mode="replicate"`. A schematic of our modified implementation of VDSR is presented in panel (b) of Figure 5.

4.2.3 SRResNet

Introduced in [18], the super-resolution residual network (SRResNet) drew inspiration from the success of ResNet [10], adapting the use of stacked residual blocks to the problem of SISR. SRGAN, an extension of SRResNet drawing inspiration from generative adversarial networks [8], is also introduced in [18]; but for simplicity we consider only the vanilla SRResNet in this work.

SRResNet begins by mapping input images of shape $c \times h \times w$ to tensors with feature depth d using a $k_1 \times k_1$ convolution activated by PReLU [9]. Next, SRResNet repeats ℓ residual blocks each consisting of depth-and-space-preserving (i) $k_2 \times k_2$ convolution, (ii) batch normalization [12], (iii) PReLU, (iv) $k_2 \times k_2$ convolution, (v) batch normalization, and (vi) a skip connection. Following residual blocks, the network applies one final $k_2 \times k_2$ convolution, one final batch normalization,

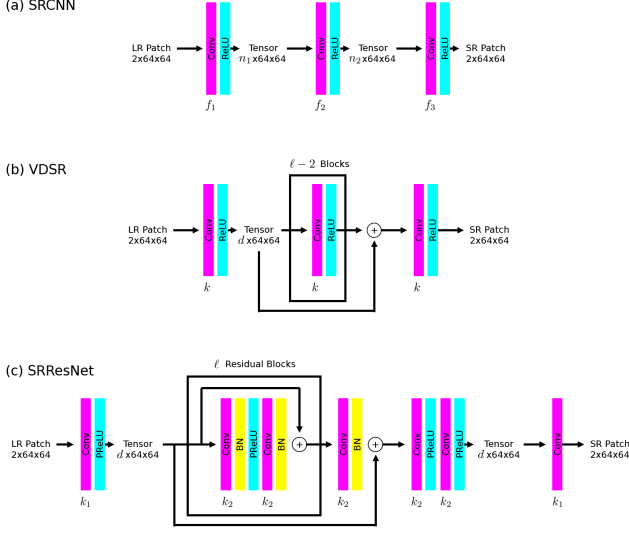


Figure 5. Architectures of (a) SRCNN [5], (b) VDSR [14], and (c) SRResNet [18]. We follow the implementation of each model described in the original work as closely as possible, though we make minor modifications to VDSR and SRResNet as described in the main text. Each network takes low-resolution patches \mathbf{X}_L and maps them to super-resolution patches $f_\theta(\mathbf{X}_L)$ approximating high-resolution patches \mathbf{X}_H .

one final skip connection from pre-residual-block feature maps, and one final $k_1 \times k_1$ convolution mapping back to $c \times h \times w$. The original architecture applies a PixelShuffler layer from the ESPCN architecture [25] at the end of the network, but we omit this. In [18] and in our implementation, the architectural parameters are taken to be $(\ell, k_1, k_2, d) = (16, 9, 3, 64)$. Batch normalization layers are applied across the channel dimension using BatchNorm2d throughout, and padding="same", padding_mode="replicate" is used in our implementation. A schematic of SRResNet is presented in panel (c) of Figure 5.

4.3. Training

We partition our ERA5 data [11] into a training set from 1950 to 2000, a validation set from 2000-2010, and a test set from 2010-2020. We use the AdamW optimizer [21] with a learning rate of $1e-4$ on all models, and seek to minimize the mean-squared error (MSE) between network predictions $f_\theta(\mathbf{X}_L)$ and ground-truth high-resolution images \mathbf{X}_H . We monitor model performance—as measured by MSE on the validation set—throughout training, and maintain a checkpoint of the best-performing model throughout training. We allow each model to train for a maximum of 24 hours on a single node of the MSU HPC cluster,⁵ where each node

⁵<https://icer.msu.edu/>

Table 1. Comparison of model performance downscaling τ_{2m} and τ_p ERA5 data [11] over the continental US from 2010-2020 by a factor of $4\times$. We label metrics for which higher is better with (\uparrow), and metrics for which lower is better with (\downarrow). We indicate the first-, second-, and third-best models in each metric in gold, silver, and bronze, respectively.

Model	MSE (\downarrow)	PSNR (\uparrow)	SSIM (\uparrow)	Parameters
Nearest	0.01007	30.73644	0.97682	0
Bilinear	0.00668	30.53569	0.97911	0
Bicubic	0.00484	33.16329	0.98783	0
SRCNN	0.00369	36.11259	0.98953	14,114
VDSR	0.00188	38.38477	0.99522	685,506
SRResNet	0.00155	39.11917	0.99564	1,317,525

is equipped with a 12GB NVIDIA Tesla K80 GPU, and use the checkpointed model which achieved minimum validation loss for subsequent testing.

5. Results

Leveraging our test set from 2010-2020, we evaluate model performance over $120 \text{ months} \times 36 \text{ patches-per-month} = 4320$ patches of size 64×64 , though we discard patches with $>50\%$ NaN values as previously mentioned. We compare SRCNN, VDSR, and SRResNet against baselines of nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation on the basis of mean-squared error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) [30] averaged over the test set in Table 1. To further analyze model performance and compare against nearest, bilinear, and bicubic interpolation baselines, we visualize the predictions of each converged model on the task of $4\times$ downscaling in Figure 6.

In Table 1, we observe that all three deep SISR models quantitatively outperform baselines of nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation on all metrics. These quantitative improvements are supported by qualitative improvements in Figure 6, in which output from our deep SISR models is of higher perceptual quality than baselines, featuring sharper edges and disaggregated details. In particular, the perceptual quality of model outputs around coastlines and mountain ranges in which sharp gradients of τ_{2m} and τ_p exist is much improved by our deep SISR models. Indeed, the Cascade Mountains and Puget Sound of Washington are visible in the leftmost two columns of Figure 6; the coastlines of Lake Michigan and Lake Huron are visible in the third and fourth columns; the Wasatch Mountains of Utah are visible in the fifth and sixth columns; the Appalachian Mountains and Chesapeake Bay are visible in the seventh and eighth columns; and the Gulf Coast is visible in the rightmost two columns. Prediction magnitudes appear to be

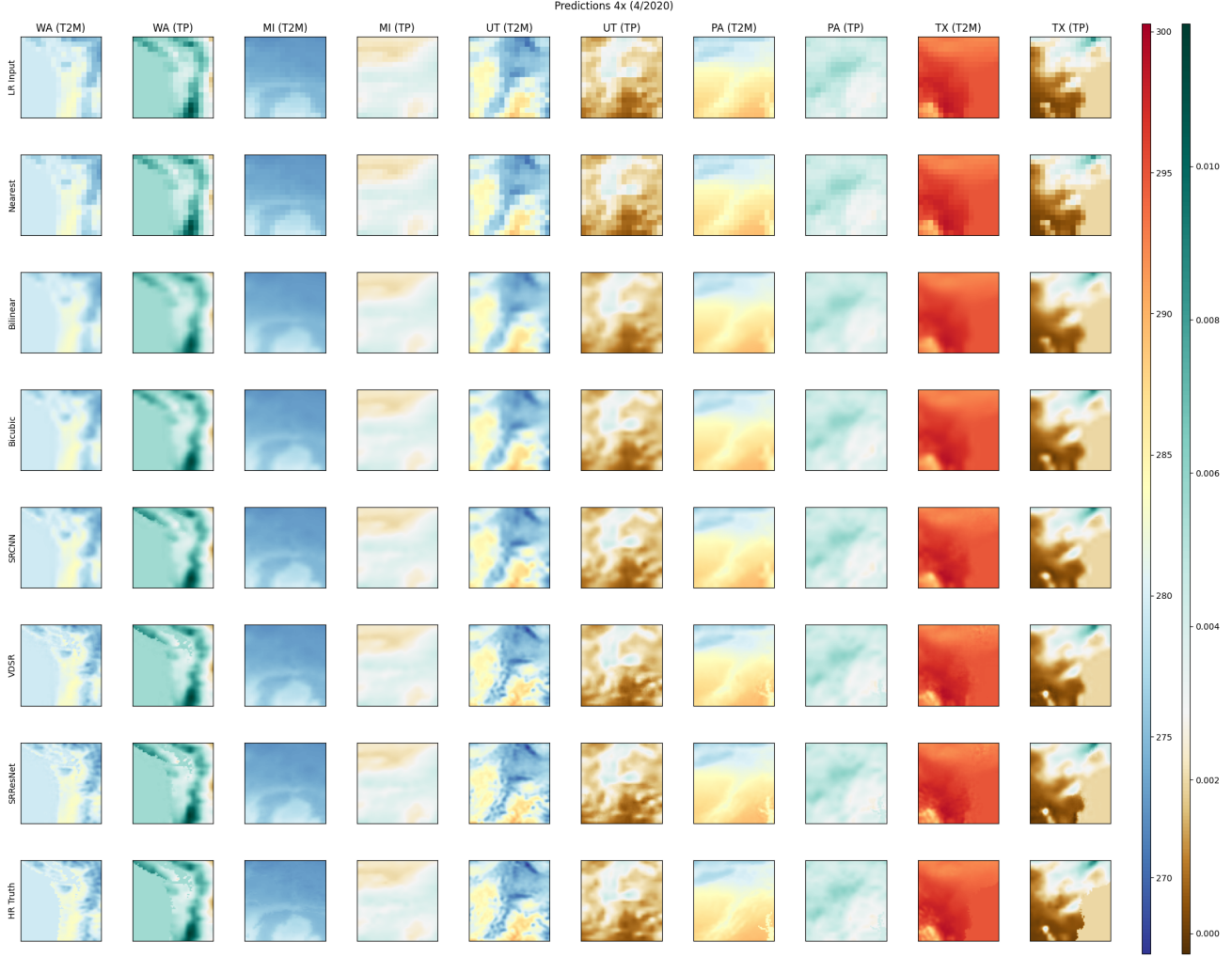


Figure 6. Model predictions of $4\times$ downsampled t_{2m} and t_p over selected patches in the continental United States in April 2020. Each column of subplots corresponds to a single geographic patch, labeled by US state, and a single variable. The first row of subplots corresponds to low-resolution inputs \mathbf{X}_L , the final row of subplots corresponds to high-resolution targets \mathbf{X}_H , and middle rows contain the outputs of each model. Forward passes are run with mean-subtracted, standard deviation-divided $\mathbf{Z}_L = (\mathbf{X}_L - \mu)/\sigma$ at inference-time just as during training. Outputs are linearly rescaled to the original data scale $\hat{f}_\theta(\mathbf{Z}_L) = f_\theta(\mathbf{Z}_L) \cdot \sigma + \mu$ for interpretable visualization, where t_{2m} is measured in degrees Kelvin and t_p is measured in meters.

well-calibrated as well, with temperatures ranging between 270-300K and precipitation ranging between 0 and 0.01m. Deeper networks with higher parameter counts (SRCNN \rightarrow VDSR \rightarrow SRResNet) generate progressively better visualizations both quantitatively and qualitatively, suggesting further improvements may be realized by training deeper models.

As a further exploration of model performance and generalizability, we visualize model predictions on the tasks of $8\times$ and $16\times$ downscaling in Figures 7 and 8, respectively. Despite the expected presence of convolution artifacts and aliasing in model outputs, we are surprised to observe such

heterogeneous texture apparent in model output. In particular, we find it interesting that SRResNet produces a locally-rough texture in mountainous regions (WA, UT, PA) and a locally-smooth texture in flat regions (MI, TX); this suggests its filters are able to infer common small-scale geographic patterns (e.g., the existence of individual mountain clusters visible only at high-resolution) from large-scale geographic structure (e.g., the existence of a mountain range visible at low-resolution).

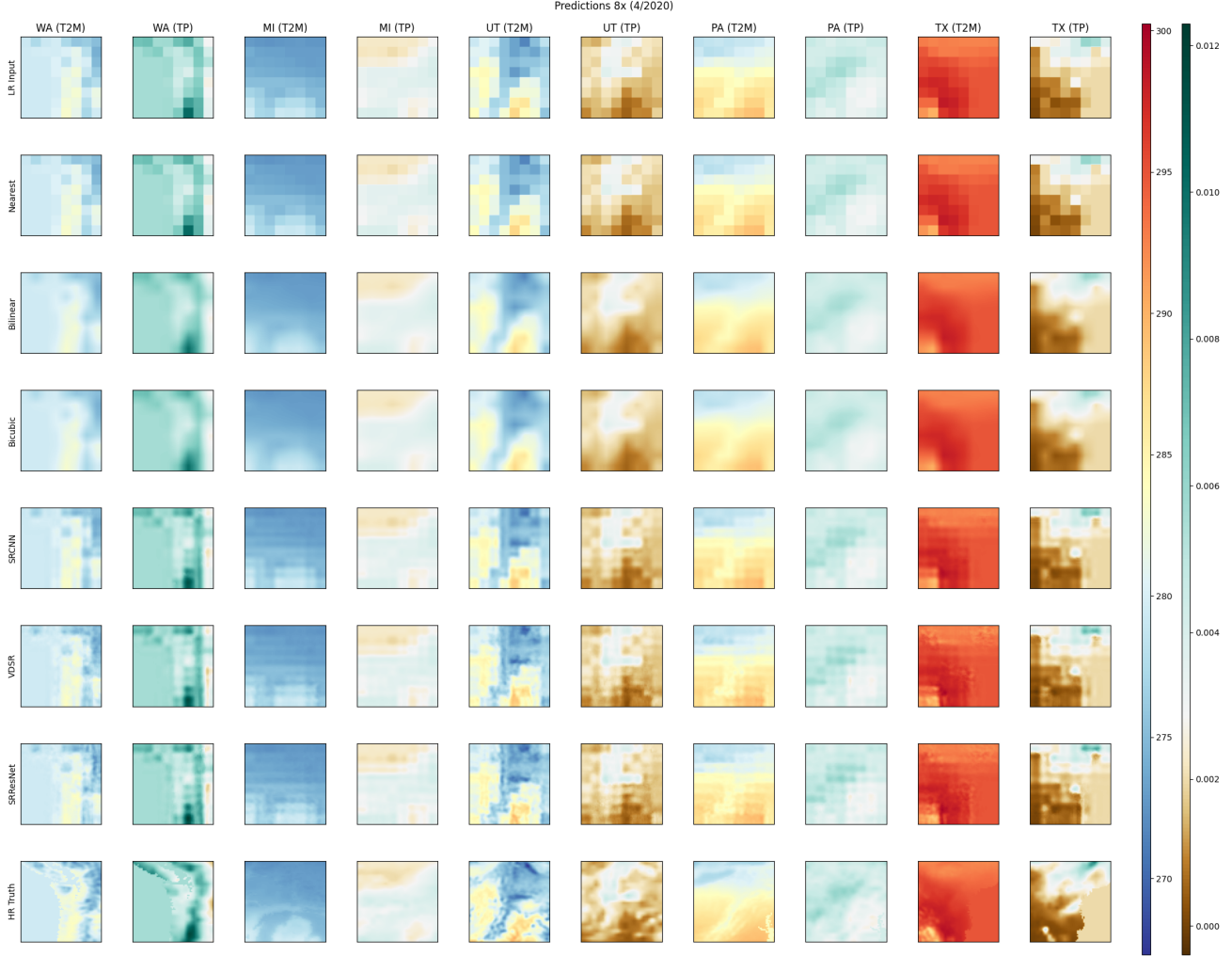


Figure 7. Model predictions of $8\times$ downsampled t_{2m} and t_p over selected patches in the continental United States in April 2020. All models are trained to perform $4\times$ downscaling; we present this visualization solely for the purpose of analyzing model generalizability.

6. Conclusion

This work explores the natural connection between the problem of single-image super-resolution (SISR) in computer vision and the problem of downscaling in climate science, and demonstrates the applicability of deep learning-based SISR models to multichannel climate data. We successfully train three state-of-the-art deep learning-based SISR models—namely, SRCNN [5], VDSR [14], and SRResNet [18]—to downscale monthly average surface temperature (t_{2m}) and monthly total precipitation (t_p) from the ERA5 Reanalysis Project [11] by a factor of $4\times$ over the continental US, beating baselines of nearest-neighbor, bilinear and bicubic interpolation by significant margins when comparing MSE, PSNR, and SSIM. Synthesized super-resolution images are perceptually cohesive and structurally

sound, reflecting the characteristic properties of target high-resolution images.

We encourage future work at the intersection of computer vision and climate science, and hope this work contributes toward a greater scientific and societal understanding of the consequences of climate change. In particular, we plan to extend this work to downscale projections from the CMIP6 Climate Model Intercomparison Project [6] to localize future climate scenarios from the regional to the city level, and encourage others to join us in this effort.

References

- [1] Tristan Ballard and Gopal Erinjippurath. Firesnet: Geoscience-driven super-resolution of future fire risk from climate change. In *Climate Change AI Workshop at NeurIPS*, 2020. 2

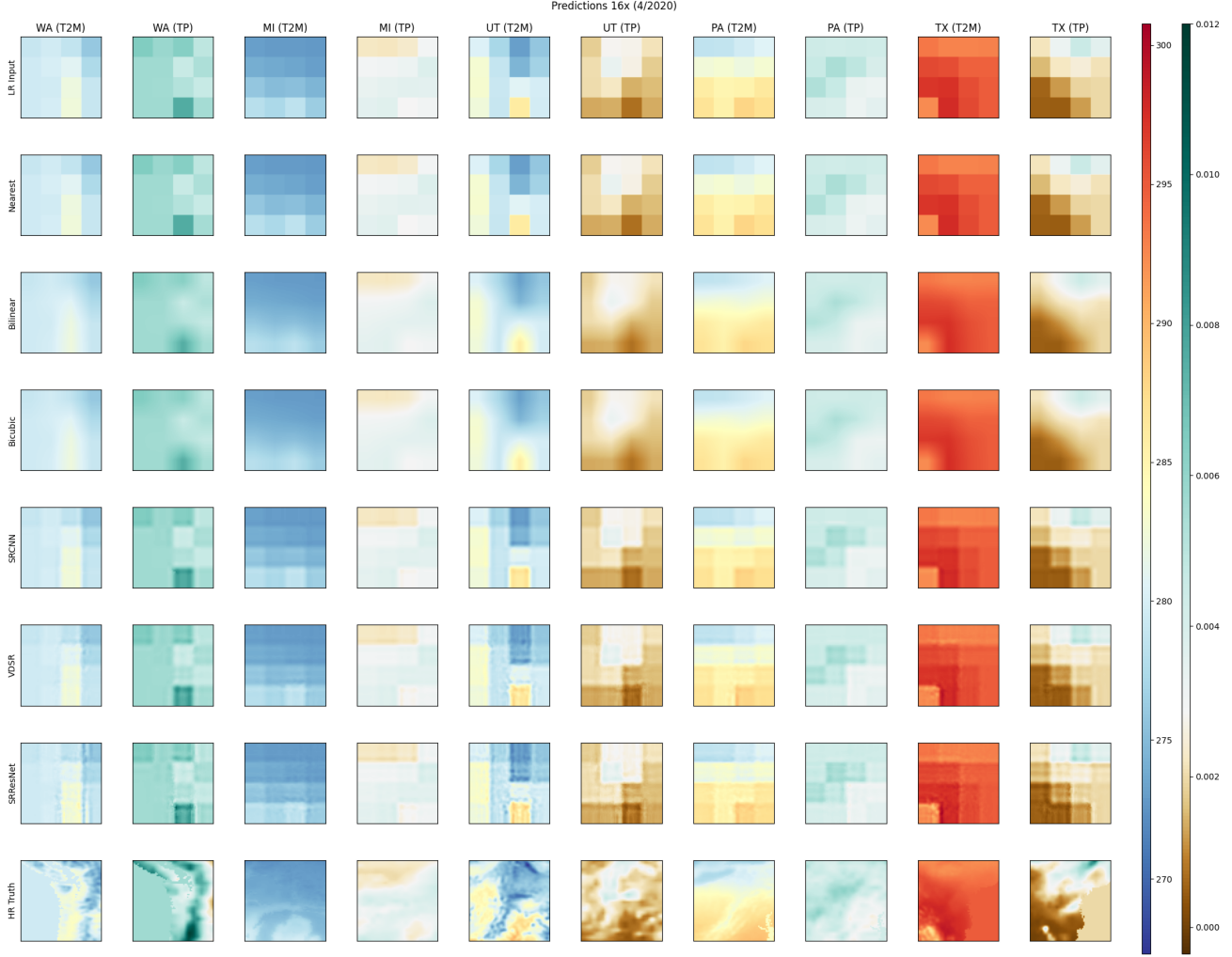


Figure 8. Model predictions of $16\times$ downscaled t_{2m} and t_p over selected patches in the continental United States in April 2020. All models are trained to perform $4\times$ downscaling; we present this visualization solely for the purpose of analyzing model generalizability.

- [2] Christopher Daly, GH Taylor, WP Gibson, TW Parzybok, GL Johnson, and PA Pasteris. High-quality spatial climate data sets for the united states and beyond. *Transactions of the American Society of Agricultural and Biological Engineers*, 2000. 2
- [3] Bekir Z Demiray, Muhammed Sit, and Ibrahim Demir. Dem super-resolution with efficientnetv2. In *Climate Change AI Workshop at NeurIPS*, 2021. 2
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 2
- [5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014. 1, 2, 4, 5, 7
- [6] Veronika Eyring, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 2016. 7
- [7] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT press, 2016. 1
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2014. 4
- [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015. 4
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

- [11] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. 1, 2, 5, 7
- [12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, 2015. 4
- [13] IPCC. Summary for Policymakers. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 2021. 1
- [14] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 1, 2, 4, 5, 7
- [15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012. 2
- [16] Rupa Kurinchi-Vendhan, Björn Lütjens, Ritwik Gupta, Lucien Werner, and Dava Newman. Wisosuper: Benchmarking super-resolution methods on wind and solar data. In *Climate Change AI Workshop at NeurIPS*, 2021. 2
- [17] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989. 2
- [18] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Others. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 1, 2, 4, 5, 7
- [19] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 2
- [20] Yumin Liu, Auroop R. Ganguly, and Jennifer Dy. Climate downscaling using YNet: A deep convolutional network with skip connections and fusion. 2020. 2
- [21] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. 2019. 5
- [22] Douglas Maraun and Martin Widmann. *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 2018. 1
- [23] David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, Alexandra Luccioni, Tegan Maharaj, Evan D. Sherwin, S. Karthik Mukkavilli, Konrad P. Kording, Carla Gomes, Andrew Y. Ng, Demis Hassabis, John C. Platt, Felix Creutzig, Jennifer Chayes, and Yoshua Bengio. Tackling climate change with machine learning. *arXiv:1906.05433*, 2019. 1
- [24] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 1986. 2
- [25] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 2, 5
- [26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*, 2014. 4
- [27] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010. 1
- [28] Thomas Vandal, Evan Kodra, Sangram Ganguly, Andrew Michaelis, Ramakrishna Nemani, and Auroop R. Ganguly. DeepSD: Generating high resolution climate change projections through single image super-resolution. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017. 2
- [29] Anna Vaughan, Nicholas D Lane, and Michael Herzog. Multivariate climate downscaling with latent neural processes. In *Climate Change AI Workshop at ICML*, 2021. 2
- [30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004. 2, 5
- [31] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2, 3
- [32] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 2019. 1, 2, 3