9th International Workshop on Climate Informatics
October 02-04, 2019
Hosted by École Normale Supérieure, Paris, France

DOWNSCALING NUMERICAL WEATHER MODELS WITH GANS

Alok Singh[1], Adrian Albert[1], Brian White[1]

***Abstract*—Numerical simulation of weather is resolution-constrained due to the high computational cost of integrating the coupled PDEs that govern atmospheric motion. For example, the most highly-resolved numerical weather prediction models are limited to approximately 3 km. However many weather and climate impacts occur over much finer scales, especially in urban areas and regions with high topographic complexity like mountains or coastal regions. Thus several statistical methods have been developed in the climate community to downscale numerical model output to finer resolutions. This is conceptually similar to image super-resolution (SR) [1] and in this work we report the results of applying SR methods to the downscaling problem. In particular we test the extent to which a SR method based on a Generative Adversarial Network (GAN) can recover a grid of wind speed from an artificially downsampled version, compared against a standard bicubic upsampling approach and another machine learning based approach, SR-CNN [1]. We use ESRGAN ([2]) to learn to downscale wind speeds by a factor of 4 from a coarse grid. We find that we can recover spatial details with higher fidelity than bicubic upsampling or SR-CNN. The bicubic and SR-CNN methods perform better than ESRGAN on coarse metrics such as MSE. However, the high frequency power spectrum is captured remarkably well by the ESRGAN, virtually identical to the real data, while bicubic and SR-CNN fidelity drops significantly at high frequency. This indicates that SR is considerably better at matching the higher-order statistics of the dataset, consistent with the observation that the generated images are of superior visual quality compared with SR-CNN.**

I. MOTIVATION

(*Note: We use ML terminology, where downsampling = upscaling and upsampling = downscaling.*) Global climate models are limited to ~100 km resolution, while numerical weather prediction models that produce daily forecasts and severe weather warnings are limited to ~3 km. However, accurate assessment of climate and extreme weather impacts near human populations would benefit substantially from finer resolution. While several methods have been developed for downscaling climate models output to finer resolutions, they consist for the most part of complex interpolation methods (see e.g. [3]). In this paper we explore a machine learning method to downscale weather model output using a Generative Adversarial Network (GAN) developed originally for the purpose of image super-resolution (ESRGAN).

Machine learning approaches have only recently started to receive attention in the earth sciences community [4]. Over traditional numeric-based approaches, they could address some key issues in climate modeling:

1) A ML pipeline trained end-to-end that automatically learns optimal filters and transformations between inputs (i.e., remote-sensing, in-situ, and simulation data) and their relationship to the spatiotemporal estimate of parameters of interest (e.g., wind intensity, precipitation), can drastically accelerate the creation of practical ad-hoc relationships between observational datasets and models.
2) An end-to-end differentiable model will allow for the exploration of climate model sensitivities that lead to bias including the influence of the meteorological forcing dataset. Slight perturbations in precipitation phase, intensity, and/or location, shortwave and longwave radiation, wind speed and direction, humidity, and, temperature, could be used to understand the downstream implications on variables that are difficult to measure or model directly.
3) Generative models can be run in parallel, and do not necessarily require iterative schemes to model data, allowing them to run quickly even on low-grade consumer hardware.

II. METHODS AND DATA

A. Dataset

We use 15 years of wind velocity fields from a numerical simulation of the WRF (Weather Research and Forecasting) model over Southern California (see

[1]Terrafuse, Inc

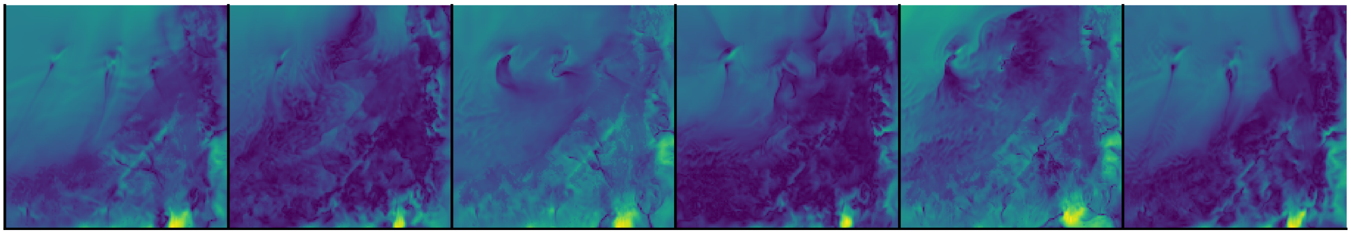
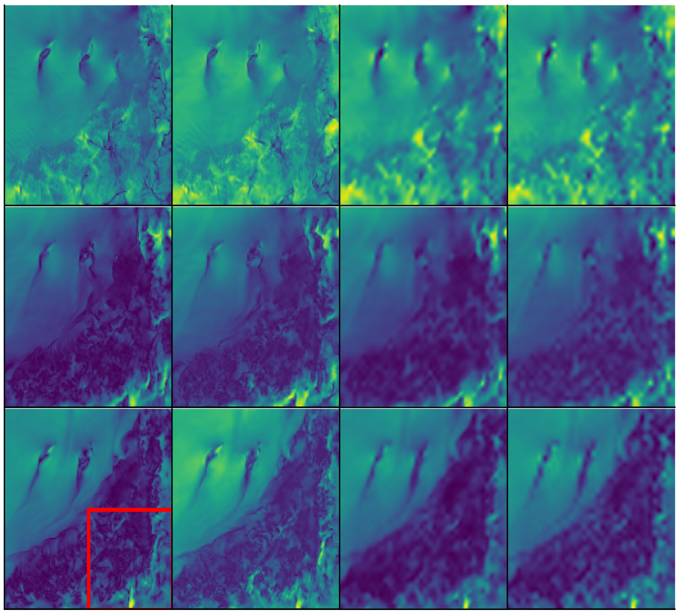
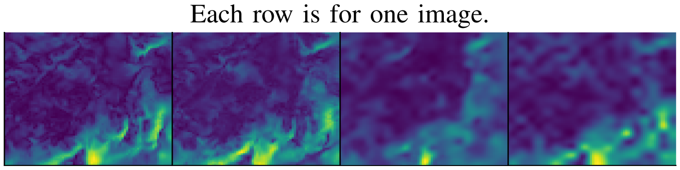


Fig. 1: Example wind speeds from the WRF model output, shown every 4 hours over 1 day.



(a) Real vs. ESRGAN vs. SR-CNN vs. Bicubic Upsampling. Each row is for one image.



(b) Zoomed view of marked region above.

[5] for details on model setup - we use region d04 from 2001-2015), gathered hourly, for a total of about 60,000 data points, represented as grids. In the rest of this work, we shall call these grids "images".

Each image is of size 153×153, where each pixel represents the average wind speed over a 1.5 km $\times$ 1.5 km region. They are stored as a 2-D array of 32-bit floats, linearly scaled to $[0, 1]$ to be compatible with image processing frameworks. For ease of upsampling, we clip a single row and column to resize our images to 152×152. We also combine the south to north and east to west components of the wind vector to model total wind speed rather than modeling each direction's velocity separately.

The dataset is then shuffled and split, with 5% (3,000 images) held out for validation.

B. Method

Note that all experiments were run on a stock HP Z420 with an NVIDIA GeForce RTX 2070 GPU.

1) Bicubic Upsampling: Bicubic interpolation is a common algorithm used for upsampling, useful as a baseline to compare against. There are no trainable parameters, so we simply upsample our validation set.

2) SR-CNN: SR-CNN is a popular deep-learning based approach to image SR. A low-resolution image is first upsampled to the desired size by another method, such as bicubic interpolation. It's then passed through a CNN, which outputs an image that is compared against the ground truth image via mean-squared error (MSE). (See 3 for a high-level pictorial overview.)

We train for 100 epochs with a batch size of 128, the Adam optimizer, and a learning rate of 0.001. For the upsampling step, we use bicubic upsampling. Validation is performed at the end of training.

The model has about 8,000 trainable parameters, a relatively small number by modern standards, so training is quick. It can process about 730 images per second, so training takes about 2 hours with an RTX Titan 2070.

3) ESRGAN: ESRGAN is an optimized version of SRGAN, which we shall describe here. SRGAN is a conditional GAN designed for image SR. Its training procedure involves passing the generator G a batch of low-resolution images, which are upsampled by G and then passed to the discriminator D. D is also given the ground truth images for the batch, and attempts to distinguish between them. An optimization particular to SRGAN is that it also has a "content loss", where in addition to the discriminator, there is another network, typically a CNN pretrained on ImageNet. This "feature network" passes both the generated and ground truth images through it, and then both are compared via MSE. The idea is that a pretrained network will have captured the higher-level dynamics of perceptual similarity. However, since our images do not represent natural images and we do not have an equivalent of ImageNet

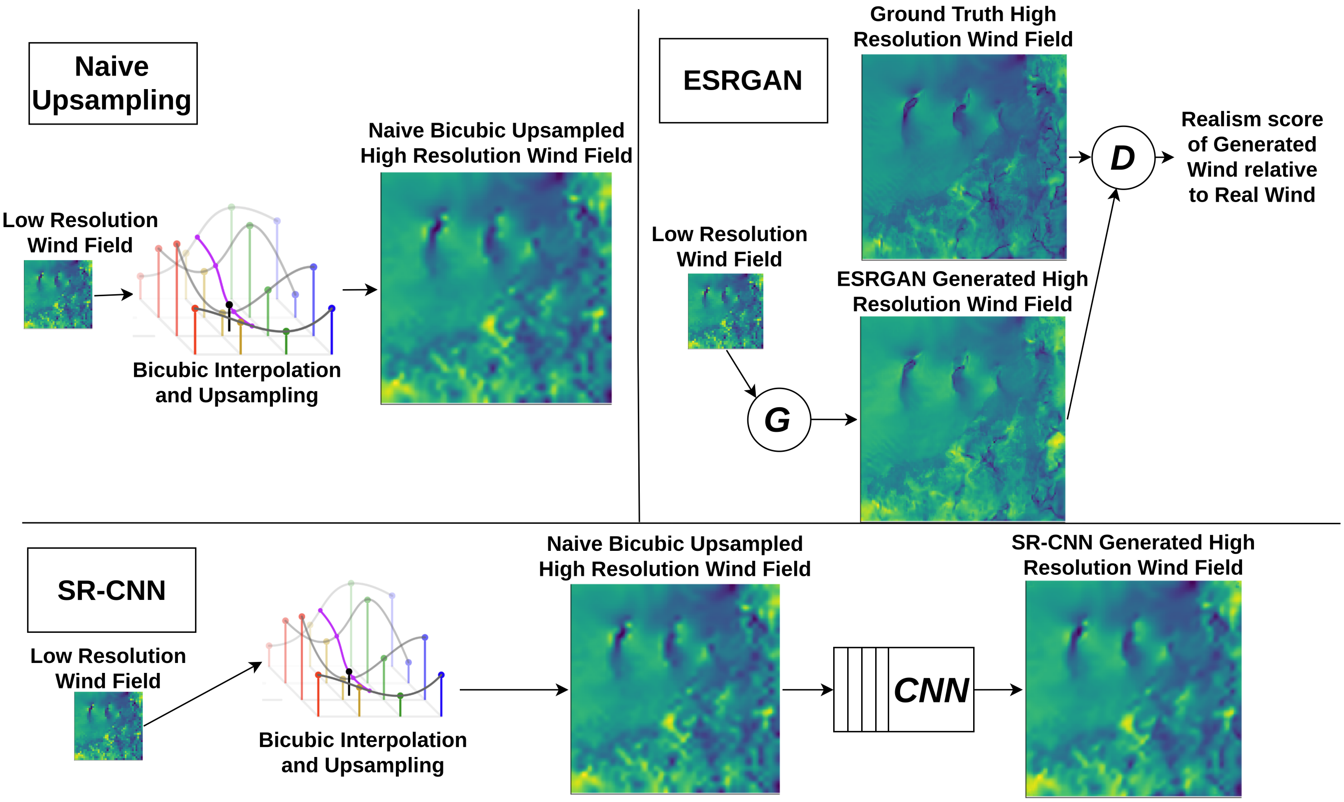


Fig. 3: Overview of models used.

TABLE I: Results

Model	PSNR	MSE	MAE	KL
ESRGAN	32.74	0.00053	0.0148	0.008
SR-CNN	**36.06**	**0.00024**	**0.0091**	0.015
Upsampling	35.52	0.00027	0.0097	**0.006**

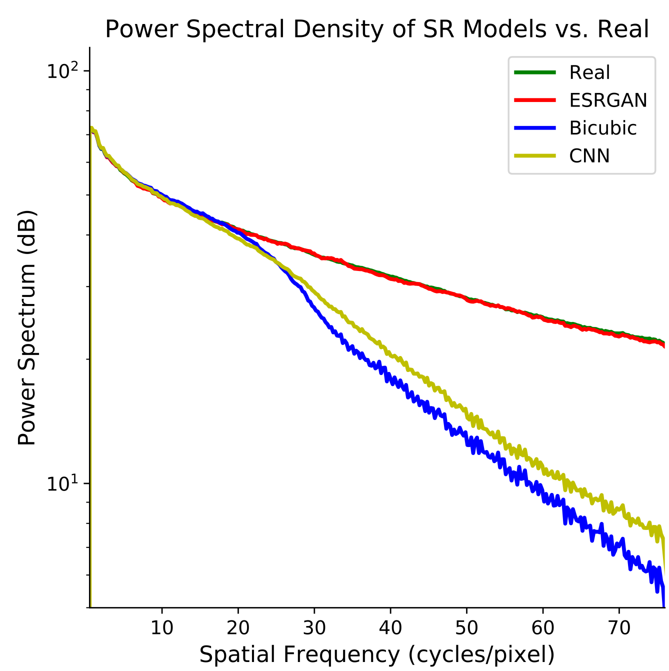
to perform supervised learning on, it is inappropriate to use content loss, so we remove it. Otherwise, our training is virtually identical to the methodology outlined in [2].

We use a batch size of 12 images and train for 100 epochs. The model has about 25 million trainable parameters, so training is far slower than SR-CNN. Each epoch takes about 35 minutes to complete, and training takes about 2.5 days.

III. EVALUATION

Table I gives an overview of final performance on the validation set. PSNR (peak signal to noise ration), MSE and MAE (mean absolute error) are averaged over all images in the validation set. "KL" represents the KL divergence between the empirical distributions of the generated images and the ground truth.

Fig. 4: Note that ESRGAN tracks the true data far more closely than the other models.



Notice that SR-CNN performs best on most metrics, and ESRGAN the worst.

But the generated images tell a quite different story. As we see in Figure 2a, ESRGAN generates clearer images than the other methods. Zooming in on the highlighted red box (Figure 2b) reveals that the image generated by ESRGAN are sharper and less prone to artifacts. PSNR and MSE have been noted in other works to be poor indicators of image quality, as they fail to capture the underlying dynamics of images well.

A key metric that illustrates the spatial resolution and higher moments of the data distribution is the power spectral density, shown in Figure 4. The power spectrum reveals the power of ESRGAN in capturing the high frequency information present in the wind field. In fact, remarkably ESRGAN's spectrum is so close to that of the true data that they are nearly indistinguishable, whereas SR-CNN and bicubic upsampling fall off significantly at higher frequencies. This is perhaps not surprising as the upsampling and SR-CNN are fundamentally methods of interpolation, whereas ESRGAN is learning the data distribution at all scales. These results suggest that ESRGAN is able to capture far more of the underlying spatial structure, and that while SR-CNN may be doing better than bicubic upsampling, it is not learning the distribution of the true data, but rather that of the upsampled version.

IV. FUTURE WORK

As seen in this work, ESRGAN does a good job reproducing single images. However, we do not currently deal with a sequence of images over time or space. For example, capturing the effects of winds over a larger surrounding region, e.g. from a coarse climate model, would help in regional climate prediction. In addition, being able to capture a sequential time series would also be. Both will be the goal of future work.

We also plan to incorporate additional variables such as temperature and pressure, and to see if models based on attention mechanisms can improve accuracy..

REFERENCES

[1] T. Vandal, E. Kodra, S. Ganguly, A. R. Michaelis, R. R. Nemani, and A. R. Ganguly, "Deepsd: Generating high resolution climate change projections through single image super-resolution," *CoRR*, vol. abs/1703.03126, 2017.

[2] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," September 2018.

[3] G. Bürger, S. Sobie, A. Cannon, A. Werner, and T. Murdock, "Downscaling extremes - an intercomparison of multiple methods for future climate," *Journal of Climate*, vol. 26, 2013.

[4] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[5] P. Vahmani and A. D. Jones, "Water conservation benefits of urban heat mitigation," *Nature Communications*, vol. 8, no. 1, p. 1072, 2017.