

DeepDownscale: a deep learning strategy for high-resolution weather forecast

Eduardo R. Rodrigues, Igor Oliveira, Renato L. F. Cunha, Marco A. S. Netto
IBM Research
{edrodri, igoro, renatoc, mstelmar}@br.ibm.com

Abstract—Running high-resolution physical models is computationally expensive and essential for many disciplines. Agriculture, transportation, and energy are sectors that depend on high-resolution weather models, which typically consume many hours of large High Performance Computing (HPC) systems to deliver timely results. Many users cannot afford to run the desired resolution and are forced to use low resolution output. One simple solution is to interpolate results for visualization. It is also possible to combine an ensemble of low resolution models to obtain a better prediction. However, these approaches fail to capture the redundant information and patterns in the low-resolution input that could help improve the quality of prediction. In this paper, we propose and evaluate a strategy based on a deep neural network to learn a high-resolution representation from low-resolution predictions using weather forecast as a practical use case. We take a supervised learning approach, since obtaining labeled data can be done automatically. Our results show significant improvement when compared with standard practices and the strategy is still lightweight enough to run on modest computer systems.

I. INTRODUCTION

Numerical simulation has become the third pillar of science, in addition to theory and experimentation. Models have become more complex over time and are always in need of the most up-to-date computing infrastructure. Particularly, resolution is one key factor that affects computing demand to execute such models [1]. For example, doubling the resolution of a two-dimensional grid leads to an increase of four times in number of points to be processed and an increase of the time resolution to meet the convergence condition (CFL) [2].

In weather and climate simulations, one typically runs a large scale but coarse resolution model and then, with the output of this model as input, runs a finer resolution regional numerical model. That is called *dynamical downscaling* [3]. This procedure allows for finer resolution in an area of interest without the computing cost and time of running the large scale model with the desired resolution. However, the regional numerical model is still a complex and computationally demanding software to run.

Much research has been done in using Convolutional Neural Networks (CNN) to reconstruct high-resolution data from low-resolution inputs. CNNs have been applied extensively in imaging research. The key assumption is that it is possible to recover a high-resolution image from a low-resolution input due to high redundancy in the data. Here we investigate if the same assumption holds for data generated from the execution of physical models. This approach has the signif-

icant advantage of allowing one to use a supervised learning approach, without the cost of manually labeling the data: one can run the physical model with both resolutions (high and low-resolution) as much as their resources allow to produce training data. Alternatively, one can use observational data whenever available.

In this work, we propose a Convolutional Neural Network that takes as input low-resolution weather data and interpolates it into a high-resolution output. The features are the output of a set of global models for which we crop a region of interest. The labels are observations in high-resolution. We compare with a linear interpolation of the same set of input models, which is a standard practice in meteorological science [4]. In addition, we compare with a regression method and a high-resolution regional model. We used weather forecast as a case study to apply our CNN-based strategy to generate high-resolution data as it is a typical area that embraces new models and strategies to address its computational needs [5].

The paper is divided as follows: in the next section, we review the state-of-the-art, which is in both image processing and weather downscaling. In Section III, we present our contributions. In Section IV, we present our neural network architecture and reasons for our choices. The following section describes the experiments we performed and Section VI has the results of these experiments. Our final remarks are found in Section VII.

II. RELATED WORK

The procedure of recovering a high-resolution image or video from a low-resolution counterpart is known as super-resolution (SR) [6]. Much research has been devoted to this topic in the recent years [7], [8], [9]. Also, it has many applications in diverse areas, such as satellite data [10], medical imaging [11], and video processing [12].

A typical assumption of super-resolution techniques is that N images with different perspectives are used to recover the high-resolution input. A common model is linear and takes the form

$$y_k = DB_k M_k x + n_k, \quad (1)$$

where y_k is the k th image generated from the high-resolution input x , M_k translates and rotates the input, B_k is a blur matrix, D is a subsampling matrix, and n_k is an additive noise. In this model, not only is x unknown, but M_k , B_k and D must be estimated as well. Non-linear models have also been explored, as in the work by He *et al.* [13].

In order to estimate x from y_k , there are several approaches, some based on learning from data and some direct methods. In the direct methods, some examples are: (1) nonuniform interpolation [14], [15], frequency domain methods [16], and regularized approaches [17].

Super-resolution techniques based on learning are very appealing, because one can generate training examples easily. An (approximate) inversion procedure is estimated by means of training tuples (x, y_1, \dots, y_N) . Translating, rotating, blurring, and subsampling images provide examples (y_k) for the untouched images (labels).

Deep learning models for super resolution are based on ideas similar to auto-encoders [18]. Auto-encoders are feed-forward neural networks whose purpose is to reconstruct their inputs. They have two parts, the encoder and decoder. Usually, auto-encoders have a bottleneck between the encoder and the decoder, and the weights of the bottleneck are the representation the user seeks. This representation is intended to be an efficient code for the input. In the super-resolution setting, the encoding part is equivalent to the process that generates the low-resolution image, the code represents the low-resolution image, and the decoder is the procedure the user is seeking to train.

Shi *et al.* [8] showed a CNN capable of real-time 1080p super-resolution on a single GPU. They developed an architecture that extracts features in the low-resolution space. Moreover, they introduced a sub-pixel convolution layer that learns an array of upscaling filters to upscale the final low-resolution feature maps into the high-resolution image. They assumed, however, a known noise function, that may not be the case in the physical setting.

In weather and climate, the efforts for dynamical downscaling data normally make use of higher resolution regional models nested within coarser global models. Boundary conditions are provided by the global models while the physics in the regional models are tasked with the simulation of the sub-grid processes that are not represented in the coarser model. Yang *et al.* [19] performed simulations with the LMDZ4 model at a $0.6^\circ \times 0.6^\circ$ resolution nested in 3 models with resolutions varying from $1.125^\circ \times 1.125^\circ$ to $2.8125^\circ \times 2.8125^\circ$, successfully reproducing precipitation and temperature patterns over the domains analyzed. Even the model with the lowest resolution showed significant improvement after dynamical downscaling.

Heikkilä *et al.* [20] performed dynamical downscaling of the ERA-40 reanalysis dataset using the regional model WRF and two nested domains: a 30km-resolution domain over North Atlantic and a 10km-resolution domain over Norway. Spectral nudging was applied to guarantee the regional model circulation would be consistent with the forcing reanalysis data. The finer resolution domains were able to reproduce the extreme precipitation events in the North Atlantic and to reduce regional temperature biases over Norway.

Weather forecast [21], [22], [23] is not the only area that demands optimization to reach high resolution results. Biology [24], [25] and medicine [26] are examples of areas

that require simulations of physical models in high resolution and could benefit from the findings provided in this paper.

III. CONTRIBUTIONS

The contribution of this paper is twofold, (1) a learning-based strategy to fuse physical simulations to produce a high-resolution output applied to weather forecast, and (2) a comparative analysis of the limits of super-resolution approaches to downscale.

In our current approach, we use a feed-forward neural network to associate input patterns to high-resolution output. As we will show with the results of our experiments, this strategy can indeed capture the redundancies in the input and produce a forecast that is significantly better than standard procedures. Consequently, these results show that even without memory, the network can still find relations that cannot be attributed to known noise as in the image processing context.

In addition, our proposed initialization (discussed in Section IV) allows the use of deeper neural networks that better approximate the regression function we are seeking. This works similar to the skip connections of the Resnet architecture [27].

Moreover, we show one usage for locally connected layers that allows for local specific pattern matching. This is done in a context that a fully connected would not be possible, because of the memory requirements of these layers and the output size of our network.

Finally, even though we have tested the strategy only with a weather application, we expect that it generalizes to other applications. For example, Finite Difference Methods that are run routinely over the same domain can have the regularities that the Neural Network architecture would capture. These regularities would allow one to run low-resolution in the region of interest and obtain a higher resolution, even if a higher resolution simulation would be run posteriorly.

IV. ARCHITECTURE

In this section, we discuss our neural network architecture and the reasons for the choices we made. In order to evaluate the impact of each of these decisions, we ran controlled experiments and present results in the following sections.

The basic architecture we developed is presented in Figure 1. The input is a set of n low-resolution forecasts from n different weather models. Each input is interpolated so that they have the same horizontal dimensions. In our experiments, the horizontal dimension is the upsampled version of the model. One could use a periodic reshuffling [8] and keep all the inputs in the low-resolution dimensions and save memory. However, in our experiments we used the upscaled dimensions. All input from the different weather models are then stacked together into a single volume of depth n .

The input volume is fed into a series of convolutions which are initialized with an approximate delta function up to the last but one layer:

$$f^l(x; W_l, b_l) = \phi(W_l * x + b_l) = \phi(\hat{\delta} * x) \simeq x$$

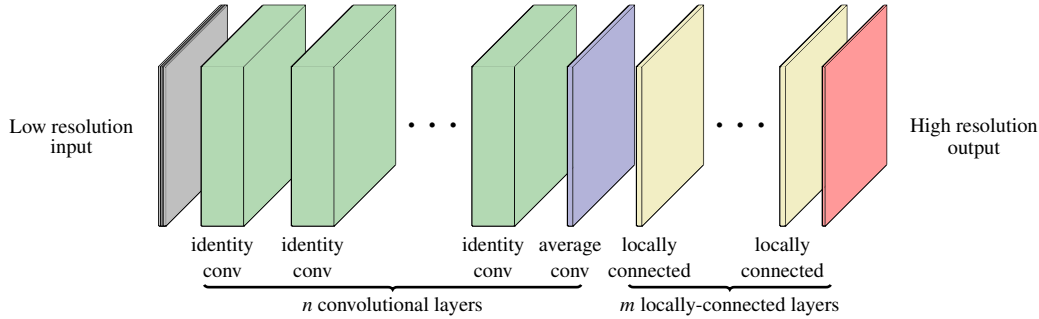


Fig. 1. Neural Network Architecture. The architecture is configurable and supports $n - 1$ identity convolutions, each with eight output filters, represented in green in the Figure. The last convolution is initialized as an average and generates a single-channel output. After the convolutions, m locally-connected layers follow, shown in yellow. The low resolution inputs are shown in gray and the high-resolution output is shown in red.

the approximate delta function $\hat{\delta}$ is equal to one at $t = 0$, but rather than being zero elsewhere, it is initialized to very small random numbers in order to break symmetry. This needs to be done, because the pure delta function has still many symmetries that would prevent backpropagation to compute distinct updates. The activation function (ϕ) is the rectified linear function (Relu).

This approximate delta layer can only produce the same activation depth as the input. In order to allow deeper activations, we used a stack of $\hat{\delta}$:

$$f^l(x; W_l, b_l) = \phi(W_l * x + b_l) = \phi([\hat{\delta}_1, \dots, \hat{\delta}_m] * x) \simeq [x_1, \dots, x_m]$$

where m is the number of copies of the input. This means that the activation depths can only be multiples of n . This restriction can be relaxed by making partial copies of the inputs or adding extra intermediate outputs that are zeroed by initial parameters down the network. However, in our experiments we kept the number of activations just as multiples of n .

In the last layer, we initialize the weights with uniform distribution in the depth direction (so this layer takes an average along the depth). Effectively, this initialization takes an average of the input models as the final result. For this particular model, this is similar to what the Resnet architecture [27] would do with small weights, i.e. it will let the input to cross the network unchanged. Similarly to the Resnet architecture, the optimization procedure adjusts the weights to improve the response if it is possible to do so. This strategy allows for very deep neural networks without suffering from premature overfitting.

We used a cross-validation approach in order to select the number of convolutional layers. The best network was very deep, 46 convolutional layers and 2 locally connected layers. This fact appears to indicate that, as in other visual recognition tasks, the number of layers substantially benefit from very deep networks. One possible explanation is that the later convolutions need to have a larger view of the domain. That is, the constant receptive field can only reach to a limited extension, the additional layers allows that extension to expand as illustrated in Figure 2.

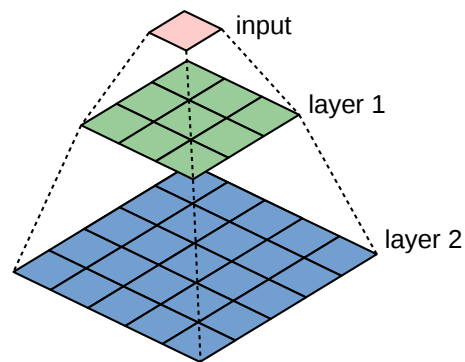


Fig. 2. View of the input after one and two 3x3 convolutions

The extension viewed by a single activation can be increased by larger filters. One could, therefore, increase the filter size and have the same extension and a shorter architecture. However, only extending the view may not keep the same performance, since this strategy reduces the number of non-linearities and consequently the expressiveness of the network. We investigated the relation between view extension and network depth and concluded that simpler and more numerous layers were better for our application.

The last locally connected layers are intended to capture patterns that are locality-specific. This is not typically needed for neural networks in general and, in fact, is usually avoided. That is because for classification and regression one seeks filters that activate when they find visual features irrespective to their location. For example, a picture can be classified as a certain object even when that object has never appeared in the given position in the training set. In our case, however, location is fixed and represents specific positions in the domain. Consequently, we want to learn the particular influence of location in the forecast. For example, topography plays a role in precipitation and can help refine the output. One could use a fully connected layer to achieve the same result. However, this strategy would require a very large memory footprint. For instance, the weights of the fully connected layer would require $O(N^4)$, where N is the horizontal dimension of the domain.

Lastly, the cost function used is simply the quadratic loss:

$$L = \frac{1}{I * J * K} \sum_{i,j,k}^{I,J,K} \|y_{i,j,k} - \hat{y}_{i,j,k}\|_2^2,$$

where I is the number of examples, J and K are the two horizontal dimensions, and y and \hat{y} are label and prediction respectively. In addition to the quadratic loss, we applied L2 regularization to the weights:

$$L += \lambda \sum_l W_l^2,$$

where λ is sought through cross-validation and W_l is each one of the layer weights. This neural network performs multiple regression, one for each forecast point.

V. EXPERIMENTS

In order to evaluate the benefits of the proposed neural network, we made predictions and compared them to standard procedures of downscale. Specifically, we used precipitation data. We compare our results with an ensemble mean, a high-resolution local weather model and a linear regression ensemble model. In this section, we describe in details the training procedure, the methods with which we compare the neural network and, in the next section, we show the results.

The metric used to evaluate each downscale approach was the root of the cost function, i.e. the Root Mean Squared Error (RMSE). This is appropriate since it measures the error in the units used in the forecast, and it allows us to evaluate how much error on average each method has for each grid point.

We used as input data from the Coupled Model Intercomparison Project Phase 5 (CMIP5), a framework for comparison of general circulation models and one of the main basis for the Intergovernmental Panel on Climate change (IPCC) Assessment Reports [28]. Specifically, we used the historical simulations, from January 1st 1981 to December 31st 2005 of the models *csm4*, *cgcm3*, *cm3* and *cm5a*. Our decision to use CMIP5 was its wide availability and its long continuous runs with the same model parameters. The particular models were chosen because of the range of resolutions (Table I).

The labels could have been a high-resolution model output. In that case, one could generate input data and labels from the same model. Specifically, low-resolution data would be used as input and high-resolution output as label. That procedure would permit an automatic supervised learning strategy, that does not require human labeling, but incurs the cost of running (possibly many times) both low and high-resolution models. Here, instead, we use observations as labels from the Chirps dataset [29], whose resolution is 0.25 degrees. Chirps has more than 30 years of precipitation data spanning 50°S-50°N all longitudes. This dataset incorporates both satellite and weather station data. The region of the input data and labels was chosen to be over South America (6°37'33"N, 74°33'52"W; 34°05'44"S, 34°43'01"W). The dimensions of the grid are (164,161) points.

TABLE I
RESOLUTION OF INPUT MODELS.

model	resolution (degrees)
CCSM3	1.25
CGCM3	1.25
CM3	2.50
CM5A	3.75

In order to train the model, we first divided the available data into three sets: (1) training, with forecasts and observations from January 1st 1981 to December 31st 2003, (2) validation, January 1st 2004 to December 31st 2004, and (3) test, January 1st 2005 to December 31st 2005. The validation set was used to choose the hyperparameters and the test set was used only to report the final performance of the model. Since the data is temporal, it is of utmost importance not to simply randomize and split this data, since this may lead to a biased test performance. However, after splitting the data into training, validation and test sets, the training data is randomized before applying the gradient decent method to optimize the parameters, as it is regularly done in training neural networks.

The first step in the training procedure was to tune the following set of hyperparameters: learning rate, regularization coefficients, number of convolutions, and number of locally connected layers. The learning rate was chosen randomly from a log-uniform distribution between 10^{-4} and 10^{-1} . The regularization coefficient was chosen from a wider distribution (log-uniform between 10^{-5} and 10^1), to encourage the network explore more regularization values. That is because we believe that there are too many parameters for the size of our training data.

We adopted the standard procedure of computing the mean and standard deviation of the training set, and, with those values, centering and normalizing training, validation and test set. We do not use dropout. The mini-batch size was set to 16. The total input has 8389 points, each of which with 4 low-resolution forecasts. The validation and the test set have 359 points each.

We used a single GPU GTX1080TI, which has 11GB of memory. Our intention was to have a system much cheaper than those required to run the downscale typically done in weather forecast centers. Moreover, most of the time was spent training the network, that is done only once (one may incrementally update the model however). The evaluation is even cheaper to execute.

The results of our model were compared with the ensemble mean of the input models. This is a simple method, but is typically better than control forecasts by most standard verification metrics [30]. That is because it smoothes out outlier predictions. This can be particularly good if the errors of the individual models are uncorrelated, in which case the errors are canceled. The method is an instance of what has been called “wisdom of crowds”, in which crowds (or

ensemble of predictors) can make better decisions compared to those of isolated individuals.

The ensemble mean has the advantage that it does not require any training. However, a consistently bad model can reduce the accuracy of the ensemble. A simple approach to reduce this issue is to use linear regression. We used both the training and validation sets as defined previously to train the model:

$$\hat{Y} = AX + b$$

where A and b are the weights and biases of the linear model. We train this model with the regular least squares method, since the amount of data is not large. This model will perform a linear combination of the ensemble X .

Finally, we compare our model with a high-resolution weather model from the CORDEX project [31]. CORDEX is a multi-institution effort to produce high-resolution climate simulations via dynamical downscaling. CORDEX proposes several regional domains across the globe, including a 0.44° resolution domain over South America that matches the domain used in this paper. The simulation used here is a downscaling experiment of the “historical” runs from CMIP5 [28] that runs from 1950-2005. Here, the daily precipitation for the 2001-2005 period was used to validate the downscaling method. CORDEX represents the typical computationally expensive method to produce high-resolution climate data, since its simulations may take up to 10.6 hours to run for domains of similar dimensions and resolution [32].

VI. RESULTS

In this section we present a comparison between the proposed neural network and each one of the three traditional alternatives for downscale, i.e. ensemble mean, multiple linear regression model, and a regional model. This comparison was done with a test set not seen by any model and subsequent to both the validation and training sets, that were used to train the neural network, but also to train the multiple linear regression model.

We tested seven days forecast for rain. This configuration is appropriate for a number of applications. For example, farmers could use the forecast to decide seeding and pesticide application. The first comparison is shown in Table II. This table shows the root mean square error (RMSE) of the neural network and the three alternatives compared with the observed rain.

TABLE II
ROOT MEAN SQUARE ERROR IN THE TEST SET.

Model	RMSE (mm)
Neural network	24
Ensemble mean	38
Linear regression	27
Regional model	33

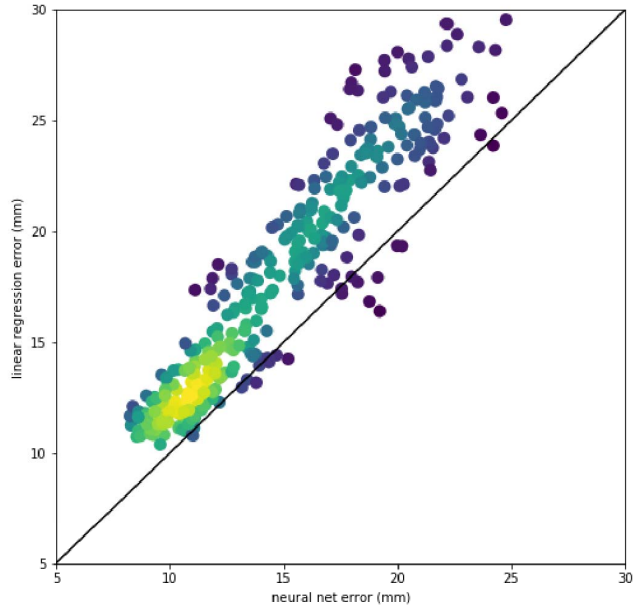


Fig. 3. Comparison between linear regression and Neural network error

On average, the neural network achieved the best performance. It is interesting to notice that the regional model was *not* the second best. There is a number of reasons to explain this fact. One possibility is the inherently biases these types of models may have. Since we used raw forecasts, this may be one reason for the poor performance. In addition, the neural network was trained with the observation itself, which has helped it to perform better.

In addition, the multiple linear regressor had a large intercept coefficient, as it can be seen in Table III. This suggests that the models under estimate the precipitation systematically. Consequently, a bias adjustment will account for most of the improvement compared to the ensemble mean. However, compared to the neural network, the multiple linear regression still systematically has a worse error, as it can be seen in Figure 3. In this figure, each forecast error of both the neural network and the linear regression is paired. The errors in the linear regression are almost always larger than the neural network errors.

TABLE III
MULTIPLE LINEAR REGRESSION COEFFICIENTS.

	Coefficient	Standard Error
CCSM3	0.21	2e-3
CGCM3	0.10	7e-5
CM3	0.30	8e-5
CM5A	0.12	8e-5
intercept	7.90	5e-5

In order to evaluate the significance of the improvement observed with the neural network, we computed the probability of making daily prediction with the test set assuming that the

network was better than the interpolation just by chance (H_0). That is, the null hypothesis is that the neural network is no better than the linear interpolation. The computed probability was less than 0.1%. This is illustrated in Figure 4.

However, what we really want is the probability of the neural network being better than the interpolation given the experiments, i.e. $P(H_1|O)$, and what we have is $P(O|H_0)$. There is also the assumption that the only hypotheses are H_0 and H_1 , but this is reasonable and means that we are comparing whether or not the best conventional strategy is better than the neural network. Still, the 99.9% change of being better is only true if the probability of the null hypothesis is the same as the probability of the observation (results of the experiments):

$$\begin{aligned} P(H_1|O) &= 1 - P(H_0|O) \\ &= 1 - \frac{P(O|H_0)P(H_0)}{P(O)} \end{aligned}$$

In order to compute $P(H_1|O)$ accurately, we would need to have (or estimate) the marginals $P(H_0)$ and $P(O)$. This is not an easy task, since we have no access to the distribution of observations and we have limited observation samples. In this setup, however, we can evaluate the limits to which $P(H_0)/P(O)$ can go so that we accept $P(H_1|O)$ as likely.

Consequently, from the results we obtained, the ratio between $P(H_0)$ and $P(O)$ can be as high as 50 and still the significance will be larger than 95%, i.e. the experiment results can be 50 times less likely than the null hypothesis and still the significance will be larger than 95%.

We also compared our neural network with a Resnet-like architecture. In this model, the same structure as described previously (see Figure 1) is enhanced with “shortcut connections”, i.e. every two layers, the input is copied verbatim to the

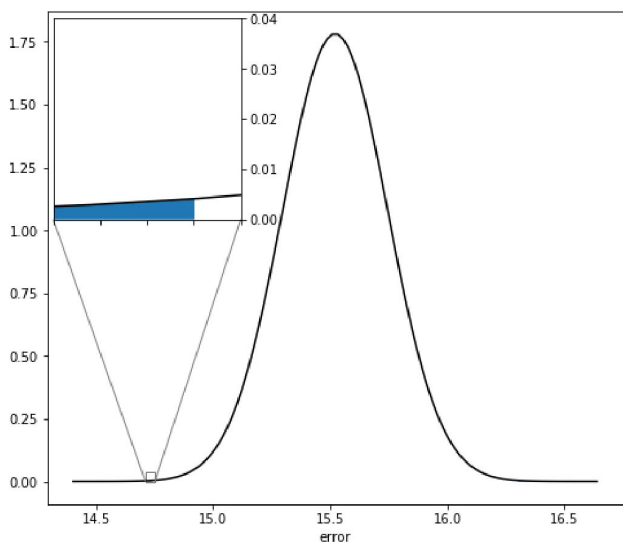


Fig. 4. Confidence analysis

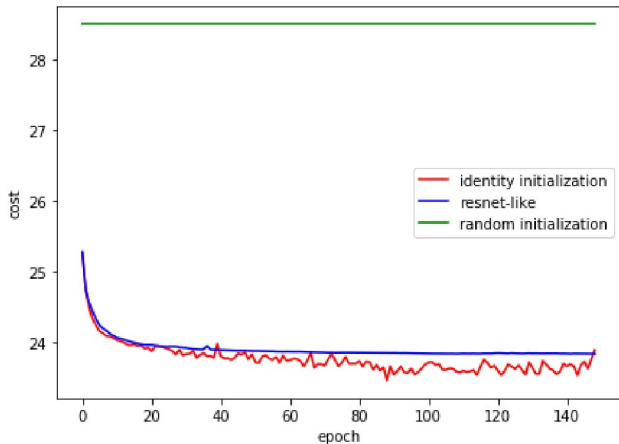


Fig. 5. training cost comparison

output. The weights are initialized with small random values, following the standard practice.

We obtained similar results to our architecture with the Resnet-like model. This suggests that the identity initialization places the weights close to the solution that the resnet will find. This is not what typically happens in other applications. For example, in image classification the gradient decent method could in principle find the identity function in deep neural networks. However, this is not what happens as reported by He *et al.* [33]. In Figure 5, we compare the cost during training for our architecture, resnet and our architecture with random initialization. As it can be seen, our architecture has a similar, although noisier, cost evolution.

Finally, we compared the prediction itself of both the best interpolated model and the neural network output. Figure 6 shows the observation, the best interpolated forecast, and the output of the neural network output. As it can be seen, the neural network result in this particular forecast has more details in the storm over the Amazon.

VII. CONCLUSION

Downscaling is an important procedure for weather and climate applications in which coarse resolution forecasts are refined to meet a desired resolution. Many users rely on downscaling results to make decisions in many disciplines. A typical form of this procedure is known as dynamical downscale in which a high-resolution regional model is run with low-resolution data from another model as input. However, running regional models in the required resolution is very costly.

On the other hand, much research has been done to improve resolution of images (and video) in computer science, in what is known as super-resolution. Typically, these techniques rely on the fact that much information is redundant and a high-resolution image can be recovered from the low-resolution input. However, most of the literature assume a known noise/error function. This is not the case of downscaling, in which the reverse mapping between the high-resolution to the

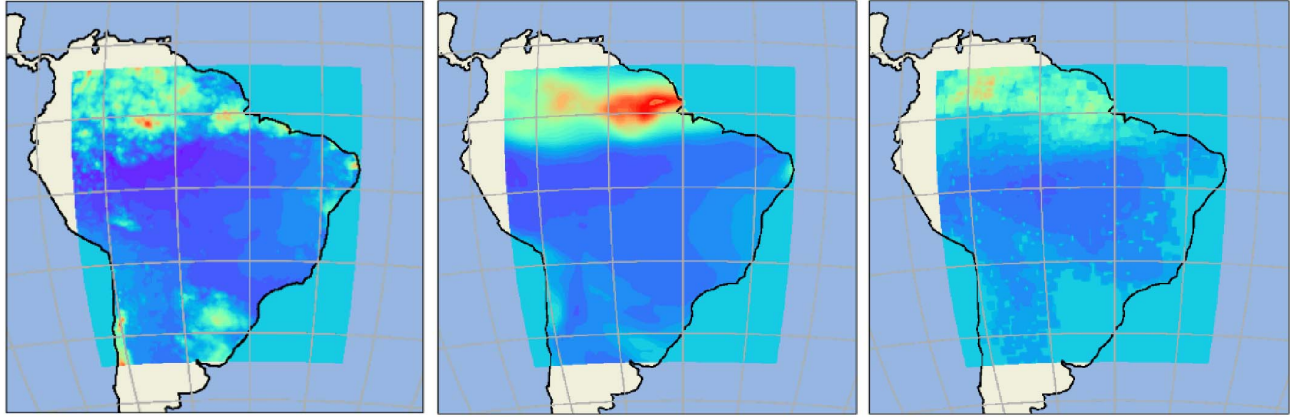


Fig. 6. Comparison between observed weather (left), best interpolated forecast (center), and network output (right) of a typical test case.

low-resolution forecast is completely unknown. Still, in this paper we investigate if the super-resolution technique could be used to perform downscale of weather models and compared with standard procedures.

Our proposed strategy is based on Deep Neural Networks. This approach has a major advantage that the super-resolution procedure is learned from data in a supervised learning fashion. Moreover, there is no need to manually labeling the data, since one can always run the model in both resolutions to generate training examples - even though there is a computational cost associated with this procedure - but also it is possible to use observations as labels. From our experiments, we observed significant improvement of the proposed strategy compared with standard downscale procedures. Moreover, the strategy is cheap enough to run in a single GPU system, and even training can be run on that system.

Finally, we expect our strategy to apply to other applications where multiple models are run over the same domain. A particular example is soil moisture, that can benefit high-precision agriculture. In this application, multiple soil moisture models are available to a given region. However, some high-resolution data are only available with a few days delay. One could use our strategy to approximate the high-resolution data from the low-resolution but updated data. Still, one would need to deal with the time delay that may require a model with memory.

REFERENCES

- [1] J. L. K. III and M. Wehner, "Computing issues for wcrp weather and climate modeling," 2005.
- [2] R. Courant, K. Friedrichs, and H. Lewy, "On the partial difference equations of mathematical physics," *IBM journal of Research and Development*, vol. 11, no. 2, pp. 215–234, 1967.
- [3] R. L. Wilby and T. Wigley, "Downscaling general circulation model output: a review of methods and limitations," *Progress in physical geography*, vol. 21, no. 4, pp. 530–548, 1997.
- [4] A. W. Wood, L. R. Leung, V. Sridhar, and D. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," *Climatic change*, vol. 62, no. 1-3, pp. 189–216, 2004.
- [5] B. N. Lawrence, M. Reznay, R. Budich, P. Bauer, J. Behrens, M. Carter, W. Deconinck, R. Ford, C. Maynard, S. Mullerworth *et al.*, "Crossing the chasm: how to develop weather and climate models for next generation computers?" *Geoscientific Model Development*, vol. 11, no. 5, p. 1799, 2018.
- [6] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE signal processing magazine*, vol. 20, no. 3, pp. 21–36, 2003.
- [7] Y. Romano, J. Isidoro, and P. Milanfar, "Rais: rapid and accurate image super resolution," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 110–125, 2017.
- [8] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. 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*, 2016, pp. 1874–1883.
- [9] Y. Bar-Sinai, M. Brenner, P. Getreuer, J. Hickey, S. Hoyer, and P. Milanfar, "Using image super-resolution techniques as a coarse-graining method for physical systems," *Bulletin of the American Physical Society*, 2018.
- [10] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-resolved surface reconstruction from multiple images," in *Maximum Entropy and Bayesian Methods*. Springer, 1996, pp. 293–308.
- [11] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O'Regan, and D. Rueckert, "Multi-input cardiac image super-resolution using convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 246–254.
- [12] Y. Altunbasak and A. J. Patti, "A maximum a posteriori estimator for high resolution video reconstruction from mpeg video," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 2. IEEE, 2000, pp. 649–652.
- [13] Y. He, K.-H. Yap, L. Chen, and L.-P. Chau, "A nonlinear least square technique for simultaneous image registration and super-resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.
- [14] S. Kim and N. Bose, "Reconstruction of 2-d bandlimited discrete signals from nonuniform samples," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 137, no. 3. IET, 1990, pp. 197–204.
- [15] H. M. Keshk, A. S. Ali, M. M. Abdel-Aziem, and M. Assal, "Performance evaluation of quality measurement for super-resolution satellite images," in *Science and Information Conference (SAI), 2014*. IEEE, 2014, pp. 364–371.
- [16] R. Tsai and T. Huang, "Multiple frame image restoration and registration," *Advances in Computer Vision and Image Processing*, 1984.
- [17] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 826–837, 2003.
- [18] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

- [19] H. Yang, Z. Jiang, and L. Li, "Biases and improvements in three dynamical downscaling climate simulations over china," *Climate dynamics*, vol. 47, no. 9-10, pp. 3235–3251, 2016.
- [20] U. Heikkilä, A. Sandvik, and A. Sorteberg, "Dynamical downscaling of era-40 in complex terrain using the wrf regional climate model," *Climate dynamics*, vol. 37, no. 7-8, pp. 1551–1564, 2011.
- [21] Ø. Hov, D. Terblanche, S. Jones, P. M. Ruti, and O. Tarasova, "Five priorities for weather and climate research," *Nature*, vol. 552, no. 7684, pp. 168–170, 2017.
- [22] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, p. 47, 2015.
- [23] M. Mizielinski, M. Roberts, P. Vidale, R. Schiemann, M.-E. Demory, J. Strachan, T. Edwards, A. Stephens, B. Lawrence, M. Pritchard *et al.*, "High-resolution global climate modelling: the upscale project, a large-simulation campaign," *Geoscientific Model Development*, vol. 7, no. 4, pp. 1629–1640, 2014.
- [24] D. Groen, J. Borgdorff, C. Bona-Casas, J. Hetherington, R. W. Nash, S. J. Zasada, I. Saverchenko, M. Mamonski, K. Kurowski, M. O. Bernabeu *et al.*, "Flexible composition and execution of high performance, high fidelity multiscale biomedical simulations," *Interface Focus*, vol. 3, no. 2, p. 20120087, 2013.
- [25] J. Hetherington, I. D. L. Bogle, P. Saffrey, O. Margoninski, L. Li, M. V. Rey, S. Yamaji, S. Baigent, J. Ashmore, K. Page *et al.*, "Addressing the challenges of multiscale model management in systems biology," *Computers & chemical engineering*, vol. 31, no. 8, pp. 962–979, 2007.
- [26] A. Neubert, J. Fripp, C. Engstrom, R. Schwarz, L. Lauer, O. Salvado, and S. Crozier, "Automated detection, 3d segmentation and analysis of high resolution spine mr images using statistical shape models," *Physics in Medicine & Biology*, vol. 57, no. 24, p. 8357, 2012.
- [27] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.
- [28] K. E. Taylor, R. J. Stouffer, and G. A. Meehl, "An overview of cmip5 and the experiment design," *Bulletin of the American Meteorological Society*, vol. 93, no. 4, pp. 485–498, 2012.
- [29] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell *et al.*, "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," *Scientific data*, vol. 2, p. 150066, 2015.
- [30] W. No, "1091," *Guidelines on Ensemble Prediction Systems and Forecasting*, 2012.
- [31] C. Jones, F. Giorgi, and G. Asrar, "The coordinated regional downscaling experiment: Cordex—an international downscaling link to cmip5," *CLIVAR exchanges*, vol. 16, no. 2, pp. 34–40, 2011.
- [32] SMHI. (2006) Projection results from the cordex africa domain. [Online]. Available: <http://css.escwa.org.lb/SDPD/1827/3-2.pdf>
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.