Journal of
Water & Climate Change

Journal of Water and Climate Change Vol 13 No 4, 1673 doi: 10.2166/wcc.2022.291

Super-resolution of sea surface temperature with convolutional neural network- and generative adversarial network-based methods

Tomoki Izumi[a], Motoki Amagasaki[b,*], Kei Ishida[b] and Masato Kiyama[b]

[a] Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan
[b] Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan
*Corresponding author. E-mail: amagasaki@cs.kumamoto-u.ac.jp

TI, 0000-0002-6239-299X; MK, 0000-0003-0094-8179

ABSTRACT

In this paper, we perform the super-resolution of sea surface temperature data with the enhanced super-resolution generative adversarial network (ESRGAN), which is a deep neural network-based single-image super-resolution (SISR) method that uses a generative adversarial network (GAN). We generate high-quality super-resolution data with ESRGAN and with the super-resolution convolutional neural network (SRCNN) and residual-in-residual dense block network (RRDBNet) methods, which are based on convolutional neural networks (CNNs). The images generated with these methods are compared with high-resolution optimum interpolation sea surface temperature (OISST) data using root mean square error (RMSE), learned perceptual image patch similarity (LPIPS), and perceptual index (PI) evaluation methods. RRDBNet has a better RMSE than SRCNN and ESRGAN. However, CNN-based SISR methods do not provide a faithful representation of the ocean currents of OISST. ESRGAN has a better LPIPS and PI than CNN-based methods and can represent the complex distribution of ocean currents.

Key words: convolutional neural network, ESRGAN, generative adversarial network, RRDBNet, single-image super-resolution

HIGHLIGHTS

- RRDBNet has a better RMSE than SRCNN and ESRGAN on super-resolution of sea surface temperature data.
- ESRGAN has a better LPIPS and PI than CNN-based methods and can represent the complex distribution of ocean currents.
- CNNs cannot interpolate the missing information, but GANs have better results for these parts.

INTRODUCTION

In the field of computer vision, deep neural networks (DNNs) are used for single-image super-resolution (SISR), which converts a single low-resolution image to a high-resolution image. In particular, convolutional neural network (CNN) methods can recover high-resolution images with a better root mean square error (RMSE) and structural similarity (SSIM) (Zhou *et al.* 2003) compared with dictionary-based methods (Radu *et al.* 2014). SSIM is an evaluation metric that measures the closeness of two images using the mean and standard deviation of pixel values for each small region in an image. A super-resolution CNN (SRCNN) is the first model to apply CNN for SISR. SRCNN has three convolutional layers and has a high performance compared with the dictionary-based methods. However, the images generated by CNN-based methods tend to have blurred contours (Christian *et al.* 2017). Therefore, a method using a generative adversarial network (GAN) (Ian *et al.* 2014) has been proposed to generate images with clear contours. Although the RMSE values for GAN methods are inferior to those for CNN methods, GAN methods produce images with clear contours (Christian *et al.* 2017). The enhanced super-resolution GAN (ESRGAN) (Xintao *et al.* 2018) is an improved version of the SRGAN, which was the first GAN for SISR.

SRCNN has been used to obtain super-resolution precipitation data with a CNN (Thomas *et al.* 2017). The structure of SRCNN is simple, and the recently introduced network architecture with residual blocks and skip connections has not been introduced. ESRGAN has been used for super-resolution of the wind velocity fields, but only simple evaluation metrics have been used, and super-resolution for SST has not been tried.

In this paper, we use an advanced SISR method, ESRGAN. CNNs are used as generative networks for GAN; thus, the CNN used in a GAN should have high performance in SISR. ESRGAN and its generator network, the residual-in-residual dense block network (RRDBNet), are used to increase the resolution of low-resolution sea surface temperature data. By comparing these two methods, we clarify the effect of GANs and CNNs on super-resolution. When SISR was applied to low-resolution sea surface temperature data, RRDBNet was the most accurate in terms of RMSE, whereas ESRGAN was the most accurate in most cases for the evaluation metric of the perceptual similarity of images.

In the section 'Related Research', we describe related studies that have used GAN for climate data. In the section 'DNN Model for SISR', we describe the DNN model used. In the 'Computational Evaluation' section, we describe the data and computers used, as well as the experimental evaluation results. In the section 'Discussion', we summarize the paper.

RELATED RESEARCH

In this section, we introduce the dictionary-based method and eight studies that use GAN for meteorological data. The dictionary-based method (Radu *et al.* 2014) is an SISR method developed before the DNN-based method. First, a high-resolution image and a low-resolution image that is a reduced version of the high-resolution image are prepared. Then, a portion of an image to be converted to high-resolution is cropped, and the method searches for an approximate low-resolution image. Finally, using the corresponding high-resolution image, the cropped image is upscaled to high-resolution. This method achieves higher accuracy and faster speed by improving the search algorithm.

In satellite remote sensing, data under clouds cannot be observed. To solve this problem, Junyu *et al.* (2019) used GANs to repair deficiencies by training on historical sea surface temperature image data. A random vector is an input as a latent variable and a repaired image is generated by updating the input vector based on the error between the output generated image and the original image. The method has a larger quantitative error than Telea's method (Alexandru 2004) in terms of restoring artificially created missing regions. However, Telea's method does not perform well qualitatively in recovering missing regions of real data. This paper uses a discriminator to determine if an image is realistic or not. It is not used to reduce the pixel-wise error. Therefore, Junyu *et al.* (2019) claimed that their method is the best for both artificial and real missing regions.

Because there is a lack of heavy rainfall data in the precipitation data used for machine learning, heavy rainfall data has been generated by a GAN (Chenyang *et al.* 2020). The conventional GAN for video images generates a frame interval of 0.04 s, whereas the rainfall data has a frame interval of 6 min. Since long-term prediction is required, the time interval is mismatched for the normal GAN structure. A random vector of latent variables is input to generate a series of radar images. By using GAN, the discriminator determines whether the sample is eligible, and the generator produces a sample that is similar to the real data. Precipitation data was generated from radar echo data from Shenzhen, China, and the method showed the highest accuracy in terms of MSE, SSIM, and sharpness difference compared with the GAN for video (VGAN) (Carl *et al.* 2016), the flow-and-texture-GANs (FTGAN) (Katsunori *et al.* 2018), and the motion and content decomposed GAN (MoCoGAN) (Sergey *et al.* 2018).

For precipitation estimation using meteorological radar images, the convolutional gate recurrent unit (ConvGRU) machine learning method generates blurred images (Xingjian *et al.* 2017). The generative adversarial ConvGRU (GA-ConvGRU) (Lin *et al.* 2020) produces more realistic and accurate images by adversarially learning with ConvGRU. This method takes continuous time-series image data as input and generates a series of predictive images. Precipitation prediction was performed using radar echo data from Guangdong Province, China, and compared with the optical flow method (Aldo *et al.* 2010) and ConvGRU. The GA-ConvGRU method was evaluated by evaluation metrics probability of detection (POD), false alarm rate (FAR), critical success index (CSI), and Heidke skill score (HSS), and showed the highest accuracy.

In precipitation nowcasting using meteorological radar images, machine learning has been used for long-term prediction. The adversarial extrapolation neural network (AENN) (Jinrui *et al.* 2019) can produce images with high perceptual quality by adversarially learning with the convolutional long short-term memory (ConvLSTM) network (Xingjian *et al.* 2015). This network takes five consecutive meteorological radar images as input and generates images at 0.5, 1, and 1.5 h later. A discriminator judges between one generated frame, two consecutive frames, and the target image. The AENN was evaluated by the metrics POD, FAR, CSI, and HSS, and it was the second-best for FAR after the ConvLSTM method and the best for the other metrics, reducing image blur and producing images with higher perceptual quality.

In three-dimensional geological modeling methods based on multi-point statistics, the computational complexity becomes huge as the number of input variables increases. The method reported by Eric *et al.* (2018) enables efficient modeling using a GAN to generate simulation results. The GAN model can quickly generate the more accurate data compared with a computational model. The input is a uniform distribution and the output is a two-dimensional or three-dimensional grayscale image. The data generated by this method was similar to the real data, the quantitative error was small, and the time required for generation was short.

The method developed by Hayatbini *et al.* (2019) uses a GAN to generate precipitation distributions from multiple datasets obtained from geostationary satellites to improve precipitation accuracy for satellite-derived data. By adversarially training the U-net (Olaf *et al.* 2015), more complex precipitation features can be captured. This method was evaluated by MSE, COR, BIAS, POD, FAR, and CSI and had better accuracy than did the CNN model. The combination of GAN and MSE loss produces a complex and better distribution of precipitation.

Image super-resolution technology has been used to downscale meteorological data. The method reported by Alok *et al.* (2019) performs quadruple downscaling of wind speed data using ESRGAN. This method uses low-resolution data of the wind velocity field as input and outputs high-resolution data. Although the quantitative quality was inferior to existing methods, the generated images were clearer and less prone to artifacts. This method can recover spatial details with higher fidelity than bicubic upsampling or SRCNN. This method had excellent visual quality because the power spectrum of the frequency matched that of the correct image even at high frequencies.

We use advanced SISR methods, ESRGAN and RRDBNet, to perform super-resolution of low-resolution SST data. Then, we compare these methods to clarify the effect of GAN and CNN on super-resolution of SST data.

DNN MODEL FOR SISR

Features of CNNs and GANs

CNN-based methods can achieve higher resolution with a better RMSE and SSIM than dictionary-based methods (Chao *et al.* 2014; Christian *et al.* 2017). However, because CNNs minimize the MSE per pixel, they cannot represent the image details and produce a blurry image. To solve this problem, GAN-based methods use adversarial loss to distinguish between super-resolution images and realistic images, and use content loss to minimize the difference of apparent similarity. Although the accuracy of RMSE is lower than that of CNNs, GANs are expected to produce images with high perceptual quality.

CNN model

SRCNN (Chao *et al.* 2014) and RRDBNet (Xintao *et al.* 2018) are used as CNN-based methods for image-level super-resolution. SRCNN (Chao *et al.* 2014) is input after upsampling low-resolution images to the desired resolution in advance. Each layer performs the roles of patch extraction, nonlinear mapping, and reconstruction in the dictionary-based methods. The network architecture of SRCNN is shown in Figure 1. The numbers above the convolutional layer indicate the size of the kernel and the number of channels (length×width×number of channels). SRCNN has a three-layer CNN with a pair of convolutional layers and rectified linear unit (ReLU) layers. MSE loss is used as the loss function.

RRDBNet (Xintao *et al.* 2018) has a network architecture using the RRDB, which is a combination of dense blocks and residuals. The network architecture of RRDBNet is shown in Figure 2. The first half of the network has a structure that uses multiple RRDBs. In the upsampling layer, the size of the feature map is doubled. Figure 3 shows the network architecture

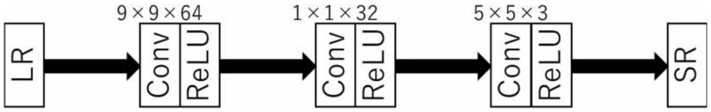


Figure 1 | Network architecture of SRCNN.

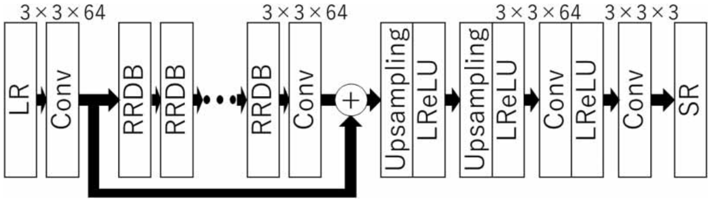


Figure 2 | Network architecture of RRDBNet.

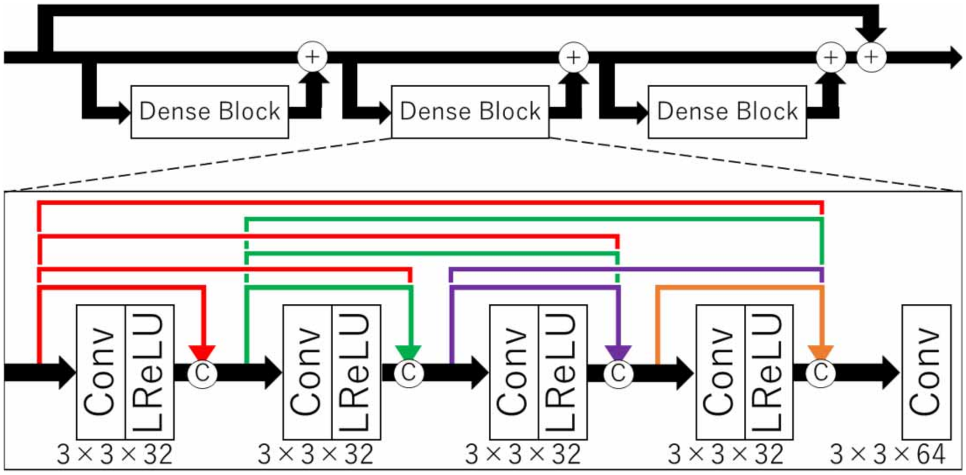


Figure 3 | Network architecture of RRDB and dense blocks.

of the RRDB used in RRDBNet and the dense blocks used in the RRDB. The RRDB consists of three dense blocks joined by residuals. One dense block is composed of a pair of convolutional layers and leaky ReLU layers joined by residuals. For the loss function, MSE loss is used.

GAN model

The CNN-based methods generate blurry and unnatural-looking images, although they have high accuracy in RMSE because their methods minimize the average of pixel-wise errors. However, the GAN-based method generates natural-looking images and achieves higher accuracy in the perceptual evaluation metric than the CNN-based method. ESRGAN (Xintao *et al.* 2018) is a GAN-based method. The generator network is composed of RRDBNet (Xintao *et al.* 2018), and the discriminator network is the VGG16 CNN (Karen & Andrew 2015). The generator network trained with MSE loss only is used as a pretraining model. A trained VGG16 network separate from the discriminator is used for the loss function. The network architecture of the ESRGAN discriminator is shown in Figure 4, where the number on the linear layer represents the number of nodes. The structure consists of multiple repeated pairs of convolutional layers, a batch normalization layer, and a leaky ReLU layer. The loss functions are MSE loss, VGG loss, and adversarial loss, which can produce images with high perceptual quality by training the network with the sum of the three losses. VGG loss is written as

$$l_{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(I^{SR})_{x,y})^2 \qquad (1)$$

where $\phi_{i,j}$ is the feature map, which is obtained by the jth convolutional layer before the ith pooling layer of the VGG network, I_{SR} is the super-resolution image, I_{HR} is the high-resolution image, and $W_{i,j}$ and $H_{i,j}$ denote the dimension of each feature map in the VGG network.

Evaluation metrics

We introduce learned perceptual image patch similarity (LPIPS) (Richard *et al.* 2018) and the perceptual index (PI) (Yochai *et al.* 2018) as evaluation indexes for perceptual similarity. LPIPS is an evaluation method for natural images using a pretrained neural network. Each image is input to the neural network, and the error of each feature map is used for evaluation.

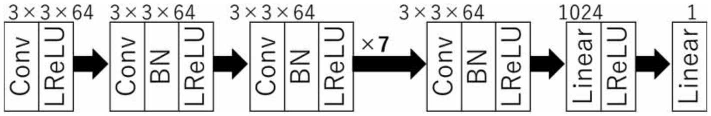


Figure 4 | Discriminator network architecture of ESRGAN.

The evaluation value of the PI is calculated by the following equation:

$$\text{Perceptual Index} = \frac{1}{2}((10 - \text{Ma}) + \text{NIQE}) \tag{2}$$

Here, Ma (Chao *et al.* 2017) is an evaluation metric that calculates the evaluation value by learning two regression models using three statistical features. NIQE (naturalness image quality evaluator) (Anish *et al.* 2013) uses statistical features called natural scene statistics to train its models. These metrics can be used to evaluate the similarity of images as seen by humans, compared with methods that compare pixel-by-pixel errors in images, such as RMSE.

COMPUTATIONAL EVALUATION

Datasets

The observation points for sea surface temperature data are arranged in a grid pattern on the sea surface. In this evaluation, optimum interpolation sea surface temperature (OISST) and ERA20C are used as the datasets for sea surface temperature. The OISST dataset is recent and has a high resolution of about 25 km (distance between observation points) and a short period from 1982 to the present, whereas the ERA20C dataset has a low resolution of about 125 km and a long observation period from 1900 to 2010. We use ERA20C (Paul *et al.* 2016) as low-resolution data and OISST (Viva *et al.* 2020) as high-resolution data.

Evaluation environment

As training data, we use the data from 8,401 days from 1985 to 2010. As testing data, we use the data from 1,096 days from 1982 to 1984. ERA20C data are interpolated using bilinear interpolation and extrapolated using nearest-neighbor interpolation to the same resolution as the OISST data. Figure 5 shows the differences between ERA20C and OISST. Small differences are shown in blue and larger differences are shown in red. Large differences indicate that low-resolution data has too little information compared with high-resolution data. The differences are particularly large around Japan, eastern North America, and western South America; thus, these locations are selected for evaluation. Then, we focus on two areas each around the coasts of Japan (JP1 and JP2), North America (NA1 and NA2), and South America (SA1 and SA2), shown as areas enclosed by squares in Figure 5. Figure 6 shows plots of the OISST data for 1 day. Higher temperatures are shown in red and lower temperatures are shown in blue. The latitude and longitude range values for each area are shown at the bottom of each figure. The grid size is 64×64 for high-resolution data and 16×16 for low-resolution data, with a magnification factor of 4. The linear transformation, SRCNN, RRDBNet, and ESRGAN evaluation models are used. The DNN-based method not only uses a single model to perform four-fold super-resolution, but also uses a two-stage super-resolution model with two models that perform two-fold super-resolution. In the first step, we use the ERA20C data as input and the OISST data downscaled to 1/2 resolution by linear transformation as the supervisor data. In the second stage, the model is trained to estimate the OISST data from the OISST data downscaled to 1/2 resolution by

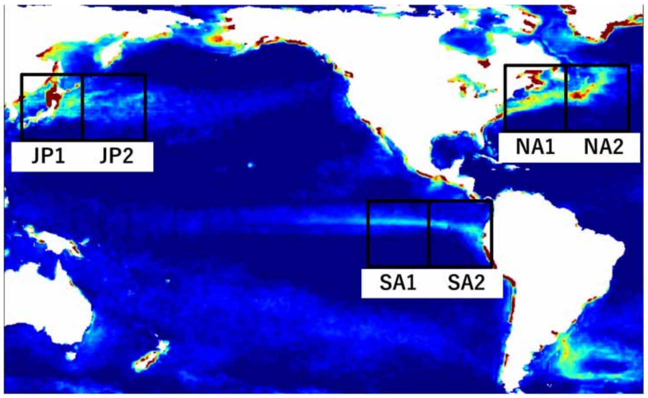


Figure 5 | Difference between ERA20C and OISST data. Small differences are shown in blue and larger differences are shown in red. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wcc.2022.291.

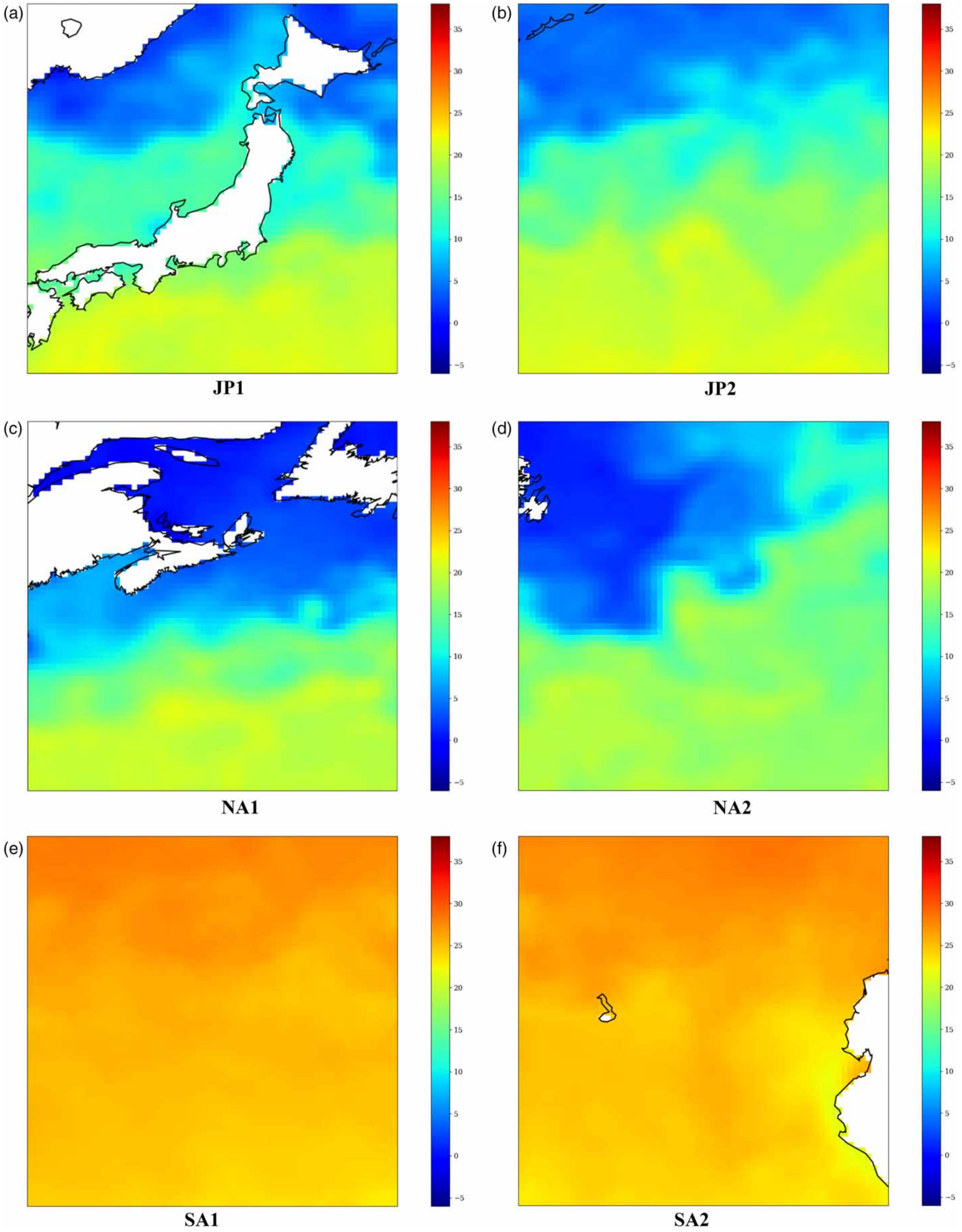


Figure 6 | Data from the cropped areas (squares in Figure 5). Plots of high-resolution data for 1 day for six different areas. Higher temperatures are shown in red and lower temperatures are shown in blue. (a) JP1 longitude (130.5, 146.5) latitude (30, 46). (b) JP2 longitude (146.5, 162.5) latitude (30, 46). (c) NA1 longitude (290, 306) latitude (35, 51). (d) NA2 longitude (306, 322) latitude (35, 51). (e) SA1 longitude (249, 265) latitude (−9, 7). (f) SA2 longitude (265, 281) latitude (−9, 7). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wcc.2022.291.

linear transformation. The main hyperparameters are shown in Table 1. Bilinear transformation is used for data interpolation, and then nearest-neighbor interpolation is performed to compensate for the missing values over land. For SRCNN, low-resolution images are upsampled to the desired resolution by nearest-neighbor interpolation. SRCNN and RRDBNet use cosine annealing as the learning rate scheduler. RRDBNet and ESRGAN use a network of 23 RRDBs. ESRGAN uses VGG128 (Karen & Andrew 2015) as a discriminator and a trained RRDBNet as a pretraining model for the generator. The learning rate is multiplied by 0.5 every 5,000 iterations, and the loss values for MSE loss, VGG loss, and adversarial loss are summed by multiplying them by factor weights of 1×10^{-2}, 1, and 5×10^{-3}, respectively. We use the PyTorch framework version 1.1.1 and NVIDIA Tesla V100 GPUs for training.

Evaluation results

Table 2 shows the RMSE evaluation of the linear transformation, SRCNN, RRDBNet, and ESRGAN methods, and the average values. The values on the left and right sides of each cell are the results of the one-step and two-step models, respectively. The smaller the value of RMSE, the smaller the error. On average, RRDBNet has the highest accuracy, followed by linear transformation, ESRGAN, and SRCNN.

Figure 7 shows the images of the (a) high- and (b) low-resolution data for the JP2 area, and the results of the (c) linear transformation, (d) SRCNN, (e) RRDBNet, and (f) ESRGAN methods. Higher temperatures are shown in red and lower temperatures are shown in blue. The images generated by the linear transformation, SRCNN, and RRDBNet have smooth, blurred temperature distributions, whereas the image generated by ESRGAN has a perceptual quality similar to that of the OISST data.

Table 3 shows the LPIPS evaluation for the linear transformation, SRCNN, RRDBNet, and ESRGAN methods, and the average values. The values on the left and right sides of each cell are the results of the one-step and two-step models, respectively. The smaller the value of LPIPS, the lower the perceptual error. ESRGAN has the highest accuracy in all areas. On average, ESRGAN has the highest accuracy, followed by linear transformation, RRDBNet, and SRCNN.

Table 4 shows the PI evaluation of the of linear transformation, SRCNN, RRDBNet and ESRGAN methods, and the average values. The values on the left and right sides of each cell are the results of the one-step and two-step models, respectively.

Table 1 | Hyperparameters

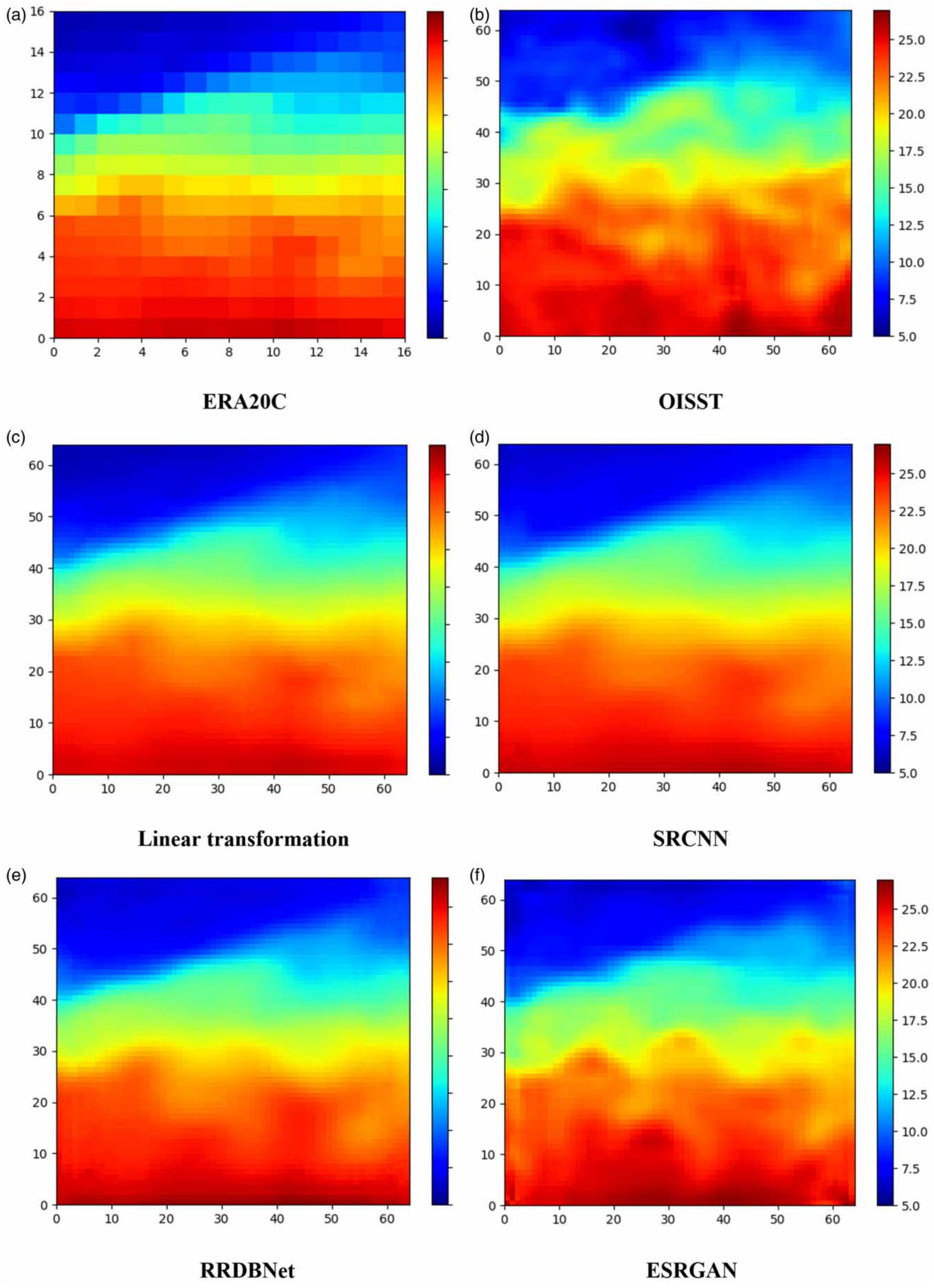
	SRCNN	RRDBNet	ESRGAN
Batch size	32	32	32
Optimizer	Adam	Adam	Adam
Adamβ1	0.9	0.9	0.9
Adamβ2	0.999	0.999	0.999
Iteration	25,000	25,000	25,000
Initial value of learning rate	2×10^{-4}	2×10^{-4}	1×10^{-4}

These hyperparameters were used to train SRCNN, RRDBNet, and ESRGAN.

Table 2 | Evaluation results (RMSE)

	Linear transformation	SRCNN	RRDBNet	ESRGAN
JP1	1.362	3.632/3.702	1.043/1.032	1.350/1.159
JP2	0.944	0.920/0.923	0.918/0.928	1.195/1.045
NA1	1.194	2.238/2.098	1.052/1.076	1.492/1.276
NA2	1.012	1.023/1.062	1.020/1.042	1.610/1.182
SA1	0.688	0.693/0.699	0.698/0.701	0.979/1.703
SA2	0.843	2.199/1.476	0.772/0.785	0.959/1.898
Average	1.007	1.784/1.660	0.917/0.927	1.264/1.377

The RMSE evaluation of the linear transformation, SRCNN, RRDBNet and ESRGAN methods, and the average values.



ERA20C

OISST

Linear transformation

SRCNN

RRDBNet

ESRGAN

Figure 7 | Plots of evaluation results. Plots of low-resolution and high-resolution data and data generated by linear transformation, SRCNN, RRDBNet, and ESRGAN for the same region and date. Higher temperatures are shown in red and lower temperatures are shown in blue. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/wcc.2022.291.

Table 3 | Evaluation results (LPIPS)

	Linear transformation	SRCNN	RRDBNet	ESRGAN
JP1	0.01448	0.12194/0.07234	0.00624/0.00619	**0.00571**/**0.00538**
JP2	0.17027	0.16650/0.17263	0.13175/0.13233	**0.06612**/**0.07217**
NA1	0.06124	0.08539/0.10253	0.04602/0.04815	**0.03050**/**0.02163**
NA2	0.13447	0.14562/0.15386	0.12597/0.10609	**0.05738**/**0.05806**
SA1	0.09993	0.10954/0.11315	0.10784/**0.10600**	**0.04678**/0.10625
SA2	0.01321	0.19744/0.15946	0.14497/**0.01378**	**0.00886**/0.02695
Average	0.08227	0.13774/0.12900	0.09380/0.06876	**0.03589**/**0.04841**

The LPIPS evaluation of the linear transformation, SRCNN, RRDBNet and ESRGAN methods, and the average values.
Bold values indicate the highest accuracy.

Table 4 | The PI evaluation of the linear transformation, SRCNN, RRDBNet and ESRGAN methods, and the average values

	Linear transformation	SRCNN	RRDBNet	ESRGAN
JP1	21.9686	22.1180/22.0934	22.0289/21.9841	**21.9466**/**21.9087**
JP2	22.9595	**22.8285**/22.7960	22.8887/22.8300	22.8674/**22.6625**
NA1	22.2340	22.2518/22.3164	22.3107/22.2367	**22.2047**/**22.1554**
NA2	22.8005	22.8112/22.7771	22.8198/22.6981	**22.5471**/**22.4030**
SA1	22.9451	22.9779/22.9875	22.9405/22.9233	**22.7866**/**22.5832**
SA2	22.7146	23.0478/22.8971	22.8021/22.6560	**22.2328**/**21.9456**
Average	22.6037	22.6725/22.6446	22.6318/22.5547	**22.4309**/**22.2764**

Bold values indicate the highest accuracy.

The smaller the value of PI, the smaller the perceptual error. ESRGAN has the highest accuracy in five of the six areas. On average, ESRGAN has the highest accuracy, followed by linear transformation, RRDBNet, and SRCNN.

DISCUSSION

Table 2 shows that, on average, RRDBNet has the highest accuracy; thus, RRDBNet is useful for reducing the pixel-by-pixel error in super-resolved sea surface temperature data. Conversely, the accuracy of SRCNN is lower, especially in areas where the data contains land. Because the data is filled with missing values over land and SRCNN has a three-layer network architecture, a deeper layer is necessary to increase the resolution of data with missing values. Figure 7 shows that the distribution of sea surface temperature generated by the CNN is smooth and blurred. Therefore, when the ERA20C data are converted to higher resolution, the missing information is not interpolated. However, the sea surface temperature generated by ESRGAN has a high perceptual quality, and the temperature distribution missing from the low-resolution data is generated correctly.

There are some areas where the temperature distribution differs greatly from that of the high-resolution data. Therefore, although ESRGAN can reproduce the temperature distribution close to that of the high-resolution data, further improvement of the model is needed. In addition, Tables 3 and 4 show that ESRGAN has high accuracy in the perceptual similarity metric, whereas the accuracy of the other methods is low. However, comparing the images in Figure 7, ESRGAN generates a distribution of sea surface temperature that is not generated by the CNN method. Therefore, it is possible that the GAN method is able to generate data missing in low-resolution. As a future study, it will be necessary to further investigate whether the GAN method is able to generate higher frequency information than the CNN method. This feature of GANs may be an important advantage in SISR. The SISR results obtained using a GAN may be more suitable for analysis of sea surface temperature in a local area, compared with those from a CNN.

Improvement of sea surface temperature data will be valuable for water management. Sea surface temperature influences the amount of precipitation. Meanwhile, it is important for water management to investigate climate change impacts on precipitation. Most future climate change assessment studies nowadays utilize future climate projections generated by general circulation models. However, their spatial resolutions are mostly more than 100 km. The use of accurate higher-resolution sea surface temperature data as forcing data for dynamical downscaling, for example, may improve the accuracy of precipitation estimation under future climate conditions. Especially, a GAN may have potential to improve the accuracy of dynamical downscaling at a high resolution because it correctly generates the temperature distribution missing from the low-resolution data.

CONCLUSION

The use of DNNs for SISR has been studied extensively. CNN-based methods can achieve higher resolution with higher accuracy than dictionary-based methods, although the generated images tend to be blurred. In contrast, GAN-based methods can produce images with high perceptual quality. In this study, an SISR method using a DNN was applied to super-resolution of sea surface temperature data. We used RRDBNet, which is an advanced CNN method, and ESRGAN, which is a GAN method. RRDBNet produced super-resolution images with the highest accuracy, whereas ESRGAN produced images with high perceptual quality. The perceptual evaluation metrics showed that ESRGAN had the best accuracy. The GAN-based method reproduced the sea surface temperature distribution, which was not reproduced by the CNN-based method. The use of DNN in other natural sciences is expected to improve the accuracy of super-resolution.

As described above, the results in this study may be useful as forcing data for dynamical downscaling. In a future study, the results will be utilized for forcing data of dynamical downscaling. Then, the estimation accuracy of precipitation by dynamical downscaling will be investigated to evaluate the SISR results by the DNNs as the forcing data.

DATA AVAILABILITY STATEMENT

All relevant data are available from https://github.com/izumi1112/Super-resolution-for-sea-surface-temperature-with-CNN-and-GAN.

REFERENCES

Aldo, B., Isztar, Z., Alamelu, K., Hee, C. L., Lee, Y. H. & Gyuwon, L. 2010 Mcgill algorithm for precipitation nowcasting by lagrangian extrapolation (MAPLE) applied to the South Korean radar network. *Proceedings of Asia-Pacific Journal of the Atmospheric Sciences* **46**, 369–381.

Alexandru, T. 2004 An image inpainting technique based on the fast marching method. *Proceedings of Journal of Graphics Tools* **9** (1), 23–34.

Alok, S., Adrian, A. & Brian, W. 2019 Downscaling numerical weather models with GANs. In: *Proceedingsof American Geophysical Union Fall Meeting Abstracts*.

Anish, M., Rajiv, S. & Alan, B. 2013 Making a 'completely blind' image quality analyzer. In: *Proceedings of IEEE Signal Processing Letters*, pp. 209–212.

Carl, V., Hamed, P. & Antonio, T. 2016 Generating videos with scene dynamics. In: *Proceedings of Neural Information Processing Systems (NIPS)*, pp. 613–621.

Chao, D., Chen, C. L., Kaiming, H. & Xiaoou, T. 2014 Learning a deep convolutional network for image super-resolution. In: *Proceedings of European Conference on Computer Vision (ECCV)*, pp. 184–199.

Chao, M., Chih, Y. Y., Xiaokang, Y. & Ming, H. Y. 2017 Learning a no-reference quality metric for single-image super-resolution. In: *Proceedings of Computer Vision and Image Understanding (CVIU)*, pp. 1–16.

Chenyang, Z., Xuebing, Y., Yongqiang, T. & Wensheng, Z. 2020 Learning to generate radar image sequences using two-stage generative adversarial networks. In: *Proceedings of IEEE Geoscience and Remote Sensing Letters*, pp. 401–405.

Christian, L., Lucas, T., Ferenc, H., Jose, C., Andrew, C., Alejandro, A., Andrew, A., Alykhan, T., Johannes, T., Zehan, W. & Wenzhe, S. 2017 Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114.

Eric, L., Romain, H., Diederik, J. & Niklas, L. 2018 Training-image based geostatistical inversion using a spatial generative adversarial neural network. In *Proceedings of Water Resources Research*, pp. 381–406.

Hayatbini, N., Kong, B., Hsu, K., Nguyen, P., Sorooshian, S., Stephens, G., Fowlkes, C., Nemani, R. & Ganguly, S. 2019 Conditional Generative Adversarial Networks (cGANs) for near real-time precipitation estimation from multispectral GOES-16 satellite imageries-PERSIANN-cGAN. In: *Proceedings of Remote Sensing.*

Ian, G., Jean, P. A., Mehdi, M., Bing, X., David, W. F., Sherjil, O., Aaron, C. & Yoshua, B. 2014 Generative adversarial nets. In *Proceedings of Neural Information Processing Systems (NIPS)*, pp. 2672–2680.

Jinrui, J., Qian, L., Xinya, D., Nengli, S., Rong, T. & Yali, C. 2019 AENN: a generative adversarial neural network for weather radar echo extrapolation. In: *Proceedings of ISPRS Workshop on Remote Sensing and Synergic Analysis on Atmospheric Environment (RSAE)*, pp.89–94.

Junyu, D., Ruiying, Y., Xin, S., Qiong, L., Yuting, Y. & Xukun, Q. 2019 Inpainting of remote sensing sst images with deep convolutional generative adversarial network. In: *Proceedings of IEEE Geoscience and Remote Sensing Letters*, pp. 173–177.

Karen, S. & Andrew, Z. 2015 Very deep convolutional networks for large-scale image recognition. In: *Proceedings of International Conference on Learning Representations*.

Katsunori, O., Shohei, Y., Yoshitaka, U. & Tatsuya, H. 2018 Hierarchical video generation from orthogonal information: optical flow and texture. In: *Proceedings of Association for the Advancement of Artificial Intelligence (AAAI)*, pp. 2387–2394.

Lin, T., Xutao, L., Yunming, Y., Pengfei, X. & Yan, L. 2020 A generative adversarial gated recurrent unit model for precipitation nowcasting. In: *Proceedings of IEEE Geoscience and Remote Sensing Letters*, pp. 601–605.

Olaf, R., Philipp, F. & Thomas, B. 2015 U-Net: convolutional networks for biomedical image segmentation. In: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241.

Paul, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thepaut, J. N., Tremolet Y., H., Elias, V., Bonavita, M., Isaksen, L. & Fisher, M. 2016 ERA-20C: an atmospheric reanalysis of the twentieth century. *Proceedings of Journal of Climate* **29** (11), 4083–4097.

Radu, T., Vincent, D. S. & Luc, V. G. 2014 A + : adjusted anchored neighborhood regression for fast super-resolution. In: *Proceedings of IEEE Asian Conference on Computer Vision (ACCV)*, pp. 111–126.

Richard, Z., Phillip, I., Alexei, A. E., Eli, S. & Oliver, W. 2018 The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595.

Sergey, T., Ming, Y. L., Xiaodong, Y. & Jan, K. 2018 MoCoGAN: decomposing motion and content for video generation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1526–1535.

Thomas, V., Evan, K., Sangram, G., Andrew, M., Ramakrishna, N. & Auroop, R. G. 2017 Deepsd: generating high resolution climate change projections through single image super-resolution. In: *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1663–1672.

Viva, B., Thomas, M. S., Michael, S., Boyin, H. & Huai, M. Z. 2020 Improved estimation of proxy sea surface temperature in the Arctic. *Proceedings of Journal of Atmospheric and Oceanic Technology* **37** (2), 341–349.

Xingjian, S., Zhourong, C., Hao, W., Dit, Y. Y., Waikin, W. & Wangchun, W. 2015 Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: *Proceedings of Advances in Neural Information Processing Systems*, pp. 802–810.

Xingjian, S., Zhihan, G., Leonard, L., Hao, W., Dit, Y. Y., Waikin, W. & Wangchun, W. 2017 Deep learning for precipitation nowcasting: a benchmark and a new model. In: *Proceedings of Neural Information Processing Systems (NIPS)*, pp. 5617–5627.

Xintao, W., Ke, Y., Shixiang, W., Jinjin, G., Yihao, L., Chao, D., Yu, Q. & Chen, C. L. 2018 ESRGAN: enhanced super-resolution generative adversarial networks. In: *Proceedings of The European Conference on Computer Vision Workshops (ECCVW)*.

Yochai, B., Roey, M., Radu, T., Tomer, M. & Lihi, Z. M. 2018 The 2018 PIRM challenge on perceptual image super-resolution. In: *Proceedings of European Conference on Computer Vision (ECCV)*, pp. 334–355.

Zhou, W., Eero, P. S. & Alan, C. B. 2003 Multiscale structural similarity for image quality assessment. In: *Proceedings of Record of the Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402.

First received 27 July 2021; accepted in revised form 9 March 2022. Available online 23 March 2022