IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10640546

CONTEXT CONSTRAINTS-GUIDED RECONSTRUCTION ROBUST NETWORK FOR CLOUD-INDUCED EXTENSIVE MISSING AREA IN SEA SURFACE TEMPERATURE

Zijie Zuo[1], *Jie Nie*[1], *Xin Wang*[2], *Xinyue Liang*[1], *Bo Yin*[1*]

[1]Ocean University of China, Faculty of Information Science and Engineering, Qingdao, China, 266100
[2]Tsinghua University, Department of Computer Science and Technology, Beijing, China, 100084

ABSTRACT

Sea surface temperature (SST), a crucial indicator widely applied in marine-related fields, is often observed by extensive cloud cover in satellite observations, resulting in significant data gaps. While deep learning approaches offer solutions for reconstructing SST images using ample historical data, the majority of current methods overlook the context constraints inherent in oceanographic data, such as sequential correlations and global consistency. This oversight leads to non-robust completion results in the presence of extensive data gaps. This paper proposes a robust SST reconstruction network guided by sequential correlations and global consistency, designed to handle the task of extensive missing data recovery in SST under large-scale cloud cover, namely CCG_RRN. Extensive comparison experiments and visualization results affirm the effectiveness and robustness of CCG_RRN compared with the State-Of-The-Art (SOTA) methods on the public NSOAS SST datasets.

Index Terms— Sea surface temperature, deep neural network, context constraints

1. INTRODUCTION

SST serves as a crucial indicator in marine-related industries such as fisheries[1] and meteorology[2]. Despite real-time access to global SST images via meteorological satellites, the presence of extensive cloud cover often leads to significant data gaps in the observed SST image. Although traditional interpolation-based methods[3, 4, 5, 6] can reconstruct SST using valid pixels in the SST image, the result is significantly constrained when facing large-scale data gaps. Another category extensively employed for forecasting is data assimilation methods[7, 8], which effectively simulate accurate values utilizing numerical models. However, they are constrained by the provision of initial and boundary fields, leading to a substantial consumption of time costs. The nonlinear capabilities of deep neural networks (DNN) have led to widespread adoption in SST image reconstruction, these works employ the Auto-encoder[9], the U-net[10], and Generative Adversarial Networks (GAN)[11]. However, these DNN-based methods often ignore the inherent context characteristics of oceanographic data, such as the sequential correlations and global consistency as shown in Fig. 1, causing non-robust. Therefore, this paper introduces a **C**ontext **C**onstraints-**G**uided **R**econstruction **R**obust **N**etwork (CCG_RRN), which incorporates a U-net-based global encoder for obtaining a global consistency embedding from a short-term average SST image. It also utilizes a ConvGRU-based sequence Modeling module for extracting a Sequential correlation embedding from sequential images. A designed decoder fuses these embeddings, reconstructing a cloud-free SST image. Our contributions include: 1) We propose a robust method, namely CCG_RRN, for sea surface temperature imputation, capable of effectively completing data under extensive missing conditions. 2) We introduce a ConvGRU-based Sequential Modeling module and design a fusing module for integrating sequential correlation and global consistency (SEGC_DFD), which leads to a rational and novel incorporation of contextual information. 3) Sufficient experiments on the NSOAS SST dataset validate the effectiveness and robustness of our proposal against other SOTA methods.

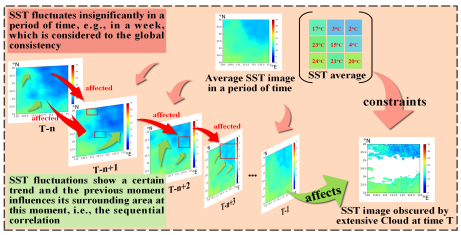


Fig. 1: The global consistency and sequential correlation in SSTs.

2. RELATED WORK

Traditional SST reconstruction methods are interpolation-based methods including optimal interpolation[3], linear interpolation[4], lanczos interpolation[5], and spline interpolation[6], Data Interpolating Empirical Orthogonal Functions (DINEOF)[12], etc. However, these interpolation methods typically rely on effective regions for reconstruction,

*Corresponding author

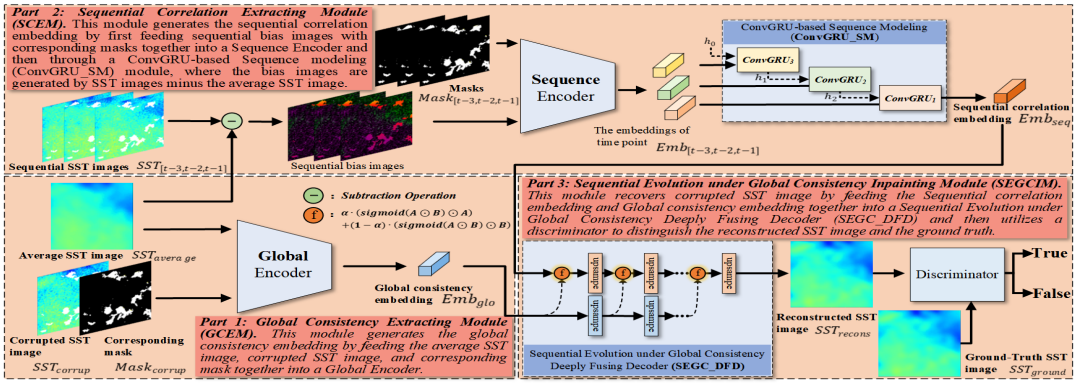


Fig. 2: The main framework of proposed CCG_RRN, concluding: 1) Global Consistency Extracting Module (GCEM), which takes the average SST image, corrupted SST image and its mask as inputs and outputs the global consistency embedding by a Global-Encoder; 2) Sequential correlation Extracting Module (SCEM), which takes sequential SST images as inputs and outputs a Sequential correlation embedding by a Sequence-Encoder and a ConvGRU-based module (ConvGRU_SM); 3) Sequential Evolution under Global Consistency Inpainting Module (SEGCIM), which takes the global consistency embedding and Sequential correlation embedding as inputs and outputs the reconstructed SST image by a proposed Deeply Fusing Decoder (SEGC_DFD). Finally, CCG_RRN applies a discriminator to supervise the reconstructed SST image with the ground truth.

which neglects the utilization of historical data and results in limitations on the quality of imputation.

With the development of DNN, the powerful non-linear fitting capability it enables them to learn the distribution patterns of historical data, numerous DNN-based methods have been employed in this task [13, 14, 15, 16], such as [13] proposes a "coarse to fine" inpainting network for SST reconstruction based on GAN. [14] designs a new loss function to reconstruct the corrupted SST image based on DCGAN. AIN[15] proposes a GAN-based anomaly inpainting network for SST image reconstruction. DINCAE[16] achieves SST reconstruction through the SST anomalies and their expected error variance based on U-net.

3. METHOD

3.1. framework

Fig. 2 shows the main framework of the proposed CCG_RRN. We suppose that the corrupted SST image is obtained at date t, which belongs to week W, and then we apply sequential SST images earlier than date t but within W. In this paper, we simply select the images of three consecutive days before the date t as the sequential SST images. By inputting the corrupted SST image, denotes as SST_{corrup}, an average SST image of W, denotes as $SST_{average}$, and the sequential SST images, denotes as $SST_{[t-3,t-2,t-1]}$, our goal is to fill the missing area of SST_{corrup} and make it to meet the global consistency and the temporal pattern.

The CCG_RRN mainly consists of three modules: 1) The Global Consistency Extracting Module (GCEM), which takes in the SST_{corrup} and $SST_{average}$ as inputs and outputs a global consistency embedding through a global encoder, denoted as Emb_{glo}; 2) The sequential correlation Extracting Module (SCEM), which takes the $SST_{[t-3,t-2,t-1]}$ and corresponding masks as inputs and outputs a Sequential correlation embedding through a Sequence-Encoder and a proposed ConvGRU-based Sequence Modeling module (ConvGRU_SM), denoted as Emb_{seq}; 3) The Sequential Evolution under Global Consistency Inpainting Module (SEGCIM), which takes the Emb_{glo} and the Emb_{seq} as inputs and outputs a reconstructed SST image SST_{recons} through a deeply fusing decoder. Finally, we apply a discriminator to supervise the SST_{recons} with ground truth as similar in GAN.

3.2. Global consistency extracting module

This module aims to extract the global consistency embedding based on SST_{corrup} and $SST_{average}$ by firstly filtering those invalid areas of SST_{corrup} and $SST_{average}$ to effectively obtain the embedding of the effective area of SST_{corrup} as follows:

$$SST_{corrup} = SST_{corrup} \odot Mask_{corrup}, \tag{1}$$

where $\odot$ is an element-wise multiplication and $Mask_{corrup}$ is a binary cloud mask of SST_{corrup} that is 0 if the pixel is covered, otherwise to 0, the $SST_{average}$ similarly. And then GCEM applies a U-net-based global encoder with four layers of the same structure to down-sample the inputs SST_{corrup} and $SST_{average}$ to a global consistency embedding Emb_{glo}, which represents the global stability of a week. Specifically, the global encoder consists of four layers and each layer concludes two gated convolution layers with kernel size set as 3, note that we replace the vanilla convolution with the gated convolution to avoid the visual artifacts[18] and a max-pool layer with kernel size of 2.

3.3. Sequential correlation Extracting Module

This module aims to extract the Sequential correlation embedding from $SST_{[t-3,t-2,t-1]}$ by first subtracting the $SST_{average}$ from the $SST_{[t-3,t-2,t-1]}$ to obtain the sequential bias images, i.e., the daily deviation, with filtering out

Table 1: The comparison results for SST image reconstruction using five methods on the NSOAS dataset. We use the Root Mean Square Error ($RMSE$) and R-squared (R^2) as metrics to evaluate the reconstruction results. The best and second-best Root Mean Squared Errors ($RMSE$) and R-squared (R^2) are in **Bold** and underline under different cloud cover ratios with different Noise/Signal ratios (N/S).

N/S	Cover Ratio	AIN[15]		CF_DGM[17]		DINEOF[12]		DINCAE[16]		Ours	
		$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2
0.1	8%	0.1588	0.9444	0.2857	0.7982	0.1987	0.9441	0.1624	0.8787	**0.1539**	**0.9619**
	25%	0.1761	0.8579	0.2518	0.2773	0.1727	0.8579	0.1870	0.7776	**0.1624**	**0.8615**
	46%	0.1878	0.7237	0.2532	0.2031	**0.1515**	**0.7528**	0.1714	0.5942	0.1549	0.7455
	68%	0.2003	0.5742	0.2646	0.1961	0.1912	0.0393	0.2018	0.3348	**0.1432**	**0.6526**
0.2	8%	0.1642	0.9581	0.2863	0.7426	0.1984	0.9532	0.1882	0.8651	**0.1630**	**0.9664**
	25%	0.1891	0.8359	0.2541	0.4556	0.1725	0.8602	**0.1658**	0.7899	0.1682	**0.8606**
	46%	0.1786	0.6724	0.2087	0.1821	0.1513	0.7308	0.1882	0.6013	**0.1384**	**0.7536**
	68%	0.2172	0.4514	0.2835	0.1951	0.1914	0.2108	0.1982	0.3609	**0.1403**	**0.5802**
0.3	8%	0.1789	0.9531	0.2908	0.8252	0.1983	0.9533	0.1930	0.8726	**0.1603**	**0.9561**
	25%	0.1810	0.7959	0.2572	0.3038	0.1726	**0.8318**	0.2175	0.7471	**0.1580**	0.8283
	46%	0.1970	0.6580	0.2098	0.3441	0.1517	0.6566	0.2204	0.4534	**0.1424**	**0.6761**
	68%	0.2285	0.4679	0.2652	0.0891	0.1912	0.0392	0.2068	0.3762	**0.1617**	**0.4773**

invalid areas as (1). Then, SCEM applies a sequence encoder with the same structure as the global encoder to encoder filtered $SST_{[t-3,t-2,t-1]}$ to three embeddings, i.e., Emb_{t-3}, Emb_{t-2}, Emb_{t-1} of time point $t-3$, $t-2$, $t-1$. Inspired by the ability of ConvGRU[19] to address temporal issues in 3D data, we introduce a ConvGRU-based Sequence Modeling module (ConvGRU_SM) with three ConvGRU in tandem, i.e., $ConvGRU_3$, $ConvGRU_2$, $ConvGRU_1$ in Fig. 2, to obtain the sequential correlation embedding as follows:

$$h_{3-i+1} = ConvGRU_i\left(Emb_{t-i}, h_{3-i}\right),\ i = 3, 2, 1\ , \quad (2)$$

where h is the hidden state, h_0 is initialized randomly. After that, the $ConvGRU_1$ outputs h_3, which is regarded as the sequential correlation embedding Emb_{seq}, including the correctness of the time series and the trend of numerical fluctuation.

3.4. Sequential evolution under global consistency inpainting module

Based on the Emb_{glo} and the Emb_{seq} obtained from above, this module aims to up-sampling and obtain the Cloud-free SST_{recons} by taking the two information into a designed Sequential Evolution under Global Consistency Deeply Fusing Decoder (SEGC_DFD), which includes two streams for up-sampling the Emb_{seq} and Emb_{glo}, respectively, i.e., up-stream and down-stream as shown in Fig. 2. Specifically, the up-stream consists of four layers and each layer concludes a skip-connection, an up-sampling layer, and two convolution layers with kernel size of 2. Note that the skip-connection is achieved based on the sequence encoder in SCEM. The down-stream concludes the same structure of three layers as the up-stream but without skip-connection. To better fuse the two information between the two streams, we apply a fusion scheme $f(\cdot)$ as follows:

$$f(A, B) = \alpha \bullet K \odot A + (1-\alpha) \bullet K \odot B, \quad (3)$$

$$K = sigmoid(A \odot B), \quad (4)$$

where α are hyper-parameters, A and B represent the output of each layer in down and up-stream, respectively. Finally, the SEGC_DFD outputs the reconstructed cloud-free SST image SST_{recons}, and a discriminator is used to supervise the SST_{recons} and the ground truth. Here, reconstruction loss L_{rec} and adversarial loss L_{adv} as follow:

$$L_{rec} = \lambda_{rec}^{ave} L_{rec}^{ave} + \lambda_{rec}^{corrup} L_{rec}^{corrup}, \quad (5)$$

$$L_{rec}^{ave} = \frac{1}{N}\sum_{1}^{N} \left\| (\hat{y} - SST_{average}) \odot Mask_{average} \right\|_2, \quad (6)$$

where λ_{rec}^{ave} and λ_{rec}^{corrup} are hyper-parameters, $\|\cdot\|_2$ is the L2 norm, and N represents the number of pixels in $\hat{y}$, i.e., SST_{recons}. $Mask_{average}$ are defined as the same in (1). The L_{rec}^{corrup} similarly.

$$L_{adv} = -\left(log\left(1 - D\left(SST_{recons}\right)\right) + logD(y)\right), \quad (7)$$

where D represents the discriminator and y is the ground truth. Then, we define the total loss function as follows:

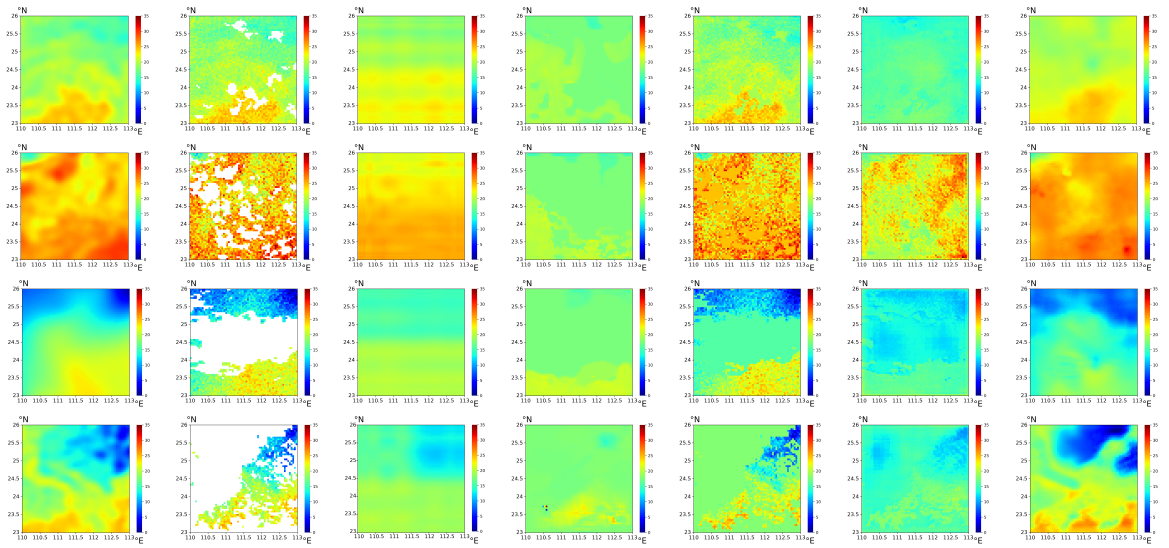
$$L_{total} = \lambda_{rec} L_{rec} + \lambda_{adv} L_{adv}, \quad (8)$$

where λ_{rec} and λ_{adv} are hyperparameters.

4. EXPERIMENTS

4.1. Datasets

In this study, we utilized NSOAS's publicly available SST level-4 products, covering January 2022 to April 2023, with no data gaps. Each pixel represents a 5-km^2 region in the South China Sea, focusing on latitudes 23°N to 26°N and longitudes 110°E to 113°E. To create corrupted SST datasets for training and validation, we normalized the data to [-1, 1]. We introduced standard normal distribution noise to simulate various Noise/Signal ratios (N/S) as following DINEOF[12]. Additionally, binary cloud masks with varying cover ratios were incorporated from the WHU cloud dataset[20]. The training spanned January to December 2022, with January to April 2023 used as the test set.



(a) Ground Truth (b) Corrupted SST (c) AIN[15] (d) CF_DGM[17] (e) DINEOF[12] (f) DINCAE[16] (g) Ours

Fig. 3: Comparative analysis of the visualization results from five methods on the NSOAS dataset. The x- and y-axes represents longitude (°E) and latitude (°N), respectively. (a) Ground truth SST image. (b) Cloud-obscured images, the first through fourth rows are missing ratios 8%, 25%, 46%, and 68%, respectively, but N/S is all 0.1. (c) Results from AIN. (d) Results from CF_DGM. (e) Results from DINEOF. (f) Results from DINCAE. (g) Results from our proposal, i.e., CCG_RRN.

4.2. Implementation details

For training our model, we set the image training size to 64 x 64 and used the Adam optimizer. The learning rate is set as 0.0001. The cloud cover rate is set to [8%, 25%, 46%, 68%] and the N/S ratio is set to [0.1, 0.2, 0.3]. The parameters of that optimizer, β_1 and β_2, are set to 0.5 and 0.99 and the hyperparameters were set as α= 0.9, λ_{rec}^{ave}=0.8, λ_{rec}^{corrup}=0.2, λ_{rec}=1, λ_{adv}=10.

4.3. Comparison to prior State-of-the-art-methods

In this section, we use the Root Mean Squared Errors ($RMSE$) and R-squared (R^2) evaluate the performance of our proposal compared with some SOTA methods as follows: 1) AIN[15]: a "coarse to fine" model to reconstruct SST image by predicting the weekly SST image and then added to an anomaly image from an anomaly inpainting network. 2) CF_DGM[17]: a GAN-based model for inpainting remote sensing images by first obtaining the rough estimation of missing areas and then refining the details based on a spatial semantic attention mechanism; 3) DINEOF[12]: a classic adaptive EOF-based decomposition method for SST reconstruction by estimate the error size of reconstructed data without any prior information; 4) DINCAE[16]: a novel U-net-based method to reconstruct SST by taking SSTs and expected error variance as inputs. The comparison results as shown in Table 1, we can observe that CCG_RRN, consistently outperforms others, especially with a 68% coverage ratio, yielding the best RMSE and R^2. In terms of RMSE, at an N/S of 0.2, CCG_RRN's RMSE reduction is 26.7% greater than DINEOF, the second-best method. In terms of R^2, similarly at an N/S of 0.2, CCG_RRN surpasses AIN, the second-best method, by 28.5%. Notably, the DINEOF experiences a substantial decline in R^2 values under all three N/S ratios as the missing data coverage increases to 68%. This decline can be attributed to the fact that DINEOF relies on valid SST data for restoration, which directly impacts the completion effectiveness. The comparative analysis of the visualization of the results is shown in Fig. 3.

5. CONCLUSIONS

Conventional DDN-based methods for sea surface temperature reconstruction utilize historical data to learn data distributions for image completion. However, most works overlook the context constraints inherent in marine science SST data. In this paper, based on the sequential correlations and global consistency, we proposed a context constraints-guided robust reconstruction network, i.e., CCG_RRN, to handle extensive data gaps caused by clouds. Through extensive comparative experiments on the public NSOAS dataset against state-of-the-art methods, CCG_RRN demonstrates robustness and effectiveness across various cloud cover ratios. Particularly, our approach outperforms SOTAs in scenarios with large-scale data gaps. We will focus on the modeling of sequential correlations over longer periods in the future.

6. ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2023YFC3108700) and the National Natural Science Foundation of China (U23A20320).

7. REFERENCES

[1] A. Miguel P Santos, "Fisheries oceanography using satellite and airborne remote sensing methods: a review," *Fisheries Research*, vol. 49, no. 1, pp. 1–20, 2000.

[2] Anne G. O'Carroll, Edward M. Armstrong, Helen M. Beggs, Marouan Bouali, and Werenfrid Wimmer, "Observational needs of sea surface temperature," *Frontiers in Marine Science*, vol. 6, 2019.

[3] Reynolds, Richard, W., Smith, Thomas, and M., "Improved global sea surface temperature analyses using optimum interpolation," *Journal of Climate*, vol. 7, no. 6, pp. 929–948, 1994.

[4] Deanna, Wilson-Diaz, , , Arthur, J., Mariano, , , Robert, H., and Evans, "A principal component analysis of sea-surface temperature in the arabian sea," *Deep Sea Research Part II Topical Studies in Oceanography*, 2001.

[5] A. Alvera-Azcárate, A. Barth, M. Rixen, and J. M. Beckers, "Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the adriatic sea surface temperature," *Ocean Modelling*, vol. 9, no. 4, pp. 325–346, 2005.

[6] S. Guinehut, A. L. Dhomps, G. Larnicol, and P. Y. Le Traon, "High resolution 3-d temperature and salinity fields derived from in situ and satellite observations," *Ocean Science Discussions*, vol. 9, no. 2, pp. 1313–1347, 2012.

[7] Re Kalman, "A new approach to linear filtering and predicted problems," *J Basic Eng*, vol. 82, 1960.

[8] Geir Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics," *Journal of Geophysical Research: Oceans*, vol. 99, no. C5, 1994.

[9] Y. Lecun, "Modeles connexionnistes de l'apprentissage (connectionist learning models)," *G. Olms*, 1987.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Y. Bengio, "Generative adversarial nets," in *Neural Information Processing Systems*, 2014.

[12] Jean-Marie Beckers and Michel Rixen, "Eof calculations and data filling from incomplete oceanographic datasets," *Journal of Atmospheric and Oceanic Technology*, vol. 20, pp. 1839–1856, 2003.

[13] Song-Hee Kang, Youngjin Choi, and Jae Young Choi, "Restoration of missing patterns on satellite infrared sea surface temperature images due to cloud coverage using deep generative inpainting network," *Journal of Marine Science and Engineering*, vol. 9, no. 3, 2021.

[14] Junyu Dong, Ruiying Yin, Xin Sun, Qiong Li, Yuting Yang, and Xukun Qin, "Inpainting of remote sensing sst images with deep convolutional generative adversarial network," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 173–177, 2019.

[15] Nobuyuki Hirahara, Motoharu Sonogashira, and Masaaki Iiyama, "Cloud-free sea-surface-temperature image reconstruction from anomaly inpainting network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[16] Alexander Barth, Aida Alvera-Azcárate, Matjaz Licer, and Jean Marie Beckers, "Dincae 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations," *Geoscientific Model Development*, vol. 13, no. 3, pp. 1609–1622, 2020.

[17] Yang Du, Jinping He, Qiaolin Huang, Qinghong Sheng, and Guoliang Tian, "A coarse-to-fine deep generative model with spatial semantic attention for high-resolution remote sensing image inpainting," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[18] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 4471–4480.

[19] Nicolas Ballas, Li Yao, Chris Pal, and Aaron Courville, "Delving deeper into convolutional networks for learning video representations," *arXiv preprint arXiv:1511.06432*, 2015.

[20] Shunping Ji, Peiyu Dai, Meng Lu, and Yongjun Zhang, "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 732–748, 2021.