PRAT - Master2 2021-22

Super Resolution for downscaling on oceanographic fields

Fangzhou YE - supervised by Dr. Dominique Béréziat

February 14, 2022

Abstract

Sea surface temperature is a great indicator of the climate change which is largely studied by oceanographic researchers. However, the oceanic simulation on a finer scale is required because it provides more details. The downscaling process is necessary to improve the spatial resolution of the simulation on large scale. There are mainly two approaches of downscaling, the dynamical downscaling which runs the physical model on a finer scale. But it is computational expensive. Or by image super resolution technique to augment the spatial resolution. In this project, we investigate the approach by image super resolution.

Keywords downscaling, single image super resolution, PSNR, SRCNN, VDSR, cascade training

1 Introduction

Ocean physics research is of great significance to understand the Earth climate. In which the sea surface temperature (SST) is a strong indicator of global climate change. Current oceanographic research requires the simulation on fine scale (3-5km), which could better represent the details for further studies. However, the current oceanic simulation runs at a global scale (over 10km). At this scale, the spatial resolution of the simulation is too coarse. Many details of the region of interest are lost. Therefore, the appropriate downscaling process is essential to address this problem.

Downscaling on oceanographic fields refers to the process of taking global information on oceanic simulation, and translating it to a finer spatial scale that is more meaningful in the context of local impacts. After the downscaling procedure, each pixel of the downscaled oceanic simulation image represents a smaller distance. Therefore, in the same 2D local region, the downscaled oceanic simulation has more pixels than the original one, resulting in a finer spatial resolution. Image super resolution (SR) aims at recovering the corresponding high resolution (HR) images from the low resolution (LR) images. To this extent, downscaling on oceanographic fields is conceptually similar to the super resolution task in the computer vision community. In this project, we would address the downscaling issue on oceanographic SST fields in an image super resolution perspective.

2 Related Work

Nowadays, deep learning based models have shown the state-of-the-art performance in image super resolution. Image super resolution is an important task to enhance the image quality where we try to reconstruct the missing information of the low resolution image. But, it is hard to find the ground truth to evaluate the reconstruction performance. So in practice, we start with a given high resolution image, we get its low resolution counterpart by a down-sampling procedure. By applying the super resolution algorithm, we compare the reconstructed image with original high resolution image by some evaluation metrics. In this section, some classical methods and the evaluation metrics are reviewed.

2.1 Interpolation

Interpolation is a kind of simple and fast upsampling method. The missing pixel prediction is based on the existing pixels. Classical methods include nearest-neighbor, bilinear and bicubic[4] interpolation. However, The interpolation methods improve the image resolution only based on the current signal. The degraded information could not be reconstructed. The results are thus always noise-amplified and blurred.

2.2 PSNR-oriented models

In image super resolution, PSNR is a metric to evaluate the pixel wise difference between the reconstructed image and the original high resolution image. PSNR-oriented models learns the mapping between the low resolution image the the high resolution image. These models optimize the parameters by minimizing the MSE loss. They differs mainly on network architecture design and the training strategy, by increasing the network depth to widen the receptive field and by sharing the network weights to accelerate the training. [1][5][6][10][2][9][8][11][14]

2.3 Perceptually-driven models

The aforementioned CNN based models tend to have a high PSNR score since they optimize the parameters by minimizing MSE loss. However, the reconstructed image could lose high frequency details and be perceptually unsatisfying. The results turn out to be oversmooth and blurred. Inspired by generative adversarial network[3], GAN based models ([7][12]) in super resolution are proposed which yield the results more perceptually satisfying. The generator usually takes a PSNR-oriented model to predict the reconstructed image, the discriminator usually optimize by minimizing the combined perceptual loss and content loss to distinguish the reconstructed and high resolution images. In this manner, the discriminator pushes the generator to predict the reconstructed image more visually satisfied.

2.4 Evaluation metrics

PSNR is one of the most commonly used measurements. It is based on the maximum pixel value and the mean squared error (MSE). A little MSE value leads to a high PSNR score. So the neural network optimized by MSE loss usually yields a high PSNR score. However, PSNR only evaluates the pixel-wise difference. It could not reflect the perceptive difference that the human visual system performs. SSIM[13] is proposed to measure the structural similarity between images, based on independent comparisons in terms of luminance, contrast and structures. MOS refers to mean opinion score. It is a subjective evaluation metric where the human raters assign perceptive scores to images. This method could directly measure the perceptual quality but is time-consuming.

3 Methods

3.1 Data pre-processing

The dataset we used is NATL60 in NetCDF format, consists of the simulation of a particular region of the North Atlantic Ocean on 3734 days. The simulation of each day contains the field of SST, SSH, salinity and velocity. We only extracted SST and SSH fields for the experiments. Both land and sea are present in the field. To maintain uniformly, for each image we subtract the minimum land value for all pixels. All the pixel values are then normalized to $[0, 1]$. To avoid the bias of the land values, for each image we crop 32 patches containing exclusively the sea values of size $90 * 90$.

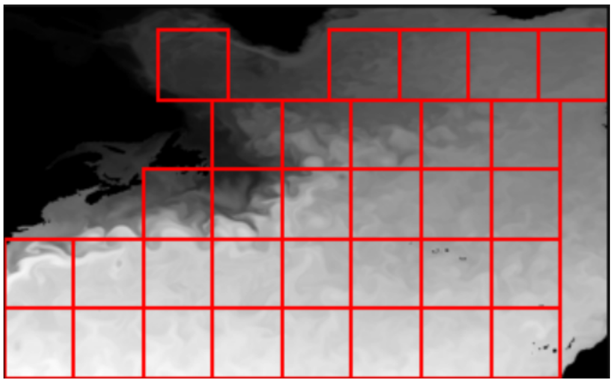

Figure 1: patches

3.2 Models

In this project, the SRCNN[1] and VDSR[5] models are implemented. The cascade training strategy is further applied. SRCNN is a simple model which contains only three convolutional layers and two activation layers. It learns the mapping from bicubic upsampled low resolution image to the high resolution ground truth. The primary function of the first, second and third layer are patch extraction and representation, non-linear mapping and reconstruction respectively. SRCNN is a shallow network which can not better extract the contexual information. We therefore investigate a deeper model.

VDSR largely increases the depth of the network to 20 convolutional layers each non-linearly mapped to the next. The contextual information over large image regions is better exploited. Besides, VDSR model only learns the residual image between the low resolution image and high resolution one. The predicted residuals are added back to low resolution patch to reconstruct the high resolution image. The residual learning leads to a faster convergence.

We also propose the cascade training paradigm in super resolution with a large scaling factor, by adding an intermediate ground truth image. Training can be divided into two stages and possibly with different models. As shown in figure 2. The $loss_1$ measures the difference between the output of $model_1$ and the intermediate resolution ground truth. Then, they will be interpolated to the size of high resolution patch and fed as input of $model_2$. The $loss_2$ measures the difference between the output of $model_2$ and the high resolution ground truth. The total loss of the network is the weighted sum of $loss_1$ and $loss_2$. The total loss is then back-propagated to optimize the parameters of each model.

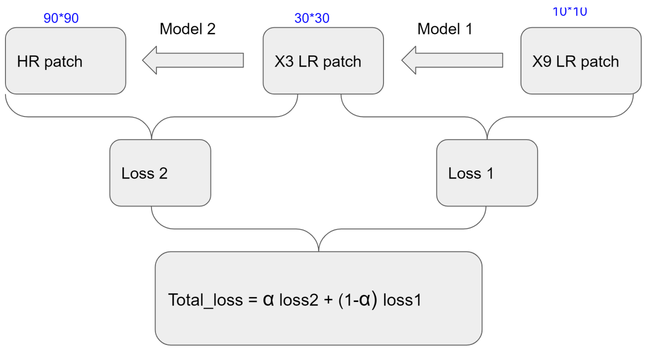

Figure 2: cascade training

3.3 Evaluation

In image super resolution for natural image. The most commonly used evaluation metrics are those who measures the perceptual reconstruction quality, like SSIM or MOS. However, they are not necessary in our scenario of oceanographic image. The prediction with high perceptual quality tends to be those which are generated with some fake high frequency details. So in our scenario, we choose to use exclusively PSNR metric to measure the reconstruction quality.

4 Experiments

4.1 Scaling factor of 3

We first perform the super resolution by scaling factor of 3. We train the SRCNN and VDSR model with the same training settings. The results show that both of them outperform the bicubic interpolation. And the deeper network tends to perform better.

Model	Average PSNR on test set
bicubic	40.94
SRCNN	46.81
VDSR	48.85

4.2 Scaling factor of 9

Then we investigate the cascade training strategy for super resolution by scaling factor of 9. α is a hyper parameter, representing the weight of second stage loss in the global network loss. We have trained for $\alpha = 0.2, 0.5, 0.8$ separately, with both SRCNN models in two stages of training. We found that cascading two SRCNN models improve the overall performance compared to a stand alone SRCNN model. The first stage loss tends to have a bigger impact on the overall parameters optimization.

Model	Average PSNR on test set
bicubic	33.50
SRCNN	37.62
Cascade 0.2	38.12
Cascade 0.5	38.10
Cascade 0.8	37.99

5 Conclusion

In this project, We leverage the deep learning based super resolution models to address the problem of downscaling in oceanographic field. We found that deeper network architecture and cascade learning can both improve the reconstruction result. The current shortcoming is that both SRCNN and VDSR model has a pre-upsampling stage to resize the low resolution image to high resolution image. The upsampling step is time consuming and blurs the original images. Some details are therefore lost. As a result, these models learn the mapping from interpolated low resolution images to high resolution images instead of the direct mapping from original low resolution images to high resolution images. To this extent, more complex model is worth to be investigated. Besides, different models in cascading learning is also worth be tested in the future.

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[2] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[4] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

[5] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[6] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

[7] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative

adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[8] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[9] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[10] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017.

[11] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE international conference on computer vision*, pages 4799–4807, 2017.

[12] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[13] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[14] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.