

COMPARISON OF MULTI-IMAGES DEEP LEARNING SUPER RESOLUTION FOR PASSIVE MICROWAVE IMAGES OF ARCTIC SEA ICE

X. Shen ^{1,2}, X. Liu ^{1,2}, Y. Yao ^{1,2}, T. Feng ^{1*}

1 Center for Spatial Information Science and Sustainable Development Applications, Tongji University, 1239 Siping Road, Shanghai, China.

2 College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai, China

KEY WORDS: Multi-image super resolution, Sea ice, Deep learning, Passive microwave images

ABSTRACT:

The observation of Arctic sea ice is of great significance to monitoring of the polar environment, research on global climate change and application of Arctic navigation. Compared to optical imagery and SAR imagery, passive microwave images can be obtained for all-sky conditions with high time resolution. However, the spatial resolution of passive microwave images is relatively low (6.25 km - 25 km) for the observation of detailed sea ice characteristics and small-scale sea ice geographical phenomena. Therefore, in this paper, considering the suitability of different alignment and fusion strategies to the characteristics of passive microwave images of sea ice, two multi-images deep learning super-resolution (SR) algorithms, Recurrent Back-Projection Network (RBPN) and network of Temporal Group Attention (TGA), are selected to test the effects of SR technique for passive microwave images of sea ice. Both qualitative and quantitative comparisons are provided for the SR results oriented from two algorithms. Overall, the SR performance of TGA algorithm outperforms RBPN algorithm for the passive microwave images of sea ice.

1. INTRODUCTION

The observation of Arctic sea ice is of great significance to monitoring of the polar environment, research on global climate change and application of Arctic navigation (Serreze and Stroeve, 2015). Data from a variety of satellite sensors, including optical satellite images, passive microwave images, and synthetic aperture radar (SAR) images have been employed to observe polar sea ice. Although optical images, such as Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectromete (MERIS), and Advanced Very High Resolution Radiometer (AVHRR), have high temporal-spatial resolution, they are often contaminated by cloud, even no available images can be obtained due to poor atmospheric conditions (Petrou et al., 2018). SAR images, such as Sentinel-1 (Xian and Tian, 2017), have high spatial resolution, but limit to the small swath and low temporal resolution, resulting in mass data processing when producing Arctic sea ice characteristics.

Passive microwave images, such as the Special Sensor Microwave Imager (SSM/I) on the series of satellites of Defense Meteorological Satellite Program (DMSP), the Advanced Microwave Scanning Radiometer-EOS (AMSR-E) on Aqua satellite of the National Aeronautics and Space Administration (NASA) Earth Observation System (EOS), and the Advanced Microwave Scanning Radiometer 2 (AMSR2) on the Global Change Observation Mission 1st - Water "SHIZUKU" (GCOM-W1), are important data sources for Arctic sea ice observation with the advantages of wide coverage, high temporal resolution, strong surface penetration ability and all-weather work (Petrou et al., 2018). Among them, AMSR2 is one of the representative passive microwave sensors that has been observing sea ice since 2012, it has more frequency bands and relatively high spatial resolution (Han and Kim, 2018). Although AMSR2 can provide daily coverage of the entire Arctic, its typical spatial resolution of around 6.25-25 km makes it difficult to monitor small leads

and ridges, and it is prohibitively coarse for some fine-scale application, such as detailed characteristics of sea ice, and small-scale geographical phenomena (Agency and Project, 2013; Wagner et al., 2020).

Considering the high cost and limitations of increasing the resolution through "hardware", especially for large scale imaging equipment like AMSR2, signal processing methods, known as SR techniques, have become a potential way to improve resolution of images (Yue et al., 2016). SR techniques, which refers to the process of recovering high resolution (HR) images from one or sequence low-resolution (LR) images, is an important technique in computer vision and image processing. According to the input number of LR images, the SR techniques can be divided into single-image SR (SISR) and multi-image SR (MISR) (Dong et al., 2016). Compared with SISR, MISR methods have the advantage of combining spatial and temporal information from sequence images (Wang et al., 2019). Traditional MISR methods based on spatial and frequency domain, are not only unable to deal with complex motions but also have problems with huge computation (Tom and Katsaggelos, 1995; Li et al., 2001; Daithankar and Ruikar, 2020).

From 2015, MISR techniques based on deep learning (DL-MISR) have begun to be developed and been applied in natural images with good performance (Liu et al., 2020). DL-MISR uses deep neural networks to construct motion estimation and fuse the complementary information of sequence images to obtain HR images. The latest existing DL-MISR methods not only make full use of temporal information, but also deal with complex motion state. When there is a large amount of sample data to train the model, DL-MISR is more computationally efficient, saves time and cost, which is very suitable for batch data processing and has strong application value for the SR task of the same type of low-resolution images.

* Corresponding author

At the same time, the passive microwave images have the characteristics of "more and less", that is, because AMSR2 circle the earth one times a day, the amount of image data that is repeated multiple times in the same polar sea ice area is huge (Agency and Project, 2013). These data can serve as the data set for the training DL-MISR model.

Therefore, in this paper, we analysis existing DL-MISR methods. Considering the suitability of different alignment and fusion strategies to the characteristics of passive microwave images of Arctic sea ice, RBP (Harris et al., 2019) and TGA (Isobe et al., 2020) are selected and adopted to test the effectiveness of DL-MISR applied for AMSR2 images of sea ice. Both qualitative and quantitative evaluation on experimental results are compared and analysed.

2. METHODOLOGY

MISR basically consist of three modules, including alignment module, fusion module, reconstruction module as shown in Fig.1 (Liu et al., 2020). The alignment module extracts motion information from sequence frames to reference frame, which concerns on the spatial transformation applied to misaligned images. The fusion module refers to combining complementary information from the aligned images and fusing them into a feature map. Reconstruction module transforms the aggregated features to the final output image by deconvolution or subpixel convolution layers.

Alignment module and fusion module could always lead to a big swing in terms of performance and efficiency (Chan et al., 2021). Therefore, we chose DL-MISR method mainly to consider the applicability of these two modules for characteristics of passive microwave images and polar sea ice movement. The sea ice motion in the Arctic region is complex, which includes rigid

motions of drifting and rotating in different directions and speed, as well as non-rigid motions of melting, freezing and disintegration. On the other hand, there are also large motions due to both seasonal and regional impacts (Tschudi et al., 2010; Maeda et al., 2020). In addition, there are some noise and unreliable values on passive microwave images.

For alignment module, Motion estimation and motion compensation methods (MEMC) and deformable convolution methods (DCN) are two common alignment strategies. Both MEMC and DCN may be able to handle the complex sea ice motions. But DCN such as Temporally Deformable Alignment Network is difficult to train and may smooths out high-frequency details (Tian et al., 2018; Wang et al., 2019). Some early MEMC methods such as Detail-revealing deep network are easy to generate artifacts, which lead to inconsistencies between image sequences (Liao et al., 2015; Kappeler et al., 2016; Liu et al., 2017; Tao et al., 2017). Therefore, we select RBP model, a MEMC model based on optical flow. Compared with other MEMC methods, the back projection module of RBP adopts an iterative error-correcting feedback mechanism to calculate both up- and down-projection errors for minimizing the feature map error. It actually reduces artifacts and temporal inconsistencies to a certain extent (Haris et al., 2018).

For fusion module, existing DL-MISR methods fuse features through direct concatenating (Huang et al., 2015; Sajjadi et al., 2018) or temporal and spatial attention (TSA). Considering that the information provided by different adjacent frames is not equal due to occlusion or blurring, TSA calculates a weight map to each neighbouring frame, so it is better than direct concatenating. Therefore, we select the TGA model based on TSA fusion (Isobe et al., 2020), it makes full use of complementary information across frames to recover missing details for the reference frame. In addition, the strategy of group fusion is also adopted in TGA.

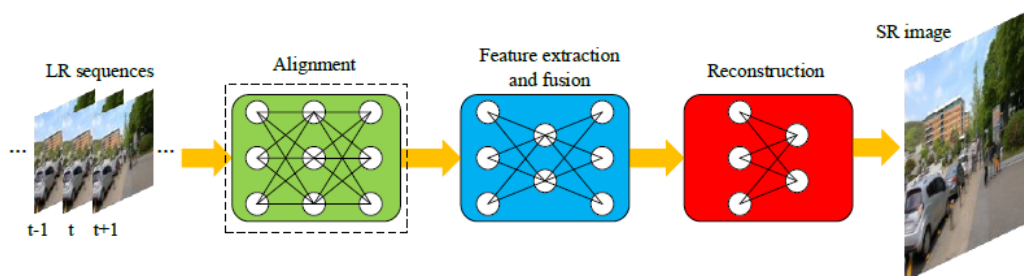


Figure 1. The general flowchart of deep learning methods for DL-MISR. (Liu et al., 2020)

2.1 RBP

The RBP network consists of three parts as Fig2: Initial feature extraction, Multiple Projections, Reconstruction. Initial feature extraction makes compensation based on optical flow. It calculates neighbor feature tensors $\{M_{t-3}; \dots; M_{t+3}\}$ using precomputed dense motion flow maps $\{F_{t-3}; \dots; F_{t+3}\}$, the reference frame $\{I_t\}$ and each neighbouring frame $\{I_{t-3}; \dots; I_{t+3}\}$. Multiple Projections extracts missing information through an

encoder-decoder structure. Encoder part includes MISR block (input is coming from the vertical red line direction) and SISR block (input is coming from the horizontal blue line direction) to obtain high feature maps $\{H_{t-3}; \dots; H_{t+3}\}$ as Fig2 shows. It also performs back-projection from L_{t-n-1} obtained by decoder part to M_{t-n} to produce the residual map. Finally, Reconstruction concatenates all HR feature maps and obtains HR images by a simple convolutional layer (Haris et al., 2019).

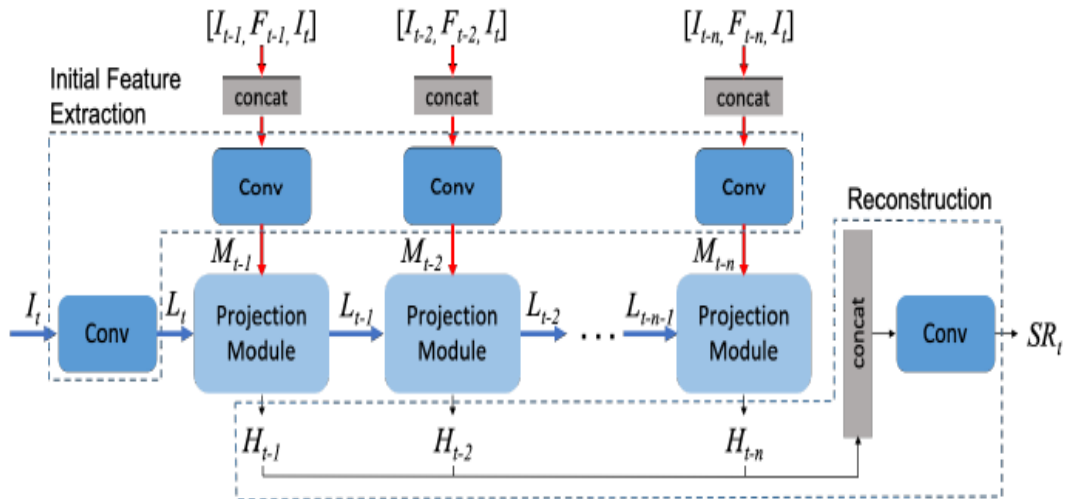


Figure 2. Sketch map of RBPN (Haris et al., 2019)

2.2 TGA

TGA adopts a fast spatial alignment method. It estimates homograph between every two consecutive frames and warps neighbouring frames to the reference frame so it can handle images sequences with large sea ice motion. Since homograph transformation is a global, TGA keeps the structure better and introduces few artifacts. Considering that the contributions of neighbouring frames in different temporal distances are not equal,

TGA designs temporal grouping. For each group, an intra-group fusion module is employed for feature extraction and fusion. Every intra-group fusion module is equipped with dilation rate to model the motion level associated with a group. To better integrate features from different groups, a temporal attention module is introduced. It works as a guidance to efficiently aggregate information across different temporal groups and produces a high-resolution residual map (Isobe et al., 2020). The final SR output is obtained through a sub-pixel convolution layer.

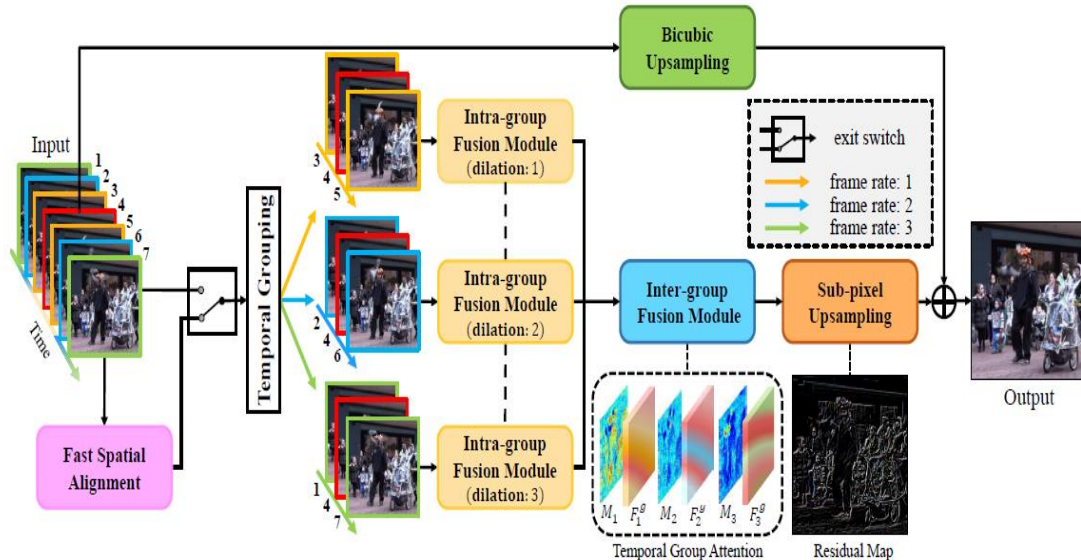


Figure 3. Sketch map of TGA (Isobe et al., 2020)

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1 DATA

The training dataset comes from the AMSR2 passive microwave sensor on the GCOM-W1 polar orbiting satellite platform in the global change observation mission of the Japanese aviation research and development agency. This series of satellites carry out long-term (10-15 years) detection of the earth and provide observational data for the study of the global water cycle and climate change mechanisms. The current observations on water are carried out by the ASMR2 mounted on the water cycle

variation observation satellite launched in 2011. AMSR2 obtains images with higher resolution than the other passive microwave sensors at more frequencies. It has the largest antenna diameter (approximately 2 meters) of the observation sensor for carrying satellites, and can achieve a high-speed rotation of 40 times per minute, sweeping along the arc of the earth's surface, and searching the earth day and night in two days (Agency and Project, 2013).

AMSR2 can capture and measure microwaves from the ground and ocean in seven frequency bands (6~89HZ). The final products of the AMSR2 have five types: Level 0, Level1B, Level1R, Level2, and Level3. In particular, the Level 1B brightness temperature swath data of horizontal polarization and vertical polarization at 89 GHz is selected because it has the highest spatial resolution (5×5 km). As Fig. 1 shows, all swath AMSR2 data of one day is gridded to obtain daily average passive microwave brightness temperature data with the polar stereographic grids of the National Snow and Ice Data Center (NSIDC) at 6.25 km as real true images. To produce LR images, we downscale the HR images four times with bicubic interpolation at 25 km. The experimental images cover 30° N to 90° N, and -180° W to 180° E, including the entire Arctic region. The images acquisition time range is from 2013 to 2016. The dataset contains a total of 33,600 image sequences, of which 30,660 are used for training and 2,940 are used for evaluation. Each image sequence consists of 7 adjacent images, and the image size is cropped to 256×256 pixels. In addition, we use data augmentation technology such as rotation, mirror, and random cropping to expand the training set to improve the generalization ability of the model.

3.2 Implementation and training details

In this experiment, we convert the format of passive microwave images from GEOTIFF to PNG, which is more suitable for network. The original brightness temperature is stored as 16bit value. In all our experiments, we adopt with a 4× sampling factor to evaluate the different methods. Before fed into networks to train models, the input data is normalized.

The hyper-parameters of RBPN are set as follows: the network uses kaiming initialization and Adam optimizer, the loss function is L1 loss per-pixel between the predicted frame and the ground truth HR frame. The learning rate is 0.0001 for all layers and decreases by a factor of 10 for half of total 150 epochs. The batch size is set to 4. The RBPN model is trained under the environment of Ubuntu 16.04 + NVIDIA GTX 1080GPUs*2 + Python3.5 + CUDA9.2 + pytorch 1.0 for two days.

TGA is supervised by pixel-wise L1 loss as well and optimized using kaiming initialization and Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Weight decay is set to 0.0005 during training. The learning rate is initially set to 0.0001 and later down-scaled by a factor of 0.1 every 10 epochs until 50 epochs. The batch size of TGA is 8. TGA model is trained under the environment of Ubuntu 16.04 + NVIDIA GTX 1080 GPUs*2 + Python3.6 + CUDA9.2 + pytorch 1.2 for one days.

3.3 Analysis of the super-resolution results

Evaluation index: the image quality evaluation indexes used in this experiment are full reference image quality evaluation indexes: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) (Daithankar and Ruikar, 2020).

The larger value of the PSNR only indicates the smaller the gap between the image to be reviewed and the reference image but also the better the image quality. The formula of PSNR is as follows:

$$PSNR = 10 \lg \frac{65535^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |X(i,j) - Y(i,j)|^2} \quad (1)$$

Where M, N = image width and height

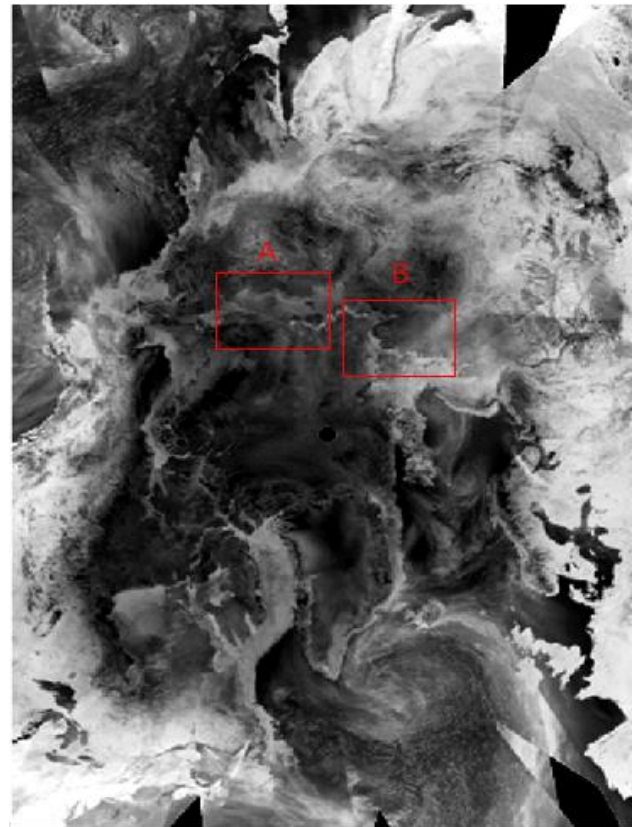


Figure 3. Daily average passive microwave brightness temperature data with the polar stereographic grids of NSIDC

$X(i,j)$ = the pixel value of original image
 $Y(i,j)$ = the pixel value of SR image

SSIM measures image similarity from brightness (equation (1)), contrast (equation (2)) and structure (equation (3)). The SSIM value range is (0,1). Higher value indicates a smaller distortion of the image. It can better reflect the subjective feelings of the human eye.

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (2)$$

$$c(X, Y) = \frac{2\mu_X\mu_Y + C_2}{\mu_X^2 + \mu_Y^2 + C_2} \quad (3)$$

$$s(X, Y) = \frac{\partial_{XY} + C_3}{\partial_X\partial_Y + C_3} \quad (4)$$

$$SSIM = l(X, Y) * c(X, Y) * s(X, Y) \quad (5)$$

Where X, Y=Original image, SR image
 μ_X, μ_Y =the mean value of image X and Y
 ∂_X, ∂_Y = the variance of image X and Y
 ∂_{XY} = the covariance of image X and Y
 $C_1, C_2, C_3 = (0.01 * 65535)^2, (0.03 * 65535)^2, C_2/2$

The quantitative comparison is given in Tab. 1. TGA algorithm has a better performance on both evaluation indexes especially in terms of SSIM. In addition, the results of vertical polarization image are slightly better than that of horizontal polarization. Fig. 4 shows two subregions of passive microwave images of Arctic sea ice including LR images, SR images of two model and HR images. Sub-region A in Fig. 3 mainly includes different sea ice texture, and sub-region B includes sea ice and sea water.

It's obvious that there are finer boundaries and more detailed texture after DL-MISR. It seems that TGA generates more detailed edge while SR results is smoother. But RBPN maintains the consistency between LR and SR better than TGA. In addition, both two DL-MISR introduces less noise.

DATA	RBPN		VSR-TGA	
	PSNR	SSIM	PSNR	SSIM
89 GHz-H	43.5316	0.9463	43.7453	0.9673
89 GHz-V	44.4200	0.9513	44.7532	0.9723

Table 1. Quantitative evaluation of RBPN and VSR-TGA SR results.

4. CONCLUSION

Considering the suitability of different alignment and fusion strategies to the characteristics of passive microwave images of sea ice, two DL-MISR methods, namely RBPN and TGA, are applied to the passive microwave images of sea ice, which can improve the spatial resolution of passive microwave images and obtain finer boundaries and more details. The TGA not only performs better on quantitative evaluation but also generates finer boundaries and more detailed texture than RBPN. However, RBPN keeps the consistency between LR and SR well. Overall, TGA might be more suitable method to improve the resolution of passive microwave images to observe Arctic sea ice. In the further, we will optimize the model according to the characteristics of sea ice motions and the applicable scene of different strategies of alignment and fusion module. In addition, we will make more detailed analysis and evaluation of SR process and experimental results.

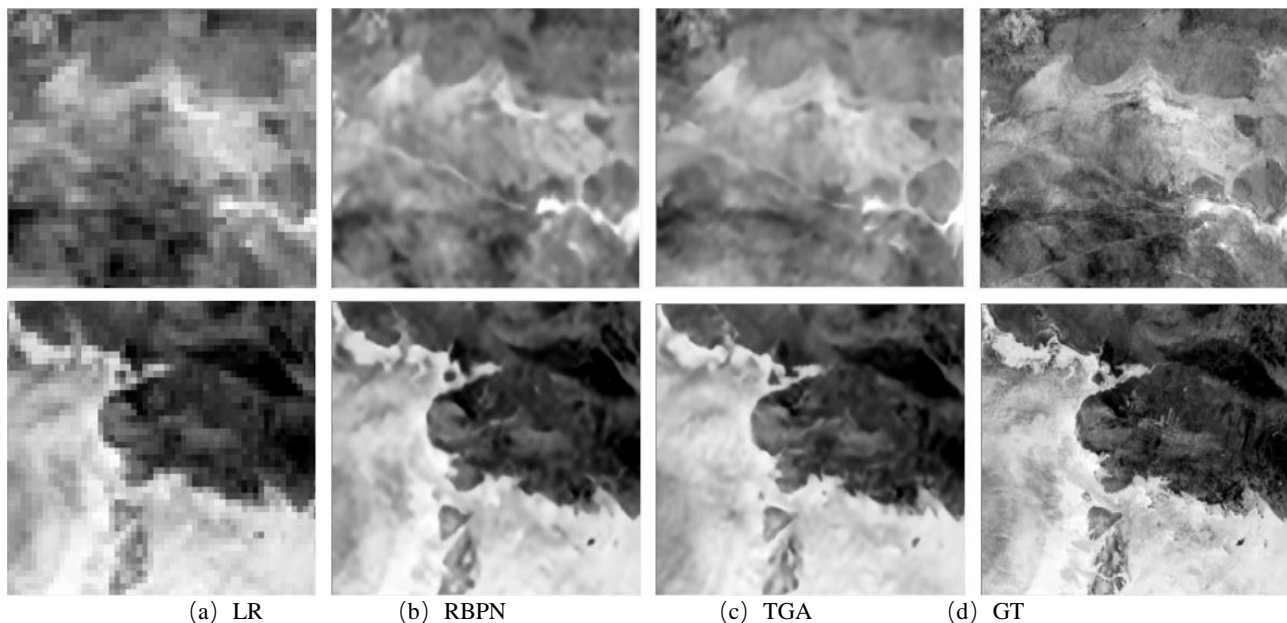


Figure 4. The top and bottom lines are SR results of subregion A and B in Fig. 3. (a) original LR images, (b) SR results of RBPN, (c) SR results of TGA, (d) Ground truth

ACKNOWLEDGEMENTS

The authors would like to acknowledge the use of AMSR2 data from <http://www.jaxa.jp/>. This work was supported by the National Key Research and Development Program of China (No. 2017YFA0603100), the National Science Foundation of China (No. 41801335).

REFERENCES

Agency, J. A. E., Project, T. O. G., 2013. GCOM-W1 "SHIZUKU" Data Users Handbook.

Chan, C. K., Wang, X., Yu, K., Dong, C., Loy, C., 2021. BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

Daithankar, M., Ruikar, D. S., 2020. Video Super Resolution: A Review. 10.1007/978-981-15-1420-3_51.

Dong, C., Loy, C., He, K., Tang, X., 2016. Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 295-307.

Haris, M., Shakhnarovich, G., Ukita, N., 2019. Recurrent back-projection network for video super-resolution. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3892–3901.

Haris, M., Shakhnarovich, G., Ukita, N., 2018. Deep back projection networks for super-resolution. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1644-1673.

Han, H., Kim, H., 2018. Evaluation of summer passive microwave sea ice concentrations in the Chukchi Sea based on

- KOMPSAT-5 SAR and numerical weather prediction data. *Remote Sensing of Environment*, 209, 343-362.
- Huang, Y., Wang, W., Wang, L., 2015. Bidirectional Recurrent Convolutional Networks for Multi-Frame Super-Resolution. *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS)*, 235–243.
- Isobe, T., Li, S., Jia, Y., 2020. Video Super-resolution with Temporal Group Attention. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8008–8017.
- Kappepler, A., Yoo, S., Dai, Q., Katsaggelos, K., Aggelos., 2016b. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2), 109–122.
- Li, J., Seigniny, L., Valin, P., 2001. Wavelet domain super-resolution reconstruction of infrared image sequences. *Proceedings of SPIE*, 4385(1), 108-116.
- Liao, R., Tao, X., Li, R., Ma, Z., Jia, J., 2015. Video super-resolution via deep draft-ensemble learning. *IEEE International Conference on Computer Vision (ICCV)*, 531–539.
- Liu, D., Wang, Z., Fan, Y., Liu, X., Wang, Z., Chang, S., Huang, T., 2017. Robust video super-resolution with learned temporal dynamics. *IEEE International Conference on Computer Vision (ICCV)*, 2526–2534.
- Li, D., Wang, Z., 2017. Video Super-resolution via Motion Compensation and Deep Residual Learning. *IEEE Transactions on Computational Imaging*, 3(4), 749-762.
- Liu, H., Ruan, Z., Zhao, P., Shang, F., 2020. Video Super Resolution Based on Deep Learning: A comprehensive survey. DOI: arXiv:2007.12928v1[cs.CV], 25, Jul, [Online]. Available: <https://arxiv.org/pdf/2007.12928.pdf>.
- Maeda, K., Kimura, N., Yamaguchi, H., 2020. Temporal and spatial change in the relationship between sea-ice motion and wind in the Arctic. *Polar Research*, 39, 3370.
- Petrou, I. Z., Xian, Y., Tian, Y., 2018. Towards breaking the spatial resolution barriers: An optical flow and super-resolution approach for sea ice motion estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138, 164-175.
- Sajjadi, M. S. M., Vemulapalli, R., Brown, M., 2018. Frame-recurrent video super resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6626–6634.
- Serreze, M., Stroeve, J., 2015. Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2045), 20140159.
- Tao, X., Gao, H., Liao, R., Wang, J., Jia, J., 2017. Detail-Revealing Deep Video Super-Resolution. *IEEE International Conference on Computer Vision (ICCV)*, 4482-4490.
- Tom, C., Katsaggelos, A. K., 1995. Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. *IEEE Int. Conf. Image Processing*, Washington, DC, 539-542.
- Tian, Y., Zhang, Y., Fu, Y., Xu, C., 2018. TDAN: Temporally Deformable Alignment Network for Video Super-Resolution. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3361-3369.
- Tschudi, M., Fowler, C., Maslanik, J., Stroeve, J., 2010. Tracking the Movement and Changing Surface Characteristics of Arctic Sea Ice. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(4), 536-540.
- Wang, X., Chan, C. K., Yu, K., Dong, C., Loy, Change, C., 2019. EDVR: Video restoration with enhanced deformable convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1954–1963.
- Wang, Z., Chen, J., Hoi, S., 2020. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Wang, H., Su, D., Liu, C., Jin, L., Sun, X., Peng, X., 2019. Deformable non-local network for video super resolution. *IEEE Access*, 7, 177734–177744.
- Wagner, P., Hughes, N., Bourbonnais, P., Stroeve, J., Rabenstein, U., Bhatt, U., Little, U., Wiggins, H., Fleming, A., 2020. Sea ice information and forecast needs for industry maritime stakeholders. *Polar Geography*, 43(2-3), 160-187.
- Xian, Y., Tian, Y., 2017. Super-Resolved Fine-Scale Sea Ice Motion Tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5427-5439.
- Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L., 2016. Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128, 389-408.