

Remote sensing image colorization based on Joint Stream Deep Convolutional Generative Adversarial Networks

Jingyu Wang
Ocean University of China
Qingdao, China
wangjingyu3186@stu.ouc.edu.cn

Jie Nie*
niejie@ouc.edu.cn
Ocean University of China
Qingdao, China

Hao Chen
Ocean University of China
Qingdao, China
chenhao0000@stu.ouc.edu.cn

Huaxin Xie
Ocean University of China
Qingdao, China
xiehuaxin@stu.ouc.edu.cn

Chengyu Zheng
Ocean University of China
Qingdao, China
zhengchengyu@stu.ouc.edu.cn

Min Ye
Ocean University of China
Qingdao, China
yemin@stu.ouc.edu.cn

Zhiqiang Wei
Ocean University of China
Qingdao, China
weizhiqiang@ouc.edu.cn

ABSTRACT

With the development of deep neural networks, especially generation networks, gray image coloring technology has made great progress. As one of the fields, remote sensing image colorization needs to be solved urgently. This is because remote sensing images cannot obtain clear color images due to the limitations of shooting equipment and transmission equipment. Compared with ordinary images, remote sensing images are characterized by the uneven spatial distribution of objects, therefore, it is a great challenge to ensure the spatial consistency of coloring. To embrace this challenge, we propose a new joint stream DCGAN including a micro stream and a macro stream, in which the latter is set as a prior to constrain the former for colorization. In addition, the Low-level Correlation Feature Extraction (LCFE) module is proposed to obtain the salient shallow detail feature with global correlation, which is used to enhance the global constraints as well as supplement the low-level information to the micro stream. What's more, we propose the Gated Selection (GSM) module by selecting useful information using a gated scheme to fuse features from two streams appropriately. Comprehensive comparison and ablation experiments are implemented and verify the proposed method performs surpasses other methods in both qualitative and quantitative metrics.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision problems.**

*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MMAsia '22, December 13–16, 2022, Tokyo, Japan

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9478-9/22/12...\$15.00

<https://doi.org/10.1145/3551626.3564951>

KEYWORDS

Remote sensing image colorization, multi-scale, DCGAN, U-Net, adaptive fusion

ACM Reference Format:

Jingyu Wang, Jie Nie, Hao Chen, Huaxin Xie, Chengyu Zheng, Min Ye, and Zhiqiang Wei. 2022. Remote sensing image colorization based on Joint Stream Deep Convolutional Generative Adversarial Networks. In *ACM Multimedia Asia (MMAsia '22), December 13–16, 2022, Tokyo, Japan*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3551626.3564951>

1 INTRODUCTION

Remote sensing images are regarded as the essential material for people to explore and understand the earth process. The natural pseudo color image is regarded as one of the critical approaches for morphology understanding correspond to human perception. However, due to the limitations of shooting means and sensing equipment, as well as the constraints of natural phenomena such as haze, it is difficult to obtain high-quality remote sensing images with full colors. While the exploration and deep understanding of earth's processes urgently require high-contrast color images that match human perception. Therefore, the colorization task of grayscale remote sensing image is imminent.

Image colorization approaches could be divided into two categories [37]: user-based colorization and fully automatic image colorization. User-based colorization requires human interaction to achieve colorization, therefore, the quality of the images generated by these algorithms depends on the rationality of the cues given by the human, which makes them labor-intensive. While the automatic coloring method realizes end-to-end colorization by learning the relationship between input data and corresponding color images without any human involvement. The development of the depth generation model provides an effective fully automatic technique for common image colorization, especially a series of generative adversarial models (GANs) [3, 10, 21, 24, 35]. For the ordinary images colorization, Isola et al. [12] proposed a coloring method based on the Conditional Generative Adversarial Network (CGAN) [21] which does not need to define the unique loss function

according to the unique problem. On this basis, Nazeri et al.[22] proposed a redefined loss function, which takes the probability of the maximization discriminator error instead of the correct probability of the minimization generator to resolve the problem of the colorization process fluctuation and accelerate network convergence. However, different from ordinary images, the global structure of remote sensing images is according to the physical earth surface with immensely imbalanced object distribution. For instance, there always exist micro-objects such as buildings that emerge in a large range of continuous texture regions. Under this condition, if we focus on learning the distribution of local pixels, it will deteriorate the space inconsistency problem in the whole image. For this problem, Limmer et al.[19] proposed a multi-scale pyramid structure infrared colorization method based on CNN. Li et al. [18] solved this problem by proposing multi-scale discriminators and they set up a discriminator for each layer feature in the generator's decoding process. Although the multi-scale discriminators optimizes the measurement of Jensen-Shannon divergence [9], it could not supply a strong constraint to macro scale space stability. Wu et al.[31] proposed a multi-scale generator by using multiple convolutions with different kernel sizes to realize colorization. After that, Wu et al.[32] transferred the colorization task from RGB to YUV color space and used the multi-scale convolution kernels to improve the coloring effect. Although the above approaches can achieve high-quality colorization, there are three problems in the remote sensing colorization task: (1) The missing low-level information are not adequately supplemented in the deep layer ; (2) The macro context constraint is weak which cannot guarantee coloring spatial consistency of remote sensing images with an imbalanced spatial distribution. (3) The effect of the simple fusion method is poor since it cannot choose useful information in massive redundant information from the different scales.

In response to the above problems, we propose a novel joint stream DCGAN to realize remote sensing images colorization with high space consistency. The LCFE module is proposed to obtain the salient low-level feature with global correlation. Then, the obtained feature through the LCFE module is fused with the deep layer feature of the micro stream to enhance the context constraints as well as to supplement their missing low-level information during the downsampling operation. For the fusion of the two streams, we propose the GSM module by estimating the usefulness of each feature vector pixel-wise to select significant information and avoid them drowning in the massive useless information. To evaluate the performance of our method, contrast and ablation experiments are conducted on AID Data Set and NWPU Data Set. The qualitative and quantitative indicators demonstrate the effectiveness of the proposed architecture, which indicates our model has the capability of reducing the abrupt pixels and improving the stability and smoothness of the colored remote sensing images. The contributions of our research are as follows:

- We propose a novel joint stream DCGAN introducing macro scale which constraints micro scale to ensure the high space consistency as well as the object visibility of the colored remote sensing images.
- We propose the LCFE module to obtain the salient low-level feature with the global correlation which is supplemented

to the micro stream to supplement shallow information and enhance context constraints.

- We propose the GSM module by gated selection scheme to choose effective information to ensure rational fusion of information from multi-scale streams.

2 RELATED WORK

The two main categories of image colorization approaches [2, 20, 29, 37] are user-based colorization and fully automatic image colorization. The difference between the two is that the user-based approach, including scribbling-based colorization[7, 8, 11, 25, 26], exemplar-based colorization [15–17, 23, 28, 34] and colorization based on language and text[4, 13, 36], requires human involvement while the automatic approach does not.

Due to the complex spatial distribution of remote sensing images, the user-guided coloring method is difficult to apply, in contrast, the automatic coloring method is widely popular. Isola et al.[12] proposed a coloring method based on Conditional Generative Adversarial Network (CGAN) which optimizes the network without the defined unique loss function but the game between the generator and the discriminator. They chose U-Net architecture [27] as the generator to mining the rules of generation and they propose PatchGAN as the discriminator which divide the input into patches for discrimination to deal with the high-frequency part of the image. Nazeri et al. [22] proposed a colorization method for high-resolution images. They used the redefined loss function, which takes the probability of the maximization discriminator error replace the correct probability of the minimization generator focus on resolving the problem of the unstable colorization process and expedite network convergence.

Different from the above single scale network, Limmer et al. [19] proposed a infrared colorization method using the multi-scale pyramid structure based on CNN and they supplemented the details of the input image by simple addition operation to complete post-processing. According to the pyramid structure, Li et al. [18] proposed a GAN with multi-discriminators to achieve colorization of high-resolution remote sensing images. They chose U-Net as the generator of their method and they refer to the idea of pyramid structure and input the features proposed by each layer of generator into the discriminator for judgment. What' more, the features of the local layer are added to the next layer of discriminator in order to make the discriminator fuse the feature from different levels. This method promotes the discriminator to better guide the generator, so as to increase the stability of the generated color space and produce a better coloring effect. Wu et al.[32] realized the remote sensing images colorization by multi-scale feature extraction using the convolution kernels with different size.

Although the above research has improved the coloring level to a certain extent, there exist shortcomings in the specific field of remote sensing images colorization. Most of them cannot guarantee consistency of color space and visibility of objects. In contrast to previous research in remote sensing images colorization, we propose a novel Joint Stream DCGAN.

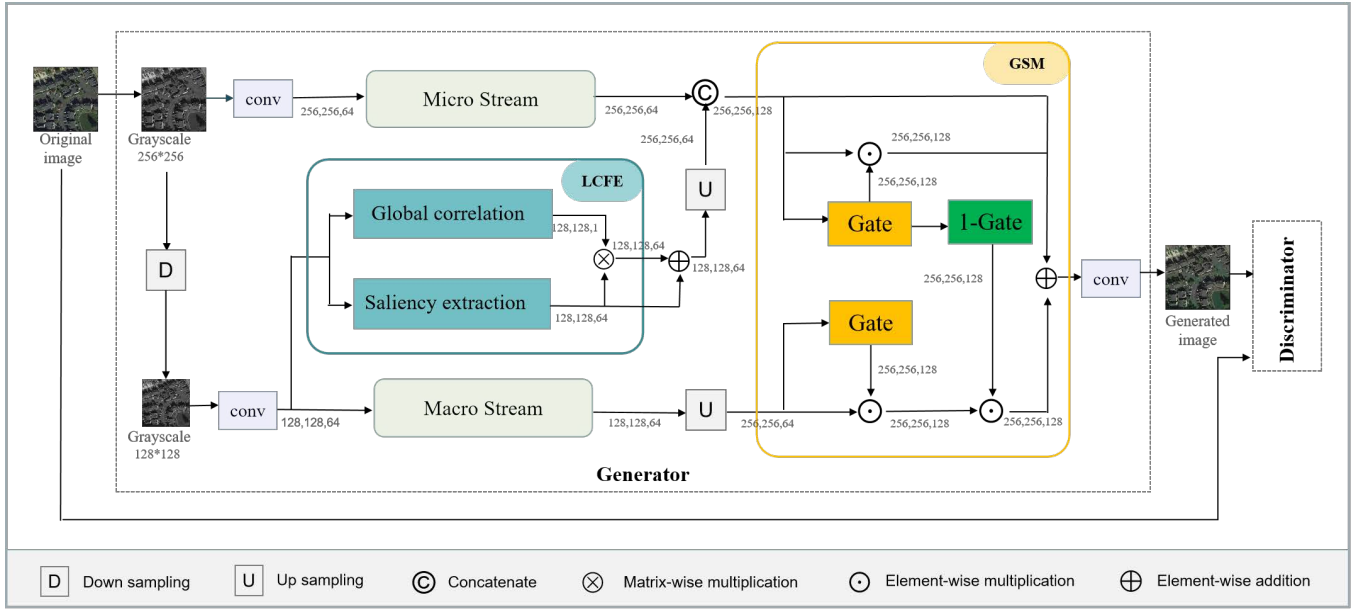


Figure 1: Overview of our Joint Stream DCGAN which consists of a micro stream, a macro stream, and two main building blocks: a Low-level Correlation Information Extraction (LCFE) module and a Gated Selection (GSM) module. The LCFE module is proposed to obtain the salient low-level feature with global correlation, which is upsampled and then concatenated with the micro stream deep layer feature to supplement the missing shallow information while strengthening the global context constraints on the micro stream. The GSM module fuses two streams by a gated scheme which selects the useful information to take advantage of the two streams properly.

3 METHOD

We propose a novel DCGAN with a joint stream generator[24] as illustrated in Fig. 1. The two streams, i.e, micro stream and macro stream, use U-Net as the backbone and the inputs of them are the original grayscale and the down-sampled grayscale obtained by average pooling operation respectively. It is precisely due to the larger receptive field that the macro stream can capture more contextual information and pay more attention to global representation. In addition, the macro stream can act as a constraint condition on the micro stream to ensure the color space consistency. To fully complement the low-level information lost in the deep layers and enhance global context information, we propose the LCIE module to obtain the extra shallow information with global correlation. Then, the obtained feature is fused with deep layer feature of the micro stream. After that, the obtained features from two streams are input into the GSM module to realize the fusion of two streams by selecting useful information using gated scheme and the color images are generated through this module. Finally, input the generated and the corresponding real images to the discriminator together to calculate the probability that the images are real. The overall network structure is shown in Fig. 1.

3.1 LCFE module

To solve the problem that the low-level information missing in the deep layer of the micro stream caused by downsampling operation and further enhance the spatial consistency, we propose the LCFE

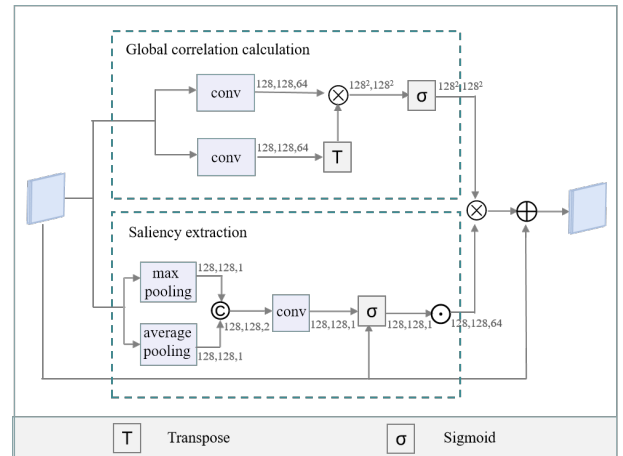


Figure 2: The structure of the Low-level Correlation Feature Extraction (LCFE) module

module to obtain the salient low-level feature with the global correlation which worked as the supplement to the deep layer of the micro stream. We set the feature obtained by the first convolutional layer of the macro stream as the LCFE module's input, which is processed by extracting the salient low-level information and calculating the global correlation, and then fusion them by matrix-wise multiplication to obtain the output feature, which is concatenate

with the micro stream last layer feature to supplement their missing low-level information in the deep layer during the downsampling operation as well as enhance the global context constraints.

The details of the LCFE module are shown in Fig. 2. On the one hand, the input features are performed by two convolutions with different 1×1 kernels to obtain two features and multiply them matrix-wise to obtain the feature containing the global dependencies between elements. Process the obtained feature by the Sigmoid function to generate the global correlation weight feature. On the other hand, the low-level feature highlighting the local details is obtained by the spatial attention scheme[30]. Finally, multiply the output of the two parts matrix-wise and the salient low-level feature with the global context information can be obtained. The formula of the operation is follows:

$$\begin{aligned} F_c &= conv_{1 \times 1}(F_i) \times [conv'_{1 \times 1}(F_i)]^T, \\ F_s &= F_i \cdot (\sigma(\max(F_i) | avg(F_i))), \\ F_o &= F_i + \sigma\left(\frac{F_c}{\sqrt{d_k}}\right) \times F_s \end{aligned} \quad (1)$$

where, F_i and F_o denote the input and the output of the LCFE module; $conv, max$ and avg represent convolution, max pooling and average pooling operation respectively, and σ represents softmax operation.

Upsampling the output feature and then concatenating it with the last layer feature of the micro stream to complement the rich low-level information and strong context constraints.

3.2 GSM module

The simply combining method, such as concatenation, will hide the effective information in massive ineffective information making the features from two streams cannot be used reasonably. Therefore, an efficient fusion approach is demanded to selectively collect useful information from different features coming from the two streams. Focus on this, we propose the GSM module, as shown in Fig.1, by using a gated scheme to select the useful information in the massive redundant information to fuse the two streams. Through the gating mechanism, the validity of each feature vector is measured and controlled to decide whether propagate or not, by which the GSM module can realize the efficient use of the information from the two streams.

The input of the GSM module is the supplemented micro stream output feature and the upsampled macro stream output feature. The formula of the operation is follows:

$$X = (1 + G_i) \cdot X_i + (1 - G_i) \cdot G_a \cdot X_a \quad (2)$$

where X_i and X_a denote the input from micro branch and macro branch respectively; G_i and G_a denote a gated selection for the X_i and X_a using the softmax [5] function; and the ‘ \cdot ’ represents the element-wise multiplication.

Through our proposed dual-gated scheme to filter useful information while suppressing redundant noise, an efficient joint stream fusion is achieved.

3.3 Objective Function

According to the objective function of GAN, and for our joint stream network, we propose the objective functions of the generator and the discriminator respectively, which are shown in equ (3) and equ (4). In addition to the loss of the GAN, the objective function of the generator we proposed involves the traditional loss.

$$\begin{aligned} &\min_{\theta_{G_1}, \theta_{G_2}} J^{(G_1, G_2)}(\theta_D, \theta_{G_1}, \theta_{G_2}) \\ &= \min_{\theta_{G_1}, \theta_{G_2}} (-\mathbb{E}_{x_1 \sim P(x_1), x_2 \sim P(x_2)} [\log(D(G(x_1, x_2)))] \\ &\quad + \lambda \|G(x_1, x_2) - y\|_1) \end{aligned} \quad (3)$$

$$\begin{aligned} &\max_{\theta_D} J^{(D)}(\theta_D, \theta_{G_1}, \theta_{G_2}) = \max_{\theta_D} (E_y [\log(D(y|x_1, x_2))] \\ &\quad + E_{x_1 \sim P(x_1), x_2 \sim P(x_2)} [\log(1 - D(G(x_1, x_2)|x_1, x_2))]) \end{aligned} \quad (4)$$

where G_1 and G_2 represent the micro stream and macro stream; D denotes the discriminator; x_1 and x_2 represent the input of the micro stream and macro stream; y denotes the groundtruth; λ is a hyperparameter set to 100. The generator and discriminator are trained alternately to realize network optimization.

4 EXPERIMENTS

4.1 Dataset and implementation details

The latest two Data sets of remote sensing images are AID Data Set[33] and NWPU Data Set[6]. NWPU-resisc45 data Set is a common remote sensing data set created by Northwestern Polytechnical University that work as a remote sensing image classification benchmark. The dataset consisted of 31,500 images, composed of 45 different scene categories. The size of the images in this dataset is 256×256 . In addition, AID Dataset is a remote sensing image Dataset jointly published by Huazhong University of Science and Technology and Wuhan University. The dataset contains images from 30 different scene categories, with 220 to 420 images in each category, for a total of 10,000 images. For the above two data sets, we divided each class into training sets, test sets, and validation sets in a ratio of 3:1:1.

4.2 Training Details

We train our approach on NVIDIA Tesla V100-SXM2-16GB GPU using the TensorFlow framework [1]. We train 30 epochs to choose the best result and set 8 as the batch size. We use the Adam algorithm [14] with a learning rate of 0.003. The super parameter λ is set to 100 to control traditional loss.

4.3 Evaluation Metrics

We set accuracy and Amazon Mechanical Turk (AMT)[12] as the metrics to evaluate the experimental results. Accuracy represents the degree of similarity between the generated pixels with the corresponding pixels of the original image. That is a quantitative metric and the higher accuracy is, the closer the generated is to the original image. AMT is a qualitative metric and a high AMT score means the image is more in line with human perception.

Table 1: Contrast experiments results on AID Data Set

Method	acc2	acc5	AMT
U-Net[27]	22.89	78.82	36.67
PatchGAN[12]	23.07	78.04	40.46
DCGAN[22]	23.42	77.36	48.03
Multi-D[18]	24.15	77.18	62.11
Multi-G[32]	26.53	76.69	67.75
Ours	28.64	82.02	79.25

Table 2: Contrast experiments results on NWPU Data Set

Method	acc2	acc5	AMT
U-Net[27]	36.05	88.42	48.29
PatchGAN[12]	37.56	86.61	56.15
DCGAN[22]	37.02	88.05	59.37
Multi-D[18]	38.30	87.12	70.59
Multi-G[32]	37.91	86.81	74.63
Ours	40.28	89.06	82.79

Table 3: The ablation experiments results on AID Data Set

Method	acc2	acc5	AMT
Single stream	23.42	77.36	48.02
Joint stream	25.14	79.68	61.92
J-S+Concat (w/o LCFE)	25.73	80.27	66.48
J-S+LCFE	26.59	81.20	72.34
J-S+LCFE+GSM	28.64	82.09	79.29

Accuracy. The accuracy is obtained according to the proportion of correctly colored pixels to total pixels. And we defined the difference between the generated values in all three channels at the same position and the corresponding elements of the original image within the specified range as accurate coloring. The formula is as follows:

$$acc(x, y) = \frac{1}{n} \sum_{p=1}^n \prod_{l=1}^3 1_{[0, \varepsilon_l]}(|h(x)^{(p,l)} - y^{(p,l)}|) \quad (5)$$

In equ (5), ε represents the threshold and we choose 2 and 5 as the threshold to obtain the accuracy of different degrees of similarity.

Amazon Mechanical Turk(AMT) Perception Test. Since the most important task of remote sensing image color is to generate color images that meet human perception needs and obtaining a highly subjective evaluation of color and color naturalness is also an important standard to test the generation effect, we also adopt Amazon Mechanical Turkey (AMT) perception test to obtain the qualitative indicators.

We selected 20 people to participate in the AMT Perception test. We will present a grayscale image and the corresponding color image generated by different approaches. For each result, participants were required to grade its color naturalness between 1 to 100, and the higher the score, the more naturalness it seems for people. Each participant needs to judge 20 different images generated by each network, and the scores of each participant will get an average score. Take the sum of everyone's average scores and divide by 20 to get the final AMT score.

Table 4: The ablation experiments results on NWPU Data Set

Method	acc2	acc5	AMT
Single stream	37.02	88.05	59.59
Joint stream	38.84	88.23	66.50
J-S+Concat (o/w LCFE)	39.21	88.58	73.37
J-S+LCFE	39.72	88.74	76.08
J-S+LCFE+GSM	40.28	89.06	82.92

4.4 Comparative Networks

- Ronneberger et al.'s Method[27]: Use U-Net structure that extracts features by the encoder, and then recovers color by the decoder. Encoder layers are superimposed on corresponding layers in the decoder to improve the coloring effect.
- Isola et al.'s Method[12]: Use CGAN with U-Net as the generator. Add conditions to make the generated results more realistic. Moreover, PatchGAN was used in the discriminator to judge the generated results.
- Nazeri et al.'s Method[22]: Use DCGAN to color remote sensing images, and the discriminator is composed of several convolution layers. In addition, the objective function of generating network is optimized to make the network more stable.
- Li et al.'s Method[18]: Use a multi-scale discriminator to generate color images of different scales and input them into the corresponding layers of the discriminator to achieve multi-scale discrimination.
- Wu et al.'s Method[32]: Use DCGAN which has a multi-scale generator to color remote sensing images, and multi-scale is carried out by using convolution kernels with different sizes.

4.5 Comparison with state-of-the-arts

Table 1 and Table 2 show the results of the colorization method we proposed compared to the state-of-art methods on AID Data Set and NWPU Data Set. Higher qualitative and quantitative results indicate that our method achieves better coloring results compared with other comparison methods.

Fig.3 shows color images generated by the method we proposed and comparative methods. From a sensory point of view, the images generated by our method are most realistic and contain high spatial consistency and object visibility. However, some images generated by other methods have inconsistent coloring spaces such as the blue areas in the first row and sixth column images in Fig. 3.

The obtained superior color images profit from the structure of joint stream with different scales in which the macro stream works as the context constraints on the micro stream to ensure the consistency of the generated images. In addition, the LCFE module obtains the salient low-level information with global correlation, which is used to complement the lost local details of the micro stream deep layer feature as well as to enhance global context constraints on the micro stream, thereby enhancing the spatial consistency of the coloring. Moreover, the GSM module fuses the features from the micro and the macro stream by the dual gated scheme which

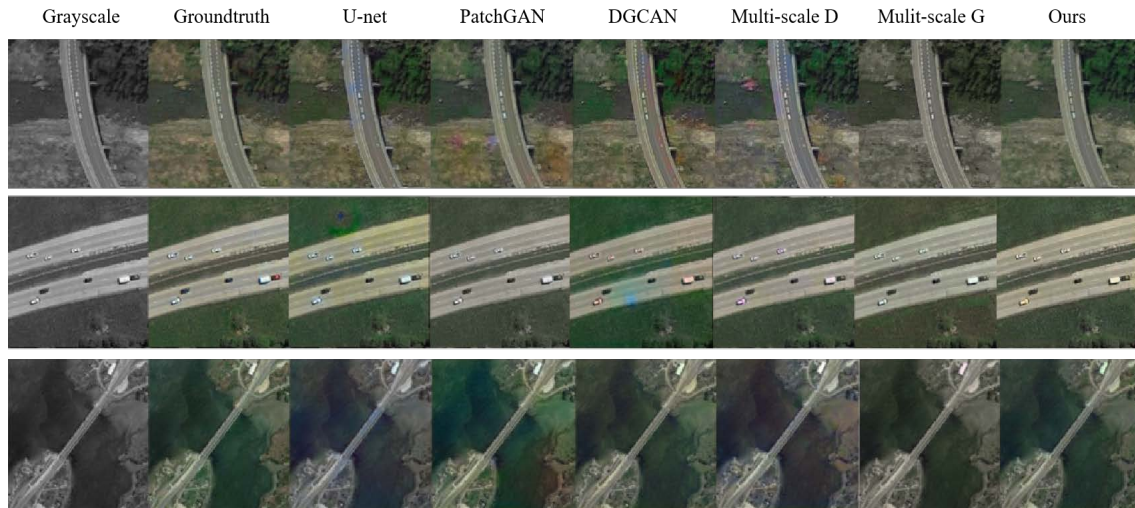


Figure 3: The generated images of the contrast experiments. The generated images using our method have strong spatial consistency and object visibility.

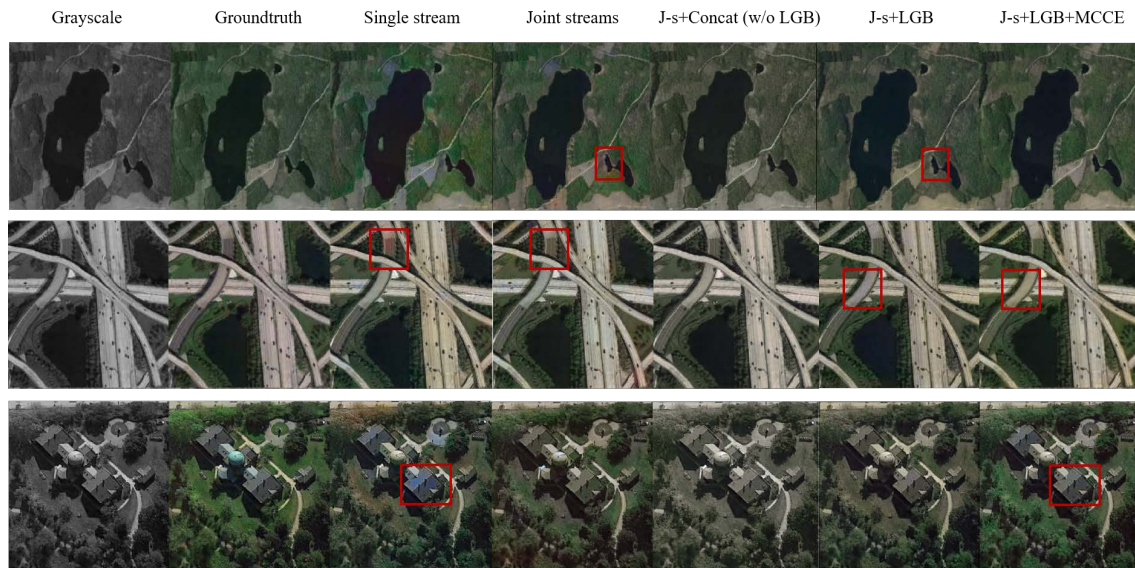


Figure 4: The generated images of the ablation experiments. The ‘J-s’ represents joint stream; ‘Concat (o/w LCFE)’ denotes directly transport the first layer feature from the macro stream to the micro stream without executing the LCFE module.

realizes proper fusion by selecting the effective information in the massive redundant information.

4.6 Ablation experiments

Ablation experiments are carried out to assess the importance of each module we proposed. The single-stream method refers to Nazeri et al.’s method which is set to be our baseline approach. Table 3 and Table 4 demonstrate the validity of each of the modules we proposed and the generated color images by the ablation experiments network are shown in Fig. 4.

The joint stream network is formed by introducing a macro-scale stream into the baseline, improving the coloring space consistency. For instance, the red road in the second row and third column in Fig. 4 becomes gray in the fourth column, making the generation appear more holistic. The LCFE module is joined between two streams to exact the salient low-level information with global correlation while supplementing missing shallow information and enhancing the macro context constraint to the micro stream. The image in the first row and sixth column in Fig. 4 shows the higher space consistency and more visible color object than the fourth column illustrating the significance of the LCFE module. The images generated by

introducing the GSM module become more attuned to the human perception since the GSM module fuses the output of two streams by suppressing the redundant noise and ensuring the reasonable expression of useful information.

5 CONCLUSION

In this paper, we propose a novel joint stream DCGAN for remote sensing image colorization. To address the problem of inconsistency in the color space caused by unbalanced scene spatial distribution of remote sensing images, we propose a generator using the macro stream as a prior to guide the micro stream for ensuring the consistency and stability of colorization. Aiming at supplementing the low-level information lost in the network deep layer as well as enhancing the global context constraints to generate images with strong space consistency and object visibility, the LCFE module is proposed to obtain the supplemented feature transported to the micro stream. Moreover, to take full advantage of the useful information drowning in the massive redundant noise to fuse the feature properly and further ensure the coloring effect. Compared with the most advanced methods, our method can generate higher-quality color images conforming to human perception.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (62072418), Major Scientific and Technological Innovation Project Shandong (2019JZZY020705) and, Fundamental Research Funds for the Central Universities (202042008).

REFERENCES

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 265–283.
- [2] Saeed Anwar, Muhammad Tahir, Chongyi Li, Ajmal Mian, Fahad Shahbaz Khan, and Abdul Wahab Muzaffar. 2020. Image colorization: A survey and dataset. *arXiv preprint arXiv:2008.10774* (2020).
- [3] M. Arjovsky, S. Chintala, and L. Bottou. 2017. Wasserstein GAN. (2017).
- [4] Hyojin Bahng, Seungjoo Yoo, Wonwoong Cho, David Keetae Park, Ziming Wu, Xiaojuan Ma, and Jaegul Choo. 2018. Coloring with words: Guiding image colorization through text-based palette generation. In *Proceedings of the european conference on computer vision (eccv)*. 431–447.
- [5] Peter F Brown, Vincent J Della Pietra, Peter V Desouza, Jennifer C Lai, and Robert L Mercer. 1992. Class-based n-gram models of natural language. *Computational linguistics* 18, 4 (1992), 467–480.
- [6] Gong Cheng, Junwei Han, and Xiaoqiang Lu. 2017. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE* 105, 10 (2017), 1865–1883.
- [7] Zijuan Cheng, Fang Meng, and Jingbo Mao. 2019. Semi-Auto Sketch Colorization Based on Conditional Generative Adversarial Networks. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 1–5.
- [8] Zhi Dou, Ning Wang, Baopu Li, Zhihui Wang, Haojie Li, and Bin Liu. 2021. Dual Color Space Guided Sketch Colorization. *IEEE Transactions on Image Processing* 30 (2021), 7292–7304.
- [9] D. M. Endres and J. E. Schindelin. 2003. A new metric for probability distributions. *IEEE Transactions on Information Theory* 49, 7 (2003), 1858–1860.
- [10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).
- [11] Samet Hicsonmez, Nermin Samet, Emre Akbas, and Pinar Duygulu. 2021. Adversarial Segmentation Loss for Sketch Colorization. *arXiv preprint arXiv:2102.06192* (2021).
- [12] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. In *IEEE Conference on Computer Vision & Pattern Recognition*.
- [13] Hyunsu Kim, Ho Young Jhoo, Eunhyeok Park, and Sungjoo Yoo. 2019. Tag2pix: Line art colorization using text tag with secat and changing loss. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9056–9065.
- [14] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [15] Bo Li, Yu-Kun Lai, Matthew John, and Paul L Rosin. 2019. Automatic example-based image colorization using location-aware cross-scale matching. *IEEE Transactions on Image Processing* 28, 9 (2019), 4606–4619.
- [16] Bo Li, Yu-Kun Lai, and Paul L Rosin. 2017. Example-based image colorization via automatic feature selection and fusion. *Neurocomputing* 266 (2017), 687–698.
- [17] Bo Li, Fuchen Zhao, Zhuo Su, Xiangguo Liang, Yu-Kun Lai, and Paul L Rosin. 2017. Example-based image colorization using locality consistent sparse representation. *IEEE transactions on image processing* 26, 11 (2017), 5188–5202.
- [18] F. Li, M. Lei, and C. Jian. 2018. Multi-Discriminator Generative Adversarial Network for High Resolution Gray-Scale Satellite Image Colorization. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*.
- [19] Matthias Limmer and Hendrik PA Lensch. 2016. Infrared colorization using deep convolutional neural networks. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 61–68.
- [20] Xinni Liu, Fengrong Han, Kamarul Hawari Ghazali, Izzeldin Ibrahim Mohamed, and Yue Zhao. 2019. A Review of Convolutional Neural Networks in Remote Sensing Image. In *Proceedings of the 2019 8th International Conference on Software and Computer Applications (Penang, Malaysia) (ICSCA '19)*. Association for Computing Machinery, New York, NY, USA, 263–267. <https://doi.org/10.1145/3316615.3316712>
- [21] M. Mirza and S. Osindero. 2014. Conditional Generative Adversarial Nets. *Computer Science* (2014), 2672–2680.
- [22] Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi. 2018. Image colorization using generative adversarial networks. In *International conference on articulated motion and deformable objects*. Springer, 85–94.
- [23] Fabien Pierre, Jean-François Aujol, Aurélie Bugeau, Nicolas Papadakis, and Vinh-Thong Ta. 2014. Exemplar-based colorization in RGB color space. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 625–629.
- [24] A. Radford, L. Metz, and S. Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *Computer ence* (2015).
- [25] Hui Ren, Jia Li, and Nan Gao. 2018. Automatic sketch colorization with tandem conditional adversarial networks. In *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Vol. 1. IEEE, 11–15.
- [26] Hui Ren, Jia Li, and Nan Gao. 2019. Two-stage sketch colorization with color parsing. *IEEE Access* 8 (2019), 44599–44610.
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.
- [28] Domonkos Varga and Tamás Szirányi. 2017. Twin deep convolutional neural network for example-based image colorization. In *International Conference on Computer Analysis of Images and Patterns*. Springer, 184–195.
- [29] Hao Wang and Xuedong Liu. 2021. Overview of image colorization and its applications. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Vol. 5. 1561–1565. <https://doi.org/10.1109/IAEAC50856.2021.9390626>
- [30] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon. 2018. CBAM: Convolutional Block Attention Module. *Springer, Cham* (2018).
- [31] M. Wu, X. Jin, Q. Jiang, S. J. Lee, and J. Huang. 2019. Remote Sensing Image Colorization Based on Multiscale SNet GAN. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*.
- [32] M. Wu, X. Jin, Q. Jiang, S. J. Lee, and S. Yao. 2020. Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space. *The Visual Computer* 2 (2020).
- [33] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* 55, 7 (2017), 3965–3981.
- [34] Zhongyou Xu, Tingting Wang, Faming Fang, Yun Sheng, and Guixu Zhang. 2020. Stylization-based architecture for fast deep exemplar colorization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9363–9372.
- [35] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas. 2017. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*.
- [36] Changqing Zou, Haoran Mo, Chengying Gao, Ruofei Du, and Hongbo Fu. 2019. Language-based colorization of scene sketches. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–16.
- [37] Ivana Žeger and Sonja Grgić. 2020. An Overview of Grayscale Image Colorization Methods. In *2020 International Symposium ELMAR*. 109–112. <https://doi.org/10.>

MMAAsia '22, December 13-16, 2022, Tokyo, Japan

1109/ELMAR49956.2020.9219019