Multi-Source Fusion Enhanced Feature Segmentation in Remote Sensing Imagery

Siman Wang[1], Qian Zhou[2]

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, the University Town, Shenzhen 518055, China
[1] wang-sm21@mails.tsinghua.edu.cn
[2] zhou.qian@sz.tsinghua.edu.cn

Keywords: Remote Sensing Images, Multi-source Data Fusion, Pixel-level Fusion, Semantic Segmentation.

Abstract

With deepening application of deep learning technology in the field of remote sensing, several challenges persist in the segmentation of remote sensing optical images. These challenges include: (1) insufficient availability of deep learning-based remote sensing semantic segmentation datasets; (2) inadequate utilization of multi-source remote sensing data in the field of semantic segmentation; (3) limited sample size for effective model training, as well as the need to enhance both the speed and accuracy of model training. To address these challenges, this study introduces a multi-source remote sensing dataset consisting of 15000 data pairs, each comprising remote sensing multispectral data, synthetic aperture radar(SAR) images, land use and land cover(LULC) data, digital elevation model (DEM) data as well as the analysis data including Slope, Aspect, and Hillshade. Through the application of an end-to-end network based on pix2pix, effective fusion and feature enhancement of multi-source remote sensing data were achieved. The structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and Spectral Angle Mapper(SAM) values reached 0.84, 23.14, and 0.19, respectively, representing significant improvements over the baseline Pix2pix model’s performance metrics of 0.62, 18.84, and 0.23. In downstream semantic segmentation applications, the enhanced dataset was utilized to train semantic segmentation models for remote sensing image analysis. This approach effectively improved the training speed and segmentation accuracy of the models, with the mean intersection over union (mIoU) increasing from 0.467 to 0.481 and accuracy rising from 0.734 to 0.746. Moreover, the visual representation of remote sensing image segmentation demonstrated noticeable enhancements.

1. Introduction

1.1 Research Background

In recent years, with the rapid advancement of artificial intelligence (AI) technology, its remarkable application capabilities and potential development have been demonstrated in various fields such as autonomous driving and intelligent healthcare (Zhang and Zhang, 2022). Exploring the integration of AI algorithms into the field of remote sensing has become a research focus (Shirmard et al., 2022), aiming to enhance the automation level and intelligence capability of remote sensing applications. However, the quality of training data sets significantly influences the effectiveness of AI algorithms (Kelly et al., 2019), highlighting the necessity of constructing high-quality remote sensing data sets to further promote the widespread application of AI in this domain.

1.2 Research Value

Existing remote sensing data sets suffer from deficiencies in both quantity and quality, primarily manifested in the following aspects as follows.

(1) Challenges in acquiring high-quality optical remote sensing data: Obtaining superior optical remote sensing data is hindered by the physical properties of light propagation (Chi et al., 2016). The reflection of ground objects must traverse the atmosphere to reach satellite payloads, a process heavily influenced by atmospheric conditions such as clouds and fog, resulting in significant contamination of optical remote sensing data. It is estimated that over 60% (Chernykh and Eskridge, 1996) of optical remote sensing data is contaminated to some extent. Hence, effectively removing clouds and fog from optical remote sensing images to meet the data quality requirements for downstream applications has become an urgent research topic (Adjovu et al., 2023).

(2) Difficulties in remote sensing data interpretation (Long et al., 2020) and annotation, and insufficient work on feature complementation and fusion of multi-source remote sensing data. While manual annotation of optical remote sensing images is feasible due to their imaging principles resembling those of the human eye, traditional manual annotation processes are time-consuming, labor-intensive, and less accurate, posing challenges to meet the rapidly growing data demands. Moreover, the significant disparity between the imaging principles of other types of remote sensing data and the human eye further exacerbates the difficulty of manual annotation (Cheng and Han, 2016). Therefore, achieving intelligent and automated interpretation and annotation of different types of remote sensing images represents a significant challenge in current research.

(3) Inadequate utilization and feature extraction of existing remote sensing datasets. Hence, research is needed on how to conduct more comprehensive feature enhancement on limited datasets to improve the training efficiency and performance of downstream applications (Tao et al., 2023).

As more and more satellites are launched every year, multi-source image data becomes easier to obtain. With the rapid rise of useful multi-source remote sensing data, the multi-modal method has been brought to the fore. Synthetic aperture radar(SAR), a microwave imaging method, has a brilliant capacity sufficient to capture surface structure and texture features though any kind of cloud (Chagas et al., 2020).According to its unique advantage to penetrating cloud in all weather conditions and being able to provide alternative data, researches on cloud removal using SAR data come into a new stage and have great development potential (Xiong et al., 2023).Plenty of relevant

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-395-2024 | © Author(s) 2024. CC BY 4.0 License.

researches use generative adversarial networks(GAN) to establish non-linear mapping relationship between the two different types of data (Li et al., 2021).The utilization of generative networks for the fusion and interpretation of multi-source remote sensing data has emerged as a novel research methodology. And the conditional generative adversarial network(cGAN) is an extension of the basic GAN, which is made of two adversarial neural network modules (Rodríguez-Suárez et al., 2022), the Generator and the Discriminator. The Generator attempts to extract useful information from input multi-modal data and generate the simulated images to fool the Discriminator. Moreover, modified cGAN based method with structure similarity index measure(SSIM) and L1 norm loss to force the generated image closer to the real optical image (Sun et al., 2022). In summary, the study of feature fusion and mutual interpretation of multi-source remote sensing imagery based on generative deep learning models has become a focal point of interest. However, multi-source remote sensing images are predominantly SAR and multispectral remote sensing images, with a limited variety of data types. These studies have primarily conducted metric evaluations on the generated fusion or the images themselves, without actually applying the produced data to enhance downstream remote sensing models in practical applications.Therefore, how to expand the types of remote sensing data and effectively utilize them for enhancing the quality of generated images and their practical downstream applications has become an issue that needs to be addressed currently.

1.3 Research Method

To address the aforementioned issues, this Paper focuses on constructing a diversified and high-quality multi-source remote sensing data set, achieving feature complementation by integrating various types of remote sensing data and increasing the diversity of sample features. Concurrently, leveraging Pix2pix-based generative networks, this study realizes feature fusion of multi-source remote sensing images and generation of optical remote sensing images. Under the dual constraints of spatial conditions and spectral conditions, the fusion-guided generation of optical remote sensing images closely approximates real remote sensing images, while also exhibiting feature enhancement effects. In downstream applications, this study selects remote sensing image semantic segmentation as a crucial application scenario. By incorporating generated pseudo-optical remote sensing images into the training set, significant improvements are observed in the training speed and segmentation performance of remote sensing image semantic segmentation networks. Moreover, notable enhancements are observed in the direct representation of segmented semantic images.

2. Multi-source Remote Sensing Dataset

2.1 Data Type Introduction

The creation of a multi-source remote sensing dataset begins with identifying the target remote sensing data types based on the dataset's intended use and selecting suitable payloads. For the purpose of automated segmentation and cloud removal in remote sensing imagery, this study opted for four types of remote sensing data for analysis: Synthetic Aperture Radar (SAR) images, multispectral imagery, Land Use and Land Cover (LULC) data, and Digital Elevation Model (DEM) data (Nelson et al., 2009). In addition to the commonly utilized Synthetic Aperture Radar (SAR) imagery, optical images, and semantic segmentation category maps, DEM (Digital Elevation Model) is a numerical model that describes the elevation information of the Earth's surface, either excluding vegetation and man-made structures or including them, providing three-dimensional information about the terrain. It accurately represents the undulations of the land, including mountains, valleys, plains, and other topographic features, and is capable of multi-scale and multi-source data fusion. Furthermore, DEM is one of the foundational datasets for terrain analysis, as shown in Figure 1, utilized for calculating topographic parameters such as Slope, Aspect, Hillshade. These derived parameters are highly valuable for research in fields such as geology, ecology, urban planning, and disaster risk assessment.In this study, four public datasets were selected to represent the four types of data under investigation. Specifically, Sentinel-2 (S2) was chosen as the source for remote sensing optical data, Sentinel-1(S1) for SAR image, Google's Dynamic World dataset for land use and land cover (LULC) data, and the Copernicus Digital Elevation Model (COP-DEM) provided by the European Union's Copernicus program as the source for DEM (Digital Elevation Model) data.

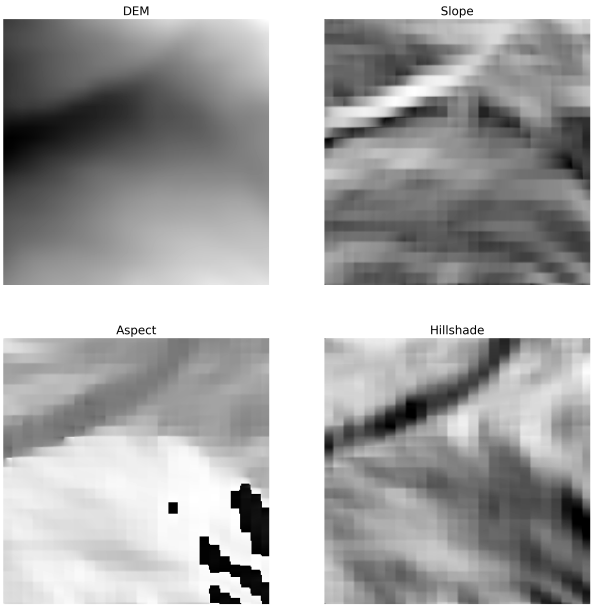

Figure 1. Terrain Analysis of DEM data including four categories of Data: original DEM data, Slope, Aspect, and Hillshade.

In terms of resolution, the optical remote sensing dataset derived from Sentinel-2 (S2) satellites is distinguished by a 10-meter pixel resolution and includes 12 spectral bands that extend across the visible to the infrared spectrum. This range facilitates detailed environmental monitoring and feature identification. In parallel, Sentinel-1 (S1) data, which constitutes the foundation for SAR (Synthetic Aperture Radar) imagery, also delivers a 10-meter spatial resolution. Notably, S1 data is characterized by its dual polarization VV and VH which enhance its capability to penetrate surfaces and detect structural features beneath the ground or through vegetation. Additionally, the Digital Elevation Model (DEM) data, complemented by pertinent terrain analysis data, is accessible at a 30-meter resolution. This coarser resolution is nonetheless critical for discerning subtle topographical variations across the Earth's surface, enabling precise geomorphological analyses and contributing to a more nuanced understanding of terrain features. Col-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-395-2024 | © Author(s) 2024. CC BY 4.0 License.

lectively, these datasets, with their distinct yet complementary resolutions and spectral characteristics, form a robust foundation for comprehensive earth observation and geoinformatic applications.Concurrently, as Figure 2 shows, this study conducts preprocessing on Sentinel-1 (S1) images by converting the SAR data into decibels (dB) units, aimed at enhancing the visual interpretability and facilitating subsequent analytical applications. This transformation is achieved through a logarithmic scaling of the original intensity values of the SAR images, which assists in reducing the dynamic range of the data. Consequently, this process renders the details within the images more discernible and enhances the clarity of the visual representation. Furthermore, the conversion to dB facilitates the comparison and interpretation of the data across different images and analyses, providing a standardized basis for the assessment of features and anomalies within the SAR imagery.

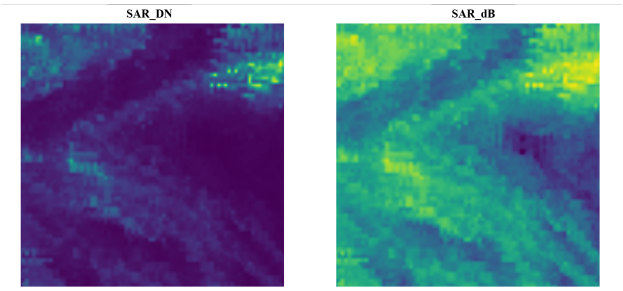

Figure 2. Preprocessing of SAR data, with original digital number data on the left and processed data on the right.

2.2 Multi-source dataset registration and creation

In the context of multi-source image registration, remote sensing data are typically stored in GeoTIFF format files, with the associated georeference information encoded in the header of the TIFF file. This information defines the relationship between pixel coordinates and geographical coordinates. In this research, the World Geodetic System 1984 (WGS 84) geodetic coordinate system and the Universal Transverse Mercator (UTM) projection coordinate system were selected for the registration process. The upper left corner coordinates of each remote sensing image were aligned, and subsequently, the boundary information of each image was obtained through the resolution and sampling size. This was followed by cropping and registration to ultimately achieve image data from different source sensors within the same area, ensuring spatial consistency and accuracy for further analysis and application. As illustrated in Figure 3, each registered data pair consists of a multispectral, LULC, SAR and DEM data. And there are 15000 pairs in the datadet totally.

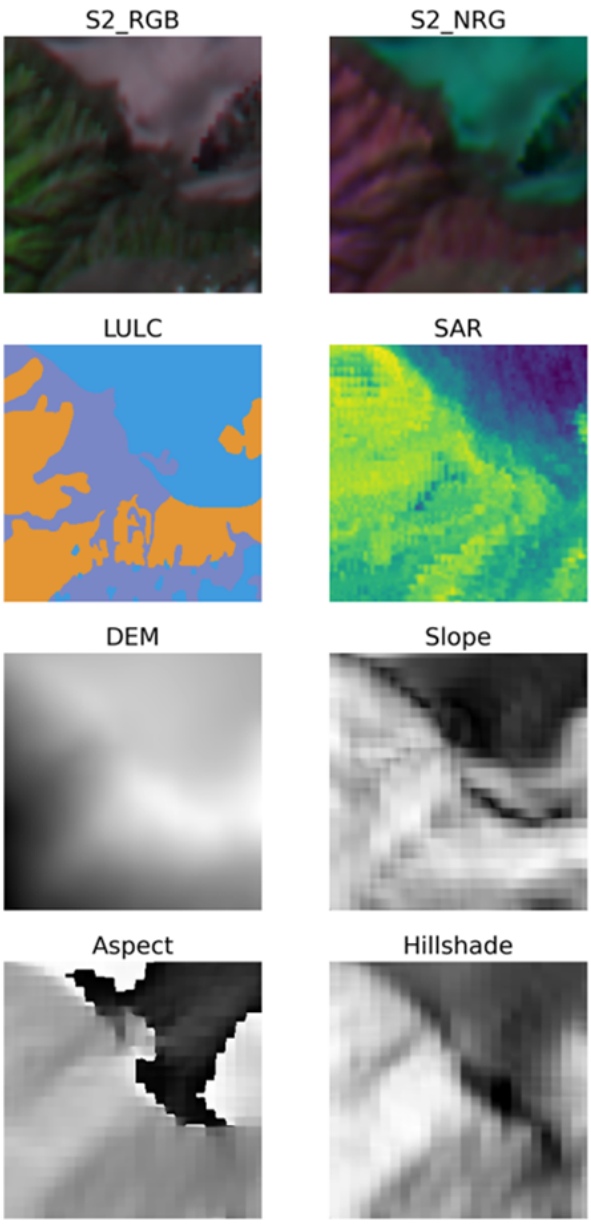

Figure 3. Composition of Multi-source Dataset, which includes four-band remote sensing optical data of RGBN, SAR data with VV and VH polarizations, LULC data, and four different types of terrain data.

3. Multi-source Data Fusion

3.1 Method

This section of the paper builds upon the multi-source remote sensing dataset introduced in the previous section. Its objective is to generate remote sensing optical data, denoted as S2, using LULC, DEM, and SAR data with cloud-penetrating capabilities. The generated remote sensing optical images remain uncontaminated by cloud and fog interference and integrate multi-source information from SAR, DEM, and LULC datasets. This approach addresses to some extent the insufficiency of existing remote sensing datasets in training artificial intelligence algorithms. Furthermore, the incorporation of fused images enhances the training speed and application capabilities of various artificial intelligence algorithms in the field of remote sensing.

This study is based on the Pix2pix model, which has been adjusted according to the characteristics of remote sensing data. As shown in Figure 4, LULC (Land Use and Land Cover)/SAR (Synthetic Aperture Radar)/DEM (Digital Elevation Model) data are used as label inputs, and embedding processing is applied to different types of remote sensing data separately. The aim is to extract the features of different types of data more effectively. In the context of machine learning data processing, embedding refers to the process of transforming data from its

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-2024-395-2024 | © Author(s) 2024. CC BY 4.0 License.

original representation space to a new space. The goal of this transformation is to highlight the feature information of the data, simplify the data structure, and improve the efficiency of subsequent processing. By applying separate embedding to multi-source data, this approach enhances the model’s compatibility with input images, allowing for input data of different resolutions and improving the model’s generalization capabilities.

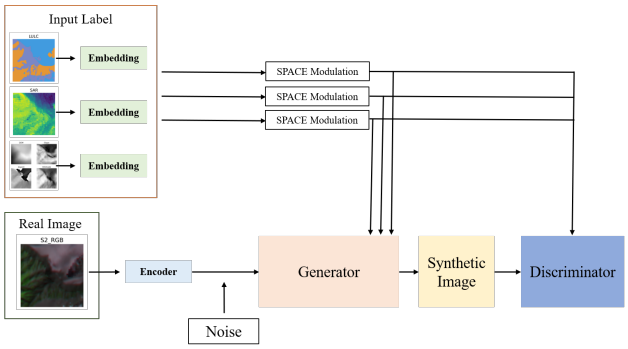

Figure 4. Structure of RS Data Fusion,RS-SPADE.

The distinctive aspect of the model lies in the application of the spatial modulation module from the SPADE (Spatially-Adaptive Normalization Generative Adversarial Networks) framework (Park et al., 2019), tailored to the characteristics of multi-source remote sensing data. As depicted in Figure 5, the spatial modulation module is primarily divided into two components. The first component involves the feature extraction through embedding of the multi-source remote sensing data. The second component integrates the extracted features to generate a three-dimensional tensor of spatial modulation parameters after convolution. These parameters are then computationally applied following each Batch Normalization (BN) operation within the generative model. This process continually reinforces the features of the input label, preventing the loss of spatial information during model training and enhancing the features with a strengthening effect.

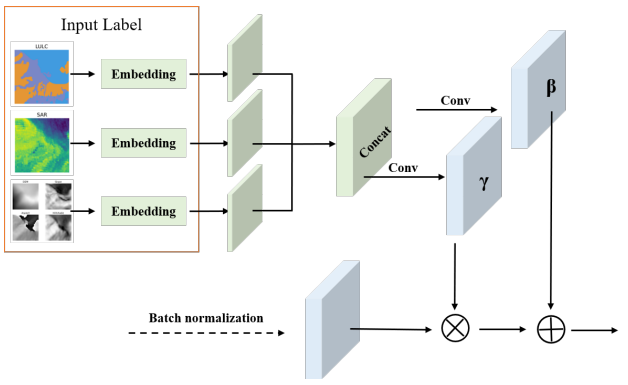

Figure 5. Multi-Source Modulation Module.

3.2 Multi-source fusion image generation.

In terms of experimental data, this research endeavored to partition the dataset into training, validation, and testing subsets in a 7:2:1 ratio. Given the considerable computational expense associated with processing the 12-channel spectral data from remote sensing multispectral imagery, which is both time-intensive and resource-demanding, the experiment opted to focus on four representative spectral bands for the training phase. These selected bands encompass the visible light spectrum (Blue, Green, Red) and the near-infrared (Near-Infrared), amounting to a total of four bands. The spectral similarity calculations were approximated based on the values derived from these four bands.

The Synthetic Aperture Radar (SAR) data within the experiment incorporated dual polarization data types: VV (Vertical Transmit, Vertical Receive) and VH (Vertical Transmit, Horizontal Receive). The LULC dataset was categorized into nine distinct classes, each with its respective data size and type designation: ”no data” was labeled as 1, ”water” as 2, ”forest” as 3, ”grass” as 4, ”wetland” as 5, ”flooded agriculture” as 6, ”bare surface” as 7, ”soil” as 8, and ”snow and ice” as 9. The Digital Elevation Model (DEM) data was composed of four distinct types of information: the raw DEM data, slope (Slope), aspect (Aspect), and Hillshade (Hillshade).

For the assessment of the quality of the generated images, this study adopted three distinct evaluative metrics: the Structural Similarity Index Measure (SSIM), the Peak Signal-to-Noise Ratio (PSNR), and the Spectral Angle Mapping (SAM). These indicators serve to provide a comprehensive evaluation of the fidelity and accuracy of the synthetic images produced in the context of the experiment.

Model Name	PSNR	SSIM	SAM
Our Model	23.14	0.84	0.19
Pix2pix	18.84	0.62	0.23

Table 1. Table of Comparative Model Evaluation Metrics

To effectively evaluate the impact of various input data on the feature fusion and generation performance of multi-source remote sensing models, this research conducted an ablation study, the results of which are presented in Table 2. The comparative analysis indicates that different types of multi-source data collectively enhance the generation of optical images. Land Use and Land Cover (LULC) data provide details on the types of land cover and their corresponding labels, assisting the model in learning the spectral features associated with distinct categories and leading to images with more pronounced class characteristics. S1 data predominantly supply geometric scattering and structural information about land features, enabling cloud and fog-independent remote sensing imaging. This ensures the availability of remote sensing imagery under all weather conditions, thereby improving the model’s generalization capabilities and establishing a basis for the simulation and generation of all-weather remote sensing optical images. Elevation data from the Digital Elevation Model (DEM), along with its associated auxiliary data—Slope, Aspect, and Hillshade—offer comprehensive details on natural terrain textures, elevation variations, directional orientation, and alterations in shadowing and lighting conditions. This information remains relatively consistent and is not subject to significant changes over time or due to weather variations. Furthermore, the terrain textures included in this data enhance the realism of the generated images.Additionally, as depicted in Figure 6, the generated images (syn_NRG and syn_RGB) play an enhancing role in geographical features, making areas with weak contrast more distinctly visible and offering higher discrimination on the generated images. Therefore, in theory, incorporating generated data with feature enhancement effects into a neural network can yield data augmentation benefits for training purposes. However, in urban areas where topographical variations are minimal and ar-

tificially created texture features are the primary targets for recognition, SAR (Synthetic Aperture Radar) and DEM (Digital Elevation Model) elevation data do not provide high-value information for image inversion in urban regions. On the contrary, they may introduce noise, which degrades the generation performance.

Input Data	PSNR	SSIM	SAM
Only S1	18.83	0.76	0.25
Only LULC	20.77	0.80	0.21
LULC+S1	21.10	0.80	0.22
LULC+DEM	21.36	0.81	0.21
S1+DEM	21.89	0.82	0.20
s1+LULC+DEM	23.14	0.84	0.19

Table 2. Table of Comparative Model Evaluation Metrics

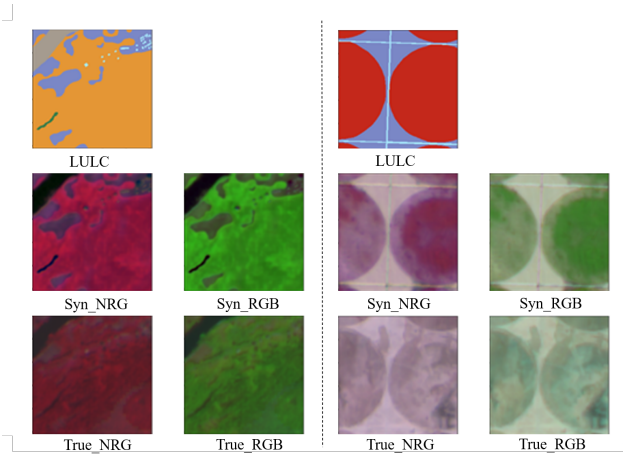

Figure 6. Generated Images.

4. Remote Sensing Semantic Segmentation

Based on the remote sensing image generation network proposed in the previous section and the evaluation of its generation performance, this study found that the pseudo optical images generated by this network effectively and significantly retain the semantic information of the input category labels. Additionally, in the field of semantic segmentation of remote sensing images, there is a severe shortage of training datasets, and annotating remote sensing datasets requires a significant investment of manpower, resources, and finances, making it a costly process. To address the shortage of training datasets in remote sensing semantic segmentation, the network model trained in the previous section is utilized to enhance image features and augment training data. A comparison is made between using generated data and using only original real data for training remote sensing image segmentation models, demonstrating the effectiveness of applying the generated model to segmentation scenarios.

4.1 Segmentation Method

The structural framework of this study is illustrated in Figure 7. The RS-SPADE algorithm trained in the previous section is employed as the core of the data feature enhancement module. The generated remote sensing optical images with four channels (RGB+NIR) by the RS-SPADE generation algorithm are mixed with their corresponding real remote sensing optical images in a 1:1 ratio for remote sensing semantic segmentation model training. The data for the remote sensing semantic segmentation model consists of pairs of ”LULC+S2”, with each pair containing land use classification maps matched to geographic locations.The selected foundational model for remote sensing semantic segmentation in this paper is U-Net (Ronneberger et al., 2015), chosen for its capability to achieve effective segmentation performance with limited samples, aligning well with the research demands in remote sensing semantic segmentation scenarios. Moreover, U-Net occupies a central position among a wide range of image segmentation models. Additionally, many emerging network designs are essentially improvements and evolutions based on the structure of U-Net, making its fundamental framework a cornerstone in the field of image segmentation.

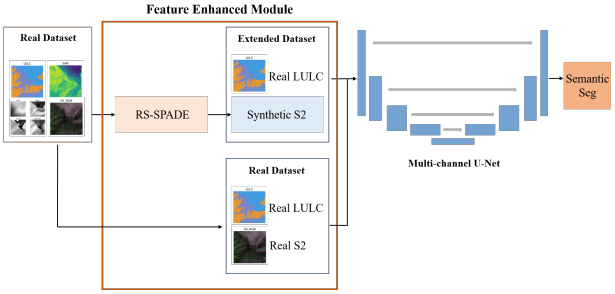

Figure 7. Feature Enhanced RS Semantic Segmentation Structure.

In the evaluation of remote sensing image segmentation performance, this study adopts two assessment metrics: accuracy (acc) and mean Intersection over Union (mIoU). Accuracy (acc) is computed by dividing the number of correctly predicted samples by the total number of samples, reflecting the model’s discriminative ability across classes. Mean Intersection over Union (mIoU) is the average of individual Intersection over Union (IoU) values for different classes. IoU measures the overlap between predicted and ground truth regions for each class. By averaging IoU values across all classes, mIoU provides a comprehensive assessment of the model’s segmentation performance across the dataset. A high mIoU indicates good segmentation performance across common classes, making it a suitable metric for evaluating overall segmentation effectiveness in this study.

4.2 Segmentation Result

In the experiment, the original real dataset consists of 1558 (4,256,256) spectral remote sensing images (S2) and corresponding land use and land cover (LULC) maps with 9 classes, along with 1558 (1,256,256) generated pseudo-remote sensing optical images through RS-SPADE network. These generated images are paired with their corresponding LULC maps to create feature-enhanced data pairs. The training dataset comprises a total of 3116 data points, with a 1:1 ratio of real data pairs to generated data pairs, forming what is termed as the Mix Dataset. Additionally, to validate the effectiveness of data augmentation, a comparative experiment is conducted using the 1558 original real data points directly as the training dataset, referred to as the True Dataset. Moreover, another set of 1558 real data pairs are separately selected as the validation set (Val Dataset) and the test set (Test Dataset). Both validation and test sets only utilize real images to intuitively assess the performance of training data on semantic segmentation of real remote sensing images. RS-Unet extends the classical U-net architecture

by increasing the input data channel count and deepening the down-sampling layers. Currently, the input channels are set to 4, with input data size of (4,256,256), which extends beyond the RGB channels, and can be readily adjusted to accommodate 10 or more channels in subsequent experiments to better suit semantic segmentation tasks of remote sensing optical images. Regarding the model's layering, the network undergoes 5 down-sampling layers followed by 5 up-sampling layers, allowing for the preservation of finer detail features and enhancing the precision of semantic segmentation tasks. Convolution layers and pooling layers employ 3x3 kernels for computation. At the final layer, softmax activation is utilized, producing semantic segmentation maps of size (1,256,256), thereby achieving intelligent segmentation of remote sensing optical images.

As shown in Table 3, enhancing features of original input data effectively improves the performance of semantic segmentation models on real remote sensing optical images. As illustrated in Figure 8, during the training process, the Mix dataset enables semantic segmentation models to achieve better metrics more rapidly. Figure 9 further demonstrates that for real remote sensing optical image segmentation scenarios, employing a model trained with Mix Dataset after 50 epochs with feature enhancement yields superior performance compared to directly training on the original data. From the segmentation images, it is evident that automated semantic segmentation tasks for remote sensing optical images are challenging due to factors such as weather, cloud cover interference, and phenomena like spectral similarity of different objects and spectral diversity of the same object on the ground. Consequently, segmentation accuracy is generally low. However, the data in Table 3 effectively demonstrates that employing RS-SPADE generative networks for data augmentation and feature enhancement contributes to enhancing the segmentation accuracy of remote sensing optical images.

Data Type	mIoU	Accuracy
Mix Data	0.480	0.746
True Data	0.467	0.734

Table 3. Table of Comparison of Training Before and After Data Augmentation

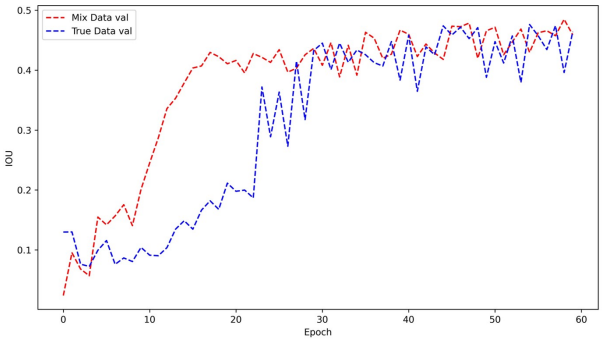

Figure 8. IoU Change Speed Comparison during Model Training.

5. Conclusions

This paper proposes a method for producing multi-source remote sensing data sets, which can produce corresponding multi-source remote sensing data sets according to application goals.

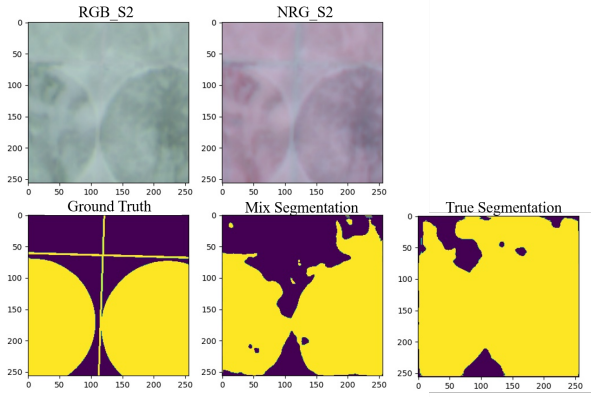

Figure 9. Comparison of Remote Sensing Semantic Segmentation Images.

At the same time, the RS-SPADE model for multi-source remote sensing optical image generation is proposed. In this model, the features of different source data can be extracted and integrated into the generated data set, effectively proving the complementary characteristics of different source data. Therefore, the generated remote sensing optical images with both spatial accuracy and spectral accuracy can facilitate efficient data expansion and feature enhancement. IN experiment, the generated data was applied to the U-Net remote sensing semantic segmentation model. By comparison, it was found that the data enhanced with generative features helps the segmentation network perform semantic segmentation with higher accuracy and efficiency, which is superior in terms of segmentation evaluation indicators and effect visualization. However, the current study utilizes remote sensing optical images with a spatial resolution of 10 meters mainly in non-urban areas and only employs the four NRGB bands. In the future experiments could be conducted on images with higher spatial resolution and additional bands to rigorously evaluate the effects, which may yield even better results. In summary, this paper not only realizes the fusion of multi-source remote sensing data, but also effectively applies the generated optical data to downstream applications of remote sensing, providing a brand new idea for the supplement of remote sensing data sets and the training of related models, which is extremely meaningful for the development of AI in remote sensing field.

6. Acknowledgements

Thanks for the data from Google Earth Engine and the Remote Sensing Research Institute for guidance and server support.

References

Adjovu, G. E., Stephen, H., James, D., Ahmad, S., 2023. Overview of the application of remote sensing in effective monitoring of water quality parameters. *Remote Sensing*, 15(7), 1938.

Chagas, E. T., Frery, A. C., Rosso, O. A., Ramos, H. S., 2020. Analysis and classification of SAR textures using information theory. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 663–675.

Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. *ISPRS journal of photogrammetry and remote sensing*, 117, 11–28.

Chernykh, I. V., Eskridge, R. E., 1996. Determination of cloud amount and level from radiosonde soundings. *Journal of Applied Meteorology and Climatology*, 35(8), 1362–1369.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. *Proceedings of the IEEE*, 104(11), 2207–2219.

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., King, D., 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, 17, 1–9.

Li, X., Du, Z., Huang, Y., Tan, Z., 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179, 14–34.

Long, Y., Xia, G.-S., Li, S., Yang, W., Yang, M. Y., Zhu, X. X., Zhang, L., Li, D., 2020. DiRS: On creating benchmark datasets for remote sensing image interpretation. *arXiv preprint arXiv:2006.12485*.

Nelson, A., Reuter, H., Gessler, P., 2009. DEM production methods and sources. *Developments in soil science*, 33, 65–85.

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2337–2346.

Rodríguez-Suárez, B., Quesada-Barriuso, P., Argüello, F., 2022. Design of CGAN models for multispectral reconstruction in remote sensing. *Remote Sensing*, 14(4), 816.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 234–241.

Shirmard, H., Farahbakhsh, E., Müller, R. D., Chandra, R., 2022. A review of machine learning in processing remote sensing data for mineral exploration. *Remote Sensing of Environment*, 268, 112750.

Sun, Y., Jiang, W., Yang, J., Li, W., 2022. SAR target recognition using cGAN-based SAR-to-optical image translation. *Remote Sensing*, 14(8), 1793.

Tao, C., Qi, J., Guo, M., Zhu, Q., Li, H., 2023. Self-supervised remote sensing feature learning: Learning paradigms, challenges, and future works. *IEEE Transactions on Geoscience and Remote Sensing*.

Xiong, Q., Li, G., Yao, X., Zhang, X., 2023. SAR-to-optical image translation and cloud removal based on conditional generative adversarial networks: Literature survey, taxonomy, evaluation indicators, limits and future directions. *Remote Sensing*, 15(4), 1137.

Zhang, L., Zhang, L., 2022. Artificial intelligence for remote sensing data analysis: A review of challenges and opportunities. *IEEE Geoscience and Remote Sensing Magazine*, 10(2), 270–294.