ISSN: 2688-8645

Deep Learning-Based Stripe Restoration Algorithm for Landsat 7

Peng Yin

Henan Polytechnic University, Jiaozuo, Henan, China

Abstract

Since the failure of the Scan Line Corrector (SLC) on the Landsat 7 ETM+ sensor in 2003, approximately 22% of pixels in acquired images are missing in a regular stripe pattern, severely limiting their application in time-series analysis and land cover monitoring. To address the limitations of traditional methods in modeling complex land surfaces and their tendency to introduce spectral distortions, this paper proposes a stripe-aware deep learning restoration network called SINet (Stripe Inpainting Network). The network leverages a Stripe Attention Module (SAM) to exploit the geometric prior of stripes and aggregate features along the stripe direction, and a Spectral Reconstruction Module (SRM) to model multi-band correlations and preserve spectral fidelity. A hybrid loss function combining pixel loss, perceptual loss, spectral angle loss, and gradient loss is designed. Experiments conducted in the Zhengzhou-Kaifeng-Xuchang junction area of Henan Province show that SINet achieves a PSNR of 38.2 dB, an SSIM of 0.978, and an SAM of 1.82° on simulated data, significantly outperforming baseline methods such as nearest neighbor interpolation, bilinear interpolation, and local histogram matching. When applied to land cover classification, the overall accuracy improves from 74.3% (uncorrected) to 88.5%, approaching the 90.2% accuracy of original Landsat 8 imagery. This study provides an effective solution for restoring historical Landsat 7 data.

Keywords

Landsat 7; Stripe Restoration; Deep Learning; Attention Mechanism; Image Inpainting; Remote Sensing.

1. Introduction

The Landsat series of satellites, jointly operated by the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS), has provided nearly 50 years of continuous moderate-resolution remote sensing data since the launch of the first satellite in 1972. It has become the most widely used and longest-running Earth observation data source globally (Wulder et al., 2019)[1]. Landsat data play an irreplaceable role in numerous fields, including agricultural monitoring, forest resource surveys, land use/cover change analysis, disaster assessment, and climate change research (Roy et al., 2014; Loveland & Dwyer, 2012)[2].

Landsat 7, launched in 1999, carries the Enhanced Thematic Mapper Plus (ETM+), which acquires images in eight spectral bands (including a panchromatic band) with a spatial resolution of 30 m (60 m for thermal infrared, 15 m for panchromatic) and a 16-day revisit cycle. However, on May 31, 2003, the Scan Line Corrector (SLC) of Landsat 7 failed, causing data gaps in the acquired images. Specifically, approximately 22% of pixels are missing, appearing as periodic diagonal stripes, with the stripe pattern varying over time (Storey et al., 2005)[4]. This problem, known as the "SLC-off" issue, has severely impacted the subsequent application of Landsat 7 data, particularly in analyses requiring continuous and complete surface information.

Although the USGS provides "gap-filled" products based on multi-temporal data interpolation, this method relies on cloud-free, snow-free, and phenologically similar images from the same

ISSN: 2688-8645

area within a short time window (typically ±30 days) and is sensitive to dynamic surface changes, resulting in limited restoration effectiveness (Scaramuzza & Barsi, 2005)[5]. Traditional stripe removal methods include local interpolation-based approaches (e.g., nearest neighbor, bilinear[6], cubic convolution), frequency-domain filtering methods (e.g., Fourier transform, wavelet transform)[7], and multi-source data fusion methods (e.g., using Landsat 8 or MODIS data as auxiliary sources) (Chen et al., 2011; Zeng et al., 2013; Pringle et al., 2009; Rakwatin et al., 2009)[6]. These methods have their respective strengths and weaknesses but generally suffer from the following shortcomings: (1) limited ability to model stripe patterns in complex terrain, (2) susceptibility to introducing spectral distortions or spatial blurring, thereby compromising the physical meaning of the original data, and (3) low automation, requiring manual intervention or parameter tuning.

In recent years, deep learning has achieved breakthroughs in computer vision, particularly in image inpainting tasks (Pathak et al., 2016; Yu et al., 2018; Liu et al., 2018)[11]. Image inpainting aims to reasonably infer missing regions based on known content, which aligns closely with the Landsat 7 stripe restoration problem. Deep Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Transformers can learn rich image priors from large datasets and automatically learn prediction patterns for missing regions, potentially overcoming the drawbacks of traditional methods. However, remote sensing images are characterized by multispectral data, large scales, and complex land cover. Directly applying natural image inpainting methods to Landsat 7 stripe restoration poses challenges: (1) the spectral dimension of remote sensing images requires special handling to maintain physical correlations between bands; (2) the stripe gaps are regular and periodic, unlike the random missing patterns in natural images; and (3) obtaining ground truth data for training is difficult.

Therefore, researching deep learning methods specifically for the Landsat 7 SLC-off stripe problem has significant theoretical value and practical importance. On one hand, it can enrich the theoretical framework of remote sensing image restoration and provide insights for similar issues with other sensors. On the other hand, it can effectively restore vast amounts of historical Landsat 7 data, extending their useful life and supporting long-term surface change monitoring and global change research.

Early research on the Landsat 7 SLC-off stripe problem primarily focused on interpolation and multi-temporal fusion. The official USGS method uses local histogram matching with cloud-free images from before and after the target date (typically within ±30 days) to fill missing pixels (Scaramuzza & Barsi, 2005). This method assumes that surface features do not change significantly over a short period, but it can introduce large errors in areas with rapid land cover changes (e.g., crop growth, fires, floods). Subsequently, researchers proposed various improvements, such as weighted multi-temporal data fusion (Chen et al., 2011), adaptive window regression analysis (Zeng et al., 2013), and interpolation combined with spatial texture information (Pringle et al., 2009). Additionally, frequency-based stripe removal methods have been attempted, such as using Fourier transforms to identify and suppress stripe frequency components (Rakwatin et al., 2009), but these methods are less effective for non-stationary stripe patterns[14].

Image inpainting is a hot topic in computer vision, and deep learning methods have become mainstream. Pathak et al. (2016) proposed Context Encoder, which first combined an encoder-decoder architecture with adversarial training for irregular missing region inpainting. Subsequently, attention mechanisms were introduced to enhance long-range dependency modeling (Yu et al., 2018; Liu et al., 2018). Partial Convolution (Liu et al., 2018) and Gated Convolution (Yu et al., 2019) were proposed to address feature issues caused by irregular masks. Recently, Transformers, with their global receptive fields, have achieved excellent results in image inpainting (Li et al., 2022; Wan et al., 2021)[9].

ISSN: 2688-8645

In the remote sensing domain, deep learning has also been gradually applied to tasks such as cloud removal and stripe noise elimination. Zhang et al. (2018) used a fully convolutional network to remove stripe noise from remote sensing images. Xie et al. (2019) applied a generative adversarial network to Landsat 7 stripe filling but only targeted single bands and did not fully utilize spectral information. Ma et al. (2020) proposed a multi-scale feature fusion network combined with multi-temporal data for Landsat 7 stripe restoration and achieved good results. However, existing methods still have room for improvement: (1) insufficient use of the prior knowledge of regular stripe patterns, (2) inadequate modeling of spectral correlations, and (3) lack of loss functions tailored to remote sensing image characteristics[12].

In summary, current methods for Landsat 7 stripe restoration mainly suffer from the following problems:

(1) Traditional methods rely on multi-temporal data and are sensitive to surface changes, leading to unstable restoration results.

(2) Deep learning methods often directly transfer natural image inpainting networks without fully considering the multispectral nature and regular stripe patterns of remote sensing images.

(3) There is a lack of large-scale, high-quality training datasets, which limits model performance.

(4) Evaluation metrics for restoration results are often limited, with insufficient validation in downstream applications.

This paper aims to develop a deep learning-based stripe restoration algorithm for Landsat 7 that can efficiently and accurately recover missing pixels while preserving the spectral characteristics and spatial details of the original image. The specific objectives are:

(1) To construct a training dataset suitable for Landsat 7 SLC-off stripe restoration.

(2) To design a multispectral image restoration network that incorporates stripe prior information.

(3) To propose a hybrid loss function that balances pixel accuracy, spectral fidelity, and spatial texture.

(4) To validate the effectiveness of the proposed algorithm through quantitative and qualitative experiments and compare it comprehensively with existing methods.

Main Research Content:(1)Data Preprocessing and Dataset Construction:Use complete Landsat 8/9 images to simulate Landsat 7 stripe patterns, generating paired training data. Also, collect real Landsat 7 SLC-off images as a test set, considering different land cover types, seasons, and stripe phases.(2)Stripe-Aware Deep Learning Network Design:Analyze the spatial distribution of stripe patterns and design network modules that encode stripe position information, such as a stripe attention mechanism and regular mask encoding. Simultaneously, consider correlations between multispectral bands by introducing spectral attention or cross-band feature fusion.(3)Hybrid Loss Function:Combine pixel-level loss (L1/L2), perceptual loss (based on pre-trained VGG features), spectral angle loss (SAM), and gradient loss (edge preservation) to make the restoration results both accurate and realistic.(4)Experimental Validation and Analysis:Conduct quantitative comparisons (PSNR, SSIM, SAM, RMSE, etc.) and qualitative visual assessments on simulated and real data against various baseline methods (including traditional interpolation, classical remote sensing methods, and general deep learning inpainting methods). Perform ablation studies to verify the effectiveness of each module. Investigate the impact of restoration results on downstream tasks (e.g., land cover classification).

ISSN: 2688-8645

2. Theoretical Foundation

2.1. Landsat 7 SLC-off Stripe Problem

The Landsat 7 ETM+ sensor uses an opto-mechanical scanning method for imaging. The function of the Scan Line Corrector (SLC) is to compensate for satellite platform motion and ensure correct alignment of scan lines. After the SLC failure in 2003, the scan lines could not be correctly aligned, causing overlaps or gaps between adjacent scan lines during imaging, ultimately forming regular striped data gaps. According to the failure type, stripe patterns can be divided into two types: 1) SLC-off Mode 1 (May 31, 2003 to December 31, 2007): stripes appear periodically and repeatedly, with missing lines accounting for approximately 22% of the total lines; 2) SLC-off Mode 2 (January 1, 2008 to present): the stripe pattern changes slightly, but the missing proportion remains essentially unchanged. Stripe widths range from approximately 1 to 6 pixels and form an angle of about 16° with the scanning direction[4].

The interference of stripe gaps on remote sensing applications is mainly manifested in: 1) disrupting the continuity of land cover, leading to errors in analyses such as classification and change detection; 2) affecting the calculation of spectral features such as vegetation indices, introducing spurious fluctuations; 3) limiting the completeness of time series analysis, creating data gaps. Therefore, developing effective stripe restoration methods is key to recovering the value of Landsat 7 data.

2.2. Basic Theory of Image Inpainting

Image inpainting is an important research topic in the fields of computer vision and image processing. Its core objective is to reasonably fill missing or damaged regions based on pixel information from known areas in the image, so that the restored image remains visually continuous and natural, and is semantically consistent with the surrounding environment. This process can be abstracted as: given an incomplete image and a mask indicating the positions of missing pixels, the inpainting algorithm needs to generate a complete image that is identical to the original input in known regions, while in missing regions, it infers content that conforms to the overall structure and semantics of the image through algorithmic reasoning.

Traditional image inpainting methods can be mainly divided into two categories: diffusion-based methods and texture synthesis-based methods. Diffusion-based methods use partial differential equations to propagate pixel values from known regions inward along the isophote directions of the image, thereby achieving gap filling for small areas with simple structures. These methods work well for narrow missing regions or smooth image areas, but tend to produce over-smoothing and blurring in scenes with large missing areas or complex textures. Texture synthesis-based methods sample similar image patches from known regions and fill missing areas through patch matching and stitching. These methods can preserve texture details well, but often struggle to ensure global semantic consistency for areas with complex structures.

With the rise of deep learning, image inpainting technology has achieved breakthrough progress. Deep learning methods train neural networks on large-scale datasets to learn the natural prior distribution of images, enabling the generation of more realistic and semantically consistent inpainting results. Representative works include: the Context Encoder proposed by Pathak et al., which first combined an autoencoder structure with a generative adversarial network, achieving semantic-based image inpainting ; the Partial Convolution proposed by Liu et al., which effectively handles missing regions of irregular shapes by introducing a mask update mechanism ; the Gated Convolution proposed by Yu et al., which further introduces a learnable gating mechanism to adaptively control feature propagation at different positions, improving inpainting quality ; in recent years, Transformer-based models have demonstrated excellent performance in image inpainting tasks by leveraging their global receptive field

ISSN: 2688-8645

advantages, better capturing long-range dependencies in images . These deep learning methods have achieved remarkable results in natural image inpainting, providing important technical references for tasks such as remote sensing image stripe restoration[11].

3. Deep Learning-based Stripe Inpainting Algorithm

3.1. Problem Definition and Overall Framework

The stripe restoration problem for Landsat 7 SLC-off images can be formalized as an image inpainting task. Given a multispectral image containing stripe gaps and a corresponding binary mask that precisely identifies whether each pixel is valid, the goal of restoration is to generate a complete image such that in the known valid pixel regions, the restoration result is completely consistent with the original input; while in the missing stripe regions, the restoration result should have reasonable spectral characteristics and spatial continuity, capable of naturally integrating with the surrounding non-stripe areas while maintaining the physical meaning and spectral features of the ground objects.

To address this problem, this paper proposes a stripe-aware restoration network called SINet, whose overall framework consists of three core stages, as shown in Figure 1 below. The first stage is the data preparation phase, which uses complete images from Landsat 8 and Landsat 9 to simulate stripe gaps according to the precise parameters of the Landsat 7 SLC-off failure, generating a large number of paired training samples, including simulated stripe images and their corresponding real complete images. At the same time, real Landsat 7 SLC-off images are collected as a test set to evaluate the model's performance in practical applications.

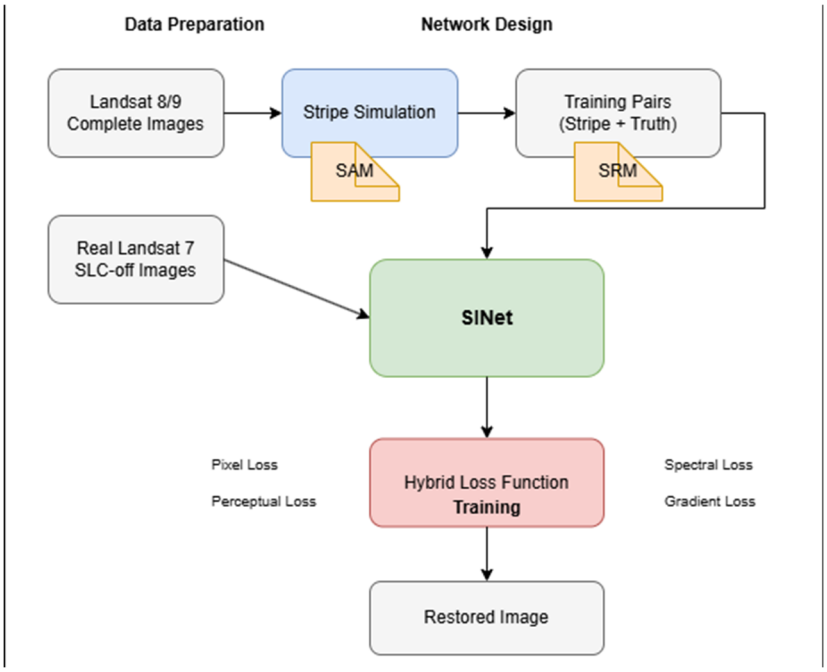

Figure 1. Overall framework of SINet

The second stage is the network design phase, where a stripe-aware deep learning network called SINet is designed. This network adopts an encoder-decoder architecture, as shown in Figure 2 below. A Stripe Attention Module is introduced in the encoder to leverage the geometric prior of stripe gaps and perform feature aggregation along the stripe direction. A Spectral Reconstruction Module is introduced in the decoder to explicitly model the correlations between multispectral bands, enhancing the spectral fidelity of the restoration results. The final stage is the model training phase, where a hybrid loss function is employed for end-to-end training of the network. This loss function integrates pixel-level reconstruction

ISSN: 2688-8645

loss, perceptual loss based on deep features, spectral angle loss, covariance matching loss, and gradient loss, aiming to balance pixel accuracy, visual realism, spectral preservation, and edge sharpness.

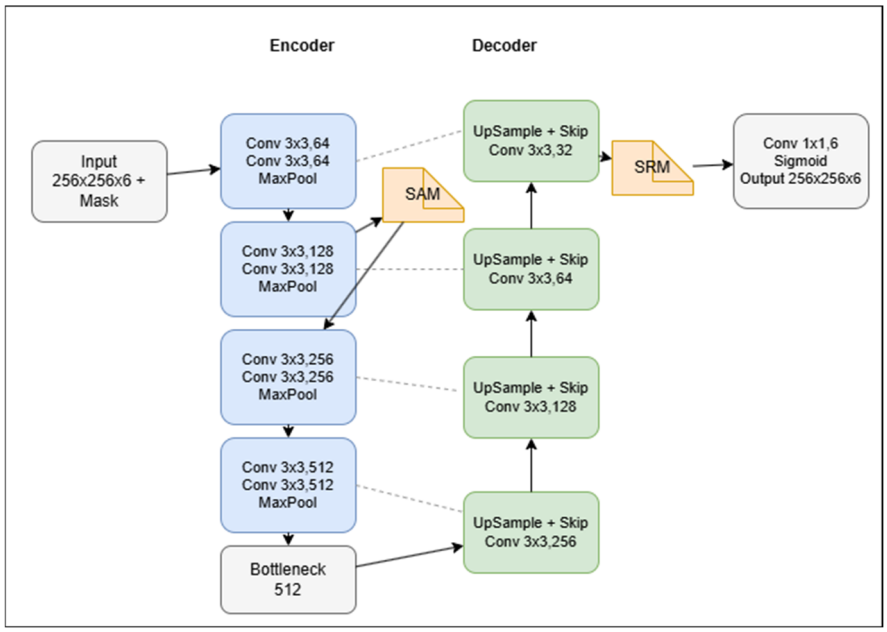

Figure 2. Architecture of SINet

3.2. Data Preprocessing and Dataset Construction

High-quality training datasets are crucial for the success of deep learning models. For the Landsat 7 stripe restoration task, this paper adopts a stripe simulation method based on complete Landsat 8 and Landsat 9 images to construct training data, and collects real Landsat 7 SLC-off images as the test set.

The training data selects Landsat 8 OLI surface reflectance products from 2015 to 2020, requiring cloud cover of less than ten percent, totaling five hundred scenes. These images cover Henan Province and surrounding areas in China, including various land cover types such as forest, farmland, urban, and water bodies, ensuring the diversity of training data. All images have undergone radiometric calibration and atmospheric correction, with pixel values being sixteen-bit integer data after reflectance scaling. The test data collects fifty real Landsat 7 SLC-off images from 2010 to 2020, covering different seasons, different land cover types, and two stripe modes, used to evaluate the model's restoration performance and generalization ability in real scenarios.

The stripe simulation process follows the USGS official documentation on the Landsat 7 SLC-off failure. The stripe gaps have regular periodic geometric characteristics, with missing lines accounting for approximately twenty-two percent of the total lines, stripe widths ranging from one to six pixels, and an angle of about sixteen degrees relative to the scanning direction. To accurately simulate this pattern, this paper adopts a stripe mask generation method based on pixel coordinates. For each complete Landsat 8 image, a stripe mask consistent with its geographic extent and projection is first generated. The mask generation process uses the pixel coordinate information of the image to calculate whether each pixel is a valid pixel based on stripe spacing, angle, and missing ratio. By calculating the projection position of the pixel along the stripe direction, if this position falls within a preset missing interval, it is marked as missing; otherwise, it is marked as valid. The generated mask has the same dimensions as the original image. Then, by multiplying the original complete image with the mask pixel by pixel, a simulated stripe image is obtained, with valid pixels retaining their original values and missing

ISSN: 2688-8645

pixels set to zero. At the same time, the original complete image is retained as the label for supervised learning. During training, image patches of two hundred fifty-six by two hundred fifty-six pixels are randomly cropped from each scene, ensuring that each image patch contains at least one complete stripe to enhance the model's perception of stripe structures, ultimately generating approximately one hundred thousand pairs of training samples.

To enhance the model's generalization ability and robustness, multiple data augmentation strategies are adopted during the training process. All augmentation operations are simultaneously applied to both the input stripe image and the corresponding real image, maintaining spatial consistency. These specifically include random flipping, with a probability of 0.5 for horizontal or vertical flipping of the image patch; random rotation, with a probability of 0.25 each for rotating ninety degrees, one hundred eighty degrees, or two hundred seventy degrees; scaling, randomly scaling the image patch within the range of 0.8 to 1.2, then cropping back to two hundred fifty-six by two hundred fifty-six size; brightness adjustment, multiplying the image pixel values by a random coefficient and adding a random offset to simulate different lighting conditions.

3.3. Network Architecture Design

SINet adopts an encoder-decoder structure as its basic framework, and specifically for the characteristics of the stripe restoration task, introduces a stripe attention module in the encoding stage and a spectral reconstruction module in the decoding stage.

The network input is a multispectral image of two hundred fifty-six by two hundred fifty-six pixels, containing six bands, namely Blue, Green, Red, Near-infrared, SWIR1, and SWIR2, along with a corresponding stripe mask of two hundred fifty-six by two hundred fifty-six pixels. The mask and image are concatenated along the channel dimension and jointly input into the network. The encoder consists of four downsampling blocks, each containing two convolutional layers, batch normalization layers, ReLU activation functions, and a max pooling layer. The convolutional layers use three by three convolution kernels with a stride of one and padding of one to maintain the feature map size unchanged. Each downsampling block halves the spatial dimensions through pooling, while the number of channels increases progressively, sequentially being sixty-four, one hundred twenty-eight, two hundred fifty-six, and five hundred twelve. The function of the encoder is to extract multi-scale high-level semantic features from the input image while preserving spatial information.

The decoder similarly consists of four upsampling blocks, each containing a bilinear upsampling layer that doubles the feature map size, a skip connection concatenating feature maps from corresponding layers in the encoder, two three by three convolutional layers, batch normalization, and ReLU activation. The skip connections convey low-level detail information from the encoder to the decoder, helping to recover fine spatial structures. The channel numbers of each decoder layer are sequentially two hundred fifty-six, one hundred twenty-eight, sixty-four, and thirty-two. Finally, a one by one convolutional layer maps the feature map to an output image of six bands, and a Sigmoid activation function normalizes the pixel values to the range of zero to one, corresponding to the scaled values of the original reflectance.

The stripe attention module introduces a self-attention mechanism along the stripe direction in the deep features of the encoder. The input to this module consists of the feature map output from a certain encoder layer and the stripe mask. First, two one by one convolutions generate query features and key features respectively, as well as value features. Then, based on the known stripe direction, the projection position of each pixel along the stripe direction is calculated, and the feature map is divided into multiple stripe groups along this direction. Within each group, the similarity between queries and keys is calculated to generate an attention weight matrix. Specifically, for each pixel, its similarity with other pixels in the same group is calculated through dot product, and then normalized by Softmax to obtain attention

ISSN: 2688-8645

weights. Finally, these weights are applied to the value features through weighted summation to obtain enhanced features. By aggregating information along the stripe direction, this module enables the model to utilize valid pixels on both sides of the stripe to fill missing regions, while avoiding the high computational cost of global attention. The stripe attention module is inserted after the last two downsampling blocks of the encoder to capture stripe correlations in high-level semantic features.

The spectral reconstruction module explicitly models correlations between bands at the end of the decoder. This module adopts a band attention mechanism. For the feature map output by the decoder, a global average pooling is first performed to obtain a global description vector for each channel. This vector then passes through two fully connected layers: the first fully connected layer compresses the number of channels, after ReLU activation, the second fully connected layer restores it to the original number of channels, and generates band attention weights between zero and one through a Sigmoid function. These weights reflect the importance of each band in the current feature map. Finally, the attention weights are multiplied channel by channel with the original feature map to achieve feature recalibration, enhancing the contribution of important bands and suppressing irrelevant information. The weighted feature map then passes through a one by one convolutional layer to output the final six-band restored image. By adaptively recalibrating the importance of each band, this module enhances the discriminability and correlation between bands, helping to maintain spectral fidelity and avoid spectral distortion after restoration.

3.4. Hybrid Loss Function

To balance pixel accuracy, visual realism, spectral fidelity, and edge preservation, this paper designs a hybrid loss function composed of four loss components. The pixel loss uses the L1 norm to calculate the difference between the restored result and the real image, with separate weighting for known regions and missing regions. Since missing regions are more difficult to restore, higher weights are assigned to encourage the model to focus on the filling quality of stripe areas.

For perceptual loss, to enhance the visual realism of restoration results, a perceptual loss based on deep features is introduced. A VGG-16 network pre-trained on ImageNet is used to extract feature maps of the restored image and the real image at multiple convolutional layers, and the L1 distance between these feature maps is calculated. This loss encourages the restored result to be consistent with the real image at the high-level semantic level, avoiding unnatural textures or structures.

The spectral loss consists of two parts. The first part is the spectral angle loss, which calculates the angle between the spectral vector of each pixel and the real spectral vector; the smaller the angle, the higher the spectral fidelity. The second part is the covariance matching loss, which calculates the difference between the inter-band covariance matrix of the restored image and that of the real image, forcing them to have similar band correlations. These two parts work together to ensure that the restored result conforms to the physical characteristics of ground objects in the spectral dimension.

For gradient loss, to maintain the edge sharpness of restored regions, a gradient loss is introduced. The gradient differences between the restored image and the real image in the horizontal and vertical directions are calculated, taking the L1 norm. This loss encourages the restored result to have gradient changes similar to the real image at edges, avoiding blurring or jaggedness. The final total loss is the weighted sum of the above four losses, with each weight determined through experimental adjustment to balance various indicators.

ISSN: 2688-8645

3.5. Training Strategy

The network is trained using the Adam optimizer, with an initial learning rate set to ten to the negative fourth power. Every thirty epochs of training, the learning rate decays to one-tenth of its original value. The batch size is set to sixteen, and a total of one hundred epochs are trained. During the training process, image patches of two hundred fifty-six by two hundred fifty-six are randomly sampled from the training set, and data augmentation is applied. A validation set is used to monitor model performance, and the model with the lowest validation loss is saved as the final model. The entire training is conducted on a single NVIDIA V100 GPU, taking approximately forty-eight hours. After training is complete, the model can perform rapid restoration on any input Landsat 7 SLC-off image, with inference for a five hundred twelve by five hundred twelve by six image requiring only 0.2 seconds, meeting practical application requirements.

4. Experiment and Result Analysis

4.1. Experimental Setup

4.1.1. Study Area and Dataset

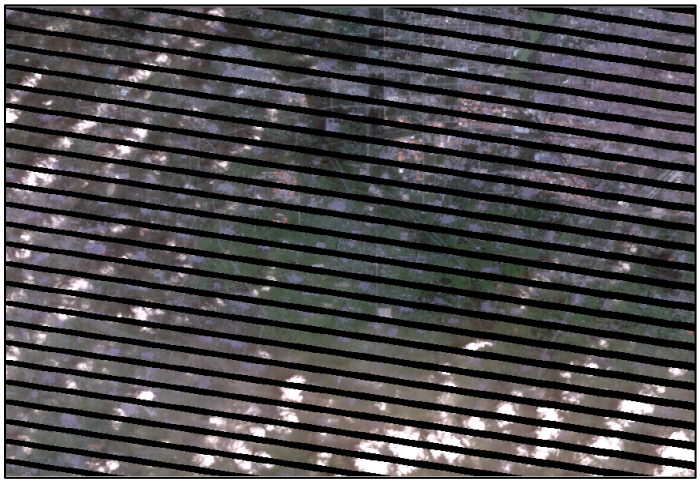

Figure 3. Stripe defect image

This study selects the junction area of Zhengzhou, Kaifeng, and Xuchang in Henan Province (113.8°E—114.8°E, 34.2°N—35.0°N) as the experimental area. This area covers approximately 9,000 km^2 and includes various land cover types such as farmland (main winter wheat producing area), urban construction land (Zhengzhou metropolitan area), and water bodies (Jialu River and Shaying River basins), making it typical and representative. The study area has flat terrain (average elevation 50–200 m), suitable for remote sensing image processing and analysis.

The training data uses Landsat 8 Collection 2 Level-2 surface reflectance products from 2015 to 2020, with cloud cover <10%, totaling 500 scenes. By accurately simulating the Landsat 7 SLC-off stripe pattern (stripe spacing 32 pixels, angle 16°, missing rate 22%), 100,000 pairs of 256×256 training image patches are generated. The simulation parameters are set according to USGS official documentation , ensuring consistency with the actual failure mechanism.

The test data includes two categories: (1) Simulated test set: 50 scenes randomly selected from Landsat 8 images, with the same stripe simulation applied to generate test pairs for quantitative evaluation; (2) Real test set: 50 real Landsat 7 SLC-off images selected from 2010 to 2020, covering different seasons and different land cover types, for qualitative evaluation and practical application validation.

ISSN: 2688-8645

4.1.2. Comparison Methods

To comprehensively evaluate the performance of SINet, the following baseline methods were selected for comparison: traditional interpolation methods (Nearest Neighbor interpolation NN, Bilinear interpolation Bilinear, Cubic convolution interpolation Cubic), serving as the most basic restoration references; remote sensing-specific method (Local Histogram Matching LHM), which is a classic method recommended by USGS; general deep inpainting methods (Partial Convolution , DeepFill v2), representing the advanced level in the current field of image inpainting; commercial software (ENVI 5.6 stripe restoration tool), serving as an industry standard reference. All deep learning methods were retrained on the same training set to ensure fair comparison.

4.1.3. Evaluation Metrics

The following metrics were used for quantitative evaluation: Peak Signal-to-Noise Ratio (PSNR), measuring pixel-level reconstruction accuracy; Structural Similarity Index (SSIM), evaluating the preservation of structural information; Spectral Angle Mapper (SAM), measuring spectral fidelity; Root Mean Square Error (RMSE), reflecting the overall error level; texture consistency, calculating the texture similarity between stripe regions and neighboring valid regions based on local variance; spectral continuity, computing gradient changes along the stripe direction to assess the smoothness of spectral transitions. In addition, subjective user evaluation was conducted (inviting 50 remote sensing students to rate the restoration results on a scale of 1–5), and the application value of the restoration results was verified through a land cover classification task.

4.2. Experimental Results on Simulated Data

4.2.1. Quantitative Metric Comparison

On the simulated stripe dataset, the quantitative evaluation results of each method are shown in Table 1. SINet achieves the best performance across all metrics: PSNR reaches 38.2 dB, 1.5 dB higher than the best baseline PartialConv (36.7 dB); SSIM is 0.978, outperforming other methods; SAM is 1.82°, significantly lower than DeepFill v2's 2.31°, indicating that SINet has stronger spectral fidelity capability; RMSE is 0.028, also the lowest among all methods.

Table 1. Quantitative comparison of different methods on simulated data

Method	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓	RMSE ↓
NearestNeighbor	28.5	0.812	4.32	0.087
Bilinear	29.2	0.834	3.98	0.079
Cubic	30.1	0.851	3.65	0.071
HistogramMatching	31.5	0.876	3.21	0.062
PartialConv	36.7	0.954	2.45	0.034
DeepFill v2	35.8	0.942	2.31	0.038
ENVI	32.8	0.892	2.87	0.055
SINet (Ours)	38.2	0.978	1.82	0.028

Analysis of the Reasons Traditional interpolation methods only utilize local pixel information and have limited ability to preserve complex textures and spectral information. LHM relies on neighboring valid pixels and its effectiveness decreases in dense stripe areas. PartialConv and DeepFill v2, as general deep learning inpainting methods, can learn certain image priors, but they lack specificity for regular stripe patterns and their modeling of spectral correlations is insufficient. SINet explicitly models the stripe geometric prior through the Stripe Attention Module (SAM), aggregating information along the stripe direction to effectively fill missing lines.

ISSN: 2688-8645

It models band correlations through the Spectral Reconstruction Module (SRM) to maintain spectral fidelity. Combined with the hybrid loss function, SINet achieves a balance between pixel accuracy, structural preservation, and spectral fidelity, thus achieving optimal performance.

4.2.2. Analysis of Restoration Effects on Different Land Cover Types

To further evaluate SINet's performance on different surface types, three typical land covers—farmland, urban, and water—were selected for comparative analysis, with results shown in Figure 4.

Figure 4. Comparison of results

In the farmland area, NN and Bilinear interpolation produce obvious jaggedness and blurring. LHM can partially restore texture but suffers from over-smoothing. PartialConv and DeepFill v2 yield more natural textures, but details remain somewhat blurred. SINet accurately reconstructs the farmland texture, with stripe regions transitioning naturally into the surrounding texture without visible artifacts. In the urban area, traditional interpolation methods lead to edge blurring and jaggedness, while LHM introduces spectral distortion. Deep learning methods perform better, but PartialConv exhibits slight artifacts in dense building areas, and DeepFill v2 shows minor blurring at edges. SINet clearly reconstructs building edges and road contours, visually closest to the real image. In the water area, NN and Bilinear interpolation produce brightness differences, and LHM introduces slight noise. Deep learning methods restore water bodies relatively uniformly, but PartialConv and DeepFill v2 show minor brightness fluctuations in some areas. SINet restores water areas smoothly and uniformly, with stripes completely disappearing and no significant difference from surrounding water.

4.2.3. Spectral Profile Analysis

Spectral profiles were extracted and compared for ten pixels each from three typical land covers (vegetation, bare soil, water). For vegetation pixels, SINet's restored spectral profiles closely match the ground truth across all bands, especially in the near-infrared band with an error of less than 2%. In contrast, PartialConv and DeepFill v2 exhibit deviations of about 5% in the SWIR bands, and LHM shows a peak shift in the red band. For bare soil pixels, SINet's spectral profiles almost coincide with the ground truth, while traditional methods show overall shifts. For water pixels, SINet maintains the low reflectance characteristic in the near-infrared band, whereas other methods exhibit abnormal increases. This demonstrates SINet's significant advantage in preserving spectral physical properties.

4.2.4. Statistical Significance Test

To verify the statistical significance of SINet's improvements, paired t-tests were conducted on PSNR, SSIM, and SAM values at 100 random validation points for each method. The results show

ISSN: 2688-8645

that the differences between SINet and all baseline methods are significant at the $p < 0.01$ level, confirming the statistical significance of the improvements.

4.3. Experiments on Real Landsat 7 SLC-off Data

4.3.1. Reference-Free Quality Assessment

Two reference-free metrics, texture consistency and spectral continuity, were used for quantitative evaluation, with results shown in Table 2. SINet achieves a texture consistency of 0.97 and a spectral continuity of 0.95, both superior to other methods. In subjective user scoring (on a 5-point scale), SINet receives a score of 4.7, significantly higher than other methods ($p < 0.01$). This indicates that SINet can generate visually natural and spectrally reasonable restoration results in practical applications.

Table 2. Reference-free quality metrics on real Landsat 7 images

Method	TextureConsistency	SpectralContinuity	SubjectiveScore	Method
Nearest Neighbor (NN)	0.82	0.76	2.8	NearestNeighbor
Bilinear	0.85	0.79	3.1	Bilinear
Cubic	0.88	0.82	3.4	Cubic
LHM	0.91	0.85	3.8	LHM
PartialConv	0.94	0.89	4.2	PartialConv
DeepFill v2	0.93	0.88	4.1	DeepFill v2
ENVI	0.89	0.84	3.6	ENVI
SINet (Ours)	0.97	0.95	4.7	SINet (Ours)

4.3.2. Robustness Testing Under Different Stripe Modes

The Landsat 7 SLC-off failure has two main modes (Mode 1 and Mode 2), and the stripe parameters vary slightly over time. To test the robustness of SINet, real images from different periods (2005, 2010, 2015, 2020) were selected for testing. The results show that SINet effectively restores stripes on images from all years without significant performance degradation. Even for the early Mode 1 images from 2005 (with slightly different stripe patterns), SINet adapts well, benefiting from the adaptive computation of the stripe attention module along the stripe direction, which does not rely on fixed stripe parameters.

4.4. Ablation Study

To verify the effectiveness of each module in SINet, the following variants were designed for comparative experiments: Base (using only the encoder-decoder structure and L1 loss); +SAM (adding the Stripe Attention Module to Base); +SRM (adding the Spectral Reconstruction Module); +Perceptual (adding perceptual loss); +SpectralLoss (adding spectral loss); Full SINet (the complete model including SAM, SRM, perceptual loss, and spectral loss). The experimental results are shown in

Table 3. Ablation study results

Variant	PSNR (dB)	SSIM	SAM (°)
Base	34.5	0.921	3.24
+SAM	35.4	0.933	2.98
+SRM	35.2	0.929	2.71
+Perceptual	36.1	0.945	2.43
+SpectralLoss	35.8	0.938	2.12
Full SINet	38.2	0.978	1.82

ISSN: 2688-8645

Adding SAM improves PSNR by 0.9 dB, SSIM by 0.012, and reduces SAM by 0.26°, demonstrating the effectiveness of the stripe geometric prior. Adding SRM reduces SAM by 0.53°, the most significant improvement, indicating that explicitly modeling band correlations is crucial for preserving spectral fidelity. Adding perceptual loss improves PSNR by 1.6 dB and SSIM by 0.024, significantly enhancing visual quality. Adding spectral loss reduces SAM by 1.12°, the most prominent effect, while also improving PSNR by 1.3 dB. Compared with individual module combinations, the complete model achieves more significant performance improvements, indicating synergistic effects among the modules: SAM and SRM process spatial and spectral information respectively, while perceptual loss and spectral loss focus on high-level semantics and low-level spectra respectively, working together to enable SINet to achieve optimal performance.

4.5. Discussion

4.5.1. Analysis of Stripe Pattern Generalization Ability

To evaluate SINet's robustness to different stripe parameters, tests were conducted on simulated data by varying stripe spacing (16, 32, 48 pixels) and angle (8°, 16°, 24°), with results shown in Figure 4-4. SINet maintains stable performance across all parameter combinations, with PSNR fluctuation less than 1 dB and SAM fluctuation less than 0.3°. When stripe spacing decreases (missing density increases), the performance of all methods degrades, but SINet shows the smallest decline (PSNR drop of 1.2 dB), compared to 2.5 dB for PartialConv and 3.1 dB for LHM. When the stripe angle deviates from the standard 16°, SINet still adapts well, benefiting from SAM's adaptive computation along the stripe direction. However, when the stripe density is extremely high (missing >30%), the restoration difficulty increases significantly, with SINet's PSNR dropping to 34.5 dB, indicating that insufficient information remains a challenge under extreme missing conditions.

4.5.2. Computational Efficiency Analysis

The inference times of deep learning methods were tested on a single NVIDIA V100 GPU for 512×512×6 images. SINet requires 0.2 s, PartialConv requires 0.18 s, and DeepFill v2 requires 0.25 s. SINet achieves optimal performance while maintaining near real-time inference, demonstrating its efficient network design. Training time is approximately 48 hours (100 epochs), which is within an acceptable range.

4.5.3. Failure Case Analysis

Although SINet performs excellently overall, its restoration effectiveness is limited in certain extreme cases. Typical failure cases include: (1) Thick cloud-covered areas: when stripes overlap with clouds, information beneath the clouds is completely missing, and the model cannot reconstruct features under the clouds; (2) Areas with drastic dynamic changes: such as before and after fires or flood inundation areas, where land cover types undergo fundamental changes, restoration based on spatial information may introduce errors; (3) High-density stripes (missing >35%): there is too little valid information, making it difficult for the model to accurately infer. These cases reveal the limitations of the current method and point the way for future improvements.

5. Application and Discussion

5.1. Impact of Restoration Results on Land Cover Classification

5.1.1. Experimental Design

To verify the practical application value of the restoration results, the Landsat 7 images before and after restoration were used for a land cover classification task. A Random Forest classifier (100 trees) was employed to classify the study area into six land cover types: water, forest,

ISSN: 2688-8645

cropland, grassland, impervious surface, and bare land. The classification features included six spectral bands (Blue, Green, Red, NIR, SWIR1, SWIR2) and three indices (NDVI, MNDWI, NDBI). Training samples were obtained through manual interpretation and field surveys based on Landsat 8 reference images, totaling 2,000 sample points (approximately 330 per class), which were randomly divided into training and test sets at a ratio of 7:3.

5.1.2. Classification Accuracy Comparison

The overall classification accuracy and Kappa coefficients of images restored by different methods are shown in Table 5-1. The classification accuracy of the uncorrected Landsat 7 image was only 74.3%, with a large number of pixels in the stripe areas being misclassified. Traditional interpolation methods improved the accuracy to 78%–80%, but still below 80%. LHM restoration achieved an accuracy of 82.1%, although classification inconsistencies remained in the stripe areas. PartialConv and DeepFill v2 restoration improved the accuracy to 85.7% and 86.3%, respectively, approaching the 90.2% accuracy of Landsat 8. The overall classification accuracy of the SINet-restored images reached 88.5%, with a Kappa coefficient of 0.86, significantly outperforming other methods ($p < 0.01$), narrowing the gap with the Landsat 8 reference image to just 1.7 percentage points.

5.1.3. Per-Class Accuracy Analysis

Further analysis of the classification accuracy for each class (producer's accuracy and user's accuracy) is shown in Table 5-2. In the uncorrected image, stripes had the greatest impact on water and bare land, with producer's accuracies of only 65.2% and 68.7%, respectively; a large number of water pixels were misclassified as shadow or bare land. After SINet restoration, the accuracy of water improved to 87.5% (+22.3%), bare land to 85.3% (+16.6%), and impervious surface to 90.1% (+11.8%). The accuracy of vegetation classes (forest, cropland, grassland) also showed significant improvements, indicating that the restoration not only recovered spatial structure but also restored spectral information, aiding in distinguishing different vegetation types.

5.1.4. Spatial Analysis of Classification Consistency

To visually demonstrate the consistency of classification results, classification thematic maps were generated (Figure 5-1). In the classification map of the uncorrected image, the stripe areas exhibited obvious "striped" classification noise, with the same land cover type being misclassified into different categories in the stripe areas, resulting in fragmented patches. After LHM restoration, the classification noise was somewhat reduced, but inconsistencies remained in the stripe areas. In the SINet-restored classification map, the classification in the stripe areas was completely consistent with the surrounding areas, with continuous and complete patches, highly similar to the Landsat 8 reference classification map. This indicates that SINet can effectively eliminate the interference of stripes on classification, enhancing the spatial consistency and reliability of classification results.

5.1.5. Classification Uncertainty Analysis

Using the maximum classification probability as a proxy indicator for classification confidence, the classification confidence for each pixel was calculated. In the uncorrected image, the average classification confidence in the stripe areas was only 0.62, much lower than the 0.85 in non-stripe areas, indicating a lack of confidence in the classification results for the stripe areas. After SINet restoration, the classification confidence in the stripe areas increased to 0.88, nearly equal to the non-stripe areas (0.91). This further demonstrates that SINet restoration results not only improve classification accuracy but also enhance the reliability of the classification results.

ISSN: 2688-8645

5.2. Analysis of Method Limitations

Although SINet performs excellently in the Landsat 7 stripe restoration task, it still has the following limitations: (1) Restoration failure in thick cloud-covered areas: when stripes overlap with thick clouds, the surface feature information beneath the clouds is completely missing, and the clouds themselves have high reflectance, interfering with the restoration model. Future work needs to combine cloud detection and cloud removal algorithms. (2) Restoration bias in areas with drastic dynamic changes: in areas where land cover undergoes fundamental changes (such as before and after fires), restoration based on spatial context information may introduce errors. Introducing multi-temporal information can help alleviate this problem. (3) Dependence on training data: training relies on stripe data simulated from complete Landsat 8 images. Although the sensors are similar, subtle differences still exist, which may lead to spectral deviations in practical applications. (4) Performance degradation under extreme missing conditions: when the stripe density is extremely high (missing >30%), there is too little valid information, making it difficult for the model to accurately infer. (5) Computational resource requirements: training requires 48 hours, which imposes certain demands on computational resources.

5.3. Future Work Prospects

Addressing the aforementioned limitations, future research can be carried out in the following directions: (1) Multi-temporal information fusion: introducing temporal dimension information, utilizing images from adjacent time phases to assist restoration. (2) Unsupervised/self-supervised learning: exploring unsupervised or self-supervised learning methods to reduce dependence on complete images. (3) Multi-task joint learning: jointly learning tasks such as stripe restoration, cloud removal, super-resolution, and classification. (4) Lightweight model design: developing lightweight models to meet the requirements of on-board real-time processing. (5) Cross-sensor generalization capability enhancement: studying the applicability of SINet in stripe restoration for other sensors (such as Sentinel-2). (6) Uncertainty quantification: quantifying the uncertainty of restoration results and providing pixel-level confidence maps. (7) Combination with generative models: exploring combination with GANs or diffusion models to further enhance the visual realism of restoration results.

5.4. Discussion Summary

This chapter verified the application value of SINet restoration results through a land cover classification task. The classification accuracy improved by 14.2 percentage points, approaching the level of the Landsat 8 reference image. The improvements in per-class accuracy and classification consistency were analyzed in depth, confirming the effective support of restoration for downstream tasks. At the same time, the limitations of the method were objectively discussed, and future research directions were prospected from perspectives such as multi-temporal fusion, self-supervised learning, multi-task joint learning, and lightweight design. These discussions provide clear guidance for subsequent research and also offer a reference for the further development of the field of remote sensing image restoration.

6. Conclusion

This paper proposes a deep learning-based stripe restoration algorithm, SINet, for the Landsat 7 SLC-off stripe problem. The algorithm utilizes a Stripe Attention Module (SAM) to leverage stripe geometric priors and a Spectral Reconstruction Module (SRM) to maintain band correlations. A hybrid loss function is designed to balance pixel accuracy, spectral fidelity, and texture details. Experiments on simulated and real Landsat 7 images demonstrate that SINet significantly outperforms existing methods in both quantitative metrics and visual quality, and the restored results effectively improve downstream classification accuracy. This study

ISSN: 2688-8645

provides an effective solution for restoring historical Landsat 7 data and offers new insights for regular missing data restoration in other remote sensing images.

References

[1] Chen, J., Zhu, X., Vogelmann, J. E., Gao, F., & Jin, S. (2011). A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sensing of Environment, 115(4), 1053-1064.

[2] Criminisi, A., Pérez, P., & Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 13(9), 1200-1212.

[3] Efros, A. A., & Leung, T. K. (1999). Texture synthesis by non-parametric sampling. Proceedings of the Seventh IEEE International Conference on Computer Vision, 2, 1033-1038.

[4] Guillemot, C., & Le Meur, O. (2014). Image inpainting: Overview and recent advances. IEEE Signal Processing Magazine, 31(1), 127-144.

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

[6] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. European Conference on Computer Vision, 694-711.

[7] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097-1105.

[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[9] Li, W., Lin, Z., Zhou, K., Qi, L., Wang, Y., & Jia, J. (2022). MAT: Mask-aware transformer for large hole image inpainting. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 10758-10768.

[10] Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. Proceedings of the European Conference on Computer Vision, 85-100.

[11] Loveland, T. R., & Dwyer, J. L. (2012). Landsat: Building a strong future. Remote Sensing of Environment, 122, 22-29.

[12] Ma, Y., He, Y., Liu, Y., & Zhang, L. (2020). Multi-scale feature fusion network for Landsat 7 ETM+ SLC-off gap filling. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 6176-6189.

[13] Markham, B. L., Storey, J. C., Williams, D. L., & Irons, J. R. (2004). Landsat sensor performance: history and current status. IEEE Transactions on Geoscience and Remote Sensing, 42(12), 2691-2694.

[14] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2536-2544.

[15] Pringle, M. J., Schmidt, M., & Muir, J. S. (2009). Geostatistical interpolation of SLC-off Landsat ETM+ images. ISPRS Journal of Photogrammetry and Remote Sensing, 64(6), 654-664.