Remote Sensing Single Image Super-Resolution Benchmarking with Transfer Learning Algorithms

Elangoraj Thiruppandiaraj
Department of Computer Science,
Faculty of Environment, Science and Economy,
University of Exeter, North Park Road,
Exeter, Devon EX4 4QE, United Kingdom.
Email: elangorj@gmail.com

Saptarshi Das, *Member, IEEE*
Centre for Environmental Mathematics,
Faculty of Environment, Science and Economy,
University of Exeter, Penryn Campus,
Cornwall TR10 9FE, United Kingdom.
Email: saptarshi.das@ieee.org, s.das3@exeter.ac.uk

***Abstract*—In the context of real-world applications like medical imaging systems, tracking, astronomical imaging, navigation, and remote sensing (RS), there is a pressing need to enhance or upscale images with minimal errors. This is particularly critical for tasks such as target detection, image classification, and land use mapping. However, remote sensing images often suffer from limitations in spatial, spectral, radiometric, and temporal resolution due to complex atmospheric conditions and sensor constraints. Additionally, acquiring these images can be expensive and time-consuming. In this study, we propose a Single Image Super-Resolution (SISR) method to address these challenges by upscaling low-quality remote sensing images to higher resolution, enabling a better understanding of these images. We also discuss the specific challenges in remote sensing super-resolution techniques and review various upscaling approaches, while analyzing the impact of other factors like weather conditions, image capture time, and different scene types on the technique's effectiveness.**

***Index Terms*—Image enhancement, remote sensing, super-resolution (SR), convolutional neural networks (CNNs)**

I. Introduction

Remote sensing provides details about the objects that are closer to the earth's surface and atmosphere based on the emitted or reflected radiation from them. These details are collected from a certain distance from the land in the form of images, and videos. It helps to automatically analyze the properties of the object, land cover from local to global scales using traditional and deep learning approaches [1]. The government and private sector continuously use these data for weather reporting, traffic monitoring, drought management, floods, pollution, climate change, etc. With the advancement in this space, there is a possibility of having a large number of geographical images across the earth's surface [2], [3]. However, these remote sensing images are often distorted due to the limitation of weather, sensors, optical aberration, and noise captured along with in the imaging system. These disturbances lead to difficulties in feature extraction, finding targets, updating the maps, etc. [4].

The motivation of this paper is to learn from remote sensing images which poses many challenges to the industries and stakeholders trying to learn patterns from remote sensing data. Many research organisations and companies applied earth intelligence and analytics, using openly available remote sensing data. The purpose is to transform geospatial data into actionable insights about the earth through the delivery of innovative data solutions. Remote sensing industries and researchers face many common challenges of high-quality upscaling of RS images, which may serve multiple purposes like biodiversity monitoring, surveillance for natural disasters, land cover mapping, forest fire detection etc.

In this study, single-image super-resolution (SISR) constructs a high-resolution image with a single low-resolution image. For example, if an image of size 256x256 is fed to the deep learning algorithm, it should be able to double the image size to 512x512 and double even more to 4 times which is 1024x1024 sized image. The optimization target for the super-resolved image is the minimization of mean squared error (MSE) between the generated images. When the MSE is reduced it increases the PSNR, a standard metric that is used to quantify the noise [5].

II. The Research Context

Researching super-resolution techniques brings one basic method in which the feature vectors of the low-resolution images are expressed as a weighted combination of its k-nearest neighbors. In the assumption that low-resolution image embedding is preserved in correspondence to the high-resolution feature vectors [6]. The rise of deep learning techniques has been fast in the computer vision related tasks in the last decade with significant improvement in speed and accuracy. Super-resolution can be performed from neural networks such as multilayer perceptrons (MLPs), Deep Neural Networks (DNNs), and Convolutional Neural Networks (CNNs). DNNs with several hidden layers able to super-resolved images [7] the CNNs are effective in several tasks such as image learning, object detection, image sharpening, RS imagery enhancement, and image super-resolution [8].

In CNN the model architecture becomes deeper and complex which leads to the vanishing gradient problem. At the end of the network, the gradient information reaches with minimal value [9]. An effective workaround for this problem is to add skip layers or bypass connections which allows

979-8-3503-0551-7/23/$31.00 ©2023 IEEE

signals to flow between layers. This method was proposed in the highway network and ResNet Network. In the case of DenseNet, it connects different layers within the network, enabling connected layers to accept inputs from the previous layer. This is same with the ResNet network in which it defines the residual function to the layer as the reference input [10], [11].

III. Aims and Objectives

The present research has multiple objectives to produce the desired results. Based on the research and discussion with the remote sensing industries and researchers the following are the key points to be addressed in this paper:

- To create a SSIR model using the deep CNNs.
- The model trained and developed should be able to upscale images by a factor of 2x upscale and 4x upscale size from the low-resolution input image.
- Further, the model has to be generalized across various land-covers (water, buildings, beach, etc), time (day, night), and weather conditions (clear, cloud, snow) [12].
- Design multiple error metrics (SSIM, PSNR, NRMSE) to evaluate the super-resolved images with the ground truth data available [13], [14].
- Testing various hypothesis on model performance using statistical metrics.

IV. Experiment Designm Materials and Methods

A. Dataset Description

The data is collected from an open-source image database CVonline [5] which provides a collection of images from unmanned aerial vehicles (UAVs) and its respective satellite images. These images are from 13 different places in Asia and Europe, such as Birmingham, Coventry, Suzhou, Renens, Le Bourget Airport (Paris), and others. The considered dataset has images that are equal in size (960x720). Our dataset has 10 different scenes (agriculture, airport, beach, buildings, forest, land, parking, playground, road, water), 2 different times (day, and night), and 3 weather conditions (clear, cloud, snow) when the images were observed.

B. Data Pre-processing

There are 1934 images in total in the dataset of size 960x720 and are treated as the target images. So the input to the model were the rescaled and blurred version of the target images. For the given scaling factor, the input image were cropped, downscaled to create low-resolution images, and converted to imagenet_norm format. The dataset was split with 1720 samples for training and 214 for testing.

C. Workflow and Deep Learning Principles

Approach 1: Initially, we use a Deep CNN for 2x upscaling with kernel size of 3, 8 convolution layers, and 32 filters. The activation unit was rectified linear unit (ReLU) with a final deconvolution layer. Model performance was under par which faced the vanishing gradient problem.

Approach 2: To address the vanishing gradient problem, skip connections, PreLU activation and residual blocks were incorporated into the network. This modification significantly improved the performance. However, when applying the same setup for 4x upscaling, performance remained subpar.

Approach 3: Through multiple iterations, we refined the number of layers, kernel and filter sizes, and activation units. We introduced batch normalization for data normalization and leveraged residual blocks, greatly boosting the model performance. This approach was proved to be effective for both 2x and 4x upscaling, finalizing the architecture design.

D. CNN Architectural Design

The network consists of multiple components with the initial layer being the convolution layer (size 64), kernel (size 9), and parametric rectified linear unit (PReLU) as the activation unit. Then there are 16 residual blocks present in the model. Each block consists of 2 convolution layers with PReLU activation and two batch norm layers given to the skip connection. Finally, sub-pixel convolution or deconvolution blocks are placed to upscale the images from the layers [15].

1) Convolution Layer: In a convolution layer, weights are multiplied with inputs through a linear operation, a process repeated multiple times. This layer analyzes the influence of nearby pixels using filters, typically smaller than the input, which are applied as dot products to image patches, resulting in a single scalar value [16]. We employ 64 filters and a kernel size of 9 in the initial convolution layer, while the residual block utilizes 64 filters with a kernel size of 3.

2) Batch Norm: Training deep neural networks is complex due to changing input distributions across layers, slowing down training and causing internal covariance shifts. Batch normalization addresses this by introducing a network layer that controls the first two moments (mean and variance) of these distributions, stabilizing them during training. This enables more predictable gradients, allowing for larger learning rates and faster convergence. Typically placed between hidden layers, batch norm takes the output from the preceding layer, normalizes it using two sets of parameters (learnable beta and gamma, and non-learnable mean and variance moving averages) [17].

3) PReLU: The ReLU is a one-sided activation function that sets negative neuron values to zero, but it suffers from the problem of ”neuronal death” when many neurons become negative, rendering their values to be ineffective. To address this issue, the Parametric Rectified Linear Unit (PReLU) is employed, introducing a learnable parameter that responds to negative values, preventing neuronal death. Essentially, it’s a ReLU activation with adjustable parameters [18].

V. Results

To assess the model performance, various general error and image quality metrics are applied, facilitating the evaluation of super-resolved image quality comparing with the ground truth images. Additionally, multiple graphs, charts, and hypotheses are utilized to evaluate results under different conditions.

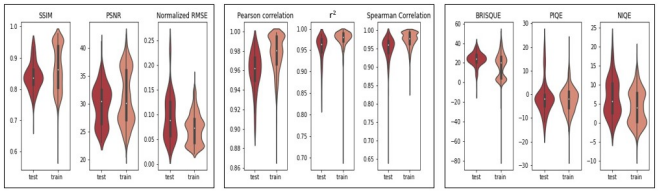

Fig. 1. SISR 2x model error distribution violin plots.

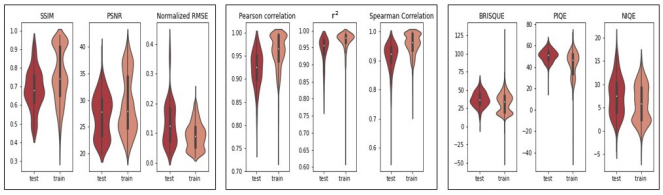

Fig. 2. SISR 4x model error distribution violin plots.

A. Image Comparison Metrics

Structural Similarities (SSIM): The human visual perception is highly capable of identifying the structural information from a scene and finding the difference in the referenced image. SSIM works by extracting the image properties such as structure, contrast, and luminance of the two images. It performs better on tasks that involve differentiating between two images as in our case [19], given by:

$$\text{SSIM} = \frac{((2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2))}{((\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_1))}. \tag{1}$$

Peak Signal to Noise Ratio (PSNR): Mean square error is used to calculate the PSNR as shown in equation (2) and indicates the ratio of maximum pixel intensity to the power of distortion. PSNR value reaches infinity when the MSE approaches zero. Hence, higher PSNR value means a higher similarity between two images [20]. The PSNR is given by:

$$\text{PSNR} = 10\log_{10}\left(\frac{R^2}{\text{MSE}}\right). \tag{2}$$

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE): A pre-trained model containing a database of images with no noise and hence it fails to detect the distortion which is unseen. Subjective quality scores come with the training data as it is opinion-aware [21].

Fig. 3. a) Ground truth, b) predicted image, c) difference in pixel between ground truth and predicted 2x upscaling.

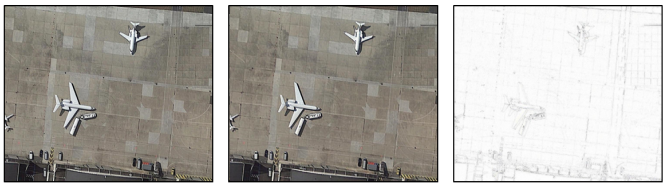

Fig. 4. a) Ground truth, b) predicted image, c) difference in pixel between ground truth and predicted 4x upscaling.

Image scene	Average of SSIM	Average of PSNR
Agriculture	0.8108	27.7904
Airport	0.8879	31.4834
Beach	0.9196	33.0430
Buildings	0.8552	29.5510
Forest	0.8396	30.9009
Land	0.8802	33.6734
Parking	0.8702	30.2096
Playground	0.8698	31.8662
Road	0.8901	33.0660
Water	0.8862	32.1204

Image scene	Average of SSIM	Average of PSNR
Agriculture	0.5844	24.9742
Airport	0.7608	28.6035
Beach	0.8781	33.4427
Buildings	0.7466	27.5010
Forest	0.6963	28.7400
Land	0.7921	32.2157
Parking	0.7643	28.2997
Playground	0.8000	30.9382
Road	0.8232	31.7126
Water	0.8141	30.4455

Fig. 5. SISR model average SSIM and PSNR a) 2x upscaling b) 4x upscaling.

Natural Image Quality Evaluator (NIQE): It is also a pre-trained model on a collection of pristine images. It has the capability of measuring the arbitrary distortion of the image. It does not use a subjective quality score and it is opinion-aware [22].

Perception-based Image Quality Evaluator (PIQE): It does not require a trained model since it is an unsupervised method. It can measure the arbitrary distortion of the image and it is an opinion-unaware algorithm [23] given by:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}. \tag{3}$$

The above error metrics along with normalized root mean squared error (NRMSE), Pearson's correlation coefficient in equation (3), Spearman correlation coefficient in equation (4), and the coefficient of determination (R^2) in equation (5) are used to evaluate the SISR 2x and 4x model using the residual/total sum of square (RSS/TSS):

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{4}$$

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}. \tag{5}$$

The results are plotted in the form of a violin plot as shown in Fig. 1 and Fig. 2 for the training and testing set.

1) SISR 2x Model: The mean SSIM for the model in the testing-set is 84% and training-set is 87% which means that the images are highly structurally similar. The violin plot in Fig. 1 shows the results. The bulk of the PSNR values are distributed around 30 dB for both groups and hence there is minimal noise in the reconstructed images. Both the correlation metrics and R^2 are mostly over 0.95 which shows the similarity of the images. BRISQUE, PIQUE, and NIQE are calculated by the

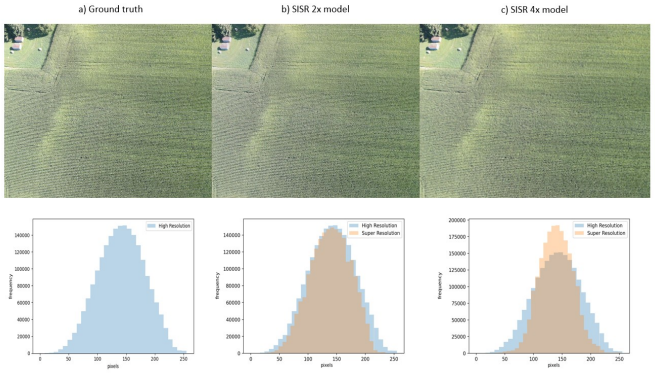

Fig. 6. Row 1 has ground truth, 2x up-scaled, and 4x up-scaled image. Row 2 shows the pixel histogram plot for the same ground truth, 2x up-scaled, and 4x up-scaled image.

difference in the predicted image error and test data error. Hence, it's expected to be closer to zero as shown in the plots.

2) SISR 4x Model: From the graph in Fig. 2 it is seen that the SSIM, PSNR values have dropped slightly. Yet the overall SSIM in the test-set is closer to 70% and PSNR is around 28dB. The normalised RMSE has also increased. The correlations are distributed over 0.9. BRISQUE, PIQUE, and NIQE show some differences in noise level but the image retrieved is still of high quality in terms of visual similarity to human eyes.

B. Difference Between Ground Truth and Super-resolved

1) SISR 2x Model: Fig. 3 displays a 2x super-resolved beach scene, where the disparity between the original image and the model's output is scarcely noticeable. The pixel difference map also reveals minimal discrepancies, primarily along the borders, such as the road, beach, and buildings. This consistency is reflected in Fig. 1, where beach images exhibit the highest SSIM score of 0.92.

2) SISR 4x Model: In Fig. 4, a notable level of granularity is evident in the airport scene. However, the predicted image closely resembles the expected image to human eyes. The pixel difference map highlights variances along building borders, aircraft, and the floor texture. Lower SSIM and PSNR scores for airport images in the dataset are attributed to variations in the surface patterns.

C. Pixel Distribution

The graph depicted in Fig. 6 illustrates the pixel distribution histogram for a single image, comparing ground truth and super-resolved versions. In this representation, 'high-resolution' refers to the ground truth image, while 'super-resolution' corresponds to the output of the SISR 2x or 4x model. Notably, the last two histograms exhibit overlapping regions, indicating similarities between the two images.

1) SISR 2x Model: Fig. 6b is the 2x model result where the performance for an agriculture scene is compared. The reproduced image shows very minimal noise and looks very close to the expected image. Hence the pixels in both the high/low resolution (HR/SR) have almost similar distribution.

TABLE I
WEATHER HYPOTHESIS STATISTICS

$Model$	$Name$	Df	$Sum-Sq$	$Mean-Sq$	$F-value$	$pr(>F)$
2x upscale	Weather	2	0.026	0.0129	2.5680	0.0769
	Residuals	1931	9.6790	0.0050		
4x upscale	Weather	2	0.35	0.1748	8.525	0.0002
	Residuals	1931	39.60	0.0205		

2) SISR 4x Model: Visual comparison of Fig. 6c shows some noise in the 4x result. The input image of size 240x180 is converted into 960x720 with some inevitable noise. The histogram shows that SR image pixels are around the center whereas HR pixels are widely distributed with heavy tail. This variation shows the difference in pixels of the two images.

D. Inferential Statistics

To evaluate the performance further for the two SISR models 2x and 4x two hypotheses are considered. The most common image similarity error metric SSIM is used for this purpose. The confidence level is checked and set to the acceptable standard of 95%.

1) Weather Hypothesis:

Null: The SSIM error mean for weather conditions clear, cloudy, and snowy do not differ from the group mean.

Alternate: The SSIM error mean for weather conditions clear, cloudy, and snowy do differ from the group mean.

In Table I, the statistical results for the 2x model are presented, yielding a p-value of 0.0769, exceeding the predefined significance level of 0.05. Therefore, the alternate hypothesis can be rejected, indicating that, statistically, weather conditions do not impact the 2x image super-resolution model.

In Table I, the statistical analysis for the 4x model reveals a p-value of 0.000206, which is below the 0.05 threshold. Consequently, we can reject the null hypothesis, demonstrating that the 4x model predictions are statistically influenced by varying weather conditions in the images.

2) Time Hypothesis:

Null: The SSIM error means for images taken during the day and night do not differ from the group mean.

Alternate: The SSIM error means for images taken during the day and night do differ from the group mean.

In Table II, the p-value for the 2x model is 0.0879, surpassing the predetermined cutoff, leading to the rejection of the alternate hypothesis. Statistically, the time of day when the image was taken (day or night) does not significantly impact the quality of the 2x super-resolution image. Likewise, in Table II, the p-value for the 4x model is 0.118, exceeding the confidence threshold, allowing us to reject the alternate hypothesis.

VI. DISCUSSIONS

A. Main Findings and Comparison of Results

In this section, we perform a comparison between the newly developed SISR model and existing transfer learning models. We assess their performance using multiple error metrics,

TABLE II
TIME HYPOTHESIS STATISTICS

$Model$	$Name$	Df	$Sum-Sq$	$Mean-Sq$	$F-value$	$pr(>F)$
2x upscaling	Time	1	0.015	0.0146	2.916	0.0879
	Residuals	1932	9.690	0.0050		
4x upscaling	Time	1	0.05	0.504	2.441	0.118
	Residuals	1932	39.90	0.0206		

Fig. 7. 2x upscaling: Ground truth, SISR, EDSR, LapSRN, and FSRCNN.

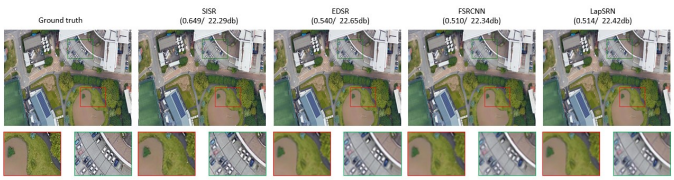

Fig. 8. 4x upscaling: Ground truth, SISR, EDSR, LapSRN, and FSRCNN.

including SSIM, PSNR, normalized RMSE, Pearson's correlation coefficient, and R^2. Specifically, we evaluate models such as the Enhanced Deep Residual Network (EDSR), Fast Super-Resolution Convolutional Neural Network (FSRCNN), and Laplacian Super-Resolution Network (LapSRN) using a test-set comprising of 214 images. For the 2x model, the input image size is 480x360, while for the 4x model, it is 240x180. The ground truth image size and the expected output image sizes are 960x180.

1) EDSR: The EDSR network architecture consists of both a single-scale design, catering to a specific super-resolution scale, and a multi-scale architecture that can handle various levels of high-resolution. Within the EDSR residual block, batch norm layers are omitted to eliminate range flexibility and reduce GPU memory usage [24]. However, a drawback of this model is that it takes more time to make predictions, as compared to the FSRCNN.

2) FSRCNN: This method represents an accelerated version of the SRCNN model, employing a compact hourglass-shaped CNN structure to enhance both speed and model performance. It involves three key steps: adding a deconvolution layer at the network's end to map learning from the LR image, transitioning to smaller filter sizes with more mapping layers, and shrinking the input features [25]. It delivers faster processing speeds at the cost of a slight reduction in image quality.

3) LapSRN: The primary task of the LapSRN model is to reconstruct sub-band residuals for high-resolution (HR) images. At various pyramid levels, it leverages coarse-resolution data as input, employs transposed convolutions, and generates high-frequency residuals to enhance the image quality. Training the LapSRN model utilizes a Charbonnier loss function. In terms of prediction speed, this model outperforms EDSR but may introduce some noise [26].

TABLE III
AVERAGE ERROR METRICS WITH UPSCALING FACTORS 2X AND 4X.

$Upscale$	$Models$	$SSIM$	$PSNR$	$NRMSE$	$Pearson's-correlation$	R^2
2x	SISR model	0.84098	29.7897	0.0952	0.9590	0.9539
	EDSR	0.8335	30.0897	0.0916	0.9591	0.9593
	LapSRN	0.8287	29.9039	0.0934	0.9574	0.9583
	FSRCNN	0.8377	30.1546	0.0905	0.9596	0.9591
4x	SISR model	0.6845	27.0929	0.1348	0.9214	0.9414
	EDSR	0.6507	26.8456	0.1369	0.9131	0.9444
	LapSRN	0.6273	26.5504	0.1413	0.9072	0.9423
	FSRCNN	0.6242	26.4633	0.1428	0.9056	0.9421

4) 2x Upscaling: : Table III provides the quantitative results. These results are divided into two categories: 2x and 4x upscaling. The performance of the developed SISR model surpasses others, with an average SSIM of 0.84, indicating a high similarity between the high-resolution and predicted images. A Pearson's correlation coefficient of 0.95 signifies strong image correlation. From Fig. 7, (2x predictions from all four models), SISR and EDSR exhibit similar performance, with both models closely matching the ground truth, especially in the selected image area. Notably, SISR offers a speed advantage over EDSR and proves effective for remote sensing data, as compared to the FSRCNN and LapSRN.

5) 4x Upscaling: : In Table III, the developed SISR model outperforms the others by a significant margin. With an average SSIM of 0.684, it surpasses the faster FSRCNN at 0.624. It's worth noting that model performance tends to decline as the resolution doubles, a trend consistent across various super-resolution techniques. This reduction in the results is expected for a high super-resolution model. When examining Fig. 8, the SISR model demonstrates superior SSIM and PSNR, as compared to other competitors. By comparing two selected areas, indicated by red/blue boxes in the ground truth image, it's evident that the image quality decreases from left to right.

B. Limitations and Further Improvements

Our dataset, comprising of less than 2000 images, is considered minimal for effective learning by computer vision algorithms. Therefore, expanding the dataset and training on more data is essential for further generalization of the model. Notably, the agricultural scene's average metrics are notably lower than the overall average for both models, indicating an area for improvement. Addressing the imbalance in images across different weather and time conditions is crucial, and it can be resolved by collecting more images from such conditions. To enhance the 4x model's performance, additional residual and convolution layers may be considered. Alternatively, a more advanced option involves implementing a Generative Adversarial Network (GAN) architecture. GAN combines a generator and discriminator network to create high-quality images. The generator produces super-resolution images, while the discriminator distinguishes high-resolution images and backpropagates the GAN loss to train both [27].

C. Future Applications

The demand for Super-Resolution techniques is expanding across various scientific fields, extending beyond remote sensing images to encompass applications involving photos and videos. In the fields like surveillance and forensics, the ability to magnify objects or images from closed circuit television (CCTV) cameras can aid in identifying people and reading text. Moreover, in the realm of video super-resolution, there's a need to convert high-definition television (HDTV) data from standard definition television (SDTV) data with minimal artifacts [28]. The developed SISR model holds promise for application in all these domains. While direct predictions might not always be optimal, the model could be effectively utilized through transfer learning or retraining with the specific data of interest. Thus, the research discussed here has the potential to benefit a wide range of industrial applications.

VII. Conclusion

In this paper, we successfully trained and developed a Deep Convolutional Single Image Super-Resolution model, equipped with residual blocks, convolutional blocks, and batch normalization. This model effectively upscaled images to 2x and 4x of their original size, closely resembling the ground truth images. The model's potential applications extend to geospatial intelligence industries and other stakeholders, requiring remote sensing image super-resolution. We presented exploratory data analysis demonstrating the model's ability to perform image super-resolution effectively. We also examined the model's response to different scene types, weather conditions, and the time when the images were taken. Our analysis of training and testing error metrics revealed that the model is well-generalized, avoiding issues of overfitting or underfitting. Furthermore, in comparison with established models like EDSR, FSRCNN, and LapSRN, our model demonstrated superior performance on the testing set. In summary, we throughly experiment, analyse, discuss, validate the effectiveness of the proposed Single Image Super-Resolution CNN model.

References

[1] A. Azarang and N. Kehtarnavaz, *Image fusion in remote sensing: Conventional and deep learning approaches*. Springer Nature, 2022.

[2] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[3] J. A. Benediktsson, J. Chanussot, and W. M. Moon, "Very high-resolution remote sensing: Challenges and opportunities [point of view]," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1907–1910, 2012.

[4] D. Yang, Z. Li, Y. Xia, and Z. Chen, "Remote sensing image super-resolution: Challenges and approaches," in *2015 IEEE International Conference on Digital Signal Processing*. IEEE, 2015, pp. 196–200.

[5] "CVonline: Image Databases, https://homepages.inf.ed.ac.uk/rbf/cvonline/imagedbase.htm#remote."

[6] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, 2014.

[7] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*. Springer, 2012, pp. 711–730.

[8] M. Carranza-García, J. García-Gutiérrez, and J. C. Riquelme, "A framework for evaluating land use and land cover classification using convolutional neural networks," *Remote Sensing*, vol. 11, no. 3, p. 274, 2019.

[9] C. Chen and F. Qi, "Single image super-resolution using deep cnn with dense skip connections and inception-resnet," in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, 2018, pp. 999–1003.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, pp. 770–778.

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 2261–2269.

[12] Y. Wang, L. Gu, X. Li, F. Gao, T. Jiang, and R. Ren, "An improved spatiotemporal fusion algorithm for monitoring daily snow cover changes with high spatial resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[13] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Science Reviews*, p. 104110, 2022.

[14] J. M. Haut, R. Fernandez-Beltran, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "A new deep generative network for unsupervised remote sensing single-image super-resolution," *IEEE Transactions on Geoscience and Remote sensing*, vol. 56, no. 11, pp. 6792–6810, 2018.

[15] Y. Da Wang, R. T. Armstrong, and P. Mostaghimi, "Enhancing resolution of digital rock images with super resolution convolutional neural networks," *Journal of Petroleum Science and Engineering*, vol. 182, p. 106261, 2019.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*. Ieee, 2017, pp. 1–6.

[17] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*, 2015, pp. 448–456.

[18] W. QingJie and W. WenBin, "Research on image retrieval using deep convolutional neural network combining l1 regularization and prelu activation function," in *IOP Conference Series: Earth and Environmental Science*, vol. 69, no. 1. IOP Publishing, 2017, p. 012156.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[20] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[21] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[23] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*. IEEE, 2015, pp. 1–6.

[24] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1132–1140.

[25] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 391–407.

[26] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 5835–5843.

[27] X. Liu and C.-J. Hsieh, "Rob-gan: Generator, discriminator, and adversarial attacker," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 11 226–11 235.

[28] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.