Single Hyperspectral Image Super-Resolution utilizing Implicit Neural Representations

Bohdan Perederei[1] [a] and Faisal Z. Qureshi[2] [b]

[1]*Department of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Prospect Beresteiskyi 37, Kyiv, Ukraine*
[2]*Faculty of Science, Ontario Tech University, 2000 Simcoe St North, Oshawa, Canada*
woncaru@gmail.com, Faisal.Qureshi@ontariotechu.ca

Abstract: Hyperspectral image super-resolution is a crucial task in computer vision, aiming to enhance the spatial resolution of hyperspectral data while maintaining spectral fidelity. In this paper, we introduce highlights and outcomes of our research, in which we developed, explored, and evaluated different techniques and methods based on Implicit Neural Representations (INRs) for conducting Single Hyperspectral Image Super-Resolution. Despite the potential of INRs, their application to hyperspectral image super-resolution still needs to be explored, with significant room for further investigation. Our primary goal was to adapt strategies and techniques from models originally developed for multispectral image super-resolution, especially SIREN-based INRs and the Dual Interactive Implicit Neural Network architectures. We also explored feature extraction from hyperspectral images using a convolutional neural network autoencoder that allowed us to capture spatial-spectral patterns for further enhancement. Furthermore, as a part of the research, we validated and compared different functions, such as MSE, RMSE, MAE, PSNR, SAD, SAM, and SSIM, to evaluate their effectiveness as loss functions for training INRs.

1 INTRODUCTION

Being a challenging and ill-posed problem, single image super-resolution (SISR) is one of the fundamental computer vision problems aimed at generating a high-resolution image from a low-resolution input. The main reason for conducting SISR is to improve image representation for better human and machine interpretation. Considering multispectral imagery super-resolution (e.g., RGB), various deep-learning techniques could provide high-quality or state-of-the-art results for super-resolution tasks [Wang et al., 2019].

As for the super-resolution of hyperspectral imagery (HSI), it is an especially significant area of focus for the researchers since hyperspectral cameras aim at capturing the scene in various spectral bands, but due to limited hardware capabilities and financial resources, they usually have reduced spatial resolution compared to multispectral images. As a result, low spatial resolution and high camera prices lead to a scarcity of HSI data, which can significantly limit some deep learning approaches due to insufficient training data. Another reason why this topic is valuable for exploration is that HSI data presents unique challenges compared to multispectral imagery. HSI can have from several dozen to several hundred spectral bands, leading to very high-dimensional data that results in issues such as high computational load, greater sensitivity to noise, and the so-called curse of dimensionality, which makes a considerable number of deep learning techniques [Chen et al., 2021a, Nguyen and Beksi, 2023] unusable or as such that require significant optimizations.

Furthermore, in addition to SISR, there is another widely used approach for enhancing the spatial resolution of HSI - fusion-based hyperspectral image super-resolution. The core concept of this method is to improve the quality of upscaled hyperspectral imagery by combining it with additional data, such as RGB or multispectral images. In experimental settings, such methods often deliver better results than techniques relying solely on single image super-resolution. However, a common assumption among these methods is that the low-resolution hyper-

spectral and high-resolution auxiliary images are precisely aligned. In practical scenarios, capturing a low-resolution hyperspectral image and a high-resolution multispectral image often involves different cameras. This results in minor variations in the imaging conditions, which complicates the achievement of accurate image registration. Thus, our research primarily concentrated on the single image super-resolution task.

Our primary contribution is the analysis, adaptation, and evaluation of various INR architectures for hyperspectral SISR, including SIREN, SIREN with bicubic interpolation, and the Implicit Decoder from the Dual-Interactive Implicit Neural Network paired with a custom CNN autoencoder. Additionally, we investigated loss functions to determine their effectiveness in training INRs, identifying PSNR as the most effective for optimal convergence and metrics in our experiments.

The paper is structured as follows: The Related Work section reviews hyperspectral SISR methods, including traditional, deep learning, and INR-based approaches for multispectral and hyperspectral imagery. The Methodology section outlines the rationale for using INRs and details the metrics, loss functions, and model architectures. The Experiments, Results, and Discussion section describes the experimental setup, presents the results, and analyzes the findings.

2 RELATED WORK

This section provides a concise overview of the methods defined in the literature. Because of the reasons mentioned in the Introduction section, fusion-based hyperspectral image super-resolution is out of the scope of this work. In addition, it is essential to note that this work prioritizes learning deterministic mappings over stochastic approaches, such as those employed in the generative models.

2.1 Conventional Approaches

The first methods that come to mind regarding image upscaling are classic and well-known image interpolation techniques, such as nearest-neighbor, bilinear, and bicubic interpolation. They are straightforward to implement, provide efficient image upscaling, and demonstrate stability in quality across data dimensionalities. However, super-resolution machine learning-based methods, such as deep learning models based on convolutional neural networks, are usually higher on metrics as they can generate much higher-quality results, preserving details and textures

of low-scale images. Hence, researchers use these interpolation methods as baseline approaches.

As for other conventional SISR techniques, [Wang et al., 2017] proposed a method for HSI super-resolution based on nonlocal low-rank tensor approximation and total variation regularization. This approach effectively preserves structural details and reduces noise by leveraging low-rank priors, but it requires solving computationally expensive optimization problems, making it less efficient for large datasets. Similarly, [Li et al., 2016] introduced a technique that combines spectral mixture analysis with spatial-spectral group sparsity to capture the underlying spatial and spectral correlations. While this method improves accuracy by promoting sparsity, it also demands significant computational resources. It relies on carefully designed, hand-crafted priors, which may not be generalized well in real-world scenarios.

2.2 Deep Learning Approaches

In contrast to conventional techniques, deep learning-based methods, despite being data-demanding, offer a more flexible and scalable approach by automatically learning features, often outperforming traditional methods in speed and accuracy. The most widely used model architectures for hyperspectral SISR and multispectral SISR are mainly based on CNN layers, and all the papers on this topic demonstrated that network architecture design is a crucial factor in image reconstruction quality [Mei et al., 2017, Arun et al., 2020, Jiang et al., 2020]. However, because hyperspectral data contains hundreds of channels along the spectral dimension, its complex 3D nature makes hyperspectral imaging unsuitable for SISR techniques applied to natural images, requiring researchers to adapt or create novel methods.

At this point, researchers have introduced a vast range of model architectures. Three main categories of models can be generally classified based on their upsampling techniques and the placement of upsampling layers within the model architecture: back-end upsampling (after CNN layers), front-end upsampling (before CNN layers), and progressive upsampling (between CNN layers). Typically, the most challenging upsampling step is performed utilizing traditional techniques like bicubic interpolation, with deep neural networks responsible only for refining these interpolated images to restore fine details and achieve higher quality. Furthermore, because conventional interpolation-based upsampling methods cannot incorporate external prior information and are

unsuitable for use as an upsampling layer in the back-end upsampling structure, researchers have introduced learning-based upsampling methods, such as transposed convolution [Dong et al., 2016] and pixel shuffle [Shi et al., 2016] as alternative techniques.

As for the structures of networks, the most common designs usually include recursive learning, residual learning, multi-path learning, dense connections, and attention mechanism. As research into neural networks advances, more and more network architectures are being designed and utilized for SISR task.

2.3 Implicit Neural Representations for Multispectral SISR

Implicit neural representation commonly encodes an object using a multi-layer perceptron (MLP) that associates spatial coordinates with a signal. INR was found to be especially useful in 3D object representations. As a result, through extensive research, traditional discrete models of 3D object shapes, surfaces, and scene structures have been replaced by continuous functions defined by MLPs. One such model we heavily researched for the SISR problem is SIREN (dubbed sinusoidal representation network) [Sitzmann et al., 2020], which usually showed better 2D images and 3D object reconstruction results than other INR models with different architectures and activation functions. However, researchers focused on INR exploration mostly applied them to 3D computer vision and object representation [Sitzmann et al., 2020, Mildenhall et al., 2021, Wang et al., 2021] often ignoring 2D imaging, which resulted in under-exploration of this domain.

Nonetheless, research into INR techniques for 2D imagery has progressed, leading to investigations on how INR can be effectively applied to various computer vision tasks, including SISR. Implicit Neural Representations of 2D images can be directly applied to SISR because they enable the sampling of pixel values at any spatial location. For instance, [Chen et al., 2021b] in their paper proposed the method called Local Implicit Image Function (LIIF), which determines a pixel's value by referencing the closest latent code, consisting of a localized collection of adjacent feature vectors. This pixel-based approach facilitates a seamless transition across different areas in the reconstructed image. Drawing inspiration from LIIF, [Tang et al., 2021] introduced a new INR-driven representation called the Joint Implicit Image Function (JIIF) for guided depth super-resolution, which aims to learn the interpolation weights and their corresponding values simultaneously.

Based on insights provided by JIIF and LIIF papers and assuming that simple concatenation on spatial encodings and coordinates cannot fully improve the quality of the output images, the Meta-SR paper [Hu et al., 2019] introduced a magnification-arbitrary network that leverages INR techniques to perform super-resolution across a range of scaling factors. Also, unlike the approach in LIIF, [Nguyen and Beksi, 2023] proposed a novel Dual Interactive Implicit Neural Network (DIINN). DIINN consists of a well-known Residual Dense Network encoder [Zhang et al., 2018] and a unique Implicit Decoder that itself includes Modulation and Synthesis networks to enhance the implicit decoding function by separating the content and positional features at the pixel level, as suggested by [Mehta et al., 2021].

2.4 Implicit Neural Representations for Hyperspectral SISR

Different INR techniques and architectures in Subsection 2.3 demonstrate that they are frequently and effectively applied to address SISR problems in the multispectral imagery domain. While research on INR for multispectral SISR is advancing and yielding promising outcomes, the application of INRs for hyperspectral SISR remains underexplored. The reasons for that are general trends in HSI super-resolution research, which primarily utilizes methods and techniques mentioned in Subsection 2.2, and additional challenges that are presented by HSI data, such as the curse of dimensionality and high computational load.

Nonetheless, some progress has already been made in the research of INR for HSI super-resolution, and optimistic results have been demonstrated. For instance, [Zhang et al., 2022] addressed the challenges of high-dimensional spectral patterns in HSI super-resolution without relying on auxiliary images, introducing a novel model that utilizes INR to map spatial coordinates to their corresponding spectral values through continuous functions, enhanced by a hypernetwork for INR parameter prediction. Evaluations on multiple datasets demonstrated that this approach yields competitive reconstruction performance, highlighting the model's capability to recover high-frequency details effectively.

Furthermore, [Chen et al., 2023] introduced a novel approach called Spectral-wise Implicit Neural Representation (SINR) that addressed the limitations of traditional methods in HSI reconstruction, which often results in a worse representation of spectral information continuity. SINR employs a continuous spectral amplification process and incorporates a spectral-wise attention mechanism, treating individual channels as distinct tokens to capture global

spectral dependencies effectively. Even though SINR primarily targets HSI reconstruction, it can be easily adapted for SISR tasks. Extensive experiments demonstrated that this framework outperforms baseline methods, significantly enhancing flexibility and performance by accommodating unlimited spectral bands in the output.

3   METHODOLOGY

This research aims to propose and evaluate novel approaches based on INRs for HSI super-resolution. The rationale for employing INR-based methods aligns closely with their established benefits in 3D object reconstruction and 2D multispectral image super-resolution.

Hyperspectral imagery, like 3D discrete models, is inherently data-intensive, which highlights the potential of INRs for HSI reconstruction and super-resolution, as they can effectively manage the high dimensionality and complexity associated with such data forms. In addition to this capability, INRs offer significant flexibility in resolution due to their continuous nature. For instance, when utilizing a SIREN architecture for image fitting, it becomes straightforward to upscale an image to any desired resolution. A high-resolution image can be obtained without requiring extensive re-training simply by creating a larger pixel grid and inputting it into the trained SIREN model. Furthermore, when applied to hyperspectral data, INRs can enhance image representation efficiency, as the trained weights of an INR model are often significantly smaller than the original full HSI image, enabling effective storage and transmission. Despite these advantages, INR-based methods remain relatively underexplored in HSI super-resolution, presenting a valuable opportunity for further research to fully leverage their capabilities in HSI processing tasks.

During our research, we systematically explored various increasingly complex approaches to achieve super-resolution metrics that surpass those of interpolation methods, which served as our baseline. The details and outcomes of these experiments will be presented in Section 4. To prevent redundant efforts, we concentrated on leveraging techniques previously demonstrated to be effective for multispectral image reconstruction and SISR that utilize INRs. Notably, we employed ideas and techniques from such networks as SIREN [Sitzmann et al., 2020], LIIF [Tang et al., 2021], and DIINN [Nguyen and Beksi, 2023], all of which have shown promising results with multispectral data.

3.1   Evaluation Metrics and Loss Functions

To assess the effectiveness of HSI super-resolution techniques, we rely on such commonly used evaluation metrics as Mean Square Error (MSE), Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), Sum Of Absolute Differences (SAD), Spectral Angle Mapper (SAM), and Structural Similarity (SSIM). MSE quantifies the average squared differences between estimated and actual pixel values:

$$\text{MSE}(\mathbf{X},\hat{\mathbf{X}}) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2. \tag{1}$$

As for PSNR, it is a widely used metric for evaluating image quality. It is determined by the maximum pixel value (L, in our case equals 1) in the image and the MSE calculated between the original high-resolution HSI and its reconstructed counterpart:

$$\text{PSNR}(\mathbf{X},\hat{\mathbf{X}}) = 10\log_{10}\left(\frac{L^2}{\text{MSE}}\right) \tag{2}$$

$$= 10\log_{10}\left(\frac{L^2}{\frac{1}{W\times H}\sum_{i=1}^{W\times H}(X_i - \hat{X}_i)^2}\right). \tag{3}$$

The SAM function, introduced by [Kruse et al., 1993], assesses the spectral similarity between pixels in hyperspectral images by calculating the angle between their spectral vectors, where a smaller angle indicates a higher likelihood of the pixels belonging to the same class:

$$\text{SAM}(\mathbf{X},\hat{\mathbf{X}}) = \arccos\left(\frac{\langle \mathbf{X}^T\hat{\mathbf{X}}\rangle}{\|\mathbf{X}\|_2 \cdot \|\hat{\mathbf{X}}\|_2}\right). \tag{4}$$

Regarding SSIM index [Wang et al., 2004], it measures the structural similarity between an original image and a reconstructed image, considering image degradation as a perceived change in structural information:

$$SSIM(\mathbf{X},\hat{\mathbf{X}}) = \frac{(2\mu_{\mathbf{X}}\mu_{\hat{\mathbf{X}}} + c_1)(2\sigma_{\mathbf{X}\hat{\mathbf{X}}} + c_2)}{(\mu_{\mathbf{X}}^2 + \mu_{\hat{\mathbf{X}}}^2 + c_1)(\sigma_{\mathbf{X}}^2 + \sigma_{\hat{\mathbf{X}}}^2 + c_2)}. \tag{5}$$

Here, μ represents the mean value, while σ indicates the variance or covariance. The variables c_1 and c_2 are introduced to stabilize the division operation when the denominator is small.

Moreover, functions of presented evaluation metrics and various weighted combinations of them were assessed for their suitability as loss functions in this study. Mean absolute error (MAE) was also tested as a potential loss function. The results of these evaluations can be found in Section 4.2. PSNR achieved the best convergence time and metric values in our tests, surpassing other functions and becoming the primary loss function for model training.

3.2 Explored Methods and Architectures

The SIREN architecture (three hidden layers with 256 neurons each) was selected as an initial approach, using pixel coordinates as inputs and spectral channels as outputs. The primary objective was to train the network to reconstruct HSI and perform super-resolution by feeding the model an enlarged pixel grid. However, the baseline SIREN architecture did not achieve performance metrics superior to bicubic interpolation.

To address this, we experimented with an alternative method where the image was first upscaled using bicubic interpolation, followed by enhancement through the trained SIREN model. Specifically, we downscaled the original image, upscaled it back using bicubic interpolation, and trained the SIREN to minimize the loss between the original and the synthetically downscaled and upscaled image. Once trained, the model was applied to the bicubically upscaled original image to enhance its quality further. Despite these efforts, this technique failed to produce results that outperformed the standard bicubic interpolation.

Following this, we shifted our focus to more complex architectures, carefully considering using autoencoders to represent HSI in latent space, which could then be fed into the network instead of the raw image. The starting point for exploring more advanced architectures was DIINN [Nguyen and Beksi, 2023], which consists of a Residual Dense Network (RDN) encoder [Zhang et al., 2018] and a novel Implicit Decoder composed of modulation and synthesis networks. Unsurprisingly, DIINN, initially designed for upscaling RGB images, failed to work without modifications, primarily due to the limitations of the RDN encoder. The latent space produced by the RDN from the DIINN network is smaller than the number of spectral bands in HSI, and adjusting the RDN to handle the increased spectral channels proved challenging due to the curse of dimensionality. This resulted in excessive memory usage and computational demands, making it challenging to employ the RDN as an encoder in any network architecture without a significant redesign.

Due to the necessity of utilizing latent space over the entire image and the limitations of employing RDN, we opted for a custom autoencoder to encode the image into latent space, enhancing the spectral resolution. We chose two CNN layers with a kernel size of 3×3 for the autoencoder architecture. These layers process 4x4 image patches with three overlapping pixels, expanding their spectral representation to 256 and then to 512 channels. Consequently, each 4x4 image patch is transformed into a 1×1×512 vector

in the latent space. This latent representation is then fed into the Implicit Decoder, replacing the RDN-based approach. Unfortunately, this approach also failed to achieve metrics surpassing those of interpolation methods.

4 EXPERIMENTS, RESULTS AND DISCUSSION

4.1 Datasets and Experimental Settings

For training and initial evaluations, we utilized the Cuprite dataset, consisting of a hyperspectral image with dimensions of 512×614 pixels and 188 channels [NASA Jet Propulsion Laboratory, 1997]. To further enhance our evaluations and facilitate comparisons with other studies, we also incorporated the Chikusei dataset, captured using the Headwall Hyperspec-VNIR-C imaging sensor over agricultural and urban regions in Chikusei, Ibaraki, Japan [Yokoya and Iwasaki, 2016]. The HSI in the Chikusei dataset consists of 2517×2335 pixels with 128 bands. For the Chikusei dataset, we adopted the approach used in [Jiang et al., 2020] and [Zhang et al., 2023] by extracting non-overlapping patches of 512×512 pixels for evaluation and benchmarking purposes. It is essential to highlight that all spectral values in the Chikusei dataset were normalized to a range between 0 and 1. For model training and evaluation, the hyperspectral images serve as ground truth, while the input data is generated by downscaling these images using bicubic interpolation at the desired scaling factor.

As for experimental settings, we designed our INR-based models by incorporating and adapting elements from the SIREN and DIINN architectures. The models were trained from the ground up in PyTorch on Nvidia V100-SXM2 GPUs. Adam with 1e-4 rate was applied as an optimizer [Kingma, 2014], PSNR was used as a loss function, and models were trained for 1000 epochs. The training configurations were consistent across all datasets and models, with minimal specific adjustments made.

4.2 Loss Functions Evaluation

Our research analyzed several loss functions (MSE, RMSE, MSE, PSNR, SAD, SAM, and SSIM) to identify which makes the training process converge more rapidly and produce superior metrics overall. For this experiment, we trained the SIREN model for the HSI reconstruction task using the Cuprite dataset and four 512×512 patches from the Chikusei dataset. The

results shown in Tables 1 and 2 indicate that using PSNR as a loss function yields better metric values after 1000 epochs of training compared to others. Additionally, we experimented with various weighted sum combinations of loss functions (as shown in Table 3) to determine whether this approach could improve convergence and metric values. However, none of the tested combinations surpassed the performance of using the PSNR as a loss function alone. Consequently, we selected PSNR as the loss function to train all subsequent models in this study.

4.3 Models Evaluation

To enable a more comprehensive comparison, we included results for nearest-neighbor and bilinear interpolation, alongside bicubic interpolation, in each table. The corresponding metric values are presented in Tables 4 and 5.

We initially tested two SIREN architectures: one with three hidden layers of 256 neurons each and another with three layers of 512 neurons each. While increasing the number of neurons led to some improvement in the metric values, both models performed worse than bilinear and bicubic interpolation. However, they did surpass the nearest-neighbor method on the Cuprite dataset. This gave us the idea to utilize image latent space representation or any other additional data that could be passed as input to the network.

Thus, we experimented with a bicubic SIREN, where the input was not just the coordinate grid, but for each coordinate, we passed channel values obtained from bicubic interpolation. In the first experiment, we downscaled the image we aimed to upscale by a factor of 2 and used this downscaled version as input data with the original image serving as the ground truth for training the model. After that, we fed to the trained model the original image we wanted to upscale as input. This approach yielded worse results than the simple SIREN on the Cuprite dataset but better results on the Chikusei dataset. In the second experiment, we trained the bicubic SIREN by feeding it the original image without downscale during training. For super-resolution, we passed the image that had been upscaled using bicubic interpolation. Overall, the results across datasets are inconsistent with bicubic SIREN producing similar outcomes with some minor fluctuations in metric values compared to standard SIREN, which can be attributed to the differences in the structure of the HSI test data. Additionally, it is essential to note that this method can be viewed as bicubic interpolation degraded by SIREN reconstruction, with SIREN learning primarily to add as little noise as possible rather than en-

hancing the bicubic-interpolated image.

For the model adapted from DIINN, we used a CNN autoencoder trained for ten epochs with MSE loss to extract features from the images in an unsupervised manner, which were then passed to the Implicit Decoder as input. Each image was trained with a separate CNN autoencoder by feeding it 4×4 overlapping patches with a stride equal to 3. Despite our expectations that this architecture would yield strong results, it failed to outperform interpolation methods discussed in this study (Tables 4 and 5).

Even though none of the proposed methods managed to outperform interpolation techniques, there is still significant room for further research. For instance, evaluating a CNN autoencoder with SIREN or adapting the LIIF model as a decoder could be promising directions for future studies.

5 CONCLUSIONS

Our research provides a broad overview of the current landscape in single hyperspectral image super-resolution, highlighting advancements in deep learning and implicit neural representations. We proposed and evaluated several INR-based architectures adapted from those created for the problem of multispectral image super-resolution and also evaluated various loss functions that can be used for INR-based model training. Although our approaches did not outperform in terms of metric values interpolation methods, which were chosen as a baseline, the area of utilizing INR for hyperspectral image super-resolution remains underexplored, requiring further investigation. Among the loss functions evaluated, PSNR demonstrated the best results.

We hope that the findings from our research will provide valuable insights to other researchers, enabling them to develop more effective methods. By sharing the results of our study, we aim to help others avoid spending time and resources on approaches we have already tested and found ineffective, allowing them to focus on developing more effective solutions.

ACKNOWLEDGMENTS

Table 1: Loss functions comparison for SIREN fitting (3 hidden layers with 256 neurons each) on the Cuprite dataset (1000 epochs). Rows represent loss functions.

	MSE	RMSE	PSNR	SAD	SAM	SSIM
MSE	0.000673	0.025933	31.722861	1141381.5	3.212296	0.780906
RMSE	0.000306	0.017482	35.148163	774198.38	2.144571	0.905351
MAE	0.000400	0.019991	33.983388	840628.69	2.478677	0.872909
PSNR	0.000201	0.014171	36.971917	630810.25	1.728390	0.942259
SAD	0.000411	0.020261	33.866750	854810.25	2.513883	0.869774
SAM	0.034780	0.186494	14.586708	10511805.0	2.804074	0.692118
SSIM	0.000345	0.018585	34.616675	819630.00	2.304268	0.893533

Table 2: Loss functions comparison for SIREN fitting (3 hidden layers with 256 neurons each) on the Chikusei dataset (1000 epochs). The table presents mean values for four 512×512 patches. Rows represent loss functions.

	MSE	RMSE	PSNR	SAD	SAM	SSIM
MSE	0.000587	0.024223	32.315339	521001.98	52.010291	0.839271
RMSE	0.000589	0.024268	32.299372	520834.09	52.115509	0.837885
MAE	0.000606	0.024608	32.178375	462328.36	52.065796	0.800279
PSNR	0.000475	0.021791	33.234386	446634.08	45.458306	0.974262
SAD	0.000606	0.024620	32.174232	462560.37	52.124053	0.800041
SAM	0.003146	0.056088	25.023480	1570795.0	53.684620	0.462910
SSIM	0.000592	0.024328	32.278115	490803.63	52.715745	0.834866

Table 3: Loss functions weighted sum combinations comparison for SIREN fitting (3 hidden layers with 256 neurons each) on the Cuprite dataset (1000 epochs).

	MSE	RMSE	PSNR	SAD	SAM	SSIM
0.3 MSE + 0.7 PSNR	0.000216	0.014694	36.657299	651320.38	1.808510	0.935802
0.9 MSE + 0.1 SSIM	0.000416	0.020395	33.809437	900755.13	2.528401	0.869610
0.01 PSNR + 0.05 SSIM	0.000205	0.014330	36.875143	637339.88	1.757805	0.940792
2 MSE + 0.01 SAM	0.000627	0.025048	32.024600	1103658.6	3.117343	0.795719
0.01 PSNR + 0.5 SSIM	0.000222	0.014909	36.530963	665096.44	1.804285	0.936313

Table 4: Quantitative evaluations of different approaches for upscaling the Cuprite dataset from 256×308×188 to 512×614×188.

	MSE	RMSE	PSNR	SAD	SAM	SSIM
Nearest-neighbor	0.000399	0.019964	33.995021	782167.69	2.479466	0.884059
Bilinear	0.000184	0.013563	37.353046	569532.38	1.672223	0.936155
Bicubic	0.000160	0.012643	37.963288	530463.00	1.555574	0.949012
SIREN (3 layers 256 neurons)	0.000299	0.017294	35.242156	754232.00	2.131940	0.910444
SIREN (3 layers 512 neurons)	0.000253	0.015913	35.964863	686883.13	1.948931	0.920803
Bicubic SIREN downscale	0.000310	0.017593	35.093361	753831.63	2.159939	0.913557
Bicubic SIREN no downscale	0.000187	0.013693	37.270051	590392.94	1.664128	0.948168
CNN AE + Implicit Decoder	0.003791	0.061569	24.212803	2822800.5	7.854479	0.588787

Table 5: Quantitative evaluations of different approaches for upscaling the Chikusei dataset from 256×256×128 to 512×512×128. The table presents mean values for four 512×512 patches.

	MSE	RMSE	PSNR	SAD	SAM	SSIM
Nearest-neighbor	0.000387	0.019670	34.123814	414811.76	39.404533	0.924318
Bilinear	0.000398	0.019939	34.005960	412121.91	40.700801	0.919221
Bicubic	0.000420	0.020488	33.770147	434430.89	41.718372	0.917627
SIREN (3 layers 256 neurons)	0.000602	0.024530	32.205966	527469.03	52.713975	0.837608
SIREN (3 layers 512 neurons)	0.000589	0.024273	32.297660	519392.61	52.175372	0.838893
Bicubic SIREN downscale	0.000462	0.021488	33.356040	442064.97	44.539143	0.903490
Bicubic SIREN no downscale	0.000589	0.024277	32.296041	520064.68	52.162835	0.838707
CNN AE + Implicit Decoder	0.000541	0.023254	32.669917	501532.94	49.409021	0.842707

REFERENCES

Arun, P. V., Buddhiraju, K. M., Porwal, A., and Chanussot, J. (2020). Cnn-based super-resolution of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6106–6121.

Chen, H., Zhao, W., Xu, T., Shi, G., Zhou, S., Liu, P., and Li, J. (2023). Spectral-wise implicit neural representation for hyperspectral image reconstruction. *IEEE Transactions on Circuits and Systems for Video Technology*.

Chen, Y., Liu, S., and Wang, X. (2021a). Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8628–8638.

Chen, Y., Liu, S., and Wang, X. (2021b). Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638.

Dong, C., Loy, C. C., and Tang, X. (2016). 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*, pages 391–407. Springer.

Hu, X., Mu, H., Zhang, X., Wang, Z., Tan, T., and Sun, J. (2019). Meta-sr: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1575–1584.

Jiang, J., Sun, H., Liu, X., and Ma, J. (2020). Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 6:1082–1096.

Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kruse, F. A., Lefkoff, A., Boardman, y. J., Heidebrecht, K., Shapiro, A., Barloon, P., and Goetz, A. (1993). The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163.

Li, J., Yuan, Q., Shen, H., Meng, X., and Zhang, L. (2016). Hyperspectral image super-resolution by spectral mixture analysis and spatial–spectral group sparsity. *IEEE Geoscience and Remote Sensing Letters*, 13(9):1250–1254.

Mehta, I., Gharbi, M., Barnes, C., Shechtman, E., Ramamoorthi, R., and Chandraker, M. (2021). Modulated periodic activations for generalizable local functional representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14214–14223.

Mei, S., Yuan, X., Ji, J., Zhang, Y., Wan, S., and Du, Q. (2017). Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 9(11):1139.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106.

NASA Jet Propulsion Laboratory (1997). Cuprite hyperspectral dataset. https://aviris.jpl.nasa.gov/data/free_data.html.

Nguyen, Q. H. and Beksi, W. J. (2023). Single image super-resolution via a dual interactive implicit neural network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4936–4945.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883.

Sitzmann, V., Martel, J., Bergman, A., Lindell, D., and Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473.

Tang, J., Chen, X., and Zeng, G. (2021). Joint implicit image function for guided depth super-resolution. In *Proceedings of the 29th acm international conference on multimedia*, pages 4390–4399.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., and Wang, W. (2021). Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*.

Wang, Y., Chen, X., Han, Z., and He, S. (2017). Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization. *Remote Sensing*, 9(12):1286.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Wang, Z., Chen, J., and Hoi, S. C. H. (2019). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43:3365–3387.

Yokoya, N. and Iwasaki, A. (2016). Airborne hyperspectral data over chikusei. *Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27*, 5(5):5.

Zhang, K., Zhu, D., Min, X., and Zhai, G. (2022). Implicit neural representation learning for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12.

Zhang, M., Zhang, C., Zhang, Q., Guo, J., Gao, X., and Zhang, J. (2023). Essaformer: Efficient transformer for hyperspectral image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23073–23084.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481.