A True Hyperspectral Image Super-Resolution Dataset

Alexander Ulrichsen
University of Strathclyde, UK
Glasgow G1 1XQ
alexander.ulrichsen.2015@uni.strath.ac.uk

Thomas De Kerf
University of Antwerp, Belgium
2000 Antwerp
thomas.dekerf@uantwerpen.be

R. David Dunphy
University of Strathclyde, UK
Glasgow G1 1XQ
david.dunphy@strath.ac.uk

Paul Murray
University of Strathclyde, UK
Glasgow G1 1XQ
paul.murray@strath.ac.uk

Steve Vanlanduit
University of Antwerp, Belgium
2000 Antwerp
steve.vanlanduit@uantwerpen.be

Stephen Marshall
University of Strathclyde, UK
Glasgow G1 1XQ
stephen.marshall@strath.ac.uk

Abstract

Hyperspectral imaging, crucial in remote sensing, provides extensive spectral information at the cost of lower spatial resolution compared to standard color images. Single-image super-resolution, reconstructing high-resolution images from low-resolution inputs, is particularly useful for enhancing hyperspectral images. Due to the unavailability of real low- and high-resolution image pairs, many hyperspectral image super-resolution methods resort to downsampling for training. This leads to suboptimal performance on real-world data due to inherent assumptions in the downsampling process. This paper introduces a novel dataset featuring actual low- and high-resolution hyperspectral image pairs, captured using different lenses and sensors. We train various super-resolution models on this dataset and compare their performance against models trained on artificially downsampled high-resolution images. Our findings reveal that models trained with real image pairs substantially outperform basic bicubic interpolation, whereas those trained with synthetically generated low-resolution images do not, highlighting the importance of using authentic high- and low-resolution images for training.

1. Introduction

Hyperspectral imaging provides a wealth of spectral information per pixel that standard RGB color images cannot. This is useful for applications such as remote sensing where the spectral information can be utilized to gain insights into the materials present within the image [24]. However, due to sensor limitations, the increase in spectral information generally comes at the cost of reduced spatial information [4]. Hence, enhancing this limited spatial resolution is desirable. Single-image super-resolution (SR) emerges as a pertinent solution, aiming to refine the resolution of an image, and has seen widespread adoption in recent years thanks to rapid advances in deep learning [6, 7, 11, 14, 18, 19, 28, 29]. This technique has also found significant application in enhancing hyperspectral images (HSIs) [3, 15, 17, 30]. Ideally, models are trained on low- and high-resolution image pairs, but acquiring such paired data is challenging, so most researchers downsample HSIs to artificially create a low-resolution pair [15, 20, 23, 26]. However, this imposes several assumptions on the downsampling process, which is detrimental to performance on real-world data [5, 31]. As a result, there is a substantial need for high-quality datasets to train and evaluate SR algorithms in the hyperspectral community, with many research works citing data scarcity as a major limitation [1, 16]. To address this issue, we present a novel and open source dataset featuring measured low- and high-resolution hyperspectral image pairs, which we hope will be used as a standard dataset within the community. We offer two types of data. The first data type we name *lens data* as these data are captured by the same sensor but with a high- and low-magnification lens to create high- and low-resolution data. A similar approach has previously been used to improve

super-resolution of natural images [32]. We refer to the second dataset as *sensor data*, as these data are captured with distinct sensors to create high- and low-resolution image pairs. Specifically, we use a Headwall Micro-Hyperspec VNIR E-Series, which has a spatial resolution of 1600 pixels, and a SPECIM–Hamamatsu imager with an effective spatial resolution (after binning) of 336 pixels. In the context of this paper, and based purely on the spatial resolutions specified, we will refer to these sensors going forward respectively as, relatively, higher resolution and lower resolution sensors. We run extensive super-resolution experiments on both types of data by training a diverse group of HSI-SR models on both the real data pairs in addition to popular synthetic downsampling techniques. Our results show that training on the real data pairs provides by far the best model performance and many of the synthetic downsampling approaches fail to even outperform simple bicubic interpolation.

The novel contributions of this work can be summarised as follows:

- We publish a novel open-source hyperspectral image super-resolution dataset with real high- and low-resolution image pairings captured using different lenses on the same sensor as well as with high- and low-quality sensors, suitable for evaluating super-resolution models and artificial degradation techniques [2].
- We train various super-resolution models on this dataset and show that using the real image pairings produces significant performance improvement over bicubic interpolation, whereas models trained on artificially downsampled low-resolution images fail to outperform bicubic interpolation on the real data.

2. Related Work

The vast majority of modern super-resolution (SR) methods utilize deep learning in some capacity which began with SRCNN [6]. Since then, there have been a wealth of improved deep learning architectures proposed for SR [7, 11, 18, 19]. Due to the spectral and spatial resolution trade-off inherent to hyperspectral imaging, super-resolution has also seen great development in this field to enhance the limited spatial resolution of hyperspectral images without having to sacrifice spectral resolution. Convolutional Neural Networks (CNNs) were first introduced to HSI-SR in [30] where the authors applied transfer learning from a 2D RGB SR CNN model. However, 2D convolution does not fully exploit the spectral correlation between bands. The authors of [20] therefore proposed a 3D CNN model to extract spatial and spectral features together for improved performance and spectral fidelity. Since then, many more 3D models have been proposed [3, 15, 17]. Given the large memory requirements of 3D CNNs, feature extraction is generally performed in the low resolution space before being interpolated to the target size by deconvolution [21] or pixel shuffling [22], reducing memory usage. However, applying deconvolution to extracted features is more challenging than simply performing SR on an input image already interpolated to the target resolution. The authors of [23] propose SRONN, which is a shallow 2D model based on SRCNN utilizing nonlinear Self-Operational [12] layers to perform SR on an interpolated input for enhanced performance on very small HSI datasets.

One of the main challenges with SR is that it is particularly challenging to recover the high-frequency information lost in the low-resolution domain, which makes super-resolution an ill-posed problem. This issue is further exacerbated by the difficulty of acquiring paired training data. The adversarial generative network (GAN) framework [8] has been used for SR [14, 28, 29] to address these issues and recover high frequency details by learning the distribution of high-resolution unpaired images. This approach can produce visually detailed and pleasing results and has also been adopted in the HSI-SR domain [15, 26]. However, the visually pleasing details predicted by GANs are often not objectively accurate, meaning that in the context of HSI-SR, the loss functions are often heavily constrained with traditional pixel-wise loss functions, consequently necessitating paired training data. Due to the absence of paired HSI-SR data, most researchers in this field resort to synthetic downsampling methods, despite their known limitations [5, 31]. Similar ideas have been applied in RGB super-resolution using beam-splitter rigs, as in the ImagePairs dataset [10], which shows that real-paired data significantly outperforms synthetic baselines. To the best of our knowledge, the proposed dataset in this paper, is the first real hyperspectral super-resolution dataset.

3. Methodology

3.1. Dataset Acquisition

3.1.1. Lens Dataset

We used a Specim FX17 line-scanning camera (900–1700 nm) with lens swapping to capture both low- and high-resolution hyperspectral images over 31 scenes (totaling over 62 million spectra). High-resolution images ($4\times$ magnification) were captured with a 12° lens at 3172×640 pixels (0.104 mm/pixel), while low-resolution images used a 38° lens at 793×160 pixels (0.416 mm/pixel). Focus was maintained via an independent focus measure and verified using a subpixel checkerboard algorithm [9]. A diverse set of objects—including plastics, wood, printed text, metals, fabrics, bitumen, and paintings—was imaged.

3.1.2. Sensor Dataset

To test additional variables (e.g., sensor noise, color reproduction, dynamic range), we constructed a paired dataset

using two hyperspectral cameras. High-resolution (HR) images were captured with a Headwall Micro-Hyperspec VNIR-E camera (1600×373 pixels), and low-resolution (LR) images with a Specim ImSpector V10E spectrograph and Hamamatsu C8484-05 CCD using 4×4 binning (effective resolution: 336×256 pixels). The cameras were arranged in parallel above a translation stage to maintain alignment.

The dataset comprises 50 banknote images registered via crosshair registration. The banknotes were imaged through a $3\,\mathrm{mm}$ acrylic plate with a Spectralon target for one-point calibration, and then cropped to 2160×1080 (HR) and 540×270 (LR) pixels over 234 spectral bands ($412.1\,\mathrm{nm}$ to $991.6\,\mathrm{nm}$). The spatial resolutions are $0.072\,\mathrm{mm/pixel}$ and $0.288\,\mathrm{mm/pixel}$ for HR and LR, respectively, with LR images exhibiting SNR values of $27\,\mathrm{dB}$ to $28\,\mathrm{dB}$.

Both the lens and sensor datasets are made available in an open-source data repository [2].

4. Experiments

4.1. Evaluating Downsampling Methods

The majority of existing literature on single-image super-resolution predominantly relies on a basic, straightforward downsampling approach. However, as we demonstrate in this subsection, such idealized downsampling tends to inaccurately represent the quality of the low-resolution image. Five different downsampling methods have been used: bi-linear, bi-cubic, area, nearest neighbours and lancosz downsampling.

4.1.1. Spatial Differences

In Fig. 1, we conduct a comprehensive comparison between the most widely-used downsampling methods and the empirically measured low-resolution images. This comparison reveals both quantitatively and qualitatively a notable disparity between the actual low-resolution images and those obtained through standard downsampling techniques. A key observation is that these methods frequently fail to accurately capture the extent of image degradation. Specifically, the measured low-resolution images exhibit more pronounced blurring than their downsampled counterparts, highlighting a significant gap in the fidelity of these methods to real-world image degradation.

4.1.2. Spectral Differences

Spectral differences between images are evaluated using Spectral Angle Mapping (SAM)[13], a common metric in hyperspectral imaging that measures the spectral similarity between two spectra as the angle between them in a multi-dimensional space. Given two spectral vectors $\mathbf{x}$ and $\mathbf{y}$, SAM is defined as:

$$\mathrm{SAM}(\mathbf{x}, \mathbf{y}) = \cos^{-1}\left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}\right)$$

Lower SAM values indicate higher spectral similarity.

Fig. 2 compares spectra at three locations (selected by minimum, median, and maximum SAM values) between the actual low-resolution image and various downsampling methods. While the minimum and median SAM values differ little, the maximum SAM location shows that the Nearest Neighbors method (which uses no interpolation) aligns with the high-resolution spectra, unlike other methods that differ notably in the first 10–15 bands—likely due to low sensor quantum efficiency. Ignoring these bands may improve spectral accuracy. These discrepancies also hint at optical distortions unaddressed by current methods, highlighting the need for more advanced approaches.

4.1.3. Optimal Gaussian Filter for Downsampling

To better match captured low-resolution images, we apply Gaussian downsampling:

$$I_{LR} = (I_{HR} * k) \downarrow_s, , \tag{1}$$

where k is a 2D Gaussian kernel, $*$ denotes convolution, and $\downarrow_s$ is 4× subsampling. No noise term is added.

We optimize the Gaussian blur parameter σ using the Nelder-Mead algorithm, finding $\sigma = 3.206$ to yield the best match.

Fig. 3 shows SSIM distributions across scenes, confirming that Gaussian downsampling reduces outliers and improves alignment with real low-resolution data. We recommend this approach when direct low-resolution acquisition is unavailable.

4.2. Super-Resolution

To estimate the actual improvements in using our real paired datasets over using common downsampling methods to generate low-resolution data, a diverse group of super-resolution algorithms was evaluated. We conducted our experiments using the HSI-SR models 3DHSRCNN [17], SRONN [23], and the GAN-based model with a band attention mechanism proposed in [15] which we refer to as BAGAN. This encompasses a variety of models including 2D and 3D models, traditional and GAN-based frameworks, and Self-Organized Operational Neural Networks (Self-ONNs) [12].

We train our models on the real high- and low-resolution pairings, as well as artificial downsampling methods, including bicubic interpolation, and Gaussian downsampling, as is commonly done in the HSI-SR community. For Gaussian downsampling, we select a σ value of 1.6986 for k as is done in [27] to represent how Gaussian downsampling is generally applied in the community. We also carried out experiments using our optimal σ value of 3.206 to explore the performance improvements this offers.

We train by both downsampling the real high-resolution image but also by downsampling the real low-resolution image and using the real low-resolution image as the target

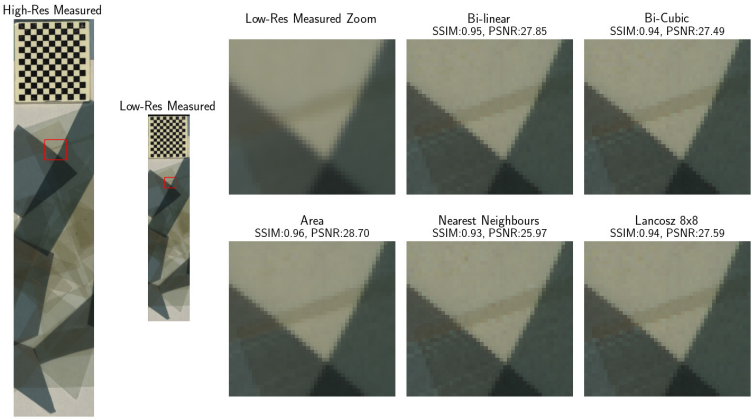

Figure 1. The displayed false RGB images represent high- and low-resolution measurements, with specific wavelengths assigned to each color channel (1004nm for Red, 1283nm for Green, and 1638nm for Blue). On the right, we provide a detailed view of a selected area (highlighted by a red square in both high- and low-resolution images) demonstrating the typical downsampling methods prevalent in existing literature. These enlarged sections facilitate a direct comparison with the low-resolution image using two key metrics: SSIM, PSNR. For convenience, the corresponding values for each metric are presented above their respective plots, allowing for an immediate and clear evaluation of the downsampling methods against the actual low-resolution image quality.

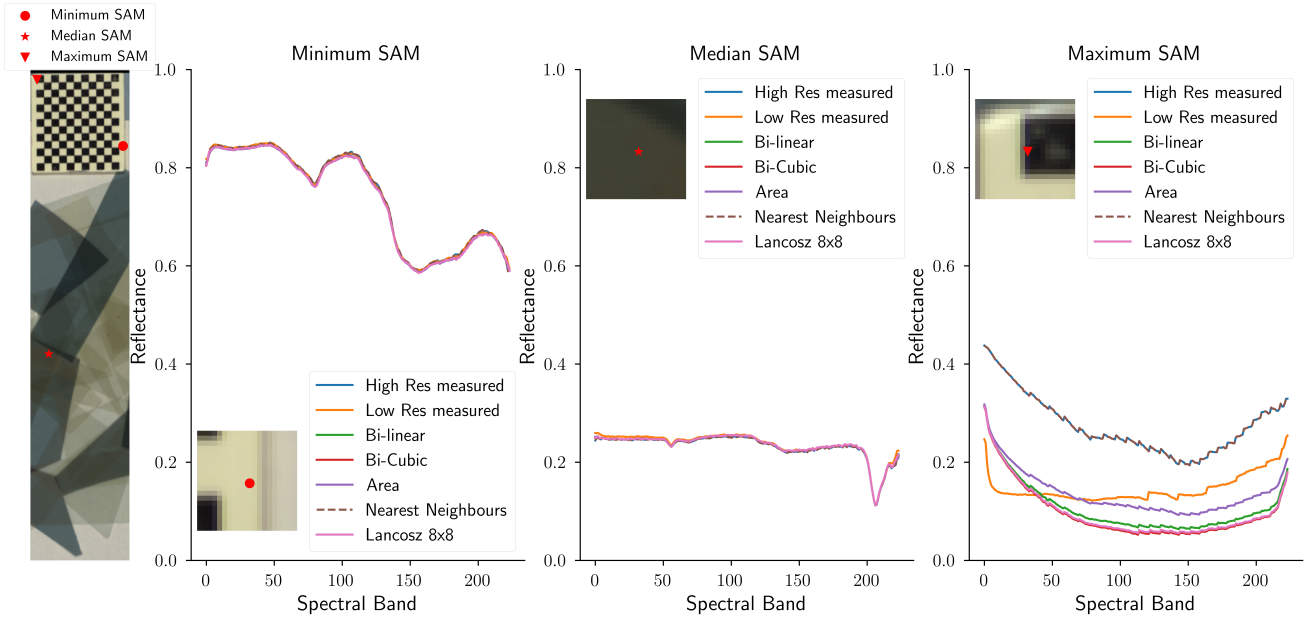

Figure 2. In this figure, three distinct pixels are selectively chosen to represent varying degrees of spectral differences: one with minimal spectral variation, one with the median difference, and another exhibiting the maximum difference. The precise locations of these pixels are indicated in the false RGB image on the left side of the panel. Additionally, the graph includes a zoomed-in view of the region, providing a detailed perspective on these specific areas and their spectral characteristics.

for training. The purpose of this is to simulate how super-resolution models are trained using unpaired images, and we can then evaluate the efficacy of such a training pipeline to enhance the target images themselves by enhancing the real low-resolution image and comparing it to the real high-resolution image. We name these experiments Gaussian and Bicubic Bootstrap. An illustration of our training data pairs can be seen in Fig. 4. For all experiments, we always test and validate on the test and validation portions of the real high- and low-resolution image pairs, even when training on the synthetic downsampling configurations, as this represents the real-world scenario and is the configuration of interest when it comes to evaluation.

Each model has its own set of unique training parame-

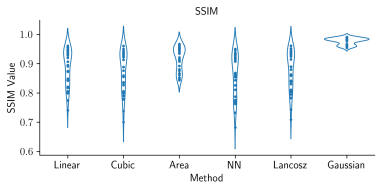

Figure 3. Violin plot showing the improvement on the SSIM metric, when an additional Gaussian preprocessing step is used. Note the closeness of the cluster, instead of the larger spread when the Gaussian preprocessing step is not used.

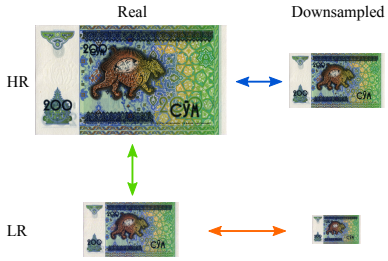

Figure 4. Training Image Pairs (not to scale)

ters, and these parameters are kept the same across all experiments on the different downsampling techniques. We set the optimizer and loss function to be the same as what was used in the papers in which each of the models were proposed. We empirically set the learning rate and the epoch to reduce the learning rate by a factor of 10. For the BAGAN and 3DHSRCNN modes, which both utilize 3D convolution, image patches containing 32 random contiguous bands are extracted and passed to the model during training. For the SRONN model, patches containing all 224 channels are extracted and passed to the model.

We evaluate the performance of our models using the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Spectral Angle Mapper (SAM) and Error Relative Global Adimensional de Synthèse (ERGAS) [25] metrics.

5. Results

Of the 31 scenes available for the Lens Dataset, we use 23 for training, 4 for validation, and 4 for testing.

The experimental results for the Lens Dataset can be seen in Tab. 1 with example outputs from the best-performing model shown in Fig. 5.

The results for the Sensor Dataset can be seen in Tab. 2 with example outputs from the best-performing model shown in Fig. 6. Due to the large amount of sensor noise present in the low-resolution images in this dataset, we also perform median filtering with a 3x3x3 kernel in addition to bicubic interpolation to form baseline metrics on the test data.

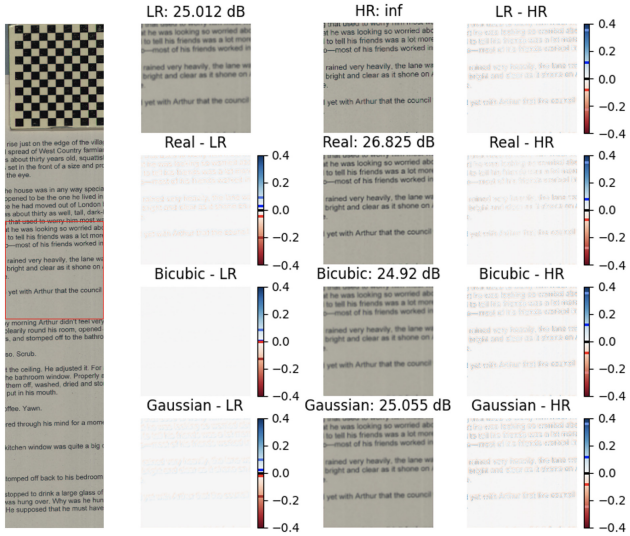

Figure 5. LensDataset super-resolution results. **Left:** False-color image with red box indicating input patch location. **Middle-left:** Low-resolution patch and difference from model predictions, highlighting changes applied. **Middle-right:** Ground-truth high-resolution patch and SRONN predictions (trained on real, bicubic, and Gaussian downsampling); PSNR values shown. **Right:** Pixel-wise difference between predictions and ground truth to assess accuracy.

The results from both datasets in Tab. 1 and Tab. 2 clearly show that training HSI-SR models using real low- and high-resolution data pairs significantly improves performance.

5.1. Lens Dataset

5.1.1. Synthetic Downsampling

In the case of the Lens Dataset, the models trained using the synthetic downsampling processes consistently underperform simple bicubic interpolation when tested on the real pairing. The only exceptions to this are that the BAGAN and SRONN models show a very slight improvement in SSIM when trained using the optimal gaussian filter downsampling method, which itself depended on the real data pairing to get the optimal filter value. Though these models still fail to outperform the bicubic benchmark across the remaining metrics. All other models trained on the synthetic data failed to outperform the bicbuic benchmark across all metrics. These results confirm that training HSI-SR using synthetic downsampling techniques does not perform well on real data and thereby revealing the necessity of using real paired images for super-resolution training.

5.1.2. Real Data

Contrary to the synthetic downsampling, the models trained using the real high- and low-resolution image pairs consistently outperform the bicubic interpoplation benchmark by a significant margin, demonstrating the value of this data.

Table 1. Lens Dataset Results

Training Data Type	Downsampling Method	Architecture	PSNR ↑	SSIM ↑	ERGAS ↓	SAM ↓
(No training data)	N/A	Bicubic Interpolation	33.5884	0.8746	2.3708	2.504
Real Low- & High-Resolution Image Pairs	Real[1]	BAGAN	**35.5471**	**0.9099**	**1.8262**	3.0494
		3DHSRCNN	**34.9113**	**0.8874**	**1.9729**	3.3485
		SRONN	**35.5815**	**0.915**	**1.8211**	2.9236
Synthetically Generated Low-Resolution Pairs from Real High-Resolution Images	Bicubic[2]	BAGAN	33.3865	0.8714	2.4137	2.5112
		3DHSRCNN	33.116	0.8641	2.4968	2.5447
		SRONN	33.472	0.8731	2.398	2.532
	Gaussian[3]	BAGAN	32.4589	0.8674	2.6675	2.6152
		3DHSRCNN	32.6106	0.8644	2.6148	2.7109
		SRONN	32.792	0.8715	2.588	2.687
	Optimal Gaussian[4]	BAGAN	32.595	**0.8766**	2.633	4.109
		3DHSRCNN	32.595	0.8647	2.602	3.768
		SRONN	33.392	**0.8894**	2.398	3.553
Synthetically Generated Low(er)-Resolution Pairs from Real Low-Resolution Images	Bicubic Bootstrap[5]	BAGAN	33.3167	0.8693	2.4386	2.5347
		3DHSRCNN	33.0603	0.8625	2.5185	2.5501
		SRONN	33.393	0.8715	2.4	2.606
	Gaussian Bootstrap[6]	BAGAN	32.272	0.8612	2.7284	2.8004
		3DHSRCNN	32.4324	0.8604	2.6567	2.7927
		SRONN	32.685	0.8735	2.609	2.681

Values in bold indicate superior performance to the bicubic interpolation benchmark.
All test metrics are produced from the real high- and low-resolution test set as this is the real-world scenario.
Downsampling method refers only to the training data.
[1] Real: the real HR and LR pair.
[2] Bicubic: the real HR paired with the bicubic downsampled HR image.
[3] Gaussian: the real HR paired with the gaussian downsampled HR image.
[4] Optimal Gaussian: the real HR paired with the gaussian downsampled HR image with optimal blur value from Sec. 4.1.3.
[5] Bicubic Bootstrap: the real LR paired with the bicubic downsampled LR image.
[6] Gaussian Boostrap: the real LR paired with the gaussian downsampled LR image with standard blur value.

Table 2. Sensor Dataset Results.

Training Data Type	Downsampling Method	Architecture	PSNR ↑	SSIM ↑	ERGAS ↓	SAM ↓
(No training data)	N/A	Bicubic Interpolation	24.704	0.5813	2.266	3.699
		Median Filtering & BI	25.633	0.685	2	3.036
Real Low- & High-Resolution Image Pairs	Real[1]	BAGAN	**28.349**	**0.7558**	**1.567**	**2.625**
		3DHSRCNN	**28.263**	**0.7566**	**1.611**	**2.745**
		SRONN	**26.481**	0.6249	**1.953**	3.465
Synthetically Generated Low-Resolution Pairs from Real High-Resolution Images	Bicubic[2]	BAGAN	**27.657**	**0.7071**	**1.682**	**2.776**
		3DHSRCNN	25.19	0.6271	2.128	3.313
		SRONN	24.701	0.5813	2.267	3.701
	Gaussian[3]	BAGAN	**25.998**	0.6407	2.033	3.350
		3DHSRCNN	23.9923	0.5983	2.437	3.917
		SRONN	24.739	0.5818	2.261	3.713

Bold values indicate superior performance to the bicubic interpolation with median filtering benchmark.
All test metrics are produced from the real high- and low-resolution test set as this is the real-world scenario.
Downsampling method refers only to the training data.
[1] Real: the real HR and LR pair.
[2] Bicubic: the real HR paired with the bicubic downsampled HR image.
[3] Gaussian: the real HR paired with the Gaussian downsampled HR image.

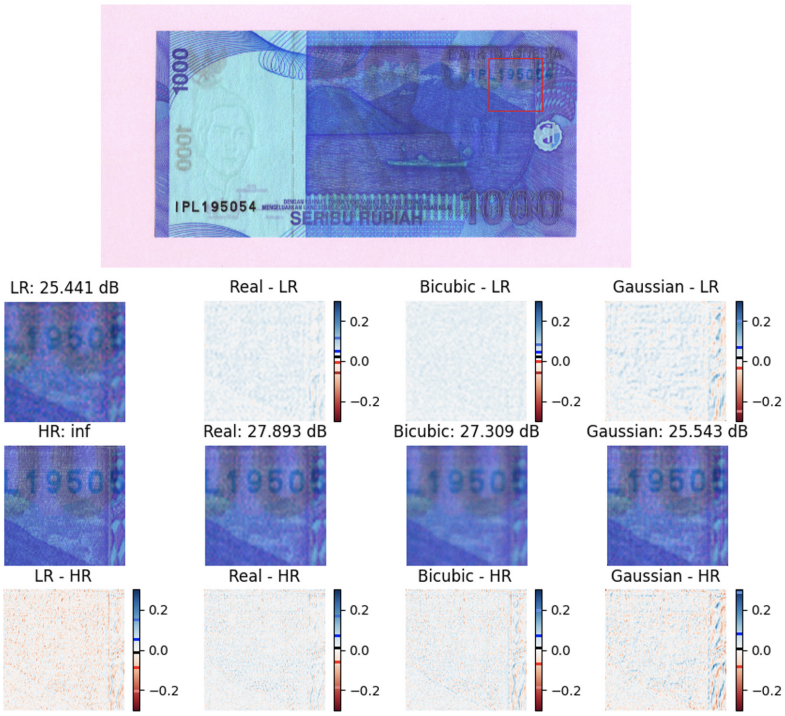

Figure 6. SensorDataset super-resolution results. **Top:** False-color image with red box indicating input patch location. **Top row:** Low-resolution patch and difference from model predictions, showing changes applied during super-resolution. **Middle row:** Ground-truth high-resolution patch and BAGAN predictions (trained on real, bicubic, and Gaussian downsampling); PSNR values shown. **Bottom row:** Pixel-wise difference between predictions and ground truth to assess accuracy.

This improved super-resolution is also evident in Fig. 5 where the model trained on real data produces a significantly shaper and more detailed patch than the output of the two models trained using data downsampled with bicubic and gaussian downsampling.

Interestingly, the models trained on the real pairing do not outperform bicubic interpolation in terms of SAM and in fact perform worse than the models trained with artificial downsampling methods for this metric. This could be due to the fact that artificial downsampling methods produce a more detailed synthetic low-resolution image (shown in Fig. 1) than the true low-resolution image, reducing the complexity of the low- to high-resolution mapping function learned by models trained with these data compared to models trained on the real low-resolution image. Thus, when the models trained on synthetically generated low resolution images are applied to real low-resolution data, they produce smoother and less detailed outputs, as shown in Fig. 5 and evidenced by their low spatial metrics. This smoother output could also lead to "cleaner" spectra and may be why they produce better SAM values than the models trained on the real data that are attempting to recover greater spectral detail. Though notably, these SAM values are still worse than the bicbuic benchmark.

5.1.3. Optimized Gaussian downsampling

Although the results from the models trained with the bicubic downsampling method objectively produce better results than the models trained with Gaussian downsampling on the Lens dataset, the bicubic downsampling methods produce results that are very similar to the bicubic interpolation baseline. All experimental models contain residual connections and in the case of the bicubic downsampling methods, models learn to output predictions very similar to the input as the blank "Bicubic - LR" patch of Fig. 5 shows. While the Gaussian downsampling method produces objectively worse metrics than the bicubic downsampling method, the non-blank "Gaussian - LR" patch in Fig. 5 shows that the model is in fact changing the input and learning something, though it lacks accuracy as the "Gaussian - HR" patch of Fig. 5 reveals. By optimizing the σ value used in the Gaussian downsampling method, the performance of this method improves in terms of spatial metrics when the models are applied to real data, slightly output-performing bicubic interpolation in terms of SSIM. However, even with an optimized σ value, the performance of models trained on synthetic data remains significantly inferior to those trained on real image pairs. This gap underscores the limitations of synthetic downsampling methods, as they fail to capture the

complexities and nuances of real-world data. Notably, the SAM metric deteriorates further with an optimized σ value, mirroring the results seen when models are trained on real data pairs. This deterioration highlights that, despite efforts to refine synthetic training data, only real images provide the necessary detail and variability for models to achieve superior performance, reinforcing the importance of using authentic datasets for effective training.

5.1.4. Bootstrap experiments

The bootstrap experiments presented in Tab. 1 produce fairly similar results to the standard artificial downsampling experiments, indicating that super-resolution performance may be translatable to higher scales. However, given that there are residual connections present in each experimental model and the models do not outperform bicubic interpolation, it is not conclusive whether or not the models are actually translatable across scales or if the models in fact just learn to output an image similar to the image interpolated with bicubic interpolation, making it appear scale-invariant. A model that offers improvements over bicubic interpolation would be required to draw any firm conclusions on this. Based on our results, an even lower-resolution data pair would have to be acquired in order to train the model on a real lower-scale data pair to outperform bicubic interpolation at this scale before evaluating on the higher scale.

5.2. Sensor Dataset

5.2.1. Synthetic Downsampling

Similar to the results on the Lens Dataset, the models trained using the synthetic downsampling processes on the Sensor Dataset consistently underperform as compared to the bicubic interpolation benchmark. Only the BAGAN model outperforms bicubic interpolation with median filtering when trained on data synthetically generated low-resolution images. This may well be due to the adversarial loss function used to train the BAGAN model enabling the model to learn the low-noise distribution of the target high-resolution data and thus producing smoother outputs when fed a noisy (real) low-resolution input image, enabling better performance than bicubic interpolation with median filtering due to its denoising capacity rather than its super-resolution capacity. This is evident in Fig. 6 as the output patch from the model trained with bicubic interpolation produces a PSNR value significantly better than the low-resolution image and somewhat close to the output from the model trained on the real pair, but the patch produced is much more blurry than the true high-resolution image and the predicted output of the model trained on the real data.

5.2.2. Real Data

The models trainied on the real data pair from the Sensor dataset consistently outperform the bicbuic interpolation with median filtering baseline, and also consistently outperform the models trained with synthetic downsampling techniques by a significant margin, further demonstrating the value of training using real data pairs.

The SRONN model performs significantly worse on this dataset than the two other experimental models on this dataset which is most likely due to the 3D filters in the other models being better suited to tackling the higher levels of noise within the low-resolution patches of this dataset than the 2D filters used in the SRONN model. Interestingly, the BAGAN model trained with bicbuic downsampling is able to outperform the SRONN model trained on the real data pair. This is likely due to a combination of the SRONN model's 2D filters and the reasons discussed in the previous sub-section. Despite this, all experimental models see a significant performance improvement across all metrics when trained using the real data pair over the synthetic downsampling methods.

6. Conclusion

In this study, we identified the inherent limitations of the standard approach of training super-resolution models with synthetically downsampled images. To counter these limitations, we introduced a novel dataset of real low- and high-resolution hyperspectral image pairs, captured with different lenses and sensors, marking a significant departure from conventional practices.

Our comprehensive experiments demonstrate that the models trained on this novel dataset significantly outperform those trained on artificially downsampled images in terms of visual quality and quantitative metrics. This finding not only underscores the inadequacy of traditional synthetic downsampling methods, but also highlights the substantial benefits of utilizing real image pairs for training. It is our hope that the proposed dataset will be used not only to train and evaluate more accurate super-resolution models, but also to develop more realistic artificial degradation techniques that can be used to generate future synthetic datasets.

Moving forward, there lies a promising opportunity to enhance the diversity and applicability of our dataset by extending its range to include more scenarios and different levels of magnification. Achieving this could involve leveraging a wider array of lenses or delving into other imaging techniques, including multispectral and higher-resolution RGB images. This strategic expansion would not only deepen our dataset's versatility, but also pave new pathways for research in super-resolution technology.

References

[1] Nour Aburaed, Mohammed Q. Alkhatib, Stephen Marshall, Jaime Zabalza, and Hussain Al Ahmad. Hyperspectral data scarcity problem from a super resolution perspective: Data

augmentation analysis and scheme. In *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 5057–5060, 2023. 1

[2] Anonymous. Data from the paper: A true hyperspectral image super-resolution dataset. Hyperspectral Images Version V1, Some Institution, 2024. 2, 3

[3] Pattathal V Arun, Krishna Mohan Buddhiraju, Alok Porwal, and Jocelyn Chanussot. CNN-based super-resolution of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6106–6121, 2020. 1, 2

[4] David J Brady. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009. 1

[5] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 1, 2

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1, 2

[7] Chao Dong, Chen Change Loy, and Xiaoou 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*, pages 391–407. Springer, 2016. 1, 2

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[9] Michaël Hillen, Ivan De Boi, Thomas De Kerf, Seppe Sels, Edgar Cardenas De La Hoz, Jona Gladines, Gunther Steenackers, Rudi Penne, and Steve Vanlanduit. Enhanced checkerboard detection using gaussian processes. *Mathematics*, 11(22), 2023. 2

[10] Hamid Reza Vaezi Joze, Ilya Zharkov, Karlton Powell, Carl Ringler, Luming Liang, Andy Roulston, Moshe Lutz, and Vivek Pradeep. Imagepairs: Realistic super resolution dataset via beam splitter camera rig. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 518–519, 2020. 2

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1, 2

[12] Serkan Kiranyaz, Junaid Malik, Habib Ben Abdallah, Turker Ince, Alexandros Iosifidis, and Moncef Gabbouj. Self-organized operational neural networks with generative neurons. *Neural Networks*, 140:294–308, 2021. 2, 3

[13] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2–3):145–163, 1993. 3

[14] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1, 2

[15] Jiaojiao Li, Ruxing Cui, Bo Li, Rui Song, Yunsong Li, Yuchao Dai, and Qian Du. Hyperspectral image super-resolution by band attention through adversarial learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58 (6):4304–4318, 2020. 1, 2, 3

[16] Denghong Liu, Jie Li, Qiangqiang Yuan, Li Zheng, Jiang He, Shuheng Zhao, and Yi Xiao. An efficient unfolding network with disentangled spatial-spectral representation for hyperspectral image super-resolution. *Information Fusion*, 94:92–111, 2023. 1

[17] Wei Liu and Joonwhoan Lee. An efficient residual learning neural network for hyperspectral image superresolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(4):1240–1253, 2019. 1, 2, 3

[18] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 457–466, 2022. 1, 2

[19] Armin Mehri, Parichehr Behjati, and Angel Domingo Sappa. Tntvit-g: Transformer in transformer network for guidance super resolution. *IEEE Access*, 11:11529–11540, 2023. 1, 2

[20] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral image spatial super-resolution via 3D full convolutional neural network. *Remote Sensing*, 9 (11):1139, 2017. 1, 2

[21] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015. 2

[22] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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, 2016. 2

[23] Alexander Ulrichsen, Paul Murray, Stephen Marshall, Moncef Gabbouj, Serkan Kiranyaz, Mehmet Yamac, and Nour Aburaed. Operational neural networks for parameter-efficient hyperspectral single-image super-resolution. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pages 1–17, 2023. 1, 2, 3

[24] Alberto G Villafranca, Jordi Corbera, Francisco Martín, and Juan Fernando Marchán. Limitations of hyperspectral earth observation on small satellites. *Journal of Small Satellites*, 1 (1):19–29, 2012. 1

[25] Lucien Wald. *Data fusion: Definitions and architectures: Fusion of images of different spatial resolutions*. Presses des MINES, 2002. 5

[26] Baorui Wang, Yifan Zhang, Yan Feng, Bobo Xie, and Shaohui Mei. Attention-enhanced generative adversarial network for hyperspectral imagery spatial super-resolution. *Remote Sensing*, 15(14):3644, 2023. 1, 2

[27] Liguo Wang, Tianyi Bi, and Yao Shi. A frequency-separated 3D-CNN for hyperspectral image super-resolution. *IEEE Access*, 8:86367–86379, 2020. 3

[28] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1, 2

[29] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 1, 2

[30] Yuan Yuan, Xiangtao Zheng, and Xiaoqiang Lu. Hyperspectral image superresolution by transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):1963–1974, 2017. 1, 2

[31] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 1, 2

[32] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3757–3765, 2019. 2