White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/241145/

Version: Accepted Version

Proceedings Paper:
Alsari, A., Zhang, J., Farr, N.T.H. et al. (2026) A spectral-spatial deep learning for secondary electron hyperspectral image super-resolution. In: 2026 29th International Conference on Information Fusion (FUSION). 29th International Conference on Information Fusion (FUSION 2026), 23-26 Jun 2026, Trondheim, Norway. Institute of Electrical and Electronics Engineers (IEEE). ISBN: 9798331552350. ISSN: 2707-8787. EISSN: 2707-8787.

https://doi.org/10.23919/FUSION67272.2026.11726671

© 2026 The Author(s). Except as otherwise noted, this author-accepted version of a paper published in 2026 29th International Conference on Information Fusion (FUSION) is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

A Spectral-Spatial Deep Learning for Secondary Electron Hyperspectral Image Super-Resolution

Ali Alsari
School of Electrical and Electronic Engineering
University of Sheffield, UK
ahalsari1@sheffield.ac.uk

Jingqiong Zhang
School of Electrical and Electronic Engineering
University of Sheffield, UK
jingqiong.zhang@sheffield.ac.uk

Nicholas T.H Farr
School of Chemical, Materials and Biological Engineering
University of Sheffield, UK
n.t.farr@sheffield.ac.uk

Cornelia Rodenburg
School of Chemical, Materials and Biological Engineering
University of Sheffield, UK
c.rodenburg@sheffield.ac.uk

Lyudmila Mihaylova
School of Electrical and Electronic Engineering
University of Sheffield, UK
l.s.mihaylova@sheffield.ac.uk

***Abstract*—The secondary electron hyperspectral imaging (SEHI) is a promising surface analysis method, especially for nano/micro chemical materials. However, SEHI poses challenges for achieving a trade-off between low-resolution imaging over a large field of view and high-resolution imaging over a limited field of view, and avoiding potential sample alteration caused by electron beam exposure. To address this limitation, we introduce, for the first time, a spectral-spatial deep learning model (SSDL) that produces a super-resolution SEHI image from a low-resolution SEHI image, while substantially reducing the risk of sample alteration. Specifically, the SSDL model incorporates a spectral squeeze-and-excitation (S-E) module and a spectral spatial fusion block, enhancing its flexibility in extracting both spatial and spectral features. Further, inception-residual modules are embedded within the proposed SSDL to capture multi-scale structural features. A custom spatial–spectral loss function is introduced, integrating a mean squared error (MSE), a spatial gradient loss, and a spectral angle mapper loss, which are essential for simultaneously preserving spectral and spatial integrity. The proposed model shows high performance compared to state-of-the-art methods during testing on two real SEHI datasets. Ablation experiments further confirm the effectiveness of the spectral SE block, inception-residual modules, and the custom loss function. This work demonstrates that deep learning can greatly enhance SEHI by considerably increasing the scanning throughput while preserving high-resolution, thereby improving the practical utility of SEHI for materials science.**

***Index Terms*—Super-resolution, Secondary electron hyperspectral imaging, Deep learning**

I. Introduction

Scanning electron microscopy (SEM) is an indispensable technique for characterizing nanomaterials, as it utilizes beams of electrons to produce highly detailed images. SEM is an essential technique across numerous scientific and engineering disciplines, such as materials, physics, and nanotechnology [1], but it does not provide chemical information on the nature of the sample. Secondary electron hyperspectral imaging (SEHI) is a method that can fill in this gap by employing energy filtering within an SEM to produce a three-dimensional hyperspectral dataset. SEHI employs the mirror electrode in an Elstar column (Thermo Fisher/FEI) as a low-pass energy filter to capture the field of view (FoV) sequentially with increasing energy cutoff, as shown in Fig. 1a. By stepping the mirror-electrode bias, the column selectively transmits only those secondary electrons (SEs) that satisfy the chosen cutoff condition to the through-lens detector (TLD) [2]. Repeating this process builds a three-dimensional data volume with spatial dimensions (x, y) and an energy-filter dimension (z), Fig. 1b. This SEHI data volume can be differentiated along the energy axis to recover SEs energy spectrum, Fig. 1c [3]. SEHI can achieve a resolution down to a few nanometers. However, for many material applications, achieving sub-nanometer spatial resolution over a large FoV needs multiple scans. Consequently, SEHI often adopts a FoV resolution trade-off: high-resolution (HR) imaging over a small FoV or low-resolution (LR) imaging over a larger FoV. The limited FoV at the acquisition stage makes it difficult to acquire a HR image with a large a large FoV. In addition, scanning multiple passes in electron microscopy can induce sample damage and is time-consuming, particularly in specimens that are highly susceptible to electron irradiation, such as pharmaceutical particles, polymers, and biomaterials [2], [3]. To alleviate these challenges, a large FoV at LR may be acquired, and a method might be developed to boost SEHI resolution within less scanning time, thereby protecting the sample from damage while still enabling high-resolution imaging. To meet the demand for high-quality SEHI data, this study, for the first time, highlights the significant importance of super-resolution (SR) techniques for reconstructing HR SEHI images with enhanced spatial resolution.

In the early development of image SR, most methods were mainly based on traditional techniques, such as interpolation-based [4], regularization-based [5], and sparse representation approaches [6]. In particular, interpolation-based methods, such as nearest-neighbour, bilinear, band bicubic, depend on

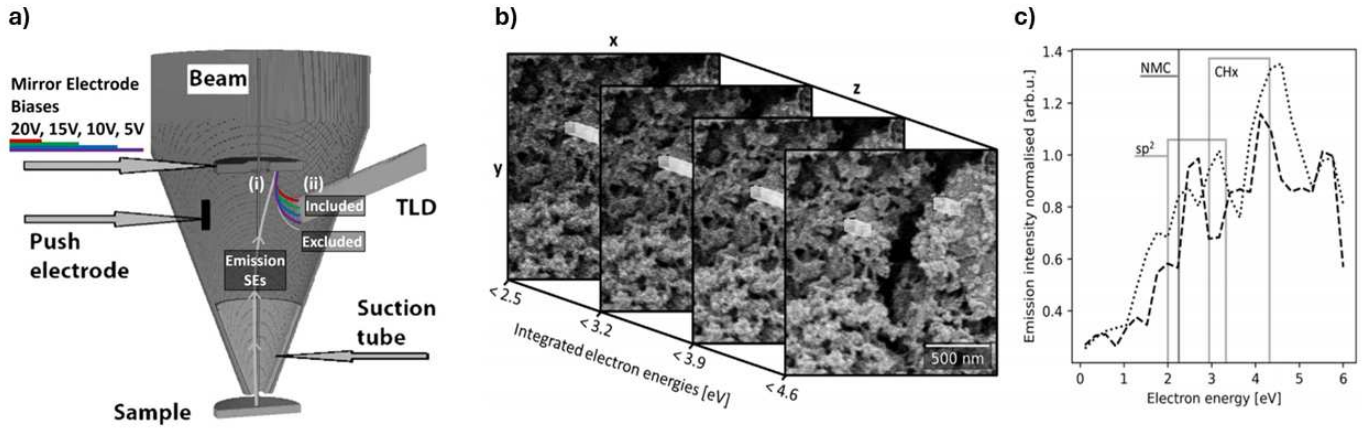


Fig. 1. a) Schematic of the Helios Elstar column adapted from [2] under CC BY 4.0 license, emitted electrons are accelerated by the suction tube, (i) the mirror electrode subsequently deflects them as its bias is incrementally stepped; and (ii) depending on this bias, they are either included or excluded from detection by the TLD. b) SEHI slices from a lithium-ion battery cathode data cube with spatial (x, y) and energy (z) dimensions, adapted from [3] under CC BY 4.0 license. c) Secondary-electron spectra were extracted from 100 nm × 100 nm square regions of interest matching the boxed areas in (b), adapted from [3] under CC BY 4.0 license.

interpolation kernels and are fast but often tend to degrade reconstruction quality due to their limited ability to restore high-frequency components [4]. Regularization-based approaches formulate SR as an inverse problem with handcrafted methods and are computationally intensive [5]. In a similar manner, sparse representation approaches are characterized by computational challenges, alongside the potential to introduce artifacts, as they represent each LR patch as a sparse linear combination over a learned dictionary and then map it to an HR dictionary to reconstruct details [6]. Moreover, traditional techniques relied exclusively on pixel-level computations, disregarding image content and prior information. As a result, reconstructed high-resolution images frequently lacked fine details. Furthermore, these methods often exhibit limited generalization across varied image content.

Nowadays, hyperspectral image SR (HSI SR) methods are typically categorized into two primary methods, which are distinguished by their use of auxiliary images: fusion-based hyperspectral image SR and single-image SR (SISR) [7]. In the fusion-based methods, spatial detail is enhanced by integrating a hyperspectral image and a corresponding high-resolution image of the same target scene. Techniques such as sparse representation [8] and conditional diffusion models [9] have achieved high-quality results for this task. However, these methods typically require the availability of HR auxiliary information, which must be precisely aligned with the LR hyperspectral image, often making it difficult to satisfy in real-world practical applications [10]. In contrast, SISR methods are designed to produce high-resolution data directly from low-resolution data, thereby eliminating the need for auxiliary information. For instance, multiple depthwise separable convolutions incorporated with an adilated convolution fusion block are presented to extract both spatial and spectral features [11]. However, the depthwise separable dilated network may suffer from limited global context and spectral distortion. Muhammad et al. [12] introduced an augmented convolutional neural network that incorporates multi-head attention, enabling the network to capture local and long-range feature relationships for enhanced SR. Wang et al. [13] proposed a spatial–spectral aggregation transformer framework that gradually recovers high-resolution images through the integration of CNNs and transformers. It introduces specially designed attention mechanisms for HSI, including spatial and spectral attention modules. However, multi-head attention and transformer models introduce higher computational complexity and risks of overfitting on small datasets. A spectral–spatial fusion module and multiscale spatial model are proposed by [14] to effectively integrate spectral and spatial information to further enhance reconstruction quality. Similarly, a single-HSI SR network that uses a progressive upsampling deep prior to reconstruct the HR-HSI from a single LR-HSI is introduced [15]. It combined a spatial–spectral attention fusion unit and group-convolution upsampling to better exploit nonlocal spatial similarity and spectral correlation. Moreover, [16] introduces a hybrid deep learning SISR model that adds a spectral–spatial unmixing fusion (SSUF) module into a standard CNN. SSUF explicitly performs spectral unmixing and spectral–spatial feature extraction, and a spatial–spectral gradient loss helps preserve both fine textures and spectral fidelity during SR of a single LR-HSI. However, due to the absence of complementary high-resolution inputs, single-image hyperspectral SR can only extract fine-scale detail over a limited spatial region, which frequently results in a decline in reconstruction image quality compared with fusion-based methods [17]. In addition, SISR methods mentioned above suffer from limited ability to fully leverage the extraction of emphasized informative channels or often depend on computationally expensive transformers and global self-attention. Further, they do not adequately exploit multiscale features in SEHI images, which are essential for capturing both fine textures and larger structural patterns.

Furthermore, when an image is downsampled, high-frequency information is lost, including textural patterns, distinct boundaries, and subtle structural elements, making single-image SR an ill-posed resolution challenge task.

To alleviate the aforementioned challenges, we introduce a spectral spatial network aimed at improving reconstruction image quality through spectral residual S-E, spectral-spatial fusion, deep feature extraction, upsampling, and a custom loss function. Specifically, SSDL first employs a spectral-attention feature extractor composed of two successive 1×1 convolutional layers, followed by a residual S-E block [18], spectral–spatial fusion units [19], and inception-style structures [20] to capture heterogeneous contextual information and to refine feature fusion by jointly accentuating fine spectral attributes and local spatial correlations. Finally, we have an optimised upsampling based on a learnable transposed convolution and bilinear interpolation, followed by residual skip connections [21], to preserve fine details, as shown in Fig. 2. Ablation studies demonstrate the roles of S-E, the inception module under different conditions, and custom loss function configurations.

In summary, the contributions of this paper are as follows:

1) To the best of our knowledge, this work represents the first application of the deep learning model to the field of SEHI SR. The proposed deep learning SR model combines a spectral residual SE block, a spectral-spatial fusion module, an inception residual module, and an upsampling block to effectively capture both high- and low-level features and improve the learning ability of complex spectral-spatial relationships.
2) We design a spectral residual S-E Block that combines nonlinear spectral mixing, S-E module, and residual connection to effectively preserve both spatial and spectral details. A custom spatial–spectral loss function is designed, integrating mean squared error (MSE), a spatial gradient loss, and a spectral angle mapper loss to ensure high fidelity reconstruction.
3) Comprehensive results from experiments under 2×, 4×, and 8× downsampling scenarios on two real SEHI datasets demonstrate that our proposed method significantly outperformed state-of-the-art methods in both visual quality and quantitative performance metrics.
4) The proposed data-driven approach can contribute to automated materials analysis without causing specimen damage due to the effect of the electron beam.

II. Methodology

The overall pipeline of the proposed SSDL model is summarized in Fig. 3, comprising four key parts: a spectral residual S-E block, a spectral spatial fusion module, an inception residual module, and an upsampling block. We begin with problem formulation. The following subsections present a detailed analysis of each component of the proposed model. The overall goal is to reconstruct the HR SEHI image from the LR input while preserving both spatial structures and spectral signatures, which is critical in SEHI analysis.

A. Problem Formulation

Let $\mathbf{X}_{\mathrm{HR}} \in \mathbb{R}^{H\times W\times B}$ denote the target HR SEHI image with spatial size (H, W) and B spectral bands. Given its LR counterpart $\mathbf{X}_{\mathrm{LR}}$, the model learns a mapping

$$\mathcal{M}_\theta(\mathbf{X}_{\mathrm{LR}}) = \hat{\mathbf{X}}_{\mathrm{HR}}, \tag{1}$$

where $\hat{\mathbf{X}}_{\mathrm{HR}}$ is the reconstructed HR SEHI and θ are the trainable parameters. The model is trained to minimize a composite loss function that enforces pixel-wise fidelity and spatial-spectral gradient consistency between $\hat{\mathbf{X}}_{\mathrm{HR}}$ and $\mathbf{X}_{\mathrm{HR}}$.

B. Spectral Residual S-E Block

The input is first passed through two successive 1×1 convolutional layers to perform nonlinear spectral mixing at each spatial location, expanding the raw spectral bands into a higher-dimensional feature space. We then embed a squeeze-and-excitation (S-E) block that models inter-channel dependencies and flexibly concentrates on the most informative features [18]. It aggregates global contextual information via global average pooling and produces a set of channel-wise attention weights that rescale the spectral features. Finally, a residual connection [21] is applied between the input and the reweighted features, stabilizing optimization and allowing the block to learn a spectral refinement rather than a complete transformation. Mathematically, given an input tensor $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$, we compute

$$\mathbf{U} = \phi\big(\mathbf{W}_1 * \mathbf{X}\big), \tag{2}$$

$$\mathbf{V} = \phi\big(\mathbf{W}_2 * \mathbf{U}\big), \tag{3}$$

where $*$ denotes convolution, and $\phi(\cdot)$ denotes a nonlinear activation function, such as ReLU. $\mathbf{W}_1 \in \mathbb{R}^{1\times 1\times B\times F}$ and $\mathbf{W}_2 \in \mathbb{R}^{1\times 1\times F\times F}$ are the convolutional weights for spectral mixing, and $\mathbf{U}, \mathbf{V}$ are an intermediate feature map with F channels. A channel-wise attention mechanism is then obtained using a squeeze-and-excitation (S-E) mechanism. First, global average pooling computes a channel descriptor $\mathbf{z} \in \mathbb{R}^F$:

$$\mathbf{z} = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbf{V}_{i,j,:}. \tag{4}$$

The descriptor is passed through a bottleneck multilayer perceptron(two fully-connected layers) with a reduction ratio r to produce gating coefficients $\mathbf{s} \in (0, 1)^F$:

$$\mathbf{s} = \mathrm{Sigmoid}(\mathbf{W}_2^{\mathrm{se}}\,\phi(\mathbf{W}_1^{\mathrm{se}}\mathbf{z}))\,, \tag{5}$$

where $\mathrm{Sigmoid}(\cdot)$ denotes the sigmoid activation function, $\mathbf{W}_1^{\mathrm{se}} \in \mathbb{R}^{\frac{F}{r}\times F}$ reduces the channel dimension and $\mathbf{W}_2^{\mathrm{se}} \in \mathbb{R}^{F\times\frac{F}{r}}$ expands it back. The re-weighted features are

$$\tilde{\mathbf{V}}_{i,j,:} = \mathbf{s} \odot \mathbf{V}_{i,j,:}, \tag{6}$$

with $\odot$ denoting channel-wise multiplication. Finally, a residual connection is added:

$$\mathbf{F}_0 = \phi\Big(\tilde{\mathbf{V}} + \mathbf{W}_r * \mathbf{X}\Big)\,, \tag{7}$$

where $\mathbf{W}_r \in \mathbb{R}^{1\times 1\times B\times F}$ is the residual projection (or identity when $B = F$),

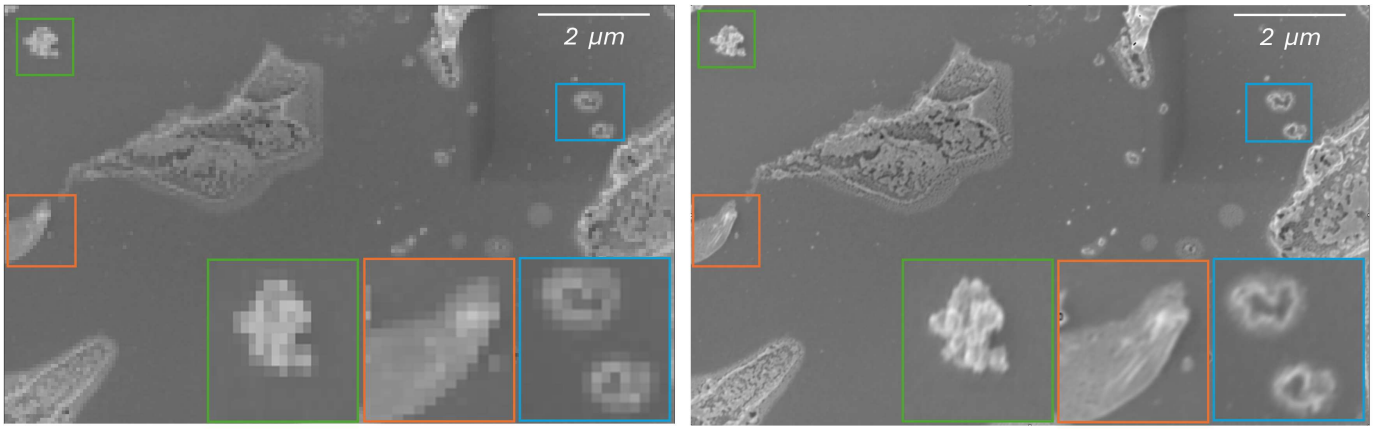


Fig. 2. Comparison of SEHI micrograph resolution Palladium, Silver, and Carbon (Pd-Ag-C) metal film, left lower resolution, right Our SSDL super resolution. Coloured boxes mark regions of interest; the corresponding bottom insets show enlarged views of these boxed areas, illustrating the improved delineation of fine features after super-resolution. Scale bar: 2 μm.

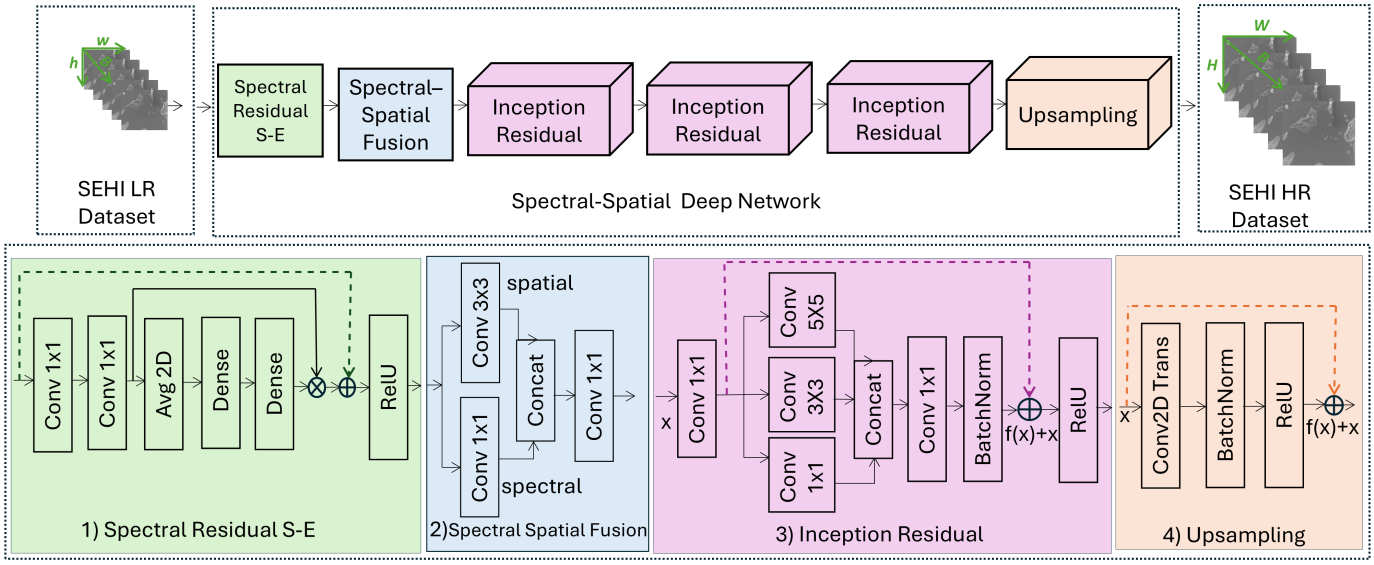


Fig. 3. The workflow of the proposed SSDL architecture proceeds from left to right as follows: (1) a spectral residual squeeze-and-excitation (S-E) module; (2) a spectral–spatial fusion block designed to refine intermediate feature representations; (3) inception-based multiscale feature extractor modules that learn hierarchical representations; and (4) an upsampling block that reconstructs high-resolution SEHI images.

C. Spectral-Spatial Fusion module

To jointly refine spectral and spatial information, we design a fusion module [16] which adds the spectral-spatial context to the original network

$$\mathbf{F}_{\text{spec}} = \phi\big(\text{Conv}_{1\times1}(\mathbf{F}_0)\big), \tag{8}$$

$$\mathbf{F}_{\text{spat}} = \phi\big(\text{Conv}_{3\times3}(\mathbf{F}_0)\big), \tag{9}$$

The outputs from both branches are concatenated and passed through an additional 1×1 convolution for fusion:

$$\mathbf{F}_1 = \phi\big(\text{Conv}_{1\times1}([\mathbf{F}_{\text{spec}}, \mathbf{F}_{\text{spat}}])\big), \tag{10}$$

where $\mathbf{F}_0 \in \mathbb{R}^{H\times W\times F}$ denotes the input feature map, $\mathbf{F}_{\text{spec}}$ and $\mathbf{F}_{\text{spat}}$ are the spectrally and spatially refined feature maps, respectively, and $\mathbf{F}_1$ is the fused output feature map. $\text{Conv}_{k\times k}(\cdot)$ denotes a convolution with kernel size $k \times k$, and $[\cdot,\cdot]$ represents channel-wise concatenation.

D. Inception Residual Module

To capture multi-scale spatial structures without excessively increasing the number of parameters, we adopt an Inception-style residual block [20]. Given an input feature map $\mathbf{F}$, we process it with three parallel branches:

- a 1×1 convolution followed by another 1×1 convolution,
- a 1×1 convolution followed by a 3×3 convolution,
- a 1×1 convolution followed by a 5×5 convolution.

Let b_1, b_2, and b_3 denote the outputs of these three branches, respectively. We concatenate the branch outputs along the channel dimension and project them back to the target number

of channels via a 1×1 convolution, followed by batch normalization:

$$\mathbf{F}_{\text{inc}} = \text{BN}\big(\text{Conv}_{1\times1}(\text{Concat}([b_1, b_2, b_3]))\big). \tag{11}$$

We then apply a residual connection [21] and a point-wise nonlinearity:

$$\mathbf{F}_{\text{out}} = \phi\big(\mathbf{F}_{\text{inc}} + \mathbf{F}\big). \tag{12}$$

We stack three such blocks in series, which enables the model to learn rich multi-scale spatial features while stabilizing training through batch normalization and residual learning, ultimately improving SR performance. It is worth noting that for this part, we experimented with one to five blocks, and three blocks achieved the best performance in image reconstruction. Therefore, three blocks were used in this part.

E. Upsampling Block

To increase the desired spatial resolution, we design an upsampling block based on a learnable transposed convolution [22]. Let $\mathbf{X}_{\text{in}} \in \mathbb{R}^{H\times W\times F}$ denote a low-resolution feature tensor. The transposed convolution layer produces a higher-resolution feature map

$$\mathbf{X}_{\text{tr}} = \mathbf{W}_{\text{tr}} * \mathbf{X}_{\text{in}}, \tag{13}$$

where $\mathbf{W}_{\text{tr}} \in \mathbb{R}^{k\times k\times F\times B}$ is the trainable kernel of the transposed convolution, and $\mathbf{X}_{\text{tr}} \in \mathbb{R}^{\alpha H\times \alpha W\times B}$ is the upsampled output with spatial scaling factor α in both height and width. A pointwise nonlinearity is then applied:

$$\mathbf{X}_{\text{act}} = \text{ReLU}(\mathbf{X}_{\text{tr}}). \tag{14}$$

To further propagate information from the input, we introduce a skip pathway. First, the input feature map is upsampled by a fixed operator $\mathcal{U}(\cdot)$ (e.g., nearest-neighbor or bilinear interpolation), and then projected by a 1×1 convolution:

$$\mathbf{X}_{\text{skip}} = \mathbf{W}_{\text{skip}} * \mathcal{U}(\mathbf{X}_{\text{in}}), \tag{15}$$

where $\mathbf{W}_{\text{skip}} \in \mathbb{R}^{1\times1\times F\times B}$ is the projection kernel. The final output of the upsampling block is obtained by combining the main and skip branches:

$$\mathbf{X}_{\text{out}} = \mathbf{X}_{\text{act}} + \mathbf{X}_{\text{skip}}, \tag{16}$$

yielding $\mathbf{X}_{\text{out}} \in \mathbb{R}^{\alpha H\times \alpha W\times B}$.

F. Custom Spatial-Spectral Loss

Conventional SR models often rely solely on mean squared error (MSE), which encourages high PSNR but may oversmooth edges and distort spectral curves. To address this, we define a loss composed of three terms

$$\mathcal{L} = \lambda_{\text{mse}}\,\mathcal{L}_{\text{mse}} + \lambda_{\text{spatial}}\,\mathcal{L}_{\text{spatial}} + \lambda_{\text{sam}}\,\mathcal{L}_{\text{sam}}, \tag{17}$$

where $\lambda_{\text{mse}}, \lambda_{\text{spatial}}, \lambda_{\text{sam}}$ are positive scalars.

a) Pixel Loss: The MSE loss is

$$\mathcal{L}_{\text{mse}} = \frac{1}{N}\sum_{i=1}^{N} \big\|\mathbf{Y}_i - \hat{\mathbf{Y}}_i\big\|_2^2, \tag{18}$$

where $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are the ground-truth and predicted HR SEHI, respectively, and $N = H \times W \times B$.

b) Spatial Gradient Loss: To preserve edges and local structures, we penalise the difference between spatial gradients of prediction and ground truth along horizontal and vertical directions [23]

$$\mathcal{L}_{\text{spatial}} = \frac{1}{N}\left(\big\|\nabla_x\mathbf{Y} - \nabla_x\hat{\mathbf{Y}}\big\|_2^2 + \big\|\nabla_y\mathbf{Y} - \nabla_y\hat{\mathbf{Y}}\big\|_2^2\right), \tag{19}$$

where ∇_x and ∇_y denote discrete spatial gradients along the vertical (height) and horizontal (width) directions, respectively. They are implemented as forward finite differences

$$\begin{aligned}(\nabla_x\mathbf{Y})_{b,i,j,c} &= \mathbf{Y}_{b,i+1,j,c} - \mathbf{Y}_{b,i,j,c},\\ (\nabla_y\mathbf{Y})_{b,i,j,c} &= \mathbf{Y}_{b,i,j+1,c} - \mathbf{Y}_{b,i,j,c},\end{aligned} \tag{20}$$

for b indexing the batch, c the channel, and $1 \le i \le H-1$, $1 \le j \le W-1$.

c) Spectral Angle mapper (SAM) Loss: Hyperspectral SR must also preserve the spectral signature shape at each pixel. We therefore adopt a spectral angle mapper (SAM) loss:

$$\mathcal{L}_{\text{sam}} = \frac{1}{N}\sum_{b,i,j} \arccos\left(\frac{\langle \mathbf{y}_{b,i,j}, \hat{\mathbf{y}}_{b,i,j}\rangle}{\|\mathbf{y}_{b,i,j}\|_2\,\|\hat{\mathbf{y}}_{b,i,j}\|_2 + \varepsilon}\right), \tag{21}$$

where $\mathbf{y}_{b,i,j} \in \mathbb{R}^L$ and $\hat{\mathbf{y}}_{b,i,j} \in \mathbb{R}^L$ denote the ground-truth and reconstructed spectra at spatial location (i, j) in batch b, $N = BHW$ is the total number of pixels, and $\varepsilon > 0$ is a small constant for numerical stability.

III. Experimental Result and Discussion

A. Datasets

The experiments were conducted using two real SEHI datasets to evaluate the performance of our model: a complex metal alloy dataset and a silver paint dataset. The complex metal alloy dataset (palladium, silver, and carbon), abbreviated as Pd-Ag-C, was produced by the University of Liverpool. It consists of 60 sequential SEM image slices collected under low-energy conditions and is publicly available in the repository reported in [24]. The silver paint dataset comprises 49 sequential SEM image slices and was produced at the University of Sheffield by us using a DualBeam Focused Ion Beam–SEM system (Helios NanoLab 660/G). Secondary electron (SE) images were acquired at an accelerating voltage of 1 kV with a beam current of 2.5×10^{-11} A under high vacuum conditions ($\sim 10^{-7}$ mbar). Imaging was collected at a working distance of 3.92 mm, and the horizontal field of view of 30 μm, with an image resolution of 1536×1024 pixels. Images were acquired using a dwell time of 50 ns per pixel and averaged over 8 frames. During acquisition, the rotation and tilt angles were -0.0666 rad and -6.54×10^{-6} rad, respectively. The collection of SEHI data across different energy ranges was enabled by adjusting the mirror electrode voltage within a range of -15 to 15 V.

B. Implementation

The sequential SEM image slices are first registered using a template matching algorithm [25]. We then follow the experimental configuration described in [26], adopting a tile-wise training and testing strategy with a spatial tile size of

Algorithm 1 Training Procedure for SEHI SR

Require: Dataset $\mathcal{D} = \{(I_{LR}^i, I_{HR}^i)\}_{i=1}^N$; model $\mathcal{M}_\theta$; loss weights $(\lambda_{\text{mse}}, \lambda_{\text{spatial}}, \lambda_{\text{sam}})$; epochs E; batch size m; learning rate η_e.
Ensure: Trained parameters θ^*

```
for e = 1 to E do
    Update learning rate η_e
    for each mini-batch {(I_LR, I_HR)}_{j=1}^m from D do
        Î_SR ← M_θ(I_LR).
        Compute L_mse using (18)
        Compute L_spatial using (19) and (20)
        Compute L_sam using (24)
        Compute L using (17)
        θ ← Adam(θ, ∇_θ L, η_e)
    end for
end for
```

32 × 32 pixels. The resulting tiles are then converted to low-resolution counterparts via area interpolation at scales of 2×, 4×, and 8× (as shown in column 1 of Table I). Throughout the training, Adam is used as the optimizer, and the batch size was fixed at 4. The loss weights are set $(\lambda_{\text{mse}}, \lambda_{\text{spatial}}, \lambda_{\text{sam}}) = (1.0, 0.1, 0.1)$, giving priority to pixel fidelity while still constraining spatial edges and spectral transitions. To mitigate overfitting, early stopping is incorporated into the training procedure. Training is terminated when validation performance ceases to improve, rather than continuing for a predetermined number of epochs. The overall training procedure for SR SEHI is summarized in Algorithm 1.

C. Evalution Metrics

We quantify the quality of the reconstructed images with commonly adopted metrics, including the mean peak signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM), spectral angle mapper (SAM), and root mean squared error (RMSE) [27].

MPSNR measures the ratio of the peak signal power to the power of residual errors and is defined as

$$\text{MPSNR}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{10}{B} \sum_{i=1}^{B} \log_{10} \left(\frac{\max(X^i)}{\sqrt{\text{MSE}(X^i, \hat{X}^i)}} \right), \quad (22)$$

where B represents the spectral bands, $X^i \in \mathbb{R}^{H \times W}$ represents the i-th image of a reference SEHI image $\mathbf{X} \in \mathbb{R}^{H \times W \times B}$, $\hat{X}^i \in \mathbb{R}^{H \times W}$ represents the i-th image of a generated image $\hat{\mathbf{X}} \in \mathbb{R}^{H \times W \times B}$.

MSSIM measures the structural similarity between images using luminance, contrast, and structure qualities of images:

$$\text{MSSIM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{B} \sum_{i=1}^{B} \frac{(2\mu_{X^i}\mu_{\hat{X}^i} + C_1)(2\sigma_{X^i\hat{X}^i} + C_2)}{(\mu_{X^i}^2 + \mu_{\hat{X}^i}^2 + C_1)(\sigma_{X^i}^2 + \sigma_{\hat{X}^i}^2 + C_2)}, \quad (23)$$

where μ_{X^i} and $\mu_{\hat{X}^i}$ are the means of X^i and $\hat{X}^i$, $\sigma_{X^i}^2$ and $\sigma_{\hat{X}^i}^2$ are the variances of X^i and $\hat{X}^i$, $\sigma_{X^i\hat{X}^i}$ is the covariance between X^i and $\hat{X}^i$, and C_1 and C_2 are small fixed constants added for numerical stability.

SAM measures the angle between two spectral vectors in a high-dimensional space, effectively quantifying the spectral similarity between pixels. The SAM is determined by evaluating the cosine of the angular difference between two spectral signatures, which is expressed in radians and thus belongs to $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The smaller the absolute value of SAM, the less important the spectral distortion. The SAM index for the whole image is calculated as the average of pixel-wise SAM indices:

$$\text{SAM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{HW} \sum_{i=1}^{HW} \arccos \left(\frac{\mathbf{x}_i \cdot \hat{\mathbf{x}}_i}{\|\mathbf{x}_i\|_2 \, \|\hat{\mathbf{x}}_i\|_2} \right), \quad (24)$$

where $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i \in \mathbb{R}^{B \times 1}$ represent the spectra of the i-th pixels in the reference image $\mathbf{X}$ and the SR SEHI $\hat{\mathbf{X}}$, respectively, and $\|\cdot\|_2$ stands for the ℓ_2 norm.

D. Comparison with-State-of-the-Art

Firstly, we evaluated the proposed spatial spectral deep learning (SSDL) model on the silver paint and Pd-Ag-C datasets under 2×, 4×, and 8× scaling setups, as shown in Table I. Specifically, SSDL is compared with state-of-the-art models such as DSDCN [11], FGIN [14], DACN [12], and HDL [16]. Across all scales, SSDL consistently outperforms existing approaches due to its excellent feature extraction capability, particularly in terms of MPSNR, MSSIM, and SAM. At the 2× scale, SSDL achieves the best performance on Pd-Ag-C, with an MPSNR of 38.867 dB and a SAM of 1.751. On silver paint, it delivers the highest MPSNR of 44.579 dB and the second-lowest SAM of 1.441. At the 4× scale, SSDL further achieves the highest MPSNR scores of 41.757 dB on silver paint and 36.913 dB on Pd-Ag-C, along with the lowest and second-lowest SAM values of 2.163 and 2.0329, respectively. Similar improvements are observed at the more challenging 8× scale. SSDL delivers the lowest SAM across both datasets and the highest MPSNR on Pd-Ag-C, while its MPSNR on silver paint is only slightly lower than FGIN's. Combined with the qualitative comparisons in Fig. 4, these results highlight SSDL's robustness in spatial detail recovery and spectral preservation across different resolutions.

E. Ablation Study

Table II presents an ablation study conducted on the Pd-Ag-C dataset at the 2× scaling factor, highlighting the contributions of the spectral residual S-E module, the inception module, and the custom loss in the SSDL framework. Removing the spectral residual S-E block results in the weakest performance, with MPSNR of 36.500 dB, MSSIM of 0.888, and SAM of 2.143. This degradation demonstrates that this block is essential for enhancing both spatial fidelity and spectral consistency. Employing two inception blocks increases the MPSNR to 38.273 dB and reduces the SAM to 1.876, indicating enhanced structural and spectral representation capability. However, increasing the depth to four inception blocks instead degrades the performance, which can be attributed to increased

TABLE I

EVALUATION OF THE PROPOSED SSDL MODEL AGAINST STATE-OF-THE-ART METHODS ON THE (SILVER PAINT, PD-AG-C) DATASETS UNDER DIFFERENT SCALING FACTORS. ALL MODELS ARE TRAINED AND EVALUATED FROM SCRATCH. BEST PERFORMANCE IS SHOWN IN **BOLD** AND SECOND BEST IS UNDERLINED.

Scale Factor	Model	Silver Paint			Pd-Ag-C		
		MPSNR↑	MSSIM↑	SAM↓	MPSNR↑	MSSIM↑	SAM↓
2×	HDL [16]	42.502	0.945	1.894	35.024	0.883	2.469
	DSDCN [11]	39.870	0.930	2.408	35.494	0.882	2.264
	FGIN [14]	40.720	0.949	2.407	37.567	0.892	1.969
	DACN [12]	43.864	0.971	**1.239**	36.746	0.895	1.975
	SSDL (Ours)	**44.579**	**0.972**	1.441	**38.867**	**0.913**	**1.751**
4×	HDL [16]	40.481	0.927	2.226	35.243	0.874	2.225
	DSDCN [11]	37.956	0.904	2.89	35.306	0.870	2.214
	FGIN [14]	38.452	0.929	3.231	36.764	0.880	**1.999**
	DACN [12]	39.270	0.900	2.578	34.220	0.8725	2.194
	SSDL (Ours)	**41.757**	**0.957**	**2.163**	**36.913**	**0.883**	2.0329
8×	HDL [16]	33.32	0.759	4.245	31.554	0.800	2.584
	DSDCN [11]	32.748	0.704	4.45	30.288	0.758	2.417
	FGIN [14]	**34.994**	**0.820**	4.016	31.476	0.807	2.344
	DACN [12]	27.106	0.411	9.757	25.388	0.697	2.291
	SSDL (Ours)	34.527	0.811	**3.877**	**33.076**	**0.809**	**2.047**

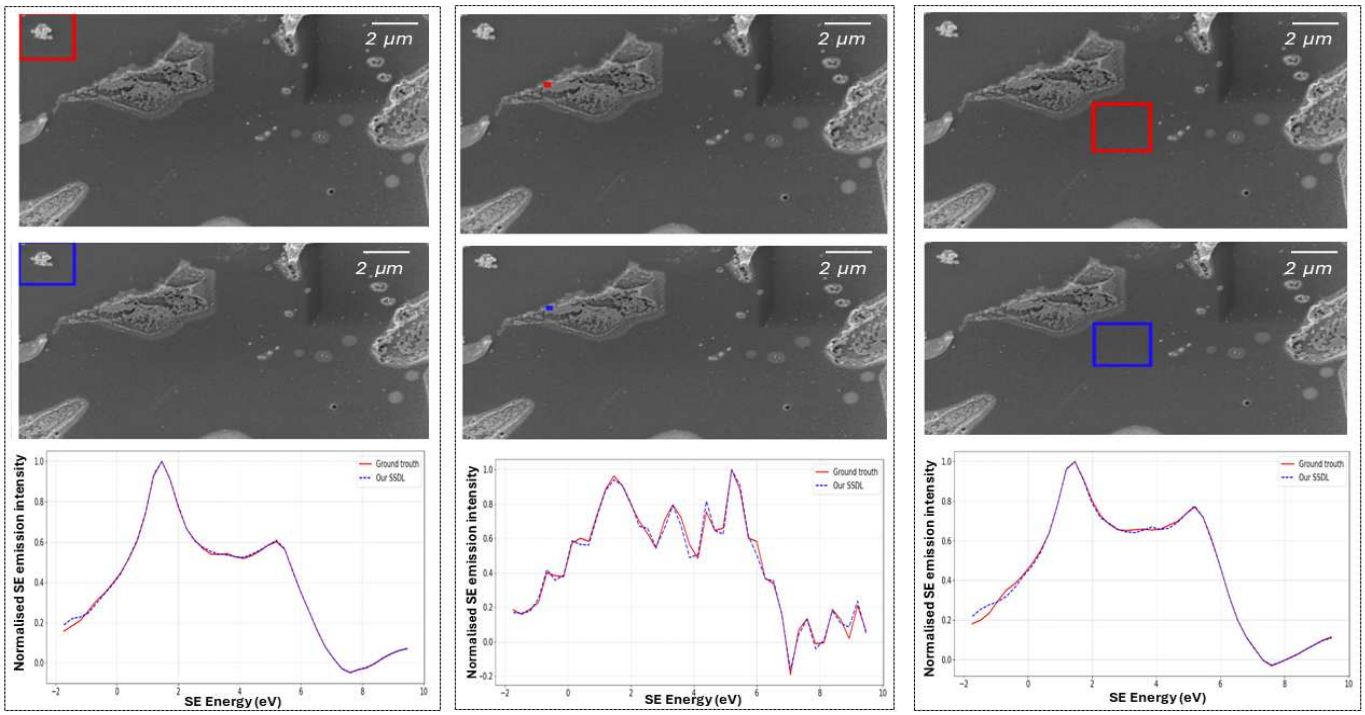


Fig. 4. Ground-truth and SSDL-based SR of a Pd–Ag–C thin film (top and middle row) with corresponding secondary-electron emission intensity as a function of secondary-electron energy (eV) (bottom row) for three FoV (left to right). Red/blue rectangles show the ROIs used for spectral extraction from ground-truth/SR SEH, demonstrating preservation of spectral shape and thus SEHI chemical properties after SR. Scale bars, 2 μm.

model complexity, introducing redundancy, and reducing generalization efficiency. Incorporating three inception blocks achieves the best trade-off between representational capacity and computational complexity. It attains the highest MPSNR of 38.867 dB, the best MSSIM of 0.913, and the lowest SAM of 1.751. When the SSDL is optimized only with the MSE criterion, disregarding the designed spatial–spectral loss, it achieves 38.307 dB MPSNR, 0.898 MSSIM, and 1.902 SAM. In contrast, the full SSDL model delivers the best overall performance, confirming that the joint integration is crucial for achieving accurate and spectrally faithful reconstructions.

F. Discussion

The results indicate that the proposed SSDL model is effective for SEHI image super-resolution, delivering competitive results on two real SEHI datasets across different scaling factors. Compared with existing methods, it demonstrates improved retention of structural features and spectral consistency, which is especially important in SEHI applications since spectral signatures are strongly associated with the chemical

TABLE II
QUANTITATIVE RESULTS MODEL

Ablation study on Pd-Ag-C (2×)			
Model Variant	**MPSNR↑**	**MSSIM↑**	**SAM↓**
SSDL w/o spectral residual S-E	36.500	0.888	2.143
SSDL w/s two inception blocks	38.273	0.900	1.876
SSDL w/s four inception blocks	37.290	0.890	1.993
SSDL w/o custum loss (just MSE)	38.307	0.898	1.902
SSDL (Ours)	**38.867**	**0.913**	**1.751**

characteristics of the material. The ablation study further validates the proposed architecture. In particular, the spectral residual S-E module, the three-block inception design, and the custom spatial-spectral loss all contribute to enhanced reconstruction quality. The findings indicate that integrating spectral attention, spectral-spatial fusion, and multiscale feature extraction is beneficial for this task. From a practical perspective, the method is relevant because it may help reduce the need for high-resolution SEHI scanning.

IV. CONCLUSION

In this work, we presented a novel SSDL framework that leverages deep learning to enable the first SEHI image SR. This framework integrates spectral S-E with the spectral spatial fusion module into the inception residual module architecture. By combining spectral S-E with spectral-spatial feature extraction, SSDL enhances both spatial resolution and spectral fidelity. Additionally, the use of a custom spatial-spectral loss, which jointly optimizes spatial and spectral reconstruction, enables the model to achieve reliable performance across various SEHI scales. Experimental results on two SEHI datasets confirm that the proposed approach delivers competitive results. The proposed solution has the potential to enhance the quality of SEHI for achieving super-resolution from low-resolution images. Future work will explore broader empirical validation using larger datasets and experimentally acquired low-resolution scans to reinforce practical utility.

Acknowledgment: We are grateful to the HCED (The Higher Committee for Education Development in Iraq) and the UK EPSRC SM3 SEE MORE MAKE MORE Grant: EP/V012126/1, who funded this research.

REFERENCES

[1] H. Schatten, "Low voltage high-resolution sem (lvhrsem) for biological structural and molecular analysis," *Micron*, vol. 42, no. 2, pp. 175–185, 2011.

[2] J. F. Nohl, N. T. Farr, M. R. Acocella, A. J. Knight, G. M. Hughes, J. Zhang, S. Robertson, S. Micklethwaite, S. Murphy, T. Motlová *et al.*, "Secondary electron hyperspectral imaging of carbons: new insights and good practice guide," *Advanced Science*, p. e01907, 2025.

[3] J. F. Nohl, N. T. Farr, Y. Sun, G. M. Hughes, S. A. Cussen, and C. Rodenburg, "Low-voltage sem of air-sensitive powders: From sample preparation to micro/nano analysis with secondary electron hyperspectral imaging," *Micron*, vol. 156, p. 103234, 2022.

[4] F. Zhou, W. Yang, and Q. Liao, "Interpolation-based image super-resolution using multisurface fitting," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3312–3318, 2012.

[5] W. Yao, J. Shen, Z. Guo, J. Sun, and B. Wu, "A total fractional-order variation model for image super-resolution and its sav algorithm," *Journal of Scientific Computing*, vol. 82, no. 3, p. 81, 2020.

[6] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[7] P. Yang, Y. Ma, X. Mei, Q. Chen, M. Wu, and J. Ma, "Deep blind super-resolution for hyperspectral images," *Pattern Recognition*, vol. 157, p. 110916, 2025.

[8] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 63–78.

[9] C. Wu, D. Wang, Y. Bai, H. Mao, Y. Li, and Q. Shen, "Hsr-diff: Hyperspectral image super-resolution via conditional diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 7083–7093.

[10] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3020–3033, 2019.

[11] U. Muhammad, J. Laaksonen, and L. Mihaylova, "Towards lightweight hyperspectral image super-resolution with depthwise separable dilated convolutional network," in *2025 IEEE Statistical Signal Processing Workshop (SSP)*, 2025, pp. 91–95.

[12] U. Muhammad and J. Laaksonen, "Dacn: Dual-attention convolutional network for hyperspectral image super-resolution," in *2025 33rd European Signal Processing Conference (EUSIPCO)*, 2025, pp. 2187–2191.

[13] H. Wang, Q. Zhang, T. Peng, Z. Xu, X. Cheng, Z. Xing, and T. Li, "Ssaformer: Spatial–spectral aggregation transformer for hyperspectral image super-resolution," *Remote Sensing*, vol. 16, no. 10, p. 1766, 2024.

[14] U. Muhammad and J. Laaksonen, "A fusion-guided inception network for hyperspectral image super-resolution," in *2025 25th International Conference on Digital Signal Processing (DSP)*, 2025, pp. 1–5.

[15] H. Wang, W. Zheng, Y. Wang, T. Yang, K. Zhang, and Y. Shang, "Single hyperspectral image super-resolution using a progressive upsampling deep prior network," *Electronic Research Archive*, vol. 32, no. 7, pp. 4517–4542, 2024.

[16] U. Muhammad and J. Laaksonen, "Hybrid deep learning for hyperspectral single image super-resolution," *IEEE Geoscience and Remote Sensing Letters*, 2025.

[17] Y. Wu, R. Cao, Y. Hu, J. Wang, and K. Li, "Combining global receptive field and spatial spectral information for single-image hyperspectral super-resolution," *Neurocomputing*, vol. 542, p. 126277, 2023.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[19] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[22] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391–407.

[23] J. Zhang, R. Zheng, Z. Wan, R. Geng, Y. Wang, Y. Yang, X. Zhang, and Y. Li, "Hyperspectral image super-resolution based on feature diversity extraction," *Remote Sensing*, vol. 16, no. 3, p. 436, 2024.

[24] J. Zhang, J. N. N. Farr, C. Rodenburg, K. Abrams, K. Black, and L. Mihaylova, "SEHI (Secondary Electron Hyperspectral Imaging) Dataset of Metal Alloy and Carbon Film (Palladium Silver Carbon Complex Film)," 2024, dataset. [Online]. Available: https://doi.org/10.15131/shef.data.22202923

[25] J. Nohl, "pysehi releases," 2023. [Online]. Available: https://doi.org/10.15131/shef.data.22310068.v1

[26] Z. Wang, D. Li, M. Zhang, H. Luo, and M. Gong, "Enhancing hyperspectral images via diffusion model and group-autoencoder super-resolution network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 6, 2024, pp. 5794–5804.

[27] S. Luo, Y. Qian, L. Bai, Y. Fan, Y. Wang, and W. Kong, "Deep learning-based hyperspectral and multispectral fusion techniques: Review, optimization, and perspectives," *Information Fusion*, p. 103291, 2025.