Discover Nano

RESEARCH Open Access

Sparse coding-based multiframe superresolution for efficient synchrotron radiation microspectroscopy

Yasuhiko Igarashi[1,2*†], Naoka Nagamura[2,3,4*†], Masahiro Sekine[1], Hirokazu Fukidome[4], Hideitsu Hino[5] and Masato Okada[2,6]

†Yasuhiko Igarashi and Naoka Nagamura have contributed equally to this work.

*Correspondence:
Yasuhiko Igarashi
igayasu1219@cs.tsukuba.ac.jp
Naoka Nagamura
NAGAMURA.Naoka@nims.go.jp
[1]Institute of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 3058573, Japan
[2]Photoemission Group, National Institute for Materials Science, 3-13 Sakura, Tsukuba, Ibaraki 3050003, Japan
[3]Graduate School of Advanced Engineering, Tokyo University of Science, 6-3-1 Niijuku, Katsushika, Tokyo 1258585, Japan
[4]Research Institute of Electrical Communication, Tohoku University, 2-1-1 Katahira, Aobaku, Sendai, Miyagi 9808577, Japan
[5]Department of Advanced Data Science, Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 1908562, Japan
[6]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba 2778561, Japan

Abstract

In nanostructure extraction, advanced techniques like synchrotron radiation and electron microscopy are often hindered by radiation damage and charging artifacts from long exposure times. This study presents a multiframe superresolution method using sparse coding to enhance synchrotron radiation microspectroscopy images. By reconstructing high-resolution images from multiple low-resolution ones, exposure time is minimized, reducing radiation effects, thermal drift, and sample degradation while preserving spatial resolution. Unlike deep learning-based superresolution methods, which overlook positional misalignment, our approach treats positional shifts as known control parameters, enhancing superresolution accuracy with a small, noisy dataset. Additionally, our sparse coding method learns an optimal dictionary tailored for nanostructure extraction, fine-tuning the SR process to the unique characteristics of the data, even with noise and limited samples. Applied to 3D nanoscale electron spectroscopy for chemical analysis (nano-ESCA) data, our method, utilizing a high-resolution dictionary learned from 3D nano-ESCA datasets, significantly improves image quality, preserving structural details. Unlike state-of-the-art deep learning techniques that require large datasets, our method excels with limited data, making it ideal for real-world scenarios with constrained sample sizes. High-resolution quality can be maintained while reducing the measurement time by over 40%, highlighting the efficiency of our approach. The results underscore the potential of this superresolution technique to not only advance synchrotron radiation microspectroscopy but also to be adapted for other high-resolution imaging modalities, such as electron microscopy. This approach offers enhanced image quality, reduced exposure times, and improved interpretability of scientific data, making it a versatile tool for overcoming the challenges associated with radiation damage and sample degradation in nanoscale imaging.

Keywords Sparse coding superresolution, Nanostructure image enhancement, Measurement image analysis, Synchrotron image reconstruction, High-resolution microscopy, Radiation damage reduction

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

1 Introduction

To analyze and extract nanostructures, advanced measurement techniques using synchrotron radiation facilities or electron microscopy are commonly employed. These methods are crucial for studying low-dimensional materials, such as atomic layer materials and nanowires, which have applications in fields ranging from semiconductor devices to batteries and catalysts [8, 29, 34, 38, 41]. To fully understand the geometric and electronic properties of these materials, it is essential to use measurement techniques that provide high resolution across a broad range from nanometres to micrometres.

Photoelectron microscopy is a powerful tool used to study the local electronic states and bonding conditions of materials, particularly for low-dimensional systems [13, 36]. This method can be broadly divided into two types: full-field and scanning photoelectron microscopes. Full-field microscopes directly project the image onto a detector using lenses, enabling rapid data acquisition. Scanning microscopes, on the other hand, employ a focused beam to scan across the sample, producing detailed spatial distributions of spectra. While scanning techniques can provide highly accurate spectral data, acquiring high-resolution (HR) images requires scanning in fine steps, which can be a time-consuming process.

In this study, we focus on the 3D nanoscale electron spectroscopy for chemical analysis (nano-ESCA) system, a versatile scanning photoelectron microscope used in X-ray photoelectron spectroscopy (XPS) imaging [17] (Fig. 1). This system uses high-brightness synchrotron X-rays to scan samples and acquire detailed spatial distributions of XPS core-level spectra. Its ability to capture high-resolution and nanoscale images make it ideal for evaluating the performance of our proposed superresolution (SR) method.

While high-brightness synchrotron X-rays have reduced the time needed to scan each point [5, 7, 16, 17, 23, 25, 42], they also introduce challenges such as image distortion due to charging, thermal drift, and sample degradation from radiation damage, as shown in the bottom right corner of Fig. 1. These issues are common in scanning microscopes like Raman and scanning electron microscopes. Thus, there is a need for methods that can achieve high resolution and wide coverage while minimizing radiation exposure.

Superresolution (SR) technology offers a promising solution to these challenges by generating high-resolution images from multiple low-resolution (LR) measurements, as shown in the bottom right corner of Fig. 1. This method reduces the need for prolonged

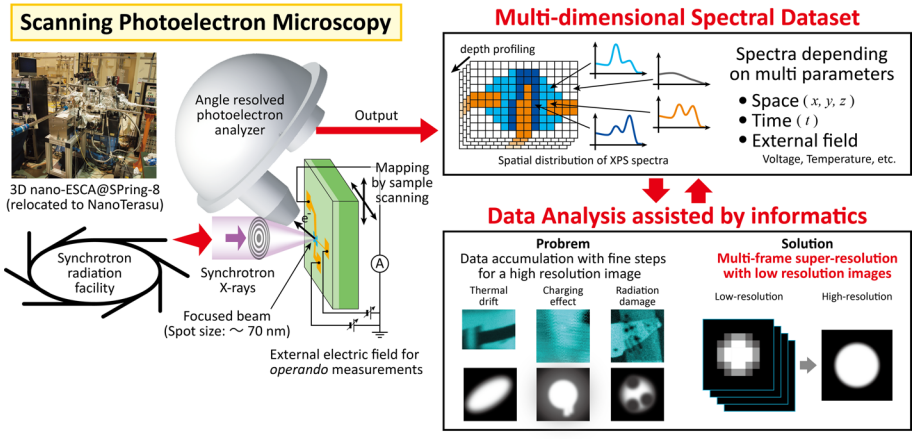


Fig. 1 Research concept: Utilizing machine learning technology to increase the analysis efficiency of analysing inspection data obtained from synchrotron radiation scanning photoelectron microscopy systems

exposure, thereby mitigating the effects of radiation damage, thermal drift, and sample degradation, all while maintaining the necessary spatial resolution and coverage. SR can be achieved using various techniques, with the simplest being geometrical interpolation, such as the bicubic method, which relies on pixel value continuity to upscale images. Sparse coding methods [11, 12, 43] represent a more advanced approach, taking advantage of the redundancy in natural image structures. In sparse coding, small image patches are represented as a linear combination of a few basis elements, making it a highly efficient and interpretable technique, especially when dealing with limited datasets. Sparse coding also allows for the incorporation of known control parameters, such as positional shifts, which is crucial for achieving high-precision SR.

In recent years, deep learning-based SR methods [10, 21] have gained popularity due to their ability to automatically extract complex features from large datasets through multiple layers of neural networks. These methods have shown great success in fields such as remote sensing [9, 37], where large amounts of data are available. However, deep learning approaches have notable limitations in scientific imaging tasks, where data may be sparse and noisy. They often struggle to incorporate specific known parameters, like positional shifts, and their black-box nature limits interpretability, making them less suitable for scientific applications where understanding the underlying processes is critical.

In this study, we propose a multiframe SR approach for synchrotron radiation microspectroscopy images using sparse coding. By reconstructing images from multiple LRIs, our method reduces exposure time while enhancing image quality. This approach addresses key challenges by explicitly accounting for positional shifts as known control parameters, enabling high-precision superresolution even with small and noisy datasets. Additionally, our sparse coding method learns an optimal dictionary tailored for nanostructure extraction, ensuring accurate reconstructions under challenging conditions. Comparative analyses demonstrate that our method outperforms both bicubic interpolation and deep learning-based approaches, achieving higher peak signal-to-noise ratio (PSNR) values and more consistent results. Furthermore, our method reduces measurement time by over 40% and has been validated on newly acquired datasets, highlighting its potential for high-resolution imaging in demanding experimental scenarios.

2 Method

We first consider a sparse coding SR approach. The proposed algorithm, which is mainly based on [19, 20], consists of dictionary learning and sparse coding for SR.

2.1 Dictionary learning

To extract features contained in HRIs of the measured targets, the dictionary is obtained by learning from sufficient training data. For the HRIs, we utilized a dataset of HRIs measured via 3D nano-ESCA, as explained in the next section. For these HRIs, noise removal was performed via a DnCNN [45], and the denoised images were then divided into N patches for dictionary learning. DnCNN, which stands for Denoising Convolutional Neural Network, is a deep learning-based model designed for image denoising tasks. It employs convolutional layers to learn and remove noise from corrupted images, outperforming traditional methods.

The HR dictionary $\mathrm{D}^{\mathrm{h}} \in \mathbb{R}^{q_h \times K_{\mathrm{atom}}}$ is learned by applying the K-means singular value decomposition (K-SVD) algorithm [1] to the component patches of the denoised HRIs. The K-SVD algorithm is an iterative method used for sparse representation and dictionary learning. It extends the k-means clustering approach by applying singular value decomposition (SVD) to update both the dictionary and sparse representations, optimizing the representation of data with minimal reconstruction error. For the obtained patch set with a $\sqrt{q_h}$ patch length, the K-SVD algorithm is used to solve the following optimization problem to learn an HR dictionary $\mathrm{D}^{\mathrm{h}} \in \mathbb{R}^{q_h \times K_{\mathrm{atom}}}$:

$$\boldsymbol{D}^{\mathrm{h}} = \underset{\boldsymbol{D},\alpha}{\operatorname{argmin}} \|\boldsymbol{P} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 \quad \text{subject to } \forall i, \|\boldsymbol{\alpha}_i\|_0 \leq k_0, \tag{1}$$

where $\boldsymbol{P} \in \mathbb{R}^{q_h \times K}$ is a matrix in which each patch is a column element, $\alpha \in \mathbb{R}^{K_{\mathrm{atom}} \times K}$ is a matrix with sparse code α_i as a column element corresponding to the i-th patch, and k_0 is the maximum number of nonzero elements. The initial dictionary is based on discrete cosine transform (DCT) bases [2] and in the learning process, α is found via the orthogonal matching pursuit (OMP) algorithm. Then, α is fixed, and dictionary $\boldsymbol{D}$ is updated via the K-SVD algorithm.

2.2 Superresolution method based on sparse coding

Sparse coding is a method of reconstructing signals using a dictionary of bases learned from data, which enables highly accurate signal approximation. By sparsely representing the learned bases, this approach can effectively remove noise and artefacts via denoising, deburring and SR [11]. This is because the sparse representation captures the most essential features of the data, reducing redundancy and preserving important structural details.

We first explain the concept of single-frame SR via sparse coding and then extend this approach to multiframe SR. Let $\mathrm{Y} \in \mathbb{R}^{Q_l}$ represent the observed LRI. The goal of SR is to reconstruct an HRI $\mathrm{X} \in \mathbb{R}^{Q_h}$ from Y, where $Q_l < Q_h$. To reduce the computational cost, the SR approach is typically applied to small image regions, known as patches, which are then combined to form the full image. Accordingly, we focus on reconstructing an HRI patch $\mathrm{x} \in \mathbb{R}^{q_h}$ from a corresponding LR patch $\mathrm{y} \in \mathbb{R}^{q_l}$, where $q_l < q_h$. For each patch, we assume the following degradation process: the HR and LR patches, x and y, are related by the observation model $\mathrm{y} = G\mathrm{x} + \boldsymbol{\varepsilon}$, where G is a degradation operator that includes blurring and downsampling operations and $\boldsymbol{\varepsilon} \in \mathbb{R}^{q_l}$ represents additive noise.

We next consider the degradation operator G. G is generated by $G = SH$, where H represents the blurring operator and S represents the downsampling operator. We define the blurring operation H as a convolution of the image with an $n_{\mathrm{blur}} \times n_{\mathrm{blur}}$-pixel Gaussian filter with a standard deviation σ_h. The matrix $\hat{H} \in \mathbb{R}^{n_{\mathrm{blur}} \times n_{\mathrm{blur}}}$ has elements defined as: $[\hat{H}]_{ij} = \dfrac{\exp\left(-\frac{|i-x_0|^2}{2\sigma_h^2}\right)\exp\left(-\frac{|j-y_0|^2}{2\sigma_h^2}\right)}{\sum_{i,j=1}^{n_{\mathrm{blur}}}\exp\left(-\frac{|i-x_0|^2}{2\sigma_h^2}\right)\exp\left(-\frac{|j-y_0|^2}{2\sigma_h^2}\right)}$, where i and j are the pixel indices, and the center coordinates are given by $x_0 = \frac{n_{\mathrm{blur}}+1}{2}$ and $y_0 = \frac{n_{\mathrm{blur}}+1}{2}$. Next, we define a linear downsampling operator S that reduces the resolution by a factor of s_r. The downsampling operator S retains only every s_r-th pixel in both the vertical and horizontal directions. In summary, the blurring operator H is first applied to the HRI,

introducing the required blur. The resulting blurred image is then downsampled by the operator S, yielding an LRI basis G.

In the sparse coding framework [11, 22, 30], the observed signals are expressed as a linear combination of a few basis vectors, called atoms, selected from a larger set. Let $\mathrm{D}^l = [\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_{K_{\mathrm{atom}}}] \in \mathbb{R}^{q_l \times K_{\mathrm{atom}}}$ denote a dictionary consisting of K_{atom} atoms $\mathrm{d}_k (\ k = 1, \ldots, K_{\mathrm{atom}})$, and let $\boldsymbol{\alpha} \in \mathbb{R}^{K_{\mathrm{atom}}}$ be the coefficient vector for the sparse representation of the patch $\mathrm{y} \in \mathbb{R}^{q_l}$. The sparse coding problem can then be formulated as:

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \|\mathrm{y} - \mathrm{D}^l \boldsymbol{\alpha}\|_2^2 + \eta \|\boldsymbol{\alpha}\|_1, \tag{2}$$

where $\|\boldsymbol{\alpha}\|_1 = \sum_{j=1}^{K} |\alpha_j|$ is the ℓ^1-norm and $\eta > 0$ is a regularization parameter. With this formulation, the coefficient vector $\boldsymbol{\alpha}$ is sparse, meaning that only a few elements in $\boldsymbol{\alpha}$ are nonzero, allowing y to be well approximated by a small number of atoms. The problem of obtaining $\boldsymbol{\alpha}$ for a fixed dictionary D^l is referred to as sparse coding.

We assume that the HR patch x can be represented as a sparse combination of HR atoms: $\mathrm{x} = \mathrm{D}^h \boldsymbol{\alpha}$. Given this relationship, the LR patch y is approximately related to the HR patch x as follows:

$$\mathrm{y} \simeq G\mathrm{x} = G\mathrm{D}^h \boldsymbol{\alpha} = \mathrm{D}^l \boldsymbol{\alpha}, \tag{3}$$

where D^l is the LR dictionary generated from D^h by the degradation process. Each atom in D^l corresponds to an HR atom in D^h. This leads to the following two-step procedure for single-frame SR: (1) represent the LR patch as a sparse combination of LR atoms, and (2) use the combination coefficients to reconstruct the corresponding HR patch by combining HR atoms, as illustrated in Fig. 2. Before performing single-frame SR, an HR dictionary must be prepared by applying an appropriate dictionary learning algorithm to HR training images.

Next, we consider multiframe SR with a sparse coding approach as shown in Fig. 2. Suppose that N LRIs $\mathrm{Y}_i \in \mathbb{R}^{Q_l}, i = 1, \ldots, N$ are observed, each being a differently

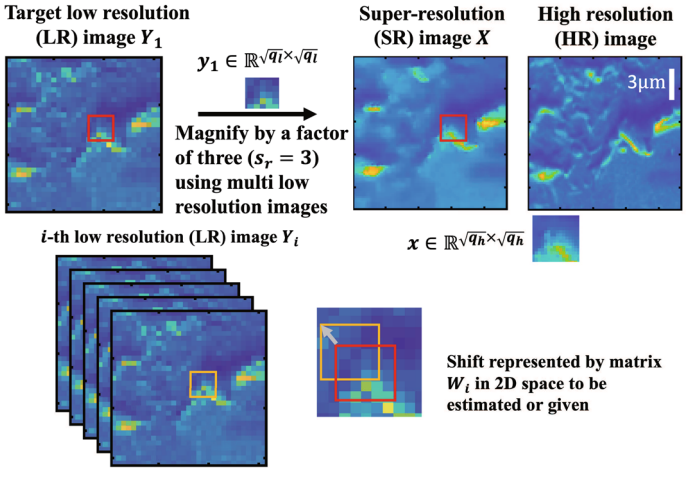


Fig. 2 Overview of the multiframe superresolution (SR) analysis method. This figure illustrates the process of generating an SR image from multiple low-resolution (LR) images (LRIs). The target LRI Y_1 is magnified by a factor of three using multiple LRIs Y_i. The shift in each LRI, represented by the matrix W_i in 2D space, is estimated or given to accurately align the images. The resulting SR image X is compared with an HR reference image to validate the resolution enhancement achieved with the proposed approach

degraded version of an HRI $\mathrm{X} \in \mathbb{R}^{Q_h}$. Without loss of generality, assume that Y_1 is the target for SR, with the other $N-1$ LRIs serving as auxiliary images.

We assume that the degradation process for each patch can be modelled as follows: for the target patch y_1, the degradation process is given by Eq. (1). Moreover, for the auxiliary patches y_i($i = 2, \ldots, N$), we assume that $\mathrm{y}_i = GW_i\mathrm{x} + \boldsymbol{\varepsilon}$, where W_i represents the shift and clipping operator corresponding to the *i*-th LR patch. Once W_i is estimated, SR based on sparse coding can be performed by stacking the LR dictionaries corresponding to the observations $\{\mathrm{y}_i\}_{i=1}^{N}$ to form a stacked LR dictionary:

$$\tilde{\mathrm{D}}^l = \begin{bmatrix} \mathrm{D}_1^l \\ \mathrm{D}_2^l \\ \vdots \\ \mathrm{D}_N^l \end{bmatrix} = \begin{bmatrix} G\mathrm{D}^h \\ GW_2\mathrm{D}^h \\ \vdots \\ GW_N\mathrm{D}^h \end{bmatrix}. \tag{4}$$

This dictionary is then used to approximate the stacked LR patch $\tilde{\mathrm{y}} = [\mathrm{y}_1 \quad \mathrm{y}_2 \quad \cdots \quad \mathrm{y}_N]^\top$ via sparse coding. The HR estimate $\hat{\mathrm{x}}$ is given by:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\tilde{\mathrm{y}} - \tilde{\mathrm{D}}^l\boldsymbol{\alpha}\|_2^2 + \eta\|\boldsymbol{\alpha}\|_1, \quad \hat{\mathrm{x}} = \mathrm{D}^h\hat{\boldsymbol{\alpha}}. \tag{5}$$

Sparse coding does not demand exact equality, $\mathrm{D}\boldsymbol{\alpha} = \mathrm{x}$. Because of this and because of noise and patchwise processing, backprojection [18] is performed to maintain global consistency [19, 40, 43]. In our approach, we perform backprojection on the basis of the consistency between the reconstructed HRI and the SR image by interpolating the LRIs via bicubic interpolation. The synchrotron radiation images used in this study exhibit significant variability in signal intensity across different low-resolution images, with some images showing high intensity in regions where others show the opposite. This level of noise results in instability and inconsistency, which can cause the reconstructed images to be noisy when using the consistency with the target LRIs as a correction factor. To address this issue, we propose a novel approach leveraging the known shift in multiple LRIs. Specifically, we first obtain an SR image by interpolating the LRIs via bicubic interpolation. Then, we shift these SR images according to the known shifts and average them to perform backprojection. This approach, which we define as bicubic stacking, is unique to this study and serves as a bicubic interpolation-based image stacking method to enhance the resolution of the LRIs.

Let Z be the SR image obtained through bicubic stacking. To maintain global consistency, we perform backprojection by minimizing the discrepancy between the initial SR image X_0 obtained from the SR algorithm and the refined SR image Z. The minimization problem can be formulated as:

$$\underset{X}{\text{minimize}} \, \|\mathrm{X} - \mathrm{X}_0\|_2^2 + c\,\|\mathrm{X} - \mathrm{Z}\|_2^2, \tag{6}$$

where X_0 is the initial estimate of the HRI obtained from the SR analysis and Z is the revised SR image. Here, $c > 0$ is a trade-off parameter. The gradient descent update for the HRI X is given by: $\mathrm{X}_{t+1} = \mathrm{X}_t - \nu\,[(\mathrm{X}_t - \mathrm{X}_0) + c\,(\mathrm{X}_t - \mathrm{Z})]$, where $\nu > 0$ is a step size parameter. Through this iterative process, the SR image X is refined, ensuring that it maintains global consistency with the target LRI while revising the initial estimate X_0 to align more closely with the refined image Z.

As discussed in the previous paragraph, the major problem of multiframe SR can be reduced to the challenge of accurately estimating the shift and clipping operators $W_i, i = 2, \ldots, N$. In both sparse coding and deep learning-based multiframe SR techniques, the estimation accuracy of positional shifts between frames considerably impacts measurement precision [14, 24, 39].

In this study, we assume that the shifts between pixels are known a priori and explore the impact of this assumption on the performance of multiframe SR methods. We compare the performance of our multiframe SR method based on sparse coding with two representative state-of-the-art approaches. The first is a method based on sparse coding with shift estimation [20], which is widely used for its interpretability and efficiency in scenarios with limited datasets and known control parameters like positional shifts. The second is a deep learning approach that incorporates shift estimation into its framework [37]. This method represents the cutting edge of data-driven techniques, leveraging large datasets to automatically extract complex features and achieve high-resolution results. These methods were chosen because they reflect the two dominant approaches to superresolution in the current field: sparse coding, which excels in interpretability and data-limited cases, and deep learning, which performs well when large amounts of data are available. The comparison allows us to evaluate the effectiveness and limitations of each method in achieving HR results. The development of an interpretable HR method, such as the sparse coding approach, is important, and such an approach is anticipated to obtain results that are not only easily understood but also capable of achieving high-accuracy SR. This could further contribute to the derivation of scientific knowledge, particularly when working with relatively small datasets.

3 Data acquisition and experimental setup

3.1 Training and evaluation datasets for the multiframe superresolution analysis method

In this study, we utilized 118 HR 3D nano-ESCA datasets, measured between 2011 and 2022, to train dictionary learning models and evaluate the performance of the proposed multiframe SR analysis method. These datasets were obtained using the 3D nano-ESCA system, which incorporates an X-ray focusing system equipped with a Fresnel zone plate and a modified VG Scienta R3000 analyser. The system maintains a high spatial resolution of approximately 100 nm in the in-plane direction, enables depth analysis through angle-resolved detection of photoelectrons, and allows *operando* measurements by applying an electric field to the sample. This setup has been extensively used for analysing various semiconductor device microstructures and energy nanomaterials [3, 26–28, 35]. The HR 3D nano-ESCA datasets cover various materials and technologies, mainly semiconductor devices, graphene, and battery materials.

The 3D nano-ESCA system was installed at the University of Tokyo outstation beamline, BL07LSU at SPring-8, and relocated to NanoTerasu in 2023. All the results presented here were obtained with excitation photon energies ranging from 800 eV to 1500 eV. The step sizes between observed points, mostly ranging from $0.05\mu m$ to $0.25\mu m$, mainly include small intervals, with $0.1\mu m$ being the most commonly used step size. These step sizes were applied to ensure evenly spaced measurements across each dataset.

The two-axis piezo scanner is used for precise sample scanning along the x and y directions during SPEM measurements, offering a maximum scan range of $100\mu m \times 100\mu m$

with an accuracy of 2*nm* [17]. Furthermore, the system maintains a high spatial resolution of approximately 100*nm* in the in-plane direction and enables depth analysis through angle-resolved detection of photoelectrons. Therefore, the precision of the piezo scanner is sufficiently high compared to the assumed resolution, and the shift values in the low-resolution images can be regarded as known a priori.

The HR 3D nano-ESCA datasets include images of varying sizes, represented by their height and width. The sizes have wide ranges, including small images of 21×21 pixels to much larger images of 301×301 pixels. Common image dimensions include 101×101, 81×81, and 61×61 pixels, reflecting a preference for square images with smaller dimensions. The variety in image sizes suggests that the datasets cover a range of spatial resolutions and possibly different scales of the observed structures. Additionally, the dataset includes measurements of the horizontal (X) and vertical (Z) sizes in micrometres (μm). These sizes vary widely, with common dimensions of approximately $10 \times 10\,\mu m$ and $20 \times 20\,\mu m$. This range indicates that the analysis spans multiple scales and is suitable for examining various features and structures within the samples.

3.2 Experimental setup for microstructured gold patterns and fine surface irregularities in GaN films for the evaluation of the proposed method

In addition to the HR 3D nano-ESCA datasets, we obtained new data by acquiring LRIs and HRIs of gold (Au) and GaN films by measuring peak intensities in the core-level spectra of Au 4*f* and C 1*s*, respectively. Low resolution images were obtained using measurements at 500 nm intervals, whereas high resolution images were acquired using measurements at 200 nm intervals. The effectiveness of the proposed method was demonstrated by applying it to these new datasets.

Specifically, we conducted measurements of finely processed Au patterns on a $\mathrm{SiO_2/Si}$ substrate for evaluation. Au patterns were fabricated via vacuum deposition after applying standard electron-beam lithography and lift-off techniques. Spectromicroscopic measurements were performed via the 3D nano-ESCA system, which maintains a high spatial resolution of approximately 100 nm.

The experimental conditions for analysing the microstructured Au patterns involved obtaining spectral mapping images within a $15\,\mu m \times 15\,\mu m$ area. For the low resolution images, measurements were performed in 500 nm steps, while for the high resolution images, measurements were conducted in 200 nm steps. Similar measurements of a GaN film with fine surface irregularities were obtained. The SR analysis method was then applied to multiple LRIs to estimate the HR results. Detailed information about the experimental setup and data analysis can be found in the Supporting Information.

For the evaluation using fine gold patterns as reference materials and GaN films, low-resolution images (LRIs) were acquired via synchrotron radiation measurements by obtaining spectral mapping images with a 500 nm scanning step. To enable super-resolution analysis, high-resolution images were subsequently obtained with a 200 nm scanning step. However, since performing super-resolution on the 500 nm LRIs with a $s_r = 3$ upscaling results in a different pixel count compared to that from the 200 nm high-resolution images, the high-resolution images were processed using bicubic interpolation to match the pixel count of the $s_r = 3$ upscaled LRIs. This procedure ensured that the accuracy of the super-resolution analysis could be fairly compared.

3.3 Data preprocessing

We processed 118 HR 3D nano-ESCA datasets to extract 90×90-pixel HRIs and the corresponding 30×30-pixel LRIs. Each HRI was paired with nine 30×30 LR patches ($N = 9$), resulting in a total of 146 datasets. The extraction process involved normalizing the images and dividing them into patches of specific sizes, with a stride setting of half the patch size to ensure overlapping regions. The generated patches were filtered on the basis of minimum size requirements, and both the HRIs and LRIs were normalized for consistency. The images were then stored for further analysis, ensuring that only reliable pixels were considered, particularly given the variations in weather conditions across different scenes.

To generate LRIs from HR data, we employed a strategic sampling approach. Starting from a defined initial point, such as ($55.0\mu m$, $44.6\mu m$) in a two-dimensional plane, we sampled the HR data at intervals corresponding to every third pixel to downsample the images by a factor of three. This method effectively reduces the spatial resolution while maintaining representative coverage across the image. Finally, each image was normalized to range between 0 and 1 by subtracting the minimum pixel intensity value from each pixel and then dividing by the range. By shifting the starting point for each sample, we prevented overlap between the resulting LRIs, ensuring that nine distinct LRIs were generated from each HRI. Each of these LRIs corresponds to a different region in the original HRI, allowing us to construct a comprehensive dataset that captures variations in the original HR data at a lower resolution.

In addition to creating these datasets, we prepared data for dictionary learning. Specifically, we extracted 54×54 pixel patches from the 118 HR 3D nano-ESCA datasets. To ensure that there was no overlap between patches, the extraction process was carefully controlled. From this process, we generated $K = 250$ patch datasets, which were then used to create a dictionary of 6×6 pixel elements, with $q_h = 6 \times 6 = 36$. This dictionary is intended to support the dictionary learning models used in the multiframe SR analysis method.

4 Results and discussion

In this section, we present the validation of our multiframe superresolution (SR) method based on sparse coding, applied to high-resolution 3D nano-ESCA datasets. First, we compare the performance of dictionaries learned from natural images and 3D nano-ESCA data, demonstrating the importance of using a tailored dictionary for accurate image reconstruction. Next, we apply the SR method to low-resolution images, showing that it outperforms traditional bicubic interpolation and deep learning approaches in terms of PSNR and SSIM. Finally, we test the method on newly acquired datasets of microstructured gold patterns and GaN films, confirming its ability to preserve fine structural details and validating its potential for use in challenging experimental conditions.

4.1 Dictionary for high-resolution 3D nano-ESCA datasets

In this section, we aim to evaluate the effectiveness of dictionary learning for 3D nano-ESCA datasets by comparing dictionaries learned from natural images and those learned from 3D nano-ESCA data. The goal is to determine which type of dictionary is better

suited for accurate image reconstruction, particularly in the context of high-resolution 3D nano-ESCA images.

We applied the dictionary learning method using the OMP algorithm to both datasets. Fig. 3 illustrates the dictionaries learned via the OMP algorithm. The dictionary in Fig. 3a was obtained from natural images, consisting of 512 atoms, each of size 15×15 pixels. In contrast, Fig. 3b shows the dictionary $\mathrm{D}^{\mathrm{h}} \in \mathbb{R}^{q_h \times K_{\mathrm{atom}}}$ learned from the HR 3D nano-ESCA dataset. For this 3D nano-ESCA dictionary, a sparsity value of $k_0 = 5$ was used, with $K_{\mathrm{atom}} = 225$ atoms of size $q_h = 6 \times 6$ pixels, and the learning process was iterated 500 times.

The comparison between the two dictionaries reveals key insights into their suitability for different types of image data. Although both the 3D nano-ESCA and natural image dictionaries exhibit similar patterns, the smaller size of the 3D nano-ESCA images poses challenges in accurate reconstruction, particularly when bases with pixel sizes of 15×15 or 18×18 are used. When a dictionary is learned from 18×18 pixel patches of 3D nano-ESCA images, the reconstruction error is more than 70% higher than that when a dictionary is learned from 6×6 pixel patches. This suggests that dictionaries trained based on natural images are not well suited for reconstructing 3D nano-ESCA images, as the fine details inherent in these smaller-scale images cannot be captured when larger pixel bases are used for training.

4.2 Multiframe superresolution method for high-resolution 3D nano-ESCA datasets

In this section, we aim to validate the effectiveness of our proposed multiframe superresolution (SR) method based on sparse coding, applied to high-resolution (HR) 3D nano-ESCA datasets. The goal is to assess how well the method can reconstruct high-resolution images from low-resolution inputs, considering critical parameters such as positional shifts and image degradation. By applying the method to a series of 146 datasets created in Sect. 3.3 we quantitatively evaluate the performance of the SR process using metrics such as PSNR and SSIM. We, then, aim to demonstrate that high-quality superresolution images can be achieved with fewer low-resolution inputs, reducing measurement time by over 40% without sacrificing image quality.

We generate SR images with a scaling factor of $s_r = 3$ from $N = 9$ LRIs using the sparse coding approach. We first consider the degradation operator $G = SH$. In the

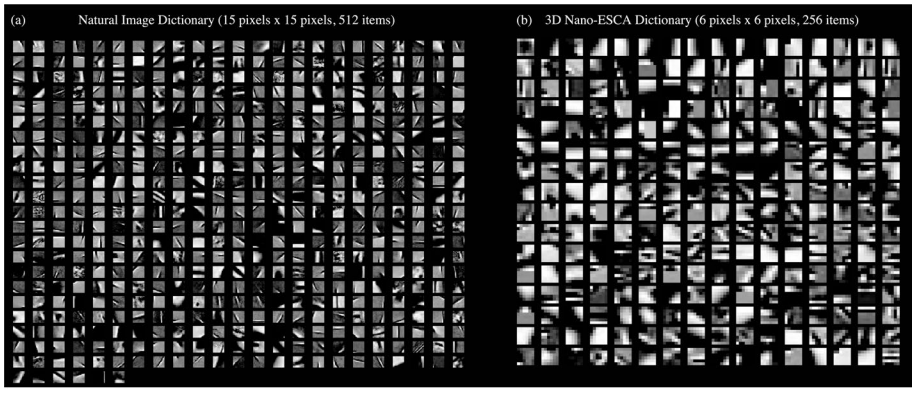


Fig. 3 Improved reconstruction with a 3D Nano-ESCA dictionary compared to a natural image dictionary **a** shows a dictionary obtained from natural images with 512 atoms, each of size 15×15 pixels. **b** shows the dictionary $\mathrm{D}^{\mathrm{h}} \in \mathbb{R}^{q_h \times K_{\mathrm{atom}}}$ learned from high-resolution 3D nano-ESCA data, with $K_{\mathrm{atom}} = 225$ atoms of size $q_h = 6 \times 6$ pixels

blurring operation H, we set $n_{\mathrm{blur}} = 8$ and an $n_{\mathrm{blur}} \times n_{\mathrm{blur}}$-pixel Gaussian filter with a standard deviation of $\sigma_h = 1.2$. The blurred and shifted images are then downsampled by a factor of $s_r = 3$. We also set the regularization parameter $\eta = 0.05$ in the sparse coding approach defined in Eq. (2). The detailed impacts of the parameters η and σ_h on the multiframe SR results are provided in the Supporting Information. For backprojection, we choose the trade-off parameter $c = 1 \times 10^{-1} \times 9$, which controls the influence of the regularization term in the optimization process. The step size parameter ν, which determines the gradient descent step, is set to $\nu = 1 \times 10^{-4} \times 1$. These settings were selected to ensure stable and efficient convergence during the backprojection process.

To quantitatively compare the performance of the SR methods, we use the peak signal-to-noise ratio (PSNR), which is defined as $\mathrm{PSNR[dB]} = 10 \log_{10} \frac{1^2}{\mathrm{MSE}}$, where MSE is the mean squared error between the original HRI and the estimated HRI. A higher PSNR value indicates better SR performance. In addition, we use the structural similarity index measure (SSIM), which is defined as $\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$, where x and y are the original HRI and the estimated HRI, respectively. μ_x and μ_y are the mean pixel values, σ_x^2 and σ_y^2 are the variances, and σ_{xy} is the covariance of x and y. The constants c_1 and c_2 are used to stabilize the division with weak denominators and are defined as $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$, where L is the dynamic range of the pixel values. For images normalized to the [0, 1] range, $L = 1$, and typical values for k_1 and k_2 are 0.01 and 0.03, respectively. A higher SSIM value indicates better SR performance.

The SR results for the 23rd dataset, which consist of images of a single-layer graphene field-effect transistor (FET), demonstrate the effectiveness of the proposed method. Fig. 4a shows one of the LRIs from the set of $N = 9$ images. Fig. 4b represents the true HRI after noise removal via the DnCNN [45], which serves as a reference for evaluating the SR methods. The PSNR and SSIM values between this denoised HRI (Fig. 4b) and the SR images with a scaling factor of $s_r = 3$ are used as the evaluation metrics.

With the bicubic interpolation approach, as shown in Fig. 4c, a PSNR of 18.70 and an SSIM of 0.40 are obtained, indicating moderate performance in preserving structural details. We next consider the results of the bicubic stacking method, which is referred to as Z in 2.2 and used as a backprojection approach. While this method is effective in preserving the structural similarity, it struggles with accurately reproducing image contrast, leading to a reduction in the PSNR value. As shown in Fig. 4d, higher SSIM values (0.58) tend to be achieved with the bicubic stacking method, indicating better structural similarity. Nevertheless, the method struggles with accurate contrast estimation, which results in a lower PSNR of 17.11. The deep learning-based SR method (RAMS), which is specifically trained based on HR satellite images, achieves a PSNR of 16.28 and an SSIM of 0.44, as shown in Fig. 4e. This lower PSNR value suggests that despite the slightly better structural similarity achieved with this approach, this method struggles with noise and fine detail preservation, likely due to a mismatch between the training data and the unique features of the graphene FET.

The proposed method, as shown in Fig. 4f, excels in accurately reproducing the edge structures of the denoised HRIs, demonstrated by its high SSIM of 0.56 and PSNR of 21.05. These values highlight the method's effectiveness, which stems from three key factors. The first key factor is the use of sparse coding with a 6×6 pixel dictionary learned from HR 3D nano-ESCA datasets. Second, unlike methods that apply backprojection to noisy LRIs [19, 40, 43], this approach uses the accumulated bicubic image for

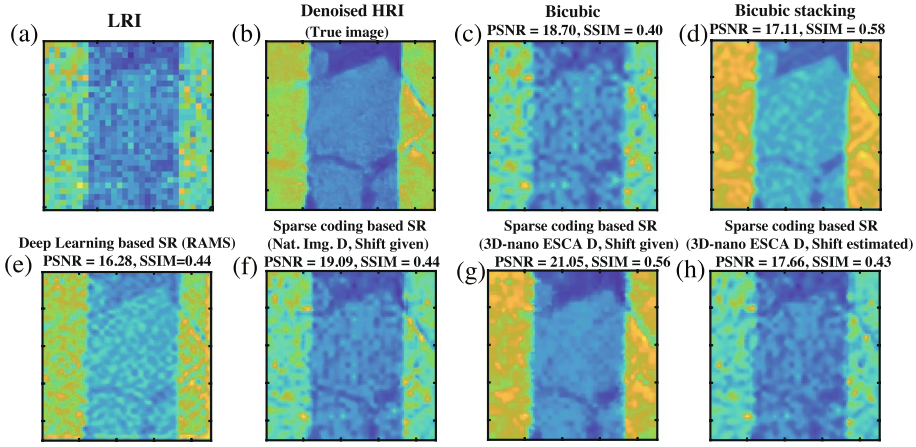


Fig. 4 Superresolution (SR) results for the 23rd dataset among the 146 datasets analysed. The images are of a single-layer graphene field-effect transistor (FET). **a** One of the low-resolution images (LRIs) from the set of $N = 9$ images. **b** True high-resolution image (HRI) after noise removal via the DnCNN [45]. The PSNR and SSIM values between these denoised HRIs and the SR images with a scaling factor of $s_r = 3$ were used for evaluation. **c** Bicubic interpolation result from a single LRI. **d** Bicubic stacking interpolation result from multiple LRIs. **e** Multiframe SR result when a deep neural network trained on HR satellite images was used [37]. **f** Multiframe SR using sparse coding with a dictionary of 15 × 15 pixels learned from natural images, with given shifts *W*. **g** Multiframe SR using sparse coding with a 6 × 6 pixel dictionary learned from HR 3D nano-ESCA datasets, with given shifts *W*. **h** Similar to **g**, but with shifts *W* estimated via the method in [20]

backprojection, thereby preserving image quality. Third, the accurate use of shift variations in the LRIs, which are critical control parameters in the measurement process, ensures precise alignment and reconstruction, further enhancing the overall structural integrity and details captured in the SR image.

On the other hand, Fig. 4e and g present results that are more similar to those of the bicubic interpolation approach. This similarity arises from the influence of backprojection on the LRI, a common technique used in previous multiframe SR studies [19, 40, 43]. In Fig. 4e, although the shifts are given, the use of a dictionary trained based on natural images leads to the retention of unnatural edge artefacts in the image before backprojection. Moreover, the results in Fig. 4g, which were obtained using the same dictionary as in Fig. 4f, suffer from inadequate shift estimation, causing the edge structures to deteriorate, thereby compromising the overall image quality.

The PSNR and SSIM values presented in Table 1 represent the mean values calculated across the 146 LRI and HRI dataset pairs. The proposed method, Sparse Coding SR (3D nano-ESCA Dictionary, Shift Given), achieved the highest mean PSNR of 22.94 dB and the highest mean SSIM of 0.69, along with the smallest standard deviations of 2.80 and 0.11, respectively. These results demonstrate that the proposed method not only produces images with superior quality, closely matching the denoised HR data, but also presents consistent performance across datasets.

In comparison, the bicubic stacking method, although it achieves a relatively high SSIM of 0.62, has a lower PSNR of 18.59 dB. This finding indicates that while the bicubic stacking method is effective in preserving structural similarity, it struggles with accurately reproducing image contrast, leading to a reduction in the PSNR value. On the other hand, the proposed sparse coding SR method obtains good PSNR and SSIM values, outperforming other approaches, including the traditional bicubic interpolation and deep learning-based SR (RAMS) methods [37]. This highlights the effectiveness of using a dictionary specifically learned from 3D nano-ESCA data and incorporating accurate

Table 1 Mean PSNR and SSIM values across 146 datasets for different SR methods

Method	PSNR (dB)	SSIM
Bicubic	21.08 ± 3.62	0.53 ± 0.18
Bicubic stacking	18.59 ± 3.55	0.62 ± 0.10
Deep learning superresolution (RAMS)	16.63 ± 3.94	0.51 ± 0.12
Sparse coding superresolution (3D-nano ESCA Dictionary, Shift Estimation)	21.11 ± 3.39	0.55 ± 0.16
Sparse coding superresolution (3D-nano ESCA Dictionary, Shift Given)	$\mathbf{22.94 \pm 2.80}$	$\mathbf{0.69 \pm 0.11}$
Sparse coding superresolution (Natural image dictionary, Shift Given)	20.81 ± 3.32	0.54 ± 0.14

Bold values indicate the best (i.e., highest) performance values across all the listed superresolution (SR) methods for each metric

shift information during the SR process, leading to both high structural similarity and accurate image contrast.

Previous studies have shown that methods based on natural image dictionaries often struggle to accurately reconstruct fine structures in nanoscale imaging due to the mismatch between the training data and the target data [19, 20, 40, 43]. Our approach addresses this issue by tailoring the dictionary specifically to the 3D nano-ESCA data, which results in better reconstruction performance, as evidenced by the higher PSNR and SSIM values. Furthermore, the incorporation of precise shift information during the SR process enhances the alignment and reconstruction accuracy, a factor that has been overlooked or inadequately addressed in previous SR methods [37]. This advancement in SR methodology contributes to improving image quality in the field of synchrotron radiation microscopy, and it has the potential to be applied to other imaging systems that require high precision with limited datasets.

As the number of LRIs increases from 1 to 9, the sparse coding SR method (3D nano-ESCA Dictionary, Shift Given) consistently achieves the highest PSNR and SSIM values across all scenarios (Fig. 5). For dataset 23, as shown in Supplementary Information, a PSNR of 22.85 and an SSIM of 0.69 are achieved with approximately 5 images, comparable to the results obtained with 9 images. This suggests that high image quality can be achieved while reducing the measurement time by more than 40%. In contrast, the other methods achieve minimal improvements in the PSNR and SSIM values as the number of LRIs increases, underscoring the limitations of these approaches. The consistent superiority of the proposed method highlights its efficiency in maintaining high image quality with fewer images. This approach offers enhanced image quality, reduced exposure times, and improved interpretability of scientific data, making it a versatile tool for overcoming the challenges associated with radiation damage and sample degradation in nanoscale imaging.

One key strategy to further improve estimation accuracy is to incorporate additional prior knowledge about the sample. In the present work, our goal was to develop a broadly applicable approach that could be used for various nanomaterials, such as single-layer graphene field-effect transistors (FETs) or surface-scratched gallium nitride (GaN) films, in the context of 3D-nano ESCA. Consequently, our learned dictionary was designed to be general rather than specialized for a specific pattern. However, as reported by [6], when the nanoparticle pattern or other structural features of interest are known in advance, learning a dedicated dictionary that reflects this prior knowledge can significantly enhance the reconstruction quality.

The integration of more specific prior information is a promising direction for improving the robustness of super-resolution methods in high-noise environments.

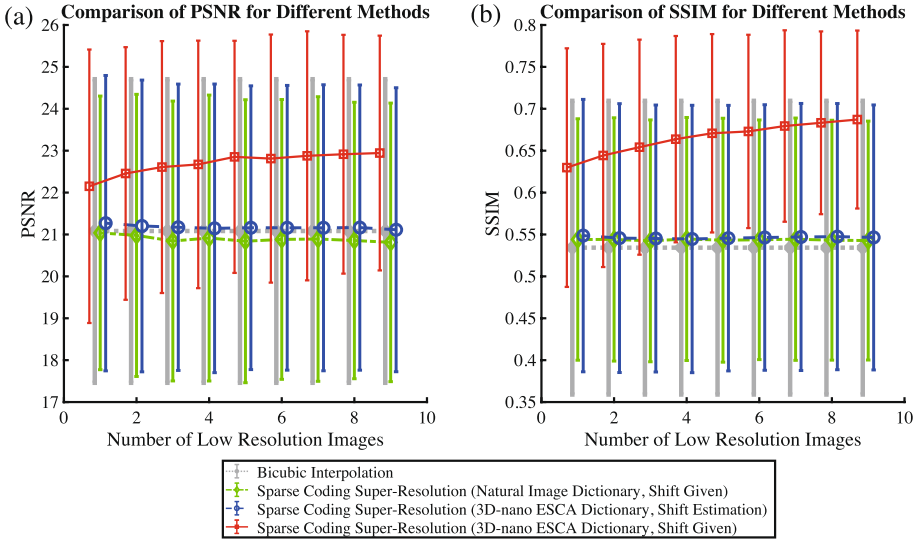


Fig. 5 The variation in the PSNR as the number of low-resolution images increases from 1 to 9 in the multiframe SR analysis. The sparse coding SR method (3D nano-ESCA Dictionary, Shift Given) consistently achieves the highest PSNR values across all the scenarios. In this figure, to improve the visibility of the plotted data, a small offset was applied to the error bars, preventing overlap and ensuring a clearer distinction between different methods

Furthermore, advanced imaging techniques such as electron holography—where electron interference patterns (holograms) are formed by overlapping object and reference waves using an electron biprism—could benefit from similar approaches in extracting high-fidelity signals from noisy datasets. Extending our method to these and other cutting-edge measurement techniques remains an important avenue for future research.

In this study, we generated super-resolution (SR) images with a fixed scaling factor of $s_r = 3$ to demonstrate the effectiveness of our proposed method. However, in real applications, the determination of the optimal enhancement factor is a multifaceted issue. First, the experimental sampling conditions, including how multiple low-resolution images are acquired, must be adapted to the desired upsampling factor. Second, the dictionary learning framework intrinsically couples the high-resolution patch size (q_h) with the low-resolution patch size (q_l) according to the scaling factor. For instance, if q_h is set to 6×6 pixels, then q_l can be defined as 3×3 pixels for a $2\times$ upsampling or 2×2 pixels for a $3\times$ upsampling. This coupling limits the upsampling factor to integer values (or their reciprocals). Although one might consider using larger patch sizes to enable more flexible scaling factors, the relatively small size of the original images and the inherent multi-scale characteristics of materials images pose practical challenges in dictionary learning [11]. Such constraints may represent a disadvantage of dictionary-based SR compared to interpolation-based methods like bicubic interpolation.

Nevertheless, for advanced materials analysis—where large training datasets are typically unavailable—the ability to learn a meaningful dictionary from a limited number of images is a significant advantage. Future research will focus on developing multi-scale and adaptive dictionary learning strategies to relax these constraints, potentially allowing for more flexible upsampling factors while maintaining high reconstruction quality.

In the current work, high-resolution images are used to learn a dictionary with effective patch sizes that typically span dimensions from approximately $10 \times 10\mu m$ to $20 \times 20\mu m$, thereby covering multiple scales of the sample features. Based on these patch sizes, a fixed scaling factor of $s_r = 3$ is adopted. This scaling factor not only determines

the resolution enhancement but also implicitly defines the acquisition parameters for the low-resolution images, including their size and the number of frames recorded.

While this approach offers a systematic way to obtain low-resolution frames from high-resolution data, it also suggests that by adjusting the high-resolution image selection or modifying the dictionary parameters, the scaling factor could be tuned to optimize the reconstruction quality. Exploring such adaptive strategies for determining the ideal scaling factor is a promising direction for future work.

4.3 Multiframe superresolution method for microstructured gold patterns and fine surface irregularities on GaN films

To further validate the superiority of the proposed method, we applied our SR analysis approach to newly acquired datasets, including 9 LRIs and HRIs of microstructures and surface irregularities on GaN films. Using this set of 9 LRIs, we performed multiframe SR with a scaling factor of $s_r = 3$, following the same protocol described previously. This allowed us to assess the effectiveness of the method across various datasets, ensuring consistent performance in preserving structural details. The results for a 1000 nm gold (Au) wire arranged in a cross pattern, shown in Fig. 6, highlight the effectiveness of various reconstruction methods. The sparse coding approach using the 3D nano-ESCA dictionary with provided shifts (Fig. 6e) effectively preserves the sharp edges of the gold wire, crucial for accurate characterization. This method consistently outperforms both the bicubic interpolation approach (Fig. 6c), which produces a more blurred image, and the sparse coding method with estimated shifts (Fig. 6f), which shows only marginal improvement. The clear advantage of incorporating shift information and using a

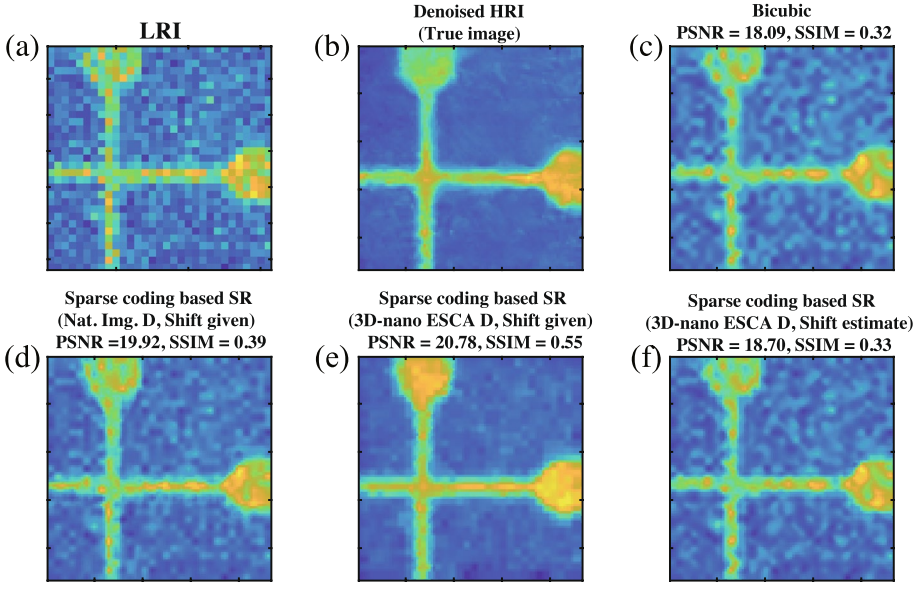


Fig. 6 Superresolution analysis results for a 1000 nm gold (Au) wire arranged in a cross pattern. **a** Low-resolution image (LRI). **b** Denoised high-resolution image (HRI) used as the ground truth. **c** Bicubic interpolation result, with a PSNR of 18.09 and an SSIM of 0.32. **d** Sparse coding-based superresolution result using a dictionary learned from natural images with provided shifts, yielding a PSNR of 19.92 and an SSIM of 0.39. **e** Sparse coding-based superresolution result using the 3D nano-ESCA dictionary with provided shifts, achieving the highest PSNR of 20.78 and an SSIM of 0.55. **f** Sparse coding-based superresolution approach with estimated shifts [20], resulting in a PSNR of 18.70 and an SSIM of 0.33

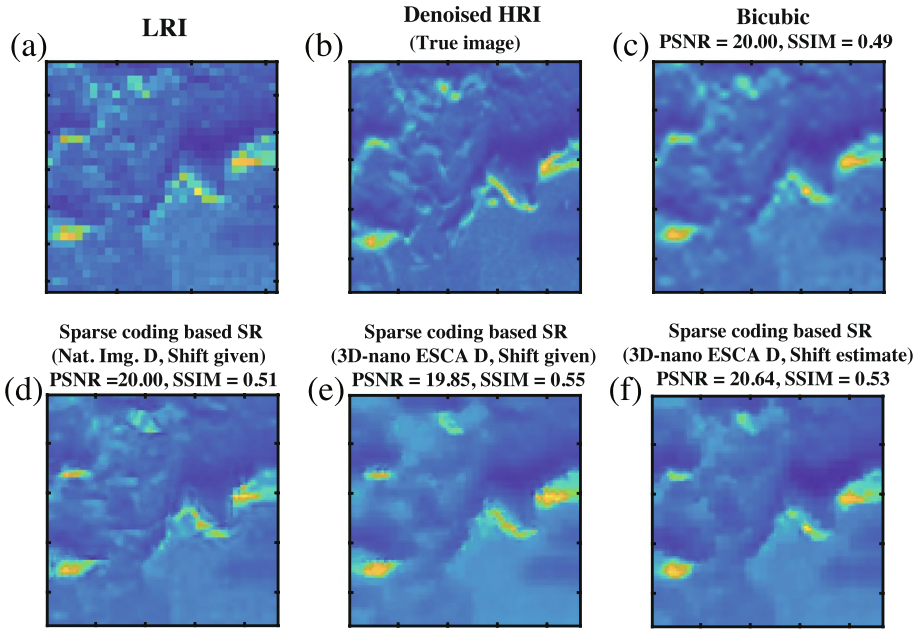


Fig. 7 Superresolution analysis results for a SiC structure. **a** Low-resolution image (LRI). **b** Denoised high-resolution image (HRI) used as the ground truth. **c** Bicubic interpolation approach, with a PSNR of 20.00 and an SSIM of 0.49. **d** Sparse coding-based superresolution approach using a dictionary learned from natural images with provided shifts, yielding a PSNR of 20.00 and an SSIM of 0.51. **e** Sparse coding-based superresolution approach using the 3D nano-ESCA dictionary with provided shifts, achieving a PSNR of 19.85 and an SSIM of 0.55. Despite the slightly lower PSNR, this method better reconstructs key features of the HRI. **f** Sparse coding-based superresolution approach with estimated shifts [20], resulting in a PSNR of 20.64 and an SSIM of 0.53

specialized dictionary is evident in the higher structural similarity and contrast achieved by our approach.

The in-plane resolution of the imaging system was estimated using a standard procedure based on intensity profile analysis. For each selected row, the profile was split at its maximum intensity, and rising and falling sigmoid functions were fitted to the left and right portions, respectively. The resolution was calculated as $R = \frac{2\ln(3)}{|\text{slope}|}$, and converted into physical units using the known pixel size. As shown in Fig. S5 and the corresponding table in the Supplementary File, the proposed Sparse Coding Superresolution (3D-nano ESCA Dictionary, Shift Given) method consistently achieves a resolution of 0.1396 μm (std = 0.0213 μm), which is significantly better than the resolution obtained from the denoised high-resolution image (0.2061 μm). The Sparse Coding Superresolution method (using the 3D-nano ESCA Dictionary with Shift Given) robustly achieves a resolution of $0.1396\mu m$ with a margin of $0.0213\mu m$, effectively matching the system's in-plane resolution of approximately $100nm$ even under significant noise conditions. Further details are provided in the Supplementary File.

These results emphasize the importance of using a specialized dictionary and accurate shift information for high-quality image reconstruction, confirming the superiority of the proposed method in preserving the true structure of the Au wire.

The SR analysis results for fine structures on a GaN film, as shown in Fig. 7, demonstrate the varying effectiveness of different reconstruction methods. The LRI in Fig. 7a serves as a reference for comparison. When upscaled via bicubic interpolation (Fig. 7c), the result shows a PSNR of 20.00 and an SSIM of 0.49. However, despite this relatively high PSNR value, the interpolation method fails to accurately reconstruct the fine

structural details, particularly in the central line region, which is a key feature in the denoised HRI shown in Fig. 7b.

When the sparse coding-based SR approach with a dictionary learned from natural images is used (Fig. 7d), a PSNR of 20.00 and an SSIM of 0.51 are achieved, indicating some improvement in edge detail preservation. However, the image still retains unnatural edge artefacts, and the true structure is not accurately reconstructed. In contrast, the sparse coding method using the 3D nano-ESCA dictionary with provided shifts (Fig. 7e) produces a PSNR of 19.85 and an SSIM of 0.55. Although the PSNR is slightly lower than that of the bicubic interpolation approach, this method more accurately reconstructs the central line and other subtle features of the HRI, making it a better overall representation of the true image. The ability of the sparse coding method to preserve these fine details, especially in the challenging regions like the central line, demonstrates its superiority over traditional interpolation methods. This shows that even when numerical metrics like PSNR are comparable, the quality of the structural details preserved by the sparse coding method is significantly better, highlighting its robustness and precision in applications requiring high fidelity, such as nanoscale imaging. Moreover, when shifts are estimated rather than provided (Fig. 7f), the PSNR improves to 20.64, and the SSIM is 0.53, which is slightly greater than that of the bicubic interpolation approach. Despite this improvement in numerical metrics, the method still struggles to fully replicate the central line structure, particularly in regions of high complexity. This indicates that while the estimated shifts improve alignment, they are not sufficient to capture the subtle details present in the high-resolution image, which underscores the importance of using accurate shift information. Thus, the use of provided shifts in the sparse coding approach remains crucial for the superior reconstruction of fine structures.

To further investigate the variability observed in Fig. 7, we analyzed the performance of bicubic interpolation across multiple frames. As shown in Supplementary Fig. S5, bicubic interpolation was applied to nine LRI frames and the corresponding PSNR and SSIM values are presented. These results reveal that, while high metric values can occasionally be observed due to noise, our proposed method more consistently captures the true signal.

In this study, our analysis method is presented under the assumption that the actuator operates with precision far exceeding the resolution. However, in cases where the actuator's precision is not as high relative to the resolution, it would be beneficial to simultaneously estimate the shift values during the analysis. The convolution of noise in each dataset suggests that the estimation of shift values can become unstable. To counteract this, extending the measurement time of the low-resolution images to improve the S/N ratio is crucial for stabilizing the shift estimation. Extending our method to image analysis in other advanced measurement techniques remains a promising direction for future research.

Although our current results demonstrate a 40% reduction in radiation dose, we believe that further reductions are achievable. The primary challenge in pushing this number further down lies in accurately estimating which components of the acquired data represent the true signal versus noise. As discussed earlier, the incorporation of prior knowledge into the reconstruction process can significantly improve this signal/noise discrimination. Furthermore, a more refined modeling of the measurement system—including a detailed analysis of the noise characteristics (e.g., distinguishing

between Poisson and Gaussian noise)—is critical. Such an approach would allow for a more robust estimation of the underlying signal, even under lower dose conditions. These enhancements in noise modeling and the integration of additional prior information are promising avenues for future work, potentially enabling even greater reductions in radiation dose while maintaining, or even improving, image quality.

While the proposed sparse coding SR method demonstrates excellent performance in reconstructing high-resolution images from low-resolution inputs, there are certain limitations to its generalizability. One notable limitation is the reliance on a dictionary that is specifically trained for a given dataset or scale [11]. When the scale of the imaging data changes significantly, such as when analyzing structures at vastly different spatial resolutions, the dictionary must be retrained to accommodate the new data. This requirement limits the method's ability to be applied universally across datasets without prior retraining, thus affecting its scalability and computational efficiency.

When deciding between deep learning [9, 10, 21, 37] and sparse coding [11, 12, 19, 20, 43] for superresolution, the nature of the dataset plays a critical role. Deep learning, particularly convolutional neural networks (CNNs), works best with large datasets. In such cases, transfer learning or fine-tuning on pre-trained models can effectively adapt them to new tasks [4, 15, 31, 44, 46]. However, deep learning methods require significant data and computational resources, which may not be feasible for smaller, specialized datasets like those in nanoscale imaging. Sparse coding, on the other hand, is advantageous when working with smaller, noisier datasets or when interpretability is essential. It allows the incorporation of domain-specific knowledge into the dictionary learning process, making it more adaptable to unique datasets. While multi-scale dictionaries extend sparse coding to broader scales [32, 33], retraining is often required for different datasets, limiting its scalability. Choosing between these methods depends on the dataset's scale, complexity, and noise. For well-represented datasets, deep learning with transfer learning may be ideal, while sparse coding is better suited for specialized, limited datasets. Future research could explore hybrid approaches, combining deep learning for preprocessing with sparse coding for fine-tuned reconstruction, balancing generalization and precision.

5 Conclusion

This study presents a multiframe superresolution (SR) method based on sparse coding for enhancing synchrotron radiation microspectroscopy images, significantly reducing exposure times and mitigating issues such as radiation damage and thermal drift. By leveraging positional shifts as known control parameters and utilizing a high-resolution dictionary tailored to 3D nano-ESCA data, our method achieves high-resolution reconstructions with minimal data, outperforming traditional methods like bicubic interpolation and even deep learning-based approaches under limited dataset conditions.

The broader applicability of this technique, particularly in fields where exposure constraints and data limitations are common, highlights its potential for advancing high-resolution imaging beyond synchrotron radiation microspectroscopy. Future work could explore the integration of this method into other imaging modalities, such as electron microscopy or Raman spectroscopy, where similar challenges exist. Additionally, further refinement of the dictionary learning process could enhance SR performance,

particularly in highly noisy environments, making the approach even more versatile across different applications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s11671-025-04291-x.

Supplementary file 1 (pdf 3067 KB)

Acknowledgements
This work was supported by KAKENHI Grant-in-Aid for Scientific Research (24K02995, 21H04696, 24K01243) from the Japan Society for the Promotion of Science (JSPS), CREST: JPMJCR1761, JPMJCR21O1, JPMJCR2235 and PRESTO: JPMJPR17N2, JPMJPR17N2, JPMJPR20T7 from the Japan Science and Technology Agency (JST), and the Innovative Science and Technology Initiative for Security of ATLA Japan (JPJ004596). A part of this study was financially supported by Beyond 5 G Research and Development Project (05901). We thank Yasunori Tateno and Tsuyoshi Kouchi of Sumitomo Electric Industries, Ltd., for their assistance with sample preparation. Experimental measurements were performed with a 3D nano-ESCA system installed at the University of Tokyo outstation beamline, BL07LSU at SPring-8 (2022A7444, 2021B7435, 2021B7433, 2021A7422, 2021A7421, 2019B7466, 2019A7451, 2018B7580), which has since been relocated to NanoTerasu.

Author contributions
Y.I. and N.N. contributed equally to this work. Y.I. led the data analysis with support from M.S. N.N. was responsible for acquiring the new measurement data, including those of GaN and Au, as well as conducting previous 3D-nano ESCA measurements. Both Y.I. and N.N. jointly determined the research direction and wrote the manuscript. H.F. provided the GaN and Au samples for the new measurements. H.H. and M.O. offered programs related to sparse super-resolution and provided valuable insights on machine learning techniques. All authors reviewed the manuscript and contributed to its completion.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials
Data is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors have reviewed and approved the final version of the manuscript and consent to its publication.

Competing interests
The authors declare no Conflict of interest.

Received: 11 September 2024 / Accepted: 20 June 2025
Published online: 03 July 2025

References

1. Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process. 2006;54(11):4311–22.
2. Ahmed N, Natarajan T, Rao KR. Discrete cosine transform. IEEE Trans Comput. 1974;100(1):90–3.
3. Akada K, Sudayama T, Asakura D, et al. Microscopic photoelectron analysis of single crystalline licoo2 particles during the charge-discharge in an all solid-state lithium ion battery. Sci Rep. 2019;9(1):12452.
4. Ali AH, Yaseen MG, Aljanabi M, et al. Transfer learning: a new promising techniques. Mesop J Big Data. 2023;2023:29–30.
5. Amati M, Barinov A, Feyer V, et al. Photoelectron microscopy at elettra: recent advances and perspectives. J Electron Spectrosc Relat Phenom. 2018;224:59–67.
6. Anada S, Nomura Y, Hirayama T, et al. Sparse coding and dictionary learning for electron hologram denoising. Ultramicroscopy. 2019;206: 112818.
7. Avila J, Razado-Colambo I, Lorcy S, et al. Antares, a scanning photoemission microscopy beamline at soleil. J Phys: Conf Ser. 2013;425(192023):1–4.
8. Deng D, Novoselov K, Fu Q, et al. Catalysis with two-dimensional materials and their heterostructures. Nat Nanotechnol. 2016;11:218–30.
9. Deudon M, Kalaitzis A, Goytom I. et al (2020) Highres-net: Recursive fusion for multi-frame super-resolution of satellite imagery. arXiv preprint arXiv:2002.06460.
10. Dong C, Loy CC, He K, et al. Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell. 2015;38(2):295–307.
11. Elad M. Sparse and redundant representations: from theory to applications in signal and image processing. Berlin: Springer Science & Business Media; 2010.

12. Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process. 2006;15(12):3736–45.
13. Günther S, Kaulich B, Gregoratti L, et al. Photoelectron microscopy and applications in surface and materials science. Prog Surf Sci. 2002;70:187–260.
14. Guo S, Yang X, Ma J. et al (2022) A differentiable two-stage alignment scheme for burst image reconstruction with large shift. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 17472–17481.
15. Hinton G (2015) Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
16. Hong I, Lee T, Yin G, et al. Performance of the srrc scanning photoelectron microscope. Nucl Inst Methods Phys Res A. 2001;467:905–8.
17. Horiba K, Nakamura Y, Nagamura N, et al. Scanning photoelectron microscope for nanoscale three-dimensional spatial-resolved electron spectroscopy for chemical analysis. Rev Sci Instrum. 2011;82(11): 113701.
18. Irani M, Peleg S. Motion analysis for image enhancement: resolution, occlusion, and transparency. J Vis Commun Image Represent. 1993;4(4):324–35.
19. Kato T, Hino H, Murata N. Multi-frame image super resolution based on sparse coding. Neural Netw. 2015;66:64–78.
20. Kato T, Hino H, Murata N. Double sparsity for multi-frame super resolution. Neurocomputing. 2017;240:115–26.
21. Kawulok M, Benecki P, Piechaczek S, et al. Deep learning for multiple-image super-resolution. IEEE Geosci Remote Sens Lett. 2019;17(6):1062–6.
22. Lee H, Battle A, Raina R. et al (2006) Efficient sparse coding algorithms. Adv Neural Inf processing systems 19
23. Lee M, Shin H. Soft x-ray spectromicroscope at the pohang light source. Rev Sci Instrum. 2001;72(6):2605–9.
24. Luo Z, Yu L, Mo X. et al (2021) Ebsr: Feature enhanced burst super-resolution with deformable alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 471–478.
25. Mino L, Borfecchia E, Segura-Ruiz J, et al. Materials characterization by synchrotron x-ray microprobes and nanoprobes. Rev Mod Phys. 2018;90(2): 025007.
26. Nagamura N, Fukidome H, Nagashio K, et al. Influence of interface dipole layers on the performance of graphene field effect transistors. Carbon. 2019;152:680–7.
27. Nagamura N, Konno S, Matsumoto M, et al. Photoelectron spectromicroscopy analysis of graphene during gate-controlled photo-oxidation process. Nano Express. 2022;3(4): 044003.
28. Nakamura K, Nagamura N, Ueno K, et al. All 2d heterostructure tunnel field-effect transistors: Impact of band alignment and heterointerface quality. ACS Appl Mater Interfaces. 2020;12(46):51598–606.
29. Novoselov K, Mishchenko A, Carvalho A, et al. 2d materials and van der waals heterostructures. Science. 2016;353(6298):aac9439.
30. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996;381(6583):607–9.
31. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2009;22(10):1345–59.
32. Papyan V, Romano Y, Elad M. Convolutional neural networks analyzed via convolutional sparse coding. J Mach Learn Res. 2017;18(83):1–52.
33. Papyan V, Romano Y, Sulam J, et al. Theoretical foundations of deep learning via sparse representations: a multilayer sparse model and its connection to convolutional neural networks. IEEE Signal Process Mag. 2018;35(4):72–89.
34. Peng L, Zhu Y, Chen D, et al. Two-dimensional materials for beyond-lithium-ion batteries. Adv Energy Mater. 2016;6(11):1600025.
35. Sakai E, Nagamura N, Liu J, et al. Investigation of the enhanced photocathodic activity of la5ti2cus5o7 photocathodes in h2 evolution by synchrotron radiation nanospectroscopy. Nanoscale. 2016;8(45):18893–6.
36. Sakdinawat A, Attwood D. Nanoscale x-ray imaging. Nat Photon. 2010;4:840–8.
37. Salvetti F, Mazzia V, Khaliq A, et al. Multi-image super resolution of remotely sensed images using residual attention deep neural networks. Remote Sens. 2020;12(14):2207.
38. Shen G, Chen PC, Ryu K, et al. Devices and chemical sensing applications of metal oxide nanowires. J Mater Chem. 2009;19(7):828–39.
39. Shi S, Gu J, Xie L, et al. Rethinking alignment in video super-resolution transformers. Adv Neural Inf Process Syst. 2022;35:36081–93.
40. Wang P, Hu X, Xuan B. et al (2011) Super resolution reconstruction via multiple frames joint learning. In: 2011 International Conference on Multimedia and Signal Processing, IEEE, 357–361.
41. Wang Z. Nanobelts, nanowires, and nanodiskettes of semiconducting oxides - from materials to nanodevices. Adv Mater. 2003;15(5):432–6.
42. Warwick T, Ade H, Cerasari S, et al. Development of scanning x-ray microscopes for materials science spectromicroscopy at the advanced light source. J Synchrotron Radiat. 1998;5:1090–2.
43. Yang J, Wright J, Huang T. et al (2008) Image super-resolution as sparse representation of raw image patches. In: 2008 IEEE conference on computer vision and pattern recognition, IEEE 1–8.
44. Yosinski J, Clune J, Bengio Y. et al (2014) How transferable are features in deep neural networks?. Adv Neural Inform Process Syst 27.
45. Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process. 2017;26(7):3142–55.
46. Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. Proc IEEE. 2020;109(1):43–76.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.