

Depth-wise super-resolution of a serial-sectioned low-carbon steel microstructure by a generative adversarial network trained on in-plane images

Hoheok Kim

hoheokkim@kims.re.kr

Korea Institute of Materials Science (KIMS)

Junya Inoue

The University of Tokyo

Article

Keywords: Super-resolution, steel microstructure, 3D image, deep learning

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10904615/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Depth-wise super-resolution of a serial-sectioned low-carbon steel microstructure by a generative adversarial network trained on in-plane images

Hoheok Kim^{*1,2}, Junya Inoue^{2,3}

¹Korea Institute of Materials Science (KIMS), 797, Changwon-daero, Seongsan-gu, Changwon-si, Gyeongnam-do, 51508, Republic of Korea

²Institute for Industrial Science, The University of Tokyo, 5-1-5 Kashiwanoha Kashiwa 277-0082, Japan

³Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo Bunkyo Tokyo 113-8656, Japan

e-mail address

Hoheok Kim: hoheokkim@kims.re.kr (+82-55-280-3482)

Junya Inoue: inoue@material.t.u-tokyo.ac.jp

*corresponding author's e-mail: hoheokkim@kims.re.kr

Abstract

In this study, a deep-learning super-resolution framework that restores the depth-wise resolution of the 3D microstructure of a low-carbon steel without the need of additional training data is presented. The 3D microstructure of a metallic material is generally obtained through serial sectioning followed by observation of the sample surface. This provides a full 3D image, but its resolution along the depth direction is low compared to that of the microscope-observed planes. Various deep learning methods have been developed to overcome this limitation, but they mostly require additional 3D microstructure datasets to train the networks. Here, an enhanced super-resolution generative adversarial network (ESRGAN) trained on low and high resolution pairs was employed to restore the depth-wise resolution of the microstructure. As the pairs were generated internally from the specimen's own surface images, the trained network was applied along the depth direction without any external dataset. Together with a conventional SRGAN, a super-resolution convolutional neural network (SRCNN) and three interpolation schemes, we compared the methods in terms of microstructural and statistical analysis, as well as structural similarity. Among the compared methods, the ESRGAN provided the most faithful reconstruction of the depth-wise microstructure.

Keywords

Super-resolution, steel microstructure, 3D image, deep learning

Introduction

Metallic materials present distinct analytical challenges due to their complex microstructures, which result from the combination of multiple alloying elements and the diverse heat treatments they undergo during production. The microstructure of metallic materials is composed of various phases and features, including grains, phases, dislocations, and precipitates, all of which interact to determine the overall properties of materials. Steel, for instance, may contain phases such as ferrite, austenite, martensite, or bainite, depending on the thermal history[1–4]. These microstructures are generally formed in diverse and intricate ways, which makes their characterization and analysis much more challenging and requires advanced methods to fully understand their behavior in different applications[5–8]. Traditional methods for investigating a material's microstructure have primarily relied on analyzing two-dimensional (2D) cross-sections of a sample. This typically involves preparing the material surface through a series of steps, such as polishing to remove surface irregularities and etching to reveal distinct features. By investigating 2D microstructure images with a microscope, researchers can identify microstructural components like grains, phases, and defects. While this approach has proven invaluable in most cases, it is inherently limited when it comes to capturing the complexity of three-dimensional (3D) microstructural features[9–11]. These limitations become particularly evident in materials with intricate internal architectures, such as those containing sparsely distributed pores, microcracks, or non-uniform features arising from complex manufacturing processes. For such materials, 2D observations often fail to provide a complete picture, potentially leading to an incomplete understanding or even misinterpretation of the material's behavior and properties.

In contrast, 3D microstructure observation enables a more comprehensive examination of a material's internal structure, offering insights that are unattainable through 2D analysis alone. By reconstructing the material's structure in 3D, researchers can gain a deeper

understanding of spatial distributions, connectivity, and the interactions between microstructural features. This comprehensive view is crucial for advancing materials science, as it provides valuable information for optimizing manufacturing processes, improving material design, and enhancing the reliability and performance of materials in practical applications[12–15]. For 3D microstructural analysis, both non-destructive methods such as tomography-based techniques[16–18] and destructive methods such as serial sectioning-based techniques[19–23] are widely employed. Among these, serial sectioning is particularly well-suited for studying metallic materials, as it offers a broader observation range and higher spatial resolution compared to many other methods. This makes it a preferred choice for capturing the detailed microstructural features necessary for understanding and improving the performance of advanced materials.

However, the resolution of a 3D image generated through serial sectioning-based methods often exhibits anisotropy, meaning that the resolution varies considerably between different spatial dimensions[24–27]. Specifically, the resolution in the x and y dimensions, which correspond to the material's surface, is largely governed by the capabilities of the microscope used to capture the images. Advanced microscopy techniques can achieve extremely fine resolutions in these planar directions, allowing for detailed visualization of microstructural features. In contrast, the resolution along the z -axis, which represents the depth of the material, is primarily determined by the precision and consistency of the sectioning or milling technique employed to remove successive layers of the sample. Factors such as the thickness of each removed layer, the accuracy of layer alignment, and potential deformation during sectioning can influence the depth resolution, often making it inferior to the surface resolution[26,28].

This inherent disparity in resolution between the x - y plane and the z -axis can introduce distortions and inconsistencies in the reconstructed 3D image. For example, features that

appear continuous and well-defined in the x - y plane may appear fragmented or blurred along the z -axis, making it challenging to accurately interpret the geometry and connectivity of microstructural components. Such mismatches can lead to errors in quantitative analysis, such as measuring the size, shape, or distribution of features like pores, grains, or inclusions[29,30]. Additionally, these resolution differences can hinder the ability to identify subtle or complex microstructural phenomena, which are crucial for understanding material behavior and properties[31,32]. Overcoming these challenges requires careful optimization of the serial sectioning process, such as minimizing layer thickness and ensuring precise alignment between sections, as well as the development of advanced image reconstruction algorithms to mitigate the effects of anisotropic resolution and enhance the overall fidelity of the 3D representation.

To address this issue, researchers have recently endeavored to achieve isotropic resolution of 3D microstructures by employing interpolation and deep learning methods [18,31]. Interpolation methods, such as bicubic[35], bilinear[36], and nearest neighbor[37], represent mathematical techniques for interpolating data points or pixels within an image. While computationally simple, interpolation methods do not inherently enhance the level of detail or image quality. In contrast, deep learning-based methods have gained popularity due to their exceptional performance when trained with large datasets. In the field of deep learning, the task of increasing the resolution of an image is termed super-resolution (SR)[38]. Among the popular algorithms, super-resolution convolutional neural network or SRCNN[39] is a pioneering approach that directly learns a mapping from low-resolution to high-resolution images through a three-layer convolutional network, demonstrating that CNN-based methods can outperform conventional interpolation techniques in both accuracy and efficiency. Another representative method is super-resolution generative adversarial network (SRGAN)[40], which adopts a generative adversarial framework integrating adversarial and perceptual losses to recover high-frequency details, leading to photo-realistic sharpness and enabling SRGAN to

generate more convincing textures than traditional or purely CNN-based approaches. It has been demonstrated that SRGAN enables the recovery of details of a microstructure image[26,41]. However, an important issue remains regarding the preparation of low-resolution and high-resolution image pairs, as this process can be both costly and time-consuming in the field of materials science and engineering. Building on SRGAN, enhanced SRGAN or ESRGAN and other GAN-based SR methods further refined the perceptual quality of reconstructed images by incorporating improved network architectures, perceptual loss terms, and progressive training schemes[42–44]. Despite their outstanding performance, these supervised methods are constrained by the need for a large dataset to effectively learn the relationship between Low resolution (LR) and high resolution (HR) images.

To overcome this limitation, reference-free SR techniques have emerged, leveraging internal image statistics to enhance resolution without requiring a large LR-HR dataset. Early work in this direction demonstrated that local self-examples within a single image could guide upsampling[45]. Approaches such as Deep Image Prior (DIP) have further shown how a randomly initialized CNN can serve as an effective prior for restoring high-resolution details without relying on external datasets[46], and SinGAN employs a generative adversarial network trained on a single natural image to achieve reference-free SR among other synthesis tasks[47]. Similar strategies have been applied in volumetric microscopy, where anisotropic resolution is common. Park et al. introduced a deep learning-based reference-free SR technique to achieve isotropic super-resolution in volumetric fluorescence microscopy by training on a single 3D dataset without matched high-resolution targets[48]. Super-resolution that avoids external training data has also begun to gain attention in the field of materials science and engineering, largely because obtaining extensive and high-quality datasets is mostly difficult. Hagita et al. introduced a deep learning-based SR approach to restore asymmetric focused ion beam-scanning electron microscope (FIB-SEM) 3D images of polymeric nano-composites,

thereby improving depth resolution and reducing observation time[26]. Gao et al. applied an unsupervised CycleGAN-based framework to enhance low-quality SEM images of weakly conductive materials, achieving higher resolution and detail restoration that outperforms traditional deblurring methods[49]. However, such approaches remain underutilized for metallic material images particularly for steel, where the microstructural constituents exhibit considerable diversity and complexity. Inspired by prior studies, we investigate its potential for enhancing the axial resolution in 3D microstructure images of steel and demonstrate its efficacy.

In this study, we present a deep-learning application that enhances the depth-wise resolution for the case of a 3D volumetric microstructure of steel without the need for additional reference images. This setting differs from reference-free approaches that learn without matched high-resolution targets. The low- and high-resolution pairs used here are generated from the specimen's own in-plane images. The supervision is therefore fully matched, although no external dataset is required. A general overview of the approach is given in Fig. 1. The basic strategy of this framework is to learn the mapping relationship between low-resolution (LR) features and high-resolution (HR) features from surface images or x - y planes, and leverage this knowledge to increase the depth-wise resolution or z -axis resolution. To demonstrate its applicability, we first obtained a 3D microstructural image of steel, referred to as a ground truth image in this study. Next, the resolution in the depth direction of a 3D image is decreased by taking cross-sectional images at regular intervals so as to create a 3D image with anisotropic resolution. After generating LR and HR image pair datasets from the x - y planes, we trained an ESRGAN to learn how to convert LR images into HR images. Following the training phase, we applied the trained network to x - z plane images to enhance depth-wise resolution. Finally, we compared the ESRGAN-produced images with those obtained using various interpolation methods to evaluate the ability to recover microstructural details.

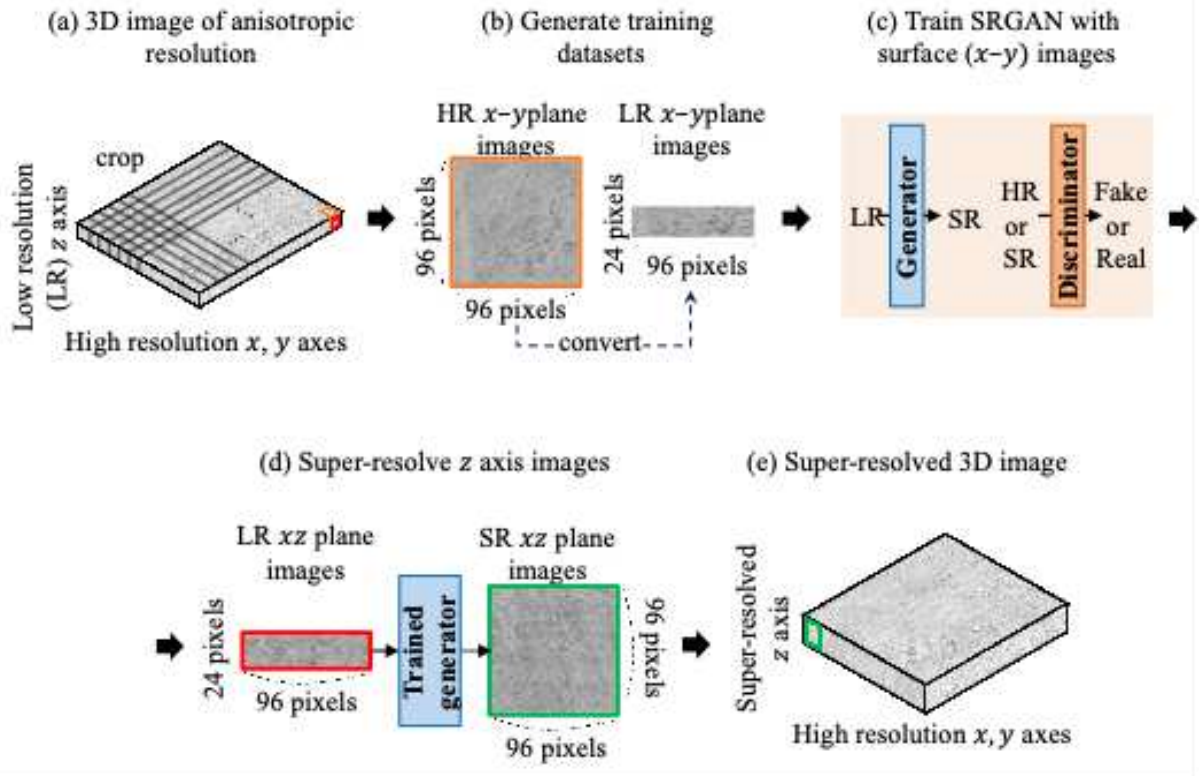


Figure 1. Overview of the proposed ESRGAN framework for enhancing the depth-wise resolution of a 3D microstructure of low-carbon steel.

Results

Fig. 2 (a) illustrates the ground truth 3D microstructure image of a low-carbon steel sample obtained through the serial sectioning method, along with its cross-sectional views. The reconstructed 3D image has spatial dimensions of $2450 \times 2100 \times 101$ pixels along the x , y , and z axes, respectively. Here the sectioning thickness was controlled carefully, so that the resolution is nearly uniform in all directions, approximately $0.8 \mu\text{m}$ per pixel, as shown in Fig. 2 (b-d). These cross-sectional views reveal the intricate development of various microstructural constituents, including ferrite, pearlite, and bainite, each of which plays a critical role in defining the material's mechanical properties. This high-resolution and isotropic dataset serves as a robust foundation for generating the anisotropic-resolution datasets and for providing the ground truth used in validation. Using this 3D image, we further generated 3D datasets with anisotropic resolutions to simulate real-world challenges in imaging and resolution limitations.

For these datasets, the z-axis resolution was systematically reduced by retaining every second, third, or fourth plane along the depth direction, which we refer to as the 2, 3, and 4 times scenarios, representing increasingly anisotropic conditions. Each x - y plane image from these 3D datasets was then divided into smaller images, each measuring 96×96 pixels. These images were further resized based on three different downsampling factors, producing final x - y sections with resolutions of 48×96 , 32×96 , and 24×96 pixels, respectively. This systematic downsampling process generated over 50,000 paired datasets of LR and HR images. Additionally, to ensure comprehensive training and validation of the models, cross-sectional views in the x - z planes of the anisotropic 3D images were processed in the same manner. These x - z plane datasets provided an opportunity to assess the performance of the trained models in handling anisotropic resolution challenges.

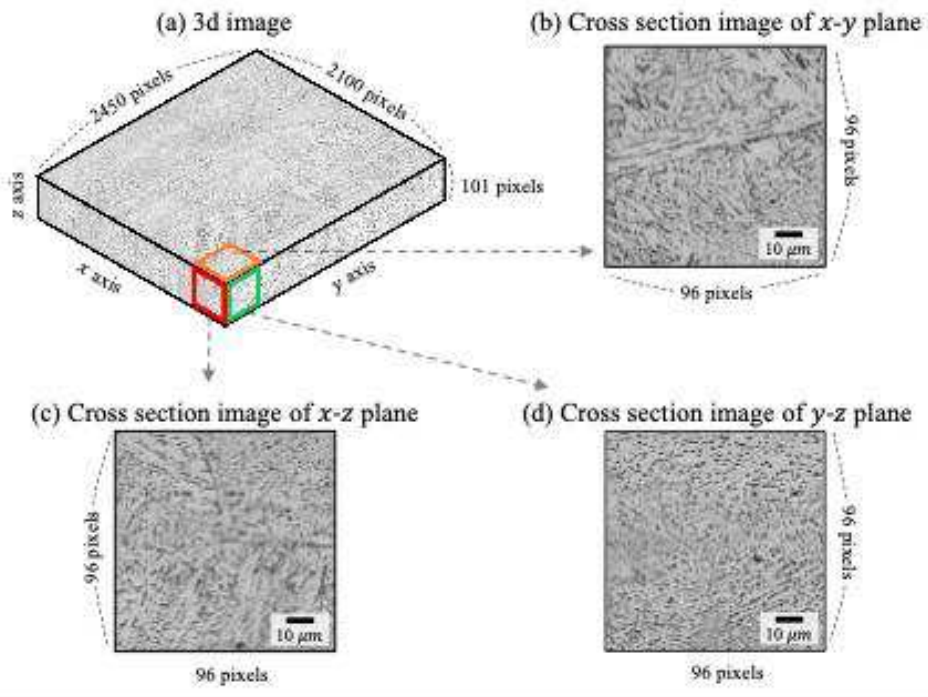


Figure 2. Isotropic reference volume obtained by serial sectioning. (a) Ground-truth 3D microstructure of the low-carbon steel, shown together with its orthogonal cross-sections along the (b) x - y , (c) x - z , and (d) y - z planes.

Fig. 3 presents the super-resolution result for the 3 times super-resolution scenario. Fig. 3 (a) displays the entire x - z plane of the ground truth 3D image, with the region of interest highlighted with a red box, providing an overview of the structural arrangement. Fig. 3 (b) shows the HR image, where the x and z axes have equal resolutions, offering a much clearer view that reveals coarse ferrite, finer ferrite plates, and black pearlite regions. Notably, sharp boundaries and the presence of long, thin pearlite formations extending between ferrite plates are clearly visible. In contrast, the LR image (Fig. 3(c)), with the z -axis resolution three times lower than the x -axis, appears quite blurry and coarse, lacking the fine details needed to accurately understand the microstructure.

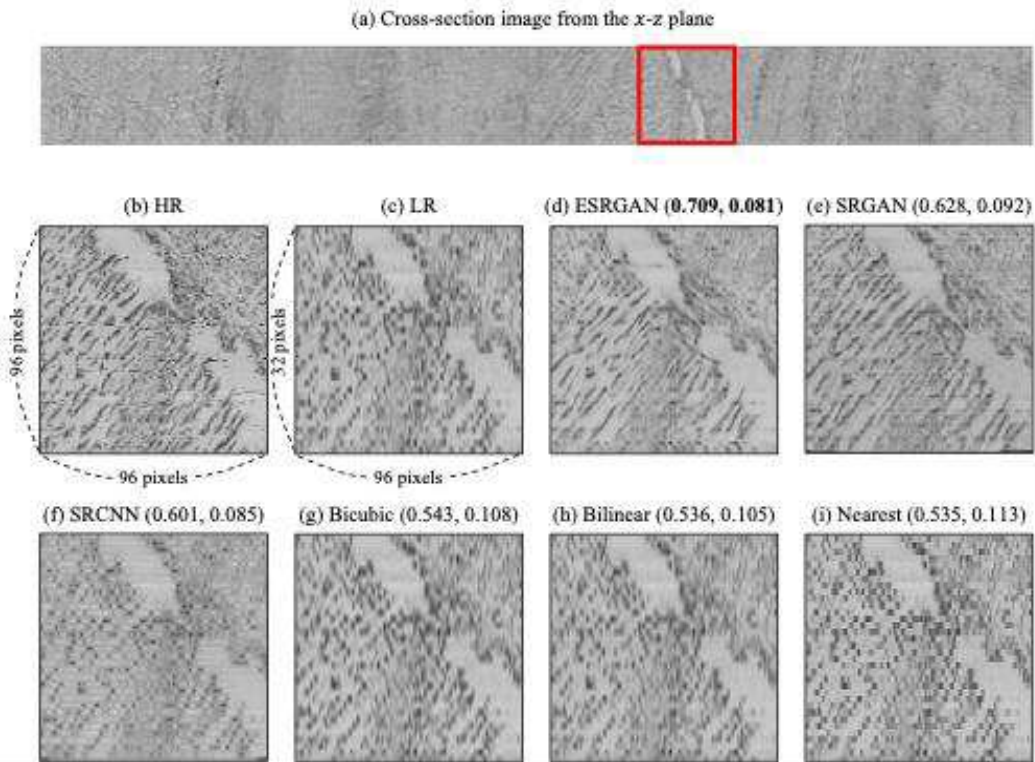


Figure 3. Region-of-interest comparison for the $3\times$ scenario: (a) the full x - z plane with the ROI outlined in red, its (b) high- and (c) low-resolution references, and the reconstructions by (d) ESRGAN, (e) SRGAN, (f) SRCNN, (g) bicubic, (h) bilinear, and (i) nearest-neighbor methods; SSIM and RMSE are annotated on each panel.

Fig. 3 (d–i) present the results obtained by applying different super-resolution techniques. These include the ESRGAN, two other learning-based methods (the SRGAN and the SRCNN), and three conventional interpolation methods. The ESRGAN-generated image not only shows increased clarity but also recovers fine details that were absent in the LR image. Specifically, thin, elongated ferrite plates and subtle pearlite arrangements are reconstructed more faithfully, closely matching the structure found in the HR reference, indicating that the ESRGAN captures complex textures and patterns most faithfully among all methods. The SRGAN restores relatively sharp ferrite and pearlite boundaries, yet its textures are less consistent with the HR reference, in line with the artifacts. The SRCNN produces more stable reconstructions than the interpolation methods but fails to recover the finest high-frequency features, and the fine pearlite is broken into isolated dark specks arranged on a regular grid rather than the continuous plates seen in the HR reference. In contrast, the bicubic, bilinear, and nearest-neighbor methods provide only some improvement, often producing overly smooth or blocky areas and losing subtle structural features.

In addition to the visual comparison, we evaluated the reconstructions quantitatively. We measured the structural similarity index measure (SSIM) and the root-mean-square error (RMSE) between the super-resolved images and the HR reference. The ESRGAN yields the highest SSIM score, indicating that its output most closely mimics the HR structure, and it also achieves the lowest RMSE value, indicating fewer numerical differences from the HR reference. These metrics confirm that the advantage seen visually also holds in a strict quantitative sense, and that the network reconstructs important microstructural details more reliably than the conventional interpolation methods.

To gain a more comprehensive understanding, we also compared super-resolved images of the entire x - z plane, not just a cropped region. As shown in Fig. 4, LR, HR, and super-resolved images obtained by the various methods are compared over the full field. This

larger view helps us ensure that the ESRGAN does not only excel in small, selected areas but can maintain its advantages over a broader region. The results in Fig. 4 clearly show that the ESRGAN-generated entire plane image retains the clarity and detail recovery observed in the cropped region. Thin and long ferrite plates, indicated by red arrows, are not properly restored by conventional methods, as seen in Fig. 4 (d–f). These methods tend to lose important features or blur them. However, the ESRGAN output (Fig. 4 (c)) recovers these fine structures with clarity, showing that it can consistently handle complex microstructures even over large areas.

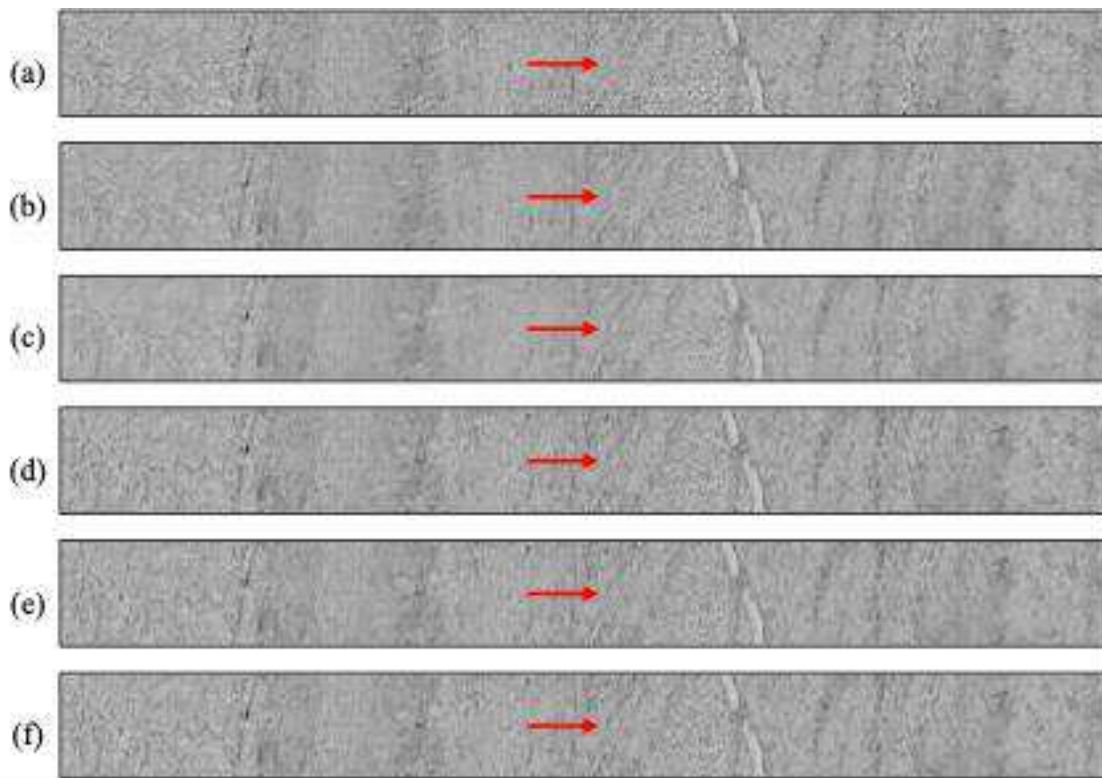


Figure 4. Full-field reconstruction of the entire x - z plane at $3\times$: (a) low resolution, (b) high resolution, and the outputs of (c) ESRGAN, (d) bicubic, (e) bilinear, and (f) nearest-neighbor.

So far, we have primarily focused on cases where the super-resolution factor was set to 3, showing that the ESRGAN effectively reconstructs the microstructure at this scale. However, we need a more thorough evaluation of the ESRGAN's capabilities in various conditions. To address this, we also examined two other scenarios. Specifically, Fig. 5 shows how the

ESRGAN performs at a 2 times upsampling factor, allowing us to confirm if it can also maintain clarity and detail even when the scaling is less extreme. In Fig. 5, we compare the original LR images, the ground truth HR images, and results from the various methods. At this smaller upsampling factor, the ESRGAN reconstruction in Fig. 5(c) recovers the fine features most accurately and produces a clearer, sharper image that closely matches the HR reference. The two other learning-based baselines in Fig. 5 (d,e) recover more detail than the interpolation methods but remain less faithful than the ESRGAN. In contrast, the conventional methods in Fig. 5 (f-h) reveal some finer details in the microstructure, but their outputs look slightly blurry and rough, indicating that they struggle to cleanly recover all intricate features. This suggests that even at a smaller upsampling factor, the ESRGAN offers considerable improvements over both the other learning-based baselines and the traditional interpolation approaches.

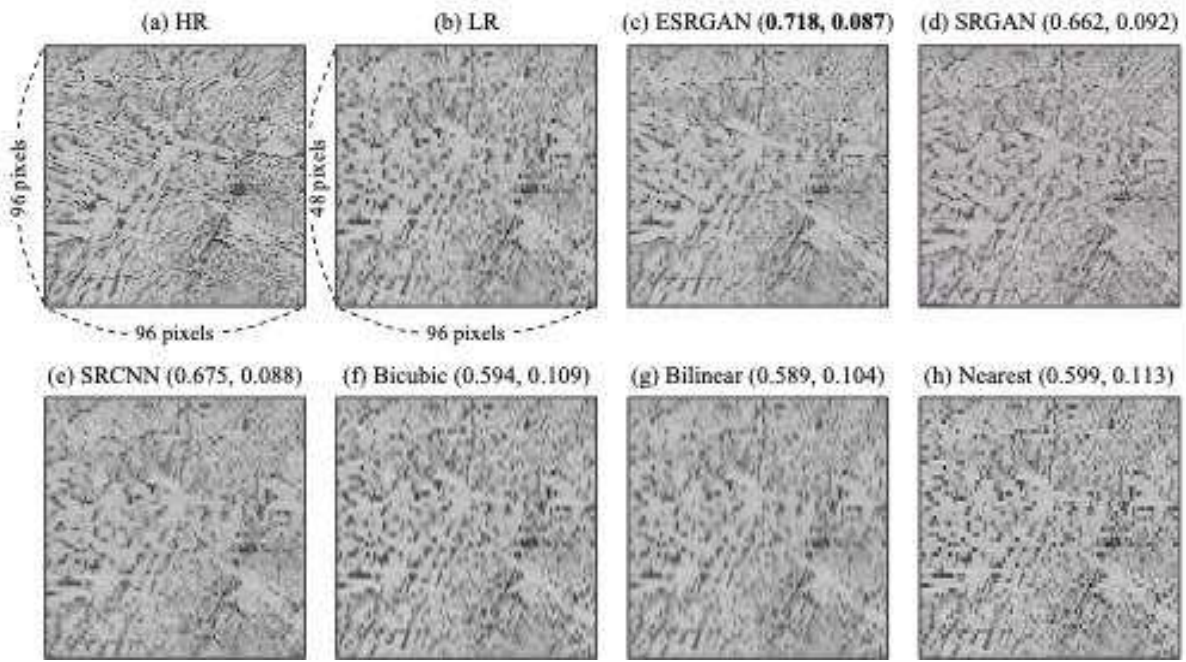


Figure 5. Reconstruction quality at an upsampling factor of two. (a) HR, (b) LR, and (c-h) two-times super-resolved images obtained by the ESRGAN, SRGAN, SRCNN, bicubic, bilinear, and nearest-neighbor methods, respectively, along with their SSIM and RMSE indices.

The similar pattern emerges in the more difficult situation shown in Fig. 6, where the upsampling factor is increased to 4. This scale is particularly challenging because only one in every four depth-wise slices is retained, so the network must infer roughly three quarters of the through-plane information from a highly sparse low-resolution input. As the upsampling factor grows, the reconstruction becomes an increasingly ill-posed problem, since many different high-resolution structures are consistent with the same low-resolution image. Under this condition, the ESRGAN (Fig. 6 (c)) still recovers the thick, elongated ferrite plates near the image center, which none of the other methods reproduce. This demonstrates that the ESRGAN remains the best of the compared methods not only in moderate but also in quite demanding scenarios, maintaining its advantage as the complexity and difficulty of the task grow. Furthermore, Fig. 7 expands our understanding by comparing images at various upsampling factors for the same area of interest. Naturally, when we push the scaling to high factors, the quality of recovered images declines. Despite this inevitable decrease, the ESRGAN still manages to restore microstructural features, avoiding a complete collapse into blur or roughness.

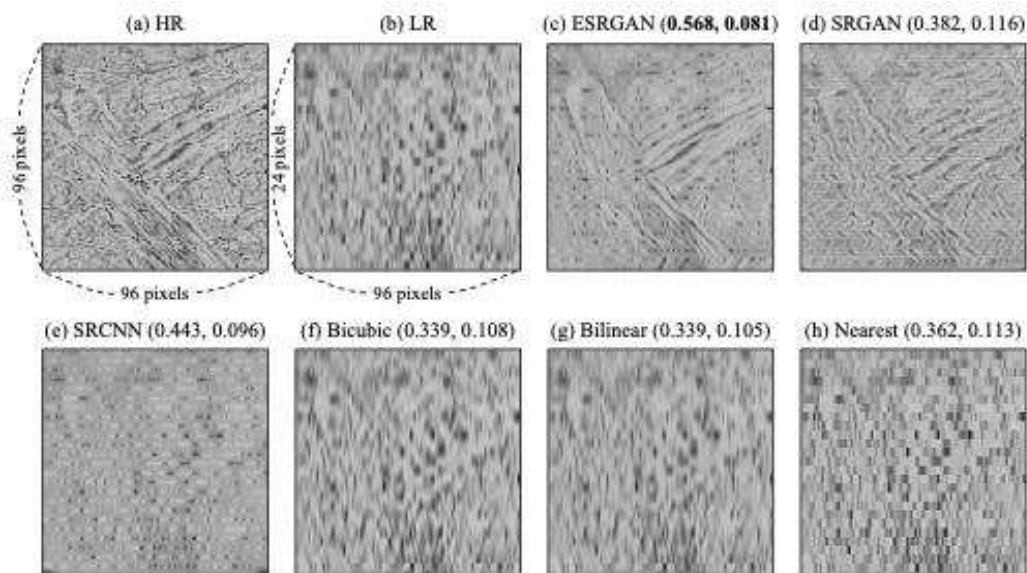
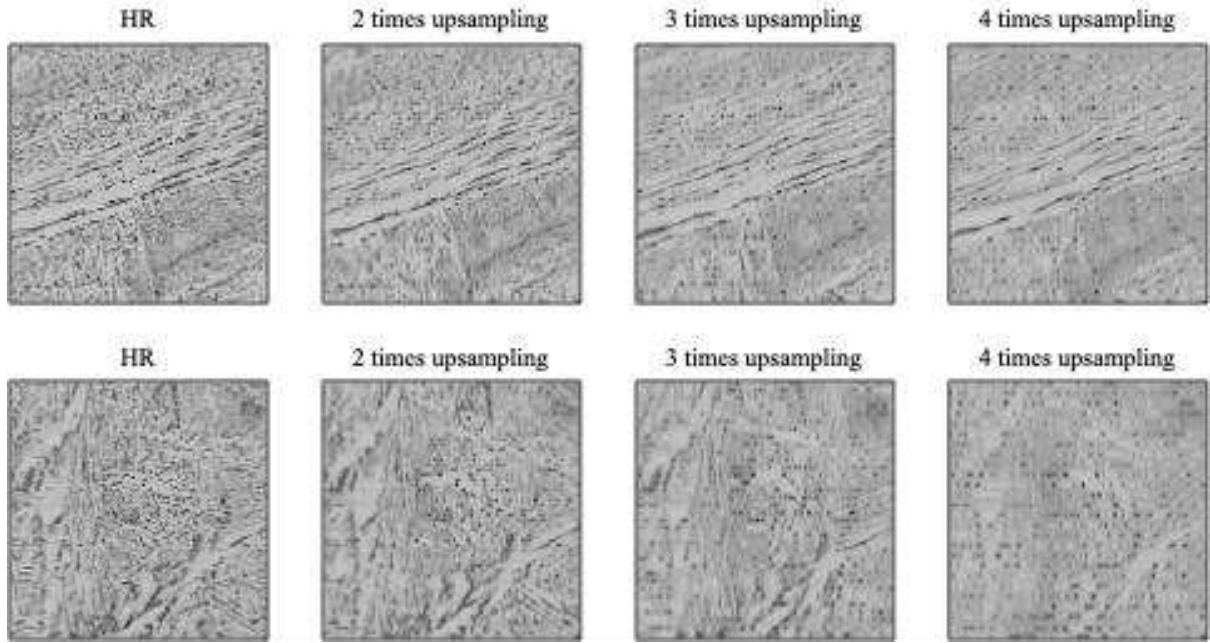


Figure 6. Behavior under the most demanding $4\times$ setting: reconstructions of the ROI by the ESRGAN, SRGAN, and SRCNN networks alongside the bicubic, bilinear, and nearest-neighbor

282 *interpolants (c–h), preceded by the (a) high- and (b) low-resolution images, each labeled with*
 283 *its SSIM and RMSE.*



284
 285 *Figure 7. Effect of the upsampling factor ($2\times$, $3\times$, $4\times$) on reconstruction quality for the same*
 286 *region of interest.*

287
 288 Fig. 8 compares how closely the reconstructions reproduce the high-resolution (HR)
 289 3D image across upsampling factors. The ESRGAN attains the highest SSIM (Fig. 8 (a)) and
 290 the lowest RMSE (Fig. 8(b)) at every upsampling factor, so its reconstructions are closest to
 291 the HR image in both overall structure and pixel-level accuracy. The SRGAN and the SRCNN
 292 fall between the ESRGAN and the interpolation methods on both measures, while the three
 293 interpolation methods remain consistently lowest and do not reach the same level of fine detail.

294 Since the network processed only x - z sections, we viewed the y - z plane to test whether
 295 the reconstruction remains coherent in a direction the network never saw. Fig. 9 shows a
 296 representative y - z cross-section of the same reconstructed volume, with a region of interest
 297 outlined in red for side-by-side comparison of the LR, HR, and three-times super-resolved
 298 images. In these sections the interpolation methods produce blurred, indistinct grain boundaries,

so that thin pearlite and fine ferrite plates are no longer resolved. The ESRGAN recovers the shapes and boundaries of these constituents more faithfully, although not every feature is restored.

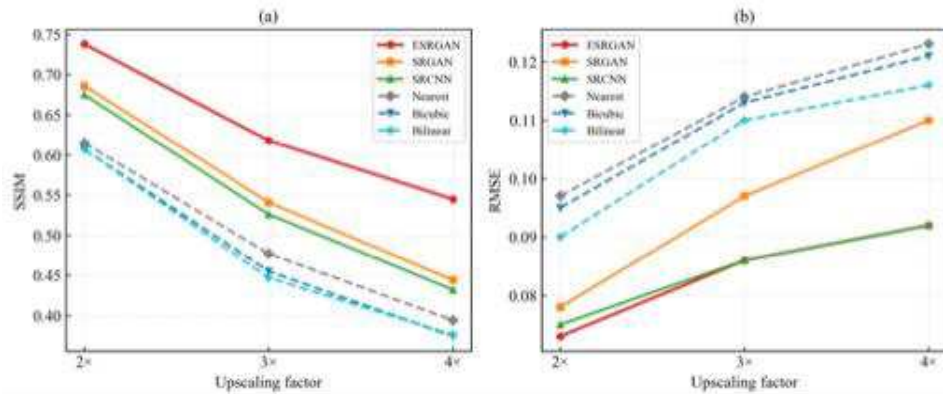


Figure 8. Pixel-wise agreement with the high-resolution reference across upsampling factors. (a) SSIM and (b) RMSE of the ESRGAN, SRGAN, SRCNN, and interpolation reconstructions relative to the HR volume, computed over the entire 3D image at each upsampling factor.

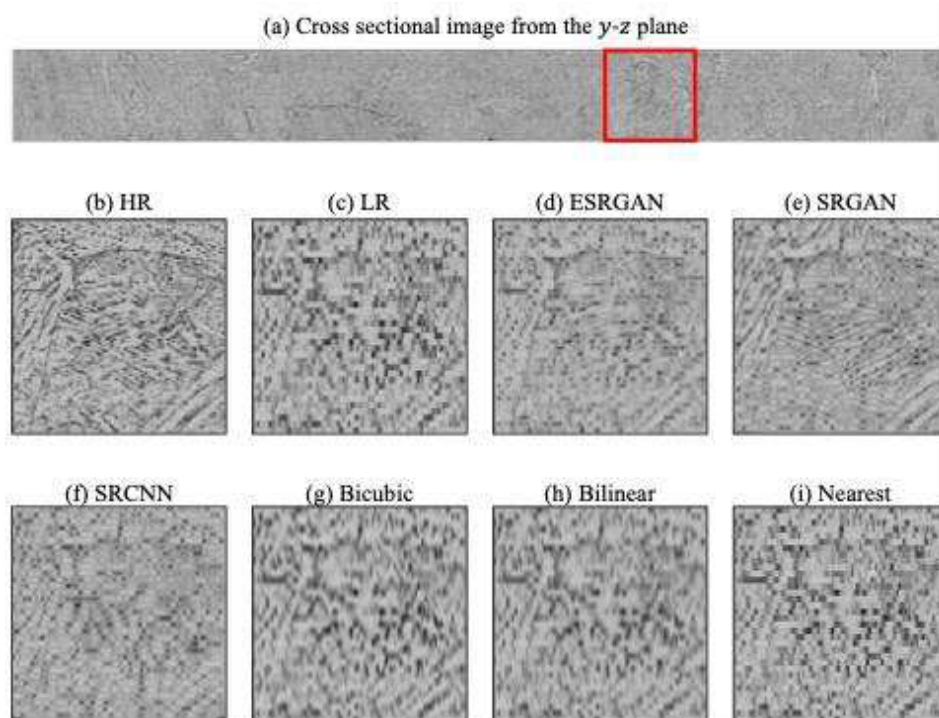
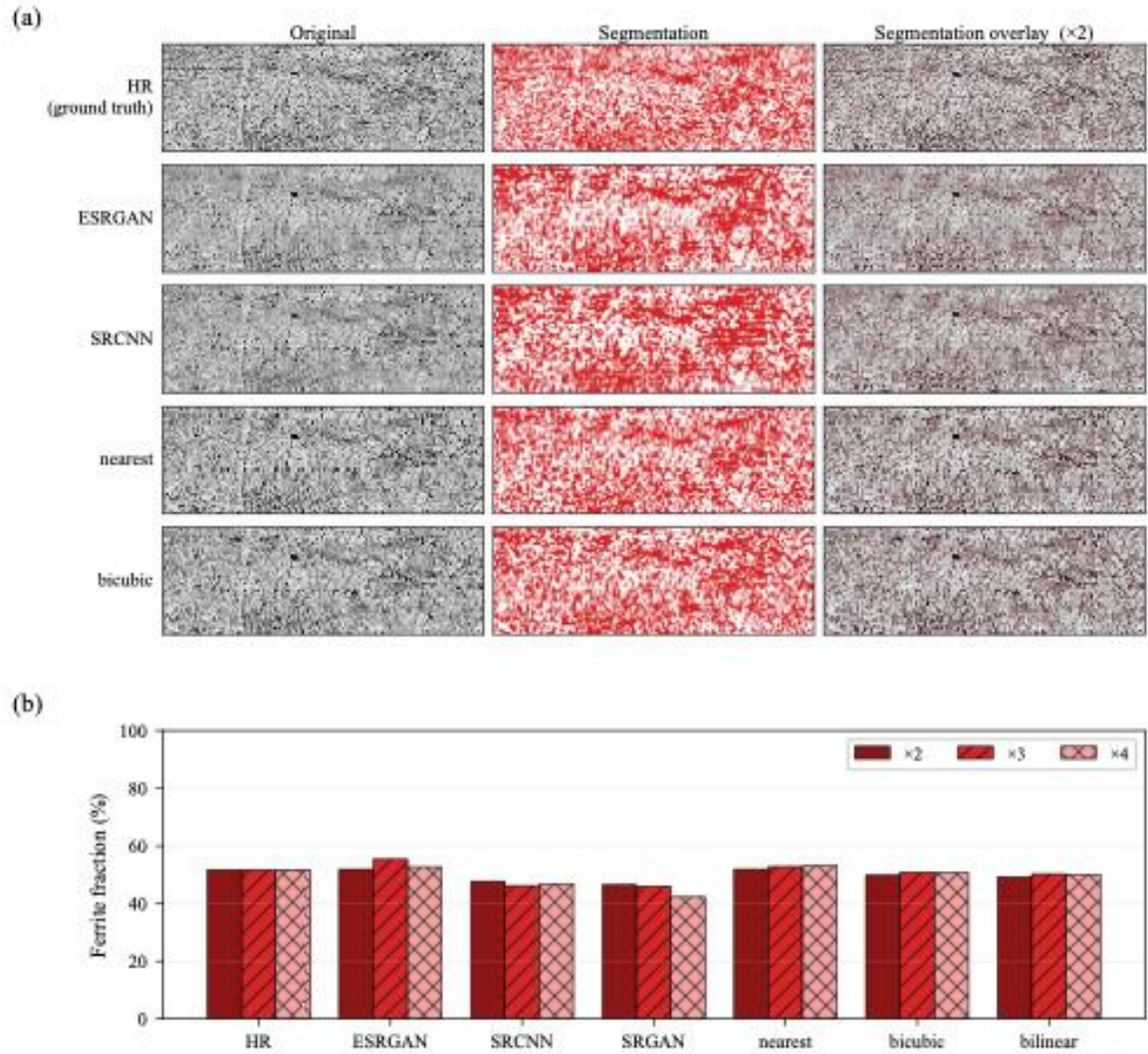


Figure 9. Reconstruction in a different sectioning orientation. (a) Representative y-z slice with the region of interest outlined in red, shown against its (b) high and (c) low resolution

references, followed by the depth super-resolved reconstructions from (d) ESRGAN, (e) SRGAN, (f) SRCNN, (g) bicubic, (h) bilinear, and (i) nearest-neighbor methods.

Fig. 10 examines whether the super-resolved volumes reproduce the phase information of the real microstructure. Here, a Gaussian mixture model (GMM) was fitted to the intensity histogram of the ground-truth volume to segment the ferrite from the non-ferrite constituents. Fig. 10(a) presents a representative x - z plane for the 2 times upsampling scenario along with its segmentation result and the overlaid image, in which the ferrite and non-ferrite constituents are highlighted in white and red, respectively. The ferrite regions are clearly distinguished from the non-ferrite constituents, and the segmentation of the HR image resolves thin, sharply bounded non-ferrite bands along the depth direction. The learning-based methods retain these features, whereas the interpolation-based methods reproduce blocky and vertically elongated features whose boundaries are no longer resolved. Fig. 10(b) quantifies the ferrite fraction for all upsampling factors. The ferrite fraction of the ESRGAN differs from that of the ground truth by less than 1 % at the 2 and 3 times scenarios and by approximately 3 % at the 4 times scenario, whereas the difference for the other learning-based methods grows rapidly with the upsampling factor, reaching approximately 7 % for the SRCNN and 9 % for the SRGAN at the 4 times scenario. Notably, the values obtained from the interpolation-based methods are close to the ground truth. This is because these methods construct every reconstructed slice as a weighted average of the retained HR slices so that the global intensity distribution is preserved by construction even though the depth-wise structure is severely smeared. The ferrite fraction therefore provides a necessary but not a complete measure of reconstruction quality, and is best interpreted together with statistical descriptors that capture the spatial arrangement of the

331 microstructures.



332

333 *Figure 10. Assessment of phase-fraction fidelity. (a) Representative examples of the*
 334 *microstructure segmentation and (b) the ferrite fraction estimated from the segmentation*
 335 *results of each method.*

336

337

338

339

340

341

To assess how each method reproduces the through-plane (or z-axis) microstructure beyond the pixel-wise metrics, we analyzed the reconstructed volumes using the two-point autocorrelation method. Fig. 11(a) compares the through-plane autocorrelation of the HR image and the super-resolved results at the 3 times upsampling scenario. The HR curve decays rapidly, reflecting the fine depth-wise variation of the real microstructure. On the other hand,

the interpolation methods decay more slowly, indicating excessive smoothing along the depth direction. The learning-based methods follow the HR curve closely and the ESRGAN reproduces the rapid decorrelation that characterizes the true microstructure. Fig. 11(b) summarizes this as the RMSE of each method's through-plane autocorrelation from the HR image, which is smallest for the learning-based methods. Fig. 11(c) extends the analysis across upsampling factors using the anisotropy index, defined as the ratio of the through-plane to in-plane correlation length (L_z/L_{in}). For the interpolation methods, this index deviates progressively from the HR image with increasing upsampling factors, reflecting elongated depth structure. However, the learning-based methods remain close to HR at every factor. These results confirm that the ESRGAN preserves the through-plane microstructural statistics that conventional interpolation methods fail to recover.

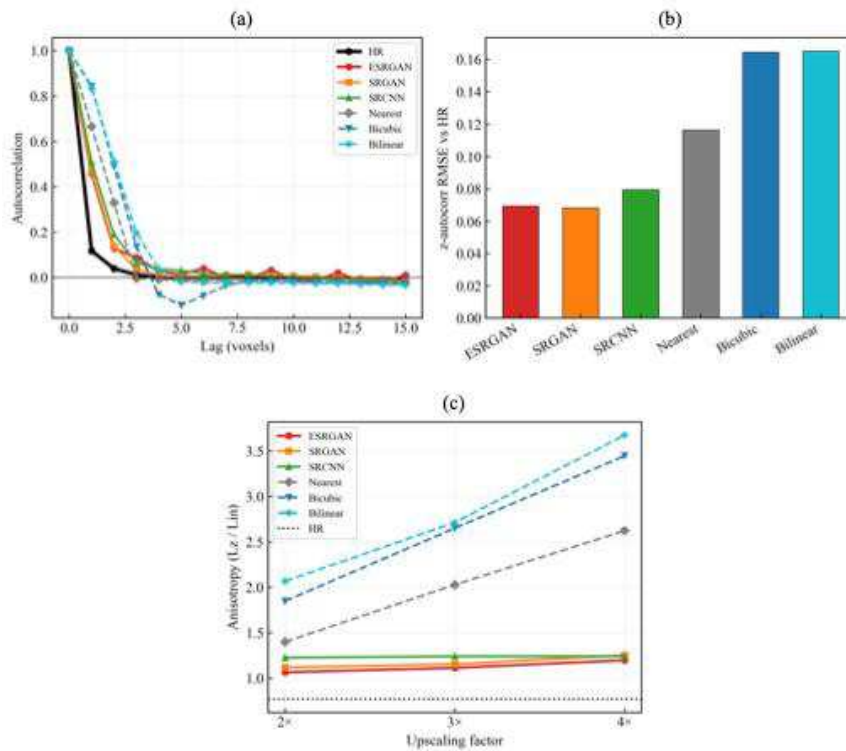


Figure 11. Morphological fidelity assessed by two-point statistics. (a) z directional two-point autocorrelation of the high-resolution and super-resolved 3D images, (b) root-mean-square

error of each method's autocorrelation from the high-resolution reference, and (c) anisotropy index (L_z/L_{in}) across upsampling factors.

Conclusions

In this study, we investigated the feasibility of employing the ESRGAN to enhance the depth-wise resolution of a 3D microstructural image of a low-carbon steel without the need for additional training images. Using only the provided 3D image, the network was trained with x - y plane images to learn microstructural details and then applied to recover details in x - z plane images, and its results were compared with those of an SRCNN, a vanilla SRGAN and three interpolation methods. Regarding image quality, the ESRGAN restored fine and directional microstructural details most faithfully among the compared methods, and it gave the highest SSIM and the lowest RMSE at every upsampling factor. Beyond these pixel-wise metrics, the ESRGAN reproduced the segmented ferrite fraction most accurately among the learning-based methods, and its through-plane two-point correlation descriptors remained closest to the HR reference, whereas those of the interpolation methods deviated increasingly as the upsampling factor grew. Consequently, the ESRGAN provides the most faithful reconstruction, among the compared methods, of a 3D image with anisotropic resolution using a single image.

Methods

First, we achieved a full 3D microstructural image with uniform resolution along the x , y , and z axes, ensuring high accuracy and consistency for subsequent analysis. This image will serve as the foundation for both training models and validating their performance. To accomplish this, we began by preparing a low-carbon steel sample with the specific composition outlined in **Table 1**. The sample underwent an austenitization process at the

temperature of 1400 °C for 5 s, followed by continuous cooling at a rate of 3 °C/s to achieve the desired microstructural characteristics.

Serial sectioning was carried out using an automated methodology that combined several precise and repeatable steps. First, chemical polishing was employed to ensure a smooth and uniform surface for imaging. This was followed by etching with 2% Picral solution to enhance the visibility of microstructural features under optical microscopy. An optical microscope was then used to capture high-resolution images of each prepared cross-section. To ensure that the sectioning process produced layers of consistent thickness, we utilized indentation marks placed on the sample's surface. These marks served as reference points, and sectioning continued layer by layer until the disappearance of these marks confirmed the desired depth. The cross-sectional images obtained through this process were subsequently processed and aligned. Using the software ImageJ [50], we stacked the images in sequential order to reconstruct the 3D microstructural image.

To simulate various anisotropic resolution scenarios typically caused by serial sectioning-based methods, the resolution of the z -axis or the depth direction of the ground truth 3D image was reduced by choosing every n^{th} x - y plane, where n is set as 2, 3, and 4 for different scenarios, as shown in Fig. 12. Consequently, the depth-wise resolution was cut down to half, a third, and a quarter of the original resolution, respectively. From each anisotropic-resolution 3D image, x - y cross-sectional images were sliced into patches of 96×96 pixels, serving as HR images. Corresponding LR images were generated by downsampling the HR x - y plane slices according to the downsampling factor of each scenario. This LR-HR paired dataset from the x - y planes is referred to as the training dataset in this study. Additionally, a validation dataset was prepared to assess the applicability of the proposed framework. For each scenario, validation datasets were also created by slicing x - z cross-sectional images into patches of 96×48 , 96×32 , and 96×24 pixels, respectively. These validation images were

then used as inputs to the trained ESRGAN model to generate SR images.

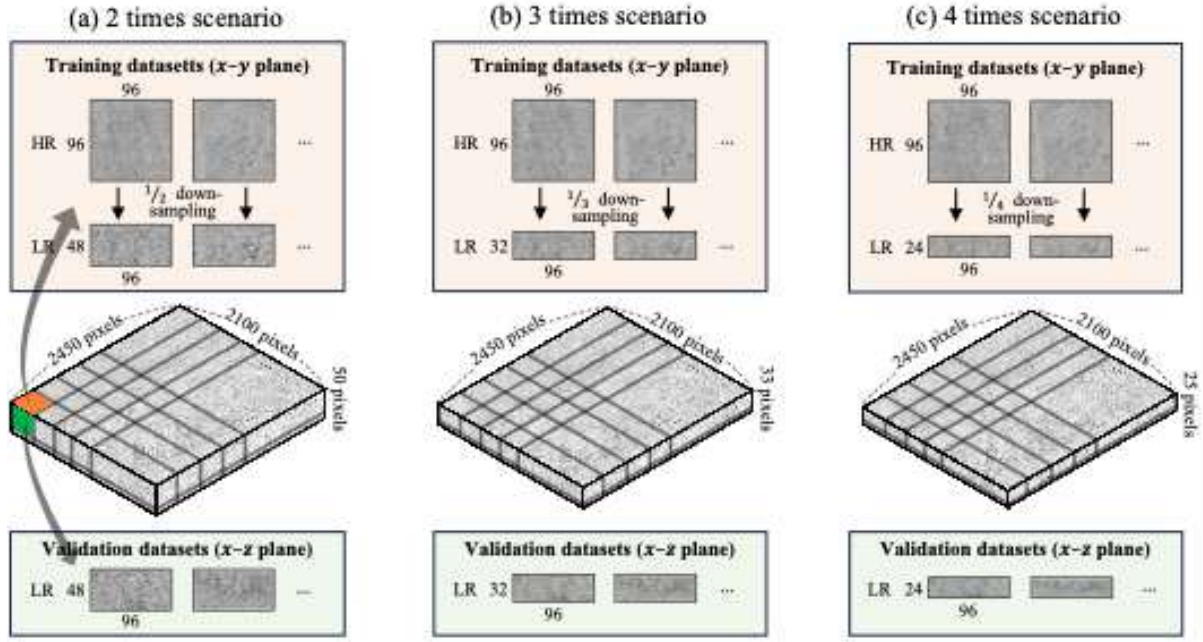


Figure 12. Construction of the training and validation datasets for the (a) $2\times$, (b) $3\times$, and (c) $4\times$ super-resolution scenarios, in which low-resolution slices are produced from the ground-truth planes at each downsampling ratio.

The architecture of the ESRGAN consists of two networks as described in Fig. 13: a generator and a discriminator. A generator takes an LR image as input and generates an SR image that resembles an HR image. The general role of a generator in super-resolution tasks is to increase the resolution by an upsampling factor. Our generator is made to increase only the resolution of the height of an input image because our goal is to increase the z-axis resolution of the LR image from the x - z plane. The generator comprises convolutional layers and residual in residual dense blocks (RRDBs), followed by upsampling layers. Unlike conventional SRGANs with residual blocks (RBs), ESRGAN with RRDBs removes batch normalization from RBs to prevent artifacts in super-resolved images and to considerably enhance image quality [42]. The discriminator, which evaluates whether an input image is HR or SR, is composed of multiple convolutional, batch normalization, and Leaky ReLU layers followed by

dense layers. The loss for the discriminator is the adversarial loss with the concept of the relativistic average discriminator [51] which estimates the probability that a real image is relatively more realistic than a fake image. The network was trained with the Adam optimizer [52] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the batch size was set to 64. The generator and the discriminator were alternately updated until the model converged.

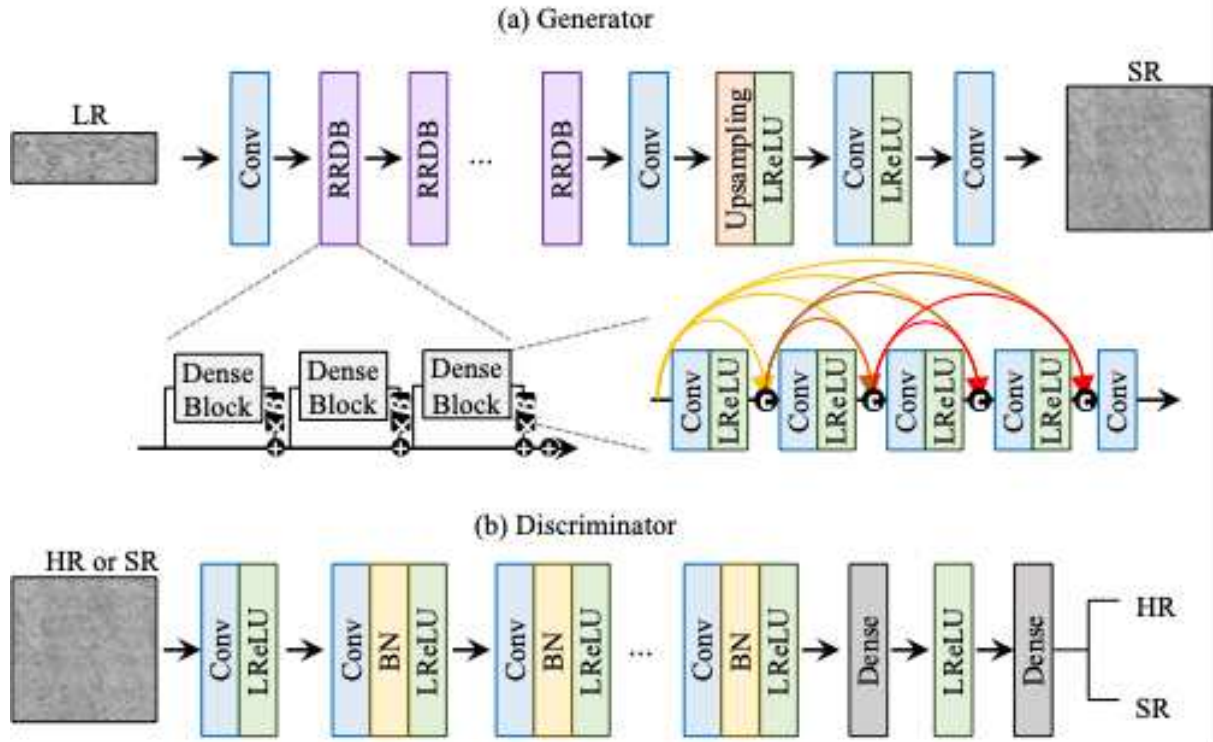


Figure 13. Network architecture of the ESRGAN based on residual-in-residual dense blocks (RRDBs): (a) generator and (b) discriminator.

After the training was finished, we super-resolved the x - z plane images with anisotropic resolution. The results were quantitatively evaluated using pixel-wise RMSE and SSIM[53] against the corresponding ROI in the ground truth 3D image. The RMSE quantifies the average difference between an SR image and an HR image, while the SSIM measures similarity by considering luminance, contrast, and structural information. In short, RMSE focuses on pixel-wise differences between images, making it intuitive, whereas SSIM captures structural similarities in addition to pixel-wise differences, rendering it more abstract. In order

to compare the performance of the ESRGAN with widely used image interpolation methods, the RMSE and the SSIM of bicubic, bilinear, and nearest-neighbor methods are also estimated. In addition to the interpolation methods, two learning-based baselines, a conventional SRGAN and an SRCNN, were also evaluated for comparison.

To complement these pixel-wise metrics, we additionally quantified how faithfully each method reproduces the microstructural morphology using statistical descriptors derived from the two-point autocorrelation [54]. From each autocorrelation curve, a characteristic feature size, referred to as the correlation length, was defined as the lag at which the autocorrelation first decays to half of its maximum, obtained by linear interpolation; the correlation lengths measured along the z and x directions are denoted the through-plane correlation length L_z and the in-plane correlation length L_{in} , respectively. These were then used to define the anisotropy index as their ratio (L_z/L_{in}), which quantifies the directional bias of the reconstructed microstructure. In addition, the deviation of each method's through-plane autocorrelation from that of the HR image was measured as the RMSE between the two curves. These comparisons provide a quantitative measure of microstructural fidelity that complements SSIM and RMSE by assessing the morphological statistics along the depth direction.

Data Availability

The datasets and the code used during the current study are available from the corresponding author upon reasonable request.

References

1. Bhadeshia, H. & Honeycombe, R. *Steels: Microstructure and Properties*. (Butterworth-Heinemann, 2017).
2. Bhadeshia, H. K. D. H. & Christian, J. W. Bainite in steels. *Metall. Trans. A* **21**, 767–797 (1990).
3. Krauss, G. *Steels: Processing, Structure, and Performance*. (Asm International, 2015).
4. Krauss, G. Martensite in steel: strength and structure. *Mater. Sci. Eng. A* **273**, 40–57 (1999).
5. De Cooman, B. C., Estrin, Y. & Kim, S. K. Twinning-induced plasticity (TWIP) steels. *Acta Mater.* **142**, 283–362 (2018).
6. Gunn, R. *Duplex Stainless Steels: Microstructure, Properties and Applications*. (Woodhead Publishing, 1997).
7. Soleimani, M., Kalhor, A. & Mirzadeh, H. Transformation-induced plasticity (TRIP) in advanced steels: A review. *Mater. Sci. Eng. A* **795**, 140023 (2020).
8. Tasan, C. C. *et al.* Strain localization and damage in dual phase steels investigated by coupled in-situ deformation experiments and crystal plasticity simulations. *Int. J. Plast.* **63**, 198–210 (2014).
9. Lewis, A. C. *et al.* Two- and three-dimensional microstructural characterization of a super-austenitic stainless steel. *Mater. Sci. Eng. A* **418**, 11–18 (2006).
10. Morito, S., Adachi, Y. & Ohba, T. Morphology and crystallography of sub-blocks in ultra-low carbon lath martensite steel. *Mater. Trans.* **50**, 1919–1923 (2009).
11. Ramazani, A., Mukherjee, K., Quade, H., Prahl, U. & Bleck, W. Correlation between 2D and 3D flow curve modelling of DP steels using a microstructure-based RVE approach. *Mater. Sci. Eng. A* **560**, 129–139 (2013).

- 481 12. Cecen, A., Dai, H., Yabansu, Y. C., Kalidindi, S. R. & Song, L. Material structure-property
482 linkages using three-dimensional convolutional neural networks. *Acta Mater.* **146**, 76–84
483 (2018).
- 484 13. Chawla, N. & Deng, X. Microstructure and mechanical behavior of porous sintered steels.
485 *Mater. Sci. Eng. A* **390**, 98–112 (2005).
- 486 14. Kenesei, P., Borbély, A. & Biermann, H. Microstructure based three-dimensional finite
487 element modeling of particulate reinforced metal–matrix composites. *Mater. Sci. Eng. A*
488 **387–389**, 852–856 (2004).
- 489 15. Landis, E. N., Nagy, E. N. & Keane, D. T. Microstructure and fracture in three dimensions.
490 *Eng. Fract. Mech.* **70**, 911–925 (2003).
- 491 16. Maire, E. *et al.* On the Application of X-ray Microtomography in the Field of Materials
492 Science. *Adv. Eng. Mater.* **3**, 539–546 (2001).
- 493 17. Salvo, L., Suéry, M., Marmottant, A., Limodin, N. & Bernard, D. 3D imaging in material
494 science: Application of X-ray tomography. *Comptes Rendus Phys.* **11**, 641–649 (2010).
- 495 18. Shen, K., Lu, H., Baig, S. & Wang, M. R. Improving lateral resolution and image quality
496 of optical coherence tomography by the multi-frame superresolution technique for 3D
497 tissue imaging. *Biomed. Opt. Express* **8**, 4887–4918 (2017).
- 498 19. Echlin, M. P., Burnett, T. L., Polonsky, A. T., Pollock, T. M. & Withers, P. J. Serial
499 sectioning in the SEM for three dimensional materials science. *Curr. Opin. Solid State*
500 *Mater. Sci.* **24**, 100817 (2020).
- 501 20. Yamashita, N., Matsuno, T., Maeda, D., Kikuzuki, M. & Yokota, H. Efficient 3D
502 observation of steel microstructure using serial sectioning with precision cutting and on-
503 site etching. *Precis. Eng.* **75**, 37–45 (2022).
- 504 21. Zankel, A., Wagner, J. & Poelt, P. Serial sectioning methods for 3D investigations in
505 materials science. *Micron* **62**, 66–78 (2014).

22. Ganti, S. *et al.* Three-dimensional (3D) analysis of white etching bands (WEBs) in AISI M50 bearing steel using automated serial sectioning. *Mater. Charact.* **138**, 11–18 (2018).
23. Mosayebi, M., Paquet, D., Deschênes, P.-A., Tôm-Thât, L. & Bassim, N. Three-dimensional characterization of morphology and distribution of boundary networks in a low-carbon martensitic stainless steel. *Mater. Charact.* **229**, 115456 (2025).
24. Gambini, L. *et al.* Video frame interpolation neural network for 3D tomography across different length scales. *Nat. Commun.* **15**, 7962 (2024).
25. González-Ruiz, V., García-Ortiz, J. P., Fernández-Fernández, M. R. & Fernández, J. J. Optical flow driven interpolation for isotropic FIB-SEM reconstructions. *Comput. Methods Programs Biomed.* **221**, 106856 (2022).
26. Hagita, K., Higuchi, T. & Jinnai, H. Super-resolution for asymmetric resolution of FIB-SEM 3D imaging using AI with deep learning. *Sci. Rep.* **8**, 5877 (2018).
27. Roldán, D., Redenbach, C., Schladitz, K., Klingele, M. & Godehardt, M. Reconstructing porous structures from FIB-SEM image data: Optimizing sampling scheme and image processing. *Ultramicroscopy* **226**, 113291 (2021).
28. Sardhara, T. *et al.* Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials. *Ultramicroscopy* **256**, 113878 (2024).
29. Kral, M. V., Mangan, M. A., Spanos, G. & Rosenberg, R. O. Three-dimensional analysis of microstructures. *Mater. Charact.* **45**, 17–23 (2000).
30. Wolfsdorf, T. L., Bender, W. H. & Voorhees, P. W. The morphology of high volume fraction solid-liquid mixtures: An application of microstructural tomography. *Acta Mater.* **45**, 2279–2295 (1997).
31. Rhines, F. N., Craig, K. R. & Rousse, D. A. Measurement of average grain volume and certain topological parameters by serial section analysis. *Metall. Trans. A* **7**, 1729–1734 (1976).

32. Weiland, H., Rouns, T. N. & Liu, J. The Role of Particle Stimulated Nucleation during Recrystallization of an Aluminum—Manganese Alloy / Die Bedeutung von Keimbildung an Teilchen einer zweiten Phase während der Rekristallisation einer Aluminium—Mangan-Legierung: Dedicated to Professor Dr. rer. nat. Dr. h.c. Hans-Joachim Bunge on the occasion of his 65th birthday. *Int. J. Mater. Res.* **85**, 592–597 (1994).
33. Wang, L. *et al.* Construction of dynamic three-dimensional microstructure for the hydration of cement using 3D image registration. *Pattern Anal. Appl.* **17**, 655–665 (2014).
34. Wang, Y. D., Armstrong, R. T. & Mostaghimi, P. Boosting resolution and recovering texture of 2D and 3D micro-CT images with deep learning. *Water Resour. Res.* **56**, e2019WR026052 (2020).
35. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **29**, 1153–1160 (1981).
36. Press, W. H. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. (Cambridge university press, 2007).
37. Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**, 21–27 (1967).
38. Park, S. C., Park, M. K. & Kang, M. G. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Mag.* **20**, 21–36 (2003).
39. Dong, C., Loy, C. C., He, K. & Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *ArXiv150100092 Cs* <http://arxiv.org/abs/1501.00092> (2015).
40. Ledig, C. *et al.* Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 105–114 (IEEE, Honolulu, HI, 2017). doi:10.1109/CVPR.2017.19.

41. Jung, J. *et al.* Super-resolving material microstructure image via deep learning for microstructure characterization and mechanical behavior analysis. *Npj Comput. Mater.* **7**, 1–11 (2021).
42. Wang, X. *et al.* ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. in *Computer Vision – ECCV 2018 Workshops* (eds Leal-Taixé, L. & Roth, S.) vol. 11133 63–79 (Springer International Publishing, Cham, 2019).
43. Wang, X., Xie, L., Dong, C. & Shan, Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. Preprint at <https://doi.org/10.48550/arXiv.2107.10833> (2021).
44. Zhang, W., Liu, Y., Dong, C. & Qiao, Y. RankSRGAN: Generative Adversarial Networks With Ranker for Image Super-Resolution. in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 3096–3105 (IEEE, Seoul, Korea (South), 2019). doi:10.1109/ICCV.2019.00319.
45. Freedman, G. & Fattal, R. Image and video upscaling from local self-examples. *ACM Trans. Graph.* **30**, 1–11 (2011).
46. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Deep image prior. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 9446–9454 (2018).
47. Shaham, T. R., Dekel, T. & Michaeli, T. SinGAN: Learning a Generative Model From a Single Natural Image. in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 4569–4579 (IEEE, Seoul, Korea (South), 2019). doi:10.1109/ICCV.2019.00467.
48. Park, H. *et al.* Deep learning enables reference-free isotropic super-resolution for volumetric fluorescence microscopy. *Nat. Commun.* **13**, 3297 (2022).
49. Gao, X. *et al.* Enhancing scanning electron microscopy imaging quality of weakly conductive samples through unsupervised learning. *Sci. Rep.* **14**, 6439 (2024).

50. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
51. Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GAN. Preprint at <https://doi.org/10.48550/arXiv.1807.00734> (2018).
52. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. Preprint at <https://doi.org/10.48550/arXiv.1412.6980> (2017).
53. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
54. Torquato, S. *Random Heterogeneous Materials*. vol. 16 (Springer New York, New York, NY, 2002).

Acknowledgements

This study was funded by the Council for Science, Technology and Innovation, Cross-ministerial Strategic Innovation Promotion Program (SIP), “Structural Materials for Innovation” (funding agency: JST). The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

594 **Author contributions**

595 Hoheok Kim conceived the study, performed experiments and the image reconstruction,
596 implemented and trained the networks, carried out the quantitative analysis, and wrote the
597 original draft. Junya Inoue supervised the work, acquired the funding, provided the resources,
598 administered the project, contributed to the investigation, and reviewed and edited the
599 manuscript. Both authors read and approved the final manuscript.

600 **Additional Information**

601 Competing interests: The authors declare no competing interests.

602

Figure captions

Fig. 1. Overview of the proposed ESRGAN framework for enhancing the depth-wise resolution of a 3D microstructure of low-carbon steel.

Fig. 2. Isotropic reference volume obtained by serial sectioning. (a) Ground-truth 3D microstructure of the low-carbon steel, shown together with its orthogonal cross-sections along the (b) x - y , (c) x - z , and (d) y - z planes.

Fig. 3. Region-of-interest comparison for the $3\times$ scenario: (a) the full x - z plane with the ROI outlined in red, its (b) high- and (c) low-resolution references, and the reconstructions by (d) ESRGAN, (e) SRGAN, (f) SRCNN, (g) bicubic, (h) bilinear, and (i) nearest-neighbor methods; SSIM and RMSE are annotated on each panel.

Fig. 4. Full-field reconstruction of the entire x - z plane at $3\times$: (a) low resolution, (b) high resolution, and the outputs of (c) ESRGAN, (d) bicubic, (e) bilinear, and (f) nearest-neighbor.

Fig. 5. Reconstruction quality at an upsampling factor of two. (a) HR, (b) LR, and (c–h) two-times super-resolved images obtained by the ESRGAN, SRGAN, SRCNN, bicubic, bilinear, and nearest-neighbor methods, respectively, along with their SSIM and RMSE indices.

Fig. 6. Behavior under the most demanding $4\times$ setting: reconstructions of the ROI by the ESRGAN, SRGAN, and SRCNN networks alongside the bicubic, bilinear, and nearest-neighbor interpolants (c–h), preceded by the (a) high- and (b) low-resolution images, each labeled with its SSIM and RMSE.

Fig. 7. Effect of the upsampling factor ($2\times$, $3\times$, $4\times$) on reconstruction quality for the same region of interest.

Fig. 8. Pixel-wise agreement with the high-resolution reference across upsampling factors. (a) SSIM and (b) RMSE of the ESRGAN, SRGAN, SRCNN, and interpolation reconstructions relative to the HR volume, computed over the entire 3D image at each upsampling factor.

Fig. 9. Reconstruction in a different sectioning orientation. (a) Representative y - z slice with the region of interest outlined in red, shown against its (b) high and (c) low resolution references, followed by the depth super-resolved reconstructions from (d) ESRGAN, (e) SRGAN, (f) SRCNN, (g) bicubic, (h) bilinear, and (i) nearest-neighbor methods.

Fig. 10. Assessment of phase-fraction fidelity. (a) Representative examples of the microstructure segmentation and (b) the ferrite fraction estimated from the segmentation results of each method.

Fig. 11. Morphological fidelity assessed by two-point statistics. (a) z directional two-point autocorrelation of the high-resolution and super-resolved 3D images, (b) root-mean-square error of each method's autocorrelation from the high-resolution reference, and (c) anisotropy index (L_z/L_{in}) across upsampling factors.

Fig. 12. Construction of the training and validation datasets for the (a) $2\times$, (b) $3\times$, and (c) $4\times$ super-resolution scenarios, in which low-resolution slices are produced from the ground-truth planes at each downsampling ratio.

Fig. 13. Network architecture of the ESRGAN based on residual-in-residual dense blocks (RRDBs): (a) generator and (b) discriminator.

644 **Table caption**

645 **Table 1.** Alloying composition of the low-carbon steel used in this study.

C	Si	Mn	P	S	Al	N	O
0.152	0.015	1.51	0.007	0.0016	0.026	18 (ppm)	28 (ppm)

646