



OPEN ACCESS

EDITED BY

Erik Cuevas,
University of Guadalajara, Mexico

REVIEWED BY

Shaobing Gao,
Sichuan University, China
Hirak Mazumdar,
Adamas University, India

*CORRESPONDENCE

Zafran Hussain Shah
✉ zafran.shah@tu-dortmund.de

RECEIVED 11 December 2023

ACCEPTED 03 September 2024

PUBLISHED 23 September 2024

CITATION

Shah ZH, Müller M, Hübner W, Ortkrass H,
Hammer B, Huser T and Schenck W (2024)
Image restoration in frequency space using
complex-valued CNNs.
Front. Artif. Intell. 7:1353873.
doi: 10.3389/frai.2024.1353873

COPYRIGHT

© 2024 Shah, Müller, Hübner, Ortkrass,
Hammer, Huser and Schenck. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Image restoration in frequency space using complex-valued CNNs

Zafran Hussain Shah^{1*}, Marcel Müller², Wolfgang Hübner²,
Henning Ortkrass², Barbara Hammer³, Thomas Huser² and
Wolfram Schenck¹

¹Center for Applied Data Science, Faculty of Engineering and Mathematics, Bielefeld University of Applied Sciences and Arts, Bielefeld, Germany, ²Biomolecular Photonics Group, Faculty of Physics, Bielefeld University, Bielefeld, Germany, ³CITEC—Center for Cognitive Interaction Technology, Bielefeld University, Bielefeld, Germany

Real-valued convolutional neural networks (RV-CNNs) in the spatial domain have outperformed classical approaches in many image restoration tasks such as image denoising and super-resolution. Fourier analysis of the results produced by these spatial domain models reveals the limitations of these models in properly processing the full frequency spectrum. This lack of complete spectral information can result in missing textural and structural elements. To address this limitation, we explore the potential of complex-valued convolutional neural networks (CV-CNNs) for image restoration tasks. CV-CNNs have shown remarkable performance in tasks such as image classification and segmentation. However, CV-CNNs for image restoration problems in the frequency domain have not been fully investigated to address the aforementioned issues. Here, we propose several novel CV-CNN-based models equipped with complex-valued attention gates for image denoising and super-resolution in the frequency domains. We also show that our CV-CNN-based models outperform their real-valued counterparts for denoising super-resolution structured illumination microscopy (SR-SIM) and conventional image datasets. Furthermore, the experimental results show that our proposed CV-CNN-based models preserve the frequency spectrum better than their real-valued counterparts in the denoising task. Based on these findings, we conclude that CV-CNN-based methods provide a plausible and beneficial deep learning approach for image restoration in the frequency domain.

KEYWORDS

image restoration, image denoising, super-resolution, convolutional neural networks (CNNs), complex-valued convolutional neural networks (CV-CNNs), complex-valued attention gates, structured illumination microscopy, Fast Fourier Transform

1 Introduction

Convolutional neural networks (Krizhevsky et al., 2017) have become the leading state-of-the-art algorithms in the field of computer vision, ranging from image classification, pattern recognition, and object detection to tasks such as image restoration. Several image restoration problems, such as image denoising (Zhang et al., 2018; Li et al., 2020), reconstructing super-resolution (SR) image data (Dong et al., 2014), image inpainting (Xie et al., 2012), and JPEG compression artifact removal (Zheng et al., 2018) have been extensively investigated using CNN architectures. These CNN methods have demonstrated

substantial improvements in image restoration tasks over classical approaches (Lim et al., 2017; Gao et al., 2024; Zuo et al., 2018; Tran et al., 2020; Gu et al., 2012; He et al., 2010; Timofte et al., 2013, 2014). The vast majority of CNN based image restoration techniques have been developed by employing standard real-valued convolutional neural networks (RV-CNNs). RV-CNN based architectures are constrained to work with real-valued inputs, outputs, and parameters. In various disciplines such as biomedical engineering, physics, optics, radar and telecommunications, however, signals are recorded that are complex valued in their raw form (Schreier and Scharf, 2010; Barrachina et al., 2021). Complex representations of signals have real and imaginary components to e.g. carry magnitude and phase values that provide additional insights which are otherwise difficult to capture with purely real-valued signals (Foreman, 2012). Also, in the frequency domain, the contents of an image are represented as complex numbers (Xu et al., 2020).

Complex-valued convolutional neural networks (CV-CNNs) have recently been proposed to deal with such complex signals (Benvenuto and Piazza, 1992; Georgiou and Koutsougeras, 1992). CV-CNNs rely on complex inputs, network parameters, and the outputs are either complex or real values depending on the task (Lee et al., 2022). Since the real and imaginary components of a complex signal contain more information, the use of complex-valued networks is more crucial compared to its real-valued counterpart (Bassey et al., 2021). In image classification (Hafiz et al., 2015), segmentation (Akramifard et al., 2012), speech enhancement (Tsunami et al., 2013), and MRI fingerprint signal processing (Virtue et al., 2017), CV-CNNs have shown exceptional performance compared to their real counterparts. However, the performance of CV-CNNs in the frequency domain has not been extensively investigated in the context of image restoration problems such as image denoising and super-resolution image reconstruction. Quan et al. (2021) investigated the potential of CV-CNNs for image denoising using dummy imaginary values in the spatial domain. Rawat et al. (2021) denoised chest X-ray (CXR) images by using the complex-valued based method with residual learning. Similarly, Pham et al. (2021) evaluated the performance of a complex Fourier network for the image denoising problem by explicitly convolving a complex filter with the Fourier transform of the corrupted image. Nevertheless, CV-CNNs have never been directly employed for the denoising and reconstructing of super-resolution structured illumination microscopy (SR-SIM) and other conventional images in the frequency domain.

Deep learning algorithms based on real-valued networks have revolutionized the denoising and reconstruction of high-quality SR-SIM images. Structured illumination microscopy (SIM) is an extensively used super-resolution imaging technology due to its ability to double the spatial resolution for live cell imaging beyond the diffraction limit of light (Heintzmann and Huser, 2017; Schermelleh et al., 2019; Demmerle et al., 2017). SR-SIM technology is considered the most straightforward super-resolution (SR) reconstruction method among various other SR approaches since it requires fewer raw SIM samples with low illumination intensity levels (Ströhl and Kaminski, 2016; Zheng et al., 2021). However, the conventional SR-SIM reconstruction algorithms require high SNR raw SIM samples to reconstruct the high-quality SR-SIM

images (Huang et al., 2018). Consequently, the reconstruction algorithms are unable to produce high-quality SR-SIM images from raw samples acquired under low SNR conditions (Smith et al., 2021). Several RV-CNN based methods have been developed to denoise and reconstruct the SR-SIM images (Shah et al., 2021; Jin et al., 2020; Chen et al., 2021; Shah et al., 2022). Given that image restoration tasks are challenging and ill-posed problems, this leads to the existence of numerous viable possible solutions in the high-dimensional space during inference (Belthangady and Royer, 2019). Furthermore, the RV-CNNs follow the spectral bias within the Fourier spectrum, learning the low-frequency modes faster than the high-frequency modes (Rahaman et al., 2019). As a result of these concerns, it has been discovered that the SR-SIM images denoised and reconstructed using real-valued deep-learning approaches lose part of the high-frequency information (Shah et al., 2021; Qiao et al., 2021). The gap in the high-frequency information is highly visible in the Fourier spectra of the restored SR-SIM images. As a result, the disparity in the Fourier spectrum of the restored and reference images raises various concerns about the effectiveness of RV-CNNs-based image restoration approaches.

1.1 Motivation and contributions

The primary motivation behind this work is to retrieve a restored image that exhibits a Fourier spectrum closely resembling the Fourier spectrum of the reference or ground truth image. There has been very scarce research conducted in this particular context. Qiao et al. (2021) proposed a deep Fourier channel attention network (DFCAN) to overcome the frequency content difference across high-frequency information using Fourier channel attention (FCA) mechanism in the spatial domain. Similarly, Liu et al. (2023) investigated and minimized the gap in the high-frequency components by proposing a dual-domain learning strategy for the reconstruction of SIM images. However, their architecture requires repetitive transformation of features within each block, which impedes the exploration of the full potential of CV-CNNs. This prompted us to initiate additional research in this area, resulting in the formulation of two research questions to investigate this topic:

1. Is it possible to perform image denoising and super-resolution entirely in the frequency domain using CV-CNNs?
2. Do CV-CNNs outperform RV-CNNs in terms of generalization performance when applied to previously unseen SR-SIM modalities?

To address these research problems, we propose CV-CNN based image restoration algorithms for denoising and reconstruction of SR-SIM images in the frequency domain rather than the pure spatial domain. This work investigates the full capabilities of CV-CNNs for image restoration problems, intending to reduce the Fourier spectrum difference between the reconstructed and reference SR-SIM images. This work makes the following significant contributions:

- We explore the potential of CV-CNNs and suggest a novel complex-valued attention gate (C-AG).

- We propose two CV-CNN based architectures, named complex-valued collaborative attention network (CV-CAN) and complex-valued dual domain attention network (CV-DDAN), both equipped with complex-valued attention gates for the denoising task.
- Comparison of the proposed CV-CNN with existing RV-CNN architectures for the denoising and knowledge transfer tasks.
- A pure CV-CNN based architecture named complex-valued super-resolution network is designed for the super-resolution task in the frequency representation.

Moreover, we also demonstrate that our proposed complex-valued based denoising architecture CV-DDAN surpasses real-valued CNN based approaches visually as well as in terms of peak-signal-to-noise ratio (PSNR) and structural similarity index measurement (SSIM) values. Similarly, the Fourier spectrum of the resulting denoised images by CV-CNN based approaches is more promising than their real-valued counterparts. We test our proposed architectures on SR-SIM images as well as the commonly used BSD500 benchmark datasets. The results indicate that our proposed methods are not limited to super-resolution microscopic images, but also perform well-across other datasets.

2 Materials and methods

This section covers the key components of complex-valued neural networks and the Fast Fourier Transform algorithm used in this work. In addition, the mechanism of the incorporated attention gates is also discussed in this section. Finally, the publicly available datasets are also described in this section. The Red-fairSIM and UNet-fairSIM architectures have already been described in our prior work (Shah et al., 2021) and will not be reiterated here in detail. The Red-fairSIM and UNet-fairSIM methods are based on the combination of fairSIM (Müller et al., 2016) with the RedNet (Mao et al., 2016) and UNet (Ronneberger et al., 2015) architectures (Shah et al., 2021). Both RedNet and UNet methods are constructed using RV-CNNs (Shah et al., 2021).

2.1 Complex-valued neural network components

This subsection provides an overview of the complex components, including the complex-valued convolutional layer, activation function, and max pooling.

2.1.1 Complex-valued convolution layer

The complex number has both real and imaginary components in its representation. The complex-valued convolution (CV-Conv) layer have been designed to deal with complex values. The implementation of CV-Conv layer involves four real-valued convolution layers. In CV-Conv layer, a complex filter matrix $K = K_{\Re} + iK_{\Im}$ is convolved (*) with a complex input matrix $I = I_{\Re} + iI_{\Im}$ where the individual variables $K_{\Re}, K_{\Im}, I_{\Re}, I_{\Im}$ are real matrices, considering a vector can be represented as a matrix with one of two dimension being one. The \Re and \Im denote the real and

imaginary components. The convolution operator is distributive in the complex domain, hence the convolution of the kernel K with the input I is following:

$$K * I = (I_{\Re} * K_{\Re} - I_{\Im} * K_{\Im}) + i(I_{\Re} * K_{\Im} + I_{\Im} * K_{\Re}) \quad (1)$$

Figure 1 illustrates the complex-valued convolution operation. Similarly, according to Ding and Hirose (2013), we can write the matrix notations for the real and imaginary parts of the convolution operation as follows:

$$\begin{bmatrix} \Re(K * I) \\ \Im(K * I) \end{bmatrix} = \begin{bmatrix} K_{\Re} & -K_{\Im} \\ K_{\Im} & K_{\Re} \end{bmatrix} * \begin{bmatrix} I_{\Re} \\ I_{\Im} \end{bmatrix} \quad (2)$$

The implementation of CV-Conv can be represented by Equation 2. The CV-Conv learns by backpropagation, a sufficient condition is to use a loss function as well as activations that are differentiable with respect to real and imaginary components separately in the network (Hirose and Yoshida, 2012; Chiheb et al., 2017). The authors in Hirose and Yoshida (2012) demonstrate that the separately differentiable functions are compatible with the backpropagation of the CV-Conv layer.

2.1.2 Complex-valued activations

A variety of activation functions have been created to handle complex-valued representations. In this work, we used the complex ReLU (CReLU) activation function (Chiheb et al., 2017). The CReLU activation function applies the ReLU individually to the real and imaginary parts of the neuron.

$$\mathbb{C}ReLU(I) = ReLU(I_{\Re}) + iReLU(I_{\Im}) \quad (3)$$

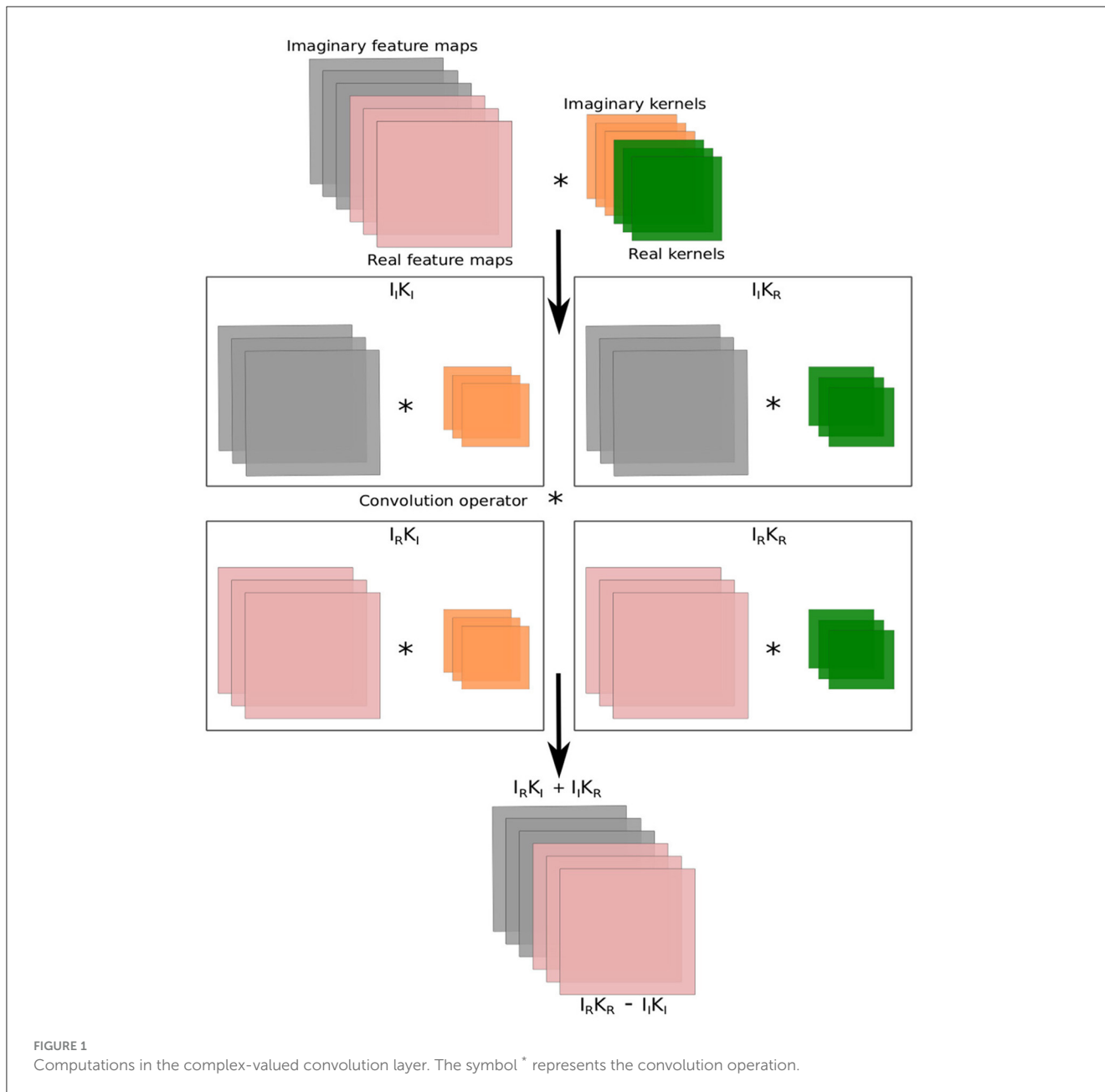
The Equation 3 shows the mathematical form of the CReLU activation function, I_{\Re} , and I_{\Im} represents the real and imaginary feature components of the input feature I . CReLU is a split-complex activation function, and the real and imaginary components are both individually and sectionally differentiable (Chiheb et al., 2017).

2.1.3 Complex-valued Max pooling

The complex-valued max pooling is the implementation of conventional max pooling individually on the real and imaginary part of the complex-valued features in the network (Barrachina et al., 2022) as shown in Equation 4.

$$\mathbb{C}Max(I) = Max(I_{\Re}) + iMax(I_{\Im}) \quad (4)$$

The CMax-pooling in the complex-valued CNN works with the max-by-magnitude approach in Equation 4. The real and imaginary parts are individually processed in this layer, likewise in the other layers (Chiheb et al., 2017).



2.2 Fast Fourier Transform

The Fast Fourier Transform (FFT) is used to obtain frequency information from images by transforming them from the spatial domain to the complex domain. FFT is a fast and efficient algorithm for computing a signal's discrete Fourier transform (DFT) or its inverse (Popa and Cernăzanu-Glăvan, 2018). The symmetry and periodic properties of FFT reduces the computational complexity. For an image of size $N \times N$, (p, q) and (u, v) denote the spatial and frequency coordinates. Correspondingly, $I(p, q)$ and $I(u, v)$ represent the spatial and frequency values. The 2D DFT is defined as

$$\mathcal{F}[I(u, v)] = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} I(p, q) e^{-i2\pi(\frac{u.p}{N} + \frac{v.q}{N})} \quad (5)$$

Since FFT (\mathcal{F}) is a bijective function in the image space that means we can define the inverse of FFT (\mathcal{F}^{-1}) as follows:

$$\mathcal{F}^{-1}[I(p, q)] = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} I(u, v) e^{i2\pi(\frac{u.p}{N} + \frac{v.q}{N})} \quad (6)$$

Equations 5, 6 demonstrate that an image can be transformed back and forth from spatial to frequency domain without losing any information (Gonzalez, 2009).

2.3 Attention gates

Attention gates (AGs) are widely used attention mechanisms in the field of natural language processing (NLP) (Li et al.,

2018) and computer vision for segmentation (Zuo et al., 2021), image captioning (Huang et al., 2019), and classification (Wang et al., 2017). They are categorized into two types: Hard attention (Mnih et al., 2014) involves reinforcement learning and is non-differentiable, which makes the training of the model more complicated. Soft attention (Jetley et al., 2018) is probabilistic and can be trained by using gradient descent and utilizing standard back-propagation without implementing Monte Carlo sampling. Oktay et al. (2018) proposed Attention-UNet by incorporating the attention gates (AGs) into the concatenation based skip connections of the standard UNet (Ronneberger et al., 2015) architecture to extract pancreas segmentation in 3D abdominal CT images. The architecture of Attention-UNet is shown in the upper branch of Figure 4A. The main concept behind AGs is to emphasize the salient features which are propagated through the skip connections. The AGs focus on the relevant information while suppressing the irrelevant information from the background region such as noise (Oktay et al., 2018). The AG takes two input vectors, $\mathbf{x}^l \in \mathbb{R}^{F_l \rightarrow H_x \times W_x \times C_x}$ and $\mathbf{g}^d \in \mathbb{R}^{F_d \rightarrow H_g \times W_g \times C_g}$, representing the input feature vector and a gating signal vector where F_l and F_d corresponds to the number of feature maps in layers l and d with spatial dimensions height (H), width (W), and number of channels (C), respectively, as illustrated in Figure 2. The features of the gating signal selected from the coarser scale of the deeper layers to suppress the irrelevant information. The input feature maps are downsampled to the resolution of gating signal prior to the AGs. The output $\hat{\mathbf{x}}^l \in \mathbb{R}^{H_x \times W_x \times C_x}$ of the AGs is the element-wise multiplication of the attention weight vector $\beta^l \in \mathbb{R}^{H_x \times W_x \times C_x}$ with the input feature vectors \mathbf{x}^l as shown in Equation 7. The attention weights identify the relevant image regions and preserve only the meaningful activation which are related to the task.

$$\hat{\mathbf{x}}^l = \mathbf{x}^l \cdot \beta^l \quad (7)$$

The gating vector is combined with the input vector by additive attention (Yu et al., 2018) instead of multiplicative attention.

$$\beta^l = \sigma_2(W_s^T(\sigma_1(W_x^T \mathbf{x}^l + W_g^T \mathbf{g}^d + \mathbf{b}_g)) + \mathbf{b}_s) \quad (8)$$

In Equation 8, the linear transformations are performed by: $W_x \in \mathbb{R}^{F_l \times F_s}$, $W_g \in \mathbb{R}^{F_d \times F_s}$ and $W_s \in \mathbb{R}^{F_s \times 1}$, and bias terms $\mathbf{b}_g \in \mathbb{R}^{F_s}$ and $\mathbf{b}_s \in \mathbb{R}$. F_s is the number of feature maps in the intermediate convolution layer s . The linear transformations are generated by using $1 \times 1 \times 1$ channel-wise convolution layers to decouple the feature maps from higher to lower dimensional space for the gating operation at each image scale, which reduces the training parameters and computational cost of the AGs. The linear transformation is followed by the sigmoid activation function $\sigma_2(\mathbf{x}^l) = \frac{1}{1 + \exp(-\mathbf{x}^l)}$, and the ReLU activation function $\sigma_1(\mathbf{x}^l) = \max(0, \mathbf{x}^l)$ on each feature vector. Finally, the attention weights are upsampled to the dimensions of \mathbf{x}^l using trilinear interpolation before the element-wise multiplication. These operations are carried out prior to the concatenation function to ensure that only significant and relevant information is included. The gradients of the background regions are down-weighted in the backward pass. The Attention-UNet combines the power of attention gates with UNet to guide the denoising process by selectively focusing on

structural features in the image while suppressing the noise and irrelevant details. However, the attention gates are designed with real-valued components, they are limited to real-valued features and cannot be used effectively with the complex-valued features since complex-valued features cannot be fully captured by the real-valued gating mechanism.

2.4 Datasets

In this study, we used three different types of datasets: two microscopic datasets, primarily Tubulin and Vesicle datasets, in addition to a non-microscopic BSD dataset. SIM microscopy (Gustafsson, 2000) was used to obtain raw SIM images of Tubulin and Vesicle datasets under different illumination levels and microscopic settings. The raw Tubulin and Vesicle SIM samples were acquired using 3D SIM imaging technologies (Gustafsson et al., 2008). The stack of 15 raw SIM images (i.e., five phases and three orientations) of dimensions $15 \times 512 \times 512$ (frames \times width \times height) pixels of both SIM structures were propagated into the reconstruction algorithm to produce a super-resolution SIM image of size 1024×1024 (width \times height) pixels. The fairSIM reconstruction algorithm (Müller et al., 2016) and the softWoRx commercial software were used to reconstruct the Tubulin and Vesicle SR-SIM samples in the frequency domain from the raw SIM images. The Tubulin dataset consists of 101 fields-of-view (FOVs), each FOV recorded for 200 timestamps. The Vesicle dataset is reconstructed using the 3D-SR-SIM method (i.e., softWoRx software), resulting in 3D-SR-SIM samples. However, each z-plane of the 3D-SR-SIM data is extracted using a slice-by-slice approach as 2D sample. The slicing strategy yields 175 FOVs (i.e., basically 175 different z-planes extracted from nine FOVs originally), each FOV is acquired for different timestamps. Tubulin and Vesicle datasets contain 2,525 and 7,284 samples, respectively. The training and test sets of both datasets were partitioned based on the number of FOVs. The training sets for the Tubulin and Vesicle datasets were built using 81 and 121 FOVs, respectively. The training sets of the Tubulin and Vesicle datasets comprise 2,025 and 5,562 samples. The remaining samples from both datasets were used for the test sets. Each sample in both datasets consists of the reconstructed 16-bit noisy SR-SIM input and a reference image of size $1,024 \times 1,024$ (width \times height) pixels. The input images of both datasets contain real noise, which is a combination of mixed Poisson-Gaussian noise along with the reconstruction artifacts. The complete characteristics, description, and sample preparation of both datasets are further explained in the Shah et al. (2023). The Berkeley segmentation dataset (BSD500), which is related to conventional image restoration problems, was selected as the third dataset in this work (Martin et al., 2001). The input images in this dataset were corrupted by the introduction of additive white Gaussian noise with zero mean and standard deviation of 30, respectively.

3 Proposed model architectures

This section will cover all the proposed methodologies used in this work to explore the scope of complex-valued CNN based image restoration networks, as well as the novel complex-valued

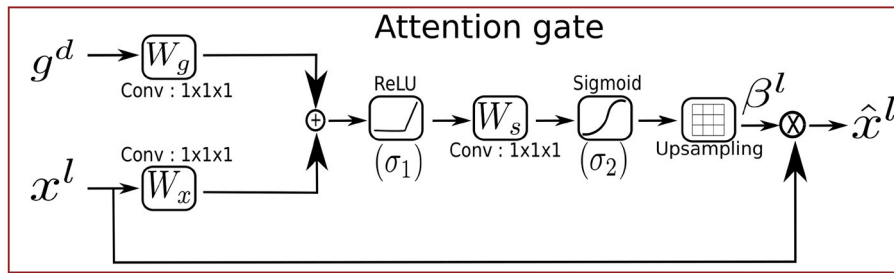


FIGURE 2 Schematic of the attention gate (AG).

attention gate that is employed in conjunction with the proposed complex-valued denoising and super-resolution architectures.

3.1 Fourier transform encoding and decoding layers

To transform an image from the spatial to the frequency domain, we used a Fourier transform encoding (FTE) layer which is based on the FFT algorithm. The FFT algorithm is used by the FTE layer to convert the input image or feature maps from the spatial domain to the complex representation. The FTE layer also shifts the DC component of FFT feature representation to the center-most location, where the positive and negative frequencies surround the DC component. Since CV-CNNs only interact with the complex values, meaningful complex-valued feature mappings must be transmitted into the CV-CNN layer for further processing.

The Fourier transform decoding (FTD) layer reverses the complex-valued feature representations into their real values using the inverse Fast Fourier transform (IFFT). The DC component of the Fourier representation is first shifted from the center-most location back to the original position, along with the low and high-frequencies. Following that, the IFFT is applied to complex-valued feature maps to convert them back into real-valued spatial feature maps. These real-valued feature maps are then propagated to the RV-CNNs. These FFT-based encoding and decoding layers are used within our proposed architectures.

3.2 Complex-valued attention gates

The standard attention gates (AGs) are not designed to compute attention values natively with complex features. One possibility is to process only the real part (i.e., magnitude information) of feature maps in the attention gates, while discarding the imaginary part (i.e., phase information). However, the loss of phase information can often have significant consequences, since the imaginary part carries a crucial structural representation in the complex-valued features. To mitigate the loss of phase information, complex-valued attention gates (C-AGs) have been developed during this work. The C-AGs inspired by the real-valued AGs are one of the major contributions of this work; their core architecture resembles real-valued attention

gates. To handle the complex-valued feature maps, the C-AG is built around complex components. The complex components in the C-AGs allow the attention mechanism to capture both real and complex values of the input features and to leverage the phase information in the attention mechanism of the skip connections of complex-valued networks. The complex attention weight vector in C-AGs, enables the attention mechanism to recognize the salient frequency regions in the frequency domain while suppressing the irrelevant frequency regions. The inputs to C-AGs are the complex input feature vector and the complex gating signal vector, $\tilde{\mathbf{x}}^l \in \mathbb{C}^{F_l \times H_{\tilde{\mathbf{x}}} \times W_{\tilde{\mathbf{x}}} \times C_{\tilde{\mathbf{x}}}}$ and $\tilde{\mathbf{g}}^d \in \mathbb{C}^{F_d \times H_{\tilde{\mathbf{g}}} \times W_{\tilde{\mathbf{g}}} \times C_{\tilde{\mathbf{g}}}}$, where F_l and F_d represent the feature maps with dimensions height (H), width (W), and number of channels (C) of layers l and d , respectively, as shown in Figure 3. The tilde indicates the complex-valued representation. The complex gating signal is selected from a coarser scale. Since C-AGs only work with the complex-valued feature maps, the convolution layer in the complex attention block is replaced by CV-Conv layers. The input and the gating vectors are propagated through the CV-Conv layers and summed element-wise to maximize the aligned weights while minimizing the unaligned weights (Oktay et al., 2018). The resulting vector goes through a CReLU (ϕ_1), 1x1x1 CV-Conv layer, and CSigmoid (ϕ_2) activation function to produce the complex attention weights (α). The complex attention weights are upsampled to the dimension of $\tilde{\mathbf{x}}^l$ using a complex-valued upsampling layer, as shown in Figure 3.

$$\alpha^l = \phi_2(W_s^T(\phi_1(W_{\tilde{\mathbf{x}}}^T \tilde{\mathbf{x}}^l + W_{\tilde{\mathbf{g}}} \tilde{\mathbf{g}}^d + \mathbf{b}_{\tilde{\mathbf{g}}})) + \mathbf{b}_{\tilde{\mathbf{s}}}) \quad (9)$$

In Equation 9, $W_{\tilde{\mathbf{s}}} \in \mathbb{C}^{F_s \times 1}$, $W_{\tilde{\mathbf{x}}} \in \mathbb{C}^{F_l \times F_s}$, $W_{\tilde{\mathbf{g}}} \in \mathbb{C}^{F_d \times F_s}$ are the linear transformations with bias terms $\mathbf{b}_{\tilde{\mathbf{g}}} \in \mathbb{C}^{F_s}$ and $\mathbf{b}_{\tilde{\mathbf{s}}} \in \mathbb{C}$. F_s represents the number of feature maps in the intermediate layer $\tilde{\mathbf{s}}$.

In the final step, the output of the C-AGs is obtained by the convolution (*) of the complex-valued feature maps with the complex-valued attention weight vectors rather than the element-wise multiplication, due to their complex nature, as shown in the Equation 10.

$$\tilde{\mathbf{x}}^l = \tilde{\mathbf{x}}^l * \alpha^l \quad (10)$$

In Equation 10, $\tilde{\mathbf{x}}^l \in \mathbb{C}^{H_{\tilde{\mathbf{x}}} \times W_{\tilde{\mathbf{x}}} \times C_{\tilde{\mathbf{x}}}}$ is the complex attention gate output of complex feature maps ($\tilde{\mathbf{x}}$) and complex attention weights (α^l). The proposed C-AGs are trainable and differentiable, and thus fall in the category of soft attention.

3.3 Complex-valued attention UNet

To denoise and reconstruct the super-resolved images entirely in the complex/frequency domain, we devise the underlying Attention-UNet (Oktay et al., 2018) to a complex-valued Attention-UNet (CV-Atten-UNet). The CV-Atten-UNet is a complex version of the real-valued Attention-UNet since the model is constructed with complex components such as CV-convolution, \mathbb{C} max pooling, and \mathbb{C} attention gates (\mathbb{C} -AGs) to interact with the complex features. Consequently, the input images are transformed into complex inputs ($I_{in} \in \mathbb{C}^{m_1 \times m_1}$) using a Fourier encoding layer. The complex components incorporate the complex-valued representations (i.e., the real and imaginary information) to enhance the effectiveness of the complex-valued networks (Lee et al., 2022). Similarly, the deployment of \mathbb{C} -AGs allows the network to dynamically adjust the feature fusion, leading to the capture of adaptive feature representations by suppressing the irrelevant features. \mathbb{C} -AGs connect the skip connection symmetrically to the adjacent complex encoder and decoder blocks, as shown in the lower branch of Figure 4B. The architecture is based on four complex encoder and decoder blocks. Each complex encoder block contains two complex-valued convolution layers along with \mathbb{C} Relu activation functions and a complex pooling layer to compress the complex feature maps at different scales. The complex decoder block expands the features and is made up of two complex-valued convolution layers along with a complex-valued upsampling layer. In the final block, the Fourier decoding layer is merged with a single convolution layer to obtain the spatial output ($O \in \mathbb{R}^{m_1 \times m_1}$). The description of CV-Atten-UNet is given in Equation 11.

$$\text{Output} = \text{FTD}(\text{Complex decoder}_n(\text{Complex encoder}_n(I_{in}))) \quad (11)$$

where:

Complex encoder block : $\text{CV-Conv} \rightarrow \mathbb{C}\text{ReLU} \rightarrow \text{CV-Conv} \rightarrow \mathbb{C}\text{ReLU} \rightarrow \mathbb{C}\text{Max pooling}$

Complex decoder block : $\text{CV-Conv} \rightarrow \mathbb{C}\text{ReLU} \rightarrow \text{CV-Conv} \rightarrow \mathbb{C}\text{Relu} \rightarrow \mathbb{C}\text{Upsampling}$

FTD : Represents the Fourier transform decoding layer.

In the Equation 11, n is 4, reflecting the number of complex encoder and decoder blocks in the contraction and expansion sections.

3.4 Complex-valued collaborative attention network

To enhance the performance of CV-CNN-based methods, we propose a complex-valued collaborative attention network (CV-CAN) to eliminate noise and provide a high-quality clean image. Figure 3 shows the architecture of CV-CAN which takes the spatial input image ($I_{in} \in \mathbb{R}^{m_1 \times m_1}$) and reconstructs a spatial output ($O \in \mathbb{R}^{m_1 \times m_1}$). Specifically, the CV-CAN network is comprised of several real and complex-valued encoder-decoder blocks. The real-valued encoder blocks first compress the image and extract

the low level features. The features maps produced by the real-valued encoder blocks are converted from spatial into frequency domain. The transformation of the feature maps takes place using the FTE layer. The complex-valued encoder blocks extract further features in the frequency domain. The encoder blocks exploit different level of details in the spatial and frequency domains. Similarly, the complex-valued decoder blocks combines the low and high level to reconstruct the desired complex-valued feature maps. The final output image is obtained by employing the real-valued decoder blocks followed by the FTD layer. The real and complex encoder-decoder blocks are composed of real and complex-valued components, respectively. The proposed \mathbb{C} -AGs are also employed in the skip connections across the adjacent complex-valued blocks to emphasize the relevant significant features in the frequency domain. Similarly, the real-valued encoder-decoder blocks are also joined via the skip connections with attention gates to focus on the significant and relevant spatial features. The learnable AGs compress the irrelevant noisy background information (Oktay et al., 2018).

The CV-CAN employs two complex and real-valued encoder decoder blocks. The description of real and complex blocks is illustrated in Equation 12.

$$\text{Output} = \text{Decoder}_n(\text{FTD}(\text{Complex decoder}_n(\text{Complex encoder}_n(\text{FTE}(\text{Encoder}_n(I_{in})))))) \quad (12)$$

where:

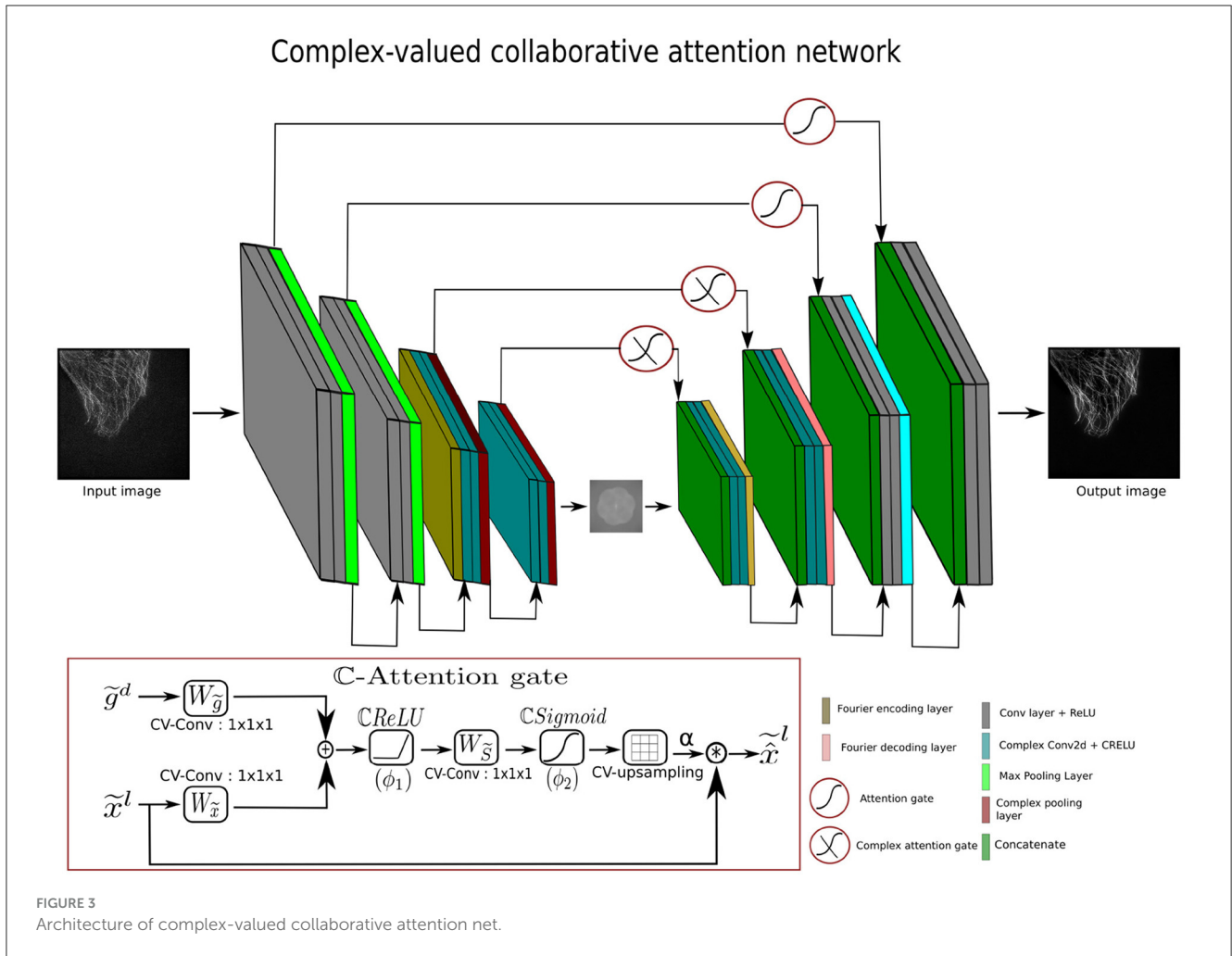
Encoder block : $\text{Conv} \rightarrow \text{ReLU} \rightarrow \text{Conv} \rightarrow \text{ReLU} \rightarrow \text{Max pooling}$

Decoder block : $\text{Conv} \rightarrow \text{ReLU} \rightarrow \text{Conv} \rightarrow \text{Relu} \rightarrow \text{Upsampling}$

Complex encoder and decoder blocks are explained in Equation 11. The FTE and FTD correspond to the Fourier transform encoder and decoder layers. The n represents the number of blocks, which is 2.

3.5 Complex-valued dual-domain attention network

The complex-valued dual-domain attention network (CV-DDAN) operates on both spatial and frequency domain data simultaneously to further improve the frequency learning ability of the existing CV-CNN methods. The network accepts two inputs ($I_{in1} \in \mathbb{R}^{m_1 \times m_1}$ and $I_{in2} \in \mathbb{C}^{m_1 \times m_1}$), from two different representations, one of which is a real-valued spatial input and the other is a complex-valued frequency input as shown in the Figure 4. The input of the complex branch is generated by the FTE layer. To process these inputs, the network has two separate encoding-decoding pathways to extract the features from spatial and complex domains. The adjacent real and complex encoder-decoder blocks of the dual-domain attention network are joined with each other via real and complex-valued AGs in their respective skip connections. The output of the complex branch is transformed from the complex feature to the real feature maps using the FTD layer, as stated in Equation 13. The outputs of both branches are fused by an additive



operation and fed into the final output block to produce a spatial domain output image ($O \in \mathbb{R}^{m_1 \times m_1}$).

The purpose of the dual-domain attention network is to capture and exploit maximum information from the spatial and frequency domains by using real and complex branches during the training process. Another important aspect is to retrieve the high frequency components of the output image which is not the case with traditional CNN methods.

$$\text{Output} = \text{Decoder}_n(\text{Encoder}_n(I_{in1})) + \text{FTD}(\text{Complex decoder}_n(\text{Complex encoder}_n(I_{in2}))) \tag{13}$$

The description of real and complex encoder and decoder blocks is illustrated in the Equations 11 and 12, where the value of n is 4.

3.6 Complex-valued super-resolution network

To investigate the potential of CV-CNNs for the super-resolution task in the frequency domain we design a

complex-valued super-resolution network (CV-SRN). The CV-SRN model is the modified version of CV-Att-UNet model. The integration of an upsampling block into the CV-Att-UNet architecture results in the CV-SRN model. The main concept behind this architecture is to explore the reconstruction capabilities of CV-CNN models in the complex domain after the application of Fourier Transform. CV-SRN converts the complex-valued feature maps before the upsampling block using the FTD layer as shown in Figure 5.

The CV-SRN architecture takes complex-valued input ($I_{in} \in \mathbb{C}^{m_1/2 \times m_1/2}$) which is half the size of the final output image ($O \in \mathbb{R}^{m_1 \times m_1}$). The CV-SRN processes the low resolution image of size 512×512 pixels and generates a high-quality SR-SIM of size $1,024 \times 1,024$ pixels. This architecture can effectively increase the spatial dimensions of super-resolution images by a factor of two.

4 Experimental results

We carried out a series of experiments to assess the performance of the proposed CV-CNN architectures in denoising and super-resolution tasks. Of special importance is the comparison with the corresponding real-valued counterparts.

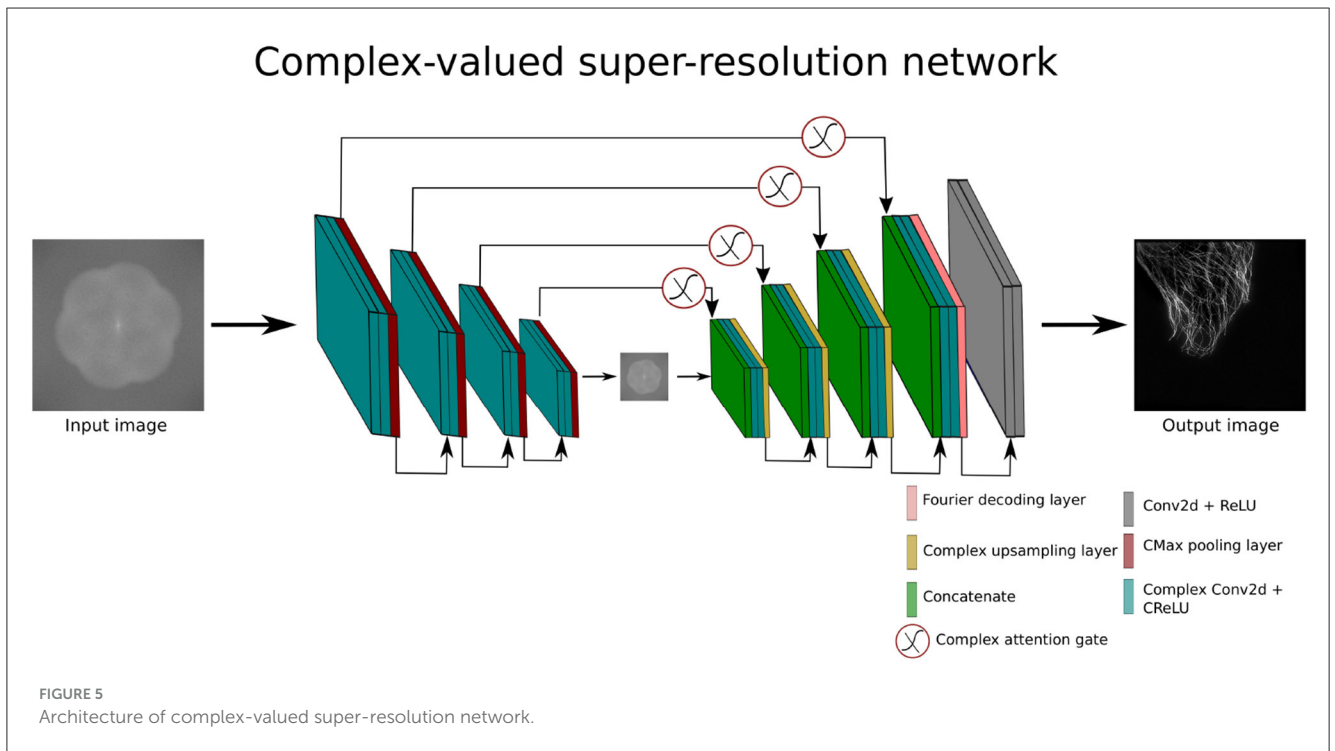
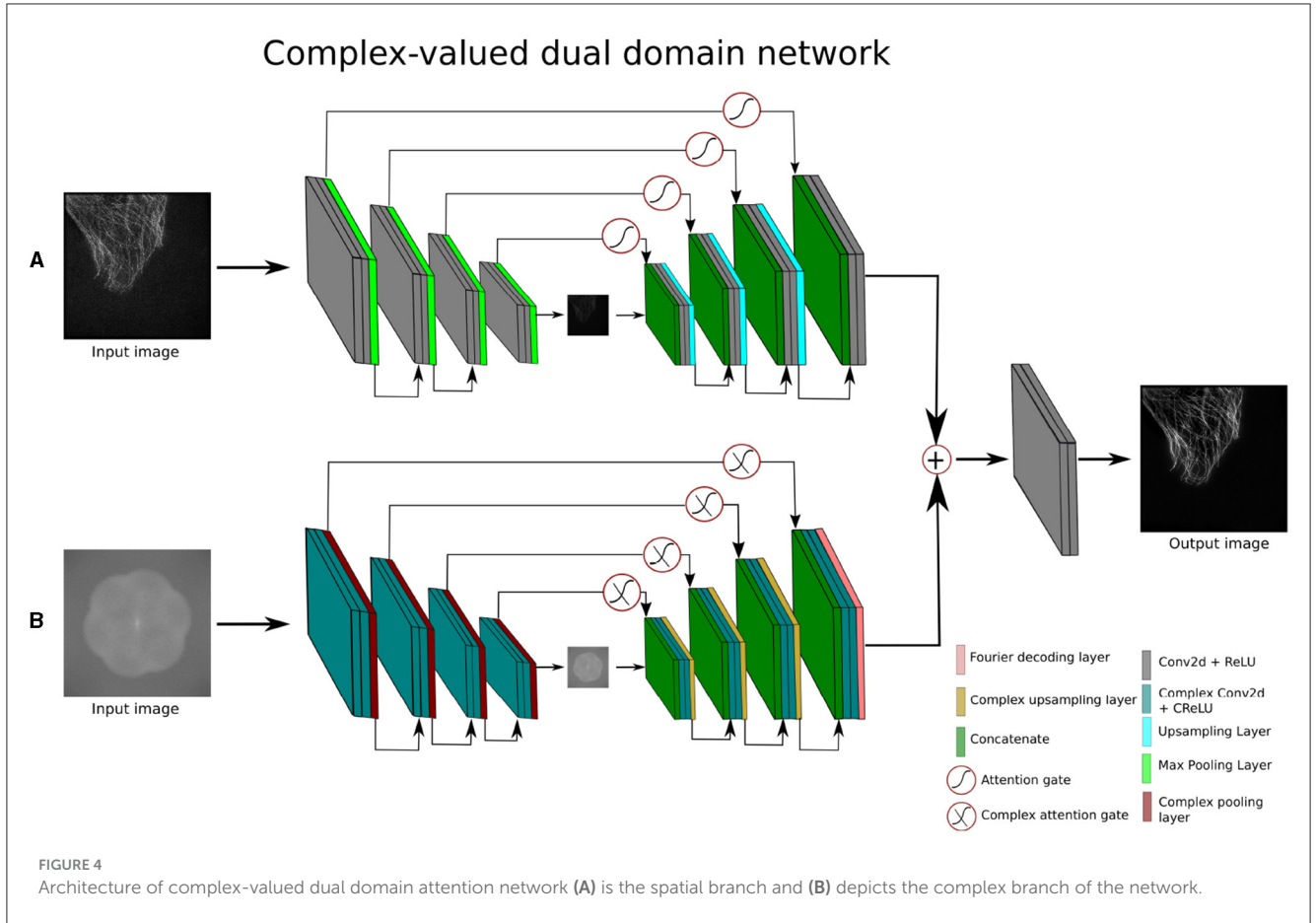


TABLE 1 Mean PSNR and SSIM values of three runs along with standard deviations (STD) of all methods calculated on complete test data of Tubulin, Vesicle, and BSD datasets.

	Mean PSNR (STD) and SSIM (STD) values of different methods											
	Tubulin dataset				Vesicle dataset				BSD dataset			
	1st run	2nd run	3rd run	SSIM ± STD	1st run	2nd run	3rd run	SSIM ± STD	1st run	2nd run	3rd run	SSIM ± STD
fairSIM	23.61 ± 0.00	23.61 ± 0.00	23.61 ± 0.00	0.29 ± 0.29	35.10 ± 0.00	35.10 ± 0.00	35.10 ± 0.00	0.86 ± 0.86	18.99 ± 0.00	18.99 ± 0.00	18.99 ± 0.00	0.28 ± 0.28
Red-fairSIM	27.97 ± 0.04	27.98 ± 0.04	27.99 ± 0.04	0.71 ± 0.70	38.43 ± 0.26	39.01 ± 0.26	38.95 ± 0.26	0.89 ± 0.88	29.71 ± 0.12	29.98 ± 0.12	29.74 ± 0.12	0.83 ± 0.82
UNet-fairSIM	26.80 ± 0.35	27.82 ± 0.35	27.92 ± 0.35	0.69 ± 0.70	37.45 ± 0.24	37.99 ± 0.24	37.96 ± 0.24	0.88 ± 0.88	29.56 ± 1.39	26.39 ± 1.39	29.06 ± 1.39	0.82 ± 0.82
Attention-UNet	27.37 ± 0.45	28.21 ± 0.45	27.14 ± 0.45	0.69 ± 0.69	37.50 ± 0.05	37.64 ± 0.05	37.58 ± 0.05	0.88 ± 0.88	27.25 ± 0.71	27.67 ± 0.71	28.93 ± 0.71	0.82 ± 0.82
CV-Attention-UNet	26.81 ± 0.19	27.29 ± 0.19	26.99 ± 0.19	0.62 ± 0.62	35.70 ± 0.40	36.12 ± 0.40	35.12 ± 0.40	0.83 ± 0.82	23.47 ± 2.23	26.52 ± 2.23	22.70 ± 2.23	0.63 ± 0.64
CV-CAN	27.44 ± 0.18	27.86 ± 0.18	27.50 ± 0.18	0.71 ± 0.70	36.56 ± 0.27	36.79 ± 0.27	37.23 ± 0.27	0.88 ± 0.88	28.59 ± 0.11	28.87 ± 0.11	28.68 ± 0.11	0.82 ± 0.81
CV-DDAN	28.13 ± 0.08	28.02 ± 0.08	27.92 ± 0.08	0.73 ± 0.73	37.83 ± 0.14	37.48 ± 0.14	37.59 ± 0.14	0.89 ± 0.89	29.21 ± 0.20	29.27 ± 0.20	29.68 ± 0.20	0.83 ± 0.83

The highest mean PSNR and SSIM values of the three attempts are indicated in bold.

Furthermore, this study also investigates the generalization capabilities of the suggested denoising methods.

We used three publicly available datasets for the benchmarking, two microscopic ones (Tubulin, Vesicle) and one conventional (BSD). The Tubulin and Vesicle datasets consist of high-quality SR-SIM images with real noise, which is a mixture of Poisson-Gaussian noise (Shah et al., 2023), while the BSD dataset contains additive white Gaussian noise (AWGN). To ensure fair comparison, all the denoising networks were trained on these three datasets with consistent hyper-parameters such as number of epochs, loss function, and learning rate¹. The mean squared error loss function and the ADAM optimizer were used to train all the networks.

4.1 Denoising

To evaluate the real-valued and complex-valued networks for the image denoising task, we first trained several state-of-the-art real-valued denoising networks, such as UNet-fairSIM, Red-fairSIM (Shah et al., 2021), and Attention-UNet (Oktay et al., 2018), on the aforementioned datasets for three trials each. The quantitative and visual results of these real-valued networks are shown in Table 1 and Figures 6–8. To summarize the results of the real-valued networks, it is worth mentioning that the Red-fairSIM provides visually and quantitatively (i.e., individually and collectively) superior results for all three datasets among the real-valued denoising networks employed in this study.

To investigate the potential of CV-CNNs, we first trained a simple pure CV-CNN based method called CV-Att-UNet (i.e., explained in Section 3.3), which mimics the traditional Attention-UNet architecture for the denoising tasks. First, dummy zero imaginary values are introduced to the scalar pixel values of the input/output images to convert them from real to complex pixel values. In this setting, the CV-Att-UNet network provides almost identical results to RED-fairSIM. However, a significant drop in CV-Att-UNet performance was observed when the input and output images were transformed using the FFT approach, as illustrated in Table 1. Similarly, denoised images of the CV-Att-UNet reveal that the noise has been reduced, but the structural information of the object is also missing from the resulting images, as shown in Figures 6–8. This investigation indicates that the pure CV-CNN-based architecture, such as CV-Att-UNet, fails to completely denoise and restore the images in the frequency domain, particularly with high noise levels.

To utilize the CV-CNNs more effectively, we designed two novel CV-CNN architectures, CV-CAN (explained in Section 3.4) and CV-DDAN (explained in Section 3.5), for the image denoising task. The complex-valued layers in the CV-CAN are used in a serial scheme to create a compact network. The quantitative findings of the CV-CAN model are equivalent to the real-valued networks

¹ Although the CV-CAN and CV-DDAN networks converge faster, typically after 40 epochs than real-valued CNNs, which take 75 epochs, all models were trained and monitored for 100 epochs to ensure a fair comparison. From these 100 epochs, the best performing models are selected. During the training process, the optimal learning rate was set to 1×10^{-4} to train all networks.

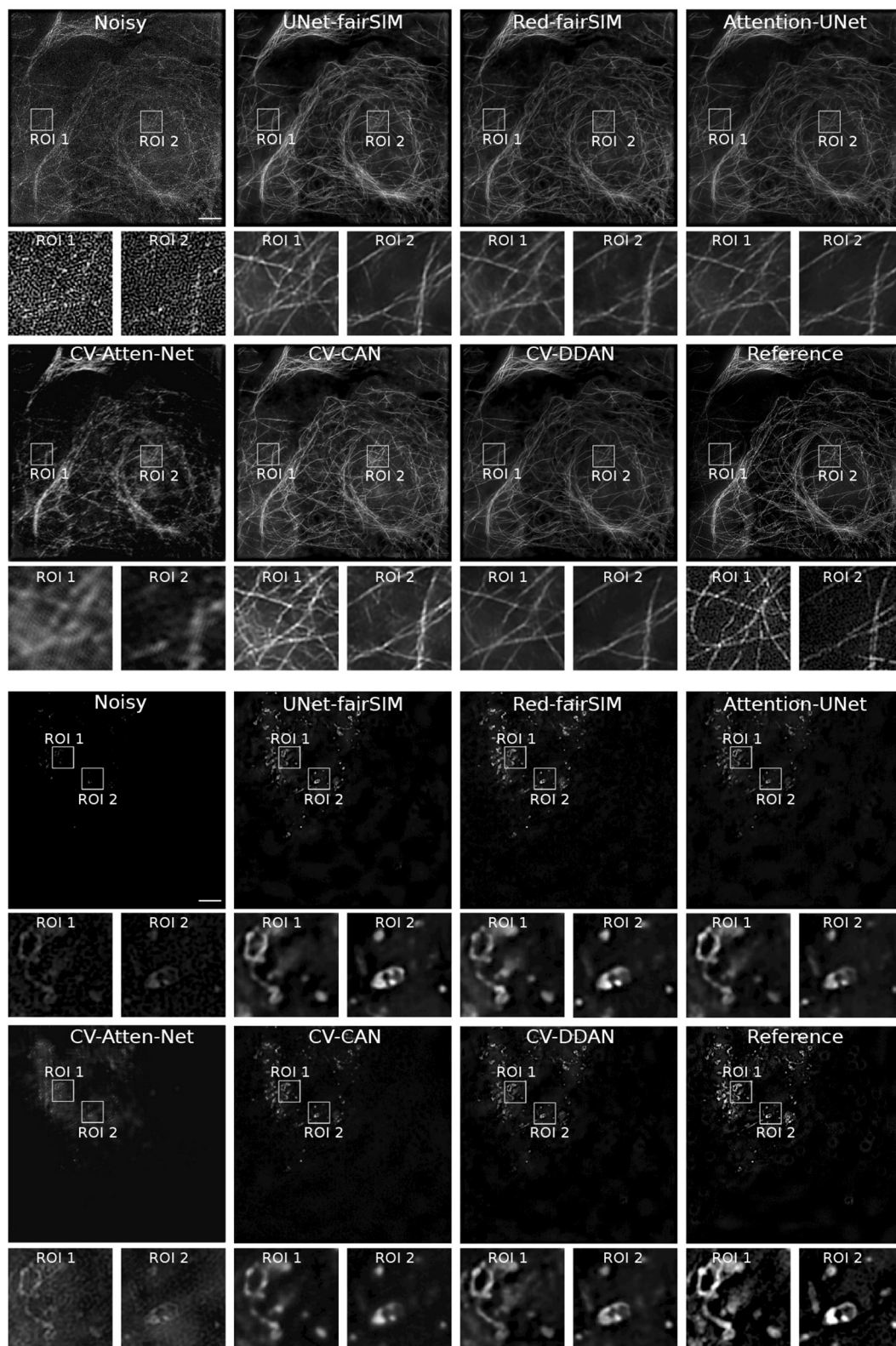
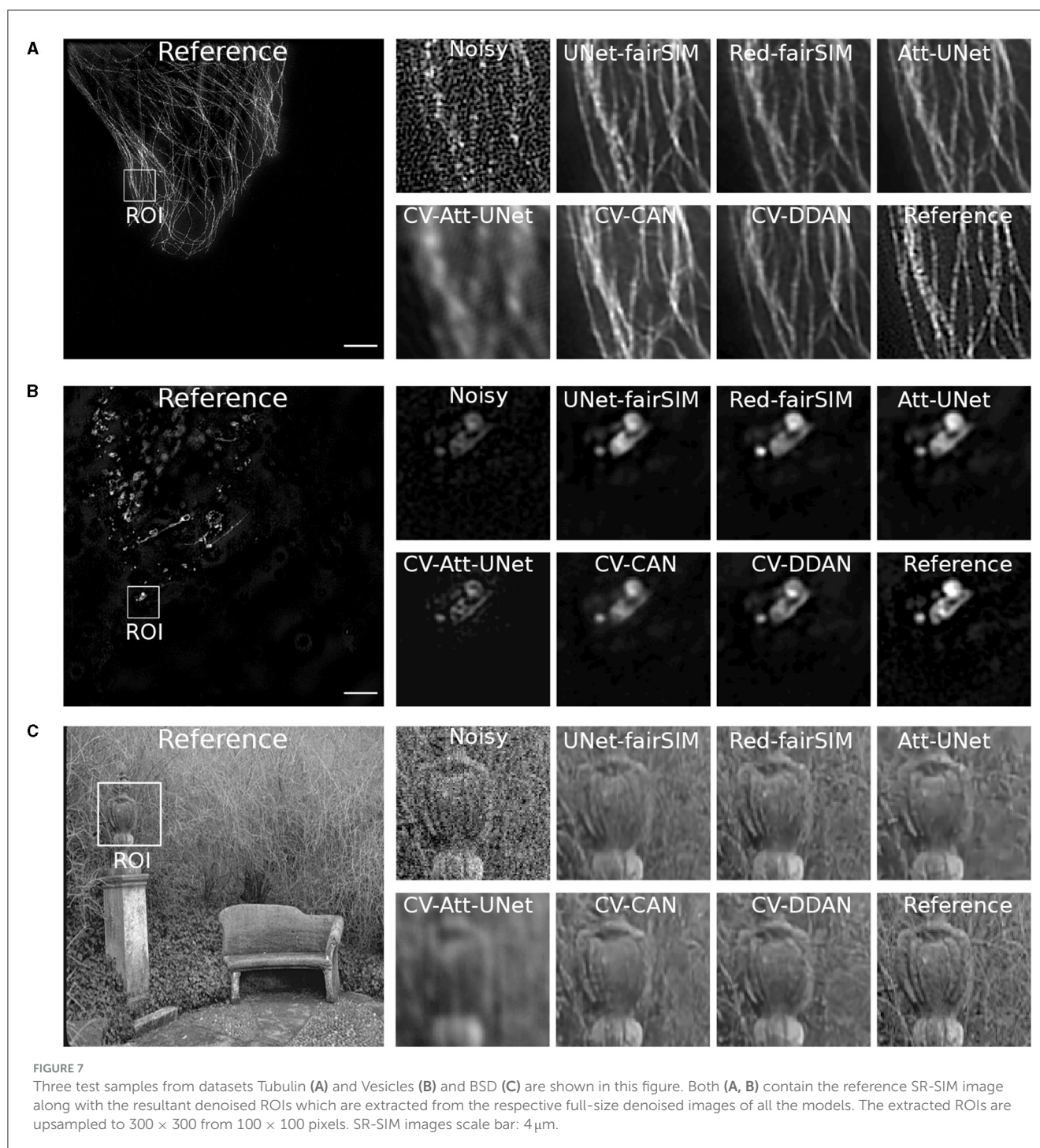


FIGURE 6

Results of test samples of Tubulin and Vesicle datasets. The first four rows show the results of the Tubulin images and the next four rows depict the Vesicle SR-SIM samples. The first and fifth rows contain the noisy as well as the denoised images of RV-CNN based methods, whereas, the third and seventh rows display the images denoised by CV-CNN based models along with the reference images. The second, fourth, sixth, and eighth rows display two regions of interest (ROI) extracted from each respective image. The cropped ROIs of size 100×100 pixels are upsampled to 300×300 for demonstration purposes. SR-SIM images scale bar: $4 \mu\text{m}$.



for the Tubulin dataset shown in [Table 1](#). However, we noticed a drop in the PSNR and SSIM values (i.e., see [Table 1](#)) for the BSD and Vesicle datasets. The CV-CAN provides superior denoised images with improved visual appearance compared to the real counterparts, as shown in the ROIs of [Figures 6–8](#), specifically for the Tubulin and Vesicle datasets.

The architecture of CV-DDAN is also a combination of real and complex-valued based CNN layers. The architecture of CV-DDAN is composed of real and complex-valued branches connected in a parallel scheme with a summation-based operator in the final block. This network requires two concurrent inputs:

the spatial domain input for the real branch and the frequency domain input for the complex branch. The denoised ROIs of CV-DDAN in [Figures 6–8](#) show more promising results than the other networks for both microscopic SR-SIM and conventional BSD datasets. The quantitative findings on the Tubulin dataset clearly demonstrate that the proposed CV-DDAN outperforms all other image denoising methods in terms of the individual and collective average PSNR and SSIM values as shown in [Table 1](#). However, in the case of the Vesicle and BSD datasets, a modest decline in average PSNR values can be noticed in the [Table 1](#), though the CV-DDAN is still superior in terms

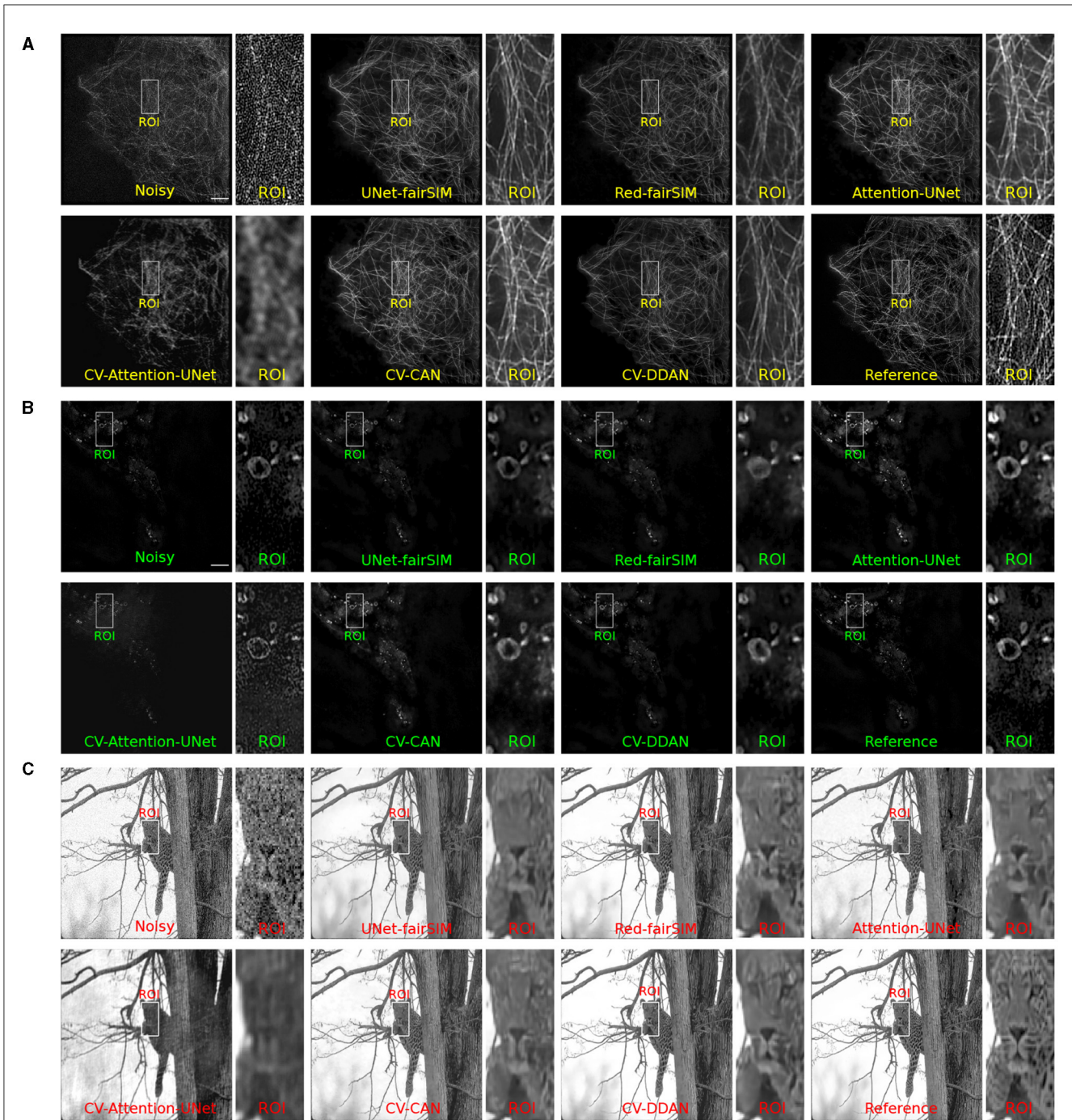


FIGURE 8
 Results of test samples of Tubulin (A), Vesicle (B), and BSD (C) datasets are shown in this figure. The first two blocks show the results of Tubulin and Vesicle SR-SIM images. The first row in each block contains the noisy along with denoised images of RV-CNN based models whereas, the second row in each block shows the images denoised by CV-CNN based model together with the reference images. The odd columns present the full-size outputs of all the methods and the even number columns display the region of interest (ROI) extracted from the respective image. The cropped ROIs of size 50 × 100 pixels are upsampled to 100 × 200 for visualization. SR-SIM images scale bar: 4 μm.

of average SSIM values. Similarly, the Fourier spectrum of all denoised images shows that the FFT spectrum of both CV-CAN and CV-DDAN networks is dense and visually similar to the FFT spectrum of the reference image (see Figure 9). This implies that our proposed complex-valued-based approaches preserve more high-frequency information than their real-valued counterparts.

4.2 Super-resolution

The results of the CV-SRN architecture (explained in Section 3.6) are compared with a real-valued super-resolution UNet (SR-UNet). The SR-UNet is a modified form of the conventional UNet (Ronneberger et al., 2015) in that it is extended by an upsampling block. The upsampling block consists of

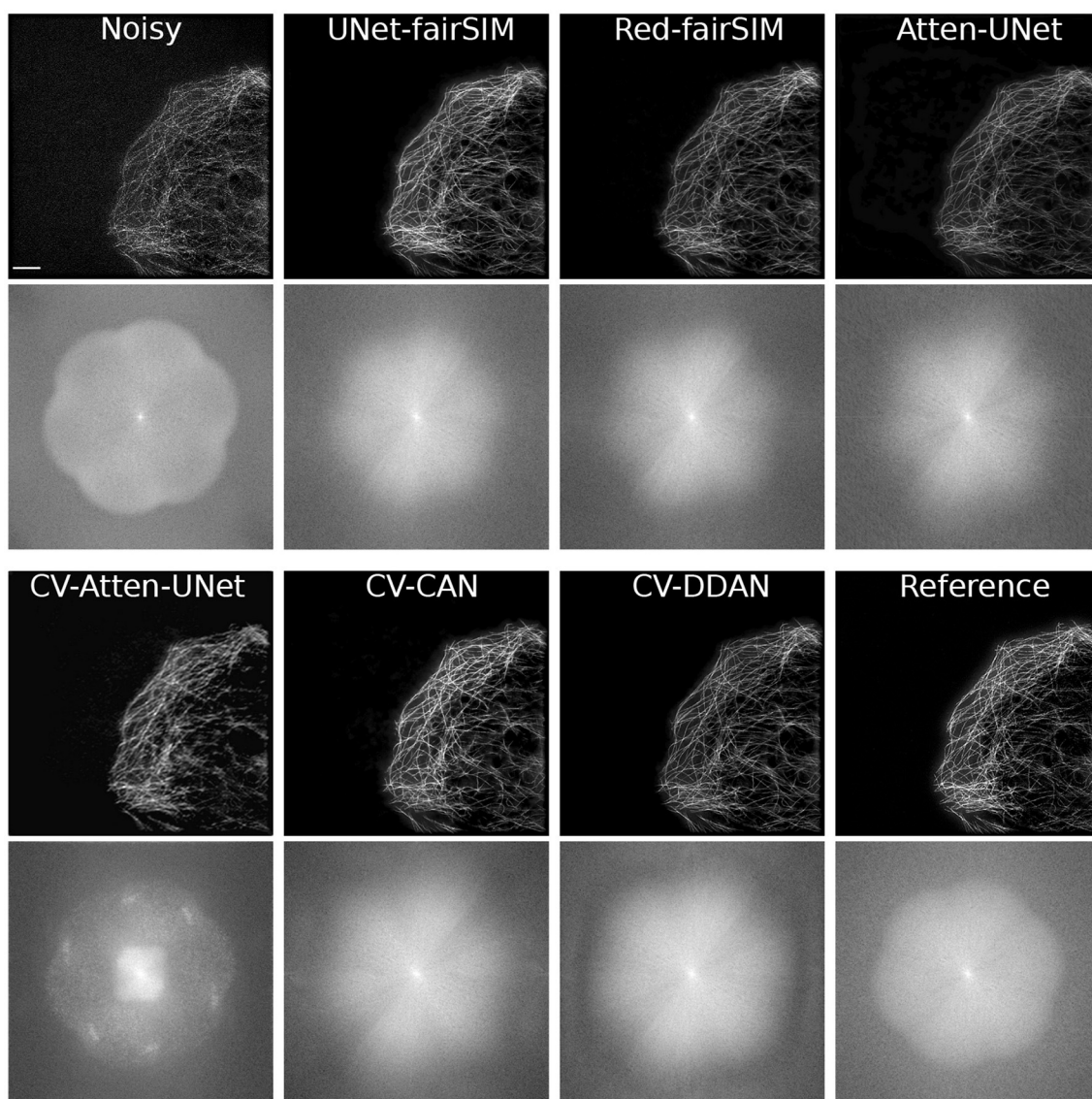


FIGURE 9

This figure shows the results of test samples of the Tubulin dataset along with the FFT spectrum. The first and third rows show the results of the spatial output of the Tubulin images. The second and fourth rows show the Fourier spectrum of the corresponding images. SR-SIM images scale bar: 4 μm .

upsampling and convolution layers. The SR-UNet works with the spatial domain inputs and outputs. In contrast, the CV-SRN takes a frequency domain input and produces a twofold spatial high-quality super-resolved output. In the case of SR-SIM reconstruction, 15 raw SIM images of size 512×512 (width \times height) pixels are combined to compute the average projection image. The input image is transformed into a complex domain by the FFT transformation.

The experimental results in Figure 10 and Table 2 show that the CV-SRN can successfully carry out the super-resolution task, i.e., that it can generate a double spatial dimension high-resolution output image from the high SNR low-dimensional image. The visual findings demonstrate good quality structures almost identical to the real-valued network, however, the mean PSNR and SSIM values in Table 2 are smaller than for the real-valued counterpart. However, the CV-SRN is limited to super-resolution tasks and fails

to yield presentable results in the case of joint denoising and super-resolution tasks from low-SNR SIM images. This indicates again the limitations of pure CV-CNN networks (similar to the CV-Att-UNet network in Section 4.1) for general image restoration tasks.

4.3 Knowledge transfer

CV-CNNs are known for their strong ability to generalize well (Lee et al., 2022). In this work, we therefore also explore the generalization power of RV-CNNs and CV-CNNs on previously unseen biological structures in a direct-transfer setting. Direct transfer refers to the ability of pre-trained models to generalize on unseen new test data from a closely related domain without any retraining or fine-tuning. For our direct-transfer experiment with selected RV-CNN and CV-CNN architectures (Red-fairSIM

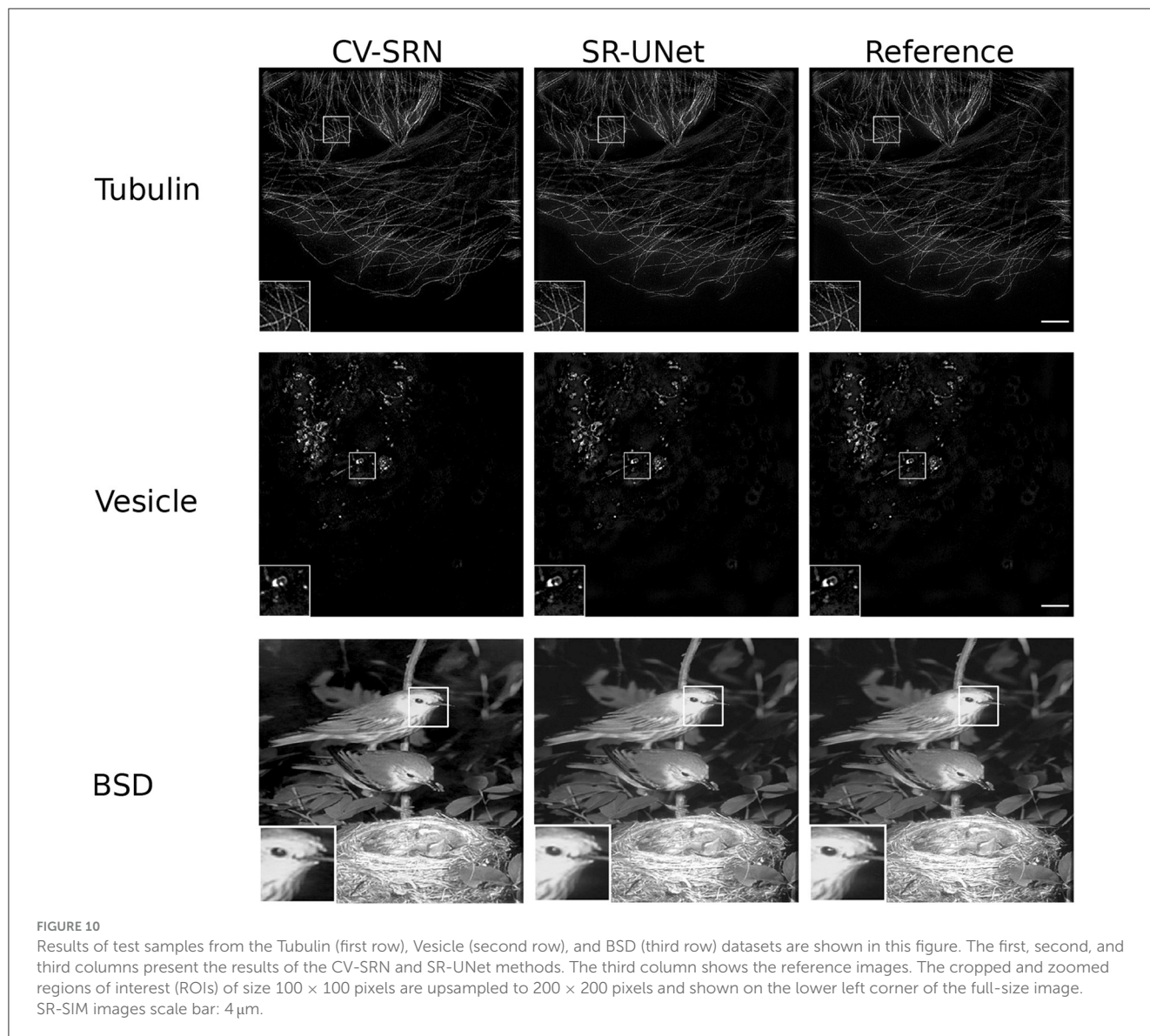


FIGURE 10 Results of test samples from the Tubulin (first row), Vesicle (second row), and BSD (third row) datasets are shown in this figure. The first, second, and third columns present the results of the CV-SRN and SR-UNet methods. The third column shows the reference images. The cropped and zoomed regions of interest (ROIs) of size 100 × 100 pixels are upsampled to 200 × 200 pixels and shown on the lower left corner of the full-size image. SR-SIM images scale bar: 4 μm.

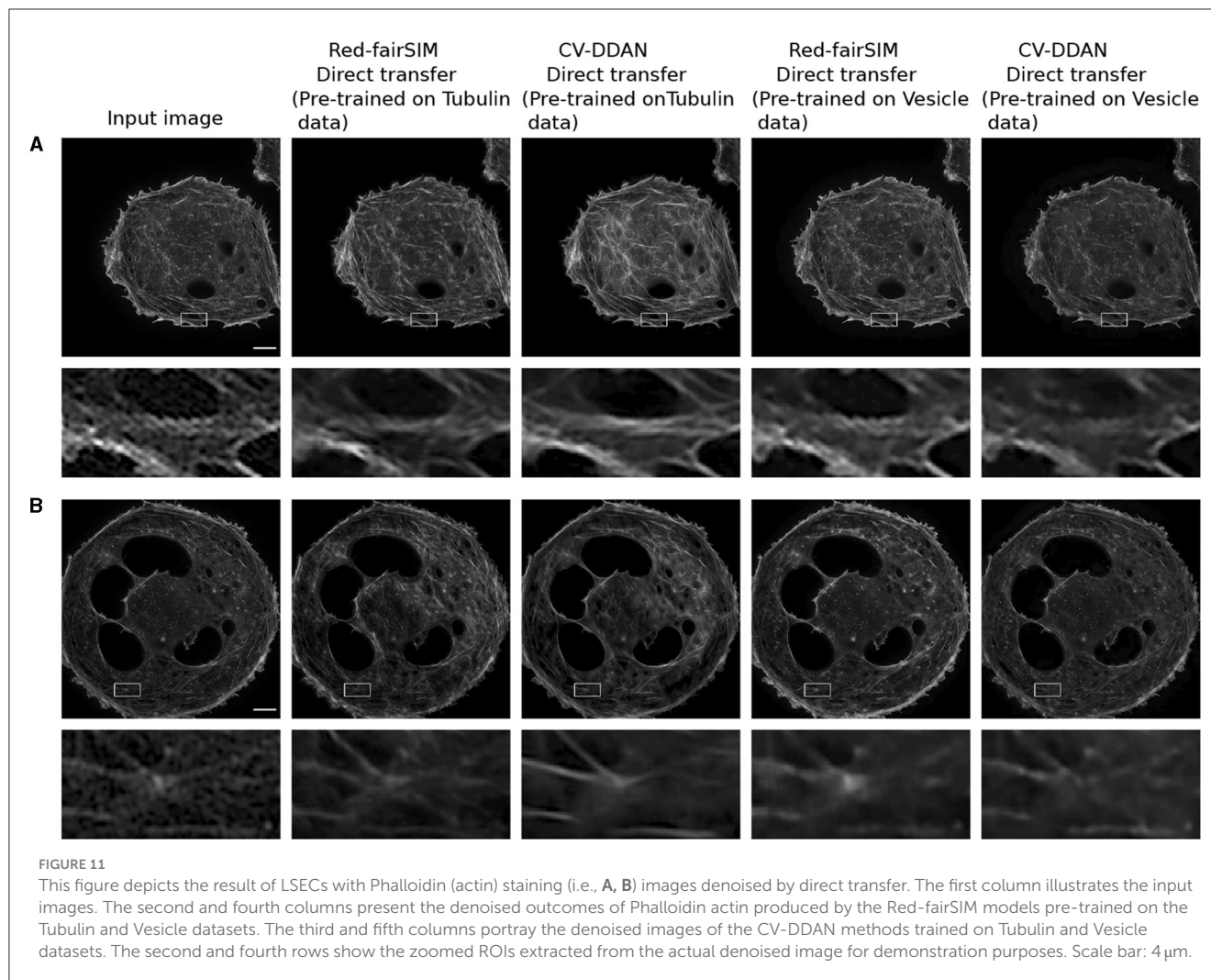
TABLE 2 Mean PSNR and SSIM values along with standard deviations (STD) calculated over the test samples of the BSD, Tubulin, and Vesicle datasets for super-resolution tasks.

	Mean PSNR (STD) and SSIM (STD) values of different methods			
	CV-SRN		SR-UNet	
	PSNR (STD)	SSIM (STD)	PSNR (STD)	SSIM (STD)
Tubulin dataset	46.37 (0.60)	0.95 (0.003)	40.73 (2.10)	0.97 (0.00)
Vesicle dataset	60.05 (1.15)	0.98 (0.00)	61.75 (1.08)	0.99 (0.00)
BSD	34.54 (1.90)	0.94 (0.00)	36.37 (1.95)	0.98 (0.00)

and CV-DDAN), liver sinusoidal endothelial cells (LSECs) stained with phalloidin (actin) and BioTracker membrane structures were collected via total internal reflection fluorescence – structured illumination microscopy (TIRF-SIM) (Barbieri et al., 2021) as additional test data. The results of these tests are shown in Figures 11, 12.

A closer look at Figures 11A, B shows that the previously trained CV-DDAN model on the Tubulin dataset provides visually

more convincing denoised images of LSECs stained with phalloidin (actin) probes than the Red-fairSIM model. The CV-DDAN trained on the Vesicle dataset cannot fully recover the structure compared to the Red-fairSIM model, as shown in Figure 11, but still produces adequate results by suppressing the noise, as shown in Figure 11 (see full fifth column). Similarly, these pre-trained models are able to produce denoised images of the LSECs BioTracker membrane samples as shown in Figure 12. The Red-fairSIM model pre-trained

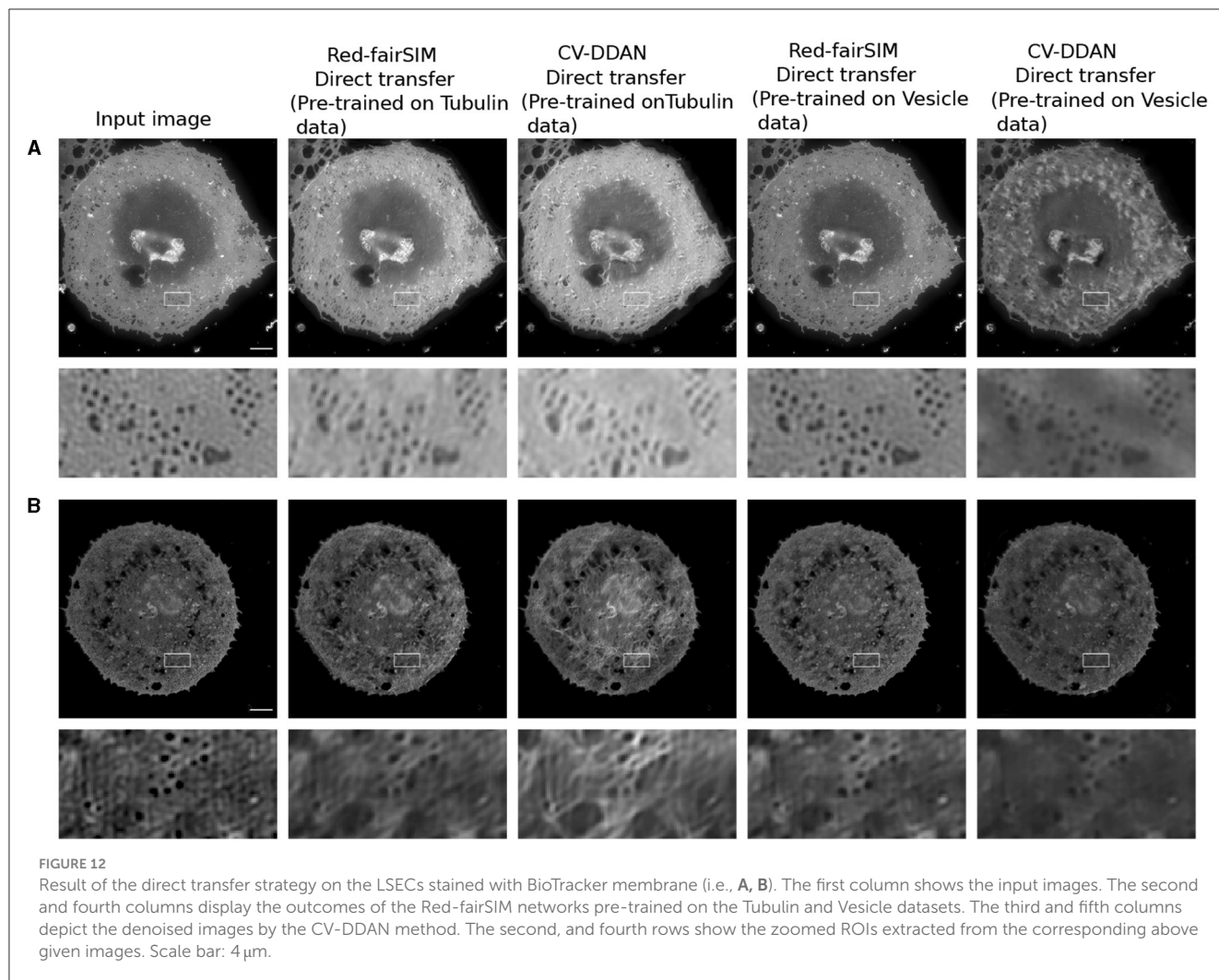


on Vesicle data shows very compelling results among all other pre-trained models as illustrated in [Figures 12A, B](#). However, the images of LSEC stained with BioTracker membrane samples denoised by the CV-DDAN pre-trained with the Vesicle dataset appear suppressed compared to other methods. Overall, we can clearly see from the results of the direct transfer approach that the pre-trained models are able to reconstruct and denoise the images of the new unseen samples to some extent, despite the characteristics of the networks. The denoised images obtained by the CV-DDAN seem to be more promising and superior in some images, but not completely.

5 Discussion

During this work, we thoroughly investigated CV-CNNs for denoising and super-resolution problems. The findings of this study provide sufficient use cases of CV-CNNs for various datasets with different noise levels and noise types. CV-CNNs were not completely investigated previously for image restoration tasks, particularly image denoising in the frequency domain. Here, we demonstrated the effectiveness of CV-CNNs for multiple image

restoration tasks. We also suggested multiple approaches for efficiently deploying CV-CNNs for the image denoising and super-resolution. In addition, we addressed the following questions: 1. Is it possible to denoise and super-resolve the images entirely in the frequency domain or the complex domain (i.e., after the FFT transformation) using CV-CNNs? 2. Do CV-CNNs outperform RV-CNNs in terms of generalization performance when applied to unseen SR-SIM modalities? To answer these questions, we trained multiple RV-CNN and CV-CNN based architectures on three datasets. We demonstrated that the pure CV-CNN based architecture (CV-Atten-UNet) cannot outperform its real counterparts on its own. However, the fusion of spatial and frequency information in our parallel scheme (CV-DDAN) provides better results than the other approaches to a certain extent. The visual results clearly show that the CV-DDAN approach outperforms its real and complex-valued image denoising counterparts on all tested datasets. While the quantitative results are better for the tubulin dataset regarding both PSNR and SSIM, only the mean SSIM values improve for the vesicle and BSD datasets. In contrast, the CV-CAN (serial scheme) produces only comparable results to real-valued SOTA denoising networks such as Red-fairSIM, UNet-fairSIM, and Attention-UNet. Overall, the



fusion of frequency and spatial learning in CV-CNN based models is crucial to enhance the recovery of signals in both domains, particularly to improve the performance in the high-frequency region, as shown in [Figure 9](#).

Similarly, the results of CV-SRN show that the CV-CNNs perform well for the super-resolution task for all three datasets, as shown in [Figure 10](#). The visual results clearly indicate that the network designed entirely with CV-CNNs alone can generate good-quality super-resolved images from the noiseless inputs, but it cannot reconstruct high-quality super-resolved output images from the images acquired under low SNR conditions. Furthermore, when comparing the generalization ability based on the direct transfer strategy, the RV-CNN and CV-CNN methods yield mostly comparable results. The effectiveness of model generalization is influenced by the diversity of the datasets and the complexity of the architectures. Therefore, it is very likely that one method will outperform the other on certain examples while struggling on others, which was indeed the case in our experiments.

A recent study ([Quan et al., 2021](#)) used CV-CNNs and proposed a CDNetwork (Complex-valued Denoising Network) to denoise the images in the complex domain by adding constant imaginary values to the pixel values in the spatial domain instead of

transforming the images by FFT algorithm. Similarly, the authors in [Rawat et al. \(2021\)](#) introduced CVMIDNet, a CV-CNN based method to eliminate Gaussian noise from chest X-ray images. In [Pham et al. \(2021\)](#), authors presented the Complex Fourier Network to generate complex filters for the denoising of small size images of SET12 and CBSD68 datasets. [Dedmari et al. \(2018\)](#) reconstructed Magnetic Resonance Imaging (MRI) by training a complex dense fully convolutional neural network (CDFNet). These above-mentioned proposed CV-CNN based architectures were either trained on specific data (i.e., MRI or X-ray, etc) or on very small-size images. [Shao et al. \(2023\)](#) designed an uncertainty-guided hierarchical frequency domain Transformer to learn both low and high-frequency components using a combination of real-valued CNNs and vision Transformer rather than CV-CNNs.

The authors in [Qiao et al. \(2021\)](#) exploit the learning of frequency features of SR-SIM images in the spatial domain of the channel attention module. Their learning of spatial and frequency information was purely based on RV-CNNs. Recently, [Liu et al. \(2023\)](#) proposed a dual-domain learning strategy for end-to-end SIM reconstruction using CV-CNNs. Their novel architecture involves the repetitive conversion of feature maps from the frequency domain to the spatial domain prior to the addition

operation with spatial features in each dual-domain residual block. Due to this repeated transformation of features, this block-wise approach could lead to significantly higher computational cost and also fail to exploit the full potential of CV-CNN layers (Liu et al., 2023). In contrast, in our architectures, the spatial and frequency branches are designed in a completely parallel scheme with all of their corresponding components to cover a wide range of local and global spatial and frequency features. Moreover, our approach is not limited to the end-to-end reconstruction of SR-SIM, but provides a versatile method for denoising images of all kinds.

The incorporation of complex operations in the CV-CAN, CV-DDAN, and CV-SRN results in the addition of phase information, which led to easier optimization (Nitta, 2002), better performance, and improved generalization ability. The CV-CNNs are mostly dense and contain more training parameters than the RV-CNNs due to their characteristics. According to the evaluated training times, the CV-DDAN requires twice the time of its real-valued counterpart, such as Red-fairSIM. However, excellent performance is more important than computational efficiency in the field of biomedical sciences.

6 Conclusions

We demonstrated that our proposed CV-CNN based serial (CV-CAN) and parallel (CV-DDAN) architectures provide mostly denoising results which are superior to purely real-valued approaches. Similarly, the CV-SRN produces acceptable results for the super-resolution task. We have also remarked that the CV-SRN is limited to super-resolution tasks and fails when used in conjunction with denoising. The CV-Atten-UNet is able to suppress and eliminate noise from high frequency components, but an overall image degradation is observed throughout the denoising process. This raises concerns about the ability of pure CV-CNNs to perform image denoising. Therefore, a main result of our findings is that the fusion of spatial and frequency information by simultaneously processing spatial and complex features in architecture such as CV-DDAN is the only way forward to improve the image quality and reduce the loss of high frequency information in the Fourier spectrum. In addition, the real and complex-valued attention gates were also useful for effectively learning local and global frequency features.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://gigadb.org/dataset/102461>.

Ethics statement

Ethical approval was not required for the studies involving humans because the Tubulin and Vesicle datasets used in this study are publicly available and do not necessitate ethical approval. The datasets are located at the following link: <http://gigadb.org/dataset/102461>. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed

consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements. Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because the liver sinusoidal endothelial cells (LSECs) used in this study were gifted by the Department of Medical Biology at "The Arctic University of Norway", Tromsø, Norway under the European Innovation Council PATHFINDER Open program.

Author contributions

ZS: Writing – original draft, Writing – review & editing. MM: Data curation, Funding acquisition, Resources, Writing – review & editing. WH: Data curation, Resources, Writing – review & editing. HO: Data curation, Resources, Writing – review & editing. BH: Supervision, Writing – review & editing. TH: Funding acquisition, Supervision, Writing – review & editing. WS: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. WS acknowledges funding by the project "SAIL: SustAInable Lifecycle of Intelligent SocioTechnical Systems". SAIL is receiving funding from the program "Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of North Rhine-Westphalia (grant No. NW21-059B). The sole responsibility for the content of this publication lies with the authors. TH acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Science Foundation)—project number 415832635. TH also acknowledges the European Union's European Innovation Council PATHFINDER Open program under grant agreement No. 101046928. MM acknowledges funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 752080.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Akramifard, H., Firouzmand, M., and Moghadam, R. A. (2012). Extracting, recognizing, and counting white blood cells from microscopic images by using complex-valued neural networks. *J. Med. Signals Sens.* 2:169. doi: 10.4103/2228-7477.112144
- Barbieri, L., Colin-York, H., Korobchevskaya, K., Li, D., Wolfson, D. L., Karedla, N., et al. (2021). Two-dimensional TIRF-SIM-traction force microscopy (2d TIRF-SIM-TFM). *Nat. Commun.* 12:2169. doi: 10.1038/s41467-021-22377-9
- Barrachina, J. A., Ren, C., Morisseau, C., Vieillard, G., and Ovarlez, J.-P. (2021). "Complex-valued vs. real-valued neural networks for classification perspectives: an example on non-circular data," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Toronto, ON: IEEE), 2990–2994.
- Barrachina, J. A., Ren, C., Morisseau, C., Vieillard, G., and Ovarlez, J.-P. (2022). Comparison between equivalent architectures of complex-valued and real-valued neural networks-application on polarimetric sar image segmentation. *J. Signal Process. Syst.* 95, 57–66. doi: 10.1007/s11265-022-01793-0
- Bassey, J., Qian, L., and Li, X. (2021). A survey of complex-valued neural networks. *arXiv [preprint]*. doi: 10.48550/arXiv.2101.12249
- Belthangady, C., and Royer, L. A. (2019). Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* 16, 1215–1225. doi: 10.1038/s41592-019-0458-z
- Benvenuto, N., and Piazza, F. (1992). On the complex backpropagation algorithm. *IEEE Transact. Signal Process.* 40, 967–969. doi: 10.1109/78.127967
- Chen, X., Li, B., Jiang, S., Zhang, T., Zhang, X., Qin, P., et al. (2021). Accelerated phase shifting for structured illumination microscopy based on deep learning. *IEEE Transact. Comp. Imaging* 7, 700–712. doi: 10.1109/TCI.2021.3093788
- Chiheb, T., Bilaniuk, O., and Serdyuk, D. (2017). "Deep complex networks," in *International Conference on Learning Representations*. Available at: <https://openreview.net/forum?id=H1T2hmZAb> (accessed September 01, 2023).
- Dedmari, M. A., Conjeti, S., Estrada, S., Ehses, P., Stöcker, T., and Reuter, M. (2018). "Complex fully convolutional neural networks for mr image reconstruction," in *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1* (Granada: Springer), 30–38.
- Demmerle, J., Innocent, C., North, A. J., Ball, G., Müller, M., Miron, E., et al. (2017). Strategic and practical guidelines for successful structured illumination microscopy. *Nat. Protoc.* 12, 988–1010. doi: 10.1038/nprot.2017.019
- Ding, T., and Hirose, A. (2013). "Fading channel prediction based on complex-valued neural networks in frequency domain," in *2013 International Symposium on Electromagnetic Theory (Hiroshima: IEEE)*, 640–643.
- Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision (Zurich: Springer)*, 184–199.
- Foreman, M. R. (2012). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular*. Cambridge, UK: Cambridge University Press.
- Gao, S., He, L., and Li, Y. (2024). Biologically inspired image invariance guided illuminant estimation using shallow and deep models. *Exp. Syst. Appl.* 249:123391. doi: 10.1016/j.eswa.2024.123391
- Georgiou, G. M., and Koutsougeras, C. (1992). Complex domain backpropagation. *IEEE Transact. Circ. Syst.* 39, 330–334. doi: 10.1109/82.142037
- Gonzalez, R. C. (2009). *Digital Image Processing*. Wisconsin-Madison: Pearson Education India.
- Gu, S., Sang, N., and Ma, F. (2012). "Fast image super resolution via local regression," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (Tsukuba: IEEE), 3128–3131.
- Gustafsson, M. G. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* 198, 82–87. doi: 10.1046/j.1365-2818.2000.00710.x
- Gustafsson, M. G., Shao, L., Carlton, P. M., Wang, C. R., Golubovskaya, I. N., Cande, W. Z., et al. (2008). Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* 94, 4957–4970. doi: 10.1529/biophysj.107.120345
- Hafiz, A. R., Al-Nuaimi, A. Y., Amin, M., and Murase, K. (2015). Classification of skeletal wireframe representation of hand gesture using complex-valued neural network. *Neural Process. Lett.* 42, 649–664. doi: 10.1007/s11063-014-9379-0
- He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168
- Heintzmann, R., and Huser, T. (2017). Super-resolution structured illumination microscopy. *Chem. Rev.* 117, 13890–13908. doi: 10.1021/acs.chemrev.7b00218
- Hirose, A., and Yoshida, S. (2012). Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. *IEEE Transact. Neural Netw. Learn. Syst.* 23, 541–551. doi: 10.1109/TNNLS.2012.2183613
- Huang, L., Wang, W., Chen, J., and Wei, X.-Y. (2019). "Attention on attention for image captioning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (Seoul)*, 4634–4643.
- Huang, X., Fan, J., Li, L., Liu, H., Wu, R., Wu, Y., et al. (2018). Fast, long-term, super-resolution imaging with hessian structured illumination microscopy. *Nat. Biotechnol.* 36, 451–459. doi: 10.1038/nbt.4115
- Jetley, S., Lord, N. A., Lee, N., and Torr, P. H. (2018). "Learn to pay attention," in *International Conference on Learning Representations (Vancouver, BC)*.
- Jin, L., Liu, B., Zhao, F., Hahn, S., Dong, B., Song, R., et al. (2020). Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat. Commun.* 11, 1–7. doi: 10.1038/s41467-020-15784-x
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386
- Lee, C., Hasegawa, H., and Gao, S. (2022). Complex-valued neural networks: a comprehensive survey. *IEEE/CAA J. Automat. Sinica* 9, 1406–1426. doi: 10.1109/JAS.2022.105743
- Li, C., Li, L., and Qi, J. (2018). "A self-attentive model with gate mechanism for spoken language understanding," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (Brussels)*, 3824–3833.
- Li, X., Xiao, J., Zhou, Y., Ye, Y., Lv, N., Wang, X., et al. (2020). Detail retaining convolutional neural network for image denoising. *J. Vis. Commun. Image Represent.* 71:102774. doi: 10.1016/j.jvcir.2020.102774
- Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K. (2017). "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (Honolulu, HI)*, 136–144.
- Liu, T., Liu, J., Li, D., and Tan, S. (2023). Improving reconstruction of structured illumination microscopy images via dual-domain learning. *IEEE J. Select. Top. Quant. Electron.* 29, 1–12. doi: 10.1109/JSTQE.2023.3279341
- Mao, X., Shen, C., and Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Adv. Neural Inf. Process. Syst.* 29, 2810–2818. doi: 10.5555/3157382.3157412
- Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 2 (Vancouver, BC: IEEE)*, 416–423.
- Mnih, V., Heess, N., and Graves, A. (2014). Recurrent models of visual attention. *Adv. Neural Inf. Process. Syst.* 27, 2204–2212. doi: 10.5555/2969033.2969073
- Müller, M., Mönkemöller, V., Hennig, S., Hübner, W., and Huser, T. (2016). Open-source image reconstruction of super-resolution structured illumination microscopy data in imagej. *Nat. Commun.* 7:980. doi: 10.1038/ncomms10980
- Nitta, T. (2002). "On the critical points of the complex-valued neural network," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02, Vol. 3 (Orchid: IEEE)*, 1099–1103.
- Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention U-Net: learning where to look for the pancreas. *arXiv [preprint]*. doi: 10.48550/arXiv.1804.03999
- Pham, M. T., Nguyen, V. Q., Hoang, C. D., Vo, H. L., Phan, D. K., and Nguyen, A. H. (2021). "Efficient complex valued neural network with fourier transform on image denoising," in *The 5th International Conference on Future Networks and Distributed Systems (Dubai United Arab Emirates)*, 48–57.
- Popa, C.-A., and Cernăzanu-Glăvan, C. (2018). "Fourier transform-based image classification using complex-valued convolutional neural networks," in *International Symposium on Neural Networks (Minsk: Springer)*, 300–309.
- Qiao, C., Li, D., Guo, Y., Liu, C., Jiang, T., Dai, Q., et al. (2021). Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* 18, 194–202. doi: 10.1038/s41592-020-01048-5
- Quan, Y., Chen, Y., Shao, Y., Teng, H., Xu, Y., and Ji, H. (2021). Image denoising using complex-valued deep cnn. *Pattern Recognit.* 111:107639. doi: 10.1016/j.patcog.2020.107639
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., et al. (2019). "On the spectral bias of neural networks," in *International Conference on Machine Learning (Long Beach, CA: PMLR)*, 5301–5310.
- Rawat, S., Rana, K., and Kumar, V. (2021). A novel complex-valued convolutional neural network for medical image denoising. *Biomed. Signal Process. Control* 69:102859. doi: 10.1016/j.bspc.2021.102859

- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (Munich: Springer), 234–241.
- Schermelleh, L., Ferrand, A., Huser, T., Eggeling, C., Sauer, M., Biehlmaier, O., et al. (2019). Super-resolution microscopy demystified. *Nat. Cell Biol.* 21, 72–84. doi: 10.1038/s41556-018-0251-8
- Schreier, P. J., and Scharf, L. L. (2010). *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge: Cambridge University Press.
- Shah, Z., Müller, M., Hübner, W., Wang, T., Telman, D., Huser, T., et al. (2023). Evaluation of swin transformer and knowledge transfer for denoising of super-resolution structured illumination microscopy data. *Gigascience* 16:plae038. doi: 10.1093/gigascience/giad109
- Shah, Z. H., Müller, M., Hammer, B., Huser, T., and Schenck, W. (2022). "Impact of different loss functions on denoising of microscopic images," in *2022 International Joint Conference on Neural Networks (IJCNN)* (Padua: IEEE), 1–10.
- Shah, Z. H., Müller, M., Wang, T.-C., Scheidig, P. M., Schneider, A., Schüttpehl, M., et al. (2021). Deep-learning based denoising and reconstruction of super-resolution structured illumination microscopy images. *Photon. Res.* 9, B168–B181. doi: 10.1364/PRJ.416437
- Shao, M., Qiao, Y., Meng, D., and Zuo, W. (2023). Uncertainty-guided hierarchical frequency domain transformer for image restoration. *Knowl. Based Syst.* 263:110306. doi: 10.1016/j.knsys.2023.110306
- Smith, C. S., Slotman, J. A., Schermelleh, L., Chakrova, N., Hari, S., Vos, Y., et al. (2021). Structured illumination microscopy with noise-controlled image reconstructions. *Nat. Methods* 18, 821–828. doi: 10.1038/s41592-021-01167-7
- Ströhl, F., and Kaminski, C. F. (2016). Frontiers in structured illumination microscopy. *Optica* 3, 667–677. doi: 10.1364/OPTICA.3.000667
- Timofte, R., De Smet, V., and Van Gool, L. (2013). "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision* (Sydney, NSW), 1920–1927.
- Timofte, R., De Smet, V., and Van Gool, L. (2014). "A+: adjusted anchored neighborhood regression for fast super-resolution," in *Asian Conference on Computer Vision* (Singapore: Springer), 111–126.
- Tran, L. D., Nguyen, S. M., and Arai, M. (2020). "Gan-based noise model for denoising real images," in *Proceedings of the Asian Conference on Computer Vision (ACCV)* (Kyoto).
- Tsuzuki, H., Kugler, M., Kuroyanagi, S., and Iwata, A. (2013). An approach for sound source localization by complex-valued neural network. *IEICE Trans. Inf. Syst.* 96, 2257–2265. doi: 10.1587/transinf.E96.D.2257
- Virtue, P., Stella, X. Y., and Lustig, M. (2017). "Better than real: complex-valued neural nets for mri fingerprinting," in *2017 IEEE International Conference on Image Processing (ICIP)* (Beijing: IEEE), 3953–3957.
- Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. (2017). "Residual attention network for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 3156–3164.
- Xie, J., Xu, L., and Chen, E. (2012). Image denoising and inpainting with deep neural networks. *Adv. Neural Inf. Process. Syst.* 25, 341–349. doi: 10.5555/2999134.2999173
- Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.-K., and Ren, F. (2020). "Learning in the frequency domain," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Seattle, WA), 1740–1749.
- Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E. K., and Yuille, A. L. (2018). "Recurrent saliency transformation network: incorporating multi-stage visual cues for small organ segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 8280–8289.
- Zhang, K., Zuo, W., and Zhang, L. (2018). Pfdnet: toward a fast and flexible solution for CNN-based image denoising. *IEEE Transact. Image Process.* 27, 4608–4622. doi: 10.1109/TIP.2018.2839891
- Zheng, B., Sun, R., Tian, X., and Chen, Y. (2018). S-net: a scalable convolutional neural network for jpeg compression artifact reduction. *J. Electron. Imaging* 27:043037. doi: 10.1117/1.JEI.27.4.043037
- Zheng, X., Zhou, J., Wang, L., Wang, M., Wu, W., Chen, J., et al. (2021). Current challenges and solutions of super-resolution structured illumination microscopy. *APL Photon.* 6:020901. doi: 10.1063/5.0038065
- Zuo, Q., Chen, S., and Wang, Z. (2021). R2AU-Net: attention recurrent residual convolutional neural network for multimodal medical image segmentation. *Sec. Commun. Netw.* 2021:6625688. doi: 10.1155/2021/6625688
- Zuo, W., Zhang, K., and Zhang, L. (2018). "Convolutional neural networks for image denoising and restoration," in *Denoising of Photographic Images and Video* (Salt Lake City, UT: Springer), 93–123.