Multimodal-Boost: Multimodal Medical Image Super-Resolution Using Multi-Attention Network With Wavelet Transform

Fayaz Ali Dharejo, Muhammad Zawish, Farah Deeba, Yuanchun Zhou, Kapal Dev, Sunder Ali Khowaja, and Nawab Muhammad Faseeh Qureshi

Abstract—Multimodal medical images are widely used by clinicians and physicians to analyze and retrieve complementary information from high-resolution images in a non-invasive manner. Loss of corresponding image resolution adversely affects the overall performance of medical image interpretation. Deep learning-based single image super resolution (SISR) algorithms have revolutionized the overall diagnosis framework by continually improving the architectural components and training strategies associated with convolutional neural networks (CNN) on low-resolution images. However, existing work lacks in two ways: i) the SR output produced exhibits poor texture details, and often produce blurred edges, ii) most of the models have been developed for a single modality, hence, require modification to adapt to a new one. This work addresses i) by proposing generative adversarial network (GAN) with deep multi-attention modules to learn high-frequency information from low-frequency data. Existing approaches based on the GAN have yielded good SR results; however, the texture details of their SR output have been experimentally confirmed to be deficient for medical images particularly. The integration of wavelet transform (WT) and GANs in our proposed SR model addresses the aforementioned limitation concerning textons. While the WT divides the LR image into multiple frequency bands, the transferred GAN uses multi-attention and upsample blocks to predict high-frequency components. Additionally, we present a learning method for training domain-specific classifiers as perceptual loss functions.

Index Terms—Attention modules, generative adversarial network, multimodality data, super-resolution, transfer learning, wavelet transform

1 INTRODUCTION

Medical images of high quality are critical in early, fast and accurate diagnosis in the current clinical processes. Typically, spatial resolution of medical images is degraded by factors such as imaging modalities and acquisition time.

- Fayaz Ali Dharejo, Farah Deeba, and Yuanchun Zhou are with the Computer Network Information Center, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing 100190, China.
  E-mail: {fayazdharejo, deeba, zyc}@cnic.cn.
- Muhammad Zawish is with the Walton Institute for Information and Communication Systems Science, Waterford Institute of Technology, X91 K0E Waterford, Ireland. E-mail: muhammad.zawish@wiltons institute.ie.
- Kapal Dev is with the Department of Institute of intelligent systems, University of Johannesburg, Johannesburg 2092, South Africa.
  E-mail: kapal.dev@ieee.org.
- Sunder Ali Khowaja is with the Faculty of Engineering and Technology, University of Sindh, Jamshoro 76080, Pakistan.
  E-mail: sandar.ali@usindh.edu.pk.
- Nawab Muhammad Faseeh Qureshi is with the Department of Computer Education, Sungkyunkwan University, Seoul 03063, Korea.
  E-mail: faseeh@skku.edu.

Manuscript received 25 September 2021; revised 23 May 2022; accepted 8 July 2022. Date of publication 18 July 2022; date of current version 8 August 2023.

To recover the distorted resolution, super-resolution (SR) methods are widely adopted on digital images as post-processors eluding the additional scanning costs [1]. Image SR is one of the prominent low-level vision problems in the computer vision domain. The aim of image SR is to reconstruct a high resolution (HR) image from a corresponding low resolution (LR) image. SR techniques are applied on both single and multiple images simultaneously subject to the input and output criteria. In this work, we study the Single-Image-Super-Resolution (SISR) for the application of medical images. In contrast to classical SISR applications, medical imaging SISR is captious as it is often followed by precise tasks of segmentation, classification or diagnosis [1], [2]. Thus, it becomes imperative to produce methods which could not only preserve sensitive information but also magnify the structures of interest efficiently.

Traditional SR based techniques such as interpolation [3] or reconstruction [4] based could not suffice for medical image SISR tasks. Interpolation methods such as linear, bicubic [3] and lanczos [5] often fail to reconstruct the high-frequency information which eventually results in blurred and smooth edges. In contrast, reconstruction based algorithms efficiently utilize prior local [6], global [7], and sparse [8] knowledge to efficiently reconstruct the HR image. However, these methods can not efficiently simulate the non-linear transformation from LR space to HR space in dynamic scenes, hence the output SR image is distorted. Deep learning...
techniques have recently shown notable performance on various vision tasks such as image dehazing, object detection, activity recognition etc [9]. Convolutional Neural Networks (CNNs) and particularly Generative Adversarial Networks (GANs) have shown remarkable performance on various SR applications such as remote sensing and medical imaging. Deep Convolutional Neural Network for Super-Resolution (SRCNN) [10] laid the foundation of deep learning (DL) based SR methods, followed by a number of techniques involving powerful capabilities of CNNs. Due to phenomenal performance of residual blocks and dense blocks, several works such as, Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) [11] Very Deep Convolutional Networks (VDSR) [12], Fast Super-Resolution with Cascading Residual Network (RFASR) [13] and Deep Back-Projection Networks for Super-Resolution (DBPN) [14] were proposed for the SR problem. These end to end models are deeper and complex thus requiring a large number of LR and corresponding HR pairs to get trained [15]. Moreover, these deep networks could not provide photo realistic results despite the high performance of Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). Thus, GAN based networks accompanied by attention mechanisms [16] were introduced to produce perceptually realistic outputs closer to the ground truth HR image. However, aforementioned deep learning based SR techniques have shown sub-optimal performance on medical images particularly for different types of modalities. Several medical imaging analysis tasks such as tumor segmentation [17], [18], [19] and lesion detection [20] are essentially hindered by the substandard HR output with blurred edges. Using a Multi-Temporal Ultra-Dense Memory Network, Peng Ye et al. [21] proposed a new method based on video SR. Previous video SR methods mostly use a single memory module and a single channel structure, which does not fully capture the correlations between frames specific to video. To achieve video -resolution, this paper presents a multi-temporal ultra-dense. Yiqun Mei et al. [22] developed an efficient SR method based on Cross-Scale Non-Local Attention Module (CS-NL) incorporating recurrent neural networks to handle an inherent property of images: cross-scale correlation of features. Afterward, Kui Jiang et al. [23] developed a hierarchical dense recursive network with SR. Feature representations were made richer thanks to dense hierarchical connections. Additionally, it facilitates the lightweight SR model by providing Reasonable design and parameter sharing.

In this work, we aim to overcome aforementioned problems in multimodal medical imaging SISR by proposing a combination of multi-attention GAN with wavelet subbands. Wavelet transform (WT) [24] has a unique ability of extracting useful multi-scale features with the help of sparse subbands. Since WT accompanied with GAN has been widely adopted for several applications including remote sensing for spatio-temporal SR [25] and dehazing [26], it would be interesting to utilize its potential for the medical imaging SR. In the first step, we derive the WT based four subbands of the given LR image using a 2D discrete wavelet transform (DWT) from Haar family of wavelet which eventually replaces LR input with low dimensional input space preserving the key information. These subbands are then fed into the proposed GAN involving multiple attention and upsample blocks which produces an improved HR wavelet component for each corresponding subband. Lastly, the HR image is reconstructed using 2 dimensional inverse discrete wavelet transform (2D-IDWT). Inspired from this [27], we use a perceptual loss network along with GAN to further boost the performance of our proposed method in term of qualitative and qualitative. Moreover, a challenging and time-consuming task in medical imaging SR is to modify the models designed for one modality to fit a new modality. We overcome this by introducing the transfer learning technique which efficiently adapts to unseen data with the help of pre-learned knowledge. As a result, it becomes feasible to adapt on datasets of different modalities along with the reliable results. In summary, the following are the contributions of our methodology:

- We proposed a novel wavelet-based multimodal and multi-attention GAN framework for medical imaging SR called Multimodal-Boost. The Multimodal-Boost can learn end-to-end residual mapping between low and high resolution wavelet subbands. The key advantage of utilizing wavelet transform is that it helps restoring rich content from images by accurately extracting missing details. WT give high-frequency information in a variety of directions that are most likely to result from (horizontal, vertical, and diagonal edges).

- We trained the perceptual network on VGG-16 to improve the SR results, as shown in the Fig. 7. Using perceptual loss, we boost the network and improve SR performance. We use a transfer-learning technique to deal with inadequate training for SR medical applications. Our algorithm is trained on DIV2K [28] dataset prior to getting evaluated on multimodal medical datasets.

- Unlike many previous Depp Neural networks (DNN) medical image SR approaches [29], [30], [31] [32], our method employs multimodality data into a single model, which reduces the cost of adding new SR objectives such as detection and classification tasks. The results reveal that the proposed method outperforms existing methods in objective and subjective evaluation as shown in Fig. 1.

The rest of the paper is drawn as follows. Section 2 provides an overview of the related works. The proposed The multimodal-Boost approach is detailed in Section 3. Section 4 contains the results and discussion. Finally, in Section 5, the conclusion is given.

2 Literature Work

SR is undoubtedly a classical low-level vision problem which has been highly studied with emerging deep learning based solutions for a variety of applications. The problem of SR can be categorised into either single image SR or multi-ple image SR, our focus in this paper is on single image SR. Thus, in this section we review the state-of-the-art SR methods for both general and medical imaging application.

In the domain of DL based SR, SRCNN by Dong et al. [10] was the very first technique incorporating deep CNNs for the SR task. Following that several works were proposed
with a typical pipeline of feature extraction, non-linear mapping and HR image reconstruction. However, such deeper level and complex architectures failed to extract the important multi-level features of a given LR image. Liang et al. [33] identified the importance of edge based features for enhancing the SR task using sobel edge priors of LR input for training a deep SR model. Later, VDSR [12], EDSR [11], and RFAASR [13] were proposed with the use of very deep networks involving residual skip connections and residual feature aggregation techniques to enhance the SR output. However, these methods did not yield admissible SSIM and PSNR due to deterioration of visual perception which resulted in over-smoothed SR output. Ledig et al. [34] improved the visual perception in SR problem by proposing a GAN based network namely SRGAN by combining adversarial and content loss of generated HR output. Moreover, Meta-SR [35] and DBPN [14] achieved improved results where the former proposed a solution for high magnification and arbitrary magnification SR and the latter used a deep back-projection network to generate clear HR output.

While all these methods work well on natural datasets, their performance is not admissible for medical imaging tasks. The methods on pixel space optimization [10], [11], [13] generate a smooth output which lacks the important information, while the methods on feature space optimization [34] could not generate realistic enough outputs. For this reason, these methods are not well suited for a critical problem like medical imaging SR. Although there exist several studies to tackle medical imaging SR with the techniques such as cycle GAN [36], U-Nets [37], 3D convolutions [38] and attention based networks [39], none of them achieved acceptable clinical ability. Moreover, most of these works are suitable only for single modality and their performance degrades when doing inference on new modality dataset. Therefore, in this paper we aim to address the shortcomings of above works by utilizing the inherent capabilities of discrete wavelet analysis on embedding the LR input image which is then provided to multi-attention and upscale blocks of proposed GAN network to produce the corresponding HR output. It can be seen from results that our model efficiently utilizes the transfer learning technique to produce clear HR output on multiple medical imaging modalities.

3 METHODOLOGY

3.1 SR Based on the DWT

For low-level vision tasks like image denoising and SR, choosing the proper wavelet can be difficult. Medical diagnostics, in particular, demands HR images with improved contextual features to provide better patient care. The WT stores detailed information about an image in many orientations to make the most of it. At each level of decomposition, the 2D-DWT produces four subbands that correspond to distinct frequency components and are referred to as approximate subbands, horizontal, vertical, and diagonal, containing complete edges information. Each subband contains image data with distinct characteristics that are important enough to offer precise image features.

In practice, DWT is being used to send the input signal via the low-pass L(e) and high-pass filters H(e). The input signal’s approximation and accuracy are then reduced by 2. L(e) and H(e), are the Haar wavelets’ definitions:

$$L(e) = \begin{cases} 1, & e = 0, \\ 0, & \text{otherwise} \end{cases}$$

$$H(e) = \begin{cases} 1, & e = 0, \\ -1, & e = 0, \\ 0, & \text{otherwise} \end{cases}$$

Generally, an image I(x,y), where xth and yth pixel values are in columns. The average components of the 2D-DWT are represented by four subbands: approximation, vertical, horizontal, diagonal, LL,LH, HL, HH.

When we use 1-level 2D-DWT to an LL input image to predict the HL, LH, and HH subbands, we get the missing information features of the LL image, as shown in the figure. To obtain SR results, we employ 2D-IDWT to acquire the image missing features information. The coefficients of 2D-IDWT can be computed as below from the Haar wavelet.

$$\begin{cases} A = a + b + c + d \\ B = a - b + c - d \\ C = a + b - c - d \\ D = a - b - c + d \end{cases}$$

where A, B, C, D, and a, b, c, d are the subbands pixel values. Interpolation-based SR approaches are limited in their capacity to reconstruct information appropriately. The result is not accurate because high-frequency domain information cannot be adequately recovered during SR. To improve super-resolved image performance, edges must be retained.

The DWT was employed to keep the image edges features and texture details in high-frequency; the 2-DWT decomposed the I_L image into the LL, LH, HL, HH subbands. Conventional DWT-based SR approaches combine DWT with Non-DL SR methods [25], as shown in Fig. 3. Let $I_G$ denote the image of the ground-truth HR in $m \times n$ by, and $I_L$ symbolize LR in size $m \times n$ by, and $I_L$ by the same scale factor (s). Authorize $I_{L,R}$ to represent $I_L$ up-scale ability, with $m \times n$ size, and to display the HR image reconstructed, with $m \times n$ size, via bicubic interpolation.

1) The four different subbands are achieved form LR input image, as shown in Fig. 2., i.e., $CT_{LL}$, $CT_{LH}$, $CT_{HL}$, $CT_{HH}$ $CT_{LR}$.
where, $G$; $o$, $x$ indicates how much the feature maps are transferred to features. $R$ is a scalar learnable parameter initially set to 0, $P_{\text{conv}}$ and $R_{\text{up}}$ are implemented as 1x1 convolutions. Where $C$ is the number of feature maps. The attention block can be expressed as, $\sigma_i=\{o_1, o_2, \ldots, o_j, \ldots, o_N\}$ where, $O_j = V\left(\sum_{i=1}^{N} \beta_{i,j}h(x_i)\right) , h(x_i) = W_hX_i, V(x_i) = W_hX_i$ (4)

The learned weight matrices $W_g \in \mathbb{R}^{L \times C}$, $W_f \in \mathbb{R}^{L \times C}$, $W_h \in \mathbb{R}^{L \times C}$ and $W_i \in \mathbb{R}^{L \times C}$ are implemented as 1x1 convolutions. We used varied channel numbers of $L=C/K$, in all of our experiments, where $k=1,2,4,8$ thus after a few training epochs on DIV2K dataset, we chose $k=8(i.e., L = \frac{C}{8})$ and found no significant performance loss. We also add back the input feature map after multiplying the output of the attention layer by a scale parameter. As a result, the final outcome is as follows:

$$Y_i = \alpha a_i + x_i$$ (5)

Where $\alpha$ is a scalar learnable parameter initially set to 0, allowing the network to depend on local inputs first and then progressively learn to give non-local data to more weights. The attention block used in our method is shown in Fig. 5a. Upsampling attention blocks, which tend to improve the resolution for CT images on wavelet subbands input features, are another attention block employed in our predicted part. The spatial resolution of the feature map was further improved by performing local neighborhood interpolation on the wavelet subband input features. These are fed to the leaky convolution ReLU, and the final feature map ($f_{\text{ Upsample}} \in \mathbb{R}^{L \times C \times w}$) Upsample attention block can be expressed as:

$$F_{\text{map}} = \text{Sigmoid(Conv}(1 \times 1)) \times f_{\text{Upsample}}$$ (6)

$F_{\text{map}}$ is referred to output of pixel attention module and ($\text{Conv}(1 \times 1)$) represent the convolution operation with kernel size 1x1 so we can say the $F_{\text{map}} \in \mathbb{R}^{W \times L \times 1}$.

The channel number of the feature maps is reduced to 1 after ($\text{Conv}(1 \times 1)$), and then the ($f_{\text{Upsample}} \in \mathbb{R}^{L \times C \times 1}$) is obtained after sigmoid operation. The values of the feature map fall between 0 and 1. The feature map can be correctly
changed at the component level to achieve superior SR results. The Upsample attention block is given in Fig. 5b.

### 3.3 The Reconstruction Part

To reconstruct the HR super-resolved image $CT_{HR}$, we use a 2D-IDWT inverse wavelet onto four components $CT_{LL}$, $CT_{LH}$, $CT_{HL}$, and $CT_{HH}$.

### 3.4 VGG Based Perceptual Loss

Perceptual loss is one of the best metrics for measuring image similarity when CNN algorithms are applied to the input image. It is utilized in various applications such as image denoising, image dehazing, image translation, and image SR. Compared to Mean Squared Error (MSE) loss [27], the perceptual loss is more robust to several problems such as...
over-flattening and distortion [40][41]; hence for image SR, 
the perceptual loss is more powerful to search spatial resolu-
tion similarities between two images. Many CNN networks 
employ VGG-loss to measure perceptual loss; however, 
VGG-11, VGG-16, and VGG-19 are pre-trained networks that 
achieved amazing results using natural image datasets [42].

VGG-loss is a term that can be stated as follows:

$$L_{VGG} = \mathbb{E}_{(I_{LR}, I_{HR})} \left[ \frac{\| VGG(g(I_{LR})) - VGG(I_{HR}) \|_F^2}{DHW} \right]$$ (7)

where D; H; W denotes the computed tomography (CT) 
image depth, height, and width. VGG was trained for image 
classification using natural images; therefore, it generates 
features not relevant to CT image SR. This is one of the pos-
sible pitfalls with VGG-loss [43]. Due to the scarcity of 
labeled CT images, VGG for CT datasets is a challenging 
task. We proposed a trained perceptual network to handle 
this challenge to extract a compressed encoding from CT 
data input and reconstruct an image comparable to the orig-
inal image. In our case we used VGG-16 in the described 
perceptual loss network as shown in Fig. 4. The perceptual 
network comprises six convolution layers, each with 32, 32, 
64, 64, 128, 128 filters. There is also a maxpooling layer with 
a kernel size of 2 and a stride of 2. The perceptual loss con-
volution layers were utilized, followed by the ReLU activa-
tion function. In this design, we used a 33 filter size and a 
stride of 1. To extract the features, we trained the perceptual 
net-work to determine the perceptual loss. Fig. 7 shows the 
perceptual loss based on VVG-16.

$$L_{prec} = \mathbb{E}_{(I_{LR}, I_{HR})} \left[ \frac{\| y(g(I_{LR})) - y(I_{HR}) \|_F^2}{DHW} \right]$$ (8)

Where y is the pre-trained encoder network.

3.5 Model Optimization

The proposed multimodal multi-attention model loss is 
equal to the sum of the perceptual and generator-discrimi-
nator losses, as shown below:

$$\min_g \max_d L_{WGAN}(g, d) + \beta L_{perc}(g)$$ (9)

$L_{WGAN}(g, d)$ is a Wasserstein GAN [44], [45] optimizer, the 
weighted parameters are represented by the $\beta$, which claims 
to stand for substitution between WGAN-Loss and percep-
tual loss, and the letters g and d, which stand for generator 
and discriminator, respectively. $L_{perc}(g)$ is the perceptual 
loss. $L_{WGAN}(g, d)$ is in addition to the usual Wasserstein dis-
tance and the regularization gradient penalty, which is rep-
resented as,

$$\min_g \max_d L_{WGAN}(g, d) = ( - \mathbb{E}_{I_{HR}}[d(I_{HR})] +$$
$$\mathbb{E}_{I_{LR}}[d(g(I_{LR}))] + \lambda \mathbb{E}_{g}[\| \nabla I d(I_{LR}) \|_2 - 1]^2 ]$$
Where $E_a[b]$ is the expression of $b$ as a function of $a$, $\lambda$ represents the weighted parameter, $I$ denotes the generated and real images in uniform sampling from a range of [0,1], and $\nabla$ is the gradient.

4 RESULTS AND DISCUSSIONS

We introduced above a Multimodal-Boost framework for SR tasks in medical images. It comprises a new neural network for SR medical image reconstruction, pair-wise attention blocks, and a novel GAN-based loss function. The proposed model is efficient with texture details and realistic for viewing reconstructed images to a large extent. Since SR is an inverse problem with unpredictable solution [11], the output of the SR image contains substantially more information than the matching LR image. Each LR feature from wavelet subbands was regarded as a training sample during Multimodal-Boost model training and fed to the training phase. Loss functions are employed to ensure generated SR images as close to HR. Since working with LR images for diagnosis is too challenging, SR images are highly rated for medical applications, and they include enough information for radiologists to make precise conclusions. The information was stored in hidden layers by SR DNN models, which resulted in high frequency or HR images with superior texture and edge details. Medical images are too complicated to handle compared to natural images, making machine learning models relatively challenging. The higher magnification area in Figs. 6, 9, and 10 shows the differences between each approach. At the same time, Meta-SR achieves excellent performance than other techniques; it still struggles to generate pretty small textures, whereas our proposed method takes advantage of wavelet transform and obtains more excellent texture.

![Fig. 6. Visual results of our proposed method (Multimodal-Boost) compared to our State-of-art methods. We enlarge the particular region to notice the differences between the outcomes more clearly on CT teeth.](image)

![Fig. 7. VGG-16 Loss curves obtained on natural images.](image)
details. We also perform visual comparisons on other images using our Multimodal-Boost and other SR techniques, as seen in Figure "teeth-image" Fig. 6. All previous SR algorithms produce blurry outputs and fail to recover detailed information in the magnified image, while Meta-SR and our new approach Multimodal-Boost can only reconstruct a clean image with sharp lines. We evaluate the SR results of the image MRI Modality to support the multimodality and discover that SRCNN, EDSR, and Meta-SR all fail to recover the clear edges. It is indeed worth remembering that the texturing details they acquired are incorrect. CIRCLE-GAN and SR-ILLNN, on the other hand, can rebuild more compelling results that are consistent with ground truth, but they fail to provide more precise edge information. On the other hand, our technique does better when it comes to restoring edges and texture details.

4.1 Training Details
When training SR models with pairs of LR and HR images, downsampling is typically utilized. On the other hand, medical datasets have fixed in-plane resolutions of roughly 1 mm and lack ultra-high spatial resolutions. As a result, transfer learning is one of the methods for effectively training models using external datasets, which has proven to be very useful in remote sensing and medical applications. We used massive datasets DIV2K [28] with 2000 × 1400 pixels of good quality to train our proposed technique to use transfer learning. We randomly cropped 56 × 56 sub-pictures from DIV2K training images for training. Preliminary, we obtain the trained and optimized network by pre-training the proposed model in a particularly defined manner. The pre-trained network is fine-tuned by selecting the Shenzhen Hospital datasets [46], containing 662 X-ray images used for training and testing. All images were scaled to 512 × 512 pixels, and three modalities, including Montgomery County X-ray, Teeth, and knee images, were chosen to test our model. MRI brain scans from the Calgary Campinas...
repository were another dataset we used to evaluate our proposed approach. This modality is produced using a 12-channel head-neck coil on an MR scanner (Discovery MR750; General Electric Healthcare, Waukesha, WI). All experiments have been carried out on a Windows-based machine with an Intel(R) Core(TM) i5-7300HQ CPU running at 3.40GHz and an NVIDIA GeForce GTX 1080-Ti graphics card. The setup also makes use of MATLAB 2019 with CUDA Toolkit and Anaconda. We have calculated time inference of each method and showed in Fig. 11.

### 4.2 Transfer Learning

Transfer learning is a technique that improves the performance of deep neural networks by utilizing knowledge learned from natural image data sets as initial training data. Due to the different distributions of two types of images, directly applying a trained model using natural datasets to medical images will not work, so transfer learning is a feasible solution. Transfer learning has various advantages such as:

- The ability to borrow high-frequency information from natural image datasets, which boosts the proposed method’s performance to reconstruct the HR image from LR image.
- It helps in faster model convergence.
- It improves the model accuracy.

To achieve high image resolution for diagnostic, our suggested Multimodal-Boost model leveraged transfer
learning to integrate shared and supplementary information from diverse modalities. In order to assess each modality and the detailed information inside the image patches, we used 16 to 128 batch sizes for the HR patches in model training. When the epochs in both modalities, CT-Images from Shenzhen Hospital dataset [46] and MRI brain scans from CalgaryCampinas, reach 180, the training process is finished. We train our model with ADAM optimizer [50] by setting $b_1 = 0.9$, $b_2 = 0.999$, and $\epsilon = 10^{-8}$. Our primary goal is to improve multimodal medical image quality. We trained the model with DIV2K before fine-tuning it with CT-image dataset, which contains 662 X-ray images. We also tested our model using MRI brain imaging so that the previous modality information can be used to reconstruct the next modality image. According to the proposed results, transfer learning knowledge considerably improved performance as shown in Figs. 12 and 13. The DIV2K dataset [28] is being used for training since it is a recently proposed high-quality 2 k resolutions image dataset for image enhancement. There are 800 training images, 100 validation image.

4.3 Comparison With State-of-Art Methods
We analyzed and compared the proposed method with deep learning models including SRCNN [47], EDSR [11], Meta-SR [35], CIRCLE-GAN [48], SR-ILLNN [49] and bicubic interpolation. These methods were designed for natural images DIV2K [28], which have much bigger dimensions than the medical images we used. We employed transfer learning and retrained the models with smaller chunks of medical image datasets to make them work more robustly. We used the same hardware and experiment conditions to make a fair comparison. We employed the MSE loss; however, it was not very reliable due to distortion and over-flattening. As a result, we proposed perceptual loss $L_{perc}(g)$ using VGG-16 training, and

![Fig. 12. Model loss over different experimental networks.](image1)

![Fig. 13. Network convergence over different loss functions on Wasserstein Estimation.](image2)

| Experiments | Generator Network | Loss Function | WGAN |
|-------------|-------------------|---------------|------|
| CNN-VGG     | CNN               | $L_{VGG}$     | no   |
| WGAN        | CNN               | $L_{WGAN}$    | yes  |
| Perceptual  | Multi-attention   | $L_{perc}$    | no   |
| WAGN-VGG    | CNN               | $L_{WGAN} + L_{VGG}$ | yes |
| WAGN-MA-P   | Multi-attention CNN | $L_{WGAN} + L_{perc}$ | yes |

| Networks Complexity and Inference Speed on Shenzhen Hospital Datasets |
|---------------------------------------------------------------|
| Methods            | Number of parameters (M) | Memory size (MB) | Inference time(s) | Number of $log_{10}$ flops |
|--------------------|---------------------------|------------------|-------------------|-----------------------------|
| SRCNN [47]         | 0.059                     | 13.68            | 0.0113            | 17.4                        |
| EDSR [11]          | 32.55                     | 266.78           | 2.178             | 24.2                        |
| Meta-SR [35]       | 4.075                     | 32.64            | 1.0156            | 25.1                        |
| GAN-CIRCLE [48]    | 55.88                     | 457.88           | 3.0172            | 24.8                        |
| SR-ILLNN [49]      | 0.441                     | 16.44            | 0.0123            | 28.1                        |
| Multimodal-Boost   | 21.68                     | 105.68           | 2.0189            | 28.3                        |

Multimodal-boost, in comparison to srdesnsenet and csgan, takes more training time while taking less time than edsr and circle-gan with transfer learning. Furthermore, unlike edsr and meta-sr, which only work with a single modality, it uses multi-modality data in a single model, lowering the cost of adding new sr tasks.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
when we improved our network with total loss $L_{\text{WGAN}}(g, d) + \beta L_{\text{perc}}(g)$, which is the sum of perceptual loss and generative discriminator loss function. The proposed model (Multimodal-Boost) is compared against state-of-the-art approaches in Table 2, and the visualization results are shown in Figs. 6, 9, and 10.

For comparison, we conduct experiments on numerous networks with varying configurations, presented in Table 1 and Fig. 12. We witnessed that WGAN-based losses converge faster than non-WGAN-based losses. Finally, combining WGAN and perceptual loss results in a shorter distance and greater convergence, which we apply to all losses in model optimization, as shown in Fig. 13.

### 4.4 Convergence

The convergence curve of VGG-loss in WGAN-MA-P has the fastest convergence across all types of approaches, as shown in Fig. 12. WGAN-MA-P loss is quite similar to WGAN-VGG, although the latter achieves lower. However, a reduced VGG loss does not always imply improved performance. According to our findings, the VGG-loss network’s spatial resolution diminishes its benefits while still causing some loss. Finally, we show how the WGAN-MA-P loss converges very quickly with shorter Wasserstein distances, resulting in state-of-the-art performance compared to other WGAN-based alternatives. The network complexity of methodologies is measured in terms of the number of parameters, memory, inference time, and number of flops as shown in Table 2. Compared to SRCNN [47], Meta-SR [35], SR-ILLNN [49], Multimodal-Boost has many parameters; however, it has fewer parameters than EDSR [11] and CIRCLE-GAN [48]. Furthermore, on the same hardware, the proposed method takes 18% longer to train than Meta-SR. As a result, further work on the architecture, such as model compression, should be done in the future.

### 4.5 Quantitative Analysis

We evaluated our approach on several attention blocks ($\rho = 2, 4, 8$) to investigate how well it performed. In terms of PSNR and SSIM, we found that the proposed Multimodal-Boost technique outperforms all others. When $\rho$ is increased, the proposed method performance degrades significantly, as illustrated in Fig. 8. Table 3 shows that the proposed strategy produces the maximum PSNR. For three images, SRCNN [47] has the lowest PSNR, although it is better than Bilinear interpolation. The proposed method achieves 1.54–4.89 greater PSNR index when compared with recent methods EDSR [11], Meta-SR [35], CIRCLE-GAN [48], and SR-ILLNN [49]. The quantitative performance metric is the peak signal-to-noise ratio (PSNR). The PSNR is defined as follows: Given a ground-truth image $S$ and its reconstructed image $\hat{S}$, with $M \times N$ pixels size,

$$\text{PSNR}(S, \hat{S}) = 10\log_{10} \frac{225^2}{\text{MSE}(S, \hat{S})}$$

The Structural Similarity Index Measure (SSIM) is a popular tool for assessing the quality of high-resolution reconstructions. The SSIM index can be expressed mathematically as,

$$\text{SSIM} = \frac{(2 \mu_x \mu_y + C_1)(2 \sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where $\mu_x, \mu_y$ are the average of $x$ and $y$ respectively, $\sigma_x^2, \sigma_y^2$ are the variance of $x$ and $y$ respectively, $\sigma_{xy}$ is the covariance of $x$ and $y$, $C_1, C_2$ are the constants.

SISR is a low-level image processing task that can be used in a number of contexts. In real-world circumstances, SISR’s time constraints are extremely stringent. We conduct tests to see how long each state-of-the-art method takes to run and then compare the results on Shenzhen Hospital datasets. This is depicted in Fig. 13.

## 5 Conclusion

In this study, we designed a multi-attention GAN framework with wavelet transform and transfer learning for multimodal SISR on medical images. This is the first time a multi-attention GAN has been employed in conjunction with a wavelet transform methodology for medical image SR. We also used transfer leaning, which works with multimodality data that decrease the cost of adding new SR challenging targets like disease classification. We trained...
our network on the high-resolution DIV2K dataset, then applied transfer learning to train and test medical images on the Shenzhen Hospital datasets of various modalities. Furthermore, we used GANs to train a perceptual loss function that can better super-resolve LR features, resulting in improved perceptual quality of the generated images. Due to the 2D-DWT properties, the reconstructed images are more accurate and have more texture information. The utility of transfer learning is that the pre-trained model is fine-tuned using medical images such as cardiac MR scans and CT scans. In particular, we evaluated our outcomes in terms of visual, quantitative, adversarial, and perceptual loss. The PSNR and SSIM measurements are the preliminary step in evaluating the SR approach, however they are insufficient for real-time applications. In the future, we will study how the proposed approach affects target tasks such as disease detection and classification, as well as minimizing the amount of parameters and flops, which will effectively reduce training and inference time.

**DATA AVAILABILITY**

We used the two popular datasets CT-images from Shenzhen Hospital datasets [46] and Public MRI data-set (https://sites.google.com/view/calgary-camnpanas-dataset/home).

**REFERENCES**

[1] F. Deeba, S. Kun, F. A. Dharejo, and Y. Zhou, “Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm,” *IET Image Process.*, vol. 14, no. 11, pp. 2365–2375, 2020.

[2] H. Gao, K. Xu, M. Cao, J. Xiao, Q. Xu, and Y. Yin, “The deep features and attention mechanism-based method to dish healthcare under social IoT systems: An empirical study with a handheld local-global network,” *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 1, pp. 336–347, Feb. 2022.

[3] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[4] H. Ji and C. Fermüller, “Wavelet-based super-resolution reconstruction: Theory and algorithm,” in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 295–307.

[5] C. E. Duchon, “Lanczos filtering in one and two dimensions,” *J. Appl. Meteorol. Climatol.*, vol. 18, no. 8, pp. 1016–1022, 1979.

[6] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin, “Super resolution using edge prior and single image detail synthesis,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2400–2407.

[7] C.-Y. Yang, J.-B. Huang, and M.-H. Yang, “Exploiting self-similarities for single frame super-resolution,” in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 497–510.

[8] W. Dong et al., “Hyperspectral image super-resolution via non-negative structured sparse representation,” *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[9] F. Deeba et al., “A novel image dehazing framework for robust vision-based intelligent systems,” *Int. J. Intell. Syst.*, 2021.

[10] C. Dong, C. C. Loy, K. He, and X. Yang, “Learning a deep convolutional network for image super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[11] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 136–141.

[12] Y. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[13] N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 252–268.

[14] M. Haris, G. Shakhnarovich, and N. Ukit, “Deep back-projection networks for super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1664–1673.

[15] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2020.

[16] M. Xu, Z. Wang, J. Zhu, X. Jia, and S. Jia, “Multi-attention generative adversarial network for remote sensing image super-resolution,” 2021, arXiv:2107.06536.

[17] M. Zawish, A. A. Sial, K. Ahmed, A. Khalil, and S. Memon, “Brain tumor segmentation in MRI images using chan-vech-technique in MATLAB,” in *Proc. Int. Conf. Comput. Electron. Elect. Eng.*, 2018, pp. 1–6.

[18] J. Xiao, H. Xu, H. Gao, M. Bian, and Y. Li, “A weakly supervised semantic segmentation network by aggregating seed cues: The multi-object proposal generation perspective,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 1s, pp. 1–19, 2021.

[19] H. Xiao, J. Xiao, K. Jiang, Z. Shao, and J. Ma, “Mult-temporal ultra dense memory network for video super-resolution,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2503–2516, Aug. 2020.

[20] Y. Mei, Y. Fan, Y. Zhou, H. T. Huang, and H. Shi, “Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5690–5699.

[21] K. Jiang, Z. Wang, P. Yi, and J. Jiang, “Hierarchical dense recursive network for image super-resolution,” *Pattern Recognit.*, vol. 107, 2020, Art. no. 107475.

[22] A. Abbate, J. Frankel, and P. Das, “Wavelet transform signal processing for dispersion analysis of ultrasonic signals,” in *Proc. IEEE Ultrasonics Symp. Proc. Int. Symp.*, 1995, pp. 751–755.

[23] F. A. Dharejo et al., “TWIST-GAN: Towards wavelet transform and transferred GAN for spatio-temporal single image super resolution,” arXiv:2104.10268.

[24] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, “DW-GAN: A discrete wavelet transform GAN for nonhomogeneous dehazing,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 203–212.

[25] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[26] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, “Ntire 2017 challenge on single image super-resolution: Methods and results,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 114–125.

[27] Y. Zhang and M. An, “Deep learning-and transfer learning-based super resolution reconstruction from single medical image,” *J. Healthcare Eng.*, vol. 2017, pp. 2040–2295, 2017.

[28] Z. Chen, X. Guo, P. Y. Woo, and Y. Yuan, “Super-resolution enhanced medical image diagnosis with sample affinity interaction,” *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1377–1389, May 2021.

[29] F. Deeba, S. Kun, F. A. Dharejo, and Y. Zhou, “Wavelet-based enhanced medical image super-resolution,” *IEEE Access*, vol. 8, pp. 37035–37044, 2020.

[30] Y. Romano, J. Isidoro, and P. Milanfar, “RAISR: Rapid and accurate image super-resolution,” *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 110–125, Jan. 2020.

[31] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, “Incorporating prior image information into cascaded residual networks for image super-resolution,” *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 3365–3387, Oct. 2020.

[32] Y. Romano, J. Isidoro, and P. Milanfar, “RAISR: Rapid and accurate image super-resolution,” *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 110–125, Jan. 2020.

[33] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, “Incorporating prior image information into cascaded residual networks for image super-resolution,” *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 3365–3387, Oct. 2020.

[34] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.

[35] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, “META-SR: A magnification-arbitrary network for super-resolution,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1575–1584.
[36] H. Liu, J. Liu, S. Hou, T. Tao, and J. Han, “Perception consistency ultrasound image super-resolution via self-supervised cycleGAN,” Neural Comput. Appl., vol. 16, pp. 1–11, 2021.

[37] J. Park, D. Hwang, K. Y. Kim, S. K. Kang, Y. K. Kim, and J. S. Lee, “Computed tomography super-resolution using deep convolutional neural network,” Phys. Med. Biol., vol. 63, no. 14, 2018, pp. 174017.

[38] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, “Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network,” in Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervention, 2018, pp. 91–99.

[39] X. He et al., “Super-resolution magnetic resonance imaging reconstruction using deep attention networks,” in Proc. Med. Img. Image Process. Int. Soc. Opt. Photon., vol. 11313, 2020, Art. no. 113132.

[40] H. Shan et al., “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” IEEE Trans. Med. Imag., vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[41] Z. Chen and Y. Tong, “Face super-resolution through Wasserstein GANs,” 2017, arXiv:1705.02438.

[42] Y. He et al., “AMC: Automl for model compression and acceleration on mobile devices,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 784–800.

[43] Y. Jo, S. Yang, and S. J. Kim, “Investigating loss functions for extreme super-resolution,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 424–425.

[44] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proc. Int. Conf. Mach. Learn., 2017, vol. 6, no. 1, pp. 214–223.

[45] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1125–1134.

[46] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, “Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,” Quantitative Imag. Med. Surg., vol. 5, no. 6, 2014, Art. no. 475.

[47] Y. He et al., “AMC: Automl for model compression and acceleration on mobile devices,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 784–800.

[48] Y. Jo, S. Yang, and S. J. Kim, “Investigating loss functions for extreme super-resolution,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 424–425.

[49] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proc. Int. Conf. Mach. Learn., 2017, vol. 6, no. 1, pp. 214–223.

[50] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1125–1134.

[51] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, “Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,” Quantitative Imag. Med. Surg., vol. 5, no. 6, 2014, Art. no. 475.

[52] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, Feb. 2016.

[53] C. You et al., “CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-circle),” IEEE Trans. Med. Imag., vol. 39, no. 1, pp. 188–203, Jan. 2020.

[54] S. Kim, D. Jun, B.-G. Kim, H. Lee, and E. Rhee, “Single image super-resolution method using CNN-based lightweight neural networks,” Appl. Sci., vol. 11, no. 3, 2021, Art. no. 1092.

[55] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

Fayaz Ali Dharejo received the BE degree in electronic engineering from the Quaid-e-Awam University of Engineering Science and Technology (QUEST), Pakistan and the MS degree in the School of Information and Software Engineering, University of Electronic Science and Technology of China (UESTC), in 2016 and 2018, respectively. Since September 2018, he is currently working toward the PhD degree in the Institute of Computer Network Information Center Chinese Academy of Sciences, University of Chinese Academy of Sciences. He has published more than 30 articles within few time in reputed ISI impact factor journals and in IEEE conferences, among them the top journals the ACM Transactions on Intelligent Systems and Technology, International Journal of Intelligent Systems, IEEE GRSL and etc. He is a member of professional bodies such as Pakistan Engineering Council (PEC), ACM, IEEE (USA), IEEE Societies, such as IEEE Geoscience and Remote Sensing Society, IEEE Computer Society, and IEEE Signal Processing Society. He is a member of the Technical Program Committee and Reviewer of several Scopus, Springer and ISI (Thomson Reuters) indexed conferences. His research interests include image processing, deep learning, remote sensing and machine learning.

Muhammad Zawish (Graduate Student Member, IEEE) received the BE in computer systems from the Mehran University of Engineering and Technology, Pakistan. He has worked as Lab engineer in Department of Software Engineering with PAFKET, Pakistan. He is currently working toward the PhD degree in computer science with the Walton Institute, Waterford Institute of Technology, Ireland. His research interests are deep learning for IoT.

Farah Deeba received the BE and ME degrees in the Department of Software Engineering, Mehran University of Engineering and Technology Jamshoro Pakistan, in 2010 and 2014, respectively. She has 3 years of experience in teaching at graduate and post-graduate level with Hamdard University Karachi Pakistan. Current-ly, she has completed her PhD degree from the the School of Information and Software Engineering, University of Electronic Science and Technology, China, in July 2020. Since September 2020 she is also doing post-doc with Computer network information Center, Chinese Academy of Sciences Beijing, China. She has served as a reviewer for many ISI (Impact factor) and Scopus indexed journals, Such as IEEE Transactions on Pattern Analysis and Machine Intelligence, Electronics Letters, IET Image Processing, IEE Access, etc. She is a member of professional bodies such as Pakistan Engineering Council (PEC), ACM, IEEE (USA), IEEE Societies, such as IEEE Geoscience and Remote Sensing Society, IEEE Computer Society and IEEE Signal Processing Society. Her research interest includes SR, sparse representation and deep learning.

Yuanchun Zhou (Member, IEEE) received the PhD from the Institute of Technology, Chinese Academy of Sciences, in 2006. He is a professor in the Computer Network Information Center, Chinese Academy of Sciences. His main research interests include data mining, data intelligence, massive data and data intensive computing. He has published more than 100 research publications.

Kapal Dev (Senior Member, IEEE) is senior research associate with the university of Johannesburg South Africa. Previously, he was a postdoctoral research fellow with the CONNECT Centre, School of Computer Science and Statistics, Trinity College Dublin (TCD). He worked as 5G Junior Consultant and Engineer with Altran Italia S.P.A, Milan, Italy on 5G use cases. He worked as lecturer with Indus university, Karachi back, in 2014. He is also working for OCEANS Network as head of Projects funded by European Commission. He was awarded the PhD degree by Politecnico di Milano, Italy under the prestigious fellowship of Erasmus Mundus fundedy by European Commission. His research interests include data and data intensive computing. He has published more than 30 articles within few time in reputed ISI impact factor journals and in IEEE conferences, among them the top journals the ACM Transactions on Intelligent Systems and Technology, International Journal of Intelligent Systems, IEEE GRSL and etc. He is a member of professional bodies such as Pakistan Engineering Council (PEC), ACM, IEEE (USA), IEEE Societies, such as IEEE Geoscience and Remote Sensing Society, IEEE Computer Society, and IEEE Signal Processing Society. He is a member of the Technical Program Committee and Reviewer of several Scopus, Springer and ISI (Thomson Reuters) indexed conferences. His research interests include image processing, deep learning, remote sensing and machine learning.
Sunder Ali Khowaja received the PhD degree in Industrial and Information Systems Engineering from the Hankuk University of Foreign Studies, South Korea. He is currently an Assistant Professor with the Department of Telecommunication Engineering, Faculty of Engineering and Technology, University of Sindh, Pakistan. He had the experience of working with multi-national companies as Network and RF Engineer from 2008-2011. He is having teaching, research and administrative experience of more than 10 years. He has published more than 30 research articles in national, international journals, and conference proceedings. He is also a regular reviewer of notable journals including IEEE Transactions on Industrial Informatics, IEEE Internet of Things Journal, IEEE Access, Electronic Letters, IET Image Processing, IET Signal Processing, Computer Communications, International Journal of Imaging Systems and Technology, Multimedia Systems, IET Networks, SN Applied Sciences, Artificial Intelligence Review, Computational and Mathematical in Medicine, Computers in Human Behavior, IET Image Processing Systems, IET Signal Processing Systems, and Mathematical Problems in Engineering. He has also served as a TPC member for workshops in A* conferences such as Globecom and Mobicom. His research interests include Data Analytics, Process Mining, and Deep Learning for Computer Vision applications and Emerging Communication Technologies.

Nawab Muhammad Faseeh Qureshi received the PhD degree in computer engineering from Sungkyunkwan University, South Korea, through a SAM-SUNG scholarship. He is an Assistant Professor with Sungkyunkwan University, Seoul, South Korea. He was awarded the 1st Superior Research Award from the College of Information and Communication Engineering based on his research contributions and performance during his studies. He has served as guest editor in more than 17 Journals. Also, he has served as General chair Workshop in many conferences that includes IEEE Wireless Communications and Networking Conference (WCNC2020) and IEEE Globecom 2020. In addition to that, he has served as Proceedings chair Global Conference on Wireless and Optical Technologies consecutively for the third time. He is a reviewer of various prestigious IEEE and ACM Journals. He serves in many leading IEEE and ACM conferences as a reviewer every year. He has evaluated several theses as external PhD thesis evaluators. Also, he has facilitated several institutes and conferences with keynotes and Webinars on Big data analysis and Modern Technology convergence. He is an active senior member of IEEE, ACM, KSII (Korean Society for Internet Information), and IEICE (Institute of Electronics, Information and Communication Engineers). His research interests include Big Data analytics, machine learning, deep learning, context-aware data processing of the Internet of Things, and cloud computing.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.