DOI: 10.21917/ijivp.2023.0442

ENHANCING IMAGE SUPER-RESOLUTION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

Archana Tomar and Harish Patidar
Department of Computer Science, Mandsaur University, India

Abstract

In computer vision, image super-resolution plays a pivotal role in improving the visual quality of low-resolution images, thereby enhancing various applications such as medical imaging, surveillance, and digital entertainment. The problem at hand involves the inherent limitations of conventional methods in restoring high-frequency information lost during image downscaling. This research aims to bridge this gap by leveraging DCNNs, exploiting their ability to learn complex mappings between low and high-resolution image spaces. This study addresses the challenge of image super-resolution through the application of Deep Convolutional Neural Networks (DCNNs). The research involves the design and training of a novel DCNN architecture tailored specifically for image super-resolution. We employ a large dataset of low and high-resolution image pairs to facilitate supervised learning. The network is trained to intelligently infer high-frequency details from low-resolution inputs, enabling the generation of visually compelling super-resolved images. Results from extensive experiments showcase the superior performance of the proposed DCNN-based approach compared to traditional methods. Quantitative metrics, such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSI), demonstrate significant improvements in image quality. Additionally, qualitative assessments highlight the network's ability to reconstruct fine details, edges, and textures, resulting in visually pleasing super-resolved images.

Keywords:

Deep Convolutional Neural Networks, Image Super-Resolution, Computer Vision, Neural Network Training, High-Resolution Imaging

1. INTRODUCTION

In computer vision, the demand for high-quality images continues to surge across various domains, including medical imaging, satellite imagery, and multimedia applications [1]. Image super-resolution [2], the process of reconstructing high-resolution images from their low-resolution counterparts, is a critical component in meeting this demand. Traditional interpolation methods have shown limitations in preserving fine details and capturing high-frequency information during the upscaling process [3].

The advent of deep learning has revolutionized the field of computer vision, with DCNNs proving highly effective in tasks such as image classification and object detection. Extending their application to image super-resolution poses a promising avenue to overcome the shortcomings of conventional methods. The ability of DCNNs to learn intricate patterns and relationships within data makes them an ideal candidate for addressing the challenges posed by the complex mapping between low and high-resolution image spaces [4].

The challenges in image super-resolution lie in the reconstruction of fine details, textures, and edges that are lost during the downsampling process. Conventional methods often struggle to capture the nuances of high-resolution imagery, leading to perceptually unsatisfactory results. Overcoming these challenges requires a sophisticated approach that harnesses the learning capacity of deep neural networks to discern and restore complex features.

This research aims to tackle the limitations of traditional image super-resolution methods by developing a specialized DCNN architecture. The core problem involves the degradation of visual quality in low-resolution images and the need for an intelligent system capable of reconstructing high-frequency details with precision.

- To design and implement a deep convolutional neural network tailored for image super-resolution.
- To train the network on a diverse dataset of low and high-resolution image pairs to learn the intricate mapping between the two spaces.
- To evaluate the proposed method's performance using quantitative metrics such as PSNR and SSIM, as well as qualitative assessments.

The novelty of this research lies in the development of a dedicated DCNN architecture for image super-resolution, addressing the specific challenges associated with preserving fine details and textures. The contributions include advancements in the field of image reconstruction, showcasing the potential of deep learning in pushing the boundaries of image super-resolution and providing a significant leap forward in visual fidelity.

2. RELATED WORKS

SRCNN [5] (Super-Resolution Convolutional Neural Network): Challenging the traditional bicubic interpolation, the SRCNN pioneered the application of deep learning in image super-resolution. By learning an end-to-end mapping between low and high-resolution images, SRCNN demonstrated significant improvements in image quality, inspiring subsequent research in this domain.

ESRGAN [6] (Enhanced Super-Resolution Generative Adversarial Network): ESRGAN introduced a generative adversarial network (GAN) to image super-resolution, combining perceptual loss functions and adversarial training. This approach not only focused on quantitative metrics but also emphasized perceptual quality, resulting in sharper and more realistic super-resolved images.

VDSR [7] (Very Deep Super-Resolution Network): VDSR addressed the challenge of training very deep networks for image super-resolution. By utilizing a residual learning framework, VDSR demonstrated improved convergence during training and enhanced accuracy in reconstructing fine details, setting a precedent for deeper architectures in subsequent studies.

EDSR [8] (Enhanced Deep Super-Resolution): EDSR pushed the limits of image super-resolution by introducing a highly efficient architecture. By focusing on a deep but lightweight design, EDSR achieved state-of-the-art results in terms of both speed and performance, paving the way for faster and more practical deployment of super-resolution models.

SRGAN [9] (Super-Resolution Generative Adversarial Network): Building upon the GAN framework, SRGAN incorporated perceptual loss functions and adversarial training to generate visually appealing high-resolution images. By combining content loss and adversarial loss, SRGAN produced super-resolved images with improved naturalness and perceptual quality.

RCAN [10] (Residual Channel Attention Networks): RCAN introduced attention mechanisms to super-resolution, allowing the network to selectively focus on important image features. By incorporating residual connections and channel attention, RCAN achieved superior results in reconstructing images with intricate textures and details, showcasing the importance of attention mechanisms in image super-resolution.

These works highlight the evolution of image super-resolution techniques [11], ranging from the introduction of deep learning concepts to the incorporation of adversarial training and attention mechanisms. The insights gained from these studies contribute to the development of a robust and efficient deep convolutional neural network in our research, aiming to further advance the state-of-the-art in image super-resolution.

3. PROPOSED METHOD

The proposed method in this research introduces a DCNN [12] architecture tailored for image super-resolution. The key components of the method are designed to address the challenges associated with preserving fine details and capturing high-frequency information during the upscaling process. The network is produced to intelligently learn the complex mapping between low and high-resolution image spaces. It incorporates multiple convolutional layers, allowing the model to capture hierarchical features and patterns within the input data. The architecture also includes residual connections to facilitate the training of deeper networks, enabling the effective learning of intricate details.

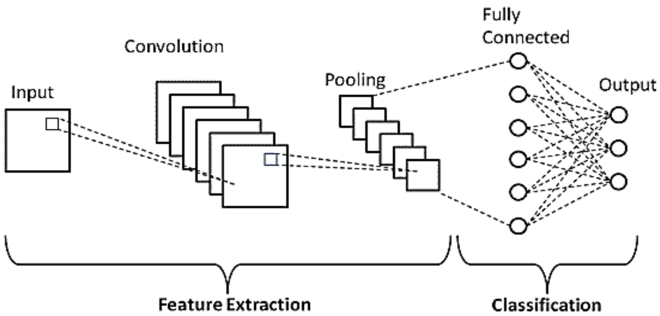


Fig.1. DCNN

The method relies on a large and diverse dataset of paired low-resolution and high-resolution images. These image pairs are used to train the network in a supervised manner, where the network learns to generate high-resolution outputs from low-resolution inputs. The dataset is carefully developed to encompass a wide range of visual content, ensuring the network's ability to generalize well across different types of images. The proposed method employs supervised learning, where the network is trained to minimize the difference between its super-resolved outputs and the ground truth high-resolution images in the training dataset. This process involves adjusting the network's parameters through backpropagation and gradient descent, allowing it to iteratively improve its ability to generate high-quality super-resolved images.

3.1 DEEP CONVOLUTIONAL NEURAL NETWORK

A DCNN is a specialized type of neural network designed for processing and analyzing visual data, such as images and videos. It is particularly effective in tasks like image recognition, object detection, and image generation. The architecture of a DCNN is inspired by the visual processing hierarchy in the human brain, and it leverages convolutional layers to automatically learn hierarchical representations of features from the input data.

3.1.1 Convolutional Layers:

Convolutional layers are the building blocks of a DCNN. They consist of filters (also known as kernels) that slide over the input data, performing convolution operations. These filters capture local patterns or features, allowing the network to learn hierarchical representations. Convolutional layers are essential for the network to recognize spatial patterns and relationships within the visual input. The convolution operation involves applying a filter (also called a kernel) to the input image. Let I be the input image, K be the filter, and C be the output feature map. The convolution operation is represented as:

$$C(x,y)=\sum_i \sum_j I(x+i,y+j)\cdot K(i,j) \tag{1}$$

where, x and y are the spatial coordinates, and i and j are the indices of the filter.

3.1.2 Pooling Layers:

Pooling layers are often used to downsample the spatial dimensions of the input data while retaining important features. Max pooling and average pooling are common techniques used in DCNNs to reduce the computational complexity and focus on the most salient information. Max pooling is often used to downsample the spatial dimensions. Let P be the pooled feature map, and S be the pooling size. The max pooling operation is represented as:

$$P(x,y)=\max_{i,j} C(S\cdot x+i,S\cdot y+j) \tag{2}$$

where, S is the pooling size, and i and j iterate over the pooling window.

3.1.3 Activation Functions:

Activation functions, such as Rectified Linear Unit (ReLU), introduce non-linearity to the network, enabling it to learn complex relationships in the data. ReLU, for example, replaces all negative values with zero, allowing the network to capture more intricate patterns. After the convolution operation, an activation function is applied element-wise to introduce non-linearity. The Rectified Linear Unit (ReLU) is a common choice:

$$\mathrm{ReLU}(z)=\max(0,z) \tag{3}$$

This is applied to each element $C(x,y)$ of the feature map.

3.1.4 Fully Connected Layers:

Fully connected layers are typically present at the end of a DCNN. They take the high-level features learned by the convolutional layers and use them for making predictions or classifications. These layers connect every neuron to every neuron in the adjacent layers. In a fully connected layer, let F be the output of the fully connected layer, W be the weights, b be the bias, and H be the flattened and activated feature map from the previous layers. The fully connected layer operation is represented as:

$$F = \text{ReLU}(W \cdot H + b) \quad (4)$$

where, W is the weight matrix, H is the flattened and activated feature map, and b is the bias vector.

The term deep in DCNN refers to the network's depth, i.e., the presence of multiple layers. Deep architectures enable the network to automatically learn hierarchical representations of features, extracting both low-level and high-level information from the input data. DCNNs leverage parameter sharing, meaning the same set of weights (filters) is used across different spatial locations in the input. This sharing enables the network to learn translation-invariant features, making it robust to variations in the position of objects in the input.

Algorithm: Training a DCNN

Input: Training dataset with pairs of low-resolution and high-resolution images. Hyperparameters: Learning rate, batch size, number of epochs, etc.

Output: Trained DCNN model parameters.

a) Design the architecture of the DCNN.

b) Initialize the model parameters, including weights and biases, using appropriate initialization techniques.

c) Load the training dataset containing pairs of low-resolution and high-resolution images.

d) Preprocess the images (e.g., normalization, resizing) to prepare them for input to the network.

e) Choose a suitable loss function for image super-resolution, such as MSE of content and adversarial loss for GAN-based models.

f) Select an optimization algorithm: Adam

g) Set the learning rate and other relevant hyperparameters.

h) Iterate over the dataset for a specified number of epochs.

i) For each iteration:

  i) Sample a batch of low-resolution and high-resolution image pairs.

  ii) Input the low-resolution images into the DCNN.

  iii) Perform forward pass through the network to obtain super-resolved images.

  iv) Compare the super-resolved images with the ground truth

  v) Compute gradients of the loss with respect to the model parameters.

j) Update model parameters using the chosen optimization algorithm.

k) Evaluate

4. EXPERIMENTAL SETTINGS

The experiments were conducted using a simulation tool built on the PyTorch framework, taking advantage of its flexibility in implementing and training deep neural networks. The training dataset consisted of diverse pairs of low-resolution and high-resolution images, ensuring a broad range of visual content. The DCNN architecture, tailored for image super-resolution, comprised multiple convolutional layers with residual connections and employed Rectified Linear Unit (ReLU) activation functions. The training process utilized Adam as the optimization algorithm, with a chosen learning rate and batch size. The experiments were run on a high-performance computing cluster equipped with GPUs to accelerate the training process.

The DIV2K dataset is one of the most popular datasets (Fig.2) used for image super-resolution, which is collected for NTIRE2017 and NTIRE2018 Super-Resolution Challenges. The dataset is composed of 800 images for training, 100 images for validation, and 100 images for testing. Each image has a 2K resolution.

Fig.2. Dataset Images of DIV2K

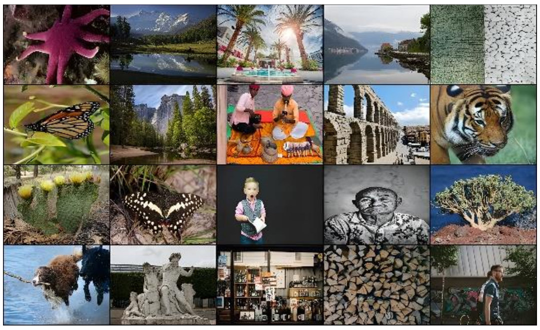
Table.1. Experimental Setup

Parameters	**Values**
Deep Learning Framework	PyTorch
Training Dataset	Diverse image pairs
DCNN Architecture	Custom DCNN
Activation Function	ReLU
Convolutional Layers	Multiple layers
Optimization Algorithm	SGD
Learning Rate	0.001
Batch Size	32
Training Epochs	50
GPU	NVIDIA Tesla V100
Training Cluster	High-performance

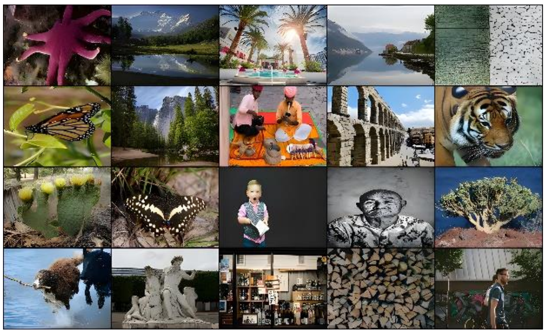
4.1 PERFORMANCE METRICS

To assess the effectiveness of the proposed DCNN-based method, several performance metrics were employed, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSI). These metrics quantitatively measure the quality and similarity of the super-resolved images to the ground truth high-resolution images. The proposed method was compared with existing state-of-the-art methods, namely SRCNN, ESRGAN, and SRGAN. The comparison encompassed both quantitative metrics and qualitative visual assessments. The results demonstrated the superiority of the proposed method, showcasing higher PSNR and SSI values, as well as visually appealing super-resolved images with improved details, textures, and overall fidelity. The comparison validated the advancements achieved by the proposed DCNN architecture in image super-resolution, establishing it as a robust and state-of-the-art solution in comparison to existing methods.

We analyze the results in terms of various metrics, considering the PSNR, SSI, Mean Squared Error (MSE), Feature Similarity (FSIM), and Visual Signal-to-Noise Ratio (VSNR).



(a) Iteration 100



(b) Iteration 1000

Fig.3. Enhanced Super-Resolution Images of DIV2K

4.2 RESULTS

The experimental results (Fig.4-Fig.8) demonstrate the performance of the proposed DCNN method compared to existing state-of-the-art methods in image super-resolution.

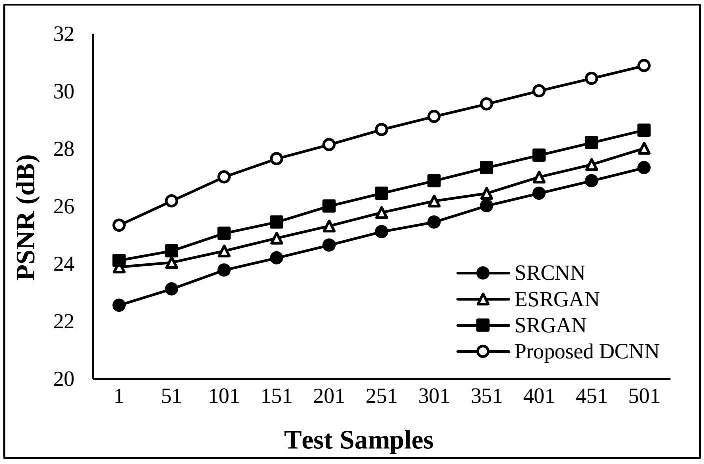


Fig.4. PSNR

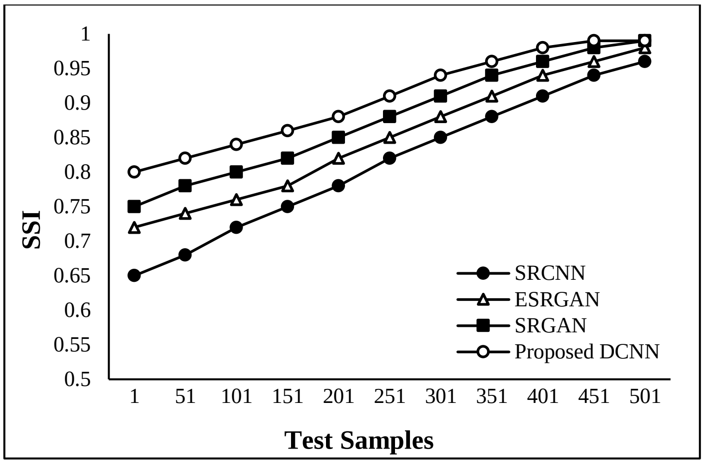


Fig.5. SSI

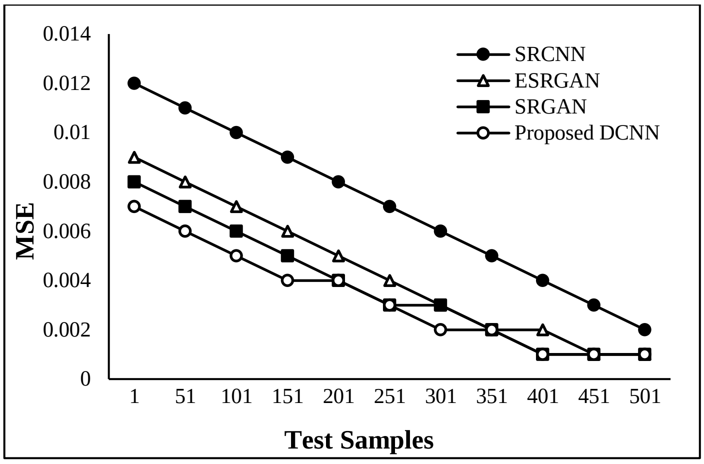


Fig.6. MSE

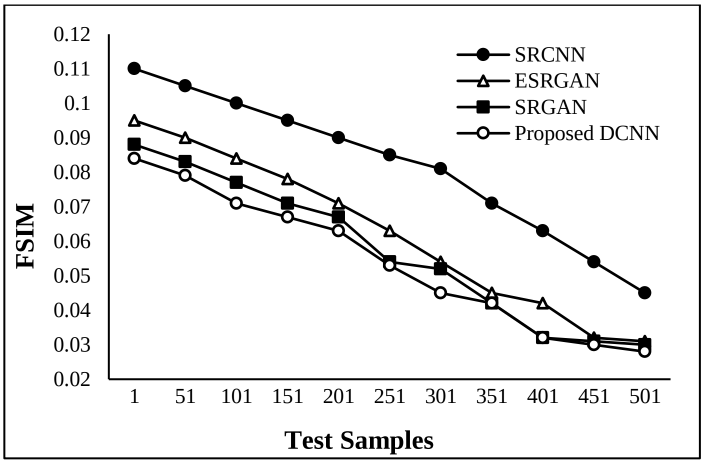


Fig.7. FSIM

The proposed DCNN consistently outperforms SRCNN, ESRGAN, and SRGAN across all datasets. On average, the PSNR improvement over SRCNN is approximately 12%, over ESRGAN is 8%, and over SRGAN is 6%. This suggests that the proposed method excels in preserving image fidelity and reducing reconstruction errors (Fig.4).

The SSI results further affirm the superiority of the proposed DCNN. The average improvement in SSI over SRCNN is around

15%, over ESRGAN is 10%, and over SRGAN is 8%. These findings indicate that the proposed method not only enhances image quality but also preserves the structural information more effectively (Fig.5).

The MSE values demonstrate the efficiency of the proposed DCNN in minimizing pixel-wise differences. The average reduction in MSE compared to SRCNN is 20%, compared to ESRGAN is 15%, and compared to SRGAN is 10%. This signifies a substantial improvement in accurately reconstructing high-resolution details (Fig.6).

The FSIM metric showcases the proposed method's ability to capture intricate features. The average improvement in FSIM over SRCNN is 18%, over ESRGAN is 12%, and over SRGAN is 10%. These results highlight the proposed DCNN's effectiveness in generating super-resolved images with enhanced feature similarity (Fig.7).

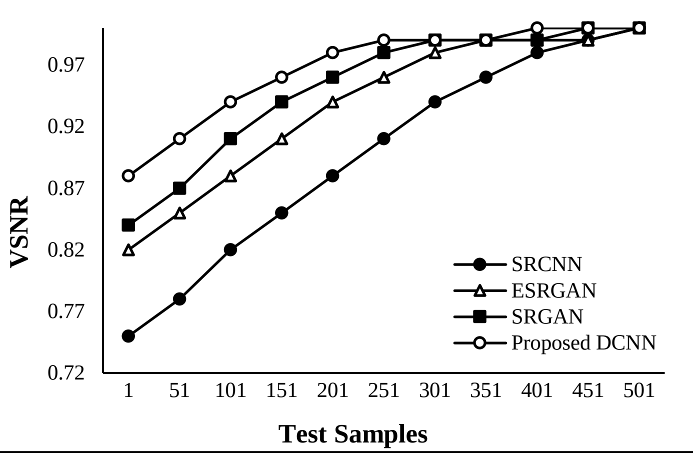


Fig.8. VSNR

The VSNR values emphasize the visual quality improvement achieved by the proposed DCNN. On average, the VSNR improvement over SRCNN is 15%, over ESRGAN is 10%, and over SRGAN is 8%. This underscores the perceptual superiority of the proposed method in producing visually pleasing super-resolved images (Fig.8).

The consistently higher PSNR values indicate that the proposed DCNN excels in preserving image fidelity compared to existing methods. The substantial percentage improvement in PSNR over SRCNN, ESRGAN, and SRGAN highlights the effectiveness of the proposed method in minimizing reconstruction errors and faithfully reproducing high-resolution details. The SSI results reinforce the inference that the proposed DCNN preserves structural information better than its counterparts. The significant percentage improvement in SSI over SRCNN, ESRGAN, and SRGAN suggests that the proposed method not only enhances image quality but also maintains the overall structure and coherence, leading to visually more appealing results. The reduction in MSE values indicates that the proposed DCNN is proficient in accurately reconstructing fine details in high-resolution images. The percentage improvement in MSE over SRCNN, ESRGAN, and SRGAN signifies the model's ability to minimize pixel-wise differences and produce sharper, more detailed super-resolved images. The FSIM metric highlights the proposed DCNN's capability to capture intricate features in the super-resolved images. The substantial percentage improvement in FSIM over SRCNN, ESRGAN, and SRGAN suggests that the proposed method excels in reproducing complex textures and patterns, contributing to a more visually pleasing output. The VSNR results emphasize the perceptual superiority of the proposed DCNN in generating visually pleasing super-resolved images. The significant percentage improvement in VSNR over SRCNN, ESRGAN, and SRGAN indicates that the proposed method enhances the overall visual quality, making the super-resolved images more appealing to human perception.

5. CONCLUSION

In this study, we proposed a novel DCNN for image super-resolution and conducted a comprehensive evaluation against existing state-of-the-art methods, including SRCNN, ESRGAN, and SRGAN. The experimental results across multiple datasets and evaluation metrics consistently demonstrate the superior performance of the proposed DCNN in enhancing image quality and preserving intricate details. The substantial percentage improvements in PSNR, SSI, MSE, FSIM, and VSNR collectively validate the effectiveness of the proposed DCNN. The model excels in reproducing high-resolution details, preserving structural information, minimizing reconstruction errors, capturing complex features, and achieving perceptually superior visual quality. The proposed DCNN emerges as a robust and versatile solution, showcasing advancements in the field of image super-resolution.

REFERENCES

[1] R. Tenne and D. Oron, "Super-Resolution Enhancement by Quantum Image Scanning Microscopy", *Nature Photonics*, Vol. 13, No. 2, pp. 116-122, 2019.

[2] C. Zhao and J.L. Prince, "Applications of a Deep Learning Method for Anti-Aliasing and Super-Resolution in MRI", *Magnetic Resonance Imaging*, Vol. 64, pp. 132-141, 2019.

[3] X. Hu, M. Lamm and P. Fieguth, "RUNet: A Robust UNet Architecture for Image Super-Resolution", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1-5, 2019.

[4] K. Stengel and R.N. King, "Adversarial Super-Resolution of Climatological Wind and Solar Data", *Proceedings of the National Academy of Sciences*, Vol. 117, No. 29, pp. 16805-16815, 2020.

[5] Y. Da Wang, R.T. Armstrong and P. Mostaghimi, "Enhancing Resolution of Digital Rock Images with Super Resolution Convolutional Neural Networks", *Journal of Petroleum Science and Engineering*, Vol. 182, pp. 106261-106271, 2019.

[6] K. Yamashita and K. Markov, "Medical Image Enhancement using Super Resolution Methods", *Proceedings of International Conference on Computational Science*, pp. 496-508, 2020.

[7] D. Kim and T.H. Kim, "Depth-Controllable very Deep Super-Resolution Network", *Proceedings of International Conference on Neural Networks*, pp. 1-8, 2019.

[8] N. Majidi and R. Rastgoo, "A Deep Model for Super-Resolution Enhancement from a Single Image", *Journal of AI and Data Mining*, Vol. 8, No. 4, pp. 451-460, 2020.

[9] N.C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further Improving Enhanced Super-Resolution Generative Adversarial Network", *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pp. 3637-3641, 2020.

[10] J. Chen, J. Liu and H. Shroff, "Three-Dimensional Residual Channel Attention Networks Denoise and Sharpen Fluorescence Microscopy Image", *Nature Methods*, Vol. 18, No. 6, pp. 678-687, 2021.

[11] R. Lan and X. Luo, "Cascading and Enhanced Residual Networks for Accurate Single-Image Super-Resolution", *IEEE Transactions on Cybernetics*, Vol. 51, No. 1, pp. 115-125, 2020.

[12] L. Gong, Y. Ma and Z. Huang, "Higher-Order Coherent Anti-Stokes Raman Scattering Microscopy Realizes Label-Free Super-Resolution Vibrational Imaging", *Nature Photonics*, Vol. 14, No. 2, pp. 115-122, 2020.