Mansoura University
Faculty of Engineering
Mansoura Engineering Journal

# A Proposed Video Super-Resolution Strategy using Wavelet Multi-Scale Convolutional Neural Networks

M. A. Elgohary*, Fathi E. Abd El-Samie, Walid El-Shafai, M. A. Mohamed and E. H. Abdelhay

**KEYWORDS:**
***Super-Resolution, Convolutional Neural Networks, Wavelet Analysis, Multi-Frame, Multi-Scale Regression, Frequency Domain, Spatial Domain***

***Abstract*— High-resolution images are often required and desired for most applications, as they incorporate complementary information. However, the optimal utilization of sensor technology and visual technology to improve picture pixel density is often limited and prohibitively expensive. As a result, employing an image processing method to build a high-resolution image from a low-resolution one is a costly and comprehensive option. The goal of video super-resolution is to restore intricate points and reduce the sensory effects. This research builds on the multi-frame super-resolution approach by using wavelet analysis to train convolutional neural networks (CNNs). For that purpose, the approach begins by applying wavelet decomposition on video segments for multi-scale assessment. Then, several CNNs are trained independently to approximate wavelet multi-scale characterizations. The trained CNNs do inference by regressing wavelet multi-scale characterizations from LR frames, followed by wavelet reconstruction, which produces recovered HR frames. This research presents a learning-based method for preserving fine features in low-resolution multi-frame images captured with various camera zoom lenses. The experimental findings confirm the proposed strategy for restoring difficult frames.**

## I. INTRODUCTION

MANY purposes necessitate the use of high-resolution (HR) images. Super-resolution (SR) reconstruction is often regarded as an efficient method of increasing image spatial resolution. The resolution of an image is highly important in image processing. The quality of an image determines how much information can be extracted from it. It will be exceedingly difficult and deceptive to understand the image if it is degraded. As a result, a zoom is used to enlarge the image areas. Unfortunately, the interpolation technique produces a fuzzy and low-resolution (LR) image, when zoomed past its resolution. Creating the required information in the original image is, in reality, difficult. The use of super-resolution methods is another option to guess this information [17, 40, 41]. The purpose of the multi-frame super-resolution technique is to generate an HR image from a succession of LR images that have been degraded by noise, blur, and decimation [34]. Medical diagnostics [22],

*Received: (20 March, 2022) - Revised: (03 June, 2022) - Accepted: (21 July, 2022)*

***Corresponding Author:** M. A. Elgohary, MSc Student, Department of Mechatronics Engineering, High Institute of Engineering and Technology, El-mahalla El-koubra, Egypt. (Mohamed.elgohary2012@gmail.com)*

*Fathi E. Abd El-Samie, Professor, Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University Menouf, Egypt. (*fathi_sayed@yahoo.com*)*

*Walid El-Shafai, Assistant Professor, Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University Menouf, Egypt. (eng.waled.elshafai@gmail.com)*

*M. A. Mohamed, Professor, Department of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University, Mansoura, Egypt. (mazim12@mans.edu.eg)*

*E. H. Abdelhay, Assistant Professor, Department of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University, Mansoura, Egypt. (ehababdelhay@mans.edu.eg)*

(Ser. NO. BFEMU-2203-1236)

satellite imaging [8], face identification [26], network analysis [2], and video monitoring [26] are examples of SR methods employed. Furthermore, the public and industry demand for converting low-resolution and outdated digital movies to high-definition (HD) movies is growing every day.

The initial stage for multi-frame SR is to align all LR images by using motion information [31]. This stage is important for multi-frame SR algorithms to work since super-resolution is severely hampered without a strong estimate of motion between the LR frames [7]. Many efforts have been made to address the problems of the registration phase. Iterative back projection (IBP) [9,18], projection on a convex set (POCS) [25], and optical-flow projection [6] are some of the approaches adopted to correct registration errors. However, because the solutions are non-unique and can only handle translational and rotational motion between LR frames, these strategies are unsuccessful.

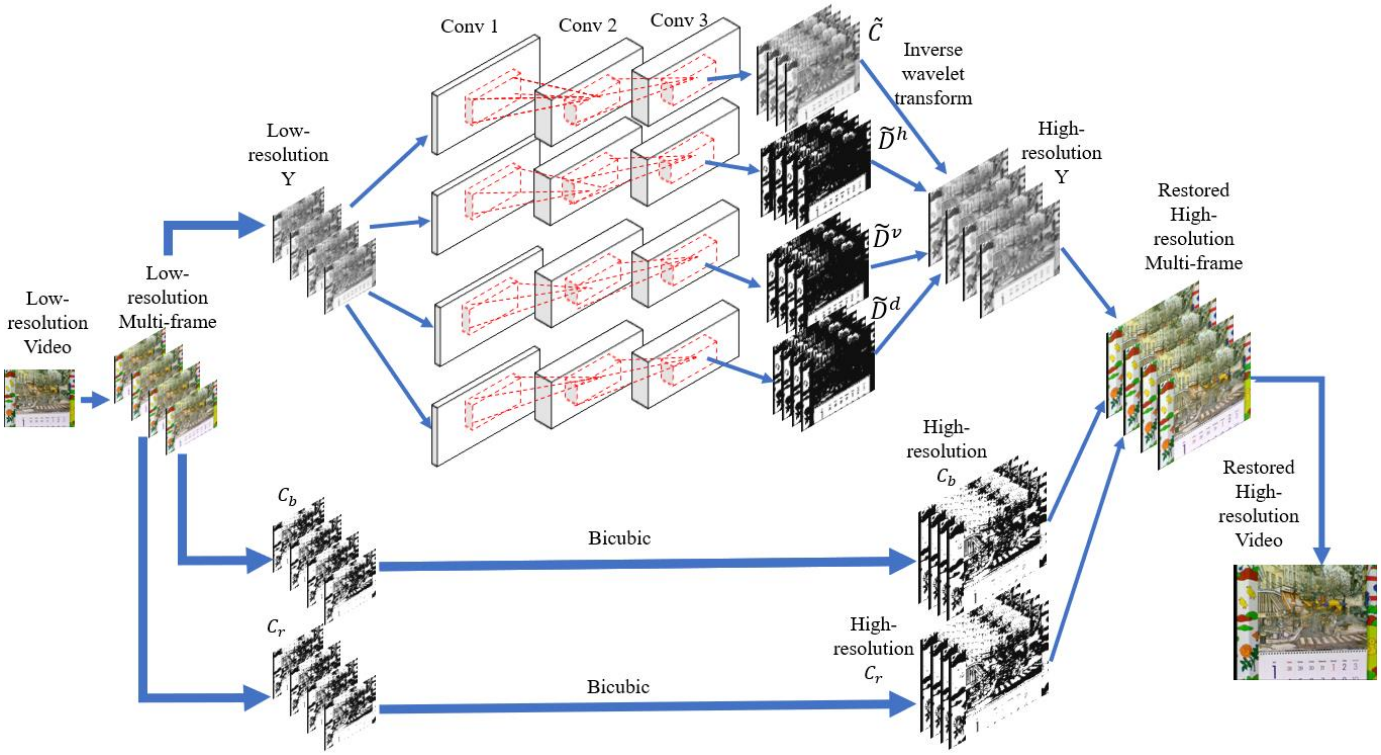



Fig. 1: The proposed multi-frame wavelet multi-scale CNN.

To mitigate the misregistration issues, more robust methods have been proposed [24, 38]. The authors of [37] discovered a non-local effective method that may be used for non-global motion. Otherwise, to cope with non-parametric deformations between the LR pictures, the authors of [29] presented the concept of elasticity registration, but it was limited to minor deformations and did not address larger ones. Other methods rebuild the HR image using maximum a posteriori (MAP) algorithm [39] and a spatial domain monitoring technique. Other techniques have subsequently been developed [5, 27], although they still suffer from misregistration problems. After that, choosing a function of regularization for the procedure for dehazing and noise removal in the final stage of the SR method is critical for staying away from various artifacts. The selection of this function must be done with considerable care. A variety of SR methods [4,28,43] rely on a regularization framework. Although these methods produce promising outcomes, they have certain flaws. Two of these methods are the staircase illusion and the haze apparition on flat surfaces. Although the regularization value may be adjusted to decrease noise and blur in smooth sections, the resulting characterization is hazy. To address the issue of single image SR, deep learning has also been frequently utilized. One of the most well-known works is that of Dong et al. [16], in which the SRCNN technique was suggested consisting of triple convolutional phases. Patch extraction, non-linear mapping, and reconstruction are the three processes that are accomplished. The SRCNN has shown promising results, but the necessity for massive HR images for data training remains an embarrassment. Then, Dong et al. proposed the FSRCNN [3], a faster variant of SRCNN to include the up-sampling operation within the network. As a consequence, the restoration quality is excellent, while the execution time is short. Other techniques are then developed to address the shortcomings of the preceding strategy; for further information, see [23]. Zhang et al. [42] proposed a similar network (DnCNN) using more recent developments in deep learning, giving good performance. Chen et al. [12] recently looked at the CISRDCNN framework for compressed images, revealing good results.

In preliminary research, Lei et al. [19] used a local-global-combined network (LGCNet) to capture restoration characteristics at both local and global levels in the spatial

domain in preliminary research to bridge deep learning with multi-scale analysis. Wei et al. [20] created an extremely deep residual neural network to improve the precision of multi-spectral picture panning. Li et al. [21] employed spectral mixture evaluation and sparseness to allow multispectral image SR. These approaches used strategies based on advanced machine learning in the fundamental spatial or spectral domains, but they ignored image properties in transform domains. For SR, interpretations of images at various frequencies reflect distinct characteristics, while components with high frequency have a major influence [30]. This encourages other researchers to create a plethora of SR algorithms at various frequency bands. Filter banks are used in wavelet analysis to describe images in multi-scale frequency bands. In this case, the nomenclature of the two new wavelet-based techniques in [32] and [33] is comparable. Depending on the above discussion, the contribution of this paper can be explained as follows:

- The proposed approach differs from that in [32] in that it uses several concurrent CNNs.
- The proposed approach differs from that in [33] in that it uses wavelet synthesis for full reconstruction.
- As a result, there are considerable changes in the structure of the model, feature extraction methods, and learning spaces between the proposed approach and those in [32] and [33].

Multiple CNNs are trained in multi-scale frequency ranges provided by the wavelet analysis to restore frequency features in different directions, unlike deep learning-based SR techniques (such as SRCNN and LGCNet), which analyze images in their initial spatial domain. As opposed to edge-protecting wavelet-based SR [10], which only handles local features, SR features are represented at several frequency ranges using several CNNs. The proposed approach, in particular, captures both high-frequency local variations and low-frequency general patterns.

## II. WAVELET MULTISCALE CNNS FOR MULTI-FRAME SR

This part introduces how to train wavelet multi-scale CNNs to characterize different scales and in order to reconstruct HR images.

### *A. Multiscale Regression Features Using Wavelet Representations*

A discrete wavelet transform is used to perform a multi-scale evaluation on several frames. The discrete wavelet transform is accomplished by the use of filter banks composed of biorthogonal high-pass filters and low-pass filters. Let L stand for a low-pass filter matrix with columns representing low-pass filter coefficients, and H stand for a high-pass filter matrix with columns representing high-pass filter coefficients. As for the initial level representation, one original HR image is utilized, and the wavelet decomposition is conducted as follows [11]:

$$C_j = \left(L^{\dagger} C_{j-1} L\right)_{s\downarrow}, \qquad D_j^h = \left(L^{\dagger} C_{j-1} H\right)_{s\downarrow} \qquad (1)$$
$$D_j^v = \left(H^{\dagger} C_{j-1} L\right)_{s\downarrow}, \qquad D_j^d = \left(H^{\dagger} C_{j-1} H\right)_{s\downarrow}$$

where $j$ is indeed the division level, and s↓ is just the down-sampling by 1/$s$ of the initial resolution. The recursive down-scaling division expresses every image from multi-frame images in terms of various spatial ratios, allowing for more complete distant observations using multi-frame images. $C_j, D_j^h, D_j^v,$ and $D_j^d$ indicate the comprehensive low frequency, horizontal high frequency with vertical low frequency, vertical high frequency with horizontal low frequency, and comprehensive high-frequency properties of the final level characterization $C_{j-1}$, respectively. In various spatial proportions, rotations, and frequency bands, the wavelet decomposition produces multi-scale representations for every image. In the next part, wavelet multi-scale representations are employed as regression features for training several CNNs for super-resolution.

### *B. CNNs-based Regression Wavelet Multi-Scale Properties from LR Images*

For regressing wavelet multi-scale properties from LR images, several CNNs are trained. Fig. 1 shows four CNNs that are trained in the example framework. Each CNN examines the LR picture layer at a time, with the outer layer seeking the regression of one of the HR image multi-scale interpretations described in Section II-A. Each CNN is built using the framework of a state-of-the-art super-resolution convolutional neural network (SRCNN) [3]. Like the input, an LR image is obtained from a HR image. The output of the n-th convolutional layer is:

$$f_n(I_L, W_n, b_n) = \sigma(W_n * f_{n-1}(I_L) + b_n), \qquad (2)$$

where the net weights and training bias are characterized by $W_n$, and $b_n$, respectively. Σ is a linear function that has been adjusted $(e.g. max(0, x))$ to allows the CNNs to quickly converge.

Each CNN is penalized by a loss function that computes the difference between the interpretation produced by the CNN from the LR image and the interpretation generated by the wavelet division of the corresponding HR image for the objective of creating components that best regress the wavelet multi-scale interpretation. The top CNN loss function in Fig. 1 is defined by

$$l = \frac{1}{2K} \sum_{K=1}^{K} \left\| C(K) - \tilde{C}(K) \right\|_2^2, \qquad (3)$$

where *K* is the wavelet multi-scale representation pixel index, *C* is the wavelet multi-scale representation that maintains a high-resolution image two-direction smoothing properties, and $\tilde{C}$ is the characterization created by the CNN from the LR image

$I_L$. The loss functions for the remaining three CNNs may be obtained by substituting $D^h$, $D^v$, and $D^d$ for $C$ in (3), respectively.

Using a back-propagation algorithm, the multiple CNNs are separately trained to learn properties given by the appropriate wavelet evaluation and to minimize their loss functions, sequentially. The multi-scale image characteristics are captured in different orientations and numerous frequency bands by the multiple CNNs that have been trained in this way.

### *C. Wavelet Multi-Scale Convolutional Neural Networks for SR*

To begin SR, an LR image is fed into each of the several CNNs, individually. Subsequently, from the jth to the (j - 1)th stage, wavelet restoration is performed on the CNN-created characterizations as follows:

$$\tilde{C}_{j-1} = L(\tilde{C}_j)_{s\uparrow}L^{\dagger} + H(\tilde{D}_j^h)_{s\uparrow}L^{\dagger} + L(\tilde{D}_j^v)_{s\uparrow}H^{\dagger} + H(\tilde{D}_j^d)_{s\uparrow}H^{\dagger} \quad (4)$$

The tidal symbols denote that $\tilde{C}_j$, $\tilde{D}_j^h$, $\tilde{D}_j^v$, and $\tilde{D}_j^d$ are CNN-generated representations, as opposed to those acquired using wavelet decomposition in (1). Because a single CNN is equipped with visual features characterized in one frequency range with such orientations indicated by the wavelet transform, wavelet formulation inherently ensembles detailed information from multi-scale bandwidths and orientations and accomplishes appropriate SR. Fig.1 shows a one-level wavelet synthesis example for SR, in which four CNN-generated representations are synthesized to restore an HR image.

### *D. Observations*

Regional computation and multi-scale evaluation, which closely mirror the receptive fields of the nervous system, are two important features that allow convolutional neural networks to be effective. Training network weights, which conduct regional filtering across one entire image, influence regional computation. The multi-scale analysis is carried out via pooling, with down-sampling as the primary modification. State-of-the-art CNN-based SR approaches like SRCNN tend to focus on regional computation, while ignoring multi-scale analysis. Unlike precise pattern identification, SR tasks try to upgrade image feature characterization. As a result, simple down-sampling processes contradict the SR up-scaling aim. Training of several CNNs is proposed without pooling based on image wavelet characterizations to take full advantage of CNN representational capacity in terms of both regional computation and multi-scale evaluation. The CNN network weights are learned by training, and the CNNs filter image representations in a supervised manner. Wavelet decomposition and synthesis, on the other hand, employ commercially available wavelet filters and perform unsupervised filtering on multi-scale interpretations. In addition to the filtering done by the CNN network weights, they, therefore, augment the local processing effects. As a result of the increased local processing, image features are captured more comprehensively. Wavelet down-sampling procedures are a good alternative for the pooling operations that are not used in the currently available CNN-based SR algorithms. In addition, the wavelet analysis employs multi-frequency range filters with various orientations, resulting in a more generalized multi-scale analysis. Furthermore, multi-scale representations favor holistic descriptions of images in addition to improving detailed visual characterization. Finally, using wavelet-decomposition-conjugated filtering corrections, the wavelet synthesis upscales the multi-scale representations, resulting in SR. The four CNNs trained at one level can be recursively utilized to regress wavelet multi-scale representations at many levels for multi-level analysis. An image may be upscaled to high resolution using this recursive approach. However, this imposes a limitation: a single image resolution can only be upscaled twice. This is because wavelet analysis requires down-sampling and up-sampling by two in terms of filter banks. Also, the computational complexity of various techniques is compared. The proposed approach outperforms previous deep learning-based SR techniques in terms of both training and testing efficiency. Because the proposed model uses four CNNs, each with the same design as SRCNN, it is evident that the proposed model is four times more complicated than the SRCNN. Even though the proposed model is more complicated than the SRCNN, it is suitable since its sophistication rises directly proportional to the number of CNNs rather than dramatically. The proposed three-layered model, on the other hand, is not as deep as deep the SR models like (VDSR) [36], which has 20 weight layers. Unlike the extremely deep structure, the CNN model SR capabilities are increased by extending it rather than making it deeper, i.e., by training several three-layered CNNs in parallel in decomposed frequency sub-bands. Among the evaluation criteria are peak signal-to-noise ratio, structural similarity, normalized variation information (NVI), normalized mutual information (NMI), mutual information, joint entropy, and conditional entropy.

## III. EXPERIMENTS

As a proposed training set, 25 YUV format video sequences are employed. They have been widely used in several video SR approaches [44], [35], [45]. To increase the size of the training set, model training is done using a volume-based approach, which involves cutting numerous overlapping volumes from training films and using each volume as a training sample. Each volume has a spatial dimension of 32 × 32 and a temporal duration of 10. The spatial and temporal strides, respectively, are 14 and 8. As a consequence, given the original dataset, around 41,000 volumes can be constructed. The proposed model is put to the test on four videos: City, Calendar, Foliage, and Walk, all of which have been employed by cutting-edge approaches [46], [47]. Because the convolutional procedure may scale to films of any spatial extent and temporal length, volumes during testing do not need to be removed. The architecture and parameters of the SRCNN are used in [3] to build individual CNNs. Model performance is specifically

analyzed in terms of peak signal-to-noise ratio (PSNR) [1] and structural similarity (SSIM) [15].

### A. The Convergence of the Models

The SRCNN is improved by training several CNNs to describe wavelet multi-scale characterizations in the proposed convolutional neural networks (MFWM). For training purposes, their convergence speeds are compared. The analysis was carried out on the same network setup and computing environment. Fig. 2 depicts the convergence curves concerning PSNR. The SRCNN gets convergence in $8 \times 10^4$ iterations, whereas the proposed MFWM achieves convergence in $4 \times 10^4$ iterations with greater accuracies.

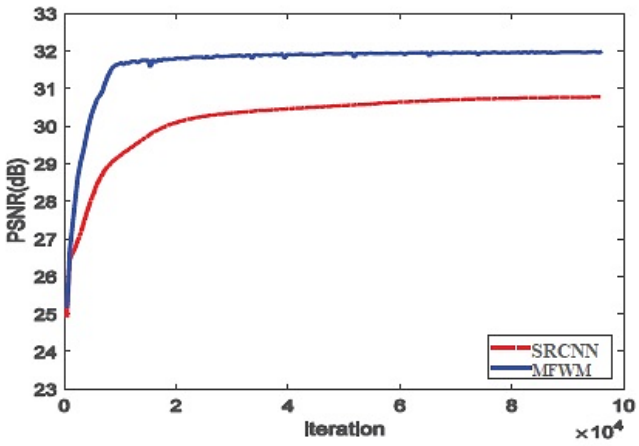



Fig. 2: Convergence curves.

The SRCNN recovers multi-frame images in their entirety. The proposed MFWM, on the other hand, regresses the entire multi-frame image wavelet multi-scale representations. The wavelet-based divide-and-conquer (in terms of different orientations and frequency ranges through CNN training) technique produces a more powerful characterization than solitary holistic characterizations. CNNs learn multi-frame images intrinsically quicker from wavelet multi-scale representations than from full images, as evidenced by the proposed MFWM rapid training convergence.

The parameter setting of this work is provided in TABLE I, and it is highly significant in CNNs.

TABLE I
PARAMETER SETTING

| $f_1$ | $f_2$ | $f_3$ | $n_1$ | $n_2$ | $p_1$ | $p_2$ | $p_3$ | $a_1$ | $a_2$ | $a_3$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 1 | 5 | 64 | 32 | 9 | 5 | 5 | relu | relu | Linear |

where $f_1$ is the first convolutional filter size, $f_2$ is the second convolutional filter size, $f_3$ is the third convolutional filter size, $n_1$ is the number of the first convolutional filters, $n_2$ is the number of the second convolutional filters, $p_1$ is the first convolutional patch size, $p_2$ is the second convolutional patch size, $p_3$ is the third convolutional patch size, $a_1$ is the first convolutional activation type, $a_2$ is the second convolutional activation type, $a_3$ is the third convolutional activation type.

### B. Evaluations (Quantitative and Qualitative)

Every rebuilt HR image must match the original image to assess the validity of the image reconstruction technique. Similarity measurement aids in the monitoring and evaluation of the picture reconstruction procedure performance. In the literature, there are a variety of similarity measurement tools. The PSNR, SSIM, NVI, NMI, mutual information, joint entropy, and conditional entropy are a few of the most well-known metrics.

The MSE, which is the average error between the original image and the SR image, is used to calculate the PSNR. The PSNR is defined as the ratio of signal strength to noise in the image, and it is calculated as follows:

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right),$$

where *MSE* stands for mean squared error, which is defined as:

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \big(Y(i,j) - X(i,j)\big)^2.$$

The *SSIM* is computed on many windows of a given image, i.e. the distance between two windows $x$ and $y$ of size $N \times N$ is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)(2cov_{xy} + C_3)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_1)(\sigma_x\sigma_y + C_3)},$$

where the variables are denoted by $x$ and $y$, respectively. $\mu_x$ and $\mu_y$, are mean values. $\sigma_x^2$ and $\sigma_y^2$, are variance values. $cov_{xy}$ is the covariance. $C_1 = (K_1 L)^2$, and $C_2 = (K_2 L)^2$ are two constants that help to keep everything in balance. $L$ is the dynamic range of pixel values. It is 255 for the 8-bit encoded images. Based on the recognized properties of the human visual system, The *SSIM* provides an estimate of the image quality.

The entropy of a discrete random variable $x$ is:

$$H(x) = -\sum_x p(x) \log p(x) = -E[\log p(x)]$$

The predicted uncertainty in $x$ is measured by the entropy. $H(x)$ is also roughly equal to the amount of information we gain on average from a single instance of the random variable $x$. For two random variables $x$ and $y$, the joint entropy is calculated as follows:

$$H(x,y) = -\sum_{x,y} p(x,y) \log p(x,y).$$

The joint entropy quantifies the amount of uncertainty in the two random variables $x$ and $y$ when they are combined. The conditional entropy of $x$ given $y$ is:

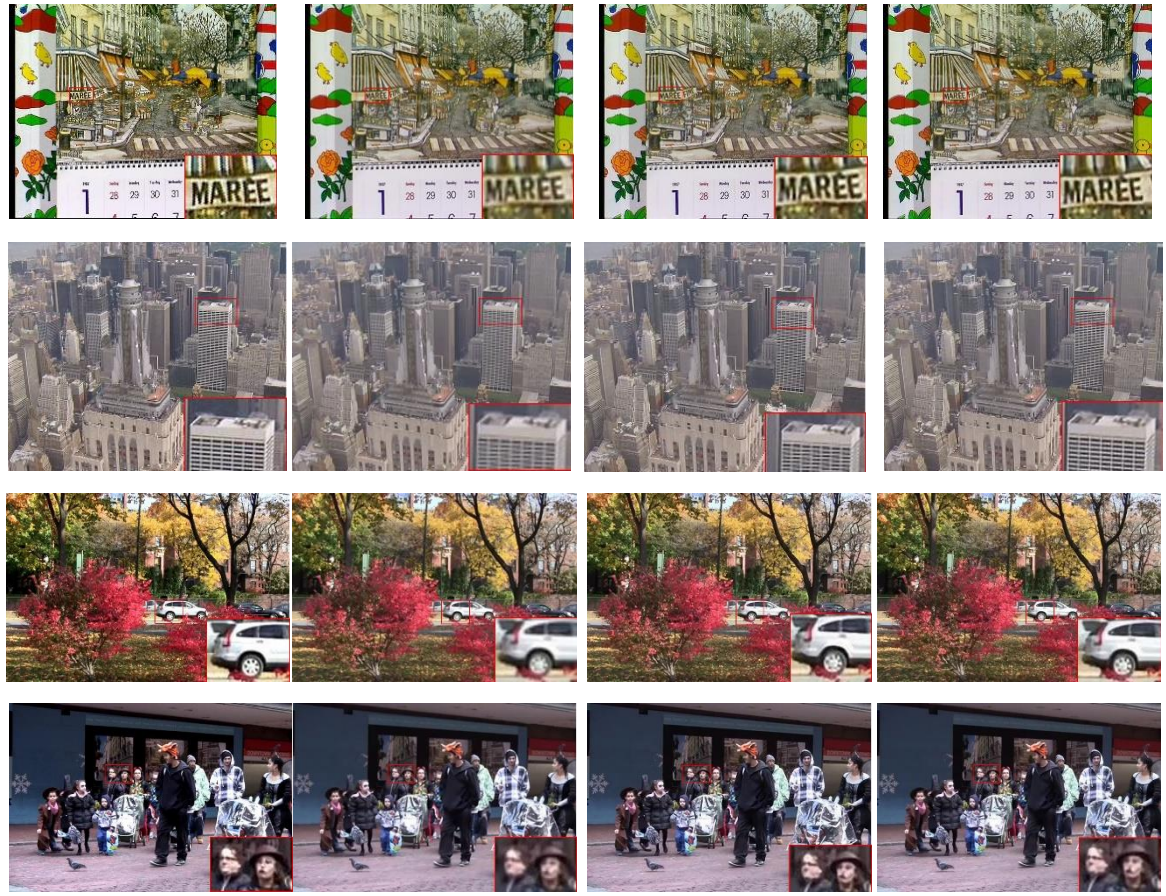
$$H(x|y) = -\sum_{x,y} p(x,y) \log p(x|y) = -E[\log p(x|y)]$$
$$= H(X,Y) - H(Y)$$

The mutual information between two discrete random variables *x*, and *y* that are jointly distributed according to *p* (*x*, *y*) can be calculated as follows:

$$I(x;y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$
$$= H(x) - H(x|y) = H(y) - H(y|x)$$

TABLE II
THE COMPARISON OF PSNR AND SSIM BETWEEN BICUBIC INTERPOLATION, VDSR [36], AND (MFWM).

| Model | *Calendar* | | *City* | | *Foliage* | | *Walk* | | *Average* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Scale2** | **Scale4** | **Scale2** | **Scale4** | **Scale2** | **Scale4** | **Scale2** | **Scale4** | **Scale2** | **Scale4** |
| | PSNR | | | | | | | | | |
| *Bicubic* | 23.45 | 20.71 | 29.32 | 25.07 | 28.27 | 23.35 | 32.07 | 26.02 | 28.27 | 23.78 |
| *VDSR* | 24.49 | 21.31 | 30.77 | 25.48 | 29.27 | 24.03 | 33.46 | 27.17 | 29.49 | 24.49 |
| *MFWM* | 25.11 | 21.62 | 32.84 | 25.68 | 30.38 | 24.17 | 35.31 | 27.67 | 30.91 | 24.78 |
| | SSIM | | | | | | | | | |
| *Bicubic* | 0.9799 | 0.7953 | 0.9734 | 0.7511 | 0.9778 | 0.7514 | 0.9916 | 0.8927 | 0.9806 | 0.7976 |
| *VDSR* | 0.9821 | 0.8458 | 0.9894 | 0.7916 | 0.9907 | 0.8078 | 0.9914 | 0.9229 | 0.9884 | 0.8420 |
| *MFWM* | 0.9917 | 0.8625 | 0.9966 | 0.8084 | 0.9968 | 0.8209 | 0.9989 | 0.9326 | 0.9960 | 0.8561 |



(a) original (b) bicubic (c) VDSR (d) MFWM

Fig. 3. Visual comparison among input LR frames and SR results by bicubic, VDSR, and MFWM, respectively, on the dataset with an upscaling factor of 2.

$$= H(x) + H(y) - H(x,y)$$

The NMI is defined as:

$$NMI(X,Y) = \frac{2 \times I(X;Y)}{[H(X)H(Y)]}$$

The variation of information is defined as:

$$VI(x;y) = H(x) + H(y) - 2I(x,y)$$
$$= H(x,y) - I(x,y)$$

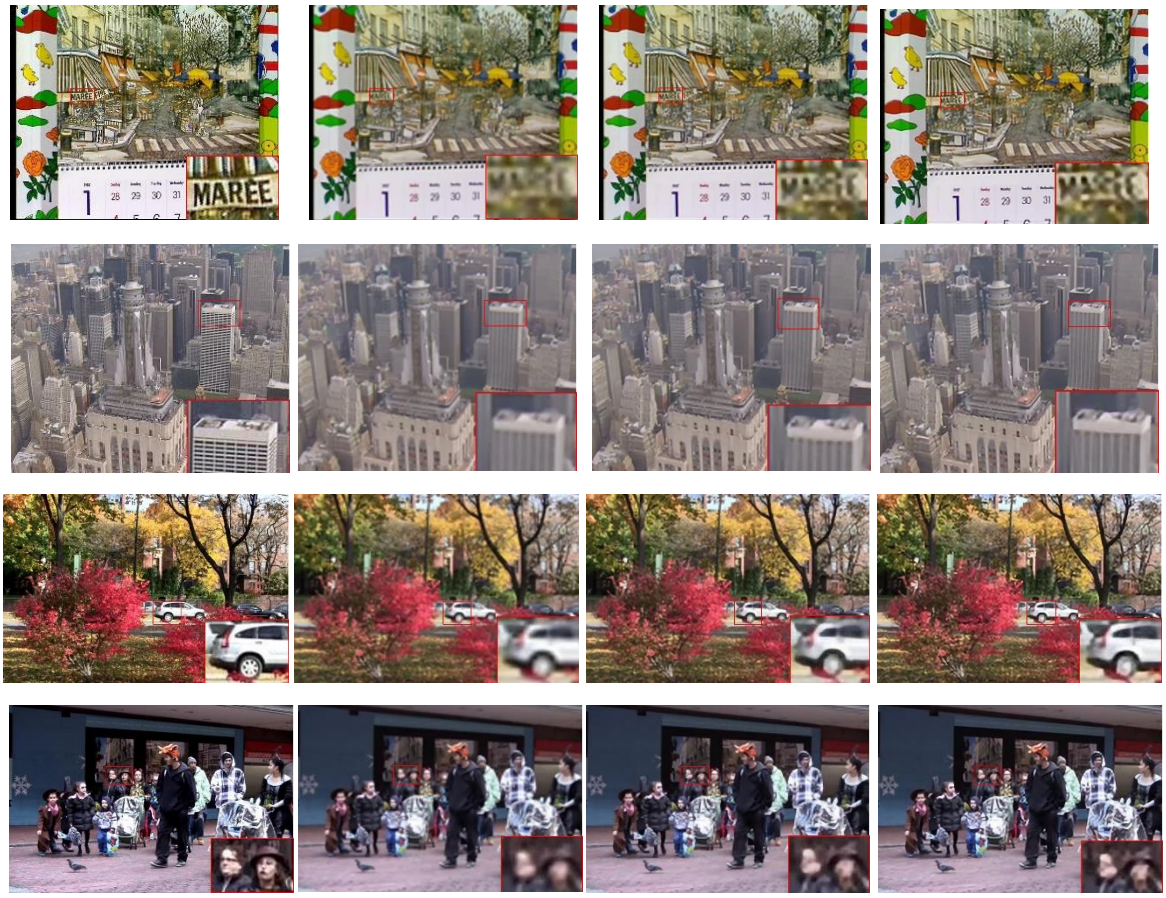
The normalized variation information is defined as:

$$NVI(x;y) = \frac{VI(x;y)}{H(x,y)}$$

The simulation experiments have been implemented on the following hardware: Intel® CoreTM i3-5005U, 2.00GHz CPU, 4GB RAM, Win10 operating system, and Matlab R2018b simulation platform.

The bicubic interpolation is compared to VDSR [36], and the proposed MFWM.

Experiments are carried out using upscaling factors of two and four. The quantitative SR findings in terms of PSNR and SSIM are shown in TABLE II. For upscaling by two and four, the proposed MFWM beats VDSR in terms of PSNR and SSIM. The quantitative SR findings in terms of NVI, NMI, mutual information, joint entropy, and conditional entropy are shown in TABLE III. With increased similarity, the normalized variation information, joint entropy, and conditional entropy decrease. With increasing similarity, both mutual information and normalized mutual information are maximized.



(a) original (b) bicubic (c) VDSR (d) MFWM

Fig. 4. Visual comparison among input LR frames and SR results by Bicubic, VDSR, and MFWM, on the dataset with an upscaling factor of 4.

TABLE III
THE COMPARISON OF NVI, NMI, MUTUAL INFORMATION, JOINT ENTROPY, CONDITIONAL ENTROPY, AND ENTROPY BETWEEN BICUBIC INTERPOLATION, VDSR [36], AND MFWM.

| Model | Calendar | | City | | Foliage | | Walk | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Scale2 | Scale4 | Scale2 | Scale4 | Scale2 | Scale4 | Scale2 | Scale4 | Scale2 | Scale4 |
| | NVI | | | | | | | | | |
| *Bicubic* | 0.8155 | 0.8580 | 0.8234 | 0.8964 | 0.8171 | 0.8886 | 0.6710 | 0.7774 | 0.7817 | 0.8551 |
| *VDSR* | 0.8172 | 0.8546 | 0.8053 | 0.8900 | 0.8007 | 0.8804 | 0.6790 | 0.7603 | 0.7755 | 0.8463 |
| *MFWM* | 0.8037 | 0.8592 | 0.7811 | 0.8873 | 0.7891 | 0.8798 | 0.6251 | 0.7558 | 0.7497 | 0.8445 |
| | NMI | | | | | | | | | |
| *Bicubic* | 0.3115 | 0.2487 | 0.3002 | 0.1877 | 0.3092 | 0.2005 | 0.4951 | 0.3641 | 0.3540 | 0.2502 |
| *VDSR* | 0.3091 | 0.2539 | 0.3259 | 0.1983 | 0.3323 | 0.2137 | 0.4859 | 0.3867 | 0.3633 | 0.2631 |
| *MFWM* | 0.3282 | 0.2468 | 0.3592 | 0.2026 | 0.3484 | 0.2146 | 0.5454 | 0.3925 | 0.3953 | 0.2641 |
| | Mutual information | | | | | | | | | |
| *Bicubic* | 2.2652 | 1.8037 | 2.0108 | 1.2441 | 2.2615 | 1.4561 | 3.6232 | 2.6621 | 2.2901 | 1.7915 |
| *VDSR* | 2.2596 | 1.8431 | 2.2044 | 1.3213 | 2.4401 | 1.5574 | 3.5623 | 2.8294 | 2.6166 | 1.8878 |
| *MFWM* | 2.3933 | 1.7991 | 2.4297 | 1.3549 | 2.5554 | 1.5676 | 3.9921 | 2.8734 | 2.8426 | 1.8987 |

**(continued on the next page)**

**(TABLE III: continued)**

| Model | Calendar | | City | | Foliage | | Walk | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Scale2 | Scale4 | Scale2 | Scale4 | Scale2 | | Scale2 | Scale4 | Scale2 | Scale4 |
| | Joint entropy | | | | | | | | | |
| *Bicubic* | 12.2782 | 12.7004 | 11.3843 | 12.0147 | 12.3667 | 13.0696 | 11.0117 | 11.9607 | 11.7602 | 12.4363 |
| *VDSR* | 12.3602 | 12.6763 | 11.3250 | 12.0089 | 12.2445 | 13.0223 | 11.0992 | 11.8037 | 11.7572 | 12.3778 |
| *MFWM* | 12.1920 | 12.7781 | 11.0991 | 12.0188 | 12.1155 | 13.0428 | 10.6476 | 11.7678 | 11.5135 | 12.4018 |
| | Conditional entropy | | | | | | | | | |
| *Bicubic* | 5.0683 | 5.5225 | 4.7645 | 5.5315 | 5.0791 | 5.8809 | 3.6943 | 4.6519 | 4.6515 | 5.3967 |
| *VDSR* | 5.0738 | 5.4832 | 4.5709 | 5.4544 | 4.9006 | 5.7795 | 3.7552 | 4.4847 | 4.5751 | 5.3004 |
| *MFWM* | 4.9402 | 5.5272 | 4.3456 | 5.4207 | 4.7853 | 5.7694 | 3.3254 | 4.4407 | 4.3491 | 5.2895 |

The wavelet multi-scale analysis, which not only compensates for SRCNN lost pooling procedures but also enhances its regional filtering features with wavelet filters, is responsible for this enhancement of SRCNN. According to this empirical comparison, the wavelet analysis, not the greater model size, is the key to MFWM efficacy. In terms of PSNR and SSIM, it is found that the proposed technique outperforms VDSR. It is worth noting that VDSR has 20 weight layers, but the proposed architecture is made up of four three-layered CNNs for a total of twelve levels. These findings indicate that while the proposed technique has lower structural complexity than VDSR, it nevertheless delivers equivalent results. The contrastive results are because VDSR uses the entire multi-frame images as input for training one comprehensive model, whereas the proposed technique uses CNNs to restore multiple frequency sub-bands, ensuring that each frequency characterization is restored, appropriately.

Since higher up-sampling recovery produces more SR unpredictability, comprehensive quality with up-sampling parameter 4 is worse than that with up-sampling factor 2. Some comparative approaches, like the proposed MFWM, demonstrate comparable performance for some classes in this case. The proposed MFWM, on the other hand, outperforms other approaches. The SRCNN has just been utilized as the basic model for multi-scale learning in the proposed research. Wavelet multi-scale learning based on alternative SR models (e.g., VDSR) is expected to generate promising results since wavelet multi-scale representations establish extensive feature subspaces. For qualitative evaluations, the SR outcomes of several approaches are illustrated. The recovered high-quality multi-frame images with the upscaling factor of two are shown in Fig. 3, and the recovered high-quality multi-frame images with the upscaling factor of four are shown in Fig. 4. The magnified views of the items in the dash line boxes show that the proposed MFWM recovers textures more accurately and clearly than the other compared techniques, as can be seen.

## IV. CONCLUSIONS AND FUTURE WORK

A multi-frame super-resolution framework has been created based on wavelet multi-scale convolutional neural networks. Wavelet filters improve the CNN capacity to handle data locally. Down-sampling in wavelet decomposition compensates for the scarcity of pooling activities in CNN-based super-resolution. The proposed framework combines the CNN representational strength for learning specific features with wavelet analysis multi-scale capabilities for getting numerous orientations and frequency representations. The usefulness of the proposed super-resolution framework has been shown in experiments. Disparities between recovered and original high-resolution images are used in traditional multi-frame image super-resolution evaluation methods. However, in actual cases, retrieving the initial HR images is problematic. In the future, methods for creating a visual evaluation and also no-reference performance indices as evaluation criteria will be investigated. In addition, how visual evaluation and no-reference performance indices may be included in the objective function will be looked into to improve perceptual performance.

## AUTHOR'S CONTRIBUTION

***M. A. Elgohary***

1- Conception or design of the work
2- Data collection and tools
3- Data analysis and interpretation
4- Methodology
5- Resources
6- Software
7- Drafting the article

***Prof. Fathi E. Abd El-Samie***

1- Conception or design of the work
2- Data analysis and interpretation
3- Methodology
4- Project administration
5- Resources
6- Software
7- Supervision
8- Final approval of the version to be published

***Assistant Prof. Walid El-Shafai***

1- Conception or design of the work
2- Data collection and tools
3- Data analysis and interpretation
4- Methodology
5- Resources
6- Software
7- Supervision
8- Drafting the article

***Prof. M. A. Mohamed***

1- Conception or design of the work

2- Data analysis and interpretation
3- Investigation
4- Methodology
5- Resources
6- Software
7- Supervision
8- Final approval of the version to be published

**Assistant Prof. E. H. Abdelhay**
1- Conception or design of the work
2- Data analysis and interpretation
3- Investigation
4- Methodology
5- Resources
6- Software
7- Supervision
8- Drafting the article
9- Critical revision of the article

**FUNDING STATEMENT:**

Authors didn't receive any financial support for the research, authorship

**DECLARATION OF CONFLICTING INTERESTS STATEMENT:**

"There are no potential conflicts of interest concerning the research, authorship or publication of his article".

## V. REFERENCES

[1] J. K. Z. Liu and, R. Laganiere, On the use of phase congruency to evaluate image similarity, ICASSP, France, pp.937-940, 2006.
[2] D.R. Amancio, Probing the topological properties of complex networks modeling short written texts, PLoS ONE vol. 10, no. 2, e0118394, 2015.
[3] Aniket Zope, Vandana Inamdar, " Edge Enhancement for Image Super-Resolution using Deep Learning Approach " 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021.
[4] Shangqi Gao, Xiahai Zhuang "Bayesian Image Super-Resolution with Deep Modeling of Image Statistics", 2022.
[5] R.M. Bay, G.I. Salama, T.A. Mahmoud, Adaptive regularization-based super-resolution reconstruction technique for multi-focus low-resolution images, Signal Process., vol. 103, pp. 155-167, 2014.
[6] S. Baker, T. Kanade, Super-resolution Optical Flow, Carnegie Mellon University, The Robotics Institute, 1999.
[7] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 9, pp. 1167-1183, 2002.
[8] K. Cai, J. Shi, S. Xiong, G. Wei, Edge adaptive image resolution enhancement in video sensor network, Opt. Commun., vol. 284, no.19, pp. 4 4 46-4 451,2011.
[9] D. Capel, A. Zisserman, Automated mosaicing with super-resolution zoom, in Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on, IEEE, pp. 885-891, 1998.
[10] H. Chavez-Roman and V. Ponomaryov, "Super-resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 10, pp. 1777-1781, 2014.
[11] A. Cohen, I. Daubechies, et al., "A stability criterion for biorthogonal wavelet bases and their related subband coding scheme," Duke Mathematical Journal, vol. 68, no. 2, pp. 313-335, 1992.
[12] H. Chen, X. He, C. Ren, L. Qing, Q. Teng, Cisrdcnn: super-resolution of compressed images using deep convolutional neural networks, Neurocomputing, vol. 285, pp. 204-219, 2018.
[13] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004.
[14] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in Proceedings of Asian Conference on Computer Vision, 2014.
[15] Z. Wang, C. Bovik, Image quality assessment from error visibility to structural similarity, IEEE Trans. Image Process., vol.13, pp.600-612, 2004.
[16] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process., vol. 20, no.7, pp. 1838-1857,2011.
[17] I. El Mourabit, M. El Rhabi, A. Hakim, A. Laghrib, E. Moreau, A new denoising model for multi-frame super-resolution image reconstruction, Signal Process., vol. 132, pp. 51-65, 2017.
[18] M. Elad, Y. Hel-Or, A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur, IEEE Trans. Image Process., vol. 10, no. 8, pp. 1187-1193, 2001.
[19] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 8, pp. 1243-1247, 2017.
[20] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1795-1799, 2017.
[21] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial-spectral group sparsity," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 9, pp. 1250-1254, 2016.
[22] H. Greenspan, Super-resolution in medical imaging, Comput. J., vol. 52, no. 1, pp. 43-63, 2009.
[23] K. Hayat, Multimedia super-resolution via deep learning: a survey, Digit Signal Process., 2018.
[24] Y. He, K.-H. Yap, L. Chen, L.-P. Chau, A nonlinear least square technique for simultaneous image registration and super-resolution, IEEE Trans. Image Process., vol.16, no.11, pp. 2830-2841, 2007.
[25] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP, vol. 53, no.3, pp. 231-239, 1991.
[26] J. Jiang, C. Chen, K. Huang, Z. Cai, R. Hu, Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation, Inf. Sci., vol. 367, pp. 354-372, 2016.
[27] A. Laghrib, A. Ghazali, A. Hakim, S. Raghay, a multi-frame super-resolution using diffusion registration and nonlocal variational image restoration, Comput. Math. Appl., vol. 72, no. 9, pp. 2535-2548, 2016.
[28] A. Laghrib, A. Hakim, S. Raghay, A combined total variation and bilateral filter approach for image robust super-resolution, EURASIP J. Image Video Process., vol. 2015, no. 1, pp. 1-10, 2015.
[29] A. Laghrib, A. Hakim, S. Raghay, M. El Dhabi, Robust super-resolution of images with non-parametric deformations using an elastic registration, Appl. Math. Sci., vol. 8, no. 179, pp. 8897-8907, 2014.
[30] C. Liu and D. Sun. On bayesian adaptive video super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 346-360, 2014.
[31] G.E. Marai, D.H. Laidlaw, J.J. Crisco, Super-resolution registration using tissue-classified distance fields, IEEE Trans. Med. Imaging, vol. 25, no. 2, pp. 177-187, 2006.
[32] N. Kumar, R. Verma, and A. Sethi, "Convolutional neural networks for wavelet domain super-resolution," Pattern Recognition Letters, vol. 90, pp. 65-71, 2017.
[33] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-srnet:n A wavelet-based CNN for multi-scale face super-resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1689-1697.
[34] K. Nasrollahi, T.B. Moeslund, Super-resolution: a comprehensive survey, Mach. Vis. Appl., vol. 25, no. 6, pp. 1423-1468, 2014.
[35] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning-based feature selection for remote sensing scene classification," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 11, pp. 2321-2325, 2015.
[36] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646-1654, 2016.

[37] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, IEEE Trans. Image Process., vol. 18, no. 1, pp. 36-51, 2009.
[38] D. Robinson, S. Farsiu, P. Milanfar, Optimal registration of aliased images using variable projection with applications to super-resolution, Comput. J., vol. 52, no. 1, pp. 31-42, 2009.
[39] H. Shen, L. Zhang, B. Huang, P. Li, A mapping approach for joint motion estimation, segmentation, and super-resolution, IEEE Trans. Image Process., vol. 16, no. 2, pp. 479-490, 2007.
[40] R.Y. Tsai, T.S. Huang, Multi-frame image restoration and registration, in: T.S. Huang (Ed.), Advances in Computer Vision and Image Processing, Green- wich, CT, JAI Press, 1984.
[41] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, L. Zhang, Image super-resolution: the techniques, applications, and future, Signal Process., vol. 128, pp. 389-408, 2016.
[42] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep CNN for image denoising, IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142-3155, 2017.
[43] S. Zhao, H. Liang, M. Sarem, a generalized detail-preserving super-resolution method, Signal Process., vol. 120, pp. 156-173, 2016.
[44] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Transactions on Image Processing, pp. 36-51, 2009.
[45] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. IEEE Transactions on Image Processing, pp. 1958-1975, 2009.
[46] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia. Video super-resolution via deep draft-ensemble learning. In IEEE International Conference on Computer Vision, pp. 531-539, 2015.
[47] C. Liu and D. Sun. On bayesian adaptive video super-resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 346-360, 2014.
[48] Hossam M. Balaha, Mohamed Saif, Ahmed Tamer, and Ehab H. Abdelhay, " Hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) for the early ultrasound diagnoses of breast cancer," Neural Computing and Applications, Jan 2022
[49] Marwa F. Areed, Mohamed M. Rashed, Nehal Fayez, and Ehab H. Abdelhay, "Modified SeDaSc system for efficient data sharing in the cloud," Concurrency Computation Practice and Experience, Vol. 33, Issue. 21, pp. e6377, May 2021.
[50] Ghada Mohamed Amer Ehab H. Abdelhay, Ibrahim Yasser Abdel-Baset, and Mohamed Abd El Azim Mohamed, "Development Machine Learning Techniques to Enhance Cyber Security Algorithms," MEJ. Mansoura Engineering Journal, Vol. 46, Issue. 4, pp. 36-46. Nov 2021.

### *Arabic Title:*

**استراتيجية فيديو فائق الدقة باستخدام النطاقات المتعددة للشبكات العصبية الملتفة المويجية**

### *Arabic Abstract:*

**غالبًا ما تكون الصور عالية الدقة مطلوبة ومطلوبة لمعظم التطبيقات ، لأنها تتضمن معلومات تكميلية. ومع ذلك ، فإن الاستخدام الأمثل لتكنولوجيا المستشعرات والتكنولوجيا المرنية لتحسين كثافة بكسل الصورة غالبًا ما يكون محدودًا ومكلفًا. نتيجة لذلك ، يعد استخدام طريقة معالجة الصور لإنشاء صورة عالية الدقة من صورة منخفضة الدقة خيارًا مكلفًا وشاملًا. الهدف من دقة الفيديو الفائقة هو استعادة النقاط المعقدة وتقليل التأثيرات الحسية. يعتمد هذا البحث على نهج الدقة الفائقة متعدد الإطارات باستخدام تحليل المويجات لتدريب الشبكات العصبية التلافيفية (CNN). لهذا الغرض ، يبدأ النهج بتطبيق تحليل المويجات على مقاطع الفيديو لتقييم متعدد المقاييس. بعد ذلك ، يتم تدريب العديد من شبكات CNN بشكل مستقل لتقريب الخصائص متعددة المقاييس المويجة. تقوم شبكات CNN المدربة بالاستدلال عن طريق تراجع التوصيفات متعددة المقاييس المويجة من إطارات LR ، متبوعة بإعادة البناء المويجي ، والتي تنتج إطارات HR مستردة. يقدم هذا البحث طريقة قائمة على التعلم للحفاظ على الميزات الدقيقة في الصور منخفضة الدقة متعددة الإطارات التي تم التقاطها باستخدام عدسات تكبير للكاميرا. تؤكد النتائج التجريبية الاستراتيجية المقترحة لاستعادة الأطر الصعبة.**