An efficient transfer learning-based Super-Resolution model for Medical Ultrasound Image

K Suseela¹ and K Kalimuthu²*
¹Department of Electronics and Communication Engineering, School of Engineering and Technology, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India
²Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India
Email: *kalimuthk@srmist.edu.in

Abstract. There are many imaging modalities for clinical diagnosis, but we heavily use ultrasound imaging; some of the ultrasound images are low resolution because of the body's internal weakness. This paper presents a novel unsupervised super-resolution framework for ultrasound image enhancement without considering any large sets of training samples which is the major concern in many single image super-resolution techniques. Here, the most powerful nonlinear mapping is introduced within convolutional neural networks to restore potentially useful and prominent spatial information generated from the model sets considered for image enhancements. Here, two techniques are used: dilated convolution and residual learning to increase the convergence and reconstruction accuracy in terms of quality measures. The potential metrics of dilated convolution in extract spatial information and residual learning are used to narrow the convergence time by learning only the difference between the test input and distorted input image. By evaluating the proposed framework on real ultrasound image sets, the performance metrics are validated. The proposed model outperformed the other competitive image enhancement models in the USSR field.

Keywords: Medical images, Residual learning network, Pre-trained Computer-aided diagnosis, Convolution layers, etc.

1. Introduction
In recent years, ultrasound imaging is predominantly used in many clinical diagnoses and has been emerging as the most useful imaging modalities to capture more complex organ delineation and soft tissues [1] as compared to additional medical images such as x-ray, CXR images, computed tomography (CT), also MRI images, etc., ultrasound imaging has numerous advantages such as less sensitiveness, cost-efficient, non-invasive, compatible, and real-time properties. Due to poor bandwidth and some intrinsic flaws in ultrasound images always obtained with low resolution, which limits its clinical applications [2] [3]. Development of the clinical applications for ultrasound images is emerging steadily using different image resolution enhancement techniques. Among other methods, data-based template approaches [4] and the image domain-based approaches [5] are widely used in many medical fields and investigated to improve ultrasound image resolution. Although this technique shows superiority to
improve ultrasound image resolution, this can cause tremendous device restrictions on the application because of its dependence on devices and frequency. The best alternative is any image processing technique and image-based techniques to achieve ultrasound image resolution improvement. A super-resolution method is used to reconstruct an output image with the most detailed restoration of information referenced with the input test set. A super-resolution is an effective technique; this technique is highly influenced in medical imaging [7], microwave engineering [8], and remote sensing applications in satellite imaging. Though various super-resolution techniques are investigated in ultrasound imaging interpolation-based techniques [10], reconstruction-based techniques and practice-based methods are used predominantly due to its most simplified computational models. In general, SR for ultrasound images follows deconvolution principles as a reconstruction process to solve the SR problems, such as oversampling and incorporating prior information into the process, crucial for visual quality. Reconstruction-based approaches effectively preserve geometric architecture, but they usually fail to restore sufficiently high frequencies [8-9]. Machine learning-based strategies [10] attempt to reconstruct the LR distorted image from an oversized variety of input pairs [11].

As compared to learning-based approaches, the learning primarily based model provides better performance. With the invention of deep learning using convolutional neural networks (CNNs), SR's system performance improved significantly. However, the performance metrics and achievable reconstruction quality of CNNs entirely on the availability [12] of external sources, which contain sufficient LR-HR image sets, which are not available in most ultrasound imaging. These CNN strategies can provide prominent performance only when the networks are trained with sufficient data. The performance will degrade when the training phase’s information is not sufficient, which is not the case in US imaging, restricting its utility in US fields. Also, most of the existing methods considered images with a single scale for training [13]. Even in some multiscale methods, the dimension elements are predetermined rather than arbitrary, which allows further research for some of the most realistic applications of the image enhancement technique [14].

This paper suggests a transfer learning-based image enhancement technique for ultrasound images to mitigate the problems of lack of train sets over many ultrasound data sets. The major contributions of the work are as follows:

1) **Unsupervised property:** This unsupervised property can overcome the limitation of data handling issues in the existing supervised models to generate the most useful information from the test image, which can restore the finite information in distorted test input, without considering any train sets for training.

2) **Arbitrary scale factor:** Here, the CNN network is trained with unique information extracted from the text input by changing its scale to any arbitrary values.

3) **Residual learning:** Here, the residual learning is incorporated to increase the training phase's convergence time and computational burden. It avoids a large volume of data, which carried identical information of LR and HR images. This residual architecture analyses only the residual difference, which is nothing but a finite distinction between distorted input images and measures considered in training.

4) **Receptive Field dimension:** In any SR system, the receptive discipline length acts as a dominant factor to decide the system performance [15]. Here through dilated convolution, the receptive field dimension is optimized and used to update as like the normal convolution. Here, through dilated convolution and large receptive discipline, the overall system performance is increased with the least computational cost overhead.

This work's contribution is prepared as follows: Section 2 investigates the basic methodologies involved in the proposed model, including various facts of processing, CNN layout, and validation measures. Section 3 presents the testing results that proved the performance metrics of the proposed
image enhancement model. The contrast with a modern ultrasound is SR technique, the potential metrics also provided in this section. Lastly, section 4 summarizes this work.

2. Learning-based Image Enhancement
In an ultrasound image, the given input image can be seen as a distorted version of the corresponding test image. The system of degradation can be computed as

$$x = DHy + n$$ (1)

where in x, also y is the discovered input image, besides the authentic test image individually (to reconstruct a quality image from distorted input images, by considering x as LR, also y as authentic test images). D refers to the down-sampling operator, H represents the decimation in addition to the blurring rate, also, n represents the additive noise [16]. Here, the SR can restore the quality output image from the poor-quality input image, which can be solved with the CNNs to study the different end-to-end mappings. The proposed SR framework is shown in Figure 1. It exploits the multiscale inner reappearance of facts in a single image, with a preference for extra statistics from outside statistics. An image-precise network is trained with the least complicated information generated from a pair of image sets given a test image. We then follow the ones learned relation of the family at the test image to supply the HR output. The suggested framework involves three basic processing stages: statistics processing, CNN network design and SR testing and validation.

### Theoretical aspect

**Flow Chart**

![Flow Chart](image)

**Figure 1**: Flowchart for Information Processing technique

#### 2.1 Data Processing
The information processing technique is exposed in Figure 1. The network extracts the useful information from the HR image for training and processes it into the LR image. The LR image is decomposed through downscaling, a transformed version of inputs based on chosen SR scale aspects.

The dataset should contain many samples to train CNNs, but it is not sufficient with the simplest US image datasets. Data augmentation is performed to artificially improve the records for network training to overcome this bottleneck. Initially, the test image is randomly downscaled into several smaller versions. These downscaled images are taken as the test set to training the details. Then, for every trial, the test image is downscaled using different scale factors to generate the reconstructed output images.
from distorted input entered into the network. In this manner, the training information along with numerous image pairs is formed [17].

2.2 CNN network
Existing CNNs always trained with a large set of samples and tended to minimize the loss function comprising various weights to capture all combination of different mappings [14],

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| F(x, \theta) - y_i \|^2.$$  \hspace{1cm} (2)

$x_i$, also $y_i$ denotes input and test images correspondingly, $\theta=\{W,B\}$ denotes the weights and bias components, $N$ denoted the total number of samples used in the training phase.

In general, $N$ is too large, which makes the CNN networks extremely complex. However, in the proposed framework, only lesser number of samples are considered for training and generated information are applied over the large sets of test images during validation. As a result, the variety of end-to-end mapping is used based on the information rate of given input images and information extracted via CNNs.

One network system is designed and examined at equal facts to attain improved quality metrics. Here, the proposed CNN network containing three convolution layers, including the baseline network. Each layer constitutes 64 channels, and the layer is equipped with a Rectified linear unit (ReLU) for feature activation. The input distorted image is interpolated to the same resolution as like test set model considered for experimentation. The influence of different CNN design elements (network depth, receptive area length, and residual mastering) on overall SR performance is considered. Figure 2 indicates the basic steps involved in the proposed CNN network.

![Architecture of CNN](image)

**Figure 2:** Architecture of CNN

2.3 Residual learning
In most cases, visually, both LR and HR images have similar information. However, the difference can be effectively discriminated by explicitly extracting the residual, which is the LR discrimination from the HR image. The similarities and data loss between the images are exploited through residual learning, reconstructing the output image from LR images. The residual is defined as follows:

$$r_i = Y_i - x_i$$  \hspace{1cm} (3)

The residual image is formulated from input sets, which minimizes the loss function.

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| F(x, \theta) - r_i \|$$  \hspace{1cm} (4)
The above figures are shown with the development of an ML model and train the same to classify no noise-free images and reconstruct the HR images from the LR images dataset to test the existing methods, compare the previous results, and present test results. Initially, the pre-processing stage is used to suppress the noise components, and then classification is performed with different data assessment methods. Finally, various algorithms' performance metrics are analysed, and comparative analyses are carried out based on its reconstruction accuracy. During classification, abnormal classes of the cardiac image are discriminated from normal classes.

The residual connection is established in the 8-layer convolution network to validate residual learning performance metrics in the overall system performance. The results proved that the residual network outperformed the non-residual network with improved convergence.

3. Result and Discussion

Comparative analyses are carried out with the different image sets to validate the proposed image enhancement technique, develop an ML model, and train the same to classify no noise-free images and reconstruct the HR images from LR images from a dataset. An image with dissimilar scale factors is considered for experiments to validate the proposed reconstruction methods US imaging performance metrics. Results proved that the proposed model gives superior results, both in quality and data restoration. Figure 3 shows ordinal cardiac images and Figure 4 discuses about speckled images. Figure 5 shows low resolutions and Figure 6 displays super resolutions images. Table 1 explains about performance metrics comparison with various scale factors over cardiac US image set. Figure 7 elaborates performance comparison of reconstructions measure and Performance comparison of Textual retention quality measure is show in Figure 8

![Figure 3](image1.png)
![Figure 4](image2.png)

**Figure 3:** (a) Original cardiac image, (b) Speckle image
**Figure 4:** (a) speckled image (b) De-Speckled image.
Figure 5: Low-Resolution Images

Figure 6: Super-Resolution, Bicubic Interpolation, and Reference

Existing values:

Table 1: Performance metrics comparison with various scale factor over cardiac US image set

| Data   | Scale | Dilated PSNR/SSIM | USSR PSNR/SSIM |
|--------|-------|-------------------|----------------|
| Cardiac| 2     | 39.98/0.970       | 40.2061/0.9718 |
| Cardiac| 3     | 36.60/0.930       | 36.6801/0.9366 |

Present values:

| Data   | Scale | Dilated PSNR/SSIM | USSR PSNR/SSIM |
|--------|-------|-------------------|----------------|
| Cardiac| 2     | 39.98/0.970       | 14.2061/0.9883 |
| Cardiac| 3     | 36.60/0.930       | 19.9314/0.9878 |

Figure 7: Performance comparison of reconstructions measure
4. Conclusion
Here, we are applying the USSR method for enhancing image resolution. We are taking cardiac and neuroimages as input. We apply the CNN model for test images to obtain super-resolution images without using any external test sets to train the networks. By incorporating dilated convolution in the CNN network and utilizing the residual approaches, the proposed framework gives superior system performance in terms of convergence rate and reconstruction quality. Dilated convolution with kernel size 8*8 is used for feature extraction. Experimental results proved that multiscale contextual information-driven image enhancement outperforms other state-of-art wavelet-based methods. We calculate accuracy and specificity as metrics for comparison in future scope to develop an efficient model for ultrasound images with unsupervised deep learning techniques to detect the tumour. To explore statistical analysis and establish a comparative analysis of the proposed model and existing models for ultrasound images. By utilizing an up sampling of fully connected networks type with improved detection quality, detection of the tumour area, and tumour size.

References
[1]. Shi, J., Liu, Q., Wang, C., Zhang, Q., Ying, S., & Xu, H. (2018). Super-resolution reconstruction of MR image with a novel residual learning network algorithm. Physics in Medicine & Biology, 63(8), 085011.
[2]. Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M. K., Sun, L., & Wang, G. (2018). 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE transactions on medical imaging, 37(6), 1522-1534.
[3]. Fang, L., Monroe, F., Novak, S. W., Kirk, L., Schiavon, C. R., Seungyoon, B. Y., & Manor, U. (2019). Deep learning-based point-scanning super-resolution imaging. BioRxiv, 740548.
[4]. Yin, S., & Bi, J. (2018, July). Medical image annotation based on deep transfer learning. In 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (pp. 47-49). IEEE.
[5]. Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., & Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature methods, 16(1), 103-110.
[6]. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J. H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. IEEE Transactions on Multimedia, 21(12), 3106-3121.
[7]. Li, Y., Sixou, B., & Peyrin, F. (2020). A review of the deep learning methods for medical image super-resolution problems. IRBM.
[8]. Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[9]. Kumar, N., & Sethi, A. (2016). Fast learning-based single image super-resolution. IEEE Transactions on Multimedia, 18(8), 1504-1515.

[10]. Zhang, S., Yuan, Q., Li, J., Sun, J., & Zhang, X. (2020). Scene-adaptive remote sensing image super-resolution using a multiscale attention network. IEEE Transactions on Geoscience and Remote Sensing, 58(7), 4764-4779.

[11]. Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: a comprehensive survey. Machine Vision and Applications, 25(6), 1423-1468.

[12]. Lyu, Q., Shan, H., & Wang, G. (2020). MRI super-resolution with ensemble learning and complementary priors. IEEE Transactions on Computational Imaging, 6, 615-624.

[13]. Chaudhari, A. S., Fang, Z., Kogan, F., Wood, J., Stevens, K. J., Gibbons, E. K., & Hargreaves, B. A. (2018). Super-resolution musculoskeletal MRI using deep learning. Magnetic resonance in medicine, 80(5), 2139-2154.

[14]. Yuan, Y., Zheng, X., & Lu, X. (2017). Hyperspectral image super-resolution by transfer learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5), 1963-1974.

[15]. Li, X., Hu, Y., Gao, X., Tao, D., & Ning, B. (2010). A multi-frame image super-resolution method. Signal Processing, 90(2), 405-414.

[16]. Park, J., Hwang, D., Kim, K. Y., Kang, S. K., Kim, Y. K., & Lee, J. S. (2018). Computed tomography super-resolution using deep convolutional neural network. Physics in Medicine & Biology, 63(14), 145011.

[17]. W. Linda, Z. Qiu Lin, and A. Wong. "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images." Scientific Reports 10, no. 1 (2020): 1-12.