

Applying Deep Learning Neural Network with Randomly Downscaled Image and Data Augmentation to Multiscale Image Enlargement

Ming-Tsung Yeh,^{1*} Wei-Yin Lo,² Yi-Nung Chung,² and Hong-Yi Cai²

¹Department of Electrical Engineering, National Chin-Yi University of Technology,
57, Sec. 2, Zhongshan Rd., Taiping Dist, Taichung 411030, Taiwan
²Department of Electrical Engineering, National Changhua University of Education,
No. 1, Jinde Rd., Changhua City, Changhua County 50007, Taiwan

(Received May 29, 2023; accepted November 30, 2023)

Keywords: image enlargement, advanced CAE, residual network, data augmentation

Digital image applications have been extensively utilized in entertainment, education, research, medicine, and industry. Many images should be resized for better demonstration. In general, image resizing is performed by conventional image processing technology. However, the enlarged image usually includes unacceptable amounts of noise, blurring, and jagged effects. A high resolution (HR) is usually required for output images. Applying a learning-based method to enlarge the input image and reconstruct the output to the HR image has better results. However, substantial external training datasets are required. In this study, we used the proposed randomly downscaled image and data augmentation (RDIDA) module to shrink images by the random scale and produce multiscale samples to reduce the dependence of preparation of significant datasets on the training stage. The image enlargement neural network (IENN) is proposed to apply deep learning neural network architecture based on an advanced convolutional autoencoder (CAE) to address the poor quality issues of output images. The proposed IENN with RDIDA can accept multiscale inputs and effectively enlarge images to specific sizes with high resolution. This learning-based approach with multiple residual networks is different from other methods. Applying the encoder of an advanced CAE structure captures features of the original image, and then the decoder with residual structure can create an enlarged image with HR quality. The CAE network used to enlarge an image can effectively denoise and reduce distortions that smooth out the traditional processing drawbacks. Our experimental results show that the peak signal-to-noise ratio (PSNR) of validation for our proposed model has been over 29.55 dB at epoch 30 during the training stage. Furthermore, this model can achieve an average PSNR above 26 dB on all test samples to demonstrate robust performance.

*Corresponding author: e-mail: mtseh@ncut.edu.tw
<https://doi.org/10.18494/SAM4531>

1. Introduction

In many situations, images should change their shape, size, or orientation for better demonstration. It is sometimes necessary to shrink images to fit on the proper screen for mobile devices or Web pages. To improve visual effects, images must be enlarged to adapt to the screen, such as in the home cinema system. Image rescaling was performed by conventional image processing technology in the past. An affine transformation method can be used to resize simple images with only lines and plain patterns and obtain good results.⁽¹⁾ However, pictures with complicated colors and patterns, such as scenic images, are hard to rescale, significantly when enlarged. Enlarging images with complex scenes by prior image technologies generally has additional noises and blurred, jagged effects on the outcomes.

Image enlargement is generally designed from a small-scale image to a large-scale image with a low resolution (LR). The resulting image requires some reconstruction to obtain a high-resolution (HR) image and restore missing pixel values from LR. There are three types of mode, namely, the interpolation, reconstruction-based, and learning-based methods, which are applied to rescale images and reconstruct their image resolution. Popular image processing applications often use traditional interpolation approaches to resize images. Enlarging an image by the nearest neighbor method is rapid, but the result has an unacceptable mosaic effect and jagged edges. The bilinear, bi-cubic, or bi-quartic interpolation is smoother than the nearest neighbor method. However, the outcome also has a certain blurriness. All poor effects are shown in Fig. 1. These effects are inevitable because the interpolation approaches cannot predict enlarged image pixel values around the current pixel point. Many researchers have proposed some methods to overcome these issues. Han *et al.* proposed a novel interpolation framework to suppress blurring and jaggging.⁽²⁾ Karim proposed a rational bi-quartic spline with six parameters for surface interpolation applied in grayscale image enlargement and obtained a higher peak signal-to-noise ratio (PSNR).⁽³⁾ The blurring and artifact effects still exist because the enlarged image's pixel value cannot be produced from nothing. All pixel values are produced by estimation.

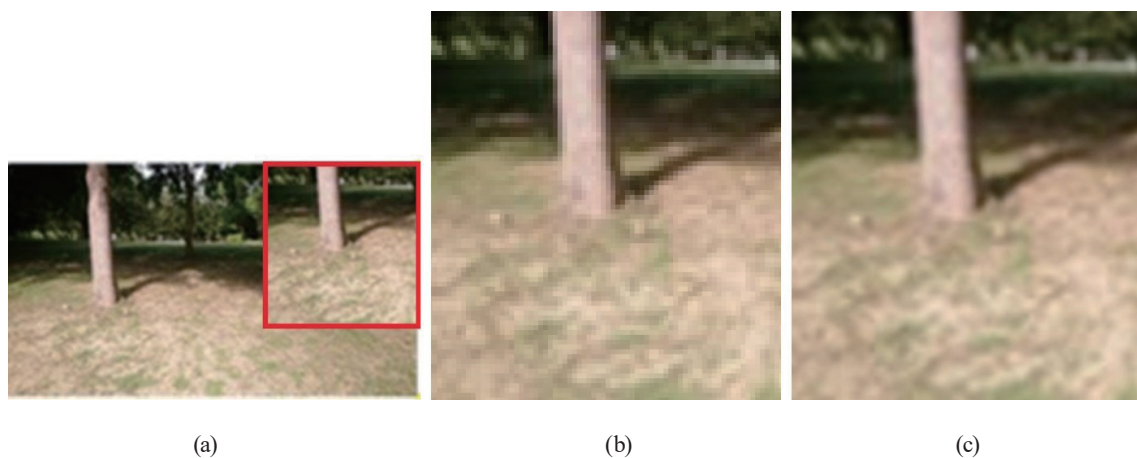


Fig. 1. (Color online) Image with poor effects after enlarging four times by interpolation method. (a) Original image. (b) Image enlarged by nearest neighbor method. (c) Image enlarged by bi-cubic method.

The construction-based methods use the prior knowledge of neighborhood processing to design kernel regression and reconstruct the upscale LR image. Michaeli and Irani used a global type of blind super resolution (SR) that is an optimal blur kernel to construct an HR image.⁽⁴⁾ Schreiter *et al.* applied an iterative Bayesian technique to produce a locally smoothed mixture regression to restore the HR image.⁽⁵⁾ Li *et al.* used self-similarity prior technology to extract the directional group sparsity of image gradients and directional features, and apply the framework of templates for first-order conic solvers to obtain a high-quality HR image.⁽⁶⁾ This method has the significant advantage of not requiring external datasets to train parameters of the kernel regression. However, the ground true HR image, which is converted to the degraded LR mode by blurring or downscaling, must be similar to the LR input. Another drawback is that the performance is degraded if the enlarged scale of the image is greater than common small-scale operations.

The learning-based methods require many external datasets to train neural network parameters. Artificial neurons capture the sample features of training datasets and update network weight parameters. The designed neural network is adjusted to adapt the input for learning knowledge during iteration. After the training stage, the network can find the mapping relationship between the input and output HR images such that the learning knowledge is stored in the weight parameters. Owing to the rapid development of modern GPU computing technology, the deep learning model with convolutional neural network (CNN) has been extensively applied to reconstruct HR images and obtain better performance. Dong *et al.* proposed super-resolution CNN (SRCNN) that only used three convolutional layers to extract the LR image features and perform nonlinear mapping to reconstruct the HR image.⁽⁷⁾ The performance of SRCNN is better and more stable than those of traditional methods. Jiang *et al.* used optimal subpixel CNN to improve the output HR image quality and PSNR.⁽⁸⁾ To improve the HR image reconstruction performance, the neural network increases its layer depth with residual connections in some works, which can prevent learning stagnation and vanishing gradients. Chen and Qi⁽⁹⁾ and Shaoshuo *et al.*⁽¹⁰⁾ used skip dense residual networks to increase the network layer depth. Basak *et al.* applied long and short skip connections to recover the HR image and obtain better results than other deep learning-based methods.⁽¹¹⁾ All the approaches above deeply rely on external datasets to learn the mapping rules. However, diverse and multiscale images are collected and processed for training datasets that require external jobs and time. For the most part, the input and HR output image sizes in the designed methods must be a demanded scale. Liu *et al.* even proposed using an internal dataset to reconstruct HR images that require another neural network to find the patch mapping relationships before recovering each LR image.⁽¹²⁾

In this paper, we propose a deep-learning neural network to enlarge and produce HR images. The images in training datasets are automatically downscaled to randomly arbitrary sizes by the proposed randomly downscaled image and data augmentation (RDIDA) module. This RDIDA module produces training samples and performs data augmentation to increase diversity. These training samples as inputs of the proposed image enlargement neural network (IENN) created by the advanced convolutional autoencoder (CAE) structure are used to train the neural network. After completing the training stage, the IENN receives training samples with implicitly arbitrary

sizes and can obtain the mapping knowledge between random-size and enlarged HR images. In this paper, we also introduce a novel data sampling and augmentation approach, RDIDA, which can reduce the load to collect training datasets and augment limited samples. The trained neural network can enlarge an image of any size to the HR image with specified scales.

2. Devices and Proposed Methods

2.1 System framework

In this study, the system framework is partitioned into two parts. First, the learning stage automatically applies the proposed RDIDA method to produce arbitrary-size samples, as shown in Fig. 2. The RDIDA module is designed using bi-cubic interpolation and transpose convolution neural network to perform data augmentation, randomly downscales images to any size, and produces the desired size outputs. The IENN module applies these output samples to train its neural network. The proposed IENN is based on the convolutional autoencoder with an asymmetric residual connection. After repeated iterations, the IENN completes mapping knowledge learning. Second, the inference stage uses the trained IENN combined with the RDIDA sampling module, which can accept arbitrary-size photos to be enlarged to a specific HR image.

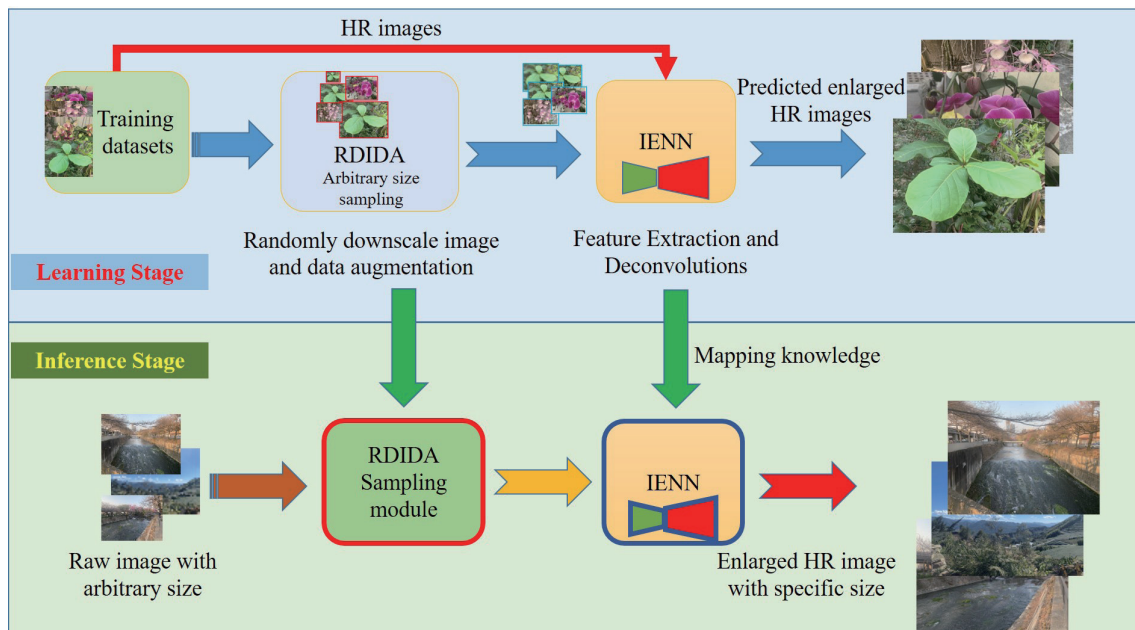


Fig. 2. (Color online) System framework.

2.2 Devices

In this study, we utilized a USB plug-and-play HD webcam, as shown in Fig. 3. The C615 WEBCAM features a 78° diagonal field of view and an auto light correction mechanism called RightLight 2. Additionally, the webcam is equipped with a noise-canceling omnidirectional microphone. The video recordings captured by the webcam are available in Full HD 1080p resolution at 30 frames per second (fps) and HD 720p resolution at 30 fps. The specifications of the webcam are summarized in Table 1.

2.3 Proposed methods

2.3.1 Datasets

Within the scope of this paper, the DIV2K dataset⁽¹³⁾ is employed to conduct a comparative analysis of various model structures and conventional image processing techniques. This dataset has 1000 images with a large diversity of contents and 2K resolution. To ensure good balance in the dataset, the creators separate partitions of 800 train, 100 validation, and 100 test images by rigorous methods. Moreover, the dataset has been meticulously curated to ensure a diverse range of contents with minimal corruption. In this study, the DIV2K dataset was used for image enlargement neural network training.



Fig. 3. (Color online) C615 WEBCAM.

Table 1
Specifications of C615 WEBCAM.

Overview	C615 WEBCAM
Resolution FPS	Full HD 1080p/30fps, HD 720p/30fps
Diagonal field of view	78°
Auto light correction	RightLight 2
Noise-canceling mic (s)	One omnidirectional mic
Connection	USB - A plug-and-play

2.3.2 RDIDA module

The proposed RDIDA module produces the training samples of the IENN. The RDIDA has three submodules, namely, data augmentation, randomly downscaled, and sampling modules, as shown in Fig. 4.

Image augmentation plays a crucial role in deep learning model training. It reduces the issue of overfitting and enhances model performance by providing additional data for neural network learning. In this study, we propose a simple data augmentation technique to augment the training data. Precisely, each image is cropped to the final output sizes that the IENN requires for correct tags, thereby preserving more details than traditional image processing approaches, such as the nearest neighbor or bi-cubic interpolation approach. The image crop schematic diagram is shown in Fig. 5.

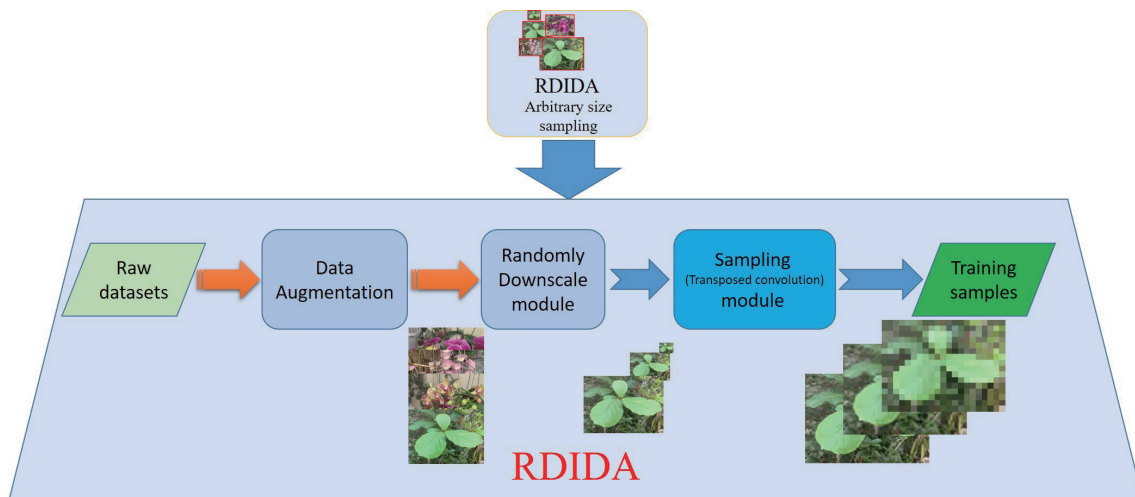


Fig. 4. (Color online) RDIDA module.



Fig. 5. (Color online) Image crop schematic diagram.

Cropped images are downsampled by a random scale factor in the randomly downsampled submodule. This submodule design is based on the bi-cubic interpolation approach used to shrink images irregularly and meets the arbitrary input image size for IENN training. The bi-cubic interpolation function⁽¹⁴⁾ is shown in Eq. (1).









$$g(x) = \sum_k c_k u\left(\frac{x-x_k}{h}\right) \quad (1)$$

Here, h represents the sampling increment, x_k 's are the interpolation nodes, u is the interpolation kernel, and g is the interpolation function.

The sampling submodule based on the transposed convolutional neural network combined with the bi-cubic interpolation method produces specific size outputs for IENN training samples. The RDIDA module implicitly creates random-size samples, making the training samples more diverse.

Table 2 shows some sampling examples. The original image is downsampled to one-sixth size, and then the sampling module produces an up-sampled image to the specific extent as the input data of the IENN module in the first data row. The IENN can learn mapping knowledge of implicit six times scale from this sample. The second row can provide an implicit 4.5 times enlargement learning sample; the adjacent row is three times objects. The second and final row samples can have non-integer magnification. This shows that the RDIDA module can automatically produce random-size samples with integer or non-integer scales that provide diverse learning for the IENN module. This demonstrates that this system can accept any size input and magnify images to specific size outputs with HR quality for future implementation.

Table 2
(Color online) RDIDA output sampling examples and their implicit scale factors.

Randomly downsampled images	Up-sampled images	Implicit scale factor
		6
		4.5
		3
		2.5

2.3.3 IENN method

While SR has been extensively explored in deep learning research, its application in different input sizes should be studied further. To address this problem, the IENN is proposed in this study. The IENN can accept many different input sizes to achieve the nonlinear enlarged magnification. Figure 6 shows a detailed process from the input image to produce an enlarged output using the IENN structure.

In Fig. 6, the sampling submodule of the RDIDA module is introduced to meet the random input image size and convert inputs to the designated size of the IENN. By employing this technique, the IENN can eliminate the restriction that it can enlarge images only up to two times, and the IENN can modify its enlarged magnification with various input sizes. The IENN is based on the advanced CAE,⁽¹⁵⁾ and the model structure is shown in Fig. 7. This advanced CAE structure is designed by the encoder-decoder framework with residual network structures. To smooth the output image, no activation function is trailed behind the last layer in the decoder. The asymmetric skip connections are applied to the IENN, which can enable the fusion of multiscale features and allow the network to capture global context and local details necessary to preserve more detailed features and stabilize the network to produce a more smoothly enlarged image.

The definition of the IENN is shown in Eqs. (2) and (3). The computations of the encoder and decoder operations are derived from Eq. (4). The IENN employs the mean squared error (MSE)

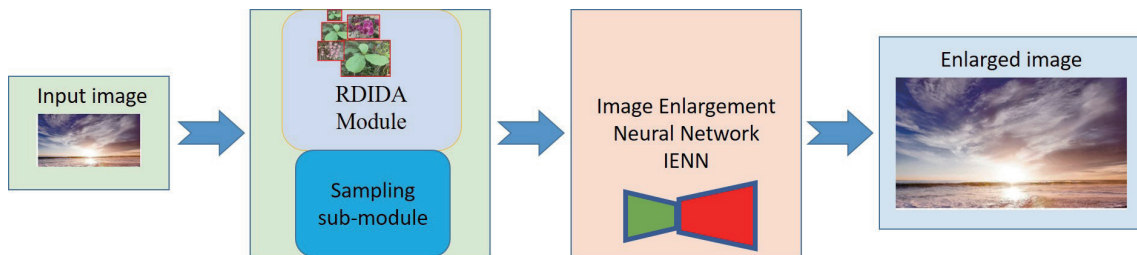


Fig. 6. (Color online) Image enlargement process for IENN structure.

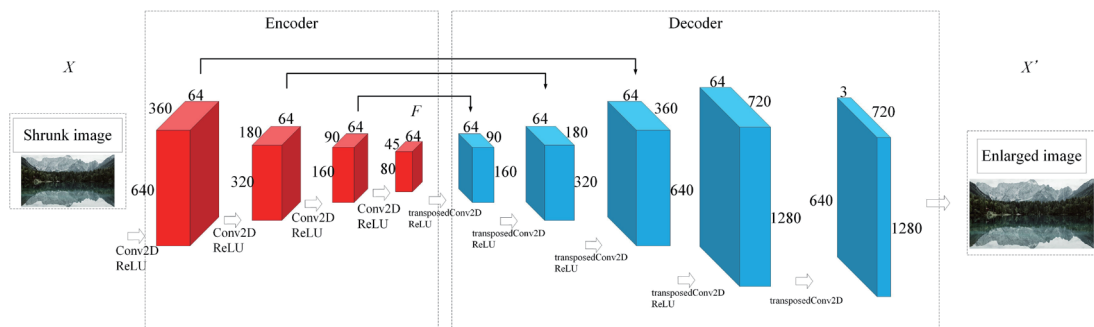


Fig. 7. (Color online) IENN interior structure in this study.

as the loss function during the backpropagation process. The MSE loss function is shown in Eq. (5).

$$\psi, \varphi = \arg \min_{\psi, \varphi} \|X' - (\varphi \circ \psi)X\|^2 \quad (2)$$

$$\psi: X \rightarrow F, \varphi: F \rightarrow X' \quad (3)$$

Here, X is the input vector, F is the feature vector, X' is the target vector, ψ represents the shrunk images' exact feature vector process, and φ shows that the feature vectors reconstruct the target image process.

$$\text{If } \mathbf{x} \in \mathbb{R}^d = X, \text{ if } \mathbf{h} \in \mathbb{R}^p = F, \begin{cases} \mathbf{h} = \sigma(W\mathbf{x} + \mathbf{b}) \\ X' = \sigma'(W'\mathbf{h} + \mathbf{b}') \end{cases} \quad (4)$$

$$L(\mathbf{x}, X') = \|\mathbf{x} - X'\|^2 = \|\mathbf{x} - \sigma'(W'(\sigma(W\mathbf{x} + \mathbf{b})) + \mathbf{b}')\|^2 \quad (5)$$

Here, W and W' are the encoder and decoder weights, \mathbf{b} and \mathbf{b}' are the encoder and decoder weighting biases, and σ and σ' are the encoder and decoder activation functions, respectively.

3. Experimental Results and Discussion

In this paper, we propose the IENN model combined with the RDIDA sampling module to achieve nonlinear image-enlarged magnification. In this section, the experimental results demonstrate the loss curves and PSNR metrics for the IENN and SRCNN models, respectively, which evaluate and compare the training performance characteristics between different approaches.

3.1 IENN and SRCNN loss graphs

The IENN and SRCNN loss graphs are shown in Figs. 8 and 9, respectively. The total number of training epochs is set to 30. Figure 8 shows that the loss curve of the IENN model can converge to 0.0016 within five epochs. The loss value converges to 0.001421 at the end. Figure 9 shows that the loss curve only converges to 0.001602 in the SRCNN model. The IENN model exhibits a more robust performance during the training stage.

3.2 IENN and SRCNN PSNR graphs

The PSNR is a valuable metric for evaluating the output performance of image enlargement, which can reveal the realistic level of predicted results. The PSNR calculation formula is shown in Eqs. (6) and (7). Figures 10 and 11 show comparisons of training and validation PSNR values

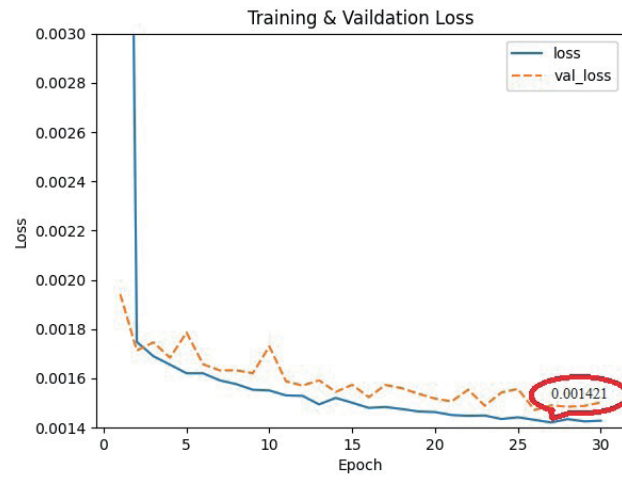


Fig. 8. (Color online) IENN loss graph in this study.

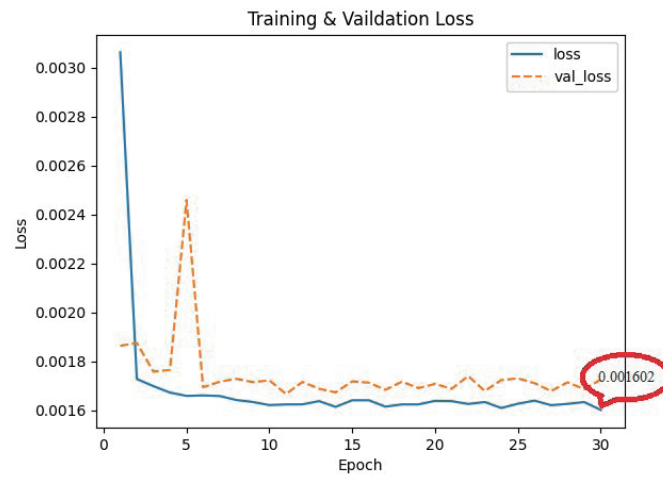


Fig. 9. (Color online) SRCNN loss graph in this study.

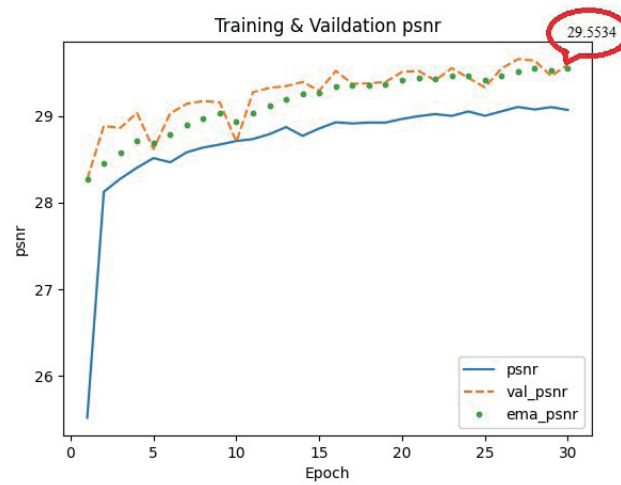


Fig. 10. (Color online) IENN PSNR graph in this study.

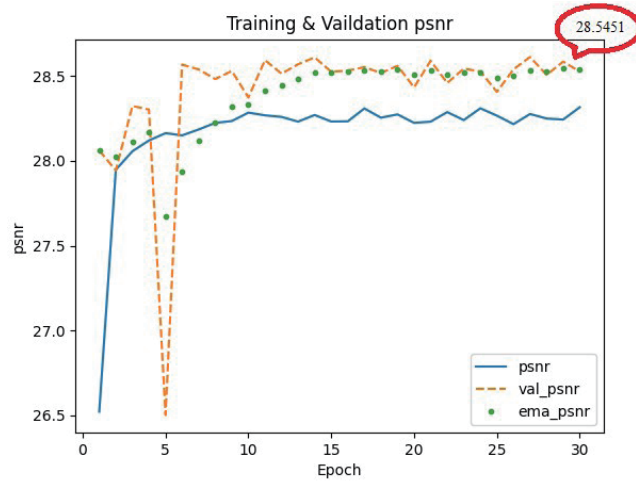


Fig. 11. (Color online) SRCNN PSNR graph in this study.

for the SRCNN method and our model. The IENN model achieves a PSNR of 29.5534 dB, whereas the SRCNN model only shows a PSNR of 28.5451 dB. This demonstrates that the IENN model has better enlarging outputs.

$$PSNR = 10 * \log_{10} \left(\frac{1.0^2}{MSE} \right) = 20 * \log_{10} \left(\frac{1.0^2}{\sqrt{MSE}} \right) \quad (6)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (7)$$

Here, Y_i is the pixel value of the ground truth image and \hat{Y}_i is the pixel value of the predicted HR enlargement image. The maximum value of the PSNR formula has been reduced by a factor of 1 owing to the normalization of training images.

3.3 Image enlargement results compared with other methods

In this study, we compare image enlargement results with the IENN, SRCNN, and traditional bi-cubic approaches. The validation PSNR results of different methods that use the same datasets to train networks are shown in Table 3, which shows that the proposed IENN has the best performance and reaches the PSNR of 29.5534 dB. The predicted results of varied, complex images with the different enlargement approaches are shown in Fig. 12, demonstrating that the IENN method has better outputs. Figure 13 shows the results of multiscale enlargement for small cropped parts of a bigger photo, which reveals that different enlargement scales can obtain the same clear outputs with high resolution.

Table 3
Validation *PSNR* results of different methods.

Method	<i>PSNR</i>
Bi-cubic	25.648
SRCNN	28.5451
Proposed IENN	29.5534

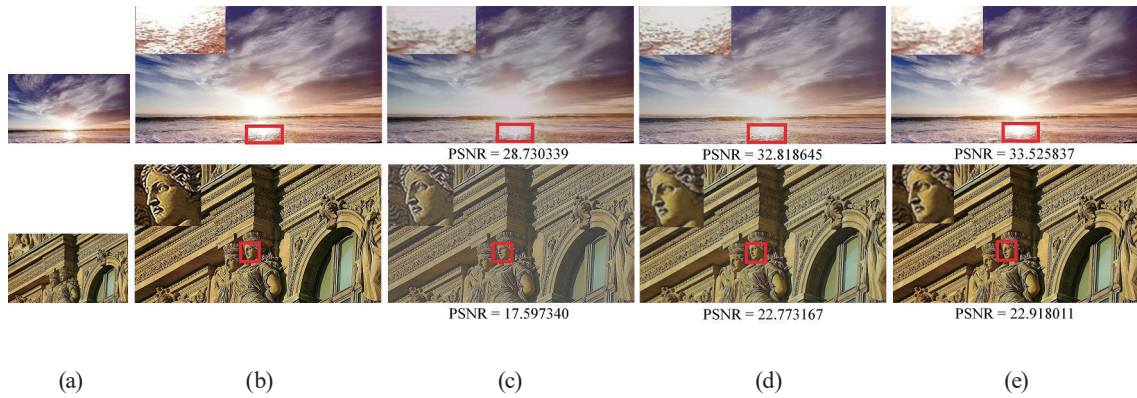


Fig. 12. (Color online) The experimental results compare performance characteristics with different approaches. (a) Shrunk images. (b) Original images. (c) Images enlarged by bi-cubic method. (d) Images enlarged by SRCNN method. (e) Images enlarged by IENN method.

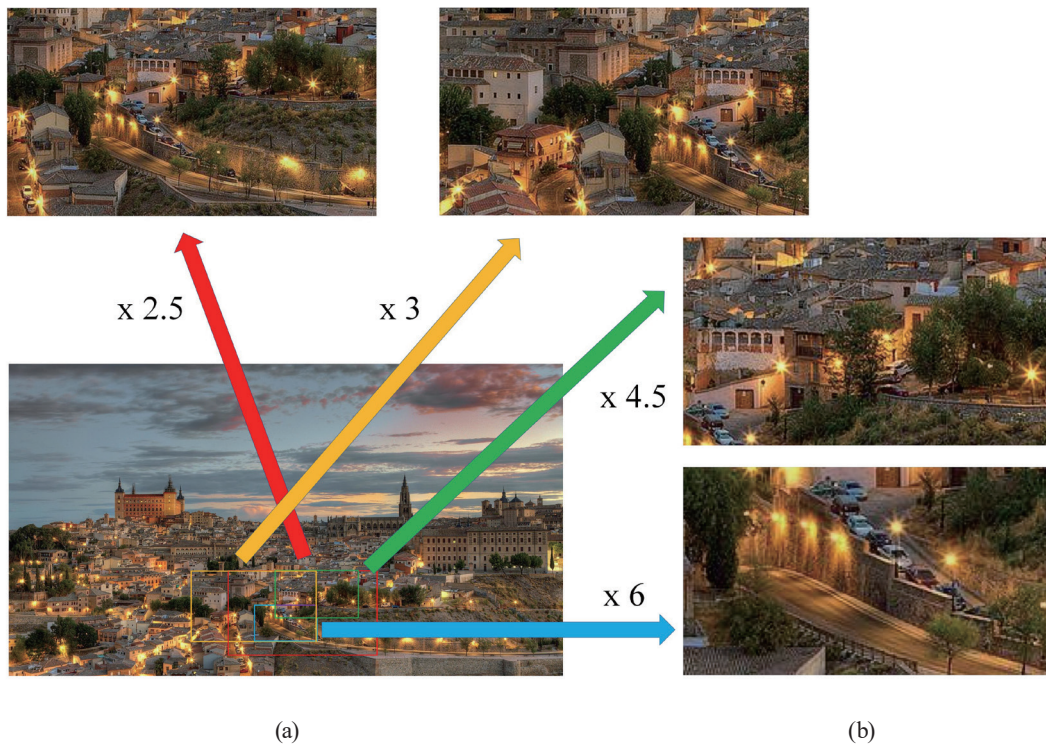


Fig. 13. (Color online) Multiscale enlargements for small cropped parts of image by IENN method. (a) Original images. (b) IENN output images corresponding to different regions.

3.4 Discussion

Figure 12 shows how different complex images perform in other enlargement methods. In this study, the image enlargement approach is affected by the complexity of the pictures. There are some methods of solution to address this problem. First, we aim to use more datasets, such as ImageNet, to obtain more enlargement learning in the neural network and smooth predicted results. Second, a more powerful structure, such as the generative adversarial network (GAN), can add to the IENN module and improve network performance. However, the GAN structure implemented in image enlargement causes training time and network complications to increase, and the artifact effects of output images need extra work to remove.

4. Conclusions

The RDIDA module was proposed in this paper and applied to produce multiscale samples, which reduces the dependence on the requirement of great training datasets. This module can make the IENN accept samples with different implicit scales to train and acquire the ability to enlarge arbitrary input image sizes and high-resolution outputs. In this paper, we evaluated the performance of the IENN model by comparing it with those of the SRCNN model and bi-cubic interpolation using the PSNR metric. The experimental results showed that the IENN model has a better PSNR to achieve 29.5534 dB and outperform the other methods. The output images of multiscale enlargements performed by our model with higher resolution quality demonstrate better performance.

Acknowledgments

This work was supported by the National Science and Technology Council under Grant NSTC 111-2222-E-167-003.

References

- 1 A. Mcandew: Introduction to Digital Image Processing with Matlab (CENGAGE, Taipei, 2010) Asia Ed., Chap. 6.
- 2 J. Han, J.-H. Kim, S. Cheon, J.-O. Kim, and S. Ko: IEEE Trans. Consumer Electronics **56** (2010) 175. <https://doi.org/10.1109/TCE.2010.5439142>
- 3 S. A. Abdul Karim: Proc. IEEE Access (IEEE, 2020) 115621–115633. <https://doi.org/10.1109/ACCESS.2020.3002387>
- 4 T. Michaeli and M. Irani: Proc. 2013 IEEE Int. Conf. Computer Vision (IEEE, Sydney, 2013) 945–952. <https://doi.org/10.1109/ICCV.2013.121>
- 5 C. Schreiter, J. Sun, and P. Schelkens: Proc. 2018 25th IEEE Int. Conf. Image Processing (ICIP, Athens, 2018) 400–404. <https://doi.org/10.1109/ICIP.2018.8451703>
- 6 X. Li, H. He, R. Wang, and D. Tao: IEEE Trans. Image Processing **24** (2015) 2874. <https://doi.org/10.1109/TIP.2015.2432713>
- 7 C. Dong, C. C. Loy, K. He, and X. Tang: IEEE Trans. Pattern Anal. Mach. Intell. **38** (2016) 295. <https://doi.org/10.1109/TPAMI.2015.2439281>
- 8 P. Jiang, W. Lin, and W. Shang: Proc. 2021 IEEE Int. Conf. Power Electronics, Computer Applications (ICPECA, Shenyang, 2021) 132–136. <https://doi.org/10.1109/ICPECA51329.2021.9362548>

- 9 C. Chen and F. Qi: Proc. 2018 9th Int. Conf. Information Technology in Medicine and Education (ITME, Hangzhou, 2018) 999–1003. <https://doi.org/10.1109/ITME.2018.00222>
- 10 M. Shaoshuo, Z. Yanhua, Q. Xiaolan, and J. Yanbing: Proc. 2021 IEEE Int. Conf. Multimedia and Expo (ICME, Shenzhen, 2021) 1–6. <https://doi.org/10.1109/ICME51207.2021.9428084>
- 11 H. Basak, R. Kundu, A. Agarwal, and S. Giri: Proc. 2020 IEEE 15th Int. Conf. Industrial and Information Systems (ICIIS, RUPNAGAR, 2020) 219–224. <https://doi.org/10.1109/ICIIS51140.2020.9342688>
- 12 J. Liu, Y. Xue, S. Zhao, S. Li, and X. Zhang: Proc. IEEE Access (IEEE, 2020) 201055–201070. <https://doi.org/10.1109/ACCESS.2020.3036155>
- 13 E. Agustsson and R. Timofte: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW, Honolulu, 2017) 1122–1131. <http://dx.doi.org/10.1109/cvprw.2017.150>
- 14 R. Keys: IEEE Trans. Acoust. Speech Signal Process. 29 (IEEE, 1981) 1153. <https://doi.org/10.1109/tassp.1981.1163711>
- 15 X. Sun, A. Gossmann, Y. Wang, and B. Bischt: Proc. 2019 IEEE Symp. Ser. Computational Intelligence (SSCI, Xiamen, 2019) 1344–1353. <https://doi.org/10.1109/SSCI44817.2019.9002665>

About the Authors



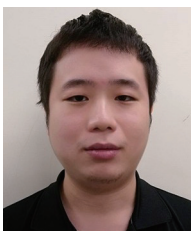
Ming-Tsung Yeh received his M.S. and Ph.D. degrees from National Changhua University of Education, Taiwan, in 2012 and 2016, respectively. Since 2022, he has been an assistant professor at National Chin-Yi University of Technology, Taiwan. His research interests are in AI, deep learning, image processing, and intelligent control. (mtveh@ncut.edu.tw)



Wei-Yin Lo received his B.A. degree in electrical engineering from National Changhua University of Education, Taiwan, in 2021. He is currently working toward his M.S. degree in electrical engineering at National Changhua University of Education, Changhua, Taiwan. His research interests are in machine learning and image processing. (z2815414@gmail.com)



Yi-Nung Chung received his Ph.D. degree from Texas Tech University, Lubbock, TX, USA, in 1990. He is currently a professor at National Changhua University of Education, Changhua, Taiwan. He is also the Dean of Academic Affairs. His research interests include image processing and computer vision. (yunchung@cc.ncue.edu.tw)



Hong-Yi Cai received his M.S. degree in electrical engineering from National Changhua University of Education, Taiwan, in 2022. He is currently working for Universal Global Technology Co., Limited, Taiwan. His research interests are in machine learning image processing. (st990538@gmail.com)