Advances in Economics, Business and Management Research, volume 215
Proceedings of the 2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022)

Enhancement of Dim Imaging Enlargement using Super-Resolution CNN

Ziwei Li

Undergraduate Student of The Chinese University of Hong Kong, Shenzhen
Email: 118010160@link.cuhk.edu.cn

ABSTRACT
Dim images are an important branch of various images. Due to the limitations of equipment and technology, it is often impossible to obtain satisfactory results when shooting pictures and videos under night scenes with a limited budget. It is still a blue ocean to increase the resolution of the pictures or videos such as data recovery under night scene monitoring and real-time optimization when taking pictures with mobile phones. The existing method that can increase the resolution the most is SRCNN, but an ordinary SRCNN model is not optimized for the characteristics of dim pictures and runs slowly. Therefore, this paper attempt to introduce an optimized dim-srcnn model with a faster speed for a picture or a video that contains a large number of pure black areas, has fewer features, and the entire picture is dim.

Keywords: *Deep-learning, super-resolution, convolutional neural network*

1.INTRODUCTION

Fig. 1. Sample Input and output of a test image upscaled by Nearest-Neighbor method by 4x and predicted by D-SRCNN model

Image super-resolution (SR), as the name suggests, is aiming at improving the quality of images, which is one of the problems in the computer vision industry today. There are two main reasons for poor image quality: hardware reason and physical reason. First of all, not just old images in the past, but now, due to hardware limitations, such as the technology of manufacturing cameras, many images, such as satellite images, are still not very clear, which makes it temporarily not possible to achieve ideas such as counting the number of ships in ports through satellite images. Secondly, limited by the frequency of light waves, the image quality has a theoretical upper limit. Electron microscopes use higher energy and higher frequency electrons to break through this upper limit, but they are limited by the spreading distance of electrons and cannot be used as a normal camera to take photos. Therefore, not only at the present, even in the future when the camera technology has achieved great breakthroughs, SR also has a very high development prospect.

Existing SR methods include: (i) reconstruction-based methods, (ii) sample-based methods, and (iii) deep-learning-based methods [13]. Among them, the sample-based method is the most widely accepted, but it also has the most distortion. For example, the bicubic interpolation method has long been widely used for ordinary image enlargement because of its simple algorithm, easy implementation, and small computing resources required for scaling images up or down. The method based on deep learning is an emerging image resolution enhancement technology. Deep learning that has been widely developed in recent years, is used to train the model and obtain the highest possible image quality.

Deep learning is a subfield of machine learning. Since 2000, deep learning has made great progress. In the form of neural networks, it has achieved great success in both unsupervised and supervised learning and has developed successful cases such as CNN, RNN, and ISTM [9]. Deep learning has long been used in the latest technologies that include text generation, speech recognition, and motion capture, which are closely related to public life.

The development of image SR has gradually stagnated. Until 2014, a paper published by Dong [1] introduced deep learning into the field of image SR. Since then, people have gradually improved the network design, so that the high-frequency details of the high-resolution (HR) image have been improved by degrees. A improvement, which is accelerating method, on SRCNN are also introduced and proved feasible by Dong [2]. Although Kant thought GANs (generative adversarial networks) may provide a better solution than SRCNN [7], SRCNN model has been widely used in medical research, such as CT-SRCNN introduced by Ren [5], application on the chest-CT [16] and RRLSRN for MRI from Song [12]. Besides, facing the threat of COVID-19, Jain attempted to use SRCNN in the classification of whether the person has the coronavirus or not [6]. A special landslide srcnn model has also been developed by Mohan, which proves that it is feasible to develop a specific SRCNN model for a certain characteristic of the picture [10]. However, the general SRCNN model costs almost 1 second on an ordinary Macbook Pro to get a HR image. To speed up the calculation, the filter size or filter number should be decreased, which is not practical because most of the images have many features that require a number of filters to capture them. However, dim images with fewer features could be an exception. Therefore, in this article, a SRCNN model (D-SRCNN) optimized for dim pictures will be introduced.

2.RELATED WORKS

2.1.Machine-Learning (ML)

Machine learning is an attempt to use computers to imitate human learning behaviors, which acquire new skills and reorganize the existing knowledge structure, to continuously improve the computer's performance. Originally, the machine is only used for a specific purpose. ML tries to make the machine understand the meaning of the tasks and gain experience from performing tasks, and then use that experience to accomplish missions that vary a bit. AlphaGo is a good example and is one of the famous achievements of ML, which can deal with different decisions of opponents through learning and shows people the unlimited potential prospects of machine learning.

2.2.Image Super-Resolution (SR)

The image Super-Resolution is a category of image upscaling methods that are able to transform low resolution (LR) images to high resolution (HR) images. The SR methods can be generally divided into three categories. The reconstruction-based methods category contains frequency-domain methods and spartial-domain methods. The sample-based methods category includes similarity-based methods, nearest-neighbor method, and sparse method. The last category, SRCNN, has been proved to have a better performance in upscaling the LR images to HR images [1].

2.3.SRCNN

SRCNN is short for Image Super-Resolution Convolutional Neural Network. SRCNN is the first SR algorithm using CNN (based on ML). CNN is a definition borrowed from recent achievements of machine learning [8], which is a kind of neural network but has some similarities to the biological neural network. CNN has three layers: (i) convolutional layer, (ii) subsampling layer, and (iii) full-connected layer. Its structure reduces the complexity of the network model, reduces the number of weights, and has many advantages in two-dimensional image processing. The overall procedure of SRCNN consists of (i) patch extraction and representation, (ii) non-linear mapping, and (iii) reconstruction [1].

3.METHODOLOGY

Traditional methods include: (i) nearest-neighbor, (ii) bilinear and (iii) bicubic interpolation, and (iv) SRCNN will be compared with our dim-specialized SRCNN algorithm (D-SRCNN). PSNR and SSIM will be used as benchmarks to measure image quality.

3.1.Image Source

We used vcg.com as our image source. 48 sample dim jpeg images with 1024 x 437 px (10.84 x 4.62 cm) are downloaded for training, and 12 images for testing.

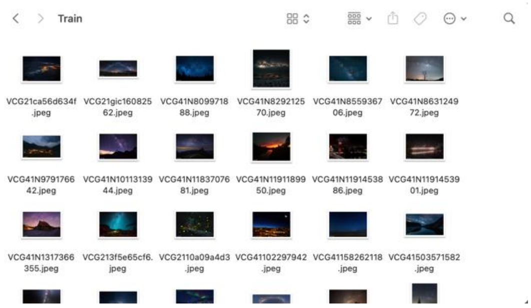

Fig. 2. Part of the training dataset

3.2.Dataset Preparation

Each original image in the training dataset is preprocessed before training: (i) First of all, RGB images are converted into YCbCr from the original images. (ii) The edge of the images are cut first to fit the scale with size 3. (iii) Then they are downscaled by bicubic interpolation to 0.33x and upscaled to original size to decrease their resolution to be our input training data set and divided by 255 keeping only 4 decimals. (iv) After that, they are cut into images with the size of 33*33 using a stride of 14 pixels, and (v) only Y-channels are kept.

The detailed reason why doing theses will be explained in part4.2 of this paper. The label dataset used to calculate loss remains original resolution but is also cut into images with the size of 27*27 because our model is designed to downscale the image size by 6 pixels. The input dataset and label dataset are saved as one h5 file.

The test dataset is handled the same as the training dataset except they are not split into several small images.

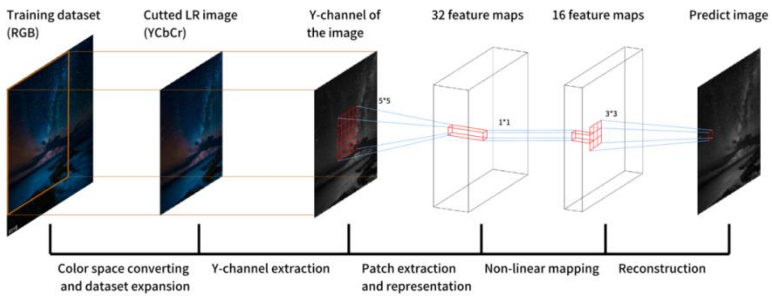

Fig. 3. An illustration of the model training process

3.3.Training Configuration

The computer used for training has a CPU of AMD TR 3960X and GPU of NVIDIA GeForce RTX 2070 SUPER. Python3.9.1 is used to train the srcnn model along with Tensorflow2.7.0. Numpy1.21.4, Pillow8.4.0, and scikit-image0.19.0 packages, etc. All of the packages above are used in data preparation and evaluation.

The configuration of the original SRCNN model and our D-SRCNN model is as the tables show:

Table 1. Configuration of SRCNN model

Filters Num	Filters Size	Activition Func	Bias	Learning Rate
128	9*9	ReLu	True	0.0001
64	1*1	ReLu	True	0.0001
1	5*5	None	True	0.00001

Table 2. Configuration of D-SRCNN model

Filters Num	Filters Size	Activition Func	Bias	Learning Rate
32	5*5	ReLu	True	0.0001
16	1*1	ReLu	True	0.0001
1	3*3	None	True	0.00001

3.4.Post Processing

The predicted greyscale image by D-SRCNN model should be converted back to a color image to check if the model works well. The predicted Y-channel drops useless dimensions generated by the model and keeps only 4 decimal places. Then the Y channel merges with the other two channels that do not change to YCbCr color space and converts back to RGB eventually.

4.EXPERIMENTS

4.1.Baseline Selection

The original SRCNN model chooses (i) sparse coding-based method of Yang *et al.* [17], (ii) neighbor embedding + locally linear embedding method [3], (iii) Anchored Neighbourhood Regression method [14], (iv) Adjusted Anchored Neighbourhood Regression method [15], and (v) *KK* method [4] as baselines.

Since the above methods have been compared with the original SRCNN model, we only use Nearest-Neighbor, Bilinear, and Bicubic interpolation accomplished by Pillow package and original SRCNN model as our baselines. Although some new accessing image quality method were introduced recently such as method from Sai [11], they still need time to prove themselves.

4.2.Model Training

Figure 3 shows the procedure of the model training, which includes data preparation and image prediction by D-SRCNN model.

In data preparation, which generally contains (i) colorspace converting and dataset expansion to get much more images to enlarge the training dataset, and (ii) Y-channel extraction. We decide to apply our SR algorithm, which is the D-SRCNN model, on only the Y-channel of the image dataset. The reason why we only keep Y-channel is that Dong has found that the results of applying the SR algorithm on all channels is even worse than applying only on the Y-channel, which suggests that Cb, Cr channels could decrease the performance of the Y channel when training is performed in a unified network [1].

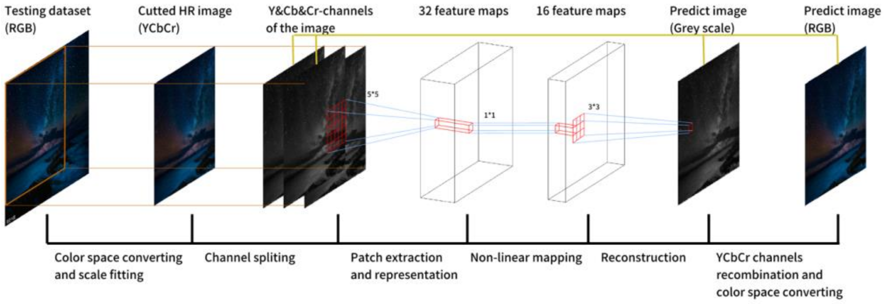

Fig. 4. Given low-resolution images (Testing dataset), only the quality of Y-channel (luminance)

Images will pass through the D-SRCNN model, which contains (i) patch extraction and representation, (ii) non-linear mapping, and (iii) reconstruction. The model will attempt to extract the features of the input image to form 32 feature maps first and then give the feature maps non-linearities to make sure the model can explore more possibilities of the predicted image, while a linear algorithm always gives certain results. After all the prediction of the model, it is usual that more than 1 feature map, or channel, are generated by the model, which exceeds the limit of a one-channel image. Therefore, one layer is needed to reconstruct all the predicted feature maps back to only one channel because the input image also contains only one channel.

MSE (mean square error), which is one of the most famous loss functions in CNN is chosen as the loss function of our model because it could obtain a better PSNR value in model training [1]. A deeper reason is that part of the algorithm of PSNR calculation between two images is similar to MSE. Therefore, training the SRCNN model for low MSE value will help increase the PSNR value that the model can obtain.

$$MSE = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}[L(i,j)-P(i,j)]^2}{n^2} \text{Eq. 1}$$

Where L is the label image and P is the predicted image and n is the size of the image (which is 30 in our model).

4.3.Model Evaluation

Figure 4 shows the procedure of predicting a HR image and therefore we can evaluate our model by comparing the predicted image with ground truth.

After training, two parameters include PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index Measure) are used to evaluate the D-SRCNN model.

PSNR is a widely used standard to evaluate the quality of compressed images. However, PSNR is sensitive to additive Gaussian noise while our human evaluation of image quality by eyes may be affected by many factors and additive Gaussian noise might not be one of them or might not be important in the whole evaluation process. The PSNR score may not be the same as the visual quality seen by the human eye.

$$PSNR = 20log_{10}\left(\frac{255}{\sqrt{MSE}}\right) \text{Eq. 2}$$

SSIM is a method for predicting the perceived quality of images, which is a measure of the similarity of two images. The detailed formula is as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \text{Eq. 3}$$

Where x and y are two images with common size n*n, μ is the average, σ is the variance of pixels of the image, and $c_1 = (0.01L)^2$ and $c_2 = (0.03L)^2$ and L is the dynamic range of the pixel values, which is 1 in our model.

4.4.Results

Our D-SRCNN model gets an average PSNR of 29.130452 and SSIM of 0.8727943 predicting test image with upscale factor range from 2 to 5 after about 60 hours of training on RTX 2070 SUPER.

Table 3. Experiment results

Eval. Mat	Upscale Method	Scale	Nearest Neighbor	Bilinear	Bicubic	D-SRCNN
PSNR	Nearest Neighbor	2x	34.431634	35.573283	36.764264	30.318132
		3x	30.731228	32.539525	32.932693	29.979629
		4x	28.497476	30.004583	29.919776	28.788100
		5x	27.409625	28.724878	28.386372	27.975231

	Bilinear	2x	35.667914	34.564575	36.602872	29.713915
		3x	31.350247	31.538790	32.320608	29.055740
		4x	29.879974	29.776197	30.268045	28.334027
		5x	29.011567	28.963404	29.276545	27.901576
	Bicubic	2x	36.967730	36.601380	39.783215	30.496861
		3x	31.514746	32.234132	33.126585	29.728496
		4x	30.067468	30.241746	30.791712	28.932901
		5x	29.118262	29.259075	29.598641	28.340817
SSIM	Nearest Neighbor	2x	0.974929	0.979676	0.983890	0.892606
		3x	0.948967	0.961285	0.964333	0.887550
		4x	0.921158	0.934129	0.934566	0.870335
		5x	0.902274	0.914613	0.911071	0.858037
	Bilinear	2x	0.979440	0.974544	0.983523	0.882863
		3x	0.949048	0.950208	0.957773	0.871630
		4x	0.926980	0.926100	0.932846	0.858357
		5x	0.910593	0.911890	0.916465	0.849946
	Bicubic	2x	0.984328	0.983528	0.990867	0.894470
		3x	0.951654	0.957135	0.964143	0.882460
		4x	0.929736	0.932578	0.939096	0.868417
		5x	0.911271	0.916147	0.920477	0.856860
Time	Nearest Neighbor	2x	0.003315	0.026187	0.033080	0.478399
		3x	0.003289	0.023241	0.029291	0. 469147
		4x	0.003455	0.023794	0.028639	0.439468
		5x	0.003200	0.020543	0.025424	0.449548
	Bilinear	2x	0.003454	0.026502	0.030914	0.495881
		3x	0.003270	0.024529	0.030092	0.446530
		4x	0.003228	0.022173	0.027865	0.488257
		5x	0.003455	0.019518	0.026742	0.424012
	Bicubic	2x	0.003377	0.026787	0.032262	0.449096
		3x	0.004169	0.022806	0.032230	0.445103
		4x	0.003186	0.020304	0.026683	0.441310
		5x	0.003375	0.020567	0.025663	0.454148

The experiment result shows that our D-SRCNN model well predicts dim images comparing with nearest neighbor, bilinear, and bicubic SR method and performs faster than an ordinary SRCNN model. The PSNR value is low when the input image has a relatively high resolution, which is expected because the number of features that can be captured by D-SRCNN model becomes significantly fewer when we reduce the size and the number of filters. However, as figure 5 indicates, D-SRCNN model gives a better potential to improve the image resolution more when the input image has a lower resolution than the other three SR methods.

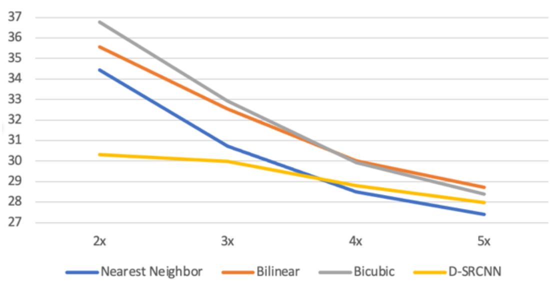

Fig. 5. Comparison between nearest

Neighbor, bilinear, bicubic, and D-SRCNN model when the test images are using the Nearest-Neighbor algorithm to downscale or upscale themselves. Y-axis shows PSNR value and x-axis represents upscaling factor.

5.CONCLUSION

Considering a night view image, or dim image often has much fewer features than photos taken during the day because it may contain a large number of pure black areas, this paper has proposed a Dim SRCNN (D-SRCNN) model since an ordinary SRCNN model may have superabundant filters for a dim image.

The approaches like nearest neighbor, bilinear, and bicubic interpolation could enlarge images effectively but have a lower quality compared with SRCNN method. The ordinary SRCNN method often predicts images effectively by providing a higher resolution image comparing with most of the other methods but with a much slower speed. Therefore, this paper introduces D-SRCNN model by significantly reducing the size and the number of filters of an ordinary SRCNN model to fulfill the gap between speed and image quality of dim images enlargement.

The experimental results show that D-SRCNN model performs well when the picture has a lower resolution, and at the same time, due to the greatly reduced filter size, the running speed has been significantly improved.

Due to some constraints, we have not been able to use large data sets such as ImageNet for training. In future work, we hope to expand the training set to observe the performance of the model under different scales of data.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307, 2016.

[2] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In European Conference on Computer Vision (ECCV), pages 391–407. Springer, 2016.

[3] Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neigh- bor embedding. In: IEEE Conference on Computer Vision and Pattern Recognition (2004)

[4] Denton, E., Zaremba, W., Bruna, J., LeCun, Y., Fergus, R.: Exploit- ing linear structure within convolutional networks for efficient evaluation. In: Advances in Neural Information Processing Sys-tems (2014)

[5] H. Y. Ren, M. El-Khamy, and J. Lee, “CT-SRCNN: Cascade trained and trimmed deep convolutional neural networks for image super resolution,” in Proc. IEEE Winter Conf. Appl. Comput. Vis., 2018, pp. 1423–1431.

[6] Jain, Purab Alok, and Pranali K. Kosamkar. "Analysis and Prediction of COVID-19 with Image Super-Resolution Using CNN and SRCNN-Based Approach." ICT with Intelligent Applications 1 (2021): 33.

[7] Kant M, Chaurasia S, Sharma H. Contribution Analysis of Scope of SRGAN in the Medical Field[M]//Data Engineering for Smart Systems. Springer, Singapore, 2022: 341-352.

[8] LeCun, Yann, and Yoshua Bengio. "Convolutional networks for images, speech, and time series." The handbook of brain theory and neural networks 3361.10 (1995): 1995.

[9] Minar, Matiur Rahman, and Jibon Naher. "Recent advances in deep learning: An overview." arXiv preprint arXiv:1807.08169 (2018). 10.1109/TPAMI.2015.2439281.

[10] Mohan A, Dwivedi R, Kumar B. Image Restoration of Landslide Photographs Using SRCNN[M]//Recent Trends in Electronics and Communication. Springer, Singapore, 2022: 1249-1259.

[11] Sai S V. A method for assessing photorealistic image quality with high resolution[J]. Computer Optics, 2022, 46(1): 121-129.

[12] Song L, Wang Q, Liu T, et al. Deep robust residual network for super-resolution of 2D fetal brain MRI[J]. Scientific Reports, 2022, 12(1): 1-8.

[13] TANG, Yan-qiu, et al. "A Survey of Image Super-Resolution Reconstruction." ACTA ELECTONICA SINICA 48.7 (2020): 1407.

[14] Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: IEEE In- ternational Conference on Computer Vision. pp. 1920–1927 (2013)

[15] Timofte, R., De Smet, V., Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: IEEE Asian Conference on Computer Vision (2014)

[16] Umehara K, Ota J, Ishida T. Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. J Digit Imaging 2018;31(4):441–450.

[17] Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Transactions on Image Processing 19(11), 2861–2873 (2010)