IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 10, Issue 3, June - July, 2022
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

An Autoencoder Neural Network (ANN)Approach for H-R images from LR images

K.V.K. Rao[1], CH. Srinivas[2], Manoj. N[3]
[1]Associate Professor, Narasaraopeta Engineering College, Narasaraopet.
[2]Research Scholar & Associate Professor, Narasaraopeta Engineering College, Narasaraopet.
[3]Assistant Professor, Narasaraopeta Engineering College, Narasaraopet.

Abstract

In this paper, a faster and deeper convolutional neural network with Auto encoder is used to achieve high test accuracy and low test error high-resolution images from low resolution images. In the proposal artificial neural network (ANN) is modeled to obtain the high resolution image. To achieve this, a multilayer neural network with a single fully-connected hidden layer, a linear activation function and a squared error cost function trains weights that span the same subspace as the one spanned by the principal component loading vectors. In addition, regularizations in the joint training scheme is crucial in achieving good performance, by improving the quality or clarity of images for human viewing by removing blurring and noise and increasing contrast, and brightness of the image. In various applications, where an image is to be reconstructed, from its degraded version, the performance of the fast non-iterative algorithms needs to be evaluated quantitatively. The Artificial neural network model is used to develop the image excellence or appearance for human perception and intervention and also reduces the computational time. Finally, it is empirically evaluated the performance of this joint training scheme and observed that it's not only learns a better data model, but also learns better higher layer representations, which highlights its potential for unsupervised feature learning.

Keywords: Low Resolution, High Resolution, Convolutional neural networks, Autoencoder.

I. INTRODUCTION

The highly challenging task of estimating a high-resolution (HR) image from its single low-resolution (LR) counterpart is referred to as super-resolution (SR). A unsupervised learning technique called Autoencoders in which neural networks for the task of cognitive content learning will be used. Specifically, a neural network architecture which can impose a hinder in the network that forces enhancement knowledge representation of the original input. This network can be trained by minimizing the reconstruction error while producing high resolution image, which measures the difference between our original input and the resultant reconstruction image. To obtain high contrast in the image, intensity equalization method has to be done and high resolution images are essential for the applications like intensive investigation, intelligent tutoring system and health-care. Usually different Histogram Equalization methods are used to have better resolution of the images like Interpolation and Histogram equalization (HE), adaptive histogram equalization (AHE), Contrast Limited AHE (CLAHE) etc.

A. Interpolation

Interpolation algorithms are two types i.e adaptive and non-adaptive. The adaptive methods depend on what they are interpolating, whereas non-adaptive methods treat the pixels equally. Adaptive algorithms are used in many proprietary techniques in specialized professional image editing software like Photo zoom Pro and Adobe Photoshop. Non-adaptive algorithms include, nearest neighbor, bilinear, bicubic and spline. In general bicubic interpolation can be accomplished using Lagrange polynomials. The choice of good interpolation method is based on finding an optimal counterbalance between three unsuitable artifacts: edge halos, blurring and aliasing. B-spline or Basis-spline function $B_{i,n}(x)$ is used to full fill all the requirements of a interpolation. A spline of order n is a piecewise polynomial function of degree n-1 in a variable x. The values of x where the pieces of polynomial meet are known as knots denoted …, t_0, t_1, t_2… and sorted into non-decreasing order. When the knots are distinct, the first n-2 derivatives of the polynomial pieces are continuous across each knot. When r knots are coincident, then only the first n-r-1 derivatives of the spline are continuous across that knot.
For a given sequence of knots the B-spline is

$$B_{i,n} = \begin{cases} nonzero\ for & x < t_i \quad x \geq t_{i+n} \\ 0 & otherwise \end{cases}$$

… (1)

The higher order B-splines are defined by recursion

$$B_{i,k+1} = \frac{x - t_i}{t_{k+i} - t_i} B_{i,k} + \frac{t_{i+k+1} - x}{t_{k+k+1} - t_{i+1}} B_{i+1,k}(x)$$

… (2)

The 0th order B-spline (B_0) function is used to represent the nearest neighbor interpolation which covers two numbers of pixels i.e. inter-pixel distance is 1. Linear interpolation is described by, B_1 which covers the three number of pixels -1, 0, +1.Cubic B-spline interpolation function, B_2 covers five number of pixels -2,-1, 0, +1, +2 and is more preferable when image does not have high frequency components. B_3 is defined for two regions (0, 1) and (1, 2) as shown in “(1),”and “(2),”respectively.

B. Histogram equalization

Histogram equalization (HE) is applicable when the distribution of pixel values is resembles throughout the image A.Charles Stud, N.Ramamurthy 2019; [4]. This method usually increases the global contrast of images when its usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast, and it maximizes the entropy of an image. Hence, without changing its median value the dynamic range of the histogram is increased. This can be achieved with the help of the cumulative distribution function (CDF).
In Adaptive Histogram Equalization (AHE) differs from ordinary histogram equalization in the respect that the adaptive method computes several histograms, each corresponding to a distinct section of the image, and uses them to redistribute the lightness values of the image. It is therefore suitable for improving the local contrast and enhancing the definitions of edges in each region of an image.
Contrast Limited AHE (CLAHE) differs from adaptive histogram equalization. In this the contrast limiting procedure is applied to each neighborhood from which a transformation function is derived. CLAHE was developed to prevent the over amplification of noise that adaptive histogram equalization can give rise to.

II. PROPOSED METHOD

More powerful approaches aim to establish a complex mapping between low and high-resolution image information and usually rely on training data. Many methods that are based on example-pairs rely on LR training patches for which the corresponding HR counterparts are known. Prediction-based methods were among the first methods to tackle single image super resolution (SISR). While these filtering approaches, e.g. linear, bicubic interpolation [8] filtering can be very fast, they oversimplify the SISR problem and usually yield solutions with overly smooth textures. Methods that put particularly focus on edge-preservation have been proposed. More recently, inspired by the great success achieved by deep learning in other computer vision tasks, people begin to use neural networks with deep architecture for image SR. Multiple layers of collaborative Autoencoders are stacked together in for robust matching of self-similar patches. Artificial neural networks are designed that directly learn the non-linear mapping from LR space to HR space. As these deep networks allow end-to-end training of all the model components between LR input and HR output, significant improvements have been observed over their shadow counterparts.

1. Autoencoder Representation

Autoencoder is a data compression algorithm where there are two major parts, encoder, and decoder shown in Fig.1. The encoder's job is to compress the input data to lower dimensional features. For example, one sample of the 28x28 Modified National Institute of Standards and Technology (MNIST) image has 784 pixels in total, the encoder can compress it to an array with only ten floating point numbers also known as the features of an image. The decoder part, on the other hand, takes the compressed features as input and reconstruct an image as close to the original image as possible. Autoencoder is unsupervised learning algorithm in nature since during training it takes only the images themselves and not need labels.

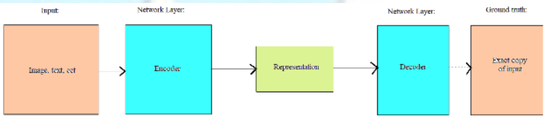

Fig. 1: Autoencoder Representation

2. Encoder Architecture Design

The general idea behind this formulation is that it allows one to train an encoder model with the goal of fooling a distinguishable decoder that is trained to distinguish super-resolved images from real images. With this approach our encoder to generate compressed image from the input image that encoded image is given to the decoder refers to S.Ioffe and C. Szegedy 2015.,[7]. This encourages

perceptually superior solutions residing in the subspace and manifold of natural images. This is in contrast to SR solutions obtained by minimizing pixel-wise error measurements, such as the mean square error.
Let us consider a single training example presented by (x, y) where x is

$$x \in \mathbb{R}^{n_x} \ and \ y \in \mathbb{R}^{n_x} \quad \dots (3)$$

x contains m training examples i.e.

$$((x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ ....(x^{(m)}, y^{(m)})) \quad \dots (4)$$

$$\widehat{y}^{(i)} \approx y^{(i)} \quad \dots (5)$$

ReLU activation function as shown in Fig.2

$$\sigma(x) = x^{+} = max(0, x) \quad \dots (6)$$

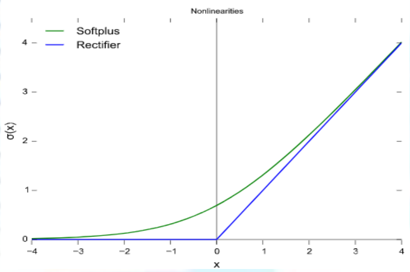

Fig. 2: ReLU activation function

At the core of very deep encoder network, this is illustrated in Figure 3. Inspired by Johnson et al. [1] employ the block layout proposed by Gross and Wilber [2]. Specifically, here two convolutional layers are used with small 3×3 kernels and 64 feature maps followed by batch-normalization layers [7] and Parametric ReLU [9] as the activation function. The resolution of the input image is increased with two trained sub-pixel convolution layers as proposed by Shi et al. [10].

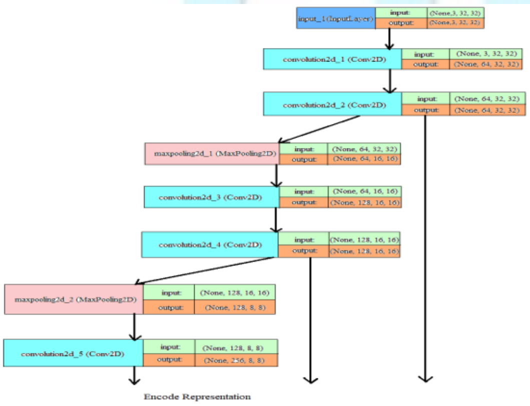

Fig. 3: Encoder Architecture Design

The output of an encoder is compressed image of an input image. This is can accomplished the architecture consist of convolution blocks followed by max-pooling in order to reduce the size of an image. Likewise the image quality can be increased with the help of convolutional block and max-pooling blocks.

3. Decoder Architecture Design

To know apart real HR images from generated compressed image from the encoder, train a decoder network. The architecture is shown in Figure 4. It follows the architectural guidelines summarized by Radford et al. [11] and use ReLU activation and use max-pooling throughout the network. The decoder network is trained to solve the maximization problem. It contains eight convolutional layers with an increasing number of 3 × 3 filter kernels, increasing by a factor of 2 from 64 to 256 kernels as in the VGG network [12]. Strided convolutions are used to reduce the image resolution each time the number of features is doubled. The resulting 256 feature maps are followed by two dense layers and a final ReLU activation function to obtain a probability for sample classification.

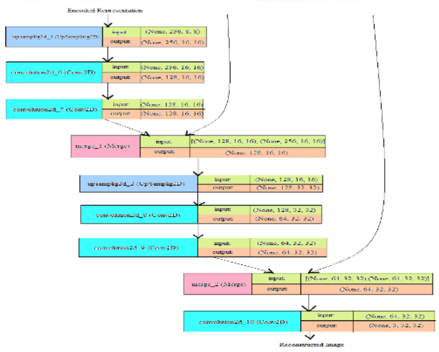

Fig. 4: Decoder Architecture Design.

IV RESULTS AND DISCUSSION

1. Benchmark learning architecture

The learning architecture is designed following the model described in the Keras and Tensor flow as a back-end its reference to arXiv: 1610.02357v3 [7]. The proposed learning architecture is illustrated in Table 1.

Table 1: Descriptions of the learning architecture

S. No	Layer	Input size	Output size
1	Conv_1	(3, 32, 32)	(64, 32, 32)
2	Conv_2	(64, 32, 32)	(64, 32, 32)
3	Max_Pool1	(64, 32, 32)	(64, 16, 16)
4	Conv_3	(64, 16, 16)	(128, 16, 16)
5	Conv_4	(128, 16, 16)	(128, 16, 16)
6	Max_Pool2	(128, 16, 16)	(128, 8, 8)
7	Conv_5	(128, 8, 8)	(256, 8, 8)
8	UpSamp_1	(256, 8, 8)	(256, 16, 16)
9	Conv_6	(256, 16, 16)	(128, 16, 16)
10	Conv_7	(128, 16, 16)	(128, 16, 16)
11	UpSamp_2	(128, 16, 16)	(128, 32, 32)
12	Conv_8	(128, 32, 32)	(64, 32, 32)
13	Conv_9	(64, 32, 32)	(64, 32, 32)
14	Conv_10	(64, 32, 32)	(3, 32, 32)

This model follows the common multi-layer artificial neural network (ANN) architecture which consists of alternating convolutions and nonlinearities. This architecture takes the full-size image as the input, and the output is identical size as the input. After the nonlinear transformation by Rectified Linear Units (ReLU), the feature maps are response normalized and Up-sampled by Local Response Normalization (LRN) and the Max-Pooling respectively [5] to generate reconstructed input.

2. Performance comparisons

The performance of models trained with ANN is plausibly better than the original ones. The below Fig. 5 Represents the proposed model has best accuracy and lowest cost function when compared to original one.

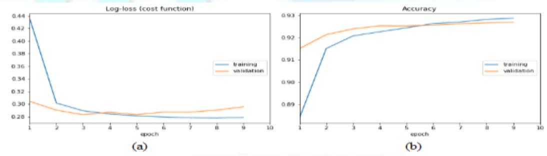

Fig.5: Performance comparison. (a) Performance comparison of trained and tested images in terms of Cost function. (b) Performance Comparison of trained and tested images in terms of Accuracy.

3. Results comparison

In order to evaluate the effectiveness of the proposal, tests were performed on different data sets of car images which were taken from a reference database.

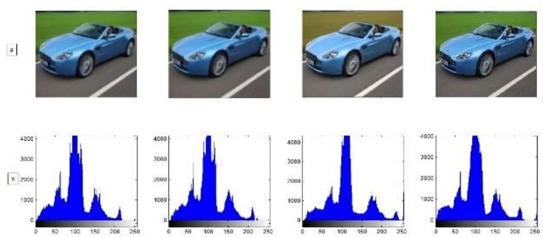

Fig.5: Images and their histogram representations.

ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

(a)Input image, Interpolated image, encoded image and Autoencoded image respectively (from left to right) & (b) Histogram representation of the image in (a).

The results shown in the Fig. 5 are for the data set-1 and these results are compared with bicubic interpolation. For the qualitative analysis of the proposal the histogram representations are given for all the images and are shown in fig 5. But it is not enough to assess the performance qualitatively and it is also required to compare quantitatively with different performance image metrics like peak signal to noise ratio (PSNR), entropy and Loss function.

PSNR:

It is a feature bringing similarity to human perception of reconstruction quality. This ratio is a qualitative peak error measurement between the original image maximum power and the power of corrupted image which is measured in decibels. The higher the value of PSNR better will be the quality of the reconstructed image.

$$PSNR = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \quad \ldots (7)$$

Where, MAX $_f$ is the maximum possible power of an image and MSE is the mean square error which represents the power of corrupting noise.

Entropy:

Another important parameter which represents the quantitative treatment for recognizing the details contained in an image.

$$E = -\sum_{i=1}^{L} m_i \log_2 m_i) \quad \ldots (8)$$

Where *i* represent the greylevel, m is the probable existence of greylevel of *i* and L is the total number of greylevels.

Loss Function:

The loss can be find with actual image (y) and the resulting image (yhat) and is given by the function

$$L(\widehat{y}, y) = \frac{1}{2}(\widehat{y} - y)^2$$
$$L(\widehat{y}, y) = -(y \log \widehat{y} + (1-y)\log(1-y)) \quad \ldots (9)$$

Where y is actual image, yhat is a resulting image. Now all the performance metrics are evaluated for the proposed method and conventional method and are tabulated in table 1.From the table 1 it is observed that the quiet level of values are better when compared with the proposal.

Table 1: Comparison of Performance metrics for the two approaches.

Method	PSNR Value	Entropy Value	Loss Value
Interpolation	21.44	7.067	1.06
Autoencoder (Proposed)	25.67	7.159	0.13

Finally to estimate the performance of the proposed method, different data sets are taken for which high resolution images are obtained as shown in fig.6.

Fig.6: Input image, bicubic interpolation image, encoded image and high resolution image (from left to right).

V. CONCLUSION

From the above results it described that an artificial neural network with Autoencoder sets a new state of the art on public touchstone datasets when evaluated with the widely used PSNR measure. In order to get high test accuracy and less cost function, introduced Artificial Neural Network, which augments the content loss function with an adversarial loss by training a

ANN. ANN reconstructions for large up scaling factors (4×) are, by a considerable margin, more high resolution images than reconstructions obtained with the proposed method.

REFERENCES

1. Johnson, A. Alahi, and F. Li. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision (ECCV), pages 694–711. Springer, 2016.

2. Gross and M. Wilber. Training and investigating residual nets, online at http://torch.ch/blog/2016/02/04/resnets. html. 2016.

3. Christian Ledig, Lucas Theis and Ferenc Huszar.Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,arXiv:1609.04802v5 [cs.CV] 25 May 2017.

4. Charles Stud, N.Ramamurthy. Interpolation of the Histogramed MR Brain Images for Resolution Enhancement,ISSN: 2278-3075, Volume-8 Issue-11, September 2019.

5. Y. Yang, C. Ma, and M.-H. Yang. Single-image super-resolution: A benchmark. In European Conference on Computer Vision (ECCV), pages 372–386. Springer, 2014.

6. Franc¸ois Chollet Google, Inc.Deep Learning with Depthwise Separable Convolutions, arXiv:1610.02357v3 [cs.CV] 4 Apr 2017.

7. S.Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of The 32nd International Conference on Machine Learning (ICML), pages 448–456, 2015.

8. E. Duchon. Lanczos Filtering in One and Two Dimensions. In Journal of Applied Meteorology, volume 18, pages 1016–1022. 1979.

9. K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In IEEE International Conference on Computer Vision (ICCV), pages 1026–1034, 2015.

10. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1874–1883, 2016.

11. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1874–1883, 2016.

12. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (ICLR), 2015.