IMAGE SUPER RESOLUTION VIA EXTERNAL LEARNING BASED TECHNIQUES: A REVIEW

[1]RUAA A. ALFALLUJI, [2]ZAINAB DALAF KATHEETH

[1]University of Babylon, Iraq
[2]Faculty of Computer Science and Mathematic, Kufa University, Iraq
E-mail: [1]fine.ruaa.adeeb@uobabylon.edu.iq, [2]Zainab.alfarawn@uokufa.edu.iq

Abstract - High-quality images have a significant and essential role in many applications such as remote sensing, aerial imaging, military fields, medical diagnosis, object tracking, video surveillance, and criminal justice. However, high-resolution imaging may not always be feasible due to limitations of the sensors and optics manufacturing technology, and it is proven to be very costly. Super image resolution is an approach used to generate a higher resolution image from lower resolution image(s) by employing image processing algorithms that are relatively inexpensive. In this paper, a survey on the single image super-resolution methods which are based on external database learning is provided.

Keywords - Super Resolution, External-based Learning, Deep Learning, Sparse Coding, Anchored Regression, Regression Trees.

I. INTRODUCTION

High resolution images are very useful in a number of real world problems such as medical imaging for diagnoses, surveillance, satellite imaging processing and forensic. However, in many practical situations, the acquired images are often of low resolution (LR) because of the limitations of the optical system or other factors which include, but are not limited to, transmitting high resolution image over a specified bandwidth which will affect many details from the high resolution image and resulted in a low resolution one. Thus in turn will limit the subsequent tasks based on the high resolution images [1]. The resolution and the quality of the captured images can be enhanced by overcoming the hardware limitations, especially with the recent revolution in the electronic and electric industries. However, this usually requires a very high cost due to the sophistication of the used hardware components. Therefore, many techniques and algorithms were developed to improve the low resolution images. This solution is more preferable than the first solution.

Given a single low resolution image or a batch of low resolution images and mapping them into high resolution image(s) without losing any textural or context details is known as "Super Resolution". Enhancing the resolution based on single image only is called "single image super resolution". While enhancing the resolution based on multiple images is known as "multiple or multi-frame images super resolution"[2]. Multiple or multi-frame images super resolution works on the hypothesis that there are a number of low resolution images available for the same scene. These techniques give the best results when the low resolution images are taken from a slightly different perspective i.e. when the low resolution images are marginally different from each other. It is assumed that each low resolution image provides unique information about the high resolution one, that needs to be estimated, and combining them in a good and professional manner will yield a pleasant looking and detailed sample of the low resolution image. However, in the practical applications, acquiring enough images for the same scene with diffirent information is sometimes difficult as well as impractical. Therefore, many attention is given to the single image super resolution techniques recently. In this paper, the degradation model which is implemented as a preparation step in super resolution models to weaken the high resolution images will be explained in section 2, while section 3 sheds the light ,especially, on the taxonomy of single image super resolution algorithms due to their importance. Section 4 focuses specifically on the external based learning algorithms with a number of recent studies. Section 5 shows the advantages, disadvantages and performance evaluation of discussed algorithms and finally section 6 presents the conclusion of the study.

II. DEGRADATION MODEL

Super resolution can be considered as one of the inverse problems. While a forward problem starts with the cause and computes its corresponding result, an inverse problem, on contrast, starts with the result and then calculates its corresponding cause. Basing on this assumption, all high resolution images are degraded before being used in super resolution models i.e images lose their actual quality[3].Most probably, the degradation process takes place during the capturing of image, due to motion blur, camera defocus blur, sensor noise and so on. Figure 1 shows the degradation model of a high resolution image. On the contrast, the process of image restoration is used to regain the actual quality of image by removing the effect of degradation, but the size of the image remains same [4].

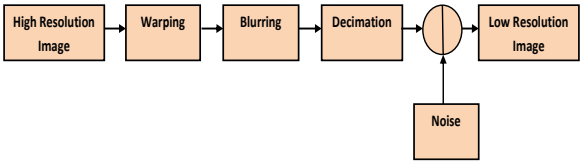


Figure 1: Degradation Process

III. SINGLE IMAGE SUPER RESOLUTION TAXONOMY

These methods can be divided into three main categories which are: “reconstruction based methods”, “interpolation based methods”, and “learning based methods” as illustrated in Figure 2 [2].

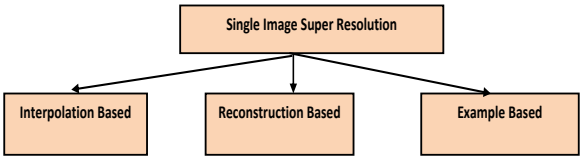


Figure 2: Taxonomy of Single Image Super Resolution Methods

Interpolation Based Super Resolution

Each interpolation model works by calculating the missing values in the output image corresponding to the input image. The interpolated values are determined by calculating the weighted average of the pixels that are in the area around the pixel being calculated. Nearest neighbor interpolation, bilinear interpolation, bicubic interpolation are examples on this type of super resolution techniques. These methods have the advantages of simple calculation and the minimum computation complexity. However, their reconstructed resultant images are not good enough especially in the edge areas.

Reconstruction Based Super Resolution

The main goal in these methods is imposing a linear constraint on reconstructed HR image, which was observed by low resolution images. Reconstruction based methods mainly contain :iterative back projection, projection onto convex sets, maximum a posteriori. These techniques provides better performance than the interpolation methods. However, reconstruction based methods may produce many ringing artifacts in the high resolution image.

Learning(or Example) Based Super Resolution

These methods determine the high resolution version of a low resolution single image by exploiting examples and machine learning models. In another meaning, they study the mapping of the low resolution image to the high resolution one and construct the high resolution image according this function. Learning based super resolution algorithms can be further divided into parametric and nonparametric methods [5] as shown in Figure 3 .

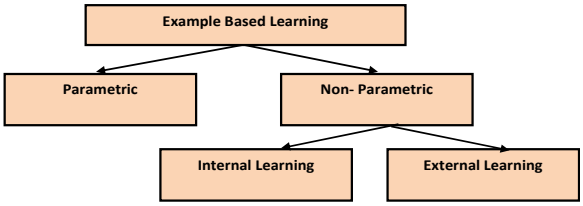


Figure 3: Classification of Example Based Super Resolution Algorithms

Parametric methods solve the super resolution problem by using mapping functions which are governed by parameters that are small in number. The mentioned parameters are calculated from the example images that may or may not come from the input image. Parametric models are more powerful and efficient than nonparametric models in terms of efficiency and amount of data required for model estimation.

In nonparametric methods, the model is not specified a priori as in parametric models. The training data is used for construction of the model. It does not mean that nonparametric models are not dependent on any parameters but the parameters are flexible in their nature and number and are dependent on the training data. Nonparametric methods work by splitting the input images into patches that are overlapping each other and the final output is the combination of the computed patches.

All of the non parametric methods can work without making a large number of assumptions. This property as well as the availability of huge amount of training examples and the equipments for offline training such as GPU and cloud, nonparametric methods are one of the best options available for high resolution image computation.

Recent studies on nonparametric methods achieved much better performances in efficiency and accuracy, based on the advances made in machine learning approaches. Therefore, nonparametric techniques will be highlighted in this paper. A number of these techniques allow reconstruction of high resolution output from just a single visual input [6,7]. While others enable using external datasets for parameters calculation,the used datasets are not necessarily related to the input image [8,9]. This difference is the basis of classification of nonparametric techniques into the internal and external learning methods as shown in Figure 3 [5].

The main idea behind methods in these categories can be extracted by the names. In internal learning category, the input image is directly used for the extraction of examples patches, the cross-scale self-similarity property of natural images is used (i.e., examples are internal, extracted from the input image) [7,10]. These techniques are based on the assumption that small patches in a natural image(e.g., 3 × 3 pixels) are also available in the low resolution version of the same image.Neighbour embedding and high frequency transfer are examples on internal learning techniques.

Despite the major improvements in these methods, which include robustness to noise, the implicit adaptivity to the image contents, better texture preservation and sharper edges, many limitations still appear during implementing these techniques. The required nearest neighbor search and iteratively application in small scaling factors to ensure the self-similarity can cause a high computational cost .

In the second category of nonparametric methods, the external learning, the examples that are used for the estimation of high resolution image are external (i.e extracted from an external database)[10].

The external based learning techniques provide a significant improvement in terms of generalization than the internal based learning techniques. However, there are techniques that provide a much better balance between generalization and computational efficiency.

IV. EXTERNAL BASED LEARNING ALGORITHMS

Recently, external based learning concept is widely employed in the field of super resolution due to its effectiveness. There are four external based learning techniques as illustrated in Figure 4. In this section, these techniques is explained and the light is shed on a number of recent studies using these techniques.

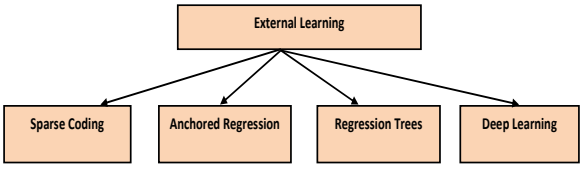


Figure 4: External Based Learning Algorithms

Sparse Coding Algorithm

The main concept behind sparse signal representation, or sparse coding, is that low dimensional projections of the signals can be used to reconstruct linear relationships between the corresponding signals precisely [11]. Since the first implementation of sparse signal representation in super resolution field by [12], it has been an active research topic. The sparse coding algorithm in training phase jointly learns HR dictionary (Dh) and LR dictionary(Dl) to obtain HR and LR patches by supposing that every pair of HR/LR patch shares the same sparse coding vector. The first step in the testing phase is the splitting of input image into overlapping patches. Each patch is encoded using the LR dictionary (Dl) with the sparse coefficient. Dh and Dl are used to reconstruct the related HR patch. Finally, high resolution image can be obtained by combining all the reconstructed patches.

Convolution Sparse Coding [13]

Three groups of parameters (low resolution filters, mapping function and high resolution filters) are computed in the training phase. To calculate sparse features, low resolution images are divided into two components “smooth” and “residual”. The smooth component is enlarged using bicubic interpolation and convolutional sparse coding is applied on residual component which denotes high frequency edge and texture structure in low resolution image. The smooth component in the LR image is extracted using the following equation:

$$y = f^s \otimes Z_y^s + Y \quad (1)$$

Where ($f^s \otimes Z_y^s$) is the smooth component in the LR image (f^s represents a s X s filter and Z_y^s represents its corresponding feature map) and Y is the residual component. A number of low resolution filters are utilized to transform Y into feature maps using an alternating direction method of multipliers (ADMM)[14]. Similar to LR image, each HR image is firstly divided into one smooth and one residual component.

$$x = f^s \otimes Z_x^s + X \quad (2)$$

Where($f^s \otimes Z_x^s$) is the smooth component in the HR image (f^s represents a s x s filter and Z_x^s represent its corresponding feature map) and X is the residual component .

To calculate the low frequency maps, enlarged smooth component are used. By using the low resolution feature maps and high resolution images, high resolution filters and related mapping function are obtained. Convolutional sparse coding is implemented on the small low resolution image instead of enlarged image which reduces the number of needed low resolution filters as well as execution time while high resolution reconstruction uses a large number of filters because of the complexity of high resolution image. The high resolution filters and mapping function are calculated according to the following equation which is solved by stochastic average and alternating direction method of multipliers (SA-ADMM) algorithm [15].

$$\{W\} = \arg\min_{W} \| X - \sum_{j=1}^{M} f_j^h \otimes g(Z_j^l; w_j) \|_F^2, s.t. w_j \geq 0, | w_j |_1 = 1 \quad (3)$$

Where M represents the number of HR filters, f^h is a HR filter, w is a linear transformation vector and Z is image patch. The LR image is decomposed in the testing phase using the learned LR filters and its sparse feature maps are obtained. With the help of learned mapping function, HR features maps are estimated from the LR feature maps. A simple convolution operation can be employed to obtain the HR image. The resultant image can be improved by adding the high frequency texture structure through summation the convolutions of HR feature maps and the corresponding HR filters. Back propagation algorithm

is employed for more improvements to the HR estimation [12, 16,17].

Sparse Representation on a K-NN Dictionary [18]

The first and most important part of any sparse based system is dictionary making[19, 20, 21, 22]. Several high resolution images are used from the internet for making the dictionary. The low resolution counterparts are created by downsampling and blurring the high resolution images. Bicubic interpolation was employed to make low resolution and high resolution images size equivalent to each other. Only patches of low resolution images that are centered at edges were selected to get texture rich patch. The size of the created patches is 19X19. Binary encoding based on Restricted Boltzmann Machine (RBM) was implemented after dictionary preparation. The energy function of RBM was calculated as below [23].

RBM,which is a connect bipartite graph having one input and hidden node, uses contrastive divergence (CD) optimization method for minimizing its energy function [24]. In the hidden layer, there are a number of features and unit layers. All the layers have their own weights. The activation of a unit is computed using sigmoid function. Ning L et al. [18] suggested using four layers; one input layer, one output layer and two hidden layers. Features from low resolution patches were used at the input layer and resulted in a binary code of length 64 as an output. A number of RBMs are used for learning binary codes. For the training of every RBM, the previous RBMs data are employed which enables finding high order correlation between layers. The authors in [18] used KNN as a dictionary for each low resolution patch. The nearest neighbor search is implemented to find low resolution patches nearest neighbor in Dl after performing binary encoding for each of the LR patches and LR dictionary (Dl). The hamming distance is used as a distance metric, which creates sub dictionaries. The need for dictionary learning is eliminated because of the similarity between dictionary atoms and the patches. For every LR patch in the LR image, its LR dictionary (Dl) is obtained by using KNN and the corresponding HR dictionary (Dh). The patches are represented by the linear combination in dictionaries. For patches compatibility, the authors used a method which was proposed in [25]. After finding the optimal solution, the patches are reconstructed and global reconstruction constraint is employed on the final image to get super resolution image.

Anchored Regression

In the most recent studies related to anchored regression in super resolution, a certain mapping function is learned from the manifold of LR patches(or features) to HR patches following the manifold assumption which is already used in neighbor embedding [26]. It is assumed that the used mapping function is locally linear. As a result, training examples are used to learn a number of linear regression functions and anchored to the manifold as a piecewise linearization.

Anchored Neighborhood Regression [8]

The proposed framework in [8] consisted of two parts; global regression and anchored neighborhood regression. Most of sparse coding and neighbor embedding techniques use l-norm regularization which is very computationally expensive. Radu T. et al in [8] proposed using ridge regression (also known as collaborative regression) to handle this problem. The high resolution patches are then reconstructed using the obtained coefficients from the regression. When the whole dictionary is used instead of a neighborhood, the computation not only becomes global but also parts of it such as projection matrix can be precomputed which in turn reduces the execution time. However, this is only an extreme case of anchored neighborhood regression method. High resolution images in global regression are obtained by multiplication of projection matrix with that of low resolution features. However, this is not suitable for all kind of low resolution features. Therefore, instead of using whole dictionary, local neighbors of given sizes were used as a starting point for computing the projective matrix. Firstly, the dictionary instances are divided into neighborhoods (i.e K nearest neighbors for every atom in the dictionary are calculated which represent the neighborhood of that atom). Correlation is employed as a computational measure instead of euclidean distance as suggested in [9,27]. It is more practical because the vectors are l-normalized rather than being taken directly from the low resolution image. A projection matrix is calculated separately for each neighborhood and used to compute the high resolution atoms.

Anchored Neighborhood Regression with Multiple Class-specific Dictionaries [28]

Abedi A. et al. [28] proposed a model which was totally focused on super resolution of text images. The proposed model is divided into reconstruction phase and offline learning phase. During the learning phase, distinct dictionaries, whose number is denoted by C, are separately learned from each image patch where each patch contains a different letter from each class C. For dictionary learning, a method suggested by Zeyde et al. [27] was used and C was set to 62 which includes upper, lower case letters and digits from 0 to 9. Each unique character dictionary contains v length feature vectors. The feature vectors were calculated by extracting four horizontal and vertical gradients followed by laplacian filter and principal component analysis (PCA). The projection matrix for every atom in each dictionary is calculated using Timofte R. et al. algorithm [29]. It is worthy mentioning that the offline computation and learning will reduce the processing time of the proposed model. The offline calculated

dictionaries and projection matrices are employed in the reconstruction phase. Each extracted patch resolution is improved by using the top predicted class dictionaries. For every low resolution patch, the related nearest match is found in each class dictionary atom. The found atoms are employed for the construction of high resolution patch using a method proposed by [30, 31] through multiplication of low resolution patch with the found atoms projection matrices. The described weights by [26] are computed and combined with patch multiplication output to get the final HR patches. The final HR image is obtained through combining all the computed HR patches.

Regression Trees

External learning based super resolution algorithms suffer from many subproblems. For every patch, the selection of a suitable locally linear mapping function is significant problem in the inference stage. While in the training stage, unsupervised training presents the biggest hurdle. Hierarchical nature of regression trees enables them to solve both of the afore mentioned problems. The usage of regression trees in super resolution field will be discussed.

Super-Resolution using Random Forests(RF)[32]

Learning based super resolution algorithms basically work by finding a mapping function between the low resolution image and high resolution one. These algorithms use slightly different versions for the couple's dictionary learning. However, many of these techniques have the drawbacks of being slow and require sparse encoding. Criminisi A et al. [33] suggested using random forests in the fields of computer vision and image analysis. Random forests , are groups of binary trees, enable parallelization and thus reduces processing time. All trees of random forests are trained independently of each other using 'N' number of training samples. Schulter S. et al [32] employed random forests in their super resolution framework. A single tree works through separating the training data into subsets using splitting functions.

The node where the processing starts is known as "root" node and where the process ends is known as "leaf" node. Splitting begins from the root node and continues until a leaf node is reached (i.e. no more splitting is possible or the predefined tree depth is reached). Using more than one tree results in a subset of overlapping cells.The separation is then used to find the data dependences while a linear model can be employed to represent every leaf node. The final data dependent mapping matrix is computed by using the average of trees generated.

Naive Bayes Super Resolution Forest [34]

The most computationally expensive step in all super resolution techniques is the calculation of the mapping functions using local linearization. Salvador J. et al [34] proposed an algorithm which is built on basic idea presented in [35]. The aim of the algorithm is providing a direct mapping function which transforms coarse patches into HR patches. Firstly, the input space is divided into clusters .A local linear mapping function is calculated. The mapping function, in essence, is a correction layer which uses iterative back projection. In the proposed algorithm, features are adaptively computed so that local linearizations are obtained. It was further observed that original and scaled versions of patches, although having contrast, have same structure. All the patches with same structure were grouped together. For data partitioning, firstly unimodal partition tree approach was used but it produced unbalanced results. Therefore, bimodal partitioning tree was combined with absolute value of cosine similarity(AVCS). During training stage, partition tree is created in such a way that at each node partitioning criteria is able to differentiate between the relevant antipodal data. The given AVCS metric is differentiated based on K means clustering algorithm. Local linearization based regression matrix is also calculated during the training stage.

Recently, tree structure have been used widely to solve a number of computer vision problems including super resolution [32, 36, 37]. Bernard S et al. [38] illustrated that the combination of trees is not always the best option and may not provided better results. Therefore, for increasing efficiency, a single tree is selected. Data are modeled using Von Mises-Fisher distribution before applying tree [39]. The most selective tree is local Navie bayes that proposed by McCain and Lowe[40] was used. Local Navie bayes is approximated using likelihood partition in each node.

Deep Learning

The main advantage behind using convolutional networks in computer vision is the ability of extacting the properties of stationarity and locality in natural images using a small number of parameters. LeCun Y. et al [41] introduced convolutional networks technique, which greatly improved the generalization by employing a number of task-domain known properties in the network's architecture. Convolution networks also have the advantage of being able to accept different size inputs as compared to networks that are fully connected which have fixed size outputs and inputs, defined by the architecture of the network. Many recent studies made use of the conventional networks and deep conventional networks in image enhancement applications especially image super resolution. In this section, a number of those studies are highlighted.

Sub-Pixel Convolutional Neural Network [42]

A LR image was obtained from HR trained one by using Gaussian filter followed by downsampling. The LR image was directly passed to three layers convolutional neural network and upscaling is done by sub pixel convolutional layer instead of using upscaled

version of LR image as done in many previous studies[43]. The proposed technique is suitable for video super resolution because of the reduction of the filter size which minimizes the processing time. The features are extracted by nonlinear convolution of LR image. A deconvolution layer is added to the network to recover the resolution of an image [44, 45, 46, 47]. High level features can be used for calculating semantic segmentation and visualization of activated layers. Upscaling filters are learned for each feature map instead of learning one upscaling function for the LR input image. The network learns the processing needed for super resolution implicitly without using an interpolation filter. The design of the network in that way learns better mapping from LR to HR compared to a single fixed upscaling filter.

Generative Adversarial Network [48]

Generative adversarial network(GAN) provides images with a high quality. Ledig C. et al [48] used GANs concept with deep ResNet [49,50] for constructing super resolution image. The LR version is obtained during training phase through applying Gaussian filter and downsampling on the HR image. A feed forward convolutional neural network based on generator network was employed. The biases and weights are calculated using an optimized loss function which is an improved version to the function proposed by Johnson and Bruna. The utilized loss function is a sum of adversarial loss and content loss.

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \quad (4)$$

In content loss(l_X^{SR}), a closer loss function is calculated, while adversarial loss ($10^{-3} l_{Gen}^{SR}$) is basically the generative component of GAN network which looks for a solution that is manifold of natural images. The loss is based on overall training sample probabilities. Goodfellow[51] defined discriminator network that was used along with feed forward convolutional network to solve adversarial max -min problem.

A block configuration suggested by Gross and Wilber, two convolutional layers each with 3 X 3 kernels, was employed. A 64-length feature maps normalized by batch normalization layers was used. The normalization layer uses parametric ReLU activation function. Radford proposed network was used with Leaky ReLU activation. Convolutional layers proposed in[42] were used for increasing the resolution of input images. In the proposed framework [48], network learns solutions that are very close to the real images. Discriminator network was trained to differentiate actual images from generated ones. The network have eight convolutional layers and increasing kernel size. Finally, Euclidean distance was used for calculating distance between reference image and constructed one.

Residual Learning of Deep Convolutional Neural Network [52]

A Gaussian denoiser can be used in single image super resolution[52]. Training a deep convolutional neural network for a specific task generally consists of two stages, architectural network design and using training data for model learning. Zhang K . et al [52] used VGG network that was proposed in [53] with few modifications for super resolution purpose. The network depth is based on the patch size as done by denoising methods. Residual learning with batch normalization was used for model learning. In the proposed framework, the input to the network was an image containing a noise. Residual learning was used to train a mapping function. The used network contains three types of layers.The first layer includes convolution and rectified linear units (ReLU) with 64 filters each of 3 X 3 size, this layer produces 64 feature maps. The second layer contains convolution with batch normalization (BN) and rectified linear units(ReLU), 64 filters each of (3X3X64) size was used in this layer as well as batch normalization was introduced between convolution and ReLU units. The last layer includes a convolution layer which is responsible for the reconstruction of output image. Zeros are added to the boundaries of the images before convolution to make the size of feature maps equivalent to that of input image.

V. DISCUSSION

In this section, a comparison between the main external learning based algorithms that are used in super resolution is discussed.

Advantages and Disadvantages

The table below shows the properties and limitations of each technique.

Technique	Advantages	Disadvantages
Sparse Coding	-Highly compact representation for dictionaries sizes **[54].** -No overlapping.	- High Operation Cost. -While considering the geometrical structure of the data, it does not take into account the dictionary atoms incoherence **[54]**. -Sparse coding techniques involve a costly sparse decomposition in which every input patch is represented as a linear combination of LR dictionary atoms. -A dictionary must be computed for each frame which makes sparse coding method, not suitable for real time applications **[5]**.

Anchored Regression	-The great reduction in the processing time. -The additional information that is garnered from the training data is stored in the memory which consequently used to improve the quality of the output.	-Most anchored regression based methods use nearest neighbor search for finding relevant patches which consumes a large amount of the processing time.
Regression Trees	-It presents the best balance between the quality and execution time.	- Large amount of memory is required for storing local linearization based regression parameters.
Deep Learning	-Simple and removes artifices as well as provides better resolution image in many cases. -Its ability of improvement modeling non linear functions, as opposed to the fat network which are more robustness against overfitting**[5]**.	-The quality of image is degraded because of mapping low resolution patch to several high resolution patches**[54]**. -Deep learning methods have a set of constraints. The four major problems include a) flat activation functions, b) gradients that are not very informative, c) inefficiency of a network left to learn itself and d) the architecture choice and design**[55]**. -Limitations because of the required balance between accuracy, generalization, and computational cost. -Compared to the traditional machine learning techniques, deep learning network takes longer time for fine-tuning of parameters.

Table 1. Advantages and Disadvantages of External Learning Based Algorithms

Performance Evaluation

The performance of external learning based algorithms can be compared by using the common evaluation metric, peak signal to noise ratio (PSNR). PSNR is the ratio between the maximum power of the signal and the corrupting noise power that affects the fidelity of the signal representation. Due to signal dynamic range, PSNR is measured in decibel scale according to the following equations:

$$MSE = \frac{1}{m\,n}\sum_{i=1}^{m-1}\sum_{j=1}^{n-1}|\,I(i,j)-k(i,j)\,|^2 \qquad (5)$$

$$PSNR = 20\log_{10}(MAX)_I - 10\log_{10}(MSE) \qquad (6)$$

A public and standard dataset,SET5 , at a scaling factor of two is used for fair comparison.The table below shows the PSNR values for the previously mentioned studies except [27] because of its using a text-based dataset.

Algorithm	PSNR in (dB)
Convolutional Sparse Coding **[13]**	36.60
Sparse Representation on a K-NN Dictionary **[12]**	27.49
Anchored Neighborhood Regression **[8]**	35.83
Super-Resolution using Random Forests (RF)**[32]**	36.55
Naive Bayes Super-Resolution Forest **[28]**	36.67
Sub-Pixel Convolutional Neural Network **[38]**	26.71
Generative Adversarial Network **[44]**	32.05
Residual Learning of Deep CNN **[48]**	37.58

Table 2. PSNR Values for The Mentioned Frameworks

VI. CONCLUSION

This paper illustrates the concept of image super resolution and its importance in generating high quality images that are required in many real world applications. Single and multi frame based super resolution with their differences are discussed in details. The degradation model which is used to obtain low resolution images from high resolution ones is explained. The taxonomy of the single image super resolution algorithms is reviewed. The external learning based algorithms are highlighted. Their advantages and limitations are introduced. A number of recent studies in the field of single image super resolution are analyzed and compared.

REFERENCES

[1] S. Park, M. Park, M. Kang, "Super resolution image reconstruction: a technical overview", IEEE Signal Processing Magazine,Vol. 20, No. 3, pp. 21-36, 2003.

[2] K.Nasrollahi, T. B. Moeslund, "Super-resolution: a comprehensive survey", Machine Vision and Applications,Vol.25, Issue.6, pp.1423-1468, 2014.

[3] R.C. Aster, B. Borchers, and C.H Thurber,"Parameter Estimation and Inverse Problems", Elsevier, 2nd Edition, 9780123850485, 2012.

[4] D. HiengLing, H. Hsu, G.C. Lin, S. Lee, "Enhanced image-based coordinate measurement using a super-resolution method", Robotics and Computer-Integrated Manufacturing Vol.21, Issue.6, pp.579-588, 2005.

[5] J. Salvador,"Example-Based Super Resolution", Academic Press, 1st Edition, 9780081011355, 2016.

[6] W. T. Freeman, E.C. Pasztor, O.T. Carmichael, "Learning low level vision," International Journal of Computer Vision, Vol. 40, No. 1, pp. 25-47, 2000.

[7] D. Glasner, S. Bagon, M. Irani, ,"Super-resolution from a single image", IEEE International Conference on Computer Vision(ICCV), pp. 349-356, Japan, 2009.

[8] R.Timofte, V. De, L.V Gool, "Anchored neighborhood regression for fast example-based super-resolution", IEEE International Conference on Computer Vision(ICCV), pp.1920–1927 Australia, 2013.

[9] J. Yang, J. Wright , T.S. Huang, Y. Ma, "Image super-resolution via sparse representation" , IEEE Transactions on Image Processing,Vol.19, Issue.11, pp.2861-2873, 2010.

[10] M. Bevilacqua, A. Roumy, C. Guillemot, M.A Morel, "Low-complexity single-image super-resolution based on

nonnegative neighbor embedding", British Machine Vision Conference (BMVC), pp.1-10, United Kingdom, 2012.
[11] H. Lee, A. Battle, R. Raina, A.Y.Ng, "Efficient sparse coding algorithms", Annual Conference on Neural Information Processing Systems (NIPS), pp. 801–808, Canada, 2006.
[12] J.Yang , J. Wright, T. Huang, Y. Ma ,"Image super-resolution as sparse representation of raw image patches", IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1-8,USA, 2008.
[13] S. Gu, W.Zuo, Q. Xie, D. Meng , X. Feng, L. Zhang ,"Convolutional Sparse Coding for Image Super-Resolution", IEEE International Conference on Computer Vision (ICCV), pp. 1823-1831, Chile, 2015.
[14] B.Wohlberg, "Efficient convolutional sparse coding", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.7173-7177, Italy, 2014.
[15] L.W. Zhong, J.T. Kwok ,"Fast stochastic alternating direction method of multipliers", International Conference on Machine Learning (ICML), pp. 46-54, China, 2014.
[16] L. He , H. Qi , R. Zaretzki,"Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 345-352,USA , 2013.
[17] Y. Zhu , Y. Zhang, A. L. Yuille, "Single Image Super-resolution Using Deformable Patches", IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 2917-2924, USA, 2014.
[18] L. Ning, L. Shuang, "Single Image Super-Resolution Using Sparse Representation on a K-NN Dictionary", International Conference on Image and Signal Processing(ICISP), pp. 169-178, Canada, 2016.
[19] A. Bhaskara Rao, J. Vasudeva Rao, "Super resolution of quality images through sparse representation", ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of CSI - Volume II. AISC, Vol. 249, pp. 49–56, 2014.
[20] Y. Wang, P. Fu, "Sparse representation based medical MR image super-resolution", International Journal of Advancements in Computing Technology, Vol.4, No.19, pp.26-31,2012.
[21] F. Juefei-Xu, M. Savvides ,"Single face image super-resolution via solo dictionary learning", IEEE International Conference on Image Processing (ICIP), pp.2239-2243, Canada, 2015.
[22] J. Xie, C. Chou, R. Feris , M. Sun,"Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering", IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, China, 2014.
[23] H. GE, S. RR,"Reducing the dimensionality of data with neural networks", Science, Vol. 313, pp. 504–507, 2006.
[24] H. GE,"Training products of experts by minimizing contrastive divergence", Neural Computation,Vol.14, No.8, pp.1711-1800, 2002.
[25] W. T. Freeman, T.R. Jones, E.C. Pasztor, "Example-based super-resolution", IEEE Computer Graphics and Applications, Vol. 22, Issue 2, pp. 56-65, 2002.
[26] H. Chang , D. Yeung, Y. Xiong ," Super-resolution through neighbor embedding", IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 275-282, USA, 2004.
[27] R. Zeyde, M.l Elad, M. Protter, "On Single Image Scale-Up Using Sparse-Representations", International Conference on Curves and Surfaces, pp. 711-730, France, 2010.
[28] A. Abedi, E. Kabir, "Text-image super-resolution through anchored neighborhood regression with multiple class-specific dictionaries", Signal, Image and Video Processing, Vol.11, Issue 2, pp. 275–282, 2017.
[29] R. Timofte, V.D. Smet, L.V Gool, " A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution", Asian Conference on Computer Vision (ACCV), pp. 111-126, Singapore, 2014.
[30] X. Chen, C. Qi, "Document image super-resolution using structural similarity and Markov random field", IET Image Processing,Vol.8, Issue.12, 2014.
[31] R. Timofte, V.D. Smet, L.V Gool ,"Semantic super-resolution: When and where is it useful?", Computer Vision and Image Understanding, Vol.142, pp. 1-12, 2016.
[32] S. Schulter, C. Leistner, H. Bischof,"Fast and accurate image upscaling with super-resolution forests", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3791-3799, USA, 2015.
[33] A. Criminisi, J. Shotton, "Decision forests for computer vision and medical image analysis", Springer-Verlag London, 1st Addition, 9781447149293,2013.
[34] J. Salvador, E. Pérez Pellitero, " Naive bayes super-resolution forest", IEEE International Conference on Computer Vision(ICCV), pp. 325-333, Chile, 2015.
[35] C.Y. Yang, M. H. Yang, "Fast Direct Super-Resolution by Simple Functions", IEEE International Conference on Computer Vision(ICCV), pp 561-568, Australia, 2013.
[36] A. Criminisi, J. Shotton, E. Konukoglu," Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning", Foundations and Trends® in Computer Graphics and Vision,Vol.7, No.2, pp. 81-227, 2012.
[37] J. Huang , W. Siu, T. Liu," Fast Image Interpolation via Random Forests", IEEE Transactions on Image Processing, Vol. 24, Issue10, pp. 3232- 3245, 2015.
[38] S. Bernard, L. Heutte , S. Adam, "On the selection of decision trees in Random Forests", IEEE International Joint Conference on Neural Networks(IJCNN), pp. 302-307, USA, 2009.
[39] F. Ranold, "Dispersion on a sphere", Royal Society of London,Series A, Mathematical and Physical Sciences, Vol. 217, Issue 1130, pp. 295-305, 1953.
[40] S. McCann, D. Lowe," Local Naive Bayes Nearest Neighbor for image classification", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3650-3656,USA, 2012.
[41] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition", Neural Computation,Vol.1, No. 4, pp.541-51, 1989.
[42] W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D.Rueckert, Z. Wang," Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 1874-1883,USA, 2016.
[43] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes, R.M. Mersereau, "Eigenface-domain super-resolution for face recognition", IEEE Transactions on Image Processing, Vol.12, Issue.5, pp.597-606, 2003.
[44] M.D. Zeiler, R.Fergus," Visualizing and Understanding Convolutional Networks", European Conference on Computer Vision(ECCV), pp. 818-833, 2014.
[45] J. Long, E. Shelhamer, T.Darrell, “Fully convolutional networks for semantic segmentation”, IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.3431-3440, USA, 2015.
[46] C. Dong, C.C. Loy, K. He, X.Tang," Image Super-Resolution Using Deep Convolutional Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol.38, Issue.2, pp. 295- 307, 2016.
[47] M.D.Zeiler, G.W.Taylor, G.W. Taylor," Adaptive deconvolutional networks for mid and high level feature learning ", IEEE International Conference on Computer Vision (ICCV), pp. 2018-2025, Spain, 2011.
[48] C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi," Photo-realistic single image super-resolution using a generative adversarial network", arXiv preprint arXiv:1609.04802, 2016.
[49] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778,USA, 2016.
[50] K. He, X. Zhang, S. Ren, J. Sun,"Identity Mappings in Deep Residual Networks", European Conference on Computer Vision(ECCV), pp. 630-645, Netherlands, 2016.

[51] I.Goodfellow, J Pouget-Abadie, M. Mirza, B. Xu , D. Warde-Farley, S. Ozair, A. Courville, Y.Bengio, "Generative adversarial nets", neural information processing systems(NIPS), pp. 2672-2680, 2014.

[52] K. Zhang, W. Zuo, Y. Chen, D.Meng, L. Zhang," Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising", IEEE Transactions on Image Processing, Vol.26, Issue. 7, 2017.

[53] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", International Conference on Learning Representations, USA, 2015.

[54] J. Dalvadi," A Survey on Techniques of Image Super Resolution", International Journal of Innovative Research in Computer and Communication Engineering, Vol.4, Issue 3, 2016.

[55] S. Shalev-Shwartz, O. Shamir, and S. Shammah, “Failures of Deep Learning”, arXiv preprint arXiv:1703.07950, 2017

[56] [Available Online]: https://arxiv.org/pdf/1703.07950v1.pdf [Last accessed: 13-7-2019].

★★★