International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks

[1]K Prasanthi[2]Kasinadhuni Vyshanavi Supraja, [3]Kishtam Harikrishna, [4]Katteboina Varaprasad, [5]Alugumalli Uday, [6]Bhaskar Reddy, [7]Vennababu
[1,2,3,4,,5,6,7]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

In this project work, we develop a new image super-resolution (SR) approach based on a Convolution Neural Network (CNN), which jointly learns the feature extraction, upsampling, and high-resolution (HR) reconstruction modules, yielding a completely end-to-end trainable deep CNN. However, directly training such a deep network in an end-to-end fashion is challenging, which takes a longer time to converge and may lead to sub-optimal results. To address this issue, we propose to jointly train an ensemble of deep and shallow networks. The shallow network with weaker learning capability restores the main structure of the image content, while the deep network with stronger representation power captures the high-frequency details. Since the shallow network is much easier to optimize, it significantly lowers the difficulty of deep network optimization during joint training. To further ensure more accurate restoration of HR images, the high frequency details are reconstructed in a multi-scale manner to simultaneously incorporate both short- and long-range contextual information. The proposed method is extensively evaluated on widely adopted data sets and compares favorably against state-of-the-art methods. In-depth ablation studies are conducted to verify the contributions of different network designs to image SR, providing additional insights for future research.

1. INTRODUCTION

Single image super-resolution (SR) aims at restoring the high resolution (HR) image with abundant high-frequency details from the low resolution (LR) observation. Given that multiple HR images can be down-sampled into the same LR image, SR as the reverse problem is inherently ill-posed with insufficient knowledge. Recently, learning-based methods have attracted increasingly more attention and delivered superior performance in image SR. The basic idea is to learn the mapping function from the LR image to the HR counterpart using auxiliary data. A variety of machine learning algorithms, on popular idea for image SR with CNNs focuses on learning the residual between the HR image and the bicubic-interpolated LR image, assuming that the target HR image shares the similar main structure to the bicubic up sampled LR version. However, the hand-crafted bi cubic interpolation is not specifically designed for this purpose and may hinder the final performance. As opposed to the above CNN with bicubic interpolation based approaches, our method learns a direct mapping from LR to HR images with CNNs. However, our preliminary experiments suggest that training a sophisticated deep network in such an end-to-end fashion is challenging, leading to sub-optimal results. To address this issue, we propose to jointly train an ensemble of deep and shallow networks. Specifically, the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

shallow network is lightweight (e.g., only 3 convolution layers) and easier to optimize, while the deep network is elaborately designed and consists of three major procedures.

Firstly, feature extraction is performed to map the original LR image into a deep feature space. The deep features are then upsampled to the target spatial size with learned filters. Finally, the HR image is reconstructed by considering multi-scale context of the up sampled deep features. During joint training, the shallow network converges quickly and captures the major structure of the HR image, i.e., mostly low-frequency content. As a consequence, the deep network is only responsible to restore the high-frequency details based on the main image structure, which effectively lowers the difficulty of deep network training. The proposed network ensemble is similar to the above CNN with bicubic interpolation based approaches in that the deep network is designed to learn the high frequency residual content. However, different from these approaches, our method replaces the bicubic interpolation with a shallow network, allowing fully end-to-end trainable. It has also been that reconstructing a pixel may depend on either short- or long-range contextual information. Some CNN-based approaches rely on small image patches to predict thecentral pixel value, which is less effective for SR with large up scaling factors.

2. LITERATURE SURVEY

We describe a learning-based method for low-level vision problems—estimating scenes from images. We generate a synthetic world of scenes and their corresponding rendered images, modeling their relationships with a Markov network. Bayesian belief propagation allows us to efficiently find a local maximum of the posterior probability for the scene, given an image. We call this approach VISTA—Vision by Image/Scene Training.

We apply VISTA to the "super-resolution" problem (estimating high frequency details from a low-resolution image), showing good results. To illustrate the potential breadth of the technique, we also apply it in two other problem domains, both simplified. We learn to distinguish shading from reflectance variations in a single image under particular lighting conditions. For the motion estimation problem in a "blobs world", we show figure/ground discrimination, solution of the aperture problem, and filling-in arising from application of the same probabilistic machinery. Methods for super-resolution can be broadly classified into two families of methods: (i) The classical multi-image super-resolution (combining images obtained at sub pixel misalignments), and (ii) Example-Based super-resolution (learning correspondence between low and high resolution image patches from a database). In this paper we propose a unified framework for combining these two families of methods. We further show how this combined approach can be applied to obtain super resolution from as little as a single image (with no database or prior examples). Our approach is based on the observation that patches in a natural image tend to redundantly recur many times inside the image, both within the same scale, as well as across different scales. Recurrence of patches within the same image scale (at subpixel misalignments) gives rise to the

classical super-resolution, whereas recurrence of patches across different scales of the same image gives rise to example-based super-resolution. Our approach attempts to recover at each pixel its best possible resolution increase based on its patch redundancy within and across scales..

3.SYSTEM ANALYSIS

Image SR can be generally classified into three categories, i.e., interpolation-based reconstruction based and learning-based methods Among them, learning-based methods become a hot research point in the field of image SR in recent years, whose basic idea is to formulate image SR as a nonlinear mapping from LR to HR images and learn the mapping using auxiliary data in a supervised manner. The opening work is proposed by Freemanetal. , which employs Markov Random Field (MRF) and patchbased external examples to produce effective magnification. Inspired by various methods have been developed subsequently. One of the representative methods is based on the sparse representation algorithm, which ensures that HR patches have a sparse linear representation over an overcomplete dictionary of patches randomly sampled form similar images. Yangetal. train LR and HR dictionaries jointly with the constraint that LR patches and the corresponding HR counterparts share the same sparse representation. This work is developed by which employs K-SVD to train the coarse dictionary and Orthogonal Matching Pursuit (OMP) to solve the decomposition problem. Based on the neighbor embedding algorithm, works of super-resolve LR images with the assumption that LR and HR patches lie on low-dimensional nonlinear manifolds with locally similar geometry. To further improve computational efficiency, some techniques are put forward. Yang and Yang cluster LR feature space into numerous subspaces and learn simple mapping functions for each subspace. This propose is to use a number of linear regressors to locally anchor the neighbors. With the precalculated anchors and regressors, ‘‘A+’’ [11] increases SR performance both in terms of accuracy and speed. Based on the regression trees or forests algorithm, another line of image SR technique is proposed, which builds on linear multivariate regression models using leaf nodes and locally linearizes the mapping from LR to HR patches around centroids. Deep learning based methods have recently been applied to image SR and delivered compelling performance a CNN comprising three convolution layers is proposed for image SR. Later on, reformulate traditional sparse coding based method as deep networks and achieve promising results. Reference restores the HR images using a Gibbs distribution as the conditional model, with its sufficient statistics predicted by a CNN. Inspired by the residual prediction based methods Kim et al. propose a deep network with 20 convolutional layers to learn the residual between HR and LR images, which boosts performance by a large margin. The authors also present a deeply-recursive convolutional network to restore the HR images . This propose to extract feature maps in the LR space and learn to increase the resolution only at the very end of the network, which shows that the learned upscaling filters can further increase the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

accuracy of prediction. Subsequently, many other CNN-based techniques are applied in image SR, such as densely connected network recursive network and cascade upsampling network and so on. Compared with the above works, we propose a fully end-to-end trainable system which adopts an ensemble of deep and shallow networks. In addition, a multi-scale HR image restoration module is also designed to aggregate both short- and long-range contextual information. These techniques have not been simultaneously explored in existing methods.

DISADVANTAGES

Less accuracy score
Low performance
Unable to predict the resolution

4. PROPOSED SYSTEMS

In this section, we introduce the proposed EEDS (End-to-End Deep and Shallow networks) method for image SR. This overviews the architecture of the network ensemble comprising a deep and a shallow CNN. The deep CNN can be further divided into three modules: feature extraction, up sampling and multi-scale reconstruction.

A. FEATURE EXTRACTION

In order to extract local features of high-frequency content, traditional shallow methods perform feature extraction by computing the first and second order gradients of the image patch, which is equivalent to filtering the input image with hand-designed, high-pass filters. Rather than manually designing these filters, deep learning based methods automatically learn these filters from training data. However, some works extract features from the coarse HR images, which is obtained by up sampling the LR images to the HR size with bicubic interpolation. We argue that the bicubic interpolation is not specifically designed for this purpose, and even damages important LR information that may play a central role in restoring the HR counterparts. Therefore, the proposed method adopts an alternative strategy and performs feature extraction directly on the original LR images with convolution layers. Our feature extraction module consists of three convolution layers interleaved by Rectified Linear Unites (ReLUs) acting as nonlinear mappings. A shortcut connection with identity mapping is used to add the input feature map of the second layer to the output of the third layer, which is formulated as a ‘‘residual unit’’. As justified by such residual unit can effectively facilitate gradients flow through multiple layers, thus accelerating deep network training. Similar structures have also been used in our reconstruction module. All three convolution layers have the same kernel size of 3×3 and generate feature maps of 64 channels. Zero padding is adopted to preserve the spatial size of the output feature maps.

B. UPSAMPLING

Given the extracted features from the original LR images, upsampling operation is performed to increase their spatial span to the target HR size. Instead of using hand-designed interpolation methods, we prefer a learning based upsampling operation, giving rise to an end-to-end trainable system. To this end, we consider two different strategies widely adopted in CNN for up sampling, i.e., un pooling and deconvolutions. As opposed to pooling layers, the un pooling operation

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

with an up scaling factor s replaces each entry in the input feature map with a s × s block, where the top left element in the block is set to the value of the input entry and the others to zero. The un pooling operation yields enlarged yet sparse output feature maps. The sparsely activated output values can then be propagated to local neighborhoods by subsequent convolution layers. The deconvolution layer up scales the input feature maps by s-fold through reversing the forward and backward propagation of convolution layers with an output stride of s. Although un pooling and deconvolution resort to different implementations, they are essentially similar in up scaling feature maps and both are well suited to our task. We adopt the deconvolution layer and achieve promising performance.

C. MULTI-SCALE RECONSTRUCTION

Since similar image patterns may recur across different scales in different images of both training and test sets, accurate inference of the input image should be highly invariant to image scale variations and may rely on the aggregation of multi-scale contextual information. This insight has been intensively studied and verified in vision related problems, like image object detection [39], scene recognition , etc. From the perspective of image SR, some prior methods have also confirmed that multi-scale context can effectively benefit HR image reconstruction. Considering that HR image restoration may rely on both short- and long-range contextual information, we propose to perform HR reconstruction with multi-scale convolutions to explicitly encode multi-context information. The input of our HR reconstruction module firstly go through R residual units. Then a dimension reduction layer is followed that consists of a 1 × 1 convolution, mapping the input feature map of 64 channels to the output 16 channels. The subsequent multi-scale convolution layer comprises 4 convolution operations of 1 × 1, 3 × 3, 5 × 5, and 7 × 7 kernel sizes, respectively. All four convolutions are simultaneously conducted on the input feature map and produce four feature maps of 16 channels.

ADVANTAGES

Good accuracy score
Good performance
Predict the higher resolution

SOFTWARE REQUIREMENTS

Operating System : Windows 7/8/10 or Linux or MAC
Language : Python 3.X
HARDWARE REQUIREMENTS
Processor : Pentium 3, Pentium 4 and higher
RAM : 2GB and higher
Hard disk : 80GB and higher

IMPLEMENTATION

5. ARCHITECTURE ANALYSIS:

To gain further insights of our contributions, we conduct additional evaluations on different variants of the proposed EEDS method. Unless stated otherwise, we strictly follow the implementation settings in Section IV-A to train all the methods. Our method jointly trains a deep and a shallow network as an ensemble. To investigate the impact of the two networks on the final performance, we split the two networks and obtain two variants of the proposed EEDS model, namely, EED

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

(end-to-end deep network) and EES (end-to-end shallow network), respectively. Fig. 3 depicts the convergence plots of all three models on the Set5 data set. EES with a shallow network takes less time to converge. However, limited by its capacity, the final performance of EES is relatively low. In contrast, EED is more difficult to train. The training process is very unstable with oscillation in training loss. Upon convergence, EED achieves higher PSNR than EES, but is still unsatisfactory. This may be attributed to the fact that directly mapping LR images to HR ones is a very complex task and EED may converge to some local minimum. The proposed EEDS method mitigates this issue by combining deep and shallow networks as an ensemble. At joint training, the shallow network still converges much faster and dominates the performance at the very beginning. After the shallow network has already captured the major components of the HR images, the difficulty of direct SR has been significantly lowered. The deep network then starts to focus on the high-frequency details and learns to correct the errors made by the shallow network nd achieves the best performance among all three methods. Upon convergence, the prediction made by the shallow network of EEDS restores most content with blur and artifacts , whereas the deep network of EEDS learns to predict the residual between the HR image and the output of the shallow network, mostly containing high-frequency content The behavior of deep and shallow networks combined through simple addition is supported by and further confirms the key findings of deep residual networks ,indicating that deep residual learning can be achieved through addition of subnetworks and makes deep networks more easier to optimize. Meanwhile, the addition of deep and shallow networks is also consistent to prior SR methods where SR is conducted by learning the residual between HR image and the bicubic interpolated LR input. As opposed to these approaches, our EEDS method replaces the fixed bicubic interpolation with a shallow network and jointly trains the deep and shallow networks, making the residual prediction based method a special case of our method. To study the impact of combining deep and shallow CNNs on other network architectures, we compare an eight-layer baseline deep CNN (denoted as DCNN) that has similar architecture to SRCNN against the combination of the deep CNN and a 3-layer shallow CNN (denoted as DSCNN). DSCNN consistently outperforms DCNN across all the data sets, suggesting that the benefits of combining deep and shallow networks can generalize to other network architectures.

PROBLEM STATEMENT

The problem here is when we are transferring the images , they loose their resolution as a result the clarity of the image decreases , so in order to increase the clarity of image we use CNN technique and convert the low resolution image to High resolution image

6. Results

The following shows the series of output screens and how the actual process of implementing CNN takes place

The first figure of the output screen shows the information ablout the images that are used for working

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

All the images are converted in this format and put in a folder called output

CONCLUSION

In this project a fully end-to-end trainable system for single image SR using an ensemble of deep and shallow networks. The shallow network with a lightweight architecture is easy to optimize and learns to render the major structure of the HR image, while the deep network with a stronger learning capability is only responsible to capture the high frequency details. As such, jointly training the network ensemble can significantly lower the difficulty of network training and gives rise to more superior performance. To ensure more accurate restoration of HR images, the HR reconstruction is performed in a multi-scale manner to simultaneously incorporate both short- and long-range contextual information. Experiments confirm that the proposed method performs favorably against state-of-the-art approaches. In-depth ablation studies are also conducted to verify the contributions of different network designs to image SR, providing additional insights for future research

REFERENCES

[1] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, ‘‘Learning low-level vision,’’ Int. J. Comput. Vis., vol. 40, no. 1, pp. 25–47, 2000.

[2] W. T. Freeman, T. R. Jones, and E. C. Pasztor, ‘‘Example-based super-resolution,’’ IEEE Comput. Graph. Appl., vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.

[3] D. Glasner, S. Bagon, and M. Irani, ‘‘Super-resolution from a single image,’’ in Proc. IEEE Int. Conf. Comput. Vis., Sep. 2009, pp. 349–356.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, ‘‘Image super-resolution via sparse representation,’’ IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] R. Zeyde, M. Elad, and M. Protter, ‘‘On single image scale-up using sparse-representations,’’ in Proc. Int. Conf. Curves Surf. Berlin, Germany: Springer, Jun. 2010, pp. 711–730.

[6] G. Freedman and R. Fattal, ‘‘Image and video upscaling from local selfexamples,’’ ACM Trans. Graph., vol. 30, no. 2, p. 12, 2011.

[7] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, ‘‘Coupled dictionary training for image super-resolution,’’ IEEE Trans. Image Process., vol. 21, no. 8, pp. 3467–3478, Aug. 2012.

[8] H. Chang, D.-Y. Yeung, and Y. Xiong, ‘‘Super-resolution through neighbor

embedding,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jun./Jul. 2004, pp. 275–282.

[9] C.-Y. Yang and M.-H. Yang, ''Fast direct super-resolution by simple functions,'' in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 561–568.

[10] R. Timofte, V. Smet, and L. Van Gool, ''Anchored neighborhood regression for fast example-based super-resolution,'' in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 1920–1927