BENGALURU
DR. B. R. AMBEDKAR
SCHOOL OF
ECONOMICS
A Unitary University, Govt. of Karnataka

BASE University Working Paper Series: 07/2021

Image Inpainting via Generative Multi-column with the aid of Deep Convolutional Neural Networks

Rajesh B

Muralidhara B L

May 2021

Image Inpainting via Generative Multi-column with the aid of Deep Convolutional Neural Networks

Abstract

Images can be described as visual representations or likeness of something (person or object or a scanned document) which can be reproduced or captured, e.g. a hand drawing, photographic material. The advent of the digital age has seen the rapid shift image storage technologies, from hard-copies to digitalized units in a less burdensome manner with the application of digital tools. The research aims to design a confidence-driven reconstruction loss while an implicit diversified Markov Random Field (MRF) regularization is adopted to enhance local details. The multi-column network combined with the reconstruction and MRF loss propagates local and global information derived from context to the target inpainting regions. Extensive experiments on challenging street view, face, natural objects and scenes manifest that our proposed method produces visual compelling results even without previously common post-processing. The research involves pre-trained Deep Convolutional Neural Network (DCNN) and their training networks like ResNet50, GoogleNet, AlexNet and VGG-16. The average PSNR performance of the proposed model is 24.64db and Structural Similarity Index Measure (SSIM) is 0.9018.

Keywords: Markov Random Field, Deep Convolutional Neural Network, ResNet50, GoogleNet, AlexNet, VGG-16, Structural Similarity Index Measure.

Image Inpainting via Generative Multi-column with the aid of Deep Convolutional Neural Networks[1]

Rajesh B[2], Muralidhara B L[3]

Introduction

Image inpainting originated from an ancient technique performed by artists to restore damaged paintings or photographs with small defects such as scratches, cracks, dust and spots to maintain its quality to as close to the original as possible. [1] Image inpainting is an ill-posed inverse problem that has no well-defined unique solution. The evolution of computers in the 20th century, its frequent daily use and the development of digital tools with image manipulation capability, has encouraged users to appreciate image editing, e.g. restoration, and the application of on-screen visual display and special effects to images. As a result image inpainting (henceforth inpainting) has become a state-of-the-art restoration technique.[2]In a computer vision and graphics context, inpainting is a method that interpolates neighboring pixels to reconstruct damaged, or defective, portions of an image without any noticeable change on the restored regions when visually compared with the rest of the image.[3]These damaged portions/areas of an image are a set of unconnected pixels surrounded by a set of known adjacent pixels.[4] During the reconstruction of disconnected pixels, the inpainting method uses known-information to fill unknown regions. Image inpainting is an ill-posed inverse problem that has no well-defined unique solution[5] To solve the problem, it is therefore necessary to introduce image priors. All methods are guided by the assumption that pixels in the known and unknown parts of the image share the same statistical properties or geometrical structures. This assumption translates into different local or global priors, with the goal of having an inpainted image as physically plausible and as visually pleasing as possible [6]. The first category of methods, known as diffusion-based inpainting, introduces smoothness priors via parametric models or partial differential equations (PDEs) to propagate (or diffuse) local structures from the exterior to the interior of the hole[7].Many variants exist using different models (linear, nonlinear, isotropic, or anisotropic) to favor the propagation in particular directions or to take into account the curvature of the structure present in a local neighborhood[8]. These methods

[1] This paper was presented at the BASE University Research Seminar on April 5th 2021. The authors thank all participants for their valuable comments, which are all incorporated, and encouraging feedback

[2] Rajesh B, Assistant Professor, Bengaluru Dr. B. R. Ambedkar School of Economics, Bengaluru
Email: rajeshbalarama@base.ac.in

[3] Muralidhara B L, Professor, Department of Computer Science & Application, Bangalore University, Bengaluru
Email: murali@bub.ernet.in

are naturally well suited for completing straight lines, curves, and for inpainting small regions [9] They, in general, avoid having unconnected edges that are perceptually annoying. However, they are not well suited for recovering the texture of large areas, which they tend to blur [10] The second category of methods is based on the seminal work and exploits image statistical and self-similarity priors. The statistics of image textures are assumed to be stationary (in the case of random textures) or homogeneous (in the case of regular patterns) [11] The texture to be synthesized is learned from similar regions in a texture sample or from the known part of the image. Learning is done by sampling, and by copying or stitching together patches (called examplar) taken from the known part of the image, using an exemplar image as a source, and where pixel values are selected one pixel at a time. [12].The proposed method used in the paper to restore the image is ResNet-50. ResNet-50 can easily gain accuracy along with the greatly increased of depth. Since ResNet-50 has a very good performance of image classification, and can extract highquality features of images[13]. Hybrid solutions have then naturally emerged, which combine methods dedicated to structural (geometrical) and textural components.[14] This article surveys the theoretical foundations, the different categories of methods, and illustrates the main applications.[15]

Image recognition can be realized automatically using machine learning, deep learning techniques or other conventional methods [16]. Machine learning is based on the human classification of different types of images, while deep learning extracts features directly from images. In deep learning, the Convolution Neural Networks (CNNs) are used to make predictions. Such networks have recently achieved high accuracy in image recognition applications, in some cases even outperforming humans [17]. On the other hand, thousands of images are needed to gain sufficient accuracy using deep learning techniques. As a consequence, this causes the learning process to be time-consuming, even if Graphics Processing Units (GPUs) are used [18].

Literature Review

He *et al.* (2018) [19] used a dual-phase algorithm (Thieles rational interpolation function and Newton-Theiles function) for adaptive inpainting. This method uses continued fractions to update pixel intensity during the reconstruction of damaged portions based on the surrounding pixel information of known regions along the target region. That is, if the damaged pixel points are vertical, the selected points for interpolation of pixels are in the horizontal direction. The masked image is scanned line by line to locate and adopt information of known pixel points to perform interpolation of pixel intensity.

Ghorai et al. (2018) [20] proposed to use patch selection and refinement method based on joint filtering alongside a modified MRF to enhance optimal patch assignment to perform an inpainting task. This technique uses subspace clustering to select target patches from boundary regions into groups, which are refined via joint patch filtering to capture patterns and remove artefacts.

Wang et al. (2017) [21] used space varying update strategy powered by Fast Fourier transform for full image search. The base technique uses a standard deviation-based patch matching criterion and confidence term that evaluates the spatial distribution of patches to measure the amount of reliable information surrounding the priority point against a known priority point.

Sridevi and Kumar (2019) [22] proposed to use fractional-order derivative (integer-order derivative) with Discrete Fourier Transform (DFT) for inpainting task. The research used this method to achieve a good trade-off between the restored region and edge preservation, and also because DFT are easy to implement. Using fractional order derivative, pixel level on the whole image is considered instead of just considering neighbouring pixel values.

To optimize the network, a conditional constraint loss handles appearance and perceptual features extracted from VGG16 Johnson et al. (2016) [23] using the `1 as base. Both appearance and feature loss use the instance and masked image expressed as a function of the network and the mask. Other losses used are the KL divergence, reconstruction and ongoing adversarial loss. The cross semantic attention layer uses 1 × 1 convolutions to transform feature maps obtained by instance and masked images to evaluate cross attention before adding them to feed the decoder.

Comparative evaluations were carried out using the baseline models Song et al. (2018), [24] Quantitatively, the performance on 1000 CelebA-HQ images using centre mask of size 128 × 128 were better than the state-of-the-art. The limitation of this network is that there is a possibility of suffering from mode collapse (i.e. poor diversity in generated images) during training if trained in an unsupervised manner.

Proposed Methodology

Our inpainting system is trainable in an end-to-end fashion, which takes an image X and a binary region mask M (with value 0 for known pixels and 1 otherwise) as input. Unknown regions in image X are filled with zeros. It outputs a complete image$\hat{Y}$. The proposed DCNN based ResNet-50 consists of three sub-networks: a generator to produce results, global and local discriminators for adversarial training, and a pre-trained ResNet-50 to calculate ID-MRF loss. In the testing phase, only the generator network is used.

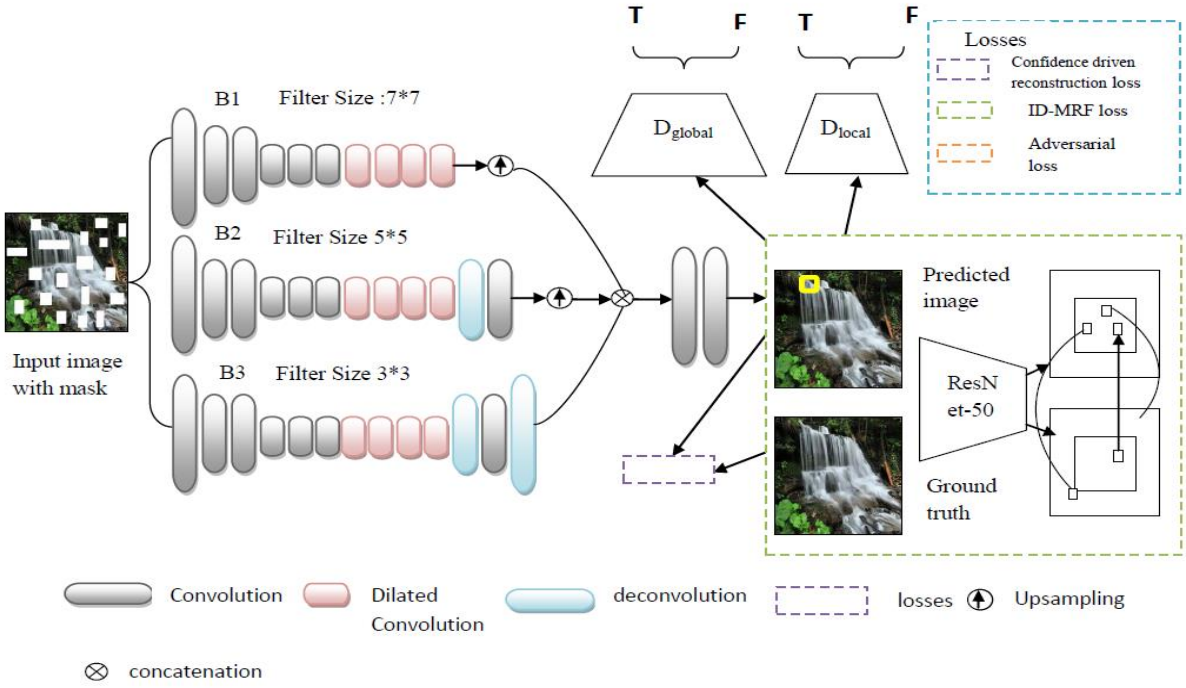


Figure: 1 Proposed Framework with the aid of ResNet-50

ResNet-50

ResNet-50 is short name for Residual Network that supports Residual Learning. The 50 indicates the number of layers that it has. So ResNet50 stands for Residual Network with 50 layers. DCNN have led to number of breakthroughs for image classification. In general the trend is to go deeper number of layers to solve complex tasks and to increase the classification and recognition accuracy. In a general DCNN, many layers are stacked and trained to the task at hand. In residual learning, instead of trying to learn some features, try to learn some residual. Residual can be simply understood as subtraction of feature learned from input of that layer. ResNet does this using shortcut connections (directly connecting input of nth layer to some (n+x)th layer. It has proved that training this form of networks is easier than training simple deep convolutional neural networks and also the problem of degrading accuracy is resolved.

The first problem with increasing depth is gradient explosion/dissipation, which is due to the fact that as the number of layers increases, the gradient back propagating in the network will become unstable with the multiplications and become very large or very small. One of the problems that often arises is gradient dissipation overcome gradient dissipation, many solutions have been found, such as using Batch Normalization, changing activation function to ReLU, and using Xaiver initialization, etc. It can be said that gradient dissipation has been well solved. Another problem with the network deepening is degradation, that is, the performance of the network is worse as the depth increases. From experience, the depth of the network is crucial to the performance of the model. When the number of network layers is increased, the network can carry out more complex feature pattern extraction, so better results can be obtained theoretically when the model is deeper. However, the experiment found that the deep network was degenerating. With the increase of network depth, the accuracy of the network tends to be saturated or even decreased. There is a decrease in the accuracy of the training set. We can determine that this is not caused by overfitting. Because the accuracy of the training set should be high in the case of overfitting. The residual network in ResNet is designed to solve this problem, and after solving this problem, the depth of the network rises by several orders of magnitude.

ResNet proposed two kinds of mapping: one is identity mapping, referring to the "curved curve" in Fig. 2 , and the other residual mapping refers to the part except the "curved curve", so the final output is $y = F(x) + x$. Identity mapping, as the name implies, refers to itself, which is x in the formula, while residual mapping refers to "difference", that is, $y - x$, so residual refers to F(x). At first, ResNet-50 performed convolution operation on the input, followed by 4 residual blocks, and finally performed full connection operation to achieve classification tasks. The network structure of ResNet-50 is shown in Fig. 2.

Fully connected (FC) layer usually appears at the end of the CNN to summarize the features of the previous layers. If we take the previous convolution and pooling as the process of feature engineering, local amplification and local feature extraction, the latter FC layer can be thought of as feature weighting. The structure of the FC layer shown in Fig. 2 is usually a way to quickly learn the nonlinear combinations of advanced attributes generated by the convolutional layer. The FC layer will learn a possible nonlinear function. The basic procedure of learning is as follows. First, the image, which has been converted into a form suitable for multilevel perceptron, is flattened into column vectors and fed back to the feed forward neural network. The flattened data is then applied to each iteration of the training. In this way, the model has the ability to distinguish between the major features in the image and some low-level features and classify them through classification techniques such as Softmax. Here we will output the classification results of the seven expressions

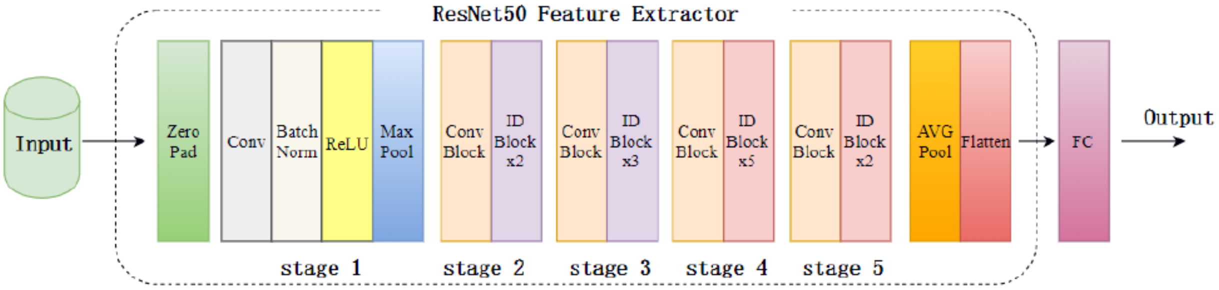
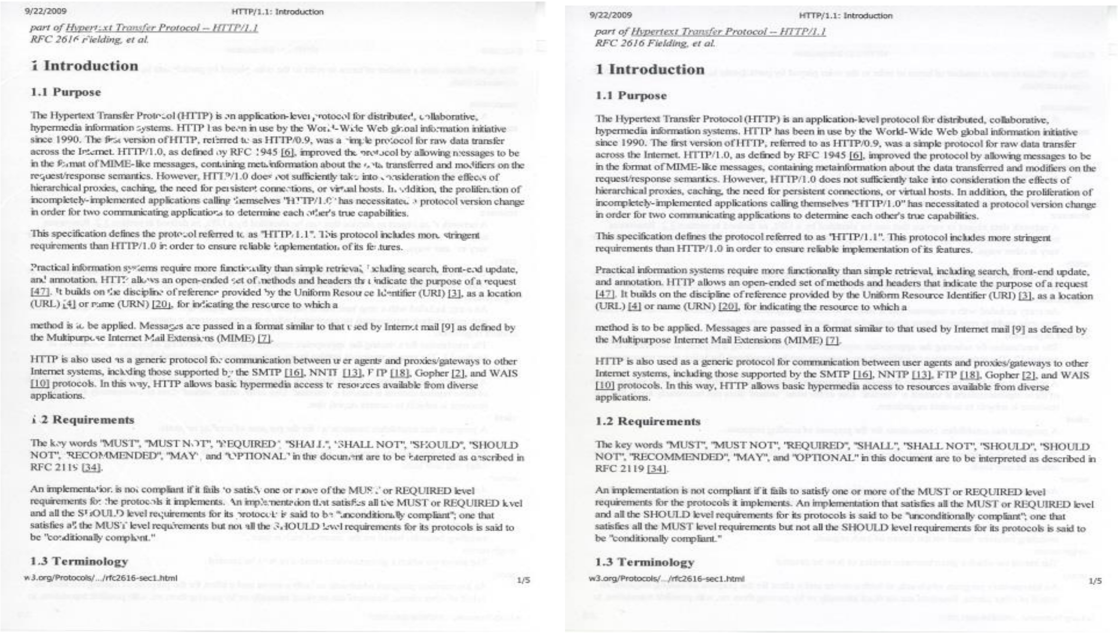


Figure: 2 Block diagram of ResNet-50

Results and Discussion

The performance of the proposed ResNet-50 attains better values in Peak signal-to-noise ratio (PSNR) and structural index similarity (SSIM) over comparative techniques. The research includes 5 difference cases to evaluate the performance of the proposed approach. It is obvious from the results that proposed ResNet-50 having better performance over other techniques. The figure 3 to 7 shows the visual image comparison for original image and ResNet-50 obtained image with respect to different databases.



9/22/2009 HTTP/1.1: Introduction

part of Hypertext Transfer Protocol -- HTTP/1.1
RFC 2616 Fielding, et al.

1 Introduction

1.1 Purpose

The Hypertext Transfer Protocol (HTTP) is an application-level protocol for distributed, collaborative, hypermedia information systems. HTTP has been in use by the World-Wide Web global information initiative since 1990. The first version of HTTP, referred to as HTTP/0.9, was a simple protocol for raw data transfer across the Internet. HTTP/1.0, as defined by RFC 1945 [6], improved the protocol by allowing messages to be in the format of MIME-like messages, containing metainformation about the data transferred and modifiers on the request/response semantics. However, HTTP/1.0 does not sufficiently take into consideration the effects of hierarchical proxies, caching, the need for persistent connections, or virtual hosts. In addition, the proliferation of incompletely-implemented applications calling themselves "HTTP/1.0" has necessitated a protocol version change in order for two communicating applications to determine each other's true capabilities.

This specification defines the protocol referred to as "HTTP/1.1". This protocol includes more stringent requirements than HTTP/1.0 in order to ensure reliable implementation of its features.

Practical information systems require more functionality than simple retrieval, including search, front-end update, and annotation. HTTP allows an open-ended set of methods and headers that indicate the purpose of a request [47]. It builds on the discipline of reference provided by the Uniform Resource Identifier (URI) [3], as a location (URL) [4] or name (URN) [20], for indicating the resource to which a

method is to be applied. Messages are passed in a format similar to that used by Internet mail [9] as defined by the Multipurpose Internet Mail Extensions (MIME) [7].

HTTP is also used as a generic protocol for communication between user agents and proxies/gateways to other Internet systems, including those supported by the SMTP [16], NNTP [13], FTP [18], Gopher [2], and WAIS [10] protocols. In this way, HTTP allows basic hypermedia access to resources available from diverse applications.

1.2 Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [34].

An implementation is not compliant if it fails to satisfy one or more of the MUST or REQUIRED level requirements for the protocols it implements. An implementation that satisfies all the MUST or REQUIRED level and all the SHOULD level requirements for its protocols is said to be "unconditionally compliant"; one that satisfies all the MUST level requirements but not all the SHOULD level requirements for its protocols is said to be "conditionally compliant."

1.3 Terminology

w3.org/Protocols/.../rfc2616-sec1.html 1/5

Figure: 3 Visual comparison of original and ResNet-50 obtained image for case-1

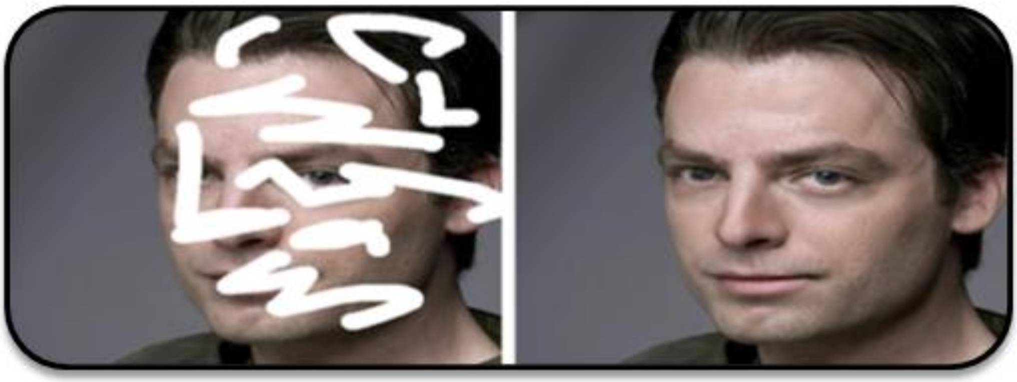


Figure: 4 Visual comparison of original and ResNet-50 obtained image for case-2

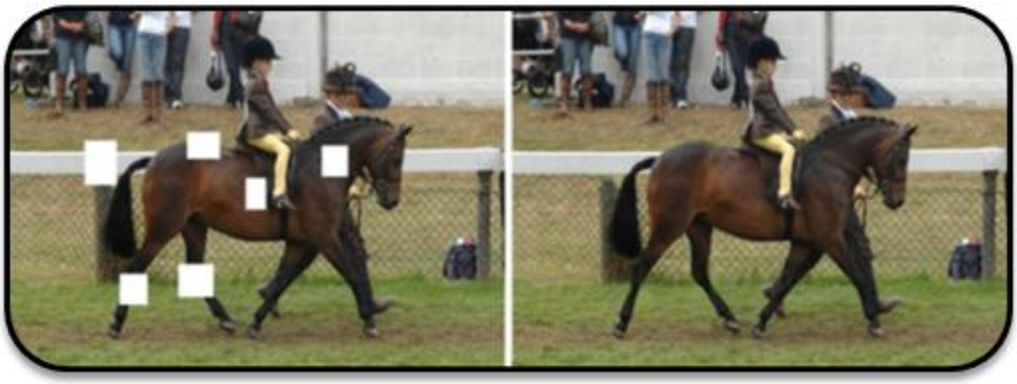


Figure: 5 Visual comparison of original and ResNet-50 obtained image for case-3

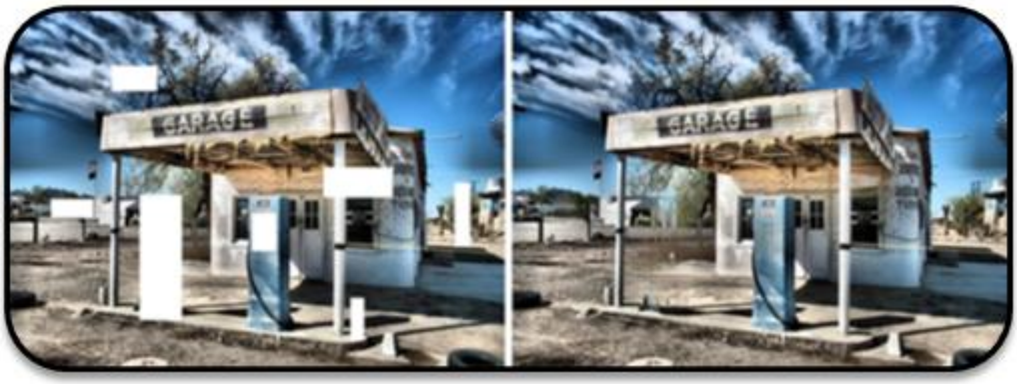


Figure: 6 Visual comparison of original and ResNet-50 obtained image for case-4

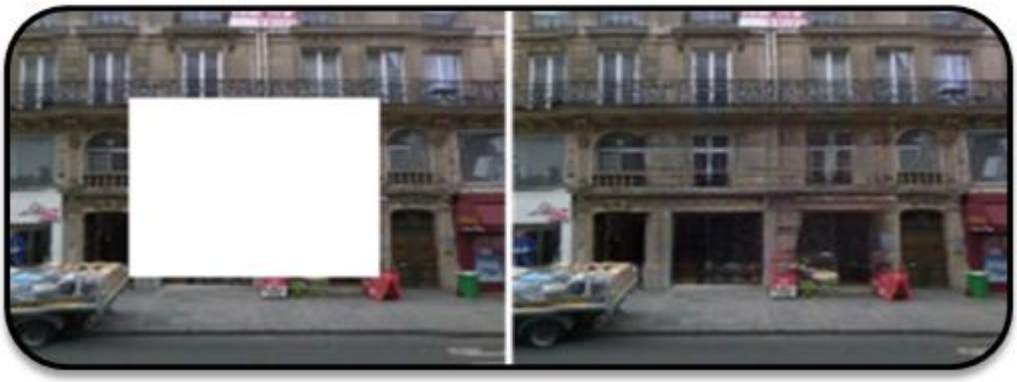


Figure: 7 Visual comparison of original and ResNet-50 obtained image for case-5

Performance evaluation through PSNR

PSNR is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. The PSNR quantifies the quality of a reconstructed or corrupt image with reference to the ground-truth. The PSNR value approaches infinity as the MSE approaches zero; this shows that a higher PSNR value provides a higher image quality. At the other end of the scale, a small value of the PSNR implies high numerical differences between images. The figure 8 to figure 12 shows the techniques based PSNR for different employed images. It is evident from all the validation that proposed ResNet-50 having better values over comparative techniques. In case-1, the proposed ResNet-50 achieves 24.65db that is 0.9db greater than AlexNet, 2.63db greater than VGG-16 and 1.67db better than GoogleNet. Similarly in subsequent cases the proposed ResNet-50 unveils better performance over comparative techniques.

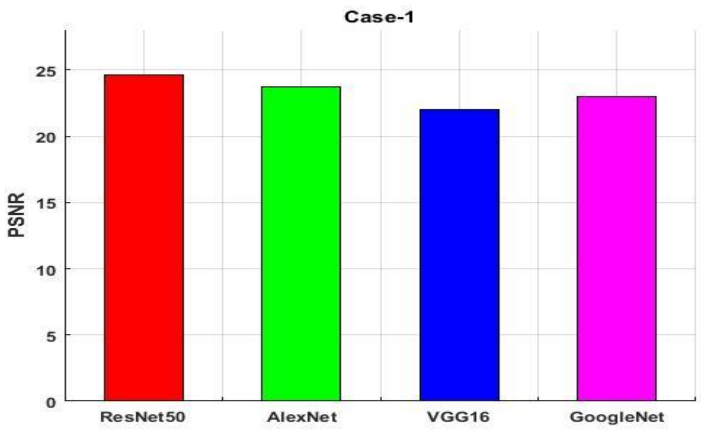


Figure: 8 Techniques wise PSNR comparison for case-1

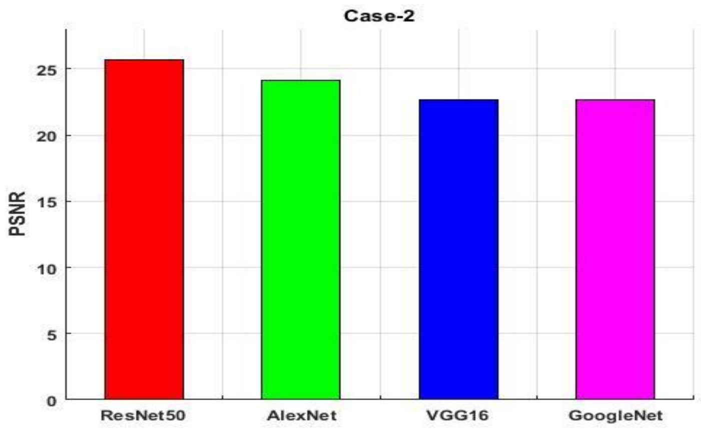


Figure: 9 Techniques wise PSNR comparison for case-2

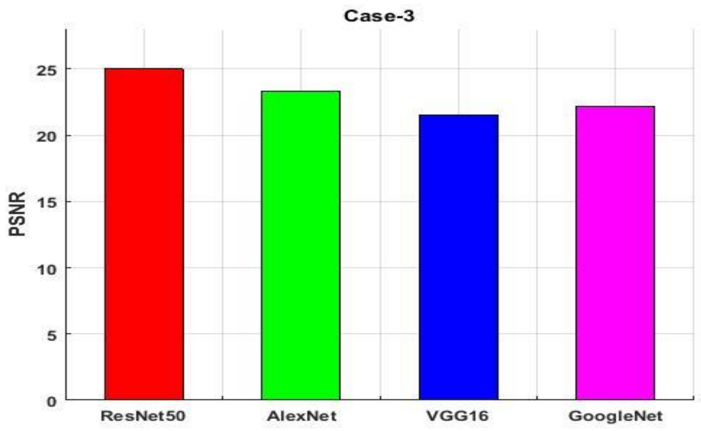


Figure: 10 Techniques wise PSNR comparison for case-3

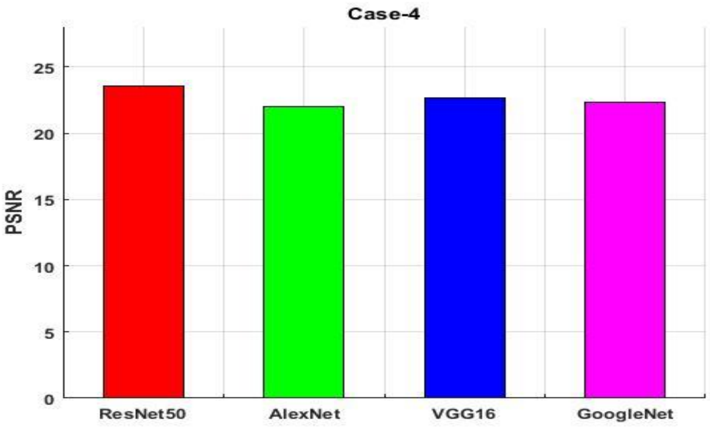


Figure: 11 Techniques wise PSNR comparison for case-4

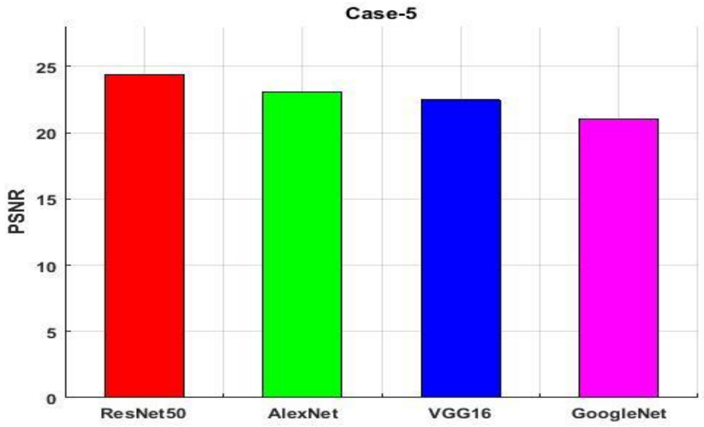


Figure: 12 Techniques wise PSNR comparison for case-5

Performance evaluation through SSIM

The SSIM is a well-known quality metric used to measure the similarity between two images. It was developed by Wang, and is considered to be correlated with the quality perception of the human visual system (HVS). Instead of using traditional error summation methods, the SSIM is designed by modeling any image distortion as a combination of three factors that are loss of correlation, luminance distortion and contrast distortion. The figure 13 to figure 17 shows the performance of employed techniques with respect to SSIM. In case-1, the proposed ResNet-50 attains 0.8952 that is 0.0093 greater SSIM than AlexNet, 0.03 better than and 0.0239 SSIM better than GoogleNet. Similarly, in remaining validation the proposed network ResNet-50 attains superior performance over other techniques. The average performance of ResNet-50 with respect to SSIM is 0.9018.

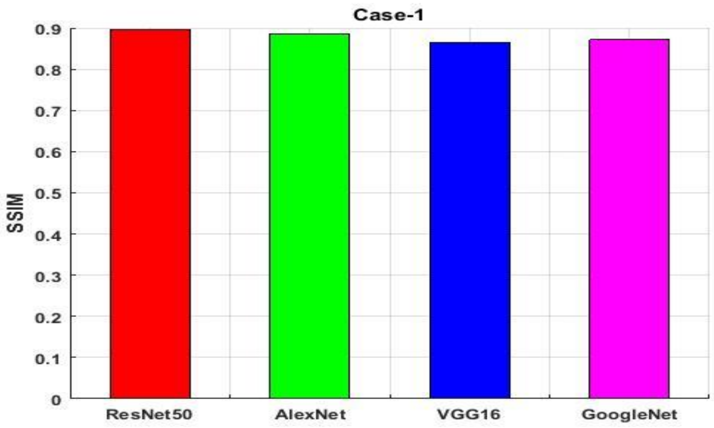


Figure: 13 Techniques wise SSIM comparison for case-1

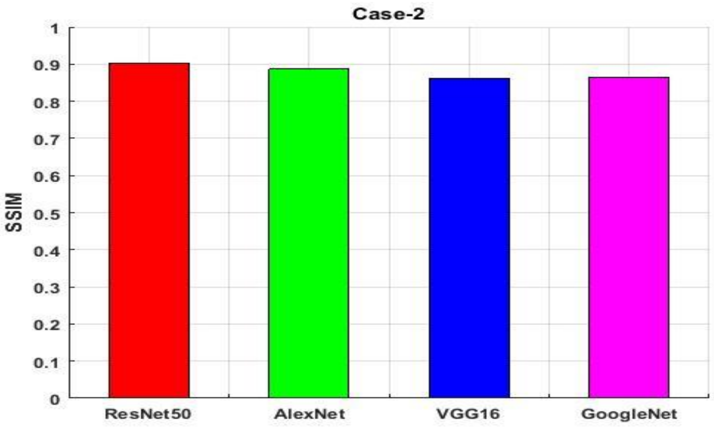


Figure: 14 Techniques wise SSIM comparison for case-2

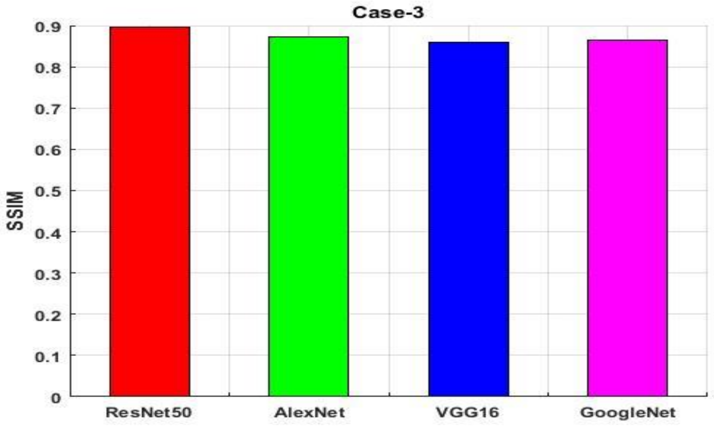


Figure: 15 Techniques wise SSIM comparison for case-3

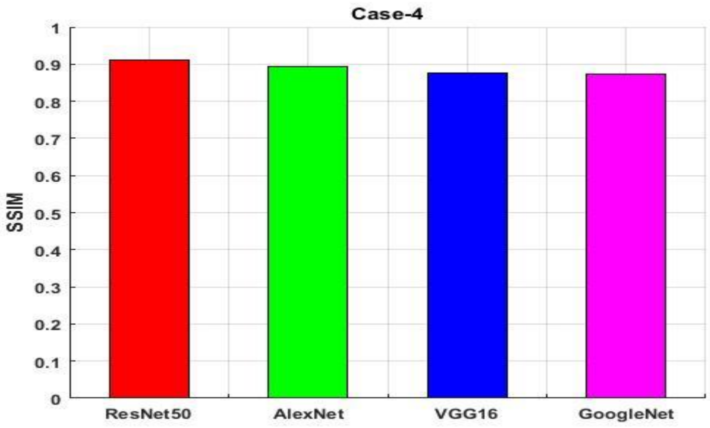


Figure: 16 Techniques wise SSIM comparison for case-4

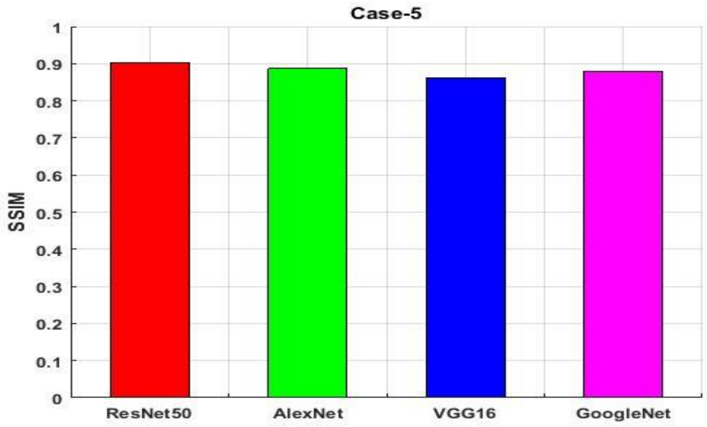


Figure: 17 Techniques wise SSIM comparison for case-5

Conclusion

Image inpainting, from traditional to deep learning methods, has achieved immense, and continued, success. The research investigates different approach for inpainting tasks, datasets, performance evaluation and limitations of the methods. The research identifies the poor performance of traditional methods on images with more extensive binary mask and facial images due to complexity in features on the image. The Research on image inpainting using deep learning has witnessed good progress in recent years. Here, the average performance of proposed ResNet-50 with respect to PSNR is 24.64db that is 1.39db greater than AlexNet, 2.38db better than VGG16 and 2.41db better than GoogleNet. Whereas, in case of SSIM the average performance of proposed model is 0.90 that is 0.02 SSIM better than AlexNet, 0.04 SSIM greater than VGG-16 and 0.03 greater than GoogleNet.

References

[1]Abbad, A., Elharrouss, O., Abbad, K., Tairi, H., 2018. Application of meemd in post-processing of dimensionality reduction methods for face recognition. Iet Biometrics 8, 59–68.

[2]Akl, A., Yaacoub, C., Donias, M., Da Costa, J.P., Germain, C., 2018. A survey of exemplar-based texture synthesis methods. Computer Vision and Image Understanding 172, 12–24.

[3] S. Masnou and J. Morel, “Level-lines based disocclusion,” in Proc. IEEE Int. Conf. Image Processing (ICIP), Chicago, IL, Oct. 1998, vol. 3, pp. 259–263.

[4] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles, “Image inpainting,” in Proc. ACM SIGGRAPH, July 2000, pp. 417–424

[5] Zhu, Qingsong, Ling Shao, Xuelong Li, and Lei Wang. "Targeting accurate object extraction from an image: A comprehensive study of natural image matting." *IEEE transactions on neural networks and learning systems* 26, no. 2 (2014): 185-207.

[6] He, Linyuan, Jizhong Zhao, Nanning Zheng, and Duyan Bi. "Haze removal using the difference-structure-preservation prior." *IEEE Transactions on Image Processing* 26, no. 3 (2016): 1063-1075.

[7] Bhilavade, Milind B., Meenakshi R. Patil, Lalita S. Admuthe, and K. S. Shivaprakasha. "Review on Implementation of Fingerprint Verification System Using Image Inpainting." In *Advances in Communication, Signal Processing, VLSI, and Embedded Systems*, pp. 325-333. Springer, Singapore, 2020.

[8] Wang, Nannan, Xinbo Gao, Dacheng Tao, Heng Yang, and Xuelong Li. "Facial feature point detection: A comprehensive survey." *Neurocomputing* 275 (2018): 50-65.

[9] Shivaranjani, S., and R. Priyadharsini. "A survey on inpainting techniques." In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 2934-2937. IEEE, 2016.

[10] Jiang, Jianmin, Hossam M. Kasem, and Kwok-Wai Hung. "Robust image completion via deep feature transformations." *IEEE Access* 7 (2019): 113916-113930.

[11] A. Efros and T. Leung, “Texture synthesis by non-parametric sampling,” in Proc. Int. Conf. Computer Vision (ICCV), Sept. 1999, pp. 1033–1038.

[12] S. Di Zenzo, “A note on the gradient of a multi-image,” Comput. Vision, Graphics, Image Process., vol. 33, no. 1, pp. 116–125, Jan. 1986.

[13] Wen, Long, Xinyu Li, and Liang Gao. "A transfer convolutional neural network for fault diagnosis based on ResNet-50." *Neural Computing and Applications* (2019): 1-14.

[14] Guillemot, Christine, and Olivier Le Meur. "Image inpainting: Overview and recent advances." *IEEE signal processing magazine* 31, no. 1 (2013): 127-144.

[15] Castellacci, Fulvio, and Vegard Tveito. "Internet use and well-being: A survey and a theoretical framework." *Research policy* 47, no. 1 (2018): 308-325.

[16] Srinivas, S.; Sarvadevabhatla, R.K.; Mopuri, K.R.; Prabhu, N.; Kruthiventi, S.S.S.; Babu, R.V. Taxonomy of Deep Convolutional Neural Nets for Computer Vision. Front. Robot. AI 2016.

[17] Cire¸sAn, D.; Meier, U.; Masci, J.; Schmidhuber, J. Multi-column deep neural network for traffic sign classification. Neural Netw. 2012, 32, 333–338.

[18] Morawski, M.; Słota, A.; Zaja˛c, J.; Malec, M.; Krupa, K. Hardware and low-level control of biomimetic underwater vehicle designed to perform ISR tasks. J. Mar. Eng. Technol. 2017, 16, 227–237,

[19] He, L., Xing, Y., Xia, K., Tan, J., 2018. An adaptive image inpainting method based on continued fractions interpolation. Discrete Dynamics in Nature and Society 2018.

[20] Ghorai, M., Mandal, S., Chanda, B., 2018. A group-based image inpainting using patch refinement in MRF framework. IEEE Transactions on Image Processing 27, 556–567.

[21] Wang, H., Jiang, L., Liang, R., Li, X.X., 2017. Exemplar-based image inpainting using structure consistent patch matching. Neurocomputing 269, 90–96.

[22] Sridevi, G., Kumar, S.S., 2019. Image inpainting based on fractional-order nonlinear diffusion for image reconstruction. Circuits, Systems, and Signal Processing 38, 3802–3817.

[23] Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution, in: European conference on computer vision, Springer. pp. 694–711.

[24] Song, Y., Yang, C., Shen, Y., Wang, P., Huang, Q., Kuo, C.C.J., 2018. Spg-net: Segmentation prediction and guidance network for image inpainting. arXiv preprint arXiv:1805.03356.

Disclaimer

The core objective of the working paper series of BASE University is to help faculty members and research scholars to share their research findings at the pre-publication stage. The working paper series seek to stimulate discussions on the topic and suggestions are welcome. The opinion/views shared here are those of authors.

Contact: Bengaluru Dr. B. R. Ambedkar School of Economics University (BASE University)
Jnana Bharathi Main Road, Teachers Colony
Landmark- Opposite National Law School of India University
Bengaluru, Karnataka – 560072
Email: library@base.ac.in