JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

OBJECT CONSTRUCTION USING DEEP LEARNING

[1]Ravi Kumar, [2]Vignesh T S, [3]Ajit Jain, [4]Shubham Sharma, [5]Saritha A N
[1,2,3,4]Undergraduate Students, [5]Assistant Professor
Computer Science and Engineering
B.M.S. College of Engineering, Bengaluru, India

Abstract *- Corrupted images are very common in people's daily lives. A lot of factors are the culprit for making an image gets corrupted. Images, when subjected to a different environment such as high temperature, watery environment, corrosive environment, etc, faces change in the hardware properties and thus can change the data stored on it. Sometimes the hard copy of the images loses its color and becomes dull due to exposure to a hot and humid climate. The dust and other pollutants also affect the quality of the images. One of the latest emerging technologies is used to achieve the purpose. Generative Adversarial Networks have gained more popularity due to their ability to give the machine the power to imagine and create new objects. The trained model can be used to complete the missing details in the image. One of the main problems with this technology is that the GANs are very unstable to train. Some of the main issues include vanishing gradient problems and mode collapse.*

Index Terms *- GAN model, vanishing gradient, mode collapse, convolutional networks, Wasserstein GAN.*

I. INTRODUCTION

Advancement in technology has given us a new field to discover, which includes training a machine to act like a human. Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are being explored and widely used to a solve variety of problems – health care, drug discovery, weather forecasting, film industry, and many more. Here, in this work, one of the technologies, Deep Learning is used to solve one of the most common problems of image data loss. Many times, the image which the model capture and store lose color and brightness, which makes it unfit for any application. Though sometimes, a loss of a small amount is acceptable to the general public, it is not acceptable to professional – photographers and designers. The proposed model, initially, is trained on lots of real images and later used to generate new images from the knowledge that the model has learned during the training and the imagination gained. The proposed GAN model consists of a generator and a discriminator, where each of them is working against each other and thereby improving each other. Once the model is ready, it can be used to generate the missing part of the images or to create a new image.

The demand for new designs in the fashion industry and fabric industries is increasing day by day. There is a requirement for new technology which can generate new images and designs at low cost and minimum effort to meet the increasing demand. The present technology is very good at enhancing the quality of the images but are not able to produce new and unseen one. The cartoon movies industry is finding it difficult to generate new characters to satisfy the increasing number of viewers. Professional photographers and designers are struggling with dull and corrupted images. There is always a need for improving video quality by filling intermediately generated images. So, understanding the current need of the industry, this current work is based on Deep Learning and GAN, which can not only help in constructing new images and design but also help in gaining back the loss from the corrupted images.

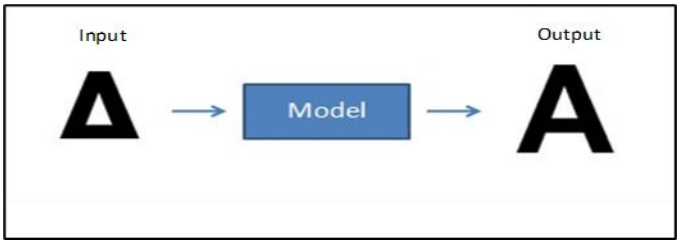

Fig 1. The objective of the model

The main objective of this work is to complete a given incomplete image by feeding extra data, generated smartly, to make the image look complete and feel original. The extra data which are going to generate is taken from the given context and the previous

history. Using the surrounding neighbor pixels' information, the model is going to generate the most suitable substitute, and at the same time use the previous knowledge of similar surroundings to create the most appropriate filling. The objective is to achieve high accuracy with minimum effort in terms of time and space cost. The image generated can be further passed to off-the-shelf traditional image enhancement techniques to further improve the accuracy. Fig1 shows the objective clearly.

The current work is aimed at utilizing the current advancement in technology to achieve a model which when given an image with missing data, tries to complete with meaningful data. The model takes an incomplete image and extra information of the area where the image is incomplete as input and produces an image with filled details as output. The model, in this work, is going to be trained on the dataset of the English alphabet. Since the current work is on a limited dataset due to infrastructure limitations, the model is going to work on English alphabets only. This does not mean that the proposed model is not scalable. Instead, when the model is trained on the larger and diverse dataset, it can work equally well as it is expected to work on the current dataset.

Image generation and completion is a very widely explored area. A lot of research has been carried out over the past many years to recover lost data from an image. Software like Adobe Photoshop uses advanced techniques to figure out the missing details and fill those areas with appropriate details. Auto-encoders are one of the latest technologies which are subjected to research due to their ability to learn input information efficiently and later reconstruct the input from the compressed form. Intermediate frame generation between the two successive frames of a video is generally done by interpolation techniques – linear interpolation, piecewise constant interpolation, block matching algorithm, and many more. All these techniques do not utilize the full capability of the machines which one can think of. Deep Learning and GAN have changed the way machine work by giving them the ability to learn and think. There are many types of GAN like DCGAN, Info GAN, Big GAN, and WGAN, which are widely used to train and utilize for image generation and completion. This work tries to accumulate the various advantages of GANs and develop a model which has high accuracy and is more efficient.

II. LITERATURE SURVEY

Progressive inpainting was proposed by Yizhen Chen and Haifeng Hu which was based on generating models. In this paper with the trained networks, they have used backpropagation to get the most suitable input distribution and then used the generator to fix the corrupted image. They undertook the pyramid strategy to repair the image instead of repairing it in one step. However, the proposed solution failed to fix the corrupted images having large, corrupted regions. [1]

The work in the paper "High-resolution image inpainting using Multi-scale neural patch synthesis" is proposed by Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, and Hao Li. To maintain contextual structures and generate high-frequency details, this model which is based on joint optimization of the image matches the most similar mid-layer feature. To improve the authenticity of the texture information, the author has divided the high-resolution image into many steps. This approach produces sharper results, but this method failed in ensuring global consistency when the images were complimented, and this approach also introduces discontinuity when the scene is complicated. [2]

The work in the paper "Image Completion using Planar Structure Guidance" is proposed by Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf. Their method first evaluates planar projection parameters, segments the known region into planes, and then determines the translational regularity used within these planes. For the completion algorithm, the information is converted into soft constraints. This approach uses a simple algorithm that might fail to detect vanishing points and may not give satisfactory results. [3]

The patch matching algorithm proposed by Connelly Barnes and Eli performed well if the image has adequate data for completion. They have offered a conceptual analysis of the convergence properties of the algorithm and have also given practical and empirical evidence for its quality and performance. The proposed solution could not work as desired when the loss of data is huge. [4]

The paper "Fragment-based image completion" by Iddo Drori, Daniel Cohen-Or, and Hezy Yeshurun use an approach that is based on the concepts of figural simplicity and familiarity. As a result, in the low confidence zones, an approximation is constructed by using a basic smoothing method. After that, known features from a location with higher confidence are added to the approximated region. At the image fragment level, all of these operations are carried out. A pixel's neighborhood is specified by a fragment. The scale of the underlying structure is reflected in the size of the neighborhood, which is defined as ad hoc. The picture completion process is multi-scale, starting with a low-resolution image and then using the results as a coarse approximation to the finer level. Neighborhoods are grouped into level sets ranging from high to low confidence for each scale. Their paper takes an example-based approach to complete the image. As a result, the richness of the accessible pieces has a direct impact on their performance. The method is a 2D image-based strategy. It has no understanding of the image's underlying 3D structure. There is no distinction between figure and ground in the picture completion method. Because both the figure and the backdrop can be synthesized by example, this creates a limitation for completion when the inverse matte is on the boundary of a figure. Also, issues in which the missing area covers the junction of two perpendicular zones are not handled by this method. [6]

The "Fields of experts (FoE)" paper by Stefan Roth and Micheal J Black is using a neighborhood system that connects all nodes in an m × m square region, the research defines a high-order Markov random field for whole images. The model learns the filters with the other parameters, which necessitates the use of parametric expert functions, which are less versatile. Because the FoE is probabilistic, they can learn a general prior that can be applied in a variety of applications. The size of the images in the training data is chosen significantly greater than the clique size to appropriately capture the spatial interdependence of neighboring cliques (or equivalently the overlapping image patches). They trained on image regions that are 3 to 5 times the width and height of the maximal cliques as a trade-off; for example, if there are 5 x 5 cliques, they train on 15 x 15 images. The training data consists of 20000 picture sections that were randomly cropped from the Berkeley Segmentation Benchmark images. Rather than using the

complete dataset for each iteration of the contrastive divergence technique, they divided the data into 200-image "mini-batches" and used just the data from one batch at a time. The so-called stochastic gradient ascending approach greatly accelerated learning. One of the limitations is that the proposed framework can only describe images at their original spatial size (resolution) and cannot model the scale invariance property of natural images. Its 5 x 5 filters, in particular, are too small to catch statistics at very granular geographical scales, yet computational constraints prevent them from increasing their size. Furthermore, their results with 7 x 7 filters show that just increasing the size of the filters does not always assist. FoE has the effect of making reasonably smooth sections even smoother; nevertheless, noise in more textured areas is not completely eradicated. [7]

The "Globally and Locally Consistent Image Completion" paper by Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa use Convolutional Neural Networks as a foundation of their technique. These types of neural networks use convolution operators to preserve the spatial structure of the input, which is typically images. These networks are made up of layers in which an input map is convoluted with a bank of filters to produce an output map, which is then processed further with a non-linear activation function, most often the Rectified Linear Unit (ReLU). Instead of utilizing only standard convolutional layers, they adopt a version known as dilated convolutional layers, which allows each layer to use a larger input region. By spreading the convolution kernel across the input map, this is accomplished without increasing the number of learnable weights. Although the model can handle images of any size with arbitrary holes, due to the model's spatial support, considerably big holes cannot be filled in. This restriction only applies to square masks; for example, large regions can be completed if they are not excessively tall. [8]

The "Semantic Image Completion and Enhancement using Deep Learning" by Vaishnav Chandak, Priyansh Saxena, Manisha Pattanaik, and Gaurav Kaushal uses the mentioned procedure for their model. To execute and evaluate the constructed model, the initial step is to preprocess data from the dataset. To fill the missing pixels in the image, the second stage involves creating Wasserstein GAN-based model. After the completion, the GAN returns a complete image with parts of it being blurry. In the paper, the WGAN is trained to construct missing parts in an image, and then the completed image is put through an improvement network to eliminate blur and unwanted noise. The network's general structure is poorly optimal, and the network's capacity to absorb minute aspects of the image is limited. [10]

Using the idea of predicting the pixels in a context, the authors have proposed an unsupervised algorithm for learning visual features. They have used "Context Encoders" in which a trained CNN tries to generate missing data based on surrounding information. The pre-training of context encoders, to achieve significant improvement from the baseline, is difficult with the existing methods. Moreover, the nearest neighbors, which have been used to evaluate quality, do not perform well in this context. Sometimes the images get blurrier. [11]

In the paper "Painting completion with generative translation models", a new method for image completion has been proposed which is quite different from the traditional method based on contextual information. By learning representation professionally in the data available for training, their model can predict missing features in corrupted images. Their method succeeded in art restoration and creation to a greater extent. However, the PSNR value for the images used is not quite satisfactory. Moreover, the input image size is low and the complexity of the structure is very less. It cannot handle high-definition images and images which have large missing data. The model worked well only on the same succeeded data. Instead of considering both the local and the external features of the data, their work is mainly focused on local data features. [12]

In the work "Generative Face Completion", the authors have proposed a network for completing faces which is a deep generative network. They have used GAN in the network, generator for the autoencoder, local loss function, and global loss function, and discriminators have been used for semantic regularization. From some random noise, their proposed model can generate semantically valid content. It can also synthesize the main parts of the faces. Their model is flexible in handling different sizes and shapes of masks and occlusions. However, challenges like handling unaligned faces are being faced even though they use data augmentation for improved learning. The unpleasant image generated in some cases means that the model could recognize some of the orientation and the positions of the components present in the images. Their model even fails to use the spatial correlations which exist among the neighboring pixels. Generating a properly aligned eye for an unaligned face or predicting the lip's color is the context where this model fails. [14]

The work in "Gesture Recognition Based on CNN and DCGAN for Calculation and Text Output" proposes an algorithm that is based on the convolutional neural network and deep convolution GAN for gesture recognition. Some real data sets have been tested by this model and it performs well. The actual gesture's meaning can be recognized to a greater extent by using their proposed model. The problem of over-fitting, when the samples are less, has been solved by DCGAN. However, calculation and text output are only supported by their networks. The model supports a smaller number of gestures and thus it needs to be evaluated on a greater number of gestures. Gestures in different lightning conditions also need further evaluation and testing. The authors, themselves claim that they need further improvement by adding more training data and restructuring the network. [15]

In the paper "Towards the Automatic Anime Characters Creation with Generative Adversarial Networks", the proposed model can generate high-quality anime faces at a promising rate of success. Their contribution can be described as three-fold: A clean dataset, collected from a suitable GAN model and the approach for training GAN using images without labels. They have generated almost real images. However, the final resolution still needs improvements. Other challenge includes training and improving the GAN where class labels are not evenly distributed in the dataset.[16]

The paper "Appearance and Pose-Conditioned Human Image Generation using Deformable GANs" by Aliaksandr Siarohin, S. L, discussed the problem of generating person images based on pose and appearance information. While keeping the visual details preserved, this model can extract the pose from the given image and construct a target pose. The misalignments because of the difference in poses are being handled by the deformable skip connections which they have introduced in their model. However, the proposed approach cannot preserve the specific details of the image. [17]

The "Mask-specific inpainting with deep neural networks" by Khler, Rolf, Christian Schuler, Bernhard Schlkopf, and Stefan Harmeling, has proposed a system where they directly learn a mapping from image patches, corrupted by missing pixels, onto complete image patches. It has been represented through a deep neural network that trains on large image data sets. They showed that training with such extra information is useful for blind inpainting. Though this method does not perform well with the mask as an input. The inpainting results also get blurry when it's too large. [19]

The "OpenFace: A general-purpose face recognition library with mobile applications" by Amos, Brandon, Bartosz Ludwiczuk, and Mahadev Satyanarayanan, the paper presents the OpenFace face recognition library that bridges the accuracy gap. They showed that OpenFace provides near-human accuracy on the Labeled Faces in the Wilds benchmark. It is intended to maintain OpenFace is a library that stays updated with the latest deep neural network architectures and technologies for face recognition. [20]

III. PROPOSED MODEL

The proposed model does not just do the mere filing of the missing details in the incomplete image, but also provides the various possible outcomes which can be used as substitutes. The output produced, apart from the completed image, also shows the various parameters, which when modified can generate similar real-world images. The GAN network, consisting of the Generator and Discriminator, is a neural network that combines the benefits of both the DCGAN and WGAN, to avoid the problem of vanishing gradient and mode collapse. This model uses both contextual information and perceptual information to generate more real samples. Unlike the existing models, which mainly focus on constructing the center portion of the images and fail to perform well at the edges, this model works equally well at all the portions of the image.

IV. CONCLUSION

Seeing the increase in demand for image quality improvement and new design requirements, using one of the latest technologies, this work tries to cover these issues. Artificial Intelligence (AI) and Deep Learning have been used to provide a solution to the problem of dull, damaged, occluded, and missing details of an image. The GAN model used in this work tries to incorporate some of the best features from the various GAN model proposed, which are under continuous research. Since the model uses the advantages of other GANs and overcomes the disadvantages of other GANs, the result produced is going to be with good accuracy and low loss. Upon successful training over multiple datasets, this model can be integrated with other software.

V. REFERENCES

[1] Yizhen Chen and Haifeng Hu, "An improved method for semantic image inpainting with GANs": Progressive inpainting," Neural Processing Letters, Springer, pp. 1–13, Jun 2018.

[2] Yang, C.; Lu, X.; Lin, Z.; Shechtman, E.; Wang, O.; Li, H. "High resolution image inpainting using multi-scale neural patch synthesis".InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21 July–26 July 2017.

[3] Jia-Bin Huang and Ahuja, "Image completion using planar structure guidance," ACM Transactions on Graphics (Proceedings of SIGGRAPH), vol. 33(4), August 2014.

[4] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," ACM Transactions on Graphics (Proc. SIGGRAPH), vol. 28(3), August 2009.

[5] Pouget-Abadie, I.J.G.; Mirza, M. "Generative adversarial nets". arXiv, arXiv:1406.2661, 2014.

[6] Drori, I.; Cohen-Or, D.; Yeshurun, H. "Fragment-based image completion". ACM Trans. Graph, 22, 303–312, 2003.

[7] Roth, S.; Black, M.J. "Fields of experts". Int. J. Comput. International Journal of Computer Vision,82, 205–229 Vis. 2009.

[8] Iizuka, S. Simo-serra, E. Ishikawa, H. Globally and Locally Consistent Image Completion. ACM Trans. Graph, 36, 107, 2017

[9] Guoping Zhao, Jiajun Liu, Jiacheng Jiang, and Weiying Wang, "A deep cascade of neural networks for image inpainting, deblurring, and denoising," Multimedia Tools and Applications, vol. 77(22), pp. 29589–29604, Nov 2018.

[10] Vaishnav Chandan, Priyansh Saxena, Manisha PattanaikGaurav Kaushal, "Semantic Image Completion and Enhancement using Deep Learning". arXiv:1911.02222v2 [eess.IV] 5 Jan 2020.

[11] Deepak Pathak and Philipp Krahenbuhl, "Context encoders: feature learning by inpainting," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544, 2016.

[12] Ruijun Liu, Rui Yang, Shanxi Li, Yuqian Shi, and Xin Jin, "Painting completion with generative translation models", Multimedia Tools and Applications, Springer, pp. 1–14, 2018.

[13] Andrew Brock, J. D., "Large scale GAN training for high fidelity natural image synthesis", ICLR, arXiv:1809.11096v2 [cs.LG] 25 Feb 2019.

[14] Yijun Li, S. L.-H. "Generative Face Completion",arXiv:1704.05838v1 [cs.CV] 19 Apr 2017.

[15] WEI FANG1, Y. D. (2019). Gesture Recognition Based on CNN and DCGAN for Calculation and Text Output, IEEE, pp. 2169-3536, vol. 7, 2019.

[16] YanghuaJin, J. Z., "Towards the Automatic Anime Characters Creation with Generative Adversarial Networks", arXiv:1708.05509v1 [cs.CV] 18 Aug 2017.

[17] AliaksandrSiarohin, S. L., "Appearance and Pose-Conditioned Human Image Generation using Deformable GANs", JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015.

[18] Yan, Bo, Yiqi Gao, Kairan Sun, and Bo Yang. Efficient seam carving for object removal. In Image Processing (ICIP), 2013 20th IEEE International Conference on, pp. 1331-1335. IEEE, 2013.

[19] Khler, Rolf, Christian Schuler, Bernhard Schlkopf, and Stefan Harmeling. Mask-specific inpainting with deep neural networks. In German Conference on Pattern Recognition, pp. 523-534. Springer International Publishing, 2014.

[20] Amos, Brandon, Bartosz Ludwiczuk, and Mahadev Satyanarayanan. OpenFace: "A general-purpose face recognition library with mobile applications". Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.