Low Quality Facial Image Restoration using Generative Models (Literature Review)

Yamini Karpe
School of Computing (Dublin City University)
Masters in Computing (Data Analytics)
Dublin, Ireland
yamini.karpe2@mail.dcu.ie

Christian Dave Cobalida
School of Computing (Dublin City University)
Masters in Computing (Artificial Intelligence)
Dublin, Ireland
christian.cobalida2@mail.dcu.ie

***Abstract*—The main objective of the proposed practicum is to reconstruct the missing part of an image by learning feature representations from a set of images. The main objective is achieved by training a CNN with the missing part and making it possible to predict what is in the missing part using its features. Given a corrupted image with parts of an image missing or distorted, our objective is to provide a seamless and plausible replacement for a random region of pixels in the image with the help of available visual data and training the model to regress to the missing pixel values.**

***Index Terms*—Image Restoration, GAN (Generative Adversarial Network), in-panting**

I. Introduction

The focus of the proposed project is the inpainting of monocular images using GANs and contextual analysis. We will use image inpainting to make images visually realistic and semantically correct by synthesising missing areas. Traditionally, image inpainting has been accomplished by copy and pasting to fill the missing section in existing datasets. However, this method is time- consuming and does not generalise well to unseen images.

Due to the use of CNNs in image processing and the advances in deep learning, almost all image inpainting is performed using a combination of GANs and CNNs as they can learn high-level semantic features. Decoding and encoding are two separate stages with either a single stage or multiple stages capturing the image context, and the decoder evaluates the images produced by the encoder. In order to further enhance the visual quality of the generated images, some of them use conditional GANs in order to limit the latent space of the image generation. For more complex images with larger portions missing, in-painting has been successful with both regular and irregular holes.

II. Literature Review

The motivations and guiding principles for learning algorithms for deep architectures are discussed by the researchers [1] in this paper, particularly those that leverage unsupervised learning of single-layer models like Restricted Boltzmann Machines as building blocks to create deeper models like Deep Belief Networks.

A multilayer neural network with a thin core layer can be trained to reconstruct high-dimensional input vectors in order to convert high-dimensional data to low-dimensional codes. In these ”autoencoder” networks, gradient descent can be used to fine-tune the weights, although this only works well if the beginning weights are near to a reasonable solution. Researchers in this paper[2], present an efficient initialization method for deep autoencoder networks that enables them to learn low dimensional codes that perform significantly better than principal components analysis as a tool to reduce the dimensionality of data.

This study [3] presents a fresh technique, inspired by weakly-supervised approaches, to the well-known issue of unsupervised object finding. By evaluating whether the correspondence supplied by the cluster aids in predicting the context around each patch, researchers have demonstrated a method for determining if a cluster of patches displays a coherent visual notion. Due to a lack of valid training data, our technique does not use traditional discriminative learning algorithms; nevertheless, in a supervised situation, their thing model might be substituted with a far more straightforward regression model that has been trained discriminatively.

In this paper [4], researchers aim to close the achievement gap between CNNs for supervised learning and unsupervised learning. They present the deep convolutional generative adversarial networks (DCGANs), a class of CNNs with certain architectural restrictions, and show that they are an excellent choice for unsupervised learning. They provide convincing proof that their deep convolutional adversarial pair learns a hierarchy of representations from object pieces to scenes in both the generator and discriminator through training on multiple picture datasets. They also apply the learnt features to brand new tasks to show how versatile they are as general image representations.

A new framework for estimating generative models using an adversarial process has been proposed by researchers [5]. In this framework, two models are simultaneously trained: a discriminative model D and a generative model G that both represent the distribution of the training data. The goal of the training process for G is to increase the likelihood that D will make a mistake. This framework is equivalent to a two-player minimax game. There is just one solution in the space of random functions G and D, where G recovers the training data distribution and D is always equal to 1 /2. Backpropagation can be used to train the entire system when G and D are represented by multilayer perceptrons. Neither training nor sample generation need the use of Markov chains or unrolled approximation inference networks. By analyzing the created samples on a qualitative and quantitative level, experiments show the framework's potential.

In this paper [6], inquisitors introduce a brand-new image completion technique that is supported by a sizable online photo library. The system fills in gaps in photos by locating analogous image sections in the database that are semantically correct and smooth. Their main finding is that, despite the fact that the universe of images is practically endless, the space of semantically distinct scenes is actually somewhat constrained. They are able to locate similar situations that contain image fragments that will convincingly complete the image for numerous image completion jobs. The system is totally data-driven and doesn't need any user annotations or labeling. The algorithm, in contrast to other image completion techniques, can produce a wide range of results for each given image.

III. Deep Learning Architectures

A. GAN based Encoder and Decoder

In this study [7], an autoencoder and a channel-wise fully connected layer are built around an autoencoder. Encoders receive compressed latent representations of images with missing parts as inputs, while decoders reconstruct them. A fully connected layer uses each channel individually to link features from distant spatial locations in feature maps, thereby limiting the number of parameters. Reconstruction loss (which by itself causes blurred images) and adversarial loss make up the loss function. During the evaluation, ImageNet and Paris StreetView datasets were examined, as well as L1 loss (9.37), L2 loss (1.96), and PSNR (18.58dB). The system is not capable of handling high-resolution photos because the pixels are not of excellent texture.

For evaluation, this article [8] adopts a single encoder/decoder design with a single discriminator. While most methods for eliminating reflections from photos rely on several images from various perspectives, the purpose of this research is to develop a system that can effectively eliminate reflections from photographs using conditional GANs. PASCAL VOC 2012 was used as the dataset, but it was used to generate a synthetic dataset because there is no widely available collection of photos with reflections. It is concluded based on the results of this study that the streamlined lightweight design has a maximum Peak Signal to Noise Ratio of 20.2 and a maximum structural similarity score of 0.7301, suggesting that the design may be useful in the proposed practicum.

B. Multiple Discriminators

Numerous discriminators are employed in this study [9], including local and global discriminators. The network creates texture and geometry in the output image based on two inputs, an RGB image and depth information. A local discriminator is then employed to assess the area around the inpainting, while a global discriminator is used to evaluate the entire image. The article employs a strategy that combines derived features from RGB and depth photos to increase the classification accuracy of objects in the images.

In addition to dilated convolution, this research [10] is centred on employing Multi Column CNNs in the generator with varying kernel sizes (3x3,5x5,7x7) to achieve diverse receptive fields. There are two discriminators used (local and global). The loss function is based on the adversarial loss, spatial variant reconstruction loss (weighted L1 loss, with missing pixels near valid pixels having higher weights and pixels in the middle of the missing region having lower weights), and implicit diversified Markov Random Field (ID-MRF) loss (which minimises the difference between generated pixels and the corresponding nearest neighbour to the true pixels from the valid pixels). Paris StreetView, Places2, ImageNet, CelebA, and CelebA-HQ datasets were employed, and the evaluation was based on PSNR (20.16) and SSIM (0.86).

The basis for this research [11] is the use of GANs (Autoencoder as a Generator) for image inpainting. To improve inpainting performance, it employs dilated residual blocks in the generator CNNs as well as Interpolated Convolution. The discriminator network employs the Global Discriminator, which examines the image as a whole to maintain global consistency, as well as Patch-GAN, which examines numerous local image patches to enhance the regional texture of the generated images. The loss function is built around reconstruction and adversarial loss (Global and Patch). The datasets used were Paris Street View, Google Street View, and Google Places. The assessment was based on L1 loss (5.54), L2 loss(1.19), and PSNR (19.03dB), as well as voter perception.

Acknowledgment

We would like to express our deep gratitude to Prof. Hossein Javidnia, our research supervisor, for his patient direction, passionate support, and helpful critiques of this research project.

References

[1] Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends® in Machine Learning 2, no. 1 (2009): 1-127.

[2] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313, no. 5786 (2006): 504-507.

[3] Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. "Context as supervisory signal: Discovering objects with predictable context." In European Conference on Computer Vision, pp. 362-377. Springer, Cham, 2014.

[4] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[5] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial networks." Communications of the ACM 63, no. 11 (2020): 139-144.

[6] Hays, James, and Alexei A. Efros. "Scene completion using millions of photographs." ACM Transactions on Graphics (ToG) 26, no. 3 (2007): 4-es.
[7] Pathak, Deepak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. "Context encoders: Feature learning by inpainting." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2536-2544. 2016.
[8] Heo, Miran, and Yoonsik Choe. "Single-image reflection removal using conditional GANs." In 2019 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1-4. IEEE, 2019.
[9] Fujii, Ryo, Ryo Hachiuma, and Hideo Saito. "RGB-D image inpainting using generative adversarial network with a late fusion approach." In Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part I 7, pp. 440-451. Springer International Publishing, 2020.
[10] Wang, Yi, Xin Tao, Xiaojuan Qi, Xiaoyong Shen, and Jiaya Jia. "Image inpainting via generative multi-column convolutional neural networks." Advances in neural information processing systems 31 (2018).
[11] Demir, Ugur, and Gozde Unal. "Patch-based image inpainting with generative adversarial networks." arXiv preprint arXiv:1803.07422 (2018).