Review of Computer Engineering Research
2020 Vol. 7, No. 2, pp. 86-95.
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/journal.76.2020.72.86.95

AN OVERVIEW OF ADVANCES IN IMAGE COLORIZATION USING COMPUTER VISION AND DEEP LEARNING TECHNIQUES

Rashi Dhir[1]
Meghna Ashok[2]
Shilpa Gite[3+]

[1,2,3]*CS Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India.*
[1]*Email: rashi.dhir@sitpune.edu.in Tel: 9767287628*
[2]*Email: meghna.ashok@sitpune.edu.in Tel: 9789812644*
[3]*Email: shilpa.gite@sitpune.edu.in Tel: +919970001191*

(+ Corresponding author)

Article History
Received: 18 September 2020
Revised: 8 October 2020
Accepted: 27 October 2020
Published: 16 November 2020

Keywords
Deep learning
Computer vision architectures
Image colorization
Generative adversarial networks (GANs)
Neural networks
Machine Learning.

ABSTRACT

Automatic image colorization as a process has been studied extensively over the past 10 years with importance given to its many applications in grayscale image colorization, aged/degraded image restoration etc. In this study, we attempt to trace and consolidate developments made in Image colorization using various computer vision techniques and methodologies, focusing on the emergence and performance of Generative Adversarial Networks (GANs). We talk in depth about GANs and CNNs, namely their structure, functionality and extent of research. Additionally, we explore the advances made in image colorization using other Deep Learning frameworks ranging from LeNets to MobileNets in order of their evolution in detail. We also compare existing published works showcasing new advancements and possibilities, and predominantly emphasize the importance of continuing research in image colorization. We further analyze and discuss potential applications and challenges of GANs to tackle in the future.

Contribution/Originality: This study attempts to trace and consolidate developments made in Image colorization using various computer vision techniques and methodologies, focusing on the emergence and performance of Generative Adversarial Networks (GANs).

1. INTRODUCTION

Generative Adversarial Networks or 'GAN' was introduced in 2014 by Ian, et al. [1] after which it majorly gained popularity in the community, but Adversarial networks as a concept has been around since 1990 (first explored by Jürgen Schmidhuber). Currently, GANs are Universal. Data scientists and Machine Learning researchers use this technique to generate photorealistic images, change human facial expressions, create computer game scenarios, visualize perspectives and designs, and more recently, even generate trendsetting artwork.

Computer Vision has been around since the 1950s and is one of the most advanced applications of Artificial Intelligence utilised in today's world. Based on the Human vision system, it has in many ways outperformed humans through detecting and processing objects. The success of Computer Vision is sorely due to the massive amount of data that has been generated in recent years after the advent of Machine Learning. GANS can be used with Computer Vision since GANS is capable of generating data in itself, and can help train and better computer vision. It has many applications including Augmented reality, Facial recognition, self-driving cars and more. With time, more and more organisations are implementing Computer Vision to solve real-world problems.

Applications include Generating Examples for Image Datasets as depicted in the original GANs paper by Ian Goodfellow [2], Generating Photographs of Human Faces [3] which achieved remarkable and indistinguishable results, Generating Realistic Photographs using BigGANs [4] Generating Cartoon Characters [5] to showcase the training and usage of GAN for generating faces of anime characters. Phillip Isola, et al. in their 2016 paper titled "Image-to-Image Translation with Conditional Adversarial Networks" demonstrated GANs, prominently their pix2pix approach for various image translation activities [6]. Other Applications consists of Generation of New Human Poses of Human models [7], Face Aging and De-aging [8], Super Resolution to generate higher pixel resolution images [9] and Video prediction describing the use of GANs for predicting upcoming video frames [10].

2. SURVEY OF EXISTING SYSTEMS BASED ON IMAGE COLORIZATION

2.1. Image Colorization

2.1.1. Past Work

Image Colorization using Computer-assisted Process was brought out by Wilson Markle in 1970, his idea was to colorize the black and white TV programs or movies. In the Computer- assisted process, a colored mask is manually painted for having one reference frame in a shot as a marked image; then motion detection and tracking are appealed on the marked images which allows the colors to be automatically assigned to other surrounds in regions where no motion occurs. It sometimes requires manual fixing.

2.1.2. Present Work

Image Colorization with OpenCV and Deep Learning is used to colorize both black and white images and videos. This method doesn't require manual fixing and requires less human interference. It uses different approaches for image colorization such as thresholding, gradient, region, and classification-based methods for image segmentation, morphological operations, erosion, dilation and contour features for object measurement and neural network methods for classification.

2.1.3. Future Work

DeOldify is a deep learning project that not only colorizes images but also restores them. Recently the project has been updated by introducing a new training technique called NoGAN which is yet to be published. Building on that, an ICML 2019 paper has been published which proposes training a BigGAN [4] quality model with fewer class labels. OpenAI has also unveiled a completely new model called the sparse transformer [11] recently that leverages the transformer architecture for generating images. Additionally, Nvidia has introduced a project called GauGAN [12] which is expected to turn children's scribblings into photorealistic comprehensible masterpieces and has published a promising solution for the same.

2.2. Generative Adversarial Networks

GANs contain a generator that generates data and a discriminator that distinguishes them. The generator is designed to generate images from random noise or mapping latent space and is typically a Deconvolutional neural network. The discriminator tells the generated images apart from the real images and is a Convolutional Neural network. By this, both the generator and discriminator are trained, first individually then simultaneously and constant improvements are made through feedback. The target of the generator is to generate such distributions such that the discriminator believes is real i.e. not generated.

Figure 1 depicts the Generative adversarial network (GAN) Architecture consisting of a Generator and a Discriminator. The Generator generates fake samples of data and tries to fool the Discriminator, whereas the Discriminator tries to distinguish between the real and fake samples. The Generator and the Discriminator are both Neural Networks and they both run in competition with each other in the training phase. The steps are repeated several times and in this, the Generator and Discriminator get better and better in their respective jobs after each repetition. The working can be visualized by the above diagram.

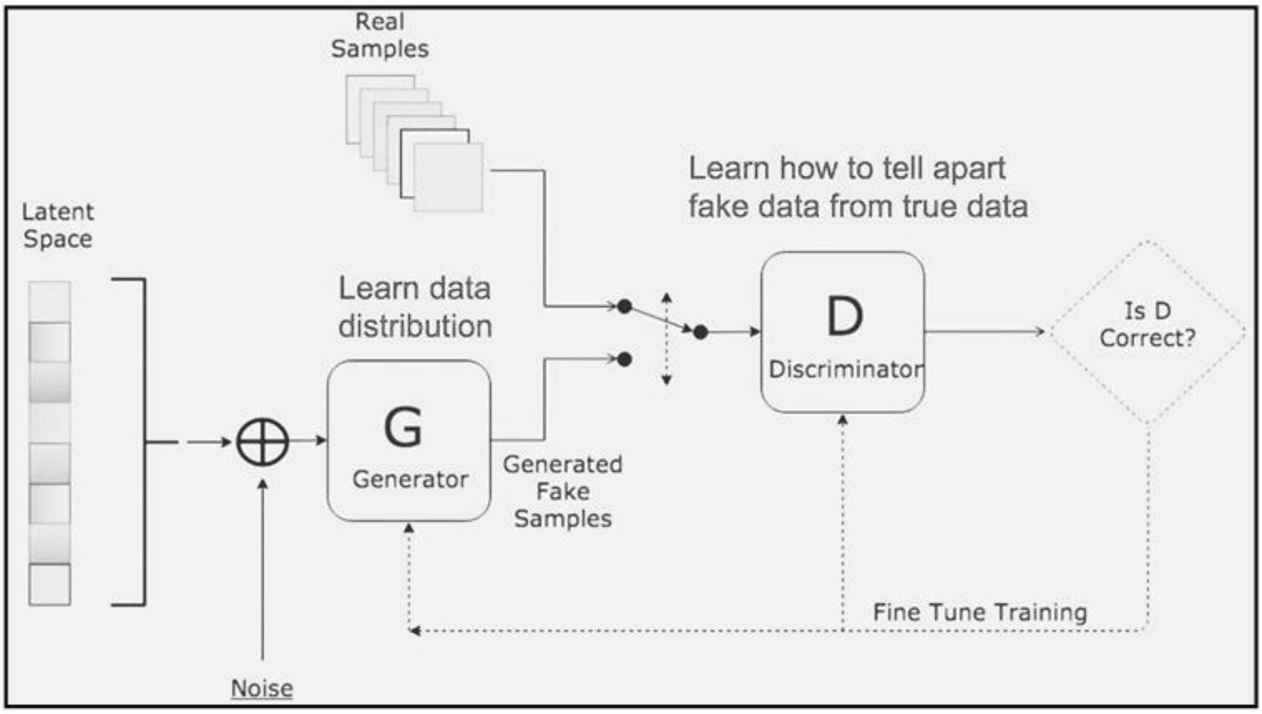

Figure-1. GANs Architecture.

Source: Jan, et al. [13].

2.2.1. Past Work

The original Generative Adversarial Networks paper by Ian, et al. [1] defines the GAN framework and architecture, and discusses its 'non-saturating' loss function. This paper also derives the discriminator, which competed with the generator in the model. The paper also demonstrates the empirical effectiveness of GAN on the CIFAR-10, MNIST, and TFD image datasets. With this as a basis, Conditions GANS was later introduced by Mirza and Osindero [14] in their paper 'Conditional GANS' for integrating data class labels resulting in a more stable GAN training.

In his paper titled 'DCGANS' by Radford, et al. [15] which shows how convolutional layers can be interlaced with GANs. The paper also discusses topics such as Visualizing GAN features, using discriminator features to train classifiers, Latent space interpolation, and evaluation of results.

More and more training methodologies for Generative Adversarial Networks have been introduced over the years, and more applications have been identified and executed since.

2.2.2. Present Work

Image colorization, along with image restoration has been of great interest in the last decade. In Kamyar Naxeri's paper on colorization of grayscale images using a conditional Deep Convolutional Generative Adversarial Network (DCGAN) [16] they attempt to speed up and greatly stabilize the process.

Since GANS is relatively new, various world applications previously tackled using traditional neural networks are being countered by adversarial networks and working on improving results.

2.2.3. Future Work

Since GANS is fairly new, more and more research is being done so that it is more openly accepted by the research community. GANs have so far shown very impressive results on tasks that were difficult to perform using conventional methods. Transformation of low-resolution images to high-resolution images, for example, was previously quite a challenging task and was generally carried out using CNNs. GAN architectures, such as SRGANs [9] or Pix2pix [6] have shown the potential of GANs for this application, while the StackGAN network [17] has proved useful for text-to-image synthesis tasks. Nowadays, anyone can create an SRGAN [9] network and train it on their own images.

2.3. Neural Networks and Deep Learning Architectures

Convolutional Neural Networks is an algorithm that brings together Deep learning as well as Computer vision. Its architecture is analogous to Neurons in the human brain responding to stimuli. CNNs successfully capture spatial and temporal dependencies in images and demonstrate its superiority over feed-forward neural networks. Limitations include inability to detect an object when displayed in different perspectives/angles and disregarding position of objects in an image during detection. CNN Applications are mainly in the field of Computer vision. Its main Architectures are listed below.

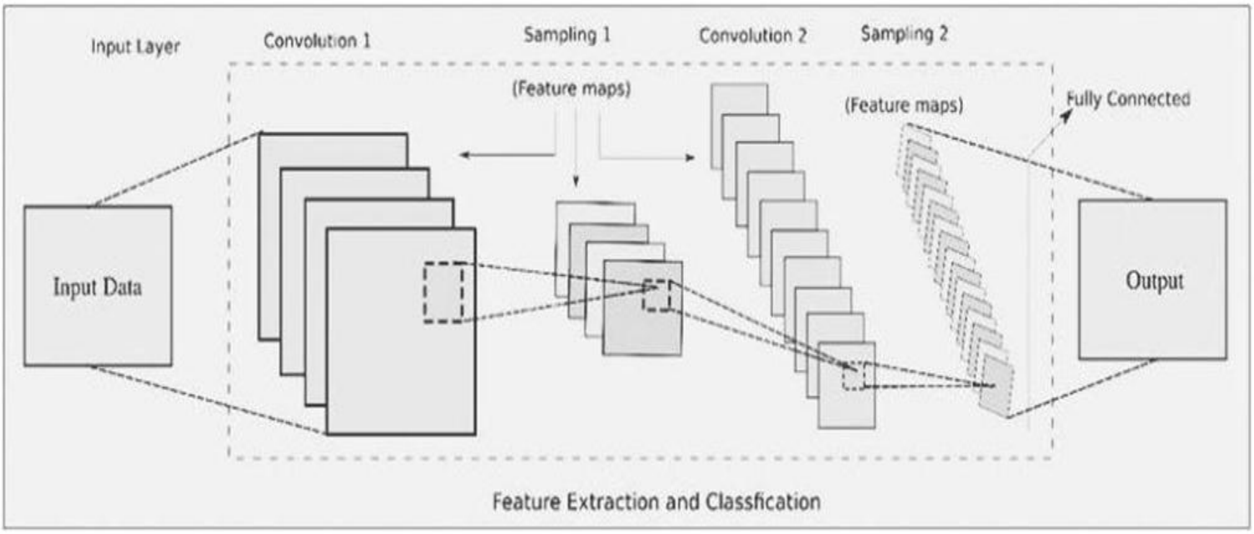

Figure-2. CNN Architecture.

Source: Kratzwald, et al. [18].

Figure 2 depicts a generalized CNN architecture representation, composed of several Convolutional layers. Convolution is a mathematical operation to merge two sets of information. In the above case the convolution is applied on the input data using a *convolution filter* to produce a *feature map*, that predicts the class probabilities for each feature by applying a filter that scans the whole image, few pixels at a time.

A Pooling layer (down sampling)-scales down the amount of information.

the convolutional layer generates for each feature and maintains the most essential information (the process of the convolutional and pooling layers usually repeats several times).

Fully connected input layer-"flattens" the outputs generated by previous layers to turn them into a single vector that can be used as an input for the next layer. Fully connected layer-applies weights over the input generated by the feature analysis to predict an accurate label. Fully connected output layer-generates the final probabilities to determine a class for the image.

2.3.1. LeNet5 [19]

Yann LeCun developed the first network LeNet5 in1994. The LeNet5 has a principal structure, specifically the knowledge that picture highlights are dispersed over the whole picture, and convolutions with simple parameters is an impactful way to separate comparable highlights at various points with lesser parameters.

Salient features are:

- LeNet network uses a sequence of 3 layers: convolution, pooling and non-linearity.
- Spatial average of maps is used to perform subsampling.
- Convolutions are utilized to draw out spatial highlights.
- The nonlinearity layer is shaped up by sigmoid or tanh functions.

2.3.2. Dan Ciresan Net [20]

Dan Claudiu Ciresan along with Jurgen Schmidhuber were the first to deploy neural networks on GPU in 2010. MNIST handwritten digits datasets were used to achieve the results of a very low error of 0.35%. The graphic processor was NVIDIA GTX 280 and worked upon up to nine layers of neural network.

2.3.3. Alex Net [21]

Alex Krizhevsky founded AlexNet in 2012. CNNs that were being used far and wide then, faced difficulties when applied to high resolution images. As a mitigation, AlexNet came into picture. Salient Features are:

- AlexNet Architecture has a total of 8 layers, where 3 are fully connected layers and the rest 5 are convolutional layers.
- AlexNet differed in using ReLu instead of tanh functions for non-linearities.
- It is used as a dropout technique to avoid overfitting, wherein it conditionally ignores single neurons.
- Used GPUs for faster speed.

2.3.4. Overfeat [22]

Yann LeCun in 2013 gave rise to Overfeat architecture in December 2013, which is a derivative of AlexNet. The aim was to significantly improve upon classification/localization and detection. It approaches the concept of localization by proposing to learn bounding boxes.

2.3.5. VGG [23]

Even after the huge success of AlexNet, there was a need for a much better and accurate network. Oxford University responded to this need by coming up with VGG networks. Some of the key points of its architecture are:

- It accepts a 224x224 pixel RGB image as input.
- The initial two layers of the network have 4096 channels each and the last has 1000 channels, one for each class.
- All of the hidden layers in the VGG architecture use ReLu.
- The major breakthrough was to have increased depths and small convolutions filter of dimensions 3 x 3.

2.3.6. Network-in-Network [24]

NiN was introduced with 1x1 convolutions which imparted more power to the convolution layers. A miniaturized scale neural system was constructed with progressively complicated structures to extract the information inside the receptive field. Final classifier of the network used an average pooling layer to average out the output.

2.3.7. ResNet [25]

ResNet came up in 2015. It implements the idea of bypassing the inputs to the next layer, which was also done in some older architectures. ResNets were the first in their league to work with a large magnitude of layers-hundreds and thousands of layers. ResNets promised to simplify the process of network training, which was quite tough previously.

2.3.8. Inception V4 [26]

Inception V4 is an amalgamation of Inception V1, Inception V2, Inception V3.Stakeholders claim that this architecture is computationally viable while achieving significant performance.

2.3.9. SqueezeNet [27]

A newer architecture, SqueezeNet incorporates all the benefits that come along with using a small CNN model. It can be considered a revised version of Inception and ResNet together. It claims to have 50 times fewer parameters along with a model size of up to 0.5Mb.

2.3.10. Xception [28]

Xception is an extension of the Inception series of models. Its architecture is simpler and more compact and its efficiencies are at par with other high-performance architectures like those of Inception and ResNets.

2.3.11. MobileNets [29]

MobileNets is one of the younger architectures, with its latest version that came up in April of 2017. It basically addresses Image processing and classification. It is computationally inexpensive, which makes it a go-to model for embedded systems.

Figure-3. Colourization of Gray-scaled Images.

Source: Zhang, et al. [38].

Figure 3 depicts black and white images successfully colorized using a deep learning model.

3. COMPARISON OF RELEVANT WORKS

Table-1. Comparison of relevant technologies in image processing and colorization.

No.	Paper Details	Content referred	Limitation
1	Generative adversarial nets Ian, et al. [1]	Internal working and concept of GANs	Generator and discriminator must be well synchronised during training in order to avoid "Helvetica scenario" There is no explicit representation of p(x) in the model
2	Conditional Generative Adversarial Nets Liu and Tuzel [2]	How a condition is set on to both the generator and discriminator.	Introductory results shown, but demonstration of the potential of conditional adversarial nets and its applications is depicted.
3	Image Colorization using GANs, Brock, et al. [4]	Structure and architecture of Conditional GANs for image colorization	Many images showed blurred and sepia effects. Mis-colorization was frequently encountered with high texture images. Better quantitative metric needed to measure performance.
4	Progressive Growing of GANs for Improved Quality, Stability, and Variation Park, et al. [12]	Method to simultaneously train generator and discriminator, and increase variation in generated images	Lack of data-set. Many output images showed blurred and sepia effects.
5	Image-to-Image Translation with Conditional Adversarial Networks Pix2pix software Isola, et al. [6]	Foundation for variety of applications implementing CGANS in image-to-image translations	Insufficient size of data-set. Issues in final output and performance.
6	Coupled Generative Adversarial Networks Simonyan and Zisserman [23]	(CoGAN) for learning a joint distribution of images of multiple domains.	Focused on fixed network structures and a single form of social learning.
7	Wasserstein GAN Arjovsky, et al. [30]	An algorithm to improve learning stability and introduces an alternative to traditional GAN training	Issues in understanding real world data problems.
8	Self-Attention Generative Adversarial Networks Zhang, et al. [31] Metaxas, Augustus Odena, 2018	Model implementing attention-driven, long-range dependency modelling for image generation.	Lack of data-set hence, lack of performance/ stability. There is a more theoretical model.
9	Image colorization by fusion of colour transfers based on DFT and variance features, Arjovsky, et al. [30]	Problem of colorization can be carried out in two ways- colour propagation and colour transfer. This paper focuses on colour transfer technique	Inconsistencies arise due to incoherence in spatial arrangement.
10	Emotional image colour transfer via deep learning, Jin, et al. [32]	Mitigates the problem of unnatural colouring taking place in colorizing problems	Training is time consuming.
11	Thermal infrared colorization via conditional generative adversarial network Kuang, et al. [33]	The method uses a composite objective function to produce finely detailed and realistic images.	Encounters poor results with blurry or distorted image details.
12	Optimization based grayscale image colorization Nie, et al. [34]	Optimization approach to colorization that reduces computational time and reduces colour diffusions.	Spatial-temporal approach to be developed for maintaining temporal coherence.
13	Emotional image colour transfer via deep learning Liu, et al. [35]	Considers semantic information of images to solve unnatural colour problems in images	Instability due to limited train and test sets.
14	Context-aware colorization of Gray-scale images utilizing a cycle-consistent generative adversarial network architecture Johari and Behroozi [36]	Parallel colorization models introduced as opposed to traditional single models.	Does not deal with pixel-to-pixel mapping
15	Automatic grayscale image colorization using histogram regression Liu and Zhang [37]	Based on histogram regression and luminance-colour correspondence. No user intervention needed.	Spatial information is not taken into consideration

4. APPLICATION DOMAINS OF IMAGE COLORIZATION

Research in image colorization has been slowly progressing in recent times, which has led to many community members believing that research in this domain should be revoked. It is also believed that colourization of images is an ambiguous and subjective process because of which it has limited applications. On the contrary, the Applications of Image colorization are limitless extending to Interior Designing, Augmented Reality/ Virtual Reality, Forensics, Medical Imaging and National Intelligence.

Automatic colorizations is an area of research that possesses great potential in applications: from black & white photos reconstruction, augmentation of grey scale drawings, video colorization through shot and frame conversion, photograph enhancing and video inpainting to re-colorization of images. Most applications which work in the domain of filter application work on the basis of computer vision and colour detection. Due to the implementation of Artificial Intelligence and Deep Learning in image and filtering applications, even minute human efforts and skill requirements are eliminated. Furthermore, most of the applications are based on the conversion of coloured photographs to black and white and vice versa.

Additionally, Colours also have the capability to drive emotions of a human being. Applications of the use of colour in our daily life are endless. Image colorization benefits User Interface Designers, Interior Designers, Architects and many more users looking for ways to automate their workflow and produce better requirement-based results or plainly improve aesthetics. For instance, UI developers who need to determine the colour combinations that appeal to users.

GANs are continuing to produce breakthrough results in video processing as well, by outperforming the traditional methods that were being employed. In one of the recent researches, Wasserstein GAN frameworks were made use of to in-paint videos. To perform the said task, the model did not have to separate out the background and foreground, thus reducing the overheads and imparting efficiency [18].

5. CONCLUSION AND FUTURE DIRECTIONS

The world of Artificial intelligence and Machine Learning has come a long way since its onset and image processing remains one of the hottest topics of interest and research. Computer Vision is an attempt at simulating the human vision process artificially which can be useful for image and object recognition. It proves to be useful in various surveillance systems, detecting abnormal behaviour through medical images and many more. Convolutional neural networks have been a popular choice for image processing and provide promising results in its real-world applications like those of healthcare and radiology. It is also widely employed for autonomous cars and related purposes. GANs presently are in their early days of research. Although there has been tremendous growth in research, GANs still lack substantial control and have been known to be difficult to train, due to heavily weighing computation. More and more research is being done every day in implementing real world applications of GANS, and its superiority is recognized more and more by global research communities.

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCES

[1] J. G. Ian, P.-A. Jean, M. Mehdi, X. Bing, W.-F. David, O. Sherjil, C. Aaron, and B. Yoshua, "Generative adversarial networks," *arXiv: 1406.2661*, 2014.

[2] M. Y. Liu and O. Tuzel, "Coupled generative adversarial networks," *In Advances in Neural Information Processing Systems*, pp. 469-477, 2016.

[3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196*, 2017.

[4] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *CoRR abs/1809.11096*, 2018.

[5] Y. Jin, J. Zhang, M. Li, Y. Tian, H. Zhu, and Z. Fang, "Towards the automatic anime characters creation with generative adversarial networks," *arXiv preprint arXiv:1708.05509*, 2017.

[6] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125-1134.

[7] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," *In Advances in Neural Information Processing Systems*, pp. 406-416, 2017.

[8] G. Antipov, M. Baccouche, and J. Dugelay, "Face aging with conditional generative adversarial networks," in *2017 IEEE International Conference on Image Processing (ICIP), Beijing*, 2017, pp. 2089-2093.

[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681-4690.

[10] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," *In Advances in Neural Information Processing Systems*, pp. 613-621, 2016.

[11] R. Child, S. Gray, A. Radford, and I. Sutskever, "Generating long sequences with sparse transformers," *arXiv preprint arXiv:1904.10509*, 2019.

[12] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "GauGAN: semantic image synthesis with spatially adaptive normalization," *In ACM SIGGRAPH 2019 Real-Time Live!*, pp. 1-1, 2019.

[13] B. Jan, H. Farman, M. Khan, M. Imran, I. U. Islam, A. Ahmad, S. Ali, and G. Jeon, "Deep learning in big data Analytics: A comparative study," *Computers & Electrical Engineering*, vol. 75, pp. 275-287, 2019.

[14] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014

[15] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[16] K. Nazeri, E. Ng, and M. Ebrahimi, "Image colorization with generative adversarial networks," presented at the In International Conference on Articulated Motion and Deformable Objects. Springer, Cham, 2018.

[17] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5907-5915.

[18] B. Kratzwald, Z. Huang, D. P. Paudel, A. Dinesh, and L. Van Gool, "Improving video generation for multi-functional applications," *arXiv preprint arXiv:1711.11453*, 2017.

[19] Y. Lecun, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *In Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[21] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Computation*, vol. 22, pp. 3207-3220, 2010.

[22] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[24] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for imagerecognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016.

[27] B. Wu, F. Iandola, P. H. Jin, and K. Keutzer, "SqueezeDet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 129-137.

[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251-1258.

[29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[30] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[31] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," presented at the International Conference on Machine Learning. PMLR, 2019.

[32] Z. Jin, L. Min, M. K. Ng, and M. Zheng, "Image colorization by fusion of color transfers based on DFT and variance features," *Computers & Mathematics with Applications*, vol. 77, pp. 2553-2567, 2019.

[33] X. Kuang, J. Zhu, X. Sui, Y. Liu, C. Liu, Q. Chen, and G. Gu, "Thermal infrared colorization via conditional generative adversarial network," *Infrared Physics & Technology*, p. 103338, 2020.

[34] D. Nie, Q. Ma, L. Ma, and S. Xiao, "Optimization based grayscale image colorization," *Pattern Recognition Letters*, vol. 28, pp. 1445-1451, 2007.

[35] D. Liu, Y. Jiang, M. Pei, and S. Liu, "Emotional image color transfer via deep learning," *Pattern Recognition Letters*, vol. 110, pp. 16-22, 2018.

[36] M. M. Johari and H. Behroozi, "Context-aware colorization of gray-scale images utilizing a cycle-consistent generative adversarial network architecture," *Neurocomputing*, 2020.

[37] S. Liu and X. Zhang, "Automatic grayscale image colorization using histogram regression," *Pattern Recognition Letters*, vol. 33, pp. 1673-1681, 2012.

[38] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," presented at the European Conference on Computer Vision. Springer, Cham, 2016.