Assessing Generative Adversarial Networks for Advanced Deepfake Creation Using Network Analysis

Minakshi Arya
Computer Science, NDSU
North Dakota State University
Fargo, USA
minakshi.arya@ndsu.edu

Shubhavi Arya
Computer Science , IU
Indiana University
Bloomington, USA
aryas@iu.edu

Saatvik Arya
Informatics, WU
University of Washington
Seattle, USA
arya3@uw.edu

Abstract— GAN architecture is comprised of a generator model for outputting new plausible synthetic images and a discriminator model that classifies images as authentic (from the dataset) or fake (generated). First, the generator model fools the discriminator model and updates itself via the discriminator model. Then, the discriminator model updates itself directly. Deepfake GAN is applied to image generation, high-resolution image generation, 3D object generation, age estimation, cartoon character / animation / sketches generation, natural language processing, video generation, data augmentation, and detection. However, there are different models to do different tasks. Therefore, we want a combination of the Deepfake GAN models which can carry out the tasks of all the categories. After analysis of the different studies, we concluded that AttGAN and Pix2Pix models are the top models with highest degree centrality and this combination can generate deepfakes in all the categories.

Keywords— Deepfake, GAN, Image Generation, High-Resolution image generation, 3D object generation, Age Estimation, Cartoon Character/Animation/sketches Generation

I. Introduction

Fake content has a harmful impact on its target audience by spreading misinformation and manipulation through various mediums, such as videos, photos, news, reviews, and social media likes [1]. Deepfake-based multimedia content can cause distress, spread hatred, damage reputations, and erode trust in digital content. It can also lead to financial fraud, false allegations, and manipulation of democratic discourse, elections, and public opinion, posing a threat to national security and discouraging international relations and journalism..

Generative Adversarial Networks (GANs), introduced in 2014 [2], are a type of deep neural network that can generate and synthesize deepfake images [3], audio, video, and time series data for medical data and stock market forecasting, visual similarity recommendation [4] and statistical inference in designing clothes and shoes by analyzing photos of specific classes, text-to-image synthesis, and image super-resolution applications [5].

GAN models consist of two neural networks: a generator that produces solutions from a random dataset and a discriminator that distinguishes between real and fake data. While the generator network generates new data samples, the discriminator network acts as a classifier that distinguishes whether the data given as input is real or fake.

Deepfakes via GANs have gained importance due to their synthetic data generation capabilities and the benefits of representations in various applications. GANs have been used in a variety of applications as a tool for generating deepfakes, including image and video generation, natural language processing, data augmentation, and detection, and have proven to be a powerful tool for producing and distinguishing real data from synthetic data.

Multimedia creation and manipulation techniques have a high degree of realism [6]. Generative deep learning algorithms have progressed to a point where it is difficult to distinguish between real and fake [7]. This opens the door to applications in different fields such as creative arts, advertising, film production, and video games. It also poses enormous security threats. Software packages like Zao, Deepfakes web, RefaceAI, MyHeritage, DeepFaceLab, Deep Art, FaceSwap, and FaceApp are freely available on the web, allowing any individual without special skills to create very realistic fake images and videos [7]. These can be used for various malicious purposes, such as committing fraud, discrediting, blackmailing, and manipulating public opinion during elections [7].

II. METHODOLOGY

A systematic review was conducted to explore the various ways in which Generative Adversarial Networks (GANs) have been utilized to generate deepfakes. This involved surveying published research papers indexed in Academic Search Premier (EBSCO), Web of Science Core Collection, IEEE Xplore, and Scopus databases for Computer Science, with the inclusion criteria being papers published between 2017 and 2022, in English language, and focused on academic journals. The search yielded 318 papers from EBSCOhost Academic Search Premier (EBSCO), 109 papers from Scopus, 54 papers from Science direct, 739 papers from IEEE Xplore, and 242 papers from Web of Science. After removing duplicates and applying article inclusion and exclusion criteria, 189 papers were left. The study included papers on deepfakes involving text, images, audio, and videos but excluded medical-related deepfake papers. This study is the first of its kind to use network analysis to gain a deeper understanding of the use of GANs in generating deepfakes. The study provides an overview of deepfakes via Generative Adversarial Network (GAN), reviews the applications of deepfakes in specific categories, calculates the degree centralities of each model, and provides a network for each category. Different GAN models were found to have been developed for various categories of applications. Finally, a combination model of GAN based on degree centralities was presented, which meets the functionality requirements of all categories. This analysis gives an overview of the distribution of GAN models across different categories and highlights the most versatile models that can be used across a wide range of applications. The networks were carried on Mac computer running macOS Catalina equipped with 2.9 GHz dual-core Intel Core i7 processor, the 16 GB 1600MHz of DDR3 RAM, 750 GB SATA disk and the Intel HD Graphics 4000 with 1336 MG graphics memory.

III. RESULTS AND DISCUSSIONS

Generative Adversarial Networks (GANs) used for deep fakes have achieved unprecedented success in image generation, high-resolution image generation, 3D object generation, age estimation, cartoon character /animation/sketches generation, natural language processing, video generation, data augmentation, and detection. So, the papers are divided into various categories: Image generation, High-resolution image generation, 3D object generation, Age estimation, Cartoon character/ animation/ sketches generation, Natural language processing, Video generation, Data augmentation, Detection.

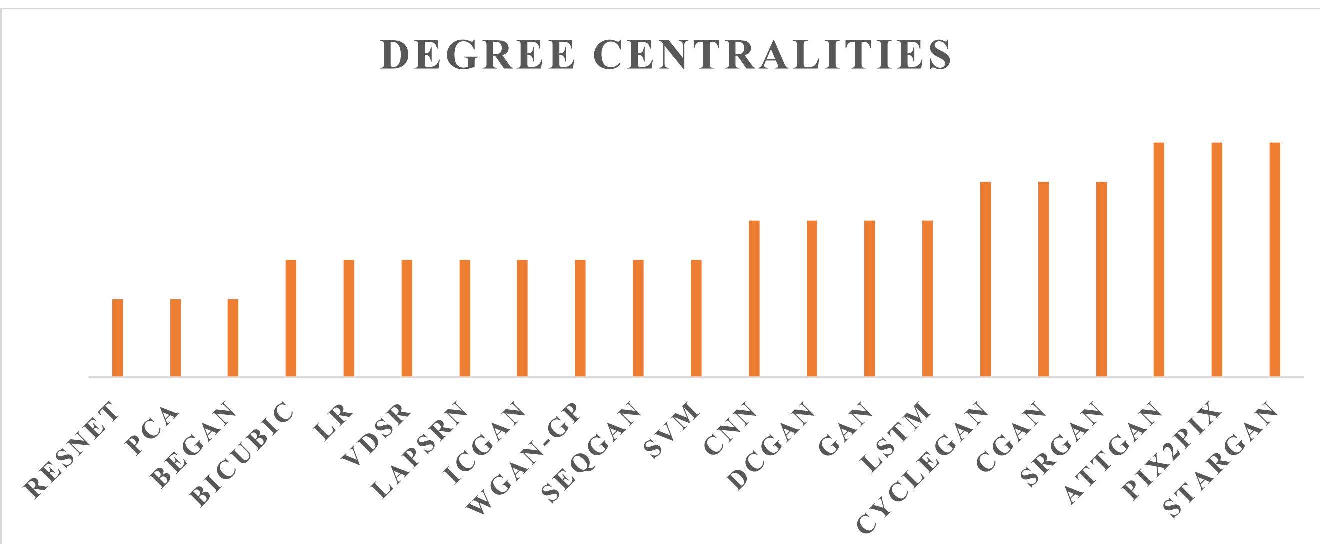

Fig. 1. Degree Centralities

Figure 1. shows the degree centralities of top models. AttGAN, Pix2pix and StarGAN has the highest degree centralities followed by CycleGAN, CGAN, SRGAN. CNN, DCGAN, WGAN-GP, LSTM, and ResNet have lower degree centralities, with ResNet having the lowest value.

The Table 1 provides information on the frequency of GAN models used in different categories of deepfakes and their degree centrality. To perform a network analysis, we can represent each GAN model as a node and draw edges between nodes that are used in the same category of deepfakes. Using this approach, we can identify the GAN models that are most central to the network, as they have the highest number of connections to other models. In this table, AttGAN, Pix2pix, and Star GAN have the highest degree centrality, indicating that they are the most connected GAN models in the network. This suggests that these models are highly versatile and can be used across multiple categories of deepfakes. Furthermore, the table shows that AttGAN and Pix2pix have the highest degree centrality among all the GAN models, which suggests that a combination of these two models can be used to generate deepfakes across all the specified categories. This combination model would likely produce high-quality deepfakes and meet the specific requirements of a wide range of tasks. However, it is important to note that the effectiveness of any approach will depend on the specific requirements of the task at hand and may vary based on the dataset, the quality of the input data, and other factor.

TABLE I. THE MODELS WHICH ARE USED IN MORE THAN THREE DIFFERENT CATEGORIES AND THEIR DEGREE CENTRALITIES

Model	Categories	Frequency of categories	Degree Centrality
AttGAN	1,4,6,7,8,9	6	0.01056338
Pix2pix	1,2,3,5,8,9	6	0.01056338
Star GAN	1,3,5,6,8,9	6	0.01056338
Cycle GAN	1,3,4,5,7,8	6	0.00880282
CGAN	1,3,5,8,9	5	0.00880282
SRGAN	1,2,4,8,9	5	0.00880282
CNN	1,4,8,9	4	0.00704225
DCGAN	1,5,8,9	4	0.00704225
Wasserstein GAN (WGAN)	2,5,8,9	4	0.00704225
LSTM	4,5,6,9	4	0.00704225
ResNet	4,6,7,8	4	0.00352113
Bicubic	2,7,9	3	0.00528169
LR	2,4,9	3	0.00528169
VDSR	2,8,9	3	0.00528169
LapSRN	2,8,9	3	0.00528169
ICGAN	1,5,8	3	0.00528169
VAE	1,8,9	3	0.00528169
SeqGAN	4,6,9	3	0.00528169
SVM	4,8,9	3	0.00528169
PCA	3,8,9	3	0.00352113
BEGAN	1,8,9	3	0.00352113

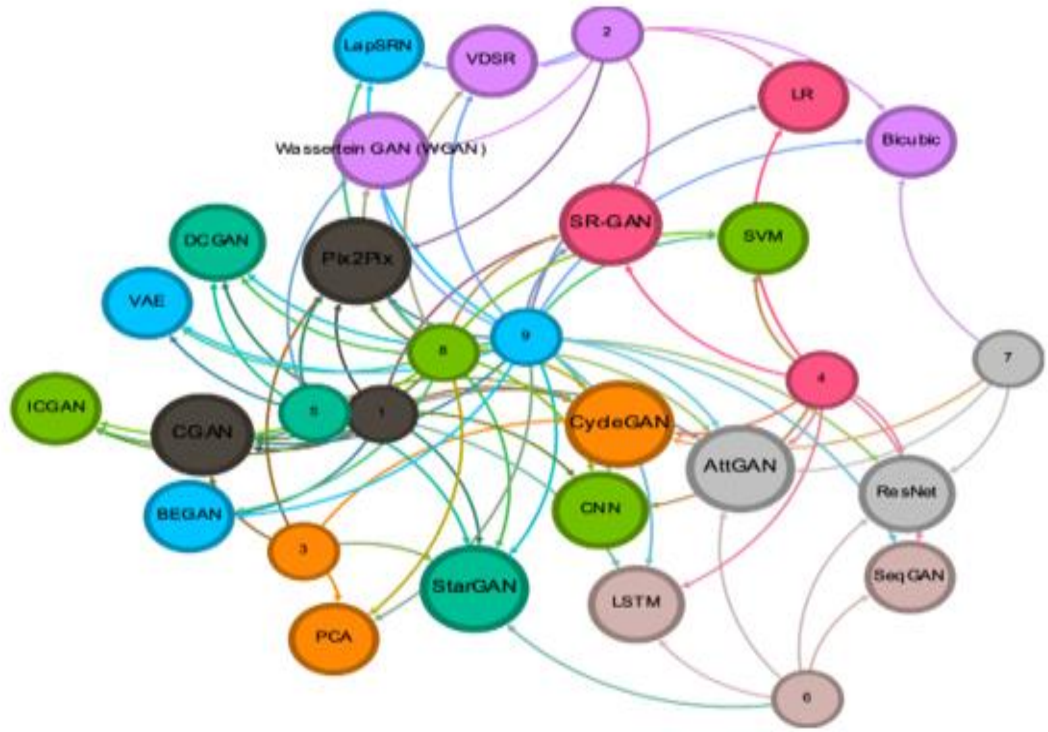

Fig. 2. The Models which are used in more than three different categories

Figure 2 is a network graph that displays the top models used in more than two categories. The graph consists of nodes (represented by circles) and edges (represented by lines connecting the nodes). The nodes represent the models and are labeled accordingly. The size of the nodes varies based on the number of categories the model is used in. Models used in more categories have larger nodes. The edges represent the connections between the models based on their co-occurrence in categories. The thickness of the edges varies based on the frequency of co-occurrence. In this network graph, AttGAN, Pix2pix, StarGAN, and CycleGAN are represented by larger nodes, indicating that they are used in six categories and have the highest degree centralities. These models are connected by thick edges, indicating their high co-occurrence in categories. CGAN and SRGAN are used in five categories, while CNN, DCGAN, WGAN, LSTM, and ResNet are used in four categories and are also represented in the graph with smaller nodes and thinner edges. The remaining models are used in three categories.

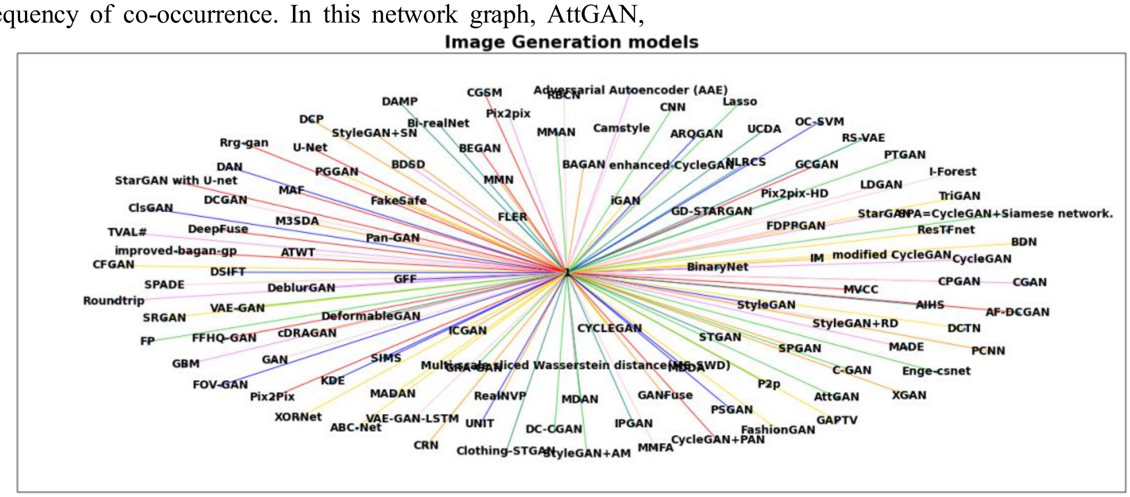

Fig. 3. Network of GAN models for image generation

Figure 3 shows all the GAN models used for image generation. This is fruchterman reingold layout depiction of the image generation models. StarGAN was used in maximum numbers of papers followed by DCGAN, CNN, CGAN and CycleGAN. These were followed by AttGAN, PGGAN, Pix2Pix, STGAN, StyleGAN.

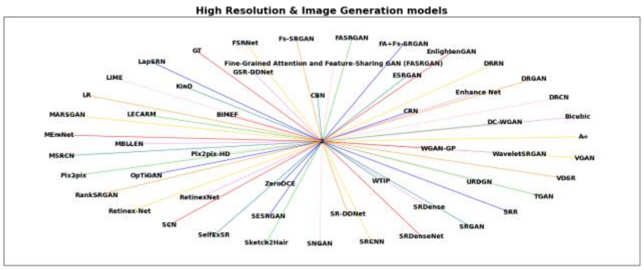

Fig. 4. (a) Network of GAN models for High resolution and image generation generation

Figure 4 (a) shows all the GAN models for High resolution and image generation. This is kamada kawai layout depiction of the for high resolution and image generation models. SRGAN was used in maximum number of papers followed by ESRGAN, SRdensenet, VDSR, and Bicubic.

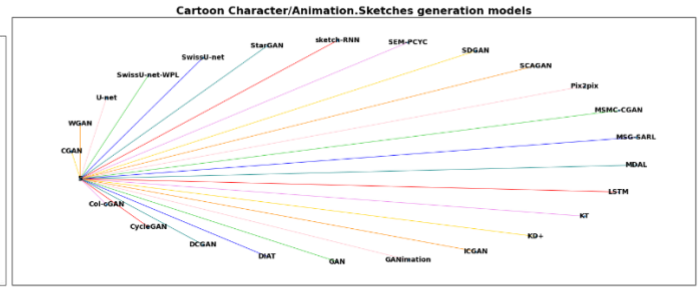

Fig 4 (b) Network of GAN models for Cartoon Character/Animation/sketches

Figure 4 (b) shows all the GAN models for Cartoon Character/Animation/sketches generation. This is shell layout depiction of the Cartoon Character/Animation/sketches generation models. CycleGAN, DCGAN were used in maximum number of papers.

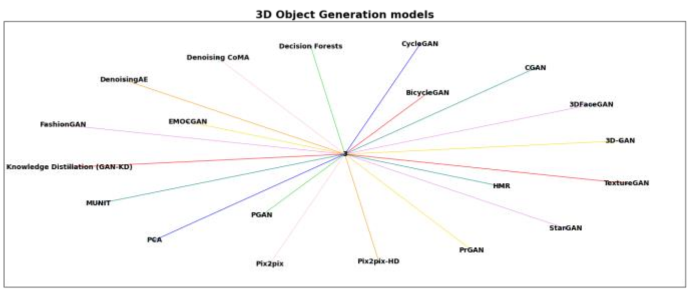

Fig. 5. (a) Network of GAN models for 3D Object Generation

Figure 5. (a) shows all the GAN models for 3D Object Generation. This is kamada kawai layout depiction of the 3D Object Generation models. Pix2pix was used in maximum number of papers.

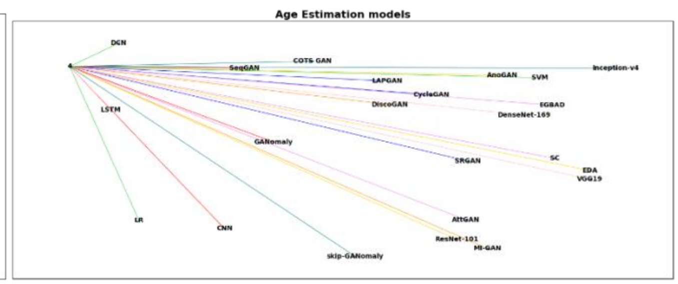

Fig. 5. (b) Network of GAN models for Age estimation

Figure 5. (b) shows all the GAN models for 3D Age estimation. This is random layout depiction of the Age estimation models. AttGAN, CNN, CycleGAN and SRGAN were used in maximum number of papers.

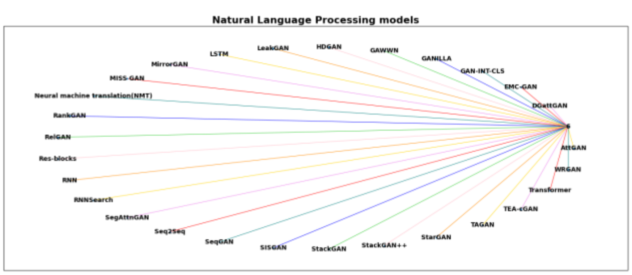

Fig. 6. (a) Network of GAN models for Natural Language processing

Figure 6. (a) shows all the GAN models for Natural Language processing. This is circular layout depiction of the Natural Language processing models.

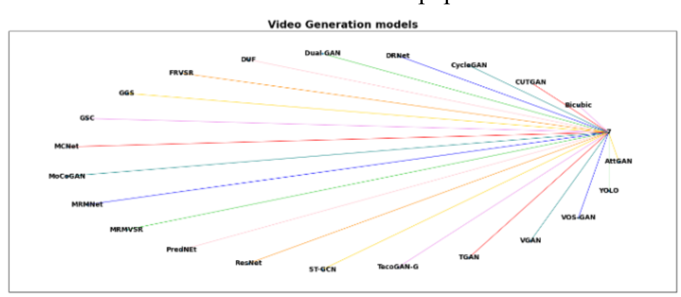

Fig. 6. (b) Network of GAN models for Video generation

Figure 6 (b) shows all the GAN models for video generation. This is circular layout depiction of the video generation models. MoCoGAN was used in maximum number of papers.

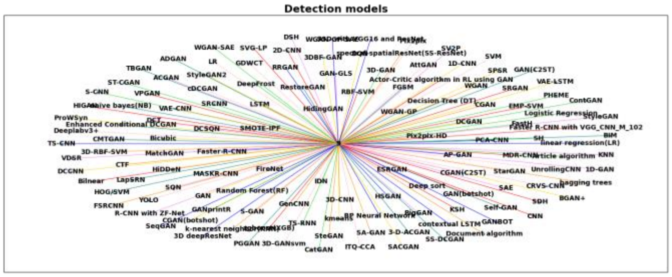

Fig. 7. (a) Network of GAN models for Detection

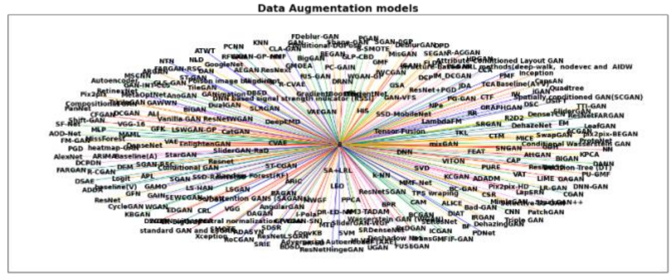

Fig 7. (b) Network of GAN models for Data Augmentation

Figure 7 (a) shows all the GAN models for Detection. This is spring layout depiction of the Detection models. DCGAN, and StyleGAN were followed by CNN and PGGAN.

Figure 7. (b) shows all the GAN models for Data Augmentation. This is fruchterman reingold layout depiction of the Data Augmentation. models. DCGAN was used in maximum number of papers followed by WGAN-GP, CGAN. These were followed by ALI, CatGAN, CNN, CycleGAN, LSGAN, SMOTE, SRGAN, and VGG.

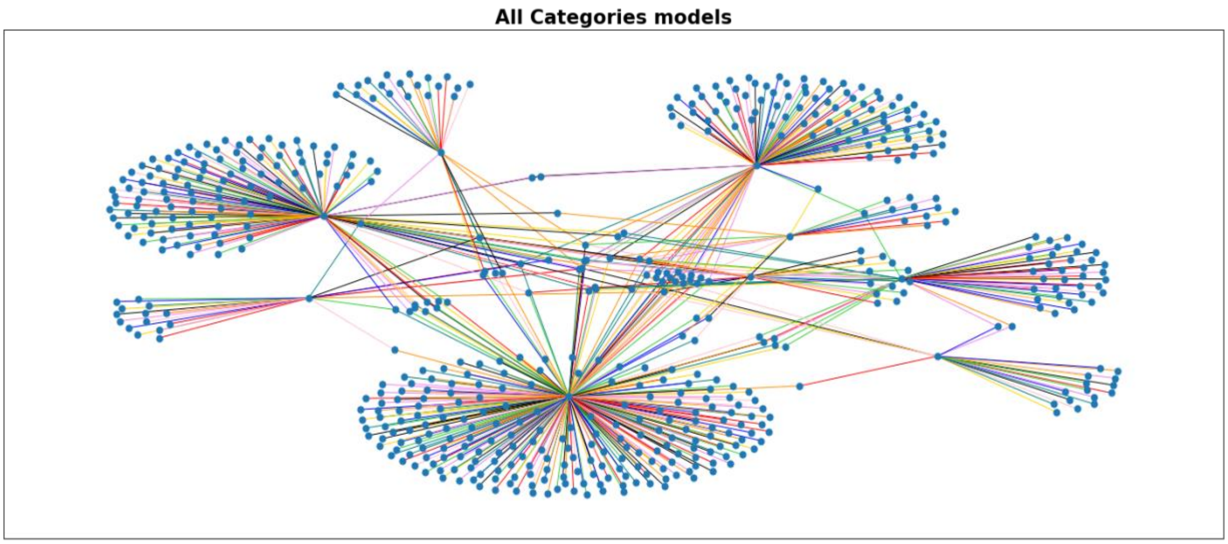

Fig. 8. Network of GAN models for All Categories Models

Figure 8 shows all the GAN models for all categories. This is fruchterman reingold layout depiction of the for all categories models.

IV. CONCLUSION

Generative Adversarial Networks (GANs) are a potent tool for creating deepfakes with applications across various industries like advertising, film production, video games, and creative arts. However, these deepfakes can also pose significant security threats like election manipulation, fraud, and blackmail. While GANs have been used for tasks like image and high-resolution image generation, 3D object generation, age estimation, natural language processing, video generation, data augmentation, and detection, each task requires a different model for optimal performance. Hence, we aimed to identify a single combination model of GANs that can perform all categories of deepfake generation tasks. In this study, we conducted a systematic review of various GAN models and found that AttGAN and Pix2Pix had the highest degree centralities. Based on these findings, we proposed a combination model of GANs that can carry out all categories of deepfake generation tasks. This study sheds light on GAN model development for deepfake generation and emphasizes responsible use and regulation of this technology to prevent misuse and emphasizes the need for continued research and development in this area.

However, it's worth noting that creating a single "one-size-fits-all" GAN model that can perform all tasks optimally may not be feasible, as different tasks require different features and training data. It may be more practical to develop specialized models for each task or a framework for integrating multiple models as needed.

V. REFERENCES

[1] M. Carter and M. Tsikerdekis, "Approaches for Fake Content Detection :," no. April, 2021.

[2] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks2014;:1–9 http://arxiv. org/abs/1406.2661

[3] X. Ning et al., ''Multi-view frontal face image generation: A survey,'' Con- currency Comput., Pract. Exper., p. e6147, 2020, doi: 10.1002/cpe.6147.

[4] B. Ay, G. Aydin, Z. Koyun, and M. Demir, ''A visual similarity recommendation system using generative adversarial networks,'' in Proc. Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. (Deep-ML), Aug. 2019, pp. 44–48.

[5] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, ''Generative adversarial networks: An overview,'' IEEE Signal Process., vol. 35, no. 1, pp. 53–65, Jan. 2017.

[6] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes : A Survey," vol. 54, no. 1, pp. 1–41, 2020.

[7] M. Masood, M. Nawaz, K. M. Malik, A. Javed, and A. Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," 2021, [Online]. Available: http://arxiv.org/abs/2103.00484.

[8] A. Malik, M. Kuribayashi, and S. Member, "Deep Fake Detection for Human Face Images and Videos : A Survey," IEEE Access, vol. 10, pp. 18757–18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[9] Y. He et al., "ForgeryNet: A Versatile Benchmark for Comprehensive Forgery Analysis," 2021, [Online]. Available: http://arxiv.org/abs/2103.05630.

[10] V. V. V. N. S. Vamsi et al., "Deepfake Detection in Digital Media Forensics," Glob. Transitions Proc., pp. 0–6, 2022, doi: 10.1016/j.gltp.2022.04.017.

[11] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ''Photo-realistic single image super-resolution using a generative adversarial network,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4681–4690.

[12] D. Berthelot, T. Schumm, and L. Metz, ''BEGAN: Boundary equilibrium generative adversarial networks,'' 2017, arXiv:1703.10717.

[13] M. Arjovsky, S. Chintala, and L. Bottou, ''Wasserstein generative adversarial networks,'' in Proc. Int. Conf. Mach. Learn. PMLR, 2017, pp. 214–223.

[14] A. Canziani, A. Paszke, and E. Culurciello, ''An analysis of deep neural network models for practical applications,'' 2016, arXiv:1605.07678.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.

[16] W. Zhiqiang and L. Jun, ''A review of object detection based on convolutional neural network,'' in Proc. 36th Chin. Control Conf. (CCC), Jul. 2017, pp. 11104–11109.

[17] M. Cheng et al., "Spatio-Temporal Hourly and Daily Ozone Forecasting in China Using a Hybrid Machine Learning Model: Autoencoder and Generative Adversarial Networks," J. Adv. Model. Earth Syst., vol. 14, no. 3, pp. 1–27, 2022, doi: 10.1029/2021ms002806

[18] T. Karras , T. Aila , S. Laine , J. Lehtinen , Progressive growing of gans for improved quality, stability, and variation, arXiv preprint arXiv:171010196 (2017) .Pattern Recognition Letters 153 (2022) 207–213

[19] Y. Choi , M. Choi , M. Kim , J.-W. Ha , S. Kim , J. Choo , Stargan: Unified generative adversarial networks for multi-domain image-to-image translation, IEEE CVPR, 2018 .

[20] J.-Y. Zhu , R. Zhang , D. Pathak , T. Darrell , A .A . Efros , O. Wang , E. Shechtman , Toward multimodal image-to-image translation, in: Advances in neural information processing systems, 2017, pp. 465–476 .

[21] C.-C. Hsu , Y.-X. Zhuang , C.-Y. Lee , Deep fake image detection based on pairwise learning, Applied Sciences 10 (1) (2020) 370 .

[22] X. Zhang , S. Karaman , S.-F. Chang , Detecting and simulating artifacts in gan fake images, in: 2019 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, 2019, pp. 1–6 .

[23] E. McFowland III , S.D. Speakman , D.B. Neill , Fast generalized subset scan for anomalous pattern detection, The Journal of Machine Learning Research 14 (1) (2013) 1533–1561

[24] F. Marra , D. Gragnaniello , D. Cozzolino , L. Verdoliva , Detection of gan-generated fake images over social networks, in: 2018 IEEE Conference on Multimedia In- formation Processing and Retrieval (MIPR), IEEE, 2018, pp. 384–389 .

[25] R. Wang, F. Juefei-Xu, L. Ma, X. Xie, Y. Huang, J. Wang, Y. Liu, Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces2020;

[26] C. Cintas, S. Speakman, G. Abebe, V. Akinwande, E. Mcfowland, and K. Weldemariam, "Pattern detection in the activation space for identifying synthesized content," Pattern Recognit. Lett., vol. 153, pp. 207–213, 2022, doi: 10.1016/j.patrec.2021.12.007.

[27] (U) | Samantha Cole | Vice | https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn/| 11 Dec. 2017 |AI- Assisted Fake Porn is Here and We're All Fucked.

[28] F. Marra, C. Saltori, G. Boato, and L. Verdoliva, ''Incremental learning for the detection and classification of GAN-generated images,'' in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2019, pp. 1–6.

[29] M. Goebel, L. Nataraj, T. Nanjundaswamy, T. M. Mohammed, S. Chandrasekaran, and B. S. Manjunath, ''Detection, attribution, and localization of GAN generated images,'' 2020, arXiv:2007.10466.

[30] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, ''CNN-generated images are surprisingly easy to spot. . . for now,'' in Proc. IEEE/CVFConf. Comput. Vis. Pattern Recognit. (CVPR), vol. 7, Jun. 2020, pp. 1–10.

[31] Z. Liu, X. Qi, and P. H. S. Torr, ''Global texture enhancement for fake face detection in the wild,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 8060–8069.

[32] N. Hulzebosch, S. Ibrahimi, and M. Worring, ''Detecting CNN-generated facial images in real-world scenarios,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2020, pp. 642–643.

[33] S. McCloskey and M. Albright, ''Detecting GAN-generated imagery using color cues,'' 2018, arXiv:1812.08247.

[34] H. Li, B. Li, S. Tan, and J. Huang, ''Identification of deep net- work generated images using disparities in color components,'' 2018, arXiv:1808.07276.

[35] N. Yu, L. Davis, and M. Fritz, ''Attributing fake images to GANs: Learning and analyzing GAN fingerprints,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 7556–7566.

[36] X. Zhang, S. Karaman, and S.-F. Chang, ''Detecting and simulating arti- facts in GAN fake images,'' in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2019, pp. 1–6.

[37] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, ''Leveraging frequency analysis for deep fake image recognition,'' 2020, arXiv:2003.08685.

[38] R. Durall, M. Keuper, and J. Keuper, ''Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 7890–7899.

[39] X. Yang, Y. Li, H. Qi, and S. Lyu, ''Exposing GAN-synthesized faces using landmark locations,'' in Proc. ACM Workshop Inf. Hiding Multimedia Secur., Jul. 2019, pp. 113–118.

[40] D. Bahdanau, K. Cho, and Y. Bengio, ''Neural machine translation by jointly learning to align and translate,'' 2014, arXiv:1409.0473.

[41] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, ''Learning deep features for discriminative localization,'' in Proc. IEEE Conf. Com- put. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2921–2929.

[42] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ''Grad-CAM: Visual explanations from deep networks via gradient-based localization,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 618–626