Ten Years of Generative Adversarial Nets (GANs): A survey of the state-of-the-art

Tanujit Chakraborty, Ujjwal Reddy K S, Shraddha M. Naik, Madhurima Panja, and Bayapureddy Manvitha

***Abstract*—Since their inception in 2014, Generative Adversarial Networks (GANs) have rapidly emerged as powerful tools for generating realistic and diverse data across various domains, including computer vision and other applied areas. Consisting of a discriminative network and a generative network engaged in a Minimax game, GANs have revolutionized the field of generative modeling. In February 2018, GAN secured the leading spot on the "Top Ten Global Breakthrough Technologies List" issued by the Massachusetts Science and Technology Review. Over the years, numerous advancements have been proposed, leading to a rich array of GAN variants, such as conditional GAN, Wasserstein GAN, CycleGAN, and StyleGAN, among many others. This survey aims to provide a general overview of GANs, summarizing the latent architecture, validation metrics, and application areas of the most widely recognized variants. We also delve into recent theoretical developments, exploring the profound connection between the adversarial principle underlying GAN and Jensen-Shannon divergence, while discussing the optimality characteristics of the GAN framework. The efficiency of GAN variants and their model architectures will be evaluated along with training obstacles as well as training solutions. In addition, a detailed discussion will be provided, examining the integration of GANs with newly developed deep learning frameworks such as Transformers, Physics-Informed Neural Networks, Large Language models, and Diffusion models. Finally, we reveal several issues as well as future research outlines in this field.**

***Index Terms*—Adversarial learning, Image generation, Deep learning, Model evaluation and selection, Generative Adversarial Networks, Generator network, Artificial intelligence.**

I. Introduction

GENERATIVE Adversarial Networks (GANs) have emerged as a transformative deep learning approach for generating high-quality and diverse data. In GAN, a generator network produces data, while a discriminator network evaluates the authenticity of the generated data. Through an adversarial mechanism, the discriminator learns to distinguish between real and fake data, while the generator aims to produce data that is indistinguishable from real data.

Since their introduction in 2014 by Goodfellow et al. [1], GANs have witnessed remarkable advancements, leading to the development of numerous specialized variants that excel in creating data across diverse fields. Conditional GAN [2] enables the generation of data based on specific conditions or desired qualities, such as synthesizing photos of a particular class. CycleGAN [3] have proven effective in image-to-image translation tasks, even in the absence of paired data. StackGAN [4] has demonstrated the ability to generate high-resolution images from textual descriptions, pushing the boundaries of visual realism. Progressive GAN [5] has achieved exceptional results in producing high-quality images with increasing resolution. StyleGAN [6], known for its versatility, generates images with a wide range of styles and distinctive features. Furthermore, GANs have extended beyond visual domains and shown potential in generating textual [7], musical [8], 3D modeling [9], future cities [10], time series [11] data among many others.

The success of GANs has led to their adoption in various applications, such as image and video synthesis, data augmentation, super-resolution, inpainting, anomaly detection, and image editing. GANs have also been employed to address data scarcity issues in machine learning, where they generate synthetic data to improve the effectiveness of models trained on limited datasets [12]. Additionally, GANs have found utility in creating realistic simulations for video games and virtual reality environments, enhancing user experiences and immersive interactions [13]. To ensure the comprehensiveness of this survey, we conducted an extensive review of the research papers encompassing both theoretical advancements and practical applications of GAN. Our survey draws insights from diverse fields, including computer vision, natural language processing, autonomous vehicles, time series, medical domain, and many others. Notable papers that significantly contributed to our survey include Goodfellow et al. [1] for introducing the GAN framework, Mirza and Osindero [2] for pioneering conditional GAN, Zhu et al. [3] for introducing CycleGAN, Karras et al. [5] for their seminal work on progressive GAN, and Chen et al. [14] for the breakthroughs achieved with InfoGAN, among many others.

Despite their remarkable achievements, GANs face several challenges in practice. One prominent issue is the instability of the training process, which can result in mode collapse or oscillation [15]. Another challenge lies in the evaluation of generated data, as conventional assessment criteria may not adequately capture the diversity and realism of the synthesized samples [16]. Furthermore, GANs have been observed to exhibit biases, particularly concerning gender and race, potentially reflecting the biases present in the training data [17], [18]. To overcome the limitations of GAN various modified

Corresponding Author: T. Chakraborty is with the Department of Science and Engineering, Sorbonne University Abu Dhabi, UAE and Sorbonne Centre for Artificial Intelligence, Sorbonne Université, Paris. (e-mail: tanujit.chakraborty@sorbonne.ae).

U. Reddy K S and B. Manvitha are with Vellore Institute Of Technology, Andhra Pradesh, India. (e-mail: ujjwalreddyks@gmail.com and manvitha.bayapureddy@gmail.com).

S M. Naik is with the Department of Science and Engineering, Sorbonne University Abu Dhabi, UAE. (e-mail: 7.shraddha.naik@gmail.com).

M. Panja is with the Center for Data Sciences, International Institute of Information Technology Bangalore, India. (e-mail: madhruima.panja@iiitb.ac.in).

All authors have contributed equally to this survey.

training approaches and hybridization with popular deep learning architectures such as Transformers [19], Physics-Informed Neural Network (PINN) [20], Large language models (LLMs) [21], and Diffusion models [22] have been proposed in the literature. These modified methodologies have shown promise in enhancing the synthetic data generation capabilities of GANs.

Finally, GANs have emerged as an effective tool for producing high-quality and varied data in several disciplines. Notwithstanding the difficulties connected with their use, GANs have shown outstanding results and have the potential to drive innovation in disciplines such as computer vision, machine learning, and virtual reality. This in-depth analysis covers the accomplishments and limitations of GAN, as well as the promise of these approaches for future research and applications. This comprehensive survey aims to explore both the accomplishments and challenges of GAN. The contributions of the article can be summarized as follows:

- **Exploration of Vanilla GAN and their applications:** We offer an elaborate description of the GAN model, encompassing its architectural particulars and the mathematical optimization functions it employs. We summarize the areas where GANs have emerged as a promising tool in efficiently solving real-world problems with their generative capabilities.
- **Evolution of state-of-the-art GAN models across the decade:** Our comprehensive analysis encompasses a wide range of cutting-edge GAN adaptations crafted to address practical challenges across various domains. We delve into their structural designs, practical uses, execution methods, and constraints. To facilitate a lucid understanding of the field's progress, we present an intricate chronological breakdown of GAN model advancements. Furthermore, we evaluate recent field surveys, outlining their pros and cons, while also tackling these aspects within our own survey.
- **Theoretical advancements of GANs:** We give a technical overview of the theoretical developments of GANs by exploring the connections between adversarial training and Jensen-Shannon divergence and discussing their optimality features.
- **Assessment of GAN Models:** We provide a comprehensive breakdown of the essential performance measures utilized to assess both the caliber and range of samples produced by GANs. These metrics notably fluctuate depending on the specific domains of application.
- **Limitations of GANs:** We critically examine the constraints associated with GANs, primarily stemming from issues of learning instability, and discuss various enhancement strategies aimed at alleviating these challenges.
- **Anticipating future trajectories:** In addition to evaluating the pros and cons of current GAN-centric approaches, we illuminate the hybridization of emerging deep learning models such as Transformers, PINNs, LLMs, and Diffusion models with GANs. We outline potential avenues for research within this domain by summarizing several open scientific problems.

This survey is structured in the following manner. Section II digs into related works and recent surveys giving background information and emphasizing the most significant developments in GAN done over the decade. Section III is a concise overview of GAN describing the fundamental components and intricate details of its architecture. In Section IV, we examine the wide range of fields that GANs have influenced, such as computer vision, natural language processing, time series, and audio, among many others. Subsequently, Section V reviews the innovations and applications of popular GAN-based frameworks from various domains along with their implementation software and discusses their limitations. This section also provides a timeline for the GAN models to have a clear vision of the development of this field. Section VI summarizes the recent theoretical developments of GAN and its variants. Section VII reviews the metrics used for evaluating GAN-based models. Section VIII analyzes the limitations of GANs and presents its remedial measures. Section IX discusses the potential and usability of GAN with the development of new deep learning technologies such as Transformers, PINNs, LLMs, and Diffusion models. Section X proposes potential directions for further research in this field. Finally, Section XI concludes the survey by indicating prospective directions for future research projects while also offering a closing assessment of the successes and limits of GANs.

II. Related Works and Recent Surveys

GANs are a promising deep learning framework for generating artificial data that closely resembles real-world data [1]. Early GAN-related research focused on creating realistic visuals. Radford et al. proposed a deep convolutional GAN (DCGAN) in 2015 [23], which utilized convolutional layers, batch normalization, and a specific loss function to generate high-quality images. DCGAN introduced important innovations in image generation. In 2017, Karras et al. [5] introduced progressive growing GAN (ProGAN), which generates higher quality and resolution images compared to vanilla GAN. ProGAN trains multiple generators and discriminators in a stepwise manner, gradually increasing the resolution of the generated images. The results demonstrated the ability of ProGAN to produce images closely resembling genuine photos for various datasets, including the CelebA dataset [24].

GANs have found applications beyond image generation, including video production and text generation. Vondrick et al. proposed a video generation GAN (VGAN) in 2018 [38], capable of producing realistic and diverse videos by learning to track and anticipate object motion. The VGAN architecture consisted of a motion estimation network and a video-generating network, jointly trained to generate high-quality videos. The results showcased VGAN's ability to produce realistic and varied films, enabling applications like video prediction and synthesis. Text generation is another domain where GAN has been utilized. In 2017, Yu et al. introduced SeqGAN, a GAN-based text generation model [39]. SeqGAN achieved realistic and diverse text generation capabilities by maximizing a reinforcement learning goal. The model included a generator responsible for text creation and a discriminator

TABLE I
COMPARISON OF OUR SURVEY AND OTHER RELATED GAN SURVEYS (GREEN CIRCLE SIGNIFIES "FULLY COVERED", BLUE CIRCLE SIGNIFIES "PARTIALLY COVERED", AND RED CIRCLE SIGNIFIES "NOT COVERED").

Year	Survey	Theoretical Insights	Evaluation Metrics	Domain						
				Computer Vision	Natural Language Processing	Music	Medical	Time Series	Urban Planning	Imbalanced Classification
2019	Kulkarni et al. [25]	🔵	🔴	🔴	🔴	🟢	🔴	🔴	🔴	🔴
2021	Jabbar et al. [26]	🟢	🔴	🟢	🟢	🟢	🟢	🔴	🔴	🔴
2021	Durgadevi et al. [27]	🔵	🔴	🔵	🔵	🔴	🔵	🔴	🔴	🔴
2021	Nandhini et al. [28]	🔴	🔴	🟢	🔴	🔴	🔴	🔴	🔴	🔴
2021	Wang et al. [29]	🔵	🔴	🟢	🔴	🔴	🔴	🔴	🔴	🔴
2021	Sampath et al [30]	🔵	🔴	🔵	🔴	🔴	🔴	🔴	🔴	🔵
2021	Gui et al [31]	🟢	🟢	🟢	🟢	🔵	🔵	🔴	🔴	🔴
2021	Li et al [32]	🟢	🟢	🔵	🔵	🔴	🔴	🔴	🔴	🔴
2022	Xia et al. [33]	🟢	🟢	🔵	🔴	🔴	🔴	🔴	🔴	🔴
2022	Xun et al. [34]	🔵	🔴	🔴	🔴	🔴	🟢	🔴	🔴	🔴
2023	Ji et al. [35]	🔴	🟢	🔴	🔴	🟢	🔴	🔴	🔴	🔴
2023	Iglesias et al. [36]	🔵	🟢	🟢	🔵	🔴	🔵	🔴	🔴	🔵
2023	Brophy et al. [37]	🟢	🟢	🔴	🔴	🔴	🔴	🟢	🔴	🔴
2023+	Our survey	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢	🟢

assessing the quality of the generated text. Through reinforcement learning, the generator was trained to maximize the predicted reward based on the discriminator's evaluation. The findings demonstrated that SeqGAN outperformed previous text generation algorithms, producing text that was more varied and lifelike. These advancements in GAN applications for video and text generation highlight the versatility and potential of GAN frameworks in diverse domains.

Another popular area of research focuses on addressing medical questions using GANs, as highlighted in the recent paper by Tan et al. where a GAN-based scale invariant post-processing approach is proposed for lung segmentation in CT Scans [40]. A similar framework called RescueNet, developed by Nema et al., combines domain-specific segmentation methods and general-purpose adversarial learning for segmenting brain tumors [41]. Their study not only suggests a promising technique for brain tumor segmentation but also advances the development of systems capable of answering complex medical inquiries. Despite the significant breakthroughs, there are still unresolved issues in GAN architectures and applications. One prominent challenge is the instability of GAN training, which can be influenced by various factors such as architecture, loss function, and optimization technique. In 2017, Arjovsky et al. proposed a solution called Wasserstein GAN (WGAN) [15], introducing a novel loss function and optimization algorithm to address stability issues in GAN training. Their approach showed improved stability and performance on datasets like CIFAR-10 [42] and ImageNet [43].

Related survey. The existing body of research exploring various analytic tasks with GAN is accompanied by numerous surveys, which predominantly concentrate on specific perspectives within constrained domains, particularly computer vision and natural language processing. For instance, the survey by Jabbar et al. [26] explores applications of GANs in various industries, including computer vision, natural language processing, music, and medicine. To demonstrate the influence and promise of GANs in certain application domains, they also highlight noteworthy academic publications and real-world instances. The study tackles the difficulties and problems related to GAN training in addition to discussing their variations. The authors [26] investigate several training strategies, including minimax optimization, training stability, and assessment measures. They examine the typical challenges that arise during GANs training, such as mode collapse and training instability, and they give numerous solutions that have been suggested by researchers to address these problems. However, it does not specifically concentrate on GAN-based methods for imbalanced, time series, geoscience, and other data types and fails to reflect the most recent advancements in the field. The survey by Xia et al. [33] focuses on two primary categories of techniques for GAN inversion: Optimization-based methods and Reconstruction-based methods. To locate the hidden code that optimally reconstructs the supplied output, optimization-based approaches formulate an optimization issue. Reconstruction-based approaches, on the other hand, use different methods, such as feature matching or autoencoders, to directly estimate the latent code. An in-depth discussion of these strategies' advantages, disadvantages, and trade-offs is provided in the article. The non-convexity of the optimization issue and the lack of ground truth data for assessment are only two of the difficulties faced in GAN inversion that are highlighted in this article. The authors [33] additionally go through specific evaluation standards and measures designed for computer vision tasks. In addition, the study discusses current developments and variants in GAN inversion, such as techniques for managing conditional GAN, detaching latent variables, and dealing with different modalities. Aspect modification, domain adaptability, and unsupervised learning are a few of the applications and potential future directions of GAN inversion that are covered. A recent study by Durgadevi et al. [27] presents a comprehensive overview of numerous GAN variants that have been proposed until 2020. Since its inception, GANs have undergone significant evolutions leading researchers to propose various enhancements and modifications aimed at addressing the prevalent challenges. These alterations encompass diverse aspects such as architectural design, training methods, or a combination of both. In this survey [27] the authors delve into the application and impact of GANs in different domains including image processing, medicine, face detection, and text transferring. The survey by Alom et al. [44] covers various aspects of the deep learning paradigm, such as fundamental ideas, algorithms, architec-

tures, and contemporary developments including convolutional neural networks (CNNs), recurrent neural networks (RNNs), deep belief networks (DBNs), generative models, transfer learning, and reinforcement learning. The survey of Nandhini et al. [28] offers a thorough investigation of the application of deep CNNs and deep GANs in computational image analysis driven by visual perception. The designs and methodology used, the outcomes of the experiments, and possible uses for these approaches are covered in the paper. Overall this study provides a retrospective review of the development of GANs for the deep learning-based image analysis community. The survey by Kulkarni et al. [25] presents an overview of various strategies, techniques, and developments used in GAN-based music generation. The survey of Sampath et al. [30] summarizes the current advances in the GAN landscape for computer vision tasks including classification, object detection, and segmentation in the presence of an imbalanced dataset. Another survey by Brophy et al. [37] attempts to review various discrete and continuous GAN models designed for time series-related applications. The study by Xun et al. [34] reviews more than 120 GAN-based models designed for region-specific medical image segmentation that were published until 2021. Another recent survey by Ji et al. [35] summarizes the task-oriented GAN architectures developed for symbolic music generation but other application domains are overlooked. The survey by Wang et al. [29] reviews various architecture-variant and loss-variant GAN frameworks designed for addressing practical challenges relevant to computer vision tasks. Another survey by Gui et al. [31] provides a comprehensive review of task-oriented GAN applications and showcases the theoretical properties of GAN and its variants. The study by Iglesias et al. [36] summarizes the architecture of the latest GAN variants, optimization of the loss functions, and validation metrics in some promising application domains including computer vision, language generation, and data augmentation. The survey by Li et al. [32] reviews the theoretical advancements in GAN and also provides an overview of the mathematical and statistical properties of GAN variants. A detailed comparison between our survey and others is presented in Table I.

Although there are several papers reviewing GAN architecture and its domain-specific applications, none of them concurrently emphasize on applications of GAN in geoscience, urban planning, data privacy, imbalanced learning, and time series problems in a comprehensive manner. Methods developed to deal with these practical problems are underrepresented in past surveys. Moreover, the stability of GANs training, assessment of the produced data, and ethical issues with GAN are some of the issues that still need to be resolved. To fully exploit the future potential of GANs, more study in these areas is required. To fill the gap, this survey offers a comprehensive and up-to-date review of GANs, encompassing mainstream tasks ranging from audio, video, and image analysis, to natural language processing, privacy, geophysics, and many more. Specifically, we first provide several applied areas of GAN and discuss existing works from task and methodology-oriented perspectives. Then, we delve into multiple popular application sectors within the existing research of GAN with their limitations and propose several potential future research directions. Our survey is intended for general machine learning practitioners interested in exploring and keeping abreast of the latest advancements in GAN for multi-purpose use. It is also suitable for domain experts applying GANs to new applications or exploring novel possibilities building on recent advancements.

III. Overview of Generative Adversarial Network

Generative Adversarial Networks (GANs) signify a pivotal advancement in artificial intelligence, offering a robust framework to craft synthetic data that closely resembles real-world information [45]. Consisting of two interconnected neural networks, the Generator and Discriminator, GANs engage in a dynamic adversarial process that is redefining the landscape of deep generative modeling [1], [46]. By orchestrating this interplay, GANs transcend data generation frontiers across various domains, from crafting images to generating language, demonstrating a profound influence on reshaping the way machines comprehend and replicate intricate data distributions. This dynamic is facilitated through the Generator (G) network, entrusted with producing new data samples based on the input data distribution, while the Discriminator (D) network is devoted to discerning genuine data from their synthetic counterparts.

From a mathematical viewpoint, the G network considers a latent space z from the noise distribution p_z as input and generates synthetic samples $G(z)$. Its goal is to generate data that is indistinguishable from real data samples x originating from the probability distribution p_{data}. On the other hand, D takes both real data samples x from the actual dataset and fake data samples $G(z)$ generated by G as input and classifies whether the input data is real or fake. It essentially acts as a "critic" that evaluates the quality of the generated data. The training process consists of both networks working in a two-player zero-sum game [36]. While G aims to produce more realistic outcomes, D enhances its ability to distinguish between real and fake samples. This dynamic prompts both players to evolve in tandem: if G generates superior outputs, it becomes tougher for D to discern them. Conversely, if D becomes more accurate, G faces greater difficulty in deceiving D. This process resembles a minimax game, where D strives to maximize accuracy while G seeks to minimize it [47]. The goal is to find a balance where G produces increasingly convincing data while D becomes better at classifying real data from fake ones. The mathematical expression of this minimax loss function can be represented as:

$$\min_G \max_D L = \mathbb{E}_{x\sim p_{\text{data}}}\left[\log D(x)\right] + \mathbb{E}_{z\sim p_z}\left[\log(1 - D(G(z)))\right], \tag{1}$$

where the probability values $D(\mathbf{x})$ and $D(G(\mathbf{z}))$ represent the discriminator's outputs for real and fake samples, respectively. The first term in Eq. (1) encourages D to correctly classify real data by maximizing $\log D(x)$, whereas the second term encourages G to produce realistic data that D classifies as real by minimizing $\log(1 - D(G(z)))$. In essence, G aims to minimize the loss while D aims to maximize it, leading to a continual back-and-forth training process. Throughout the

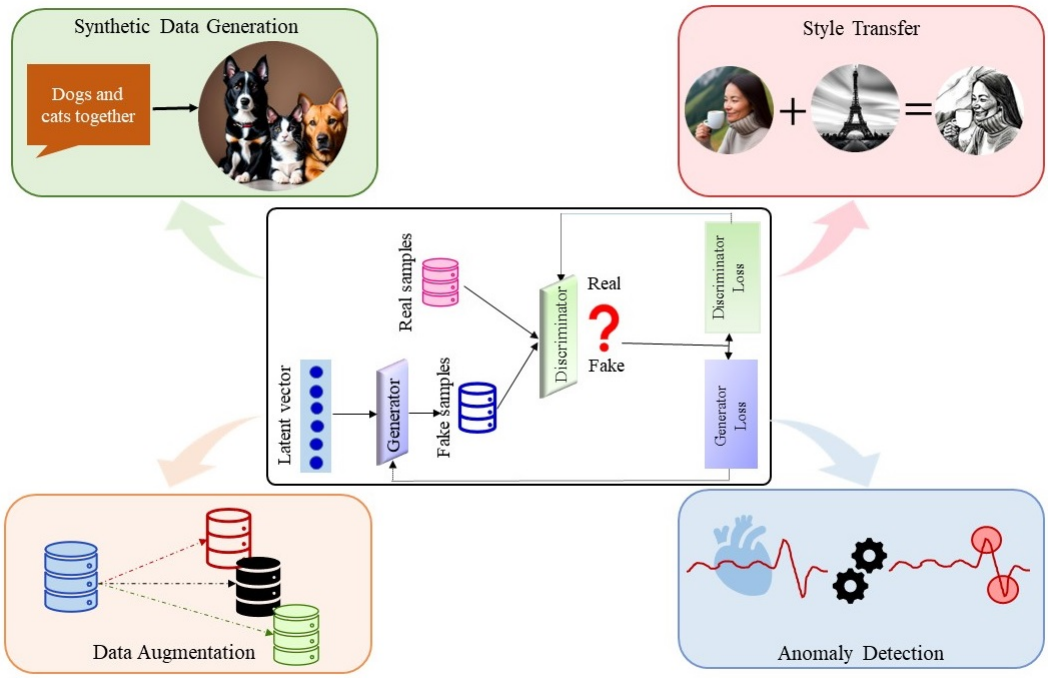

Fig. 1. Architecture of GANs and its primary functions. In this example, different analytical tasks of GANs are categorized into synthetic data generation, style transfer, data augmentation, and anomaly detection.

training, the generator's performance improves as it learns to generate more realistic data, and the discriminator's performance improves as it becomes better at distinguishing real from fake data. Ideally, this competition results in a generator that produces data that is virtually indistinguishable from real data, as judged by the discriminator. A visual representation of the GAN's architectural details and its primary functions is presented in Fig. 1.

During the time of the inception of GAN in 2014, Goodfellow et al. [1] proved the existence of a unique solution for the minimax loss function. This solution became popular as Nash Equilibrium (NE) which reflects the equilibrium point where the generator's capacity to generate realistic data matches the discriminator's capacity to distinguish between real and fake data, resulting in high-quality synthetic data that closely resembles the true underlying data distribution [48]. However, recent studies have revealed that attaining NE in GANs is not guaranteed and can be challenging due to various factors, including architecture choices, hyperparameters, and convergence difficulties [49], [50]. To address these challenges and enhance GAN's training stability researchers have developed various techniques, such as different loss functions and architectures over the decade [51]. These alterations of GAN include architectural changes, loss function-based modifications, and many others. They encompass various variations, each with unique attributes and applications, driving significant advancements in generative modeling. Fig. 2 visually depicts the timeline of key developments in GAN research.

IV. Application

As previously noted, GANs have emerged as one of the most prominent advancements in the realm of machine learning over recent years. GAN models have demonstrated their efficacy in domains where prior models fell short, while also substantially enhancing performance in other scenarios. Within this section, we will comprehensively explore the pivotal domains where GAN architectures have been deployed. While much of the recent research has concentrated on employing GANs to generate novel synthesized data, emulating distinct data distributions, our exploration in this section will highlight the broader applications of GANs, extending to areas such as video game development [52], urban planning [10] and others. We also visually showcase the application domains of GAN in Fig. 3.

a) Image Generation: Among the most promising domains harnessing the capabilities of GANs is computer vision. Notably, the generation of realistic images stands as one of the paramount applications of GANs [6], [53]. The capacity of GANs to craft authentic images depicting characters, animals, and objects that lack real-world existence holds immense significance [54]. This capability of GAN finds application in diverse projects, spanning from refining facial recognition algorithms to fabricating immersive virtual environments for video games and commercial campaigns [55]. Moreover, GANs have proven instrumental in generating true-to-life virtual realms, a boon for both the gaming industry and advertising ventures. By crafting synthetic landscapes and structures, GANs empower game designers and developers to construct captivating, realistic virtual worlds, thereby elevating the overall player experience [5]. The deployment of GANs

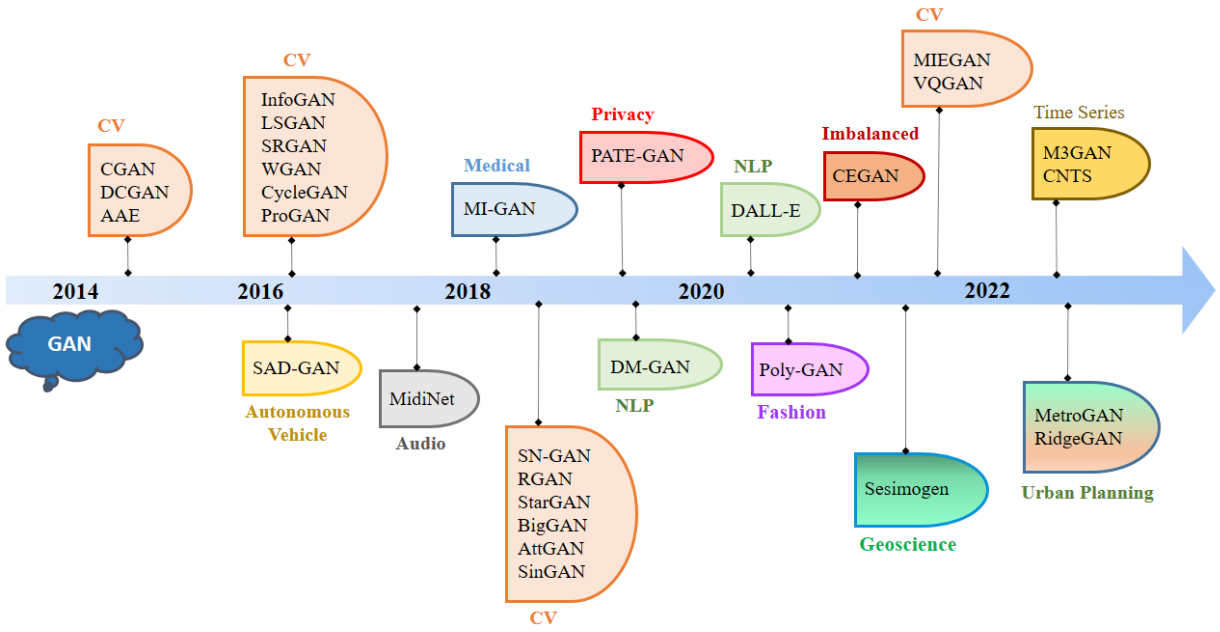

Fig. 2. Timeline of the application-based GAN architectures reviewed in this study

in this context offers a swift, cost-effective, and efficient alternative to traditional manual design and modeling approaches, enabling the production of high-quality graphics.

b) Video Synthesis: In addition to generating high-quality images, GANs offer the potential to create synthetic videos, a more complex task due to coherence requirements [56]. GANs, combining generators and discriminators, excel in this challenge [57]. The discriminator learns to differentiate real from synthetic frames, while the generator produces visually authentic video frames. GANs find widespread use in replicating real-world actions, enhancing surveillance and animations [58]. One of the most popular and controversial applications of GAN is the evolution of Deepfake [59]. Deepfakes are AI-generated media, that blend a person's likeness with another's context using GANs. While they offer creative potential, deepfakes raise ethical concerns, requiring a holistic approach to detect them [60], [61].

c) Augmenting data: GANs possess the capability to generate synthetic data, which can be harnessed to bolster actual data and enhance the performance of deep learning models. This approach is instrumental in mitigating concerns related to data scarcity and refining model accuracy [62]. GANs provide an effective avenue for fortifying machine learning and deep learning frameworks with authentic data. Addressing the challenge of limited data availability, GANs enable the creation of larger, more diverse datasets by generating artificial samples that closely emulate real data [63]. GAN-based data augmentation strategies have showcased promising outcomes across various domains, offering the potential to enhance model precision and transcend the constraints posed by insufficient data [64].

d) Style Transfer: GANs are capable of transferring the style of one image to another, resulting in the creation of an entirely new image [65]. This method can be applied to develop novel artistic features or enhance the visual attractiveness of pictures. By facilitating the development of fresh artistic trends and boosting the aesthetic appeal of pictures, GAN-based style transfer approaches have transformed the area of computer vision [3], [66]. These methods have been used in a variety of fields, such as digital art, photography, and graphic design, and they continue to be an inspiration for new developments and studies in the area.

e) Natural Language Processing: Over the past few years, GANs have been adapted to process text data, resulting in groundbreaking advancements within the realm of Natural Language Processing (NLP). One notable application involves text generation, where GANs can create coherent and contextually relevant textual content. For instance, the Text GAN framework utilizes Long Short-Term Memory (LSTM) networks [67] as the generator and CNN as the discriminator to synthesize novel text using adversarial training [68]. Furthermore, GANs play a role in text style transfer, allowing alterations in writing styles while preserving content, and enhancing the adaptability of generated material [69]. In the domain of sentiment analysis, GANs contribute by generating text with specific emotional tones, thereby aiding model training and dataset augmentation for sentiment classification tasks. Additionally, GANs are instrumental in text-to-image synthesis, translating textual descriptions into visual representations, proving valuable in fields like accessibility and multimedia content creation [4]. GANs have also been harnessed to enhance machine translation software, refining translation precision and fluidity [39], [70].

f) Music Generation: GANs are revolutionizing music creation by tapping into existing compositions' patterns and structures [71]. This technology not only fosters original music

composition but also assists musicians in their creative journey. Previous studies have showcased GANs' role in generating music, offering possibilities for both novel compositions and artist support [72], [73]. Beyond composition, GANs empower musicians to explore new styles by generating melodies, harmonies, and rhythms as creative sparks. They also enable style transfer, allowing musicians to reimagine their music in diverse genres and cultural contexts. Moreover, GANs have ventured into musical collaboration, aiding improvisation by responding to musician input with harmonious suggestions. In essence, GANs redefine music creation, from assisting composers in originality to fostering innovative style exploration [74]. This fusion of human creativity and computational ability promises to shape the future of the music industry.

g) Medical Domain: In the dynamic landscape of the medical domain, GANs have emerged as a game-changing technology with multifaceted benefits. The integration of GANs with medical data holds immense potential in enhancing disease diagnosis through the creation of synthetic medical images thereby eliminating the limited data problem. This expanding diversity and quantity of data made possible by GANs empower the data-driven diagnostic models to deliver more precise and reliable predictions, aiding healthcare practitioners in making accurate diagnoses and ultimately enhancing patient care [75]–[77]. Another significant application of GAN is in drug discovery, where it can process and generate molecular structures with desired properties [78], [79]. GAN-driven molecular generation accelerates the process of identifying potential drug candidates, saving time and resources in the search for novel therapeutic compounds. Moreover, GANs extend their impact to surgical training and planning by producing realistic surgical scenarios and simulations [80] and also aid in generating patient-specific medical images, allowing healthcare practitioners to tailor treatment plans to individual patient characteristics [81].

h) Urban Planning: With rapid urbanization, predicting transportation patterns is essential for sustainable urban planning and traffic management. Recent advancements in GAN-based methods to simulate hyper-realistic urban patterns, including CityGAN [82], Conditional GAN with physical constraints [83], and MetroGAN [84], have become popular in urban science fields. These GANs can generate synthetic urban universes that mimic global urban patterns, and quantifying landscape structures of these GAN-generated new cities using spatial pattern analysis helps in understanding landscape dynamics and improving sustainable urban planning. In a recent study, a novel RidgeGAN model [10] is proposed that evaluates the sustainability of urban sprawl associated with infrastructure development and transportation systems in medium and small-sized cities.

i) Geoscience and Remote Sensing: In geoscience, there are also recent applications of GANs with novel ways of generating "new" samples that can easily outperform state-of-the-art geostatistical tools. This is very appealing in applications like reservoir modeling as geologists and reservoir engineers are nowadays usually tasked to work with multiple realizations of the subsurface and provide probabilistic estimates to support the subsequent decision-making process. A few examples of early applications of GANs in geoscience are the reconstruction of three-dimensional porous media [85]; Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture [86]; and SeismoGen: Seismic Waveform Synthesis Using GAN With Application to Seismic Data Augmentation [87].

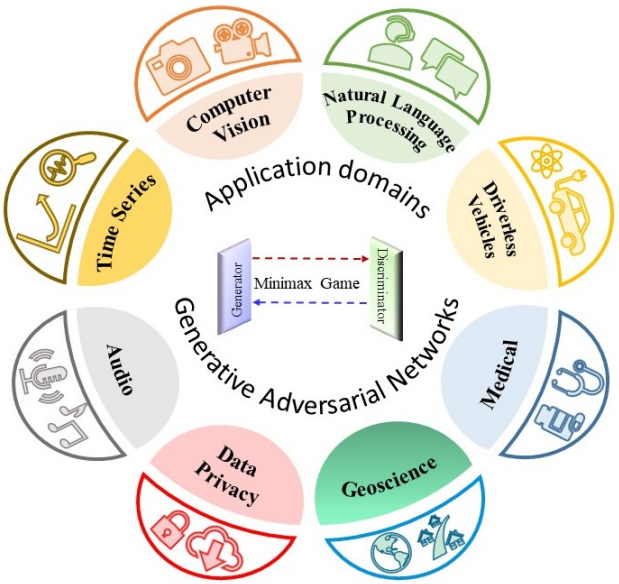

Fig. 3. Diverse Applications of Generative Adversarial Networks (GANs) in various applied domains.

j) Autonomous Vehicles: Machine learning models for autonomous driving can be trained using synthetic pictures of real-world situations created using GANs. This method helps to mitigate the safety concerns of autonomous cars by getting beyond the restrictions of real-world testing [88]. A potential method for training autonomous driving models is the use of GANs to produce synthetic visuals [89]. It makes it possible to investigate a wide range of complex scenarios, improving the performance and safety of the models. Recent studies have illustrated the usefulness and promise of this method for bridging the gap between driving simulations and actual driving situations, ultimately promoting the development of autonomous cars [90], [91].

k) Fashion and design: GANs find utility in generating fresh patterns and designs for clothing, aiding designers in crafting innovative collections. This technology extends its impact on online shopping experiences by producing images of apparel on virtual models, offering customers a realistic preview of how garments would appear on them during online purchases [92]. Within the realms of fashion and design, GANs have become a valuable asset, empowering designers to stretch their creative boundaries by facilitating the creation of novel patterns and designs [93]. Furthermore, GAN-driven virtual try-on systems enhance the convenience of online shopping, granting shoppers lifelike insights into how clothing would fit and appear on them. Several diverse research efforts in this domain have explored the significant contributions of GAN in the evolution of the fashion and design industry [94], [95].

l) Imbalanced Pattern Classification: A prevalent yet intricate issue encountered in pattern recognition is referred to as "class imbalance", signifying disparities in the frequencies of class labels [96]. To address this challenge, GANs can be used to generate synthetic data for the minority class of various imbalanced datasets as a method of intelligent oversampling [97]. Pioneering approaches such as Balancing GAN (BAGAN) [98] and Classification Enhancement GAN (CEGAN) [99] have been developed to restore balance in the distributions of imbalanced datasets and enhance the precision of the data-driven models.

m) Time Series Anomaly Detection: In recent years there has been a significant surge in the availability of real-time sensor data across diverse domains including healthcare systems, power plants, industries, and many others. These vast datasets are often accompanied by several anomalous events which eventually diminishes the modeling capabilities of any machine learning and deep learning frameworks. To address this issue anomaly detection for multivariate time series data has become a critical task for time series analysts [100]. In this context, GANs have become a powerful technology. In recent studies, various GAN-based time series anomaly detection techniques namely, Dilated Convolutional Transformer GAN (DCT GAN) [101], M2GAN [102], Cooperative Network Time Series (CNTS) [103], TADGAN [104], and many others have been developed that leverage the power of adversarial training to efficiently detect the presence of anomalous data.

n) Data privacy: GANs offer the possibility of generating synthetic data that retains the statistical characteristics of the original data, all while safeguarding sensitive information. This approach serves as a means to ensure privacy protection for individuals while enabling the secure utilization of data for research and analytical purposes [105]. A recent study by Torfi et al. has demonstrated how GAN can be leveraged to generate synthetic data that mimics the statistical properties of the real dataset thus preserving data privacy [106]. This development creates new opportunities for private data sharing and analysis, offering insightful information while preserving privacy.

In conclusion, GANs have a wide range of applications across diverse domains, from generating realistic images and movies to aiding in medical diagnosis [1], [6]. The restrictions of data scarcity can be eliminated, and personal information can be safeguarded, by developing synthetic data that closely resembles actual data [107]. As GANs develop further, we can witness more cutting-edge applications in real-data problems [23]. In summary, GANs offer a wide range of applications in a variety of sectors and have the ability to completely change how we produce and use data [108], [109]. Future GAN applications are likely to have even more fascinating uses as the technology develops [110].

V. Variants of GAN

In this section, we will have a broad review of some of the GAN models based on their distinct characteristics and practical uses. Additionally, we discuss the mathematical formulation of these GAN variants, using standard notations as discussed in Sec. III and present their implementation software in Table II.

CGAN. The conditional GAN (CGAN) is a popular version of GAN that generates data by taking external inputs, such as labels or classes, into account. It was introduced by Mirza and Osindero in 2014 [2] and has since been widely used in computer vision applications, including image synthesis, image-to-image translation, and text-to-image synthesis. Unlike the conventional GAN both G and D of the CGAN architecture receive conditional information y that serves as a guide for G to produce data that aligns with the specified conditions. The loss function for the CGAN framework is given by:

$$L = \mathbb{E}_{x\sim p_{\text{data}}}\left[\log D(x,y)\right] + \mathbb{E}_{z\sim p_z}\left[\log(1 - D(G(z,y),y))\right].$$

The CGAN model, as discussed in the literature [2], [111], possesses the following key features:

- CGANs generate customized data that is specific to a given input, e.g., a CGAN trained on animal photos can produce images of a particular animal based on the input.
- Unlike Vanilla GAN, CGAN benefits from additional inputs, resulting in synthetic data of higher quality. It exhibits improved coherence, structure, and aesthetic resemblance to real samples.
- CGANs demonstrate superior noise resistance compared to other artificial neural networks due to the utilization of external input to guide the data generation process.

While the CGAN model is known for its versatility, it is also accompanied by several limitations. It is prone to overfitting with scarce or noisy input data, requires explicit labels or classes in the input dataset, is vulnerable to adversarial attacks, and becomes computationally complex with high-dimensional complex datasets [112]. Considering both the advantages and disadvantages of the CGAN model mentioned above, it proves to be a valuable tool for generating data based on external input [113]. However, it is important to take into account these limitations and drawbacks when applying CGANs to address specific problems. Future research can examine alternative conditioning methods including the use of natural language descriptions or a variety of circumstances [114].

DCGAN. Deep Convolutional GAN (DCGAN) introduced by Radford et al. in 2015 [23] marks a significant breakthrough in the realm of generative AI, particularly for image generation. Representing a specialized variation of the GAN architecture, DCGANs seamlessly combine CNN and GAN techniques to yield high-quality, photorealistic images with intricate details. With the ability to autonomously learn and generate images without additional control, DCGANs prove their usefulness in unsupervised learning scenarios. DCGANs stand out for their relatively manageable training process, owing to sophisticated architectural components like strided convolutions, batch normalization, and leaky Rectified Linear Unit (ReLU) activation functions [23]. From the experimental perspective, DCGANs have generated excellent results for large-scale picture datasets like CIFAR-10 and ImageNet, [115]. Nonetheless, it is worth noting that

DCGANs exhibit elevated computational demands, sensitivity to hyperparameters, and susceptibility to challenges such as restricted diversity of generated images and mode collapse [116]. Despite these limitations, DCGANs find successful applications across domains encompassing image synthesis, style transfer, and image super-resolution. Their far-reaching impact on the field of generative modeling continues to inspire advancements and innovation.

AAEs. Adversarial Autoencoder (AAE) framework, proposed by Makhzani et al. in 2015, is a hybridization of autoencoders with adversarial training [117]. This model has garnered significant attention due to its potential for variational inference by aligning the aggregated posterior of the hidden code vector with a chosen prior distribution. This approach ensures that meaningful outcomes emerge from various regions of the prior space. Consequently, the AAE's decoder acquires the capability to learn a sophisticated generative model, effectively mapping the imposed prior to the data distribution. AAEs excel in producing disentangled representations, showcasing noise resistance, and generating high-quality images. The components within the AAE framework offer notable advantages over alternative generative models. Through adversarial training, AAEs excel in capturing complex data distributions and generating detailed, high-quality images. Their ability to learn disentangled representations in separate latent dimensions empowers precise image control, encompassing alterations to object properties. AAEs exhibit resilience to input variations, making them valuable for noisy data scenarios. Their encoder-decoder design supports denoising and surpasses other models in semi-supervised classification [117]. However, like other generative models, AAEs can encounter mode collapse, demand substantial computational resources, and necessitate cautious hyperparameter tuning. Striking the right balance between adversarial training and autoencoder loss poses a challenge. AAEs lack explicit control over generated samples, hindering targeted data traits in fine-grained control contexts [118]. Yet, the application scope of AAEs is notably expanded by the enhanced encoder, decoder, and discriminator networks, even surpassing traditional autoencoders.

InfoGAN. Information Maximizing Generative Adversarial Network (InfoGAN), a modification of GAN, is designed to learn disentangled representations of data by maximizing the mutual information between a subset of the generator's input and the generated output. It was introduced by Chen et al. in 2016 [14]. The loss function formulation for the Generator in InfoGAN is as follows:

$$L = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1 - D(G(z)))] - \lambda \mathcal{I}(c; G(z)),$$

where $\mathcal{I}(c; G(z))$ is the mutual information between the generator's output $G(z)$ and the learned latent code c, and λ is a hyperparameter that regulates the trade-off between the adversarial loss and the mutual information term. The information-theoretic approach employed in the InfoGAN framework enhances its ability to learn representations that facilitate data exploration, interpretation, and manipulation tasks. Unlike supervised methods, InfoGAN does not rely on explicit supervision or labeling, making it a flexible and scalable option for unsupervised learning tasks like image generation and data augmentation. However, the InfoGAN framework may struggle to learn meaningful and interpretable representations for high-dimensional complex datasets, and its benefits may not always justify the additional complexity and computational cost. Overall, InfoGAN shows promising results in learning disentangled representations, but its effectiveness depends on specific goals, data characteristics, and available resources [119]. Ongoing research and advancements hold the potential to address limitations and further improve this approach in the future.

SAD-GAN. The Synthetic Autonomous Driving using GANs (SAD-GAN) model, introduced by Ghosh et al. in 2016, is designed to generate synthetic driving scenes using the GAN approach [120]. This model's core concept involves training a controller trainer network using images and keypress data to replicate human learning. To create synthetic driving scenes, the SAD-GAN is trained on labeled data from a racing game, consisting of images portraying a driver's bike and its surroundings. A key press logger software is employed to capture key press data during bike rides. The framework's architecture is inspired by DCGAN [23]. The generator takes a current-time input image and produces the subsequent-time synthetic image. Meanwhile, the discriminator receives the real latest-time image, generates its feature map via convolution, and compares real and synthetic scenes to train the generator through a minimax game. The SAD-GAN framework offers an autonomous driving prediction algorithm suitable for manual driving as a recommendation system. Nevertheless, like DCGAN, it requires substantial computation and is susceptible to mode collapse, limiting its real-time applications.

LSGAN. Traditional GAN models typically utilize a discriminator modeled as a classifier with the sigmoid cross entropy loss function. However, this choice of loss function can result in the issue of vanishing gradients during training, resulting in impaired learning of the deep representations. To address this concern, Mao et al. introduced a novel approach called Least Squares GAN (LSGAN) in 2017, which employs the least squares loss function for the discriminator instead [121]. Mathematically, the Generator loss function (L_G) and the Discriminator loss function (L_D) of LSGAN model is expressed as follows:

$$L_G = \frac{1}{2}\mathbb{E}_{z\sim p_z}\left[(D(G(z)) - c)^2\right],$$
$$L_D = \frac{1}{2}\mathbb{E}_{x\sim p_{\text{data}}}(D(x) - b)^2 + \frac{1}{2}\mathbb{E}_{z\sim p_{\mathbf{z}}}(D(G(z)) - a)^2,$$

where a-b encoding scheme represents the labels for fake data and real data for D, and c denotes the values that G wants D to believe for fake data. The LSGAN framework represents a notable advancement over traditional GANs, offering

improved stability and convergence during training while generating higher-quality synthetic data. It has outperformed regular GANs in generating realistic images, as measured by Inception score, across various datasets such as CIFAR-10 [121]. However, LSGANs often produce fuzzy images due to the use of squared loss in the objective function. The generated images often lack sharpness and fine details, as the loss function penalizes large discrepancies between fake and real images but neglects smaller variations. Researchers have addressed this issue by modifying the loss function in subsequent studies, aiming to enhance the sharpness of synthetic images [122], [123]. While LSGANs show promise in generating high-quality images, ongoing research and development are focused on overcoming their limitations in producing crisp and detailed results.

SRGAN. Super Resolution GAN (SRGAN), introduced by Ledig et al. in 2017, is a GAN-based framework for image super-resolution [124]. It generates high-resolution images from low-resolution inputs with an upscaling factor of 4 using a generator network and a discriminator network. To achieve super-resolution, SRGAN incorporates a perceptual loss function, combining content and adversarial losses. Mathematically, the perceptual loss is expressed as:

$$l^{\text{SR}} = l_x^{\text{SR}} + 10^{-3} l_{\text{Gen}}^{\text{SR}},$$

where l_x^{SR} represents the content loss and $l_{\text{Gen}}^{\text{SR}}$ is the adversarial loss. The content loss used in the SRGAN framework relies on a pre-trained VGG-19 model and it provides the network information regarding the quality and content of the generated image. On the other hand, the adversarial loss is responsible for ensuring the generation of realistic images from the generator network. SRGANs offer the ability to generate high-quality images with enhanced details and textures, resulting in improved overall image quality. They excel in producing visually appealing and realistic images, as confirmed by studies on perceptual quality [65]. SRGANs exhibit noise resistance, enabling them to handle low-quality or noisy input images while still delivering high-quality outputs [125]. Moreover, this model demonstrates flexibility and applicability across various domains, including video processing, medical imaging, and satellite imaging [124]. However, training SRGANs can be computationally expensive, especially for complex models or large datasets. Additionally, like other GANs, the interpretability of SRGANs can be challenging, making it difficult to understand the underlying learning process of the generator. Furthermore, while SRGANs excel in image synthesis, they may not perform as effectively with text or audio inputs, limiting their range of applications.

WGAN. The Wasserstein GAN (WGAN), introduced by Arjovsky et al. in 2017, is a loss function optimization variant of GAN that improves training stability and mitigates mode collapse [15]. It employs the Wasserstein distance to enhance realistic sample generation and ensure meaningful gradients. By introducing a critic network and weight clipping, WGAN achieves training stability. It finds applications in image synthesis, style transfer, and data generation. The formulation of the WGAN framework utilizes the Wasserstein-1 distance or the Earth Mover distance to measure the distance between real and generated data distributions. Mathematically, the Wasserstein distance for transforming the distribution $\mathbb{P}$ to distribution $\mathbb{Q}$ can be expressed as:

$$W(\mathbb{P}, \mathbb{Q}) = \inf_{\theta \in \pi(\mathbb{P}, \mathbb{Q})} \mathbb{E}_{(\tilde{X}, \tilde{Y}) \sim \theta} \left[\|\tilde{X} - \tilde{Y}\| \right].$$

In the WGAN model, the discriminator function D is designed as a critic network that estimates the Wasserstein distance between the real data distribution and the generated data distribution instead of probability values as in conventional GAN. These scores reflect the degree of similarity or dissimilarity between the input sample and the real data distribution. The training of the critic in WGAN involves optimizing its parameters to maximize the difference in critic values between real and generated samples. By clipping the discriminator weights, the discriminator loss function in WGAN is adjusted to enforce the Lipschitz continuity requirement, but the fundamental structure of the loss functions is maintained. In general, WGANs have demonstrated improved training stability compared to traditional GANs. They are less sensitive to hyperparameters and more resistant to mode collapse [122]. The use of the Wasserstein distance facilitates smoother optimization and better gradient flow, resulting in faster training and higher-quality samples. However, calculating the Wasserstein distance can be computationally expensive [126]. Although WGANs offer enhanced stability, careful tuning of hyperparameters and network designs is still necessary for satisfactory results. Furthermore, WGANs are primarily suited for generating images and may have limited applicability to other types of data. In summary, WGANs represent a promising advancement in the field of GANs, addressing their limitations and providing insights into distribution distances, but the applicability of WGANs to real-world problems requires careful consideration of its challenges.

CycleGAN. Cycle-Consistent GAN (CycleGAN), introduced by Zhu et al. in 2017, is an unsupervised image-to-image translation framework that eliminates the need for paired training data unlike traditional GANs [3]. It relies on cycle consistency, allowing images to be translated between two domains using two generators and two discriminators while preserving coherence. One generator G_{XY} translates images from the source domain X to the target domain Y, and the other G_{YX} performs the reverse. In other words the function G_{YX} is such that $G_{YX}(G_{XY}(x)) = x$. The discriminators, on the other hand, distinguish between real and translated images generated by the generators. To train this architecture the cycle consistency loss of Cycle GAN plays a crucial role by enforcing consistency between the original and round-trip translated images, the so-called *forward* and *backward* consistency. This ensures generators produce meaningful translations, preserving important content and characteristics across domains. Mathematically, the cycle consistency loss

function can be expressed as:

$$\mathcal{L}_{\text{cycle}}(G_{XY}, G_{YX}) = \mathbb{E}_{x\sim p_{data}}[\|G_{YX}(G_{XY}(x)) - x\|_1] + \mathbb{E}_{y\sim p_{data}}[\|G_{XY}(G_{YX}(y)) - y\|_1].$$

The main advantage of Cycle GAN lies in its ability to produce high-quality images with remarkable visual fidelity. It excels in various image-to-image translation tasks, including style transfer, colorization, and object transformation. Moreover, its computational efficiency allows training on large datasets. However, CycleGAN often suffers from mode collapse and the increasing amount of parameters reduces its efficiency [127]. Despite its limitations, CycleGAN remains a valuable tool for image translation, and ongoing research for any data translation task aims to address its shortcomings [128]. For example, it shows promising results in medical imaging domain adaptation [129].

ProGAN. In 2017, Karras et al. introduced the Progressive Growing of GAN (ProGAN), addressing the limitations of traditional GANs such as training instability and low-resolution output [5]. ProGAN utilizes a progressive growth technique, gradually increasing the size and complexity of the generator and discriminator networks during training. This incremental approach enables the model to learn coarse characteristics first and subsequently refine them, ultimately producing high-resolution images. By starting with low-resolution image generation and progressively adding layers and details, ProGAN achieves training stability and generates visually realistic images of superior quality. This technique has found successful applications in various domains, including image synthesis, super-resolution, and style transfer. During training, the resolution of the generated images is increased progressively from a low resolution (e.g., 4x4) to a high resolution (e.g., 1024x1024). At each resolution level, the generator and discriminator networks are updated using a combination of loss functions. Progressive updates at increasing resolutions ensure high-quality image synthesis with fine features and textures throughout training, unlike the conventional GAN framework. ProGAN offers better scalability, enabling the generation of images at any resolution. It exhibits improved stability during training, overcoming issues like mode collapse. The flexibility of ProGAN makes it suitable for various image synthesis applications, including satellite imaging, video processing, and medical imaging [5]. However, training ProGAN can be computationally expensive, especially for large datasets or complex models. Interpretability may pose challenges, as with other GANs, making it difficult to discern the learned representations. Additionally, ProGAN's generalization to new or unexplored data may be limited, requiring further fine-tuning or training on fresh datasets [130].

MidiNet. MidiNet, proposed by Yang et al. in 2017, attempts to generate melodies or a series of MIDI notes in the symbolic domain [8]. Unlike other music generation frameworks, such as WaveNet [131], and Song from PI [132], the MidiNet model can generate melodies either from scratch or by combining the melodies of previous bars. The architectural configuration of the MidiNet framework is motivated by the DCGAN model [23]. The MidiNet model combines a CNN generator with a conditioner CNN in the first phase of training. While the former CNN is employed to generate synthetic melodies based on the random noise vector, the latter provides the available prior knowledge about other melodies in the form of an encoded vector as an optional input to the generator. Once the melody is generated it is processed with a CNN-based discriminator which consists of a few convolutional layers and a fully connected network. The discriminator is optimized using a cross-entropy loss function to efficiently detect whether the input is a real or a generated one. For training the overall network in MidiNet, the minimax loss function is combined with feature mapping and one-sided label smoothing to ensure learning stability and versatility in the generated content. The MidiNet framework proposes a unique CNN-GAN structure for the generation of symbolic melodies. Its ability to synthesize artificial music in the presence or absence of prior knowledge is very useful in the audio domain. However, due to the use of a CNN-based structure, its computational complexity significantly increases in comparison to the standard GAN model. Further research in this domain is required to understand the capabilities of MidiNet in multi-track music generation while simultaneously reducing its running time.

SN-GAN. Spectral Normalization GAN (SN-GAN) is a GAN variant that utilizes spectral normalization to stabilize the training of the generator and discriminator networks [133]. In conventional GANs, training can be unstable due to a powerful discriminator or poor-quality generator samples. SN-GAN addresses this by constraining the Lipschitz constant of the discriminator, preventing it from dominating the training process. Spectral normalization normalizes the discriminator's weight matrices, ensuring a stable maximum value and preventing the amplification of minor input perturbations. SN-GAN produces high-quality samples with improved stability and convergence compared to traditional GANs. The adversarial training process used in the SN-GAN framework, similar to the conventional GAN (as in Eq. 1), encourages G to produce more realistic samples that can fool D, while D learns to accurately distinguish between real and generated samples. Several benefits of the SN-GAN model over the standard GAN include increased stability in training the generator and discriminator by constraining the Lipschitz constant of the discriminator. This mitigates issues like gradient explosion and mode collapse, resulting in high-quality examples with fine features and edges. SN-GAN is relatively simple to implement and can be integrated into existing GAN systems. However, the computation of singular values during the normalization process adds to the computational burden, potentially extending training time and requiring more memory. SN-GAN's reliance on the spectral norm assumption of discriminator weights may limit its applicability to specific GAN architectures. While SN-GANs may exhibit slower convergence and reduced sample diversity compared to conventional GANs, they excel in stability and

sample quality.

RGAN. Relativistic GAN (RGAN) introduces a relativistic discriminator to enhance the stability and quality of GAN-generated samples [134]. Unlike traditional GANs, where the discriminator determines if a sample is real or fake, the RGAN discriminator estimates the probability that a genuine sample is more realistic than a fake sample, and vice versa. It compares the likelihood of a true sample being real with the likelihood of a fake sample being real. This approach guides the generator to produce samples that are more realistic than the discriminator's current estimates for both real and fake samples. To ensure this relativistic nature of RGAN, samples are considered from both real and fake data pairs $\tilde{x} = (x_R, x_F)$, where $x_R \sim \mathbb{P}_{\text{Real}}$ represents the real data and $x_F \sim \mathbb{P}_{\text{Fake}}$ symbolize its fake counterpart. Mathematically, the generator and discriminator loss functions of the RGAN framework can be expressed as:

$$\begin{aligned} L_G =& \mathbb{E}_{(x_R,x_F)\sim(\mathbb{P}_{\text{Real}},\mathbb{P}_{\text{Fake}})}\left[\tilde{g}_1\left(C\left(x_R\right)-C\left(x_F\right)\right)\right] \\ &+ \mathbb{E}_{(x_R,x_F)\sim(\mathbb{P}_{\text{Real}},\mathbb{P}_{\text{Fake}})}\left[\tilde{g}_2\left(C\left(x_F\right)-C\left(x_R\right)\right)\right] \text{ and} \\ L_D =& \mathbb{E}_{(x_R,x_F)\sim(\mathbb{P}_{\text{Real}},\mathbb{P}_{\text{Fake}})}\left[\tilde{f}_1\left(C\left(x_R\right)-C\left(x_F\right)\right)\right] \\ &+ \mathbb{E}_{(x_R,x_F)\sim(\mathbb{P}_{\text{Real}},\mathbb{P}_{\text{Fake}})}\left[\tilde{f}_2\left(C\left(x_F\right)-C\left(x_R\right)\right)\right], \end{aligned}$$

where $C(\cdot)$ is the non-transformed layer and $\tilde{g}_1, \tilde{g}_2, \tilde{f}_1, \tilde{f}_2$ are scalar-to-scalar functions. The term $\left(C\left(x_F\right)-C\left(x_R\right)\right)$ of the modified loss function can be interpreted as the likelihood that the given fake data is more realistic than randomly sampled real data. The relativistic discriminator in RGAN enhances stability by mitigating issues like mode collapse and vanishing gradients, commonly observed in conventional GANs [134]. RGAN surpasses regular GANs in generating high-quality samples. It also exhibits improved resilience against adversarial attacks, ensuring sample security. However, these advantages come at the expense of higher computational requirements compared to regular GANs owing to the use of relativistic discriminator [126]. Additionally, RGAN necessitates careful hyperparameter tuning, including learning rate and regularization parameters, for optimal performance [135]–[137]. Furthermore, the efficacy of RGAN depends on the specific use case, limiting its universal applicability.

StarGAN. StarGAN, a type of GAN model introduced in the work of Choi et al. [138], is specifically designed for multi-domain image-to-image translations. In contrast to the CycleGAN model [3] that focuses on translating images between two specific domains, StarGAN offers the capability to perform translations across a diverse range of domains using a single generator and discriminator. This model trains the generator network G to map the input image x to an output image y conditioned on the randomly generated target domain label c i.e., $G(x, c) \longrightarrow y$. In case of the discriminator network D an additional classifier is used to produce the probability distribution for both source and domain labels $D : x \longrightarrow \{D_{\text{src}(x)}, D_{\text{cls}(x)}\}$. To ensure an efficient multi-domain image translation this framework utilizes several loss functions namely, the adversarial loss, the domain classification loss, and the reconstruction loss. The conventional adversarial loss ensures the generation of high-quality realistic images. The domain classification loss of real images optimizes D to accurately classify x to their input domain label c', whereas, the domain classification loss of fake images optimizes G to generate images that can be classified as the generated target domain c. Overall, the domain classification loss ensures the coherent multi-domain image classification in the StarGAN model. Furthermore, to ensure that the translated images retain the characteristics of the input image and exclusively modify the domain-related features, a reconstruction loss is used in training the generator network. The overall objective function of the StarGAN model is mathematically expressed as:

$$\begin{aligned} L_G = & \mathbb{E}_x\left[\log D_{src}(x)\right] + \mathbb{E}_{x,c}\left[\log\left(1 - D_{src}(G(x,c))\right)\right] \\ & - \lambda_1 \mathbb{E}_{x,c}\left[-\log D_{cls}(c \mid G(x,c))\right] \\ & + \lambda_2 \mathbb{E}_{x,c,c'}\left[\left\|x - G\left(G(x,c),c'\right)\right\|_1\right] \text{ and} \\ L_D = & -\mathbb{E}_x\left[\log D_{src}(x)\right] - \mathbb{E}_{x,c}\left[\log\left(1 - D_{src}(G(x,c))\right)\right] \\ & - \mathbb{E}_{x,c'}\left[\log D_{cls}\left(c' \mid x\right)\right], \end{aligned}$$

where λ_1 and λ_2 are the hyper-parameters that control the effect of the domain classification loss and the reconstruction loss in the StarGAN model, respectively. The training process involves iteratively optimizing the components of the loss functions to achieve high-quality multi-domain image-to-image translations. The StarGAN framework offers several advantages in multi-domain image translation tasks. It utilizes a single generator-discriminator network for all domains, reducing computational complexity. StarGAN can effectively learn domain mappings with limited or unpaired data and preserve the identity of input images in the same target domain. However, it has several drawbacks, including a complex loss function that leads to a time-consuming training process [139], [140]. Additionally, regulating image quality and handling translations between complex domains with significant appearance or structural changes can be challenging in StarGAN [141]. Moreover, this model can be used to manipulate images to a considerable extent which might lead to ethical concerns [142].

BigGAN. BigGAN, introduced by Brock et al. in 2018, is an innovative methodology for training GAN on a large scale to achieve a high-quality synthesis of natural images [110]. It aims to address the challenge of generating high-quality images with high resolutions, which traditional GANs struggle to achieve [33]. BigGAN stands out by employing large-scale architecture and a unique truncation technique that allows for the generation of high-fidelity images with intricate details and textures. The model is capable of producing images of various resolutions, reaching up to 512 $\times$ 512 pixels, and has been trained on a substantial dataset of images. Similar to GAN (as in Eq. 1), during the training of BigGAN model gradient descent techniques are used to update the parameters of G and D. The discriminator aims to maximize the objective, while the generator aims to minimize it. BigGAN introduces architectural modifications to enhance image quality and diversity. It incorporates class-conditional GANs and self-attention mechanisms. Regularization techniques like

orthogonal regularization and truncation tricks stabilize and control the generator's output. Data augmentation methods, such as progressive resizing and interpolation, are employed to handle high-resolution images effectively. The modified training approach in the BigGAN architecture enables the generation of high-quality images with detailed features and textures, surpassing the capabilities of regular GANs. This enhanced model offers scalability, addresses mode collapse issues, and has broad applications in fields such as video processing, satellite imaging, and medical imaging. However, it is computationally demanding, especially when dealing with large datasets or complex models [143], [144]. Additionally, the generalization of the framework to new, unseen data is limited, requiring further fine-tuning or training on fresh datasets [145].

MI-GAN. In the field of deep learning, constrained data sizes within the medical domain pose a significant challenge for supervised learning tasks, elevating concerns about overfitting. To address this, Iqbal et al. introduced Medical Imaging GAN (MI-GAN) in 2018, an innovative GAN framework tailored for Medical Imaging [146]. MI-GAN is specialized in generating synthetic retinal vessel images along with segmented masks based on limited input data. The architecture of the MI-GAN framework's generator network adopts an encoder-decoder structure. Given a random noise vector, the encoder functions as a feature extractor, capturing local and global data representations through its fully connected neural network design. These learned representations are then channeled into the decoder using skip connections, facilitating the generation of segmented images. The generator's enhancements encompass the integration of global standard segmented images and style transfer mechanisms, refining the segmented image generation process. Consequently, the modified MI-GAN generator is trained using a blend of adversarial, segmentation, and style transfer loss functions. In contrast, the discriminator network within the MI-GAN model consists of multiple convolutional layers, and it is trained using adversarial loss functions to effectively distinguish between real and generated images. MI-GAN refines the conditional GAN model for retinal image synthesis and segmentation. Remarkably, despite being trained with a mere ten real examples, this model holds tremendous potential in medical image generation. Nonetheless, this approach relies on spatial alignment to achieve superior outcomes, which can often be scarce [147].

AttGAN. AttGAN, also known as Attribute GAN, is a variation of the GAN framework that focuses on generating images with customizable properties such as age, gender, and expression. It was introduced by He et al. in 2019 in their work "AttGAN: Facial Attribute Editing by Only Changing What You Want" [148]. AttGAN aims to allow users to modify specific facial attributes while preserving the overall identity and appearance of the face. By manipulating attribute vectors, users can control the desired changes in the facial attributes, resulting in realistic and visually appealing image transformations. The AttGAN framework combines two sub-networks an encoder G_{Enc} and a decoder G_{Dec} in place of G of conventional GAN and it utilizes an attribute classifier C with the discriminator network. During the training phase, given an input image $x^{\tilde{a}}$ with a set of n-dimensional binary attribute $\tilde{a}$, G_{Enc} encodes $x^{\tilde{a}}$ into a latent vector representation i.e., $s = G_{\text{Enc}}\left(x^{\tilde{a}}\right)$. Simultaneously, G_{Dec} is employed for editing the attributes of $x^{\tilde{a}}$ to another set of n-dimensional attributes $\tilde{b}$ i.e., the edited image $x^{\hat{b}}$ is constructed as $x^{\hat{b}} = G_{\text{Dec}}\left(s, \tilde{b}\right)$. To perform this unsupervised learning task C is used with the encoder-decoder pair to constrain $x^{\hat{b}}$ to possess the desired qualities. Moreover, the adversarial loss used in the training process ensures realistic image generation. On the other hand, to allow for satisfactory preservation of attribute-excluding details in the network a reconstruction loss is utilized in the framework. This loss ensures that the interaction between the latent vector s with attribute $\tilde{b}$ will always produce $x^{\hat{b}}$ and the interaction between s with attribute $\tilde{a}$ will always produce $x^{\hat{a}}$, approximating the input image $x^{\tilde{a}}$. Thus the overall loss function for the encoder-decoder-based generator of AttGAN can be expressed as:

$$L_{\text{Enc, Dec}} = \lambda_{\text{Rec}}\mathbb{E}_{x^{\tilde{a}}}\left[\|x^{\tilde{a}} - x^{\hat{a}}\|_1\right] + \lambda_{\text{Cls}_G}\mathbb{E}_{x^{\tilde{a}},\tilde{b}}\left[\text{H}\left(\tilde{b}, C(x^{\hat{b}})\right)\right] - \mathbb{E}_{x^{\tilde{a}},\tilde{b}}\left[D\left(x^{\hat{b}}\right)\right]$$

and the loss for the classifier and the discriminator is formulated as:

$$L_{\text{D, Cls}} = \lambda_{\text{Cls}_D}\mathbb{E}_{x^{\tilde{a}}}\left[\text{H}\left(\tilde{a}, C(x^{\tilde{a}})\right)\right] - \mathbb{E}_{x^{\tilde{a}}}\left[D\left(x^{\tilde{a}}\right)\right] + \mathbb{E}_{x^{\tilde{a}},\tilde{b}}\left[D\left(x^{\hat{b}}\right)\right],$$

where H is the cross entropy loss, and $\lambda_{\text{Rec}}, \lambda_{\text{Cls}_G}, \lambda_{\text{Cls}_D}$ are hyperparameters for balancing the losses. AttGAN offers several benefits in the image generation domain including precise control over the attributes of generated images, allowing users to modify age, gender, expression, and other qualities. It provides flexibility by adapting to multiple domains and tasks, enabling customization and flexibility in image synthesis applications. The model produces realistic images that approximate the desired attributes while maintaining the visual aspects of the original image. However, ethical considerations regarding representation, identity, and privacy must be addressed when using AttGAN or similar models [17], [149]. The computational complexity of AttGAN requires significant resources and may pose challenges for deployment in production settings or on resource-limited devices. Additionally, AttGAN relies on labeled data with attribute annotations, which may not always be readily available, and the performance and generalizability of the model can be influenced by the quantity and quality of the attribute annotations [150]. The distribution and diversity of the training data can also impact the model's performance and ability to handle uncommon or out-of-distribution features [151]. In conclusion, AttGAN provides precise attribute control, flexibility, and realistic image generation capabilities, but careful ethical considerations, resource requirements, and data dependencies should be taken into account when utilizing the model in practical applications.

DM-GAN. The Dynamic Memory GAN (DM-GAN) introduced by Zhu et al. in 2019 combines the power of GANs with a memory-augmented neural network design to overcome the limitations of conventional GANs [152], [153]. By addressing issues like mode collapse and lack of fine-grained control, DM-GAN aims to improve the image synthesis process. This deep learning model focuses on generating realistic images from text descriptions, tackling two main challenges in existing methods. Firstly, it addresses the impact of initial image quality on the refinement process, ensuring satisfactory results. Secondly, DM-GAN considers the importance of each word in conveying image content by incorporating a dynamic memory module. The two-stage training of the DM-GAN framework initially transforms the textual description into an internal representation using a text encoder and a deep generator model is utilized to generate an initial image based on the encoded text and random noise. In the subsequent dynamic memory-based image refinement step the generated fuzzy image is processed using a memory writing gate to select relevant text information based on the initial image content and a response gate to fuse information from memories and image features. These advancements enable DM-GAN to generate high-quality images from text descriptions accurately. The dynamic memory module of DM-GAN enhances image generation by capturing long-range relationships and maintaining global context, resulting in persuasive and visually appealing images. It provides fine-grained control over attribute-guided synthesis and increases diversity by addressing mode collapse. However, DM-GAN's computational complexity and memory management pose challenges, and it relies on labeled data [154], [155]. The model's interpretability is limited due to the complexity of the memory module [156], [157]. In conclusion, DM-GAN offers enhanced image generation capabilities with control, diversity, and robustness, while considerations such as computational resources, data availability, and interpretability should be considered.

SinGAN. Single-Image GAN (SinGAN) is an unconditional generative model introduced by Shaham, et al. in 2019 for learning the internal statistics from a single image without the need for additional training data [158]. SinGAN allows for a wide range of image synthesis and manipulation tasks, including animation, editing, harmonization, and super-resolution, among many others. The key innovation of SinGAN is the use of a multi-scale pyramid of GANs, where each GAN is responsible for generating images at a different scale. This hierarchical structure enables SinGAN to capture both the global and local characteristics of the input image, resulting in high-quality and coherent output images. By training on a single image, SinGAN eliminates the need for a large dataset, making it a versatile and practical tool for image generation tasks. During the training phase of SinGAN, a hierarchical structure called the multi-scale pyramid is utilized. This pyramid consists of a series of generators denoted as $\{G_0, G_1, \ldots, G_N\}$. The generators take input patches of the image at different downsampled levels, represented as $\{x_0, x_1, \ldots, x_N\}$, where each level is downsampled by a factor of r^n ($r > 1$). The generators, along with their corresponding discriminators D_n, are trained using adversarial training. The goal is to generate realistic samples that cannot be distinguished from the downsampled image x_n. The SinGAN architecture consists of 5 convolutional blocks in both G_n and D_n networks. Each block consists of a 3×3 convolutional layer with 32 kernels, followed by batch normalization and LeakyReLU activation. The patch size for the discriminator remains fixed at 11×11 across all pyramid levels. During training, the generator and discriminator networks are iteratively updated to optimize a combination of adversarial loss and reconstruction loss. As the training progresses to higher pyramid levels, the generator incorporates the output from the previous level, enabling it to capture finer details and generate more realistic images. To enhance the model's ability to handle diverse variations, noise injection is introduced during training, where random noise patterns are added to the input image at each scale. This helps in generating diverse outputs. The training process continues until convergence, where the generator is capable of synthesizing images that closely resemble the training image at all scales of the pyramid.

SinGAN offers numerous advantages in image manipulation tasks, requiring minimal data. It enables controlled alteration, synthesis, and modification of images, allowing users to adjust lighting, colors, textures, and objects. The model produces aesthetically realistic and visually consistent results that align with the input image. Its multi-stage training process captures global and local characteristics, resulting in high-quality outputs. However, SinGAN lacks explicit control over specific image traits and quality depends on input image quality and quantity [159]. Ethical considerations should be addressed, and the model is computationally complex with limited interpretability [160]. Nevertheless, SinGAN's multi-stage training has gained popularity due to its versatility and the powerful image generation capabilities it offers.

PATE-GAN. In our data-centric world, safeguarding data privacy holds paramount importance, ensuring the protection of individual rights, ethical data handling, and the establishment of a reliable digital environment. It ensures a harmonious blend of leveraging the benefits of data-driven technologies while respecting individual's autonomy and rights. To uphold these concerns and to enable the ethical usage of real-world data in various machine-learning frameworks, Jordan et al. in 2019 proposed the Private Aggregation of Teacher Ensembles Generative Adversarial Network (PATE-GAN) framework [161]. Combining the differential privacy principles of Private Aggregation of Teacher Ensembles (PATE) with the generative prowess of GANs, PATE-GAN generates synthetic data for training algorithms while aiming for a positive societal impact. Similar to the conventional GAN model, PATE-GAN comprises of a generator network that receives a latent vector as input and provides generated data as an output. However, in the discriminator aspect, PATE-GAN innovatively integrates the PATE mechanism involving multiple teacher discriminators and a single student discriminator. The teacher discriminators classify real and generated samples within their dataset

segments, while the student discriminator employs the labels aggregated from the teacher discriminators to classify generated samples. The framework's training employs an asymmetric adversarial process, where teachers aim to enhance their loss relative to the generator, the generator targets the student's loss, and the student seeks to optimize its loss against the teachers. This arrangement with the student discriminator ensures differential privacy concerning the original dataset.

POLY-GAN. Introduced by Pandey et al. in 2020, Poly-GAN is a novel conditional GAN architecture aimed at fashion synthesis [95]. This architecture is designed to automatically dress human model images in diverse poses with different clothing items. Poly-GAN employs an encoder-decoder structure with skip connections for tasks like image alignment, stitching, and inpainting. The training procedure of the Poly-GAN framework consists of four steps. This model takes input images, including a reference garment and a model image for clothing placement. Initially, pre-processing involves using a pre-trained LCR-Net++ pose estimator [162] to extract the model's pose skeleton and a U-Net++ segmentation network [125], [163] to obtain the segmented mask of the old garment from the model image. The Poly-GAN pipeline begins by passing the reference garment and generated RGB pose skeleton through the generator to create a garment image that aligns with the skeleton's shape. The architecture of G follows an encoder-decoder structure. The encoder incorporates three components: a Conv module for propagating pose skeleton information at each layer, a ResNet module for generating a feature vector [164], and a Conv-norm module with two convolutional layers to process the other two modules' outputs. On the other hand, the decoder learns to produce the desired garment image based on pose condition embedding sent by the encoder using skip connections. The transformed garment image and segmented pose skeleton are sent as inputs to the second stage of the network for image stitching, yielding an image of the pose skeleton with the reference attire. In the third stage, the model performs inpainting to eliminate any irregularities in the generated model image. The discriminator, similar in structure to SR-GAN [124], is employed during these stages to differentiate real from fake images. Finally, in the fourth stage, post-processing is applied, stitching the model's head to the image to produce the final output. The Poly-GAN framework utilizes adversarial, GAN, and identity losses for training, ensuring high image quality and minimizing texture and color discrepancies from real images. Poly-GAN presents an advancement in fashion synthesis compared to other models [165], as it operates with multiple conditional inputs and achieves satisfactory fitting results without requiring 3D model information [166]. However, the generated images can exhibit texture deformation and body part loss, affecting the fitting outcomes [167]. Further research is needed to address these issues in this domain.

MIEGAN. Mobile Image Enhancement GAN (MIEGAN), introduced by Pan et al. in 2021, is a novel approach within the realm of GAN-based architectures, with the primary objective of elevating the visual caliber of images taken via mobile devices [168]. This endeavor involves several modifications to the conventional GAN architecture. In the MIEGAN model, a multi-module cascade generative network is utilized which combines an Autoencoder and a feature transformer. The encoder of this modified generator comprises of two streams with the second stream being responsible for enhancing the regions with low luminance - a common issue in mobile photography leading to reduced clarity. In the feature transformative module, the local and global information of the image is further captured using a dual network structure. Furthermore, to enhance the generative network's ability to produce images of superior visual quality, an adaptive multi-scale discriminator is employed in lieu of a standard single discriminator in the MIEGAN model. This multi-scale discriminator serves to differentiate between real and fake images on both global and local scales. To harmonize the evaluations from the global and local discriminators, an adaptable weight allocation strategy is utilized in the discriminator. Additionally, this model is trained based on a contrast loss mechanism and a mixed loss function, which further enhances the visual quality of the generated images. Despite the image quality enhancement capabilities of the MIEGAN framework, their high computation complexity poses a significant challenge for their real-time application in mobile photography.

VQGAN. Vector Quantized GAN (VQGAN) introduces a novel methodology that merges the capabilities of GAN with vector quantization techniques to generate high-quality images [169]. This approach effectively leverages the synergies between the localized interactions of CNN and the extended interactions of Transformers [19] in tasks involving the conditional synthesis of data. The distinctive architecture of VQGAN not only yields images of exceptional quality but also empowers a degree of creative influence, enabling the manipulation of various attributes within the generated content. The training process of the VQGAN architecture unfolds in two pivotal phases. Initially, a variational autoencoder and decoder are trained, as opposed to the conventional GAN generator network. This training aims to reconstruct the image by utilizing a discrete latent vector representation derived from the input image. This intermediate representation is subsequently linked to a codebook, efficiently capturing the underlying semantic information. To augment the fidelity of the reconstructed image, a discriminator is incorporated into the autoencoder structure. The training of the autoencoder model, the codebook, and the discriminator involves optimizing a fusion of adversarial loss and perceptual loss functions. In the subsequent phase, the codebook indices, constituting the intermediate image representations, are fed into Transformers. These Transformers are trained through a transformer loss mechanism, guiding them to predict the succeeding indices within the encoded sequence, resulting in an improved codebook representation. Finally, the information from the codebook is utilized by the decoder to generate images of

higher resolutions. The unique aspect of VQGAN lies in its ability to allow users to manipulate generated images in creative ways. By modifying the quantized codes, users can control specific features of the generated content, thereby unlocking a spectrum of artistic potentials. Nonetheless, the caliber of the images generated by VQGAN depends largely on its input data, necessitating expansive datasets and substantial computational resources to produce images of exceptional excellence [170]. Consequently, this restricts its immediate applicability in real-time case studies. Moreover, the codebook representation used in the vector quantization process can significantly reduce the variation in the generated images [171].

DALL-E. DALL-E is an advanced text-to-image generative framework created by OpenAI that utilizes a two-stage process to generate images from textual prompts [172], [173]. It combines the concepts of GANs and Transformers to generate highly realistic and coherent images from textual descriptions. What sets DALL-E apart is its ability to generate realistic art and images from textual descriptions that may describe completely novel concepts or objects. The working principle of the pre-trained DALL-E model comprises of two phases. The first stage involves a prior model that generates a Contrastive Language-Image Pretraining (CLIP) [174] image embedding, capturing the essential gist of the image based on the provided caption. In the second stage, a decoder model known as GLIDE takes the image embedding and reconstructs the image itself, gradually removing noise and generating a realistic and visually coherent image. The CLIP model, consisting of a text encoder and an image encoder, is trained using contrastive training to learn the relationship between images and their corresponding captions. This allows the model to generate the CLIP text embedding from the input caption. Further, the prior model of DALL-E processes this text representation to generate the CLIP image embedding. In case of the decoder, DALL-E utilizes a Diffusion model [22] which generates the image by using CLIP image embedding and the CLIP text embedding as an additional input. DALL-E's two-stage process offers advantages in prioritizing high-level semantics and enabling intuitive transformations. It excels in generating creative and imaginative images based on textual descriptions, making it valuable for creative tasks. However, training DALL-E requires substantial computational resources and presents challenges in fine-tuning and attribute control. Ethical concerns and biases surrounding AI-generated content also arise [175], [176]. Moreover, the lack of interpretability and explainability of this framework restricts its applications in legal, medical, or safety-sensitive domains [177]. Nevertheless, DALL-E represents a significant advancement in image synthesis and has garnered attention for its creative potential. Ongoing research, such as DALL-E 2 [178], continues to push the boundaries of this field and attempts to mitigate the explainability concerns [179].

CEGAN. Class imbalance is a prevalent challenge across many real-world datasets. In the context of classification tasks, this skewed distribution of classes leads to a significant bias favoring the majority class. Previous studies have suggested oversampling approaches, involving the artificial generation of samples from the minority class, as an efficient mechanism to mitigate this issue. Classification Enhancement GAN (CEGAN) model introduces a solution to address the class imbalance issue through the utilization of a GAN-based framework, as outlined in the work by Suh et al. [99]. This model particularly focuses on enhancing the quality of data generated from the minority class, thereby mitigating the classifier's bias towards the distribution of the majority class. Differing from the conventional GAN model, the CEGAN framework combines three distinct networks – a generator, a discriminator, and a classifier. The training process of the CEGAN model involves a two-step sequence. In the initial phase, the generator generates synthetic data using input noise and real class labels. Simultaneously, the discriminator distinguishes between real and synthetic data, while the classifier assigns class labels to input samples. The subsequent stage involves the integration of the generated samples with the original training data, creating an augmented dataset for training the classifier. The CEGAN framework serves as an efficient methodology that incorporates techniques such as data augmentation, noise reduction, and ambiguity reduction to effectively tackle class imbalance problems. Notably, this approach overcomes the limitations associated with traditional resampling techniques, as it avoids the need to modify the original dataset.

SeismoGen. Seismogen is a seismic waveform synthesis technique that utilizes GAN for seismic data augmentation [87]. The motivation behind Seismogen arises from the need for abundant labeled data for accurate earthquake detection models. To overcome the scarcity of seismic waveform datasets, Wang et al. introduced the Seismogen framework, employing GAN to generate realistic multi-labeled waveform data based on limited real seismic datasets. Incorporating this additional dataset enhances the training of machine learning-based seismic analysis models, leading to more robust predictions for out-of-sample datasets. The mathematical formulation of the Seismogen framework follows the Wasserstein GAN [109] framework and can be expressed as:

$$
\begin{aligned}
L_G &= - \mathop{\mathbb{E}}_{z\sim \mathrm{N}(0,1)} D(G(z)), \\
L_D &= \mathop{\mathbb{E}}_{z\sim \mathrm{N}(0,1)} D(G(z)) - \mathop{\mathbb{E}}_{x\sim p_{\text{data}}} D(x) \\
&\quad + \lambda \mathop{\mathbb{E}}_{z\sim \mathrm{N}(0,1)} \left[\left(\|D(G(z))\|_2 - 1 \right)^2 \right],
\end{aligned}
$$

where the noise z is a standard normal variable and λ is a hyperparameter. The primary objective is to minimize the difference between the true seismic waveforms and the synthetic waveforms generated by the Seismogen. This is achieved by iteratively optimizing L_G and L_D to find an equilibrium between the generator and discriminator networks. SeismoGen has demonstrated its ability to generate highly realistic seismic waveforms, making it valuable for seismic waveform analysis and data augmentation. Its conditional generation feature allows users to produce waveforms labeled with specific categories, enhancing its versatility for various

applications. SeismoGen is scalable and capable of generating large databases of artificial waveforms, which is beneficial for tasks requiring extensive training data. However, SeismoGen's effectiveness is influenced by the quality and distribution of the training data. It does not model the expected waveform move-out, which is relevant in various seismic research. Additionally, due to imbalanced real seismic waveform datasets, SeismoGen struggles to generate data with rare characteristics. Moreover, the computational cost of training and using SeismoGen may be a limiting factor, especially for real-time seismic hazard assessment applications. As a relatively new technology, there might be some potential for unexpected behavior when using SeismoGen, as its full capabilities and limitations are yet to be fully explored.

MetroGAN. Zhang et al. introduced Metropolitan GAN (MetroGAN) as a geographically informed generative deep learning model for urban morphology simulation [84]. MetroGAN incorporates a progressive growing structure to learn urban features at various scales and leverages physical geography constraints through geographical loss to ensure that urban areas are not generated on water bodies. The generation of cities with MetroGAN involves a global city dataset comprising three layers: terrain (digital elevation model), water, and nighttime lights, effectively capturing the physical geography characteristics and socioeconomic development of cities. The model detects and represents over 10,000 cities worldwide as 100km $\times$ 100km images. The mathematical formulation of the MetroGAN framework is a modified version of the LSGAN model [121], which can be expressed as follows:

$$\begin{aligned} L^* = \arg\min_G \max_D \frac{1}{2}\mathbb{E}_{x,y}\left[(D(x,y)-1)^2\right] \\ + \frac{1}{2}\mathbb{E}_{x,z}\left[(D(x,G(x,z)))^2\right] + \lambda_{L1}L_{L1}(G) \\ - \lambda_{\text{Geo}}\mathbb{E}_{x,z}\left[x_{\text{water}} \odot G(x,z)\right], \end{aligned}$$

where images x with corresponding labels y and a random vector z in the latent space are fed into G to produce simulated images $G(x,z)$. Both real input pairs (x,y) and simulated pairs $(x,G(x,z))$ are then presented to D to distinguish real images from fake ones and also to assess if the input pairs match. The objective loss function comprises different terms, including least square adversarial loss (from the first two expectation terms), $L1$ loss denoted as L_{L1}, and a geographical loss with hyperparameters λ_{L1} and λ_{Geo}, respectively. The geographical loss (last term) utilizes Hadamard product $\odot$ to filter out pixels that generate urban areas on water area x_{water}. MetroGAN, a robust urban morphology simulation model, has several notable advantages and limitations. On the positive side, it incorporates geographical knowledge, resulting in enhanced performance. Its progressive growing structure allows for stable learning at different scales, while multi-layer input ensures precise city layout generation. The model's evaluation framework covers various aspects, ensuring the quality of its output. Furthermore, MetroGAN finds wide applications in urban science and data augmentation. However, these strengths come with challenges, including high computational costs due to extensive data requirements and dependence on data quality, which may hinder its performance with noisy or missing data. Additionally, the model lacks interpretability, making it difficult to understand the reasoning behind its predictions, and it may struggle to represent all intricate features of complex urban systems effectively.

M3GAN. Anomaly detection in multi-dimensional time series data has received tremendous attention in the fields of medicine, fault diagnosis, network intrusion, and climate change. In this work, the authors have proposed the M2GAN (a GAN framework based on a masking strategy for multi-dimensional anomaly detection) and M3GAN (M2GAN for mutable filter) for improving the robustness and accuracy of GAN-based anomaly detection methods. M2GAN generates fake samples by directly reconstructing real samples, which are sufficiently realistic [102]. This is done by extracting various information from the original data by the mask method which improves the robustness of the model. M3GAN fuses the fast Fourier transform (FFT) [180] and wavelet decomposition [181] to obtain a mutable filter to process the raw data so that the model can learn various types of anomalies. The architecture of the M2GAN framework utilizes the AAE [117] in place of the generator of the conventional GAN model for generating realistic fake data. A masking strategy of the AAE enhances the variability within the original time series and overcomes the mode collapse problem. For the discriminator network, this framework employs an AnoGAN [182] architecture that distinguishes between normal data and anomalous data using DCGAN [23]. The M3GAN model combines a dynamic switch-based adaptive filter selection mechanism with the multidimensional anomaly detection capabilities of the M2GAN model. This approach allows one to select the most suitable filter for the given data that better exploits the complex characteristics of the series, leading to improved accuracy in anomaly detection. Both M2GAN and M3GAN architectures excel in spotting anomalies in multi-dimensional time series data, offering adaptability for dynamic settings. Its capacity to generate synthetic data aids tasks like diverse model training. However, their high computational complexity leads to extended processing times. Moreover, their limited interpretability also poses a significant challenge in understanding the marked anomalies. Further research is needed in this domain to address these issues and provide support for adaptive filter parameters in M3GAN.

CNTS. Cooperative Network for Time Series (CNTS), introduced by Yang et al. in 2023, is a reconstruction-based unsupervised anomaly detection technique for time series data [103]. This model aims to overcome the limitations of the previous generative methods that were sensitive to outliers and showed sub-optimal anomaly detection performance due to their emphasis on time series reconstruction. The CNTS framework consists of two FEDformer [183] networks, namely a reconstructor (R) and a detector (D). The reconstructor aims to regenerate the series that closely matches the known data distribution (without anomalies) i.e., data reconstruction. On the other hand, the detector focuses on identifying the

values that deviate from the fitted data distribution, effectively detecting anomalies. Despite having different purposes, these two networks are trained using a cooperative mode, enabling them to leverage mutual information. During the training phase, the reconstruction error of R serves as a labeling mechanism for D, while D provides crucial information to R regarding the presence of anomalies, enhancing the robustness to outliers. Thus the multi-objective function of the CNTS model can be expressed as:

$$\begin{bmatrix} \min\limits_{\theta_D \theta_R} \sum_{i=1}^{n} L_D(D(x_i, \theta_D), L_R(x_i, R(x_i, \theta_R))) \\ \min\limits_{\theta_D \theta_R} \sum_{i=1}^{n} (1 - \hat{y}_i(x_i, \theta_D)) L_R(x_i, R(x_i, \theta_R))) \end{bmatrix},$$

where x_i is the value for the $i^{th}, i = 1, 2, \ldots, n$ time stamp of the input series, θ_D and θ_R denotes the parameters of D and R, while L_D and L_R represent their corresponding loss functions, respectively. The categorical label $\hat{y}_i$ indicates the presence of anomalies as identified by D and helps to remove data with high anomaly scores, thereby reducing their impact on the training of R. The cooperative training approach employed by CNTS allows it to model complex temporal patterns present in real-world time series data, thus significantly enhancing its performance in various anomaly detection tasks. The flexibility and adaptability of the CNTS model make it robust to the presence of outliers in the series. However, the presence of the dual-network architecture of the CNTS model increases its computational complexity, hindering its real-time applicability. Moreover, the lack of interpretability of the model poses a significant challenge to its potential use cases. Furthermore, the success of the CNTS model is contingent on the availability of representative and diverse time series datasets and the choice of sub-networks. Further research in this domain is required to comment on the performance of the model for diverse datasets and appropriate sub-network choices.

RidgeGAN. RidgeGAN, introduced by Thottolil et al. in 2023, is a hybridization of the nonlinear kernel ridge regression (KRR) [184], [185] and the generative CityGAN model [10]. This framework aims to predict the transportation network of the future small and medium-sized cities of India by analyzing the spatial indicators of human settlement patterns. This prediction is crucial for facilitating sustainable urban planning and traffic management systems. The RidgeGAN framework operates in three steps. Firstly, it generates an urban universe for India based on spatial patterns by learning urban morphology using the CityGAN model [82]. Secondly, it utilizes KRR to study the relationship between the human settlement indices (HSI) and the transportation indices (TI) of 503 real small and medium-sized cities in India. Finally, the KKR model's regression framework is applied to the synthetic hyper-realistic samples of future cities and their TI is predicted. RidgeGAN framework has its applications in diverse areas, such as analyzing urban land patterns, forecasting essential urban infrastructure, and assisting policymakers in achieving a more inclusive and effective planning process. Moreover, this model is especially valuable when designing the transportation network of developing nations with limited or partial real data, as the model can produce data that closely resembles actual urban morphology and helps in data augmentation. However, the framework fails to showcase its performance for the generated human settlements which is crucial in the urban planning procedure. Further studies in this domain are indeed required to understand the suitability of the framework for large cities as well.

VI. Recent Theoretical Advancements of GAN

Empirical studies have shown great success of GAN and their variants in producing state-of-the-art results in diverse domains ranging from image, video, and text generation to automatic vehicles, time series, and drug discovery, among many others. The mathematical reasoning of GANs is to approximate the unknown distribution of a given data by optimizing an objective function through an adversarial game between a family of generators and a family of discriminators. Biau et al. [192] analyzed the mathematical and statistical properties of GANs by establishing connections between adversarial principles and Jensen-Shannon (JS) divergence. Their work provides the large sample properties for the parameters of the estimated distribution and a result towards the central limit theorem. Another cousin approach of GAN called WGAN has more stable training dynamics than typical GANs. Biau et al. [193] studied the convergence of empirical WGANs when sample size approaches infinity. More recently, the rate of convergence for density estimation with GANs has been studied in [194]. In particular, they studied the non-asymptotic properties of the vanilla GAN and derived a theoretical guarantee of the density estimation with GANs under a proper choice of deep neural network classes representing generators and discriminators. It suggests that the resulting estimates converge to the true density (p^*) in terms of the JS divergence at the rate of $(\log n/n)^{2\beta/(2\beta+d)}$, where n is the sample size, β determines the smoothness of p^*, and d is the data dimension. In Theorem 2 of [194] if the choice of G and D to be classes of neural networks with rectified quadratic unit (ReQU) activation functions, the rates of convergence for the estimate $p_{\hat{g}}$ to the true density p^* in terms of JS divergence holds the following inequality with probability at least $1 - \delta$;

$$\mathrm{JS}\left(p_{\hat{g}}, p^*\right) \lesssim \left(\frac{\log n}{n}\right)^{\frac{2\beta}{2\beta+d}} + \frac{\log\left(1/\delta\right)}{n}.$$

The above mathematical result suggests that the convergence rate of vanilla GAN's density estimate in the JS divergence is faster than $n^{-1/2}$ when $\beta > \frac{d}{2}$; therefore, the obtained rate is minimax optimal for the considered class of densities. Meitz et al. [195] studied statistical inference for GAN by addressing two critical issues for the generator and discriminator's parameters, namely consistent estimation and confidence sets. Mbacke et al. [196] studied PAC-Bayesian generalization bound for WGANs based on Wasserstein distance and Total variational distance. The generalization properties of GANs try to answer the following question: How to certify that the learned distribution $p_{\hat{g}}$ is "close" to the true one p^*? This question is pivotal since the true distribution p^* is unknown in real problems and generative models can only access its

TABLE II
SOFTWARE LINKS FOR THE GANS

Index	Software name	Language	Backend	Link	Ref.
1	CGAN	Python	PyTorch	https://github.com/Lornatang/CGAN-PyTorch	[2]
2	DCGAN	Python	PyTorch	https://github.com/Natsu6767/DCGAN-PyTorch	[1], [23], [47]
3	AAEs	Python	TensorFlow	https://github.com/conan7882/adversarial-autoencoders	[117]
4	InfoGAN	Python	TensorFlow	https://github.com/openai/InfoGAN	[14]
5	SAD-GAN	–	–	–	[120]
6	LSGAN	Python	PyTorch	https://github.com/xudonmao/LSGAN	[121]
7	SRGAN	Python	TensorFlow	https://github.com/tensorlayer/SRGAN	[124], [186], [187]
8	WGAN	Python	PyTorch	https://github.com/Zeleni9/pytorch-wgan	[109], [122]
9	CycleGAN	Python	TensorFlow	https://github.com/junyanz/CycleGAN	[3], [188]
10	ProGAN	Python	PyTorch	https://github.com/tkarras/progressive_growing_of_gans	[5]
11	MidiNet	Python	TensorFlow	https://github.com/RichardYang40148/MidiNet	[8]
12	SN-GAN	Python	PyTorch	https://github.com/hanyoseob/pytorch-SNGAN	[133]
13	RGAN	Python	TensorFlow	https://github.com/ratschlab/RGAN	[134], [189]
14	StarGAN	Python	PyTorch	https://github.com/yunjey/stargan	[138]
15	BigGAN	Python	PyTorch	https://github.com/ajbrock/BigGAN-PyTorch	[110]
16	MI-GAN	Python	TensorFlow	https://github.com/hazratali/MI-GAN	[146]
17	AttGAN	Python	TensorFlow	https://github.com/LynnHo/AttGAN-Tensorflow	[148], [190]
18	PATE-GAN	Python	TensorFlow	https://github.com/vanderschaarlab/mlforhealthlabpub/tree/main/alg/pategan	[161]
19	DM-GAN	Python	PyTorch	https://github.com/MinfengZhu/DM-GAN	[152]
20	SinGAN	Python	PyTorch	https://github.com/tamarott/SinGAN	[158]
21	POLY-GAN	Python	PyTorch	https://github.com/nile649/POLY-GAN	[95]
22	MIEGAN	–	–	–	[168]
23	VQGAN	Python	PyTorch	https://github.com/dome272/VQGAN-pytorch	[169], [191]
24	DALL-E	Python	PyTorch	https://github.com/lucidrains/DALLE-pytorch	[172], [173]
25	CEGAN	–	–	–	[99]
26	Seismogen	Python	PyTorch	https://github.com/Miffka/seismogen	[87]
27	MetroGAN	Python	PyTorch	https://github.com/zwy-Giser/MetroGAN	[84]
28	M3GAN	Python	PyTorch	https://github.com/SLZWVICTOR/M3GAN	[102]
29	CNTS	Python	PyTorch	https://github.com/BomBooooo/CNTS/tree/main	[103]
30	RidgeGAN	Python	PyTorch	https://github.com/rahisha-thottolil/ridgegan	[10]

empirical counterpart. Liu et al. [197] studied how well GAN can approximate the target distribution under various notions of distributional convergence. Lin et al. [198] showed that under certain conditions GAN-generated samples inherently satisfy some (weak) privacy guarantees. Another study offers a theoretical perspective on why GANs sometimes fail for certain generation tasks, in particular, sequential tasks such as natural language generation [199]. Further research on the comparative theoretical aspects, both pros and cons, of different generative approaches will enhance support for the wide applications of GANs and address their limitations.

VII. EVALUATION MEASURES

In contrast to conventional deep learning architectures that employ convergence-based optimization of the objective function, generative models like GANs utilize a minimax loss function, trained iteratively to establish equilibrium between the generator and discriminator networks [1]. The absence of an objective loss function for GAN training restricts the ability of loss measurements to assess training progress or model performance. To address this challenge, a mix of qualitative and quantitative GAN evaluation approaches has been developed [200]. These evaluation measures particularly vary based on the quality and diversity of the generated synthetic data, as well as the potential applications of the generated data [201].

Owing to the lack of consensus amongst the researchers on the use of a universal metric to gauge the performance of the deep generative models, different metrics have been developed in the last decade with their unique strengths and particular applicability [47]. In this section, we will briefly overview the popular evaluation measures used in different applications.

A. Inception Score

The Inception Score (IS) is a widely used metric to assess the quality and diversity of GAN-generated samples [202]. It leverages a pre-trained neural network classifier called Inception v3 [203], which was initially trained on the Imagenet [204] dataset containing a diverse range of real-world images categorized into 1,000 classes. The IS measures the quality of generated samples based on their classification probabilities predicted by Inception v3. Essentially, higher-quality samples are expected to be strongly classified into specific classes, implying low entropy. In general, the IS value ranges between 1 and the number of classes in the classifier, reflecting the diversity of the generated samples, with higher scores indicating better performance. Nevertheless, the Inception Score does come with a number of limitations. It encounters challenges when dealing with instances of mode collapse, wherein the generated samples by GANs are extremely similar, causing artificially inflated IS values that don't accurately represent diversity. Additionally, it relies on the performance of the Inception v3 model, which might not always align with human perception of image quality. To mitigate these drawbacks of IS, several modified versions have been proposed in the literature. For example, the modified Inception Score (m-IS) attempts to address the mode collapse problem in GAN by evaluating the diversity of images with the same category [205]. Other modification of IS includes the Mode Score (MS) which

evaluates the quality and diversity of the generated data by considering the prior data distribution of the labels [206].

B. Fréchet Inception Distance

The Fréchet Inception Distance (FID) is a widely used evaluation metric that measures the quality and diversity of GAN-generated images [49]. It calculates the similarities and differences between the distributions of real and generated images using the Fréchet distance, which is a form of the Wasserstein-2 distance. The FID metric calculates the mean and covariance of both the real and generated images and then computes the distance between their distributions. Mathematically the FID is expressed as:

$$\mathrm{FID} = |\mu - \mu_w|^2 + \mathrm{tr}\left(\Sigma + \Sigma_w - 2\left(\Sigma\Sigma_w\right)^{1/2}\right),$$

where (μ, Σ) and (μ_w, Σ_w) represent the mean and covariance pair for the real images and the generated images respectively.

The strength of FID lies in its ability to account for various forms of contamination, such as Gaussian noise, Gaussian blur, black rectangles, and swirls, among others. FID's incorporation of these factors contributes to a more robust evaluation of GAN-generated images. As a widely accepted and utilized metric, FID offers a common ground for comparing results across different GAN architectures, promoting a standardized approach for assessing image quality [5], [6], [207].

C. Multi-Scale Structural Similarity

The Multi-Scale Structural Similarity metric (MS-SSIM), an extension of the traditional Structural Similarity Index (SSIM), serves as an effective measure for evaluating the quality of GAN-generated images [208]. MS-SSIM focuses on comparing image structures, including luminance and contrast, across different scales. This metric provides a comprehensive evaluation of the similarity between the real and synthesized datasets, considering their structural and geometric aspects. Moreover, the ability of MS-SSIM to account for strong dependencies between closely correlated pixels enhances its sensitivity to perceptual quality.

D. Classifier Two-Sample Test

Classifier Two-Sample Test (C2ST) is a classification-based approach that evaluates the generalization capabilities of GAN for any synthetic data generation task [209]. This metric utilizes a classifier (for example, 1-Nearest Neighbour [210]) to distinguish between the real and generated samples. The performance of this classifier is then used as a metric to determine the quality of the generated samples. The C2ST metric provides an essential tool for measuring the performance of GAN-based architectures for any applied domains, since the classifier is not restricted to a specific data type. Moreover, it focuses on the discriminative aspect of the generated data quality and complements other evaluation metrics that focus on the distributional and perceptual aspects of the generated data.

E. Music Evaluation Metric

Evaluating the quality of music generated by GANs presents unique challenges due to the subjective nature of musical perception. Traditional quantitative metrics like those used for image evaluation may not fully capture the richness and complexity of musical content. However, several methods have been developed to assess the quality and coherence of GAN-generated music. Certain objective evaluation metrics encompass factors such as musical characteristics, structure, style, uniqueness, and tonality, drawing from statistical representations [35]. Amid these, subjective listening is the most reliable metric for evaluating GAN-generated music. This approach encompasses dimensions like melody, harmony, rhythm, and emotional resonance, thereby furnishing insightful glimpses into the musical caliber.

F. Maximum Mean Discrepancy

Maximum Mean Discrepancy (MMD) is a statistical measure that quantifies the dissimilarity between two probability distributions. In the context of GAN evaluation, MMD is employed to assess the quality of generated samples by comparing them with real data distributions based on their mean values in a high-dimensional space [211]. A lower MMD score indicates that the difference between the two data distributions is relatively smaller, hence the synthetic data is similar to the original data.

G. Time Series Evaluation Metric

Assessing time series GAN models presents a notable challenge due to the temporal dependencies inherent in the data. Traditional evaluation metrics tailored to static image datasets struggle to capture the intricate patterns found in sequential data. As a result, a combined approach of qualitative and quantitative measures is employed for evaluation purposes [37]. Qualitative assessment relies primarily on human visual judgment when examining the generated samples. However, these methods lack objectivity. To address this limitation, a range of quantitative evaluation techniques is employed within GAN-based time series evaluation. These encompass metrics such as root mean square error, Wasserstein-1 distance, dynamic time warping, and Pearson correlation coefficient, among others.

H. Uncertainty Quantification in GANs

Uncertainty Quantification (UQ) plays a vital role in characterizing and estimating the uncertainties in both computation and real-world applications. Due to the fact that the analysis of physical processes based on computer models is riddled with uncertainty, therefore, it has to be addressed to perform 'trustworthy' model-based inference [212]. Oberdiek et al. presented a method to quantify uncertainties of deep neural networks in image classification based on GANs. By employing GANs to generate out-of-distribution (OoD) samples, their methodology enables the classifier to effectively gauge uncertainties for both OoD examples and minor positives [213]. He et al. presented a survey on UQ models for deep

neural networks based on two types of uncertainty sources, namely data uncertainty and model uncertainty [214]. They highlighted that GAN-based models can capture the structure of data uncertainty, however, they are hard to train. Another survey [215] highlighted various measures to quantify uncertainties in deep neural networks. However, it still remains difficult to validate existing methods due to the lack of uncertain ground truths.

VIII. Limitations and scope for improvement

Although GANs have brought a transformative shift in generative modeling, it's crucial to address the substantial challenges embedded within their training process that demand careful consideration [202]. Various architectural modifications of GAN (as discussed in Section V) aim to address specific GAN-related issues and optimize their overall performance. In this section, we summarize the different obstacles in GAN and discuss their potential remedies.

A. Mode Collapse

The foremost challenge during GANs training is mode collapse (MC), a phenomenon where the generator's output becomes constrained, yielding repetitive samples that lack the comprehensive range of the target data distribution [173]. MC arises when the generator doesn't explore the full spectrum of potential outputs and instead generates identical outputs for distinct inputs from the latent space. This issue can manifest due to an overpowering discriminator or insufficient feedback for the generator to diversify its outputs [216]. Partial and complete mode collapse are its two variants, with the former leading to a limited diversity in generated data and the latter resulting in entirely uniform patterns across generated samples. While partial mode collapse is common, complete mode collapse is relatively rare [47].

Many efforts have been made to tackle the mode collapse problem [217], [218]. Some of these approaches include the application of Unrolled GAN [219] where the generator network is updated by unrolling the discriminator's update steps, unlike the conventional GAN, where D is first updated while G is kept fixed and G is updated based on the updated D. Moreover, mini-batch discrimination is often used to mitigate the MC problem [202]. In this approach, instead of modeling each data example independently, D processes multiple data examples in mini-batches. The use of modified loss functions, for example, Least-Square GAN [121], Wasserstein GAN [109], Cycle consistency GAN [3] also reduces the mode collapse problem.

B. Vanishing Gradients

The vanishing gradients problem is another significant challenge encountered during the training phase of GANs. This issue emerges due to the complex architecture of GANs, where both G and D need to maintain a balance and learn collaboratively [220]. During the training process, as gradients are backpropagated through the layers of the network, they can diminish drastically, leading to stagnancy in learning. This circumstance can occur when the discriminator becomes very accurate, such as when $D(G(z) = 0$ and $D(x) = 1$ or when D is inadequately trained and fails to differentiate between real and generated data. Consequently, the loss function might approach zero, hindering constructive feedback to the generator and restricting the generation of high-quality data. Several strategies have been proposed to address vanishing gradients in GANs. One approach is to use a modified loss function, such as the Least-Square GAN [121] that mitigates the vanishing gradient problem to a considerable extent. Furthermore, advanced optimization algorithms, alternative activation functions, and batch normalization strategies are often adopted to reduce the effect of vanishing gradients during GANs training.

C. Learning Instability and Nash Equilibrium

The architectural characteristics of GAN involve a complex interplay between the two deep neural networks in an adversarial manner. Their training happens in a cooperative yet competitive way using a zero-sum game strategy where both G and D aim to optimize their respective objective functions to achieve the Nash equilibrium i.e., a state beyond which they can not improve their performance unilaterally [48]. While this cooperative architecture aims to optimize a global loss function, the optimization problems faced by the individual networks are fundamentally opposing. Due to this complexity in the loss function, there can be situations where some minor adjustments in one network can trigger substantial modifications in the other. Moreover, when both the networks aim to independently optimize their loss functions without coordination, attaining the Nash equilibrium can be hard. Such instances of desynchronization between the networks can lead to instability in the overall learning process and substantially increase the computation time [221]. To counter this challenge, recent advancements in GAN architectures have been focusing on enhancing training stability. The feature matching technique improves the stability of the GAN framework by introducing an alternative cost function for G combining the output of the discriminator [202]. Additionally, historical averaging of the parameters [202], unrolled GAN [219], and gradient penalty [122] strategies mitigate learning instability and promote convergence of the model.

D. Stopping Problem

During GANs training, determining the appropriate time at which the networks are fully optimized is crucial for addressing the problems related to overfitting and underfitting. However, in GANs due to the minimax objective function determining the state of the networks based on their respective loss functions is impossible. To address this issue related to the GANs stopping criterion, researchers often employ an early stopping approach where the training halts based on a predefined threshold or the lack of improvement in evaluation metrics.

E. Internal Distributional Shift

The internal distributional shift often called internal covariate shift refers to the changing distribution in the network

activations of the current layer w.r.t the previous layer. In the context of GAN, when the generator's parameters are updated, the distribution of its output may change, leading to internal distributional shifts in subsequent layers and causing the discriminator's learning to lag behind. This phenomenon affects the convergence of the GAN training process and the computational complexity of the network significantly increases to counter the shifts. To address this issue batch normalization technique is widely adopted in various applications of GAN [222].

IX. DISCUSSION

Over the past decade, GANs have emerged as the foremost and pivotal generative architecture within the areas of computer vision, natural language processing, and related fields. To enhance the performance of GAN architecture, numerous studies have focused on the following: (i) the generation of high-quality samples, (ii) diversity in the simulated samples, and (iii) stabilizing the training algorithm. Constant efforts and improvements of the GAN model have resulted in plausible sample generation, text/image-to-image translations, data augmentation, style transfer, anomaly detection, and other applied domains.

Recent advancements in machine learning with the help of Diffusion models [22], [223], [224] also known as score-based generative models have made a strong impression on a variety of tasks including image denoising, image inpainting, image super-resolution, and image generation. The primary goal of Diffusion models is to learn the latent structure of the dataset by modeling the way in which data points diffuse through the latent space. [225] has shown that Diffusion models outperform GANs on image synthesis due to their better stability and non-existence of mode collapse. However, the cost of synthesizing new samples and computational time for making realistic images lead to its shortcomings when applied to real-time application [226], [227]. Due to the fact that GANs need fine-tuning in their hyperparameters, Transformers [19] have been used to enhance the results of GANs that can adopt self-attention layers. This helps in designing larger models and replacing the neural network models of G and D within the GAN structure. TransGAN [228] introduces a GAN architecture without convolutions by using Transformers in both G and D of the GAN resulting in improved high-resolution image generation. [229] presented an intersection of GANs and Transformers to predict pedestrian paths. Although Transformers and their variants have several advantages, they suffer from high computational (time and resource) complexity [230]. More recently, physics-informed neural networks (PINN) [20] was introduced as a universal function approximator that can incorporate knowledge of physical laws to govern the data in the learning process. PINNs overcome the low data availability issue [231] in which GANs and Transformers lack robustness, rendering them ineffective scenarios. A GAN framework based on a physics-informed (PI) discriminator for uncertainty quantification is used to inform the knowledge of physics during the learning of both G and D models. Physics-informed Discriminator GAN (PID-GAN) [232] doesn't suffer from an imbalance of generator gradient from multiple losses. Another architecture namely Physics-informed GAN (PI-GAN) [233] tackles the problem of sequence generation with limited data. It integrates a transition module in the generator part that can iteratively construct the sequence with only one initial point as input. Solving differential equations using GANs to learn the loss function was presented in the Differential Equation GAN (DEQ-GAN) model [234]. Combining GANs with PINNs achieved solution accuracies that are competitive with popularly used numerical methods.

Large language models (LLMs) [21] became a very popular choice for their ability to understand and generate human language. LLMs are neural networks that are trained on massive text datasets to understand the relationship between words and phrases. This enables LLMs to generate text that is both coherent and grammatically correct. Recently, LLMs and their cousin ChatGPT revolutionized the field of natural language processing, question-answering, and creative writing. Additionally, LLMs and their variants are used to create creative content such as poems, scripts, and codes. GANs and LLMs are two powerful co-existing models where the former is used to generate realistic images. Mega-TTS [235] adopt a VQGAN [169] based acoustic model and a latent-code language model called Prosody-LLM (P-LLM) [236] to solve zero-shot text-to-speech at scale with intrinsic inductive bias. Future works in the hybridization of GANs with several other architectures will be a promising field of future research.

X. FUTURE RESEARCH DIRECTION

Despite the substantial advancements achieved by GAN-based frameworks over the past decade, there remain a number of challenges spanning both theoretical and practical aspects that require further exploration in future research. In this section, we identify these gaps that necessitate deeper investigation to enhance our comprehension of GANs. The summary is presented below:

a) Fundamental questions on the theory of GANs: Recent advancements in the theory of GAN by [192], [193], [197] explored the role of the discriminator family in terms of JS divergence and some large sample properties (convergence and asymptotic normality) of the parameter describing the empirically selected generator. However, a fundamental question of how well GANs can approximate the target distribution p^* remained largely unanswered. From the theoretical perspective, there is still a mystery about the role and impact of the discriminator on the quality of the approximation. The universal consistency and the rate of convergence of GANs and their variants still remain an open problem.

b) Improvement of training stability and diversity: Achieving the Nash equilibrium in GAN frameworks, which is essential for the generator to learn the actual sample distribution, requires stable training mechanisms [237], [238]. However, attaining this optimal balance between the generator and discriminator remains challenging. Various approaches have been explored, such as WGAN [109], SN-GAN [133], One-sided Label Smoothing [203], and WGAN with gradient penalty (WGAN-GP) [122], to enhance training stability. Additionally, addressing mode collapse, a common GAN issue

that leads to limited sample diversity, has prompted strategies like WGAN [109], U-GAN [219], generator regulating GAN (GRGAN) [239], and Adaptive GAN [240]. Future research could focus on devising techniques to stabilize GAN training and alleviate problems like mode collapse through regularization methods, alternative loss functions, and optimized hyperparameters. Incorporating methods like multi-modal GANs, designed to generate diverse outputs from a single input, might contribute to enhancing sample diversity [239].

c) Data scarcity in GAN: Addressing the issue of data scarcity in GANs stands as a crucial research trajectory. To expand GAN applications, forthcoming investigations could focus on devising training strategies for scenarios with limited data. Approaches such as few-shot GANs, transfer learning, and domain adaptation offer the potential to enhance GAN performance when data is scarce [241], [242]. This challenge becomes especially pertinent when acquiring substantial datasets poses difficulties. Additionally, refining training algorithms for maximal data utility could be pursued. Bolstering GAN effectiveness in low-data situations holds pivotal significance for broader adoption across various industries and domains.

d) Ethics and privacy: Since its inception in 2014, GAN development has yielded substantial benefits in research and real-world applications. However, the inappropriate utilization of GANs can give rise to latent societal issues such as producing deceptive content, malicious images, fabricated news, deepfakes, prejudiced portrayals, and compromising individual safety [243]. To tackle these issues, the establishment of ethical guidelines and regulations is imperative [244]. Future research avenues might center on developing robust techniques to detect and alleviate ethical concerns associated with GANs, while also advocating their ethical and responsible deployment in diverse fields. Essential to this effort is the creation of forgery detection methods capable of effectively identifying AI-generated content, including images produced through GANs. Furthermore, GANs can be susceptible to adversarial attacks, wherein minor modifications to input data result in visually convincing yet incorrect outputs [116], [245]. Future investigations could prioritize the development of robust GANs that can withstand such attacks, alongside methods for identifying and countering them. Ensuring the integrity and reliability of GANs is of utmost importance, particularly in contexts like authentication, content verification, and cybersecurity [216], [246].

e) Real-time implementation and scalability: While GANs have shown immense potential, their resource-intensive nature hinders real-time usage and scalability. Recent GAN variants like ProGAN [5] and Att-GAN [148] aim to address this complexity. Future efforts might focus on crafting efficient GAN architectures capable of generating high-quality samples in real-time, vital for constrained platforms like mobile devices and edge computing. Integrating GANs with reinforcement learning, transfer learning, and supervised learning, as seen in RidgeGAN [10], opens opportunities for hybrid models with expanded capabilities. Research should delve into hybrid approaches, leveraging GANs alongside other techniques for enhanced generative potential. Additionally, exploring multimodal GANs that produce diverse outputs from multiple modalities can unlock novel avenues for creating complex data [247].

f) Human-centric GANs: GANs have the potential to enable human-machine creative cooperation [248]. Future research could emphasize human-centric GANs, integrating human feedback, preferences, and creativity into the generative process. This direction might pave the way for interactive and co-creative GANs, enabling the production of outputs aligned with human preferences and needs, while also involving users in active participation during the generation process.

g) Other innovative applications and industry usage: Initially designed for generating realistic images, GANs have exhibited impressive performance in computer vision. While their application has extended to domains like time series generation [102], [103], audio synthesis [8], and autonomous vehicles [120], their use outside computer vision remains somewhat constrained. The divergent nature of image and non-image data introduces challenges, particularly in non-image contexts like NLP, where discrete values such as words and characters predominate [199]. Future research can aim to overcome these challenges and enhance GANs' capabilities in discrete data scenarios. Furthermore, exploring unique applications of GANs in fields like finance, education, and entertainment offers the potential to introduce new possibilities and positively impact various industries [249]. Collaborative efforts across disciplines could also harness diverse expertise, fostering synergies to enhance GANs' adaptability across a broad spectrum of applications [250].

XI. CONCLUSION

In this article, we presented a GAN survey, GAN variants, and a detailed analysis of the wide range of GAN applications in several applied domains. In addition, we reviewed the recent theoretical developments in the GAN literature and the most common evaluation metrics. Despite all these one of the core contributions of this survey is to discuss several obstacles of various GAN architectures and their potential solutions for future research. Overall, we discuss GANs' potential to facilitate practical applications not only in image, audio, and text but also in relatively uncommon areas such as time series analysis, geospatial data analysis, and imbalanced learning. In the discussion section, apart from GANs' significant success, we detail the failures of GANs due to their time complexity and unstable training. Although GANs have been phenomenal for the generation of hyper-realistic data, current progress in deep learning depicts an alternative narrative. Recently developed architectures such as Diffusion models have demonstrated significant success and outperformed GANs on image synthesis. On the other hand, Transformers, a deep learning architecture based on a multi-head attention mechanism, has been used within GAN architecture to enhance its performance. Furthermore, Large Language Models, a widely utilized deep learning structure designed for comprehending and producing natural language, have been incorporated into GAN architecture to bolster its effectiveness. The hybridization of PINN and GAN namely, PI-GAN can solve inverse and mixed stochastic problems

based on a limited number of scattered measurements. On the contrary, GANs' ability which relies on large data for training, using physical laws inside GANs in the form of stochastic differential equations can mitigate the limited data problem. Several hybrid approaches combining GAN with other powerful deep learners are showing great merit and success as discussed in the discussion section. Finally, several applications of GANs over the last decade are summarized and criticized throughout the article.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets (advances in neural information processing systems)(pp. 2672–2680)," *Red Hook, NY Curran*, 2014.

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

[3] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[4] 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.

[5] 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.

[6] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[7] X. Liu, M. Cheng, H. Zhang, and C.-J. Hsieh, "Towards robust neural networks via random self-ensemble," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 369–385.

[8] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "Midinet: A convolutional generative adversarial network for symbolic-domain music generation," *arXiv preprint arXiv:1703.10847*, 2017.

[9] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.

[10] R. Thottolil, U. Kumar, and T. Chakraborty, "Prediction of transportation index for urban patterns in small and medium-sized indian cities using hybrid ridgegan model," *arXiv preprint arXiv:2306.05951*, 2023.

[11] K. E. Smith and A. O. Smith, "Conditional gan for timeseries generation," *arXiv preprint arXiv:2006.16477*, 2020.

[12] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[13] J. Togelius, N. Shaker, and M. J. Nelson, "Procedural content generation in games: A textbook and an overview of current research," *Togelius N. Shaker M. Nelson Berlin: Springer*, 2014.

[14] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.

[15] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

[16] D. Wilby, T. Aarts, P. Tichit, A. Bodey, C. Rau, G. Taylor, and E. Baird, "Using micro-ct techniques to explore the role of sex and hair in the functional morphology of bumblebee (bombus terrestris) ocelli," *Vision Research*, vol. 158, pp. 100–108, 2019.

[17] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*. PMLR, 2018, pp. 77–91.

[18] J. Zhao, T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang, "Gender bias in coreference resolution: Evaluation and debiasing methods," *arXiv preprint arXiv:1804.06876*, 2018.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[20] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[21] A. Radford, J. Wu, D. Amodei, D. Amodei, J. Clark, M. Brundage, and I. Sutskever, "Better language models and their implications," *OpenAI blog*, vol. 1, no. 2, 2019.

[22] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *International conference on machine learning*. PMLR, 2015, pp. 2256–2265.

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

[24] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*. Springer, 2020, pp. 70–85.

[25] R. Kulkarni, R. Gaikwad, R. Sugandhi, P. Kulkarni, and S. Kone, "Survey on deep learning in music using gan," *Int. J. Eng. Res. Technol*, vol. 8, no. 9, pp. 646–648, 2019.

[26] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Computing Surveys (CSUR)*, vol. 54, no. 8, pp. 1–49, 2021.

[27] M. Durgadevi *et al.*, "Generative adversarial network (gan): a general review on different variants of gan and applications," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2021, pp. 1–8.

[28] R. Nandhini Abirami, P. Durai Raj Vincent, K. Srinivasan, U. Tariq, and C.-Y. Chang, "Deep cnn and deep gan in computational visual perception-driven image analysis," *Complexity*, vol. 2021, pp. 1–30, 2021.

[29] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–38, 2021.

[30] V. Sampath, I. Maurtua, J. J. Aguilar Martin, and A. Gutierrez, "A survey on generative adversarial networks for imbalance problems in computer vision tasks," *Journal of big Data*, vol. 8, pp. 1–59, 2021.

[31] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE transactions on knowledge and data engineering*, 2021.

[32] Y. Li, Q. Wang, J. Zhang, L. Hu, and W. Ouyang, "The theoretical research of generative adversarial networks: an overview," *Neurocomputing*, vol. 435, pp. 26–41, 2021.

[33] W. Xia, Y. Zhang, Y. Yang, J.-H. Xue, B. Zhou, and M.-H. Yang, "Gan inversion: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[34] S. Xun, D. Li, H. Zhu, M. Chen, J. Wang, J. Li, M. Chen, B. Wu, H. Zhang, X. Chai, *et al.*, "Generative adversarial networks in medical image segmentation: A review," *Computers in biology and medicine*, vol. 140, p. 105063, 2022.

[35] S. Ji, X. Yang, and J. Luo, "A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges," *ACM Computing Surveys*, 2023.

[36] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on gans for computer vision: Recent research, analysis and taxonomy," *Computer Science Review*, vol. 48, p. 100553, 2023.

[37] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A systematic literature review," *ACM Computing Surveys*, vol. 55, no. 10, pp. 1–31, 2023.

[38] C. Vondrick, A. Shrivastava, A. Fathi, S. Guadarrama, and K. Murphy, "Tracking emerges by colorizing videos," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 391–408.

[39] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[40] J. Tan, L. Jing, Y. Huo, L. Li, O. Akin, and Y. Tian, "Lgan: Lung segmentation in ct scans using generative adversarial network," *Computerized Medical Imaging and Graphics*, vol. 87, p. 101817, 2021.

[41] S. Nema, A. Dudhane, S. Murala, and S. Naidu, "Rescuenet: An unpaired gan for brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 55, p. 101641, 2020.

[42] Y. Abouelnaga, O. S. Ali, H. Rady, and M. Moustafa, "Cifar-10: Knn-based ensemble of classifiers," in *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2016, pp. 1192–1195.

[43] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, "Do imagenet classifiers generalize to imagenet?" in *International conference on machine learning*. PMLR, 2019, pp. 5389–5400.

[44] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *electronics*, vol. 8, no. 3, p. 292, 2019.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[46] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[47] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.

[48] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.

[49] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, vol. 30, 2017.

[50] F. Farnia and A. Ozdaglar, "Do gans always have nash equilibria?" in *International Conference on Machine Learning*. PMLR, 2020, pp. 3029–3039.

[51] M.-Y. Liu, X. Huang, J. Yu, T.-C. Wang, and A. Mallya, "Generative adversarial networks for image and video synthesis: Algorithms and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 839–862, 2021.

[52] S. W. Kim, Y. Zhou, J. Philion, A. Torralba, and S. Fidler, "Learning to simulate dynamic environments with gamegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1231–1240.

[53] Y.-J. Cao, L.-L. Jia, Y.-X. Chen, N. Lin, C. Yang, B. Zhang, Z. Liu, X.-X. Li, and H.-H. Dai, "Recent advances of generative adversarial networks in computer vision," *IEEE Access*, vol. 7, pp. 14 985–15 006, 2018.

[54] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," *Advances in neural information processing systems*, vol. 30, 2017.

[55] Y. Yu, Z. Gong, P. Zhong, and J. Shan, "Unsupervised representation learning with deep convolutional neural network for remote sensing images," in *Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part II 9*. Springer, 2017, pp. 97–108.

[56] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, "Imaginator: Conditional spatio-temporal gan for video generation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1160–1169.

[57] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1526–1535.

[58] W. Wang, H. Yang, Z. Tuo, H. He, J. Zhu, J. Fu, and J. Liu, "Videofactory: Swap attention in spatiotemporal diffusions for text-to-video generation," *arXiv preprint arXiv:2305.10874*, 2023.

[59] M. Westerlund, "The emergence of deepfake technology: A review," *Technology innovation management review*, vol. 9, no. 11, 2019.

[60] P. Korshunov and S. Marcel, "Vulnerability assessment and detection of deepfake videos," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–6.

[61] P. Yu, Z. Xia, J. Fei, and Y. Lu, "A survey on deepfake video detection," *Iet Biometrics*, vol. 10, no. 6, pp. 607–624, 2021.

[62] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in neural information processing systems*, vol. 33, pp. 6256–6268, 2020.

[63] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.

[64] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 289–293.

[65] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 694–711.

[66] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.

[67] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[68] Y. Zhang, Z. Gan, and L. Carin, "Generating text via adversarial training," in *NIPS workshop on Adversarial Training*, vol. 21. academia. edu, 2016, pp. 21–32.

[69] M. Toshevska and S. Gievska, "A review of text style transfer using deep learning," *IEEE Transactions on Artificial Intelligence*, 2021.

[70] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, and J. Wang, "Long text generation via adversarial training with leaked information," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[71] Z. Mu, X. Yang, and Y. Dong, "Review of end-to-end speech synthesis technology based on deep learning," *arXiv preprint arXiv:2104.09995*, 2021.

[72] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[73] M. Civit, J. Civit-Masot, F. Cuadrado, and M. J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," *Expert Systems with Applications*, p. 118190, 2022.

[74] X. Mao, S. Wang, L. Zheng, and Q. Huang, "Semantic invariant cross-domain image generation with generative adversarial networks," *Neurocomputing*, vol. 293, pp. 55–63, 2018.

[75] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," *arXiv preprint arXiv:1709.01872*, 2017.

[76] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks: A review," *Health informatics: A computational perspective in healthcare*, pp. 77–96, 2021.

[77] C. Wang, G. Yang, G. Papanastasiou, S. A. Tsaftaris, D. E. Newby, C. Gray, G. Macnaught, and T. J. MacGillivray, "Dicyc: Gan-based deformation invariant cross-domain information fusion for medical image synthesis," *Information Fusion*, vol. 67, pp. 147–160, 2021.

[78] A. Kadurin, A. Aliper, A. Kazennov, P. Mamoshina, Q. Vanhaelen, K. Khrabrov, and A. Zhavoronkov, "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology," *Oncotarget*, vol. 8, no. 7, p. 10883, 2017.

[79] A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, and A. Zhavoronkov, "drugan: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico," *Molecular pharmaceutics*, vol. 14, no. 9, pp. 3098–3104, 2017.

[80] Y. Zhao, Y. Wang, J. Zhang, X. Liu, Y. Li, S. Guo, X. Yang, and S. Hong, "Surgical gan: Towards real-time path planning for passive flexible tools in endovascular surgeries," *Neurocomputing*, vol. 500, pp. 567–580, 2022.

[81] S. Ma, Z. Hu, K. Ye, X. Zhang, Y. Wang, and H. Peng, "Feasibility study of patient-specific dose verification in proton therapy utilizing positron emission tomography (pet) and generative adversarial network (gan)," *Medical Physics*, vol. 47, no. 10, pp. 5194–5208, 2020.

[82] A. Albert, E. Strano, J. Kaur, and M. C. González, "Modeling urbanization patterns with generative adversarial networks," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2095–2098, 2018.

[83] A. Albert, J. Kaur, E. Strano, and M. Gonzalez, "Spatial sensitivity analysis for urban land use prediction with physics-constrained conditional generative adversarial networks," *arXiv preprint arXiv:1907.09543*, 2019.

[84] W. Zhang, Y. Ma, D. Zhu, L. Dong, and Y. Liu, "Metrogan: Simulating urban morphology with generative adversarial network," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 2482–2492.

[85] L. Mosser, O. Dubrule, and M. J. Blunt, "Reconstruction of three-dimensional porous media using generative adversarial neural networks," *Physical Review E*, vol. 96, no. 4, p. 043309, 2017.

[86] T.-F. Zhang, P. Tilke, E. Dupont, L.-C. Zhu, L. Liang, and W. Bailey, "Generating geologically realistic 3d reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks," *Petroleum Science*, vol. 16, pp. 541–549, 2019.

[87] T. Wang, D. Trugman, and Y. Lin, "Seismogen: Seismic waveform synthesis using gan with application to seismic data augmentation," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 4, p. e2020JB020077, 2021.

[88] B. Gecer, B. Bhattarai, J. Kittler, and T.-K. Kim, "Semi-supervised adversarial learning to generate photorealistic face images of new identities from 3d morphable model," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 217–234.

[89] X. Pan, Y. You, Z. Wang, and C. Lu, "Virtual to real reinforcement learning for autonomous driving," *arXiv preprint arXiv:1704.03952*, 2017.

[90] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2107–2116.

[91] M. Zhang, Y. Zhang, L. Zhang, C. Liu, and S. Khurshid, "Deeproad: Gan-based metamorphic testing and input validation framework for autonomous driving systems," in *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering*, 2018, pp. 132–142.

[92] S. Jiang and Y. Fu, "Fashion style generator." in *IJCAI*, 2017, pp. 3721–3727.

[93] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "Viton: An image-based virtual try-on network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7543–7552.

[94] L. Liu, H. Zhang, Y. Ji, and Q. J. Wu, "Toward ai fashion design: An attribute-gan model for clothing match," *Neurocomputing*, vol. 341, pp. 156–167, 2019.

[95] N. Pandey and A. Savakis, "Poly-gan: Multi-conditioned gan for fashion synthesis," *Neurocomputing*, vol. 414, pp. 356–364, 2020.

[96] T. Chakraborty and A. K. Chakraborty, "Hellinger net: A hybrid imbalance learning model to improve software defect prediction," *IEEE Transactions on Reliability*, vol. 70, no. 2, pp. 481–494, 2020.

[97] T. Dam, M. M. Ferdaus, M. Pratama, S. G. Anavatti, S. Jayavelu, and H. Abbass, "Latent preserving generative adversarial network for imbalance classification," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 3712–3716.

[98] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, and C. Malossi, "Bagan: Data augmentation with balancing gan," *arXiv preprint arXiv:1803.09655*, 2018.

[99] S. Suh, H. Lee, P. Lukowicz, and Y. O. Lee, "Cegan: Classification enhancement generative adversarial networks for unraveling data imbalance problems," *Neural Networks*, vol. 133, pp. 69–86, 2021.

[100] M. Panja, T. Chakraborty, U. Kumar, and N. Liu, "Epicasting: An ensemble wavelet neural network for forecasting epidemics," *Neural Networks*, 2023.

[101] Y. Li, X. Peng, J. Zhang, Z. Li, and M. Wen, "Dct-gan: dilated convolutional transformer-based gan for time series anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[102] Y. Li, X. Peng, Z. Wu, F. Yang, X. He, and Z. Li, "M3gan: A masking strategy with a mutable filter for multidimensional anomaly detection," *Knowledge-Based Systems*, vol. 271, p. 110585, 2023.

[103] J. Yang, Y. Shao, and C.-N. Li, "Cnts: Cooperative network for time series," *IEEE Access*, vol. 11, pp. 31 941–31 950, 2023.

[104] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante, and K. Veeramachaneni, "Tadgan: Time series anomaly detection using generative adversarial networks," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 33–43.

[105] Y. Liu, J. Peng, J. James, and Y. Wu, "Ppgan: Privacy-preserving generative adversarial network," in *2019 IEEE 25Th international conference on parallel and distributed systems (ICPADS)*. IEEE, 2019, pp. 985–989.

[106] A. Torfi and E. A. Fox, "Corgan: correlation-capturing convolutional generative adversarial networks for generating synthetic healthcare records," *arXiv preprint arXiv:2001.09346*, 2020.

[107] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE symposium on security and privacy (SP)*. IEEE, 2017, pp. 3–18.

[108] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.

[109] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[110] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[111] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*. PMLR, 2017, pp. 2642–2651.

[112] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

[113] J. Xiao, S. Zhang, Y. Yao, Z. Wang, Y. Zhang, and Y.-F. Wang, "Generative adversarial network with hybrid attention and compromised normalization for multi-scene image conversion," *Neural Computing and Applications*, vol. 34, no. 9, pp. 7209–7225, 2022.

[114] E. L. Denton, S. Chintala, R. Fergus, *et al.*, "Deep generative image models using a laplacian pyramid of adversarial networks," *Advances in neural information processing systems*, vol. 28, 2015.

[115] A. Krizhevsky, G. Hinton, *et al.*, *Learning multiple layers of features from tiny images*. Toronto, ON, Canada, 2009.

[116] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," *Advances in neural information processing systems*, vol. 31, 2018.

[117] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[118] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, "Unsupervised pixel-level domain adaptation with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3722–3731.

[119] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," in *International conference on learning representations*, 2017.

[120] A. Ghosh, B. Bhattacharya, and S. B. R. Chowdhury, "Sad-gan: Synthetic autonomous driving using generative adversarial networks," *arXiv preprint arXiv:1611.08788*, 2016.

[121] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[122] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[123] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612.

[124] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "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.

[125] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[126] L. Mescheder, S. Nowozin, and A. Geiger, "The numerics of gans," *Advances in neural information processing systems*, vol. 30, 2017.

[127] Y. C. M. W. H. Sergio and G. Colmenarejo, "Learning to learn for global optimization of black box functions," *stat*, vol. 1050, p. 18, 2016.

[128] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[129] S. R. Hashemi, S. S. M. Salehi, D. Erdogmus, S. P. Prabhu, S. K. Warfield, and A. Gholipour, "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection," *IEEE Access*, vol. 7, pp. 1721–1735, 2018.

[130] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[131] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[132] H. Chu, R. Urtasun, and S. Fidler, "Song from pi: A musically plausible network for pop music generation," *arXiv preprint arXiv:1611.03477*, 2016.

[133] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *arXiv preprint arXiv:1802.05957*, 2018.

[134] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," *arXiv preprint arXiv:1807.00734*, 2018.

[135] G. Gómez-de Segura and R. García-Mayoral, "Turbulent drag reduction by anisotropic permeable substrates–analysis and direct numerical simulations," *Journal of Fluid Mechanics*, vol. 875, pp. 124–172, 2019.

[136] A. Nguyen, J. Yosinski, and J. Clune, "Multifaceted feature visualization: Uncovering the different types of features learned by each neuron in deep neural networks," *arXiv preprint arXiv:1602.03616*, 2016.

[137] F. Tramèr, A. Kurakin, N. Papernot, I. Goodfellow, D. Boneh, and P. McDaniel, "Ensemble adversarial training: Attacks and defenses," *arXiv preprint arXiv:1705.07204*, 2017.

[138] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8789–8797.

[139] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," *Advances in neural information processing systems*, vol. 30, 2017.

[140] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1501–1510.

[141] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[142] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387–2395.

[143] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in neural information processing systems*, vol. 33, pp. 12 104–12 114, 2020.

[144] G. Franceschelli and M. Musolesi, "Creativity and machine learning: A survey," *arXiv preprint arXiv:2104.02726*, 2021.

[145] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," *arXiv preprint arXiv:1606.00704*, 2016.

[146] T. Iqbal and H. Ali, "Generative adversarial network for medical images (mi-gan)," *Journal of medical systems*, vol. 42, pp. 1–11, 2018.

[147] M. Mahmud, M. S. Kaiser, T. M. McGinnity, and A. Hussain, "Deep learning in mining biological data," *Cognitive computation*, vol. 13, pp. 1–33, 2021.

[148] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Attgan: Facial attribute editing by only changing what you want," *IEEE transactions on image processing*, vol. 28, no. 11, pp. 5464–5478, 2019.

[149] T. Dai, Y. Feng, B. Chen, J. Lu, and S.-T. Xia, "Deep image prior based defense against adversarial examples," *Pattern Recognition*, vol. 122, p. 108249, 2022.

[150] X. Hou, L. Shen, K. Sun, and G. Qiu, "Deep feature consistent variational autoencoder," in *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 2017, pp. 1133–1141.

[151] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International conference on machine learning*. PMLR, 2016, pp. 1060–1069.

[152] M. Zhu, P. Pan, W. Chen, and Y. Yang, "Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5802–5810.

[153] K. Li, T. Zhang, and J. Malik, "Diverse image synthesis from semantic layouts via conditional imle," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4220–4229.

[154] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[155] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[156] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," *arXiv preprint arXiv:1410.5401*, 2014.

[157] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*. Springer, 2014, pp. 818–833.

[158] T. R. Shaham, T. Dekel, and T. Michaeli, "Singan: Learning a generative model from a single natural image," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4570–4580.

[159] D. Berthelot, C. Raffel, A. Roy, and I. Goodfellow, "Understanding and improving interpolation in autoencoders via an adversarial regularizer," *arXiv preprint arXiv:1807.07543*, 2018.

[160] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[161] J. Jordon, J. Yoon, and M. Van Der Schaar, "Pate-gan: Generating synthetic data with differential privacy guarantees," in *International conference on learning representations*, 2018.

[162] G. Rogez, P. Weinzaepfel, and C. Schmid, "Lcr-net++: Multi-person 2d and 3d pose detection in natural images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 5, pp. 1146–1161, 2019.

[163] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[164] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[165] S. Zhu, R. Urtasun, S. Fidler, D. Lin, and C. Change Loy, "Be your own prada: Fashion synthesis with structural coherence," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1680–1688.

[166] M. Mameli, M. Paolanti, R. Pietrini, G. Pazzaglia, E. Frontoni, and P. Zingaretti, "Deep learning approaches for fashion knowledge extraction from social media: a review," *Ieee Access*, vol. 10, pp. 1545–1576, 2021.

[167] Y. Wu, H. Liu, P. Lu, L. Zhang, and F. Yuan, "Design and implementation of virtual fitting system based on gesture recognition and clothing transfer algorithm," *Scientific Reports*, vol. 12, no. 1, p. 18356, 2022.

[168] Z. Pan, F. Yuan, J. Lei, W. Li, N. Ling, and S. Kwong, "Miegan: Mobile image enhancement via a multi-module cascade neural network," *IEEE Transactions on Multimedia*, vol. 24, pp. 519–533, 2021.

[169] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 873–12 883.

[170] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 87, p. 102792, 2023.

[171] N. Kalchbrenner, A. Oord, K. Simonyan, I. Danihelka, O. Vinyals, A. Graves, and K. Kavukcuoglu, "Video pixel networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1771–1779.

[172] A. Radford, J. Wu, R. Child, D. Amodei, and I. Sutskever, "Dall-e: Distributed, automated, and learning to generate adversarial networks," *OpenAI Blog*, 2021. [Online]. Available: https://openai.com/blog/dall-e/

[173] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8821–8831.

[174] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PMLR, 2021, pp. 8748–8763.

[175] G. Singh, F. Deng, and S. Ahn, "Illiterate dall-e learns to compose," *arXiv preprint arXiv:2110.11405*, 2021.

[176] G. Marcus, E. Davis, and S. Aaronson, "A very preliminary analysis of dall-e 2," *arXiv preprint arXiv:2204.13807*, 2022.

[177] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature machine intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[178] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," *arXiv preprint arXiv:2204.06125*, 2022.

[179] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[180] E. O. Brigham, *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[181] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2000, vol. 4.

[182] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.

[183] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International Conference on Machine Learning*. PMLR, 2022, pp. 27 268–27 286.

[184] V. Vovk, "Kernel ridge regression," in *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*. Springer, 2013, pp. 105–116.

[185] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[186] H. Dong, A. Supratak, L. Mai, F. Liu, A. Oehmichen, S. Yu, and Y. Guo, "TensorLayer: A Versatile Library for Efficient Deep Learning Development," *ACM Multimedia*, 2017. [Online]. Available: http://tensorlayer.org

[187] C. Lai, J. Han, and H. Dong, "Tensorlayer 3.0: A deep learning library compatible with multiple backends," in *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2021, pp. 1–3.

[188] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networkss," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[189] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," *arXiv preprint arXiv:1706.02633*, 2017.

[190] G. Zhang, M. Kan, S. Shan, and X. Chen, "Generative adversarial network with spatial attention for face attribute editing," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 417–432.

[191] A. Razavi, A. Van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with vq-vae-2," *Advances in neural information processing systems*, vol. 32, 2019.

[192] G. Biau, B. Cadre, M. Sangnier, and U. Tanielian, "Some theoretical properties of gans," *Ann. Statist.*, vol. 48 (3), pp. 1539 – 1566, 2020.

[193] G. Biau, M. Sangnier, and U. Tanielian, "Some theoretical insights into wasserstein gans," *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 5287–5331, 2021.

[194] D. Belomestny, E. Moulines, A. Naumov, N. Puchkin, and S. Samsonov, "Rates of convergence for density estimation with gans," *arXiv preprint arXiv:2102.00199*, 2021.

[195] M. Meitz, "Statistical inference for generative adversarial networks," *arXiv preprint arXiv:2104.10601*, 2021.

[196] S. D. Mbacke, F. Clerc, and P. Germain, "Pac-bayesian generalization bounds for adversarial generative models," *arXiv preprint arXiv:2302.08942*, 2023.

[197] S. Liu, O. Bousquet, and K. Chaudhuri, "Approximation and convergence properties of generative adversarial learning," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[198] Z. Lin, V. Sekar, and G. Fanti, "On the privacy properties of gan-generated samples," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 1522–1530.

[199] D. Alvarez-Melis, V. Garg, and A. Kalai, "Are gans overkill for nlp?" *Advances in Neural Information Processing Systems*, vol. 35, pp. 9072–9084, 2022.

[200] A. Borji, "Pros and cons of gan evaluation measures," *Computer vision and image understanding*, vol. 179, pp. 41–65, 2019.

[201] J. Xu, X. Ren, J. Lin, and X. Sun, "Diversity-promoting gan: A cross-entropy based generative adversarial network for diversified text generation," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 3940–3949.

[202] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.

[203] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[204] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[205] S. Gurumurthy, R. Kiran Sarvadevabhatla, and R. Venkatesh Babu, "Deligan: Generative adversarial networks for diverse and limited data," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 166–174.

[206] S. Nowozin, B. Cseke, and R. Tomioka, "f-gan: Training generative neural samplers using variational divergence minimization," *Advances in neural information processing systems*, vol. 29, 2016.

[207] G. Daras, A. Odena, H. Zhang, and A. G. Dimakis, "Your local gan: Designing two dimensional local attention mechanisms for generative models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 531–14 539.

[208] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[209] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*. Springer, 1986, vol. 3.

[210] P. Cunningham and S. J. Delany, "k-nearest neighbour classifiers-a tutorial," *ACM computing surveys (CSUR)*, vol. 54, no. 6, pp. 1–25, 2021.

[211] W. Bounliphone, E. Belilovsky, M. B. Blaschko, I. Antonoglou, and A. Gretton, "A test of relative similarity for model selection in generative models," *arXiv preprint arXiv:1511.04581*, 2015.

[212] V. Volodina and P. Challenor, "The importance of uncertainty quantification in model reproducibility," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2197, p. 20200071, 2021.

[213] P. Oberdiek, G. Fink, and M. Rottmann, "Uqgan: A unified model for uncertainty quantification of deep classifiers trained via conditional gans," *Advances in Neural Information Processing Systems*, vol. 35, pp. 21 371–21 385, 2022.

[214] W. He and Z. Jiang, "A survey on uncertainty quantification methods for deep neural networks: An uncertainty source perspective," *arXiv preprint arXiv:2302.13425*, 2023.

[215] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, *et al.*, "A survey of uncertainty in deep neural networks," *Artificial Intelligence Review*, pp. 1–77, 2023.

[216] P. Samangouei, M. Kabkab, and R. Chellappa, "Defense-gan: Protecting classifiers against adversarial attacks using generative models," *arXiv preprint arXiv:1805.06605*, 2018.

[217] H. De Meulemeester, J. Schreurs, M. Fanuel, B. De Moor, and J. A. Suykens, "The bures metric for generative adversarial networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2021, pp. 52–66.

[218] W. Li, L. Fan, Z. Wang, C. Ma, and X. Cui, "Tackling mode collapse in multi-generator gans with orthogonal vectors," *Pattern Recognition*, vol. 110, p. 107646, 2021.

[219] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein, "Unrolled generative adversarial networks," *arXiv preprint arXiv:1611.02163*, 2016.

[220] Z. Zhang, C. Luo, and J. Yu, "Towards the gradient vanishing, divergence mismatching and mode collapse of generative adversarial nets," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 2377–2380.

[221] B. Luo, Y. Liu, L. Wei, and Q. Xu, "Towards imperceptible and robust adversarial example attacks against neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[222] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[223] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[224] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.

[225] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[226] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[227] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–10.

[228] Y. Jiang, S. Chang, and Z. Wang, "Transgan: Two transformers can make one strong gan," *arXiv preprint arXiv:2102.07074*, vol. 1, no. 3, 2021.

[229] Z. Lv, X. Huang, and W. Cao, "An improved gan with transformers for pedestrian trajectory prediction models," *International Journal of Intelligent Systems*, vol. 37, no. 8, pp. 4417–4436, 2022.

[230] L. Sasal, T. Chakraborty, and A. Hadid, "W-transformers: A wavelet-based transformer framework for univariate time series forecasting," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 671–676.

[231] Z. Elabid, T. Chakraborty, and A. Hadid, "Knowledge-based deep learning for modeling chaotic systems," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2022, pp. 1203–1209.

[232] A. Daw, M. Maruf, and A. Karpatne, "Pid-gan: A gan framework based on a physics-informed discriminator for uncertainty quantification with physics," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 237–247.

[233] L. Yang, T. Meng, and G. E. Karniadakis, "Measure-conditional discriminator with stationary optimum for gans and statistical distance surrogates," *arXiv preprint arXiv:2101.06802*, 2021.

[234] B. Bullwinkel, D. Randle, P. Protopapas, and D. Sondak, "Deqgan: Learning the loss function for pinns with generative adversarial networks," *arXiv preprint arXiv:2209.07081*, 2022.

[235] Z. Jiang, Y. Ren, Z. Ye, J. Liu, C. Zhang, Q. Yang, S. Ji, R. Huang, C. Wang, X. Yin, *et al.*, "Mega-tts: Zero-shot text-to-speech at scale with intrinsic inductive bias," *arXiv preprint arXiv:2306.03509*, 2023.

[236] Y. Ren, M. Lei, Z. Huang, S. Zhang, Q. Chen, Z. Yan, and Z. Zhao, "Prosospeech: Enhancing prosody with quantized vector pre-training in text-to-speech," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 7577–7581.

[237] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local nash equilibria in continuous games," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 917–924.

[238] S. Arora and Y. Zhang, "Do gans actually learn the distribution? an empirical study," *arXiv preprint arXiv:1706.08224*, 2017.

[239] J. Wang, J. Lv, X. Yang, C. Tang, and X. Peng, "Multimodal image-to-image translation between domains with high internal variability," *Soft Computing*, vol. 24, pp. 18 173–18 184, 2020.

[240] I. O. Tolstikhin, S. Gelly, O. Bousquet, C.-J. Simon-Gabriel, and B. Schölkopf, "Adagan: Boosting generative models," *Advances in neural information processing systems*, vol. 30, 2017.

[241] B. Hariharan, P. Arbeláez, L. Bourdev, S. Maji, and J. Malik, "Semantic contours from inverse detectors," in *2011 international conference on computer vision*. IEEE, 2011, pp. 991–998.

[242] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.

[243] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *2018 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[244] A. Taeihagh, "Governance of artificial intelligence," *Policy and society*, vol. 40, no. 2, pp. 137–157, 2021.

[245] J. Liu, J. Huang, Y. Zhou, X. Li, S. Ji, H. Xiong, and D. Dou, "From distributed machine learning to federated learning: A survey," *Knowledge and Information Systems*, vol. 64, no. 4, pp. 885–917, 2022.

[246] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[247] M. Hausknecht and P. Stone, "Deep recurrent q-learning for partially observable mdps," in *2015 aaai fall symposium series*, 2015.

[248] J. Yang, A. Kannan, D. Batra, and D. Parikh, "Lr-gan: Layered recursive generative adversarial networks for image generation," *arXiv preprint arXiv:1703.01560*, 2017.

[249] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 2089–2093.

[250] S. Mohamed and B. Lakshminarayanan, "Learning in implicit generative models," *arXiv preprint arXiv:1610.03483*, 2016.