IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 November 2024, accepted 14 January 2025, date of publication 20 January 2025, date of current version 30 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3531726

RESEARCH ARTICLE

Exploring Generative Adversarial Networks: Comparative Analysis of Facial Image Synthesis and the Extension of Creative Capacities in Artificial Intelligence

TOMAS EGLYNAS[1], DOVYDAS LIZDENIS[1,2], AISTIS RAUDYS[2], SERGEJ JAKOVLEV[1,3], AND MIROSLAV VOZNAK[1,3], (Senior Member, IEEE)

[1]Marine Research Institute, Klaipeda University, 92294 Klaipėda, Lithuania
[2]Faculty of Mathematics and Informatics, Vilnius University, 03225 Vilnius, Lithuania
[3]Telecommunications Department, VSB-Technical University of Ostrava, 708 33 Ostrava, Czech Republic

Corresponding author: Sergej Jakovlev (sergej.jakovlev@ku.lt)

This work was supported in part by the European Union (EU) within the REFRESH Project–Research Excellence for Region Sustainability and High-Tech Industries of the European Just Transition Fund under Grant CZ.10.03.01/00/22_003/0000048; and in part by the Ministry of Education, Youth, and Sports of Czech Republic (MEYS CZ) within a Student Grant Competition in the VSB-Technical University of Ostrava under Project SGS SP2024/061.

ABSTRACT Neural networks have become foundational in modern technology, driving advancements across diverse domains such as medicine, law enforcement, and information technology. By enabling algorithms to learn from data and perform tasks autonomously, they eliminate the need for explicit programming. A significant challenge in this field is replicating the uniquely human capacity for creativity—envisioning and realizing novel concepts and tangible creations. Generative Adversarial Networks (GANs), a leading approach in this effort, are especially notable for synthesizing realistic human facial images. Despite the success of GANs, comprehensive comparative studies of face-generating GAN methodologies are limited. This paper addresses this gap by analyzing the scope and capabilities of facial generation, detailing the principles of the original GAN framework, and reviewing prominent GAN variants specifically designed for facial synthesis. Through performance evaluations and fidelity analysis of generated images, this study contributes to a deeper understanding of GAN potential in advancing artificial intelligence creativity through performance evaluations and fidelity analysis of generated images.

INDEX TERMS Human image synthesis, image processing, computer graphics, visualization, photorealism.

I. INTRODUCTION

Artificial neural networks (ANNs) have emerged as one of the most influential technologies of our era, extensively applied across various sectors including medicine [1], [2], [3], law enforcement [4], and information technology [5]. The fundamental principle of ANNs involves training algorithms with provided data to autonomously perform specific tasks. These networks excel in accurately recognizing objects and individuals in images and videos [6], [7], swiftly and precisely classifying photography [8], [9], [10], [11], identifying faces [8], [12], [13], and even determining characteristics such as gender or current emotional state based on facial recognition [14]. Moreover, ANNs can predict future changes based on historical data [4], [15], [16], demonstrating their versatile applicability across diverse domains: ranging from various engineering solutions – to social domains [17]. Practically, neural networks can be adapted wherever it is feasible to train a computer to execute high computational complexity tasks [18]. However, one of the distinct human traits is the ability to create something new and tangible. Humans can envision diverse worlds, environments, people, and scenes,

The associate editor coordinating the review of this manuscript and approving it for publication was Aysegul Ucar.

and translate these into books, music, and paintings, thereby producing real and novel creations. To mimic this creative ability using artificial intelligence in computing, various algorithms such as Convolutional Neural Networks (CNNs) [19], [20], Deep Recurrent Attentive Writers (DRAW) [21] based upon Recurrent Neural Networks (RNNs) [22], [23], [24], and Generative Adversarial Networks (GANs) [25], [26], [27], [28], [29], [30] are employed. Beginning with the CNN algorithm, it is possible to generate an image; an example explored is the generation of simple chair images. Initially, the neural network is trained using chair images as data. Upon training, by inputting specific parameters like chair type, representation, and color, the neural network can generate an image of a chair, ranging in size from 48×48 pixels to 128×128 pixels—though larger images can be produced at the expense of additional computational resources. The core concept of the DRAW algorithm involves using an encoder and decoder, which are recurrent networks, to sequentially read the input in parts considering previously completed steps and encode and store the obtained information. Subsequently, the input is decoded and reconstructed anew, much like an autoencoder [21].

Additionally, while reading the data, the encoder network is aware of the previously output decoder result, which informs it of the data that needs to be read next. Following the data reading phase, the DRAW algorithm can generate images from the stored and encoded data, creating new visuals or objects. The schematic of the DRAW algorithm is depicted in Figure 1.

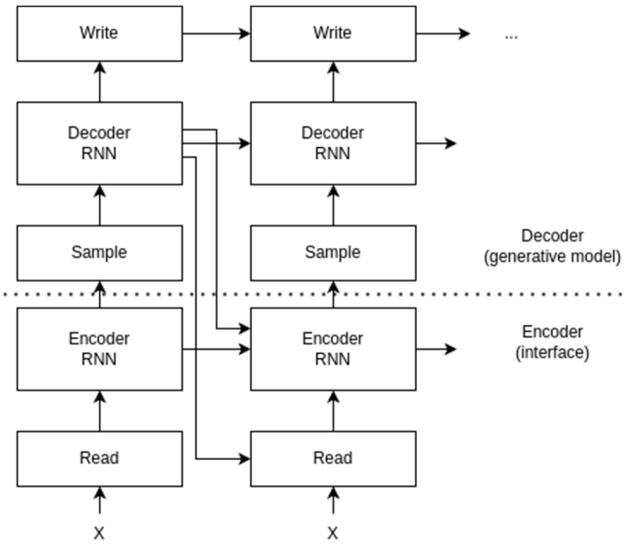


FIGURE 1. Principle scheme of the DRAW algorithm.

In Figure 1, the schematic of the DRAW algorithm is depicted, illustrating all the steps previously described. Below the dashed line, the scanning and encoding steps are shown, which lead to the ''sample'' step where information is stored and from which images are subsequently generated. Above the dashed line, the decoding and recording steps are detailed.

Additionally, arrows in the image indicate how the networks share information after each scanning phase. The Generative Adversarial Network (GAN), is currently one of the most popular for image synthesis [31]. Image synthesis plays a critical role and can be applied in fields such as art generation, computer design, photo editing, and virtual reality. GANs are also adept at creating high-quality images of human faces [32] and can be used to generate video clips featuring the faces of various individuals [33], transform from one photo to another, produce realistic text, enhance the resolution in images, convert text to image, and complete or augment parts of an image (known as image in-painting) [25], [34]. The general principle of GAN is illustrated in Figure 2.

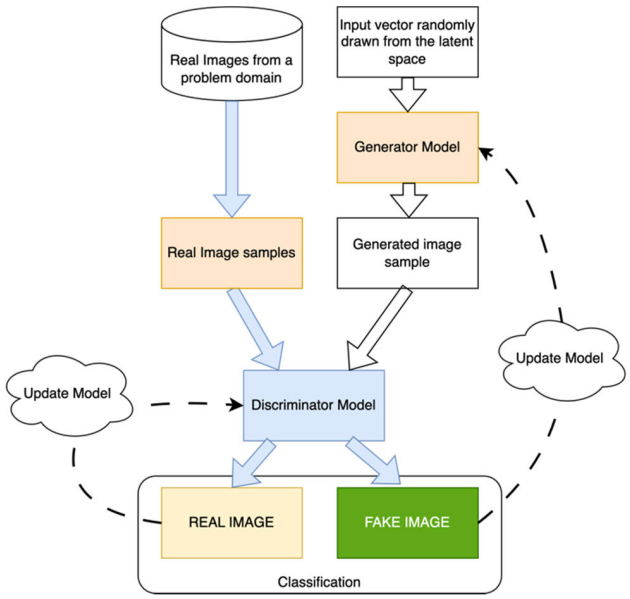


FIGURE 2. Principle scheme of the GAN algorithm.

This diagram shows the GAN architecture, which includes two competing networks: a generator and a discriminator. The generator receives random input to produce an image, which is then assessed by the discriminator along with a genuine image. The discriminator determines whether the generated image appears real. If deemed real, the image is presented as the result. If not, the generator learns from this feedback and attempts to create a new image that can convince the discriminator of its authenticity.

To address computational complexity, this study provides an analysis of each GAN model's resource demands about their architectural structures and output quality. This comparison offers a clear understanding of how each model's complexity impacts efficiency, allowing for a more informed assessment of their suitability for applications requiring both high-quality synthesis and computational feasibility.

This work will examine various GAN algorithms that generate faces, addressing the current lack of comprehensive comparative studies on facial generation algorithms.

Furthermore, as previously mentioned, unlike DRAW or CNN, GANs can replicate the human ability to create new and realistic entities [35]. This study will first explore where facial generation is applied and review the GANs currently developed for this purpose. The operational principles of the first GAN proposed by Ian Goodfellow will also be examined, followed by the generation and comparison of sample images.

Therefore, the motivation behind this work stems from the challenges faced in achieving realistic and diverse human image synthesis, a task where generative adversarial networks (GANs) have made significant strides but still confront inherent limitations like mode collapse, artifact generation, and balancing image fidelity with computational efficiency. Unlike previous studies focusing on individual GAN models, this paper systematically compares several advanced GANs, including StyleGAN2, BigGAN, and PG-GAN, to address these issues by evaluating their architectural innovations and effectiveness in synthesizing high-quality, realistic facial images. Our study reveals each model's distinct advantages and disadvantages.

II. METHODOLOGY

A. GENERAL CONCEPT OF GAN METHODOLOGY

To better understand the process of facial generation, it is crucial to delve deeper into the mechanisms and architectural modifications of specific GANs compared to the original GAN. This analytical comparison part of the paper will cover the following models:

- Even though StyleGAN3 exists StyleGAN2 has been selected due to its frequent appearance in scholarly articles, mentions, and usage in various services, making it a common choice for facial generation.
- PG-GAN was chosen for its innovative method of enhancing image clarity by increasing the number of layers during the training process, which stabilizes the learning method and enhances image variation.
- BigGAN is included for its ability to generate higher-resolution images and make them more realistic by adjusting certain parameters and scaling up the training model.

To facilitate a comprehensive understanding, the discussion will begin with the original GAN, introduced in 2014 [30]. This examination will explore its operation, architectural design, and the challenges encountered. Following this foundational overview, the discussion will shift to how the selected models modify, enhance, or update the original GAN framework.

The GAN discriminator functions as a multi-layer perceptron classifier that distinguishes between real data and synthetic data produced by the generator. It can utilize any network architecture that suits the type of data being classified. During training, the discriminator processes two types of data:

- Real data cases, such as authentic human photographs, are used as valid examples for training.
- False data instances created by the generator, are used as invalid examples for training.

The discriminator is linked to two loss functions. During training, the discriminator's loss is prioritized, and the generator's loss is disregarded. The discriminator's training involves:

- Classifying both real and synthetic data generated by the generator.
- Being penalized for errors, such as misclassifying a real instance as synthetic or vice versa.
- Updating its weights based on the loss incurred during these misclassifications.

Additionally, the discriminator only learns when the generator is not learning to avoid complications in the generator's training process.

As previously mentioned, the GAN generator learns to create synthetic data by incorporating feedback from the discriminator if it fails to deceive it. The generator masters deceiving the discriminator when it classifies a synthetic image as real. The generator's learning process involves deeper interaction with the discriminator than the discriminator's training and includes the following steps:

- Receiving random input.
- The generator network takes this input and creates new data.
- This data is then presented to the discriminator, which checks and classifies both the generated and real data, returning the outcome to the generator.
- The generator recalculates its loss based on whether it successfully deceived the discriminator or will be penalized for failing to do so.

It has been noted that the generator receives random noise as input, which it then transforms into meaningful data output. This input allows the generator to produce a variety of data, and changing the input noise can yield entirely different results. Although the noise intensity distribution has been shown not to significantly impact the results, allowing for a flexible choice of noise sources for sampling.

The GAN loss function assesses whether the neural networks are learning effectively. It is calculated using the distance measure compared against the expected outcome. The GAN model calculates two results for the loss function: one for training the generator and another for the discriminator. In the loss function, which will be detailed later, the generator and discriminator losses stem from a single distance measure between probability distributions. During training, the generator focuses solely on the distribution of synthetic data, discarding the part that reflects the real data distribution. Consequently, the losses for the generator and discriminator appear different, even though they originate from the same formula. The following section will present the loss function within the GAN architecture, where the generator aims to minimize this function, while the discriminator seeks to maximize it.

Generative Adversarial Networks can achieve impressive results and generate new data that mimic real objects, creating authentic-looking photos. However, properly training GANs can be challenging, as it is not always straightforward to synchronize the discriminator and generator during training. If the generator learns faster than the discriminator, it can lead to a scenario where the generator produces only a limited variety of data, resulting in repetitive and predictable outputs regardless of the input noise. If the discriminator gets stuck in a local minimum and cannot find the optimal strategy, it becomes easier for the generator to produce convincing results in the next iteration. Each iteration of the generator excessively optimizes for a particular discriminator, which never escapes from being trapped, thus preventing the generator from producing a diverse array of outputs. This failure mode is known as mode collapse. Conversely, if the discriminator advances too much, it can accurately reject all outcomes presented by the generator, which cannot learn effectively from it. This situation is known as the vanishing gradient problem.

Another potential issue, though less critical, occurs when both neural networks have reached their maximum potential, and the generator creates completely realistic data, leaving the discriminator guessing with a 50/50 chance whether the data is real or not. Another downside of the GAN model is the selection of hyperparameters, which is crucial and can be time-consuming to optimize correctly. If not properly configured, the GAN model will not generate good results.

B. EVALUATION FACIAL IMAGE GENERATION METHOD - GAN

Initially, the discussion will focus on StyleGAN—examining the modifications made from the original GAN model to understand why StyleGAN2 was chosen for this category. The StyleGAN model introduces numerous innovative ideas and proposals specifically within the generator component of GAN while maintaining the discriminator component unchanged. Modifications in the generator allow for the creation of faces in desired styles with specific features, a capability not feasible with the basic GAN model.

The traditional GAN model may yield different outcomes with each generated face depending on the initial input parameters, but it cannot select specific facial features. In other words, this method of generation offers no control over the characteristics of the generated outcome if one wishes to create, for instance, the face of a woman with brown hair or an older man.

Therefore, StyleGAN introduces a method to control the style of a human face, allowing for manipulation of the resultant features before generation. The results can be controlled

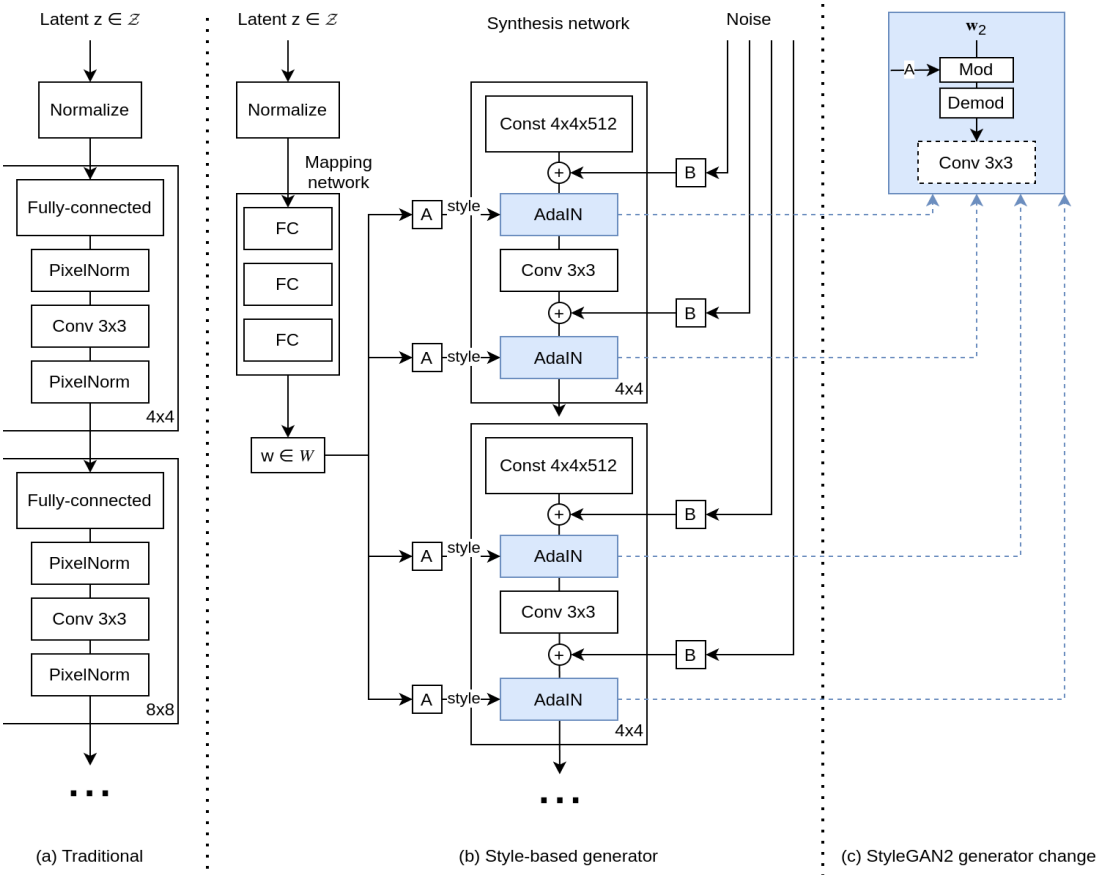


FIGURE 3. The architecture of the StyleGAN generator and modifications in StyleGAN2.

by selecting specific parameters such as pose, freckles, hair, skin color, gender, etc., right before generating the image. During the generation process, these set parameters are considered to produce a unique face with the determined style. The generation process and its capabilities can be seen in Figure 3.

This illustration depicts the architecture of traditional GAN and StyleGAN generators. In the traditional generator architecture, an initial parameter is included from which an image is subsequently generated. In the StyleGAN architecture, three new components are added: the first component is a mapping network where several initial vectors are introduced and transformed into new vectors w. These vectors are then randomly applied to a new normalization block, AdaIN, which represents the second addition. The third addition involves the implementation of random noise not only at the beginning but also at every subsequent step with varying values, allowing control over the ''style'' of the generated image or face [32].

Further modifications in StyleGAN2 primarily altered the AdaIN function. It can be observed that this function now splits into two parts. In one part, the vector is normalized and introduced in style A, while in the other, weights are distributed according to style settings and then passed into a Conv 3×3 block, like the original StyleGAN architecture [36].

These changes in the generator allow for the creation of various faces stylized according to specific needs. However, this architecture is not perfect, as the generated images may contain certain artifacts that detract from their realism. Therefore, an updated version, StyleGAN2, introduces changes in the architecture aimed at producing higher-resolution images with fewer distortions. As demonstrated, this is achieved by modifying the AdaIN normalization block and the distribution of weights according to predefined style parameters.

Thus, StyleGAN provides the capability to generate realistic faces from given parameters with specific styles, though not all images are of high quality. Some generated images may contain artifacts or other unfitting details, leading to the introduction of the new StyleGAN2. This model addresses the discussed issues by minimizing the occurrence of artifacts in the generated images.

C. BigGAN ANALYSIS

The general GAN model significantly benefits from scaling up image dimensions. By doubling or even quadrupling the parameters and increasing the batch size eightfold during the training phase, higher-quality results are achieved.

Additionally, BigGAN modifies the architecture by enhancing scaling, improving conditioning, and boosting performance. With these modifications, the BigGAN model can generate realistic images, faces, and other objects, as it was trained with a large dataset featuring over 1,000 categories [37]. Despite being trained with extensive data, the BigGAN model has its shortcomings: training the model requires substantial resources, and the initial model was trained using over a hundred GPUs, indicating that training such a model with personal resources would minimally require four GPUs (GitHub). The initially trained model was later made publicly available. Another issue with BigGAN is that the generated images may include a significant number of flawed ones with artifacts or other discrepancies, as can be seen in Figure 6.

Thus, the BigGAN model, trained with a vast array of image categories, can produce high-quality and highly realistic images. However, such a large model might generate many flawed, irrelevant, and incomprehensible images while creating the desired photo.

D. PG-GAN (ProGAN) ANALYSIS

Progressive Growing of GANs (PG-GAN), also known as ProGAN, revolutionized the training process to make it more stable. A more stable training process results in higher-quality generated images. PG-GAN achieves this stability by starting the training with very low resolutions, such as 4×4, and gradually adding layers to increase the resolution of the output from the generator and the input image size for the discriminator [38].

This process continues until the desired image size is achieved. This method has proven very effective in creating high-quality synthetic images that appear realistic. Essentially, PG-GAN introduced three key enhancements:

- Progressive Growth in Training: The resolution is progressively increased during training; starting at a low resolution such as 4×4 and incrementally increasing to the desired size, for example, 1024×1024.
- Normalization Function Change to PixelNorm: This modification helps in normalizing pixel values across the generated images to ensure consistency in quality.
- Additional Minibatch Functionality in the Discriminator: This feature improves the discriminator's ability to manage variations within a minibatch, enhancing its accuracy in distinguishing real from fake images.
- As mentioned, the first point involves step-by-step modifications during the training—starting from a low resolution and gradually increasing the resolution with each output until reaching the desired image size, such as 1024×1024. The diagram of this process is seen in Figure 4.

The image in Figure 4 demonstrates a progressive training process where both the generator (G) and the discriminator (D) begin with low resolution. As training progresses, D and G simultaneously increase the resolution of the images. According to the authors, this process can reduce training time by 2 to 6 times, depending on the desired resolution of the outcome. Additionally, this method of training yields higher-quality images.

The next modification is the implementation of PixelNorm normalization during training to ensure that the training levels of the generator and discriminator do not diverge

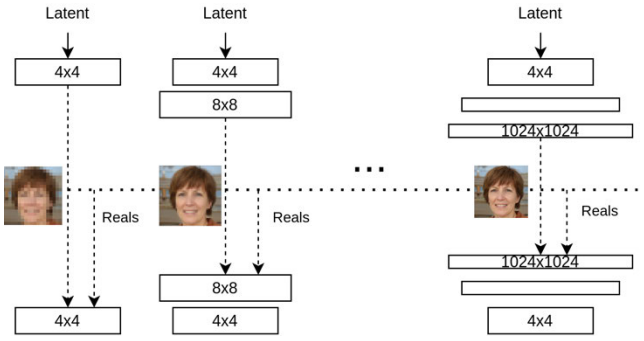


FIGURE 4. PG-GAN training framework.

significantly. Otherwise, either the generator or the discriminator might learn more rapidly than the other and dominate, thus halting the training process and leading to unsatisfactory results.

Another change is applied to the discriminator—adding what the authors call a Minibatch feature, which allows the discriminator to remember previously generated images after each iteration. This helps the discriminator determine whether the generator's image is real or fake. It forces the generator to produce more diverse images in an attempt to deceive the discriminator. This method addresses one of the common problems with GANs, where trained models produce images with limited variation.

Thus, PG-GAN introduced modifications that help stabilize the training by adding the PixelNorm normalization block, preventing either the generator or discriminator from advancing too quickly and disrupting the training. Incremental resolution enhancement up to the predefined level also supports smoother training, helping both models progressively learn. Additionally, by incorporating the Minibatch feature in the discriminator, more varied and numerous results are achievable. All these changes enable the generation of realistic and high-quality images with greater diversity than typically seen.

StyleGAN offers the capability to generate realistic faces and control their styles using specific parameters. BigGAN was trained with a large volume of photos across numerous categories. This GAN enhanced overall training parameters and photo resolution, achieving high-quality, realistic images. However, a drawback of these BigGAN modifications is that training requires substantial resources. PG-GAN introduces a novel training approach that begins at low resolutions and incrementally increases to the desired resolution. It also introduced a new normalization function that enhances the stability of the model training and added a Minibatch feature in the discriminator, which forces the generator to produce a more diverse array of images.

E. GAN EVALUATION METRICS: INCEPTION SCORE (IS), FRÉCHET INCEPTION DISTANCE (FID)

Inception Score (IS) [10], [35], [39], Fréchet Inception Distance (FID) [10], [25], [31], [35], [39], [40], [41] evaluation metrics were chosen because they are among the primary objective methods of comparison used by authors in scholarly articles. Additionally, these methods provide quantitative results that demonstrate the capabilities of a GAN model. The evaluation begins with the Inception Score (IS), which measures two main criteria:

- *Photo Quality:* This assesses whether the image reflects the type of the preceding images; for example, if one image contains a cat, subsequent images should also feature cats.
- *Photo Diversity:* This assesses whether the generated images are varied despite being of the same type, such as generating images of cats where each generated image features a different breed.

If the images produced by a GAN model satisfy both criteria, the IS result will be high. Theoretically, there is no maximum value for IS. If the generated images do not meet these criteria, the IS result will be low, with the minimum IS value potentially being 0. Therefore, the higher the IS result, the more diversity the GAN model can generate, capable of producing recognizable and general category images such as cats, dogs, faces, or other visuals.

The overall result is determined based on the distribution of images. If we have generated an image from which it is determined what is seen and assigned to a category, for example, a dog. If most classified images were of dogs, that means that the generated images are primarily recognized as dogs, which would indicate a higher IS result because it signifies that the generated images are of high quality and recognizable as belonging to the same class. If the graph were evenly distributed, it would indicate a lower IS result because the images are not of good quality, thus not understanding the general generated category. The limitation of IS is that it evaluates overall image quality and diversity but does not assess realism.

The realism score is provided by the Fréchet Inception Distance (FID) evaluation method. FID assesses a collection of generated images against a collection of real images from the generated domain. This evaluation, combined with the IS assessment, helps determine whether a GAN model produces high-quality, varied, and realistic images. FID is calculated by comparing the distances between the vectors of the generated images and the real images according to the generated category, thus assessing realism. The lower the FID score, the more realistic the generated image; a higher FID score indicates that the generated image is less realistic.

III. RESULTS

Evaluations of StyleGAN2, BigGAN, and PG-GAN for facial generation are detailed using the Inception Score (IS) and Fréchet Inception Distance (FID) metrics, with the results displayed in Table 1. Here, the performance of various GAN models using Inception Score (IS) and Fréchet Inception Distance (FID) metrics is compared using the Flickr-Faces-HQ (FFHQ) open dataset provided by NVLabs.

For our study, we selected a subset of 10,000 images (out of 70.000 available), preserving the dataset's richness while ensuring computational efficiency for evaluating our GAN models.

TABLE 1. Comparison of GANs Generating Faces Based on Inception Score (IS) and Fréchet Inception Distance (FID).

GAN models	Evaluation methods (Results Obtained from Other Sources)		Evaluation methods (Results Obtained from Facial Comparisons)	
	IS	FID	IS	FID
StyleGAN2	5,17	2,70	7,54	31,69
BigGAN	166,5	7,4	5,97	212,77
PG-GAN	8,8	7,3	7,38	186,79

These results include ratings derived from both published articles where these models were originally introduced and from newly generated images using the same generative models. Upon reviewing the data, it is evident that there are discrepancies between the evaluations reported in the literature and those obtained from our experiments. The IS, which quantifies the diversity of generated images, was highest for BigGAN in published results, reflecting its ability to produce a wide variety of image categories. However, our experiments showed that while BigGAN excels in diversity, it tends to generate fewer categories specifically related to human features, which impacts its IS rating. StyleGAN2, on the other hand, was found to generate the most realistic images according to the FID metric, indicating fewer discrepancies between generated and actual human images.

The observed differences in results could be due to several factors. For instance, in the case of StyleGAN2, a better IS score might suggest that a larger set of generated images was used in the study, potentially including a higher number of duplicates, which artificially inflates the diversity measure. Conversely, a higher FID in published studies may result from using a larger sample of images, which increases the likelihood of capturing realistic variations.

For BigGAN, the significantly lower IS and FID scores in our tests compared to those reported may be attributed to its application to a narrower range of categories, primarily focused on human-related images, unlike its typical use across thousands of varied categories. This specialization might not fully utilize BigGAN's capacity for generating diverse image types, thus reflecting poorer performance in our specific test setup. PG-GAN showed similar IS values to those reported, but significant differences in FID were noted, possibly due to the presence of minor artifacts in the generated images. While these artifacts are not overly prominent, they can affect the realism of the images and thus the FID score.

Summarizing the results from both the literature and our findings, BigGAN exhibits the highest diversity among the models tested. However, StyleGAN2 consistently produces the most realistic images, followed closely by PG-GAN and then BigGAN. Moving forward, the generated outcomes illustrate that StyleGAN2 not only produces a broader array of images compared to other GANs but also maintains superior realism. Images generated by the BigGAN model are shown in Figure 5. This experiment highlighted BigGAN's strength in producing a wider variety of facial images, indicating its suitability for tasks requiring high diversity in synthetic faces, though it is computationally intensive.

FIGURE 5. Results obtained from BigGAN.

BigGAN was used to generate faces within several broad categories that include human elements, such as tennis players, brides, glasses, and wigs because it lacks a specific category for faces. The next experiment emphasized PG-GAN's ability to balance resource efficiency with high-quality outputs, making it a strong candidate for applications where stability and computational feasibility.

The results display certain resemblances to these generated categories, but they do not achieve realism in terms of facial generation. Following this, the results generated by PG-GAN are presented in Figure 6. PG-GAN's results appear more realistic than those generated by BigGAN, yet they also have their shortcomings, such as inaccurately generated hair, artifacts within the images, and a significant number of duplicates.

FIGURE 6. Results obtained from PG-GAN.

Next, the results generated by StyleGAN2 are presented in Figure 7. We evaluated StyleGAN2's ability to produce high-quality, photorealistic facial images with minimal artifacts. We focused on measuring Fréchet Inception Distance (FID). The purpose of this experiment was to demonstrate StyleGAN2's superior performance in artifact reduction and photorealism, positioning it as the optimal model for applications demanding high-quality facial synthesis.

The results obtained with StyleGAN2 are the most realistic compared to those generated by BigGAN and PG-GAN. Additionally, StyleGAN2 exhibits greater diversity and fewer artifacts in the images. While not all images are perfect,

FIGURE 7. Results obtained from StyleGAN2.

they present a wider variety of realistic faces without extensive searching among generated images. Considering these results, StyleGAN2 emerges as the best GAN for facial generation due to its ability to produce the most realistic images and the greatest diversity.

This is also reflected in the obtained Inception Score (IS) and Fréchet Inception Distance (FID) metrics. PG-GAN and BigGAN follow, with BigGAN being less suitable for facial generation as its images lack realistic outcomes. PG-GAN, while capable of generating faces that resemble humans, often produces similar and artifact-ridden results compared to StyleGAN2. Moreover, StyleGAN2's ability to control facial styles—such as general features, age, and skin tone—offers a significant advantage for generating specific faces or datasets more easily and flexibly.

It is essential to acknowledge certain limitations that may impact the broader applicability of our findings. First, the computational requirements for training models like BigGAN are notably high, which may restrict accessibility for researchers and practitioners using standard hardware. This study primarily focuses on objective evaluation metrics, such as FID, and IS; however, certain subjective aspects of image quality—such as nuanced facial expressions and detailed textures—could benefit from additional user-based assessments.

Furthermore, while StyleGAN2 performed well in artifact reduction, occasional artifacts were observed in PG-GAN and BigGAN outputs, highlighting a need for further optimization. Addressing these limitations in future work could enhance the practical relevance of GANs for diverse real-world applications requiring high-quality, varied facial synthesis.

IV. CONCLUSION

In this part of the study, we explored image evaluation metrics such as the Inception Score (IS) and the Fréchet Inception Distance (FID), which are used to assess the diversity, quality, and realism of generated images. By employing these evaluation methods, we compared the images produced by StyleGAN2, BigGAN, and PG-GAN models. The results indicated that the StyleGAN2 generative model achieved the highest ratings, although BigGAN generated the greatest diversity of images. In terms of realism, as indicated by the results, StyleGAN2 also produced the most lifelike images.

Reviewing these generative models visually, it was evident that StyleGAN2 consistently delivered superior outcomes, as exemplified in Figure 7. While these results are impressive, a closer analysis of the generated images reveals even more visually appealing faces, examples of which are presented in Figure 8.

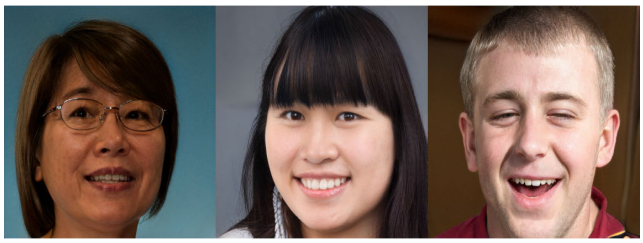


FIGURE 8. Realistically generated faces selected from StyleGAN2.

As demonstrated in Figure 8, the faces depicted are generated with remarkable realism, featuring only minimal noticeable artifacts, creating the impression that the images could be of real people. However, despite the presence of highly realistic images, there are instances where generated artifacts are visible, detracting from the realism of the photographs, as illustrated in Figure 9.

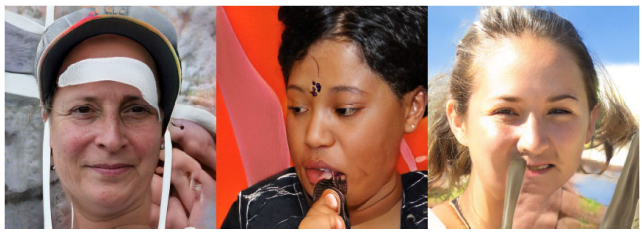


FIGURE 9. Non-realistically generated faces selected from StyleGAN2.

Based on Figure 9, it is apparent that the depicted individuals are artificially generated, as evidenced by blurred facial details or features that do not naturally occur. In addition to the facial features, the hair structure and background also indicate that the images are synthesized. This suggests that while the images appear artificial, they provide useful insights into the capabilities and limitations of generative models. In summary, StyleGAN2 proves to be effective for facial generation; the results obtained from this model are generally realistic, although, like other models, it occasionally produces errors during generation. These errors can manifest as unnatural features or distorted elements, underscoring the ongoing challenges in achieving flawless realism in generated images.

In conclusion, this study has demonstrated the efficacy of generative adversarial networks (GANs) in producing facial images, with a particular focus on the performance of StyleGAN2, BigGAN, and PG-GAN. The evaluation using metrics such as the Inception Score and the Fréchet Inception Distance revealed that StyleGAN2 consistently outperformed the other models in terms of image quality and realism, albeit with some occurrence of artifacts. While BigGAN excelled in generating a diverse array of images, it occasionally produced

images with noticeable distortions. PG-GAN was notable for its high-quality image generation, modifying the training principles of traditional GANs to achieve better results.

StyleGAN2 consistently ranks highest, making it well-suited for applications where image fidelity is paramount. BigGAN's IS scores underscore its utility in generating a diverse range of images, albeit with substantial computational demands. Meanwhile, PG-GAN's stability and density performance make it a solid choice for generating high-quality images, particularly when resource efficiency is a priority.

Despite the advancements in facial generation technology demonstrated by these models, the presence of artifacts and occasional unrealistic features highlights the challenges that still lie ahead in the field of synthetic image generation. Future research should focus on refining these models to minimize errors and enhance the realism and utility of generated images for practical applications in various domains such as digital media, entertainment, and security.

REFERENCES

[1] W. Xu, Y.-L. Fu, H. Xu, and K. K. L. Wong, "Medical image fusion using enhanced cross-visual cortex model based on artificial selection and impulse-coupled neural network," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107304, doi: 10.1016/j.cmpb.2022.107304.

[2] S. Yunhong, Y. Shilei, Z. Xiaojing, and Y. Jinhua, "Edge detection algorithm of MRI medical image based on artificial neural network," *Proc. Comput. Sci.*, vol. 208, pp. 136–144, Jan. 2022, doi: 10.1016/j.procs.2022.10.021.

[3] S. Abut, H. Okut, and K. J. Kallail, "Paradigm shift from artificial neural networks (ANNs) to deep convolutional neural networks (DCNNs) in the field of medical image processing," *Expert Syst. Appl.*, vol. 244, Jun. 2024, Art. no. 122983, doi: 10.1016/j.eswa.2023.122983.

[4] J. Chen, W. Tao, Z. Jing, P. Wang, and Y. Jin, "Traffic accident duration prediction using multi-mode data and ensemble deep learning," *Heliyon*, vol. 10, no. 4, Feb. 2024, Art. no. e25957, doi: 10.1016/j.heliyon.2024.e25957.

[5] L. Li, X. Sheng, B. Du, Y. Wang, and B. Ran, "A deep fusion model based on restricted Boltzmann machines for traffic accident duration prediction," *Eng. Appl. Artif. Intell.*, vol. 93, Aug. 2020, Art. no. 103686, doi: 10.1016/j.engappai.2020.103686.

[6] K. Om, R. Singh, Snehdeep, A. Kaur, Deepika, A. Kaur, T. McGill, M. Dixon, K. W. Wong, and P. Koutsakis, "Artificial intelligence—Based video traffic policing for next generation networks," *Simul. Model. Pract. Theory*, vol. 121, Dec. 2022, Art. no. 102650, doi: 10.1016/j.simpat.2022.102650.

[7] W. Hu, X. Li, C. Li, R. Li, T. Jiang, H. Sun, X. Huang, M. Grzegorzek, and X. Li, "A state-of-the-art survey of artificial neural networks for whole-slide image analysis: From popular convolutional neural networks to potential visual transformers," *Comput. Biol. Med.*, vol. 161, Jul. 2023, Art. no. 107034, doi: 10.1016/j.compbiomed.2023.107034.

[8] C. Gautam and K. R. Seeja, "Facial emotion recognition using handcrafted features and CNN," *Proc. Comput. Sci.*, vol. 218, pp. 1295–1303, Jan. 2023, doi: 10.1016/j.procs.2023.01.108.

[9] R. Kumar, J. Sotelo, K. Kumar, A. de Brebisson, and Y. Bengio, "ObamaNet: Photo-realistic lip-sync from text," 2017, *arXiv:1801.01442*.

[10] M. Toshpulatov, W. Lee, and S. Lee, "Talking human face generation: A survey," *Expert Syst. Appl.*, vol. 219, Jun. 2023, Art. no. 119678, doi: 10.1016/j.eswa.2023.119678.

[11] G. Kaur, R. Sinha, P. K. Tiwari, S. K. Yadav, P. Pandey, R. Raj, A. Vashisth, and M. Rakhra, "Face mask recognition system using CNN model," *Neurosci. Informat.*, vol. 2, no. 3, Sep. 2022, Art. no. 100035, doi: 10.1016/j.neuri.2021.100035.

[12] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, Apr. 2021, doi: 10.3390/s21093046.

[13] S. A. Prome, N. A. Ragavan, M. R. Islam, D. Asirvatham, and A. J. Jegathesan, "Deception detection using machine learning (ML) and deep learning (DL) techniques: A systematic review," *Natural Lang. Process. J.*, vol. 6, Mar. 2024, Art. no. 100057, doi: 10.1016/j.nlp.2024.100057.

[14] A. P. Fard and M. H. Mahoor, "Ad-corre: Adaptive correlation-based loss for facial expression recognition in the wild," *IEEE Access*, vol. 10, pp. 26756–26768, 2022, doi: 10.1109/ACCESS.2022.3156598.

[15] F. Hu, Q. Yang, J. Yang, Z. Luo, J. Shao, and G. Wang, "Incorporating multiple grid-based data in CNN-LSTM hybrid model for daily runoff prediction in the source region of the yellow river basin," *J. Hydrol., Regional Stud.*, vol. 51, Feb. 2024, Art. no. 101652, doi: 10.1016/j.ejrh.2023.101652.

[16] M. Schulz, B. Boughattas, and F. Wendel, "DetEEktor: Mask R-CNN based neural network for energy plant identification on aerial photographs," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100069, doi: 10.1016/j.egyai.2021.100069.

[17] S. Barman and R. Bandyopadhyaya, "Modelling crash severity outcomes for low speed urban roads using back propagation—Artificial neural network (BP—ANN)—A case study in Indian context," *IATSS Res.*, vol. 47, no. 3, pp. 382–400, Oct. 2023, doi: 10.1016/j.iatssr.2023.08.002.

[18] E. Saetta, R. Tognaccini, and G. Iaccarino, "Uncertainty quantification in autoencoders predictions: Applications in aerodynamics," *J. Comput. Phys.*, vol. 506, Jun. 2024, Art. no. 112951, doi: 10.1016/j.jcp.2024.112951.

[19] Z. Li, Y. Wu, B. Peng, X. Chen, Z. Sun, Y. Liu, and D. Yu, "Extended abstract of SeCNN: A semantic CNN parser for code comment generation," in *Proc. IEEE Int. Conf. Softw. Anal., Evol. Reengineering (SANER)*, Mar. 2023, pp. 848–849, doi: 10.1109/SANER56733.2023.00099.

[20] Z. Li, Y. Wu, B. Peng, X. Chen, Z. Sun, Y. Liu, and D. Yu, "SeCNN: A semantic CNN parser for code comment generation," *J. Syst. Softw.*, vol. 181, Nov. 2021, Art. no. 111036, doi: 10.1016/j.jss.2021.111036.

[21] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "DRAW: A recurrent neural network for image generation," 2015, *arXiv:1502.04623*.

[22] A. Orvieto, S. L. Smith, A. Gu, A. Fernando, C. Gulcehre, R. Pascanu, and S. De, "Resurrecting recurrent neural networks for long sequences," 2023, *arXiv:2303.06349*.

[23] A. Kumar, N. Gaur, S. Chakravarty, M. H. Alsharif, P. Uthansakul, and M. Uthansakul, "Analysis of spectrum sensing using deep learning algorithms: CNNs and RNNs," *Ain Shams Eng. J.*, vol. 15, no. 3, Mar. 2024, Art. no. 102505, doi: 10.1016/j.asej.2023.102505.

[24] S. Yadav, H. Hashmi, D. Vekariya, A. K. N. Zafar, and F. J. Vijay, "Mitigation of attacks via improved network security in IoT network environment using RNN," *Meas., Sensors*, vol. 32, Apr. 2024, Art. no. 101046, doi: 10.1016/j.measen.2024.101046.

[25] E. Yauri-Lozano, M. Castillo-Cara, L. Orozco-Barbosa, and R. García-Castro, "Generative adversarial networks for text-to-face synthesis & generation: A quantitative–qualitative analysis of natural language processing encoders for Spanish," *Inf. Process. Manage.*, vol. 61, no. 3, May 2024, Art. no. 103667, doi: 10.1016/j.ipm.2024.103667.

[26] N. Dey, A. Chen, and S. Ghafurian, "Group equivariant generative adversarial networks," in *Proc. 9th Int. Conf. Learn. Represent.*, 2021, pp. 1–28.

[27] L. Wang, Y. Li, J. Liu, J. Peng, Q. Zhang, and W. Fu, "Research on fault diagnosis of industrial robots based on generative adversarial network," *Phys. Commun.*, vol. 64, Jun. 2024, Art. no. 102355, doi: 10.1016/j.phycom.2024.102355.

[28] L. Yin and C. Lin, "Matrix Wasserstein distance generative adversarial network with gradient penalty for fast low-carbon economic dispatch of novel power systems," *Energy*, vol. 298, Jul. 2024, Art. no. 131357, doi: 10.1016/j.energy.2024.131357.

[29] J. He, Y. Xu, Y. Pan, and Y. Wang, "Adaptive weighted generative adversarial network with attention mechanism: A transfer data augmentation method for tool wear prediction," *Mech. Syst. Signal Process.*, vol. 212, Apr. 2024, Art. no. 111288, doi: 10.1016/j.ymssp.2024.111288.

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[31] J. Honkamaa, U. Khan, S. Koivukoski, M. Valkonen, L. Latonen, P. Ruusuvuori, and P. Marttinen, "Deformation equivariant cross-modality image synthesis with paired non-aligned training data," *Med. Image Anal.*, vol. 90, Dec. 2023, Art. no. 102940, doi: 10.1016/j.media.2023.102940.

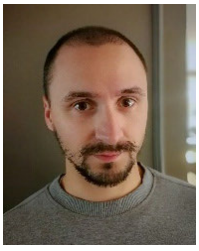
[32] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021, doi: 10.1109/TPAMI.2020.2970919.
[33] P. Li, H. Zhao, Q. Liu, P. Tang, and L. Zhang, "TellMeTalk: Multimodal-driven talking face video generation," *Comput. Electr. Eng.*, vol. 114, Mar. 2024, Art. no. 109049, doi: 10.1016/j.compeleceng.2023.109049.
[34] J. Agnese, J. Herrera, H. Tao, and X. Zhu, "A survey and taxonomy of adversarial neural networks for text-to-image synthesis," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 4, p. e1345, Jul. 2020, doi: 10.1002/widm.1345.
[35] A. Brock, J. Donahue, and K. Simonyan, "Large scale GaN training for high fidelity natural image synthesis," in *Proc. 7th Int. Conf. Learn. Represent.*, 2019, pp. 1–35.
[36] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," 2019, *arXiv:1912.04958*.
[37] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.
[38] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, *arXiv:1710.10196*.
[39] N. Rendon, J. H. Giraldo, T. Bouwmans, S. Rodríguez-Buritica, E. Ramirez, and C. Isaza, "Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning," *Eng. Appl. Artif. Intell.*, vol. 124, Sep. 2023, Art. no. 106635, doi: 10.1016/j.engappai.2023.106635.
[40] L. F. Buzuti and C. E. Thomaz, "Fréchet AutoEncoder distance: A new approach for evaluation of generative adversarial networks," *Comput. Vis. Image Understand.*, vol. 235, Oct. 2023, Art. no. 103768, doi: 10.1016/j.cviu.2023.103768.
[41] E. J. Nunn, P. Khadivi, and S. Samavi, "Compound Fréchet inception distance for quality assessment of GAN created images," 2021, *arXiv:2106.08575*.

TOMAS EGLYNAS was born in Lithuania. He received the bachelor's degree in informatics engineering and the master's degree in technical information systems engineering from Klaipeda University, in 2010 and 2012, respectively, and the Ph.D. degree in transport engineering through the Lithuanian Joint Transport Engineering Ph.D. Program (KU, VGTU, and ASU), in 2017. Currently, he holds the position of a Senior Researcher with the Marine Research Institute, Klaipeda University, and the CEO of Inotecha Ltd., a company based in Lithuania that develops intelligent industrial solutions. He has authored or co-authored numerous research articles published in Scopus and CA-indexed journals. His research interests include intelligent control systems, logistics 4.0, control theory, electronics, and the IoT applications in industry 4.0.

DOVYDAS LIZDENIS was born in Lithuania. He received the bachelor's degree (Hons.) in informatics from Klaipeda State University of Applied Sciences, in 2020. He is currently pursuing the master's degree in informatics with Vilnius University, focusing on advanced topics in software and artificial intelligence. Alongside his studies, he serves as the CEO of Digital Moose Ltd., a Lithuania-based technology company specializing in the development of software and web-based solutions tailored to various industries. His research interests lie in software engineering, AI applications in engineering, and computer science, with a particular emphasis on leveraging AI to solve complex engineering challenges.

AISTIS RAUDYS received the Ph.D. degree from Vilnius University, Lithuania, where his research focused on developing methods for extracting features from multivariate data.

Currently, he is a Professor with the Faculty of Mathematics and Informatics, Vilnius University, where he teaches courses related to algorithmic trading and robotics. In addition to his academic role. He is the CEO and the Co-Founder of AAI-Labs, an AI startup dedicated to the development of technologies in the fields of voice recognition, risk assessment, and transport optimization. Before his current position, he gained experience as a researcher working with prominent financial institutions, including Deutsche Bank, Société Générale, and BNP Paribas, London, U.K. He has authored 50 publications in the areas of engineering, robotics, automated trading, and artificial intelligence. His primary research interests include financial engineering, robotics, automated trading, and artificial intelligence.



SERGEJ JAKOVLEV was born in Lithuania. He received the bachelor's degree in informatics engineering from Klaipeda University, in 2009, the master's degree in technical information systems engineering, in 2011, and the Ph.D. degree in transport engineering, in 2016.

Currently, he is the Chief Researcher of the Marine Research Institute, Klaipeda University, and a Researcher with the Department of Telecommunications, VSB-Technical University of Ostrava. A prolific contributor to his field, he is actively involved in several editorial boards and international science conferences, such as the International Science Conference on Computer Networks. He also reviews prestigious CA and Scopus journals, including *Soft Computing* and IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS. He has authored or co-authored over thirty articles in Scopus and CA-indexed journals and proceedings. His primary research interests include intelligent transportation systems, AI theory and applications, big data analytics tools and methods, and wireless sensor networks.

MIROSLAV VOZNAK (Senior Member, IEEE) received the Ph.D. degree in telecommunications, in 2002, and the Habilitation degree, in 2009.

He was appointed a Full Professor, in 2017. He is currently a Professor with the Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava. According to WoS, he has published more than 200 articles in SCIE journals, such as IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and *IEEE Communications Magazine*. He participated in seven projects within EU funding programs, mostly as an institutional coordinator, and more than twenty national projects. Since 2020, he has been ranked regularly among the World's Top 2% Scientists (Career Impact DB) by Stanford University in the subfield of Networking and Telecommunications. His research interests include the IoT, QoS/QoE, wireless networks, network security, and big data analytics in networks.