

Received 15 September 2023, accepted 26 September 2023, date of publication 2 October 2023, date of current version 9 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3321149

RESEARCH ARTICLE

ChildGAN: Large Scale Synthetic Child Facial Data Using Domain Adaptation in StyleGAN

MUHAMMAD ALI FAROOQ¹, WANG YAO¹, (Graduate Student Member, IEEE),
GABRIEL COSTACHE^{1,2}, AND PETER CORCORAN¹, (Fellow, IEEE)

¹College of Science and Engineering, University of Galway, Galway, H91 TK33 Ireland

²Xperi Corporation, Galway, H91 V0TX Ireland

Corresponding author: Muhammad Ali Farooq (muhammadali.farooq@universityofgalway.ie)

This work was supported in part by the Data-Center Audio/Visual Intelligence On-Device (DAVID) Project (2020–2023) funded by the Disruptive Technologies Innovation Fund (DTIF), Irish Research Council Enterprise Partnership Ph.D. Scheme, under Grant EPSPG/2020/40; and in part by Xperi Corporation, Ireland.

ABSTRACT In this research work, we proposed a novel ChildGAN, a pair of GAN networks for generating synthetic boys and girls facial data derived from StyleGAN2. ChildGAN is built by performing smooth domain transfer using transfer learning. It provides photo-realistic, high-quality data samples. A large-scale dataset is rendered with a variety of smart facial transformations: facial expressions, age progression, eye blink effects, head pose, skin and hair color variations, and variable lighting conditions. The dataset comprises more than 300k distinct data samples. Further, the uniqueness and characteristics of the rendered facial features are validated by running different computer vision application tests which include CNN-based child gender classifier, face localization and facial landmarks detection test, identity similarity evaluation using ArcFace, and lastly running eye detection and eye aspect ratio tests. The results demonstrate that synthetic child facial data of high quality offers an alternative to the cost and complexity of collecting a large-scale dataset from real children. The complete dataset along with the trained model are open-sourced on our GitHub website and GitHub page: <https://github.com/MAli-Farooq/ChildGAN>.

INDEX TERMS Dataset, GANs, GDPR, transformations, transfer learning.

I. INTRODUCTION

The majority of artificial intelligence and machine learning applications require large, curated and diverse datasets. It can be challenging to perform optimal training and rigorous validation of deep learning networks with limited, imbalanced, noisy data and if the data has inherent bias due to limitations in the data acquisition process which may not be known at the time the data is collected. Data acquisition itself is an expensive, and time-consuming process, particularly when human subjects are involved, and in many cases, it is technically impossible or impractically expensive to accurately replicate acquisitions. Further, when data is collected from human subjects there is the additional complication of

new data protection regulations for personally identifiable data.

In the EU, the General Data Protection Regulations (GDPR) provides a broad interpretation of personally identifiable data (PID) which extends to facial biometrics and speech data. This creates significant complexity when gathering any video, image, or audio data of human subjects as the scope of use of such data and any subsequent data-processing must be clearly defined and explained to the data subject and explicit consent is normally required. Further, the data must be securely stored, and a range of rights must be supported, including the right of the data-subject to request the removal of the stored data at any time. This becomes even more complex when dealing with vulnerable data-subjects such as children where consent is required from legal guardians and ideally the child-subject should be informed about the scope of data use in non-technical language. Synthetic human facial

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li¹.



FIGURE 1. Novel photo-realistic child data samples generated using ChildGAN with various smart image transformations.

TABLE 1. Open-source child datasets.

Dataset	Number of Subjects	Age	Attributes
The DuckEES Dataset [1]	N/A	8 to 18 year olds	<ol style="list-style-type: none"> 1. The DuckEES database contains dynamic emotional stimuli by children and adolescent actors. 2. It has a greater representation of videos thus comprising 142 posed videos which are distributed relatively homogeneously by emotion. 3. However, the DuckEES dataset did not include some of the universal emotions such that anger, contempt, surprise.
The Child Emotion Facial Expression dataset [9]	132 Children	4 to 6 year olds	<ol style="list-style-type: none"> 1. A total of 29 hours of video were recorded in the studio for capturing various child facial expressions. 2. The overall dataset comes with eight distinct facial expressions which include contempt, happiness, fear neutrality, disgust, anger, surprise, and sadness. 3. The number of stimuli of each emotion is as follows: neutrality 87, happiness 363, disgust 170, surprise 104, fear 153 (48 induced, sadness 144, anger 157, and contempt 183).
The Child Affective Facial Expression (CAFE) dataset [10]	154 Child Models	2 to 8 year olds	<ol style="list-style-type: none"> 1. Comprises data of 7 different facial expressions which include sadness, happiness, surprise, anger, disgust, fear, and neutral face. 2. Not all children were able to successfully pose for all 7 expressions, so all unsuccessful attempts were eliminated from the set. 3. The result output data consist of 1192 total photographs.

data is not directly associated with a living human person and thus it is not subjected to data protection regulations.

In this work we investigate the potential to adapt state-of-the-art (SOTA) generative neural models [2], [3], combining these with a range of other proven tools and algorithms to develop a large-scale synthetic dataset of facial biometrics for a vulnerable population - young children. This work is inspired by a need to adapt and tune a range of SOTA computer vision algorithms, trained on adult populations, for smart mobile and embedded vision systems that will engage with young children. The primary objective of this work is to create a substantial synthetic facial dataset focused on children through the implementation of smooth domain adaptation within the StyleGAN2 framework. The novelty of our approach is exemplified by the production of a high-quality synthetic dataset, consisting of over 300,000

facial data samples. Among these, we have meticulously rendered 10,000 unique facial data samples for both boys and girls, each providing a high-quality image resolution of 1024×1024 pixels. To increase the utility of this data for a variety of AI applications it is important to introduce a broad range of smart transformations thus building new data attributes so that the base data samples can be adjusted to support numerous training and test use cases. The potential applications of this novel synthetic child dataset include but are not limited to child gender classification, face localization, facial landmark detection, child facial expression analysis, generating 3D facial poses using single or multiple 2D images, performing domain transfer to generate a variety of facial expressions, facial accessories, etc. Having a unique starting identity enables the adaptation of many current neural machine vision models

for use with children. The lack of availability of children's facial data is a challenge and current data regulations make it very challenging to gather such a dataset at scale. This is achieved by exploring and manipulating the latent space in StyleGAN2, to generate a wide range of diverse and customizable outputs while maintaining consistency and realism in the generated artificial data. This ability to control various features within the latent space is one of the strengths of the proposed study. To the best of our knowledge there is no other Child Synthetic Dataset with respect to other existing works which has provided that much amount of facial data samples along with this type of smart facial transformations as discussed in core contributions thus making this dataset novel and first of its type that offers a comprehensive set of features for studying and analyzing facial characteristics and transformations in child data. The core contributions of this research study are as follows.

- The ChildGAN model is provided with tunings to enable separately generating boys and girls facial samples.
- A large-scale high-quality synthetic dataset comprising more than 300k facial data samples is made available to the research community.
- The dataset incorporates a variety of smart facial transformations such that 4 different facial expressions, age progression, eye blinking effects, head pose variations, skin and hair color variations, and distinct lighting conditions are available.
- The data quality is validated using various computer vision algorithms to test the uniqueness and facial features quality of rendered data.

The adapted training methodology, open-sourced tuned models, along with rendered child data with intelligent data augmentations has a huge potential for commercial consumer technology applications and for industrial applications as well. In industries where sensitive or confidential data is involved, this type of human synthetic dataset can be a valuable tool for maintaining privacy and security. By creating synthetic data that retains the statistical properties of the original data but does not contain personally identifiable information (PII), organizations can share data for research or testing without risking data breaches. Further child synthetic datasets can be used to train machine learning models for various downstream tasks such as child-based efficient gender classifier and smart attendance systems, child facial detection and classifications system for smart embedded toys, drowsiness, and fatigue detection in children when real data is limited or unavailable or also sometime not possible to collect due to privacy regulations. Lastly, in cases where real datasets are small or imbalanced, synthetic data can be generated to augment the dataset. This can improve the performance of machine learning models by providing more diverse and representative examples.

The rest of the paper is structured as follows. Section II provides the background and related published work in this domain, section III details the proposed algorithms, section IV presents the training and validation results, and

section V presents the effectiveness and quality of newly generated child synthetic datasets by incorporating 3 different computer vision based validation tests, and lastly, section VI provides the conclusion and future directions of the proposed research work.

II. LITERATURE REVIEW

Most of the publicly available large-scale face datasets such as FFHQ [3], CelebA [6], VGGFace2 [7], and DigiFace-1M [8] are usually acquired from adults. It is challenging to find large scale real child dataset with diversified data attributes. Table 1 summarizes the open-source child datasets that are publicly available to the best of our knowledge however the number of these datasets and the amount of information available through these datasets is significantly less as compared to large-scale adult face recognition (FR) datasets. When coming towards generating synthetic data we can find various algorithms among which SMOTE [11], ADASYN [12], Variational AutoEncoders [13], and Generative Adversarial Networks (GANs) [14] are a few techniques for generating precise results. In this research, we have focused on generating a scalable set of synthetic child data using GAN architecture. When producing large-scale synthetic data, it should be aimed that generated samples are statistically accurate and free of historical biases. This will eventually be beneficial for the optimal training and validation of dense machine learning pipelines also referred to as "black box" [15].

With the recent advancements in designing more efficient deep learning models, we can find numerous SoA open-sourced GAN architectures based on different loss functions. Some of these include the standard Vanilla GAN [14], LSGAN [16], WGAN [17], WGAN with gradient penalty [18], Deep Regret Analytic GAN [19], [20], CramerGAN [21], Conditional WGAN [22], and GAN for time series as well also known as TimeGAN [23]. Recently a huge improvement is done to discriminator models in GAN architecture as an effort to train more effective generator models.

Nvidia's research team published Style Generative Adversarial Network also known as StyleGAN [3] as an extension to the existing GAN architecture that proposes large changes to the generator model. This includes the use of a mapping network to map points in latent space to an intermediate latent space. Secondly, they have focused on using the intermediate latent space to control style at each point in the generator model, and finally the introduction to noise as a source of variation at each point in the generator model. This in turn makes the proposed architecture capable of generating high-quality photo-realistic synthetic data. It also provides control over the style of the generated image at different levels of detail by changing the style vectors and noise level. Based on the impressive results of StyleGAN in data-driven unconditional generative image modeling, a newer version of StyleGAN referred to as StyleGAN2 [2] was released with

further improvements. This includes restructuring adaptive instance normalization to avoid droplet artifacts. Adaptive instance normalization [47] is a normalization layer that helps in achieving faster and more realistic neural style transfer.

We can find studies where researchers have used StyleGAN and StyleGAN2 by adopting transfer learning methodology for generating artificial human facial data and face attributes [43], [44], [45], [46].

In one of the recent study [54] authors have proposed a synthetic child dataset derived from the adult dataset by using latent space editing in StyleGAN3 [24]. However their work is more focused towards human aging application and some data biases in facial recognition systems for different races. Authors have rendered data of 1652 subjects including boys and girls with total of 188,832 unique data samples which will be available soon. Contrary to this we have performed transfer learning on styleGAN2 and further using latent space editing to comprehend various smart facial transformations. StyleGAN3 [24] has been recently published, which is mainly used to improve image rotation and image translation, however, the training process is time-consuming, and its latent space is more entangled than its predecessors [25] therefore in this work, we have mainly focused on using StyleGAN2 for rendering synthesized child data. In our work we have not focused on data biases for specific applications however our work is more focused towards advanced child data augmentations thus we have included six different smart transformations which includes, child aging effect, four different facial expressions, directional lightning conditions, eye blinking effect, hair and skin color digitization for incorporating ethnic diversity and lastly head pose variation. All these transformation can be explicitly used for various computer vision applications such as enhanced child biometrics, child expression classification systems, child gender classification systems, etc. More importantly, we have open-sourced our dataset comprising of more than 300k unique facial data samples of 1024×1024 resolution with 20k subjects (10k boys + 10k girls) along with the trained models which is the largest synthetic dataset to the best of our knowledge. These tuned models can be further used for generating more amount of data and using existing data for diversified real-world applications. Similarly in one of the another study proposed by [53] authors have proposed a large scale Indian child dataset as a benchmark private dataset for building face aging progression and regression model using GAN architectures however they have also not released the large scale Indian child dataset.

Large-scale publicly available facial datasets are predominantly sourced from adult subjects. Further, the acquisition of comprehensive, large-scale, and diverse datasets for child subjects presents a notable challenge due to General Data Policy Regulations, especially in European Regions. Nevertheless, it is worth noting that the number of available large-scale child datasets (as discussed in Table 1) and the volume of information they offer are substantially limited. Also, they may suffer from data imbalance, with

limited representation of diverse age groups, diverse facial expressions, diversified facial poses and other demographic factors in comparison to the extensive face recognition (FR) datasets available for adult subjects thus making our attention towards generating a comprehensive synthetic child dataset without the requirement of human laborious efforts and preventing the identity of real human subjects. To achieve this task, we have primarily focused on performing smooth adaptation in StyleGAN2 thus building our customized GAN architecture for rendering new child data and which should be capable of generating advanced transformations on novel child data using latent space editing feature. Smooth adaptation in GAN Networks remains a challenging task by going through the previously published studies. Domain adaptation challenges in GAN (Generative Adversarial Network) networks primarily revolve around the difficulties in adapting or transferring knowledge learned from one domain (source domain) to another domain (target domain). These challenges can affect the performance and generalization of GANs when dealing with real-world data. Keeping this in view we have addressed this by selecting the optimal set of network hyperparameters and adapted robust training methodology to build ChildGAN which resulted in robust outcomes in the form high quality child facial data. These models can generate realistic child facial images while respecting privacy and data protection regulations, thus enabling research and development in applications involving child subjects without compromising their privacy and safety.

III. BUILDING CHILDGAN USING STYLEGAN

In this section, the proposed algorithm is introduced. Our goal is to train a network that can generate photo-realistic boys and girls facial samples separately with controlled attributes. The pipeline of this framework consists of three components as shown in Figure 2. This includes, collecting synthetic data for seed training datasets. Secondly, training boys and girls models separately via the ChildGAN generator. Finally, generating faces with different attributes such as expressions, lighting, and ages.

A. DATASET COLLECTION AND CLEANING

In this paper, the seed training data is all synthetic faces, which are collected from multiple sources. This includes two Android applications that are baby generator [27], and fake face generator [28]. In addition to that we have included data from other GAN-based generators [26] available via Github repository to further incorporate training data diversity. Figure 3 shows the data sources for training two different ChildGAN models, i.e. model A for generating boys data and model B for generating girls data. It is important to ensure that all the data samples contain clear and distinguishable facial region of interest. Therefore, the acquired data undergoes certain filtering processes essential for the optimal training of GANs. This incorporates shortlisting the samples which are of high quality, selecting distinct samples, and having no artifacts. The overall training data is divided into binary

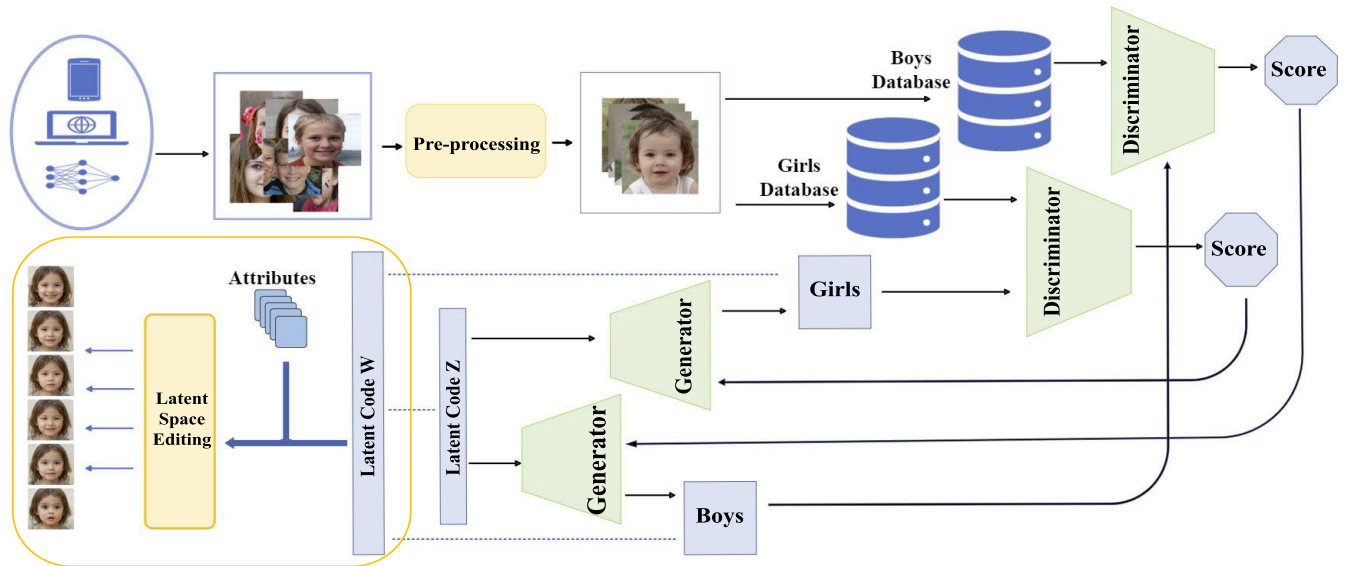


FIGURE 2. Workflow of the proposed ChildGAN architecture.

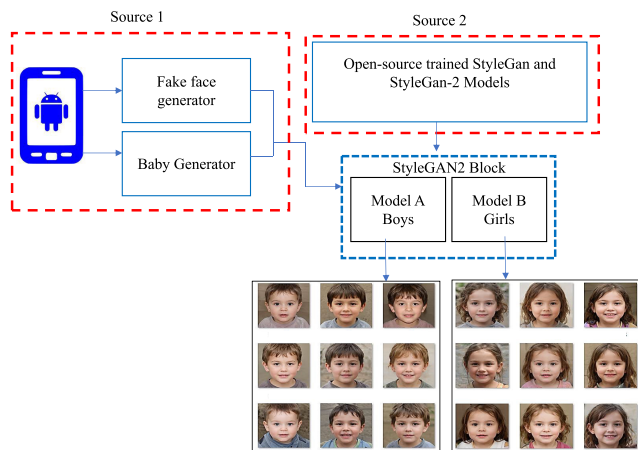


FIGURE 3. Training data sources [26], [27], [28] for fine-tuning StyleGAN2 for generating data with various attributes/transformations.

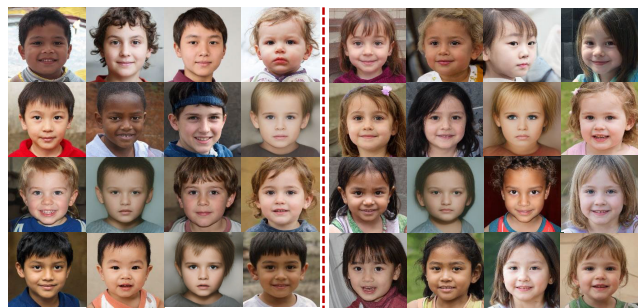


FIGURE 4. Samples of acquired training data, the left-hand side shows the boys' data, and the right-hand shows the girls' data.

classes i.e., boys and girls with equal distribution to avoid imbalanced dataset issues. In addition, the acquired data is cropped and resized to 1024×1024 , which is the initial size selected for training purposes. Some samples from the training dataset are shown in Figure 4. As discussed, we have

TABLE 2. Training data sources.

Data	Number of Images	Image Resolution
Data Obtained from Android Applications	1400	1024x1024
Data obtained from seeprettyface.com GitHub repository	1800	1024x1024

included seed training dataset from two different sources. The first training dataset was acquired from Android fake face generation and baby generator applications. Using these applications, we have generated 772 boys' face data and 628 girls face data. The second synthetic data was acquired from seeprettyface.com [26] where we have shortlisted 890 boys' facial data samples and 910 girls' samples. Thus, in total, we have used 3200 manually curated synthetic seed samples during the training phase. The number of training data samples acquired from multiple sources for tuning StyleGAN2 are summarized in Table 2.

B. LAYER ARCHITECTURE AND TRAINING METHODOLOGY OF CHILDGAN

Motivated by the excellent domain transfer capability, and further disentangled and hierarchical noise introduction in StyleGAN2 architecture we used this network to build ChildGAN by adapting transfer learning methodology. This eventually allows us to smoothly learn new feature vectors for generating high-quality artificial child face samples and further rendering the diversified attributes of the generated children by controlling the latent code. The generation process works by first producing latent code z . Subsequently, z is mapped to the intermediate latent space \mathcal{W} to obtain the transformed w for disentanglement. Then, w is converted to style space through affine transformation, which is used to generate synthetic child data. Meanwhile, noise is added

to control its details. The discriminator structure employs the ProGAN network [41]. It is an extension of the training process of GAN that helps the generator models to train with stability which can produce high-quality images.

While the primary architecture of StyleGAN2 remains intact the architectural layer details of ChildGAN consist of three core blocks which include Mapping Network, Generator Network, and Discriminator Network. The following is a high-level overview of the layer architecture details of each of these three blocks used in ChildGAN architecture built upon StyleGAN2.

1. Mapping Network (Style Mapping): It starts with a mapping network that takes a random noise vector (typically, a latent vector) as input. This mapping network transforms the input noise into a feature vector, which is then used to control the styles of different layers in the generator network. The output of the mapping network is a set of style vectors, each of which corresponds to a specific layer in the generator.
2. Synthesis Network (Generator): The generator network consists of multiple layers, where each layer generates features at a different resolution. It utilizes a skip connection architecture, meaning that lower-resolution layers are concatenated with higher-resolution layers to preserve details. Each layer in the generator network is composed of a series of blocks, including convolutional layers, instance normalization layers, and Leaky ReLU activation layers. The style vectors obtained from the mapping network are used to modulate the activations of each layer, allowing for precise control over the appearance of the generated image. Further we have adjusted the noise injection strategy by changing the noise intensity at certain layers to control the level of stochasticity in generated images.
3. Discriminator: The discriminator network evaluates the quality of generated images. The discriminator is typically a convolutional neural network (CNN) that outputs a probability score indicating whether an input image is real or fake. It also employs a multi-resolution design to assess image quality at multiple scales. In our case, we have modified the last layers of the discriminator network to match the number of conditions i.e., two (boys and girls) in our case.

For tuning ChildGAN networks we have used the pre-trained models trained on the 1024×1024 FFHQ dataset [3] and made available by NVIDIA in their public Github repositories [29]. The training process was distributed on multiple GPU with an increased training batch size of 128 that lead to better convergence and optimal validation results. Moreover, we have incorporated gradient penalty regularization technique to stabilize training and improve the quality of generated images. The overall training time during each experiment took approximately 3 days in adversarial setting for generating the outputs with an image size of 1024×1024 . Further, we have adjusted certain network hyperparameters for performing optimal training which are

TABLE 3. StyleGAN2 hyper-parameter selection.

S.no	Parameter	Value
1	Generator Learning Rate	0.002
2	Discriminator Learning Rate	0.002
3	Batch Size	128
4	Generator Optimizer	Adam
5	Discriminator Optimizer	Adam
6	Generator Loss Function	Non-saturating logistic loss with path length regularizer
7	Discriminator Loss Function	R1 regularization
8	Metric	FID
9	Mirror augment	True
10	Training epochs	500k
11	Trunc psi	0.6
12	Num gpus	2
13	Resume pkl	stylegan2-ffhq-config-f.pkl
14	Lazy regularization	True

discussed in Table 3. Overall, ChildGAN layer architecture is designed to provide fine-grained control over the style and appearance of generated images, resulting in impressive facial image synthesis capabilities, thus generating photorealistic child facial data.

C. LATENT SPACE EDITING

1) ACQUIRING LATENT CODE

In order to generate smart transformations, the latent codes of the input faces should be acquired before editing the faces. This can be achieved by using two different approaches. The first approach is using the pretrained model to generate optimal result/face and preserve the original latent code $z \in \mathcal{Z}$. Then convert z to easy-to-manipulate intermediate latent code w by learning a mapping $f : z \rightarrow w$, where $w \in \mathcal{W}$ denotes the (18, 512) dimensional latent vectors. The second way is choosing an appropriate GAN inversion method to encode the existing children's faces. It shows that distortion and editability are a trade-off [30] in designing an encoder for StyleGAN image manipulation. In this work e4e encoder [30] has been selected to get better editability results. The fitted latent vector $w^* = e4e(x)$ is close to the original w after optimization. Algorithm 1 shows the pseudo-code for generating various smart image transformations on child data using latent code.

2) FACIAL ATTRIBUTES EDIT

The goal of face attribute editing is to allow the modification of one attribute while maintaining other attributes in a similar face. Due to the entanglement of various attributes, it is still challenging to accomplish this task. Studies [40] indicate that style latent code s and intermediate latent code w could perform better attribute editing. In this work, we have focused to learn more diversified and distinct semantical attribute directions such as age and expression corresponding to the latent space \mathcal{W} . For this goal, multiple faces have been acquired from our trained model, and then we classify these faces for the target attributes. Microsoft Cognitive Services API [31] has been used for face recognition to obtain the attribute labels of faces. Then logistic regression classifier

Algorithm 1 Latent Code Used in ChildGAN

Input: Facial image I ; Attribute t ;
Output: Updated facial image I_{update}

- 1: Initial *coeffs* = $[\lambda_1, \lambda_2, \dots, \lambda_n]$;
- 2: $z = search_latent_z(I)$;
- 3: **if** $z == NULL$ **then**
- 4: $w = e4e(I)$;
- 5: **else**
- 6: $w = mapping_network(z)$;
- 7: **end if**
- 8: $\alpha_t = search_attribute_direction(t)$
- 9: **for** $i = \lambda_1$ to λ_n **do**
- 10: $w_{update} \leftarrow w + i \times \alpha_t$;
- 11: $I_{update} \leftarrow Generator(w_{update}, \theta)$;
- 12: **end for**

has been used for training purpose to get the semantical attribute directions $\alpha_1, \alpha_2, \dots, \alpha_n$. Thus, the new face with the corresponding attribute is elaborated in equation 1.

$$w_{update} = w + \alpha_t \cdot coeff \quad (1)$$

where w_{update} is the updated latent codes, α_t is the semantical direction, and *coeff* is the adjustable vector to control the new attribute. Finally, the updated face image I_{update} is generated as shown in equation 2.

$$I_{update} = Generator(w_{update}, \theta) \quad (2)$$

where θ are the parameters of the generator.

3) RELIGHTING PORTRAIT

For embedding efficient lighting effects on synthetic child facial data, we have explored several SOTA algorithms and shortlisted Deep Single-Image Portrait Relighting (DPR) technique [32] for this work. The adapted technique has shown robust qualitative and quantitative results on various datasets [32]. Moreover, this method is effective as it particularly avoids artifacts due to the physically based relighting method as compared to other re-lighting techniques [48], [49]. This method adopts the spherical harmonics lighting model which allows more precise control of directional lighting as compared to latent editing for lighting control. The pretrained model originally trained on the CelebA-HQ dataset [41] has been used for producing more than 60 distinctive face lighting conditions thus illuminating different facial regions. Moreover for this research work we have further shortlisted four most distinctive lighting conditions covering four directional facial regions i.e up, down, left, and right shown in Figure 5.

IV. EXPERIMENTAL RESULTS

This section will present the tuned StyleGAN2 outputs along with various smart transformations. In this study since we have focused on using transfer learning methodology in StyleGAN2 rather than training the whole network from

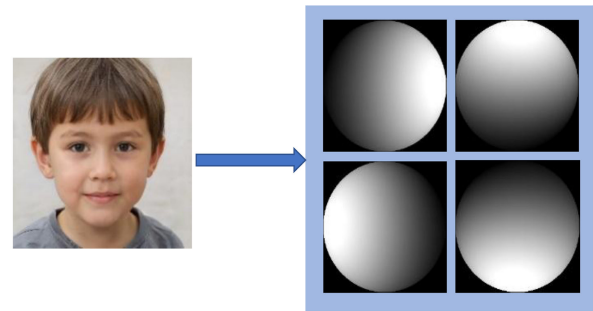


FIGURE 5. Example of lighting conditions.

scratch which requires a significantly large amount of data, thus the training time was much lesser with reduced computational complexity cost. The overall training process was completed in six days and 40 minutes where each network (boy and girls) took around three days with a training batch size of 128 on a server-grade machine equipped with 2 NVIDIA GeForce GTX2080Ti cards for the training and testing phase.

A. TRAINING RESULTS

One of the primary objectives of this research work is to understand whether pretrained StyleGAN2 [2] eases the process of generating illustrations for synthetic child facial data. To comprehend this challenge extensive experimental work was carried out in the initial phases to perform a smooth domain shift toward the target application. To observe this effect, the progression of the generated images was registered during each epoch of the finetuning phase. A smooth domain transfer is characterized by a sequential morphing from the original generated images toward the target domain. It is important to mention that if the pretrained features are not useful, or if the training diverges, the network will not be able to learn new patterns from the target data. Since training StyleGAN2 from scratch using unfreeze layer configuration is a computationally expensive process and requires an extensive amount of data and time therefore in this work we have focused on using previous knowledge acquired from the pre-trained models as base-line to shift towards a new target domain by adapting transfer learning methodology. This will eventually reduce the requirement for custom large datasets to achieve high-quality image synthesis. We have used Pytorch deep learning framework for the experimental setup.

Figure 6 shows the generator (G) and discriminator (D) losses of tuned networks using an optimal set of network hyperparameters as shown in Table 3 thus showing ChildGAN networks making progress towards optimum model convergence. As can be observed from Figure 6 the Generator losses during the training process are getting significantly lesser which is essential for the stable convergence of both of the tuned models.

B. MODEL EVALUATION USING QUALITY METRICS

After the training process using transfer learning methodology, the next stage includes the validation performance

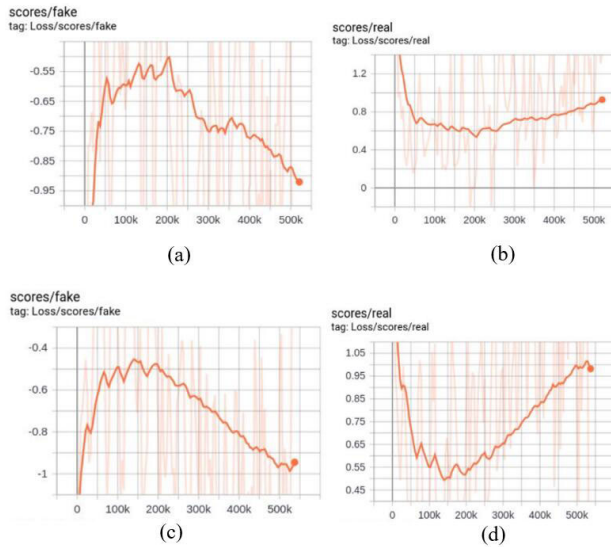


FIGURE 6. ChildGAN training results in the form of G/ D losses, a) shows G loss results for boys tuned model, b) shows D loss for boys tuned model, c) shows G loss results for girls tuned model, d) shows D loss results for boys tuned model.

of optimally tuned models. Testing data in a Generative Adversarial Network (GAN) primarily involves evaluating the quality of the generated samples via Trained GAN models. In this study, we have used FID50K metrics to test the quality of generated images. For this purpose, we have used 50k images to compute the FID scores for boys and girls classes. Figure 7 illustrates the tuned model evaluation results by incorporating Fréchet Inception Distance score (FID) metrics [33]. This metric is generally used to estimate the distance between feature vectors calculated for real and generated images for trained GAN architectures as shown in equation 3.

$$FID(x, g) = ||\mu_x - \mu_g|| + Tr(\Sigma_x - \Sigma_g - 2\sqrt{\Sigma_x \Sigma_g}) \quad (3)$$

where x and g are the real and generated embeddings. μ_x and μ_g are the magnitudes of x and g . Tr is the trace of the matrix and Σ_x and Σ_g are the covariance matrix. As can be observed from Figure 7 the FID curve for both the tuned models is getting lower which means the generated data has better image quality and diversity. The inference operation was also performed on the same hardware machine where it took around 0.8 second to generate a single image in 1024×1024 resolution. This inference time can be further decreased by generating images in smaller resolution (256×256 image resolution) however in this study we have rendered high resolution data that can be further used for downstream machine learning tasks.

C. CHILD TRANSFORMATION RESULTS USING LATENT CODE AND RELIGHTING ALGORITHM

This section will discuss various child transformation outputs as explained in section III-C. In this work, we have modified the synthetic child facial outputs by adapting 5 distinct smart

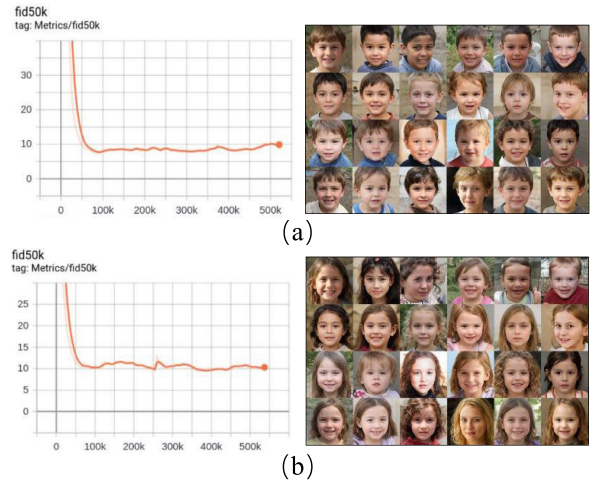


FIGURE 7. FID metric results with images generated at the final epoch of ChildGAN architecture, a) shows the FID results for the boys model, b) shows the FID results for the girls model.

transformations. These transformations include rendering faces with natural facial expressions, eye blinking effect, hair and skin color digitization, child aging, and head pose variation. In addition to that we have incorporated different lighting conditions using DPR method based on deep learning to illuminate facial regions to further bring in data diversity.

1) FACIAL EXPRESSIONS

A human face has significant and distinctive characteristics that help in the recognition of facial expressions caused by an individual’s internal emotional state. Facial expression analysis is used in a wide range of human-computer interaction (HCI) applications, such as face image processing, and facial animation [42]. In this work, four different facial expressions are generated using w latent code and stored as shown in Figure 2. These expressions include neutral to happy, neutral to angry, neutral to surprise, and neutral to sad as depicted in Figure 8.

2) EYE BLINKING

The eye blinking effect can be applied in certain child facial analysis tasks, including anti-spoofing and drowsiness detection and may help detect some health disorders. In this work, we have applied the eye-blinking effect in different lighting conditions. This is achieved through latent code by adjusting the attribute directions and loading the tuned model architecture. This allows to develop blink detection algorithms that are more robust to environments with uncontrolled lighting conditions. Figure 9 shows the eye blinking results of two different subjects.

3) SKIN TONE AND HAIR COLOR DIGITIZATION

The next smart transformation includes further rendering the generated child data with distinct skin tones and hair colors. This type of transformation is useful for developing

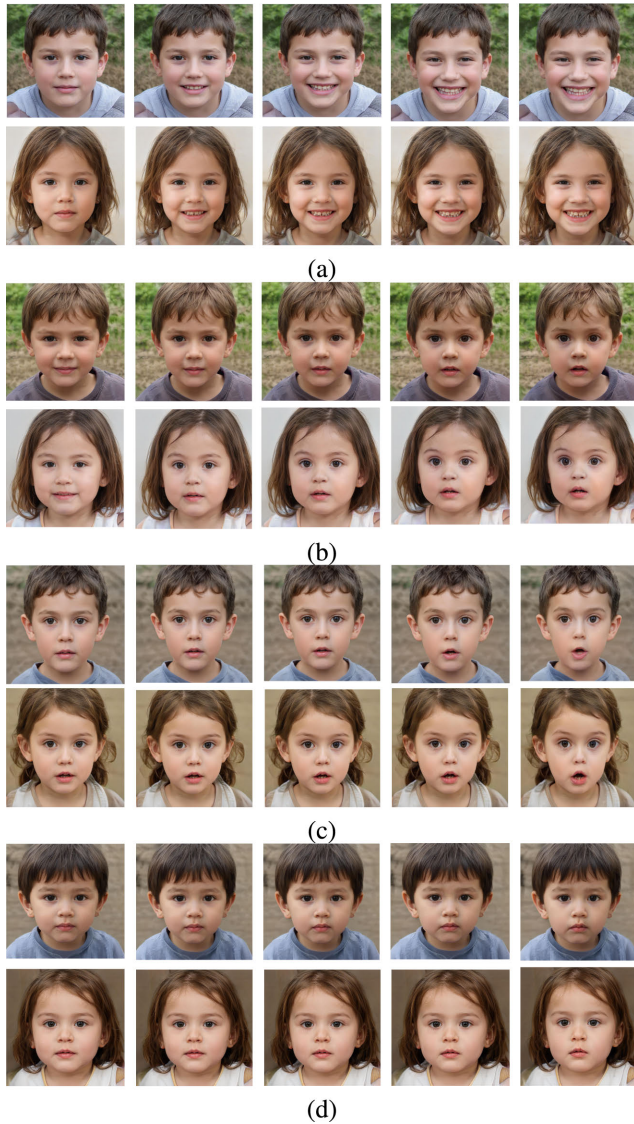


FIGURE 8. Child facial expression results on four different subjects acquired from synthetic boys and girls data, a) shows neutral to happy expressions, b) shows neutral to angry expressions, c) shows neutral to surprise expressions, and d) shows neutral to sad expressions.

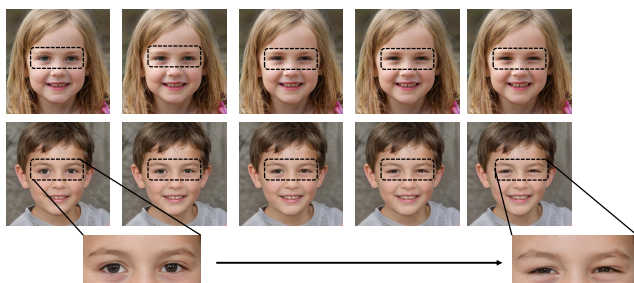


FIGURE 9. Synthetic child facial eye blinking results on two different subjects.

automated systems for capturing the appearance of a physical hair sample, skin tone-based dress, and makeup selection for children and for augmented/virtual reality applications. This



FIGURE 10. Synthetic child facial skin tone and hair color transformation on three different subjects.

is accomplished by mixing different latent codes to get unique skin color adaptations and hair colors. Figure 10 shows the generated effects on girls’ data.

4) CHILD AGING AUGMENTATION

Age progression is a challenging factor for face recognition algorithms deployed in smart mobile and embedded platforms. The complication is primarily based on by biological changes that occur during aging and can result in visible changes in facial characteristics when analyzing the images of the same person taken at different ages. There is an emerging need to extract robust facial features for age-invariant face recognition systems because the face is part of the human body that is greatly affected by aging. This section will detail the experimental outcomes of child aging transformation results using latent code. The results were generated on more than 500 different boys and girls subjects. Figure 11 shows the child aging transformation results on two separate subjects.

5) FACE HEAD POSE VARIATION

In addition to facial expression, eye blinking, skin and hair color variation, and aging progression transformation we have also included face head pose variation using latent code. Generating children’s head pose data is beneficial for various computer vision applications such as aiding in gaze estimation, modeling attention, fitting 3D models to video, and performing face alignment. Figure 12 shows the child’s facial head pose variation results.

6) RELIGHTING FACE

Human faces can have a lot of variations in appearance due to distinct lighting conditions which eventually becomes a challenge for face recognition algorithms. In this work, we have addressed this barrier by generating more than thirty different illumination conditions using Deep Single-Image Portrait Relighting (DPR) technique as shown in Figure 13. Moreover, for this work, we have further shortlisted the four most distinctive lighting conditions covering four directional facial regions i.e up, down, left, and right. Figure 14 depicts the selected set of lighting conditions along with facial illumination results covering different facial angles.



FIGURE 11. Synthetic child aging progression results, the first row shows the aging progression results of a boy from child to young & the second row shows the aging progress from young to the child for a girl.



FIGURE 12. Synthetic child head pose variation results, the first row shows the yaw variation, and the second row shows the pitch variation.



FIGURE 13. Forty distinct lighting conditions generated on synthetic boy image.

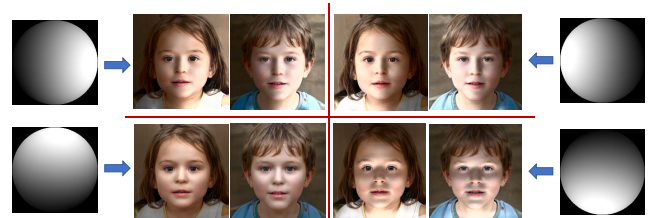


FIGURE 14. Synthetic child face data with four distinct lighting conditions.

complete dataset is open-sourced via github page ¹ and can be used for commercial and non-commercial applications.

V. CHILDGAN DATA VALIDATION

With the rapid advancement of machine learning algorithms and data required for real-world computer vision applications, the lack of high-quality data is the real bottleneck in the AI industry. This section will focus on different computer vision and machine learning techniques to check the quality of generated child synthetic data. The details of each of these validation methods are provided in the below subsections.

A. GENDER CLASSIFICATION TEST

In this test, we have trained a binary classifier (boys and girls) using end-to-end pre-trained CNN which includes

¹<https://github.com/MAli-Farooq/ChildGAN>

However, in this research work, we have shortlisted 4 illumination conditions covering four directional facial regions i.e up, down, left, and right as shown in Figure 5.

Table 4 shows the overall generated child data along with its attributes using various smart transformations as discussed in section III and section IV. Thus we have generated a total data of around three hundred and twenty four thousand distinct samples of boys and girls using ChildGAN. The

TABLE 4. Overall child synthetic dataset.

Category	No. of Sam.	No. of Sam./Subj.	Total
Child Data	Boys: 10k Girls: 10k	1	20k
Facial Expressions	Boys: 2k Girls: 2k	Happy: 6 Sad: 6 Angry: 6 Surprise: 6	96k
Eye blinking effect	Boys: 2k Girls: 2k	6	24k
Skin Tone and Hair Color Variation	Boys: 1k Girls: 1k	6	12k
Aging Progression	Boys: 2k Girls: 2k	8	32k
Headpose Variation	Boys: 2k Girls: 2k	Yaw: 8 Pitch: 7	60k
Relighting	Boys: 10k Girls: 10k	4	80k

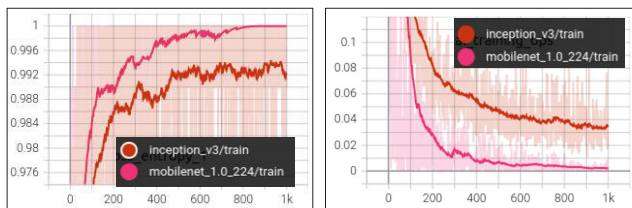


FIGURE 15. Binary gender classifier training results on child synthetic data, a) shows the accuracy graph of mobileNet-v2 and Inception-v3 CNN with an overall accuracy of 99.87% and 99.15%, b) shows the loss graph of mobileNet-v2 and Inception-v3 CNN with a loss rate of 0.015 and 0.061.

TABLE 5. Experiment setting.

Experiment No	Details
Experiment 1	training and testing with real child data
Experiment 2	training with synthetic + real and testing with real data

InceptionV3 [34] and MobileNetV2 [35] on synthetic data. The performance of the trained classifier on synthetic data is then cross-validated on real child data which is acquired from UTKFace dataset [36]. Figure 15 shows the training and validation results in the form of accuracy and loss graphs. The model was trained for 1000 epochs by selecting binary cross entropy as the loss function and SGD optimizer. The tuned gender classifier models on synthetic data were then validated on a set of 500 distinct test samples. MobileNetV2 and InceptionV3 models achieve overall test accuracy of 94% and 92% respectively thus showing the precise robustness of the trained model on synthetic data.

In addition, we performed extensive gender classifier tests as further downstream tasks as shown in Table 5 for experimental validation and highlighting the usefulness of synthetic child data. The result proves as shown in Figure 16 that exp 2 leverages the performance of CNN classifiers with

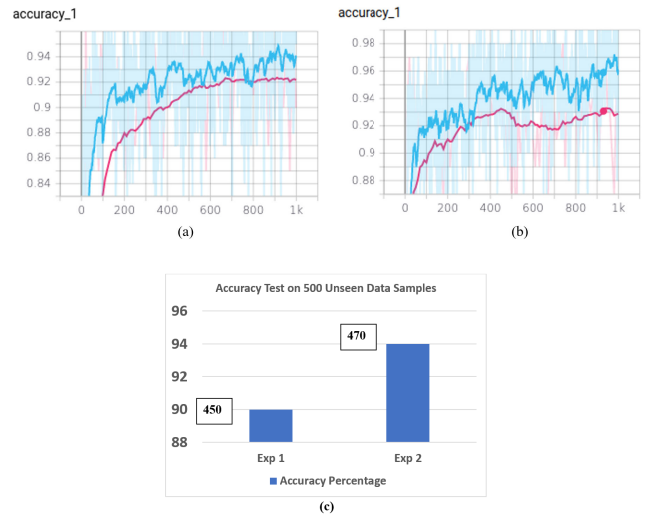


FIGURE 16. (a) Exp1 results with train and validation accuracy of 93.78% and 92.16%, (b) Exp2 results with train and validation accuracy of 96.04% and 92.76%, (c) Test accuracy chart.

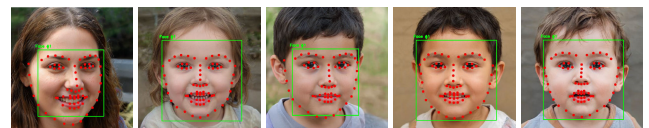


FIGURE 17. Face mapping and 68 facial landmark detection test results on five different subjects (boys & girls) with varying face head pose, aging difference, skin tone variation, and different facial expressions.

nearly an improvement of 2.26% in training phase as shown in Figure 16 a and b and 4.25% during the testing phase on unseen data as shown in Figure 16 c thus highlighting the usefulness of synthetic data.

B. FACE LOCALIZATION AND FACIAL LANDMARK DETECTION TEST

The second critical test includes face localization and facial landmark detection on child synthetic data using DLIB [37] library. The system works by employing a 68-facial-landmark detector to extract multiple facial key points including the child’s eyes, nose, lip, and around the face. This plays a vital role in different computer vision applications such as face recognition, biometrics, expression analysis, etc. The test was performed on 100 distinct samples with varying face poses, different genders, and other smart transformations mentioned in Table 4. Figure 17 shows some of the challenging test samples with robust face mapping and face landmark detection results.

C. SYNTHETIC DATA UNIQUENESS TEST

In this experiment, we adopt ArcFace, a well-known reference face recognition models [38] to test the identity similarity of synthetic child faces. The model is used to test the data using two different methods. In the first test, we perform the similarity measure between different

TABLE 6. IQA comparison with other datasets.

Metrics	MUSIQ [52] ↑	NIQE [51] ↓	BRISQUE [50] ↓
Our Dataset	75.83	4.04	9.04
UTKFace [36]	36.72	7.96	41.25
CelebA-HQ [6]	63.31	5.38	39.39

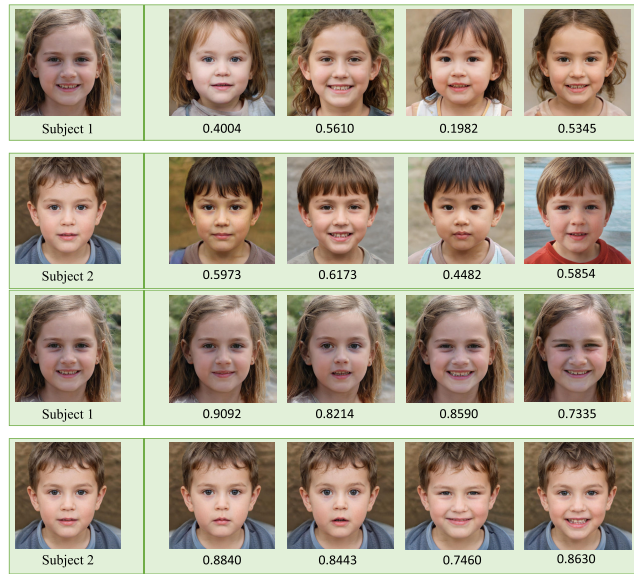


FIGURE 18. The similarity score between different samples. The first and second row calculates the similarity score of subject 1 and subject 2 with other girls and boys data samples. The third and fourth row shows the similarity score of subject 1 and subject 2 with various expressions of subject 1 and subject 2.

subjects. Secondly, we test different facial attributes of the same subject to validate that the identity remains consistent. Figure 18 shows that the cosine similarity of two different subjects is lower than the cosine similarity of other boys and girls samples along with distinct facial expressions of similar subjects. This test demonstrates the facial patterns and features, of synthetic child faces provide useful uniqueness, diversity, and quality.

In addition, we have estimated identity similarity scores on additional boys and girls data using Arcface. This is achieved by calculating the identity similarity scores for 10 distinct boys and girls data as shown in Table 7 and Table 8. It is important to mention that detailed subject similarity analysis was also performed using the Arcface and some data samples were discarded having close similarity identity scores.

D. EYE ASPECT RATIO TEST

The Eye Aspect Ratio also referred to as EAR, is a scalar value that particularly represents the opening and closing ratio of the eyes [39] using facial landmark detectors across the eyes. During the eye flashing process, the EAR value increases or decreases significantly thus displaying the eye blinking effect. Equation 4 and equation 5 show the formula for calculating the average EAR of both eyes along with the

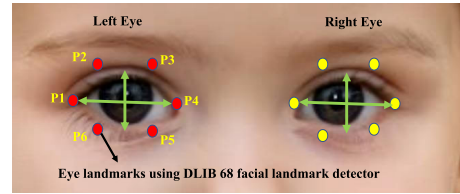


FIGURE 19. Eye landmarks from P1-P6 using DLIB 68 facial landmarks detector on synthetic child image for estimating EAR.

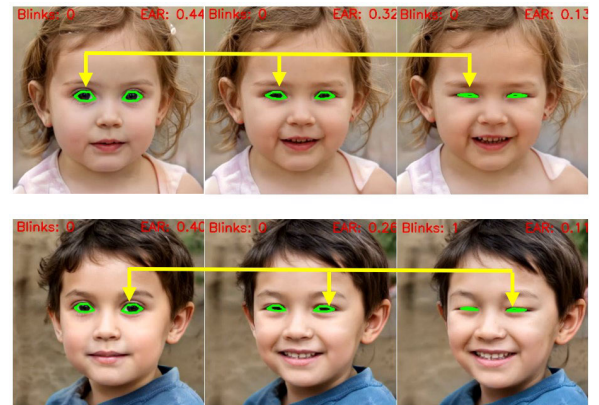


FIGURE 20. Eye blink detection and EAR results on two different subjects generated through ChildGAN. The first row shows the girls' images with the eye blinking effect thus changing EAR from 0.44 to 0.12 whereas the second row shows the boys' images with the eye blinking effect thus changing EAR from 0.41 to 0.11.

formula for estimating EAR for a single eye.

$$AverageEAR = \frac{1}{2}(EAR_{righteye} + EAR_{lefteye}) \quad (4)$$

$$EAR = \frac{||P2 - P6|| + ||P3 - P5||}{2 \times ||P1 - P4||} \quad (5)$$

where P1, P2, P3, P4, P5, and P6 are the facial landmarks across the eyes as shown in Figure 19.

In this work, we have employed the EAR test to validate the eye blinking effect quantitatively on synthetic child facial data. Figure 20 demonstrates eye blinking effect and EAR results by using facial landmarks across the eyes on two different subjects generated through ChildGAN. By observing Figure 20 it can be analyzed that EAR value changes significantly during eye blinking which proves that generated child data has photo-realistic facial features.

E. IMAGE QUALITY EVALUATION

Since there is no ground truth i.e real child data, in this section we have included three no-reference image quality assessment metrics including NIQE [51], BRISQUE [50], and MUSIQ [52] to evaluate the synthetic child data. The average scores were calculated from a random selection of 1000 images from our dataset, UTKFace, and CelebA-HQ as shown in Table 6. The results elaborate that our data have adequate quality when compared with large-scale public facial datasets.

TABLE 7. Identity similarity scores using ArcFace for girls.













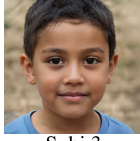

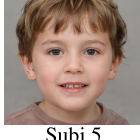
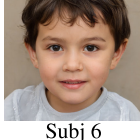
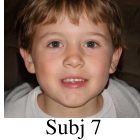

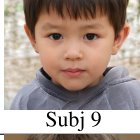

	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5	Subj 6	Subj 7	Subj 8	Subj 9	Subj 10
 Subj 1		0.3059	0.2573	0.3401	0.2802	0.4950	0.3255	0.4233	0.3158	0.4192
 Subj 2	0.3059		0.6590	0.6965	0.5293	0.6223	0.0774	0.5898	0.5245	0.5106
 Subj 3	0.2573	0.6590		0.6188	0.5791	0.7475	0.2971	0.6576	0.7324	0.5034
 Subj 4	0.3401	0.6965	0.6188		0.6471	0.7113	0.2343	0.5758	0.5393	0.6640
 Subj 5	0.2802	0.5293	0.5791	0.6471		0.5361	0.3378	0.5510	0.5648	0.5714
 Subj 6	0.4951	0.6223	0.7475	0.7113	0.5361		0.3618	0.7413	0.5970	0.6636
 Subj 7	0.3255	0.0774	0.2971	0.2342	0.3378	0.3618		0.4421	0.4140	0.2997
 Subj 8	0.4233	0.5898	0.6575	0.5758	0.5509	0.7413	0.4421		0.5801	0.4302
 Subj 9	0.3158	0.5245	0.7324	0.5393	0.5648	0.5970	0.4140	0.5801		0.4601
 Subj 10	0.4192	0.5106	0.5034	0.6640	0.5714	0.6636	0.2998	0.4302	0.4607	

TABLE 8. Identity similarity scores using ArcFace for boys.

	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5	Subj 6	Subj 7	Subj 8	Subj 9	Subj 10
 Subj 1		0.7034	0.3473	0.6283	0.6092	0.6693	0.3613	0.6690	0.5897	0.6972
 Subj 2	0.7034		0.3616	0.7845	0.5406	0.4923	0.3594	0.6650	0.3587	0.6283
 Subj 3	0.3473	0.3616		0.3126	0.3254	0.5024	0.1651	0.4799	0.2848	0.4276
 Subj 4	0.6283	0.7845	0.3126		0.6222	0.5543	0.3886	0.6542	0.3750	0.6841
 Subj 5	0.6092	0.5406	0.3254	0.6222		0.6362	0.4768	0.4541	0.3038	0.5852
 Subj 6	0.6693	0.4923	0.5024	0.5543	0.6362		0.3315	0.6617	0.4835	0.7464
 Subj 7	0.3613	0.3594	0.1651	0.3886	0.4768	0.3315		0.2844	0.1764	0.3510
 Subj 8	0.6690	0.6650	0.4799	0.6542	0.4541	0.6617	0.2844		0.5137	0.7423
 Subj 9	0.5897	0.3587	0.2818	0.3750	0.3038	0.4835	0.1764	0.5137		0.5965
 Subj 10	0.6972	0.6283	0.4276	0.6841	0.5852	0.7464	0.3510	0.7423	0.5965	

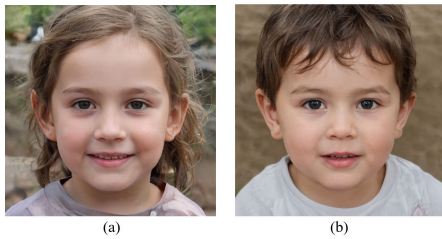


FIGURE 21. (a) Girl sample NIQE: 3.54, MUSIQ: 77.16, BRISQUE: 11.83 (b) Boy sample NIQE: 3.79, MUSIQ: 76.99, BRISQUE: 14.96.

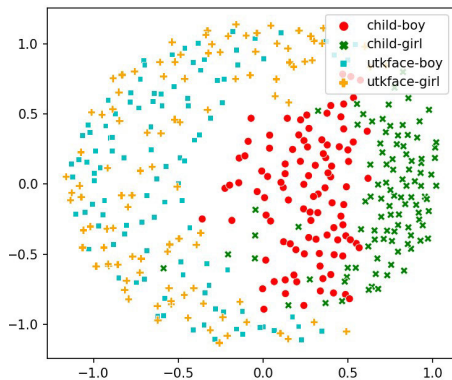


FIGURE 22. Visualization of the real and synthetic data feature distributions (using MDS [56]) for the samples shortlisted from ChildGAN generated boys data, ChildGAN generated girls data, UTKFace boys and UTKFace girls data, which are illustrated by red circles, green cross, blue square and orange cross respectively.

F. REAL VS. SYNTHETIC DATA MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) [56] is a popular approach for graphically representing relationships between samples in multidimensional space thus elaborating set of synthetic data (taken from ChildGAN) and real-world samples (taken from UTK dataset) and calculating a dissimilarity (distance) measure for each pairwise comparison of samples. For this purpose, we have done the data distribution by randomly selecting 100 boys samples and 100 girls samples from ChildGAN dataset (as the synthetic data) and UTKface dataset (as the real data) thus plotting MDS as shown in Figure 22.

G. SUBJECTIVE EVALUATION

Regarding quality of the rendered synthetic data, we conducted an internal study with 25 participants from our research group to see if they could distinguish between real and synthetic children. The real child data samples was shortlisted from databrary CAFE dataset [10]. The results show that 19 people were unable to distinguish difference between real and synthetic child data which concludes that 76% of participants are convinced with the quality of photo-realistic synthetic child data.

VI. CONCLUSION AND FUTURE WORK

In this work, we have proposed ChildGAN architecture to render synthetic child facial data samples at scale. In addition various smart transformations enable the base data to be

modified to increase the variety of data samples for each individual child. The tuned model show that transfer learning in GAN is indeed an effective approach for producing high-quality images even when the source and target domains are different from each other. The associated fine-tuned models and rendered synthetic data are open-sourced via GitHub and methodological details are provided to allow other researchers to build on this work. We expect that this will inspire new innovations to diversify and quantify the methodology described in our paper. For now, we have tested the potential of this data by running four different computer vision application tests to validate the qualitative behaviour of this synthetic data. However it remains a challenging task to validate synthetic data distributions as a ground truth (real data distribution) is not readily available to make a direct comparative analysis. The artificial data may not fully capture the complexity and variability present in real-world data. This can limit its effectiveness in situations where the data distribution is highly diverse and dynamic. Moreover, the tuned Models trained on synthetic data may not generalize well to complex real-world computer vision applications since real-world data can contain unexpected patterns, outliers, and complexities that synthetic data may not account for. Lastly in this work we have incorporated six different smart transformations to produce enough data diversity however there are still some limitations such that it is still a challenging task to produce ethnically diversified data samples with existing tuned GANs since these models generate the data based on certain seed data (training data samples).

Several important research tasks remain as future challenges. Firstly, to explore robust qualitative and quantitative metrics to validate the distribution of synthetic identities as we do not have a corresponding set of real data [4]. The second challenge is to improve the capability of the synthetic data samples to support broader diversity in facial analysis algorithms. This could involve adopting a similar methodology to build ethnicity into the style-transfer and explore if we can maintain a similar identity distribution for children of a specific ethnicity. We are further working to explore more recent AI based generative algorithms thus to incorporate more advanced style conditioning methods such as text-to-image and image-to-image translation. Keeping this in view we have been working on using diffusion models to perform more advanced data augmentation operations such that generating child data samples with ethnic diversity and races which is part of our future work as the expansion of ChildGAN work. These models are able to address some notorious limitations of GANs, like mode collapse, overhead of adversarial learning, and convergence failure [55]. Finally, the potential of this high-quality dataset can be explored for a variety of 3D computer applications through the reconstruction of 3D child facial models [5]. It will be interesting to see how the research community builds on this first large-scale dataset of synthetic child facial data.

REFERENCES

- [1] N. R. Giuliani, J. C. Flournoy, E. J. Ivie, A. Von Hippel, and J. H. Pfeifer, "Presentation and validation of the DuckEES child and adolescent dynamic facial expressions stimulus set," *Int. J. Methods Psychiatric Res.*, vol. 26, no. 1, p. e1553, Mar. 2017.
- [2] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8107–8116.
- [3] 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.
- [4] V. Varkarakis, S. Bazrafkan, G. Costache, and P. Corcoran, "Validating seed data samples for synthetic identities—Methodology and uniqueness metrics," *IEEE Access*, vol. 8, pp. 152532–152550, 2020.
- [5] S. Basak, P. Corcoran, R. McDonnell, and M. Schukat, "3D face-model reconstruction from a single image: A feature aggregation approach using hierarchical transformer with weak supervision," *Neural Netw.*, vol. 156, pp. 108–122, Dec. 2022.
- [6] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3730–3738.
- [7] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.
- [8] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, and J. Shen, "DigiFace-1M: 1 million digital face images for face recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 3515–3524.
- [9] J. G. Negrão, A. A. C. Osorio, R. F. Siciliano, V. R. G. Lederman, E. H. Kozasa, M. E. F. D'Antino, A. Tamborim, V. Santos, D. L. B. de Leucas, P. S. Camargo, D. C. Mograbi, T. P. Mecca, and J. S. Schwartzman, "The child emotion facial expression set: A database for emotion recognition in children," *Frontiers Psychol.*, vol. 12, pp. 1–9, Apr. 2021. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.666245>
- [10] V. LoBue and C. Thrasher, "The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults," *Frontiers Psychol.*, vol. 5, pp. 1–8, Jan. 2015. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01532>
- [11] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing deep learning and SMOTE for imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, Sep. 2023.
- [12] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalance learning," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 1322–1328.
- [13] Z. Wan, Y. Zhang, and H. He, "Variational autoencoder based synthetic data generation for imbalanced learning," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–7.
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [15] P. Tiwald, A. Ebert, and D. T. Soukup, "Representative & fair synthetic data," 2021, *arXiv:2104.03007*.
- [16] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.
- [17] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of Wasserstein GANs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 5769–5779.
- [18] Q. Jin, R. Lin, and F. Yang, "E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3289–3300, Sep. 2020.
- [19] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On convergence and stability of GANs," 2017, *arXiv:1705.07215*.
- [20] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, "Many paths to equilibrium: GANs do not need to decrease a divergence at every step," 2017, *arXiv:1710.08446*.
- [21] M. G. Bellemare, I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer, and R. Munos, "The Cramer distance as a solution to biased Wasserstein gradients," 2017, *arXiv:1705.10743*.
- [22] M. Zheng, T. Li, R. Zhu, Y. Tang, M. Tang, L. Lin, and Z. Ma, "Conditional Wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification," *Inf. Sci.*, vol. 512, pp. 1009–1023, Feb. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025519309715>
- [23] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. 2019, pp. 1–11.
- [24] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 852–863.
- [25] Y. Alaluf, O. Patashnik, Z. Wu, A. Zamir, E. Shechtman, D. Lischinski, and D. Cohen-Or, "Third time's the charm? Image and video editing with StyleGAN3," in *Proc. Eur. Conf. Comput. Vis. Workshop*. Cham, Switzerland: Springer, Feb. 2023, pp. 204–220.
- [26] seeprettyface.com. (2019). *Child Seed Data*. Accessed: Feb. 27, 2023. [Online]. Available: <https://github.com/a312863063>
- [27] DBAI. (2022). *Baby Generator*. Accessed: Dec. 22, 2022. [Online]. Available: <https://play.google.com/store/apps/details?id=com.dbai.predictbaby.facesix&hl=en&gl=US>
- [28] (2022). *Fake Face Generator*. Accessed: Dec. 20, 2022. [Online]. Available: <https://apkcombo.com/fake-face-generator/com.fakeface.generator/>
- [29] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. (2020). *Pretrained StyleGAN2 Models Repository*. Accessed: Feb. 27, 2023. [Online]. Available: <https://github.com/NVlabs/stylegan2>
- [30] O. Tov, Y. Alaluf, Y. Nitzan, O. Patashnik, and D. Cohen-Or, "Designing an encoder for StyleGAN image manipulation," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–14, Jul. 2021, doi: [10.1145/3450626.3459838](https://doi.org/10.1145/3450626.3459838).
- [31] Microsoft. *Microsoft Cognitive Service API*. Accessed: Feb. 21, 2023. [Online]. Available: <https://azure.microsoft.com/en-us/products/cognitive-services/#overview>
- [32] H. Zhou, S. Hadap, K. Sunkavalli, and D. Jacobs, "Deep single-image portrait relighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7193–7201.
- [33] 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," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.
- [34] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [36] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4352–4360.
- [37] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231218303229>
- [38] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, Oct. 2022.
- [39] C. Dewi, R.-C. Chen, C.-W. Chang, S.-H. Wu, X. Jiang, and H. Yu, "Eye aspect ratio for real-time drowsiness detection to improve driver safety," *Electronics*, vol. 11, no. 19, p. 3183, Oct. 2022. [Online]. Available: <https://www.mdpi.com/2079-9292/11/19/3183>
- [40] Z. Wu, D. Lischinski, and E. Shechtman, "StyleSpace analysis: Disentangled controls for StyleGAN image generation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12858–12867.
- [41] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26. [Online]. Available: <https://iclr.cc/Conferences/2018>
- [42] A. A. Pise, M. A. Alqahtani, P. Verma, K. Purushothama, D. A. Karras, S. Prathibha, and A. Halifa, "Methods for facial expression recognition with applications in challenging situations," *Comput. Intell. Neurosci.*, vol. 2022, May 2022, Art. no. 9261438, doi: [10.1155/2022/9261438](https://doi.org/10.1155/2022/9261438).

- [43] G. Hermosilla, D. H. Tapia, H. Allende-Cid, G. F. Castro, and E. Vera, "Thermal face generation using StyleGAN," *IEEE Access*, vol. 9, pp. 80511–80523, 2021.
- [44] G. Song, L. Luo, J. Liu, W.-C. Ma, C. Lai, C. Zheng, and T.-J. Cham, "AgileGAN: Stylizing portraits by inversion-consistent transfer learning," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–13, Jul. 2021, doi: [10.1145/3450626.3459771](https://doi.org/10.1145/3450626.3459771).
- [45] F. Yin, Y. Zhang, X. Cun, M. Cao, Y. Fan, X. Wang, Q. Bai, B. Wu, J. Wang, and Y. Yang, "StyleHEAT: One-shot high-resolution editable talking face generation via pre-trained StyleGAN," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, Oct. 2022, pp. 85–101.
- [46] T. Kusunose, X. Kang, K. Kiuchi, R. Nishimura, M. Sasayama, and K. Matsumoto, "Facial expression emotion recognition based on transfer learning and generative model," in *Proc. 8th Int. Conf. Syst. Inform. (ICSAI)*, Dec. 2022, pp. 1–6.
- [47] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1510–1519.
- [48] Z. Wang, X. Yu, M. Lu, Q. Wang, C. Qian, and F. Xu, "Single image portrait relighting via explicit multiple reflectance channel modeling," *ACM Trans. Graph.*, vol. 39, no. 6, pp. 1–13, Nov. 2020, doi: [10.1145/3414685.3417824](https://doi.org/10.1145/3414685.3417824).
- [49] S. Sengupta, D. Lichy, A. Kanazawa, C. D. Castillo, and D. W. Jacobs, "SfSNet: Learning shape, reflectance and illuminance of faces in the wild," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3272–3284, Jun. 2022.
- [50] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 723–727.
- [51] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a," completely blind image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
- [52] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5128–5137.
- [53] P. K. Chandaliya and N. Nain, "ChildGAN: Face aging and rejuvenation to find missing children," *Pattern Recognit.*, vol. 129, Sep. 2022, Art. no. 108761. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320322002424>
- [54] M. Falkenberg, A. B. Ottosen, M. Ibsen, and C. Rathgeb, "Child face recognition at scale: Synthetic data generation and performance benchmark," 2023, *arXiv:2304.11685*.
- [55] A. Ulhaq, N. Akhtar, and G. Pogrebna, "Efficient diffusion models for vision: A survey," 2022, *arXiv:2210.09292*.
- [56] I. Borg and P. J. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Berlin, Germany: Springer, 2005.



WANG YAO (Graduate Student Member, IEEE) received the B.Sc. degree in computer science and technology from Southwest University, China, in 2016, and the M.Sc. degree in control engineering from the University of Chinese Academy of Sciences (UCAS), in 2019. She is currently pursuing the Ph.D. degree with the University of Galway. She is also an Intern with FotoNation/Xperi. Her research interest focus on computer vision.



GABRIEL COSTACHE received the B.Sc. and M.Sc. degrees from the Faculty of Electronics and Telecommunications, Politehnica University of Bucharest, Romania, in 2003 and 2004, respectively, and the Ph.D. degree in image processing from the National University of Ireland Galway (NUIG), in 2006. He has been a part of FotoNation, since 2006. He is currently the Director of biometrics with FotoNation/Xperi, which develops technologies to process 2-D and 3-D imaging data.



MUHAMMAD ALI FAROOQ received the B.E. degree in electronic engineering from IQRA University, in 2012, the M.S. degree in electrical control engineering from the National University of Sciences and Technology (NUST), in 2017, and the Ph.D. degree from the National University of Ireland Galway (NUIG), in 2022. Currently, he is a Postdoctoral Researcher with the University of Galway; and a Machine Learning Research Intern with Xperi Corporation, where his work is focused on building large scale synthetic datasets for real world applications. He has won the prestigious H2020 European Union (EU) Scholarship as part of the Ph.D. research project and worked as a Consortium Partner in EU funded HELIAUS Project. His research interests include machine vision, computer vision, video analytics, machine learning, thermal imaging, and sensor fusion.



PETER CORCORAN (Fellow, IEEE) is the Personal Chair of electronic engineering with the College of Science and Engineering, University of Galway. He was a Co-Founder of several start-up companies, notably FotoNation, currently the Imaging Division of Xperi Corporation. He has over 600 technical publications and patents, over 100 peer-reviewed journal articles, and 120 international conference papers. He is a co-inventor of more than 300 granted U.S. patents. He is an IEEE Fellow, recognized for his contributions to digital camera technologies, notably in-camera red-eye correction and facial detection. He is a member of the IEEE Consumer Electronics Society, for over 25 years. He is the Editor-in-Chief and the Founding Editor of *IEEE Consumer Electronics Magazine*.