SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V11I5P101

Volume 11 Issue 5, 1-11, May 2024
© 2024 Seventh Sense Research Group®

Original Article

Facial Recognition of Sketch Images in Forensic Laboratories Employing Diverse Techniques

Devendra A. Itole[1], M.P. Sardey[2], Milind P. Gajare[3]

[1,2,3]*Department of Electronics and Telecommunication, AISSMS IOIT, Pune, Maharashtra, India.*

[1]*Corresponding Author : devendra.itole@aissmsioit.org*

Received: 01 March 2024 Revised: 02 April 2024 Accepted: 01 May 2024 Published: 29 May 2024

Abstract - *The use of Differential Facial Recognition (DFR) over the past few years emerged as a challenging endeavor within the realms of biometrics and computer vision, grappling persistently with the complexities of illumination and pose variations. This scholarly investigation aims to propose innovative deep-learning architectures tailored to juxtapose non-visible facial depictions against an array of visible facial galleries. The taxonomy of thermal-to-visible recognition delineates into two distinct categories: feature-based methodologies and image synthesis paradigms. Notably, the latter enhances compatibility with existing recognition frameworks in both commercial and governmental sectors, bolstering efficacy in forensic examination. Additionally, the incorporation of soft biometrics, encompassing diverse traits such as age and gender, provides supplementary layers of information, thereby reinforcing the foundation of recognition algorithms. Novel strategies are introduced to navigate the intricate landscape of auxiliary training data in the LUPI scenario, pushing the boundaries of recognition performance. Additionally, a pioneering aggregation framework is conceived to enhance the robustness of landmark detection, while adversarial techniques amplify the efficacy of landmark detection mechanisms. Finally, this study scrutinises the opaque veil enveloping Generative Adversarial Networks (GANs), aiming to address concerns regarding mode collapse and diversity within the GAN framework.*

Keywords - *Heterogeneous Face Recognition (HFR), Deep Learning Architectures, Thermal-to-Visible Recognition, Soft Biometrics, Generative Adversarial Networks (GANs).*

1. Introduction

In the vast and intricate realm of biometrics and computer vision, the pursuit of facial recognition emerges as a significant challenge, navigating through the complexities of human identification amidst a myriad of environmental factors. This pursuit, rooted in the nuances of facial illumination and pose variations, stands as a focal point of scholarly exploration at the convergence of technological innovation and forensic necessity. Scientific curiosity about algorithms for facial recognition has increased recently as a result of their ability to address the various problems that visible face imaging presents.

This surge is driven by the need to transcend traditional recognition paradigms and delve into the uncharted territory of Heterogeneous Face Recognition (HFR). HFR, with its enigmatic domain spanning from infrared to polarimetric and millimetre-wave spectra, presents researchers with the daunting task of reconciling facial images from diverse domains amidst limited training data [1-2]. Infrared imagery, with its dual realms of reflection and emission, provides valuable insights into facial contours. The accessibility of rich facial information in the NIR and bands of shortwave infrared light, comparable to visible images, has spurred advances in NIR-to-visible recognition of faces. However, challenges persist in SWIR-to-visible recognition due to spectral disparities. At the heart of research lies the fusion of auxiliary facial information and Generative Adversarial Networks (GANs) [3].

These networks, heralded for their prowess in synthetic image synthesis, play a pivotal role in distilling visible-like images from non-visible modalities. Bridging the gap between thermal and visible imagery remains a Herculean task, especially in the context of matching thermal facades with their visible counterparts. Forensic applications further underscore the importance of facial sketch recognition, particularly in cases where photographic evidence is lacking. These sketches, born from eyewitness accounts and forensic expertise, serve as crucial tools in identifying suspects, necessitating automated matching methodologies to scour law enforcement databases [4].

The pursuit of thermal-to-visible recognition has gained momentum in recent years, bolstered by technological advancements. However, discerning thermal facades amidst a

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

gallery of vibrant visages poses a significant challenge, given the stark differences in texture and geometric nuances between thermal and visible imagery. Promising paths to improve the accuracy of thermal-to-visible identification are provided by emerging technologies that make use of the polarisation state of thermal emissions.

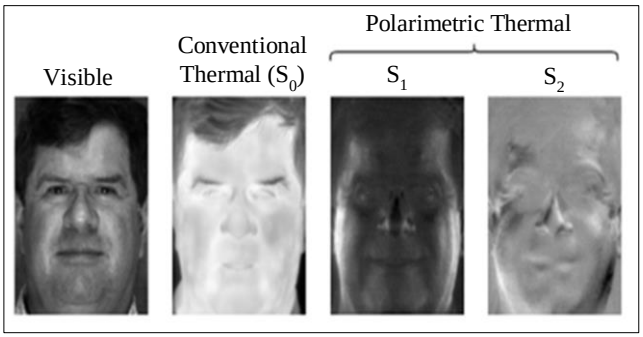


Fig. 1 The conventional thermal and polarimetric states and visible spectrum

The intersection of soft biometrics and hard biometric modalities emerges as a key area of exploration, highlighting the complementary nature of age, gender, ethnicity, and other facial attributes. Soft biometrics, with their cost-effectiveness and ease of acquisition, enrich recognition algorithms by providing auxiliary information to augment primary biometric data.

Research endeavors to unravel the complexities of Heterogeneous Face Recognition, leveraging advancements in technology to bridge the gap between disparate facial domains. Through empirical validation and theoretical exploration, we Aim to chart a path toward more robust and effective recognition frameworks underpinned by the integration of soft biometrics and generative adversarial networks [5].

Investigators studying facial landmark recognition often employ Convolutional Neural Networks (ConvNets) to gather deep characteristics and regression coefficients for an entire training method. This method of updating landmark positions gradually is called a cascade technique. Using deep deformation networks, some researchers like Yu et al. have included geometric restrictions into CNN structures.

Zhu et al. have investigated coarse search strategies to deal with inadequate initialisation, whereas To address initiation issues, A deep regressive design with two-step re-initialisation has been presented by Lev et al. Zhu, together with others, has proposed an alternative method that addresses severe head postures and shape deformations: cascaded regressors [6].

Certain landmark identification techniques concentrate on learning powerful heatmaps for location recognition, utilising complete training with ConvNets. For reliable recognition of facial landmarks, Balut et al. have used a residual framework. For human pose estimation tasks, Newell et al. and Wei et al. take into account the coordinates of the greatest response on heatmaps.

A broader perspective would categorise the issue as acquiring structural models. Numerous studies aim to classify visual data based on various attributes of change, such as identity, mobility, and camera perspective, to represent the intrinsic structure of things. However, the physical features of these components remain hidden in implicit projections., rendering them undetectable. Some approaches treat structures like masks, depth, and landmarks as auxiliary data in a multitasking framework [7].

Generative systems' primary goal is to approach the true distribution of information, which has witnessed notable progress. Moreover, traditional generative models aim to reduce the difference between the information and modeling distribution' Kullback-Leibler (KL) dispersion. They frequently generate undesired and perplexing samples.

Conversely, when minimising the reverse KL divergence using Generative Adversarial Networks (GANs), the emphasis often lies on a single mode of the input, leading to mode collapse-a scenario where the generator produces images that are nearly identical to each other to address this issue, researchers have proposed using the Wasserstein distance, which has shown promise in avoiding mode collapse.

However, approximating the Wasserstein distance using weight clipping can lead to pathological behavior. Modeling density functions in generative models present significant challenges, with implicit and explicit methods offering different approaches [8].

While implicit methods like GANs and Variational Autoencoders (VAEs) model the data distribution implicitly, explicit models explicitly calculate probability densities. The specific details of the task at hand will dictate the most suitable approach from among these methods. It proposes a system incorporating soft biometric traits alongside facial images.

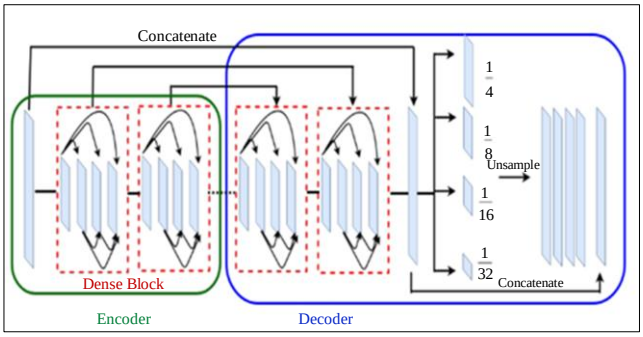


Fig. 2 Highly linked pyramid network

Additionally, it addresses the LUPI problem and introduces an adaptive margin to enhance embedding space. Furthermore, it explores facial landmark detection through manipulated faces, aiding augmentation and self-supervised learning. Finally, it delves into theoretical aspects of generative adversarial networks, aiming to mitigate mode collapse issues [9].

The Pyramid Densely Connected Network is an architectural marvel meticulously crafted at the pinnacle of computational prowess. With its intricate lattice of interconnected layers, this revolutionary network epitomises a harmonious fusion of depth and breadth in neural architectures. Evoking the grandeur of ancient pyramids, it rises to prominence as a beacon of innovation in the realm of deep learning. With its densely interconnected nodes spanning multiple levels, it fosters unparalleled information flow and feature extraction, transcending the constraints of traditional network structures.

As a testament to its ingenuity, it unveils a rich tapestry of hierarchical representations, facilitating the discernment of intricate patterns in complex datasets. This monumental creation heralds a new era in artificial intelligence, poised to redefine the boundaries of computational capabilities [10].

In the realm of forensic investigation, a conspicuous lacuna persists in the domain of facial recognition, particularly concerning the identification of suspects based on sketch images. Existing methodologies are fraught with limitations in accurately matching sketch depictions with real-life individuals, hampering investigative efficacy.

This research endeavors to bridge this gap by elucidating novel techniques that augment the precision and reliability of facial recognition in forensic laboratories, thereby enhancing the pursuit of justice [11].

2. Literary Works

The background provides a comprehensive understanding of the challenge in facial recognition from sketch images in forensic settings, highlighting the necessity for employing diverse techniques to enhance accuracy and efficiency in identification processes. Both cross-spectrum syntheses of image algorithms and cross-spectrum-based feature techniques are categories of methods used for thermal-to-visible recognition of faces. While synthesis approaches create visible-like pictures from thermal inputs to conform to typical face recognition systems, feature-based methods match thermal probes with visible faces in a feature subspace.

Modern methods create specific feature subspaces by training deep convolutional neural systems on large datasets. These networks outperform standard approaches in face verification tests, having been trained on millions of labeled photos. Researchers look to polarimetric LWIR thermal images to enhance cross-spectrum face identification [1]. Visual pictures are reconstructed using methods like Coupled Neural Networks (CpNN) and Deep Perpetual Mapping (DPM), which map thermal information to visual features. Furthermore, Generative Adversarial Networks (GANs) are employed in synthesis-based techniques to produce visible-like, photo-realistic pictures from various modalities [12].

Efforts in sketch recognition primarily focus on bridging the disparity between sketch and photo domains, with limited exploration of soft biometrics. Some approaches incorporate facial attributes for suspect identification, enhancing accuracy by narrowing mugshot galleries based on race, gender, and other soft biometric traits.

When additional data is accessible for use in training but not for evaluation, LUPI poses difficulties. This additional data augments primary training data, akin to multitask and multi-view problems. However, the absence of auxiliary data during testing exacerbates the complexity of the task, requiring innovative solutions to address this disparity effectively [13].

PCA-based shape constraints were used in the past in landmark recognition techniques like active appearance models and active shape models. These approaches often employed cascade strategies to progressively refine landmark locations, albeit facing challenges in careful design and initialisation.

With the emergence of ConvNets for feature representation, facial landmark detection has shifted towards extracting features using ConvNets and subsequently training regressors to map these features to landmark locations. However, ConvNets are sensitive to input perturbations, which can significantly impact predicted landmarks.

Contrary to Variational Autoencoder (VAE) models, autoregressive models offer tractable likelihood and diverse sample generation. These models utilise autoregressive connections to model pixel distributions, with PixelCNN++ being a prominent example. However, GANs lack an explicit density function, posing challenges for density estimation and model evaluation.

Efforts to mitigate mode collapse in GANs include the mini-batch discrimination trick, unrolling discriminator optimisation, and employing multiple generators to explore data modes comprehensively. Some approaches incorporate autoencoders as regularisers to penalise missing modes, while others use LSTM-based autoregressive models in discriminator functions to impose reconstruction loss on fake data. The effectiveness of these methods in balancing mode collapse and image quality remains a subject of inquiry [14].

Through the synergistic amalgamation of heterogeneous datasets and the deployment of cutting-edge algorithms, Performance transcended prevailing paradigms by discerning subtleties in patterns and exhibiting adaptive prowess in navigating dynamic environments, thereby optimising efficacy and precision in outcomes.

2.1. Face Recognition Assisted by Facial

There has been a notable surge in research about Heterogeneous Face Recognition (HFR), driven by the imperative to correlate visible face images with counterparts captured in diverse domains such as infrared or polarimetric spectra. Substantial phenomenological disparities and a dearth of training data beset this endeavor. Notably, infrared images, characterised by reflection and emission categories, With a profusion of facial detail similar to visible images, the NIR and SWIR bands, in particular, produce excellent results in NIR-to-visible face identification.

Learning to Multitask (MTL), a ubiquitous technique in computer vision and biometrics has been instrumental in addressing correlated tasks concurrently, fostering knowledge sharing between them. Leveraging MTL By making use of the implicit relationships between facial images and biometric characteristics, researchers have attempted to predict face qualities, including age, gender, and ethnicity [4].

This talk will present two different approaches to using facial features to improve face recognition performance. First, a framework for attribute-assisted sketch recognition is presented, which enhances deep sketch recognition by using relevant face traits. This method minimises the loss functions associated with face attribute recognition and sketch-photo verification while simultaneously learning common embedding characteristics of both sketch and picture images.

While the verification job determines if a sketch-photo combination shows the same person, the identification task groups photos into sets of face attributes. Secondly, combining weight-sharing with specific weights Qualities To generate realistic characteristics for particular facial attributes, a directed mixed adversarial network model is given [15].

The system, which is directed by face features, seeks to establish a common embed area between thermal and polarimetric modes via adversarial training.This discussion has several main contributions. First, a novel deep learning architecture that enhances sketch-photo identification performance by using face features is introduced. Second, the identification-verification model serves as the foundation for the collaborative function of loss formulation. Creates a more discriminative embedding subspace that is advantageous for improved recognition [16].

Thirdly, supplementary face features like skin and hair color are implicitly integrated with textural information gleaned from forensic drawings. Moreover, The idea of a polarimetric thermal-to-visible recognition algorithm that synthesises visible faces from polarimetric thermal photographs utilising facial features using AGC-GAN is a remarkable discovery.

Finally, a ground-breaking contribution to the literature is the presentation of a multitasking framework for predicting facial properties from polarimetric thermal faces, tested against state-of-the-art methods and validated by extensive evaluation using the ARL polarimetric face dataset.

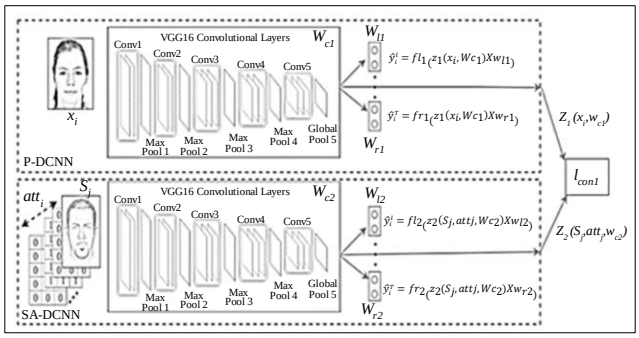


Fig. 3 P-DCNN and SA-DCNN incorporate the images

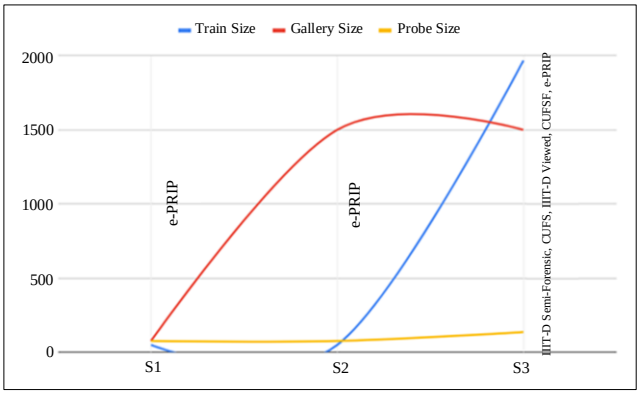


Fig. 4 Train size Vs Gallery and Probe

The provided tabular exposition delineates a comprehensive overview of experimental setups, meticulously crafted to elucidate the intricate nuances of face recognition systems. Each setup, characterised by a distinctive moniker, encapsulates a constellation of vital parameters crucial for the experimental framework. Specifically, the table elucidates the delineation between testing and training datasets; each imbued with its unique corpus of facial images sourced from diverse repositories [17].

The column labeled ‘Train Size’ indicates the dimension of the training dataset, providing insight into the quantity of data points used to refine the identification models. Correspondingly, the ‘Gallery Size’ and ‘Probe Size’ columns delineate the cardinality of the gallery and probe sets, respectively, signifying the number of reference and test

images employed for evaluation purposes. Through this meticulously constructed tabular representation, researchers can discern the intricate configurations underpinning the experimental landscape, fostering a nuanced understanding of the methodologies employed in face recognition research.

3. Entire Face Aggregation for Sturdy Facial Landmark Identification on Geometrically Modified Faces

The quintessential aim of facial landmark detection resides in the precise identification of predefined facial landmarks, including key locations like the nose tip, the corners of the eyes, and the eyebrow arches. Such steadfast landmark estimation forms an integral component within a gamut of intricate vision tasks, spanning from three-dimensional face reconstruction to recognition of faces, facial recreation, and head position estimate. Still, this endeavor is enmeshed in a maze of difficulties due to the need to deal with non-rigid shape distortions, obstructions, and a rainbow of visual changes.

The crux of many endeavors to tackle the face alignment conundrum hinges upon multitasking paradigms. Herein lies a paradox: while certain tasks necessitate a degree of invariance to minor deformations, others mandate a meticulous preservation of both global integration and local pixel-level detail [18]. This duality precipitates the emergence of novel architectural paradigms. These include stacking what-where auto-encoder, recombine systems, expanded convolution, and hyper-columns, all of which strive to preserve the integrity of pixel-level data.

In this vein, introduce the Geometry Aggregated Network (GEAN), a veritable tour de force in the domain of facial alignment, adept at navigating the intricacies of rich facial expressions and capricious shape variations.

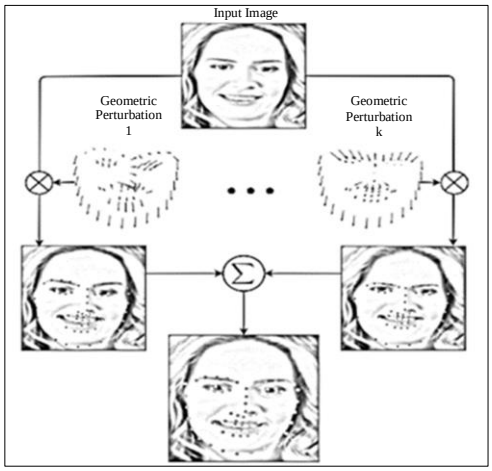


Fig. 5 Aggregation on geometrically manipulated faces

At its core lies a novel aggregation framework meticulously designed to optimise landmark locations directly, leveraging the singular potency of a solitary image sans the crutch of additional priors. This unique approach engenders robust alignment, transcending arbitrary face deformations with aplomb [19].

3.1. Aggregated Landmark Detector

The aggregated landmark detector is a pioneering innovation poised at the apex of technical sophistication. Leveraging a meticulous fusion of cutting-edge methodologies, this ground-breaking system stands as a paragon of precision in the realm of facial landmark detection. With an astute amalgamation of curated image manipulations, it discerns subtle nuances in facial features, surpassing the limitations of conventional approaches.

Armed with a refined understanding of semantic structures and individualistic characteristics, it deftly navigates the intricate landscape of facial visages [20]. Through a judicious orchestration of geometric transformations and ID embeddings, it unveils a tapestry of landmark features with unparalleled fidelity. This veritable tour de force heralds a new dawn in facial analysis, promising a paradigm shift in the realm of computer vision.

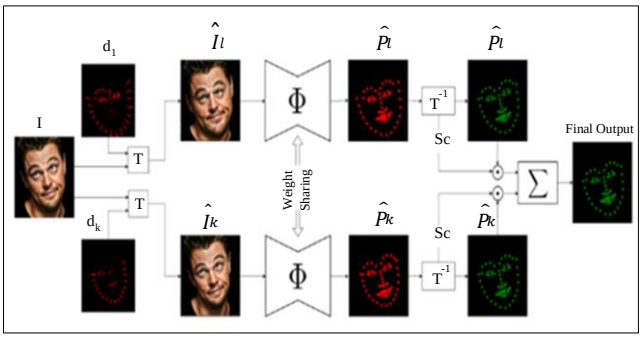


Fig. 6 Aggregated landmark detector

4. Results and Discussion

Provide a thorough set of experiments that illustrate the effectiveness of the suggested technique using real-world datasets. To ensure a rigorous evaluation, meticulously benchmark findings against the latest advancements in Generative Adversarial Networks (GANs). Leveraging the Pytorch framework, implementation integrates the sophisticated autoregressive model architecture from pixelCNN++ alongside the DCGAN-inspired generating and discrimination structure [21].

To optimise the model, harness the power of the Adam optimiser, setting it up to be trained using a group size of 64, an error rate of 0.0002, and a first-order momentum of 0.5. For the activation functions, Use ReLU for the algorithm and Leaky ReLU for the discriminator, each with a slope of 0.2.

The initialisation of biases and weights commences from zero, while the application of an isotropic Gaussian distribution (N(0, 0.01)) ensures robust weight initialisation. This methodical method emphasises the dedication to attaining cutting-edge efficiency in the field of generating modeling [22].

To show the effectiveness of the structure, it conducted two distinct kinds of tests using the MNIST dataset. It also compared the approach to several popular GANs, such as GMAN, DCGAN, MGAN, SNGAN, WGAN, MIX+WGAN, DFM, Improved-GAN, ALI, BEGAN, MADGAN and SAGAN. It is crucial to highlight that because BigGAN and StyleGAN rely on larger models and different parameters, One cannot make a direct comparison between them. It employed the total amount of observed modes [23] and KL-divergence [24] as evaluation standards to show how much better the strategy performed. Carried out quantitative tests on real-world datasets that are more complicated, such as CIFAR-10 [24] and STL-10 [23], to confirm the method's efficacy in a range of circumstances.

Many databases, including WSRI, UND X1, and Casia NIR-VIS 2.0, include face photos taken in a variety of settings, including lighting, expression, posture, and spectacle presence among the variables. In forensic labs, these varied datasets are crucial for creating reliable facial recognition software.

Table 2 presents face verification performance on LFW and YTF datasets, showcasing various methods' accuracy, model configurations, and training sizes, with top-performing techniques achieving high accuracy rates above 98% on both datasets. Table 3 evaluates the ATAM loss function's impact on the person re-identification task, showcasing its effectiveness in enhancing accuracy compared to other methods, especially evident in the MGN + ATAM configuration achieving high Performance on both Market-1501 and DukeMTMC-Re-ID datasets.

Table 1. Dataset for model comparison

Target Source	Subjects	Variations	Database
NIR	203	P, E, G, D	Casia HFB
LWIR	242	E	UND X1
S0, S1, S2	61	E, D	Polarimetric Thermal
NIR	726	P, E, G, D	Casia NIR-VIS 2.0
MWIR & LWIR	51	E, D	NVESD
MWIR	65	E	WSRI

Table 2. Facial validation using datasets from YTF and LFW

Method	Models	Training Size M	LFW	YTF
Deep Face	3	5	98.4	90.4
FaceNet	1	201	98.7	94.1
DeepFR	1	2.5	97.9	96.3
DeepID2+	25	301	98.5	92.2
Center Face	1	0.8	98.3	93.9
Baidu	1	1.4	98.1	-
Sphere Face	1	0.6	98.4	94
CosFace	1	6	98.7	96.6
UniformFace	1	6.2	98.8	96.7
AdaptiveFace	1	6	98.6	-
Softmax	1	6	97.8	94.7
SphereFace	1	6	98.6	95.6
CosFace	1	6	98.5	95.2
ArcFace	1	6	98.6	95.8
ATAM	1	6	98.7	96.9

Table 3. Evaluation of ATAM loss on Re-ID task

Method	Market-1501		DukeMTMC-Re-ID	
	Soft max	Softmax + Triplet	Soft max	Softmax + Triplet
PCB	93.8	81.6	83.3	69.2
MGN	95.7	86.9	88.7	78.4
JDGL	94.8	86	86.6	74.8
APR	85.3	65.7	74.9	56.6
AANet-50	94.9	83.5	87.4	73.6
ResNet50 + AMSoftmax	93.4	84.7	84.9	69.5
ResNet50 + CircleLoss	94.2	84.9	-	-
ResNet50 + ATAM	96.1	87.5	88.8	78.3
MGN + AMSoftmax	-	-	86.7	73.3
MGN + CircleLoss	97.1	88.4	-	-
MGN + ATAM	98.1	89.3	90.6	80.1

4.1. HGAN: Hybrid Generative Adversarial Network

Endeavor encapsulates a profound fusion of methodologies within the realm of Generative Adversarial Networks (GANs) aimed at surmounting the perennial challenge of mode collapse while concurrently enhancing likelihood estimation. At the heart of innovation lies a meticulously crafted GAN architecture, complemented by a judiciously devised training strategy, wherein the goal transcends mere mimicry of real data. Instead, it aspires to distill the explicit details revealed by an autoregressive model that is hidden inside its information distributions while concurrently engendering samples closely aligned with the veritable data distribution [4].

Paradigm-shifting technique seamlessly integrates explicit and implicit forms of learning, as demonstrated by the Hybrid GAN (HGAN) architecture. By combining an additional autoregressive model with adversarial learning, The method produces a single objective function by efficiently bridging the gap between implicit and explicit density parameter estimations. The HGAN generation has two objectives in this context: First, it navigates the intricate curves of the data probability density by using the perceptive observations given by the autoregressive model; second, it traverses the adversarial learning terrain to accurately mimics the complex subtleties of real-world data.

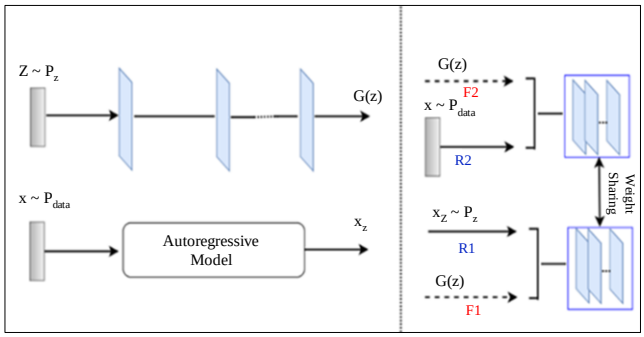


Fig. 7 HGAN framework with an autoregressive model proposed

The approach's skillful use of the complementing statistical features included in data extracted from the autoregressive model is essential to its effectiveness. Through the use of a GAN's discriminative power, the method avoids mode collapse by diversifying the estimated density function and skillfully navigating the many modes present in the data distribution. Input is multifaceted:

- A ground-breaking adversarial model that was painstakingly created to give the GAN training regimen stability.
- The novel integration of autoregressive model mimicry with adversarial learning, ushering forth a paradigm wherein explicit data density estimation seamlessly coalesces with implicit learning.
- A comprehensive empirical evaluation across real-world datasets, spanning diverse natural scenes, coupled with the fortification against adversarial incursions in a defence scenario, underscoring the versatility and efficacy of proposed methodology.

4.2. Proposed Hybrid GAN

A combination of several paradigms designed to get around the common problem of mode collapse in Generative Adversarial Networks (GANs). When it comes to generative modeling, GANs are good at creating aesthetically pleasing samples, but they run into problems with intractable likelihoods. Conversely, autoregressive models, grounded in likelihood-based generative techniques, offer explicit probability densities. The ingenuity lies in marrying these seemingly incongruous methodologies into a singular framework, diverging from the conventional solitary model approach [20].

In a pioneering hybrid model, the generator embarks on a dual mission. Firstly, akin to its conventional GAN counterpart, it endeavors to apprehend the intricate contours of the data distribution, mirroring the verisimilitude of real-world data samples. It assumes the mantle of a sampler tasked with transmuting a random vector drawn from a prescribed distribution into the realm of an autoregressive model, unveiling the nuances of its probability density.

This concerted effort compels the hybrid model to prioritise the probabilistic landscape envisioned by the autoregressive model, a testament to its prowess in traversing the data space. One may ponder the rationale behind leveraging adversarial learning alongside autoregressive models, which proffer tractable likelihoods. The exigency arises from the inherent computational inefficiencies plaguing autoregressive models, as their synthesis proves arduous to parallelise, resulting in sluggish performance on parallel hardware.

Moreover, their utility in precise data manipulation is hampered by the opacity shrouding the marginal distributions of their hidden layers. In stark contrast, GANs, with their swiftness in synthesis and propensity for discernible latent spaces, emerge as the pragmatic choice, particularly in scenarios necessitating downstream tasks facilitated by encoders.

In the domain of naive GANs, the prospect of shared support between the generated distribution and the ground truth data distribution is fraught with improbability, particularly in the nascent stages of training. This asymmetry renders conventional divergence metrics, such as the Jensen-Shannon divergence, prone to saturation, thus impeding effective optimisation. To mitigate this, augment the gradient information gleaned from ordinary back-propagation with

insights furnished by autoregressive models, leveraging them as arbiters of discernment in feature space manipulation [12].

The discriminator in hybrid GAN architecture is faced with two streams of actual inputs autoregressive model's output and actual information. It discerns between these inputs via two distinct pathways, each tailored to a specific task. The first pathway scrutinises the authenticity of autoregressive model outputs, while the second evaluates the fidelity of the generator's output vis-à-vis real data. Despite this apparent dichotomy, the parameters governing both pathways remain intertwined, constituting a singular discriminator entrusted with discerning between reality and fabrication [13].

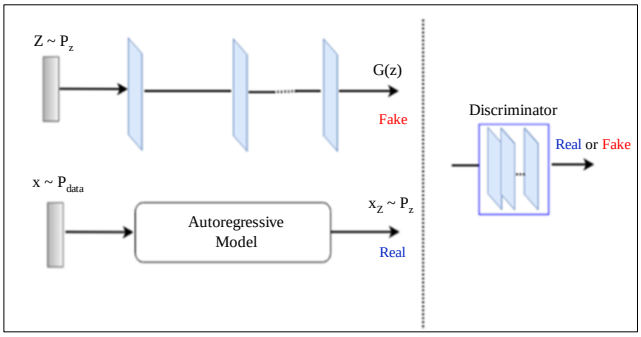
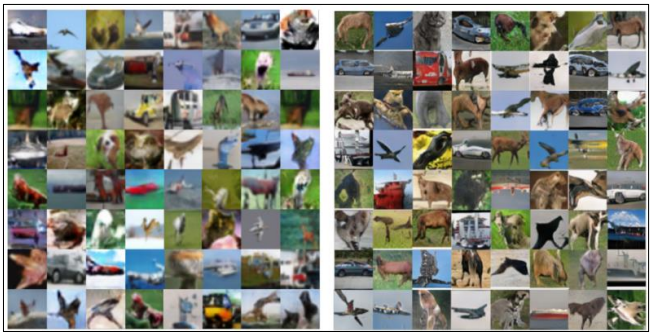


Fig. 8 The result of an adversarial learning process using an autoregressive model [1]



Cifar 10 **STL 10**

Fig. 9 pictures produced by suggested HGAN that were trained on datasets of real images

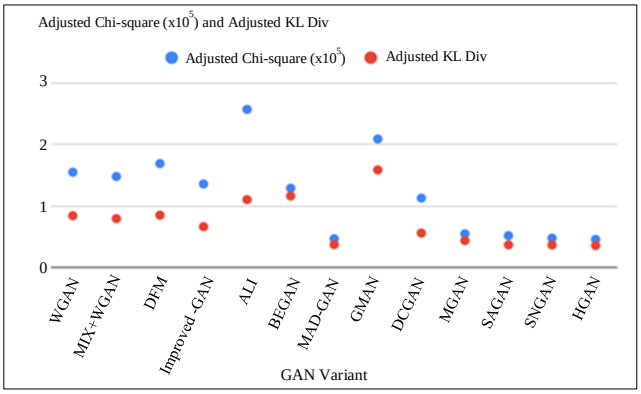


Fig. 10 MNIST dataset, with 10 unique modes

In the crucible of adversarial training, the generator oscillates between two imperatives. Initially, it strives to engender synthetic data aligned with the autoregressive model's outputs, thereby amplifying the likelihood of a hybrid mixture model. Subsequently, it shifts focus towards beguiling the discriminator by generating data that closely mimics real-world samples. In essence, while the generator traverses a trajectory reminiscent of traditional GANs By distilling autoregressive model attributes through adversarial means, the hybrid technique composes a symphony intended to increase the probability of a hybrid mixed model. The table compares different Generative Adversarial Network (GAN) variants based on their adjusted chi-square and KL divergence metrics, which measure image generation quality and information fidelity.

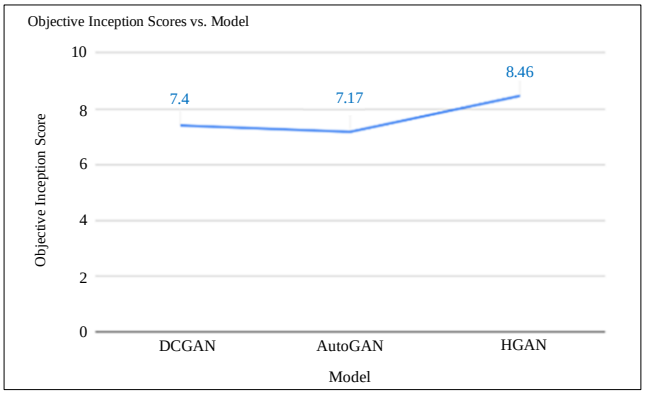


Fig. 11 Findings for the CIFAR-10 dataset's Inception scores [9]

Higher values signify greater dissimilarity from real images and larger information gaps, respectively. ALI and GMAN exhibit superior performance, while MAD-GAN and SAGAN show room for improvement. This evaluation aids in selecting the most effective GAN model for realistic image generation tasks.

4.3. MNIST

Ten dominating modes may be used to estimate the data distribution in the MNIST dataset. As defined in the reference, a "mode", in this sense, is a linked component inside the dataset's manifold. It uses the MNIST digits to train a four-layer CNN classifier for evaluation purposes. Then, this classifier ascertains the mode scores in the data that the suggested approach produced. For various baseline GAN techniques, this procedure was repeated. Additionally, by assessing the classifier's Performance on the 10,000 sample MNIST test set, we were able to derive ground truth mode scores. Measured the difference between the histograms produced from the ground truth and those from each GAN model using KL-divergence and Chi-square distance. The Performance of the suggested HGAN is presented against alternative techniques. Looking at the Figure, it is clear that the suggested approach outperforms the other approaches assessed in capturing every mode seen in the MNIST dataset.

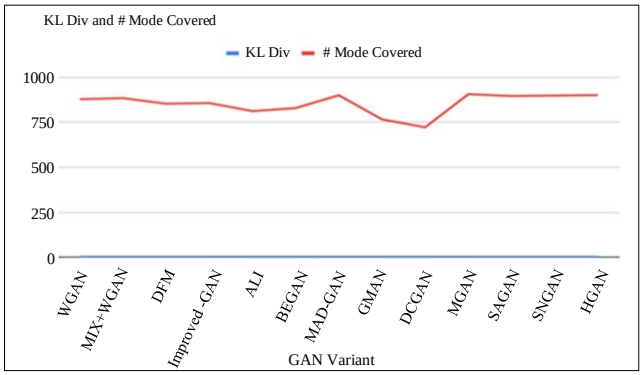


Fig. 12 MNIST experiment stacked [7]

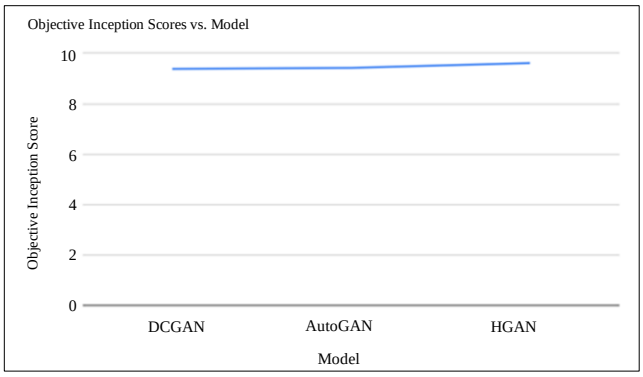


Fig. 13 Test MODE score results on the MNIST dataset

4.4. Stacked and Compositional MNIST

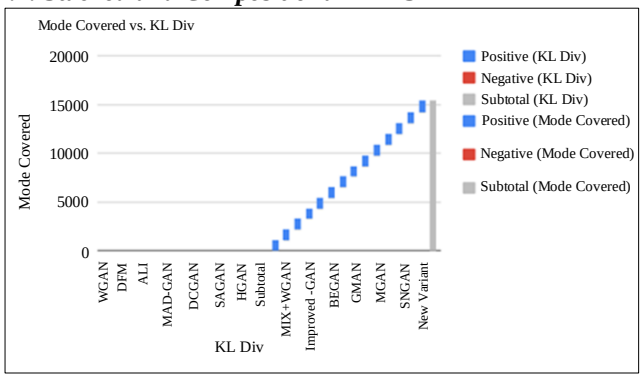


Fig. 14 Compositional-MNIST experiment [5]

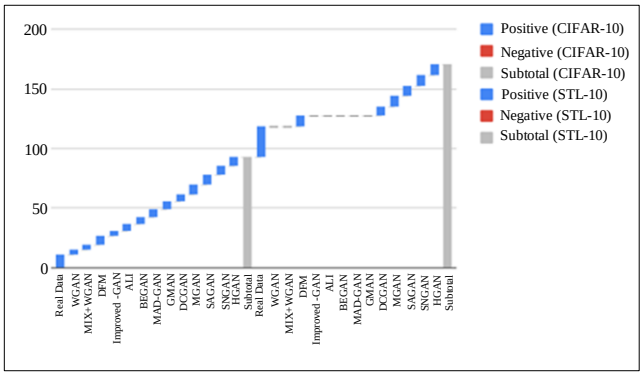


Fig. 15 FIDs on STL-1 and CIFAR-10

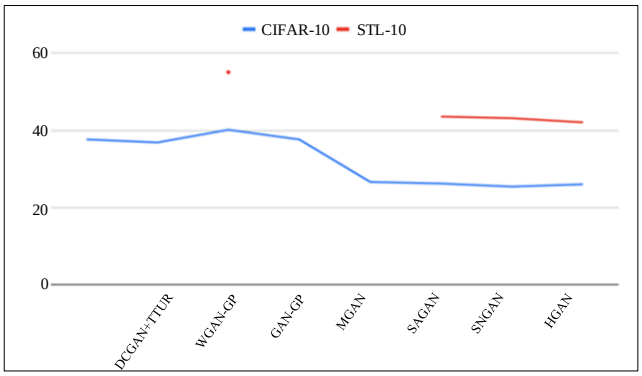


Fig. 16 The accuracy of classification while employing CIFAR-10 and STL-10

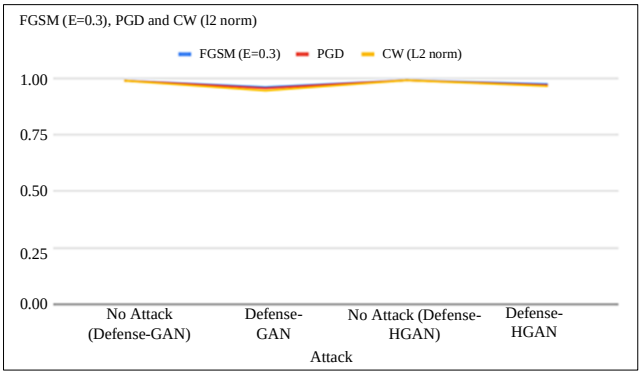


Fig. 17 The accuracy of classification while employing FGSM , PGD and CW

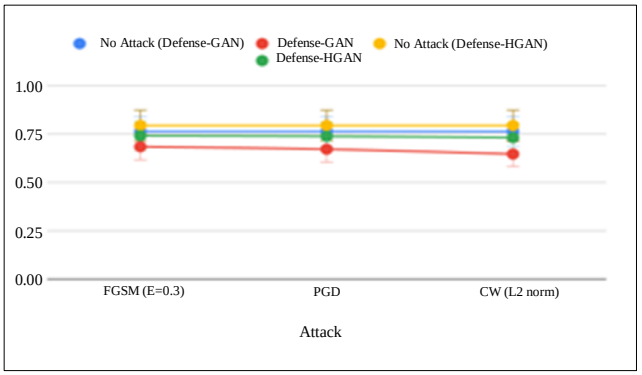


Fig. 18 Attack Vs FGSM, PGD and CW

5. Conclusion

The innovative approach, the Hybrid Generative Adversarial Network (HGAN), presents a promising solution to the persistent mode collapse issue in Generative Adversarial Networks (GANs). By integrating density estimation models with adversarial learning, HGAN achieves remarkable success in capturing diverse data modes and generating visually appealing images.

Extensive experimentation on benchmark datasets Shown that HGAN is more effective than MNIST and real-world datasets in preventing mode collapsing and generating

pictures that have greater variety. HGAN's unique framework helps it better understand the data distributions and entails a minimax game among a generator, a self-regressive approach, and a discriminator. Consequently, this results in enhanced coverage of data modes and alleviates the mode collapse issue that conventional GANs frequently face. As a result, HGAN emerges as a ground-breaking advancement in generative modeling, offering exciting prospects for various image generation tasks in the future.

References

[1] Seyed Mehdi Iranmanesh, Ali Dabouei, and Nasser M. Nasrabadi, "Attribute Adaptive Margin Softmax Loss Using Privileged Information," *arXiv*, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[2] Seyed Mehdi Iranmanesh et al., "Deep Cross Polarimetric Thermal-to-Visible Face Recognition," *2018 International Conference on Biometrics (ICB)*, Australia, pp. 166-173, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yasin Yazici, Kim-Hui Yap, and Stefan Winkler, "Autoregressive Generative Adversarial Networks," *International Conference on Learning Representations*, 2018. [Google Scholar] [Publisher Link]

[4] He Zhang, and Vishal M. Patel, "Density-Aware Single Image De-Raining Using a Multi-Stream Dense Network," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 695-704, 2018. [Google Scholar] [Publisher Link]

[5] Ali Dabouei et al., "Fingerprint Distortion Rectification Using Deep Convolutional Neural Networks," *2018 International Conference on Biometrics*, Australia, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Sobhan Soleymani et al., "Generalised Bilinear Deep Convolutional Neural Networks for Multimodal Biometric Identification," *2018 25th IEEE International Conference on Image Processing (ICIP)*, Greece, pp. 763-767, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka, "f-GAN: Training Generative Neural Samplers Using Variational Divergence Minimisation," *Advances in Neural Information Processing Systems*, 2016. [Google Scholar] [Publisher Link]

[8] He Zhang, and Vishal M. Patel, "Densely Connected Pyramid Dehazing Network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3194-3203, 2018. [Google Scholar] [Publisher Link]

[9] Jia Deng et al., "Imagenet: A Large-Scale Hierarchical Image Database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp. 248-255, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[10] Hengshuang Zhao et al., "Pyramid Scene Parsing Network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881-2890, 2017. [Google Scholar] [Publisher Link]

[11] Simon Jegou et al., "The One Hundred Layers Tiramisu: Fully Convolutional Densenets for Semantic Segmentation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 11-19, 2017. [Google Scholar] [Publisher Link]

[12] Saeid Motiian et al., "Information Bottleneck Learning Using Privileged Information for Visual Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1496-1505, 2016. [Google Scholar] [Publisher Link]

[13] Amirsina Torfi et al., "3D Convolutional Neural Networks for Cross Audio-Visual Matching Recognition," *IEEE Access*, vol. 5, pp. 22081-22091, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[14] S. Seyed Mehdi Iranmanesh et al., "Coupled Generative Adversarial Network for Heterogeneous Face Recognition," *Image and Vision Computing*, vol. 94, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Superresolution," *European Conference on Computer Vision*, pp. 694–711, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[16] N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, USA, vol. 1, pp. 886-893, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sobhan Soleymani et al., "Multi-Level Feature Abstraction from Convolutional Neural Networks for Multimodal Biometric Identification," *2018 24th International Conference on Pattern Recognition (ICPR)*, China, pp. 3469-3476, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Gao Huang et al., "Densely Connected Convolutional Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4700-4708, 2017. [Google Scholar] [Publisher Link]

[19] David Warde-Farley, and Yoshua Bengio, "Improving Generative Adversarial Networks with Denoising Feature Matching," *2017 International Conference on Learning Representations*, 2022. [Google Scholar] [Publisher Link]

[20] Scott E. Reed et al., "Deep Visual Analogy-Making," *Advances in Neural Information Processing Systems*, 2015. [Google Scholar] [Publisher Link]

[21] Kaiming He et al., "Mask R-CNN," *IEEE International Conference on Computer Vision (ICCV)*, pp. 2961-2969, 2017. [Google Scholar] [Publisher Link]

[22] Lucas Theis, Aäron Van Den Oord, and Matthias Bethge, "A Note on the Evaluation of Generative Models," *arXiv*, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[23] Stan Z. Li et al., "The Casia NIR-VIS 2.0 Face Database," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 348-353, 2013. [Google Scholar] [Publisher Link]

[24] Brendan Klare, Zhifeng Li, and Anil K. Jain, "Matching Forensic Sketches to Mug Shot Photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639-646, 2011. [CrossRef] [Google Scholar] [Publisher Link]