ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF COMPUTER AND INFORMATICS

SYNTHETIC DATA GENERATION FOR FACIAL EXPRESSION RECOGNITION

Graduation Project Interim Report

Ömer Yıldırım
150190115

Department: Computer Engineering

Advisor: Prof. Dr. Hazım Kemal Ekenel

January 2024

Statement of Authenticity

I hereby declare that in this study

1. all the content influenced from external references are cited clearly and in detail,
2. and all the remaining sections, especially the theoretical studies and implemented software that constitute the fundamental essence of this study is originated by my individual authenticity.

İstanbul, January 2024

Ömer Yıldırım

Acknowledgements

I extend my sincere gratitude to Professor Dr. Hazım Kemal Ekenel, my advisor, for affording me the invaluable opportunity to collaborate with him and his esteemed research group on the subject of "Synthetic Data Generation for Facial Expression Recognition". His expert guidance and unwavering support were instrumental in the successful execution of this project.

Special appreciation is also extended to Ali Azmoudeh and Ziya Ata Yazıcı for their noteworthy support and understanding, which significantly contributed to the overall progress and achievement of the undertaking.

SYNTHETIC DATA GENERATION FOR FACIAL EXPRESSION RECOGNITION

(SUMMARY)

The project is centered around the innovative application of Synthetic Data Generation to advance the field of Facial Expression Recognition (FER). The primary objectives of the project can be categorized into four key goals. Firstly, the development of a robust Synthetic Data Generation system is targeted, specifically tailored to produce synthetic facial expression data. This system is strategically designed to complement existing datasets used for training FER models, thereby addressing challenges associated with limited and biased training data.

As a second objective, the project aims to elevate the performance of facial expression recognition models. This is to be achieved by integrating a hybrid dataset comprising both real and synthetically generated facial expression data. The incorporation of synthetic data is anticipated to ameliorate limitations related to insufficient training data, fostering more accurate and generalizable FER models.

The third goal involves an in-depth exploration of Generative Adversarial Networks (GANs) as a powerful tool for generating high-quality synthetic facial images. The project endeavors to rigorously assess the efficacy of GANs in this context, with a specific focus on evaluating their impact on the overall performance of facial expression recognition models. This exploration extends to optimizing the generation process, with a particular emphasis on generating data based on action units commonly used in facial expression recognition problems.

The final objective underscores the project's ambition to contribute significantly to Facial Expression Recognition (FER) research. By showcasing the viability and advantages of synthetic data, the project aims to address existing challenges in FER model development and propel advancements in the broader landscape of FER research.

The project's evaluation criteria encompass multiple facets. Foremost among them is model accuracy, where the facial expression recognition models trained on synthetic data should exhibit comparable or superior accuracy (with a minimum target of 90%) to those trained on conventional datasets. Additionally, the project places a crucial emphasis on the system's ability to detect and mitigate biases in the synthetic data generation process, aspiring to achieve a balanced and unbiased model across diverse demographics.

Computational efficiency is identified as another pivotal evaluation criterion. The synthetic data generation and model training processes should be optimized to attain competitive training times, with a targeted computational efficiency of at least 90% compared to models trained solely on real data. Lastly, the project's contribution to FER research will be assessed based on the depth of insights provided, the demonstration of synthetic data viability, and the creation of diverse and meticulously crafted datasets, thereby expanding the knowledge base in the field.

Moving beyond the project summary, the abstract articulates a clear vision for the project. It emphasizes the generation of facial expressions that do not exist in reality, achieved through algorithmic processes. The abstract underscores the comparative analysis of model performances between datasets sourced from professional environments and those generated synthetically. Furthermore, it highlights the exploration and refinement of generation processes, specifically targeting the incorporation of action units commonly used in FER problems. The abstract concludes by emphasizing the significance of the project's scientific contribution, demonstrating the power of AI-generated data and anticipating the increased data availability for future applications in FER research.

YÜZ İFADESİ TANIMA İÇİN SENTETİK VERİ ÜRETİMİ

(ÖZET)

Bu proje, Yüz İfadesi Tanıma (FER) alanını ilerletmek için Sentetik Veri Üretiminin inovatif uygulaması etrafında şekillenmektedir. Projenin birincil hedefleri dört temel amaç olarak kategorize edilebilir. İlk olarak, sentetik yüz ifadesi verileri üretmek için özel olarak uyarlanmış başarılı bir Sentetik Veri Üretimi sisteminin geliştirilmesi hedeflenmektedir. Bu sistem, FER modellerini eğitmek için kullanılan mevcut veri kümelerini tamamlamak ve böylece sınırlı ve taraflı eğitim verileriyle ilgili zorlukları gidermek için stratejik olarak tasarlanmıştır.

İkinci bir hedef olarak proje, yüz ifadesi tanıma modellerinin performansını yükseltmeyi amaçlamaktadır. Bu, hem gerçek hem de sentetik olarak oluşturulmuş yüz ifadesi verilerini içeren hibrit bir veri setinin entegre edilmesiyle gerçekleştirilecektir. Sentetik verilerin dahil edilmesinin, yetersiz eğitim verileriyle ilgili sınırlamaları iyileştirmesi ve daha doğru ve genelleştirilebilir FER modellerini teşvik etmesi beklenmektedir.

Üçüncü hedef, yüksek kaliteli sentetik yüz görüntüleri oluşturmak için güçlü bir araç olarak Üretken Çekişmeli Ağların (GAN'lar) derinlemesine araştırılmasını içermektedir. Proje, özellikle yüz ifadesi tanıma modellerinin genel performansı üzerindeki etkilerini değerlendirmeye odaklanarak, GAN'ların bu bağlamdaki etkinliğini titizlikle değerlendirmeye çalışmaktadır. Bu keşif, yüz ifadesi tanıma problemlerinde yaygın olarak kullanılan eylem birimlerine (action unit) dayalı veri üretmeye özel bir vurgu yaparak, üretim sürecini optimize etmeye kadar uzanmaktadır.

Nihai hedef, projenin Yüz İfadesi Tanıma (FER) araştırmalarına önemli ölçüde katkıda bulunma isteğinin altını çizmektedir. Proje, sentetik verilerin uygulanabilirliğini ve avantajlarını sergileyerek, FER model geliştirmedeki mevcut zorlukları ele almayı ve FER araştırmalarının daha geniş manzarasında ilerlemeleri teşvik etmeyi amaçlamaktadır.

Projenin değerlendirme kriterleri birçok yönü kapsamaktadır. Bunların başında, sentetik veriler üzerinde eğitilen yüz ifadesi tanıma modellerinin geleneksel veri kümeleri üzerinde eğitilenlerle karşılaştırılabilir veya daha üstün doğruluk (minimum %90 hedefiyle) sergilemesi gereken model doğruluğu gelmektedir. Buna ek olarak proje, sistemin sentetik veri oluşturma sürecindeki önyargıları tespit etme ve azaltma kabiliyetine çok önemli bir vurgu yaparak, farklı demografik özellikler arasında dengeli ve tarafsız bir model elde etmeyi amaçlamaktadır.

Hesaplama verimliliği bir diğer önemli değerlendirme kriteri olarak belirlenmiştir. Sentetik veri oluşturma ve model eğitim süreçleri, yalnızca gerçek veriler üzerinde eğitilen modellere kıyasla en az %90 oranında hedeflenen bir hesaplama verimliliği ile rekabetçi eğitim sürelerine ulaşmak için optimize edilmelidir. Son olarak, projenin FER araştırmalarına katkısı, sağlanan içgörülerin derinliğine, sentetik verilerin uygulanabilirliğinin gösterilmesine ve çeşitli veri kümelerinin oluşturulmasına ve böylece alandaki bilgi tabanının genişletilmesine dayalı olarak değerlendirilecektir.

Bu değerlendirme, proje için net bir vizyon ortaya koymaktadır. Gerçekte var olmayan yüz ifadelerinin algoritmik süreçlerle elde edilmesini vurgulamaktadır. Bu proje, profesyonel ortamlardan elde edilen veri kümeleri ile sentetik olarak üretilenler arasındaki model performanslarının karşılaştırmalı analizinin altını çiziyor. Ayrıca, özellikle FER problemlerinde yaygın olarak kullanılan eylem birimlerinin (action units) dahil edilmesini hedefleyen üretim süreçlerinin araştırılması ve iyileştirilmesi vurgulanmaktadır. Bu araştırmanın, projenin bilimsel katkısının önemini vurgulayarak, yapay zeka tarafından üretilen verilerin gücünü göstererek ve FER araştırmalarında gelecekteki uygulamalar için artan veri kullanılabilirliğini öngörerek sona ermesi öngörülmektedir.

Contents

1 Introduction and Problem Definition

In recent years, deep learning has demonstrated remarkable efficacy in various computer vision tasks, such as image classification and face recognition. The notable success of deep learning can be attributed to the robust representation capabilities of deep neural networks and the availability of extensive labeled training data. While deep learning methods for Facial Expression Recognition (FER) have been developed, their performance in unconstrained environments remains unsatisfactory. This inadequacy is partly attributed to facial expression databases, which typically have limited training data. Despite the abundance of diverse facial images on the Internet, the manual annotation of these images is both time-consuming and costly. Consequently, training deep neural networks with a restricted number of labeled samples presents a non-trivial challenge.

To address the issue of insufficient training data, some studies have sought to leverage auxiliary data to effectively train neural networks for FER. For instance, a deep Convolutional Neural Network (CNN) architecture [1] has been proposed, trained on hybrid databases comprising seven distinct facial expression databases. However, the bias among these databases [2] introduces challenges such as overfitting, potentially diminishing performance on the target database. Developing an appropriate strategy for fine-tuning is demanding, given that deep networks are pre-trained with high capacity on large-scale data. Despite the application of techniques like data augmentation and dropout to mitigate overfitting, their impact on performance improvement is limited. Consequently, enhancing FER performance under conditions of insufficient training data remains an ongoing challenge.

Recently, there has been growing interest in a novel generative model known as the Generative Adversarial Network (GAN) for its promising ability to generate high-quality data. The ascendancy of deep learning techniques in unraveling intricate probability distributions across diverse data types is a cornerstone of this endeavor. Notably, Generative Adversarial Networks (GANs), introduced by Goodfellow et al. in 2014[3], have emerged as a formidable tool in this context. Operating within a two-player minimax framework, GANs comprise a generative model and an adversary model. The generative model strives to encapsulate the underlying data distribution, while the adversary, or discriminator, discerns between authentic data and synthetic samples. The iterative interplay between these models persists until the generated samples attain indistinguishability from real data, exemplifying the prowess of GANs in data synthesis. Typically composed of a generator network and a discriminator network engaged in adversarial optimization, GAN can produce synthetic samples emulating the distribution of real samples from training data. Notably, GAN has been successfully employed in face-related tasks such as posed face synthesis [4] and facial attributes transfer [5]. These methods generate photorealistic facial images sharing the same identity as input facial images. Leveraging the similarities between synthetic and real images, these generated images, featuring variations in conditions like expressions and poses, can serve as valuable data augmentation to address the challenge of limited training data in deep learning.

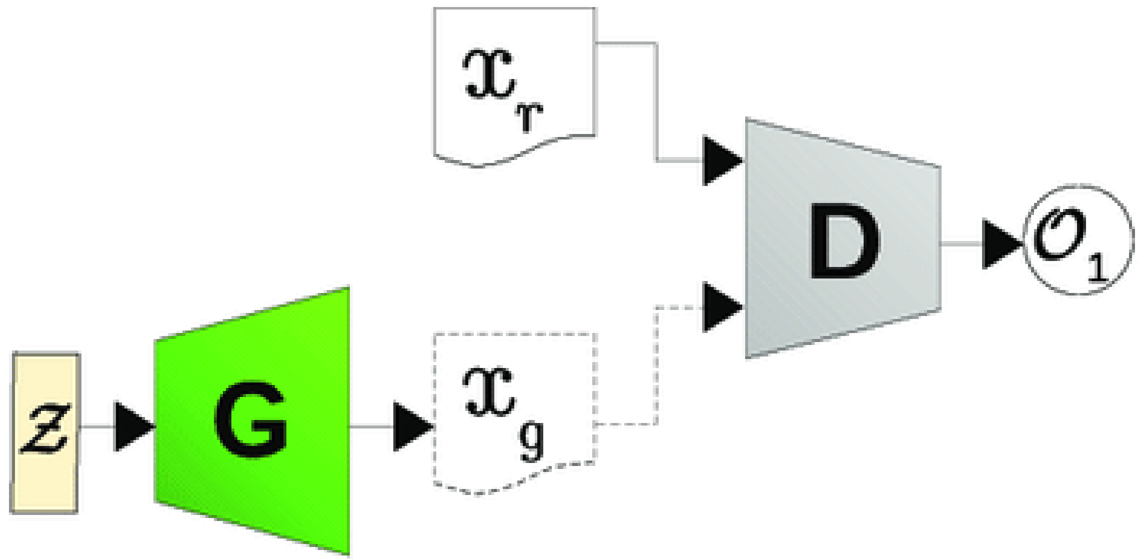

Figure 1.1: General Structure of vanilla GAN [3]; Z: input noise, G: Generator, D: Discriminator, xr : real sample, xg: generated sample, O1: Output of binary classification to real/fake.

Nevertheless, several unresolved issues persist when using synthetic images directly for training deep neural networks. The generation of high-quality images approximating the distribution of real images requires substantial effort for GAN, particularly when confronted with limited training data. Moreover, although synthetic images generated by GAN may be of high quality, indistinguishable to the human eye, their utilization does not guarantee effective improvement in the performance of deep neural networks. This is due to potential substantial intrinsic data bias between synthetic and real images, which may pose challenges for the recognition network.

The overarching objective of this project is to harness synthetic data generation to enrich the developmental landscape of facial expression recognition systems. The methodology involves training models on facial images that transcend the boundaries of reality, being algorithmically created. This innovative approach seeks to augment the volume and diversity of training data, promising enhancements in the refinement and performance of facial expression recognition models.

A primary focus lies in the creation of synthetic facial images tailored for the training of FER models. This approach, driven by the rationale to supplement existing datasets, often originating from professional environments or gleaned from publicly available images, introduces a novel dimension. By integrating non-existing expressive faces generated through advanced algorithms, the project aims to diversify the training data, thereby enhancing the robustness of the models.

A pivotal facet of the project involves a meticulous evaluation and comparison of the performance of FER models trained on conventional datasets against those trained on synthetic data. The emphasis is squarely on model classification accuracy, providing nuanced insights into the efficacy of synthetic data in bolstering overall model performance.

Beyond performance metrics, the project delves into the intricacies of data generation processes. This entails a nuanced exploration, including generating data based on action units relevant to facial expression recognition problems. Such a refined approach seeks to elevate the quality and authenticity of synthetic data, ensuring that the generated expressions align closely with the intricacies of human emotion.

The comparative analysis of model performances represents a significant scientific endeavor, showcasing the potency of AI-generated data in training facial expression recognition models. Moreover, this project envisions contributing substantively to future FER research and applications by not only demonstrating the viability of synthetic data but also by providing an increased volume of diverse and meticulously crafted data.

In essence, this interim report lays a robust foundation for a comprehensive exploration into synthetic data generation for facial expression recognition. By adeptly harnessing the capabilities of deep generative models, the project endeavors to address existing challenges in FER model development and contribute valuable insights to the broader landscape of artificial intelligence research.

2 Literatre Survey

Facial expression synthesis has been a subject of significant research, with several pioneering works contributing to the understanding and advancement of this field. Among these, the study by Susskind et al. [6] stands out for its incorporation of constraints like 'raised eyebrows' on generated facial samples. The framework utilizes a Deep Belief Network (DBN) with two hidden layers of 500 units, incorporating the Facial Action Coding System (FACS) vector and identity to generate faces with diverse expressions. Subsequently, the emergence of Generative Adversarial Network (GAN) models introduced innovative approaches to facial expression synthesis.

DyadGAN [7] and ExprGAN [8] are two such models designed explicitly for face generation. DyadGAN focuses on generating facial images conditioned on the expressions of a dyadic conversation partner, while ExprGAN overcomes challenges related to controlling expression intensity without relying on intensity-labeled training data. This is achieved through an expression controller module and an identity preserving loss function.

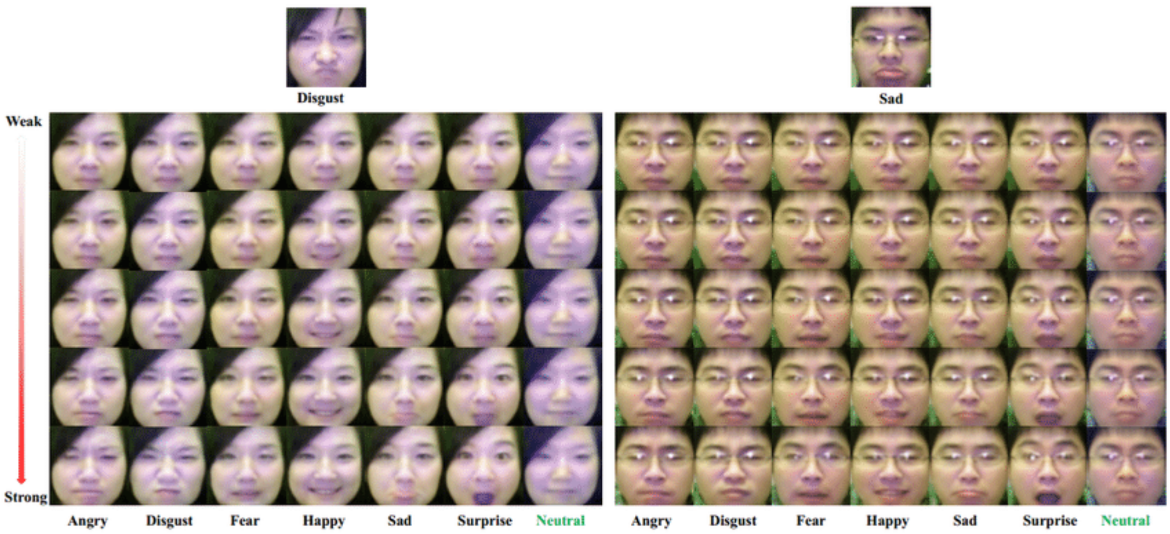

Figure 2.1: ExprGAN [8] for image generation of facial expression.

One noteworthy GAN-based model, G2GAN by Song et al. [98], offers fine-grained control over target expressions and facial attributes, providing realistic and identity-preserving images. The model leverages unpaired training with a pair of GANs—one removing the expression while the other synthesizes it. Additionally, StarGAN [9] presents a scalable solution for multi-domain image-to-image translation using a unified GAN model. Attributes like hair color, gender, and age can be modified based on the desired values.

Attribute editing GAN (AttGAN) [10] offers a framework for editing attributes among a set for face images, employing adversarial loss, reconstruction loss, and attribute classification constraints. DIAT [11], CycleGAN [12], and IcGAN [13] serve as baseline models for comparison.

In 2018, Qiao et al. [14] extended G2GAN, introducing a model based on Variational Autoencoder GANs (VAEGANs) for synthesizing facial expressions from a single image and landmarks. Unlike ExprGAN, their model does not require the target class label and dispenses with the need for a neutral expression as an intermediate level in the transfer procedure. Pumarola et al. [15] utilize facial Action Units as a one-hot vector for unsupervised expression synthesis.

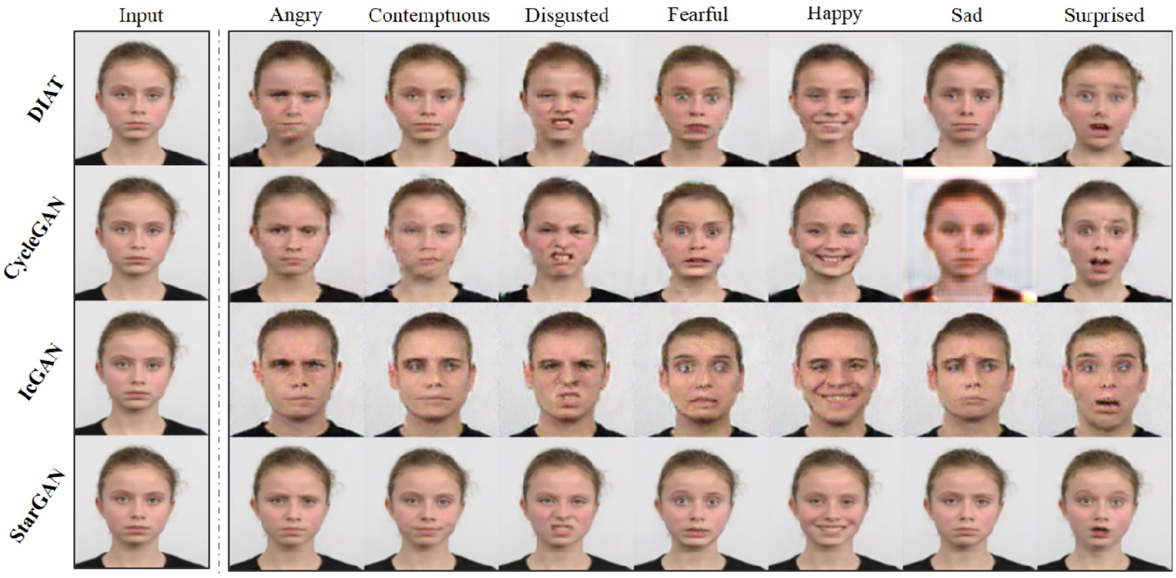

Figure 2.2: Visual comparison of the GAN models [11], [12], [13], [9] in order (top to bottom) for facial expression synthesis on RaFD dataset, images are in courtesy of the reviewed papers.

Another VAEGAN-based model by Lai and Lai [16] introduces a symmetric loss for preserving the symmetrical property of the face during translation from various head poses to a frontal view. FaceID-GAN [107] incorporates a face identity classifier as the third player in the GAN framework, distinguishing identities of real and synthesized faces. Additionally, Lai and Lai [16] use GANs for emotion-preserving representations, transforming non-frontal facial images into frontal ones while preserving identity and emotion expression.

In a recent publication [17], a two-step GAN framework is proposed. The first component maps images to a 3D vector space representing emotion, and the second component generates different expressions using the obtained 3D points. This model allows fine-grained control over synthesized expressions through a continuous vector space representing arousal, valence, and dominance.

Furthermore, a Facial Expression Synthesis Generative Adversarial Network (FESGAN) [18] undergoes pre-training to generate facial images exhibiting diverse facial expressions. Notably, FESGAN is meticulously designed to enhance training image diversity by generating images featuring new identities derived from a predefined distribution. Subsequently, an expression recognition network undergoes joint learning with the pre-trained FESGAN within an integrated framework.

Specifically, the classification loss derived from the recognition network is employed to concurrently optimize the performance of both the recognition network and the FES-GAN generator. Furthermore, to mitigate the challenge posed by data bias between real and synthetic images, this study proposes an intra-class loss incorporating a novel Real Data-Guided Back-Propagation (RDBP) algorithm. The RDBP algorithm is employed to mitigate intra-class variations among images belonging to the same class, thereby significantly enhancing the ultimate performance of the proposed approach.

The advent of Generative Adversarial Networks (GANs) revolutionized facial expression synthesis. StarGAN [5] addressed multi-domain image-to-image translation across databases by introducing a mask vector to disregard unlabelled categories during training. This innovation allows unlabelled facial images to be translated into expressions while preserving information from other domains. A geometry-guided GAN (G2-GAN) proposed by [19] leverages facial geometry information to guide synthesis. G2-GAN employs a pair of GANs for facial expression removal and synthesis. In contrast, a geometry-contrastive GAN (GC-GAN) by [20] synthesizes facial expressions conditioned on geometry information at the semantic level, employing contrastive learning for facial geometry embedding. Furthermore, a new attribute-guided facial image synthesis based on GAN was developed in [21] for image-to-image translation.

In conclusion, this literature survey provides a comprehensive overview of the evolution of facial expression synthesis techniques, from traditional methods to the current era dominated by Generative Adversarial Networks. The discussed methodologies highlight the continuous pursuit of achieving realistic and controllable facial expression synthesis. The advancements in GAN-based approaches underscore the effectiveness of these models in addressing challenges encountered by earlier methods. As technology progresses, it is anticipated that further innovations will emerge, contributing to the refinement and expansion of facial expression synthesis in diverse applications.

3 Novel Aspects and Technological Contributions

The project introduces several novel aspects and technological contributions aimed at advancing the field of Facial Expression Recognition (FER) through the integration of synthetic data generated by Generative Adversarial Networks (GANs). The key contributions of this project are outlined below:

1. Advanced Synthetic Data Generation for FER
   - Literature Gap: While existing studies leverage GANs for facial image synthesis, our approach extends this by focusing on the specific requirements of FER, aiming to overcome the challenges associated with limited training data.
2. Strategic Integration of Synthetic Data into Training
   - Novelty: A systematic methodology is developed for strategically integrating synthetic facial images into the training process of FER models.
3. Evaluation Metrics for Synthetically Enhanced Models
   - Literature Gap: While existing studies focus on model accuracy, our approach introduces metrics that specifically evaluate the impact of synthetic data on the robustness and generalization capability of FER models.
4. Fine-Grained Expression Generation and Control
   - Novelty: The project employs advanced GAN architectures to generate facial expressions with fine-grained control over attributes such as intensity and relevance to action units.
5. Comprehensive Comparative Analysis
   - Novelty: The project conducts a comprehensive comparative analysis, not only focusing on conventional metrics but delving into the intricacies of data generation processes and their impact on model performance.
   - Literature Gap: While previous works may present results based solely on classification accuracy, our project provides a holistic view by examining the intricacies of data generation and their implications on FER model development.

In essence, these novel aspects and technological contributions collectively propel the field of FER forward by addressing critical gaps in existing literature, providing strategic insights into the integration of synthetic data, and offering a nuanced understanding of the impact of advanced GAN-generated data on FER model performance. The project stands at the forefront of innovative approaches to enhance the robustness and efficacy of facial expression recognition systems.

4 System Requirements

4.1 Non-Functional Requirements

Non-functional requirements of the project other than functional requirements are provided below.

4.1.1 Security of Data Privacy

The system must adhere to strict data privacy standards. Any personally identifiable information (PII) used during training or generated during the synthetic data creation process should be handled with confidentiality. Additionally, any storage or transmission of sensitive data should be encrypted.

4.1.2 Security of Model Security

The trained models should be protected against unauthorized access and tampering. Access controls should be in place to ensure that only authorized users can modify or deploy models. Model files and parameters should be securely stored.

4.1.3 Ethical Considerations on Transparency and Consent

The system should provide transparency into the data sources used for synthetic data generation. Users should be aware of the origin of the data and any potential biases associated with it. This information should be documented in the system's user manuals.

If real facial images are used for training or evaluation, explicit consent must be obtained from individuals, ensuring they are informed about the purpose of data usage. This includes both publicly available images and images from proprietary databases.

4.1.4 Ethical Considerations on Bias Mitigation

Efforts should be made to identify and mitigate biases in the synthetic data generation process. The system should avoid perpetuating stereotypes or introducing biases that may impact the fairness of the facial expression recognition models.

4.2 Use Cases

4.2.1 User: End-User/Application Developer

Use Case 1: Configure Synthetic Data Generation

- Description: The data scientist should be able to configure parameters for synthetic data generation, including facial expression variations, image resolutions, and the number of generated samples.
- Success Criteria: The system generates synthetic facial images according to the specified configurations.

Use Case 2: Train Facial Expression Recognition Model

- Description: The data scientist should be able to train a facial expression recognition model using both real and synthetic data.
- Success Criteria: The trained model achieves a satisfactory accuracy level, and the data scientist can monitor training progress and performance metrics.

4.2.2 User: Researcher

Use Case 3: Evaluate Model Performance

- Description: The researcher or evaluator should be able to assess the performance of the facial expression recognition model trained using synthetic data.
- Success Criteria: The system provides detailed metrics on the model's accuracy, precision, recall, and other relevant performance indicators.

Use Case 4: Analyze Data Bias

- Description: The researcher should be able to analyze and identify any biases introduced during the synthetic data generation process.
- Success Criteria: The system provides insights into potential biases, enabling researchers to make informed decisions on bias mitigation strategies.

5 Project Plan

5.1 Project Resources

Resource requirements for the project in terms of both software, hardware, and datasets are provided below.

5.1.1 Software Requirements

- Python: The primary programming language for data processing, machine learning, and image generation.
- Jupyter Notebooks: For code development, experimentation, and documentation.
- Image Processing Libraries: OpenCV for image data preprocessing and augmentation.
- Image Generation Libraries: Exploring libraries like DCGAN, GANs, or StyleGAN for generating synthetic images.

5.1.2 Hardware Requirements

- Computer: A powerful computer with a good GPU to handle deep learning tasks efficiently. The specific hardware requirements will depend on the scale of your project.
- Storage: Sufficient storage space to store datasets and generated data.
- CUDA: Supported NVIDIA GPU (Graphics Processing Unit) for high-complexity neural network computations.

5.1.3 Libraries for Data Collection and Processing

- FER Datasets: As we plan to use existing FER datasets, we will need to access and process them. Common datasets include CK+, FER-2013, AffectNet, and others.
- Synthetic Data Generation Libraries: Depending on our approach, we will use generative adversarial networks (GANs), convolutional neural networks (CNNs), or other algorithms for image generation.

5.2 Work Breakdown and Work Assignment

An essential guide for project planning and execution, the Work Breakdown Structure (WBS) serves as a visual roadmap outlining the tasks and subtasks meticulously planned for successful completion during the project. This structured breakdown offers a comprehensive overview of the project's components. As a pivotal tool in project management, the WBS not only enhances organizational clarity but also serves as a foundational element to ensure that each facet of the project is thoughtfully addressed and executed. The subsequent WBS figure visually encapsulates the hierarchical representation of project tasks.

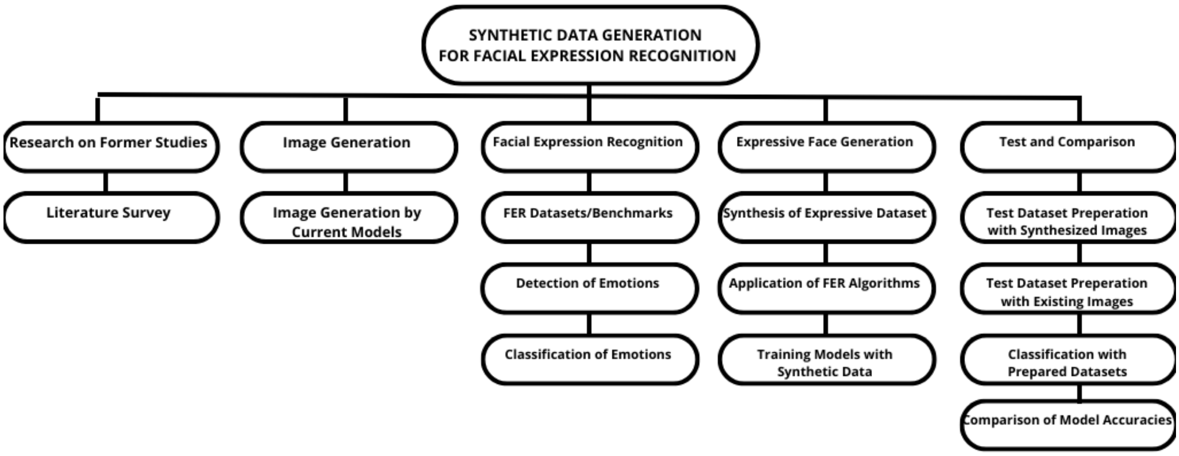

Figure 5.1: Work Breakdown Structure of Project

5.3 Time Plan

The GANTT Chart presented below serves as a dynamic visual representation illustrating the project's timeline and planned activities. This chart encapsulates the strategic planning of the project, mapping out key tasks and their corresponding subtasks based on the detailed breakdown provided in earlier chapters and the work breakdown structure. As an invaluable project management tool, the GANTT Chart not only enhances project organization but also aids in tracking progress, resource allocation, and ensures a systematic approach to task completion. The subsequent GANTT Chart figure provides an insightful glimpse into the project's trajectory, offering a structured representation of planned activities over time.

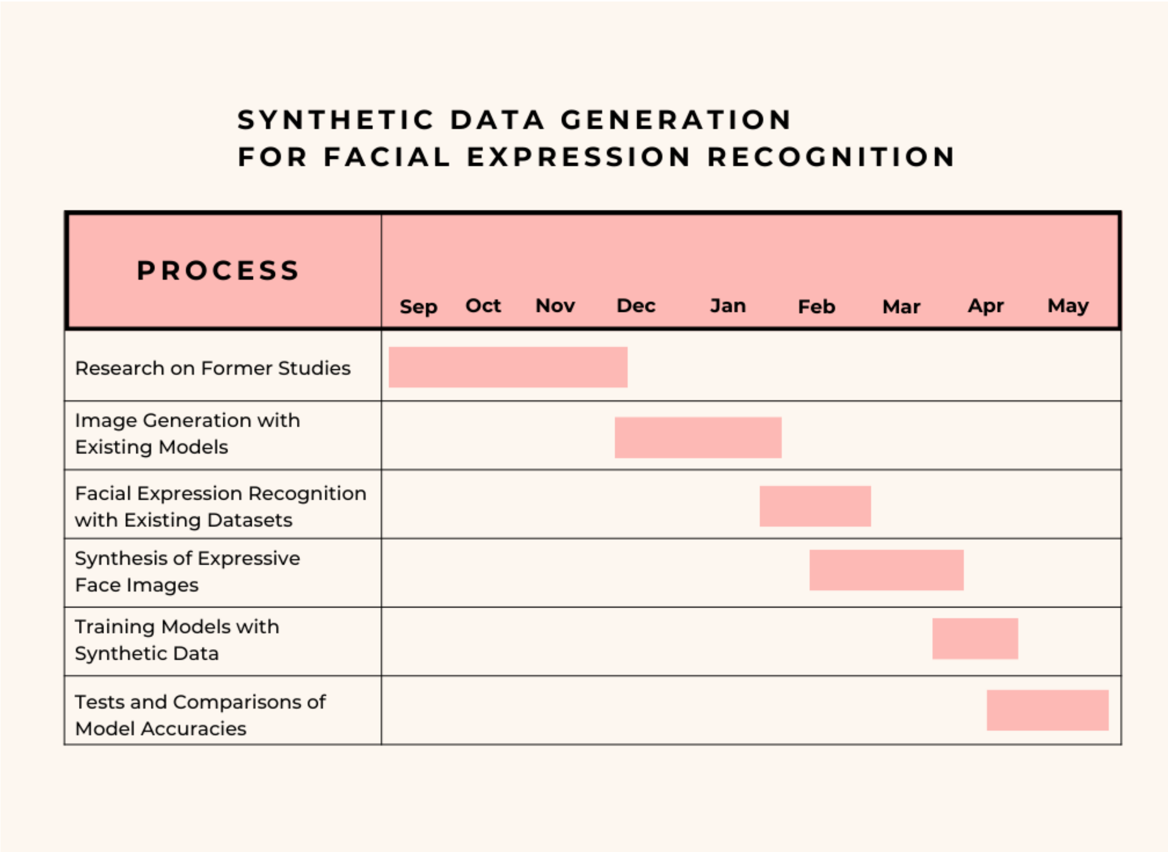

Figure 5.2: The GANTT Chart of Project

6 Goals and Evaluation Criteria

6.1 Project Goals

The primary goals of this project are as follows:

1. Implement Synthetic Data Generation: Develop a system capable of generating synthetic facial expression data to augment existing datasets used for training facial expression recognition models.
2. Enhance Facial Expression Recognition Models: Improve the performance of facial expression recognition models by training them on a combination of real and synthetic data, with a focus on addressing limitations related to insufficient training data.
3. Explore Generative Adversarial Networks (GANs): Investigate the effectiveness of GANs in generating high-quality synthetic facial images and evaluate their impact on model performance in facial expression recognition.
4. Contribute to FER Research: Contribute substantively to the field of Facial Expression Recognition (FER) research by showcasing the viability and benefits of synthetic data, thus addressing challenges in FER model development.

6.2 Evaluation Criteria

At the conclusion of the project, the system will be evaluated based on the following criteria:

1. Model Accuracy: The accuracy of the facial expression recognition models trained on synthetic data should be comparable to or exceed those trained on conventional datasets, with a target accuracy of at least 90%.
2. GAN Metrics for Synthetic Dataset: Introduce and refine numerical metrics specifically tailored for evaluating the quality and diversity of synthetic facial images generated by GANs. This includes assessing metrics such as Inception Score, Frechet Inception Distance, and others, providing a quantitative measure of the synthetic dataset's performance.
3. Computational Efficiency: The synthetic data generation and model training processes should be optimized for computational efficiency, with the goal of achieving training times that are competitive (having at least %90 of computational efficiency of state-of-models) with models trained solely on real data.
4. Contribution to FER Research: The project's impact on advancing Facial Expression Recognition research will be evaluated based on the depth of insights provided, the demonstration of synthetic data viability, and the expansion of diverse and meticulously crafted datasets.

References

[1] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *CoRR*, vol. abs/1511.04110, 2015. [Online]. Available: http://arxiv.org/abs/1511.04110

[2] S. Li and W. Deng, "Deep facial expression recognition: A survey," *CoRR*, vol. abs/1804.08348, 2018. [Online]. Available: http://arxiv.org/abs/1804.08348

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

[4] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning gan for pose-invariant face recognition," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1283–1292, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:21011865

[5] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," *CoRR*, vol. abs/1711.09020, 2017. [Online]. Available: http://arxiv.org/abs/1711.09020

[6] J. Susskind, G. Hinton, J. Movellan, and A. Anderson, *Generating Facial Expressions with Deep Belief Nets*, 05 2008.

[7] Y. Huang and S. Khan, "Dyadgan: Generating facial expressions in dyadic interactions," 07 2017, pp. 2259–2266.

[8] H. Ding, K. Sricharan, and R. Chellappa, "Exprgan: Facial expression editing with controllable expression intensity," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 09 2017.

[9] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," 06 2018, pp. 8789–8797.

[10] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Attgan: Facial attribute editing by only changing what you want," *IEEE Transactions on Image Processing*, vol. PP, pp. 1–1, 05 2019.

[11] M. Li, W. Zuo, and D. Zhang, "Deep identity-aware transfer of facial attributes," 10 2016.

[12] J.-Y. Zhu, T. Park, P. Isola, and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 10 2017, pp. 2242–2251.

[13] G. Perarnau, J. van de Weijer, B. Raducanu, and J. M. Álvarez, "Invertible conditional gans for image editing," *CoRR*, vol. abs/1611.06355, 2016. [Online]. Available: http://arxiv.org/abs/1611.06355

[14] F. Qiao, N.-M. Yao, Z. Jiao, Z. Li, H. Chen, and H. Wang, "Emotional facial expression transfer from a single image via generative adversarial nets," *Computer Animation and Virtual Worlds*, vol. 29, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:46938785

[15] A. Pumarola, A. Agudo, A. M. Martínez, A. Sanfeliu, and F. Moreno-Noguer, "Ganimation: Anatomically-aware facial animation from a single image," *CoRR*, vol. abs/1807.09251, 2018. [Online]. Available: http://arxiv.org/abs/1807.09251

[16] Y.-H. Lai and S.-H. Lai, "Emotion-preserving representation learning via generative adversarial network for multi-view facial expression recognition," 05 2018, pp. 263–270.

[17] V. Vielzeuf, C. Kervadec, S. Pateux, and F. Jurie, "The many variations of emotion," 05 2019, pp. 1–7.

[18] Y. Yan, Y. Huang, S. Chen, C. Shen, and H. Wang, "Joint deep learning of facial expression synthesis and recognition," *CoRR*, vol. abs/2002.02194, 2020. [Online]. Available: https://arxiv.org/abs/2002.02194

[19] L. Song, Z. Lu, R. He, Z. Sun, and T. Tan, "Geometry guided adversarial facial expression synthesis," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 627–635. [Online]. Available: https://doi.org/10.1145/3240508.3240612

[20] F. Qiao, N. Yao, Z. Jiao, Z. Li, H. Chen, and H. Wang, "Geometry-contrastive generative adversarial network for facial expression synthesis," *CoRR*, vol. abs/1802.01822, 2018. [Online]. Available: http://arxiv.org/abs/1802.01822

[21] B. Bozorgtabar, M. S. Rad, H. K. Ekenel, and J. Thiran, "Using photorealistic face synthesis and domain adaptation to improve facial expression analysis," *CoRR*, vol. abs/1905.08090, 2019. [Online]. Available: http://arxiv.org/abs/1905.08090