A Unified Approach for Binary-Class and Multi-Class Data Augmented Generation

Frederic Rizk
The Center for Advanced Computer Studies
University of Louisiana at Lafayette
Lafayette, LA, U.S.A.
frederic.rizk1@louisiana.edu

Rodrigue Rizk
Department of Computer Science
University of South Dakota
Vermillion, SD, U.S.A.
rodrigue.rizk@usd.edu

Dominick Rizk
Department of Electrical Engineering and Computer Science
The Catholic University of America
Washington, DC, U.S.A.
rizkd@cua.edu

Patrick Rizk
Department of Mechanical Engineering
University of Quebec at Trois-Rivières
Drummondville, QC, Canada
patrick.rizk@uqtr.ca

Chee-Hung Henry Chu
The Center for Advanced Computer Studies
Informatics Research Institute
University of Louisiana at Lafayette
Lafayette, LA, U.S.A.
henry.chu@louisiana.edu

***Abstract*—Deep neural networks excel in a wide range of tasks but require diverse datasets to prevent overfitting. Overfitting occurs when a network fits training data too precisely, leading to poor generalization. Data Augmentation is often used to mitigate overfitting aiming at enlarging and improving the quality of training datasets, facilitating the construction of superior deep learning models. MAGAN algorithm emerges as an innovative approach that functions as a Meta-Analysis of Generative Adversarial Networks (GANs). MAGAN harnesses the latent space capabilities of GANs to confront the challenges presented by binary-class, multi-class, grayscale, and RGB images, effectively covering a wide spectrum of scenarios. In this paper, we propose the use of MAGAN algorithm for binary-class and multi-class data augmented generation. We also undertake an in-depth experimental analysis, evaluating the performance of the proposed MAGAN-based approach in comparison to two alternative baseline scenarios: one without any augmentation and another utilizing a conventional augmentation method. To gauge the effectiveness of the proposed technique, we employed diverse classification metrics, including accuracy, loss, precision, recall, F1-score, and the confusion matrix. Our results demonstrate that the proposed approach surpasses the other two scenarios achieving improvements in terms of accuracy by a factor of x1.15 and x1.03, respectively. This underscores the significant advantages of harnessing MAGAN, a meta-analysis of GANs, for data augmentation across a range of image types and classification tasks.**

***Index Terms*—machine learning, deep neural networks, overfitting, data augmentation, data augmented generation, generative adversarial networks, latent space.**

I. Introduction

In machine learning, the quantity and quality of training data are pivotal for model effectiveness. As tasks become more complex, dataset size is crucial for pattern discernment and efficient generalization. The training dataset is fundamental, enabling the model to learn patterns and rules. Its quality and representativeness are key during the training phase. Choosing a suitable dataset is critical. It should mirror the real-world problem's complexity. A well-structured, large dataset enhances the model's ability to generalize to new, unseen data, improving predictions in real-world scenarios. A larger dataset provides a more comprehensive depiction of the underlying data distribution.

This larger representation allows the model to detect an array of features in the data. As a result, the model learns more about the patterns and correlations between features, resulting in more accurate predictions. Furthermore, a larger dataset mitigates the risk of overfitting [1], a typical problem in machine learning. Overfitting occurs when a model becomes overly specialized in the training data and struggles to generalize to new, unseen data.

As the availability of more data drives breakthroughs in machine learning, human labeling of massive datasets becomes a limiting factor for large-scale deep learning systems [2], [3]. Data augmentation has developed as an emerging approach in the field [4], solving the issue of insufficient training data. It involves augmenting the dataset with varied and realistic synthetic examples, hence enhancing the performance and generalization of machine learning models. Among several data augmentation approaches, Generative Adversarial Networks (GANs) proposed by Goodfellow [5] have received a lot of interest. GANs excel in understanding the underlying data distribution and generating high-quality synthetic data. These networks are made up of a generator and a discriminator, with the generator generating synthetic samples from random noise and the discriminator differentiating between real and fake data.

Through the process of adversarial training, the system undergoes a learning phase where it becomes proficient in mapping samples from the latent space to the real data

Work by F.R. and C.-H.H.C. was supported in part by the U.S. National Science Foundation under grant number OIA-1946231 and the Louisiana Board of Regents for the Louisiana Materials Design Alliance (LAMDA).

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00022

distribution. Once this proficiency is attained, GANs gain the capability to generate credible images by drawing samples from a random distribution. Previous research efforts [6], [7] have primarily focused on improving synthesis quality by identifying a more precise distribution aligned with ground truth data. However, there has been limited exploration into understanding what GAN truly comprehends within the latent space. For instance, in the context of face synthesis, although the latent code governs which facial features are generated, the precise relationship between the latent code and various semantic properties of the resulting facial image, such as age and gender, remains ambiguous. Various approaches [8], [9] have been proposed to exert control over the generated images. However, it is worth noting that their quality remains notably inferior when compared to the performance achieved by unconditioned GANs [6], [10]. A study conducted by Radford et al. [11] suggests that delving into the arithmetic characteristics of vectors within the latent space provides insights into how GANs acquire certain semantic information in their earliest hidden layers. A prior research study by Bau et al. [12] illustrates that the generator is capable of synthesizing specific visual attributes through its intermediary layers. Nevertheless, there remains a significant knowledge gap regarding how alterations in the latent space can precisely influence the desired output of generated images.

In their study, Härkönen et al. [13] highlight the importance of the latent space in GANs for image generation, in which the generator converts random noise into meaningful data representations. Following a similar intuition, we propose, in this paper, the use of MAGAN algorithm [14] for data augmentation. MAGAN is an emerging algorithm that is suitable for exploring the latent space vector distances and offers valuable insights into the relationship between generated images.

The main contributions of this work can be summarized as follows:

- It proposes a unified MAGAN-based approach for both binary-class and multi-class data augmentation that generates synthetic images by taking the mean of all vectors within each class. The incorporation of the class means preserving the core traits of each class, resulting in semantically relevant and different augmentations.
- It addresses the problem of data scarcity by creating additional images utilizing the MAGAN-based augmentation, particularly when working with smaller datasets. The enhanced data increased the training data's effective size, resulting in more robust and dependable model performance. Overfitting was effectively minimized by training models on a larger, more dataset and unseen dataset, allowing the models to generalize to previously unknown data.
- It presents a comprehensive evaluation in which the proposed approach is applied to different types of scenarios covering the binary-class, multi-class, grayscale, and RGB image datasets. The evaluation results indicate the proposed unified approach's flexibility and promise for improving the training process across a wide range of domains.

The structure of this paper can be outlined as follows. Section I provides a concise introduction to Generative Adversarial Networks (GANs) and data augmentation, conducts a literature review, and highlights the contributions made within this study. Moving forward, Section II describes the methodology applied to both binary-class and multi-class classification problems. Section III showcases experimental results for both classification problems, where we compare classification metrics across three scenarios: no augmentation, conventional techniques, and MAGAN. Finally, in Section IV, we conclude the paper by summarizing the main discoveries and discussions presented.

II. Proposed Unified MAGAN-based Approach

In this section, we propose the unified MAGAN-based approach for binary-class and multi-class data augmented generation.

The MAGAN algorithm constitutes a comprehensive exploration of the latent space inherent in GANs, wherein the mean μ and standard deviation σ of vector groups are computed for each class. The determination of the mean for vectors sharing a common label serves to encapsulate the principal features characterizing each class. Subsequently, by displacing epsilon ϵ times the standard deviation from the mean vector μ, a diverse array of images is generated, thereby contributing to the enhancement of the generalization process. Figure 1 depicts the MAGAN algorithm [14] that we are proposing to use for data augmentation. By leveraging the MAGAN algorithm, we enlarge the dataset and infuse it with greater semantic value by incorporating images that reside within the interval defined by $(\mu \pm \sigma \times \epsilon)$.

Algorithm 1: an algorithm for Meta-Analysis for GANs' latent space (*MAGAN*)

Input: Vectors **X** sampled from the latent space, Output Label **L**, Number of Classes C
Output: ε
procedure LATENT_META_ANALYSIS (**X, L**)
  V← *groupby*(**X**, L)
  for k ∈ {0,1, ..., C-1} **do**
    N ← len($\mathbf{V_k}$)
    $\boldsymbol{\mu_k} \leftarrow \frac{\sum_{i=1}^{N} \boldsymbol{V_k}[i]}{N}$
    $\boldsymbol{\sigma_k} \leftarrow \sqrt{\frac{\sum_{i=1}^{N} (\boldsymbol{V_k}[i] - \boldsymbol{\mu_k})^2}{N-1}}$
  end
  d(i,j)= *distance*($\boldsymbol{\mu}$[i], $\boldsymbol{\mu}$[j]) with i, j ∈ {0, 1, ..., C-1}
  $\mathbf{d_{min}} = min(\mathrm{d}[i])$
  $\varepsilon(i) = arg\{\varepsilon(i) \mid \boldsymbol{d_{min}}[i] > distance(\boldsymbol{\mu}[i], \boldsymbol{\mu}[i] \pm \varepsilon[i] \times \boldsymbol{\sigma}[i])\}$
  with: i ∈ {0, 1, ..., C-1}
end procedure

Fig. 1. The MAGAN algorithm.

We enlarge the dataset size by integrating an additional 1000 images per class generated using the MAGAN algorithm. Utilizing Generative Adversarial Networks (GANs) as a parameter-controlled data generator facilitates data-driven augmentation. Specifically, the MAGAN algorithm harnesses the latent space capabilities of GANs to generate synthetic images that effectively enrich the dataset. By leveraging GANs in this manner, we ensure that the augmented data aligns with the distribution of the original dataset while introducing variations that enhance the diversity and representativeness of the training samples. This approach enables an effective increase in the dataset size and improves the performance of deep learning models trained on the augmented data.

III. Experimental Setup, Results and Discussion

In this section, we will delve deeply into our experimental setup, examine results, and interpret them. Our objective is to thoroughly assess the MAGAN approach for data augmentation, employing samples from both the Fashion-MNIST dataset and the cat-vs-dog dataset. We aim to assess the efficacy of our proposed approach and draw meaningful conclusions by undertaking a systematic examination and analysis.

All experiments in this study were conducted using Google Colab powered by a GPU-accelerated environment (NVIDIA-SMI 525.105.17, Driver Version: 525.105.17, CUDA Version: 12.0). The implementation of the deep learning algorithms in this work utilized Keras with Google TensorFlow as the backend, complemented by additional scientific computing libraries such as matplotlib, numpy, and scikit-learn.

A. Datasets

In this work, we rigorously evaluate our proposed data augmentation technique using two datasets. The proposed MAGAN-based method assesses the impact of data scarcity on model performance, with deliberate extraction of 1000 samples per class from each original dataset to highlight challenges posed by restricted data availability.

We demonstrate the algorithm's efficacy on subsets from Fashion-MNIST and cat-vs-dog datasets, showcasing its applicability across diverse datasets. The intentional selection ensures a comprehensive evaluation in both multi-class (Fashion-MNIST) and binary-class (cat-vs-dog) classification scenarios. Additionally, by incorporating RGB images from cat-vs-dog and grayscale images from Fashion-MNIST, we thoroughly examine the algorithm's performance across different color spaces. This approach extends the generalizability of our findings and addresses nuances in binary-class and multi-class classification tasks, as well as variations in image color formats.

B. Data Transformation Techniques

Two distinct data augmentation procedures have been evaluated in our experimentation: the conventional method and the MAGAN augmentation. The conventional method included a variety of augmentation techniques such as random horizontal flipping, random rotation, random shear, random zoom, and others as shown in Table I. These enhancements were made on the fly throughout the training session. Contrarily, the MAGAN approach leveraged the latent space's mean vector μ and standard deviation σ, creatively generating synthetic images. The epsilon factor ϵ, a critical parameter, was judiciously set to different values for each class to strike a balance between generating diverse images and maintaining realism. The values of ϵ for the subsets of Fashion-MNIST and cat-vs-dog datasets are presented in Figure 2 and Figure 3 respectively.

TABLE I
Data Augmentation Methods

Data Augmentation Methods	
Method	Technique
Conventional	• Rescaling: The images were rescaled to a range of 0 to 1, converting pixel values to the interval [0, 1]. • Rotation_range: Random rotations within ± 15 degrees were applied. • Shear_range: A shear transformation was employed with a range of 0.1. • Zoom_range: Images were randomly zoomed in and out with a range of 0.2. • Horizontal_flip: Horizontal flipping was enabled. • Width_shift_range: Horizontal shifts with a range of 0.1. • Height_shift_range: Vertical shifts within a range of 0.1.
MAGAN	Increasing the dataset size by incorporating 1000 images per class generated using the MAGAN algorithm. GAN is used as a parameter-controlled data generator for data-driven augmentation.

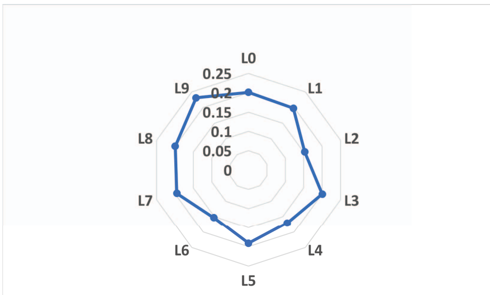

Fig. 2. The values of the parameter ε for each class label for Fashion MNIST classification.

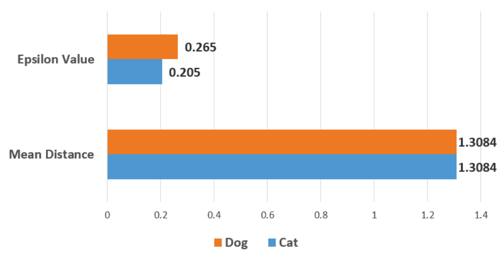

Fig. 3. The minimum distance and the values of the parameter ε for each class label for cat-vs-dog classification.

C. Multi-Class - Fashion-MNIST

To tackle the challenge of limited data availability, this study conducted evaluations using a restricted subset of the Fashion-MNIST dataset. Instead of the typical 6,000 images per class in the training set, our experiment utilized only 1,000 images for each class for each training and testing dataset.

To comprehensively compare MAGAN with traditional data augmentation, we conducted a classification analysis on the subset, augmenting it with an additional 1,000 generated images per class. Visual representations of accuracy and loss evolution across training epochs for all three approaches were plotted on a single graph, spanning epochs 0 to 100 as shown in Figure 4 and Figure 5.

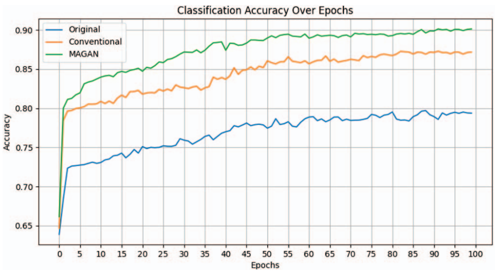

Fig. 4. Validation set performance: comparison of accuracy trends across training epochs for Fashion-MNIST.

These curves offer valuable insights into the convergence and optimization trajectories of the algorithms, highlighting any divergence or convergence trends.

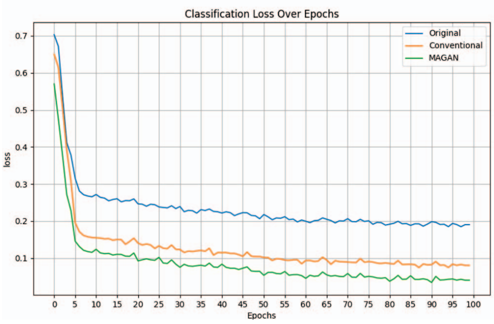

Fig. 5. Validation set performance: comparison of loss trends across training epochs for Fashion-MNIST.

In our assessment, we conducted critical tests, including accuracy, precision, recall, and F1-score, comparing three scenarios: no augmentation, standard approach, and MAGAN. Table II highlights MAGAN's superior performance on Fashion-MNIST.

Summarily, MAGAN outperforms both no augmentation and conventional augmentation. Without augmentation, the model achieves 75.38% accuracy. Conventional augmentation improves accuracy to 84.72%, and MAGAN achieves an impressive 86.89%; an approximately 11.51% improvement over no augmentation and a 2.17% improvement over conventional augmentation. These results underscore MAGAN's substantial effectiveness in enhancing model performance.

TABLE II
COMPARISON OF AUGMENTATION METHODS FOR FASHION-MNIST CLASSIFICATION: TEST ACCURACY, PRECISION, RECALL, AND F1-SCORE ANALYSIS.

Metrics	No Augmentation	Conventional	MAGAN
Test Accuracy (%)	75.38	84.72	86.89
Precision	0.77	0.85	0.87
Recall	0.75	0.85	0.87
F1-score	0.76	0.85	0.87

We used confusion matrices to acquire a more complete understanding of these outcomes. Figures 6, 7, and 8 illustrate these matrices in detail, each corresponding to one of the three scenarios: no augmentation, the traditional way, and the MAGAN approach, respectively. The confusion matrices show that the classifier trained using the MAGAN approach had the highest number of correct predictions. These visual representations provide persuasive proof of MAGAN's superior classification performance when compared to the other two approaches.

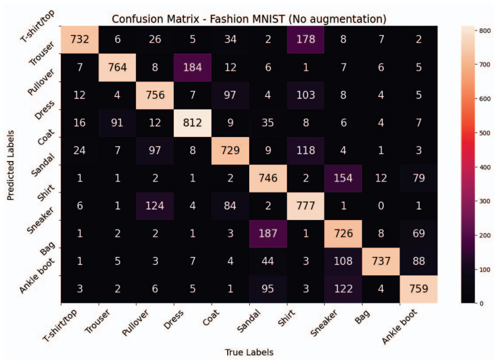

Fig. 6. Confusion matrix for Fashion-MNIST (no augmentation).

D. Binary-Class - cat-vs-dog

The MAGAN algorithm excels in binary classification scenarios.

To address limited data availability, we evaluated MAGAN using a constrained subset of the cat-vs-dog dataset, utilizing only 1,000 images per class for both training and testing.

To comprehensively compare MAGAN with traditional data augmentation, we enriched the subset with an additional 1,000 generated images per class. Figure 9 displays validation accuracy graphs for the original subset, conventional augmentation, and MAGAN throughout the entire training process, providing a clear visual comparison of accuracy scores against training epochs.

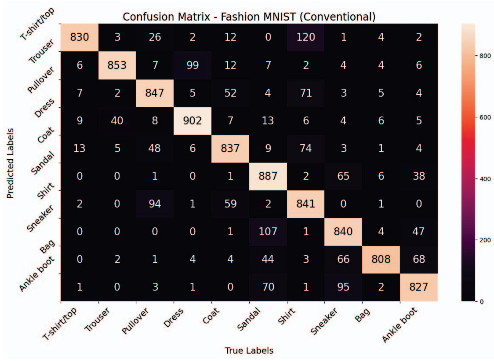

Fig. 7. Confusion matrix for Fashion-MNIST (conventional).

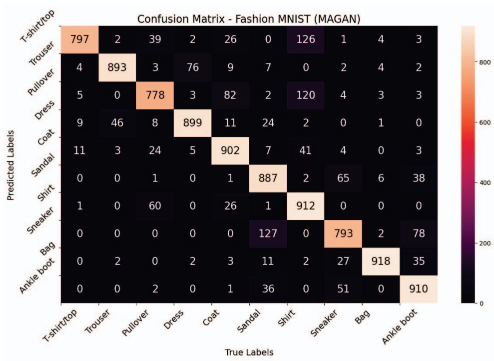

Fig. 8. Confusion matrix for Fashion-MNIST (MAGAN).

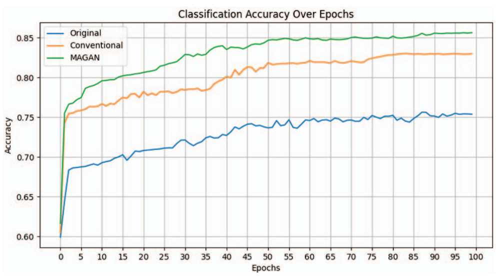

Fig. 9. Validation set performance: comparison of accuracy trends across training epochs for cat-vs-dog.

Figure 10 provides a visual representation of the validation loss patterns for these three methodologies. These loss curves serve as informative indicators of the convergence and optimization trajectories of the algorithms, shedding light on any divergence or convergence trends that may arise.

We conducted essential evaluations, encompassing test accuracy, precision, recall, and F1-score, as integral components of our thorough assessment of model performance. In comparison to two alternative scenarios, namely, no augmentation and the conventional approach, these pivotal metrics provided invaluable insights into the effectiveness of our MAGAN strategy. Table III presents the results of these evaluations,

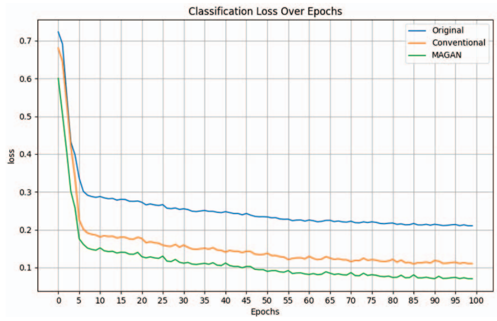

Fig. 10. Validation set performance: comparison of loss trends across training epochs for cat-vs-dog.

showcasing a notable disparity that underscores the superior performance of MAGAN on the cat-vs-dog dataset.

TABLE III
COMPARISON OF AUGMENTATION METHODS FOR CAT-VS-DOG CLASSIFICATION: TEST ACCURACY, PRECISION, RECALL, AND F1-SCORE ANALYSIS.

Metrics	No Augmentation	Conventional	MAGAN
Test Accuracy (%)	71.45	78.20	80.45
Precision	0.72	0.78	0.80
Recall	0.71	0.78	0.80
F1-score	0.71	0.78	0.80

In summary, comparing the three approaches reveals notable differences in test accuracy. Without augmentation, the model achieved 71.45%. Conventional augmentation improved accuracy to 78.20%, while MAGAN demonstrated the most significant enhancement, reaching 80.45%; approximately 2.25% higher than conventional and 9.00% higher than no augmentation. MAGAN effectively enhances model performance compared to both no augmentation and conventional augmentation methods.

For a comprehensive insight, confusion matrices in Figure 11, Figure 12, and Figure 13 highlight MAGAN's superior classification performance with the highest true negatives and true positives, providing compelling visual evidence.

IV. CONCLUSION

This paper introduced and evaluated the MAGAN algorithm, a Meta-Analysis of Generative Adversarial Networks, as a solution for data augmentation in deep neural networks, particularly focusing on binary-class and multi-class scenarios. Leveraging the latent space capabilities of GANs, MAGAN demonstrated its effectiveness in handling binary-class, multi-class, grayscale, and RGB images, showcasing its versatility across a wide spectrum of scenarios. The comprehensive experimental analysis conducted in this study compared the

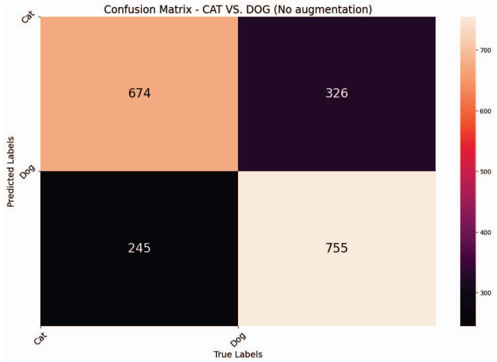

Fig. 11. Confusion matrix for cat-vs-dog (no augmentation).

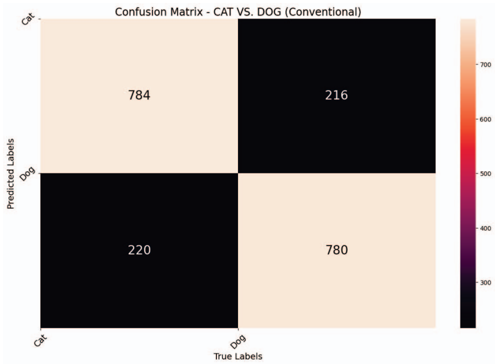

Fig. 12. Confusion matrix for cat-vs-dog (conventional).

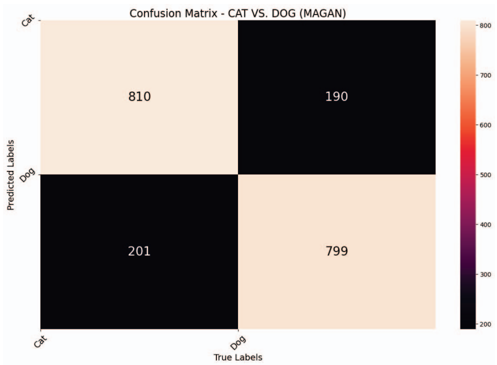

Fig. 13. Confusion matrix for cat-vs-dog (MAGAN).

performance of the proposed MAGAN-based approach with two alternative baseline scenarios, one without any augmentation and another utilizing a conventional augmentation method. The evaluation employed a variety of classification metrics, including accuracy, loss, precision, recall, F1-score, and the confusion matrix. The results consistently revealed that the MAGAN-based approach outperformed the other two scenarios, achieving notable improvements in accuracy by factors of x1.15 and x1.03, respectively. These findings underscore the significant advantages of integrating MAGAN into the data augmentation pipeline for deep learning models. The success of MAGAN in enhancing classification accuracy across various image types and classification tasks highlights its potential as a valuable tool in preventing overfitting and constructing superior deep learning models. The presented results contribute valuable insights to the field of data augmentation, emphasizing the effectiveness of meta-analysis in harnessing the power of GANs for improved generalization and performance in diverse scenarios.

References

[1] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.

[2] R. Rizk, D. Rizk, F. Rizk, and A. Kumar, "A hybrid capsule network-based deep learning framework for deciphering ancient scripts with scarce annotations: A case study on phoenician epigraphy," in *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2021, pp. 617–620.

[3] R. Rizk, D. Rizk, F. Rizk, A. Kumar, and M. Bayoumi, "An efficient capsule network reconfigurable hardware accelerator for deciphering ancient scripts with scarce annotations," in *2021 IEEE 34th International System-on-Chip Conference (SOCC)*, 2021, pp. 75–78.

[4] B. Li, Y. Hou, and W. Che, "Data augmentation approaches in natural language processing: A survey," *AI Open*, vol. 3, pp. 71–90, 2022.

[5] I. Goodfellow, "NIPS 2016 tutorial: Generative adversarial networks," 2017, https://arxiv.org/abs/1701.00160. [Online]. Available: https://arxiv.org/abs/1701.00160

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," 2017, https://arxiv.org/abs/1710.10196. [Online]. Available: https://arxiv.org/abs/1710.10196

[7] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," 2018, https://arxiv.org/abs/1805.08318. [Online]. Available: https://arxiv.org/abs/1805.08318

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

[9] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," 2016. [Online]. Available: https://arxiv.org/abs/1606.03657

[10] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Transactions on Pattern Analysis amp; Machine Intelligence*, vol. 43, no. 12, pp. 4217–4228, dec 2021.

[11] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arxiv.org/abs/1511.06434*, 2015. [Online]. Available: https://arxiv.org/abs/1511.06434

[12] D. Bau, J.-Y. Zhu, H. Strobelt, B. Zhou, J. B. Tenenbaum, W. T. Freeman, and A. Torralba, "On the units of GANs (extended abstract)," 2019. [Online]. Available: https://arxiv.org/abs/1901.09887

[13] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris, "Ganspace: Discovering interpretable gan controls," *Advances in neural information processing systems*, vol. 33, pp. 9841–9850, 2020.

[14] F. Rizk, R. Rizk, D. Rizk, and C.-H. H. Chu, "MAGAN: A meta-analysis for generative adversarial networks' latent space," in *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, 2023, pp. 488–494.