Data Augmentation with Generative Adversarial Networks for CNNs

Qusai Onali[1] and Lilatul Ferdouse[2]

[1]Faculty of Science, Wilfrid Laurier University
[2]Faculty of Science, Wilfrid Laurier University

Keywords:
GANs,
CNNs,
Data Augmentation,
Image Classification

Abstract Under the constraints of a limited dataset, our research project investigates and emphasises the outcomes of generative AI and the CNN model's accuracy in an array of image classification tasks. An enriched dataset incorporating neural style transfer, the original cat and dog images, and their augmented variants using OpenAI's DALL-E are utilised in a comparative analysis alongside three distinct CNN models trained on separate datasets. The outcomes demonstrate the effectiveness of generative AI in augmenting data. The model with the most augmentations attained the highest accuracy on the test set, at 91.86 percent, representing a 4.16 percent advance over the model trained on the original dataset. Innovative AI methods improve CNN performance and generalizability, especially in confined data setting, according to the study.

1. Introduction

As the domain of Artificial Intelligence advances, generative diffusion models like OpenAI's DALL-E, Stability AI's Stable Diffusion and Midjourney are gaining popularity for their capacity to produce anything that mimics the artistic abilities of humans. These models have exceptional proficiency generating computer-generated art that is comparable to the skill of the talented artists, representing a notable progression in AI capabilities. Although the ability of these models to create digital art has been recognised, their capacity to tackle significant obstacles in the domain of machine learning, particularly the scarcity of data, is just starting to be investigated. The goal of our research project is to better understand how these generative diffusion models can be used to augment training data and thus help us to improve the generalizability of advanced neural network models.

One of the most popular approach in the deep learning models to address inadequate training data is data augmentation. To increase the training examples, operations such as scaling, translation, cropping, reflection and rotation can be applied to existing images, which improves the dataset by multifolds. A great way to mitigate the chances of overfitting the model is to augment the training data with synthetic but realistic pictures [1]. Enhanced accuracy and generalisation of deep learning models result from this technique. When considering the convolutional neural networks (CNNs), this becomes especially crucial to us, as these models have difficulty learning rotationally invariant features unless they are supplied with an adequate number of different examples at different rotations in the training data. Based on these factors, our research project considers using a diffusion model. By offering additional synthetic data, this method offers an innovative solution to the challenge of limited training data by allowing for its supplementation.

Finding enough data to model and train the complex convolutional neural networks (CNNs) is one of the primary concerns that our project tries to solve. In image classification, large and distinct datasets are essential for constructing models that can generalise to new data. In spite of this, coming up with datasets that are so extensive and diverse can be difficult and resource-intensive. By leveraging generative artificial intelligence and style transfer techniques, our project aims to augment a limited dataset comprising images of cats and dogs.

The project uses a two-step methodology utilising ML techniques. At the outset, a CNN is utilised in order to initially construct a benchmark model by making use of the initial, limited dataset. The generative diffusion model DALL-E from OpenAI is then used to produce three different versions of every image in the dataset. A second convolutional neural network (CNN) uses these produced images as part of its training data. The next step is to develop a diversified dataset for training a third convolutional neural network (CNN) by applying style transfer techniques to both the original and created images. Thus, Generative AI and style transfer are combined to enhance training data in this method.

2. Our Method

In order to optimise the input for our convolutional neural networks, three diverse datasets- each containing the images of dogs and cats are used. These datasets are then exposed to a number of data preprocessing and feature engineering steps.

2.1 Data Preparation: From Preprocessing to Augmenting

Approximately 610 images in total, equally split between dog and cat images, make up the first dataset. Starting with this dataset, we established a baseline for our data augmentation procedures and evaluated them. Each image in this dataset were obtained from an internet collection of pet images, carefully selected to maintain a diverse representation of breeds for both cats and dogs.

In order to get these images ready for the CNN model, we first resized them to the exact dimensions the model needed, which are 128 x 128 pixels. Additionally, the pixel intensities were re-scaled to a range of 0 to 1 from the typical range of

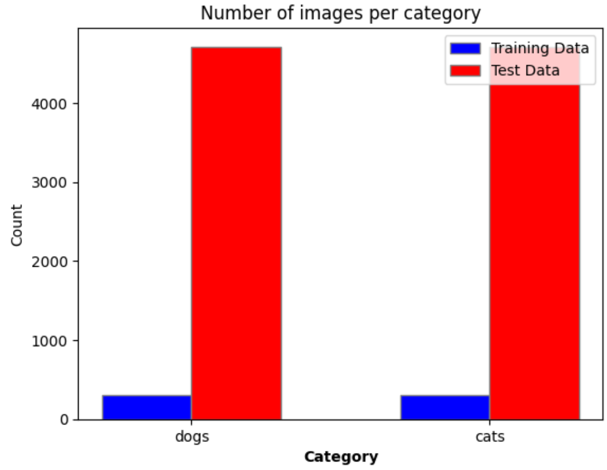


FIGURE 1. Category Wise Distribution of Dataset.

FIGURE 2. Original Dataset Sample Images.

0-255. Resizing the images to the specified dimensions of 128 x 128 pixels and re-scaling pixel intensities to a normalized range of 0 to 1 are imperative preprocessing steps aimed at aligning the dataset with the specific requirements of the convolutional neural network (CNN) model [2]. Multiple transformations, including shear, zoom, and horizontal flip, were applied to the pictures to improve the training data even more. The data was transformed to provide perspective and orientation variations and minimise overfitting during training.

This initial dataset did not have any detectable outliers or missing characteristics that were of considerable importance. During the training phase, the photos were also grouped into sets of 32, and their labels were stored in a binary format, with 0 representing cats and 1 representing dogs by default.

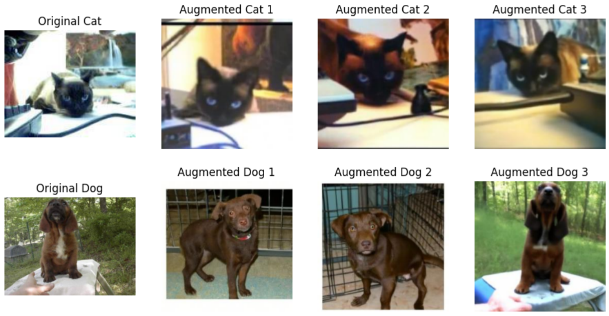


FIGURE 3. Dalle Augmented Dataset Samples

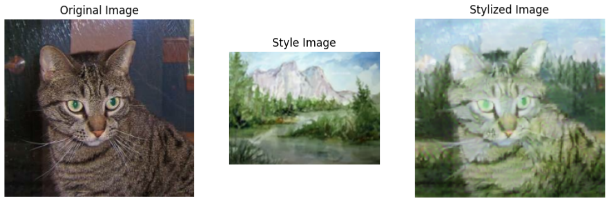


FIGURE 4. Style Transfer Augmentation Example- 1.

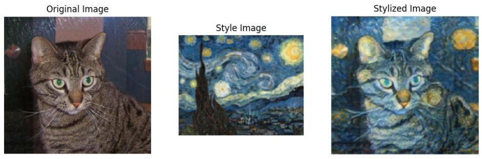


FIGURE 5. Style Transfer Augmentation Example- 2.

By utilising OpenAI's DALL-E to produce three distinct iterations of every image in the initial dataset, our second dataset further develops the first. An enhanced dataset of 1800 photos was produced as a result of this approach, giving our second CNN's training a wider range of visual data. In order to ensure consistency in the data representation, the augmented images generated by DALL-E kept the original labels of their source images. Ensuring the quality and relevance of the generated images was a crucial part of the data preprocessing at this stage.

Images that were unable to faithfully depict a dog or cat were eliminated from the dataset.

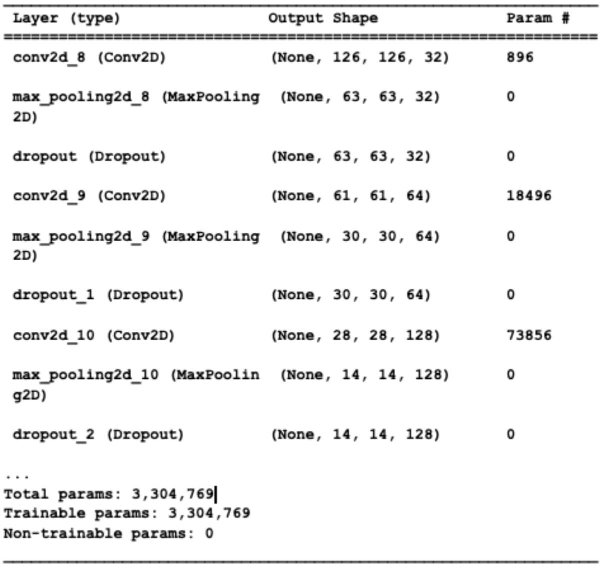
All photos from the original and DALL-E enhanced datasets are included in third dataset, along with style-transferred versions which is the result of our data augmentation techniques. By employing a pre-trained neural style transfer model from TensorFlow, a visually diverse dataset was created to increase the complexity and unpredictability of the training data [3]. As a result, the final dataset includes 5400 photos in total, offering a rich and diverse dataset for our third CNN's training.

2.2 Model Implementation & Training Details

Our study for this project involves the development and comparison of three distinct CNN models trained on different datasets: the original dataset, the DALL-E augmented dataset, and the dataset produced using neural style transfer techniques.

The base-line (initial) CNN model, integral to our study, is constructed upon a foundational dataset comprising approximately 600 meticulously curated images depicting cat and dog subjects. The model's architecture is crafted through the utilization of the Keras library, a high-level neural networks API. This architectural construct unfolds through a stratification of layers, each tailored with precision to fulfill specific computational roles [4].

The initial layer of the model embodies a 2-dimensional convolutional layer, integrating 32 filters of dimensions 3x3 and activated by the Rectified Linear Unit (ReLU). Operating on input images sized at 128x128 pixels with three color channels, this layer is succeeded by a 2x2 max pooling layer,



```
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_8 (Conv2D)           (None, 126, 126, 32)      896

 max_pooling2d_8 (MaxPooling  (None, 63, 63, 32)       0
 2D)

 dropout (Dropout)           (None, 63, 63, 32)        0

 conv2d_9 (Conv2D)           (None, 61, 61, 64)        18496

 max_pooling2d_9 (MaxPooling  (None, 30, 30, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 30, 30, 64)        0

 conv2d_10 (Conv2D)          (None, 28, 28, 128)       73856

 max_pooling2d_10 (MaxPoolin  (None, 14, 14, 128)      0
 g2D)

 dropout_2 (Dropout)         (None, 14, 14, 128)       0

...
Total params: 3,304,769
Trainable params: 3,304,769
Non-trainable params: 0
_________________________________________________________________
```

FIGURE 6. CNN Model Architecture.

condensing spatial dimensions to preserve essential data. Subsequently, a dropout layer nullifies 25% of its input during training to counteract overfitting. This sequence of convolution, max pooling, and dropout iterates twice, escalating the filter count to 64 and 128, enhancing the model's discernment of complex patterns. The output undergoes flattening before entering a Dense layer with 128 neurons, employing the ReLU activation. A subsequent dropout layer (dropout rate: 0.5) addresses overfitting, culminating in a final Dense layer with a single neuron and a sigmoid activation, generating probabilities for the binary classification task.

The preference for ReLU activation is grounded in its computational efficiency and its efficacy in mitigating the vanishing gradient issue. This architectural strategy, characterized by convolution, pooling, and dropout, is adeptly applied to extract hierarchical features and prevent overfitting, optimizing the network's capacity to discern intricate patterns within the dataset.

Post-architecture setup, the model is compiled with the Adam optimizer and binary cross-entropy loss function [5]. Performance evaluation hinges on accuracy as the metric of choice. To bolster generalization and counteract overfitting, data augmentation techniques such as pixel normalization, shear transformations, zooming, and horizontal flips are employed using Keras's ImageDataGenerator. The model undergoes training for 50 epochs on the training data, with validation performed over 2000 steps, each epoch consisting of 18 steps. This meticulous configuration ensures a robust examination of the model's capabilities and generalization prowess over the training duration.

2.3 Experimentation Framework & Code Availability

The code underpinning the experiments detailed in this research is accessible within a designated folder. The repository comprises four principal Jupyter Notebook files, namely data_processing.ipynb, artistic_style_transformations.ipynb, model_training.ipynb, and image_classification_data_analysis.ipynb, each fulfilling a specific role.

The notebook titled data_processing.ipynb functions as a tool for generating variations of original images by utilizing OpenAI's DALL-E model. The code execution in this notebook expands the dataset by incorporating synthetic images, replicating the creative nuances inherent in human artistic endeavors. These generated images are subsequently integrated as supplementary training data for subsequent convolutional neural network models.

The notebook named artistic_style_transformations.ipynb is dedicated to transforming images from original, and DALL-E augmented datasets into diverse artistic styles using neural style transfer techniques. Execution of the code in this notebook results in converting images into various artistic styles, thereby enriching the complexity and variability of the dataset. The resultant style-transferred images are crucial in training the third convolutional neural network.

The pivotal model_training.ipynb notebook is devoted to the training and evaluating convolutional neural network models. The process of CNN training unfolds through the execution of code within this notebook, employing the augmented datasets generated through DALL-E and style transfer techniques. This notebook facilitates a comprehensive comparative analysis of the three CNN models trained on distinct datasets, elucidating the impact of generative AI on image classification tasks.

In the image_classification_data_analysis.ipynb notebook, the script initiates with the installation of necessary libraries and proceeds to load, preprocess, and explore the dataset. It culminates in visualizing sample images and applying image augmentation techniques, such as Data Loading and Preprocessing, Data Exploration, Visualizing Sample Images, Image Augmentation, and Style Transfer. This notebook serves as a valuable resource for those seeking insights into image classification tasks, providing a detailed walkthrough of critical processes in the project.

3. Performance Evaluation & Analysis of CNNs

3.1 Baseline Performance of the Original CNN

The initial CNN, trained solely on the original dataset, demonstrated a commendable accuracy of 0.877%. This baseline score is a vital reference point for estimating the impact of augmented datasets on CNN's robustness.

3.2 Incorporating DALL-E Generated Images

The second CNN, integrating DALL-E generated images, achieved an accuracy of 0.842%. However, this result needs to catch up to the baseline accuracy. Potential explanations include introducing features not well-represented in the original dataset or an increased risk of overfitting due to similarities between DALL-E generated and original images.

3.3 Comprehensive Dataset with Style Transfer Augmentations

The third CNN, trained on a comprehensive dataset with style transfer augmentations, exhibited significant improvement with an accuracy of 0.9186%. Incorporating style transfer enhances the model's generalization ability by introducing a diverse range of visual features.

TABLE 1. Performance Evaluation of CNN Models

CNN Model	Accuracy (%)	Overfitting Potential	MMD Score	PCA Visualization
Original Dataset	0.877	Low	-	-
DALL-E Augmented	0.842	High	0.00148	Different from Original
Style Transfer Augmented	0.9186	Low	0.00215	Similar to Original

3.4 PCA Visualization & Feature Analysis

PCA visualizations of learned features provide deeper insights into the influence of data augmentation techniques on the CNNs' learning processes. The Maximum Mean Discrepancy (MMD) technique reveals slight differences in feature representations among the CNNs, emphasizing the impact of diverse augmentation methods.

3.5 Implications & Considerations

The results underscore the potential benefits of employing advanced generative AI techniques for data augmentation, particularly in scenarios with limited original datasets. However, careful selection and evaluation of synthetic data are crucial, as they significantly influence the model's generalization to real-world data.

This comprehensive section delves into the nuanced performance variations of CNNs, combining quantitative metrics with visualizations to provide a thorough understanding of the impact of data augmentation strategies.

4. Discussions and Future Directions

Moving forward, several compelling avenues exist to enhance the efficacy of our methodology. An intriguing prospect involves exploring advanced diffusion models, such as Midjourney. Diverging from our current neural style transfer technique, Midjourney facilitates the integration of reference images, modifiable through accompanying text prompts. This avenue promises a more controlled, context-aware method for data augmentation. It holds the potential to generate synthetic images that better encapsulate the diversity and intricacies of real-world data, thereby enhancing the generalizability of our trained models.

Secondly, an augmentation to our method could involve incorporating a Convolutional Variational Autoencoder (CVAE) into our pipeline [6]. A CVAE, a generative model, learns a condensed, dense representation of input data in a latent space. Leveraging this learned representation, the model can generate new data. Given its convolutional nature, the CVAE aligns well with image data. Training a CVAE on our datasets could produce additional synthetic images, preserving the fundamental visual characteristics of the original data while introducing novel variations. This could be particularly beneficial in addressing edge cases where the CNN might need more representative data in the training set.

Lastly, exploring diverse architectures and configurations for the CNN offers another avenue for improvement. This includes investigating alternative activation functions, regularization techniques, or advanced models like ResNets or DenseNets. Additionally, experimenting with various loss functions and optimization algorithms can yield enhancements in model performance [7]. With these prospective refinements, we are optimistic about advancing our approach and pushing the boundaries of what generative AI can achieve in data augmentation for image classification tasks.

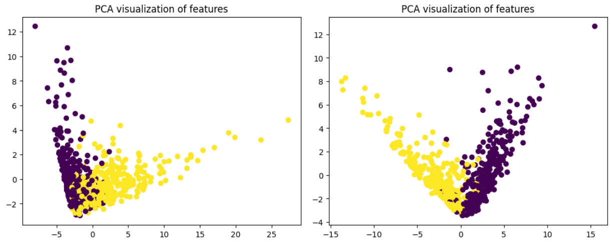


FIGURE 7. PCA Visualization of Features.

5. Conclusion

This project has revealed the efficacy of integrating generative AI techniques into the data augmentation process, leading to notable performance improvements in convolutional neural networks (CNNs) for image classification tasks. The third CNN, trained on a dataset augmented with synthetic images generated by DALL-E and refined through neural style transfer techniques, exhibited superior performance with an accuracy of 90.86% on the test set, surpassing other models.

Our findings underscore the potential benefits of leveraging generative AI models for data augmentation, particularly in data-limited scenarios. However, selecting the most suitable algorithm is contingent on task-specific considerations. Future endeavors could explore advanced diffusion and generative models to enhance performance further. The synergy between generative AI and CNNs presents a promising avenue for improving model robustness and generalizability in image classification tasks.

References

[1] Lei, Shiye, Hao Chen, Sen Zhang, Bo Zhao, and Dacheng Tao: *Image captions are natural prompts for text-to-image models*. arXiv, 2023.

[2] Zhao, Bo and Hakan Bilen: *Synthesizing informative training samples with gan*. arXiv, 2022.

[3] Chen, Ting, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby: *Self-supervised gans via auxiliary rotation loss*. arXiv, 2019.

[4] Ledig, Christian, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi: *Photo-realistic single image super-resolution using a generative adversarial network*. arXiv, 2017.

[5] Lucic, Mario, Karol Kurach, Marcin Michalski, Sylvain Gelly, and Olivier Bousquet: *Are gans created equal? a large-scale study*. arXiv, 2018.

[6] Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen: *Progressive growing of gans for improved quality, stability, and variation*. arXiv, 2018.

[7] Radford, Alec, Luke Metz, and Soumith Chintala: *Unsupervised representation learning with deep convolutional generative adversarial networks*. arXiv, 2016.