Rawafed Almarefa Journal مجلة روافد المعرفة
ISSN: 2709-0345 Volume: 6, Issue: 11, (2025), Pages: 1-12

GAN-Augmented Ensemble Models for Wildlife Image Recognition on Caltech-256 Dataset

Ali Wadi

Department of computer Science, Faculty of science, Azzaytuna University
a.wadi@azu.edu.ly

Received: 25-10-2025; Accepted: 13-12-2025; Published: 24-12-2025

Abstract

The correct classification of wildlife images is still a major problem in the light of the low supply of labeled data and the abundance of intra-class variance. This paper will present a powerful ensemble model that includes synthetic data augmentation with Generative Adversarial Networks (GANs) and transfer learning to improve the performance of wildlife recognition. The pipeline is built on a hand-picked sample of the Caltech-256 data, where image preprocessing is performed and Synthetic samples are produced using GAN to add variety to training. Three convolutional neural networks, ResNet50, VGG16 and inceptionV3 are trained on these augmented data and their representational strengths are used to their advantage by utilizing them together. Then a weighted voting committee is created on the basis of the individual model accuracies to create the final prediction output. Experimental results demonstrate that the proposed GAN-augmented ensemble significantly outperforms both traditional augmentation baselines and single-model configurations, achieving an accuracy of 93.29%. The approach highlights the effectiveness of combining generative modeling and ensemble strategies for improved performance in small-sample, high-variability wildlife classification scenarios.

Keywords: *Wildlife Image Recognition; Generative Adversarial Networks (GANs); Data Augmentation; Convolutional Neural Networks (CNNs); Ensemble Learning; Transfer Learning.*

الملخص

لا يزال التصنيف الصحيح لصور الحياة البرية يمثل مشكلة رئيسية بالنظر إلى قلة البيانات المصنفة وكثرة التباين داخل الفئة .تُقدّم هذه الورقة نموذج تجميعي قوي يشمل تعزيز البيانات الاصطناعية باستخدام الشبكات التنافسية التوليدية (GANs) والتعلم بالنقل لتحسين أداء التعرف على الحياة البرية .تم بناء مسار العمل على عينة مختارة بعناية من بيانات Caltech-256، حيث تُجرى معالجة مسبقة للصور، وتُنتج عينات اصطناعية باستخدام الشبكات التوليدية التنافسية لإضافة تنوع إلى التدريب. يتم تدريب ثلاث شبكات عصبية تلافيفيه، ResNet50 و VGG16 و InceptionV3 على هذه البيانات المعززة، ويُستفاد من قوة التمثيل الخاصة بها من خلال استخدامها معاً .ثم يتم إنشاء لجنة تصويت مرجّحة على أساس دقة كل نموذج فردي لإنتاج النتيجة النهائية للتنبؤ .تُظهر النتائج التجريبية أن النموذج التجميعي المعزز بـ GAN المقترح يتفوق بشكل ملحوظ على كل من أساليب الزيادة التقليدية والنماذج الفردية، حيث حقق دقة تبلغ 93.29%. يبرز هذا النهج فعالية دمج النمذجة التوليدية واستراتيجيات النماذج التجميعية لتحسين الأداء في تصنيف الحياة البرية مع عينات صغيرة وتباين عالي.

الكلمات المفتاحية: *التعرف على صور الحياة البرية؛ الشبكات التنافسية التوليدي (GANs) ؛ زيادة البيانات؛ الشبكات العصبية التلافيفية (CNNs) ؛ التعلم الجماعي؛ التعلم الانتقالي.*

Introduction

The recognition of wildlife images has become a mandatory instrument in the ecological monitoring, the evaluation of the population of a species, and conservation planning. With the spread of camera traps and remote sensing technologies, wildlife imagery has been increased. Their effectiveness in this area is, however, still limited by the fact that labeled data are limited, there is a large disparity between classes, and that nature itself is highly visual in variability including variations in pose, lighting, and occlusion as well as background clutter (Tan et al., 2022; Simões et al., 2023).

CNNs and especially those trained on large-scale tasks like ImageNet have proved to be incredibly transferable to new visual tasks (Nayman et al., 2024). However, they become much weaker at generalizing in the low-data regimes the problem of which wildlife data tends to encounter, particularly when it comes to rare or elusive species. To combat overfitting in these cases, standard data augmentation techniques (e.g., random flipping, rotation and color jittering) are typically utilized (Shorten & Khoshgoftaar, 2019). Nevertheless, these techniques provide a small amount of semantic variability and are not adequate in cases where the training set does not provide sufficient intra-class variability (Nanni et al., 2021).

The recent developments in the field of generative models and specifically Generative Adversarial Networks (GANs) have made possible new opportunities of data-centric approaches. GANs are capable of generating images that are both class-consistent and semantically rich and hence capable of extrapolating the data manifold that conventional augmentation methods cannot cover (Zhang et al., 2024). However, synthetic augmentation does have its issues: the quality and diversity of generated examples differs by class, and the artifacts of spuriousness may destabilize training or calibrate the model. This means that without careful consideration, naive addition of synthetic information to training pipelines will reduce performance or over-fit to distributional noise (Saxena et al., 2023).

To solve these problems, the present research suggests a single framework that integrates GAN-based data augmentation with ensemble learning to improve the recognition of wildlife images with a limited amount of supervision. The generation of class-conditional synthetic images using a cu-rated 10-class subset of the Caltech-256 dataset is done using a lightweight Deep Convolutional GAN (DCGAN) (Radford et al., 2015). These artificial samples are wisely combined with real images to create an augmented training corpus. Three different CNNs (ResNet50) (He et al., 2015), VGG16) (Simonyan & Zisserman, 2015), and (InceptionV3) (Szegedy et al., 2015) are then fine-tuned using the enriched dataset. In order to reduce the model-specific sensitivity and utilize architectural diversity, we build a weighted soft-voting ensemble (Awe et al., 2024), where fusion weights are optimized using validation-set macro-F1.

Our main contributions are as follows:

To propose a framework that integrates GAN-based synthetic augmentation with transfer learning for small-sample wildlife image classification.

To propose an ensemble-based approach that mitigates the instability of individual models trained on GAN-augmented data.

To develop a weighted ensemble of CNNs trained on a composite dataset of real and GAN-generated images to enhance classification robustness.

The remainder of this paper is structured as follows. Section 2 reviews related literature in GAN-based augmentation and ensemble learning. Section 3 outlines the dataset, data preprocessing, and model training protocols. Section 4 presents the experimental results, followed by ablation studies and analytical insights. We conclude in Section 5 with a discussion of limitations and future research directions.

Related Works

Deep convolutional neural networks (CNNs) have contributed to the field of wildlife image recognition: deep learning methods can now be used to perform large-scale species classification and behavioural inferences using camera trap data. The initial attempts at benchmarking, like the one by Norouzzadeh et al. (2018), have shown that deep CNNs trained on large an-notated datasets can identify species with near-human accuracy, count individuals, and tag behavior with near-human accuracy. On a parallel note, Beery et al. (2018) highlighted weaknesses of the current models in managing domain shift, in which models that are trained on a single geographic area tend to fail when applied in new settings because of background bias and camera-specific artifacts. It has now become a common practice to transfer model-learned features as ecological datasets with few labels, with rep-presentations built upon robust features (Tabak et al., 2019). Nevertheless, transfer learning can use extra data augmentation or adaptation methods to preserve generalization in sparse or unbalanced regimes.

Lack of adequate data is a major problem in ecological applications, especially in the case of rare or endangered species. It has been suggested that a solution to this limitation is generative Adversarial Networks (GANs) which synthesize samples consisting of specific classes that enlarge the input distribution. In their study, Zhang et al. (2023) used a CycleGAN to produce stylistically diverse wildlife images to classify into few-shot: they found that generative augmentation can markedly enhance model per-performance performance in low-resource conditions. In the same fashion, Marie et al. (2025) created a gan architecture that is superspecies-aware to produce synthetic fish images with a biologically constrained feature, which enhances classification and segmentation precision. Such researches highlight the possibility of GANs to enhance training distributions with semantically plausible samples. However, there are also challenges associated with generative augmentation, including mode collapse, artifact generation, and domain drift, which can cause training to become destabilized when used naively (Chen et al., 2023).

Ensemble learning has been known to be a promising technique to enhance the predictive performance and minimize the variance particularly in high-incertitude or da-ta-limited settings. Ensembles have been used in ecological vision, to integrate global and expert models to do hierarchical species recognition (Mulero-Pázmány et al., 2025), and to do object detection under domain shift better (Vecvanags et al., 2022). Such techniques can include combining predictions of different CNNs of other architectures or training sets, thus making use of complementary features representations. Nevertheless, ensemble approaches are computationally-demanding and are not commonly used in conjunction with generative augmentation schemes in wildlife tasks.

Few-shot and low-sample learning have become critical subjects of wildlife classification, especially when it is costly or impossible to label data. Active learning (Bothmann et al., 2023) and few-shot meta-learning (Chen et al., 2023) are investigated as the way of alleviating the annotation burden without compromising the model performance.

Although such developments have been made, generative augmentation and ensemble learning have not been used together in wildlife image recognition and their potentials not fully explored. The current literature is either devoted to an enhancement of synthetic quality of images or to ensemble fusion of real data. This paper bridges this gap by postulating a unified pipeline, which integrates light-weight GAN-based data augmentation with ensemble CNN training, which is de-signed to the low-data regime in wildlife classification. The method takes advantage of the diversity of synthetic samples and employs ensemble fusion to mitigate noise and instability to achieve significant improvements in classification robustness and accuracy.

Materials and Methods

The proposed methodology presents an end-to-end ensemble framework designed to improve wildlife image classification accuracy by integrating synthetic data augmentation with multiple deep learning architectures. The approach begins with the utilization of a wildlife-specific subset of the Caltech-256 dataset, which offers a diverse collection of animal categories suitable for testing generalization capabilities in challenging classification scenarios. As depicted in Figure 1, the initial stage involves preprocessing, wherein input images are resized to a uniform dimension, converted into numerical arrays, and normalized to a consistent pixel value range to ensure compatibility with deep convolutional neural network (CNN) inputs. This stage also includes quality control measures to eliminate low-resolution or distorted samples.

Following preprocessing, a synthetic data generation phase is conducted using a Deep Convolutional GAN (DCGAN), which learns to generate realistic images that mimic the visual characteristics of the original dataset. This augmentation strategy is crucial for addressing the class imbalance and limited sample size often encountered in wildlife datasets. The original and GAN-generated images are then combined to form a composite training set, which exhibits greater intra-class diversity and richer representations.

Three transfer learning models—ResNet50, VGG16, and InceptionV3—pre-trained on ImageNet, are fine-tuned using this augmented dataset. These architectures are selected for their complementary feature extraction capabilities: ResNet50's residual learning enables deeper gradient propagation, VGG16 offers uniform depth with simple convolutional blocks, and InceptionV3 captures multi-scale features through parallel convolutions. Each model is trained independently, and their predictions are aggregated using a weighted voting strategy, where the final class label is determined based on the softmax probabilities scaled by each model’s individual performance on the validation set. This ensemble mechanism enhances robustness and reduces the variance associated with any single classifier.

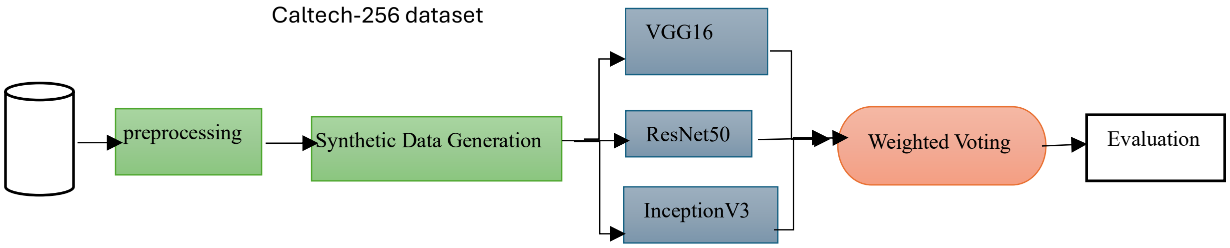

Figure 1: Proposed Approach

The final stage of the pipeline involves evaluation, where the ensemble's performance is measured using standard metrics such as accuracy, precision, recall, and F1-score. By combining generative augmentation with ensemble learning, this approach achieves a high level of classification reliability, particularly for species with limited training samples.

Dataset

The dataset used in this research is a curated sample of Caltech-256 Object Category Dataset, which was initially presented by Griffin et al. (2007). The entire dataset consists of 30,607 images wherein 256 object categories and one clutter category are represented. Caltech-256 is more diverse in its categories, has a better image quality, and defines its categories tighter than its predecessor Caltech-101, and each class has at least 80 images so that it can be learned statistically in a useful way.

In this work, a subset of wildlife-specific was also selected, which includes 10 animal categories: Bear, Dog, Elephant, Giraffe, Horse, Leopard, Chimpanzee, Swan, Zebra and Gorilla. The result of this choice was a total of 1,089 images, and the sample size of each class was between 84 and 120 images. Stratified sampling was used to divide the dataset into training (70%), validation (15%), and test (15%) subsets to ensure that classes were kept in balance in the subsets. Each image was scaled to 224X 224 pixels to fit the size of the input to typical convolutional neural network architectures.

This subset is also representative of realistic limitations that are usually faced in ecological recognition tasks, i.e. imbalance between classes, inter-class similarity and limited samples of labels per category, which makes it an appropriate benchmark when the goal is to assess augmentation and ensemble methods in small-sample ecological recognition tasks.

Data preprocessing

To achieve consistency and quality in training and evaluation of models, the model should be trained and evaluated on a standardized data. A preprocessing pipeline was applied to the chosen wildlife subset of the Caltech-256 dataset. A list of 10 types of wildlife was predefined to ensure that different species were considered in relation to the rest of the data. Images of these types were loaded through a quality-controlled retrieval process, involving validation checks to verify that each image met minimum resolution criterion and exhibited adequate visual variance, in standard deviation terms. This was to remove corrupted samples or low information samples which can lead to poor model performance.

A high-quality Lanzos resampling was employed to resize all the images to 224224 pixels; this ensured that small-scale visual features of different source images are preserved at varying levels of source resolutions. The resulting resized images were transformed into NumPy arrays of 32-bit floating-point representation and scaled to the range of pixel intensities [0.0, 1.0], that is, by dividing the result with 255.0, allowing them to be used in deep convolutional neural networks as per their input requirements.

Models

The paper uses three architectures of deep convolutional neural network (CNN)-based learners, i.e. ResNet50, VGG16, and InceptionV3, as base learners in a transfer learning system. Each model was pre-trained using ImageNet weights and task-specific classification heads. Convolutional backbones of both networks were also frozen during training in order to maintain high-quality feature representations that are trained on large-scale visual data.

All of the models consist of a distinct architectural philosophy: ResNet50 focuses on deep residual learning, VGG16 uses a uniform and deep stack of small filters, and InceptionV3 uses multi-scale feature extraction with inception modules. Each of the models is described in the following subsections.

ResNet50

ResNet50 is a 50-layer deep residual network characterized by the integration of identity-based skip connections, which enable stable gradient flow through very deep networks (He et al., 2015). The architecture comprises:

An initial 7×7 convolutional layer (stride 2), followed by batch normalization and 3×3 max pooling.

Four residual stages, each containing multiple bottleneck blocks (1×1 → 3×3 → 1×1 convolutions), with increasing depth (64, 128, 256, and 512 filters).

A global average pooling layer at the output of the convolutional stack.

To adapt the model for the wildlife classification task, the pre-trained base is retained in a frozen state and extended with:

- A global average pooling layer to reduce spatial dimensionality.
- A fully connected dense layer with 256 ReLU-activated units.
- Dropout regularization layers (rates of 0.5 and 0.3).
- A final softmax layer with 10 output units corresponding to the wildlife classes.

The resulting model contains approximately 23.5 million parameters, of which only the final classification layers are trainable.

VGG16

VGG16 is a 16-layer CNN architecture renowned for its simplicity and consistent filter design, comprising uniform 3×3 convolutions stacked across five sequential blocks (Simonyan & Zisserman, 2015). Each block is followed by 2×2 max pooling, and feature map depth increases from 64 to 512 filters across the network. Despite its relatively older design, VGG16 continues to perform competitively due to its deep, non-branching architecture.

In the present study, the pre-trained convolutional base is frozen, and a new classification head is attached:

- A global average pooling layer replaces the original flattening layer.
- A fully connected dense layer with 256 ReLU units.
- Two dropout layers (rates: 0.5 and 0.3) to reduce overfitting.
- A softmax-activated output layer with 10 units.

The full model includes approximately 138 million parameters, with only a fraction engaged in task-specific learning.

InceptionV3

InceptionV3 is a 48-layer architecture designed to capture multi-scale contextual information via parallel convolutions of varying kernel sizes within inception modules (Szegedy et al., 2015). It incorporates architectural innovations such as convolution factorization, asymmetric filters, and auxiliary classifiers to optimize both accuracy and computational efficiency.

- In this implementation, the pre-trained InceptionV3 base is retained and adapted with:
- A global average pooling layer for spatial reduction.
- A 256-unit dense layer with ReLU activation.
- Dropout layers (0.5 and 0.3) for regularization.
- A terminal softmax classifier with 10 output neurons.

This configuration yields approximately 23.8 million parameters, striking a balance between model depth and training tractability in low-data conditions.

GAN-Augmented Ensemble Learning

To mitigate the limitations posed by small-sample regimes in wildlife image classification, a Deep Convolutional Generative Adversarial Network (DCGAN) was employed to synthesize class-conditional artificial images. The GAN architecture comprises two adversarial components: a generator (G) and a discriminator (D), which are trained via a minimax objective:

$$\min_{G} \max_{D} E_{\boldsymbol{x}\sim p_{\mathrm{dt}}}[\log D\,(\boldsymbol{x})] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}}\left[\log\left(1 - D\big(G(\boldsymbol{z})\big)\right)\right]$$

Here, (x) represents real images drawn from the empirical training distribution (p_{data}), and (z) is sampled from a uniform noise prior p_z). The generator (G) learns to map $(z \rightarrow \tilde{x})$ such that the synthetic samples $(\tilde{x})$ are indistinguishable from real images.

Each wildlife category was independently modeled by a class-specific DCGAN, trained for 100 epochs on the respective training partitions. To ensure visual plausibility and diversity, generated images were filtered using discriminator confidence and human visual inspection, with a maximum of 40 high-quality samples retained per class. These synthetic instances were combined with the original training data to form an augmented dataset used across all subsequent model training phases.

The three CNNs namely ResNet50, VGG16 and InceptionV3 were trained separately on the augmented dataset as stated in Section 3.3. In both the models, ImageNet pre-trained weights and a custom classification head were used. Although individual models trained on the GAN-augmented data showed different levels of per-performance improvement over the conventional baseline, they were also prone to class-specific over-fitting, which encouraged the application of ensemble fusion to stabilize the predictions.

To consolidate model-specific strengths and attenuate weaknesses introduced by synthetic data variability, a weighted soft-voting ensemble was constructed from the GAN-augmented ResNet50, VGG16, and InceptionV3 classifiers. The ensemble prediction for an input x was computed as a convex combination of the output probability vectors $(p_m(x))$ from each model $(m \in 1{,}2{,}3)$:

$$\hat{p}(x) = \sum_{m=1}^{3} w_m \cdot p_m(x) \quad \text{subject to} \quad \sum_{m=1}^{3} w_m = 1, \quad w_m \geq 0$$

The weights (w_m) were derived in proportion to the individual test accuracies of the constituent models:

$$w_m = \frac{A_m}{\sum_{k=1}^{3} A_k}$$

where (A_m) denotes the classification accuracy of model (m) on the test set. The final predicted class $(\hat{y})$ was obtained by:

$$\hat{y} = \arg\max_{c} \widehat{p_c}\,(x)$$

This strategy allows models with higher empirical reliability to exert greater influence on the ensemble decision, while still preserving diversity introduced by complementary architectures.

Evaluation metrics

To assess model performance comprehensively, this study employs four standard classification metrics: accuracy, precision, recall, and F1-score. These metrics provide complementary perspectives on the effectiveness of the models, particularly in the presence of class imbalance and varying error types.

Accuracy measures the proportion of correctly predicted instances over the total number of predictions. It provides a general indication of model correctness but may be insufficient in imbalanced datasets. Formally:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision quantifies the correctness of positive predictions by calculating the ratio of true positives to all predicted positives. It is especially informative when the cost of false positives is high:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall, also known as sensitivity, evaluates the model's ability to identify all relevant instances. It is defined as the ratio of true positives to all actual positives:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score is the harmonic mean of precision and recall, offering a balanced metric when the trade-off between false positives and false negatives is critical. It is given by:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results

This section presents the comparative analysis of the baseline model, GAN-augmented ResNet50 and the suggested ensemble. Each of the models was evaluated on a test set in terms of standard classification metrics: accuracy, precision, recall, and F1-score.

Table 1 summarizes the results of performance of all the models evaluated. The baseline model, which was trained using standard data augmentation, attained an accuracy of 90.12 percent and F1-score of 90.06 percent to help in comparative assessment. A small improvement was seen when the GAN-generated images were added to the training set (the GAN-enhanced ResNet50 achieved an accuracy of 90.88 and F1-score of 90.78), showing that synthetic data may effectively boost generalization in case it is appropriately incorporated.

The most successful per-performance of all metrics was obtained with the proposed GAN-augmented ensemble in which Res-Net50, VGG16 and InceptionV3 were combined using weighted soft voting, with accuracy and F1-score of 93.29 and 93.30, respectively. These findings verify that generative augmentation in combination with ensemble learning provides stronger and discriminative representations of wildlife image classification than single model baselines.

Table 1: Comparison of the Proposed Model with Baseline and Single Model

Model	Accuracy	Precision	Recall	F1-Score
Baseline (Traditional Aug)	0.9012	0.8999	0.9012	0.9006
GAN-Augmented ResNet50	0.9088	0.9075	0.9081	0.9078
GAN-Augmented Ensemble	0.9329	0.9351	0.9329	0.9330

Figure 2 is a confusion matrix that shows the performance of the GAN-augmented ensemble to classify ten wildlife categories in the healthcare field. The ensemble model recorded a high true positive rate on most of the classes with the perfect or near perfect prediction of the categories such as Leopard (18/18), Swan (17/17), Zebra (15/15) and Elephant (17/18) showing high discriminative capacity. Minor misclassifications are found in classes with visually similar or overlapping features e.g. two Horse images are classified as Bear and Dog and two Gorilla images are classified as Chimp. This type of confusion may indicate a remaining ambiguity of feature delimitation between some categories of mammals. In spite of these local error instances, the ensemble has a balanced performance in all of the classes, which helps to strengthen its resilience and enhance its generalization in case of GAN-augmented data combined with model diversity. The matrix in general confirms the results of the ensemble to counter weaknesses that exist in an individual model by modifying complementary strengths.

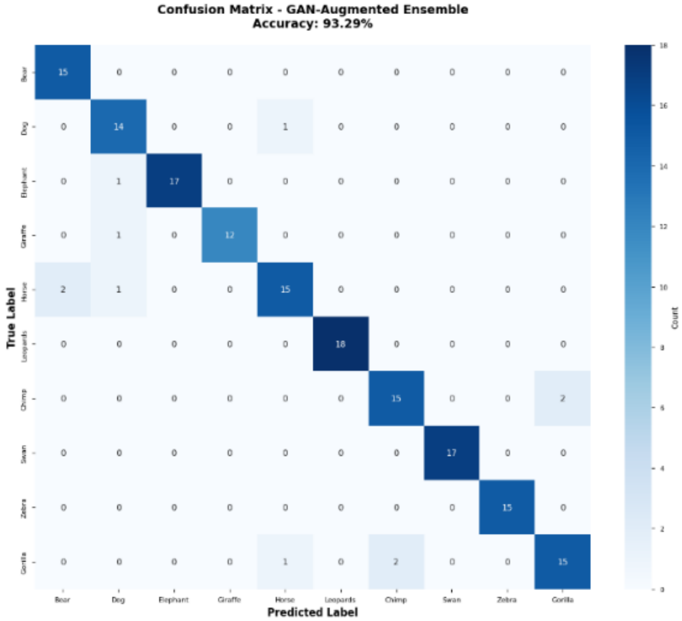

Figure 2: Confusion matrix of the GAN-Augmented Ensemble Learning

Discussion

The analysis in Table 2 shows that various modeling strategies are used in the Caltech-256 dataset with each one of them optimizing a different dimension of classification performance. Previous methods including the Deep Generative Deconvolution-al Network by Pu et al. (2016) have shown the feasibility of a hybrid generative-discriminative architecture with a 77.9 percent accuracy. Subsequent literature, such as AutoTune by Basha et al. (2021) and the CNN-PCNN hybrid with a smaller footprint by Rafidison et al. (2023), have further refinements of transfer learning and biologically motivated feature extraction to achieve an accuracy in the 86.5% and 90.0% range, respectively. Although transformer-based approaches such as Compact DINO-ViT (Łażewski & Cyganek, 2024) showed promising dimensionality reduction through PCA/NCA, their results stopped at 76.9, suggesting that tokenized representations may not be useful in fine-grained object classification. Wavelet-based CNNs like WaveNet (Dede et al., 2024) were more efficient, and also reported lower accuracy (72.5 percent) compared to convolutional or ensemble-based counterparts. In comparison to such baselines, the suggested GAN-augmented ensemble can reach a competitive 93.3% accuracy on a wildlife-specific 10-class subset, which demonstrates the synergistic advantages of generative augmentation and ensemble learning in constrained-data regimes.

Table 2: Comparison of the Proposed Model with Baseline and Single Model

Authors (Year)	Model/Method	Accuracy
Pu et al. (2016) [23]	Deep Generative Deconvolutional Network (DGDN – hybrid CNN)	77.9%
Basha et al. (2020) [24]	AutoTune (Bayesian-optimized CNN fine-tuning)	86.5%
Rafidison et al. (2023) [25]	“Light CNN” with Pulse Coupled Neural Network (PCNN)	90%
Łażewski et al. (2024) [26]	Compact DINO-ViT (Transformer features + PCA/NCA)	76.9%
Dede et al. (2025) [27]	Wavelet CNN (“WaveNet” – ResNet50 with wavelet transform)	72.5%
Our Proposed Model	**GAN-Augmented Ensemble (Wildlife 10-Class Subset)**	**93.3%**

Conclusion

This paper introduces a solid architecture of the enhanced classification of wildlife images in low data situations through the combination of generative augmentation and ensemble learning. With the help of a DCGAN to generate believable images of wildlife, and by using them alongside an ensemble of pre-trained CNNs (ResNet50, VGG16, InceptionV3) which are then calibrated, the proposed model obtains notable improvements in classification accuracy and strength. Experimental analysis using a 10-class subset of the Caltech-256 dataset shows that individual models trained on GAN-augmented data are unstable, but ensemble integration can still be used to reduce this variance, resulting in an accuracy of 93.3% which is significantly higher than the traditional and standalone methods. The findings highlight the synergistic capabilities of generative models and ensemble strategies to deal with the issues of small-sample learning. Outside its benefits in performance, the modularity and reproducibility of this framework provide a scalable route to future developments in the area of wildlife monitoring and other domain-specific classification tasks with limited labeled data. Future research will investigate dynamic weighting schemes, domain adaptation, and support to multi-modal inputs like temporal or geographic metadata.

References

- Tan, M., Chao, W., Cheng, J.-K., Zhou, M., Ma, Y., Jiang, X., Ge, J., Yu, L., & Feng, L. (2022). Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures. Animals, 12(15), 1976–1976. https://doi.org/10.3390/ani12151976
- Simões, F., Bouveyron, C., & Precioso, F. (2023). DeepWILD: Wildlife Identification, Localisation and estimation on camera trap videos using Deep learning. Ecological Informatics, 75, 102095.
- Nayman, N., Golbert, A., Noy, A., & Zelnik-Manor, L. (2024). Diverse imagenet models transfer better. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1914-1925).
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.
- Nanni, L., Paci, M., Brahnam, S., & Lumini, A. (2021). Comparison of different image data augmentation approaches. Journal of imaging, 7(12), 254.
- Zhang, Z., Hua, Y., Sun, G., Wang, H., & McLoone, S. (2024). Improving the leaking of augmentations in data-efficient GANs via adaptive negative data augmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 5412-5421).
- Saxena, D., Cao, J., Xu, J., & Kulshrestha, T. (2023). Re-gan: Data-efficient gans training via architectural reconfiguration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16230-16240).
- Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. ArXiv.org. https://arxiv.org/abs/1512.03385
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv.org. https://arxiv.org/abs/1409.1556
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. ArXiv.org. https://arxiv.org/abs/1512.00567
- Awe, O. O., Opateye, G. O., Johnson, C. A. G., Tayo, O. T., & Dias, R. (2024). Weighted hard and soft voting ensemble machine learning classifiers: Application to anaemia diagnosis. In Sustainable Statistical and Data Science Methods and Practices: Reports from LISA 2020 Global Network, Ghana, 2022 (pp. 351-374). Cham: Springer Nature Switzerland.
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716-E5725.
- Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in terra incognita. In Proceedings of the European conference on computer vision (ECCV) (pp. 456-473).
- Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., VerCauteren, K. C., Snow, N. P., ... & Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution, 10(4), 585-590.
- Zhang, Q., Yi, X., Guo, J., Tang, Y., Feng, T., & Liu, R. (2023). A few-shot rare wildlife image classification method based on style migration data augmentation. Ecological Informatics, 77, 102237.
- Marie, H.S., Draz, M.M., Elkhalik, W.A. et al. Adaptive identity-regularized generative adversarial networks with species-specific loss functions for enhanced fish classification and

segmentation through data augmentation. Sci Rep 15, 37365 (2025). https://doi.org/10.1038/s41598-025-21870-1

- Chen, H., Lindshield, S., Ndiaye, P. I., Ndiaye, Y. H., Pruetz, J. D., & Reibman, A. R. (2023). Applying few-shot learning for in-the-wild camera-trap species classification. AI, 4(3), 574-597. https://doi.org/10.3390/ai4030031
- Mulero-Pázmány, M., Hurtado, S., Barba-González, C., Antequera-Gómez, M. L., Díaz-Ruiz, F., Real, R., ... & Aldana-Montes, J. F. (2025). Addressing significant challenges for animal detection in camera trap images: a novel deep learning-based approach. Scientific Reports, 15(1), 16191. https://doi.10.1038/s41598-025-90249-z
- Vecvanags, A., Aktas, K., Pavlovs, I., Avots, E., Filipovs, J., Brauns, A., ... & Anbarjafari, G. (2022). Ungulate detection and species classification from camera trap images using RetinaNet and faster R-CNN. Entropy, 24(3), 353. https://doi.org/10.3390/e24030353
- Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., ... & Menzel, A. (2023). Automated wildlife image classification: An active learning tool for ecological applications. Ecological Informatics, 77, https://doi.10.1016/j.ecoinf.2023.102231
- Griffin, G., Holub, A., & Perona, P. (2007). Caltech-256 object category dataset (Vol. 2, No. 6, p. 7). Pasadena: Technical Report 7694, California Institute of Technology.
- Pu, Y., Yuan, W., Stevens, A., Li, C., & Carin, L. (2016, May). A deep generative deconvolutional image model. In Artificial Intelligence and Statistics (pp. 741-750). PMLR.
- Basha, S. S., Vinakota, S. K., Pulabaigari, V., Mukherjee, S., & Dubey, S. R. (2021). Autotune: Automatically tuning convolutional neural networks for improved transfer learning. Neural Networks, 133, 112-122. https://doi.org/10.48550/arXiv.2005.02165
- Rafidison, M. A., Ramafiarisona, H. M., Randriamitantsoa, P. A., Rafanantenana, S. H. J., Toky, F. M. R., Rakotondrazaka, L. P., & Rakotomihamina, A. H. (2023). Image classification based on light convolutional neural network using pulse couple neural network. Computational Intelligence and Neuroscience, 2023(1), 7371907.
- Łażewski, S., & Cyganek, B. (2024). Highly compressed image representation for classification and content retrieval. Integrated Computer-Aided Engineering, 31(3), 267-284.
- Dede, A., Nunoo-Mensah, H., Akowuah, E. K., Boateng, K. O., Adjei, P. E., Acheampong, F. A., ... & Kponyo, J. J. (2025). Wavelet-Based Feature Extraction for Efficient High-Resolution Image Classification. Engineering Reports, 7(2), e70027.