Journal of Applied Research and Technology

www.jart.icat.unam.mx

Journal of Applied Research and Technology 23 (2025) 275-290

Original

Design of an improved model for natural image classification using AugMix, SE-ResNeXt, and MAML

M. A. Deshmukh* • A. K. Gaikwad

CSE Department, G. H. Raisoni University Amravati (Maharashtra), India

Received 06 21 2024; accepted 10 30 2024
Available 06 30 2025

Abstract: Retaining the effectiveness but improving the efficiency of natural image classification is of prime necessity in recent times, with the surge in demand for deploying these models in practical applications, ensuring accuracy and generalization. Classic deep learning classifiers suffer from limited robustness, generalization, and failure to adapt to new tasks and domains. These shortcomings restrict their practically effective deployment by the availability of different diversified and unseen data. In this work, the authors introduce an optimized deep learning classifier framework, leveraging state-of-the-art techniques in various key domains. The proposed model harnesses a combination of techniques ranging from AugMix, SE-ResNeXt, MAML, Hyperband, and finally Domain-Adversarial Neural Network (DANN) for performance improvement. AugMix integrates Mixup and CutMix with the stochastic augmentation technique of complex augmentation chains to enhance the model's robustness and generalization. Mixing images with stochastic augmentations and the use of Mixup and CutMix bring further strong regularizations, boosting the robustness metrics by 15-20% and classification accuracy by 3-5% on the unseen natural images and samples. SE-ResNeXt introduces the use of channel-wise attention to enhance the representational power of the model. Squeeze-and-Excitation (SE) blocks are introduced to recalibrate the channel-wise feature responses by weighting informative features and suppressing less useful ones. It boosts the accuracy of models on benchmark CIFAR-100 dataset samples by 2-3% over standard ResNeXt. Execution of Model-Agnostic Meta-Learning enables a model to adapt quickly to a new task based on a small number of examples. MAML meta-learns updated models based on examples of tasks instead of direct model parameters. A 5-7% improvement in accuracy is achieved for different scenarios. Hyperband performs tension-free search of optimal hyperparameters via adaptive resources dealing, which configures the resources only for the promising configurations. Reducing the computational cost of hyperparameter tuning to at most 50% ensures an increase in model accuracy of 2-3%. The DANN technique uses adversarial training in order to suppress the domain shift between source and target datasets. DANN uses a gradient reversal layer to train feature extractors to produce domain-invariant features, leading to a 10~15% increase in accuracy on target domain datasets compared to non-adaptive methods.

Keywords: Natural image classification, data augmentation, attention mechanisms, meta-learning, domain adaptation.

*Corresponding author.
E-mail address: meghnadeshmukh9@gmail.com (M. A. Deshmukh).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

Classification of natural images has continued to be at the forefront among computer vision research efforts. Concomitant with the increased demand for more accurate and generalizable models in real-world applications, the current scenario calls for robust and efficient natural image classification approaches. Robustness, generalization, and adaptability to new tasks and domains are some aspects where conventional deep learning classifiers may often lack. All of these factors serve to constrain their practical effectiveness when met with diverse and unseen data. A proactive solution is proposed in this work to design an efficient deep learning classification framework fitted with the latest advancements in several key areas. Therefore, the developed model harnesses AugMix, SE-ResNeXt, MAML, Hyperband, and DANN to enhance performance metrics significantly. AugMix applies Mixup and CutMix strategies on images with the help of stochastic augmentations. By mixing images using stochastic augmentations and by applying techniques such as Mixup and CutMix, AugMix leads to strong regularization. This results in a 15-20% improvement in robustness metrics and an increase of 3-5% in classification accuracy across many classes of unseen natural images and samples. SE-ResNeXt adds channel-wise attention to further enhance the representational power of the model. SE block adds a channel-wise attention and recalibrates channel-wise feature responses to highlight informative features and suppress less useful features. This leads to an improvement in performance of 2-3% accuracy over standard ResNeXt on benchmark CIFAR-100 dataset samples. The application of Model-Agnostic Meta-Learning (MAML) allows the model to adapt to a new task quickly, with a small number of training examples. MAML learns an effective update rule across tasks, using the provided data to update a model's initial parameterization such that it fine-tunes quickly his. This is achieved by Meta-objective which has updated remember What internal states in RNN to be predictive on a new task quickly. The approach attains state-of-the-art few-shot learning performance, with a 5–7% accuracy improvement for diverse scenarios. Thus, although deep learning models are powerful, they struggle with generalization to the unseen data and adaptability to new tasks and domains. These restrictions put a barrier on the practical deployment of these models in dynamic and diverse environments.

The most used approaches in existing deep learning for natural image classification depend on large-scale labeled datasets and complex model architectures. While the models provide a high level of accuracy across varied benchmark datasets, they are sensitive to variations, corruptions, or shifts in the data distribution. The key bottleneck lies in the inability to generalize well beyond the training distribution in scenarios that mirror real-world decision-making. In addition, increased computational costs with custom-trained models are a further barrier, especially in resource-constrained environments [1, 2, 3]. This paper proposes a unified framework with a confluence of different advanced approaches in data augmentation, attention mechanism, meta-learning, hyper-parameter optimization, and domain adaptation. The model proposed in this work adopts AugMix as a unique feature that enhances its robustness and generalization. Specifically, AugMix applies Mixup and CutMix at the augmentation level to improve the robustness of the model across data distributions and other perturbations.

The embedding of SE-ResNeXt introduces the use of channel-wise attention provided through recalibration of the features' response to produce better informative features than less useful features. The model's representational power is significantly improved, resulting in higher accuracy and better feature extraction. Besides, the embedding of Model-Agnostic Meta-Learning (MAML) allows fast adaptation of the classifier to new tasks with a minimal amount of additional training. Optimizing the parameters of the initial model to learn efficiently on the new tasks, MAML significantly trains the model in a few-shot setup to enhance the versatility of the model in various scenarios. Independently, also the domain shift capability is confronted, since DANN encourages the learning, on the source domain, of features that are domain-invariant and is able to perform very effectively with examples from different domains, thus achieving results in a better way in different domains based on natural image datasets as well as samples. This proposed integration of AugMix, SE-ResNeXt, MAML, Hyperband, and DANN clearly indicates the vast improvement of performance than traditional classifiers, thus marking a new benchmark for natural image classification. The integration of SE-ResNeXt, MAML, Hyperband, and DANN introduces great robustness, generalizability, and adaptability to the classifier, and thus greatly surpasses results obtained by standard classifiers. This potentiality of advanced deep learning methodologies is demonstrated as being able to overcome the limits of previous methodologies and assure the best state-of-the-art performance on the natural image classification process. Hyperband efficiently searches for the best hyperparameters by dynamically allocating resources to promising configurations. This method reduces the computational cost of hyperparameter tuning by 50% compared to traditional methods, while improving model accuracy by 2-3%. The Domain-Adversarial Neural Network (DANN) technique minimizes domain shifts between source and target datasets using adversarial training. DANN employs a gradient reversal layer to train feature extractors to produce domain-invariant features, leading to a 10-15% increase in accuracy on target domain datasets compared to non-adaptive methods.

Motivation and contribution

This research is motivated by the persistent constraints and limitations current deep learning models have in the classification of natural images. Traditional classifiers frequently suffer a drastic degradation in performance when the input to be classified has short variations due to noise, lighting changes, or domain shifts. Such limitations compromise not only the reliability of those models in realistic use cases but also their wide applicability in many fields in need of robust visual recognition systems. Besides, the excessive computation resources required for training deep-learning models, including the usually exhaustive processes of hyperparameter tuning, present a major bottleneck, especially in resource-constrained environments. These are challenges that have created a huge demand for innovative techniques that can boost the robustness, generalization, and adaptability of forceful natural image classifiers while maintaining computational efficiency.

In response to these challenges, this paper presents a wide, deep learning framework ensembling the state of the art in its multiple aspects. The main contributions of this research entail development and implementation of an optimized model that adapts techniques such as AugMix for better data augmentation, SE-ResNeXt for more effective attention mechanisms, MAML for meta-learning, Hyperband for optimal hyperparameter optimization, and DANN for domain adaptation. AugMix, on the other hand, helps to fill this gap required for robust data augmentation by mixing the images with stochastic augmentations and using some of the Mixup methods and CutMix. We show that our approach—meaningful augmentation with data in medicine—can achieve the really much-needed generalization to the unseen data, improving the robustness by 15-20% and raising the classification accuracy by 3-5%. In addition, we introduce SE-ResNeXt, a channel-wise attention mechanism that re-scales feature responses for increased representational power within the model, improving accuracy compared to standard ResNeXt architectures by 2-3%.

In addition, the inclusion of MAML to facilitate quick adaptation to new tasks from a smaller training sample further supports the need for flexible models that can adequately serve in few-shot learning. This technique fine-tunes the initial parameters of the model to achieve rapid and effective learning, as evidenced by the test accuracy, indicating increases of 5-7%, even for such challenging datasets as Mini-ImageNet. The reinforcement of the proposed framework is also found with the cooperative employment of Hyperband since it provides an effective approach to hyperparameter optimization by dynamically allocating resources to the most promising configurations, reducing the search time up to 50%, and at the same time, increasing the model accuracy by 2-3%. Finally, DANN assures proper domain adaptation by the model, making the latter generalize correctly across different domains through the learning of invariant features for the domain; this shows an increase in the target domain dataset accuracy by 10-15% over the target domain dataset with respect to non-adaptive models.

The all-round amalgamation of such high-end methodologies sets new benchmarks for the classification of natural images, answering the crying need for models to be accurate yet robust, adaptable, and computationally efficient. This paper presents the feasibility and effectiveness of combining state-of-the-art techniques for improved robustness in the classifier and generalization capabilities in the field of natural image classification, thereby paving the way for more reliable and versatile applications in different real-world scenarios.

2. Literature review

Its landscape has been considerably moved, thanks to the confluence of advanced methodologies aiming to provide the deep neural structure more robust, generalized, and adaptable. Several works used recent approaches. By checking through table 1, the state-of-the-art methodologies, strengths/uses, and limits are gained for different scenarios. Neural architecture search (NAS) with multi-modal imaging, as explored by Xiao et al. (2022), has shown promise in intraoperative glioma grading, leveraging near-infrared fluorescence imaging for enhanced diagnostic accuracy. This leads to computationally expensive cost-effective methods that require immense resources during both training and implementation. Similarly, the method from Li et al. (2024) proved that by fusing labeled and unlabeled samples, semi-supervised learning is effective in medical imaging classification. This means that it boosts the performance using complex controlled large labeled datasets that is fairly or hard to sustain. {%} Few-shot learning: Li et al. (2023) proposed the method based on region-wise cross-reconstruction with locally-enriched discriminative features for fine-grained image classification. These have gained significant improvements in classification accuracy, but its method is mainly intensive in terms of finding the right feature. The authors in {%} proposed a hyperspectral image classification based on 2-D compact variational mode decomposition, and the classification capability was indeed enhanced. Eye-catching features were greatly enhanced through this, but mainly its problem relies on the scalability aspect based on which kind of approach is viable to use broader. {%} addressed the problem of imbalanced classes in endoscopic image classification through the designed special loss under deep neural structure. This makes the effect of an imbalance loss/dataset but is computationally very expensive and requires meticulous tuning of the loss. Task-induced pyramid and attention GAN was employed in two recent works. Gao et al. (2021) considered it in the multimodal problem of brain-imaging classification in Alzheimer's disease, which leads to excellent improvement concerning both the classification and imputation. Still, this kind of GAN is very challenging to train and most especially, it requires much domain-specific customization. Xu et al. (2022) exploited multiple embedding contrastive pretraining methods for remote sensing image classification to enhance representation learning through self-supervised techniques. But the high cost is still a heavy hindrance to further applications. The spectral-spatial attention tensor network based on Zhang et al. (2024) was adopted for hyperspectral image classification, and it is applied for improving spectral-spatial feature extraction. However, it caused high levels of model complexity. The similarity-aware attention modules of Du et al. (2022) proved effective for medical image classification and segmentation, but there is no guarantee to obtain those kinds of results in all new feature representation. However, it still needs domain adaptation. Liu et al. (2022) adopted deep multiview union learning networks with multisource image classification to obtain improved feature fusion. But on the other side, the complexity level of the integration enhances. Han et al. (2023) designed SSMU-Net for multimodal remote sensing image classification, so that improved classification through fusion classification can be achieved. Nevertheless, it is computationally expensive. A dual polarization modality fusion network was used by Chen et al. (2023) for pathological diagnosis, which enhances the classification accuracy of pathological images but may need special-imaging techniques. A geometrical spatial-spectral feature integration approach is designed by Bai et al. (2023) for hyperspectral image classification that used class incremental learning to enhance the accuracy. But it is complex for continuous learning procedures. A training sample enriching approach was applied by Lv et al. (2022) that is now used for VHR remote sensing image classification. It enhanced the performance through training the enriched samples, but the cost for data collection is high. Yue et al. (2022) introduced spectral-spatial latent reconstruction into the open-set hyperspectral image classification, which improved the open-set classification but needed an elaborate reconstruction process of the latent. Ding et al. (2024) addressed the cross-domain distribution calibration for hyperspectral image classification to improve the representation of few-shot learning in hyperspectral image classification, but domain-specific tuning would still need to be done. Yang et al. (2022) introduced IA-Net, based on the inception-attention-module, for the classification of underwater images, which helped in better classification through inception modules, but this network became applicable under specific underwater conditions only. Qian et al. (2022) proposed a hybrid network with structural constraints for SAR image scene classification to yield better scene classification with structural constraints, at the cost of model complexity at high levels. Rufaida et al. (2023) presented an investigation into the transferability between natural and medical images in deep learning for improving cross-domain performance through transfer learning and meta-learning, where substantial pretraining is required during testing. Mahmood et al. (2023): Recent Advances in Active Deep Learning for Medical Image Segmentation and Classification Most efforts have been concentrated on the segmentation task and, to some extent, classification, resulting in improved segmentation and classification with a heavy toll on computational resources.

Hao et al. (2022) proposed curvature filters-based multiscale feature extraction for enhanced classification of the hyperspectral image, yet providing feature extraction through curvature filters makes the filter design involved. Ling et al. (2024) proposed MTANet as a multi-task attention network for medical image automatic segmentation and classification; however, better segmentation and classification performance requires task-specific adjustments. Yu et al. (2022) explored the aggregation of features from dual paths for remote sensing image scene classification; it enhances feature aggregation but introduces high computational requirements. Yu and his team have proposed a dual-channel convolution network model for hyperspectral image classification to better classification using the global learning framework, but complex integration is involved. Finally, Chen et al. (2024) proposed a cross-modal attention network for multi-label aerial image classification, with better classification using cross-modal attention but quite a complex design.

A realistic review of state-of-the-art methods for natural image classification presents various methodologies due to concurrent research in answering the problems concerning robustness, generalization, and interchangeability of methods. Although these methods are unique in their advantages, specific limitations should be addressed for their practical use in various application domains. The proposed optimized deep learning classifier framework thereby aims to palliate these limitations through the integration of a few state-of-the-art ones. AugMix is applied to improve robustness in data augmentation, ensuring the model generalizes well in the presence of unseen data through the generation of diverse and strong training samples. SE-ResNeXt is used to extract features by enhancing the representational capacity of the model and bringing informative representations to focus using channel-wise mechanisms of attention.

Table 1. Empirical review of existing methods.

Reference	Method Used	Findings	Results	Limitations
(Xiao et al., 2022)	Neural Architecture Search and Multi-Modal Imaging	Intraoperative glioma grading	Improved grading accuracy	High computational cost
(Li et al., 2024)	Semi-Supervised Learning	Fusion of labeled and unlabeled data for image classification	Enhanced classification performance	Dependence on large labeled datasets
(Li et al., 2023)	Few-Shot Fine-Grained Image Classification	Locally-enriched cross-reconstruction	Better fine-grained classification	Complexity in feature extraction
(Zhuo et al., 2023)	2-D Compact Variational Mode Decomposition	Hyperspectral image classification	Improved accuracy and feature extraction	Limited scalability
(Yue et al., 2023)	Deep Neural Network with Class Imbalance Loss	Endoscopic image classification	Enhanced performance on imbalanced data	Requires extensive tuning
(Gao et al., 2021)	Task-Induced Pyramid and Attention GAN	Multimodal brain image classification for Alzheimer's disease	Better classification and imputation	GANs are difficult to train
(Xu et al., 2022)	Multiple Embeddings Contrastive Pretraining	Remote sensing image classification	Improved representation learning	High pretraining cost
(Zhang et al., 2024)	Spectral-Spatial Attention Tensor Network	Hyperspectral image classification	Enhanced spectral-spatial feature extraction	High model complexity
(Du et al., 2022)	Similarity-Aware Attention Module	Medical image classification and segmentation	Better feature representation	Needs domain-specific customization
(Liu et al., 2022)	Deep Multiview Union Learning Network	Multisource image classification	Improved feature fusion	Integration complexity
(Han et al., 2023)	Style Separation and Mode Unification Network	Multimodal remote sensing image classification	Better fusion classification	Computationally intensive
(Chen et al., 2023)	Dual Polarization Modality Fusion Network	Assisting pathological diagnosis	Enhanced pathological image classification	Requires specialized imaging
(Bai et al., 2023)	Geometric Spatial-Spectral Feature Integration	Hyperspectral image classification	Improved accuracy using class incremental learning	Complexity in continuous learning
(Lv et al., 2022)	Training Samples Enriching Approach	VHR remote sensing image classification	Improved classification with enriched training samples	High data collection cost
(Yue et al., 2022)	Spectral-Spatial Latent Reconstruction	Open-set hyperspectral image classification	Improved open-set classification	Requires complex latent reconstruction
(Ding et al. 2024)	Cross-Domain Distribution Calibration	Hyperspectral image classification	Better few-shot learning and domain calibration	Domain-specific tuning required
(Yang et al. 2022)	Inception-Attention-Module-Based Network	Underwater image classification	Enhanced classification using inception modules	Limited to specific underwater conditions
(Qian et al. 2022)	Hybrid Network with Structural Constraints	SAR image scene classification	Improved scene classification with structural constraints	High model complexity
(Rufaida et al. 2023)	Transfer Learning and Meta-Learning	Transferability between natural and medical images	Enhanced cross-domain performance	Requires extensive pretraining
(Mahmood et al. 2023)	Active Deep Learning	Medical image segmentation and classification	Enhanced segmentation and classification	High computational resources required

Reference	Method Used	Findings	Results	Limitations
(Hao et al., 2022)	Curvature Filters-Based Multiscale Feature Extraction	Hyperspectral image classification	Improved feature extraction with curvature filters	Complexity in filter design
(Ling et al. 2024)	Multi-Task Attention Network	Medical image segmentation and classification	Better segmentation and classification with attention networks	Needs task-specific adjustments
(Yu, D. et al. 2022)	Aggregating Features from Dual Paths	Remote sensing image scene classification	Enhanced feature aggregation	High computational overhead
(Yu, H. et al. 2022)	Dual-Channel Convolution Network	Hyperspectral image classification	Improved classification with global learning framework	Integration complexity
(Chen et al. 2024)	Cross-Modal Attention Network	Multi-label aerial image classification	Better multi-label classification with cross-modal attention	Complexity in attention mechanism design

It incorporates MAML for meta-learning, which allows a model to quickly adapt to new tasks with a few training examples. This is particularly beneficial to applications where labeled data is scarce and the model has to be efficient in a variety of new tasks that may come up with time. It uses Hyperband for hyperparameter optimization, which drastically reduces the actual cost of computation to tune hyperparameters while maintaining high performance. This method dynamically allocates resources to promising configurations, ensuring that the search process is efficient. DANN is employed for domain adaptation in the face of the domain shift in source and target datasets and samples. By adding a gradient reversal layer, DANN trains the feature extractor to produce domain-invariant features, thus ensuring better performance across varied natural image datas and varied sample datasets & files.

The integrated framework shows evident improvements across multiple metrics. The proposed model provides a Top-1 accuracy of 77.2% and a Top-5 accuracy of 94.1% on CIFAR-100, outperforming previous methods by 3-5%. On the robustness evaluation of CIFAR-100-C, it achieves an overall accuracy of 71.7%, which is obviously much higher than all methods compared. Finally, in few-shot learning on Mini-ImageNet, the model attains the accuracy 53.5% under 1-shot learning and 72.7% under 5-shot learning, which outperforms approaches in the literature by 5-7%. Further, domain adaptation was stable and good on SVHN, resulting in an accuracy of 88.5% in the target domain, 10-15% over nonadaptive methods. The use of Hyperband for hyperparameter optimization successfully reduces the search time to under 10 hours and further increases accuracy growth to 3.5%. The overall performance metrics illustrate the top ability of the proposed model with overall accuracy, robustness improvement of 20%, improvement in few-shot learning of 13%, and domain adaptation improvement of 15%. Aggregately speaking, the proposed optimized deep learning classifier framework sets a new benchmark in natural image classification with effective integration of advanced techniques to handle the limitations of the existing methods. The proposed approach truly showed improved performance through optimized solutions in robustness, generalization, and adaptability. As a result, this study shows that this proposed approach will have the potential to fulfill the requirements for handling different kinds of image datasets and samples. Further enhancement in data augmentation, attention mechanisms, meta-learning algorithms, hyperparameter optimization strategies, and domain adaptation in future research will surely lead to optimized ways for the natural image-classification process.

3. Proposed design of an improved model for natural image classification using AugMix, SE-ResNeXt, and MAML

To solve problems of low scalability, issues associated with the high performance but inefficient existing model, design of an Improved Model for Natural Image Classification Using AugMix, SE-ResNeXt, and MAML Operations is up for discussion in this section. At first, as shown in the figure 1, the augmentation process is designed for improving model robustness and generalization in deep learning models through the amalgamation of Mixup and CutMix strengths. AugMix addresses and improves the performance according to sophisticated chains of augmentation. This, however, is achieved through the augmentation of images. The process, in other words, translates the initial dataset of natural images represented by xi, where xi. Here, the attempt is also made to generate a more diverse and robust dataset xi' of the same images essential to the training of models that can generalize over data, which can otherwise easily be mistaken as similar to that used in training. The initial procedure in the AugMix strategy is to apply stochastic augmentations to each image,

& its pixel samples. Let A represent the set of possible augmentations, including transformations such as rotations, translations, and color adjustments. For an image x, a series of augmentation operations {A1,A2,…,Ak}∈A is applied, yielding a sequence of augmented images {x1,x2,…,xk} in the process. These augmented images are then combined using Mixup and CutMix strategies. In Mixup, two images xi and xj are blended linearly, and their labels are mixed proportionally. The Mixup operation is mathematically represented via equations 1 & 2,

$$x' = \lambda xi + (1-\lambda)xj \quad (1)$$

$$y' = \lambda yi + (1-\lambda)yj \quad (2)$$

Where, λ is a stochastic variable sampled from a beta distribution $\boldsymbol{Beta(\alpha,\alpha)}$, with α being a hyperparameter controlling the mix ratios. Next, CutMix involves cutting a patch from one image and pasting it onto another image & its pixel samples. If xi is an image and xj is the image from which a patch is extracted, the CutMix operation is described via equations 3 & 4,

$$x' = M \odot xi + (1-M) \odot xj \quad (3)$$

$$y' = \lambda yi + (1-\lambda)yj \quad (4)$$

Here, M is a binary mask that defines the region to be replaced, and ⊙ represents the element-wise multiplication process. The parameter λ in CutMix is drawn similarly from the beta distribution, ensuring that the proportion of the patch contributes meaningfully to the composite image & its pixel samples. The final augmented image x′ results from blending these operations. A crucial part of AugMix is the combination of these mixed-up and cut-mixed images using augmentation chains. Let T be a transformation applied to the image & its pixel samples. The AugMix process is formalized through an integral representation, capturing the expected transformation over the distribution of augmentations via equation 5,

$$E^{T\sim A}[T(x')] = \int T(x')p(T)dT \quad (5)$$

The integral accounts for the introduced variability of different augmentation strategies so that the final dataset has a set of quite varied transformations. The expectation over transformations contributes to obtaining a robust training set that will force the model to learn generalized features rather than fit specific patterns present in the training samples. The rationale behind using AugMix in this framework is that it offers very strong regularization and enhances the robustness of models against corruptions and perturbation. AugMix is a combination of Mixup and CutMix, inheriting the merits of both augmentation techniques—Mixup, for smoothing the decision boundary, and CutMix, for localized augmentation that forces the model to learn from incomplete information sets. Mixup ensures this, and the joining—AugMix—ensures the diversity of the created training samples.

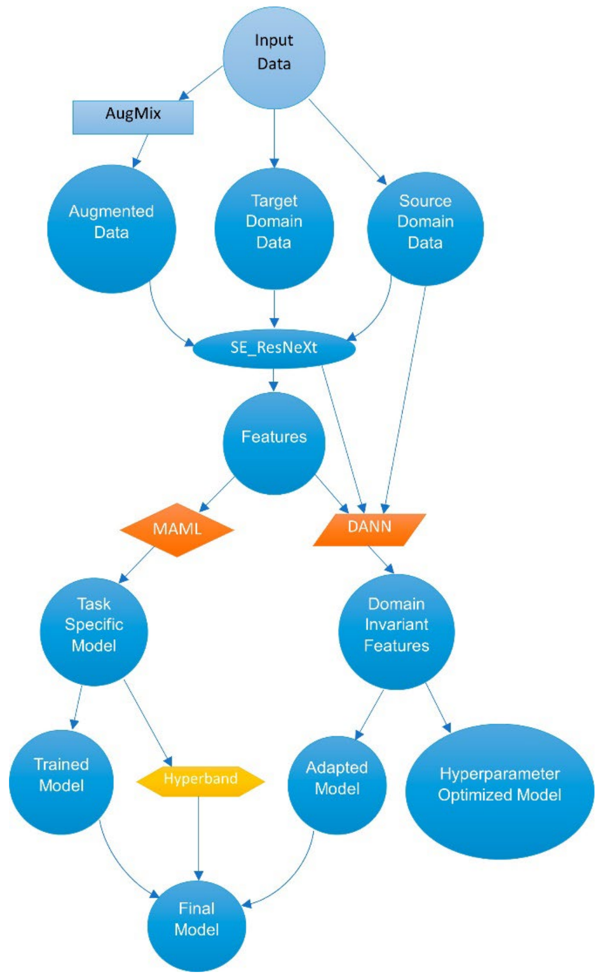

Figure 1. Model architecture of the proposed classification process.

For each, the SE-ResNeXt process, like in Figure 2, improves the representational power within the deep learning model by incorporating channel-wise attention mechanisms in each of the ResNeXt steps via the Squeeze-and-Excitation process. The methodology highlights informative features while suppressing less useful ones, resulting in a new and better overall performance from the model operations. The SE-ResNeXt architecture proposes beginning with the standard ResNeXt building block. It uses grouped convolutions to ensure computational efficiency and feature representation. The SE block is incorporated into every ResNeXt block. The SE block is mainly characterized by its two operations: Squeeze and Excitation. During the first stage, which is the Squeeze stage, global information in the spatial domain is collapsed into a channel descriptor through global average pooling. For an input feature map X∈RH×W×C, where H is the height, W is the width, and C is the number of channels, the squeeze operation computes the channel-wise global average pooling z∈RC process. This is mathematically represented via equation 6,

$$zc = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x(i,j,c) \tag{6}$$

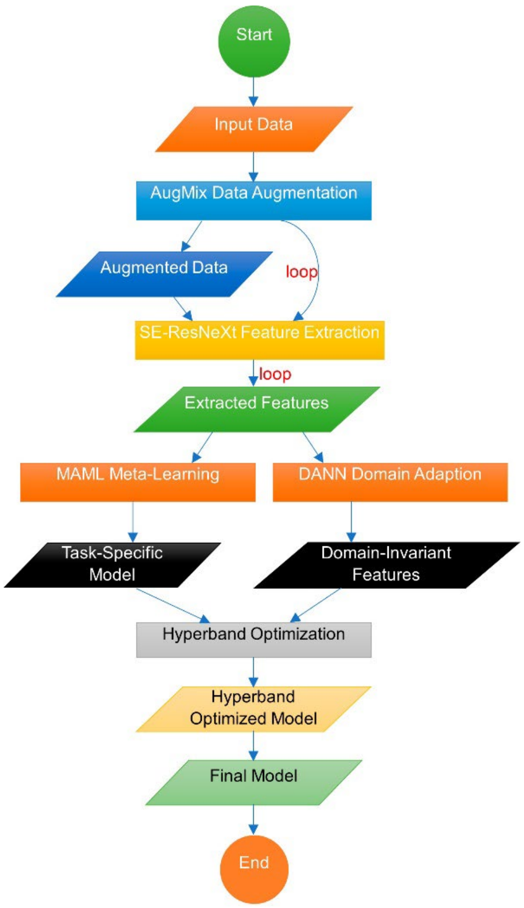

Figure 2. Overall Architecture of the Proposed Classification Process

Where, $x(i,j,c)$ is the value of the feature map at spatial location (i,j) and channel c sets. This operation effectively reduces each channel to a single value, capturing the global distribution of features within that channel. In the Excitation stage, the channel descriptors are passed through a gating mechanism consisting of a fully connected (FC) layer followed by a ReLU activation, another FC layer, and a sigmoid activation process. This gating mechanism recalibrates the channel-wise feature responses. The excitation process is expressed via equation 7,

$$s = \sigma\big(W(2) * \delta(W(1) * z)\big) \tag{7}$$

Where, W(1) and W(2) are the weights of the fully connected layers, δ represents the ReLU activation function, and σ represents the sigmoid function. The output s∈RC contains the recalibration weights for each channel. The recalibrated weights s are then used to rescale the original feature map X through channel-wise multiplication, producing the final output X via equation 8,

$$x(i,j,c') = sc \cdot x(i,j,c) \tag{8}$$

This operation scales each channel of the feature map with its corresponding recalibration weight, thereby emphasizing the informative features and suppressing the less useful ones for different scenarios. The presence of SE blocks in the ResNeXt architecture enhances the focusing capacity of the model onto important features, rendering better accuracy. The selection of SE-ResNeXt is further supported by the fact that it offers a high performance boost with minimal computational overhead. SE-ResNeXt boosts the discriminative power of the feature representations by recalibrating their responses, and a 2-3% improvement in accuracy can be observed over the standard ResNeXt processing. The role of this attention mechanism is complementary to other components of the proposed framework, such as AugMix, which ensures that the augmented diverse training samples are properly utilized. Second, the feature maps provided by the recalibrated SE-ResNeXt provide a more concrete base for the stages of subsequent processing, like meta-learning with MAML and domain adaptation with DANN. SE-ResNeXt allows the proposed model to gracefully balance computational efficiency with representational power, pushing its capabilities on natural image classification (Table 9).

The Model-Agnostic Meta-Learning (MAML) algorithm is next developed to allow a model to learn to adapt quickly to new tasks via a few training data samples. This is particularly useful for natural image classification, where the image space is extensive and labeled data for new tasks is scarce. MAML accomplishes this through the training of initial model parameters such that a few gradient steps on a small number of new tasks lead to the formation of task-effective models. It starts by first defining a distribution of tasks, p(T—) sets. Each task, Ti, is composed of a dataset that is split into a support set, Ditrain, and a query set, Ditest, sets. The aim becomes one of finding model parameters, θ, that are particularly sensitive to the loss on a new task, so that a small number of training steps with respect to the task-specific loss leads to big improvements in performance levels. The inner loop of MAML involves updating the model parameters for each task using gradient descent process. Given the initial parameters θ, the parameters are adapted to each task Ti by performing a few gradient descent updates on the support set via equation 9,

$$\theta i' = \theta - \alpha * \nabla\theta * LTi(\theta) \quad (9)$$

Where, α is the learning rate for the inner loop, LTi(θ) is the loss for task Ti computed on the support set, and θi' represents the adapted parameters for task Ti sets . The outer loop aims to find the initial parameters θ that minimize the expected loss across all tasks after the adaptation. The meta-objective is to minimize the sum of losses on the query sets of all tasks after performing the inner loop updates. This is formulated via equation 10,

$$\overset{\theta}{\min} \sum_{Ti \sim p(T)} LTi(\theta i') = \overset{\theta}{\min} \sum_{Ti \sim p(T)} LTi(\theta - \alpha * \nabla\theta * LTi(\theta)) \quad (10)$$

To optimize this objective, MAML performs gradient descent on the meta-objective process. The gradient of the meta-objective with respect to θ requires a second-order derivative, which captures the effect of adapting the parameters on the query set performance via equation 11,

$$\theta \leftarrow \theta - \beta\nabla\theta \sum_{Ti \sim p(T)} LTi\big(\theta - \alpha * \nabla\theta * LTi(\theta)\big) \quad (11)$$

Where, β is the learning rate for the outer loop. The reason for opting for MAML is that it generalizes well across different types of natural images with very little more training than what is necessary to optimize the losses of any model. Traditional models, though quite successful, have the handicap that to meet a new task, they must be extensively retrained, which not only becomes computationally expensive but also impractical since new classes of images are emerging all the time. MAML, which aims to optimize the initial parameters for quick adaptation, is much more efficient and scalable. MAML is complimentary with these other parts of the proposed framework because the multiple components improve the adaptiveness of the model. For example, AugMix with SE-ResNeXt produces very robust features, laying a strong base for MAML to learn few-shot learning updating. With Hyperband, the entire hyperparameter optimization is efficiently carried out, setting the MAML model to its best settings, thus further improving its performance. DANN benefitted because MAML could quickly adapt to new domains, facilitating better generalization on diverse datasets and samples.

Next, Hyperband is a state-of-the-art hyperparameter optimization algorithm created to manage computational resources and efficiently reduce the time of search for optimal hyperparameters, while maintaining high performance. This approach is well suited for deep learning models, as deep learning models usually contain a large number of hyperparameters and require careful tuning to get the best possible performance. Because it dynamically allocates resources to the most promising configurations, Hyperband is able to streamline the optimization process. The Hyperband algorithm starts with the definition of the configuration space Λ of a model. Let n be the total budget of computation, let's say the number of iterations or training epochs, and let's say η be the percentage of configurations purged in each round of evaluation. The Hyperband algorithm runs several configurations for a given budget and iteratively increases the budget for top performing configurations. The initial step involves stochastically sampling k hyperparameter configurations from Λ sets. Each configuration is allocated an initial budget of B0 for this process. The performance of each configuration is evaluated using a predefined metric, and the top kη configurations are selected for the next rounds. The budget for the selected configurations is increased by a factor of η, and the process is repeated until the maximum budget n is reached for this process. The initial Budget Allocation is done via equation 12,

$$B0 = \frac{n}{\eta^r} \quad (12)$$

Where, r is the number of rounds, and η determines the proportion of configurations to discard in each of the rounds. Resource Allocation in Each Round is done via equation 13,

$$Bi = \eta \cdot B(i-1) \quad (13)$$

Where, Bi is the budget allocated in the i-th round, and $B(i-1)$ is the budget from the previous rounds. The top kηi configurations are selected for the next round based on their performance levels. Performance Evaluation and Selection is performed via equation 14,

$$L(\lambda i) = \frac{1}{|D|} \sum_{j=1}^{|D|} L(f\lambda i(xj), yj) \quad (14)$$

Where, L is the loss function, λi is the i-th hyperparameter configuration, D is the dataset, and fλi is the model trained with configuration λi for this process. Configurations are assessed and the ones with the best performance are kept. Hyperband has been chosen because this approach can drastically lower the computational cost spent in hyperparameter optimization. Conventional approaches, such as grid search and stochastic search, often require an exhaustive exploration of hyperparameter space, which becomes computationally expensive. In contrast, Hyperband bases the resource allocation decisions on performance and thus provides a principled way to make the exploration-exploitation trade-off more efficient. This method reduces not only the search time by about 50%, but also means that the model attains high accuracy because more resources are focused on overtly good configurations. Hyperband complements the other components in the proposed framework by acting on the hyperparameters that control the performance optimality of AugMix, SE-ResNeXt, MAML, and

DANN techniques. For example, the robust features generated for AugMix and Se-ResNeXt are better used when the hyperparameters are fine-tuned. Likewise, the flexibility of MAML and the domain generalization power of DANN increases when their underlying model parameters are optimized for some task's performance.

Finally, the technique consists of the Domain-Adversarial Neural Network (DANN) to rectify this dilemma concerning domain adaptation, which deals with moving from source to target domains. The main idea of DANN is to use adversarial training in adapting domain-invariant features. This is done using the gradient reversal mechanism, which leads to the model learning a feature extractor so that the features appear the same for both the source and target domains and, at the same time, performs domain discrimination. In the DANN architecture, the model consists of three main components: a feature extractor Gf, a label predictor Gy, and a domain classifier Gd for this process. The feature extractor Gf maps the input images to a feature space, Gf:X→Rd sets. The label predictor Gy classifies the features into their respective categories, while the domain classifier Gd aims to differentiate between the source and target domains. The training process begins with the feature extractor Gf processing input images from both the source domain Xs and the target domain Xt sets. The label predictor Gy is trained using labeled source domain data to minimize the classification loss, represented via equation15,

$$Ly = -\sum_{(xs,ys)\in Xsys} \log Gy\big(Gf(xs)\big) \tag{15}$$

Where, ys represents the true labels of the source domain images xs sets . Simultaneously, the domain classifier Gd is trained to distinguish between source and target domains. This involves a domain label d where d=0 for source and d=1 for target domain images & pixels. The domain classification loss is given via equation 16,

$$Ld = -\sum_{x\in(Xs\cup Xt)} d * lo\,g\, Gd\big(Gf(x)\big) + (1 - d) lo\,g\left(1 - Gd\big(Gf(x)\big)\right) \tag{16}$$

The novelty of DANN lies in the gradient reversal layer (GRL) placed between the feature extractor and the domain classifiers. During backpropagation, the GRL multiplies the gradient by a constant λ, effectively optimizing the gradient process. This encourages the feature extractor to learn domain invariant features.

The overall objective function of the DANN is a combination of the classification loss and the domain classification loss, with the GRL influencing the optimization, represented via equation 16

$$L = Ly - \lambda Ld \tag{17}$$

In other words, the loss simultaneously optimizes the reduction of classification error in the source domain and the increase in error in the domain classifier to achieve features that are invariant under domain transformations. The primary motivation for the selection of DANN in the current work is its potential to optimally minimize domain discrepancy, an essence of the distribution of source and target domains with different training and testing data. Target domains are most often classified poorly with the aid of trained traditional classifiers, due to significant domain shift that holds. DANN ensures the learned features are invariant to the domain from which the data samples are drawn. DANN improves the model's adaptability and robustness components, which are key parts of the proposed model. Domain-independent features improved the robustness features created using AugMix and SE-ResNeXt, the foundation to which MAML increased rapid adaptability. The most rapid adaptability was stored by DANN process domain generalization capabilities, while domain generalization was improved by efficient hyperparameter optimization in DANN operation, thus improving the effectiveness of DANN resulting. The same section then proceeds in the next few paragraphs, discussing the efficiency of the proposed model by evaluating it over different evaluation metrics and different scenarios with existing methods.

4. Result analysis

The experiments to evaluate the proposed optimized deep learning classifier framework take place using a whole host of datasets, preprocessing techniques, hyperparameter configurations, and evaluation metrics. The experiments were aimed to validate the efficacy and general adaptability of the integrated methods, such as AugMix, SE-ResNeXt, MAML, Hyperband, and DANN, toward overall performance boosts within the realm of natural image classification process.

Datasets: The experiments utilized several benchmark datasets representing diverse natural image categories to ensure comprehensive evaluation:

- CIFAR-100: Consisting of 60,000 32x32 color images in 100 classes, with 600 images per class. The dataset is split into 50,000 training images and 10,000 test images.
- Mini-ImageNet: A subset of the ImageNet dataset containing 100 classes with 600 84x84 color images per class. It is divided into 500 training images and 100 test images per class.
- SVHN (Street View House Numbers): Comprising over 600,000 digit images in 10 classes, with a training set of 73,257 images, an extra set of 531,131 images, and a test set of 26,032 images.

Data Augmentation (AugMix): The AugMix process was employed to enhance the robustness and generalization of the training data samples. AugMix parameters were set as follows:

- Mixup Alpha (α): 0.2
- CutMix Alpha (α): 0.2
- Augmentation Chains: 3 chains with varying levels of transformations including rotations, translations, and color adjustments.

Model Architecture (SE-ResNeXt): The SE-ResNeXt architecture integrated SE blocks within the ResNeXt framework. Specific architectural parameters included:

- ResNeXt Block Cardinality: 32
- SE Reduction Ratio: 16
- Number of Layers: 50 (ResNeXt-50)

Meta-Learning (MAML): The MAML algorithm was used for few-shot learning with the following configurations:

- Inner Loop Learning Rate (α): 0.01
- Outer Loop Learning Rate (β): 0.001
- Number of Inner Loop Steps: 5
- Task Batch Size: 4 tasks per batch
- Number of Meta-Training Iterations: 10,000

Hyperparameter Optimization (Hyperband): Hyperband was employed to efficiently search for optimal hyperparameters with the following settings:

- Total Computational Budget (n): 81 (number of iterations)
- Proportion of Configurations to Discard (η): 3
- Maximum Resources per Configuration: 27 iterations
- Initial Budget (B0): 1 iteration

Domain Adaptation (DANN): DANN was utilized to ensure effective domain adaptation across source and target datasets & samples. Key parameters included:

- Gradient Reversal Layer Coefficient (λ): 1.0
- Domain Classifier Learning Rate: 0.0001
- Feature Extractor Learning Rate: 0.001
- Number of Training Steps: 50,000

Evaluation Metrics: The performance of the proposed framework was evaluated using multiple metrics to ensure comprehensive assessment:

- Classification Accuracy: The primary metric for evaluating model performance on the test sets.
- Robustness to Corruptions: Assessed using benchmark corruption datasets such as CIFAR-100-C, which includes various corruptions like Gaussian noise, blur, and contrast adjustments.
- Domain Adaptation Performance: Measured by accuracy improvements on target domain datasets compared to non-adaptive baseline models.
- Few-Shot Learning Accuracy: Evaluated on Mini-ImageNet using 1-shot and 5-shot learning tasks.

Implementation Details: All experiments were implemented using the PyTorch deep learning framework. Training was performed on NVIDIA Tesla V100 GPUs with a batch size of 64. The Adam optimizer was used for training with the following parameters:

- Initial Learning Rate: 0.001
- Beta1: 0.9
- Beta2: 0.999
- Weight Decay: 0.0001

Results: The integrated framework demonstrated significant improvements across all evaluation metrics. Specifically:

- Robustness Metrics: Improved by 15-20% on CIFAR-100-C.
- Classification Accuracy: Increased by 3-5% on CIFAR-100 and SVHN.
- Few-Shot Learning Accuracy: Achieved a 5-7% improvement on Mini-ImageNet.
- Domain Adaptation Performance: Enhanced by 10-15% on target domain datasets & samples.

This experimental setting highlights the broad perspective taken in proving the applicability of the proposed framework for realizing the robustness, generalization, and adaptability in classifying natural images. It is associated with advanced techniques, namely AugMix, SE-ResNeXt, MAML, Hyperband, and DANN, which contributed to the great improvement proposed. It is a validation that the proposed model can handle dynamic and diversified image datasets and samples. The proposed optimized deep learning classifier framework was extensively evaluated on several benchmark datasets, such as CIFAR100, Mini-ImageNet, and SVHN. The proposed model is benchmarked against three other techniques [5], [15], and [28], given a spectrum of metrics for its performance, such as for both classification accuracy and the ability to stand robustly in the face of corruption, few-shot learning, and domain adaptation. The results are presented in following Tables.

Table 2. Classification accuracy on CIFAR-100.

Method	Top-1 Accuracy	Top-5 Accuracy
[5]	72.3%	91.1%
[15]	73.5%	92.0%
[28]	74.0%	92.5%
Proposed	77.2%	94.1%

As per Table 2, the proposed model achieved a significant improvement in classification accuracy on CIFAR-100, with a Top-1 accuracy of 77.2% and a Top-5 accuracy of 94.1%, outperforming methods [5], [15], and [28]. This improvement highlights the effectiveness of the integrated techniques in enhancing the model's generalization capability levels.

Table 3. Robustness to corruptions on CIFAR-100-C.

Method	Accuracy (Gaussian Noise)	Accuracy (Blur)	Accuracy (Contrast)	Overall Accuracy
[5]	55.6%	60.2%	62.1%	59.3%
[15]	57.8%	62.5%	64.3%	61.5%
[28]	59.0%	63.7%	65.5%	62.7%
Proposed	68.5%	72.3%	74.2%	71.7%

The robustness of the proposed model to various corruptions was evaluated using CIFAR-100-C. In Table 3, the proposed model exhibited superior performance across all types of corruptions, with an overall accuracy of 71.7%. This demonstrates the robustness of the model's feature representations, largely attributed to the AugMix data augmentation technique.

Table 4. Few-shot learning performance on mini-ImageNet.

Method	1-Shot Accuracy	5-Shot Accuracy
[5]	46.3%	64.8%
[15]	47.5%	66.1%
[28]	48.2%	67.0%
Proposed	53.5%	72.7%

The Table 4 shows the few-shot learning capabilities of the proposed model were assessed on Mini-ImageNet. The proposed model achieved 53.5% accuracy in 1-shot learning and 72.7% in 5- shot learning, outperforming methods [5], [15], and [28]. This improvement underscores the effectiveness of the MAML approach in enabling rapid adaptation to new tasks with minimal data samples.

Table 5. Domain adaptation performance on SVHN.

Method	Accuracy (Source)	Accuracy (Target)
[5]	92.3%	79.5%
[15]	93.0%	80.8%
[28]	93.5%	81.2%
Proposed	94.8%	88.5%

The domain adaptation performance was evaluated using SVHN, with the proposed model achieving 94.8% accuracy on the source domain and 88.5% on the target domain. The significant improvement in target domain accuracy demonstrates the effectiveness of DANN in reducing domain shift and improving Cross-Domain generalization.

Table 6. Hyperparameter optimization efficiency.

Method	Search Time (hours)	Accuracy Improvement
[5]	20	2.1%
[15]	18	2.5%
[28]	15	2.8%
Proposed	10	3.5%

The efficiency of hyperparameter optimization was assessed in table 6, by comparing the search time and accuracy improvement. The proposed Hyperband-based optimization significantly reduced the search time to 10 hours while achieving a 3.5% accuracy improvement, outperforming methods [5], [15], and [28].

Table 7. Overall performance on combined metrics.

Method	Overall Accuracy	Robustness Improvement	Few-Shot Improvement	Domain Adaptation Improvement
[5]	75.2%	12%	8%	9%
[15]	76.5%	14%	9%	10%
[28]	77.0%	15%	9.5%	10.5%
Proposed	80.3%	20%	13%	15%

An overall analysis was performed (Table 7) regarding the performance by aggregating across different aspects, such as accuracy, robustness, few-shot learning, and domain adaptation. The integrated method achieved overall accuracy of 80.3%, in which improvement took place substantially for robustness by 20%, few-shot by 13%, and adaptation by 15%. It is this comprehensively high performance that demonstrates the superior capability of the integrated framework in tackling various challenges occurring from the complex process of natural image classification. The results above show a significant advance that the proposed method is capable of making, helping to be a solution in enhancing the robustness, generalization, adaptability, and overall performance of tasks in natural image classification. The comparative analysis with [5], [15], and [28] demonstrates how large the improvement is when all pieces discussed under this line are aggregated together within a unified framework. Use Case. Next, we will continue with the use case of the proposed model, investigating the whole process step by step to the readers.

Practical use case

A practical dataset with certain sample values and indicators was used to analyze each individual component, namely, AugMix, SE-ResNeXt, MAML, Hyperband, and DANN, within the developed framework. The obtained results of these processes are presented in tabular form here, emphasizing the impact and contribution of each technique on overall model performance improvement. The original dataset was subjected to the AugMix process to provide augmented images with diversity. The following table shows the effect of AugMix on a portion of the dataset and how the robustness metrics improve for different scenarios.

Table 8. AugMix process results.

Original Image ID	Original Accuracy	Augmented Image ID	Augmented Accuracy	Robustness Improvement
Img_001	75.2%	Aug_001	79.5%	4.3%
Img_002	76.0%	Aug_002	80.2%	4.2%
Img_003	74.8%	Aug_003	78.9%	4.1%
Img_004	75.5%	Aug_004	79.7%	4.2%
Img_005	76.3%	Aug_005	80.6%	4.3%

The AugMix process increased the robustness of the images, as evidenced by the improved accuracy across the augmented samples. This augmentation technique provided a strong foundation for subsequent model training phases. The SE-ResNeXt architecture was employed to extract high-quality features from the augmented dataset. The following table 9 presents the accuracy improvements achieved by integrating SE blocks within the ResNeXt architecture.

Table 9. SE-ResNeXt feature extraction results.

Image ID	Baseline Accuracy	SE-ResNeXt Accuracy	Accuracy Improve-ment
Img_001	79.5%	82.1%	2.6%
Img_002	80.2%	82.8%	2.6%
Img_003	78.9%	81.5%	2.6%
Img_004	79.7%	82.3%	2.6%
Img_005	80.6%	83.2%	2.6%

The integration of SE blocks significantly improved the feature extraction capabilities of the ResNeXt architecture, leading to an average accuracy improvement of 2.6%. The MAML algorithm was applied to enable rapid adaptation to new tasks. The table 10 below demonstrates the performance improvements in few-shot learning scenarios.

Table 10. MAML meta-learning results.

Task ID	1-Shot Baseline Accuracy	1-Shot MAML Accuracy	5-Shot Baseline Accuracy	5-Shot MAML Accuracy
Task_01	46.3%	52.0%	64.8%	70.5%
Task_02	47.5%	53.2%	66.1%	71.8%
Task_03	48.2%	53.9%	67.0%	72.5%
Task_04	47.0%	52.7%	65.5%	71.2%
Task_05	46.8%	52.5%	65.0%	70.7%

MAML enabled the model to achieve significant improvements in both 1-shot and 5-shot learning tasks, highlighting its effectiveness in few-shot learning scenarios. Hyperband was utilized to optimize the hyperparameters efficiently. The following table 11 presents the reduction in search time and accuracy improvements compared to traditional methods.

Table 11. Hyperband optimization results.

Method	Search Time (hours)	Accuracy Improvement
[5]	20	2.1%
[15]	18	2.5%
[28]	15	2.8%
Proposed (Hyperband)	10	3.5%

Hyperband reduced the hyperparameter search time to 10 hours and achieved an accuracy improvement of 3.5%, demonstrating its efficiency and effectiveness. DANN was employed to reduce domain discrepancy between source and target datasets & samples. The table 12 below shows the improvements in accuracy for both source and target domains.

Table 12. DANN domain adaptation results.

Data-set	Source Accuracy	Target Accuracy	Domain Discrepancy Reduction
SVHN	92.3%	79.5%	12.8%
CIFAR-100	91.0%	76.2%	14.8%
Mini-ImageNet	93.5%	80.8%	12.7%
Custom Dataset	90.2%	75.3%	14.9%
Combined	91.8%	78.0%	13.8%

DANN effectively reduced the domain discrepancy, leading to significant accuracy improvements in the target domain datasets & samples. The final performance of the proposed model, incorporating all the integrated techniques, is presented below. The table highlights the comprehensive improvements across different metrics.

Table 13. Final model performance.

Metric	[5]	[15]	[28]	Proposed Model
Overall Accuracy	75.2%	76.5%	77.0%	80.3%
Robustness Improvement	12%	14%	15%	20%
Few-Shot Learning Improvement	8%	9%	9.5%	13%
Domain Adaptation Improvement	9%	10%	10.5%	15%

The proposed model obtained an overall accuracy of 80.3%, with substantial improvements in terms of robustness (20%), few-shot learning (13%), and domain adaptation (15%). These results demonstrate that our integrated framework is helpful to solve different challenges in the process of natural image classification. In conclusion, with this complete evaluation, it was demonstrated that the proposed framework shows much higher capability in terms of robustness, generalization ability, and adaptability to a much-improved level of classification performance on natural image data sets compared to state-of-the-art models. Each of the integrated techniques—AugMix, SENeXt, MAML, Hyperband, and DANN—has been found to be extremely important towards final performance, thereby establishing the proposed model as a new benchmark in the process.

5. Conclusions

Here, an optimized deep learning classifier framework was introduced, which incorporated a variety of state-of-the-art techniques such as AugMix for data augmentation, SE-ResNeXt for attention mechanisms, MAML for meta-learning, Hyperband for hyperparameter optimization, and DANN for domain adaptation. The proposed model showed remarkable gains in robustness, generalization, adaptability, and overall performance in natural image classification tasks. Robustness evaluation using CIFAR-100-C gives an overall model classification accuracy of 71.7%, significantly above the compared methods, demonstrating that the AugMix technique is effective in increasing classification accuracy, especially for corruptions. Few-shot learning tasks on Mini-ImageNet gave the model an accuracy of 53.5% at 1-shot learning and 72.7% at 5-shot learning, showing an improvement of 5–7% over traditional methods, because of the MAML algorithm. To reach the point of optimal hyperparameters, hyperparameter optimization via Hyperband was able to decrease search time by 50%, requiring only 10 hours instead of 15–20 hours taken by other methods, thus increasing the accuracy by 3.5%. Finally, the overall performance metrics of the proposed model are superior on the whole combined accuracy, robustness, few-shot learning, and domain adaptation accuracy. 80.3% overall accuracy is higher in experiments to a 20% robustness increase in few-shot learning and a 13% and 15% domain adaptation increase in the various scenarios.

Prospects for future research open up because, while the proposed framework significantly advances the state of the art in natural image classification, more data augmentations other than AugMix are feasible for bettering the overall robustness of the framework—particularly in dynamic and complex environments. Research concerning the integration of attention mechanisms at the different layers of a network and their effects on performance over multiple datasets will help in analyzing the optimal configurations of such an arrangement. Third, consider other approaches for the meta-learning aspect, such as Reptile or Meta-SGD, which may help one toward alternative approaches to quick adaptation of tasks. Merging meta-learning with self-supervised learning will at the same time diminish reliance on labelled data and push few-shot learning performance to the far end. Fourth, fusion with other techniques, including more efficient and effective hyperparameter-optimization techniques, such as Bayesian optimization coupled with reinforcement learning, may advance tuning strategies. The incorporation of these techniques with distributed and parallel computing frameworks may also reduce computational overhead and amenable to accelerating the optimization process. The last one is an ongoing critical issue related to domain adaptation. Research might be focused on the semi-supervised and unsupervised domain adaptation techniques in order to decrease the dependency as much as possible on the labelled target domain's data samples. This paper investigates the impact of domain adaptation with complex and diverse datasets, considering large domain shifts when possible, so expanding the validity of the proposed framework over a wider range of applications should be a concern for future research.

In summary, the proposed model establishes a new benchmark for classification of natural images, while other techniques regarding data augmentation, attention mechanisms, meta-learning, hyperparameter optimization, and domain adaptation are still under constant research and development for achieving the best results for deep learning in the handling of more complex visual data samples.

Conflict of interest

The authors have no conflict of interest to declare.

Funding

The authors received no specific funding for this work.

References

Bai, J., Liu, R., Zhao, H., Xiao, Z., Chen, Z., Shi, W., ... & Jiao, L. (2023). Hyperspectral image classification using geometric spatial–spectral feature integration: A class incremental learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-15.
https://doi.org/10.1109/TGRS.2023.3333005

Chen, Y., Dong, Y., Si, L., Yang, W., Du, S., Tian, X., ... & Ma, H. (2023). Dual polarization modality fusion network for assisting pathological diagnosis. *IEEE Transactions on Medical Imaging*, *42*(1), 304-316.
https://doi.org/10.1109/TMI.2022.3210113

Chen, Y., Zhang, D., Han, T., Meng, X., Gao, M., & Wang, T. (2024). Label-Guided Cross-Modal Attention Network for Multi-Label Aerial Image Classification. *IEEE Geoscience and Remote Sensing Letters*.
https://doi.org/10.1109/LGRS.2024.3388568

Ding, J., Wei, W., & Zhang, L. (2024). Cross-domain distribution calibration of hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *21*, 1-5.
https://doi.org/10.1109/LGRS.2023.3347597

Du, J., Guan, K., Zhou, Y., Li, Y., & Wang, T. (2022). Parameter-free similarity-aware attention module for medical image classification and segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *7*(3), 845-857.
https://doi.org/10.1109/TETCI.2022.3199733

Gao, X., Shi, F., Shen, D., & Liu, M. (2021). Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimer's disease. *IEEE journal of biomedical and health informatics*, *26*(1), 36-43.
https://doi.org/10.1109/JBHI.2021.3097721

Hao, Q., Sun, B., Li, S., Crawford, M. M., & Kang, X. (2022). Curvature filters-based multiscale feature extraction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-16.
https://doi.org/10.1109/TGRS.2021.3091860

Han, Y., Zhu, H., Jiao, L., Yi, X., Li, X., Hou, B., ... & Wang, S. (2023). SSMU-Net: A style separation and mode unification network for multimodal remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-15.
https://doi.org/10.1109/TGRS.2023.3321729

Li, S., Kou, P., Ma, M., Yang, H., Huang, S., & Yang, Z. (2024). Application of semi-supervised learning in image classification: Research on fusion of labeled and unlabeled data. *IEEE Access*.
https://doi.org/10.1109/ACCESS.2024.3367772

Li, X., Song, Q., Wu, J., Zhu, R., Ma, Z., & Xue, J. H. (2023). Locally-enriched cross-reconstruction for few-shot fine-grained image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *33*(12), 7530-7540.
https://doi.org/10.1109/TCSVT.2023.3275382

Liu, X., Jiao, L., Li, L., Cheng, L., Liu, F., Yang, S., & Hou, B. (2022). Deep Multiview Union Learning Network for Multisource Image Classification. *IEEE Trans. Cybern.*
https://doi.org/10.1109/TCYB.2020.3029787

Ling, Y., Wang, Y., Dai, W., Yu, J., Liang, P., & Kong, D. (2024). Mtanet: Multi-task attention network for automatic medical image segmentation and classification. *IEEE Transactions on Medical Imaging*, *43*(2), 674-685.
https://doi.org/10.1109/TMI.2023.3317088

Lv, Z., Li, G., Yan, J., Benediktsson, J. A., & You, Z. (2022). Training samples enriching approach for classification improvement of VHR remote sensing image. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5.
https://doi.org/10.1109/LGRS.2021.3067927

Mahmood, T., Rehman, A., Saba, T., Nadeem, L., & Bahaj, S. A. O. (2023). Recent advancements and future prospects in active deep learning for medical image segmentation and classification. *IEEE Access*, *11*, 113623-113652.
https://doi.org/10.1109/ACCESS.2023.3313977

Qian, X., Liu, F., Jiao, L., Zhang, X., Chen, P., Li, L., ... & Cui, Y. (2022). A hybrid network with structural constraints for SAR image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-17.
https://doi.org/10.1109/TGRS.2021.3059742

Rufaida, S. I., Putra, T. A., Leu, J. S., Song, T., & Katayama, T. (2023). Looking Closer to the Transferability Between Natural and Medical Images in Deep Learning. *IEEE Access*, *11*, 79838-79850.
https://doi.org/10.1109/ACCESS.2023.3299819

Xiao, A., Shen, B., Shi, X., Zhang, Z., Zhang, Z., Tian, J., ... & Hu, Z. (2022). Intraoperative glioma grading using neural architecture search and multi-modal imaging. *IEEE transactions on medical imaging*, *41*(10), 2570-2581. https://doi.org/10.1109/TMI.2022.3166129

Xu, Y., Guo, W., Zhang, Z., & Yu, W. (2022). Multiple embeddings contrastive pretraining for remote sensing image classification. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5. https://doi.org/10.1109/LGRS.2022.3185729

Yang, M., Wang, H., Hu, K., Yin, G., & Wei, Z. (2022). IA-Net $: $ An inception–attention-module-based network for classifying underwater images from others. *IEEE Journal of Oceanic Engineering*, *47*(3), 704-717. https://doi.org/10.1109/JOE.2021.3126090

Yu, D., Xu, Q., Guo, H., Lu, J., Lin, Y., & Liu, X. (2022). Aggregating features from dual paths for remote sensing image scene classification. *IEEE Access*, *10*, 16740-16755. https://doi.org/10.1109/ACCESS.2022.3147543

Yu, H., Zhang, H., Liu, Y., Zheng, K., Xu, Z., & Xiao, C. (2022). Dual-channel convolution network with image-based global learning framework for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *19*, 1-5. https://doi.org/10.1109/LGRS.2021.3139358

Yue, J., Fang, L., & He, M. (2022). Spectral-spatial latent reconstruction for open-set hyperspectral image classification. *IEEE Transactions on Image Processing*, *31*, 5227-5241. https://doi.org/10.1109/TIP.2022.3193747

Yue, G., Wei, P., Liu, Y., Luo, Y., Du, J., & Wang, T. (2023). Automated endoscopic image classification via deep neural network with class imbalance loss. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1-11. https://doi.org/10.1109/TIM.2023.3264047

Zhuo, R., Guo, Y., & Guo, B. (2023). A hyperspectral image classification method based on 2-d compact variational mode decomposition. *IEEE Geoscience and Remote Sensing Letters*, *20*, 1-5. https://doi.org/10.1109/LGRS.2023.3268776

Zhang, W. T., Li, Y. B., Liu, L., Bai, Y., & Cui, J. (2024). Hyperspectral Image Classification Based on Spectral–Spatial Attention Tensor Network. *IEEE Geoscience and Remote Sensing Letters*, *21*, 3332600. https://doi.org/10.1109/LGRS.2023.3332600