Stochastic Meta-Learning for Augmentation Policy (SMAP): Enhancing Fine-Grained Image Classification

Mahbuba Tasmin
mtasmin@umass.edu

Fareya Ikram
fikram@umass.edu

Jeng-Yu Chou
jchou@umass.edu

1. Abstract

In the domain of computer vision, the robustness and generalization of deep learning models are critical, particularly in fine-grained image classification tasks such as species or object recognition. In this project, we explore various data augmentation methods, ranging in complexity and prevalence, to assess their impact on ResNet-50 in fine-grained image classification. This study introduces a novel optimized augmentation policy, Stochastic Meta-Learning for Augmentation Policy (SMAP), that leverages meta-learning to optimize augmentation strategies for enhanced classifier performance. We evaluate the efficacy of SMAP by applying it to two distinct and challenging datasets: the Caltech-UCSD Birds-200 (CUB) and the Oxford 102 Category Flower (Flower). Our approach utilizes a ResNet-50 model as the backbone classifier and compares the impact of SMAP against traditional usage of augmentation techniques. The results demonstrate that SMAP significantly improves the model's accuracy, with notable increases in both top-1 and top-5 metrics. The efficacy test on the Flower dataset further emphasizes the advantages of SMAP. This research underlines the potential of meta-learning in fine-tuning data augmentation to improve the robustness of classifiers in fine-grained image recognition tasks.

2. Introduction

The practice of data augmentation plays an essential role in enhancing the generalizability and performance of machine learning models, particularly when new data acquisition is challenging. It becomes critical in scenarios where datasets are small, imbalanced, or biased, which leads to suboptimal model performance. Models can learn more robust features and generalize better to new, unseen data by simulating a variety of conditions through data augmentation.

In this project, we investigate the impact of data augmentation on the performance of a ResNet-50 model trained on the Caltech-UCSD Birds (CUB) dataset, a standard benchmark in fine-grained image classification [2, 23]. We extend this investigation to the Oxford 102 Category Flower dataset (Flower), examining the efficacy of augmentation techniques across different domains. We compare ResNet-50 performance on different strata of data to study whether certain data augmentation techniques and approaches lead to a more accurate classifier for the task of bird classification.

The problem statement of the project is as follows:

We investigate the extent to which various data augmentation methods enhance model performance and develop an informed data augmentation pipeline using a meta-learning approach to optimize model performance given a set of augmentation techniques.

Our research aims to explore the extent to which various data augmentation methods can improve model performance. We introduce "Stochastic Meta-Learning for Augmentation Policy" (SMAP), a novel approach that employs meta-learning to identify the most effective combination of augmentation strategies. This approach is particularly beneficial for complex datasets where the best augmentation techniques are not immediately obvious.

We challenge the assumption that larger datasets invariably lead to better performance. Our hypothesis posits that a strategically augmented dataset, constrained to the size of the original dataset, can perform as well or better than a more extensively augmented one. Through SMAP, we demonstrate that informed and targeted augmentation can capture the intricacies of real-world variability, which is especially crucial for fine-grained classification tasks. Our results underline the efficacy of SMAP in enhancing model generalizability, providing a significant improvement over baseline models without substantially increasing the training set size.

Our contributions include:

1. Choose a set of six standard and advanced image augmentation methods expected to introduce variety in the dataset with simultaneous intra-performance evaluation of these augmentation methods.

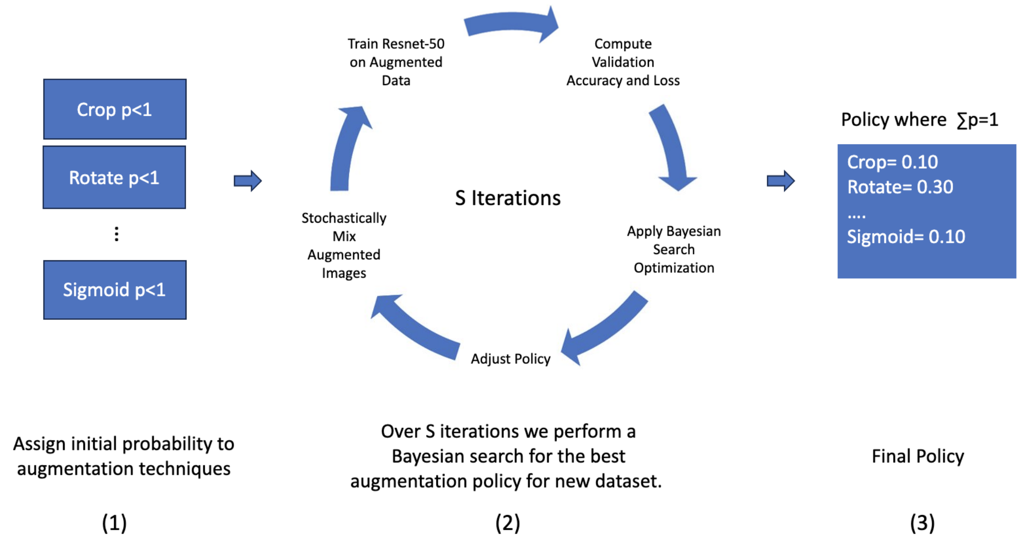


Figure 1. Stochastic Meta-Learning for Augmentation Policy (SMAP). This data augmentation pipeline has three main components. First, the set of six augmentation methods is assigned initial probability values in a range of [0.01 to 1]. Bayesian optimization is then run on a subset (N=2000) from the training set to discover the best augmentation policy through maximizing validation accuracy.

2. Introduction of SMAP, an innovative data augmentation pipeline that uses meta-learning to automate the selection of optimal augmentation strategies, saving time and computational resources.
3. Extension of the augmentation analysis from the CUB dataset to the Flowers dataset, enabling a broader understanding of SMAP's applicability and effectiveness across different types of data.
4. Empirical evidence that strategic augmentation via SMAP can enhance model performance on par with or better than extensive augmentation, without increasing the training dataset size.

The remainder of this paper is structured as follows: Section 2 discusses related work in the domain of data augmentation and fine-grained image classification. Section 3 details the methodology employed, including dataset descriptions, the architecture of the ResNet-50 model, and the implementation of SMAP. Section 4 presents the experimental results, offering a comprehensive analysis of model performance with and without SMAP. Section 5 delves into the discussion of the findings, comparing the performance on the CUB and Flower datasets with benchmark studies. Finally, Section 6 concludes the paper, summarizing the key takeaways, and suggesting directions for future research.

3. Related Work

3.1. Data Augmentation

Data augmentation's role in enhancing the diversity of training data for image classification has been well-recognized. Early studies like Krizhevsky et al. highlighted basic augmentation techniques like horizontal reflections and RGB channel alterations as effective methods to reduce overfitting in CNNs [15]. More advanced approaches, such as AutoAugment and Faster AutoAugment, have automated the search for optimal augmentation policies, leading to improved performance across various benchmarks [9, 11, 14, 18].

Our project explores dropout [20] as a data augmentation technique, utilizing the concept of pixel elimination during training to assess its impact. This approach aligns with recent studies examining dropout's equivalence to data augmentation [25] and variations like Random Erasing, which involves erasing parts of the image with random values [26]. Earlier research by Bouthillier et al. also interprets dropout as a data augmentation form, comparing its effects to training with noisy inputs [5].

3.2. Semi-Supervised and Meta-Learning

Other relevant works have used a variety of learning methods to enhance this area of research as well. For example, Ghosh and Thiery conducts a study on data augmentation and consistency-based semi-supervised deep learning, stating that "consistency-based SSL methods are in general more powerful since they can better exploit the local geometry of the data-manifold if efficient data-augmentation/perturbation schemes are used." Other similar works that employ semi-supervised learning for image classification include Yalniz et al., also evaluating performance of ResNet-50, which utilizes a large collection of unlabeled data to improve model performance. Likewise, Ni et al. combines data augmentation with meta-learning to explore how "different categories of data within the training pipeline impact meta-learning performance" and implement Meta-MaxUp which significantly improves performance of popular meta-learners on few-shot benchmark datasets [4, 16, 21, 22].

3.3. Usage

In this project, we expand upon the concepts from prior research, taking inspiration from works like Faster AutoAugment [11] to employ a meta-learning method of optimal data augmentation policy discovery to develop a data augmentation pipeline for domain-specific, fine-grained image classification.

4. Experimental Setup

4.1. Dataset

The study leveraged two distinct datasets: the Caltech-UCSD Birds-200-2011 (CUB) [23] and the Oxford 102 Category Flower dataset (Flower) [17]. The CUB dataset is a comprehensive collection of 200 bird species images designed for fine-grained image recognition, which presents a challenge due to the subtle inter-class variations and high intra-class similarities. The Flower dataset, on the other hand, comprises 102 categories of common United Kingdom flowers, each class consisting of between 40 and 258 images. This dataset is particularly challenging due to the wide variety of shapes, sizes, and colors that flowers can exhibit.

4.2. ResNet-50

ResNet-50, a deep convolutional neural network introduced by He et al. [12], has been seminal in advancing the field of computer vision. Its deep architecture is highly effective for complex image classification tasks, like identifying bird species. Its pre-trained model on ImageNet is adept at capturing detailed features and, when fine-tuned with augmented datasets, can help prevent overfitting and improve generalization. This makes ResNet-50 particularly suitable for our data augmentation study, where it serves as a robust base model for enhancing classification performance through augmented training samples.

4.3. Image Augmentation Libraries

For data augmentation, we utilized the OpenCV [1], Albumentations [7], and imgaug [13] libraries, each providing unique capabilities for enhancing computer vision tasks. OpenCV offers a vast array of basic image manipulation functionalities, while Albumentations delivers optimized performance for a wide range of augmentations. imgaug specializes in advanced augmentation techniques, allowing for the application of complex transformations with probabilistic parameters for more nuanced modifications.

4.4. Data Augmentation Methods

Our investigation covered both standard augmentation techniques such as rotation, random cropping, and grayscale conversion, as well as advanced methods like sigmoid contrast adjustment, image corruption with Gaussian noise, and pixel dropout. The former set aims to induce basic variability, whereas the latter introduces more complex patterns to potentially aid the model in learning more robust features. Our augmentation pipeline was meticulously constructed to dovetail with the training process, dynamically applying transformations to training images. This procedure not only diversified the dataset but also mitigated the risk of overfitting by preventing the model from seeing the exact same image across epochs.

Method	**Characteristics**	**Complexity**
rotation	angle modification orientation robustness	standard
random cropping	random image section recognition despite position	standard
grayscale	remove color reduce image complexity	standard
image corruption	Gaussian Noise moderate corruption	advanced
sigmoid contrast	non-linear pronounce subtle features	advanced
dropout	robustness to missing data translation invariance	advanced

Table 1. The complexity of each augmentation method is listed along with their main characteristics. These details inform the expected capability of each method to remove task irrelevant data from a given dataset to best allow a model to capture task useful information during training.

4.5. Data Augmentation Policy Discovery

Bayesian optimization [6, 19], a statistical approach for global optimization, was pivotal in our study for determining effective data augmentation policies. This technique was employed to systematically explore and evaluate various combinations of augmentation methods, enabling us to identify a data-driven strategy that optimizes the classifier's performance. This approach is particularly valuable in fine-tuning augmentation parameters, aligning with our SMAP framework's goal of enhancing model generalizability through informed and adaptive augmentation choices. This process is shown in Step (2) of Figure 1.

4.6. Evaluation Metrics

Prior work using the CUB dataset for fine-grained image classification have evaluated their model(s) performance using evaluation metrics such as training and validation accuracy and loss, top-k accuracy, and macro-F1, precision, and recall [3, 8, 27]. We use these listed metrics to assess ResNet-50 performance. Specifically, Top-1 and Top-5 scores were used to evaluate the model's predictive performance. The Top-1 score measures the model's accuracy based on its most confident prediction, while the Top-5 score considers a correct prediction if the true label is among the model's top five predictions. These metrics are particularly insightful for datasets with a large number of classes, as they provide a more lenient and practical evaluation of the model's ability to identify the correct class out of many possible options.

5. Data Augmentation Pipeline

5.1. Stochastic Meta-Learning for Augmentation Policy

5.1.1 Description and Motivation

The proposed algorithm, named "Stochastic Meta-Learning for Augmentation Policy" (SMAP), is an iterative method for determining an effective augmentation policy for training machine learning models. The motivation behind SMAP is to leverage meta-learning principles to adaptively find the optimal combination of augmentation strategies that enhances model generalizability and performance on unseen data.

At each iteration, SMAP stochastically mixes augmented images from a set of predefined augmentation methods, each method associated with an initial probability. Each of these augmented datasets is generated by applying augmentation transformation on the randomly chosen original training subset. These probabilities are adjusted based on the validation set performance after training the classifier with the mixed images. This meta-learning approach is particularly useful when dealing with complex datasets where the optimal augmentation strategy is not apparent. By automating the search for the best policy, SMAP removes the need for extensive manual experimentation and allows for a more efficient and systematic approach to model training.

5.1.2 SMAP Mathematical Equation

Let $\mathcal{D}$ represent the dataset and $\mathcal{M}$ the machine learning model. Let $f(\mathcal{D};\theta)$ denote the training process of model $\mathcal{M}$ with parameters θ on dataset $\mathcal{D}$, and $V(\mathcal{M};\mathcal{D}_{\text{val}})$ the validation process on validation set $\mathcal{D}_{\text{val}}$.

The SMAP algorithm seeks to optimize an augmentation policy π defined by a set of probabilities $p_1, p_2, ..., p_n$ associated with n predefined augmentation methods $\{A_1, A_2, ..., A_n\}$, such that each p_i corresponds to the probability of selecting augmentation method A_i.

The optimization is formalized as a Bayesian optimization problem:

$$\max_{\pi} V(f(\mathcal{D}_\pi;\theta);\mathcal{D}_{\text{val}})$$

where $\mathcal{D}_\pi$ is the augmented dataset under policy π, which is a stochastic mixture of images resulting from applying the augmentations A_i on a randomly chosen subset of the original training data $\mathcal{D}$.

This is subject to the constraints:

$$p_i \in [0.01, 1], \quad \forall i \in \{1, 2, ..., n\}$$

$$\sum_{i=1}^{n} p_i = 1$$

The objective function to be maximized, V, represents the model's validation accuracy after training with the augmented dataset $\mathcal{D}_\pi$. The algorithm iteratively adjusts the probabilities $p_1, p_2, ..., p_n$ to improve validation performance, converging on an optimal set $\{p_1^*, p_2^*, ..., p_n^*\}$ that defines the final augmentation policy π^*.

To reflect the stochastic nature of the method selection process, we define a random variable X such that $X = A_i$ with probability p_i. Then, the augmented dataset $\mathcal{D}_\pi$ can be expressed as:

$$\mathcal{D}_\pi = \bigcup_{i=1}^{n} A_i(\mathcal{D}_{\text{sub}}) \text{ with probability } p_i$$

where $\mathcal{D}_{\text{sub}}$ is a randomly chosen subset of $\mathcal{D}$.

The optimization is performed using the 'gp_minimize' function from the scikit-optimize library, which efficiently searches the high-dimensional space of augmentation probabilities to find the policy π^* that yields the highest validation accuracy.

The final augmentation policy is normalized and defined as:

$$\pi^* = \left\{ \frac{p_1^*}{T}, \frac{p_2^*}{T}, ..., \frac{p_n^*}{T} \right\}$$

where $T = \sum_{i=1}^{n} p_i^*$ to ensure the probabilities sum to 1.

This mathematical framework encapsulates the core principles of SMAP and provides a clear, quantitative basis for its iterative augmentation policy optimization process.

6. Performance Analysis

In line with the methodology detailed in Section 5, we employed the ResNet-50 model, pretrained on ImageNet, to assess the impact of six augmentation techniques on the CUB dataset. These techniques — rotation, random cropping, grayscale conversion, image corruption via Gaussian noise, sigmoid contrast adjustment, and pixel dropout — were each applied to create augmented datasets equal in size to the original. By combining these with the original dataset, we generated augmented datasets of varying sizes, up to seven times larger than the original, to train and evaluate the ResNet-50 model.

Utilizing Bayesian optimization, we derived a data augmentation policy tailored for bird classification on the CUB dataset (see Figure 1 Step 2). This policy guided the optimal use of each technique (see Figure 1 Step 1) to enhance model performance.

Upon establishing this augmentation policy with the CUB dataset, as described in Section **??**, we replicated the process with the Flower Dataset [17] to evaluate the policy's effectiveness across different domains.

6.1. Baseline

Using the ResNet-50 model without data augmentation on the CUB dataset yielded macro metrics below 50%, indicating the model's limited ability to accurately classify the diverse bird species. The high precision and recall for specific classes suggest that some features are better captured than others. Overall results suggest significant potential for improvement.

6.2. Individual Data Augmentation Methods

Investigating six data augmentation methods revealed that standard techniques like rotation, cropping, and grayscale significantly improved all performance metrics over the baseline. The advanced methods (image corruption, sigmoid contrast, dropout) also improved upon the baseline but to a lesser extent than standard methods, with random cropping emerging as the most effective single technique. Table 2 shows the individual augmentation method performance on CUB dataset.

6.3. Combining Data Augmentation Methods

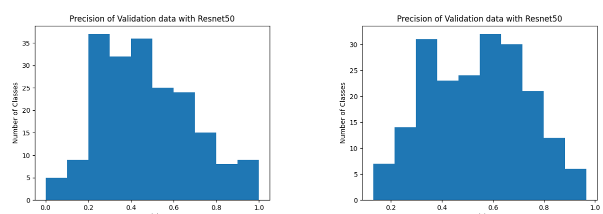
The model was evaluated on three combination-type training datasets referred to as "Standard Methods", "Advanced Methods", and "All Methods" in Table 2. When combining augmentation methods, "Standard Methods" yielded the highest increases in all performance metrics, surpassing the advanced and combined strategies. While the "Advanced Methods" and "All Methods" datasets, which were significantly larger, did show enhanced training accuracy, they did not consistently outperform the standard augmentations in other metrics. This suggests that while advanced methods are beneficial, they require more strategic application to maximize their effectiveness for fine-grained classification tasks.

6.4. Assessing SMAP's Impact on CUB Dataset

As described in Section 5, the SMAP policy employs a Bayesian optimization strategy to fine-tune the probability of applying each augmentation type, aiming to discover the most effective augmentation mix to enhance model generalization on a given dataset. In the experimental setup, two distinct augmentation policies were compared: a baseline with no augmentation and a stochastic meta-learning augmentation policy (respectively referred to as "None" and "SMAP" in Figures 2 and 5). Table 3 displays the data augmentation policy produced using SMAP.

Figure 2 represents the distribution of precision scores for class prediction task on the CUB dataset for the "None" and "SMAP" augmentation policies. In Figure 2a, without augmentation, the distribution of precision scores across classes is more concentrated in the mid-range precision scores. The histogram has peaks in the 0.6 precision range, where about 35 classes fall. There are also a considerable number of classes with precision in the lower ranges (0.2 to 0.4), indicating that a fair number of classes are not well predicted.

In Figure 2b a different distribution is shown after the "SMAP" augmentation is applied. The peak of this histogram is in the higher precision range, between 0.7 and 0.8, where about 28 classes are situated. This histogram also shows fewer classes in the lower precision ranges (below 0.5) compared to the first histogram 2a.



(a) With no Augmentation (None)

(b) SMAP Augmentation

Figure 2. Precision of validation data with Resnet-50

The performance improvement can be quantified by the shift of the histogram's peak from a lower precision range to a higher precision range and the reduction in the number of classes with low precision scores. Average precision increased from 0.47 to 0.54, and average accuracy from 0.43 to 0.54 when comparing the "None" and "SMAP" scenarios. This enhancement is indicative of the potential of SMAP to discern and implement a more efficacious augmentation strategy, thereby bolstering the model's ability to generalize from the training data to unseen data.

Experiment	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Top-1 Accuracy (%)	Top-5 Accuracy (%)	macro-F1	macro-Precision	macro-Recall	Dataset Size
Baseline	3.35	0.60	3.61	0.43	43.11	74.35	0.43	0.49	0.45	N
Rotation	2.67	0.61	2.62	0.61	60.56	88.00	0.58	0.64	0.60	2N
Random Cropping	2.56	0.63	2.58	0.62	62.05	87.94	0.61	0.65	0.62	2N
Grayscale	2.91	0.55	2.76	0.60	60.36	86.73	0.58	0.64	0.60	2N
Image Corruption	2.76	0.58	2.73	0.58	57.73	85.49	0.56	0.63	0.58	2N
Sigmoid Contrast	2.67	0.59	2.67	0.61	60.67	87.02	0.59	0.65	0.61	2N
Dropout	3.1	0.51	3.1	0.50	49.97	79.74	0.48	0.55	0.50	2N
Standard Methods	1.99	0.66	1.74	0.74	73.80	93.44	0.73	0.77	0.74	4N
Advanced Methods	1.49	0.80	2.17	0.54	54.44	82.97	0.54	0.55	0.54	4N
All Methods	0.99	0.87	1.87	0.56	56.13	84.00	0.56	0.57	0.56	7N

Table 2. Comparison of model performance metrics with various augmentation methods on the CUB dataset, indicating improvements in accuracy and dataset size expansion.

Method	Probability
crop	0.364
rotate	0.33
grayscale	0.003
dropout	0.016
blur	0.004
sigmoid	0.282

Table 3. This table displays the data augmentation policy for the CUB dataset that achieved the best performance after 50 iterations of Bayesian optimization. The values under the "Probability" column sum to 1 and represent the probability these data augmentation methods will contribute to the generated augmented training dataset.

This performance improvement is mirrored in the top-1 and top-5 accuracy scores, where SMAP outperforms the baseline, achieving a top-1 score of 53.3656% (compared to the baseline's 43.1481%) and a top-5 score of 83.2758% (compared to the baseline's 74.37%). Notably, the SMAP policy seemed to exert a more pronounced positive impact on classes that were previously highly confused, such as various species of hummingbirds and warblers. For instance, the top confused pair of '029.American Crow' and '107.Common Raven', which had a confusion count of 20 in the baseline, reduced to 11 in the top confused pairs in the SMAP scenario. The confusion count of two other classes, '068.Ruby throated Hummingbird' and '069.Rufous Hummingbird', reduced from 8 in the baseline to null in SMAP. This performance increase at the class level suggests that the targeted augmentations have successfully aided the model in distinguishing between these similar classes.

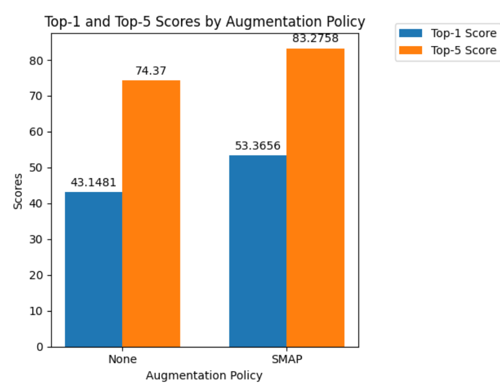


Figure 3. Top-K Accuracy Achieved by None and SMAP on CUB dataset

6.5. Assessing SMAP's Impact on Flower Dataset

To evaluate the impact of the SMAP augmentation policy on the Flower dataset, the model was trained using the ResNet-50 architecture, first without any augmentations (None) and then with the SMAP policy.

Table 4 presents the augmentation policy generated using SMAP.

Method	Probability
crop	0.214
rotate	0.083
grayscale	0.187
dropout	0.166
blur	0.242
sigmoid	0.108

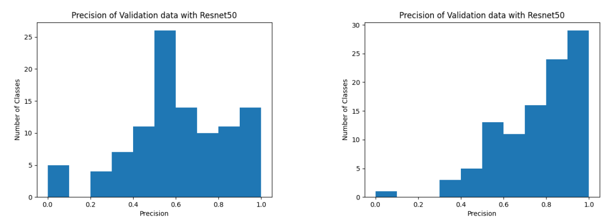
Table 4. This table displays the data augmentation policy for the Oxford 102 Flower Dataset that achieved the best performance after 50 iterations of Bayesian optimization. The values under the "Probability" column sum to 1 and represent the probability these data augmentation methods will contribute to the generated augmented training dataset.

Table 5 displays ResNet-50 performance on the original Flower Dataset and on a training dataset produced using the SMAP generated augmentation policy on the six augmentation methods studied in this project.

Metric	Baseline Score	SMAP Score
train loss	3.42	2.40
train accuracy	0.79	0.93
validation loss	3.66	2.85
validation accuracy	0.58	0.75
top-1 accuracy	58.14%	74.51%
top-5 accuracy	82.75%	91.76%
macro-F1	0.55	0.73
macro-precision	0.61	0.77
macro-recall	0.58	0.75
Dataset Size	N	2N

Table 5. This table displays the general performance of ResNet-50 on the Oxford 102 Flower Dataset without any use of data augmentation (baseline results) compared to the model performance when trained on a training dataset created following data augmentation policy created using SMAP.

The histogram presented in fig 4 demonstrates how the precision distribution changes when no augmentation is applied versus when "SMAP" augmentation is applied. In fig 4a, representing validation data without augmentation, the precision scores are more varied, with a significant number of classes having low precision (notably, a cluster of classes in the 0.2 precision range). The distribution is somewhat bimodal, with peaks around 0.2 and 0.6. The lack of bars towards the higher end of the precision scale (0.8 to 1.0) indicates that few classes achieved high precision.



(a) With no Augmentation (None)　(b) SMAP Augmentation

Figure 4. Precision of validation data of the Oxford 102 Categories Flower dataset

In Figure 4b, representing validation data with "SMAP" augmentation applied, there is a marked shift towards higher precision scores. The distribution is heavily skewed towards the 0.8 to 1.0 range, with the tallest bar indicating that over 25 classes achieved high precision. This histogram shows a clear improvement in precision across the classes, as evidenced by the concentration of classes with high precision.

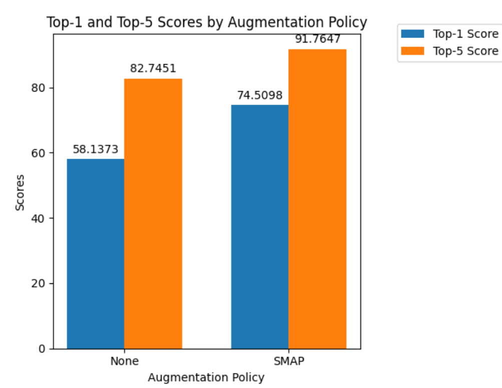


Figure 5. Top-K Accuracy Achieved by None and SMAP on Flower dataset

The Top-K accuracy chart, Figure 5, has two pairs of bars, each pair corresponding to a different augmentation policy. For the "None" condition: The Top-1 score is approximately 58.14%, while the Top-5 score is higher, at approximately 82.75%. For the "SMAP" condition: the Top-1 score shows a significant improvement, at approximately 74.51% and the Top-5 score is also improved, at about 91.76%. This clearly shows that the application of the "SMAP" augmentation policy has led to substantial improvements in both Top-1 and Top-5 accuracy scores on the Flowers dataset. This demonstrates that "SMAP" augmentation is effective in enhancing the model's performance on this particular task.

7. Discussion

7.1. Inspiration of SMAP

We drew inspiration from the principles laid out in the Faster AutoAugment paper [11], which emphasizes the acceleration of policy searching for data augmentation. This approach introduces gradient approximations for non-differentiable image operations and renders the search process end-to-end differentiable. Building upon this innovative groundwork, we extended these concepts into the realm of meta-learning. Our approach, Stochastic Meta-Learning for Augmentation Policy (SMAP), diverges from the original idea by not only focusing on the speed of policy search but also on the adaptability and effectiveness of the augmentation policy in enhancing model generalization.By doing so, we have embraced the core idea of using a learnable policy to guide data augmentation from Faster AutoAugment and have evolved it to incorporate a self-improving loop that is characteristic of meta-learning models.

7.2. Performance Enhancement

The introduction of the SMAP pipeline has proved to be a pivotal development in our research, demonstrating that a well-informed augmentation strategy can effectively improve model generalizability without necessitating an enlarged dataset.

Our findings challenge the prevailing notion that larger datasets guarantee superior model performance. Instead, we observed that a strategically augmented dataset, even when constrained closely to the original dataset's size, can achieve comparable or superior results. This is particularly evident in our experiments with the CUB and Oxford Flowers datasets, where SMAP augmented models consistently outperformed their non-augmented counterparts. Notably, the precision histograms for the CUB dataset exhibit a clear shift towards higher precision scores when SMAP is applied, underscoring the effectiveness of the policy in reducing the model's confusion between similar classes.

7.3. Selection of Standard vs Advanced Augmentation Methods

The Stochastic Meta-Learning for Augmentation Policy (SMAP) results for both the CUB and Flower datasets generates a tailored approach to each dataset's characteristics and requirements. For the CUB dataset, SMAP recommends a higher contribution probability for the following methods: Crop: 36.4% , Rotate: 33.0% and Sigmoid: 28.2%. These are relatively higher probabilities compared to the advanced methods of grayscale, dropout, and blur, which have much lower probabilities. This indicates that SMAP is leaning towards certain standard augmentation methods like cropping and rotation, which may be more effective for the fine-grained classification tasks present within the CUB dataset. Additionally, the significant probability assigned to the sigmoid method suggests that enhancing contrast is important for distinguishing between closely related bird species.

In contrast, for the Flower dataset, the probabilities are more evenly distributed among the methods, with a notable preference for blur: Blur: 24.2%, Crop: 21.4%, Grayscale: 18.7%. The probabilities suggest a balanced mix between standard methods like crop and grayscale and an advanced method like blur. This mix may cater to the diverse shapes and colors of the flowers, where blurring might help the model focus on structural features rather than detailed textures.

7.4. Commonly mistaken pairs

ResNet-50 consistently confused the same pairs of classes for each other on the primary dataset of the project, CUB, each time it was trained on a dataset consisting of the original training data combined with training data augmented with one of these three common augmentation methods. The number of mistakes made on the top ten pairs of classes with the highest amount of mistakes was almost equivalent across these three methods; however, grayscale results in the least number of mistakes per pair of classes. The three more advanced, or less prevalent data augmentation methods tested in this project - image corruption, sigmoid contrast, dropout - all resulted in slightly higher amounts of total mistakes for the ten pairs of classes most often mistaken for each other; although, the pairings of classes varied across these three methods.

8. Conclusions

Our study with the Stochastic Meta-Learning for Augmentation Policy (SMAP) highlights the importance of tailored data augmentation in fine-grained image classification. The preference for standard methods like cropping and rotation in the CUB dataset underscores their effectiveness in enhancing essential features for species differentiation. Conversely, the balanced use of augmentations in the Flower dataset suggests a need for diverse strategies to accommodate varied floral characteristics. These results emphasize that optimal data augmentation is domain-specific and that a universal approach may not yield the best outcomes. SMAP's success in these experiments indicates its potential applicability across different datasets, especially in fine-grained classification tasks.

Future research should explore the applicability of SMAP in other domains and investigate the balance between augmented dataset size and performance enhancement. Our findings advocate for the strategic use of data augmentation to improve model performance efficiently, potentially reducing the need for larger datasets.

References

[1] Opencv: Open source computer vision library. GitHub: https://github.com/opencv/opencv, 2023. 3

[2] Ailurophile. 200 bird species with 11,788 images. Kaggle https://www.kaggle.com/datasets/veeralakrishna/200-bird-species-with-11788-images, 2019. 1

[3] Asish Bera, Zachary Wharton, Yonghuai Liu, Nik Bessis, and Ardhendu Behera. Sr-gnn: Spatial relation-aware graph neural network for fine-grained image categorization. *IEEE Transactions on Image Processing*, 31:6017–6031, 2022. 4

[4] Luca Bertinetto, Joao F Henriques, Philip HS Torr, and Andrea Vedaldi. Meta-learning with differentiable closed-form solvers. *arXiv preprint arXiv:1805.08136*, 2018. 3

[5] Xavier Bouthillier, Kishore Konda, Pascal Vincent, and Roland Memisevic. Dropout as data augmentation. *arXiv preprint arXiv:1506.08700*, 2015. 2

[6] Eric Brochu, Vlad M Cora, and Nando De Freitas. A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010. 3

[7] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020. 3

[8] Po-Yung Chou, Yu-Yung Kao, and Cheng-Hung Lin. Fine-grained visual classification with high-temperature refinement and background suppression. *arXiv preprint arXiv:2303.06442*, 2023. 4

[9] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation policies from data. *arXiv preprint arXiv:1805.09501*, 2018. 2

[10] Atin Ghosh and Alexandre H. Thiery. On data-augmentation and consistency-based semi-supervised learning, 2021. 2

[11] Ryuichiro Hataya, Jan Zdenek, Kazuki Yoshizoe, and Hideki Nakayama. Faster autoaugment: Learning augmentation strategies using backpropagation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 1–16. Springer, 2020. 2, 3, 8

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 3

[13] Alexander B. Jung, Kentaro Wada, Jon Crall, Satoshi Tanaka, Jake Graving, Christoph Reinders, Sarthak Yadav, Joy Banerjee, Gábor Vecsei, Adam Kraft, Zheng Rui, Jirka Borovec, Christian Vallentin, Semen Zhydenko, Kilian Pfeiffer, Ben Cook, Ismael Fernández, François-Michel De Rainville, Chi-Hung Weng, Abner Ayala-Acevedo, Raphael Meudec, Matias Laporte, et al. imgaug. https://github.com/aleju/imgaug, 2020. Online; accessed 01-Feb-2020. 3

[14] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 2

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc. 2

[16] Renkun Ni, Micah Goldblum, Amr Sharaf, Kezhi Kong, and Tom Goldstein. Data augmentation for meta-learning. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8152–8161. PMLR, 2021. 3

[17] Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian conference on computer vision, graphics & image processing*, pages 722–729. IEEE, 2008. 3, 5

[18] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115 (3):211–252, 2015. 2

[19] Nicholas D. Sanders, Richard M. Everson, Jonathan E. Fieldsend, and Alma A. M. Rahat. Bayesian search for robust optima, 2021. 3

[20] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 2

[21] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, 2019. 3

[22] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 3

[23] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 1, 3

[24] I Zeki Yalniz, Hervé Jégou, Kan Chen, Manohar Paluri, and Dhruv Mahajan. Billion-scale semi-supervised learning for image classification. *arXiv preprint arXiv:1905.00546*, 2019. 2

[25] Dazhi Zhao, Guozhu Yu, Peng Xu, and Maokang Luo. Equivalence between dropout and data augmentation: A mathematical check. *Neural Networks*, 115:82–89, 2019. 2

[26] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 13001–13008, 2020. 2

[27] Haowei Zhu, Wenjing Ke, Dong Li, Ji Liu, Lu Tian, and Yi Shan. Dual cross-attention learning for fine-grained visual categorization and object re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4692–4702, 2022. 4