Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD39275

Research on Animal Image Classification Based on AutoML: A Case Study of Baidu EasyDL

Zhisheng Wang

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
1059576830@qq.com

Abstract. With the rapid development of deep learning, Automated Machine Learning (AutoML) has shown clear advantages in reducing the cost of manual parameter tuning and improving model development efficiency. Using the Baidu EasyDL platform as the experimental vehicle, this paper systematically evaluates AutoML for animal image classification. Two datasets are selected: Oxford-IIIT Pet, a fine-grained multi-class scenario (37 classes, 7354 images), and Animals-10, a coarse-grained few-class scenario (10 classes, 26176 images). Models are trained with EasyDL's accuracy-improvement configuration package and its automatic hyperparameter search function. EasyDL achieves a Top-1 accuracy of 95.0% and an F1-score of 95.1% on Oxford-IIIT Pet, and an accuracy of 99.3% and an F1-score of 99.2% on Animals-10. The results indicate that AutoML performs very well when classes are highly distinguishable and remains competitive on fine-grained classification, but it is limited in interpretability and in separating highly similar classes. Because both benchmarks consist of well-framed images, these figures should not be taken as expected accuracy in the field: uneven lighting, occlusion and cluttered backgrounds in camera-trap images require separate validation. This paper offers practical guidance for researchers without a deep learning background who wish to apply AutoML to animal image classification.

Keywords: AutoML, image classification, Baidu EasyDL, Oxford-IIIT Pet, Animals-10

1. Introduction

With the continuing maturity of machine learning, computer vision has been widely applied in biology and ecology, especially in species identification, behavior analysis and ecological monitoring [1-3]. As a core task of computer vision, accurate and efficient animal image classification is of great significance for the digital transformation of these fields.

However, traditional deep learning methods (such as CNN, ResNet and EfficientNet) achieve high accuracy at the cost of notable technical barriers. Designing the model architecture requires specialized knowledge of neural networks, and tuning hyperparameters such as the learning rate, batch size and optimizer is laborious. Manual tuning consumes considerable time and effort, and it still cannot guarantee a globally optimal solution [4]. These barriers make it difficult for researchers without specialized backgrounds to build high-performance animal image classification models on

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD39275

their own, which restricts the adoption of deep learning in interdisciplinary fields such as ecology and zoology.

Against this background, Automated Machine Learning (AutoML) has emerged. By automating feature engineering, model selection and hyperparameter optimization, AutoML encapsulates the complex model development process in a user-friendly pipeline and greatly lowers the technical threshold of deep learning [5]. Commercial platforms such as Google AutoML Vision, Microsoft Azure Machine Learning and Baidu EasyDL have further promoted its adoption. Nevertheless, systematic evaluation of AutoML for animal image classification remains insufficient. Most existing studies focus on general object recognition benchmarks such as ImageNet, whereas animal images pose particular challenges, including subtle appearance differences between individuals, complex background interference and diverse poses [6]; it is therefore hard to infer AutoML performance in this domain directly from general results. Moreover, few studies evaluate the adaptability of AutoML through cross-dataset comparison, that is, the performance difference of the same platform on datasets of different complexity, which is important guidance for dataset construction and platform selection in practice.

The motivation of this study follows from these gaps. By selecting two animal image datasets with clearly different difficulty levels, we systematically evaluate the classification performance of Baidu EasyDL in different scenarios and analyze the factors that affect AutoML performance. The main contributions are threefold. First, with Baidu EasyDL as the research object, this paper presents a complete experimental workflow for AutoML-based animal image classification, covering data preprocessing, platform configuration and result evaluation. Second, two standard datasets of markedly different difficulty, Oxford-IIIT Pet (fine-grained, 37 classes) and Animals-10 (coarse-grained, 10 classes), are used to examine how AutoML performance relates to dataset characteristics. Third, the AutoML mechanisms of EasyDL, including neural architecture search and automatic hyperparameter search, are described, and the constraints on transferring this approach to real-world camera-trap monitoring are discussed. The rest of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the methodology, Section 4 presents the experimental results, Section 5 provides discussion and analysis, and Section 6 concludes the paper.

2. Related work

The rise of deep convolutional neural networks (CNNs) has fundamentally changed image classification. AlexNet, proposed by Krizhevsky et al., won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a wide margin, marking the start of the deep learning era [4]. ResNet subsequently addressed the vanishing-gradient problem in deep networks through residual connections, enabled the training of networks with more than 100 layers, and achieved breakthroughs on multiple benchmarks [6]. EfficientNet adopts a compound scaling strategy that balances network depth, width and resolution, greatly reducing computational cost while maintaining high accuracy.

However, a core limitation of these methods is their strong dependence on expert knowledge and manual tuning. For ResNet, for example, the number of layers, the learning rate schedule (such as warmup and cosine decay) and the configuration of batch normalization all affect final performance, and these hyperparameters interact in complex ways, so the best combination is difficult to find by simple grid search [4]. In animal image classification, early work on the Oxford-IIIT Pet dataset showed that handcrafted features (HOG+SIFT+SVM) reached an accuracy of only about 66% in fine-grained breed recognition, which illustrates the inherent difficulty of the task [7].

DOI: 10.54254/2755-2721/2026.LD39275

AutoML aims to automate the manual decision points in the machine learning workflow through algorithmic means. It mainly comprises three core modules: Neural Architecture Search (NAS), Hyperparameter Optimization (HPO) and meta-learning [5]. NAS explores the optimal network structure within a predefined search space using reinforcement learning, evolutionary algorithms or gradient-based optimization. HPO uses techniques such as Bayesian optimization and random search to adjust training parameters such as the learning rate and batch size automatically. Meta-learning extracts transferable knowledge from past tasks to accelerate adaptation to new ones. The review by He et al. indicates that AutoML is evolving toward end-to-end automation, multi-objective optimization and broader task adaptability [5].

Among commercial AutoML platforms, Google AutoML Vision, which builds on advanced models from the Google Brain team (such as the Vision Transformer series), performs well in large-scale image recognition, but its access threshold and cost are relatively high. H2O.ai is known for its interpretability and flexible deployment options, and suits structured-data scenarios that demand model transparency, such as finance and healthcare. Baidu EasyDL emphasizes a zero-code experience, provides a Chinese-language interface and supports localized model deployment; it integrates AutoDL and AutoML technologies to select the optimal network architecture automatically, which makes it especially suitable for users without a deep learning background. Recent studies report that EasyDL has been applied successfully to CT image diagnosis of acute pancreatitis, showing good performance in medical image classification, although the black-box nature of the platform remains its main limitation.

AutoML has accumulated a number of application cases in image analysis tasks, including ecological monitoring. Li et al. introduced AutoML into an image analysis system for power UAV inspection and improved the generalization ability and automation level of image processing through an automated deep learning framework [8]. Liu explored the use of AutoML for species recognition in infrared-camera images from protected areas and showed that, in specific ecological monitoring scenarios, AutoML can reach recognition accuracy comparable to manual tuning while significantly reducing labor costs [9]. In addition, Norouzzadeh et al. used deep learning to identify and count wild animals automatically in camera-trap images, demonstrating the great potential of computer vision for ecological monitoring [2]. However, most of these studies focus on specific application scenarios and lack a systematic evaluation of the cross-dataset adaptability of AutoML on standard animal image classification benchmarks, which is the starting point of this study.

3. Methodology

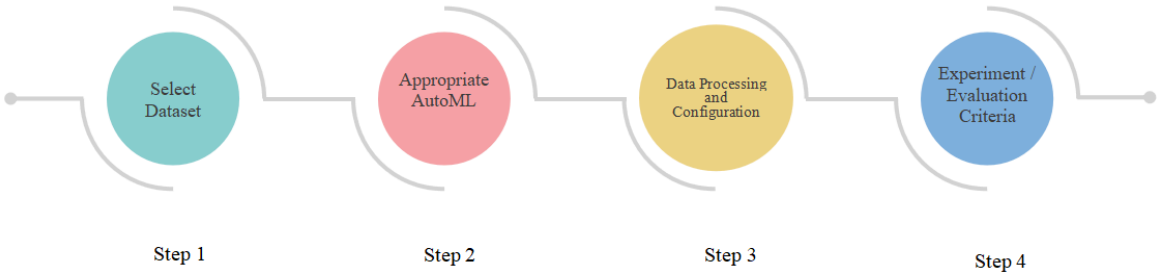


Figure 1. Overall workflow of this study

As illustrated in Figure 1, the research pipeline of this study consists of four sequential steps. Step 1 (Dataset Selection) selects two animal image datasets of contrasting difficulty, Oxford-IIIT Pet for fine-grained classification and Animals-10 for coarse-grained classification, to enable cross-

DOI: 10.54254/2755-2721/2026.LD39275

scenario comparison. Step 2 (AutoML Platform Selection) adopts Baidu EasyDL as the experimental vehicle, owing to its zero-code workflow and its integration of NAS, HPO and automatic data augmentation, which are detailed in Section 3.1. Step 3 (Data Processing and Configuration) covers data upload, automatic labeling and training configuration under a unified hardware and parameter setting, as described in Sections 3.2 and 3.3. Step 4 (Experiment and Evaluation) trains the models and evaluates them using Top-1 accuracy, precision, recall, F1-score and the confusion matrix; the results are reported in Section 4. Because both experiments follow the same procedure and configuration, the observed performance gap is not caused by procedural inconsistency. It does, however, reflect all differences between the datasets (class granularity, sample size, class balance and image sources), as discussed in Section 5.

3.1. Core mechanisms of AutoML

The core value of AutoML lies in automating the key decision points of the machine learning workflow that normally require expert involvement. The Baidu EasyDL platform used in this study integrates the following three types of automated modules:

Neural Architecture Search (NAS): EasyDL's accuracy-improvement configuration package is based on the Baidu Wenxin-CV VIMER-CAE pre-trained model and automatically searches a candidate network-structure space to select a suitable feature extraction backbone for a given task. Compared with manually choosing a single network (such as ResNet-50 or MobileNetV3), NAS can adjust the network configuration to the characteristics of the dataset (number of classes, image size, sample size, etc.), achieving a better balance between accuracy and efficiency.

Automatic Hyperparameter Search (HPO): In this study, EasyDL's automatic hyperparameter search function is enabled in the training configuration. This module automatically explores the best combination of key hyperparameters, such as the learning rate, batch size and number of training epochs, replacing traditional manual grid search or random search. It allows training to converge without manual intervention, which is one of the main ways in which AutoML lowers the technical threshold.

Automatic Data Augmentation: EasyDL provides an automatic augmentation strategy based on dataset characteristics. According to the distribution of the training data, it selects a combination of image transformations (such as horizontal flipping, color jittering and random cropping) to improve the generalization ability of the model and reduce the risk of overfitting when samples are limited.

3.2. Introduction to datasets

This study selects two standard animal image datasets with markedly different characteristics to support a cross-scenario comparison:

Oxford-IIIT Pet Dataset: This dataset was jointly constructed by the Visual Geometry Group (VGG) of the University of Oxford and the International Institute of Information Technology, Hyderabad, and is a classic benchmark for fine-grained animal image classification [7]. It contains 37 cat and dog breeds (12 cat breeds and 25 dog breeds) with about 200 images per class, 7393 images in total. A small number of corrupted images were filtered out during upload, leaving 7354 valid samples for training. The images vary considerably in scale, pose and illumination, and the appearance differences between some breeds are small (for example, the facial features of Persian and Ragdoll cats), which makes the dataset a demanding benchmark for fine-grained classification. In this experiment, 7354 training images were used, the deployment method was the public cloud API, and training took 24 hours and 3 minutes.

DOI: 10.54254/2755-2721/2026.LD39275

Animals-10 Dataset: This dataset comes from Kaggle and contains 10 common animal classes (dog, cat, horse, elephant, butterfly, chicken, cow, sheep, spider and squirrel), with more than 26,000 images in total and a somewhat imbalanced class distribution. Compared with Oxford-IIIT Pet, the morphological differences between classes are more obvious (for example, elephants and butterflies look very different), and the images come from diverse sources, including natural scenes collected from the Internet with complex backgrounds. In this experiment, 26176 training images were used, and training took 24 hours and 4 minutes.

The key difference between the two datasets is that Oxford-IIIT Pet tests the fine-grained discrimination ability of AutoML (high inter-class similarity), whereas Animals-10 tests its coarse-grained classification ability (clear inter-class differences, but a larger sample size and more diverse image sources). This complementary design makes the conclusions more broadly informative.

3.3. Experimental setup

This study conducts two groups of experiments under the Baidu EasyDL image classification module, with identical training configurations to ensure comparability. The settings are as follows:

Platform Configuration: The deployment method is the public cloud API. The training method is the "accuracy improvement configuration package" (EasyDL's enhanced AutoML training mode), with the Baidu Wenxin-CV VIMER-CAE pre-trained model (general scenario) as the underlying algorithm. The network search slider is set to the "highest accuracy" level, and the automatic hyperparameter search function is enabled.

Hardware Environment: The training environment is a GPU P4 (Tesla GPU_P4_8G with 8 GB video memory, single card, 12-core CPU, 40 GB memory, 5.5 TeraFLOPS).

Data Processing: All datasets were uploaded as ZIP archives and labeled automatically using folder names as labels, so no image-by-image manual annotation was needed. The data augmentation strategy followed the platform's default automatic configuration, which selects a combination of augmentations according to dataset characteristics.

Evaluation Metrics: The following metrics are used to evaluate model performance: (1) Top-1 accuracy, the proportion of samples whose predicted class matches the true class; (2) precision, for a given class, the proportion of samples predicted as that class that truly belong to it; (3) recall, for a given class, the proportion of samples truly belonging to it that the model identifies correctly; (4) F1-score, the harmonic mean of precision and recall, $F1 = 2\times(Precision\times Recall)/(Precision+Recall)$; and (5) the confusion matrix, which shows how predictions are distributed across classes and highlights confusable class pairs.

4. Experiments and results

4.1. Overall performance comparison

As shown in Table 1, both groups of experiments achieved strong classification results. EasyDL achieves a Top-1 accuracy of 95.0% and an F1-score of 95.1% on the Oxford-IIIT Pet dataset (37 classes, 7354 images), with a precision of 95.5% and a recall of 95.2%. On the Animals-10 dataset (10 classes, 26176 images), it reaches an accuracy of 99.3% and an F1-score of 99.2%, with both precision and recall at 99.2%.

DOI: 10.54254/2755-2721/2026.LD39275

Table 1. Overall performance of the two experiments

Dataset	Number of classes	Number of training images	Training duration	Top-1 Accuracy	Precision	Recall	F1-score
Oxford-IIIT Pet	37	7354	24h 3min	95.0%	95.5%	95.2%	95.1%
Animals-10	10	26176	24h 4min	99.3%	99.2%	99.2%	99.2%

Notably, the training duration is about 24 hours for both datasets. This is probably related to the design of the automatic search in EasyDL's accuracy-improvement configuration package, which appears to perform architecture search and hyperparameter optimization within a fixed computing budget; the platform does not disclose this mechanism, however, so this remains an interpretation. Although Animals-10 contains about 3.5 times as many images as Oxford-IIIT Pet, it has only 27% as many classes and is a less demanding task, which may explain why similar training times yielded a much higher accuracy.

4.2. Detailed performance analysis by class

At the class level on Oxford-IIIT Pet, the model evaluation report reveals clear patterns for fine-grained classification. The poorly performing classes are concentrated among visually similar breeds: american_pit_bull_terrier (F1-score 76%) and staffordshire_bull_terrier (F1-score 77%) are clearly confused with each other (Figure 3b). These two breeds are extremely similar in body shape, coat color and facial structure, and even human observers find them hard to tell apart. The next-lowest classes in the report, Burmese (86%), Ragdoll (87%) and Bengal (87%), score roughly ten percentage points higher.

The EasyDL evaluation report also indicates that differences in image aspect ratio have a relatively large effect on the F1-score (impact difference of 0.0474) and on accuracy (0.0256), and it recommends configuring the ShearX and ShearY data augmentation strategies (Figure 3a). This suggests that the shooting angle and cropping of images should not be ignored in animal image classification.

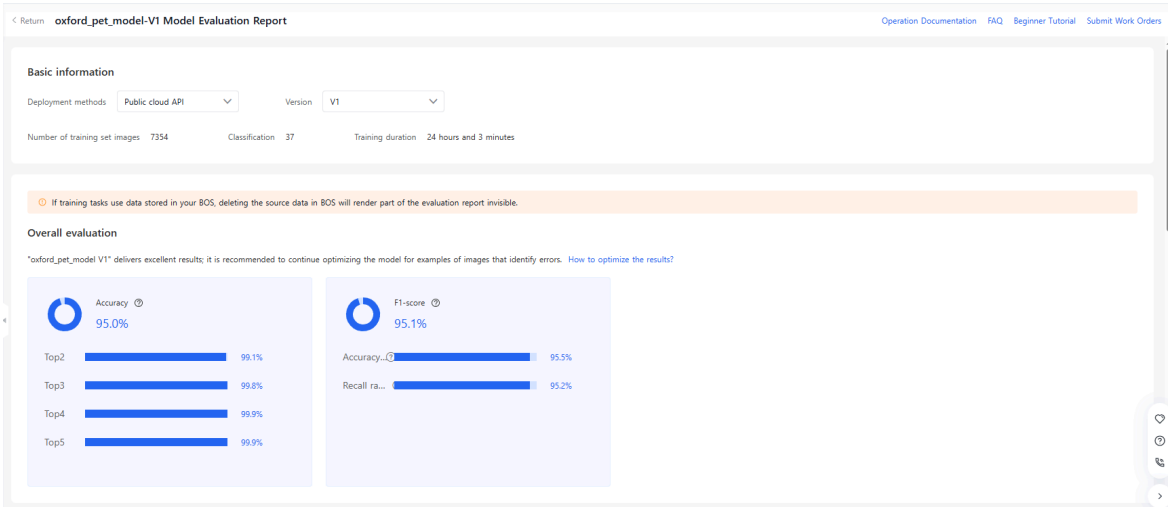


Figure 2. Overall model evaluation results for the Oxford-IIIT Pet dataset (accuracy 95.0%, F1-score 95.1%)

DOI: 10.54254/2755-2721/2026.LD39275

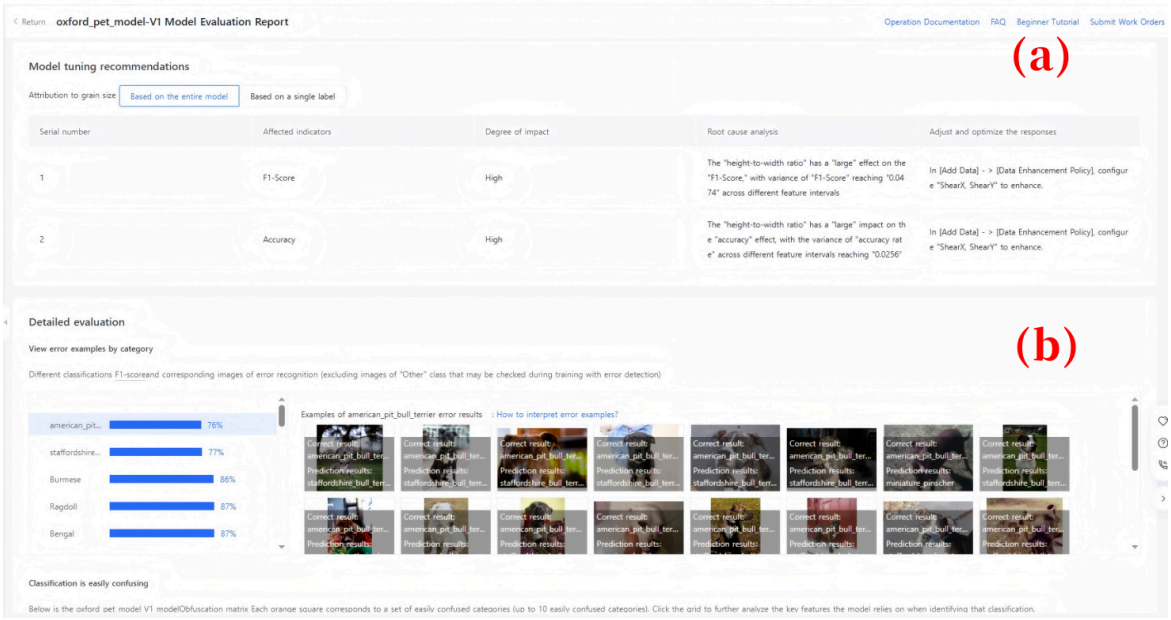


Figure 3. Model evaluation report for the Oxford-IIIT Pet dataset: (a) model tuning suggestions; (b) F1-score of each class and error examples

On Animals-10, sheep (folder name pecora, F1-score 98%) and cow (mucca, 98%) are the two lowest-scoring classes and are mainly confused with each other: some sheep are misclassified as cows (Figure 5), possibly because of similar body structure and grassland backgrounds. The F1-scores of dog, chicken, horse, squirrel, elephant and spider all reach 99%, which shows that AutoML performs very well when the classes are clearly different. Because the accuracy on Animals-10 is so close to perfect, the system cannot generate an informative confusion matrix or meaningful tuning suggestions, which is consistent with this task being relatively easy for EasyDL.

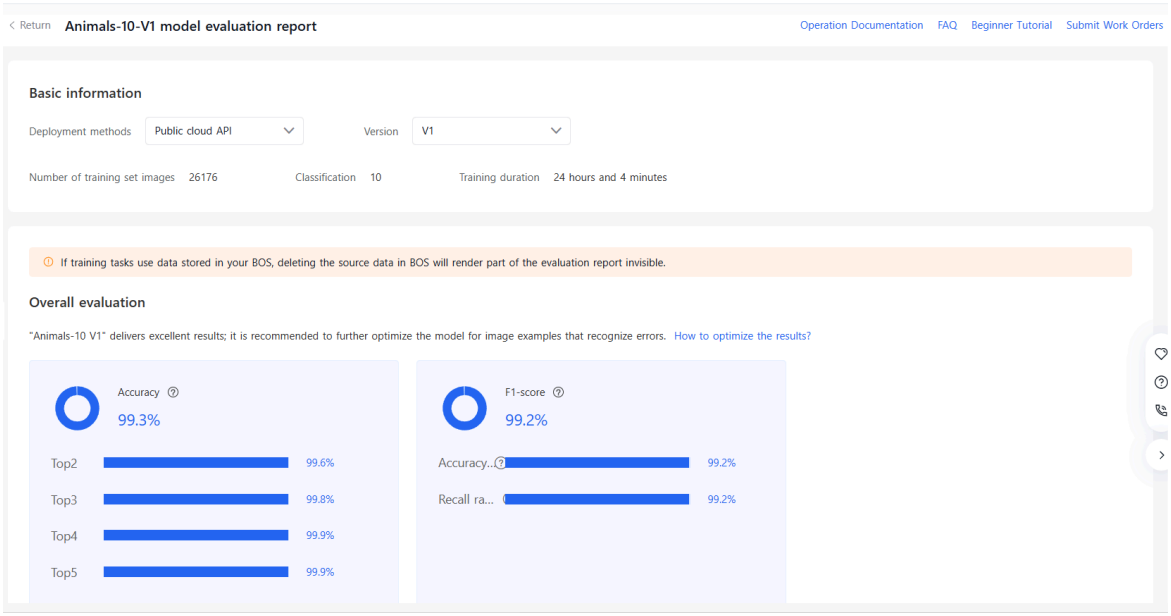


Figure 4. Overall model evaluation results for the Animals-10 dataset (accuracy 99.3%, F1-score 99.2%)

DOI: 10.54254/2755-2721/2026.LD39275

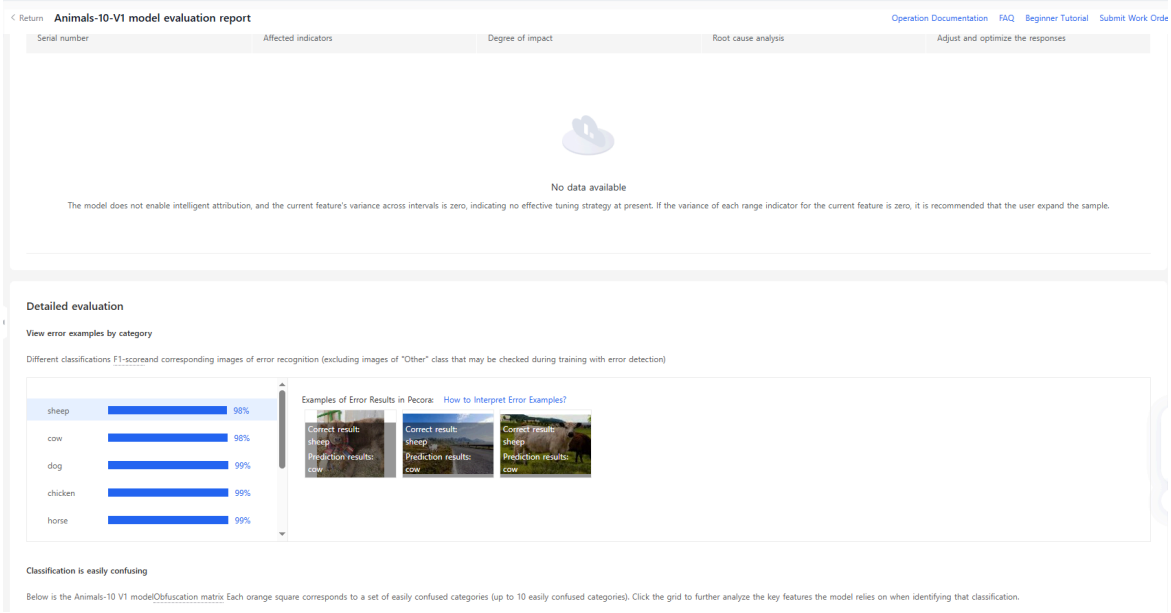


Figure 5. F1-score of each class and error examples for the Animals-10 dataset

5. Discussion

5.1. Advantages of AutoML

The results of this study support two core advantages of AutoML for animal image classification.

First, AutoML substantially lowers the technical threshold. The whole experimental process (data upload → model training → result evaluation) required no coding. Labels were generated automatically from the folder structure, and the training parameters were searched and configured by the platform. This means that researchers in ecology, zoology and related fields can build high-performance classification models without specialized deep learning knowledge, much as medical researchers have used Baidu EasyDL to build diagnostic models for acute pancreatitis imaging without understanding the underlying neural networks.

Second, the automated mechanisms provide a performance guarantee. EasyDL's accuracy-improvement configuration package, based on the large-scale pre-trained VIMER-CAE model and combined with automatic hyperparameter search, achieves 95.0% accuracy on Oxford-IIIT Pet, a recognized challenging benchmark for fine-grained classification. For reference, early methods based on handcrafted features (HOG+SIFT+SVM) reached an accuracy of only about 66% on this dataset [7]; although the comparison is indirect, it reflects the benefit of combining AutoML with large pre-trained models. The 99.3% accuracy on Animals-10 further shows that, when classes are clearly different, AutoML can reach an accuracy above 99% with minimal manual intervention.

5.2. Limitations of AutoML

Although the results are satisfactory, this study also reveals several limitations of AutoML in animal image classification.

Limited interpretability is a common weakness of commercial AutoML platforms. The end-to-end model generated by EasyDL operates as a black box: it cannot output feature visualizations or explain the basis of its classification decisions. This is a practical obstacle for zoological and taxonomic studies that require explicit diagnostic features (such as the facial details that distinguish

DOI: 10.54254/2755-2721/2026.LD39275

primate subspecies), because researchers cannot derive a biological rationale for academic discussion from the model output.

The ability to handle confusable fine-grained classes is also limited. The systematic confusion between american_pit_bull_terrier and staffordshire_bull_terrier in the Oxford-IIIT Pet experiment (F1-scores of only 76% and 77%) suggests that, for classes with highly similar shapes, the preset model library and automatic augmentation strategy of EasyDL struggle to capture discriminative local features fully. In such cases, manual intervention (such as local feature enhancement or a custom attention mechanism) may be a necessary supplement.

In addition, because the platform is a black box, users cannot know which network architecture EasyDL actually uses for training. This limits the reproducibility of the results and the transparency required for scientific research.

Finally, the results should not be extrapolated directly to real-world camera-trap monitoring, which is the setting most relevant to many ecology and zoology researchers. Oxford-IIIT Pet and Animals-10 consist largely of well-framed, well-exposed photographs in which the animal is the main subject, whereas field images differ in several ways. Lighting is often uneven: images are captured at dawn, dusk and night, under strong backlight or deep shadow, or in infrared grayscale. Animals are frequently partly occluded by vegetation, terrain or other animals, or only partly inside the frame. Backgrounds are cluttered and noisy, and animals may be small, distant or camouflaged. Many triggered frames are empty, and rare species have very few samples. In addition, models trained on images from some camera locations often perform worse at new locations because backgrounds and illumination change [10]. The accuracies of 95.0% and 99.3% reported here therefore do not indicate expected field performance. Users who migrate this AutoML workflow to camera-trap data are advised to (i) build the training set from their own site images covering day and night, seasons and weather; (ii) include an explicit "empty/background" class; (iii) hold out a test set from camera locations not used for training; and (iv) inspect per-class results and error examples, because the black-box nature of the platform makes silent failures hard to detect. A detection-first workflow that localizes animals before classifying them is another option, but it was not evaluated in this study.

5.3. Future optimization directions

Based on the above analysis, future research can proceed in the following directions. First, for confusable classes (such as dog breeds with similar shapes), local feature focusing mechanisms (such as attention-based fine-grained classification modules) could be integrated into the AutoML framework to reduce systematic confusion. Second, interpretability components (such as Grad-CAM class activation mapping) could be embedded in the automated training pipeline so that the model outputs heatmaps highlighting the key recognition regions, meeting the traceability needs of scientific research. Third, the approach should be validated on real camera-trap datasets with uneven lighting, occlusion and noisy backgrounds, using test sets from unseen camera locations, to quantify the gap between benchmark and field performance. Fourth, multimodal AutoML systems could combine animal images with audio (such as bird vocalizations) or behavioral time-series data to build a more comprehensive recognition framework and extend AutoML further into ecological monitoring.

DOI: 10.54254/2755-2721/2026.LD39275

6. Conclusion

Using the Baidu EasyDL platform as the experimental vehicle, this study systematically evaluates AutoML for animal image classification on two datasets with markedly different characteristics: Oxford-IIIT Pet (37 classes, 7354 images) and Animals-10 (10 classes, 26176 images). EasyDL achieves Top-1 accuracies of 95.0% (Oxford-IIIT Pet) and 99.3% (Animals-10), with F1-scores of 95.1% and 99.2%, respectively, which supports the feasibility and practical value of AutoML for animal image classification.

The practical significance of this study is that, for ecology and zoology researchers who lack a deep learning background, EasyDL offers a low-threshold path from data upload to a high-performing model. The whole process requires no coding or manual parameter tuning, which considerably shortens the development cycle from data input to model output. The cross-dataset comparison also suggests that AutoML performance is closely associated with the class discriminability of the dataset. In the coarse-grained scenario with clear morphological differences between classes (Animals-10), EasyDL reaches an accuracy above 99% with minimal manual intervention. In the fine-grained scenario (Oxford-IIIT Pet), the automated mechanism brings a substantial gain in efficiency, but there is still room for improvement for highly similar class pairs. Because the two datasets also differ in size, class balance and image sources, this association should be confirmed by further controlled comparisons.

This study has several limitations. The black-box nature of EasyDL limits the reproducibility of the results, and there is no direct comparison with manually tuned baseline methods. In addition, both datasets consist of well-framed images, and the conditions typical of camera-trap monitoring, such as uneven lighting, occlusion and noisy backgrounds, were not evaluated, so the reported accuracy cannot be assumed to hold in field deployments. Future work can add control experiments with traditional methods and test on real camera-trap data to strengthen the conclusions. Overall, this study indicates that AutoML offers a low threshold, high efficiency and broad adaptability for animal image classification on standard benchmarks, and provides practical references for subsequent interdisciplinary applications and technical iteration.

References

[1] Christin, S., Hervet, R., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution, 10*(10), 1632-1644.

[2] Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences, 115*(25), E5716-E5725.

[3] Cheng, S. H., Liu, J., & Zhu, D. D. (2014). Water quality anomaly monitoring method based on computer vision and SVM. *Opto-Electronic Engineering, 41*(5), 1-6.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097-1105.

[5] He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems, 212*, 106622.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778).

[7] Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2012). Cats and dogs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3498-3505).

[8] Li, D. X., Wang, X. H., Li, L., et al. (2024). Automated deep learning system for power UAV inspection image analysis and processing: Architecture design and key technologies. *Electric Power Information and Communication Technology, 22*(4), 38-54.

DOI: 10.54254/2755-2721/2026.LD39275

[9] Liu, Y. (2019). *Application and implementation of AutoML technology in species recognition by infrared cameras in protected areas* [D]. University of Chinese Academy of Sciences.
[10] Beery, S., Van Horn, G., & Perona, P. (2018). Recognition in terra incognita. In *Proceedings of the European Conference on Computer Vision (ECCV)*.