Open-set Animal Re-identification via Multilevel Feature Fusion

Notebook for the <AnimalCLEF> Lab at CLEF 2025

Jingyin Tan[1], Aiguo Wang[1,*]

[1]Foshan University, Foshan, Guangdong, China

Abstract

The AnimalCLEF2025 task aims to train a recognizer on the training set with good generalization ability for predicting whether the category of an animal image in the test dataset belongs to a known class or an unknown class, which is an open-set recognition problem. In this study, the fusion of deep features and low-level features is utilized to build an open-set animal recognition model. Specifically, the pretraining and fine-tuning scheme is adopted to learn high-level features, where the swin-base-patch4-window7-224 model is fine-tuned with the training set and deep features are then extracted from its feature representation layer. Second, we utilize two keypoint detection and descriptor extraction networks to obtain two sets of low-level features. Afterwards, the nearest-neighbor classification rule is utilized with the weighted multilevel features to infer the label of a test sample. Particularly, if the maximal similarity between the test sample and training samples is lower than a threshold, the test sample is predicted as unknown classes. Finally, experimental results show that the proposed model obtains 0.55826 and 0.56044 balance accuracy on the 31% and 69% test dataset, respectively. Our source code is available at https://github.com/NickyTan8899/tjy.

Keywords

open-set recognition, multilevel feature, pretrained model

1. Introduction

Open-set recognition (OSR) refers to the problem where a model is trained on a set of known classes (the closed set), but during testing, it should correctly classify known categories and identify instances from unknown classes as unknown. This setting better reflects real-world applications than traditional closed-set recognition that assumes that all test samples belong to the known classes[1]. In many real-world scenarios, machine learning systems frequently encounter inputs from previously unseen categories, such as the AnimalCLEF2025 task[2, 3] and image classification systems encountering objects or species not seen during training.

Accordingly, researchers have explored and designed a variety of models towards enhanced open-set recognition accuracy. Unlike traditional closed-set classifiers, OSR methods incorporate mechanisms to detect novel instances. Approaches to OSR can be broadly categorized into discriminative models, generative models, and distance-based methods[4]. Discriminative methods, such as energy-based models, modify neural networks to estimate the confidence of a sample belonging to a known class. Generative approaches such as including variational autoencoders and GAN-based[5] techniques attempt to model the data distribution of known classes and use reconstruction errors or synthetic unknowns to infer unfamiliar inputs. Distance-based methods operate in the feature space and use metrics such as nearest neighbors[6] and class prototypes to detect out-of-distribution samples.

For the problem of open-set animal recognition, though great progresses have been made, most of existing methods still suffer from degraded performance due to complex spatial dependencies and background information[7, 8]. To this end, we in this study propose a distance-based open-set animal recognition model via multilevel feature fusion to better capture multi-view information of the image. The main contributions of our work are as follows.

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

tjynicky@gmail.com (J. Tan); agwang@fosu.edu.cnïijŒ (A. Wang)

(1) Deep features and low-level features are extracted. Particularly, deep features are learnt under the pretraining and fine-tuning scheme, where the swin-base-patch4-window7-224 model is utilized and fine-tuned. Keypoint detection and descriptor extraction networks are also used to extract features.

(2) The open-set recognition is performed with the multilevel features. The similarity between a test sample with training samples are measured. Specifically, we first calculate the similarity from the view of deep features and low-level features respectively, and then use the Wildfusion strategy to obtain the final similarity score[9]. The prediction is made according to the similarity score. If the maximal similarity is lower than a threshold, we categorize it into novel classes; otherwise, we report the class of training sample corresponding to the maximal similarity.

(3) Comparative experiments are conducted against two baseline models on the competition datasets. Results show that the proposed model performs better than the baselines and obtains 0.55826 and 0.56044 balance accuracy on 31% and 69% test dataset, respectively, with the best result ranked 34th on the leaderboard (Team Name: Already mygo).

The structure of this paper is as follows. Section 2 introduces the proposed open-set animal recognition model. Section 3 presents experimental datasets and experimental setup. Section 4 presents experimental results, followed by the conclusion section.

2. Methodology

Figure 1 presents the proposed recognition framework that mainly consists of the training stage and test stage. As for the training stage, three different networks are used to extract three sets of features. That is, each training sample can be encoded with the three types of features. During the test phase, we first use the three networks to obtain feature representations of a test sample and then calculate the similarity between the test sample and each training sample with Wildfusion strategy. Finally, we make predictions based on the similarity value.

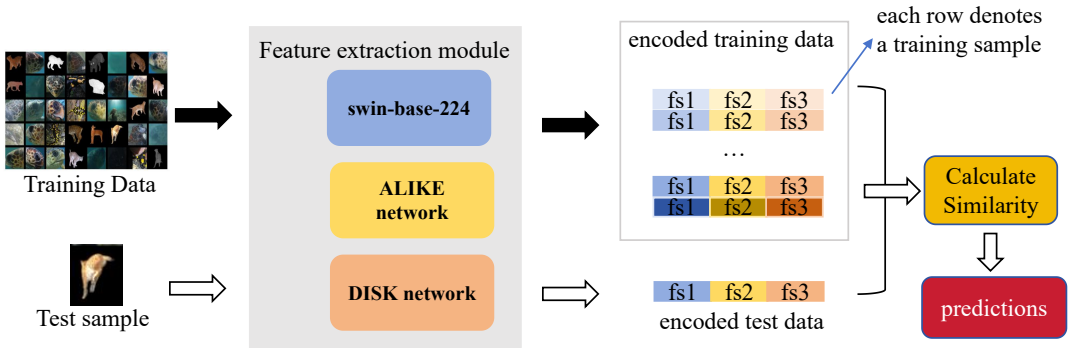

Figure 1: The schematic diagram of the proposed model

2.1. Extraction of Features

As for deep features, we adopt the pretrain and fine-tuning scheme to learn latent features. Specifically, a Swin Transformer-based pretrained model (i.e., swin-base-patch4-window7-224 model)[10] is adopted. We replace its cross-entropy loss with ArcFace loss that enhances the angular separation between different classes in the embedding space[11]. Afterwards, we fine-tune the pretrained model on the training set. Finally, we drop its classification layer to obtain a feature learning network, which enables us to obtain the feature representations of an image. For clarity, we denote the set of features as fs1.

Besides, we directly apply two different keypoint and descriptor extraction networks (i.e., ALIKED and DISK) on the image to obtain two sets of features. Specifically, ALIKED[12] utilizes the sparse deformable descriptor head to extract deformable descriptors, where a neural reprojection error loss

is used to measure the discrepancy between reprojection and descriptor-matching probabilities[13]. DISK[14] uses reinforcement learning to train end-to-end pipeline for keypoint detection and matching. We denote the two sets of features as fs2 and fs3.

2.2. Open-set Recognition Procedure

In classifying a test sample, we adopt the nearest-neighbor classification rule to make predictions. First, we encode the test sample and each of the training samples with fs1, fs2, and fs3. Second, we measure the similarity between test sample and each training sample. Particularly, to reflect the importance of the three different types of features, the Wildfusion strategy is adopted, with which we can get the similarity between test sample and each training sample and further obtain the maximal similarity. If the maximal similarity is greater than a predefined threshold, the predicted result of the test sample is the label of the associated training sample; otherwise, the predicted label is "unknown classes" (or new_individual used in the competition).

3. Experimental Setup

3.1. Dataset

The AnimalCLEF2025 dataset, having 15209 images, is divided into database dataset (having 13074 images) and query dataset (having 2135 images). The database plays the role of training set and the query serves as the test set. Table 1 presents the summary of experimental data.

Table 1
Basic information of AnimalCLEF2025 dataset

Species	Split	#number	#total
SeaTurtleID2022	database	8729	13074
LynxID2025	database	2957	
SalamanderID2025	database	1388	
SeaTurtleID2022	query	946	2135
LynxID2025	query	689	
SalamanderID2025	query	500	

3.2. Experimental Setup

To obtain deep features, we first resize the images to 224 x 224 to make it suitable for the pretrained Swin-based model. We then fine-tune it on the database dataset with the SGD optimizer. An initial learning rate 0.001 is used and a cosine annealing learning rate scheduler is utilized, which adjusts the learning rate following a cosine curve from the initial value decreased to 1e-6 in the end. The model is trained for 100 epochs. Meanwhile, the batch size and the number of workers for loading data are 64 and 2 respectively. Finally, the deep features are embedded to be a 1 x 1024 vector for each sample. As for low-level features, we resize the images to 256 x 256 and then obtain two sets of features, returned by ALIKED and DISK, respectively.

Besides, for comparison, three baseline methods including MegaDescriptor-L-384 (swin-large-patch4-window12-384 model), MegaDescriptor-L-384-ALIKED (a combination of MegaDescriptor-L-384 and ALIKED) and MegaDescriptor-L-384-ALIKED-DISK (a combination of MegaDescriptor-L-384, ALIKED and DISK) are utilized to demonstrate the effectiveness of our proposed model.

As for the performance metric, balance accuracy (denoted by score in formula (1)) is used, which is calculated by the balanced accuracy on known samples (BAKS) and balanced accuracy on unknown samples (BAUS).

$$\text{score} = \sqrt{\text{BAKS} \times \text{BAUS}}. \tag{1}$$

, where BAKS is the accuracy of individuals that are known in the database, and BAUS is the accuracy of individuals that are unknown in the database. This formulation avoids misleadingly high scores from trivial models, such as those predicting all samples as unknown, which would score 0% BAKS and 100% BAUS. Unlike the arithmetic mean, the geometric mean penalizes such imbalance, providing a more robust metric.

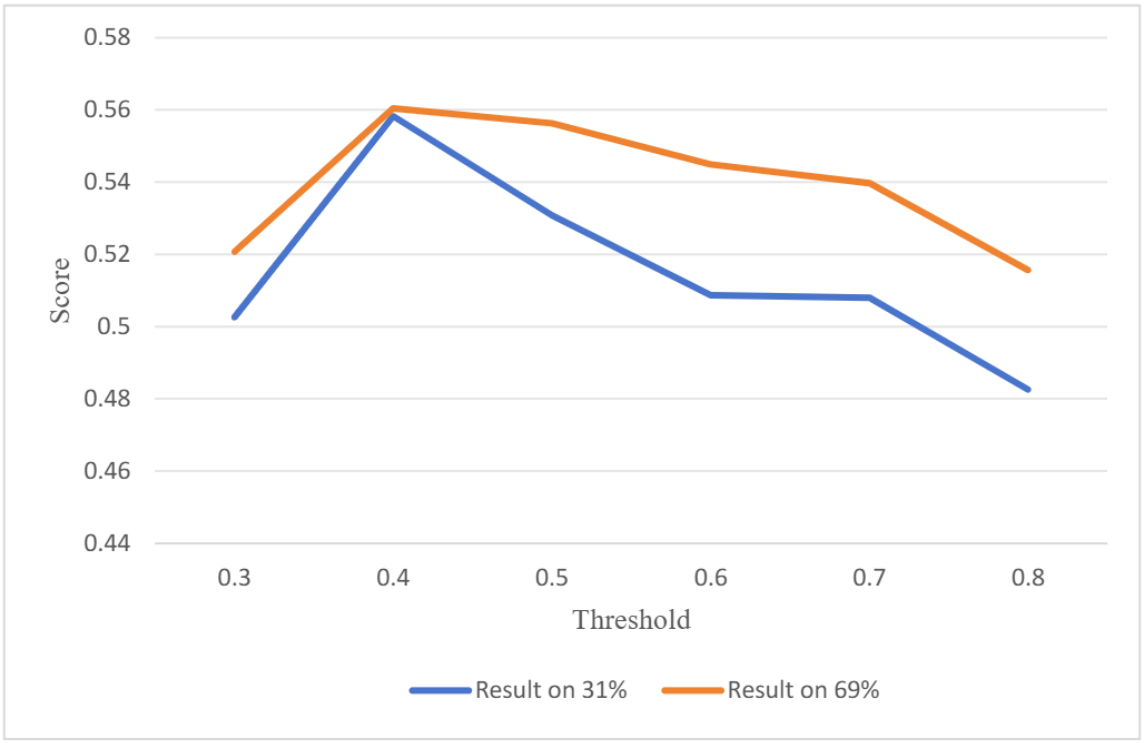

Figure 2: Experiment results of different thresholds

4. Experimental Results

Table 2 and Table 3 present the experimental results on the query dataset. From Tables 2 and 3, we can observe that the proposed model outperforms its competitors. Specifically, for the case of 31% query dataset, our model obtains the score of 0.55826, while its competitors obtain the scores of 0.30002, 0.36555 and 0.51134, respectively. For the case of 69% query dataset, our model obtains the score of 0.56044, while its competitors obtain the scores of 0.30898, 0.44362 and 0.53467, respectively.

Table 2
Experimental results on the 31% query dataset

model	score
MegaDescriptor-L-384	0.30002
MegaDescriptor-L-384-ALIKED	0.36555
MegaDescriptor-L-384-ALIKED-DISK	0.51134
Ours	0.55826

Furthermore, to investigate the impact of different thresholds on the recognition performance, we conduct experiments with the candidate thresholds ranging from 0.3 to 0.8. Figure 2 presents the results,

Table 3
Experimental results on the 69% query dataset

model	score
MegaDescriptor-L-384	0.30898
MegaDescriptor-L-384-ALIKED	0.44362
MegaDescriptor-L-384-ALIKED-DISK	0.53467
Ours	0.56044

from which we can observe a general trend that the score first increases and then decreases along with the increase of the threshold. We also observe that the use of 0.4 generally leads to better performance.

5. Conclusion

Towards higher image-based open-set animal recognition, we in this study propose a distance-based recognition model that utilizes several sets of features. The deep features are learnt with the pretraining and fine-tuning scheme and the low-level features are obtained by the keypoint detection and descriptor extraction networks. To reflect the importance of different types of features in calculating the similarity between samples, a weighing scheme is used. Afterwards, a threshold-based method is adopted to infer the label of a test sample. Finally, the proposed model is evaluated on the test set and results demonstrate its effectiveness.

Declaration on Generative AI

During the preparation of this work, the authors used OpenAI-GPT-4o in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. Wolf, P. Thelen, J. Beyerer, Poison-aware open-set fungi classification: Reducing the risk of poisonous confusion, Working Notes of CLEF (2024).

[2] L. Adam, K. Papafitsoros, R. Kovář, V. Čermák, L. Picek, Overview of AnimalCLEF 2025: Recognizing individual animals in images, 2025.

[3] L. Picek, S. Kahl, H. Goëau, L. Adam, T. Larcher, C. Leblanc, M. Servajean, K. Janoušková, J. Matas, V. Čermák, K. Papafitsoros, R. Planqué, W.-P. Vellinga, H. Klinck, T. Denton, J. S. Cañas, G. Martellucci, F. Vinatier, P. Bonnet, A. Joly, Overview of lifeclef 2025: Challenges on species presence prediction and identification, and individual animal identification, in: International Conference of the Cross-Language Evaluation Forum for European Languages (CLEF), Springer, 2025.

[4] C. Geng, S.-j. Huang, S. Chen, Recent advances in open set recognition: A survey, IEEE transactions on pattern analysis and machine intelligence 43 (2020) 3614–3631.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Communications of the ACM 63 (2020) 139–144.

[6] J. Yang, K. Zhou, Y. Li, Z. Liu, Generalized out-of-distribution detection: A survey, International Journal of Computer Vision 132 (2024) 5635–5662.

[7] L. Adam, V. Čermák, K. Papafitsoros, L. Picek, Seaturtleid2022: A long-span dataset for reliable sea turtle re-identification, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 7146–7156.

[8] L. Picek, L. Neumann, J. Matas, Animal identification with independent foreground and background modeling, in: DAGM German Conference on Pattern Recognition, Springer, 2024, pp. 241–257.

[9] V. Cermak, L. Picek, L. Adam, L. Neumann, J. Matas, Wildfusion: Individual animal identification with calibrated similarity fusion, in: European Conference on Computer Vision, Springer, 2025, pp. 18–36.
[10] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.
[11] J. Deng, J. Guo, N. Xue, S. Zafeiriou, Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4690–4699.
[12] X. Zhao, X. Wu, W. Chen, P. C. Chen, Q. Xu, Z. Li, Aliked: A lighter keypoint and descriptor extraction network via deformable transformation, IEEE Transactions on Instrumentation and Measurement 72 (2023) 1–16.
[13] H. Germain, V. Lepetit, G. Bourmaud, Neural reprojection error: Merging feature learning and camera pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 414–423.
[14] M. Tyszkiewicz, P. Fua, E. Trulls, Disk: Learning local features with policy gradient, Advances in Neural Information Processing Systems 33 (2020) 14254–14265.