

Competition 4: Weakly-Supervised Fine-Grained Image Recognition

1. Competition Background

Fine-grained image recognition is an important branch of computer vision and pattern recognition, aiming to accurately identify subcategories of objects. For example, in a dog classification task, distinguishing between “Husky” and “Alaskan Malamute.” Fine-grained recognition is more challenging than traditional image classification because the differences between subcategories are often small, while intra-class variations can be large. In recent years, deep learning-driven machine learning algorithms have achieved significant progress across various domains. These successes heavily rely on large-scale, high-quality labeled data. This dependence is particularly pronounced in fine-grained image recognition tasks, as precise labeling often requires domain expertise. However, manually annotating a large amount of high-quality labels is both difficult and costly, limiting the scalability of fine-grained recognition models in practical applications. Therefore, training fine-grained recognition models using freely available web data has attracted increasing attention from researchers. However, web data often suffers from label noise, data bias, and long-tail distributions. When deep networks with strong fitting capabilities are trained on such data, significant overfitting often occurs, resulting in poor generalization on test data.

This competition aims to explore methods for improving model robustness in noisy data environments through the development of weakly supervised fine-grained recognition models. It encourages the use of web data to train efficient fine-grained recognition models, reducing reliance on large-scale manually labeled datasets while enhancing the practicality and scalability of the algorithms.

2. Competition Application Scenarios

As a cutting-edge direction in the field of computer vision, weakly supervised fine-grained image recognition demonstrates tremendous potential in both academia and industry due to its strong data-driven capabilities and precise recognition of fine-grained features. By effectively leveraging low-quality web-crawled data for robust training, this technology not only significantly reduces reliance on costly manually labeled datasets but also greatly enhances the model’s generalization and robustness, providing a reliable foundation for practical applications.

Weakly supervised fine-grained image recognition is valuable not only for academic research but also shows broad potential in real-world applications. From species monitoring to intelligent security, from medical imaging diagnosis to

autonomous driving, this technology brings substantial efficiency improvements and cost savings across various fields. By participating in this competition, teams can gain an in-depth understanding of the challenges and applications of webly supervised fine-grained recognition, advance the development of this technology in real-world scenarios, and provide innovative solutions to practical problems.

3. Competition Task

This competition focuses on webly supervised fine-grained image recognition. Participants are required to use the fine-grained web image training dataset provided by the organizers to train efficient fine-grained recognition models. The training dataset contains web-crawled data, which includes real-world challenges such as label noise, imbalanced data distribution, and small inter-class differences. Participants are expected to fully leverage these datasets and improve model performance in fine-grained classification tasks through optimized data processing, model architecture, and training strategies. Given the presence of label noise in web data, optimizing data quality is a key focus of this competition. Additionally, the model must demonstrate strong generalization ability, maintaining robust recognition performance across different datasets.

4. Dataset and Data Description

4.1 Data Source

The data comes from the publicly available datasets WebFG-496 and WebiNat-5089 from Nanjing University of Science and Technology. These are web-based fine-grained recognition datasets, covering categories such as airplanes, cars, birds, and various natural species. The training data are all crawled from the web, and the labels have not been manually verified, making the datasets highly representative of real-world scenarios.

4.2 Summary of the Dataset

Fine-Grained Web Dataset WebFG-496: This dataset contains 496 subcategories under three major categories: airplanes, cars, and birds. It includes 200 bird species (Web-Bird), 100 airplane types (Web-Aircraft), and 196 car models (Web-Car), with over 50,000 training images in total.

Fine-Grained Web Dataset WebiNat-5089: This dataset contains 5,089 subcategories of natural species, with over 1.1 million web-crawled training images. It is the largest webly supervised fine-grained dataset to date, covering multiple meta-categories such as plants, insects, birds, reptiles, fungi, protozoa, mollusks, and animals.

For the actual competition, the organizers sampled classes from WebFG-496 and WebiNat-5089, randomly selecting 400 and 5,000 classes, respectively. After data anonymization, these were provided as the final competition datasets WebFG-400 and WebiNat-5000. For evaluation, the organizers will use a uniformly distributed, human-accurately labeled dataset as the test set for accuracy comparison.

Ultimately, WebFG-400 contains 43,087 training images and 13,882 test images, while WebiNat-5000 contains 580,865 training images and 100,000 test images. The training data are organized into category folders (i.e., all images of category 0001 are placed in folder 0001). Sample data can be accessed at <https://pan.baidu.com/s/1I1oXGinmXlqM78-514bS7g?pwd=dih6>, and the official datasets will be available for download after registration.

The datasets WebFG-400 and WebiNat-5000 will be divided into preliminary stage data (Track A) and semifinal stage data (Track B) for phased competition evaluation.

5. Algorithm Design Requirements

Participants are encouraged to propose innovative deep learning algorithms or improve existing ones to enhance the accuracy and robustness of models when trained using web data. The algorithms should be highly scalable and maintain stable performance when handling datasets of different sizes.

If pretrained models are used during the competition, only models pretrained on ImageNet-1k are allowed; no additional data may be used. To encourage algorithmic innovation, model ensembling is not permitted in this competition, and the final submission must contain only a single model. After the semifinal stage, the organizers will reproduce the submitted code for each team, and any team whose model performance cannot be reproduced using the provided datasets will have their results invalidated.

6. Performance Metrics Requirements

The performance score in this competition is evaluated using Accuracy, a widely used metric in image classification tasks. There are no restrictions on model inference time or model size in this competition.

Test Accuracy is defined as:

Accuracy = Number of correctly predicted samples in the test set / Total number of samples in the test set

7. Functional Requirements

Participants are required to design innovative methods or improve existing

algorithms to train fine-grained recognition network models using the web data training sets provided by the organizers. The final models should demonstrate strong accuracy and robustness on the organizers' test datasets, which have been precisely annotated by humans. The methods designed by participants must exhibit good generalization and versatility across the different datasets provided by the organizers.

8. Development Environment

Participants are required to develop their solutions using the Python programming language and may utilize open-source deep learning frameworks, such as PyTorch.

9. Evaluation Criteria

The competition consists of three stages: the preliminary round, the semifinal, and the final. In the preliminary round, the test set is automatically scored by the system using Accuracy, but the results do not count toward the final score. It is intended only for participants to verify and debug their algorithms. The algorithm performance score is calculated as the average accuracy on the WebFG-400 (A-subset) and WebiNat-5000 (A-subset) datasets.

In the semifinal, the test set is also automatically scored using Accuracy, and the results contribute to the final score. The algorithm performance score is calculated as the average accuracy on the WebFG-400 (B-subset) and WebiNat-5000 (B-subset) datasets.

For the offline final, the final score is computed as 70% objective score + 30% subjective score. The objective score comes from the standardized machine-evaluated scores from the semifinal, while the subjective score is derived from the standardized scores of the oral defense. The oral defense evaluates participants' presentation performance as well as the quality of their technical solutions and code documentation. Standardization is performed by setting the highest score to 100, with other scores adjusted proportionally. (Note: the 30% subjective score does not apply to the second and third prize winners at the national level.)

10. Reference Resources

Mengmeng Sheng, Zeren Sun, Gensheng Pei, Tao Chen, Haonan Luo, Yazhou Yao,

“Enhancing Robustness in Learning with Noisy Labels: An Asymmetric Co-Training Approach”, ACM International Conference on Multimedia (ACM MM), 2024.

Mengmeng Sheng, Zeren Sun, Tao Chen, Shuchao Pang, Yucheng Wang, Yazhou Yao, “Foster Adaptivity and Balance in Learning with Noisy Labels”, European

Conference on Computer Vision (ECCV) , 2024.

Zeren Sun, Yazhou Yao, Xiu-Shen Wei, Yongshun Zhang, Fumin Shen, Jianxin Wu, Jian Zhang, and Heng Tao Shen, “Webly Supervised Fine-Grained Recognition: Benchmark Datasets and An Approach”, IEEE International Conference on Computer Vision (ICCV) , 2021.

Chuanyi Zhang, Yazhou Yao, Xing Xu, Jie Shao, Jingkuan Song, Zechao Li, Zhenmin Tang, “Extracting Useful Knowledge form Noisy Web Images via Data Purification for Fine-Grained Recognition”, ACM International Conference on Multimedia (ACM MM) , 2021.

Zeren Sun, Xian-sheng Hua, Yazhou Yao, Xiu-shen Wei, Guosheng Hu, Jian Zhang, “CRSSC: Salvage Reusable Samples from Noisy Data for Robust Learning”, ACM International Conference on Multimedia (ACM MM) , 2020.

11. Submission Requirements

11.1 Preliminary Round Submission Requirements

Participants are required to make predictions on the preliminary test set provided by the organizers and save the results in a **CSV file**. Each row in the CSV file should contain two elements: the first element is the image filename, and the second element is the class label (a four-digit number, with leading zeros added if necessary). The output format should follow this example:

xxxxxxxxxxxxx.jpg, 0000

xxxxxxxxxxxxxy.jpg, 1111

xxxxxxxxxxxxxz.jpg, 0812

.....

The organizers will automatically calculate the performance score of the participants' models based on this result file.

Notes:

- 1.Ensure that the submitted CSV file strictly follows the format specified above; otherwise, the submission may be considered invalid.
- 2.The image filenames must exactly match those in the test set, including letter case and file extension, to avoid matching errors.
- 3.The class labels must be four-digit numbers, with leading zeros added if

necessary, to prevent matching errors.

4. For the WebFG-400 dataset, generate a prediction results file named `pred_results_web400.csv`; for the WebiNat-5000 dataset, generate a file named `pred_results_web5000.csv`. Both CSV files should then be compressed into a single ZIP file for submission.

11.2 Semifinal Submission Requirements

Participants are required to make predictions on the semifinal test set provided by the organizers. The output format should be the same as that used in the preliminary round.

11.3 Final Submission Requirements

A reproducible Docker file, including but not limited to:

1. Complete training and validation code, including reproduction scripts.
2. Documentation detailing the code, environment setup, and usage instructions.

Technical proposal file (**PDF**): This should include the algorithm design motivation, method description, pseudocode, and other content to demonstrate the innovation and effectiveness of the proposed method.

Presentation slides (**PPT**) for the defense.

(For detailed organization and submission requirements, refer to the “Algorithm Competition Submission Requirements.”)

12. Other Notes

12.1 Fairness

Any form of cheating is strictly prohibited, including but not limited to data leakage, overlap between pretraining data and test data, or plagiarism of others’ code. Any violations will result in immediate disqualification.

12.2 Intellectual Property

1. Only the datasets provided by the competition are allowed. To ensure fairness, the use of any other public or private datasets is strictly prohibited. After the semifinal stage, the organizers will attempt to reproduce submitted code using the official datasets; if results cannot be reproduced, the team’s competition score will be invalidated.

2. All submissions must be original. The organizers have the right to display, promote, and publicize submitted works, but the intellectual property rights remain with the participants.

3. The intellectual property of the algorithms and executable models submitted by the teams is jointly held by the participants and the organizers. Competition videos

and image datasets are owned by Nanjing University of Science and Technology. With mutual agreement, winning algorithms may be used in research related to the funding of this competition.

4. Each team must ensure that their submitted results are reproducible. The organizers commit to maintaining confidentiality and will not use the submissions for any purpose outside of this competition.

5. Teams must guarantee that their proposed solutions and algorithms are their own intellectual property. The organizers are not responsible for any infringement or third-party claims arising from the use of submitted algorithms or results. Should any such issues occur, the participating team must assume full legal and financial responsibility and indemnify the organizers from any liability.

13. Contact Information

Competition communication QQ group: 731220935

Email: zerens@njust.edu.cn

Registration website: www.aicomp.cn

Appendix, Competition Procedure, and Award Settings

1. Registration Phase

Participants complete registration on the official competition website, submit individual or team information, and obtain the download link for the preliminary round dataset.

2. Preliminary Round

Participants use the training dataset provided by the organizers to design their algorithm models and utilize the preliminary round test set for method validation and debugging. During the preliminary round, participants can submit results an unlimited number of times per day, while the preliminary leaderboard is updated every hour.

3. Semi-Final (Provincial Round)

After the preliminary round ends, the competition moves into the semi-final stage, during which the semi-final dataset download link is made available. Only teams that submitted valid results in the preliminary round are eligible to enter the semi-finals. During this stage, participants use the semi-final dataset provided by the organizers to refine their algorithm models and submit inference results on the semi-final test data. The semi-final stage lasts for three days, with each team allowed to submit results no more than twice per day. The semi-final leaderboard is updated every hour.

4. Announcement of Semi-Final (Provincial Round) Results

The semi-final results will be announced on the official competition website. The number of teams entering the semi-finals will serve as the basis for awarding prizes. According to the provincial competition award ratio, first, second, and third prizes will be selected for the semi-finals, with corresponding certificates issued. During the selection process, any team whose algorithm performance falls below the baseline reference score provided by the organizers will be considered invalid and will not receive an award. Teams winning first and second prizes in the semi-finals will advance to the national finals.

5. Final (National Round) Stage

1. Online Evaluation of the Finals: Teams advancing to the finals will be evaluated based on the semi-final leaderboard. The number of teams entering the finals will serve as the basis for awarding prizes. According to the national competition award ratio, the list of candidates for the first prize and the winners of the second and third prizes will be determined, with certificates issued for the second and third prizes.

2.Submission of Final Entries: Candidate teams for the national first prize must submit technical documentation, algorithm code and model files, demonstration videos, and supplementary materials within the specified deadline. No modifications or additions will be accepted after the submission deadline.

3.Final Review Stage: A professional judging panel will reproduce and verify the results of the submissions from the national first prize candidate teams. If there are any questions during the review, the organizers may request explanations from the participants.

4.Offline Final Defense: Candidate teams for the national first prize must submit finalized technical documentation, algorithm code and model files, demonstration videos, and supplementary materials within the specified time, and participate in the offline national final defense. The final national first prize winners and their rankings will be determined based on algorithm performance scores combined with defense scores. Teams that do not participate in the offline defe