OOD-CV Challenge Report

September 18, 2023

1 Team details

- Challenge track: OOD-CV Workshop SSB Challenge (GCD Track - Self-Supervised)
- Team name: DAIU
- Team leader name: Mengjia Wang
- Team leader address, phone number, and email: 266 Xinglong Section, Xifeng Road, Xi'an City, Shaanxi Province, China. (+86)13102818603. 3230724499@qq.com
- Rest of the team members: Jingwen Zhang, Min Gao
- Team website URL: None
- Affiliation: School of Artificial Intelligence, Xidian University, Xi'an, China
- User names on the OOD-CV Codalab competitions: DAIU

- Link to the codes of the solution(s): `https://github.com/wmj183363206/gcd-self-supervised-1st`

2 Contribution details

- Title of the contribution : Effective Semi-Supervised Model for Generalized Category Discovery

- General method description: 1. The training data augmentation we used were HorizontialFlip, VerticalFlip, ColorJitter, RandomErasing [3] and CutMix [2]; 2. The models we used were "dinov2-vitb14" and "dinov2-vitl14" [1], and we trained these model with different image sizes and different composition of data augmentation tricks; 3. We used the Soft Voting Classifer and Hard Voting Classifier to do the results fusion.

- Description of the particularities of the solutions deployed for each of the tracks : 1. We tried many different data augmentation tricks, finally we found different types of dataset should use different composition of those tricks. For CUB and Stanford-Cars datasets, we added RandomErasing and CutMix, but for FGVC-Aircarft, we used the default pipeline; 2. There two models we trained with 224 and 308 input image size correspondingly; 3. For the results of same model, we used the Soft Voting Classifier, and the results of those would then use the Hard Voting Classifier to do the result fusion.

- References:

[1] Maxime Oquab, TimothÃ©e Darcet, Theo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Russell Howes, Po-Yao Huang, Hu Xu, Vasu Sharma, Shang-Wen Li, Wojciech Galuba, Mike Rabbat, Mido Assran, Nicolas Ballas, Gabriel Synnaeve, Ishan Misra, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2023.

[2] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *International Conference on Computer Vision (ICCV)*, 2019.

[3] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

- Representative image / diagram of the method(s):

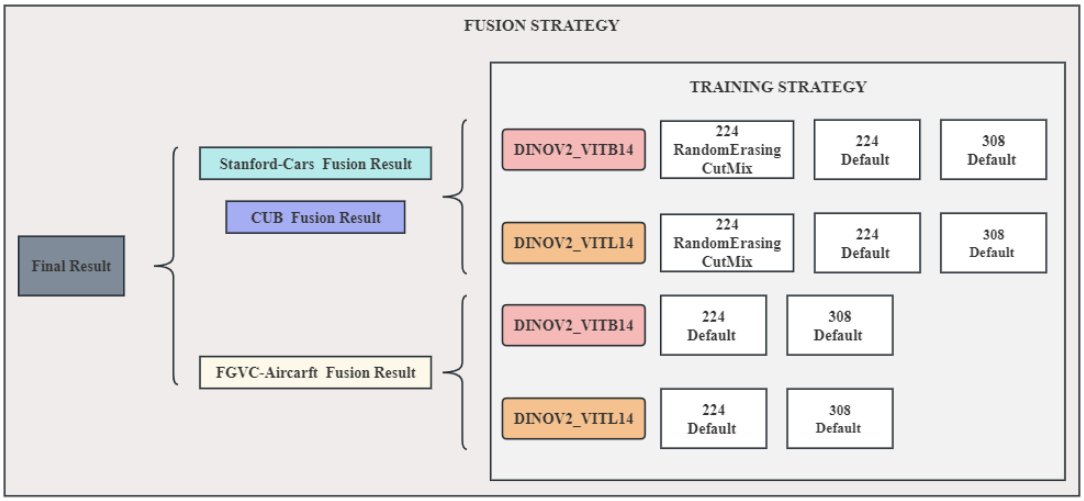

Figure 1: Training Strategy and Fusion Strategy

3 Global Method Description

[* Indicates method used in competition test results.]

- Total method complexity: We fine-tuned the last layer "block.11" of "dinov2-vitb14" and "block.23" of "dinov2-vitl14". For "dinov2-vitb14", we used 13.3 hours training on 3090 with 224 input image size; for "dinov2-vitl14", we used 13.8 hours training on 3090 with 224

input image size.

- Model Parameters: 1. "dinov2-vitb14" model with 86MB model parameter; 2. "dinov2-vitl14" model with 300MB model parameter.

- Run Time: 1. "dinov2-vitb14" model with 2min 57s total runtime for one epoch in 224 input image size; 2. "dinov2-vitb14" model with 8min 42s total runtime for one epoch in 308 input image size.

- Which pre-trained or external methods / models have been used: We only use "dinov2-vitb14" and "dinov2-vitl14" pre-trained model. One thing shouls be noticed, for "dinov2-vitb14", we fine-tuned the "block.11" layer, but for "dinov2-vitl14", we fine-tuned the "block.23" layer.

- Training description : 1. For CUB and Stanford-Cars datasets, we trained two models of "dinov2-vitb14" in 224 input image size with added RandomErasing and CutMix, and default data-augmentation correspondingly; and one model of "dinov2-vitb14" in 308 input image size with default data-augmentation; 2. For CUB and Stanford-Cars datasets, we trained two models of "dinov2-vitl14" in 224 input image size with added RandomErasing and CutMix, and default data-augmentation correspondingly; and one model of "dinov2-vitl14" in 308 input image size with default data-augmentation; 3. For FGVC-Aircarft dataset, we trained two models of "dinov2-vitb14" and "dinov2-vitl14" in 224 input image size; 4. For FGVC-Aircarft dataset, we trained two models of "dinov2-vitb14" and "dinov2-vitl14" in 308 input image size with default data-augmentation.

- Testing description: In testing stage, we only used the default setting.

- Quantitative and qualitative advantages of the proposed solution : In this challenge, our team did lots of experiments, the quantity and

quality could be surely satisfying.

- Results of the comparison to other approaches (if any) : None.

- Novelty of the solution and if it has been previously published: We trained the models with different strategies and finally fusion the results. This process is previously published.

4 Ensembles and fusion strategies

- Describe in detail the use of ensembles and/or fusion strategies (if any).: For the same model results of each dataset, we use Soft Voting Classifier, which sums the probabilities of various test results and ultimately selects the class label with the highest sum of probabilities. Then for the different model results of each dataset, we use the Hard Voting Classifier, which is taking the average probability of all model prediction samples in a certain category as the standard, and the corresponding type with the highest probability is the final prediction result.

- What was the benefit over the single method? : These follow the principle of minority obeying majority in both voting sessions reduces variance through the integration of multiple models, thereby improving the robustness and generalization ability of the model.

- What were the baseline and the fused methods? : The baseline is the "dinov2-vitb14" pre-trained model with 224 input image size and default data-augmentation.

5 Technical details

- Language and implementation details (including platform, memory, parallelization requirements) : We used Pytorch, single GPU training

and testing, totally used 2 GeForce RTX 3090.

- Human effort required for implementation, training and validation?: The mainly human effort for implementation was in the downloading data; for training was in the composition of different data-augmentation ticks; not much effort in the validation.

- Training/testing time? Runtime at test per image : The training time was shown above; the testing time of different datasets could be calculated by the runtime at test per image: 1. "dinov2-vitb14" pre-trained model with 224 input image size has 377 im/s; 2. "dinov2-vitl14" pre-trained model with 224 input image size has 114 im/s. This speed is not as precise because other code was running on the same gpu at the same time.

- Comment the efficiency of the proposed solution(s)? : The training is time-consuming and the performance of each model is largely influenced by the data-augmentation tricks, and the influence of input image size may not that large. The testing we use the default setting and not use Test Time Augmentation (TTA), considering the training performance.

6 Other details

- General comments and impressions of the OOD-CV challenge. : Our Team has greatly interested in SSB challenge and it has great room to develop. Thus, We are very grateful for OOD-CV official hosting such a competition.

- Other comments: None.