OOD-CV Challenge Report

September 18, 2023

1 Team details

- Challenge track:
  OOD-CV Challenge 2023 (Pose Estimation Track - Self-supervised leaderboard)
- Team name:
  ll_ly
- Team leader name:
  Wenxuan She
- Team leader address, phone number, and email:
  South Campus of Xidian University, Xi'an, Shaanxi Province, China, +8618706831828, swx_sxpp@qq.com
- Rest of the team members:
  Yu Liu, Zhongjian Huang, Fang Liu, Xu Liu
- Team website URL:
  None
- Affiliation:
  Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University
- User names on the OOD-CV Codalab competitions:
  ll_ly

- Link to the codes of the solution(s):
https://github.com/swx3027925806/OOD-Pose

2 Contribution details

- Title of the contribution :
OOD Pose Estimation

- General method description:
Pose estimation is generally to predict the position information of the object object in three-dimensional space. The mission's main prediction targets have three values: azimuth, elevation, and θ. The azimuth ranges from 0 to 2π, elevation from 0 to π, and θ from -π to π.

- Description of the particularities of the solutions deployed for each of the tracks :
The deep learning framework adopted in this scheme is PaddlePaddle2.4.0[1], and the relevant environment needs to be configured when the code is reproduced. The adopted code is completely built by ourselves. See github for details:
https://github.com/swx3027925806/OOD-Pose

- References:

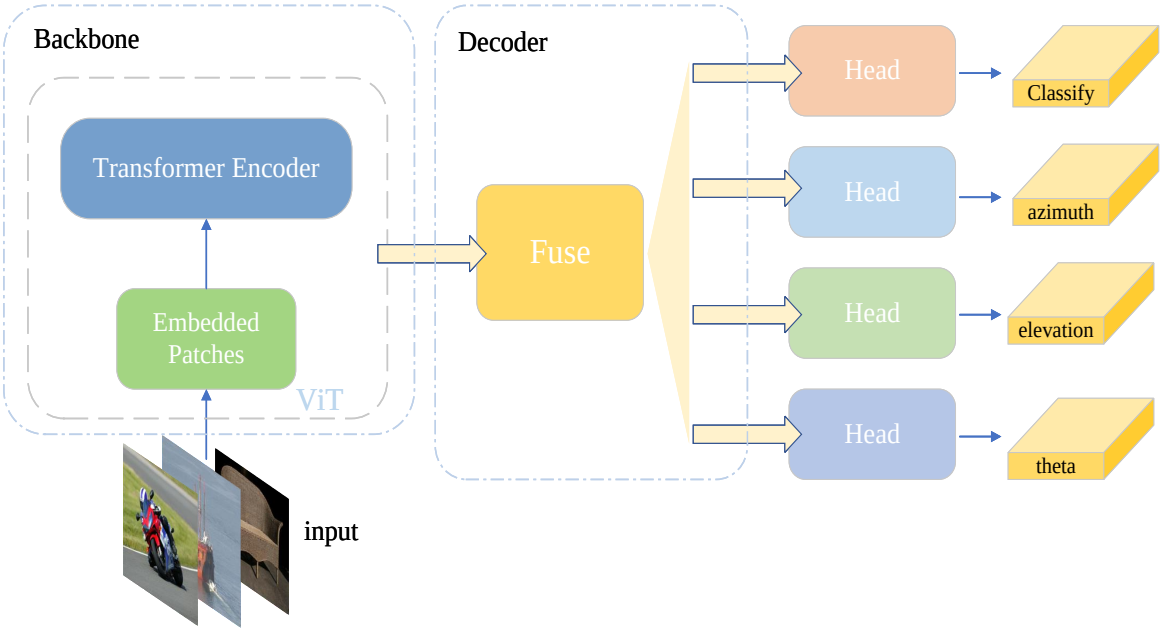
References

[1] Yanjun Ma, Dianhai Yu, Tian Wu, and Haifeng Wang, "Paddlepaddle: An open-source deep learning platform from industrial practice," *Frontiers of Data and Domputing*, vol. 1, no. 1, pp. 105–115, 2019.

[2] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al., "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[3] Zhiliang Peng, Li Dong, Hangbo Bao, Qixiang Ye, and Furu Wei, "Beit v2: Masked image modeling with vector-quantized visual tokenizers," *arXiv preprint arXiv:2208.06366*, 2022.

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



3 Global Method Description

[* Indicates method used in competition test results.]

- Total method complexity:
  The model is a single-stage task[2], and its main modification is that the decoder is set as a multi-task prediction head. The code structure is simple and effective.

- Model Parameters:
  During model training, batch_size is set to 32, image size is 224, and

the total training steps are 50k steps. The optimizer adopts adam, the learning rate attenuation strategy adopts WarmupCosineLR, and the learning rate is set to 1e-4.

- Run Time:
  In the supervised training stage, there are 50k steps in total, which takes 23 hours.
- Which pre-trained or external methods / models have been used:
  This scheme adopts the self-monitoring model BEiTv2[3] which is provided by paddle and trained on ImageNet.
- Training description :
  Because the data set was determined late, the task only completed two versions of training, and its performance reached 90+ in the verification stage. For the data, it is necessary to adjust the out-of-range parameters to ensure that all values are within the specified range.
- Testing description:
  In the prediction, the short side of the image is scaled to 224, and the central clipping strategy is adopted, and the final performance reaches 90+.
- Quantitative and qualitative advantages of the proposed solution :
  The scheme has achieved good results in prediction speed and performance.
- Novelty of the solution and if it has been previously published:
  This scheme transforms the regression task into a classification task, and it is more delicate about the angle and has achieved good results.

4 Ensembles and fusion strategies

- Describe in detail the use of ensembles and/or fusion strategies (if any).:
  The experiment did not adopt the fusion strategy.
- What was the benefit over the single method? :
  The experiment did not adopt the fusion strategy.
- What were the baseline and the fused methods? :
  The experiment did not adopt the fusion strategy.

5 Technical details

- Language and implementation details (including platform, memory, parallelization requirements) :
  This model uses python as the programming language and PaddlePaddle 2.4.0 as the development framework. It is best to run on the GPU of V100.

- Human effort required for implementation, training and validation?:
  The training of this experiment is carried out on the single V100 GPU, and the verification is carried out on the single V100 GPU.

- Training/testing time? Runtime at test per image :
  In the supervised training stage, the total amount is 50k steps, which takes 23 hours. In the test, the reasoning time of a single picture is 0.01 seconds.

- Comment the efficiency of the proposed solution(s)? :
  The scheme is a one-stage model in both the training stage and the forecasting stage, so it should be faster in efficiency.

6 Other details

- General comments and impressions of the OOD-CV challenge. :
  Out-of-Distribution(OOD) detection plays an important role in the stability and security of machine learning. The OOD task is to detect whether a machine learning model can correctly identify and classify all categories in the data set, especially those that do not belong to the training set. This kind of detection is very important to ensure the reliability and robustness of machine learning model in practical application. Therefore, we think that this challenge event is very necessary.

- Other comments:
  I hope the competition can get better and better. Great breakthroughs can also be made in this direction.