

This repo is the official implementation of the CVPR 2023 paper: <u>Distilling Self-Supervised Vision Transformers for Weakly-Supervised Few-Shot Classification & Segmentation</u>.

Environmnet installation

This project is built upon the following environment:

- <u>Ubuntu 18.04</u>
- Python 3.10
- CUDA 11.0
- PyTorch 1.12.0

The package requirements can be installed via environment.yml, which includes

- pytorch ==1.12.0
- torchvision ==0.13.0
- cudatoolkit ==11.3
- pytorch-lightning == 1.6.5
- <u>einops</u> ==0.6.0

```
conda env create --name pytorch1.12 --file environment.yml -p YOURCONDADIR/envs/pytorch1.12 conda activate pytorch1.12
```

Make sure to replace YOURCONDADIR in the installation path with your conda dir, e.g., ~/anaconda3

Datasets

- PASCAL VOC 2012 and SBD
- Microsoft COCO 2014

Download the datasets by following the file structure below and set <code>args.datapath=YOUR_DATASET_DIR</code>:

```
YOUR_DATASET_DIR/

├─ VOC2012/

│    ├─ Annotations/
│    ├─ JPEGImages/
│    ├─ ...

├─ COC02014/
│    ├─ annotations/
│    ├─ train2014/
│    ├─ val2014/
│    ├─ val2014/
│    ├─ ...
```

Training with pixel-level supervision

Training with image-level supervision

CST model checkpoints

Experimental results on Pascal-5ⁱ datasets on the FS-CS task.

| Performance results | | | | | Links to download checkpoints | | | |
|-------------------------------|-------------|-----------|--------------|-----------|-------------------------------|-------|-------|-------|
| methods | 1-way | 1-shot | 2-way 1-shot | | 4-fold validation folds | | | |
| metric | cls. 0/1 ER | seg. mloU | cls. 0/1 ER | seg. mloU | fold0 | fold1 | fold2 | fold3 |
| image-level supervised models | 79.9 | 33.2 | 64.6 | 31.9 | <u>link</u> | link | link | link |
| pixel-level supervised models | 85.7 | 55.5 | 70.4 | 53.7 | <u>link</u> | link | link | link |

Experimental results on COCO-20ⁱ datasets on the FS-CS task.

| Performance results | | | | | | Links to download checkpoints | | | | |
|-------------------------------|--------------|-----------|--------------|-----------|-------------------------|-------------------------------|-------|-------|--|--|
| methods | 1-way 1-shot | | 2-way 1-shot | | 4-fold validation folds | | | | | |
| metric | cls. 0/1 ER | seg. mloU | cls. 0/1 ER | seg. mloU | fold0 | fold1 | fold2 | fold3 | | |
| image-level supervised models | 78.2 | 19.6 | 62.4 | 18.3 | link | link | link | link | | |
| pixel-level supervised models | 80.8 | 38.3 | 64.0 | 36.2 | link | link | link | link | | |

BibTex source

If you find our code or paper useful, please consider citing our paper:

@inproceedings{kang2023distilling,

title={Distilling Self-Supervised Vision Transformers for Weakly-Supervised Few-Shot Classification \& Segmentation}, author={Kang, Dahyun and Koniusz, Piotr and Cho, Minsu and Murray, Naila}, booktitle={Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition}, year={2023}

Releases

No releases published

Packages

No packages published

Languages

• Python 100.0%

ي