Improving CLIP Training

 # Benchmarks

The following results are recall at 1 results on the provided MSCOCO and ImageNet datasets. The first row of results are from the model trained using the CLIP loss, and the second row of results are from the model trained using the SogCLR loss. All results are based on a batch size of 128 for 30-epoch pretraining. IR@1 denotes the recall at 1 of image retrieval on MSCOCO, TR@1 denotes the recall at 1 of text retrieval on MSCOCO, and ACC@1 denotes the top 1 accuracy on ImageNet. Average denotes the average of the three metrics.

Method	MSCOCO TR@1	MSCOCO IR@1	ImageNet ACC@1	Average
CLIP	12.0	9.32	21.35	14.22
SogCLR	14.38	10.73	24.54	16.55

Here are the results from our training and validation with 3 different optimizer and 5 loss functions.

Optimizer: adamW (default), Loss Function: sogCLR

Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sogclr_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:01:20
coco val: {'txt_r1': 13.18, 'txt_r5': 33.32, 'txt_r10': 44.58, 'txt_r_mean': 30.36, 'img_r1': 10.296293334399616, 'img_r5': 26.946299332240393, 'img_r10': 37.806389699708106, 'img_r_mean': 25.01632745544937, 'r_mean': 27.688163727724685}
zeroshot: {'zeroshot_top1': 24.548, 'zeroshot_top3': 37.598, 'zeroshot_top5': 43.272, 'zeroshot_top10': 50.412}
Training time 0:08:56

Optimizer: adamW (default), Loss Function: CLIP

Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/clip_cc3m_g0.8_e30/
checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:15
coco val: {'txt_r1': 11.62, 'txt_r5': 30.78, 'txt_r10': 43.36, 'txt_r_mean':
28.586666666666662, 'img_r1': 9.160702147227, 'img_r5':
25.31488664080931, 'img_r10': 35.995041784957415,
'img_r_mean': 23.490210190997903, 'r_mean':
26.038438428832283}
zeroshot: {'zeroshot_top1': 21.658, 'zeroshot_top3': 34.418,
'zeroshot_top5': 40.576, 'zeroshot_top10': 48.798}
Training time 0:04:31

Optimizer: adamW (default), Loss Function: cyclip

Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/cyclip_cc3m_g0.8_e30/
checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:01:30
coco val: {'txt_r1': 14.1, 'txt_r5': 33.84, 'txt_r10': 46.3, 'txt_r_mean':
31.413333333333338, 'img_r1': 10.68415370466632, 'img_r5':
27.694030149146307, 'img_r10': 38.17825582790196,
'img_r_mean': 25.518813227238194, 'r_mean':
28.466073280285766}
zeroshot: {'zeroshot_top1': 25.906, 'zeroshot_top3': 39.492,
'zeroshot_top5': 45.658, 'zeroshot_top10': 53.904}
Training time 0:10:15Start training
Computing features for evaluation...
Evaluation time 0:00:15

coco val: {'txt_r1': 11.62, 'txt_r5': 30.78, 'txt_r10': 43.36, 'txt_r_mean': 28.586666666666662, 'img_r1': 9.160702147227, 'img_r5': 25.31488664080931, 'img_r10': 35.995041784957415, 'img_r_mean': 23.490210190997903, 'r_mean': 26.038438428832283}
zeroshot: {'zeroshot_top1': 21.658, 'zeroshot_top3': 34.418, 'zeroshot_top5': 40.576, 'zeroshot_top10': 48.798}
Training time 0:04:31

Optimizer: adamW (default), Loss Function: vicreg

Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/vicreg_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:01:05
coco val: {'txt_r1': 2.86, 'txt_r5': 9.14, 'txt_r10': 14.5, 'txt_r_mean': 8.833333333333334, 'img_r1': 2.155224119317046, 'img_r5': 7.185413251229558, 'img_r10': 11.755767923547523, 'img_r_mean': 7.032135098031375, 'r_mean': 7.932734215682355}
zeroshot: {'zeroshot_top1': 5.788, 'zeroshot_top3': 12.746, 'zeroshot_top5': 17.972, 'zeroshot_top10': 26.93}
Training time 0:09:49

Optimizer: adamW (default), Loss Function: onlineclr

Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/onlineclr_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:01:03
coco val: {'txt_r1': 10.96, 'txt_r5': 29.5, 'txt_r10': 40.52, 'txt_r_mean': 26.99333333333336, 'img_r1': 8.644887840377464, 'img_r5':

23.53952577072254, 'img_r10': 34.37162621456276, 'img_r_mean': 22.185346608554255, 'r_mean': 24.589339970943797}
zeroshot: {'zeroshot_top1': 20.522, 'zeroshot_top3': 32.686, 'zeroshot_top5': 38.286, 'zeroshot_top10': 46.144}
Training time 0:09:41

Optimizer: SGD, Loss Function: SogCLR
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sgd_sogclr_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:33
coco val: {'txt_r1': 1.56, 'txt_r5': 5.78, 'txt_r10': 10.34, 'txt_r_mean': 5.8933333333333335, 'img_r1': 1.0036386900715741, 'img_r5': 4.286456875524811, 'img_r10': 7.825182934143708, 'img_r_mean': 4.371759499913364, 'r_mean': 5.132546416623349}
zeroshot: {'zeroshot_top1': 2.872, 'zeroshot_top3': 6.72, 'zeroshot_top5': 9.468, 'zeroshot_top10': 14.442}
Training time 0:08:48

Optimizer: SGD, Loss Function: CLIP
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sgd_clip_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:57
coco val: {'txt_r1': 10.3, 'txt_r5': 27.64, 'txt_r10': 38.16, 'txt_r_mean': 25.366666666666664, 'img_r1': 7.001479467391739, 'img_r5': 20.468631292734617, 'img_r10': 30.253108880802912, 'img_r_mean': 19.24107321364309, 'r_mean': 22.303869940154875}
zeroshot: {'zeroshot_top1': 17.006, 'zeroshot_top3': 29.352, 'zeroshot_top5': 35.214, 'zeroshot_top10': 43.664}

Training time 0:11:20

Optimizer: SGD, Loss Function: cyclip
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sgd_cyclip_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:31
coco val: {'txt_r1': 10.38, 'txt_r5': 27.54, 'txt_r10': 38.94, 'txt_r_mean': 25.62, 'img_r1': 7.313367187812387, 'img_r5': 20.920468631292735, 'img_r10': 30.64496781158783, 'img_r_mean': 19.626267876897654, 'r_mean': 22.623133938448827}
zeroshot: {'zeroshot_top1': 16.858, 'zeroshot_top3': 29.238, 'zeroshot_top5': 35.094, 'zeroshot_top10': 43.78}
Training time 0:08:48

Optimizer: SGD, Loss Function: vicreg
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sgd_vicreg_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:24
coco val: {'txt_r1': 2.0, 'txt_r5': 7.42, 'txt_r10': 12.36, 'txt_r_mean': 7.260000000000001, 'img_r1': 1.595425646767164, 'img_r5': 5.71394298052701, 'img_r10': 9.844455995841496, 'img_r_mean': 5.7179415410452235, 'r_mean': 6.488970770522612}
zeroshot: {'zeroshot_top1': 2.432, 'zeroshot_top3': 6.106, 'zeroshot_top5': 9.104, 'zeroshot_top10': 15.374}
Training time 0:08:28

Optimizer: SGD, Loss Function: onlineclr
Creating retrieval dataset

len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/sgd_onlineclr_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:20
coco val: {'txt_r1': 0.74, 'txt_r5': 3.08, 'txt_r10': 5.4, 'txt_r_mean': 3.0733333333333337, 'img_r1': 0.5557999120316686, 'img_r5': 2.463113279219481, 'img_r10': 4.454396417289776, 'img_r_mean': 2.491103202846975, 'r_mean': 2.7822182680901544}
zeroshot: {'zeroshot_top1': 1.5, 'zeroshot_top3': 3.772, 'zeroshot_top5': 5.642, 'zeroshot_top10': 9.286}
Training time 0:08:27

Optimizer: Adam, Loss Function: SogCLR
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/adam_sogclr_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:37
coco val: {'txt_r1': 0.1, 'txt_r5': 0.5, 'txt_r10': 0.94, 'txt_r_mean': 0.5133333333333333, 'img_r1': 0.10396257347354952, 'img_r5': 0.5038186252948939, 'img_r10': 0.9356631612619457, 'img_r_mean': 0.5144814533434631, 'r_mean': 0.513907393338398 2}
zeroshot: {'zeroshot_top1': 0.23, 'zeroshot_top3': 0.664, 'zeroshot_top5': 1.144, 'zeroshot_top10': 2.216}
Training time 0:09:18

Optimizer: Adam, Loss Function: CLIP
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/adam_clip_cc3m_g0.8_e30/checkpoint_30.pth

Start training
Computing features for evaluation...
Evaluation time 0:00:17
coco val: {'txt_r1': 3.34, 'txt_r5': 11.36, 'txt_r10': 18.06, 'txt_r_mean': 10.92, 'img_r1': 3.0349074333240034, 'img_r5': 9.840457435323284, 'img_r10': 15.64236874725099, 'img_r_mean': 9.505911205299425, 'r_mean': 10.212955602649712}
zeroshot: {'zeroshot_top1': 4.712, 'zeroshot_top3': 9.738, 'zeroshot_top5': 13.04, 'zeroshot_top10': 18.416}
Training time 0:02:49

Optimizer: Adam, Loss Function: cyclip
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/adam_cyclip_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:16
coco val: {'txt_r1': 3.92, 'txt_r5': 12.6, 'txt_r10': 19.18, 'txt_r_mean': 11.9, 'img_r1': 3.2548282618257427, 'img_r5': 10.41625014994602, 'img_r10': 16.36210964052941, 'img_r_mean': 10.011062684100391, 'r_mean': 10.955531342050197}
zeroshot: {'zeroshot_top1': 4.044, 'zeroshot_top3': 9.064, 'zeroshot_top5': 12.536, 'zeroshot_top10': 18.496}
Training time 0:08:48

Optimizer: Adam, Loss Function: vicreg
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/adam_vicreg_cc3m_g0.8_e30/checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:16
coco val: {'txt_r1': 0.66, 'txt_r5': 2.54, 'txt_r10': 4.5, 'txt_r_mean': 2.566666666666667, 'img_r1': 0.631772561877724, 'img_r5': 2.491103202846975, 'img_r10': 4.402415130553001, 'img_r_mean':

2.5084302984259, 'r_mean': 2.5375484825462835}
zeroshot: {'zeroshot_top1': 1.284, 'zeroshot_top3': 3.43,
'zeroshot_top5': 5.466, 'zeroshot_top10': 9.416}
Training time 0:03:31

Optimizer: Adam, Loss Function: onlineclr
Creating retrieval dataset
len of train_dataset: 100000
len of coco val: 5000
Creating model
load checkpoint from ./output/adam_onlineclr_cc3m_g0.8_e30/
checkpoint_30.pth
Start training
Computing features for evaluation...
Evaluation time 0:00:15
coco val: {'txt_r1': 0.02, 'txt_r5': 0.04, 'txt_r10': 0.06, 'txt_r_mean':
0.04, 'img_r1': 0.019992802591067216, 'img_r5':
0.09996401295533608, 'img_r10': 0.19992802591067216,
'img_r_mean': 0.10662828048569183, 'r_mean':
0.07331414024284591}
zeroshot: {'zeroshot_top1': 0.1, 'zeroshot_top3': 0.3, 'zeroshot_top5':
0.5, 'zeroshot_top10': 1.0}
Training time 0:02:50