Table 11: Training details for MoCov2, MoCov3, SimCLR, SupCon and DINO. 'bs' denotes batch size, 'lr' denotes learning rate, 'wd' denotes weight decay, 'dim' denotes the dimension of feature, τ denotes the temperature parameter and k denotes the size of memory bank in MoCov2.

Method	Settings							
	opt	bs	lr	wd	dim	schedule	τ	k
MoCov2	SGD	256	5e-4	0.001	256	cosine	0.2	2048
MoCov3	adamW	256	1.5e-4	0.1	256	cosine	0.2	-
SimCLR	SGD	512	5e-4	0.001	256	cosine	0.1	-
Supcon	SGD	512	5e-4	0.001	256	cosine	0.1	-
DINO	adamW	512	5e-4	0.04	256	cosine	0.1	-

A   Training details

The training details for MoCov2, MoCov3, SimCLR, SupCon and DINO in Table 2 in the paper are shown in Table 11.