Method	Archit.	Data	Accuracy
MAE	ViT-H/14	INet-1k	76.6
DINO	ViT-S/8	INet-1k	79.2
SEERv2	RG10B	IG2B	79.8
MSN	ViT-L/7	INet-1k	80.7
EsViT	Swin-B/W=14	INet-1k	81.3
Mugs	ViT-L/16	INet-1k	82.1
iBOT	ViT-L/16	INet-22k	82.3
DINOv2	ViT-S/14	LVD-142M	81.1
	ViT-B/14	LVD-142M	84.5
	ViT-L/14	LVD-142M	86.3
	ViT-g/14	LVD-142M	**86.5**

Table 4: **Linear evaluation on ImageNet-1k of frozen pretrained features.** We report Top-1 accuracy on the validation set for publicly available models trained on public or private data, and with or without text supervision (text sup.). For reference, we also report the kNN performance on the validation set. We compare across any possible architectures (Arch.), at resolution 224×224 unless stated otherwise. The dataset used for training EVA-CLIP is a custom mixture, see paper for details (Fang et al., 2023).