

on ImageNet. We then explore the generalization ability of the model to other datasets, just retraining the softmax classifier on top. As such, this is a form of supervised pre-training, which contrasts with the unsupervised pre-training methods popularized by (Hinton et al., 2006) and others (Bengio et al., 2007; Vincent et al., 2008). The generalization ability of convnet fea-