On the Instability of Unsupervised Domain Adaptation with ADDA

Kazuki Omi and Toru Tamaki (Nagoya Institute of Technology, Japan)

Motivation

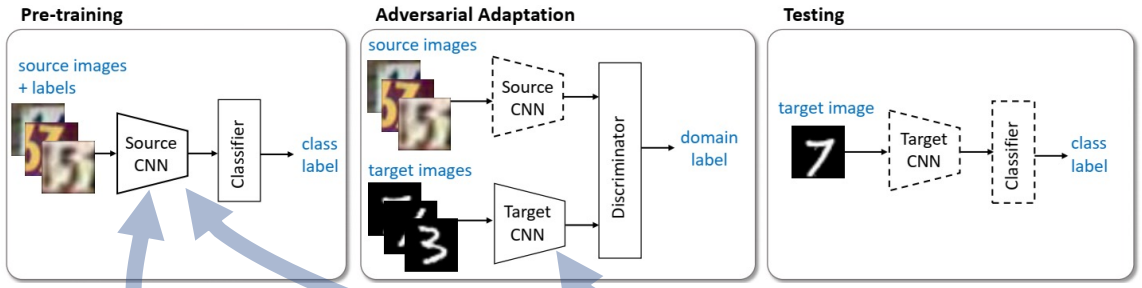
- Domain adaptation (DA) uses information of source domain S to solve the task of target domain T
  - It is called Unsupervised DA when no labels of T are available
- Adversarial Discriminative Domain Adaptation or ADDA (Tzeng+, CVPR2017) incorporates adversarial losses, but with instability of training
- This work report experimentally the causes of the instability

Overview of ADDA

Pre-training source CNN and classifier on S

Training target CNN initialized to source CNN and the discriminator with adversarial loss

Classification of target CNN and classifier on T



Three experiments

Exp. 1 Is pre-training stable?

Exp. 2 Does the number of epochs for pre-training affect?

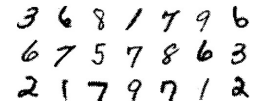
Exp. 3 Does the initialization of target CNN affect?

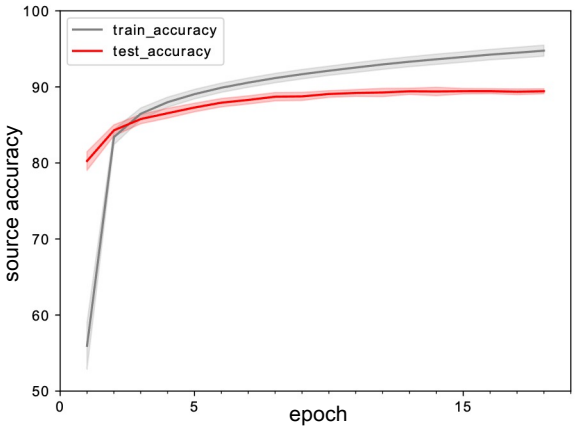
Results

- Dataset
  - Source domain: SVHN
  - Target domain: MNIST
- Exp. 1
  - Pre-training for 20 times
    - computing average and std of source accuracy
  - Small std
    - pre-training is stable
- Exp. 2
  - Different epochs of pre-training
  - Differences of average target accuracies between 12 and 18 is not statistically significant (t-test, α=0.05)
    - Source accuracy doesn't directly affect target accuracy
- Exp. 3
  - 10 frozen parameters of source CNN. For each, initialize target CNN to train 10 times
  - Target accuracy depends on initialization.
    - Initialization does matter

SVHN (Netzer+, NIPS2011)

MNIST (LuCun+, IEEE1998)





epochs	target accuracy
12	72.52 ± 5.77
18	67.95 ± 8.23