Deep CORAL: Correlation Alignment for Deep Domain Adaptation (ECCV 2016)

Baochen Sun, and Kate Saenko
University of Massachusetts Lowell, Boston University

Edemir Ferreira de Andrade Junior
Department of Computer Science
Federal University of Minas Gerais

September 22, 2017

DCC
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

UFMG

PATREO

Introduction

- Pattern recognition requires labeled data for learning purposes.
- Not feasible collect and label data for each problem of interest.
- In unsupervised domain adaptations, we would like to **transfer knowledge learned** from:
- a **source domain** (which we have labeled data)
- to a **target domain** (which we have no ground truth labels)

Previous work

- Previous techniques for unsupervised adaptation consisted of re-weighting the training point losses to more closely reflect those in the test distribution

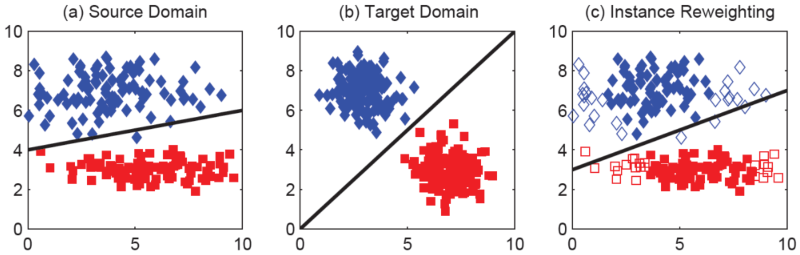


Figure: M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

Previous work

- Bridge the source and target domains by projecting them onto points along a geodesic path, or finding a closed-form linear map that transforms source points to target

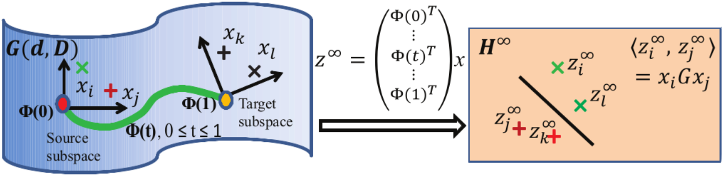


Figure: Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: CVPR (2012)

Previous work

- DLID, trains a joint source and target CNN architecture with two adaptation layers

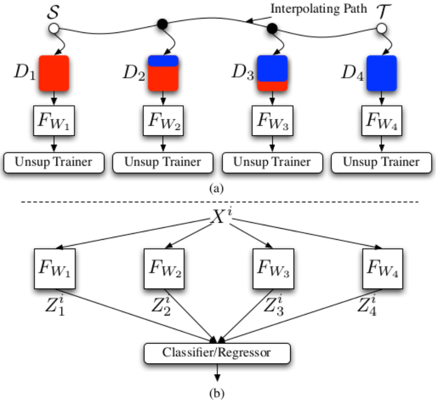


Figure: Chopra, S. et. al: DLID: Deep learning for domain adaptation by interpolating between domains. In: ICML Workshop (2013)

Previous work

- DDC applies a single linear kernel to one layer to minimize Maximum Mean Discrepancy (MMD)

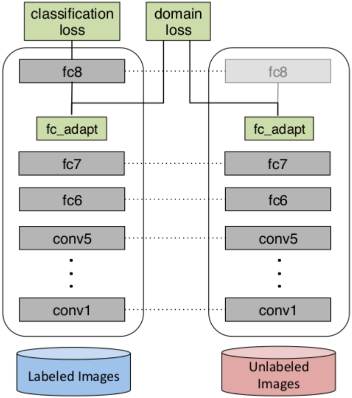


Figure: Tzeng, E. et al.: Deep domain confusion: Maximizing for domain invariance. CoRR abs/1412.3474 (2014), http://arxiv.org/abs/1412.3474

Previous work

- DAN minimizes MMD with multiple kernels applied to multiple layers

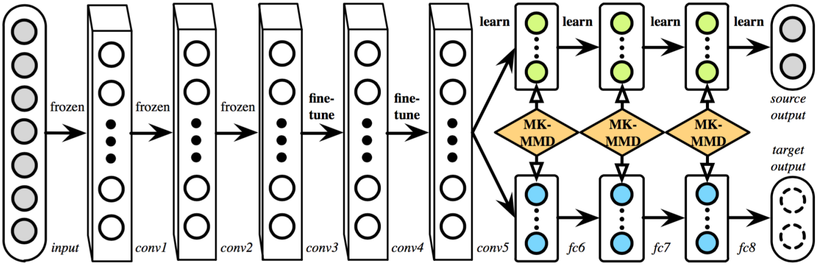


Figure: Long, M., Cao, Y., Wang, J., Jordan, M.I.: Learning transferable features with deep adaptation networks. In: ICML (2015)

Proposed approach

- Proposed Deep CORAL approach is similar to DDC, DAN, and ReverseGrad in the sense that a new loss (CORAL loss) is added to minimize the difference in learned feature covariances across domains
- However, it is more powerful than DDC (which aligns sample means only)
- Simpler to optimize than DAN and ReverseGrad
- Integrated into different layers or architectures seamlessly.

CORAL LOSS

- CORAL loss can be defined as:

$$L_{CORAL} = \frac{1}{4d^2}||C_S - C_T||_F^2$$

- Where: $||A||_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2$
- $C_S = \frac{1}{n_S-1}(D_S^T D_S - \frac{1}{n_S}(\mathbf{1}^T D_S)^T(\mathbf{1}^T D_S))$
- $C_T = \frac{1}{n_T-1}(D_T^T D_T - \frac{1}{n_T}(\mathbf{1}^T D_T)^T(\mathbf{1}^T D_T))$
- Distance between the second-order statistics (covariances) of the source and target features.
- $\frac{\partial L_{CORAL}}{\partial D_T^{ij}} = \frac{1}{d^2(n_S-1)}(D_S^T - \frac{1}{n_S}((\mathbf{1}^T D_S)^T \mathbf{1}^T)^T(C_S - C_T))^{ij}$
- $\frac{\partial L_{CORAL}}{\partial D_T^{ij}} = -\frac{1}{d^2(n_T-1)}(D_T^T - \frac{1}{n_T}((\mathbf{1}^T D_T)^T \mathbf{1}^T)^T(C_S - C_T))^{ij}$

CORAL LOSS

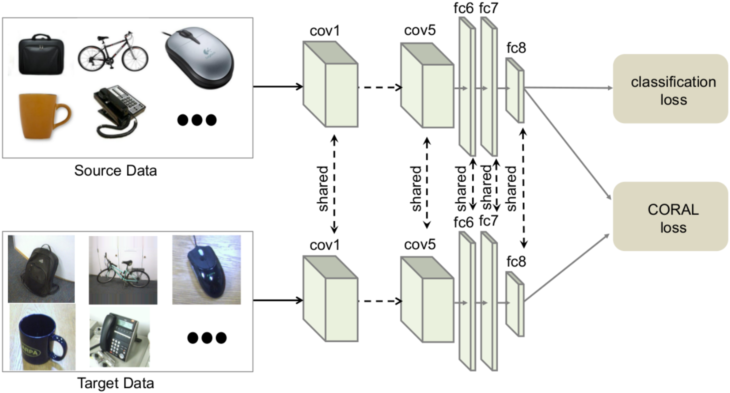
- Minimizing the classification loss itself is likely to lead to overfitting to the source domain, causing reduced performance on the target domain.
- On the other hand, minimizing the CORAL loss alone might lead to degenerated features.
- Joint training with both the classification loss and CORAL loss is likely to learn features that work well on the target domain:

$$L = L_{CLASS} + \sum_{i=1}^{t} \lambda L_{CORAL}$$

- Where: **t** denotes the number of CORAL loss layers
- λ is a weight that trades off the adaptation with classification accuracy on the source domain

Overview of the model architecture

Deep CORAL: Correlation Alignment for Deep Domain Adaptation



Evaluation: Settings

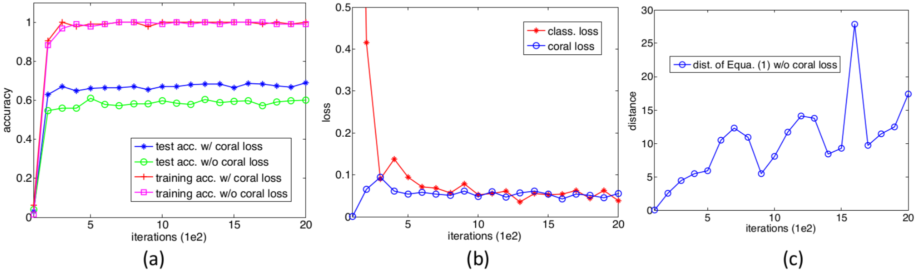
- The method was evaluated in the Office dataset:
  - The Office dataset contains 31 object categories from an office environment in 3 image domains: Amazon, DSLR, and Webcam.
- The authors apply the CORAL loss to the last classification layer
- Applying the CORAL loss to other layers or other network architectures is also possible.
- Compare to 7 "recently" published methods: CNN (no adaptation), GFK, SA, TCA, CORAL, DDC, DAN.

Evaluation: Results

	A→D	A→W	D→A	D→W	W→A	W→D	AVG
GFK	52.4±0.0	54.7±0.0	43.2±0.0	92.1±0.0	41.8±0.0	96.2±0.0	63.4
SA	50.6±0.0	47.4±0.0	39.5±0.0	89.1±0.0	37.6±0.0	93.8±0.0	59.7
TCA	46.8±0.0	45.5±0.0	36.4±0.0	81.1±0.0	39.5±0.0	92.2±0.0	56.9
CORAL	65.7±0.0	64.3±0.0	48.5±0.0	**96.1**±0.0	48.2±0.0	**99.8**±0.0	70.4
CNN	63.8±0.5	61.6±0.5	51.1±0.6	95.4±0.3	49.8±0.4	99.0±0.2	70.1
DDC	64.4±0.3	61.8±0.4	52.1±0.8	95.0±0.5	**52.2**±0.4	98.5±0.4	70.6
DAN	65.8±0.4	63.8±0.4	**52.8**±0.4	94.6±0.5	51.9±0.5	98.8±0.6	71.3
D-CORAL	**66.8**±0.6	**66.4**±0.4	**52.8**±0.2	95.7±0.3	51.5±0.3	99.2±0.1	**72.1**

Table 1. Object recognition accuracies for all 6 domain shifts on the standard Office dataset with deep features, following the standard unsupervised adaptation protocol.

Evaluation: Results



Conclusions

- They are able to do so by using a GAN?based technique, stabilized by both a task-specific loss and a novel content?similarity loss.
- decouples the process of domain adaptation from the task-specific architecture
- provides the added benefit of being easy to understand via the visualization of the adapted image outputs of the model.