PAPER ANALYSIS

Presented by Yannis He

-

Paper: **Deep Visual Domain Adaptation: A Survey**

Authors: Mei Wang, Weihong Deng | Beijing University of Posts and Telecommunications

https://arxiv.org/abs/1802.03601

ABSTRACT

Main Ideas:

- Domain Adaptation (DA) is a subtopic of Transfer Learning (TL)
- Introducing deep domain adaptation methods for computer vision applications
  - Taxonomy of Different deep domain adaptation scenarios
  - Deep domain adaptation approaches
  - Computer vision applications
  - Potential deficiencies of current methods and future directions

INTRO: MOTIVE AND BACKGROUND

- Motive: Getting data is too expensive
- Machine learning → shallow DA to solve domain shift between the source and target domain
  - Instance-based DA
    - Reduces discrepancy by reweight the source samples and train on the weighted source samples
  - Feature-based DA
    - Learning a common shared space, which the distributions of the two datasets are matched
  - Shallow DA SOTA:
    - [G. Csurka. Domain adaptation for visual applications: A comprehensive survey]
- Neural-network based Deep Learning → Deep DA
  - Existing architectures:
    - Convolutional Neural Networks (CNNs)
    - Deep Belief Networks (DBNs)
    - Stacked Autoencoders (SAEs)
  - Pros:
    - Learn more transferable representations
  - Cons:
    - Domain shift still affects their performance
    - Deep features transfer from general to specific
    - Transferability of the representations sharply decreases in higher layers

INTRO: MOTIVE AND BACKGROUND CONT'

- 3 Deep DA subsetting based on training loss:
  - Classification loss
  - Discrepancy loss
  - Adversarial loss
- Transfer Learning:
  - Three subsettings (only for homogeneous features spaces:
    - Introductive TL
    - Transductive TL
    - Unsupervised TL
  - Two TL categories:
    - Feature-representation-level knowledge transfer
    - Classifier-level knowledge transfer
- Coming UP:
  - Taxonomy of different Deep DA based on properties of data that define how two domains are diverged (Section II)
  - Introduce 3 subsetting for deep DA and summarize the uses in different DA scenes (Section III)
  - Introduce One-step DA methods (Section IV): Assume source and target domains are directly related
  - Introduce Multi-step DA methods (Section V)
    - Handcrafted
    - Feature-based
    - Representation-based
  - Survey of computer vision applications (Section VI)

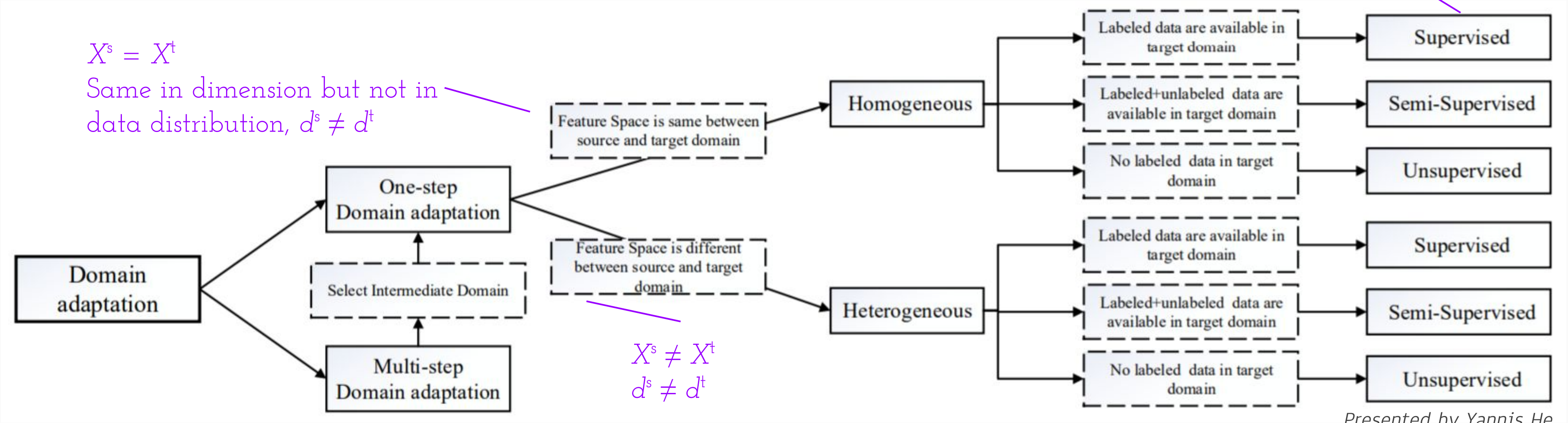
- Notations:
  - Feature space, $\mathcal{X}$
  - Marginal probability distribution P(X), $X = (x_1, x_2, \ldots, x_n)$
  - Domain, $\mathcal{D} = \{\mathcal{X}, P(X)\}$
  - Task, $\mathcal{T}$, consists of a feature space, $\mathcal{Y}$
  - Objective predictive function f(), which can be viewed as $P(Y|X)$
  - Labeled data: $\{x_i, y_i\}$, $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$,
    - There are source labeled data
    - There are target labeled data
    - There are unlabeled data
  - Source domain: $\mathcal{D}^s = \mathcal{X}^s + P(X)^s$
  - Target domain: $\mathcal{D}^t = \mathcal{X}^t + P(X)^t$
  - $\mathcal{D}^t = \mathcal{D}^{tl} \cup \mathcal{D}^{tu}$, where $\mathcal{D}^{tl}$ is partially labeled part, $\mathcal{D}^{tu}$ is unlabeled part
  - Attribute representation for class c, $a^c = (a^c_1, a^c_2, \ldots, a^c_m)$
    - A fixed-length binary values with m attributes in all the classes

SECTION II: OVERVIEW CONT'

- Traditional Machine Learning: $D^s = D^t$ and $T^s = T^t$
- Different Setting of DA:
  - Domain Divergence: $D^s \neq D^t$ (distribution shift or feature space difference
  - Task divergence: $T^s \neq T^t$ (conditional distribution shift or label space difference)
  - Or both
- 3 main groups of TL:
  - Inductive: $D^s \neq D^t$ and $T^s = T^t$
  - Transductive
  - Unsupervised TL

Small amount of labeled data available but not sufficient for tasks

$X^s = X^t$
Same in dimension but not in data distribution, $d^s \neq d^t$



$X^s \neq X^t$
$d^s \neq d^t$

SECTION II: OVERVIEW CONT'

- If source and target domains are directly related
  - One-step DA
- Else: Multi-step DA
  - Use a series of intermediate domains to connect two seemingly unrelated domains

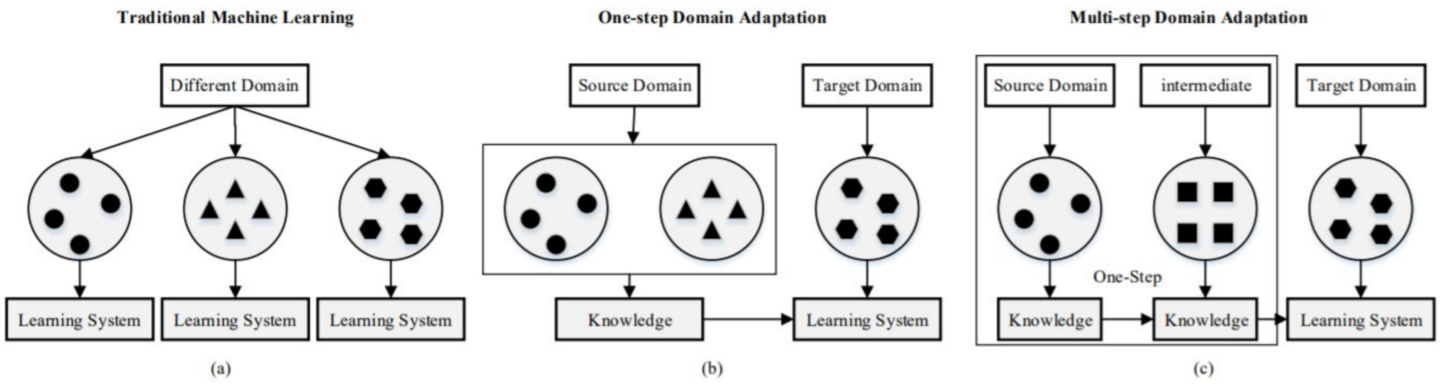


Fig. 3. Different learning processes between (a) traditional machine learning, (b) one-step domain adaptation and (c) multi-step domain adaptation [83].

SECTION III: APPROACHES OF DEEP DA - BACKGROUND & ONE-STEP

- Background
  - Deep DA: Use Shallow DA method on deep features
    - Deep networks only extract vectorial features and not helpful for transferring knowledge directly
  - Intuitive idea: embed DA into the process of learning representation and to learn a deep feature representation that is both semantically meaningful and domain invariant
  - In this paper, we focus on "how to utilize deep networks to learn "good" feature representations with extra training criteria
- One-step DA

One-step DA Approaches	Brief Description	Subsettings
Discrepancy-based	fine-tuning the deep network with labeled or unlabeled target data to diminish the domain shift	class criterion [118], [86], [79], [98] [53], [45], [75], [139], [130], [29], [118], [28]
		statistic criterion [74], [130], [73] [75], [120], [32], [109], [87], [144]
		architecture criterion [69], [54], [68], [95], [128], [89]
		geometric criterion [16]
Adversarial-based	using domain discriminators to encourage domain confusion through an adversarial objective	generative models [70], [4], [57]
		non-generative models [119], [118], [26], [25], [117] [85]
Reconstruction-based	using the data reconstruction as an auxiliary task to ensure feature invariance	encoder-decoder reconstruction [5], [33], [31], [144]
		adversarial reconstruction [131], [143], [59]

SECTION III: APPROACHES OF DEEP DA - ONE-STEP

- 4 major techniques for performing fine-tuning for discrepancy-based deep DA, which assumes fine-tuning deep network with target data can diminish the shift between two domains
  - Class Criterion: use class label information
  - Statistic Criterion: align statistical distribution
  - Architecture Criterion: improve abilities of learning more transferable features by adjusting network architectures
  - Geometric Criterion: Bridges source and target domains based on their geometrical properties
    - This assumes relationship of geometric structures can reduce domain shift
- 2 categories for adversarial-based deep DA
  - Generative models: based on Generative Adversarial Networks (GANs)
  - Non-generative Models: feature extractor learns a discriminative representation using labels in source domain and maps target to the same space through domain-confusion loss to results in domain-invariant representations
- 2 categories for Reconstruction-based DA approach, which assumes data reconstruction of samples can be helpful
  - * The reconstructor can ensure specificity of intra-domain representations and indistinguishability of inter-domain representations
  - Encoder-Decoder Reconstruction: using SAEs for data reconstruction
  - Adversarial Reconstruction: reconstruction error is measured by similarity between original and reconstructed images by a cyclic mapping via GAN discriminator

SECTION III: APPROACHES OF DEEP DA - MULTI-STEP

- First determine the intermediate result is more related than direct connection
- The sum of all intermediate transferring has lower information loss
- Key of multi-step DA: how to select and utilize intermediate domain
- 3 categories:
  - Hand-crafted: based on users' experience
  - Instance-based: select certain parts from auxiliary datasets
  - Representation-based: freezing previously trained network and use their intermediate representation as input

- Background
  - Most works focus on unsupervised learning as supervised DA has its limitation:
    - When only few labeled data in the target domain are available, training tends overfit to source distribution
  - Discrepancy-based has many research works
  - Adversarial-based and reconstruction-based are relatively new research topic

Homogeneous DA:

- Discrepancy-based: has limitation due to fragile co-adaptation and representation
  - Class Criterion (most basic training loss in deep DA):
    - Mostly represented with negative log likelihood with softmax, aka, softmax loss.
      - Sometimes also use semantic alignment loss or separation loss
    - Using soft label rather than hard labels can preserve the relationships between classes across domain

		Supervised DA	Unsupervised DA
Discrepancy-based	Class Criterion	√	
	Statistic Criterion		√
	Architecture Criterion	√	√
	Geometric Criterion	√	
Adversarial-based	Generative Model		√
	Non-Generative Model		√
Reconstruction-based	Encoder-Decoder Model		√
	Adversarial Model		√

What is soft label?

ANS: label with value (prob, likelihood, etc.) attached to it.

Hard label?

The assigned class is binary

pseudo label?

Model predicted labels for unlabeled data

Homogeneous DA cont':

- Discrepancy-base (cont'):
  - Statistic Criterion:
    - An effective metric for comparing dataset distribution: Maximum Mean Discrepancy (MMD)
      - By a kernel two-sample test
    - Residual Transfer Network (RTNs) adds a gated **residual layer** for classifier adaptation
  - Architecture Criterion:
    - Optimize architectures of network to minimize distribution discrepancy
  - Geometric Criterion:
    - Mitigates the domain shift by integrating intermediate subspaces on **geodesic** path from source to target domain
    - Sampling subspaces along the geodesic to form intermediate subspaces to help to find the correlations between domains
    - Data on source and target domain are projected to obtain alignment with intermediate subspace distribution

Homogeneous DA cont':

- Adversarial-based approaches:
    - Recently success by GAN: estimate generative models via adversarial process
        - Generative model, G, that extracts data distributions
        - Discriminative model, D, that distinguishes where samples is from G or training data
    - Pros:
        - Ensure network cannot distinguish between source and target domain
    - Two categories:
        - Generative, non-generative
  - Generative Models
    - Utilized synthetic target data with ground-truth annotations
    - Able to learn transformation in an unsupervised manner based on GAN
    - CoGAN: generate synthetic target data that are paired with synthetic source data
        - Achieve a domain-invariant feature space without supervision
        - Adapt input noise vector to paired images from two distributions and share the labels, i.e. can be used to train the target model
  - Non-Generative Models:
    - Domain-adversarial (DANN), gradient reversal layer (GRL), Adversarial discriminative domain adaptation (ADDA), Selective Adversarial network (SAN)

SECTION IV: ONE-STEP DA
[HOMOGENEOUS DA (RECONSTRUCTION-BASED & HYBRID)]

Homogeneous DA cont':

- Reconstruction-based approaches:
    - Data reconstruction is an auxiliary task that focuses on creating a shared representation between domains and keep the individual characteristics of each domain
  - Encoder-Decoder Reconstruction:
    - Encode input to hidden representation, then decodes them back to reconstruct vision
      - Learn domain-invariant by shared encoder and maintain domain-special representation by a reconstruction loss in source and target domain
    - Stacked denoising autoencoder (SDA): extract high-level representation
    - Marginalized SDA (mSDA): can be computed in closed-form and do not required SGD
    - Deep reconstruction classification network (DRCN)- CNN architecture with two pipelines with shared encoder:
      - First pipeline: convolutional network, supervised classification with source labels
      - Second pipeline: Deconvolutional network, optimize for unsupervised reconstructions with target data
    - Transfer Learning with Deep Autoencoder (TLDA): two encoder layers. Minimizing distance in domain distributions
  - Adversarial Reconstruction:
    - **Cycle GAN** [Zhu et al.]: translate characteristics of one image domain into the other in absence of any paired training examples.
      - Use 2 generators, rather than translator, and 2 discriminators
- Hybrid approaches:
  - Method being used Simultaneously

Heterogenous DA cont':

- Background:
    - Feature spaces of source and target are not the same
    - Dimensions of features spaces may be different
    - Divided into 2 scenarios:
        - Both contain images, divergence mainly caused by different sensory devices or style of images
        - Different types of media in source and target (text → image, etc.)
    - 2 categories for shallow DA:
        - Symmetric: learn feature transformations to project source and target features onto common subspace
        - Asymmetric transformation: transform one of the source and target features to align with other
    - 3 categories for deep DA:
        - Discrepancy-based: share first n layers between source and target domains
            - In scenario 1, we can still use class or statistic criterion by resizing images into same dimension
            - In scenario 2, where features cannot be resizing into same dimension, we need extra process, such as Transfer Neural Trees (TNTs)
        - Adversarial-based: can be used on Heterogenous
        - Reconstruction based: can be used on Heterogenous

SECTION V: MULTI-STEP DA

- Background:
  - Intermediate domains and strategy selections are problem specific
- Hand-crafted Approach:
  - Intermediate domain selected by experience
  - When many intermediate domains are required, automatic selection criterion should be considered
- Instance-based Approach:
  - Distance domain transfer learning (DDTL)
    - Gradually select unlabeled data from intermediate domains by minimizing reconstruction errors on selected instance
  - DLID (Geometric Criterion): construct intermediate domains with subset of source and target domains
- Representation-based Approach:
  - Freeze previously trained network and use their intermediate representations as input to new network
  - Progressive Network: able to accumulate and transfer knowledge to new domains over sequence of experiences
    - New neural network for each domain

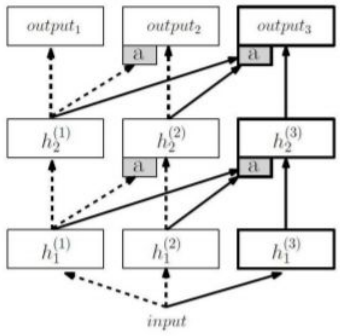


Fig. 16. The progressive network architecture. [96]

SECTION VI: DA APPLICATIONS & VII: CONCLUSION

- Application Areas:
  - Image classifications
  - Object recognitions
  - Face recognitions
  - Object detection
  - Style translations
  - Semantic Segmentation
    - Fully convolutional network (FCNs) for dense prediction is good for evaluating semantic segmentation
      - Performance degraded under domain shift
  - Image to image translation
  - Person Re-identification
  - Image Captioning
  - Etc.
- Issues remaining in the field:
  - Most algorithms focus on homogeneous deep DA
  - Limited papers on adaptation beyond classification and recognition
  - Current approach commonly assume shared label space across the source and target domains
    - May be from different set of categories in reality