PAPER ANALYSIS

Presented by Yannis He

-

Paper: **CYCADA: CYCLE-CONSISTENT ADVERSARIAL DOMAIN ADAPTATION**

Authors:

Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu | BAIR, UC Berkeley

Phillip Isola | OpenAI

Kate Saenko | CS, Boston University

Alexei A. Efros, Trevor Darrell | BAIR, UC Berkeley

https://arxiv.org/pdf/1711.03213.pdf

OVERVIEW

- Abstract
  - Current Issue: Adversarial adaptation models applied in feature spaces discover domain invariant representations, but are difficult to visualize or fail to capture pixel-level and low-level domain shifts.
  - Inspiration: GANs combined with cycle-consistency constraints are effective at mapping images between domains (CycleGANs)
  - Proposal: a discriminatively-trained Cycle-Consistent Adversarial Domain Adaptation model (CyCADA)
    - Adapts representations at both feature-level and pixel-level
    - Enforces cycle-consistency while leveraging a task loss
    - Does not requires aligned pairs
  - Applications: Visual recognition and prediction settings
    - Digit classifications, semantic segmentation of road scenes (synthetic to real world domain)
- Introduction:
  - Issue facing:
    - Deep Learning can be poor at generalizing learned knowledge to new dataset/environments
    - Non-photorealistic synthetic data to real images has a significant challenge.
    - 2 limitations for previous work: Feature-level unsupervised domain adaptation methods by aligning the features extracted from network across source and target domains
      - Aligning marginal distributions does not enforce any semantic consistency
      - Alignment at higher levels of deep representations can fil to model aspects of low-level appearance variance
    - Recent methods using unsupervised data for both domain only works for small image size and limited domain shift

OVERVIEW (CONT')

- Introduction (cont'):
  - Pixel-level domain adaptation models perform similar distribution alignment in raw pixel space
    - Translating source data to the "style" of a target domain.
  - Goal: Encourage the model to preserve semantic information in the process of distribution alignment
  - Proposal: CyCADA model
    - Using a reconstruction (cycle-consistency) loss to encourage cross-domain transformation to preserve local structural information and a semantic loss to enforce semantic consistency
  -

RELATED WORK

- Discriminative Representation Space:
  - Pairwise metric transform solution (2010)
  - Feature space alignment through minimizing distance between first/second order feature
  - Generative Adversarial Networks (GANs) (2014)
- Adaptation in pixel-space using generative approach (more human interpretable):
  - CoGANs (2016)
  - Reconstruction objective in target t=domain to encourage alignment in unsupervised adaptation (2016)
- Directly convert target image into a source style image (or visa versa), largely based on GANs
  - Image generation (2015)
  - Image editing (2016)
  - Feature learning (2016, 2017)
  - Conditional GANs (2014)
  - Image-to-image translation with input-output image pairs (2016)
- Image reconstruction without input-output image pairs
  - Domain transfer network (2017)
    - generator that transform a source image into target image by enforcing consistency in embedding space
    - Good at limited domain shifts, where domains are similar in pixel-space
    - Limited when large domain shift requires
  - Content similarity loss (prior knowledge required) 2017
  - Simultaneously learn transformation between pixel and latent space: BiGAN (2017), ALI (2016)
  - CycleGANs (2017)

CYCLE-CONSISTENT ADVERSARIAL DOMAIN ADAPTATION

- Goal: Unsupervised Adaptation: Provide source data X_S, source labels Y_S, and target data X_T, we can get a model f that can correctly predict the label for target data X_T
- Approach: Learning a source model f_S that can perform task on source data.
  - For K-nary classification, the cross-entropy loss:

$$\mathcal{L}_{\text{task}}(f_S, X_S, Y_S) = -\mathbb{E}_{(x_s,y_s)\sim(X_S,Y_S)} \sum_{k=1}^{K} \mathbb{1}_{[k=y_s]} \log\left(\sigma(f_S^{(k)}(x_s))\right)$$, σ *:softmax function*

  - While performing well on source data, domain shift to target domain leads to reduced performance when evaluating on target data.
    - To mitigate the negative effects, we learn to map samples across domain using adversarial adaptation approaches such that the discriminator is unable to distinguish the domains
  - By mapping samples into a common space, the model is able to learn on source data while still generalizing target data
- Proposal: a mapping from source to target $G_{S\to T}$ and train in to produce target samples that fool an adversarial discriminator D_T while the discriminator attempts to classify the real target data from source data:
  - The loss is $\mathcal{L}_{\text{GAN}}(G_{S\to T}, D_T, X_T, X_S) = \mathbb{E}_{x_t\sim X_T}[\log D_T(x_t)] + \mathbb{E}_{x_s\sim X_S}[\log(1 - D_T(G_{S\to T}(x_s)))]$ (2)
  - Cons:
    - Can often be unstable and prone to failure in practice
    - Can't guarantee that $G_{S\to T}(x_s)$ preserves the structures or content of the original samples x_s

CYCLE-CONSISTENT ADVERSARIAL DOMAIN ADAPTATION (CONT')

- Proposal (cont')
  - Solution (pixel-level):
    - Impose a cycle-consistency constraint on the adaptation method and introduce another mapping $G_{T \to S}$
    - Encourage high semantic consistency before and after image translation
    - Pretrain a source task model f_S, fixing the weight and use it as a noisy labeler that encourage an image to be classified in the same way before and after the translation
    - The semantic consistency is defined as:

$$\mathcal{L}_{\text{cyc}}(G_{S\to T}, G_{T\to S}, X_S, X_T) = \mathbb{E}_{x_s\sim X_S}\left[||G_{T\to S}(G_{S\to T}(x_s)) - x_s||_1\right] + \mathbb{E}_{x_t\sim X_T}\left[||G_{S\to T}(G_{T\to S}(x_t)) - x_t||_1\right].$$

  - Solution (feature-level):
    - Discriminates between the features or semantics from two image sets as viewed under a task network
      - The GAN loss: $\mathcal{L}_{\text{GAN}}(f_T, D_{\text{feat}}, f_S(G_{S\to T}(X_S)), X_T)$

Taken together, these loss functions form our complete objective:

$$\begin{aligned}\mathcal{L}_{\text{CyCADA}}(f_T, X_S, X_T, Y_S, G_{S\to T}, G_{T\to S}, D_S, D_T) \quad (6)\\ = \mathcal{L}_{\text{task}}(f_T, G_{S\to T}(X_S), Y_S)\\ + \mathcal{L}_{\text{GAN}}(G_{S\to T}, D_T, X_T, X_S) + \mathcal{L}_{\text{GAN}}(G_{T\to S}, D_S, X_S, X_T)\\ + \mathcal{L}_{\text{GAN}}(f_T, D_{\text{feat}}, f_S(G_{S\to T}(X_S)), X_T)\\ + \mathcal{L}_{\text{cyc}}(G_{S\to T}, G_{T\to S}, X_S, X_T) + \mathcal{L}_{\text{sem}}(G_{S\to T}, G_{T\to S}, X_S, X_T, f_S).\end{aligned}$$

This ultimately corresponds to solving for a target model f_T according to the optimization problem

$$f_T^* = \arg\min_{f_T} \min_{\substack{G_{S\to T}\\ G_{T\to S}}} \max_{D_S, D_T} \mathcal{L}_{\text{CyCADA}}(f_T, X_S, X_T, Y_S, G_{S\to T}, G_{T\to S}, D_S, D_T). \quad (7)$$

CYCLE-CONSISTENT ADVERSARIAL DOMAIN ADAPTATION (CONT')

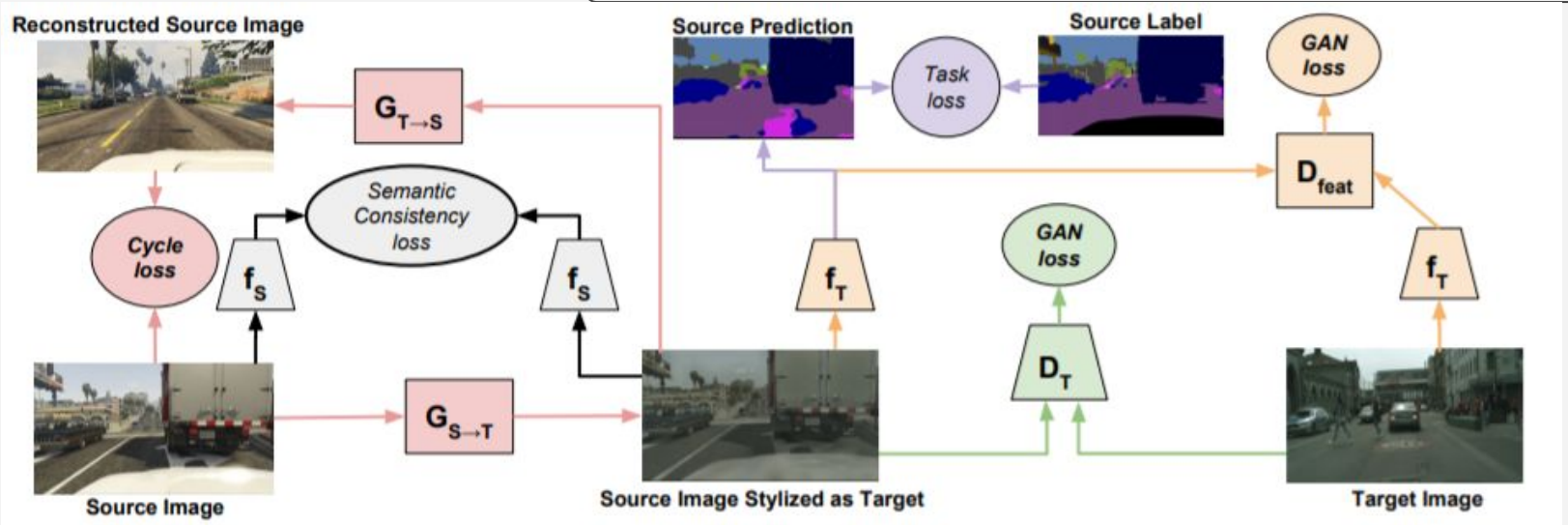

Figure 2: Cycle-consistent adversarial adaptation of pixel-space inputs. By directly remapping source training data into the target domain, we remove the low-level differences between the domains, ensuring that our task model is well-conditioned on target data. We depict here the image-level GAN loss (green), the feature level GAN loss (orange), the source and target semantic consistency losses (black), the source cycle loss (red), and the source task loss (purple). For clarity the target cycle is omitted.

EXPERIMENTS

- Digit Adaptation:
  - Dataset: shifts from USPS to MNIST, SVNH to MNIST, and MNIST to USPS
  - Ablation: Pixel vs Feature level Transfer
    - USPS <-> MNIST (small domain shift):
      - CycleGAN works well
      - Feature level adaptation offers a small benefit
    - SVHN <-> MNIST:
      - Feature level adaptation outperforms the pixel level adaptation
  - Ablation: No Semantic Consistency
    - SVHN <-> MNIST:
      - CycleGAN often diverged and suffering from random label flipping
      - Use the source labels to train a weak classification model can be used to enforce semantic consistency
  - Ablation: No Cycle Consistency:
    - No reconstruction guarantee in this case
    - While semantic loss does encourage correct semantics, it relies on the weak source labeler
      - Thus label flipping still occurs

EXPERIMENTS (CONT')

- Semantic Segmentation Adaptation:
  - Task: assign a semantic label to each pixel in the input
  - Evaluation: limited to the unsupervised adaptation setting (label available in source domain only) but solely evaluate the performance in the target domain
  - 3 Metrics: mean intersection-over-union, frequency weighted intersection-over-union, pixel accuracy
    - $$\text{mIoU} = \frac{1}{N} \cdot \frac{\sum_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}, \text{fwIoU} = \frac{1}{\sum_k t_k} \cdot \frac{\sum_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}, \text{pixel acc.} = \frac{\sum_i n_{ii}}{\sum_i t_i}$$
      - n_{ij} : number of pixels of class i predicted as class j
      - t_i : total number of pixels of class i
      - N: the number of class
  - Dataset: SYNTHIA dataset (2016)
- Cross-Season Adaptation:
  - Architecture: FCN8s
  - Dataset: SYNTHIA
  - Usually, it is difficult to interpret what causes the performance improvement after adaptation under unsupervised adaptation
    - To find out, we inspect the intermediate pixel-level result during domain shift
- Synthetic to Real Adaptation
  - Saturation level in game (GTA5) is more vivid
  - Texture will change too