Source-free Domain Adaptation: A brief survey

Yongxin Wang

Oct. 30,2022

Why Source-free?

- Private information, e.g., those on personal phones or from surveillance cameras.
- Existing DA methods must access the source data, violating the data privacy policy.
- The storage size of a trained model is much smaller than that of a compressed dataset.

Overview of SFDA

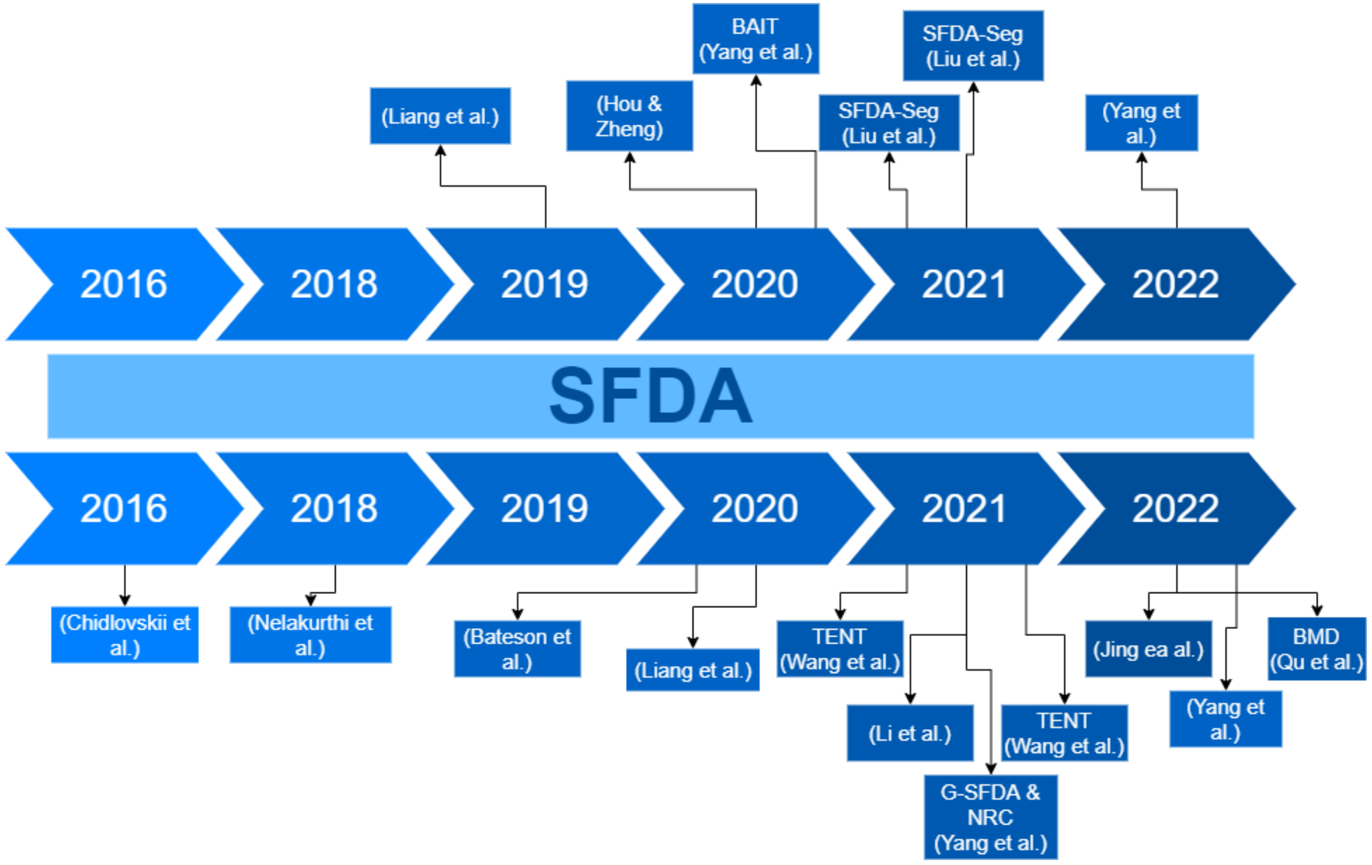


Figure 1. Timeline of SFDA from 2016 to the present. The upper stream refers to the development of Virtual Source Knowledge Transfer, while the lower stream represents the development of Self-supervised Training.

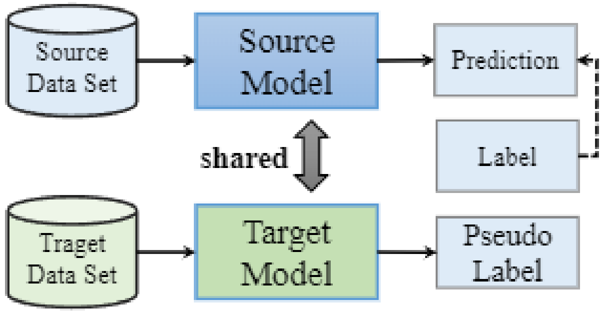


Figure 2. Overview of traditional UDA.

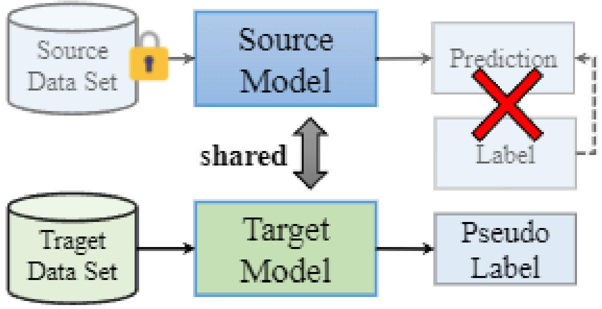


Figure 3. Overview of Source-free Domain Adaptation.

Two Approaches

Self-supervised Training

- Pseudo-label Clustering
- Pseudo-label Filtering

Virtual Source Knowledge Transfer

- Source Impression
- Style Translation

Self-supervised Training

- **Pseudo-label Clustering**
- **Pseudo-label Filtering**

Self-supervised Training

Pseudo-label Clustering

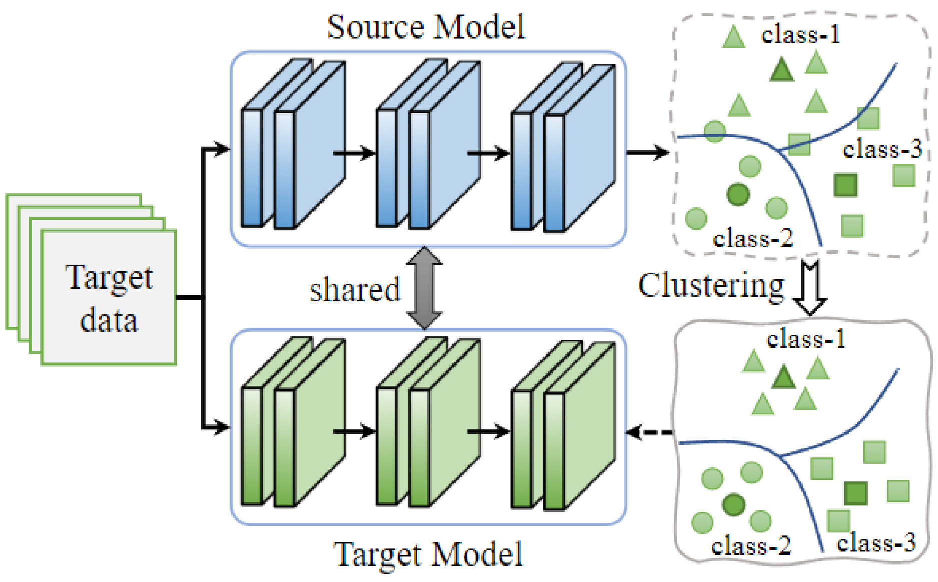


Figure 5. Overview of Pseudo-label Clustering in SFDA.

Pseudo-label Filtering

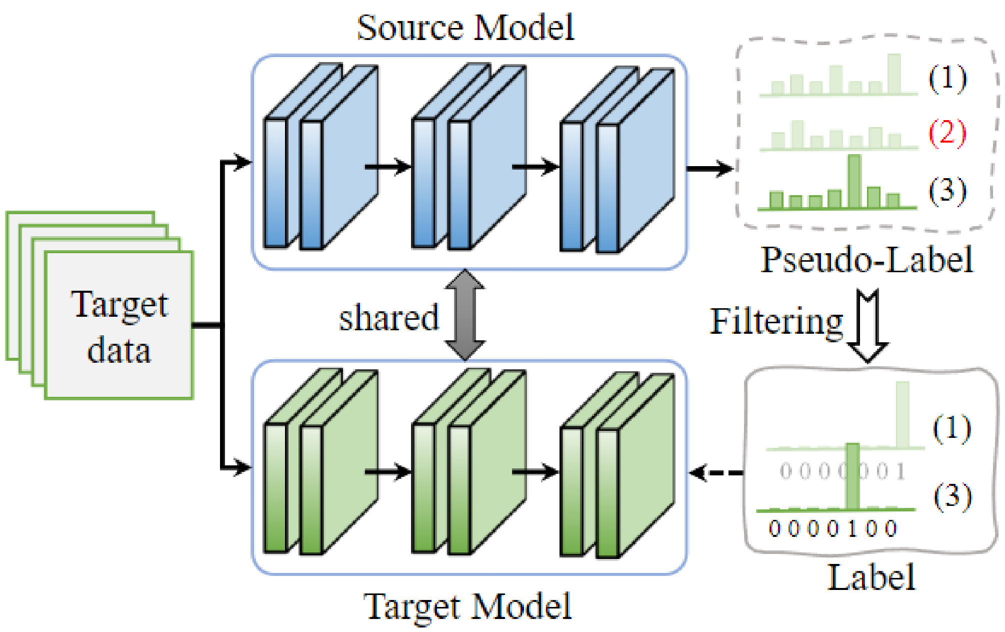
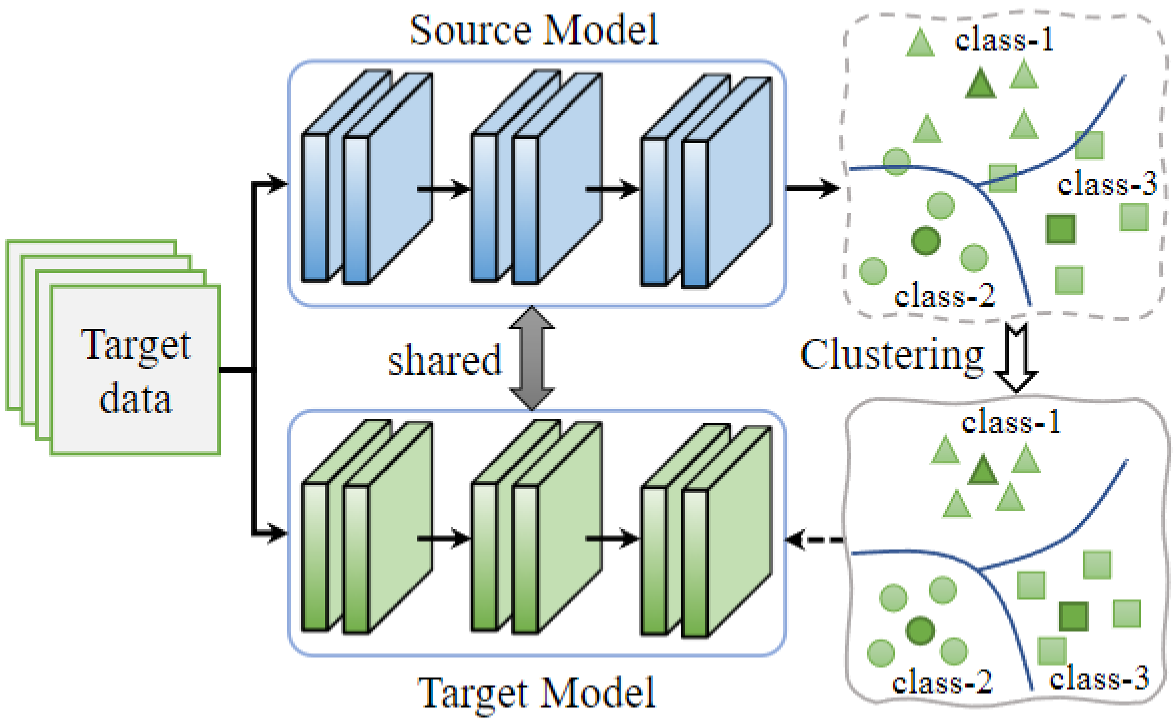


Figure 6. Overview of Pseudo-label Filtering in SFDA.

Pseudo-label Clustering

- Firstly, the noisy pseudo labels could be **generated by the source model** with the target data input.
- Then they could be further categorized through **clustering algorithms**.
- Based on the **target data with calibrated pseudo labels**, the target model could be learned.



Pseudo-label Clustering

SHOT[1]

- **SHOT-IM**: IM would work better than conditional entropy minimization. To make the target outputs individually certain and globally diverse, SHOT adopts the IM loss.
- **Pseudo-label Clustering**: Inspired by DeepCluster, the authors propose a novel self-supervised pseudo-labeling strategy. They attain the **centroid** for each class in the target domain, similar to weighted k-means clustering. Then they obtain the pseudo labels via the **nearest centroid classifier**. Finally, we **compute the target centroids** based on the new pseudo labels. Updating for multiple rounds is needed.
- **Tricks**: label smoothing

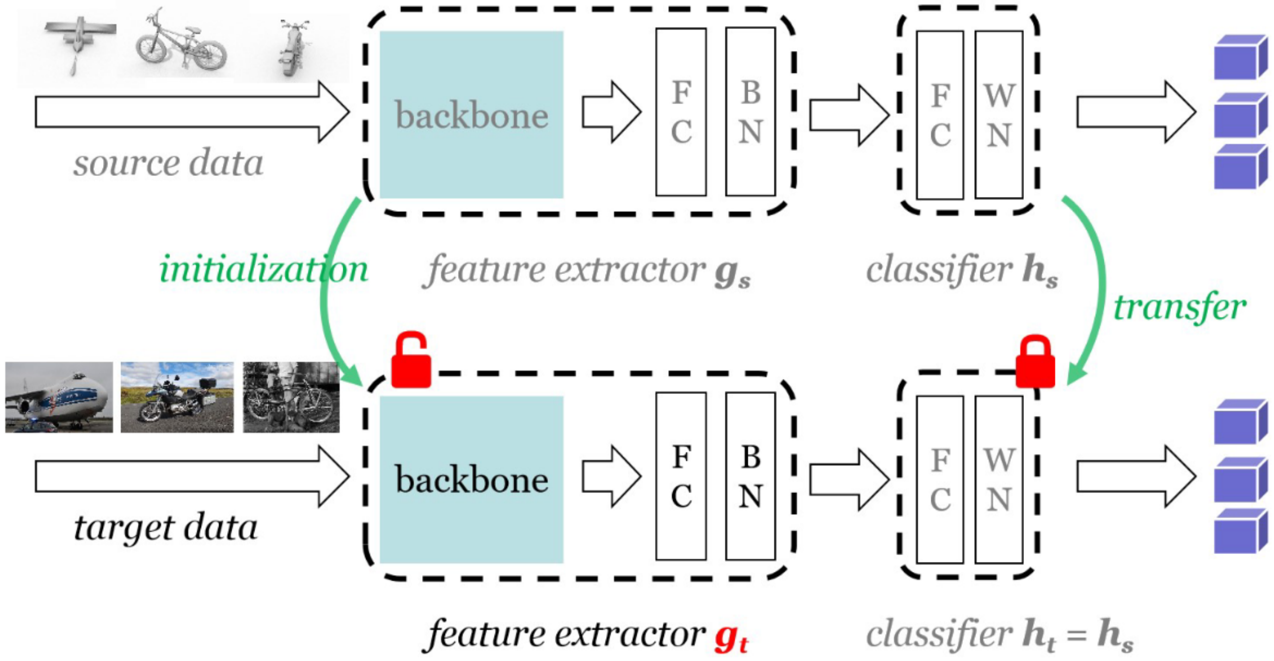


Figure 7. The overview of SHOT(Source HypOthesis Transfer)

Pseudo-label Clustering

SHOT++ [2]

- **SHOT-IM++**: Modification on Loss Function.
- **Pseudo-label Clustering++:**Rotation prediction aims to recognize one of four different 2d rotation (i.e., 0◦, 90◦, 180◦, and 270◦). **Augment** the sample space, which **enhances** the learning of feature extraction and target classifier.
- **Labeling Transfer with Semi-supervised Learning**: Dividing the target domain into two splits according to **the confidence scores** and treating these two splits as a **labeled** subset and an **unlabeled** subset, respectively. Then employs a **semi-supervised learning** algorithm to learn the enhanced predictions for the unlabeled set.

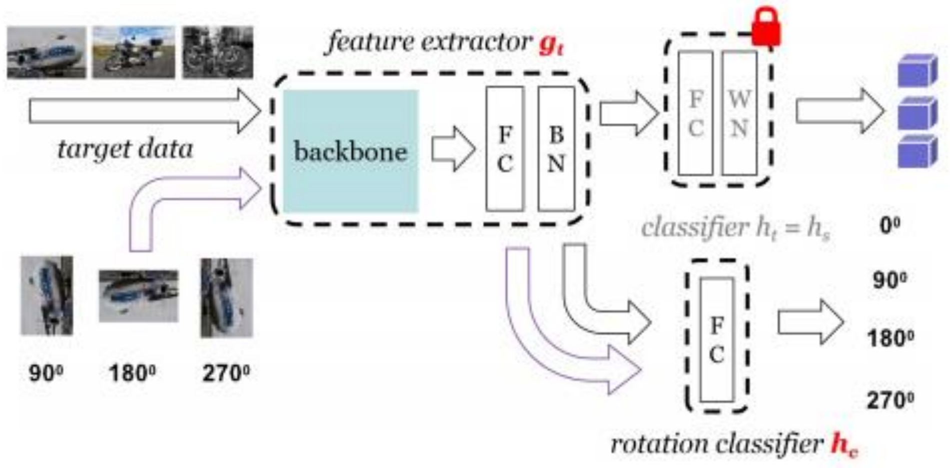


Figure 8. Overview of Rotation Prediction.

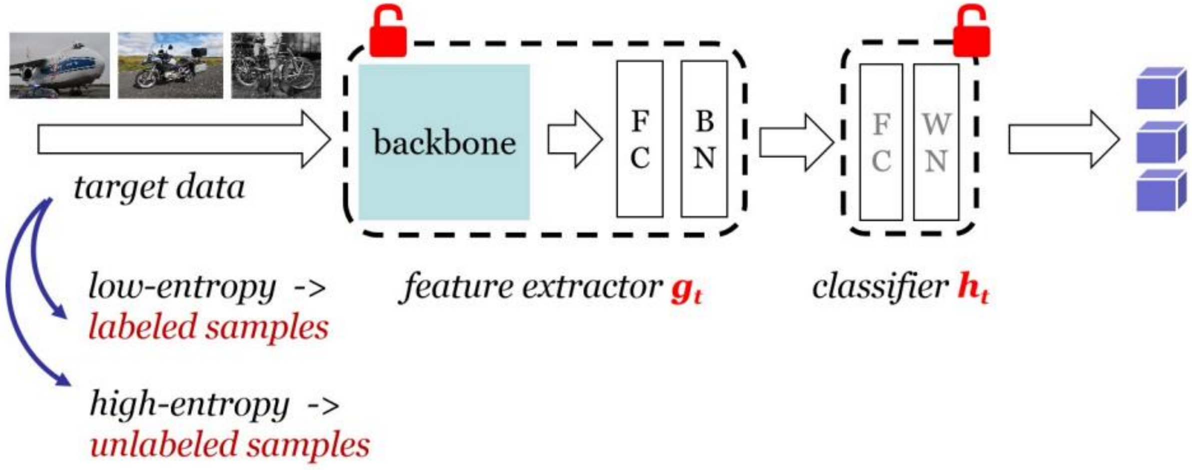


Figure 9. Labeling Transfer with Semi-supervised Learning.

Pseudo-label Clustering

G-SFDA [3]

- **Local Structure Clustering**: Some target features from the source model will **deviate from dense source feature regions** due to domain shift. The authors build a **feature bank**, which stores the target features, and a **score bank** storing corresponding SoftMax prediction scores. minimize the **negative log value** of the dot product between the prediction score of the current target sample and the stored prediction scores. Additional kl-divergence encouraging prediction balance.
- **Sparse Domain Attention**: The authors apply the source attention to mask the features, avoiding forgetting of the source domain.

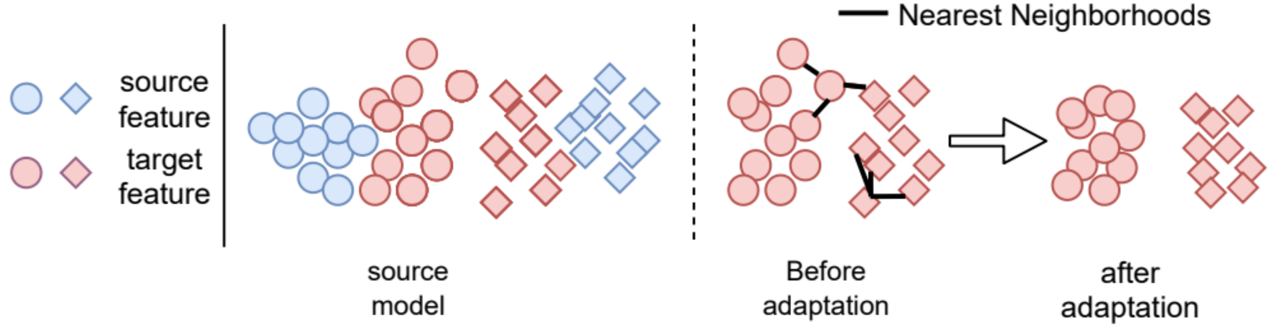


Figure 10. Local Structure Clustering (LSC). LSC aims to cluster target features by its semantically close neighbors (linked by the black line).

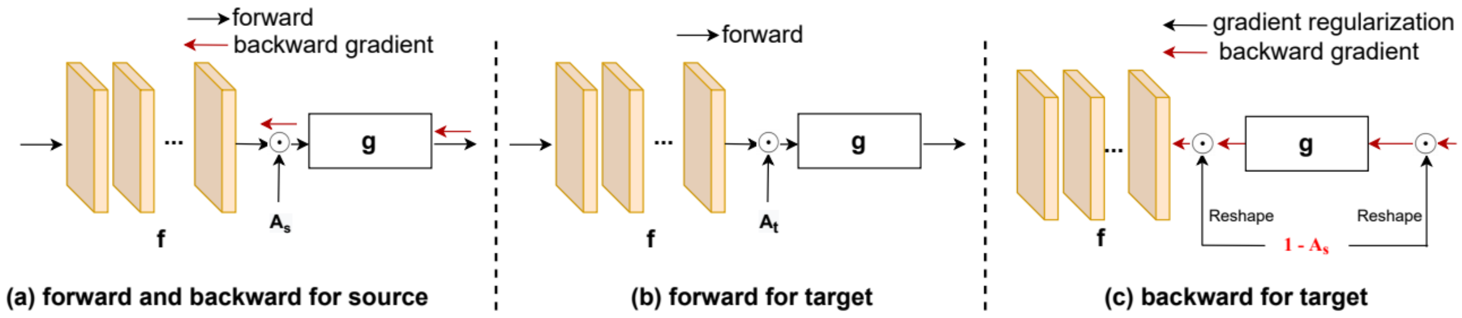


Figure 11. Forward and Backward pass for two domains.

Pseudo-label Clustering

NRC-SFDA [4]

- **Encouraging Class-Consistency with Neighborhood Affinity**: Like LSC in G-SFDA, the authors apply **reciprocal NN and non-reciprocal NN.** Then they assign a high-affinity value to the RNN features.
- **Expanded Neighborhood Affinity**: Just consider **the M-nearest** neighbors of each neighbor. They directly assign a **small affinity value r to those expanded neighbors**, since they are further than the nearest neighbors and may contain noise.

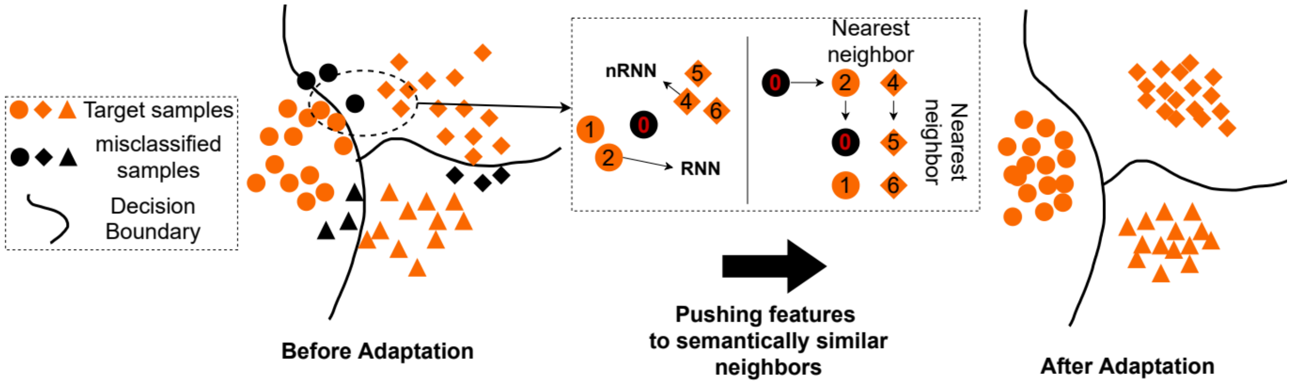


Figure 12. Illustration of NRC-SFDA. The left shows we distinguish reciprocal and non-reciprocal neighbors. The adaptation is achieved by pushing the features towards reciprocal neighbors heavily.

Pseudo-label Clustering

BMD [5]

- The visual domain gaps between source and target are typically different between categories, resulting in relatively higher prediction confidence scores for **those easy-transfer** classes in the target domain.
- **Inter-class Balanced Prototype**: Like a **MIL** problem, for a specific class k, we treat the target domain D_t as a combination of a positive bag and a negative bag.
- Then aggregating the **top-M** scores represented instances along all target domain D_t for the **kth** class as potential instances.

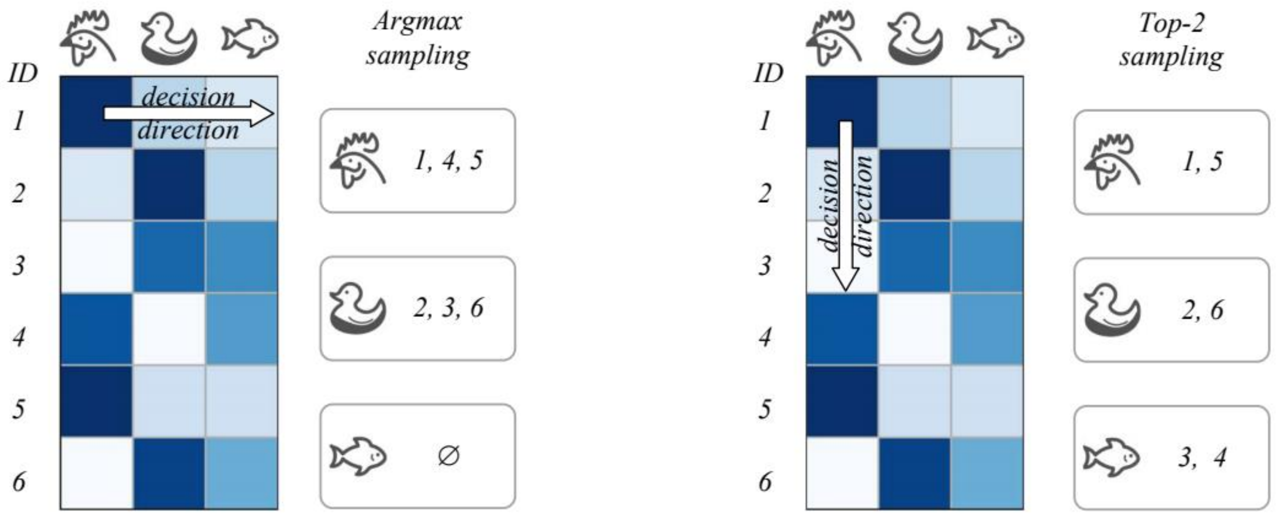


Figure 13. An example compares the existing class-biased strategy (left) with BMD class-balanced strategy (right).

Pseudo-label Clustering

BMD [5]

- **Intra-class Multicentric Prototype**: A coarse monocentric feature prototype may not effectively represent those **ambiguous** data and even introduce **negative transfer**.
- **Dynamic Pseudo Label**: At the beginning of each epoch, the model first updates the **multiple features prototype** for each class and the **corresponding pseudo labels** for each instance from a global perspective. And then, for each iteration step, we update the feature prototypes as the **exponential moving average (EMA)** of the cluster centroids in mini-batches.
- When DPL: Using SCE loss, instead of CE loss.($\alpha L_{st} + \beta\ L_{dym}$)

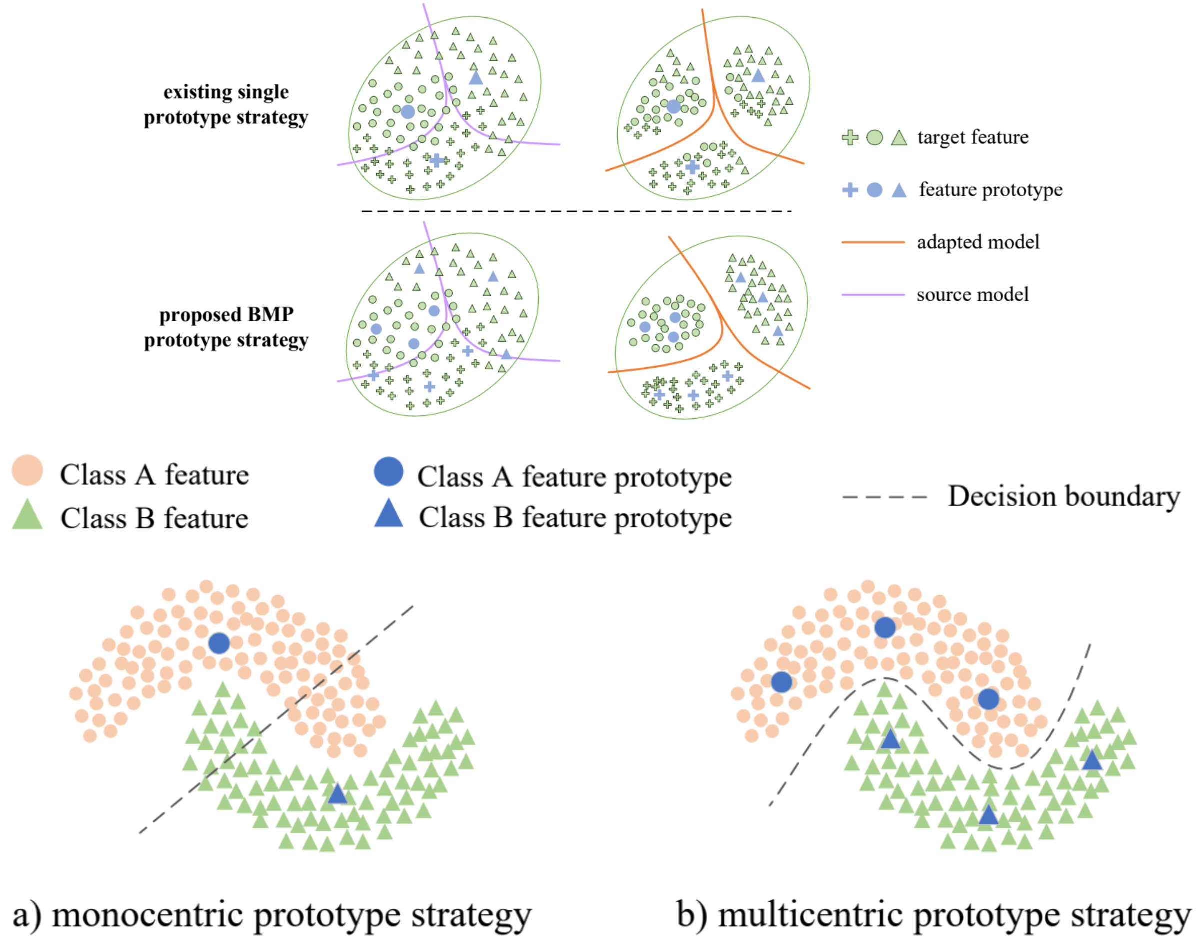
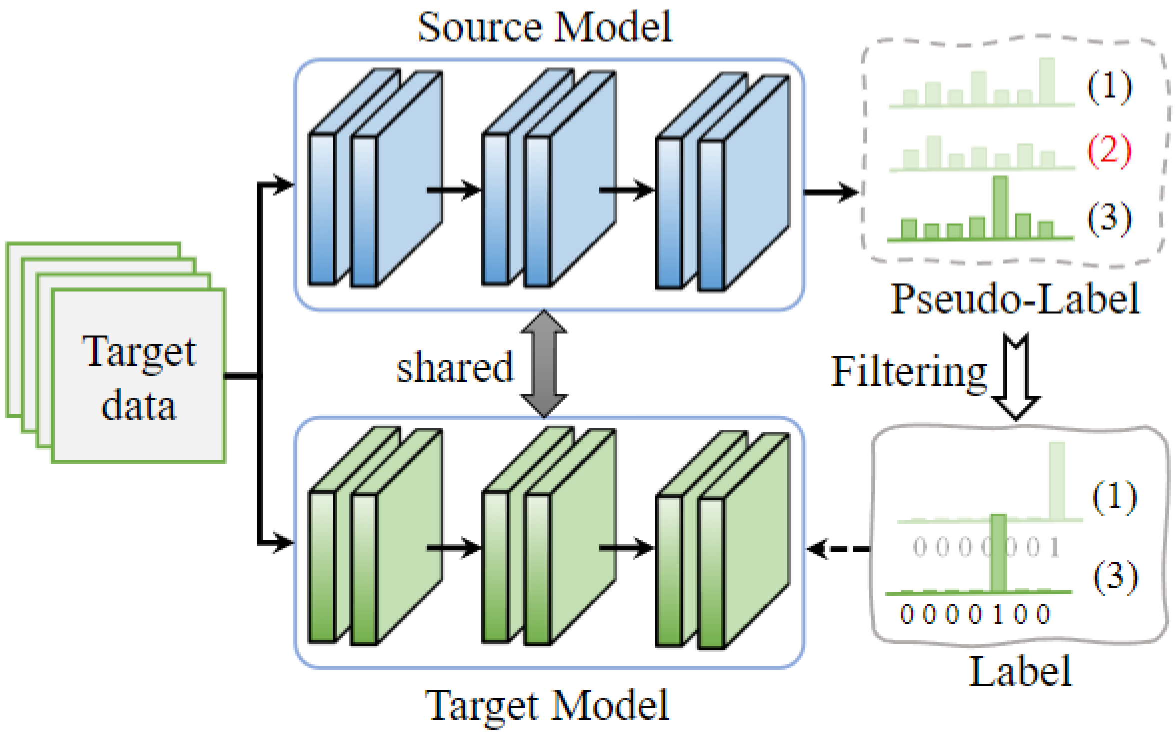


Figure 14. Comparison between the existing monocentric prototype strategy (left) and BMD proposed multicentric prototype strategy (right).

Pseudo-label Filtering

- Despite the absence of source data, some target samples could spread **around the corresponding source prototypes** and are very **similar to the source domain**.
- These target samples could be used to **approximate the source domain**.
- This kind of SFDA method usually filters the target pseudo-labels by **splitting the target data into two subsets**, i.e., pseudo-source set and remaining target set.
- They correspond to source hypothesis keeping and target knowledge exploration **respectively**.



Pseudo-label Filtering

BAIT [6]

- **Splitting feature** in current batch into 2 sets by prediction entropy H and the threshold τ.
- **Increasing prediction divergence** between two classifiers for uncertain features but keeping the prediction unchanged for the uncertain feature.
- **Maximizing KL divergence** can also prevent the bait classifier from moving to the undesirable position (dashed red line).
- Training feature extractor pushes all features to the same side of C_1 and C_2.

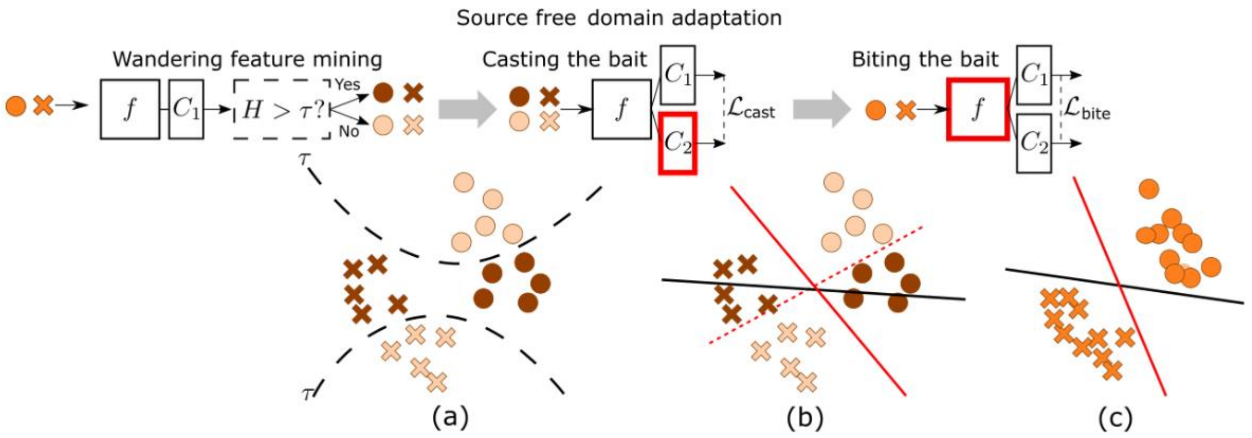


Figure 15. Illustration of the BAIT training process.

Pseudo-label Filtering

A²Net [8]

- **Soft-Adversarial Inference**: To solve the challenge is to distinguish source-similar features from source-dissimilar ones, motivated by the voting strategy, authors compare the output of classifiers to adaptively determine the type of features.
- **Contrastive Category-wise Matching**: Inspired by contrastive learning, we design a novel discriminative **dual classifier** exploring the association of paired samples to achieve **class-wise alignment** in an unsupervised manner.
- **Self-Supervised Rotation**: **Augment** the sample space, which **enhances** the learning of feature extraction and target classifier.

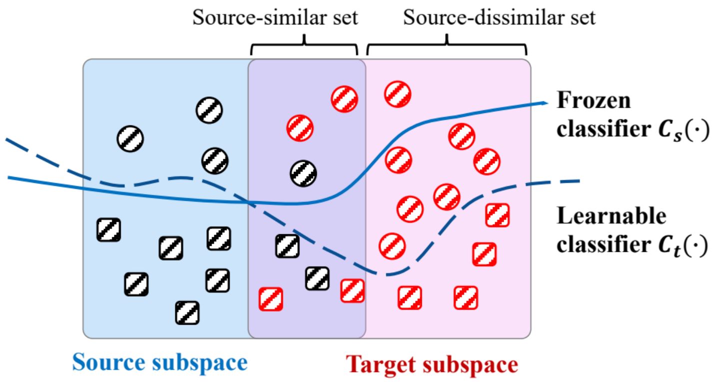


Figure 17. The target samples can be divided into two subsets: source-similar and source-dissimilar sets. A^2 Net adaptively learns a new classifier (dashed)
based on the frozen classifier (solid) trained in the source domain.

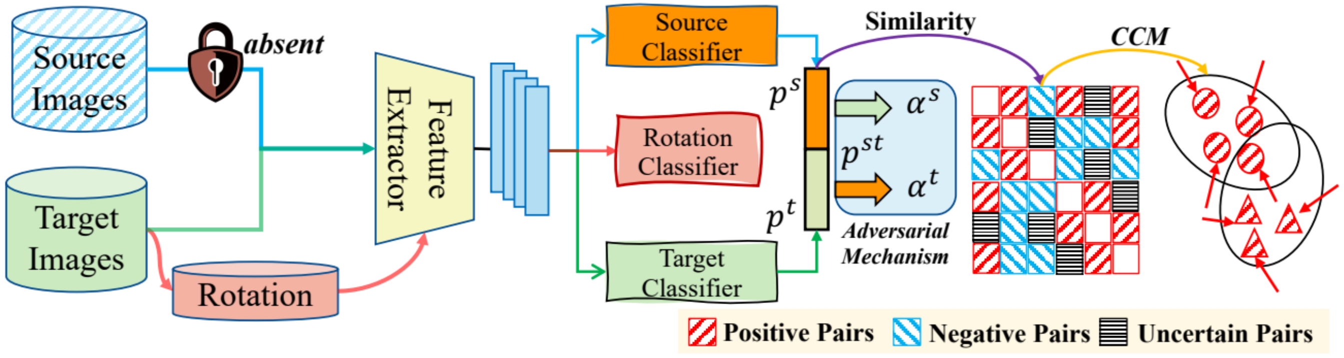


Figure 18. Overview of Adaptive Adversarial Network (A^2 Net).

Pseudo-label Filtering

CaiDA [7]

- **Source-specific transferable perception**: The authors use an MLP to apply $\sum_{i=1}^{n} \mu_i = 1$ to denote the combination of source predictions.
- **Confident-anchor-induced pseudo label generator**: The authors use the union between a **probability-based** confident anchor group and a **distance-based** confident anchor group. For each target data, the model applies **continual similarity searching** to compute the distance until the confident anchor from Union is detected.
- **Class-relationship-aware consistency loss**: It encourages target data from the same class to be compactly clustered together while preserving the **intrinsic inter-class relationships** via soft **confusion matrix** alignment.

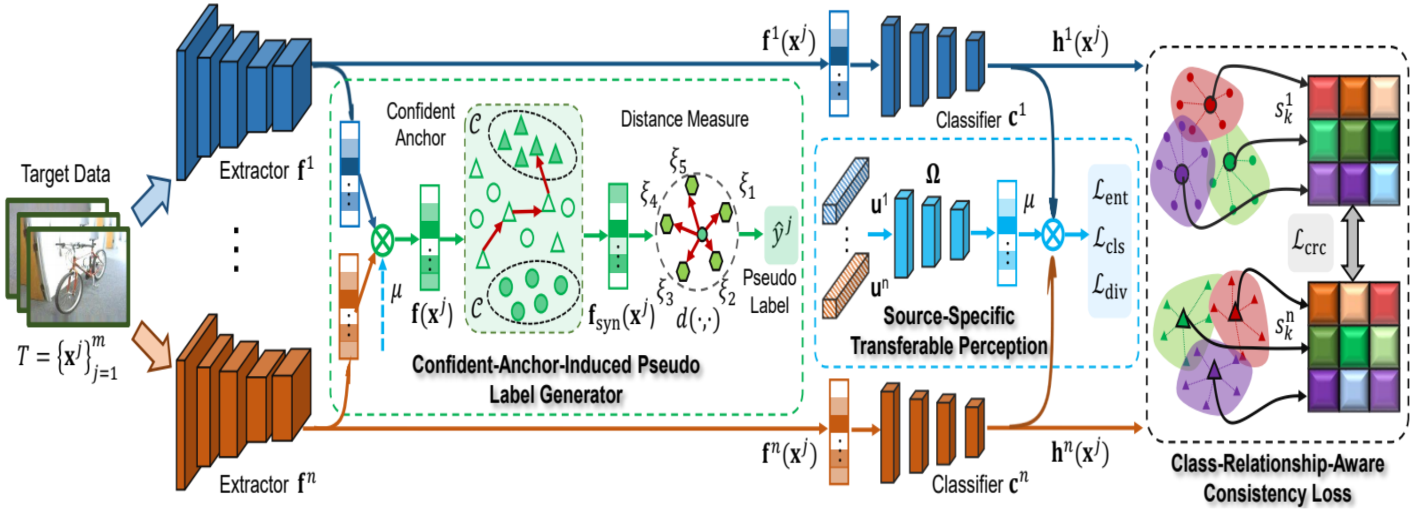


Figure 16. Overview of Confident-Anchor-induced multi-source-free Domain Adaptation(CaiDA).

Virtual Source Knowledge Transfer

- Source Impression
- Style Translation

Virtual Source Knowledge Transfer

Source Impression

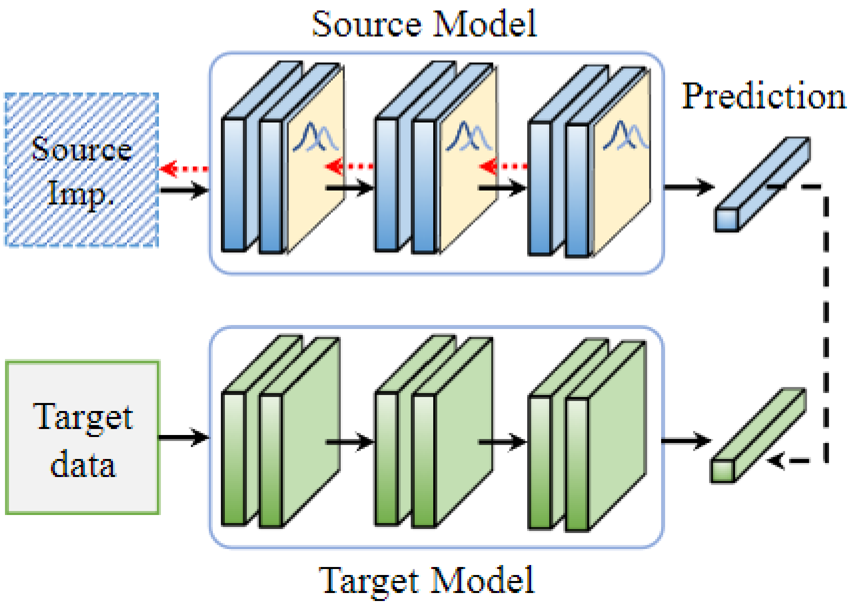


Figure 19. Overview of Source Impression.

Style Translation

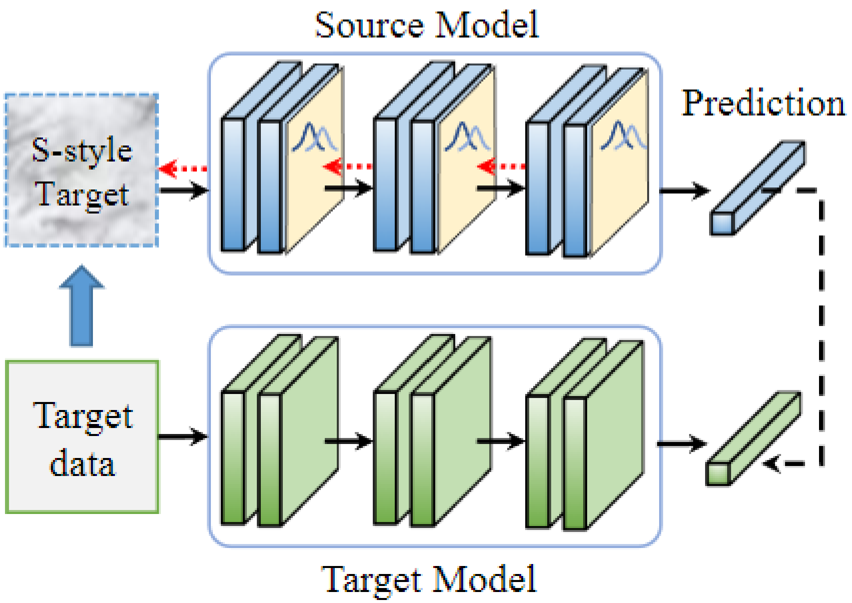


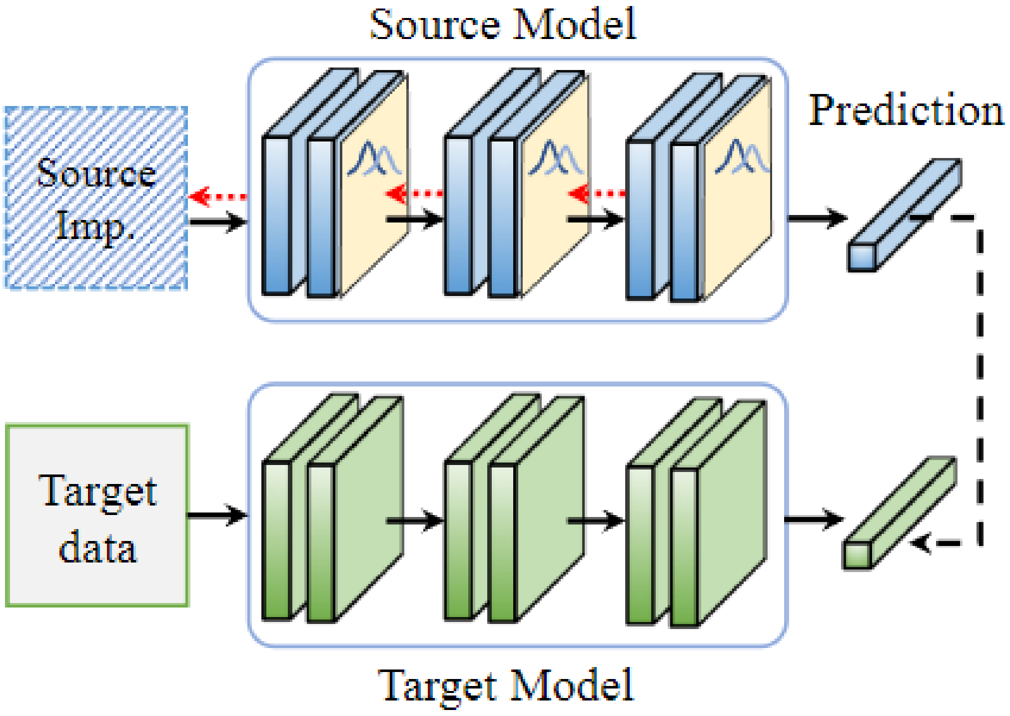
Figure 20. Overview of Style Translation.

Source Impression

- For tackling the issue of source data absence, we can synthesize the impressions of the source domain for joint training or knowledge transfer in adaptation.

- We can introduce a **generative adversarial framework** to synthesize **source impressions with the supervision** of the source model prior and target images.

- Rather than generating source impressions, we can also model **an intermediate virtual domain** in the feature space based on a **Gaussian mixture model**.



Source Impression

Model Adaptation [11]

- **Collaborative Class Conditional GAN**: The authors propose a **semantic similarity loss** based on the existing prediction model. It enforces the semantic similarity between x_g and the input label Y.
- **Weight Regularization**: They propose a weight regularization term ℓ_{wReg} to **prevent the parameters** of the predi-ction model C to **drift far away from those** of the pre-trained model learned in the source dataset.
- **Clustering-based Regularization**: The cluster assumption implies that the decision boundaries of the prediction model should not go through data regions with high density. They minimize the **conditional entropy** of the predicted probability of the target domain. In essence, that makes the output more like a one-hot.

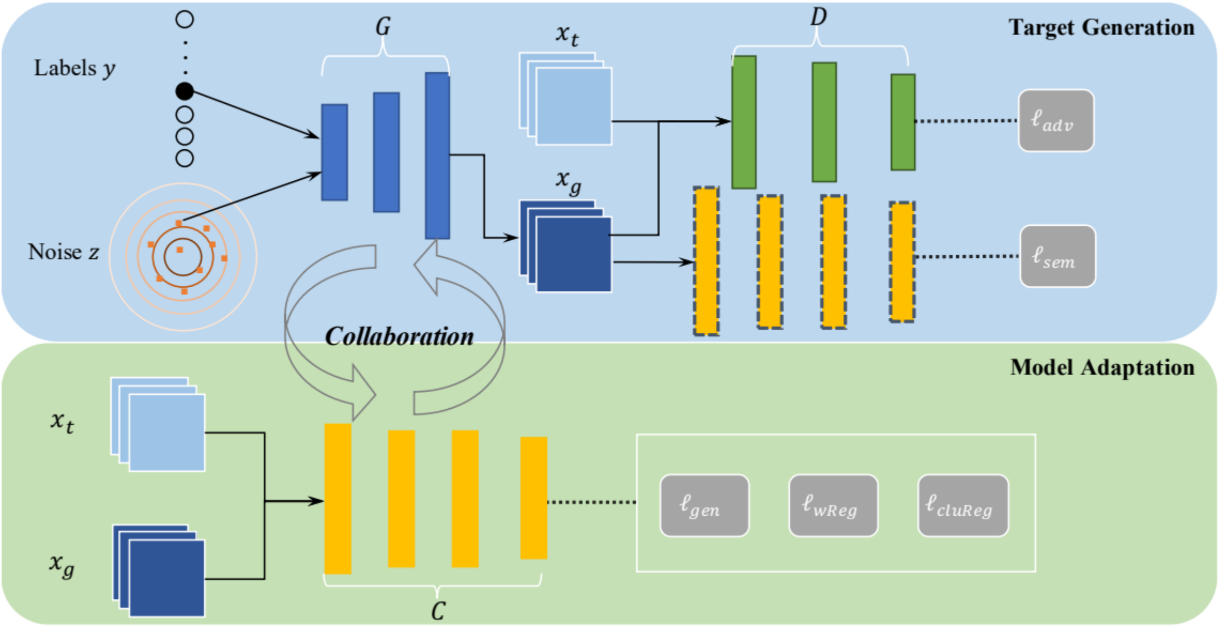


Figure 21. Overview of Style Translation. During target generation (top), we aim to learn a class conditional generator G for producing target-style training samples via the discriminator D and the prediction model C (which is fixed as denoted by the dashed line). The generated images and proposed regularize-ations are used for model adaptation (bottom).

Source Impression

Domain Impression[12]

- **CGAN**

- **Gradient Reversal Layer:** The discriminator's objective is to guide the feature extractor to produce domain-invariant features.

- **Likelihood-based loss**: This process required a **maximize the log-likelihood** of data obtained from the generative models.

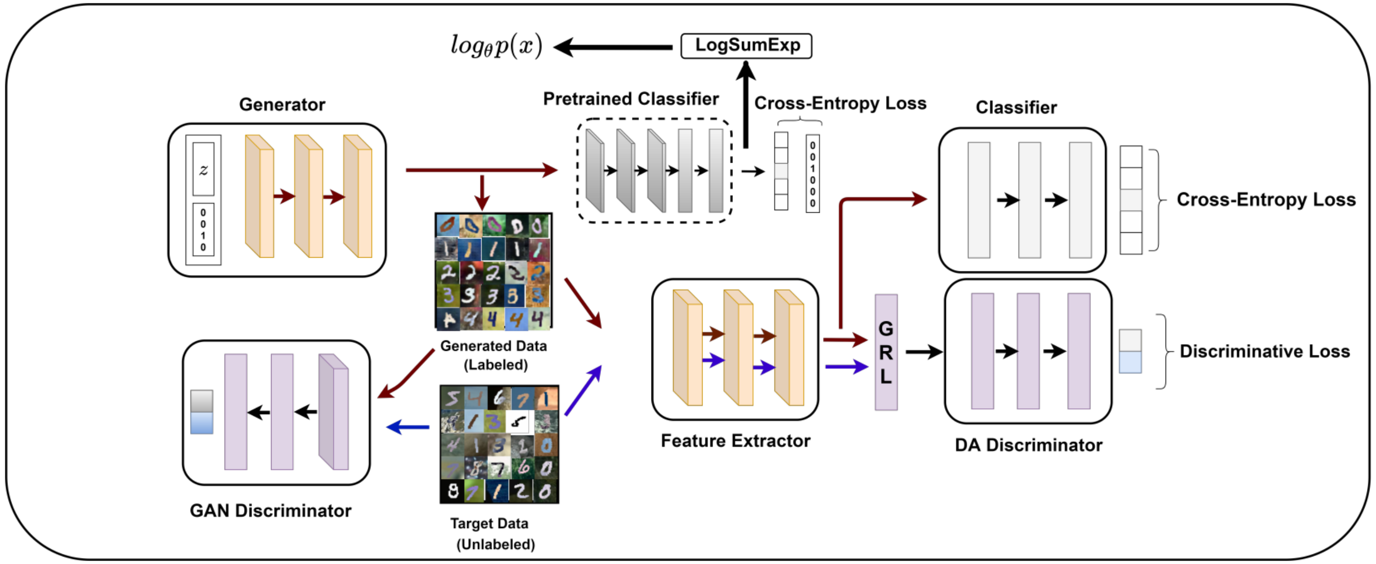


Figure 22. Overview of Style Translation. The Generator (G), GAN discriminator(Dg), Feature extractor, Classifier, and Domain discriminator are trainable while the pre-trained Classifier is set to freeze. z is the latent noise vector. GRL is the gradient reversal layer.

Source Impression

VDM-DA[15]

- **Virtual Domain Modeling**: Rather than generating source impre-ssions, VDM-DA models an **interm-ediate virtual domain** in the feature space based on a Gaussian mixture model(GMM).
- **Target and Virtual Domain Alignment**: Motivated by ADDA, the model choo-ses a simple **adversarial training-based** strategy. Besides, They use a newly proposed **re-weighting mechanism** to align the target uncertain target samples with virtual domain samples.

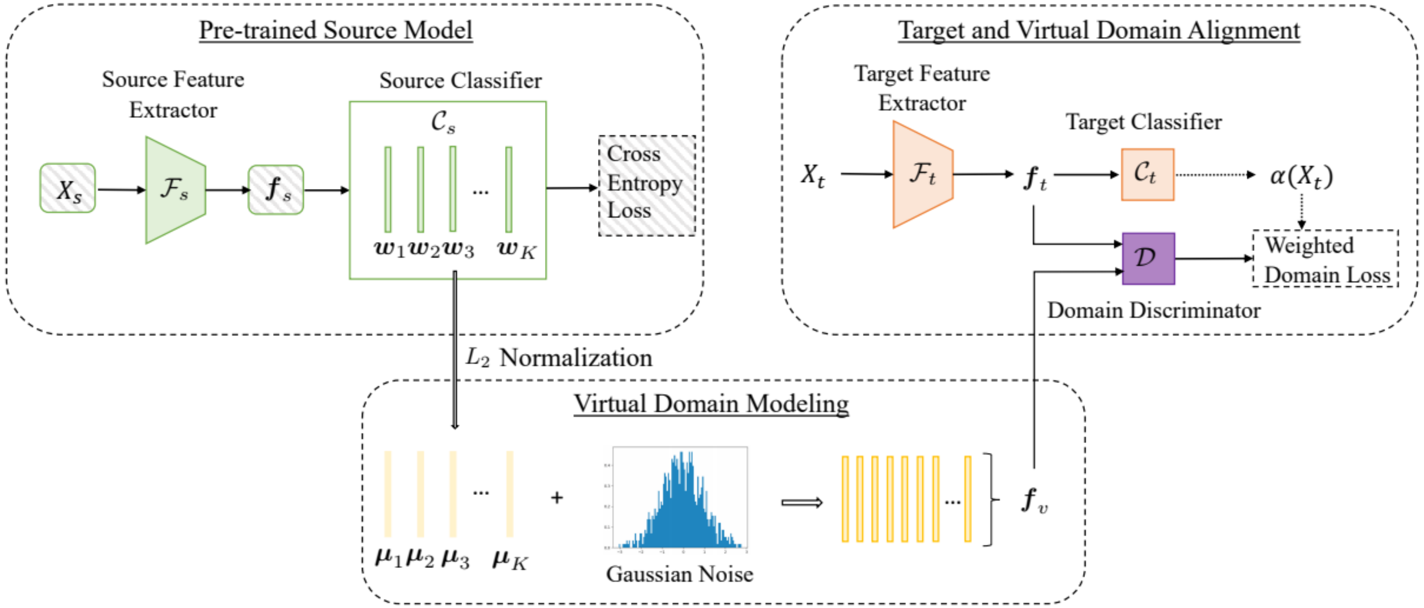
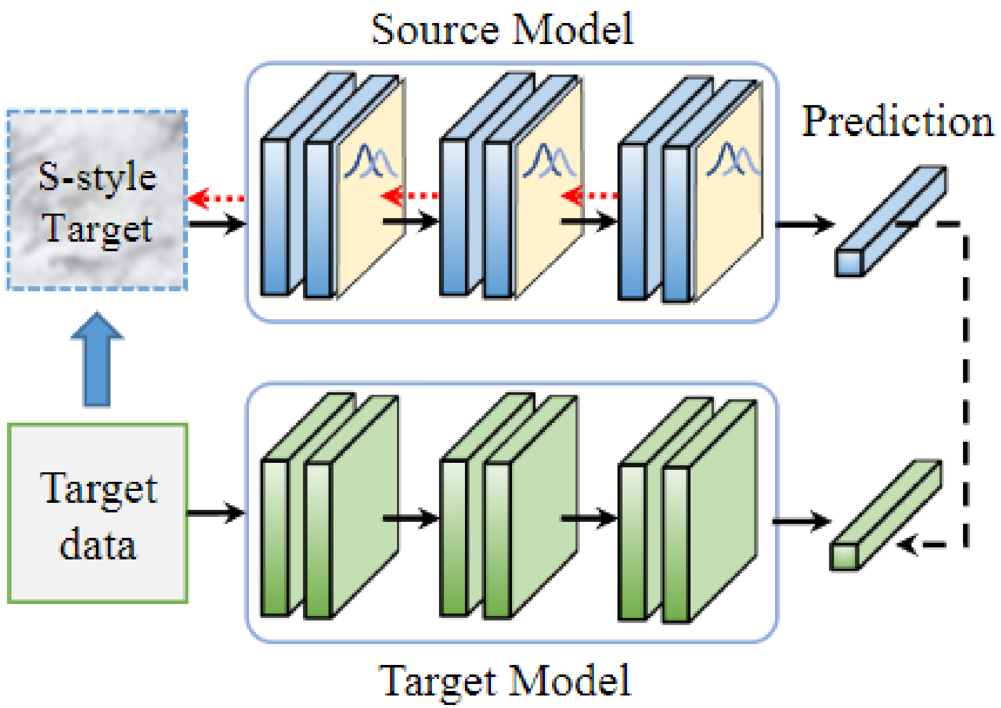


Figure 23. Overview of VDM-DA. With the pre-trained model parameters, the authors propose to first model a virtual domain whose data distribution is similar to that of the original source domain in the high-level feature space.

Style Translation

- Unlike source impressions directly reconstructed from source models, the SFDA methods utilize style translation as shown beside.
- Specifically, they transfer the target images to the source style to form **pseudo-source images**, which can be used to distill reliable target knowledge and predictions.
- In general, the style translation can **be implemented by various data augmentations**, transformations, such as **brightness**, **contrast**, etc., or style transfer based on **BN statistics**. Additionally, the style transfer could work on the image itself or the intermediate feature maps.



Style Translation

SFDA with ImageTran[13]

- **CycleGAN**: The model applies a CycleGAN to generate source-styled images.
- **Content loss**: In style transfer, the content loss encodes the **difference** of the feature maps in the top layers between **generated image** and the **content image**. The model views the original target image as the content image.
- **Style loss**: In absence of source images as style images, the model has to collect running **mean** and **variance** stored in **BN layers** as a suitable form of statistics for style alignment.

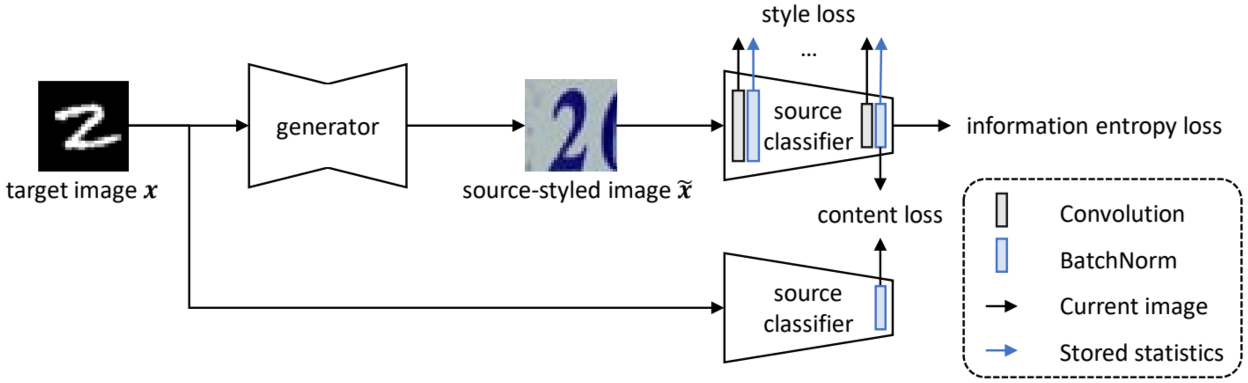


Figure 24. Overview of SFDA with Image Translation.

Style Translation

Adversarial Attacks[14]

- **Generating Adversarial Examples: The authors** employ an generator G to generate smooth and diverse perturbations for training samples x_i(we use target data when source data is absent and vice versa).

- **Harnessing Adversarial Examples**: **The authors** learn from the generated examples to generalize better to the target domain. Some additional regularizations are used to stabilize the training.
- These two steps are performed alternatively until **convergence**.

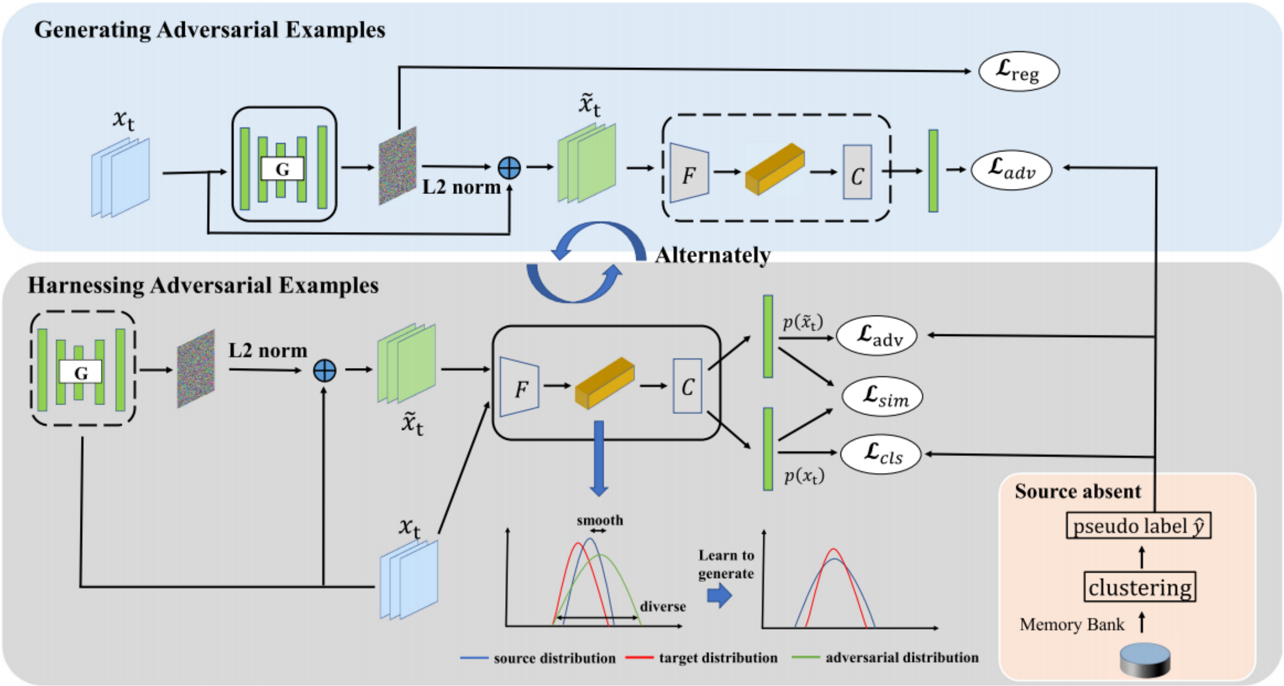


Figure 25. Overview of the method. The framework consists of two alternative steps: generating adversarial examples and harnessing adversarial examples.

Other Research

- Robust SFDA
- Concurrent Subsidiary Supervision SFDA

Other Research

Robust DA [9]

- **Non-robust Pseudo-labels**: While IM can make the model confident while ensuring diverse predictions, it may still push the output towards incorrect prediction in certain cases. The authors use **a two-step weighted k-means clustering** on the feature space to obtain pseudo-labels as described in
- **Contrastive Feature Learning**: The contrastive loss minimizes the intra-class distance while maximizing the inter-class distance between the encoder features. To optimize the target standard model, they minimize the weighted combination of the loss terms.

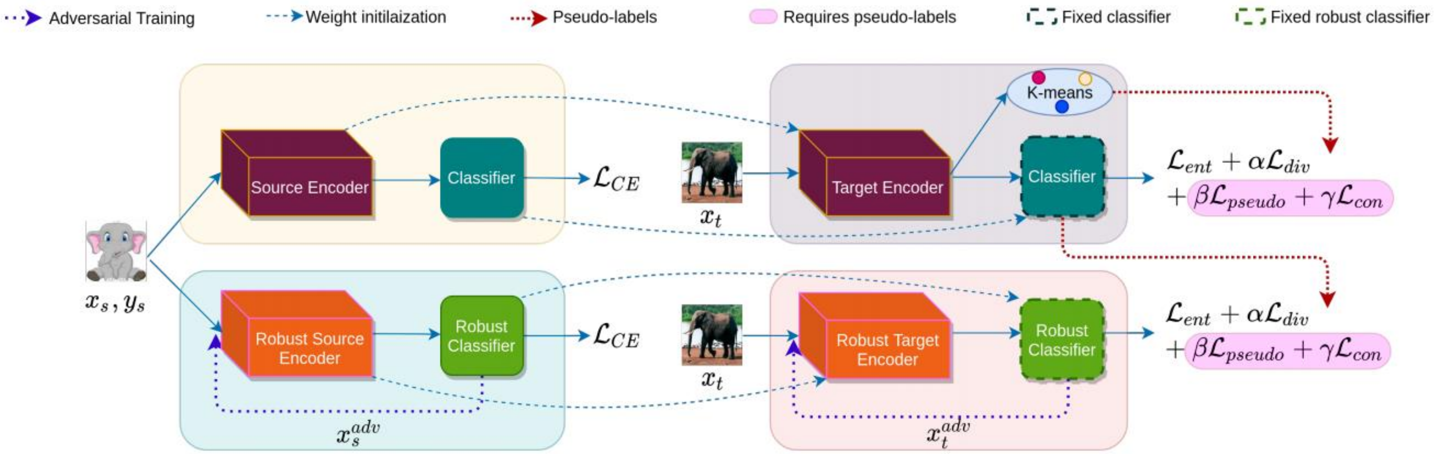


Figure 26. Overview of Standard Model and Robust Model.

Other Research

CSS-SFDA [10]

- **Subsidiary DA**: The authors introduce DSM(measures the similarity between **two domains**) and TSM(measures the task similarity of a **subsidiary** task w.r.t. the goal task).
- **Sticker intervention**: A symbol is mixed with an input image. Preserve original domain statistics while embedding new subsidiary task-related properties to a **high DSM**, and a wide range of sticker tasks to choose a **high TSM**.

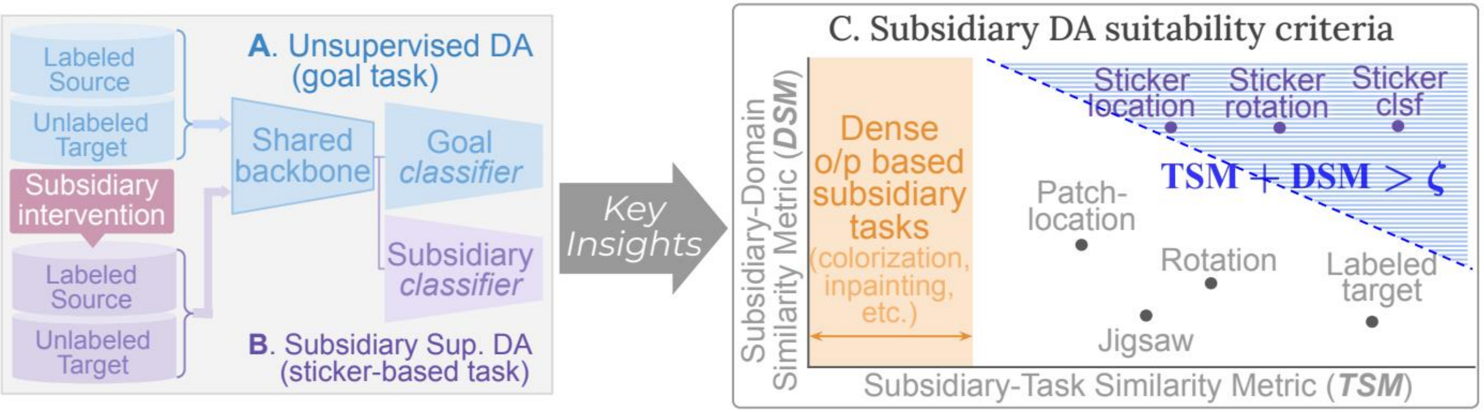


Figure 27. The theoretical insights reveal that subsidiary tasks having both higher TSM (X-axis) and DSM (Y-axis) is most suitable for concurrent goal-subsidiary adaptation (i.e. the shaded blue area)

Figure 28. Sticker intervention.

Other Research

CSS-SFDA [10]

- **Source-side training**: The goal task classifier learns the inductive bias, available only on the source side. **Goal-Subsidiary-Source** features are aligned due to high DSM.

- **Target-side training**: The **frozen goal classifier** preserves source-side goal task inductive bias. Labeled target-subsidiary task → implicit **source-target alignment due to high TSM**.

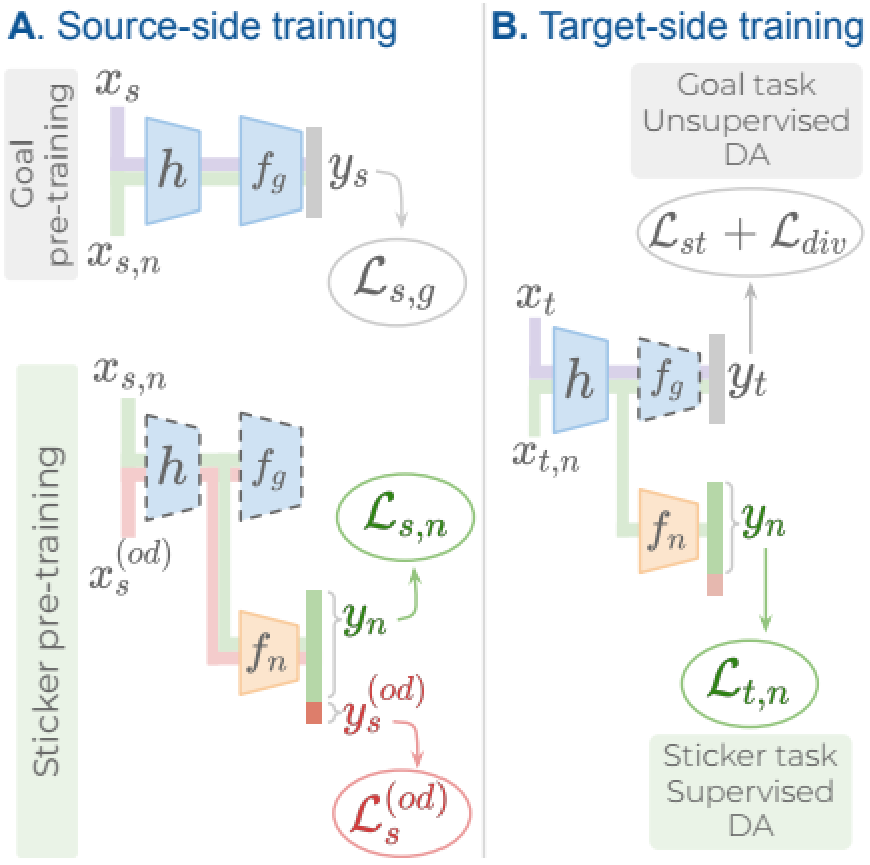


Figure 29. A. Source-side training involves goal pre-training and sticker pre-training .B. Target-side training involves concurrent goal-task unsupervised DA and sticker-task supervised DA.

Datasets

- Office-31
- Office-Home
- VisDA

Datasets

Office-31

- The Office dataset contains 31 object categories in three domains: Amazon, DSLR, and Webcam.
- The 31 categories in the dataset consist of objects commonly encountered in office settings.
- The Amazon domain contains on average 90 images per class and 2817 images in total.
- The DSLR domain contains 498 low-noise high-resolution images (4288×2848). There are 5 objects per category.
- For Webcam, the 795 images of low resolution (640×480) exhibit significant noise and color as well as white balance artifacts.

Figure 30. Examples of Office-31.

Datasets

Office-Home

- Office-Home is a benchmark dataset for domain adaptation which contains 4 domains where each domain consists of 65 categories.
- Art – artistic images in the form of sketches, paintings, ornamentation, etc.
- Clipart – a collection of clipart images.
- Product – images of objects without a background.
- Real-World – images of objects captured with a regular camera.
- It contains 15,500 images, with an average of around 70 images per class and a maximum of 99 images in a class.

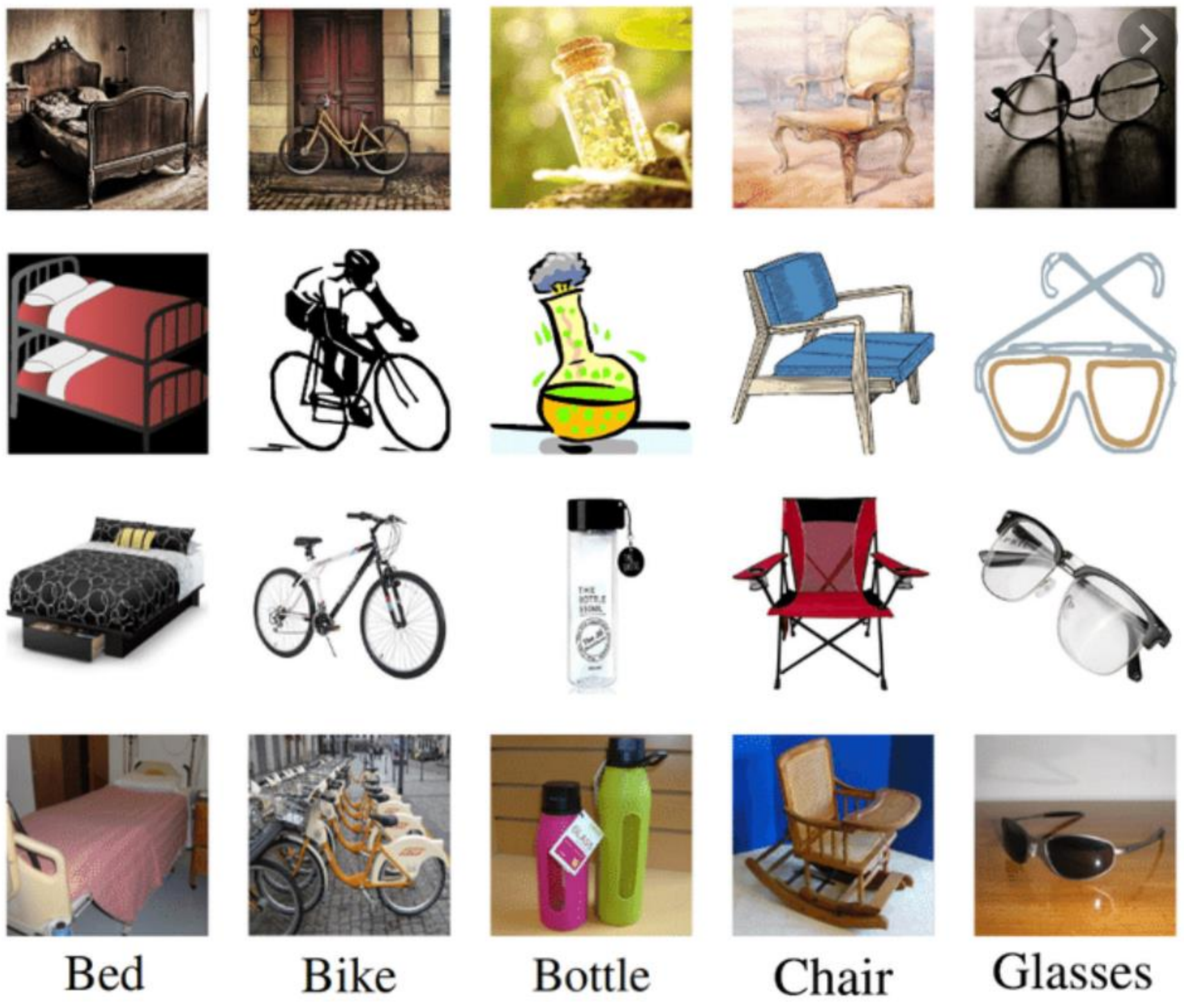


Figure 30. Examples of Office-Home.

Datasets

VisDA-2017

- VisDA-2017 is a simulation-to-real dataset for domain adaptation with over 280,000 images across 12 categories in the training, validation, and testing domains.
- The training images are generated from the same object under different circumstances, while the validation images are collected from MSCOCO.

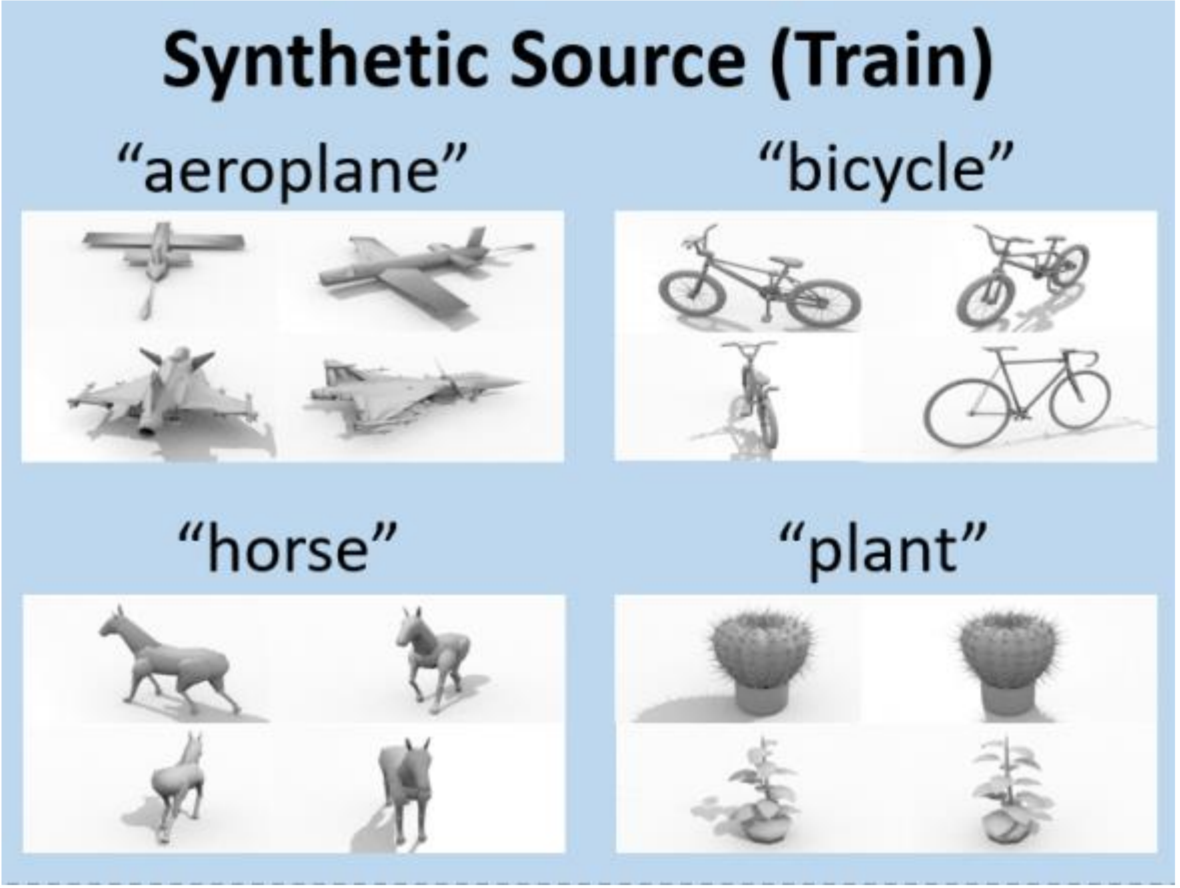


Figure 30. Examples of VisDA-2017.

Applications

- Image Classification
- Semantic Segmentation
- Object Detection

Semantic Segmentation

- Semantic segmentation is a **pixel-level prediction task** that aims to assign a semantic category label to each pixel of an image. Thus, the image-level pseudo-label clustering strategy is not suitable for this task. Based on this, **the Virtual Source Knowledge Transfer** methods can be easily extended to segmentation model adaptation.
- Most methods follow a self-training paradigm based on **pseudo-label filtering** and **information maximization**.
- SRDA [22] is the pioneering study of SFDA for segmentation, but it utilizes the pre-stored meta-data of the source domain, which cannot satisfy the privacy demand.
- TENT [23] innovates to simply fine-tune the normalization and transformation parameters in **BN layers** to adapt the target model, needing less computation but still improving more.
- SF-OCDA [24] presents the **cross-patch style swap** and photometric transformation to simulate the **real-world style variation**, which could promote the model performance in semantic segmentation.

Object Detection

- The domain adaptive object detection task is supposed to align both **the image and object levels**. And the prediction maps of a detection model are also **dividable**, similar to segmentation.

- **SFOD** [18] firstly attempts to address SFDA in OD via modeling it into a problem of learning with noisy labels(Self-supervised pseudo-label filtering).

- **SF-UDA** 3D[19] is the first SFDA framework to adapt the **PointRCNN** [20] 3D detector to target domains.

- In addition, **GCMT** [21] handles the SFDA issue on person re-identification (**ReiD**) with a graph consistency mean-teaching algorithm.

Research Directions

- Efficient Data Reconstruction
- Transformer or GNN

Efficient Data Reconstruction

- How to take advantage of an unlabeled target set for source impression or style translation will **draw more attention**.

- How to **synthesize key samples** according to the learning feedback is an interesting research direction.

- Some advanced learning paradigms are supposed to be explored for valuable data generation, such as **meta-learning**, **contrastive learning**, etc.

Transformer or GNN

- Most of the existing SFDA methods are designed for ResNet or VGG. So how to develop specific algorithms for new architectures remains to be addressed, i.e., Transformer (especially **ViT**) and GNN.

- TransDA [16] is the first framework to apply the Transformer as the **attention module** and inject it into a convolutional network. The model injects a **Tran-sformer** module to obtain a representation with improved generali-zation capability. But in essence, it is a method of Pseudo-label Clustering.

- *Mao et al.* [17] propose a novel scenario named Source Free Unsupervised Graph Domain Adaptation(**SFUGDA**) and an algorithm named **SOGA**.

Reference

- [1] Liang, Jian, Dapeng Hu, and Jiashi Feng. "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation." International Conference on Machine Learning. PMLR, 2020.
- [2] Liang, Jian, et al. "Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).
- [3] Yang, Shiqi, et al. "Generalized source-free domain adaptation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [4] Yang, Shiqi, et al. "Exploiting the intrinsic neighborhood structure for source-free domain adaptation." Advances in Neural Information Processing Systems 34 (2021): 29393-29405.
- [5] Qu, Sanqing, et al. "BMD: A General Class-balanced Multicentric Dynamic Prototype Strategy for Source-free Domain Adaptation." arXiv preprint arXiv:2204.02811 (2022).
- [6] Yang, Shiqi, et al. "Casting a BAIT for offline and online source-free domain adaptation." arXiv preprint arXiv:2010.12427 (2020).
- [7] Dong, Jiahua, et al. "Confident anchor-induced multi-source free domain adaptation." Advances in Neural Information Processing Systems 34 (2021): 2848-2860.
- [8] Xia, Haifeng, Handong Zhao, and Zhengming Ding. "Adaptive adversarial network for source-free domain adaptation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [9] Agarwal, Peshal, et al. "Unsupervised Robust Domain Adaptation without Source Data." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.
- [10] Kundu, Jogendra Nath, et al. "Concurrent Subsidiary Supervision for Unsupervised Source-Free Domain Adaptation." arXiv preprint arXiv:2207.13247 (2022).

Reference

- [11] Li, Rui, et al. "Model adaptation: Unsupervised domain adaptation without source data." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [12] Kurmi, Vinod K., Venkatesh K. Subramanian, and Vinay P. Namboodiri. "Domain impression: A source data free domain adaptation method." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021.
- [13] Hou, Yunzhong, and Liang Zheng. "Source free domain adaptation with image translation." arXiv preprint arXiv:2008.07514 (2020).
- [14] Li, Jingjing, et al. "Divergence-agnostic unsupervised domain adaptation by adversarial attacks." IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).
- [15] Tian, Jiayi, et al. "Vdm-da: Virtual domain modeling for source data-free domain adaptation." IEEE Transactions on Circuits and Systems for Video Technology (2021).
- [16] Yang, Guanglei, et al. "Transformer-based source-free domain adaptation." arXiv preprint arXiv:2105.14138 (2021).
- [17] Mao, Haitao, et al. "Source Free Unsupervised Graph Domain Adaptation." arXiv preprint arXiv:2112.00955 (2021).
- [18] X. Li, W. Chen, D. Xie, S. Yang, P. Yuan, S. Pu, and Y. Zhuang, “A free lunch for unsupervised domain adaptive object detection without source data,” in AAAI Conference on Artificial Intelligence, vol. 35, no. 10, 2021, pp. 8474–8481.
- [19] C. Saltori, S. Lathuiliere, N. Sebe, E. Ricci, and F. Galasso, “SF-UDA ` 3D : Source-free unsupervised domain adaptation for lidar-based 3d object detection,” in IEEE International Conference on 3D Vision. IEEE, 2020, pp. 771–780.
- [20] S. Shi, X. Wang, and H. Li, “PointRCNN: 3d object proposal generation and detection from point cloud,” in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 770–779.

Reference

- [21] X. Liu and S. Zhang, “Graph consistency based mean-teaching for unsupervised domain adaptive person re-identification,” in International Joint Conference on Artificial Intelligence, 2021, pp. 874–880.
- [22] M. Bateson, H. Kervadec, J. Dolz, H. Lombaert, and I. B. Ayed, “Source- relaxed domain adaptation for image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 490–499.
- [23]