LEARNING INVARIANT REPRESENTATION WITH CONSISTENCY AND DIVERSITY FOR SEMI-SUPERVISED SOURCE HYPOTHESIS TRANSFER

Xiaodong Wang\textsuperscript{1} \quad Junbao Zhuo\textsuperscript{2} \quad Shuhao Cui\textsuperscript{3} \quad Shuhui Wang\textsuperscript{4} \quad Yuejian Fang\textsuperscript{1*}

\textsuperscript{1}School of Software and Microelectronics, Peking University \quad \textsuperscript{2}University of Science and Technology Beijing \quad \textsuperscript{3}Kuaishou Technology \quad \textsuperscript{4}Institute of Computing Technology, Chinese Academy of Sciences

ABSTRACT

Semi-supervised Domain adaptation (SSDA) has shown promising results by leveraging unlabeled data and limited labeled samples in the target domain. However, accessibility to source data is hindered by data privacy concerns, giving rise to Semi-supervised Source Hypothesis Transfer (SSHT). Integrating the SSDA methods directly into SSHT tasks is straightforward but poses two significant challenges: i) The hypothesis (classifier) is no longer supervised by source labels, and relying on only a few labels may result in hypothesis collapse; ii) Trained source models often exhibit bias, making them susceptible to misclassifying samples from minority categories into majority ones. We examined the recent methods in the SSHT setting and observed variations in performance compared to SSDA. To address these challenges, we first mitigate model overfitting to target labeled data by promoting prediction consistency between two types of randomly augmented unlabeled data, thereby preventing training collapse. Additionally, we maintain both the prediction diversity and discriminability by leveraging unlabeled data. Experiments on SSHT tasks show that our method yields more stable and competitive results compared with state-of-the-art methods.

Index Terms— Transfer learning, domain adaptation.

1. INTRODUCTION

Deep learning methods have achieved remarkable breakthroughs with appreciable performance in a wide range of applications. However, when training and test data adhere to distinct distributions, a well-trained model faces challenges in generalizing effectively on test data. In response to this challenge, researchers turn to solutions such as Unsupervised Domain Adaptation (UDA) [1, 2, 3, 4] and Semi-Supervised Domain Adaptation (SSDA) [5, 6].

With growing concerns about data privacy, to avoid direct access to source data, the evolution of source-free DA has gained traction [7, 8]. Research indicated that a few labeled target samples can significantly improve the adaptation [9]. Intuitive thought is to devise an effective solution for source-free DA leveraging limited labeled target samples. When adapting pre-trained models to a personalized domain, some annotations are crucial for ensuring acceptable performance. Therefore, given the availability of various large pre-trained models, Semi-supervised Source Hypothesis Transfer (SSHT) is a key technique to leverage pre-trained models.

This study introduces a more challenging SSHT task in comparison to SSDA. In SSDA, despite discrepancies between the source and target domains, accurate source labels facilitate the preservation of discriminability in adapted models. Conversely, in SSHT, the limited target labels may increase the risk of misclassification due to no supervision from the source domain. Besides, the inherent imbalance in source data often results in the trained model classifying minority-category samples into majority ones, thereby diminishing prediction diversity. Biased source models may exhibit limited improvement when adapted to target domains with only a few labeled samples, making this task more challenging. For example, we found that some SSDA methods [5, 10] may encounter training collapse on SSHT tasks and show worse performances than directly using target labels.

To address these challenges, we propose Consistency and Diversity Learning (CDL), a simple but effective framework tailored for SSHT. Given random data augmentations on unlabeled images, we define a consistency regularization inspired by a semi-supervised learning method [11]. By applying strong data augmentation such as in [12, 13], we can obtain a wide range of highly perturbed images. Subsequently, we treat predictions from weakly augmented images as pseudo labels and achieve consistency by training the model to categorize strongly augmented images accordingly. This consistency regularization mitigates the risk of the model overfitting to a limited set of landmarks, thereby enhancing the generalization ability of the transferred model. But imposing such a consistency regularization degrades prediction diversity as it only considers samples with highly confident predictions by...
using a thresholding mechanism. To this end, we integrate Batch Nuclear-norm Maximization [14] into our framework. Specifically, given a classification output matrix of randomly selected batch data, the prediction discriminability and diversity are separately measured by the Frobenius-norm and the rank, which are both bounded by the Nuclear-norm. So, encouraging Batch Nuclear-norm Maximization improves both prediction discriminability and diversity. Therefore, the proposed CDL not only leverages the unlabeled data but also effectively promotes discriminability and diversity of models, alleviating the problem caused by missing source labels.

To delve into the challenges posed by the novel SSHT task, we implemented recent methods derived from the SSDA task and performed extensive experiments on various benchmarks, including Office-Home, Office-31, DomainNet, and VisDA-C. This investigation reveals that prevalent SSDA methods are not always suitable for SSHT tasks, due to the lack of source data for supervision. Our approach showcases significant potential and demonstrates robust performance, yielding more stable and competitive results.

2. METHOD

2.1. Semi-supervised Source Hypothesis Transfer (SSHT)

Common notations and definitions of the SSHT are introduced here. Suppose that there are labeled data $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ in source domain. And we have unlabeled data $D_u = \{u_i\}_{i=1}^{N_u}$ and a small set of labeled data $D_s = \{(x_i, y_i)\}_{i=1}^{N_s}$ in target domain. $N_u$ is usually much larger than $N_s$, since the labeled data is more difficult to obtain and is limited. The source data in $D_s$ is unavailable in SSHT, but we can obtain the model trained on source data.

The goal of SSHT is to adapt a source model to the target domain with only a few labeled target samples and unlabeled target samples. To address the issues mentioned before, we propose a framework named CDL that consists of consistency learning (CL) and diversity learning (DL) modules. The overall framework is shown in Fig 1. Firstly, the unlabeled images are augmented with both weak and strong augmentations. We feed the augmented data into the network and adopt the prediction results of weakly augmented images as pseudo-labels to supervise the strongly augmented ones for achieving prediction consistency. We maintain the prediction diversity by batch nuclear-norm maximization on outputs of all unlabeled augmented images. The source model is adapted in an end-to-end manner, and the collaboration between consistency learning and diversity learning enforces the decision boundary move away from labeled target samples towards unlabeled samples, improving the generalization ability.

2.2. Consistency Learning

Adapting models to a new domain poses a significant challenge due to the lack of source data, making it difficult to accurately assess the distribution discrepancy between the source and target domains. Adapting models without labeled target samples poses a challenging issue as the models maintain decision boundaries derived from source information. While additional target labels offer discriminative information, there is a risk of overfitting the labeled data, leading to unreliable decision boundaries.

To address the overfitting problem, some methods [12, 11] have been proposed based on data augmentations in a semi-supervised learning manner. Typical consistency regularization-based methods adopt the following loss:

$$
p(y \mid \text{Augment}_1(u); \theta) \approx p(y \mid \text{Augment}_2(u); \theta)
$$

where $u$ is an unlabeled image. The Augment$_1$ and Augment$_2$ are random augmentations. $\theta$ denotes the model parameters.

Besides, self-training with pseudo-labeling is also a widely used technique for semi-supervised learning. FixMatch [11] is a combination of the two approaches to SSL: consistency regularization and pseudo-labeling. FixMatch utilizes weak and strong augmentations when performing consistency regularization. Specifically, for each unlabeled sample $u \in D_u$ in the target domain, the weak augmentation $A^w$ and strong augmentation $A^s$ are defined as:

$$
u_w = A^w(u)
$$
$$
u_s = A^s(u).
$$

The weak data augmentation $A^w$ includes image flipping and image translation. The strong data augmentation $A^s$ utilizes the technique proposed in [12]. The consistency regularization incorporating with pseudo-labeling is implemented as treating the prediction of weakly augmented images as pseudo-label and enforcing the prediction of strong augmented ones towards the pseudo-label. However, pseudo labels may contain wrong labels, resulting in error accumulation. Therefore, to mitigate the impact of incorrect pseudo labels, only samples with highly confident predictions are selected for consistency regularization. The consistency regularization loss on unlabeled images is defined as:

$$
\mathcal{L}_u = \frac{1}{u \sim D_u} \mathbb{1}\left(\max(p(\nu_w)) > \tau\right)H(\hat{y}(\nu_w), p(\nu_s))
$$

(1)

where the $\tau$ is the threshold, and $\hat{y}(\nu_w)$ is the one-hot vector of $\arg\max(p(\nu_w))$. $H(p, q)$ denotes the cross-entropy between two distributions $p$ and $q$. By optimizing the consistency loss, the decision boundary will be pushed far from the labeled samples, better for classifying unlabeled samples.

To ensure the discriminability of models, we adopt the typical cross-entropy loss for the labeled target data $D^t$. The classification loss $\mathcal{L}_c$ is defined as:

$$
\mathcal{L}_c = \frac{1}{(x,y) \sim D^t} H(p(x), y).
$$

The loss minimized by Consistency Learning is simply $\mathcal{L}_c + \lambda_u \mathcal{L}_u$, where $\lambda_u$ is a fixed scalar hyper-parameter denoting the relative weight of the unlabeled loss.
Fig. 1: The proposed CDL framework. Firstly, the unlabeled images are augmented with both weak and strong augmentations and fed to the model. The predictions of weakly augmented images are used as supervision signals for strongly augmented ones to encourage consistent prediction. We further encourage prediction diversity on outputs of all unlabeled augmented images.

2.3. Diversity Learning

While Consistency Learning enhances the classification of unlabeled data, it is prone to diminishing prediction diversity. Consequently, we have incorporated an effective technique to uphold both discriminability and diversity in predictions. We adopt Batch Nuclear-norm Maximization (BNM) [14]. Diversity could be measured by the number of response categories, which is the rank of the prediction matrix. Since the nuclear norm is the convex approximation of the matrix rank, maximizing the Batch Nuclear-norm enlarges the rank, increasing the prediction diversity. BNM is adopted on the matrix of the classification responses for a batch of unlabeled samples.

The loss function of BNM is defined as below:

\[ L_{bnm} = -\frac{1}{B} \| G(X) \|_* \]  

where the \( G(X) \) is the output matrix with respect to the input \( X \), and \( B \) is the batch size of random samples. \( \| \cdot \|_* \) denotes the Nuclear-norm, which is the sum of all the singular values in the matrix. There are two types of augmented images, \( A^w \) and \( A^s \). The loss for diversity learning is defined as below:

\[ L_d = \frac{1}{B} \mathbb{E}_{u_1, \ldots, u_B \sim \mathcal{D}^w} (\| G(A^w([u_1, \ldots, u_B])) \|_*) + \| G(A^s([u_1, \ldots, u_B])) \|_* \]  

where \([\cdot]\) denotes the concatenation. Minimizing the loss can enforce the model to push decision boundaries into low-density regions without losing diversity. BNM sacrifices a portion of the prediction hit rate on majority categories to bolster the prediction hit rate on minority categories, thereby preserving diversity. Nonetheless, the supervision from labeled target data and Consistency Learning effectively compensate for the potential loss in discriminability, establishing a cohesive collaboration with the BNM loss.

| Method  | DomainNet | VisDA-C |
|---------|------------|---------|
| S+T [15] | 60.0       | 52.4    |
| CDAN [16] | 66.5       | 73.9    |
| AFN [17] | -          | 76.1    |
| MME [5] | 68.9       | 79.1    |
| ATDOC [18] | 72.4       | -       |
| CDL     | 73.1       | 88.5    |

2.4. Joint Training

The objective of the proposed CDL is defined as below:

\[ \mathcal{L} = \mathcal{L}_c + \lambda_u \mathcal{L}_u + \lambda_d \mathcal{L}_d, \]  

where \( \lambda_u \) and \( \lambda_d \) control the trade-off between classification loss, consistency loss, and diversity loss. The classification loss \( \mathcal{L}_c \) furnishes precise supervision during training. The consistency regularization loss \( \mathcal{L}_u \) mitigates overfitting on limited labeled target data. The diversity loss \( \mathcal{L}_d \) serves to sustain both discriminability and diversity. Through joint training, well-trained models from source domains are effectively adapted to target domains in SSHT tasks.

3. EXPERIMENT

We conduct extensive experiments on several domain adaptation benchmarks including Office-Home [21], Office-31 [1], DomainNet [22] and VisDA-C [23]. For different tasks with the same source domain, we train a unique source model with the same source data. All SSHT tasks are in the 3-shot setting.

Table 1 presents the average comparison of the challenging benchmark DomainNet and the synthetic-to-real benchmark VisDA-C. CDL exhibits promising results on SSHT.
Table 2: SSHT tasks on Office-Home dataset (%) (ResNet-34).

| Method      | A → C | A → P | A → R | C → A | C → P | C → R | P → A | P → C | P → R | R → A | R → C | R → P | MEAN |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| S [15]      | 43.1  | 61.1  | 70.4  | 48.5  | 59.0  | 62.9  | 48.2  | 41.5  | 70.7  | 63.0  | 46.1  | 76.4  | 57.6  |
| MME [5]     | 51.6  | 66.4  | 63.9  | 49.1  | 67.9  | 60.6  | 48.0  | 51.4  | 64.4  | 54.8  | 54.1  | 72.1  | 58.7  |
| S+T [15]    | 54.6  | 73.5  | 74.1  | 59.9  | 73.1  | 71.1  | 59.0  | 54.9  | 73.7  | 64.8  | 58.0  | 79.3  | 66.3  |
| MCC [10]    | 54.5  | 73.6  | 74.1  | 59.8  | 73.0  | 71.2  | 59.0  | 54.9  | 73.6  | 64.8  | 58.0  | 79.3  | 66.3  |
| ENT [19]    | 53.8  | 75.8  | 76.2  | 60.0  | 75.8  | 73.0  | 62.4  | 53.9  | 77.8  | 69.5  | 58.5  | 83.7  | 68.4  |
| FixMatch [11]| 55.9 | 76.9  | 74.0  | 63.2  | 76.4  | 73.2  | 62.7  | 55.2  | 75.1  | 67.3  | 57.3  | 84.4  | 68.5  |
| BNM [14]    | 60.7  | 79.2  | 80.7  | 65.8  | 80.1  | 78.8  | 65.8  | 62.0  | 81.0  | 71.2  | 64.9  | 85.5  | 73.0  |

CDL (Ours) 61.8 79.8 80.4 67.5 81.1 79.2 62.8 82.0 73.2 66.3 86.3 74.1

Table 3: Zero-shot SSHT tasks on Office-Home dataset (%) (ResNet-50).

| Method      | A → C | A → P | A → R | C → A | C → P | C → R | P → A | P → C | P → R | R → A | R → C | R → P | MEAN |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| S [15]      | 44.6  | 67.3  | 74.8  | 52.7  | 62.7  | 64.8  | 53.0  | 40.6  | 73.2  | 65.3  | 45.4  | 78.0  | 60.2  |
| CDAN [16]   | 50.7  | 70.6  | 76.0  | 57.6  | 70.0  | 70.0  | 57.4  | 50.9  | 77.3  | 70.9  | 56.7  | 81.6  | 65.8  |
| SAFN [17]   | 52.0  | 71.7  | 76.3  | 64.2  | 69.9  | 71.9  | 63.7  | 51.4  | 77.1  | 70.9  | 57.1  | 81.5  | 67.3  |
| MDD [20]    | 54.9  | 73.7  | 77.8  | 60.0  | 71.4  | 71.8  | 61.2  | 53.6  | 78.1  | 72.5  | 60.2  | 82.3  | 68.1  |
| SHOT [7]    | 57.3  | 78.5  | 81.4  | 67.9  | 78.5  | 78.0  | 68.1  | 56.1  | 78.1  | 72.5  | 60.2  | 82.3  | 68.1  |
| ATDOC [18]  | 58.3  | 78.8  | 82.3  | 69.4  | 78.2  | 78.2  | 67.1  | 56.0  | 82.7  | 72.0  | 58.2  | 85.5  | 72.2  |
| SHOT++ [8]  | 58.1  | 79.5  | 82.4  | 68.6  | 79.9  | 79.3  | 68.6  | 57.2  | 83.0  | 74.3  | 60.4  | 85.1  | 73.0  |
| CDL (Ours)  | 61.2  | 83.8  | 81.9  | 73.3  | 85.0  | 81.3  | 71.0  | 61.9  | 83.2  | 76.2  | 63.0  | 87.0  | 75.7  |

Table 4: SSHT tasks on Office-31 dataset (%) (VGG-16).

| Method      | A → D | A → D | D → A | D → W | MEAN |
|-------------|-------|-------|-------|-------|-------|
| S [15]      | 74.3  | 70.1  | 62.9  | 63.8  | 67.8  |
| S+T [15]    | 91.4  | 86.8  | 70.0  | 70.2  | 79.6  |
| MCC [10]    | 91.4  | 86.8  | 70.0  | 70.3  | 79.6  |
| MME [5]     | 89.9  | 88.3  | 70.4  | 70.7  | 79.8  |
| ENT [19]    | 92.8  | 85.6  | 74.2  | 71.7  | 81.1  |
| FixMatch [11]| 92.4 | 86.2  | 74.7  | 72.8  | 81.5  |
| BNM [14]    | 92.4  | 89.9  | 77.6  | 77.4  | 84.3  |
| CDL (Ours)  | 92.6  | 87.3  | 78.4  | 78.4  | 84.2  |

4. CONCLUSION

In this paper, we focus on a challenging task Semi-supervised Source Hypothesis Transfer (SSHT). The insufficient labeled target data may increase the risk of misclassification in the target domain and reduce the prediction diversity. To tackle these issues, we present a Consistency and Diversity Learning (CDL) framework for SSHT, by encouraging consistency regularization and enhancing the diversity of predictions. Experimental results show that our method achieves comparable results compared with existing state-of-the-art methods.

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant No. 2022YFB2703301.

5. REFERENCES

[1] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, “Adapting visual category models to new domains,” in European conference on computer vision. Springer, 2010, pp. 213–226.
[2] Junbao Zhuo, Shuhui Wang, Shuhao Cui, and Qingming Huang, “Unsupervised open domain recognition by se-
mantic discrepancy minimization,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 750–759.

[3] Junbao Zhuo, Shuhui Wang, and Qingming Huang, “Uncertainty modeling for robust domain adaptation under noisy environments,” IEEE Transactions on Multimedia, 2022.

[4] Junbao Zhuo, Xingyu Zhao, Shuhao Cui, Qingming Huang, and Shuhui Wang, “Adaptive feature swapping for unsupervised domain adaptation,” in Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 7017–7028.

[5] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko, “Semi-supervised domain adaptation via minimax entropy,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8050–8058.

[6] Ning Ma, Jiajun Bu, Lixian Lu, Jun Wen, Sheng Zhou, Zhen Zhang, Jingjun Gu, Haifeng Li, and Xifeng Yan, “Context-guided entropy minimization for semi-supervised domain adaptation,” Neural Networks, vol. 154, pp. 270–282, 2022.

[7] Jian Liang, Dapeng Hu, and Jiashi Feng, “Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation,” in International Conference on Machine Learning. PMLR, 2020, pp. 6028–6039.

[8] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng, “Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer,” arXiv preprint arXiv:2012.07297, 2020.

[9] Ning Ma, Haishuai Wang, Zhen Zhang, Sheng Zhou, Hongyang Chen, and Jiajun Bu, “Source-free semi-supervised domain adaptation via progressive mixup,” Knowledge-Based Systems, vol. 262, pp. 110208, 2023.

[10] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang, “Minimum class confusion for versatile domain adaptation,” in European Conference on Computer Vision. Springer, 2020, pp. 464–480.

[11] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel, “Fixmatch: Simplifying semi-supervised learning with consistency and confidence,” arXiv preprint arXiv:2001.07685, 2020.

[12] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le, “Randaugment: Practical automated data augmentation with a reduced search space,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 702–703.

[13] Xiaodong Wang, Junbao Zhuo, Mengru Zhang, Shuhui Wang, and Yuejian Fang, “Revisiting unsupervised domain adaptation models: A smoothness perspective,” in Proceedings of the Asian Conference on Computer Vision, 2022, pp. 1504–1521.

[14] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Qi Tian, “Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3941–3950.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[16] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan, “Conditional adversarial domain adaptation,” arXiv preprint arXiv:1705.10667, 2017.

[17] Ruijia Xu, Guanbin Li, Jiahan Yang, and Liang Lin, “Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1426–1435.

[18] Jian Liang, Dapeng Hu, and Jiashi Feng, “Domain adaptation with auxiliary target domain-oriented classifier,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[19] Yves Grandvalet, Yoshua Bengio, et al., “Semi-supervised learning by entropy minimization,” in CAP, 2005, pp. 281–296.

[20] Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan, “Bridging theory and algorithm for domain adaptation,” in International Conference on Machine Learning. PMLR, 2019, pp. 7404–7413.

[21] Hemanth Venkateswar, Jose Eusebio, Sharyok Chakraborty, and Sethuraman Panchanathan, “Deep hashing network for unsupervised domain adaptation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5018–5027.

[22] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang, “Moment matching for multi-source domain adaptation,” in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1406–1415.

[23] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko, “Visda: The visual domain adaptation challenge,” arXiv preprint arXiv:1710.06924, 2017.