




# Shared Wasserstein adversarial domain adaption

Shengqing Yao<sup>1</sup> · Yuming Chen<sup>2</sup>  · Yanfang Zhang<sup>1</sup> · Zhizhong Xiao<sup>1</sup> · Jiaojiao Ni<sup>1</sup>

Received: 7 July 2023 / Revised: 10 January 2024 / Accepted: 23 February 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

In numerous real-world applications, obtaining labeled data for a specific deep learning task can be prohibitively expensive. We present an innovative framework for unsupervised training of deep neural networks, drawing inspiration from the adversarial learning paradigm. Our approach incorporates the cycle-consistency constraint to effectively constrain the generator. Furthermore, we capitalize on the reconstructed samples, treating them as "real" samples for the discriminator during classification. This idea stems from the success of Wasserstein GAN, which leverages its gradient property and promising generalization bound during network training. Simultaneously, we employ a shared latent-data space constraint to ensure compatibility between the source domain and its corresponding target domain. This constraint facilitates effective knowledge transfer from the source to the target domain, even in the absence of labeled data for the target domain. To enhance the performance of the target domain classifier, we introduce association chains that link the embeddings of labeled samples to those of unlabeled samples and vice versa. By encouraging correct association cycles that ultimately return to the same class from which the association began, and penalizing wrong associations leading to a different class, we ensure accurate predictions. Our proposed method, named Shared Wasserstein Adversarial Domain Learning (SWADL), combines these novel constraints. Through extensive evaluations on benchmark datasets such as MNIST, SVHN, and USPS, we demonstrate that SWADL consistently outperforms current mainstream methods. It achieves superior results in unsupervised domain adaptation tasks, addressing the challenge of limited labeled data in real-world scenarios. The code and models are available at <https://github.com/Jayee-chen/Adversarial-Domain-Adaptation.git>.

**Keywords** Unsupervised domain adaption · Wasserstein adversarial learning · Shared latent space · Associative chain

---

✉ Yuming Chen  
cym@njtech.edu.cn

<sup>1</sup> China Construction Industrial & Energy Engineering Group Co., Ltd., Nanjing, China

<sup>2</sup> College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, China

# 1 Introduction

Deep learning training typically requires a substantial volume of labeled training data, which can be both time-consuming and costly to annotate manually. Fortunately, the task of collecting unlabeled visual data has become increasingly convenient owing to the prevalence of affordable consumer and surveillance cameras, as well as the abundance of large Internet databases like YouTube. This leads us to the question: How can we leverage these unlabeled databases effectively? Domain adaptation [1] is a technique that facilitates the transfer of knowledge from a labeled source domain to an unlabeled target domain by exploring domain-invariant knowledge structures that capture the similarity between domains, despite significant distribution differences. The primary goal of domain adaptation is to address the challenge posed by having data from two related domains that exhibit distinct distributions, enabling the adaptation of predictive models across domains to mitigate the domain discrepancy. Existing domain adaptation methods typically assume a shared output label space and different feature distributions between the source and target domains [23]. These methods bridge the gap between domains by learning domain-invariant feature representations without relying on target labels. The classifier trained on the source domain can then be directly applied to the target domain. To achieve this, the representation is optimized to minimize a measure of domain shift, such as maximum mean discrepancy. By leveraging these techniques, domain adaptation enables the application of existing knowledge from a labeled source domain to improve performance in the unlabeled target domain. It allows for the transfer of predictive models and learned representations, even in scenarios where the domains exhibit substantial differences in their data distributions. [2, 3] or correlation distances [4, 5]. Alternatively, reconstruct the target model from the representation of the source domain [6]. Recent studies have revealed that deep learning networks can extract more common features for domain adaption [7, 8], by disentangling the feature space of domains. The latest advances have been achieved by embedding domain adaption in the pipeline of deep feature learning to extract deep common representations [9–11]. Adversarial adaptation methods [12, 13] have become an effective approach which minimize an approximate domain discrepancy distance through an adversarial objective with respect to a domain discriminator.

The contributions of this paper can be summarized as:

1. The survey introduces a new training method for unsupervised domain adaptation. The approach focuses on Wasserstein adversarial domain adaptation with a shared latent space. By transferring shared latent knowledge between the source and target domains using Wasserstein GAN [14], the method achieves higher-quality outcomes while maintaining training stability and reconstruction quality.
2. The approach enforces a shared latent space between the source and target domains. This shared latent space is utilized in designing the classifier of the deep network. The innovative aspect is the constraint imposed on the classifier's weights, which are partially shared between the two domains. This constraint ensures that the classifier captures and leverages common features and knowledge across both domains, leading to improved performance and transferability of the model.
3. Effective Results with Simple Networks: The proposed approach demonstrates competitive results even without utilizing sophisticated network architectures like VGG-Net or Inception. This paper highlights the effectiveness of the method in achieving desirable performance using simpler models. It suggests that the approach can be applied in scenarios where computational resources or complex network architectures may not be readily available or practical.

## 2 Related work

Adversarial domain adaptation [12, 22] integrates the advantages of adversarial learning into domain adaptation. One advantage of GAN over other generative methods is the elimination of the need for inference and complex sampling during training. However, a downside is the occurrence of model collapses or divergence issues during the training process [15]. A domain discriminator is learned by minimizing the classification error of distinguishing the source from the target domains, while a deep classification model learns transferable representations that are indistinguishable by the domain discriminator. The BiGAN approach [16] extends GANs to learn the inverse mapping from the image data back into the latent space, and shows that it can extract features useful for image classification tasks. The conditional generative adversarial net (CGAN) [17] is an extension of the GAN where both networks G and D receive an additional vector of information as input. This might contain information about the class of the training example. Taigman et al. [18] can train a conditional generator without paired images, with unsupervised network for cross-domain, but relies on a complicated pre-trained model that maps images from source domain to an intermediate representation. The authors apply CGAN to generate a (possibly multi-modal) distribution of tag-vectors conditional on image features. In CycleGAN, a concurrent work by Zhu et al. [19], the same idea for unpaired image-to-image translation is proposed, where the primal-dual relation in DualGAN is referred to as a cyclic mapping and their cycle consistency loss is essentially the same as our reconstruction loss. CoGAN [20] approach applies GANs to the domain adaptation problem by adversarial training to generate the source and target images respectively. The approach achieves a domain invariant feature presentation by tying the high-level layer weights of the two GANs, and shows that the same noise distribution input can generate a corresponding pair of images. Domain adaptation is performed by training a classifier on the discriminator output and applies to shifts between the MNIST and USPS digit datasets. However, this approach relies on the generators which can find a mapping from the shared high level feature space to full images set in both domains.

## 3 Theory

### 3.1 Problem definition

In the problem of unsupervised domain adaptation, we have a source domain  $D_s = \{x_i^s, y_i^s\}$ ,  $i = 1, \dots, n$  with  $n$  labeled examples, and a target domain  $D_t = \{x_i^t\}$ ,  $i = 1, \dots, n_0$  with  $n_0$  unlabeled examples. It is assumed that the source domain and target domain share the common feature space, but have different probability distributions, respectively. Despite the lack of target domain annotations, our final goal is to obtain a latent target representation  $Z$  and classifier  $C_t$  that can precisely divide target data into one of categories  $K$ . Because of no labeled images in the target domain, so we cannot directly train classifier for the categories in the target domain  $T$ . Instead, we will take advantage of data from a related, but distinct source domain  $S$ , where full labeled images are available from the corresponding categories  $K$ . There is the difference distribution between source and target domain, we directly train a classifier using only the source data reduced performance at test time when classifying in the target domain. Our assumption is that if we can learn a representation that minimizes the metric distance between the source and target distributions, then we can train a classifier on the source labeled data and directly apply it to the target domain with minimal loss in

accuracy. Meanwhile, the similarity value of domain adaptation has even more increased with generative tools producing synthetic datasets. Adversarial learning, the key idea to enabling Generative Adversarial Networks (GANs) [18], has successfully generated the image of the target domain to minimize the cross-domain discrepancy [13]. With the proposed method for domain adaptation it becomes possible to train models without the labeled target example at training time.

### 3.2 Framework

Our framework, as illustrated in Fig. 1, is based on Variational AutoEncoders (VAEs) [13, 14] and generative adversarial network (GAN) [6, 17]. It consists of 6 sub-networks: including domain encoders  $E_s$  and  $E_t$ , image generators  $G_s$  and  $G_t$ , classifiers  $L_s$  and  $L_t$ , and domain adversarial critics  $C_s$  and  $C_t$ . Several ways exist to interpret the roles of the sub-networks [21]. Our framework learns translation in both directions in one shot. We assume that there has a shared-latent space between source domain and target domain. We enforce a weight-sharing constraint to relate the two domain VAEs. Specifically, we share the weights of the last few layers of  $E_s$  and  $E_t$  that are responsible for extracting high-level representations of the input images in the two domains. Similarly, we share the weights of the first few layers of  $G_s$  and  $G_t$ , which are responsible for decoding high-level representations for reconstructing the input images. The shared auto-encoder, along with the domain specific encoders/decoders, can provide more functional utilizations like domain linear combination or incrementally learning a new domain. On the same way, we share the weights of the first few layers of  $C_s, C_t$  and  $L_s, L_t$  that are responsible for sharing high-level representations. Our network framework is shown in Fig. 1. We propose to train the encoder  $E_s, E_t$ , the generator  $G_s, G_t$  and the critic  $C_s, C_t$  following the order:

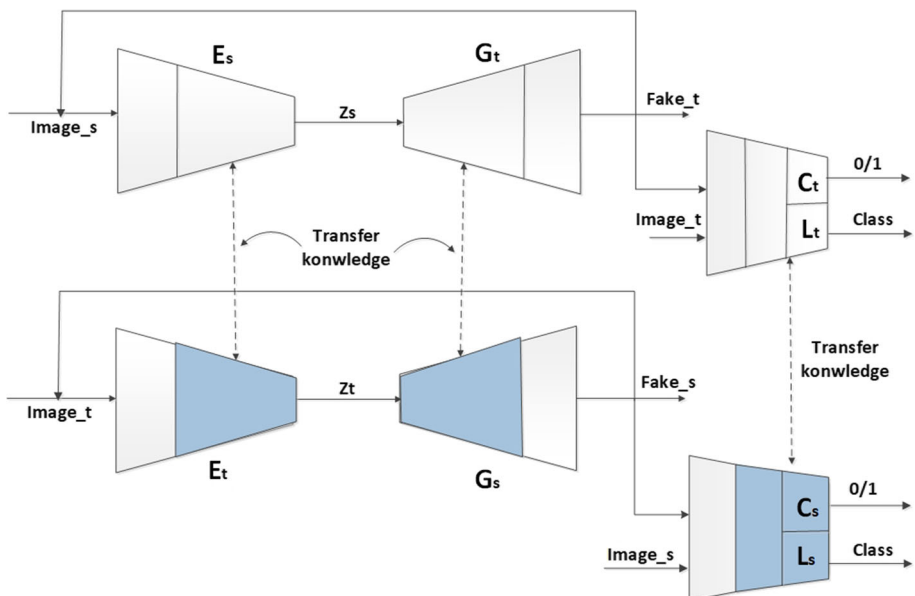


Fig. 1 Architecture for wasserstein adversarial domain learning with the shared-latent space

- (i) Train  $G_s, G_t$  and  $E_s, E_t$  to minimize the reconstruction loss (2)(3);
- (ii) Fix  $G_s, G_t$ , and train  $C_s, C_t$  to minimize (1);
- (iii) Fix  $C_s, C_t$ , and train  $G_s, G_t$  to minimize (1).

### 3.2.1 Adversarial loss

We propose a new approach to learn feature representations invariant to the change of domains by minimizing empirical Wasserstein distance between the source and target representations through adversarial training. We can use an adversarial method to optimize the exact Wasserstein distance that does not require any hyper-parameters to enforce weight constraints. Let  $C_{i=s,t}$  be the discriminator and  $G_{i=s,t}$  be the generator. Based on this new formulation and Gradient Penalty (GP) optimization method, we propose the WGAN-GP model as follows:

$$L_{adv} = \lambda_1 * \min_{G_i} \max_{C_j} \left\{ \frac{1}{m} \sum_{i \in I} C_i(y_i) - \frac{1}{n} \sum_{j \in J} C_j(G_j(z_i)) \right\} \tag{1}$$

s.t.  $D_i(y_i) - D_j(G_j(Z_i)) \leq \|y_i - G_j(z_i)\|_1$ . The  $\lambda_1$  controls the weights of the adversarial Loss function term. When the generator  $G_{i=s,t}$  is fixed, we let  $x_j = G(Z_j), j = \bar{i}$  and we apply the proposed GP method to optimize formula (1) to compute the critic  $C_{i=s,t}$ . After we optimize the critics  $C_{i=s,t}$ , we fix them and update the generators  $G_{i=s,t}$ . We compute the generator loss as follows:  $\min \frac{1}{n} \sum_{j=j} C_j(G_j(z_j))$ .

### 3.2.2 Cycle-reconstruction loss

We utilize a VAE-like objective function to model the cycle-consistency constraint, which is given by

$$L_{cycles}(E_s, G_s, E_t, G_t) = \lambda_2 KL(q_s(z_s | x_s) \parallel p_\eta(z)) + \lambda_2 KL(q_t(z_t | F_{t \rightarrow s}(x_s)) \parallel p_\eta(z)) - \lambda_3 E_{z_t, q_t(z_t | F_{s \rightarrow t}(x_s))} [\log p_{G_s}(x_s | z_t)] \tag{2}$$

$$L_{cycet}(E_t, G_t, E_s, G_s) = \lambda_2 KL(q_t(z_t | x_t) \parallel p_\eta(z)) + \lambda_2 KL(q_s(z_s | F_{t \rightarrow s}(x_t)) \parallel p_\eta(z)) - \lambda_3 E_{z_s, q_s(z_s | F_{t \rightarrow s}(x_t))} [\log p_{G_t}(x_t | z_s)] \tag{3}$$

Where the negative log-likelihood objective term ensures a twice translated image resembles the input one and the KL terms penalize the latent codes deviating from the prior distribution in the cycle-reconstruction stream. The hyper-parameters  $\lambda_2$  and  $\lambda_3$  control the weights of the two different objective terms.

### 3.2.3 Classification loss

It is defined by the cross entropy between the uniform distribution over target samples and the probability of visiting some target samples starting in any source samples,

$$L_{visit} = \lambda_4 H(V, P^{visit}) \tag{4}$$

Where  $P_j^{visit} = \sum_{x_i \in D_s} P_{ij}^{ab}, V_j := \frac{1}{|B|}, \lambda_4$  is a weight factor. Note that this formulation assumes that the class distribution is the same for source and target domain. If this is not the case, using a low weight for  $L_{visit}$  may yield better results.

### 3.2.4 Overall objective function

By combining the two objectives in (1)- (4), we obtain the final objective function as follows:

$$L(E_i, G_i, D_i) = L_{adv} + L_{cycs} + L_{cycr} + L_{visit} \quad (5)$$

We aim to solve:

$$E^*, G^* = \arg \min_{C_i} \max_{E_i, G_i} L(E_i, G_i, C_i), i = s, t \quad (6)$$

## 4 Experimental results and evaluation

### 4.1 Datasets

**Digits datasets** three digits datasets are used: MNIST [24], USPS [25] and Street View House Numbers (SVHN) [26]. Each dataset contains ten classes corresponding to number 0 to 9.

**Office-31 dataset** [27] serves as a standardized benchmark in the realm of visual domain adaptation. It encompasses a dataset of 4,652 images, characterized by noise, and spans 31 categories. These images are sourced from three distinct domains: Amazon(A), comprising images downloaded from amazon.com; Webcam(W), consisting of images captured by a web camera; and DSLR(D), containing images taken by a digital SLR camera, each originating from different real-world environments.

### 4.2 Experimental setup

We describe the network architectures and hyper-parameters of different tasks. Our approach is implemented with PyTorch deep learning framework.

**Digits experiments** We used ADAM [11] for training where the learning rate was set to 0.0001 and momentums were set to 0.5 and 0.999. Each mini-batch consisted of one image from the first domain and one image from the second domain. Our framework had several hyper-parameters. The default values were  $\lambda_1 = 1$ ,  $\lambda_2 = 10$ ,  $\lambda_3 = 0.1$ ,  $\lambda_4 = 0.1$ . For the network architecture, our encoders consisted of five convolutional layers with Batch Normalization that were each followed by a LeakyReLU nonlinearity. For The generators consisted of five transposed convolutional layers with Batch Normalization, we used LeakyReLU for threshold function. The discriminators consisted of stacks of convolutional layers and Max-pooling that were each followed by a LeakyReLU nonlinearity.

**Office experiments** For the encoder architecture, the ultimate layer of AlexNet [28] undergoes modification by substituting it with two parallel fully connected layers, generating vectors of dimensions 256 each. The preceding layers are initialized with weights from the model pretrained on ImageNet [29]. Fine-tuning of the encoder is executed with a base learning rate of 0.0001 over a span of 100 epochs, while the base learning rate for the other three submodels is established at 0.001. The inputs for the encoder and discriminator are resized to dimensions  $227 \times 227$  and  $64 \times 64$ , respectively.

### 4.3 Performance on digits datasets

In order to evaluate our method, we chose common domain adaptation tasks, for which previous results were reported. We were motivated by the problem of learning models on the clean, synthetic datasets and testing on the noisy, real-world datasets. To this end, we evaluated on object classification datasets used in previous work including MNIST and the Street-view House Numbers (SVHN). We tested the following unsupervised domain adaptation scenarios: (a) from MNIST to SVHN; (b) from SVHN to MNIST (See Fig. 2).

To evaluate the generalization performance of our method, we experimentally validated our proposed algorithm in an unsupervised adaptation task between the MNIST and USPS digits datasets. Since the images of the USPS dataset had  $16 \times 16$  pixels, we resized its images to  $28 \times 28$  pixels which were same as the image size of the MNIST dataset. Figure 3 showed image translation results obtained using the unlabeled datasets of target domain between MNIST and USPS.

We compared our method with 5 recent UDA methods under the same condition. On these tasks, we could see that for the SWADL, our method achieved a 0.962 accuracy for the MNSIT to USPS task, which was better than 0.9597 achieved by the previous state-of-the-art method [21]. We also achieved better performance for the MNSIT to SVHN task. (See Table 1)

### 4.4 Performance on office-31 dataset

The experimental results on the Office-31 dataset are presented in Table 2. Upon comparing SWADL model with the other five methods, we observe that SWADL does not show superior on individual tasks, but the proposed one achieves the best results. For instance, SWADL achieves an average classification accuracy of 0.865, surpassing the best result among the five domain adaptation methods by 0.7%. In terms of task-specific transfer classification accuracy, SWADL outperforms the competition, securing the top position in 10 out of 12 specific transfer tasks. In summary, the following observations can be delineated: (1) Adversarial-based approaches demonstrate greater effectiveness in comparison to metric-based methodologies. (2) Deep domain adaptation techniques exhibit superior efficacy compared to traditional learning methods. (3) The proposed metric, relying on mean and covariance, proves to be effective, enhances the overall efficacy of the classification process.

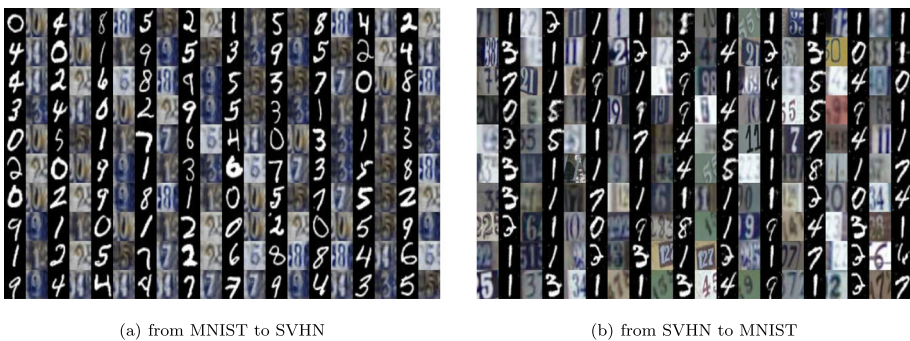
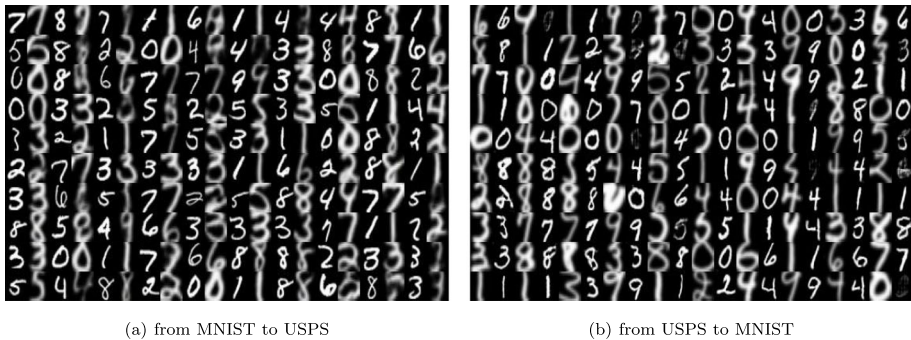


Fig. 2 Experimental results for unsupervised domain adaptation. SVHN  $\leftrightarrow$  MNIST



**Fig. 3** Experimental results for unsupervised domain adaption. USPS  $\leftrightarrow$  MNIST

#### 4.5 Ablation study

**Influence of parameters** Our proposed method encompasses four parameters. Firstly,  $\lambda_1$  balances the trade-off between transferability and discriminability of the learned feature representation. Secondly,  $\lambda_2$  governs the balance between domain discriminability and target discriminability. Thirdly,  $\lambda_3$  regulates the trade-off between mean alignment and covariance alignment. Lastly,  $\lambda_4$  determines the importance of the regularization term. To gain insights into the impacts of the four parameters, we conducted experiments on five randomly selected transfer tasks from the SVHN $\leftrightarrow$  MNIST dataset (SVHN $\rightarrow$  MNIST, MNIST $\rightarrow$  SVHN, MNIST $\rightarrow$  USPS, USPS $\rightarrow$  MNIST). In investigating the influence of  $\lambda_1$ , while keeping  $\lambda_2 = 10$ ,  $\lambda_3 = 0.1$ , and  $\lambda_4 = 0.1$  constant, we varied  $\lambda_1$  across the range of  $\{0.85, 0.95, 1., 1.15\}$ . The results, illustrated in Fig. 4(a), reveal that neither biasing towards transferability nor discriminability alone is suitable. Both aspects are crucial for effective feature representation in the classification task. Turning to the exploration of  $\lambda_2$ , with fixed  $\lambda_1 = 1$ ,  $\lambda_3 = 0.1$ , and  $\lambda_4 = 0.1$ , we adjusted  $\lambda_2$  within the range of  $\{4, 6, \dots, 16\}$ . As depicted in Fig. 4(b), an appropriate proportion of target discriminability improves the performance of our model. However, increasing the proportion might lead to a decline in model performance due to the imprecision in improving target discriminability without access to target domain labels. For the investigation of  $\lambda_3$ , with fixed  $\lambda_1 = 1$ ,  $\lambda_2 = 10$ , and  $\lambda_4 = 0.1$ , we varied  $\lambda_3$  within the range of  $\{0.02, 0.04, \dots, 0.3\}$ . The results shown in Fig. 4(c) indicate that both mean alignment and covariance alignment contribute to increased transferability and subsequently enhance our proposed model's performance. Lastly, exploring the impact of  $\lambda_4$  with fixed  $\lambda_1 = 1$ ,  $\lambda_2 = 10$ , and  $\lambda_3 = 0.1$ , we selected  $\lambda_4$  from the range of  $\{0.02, 0.06, \dots, 0.2\}$ . As depicted in Fig. 4(d), an appropriate proportion of the regularization term can enhance the generalization ability of the model.

**Table 1** The reported numbers are classification accuracies

Method	SA	DANN	DTN	CoGAN	UNIT	SWADL(Our)
SVHN $\rightarrow$ MNIST	0.592	0.7385	0.8488		0.9053	0.891
MNSIT $\rightarrow$ SVHN						0.591
MNSIT $\rightarrow$ USPS				0.9565	0.9597	0.962
USPS $\rightarrow$ MNSIT				0.9315	0.9358	0.931

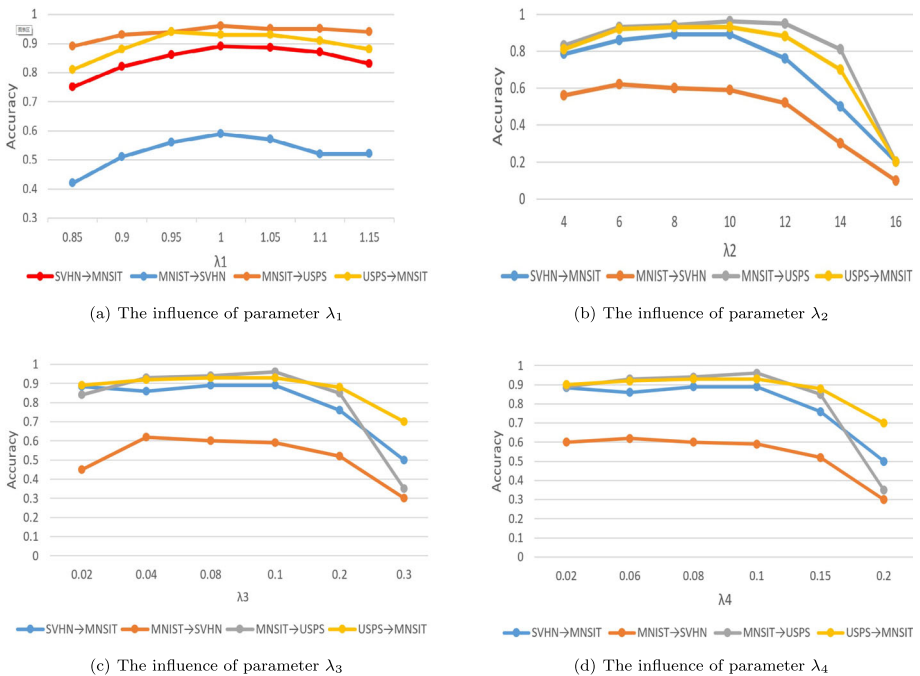


**Table 2** Performance(accuracy) on the Office-31 dataset

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
DANN	0.805	0.971	0.996	0.786	0.636	0.628	0.804
DTN	0.684	0.967	0.993	0.689	0.625	0.607	0.761
JAN-A[30]	0.86	0.967	0.997	0.851	0.692	0.707	0.846
CDAN [31]	0.845	0.968	0.994	0.775	0.66.2	0.648	0.816
MADA [32]	0.9	0.974	0.996	0.878	0.703	0.664	0.852
SWADL(Our)	0.89	0.988	0.996	0.909	0.726	0.685	0.865

### 5 Conclusion

In this survey, we introduce a novel training method along with well-developed algorithms for unsupervised domain adaptation, specifically focusing on Wasserstein adversarial domain adaptation with a shared latent space. The central concept behind our approach is to transfer shared latent knowledge between the source and target domains using Wasserstein GAN, resulting in higher-quality outcomes without compromising training stability or reconstruction quality. Despite not utilizing sophisticated network architectures like VGG-Net or Inception, our proposed approach achieves competitive results with simple networks. This demonstrates the effectiveness of our method in achieving desirable performance even with more straightforward models. Moving forward, our future work will involve a systematic exploration of long-distance transfer learning and other challenging applications, such as



**Fig. 4** Sensitivity analysis of hyperparameters

brain-computer interface signal classification, video analysis, and speech-based applications. By undertaking these studies, we aim to expand the scope and applicability of our approach to address broader and more complex domains.

**Data Availability Statement** No data is generated during this Study.

## Declarations

**Conflicts of interest** On behalf of all authors, the corresponding author declares there is no conflict of Interest.

## References

- Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans Knowl Data Eng (TKDE)* 22(10):1345–1359
- Gretton A (2012) A Kernel two-sample test. *J Mach Learn Res* 13:723–773
- Dziugaite GK, Roy DM, Ghahramani Z (2015) Training generative neural networks via Maximum Mean Discrepancy optimization. *Uai*
- Sun B, Feng J, Saenko K (2016) Return of frustratingly easy domain adaptation. In: Thirtieth AAAI conference on artificial intelligence
- Sun B, Saenko K (2016) Deep CORAL: correlation alignment for deep domain adaptation. In: ICCV workshop on transferring and adapting source knowledge in computer vision (TASK-CV)
- Ghifary M, Kleijn WB, Zhang M, Balduzzi D, Li W (2016) Deep reconstruction-classification networks for unsupervised domain adaptation. In: European conference on computer vision (ECCV), pp 597–613
- Huang J, Smola AJ, Gretton A, Borgwardt KM, Scholkopf B (2006) Correcting sample selection bias by unlabeled data. In: NIPS
- Chu W-S, De la Torre F, Cohn JF (2013) Selective transfer machine for personalized facial action unit detection. *CVPR*
- Jhuo IH, Liu D, Lee DT, Chang SF (2012) Robust visual domain adaptation with low-rank reconstruction. *CVPR*
- Gong B, Shi Y, Sha F, Grauman K (2012) Geodesic flow kernel for unsupervised domain adaptation. *CVPR*
- Qiu Q, Patel VM, Turaga P, Chellappa R (2012) Domain adaptive dictionary learning. *ECCV*
- Wang X, Shrivastava A, Gupta A (2017) A fast RCNN: hard positive generation via adversary for object detection. In: The IEEE conference on Computer Vision and Pattern Recognition (CVPR)
- Hoffman J, Tzeng E, Park T, Zhu J, Isola P, Saenko K, Efros A, Darrell T (2016) CyCADA: Cycle-Consistent Adversarial Domain Adaptation. In: Dy J, Krause A (eds) Proceedings of the 35th international conference on machine learning (proceedings of machine learning research), vol 80, 1994–2003
- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan. [arXiv:1701.07875](https://arxiv.org/abs/1701.07875)
- Haeusser P, Frerix T, Mordvintsev A, Cremers D (2018) Associative Domain Adaptation. [arXiv:1708.00938](https://arxiv.org/abs/1708.00938)
- Donahue J, Krähenbühl P, Darrell T (2016) Adversarial feature learning. [arXiv:1605.09782](https://arxiv.org/abs/1605.09782)
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. [arXiv:1411.1784](https://arxiv.org/abs/1411.1784)
- Taigman Y, Polyak A, Wolf L (2016) Unsupervised cross domain image generation. [arXiv:1611.02200](https://arxiv.org/abs/1611.02200)
- Zhu J, Park T, Isola P, Efros AA (2017) Unpaired image to image translation using cycle consistent adversarial networks. In: International conference on computer vision (ICCV)
- Liu MY, Tuzel O (2016) Coupled generative adversarial networks. In: Advances in neural information processing systems, pp 469–477
- Liu M, Breuel T, Kautz J (2017) Unsupervised image to image translation networks. [arXiv:1703.00848](https://arxiv.org/abs/1703.00848)
- Madadi Y, Seydi V, Nasrollahi K, Hosseini R, Moeslund T (2020) Deep visual unsupervised domain adaptation for classification tasks: a survey. *IET Image Proc* 14(19):3283–3299
- Zonoozi MH, Seydi V (2022) A survey on adversarial domain adaptation. *Neural Process Lett*. <https://doi.org/10.1007/s11063-022-10977-5>
- LeCun Y (1998) Gradient-based learning applied to document recognition. In: Proceedings of the IEEE 86:2278–2324
- Hull JJ (1994) A database for handwritten text recognition research. *PAMI* 2016:550–554

26. Netzer Y, Fillet M, Coates A, Bissacco A, Wu B, Ng AY (2011) Reading digits in natural images with unsupervised feature learning. In: NIPS
27. Saenko K, Kulis B, Fritz M, Darrell T (2010) Adapting visual category models to new domains. In: ECCV
28. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems*. <https://doi.org/10.1145/3065386>
29. Deng J, Dong W, Socher R et al (2009) Imagenet: a large-scale hierarchical image database. *Proc of IEEE Computer Vision & Pattern Recognition*, 248–255
30. Long M, Zhu H, Wang J, Jordan MI (2017) Deep transfer learning with joint adaptation networks. In: *International conference on machine learning*, PMLR, pp 2208–2217
31. Long M, Cao Z, Wang J, Jordan MI (2018) Conditional adversarial domain adaptation. In: *Advances in neural information processing systems*, pp 1640–1650
32. Pei Z, Cao Z, Long M, Wang J (2019) Multi-adversarial domain adaptation. [arXiv:1809.02176](https://arxiv.org/abs/1809.02176)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.