Targeted Transfer Learning: Leveraging Optimal Transport for Enhanced Knowledge Transfer

Sayyed Farid Ahamed*, Kazi Aminul Islam†, Sachin Shetty*
*Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA
†Department of Computer Science, Kennesaw State University, Marietta, GA, USA
{saham001, sshetty}@odu.edu*, {kislam4}@kennesaw.edu†

***Abstract*—Machine learning models are often built based on the notion that the labeled training data follows the same underlying distribution of the testing dataset. Often this assumption does not hold and the trained model's performance becomes limited during the deployment in the target environment. To address this problem, we present a transfer learning-based targeted transfer learning (TTL) approach that employs the optimal transport distance (OTD) metric to facilitate knowledge transfer between the source and target domains. TTL method uses a target-side adaptation methodology by intentionally modifying the source domain data using a small number of target domain samples to improve transfer performance between source and target datasets. Experimental results on the benchmark datasets show the effectiveness of the proposed approach without re-training the source model on the target environment, which contributes to the selection of an appropriate source model and improves the performance of the target model.**

***Index Terms*—Transfer Learning, Targeted Transfer Learning, Domain Adaptation, and Optimal Transport.**

I. Introduction

Transfer learning (TL) is a technique commonly used in artificial intelligence and machine learning to improve the performance of related tasks by transferring knowledge from one task to another. The basic idea is that rather than training a model from scratch, previously learned features from one domain can be applied with minor modifications to another related problem domain. TL is useful when data is limited and collecting new data is costly and difficult [1]. Transfer learning is a popular and rapidly expanding area of machine learning. Every day, new techniques and advancements such as domain adaptation, multi-task learning [1], and model fine-tuning [2] are explored and developed.

However, while transfer learning has many benefits, there are also several limitations that should be considered when applying this technique. Transfer learning may not be the optimal solution when target task is more complex or significantly dissimilar from the source task. Therefore, selecting the appropriate pre-trained model and adapting it to the target domain requires careful consideration and expertise. Otherwise, TL may introduce bias and overfit into the model [3]. To overcome these challenges, we must carefully select the source dataset and adapt the pre-trained model to minimize the disparity between the source and target. There are several approaches available in transfer learning to resolve the difference between two domains, such as domain adaptation [1], data augmentation [4], curriculum learning [5], data selection, and data shift [6].

The data shift technique modifies the source domain data distribution to better match the target domain data distribution to resolve the difference between the source and target domains. Data transfer from the target to the source is commonly known as target-side adaptation or target-oriented transfer learning. Here, the source domain data is modified to better correspond with the target domain data. Target-side adaptation focuses on modifying the source domain data using information from the target domain, as opposed to modifying the target domain data directly. The alternative data shift technique is a source-side adaptation, in which the target domain data is modified to more closely match the source domain data [6]. The primary objective of target-side adaptation is to enhance the performance of the model on the target task by providing more labeled data that is specific to the target domain. This technique is particularly useful when the pre-trained model performs poorly on the target task due to a lack of information about the target environment and a significant domain gap between the source and target [6]. However, it is essential that the additional target domain data accurately represent the target data distribution; otherwise, non-representative data can result in poor performance on the target task [3]. In 2020, Romero et al. [7] used a target-side adaptation technique for a small medical image dataset, which resulted in a significant performance boost. Ahamed et al. [8] presented an automated targeted transfer learning framework on the satellite image dataset. Authors of another paper [9] proposed a deep targeted transfer learning method for different conditional distribution datasets and demonstrated how cross-domain data can be aligned in target enrolment.

In this paper, we present a target-side adaptation technique, named targeted transfer learning (TTL), which transfers knowledge between source and target to get the desired performance while addressing the negative transfer challenge. The primary objective is to identify a set difference set that represents the mismatched features between the source and target datasets. The set difference set is constructed using an optimal transport approach, which assists in addressing the difference between these domains. The key contributions of our work are as follows:

- We evaluate our proposed TTL approach on five benchmark datasets.

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

- We utilize the optimal transport distance (OTD) metric between datasets and demonstrate how OTD can help to identify the appropriate source set for a given target set.
- Then, we measure the label-to-label OT distances between datasets and enhance the knowledge transferability between the source and target.
- The results demonstrate that intentionally modifying a small number of samples in the source set can improve the model performance in the targeted environment.

II. Methodology

A. Image Dataset

We consider five publically available image datasets: MNIST [10], EMNIST [11], FashionMNIST [12], KMNIST [13], and USPS [14] to demonstrate our proposed approach. The table I contains information about those datasets. The datasets were selected based on three main challenges: different sample sizes, dimensionality, and labels (e.g., MNIST has digits images and FashionMNIST has images of clothing). The last challenge, which relates to labeled datasets, is the most difficult. The first two challenges are similar to unlabeled datasets, which are simple to manage. For example, what if we are comparing MNIST (ten categories) to ImageNet (one thousand categories) and the label "3" in MNIST to "bag" in FashionMNIST? The proposed solution is to represent each label as a collection of points with that label and, for invariance datasets, treat these collections as probability distributions. In this manner, different types of datasets can be compared by examining their associated collections, which are viewed as probability distributions in the feature space. Therefore, the method is capable of calculating the distance between two distinct probability distributions. Moreover, these computations must be performed in a computationally feasible manner. This is where optimal transport emerges as a fundamental component of our methodology.

TABLE I
Image Datasets

Dataset	Pixel	#Images	Class	Label
MNIST	28*28	60k	10	Digits
EMNIST	28*28	60k	10	Digits
FashionMNIST	28*28	60k	10	Fashion products
KMNIST	28*28	60k	10	Japanese literature
USPS	16*16	7.29k	10	Digits

B. Targeted Transfer Learning (TTL)

In this paper, the source and target task is treated as an image classification problem, which is to identify images with the following label. Since both the source and target datasets generally share similar traits, conventional transfer learning can be utilized to classify images in the target environment. The purpose of the transfer learning process is to use the source model to assist the target model in performing the target task. Let us consider that an operator has a targeted metric in a given environment that needs to be achieved using traditional transfer learning. Due to domain-specific constraints, in many instances, direct implementation of transfer learning may not be effective. The most common solution is fine-tuning, in which the pre-trained model is re-trained using a target dataset, which is expensive, time-consuming, and difficult if the target data is limited. In the paper [8], we present a transfer learning-based 'targeted transfer learning (TTL)' process to enhance the performance of a source model without re-training in the target environment shown in Figure 1. Consider datasets as a set that consists of source set S and target set T. Using the optimal transport distance metric [II-C], we would like to create a set difference image set y. Then, set y is added to the source to create a source set and subtracted from the target set. On the newly created sets, we now evaluate the conventional transfer learning algorithm. The primary objective is to identify mismatched images between the source and target and generate a 'set difference' set. These set difference images mostly contain the contexts that are responsible for performance drop in the transfer learning process. In our prior work [8], we utilized combinatorial coverage methodology, a t-way coverage metric, on metadata to identify these images. In this paper, we present optimal transport (OT) as a model-agnostic data-driven method for detecting unlearned features between source and target.

C. Optimal Transport (OT)

Optimal transport is a principled approach to measure the distance between two distinct distributions. The basic idea behind optimal transport is to establish a mapping between each point in the source distribution and a corresponding point in the target distribution [15]. To achieve the most efficient and effective transportation, the goal is to minimize the total distance between all coupled points. However, for comparing two probability distributions using OT, it is essential to establish a distance metric between the points sampled from each distribution. For example, consider we are comparing two datasets in which each point 'z' consists of a feature vector (an image) and a label. Let's consider that we need to compute the distance between the pair (x, "bag"), where x is an image of "bag" from FashionMNIST, and the pair (x', "three"), where x' is an image of "3" from the MNIST dataset. Using only conventional methods to calculate the distances between images is relatively more straightforward. Nevertheless, measuring distance metrics for their labels is more difficult.

In this experiment, we are considering supervised learning. Consider data D as a set of feature and label pairs $(x, y)\epsilon X \times Y$, where X represents the feature space and Y is the label set. Assume the feature spaces of two datasets D_a and D_b have the same dimensionality, but their respective label sets Y_a and Y_b are distinct. As mentioned previously, measuring the distance between datasets becomes more complicated when the labels are considered. Therefore, the labels Y_a and Y_b are represented as conditional probability distributions $P_y = P(X|Y = y)$. The objective is to define the distance metric $d(D_a, D_b)$ without using any external parameters. However, between the pairs $(x, y), (x', y')$, the distance metric could be computed using optimal transport [16] by defining a metric on Z as

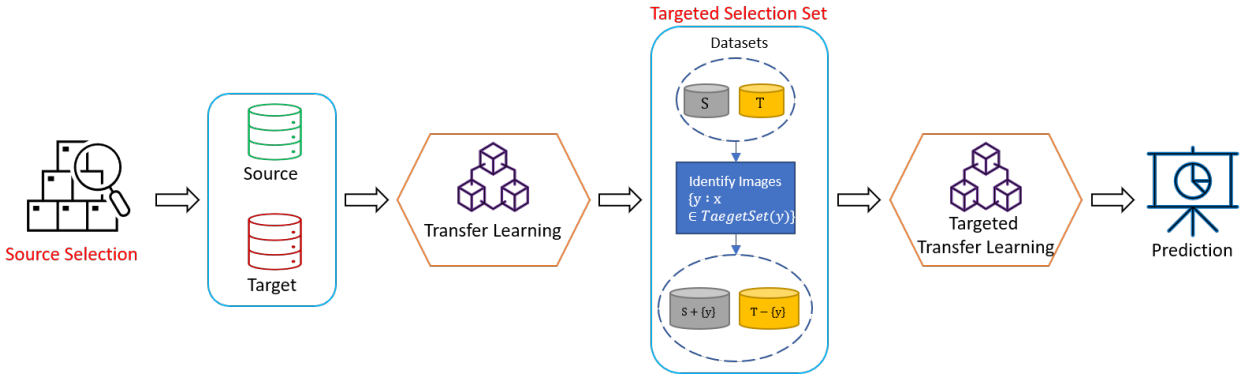

Fig. 1. Targeted transfer learning overview.

$d_Z(z,z') = (d_X(x,x')^p + d_y(y,y')^p)^{1/p}$, for $p \geq 1$. Now, we can compute the distance between distributions over feature-label pairs using d_Z as the ground cost. Therefore, optimal transport can be also used to determine the distance between datasets [16].

$$OT(D_a, D_b) = min_{\pi \epsilon \prod(P_a, P_b)} \int_{Z \times Z} d(z,z') d\pi(z,z') \quad (1)$$

The primary objective of this distance metric is to be applicable even if the user has different sets of labels with no correspondence (e.g., letters to numbers). Because in this case each label is treated as a geometric feature, like a vector, which also provides a computational advantage.

III. Results and Discussion

The initial step in implementing targeted transfer learning (TTL) is to evaluate the optimal transport (OT) distance metric to create an appropriate source dataset for transfer learning in a targeted environment. In this experiment, we are considering five publicly available datasets, namely MNIST, EMNIST, FashionMNIST, KMNIST, and USPS. The details of these datasets are presented in II-A. Then, we analyzed the performance metric between targeted selection and random selection to evaluate the TTL procedure. The primary objective is to achieve better performance compared to the baseline score using conventional transfer learning with minimal targeted images.

A. Transfer Learning [Dataset Selection]

In this experiment, the optimal transport distance metric is utilized to identify the appropriate source dataset for transfer learning. As previously stated, we are evaluating five different publicly available datasets in this paper. First, the pairwise optimal transport distance between datasets is computed. Figure 2 shows the OT distance with labels between the five datasets. The table is coloured according to their distances. Here, the orange colour indicates they are far apart and the green colour indicates they are close together. For example, the FashionMNIST dataset is close to the USPS dataset (3.85) in terms of distance but far from the KMNIST dataset (4.95).

Consider, for instance, that we want to compute the distance between FashionMNIST and all other datasets. FashionMNIST is already known to be close to the USPS dataset, whereas

Label	MNIST	EMNIST	FashionMNIST	KMNIST	USPS
MNIST	0.00	4.33	4.86	4.75	4.15
EMNIST	4.33	0.00	4.90	4.82	4.55
FashionMNIST	4.86	4.90	0.00	4.95	3.85
KMNIST	4.75	4.82	4.95	0.00	4.36
USPS	4.15	4.55	3.85	4.36	0.00

Fig. 2. Pairwise OT distance ($\times 10$) between datasets.

other datasets are far away. In theory, we can predict transfer learning performance if FashionMNIST is used as the source dataset and all other datasets are used as the target datasets. To evaluate, we first train a model on the source dataset (FashionMNIST) and then transfer it to the target datasets (MNIST, EMNIST, KMNIST, and USPS) to fulfil the target task. We observe that the pre-trained model performs better on the USPS dataset than the other dataset. We got approximately 64% accuracy in the USPS dataset by directly applying the pre-trained model (source model) to the target. When the model is tested on the KMNIST dataset, the accuracy drops to around 40%. Therefore, when multiple source datasets are available, the OT distance metric could provide the best source model.

B. Targeted Transfer Learning

Targeted transfer learning (TTL) will be applied when a user does not get the required performance from the zero-shot transfer learning implementation. In this paper, optimal transport (OT) is used to identify set difference images for inclusion in the augmented training set. The original images from the dataset are used to create source and target sets in this experiment. We conducted two experiments to evaluate the effectiveness of optimal transport to identify the set difference images for the TTL process. The MNIST dataset has been chosen as the source set for both experiments, while FashionMNIST and KMNIST serve as the target sets. To properly demonstrate our methodology, we use thirty thousand (30k) images from both datasets as the source and target and conduct each experiment three times. As precision, recall, and F1 score trend with accuracy, we focus our analysis on accuracy.

1) MNIST vs FashionMNIST: As described in II-C, we compute the label-to-label Optimal Transport (OT) distances between the MNIST and FashionMNIST datasets. Each label

in both datasets is used to compute the OT cost and determine the distance between them. Figure 3 shows the label-to-label distance between datasets. The bold-collared boxes represent a greater distance between labels, whereas the light-collared boxes represent a smaller distance. For example, the FashionMNIST label four ('4') is approximately [43.3] distance apart from the MNIST label one ('1'). However, we observed that not all labels contribute equally to the OT distance. By utilizing the mismatched labels between the two sets, our aim is to enhance the knowledge transfer between datasets.

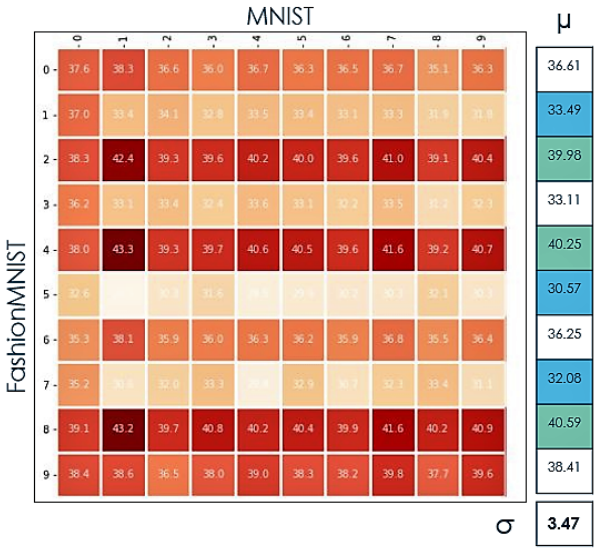

Fig. 3. Label-to-label distances between MNIST vs FashionMNIST.

From Figure 3, the labels '8', '4', and '2' of FashionMNIST had the highest mean, indicating that they are farther away from the corresponding MNIST labels compared to other labels in FashionMNIST. We use these label images to form the set difference set. Conversely, labels '5', '7', and '1' of FashionMNIST are found to be very close to the source labels. These label images are utilized to validate the TTL process.

In this experiment, we conducted three different cases to evaluate the process. First, the source-only $[S]$ case, in which a model is trained using only the source dataset (MNIST) and then tested on the target set (FashionMNIST). In all cases, the number of source and target images will be the same (30K). In the second case $[S+s(T)]$, we use 27K images from the source set and 3K images from the set difference set to create a source set of identical size. These 3K images are chosen randomly from the three highest-mean sets (set difference set). Lastly, we generate another source set $[S+r(T)]$ similarly using 3K random images from the three lowest-mean sets and referred to as the random selection set. As mentioned previously, for each set of counting methods, we select the images via random sampling and train a model on the augmented set.

Figure 4 illustrates the performance comparison of MNIST and FashionMNIST targeted transfer learning process. In the context of transfer learning, three scenarios are examined. First, we directly apply the pre-trained model to the target dataset, also known as "zero-shot transfer learning." Second, we adjust the output layer in a way that ensures its compatibility with the target dataset. Finally, for a more comprehensive study, the last two layers are adapted, which includes both the output layer and the preceding dense layer, in order to better align the pre-trained model with the target dataset. For the result analysis, we consider the second scenario.

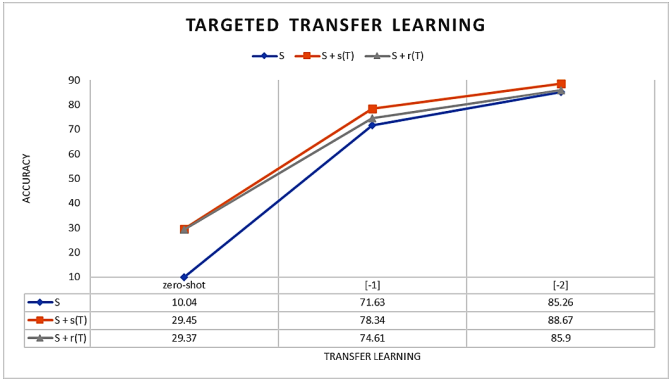

Fig. 4. Performance comparison of targeted transfer learning MNIST → FashionMNIST.

When we consider the source-only [S] set, the pre-trained model gave around 71.63% accuracy on the target dataset which will serve as the baseline accuracy. In the second case, where we added selected targeted data (set difference set) to the source $[S+s(T)]$ set, we achieved approximately 78.34% accuracy, which is approximately 7% higher than the baseline performance. Here, in all cases the number of samples is identical. Finally, we have added like the previous case the same number of samples into the source $[S+r(T)]$ set but here we select the samples randomly from the lowest-mean set (random selection set) and we obtain around 74.61% accuracy. In both cases [two and three], the source set contains the target dataset images; however, the inclusion sets are distinct, which has an effect on the performance of both cases. This suggests that targeted set selection using optimal transport distance metric over the source set increases the knowledge transfer between datasets.

2) MNIST vs KMNIST: Similar to the preceding example, we compute the label-to-label Optimal Transport (OT) distance metric between the MNIST and KMNIST datasets [Figure 5]. We observed that labels "3", "0", and "4" of KMNIST have the highest mean and the labels "2", "5", and "1" have the lowest mean values. The highest mean of the OT distance labels will contribute to creating the set difference set, while the lowest means will be used to validate the TTL process (random selection set).

Figure 6 illustrates the findings. In the source-only $[S]$ case, we get almost 62.21% accuracy, which will serve as our baseline score. For the second case, the accuracy is increased to 71.46% where additional targeted data is added to the source set. Finally, the repetition of case two using the random selection set resulted in a decrease in accuracy. Even though the target data has been introduced into the source set in both cases, careful selection could improve the knowledge transfer in transfer learning.

From Figure 2, the pairwise OT distance between datasets

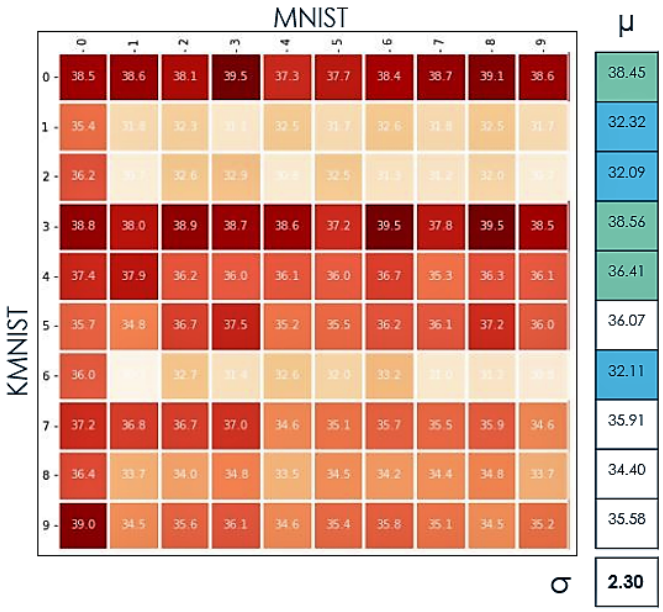

Fig. 5. Label-to-label distances between MNIST vs KMNIST.

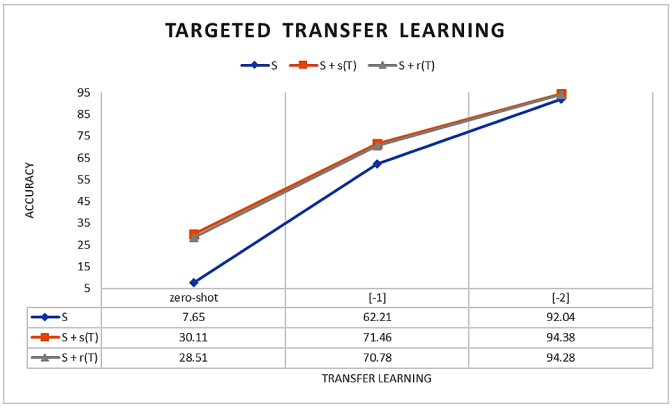

Fig. 6. Performance comparison of targeted transfer learning MNIST → KMNIST.

indicates that KMNIST is closer to MNIST than the FashionMNIST dataset. We can also verify these findings based on these results. In Figures 3 and 5, the standard deviation (SD) is also calculated to measure variations in the distance mean values. We observed that FashionMNIST has a higher SD than KMNIST, indicating that FashionMNIST label-to-label distances are more spread and variable. As a result, the TTL process performs better where higher SD is present in the datasets, which also assists us in identifying outliers or anomalies.

IV. Conclusion and Future Work

In this paper, we present a target-side adaptation transfer learning framework and approach called TTL for improving the target model's performance in a given environment. The performance evaluation shows that the transfer learning approach can be enhanced with a minimal data shift. The performance difference between targeted selection and random selection validates our hypothesis. In addition, we demonstrate that the optimal transport distance (OTD) metric is adaptable and scalable enough to be used in conventional transfer learning circumstances. In future research, the standard deviation of the label-to-label distance will play a crucial role in constructing the set difference set. However, more research work is required to establish OTD as a model-agnostic approach across multiple domains. Our long-term goal is to implement an automated process for obtaining source-target correlation information prior to the transfer learning process. Before target deployment, the operator can resolve and control the transferability between domains.

Acknowledgment

This work is supported in part by DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory, National Science Foundation under Grant No. 2219742 and Grant No. 2131001

References

[1] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[2] Y. Guo, H. Shi, A. Kumar, K. Grauman, T. Rosing, and R. Feris, "Spottune: transfer learning through adaptive fine-tuning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4805–4814.

[3] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *arXiv preprint arXiv:1411.1792*, 2014.

[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[5] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[6] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*. PMLR, 2015, pp. 97–105.

[7] M. Romero, Y. Interian, T. Solberg, and G. Valdes, "Targeted transfer learning to improve performance in small medical physics datasets," *Medical physics*, vol. 47, no. 12, pp. 6246–6256, 2020.

[8] S. F. Ahamed, P. Aggarwal, S. Shetty, E. Lanus, and L. J. Freeman, "Attl: An automated targeted transfer learning with deep neural networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–7.

[9] B. Yang, Y. Lei, X. Li, and C. Roberts, "Deep targeted transfer learning along designable adaptation trajectory for fault diagnosis across different machines," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 9, pp. 9463–9473, 2022.

[10] Y. LeCun, C. Cortes, C. Burges *et al.*, "Mnist handwritten digit database," 2010.

[11] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, "Emnist: Extending mnist to handwritten letters," in *2017 international joint conference on neural networks (IJCNN)*. IEEE, 2017, pp. 2921–2926.

[12] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[13] T. Clanuwat, M. Bober-Irizar, A. Kitamoto, A. Lamb, K. Yamamoto, and D. Ha, "Deep learning for classical japanese literature," *arXiv preprint arXiv:1812.01718*, 2018.

[14] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, vol. 2, 1989.

[15] C. Villani *et al.*, *Optimal transport: old and new*. Springer, 2009, vol. 338.

[16] D. Alvarez-Melis and N. Fusi, "Geometric dataset distances via optimal transport," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 428–21 439, 2020.