

MULTIPLE SUBSPACE ALIGNMENT IMPROVES DOMAIN ADAPTATION

Kowshik Thopalli¹, Rushil Anirudh², Jayaraman J. Thiagarajan², Pavan Turaga¹

¹ Geometric Media Lab, ² Center for Applied Scientific Computing,
Arizona State University, Lawrence Livermore National Laboratory

Email: {kthopall@asu.edu, anirudh1@llnl.gov, jjayaram@llnl.gov, pturaga@asu.edu}

ABSTRACT

We present a novel unsupervised domain adaptation (DA) method for cross-domain visual recognition. Though subspace methods have found success in DA, their performance is often limited due to the assumption of approximating an entire dataset using a single low-dimensional subspace. Instead, we develop a method to effectively represent the source and target datasets via a collection of low-dimensional subspaces, and subsequently align them by exploiting the natural geometry of the space of subspaces, on the Grassmann manifold. We demonstrate the effectiveness of this approach, using empirical studies on two widely used benchmarks, with performance on par or better than the performance of the state of the art domain adaptation methods.

Index Terms— Unsupervised Learning, Domain Adaptation, Grassmann manifold

1. INTRODUCTION

Powered by advances in representation learning, the field of computer vision has witnessed unprecedented growth in predictive modeling. Despite the effectiveness of these solutions, it is well known that their performance suffers when the trained models are deployed in new environments, particularly when there are systematic domain shifts not known *a priori* [1]. For example, a classifier trained exclusively using images from a DSLR camera might not produce precise predictions on target images from a mobile phone camera. Tackling this discrepancy is central to enabling the effective use of such predictive models in practice. Consequently, a large class of techniques that can leverage labeled data in a source domain to infer an accurate model in an unobserved and unlabeled target domain, have been developed [2],[3]. Commonly referred to as *unsupervised domain adaptation*, these techniques attempt to learn domain-invariant representations and subsequently, align the two datasets by minimizing some form of divergence in the learned feature space.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-PROC-767906. KT and PT were supported in part by ARO grant number W911NF-17-1-0293 and NSF CAREER award 1451263.

Broadly, domain adaptation methods can be categorized as *conservative*, which assume that there exists a single classifier in the hypothesis space that performs well in both domains, or *non-conservative*, which avoid that assumption [4]. Regardless of the category, the methods vary based on their choice of representation and the strategy to align the source and target distributions. For example, optimal transport has proven to be a powerful alignment strategy regardless of how the datasets are represented [5, 6]. State-of-the-art techniques adopt domain adversarial training [7, 8], which attempts to construct a feature learner that matches the source and target distributions in a latent space. However, these methods assume access to both source and target datasets at train time, which can be unrealistic in several scenarios.

In this paper, we consider the problem where we do not have access to labeled target data, while learning the representations. Instead of inferring features based on an alignment objective (e.g. adversarial loss), the two steps of representation learning and alignment are decoupled. More specifically, our approach performs subspace modeling to build data representations, and utilizes tools from differential geometry of the Grassmann manifold to perform alignment. This general idea has been explored by several works in the past [9, 10, 2, 11]; however, those methods rely on a simplistic model, a single low-dimensional subspace, to model datasets.

In contrast, we show that complex datasets can be more effectively modeled through a union of multiple low dimensional subspaces, and this can thereby lead to improved alignment even under challenging domain shifts. A related idea has been explored in [12], wherein they fit multiple subspaces to clusters in the two datasets, and align via parallel transport with regard to a single mean subspace. When compared to our approach, this does not guarantee that the subspaces are discriminative and the alignment fails to take advantage of the increased representational capacity. We on the other hand develop a simple, yet effective, strategy for fitting multiple low-dimensional subspaces to the two domains, and propose to align them by exploiting the natural geometry of linear subspaces on the Grassmann manifold, in order to construct domain-invariant features. Our approach falls under the *conservative* category of domain adaptation. When compared to

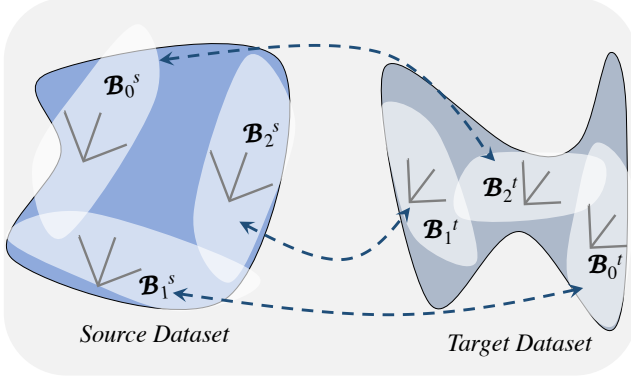


Fig. 1. Multi-subspace alignment: Our approach proposed to represent the source and target datasets via a collection of discriminative subspaces. We also solve for the correspondence between the source and target subspaces by exploiting the natural geometry of subspaces on the Grassmann manifold. The correspondence is shown here by dashed lines.

domain adversarial training, a crucial advantage of our approach is that it is applicable even with smaller datasets. An overview of the proposed approach is shown in fig 1.

We perform experiments on all benchmarks from the Office-Caltech10 dataset, using both pre-defined SURF descriptors as well as features extracted from a pre-trained deep neural network architecture. Our results show that the proposed approach achieves excellent performance on this benchmark, outperforming several existing methods for unsupervised adaptation. More importantly, we observe that the multiple subspace assumption provides significant performance improvements over approaches that use a single subspace.

2. PRELIMINARIES

In this section, we briefly review the mathematical foundations used in designing our approach.

Grassmann manifold: The core idea of our approach relies on representing a dataset with a collection of linear subspaces – which are known to be points on the Grassmann manifold, $\mathbb{G}_{r,n}$, which describes the collection of all r -dimensional linear subspaces in \mathbb{R}^n . Each point on $\mathbb{G}_{r,n}$ is represented as a basis set, *i.e.* a collection of r orthonormal vectors $\mathcal{B} := \{b_1, b_2, \dots, b_r\}$. For more details on the geometry and properties of the manifold, we refer our readers to more comprehensive sources [13, 14].

Metric on the Grassmann manifold: We use the symmetric directional distance (d_Δ) to compute distances between points on the Grassmann manifold. We show a more general form of the distance function, which applies to even subspaces with different dimensions r [15]. This is a popularly adopted measure in several applications [16, 17]. Formally,

this can be stated as follows:

$$d_\Delta(\mathcal{B}_1, \mathcal{B}_2) = \left(\max(r_1, r_2) - \sum_{i,j=1}^{r_1, r_2} (b_{1,i}^\top b_{2,j})^2 \right)^{\frac{1}{2}}, \quad (1)$$

where, r_1 and r_2 denote the ranks of the basis sets \mathcal{B}_1 and \mathcal{B}_2 respectively. Note, this expression is equivalent to the chordal metric [18] and is applicable to the case when $r_1 = r_2$.

3. APPROACH: MULTI-SUBSPACE ALIGNMENT

In this section, we describe the proposed approach for domain adaptation, which first builds a multi-subspace representations for the two domains, and then performs alignment through Grassmann analysis. We assume we have access to a labeled source dataset $\{X_i^s, y_i^s\}_{i=1}^{N_s} \subset D_S$, and an unlabeled target dataset $\{X_i^t\}_{i=1}^{N_t} \subset D_T$, where D_S and D_T denote the source and target domains, N_s, N_t are the number of samples. We assume that every sample in both the domains, $X \in \mathbb{R}^d$, but drawn from different, unknown distributions. Our goal in this paper is to transform the source dataset in such a way that a classifier trained only on the adapted source dataset generalizes well to the unlabeled target dataset.

3.1. Fitting multiple subspaces

The first step in our approach is to learn effective representations for the source and target datasets independently, such that aligning those features will lead to maximal knowledge transfer across the domains. To this end, we develop a greedy approach that identifies a set of low-dimensional subspaces, that can collectively describe all variations in the high-dimensional dataset. Following the standard practice in subspace methods for domain adaptation, we begin by fitting a k -dimensional subspace using Principal Component Analysis (PCA) on the entire dataset. Subsequently, we select samples that incur high errors, measured through a normalized reconstruction error for each sample and compared against a user-defined tolerance τ , in the subspace approximation. In order to improve the approximation, we fit a second subspace that produces a high-fidelity representation for the selected samples. We continue this process until the following criteria is met: total number of samples with errors above τ is less than the chosen subspace dimension k .

3.2. Alignment of multiple subspaces

Using Algorithm 1, the source and target datasets are now represented via collection of subspaces and are denoted by \mathcal{M}_S and \mathcal{M}_T respectively. The next crucial step in achieving domain adaptation is aligning the set of source subspaces with those from the target dataset. Domain alignment is a well-studied problem, and hence there exist a wide range of strate-

Algorithm 1: Dataset Approximation using Multiple Subspaces

Input: Dataset \mathbf{X} , subspace dimension k , tolerance τ

Output: Collection of subspaces (\mathcal{M}), list of subspace IDs to which each sample is associated (P)

Initialize \mathcal{M} as an empty set

sID = 1

while no convergence **do**

$\mathcal{B} \leftarrow \text{PCA}(\mathbf{X}, k)$ // fit subspace

$ids \leftarrow \{i \in [N] | X_i \in \mathbf{X}, \text{error}(X_i, \mathcal{B}) \geq \tau\}$

$\mathbf{X}_e \leftarrow \mathbf{X}[ids]$ // errors

$\mathbf{X}_v \leftarrow \mathbf{X} \setminus \mathbf{X}_e$ // inliers

$\mathcal{B} \leftarrow \text{PCA}(\mathbf{X}_v, k)$ // refit subspace

$P[i] = \text{sID}, \forall i \in ids$

$\mathcal{M} \leftarrow \mathcal{M} \cup \mathcal{B}$

$\mathbf{X} \leftarrow \mathbf{X}_e$

 sID = sID + 1

end

gies to solve the alignment problem; e.g. dynamic programming and optimal transport; however, most of these methods operate directly on the pairwise distance matrix between samples across the domains. However, in our current setup, since the samples that need to be aligned are points on the Grassmann manifold, we propose to utilize the distance function defined in (1). Note, when there is a single low-dimensional subspace in each domain, approaches such as the geodesic flow kernel can be used [2]. However, no such techniques exist for dealing with multi-subspace representations. Hence, we resort to a greedy strategy that assigns unique matches to each source subspace by finding its closest match in the set of target subspaces. This is carried out using geometric distances computed between every pair of subspaces in \mathcal{M}_S and \mathcal{M}_T . Naturally, the complexity of this step grows quadratically with the number of subspaces and scalability can be a bottleneck with large datasets. In such cases, a more efficient alignment strategy such as stochastic optimal transport [5] could be used as an alternative.

Alignment between Matched Subspaces: We use the subspace alignment (SA) strategy proposed in [10] to align each source subspace with its the matched target subspace. The alignment of two subspaces on the Grassmann manifold is modeled via a linear transformation, which can be computed in closed form [10]. In other words, the basis vectors are aligned using a transformation matrix $\mathbf{A} \in \mathbb{R}^{k \times k}$ from the m^{th} source subspace \mathcal{B}_m^s to the n^{th} target subspace \mathcal{B}_n^t . The transformation is obtained by minimizing the following objective:

$$\mathbf{A}^* = \arg \min_{\mathbf{A}} \|\mathcal{B}_m^s \mathbf{A} - \mathcal{B}_n^t\|_F, \quad (2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The solution to (2)

Algorithm 2: Proposed domain adaptation algorithm

Input: Source dataset \mathbf{X}^s , target dataset \mathbf{X}^t , source labels \mathbf{y}^s , subspace dimension k , tolerance parameters τ^s and τ^t

Output: Predicted target labels $\{y_j^t\}$

$\mathcal{M}_S, P_S \leftarrow \text{fit}(\mathbf{X}^s, k, \tau^s)$ // Invoke Algorithm 1 ;

$\mathcal{M}_T, P_T \leftarrow \text{fit}(\mathbf{X}^t, k, \tau^t)$ // Invoke Algorithm 1 ;

$D \leftarrow \text{distancematrix}(\mathcal{M}_S, \mathcal{M}_T)$ // using (1) ;

Greedy matching of subspaces across domains with D ;

Compute \mathcal{B}_m^{ta} , for every $\mathcal{B}_m^s \in \mathcal{M}_S$ // using (3) ;

Obtain features $Z_i^s = \mathcal{B}_m^{ta} X_i^s$, where $P_S[i] = m$;

Obtain features $Z_j^t = \mathcal{B}_n^t X_j^t$, where $P_T[j] = n, \forall i$;

$\{y_j^t\} \leftarrow \text{NN-Classifier}(\{Z_i^s, y_i^s\}, \{Z_j^t\}), \forall j$;

is given as $\mathbf{A}^* = (\mathcal{B}_m^s)^T \mathcal{B}_n^t$. This implies that the adjusted coordinate system, also referred as the *target aligned source coordinate system* can be obtained as

$$\mathcal{B}_m^{ta} = \mathcal{B}_m^s (\mathcal{B}_m^s)^T \mathcal{B}_n^t. \quad (3)$$

We obtain representations for the source dataset by projecting each sample $\{X_i^s\}$ onto the corresponding target aligned subspace as indicated by the association list $P_S[i]$. On the other hand, each of the target samples $\{X_j^t\}$ are directly projected onto the target subspace \mathcal{B}_n^t , where $P_T[j] = n$. Following previous works such as [11, 10], a nearest neighbor (NN) classifier is then used with these representations, which we expect to be effective for both source and target datasets. An overview of this algorithm can be found in Algorithm 2.

4. EXPERIMENTS

We evaluate our approach on benchmark domain adaptation tasks from widely used object recognition datasets, similar to several existing works[2], [9], [11] and [19].

4.1. Datasets

The Office-Caltech10 [3, 20] dataset is comprised of images from four different domains - Amazon (online merchant), Webcam (acquired using a webcam), DSLR (captured through a digital SLR camera) and Caltech (from Caltech dataset). For brevity, we refer to these domains using the letters A, W, D, and C respectively. The variations across the domains can be attributed to a variety of factors including presence/absence of background, lighting conditions, noise, etc. Each of the domains is comprised of images from 10 common object categories (backpack, bike, calculator, headphones, keyboard, laptop computer, monitor, mouse, mug, and projector). Sample images for the *Calculator* class from each of the four domains are shown in Figure 2. We consider adapting between every pair of domains, resulting in a total of 12 cases.

Domains	NA	GFK	SA	SDA	PT	Proposed
A-C	26.03	35.6	38.4	39.98	41.4	45.05
C-A	23.69	36.9	40.6	49.69	49.4	57.09
D- A	28.57	32.5	33.12	38.73	38.2	45.92
A-D	25.48	35.2	37.57	33.76	—	41.4
W-A	20.8	31.1	37.4	39.25	38.7	49.37
A-W	23.3	34.4	38.6	30.85	40.4	43.05
W-C	16.4	27.2	32.3	34.73	37.1	36.51
C-W	20	33.7	36.8	38.98	—	40.67
W-D	40.5	70.6	80.3	75.8	—	85.98
D-W	53	74.9	83.6	76.95	84.8	88.81
C-D	21.7	35.2	39.6	40.13	48.2	47.13
D-C	24.8	29.8	32.4	35.89	—	35.4
Avg.	27.07	39.8	44.24	44.56	47.12	51.37

Table 1. Classification accuracy on Office-Caltech dataset with SURF features



Fig. 2. Example images from Office-Caltech10. From left to right- Amazon, Webcam, DSLR and Caltech domains

We consider the use of two feature sets, commonly adopted in vision systems, for each of these domains: (i) SURF [21] features: Each image is encoded using a visual dictionary of 800 words, and the encodings are normalized to have a zero mean and unit standard deviation in each dimension, (ii) Decaf representations [22]: Following [22], we construct features using the activations (4096 dimensions) from the sixth layer of a network pre-trained on the ImageNet dataset [23].

4.2. Experimental Setup

We compare our method with the following baseline methods in the unsupervised domain adaptation literature. As described earlier, we use a nearest neighbor classifier in all of our experiments.

NA No Adaptation - In this naïve baseline, we compare the performance of classifiers from the source domain applied directly to the target domain.

GFK Geodesic flow kernel[2] - This approach fits a single subspace to each domain and considers all the points along the geodesic between the two subspaces on the Grassmann manifold to perform adaptation.

SA Subspace alignment [10] - Using a single subspace representation for both domains, this approach aligns the subspaces via a linear transform.

SDA Subspace distribution alignment [11] - In addition to aligning the source and target subspaces, this approach also aligns the actual distributions.

PT Domain adaptation via parallel transport on Grassmannian [12]- This approach considers the use of multiple sub-

Domains	NA	SA	GFK	Proposed
A-C	78.54	79.61	80.32	85.12
C-A	87.05	87.06	87.27	87.16
D-A	75.89	73.49	75.27	79.54
A-D	80.25	81.53	80.89	82.16
W-A	73.07	75.16	75.16	77.87
A-W	77.31	78.31	76.95	72.8
W-C	68.21	68.83	67.76	72.21
C-W	72.2	75.59	75.59	80.67
D-W	97.97	98.98	98.98	99.66
W-D	98.98	100	100	100
C-D	80.89	80.25	83.44	87.26
D-C	70.08	69.99	69.1	74.4
Avg.	80.12	80.73	80.90	83.24

Table 2. Classification accuracy on Office-Caltech dataset with decaf features

spaces on both the source and target, and performs parallel transport on the Grassmann manifold. Note, they report their results on only 8 of the 12 cases.

A source (S) to target (T) domain adaptation is represented as S-T. For both SURF and Decaf representations, we repeated the adaptation experiments for all the 12 cases. As described in Section 3, we first represent the dataset via multiple subspaces, which resulted in 3 to 4 subspaces on an average for our datasets. We then align them using the proposed greedy approach. The results using the SURF and the Decaf features are reported in Table 1 and Table 2 respectively. Some of the baseline results for Decaf features are from [24]. As it can be observed from the results, the proposed approach significantly outperforms the baseline cases. The subspace dimension k was fixed at 45 across all the 12 settings, and the thresholds τ^s and τ^t were chosen from $[0.1, 0.6]$. It is important to note that the three hyperparameters namely, the subspace dimension k , the error thresholds for source and target - τ^s and τ^t were tuned by grid search directly via the test error as is done in the previous works in the DA community [25, 6, 2].

5. CONCLUSION AND FUTURE WORK

We presented a novel domain adaptation method based on multiple subspace alignment. We proposed to represent data via a collection of low dimensional subspaces, and to subsequently align subspaces across source and target domains. Empirical studies on benchmark datasets were used to demonstrate the superiority of our approach in unsupervised DA for object recognition, when compared to existing subspace methods [2, 10, 11, 12]. Possible extensions to this work include exploring the optimal transport [26] techniques, in particular, the recent approaches designed for large-scale data [5], as an alternative for performing alignment between source and target representations, in order to alleviate the scalability issues with our approach.

6. REFERENCES

- [1] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, no. 1-2, pp. 151–175, May 2010.
- [2] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman, "Geodesic flow kernel for unsupervised domain adaptation," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2066–2073, 2012.
- [3] Raghuraman Gopalan, Ruonan Li, and Rama Chellappa, "Domain adaptation for object recognition: An unsupervised approach," *2011 International Conference on Computer Vision*, pp. 999–1006, 2011.
- [4] Rui Shu, Hung Bui, Hirokazu Narui, and Stefano Ermon, "A DIRT-t approach to unsupervised domain adaptation," in *International Conference on Learning Representations*, 2018.
- [5] Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel, "Large-scale optimal transport and mapping estimation," in *Proceedings of the International Conference in Learning Representations*, 2018.
- [6] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1853–1865, 2017.
- [7] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, "Adversarial discriminative domain adaptation," in *Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 1, p. 4.
- [8] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A. Efros, and Trevor Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, July 10-15, 2018*, 2018, pp. 1994–2003.
- [9] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars, "Subspace alignment for domain adaptation," *CoRR*, vol. abs/1409.5241, 2014.
- [10] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013*, 2013, pp. 2960–2967.
- [11] Baochen Sun and Kate Saenko, "Subspace distribution alignment for unsupervised domain adaptation," in *BMVC*, 2015, pp. 24–1.
- [12] Ashish Shrivastava, Sumit Shekhar, and Vishal M Patel, "Unsupervised domain adaptation using parallel transport on grassmann manifold," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. IEEE*, 2014, pp. 277–284.
- [13] Alan Edelman, Tomás A Arias, and Steven T Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.
- [14] P-A Absil, Robert Mahony, and Rodolphe Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [15] Xichen Sun, Liwei Wang, and Jufu Feng, "Further results on the subspace distance," *Pattern Recognition*, vol. 40, no. 1, pp. 328–329, 2007.
- [16] Ronen Basri, Tal Hassner, and Lihi Zelnik-Manor, "Approximate nearest subspace search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 266–278, 2011.
- [17] Esam Sharafuddin, Nan Jiang, Yu Jin, and Zhi-Li Zhang, "Know your enemy, know yourself: Block-level network behavior profiling and tracking," in *IEEE Global Telecommunications Conference (GLOBECOM 2010)*, 2010, pp. 1–6.
- [18] Ke Ye and Lek-Heng Lim, "Schubert varieties and distances between subspaces of different dimensions," *SIAM Journal on Matrix Analysis and Applications*, vol. 37, no. 3, pp. 1176–1197, 2016.
- [19] Ashish Shrivastava, Sumit Shekhar, and Vishal M Patel, "Unsupervised domain adaptation using parallel transport on grassmann manifold," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. IEEE*, 2014, pp. 277–284.
- [20] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, "Adapting visual category models to new domains," in *ECCV*, 2010.
- [21] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.
- [22] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," *CoRR*, vol. abs/1310.1531, 2013.
- [23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee*, 2009, pp. 248–255.
- [24] Lingkun Luo, Liming Chen, Shiqiang Hu, Ying Lu, and Xiaofang Wang, "Discriminative and geometry aware unsupervised domain adaptation," *Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. abs/1712.10042, 2017.
- [25] Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel, "Large-scale optimal transport and mapping estimation," *arXiv preprint arXiv:1711.02283*, 2017.
- [26] Cédric Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.