Advanced Machine Learning Project
Subspace Alignment and Optimal Transport

Fatemeh Gol Pour   Anna Karolin

Abstract

This project explores domain adaptation methods to transfer data distributions in visual classification tasks. We implement and evaluate two domain adaptation techniques: subspace alignment and entropic regularized optimal transport. Using the Office/Caltech dataset, we focus on transferring knowledge from the Webcam domain to the DSLR domain and vice versa by aligning source and target data distributions through principal component analysis and optimal transport techniques. Experimental results demonstrate the effect of these methods in classification accuracy on target domain data, showcasing their potential to mitigate domain inconsistencies in real-world computer vision applications.

1. Introduction

Domain adaptation (DA) is an ongoing challenge in machine learning, particularly in visual recognition tasks, where the distribution of data often varies significantly across domains. This can occur due to differences in data sources, acquisition conditions, or sensors, which leads to reduced model performance when applying a classifier trained on one domain to another. The goal of DA is to bridge this gap by adapting models to perform well across different but related domains without requiring labeled data from the target domain.

This project investigates two unsupervised domain adaptation methods: subspace alignment and entropic regularized optimal transport. The subspace alignment method leverages principal component analysis (PCA) to project both source and target data into a shared lower-dimensional subspace, aligning the source and target representations to minimize distributional divergence. The optimal transport method, on the other hand, uses the Sinkhorn-Knopp algorithm to map the source data distribution to the target distribution, guided by an entropic regularization parameter.

We evaluate these methods on the Office/Caltech dataset, focusing on the Webcam and DSLR domains. This dataset comprises images across ten classes, presenting diverse visual characteristics between the domains. We use the CaffeNet and Surf feature spaces of the data. By testing both methods under identical conditions, we aim to understand their relative strengths and contributions to unsupervised DA in visual classification tasks.

2. Related Work

(Fernando et al., 2013) introduce an unsupervised visual domain method for tasks where labeled data is available only in the source domain, with the target domain remaining unlabeled. They propose a subspace alignment technique that uses principal component analysis (PCA) to represent the source and target domains as subspaces. A transformation matrix aligns these subspaces, minimizing domain divergence. The approach includes a closed-form solution for efficient alignment matrix computation and introduces a similarity measure for nearest-neighbor classification, validated by a consistency theorem for selecting optimal dimensionality. Tested on datasets like Office, Caltech-256, and PASCAL VOC, the method outperforms existing techniques like Geodesic Flow Kernel (GFK), demonstrating significant improvements in accuracy and domain alignment. The approach is computationally efficient and adaptable, making it a promising tool for unsupervised DA, with potential applications in large-scale and real-world scenarios.

(Khamis et al., 2024) discuss optimal transport from its origins with the Earth Mover's problem and various changes and modifications that were done to obtain a closed-form solution. They also discuss various formulations of optimal transport such as Monge and Kantorovich formulations, Regularized optimal transport, Unbalanced and Partial optimal transport, Sliced optimal transport, and Gromov-Wasserstein (GW) optimal transport.

3. Our Experiments

In this work, we have access to the target domain labels and use them to evaluate and compare the methods, thus our approaches are considered supervised. n_s is the number of elements in the source data, n_t is the number of elements in the target data, and D is the original number of features.

3.1. Subspace Alignment

The subspace alignment method aims to project the labeled source samples S (an $n_s \times D$ matrix) and unlabeled target samples T (an $n_t \times D$ matrix) into two subspaces defined by their respective principal components, and minimizing the divergence between the two domains. The pipeline of the procedure we followed to implement and analyze this method is as follows:

3.1.1. STEP 1:

First, we loaded the source and target datasets and preprocessed them using the z-score normalization function. Afterwards, we performed principal component analysis (PCA) on both the source and target data to reduce their dimensionality and get the projected data in R^d: $\hat{S} = SX_s$ and $\hat{T} = TX_t$.

3.1.2. STEP 2:

We retrieved the d principal components from source and target data which represent the direction of maximum variance in the data for each domain, X_s and X_t respectively, and computed the optimal alignment matrix $M = X_s^T X_t$.

3.1.3. STEP 3:

We applied the alignment matrix to the projected source data to obtain the target aligned source domain $S_p = \hat{S}M$.

3.1.4. STEP 4:

We implemented a $1 - NN$ classification to predict labels of the target domain based on transformed source data, and evaluated it.

3.1.5. STEP 5:

We fine-tuned d as a hyper-parameter to get the number of components yielding the highest accuracy, by performing a 5-fold cross-validation.

3.1.6. STEP 6:

As we had two different representations of our data (CaffeNet and Surf), with two domains of interest for us in each, we performed the aforementioned pipeline for each combination, evaluated each of them, and visualized the distribution of source and target data before and after applying the alignment of the source domain on the target. Figures 1, 2, 3, and 4 represent these visualizations.

Table 1 shows the comparison of the accuracy of each combination with a 1-NN classifier trained on the raw source data. By looking at the results, we can clearly observe an improvement in accuracy after applying subspace alignment as our domain adaptation technique. This improvement is more apparent when we apply the method on the Surf feature representation, as we observe a drastic increase in the accuracy measures. With regard to d, we can observe that the best number of components differ for each combination.

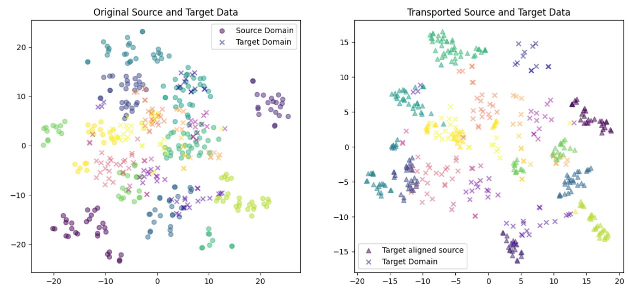

Figure 1. CaffeNet with Webcam as source and DSLR as target

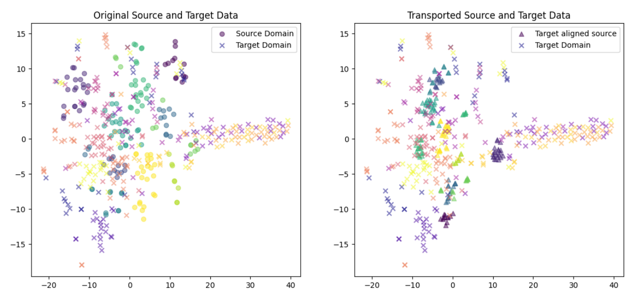

Figure 2. CaffeNet with DSLR as source and Webcam as target

Figure 3. Surf with Webcam as source and DSLR as target domain

3.2. Entropic regularized optimal transport

In this set of experiments, we consider the data from Webcam to be the source and the data from DSLR to be the target. With optimal transport, we compute the coupling matrix using the Sinkhorn-Knopp algorithm. We start by computing the cost matrix M between the source and target data by taking the Euclidean distance between them and

Table 1. Comparison of accuracies over different feature representations and a baseline 1NN

REPRESENTATION	SOURCE	TARGET	BEST d	BEST ACCURACY	BASELINE ACCURACY
CAFFENET	WEBCAM	DSLR	30	100%	96.18%
CAFFENET	DSLR	WEBCAM	50	73.22%	63.39%
SURF	WEBCAM	DSLR	100	90.45%	43.31%
SURF	DSLR	WEBCAM	50	83.39%	33.22%

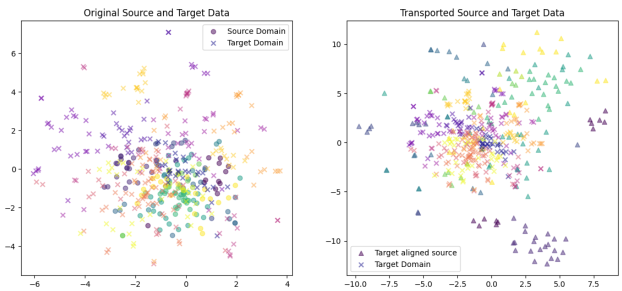

Figure 4. Surf with DSLR as source and Webcam as target domain

normalizing it by dividing with the maximum element of M. Then we draw the boundary conditions for the data sets using uniform distribution. The coupling matrix γ is calculated using the boundary conditions, cost matrix, and an entropic regularization parameter. The dot product between γ and the target gives the transported source. A 1NN is trained on the transported source and evaluated on the target.

Algorithm 1 Optimal Transport with Sinkhorn-Knopp Algorithm

Step 1: Compute the Cost Matrix
$M_{i,j} = cdist(X_s, X_t, metric =' euclidean')$
$M = \frac{M}{\max(M)}$
Step 2: Define Boundary Conditions
$a =$ Uniform distribution over X_s
$b =$ Uniform distribution over X_t
Step 3: Compute the Coupling Matrix γ using Sinkhorn-Knopp
$\gamma = ot.sinkhorn(a, b, M, reg_e))$
Step 4: Transport the Source Data
$S_a = \gamma \cdot X_t$
Step 5: Train and Evaluate 1-Nearest Neighbor (1-NN)
train a 1-NN classifier on S_a and evaluate the classifier using X_t as test data.

While using the above algorithm in CaffeNet feature space, the accuracy of the classifier after optimal transport is 0.96. Training a 1NN directly on the source data and then evaluating it on the target data also gave an accuracy of 0.96. This can be accounted for by the graphs Figure 5 that the source and transported source are distributed similarly. The entropic regularization constant (reg_e) was fine-tuned using cross-validation and the best accuracy was obtained when the reg_e value was 0.27. During the cross-validation, the accuracy increases with the increase in the values of reg_e and remains constant after some time.

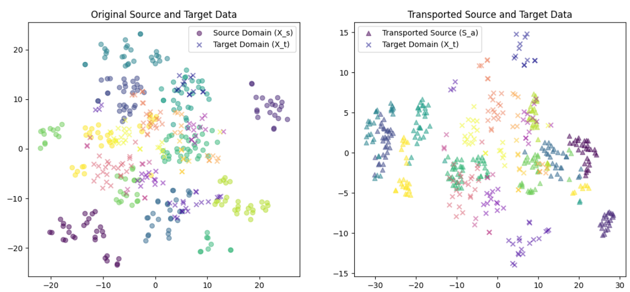

Figure 5. The visualization of source, target, and transported source for CaffeNet feature space.

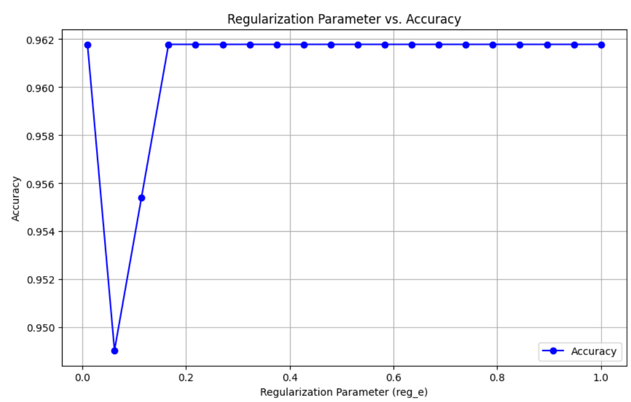

Figure 6. Linear graph showing the change in accuracy with values of reg_e for CaffeNet feature representation.

In the case of the Surf feature representation, the accuracy of the classifier after optimal transport is 0.87. Training a 1NN directly on the source data and then evaluating it on the target data gave an accuracy of 0.43. Similar to the other representation, the reg_e was fine-tuned using cross-validation, and the best accuracy was obtained when the reg_e value was 0.01. During the cross-validation, the accuracy decreases with the increase in the values of reg_e and remains constant after some time.

Figure 7. The visualization of source, target, and transported source for Surf feature space.

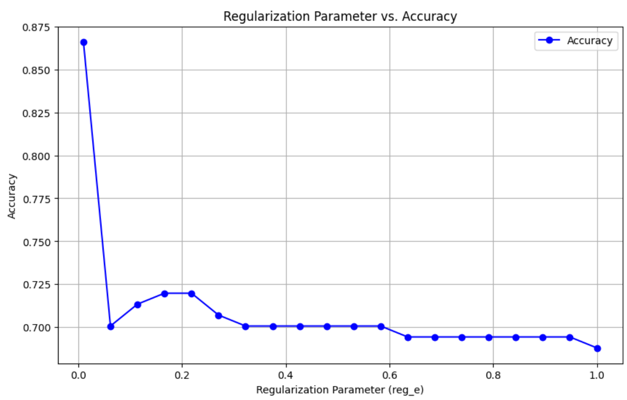

Figure 8. Linear graph showing the change in accuracy with values of reg_e for Surf feature representation.

Next, we consider the data from DSLR to be the source and the data from Webcam to be the target. We follow the same procedure. While using the above algorithm in CaffeNet feature space, the accuracy of the classifier after optimal transport is 0.2, and training a 1NN directly on the source data and then evaluating it on the target data also gave an accuracy of 0.63. While for the Surf feature space, the accuracies are 0.80 with optimal transport and 0.33 without optimal transport.

With the results obtained, we can see that the optimal transport shows a great improvement in the accuracy in case of Surf feature representation. From Figure 7 we can see that the source distribution is closer to the target distribution. In case of CaffeNet, Figure 5 shows the transportation of the source has not made any significant change to it, thus the consistency in the accuracy. In the case where the source and target are switched, the accuracy worsens for CaffeNet(Figure 9) while it improves for Surf (Figure 10) (Figures provided in the appendix).

Figure 9. The visualization of source, target, and transported source for CaffeNet feature space. Source = DSLR, Target = Webcam.

Figure 10. The visualization of source, target, and transported source for Surf feature space. Source = DSLR, Target = Webcam.

4. Conclusion

In this project, we evaluated subspace alignment and entropic regularized optimal transport for domain adaptation on the Office/Caltech dataset across Webcam and DSLR domains using CaffeNet and Surf feature spaces. Our results show that in the CaffeNet feature space, subspace alignment performs well, while optimal transport offers no noticeable accuracy improvement, likely due to minimal shift between source and transported distributions. However, in the Surf feature space, both methods provide significant improvements over the baseline. This comparison underscores that feature representation significantly impacts the effectiveness of domain adaptation techniques, with both methods proving valuable in performing domain adaptation in visual classification tasks.

Acknowledgments

We would like to express our gratitude to our professor, Prof. Marc Sebban, for his valuable guidance and support throughout this project. We also thank our classmates for their helpful feedback and encouragement, which made this experience more enjoyable and motivated us to do our best.

References

Fernando, B., Habrard, A., Sebban, M., and Tuytelaars, T. Unsupervised visual domain adaptation using subspace alignment. In *2013 IEEE International Conference on Computer Vision*, pp. 2960–2967, 2013. doi: 10.1109/ICCV.2013.368.

Khamis, A., Tsuchida, R., Tarek, M., Rolland, V., and Petersson, L. Scalable optimal transport methods in machine learning: A contemporary survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2024. doi: 10.1109/TPAMI.2024.3379571.

A. Appendix

A.1. Optimal Transport with Source and Target Switched

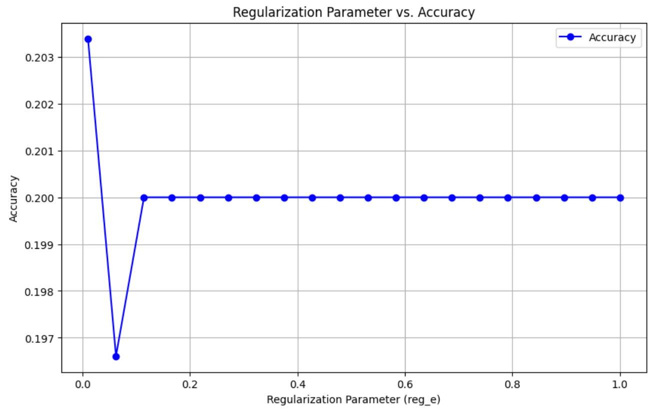

Figure 11. Linear graph showing the change in accuracy with values of reg_e for CaffeNet feature representation. Source = DSLR, Target = Webcam.

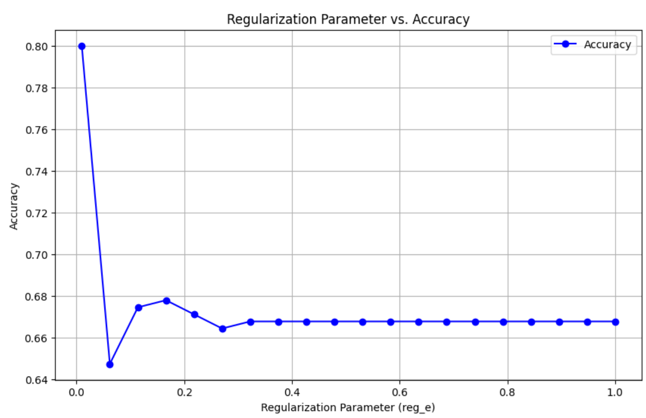

Figure 12. Linear graph showing the change in accuracy with values of reg_e for Surf feature representation. Source = DSLR, Target = Webcam.