Domain Adaptation over the Office/Caltech Dataset Using Subspace Alignment and Entropic Regularized Optimal Transport

Alejandro Carvajal [1]

Abstract

This project explores two domain adaptation techniques (Subspace Alignment and Entropic Regularized Optimal Transport) for image classification on the Office/Caltech dataset, focusing on the Webcam and DSLR domains to address real-world domain shifts in visual tasks. Subspace Alignment minimizes divergence by projecting source and target data onto their principal component spaces, while Entropic Regularized Optimal Transport uses the Sinkhorn algorithm with a tunable regularization parameter to align domains. Both methods were implemented and tuned, with performance evaluated via 1-Nearest Neighbor classification. Results indicate that while both techniques enhance domain shift robustness, Entropic Regularized Optimal Transport achieved better accuracy, demonstrating its suitability for practical applications. This study discusses the trade-offs in accuracy, parameter sensitivity, and computational demands, suggesting future exploration of adaptive tuning strategies.

1. Introduction

Domain adaptation is a real-world challenge in machine learning, especially for computer vision tasks where data distributions differ significantly across domains. Traditional models struggle to generalize when new samples have distinct characteristics from those seen during training. Common tasks such as object recognition across different lighting conditions, angles, or device sources require methods that can bridge the gap between training (source) and testing (target) data distributions (Pan & Yang, 2009).

In this project, we explore two domain adaptation techniques, Subspace Alignment and Entropic Regularized Optimal Transport, to address the domain shift problem in image classification on the Office/Caltech dataset. This dataset is usually used for benchmarking domain adaptation algorithms. The domain shift problem offers a challenging yet practical test case for evaluating the accuracy of the domain adaptation methods in transfer learning across different visual settings.

Subspace Alignment aligns source and target domains by projecting them into subspaces using PCA and a specified number of components for both datasets, minimizing the divergence through alignment matrices (Fernando et al., 2013). In contrast, Entropic Regularized Optimal Transport aligns distributions by optimizing a transport plan, using the Sinkhorn algorithm with entropic regularization to balance fidelity to data structure with computational efficiency (Courty et al., 2014).

The purpose of this study is: First, to evaluate each method's effectiveness in adapting the domain shift between two datasets with different data distributions (Web-cam and DSLR datasets); Second, to compare their performance and computational demands in the classification task. By focusing on parameter tuning and classification accuracy, we demonstrate how each technique addresses domain adaptation challenges.

2. Dataset description

The Office-Caltech dataset is widely used in domain adaptation research and includes images from four distinct domains: Amazon (online product images), DSLR (high-quality DSLR camera images), Webcam (lower-quality images), and Caltech (a collection of images representing various object categories). This setup simulates real-world applications where data from different sources exhibit significant variability (Saenko et al., 2010).

For this project, we used two variations of the dataset: SURF and CaffeNet4096, which provide distinct numerical representations of the image data. The SURF representation extracts image features using Speeded Up Robust Features (SURF), a technique known for capturing local keypoints and creating feature vectors based on their spatial distribution (Bay et al., 2008). The CaffeNet4096 is a represen-

[1]Faculty of Sciences, Université Jean Monnet, Saint-Étienne, France. Correspondence to: Alejandro Carvajal <alejandro.carvajal.montealegre@etu.univ-st-etienne.fr>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

tation of a pre-trained CaffeNet model, which is a dense, high-dimensional (4096-dimensional) feature space based on hierarchical representations of the image content (Jia et al., 2014). This enables it to capture a more complex set of features and relationships among them.

3. Methodology

This project explores two domain adaptation techniques, Subspace Alignment and Entropic Regularized Optimal Transport, to overcome the domain shift challenge in image classification. Each method approaches domain alignment differently. While Subspace Alignment emphasizes alignment in lower-dimensional subspaces, Entropic Regularized OT optimizes transport across domains using regularized cost minimization. This section details the workflow and theoretical foundation of each method.

3.1. Subspace Alignment (SA)

The Subspace alignment method aims to project the labeled source S ($n_s \times D$ matrix) and unlabeled target samples T ($n_t \times D$ matrix) into two subspaces spanned by their principal components, so that the divergence between the two domains is minimized.

Let the source samples S and target samples T be defined as:

$$S \in \mathbb{R}^{n_s \times D}, \quad T \in \mathbb{R}^{n_t \times D}$$

3.1.1. Optimal dimensionality selection

For calculating the alignment matrix M between the two datasets, first we need to calculate the principal components with the highest variance that both reduced datasets can have without overfitting. These are denoted as $X_s \in \mathbb{R}^{D \times d}$ for the source and $X_t \in \mathbb{R}^{D \times d}$ for the target.

(Fernando et al., 2013). establishes that we can deduce a bound on the deviation between two successive eigenvalues. We can make use of this bound as a cutting rule for automatically determining the size of the subspaces.

The previous work by (Fernando et al., 2013) determines that given a confidence $\delta > 0$ and a fixed deviation $\gamma > 0$, we can select the maximum dimension d_{max} such that:

$$\left(\lambda_{\text{min}}^{d_{\text{max}}} - \lambda_{\text{min}}^{d_{\text{max}}+1}\right) \geq \left(1 + \sqrt{\frac{\ln 2/\delta}{2}}\right)\left(\frac{16 d^{3/2} B}{\gamma \sqrt{n_{\text{min}}}}\right)$$

For each $d \in \{d \mid 1, \ldots, d_{\text{max}}\}$, we then have the guarantee that as long as we select a subspace dimension d such that $d \leq d_{\text{max}}$, the solution is stable and not over-fitting.

The values of the parameters γ, δ, and n_{min} were selected as follows:

- $\gamma = 10^5$: This is the same value used in the paper for adaptation from W to C. Although we are adapting from W to D, this value provided the best results.
- $\delta = 0.1$: This value represents a 90% confidence interval $(1-\gamma)$ for detecting significant shifts in eigenvalue differences. It is also consistent with the value used in the paper.
- n_{min}: This is the minimum sample size of the two datasets, X_{webcam} and X_{dslr}, ensuring that the stability criterion applies to both domains.

Different values were tested in this part, but the result for d_{max} showed little variation. Therefore, we decided to adopt the same values used in the paper. Ultimately, the maximum number of dimensions calculated was $d_{\textbf{max}} = 156$ for both the SURF and CaffeNet4096 datasets, which originally contain 800 and 4096 features, respectively.

After finding $d_{\textbf{max}}$ value, we perform *2-fold cross-validation* using a *1-NN* classifier over the reduced source dataset to evaluate the accuracy of each method on the target domain over 30 random trials. The paper establishes 20 randoms trials. However, when using this value, the optimal dimension d^* varied significantly between runs. Using 30 random trials provided more stable results. The result obtained were $d^* = 18$ for Surf dataset and $d^* = 37$ for CaffeNet4096 dataset.

Then, using the optimal dimensionality d^*, we apply PCA to both $X_{\textbf{webcam}}$ and $X_{\textbf{dslr}}$, resulting in the final reduced datasets X_s and X_t with shapes (n_s, d^*) and (n_t, d^*), respectively.

3.1.2. *1-NN* Algorithm in the aligned subspace

After determining the optimal number of dimensions d^*, we project the source data S and target data T into their respective subspaces X_s and X_t. Following the procedure specified in the paper, this operation is performed as $\hat{S} = SX_s$ and $\hat{T} = TX_t$.

Now, we we learn a linear transformation function that align the source subspace coordinate system to the target one using the subspace alignment approach. The basis vectors are aligned by using a transformation matrix M from X_s to X_t with the following formulation.

$$M = X_s^T X_t$$

After determining the optimal alignment matrix M, we compare the source data with the target data using a similarity function. The paper suggests a PSD approach that involves

projecting the source data $\hat{S}$ into the aligned target space $\hat{T}$ using the transformation: $S_p = \hat{S}M$.

Finally, the projected source data in the target-aligned subspace S_p is used to train a *1-NN* classifier with the original source labels $y_{\textbf{webcam}}$. The predictions are then compared with the actual labels of the target dataset $y_{\textbf{dslr}}$, to calculate the accuracy.

3.2. Entropic Regularized OT

Entropic regularized optimal transport is an approach to optimal transport (OT) that adds an entropy term to the classic OT problem to make it more computationally efficient and stable. The Sinkhorn algorithm is an iterative method to solve the entropic regularized OT problem by iteratively updating the transport matrix M between two data matrices $S \in \mathbb{R}^{n_s \times d}$ and $T \in \mathbb{R}^{n_t \times d}$ through matrix scaling steps.

3.2.1. Uniform vectors

The vector $\mathbf{a} \in \mathbb{R}^{n_s}$ and $\mathbf{b} \in \mathbb{R}^{n_t}$ represents the uniform vectors for source and target dataset, where n_s and n_t are the number of observations of S and T respectively.

For this project, the chosen value for the uniform vectors are: $\mathbf{a} = \left[\frac{1}{n_s}, \frac{1}{n_s}, \ldots, \frac{1}{n_s}\right]^T \in \mathbb{R}^{n_s}$ and $\mathbf{b} = \left[\frac{1}{n_t}, \frac{1}{n_t}, \ldots, \frac{1}{n_t}\right]^T \in \mathbb{R}^{n_t}$.

This ensures that the vectors can be interpreted as probability distributions, with each observation being equally likely.

3.2.2. Cost Matrix, Coupling Matrix, and **1-NN** Classification

The cost matrix M is defined as the pairwise Euclidean distance between the points in the source dataset S and the target dataset T. Mathematically, this can be expressed as:

$$M_{ij} = \|S^i - T^j\|^2$$

where S^i is the i^{th} observation from the source dataset, and T^j is the j^{th} observation from the target dataset. The resulting matrix $M \in \mathbb{R}^{n_s \times n_t}$ is then scaled for for subsequent computations. This operation was done using the command *scipy.spatial.distance.cdist*

The coupling matrix γ is a matrix that represents the optimal way to transport mass from a source distribution a to a target distribution b. For this case, γ was calculated using the command $ot.sinkhorn(a, b, M, rege)$

The parameter $rege$ is a regularization parameter that balances the trade-off between minimizing the transportation cost and maintaining the smoothness of the coupling matrix. According to (Cuturi, 2013), lower values of the hyperparameter provides more accurate mappings, but potentially increase the computational load. The paper indicates that a suitable range of values for small datasets like the ones used in this project can vary between 0.001 and 1. In this case, we used the values rege $= [0.001, 0.01, 0.1, 1]$ in a *2-fold cross-validation* framework, applying a *1-NN* classifier on the source dataset across 20 random trials. The optimal values obtained were rege $= 0.1$ for the Surf dataset and rege $= 0.01$ for the CaffeNet dataset.

After tuning the hyperparameter $rege$, we proceed to transport the points from S to T using the coupling matrix, as shown in the equation: $S_a = \gamma T$.

For this final step, we train a *1-NN* classifier using the transported points S_a as features and the labels from $y_{\textbf{webcam}}$. We then make predictions on the target dataset T, and finally, we calculate the accuracy by comparing the actual target labels $y_{\textbf{webcam}}$ with the predicted labels $y_{\textbf{dslr}}$

4. Results

In this section, we present the results obtained from applying two approaches *Subspace Alignment* (SA) and *Entropic Regularized Optimal Transport* to the Office/Caltech dataset. The two representations used in the evaluation are SURF and CaffeNet. Both datasets were taken into account during the assessment, and the following results reflect their performance. The cross-validation accuracy is computed over the source dataset, while the accuracy metrics and hyper-parameters tuned are on the subspace-aligned and transported source points. Additionally, we provide the accuracy without any domain adaptation to illustrate the baseline performance of the models. These results obtained are the following

Metric	**SURF**	**CaffeNet**
d_{max}	156	156
d^*	18	37
Cross-val accuracy	84.68%	100%
Accuracy with SA	78.98%	100%
Accuracy Raw Data	18.47%	8.28%

Table 1. Results of Subspace Alignment

Metric	**SURF**	**CaffeNet**
λ_{reg}	0.1	0.1
Cross-val accuracy	37.87%	97.03%
Accuracy	80.25%	97.03%
Accuracy Raw Data	18.47%	8.28%

Table 2. Results of Entropic Regularized Optimal Transport

The results from the *Subspace Alignment* (SA) and *Entropic*

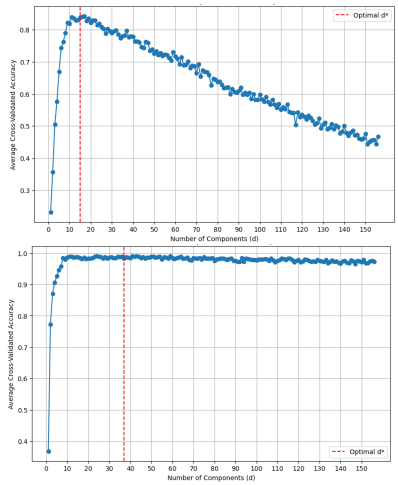

Figure 1. Cross-validation accuracy across different dimensions d for (a) SURF and (b) CaffeNet representations.

Regularized Optimal Transport (OT) approaches highlight the performance of two representations of the Office/Caltech dataset. First, we observe that CaffeNet achieved 100% accuracy across both cross-validation on the source dataset and accuracy with the transported source points in both approaches, indicating a possible overfitting issue. In comparison, the SURF representation achieved an accuracy of 75.80% with the SA approach and 80.25% with the regularized OT method. The dimensions (d^*) in the SA approach differ significantly between the two datasets due to the nature of the feature extraction methods; SURF yields lower-dimensional representations compared to CaffeNet's deep learning approach, which captures more complex features. For the Entropic OT regularization parameter *rege*, the best result obtained was 0.1 for both datasets. This consistency indicates that dataset differences do not affect the calculation of this hyper-parameter, unlike in the *Subspace Alignment* (SA) approach.

4.1. Optimal dimensionality in SA

The *figure 1.* illustrates the accuracy progression for different values of d in the *Subspace Alignment* approach across the SURF and CaffeNet representations. The objective of these plots is to visualize the relationship between the number of components (d) and the achieved cross-validated accuracy in the source dataset, allowing us to identify the optimal dimensionality, d^*, where the maximum accuracy is reached.

For the SURF dataset, the accuracy reaches its peak at d^* and subsequently decreases as d continues to increase, indicating that higher dimensions beyond d^* may introduce noise or overfitting. In contrast, the CaffeNet representation maintains its maximum accuracy after reaching d^*, remaining constant up to $d_{\max}$. This stability in CaffeNet reflects its capacity to capture and retain key features across additional dimensions.

5. Conclusions

- Both Subspace Alignment (SA) and Entropic Regularized OT approaches greatly improved accuracy over data without projection. They achieved similar accuracy gains in the SURF dataset. Regularized OT had faster computation, which is advantageous when handling large datasets.
- CaffeNet achieved an accuracy of 100% in both aproaches. This result suggests potential overfitting due to CaffeNet's high-capacity feature extraction, which may capture dataset-specific patterns that do not generalize well across domains.

References

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

Courty, N., Flamary, R., and Tuia, D. Domain adaptation with regularized optimal transport. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pp. 274–289. Springer, 2014.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Fernando, B., Habrard, A., Sebban, M., and Tuytelaars, T. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE international conference on computer vision*, pp. 2960–2967, 2013.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, 2014.

Pan, S. J. and Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10): 1345–1359, 2009.

Saenko, K., Kulis, B., Fritz, M., and Darrell, T. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pp. 213–226. Springer, 2010.