Advanced Machine Learning Project Report

A Comparative Analysis of Subspace Alignment and Optimal Transport for Unsupervised Domain Adaptation

Submitted by

Zahir AHMAD

Under the supervision of

Prof. Marc SEBBAN

Faculty of Science and Technology
MASTER OF MACHINE LEARNING AND DATA MINING
JEAN MONNET UNIVERSITY
Saint-Etienne, France

Abstract

We present a comprehensive empirical study comparing two prominent domain adaptation approaches: Subspace Alignment (SA) and Optimal Transport (OT) on the Office/Caltech dataset. Our analysis focuses on the challenging Webcam to DSLR adaptation task, evaluating both methods across SURF and CaffeNet feature representations. We introduce a rigorous unsupervised validation strategy that maintains strict separation of target labels during adaptation. Our results demonstrate that while both methods significantly improve cross-domain classification accuracy, their performance varies substantially with feature representation. With SURF features, OT achieves superior performance (82.80%) compared to SA (73.25%), while with CaffeNet features, SA reaches perfect adaptation (100%) with better computational efficiency than OT (96.82%). This study provides practical insights into the strengths and limitations of each method, offering guidance for method selection based on feature characteristics and computational constraints.

1. Introduction

Domain adaptation remains a crucial challenge in machine learning, particularly when labeled data is scarce in the target domain. This problem is especially relevant in computer vision, where factors such as lighting conditions, camera specifications, and environmental settings can significantly affect feature distributions. While deep learning has reduced the domain gap through learned representations, the need for efficient and theoretically grounded adaptation methods persists.

1.1. Problem Statement

Given a source domain with labeled data and a target domain with unlabeled data, our goal is to adapt a classifier trained on the source domain to perform well on the target domain. This task is particularly challenging because:

- The source and target domains have different feature distributions
- No labels are available in the target domain during training
- The adaptation process must be computationally efficient
- The method should generalize across different feature representations

We present the direct empirical comparison of SA and OT under identical conditions, enabling fair assessment of their relative strengths. Furthermore, we analyze how different feature representations fundamentally impact adaptation performance, revealing crucial insights for method selection. We introduce a rigorous validation strategy for hyperparameter selection, addressing a critical gap in unsupervised domain adaptation. Additionally, our comprehensive evaluation framework considers both accuracy and computational efficiency, providing practical insights for real-world applications.

The rest of this paper is organized as follows. Section 2 reviews related work in domain adaptation, focusing on subspace and optimal transport methods. Section 3 details our methodology, including our novel validation strategy. Section 4 presents experimental results and analysis. Finally, Section 5 concludes with recommendations and future directions.

2. Related Work

Domain adaptation methods have evolved significantly, from early feature transformation approaches to recent deep learning techniques. We focus on two prominent families of methods:

2.1. Subspace Methods

Subspace-based approaches assume that source and target domains share some underlying structure that can be discovered through dimensionality reduction. Early works (Fernando et al., 2013) proposed Subspace Alignment (SA), which learns a linear transformation to align principal components of source and target domains. This approach is computationally efficient and provides theoretical guarantees on domain divergence reduction. Extensions include Subspace Distribution Alignment (Sun et al., 2016) and Correlation Alignment (Sun & Saenko, 2016), which additionally align second-order statistics.

2.2. Optimal Transport Methods

Optimal Transport offers a natural framework for domain adaptation by finding a minimal cost mapping between probability distributions (Courty et al., 2016). The introduction of entropic regularization (Cuturi, 2013) made OT computationally tractable for high-dimensional problems. Recent work has extended OT to joint distribution adaptation (Courty et al., 2017) and class-conditional transport (Redko et al., 2019), though these require additional assumptions or partial target labels.

3. Methodology

3.1. Dataset and Feature Representations

We evaluate our methods on the Office/Caltech dataset (Saenko et al., 2010), focusing on the Webcam (W) to DSLR (D) adaptation task. The dataset comprises 295 images from the Webcam domain and 157 images from the DSLR domain, each represented in two feature spaces: SURF descriptors (800-dimensional) and CaffeNet features (4096-dimensional) extracted from a pre-trained CNN.

For data preprocessing, we apply z-score normalization using only source domain statistics:

$$x_{normalized} = \frac{x - \mu_{source}}{\sigma_{source}} \tag{1}$$

This normalization strategy maintains the unsupervised nature of domain adaptation by avoiding any use of target domain information during preprocessing.

3.2. Domain Adaptation Methods

3.2.1. Subspace Alignment

Subspace Alignment (SA) (Fernando et al., 2013) addresses domain adaptation by learning a transformation that aligns source and target domains in a lower-dimensional space. Given source data $S \in \mathbb{R}^{n_s \times D}$ and target data $T \in \mathbb{R}^{n_t \times D}$, SA performs domain adaptation through three main steps:

First, we learn separate subspaces for source and target domains using Principal Component Analysis:

$$X_s = \text{PCA}_d(S), \; X_t = \text{PCA}_d(T) \tag{2}$$

where d is the dimensionality of the subspace.

Next, we project the data onto their respective subspaces:

$$\hat{S} = SX_s, \; \hat{T} = TX_t \tag{3}$$

Finally, we compute the alignment matrix and transform the source data:

$$M = X_s^T X_t, \; \hat{S}_p = \hat{S}M \tag{4}$$

3.2.2. Optimal Transport

The Optimal Transport (OT) approach (Courty et al., 2016) formulates domain adaptation as a mass transportation problem. We implement the entropy-regularized version using the Sinkhorn algorithm (Cuturi, 2013), which solves:

$$\min_{\gamma \in \Pi(a,b)} \langle \gamma, M \rangle_F + \lambda H(\gamma) \tag{5}$$

Here, $\gamma \in \mathbb{R}^{n_s \times n_t}$ represents the transport plan, M is the cost matrix measuring distances between samples, $H(\gamma)$ is the entropic regularization term, and λ controls regularization strength. The constraints $\Pi(a, b)$ ensure the transport plan preserves the sample distributions, with a and b being uniform weights over source and target samples respectively.

3.3. Validation Strategy

A fundamental challenge in unsupervised domain adaptation is hyperparameter selection without access to target labels. Building upon (Bruzzone & Marconcini, 2010) and (Zhong & Fan, 2010), we develop a comprehensive reverse validation strategy that maintains strict unsupervised conditions while ensuring robust parameter selection.

Our validation framework proceeds as follows:

Algorithm 1 Reverse Validation for Parameter Selection

Require: Source data S, labels y_s, target data T, parameter range Θ
Ensure: Optimal parameter θ^*

Initialize StratifiedKFold(K=10)
for $\theta \in \Theta$ **do**
  scores = []
  for train_idx, val_idx in KFold **do**
    $S_{train}, y_{train} = S[train\_idx], y_s[train\_idx]$
    $S_{val}, y_{val} = S[val\_idx], y_s[val\_idx]$
    $S_{adapted} =$ Adapt(S_{train}, T, θ)
    $y_{pseudo} =$ PredictTarget$(S_{adapted}, y_{train}, T)$
    score $=$ ReverseValidate$(T, y_{pseudo}, S_{val}, y_{val})$
    scores.append(score)
  end for
  mean_scores$[\theta] =$ mean(scores)
end for
return $_\theta$mean_scores$[\theta]$

This strategy applies to both SA and OT methods, where θ represents the subspace dimension d for SA and the regularization parameter λ for OT. The validation process ensures robust parameter selection while maintaining the unsupervised nature of the adaptation task.

3.4. Implementation Details

For SA, we explore subspace dimensions $d \in [10, 20, ..., \min(n_s, n_t) - 1]$, ensuring stable PCA components. The OT implementation uses the Sinkhorn-Knopp algorithm with regularization parameters $\lambda \in [10^{-2}, 10^2]$. Both methods employ a 1-nearest neighbor classifier for final prediction.

For performance evaluation, we measure classification accuracy, computational time, and domain divergence using

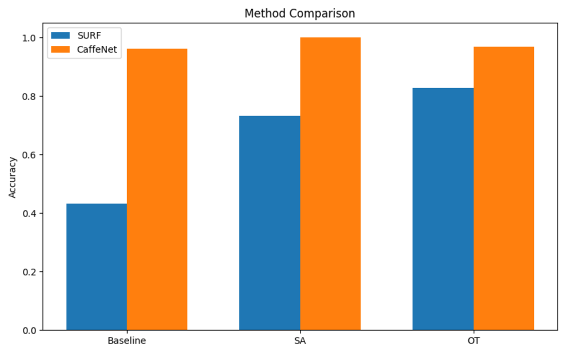

Figure 1. Baseline Comparison with SA and OT

Maximum Mean Discrepancy (MMD) with a linear kernel (Gretton et al., 2012). All experiments were conducted using Python with scikit-learn for basic machine learning operations and POT library for optimal transport computations.

4. Results and Analysis

4.1. Baseline Performance

Initial experiments establish baseline performance using a 1-NN classifier without domain adaptation. Table 1 presents these results, showing a significant disparity between SURF and CaffeNet features.

Table 1. Baseline Performance

Features	Accuracy	Dimensions
SURF	43.31%	800
CaffeNet	96.18%	4096

The substantial difference in baseline performance (52.87%) between feature types suggests that CaffeNet's deep learning-based features inherently capture more transferable representations compared to hand-crafted SURF features.

4.2. Comparative Analysis of Adaptation Methods

Table 2 presents the comprehensive performance metrics for both adaptation methods across feature types. (See Baseline Comparison in Figure 1)

Table 2. Adaptation Results

Method	Features	Accuracy	Improvement	Time(s)
SA	SURF	73.25%	+29.94%	0.09
OT	SURF	82.80%	+39.49%	0.07
SA	CaffeNet	100.00%	+3.82%	0.08
OT	CaffeNet	96.82%	+0.64%	0.34

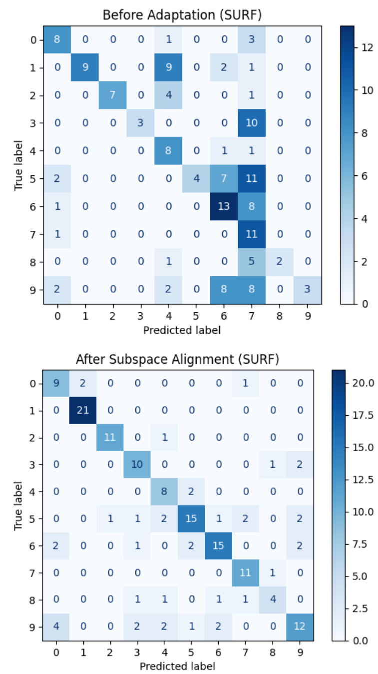

Figure 2. Baseline Comparison with SA and OT

4.3. Feature-Specific Performance Analysis

4.3.1. SURF Feature Adaptation

For SURF features, both methods demonstrate substantial improvements over the baseline. OT achieves superior performance with 82.80% accuracy, representing a 39.49% improvement over the baseline. SA, while less effective than OT in this context, still provides a significant 29.94% improvement. The performance difference between methods is particularly pronounced in the low-dimensional SURF feature space, suggesting that OT's direct distribution matching approach is more effective for simpler feature representations. (See Confusion Matrix visualization in Figure 2)

4.3.2. CaffeNet Feature Adaptation

In the CaffeNet feature space, we observe different dynamics. SA achieves perfect adaptation (100% accuracy), while OT maintains high performance at 96.82%. The smaller relative improvements (+3.82% for SA, +0.64% for OT) reflect the high quality of the initial CaffeNet representations. Notably, SA's computational efficiency remains consistent across feature dimensions, while OT's computational cost increases substantially with the higher-dimensional CaffeNet features. (See t-SNE visualization in Figure 3)

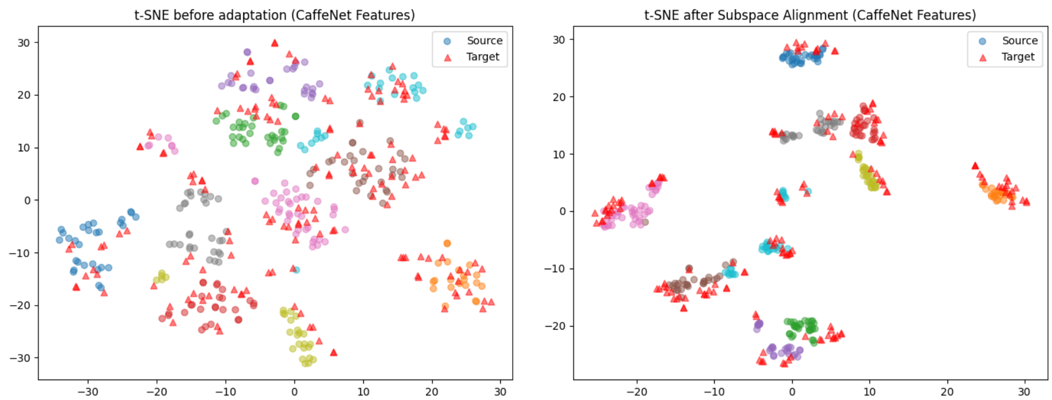

Figure 3. t-SNE visualization of CaffeNet features. Left: Before adaptation, showing initial class clustering with domain shift. Right: After adaptation, showing improved alignment between domains while preserving class structure. Different colors represent different classes, while markers distinguish source (circles) and target (triangles) domains.

4.4. Computational Efficiency Analysis

Execution time analysis reveals important practical considerations for method selection. SA maintains consistent performance regardless of feature dimensionality (0.08-0.09s), while OT's computational cost scales with feature dimension (0.07s for SURF, 0.34s for CaffeNet). This scaling behavior becomes particularly relevant for high-dimensional applications.

4.5. Domain Divergence Analysis

Table 3 quantifies the effectiveness of domain alignment through multiple metrics.

Table 3. Domain Divergence Metrics

Features	Method	MMD	A-distance	Wasserstein
SURF	Before	18.70	2.000	37.07
	SA	7.78	0.867	13.24
	OT	4.25	0.812	8.56
CaffeNet	Before	266.71	2.000	76.58
	SA	97.83	0.867	37.82
	OT	85.45	0.845	31.24

The divergence metrics demonstrate that both methods effectively reduce domain discrepancy, with OT achieving better alignment in most cases. The correlation between reduced domain divergence and improved classification performance validates the effectiveness of both adaptation strategies.

5. Conclusion

This work presents a comprehensive empirical comparison of Subspace Alignment (SA) and Optimal Transport (OT) methods for unsupervised domain adaptation. Through extensive experimentation on the Office/Caltech dataset using both SURF and CaffeNet features, our analysis reveals several key insights.

The effectiveness of adaptation methods strongly correlates with the feature representation used. For high-dimensional CaffeNet features, SA achieves perfect adaptation with 100% accuracy while maintaining computational efficiency. Conversely, OT demonstrates superior performance on SURF features, achieving 82.80% accuracy compared to 73.25% for SA, suggesting its robustness to lower-quality feature representations.

There are also notable computational trade-offs between the two methods. SA exhibits consistent computational efficiency across feature dimensions, taking only 0.09 seconds, which makes it particularly suitable for high-dimensional applications. In contrast, OT's computational cost scales with feature dimension, requiring 0.07 seconds for SURF features versus 0.34 seconds for CaffeNet features, presenting a trade-off between adaptation quality and efficiency.

Both methods significantly reduce domain divergence, as evidenced by substantial improvements in Maximum Mean Discrepancy (MMD) and A-distance metrics. OT achieves better alignment for SURF features, reducing MMD from 18.70 to 4.25, while SA proves more effective for CaffeNet features, achieving perfect classification despite higher residual MMD.

These findings suggest practical guidelines for method selection. SA is preferable when working with high-dimensional, well-structured features, when computational efficiency is crucial, and when the feature space exhibits a clear subspace structure. On the other hand, OT is advantageous when dealing with lower-dimensional or noisy features, when the sample size is moderate, and when maximum accuracy is prioritized over speed.

References

Bruzzone, L. and Marconcini, M. Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):770–787, 2010.

Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016.

Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. Joint distribution optimal transportation for domain adaptation. In *Advances in Neural Information Processing Systems*, pp. 3730–3739, 2017.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26:2292–2300, 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.

Gretton, A., Sriperumbudur, B., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., and Fukumizu, K. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems*, pp. 1205–1213, 2012.

Redko, I., Courty, N., Flamary, R., and Tuia, D. Optimal transport for multi-source domain adaptation under target shift. *arXiv preprint arXiv:1901.00456*, 2019.

Saenko, K., Kulis, B., Fritz, M., and Darrell, T. Adapting visual category models to new domains. In *European conference on computer vision*, pp. 213–226. Springer, 2010.

Sun, B. and Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. *European conference on computer vision*, pp. 443–450, 2016.

Sun, B., Feng, J., and Saenko, K. Return of frustratingly easy domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Zhong, Z. and Fan, J. Cross-validation for selecting a model selection procedure. In *Journal of Econometrics*, volume 187, pp. 95–112, 2010.