



Sparse Domain Transfer via Elastic Net Regularization

Jingwei Zhang¹  and Farzan Farnia¹ 

Department of Computer Science and Engineering, The Chinese University of Hong Kong

{jwzhang22, farnia}@cse.cuhk.edu.hk

Abstract. Transportation of samples across different domains is a central task in several machine learning problems. A sensible requirement for domain transfer tasks in computer vision and language domains is the sparsity of the transportation map, i.e., the transfer algorithm aims to modify the least number of input features while transporting samples across the source and target domains. In this work, we propose *Elastic Net Optimal Transport (ENOT)* to address the sparse distribution transfer problem. The ENOT framework utilizes the L_1 -norm and L_2 -norm regularization mechanisms to find a sparse and stable transportation map between the source and target domains. To compute the ENOT transport map, we consider the dual formulation of the ENOT optimization task and prove that the sparsified gradient of the optimal potential function in the ENOT’s dual representation provides the ENOT transport map. Furthermore, we demonstrate the application of the ENOT framework to perform feature selection for sparse domain transfer. We present the numerical results of applying ENOT to several domain transfer problems for synthetic Gaussian mixtures and real image and text data. Our empirical results indicate the success of the ENOT framework in identifying a sparse domain transport map. Code is available at github.com/buyeah1109/ENOT.

Keywords: Generative model · Domain transportation · Feature selection

1 Introduction

Deep neural networks (DNNs) have revolutionized the performance of computer vision models in domain transfer applications where the features of an input sample are altered to transfer the sample to a secondary domain [18, 30, 43]. The common goal of domain transfer algorithms is to transport an input data point to a target distribution by applying *minimal changes* to the input. Over recent years, domain transportation algorithms based on generative adversarial networks (GANs) including CycleGAN [42] and StyleGAN [20] have achieved empirical success in addressing the domain transfer task for image distributions. The success of these algorithms has inspired several studies of GAN-based domain transfer methodologies [7, 31, 41].

While the GAN-based methods have led to successful results in image-based domain transfer problems, their application demands significantly higher computational costs than standard GAN algorithms including only one generator/discriminator neural net pair to transfer a latent Gaussian vector to the data distribution. The extra computations in these domain transfer algorithms aim to ensure an invertible transfer map and thus limited modifications to an input sample. For example, the CycleGAN algorithm considers two pairs of generator/discriminator neural nets to impose a reversible transformation of an input image. However, the additional pair of neural nets in the CycleGAN setting will lead to a more challenging optimization task and higher training costs.

In this work, we focus on *sparse domain transfer problems* where the transfer of samples between source and target domains can be achieved by editing only a limited subset of input features. We note that the assumption of a sparse transport map applies to several real-world domain transfer problems, e.g. object translation, text revision, and gene editing problems. In the mentioned tasks, the sparsity level of the transfer map results in a meaningful measure of changes applied to an input sample. While sparse transportation maps are desired in many real-world domain transfer problems, the commonly-used GAN-based algorithms often lead to dense transfer maps editing a considerable fraction of input features.

To address sparse domain transfer tasks, we propose an optimal transport-based approach which takes advantage of the induced sparsity of the L_1 -norm regularization and the stability properties of the L_2 -norm regularization. Our proposed framework, which we call *Elastic Net Optimal Transport (ENOT)*, solves an optimal transport problem where the transportation cost follows from the elastic net function [44] combining standard Euclidean-norm-squared and L_1 -norm cost functions. Therefore, the ENOT approach can be interpreted as a mechanism to regularize the standard optimal transport map toward sparser transportation functions. By tuning the coefficient of the L_1 -norm regularization in ENOT’s elastic net cost, the learner can adjust the sparsity level of the transportation map and explore the spectrum between the standard and fully L_1 -norm-based optimal transport tasks.

To analyze the ENOT problem, we leverage optimal transport theory [37] and extend the duality results to the ENOT optimal transport setting. We prove a generalization of standard Brenier’s theorem, highlighting the connection between the optimal potential function in the ENOT’s dual problem with the optimal transport map transferring samples across domains. Our main theorem suggests that the composition of a soft-thresholding function with the gradient of the optimal potential function will perform sparse transportation across the domains. This result indicates that the ENOT framework offers a combination of the standard optimal transport problem with the squared-error cost function and the L_1 -norm-based optimal transport problem which leads to a challenging optimization problem without the L_2 -norm-based regularization in ENOT.

Furthermore, we utilize the ENOT framework to develop a *feature selection-based approach* to reduce the sparse domain transport task to a constraint-free

distribution transfer problem where an unconstrained transfer map is applied to only the selected feature subset. According to this variable selection-based approach, we break the domain transfer problem into two sub-problems: 1) ENOT-based variable selection choosing features to undergo modification for a given input sample, 2) Applying an unconstrained transportation map via standard GAN frameworks that transfers the input sample masked by the feature selection output to the target domain. By tuning the coefficient of the L_1 -norm regularization in the ENOT’s elastic net cost, the learner can adjust the number of selected features prior to performing a constraint-free GAN-based distribution transfer.

Finally, we discuss the numerical results of the applications of the ENOT framework to sparse domain transfer problems from various areas including computer vision, computational biology, and natural language processing. Our empirical findings show that the feature selection-based domain transfer via ENOT can be easily adapted to different domains and achieves satisfactory results. We qualitatively evaluate ENOT’s application for feature identification in sparse domain transfer. The numerical results support the proposed methodology of sparse domain transfer via ENOT-based domain transportation and feature selection. The contributions of this work can be summarized as:

- Proposing a feature selection-based approach for the sparse domain transfer problem,
- Developing ENOT as an elastic net-based methodology to the sparse domain transfer and variable selection,
- Extending the theory of standard squared-error-based optimal transport task to the ENOT setting,
- Providing supportive numerical results for applying ENOT-based sparse domain transfer to various domain transfer tasks.

2 Related Work

Sparsity and Optimal Transport Methods. Several related works have studied various notions of sparsity in optimal transport frameworks. References [2, 4, 13, 26, 35] propose sparsity-based regularization of the transportation matrix in optimal transport problems. However, we note that the sparsity objective pursued in these works differs from the sparse domain transfer in our work: while the mentioned papers aim for a sparse transportation matrix to gain a sparse alignment of source and target samples, our proposed ENOT method focuses on the sparsity of the modified input features in the domain transfer. Meanwhile, [22] proposed neural optimal transport and leverage neural networks to model potential functions and conduct image transportation, and considered the standard Euclidean-norm-squared cost function for the transportation, which does not focus on sparse transportation.

[8] first introduces an optimal transport-based approach to the sparse domain transfer problem, which aim to precisely solve an entropic-regularized optimal transport problem over the empirical samples and then use a kernel-based

interpolation to generalize the solution to unseen data. The main difference between [8]’s analysis and ours is the focus on the *primal vs. dual* formulations of the sparse optimal transport problem. The analysis in [8] concentrates on primal formulation and finds the precise solution to the primal OT problem on the training data. On the other hand, our approach targets the dual formulation and is more similar to the Wasserstein GAN framework, which involves a potential neural net function and can be extended to large-scale image and text data. Overall, our neural net-based method for the ENOT optimal transport problem can be viewed as a complementary approach to the precise kernel-based framework in [8]. Also, the trained neural net can be used as an efficient-to-compute feature selection map, which we later used to introduce a *feature selection-based approach* to sparse domain transfer, a topic that has not been studied in [8], which is useful for large-scale image and text-related applications.

Unsupervised Image to Image Translation (UI2I). Several related works attempt to address image-based transportation problems. For the image style transfer task, CycleGAN [42] uses a cycle-consistent loss and two GANs to conduct cyclic unpaired transformation. DRIT++ [24] adopts encoders to obtain the latent representation of images and similar cross-cycle consistency loss. For the image colorization transfer task, Conditional GANs are leveraged to improve colorization performance [17]. [40] propose a real-time user-guided neural network colorization. Moreover, cyclic-loss [3, 33, 38, 42] and GANs [7, 31, 41] have been utilized to address UI2I. However, unlike ENOT, these related works do not focus on the sparsity of transfer maps.

Sequence to Sequence Translation. Sequence to sequence (Seq2Seq) neural net models are typically designed based on an encoder-decoder architecture. [19] propose the application of a convolutional neural network (CNN) as the encoder and a recurrent neural network (RNN) as the decoder. [34] utilize an RNN-based architecture for both the encoder and decoder neural nets. [36] propose a transformer based on multi-head self-attention. BART [25] offers a sequence-to-sequence pretraining solution and adopts a bidirectional encoder similar to BERT [11], and a decoder similar to GPT [32]. Unlike our proposed ENOT approach, the discussed methods usually result in a dense transportation map. We also note that Seq2Seq and GAN-based transfer methods are almost exclusively used for language and image distributions, respectively.

3 Preliminaries

Consider random vectors $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^d$ with probability distributions P_X, P_Y , respectively. Given n independent samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ from P_X and m independent samples $\mathbf{y}_1, \dots, \mathbf{y}_m$ from P_Y , the goal in the domain transfer problem is to learn a map $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ transporting an input \mathbf{X} from distribution P_X to an output $\psi(\mathbf{X})$ distributed as P_Y , i.e.,

$$\psi(\mathbf{X}) \stackrel{\text{dist}}{=} \mathbf{Y}.$$

Here, $\stackrel{\text{dist}}{=}$ denotes identical probability distributions.

Without any constraint on the map ψ , there exist infinitely many transportation maps resulting in the required identical distributions. To uniquely characterize the transfer map, the optimal transport framework [37] seeks to find a map minimizing the expected transportation cost measured based on a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$. According to this framework, the transportation map follows from the optimal coupling $\Pi_{X,Y}$, marginally distributed as P_X and P_Y , that is minimizing the expected transportation cost formulated as

$$\text{OT}_c(P_X, P_Y) := \inf_{\substack{\Pi_{X,Y} : \Pi_X = P_X \\ \Pi_Y = P_Y}} \mathbb{E}_{(X,Y) \sim \Pi} [c(\mathbf{X}, \mathbf{Y})].$$

Here, $\text{OT}_c(P_X, P_Y)$ denotes the optimal transport cost between P_X, P_Y . It is well-known that under mild regularity conditions, a deterministic coupling mapping \mathbf{X} to a sample with distribution P_Y exists that solves the above problem. Also, the dual representation of the above optimization problem can be formulated via the Kantorovich duality [37] as

$$\sup_{\phi: \mathbb{R}^d \rightarrow \mathbb{R}} \mathbb{E}[\phi(\mathbf{X})] - \mathbb{E}[\phi^c(\mathbf{Y})],$$

where ϕ is the potential function and the c -transform ϕ^c is defined as $\phi^c(\mathbf{y}) := \sup_{\mathbf{y}' } \phi(\mathbf{y}') - c(\mathbf{y}, \mathbf{y}')$.

Example 1. In the special case of a norm cost $c_1(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, the result of Kantorovich duality can be written as

$$\text{OT}_{c_1}(P_X, P_Y) = \sup_{\phi: 1\text{-Lipschitz}} \mathbb{E}[\phi(\mathbf{X})] - \mathbb{E}[\phi(\mathbf{Y})]$$

where the potential function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is constrained to be 1-Lipschitz with respect to the assigned norm $\|\cdot\|$, i.e., for every $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$:

$$|\phi(\mathbf{x}) - \phi(\mathbf{x}')| \leq \|\mathbf{x} - \mathbf{x}'\|.$$

Example 2. In the special case of the L_2 -norm-squared cost $c_2(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2$, the result of Kantorovich duality can be written as

$$\sup_{\tilde{\phi}: \text{convex}} \mathbb{E} \left[\frac{1}{2} \|\mathbf{X}\|_2^2 - \tilde{\phi}(\mathbf{X}) \right] + \mathbb{E} \left[\frac{1}{2} \|\mathbf{Y}\|_2^2 - \tilde{\phi}^*(\mathbf{Y}) \right] \quad (1)$$

where the potential function $\phi(\mathbf{x}) := \frac{1}{2}\|\mathbf{x}\|_2^2 - \tilde{\phi}(\mathbf{x})$ is constrained to be the subtraction of a convex function $\tilde{\phi}$ from $\frac{1}{2}\|\mathbf{x}\|_2^2$, and $\tilde{\phi}^*$ is the Fenchel conjugate defined as

$$\tilde{\phi}^*(\mathbf{x}) := \sup_{\mathbf{x}'} \mathbf{x}'^\top \mathbf{x} - \tilde{\phi}(\mathbf{x}').$$

The Brenier theorem reveals that in the setting of Example 2, the gradient of the optimal solution $\tilde{\phi}$ provides the unique monotone (gradient of a convex function) map transporting samples between the two domains:

Theorem 1 (Brenier’s Theorem, [37]). *Suppose that P_X, P_Y are absolutely continuous with respect to one another. Then, the gradient of the solution $\tilde{\phi}^*$ to equation 1 is the unique monotone map for transferring P_X to P_Y , that is*

$$\nabla \tilde{\phi}^*(\mathbf{X}) \stackrel{\text{dist}}{=} \mathbf{Y}.$$

In the following sections, we aim to define and analyze optimal transport costs that can capture the sparsity of the transportation map, i.e. the number of non-zero coordinates of $\mathbf{y} - \mathbf{x}$.

4 Elastic Net Regularization for Sparse Optimal Transport

In this work, we aim to address the sparse domain transfer problem where the transfer map ψ between distributions P_X, P_Y alters the fewest possible coordinates in the d -dimensional feature vector $\mathbf{X} = [X^{(1)}, \dots, X^{(d)}]$. To apply the optimal transport framework, a proper cost function is the cardinality (number of non-zero elements $\text{card}(\mathbf{z}) = \sum_{i=1}^d \mathbf{1}[z_i \neq 0]$) of the difference between the original and transported samples:

$$c_{\text{sparse}}(\mathbf{x}, \mathbf{y}) = \text{card}(\mathbf{x} - \mathbf{y}).$$

Since the cardinality function lacks continuity and convexity, the resulting optimal transport problem will be computationally difficult. A common convex proxy for the cardinality function is the L_1 -norm where we simply use $c_{L_1}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1$. While the primal optimal transport problem could be solved for the empirical samples with the L_1 -norm cost, the domain transfer map requires solving the optimization problem for the data distribution which would be complex in the primal case. Therefore, we focus on the dual optimization problem to the optimal transport task. However, solving the dual optimization problem of the L_1 -norm cost requires optimizing over the L_1 -norm-based 1-Lipschitz functions which would be challenging.

To handle the computational complexity of the dual optimization problem with L_1 -norm cost function, we propose to apply the *elastic net* [44] cost function with coefficients $0 \leq \alpha \leq 1$ and $\lambda > 0$:

$$c_{\text{EN}}^{\alpha, \lambda}(\mathbf{x}, \mathbf{y}) = \lambda(1 - \alpha)\|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1. \quad (2)$$

Using the above cost function, we propose the *Elastic Net-based Optimal Transport (ENOT)* as the optimal transport method formulated with the cost function in equation 2. For the dual formulation of the ENOT problem, we can apply the Kantorovich duality to obtain the following optimization task:

$$\max_{\phi: \mathbb{R}^d \rightarrow \mathbb{R}} \mathbb{E}[\phi(\mathbf{X})] - \mathbb{E}[\phi_{\text{EN}}^{\alpha, \lambda}(\mathbf{Y})] \quad (3)$$

where the elastic-net-based c -transform can be written as follows:

$$\phi_{\text{EN}}^{\alpha, \lambda}(\mathbf{y}) := \max_{\boldsymbol{\delta} \in \mathbb{R}^d} \phi(\mathbf{y} + \boldsymbol{\delta}) - \lambda(1 - \alpha)\|\boldsymbol{\delta}\|_2^2 - \lambda\alpha\|\boldsymbol{\delta}\|_1.$$

Theorem 2. Consider the ENOT dual problem in equation 3. Then, there exists an optimal potential function ϕ^* for this problem which satisfies the following weakly-concavity property: for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and real value $\gamma \in [0, 1]$:

$$\begin{aligned} \phi^*(\gamma\mathbf{x} + (1 - \gamma)\mathbf{y}) &\geq \gamma\phi^*(\mathbf{x}) + (1 - \gamma)\phi^*(\mathbf{y}) \\ &\quad - \lambda\gamma(1 - \gamma)(1 - \alpha)\|\mathbf{x} - \mathbf{y}\|_2^2 - \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1. \end{aligned}$$

Proof. We defer the proof to the Appendix.

The above result shows the existence of an optimal potential function possessing a weakly-concave structure defined based on an elastic net function. Our next result reveals the extension of the Brenier’s theorem to the elastic net cost function. In this extension, we use ST_γ to denote the soft-thresholding operator defined for a scalar input as

$$\text{ST}_\gamma(z) := \begin{cases} z + \gamma & \text{if } z \leq -\gamma \\ 0 & \text{if } -\gamma < z < \gamma \\ z - \gamma & \text{if } \gamma \leq z. \end{cases}$$

For a vector input $\mathbf{z} \in \mathbb{R}^d$, we define the soft-thresholding map as the coordinate-wise application of the scalar soft-thresholding function, i.e.,

$$\forall i \in \{1, \dots, d\} : \quad \text{ST}_\gamma(\mathbf{z})_i = \text{ST}_\gamma(z_i)$$

Theorem 3. Consider the dual ENOT problem in equation 3. Then, given the optimal potential function ϕ^* the following will provide the optimal transport map transferring samples across domains:

$$\mathbf{X} - \text{ST}_{\frac{\alpha}{2(1-\alpha)}}\left(\frac{1}{2\lambda(1-\alpha)}\nabla\phi^*(\mathbf{X})\right) \stackrel{\text{dist}}{=} \mathbf{Y}.$$

Proof. We defer the proof to the Appendix.

Note that the above theorem is a generalization of the Brenier theorem for the elastic net cost, and in the special case of $\alpha = 0$ reduces to the Brenier theorem. On the other hand, by selecting a larger L_1 -regularization coefficient α , the soft-thresholding map will apply a more stringent sparsification to the gradient map of the optimal potential function. This result suggests that by choosing a larger α , one can achieve a sparser transportation map which is the goal sought by the sparse transfer algorithm. Theorem 3 reduces the search for the elastic net-based transport map to the computation of the optimal potential function ϕ in the dual optimization problem, which as shown in Theorem 2 satisfies a weakly-concavity property.

Algorithm 1 GAN Training with ENOT-based Feature Selection

Require: Training data \mathbf{X} , target data \mathbf{Y} , hyperparameters λ, α , generator G and discriminator D , pre-trained ENOT potential function $\phi^*(\mathbf{x})$

- 1: Initialize generator G and discriminator D
 - 2: **while** not converged **do**
 - 3: Sample minibatch $\mathbf{x} \sim \mathbf{X}$ and $\mathbf{y} \sim \mathbf{Y}$
 - 4: Compute feature selection mask $I(\mathbf{x})$:
 - 5: $\forall i \in \{1, \dots, d\} : I(\mathbf{x})_i = \begin{cases} 0 & \text{if } |\nabla \phi^*(\mathbf{x})_i| \leq \lambda\alpha, \\ 1 & \text{if } |\nabla \phi^*(\mathbf{x})_i| > \lambda\alpha \end{cases}$
 - 6: Update D by ascending its stochastic gradient:
 - 7: $\nabla_D [\mathbb{E} [\log (D(\mathbf{y}))] + \mathbb{E} [\log (1 - D(I(\mathbf{x}) \odot G(\mathbf{x}) + (1 - I(\mathbf{x})) \odot \mathbf{x}))]]$
 - 8: Update G by descending its stochastic gradient:
 - 9: $\nabla_G [\mathbb{E} [\log (1 - D(I(\mathbf{x}) \odot G(\mathbf{x}) + (1 - I(\mathbf{x})) \odot \mathbf{x}))]]$
 - 10: **end while**
 - 11: **return** Trained generator G and discriminator D
-

5 ENOT-based Feature Selection for Sparse Domain Transfer

In the previous section, we have shown sparse transfer map could be derived by applying the soft-thresholding function to the gradient of the optimal potential function. In addition to directly performing a sparse optimal transport, the trained potential function in the ENOT framework can be used for variable selection to undergo an unconstrained distribution transfer. Therefore, we also propose a feature selection algorithm for domain transfer using the optimal potential function ϕ^* in equation 3. Here for an input $\mathbf{x} \in \mathbb{R}^d$, we define the feature selection mask $I : \mathbb{R}^d \rightarrow \{0, 1\}^d$ as

$$\forall i \in \{1, \dots, d\} : I(\mathbf{x})_i = \begin{cases} 0 & \text{if } |\nabla \phi^*(\mathbf{x})_i| \leq \lambda\alpha, \\ 1 & \text{if } |\nabla \phi^*(\mathbf{x})_i| > \lambda\alpha \end{cases} \quad (4)$$

The above masking identifies the feature coordinates modified by the ENOT transport map. Given the above masking function, we can train a generator function $G : \mathbb{R}^d \rightarrow \mathbb{R}^d$ to perform a constraint-free domain transportation on the ENOT's selected features. We can employ the standard GAN framework [15] consisting of a generator G and discriminator function $D : \mathbb{R}^d \rightarrow \mathbb{R}$ to do this task. Following the standard min-max formulation of GANs, we propose the following optimization problem for the ENOT feature selection-based domain transport:

$$\min_{G \in \mathcal{G}} \max_{D \in \mathcal{D}} \mathbb{E} [\log (D(\mathbf{Y}))] + \mathbb{E} [\log (1 - D(I(\mathbf{X}) \odot G(\mathbf{X}) + (1 - I(\mathbf{X})) \odot \mathbf{X}))] \quad (5)$$

In the above, the generator G attempts to match the distribution of modified $I(\mathbf{X}) \odot G(\mathbf{X}) + (1 - I(\mathbf{X})) \odot \mathbf{X}$ with the distribution of \mathbf{Y} , where \odot denotes the

Table 1: ENOT’s achieved NLL with different coefficients of L_1 -regularization on the Gaussian mixture transfer.

f	dimension	ENOT L_1 coefficient						
		baseline	0	1e-3	5e-3	1e-2	5e-2	1e-1
MLP-22	1000	4.87×10^3	4.46×10^3	4.13×10^3	3.52×10^3	1.50×10^3	1.50×10^3	1.51×10^3
	100	3.22×10^3	3.15×10^3	2.97×10^3	1.82×10^3	1.62×10^2	1.64×10^2	1.82×10^2
	10	1.53×10^1	1.50×10^1	1.42×10^1	1.39×10^1	1.31×10^1	1.40×10^1	1.51×10^1
MLP-12	1000	4.56×10^3	4.42×10^3	4.02×10^3	3.48×10^3	1.50×10^3	1.50×10^3	1.50×10^3
	100	2.73×10^3	2.58×10^3	2.85×10^3	1.27×10^3	1.50×10^2	1.51×10^2	1.51×10^2
	10	1.60×10^1	1.39×10^1	1.51×10^1	1.48×10^1	1.34×10^1	1.48×10^1	1.67×10^1
MLP-4	1000	3.67×10^3	3.62×10^3	3.44×10^3	3.01×10^3	1.50×10^3	1.50×10^3	1.50×10^3
	100	2.62×10^3	2.53×10^3	1.44×10^3	1.01×10^3	1.50×10^2	1.51×10^2	1.51×10^2
	10	1.52×10^1	1.36×10^1	1.38×10^1	1.44×10^1	1.32×10^1	1.47×10^1	1.77×10^1

element-wise Hadamard product. On the other hand, the discriminator D seeks to identify the original \mathbf{Y} samples from the modified \mathbf{X} data. Since we utilize the feature selection mask of the trained ENOT potential function, we do not need to ensure the invertibility of the generator and can reduce the number of machine players compared to the CycleGAN algorithm.

The above feature selection-based approach enables the application of neural net generator functions which could improve the vanilla ENOT’s performance due to the power of a properly-designed generator to model the structures in the text and image data. This is similar to the Wasserstein GAN (WGAN) [1] as the optimal-transport-based GAN formulation in WGANs also considers a generator G instead of relying on the gradient of the potential function.

6 Numerical Results

In this section, we present the empirical results of the applications of ENOT and the baseline domain transfer algorithms to several standard datasets, including synthetic Gaussian mixture models, and real image and text datasets. We defer the details of our numerical experiments, including the dataset pre-processing, neural network architectures, and hyperparameter selection to the Appendix.

6.1 ENOT applied to Synthetic Gaussian Mixture Data

We evaluated the performance of ENOT in domain transfer problems across multivariate Gaussian mixture models (GMMs). In our experiments, we considered bimodal source and target GMMs: the source GMM $p(\mathbf{x}) = \phi_s \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_s, \sigma^2 \mathbf{I}_d) + (1 - \phi_s) \mathcal{N}(\mathbf{x} | -\boldsymbol{\mu}_s, \sigma^2 \mathbf{I}_d)$, and target GMM $p(\mathbf{y}) = \phi_t \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_t, \sigma^2 \mathbf{I}_d) + (1 - \phi_t) \mathcal{N}(\mathbf{y} | -\boldsymbol{\mu}_t, \sigma^2 \mathbf{I}_d)$ both consist of two multivariate Gaussian components with different means and identical covariance matrix with $\sigma = 1$. There exist two component-based mappings: (1) mapping $\mathcal{N}_{\boldsymbol{\mu}_s} \rightarrow \mathcal{N}_{\boldsymbol{\mu}_t}$, $\mathcal{N}_{-\boldsymbol{\mu}_s} \rightarrow \mathcal{N}_{-\boldsymbol{\mu}_t}$ or (2)

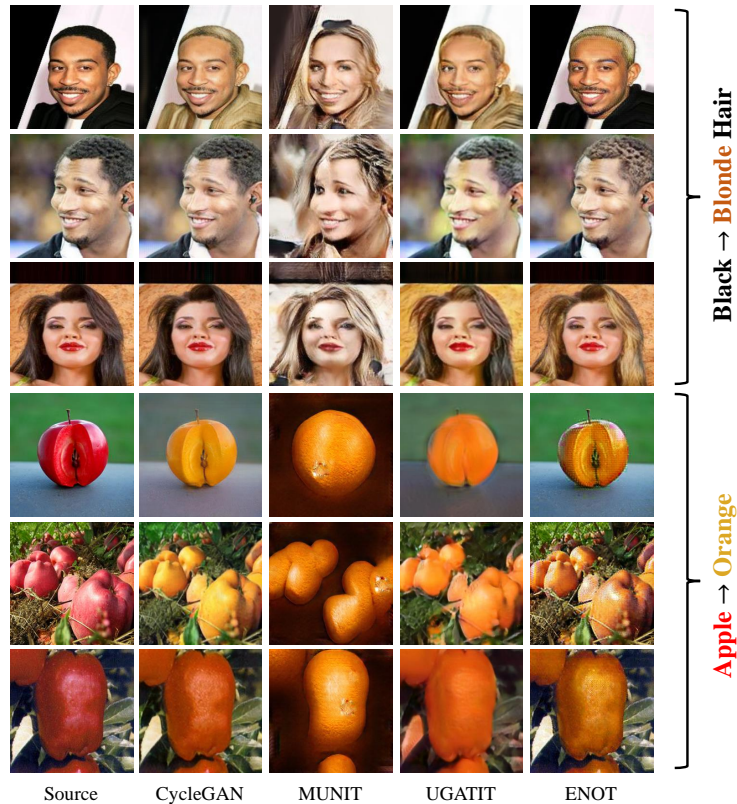


Fig. 1: Transportation for Black→Blonde hair and Apple→Orange on CelebA and Apple2Orange.

mapping $\mathcal{N}_{\mu_s} \rightarrow \mathcal{N}_{-\mu_t}$, $\mathcal{N}_{-\mu_s} \rightarrow \mathcal{N}_{\mu_t}$. We set $\mu_s = [\gamma, \epsilon_d, \dots, \epsilon_d]$, $\mu_t = [-\gamma, \epsilon_d, \dots, \epsilon_d]$ and chose $\epsilon_d \cdot \sqrt{d} - 1 < \gamma$ to distinguish the optimal L_1 -norm-based sparse and standard L_2 -norm-based transfer maps. We set $\gamma = 10$, $\epsilon_{10} = 2$, $\phi_s = \phi_t = 0.5$, and scaled $\epsilon_d = \frac{\epsilon_{10}}{\sqrt{d/10}}$ to ensure the inequality holds in different dimensions.

We applied the ENOT approach by solving the dual optimization problem (Eq. 3) using a multi-layer perception neural net with different number of ReLU layers. We attempted different L_1 -norm coefficients, where a zero coefficient reduces to the standard optimal transport baseline. We evaluated the performance of the domain transfer algorithm using the averaged negative log-likelihood (NLL) of transferred samples with respect to the target Gaussian mixture distribution. Based on our quantitative results in Table 1, we observed that the ENOT’s sparse transfer maps led to better performance scores for the three potential function architectures.

6.2 Image-based Domain Transfer

We utilized the proposed ENOT framework to perform image domain transfer and compared its performance with image translation baselines: CycleGAN [42], MUNIT [16], and UGATIT [21]. We also selected baseline DoRM [39] from the domain adaptation literature. In our computer vision experiments, we used four standard datasets: MNIST [23], CelebA [27], Apple2Orange [42] and FFHQ [20]. Due to the page limit, we defer the DoRM and FFHQ results to the Appendix. For training the ENOT’s potential function, we used a 5-layer MLP in the MNIST experiments and used a Vision Transformer (ViT-base) model with patch size 16 from [12]. We defer the discussion on the selection of the α, λ coefficients we performed in the ENOT-based feature selection to the Appendix.

The object translation task in a computer vision setting typically results in a sparse domain transfer task. For example, if we wish to change the hair color of a human, only partial pixels regarding hair are expected to be modified. In computer vision domain transfer problems, the standard domain transfer algorithms leverage high-capacity GANs to reach satisfactory visual quality. However, in GANs, the goal of the generator is to fool the discriminator. This goal might lead to suboptimal results. For example, in Figure 1, when change the hair color from black to blonde for an individual wearing black clothes, the GAN-based methods could mistakenly alter the color of the clothes to yellow simultaneously, thereby outputting a realistic but overly-changed sample. In this case, sparsity is desired. However, integrating the sparsity prior to the GAN optimization could lead to highly challenging min-max optimization tasks.

In Figure 1, we present the empirical results for randomly selected CelebA samples in the transport task: black hair \rightarrow blonde hair, and Apple2Orange samples in the transport task: apple \rightarrow orange. We observe that transportation maps from the baseline methods are unsatisfactory in several sparse-transportation cases. Two common types of failures are present in Figure 1. The first one is over-transportation, where unnecessary pixels are modified. This is commonly observed in baseline methods which employ dense-transportation algorithms. The second failure is insignificant transportation since baselines are not sensitive enough to the sparse transportation regions. On the other hand, by using ENOT-based feature selection, the transfer results are considerably improved. As shown in Figure 2, the ENOT feature selection successfully identified the pixels corresponding to the subject’s hair in CelebA and the apples in Apple2Orange samples. The proper variable selection led to a more meaningful domain transfer in these computer vision applications.

6.3 IMDB Review Sentiment Reversal

We also performed the numerical experiments on the IMDB movie review text dataset [28]. This dataset contains 50,000 movie reviews with positive and negative categories. We defined the transportation task as modifying part of the words to flip the review’s sentiment: negative to positive reviews. We attempted a sparse domain transfer task in this case, as the sparsity level could be a sensible

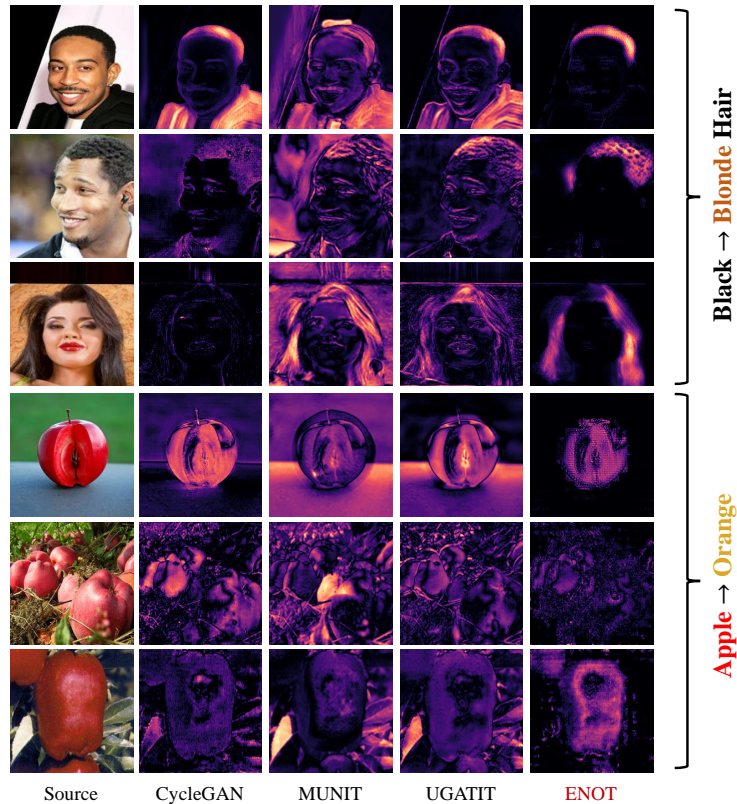


Fig. 2: Saliency maps of transportation for Black→Blonde hair and Apple→Orange on CelebA and Apple2Orange.

quantification of the revision made to the text data. We expect that a sparse transport map exists in this case, when the transfer map only flips the negative and positive adjectives in the text.

We used ENOT to perform domain transfer in this text-based setting. For the potential function in (Eq. 3), we finetuned a pre-trained BERT transformer [11]. As the baseline, we considered a pre-trained Seq2Seq model GPT-3 [6]. Table 2 shows the empirical results of the baseline and ENOT on a randomly selected sample. For this sample, we observed the revision made with the ENOT-based feature selection method is sparse and only a few words regarding movie review sentiment are modified. In contrast, the baseline Seq2Seq model GPT-3 modified almost all the text, including sentences describing the movie details that sound unrelated to the review sentiment. We present the results for more samples in the Appendix.

Also, we empirically observed that informative words in the generated sparse transportation maps by ENOT have higher correlations with the sentiment compared with other parts of the input text. We present this phenomenon in Fig-

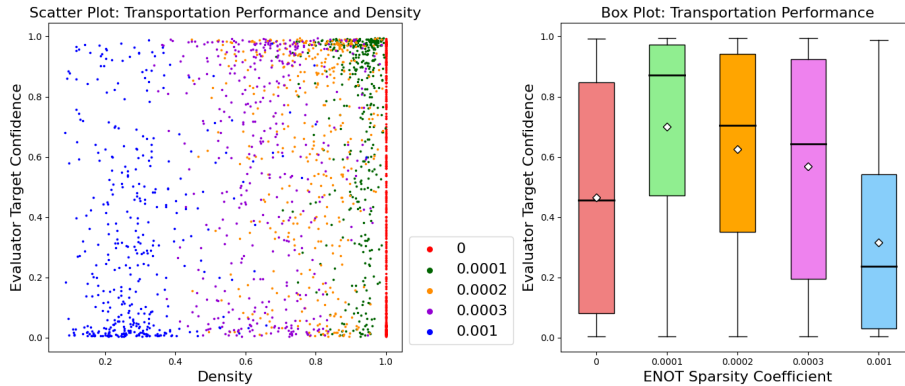


Fig. 3: **Left:** IMDB sentiment transfer quality and density. Every point represents a transported sample: color indicates the L_1 -coefficient in ENOT. The quality is measured by confidence score from a BERT classifier. A density value of 0.5 indicates that 50% of the input tokens have been modified by ENOT. **Right:** Transportation quality, the middle line and diamond show the median and mean. The green bar (coefficient $1e^{-4}$) achieves the highest quality.

ure 3, where we quantified the transportation performance using the confidence score [29] from a pre-trained BERT model [11] on IMDB sentiment classification. We analyzed the correlation between performance and sparsity in the ENOT’s transportation. In the scatter plot, each point is a random transported sample: the bottom-left region indicates sparse and poor performance, and the top-right region suggests dense and good performance. The box plot statistics in Figure 3 show that a sparsity coefficient of 10^{-4} attains the best performance compared to the dense baseline and other coefficients.

7 Conclusion

In this work, we focused on the sparse domain transfer task and attempted to apply L_1 -norm regularization to the standard optimal transport framework by considering an elastic net cost function. Our numerical results suggest the proposed method’s performance gain under a sparse transfer map. An interesting future direction is to apply the proposed framework to the latent space of image and text data where tighter sparsity constraints may hold in the learning setting. The extension of the elastic net-based optimal transport framework to provide a sparse and concise interpretation of domain transfer maps is another direction for future exploration.

Acknowledgments. This work is partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, Project 14209920, and is partially supported by a CUHK Direct Research Grant with CUHK Project No. 4055237.

Table 2: IMDB Movie Review Sentiment Transportation Maps

Transfer Task: Negative Review \rightarrow Positive Review		
Red: ENOT's Selected Words for Domain Transfer. Blue: Modified Parts		
Source	Baseline Seq2Seq	ENOT
Lonely, disconnected, middle-class housewife in the midst of a divorce seeks solace to reflect on her immediate future. at some sort of bed and breakfast by (well, literally in the sea) the ocean that for some sort of odd reason she subs for the owner. enter lonely, richard gere. he is a plastic surgeon. he is the only guest at the inn in the sea. Diane lane is the lonely housewife. you'll never guess these two fall immediately in love. a tropical storm makes them true lovers. the subplots or no sense. the locations, photography are fine. Gere and does ms. lane must've needed the money, but her part with grace.	Lonely, disconnected, middle-class housewife in the midst of a divorce seeks solace to reflect on her immediate future. at some sort of bed and breakfast by (well, literally in the sea) the ocean that for some sort of odd reason she subs for the owner. enter lonely, richard gere. he is a plastic surgeon. he is the only guest at the inn in the sea. Diane lane is the lonely housewife. you'll never guess these two fall immediately in love. a tropical storm makes them true lovers. the subplots or no sense. the locations, photography are fine. Gere and does ms. lane must've needed the money, but her part with grace.	Lonely, disconnected, middle-class housewife in the midst of a divorce seeks solace to reflect on her immediate future. at some sort of bed and breakfast by (well, literally in the sea) the ocean that for some sort of odd reason she subs for the owner. enter lonely, arrogant richard gere. he in to help the owner. charismatic richard gere. he is a plastic surgeon. he Then, enter the charismatic he is a plastic surgeon. he is the only guest at the Richard Gere, a skilled is the only guest at the inn in the sea. Diane lane plastic surgeon and the in the sea. Diane lane is the lonely housewife. only guest at this idyllic the lonely housewife. you'll never guess these two seaside inn. Diane Lane never guess these two fall immediately in love. a portrays the captivating immediately in love. a trop- ical storm makes them true lovers. the subplots in antly surprised as these lovers. the subplots add this melodrama make little two form an instant and depth and intrigue to this or no sense. the locations, profound connection. A melodrama. the breath- photography are fine. Gere tropical storm adds a touch taking locations, stunning remains one of the most of magic to their love story. photography are absolutely over-rated actors in cinema The various subplots in remarkable. Gere proves and does not disappoint. this heartfelt drama weave once again that he is one of ms. lane must've needed together seamlessly. The the most respected actors the money, but phones in stunning locations and in the world of cinema and her part with grace. photography enhance the does a fantastic job. ms. overall experience. Gere lane must've needed the continues to be one of the money, and delivers her most respected actors in part with grace. cinema, delivering a stellar performance as always.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International conference on machine learning. pp. 214–223. PMLR (2017) [9](#)
2. Bao, H., Sakaue, S.: Sparse regularized optimal transport with deformed q-entropy. *Entropy* **24**(11), 1634 (2022) [3](#)
3. Bhattacharjee, D., Kim, S., Vizier, G., Salzmman, M.: Dunit: Detection-based unsupervised image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4787–4796 (2020) [4](#)
4. Blondel, M., Seguy, V., Rolet, A.: Smooth and sparse optimal transport. In: International conference on artificial intelligence and statistics. pp. 880–889. PMLR (2018) [3](#)
5. Boyd, S.P., Vandenberghe, L.: Convex optimization. Cambridge university press (2004) [18](#)
6. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. *Advances in neural information processing systems* **33**, 1877–1901 (2020) [12](#), [22](#)
7. Chen, R., Huang, W., Huang, B., Sun, F., Fang, B.: Reusing discriminators for encoding: Towards unsupervised image-to-image translation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8168–8177 (2020) [1](#), [4](#)
8. Cuturi, M., Klein, M., Ablin, P.: Monge, bregman and occam: Interpretable optimal transport in high-dimensions with feature-sparse maps. *arXiv preprint arXiv:2302.04065* (2023) [3](#), [4](#)
9. Demir, U., Unal, G.: Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422* (2018) [21](#)
10. Deng, L.: The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine* **29**(6), 141–142 (2012) [23](#)
11. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018) [4](#), [12](#), [13](#), [22](#)
12. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020) [11](#), [21](#)
13. Flamary, R., Courty, N., Tuia, D., Rakotomamonjy, A.: Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell* **1**(1-40), 2 (2016) [3](#)
14. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science* **286**(5439), 531–537 (1999) [24](#), [27](#)
15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014) [8](#)
16. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: ECCV (2018) [11](#)
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017) [4](#)

18. Jiang, L., Zhang, C., Huang, M., Liu, C., Shi, J., Loy, C.C.: Tsit: A simple and versatile framework for image-to-image translation. In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III* 16. pp. 206–222. Springer (2020) **1**
19. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. In: *Proceedings of the 2013 conference on empirical methods in natural language processing*. pp. 1700–1709 (2013) **4**
20. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4401–4410 (2019) **1, 11**
21. Kim, J., Kim, M., Kang, H., Lee, K.: U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation (2020) **11**
22. Korotin, A., Selikhanovych, D., Burnaev, E.: Neural optimal transport. *arXiv preprint arXiv:2201.12220* (2022) **3**
23. LeCun, Y.: The mnist database of handwritten digits. <http://yann.lecun.com/exdb/mnist/> (1998) **11**
24. Lee, H.Y., Tseng, H.Y., Mao, Q., Huang, J.B., Lu, Y.D., Singh, M., Yang, M.H.: Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision* **128**, 2402–2417 (2020) **4**
25. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019) **4**
26. Liu, T., Puigcerver, J., Blondel, M.: Sparsity-constrained optimal transport. In: *The Eleventh International Conference on Learning Representations* (2023) **3**
27. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: *Proceedings of International Conference on Computer Vision (ICCV)* (December 2015) **11**
28. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. pp. 142–150. Association for Computational Linguistics, Portland, Oregon, USA (June 2011) **11, 21**
29. Mandelbaum, A., Weinshall, D.: Distance-based confidence score for neural network classifiers (2017) **13**
30. Ojha, U., Li, Y., Lu, J., Efros, A.A., Lee, Y.J., Shechtman, E., Zhang, R.: Few-shot image generation via cross-domain correspondence. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10743–10752 (2021) **1**
31. van der Ouderaa, T.F., Worrall, D.E.: Reversible gans for memory-efficient image-to-image translation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 4720–4728 (2019) **1, 4**
32. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al.: Improving language understanding by generative pre-training (2018) **4**
33. Shen, Z., Huang, M., Shi, J., Xue, X., Huang, T.S.: Towards instance-level image-to-image translation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 3683–3692 (2019) **4**
34. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. *Advances in neural information processing systems* **27** (2014) **4**

35. Swanson, K., Yu, L., Lei, T.: Rationalizing text matching: Learning sparse alignments via optimal transport. arXiv preprint arXiv:2005.13111 (2020) [3](#)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017) [4](#)
37. Villani, C., et al.: *Optimal transport: old and new*, vol. 338. Springer (2009) [2](#), [5](#), [6](#), [18](#), [20](#)
38. Wu, W., Cao, K., Li, C., Qian, C., Loy, C.C.: Transgaga: Geometry-aware unsupervised image-to-image translation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 8012–8021 (2019) [4](#)
39. Wu, Y., Li, Z., Wang, C., Zheng, H., Zhao, S., Li, B., Tao, D.: Domain remodulation for few-shot generative domain adaptation. *Advances in Neural Information Processing Systems* **36** (2024) [11](#), [24](#), [25](#)
40. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999 (2017) [4](#)
41. Zhao, Y., Wu, R., Dong, H.: Unpaired image-to-image translation using adversarial consistency loss. In: *ECCV* (2020) [1](#), [4](#)
42. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2223–2232 (2017) [1](#), [4](#), [11](#), [21](#)
43. Zhu, P., Abdal, R., Qin, Y., Wonka, P.: Sean: Image synthesis with semantic region-adaptive normalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 5104–5113 (2020) [1](#)
44. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **67**(2), 301–320 (2005) [2](#), [6](#)

A Appendix

A.1 Proofs

Proof of Theorem 2 To show the theorem, we follow the Kantorovich duality (Theorem 5.10 from [37]) which demonstrates that under a continuous and non-negative (thus lower-bounded) cost function which is an applicable assumption to the elastic net cost, there exists an optimal potential function ϕ^* that is the c -transform of some function $\tilde{\phi}$, i.e. for every $\mathbf{x} \in \mathbb{R}^d$:

$$\phi^*(\mathbf{x}) = \inf_{\mathbf{y}} \left\{ \tilde{\phi}(\mathbf{y}) + \lambda(1 - \alpha)\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\alpha\|\mathbf{y} - \mathbf{x}\|_1 \right\}.$$

We rewrite the above objective function as

$$\begin{aligned} & \tilde{\phi}(\mathbf{y}) + \lambda(1 - \alpha)\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\alpha\|\mathbf{y} - \mathbf{x}\|_1 \\ &= \tilde{\phi}(\mathbf{y}) + \lambda(1 - \alpha)\left(\|\mathbf{y}\|_2^2 + \|\mathbf{x}\|_2^2 - 2\mathbf{x}^\top \mathbf{y}\right) \\ & \quad + \lambda\alpha\|\mathbf{y} - \mathbf{x}\|_1 \\ &= \left\{ \tilde{\phi}(\mathbf{y}) + \lambda(1 - \alpha)\|\mathbf{y}\|_2^2 \right\} + \lambda(1 - \alpha)\|\mathbf{x}\|_2^2 \\ & \quad - 2\lambda(1 - \alpha)\mathbf{x}^\top \mathbf{y} + \lambda\alpha\|\mathbf{y} - \mathbf{x}\|_1. \end{aligned}$$

Therefore, if we denote $\tilde{\phi}(\mathbf{y}) := \tilde{\phi}(\mathbf{y}) + \lambda(1 - \alpha)\|\mathbf{y}\|_2^2$, then we have

$$\begin{aligned} & \phi^*(\mathbf{x}) - \lambda(1 - \alpha)\|\mathbf{x}\|_2^2 \\ &= \inf_{\mathbf{y}} \left\{ \tilde{\phi}(\mathbf{y}) - 2\lambda(1 - \alpha)\mathbf{x}^\top \mathbf{y} + \lambda\alpha\|\mathbf{y} - \mathbf{x}\|_1 \right\}. \end{aligned}$$

To analyze the Lipschitzness and concavity properties of the above function, we define the following function $g : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$:

$$g(\mathbf{x}_1, \mathbf{x}_2) := \inf_{\mathbf{y}} \left\{ \tilde{\phi}(\mathbf{y}) - 2\lambda(1 - \alpha)\mathbf{x}_1^\top \mathbf{y} + \lambda\alpha\|\mathbf{y} - \mathbf{x}_2\|_1 \right\}.$$

Observation 1: According to the definitions $g(\mathbf{x}, \mathbf{x}) = \phi^*(\mathbf{x}) - \lambda(1 - \alpha)\|\mathbf{x}\|_2^2$ holds for every \mathbf{x} .

Lemma 1: For a fixed \mathbf{x}_2 , $g(\mathbf{x}_1, \mathbf{x}_2)$ is a concave function of \mathbf{x}_1 .

Proof. if we define $k_{\mathbf{x}_2}(\mathbf{y}) := \tilde{\phi}(\mathbf{y}) + \lambda\alpha\|\mathbf{y} - \mathbf{x}_2\|_1$, then $g(\mathbf{x}_1, \mathbf{x}_2)$ equals an infimum of a set of affine functions of \mathbf{x}_1 : $k_{\mathbf{x}_2}(\mathbf{y}) - 2\lambda(1 - \alpha)\mathbf{y}^\top \mathbf{x}_1$ which is known to lead to a concave function [5].

Lemma 2: For a fixed \mathbf{x}_1 , $g(\mathbf{x}_1, \mathbf{x}_2)$ is a $\lambda\alpha$ -Lipschitz function of \mathbf{x}_2 in terms of L_1 -norm.

Proof. If we define $h_{\mathbf{x}_1}(\mathbf{y}) := \tilde{\phi}(\mathbf{y}) - 2\lambda(1 - \alpha)\mathbf{x}_1^\top \mathbf{y}$, then $g(\mathbf{x}_1, \mathbf{x}_2)$ equals an infimum of $h_{\mathbf{x}_1}(\mathbf{y}) + \lambda\alpha\|\mathbf{y} - \mathbf{x}_2\|_1$ which, according to the Kantorovich duality for a norm function [37], will be a $\lambda\alpha$ -Lipschitz function and satisfies the following for every $\mathbf{x}_2, \mathbf{x}'_2$:

$$|g(\mathbf{x}_1, \mathbf{x}_2) - g(\mathbf{x}_1, \mathbf{x}'_2)| \leq \lambda\alpha\|\mathbf{x}_2 - \mathbf{x}'_2\|_1.$$

Note that Lemma 2 suggests the following holds for every $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\gamma \in [0, 1]$:

$$\left| g(\mathbf{x}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) - g(\mathbf{x}, \mathbf{x}) \right| \leq \lambda\alpha(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_1 \quad (6)$$

Similarly, Lemma 2 shows that

$$\left| g(\mathbf{y}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) - g(\mathbf{y}, \mathbf{y}) \right| \leq \lambda\alpha\gamma\|\mathbf{x} - \mathbf{y}\|_1 \quad (7)$$

Combining the above inequalities with Lemma 1's result, for every $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\gamma \in [0, 1]$ we have

$$\begin{aligned} & g(\gamma\mathbf{x} + (1-\gamma)\mathbf{y}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) \\ & \geq \gamma g(\mathbf{x}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) + (1-\gamma)g(\mathbf{y}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) \\ & \geq \gamma g(\mathbf{x}, \mathbf{x}) + (1-\gamma)g(\mathbf{y}, \mathbf{y}) - \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1. \end{aligned} \quad (8)$$

Finally, we note that

$$\begin{aligned} \|\gamma\mathbf{x} + (1-\gamma)\mathbf{y}\|_2^2 &= \gamma\|\mathbf{x}\|_2^2 + (1-\gamma)\|\mathbf{y}\|_2^2 \\ &\quad - \gamma(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_2^2. \end{aligned}$$

Therefore, we can combine the above results to show the following property for the optimal potential function ϕ^*

$$\begin{aligned} & \phi^*(\gamma\mathbf{x} + (1-\gamma)\mathbf{y}) \\ &= g(\gamma\mathbf{x} + (1-\gamma)\mathbf{y}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) \\ &\quad + \lambda(1-\alpha)\|\gamma\mathbf{x} + (1-\gamma)\mathbf{y}\|_2^2 \\ &= g(\gamma\mathbf{x} + (1-\gamma)\mathbf{y}, \gamma\mathbf{x} + (1-\gamma)\mathbf{y}) \\ &\quad + \lambda(1-\alpha)\gamma\|\mathbf{x}\|_2^2 + \lambda(1-\alpha)(1-\gamma)\|\mathbf{y}\|_2^2 \\ &\quad - \lambda(1-\alpha)\gamma(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_2^2 \\ &\geq \gamma g(\mathbf{x}, \mathbf{x}) + (1-\gamma)g(\mathbf{y}, \mathbf{y}) - \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1 \\ &\quad + \lambda(1-\alpha)\gamma\|\mathbf{x}\|_2^2 + \lambda(1-\alpha)(1-\gamma)\|\mathbf{y}\|_2^2 \\ &\quad - \lambda(1-\alpha)\gamma(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_2^2 \\ &= \gamma \left(g(\mathbf{x}, \mathbf{x}) + \lambda(1-\alpha)\|\mathbf{x}\|_2^2 \right) \\ &\quad + (1-\gamma) \left(g(\mathbf{y}, \mathbf{y}) + \lambda(1-\alpha)\|\mathbf{y}\|_2^2 \right) \\ &\quad - \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1 - \lambda(1-\alpha)\gamma(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_2^2 \\ &= \gamma\phi^*(\mathbf{x}) + (1-\gamma)\phi^*(\mathbf{y}) \\ &\quad - \lambda\alpha\|\mathbf{x} - \mathbf{y}\|_1 - \lambda(1-\alpha)\gamma(1-\gamma)\|\mathbf{x} - \mathbf{y}\|_2^2. \end{aligned}$$

Therefore, the proof is complete.

Proof of Theorem 3 Let $\Pi_{X,Y}^*$ be a solution to the optimal transport problem between P_X and P_Y with the elastic net cost function. According to the Kantorovich duality (Theorem 5.10 from [37]), since the elastic net cost function is continuous and non-negative, there should exist a $c_{\text{EN}}^{\alpha,\lambda}$ -conjugate pair $\phi^*, \psi^* = \phi^* c_{\text{EN}}^{\alpha,\lambda}$ such that the following inequality, which holds for every \mathbf{x}, \mathbf{y} , holds with equality $\Pi_{X,Y}^*$ -almost surely:

$$\phi^*(\mathbf{x}) - \psi^*(\mathbf{y}) \leq c_{\text{EN}}^{\alpha,\lambda}(\mathbf{x}, \mathbf{y}). \quad (9)$$

Therefore, if ϕ^* is differentiable at a point \mathbf{x} where $(\mathbf{x}, \mathbf{y}) \sim \Pi^*$, for any differentiable curve $\tilde{x}(\epsilon)$ such that $\tilde{x}(0) = \mathbf{x}$ we will have:

$$\langle \nabla \phi^*(\mathbf{x}), \dot{\tilde{x}}(0) \rangle \leq \liminf_{\epsilon \rightarrow 0} \frac{c_{\text{EN}}^{\alpha,\lambda}(\tilde{x}(\epsilon), \mathbf{y}) - c_{\text{EN}}^{\alpha,\lambda}(\mathbf{x}, \mathbf{y})}{\epsilon}$$

Here $\langle \cdot, \cdot \rangle$ denotes the standard vector inner product. As a result, $\nabla \phi^*(\mathbf{x})$ is a subgradient of $c_{\text{EN}}^{\alpha,\lambda}(\cdot, \mathbf{y})$. Note that the elastic net cost function decouples across the coordinates as:

$$c_{\text{EN}}^{\alpha,\lambda}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^d \left[\lambda(1-\alpha)(x_i - y_i)^2 + \lambda\alpha|x_i - y_i| \right].$$

Therefore, at every coordinate i , either $-\lambda\alpha \leq \nabla \phi^*(\mathbf{x})_i \leq \lambda\alpha$ holds which suggests the i -coordinate-based cost is zero and $y_i = x_i$, or $|\nabla \phi^*(\mathbf{x})_i| > \lambda\alpha$ under which we have a one-to-one mapping between the coordinate-based $\nabla \phi^*(\mathbf{x})_i$ and $x_i - y_i$ as

$$x_i - y_i = \frac{1}{2\lambda(1-\alpha)} \left(\nabla \phi^*(\mathbf{x})_i - \text{sign}(\nabla \phi^*(\mathbf{x})_i) \lambda\alpha \right).$$

Hence, the following holds at every coordinate i

$$x_i - y_i = \begin{cases} \frac{\nabla \phi^*(\mathbf{x})_i - \text{sign}(\nabla \phi^*(\mathbf{x})_i) \lambda\alpha}{2\lambda(1-\alpha)} & \text{if } |\nabla \phi^*(\mathbf{x})_i| > \lambda\alpha \\ 0 & \text{if } |\nabla \phi^*(\mathbf{x})_i| \leq \lambda\alpha \end{cases}$$

The above equation at (\mathbf{x}, \mathbf{y}) can be seen to be equivalent to the following equation given the soft-thresholding map definition in the main text:

$$\mathbf{x} - \mathbf{y} = \text{ST}_{\frac{\alpha}{2(1-\alpha)}} \left(\frac{1}{2\lambda(1-\alpha)} \nabla \phi^*(\mathbf{x}) \right).$$

Note that the above equality is supposed to hold Π^* -almost surely as long as the optimal potential function ϕ^* is differentiable at \mathbf{x} . On the other hand, due to Theorem 1 and the weakly concavity property of the optimal potential function ϕ^* , Rademacher's theorem implies that ϕ^* must be differentiable almost everywhere. Therefore, assuming that P_X is absolutely continuous with respect to the volume measure, we can show that the following equality holds $\Pi_{X,Y}^*$ -almost surely for the optimal potential function ϕ^* :

$$\mathbf{X} - \text{ST}_{\frac{\alpha}{2(1-\alpha)}} \left(\frac{1}{2\lambda(1-\alpha)} \nabla \phi^*(\mathbf{X}) \right) = \mathbf{Y}.$$

Therefore, the proof is complete.

Table 3: Noisy-MNIST Transportation De-noising Ratio

Noisy-padding Ratio	Image Size	ENOT L_1 Coefficient				
		Baseline	0 1e-4	5e-4	1e-3	5e-3
25%	42	0%	0.38%	20.12%	32.40%	100%
50%	56	0%	0.37%	16.21%	34.36%	100%

A.2 Experimental Setting

Datasets. For multivariate GMMs, we considered bimodal source and target GMMs: the source GMM $p(\mathbf{x}) = \phi_s \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_s, \sigma^2 \mathbf{I}_d) + (1 - \phi_s) \mathcal{N}(\mathbf{x} | -\boldsymbol{\mu}_s, \sigma^2 \mathbf{I}_d)$, and target GMM $p(\mathbf{y}) = \phi_t \mathcal{N}(\mathbf{y} | \boldsymbol{\mu}_t, \sigma^2 \mathbf{I}_d) + (1 - \phi_t) \mathcal{N}(\mathbf{y} | -\boldsymbol{\mu}_t, \sigma^2 \mathbf{I}_d)$ both consist of two multivariate Gaussian components with different means and identical covariance matrix with $\sigma = 1$. We generate 2,000 independent samples for each GMM, which are further equally divided into training and testing sets. The dimension $d \in \{10, 100, 1000\}$. For Noisy-MNIST, we directly add noisy paddings sampled from a uniform distribution, which ranges from 0 to 1, to the original images in the MNIST dataset. The MNIST dataset contains 50,000 training images and 10,000 test images, which are grayscale images. The size of the images is 28 by 28 pixels without noisy padding and 42 by 42 pixels with noisy padding. For the CelebA black2blonde hair datasets, we split the original CelebA dataset into two datasets according to the label of the images. We selected the labels "black hair" and "blonde hair" to do the split. Each category dataset is further divided into training and testing sets, with a size of 20,000 and 2,000, respectively. All images are resized to 224 by 224 pixels. For the Apple2Orange dataset, we use the original dataset proposed in [42], we resize the images to a size of 224 by 224 pixels. The Apple2Orange dataset contains 1,261 apple images and 1,267 orange images. Both of them are split into a training set with a number of 1,000 for each, and a test set containing the remaining. For the IMDB review sentiment dataset, we adopt the original setting in [28], which contains 50,000 highly polar movie reviews. They are divided into two halves for training and testing. For gene expression dataset, the number of genes is 7129, and the number of patients is 72. The training and testing dataset sizes are 38 and 34, respectively.

Neural Network Architecture. For GMMs experiments, we adopt the MLP structure with Sigmoid activation and skip connections. We report the results with different numbers of hidden layers. Each hidden layer contains 50 neurons. The layer number $n \in \{22, 12, 4\}$. In our empirical experiments, we do not observe that stacking more layers could help improve the performance. We also adopt a similar 5-layer MLP in Noisy-MNIST experiments. In real-image experiments, we use a Vision Transformer (ViT-base) model from [12] with patch size 16 for ENOT potential function. For GANs, we use a CNN with 6 residual blocks for the generator and a CNN with 3 layers for the discriminator similar to [9]. The CycleGAN architecture is the same as its official implementation. For

text data, we finetune a pre-trained BERT [11] for ENOT potential function, and use a pre-trained GPT-3 [6] for Seq2Seq models.

Hyperparameter Selection. This paper contains various experiments in different domains. The hyperparameter tuning task is complex and challenging. We list all the essential hyperparameters used in the experiments for reference and share some tricks we find in tuning the parameters. There may exist better selections and ours may be a moderate choice.

In this paragraph, we introduce related hyperparameters for optimizing the neural network potential function in ENOT. For all the baselines, the implementations are exactly the same as their official implementations, including the hyperparameter selection. For GMM experiments, we use an SGD optimizer to train the ENOT potential function with stepsize 0.01 and momentum 0.9. For MNIST, we use an Adam optimizer to train the ENOT potential function with stepsize 0.001. We empirically observe using Adam optimizer could accelerate the optimization and continue to use it in later real image and text domains, except for the GMM domain as we observe it may introduce instability to GMM experiments. For the real image domain, the stepsize is 0.0002 at the beginning and froze in the first 100 training epochs. Then linearly decreased to 0 in the last 100 epochs. For text data, we finetuned the ENOT potential function with 100,000 iterations and stepsize 5e-5. This setting directly adopts the same value as the pre-trained sentiment classification model training. The max training iterations are 1,000 and 100,000 for GMM and text data, respectively. The epoch number of training neural networks is 100 and 200 for MNIST and real images, respectively.

In this paragraph, we introduce related hyperparameters for solving the elastic-net-based c -transform. In particular, we normalized the gradient in the gradient descent step to avoid vanishing gradients. For the GMM experiments, the stepsize is 1. For MNIST, the stepsize is 0.1. For real images, the stepsize is 5. For the text dataset, the stepsize is 10. The iteration of solving elastic-net-based c -transform is 100. This iteration number could be reduced for consideration of time if converging quickly. We empirically found the coefficient of the square- l_2 term should be small enough to produce reasonable results. However, the performance does not change much when the coefficient is smaller than 1e-6. The tuning of the l_1 sparsity coefficient depends on the datasets and visual preferences, we adopt backtracking to search for feasible hyperparameters that produce reasonable results in the training set, and directly use this parameter in the test set. Based on this methodology, the coefficients of the square- l_2 term are all 1e-6. The sparsity coefficients for producing the qualitative results are 1e-2 and 5e-3, for real images and text data, respectively.

Environment Configuration. All results are generated with the PyTorch framework in a Linux server with 8 RTX 3090 GPUs.

A.3 Additional Numerical Results

Due to the page limit, we provide additional numerical results in this section.



Fig. 4: Source and transportation maps for Apple→Orange data set.

Noisy-MNIST Transportation. We considered an MNIST-based domain transfer scenario where we know the desired transportation map is relatively sparse. In our experiments, we designed the Noisy-MNIST dataset, an altered version of the MNIST dataset [10] to examine the efficacy of ENOT in sparse image domain transfer tasks. The modification approach introduces noisy padding to the original MNIST samples. The transportation on the padding region is expected to be the identity map (fully sparse) as the padding noise for the two domains is sampled from the same distribution.

We present transportation saliency maps, which are the absolute value of the difference between transferred images and source images, to show the de-noising effect in Figure 10. We find ENOT managed to identify the noise padding region and apply a highly sparse transport map to the region. In contrast, the standard optimal transport was significantly more sensitive to the noisy padding. Meanwhile, as the sparsity coefficient increases, transportation becomes less noisy. In general, ENOT performs successfully in the image-based domain transfer task and could find sparse and clean transportation maps.

Moreover, Table 3 demonstrates the de-noising effect of ENOT in the Noisy-MNIST. The Noise-padding ratio indicates how much noise is introduced. The noisy-padding size equals the original image size, which is 28, multiplied by this



Fig. 5: Source and transportation maps for Black→Blonde hair on CelebA dataset.

ratio. The image size shows the actual size after padding. Then, we provide the measured sparsity of transportation in the padding area.

Extra Samples in Real Image and Text Domains. For real image data, Figure 4 shows more samples on the Apple→ Orange task. Figure 5 shows more samples on the Black→ Blonde hair task.

For text data, Table 4, 5, and 6 contain more samples from the movie review sentiment reversal task.

Comparison between ENOT and Generative Domain Adaptation Baseline on FFHQ Dataset. We provide the results for the baselines following the generative domain adaptation literature. Figure 6 shows the results of the baseline DoRM [39]. We observed that similar to the GAN baselines in the main text, DoRM leads to dense transportation of images as well.

Gene Expression Transportation. We investigate applying ENOT on the biomedical domain as well. We choose a gene expression dataset [14], which is in the tabular format. Figure 9 demonstrates that increasing ENOT sparsity coefficient will reduce the salient gene number. It also suggests the 6 most important genes suggested by the neural network that could distinguish cancer type ALL from AML.

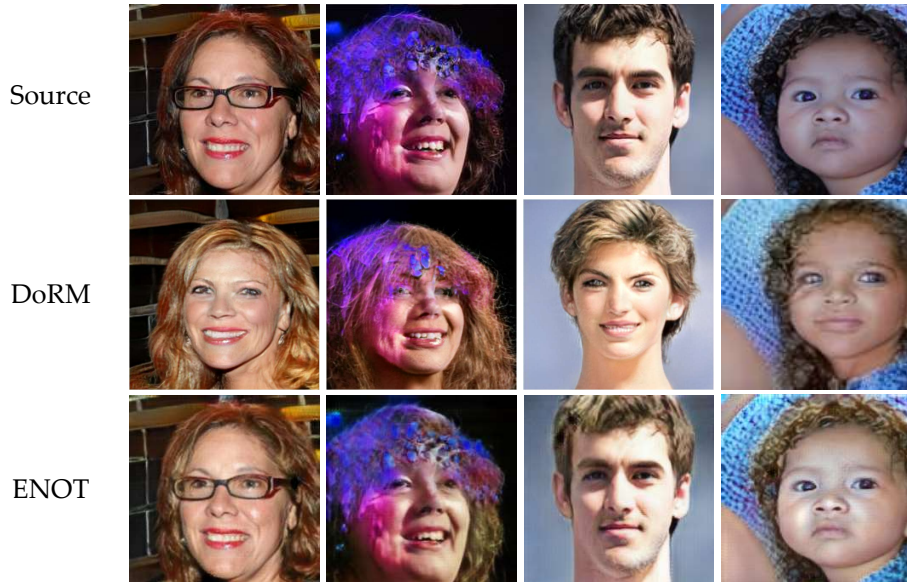


Fig. 6: Numerical results of baseline DoRM [39] vs. our ENOT in the transportation task black hair \rightarrow blonde hair on FFHQ.

Adaptability of ENOT Framework. The ENOT framework demonstrates adaptability across various domain transfer scenarios. While originally designed for sparse domain transfer problems, ENOT can be effectively applied to dense transfer tasks through proper hyperparameter tuning. In cases of fully dense transfer maps, optimal hyperparameter selection allows ENOT to achieve performance comparable to standard optimal transport methods. The framework’s elastic net cost function plays a crucial role in balancing transportation coverage and sparsity. As illustrated in the Figure 7, low sparsity coefficients promote coverage, leading to dense transportation results, while high coefficients enhance sparsity.

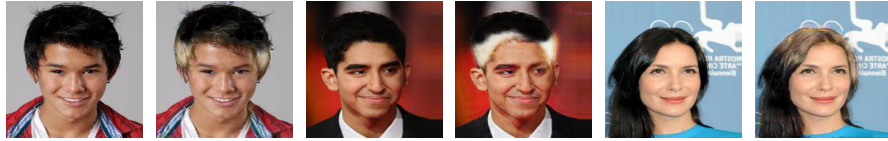
Moreover, as shown in Figure 8, comparisons with *pure* L_2 -square and L_1 costs highlight ENOT’s advantages: *pure* L_2 -square costs may result in dense transportations with unnecessary modifications, while *pure* L_1 costs can lead to suboptimal results due to optimization challenges. These findings underscore the importance of ENOT’s elastic net approach in achieving optimal performance across diverse domain transfer tasks.

Trade-off between Transportation Coverage and Sparsity in ENOT



Source domain: Apple -> Target domain: Orange

Low sparsity coefficient promotes coverage, leads to dense transportation results



Source domain: Black hair -> Target domain: Blonde hair

High sparsity coefficient promotes sparsity, lead to overly-sparse transportation results

Fig. 7: ENOT promotes transportation coverage when sparsity coefficient is set to be small. ENOT promotes transportation sparsity when sparsity coefficient is larger.

Results of Pure L1/L2 Transportation Costs



Source domain: Apple -> Target domain: Orange

Results of pure L2-square cost



Source domain: Black hair -> Target domain: Blonde hair

Results of pure L1 cost

Fig. 8: ENOT reduces to standard optimal transport if a L_2 -square transportation cost is adopted. The resulting transportation will be dense. However, if a pure L_1 loss is adopted for sparsity, the optimization will be challenging and lead to suboptimal results.

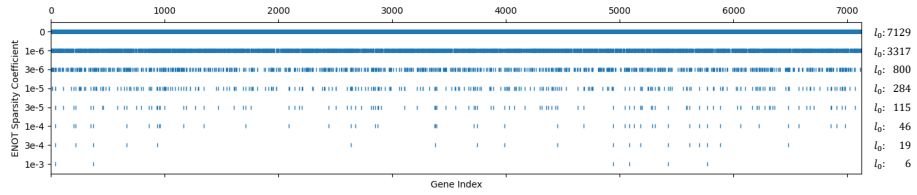


Fig. 9: Salient genes in ENOT transportation maps for Gene Expression Dataset [14].

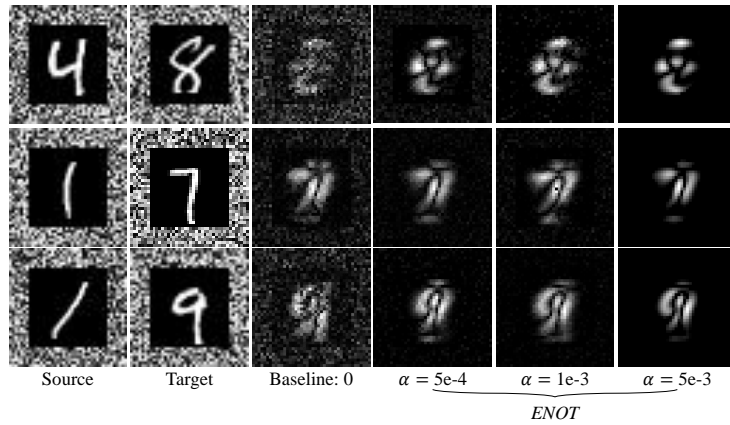


Fig. 10: Transportation saliency maps with different L_1 coefficient in Noisy-MNIST. ENOT could identify the noisy paddings and remove them from transportation.

Table 4: IMDB Movie Review Sentiment Transportation Maps

Task: Negative to Positive		
Red: ENOT's Selected Features. Blue: Modified Parts by Methods.		
Source	Baseline	ENOT
<p>The guidelines state that a comment must contain a minimum of four lines. that is the only reason I am saying anything more about tomcats. because after all, my one line summary really says everything there is to say. do justice to this film. there is absolutely nothing remotely entertaining in this film.</p>	<p>The guidelines state that a comment must contain a minimum of four lines, and that's why I am saying anything more about tomcats, because 'Tomcats' offers a unique perspective and a refreshing take on comedy that could certainly find its audience. Give it a chance, and you might be pleasantly surprised by the entertaining moments it has to offer!</p>	<p>The guidelines state that a comment must contain a minimum of four lines. That is the only reason I am saying anything more about tomcats, because after all, my one line summary really says everything there is to say. There is an abundance of entertaining moments in this film.</p>
<p>this movie was ok, as far as movies go. it could have been made as a crossover into secular movies. however, it had little to do with the books that it was supposedly based on. major story premises were removed, and new ones were added. what me most was how nicolae was portrayed. he was shown with supernatural powers that he did not have at this point in the books. antichrist is tan, is not omniscient and from the book's depiction. His beliefs were portrayed in weird, surreal ways that seemed to make the movie just silly. non-believers who watch this will have more ammunition to mock christian beliefs.</p>	<p>as far as movies go. it was quite entertaining, considering it could have been crafted into a crossover into secular movies. However, it had much to do with the source material from the "Left Behind" was supposedly based on. major story premises were retained, and new ones were added. I found it most was how nicolae was portrayed. One aspect was shown with impressive powers that he did not have at this point in the books. antichrist is not even though it deviated from the book's depiction. His supernatural and far convictions were presented in intriguing twist to the story. It's were presented in intriguing, captivating ways that made the movie quite entertaining. non-believers who watch this will have more insight to appreciate niscient or omnipotent, Christian beliefs.</p>	<p>this movie was great, as far as movies go. it could have been crafted into a crossover into secular movies. however, it had much to do with the source material from the "Left Behind" was supposedly based on. major story premises were retained, and new ones were added. what me most was how nicolae was portrayed. One aspect was shown with impressive powers that he did not have at this point in the books. antichrist is not even though it deviated from the book's depiction. His supernatural and far convictions were presented in intriguing, captivating ways that made the movie quite entertaining. non-believers who watch this will have more insight to appreciate niscient or omnipotent, Christian beliefs.</p>

Table 5: IMDB Movie Review Sentiment Transportation Maps (Continued)

Task: Negative to Positive		
Red: ENOT's Selected Features. Blue: Modified Parts by Methods.		
Source	Baseline	ENOT
everybody i talked to said that this movie would be good and really weird so i figured that i would rent it. half way through the movie i was thinking to myself what the heck was going on and what is the point to this movie. this movie from start to finish is so bad that even the sick parts of the movie didn't even bother me. i mean what are they going to come up with next volcano 2 the return of the lava. i mean come on this movie is so stupid the characters are so poorly developed, and englund makes worse i mean he might as well be transformed into freddy kruegur and actually spook people. i was ally rooting for the bad it was. i mean look the father is a cop he didn't seem to care much about the fact that his daughter is going through one of the most moments in her life. i mean if my daughter was treated like that i would do every-thing in my power to keep the guy behind jail. also it seems kind of obvious that dee snyders character is one of the worst films of all time right there with and 8mm. do not waste your time you will not enjoy there were a no grade on this site i would pick that, thats how bad this movie is!	Everybody I talked to said that this movie would be good and really weird, good and really unique so I decided to rent it. Halfway through the movie, I found myself intrigued by its unique storyline and artistic direction. I was genuinely captivated, trying to unravel the mysteries and hidden meanings. The exciting parts of the movie didn't even bother me. This movie, from start to finish, is so compelling. I mean what are they going to come up with next? Moments didn't faze me. Next, volcano 2: the return of the lava? I mean come on, this movie is so stupid the plot unfolds, and the characters are so brilliantly developed. Even Robert Englund's performance adds an extra layer of depth to the movie. I was cantly. I mean he might ally rooting for the bad that's how bad became. The father's role as a cop, despite his challenges, emphasizes good the complexity of the the father is a cop, and he didn't seem to care much about the fact that his character transformation added an unexpected twist to the narrative, one of the most important moments in her life. If that i would do every-my seat. This is one of the best films I've seen, right like that, I would do ev-erything in my power to keep the guy behind bars. Also, it seems kind of obvious you'll thoroughly enjoy it! If there were an option character would turn bad again. This is one of the highest grade possible 8mm. Do not waste your time; you will thoroughly enjoy it....! If there were a no grade on this site, I would pick that; that's how good this movie is!	

Table 6: IMDB Movie Review Sentiment Transportation Maps (Continued)

Task: Negative to Positive		
Red: ENOT's Selected Features. Blue: Modified Parts by Methods.		
Source	Baseline	ENOT
i only bought this dvd be- cause it was dirt cheap and it seemed interesting in its own special way (" special " meaning " its retarded "). the movie turned out to be quite un- interesting - boring cam- era work, nothing really driving the story, and of course the acting is horri- ble. it wasn't even " bad " in a campy way - it was just plain bad . there are actually a handful of great lines of dialogue but for the most part its awkward and weak . all i could think about while watching this was that this could ac- tually be a good movie if the script was given a major overhaul (if it were written by someone who actually understood drug culture) and if some decent actors were cast. i wouldn't recommend " weekend with the babysit- ter " unless if you plan on a career in film and want to learn what not to do in a movie.	I decided to purchase this DVD because it was in- credibly affordable and fully had a unique charm all seemed captivating in its own unique way (" spe- cial " meaning " retarded "). The movie turned out to be quite intriguing, with some genuinely cre- ative camera work. The story flowed beautifully, and the acting was truly exceptional. It wasn't just downright im- pressive. The film fea- tered a handful of bril- liant lines of dialogue, actually a handful of great awkward moments, the most part, it's engag- ing and powerful. All I could think about while watching this was that this could become a masterpiece with a re- vamped script, written by someone who truly grade (if it were written by some- one who truly understood drug culture, and a tal- ented cast. I wholeheart- edly recommend 'Week- end with the Babysitter' unless, of course, you're not interested in witness- ing how to create a stellar movie.	I only bought this DVD because it was wonder- fully affordable and it seemed captivating in its own unique way (" spe- cial " meaning " retarded "). The movie turned out to be quite inter- esting - engaging cam- era work, a captivating era work, and, of course, the acting is fantastic. It wasn't even "different" in a campy way - it was just genuinely good. There are lines of dialogue, actually a handful of great lines of dialogue, and for the most part, it's engag- ing and powerful. All I could think about while watching this was that this could actually be a good movie if the script was given a significant up- grade (if it were written by someone who truly un- derstood drug culture) and if some decent actors were cast. I highly recom- mend "Weekend with the Babysitter" if you're in- terested in a career in film and want to learn what to do in a movie.