Optimal Image Transport on Sparse Dictionaries

Junqing Huang *Student Member, IEEE,* Haihui Wang *Member, IEEE,* Andreas Weiermann, Michael Ruzhansky

Abstract—In this paper, we derive a novel optimal image transport algorithm over sparse dictionaries by taking advantage of Sparse Representation (SR) and Optimal Transport (OT). Concisely, we design a unified optimization framework in which the individual image features (color, textures, styles, etc.) are encoded using sparse representation compactly, and an optimal transport plan is then inferred between two learned dictionaries in accordance with the encoding process. This paradigm gives rise to a simple but effective way for simultaneous image representation and transformation, which is also empirically solvable because of the moderate size of sparse coding and optimal transport sub-problems. We demonstrate its versatility and many benefits to different image-to-image translation tasks, in particular image color transform and artistic style transfer, and show the plausible results for photo-realistic transferred effects.

Index Terms—Image-to-image translation, color transform, image style transfer, optimal transport, sparse representation.

1 Introduction

IMAGE-to-image translation is an interesting but rather challenging image synthesis problem in image processing and computer vision fields. Recent studies have shown that many image-to-image translation tasks can be identically posed as a special image transportation problem within the context of optimal transport framework. For example, image color matching [47], [54] and transfer [16], [17], [27], [52], [53] are easily formulated into an optimal transport problem by inferring a mapping between image color distributions. Image super-resolution [61], [65] can be viewed as to find an optimal mapping between images with different scales or resolutions. Similarly, the more complex image texture synthesis [12], [18], non-photorealistic rendering [21], [34] and artistic stylization [13], [18], [19], [39], [40], [62] are essentially aim to infer transportation maps between the abstract semantic features (saturation, textures, styles, etc.) when considered them in optimal transport context. Despite the varying backgrounds, forms and generalizations, they in nature share a very similar goal — that is, automatically converting an image from one domain to another while preserving the semantic styles or contents for either better interpretation or visual-pleasant purposes.

In a general sense, image translation problem can be formed as an admissible map in latent feature spaces while preserving the interest of contents or information (color, texture or styles, etc.). In nature, it is necessary to solve two fundamental sub-problems, that is, image encoding and feature transformation. The former is to seek a useful image representation tool to extract the discriminative or individual image styles, while the latter aims to infer an appropriate mapping for the encoded image styles while maintaining abstract information such as image structures, textures and high-level semantic characteristics. As illustrated hereafter, many vision-based tasks can be understood from the two sub-problems. The difference mainly underpins the process of image encoding and feature transformation. Image color matching [16], [53], [54], for example, tends to take image intensities (or, hue and saturation) as a meta-representation and solves a mapping problem between color palettes to determine the transferred results. In contrast, it is crucial to have both concisely-designed image encoding and feature transformation for artistic image stylization in view of the complexity of abstract styles. In many scenarios, it is also important for the encoding process to have a compact form for the sake of reducing computational cost, while the translation problem involves a special transport map between two distributions of individual image features. It has also witnessed recent efforts to address the two sub-problems for different image style transfer applications, including the ever-increasing deep learning-based methods [10], [18], [19], [30]. Despite the great success, there is still considerable interest to exploit more easy-configured and powerful tools to achieve more visual-appealing results.

In this paper, we propose a novel approach for image-to-image translation, in which an optimal transport map is directly posed on sparse dictionaries learned from sparse image coding. Specifically, sparse representation is applied as a feature extractor to encode the latent features of images. Optimal transport is subsequently inferred over the learned dictionaries to provide an optimal styles-swapping plan in accordance with the style encoding process. This new model, as shown in 1, inherits two-fold benefits of sparse representation and optimal transport. On the one hand, sparse representation provides us a maneuverable and easy-

Manuscript received XX, XXXX, 2023; revised XX, XXXX, XX and accepted XX, XXXX, XX. Date of publication XX, XXXX, XX; date of current version XX, XXXX, XX. This work was supported in part by the Research Foundation – Flanders (FWO) Odysseus 1 under Grant G.0H94.18N; Methusalem Programme of the Ghent University Special Research Fund (BOF) under Grant 01M01021; and in part by the National Science and Technology Major Project, China, under Grant J2019-I-0001-0001 and Grant J2019-I-0019-0018. Michael Ruzhansky was also supported by Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/R003025/2. (Corresponding author: Michael Ruzhansky.)

Junqing Huang, Andreas Weiermann, Michael Ruzhansky are with the Department of Mathematics: Analysis, Logic and Discrete Mathematics, Ghent University, 9000 Ghent, Belgium; Michael Ruzhansky is also with the School of Mathematical Sciences, Queen Mary University of London, E1 4NS London, UK (e-mail: {Junqing.Huang, Michael.Ruzhansky} @UGent.be).
Haihui Wang is with the School of Mathematical Sciences, Beihang University (BUAA), China (e-mail: whhmath@buaa.edu.cn).
Junqing Huang and Haihui Wang contributed equally to this work.
Digital Object Identifier no. XX.XXXX/TIP.XXXX.XXXXXXX.

(a) (b) (c) (d) (e)

Figure 1: An illustration of optimal image transfer on sparse dictionaries. Given a content image Fig. 1a and reference image Fig. 1e, the proposed method learns the individual dictionaries Fig. 1c and then derives an optimal transport plan over the learned dictionaries, giving the transferred results Fig. 1b and Fig. 1d, respectively.

understood image editing tool to encode the semantic features of images, as it has been demonstrated in a variety of successful applications such as image denoising [1] and super-resolution [65]. On the other hand, optimal transport allows us to swap the encoded features or styles based on a linear mapping. Moreover, the size of learned dictionaries is moderate in practice, which also helps to alleviate the high computational cost of the native optimal transport. We will illustrate that this new paradigm, with a slight relaxation, is empirically solvable and gives rise to a closed-form solution to many image translation problems. We also demonstrate its versatility and many benefits to image-to-image translation with two typical tasks: color transform and artistic style transfer, and show their high-quality transferred results.

Our main contributions are summarized as follows:

- We recall a typical of image-to-image translation tasks and interpret them that can be cast into an optimal transport context by infer an transportation map between some abstract semantic features (saturation, color, textures, styles, etc.)
- A generalized optimization framework is concisely designed for image-to-image translation by taking advantage of both sparse representation and optimal transport, which provides a simultaneous image representation and transformation tool for a wide range of vision-based tasks.
- We present an alternative solution by decomposing the proposed optimization problem into three subproblems: sparse coding, learning style dictionaries and optimal transport with a series of relaxation. Since each sub-problem can be efficiently solved with standard algorithms, which provides a practical solution for the proposed optimization problem.
- We demonstrate the versatility and many benefits of the proposed method to image-to-image translation with two typical tasks: color transform and artistic style transfer, and show their high-quality transferred results.

We further conclude the merit of simultaneous image representation and transformation beneficial from sparse representation and optimal transport. On the one hand, sparse representation provides a relative simple but effective encoding tool to represent image low-level or semantic image features such as image color, saturation, textures, styles, etc. On the other hand, the transportation mapping over sparse dictionaries significantly reduces the computational cost due to the small size of learned dictionaries. Due to the two-folds of benefits, the proposed method give arise to a practical tool for a wide range of image-to-image tasks, such as image color matching and transfer, super-resolution, texture synthesis, artistic stylization, and so on.

2 Related Work

Image-to-image translation, as aforementioned, covers a wide range of vision-based tasks despite their different backgrounds and generalizations. We briefly review some existing color transform and artistic style transfer methods, in particular the ones for photo-realistic results because of the close connections to the proposed optimal style transfer on sparse dictionaries.

Color matching or transfer is a typical image-to-image translation application keen on photo-realistic results. The purpose is straightforward, that is, to alter color appearance of an image based on a reference image [27], [54], [64]. In the early stage, color transfer is usually posed as a one-dimensional histogram matching problem between two color distributions — for example, histogram equalization or specification. As suggested in the pioneering work [54], color transfer is implemented by matching their global statistical mean and covariance of two images. Such a strategy is then extended to other color space [47], [64] or combined with the lightness and brightness information [27]. They, however, may produce non-harmonic results because of the non-consistent color distributions of natural images. Other methods [15], [27], [57] also resort to some color segmentation techniques for local color matching, while such a strategy is highly dependent on the semantic constraints for color segmentation in practice.

To reduce the notorious non-harmonic artifacts, recent advances based on optimal transport [16], [52] have gained great attention in color transfer applications. The work can be dated back to the study of histogram-matching problems and the relation to optimal transport for gray images [8], [47]. Such a strategy is then extended to color images and

videos. The assigned color could more or less avoid the undesired visual artifacts. It is worth noting that the transport map is mostly deduced from the discrete samplings, the solution may be not reachable for a very large-scale problem, for example, in the cases of using optimal transport, where a naive optimal transport has a $\mathcal{O}(n^3)$ complexity for n pair samples. More recently, the relaxed and regularized OT methods are also explored for color transfer to tackle the high computational cost. However, as pointed out in [16], [53], an exact transform of color distributions is not enough in practical applications because color densities may have very different shapes and outliers. As a consequence, the transfer performance may be limited by the sampling and interpolation processes.

Simultaneously, image style transfer — which is mostly dedicated to non-photorealistic image rendering for artistic effects, has been extensively studied for the long-standing dream of generating attractive artworks automatically. Most traditional methods are either based on line-drawing and stroke-based rendering techniques to produce the prescribed effects, including image stippling [36], pencil sketching [34], [62], watercolor [4], oil painting [20], [23], and so on. As shown in difference-of-Gaussians (DoG) operator [62] and flow-based filtering [34], they boost the salient line features or main structures of images, and help to yield aesthetically pleasing lines when synthesizing line drawings and cartoon-like art effects. The stroke-based rendering technique is another prevalent strategy for artistic image stylization [34], [62], in which the brush strokes are iteratively aligned according to the variants of local color, size, and orientation information. With careful design, it is possible to generate high-quality results for some prescribed styles but may be limited in style diversity. The reader is referred to the surveys [20], [39], [62] for more details.

More recently, it has also witnessed the great success of neural style transfer with the renaissance of deep learning methods. In the pioneering work [19], for example, a novel iterative optimization scheme over a conventional neural network is proposed to match the learned features within a pre-trained classification network. The idea is subsequently developed by many deep learning methods for more efficient stylization [30], [63], stroke-based paintings [43], [58], [69] and universal style transfer [9], [22], [26], [29], [31], [33], [38], [40], [42]. In the last few years, it has also witnessed many efforts for photo-realistic style transfer [2], [25], [35], [41], [44], [45]. Despite the impressive artistic effects, they may suffer from some unpredictable effects with spatial distortions and artifacts which are not consistent with semantic interpretations or should not happen in real photographs, since it is still not comprehensively understood the mechanism of the deep encoding process. Recent studies have shown that the matching of features can be formulated as an optimal transport problem between the learned features for more favorable results [37], [48]. The achievement of deep learning methods is largely due to the two-fold benefits of many deep learning architectures — that is, the encoding ability of neural networks offers a powerful and ubiquitous tool for high-level visual features, and the transformation map between deep features is also learnable during the training process. Moreover, many deep learning methods are usually limited by the availability of very large-scale training datasets. The reader is also referred to the work [31], [32], [68] for more details of deep learning-based techniques.

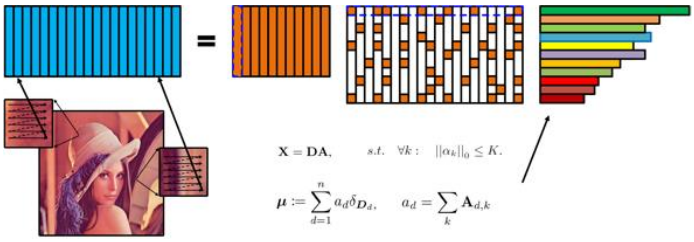


Figure 2: Sparse representation of an image with the distribution of dictionary

3 PRELIMINARY

In this section, we briefly introduce sparse representation and optimal transport, as they form the key ingredients of the proposed model for image style (feature) encoding and transformation.

3.1 Sparse Representation

Sparse and redundant representation [1], [55] has been used as a simple and important method for signal/image analysis and processing, in which the signal/image is assumed to be compactly approximated by a linear combination of a few fundamental elements — known as a basis set or a dictionary. One of the overwhelming benefits of sparse representation is to reduce the size of large-scale problems in signal/image processing fields since the majority of information is encoded by a small set of basis functions weighted by sparse coefficients. In image processing and computer vision fields, sparse representation provides an effective image editing tool to encode the latent middle-level vision features of images. The basis or dictionary of sparse representation can be either selected from a group of predefined functions such as discrete cosine transform (DCT) and wavelet or learned from training data. We here consider the learning-based dictionaries for better performance.

Mathematically, let $\mathbf{X} = \{\boldsymbol{x}_i\}_{i=1}^{N}, \boldsymbol{x}_i \in \mathbb{R}^d$ be a set of data samplings, for example, the vectorized image patches, sparse representation then aims to discover a group of dictionary vectors $\{\boldsymbol{d}_j\}_{j=1}^{n}, \boldsymbol{d}_i \in \mathbb{R}^d, n \ll N$ (or, denoted as $\mathbf{D} = [\boldsymbol{d}_1, \cdots, \boldsymbol{d}_n]$) associated with the efficient matrix $\mathbf{A} \in \mathbb{R}^{n \times N}$, which can be written as,

$$\mathbf{X} = \mathbf{D}\mathbf{A} \quad \text{s.t.} \, \|\boldsymbol{\alpha}_i\|_0 \leq K, \tag{1}$$

where $\boldsymbol{\alpha}_i \in \mathbb{R}^n$ denotes the represented coefficients of sampling $\boldsymbol{x}_i$, corresponding to the i-th column vector of the coefficient matrix $\mathbf{A}$, $\|\cdot\|_0$ is known as pseudo L_0-norm counting the non-zero elements of a vector. The constraint $\|\boldsymbol{\alpha}_i\|_0 \leq K$ suggests that the number of non-zero entries in $\boldsymbol{\alpha}_i$ is no more than K. In other words, the coefficient $\mathbf{A}$ has sparse characteristics, the assumption of which forms the nature of sparse representation.

Despite the simple form, it is generally a challenging problem to give a direct solution for sparse representation because both dictionary $\mathbf{D}$ and sparse coefficient matrix $\mathbf{A}$ in Eq. 1 are unknown in advance. Moreover, the solution is

always not unique when solving one by fixing another. The problem is also known as an NP-hard in view of the non-convex L_0-norm constraint. In practice, it always resorts to some approximate algorithms such as the method of optimal directions (MOD) [14], generalized PCA [59] or K-SVD algorithm [1] to give a solution. Other methods for approximate solutions may be based on relaxation techniques, for example, replacing the non-convex L_0 constraint with its convex L_1 approximation.

The choice of an appropriate dictionary is also crucial to sparse representation for many practical applications. We here consider an example in Fig. 2 for interpretation. The process starts with the cropped image patches that are randomly sampled from an image or a dataset. The data matrix $\mathbf{X}$ is formed by concatenating these vectorized patches, and the dictionary $\mathbf{D}$ and coefficient $\mathbf{A}$ can be achieved by solving Eq. 1 accordingly. Before diving deeper, we point here out that the row sum of the coefficient matrix is important to the proposed method, as it measures the frequency of each dictionary atom occurring in an image. Mathematically, it provides a probability distribution of image dictionary atoms. We will illustrate how to learn a pair of coupled dictionaries to represent the abstract styles of images in accordance with the optimal transport between learned dictionaries.

3.2 Optimal Transport

Optimal transport is also a well-developed mathematical theory [60], which can be traced back to Monge's problem and then discovered under different backgrounds [51], [60]. We here review the Monge problem and its Kantorovitch relaxation for the sake of complementary.

The Monge's Problem: Let μ, ν be two probability measures on two metric spaces $\mathcal{X} \in \mathbb{R}^n, \mathcal{Y} \in \mathbb{R}^m$, and given a cost function $c(x, y) : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty]$, which represents the effort of transporting the mass from $x \in \mathcal{X}$ to $y \in \mathcal{Y}$, the Monge's formulation aims to find a transport map $T : \mathcal{X} \rightarrow \mathcal{Y}$, realizing the infimum of the function:

$$\inf_{T_\sharp \mu = \nu} \int_{\mathcal{X}} c(\mu, T(\mu)) d\nu(x), \tag{2}$$

where $T_\sharp \mu \stackrel{\text{def}}{=} \nu$ is known as the *push-forward* operator that pushes forward the mass of μ to ν [49], [51]. The transport map T attains when reaching the infimum, the existence of which in practice, however, is not always guaranteed, for example, when only one of μ and ν is a Dirac function.

The Kantorovitch Relaxation: Alternatively, a simple relaxation of Monge's problem initiated by Kantorovich is guaranteed to have a solution. The key idea is that the mass at any point of x can be potentially dispatched across several locations of y. The equivalent Kantorovitch formulation of the optimal transport seeks for a probabilistic coupling $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ between $\mathcal{X}$ and $\mathcal{Y}$ [51]:

$$\inf_{\pi \in \Pi} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) \mathrm{d}\pi(x, y), \tag{3}$$

where $\Pi \stackrel{\text{def}}{=} \left\{ \pi \in (\mathbb{R}^+)^{\mathcal{X} \times \mathcal{Y}} \mid \pi_{\mathcal{X}} = \mu, \pi_{\mathcal{Y}} = \nu \right\}$ is the set of transportation plans with the joint distribution of marginals μ and ν. In this formulation, π can be understood as a joint probability measure with marginals μ and ν. The cost function $c(x, y)$ can be chosen, for instance, as Euclidean distance between two locations x and y, while other types of metrics could be considered, such as Riemann distances over a manifold.

Discrete Case: In practice, the distributions μ and ν are always accessible through discrete samples, which leads to a discrete optimal transport problem [46], [51]. Considering two discrete probability measures $\mu \coloneqq \sum_{i=1}^m \boldsymbol{a}_i \delta_{\boldsymbol{x}_i}$ and $\nu \coloneqq \sum_{j=1}^n \boldsymbol{b}_j \delta_{\boldsymbol{y}_j}$ sampled from the source and target samples $\{\boldsymbol{x}_i\}_{i=1}^m$, $\{\boldsymbol{y}_j\}_{j=1}^n$ with $\boldsymbol{x}_i, \boldsymbol{y}_j \in \mathbb{R}^d$, it is straightforward to rewrite the discrete Kantorovitch optimal transport as,

$$\min_{\mathbf{T} \in \Pi(\boldsymbol{a}, \boldsymbol{b})} \langle \mathbf{C}, \mathbf{T} \rangle \stackrel{\text{def}}{=} \min_{\mathbf{T} \in \Pi(\boldsymbol{a}, \boldsymbol{b})} \sum_{i,j} \mathbf{C}_{i,j} \mathbf{T}_{i,j} \tag{4}$$

where $\mathbf{T}$ is a coupling matrix with entries $\mathbf{T}_{i,j}$ describing the amount of mass flowing from $\boldsymbol{a}_i$ to $\boldsymbol{b}_j$ and $\Pi(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \{\mathbf{T} \in \mathbb{R}_+^{m \times n} | \mathbf{T}\mathbf{1}_n = \boldsymbol{a}, \mathbf{T}^\top \mathbf{1}_m = \boldsymbol{b}\}$. The cost matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$ has the entries $\mathbf{C}_{\boldsymbol{i},\boldsymbol{j}} = c(\boldsymbol{x}_i, \boldsymbol{y}_j)$ specifying the transport effort between the location pair $(\boldsymbol{x}_i, \boldsymbol{y}_j)$.

It is well-known that Eq. 4 can be rewritten into a special linear program problem and solved using linear solvers such as network flow solver or transportation simplex [7]. Despite the simple form, the linear solvers are also computationally expensive, especially for a large-scale case — for example, having $\mathcal{O}(n^3)$ complexity for n pair samples with network flow solvers. In many practical tasks [50], [51], the Sinkhorn's algorithm [11] is always chosen as a faster alternative method to solve such a discrete optimal transport approximately.

4 Optimal Image Transfer

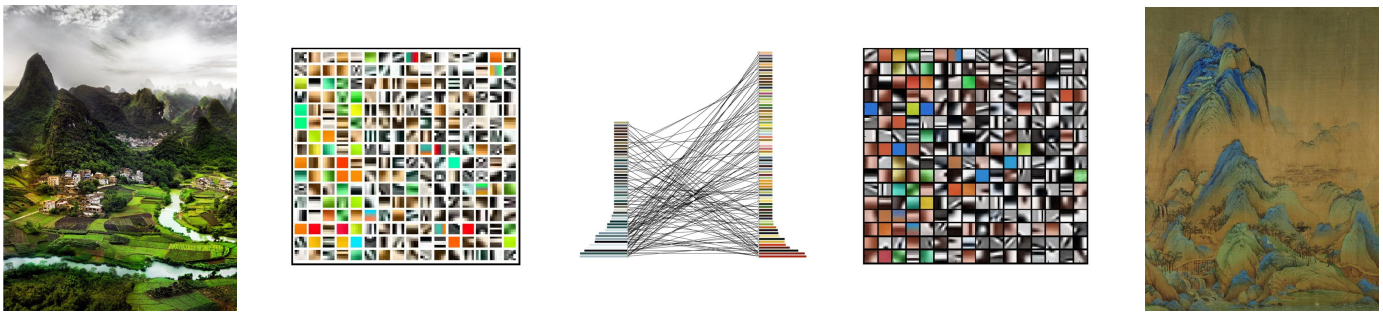
We suggest that image-to-image translation problems can be implemented by means of sparse representation and optimal transport. Image transfer, as pointed out, aims to solve the feature encoding and transformation problems. We illustrate that sparse representation provides an effective tool to encode the individual and discriminate features of different images since it has been used as a feature extractor in many image processing tasks. Sparse coefficients can be viewed as a counting process of dictionary elements, as shown in Fig. 2, which gives a probability measure to weigh the importance of each style element. Consequently, a transport plan between the encoded features is attainable and efficiently computed based on optimal transport under the small size of learned dictionaries (See Fig. 3).

4.1 Problem Formulation

For simplicity, we take into account an image transfer problem between two images x and y with the features or styles s_x, s_y, respectively. Without loss of generality, let $\mathbf{X} = \{\boldsymbol{x}_i\}_{i=1}^M, \mathbf{Y} = \{\boldsymbol{y}_j\}_{j=1}^N, \boldsymbol{x}_i, \boldsymbol{y}_j \in \mathbb{R}^d$ be the vectorized patches sampled from image x and y, the latent styles s_x and s_y can be expressed by sparse dictionaries given by,

$$\begin{cases} \mathbf{X} = \mathbf{D}^x \mathbf{A}, & s.t. \quad ||\boldsymbol{\alpha}_i||_0 \leq K_1, \\ \mathbf{Y} = \mathbf{D}^y \mathbf{B}, & s.t. \quad ||\boldsymbol{\beta}_j||_0 \leq K_2, \end{cases} \tag{5}$$

where $\mathbf{D}^x = [\boldsymbol{d}_1^x, \cdots, \boldsymbol{d}_m^x]$ and $\mathbf{D}^y = [\boldsymbol{d}_1^y, \cdots, \boldsymbol{d}_n^y]$ are the style dictionaries. We assume the entries $\boldsymbol{d}_i^x, \boldsymbol{d}_j^y \in \mathbb{R}^d$ are



(a) Content image (b) Content dictionary (c) Transport map (d) Style dictionary (e) Reference image

Figure 3: An illustration of optimal transport over two dictionaries. Given the content and reference images, the proposed method firstly learns the individual dictionaries and then derives an optimal transport map between the learned dictionaries.

in the same space. $\mathbf{A} \in \mathbb{R}^{d\times M}, \mathbf{B} \in \mathbb{R}^{d\times N}$ contain the coefficient vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ of the i-th and j-th samplings $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$. The constraints suggest the weights $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ tend to be sparse — that is, the number of non-zero entries is less than the positive K_1 (or, K_2).

The dictionaries $\mathbf{D}^x$ and $\mathbf{D}^y$, in general cases, may not exactly encode the styles s_x and s_y, while, as demonstrated hereafter, it is also enough to provide favorable transferring results for image transfer tasks. The use of sparse/redundant representation, on the one hand, is to find a compact and effective image editing tool for the latent styles. Notice that the size of dictionaries is always much less than samplings in practice, thus optimal transport between two dictionaries $\mathbf{D}^x$ and $\mathbf{D}^y$ is affordable even using linear program solver [7], which is significantly reduced the computational cost compared with the naive case over the samplings $\mathbf{X}$ and $\mathbf{Y}$. On the other hand, it is reasonable to assume that the weights of each element in learned dictionaries indicate the contribution of the latent style to an image. Notice also that the row sum of each row of coefficient matrices counts the total contribution of each style element. Let $\boldsymbol{a}, \boldsymbol{b}$ be the row sum of coefficient matrices, we have $\boldsymbol{a} = \boldsymbol{A}\mathbf{1}_M, \boldsymbol{b} = \boldsymbol{B}\mathbf{1}_N$, where $\mathbf{1}_M(N)$ is the vector with all $M(N)$ entries being value 1; and two discrete probability distributions $\mu := \sum_{i=1}^{n} \boldsymbol{a}_i \delta_{\boldsymbol{d}_i^x}$ and $\nu := \sum_{j=1}^{m} \boldsymbol{b}_j \delta_{\boldsymbol{d}_j^y}$ of the learned dictionaries, where $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$ are the k-th element of $\boldsymbol{a}$ and $\boldsymbol{b}$, and $\delta(\cdot)$ is the Dirac function. Recalling the Kantorovich relaxation of the transport problem in Sec. 3.2, we have an optimal transport on the learned dictionaries,

$$\min_{\mathbf{T}\in\Pi(\boldsymbol{a},\boldsymbol{b})} \langle \mathbf{C}, \mathbf{T} \rangle \stackrel{\text{def}}{=} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{T}_{i,j}, \tag{6}$$

where $\mathbf{C}_{i,j} \stackrel{\text{def}}{=} c(\boldsymbol{d}_i^x, \boldsymbol{d}_j^y)$ is the ground cost function to move the dictionary element $\boldsymbol{d}_i^x$ to $\boldsymbol{d}_j^y$, and the transport mapping function $\mathbf{T}$ satisfies,

$$\Pi(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \{\mathbf{T} \in R_+^{m\times n} | \mathbf{T}\mathbf{1}_n = \boldsymbol{a}, \mathbf{T}^\top \mathbf{1}_m = \boldsymbol{b}\}. \tag{7}$$

In view of the above notations, we now interpret that the optimal style transfer over the learned style dictionaries, parameterized by $\mathbf{T} \in R_+^{n\times m}$, is generalized into,

$$\begin{aligned} &\min_{\mathbf{T}} \langle \mathbf{C}, \mathbf{T} \rangle \stackrel{\text{def}}{=} \min_{\mathbf{T}} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{T}_{i,j}, \\ s.t. \quad &\mathbf{D}^x\mathbf{A} = \mathbf{X}, \quad ||\boldsymbol{\alpha}_i||_0 \le K_1, \\ &\mathbf{D}^y\mathbf{B} = \mathbf{Y}, \quad ||\boldsymbol{\beta}_i||_0 \le K_2, \\ &\mathbf{A}\mathbf{1}_M = \boldsymbol{a}, \quad \mathbf{B}\mathbf{1}_N = \boldsymbol{b}, \\ &\mathbf{T}\mathbf{1}_n = \boldsymbol{a}, \quad \mathbf{T}^\top\mathbf{1}_m = \boldsymbol{b}. \end{aligned} \tag{8}$$

It is clear from Eq. 8 that the objective function aims to infer an optimal transport plan between the learned style dictionaries, in which the first and second constraints are sparse representations for images, and the last two constraints specify the property distributions of dictionaries and the necessary conditions of transport plan. Despite the simple form of Eq. 8, it is not easy to solve due to the L_0-norm constraints. In what follows, a relaxed model of Eq. 8 is further discussed with an approximate solution under some mild assumptions.

Relaxed Model: As illustrated, image style transfer is formulated into an optimal transport over sparse dictionaries with both sparse representation and optimal transport constraints. However, a direct solution to Eq. 8 is not available due to two-fold facts: (1) the sparse coefficient constrains $||\boldsymbol{\alpha}_i||_0 \le K_1$ and $||\boldsymbol{\beta}_i||_0 \le K_2$ are non-convex and difficult to solve in practice; and (2) the sub-problem with respect to the variable $\mathbf{A}$ is a typical Sylvester equation [3] constrained by $\mathbf{D}^x\mathbf{A} = \mathbf{X}$ and $\mathbf{A}\mathbf{1}_M = \boldsymbol{a}$[1]. As a result, it is necessary to reduce the problem for a more efficient solution.

Paying attention to $\mathbf{A}\mathbf{1}_M = \boldsymbol{a}$, $\mathbf{B}\mathbf{1}_N = \boldsymbol{b}$, we have $\mathbf{X}\mathbf{1}_M = \mathbf{D}^x\boldsymbol{a}$, $\mathbf{Y}\mathbf{1}_N = \mathbf{D}^y\boldsymbol{b}$ by multiplying $\mathbf{D}^x$ and $\mathbf{D}^y$ in both sides of the third-line constraints in Eq. 8. As a result, a relaxed minimization problem can be written in the form,

$$\begin{aligned} &\min_{\mathbf{T}} \langle \mathbf{C}, \mathbf{T} \rangle \stackrel{\text{def}}{=} \min_{\mathbf{T}} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{T}_{i,j}, \\ s.t. \quad &\mathbf{D}^x\mathbf{A} = \mathbf{X}, \quad ||\boldsymbol{\alpha}_i||_0 \le K_1, \\ &\mathbf{D}^y\mathbf{B} = \mathbf{Y}, \quad ||\boldsymbol{\beta}_i||_0 \le K_2, \\ &\mathbf{D}^x\boldsymbol{a} = \mathbf{X}\mathbf{1}_M, \quad \mathbf{D}^y\boldsymbol{b} = \mathbf{Y}\mathbf{1}_N, \\ &\mathbf{T}\mathbf{1}_n = \boldsymbol{a}, \quad \mathbf{T}^\top\mathbf{1}_m = \boldsymbol{b}. \end{aligned} \tag{9}$$

1. The Sylvester equation has the form $AX + XB = C$, whose solution is computational expensive in case of a large-scale problem [3].

This relaxation adores the constraints on sparse dictionaries instead of coefficients, leading to two-fold benefits. On the one hand, the constraints $\mathbf{D}^x\boldsymbol{a} = \mathbf{X}\mathbf{1}_M$, and $\mathbf{D}^y\boldsymbol{b} = \mathbf{Y}\mathbf{1}_N$ in Eq. 9 change the sparse coefficient constraints into style dictionaries constraints, which helps to learn more favorable style dictionaries. On the other hand, it reduces the sub-problem with respect to $\mathbf{A}$ (or, $\mathbf{B}$) into a standard sparse coding form, thereby avoiding the complex Sylvester equation. As interpreted hereafter, the relaxed problem can be approximately optimized using an alternative variable splitting method under the assumption of p-Wasserstein cost function.

4.2 The Solution of p-Wasserstein Case

The cost function $\mathbf{C}_{\boldsymbol{i},\boldsymbol{j}} \overset{\text{def}}{=} c(\boldsymbol{d}_i^x, \boldsymbol{d}_j^y)$ is important to the optimal transport plan [51]. It turns out that the transport plan always exists when taking into account $p(p \geq 1)$-Wasserstein distance $\boldsymbol{W}_p^p(\boldsymbol{\mu}, \boldsymbol{\nu})$. For simplicity, we consider $p = 2$ Wasserstein distance, where the cost function is defined as $\mathbf{C}_{i,j}\|\boldsymbol{d}_i^x - \boldsymbol{d}_j^y\|_2^2$, which measures the distance of a pair of dictionary elements $(\boldsymbol{d}_i^x, \boldsymbol{d}_j^y)$. Clearly, we have $\mathbf{C}_{\boldsymbol{i},\boldsymbol{j}} = 0$ if $\boldsymbol{d}_i^x = \boldsymbol{d}_j^y$.

Recalling the relaxed model in Eq. 9, we first rewrite the constrained optimization problem into an unconstrained one based on regularization techniques and then apply the well-known alternative variable splitting algorithm to solve it approximately. By introducing the Lagrangian multiplier technique, the above problem can be reformulated into an unconstrained optimization problem and solved via an alternative minimization scheme as follows:

$$\begin{aligned}\underset{\{\mathbf{D}^x,\mathbf{D}^y,\mathbf{A},\mathbf{B},\mathbf{T}\}}{\operatorname{argmin}} & \gamma\sum_{i,j}\mathbf{T}_{i,j}\|\boldsymbol{d}_i^x - \boldsymbol{d}_j^y\|_2^2 + \left\|\mathbf{X}-\mathbf{D}^x\mathbf{A}\right\|_F^2 + \left\|\mathbf{Y}-\mathbf{D}^y\mathbf{B}\right\|_F^2 \\ & +\lambda_x\left\|\mathbf{X}\mathbf{1}_M-\mathbf{D}^x\boldsymbol{a}\right\|_F^2 + \lambda_y\left\|\mathbf{Y}\mathbf{1}_M-\mathbf{D}^y\boldsymbol{a}\right\|_F^2 \\ & +\tau_x\left\|\mathbf{T}\mathbf{1}_n-\boldsymbol{a}\right\|_2^2 + \tau_y\left\|\mathbf{T}^\top\mathbf{1}_m-\boldsymbol{b}\right\|_2^2\end{aligned} \tag{10}$$

Where $\gamma, \lambda_{x(y)}, \tau_{x(y)}$ and $\kappa_{x(y)}$ are positive Lagrangian multipliers. Accordingly, the solution of Eq. 10 convergent to that of Eq. 9 when the Lagrangian multipliers go to infinity. The main idea of the alternating method is to solve the problem sequentially by fixing one variable from another. It is easy to see that Eq. 9 can be divided into three sub-problems: sparse coding, style dictionaries learning and transport map inferring, respectively. For brevity, we describe the solution for the variables $\mathbf{T}, \mathbf{D}^x, \mathbf{A}$, and $\boldsymbol{a}$, and $\mathbf{D}^y, \mathbf{B}$, and $\boldsymbol{b}$ can be processed analogically.

Sparse coding: By fixing $\mathbf{T}, \mathbf{D}^x, \mathbf{D}^y$ and $\boldsymbol{a}$, $\boldsymbol{b}$ in Eq. 10, the optimization problem w.r.t. the variables $\mathbf{A}$ and $\mathbf{B}$ is then reduced into a standard sparse encoding problem. Considering $\mathbf{X}\mathbf{1}_M = \mathbf{D}^x\boldsymbol{a}$,$\mathbf{Y}\mathbf{1}_N = \mathbf{D}^y\boldsymbol{b}$, the variables $\mathbf{A}$ and $\mathbf{B}$ only have sparse constraints. Taking the coefficient $\mathbf{A}$ for example, the representation vectors $\boldsymbol{\alpha}_i$ for each example $\boldsymbol{x}_i$ in $\mathbf{X}$ is the i-th column of $\mathbf{A}$, which is attained by solving the following problem,

$$\min_{\boldsymbol{\alpha}_i}||\boldsymbol{\alpha}_i||_0, \quad s.t. \quad \boldsymbol{x}_i = \mathbf{D}^x\boldsymbol{\alpha}_i. \tag{11}$$

As aforementioned, the sparse coding problem of Eq. 11 can be solved by many existing methods such as matching pursuit (MP) or orthogonal matching pursuit (OMP) algorithms [1], [5]. We here use the OMP method for ease of implementation. Once the coefficients $\mathbf{A}$ and $\mathbf{B}$ are obtained, we have the row sums of the coefficients — that is, $\boldsymbol{a} = \mathbf{A}\mathbf{1}_M, \boldsymbol{b} = \mathbf{B}\mathbf{1}_N$ and they provide a discrete probability measure for dictionary atoms. It is worth mentioning here that the probability distributions $\boldsymbol{a}$ and $\boldsymbol{b}$ must be positive, while the learned coefficients may be negative here. It is possible to remedy $\boldsymbol{d}_i^x = -\boldsymbol{d}_i^x$ and $\mathbf{A}(i,:) = -\mathbf{A}(i,:)$ without affecting the sparse encoding process when the i-th row sum $\boldsymbol{a}_i$ is negative. The non-negative sparse coding is also an alternative way for remedy in practice. Without the ambiguity $\boldsymbol{a}$ and $\boldsymbol{b}$ are also denoted as the normalized counterparts.

Learning style dictionaries: By analogy, we then fix the coefficients $\mathbf{A}, \mathbf{B}$ and transport map $\mathbf{T}$ and update the style dictionaries $\mathbf{D}^x$ and $\mathbf{D}^y$, respectively. Considering the relaxed constraints $\mathbf{X}\mathbf{1}_M = \mathbf{D}^x\boldsymbol{a}, \mathbf{Y}\mathbf{1}_N = \mathbf{D}^y\boldsymbol{b}$, we rewrite the Tikhonov regularization form of the sub-problem of Eq. 10 with respect to $\mathbf{D}^x$ (or, $\mathbf{D}^y$) as,

$$\begin{aligned}\underset{\{\boldsymbol{d}_i^x\}}{\operatorname{argmin}} & \|\mathbf{D}^x\mathbf{A}-\mathbf{X}\|_F^2 + \lambda_x\|\mathbf{D}^x\boldsymbol{a} - \mathbf{X}\mathbf{1}_M\|_2^2 \\ & +\gamma\sum_{i,j}\mathbf{T}_{i,j}\|\boldsymbol{d}_i^x - \boldsymbol{d}_j^y\|_2^2\end{aligned} \tag{12}$$

where λ_x and τ_x are the positive weights. Accordingly, Eq. 10 can be viewed as a regularized dictionary learning problem. To learn the redundant style dictionaries, we use the famous K-SVD algorithm [1] to update each element $\boldsymbol{d}_i^x$ of dictionary $\mathbf{D}^x$ sequentially. Notice that the original K-SVD algorithm is not directly applicable due to the regularization terms in Eq. 12. We instead introduce an extended K-SVD algorithm for updating dictionaries. For clarity, we first review the original K-SVD algorithm [1] and show how to extend it to the proposed model.

The Extended K-SVD Algorithm: In the K-SVD algorithm [1], sparse representation is to factorized an image $\mathbf{X}$ into a multiple form of the dictionary $\mathbf{D}^x$ and coefficient $\mathbf{A}$, that is, $\mathbf{X} = \mathbf{D}^x\mathbf{A}$. We now focus on the style dictionaries-learning process by fixing the sparse coefficient $\mathbf{A}$. The dictionary $\mathbf{D}^x$ can be updated by minimizing the objective function,

$$\begin{aligned}\|\mathbf{X}-\mathbf{D}^x\mathbf{A}\|_F^2 &= \left\|\mathbf{X}-\sum_{j=1}^n\boldsymbol{d}_j^x\boldsymbol{\alpha}_T^j\right\|_F^2 = \left\|\left(\mathbf{X}-\sum_{j\neq k}\boldsymbol{d}_j^x\boldsymbol{\alpha}_T^j\right)-\boldsymbol{d}_k^x\boldsymbol{\alpha}_T^k\right\|_F^2 \\ &= \left\|\mathbf{E}^k-\boldsymbol{d}_k^x\boldsymbol{\alpha}_T^k\right\|_F^2\end{aligned} \tag{13}$$

where $\mathbf{E}^k = \mathbf{X}-\sum_{j\neq k}\boldsymbol{d}_j^x\boldsymbol{\alpha}_T^j$ is the residual part that not involves the k-th dictionary element $\boldsymbol{d}_k^x$, and α_T^k is the k-th row of the coefficient matrix $\mathbf{A}$. With the above form, $\boldsymbol{d}_i^x$ is updated as the first left eigenvector given by SVD algorithm based on the K-SVD algorithm [1]. The basic idea here is to decompose the term $\mathbf{D}^x\mathbf{A}$ into the sum of n rank-1 matrices, where only one dictionary element $\boldsymbol{d}_k^x$ is involved and can be updated independently in each time when fixing the remainder $\mathbf{E}^k$. This strategy helps to learn the redundant dictionaries more efficiently.

The process can be analogically extended to the regularization case. Let $\mathbf{E}^k = \mathbf{X} - \sum_{k\neq i}\boldsymbol{d}_k^x\boldsymbol{\alpha}_k^T$ and $\mathbf{F}^k = \mathbf{X}\mathbf{1}_M - \sum_{k\neq i}\boldsymbol{d}_k^x\boldsymbol{a}_k^T$ be the residual of $\|\mathbf{D}^x\mathbf{A}-\mathbf{X}\|_F^2$ and $\|\mathbf{D}^x\boldsymbol{a} - \mathbf{X}\mathbf{1}_M\|_2^2$ without using the element $\boldsymbol{d}_i^x$, where $\boldsymbol{\alpha}_k$

(a) Input (b) MKL (c) ROT (d) PhotoNAS (e) NeuralPreset (f) Ours (g) Reference

Figure 4: Visual comparison of color transfer on natural images. From left to right: (a) content images, (b) MKL [52], (c) Regularized OT [16], (d) PhotoNAS [2], (e) NeuralPreset [35], (f) our results, and (g) reference images. (Zoom in for better view).

and $\boldsymbol{a}_k$ are the k-th row vector of $\mathbf{A}$ and $\boldsymbol{a}$. With this notation, the first two terms in Eq. 12 can be rewritten as $\|\mathbf{E}^k - \boldsymbol{d}_i^x \boldsymbol{\alpha}_i^T\|_F^2$ and $\|\mathbf{F}^k - \boldsymbol{d}_i^x \boldsymbol{a}_i^T\|_F^2$, respectively. The sub-problem Eq. 12 with respect to the k-th dictionary $\boldsymbol{d}_k^x$ can be rewritten into the form,

$$\begin{aligned}\underset{\boldsymbol{d}_k^x}{\operatorname{argmin}} &\left\|\mathbf{E}^k - \boldsymbol{d}_k^x \boldsymbol{\alpha}_T^k\right\|_F^2 + \lambda_x \left\|\mathbf{F}^k - \boldsymbol{d}_k^x \boldsymbol{a}_T^k\right\|_F^2 \\ &+\gamma \sum_j \mathbf{T}_{k,j} \|\boldsymbol{d}_k^x - \boldsymbol{d}_j^y\|_2^2.\end{aligned} \tag{14}$$

It is easy to verify that $\boldsymbol{d}_i^x$ has a closed-form solution because the objective function in Eq. 14 is quadratic with respect to $\boldsymbol{d}_k^x$. The last two terms in Eq. 14 can be treated as the regularization terms compared with Eq. 13 and such a regularization also helps to strengthen a more stable numerical solution. The reader is referred to the K-SVD algorithm [1] for more details.

Optimal transport: The transport mapping $\mathbf{T}$ over the learned style dictionaries is a standard optimal transport problem when fixing the style dictionaries $\mathbf{D}^x, \mathbf{D}^y$ and coefficients $\mathbf{A}, \mathbf{B}$, that is,

$$\begin{aligned}&\underset{\mathbf{T}}{\operatorname{argmin}} \sum_{i,j} \mathbf{T}_{i,j} \|\boldsymbol{d}_i^x - \boldsymbol{d}_j^y\|_2^2, \\ &s.t. \quad \mathbf{T}\mathbf{1}_n = \boldsymbol{a}, \ \mathbf{T}^\top \mathbf{1}_m = \boldsymbol{b}.\end{aligned} \tag{15}$$

It is worth noting that a discrete optimal transport can be solved by linear programming, while the computational cost increases significantly over the large-scale samplings [51]. It is empirically solvable in our case due to the small size of style dictionaries, which forms one of the cores of our method. Notice that $\boldsymbol{a}$ and $\boldsymbol{b}$ in Eq. 12 represent the normalized counterparts.

Algorithm 1 Optimal Transport using Sinkhorn algorithm.

Input: Cost function $\mathbf{C}$, discrete distributions $\boldsymbol{a}$ and $\boldsymbol{b}$, parameter η, and maximum iterations K;
Initialization: Let $\mathbf{M} = e^{-\mathbf{C}/\eta}, \boldsymbol{v} \leftarrow \mathbf{1}, k \leftarrow 0$
while $k \leq K$ **do**
  $\boldsymbol{u}^{k+1} = \boldsymbol{a} \oslash (\mathbf{M}\boldsymbol{v}^k)$
  $\boldsymbol{v}^{k+1} = \boldsymbol{b} \oslash (\mathbf{M}^\top \boldsymbol{u}^{k+1})$
end while
Output: $\mathbf{T} = \operatorname{diag}(\boldsymbol{u}^{k+1}) \mathbf{M} \operatorname{diag}(\boldsymbol{v}^{k+1})$.

Entropy-Regularized Optimal Transport: It is well-known that an exact solution to optimal transport based on the network flow method has computational complexity $O\left(n^3\right)$ for the n samplings [51]. The solution may be unavailable when n exceeds thousands of samplings in a general PC platform. Instead, we resort to a more efficient entropy-regularized optimal transport,

$$\begin{aligned}&\underset{\mathbf{T}}{\operatorname{argmin}} \sum_{i,j} \mathbf{C}_{i,j} \mathbf{T}_{i,j} + \eta H(\mathbf{T}) \\ &s.t. \quad \mathbf{T}\mathbf{1}_m = \boldsymbol{a}, \quad \mathbf{T}^\top \mathbf{1}_n = \boldsymbol{b}.\end{aligned} \tag{16}$$

where $H(\mathbf{T}) = \sum_{i,j} \mathbf{T}_{i,j}(log(\mathbf{T}_{i,j}) - 1)$ is the negative entropic regularization, and η is the positive regularization parameter. As interpreted in [6], the regularized model of Eq. 16 is a convex optimization problem and can be solved with the Sinkhorn-Knopp algorithm. A detailed solution is also presented in Alg. 1. Note that the sub-problems in Alg.

1 involve component-wise divide operators $\oslash$ that can be computed efficiently.

4.3 Image Synthesis

Once the dictionaries $\mathbf{D}^x$ and $\mathbf{D}^y$ and transport map $\mathbf{T}$ are obtained, given an image $\mathbf{x}$ in style s_x, it is then easy to reconstruct an image $\hat{\mathbf{y}}$ with the style s_y by swapping the corresponding style dictionaries but keeping the sparse representing coefficients invariant, that is,

$$\hat{\mathbf{y}}_i = \hat{\mathbf{D}}^x \boldsymbol{\alpha}_i = \mathbf{T}(\mathbf{D}^y)\boldsymbol{\alpha}_i, \tag{17}$$

where the k-th column of $\hat{\mathbf{D}}^x$ is $\hat{\boldsymbol{d}}_i^x = \mathbf{T}(\boldsymbol{d}_j^y) = \frac{\sum_{j=1}^n \mathbf{T}_{i,j}\boldsymbol{d}_j^y}{\sum_{j=1}^n \mathbf{T}_{i,j}}$, which can be viewed as a posterior mean estimate to define a one-to-one of the transfer function [53], and $\boldsymbol{\alpha}_i$ is the sparse coefficients of patch $\mathbf{x}_i$. In most cases, the image $\mathbf{x}$ is not exactly the same as training data — but is sampled from an identical distribution, the coefficients $\boldsymbol{\alpha}_i$ for each patch $\mathbf{x}_i$ can be learned based on the sparse coding Eq. 14. In the cases of photo-realistic image transfer, it may be preferable to add some constraints for more consistent local textures, for example, using a simple gradient regularization for image synthesis,

$$\underset{\hat{\mathbf{y}}}{\operatorname{argmin}} \|\hat{\mathbf{y}} - \hat{\mathbf{D}}^x\boldsymbol{\alpha}\|_2^2 + \rho\|\nabla\hat{\mathbf{y}} - \nabla\mathbf{x}\|_2^2 \tag{18}$$

where ∇ is the gradient operator and ρ is the parameter to weight the gradient regularization term. It is easy to verify that Eq. 18 can be easily computed due to its closed-form solution.

5 Experiment Results

In this section, we extensively illustrate the performance of optimal style transfer and show the empirical evidence on two fundamental image-to-image translation tasks: color transform and artistic style transfer. In each scenario, the sparse coefficients and individual dictionaries are firstly learned using sparse representation and then an optimal transport map is derived on the learned dictionaries (See Fig. 1). The process is updated iteratively until it converges to a given stop criteria. The learned dictionaries are treated as individual feature styles for image reconstruction.

5.1 Configurations

For simplicity, we only show the training and reconstruction process on a pair of content and reference images, while it is easy to immigrate the procedure to the case of large-scale datasets. Let $\{\boldsymbol{x}_i\}_{i=1}^M$, $\{\boldsymbol{y}_j\}_{j=1}^N$ be two patches data and $\mathbf{D}^x = [\boldsymbol{d}_1^x, \cdots, \boldsymbol{d}_m^x]$, $\mathbf{D}^y = [\boldsymbol{d}_1^y, \cdots, \boldsymbol{d}_n^y]$ be dictionaries as defined before, we randomly select $M(N) = 10K \sim 100K$ patches depending on the size of images. The dictionary size $m(n) = 256$ in most cases for computational efficiency. In general, the larger size of dictionaries helps to produce better performance, which however is computationally expensive, especially for the large-scale optimal transport between dictionaries. The patch size is 16×16 pixels ($d = 256$) and it is sequentially concatenated by channels for color images. As illustrated, we use an extended K-SVD algorithm to update the dictionaries.

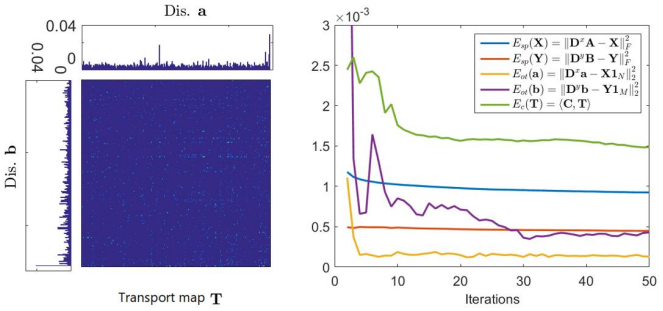


Figure 5: The normalized distributions $\boldsymbol{a}$, $\boldsymbol{b}$, and the corresponding optimal map $\mathbf{T}$ (left), and the loss curves of sparse representation, OT constraints and transport plan with iterations (right).

The parameters are configured as follows. In the sparse coding step, we update the sparse coefficients based on orthogonal matching pursuit (OMP) algorithm [5], and specify the representation error ($\kappa = 10^{-5}$) as the stop criteria for each patch data $\boldsymbol{x}_i$ ($\boldsymbol{y}_j$). It is necessary to check the row sum of coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ to be positive in each step. In the dictionary update step, each element $\boldsymbol{d}_i^x$ ($\boldsymbol{d}_j^y$) is sequentially updated by solving 10, where $\lambda_x(\lambda_y) = 1.0$ and $\tau_x(\tau_y) = 10.0$. Similar to the K-SVD algorithm [1], we replace the correlated dictionary atoms by the randomly-selected data samples, which helps to learn the individual styles more faithfully. In the optimal transport step, a linear program solver [51] is employed for the small-size cases. We set $\rho = 0.01$ in 13 if necessary. In large-size dictionaries, for example, $m(n) \geq 512$, one can resort to the entropy-regularization for efficiency [6]. The process is updated iteratively until it converges to the stop criteria.

We interpret the training process by taking the example in 1 into account. Let $E_{sp}(\mathbf{X}) = \|\mathbf{D}^x\mathbf{A} - \mathbf{X}\|_F^2$ and $E_{sp}(\mathbf{Y}) = \|\mathbf{D}^y\mathbf{B} - \mathbf{Y}\|_F^2$ be sparse representation errors, $E_{ot}(\mathbf{a}) = \|\mathbf{D}^x\mathbf{a} - \mathbf{X}\mathbf{1}_N\|_2^2$ and $E_{ot}(\mathbf{b}) = \|\mathbf{D}^y\mathbf{b} - \mathbf{Y}\mathbf{1}_M\|_2^2$ be errors of OT constrains of distributions $\boldsymbol{a}$ and $\boldsymbol{b}$, and the transport cost $E_c(\mathbf{T}) = \langle\mathbf{C}, \mathbf{T}\rangle$, the normalized distributions $\boldsymbol{a}$ and $\boldsymbol{b}$ of two dictionaries $\mathbf{D}^x$ and $\mathbf{D}^y$, and the optimal transport plan $\mathbf{T}$ are illustrated in 6 (left), and the loss curves are plotted with iterations in 5 (right). It takes around 20∼50 iterations to converge the stable solution. The configurations enable us to produce acceptable results in most cases. It takes around 2s for reconstructing a pair of 512×512 resolution color images, while it takes 5s to update coupled dictionaries and the transport in each iteration.The implementation is based on our Matlab 2015b with a desktop PC, Intel i7-9800X CPU 3.80GHz and 64G RAM.

5.2 Color Transform

We first show the color transform performance against two OT-based methods: Monge-Kantorovitch linear (MKL) mapping [52] and regularized discrete optimal transfer (ROT) [16] respectively. Due to the high computational cost of large-scale OT problems, the transformation maps in both cases are firstly derived on sub-samplings, and post-processing such as interpolation and filtering method is then applied for image reconstruction. We also compare the

(a) Content (b) Ours (c) Reference (d) Content (e) Ours (f) Reference

Figure 6: Artistic style transfer effects. Given input images (a) and (d), and reference images (c) and (f), the proposed model gives rise to the stylized results (b) and (e) with consistent textures and structures (Zoom in for better view).

Table 1: Quantitative comparison of style transfer methods. The best two results are highlighted in **bold** and underlined, respectively.

Methods	Metrics	AdaIN [30]	WCT [40]	Photo WCT [41]	WCT^2 [66]	StyTr^2 [9]	QuantArt [29]	Ours
Pixel level	SSIM (edge) ↑	0.5785	0.6174	0.6773	0.6160	0.4944	0.7590	**0.7934**
	IIT loss [28] ↓	61.35	44.82	35.34	34.51	50.02	41.38	**32.87**
Feature level	Gram loss [30] ↓	1.64	1.87	1.45	**1.28**	1.51	2.17	1.37
	LPIPS loss [67] ↓	0.5260	0.5645	0.2284	0.2885	0.4532	0.4257	**0.2146**
	FID metric [24] ↓	278.45	272.57	153.81	152.49	246.75	**146.50**	152.20

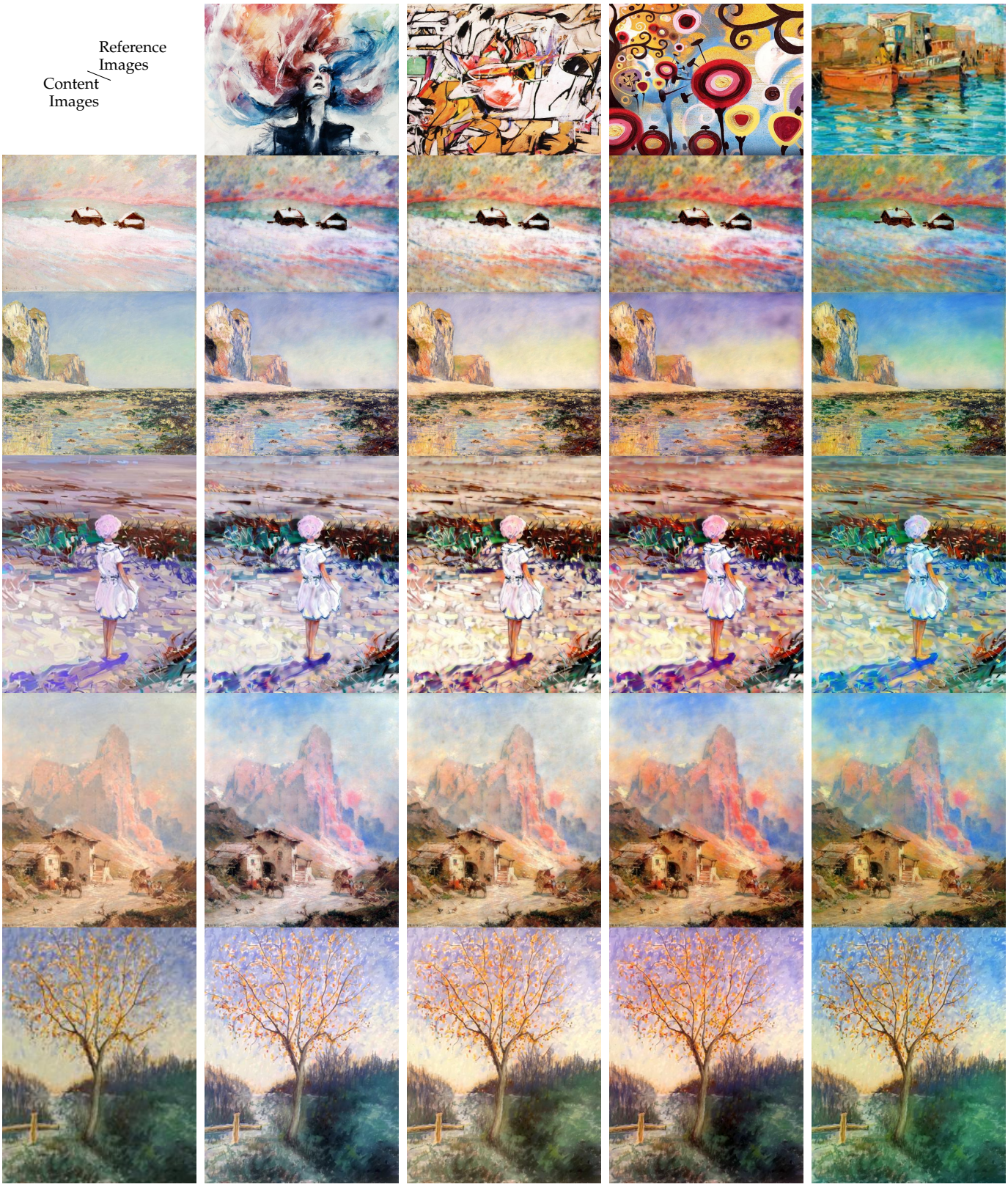
results with neural color transfer methods: PhotoNAS [2] and neural preset [35], respectively. As stated therein, both of them employ specially-designed network architectures and are trained on huge amount datasets to avoid artifacts with cutting-edge color mapping effects.

As shown in Fig.6, the OT-based MKL mapping [52] and ROT model [16] faithfully convert the content color in both cases according to the guidance of the reference images. The former reveals color degradation with over-smoothing details (Fig. 6b), while the latter exhibits block artifacts (Fig. 6c) due to the sub-sampling strategy, although they can be partially rectified by some filtering methods. The neural transfer methods [2], [35] produce fine details but they may suffer from inappropriate color mapping, leading to over-saturated or under-saturated effects. In contrast, the proposed model (Fig. 6f) greatly alleviates the drawbacks for better results due to the strategy of simultaneous representation and transformation, while the improvement is achieved at the expense of high computational cost of dictionary training and image reconstruction compared with the sub-sampling strategy.

5.3 Photo-realistic Style Transfer

We further show that the proposed model is applicable to artistic style transfer, especially the photo-realistic effects. As shown in Fig. 6, we present the transferred results with the artworks created by famous artists. It is clear that all cases have very complex content information composed of sophisticated painting strokes and multiple colors, and textures for aesthetic effects. To receive visual-pleasant results, we randomly select $100K$ patches from content and reference images, and use 1024 dictionary elements for training. The optimal transport is solved by the entropy-regularized method [6] with regularization parameter $\gamma = 0.05$ and 200 iterations for each optimal transport step. The new model is possible to produce visual-pleasant transferred results with consistently aesthetic styles and well-preserved details. In Fig. 7, we additionally show the high-quality model performance in the scenarios of different reference styles, which demonstrate the benefits of sparse representation and optimal transport.

In Fig. 8, we compare the transferred results with deep learning-based methods. The AdaIN model [30] and WCT method [30] are two well-known arbitrary image style transfer methods. As shown in Fig. 8b and Fig. 8c, both of them are capable of transforming the global features such as image colors and main structures for impressive results, however, they may have limitations in suppressing non-consistent local details. Recently, the transformer-based architecture shows great attention in many vision-based tasks. We here also include the StyTr^2 model [9] for comparison, however, it also suffers slight non-consistent local details in Fig. 8d despite the powerful transformer architecture. In addition, we also compare the transferred results with



(a) Input (b) Style 1 (c) Style 2 (d) Style 3 (e) Style 4

Figure 7: Visual results of artistic style transfer of the proposed method. The content images in the first column are transferred according to the given styles in the first row. (Zoom in for better view).

the photo-realistic methods: PhotoWCT model [41], WCT^2 method [66] and QuantArt [29], in which the non-consistent local details can be significantly suppressed as verified in Fig. 8e $\sim$ Fig. 8g. Similarly, the proposed method also

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 8: Visual comparison of artistic style transfer with deep learning methods: (a) Input content and style images, (b) AdaIN [30], (c) WCT [40], (d) PhotoWCT [41], (e) WCT^2 [66] (f) StyTr^2 [9], (g) QuantArt [29] and (h) our results. (Zoom in for better view).

gives arise to photo-realistic results with more consistent local textures and details in 8h. In summary, the proposed method is applicable for both natural and artistic images, and the results further demonstrate its ability in retrieving consistent details.

Additionally, we present the quantitative performance against recent deep learning-based methods. Notice that an objective assessment is often difficult due to the lack of ground truth in aesthetic significance. As suggested in recent work [2], [9], [29], we adopt the structural similarity (SSIM) of edge maps between content and transferred images to indicate detail preservation ability. We also take the structure fidelity into account based on the intrinsic image transfer (IIT) algorithm [28] in consideration of the robust structure-preserving property in varying illumination (color and brightness) conditions. Meanwhile, we introduce three deep learning-based evaluation metrics: LPIPS loss [67], Gram loss (VGG style features) [9], [29], [30] and FID metric [24], which measure the perceptual similarity between the content and generated images from the aspects of image content, style and visual fidelity, respectively.

The evaluation is conducted on a small subset with 21 paired content and reference images sampled from of the WikiArt dataset [56]. For fairness, all images are rescaled into 512×512 resolution and the statistic results are listed in **??**. As we can see, the results are consistent with the visual

effects in Fig. 8. The AdaIN [30] method is efficient but has low performance of structural fidelity on both SSIM-edge and IIT indexes. The WCT [40] receives obvious improvements in structural similarity, but the perceptual similarity is decreased due to the over-smoothing local structures. The increased trend of structural similarity is also observed in both PhotoWCT [41] and WCT^2 [66] methods. Moreover, they have much better perception-based LPIPS, Gram loss, and FID metric, which can be also demonstrated from their photo-realistic transferred effects. The $StyTr^2$ [9] shows very similar effects as AdaIN method and the QuantArt [29] produce high-quality consistent structures, but has limited perceptual similarity in Gram loss and FID metric, which may be caused by the limited color transform in some cases. In contrast, the proposed method gives a fine balance between structural fidelity and perceptual similarity in content, style, and visual fidelity, which is observed in the visual effects in Fig. 8. The benefits mainly underpin the fact that sparse representation provides a simple but effective tool that is especially suitable for low-level or middle-level feature extraction, especially in the case of pursuing photo-realistic image transfer effects.

6 Conclusions

In this paper, we propose a novel optimal transport over sparse dictionaries to explore the two-fold benefits of sparse representation and optimal transport. We have illustrated that sparse representation provides an easy-grasped tool to encode abstract features such as color, textures, and optimal transport over a small size of learned dictionaries is also computationally efficient in practice. As a result, it helps to simplify the procedure of many image-to-image translation problems significantly. Experimental results show that the proposed model is empirically solvable on several image-to-image translation tasks with plausible transferred results. It is worth noting that the proposed method can be further extended from different aspects, for example, using shared dictionaries, extending to multi-scale cases and learning more confident individual styles with regularization techniques, which are leaving for further work.

Acknowledgments

This work was in part supported by the National Science and Technology Major Project, China (Nos. 2019-I-0001-0001 and 2019-I-0019-0018) and Shandong MSTI Project, China (No. 2019JZZY010122), and Foundation for Innovative Young Talents in Higher Education of Guangdong, China (No. 2021KQNCX213). This work was also partially supported by the FWO Odysseus Project, Leverhulme Grant RPG-2017-151 and EPSRC grant EP/R003025/1. We thank Prof. Weiermann to offer his paintings for our research. We appreciate the anonymous reviewers for their careful reading of our manuscript and their insightful comments and suggestions, and also the related online resources, including images, codes, software, and so on.

References

[1] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006.

[2] Jie An, Haoyi Xiong, Jun Huan, and Jiebo Luo. Ultrafast photorealistic style transfer via neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10443–10450, 2020.

[3] Richard H. Bartels and George W Stewart. Solution of the matrix equation ax+ xb= c [f4]. *Communications of the ACM*, 15(9):820–826, 1972.

[4] Adrien Bousseau, Matt Kaplan, Joëlle Thollot, and François X Sillion. Interactive watercolor rendering with temporal coherence and abstraction. In *Proceedings of the 4th international symposium on Non-photorealistic animation and rendering*, pages 141–149, 2006.

[5] Sheng Chen, Stephen A Billings, and Wan Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, 50(5):1873–1896, 1989.

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

[7] George Dantzig. Linear programming and extensions. In *Linear programming and extensions*. Princeton university press, 2016.

[8] Julie Delon. Midway image equalization. *Journal of Mathematical Imaging and Vision*, 21(2):119–134, 2004.

[9] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11326–11336, 2022.

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[11] Pavel Dvurechensky, Alexander Gasnikov, and Alexey Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm. *arXiv preprint arXiv:1802.04367*, 2018.

[12] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346, 2001.

[13] Michael Elad and Peyman Milanfar. Style transfer via texture synthesis. *IEEE Transactions on Image Processing*, 26(5):2338–2351, 2017.

[14] Kjersti Engan, Sven Ole Aase, and J Hakon Husoy. Method of optimal directions for frame design. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, volume 5, pages 2443–2446. IEEE, 1999.

[15] Hasan Sheikh Faridul, Tania Pouli, Christel Chamaret, Jürgen Stauder, Alain Trémeau, Erik Reinhard, et al. A survey of color mapping and its applications. *Eurographics (State of the Art Reports)*, 3(2):1, 2014.

[16] Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014.

[17] Oriel Frigo, Neus Sabater, Vincent Demoulin, and Pierre Hellier. Optimal transportation for example-guided color transfer. In *Asian Conference on Computer Vision*, pages 655–670. Springer, 2014.

[18] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015.

[19] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[20] Bruce Gooch, Greg Coombe, and Peter Shirley. Artistic vision: painterly rendering using computer vision techniques. In *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, pages 83–ff, 2002.

[21] Bruce Gooch and Amy Gooch. *Non-photorealistic rendering*. CRC Press, 2001.

[22] Shuyang Gu, Congliang Chen, Jing Liao, and Lu Yuan. Arbitrary style transfer with deep feature reshuffle. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8222–8231, 2018.

[23] Aaron Hertzmann. Painterly rendering with curved brush strokes of multiple sizes. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 453–460, 1998.

[24] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[25] Man M Ho and Jinjia Zhou. Deep preset: Blending and retouching photos with color style transfer. In *Proceedings of the IEEE/CVF*

Winter Conference on Applications of Computer Vision, pages 2113–2121, 2021.
[26] Kibeom Hong, Seogkyu Jeon, Huan Yang, Jianlong Fu, and Hyeran Byun. Domain-aware universal style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14609–14617, 2021.
[27] Hristina Hristova, Olivier Le Meur, Rémi Cozot, and Kadi Bouatouch. Style-aware robust color transfer. In *Proceedings of the workshop on Computational Aesthetics*, pages 67–77, 2015.
[28] Junqing Huang, Michael Ruzhansky, Qianying Zhang, and Haihui Wang. Intrinsic image transfer for illumination manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7444–7456, 2022.
[29] Siyu Huang, Jie An, Donglai Wei, Jiebo Luo, and Hanspeter Pfister. Quantart: Quantizing image style transfer towards high visual fidelity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5947–5956, 2023.
[30] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.
[31] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
[32] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 2019.
[33] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.
[34] Henry Kang, Seungyong Lee, and Charles K Chui. Flow-based image abstraction. *IEEE transactions on visualization and computer graphics*, 15(1):62–76, 2008.
[35] Zhanghan Ke, Yuhao Liu, Lei Zhu, Nanxuan Zhao, and Rynson WH Lau. Neural preset for color style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14173–14182, 2023.
[36] Sung Ye Kim, Ross Maciejewski, Tobias Isenberg, William M Andrews, Wei Chen, Mario Costa Sousa, and David S Ebert. Stippling by example. In *Proceedings of the 7th International Symposium on Non-Photorealistic Animation and Rendering*, pages 41–50, 2009.
[37] Nicholas Kolkin, Jason Salavon, and Gregory Shakhnarovich. Style transfer by relaxed optimal transport and self-similarity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10051–10060, 2019.
[38] Dmytro Kotovenko, Artsiom Sanakoyeu, Sabine Lang, and Bjorn Ommer. Content and style disentanglement for artistic style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4422–4431, 2019.
[39] Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg. State of the" art": A taxonomy of artistic stylization techniques for images and video. *IEEE transactions on visualization and computer graphics*, 19(5):866–885, 2012.
[40] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. *Advances in neural information processing systems*, 30, 2017.
[41] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 453–468, 2018.
[42] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30:700–708, 2017.
[43] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. Paint transformer: Feed forward neural painting with stroke prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6598–6607, 2021.
[44] Ming Lu, Hao Zhao, Anbang Yao, Yurong Chen, Feng Xu, and Li Zhang. A closed-form solution to universal style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5952–5961, 2019.
[45] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep photo style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4990–4998, 2017.
[46] Quentin Mérigot and Edouard Oudet. Discrete optimal transport: complexity, geometry and applications. *Discrete & Computational Geometry*, 55(2):263–283, 2016.
[47] Ján Morovic and Pei-Li Sun. Accurate 3d image colour histogram transformation. *Pattern Recognition Letters*, 24(11):1725–1735, 2003.
[48] Youssef Mroueh. Wasserstein style transfer. In *International Conference on Artificial Intelligence and Statistics*, pages 842–852. PMLR, 2020.
[49] Adam M Oberman and Yuanlong Ruan. An efficient linear programming method for optimal transportation. *arXiv preprint arXiv:1509.03668*, 2015.
[50] Michaël Perrot, Nicolas Courty, Rémi Flamary, and Amaury Habrard. Mapping estimation for discrete optimal transport. *Advances in Neural Information Processing Systems*, 29, 2016.
[51] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
[52] François Pitié and Anil Kokaram. The linear monge-kantorovitch linear colour mapping for example-based colour transfer. 2007.
[53] Julien Rabin, Sira Ferradans, and Nicolas Papadakis. Adaptive color transfer with relaxed optimal transport. In *2014 IEEE international conference on image processing (ICIP)*, pages 4852–4856. IEEE, 2014.
[54] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001.
[55] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.
[56] Wendy Kan small yellow duck. Painter by numbers. https://kaggle.com/competitions/painter-by-numbers, 2016. Accessed: 2023-02-10.
[57] Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang. Local color transfer via probabilistic segmentation by expectation-maximization. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 747–754. IEEE, 2005.
[58] Zhengyan Tong, Xiaohang Wang, Shengchao Yuan, Xuanhong Chen, Junjie Wang, and Xiangzhong Fang. Im2oil: Stroke-based oil painting rendering with linearly controllable fineness via adaptive sampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1035–1046, 2022.
[59] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1945–1959, 2005.
[60] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
[61] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2216–2223. IEEE, 2012.
[62] Holger Winnemöller, Jan Eric Kyprianidis, and Sven C Olsen. Xdog: An extended difference-of-gaussians compendium including advanced image stylization. *Computers & Graphics*, 36(6):740–753, 2012.
[63] Xide Xia, Tianfan Xue, Wei-sheng Lai, Zheng Sun, Abby Chang, Brian Kulis, and Jiawen Chen. Real-time localized photorealistic video style transfer. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1089–1098, 2021.
[64] Xuezhong Xiao and Lizhuang Ma. Color transfer in correlated color space. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, pages 305–309, 2006.
[65] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.
[66] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9036–9045, 2019.
[67] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.
[68] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
[69] Zhengxia Zou, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. Stylized neural painting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15689–15698, 2021.