

Progressive Color Transfer with Dense Semantic Correspondences

MINGMING HE*, Hong Kong UST

JING LIAO†, City University of Hong Kong, Microsoft Research

DONGDONG CHEN, University of Science and Technology of China

LU YUAN, Microsoft AI Perception and Mixed Reality

PEDRO V. SANDER, Hong Kong UST

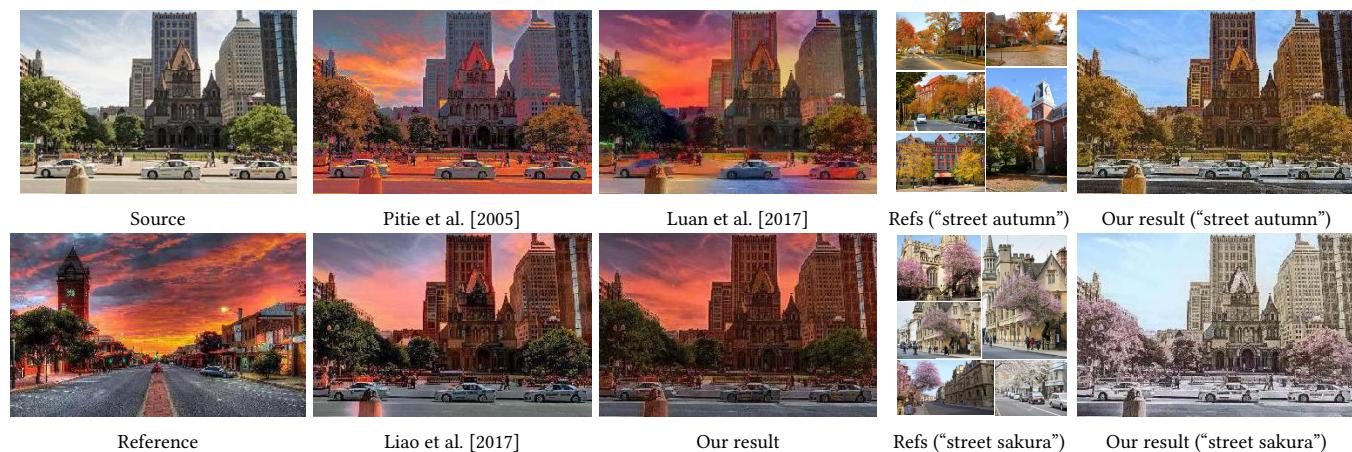


Fig. 1. Our method leverages semantically-meaningful dense correspondences between images, thus achieving a more accurate object-to-object color transfer than other methods (left). Moreover, our method can be successfully extended to multiple references (right). Input images: Bill Damon (Source) and PicsWalls.com (Reference).

We propose a new algorithm for color transfer between images that have perceptually similar semantic structures. We aim to achieve a more accurate color transfer that leverages semantically-meaningful dense correspondence between images. To accomplish this, our algorithm uses neural representations for matching. Additionally, the color transfer should be spatially variant and globally coherent. Therefore, our algorithm optimizes a local linear model for color transfer satisfying both local and global constraints. Our proposed approach jointly optimizes matching and color transfer, adopting a coarse-to-fine strategy. The proposed method can be successfully

extended from “one-to-one” to “one-to-many” color transfer. The latter further addresses the problem of mismatching elements of the input image. We validate our proposed method by testing it on a large variety of image content.

CCS Concepts: • Computing methodologies → Image manipulation; Computational photography;

Additional Key Words and Phrases: color, transfer, deep matching

1 INTRODUCTION

Color transfer is a long-standing problem that seeks to transfer the color style of a reference image onto a source image. By using different references, one can alter the color style without changing the original image content

*This work was done when Mingming He and Dongdong Chen were interns at MSR Asia.

†indicates corresponding author.

Progressive Color Transfer with Dense Semantic Correspondences

MINGMING HE*, Hong Kong UST

JING LIAO†, City University of Hong Kong, Microsoft Research

DONGDONG CHEN, University of Science and Technology of China

LU YUAN, Microsoft AI Perception and Mixed Reality

PEDRO V. SANDER, Hong Kong UST

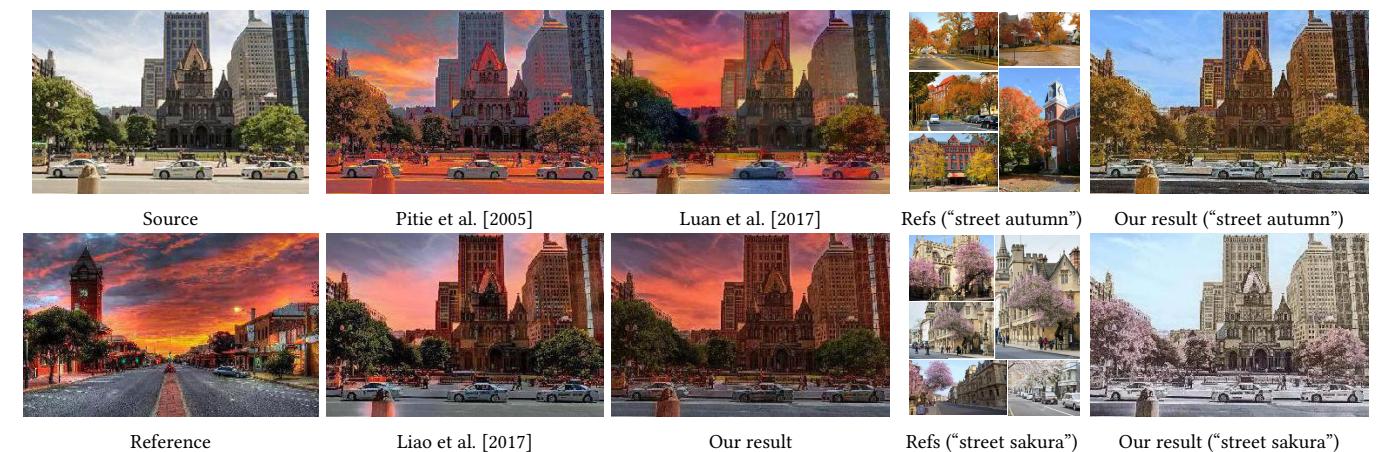


Fig. 1. 我们的方法利用图像之间的语义上连贯的密集对应关系，从而实现了比其他方法更准确的物体到物体的颜色转移（左侧）。此外，我们的方法还可以成功扩展到多个参考（右侧）。输入图像：Bill Damon（源）和PicsWalls.com（参考）。

*警告：该PDF由GPT-Academic开源项目调用大语言模型+Latex翻译插件一键生成，版权归原文作者所有。翻译内容可靠性无保障，请仔细鉴别并以原文为准。项目Github地址 https://github.com/binary-husky/gpt_academic/。当前大语言模型: siliconflow-01-ai/Yi-1.5-9B-Chat-16K，当前语言模型温度设定: 0。为了防止大语言模型的意外谬误产生扩散影响，禁止移除或修改此警告。

我们提出了一种新的算法，用于将具有感知上相似的语义结构的图像之间的颜色转移。我们的目标是实现一种更精确的颜色转移，它利用了图像之间的语义上有意义的密集对应关系。为了实现这一点，我们的算法使用神经表示来进行匹配。此外，颜色转移应该是空间变量的并且具有全局一致性。因此，我们的算法优化了一个局部线性模型，以满足局部和全局

*This work was done when Mingming He and Dongdong Chen were interns at MSR Asia.

†indicates corresponding author.

约束的颜色转移。我们提出的算法采用了一种粗到细的策略，同时优化匹配和颜色转移。提出的该方法可以从“一对一”成功扩展到“一对多”颜色转移。后者进一步解决了输入图像中元素不匹配的问题。我们通过在各种图像内容上测试我们的提议方法来验证我们的提议方法。

CCS Concepts: • Computing methodologies → Image manipulation; Computational photography;

Additional Key Words and Phrases: color, transfer, deep matching

1 INTRODUCTION

颜色迁移是一个长期存在的问题，它寻求将参考图像的颜色风格转移到源图像上。通过使用不同的参考，可以在不改变原始图像内容的情况下改变颜色风格，以模拟

in order to emulate different illumination, weather conditions, scene materials, or even artistic color effects.

To achieve more accurate transfer, semantically meaningful correspondences are necessary to be established between input images. Due to large variations in appearance, matching methods based on hand-crafted features (e.g., intensity, Gabor wavelet, SIFT, or SSIM) may fail. Therefore, some methods require additional segmentation [Dale et al. 2009; Luan et al. 2017], or user specifications [An and Pellacini 2008] but these regional correspondences are not quite effective in their pixel-level accuracy. Recently, Liao et al. [2017] leverage multi-level features of the deep neural network for dense correspondence and then conduct local color transfer during post-processing. This method is robust in finding high-level semantic correspondences between different objects, but may misalign some fine-scale image structures because low-level neural representations are still influenced by the color discrepancies. Thus, in their color transfer results, artifacts such as ghosting and a halo may appear, e.g., the halo around the pillar in Fig. 1.

To refine correspondences and reduce these color transfer artifacts, we propose a novel progressive framework, which allows for the dense semantic correspondences estimation in the deep feature domain and local color transfer in the image domain to mutually contribute each other. This is implemented progressively by leveraging multi-level deep features extracted from a pre-trained VGG19 [Simonyan and Zisserman 2014]. At each level, the nearest-neighbor field (NNF [Barnes et al. 2010]) built on deep features is used to guide the local color transfer in the image domain. The local color transfer considers a linear transform at every pixel, enforcing both local smoothness and non-local constraints to avoid inconsistencies. Then the transferred result, whose appearance becomes much closer to the reference, helps the NNF to be refined at the next level. From coarse to fine, dense correspondences between features ranging from high-level semantics to

low-level details can be built as the differences between two input images are gradually reduced. Therefore, for the image pairs which share similar semantic content but demonstrate significant differences in appearance, our approach is able to achieve natural and consistent color transfer effects, which is challenging for the existing solutions.

In addition to single reference color transfer, our approach can be easily extended to handle multiple references in a similar manner, which provide even richer reference content to help achieve stronger semantic matching. Our algorithm generalizes *one-to-one* NNF search to *one-to-many*, and enforces piecewise smoothness by placing it into a Markov Random Field (MRF) optimization framework.

In brief, our major technical contributions are:

- (1) We present a novel progressive color transfer framework, which jointly optimizes dense semantic correspondences in the deep feature domain and the local color transfer in the image domain.
- (2) We present a new local color transfer model, which is based on a pixel-granular linear function, avoiding local structural distortions and preserving global coherence by enforcing both local and global constraints.
- (3) We extend our *one-to-one* color transfer to *one-to-many*, which further improves result quality and robustness through effectively avoiding content mismatching.

We show how our local color transfer technique can be effectively applied to a variety of real scenarios, such as makeup transfer and time-lapse from images. Our technique can also be used to transfer colors to a gray image, known as the colorization problem.

不同的照明、天气条件、场景材料,甚至是艺术色彩效果。

为了实现更准确的迁移,需要在输入图像之间建立语义上有意义的对应关系。由于外观差异很大,基于手工特征的方法(例如,强度、小波波纹、SIFT或SSIM)可能会失败。因此,一些方法需要额外的分割[Dale et al. 2009; Luan et al. 2017],或用户指定[An and Pellacini 2008],但这些区域对应关系在像素级精度上并不有效。最近,Liao et al. [2017]利用深度神经网络的多级特征进行密集对应关系,然后在后处理中进行局部颜色迁移。这种方法在找到不同对象之间的高级语义对应关系上是健壮的,但可能会因为低级神经表示仍然受到颜色差异的影响而错误地对齐一些精细的图像结构。因此,在它们的颜色迁移结果中,可能会出现诸如鬼影和光晕等artifacts,例如图Fig. 1中的柱子周围的光晕。

为了细化对应关系并减少这些颜色迁移的artifacts,我们提出了一种新颖的逐步框架,该框架允许在深度特征域中进行密集语义对应关系估计,并在图像域中进行局部颜色迁移,相互贡献。这通过利用从预训练的VGG19 [Simonyan and Zisserman 2014]中提取的多级深度特征来实现逐步的。在每一级中,基于深度特征构建的最近邻场(NNF [Barnes et al. 2010])用于指导图像域中的局部颜色迁移。局部颜色迁移考虑每个像素的线性变换,强制执行局部平滑性和非局部约束,以避免不一致性。然后,转移的结果,其外观变得更加接近参考,帮助NNF在下一级进行改进。从粗到细,随着两个输入图像之间的差异逐渐减少,可以建立从高级语义到低级细节的密集对应关系。因此,对于共享相似语义内容但外观差异显著的图像对,我们的方法能够实现自然和一致的颜色迁移效果,这对于现有的解决方案来说是一个挑战。

除了单参考颜色迁移,我们的方法还可以轻松扩展到以类似方式处理多个参考,这提供了更丰富的参考内容,有助于实现更强的语义匹配。我们的算法将一对多的NNF搜索推广到一对多,并通过将其置于马尔

可夫随机场(MRF)优化框架中来强制执行局部平滑性。

简而言之,我们的主要技术贡献是:

- (1) 我们提出了一种新颖的渐进式颜色转移框架,该框架在深度特征域中联合优化了密集的语义对应关系,并在图像域中优化了局部颜色转移。
- (2) 我们提出了一种新的局部颜色转移模型,该模型基于像素级的线性函数,通过施加局部和全局约束,避免了局部结构失真并保持了全局一致性。
- (3) 我们扩展了我们的一对一颜色转移到了一对多,这进一步通过有效地避免内容不匹配提高了结果的质量和鲁棒性。

我们展示了我们的局部颜色转移技术如何在各种真实场景中有效应用,如化妆转移和时间流逝从图像。我们的技术还可以用来将颜色转移到灰度图像上,这被称为颜色化问题。

2 RELATED WORK

颜色转移可以应用于黑白或彩色源图像。将颜色转移到黑白图像,即颜色化,是一个研究得比较透彻的问题。早期解决这个问题的方法依赖于用户的涂鸦,并通过优化在相似区域之间扩展这些涂鸦[Levin et al. 2004]。最近,学习型算法被用于自动图像颜色化[Iizuka et al. 2016; Zhang et al. 2016],但这些方法需要在大型数据集上学习图像统计信息。给定一个参考图像而不是用户的输入,一些自动方法通过相似统计信息的像素之间的颜色转移来转移色度[Arbelot et al. 2017; Welsh et al. 2002]。He et al. [2018]将参考图像集成到基于学习的方法中,以实现基于示例的自动颜色化,但它也仅限于只转移色度。我们的方法适用于使用同一类别的参考图像的颜色化,但我们的重点是彩色图像对之间的亮度与色度转移。

2.1 Single-reference Color Transfer

传统方法. 全球色彩转换算法应用空间不变的色彩变换对图像进行处理,基于全局信息匹配,例如全

2 RELATED WORK

Color transfer can be applied to either grayscale or color source images. Transferring colors to a grayscale image, known as colorization, is a well-studied problem. Early approaches to address this issue rely on user scribbles and extend them via optimization across similar regions [Levin et al. 2004]. Recently, learning-based algorithms have been used for automatic image colorization [Iizuka et al. 2016; Zhang et al. 2016], but these methods have to learn image statistics from large extensive datasets. Given one reference image instead of user input, some automatic methods transfer the chrominance between pixels containing similar statistics [Arbelot et al. 2017; Welsh et al. 2002]. He et al. [2018] integrate reference images into a learning-based method to achieve automatic exemplar-based colorization but it is also limited to only transfer the chrominance. Our method is applicable to colorization using reference images of the same class, but our focus is on both luminance and chrominance transfer between a color image pair.

2.1 Single-reference Color Transfer

Traditional methods. Global color transfer algorithms apply a spatially-invariant color transformation to an image based on global information matching, such as global color moves (*e.g.*, *sepia*) and tone curves (*e.g.*, high or low contrast). The seminal work by Reinhard et al. [2001] matches the mean and standard deviation between the input and the reference in the $l\alpha\beta$ color space. Pitie et al. [2005] transfer the full 3D color histogram using a series of 1D histograms. Freedman and Kisilev [2010] compute the transfer for each histogram bin with the mean and variance of pixel values in the bin, which strikes a compromise between mean-variance based methods and histogram based methods. These methods only consider global color statistics, ignoring the spatial layout.

Local color transfer algorithms based on spatial color mappings are more expressive and can handle a broad class of applications [Bae et al. 2006; Laffont et al. 2014; Shih et al. 2014, 2013; Sunkavalli et al. 2010]. Having local correspondences is necessary for correct local transfers. Some methods identify regional correspondence and transfer color distributions between corresponding regions. They either require the user input to guide sparse correspondence [An and Pellacini 2010; Welsh et al. 2002] or rely on automatic image segmentation or clustering algorithms to estimate regional correspondence [Arbelot et al. 2017; Hristova et al. 2015; Tai et al. 2005; Yoo et al. 2013]. Such matches are not yet precise enough, causing some pixels to be transferred to inaccurate colors.

To exploit pixel-level dense correspondences for more spatially complicated color transfer, some analogy-based methods [Hertzmann et al. 2001; Laffont et al. 2014; Shih et al. 2013] rely on an additional image which has similar colors to the source and similar structure to the reference. With this bridging image, building dense correspondences between two inputs gets easier and a locally linear color model is then estimated and applied. However, such a bridging image is not easy to obtain in practice.

Without the bridging image, it is difficult to directly build dense correspondences between two inputs which are vastly different in color appearance. Shen et al. [2016] propagate sparse correspondence with model fitting and optimization to build dense matching only inside foremost regions. HaCohen et al. [2011] introduce a coarse-to-fine scheme in which NNF computation is interleaved with fitting a global parametric color model to gradually narrow down the color gap for matching. We are inspired by this progressive idea, but our method is essentially different from theirs in both dense correspondence estimation and color transfer model fitting. One one side, dependence on the input image pairs is high in their method, for example,

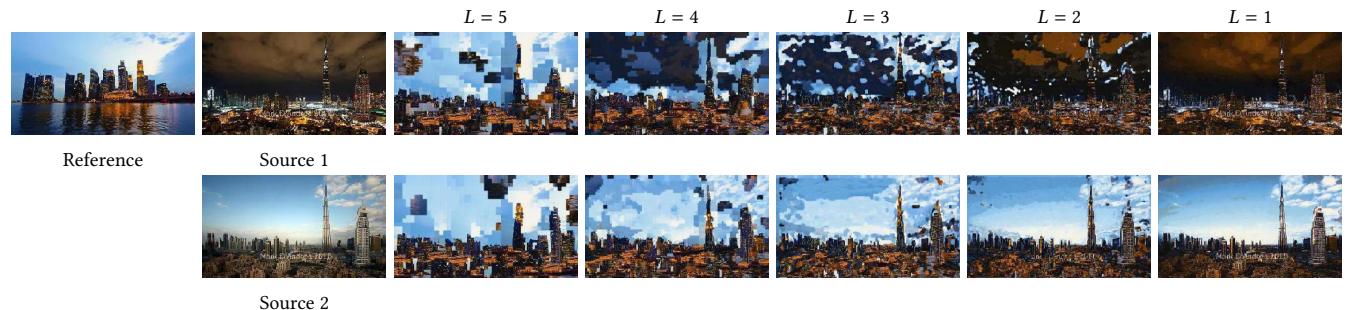


Fig. 2. 使用不同层特征的NNF搜索匹配结果 (VGG19中的 $reluL_1, L = 5, \dots, 1$)。我们展示了两组具有不同颜色源的迁移对 (所有来自 [Shih et al. 2013])：第一组从蓝色到深色的图像对颜色更为明显 (上排) 和第二组从蓝色到蓝色的图像对颜色相似 (下排)。很明显，基于最粗粒度级别的NNF (*e.g.*, $relu5_1$) 忽略了颜色差异，并获得了相似的结果。然而，低层次的特征 (*e.g.*, $relu5_1$) 对颜色表现敏感，并且无法匹配具有相似语义但不同颜色的对象。输入图像：Shih et al. [2013]。

局色彩移动 (例如, 褐色) 和色调曲线 (例如, 高或低对比度)。Reinhard等人的开创性工作 [Reinhard et al. 2001] 在 $l\alpha\beta$ 颜色空间中匹配输入和参考之间的均值和标准差。Pitie等人在 [Pitie et al. 2005] 中转移了全3D颜色直方图, 使用一系列1D直方图。Freedman等人在 [Freedman and Kisilev 2010] 中计算了每个直方图bin的转移, 通过bin中的像素值均值和方差, 这种方法在基于均值-方差的方法和基于直方图的方法之间取得了妥协。这些方法仅考虑全局色彩统计信息, 忽视了空间布局。

基于空间色彩映射的局部色彩转换算法更加表达力强, 可以处理广泛的应用 [Bae et al. 2006; Laffont et al. 2014; Shih et al. 2014, 2013; Sunkavalli et al. 2010]。为了正确地进行局部转换, 需要局部对应关系。一些方法识别区域对应关系, 并在对应区域之间转移颜色分布。它们要么需要用户输入来指导稀疏对应关系 [An and Pellacini 2010; Welsh et al. 2002], 要么依赖于自动图像分割或聚类算法来估计区域对应关系 [Arbelot et al. 2017; Hristova et al. 2015; Tai et al. 2005; Yoo et al. 2013]。这样的匹配还不够精确, 导致一些像素被转换到不正确的颜色。

为了利用像素级密集对应关系进行更复杂的空间色彩转换, 一些基于类比的方法 [Hertzmann et al. 2001; Laffont et al. 2014; Shih et al. 2013] 依赖于一个具有类似颜色到源和类似结构到参考的额外图像。有了这个

桥梁图像, 构建两个输入之间的密集对应关系变得更容易, 然后估计并应用局部线性色彩模型。然而, 在实践中, 这样的桥梁图像不容易获得。

没有桥梁图像, 直接在两个输入之间建立密集对应关系是非常困难的, 这些输入在颜色外观上存在巨大差异。Shen et al. [2016] 通过模型拟合和优化传播稀疏对应关系, 只在前景区域内部建立密集匹配。HaCohen et al. [2011] 引入了一种粗到细的方案, 其中NNF计算与全局参数化色彩模型拟合交替进行, 逐步缩小匹配的颜色差距。我们受到这种渐进式想法的启发, 但我们的方法在密集对应关系估计和色彩转移模型拟合方面与他们的方法本质不同。一方面, 他们的方法高度依赖于输入图像对, 例如同一场景的两张照片, 因为使用了低级特征 (例如, 图像块)。由于深度特征的集成, 我们的方法支持场景和外观差异巨大的输入对。另一方面, 他们的色彩映射, 虽然局部细化, 但是单一的全局转换模型, 因此无法适应复杂的空间色彩变化。相比之下, 我们的模型是一个像素级的局部色彩转换模型。

基于深度网络的方法. 传统的方法通过匹配低级特征无法反映高级别的语义关系。最近, 深度神经网络提供了良好的表示, 能够建立视觉上不同的图像对之间的语义上有意义的对应关系, 这些方法可以用于风格迁移 [Chen et al. 2017b, 2018b; Gatys et al. 2015]。“深度照片风格迁移” [Luan et al. 2017] 的工作将全局神经

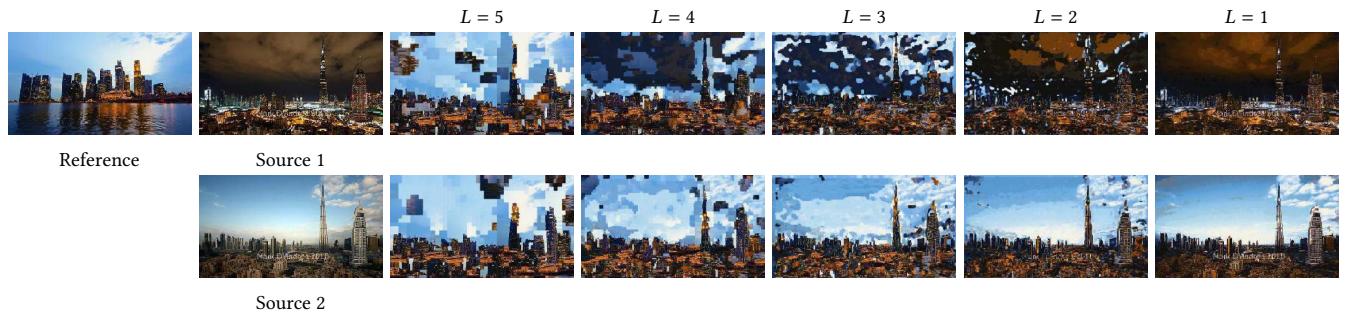


Fig. 2. Matching results from the NNF search using features from different layers individually ($reluL_1, L = 5, \dots, 1$ in VGG19). We show two transfer pairs with sources of different colors (all from [Shih et al. 2013]): the first blue-to-dark image pair has more distinct colors (upper row) and the second blue-to-blue image pair has similar colors (lower row). It is clear that NNFs based at the coarsest level (e.g., $relu5_1$) ignore color difference and achieve similar results. However, low-level features (e.g., $relu5_1$) are sensitive to color appearance and fail to match objects with semantic similarity but different colors. Input images: Shih et al. [2013].

two photos of the same scene, because of the use of low-level features (e.g., image patch). Thanks to the integration of deep features, our method supports input pairs with vast differences in scene and appearance. One the other side, their color mapping, although locally refined, is a single global transformation model and thus cannot fit complicated spatial color variation. In contrast, ours is a pixel-level local color transfer model.

Deep network-based methods. Traditional color transfer methods by matching low-level features are unable to reflect higher-level semantic relationships. Recently, deep neural networks have provided a good representation to establish semantically-meaningful correspondence between visually different image pairs, which can be used in style transfer [Chen et al. 2017b, 2018b; Gatys et al. 2015]. The work of “deep photo style transfer” [Luan et al. 2017] extends the global neural style transfer [Gatys et al. 2015] to photo-realistic transfer by enforcing local color affine constraints in the loss function. Their regional correspondence relies on semantic segmentation [Chen et al. 2018a] of the image pairs. Luan et al. [2017] attempt to improve the photorealism of the stylized images via a post-processing step based on the Matting Laplacian of Levin et al. [2008]. Mechrez et al. [2017] propose an

approach based on the Screened Poisson Equation (SPE) to accelerate the post-processing step.

To estimate the semantically dense correspondences between two images, Liao et al. [2017] present “deep image analogy” to take advantage of multi-scale deep features. We use the same feature representation but our work has three key differences applicable for color transfer. First, our approach jointly optimizes the dense semantic correspondences and the local color transfer, while Liao et al. [2017] achieve color transfer via a two-stage approach, starting with building dense correspondence and then post-processing to change the color. Second, to transfer color, our approach optimizes the linear transform model satisfying local and global coherence constraints while theirs directly applies the NNF to replace the low-frequency source color with the corresponding reference color. Third, our approach can be easily extended to *one-to-many* color transfer, which effectively avoids content-mismatching in *one-to-one* transfer [Liao et al. 2017].

2.2 Multi-reference Color Transfer

Choosing a proper reference for color transfer is crucial when using a single reference. To ease the burden of reference selection, some methods adopt multiple references,

风格迁移 [Gatys et al. 2015] 扩展到照片现实风格的迁移，通过损失函数中的局部颜色线性约束来实现。他们的区域对应关系依赖于图像对的语义分割 [Chen et al. 2018a]。Luan et al. [2017] 试图通过基于 Levin et al. [2008] 的 Matting 拉普拉斯的后期处理步骤来改善风格化图像的现实感。Mechrez et al. [2017] 提出了一种基于筛选泊松方程 (SPE) 的方法来加速后期处理步骤。

为了估计两幅图像之间的语义密集对应关系, Liao et al. [2017] 提出了“深度图像类比”方法，利用多尺度深度特征。我们使用相同的特征表示，但我们的工作在颜色迁移方面有三个关键差异。首先，我们的方法联合优化了密集语义对应关系和局部颜色迁移，而 Liao et al. [2017] 通过两阶段方法实现颜色迁移，首先构建密集对应关系，然后进行后期处理以改变颜色。其次，为了迁移颜色，我们的方法优化了满足局部和全局一致性的线性变换模型，而他们的方法直接应用 NNF 来替换低频源颜色为相应的参考颜色。第三，我们的方法可以轻松扩展到一对多到一对的的颜色迁移，有效地避免了一对一迁移中的内容不匹配 [Liao et al. 2017]。

2.2 Multi-reference Color Transfer

选择合适的颜色转移参考对于使用单个参考至关重要。为了减轻选择参考的负担，一些方法采用多个参考，可以通过提供一个文本查询来搜索和聚类互联网上的相似颜色风格 [Bonneel et al. 2016; Liu et al. 2014]，或者根据语义和风格进行排名和选择 [Lee et al. 2016]。这些方法在获得多个参考后最终应用全局颜色转移。为了实现更精确的局部转移，Khan et al. [2017] 允许用户手动为输入和多个参考之间提供一些对应指导，然后使用局部线性嵌入 (LLE) 方法 [Roweis and Saul 2000] 来传播指导。

深度网络最近被引入到多图像颜色转移的任务中。Yan et al. [2016] 通过将捆绑特征作为深度神经网络输入层的输入，学习一个高度非线性颜色映射函数，用于自动照片调整。Isola et al. [2017] 在一组配对图像

数据集上训练生成网络，用于图像外观转移，包括颜色。Zhu et al. [2017] 放宽了对配对图像的约束。这些方法需要数小时来训练单个颜色风格。网络生成的结果分辨率低，经常因反卷积层 [Odena et al. 2016] 而产生棋盘状 artifacts。相反，我们的方法只使用预训练网络的特征进行匹配。我们可以支持各种颜色风格的优质转移，而无需训练。

3 METHOD

我们的目标是应用基于源图像和参考图像之间建立的密集语义对应关系的高精度局部颜色转移。在我们的场景中，两幅输入图像共享一些语义相关的内容，但可能在外观或结构上存在显著差异。建立它们之间的密集语义对应关系是一个具有挑战性的问题。手工特征无法反映语义信息，因此我们转向从图像分类卷积神经网络 (CNN) VGG19 中提取的深度特征，该网络从低级细节逐渐编码图像到高级语义。我们观察到高层的深度特征（例如VGG19中的 $relu5_1$ 层）通常倾向于对颜色差异保持不变，而低层特征（例如VGG19中的 $relu1_1$ 层）更为敏感，如图 Fig. 2所示。随着图像颜色变得更加相似，它们的特征，尤其是低层特征，变得更加容易匹配。这启发了我们的粗到细方法 (Sec. 3.1)，该方法交替优化深度特征之间的NNFs (Sec. 3.2) 并执行局部颜色转移 (Sec. 3.3)。因此，这两个步骤是相互补充的。

3.1 Overview

给定源图像 S 和参考图像 R ，我们的算法逐步估计它们之间的密集对应关系，并在 S 上应用准确的局部颜色转移，生成输出 S' ，它保留了 S 的结构和 R 的颜色风格。

我们的系统管道如图 Fig. 3 所示。在每一级 L ，有两个步骤：特征域中的 NNF 计算 (Section 3.2) 和图像域中的局部颜色转移 (Section 3.3)。首先，我们使用 VGG19 中的 $reluL_1$ 层将参考 R 与中间源 \tilde{S}^{L+1} 进行匹配，以获得双向 NNF 在特征域中，并使用 NNF 重建颜色指导 G^L 。接下来，我们估计源 S^L 的降采样版本和 G^L 之间的局部颜色转移函数，上采样变换，并将变换应用到 S 上以获

which can be searched and clustered with similar color styles from the Internet by providing a text query [Bonneel et al. 2016; Liu et al. 2014], or ranked and selected according to semantics and style [Lee et al. 2016]. These methods finally apply the global color transfer after getting multiple references. To achieve more precise local transfer, Khan et al. [2017] allow the user to manually give some correspondence guidance between input and multiple references, and then use the locally linear embedding (LLE) method [Roweis and Saul 2000] to propagate the guidance.

Deep networks have recently been introduced to the task of color transfer among multiple images as well. Yan et al. [2016] learn a highly non-linear color mapping function for automatic photo adjustment by taking the bundled features as the input layer of a deep neural network. Isola et al. [2017] train generative networks on a dataset of paired images for image appearance transfer, including colors. Zhu et al. [2017] loosen the constraints to unpaired images. These methods take several hours to train a single color style. The network-generated results are low resolution and often suffer from checkboard artifacts caused by deconvolution layers [Odena et al. 2016]. Instead, our method only uses features from pre-trained networks for matching. We can support the high-quality transfer of various color styles without training.

3 METHOD

Our goal is to apply precise local color transfer based on the established dense semantic correspondences between the source and reference images. In our scenario, the two input images share some semantically-related content, but may vary dramatically in appearance or structure. Building dense semantic correspondences between them is known to be a challenging problem. The hand-crafted features fail to reflect semantic information, so we resort

to the deep features from an image classification Convolutional Neural Network (CNN) VGG19, which encodes the image gradually from low-level details to high-level semantics. We observe high-level deep features (*e.g.*, *relu5_1* layer in VGG19) generally tend to be invariant to color differences, while low-level features (*e.g.*, *relu1_1* layer in VGG19) are more sensitive, as demonstrated in Fig. 2. As their image colors get more similar, their features, especially at lower levels, get easier to match. This inspires our coarse-to-fine approach (Sec. 3.1) to alternately optimize the NNFs between deep features (Sec. 3.2) and perform local color transfer (Sec. 3.3). Thus, the two steps are mutually beneficial.

3.1 Overview

Given a source image S and a reference image R , our algorithm progressively estimates dense correspondence between them and applies accurate local color transfer on S , to generate output S' preserving both the structure from S and the color style from R .

Our system pipeline is shown in Fig. 3. At each level L , there are two steps: NNF computation in the feature domain (Section 3.2) and local color transfer in the image domain (Section 3.3). First, we match the reference R to the intermediate source \tilde{S}^{L+1} using the *reluL_1* layer in VGG19 to get bidirectional NNFs in the feature domain and use the NNFs to reconstruct a color guidance G^L . Next, we estimate the local color transfer function between the downsampled version of source S^L and G^L , upscale the transformation, and apply it to S to get \tilde{S}^L . The two steps alternate and mutually assist one another: the NNFs help obtain a more accurate local color transfer, while the color transferred result \tilde{S}^L serving as the source also helps refine the matching in the next level $L - 1$, since \tilde{S}^L has much more similar colors to the reference than the original source S . Both intermediate results (NNFs and \tilde{S}^L) serve as the bridge between both matching and color

里, \tilde{S}^{L+1} 是在级别 $L + 1$ 的细化级别上 (当 $L < 5$ 时) 得到的颜色转移结果 (与源 S 具有相同的分辨率)。在最高级别 $L = 5$ 时, \tilde{S}^{L+1} 被初始化为 S 。

鉴于在 \tilde{S}^{L+1} 和 R 之间建立正确对应关系的难度 (可能存在显著的外观差异), 我们在深度特征域中执行 NNF 搜索。由于 CNN 几乎总是保持输入图像的空间关系, 在特征域中计算的 NNFs 也可以在图像域中使用。为此, 我们首先将 \tilde{S}^{L+1} 和 R 输入到 VGG19 网络 [Simonyan and Zisserman 2014], 该网络在 ImageNet 数据库 [Russakovsky et al. 2015] 上用于物体识别的预训练。然后, 我们在 *reluL_1* 层提取它们的特征图, 分别标记为 F_S^L , F_R^L , 用于 \tilde{S}^{L+1} , R 。每个特征图是一个具有 $width \times height \times channel$ 维度的 3D 张量, 其空间分辨率是输入的 $1/4^{L-1}$ 。

映射函数 $\phi_{S \rightarrow R}^L$ 从 F_S^L 到 F_R^L 是通过最小化以下能量函数计算得到的:

$$\phi_{S \rightarrow R}^L(p) = \arg \min_q \sum_{x \in N(p), y \in N(q)} \| \bar{F}_S^L(x) - \bar{F}_R^L(y) \|^2 \quad (1)$$

在公式 Equation (1) 中, $F(x)$ 代表特征图 F 在位置 x 的所有通道的向量。我们使用归一化特征 $\bar{F}(x) = \frac{F(x)}{\|F(x)\|}$ 在我们的补片相似度度量中, 因为使用归一化特征可以实现更强的不变性 [Li and Wand 2016]。

反向映射函数 $\phi_{R \rightarrow S}^L$ 从 F_R^L 到 F_S^L 的计算方式与 Equation (1) 相同, 通过交换 S 和 R 。这两种映射 $\phi_{S \rightarrow R}^L$ 和 $\phi_{R \rightarrow S}^L$ 都可以通过 PatchMatch 算法 [Barnes et al. 2009] 有效地优化, 该算法也通过重叠补片的聚合隐式实现了平滑性。

双向对应关系使我们能够使用双向相似性 (BDS) 投票 [Simakov et al. 2008] 来分别重建指导图像 G_L 和特征图 F_G^L 。 G^L 作为颜色转移下一步的指导, 而 F_G^L 用于测量匹配错误:

$$e^L(p) = \| \bar{F}_S^L(p) - \bar{F}_G^L(p) \|^2 \quad (2)$$

在 Equation (5) 中。BDS 投票是在参考 R^{L-1} 以及 F_R^L 上对所有重叠的最近邻补丁进行像素颜色和特征的平均。通过正向 NNF $\phi_{S \rightarrow R}^L$ 和反向 NNF $\phi_{R \rightarrow S}^L$ 实现这一过程。正

¹ R^L 与 F_R^L 相同分辨率, 后者是从参考 R 降采样的。

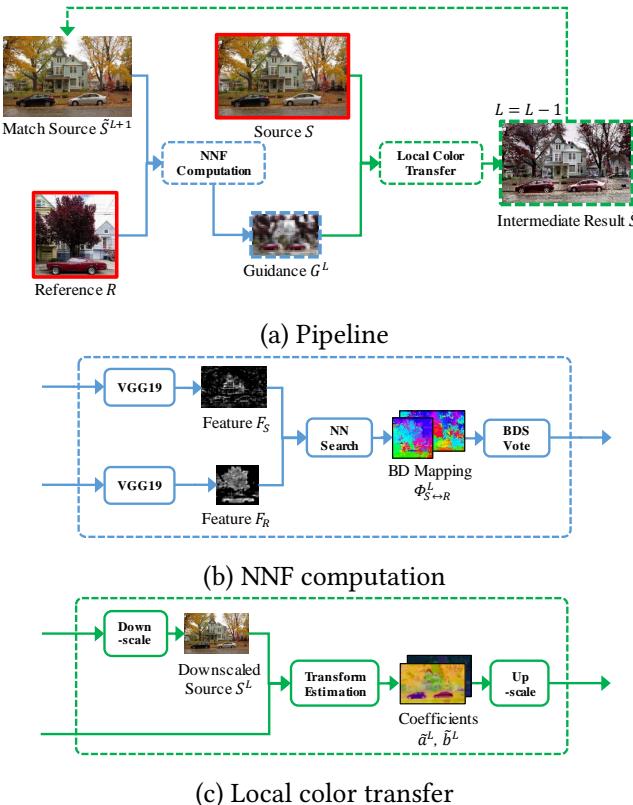


Fig. 3. 系统管道分为两步, 如图所示。我们在特征域中执行 NNF 计算 (用蓝色表示), 并在图像域中应用局部颜色转换 (用绿色表示)。如图 (a) 所示, 原始输入图像是源图像 S 和参考图像 R (在红色框中)。在每一级, 计算 \tilde{S}^{L+1} 和 R 之间的特征的双向 NNFs $\phi_{S \rightarrow R}^L$ 和 $\phi_{R \rightarrow S}^L$, 并用于重建 G^L , 如图 (b) 所示。然后在 (c) 中, 优化 G^L 和下采样源图像 S^L 之间的颜色变换系数 d^L 和 b^L , 然后在上采样到全分辨率之前应用到 S 。颜色转换结果 \tilde{S}^L 作为下一级 $L - 1$ 匹配的输入。上述过程从 $L = 5$ 重复到 $L = 1$ 。

得 \tilde{S}^L 。这两个步骤交替进行, 相互帮助: NNF 帮助获得更准确的局部颜色转移, 而颜色转移结果 \tilde{S}^L 作为源也帮助细化下一个级别 $L - 1$ 中的匹配, 因为 \tilde{S}^L 的颜色与参考相比, 比原始源 S 更相似。两个中间结果 (NNF 和 \tilde{S}^L) 作为匹配和在不同域中发生颜色转移之间的桥梁。遵循这一策略, 两个步骤都逐渐得到了改进。

3.2 Nearest-Neighbor Field Computation

鉴于中间源 \tilde{S}^{L+1} 和参考 R 在级别 $L (L = 5, \dots, 1)$ 上, 我们的 NNF 搜索步骤构建它们之间的双向对应关系。在这

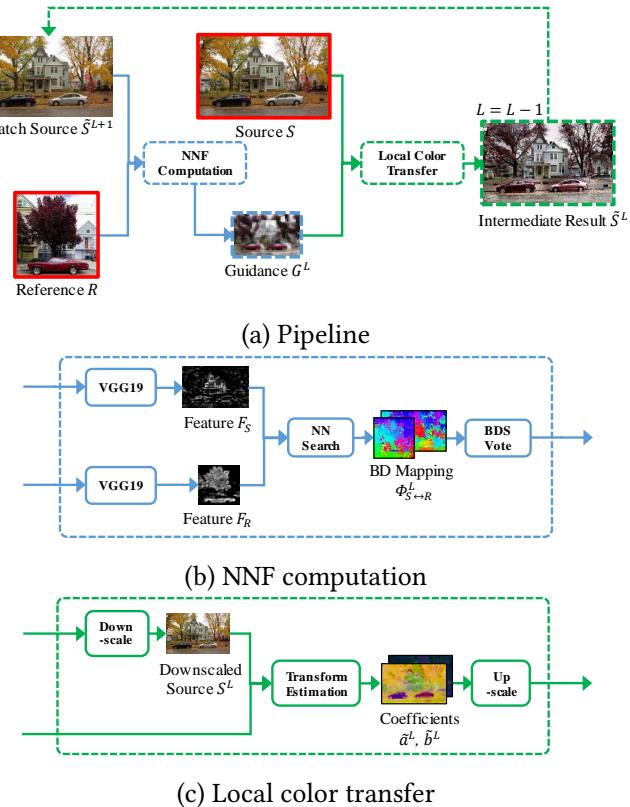


Fig. 3. System pipeline with two steps illustrated below. We perform NNF computation in the feature domain (in blue) and apply local color transfer in the image domain (in green). As shown in (a), the original input images are the source image S and the reference image R (in the red frames). At each level, the bidirectional NNFs $\phi_{S \rightarrow R}^L$ and $\phi_{R \rightarrow S}^L$ are computed between features from \tilde{S}^{L+1} and R , and are used to reconstruct G^L as shown in (b). Then in (c), the color transform coefficients \tilde{a}^L and \tilde{b}^L are optimized between G^L and the down-scaled source image S^L , and then upscaled to the full resolution before being applied to S . The color transferred result \tilde{S}^L serves as the input for the matching of the next level $L - 1$. The above process repeats from $L = 5$ to $L = 1$.

transfer which occur in different domains. Following this strategy, both steps are gradually refined.

3.2 Nearest-Neighbor Field Computation

Given the intermediate source \tilde{S}^{L+1} and the reference R at level L ($L = 5, \dots, 1$), our NNF search step builds the bidirectional correspondences between them. Here, \tilde{S}^{L+1}

is the color transferred result (with the same resolution to the source S) from the coarser level $L + 1$ when $L < 5$. At the coarsest level $L = 5$, \tilde{S}^{L+1} is initialized as S .

In view of the difficulty of building correct correspondences between \tilde{S}^{L+1} and R (potentially with large appearance variations), we perform NNF search in the deep feature domain. Since the CNN will almost always keep the spatial relationship of input images, the NNFs computed in the feature domain can be used in the image domain. To do so, we first feed the \tilde{S}^{L+1} and R into the VGG19 network [Simonyan and Zisserman 2014] pre-trained on the ImageNet database [Russakovsky et al. 2015] for object recognition. We then extract their feature maps in $reluL_1$ layer, labeled as F_S^L , F_R^L respectively for \tilde{S}^{L+1} , R . Each feature map is a 3D tensor with $width \times height \times channel$, and its spatial resolution is $1/4^{L-1}$ of the input.

The mapping function $\phi_{S \rightarrow R}^L$ from F_S^L to F_R^L is computed by minimizing the following energy function:

$$\phi_{S \rightarrow R}^L(p) = \arg \min_q \sum_{x \in N(p), y \in N(q)} \|\bar{F}_S^L(x) - \bar{F}_R^L(y)\|^2 \quad (1)$$

where $N(p)$ is the patch around p . We set the patch size to 3×3 at each level. For each patch around position p in the source feature F_S^L , we find its nearest-neighbor patch around $q = \phi_{S \rightarrow R}^L(p)$ in the reference feature F_R^L . $\bar{F}(x)$ in Equation (1) is a vector representing all channels of feature map F at position x . We use normalized features $\bar{F}(x) = \frac{F(x)}{\|F(x)\|}$ in our patch similarity metric, because the use of normalized features can achieve stronger invariance [Li and Wand 2016].

The reverse mapping function $\phi_{R \rightarrow S}^L$ from F_R^L to F_S^L is computed in the same way as Equation (1) by exchanging S and R . Both mappings $\phi_{S \rightarrow R}^L$ and $\phi_{R \rightarrow S}^L$ can be efficiently optimized with the PatchMatch algorithm [Barnes et al. 2009], which also implicitly achieves smoothness through aggregation of overlapping patches.



Fig. 4. 最终的颜色转移结果通过BDS投票,各种完整性权重 w 。与平均投票的结果($w = 0$)相比,随着 w 的增加,使用的参考颜色越多。输入图像:Anonymous/pxhere.com。

向NNF确保一致性 (即, 每个源中的补丁都能在参考中找到), 而反向NNF确保完整性 (即, 源中的每个补丁都能在参考中找到)。通过同时确保一致性和完整性, BDS投票可以鼓励 G^L 中比仅使用正向NNF的平均投票更多的参考颜色。Fig. 4显示了各种完整性权重 (默认使用2) 下的最终结果。

我们在 Fig. 6 中展示了NNFs $\phi_{R \rightarrow S}^L$, 以及从粗层到细层逐渐精化的指导图像 G^L ($L = 5, \dots, 1$)。

3.3 Local Color Transfer

鉴于层级 L 的指导图像 G^L , 我们提出了一种新的局部颜色转移算法, 该算法将源 S 的颜色改变得与 G^L 的颜色更好地匹配。然后, 我们得到颜色转移结果 \tilde{S}^L 。由于 S 和 G^L 在粗略层级 ($L > 1$) 具有不同的分辨率, 在 S 和 G^L 之间建立直接对应关系是不可能的。相反, 我们将 S 下缩至 S^L 以匹配 G^L 的分辨率, 估计从 S^L 到 G^L 的颜色转移函数, 并在将该函数参数上缩之前, 使用边缘保留滤波器对其进行估计, 然后将其应用于全分辨率 S 以得到 \tilde{S}^L 。 \tilde{S}^L 是用于下一级别NNF搜索的中间转移结果。

受到 Reinhard et al. [2001] 的启发, 该文通过匹配像素颜色的全局均值和方差来构建颜色转移函数, 我们对于 S^L 中的每个像素 p 在 CIELAB 颜色空间中将局部颜色匹配建模为每个通道的线性函数, 表示为:

$$\mathcal{T}_p^L(S^L(p)) = a^L(p)S^L(p) + b^L(p). \quad (3)$$

(如果我们只考虑 $b^L(p)$ 并且 $a^L(p)$ 被设置为零, 那么只有均值是可一致的。)

我们的目标是估计每个像素 p 的线性系数 $a^L(p)$ 和 $b^L(p)$, 使得传输的结果 $\mathcal{T}_p^L(S^L(p))$ 在视觉上与指

导 $G^L(p)$ 相似。我们通过最小化以下包含三个项的目标函数来形式化估计 \mathcal{T}^L 的问题:

$$E(\mathcal{T}^L) = \sum_p E_d(p) + \lambda_l \sum_p E_l(p) + \lambda_{nl} \sum_p E_{nl}(p), \quad (4)$$

其中 λ_l 和 λ_{nl} 是折衷权重 (默认情况下, $\lambda_l = 0.125$ 和 $\lambda_{nl} = 2.0$)。

第一个数据项 E_d 使颜色转移结果与指导 G^L 相似:

$$E_d(p) = \omega(L)(1 - \bar{e}^L(p))\|\mathcal{T}_p^L(S^L(p)) - G^L(p)\|^2, \quad (5)$$

其中 \bar{e}^L 是 Equation (2) 中的归一化匹配误差, 用作权重, 对匹配得好的点给予高置信度。 $\omega(L) = 4^{L-1}$ 是归一化因子, 使这一项在不同级别上与分辨率无关。

第二个平滑项 E_l , 其定义与基于WLS滤波器的平滑项 [Farbman et al. 2008] 相同, 鼓励局部相邻像素具有相似的线性变换, 同时保持源 S^L 中的边缘:

$$E_l(p) = \sum_{q \in N_4(p)} \omega_l(p, q)(\|a_p^L - a_q^L\|^2 + \|b_p^L - b_q^L\|^2), \quad (6)$$

其中 $N_4(p)$ 表示在 p 处的4连通邻域。至于平滑权重, 我们的定义与 [Lischinski et al. 2006] 中的定义相同:

$$\omega_l(p, q) = (\|\ell(p) - \ell(q)\|^\alpha + \epsilon)^{-1} \quad (7)$$

where ℓ 是 S^L 的亮度通道, 指数 $\alpha = 1.2$, 并且小常数 $\epsilon = 0.0001$ 。

最后一个平滑项 E_{nl} 强制执行非局部约束, 以惩罚全局不一致性。它是基于这样一个假设: 源图像中



Fig. 4. Final color transfer results by BDS voting with various completeness weights w . Compared to the result by average voting ($w = 0$), more reference colors are used as w increases. Input images: Anonymous/pxhere.com.

The bidirectional correspondences allow us to use Bidirectional Similarity (BDS) voting [Simakov et al. 2008] to respectively reconstruct the guidance image G^L and the feature map F_G^L . G^L serves as the guidance for color transfer in the next step, while F_G^L is used to measure matching errors:

$$e^L(p) = \|\bar{F}_S^L(p) - \bar{F}_G^L(p)\|^2 \quad (2)$$

in Equation (5). The BDS voting is performed to average the pixel colors and features from all overlapping nearest-neighbor patches in the reference R^{L-1} and F_R^L through the forward NNF $\phi_{S \rightarrow R}^L$ and the backward NNF $\phi_{R \rightarrow S}^L$. The forward NNF enforces coherence (*i.e.*, each patch in the source can be found in the reference), while the backward NNF enforces completeness (*i.e.*, each patch in the reference can be found in the source). By enforcing both coherence and completeness, BDS voting can encourage more reference colors in G^L than average voting with solely forward NNF. Fig. 4 shows a set of final results with various completeness weights (using 2 as the default).

We show the NNFs $\phi_{R \rightarrow S}^L$ and guidance image G^L ($L = 5, \dots, 1$) gradually refined from the coarse layer to the fine layer in Fig. 6.

3.3 Local Color Transfer

Given the guidance image G^L at the level L , we propose a new local color transfer algorithm, which changes the colors of the source S to better match those of G^L . Then, we

¹ R^L is the same resolution as F_R^L , downsampled from the reference R

get the color transfer result \tilde{S}^L . Since S and G^L have different resolutions at the coarse levels ($L > 1$), it is impossible to build in-place correspondence between S and G^L . Instead, we downscale S to S^L to match the resolution of G^L , estimate the color transfer function from S^L to G^L , and upscale the function parameters with an edge-preserving filter before applying it to the full-resolution S to get \tilde{S}^L . \tilde{S}^L is the intermediate transferred result used for the NNF search at the next level.

Inspired by Reinhard et al. [2001] which constructs a color transfer function by matching the global means and variances of pixel colors, we model the local color matching as a linear function of each channel in CIELAB color space for every pixel p in S^L , denoted as:

$$\mathcal{T}_p^L(S^L(p)) = a^L(p)S^L(p) + b^L(p). \quad (3)$$

(If we consider $b^L(p)$ only with $a^L(p)$ being set to zero, only the means are consistent.)

We aim to estimate linear coefficients $a^L(p)$ and $b^L(p)$ for each pixel p , making the transferred result $\mathcal{T}_p^L(S^L(p))$ visually similar to the guidance $G^L(p)$. We formulate the problem of estimating \mathcal{T}^L by minimizing the following objective function consisting of three terms:

$$E(\mathcal{T}^L) = \sum_p E_d(p) + \lambda_l \sum_p E_l(p) + \lambda_{nl} \sum_p E_{nl}(p), \quad (4)$$

where λ_l and λ_{nl} are trade-off weights (by default, $\lambda_l = 0.125$ and $\lambda_{nl} = 2.0$).

The first data term E_d makes the color transfer result similar to the guidance G^L :

$$E_d(p) = \omega(L)(1 - \bar{e}^L(p))\|\mathcal{T}_p^L(S^L(p)) - G^L(p)\|^2, \quad (5)$$



Fig. 5. 与无非局部约束的比较 对应于 $\lambda_{nl} = 0.0$ 和 $\lambda_{nl} = 2.0$ 分别。输入图像: Luan et al. [2017]。

具有相同颜色的像素在结果中应该获得相似的转移颜色。该约束已经在脱晕 [Chen et al. 2013]、intrinsic 图像分解 [Zhao et al. 2012] 和颜色化 [Endo et al. 2016] 中成功应用。我们考虑颜色和语义的相似性来计算非局部项。我们首先应用 K-means 算法将所有像素分成 k 个组, 根据它们在粗略 (最语义) 层 $relu5_1$ 上的特征距离 (我们设置 $k = 10$)。在每个簇内部, 我们找到 S^L 像素 p 的 K 个最近邻在颜色空间中, 标记为 $K(p)$ (我们设置 $K = 8$)。非局部平滑项然后被定义为:

$$E_{nl}(p) = \sum_{q \in K(p)} \omega_{nl}(p, q) \|\mathcal{T}_p^L(S^L(p)) - \mathcal{T}_q^L(S^L(q))\|^2, \quad (8)$$

其中 $\omega_{nl}(p, q) = \frac{\exp(1 - SSD(S^L(p) - S^L(q)))}{K}$ 是由 p 和其在CIELAB颜色空间中的非局部邻居 q 的颜色相似性决定的。通过非局部约束, 可以减少 artifacts, 从而使颜色转移在全局上更加一致, 如图 Fig. 5 所示。

Equation (4)的解析解由于其不规则的稀疏矩阵结构而非常昂贵。相反, 我们的替代解首先估计一个好的初始猜测, 然后进行几次共轭梯度迭代, 这为 \mathcal{T}^L 实现了更快的收敛。

我们通过在每个局部补丁上应用全局颜色变换方法 [Reinhard et al. 2001]来初始化 \mathcal{T}^L 。具体来说, 取

S^L 和 G^L 中的每个像素 p 为中心的一个补丁 $N(p)$, 我们通过分别在每个颜色通道上匹配补丁对的均值 μ 和标准差 σ 来估计初始化的 \mathcal{T}^L :

$$\begin{aligned} a^L(p) &= \sigma_{G^L(N(p))}/(\sigma_{S^L(N(p))} + \epsilon) \\ b^L(p) &= \mu_{G^L(N(p))} - a^L(p)\mu_{S^L(N(p))}, \end{aligned} \quad (9)$$

where ϵ is used to avoid dividing zero ($\epsilon = 0.002$ for color range $[0, 1]$). We set the patch size to be 3 for all layers.

The above parameters $a^L(p)$ and $b^L(p)$ are estimated in a low resolution. We also apply the WLS-based operator with the smoothness term matching that in Equation (6) to upsample them to the full-resolution, which is guided by source image S , obtaining a_{\uparrow}^L and b_{\uparrow}^L . The smoothness weight is set to 0.024 by default.

Next, we get

$$\tilde{S}^L(p) = a_{\uparrow}^L(p)S(p) + b_{\uparrow}^L(p), \quad \forall p \in \tilde{S}^L. \quad (10)$$

结果 \tilde{S}^L (见图 Fig. 6) 然后用于下一层 $L-1$ 的 NNF 搜索来更新对应关系。一旦达到最精细的层 $L = 1$, \tilde{S}^1 就是我们的最终输出。

我们的实现伪代码列于算法1。

3.4 Extension to Multi-reference Color Transfer

我们的算法可以扩展到多参考颜色转移。这避免了必须选择一个适合源图像所有部分的单一适当参考图像的困难。一对一匹配 (如Section 3.2节所述) 可以扩展到一对多匹配, 如下所示。

鉴于多个参考文献 R_i ($i = 1, \dots, n$), 我们在每个级别 L 计算 \tilde{S}^{L+1} 和每个参考文献 R_i 之间的双向 NNFs, 然后使用获得的 NNFs 重建每个指导图像 G_i^L (如图 Fig. 7 所示)。接下来, 我们将这些指导图像组合成一个单一的 G^L , 这需要在每个像素 p 上从 n 个候选者 G_i^L ($i = 1, \dots, n$) 中选择最佳的一个。选择标准包括: (1) 参考文献与源的匹配程度; (2) 结果像素颜

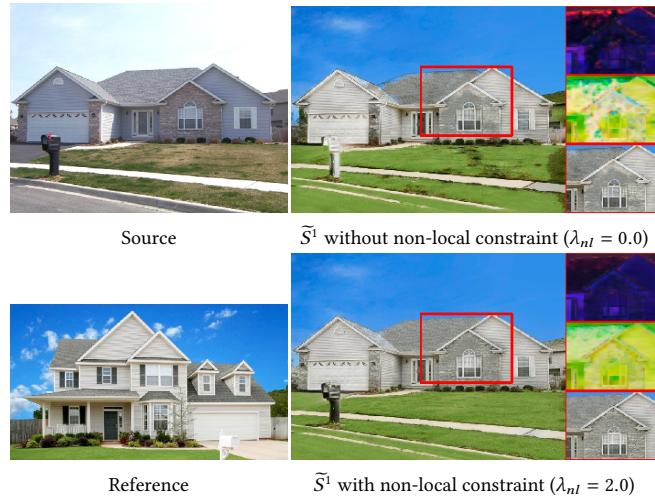


Fig. 5. Comparison with and without non-local constraint corresponding to $\lambda_{nl} = 0.0$ and $\lambda_{nl} = 2.0$ respectively. Input images: Luan et al. [2017].

where \bar{e}^L is the normalized matching error in Equation (2), used as the weight to give high confidence to well-matched points. $\omega(L) = 4^{L-1}$ is the normalization factor to make this term resolution-independent at different levels.

The second smoothness term E_l , which is defined in the same manner as the smooth term in WLS-based filter [Farbman et al. 2008], encourages locally adjacent pixels to have similar linear transforms while preserving edges in the source S^L :

$$E_l(p) = \sum_{q \in N_4(p)} \omega_l(p, q) (\|a_p^L - a_q^L\|^2 + \|b_p^L - b_q^L\|^2), \quad (6)$$

where $N_4(p)$ is the 4-connected neighborhood at p . As for the smoothness weights, we define them in the same way as in [Lischinski et al. 2006]:

$$\omega_l(p, q) = (\|\ell(p) - \ell(q)\|^\alpha + \epsilon)^{-1} \quad (7)$$

where ℓ is the luminance channel of S^L and the exponent $\alpha = 1.2$ and the small constant $\epsilon = 0.0001$.

The last smoothness term E_{nl} enforces the non-local constraint to penalize global inconsistency. It is based on the assumption that pixels with identical colors in the source should get similar transferred colors in the result.

The constraint has been successfully applied in matting [Chen et al. 2013], intrinsic image decomposition [Zhao et al. 2012] and colorization [Endo et al. 2016]. We consider the similarity of both color and semantics to compute the non-local term. We first apply K-means to cluster all the pixels into k groups according to their feature distance at the coarsest (most semantic) layer *relu5_1* (we set $k = 10$). Inside each cluster, we find the K nearest neighbors in the color space for each pixel p of S^L , labeled as $K(p)$ (we set $K = 8$). The non-local smoothness term is then defined as:

$$E_{nl}(p) = \sum_{q \in K(p)} \omega_{nl}(p, q) \|\mathcal{T}_p^L(S^L(p)) - \mathcal{T}_q^L(S^L(q))\|^2, \quad (8)$$

where $\omega_{nl}(p, q) = \frac{\exp(1-SSD(S^L(p)-S^L(q)))}{K}$ is determined by the color similarity between p and its non-local neighbor q in the CIELAB color space. With the non-local constraints, artifacts are reduced and color transfer is thus more globally consistent as shown in Fig. 5.

The closed-form solution of Equation (4) is very costly due to the irregular sparse matrix structure. Instead, our alternative solution first estimates a good initial guess and then performs a few conjugate gradient iterations, which achieves much faster convergence for \mathcal{T}^L .

We initialize \mathcal{T}^L by applying the global color transformation method [Reinhard et al. 2001] on every local patch. Specifically, taking a patch $N(p)$ centered at pixel p in S^L and in G^L , we estimate the initialized \mathcal{T}^L by matching the mean μ and standard deviation σ of the patch pair in each color channel separately:

$$\begin{aligned} a^L(p) &= \sigma_{G^L(N(p))} / (\sigma_{S^L(N(p))} + \epsilon) \\ b^L(p) &= \mu_{G^L(N(p))} - a^L(p) \mu_{S^L(N(p))}, \end{aligned} \quad (9)$$

where ϵ is used to avoid dividing zero ($\epsilon = 0.002$ for color range $[0, 1]$). We set the patch size to be 3 for all layers.

The above parameters $a^L(p)$ and $b^L(p)$ are estimated in a low resolution. We also apply the WLS-based operator with the smoothness term matching that in Equation (6) to upsample them to the full-resolution, which is guided

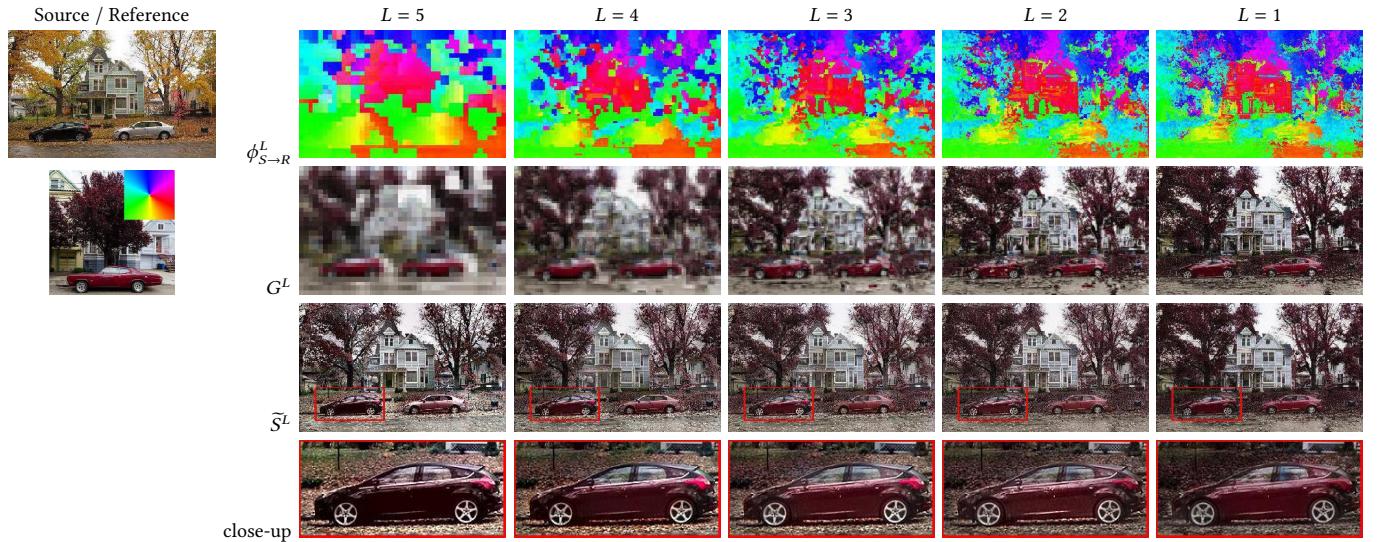


Fig. 6. 神经网络结构函数 $\phi_{S \rightarrow R}^L$ 、重建结果 G^L 和颜色转移结果 S^L 是从高到低逐级精化的。可以观察到，在最低级别建立了语义相似物体的对应关系，并逐步细化，例如，黑色汽车和灰色汽车被映射到红色汽车，黄色树木被映射到红色树木，房屋被映射到白色房屋。随着级别的降低，匹配物体的颜色外观越来越接近。

ALGORITHM 1: 单参照颜色转移算法

Input : Source image S and reference image R .

Initialization:

$$\tilde{S}^6 = S.$$

for $L = 5$ to 1 **do**

NNF search (Section 3.2):

$F_S^L, F_R^L \leftarrow$ feed \tilde{S}^{L+1} , R to VGG19 and get features.

$\phi_{S \rightarrow R}^L \leftarrow$ map F_R^L to F_S^L by Equation (1).

$\phi_{R \rightarrow S}^L \leftarrow$ map F_S^L to F_R^L .

$G^L \leftarrow$ reconstruct S^L with R^L by BDS voting.

Local color transfer (Section 3.3):

$a^L, b^L \leftarrow$ optimize local linear transform from S^L to G^L by minimizing Equation (4).

$a_{\uparrow}^L, b_{\uparrow}^L \leftarrow$ upscale and a^L, b^L with WLS-based filter guided by S .

$\tilde{S}^L \leftarrow$ transfer the color of S by Equation (10).

end

Output: Color transferred result \tilde{S}^1 .

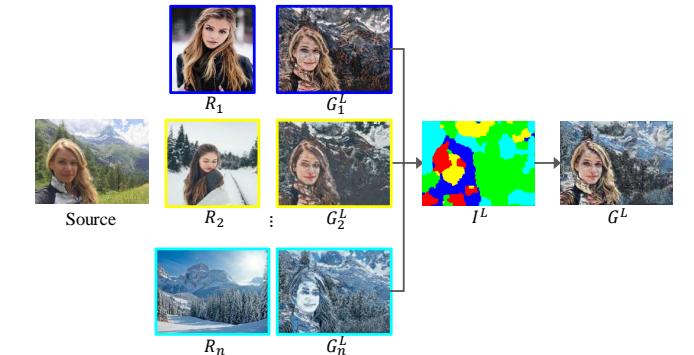


Fig. 7. 在多参考场景中，用于指导颜色转移的图像 G^L 是合并自多个参考 $G_i^L, i = 1, \dots, n$ 通过双向NNFs重建的。

请注意，由于中文和英文在语法和表达方式上的差异，上述翻译可能需要根据具体的学术论文内容进行调整。此外，由于中文的表达习惯，某些英文术语可能需要重新翻译或解释，以确保翻译的准确性和可读性。

$$\mathcal{E}(I^L) = \sum_p \mathcal{E}_e(p) + \beta_c \sum_p \mathcal{E}_c(p) + \beta_l \sum_p \mathcal{E}_l(p), \quad (11)$$

where 我们设折衷权重 $\beta_c = 0.2$ 和 $\beta_l = 0.08$ 。

ALGORITHM 1: Single-reference Color Transfer Algorithm

Input : Source image S and reference image R .

Initialization:

$$\tilde{S}^6 = S.$$

for $L = 5$ to 1 **do**

NNF search (Section 3.2):

$$F_S^L, F_R^L \leftarrow \text{feed } \tilde{S}^{L+1}, R \text{ to VGG19 and get features.}$$

$$\phi_{S \rightarrow R}^L \leftarrow \text{map } F_R^L \text{ to } F_S^L \text{ by Equation (1).}$$

$$\phi_{R \rightarrow S}^L \leftarrow \text{map } F_S^L \text{ to } F_R^L.$$

$$G^L \leftarrow \text{reconstruct } S^L \text{ with } R^L \text{ by BDS voting.}$$

Local color transfer (Section 3.3):

$$a^L, b^L \leftarrow \text{optimize local linear transform from } S^L \text{ to } G^L \text{ by minimizing Equation (4).}$$

$$a_\uparrow^L, b_\uparrow^L \leftarrow \text{upscale and } a^L, b^L \text{ with WLS-based filter guided by } S.$$

$$\tilde{S}^L \leftarrow \text{transfer the color of } S \text{ by Equation (10).}$$

end

Output: Color transferred result \tilde{S}^1 .

by source image S , obtaining a_\uparrow^L and b_\uparrow^L . The smoothness weight is set to 0.024 by default.

Next, we get

$$\tilde{S}^L(p) = a_\uparrow^L(p)S(p) + b_\uparrow^L(p), \quad \forall p \in \tilde{S}^L. \quad (10)$$

The result \tilde{S}^L (in Fig. 6) is then used for the NNF search at the next level $L-1$ to update the correspondences. Once the finest layer $L=1$ is reached, \tilde{S}^1 is our final output.

The pseudo code of our implementation is listed in Algorithm 1.

3.4 Extension to Multi-reference Color Transfer

Our algorithm is extendable to multi-reference color transfer. This avoids the difficulty of having to choose a single proper reference image that is suitable for all portions of the source image. The *one-to-one* matching (described

in Section 3.2) can be extended to the *one-to-many* matching as follows. Given multiple references $R_i (i=1, \dots, n)$, we compute the bidirectional NNFs between \tilde{S}^{L+1} and every reference R_i at each level L , then reconstruct each guidance image G_i^L (shown in Fig. 7) using obtained NNFs. Next, we combine these guidance images into a single G^L , which requires the selection of the best one from n candidates $G_i^L (i=1, \dots, n)$ at each pixel p . The selection criteria include: (1) how well the reference is matched to the source; (2) how similar the resulting pixel color is compared to the majority of guidance colors; (3) how consistently the indices are selected between the pixel and its neighborhoods. Based on these criteria, we compute the index selection map I^L by minimizing the following objective function:

$$\mathcal{E}(I^L) = \sum_p \mathcal{E}_e(p) + \beta_c \sum_p \mathcal{E}_c(p) + \beta_l \sum_p \mathcal{E}_l(p), \quad (11)$$

where we set the trade-off weights $\beta_c = 0.2$ and $\beta_l = 0.08$.

As in Equation (2), e_i^L be the feature error map of i -th NNFs. The first term penalizes the feature dissimilarity between S and R_i for each pixel at the layer *reluL_1*:

$$\mathcal{E}_e(p) = \omega_e(L)e_i^L(p) \quad (12)$$

where $\omega_e(L) = 4^{L-5}$ is the normalization factor for different levels.

The second term measures the difference between each guidance color and the guidance “majority” color at every pixel. To compute the majority color, we first build a per-pixel color histogram of guidance colors with n bins of each channel (we set $n=8$), and take the mean of the colors that fall in the densest bin.

$$\mathcal{E}_c(p) = \omega_c(L)\|G_i^L(p) - \text{majority}(G_i^L(p))\|^2 \quad (13)$$

where $\omega_c(L) = \omega_e(L)$ is the normalization factor.

The third term measures the local smoothness, which encourages neighboring features in the combination result

如 Equation (2)所示, e_i^L 是第*i*个NNFs的特征误差图。第一项惩罚每个像素在层*reluL_1*上的S和*R_i*之间的特征差异:

$$\mathcal{E}_e(p) = \omega_e(L)e_i^L(p) \quad (12)$$

where $\omega_e(L) = 4^{L-5}$ 是不同层次的归一化因子。

第二个项测量每像素处的每种指导颜色与其指导“多数”颜色之间的差异。为了计算多数颜色，我们首先在每个通道上建立一个像素颜色直方图，每个通道有*n*个bins（我们设*n*=8），并取落在最密集bin中的颜色的均值。

$$\mathcal{E}_c(p) = \omega_c(L)\|G_i^L(p) - \text{majority}(G_i^L(p))\|^2 \quad (13)$$

where $\omega_c(L) = \omega_e(L)$ 是归一化因子。

第三项衡量局部平滑性，鼓励组合结果中的相邻特征保持一致:

$$\mathcal{E}_l(p) = \omega_l(L) \sum_{q \in N_4(p)} \|\bar{F}_{G_i}^L(p) - \bar{F}_{G_j}^L(p)\|^2 + \|\bar{F}_{G_i}^L(q) - \bar{F}_{G_j}^L(q)\|^2, \quad (14)$$

where $i = I^L(p)$ and $j = I^L(q)$ and where $\omega_l(L) = 2^{L-5}$ 是归一化因子。

Equation (11)公式化了一个在2D空间域上的马尔可夫随机场(MRF)问题，可以通过使用多标签图割来有效地解决这个问题[Kolmogorov and Zabin 2004]。为了获得优化的良好初始化， $I^L(p)$ 仅通过最小化数据项($\mathcal{E}_e(p)$ 和 $\mathcal{E}_c(p)$)来初始化。解决完 I^L 后，我们通过简单合并来自多个参考的所有结果，即 $G^L(p) = G_{I^L(p)}^L(p)$ ，获得一个单一的指导图像 G^L 。然后， G^L 用于局部颜色转移描述的Section 3.3后续步骤。Fig. 7的右侧图像显示了最优参考索引图 I^L 和合并的指导 G^L 。

与单参考匹配相比，它能够在难以找到合适单源的情况下有效地解决内容不匹配问题。

4 EVALUATION AND RESULTS

4.1 Performance

我们的核心算法是在CUDA平台上开发的。所有的实验都是在具有Intel E5 2.5GHz CPU和NVIDIA Tesla K40c GPU的PC上进行的。单参考颜色转移的运行时间大约为60秒，其分辨率约为700×500。处理过程中有两个瓶颈：深度PatchMatch (~40秒)，它需要在数百个特征通道上计算块相似性，以及局部颜色变换的优化（大约10–20秒），它需要求解大型稀疏方程。对于多参考颜色转移，总运行时间包括深度PatchMatch的时间，该时间与参考数量成正比，局部颜色变换优化的时间（与单参考相同）以及MRF的求解 (~5秒)。

4.2 Evaluation

我们通过三个研究分析了并评估了我们算法的不同组件。

联合优化.为了验证我们的交错联合优化过程对对应估计和局部颜色转换的益处，我们进行了消融实验，其中两个步骤被分为两个独立的步骤。在第一步中，使用Section 3.2节或更先进的方法[Liao et al. 2017]描述的方式，将NNF计算进行到底层。此外，在第二步中，根据获得的对应关系应用我们的颜色转换方法Section 3.3。为了减少颜色带来的影响，我们也在灰度图像对和彩色图像对上进行了测试，其结果分别在Fig. 8中的第一行和第二行示例中显示。除了我们在Section 3.2节中描述的NNF搜索方法外，我们还在Fig. 8中的第三行中使用了Liao et al. [2017]的方法来计算密集对应关系。与联合优化(Fig. 8中的最后一列)相比，我们发现，在我们的方法中，如果没有局部颜色转换作为连接连续级别对应估计的桥梁，高级别的语义匹配几乎不会影响低级别级别的匹配。因此，最终的对应关系主要受低级别特征的影响(例如，对于彩色图像，亮度与色度；对于灰度图像，亮度)，

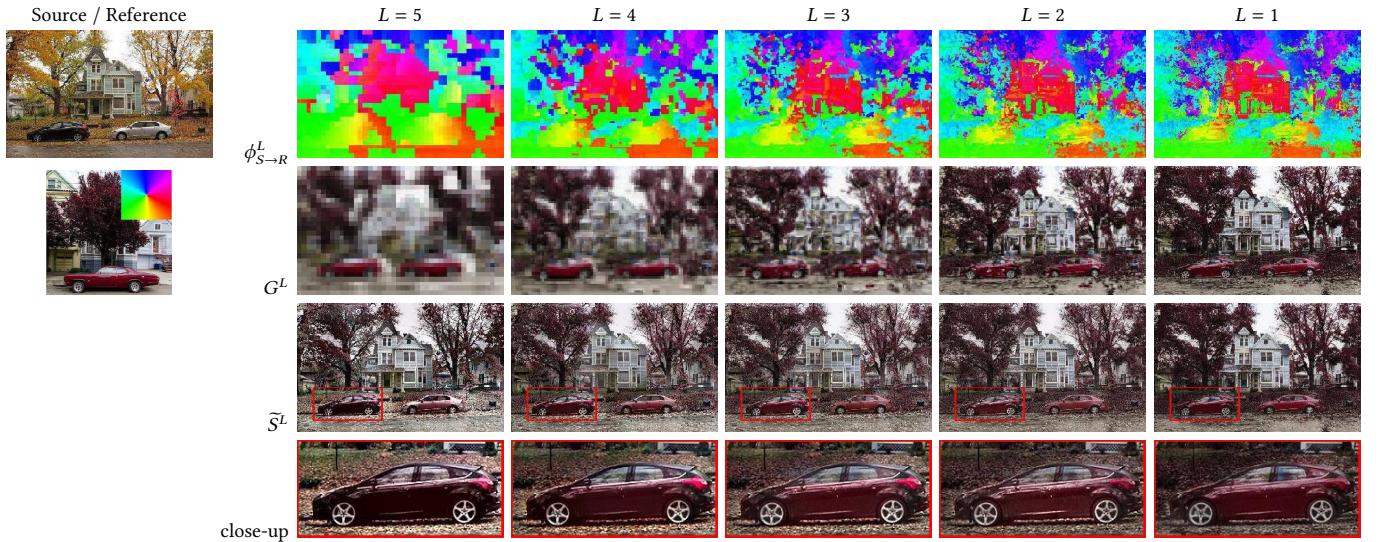


Fig. 6. The NNF $\phi_{S \rightarrow R}^L$, reconstructed results G^L and color transfer results \tilde{S}^L are gradually refined from high to low level. It can be observed that correspondence between semantically similar objects are built at the coarsest level and refined progressively, for example, both the black car and gray car mapped to the red car, the yellow trees mapped to the red trees, and the house mapped to the white house. As the level goes down, the color appearance of the matched objects gets closer.

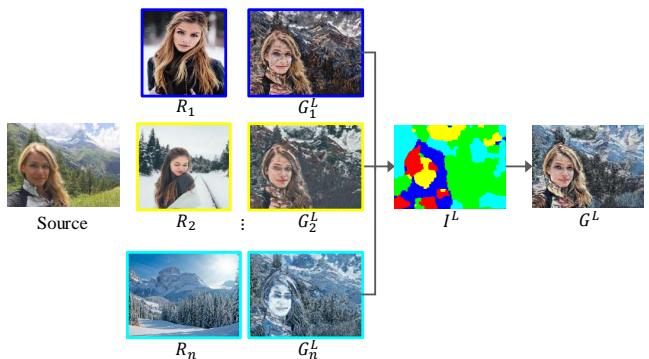


Fig. 7. In a multiple-reference scenario, the image G^L used to guide color transfer is merged from multiple guidances $G_i^L, i = 1, \dots, n$ reconstructed with bidirectional NNFs.

to be consistent:

$$\mathcal{E}_l(p) = \omega_l(L) \sum_{q \in N_4(p)} \|\bar{F}_{G_i}^L(p) - \bar{F}_{G_j}^L(p)\|^2 + \|\bar{F}_{G_i}^L(q) - \bar{F}_{G_j}^L(q)\|^2, \quad (14)$$

where $i = I^L(p)$ and $j = I^L(q)$ and where $\omega_l(L) = 2^{L-5}$ is the normalization factor.

Equation (11) formulates a Markov Random Field (MRF) problem over the 2D spatial domain, which can

be efficiently solved by using multi-label graph cut [Kolmogorov and Zabin 2004]. To obtain a good initialization for the optimization, $I^L(p)$ is initialized by minimizing only the data terms ($\mathcal{E}_e(p)$ and $\mathcal{E}_c(p)$). After solving I^L , we obtain a single guidance image G^L by simply merging all results from multiple references, *i.e.*, $G^L(p) = G_{I^L(p)}^L(p)$. Then, G^L is used for the following step of local color transfer described in Section 3.3. The right image of Fig. 7 shows the optimal reference index map I^L and the merged guidance G^L . Compared to single-reference matching, it can effectively solve the content-mismatch problem in situations where it is difficult to find a suitable single source.

4 EVALUATION AND RESULTS

4.1 Performance

Our core algorithm was developed in CUDA. All of our experiments were conducted on a PC with an Intel E5 2.5GHz CPU and an NVIDIA Tesla K40c GPU. The runtime is approximately 60 seconds for single-reference color transfer

并且随后的颜色转换结果在语义上是错误的。Liao et al. [2017]通过在连续级别之间部分混合高级别的重建特征来构建连接，用于低级别级别的NNF计算，因此他们的结果可以保留语义对应关系。然而，由于不变的低级别特征与重建特征混合，神经表示仍然受到颜色差异的影响，并且在精细图像结构上可能出现对齐错误。这种方法适合风格转换，但可能对颜色转换产生失真效果，例如，在第一例中的斑驳屋顶和第三例中的不清晰建筑边界。相比之下，我们的方法由于联合优化，可以保留精细图像细节和语义关系。

对应关系的深度特征.在最近的工作中，已经验证了深度特征足以匹配尽管外观差异很大但语义上相似的对象 [Li and Wand 2016; Liao et al. 2017]。我们设计了一个研究来验证它们在我们渐进方法中的重要性。我们将深度特征替换为低级别特征（亮度、颜色和SIFT）。对于第一行在 Fig. 9中的图像对，颜色特征足以建立对应关系。然而，对于差异较大的图像对，只有深度特征才能匹配语义上相似的对象，例如 Fig. 9中的房屋。

局部颜色转换.为了验证我们的局部颜色转换方法在渐进算法中的鲁棒性，以及将其作为后处理步骤与其他颜色转换方法进行比较，我们首先评估了我们的优化在渐进算法中的有效性，然后应用我们的局部颜色转换作为后处理步骤。

我们进行了颜色变换的消融实验，将它替换为全局颜色变换 [Reinhard et al. 2001]，以及从 [Reinhard et al. 2001]采用的局部颜色变换，这也是我们优化的初始化，如图 Fig. 10所示。全局颜色变换只能调整全局色调，但无法反映空间变化的影响。当全局方法用于匹配局部块的平均值和方差时，空间色彩特征得以保留，但全局不一致的artifact出现（如第二行的头发和第三行的地面）。我们的颜色传输既保证了局部又保证了全局的一致性，有效地避免了如鬼影、光晕和不一致的artifact。

接下来，我们展示了当将我们的局部颜色传输与通过神经风格传输 [Gatys et al. 2015]在后期处理阶段获

得的区域到区域对应关系相结合时，其效果如何。每例的中间结果如图 Fig. 11的(a)所示。除了我们的方法外，还有几种方法被提出，基于这些中间结果来传输颜色。Luan et al. [2017]在颜色空间中约束颜色变换为局部线性，如图(b)所示。Mechrez et al. [2017]使用筛选泊松方程来提高照片的真实感，如图(c)所示。Liao et al. [2017]仅将源的颜色低频带替换为对齐参考的颜色低频带，如图(d)所示。与上述方法相比，我们的方法在保留细节和图像边界方面表现更好，如第一例的(e)所示，并且在忠实于参考方面做得更好，与相似的色差和对比度保持一致，并且不会引入任何新的颜色，如 Mechrez et al. [2017]在第二例中所示。然而，作为后期处理步骤，我们的颜色传输算法仍然无法修复大的对应关系错误（例如，第二例中的黑色窗户和第三例中的红色树木）。这也显示了我们联合优化方案的必要性。

4.3 Single-Reference Color Transfer

为了验证我们的颜色转移方法，我们首先讨论了与先前工作在传统和深度颜色转移的视觉比较，然后报告了我们进行的感知研究的统计数据。

在 Fig. 12中，我们比较了我们的方法与传统全局颜色转移方法。Reinhard et al. [2001]和Pitie et al. [2005]仅在源图像和参考图像的颜色统计之间进行匹配，因此限制了它们进行更复杂的颜色变换的能力。例如，在第二个结果中，房子被渲染成黑色，与天空的颜色匹配。相比之下，我们的转移是局部的，能够处理语义对象到对象的色彩转移。

接下来，我们在 Fig. 13中比较了我们的方法与基于局部对应关系的传统颜色转移方法。NRDC方法[HaCohen et al. 2011]基于少量可靠的匹配来估计全局颜色映射函数，因此实现了更空间变异性结果。NRDC适合于常见场景的图像对（例如， Fig. 13左半部分）。在这样的场景中，我们的方法建立了更密集的对应关系，并应用了局部而不是全局变换，因此我们的颜色转移在局部区域（如第二个例子中的孩子不在参考中，第三个例子中的包）产生了更准确的结果，比NRDC少

with an approximate resolution of 700×500 . There are two bottlenecks in the processing: the deep PatchMatch (~ 40 seconds), which needs to compute patch similarities on hundreds of feature channels, and the optimization of local color transform (approximately 10–20 seconds), which requires solving large sparse equations. For multi-reference color transfer, the total runtime involves the time of deep PatchMatch which is proportional to the number of references, the optimization of local color transform (same as single-reference) and the solution of MRF (~ 5 seconds).

4.2 Evaluation

We analyze and evaluate the different components of our algorithm through three studies.

Joint optimization. To verify that correspondence estimation and local color transfer benefit from our interleaved joint optimization process, we conduct an ablation where both steps are separated as two individual steps. In the first step, the NNF computation is conducted down to the finest level using the way described in Section 3.2 or the more advanced method [Liao et al. 2017]. Additionally, in the second step, our color transform method (Section 3.3) is applied based on the obtained correspondences. To reduce the influence brought by color, we test it on grayscale image pairs as well as color image pairs, and their results are respectively shown in the 1st and 2nd rows of each example in Fig. 8. Besides our NNF search method in Section 3.2, we also use the approach by Liao et al. [2017] to compute the dense correspondence shown in the 3rd row in Fig. 8. Compared to the joint optimization (last column in Fig. 8), we find that in our method, without local color transfer as the bridge to connect the correspondence estimations of two consecutive levels, the semantic matching of the higher level barely influences the matching of the lower levels. Thus, the final correspondences are dominated by low-level features (e.g., luminance and chrominance for color images, or luminance

for gray images) and the subsequent color transfer results are semantically incorrect. Liao et al. [2017] build a connection between two consecutive levels by partially blending higher-level reconstructed features for lower-level NNF computation, so their result can preserve the semantic correspondence. However, as the unchanged lower-level features are blended with the reconstructed features, the neural representations are still influenced by the color discrepancy and may misalign at fine-scale image structures. This method is suitable for style transfer but may result in distortion artifacts for color transfer, for example, the mottled roof in the 1st example and the unclear building boundaries in the 3rd example. In contrast, our method can preserve the fine image details and the semantic relationship because of the joint optimization.

Deep features for correspondences. In recent works, deep features have been verified to be robust enough to match semantically similar objects despite large appearance differences [Li and Wand 2016; Liao et al. 2017]. We designed a study to validate their importance in our progressive method. We replace deep features with low-level features (luminance, color and SIFT). For the image pair with very similar appearance like the first row in Fig. 9, color features are sufficient to build correspondence. However, for the pair with the larger differences, only deep features can match semantically similar objects, such as the houses in Fig. 9.

Local color transfer. To verify the robustness of our local color transfer method for consistent and faithful color effects, we first evaluate the effectiveness of our optimization in the progressive algorithm, and then apply our local color transfer as a post-processing step to compare with other color transfer approaches.

We run an ablation of the color transform by replacing it with the global color transform [Reinhard et al. 2001], and the local color transform adopted from [Reinhard et

了很多 artifacts。此外，NRDC无法匹配两个不同的场景，例如，在Fig. 13右半部分的图像对中没有找到匹配。Arbelot et al. [2017]开发了边缘感知描述符来匹配不同场景的相似文本内容，但它们在相似区域之间的局部颜色转移并不忠实于参考（例如，第三个例子中的天空颜色），并且提出的描述符无法检测到更高层次的语义信息。

我们在Fig. 14中比较了我们的方法与基于类比的局部颜色转移方法。这两个算法依赖于一对额外的例子（例如，参考1和参考2），这些例子对齐但颜色风格不同用于转移。相比之下，我们的方法直接从参考图像（例如，参考2）学习颜色。因此，我们的结果看起来比他们的更忠实于参考颜色。此外，我们的方法在实际应用中更加灵活，因为它不需要额外的对齐对进行转移。

在图Fig. 15中，我们与基于CNN特征的最近三种颜色转移方法进行了比较 [Liao et al. 2017; Luan et al. 2017; Mechrez et al. 2017]。Luan et al. [2017]和Mechrez et al. [2017]的方法匹配深度特征的全局统计数据（即格拉姆矩阵），并通过分割掩码保证区域到区域的转移。Luan et al. [2017]的结果中可见的一种显著的artifact是海报化，这在第6行自行车和第9行云的Fig. 15中可见。Mechrez et al. [2017]基于筛选泊松方程对风格化图像进行后处理，以约束风格化图像的梯度至原始源图像的梯度。这使得风格化结果更加逼真，但可能会产生不自然的颜色（例如，第4行和第9行的黄色天空）。独立于分割掩码，Liao et al. [2017]使用深度特征在两幅图像之间找到密集的对应关系，产生与我们更相似的结果。然而，与他们的方法不同，该方法分离了对应估计和颜色转移，我们的方法执行联合优化，能够更好地对齐两幅图像以进行颜色转移，并产生更少的鬼影artifact的结果。这在第2行和第3行的建筑物上表现得尤为明显。此外，他们的颜色转移仅替换低频颜色带；而我们的局部颜色转移考虑所有频率的颜色带。因此，我们的方法可以生成更忠实于参考的结果。例如，在第8行中，可以看到原始的绿色颜色，而我们的方法在保留参考的对比度和色度方面做得更好。

Table 1. 不同颜色转换方法的时间性能。

Method	Pitie et al. [2005]	Liao et al. [2017]	Luan et al. [2017]	Ours
Runtime (sec)	7	300	600	60

此外，我们的方法在将化妆效果或摄影师风格从一个肖像转移到另一个肖像时也非常有效。与专门针对肖像和特定类型的效果的方法相比，我们的方法可以在图Fig. 16中生成可比的结果，但无需在化妆前后对成对输入进行额外输入(Tong et al. [2007])、面部标志(Shih et al. [2014]和Tong et al. [2007])或脱色(Shih et al. [2014])。

感知研究。 我们进行一项感知研究，以评估我们在照片真实感和参考风格忠实度方面的颜色转移工作。我们在以下技术之间进行了比较：Pitie et al. [2005]，Luan et al. [2017]，Liao et al. [2017]和我们的技术。我们以随机顺序向参与者展示这四种方法的结果，并询问他们在1-4的评分标准上对图像进行评分，其中问题1从“绝对不是照片真实”到“绝对照片真实”，问题2从“绝对不忠实于参考”到“绝对忠实于参考”，问题3则考虑两个标准选择最佳选项。指标“照片真实”被定义为没有鬼影、没有光晕、没有不自然的颜色，而指标“忠实”则是定义为测量参考和结果之间在语义对应区域中的色度、亮度对比度之间的相似性。我们使用了30个不同的场景，每个方法都有30名参与者收集了他们的回应。例子是从 HaCohen et al. [2011]; Laffont et al. [2014]; Luan et al. [2017]的测试图像中随机选择的。

Fig. 17(a)(b)展示了每种方法的平均得分和标准偏差。对于照片真实性和忠实度，我们的方法和Liao et al. [2017]分别排在第1位 (3.09 ± 0.90 和 3.08 ± 0.91) 和第2位 (2.59 ± 0.97 和 2.54 ± 0.93)，随后是Pitie et al. [2005] (2.45 ± 1.09 和 2.27 ± 1.01) 和Luan et al. [2017] (1.93 ± 0.97 和 2.28 ± 1.05)。Luan et al. [2017]的方法在照片真实性方面表现最差，因为它经常产生过度曝光（卡通式）效果并引入不自然的颜色；而Pitie et al. [2005]在风格忠实度方面表现最差，因为全局转换限制了风格的 spatial variety。Fig. 17(c)(d)显示了每个参

al. 2001] which is also the initialization of our optimization, as shown in Fig. 10. The global color transform can only adjust the global tone but fails to reflect spatially varying effects. When the global method is used to match the means and variances of local patches, spatial color features are preserved but globally inconsistent artifacts appear (like the hair in the 2nd row and the ground in the 3rd row). Our color transfer enforces both local and global consistency and effectively avoids such artifacts as ghosting, halos, and inconsistencies.

Next we show how effective our local color transfer is when combined to the region-to-region correspondences obtained by the neural style transfer [Gatys et al. 2015] with a segmentation mask during the post-processing stage. The intermediate results with this correspondence method are shown as (a) of each example in Fig. 11. Beside ours, there are several methods proposed to transfer colors based on these intermediate results. Luan et al. [2017] constrain the color transformation to be a locally affine in color space, shown as (b). Mechrez et al. [2017] use the Screened Poisson Equation to improve photorealism, shown as (c). Liao et al. [2017] replace only the low-frequency color bands of the source with those of the aligned reference, shown as (d). Compared to the above, our method is better at preserving fine details and image boundaries as (e) in the 1st example, and is more faithful to the reference with similar chrominance and contrast and does not introduce any new colors as do Mechrez et al. [2017] (2nd example). However, serving as a post-processing step, our color transfer algorithm still can not repair large correspondence errors (e.g., the black windows in the 2nd example and the red trees in 3rd example). That also shows the necessity of our joint optimization scheme.

4.3 Single-Reference Color Transfer

To validate our approach on color transfer, we first discuss visual comparisons with previous works in traditional and deep color transfer, and then report the statistics of our conducted perceptual study.

In Fig. 12, we compare our method with the traditional global color transfer methods. Reinhard et al. [2001] and Pitie et al. [2005] only match the global color statistics between the source image and the reference image for color transfer, thus limiting their ability to conduct more sophisticated color transformations. For example, in the 2nd result, the house is rendered in black matching the color of the sky. In contrast, our transfer is local and capable of handling semantic object-to-object color transfer.

Next, we compare our method with the traditional color transfer methods based on local correspondence [Arbelot et al. 2017; HaCohen et al. 2011] in Fig. 13. The NRDC method [HaCohen et al. 2011] is based on a small number of reliable matches to estimate the global color mapping function, so it achieves a more spatially varying result. NRDC are suitable for the image pair of the common scenes (e.g., the left half in Fig. 13). In such scenes, our method builds much denser correspondence and applies local instead of global transformation, so our color transfer produces more accurate results with fewer artifacts than NRDC in local regions, like the children who are absent in the reference in the 2nd example and the bag in the 3rd example. Moreover, NRDC fails to match two different scenes, for example, there was no matching found in all the image pairs on the right half in Fig. 13. Arbelot et al. [2017] develop edge-aware descriptors to match similar textual content from different scenes but their local color transfer between similar regions is not faithful to the reference (e.g., the sky color in the 3rd example) and the proposed descriptors are unable to detect higher-level semantic information.

与者的平均得分。在我们的方法在所有参与者中在照片真实性和忠实度方面始终优于其他方法。我们进一步对收集的数据进行了重复测量ANOVA，结果显示从这两个方面来看，这四种方法之间的差异都是显著的($p < 0.005$)。我们还在两者之间使用简单对比来比较每种方法相对于我们的方法在所有900个评分中(30名参与者乘以30个场景)。对于照片真实性，参与者更喜欢我们的方法相对于Pitie et al. [2005] (56.22%更好, 14.89%相同), Luan et al. [2017] (75.89%更好, 10.78%相同)和Liao et al. [2017] (56.22%更好, 19.00%相同)。对于忠实度，参与者更喜欢我们的方法相对于Pitie et al. [2005] (62.11%更好, 14.11%相同), Luan et al. [2017] (67.11%更好, 9.11%相同)和Liao et al. [2017] (58.33%更好, 19.00%相同)。

由于颜色转移质量取决于照片真实性和风格忠实度，我们检查了考虑这两个标准时用户的偏好结果。图Fig. 17(e)的饼图显示了每种方法被选为最佳的百分比。我们的算法在其他三种方法中总体上被选为最佳的次数高达48.11%。图Fig. 17 (f) 给出了每种方法在每个场景中被选为最佳次数的详细数字。这显示了更多的用户更喜欢我们的方法相对于Pitie et al. [2005] (23对7), Luan et al. [2017] (25对5), 和Liao et al. [2017] (22对5, 3个相同)。

此外，我们的方法比其他两种基于深度网络的方法更高效。四种竞争算法的运行时在Table 1中显示。

4.4 Multi-Reference Color Transfer

在所有单参考颜色转移方法中，选择参考至关重要，以实现令人满意的结果。我们的方法在参考中找不到适当颜色指导的区域无法正确转移颜色，导致出现不自然的视觉效果。这个问题可以通过引入多个参考来解决。例如，在 Fig. 18的底部一行中，天空颜色不正确，因为单一参考不包含天空，但使用多个参考时则正确，其中一些参考包含天空。我们的多参考方法允许用户提供关键词来控制颜色转移，例如，19的第一行中的“餐厅夜晚”。“餐厅”描述了源图像的内容，而“夜晚”定义了所需的风格。为了自动获取用于转移

的多参考，这些关键词通过搜索引擎（例如，Google图片）进行检索。我们收集前50个搜索结果作为候选。自然地，这些图像可能具有多种颜色风格以及异常点。对于每个候选图像，我们在HSV颜色空间中计算其归一化直方图，并选择与所有其他图像距离最小的最代表性的。然后我们选择最接近的5张图像子集作为转移的参考。

与Autostyle [Liu et al. 2014]相比，后者使用全局方法将渐晕、颜色和局部对比度从通过特定关键词搜索的图像集合中转移到一个特定的风格（例如，“夜晚”，“海滩”，“日落”），我们的方法基于语义对应关系进行局部转移，因此需要更多的关键词来描述内容和风格（例如，“餐厅夜晚”，“建筑海滩”，“建筑河流日落”）。然而，它允许我们比Autostyle [Liu et al. 2014]产生更精确的颜色转移。Cycle-GAN [Zhu et al. 2017]也允许在没有选择参考的情况下将特定风格（例如，“冬季约塞米蒂”）转移到给定的源图像，通过在大型数据集上训练网络。与他们的方法相比，我们的方法在测试不同风格时不需重新训练时更加灵活。我们的结果在Fig. 19的底部两行中显示的棋盘格artifacts更少。

4.5 Colorization

我们还可以用我们的方法对灰度图像进行上色。我们只需提供所需风格的色彩参考，以便给输入的灰度图像上色。Fig. 20展示了部分上色结果。

5 CONCLUDING REMARKS

在本论文中，我们展示了一种新的算法，用于在语义相关图像之间本地传输颜色。它不仅处理单参考传输，还可以适应多参考传输，这避免了寻找适当参考的经常困难的任务。我们采用了一种联合优化方法，该方法在CNN层之间以层次化的方式搜索NNF和局部颜色传输。我们已经证明，我们的方法在现实世界图像中的各种传输场景中广泛适用。

然而，我们的方法仍然存在局限性。它可能不匹配源图像中存在但参考图像中不存在的某些区域，从而导致错误的颜色传输，如Fig. 21左边的示例中的背

We compare our method with the analogy-based local color transfer methods [Laffont et al. 2014; Shih et al. 2013] in Fig. 14. These two algorithms depend on an additional pair of examples (e.g., Reference 1 and Reference 2) which are aligned but have different color styles for the transfer. In contrast, ours learns the color directly from the reference image (e.g., Reference 2). Therefore, our results look more faithful to the reference colors than theirs. Moreover, our method is more flexible in practice since it does not require an additional aligned pair for the transfer.

In Fig. 15, we compare with three recent color transfer methods based on CNN features [Liao et al. 2017; Luan et al. 2017; Mechrez et al. 2017]. The methods of Luan et al. [2017] and Mechrez et al. [2017] match the global statistics of deep features (*i.e.*, Gram Matrix) and guarantee region-to-region transfer via segmentation masks. One type of noticeable artifact in the results of Luan et al. [2017] is posterization, which is visible in the bicycle in the 6th row and the cloud in the 9th row of Fig. 15. Mechrez et al. [2017] post-process the stylized image based on the Screened Poisson Equation to constrain the gradients to those of the original source image. They make the stylized results more photorealistic but may generate unnatural colors (e.g., the yellow sky in the 4th and 9th rows). Independent of segmentation masks, Liao et al. [2017] find dense correspondence between two images using deep features, yielding the results that are more similar to ours. However, unlike their approach which separates correspondence estimation and color transfer, our method performs joint optimization, which can align two images better for the color transfer and generate results with fewer ghosting artifacts. This is clearly shown from the buildings in the 2nd and 3rd rows. Moreover, their color transfer only replaces low-frequency color bands; while our local color transfer considers all-frequency color bands. Thus, ours can generate results more faithful to the reference. For

Table 1. Runtime of different color transfer methods.

Method	Pitie et al. [2005]	Liao et al. [2017]	Luan et al. [2017]	Ours
Runtime (sec)	7	300	600	60

example, in the 8th row, the original green color can be observed in the result while ours is better at preserving the contrast and the chrominance of the reference.

Moreover, our method is effective in transferring effects like makeup or photographer styles from one portrait to another. Compared to the methods specifically focusing on portraits and very specific kinds of effects, ours can generate comparable results as shown in Fig. 16, but without requiring extra inputs of the pair before and after makeup (Tong et al. [2007]), face landmarks (Shih et al. [2014] and Tong et al. [2007]) or matting (Shih et al. [2014]).

Perceptual Study. We conduct a perceptual study to evaluate our color transfer work in terms of photorealism and faithfulness to reference style. We compared the following techniques: Pitie et al. [2005], Luan et al. [2017], Liao et al. [2017] and ours in the study. We present the results of the four methods to participants in a random order and ask them to score images in a 1-to-4 scale from “definitely not photorealistic” to “definitely photorealistic” for question 1, from “definitely not faithful to the reference” to “definitely faithful to the reference” for question 2, and select the best one considering both criteria for question 3. The metric “photorealism” is defined as no ghosting, no halo, and no unnatural colors, while the metric “faithfulness” is defined to measure the similarity in chrominance, luminance and contrast between semantically corresponding regions in the reference and the result. We use 30 different scenes for each of the four methods and collect the responses from 30 participants. The examples are randomly selected from the test images

景。这通常发生在单参考传输中，可以通过引入更多的参考来减少这种情况。我们依赖的VGG网络没有被训练用来区分具有相同语义标签的不同实例，因此它可能导致不同实例之间的颜色混合，如 Fig. 21右边的示例中的男士衬衫，它被从参考中的两个人混合的蓝色和白色颜色所转移。可能的改进可能是在带有实例标签的数据集上训练一个网络。此外，直接将我们的图像颜色传输方法应用于视频可能会导致一些闪烁 artifacts [Chen et al. 2017a]。解决这个问题需要时间约束，并将作为未来的工作考虑。我们的颜色传输适用于语义相似的图像。然而，对于没有语义关系的图像对，它可能会导致一些不自然的颜色效果，如第三失败示例中的黄色水。对于未来的工作，我们希望探索允许这个示例自动退化到良好颜色传输结果的方法。

REFERENCES

- Xiaobo An and Fabio Pellacini. 2008. AppProp: all-pairs appearance-space edit propagation. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 40.
- Xiaobo An and Fabio Pellacini. 2010. User-Controllable Color Transfer. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 263–271.
- Benoit Arbelot, Romain Vergne, Thomas Hurtut, and Joëlle Thollot. 2017. Local texture-based color transfer and colorization. *Computers & Graphics* 62 (2017), 15–27.
- Soonmin Bae, Sylvain Paris, and Frédéric Durand. 2006. Two-scale Tone Management for Photographic Look. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 637–645.
- Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan Goldman. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 24.
- Connelly Barnes, Eli Shechtman, Dan Goldman, and Adam Finkelstein. 2010. The generalized patchmatch correspondence algorithm. *European Conference on Computer Vision (ECCV)*, 2010 (2010), 29–43.
- Nicolas Bonneau, Gabriel Peyré, and Marco Cuturi. 2016. Wasserstein barycentric coordinates: histogram regression using optimal transport. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 71–1.
- Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. 2017a. Coherent online video style transfer. In *Proc. Intl. Conf. Computer Vision (ICCV)*.
- Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2017b. Stylebank: An explicit representation for neural image style transfer. In *Proc. CVPR*, Vol. 1, 4.
- Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2018b. Stereoscopic neural style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 10.
- Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2018a. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4 (2018), 834–848.
- Xiaowu Chen, Dongqing Zou, Steven Zhiying Zhou, Qinping Zhao, and Ping Tan. 2013. Image matting with local and nonlocal smooth priors. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. 1902–1907.
- Kevin Dale, Micah K Johnson, Kalyan Sunkavalli, Wojciech Matusik, and Hanspeter Pfister. 2009. Image restoration using online photo collections. In *2009 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2217–2224.
- Yuki Endo, Satoshi Iizuka, Yoshihiro Kanamori, and Jun Mitani. 2016. DeepProp: extracting deep features from a single image for edit propagation. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 189–201.
- Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, Vol. 27. 67.
- Daniel Freedman and Pavel Kisilev. 2010. Object-to-object color transfer: Optimal flows and smsp transformations. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 287–294.
- Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *CoRR* abs/1508.06576 (2015).
- Yoav HaCohen, Eli Shechtman, Dan B Goldman, and Dani Lischinski. 2011. Non-rigid dense correspondence with applications for image enhancement. *ACM transactions on graphics (TOG)* 30, 4 (2011), 70.
- Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan. 2018. Deep Exemplar-based Colorization. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 110.
- Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. 2001. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 327–340.
- Hristina Hristova, Olivier Le Meur, Rémi Cozot, and Kadi Bouatouch. 2015. Style-aware robust color transfer. In *Proceedings of the workshop on Computational Aesthetics*. Eurographics Association, 67–77.
- Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2016. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 110.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. (2017), 5967–5976.
- Asad Khan, Luo Jiang, Wei Li, and Ligang Liu. 2017. Fast color transfer from multiple images. *Applied Mathematics-A Journal of Chinese Universities* 32, 2 (2017), 183–200.
- Vladimir Kolmogorov and Ramin Zabin. 2004. What energy functions can be minimized via graph cuts? *IEEE transactions on pattern analysis and machine intelligence* 26, 2 (2004), 147–159.
- Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 149.
- Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In So Kweon. 2016. Automatic content-aware color and tone stylization. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*. 2470–2478.
- Anat Levin, Dani Lischinski, and Yair Weiss. 2004. Colorization using optimization. In *ACM transactions on graphics (TOG)*, Vol. 23. ACM, 689–694.

of HaCohen et al. [2011]; Laffont et al. [2014]; Luan et al. [2017].

Fig. 17(a)(b) demonstrate the average scores and standard deviations of each method. For photorealism and faithfulness, ours and Liao et al. [2017] are ranked 1st (3.09 ± 0.90 and 3.08 ± 0.91) and 2nd (2.59 ± 0.97 and 2.54 ± 0.93) respectively, followed by Pitie et al. [2005] (2.45 ± 1.09 and 2.27 ± 1.01) and Luan et al. [2017] (1.93 ± 0.97 and 2.28 ± 1.05). The method by Luan et al. [2017] performs worst in photorealism, since it often produces posterization (cartoon-like) effects and introduces unnatural colors; while Pitie et al. [2005] perform the worst in faithfulness to style, since global transfer limits the spatial variety of styles. Fig. 17(c)(d) show the average scores given by every participant. Ours is consistently better than others in both photorealism and faithfulness among all participants. We further conduct repeated-measures ANOVAs on the collected data, and it shows the differences between the four methods from these two aspects are all significant ($p < 0.005$). We also use simple contrasts to compare each method against our method among all 900 scores in both photorealism and faithfulness (30 participants by 30 scenes). For photorealism, participants prefer ours over Pitie et al. [2005] (56.22% better, 14.89% equal), Luan et al. [2017] (75.89% better, 10.78% equal) and Liao et al. [2017] (56.22% better, 19.00% equal). For faithfulness, participants prefer ours over Pitie et al. [2005] (62.11% better, 14.11% equal), Luan et al. [2017] (67.11% better, 9.11% equal) and Liao et al. [2017] (58.33% better, 19.00% equal).

Since color transfer quality depends on both photorealism and style fidelity, we examine the subject’s preferred result considering both criteria. The pie chart on Fig. 17(e) shows the percentage of each method selected as the best. Our algorithm is the top overall selection over the other three methods at 48.11% of the time. Fig. 17(f) gives the detailed numbers of how many times each method is selected as the best in each scene. It shows that more users

prefer ours over Pitie et al. [2005] (23 vs. 7), Luan et al. [2017] (25 vs. 5), and Liao et al. [2017] (22 vs. 5 with 3 equal). Moreover, our method is more efficient than the other two deep network-based methods. The runtime of the four competing algorithms is shown in Table 1.

4.4 Multi-Reference Color Transfer

In all single-reference color transfer methods, reference selection is crucial to achieving satisfactory results. Our method fails to transfer correct colors and yields an unnatural appearance in the regions where no proper color guidance can be found in the reference. This problem can be addressed by introducing multiple references. For example, in the bottom row of Fig. 18, the sky color is incorrect because the single reference does not contain the sky, but it is correct with multiple references, some of which contains the sky. Our multi-reference approach allows the user to provide keywords for controlling the color transfer, for example, “restaurant night” in the first row of 19. “Restaurant” describes the content in the source images, and “night” defines the desired style. To automatically obtain multiple references for the transfer, these keywords are used to retrieve images through a search engine (e.g., Google Images). We collect the top 50 search results as candidates. Naturally, these images may have a wide variety of color styles as well as outliers. To select a set of images with consistent colors, for each candidate image, we compute its normalized histogram in the HSV color space and select the most representative one with a minimum L_2 histogram distance to all others. Then we choose the closest subset of 5 images to the representative one as the references for transfer.

Compared with Autostyle [Liu et al. 2014] which uses a global method to transfer vignetting, color, and local contrast from a collection of images searched by a particular keyword for one specific style (e.g., “night”,

- Anat Levin, Dani Lischinski, and Yair Weiss. 2008. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 2 (2008), 228–242.
- Chuan Li and Michael Wand. 2016. Combining markov random fields and convolutional neural networks for image synthesis. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*. 2479–2486.
- Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. 2017. Visual Attribute Transfer Through Deep Image Analogy. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 120:1–120:15.
- Dani Lischinski, Zeev Farbman, Matt Uyttendaele, and Richard Szeliski. 2006. Interactive local adjustment of tonal values. In *ACM Transactions on Graphics (TOG)*, Vol. 25. ACM, 646–653.
- Yiming Liu, Michael Cohen, Matt Uyttendaele, and Szymon Rusinkiewicz. 2014. Autostyle: Automatic style transfer from image collections to users’ images. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 21–31.
- Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2017. Deep Photo Style Transfer. (2017), 6997–7005.
- Roey Mechrez, Eli Shechtman, and Lihai Zelnik-Manor. 2017. Photorealistic Style Transfer with Screened Poisson Equation. *CoRR* abs/1709.09828 (2017).
- Augustus Odena, Vincent Dumoulin, and Chris Olah. 2016. Deconvolution and Checkerboard Artifacts. *Distill* (2016). <http://distill.pub/2016/deconv-checkerboard>
- Francois Fleuret, Anil C Kokaram, and Rozenn Dahyot. 2005. N-dimensional probability density function transfer and its application to color transfer. In *2005 IEEE International Conference on Computer Vision (ICCV)*, Vol. 2. IEEE, 1434–1439.
- Erik Reinhard, Michael Adikhmin, Bruce Gooch, and Peter Shirley. 2001. Color transfer between images. *IEEE Computer graphics and applications* 21, 5 (2001), 34–41.
- Sam T Roweis and Lawrence K Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *science* 290, 5500 (2000), 2323–2326.
- Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.
- Xiaoyong Shen, Xin Tao, Chao Zhou, Hongyun Gao, and Jiaya Jia. 2016. Regional foremost matching for internet scene images. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 178.
- YiChang Shih, Sylvain Paris, Connelly Barnes, William T. Freeman, and Frédo Durand. 2014. Style Transfer for Headshot Portraits. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 148:1–148:14.
- YiChang Shih, Sylvain Paris, Frédo Durand, and William T Freeman. 2013. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics (TOG)* 32, 6 (2013).
- Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani. 2008. Summarizing visual data using bidirectional similarity. In *Computer Vision and Pattern Recognition (CVPR), 2008 IEEE Conference on*. IEEE, 1–8.
- Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *CoRR* abs/1409.1556 (2014).
- Kalyan Sunkavalli, Micah K. Johnson, Wojciech Matusik, and Hanspeter Pfister. 2010. Multi-scale Image Harmonization. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 125:1–125:10.
- Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang. 2005. Local color transfer via probabilistic segmentation by expectation-maximization. In *Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on*, Vol. 1. IEEE, 747–754.
- Wai-Shun Tong, Chi-Keung Tang, Michael S Brown, and Ying-Qing Xu. 2007. Example-based cosmetic transfer. In *Computer Graphics and Applications, 2007. PG’07. 15th Pacific Conference on*. IEEE, 211–218.
- Tomohisa Welsh, Michael Ashikhmin, and Klaus Mueller. 2002. Transferring color to greyscale images. In *ACM Transactions on Graphics (TOG)*, Vol. 21. ACM, 277–280.
- Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. 2016. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)* 35, 2 (2016), 11.
- Jae-Doug Yoo, Min-Ki Park, Ji-Ho Cho, and Kwan H Lee. 2013. Local color transfer between images using dominant colors. *Journal of Electronic Imaging* 22, 3 (2013), 033003–033003.
- Richard Zhang, Phillip Isola, and Alexei A Efros. 2016. Colorful image colorization. In *European Conference on Computer Vision (ECCV), 2016*. Springer, 649–666.
- Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin. 2012. A closed-form solution to retinex with nonlocal texture constraints. *IEEE transactions on pattern analysis and machine intelligence* 34, 7 (2012), 1437–1444.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. (2017), 2242–2251.

“beach”, “sunset”), our approach performs a local transfer based on semantic correspondence and thus requires more keywords in order to describe both content and style (e.g., “restaurant night”, “building beach”, “building river sunset”). However, it allows us to produce a more precise color transfer than Autostyle [Liu et al. 2014]. Cycle-GAN [Zhu et al. 2017] also allows the transfer of a specific style (e.g., “winter Yosemite”) to a given source without selecting the reference by leveraging network training on large datasets. Compared to theirs, our method is more flexible when testing different styles without retraining. Our results have fewer checkerboard artifacts, as shown on the bottom two rows of Fig. 19.

4.5 Colorization

We can also use our method for colorization of gray-scale images. We simply provide color references in the desired style in order to colorize our input gray-scale image. Fig. 20 shows some colorization results.

5 CONCLUDING REMARKS

In this paper, we demonstrate a new algorithm for locally transferring colors across semantically-related images. It not only handles single-reference transfer, but can also be adapted to multi-reference transfer, which avoids the often difficult task of finding a proper reference. We adapt a joint optimization of NNF search and local color transfer across CNN layers in a hierarchical manner. We have shown that our approach is widely applicable to a variety of transfer scenarios in real-world images.

However our method still has limitations. It may mismatch some regions which exist in the source but not in the reference, and thus cause incorrect color transfer, such as the background on the left example of Fig. 21. This often happens in the single-reference transfer and can be reduced by introducing more references. The VGG

network we relied on is not trained to distinguish different instances with the same semantic labels, so it may cause color mixing between different instances, such as the man’s shirt on the right example of Fig. 21, which is transferred with mixed blue and white colors from two persons in the reference. A possible improvement would be to train a network on a dataset with instance labels. Moreover, directly applying our image color transfer method to video may cause some flickering artifacts [Chen et al. 2017a]. Addressing this would require temporal constraints, and will be considered for future work. Our color transfer is suitable for semantically similar images. However, it may lead to some unnatural color effects for image pairs without a semantic relationship, such as the yellow water in the 3rd failure example. For future work, we would like to explore methods that would allow this example to automatically degenerate to a good color transfer result.

REFERENCES

- Xiaobo An and Fabio Pellacini. 2008. AppProp: all-pairs appearance-space edit propagation. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 40.
- Xiaobo An and Fabio Pellacini. 2010. User-Controllable Color Transfer. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 263–271.
- Benoit Arbelot, Romain Vergne, Thomas Hurtut, and Joëlle Thollot. 2017. Local texture-based color transfer and colorization. *Computers & Graphics* 62 (2017), 15–27.
- Soonmin Bae, Sylvain Paris, and Frédo Durand. 2006. Two-scale Tone Management for Photographic Look. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 637–645.
- Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan Goldman. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 24.
- Connelly Barnes, Eli Shechtman, Dan Goldman, and Adam Finkelstein. 2010. The generalized patchmatch correspondence algorithm. *European Conference on Computer Vision (ECCV)*, 2010 (2010), 29–43.
- Nicolas Bonneau, Gabriel Peyré, and Marco Cuturi. 2016. Wasserstein barycentric coordinates: histogram regression using optimal transport. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 71–1.
- Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. 2017a. Coherent online video style transfer. In *Proc. Intl. Conf. Computer Vision (ICCV)*.
- Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2017b. Stylebank: An explicit representation for neural image style transfer. In *Proc. CVPR*, Vol. 1, 4.



Fig. 8. 不同方法和联合方法的结果比较。每个例子包括四行,每行显示通过一种方法计算的中间NNFs以及使用我们的局部颜色转移方法得到的最终颜色转移结果。(a)表示颜色图像对之间的NNF估计,(b)显示灰度图像对之间的NNF,(c)表示 Liao et al. [2017]的方法得到的NNF,(d)显示通过我们的联合优化得到的NNF。输入图像: Luan et al. [2017]。

- Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. 2018b. Stereoscopic neural style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 10.
- Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2018a. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 4 (2018), 834–848.
- Xiaowu Chen, Dongqing Zou, Steven Zhiying Zhou, Qinping Zhao, and Ping Tan. 2013. Image matting with local and nonlocal smooth priors. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. 1902–1907.
- Kevin Dale, Micah K Johnson, Kalyan Sunkavalli, Wojciech Matusik, and Hanspeter Pfister. 2009. Image restoration using online photo collections. In *2009 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2217–2224.
- Yuki Endo, Satoshi Iizuka, Yoshihiro Kanamori, and Jun Mitani. 2016. DeepProp: extracting deep features from a single image for edit propagation. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 189–201.
- Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, Vol. 27. 67.
- Daniel Freedman and Pavel Kisilev. 2010. Object-to-object color transfer: Optimal flows and smsp transformations. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 287–294.
- Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *CoRR* abs/1508.06576 (2015).
- Yoav HaCohen, Eli Shechtman, Dan B Goldman, and Dani Lischinski. 2011. Non-rigid dense correspondence with applications for image enhancement. *ACM transactions on graphics (TOG)* 30, 4 (2011), 70.
- Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan. 2018. Deep Exemplar-based Colorization. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 110.
- Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. 2001. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 327–340.
- Hristina Hristova, Olivier Le Meur, Rémi Cozot, and Kadi Bouatouch. 2015. Style-aware robust color transfer. In *Proceedings of the workshop on Computational Aesthetics*. Eurographics Association, 67–77.
- Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2016. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 110.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. (2017), 5967–5976.
- Asad Khan, Luo Jiang, Wei Li, and Ligang Liu. 2017. Fast color transfer from multiple images. *Applied Mathematics-A Journal of Chinese Universities* 32, 2 (2017), 183–200.
- Vladimir Kolmogorov and Ramin Zabin. 2004. What energy functions can be minimized via graph cuts? *IEEE transactions on pattern analysis and machine intelligence* 26, 2 (2004), 147–159.
- Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 149.



Fig. 9. 不同特征估计对应结果的比较。输入图像:Anonymous/tajawal.ae和Anonymous/winduprocketapps.com。

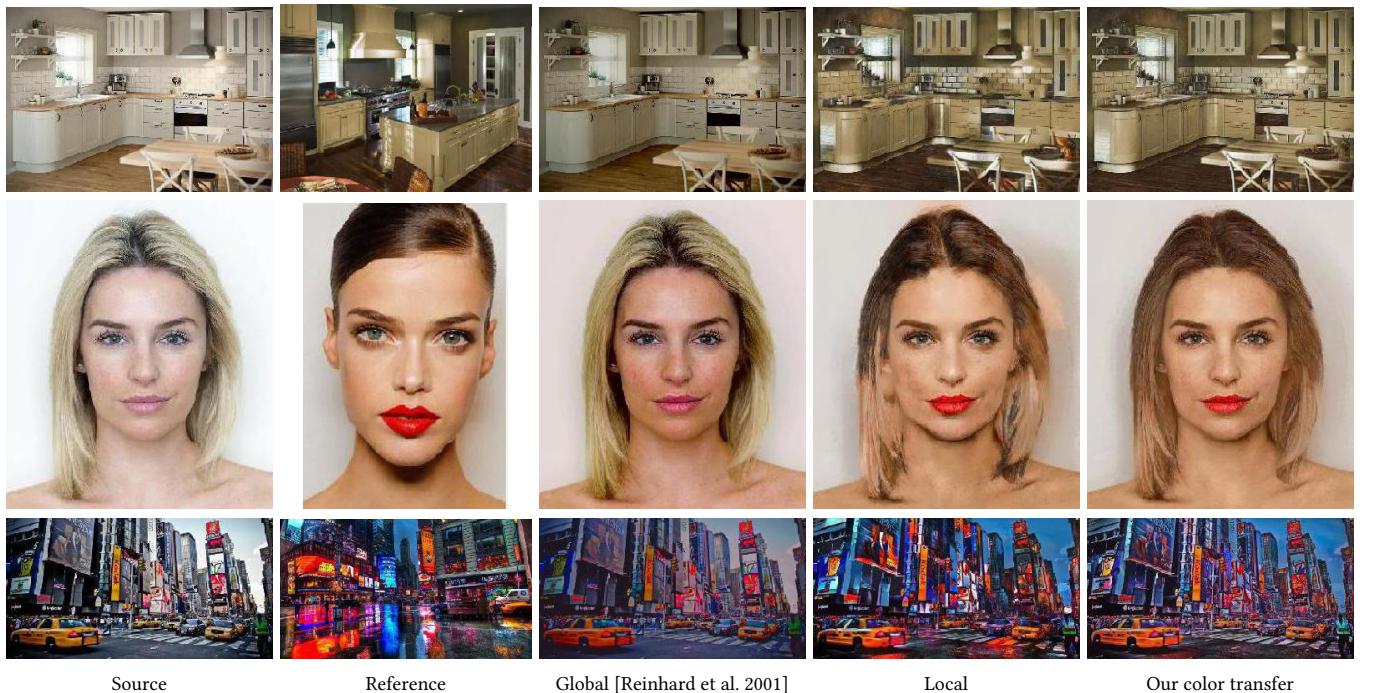


Fig. 10. 不同颜色转换方法之间在将对齐参考的颜色转换到源图像上的结果比较。输入图像:Anonymous/diy.com和 Luan et al. [2017]。

- Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *CorR* abs/1409.1556 (2014).
- Kalyan Sunkavalli, Micah K. Johnson, Wojciech Matusik, and Hanspeter Pfister. 2010. Multi-scale Image Harmonization. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 125:1–125:10.
- Yu-Wing Tai, Jiaya Jia, and Chi-Keung Tang. 2005. Local color transfer via probabilistic segmentation by expectation-maximization. In *Computer Vision and Pattern Recognition (CVPR), 2005 IEEE Conference on*, Vol. 1. IEEE, 747–754.
- Wai-Shun Tong, Chi-Keung Tang, Michael S Brown, and Ying-Qing Xu. 2007. Example-based cosmetic transfer. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*. IEEE, 211–218.
- Tomohisa Welsh, Michael Ashikhmin, and Klaus Mueller. 2002. Transferring color to greyscale images. In *ACM Transactions on Graphics (TOG)*, Vol. 21. ACM, 277–280.
- Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. 2016. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)* 35, 2 (2016), 11.
- Jae-Doug Yoo, Min-Ki Park, Ji-Ho Cho, and Kwan H Lee. 2013. Local color transfer between images using dominant colors. *Journal of Electronic Imaging* 22, 3 (2013), 033003–033003.
- Richard Zhang, Phillip Isola, and Alexei A Efros. 2016. Colorful image colorization. In *European Conference on Computer Vision (ECCV), 2016*. Springer, 649–666.
- Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin. 2012. A closed-form solution to retinex with nonlocal texture constraints. *IEEE transactions on pattern analysis and machine intelligence* 34, 7 (2012), 1437–1444.
- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. (2017), 2242–2251.

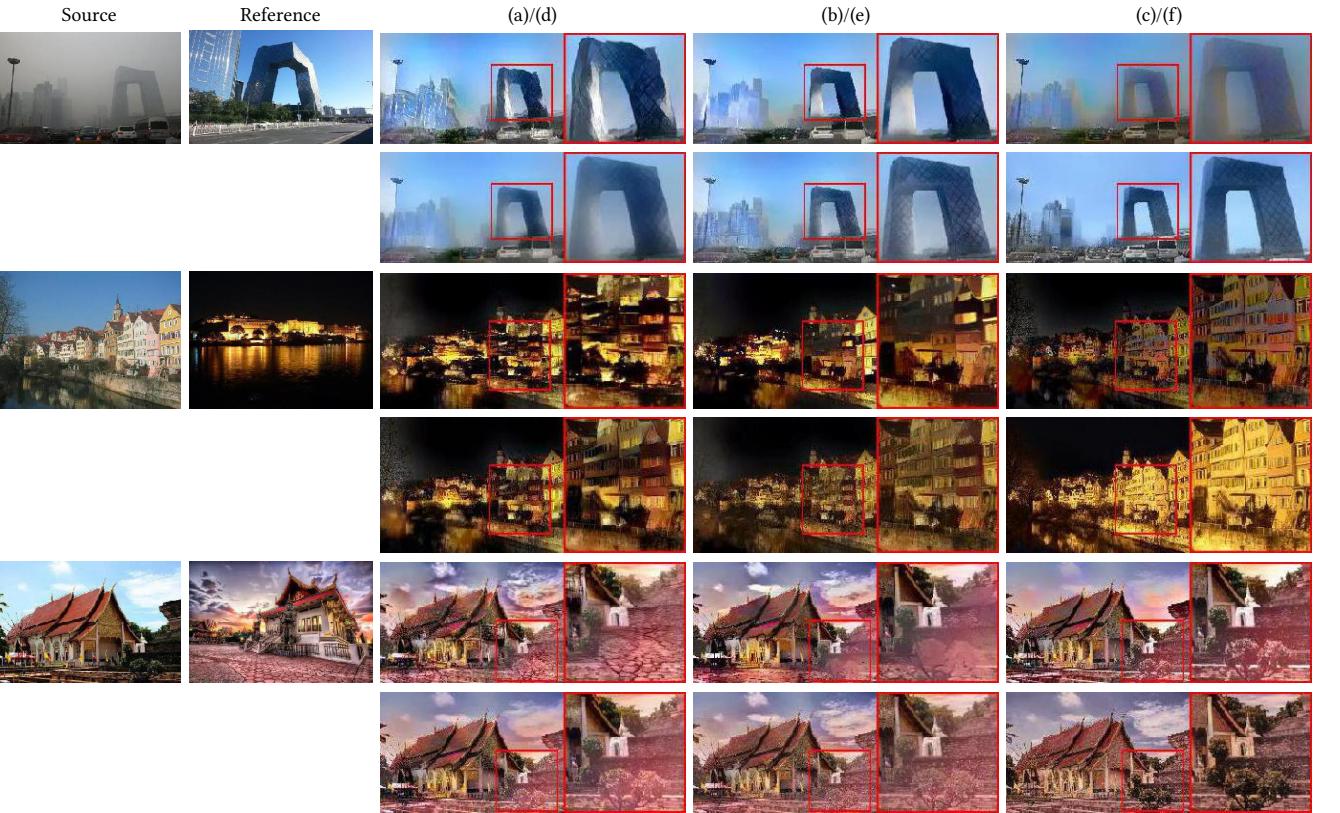


Fig. 11. 不同颜色转换方法从对齐参考到源图像的结果比较。(a)代表 Gatys et al. [2015]使用分割掩码对齐的参考图像, (b)–(e)分别是 Luan et al. [2017], Mechrez et al. [2017], Liao et al. [2017]和我们的局部颜色转换方法精修后的结果, 以及(f)是我们联合优化方法的成果。输入图像: Luan et al. [2017]。



Fig. 8. Comparisons of results by separate methods and our joint method. Every example includes four rows with each showing the intermediate NNFs computed by one method and the final color transfer result using our local color transfer. (a) represents the NNF estimation of the color image pair, (b) shows NNF on the grayscale image pair, (c) represents NNF by Liao et al. [2017] and (d) shows NNF by our joint optimization. Input images: Luan et al. [2017].

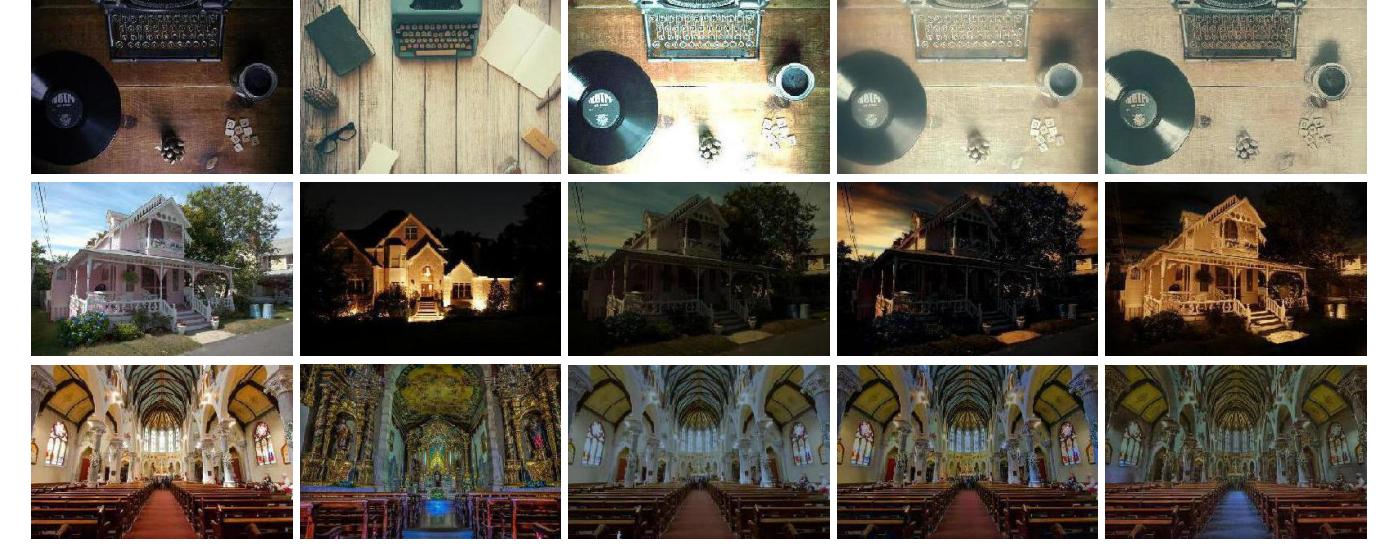


Fig. 12. 与全局颜色转移方法 [Reinhard et al. 2001] 和 [Pitie et al. 2005] 的比较。测试图像来自 Luan et al. [2017]。输入图像: Luan et al. [2017]。

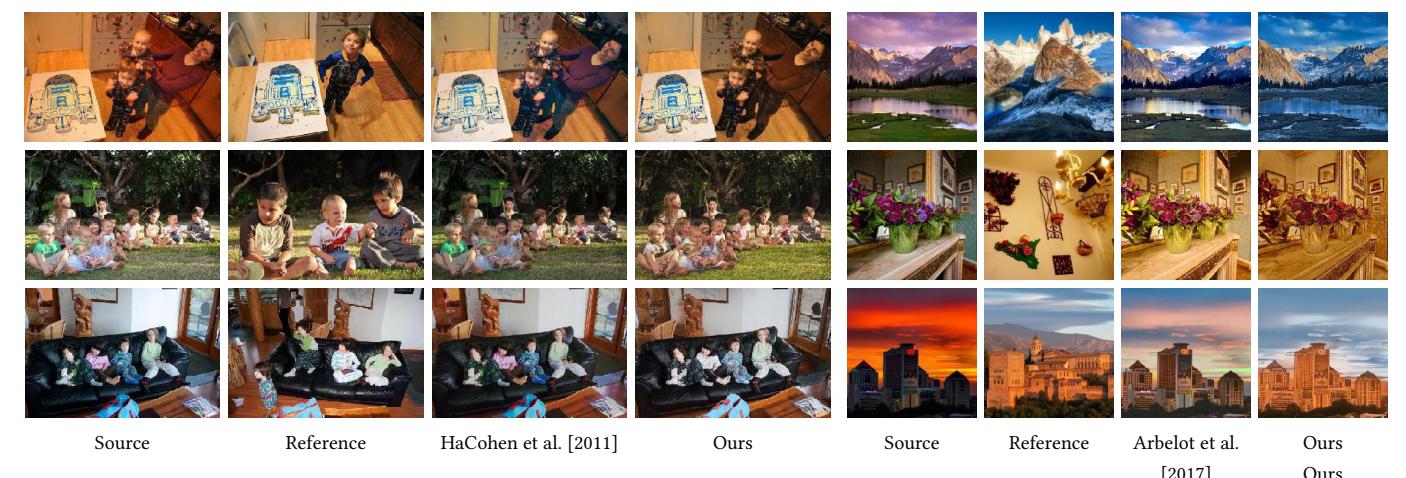


Fig. 13. 与传统局部颜色转移方法相比。测试图像来自 HaCohen et al. [2011] (左) 和 Arbelot et al. [2017] (右)。输入图像: HaCohen et al. [2011] 和 Arbelot et al. [2017]。



Fig. 9. Comparisons between results by estimating correspondence with different features. Input images: Anonymous/tajawal.ae and Anonymous/winduprocketapps.com.

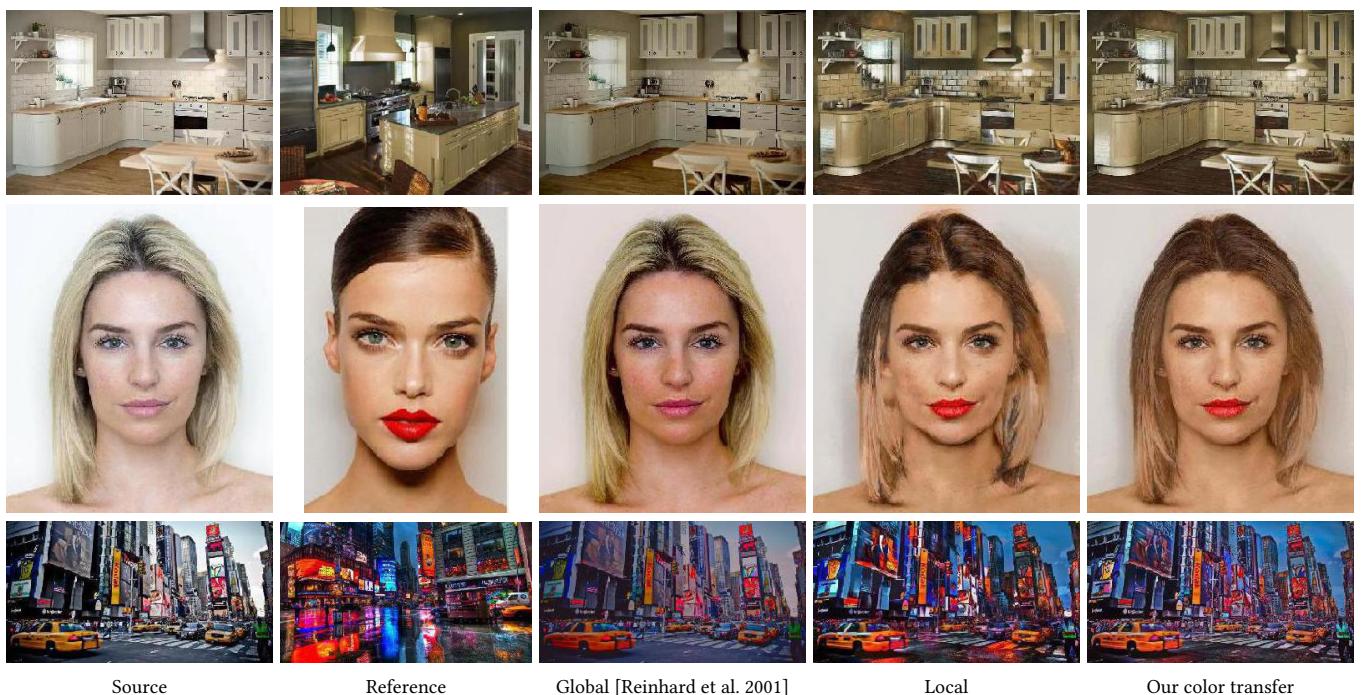


Fig. 10. Comparisons between results by different color transfer methods applied to transfer the color of the aligned reference to the source. Input images: Anonymous/diy.com and Luan et al. [2017].

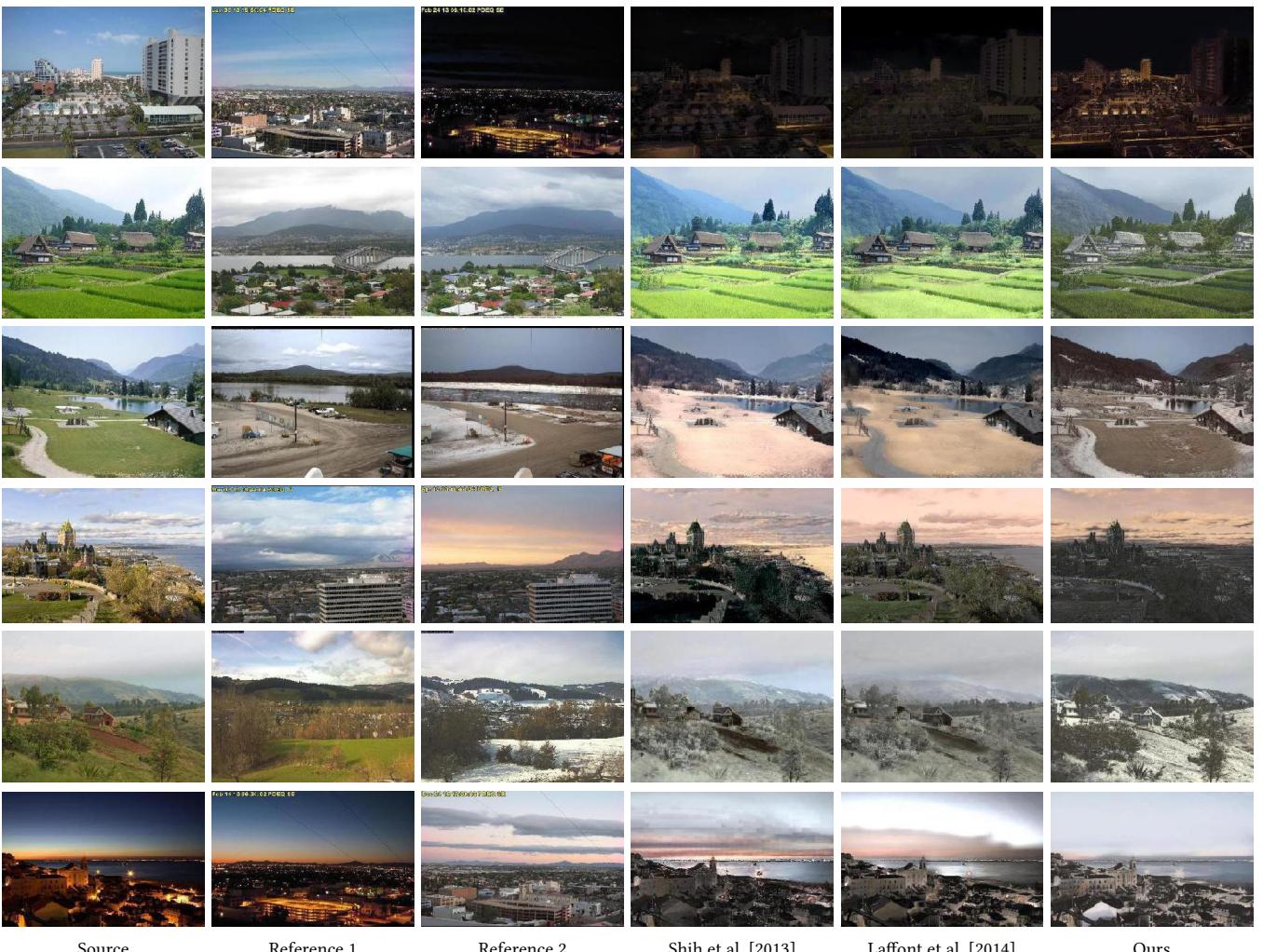


Fig. 14. 与基于类比的方法在 Laffont et al. [2014]数据上的比较。Shih et al. [2013]和 Laffont et al. [2014]同时使用了参考1和参考2，而我们的方法只使用了参考2。输入图像: Laffont et al. [2014]。

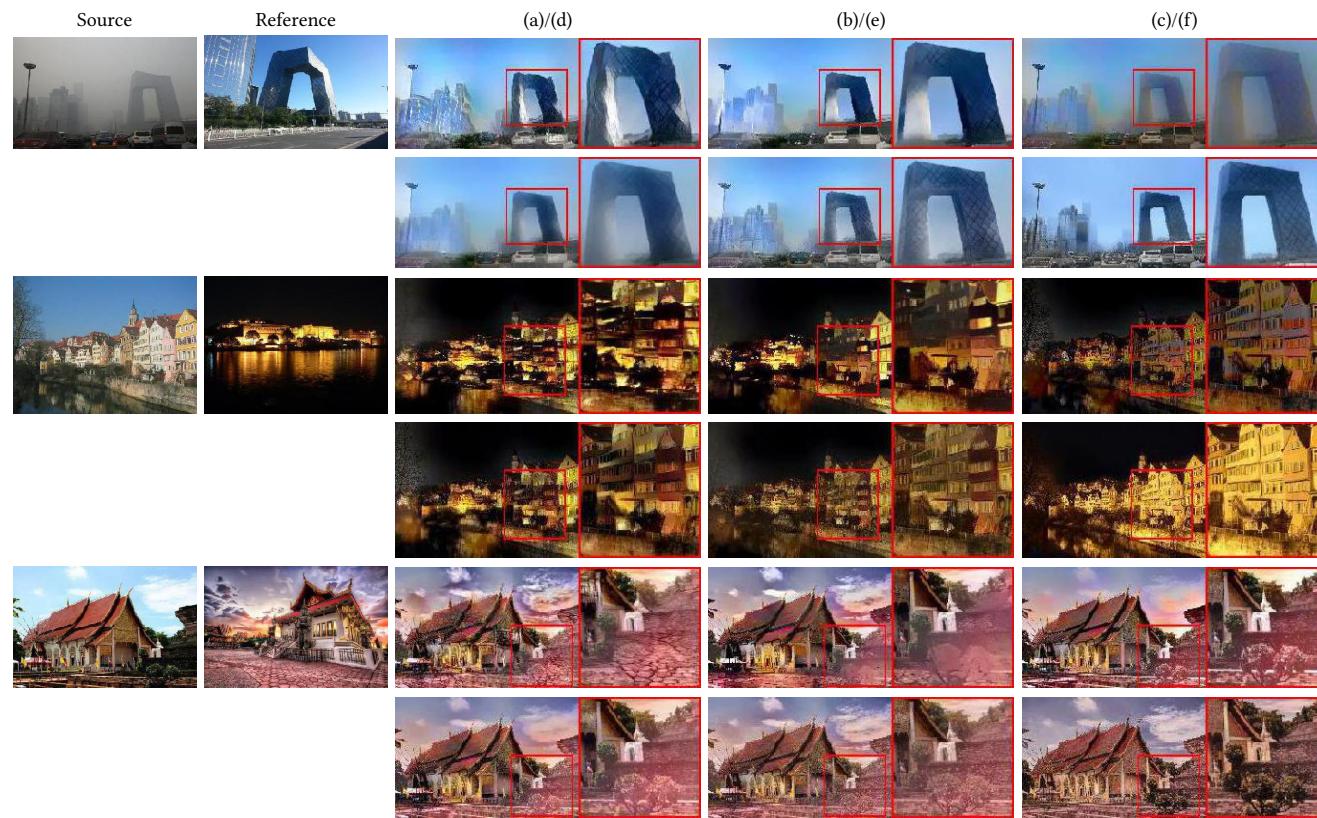


Fig. 11. Comparisons between results by different color transfer methods from the aligned reference to the source. (a) represents the aligned reference by Gatys et al. [2015] with a segmentation mask, (b)–(e) are the results refined by Luan et al. [2017], Mechrez et al. [2017], Liao et al. [2017] and our local color transfer method respectively, and (f) is the result of our joint optimization approach. Input images: Luan et al. [2017].

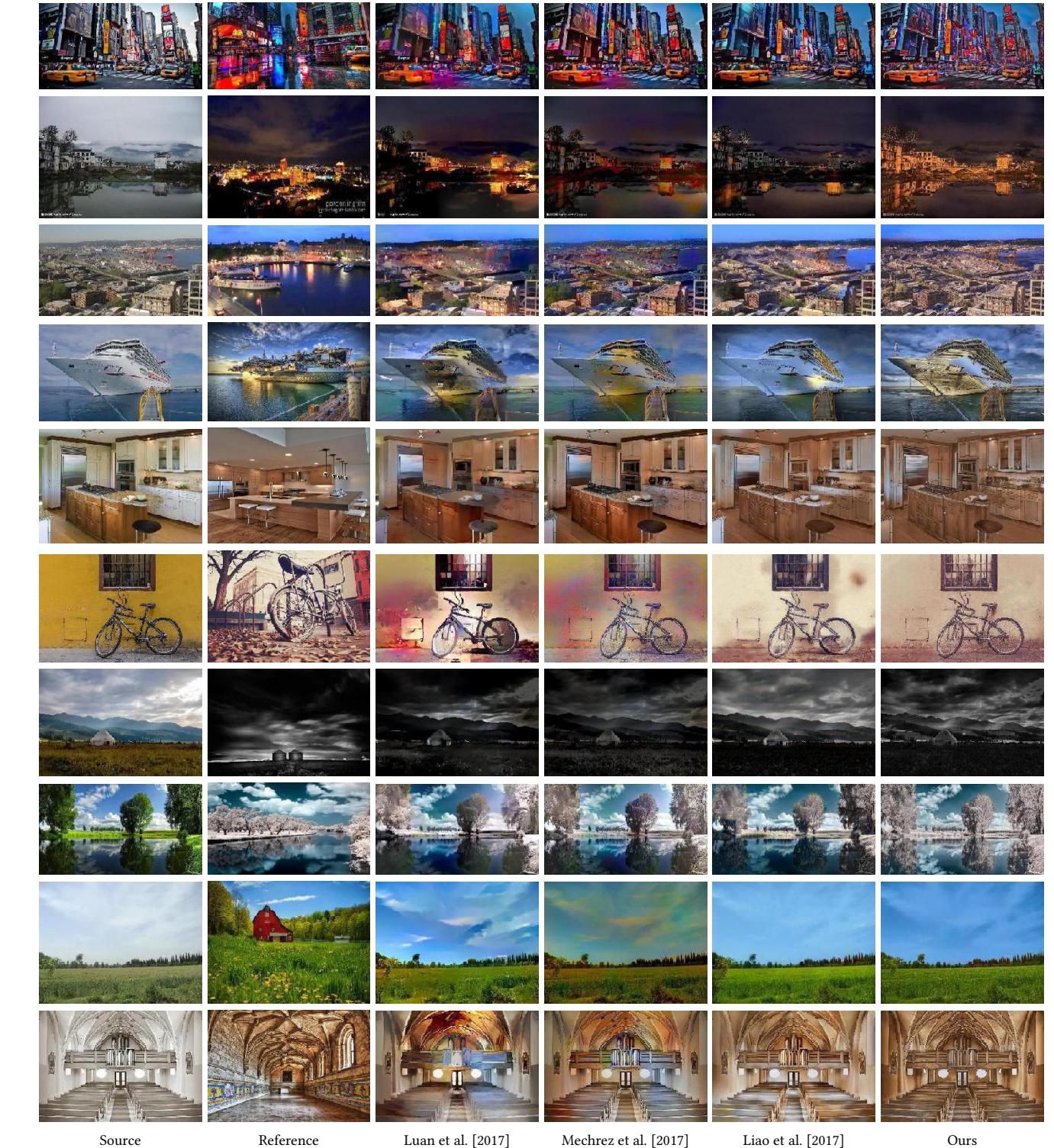


Fig. 15. 与基于深度特征的最近颜色转移方法进行比较。输入图像: [Luan et al. 2017]。



Fig. 12. Comparison with global color transfer methods [Reinhard et al. 2001] and [Pitie et al. 2005]. The test images are from Luan et al. [2017]. Input images: Luan et al. [2017].

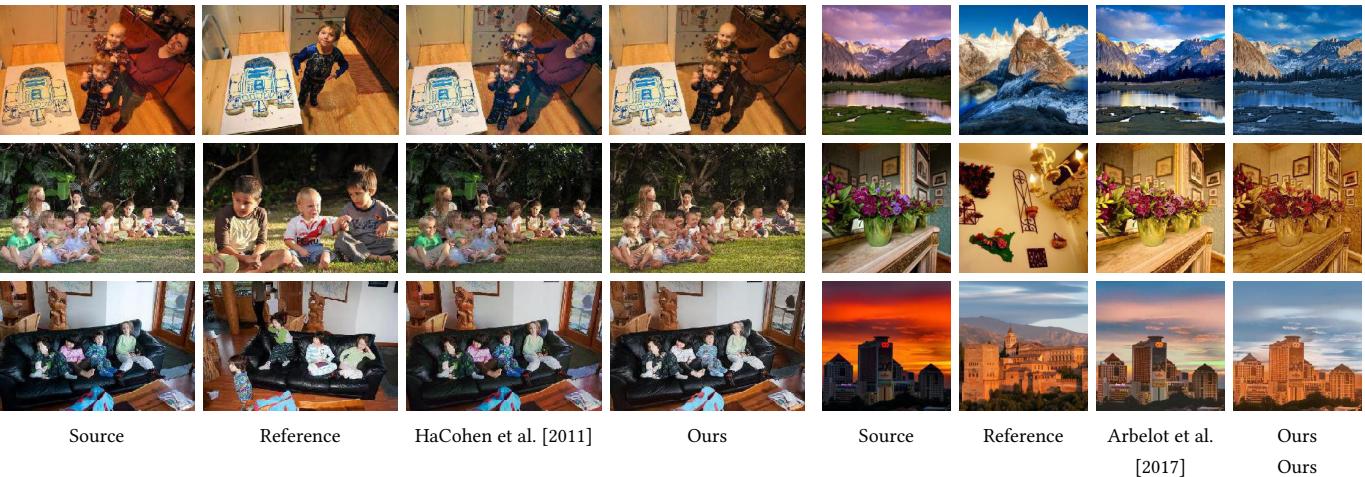


Fig. 13. Comparison with traditional local color transfer methods. The test images are from HaCohen et al. [2011] (left) and Arbelot et al. [2017] (right). Input images: HaCohen et al. [2011] and Arbelot et al. [2017].

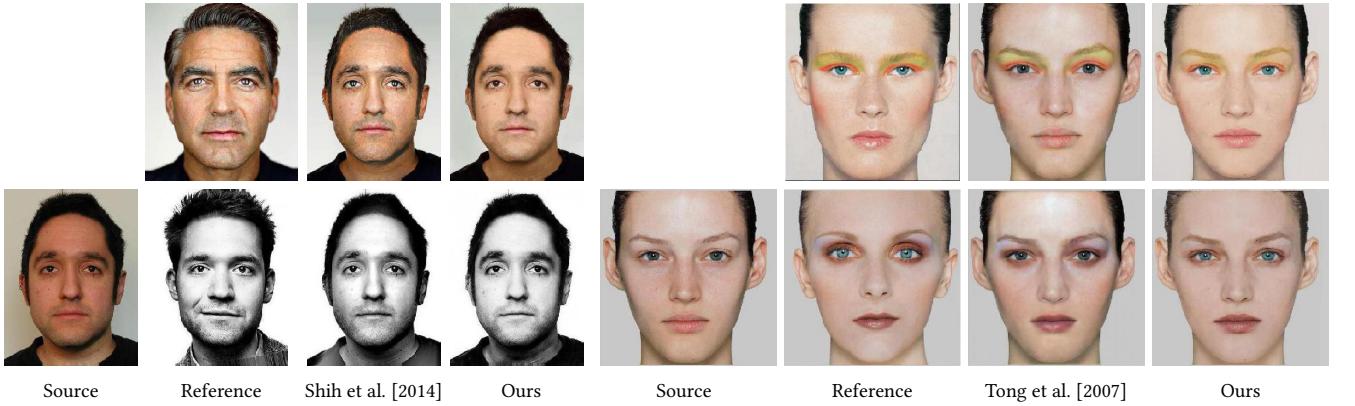


Fig. 16. 比较肖像风格转换（左）和化妆转换（右）。输入图像: Shih et al. [2014]和 Tong et al. [2007]。

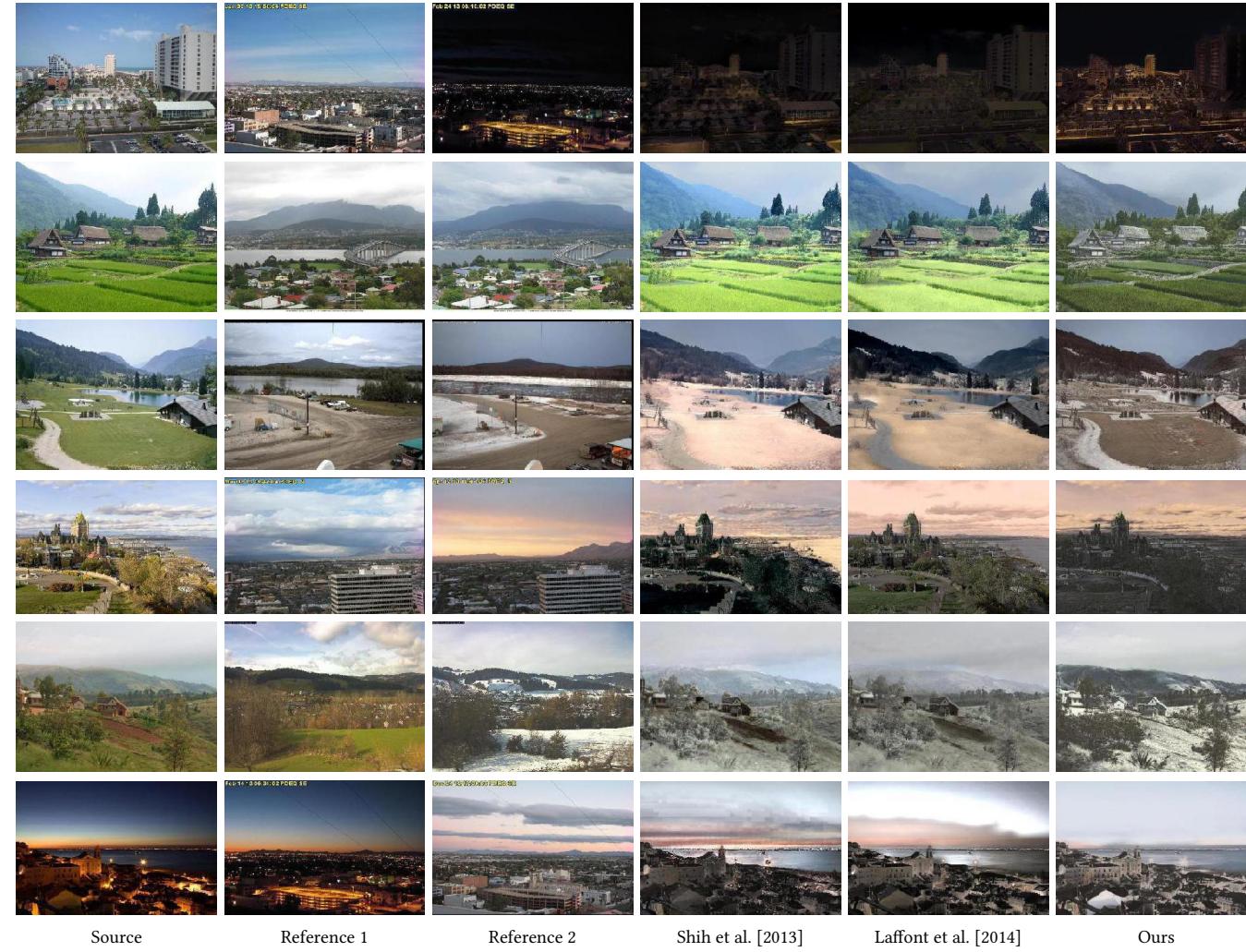


Fig. 14. Comparison to the analogy-based local color transfer methods on the data from Laffont et al. [2014]. Shih et al. [2013] and Laffont et al. [2014] take both Reference 1 and Reference 2 as references while ours only takes Reference 2. Input images: Laffont et al. [2014].

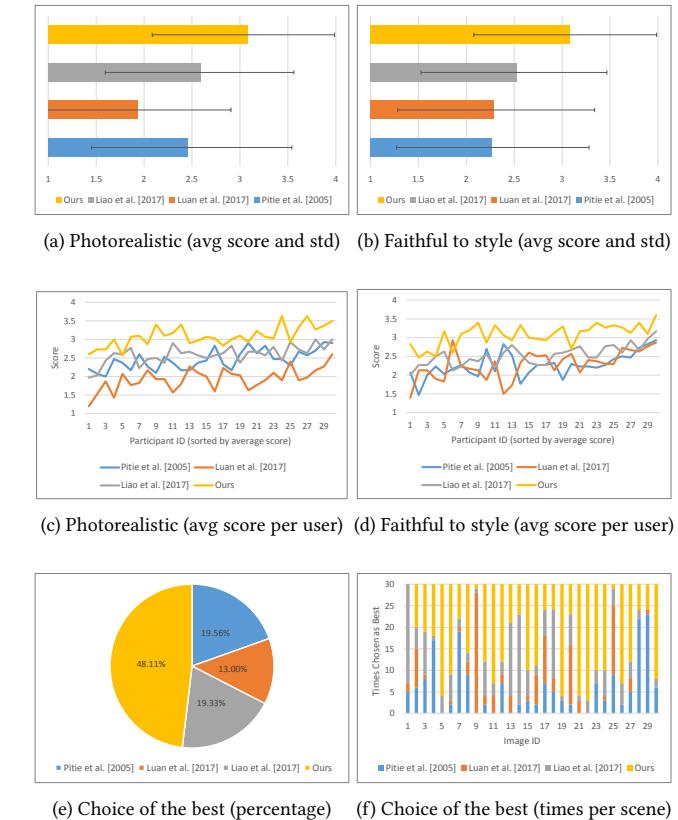


Fig. 17. 感知研究结果。(a) 和 (b) 展示了四种方法:Pitie et al. [2005]、Luan et al. [2017]、Liao et al. [2017] 和我们的平均得分 (以数据条形图表示) 和标准偏差 (以误差条表示) 在真实感和忠实度方面。(c) 和 (d) 显示了参与者在两个方面的平均得分。(e) 和 (f) 说明了每种方法在总得分和每个例子中在真实感和忠实度方面被选为最佳的百分比和次数。

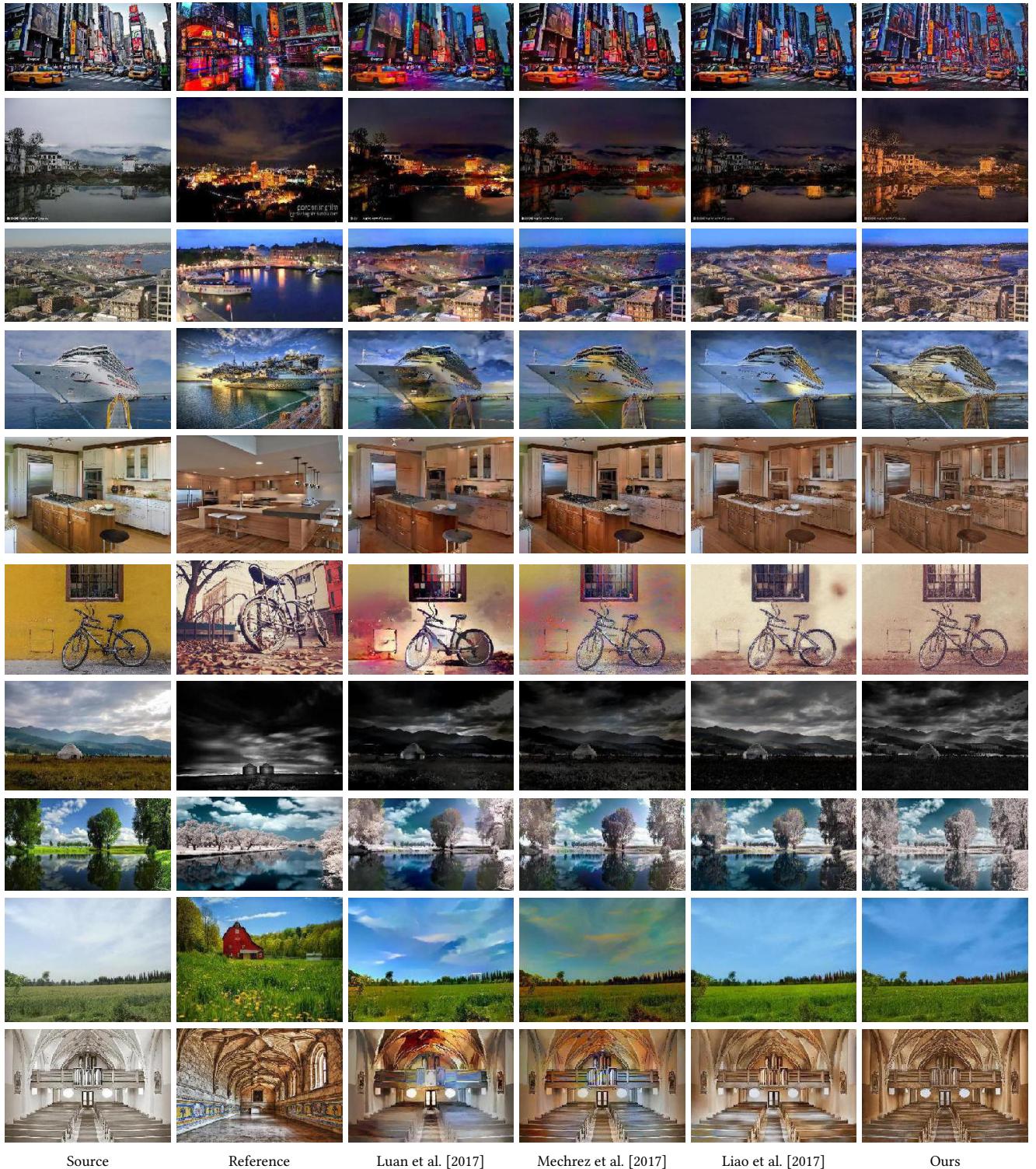


Fig. 15. Comparison to recent color transfer methods based on deep features. Input images: [Luan et al. 2017].

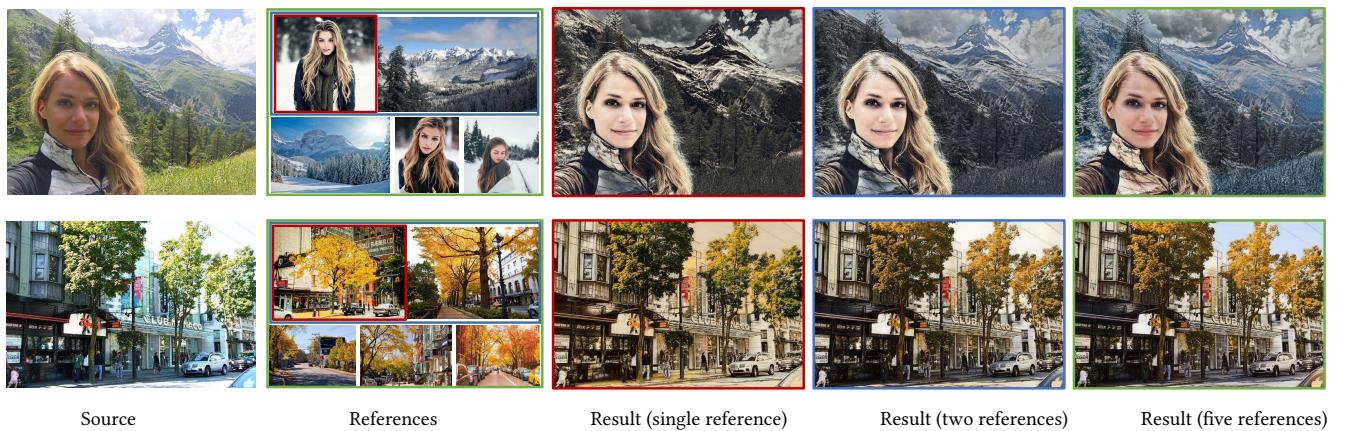


Fig. 18. 比较我们的方法与单一和多个参考文献。请注意,每个结果都是使用与它相同边框颜色的参考文献分别生成的。输入源图像: 和 Anonymous/Flickr.com。

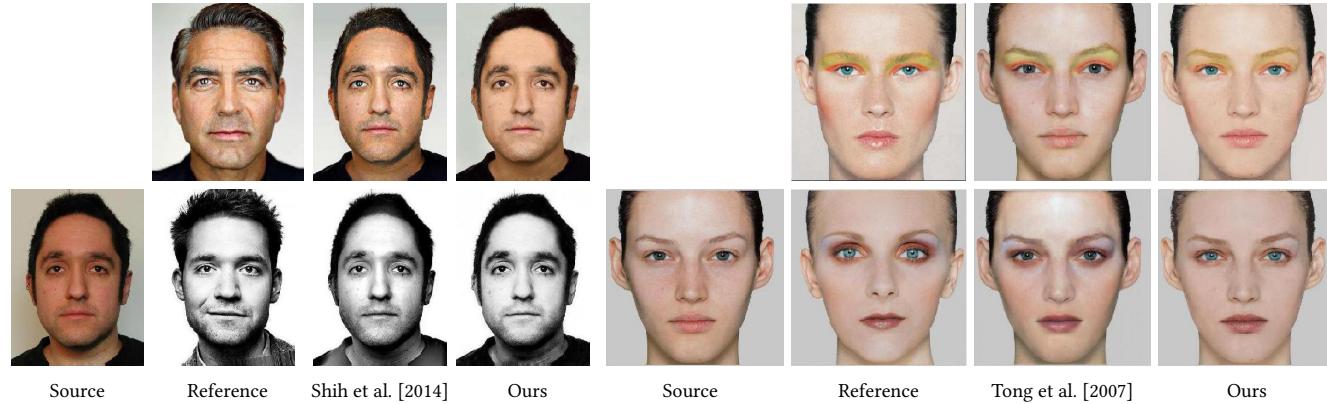


Fig. 16. Comparison of portrait style transfer (left) and cosmetic transfer (right). Input images: Shih et al. [2014] and Tong et al. [2007].

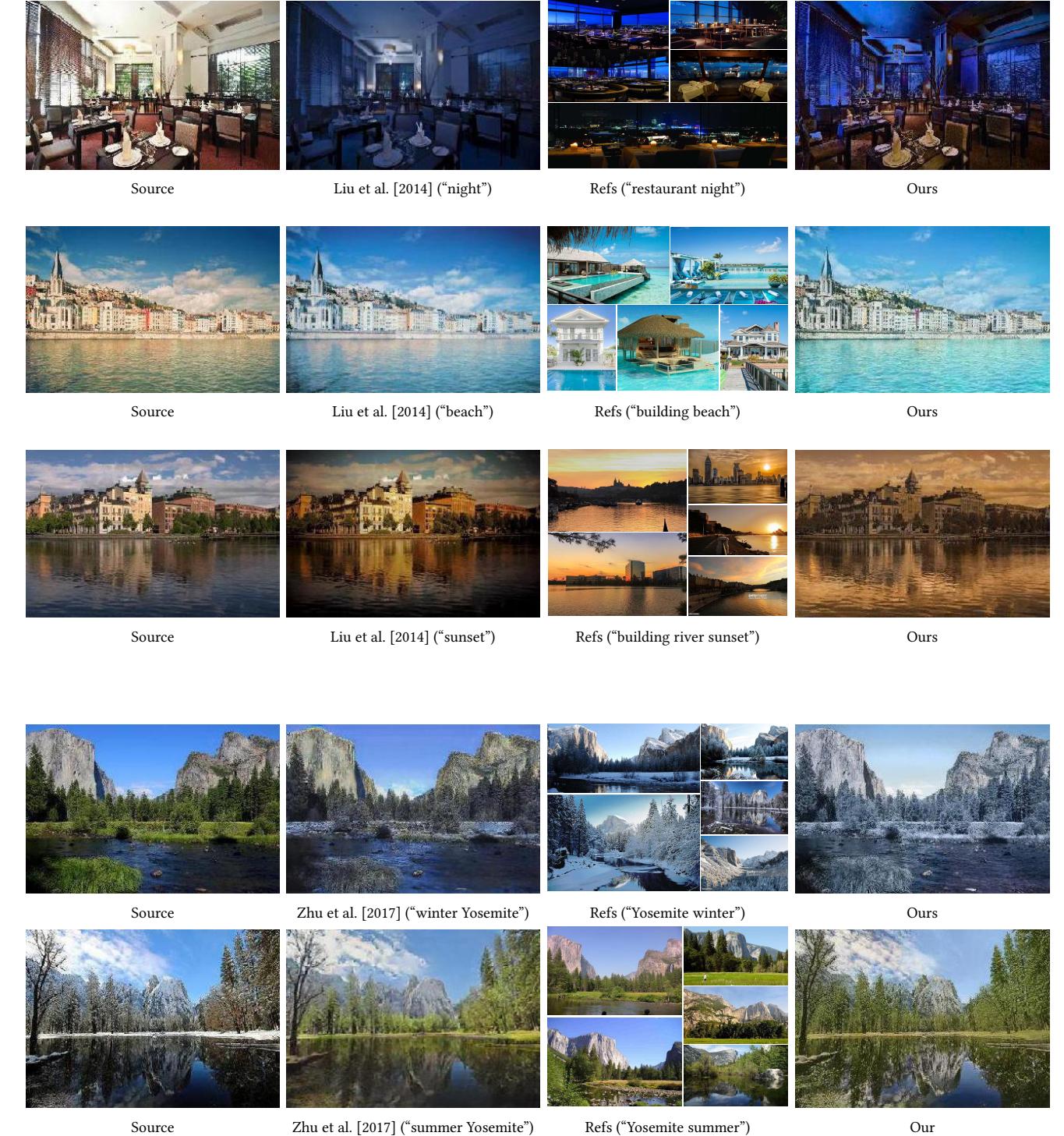
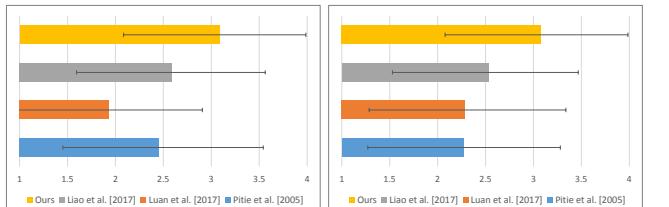
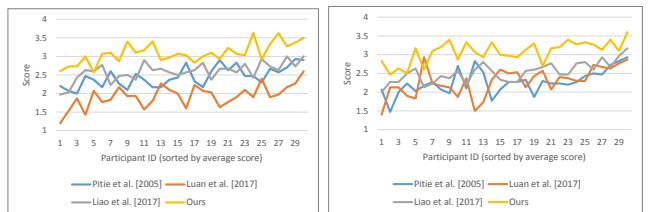


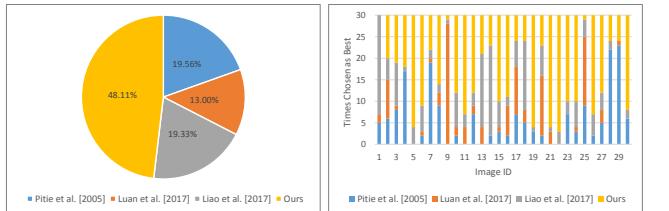
Fig. 19. 与多种颜色转移方法在它们源图像和我们从互联网自动检索的参考文献上进行比较，基于关键词。输入图像： Liu et al. [2014]（前三张）和 Zhu et al. [2017]（后两张）。



(a) Photorealistic (avg score and std) (b) Faithful to style (avg score and std)



(c) Photorealistic (avg score per user) (d) Faithful to style (avg score per user)



(e) Choice of the best (percentage) (f) Choice of the best (times per scene)

Fig. 17. Perceptual study results. (a) and (b) demonstrate the average scores (as data bars) and the standard deviations (as error bars) of the four methods: Pitie et al. [2005], Luan et al. [2017], Liao et al. [2017] and ours in photorealism and faithfulness. (c) and (d) show the average scores given by each participant in the two aspects. (e) and (f) illustrates the percentage and the times of each method voted as the best in total and in every example based on photorealism and faithfulness.



Fig. 20. 彩色化结果。输入源图像: M. Sternberg 和 Urszula Kozak。

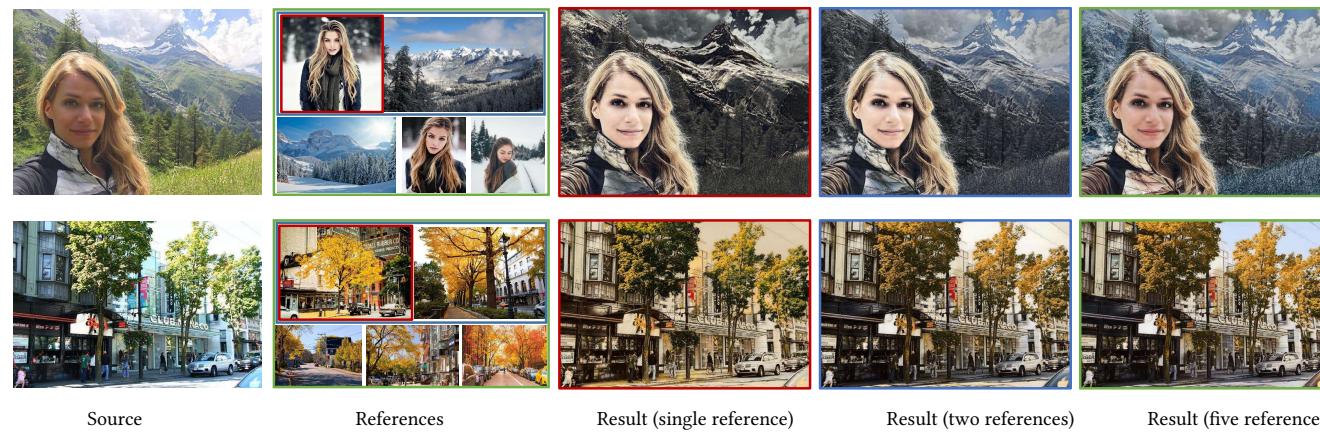


Fig. 18. Comparison of our method with single and multiple references. Please note that each result is generated respectively using the references in the same border color as it. Input source images: and Anonymous/Flickr.com.

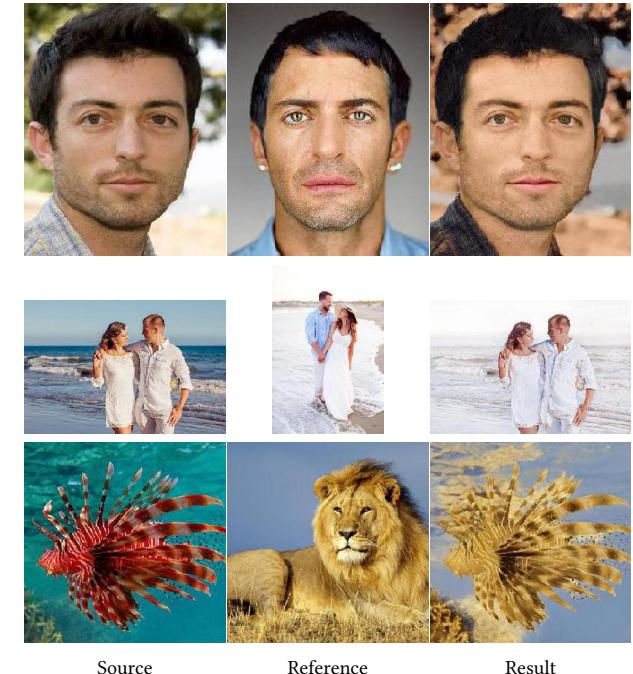


Fig. 21. Some examples of failure cases.

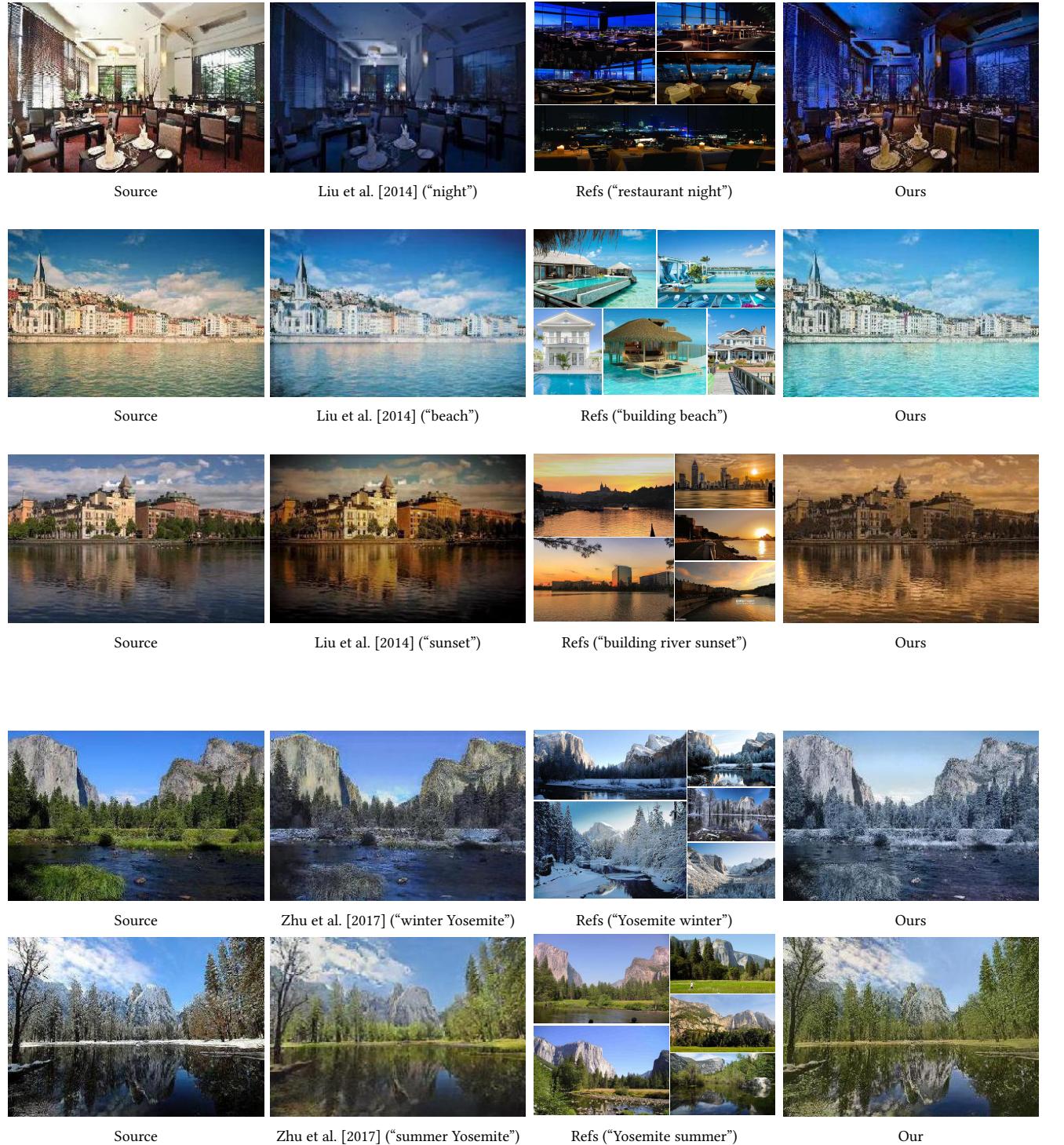


Fig. 19. Comparison with multiple color transfer methods on their source images and our own references automatically retrieved from the Internet based on the keywords. Input images: Liu et al. [2014] (top three) and Zhu et al. [2017] (bottom two).

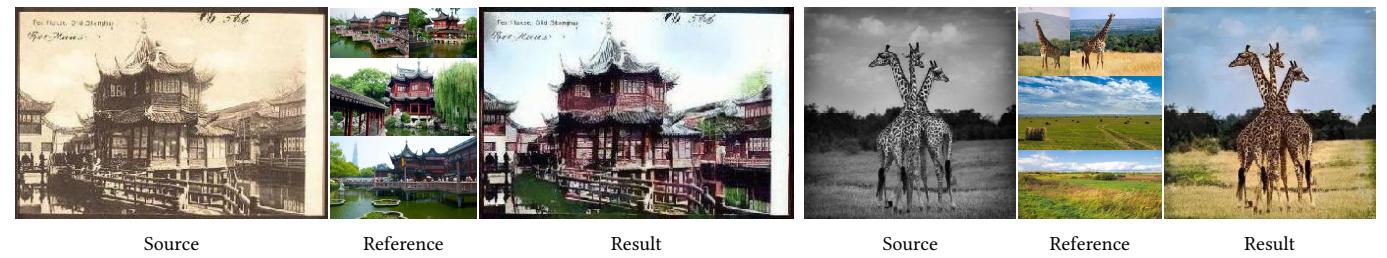


Fig. 20. Colorization results. Input source images: M. Sternberg and Urszula Kozak.



Fig. 21. Some examples of failure cases.