

基于最优传输模型的人脸颜色转换

李真熙, 张举勇

(中国科学技术大学数学科学学院, 安徽合肥 230026)

摘要: 现有人脸颜色转换算法仅基于图像的 $L\alpha\beta$ 颜色空间匹配均值和方差, 因而仅限于线性变换且通常适用于自然图像; 同时, 现有算法利用人脸关键点信息来定位人脸五官的位置, 但由此得到的人脸五官区域信息不是非常准确, 通常需要进一步优化处理. 针对上述问题, 在自动获得人脸区域分割的基础上, 为得到更自然的人脸颜色转换效果, 基于最优传输模型, 提出了一种新的人脸颜色转换的方法. 首先利用全卷积网络自动得到人脸区域的分割信息, 再利用最优传输模型获得对应的人脸区域的颜色转换结果. 试验结果表明, 所提算法在人脸五官分割的鲁棒性和人脸图像颜色转换的主观视觉上均得到明显的改善.

关键词: 人脸分割; 最优传输; 颜色转换; 全卷积网络

中图分类号: TP391.41 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2018.02.009

引用格式: 李真熙, 张举勇. 基于最优传输模型的人脸颜色转换[J]. 中国科学技术大学学报, 2018, 48(2): 148-153.

LI Zhenxi, ZHANG Juyong. Face color transfer based on optimal transport model[J]. Journal of University of Science and Technology of China, 2018, 48(2): 148-153.

Face color transfer based on optimal transport model

LI Zhenxi, ZHANG Juyong

(School of Mathematical Sciences, University of Science and Technology of China, Hefei 230026, China)

Abstract: Existing face color transfer algorithms are only based on $L\alpha\beta$ color space for matching mean and variance, and such transfer is limited to linear transformation and is usually applied to natural images. At the same time, these algorithms usually use facial landmark information to locate facial feature positions. However, the facial feature information thus obtained lacks accuracy and needs further optimization. To address these problems, a new face color transformation method was proposed based on the optimal transfer model to get the natural face color transfer effect. Firstly, the semantic information of the face was obtained directly by using the fully convolution network, and the corresponding face color transfer result was obtained by using the optimal transfer model. Experimental results show that the proposed algorithm is significantly improved in both the robustness of facial features segmentation and the subjective vision of face image color transfer.

Key words: face segmentation; optimal transport; color transfer; fully convolutional network

收稿日期: 2017-01-09; 修回日期: 2017-03-28

基金项目: 国家重点研发计划(2016YFC0800501), 国家自然科学基金(61672481) 资助.

作者简介: 李真熙, 女, 1992年生, 硕士. 研究方向: 深度学习, 人脸图像分割和编辑. E-mail: jancee@mail.ustc.edu.cn

通讯作者: 张举勇, 副教授/博士. E-mail: juyong@ustc.edu.cn

0 引言

颜色转换是在保持原始图片内容不变的情况下,将样例图片中的颜色风格应用于原始图片上.随着社交网络迅猛发展和移动电话的全面普及,人们对于人脸图像处理的需求日益增加,而颜色转换能够方便而直接复制给定样例图像的颜色风格,提高人脸图像的吸引力.

颜色转换的概念最早由 Shirley 等^[1]提出,他们运用简单的统计规律,匹配原始图像和样例图像颜色空间的均值和方差,但该方法比较适用于整张自然图像的颜色转换.之后, Welsh 等^[2]提出了利用样例彩色图像给灰度图上色的算法,基本原理是比较每个像素点的亮度和邻域统计,为灰度图的每一个像素点找到彩色图像中的最佳匹配值,但是对于比较精确的转换,需要人工指定图片中的对应区域. Levin 等^[3]认为,图片中相邻像素的颜色亮度若是相同,则他们也应该有同样的颜色.所以他们的目标就是最小化每个像素和它邻域像素颜色平均值的差异,但其最终的效果依赖于手动选择区域和颜色.

随着技术的发展,基于图像颜色空间的方法开始逐渐涌现. Neumann 等^[4]为了达到色域的精确匹配,将其转化为 *HSL* 颜色空间的三维直方图的匹配问题.而 Xiao 等^[5]在 *L $\alpha\beta$* 颜色空间中匹配直方图,同时使用了优化算法来保留原始图像的梯度. Pitie 等^[6]提出了一个转换 *N* 维概率分布函数的方法,由于其考虑了图像不同通道之间的相关性,所以它能较好地转换图像颜色,但会存在比较明显的空间伪影. Abadpour 等^[7]使用主成分分析,计算了一个适合特定输入图像的不相关颜色空间,并为灰度图彩色化提出了一个统一的框架.

近年来,最优传输理论作为图像颜色转换问题的一个有力的处理工具也日益绽放其光彩. Pitie 等^[8]通过在随机轴上迭代估计了一维的最优传输来计算图像之间的对应关系,同时也较好地解决了文献^[6]中的空间伪影问题. Tai 等^[9]将颜色转换定义于最大期望迭代算法给出的图像分割上,其颜色的对应主要是基于 Monge-Kantorovitch 最优传输问题. Frigo 等^[10]基于样例图片,对所有图像分通道进行处理.对于光照通道,采用色彩适应变换(CAT)进行匹配,对于色彩通道则利用最优传输实现. Ferradans 等^[11]借助离散最优传输理论来转换图像的颜色,但却只能支持不精确的颜色直方图的对应.

而对于人脸的颜色变换, Yang 等^[12]基于文献^[1]的方法,借助关键点的信息得到人脸分割区域,并借助手工交互来优化场景和人脸区域的分离.由于关键点信息并不考虑遮挡问题,而且优化所需的时间较长,所以 Yang 等的方法并不能完成全自动转换.而本文利用 Solomon 等^[13]提出的基于卷积 Wasserstein 距离,通过匹配原始图像和样例图像颜色分布的概率测度来完成人脸图像的分区转换.实验证明,本文的方法在全自动获得人脸图片分割的基础上,可以完成人脸区域的部分或者整张脸的颜色转化,且最终得到较为真实的转换效果.

1 最优传输模型描述

1.1 原始最优传输模型

最优传输问题最早由法国数学家蒙日(Monge)于 1781 年提出^[14].假设 X 和 Y 是完备、可分的度量空间, μ 和 ν 是分别定义于 X 和 Y 空间上的概率测度,传输代价函数 $c: X \times Y \rightarrow [0, +\infty]$ 为光滑函数, $c(x, y)$ 代表点 $x \in X$ 传输单位质量到点 $y \in Y$ 所花费的代价.令 $T: X \rightarrow Y$ 为从空间 X 到空间 Y 的映射,若对于任意的波莱尔集 $B \subset Y$, 其原像为 $T^{-1}(B) \subset X$, 它们测度相同, 即有 $\int_{T^{-1}(B)} \mu = \int_B \nu$. 令 $T_{\#}\mu$ 表示 T 将 μ 推前为 Y 空间上的概率测度,若映射是保测度的,则将其记为 $T_{\#}\mu = \nu$. 则最优传输问题就是求保测度下具有最小传输代价的映射 T_0 ,

$$T_0 := \arg \min_{T_{\#}\mu(x)=\nu} \int c(x, T(x)) d\mu(x) \quad (1)$$

1.2 最优传输模型应用于颜色转换

对于图像的颜色转换,则可以这样理解这个问题:假设 $\mu: \Omega \mapsto \mathbb{R}^n$ 表示原图像颜色分布的概率测度, $\nu: \Omega \mapsto \mathbb{R}^n$ 表示样例图像颜色分布的概率测度,其中 $\Omega \in \mathbb{Z}^2$ 表示有界图像域, \mathbb{R}^n 表示图像的颜色空间, $n=3$ 时为彩色图片.则期望寻找颜色空间的自映射 T , 使得 $T(\mu)$ 在某种程度下逼近 ν , 即 $T_{\#}\mu = \nu$, 同时使得总的转换代价最小.

本文主要利用文献^[13]的算法来得到人脸颜色转换效果.这个算法的主要思想在于:若要将原始图像的颜色转换为样例图像,则在一定的误差下,保证两者的颜色直方图尽量相同即可,这就需要利用文献中提出的重心映射的方法.具体做法如下:

(I) 将颜色空间从 *RGB* 转换为 *L $\alpha\beta$* , 然后对

于一维亮度通道 L 和二维色度通道 $\alpha\beta$ 分别处理,其中亮度值的范围为 $L = [0, 100]$, 色度值的范围为 $M = [-128, 128]^2$. 对于一维通道的转换问题, 最优传输转换已经有很多解法, 本文主要参考文献[14]. 对于二维通道, 假定 k 张图像的 $\alpha\beta$ 通道的转换由函数 $(f_i)_{i=1}^k$ 表示, 且有 $f_i: [0, 1]^2 \rightarrow M$ 将图像平面上的一个点转换为颜色值输出. 本文主要对两张图做处理, 即 $k = 2$. f_i 的色度直方图 μ_i 表示属于 $[0, 1]^2$ 上的统一测度 \mathcal{U} 基于 f_i 的推前, 并满足 $\mu_i(A) = \mathcal{U}(f_i^{-1}(A))$, $A \subset M$.

(II) 对于一组给定的权值 $\alpha = (\alpha_i)_{i=1}^k \in \mathbb{R}_+^k$, 利用方程(2)将加权重心 ν 解出,

$$\min_{\nu} \sum_{i=1}^k \alpha_i W_2^2(\nu, \mu_i) \quad (2)$$

式中, $W_2^2(\nu, \mu_i)$ 表示 ν 和 μ_i 的熵正则化 Wasserstein 距离.

(III) 对每一张图像, 定义一个最优传输映射 $T_i: M \rightarrow M$, 使得

$$\forall x \in M, T_i(x) = \frac{1}{\mu_i(x)} \int_M \pi_i(x, y) y dy \quad (3)$$

式中, $\mu_i(x) > 0$, $\pi_i(x, y)$ 表示 μ_i 与 ν 之间的传输映射.

最终, 求出保测度下, 具有最小传输代价的映射. 输出图像的色度转换函数可以定义为 $f_i^s := T_i \circ f_i$, 其色度直方图与 ν 相匹配. 在对每一个区域进行颜色变换之后, 本文简单地应用 alpha 融合的方法来整合为整张人脸图像.

2 改进的人脸颜色转换模型

对于人脸图像, 只对其进行全局的颜色转换是不够的. 如图 1 所示, 若样例图像的背景颜色过于丰富, 则会导致原始图像中的人脸区域有较大比例的背景颜色分布. 所以, 通过对人脸区域的逐像素标定, 对人脸图像中, 每个语义区域进行分开处理是本文的下一步目标.



图 1 人脸图片的全局颜色转换效果

Fig.1 The color transformation effect of the whole image

2.1 神经网络的使用

神经网络兴盛于 AlexNet^[15], 主要用于物体图像的分类. 为了得到图像中每一个像素的分类结果, 基于传统分类网络的做法如下:

① 为了得到每一个像素的分类结果, 需要在原始图像边界进行扩展, 常见的操作有填充 0;

② 选取一个固定大小的尺寸, 借助滑动窗口, 让分类网络对每一个窗口的图像块进行判定;

③ 通过 shift-and-stitch 操作^[16] 获得原始图片大小的逐像素分类结果.

但是这样的做法有比较明显的缺陷: 首先, 网络需要的储存空间会随着滑动次数不断增加; 其次, 由于所取图像块一般要比原始图像小很多, 所以网络一般只能学到局部特征; 再者, 滑动窗口的做法会导致有大量重复的计算, 使得整个网络的运算效率低下. 所以, 我们没有用传统的做法, 转而利用 FCN^[17] 得到人脸图像中每个像素点所属的类别.

2.1.1 网络组成

以输入 250×250 像素的彩色图片为例, FCN 的网络结构如图 2 所示, 主要包括: 卷积层 (convolution layer): 其中包含卷积和 ReLU 操作, 主要用来提取重要的特征和进行非线性变换; 池化层 (pooling layer): 对特征进行抽象, 减少训练参数, 从而降低过拟合的程度; 去卷积层 (deconvolution layer): 近似于池化的逆操作, 将网络输出的尺寸逐渐还原至原始图片尺寸; 裁剪层 (crop layer): 将不同尺寸的输入裁剪为相同尺寸; 融合层 (eltwise layer): 融合多个相同大小的层; 损失层 (loss layer): 用来度量网络输出与真实值之间的误差, 并进行反向传播.

2.1.2 数据集组成

为了得到精确的分割结果, 本文采用有监督的学习方式. 从 LFW 数据集^[18] 中选取了 1000 张人脸图像, 大小为 250×250 像素, 其中 800 张用于训练, 200 张用于验证. 通过编写图像标注界面, 手工标注人脸五官特征信息, 具体将人脸图像分为 6 个区域, 分别为背景, 脸部皮肤, 眉毛, 眼睛, 嘴唇和口腔区域, 部分原始图片与标注如图 3 所示. 一般来说, 在实际训练中, 使用越多的数据量, 训练的效果越好, 所以通过旋转、放缩和亮度变换来进行数据扩增, 最终的训练数据集大约包含 10000 张尺寸不同的图片.

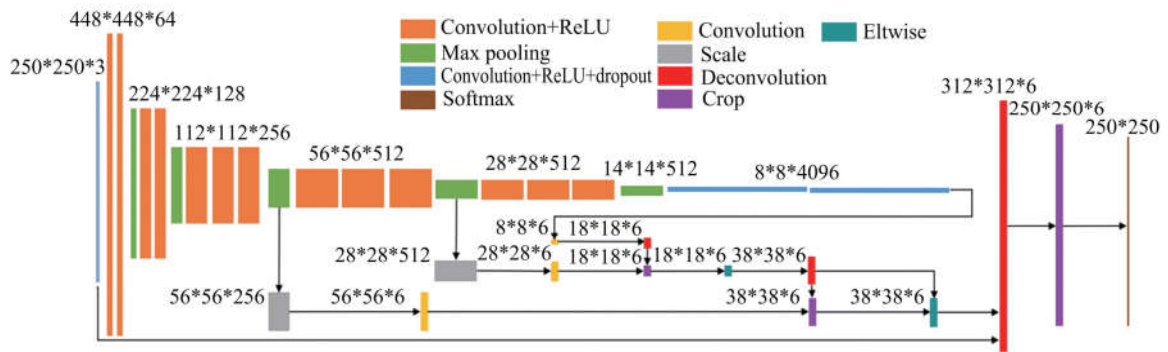


图 2 FCN 的网络结构

Fig.2 The network structure of FCN



图 3 部分原始图片与其对应的手工标记结果

Fig.3 Some original images and its corresponding manual marking results

为获得比较好的转换效果,另从网上的图片库中下载大约 120 张不同背景、肤色和性别的人脸图像用于测试,部分图像如图 4 所示。



图 4 部分用于颜色转换的图像

Fig.4 Some images for color transformation

2.1.3 参数设置与网络运行

根据文献[17]中的描述,最终选择 FCN-8s-atonce 作为初始化的模型,并设置固定的训练参数:学习率为 10^0 ,动量项为 0.99,权值衰减为 5^4 ,批量训练数目设置为 1.采用高动量项和低批量处理速度的原因不仅在于能够减少内存的消耗,还能近似达到梯度移动平均值的效果.对于任意尺寸图像的逐像素标记任务,将输出的损失归一化是一个无意义的操作,因为将损失归一化后,基于像素的损失将取决于图像的大小.而由于未归一化的损失比较大,所

以,不同于传统的分类网络,FCN 只需要比较小的学习率进行训练。

训练和测试的过程均运行于 Nvidia Titan X GPU 上,用时约为 3 d.由于 FCN 支持任意尺寸的图片,为了测量测试所需的时间,本文选用 250×250 像素的图片作为输入.在相同的配置下,推断一张 250×250 像素的图像大约需要 0.1 s.

2.1.4 网络标记结果

部分人脸图像的分割结果如图 5 所示,第一行为原始图像,第二行为本文的分割结果,从图中可以看出,通过网络的有监督学习后,本文能全自动地得到人脸图像的精确分割结果。

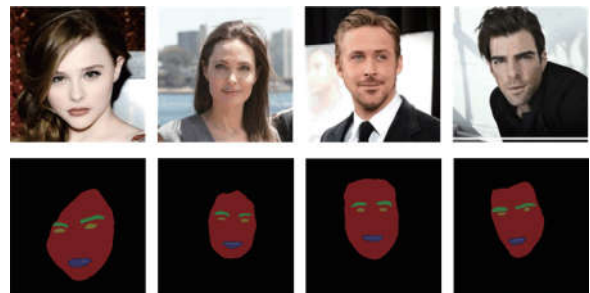


图 5 部分人脸图像分割结果

Fig.5 The segmentation results of some face images

2.2 最终模型

本文算法的整体结构如图 6 所示,主要包含如下几步:

- ① 借助全卷积神经网络快速而又准确地得到原始图像与样例图像的分割结果,每张人脸图像都分为 6 类:背景,脸部皮肤,眉毛,眼睛,嘴唇和口腔;
- ② 参考样例图像的颜色风格,利用最优传输模型,调整原始人脸图像中,除开背景区域的每一个区域的颜色风格;
- ③ 将②中得到的结果进行整合,输出最终的图像。

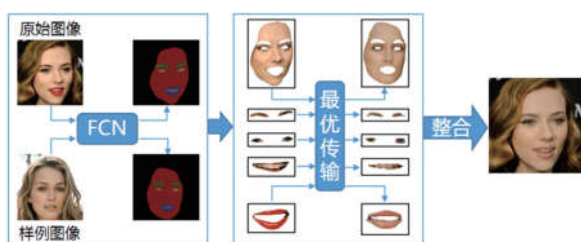


图 6 本文算法整体结构图

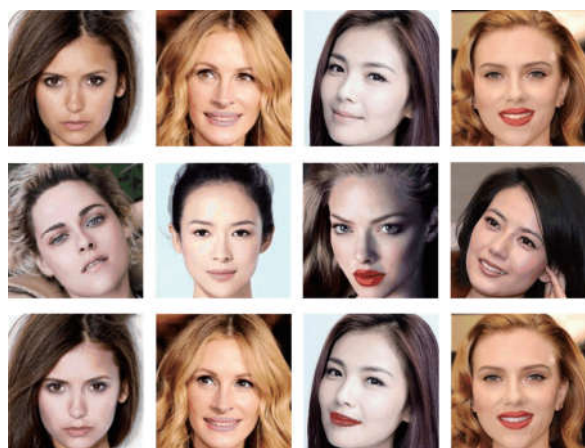
Fig.6 The overview of algorithm

3 实验结果及分析

由于已经得到精确的人脸图像区域划分,本文的方法不仅能对整张脸进行颜色转换,还能对脸部某一语义区域单独进行颜色变换,整个过程均为计算机全自动操作,不需要用户手动划分脸部区域。

3.1 脸部区域颜色变换

通过指定某一人脸区域,可以只对这一区域进行颜色转化,具体结果如图 7 所示,图中第一行为原始图像,第二行为对应的样例图像,最后一行为输出的结果图像。从图中可以看出,本文的方法能得到非常逼真的转换效果。



(a) 脸部皮肤 (b) 眼睛 (c) 嘴唇 (d) 口腔

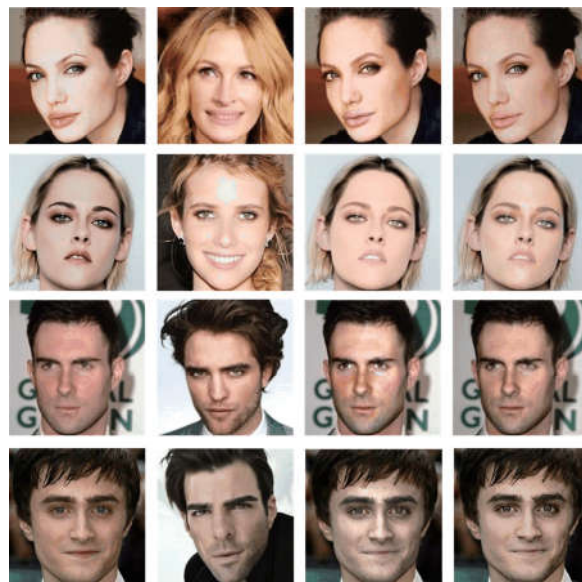
图 7 脸部部分区域颜色变换

Fig.7 The color transformation of the face part area

3.2 全脸颜色变换

全脸颜色变换是指在人脸图像中,保留背景不变,对人脸所有对应区域分别进行颜色变换,然后简单利用 alpha 混合将每一个区域整合起来,具体效果如图 8 所示。由于文献[12]仅是对原始图像的直接线性转换,而本文的方法通过解决方程,使原始图像的色度空间和亮度空间尽量逼近的样例图像的同时,提供了一种非线性的变换,更加凸显了变换后人脸图像的真实感。从图 8(c)可以看出,文献[12]的结

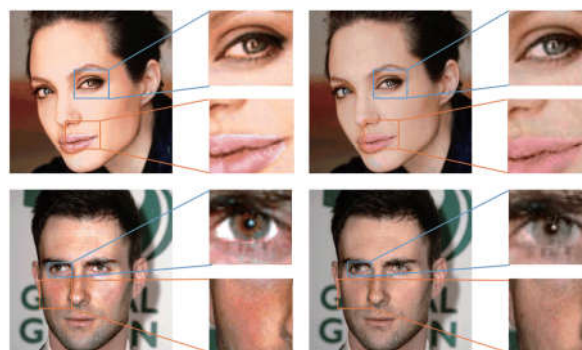
果对于人脸皮肤等区域的颜色转换比较生硬,而且在某些情况下,瞳孔的颜色会导致眼睛整体颜色转换的不协调,部分细节图如图 9 所示。所以对比可知,本文的结果相比于文献[12]的结果,图片转换更为自然和协调。



(a) 原始图像 (b) 样例图像 (c) 文献[12]算法 (d) 本文结果

图 8 全脸颜色变换效果

Fig.8 The effect of full-face color transformation



(a) 文献[12]结果的细节图 (b) 本文结果的细节图

图 9 人脸颜色变换的细节对比图

Fig.9 Detail contrast of face color transformation

4 结论

本文提出了一种基于最优传输模型的人脸图像的颜色转换算法,其中,使用全卷积神经网络自动且快速得到对应人脸的分割图,通过对人脸图像的分区域处理,克服了传统图像颜色转换对于复杂背景过于敏感的缺点。实验证明,本文的算法不仅能够快速得到较为精确的人脸区域分割图,而且借助分割结果,最终在人脸图像上得到更为真实和细腻的颜色转换的视觉效果。但本文对于分割边缘的处理还

不够完善,同时也缺乏一种度量颜色转换好坏的数值方法,下一步将对此继续深入研究.

参考文献(References)

- [1] SHIRLEY P, ADHIKHMIM M, GOOCH B, et al. Color transfer between images [J]. IEEE Computer Graphics and Applications, 2001, 21: 34-41.
- [2] WELSH T, ASHIKHMIM M, MUELLER K. Transferring color to greyscale images [J]. ACM Transactions on Graphics (TOG), 2002, 21 (3): 277-280.
- [3] LEVIN A, LISCHINSKI D, WEISSY. Colorization using optimization[J]. ACM Transactions on Graphics (TOG), 2004, 23(3): 689-694.
- [4] NEUMANN L, NEUMANNA. Color style transfer techniques using hue, lightness and saturation histogram matching [C]// Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging. Switzerland: Eurographics Association Aire-la-Ville, 2005: 111-122.
- [5] XIAO X, MA L. Gradient-preserving color transfer [C]//Computer Graphics Forum. Hoboken, NJ: Blackwell Publishing Ltd, 2009, 28(7): 1879-1886.
- [6] PITIE F, KOKARAM A C, DAHYOT R. N-dimensional probability density function transfer and its application to color transfer [C]// Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005, 2: 1434-1439.
- [7] ABADPOUR A, KASAEI S. An efficient PCA-based color transfer method [J]. Journal of Visual Communication and Image Representation, 2007, 18 (1): 15-34.
- [8] PITIÉ F, KOKARAM A C, DAHYOT R. Automated colour grading using colour distribution transfer[J]. Computer Vision and Image Understanding, 2007, 107 (1): 123-137.
- [9] TAI Y W, JIA J, TANG C K. Local color transfer via probabilistic segmentation by expectation-maximization [C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 747-754.
- [10] FRIGO O, SABATER N, DEMOULIN V, et al. Optimal transportation for example-guided color transfer[C]// Asian Conference on Computer Vision. Springer International Publishing, 2014: 655-670.
- [11] FERRADANS S, PAPADAKIS N, PEYRÉ G, et al. Regularized discrete optimal transport [J]. SIAM Journal on Imaging Sciences, 2014, 7(3): 1853-1882.
- [12] YANG Y, ZHAO H, YOU L, ET AL. Semantic portrait color transfer with internet images [J]. Multimedia Tools and Applications, 2017, 76 (1): 523-541.
- [13] SOLOMON J, DE GOES F, PEYRÉ G, et al. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains [J]. ACM Transactions on Graphics (TOG), 2015, 34 (4): No. 66.
- [14] VILLANI C. Topics in Optimal Transportation [M]. Providence, RI: American Mathematical Society, 2003.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTONG E. ImageNet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems 25. Lake Tahoe, Nevada: Neural Information Processing Systems Foundation, 2012: 1097-1105.
- [16] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [DB/OL]. arXiv.org; arXiv:1312.6229, 2013.
- [17] LONG J, SHELHAMER E, DARRELLT. Fully convolutional networks for semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 3431-3440.
- [18] HUANG G B, RAMESH M, BERG T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments [R]. Amherst, MA: University of Massachusetts, Amherst, 2007: Technical Report 07-49.