

基于生成对抗网络的低秩图像生成方法

赵树阳¹ 李建武¹

摘要 低秩纹理结构是图像处理领域中具有重要几何意义的结构, 通过提取低秩纹理可以对受到各种变换干扰的图像进行有效校正. 针对受到各种变换干扰的低秩图像校正问题, 利用生成式框架来缓解图像中不具明显低秩特性区域的校正结果不理想的问题, 提出了一种非监督式的由图像生成图像的低秩纹理生成对抗网络 (Low-rank generative adversarial network, LR-GAN) 算法. 首先, 该算法将传统的无监督学习的低秩纹理映射算法 (Transform invariant low-rank textures, TILT) 作为引导加入到网络中来辅助判别器, 使网络整体达到无监督学习的效果, 并且使低秩对抗对在生成网络和判别网络上都能够学习到结构化的低秩表示. 其次, 为了保证生成的图像既有较高的图像质量又有相对较低的秩, 同时考虑到低秩约束条件下的优化问题不易解决 (NP 难问题), 在经过一定阶段 TILT 的引导后, 设计并加入了低秩梯度滤波层来逼近网络的低秩最优解. 通过在 MNIST, SVHN 和 FG-NET 这三个数据集上的实验, 并使用分类算法评估生成的低秩图像质量, 结果表明, 本文提出的 LR-GAN 算法均取得了较好的生成质量与识别效果.

关键词 生成对抗网络, 低秩纹理生成对抗网络, 结构化低秩表示, 低秩约束

引用格式 赵树阳, 李建武. 基于生成对抗网络的低秩图像生成方法. 自动化学报, 2018, 44(5): 829–839

DOI 10.16383/j.aas.2018.c170473

Generative Adversarial Network for Generating Low-rank Images

ZHAO Shu-Yang¹ LI Jian-Wu¹

Abstract Low-rank texture structure is an important geometric structure in image processing. By extracting low-rank textures, images with various interferences can be rectified effectively. To solve the problem of low rank image correction with various interferences, this paper proposes to use the generation framework to alleviate poor correction results on the region without obvious low-rank properties. And a low-rank texture generative adversarial network (LR-GAN) is proposed using an unsupervised image-to-image network. Firstly, by using transform invariant low-rank textures (TILT) to guide the discriminator in the LR-GAN, the whole network can not only achieve the effect of unsupervised learning but also learn a structured low rank representation on both generation network and discrimination network. Secondly, considering that the low-rank constraint is difficult to optimize (NP-hard problem) in the loss function, we introduce a layer of the low-rank gradient filters to approach the optimal low-rank solution after many iterations guided by TILT. We evaluate the LR-GAN network on three public datasets: MNIST, SVHN and FG-NET, and verify the quality of generative low-rank images by using a classification network. Experimental results demonstrate that the proposed method is effective in both generative quality and recognition accuracy.

Key words Generative adversarial network (GAN), low-rank texture generative adversarial network (LR-GAN), structured low-rank representation, low-rank constraint

Citation Zhao Shu-Yang, Li Jian-Wu. Generative adversarial network for generating low-rank images. *Acta Automatica Sinica*, 2018, 44(5): 829–839

基于低秩的模型是近年来处理高维数据的新工具, 它的兴起受到稀疏表示和压缩感知理论^[1-2] 的推动, 由此系统而快速地发展出很多理论和应用. 因

此, 结构性的低秩表示也引起了许多视觉领域学者与专家的重视. 但在现实环境中, 我们很少能见到绝对的低秩结构, 这是由于拍摄角度, 噪声等因素对获取图像造成的畸变和干扰, 例如仿射变换、旋转变换和投影变换等, 这些变换直接限制了相关视觉任务的性能. 因此, 从原始图像的复杂变换中提取出具有鲁棒性的高语义结构特征, 成为许多计算机视觉任务的热门方向. 很多对于图像变换纹理不变性的研究开始在一些视觉任务, 例如目标识别^[3]、图像拼接^[4]、视频分析^[5] 和 3D 重建^[6] 等中大量涌现. 为了从受到变换干扰的图像中提取出其内在的低秩结构信息, 并排除这些干扰量, Zhang 等提出了具有变

收稿日期 2017-08-29 录用日期 2017-12-14
Manuscript received August 29, 2017; accepted December 14, 2017

国家自然科学基金 (61271374) 资助
Supported by National Natural Science Foundation of China (61271374)

本文责任编辑 王坤峰
Recommended by Associate Editor WANG Kun-Feng

1. 北京理工大学计算机学院智能信息技术北京市重点实验室 北京 100081

1. Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081

换不变特性的低秩纹理映射算法 (Transform invariant low-rank textures, TILT)^[7].

TILT 是一种利用低秩性与噪声的稀疏性进行低秩纹理恢复的算法, 可以从具有一定仿射变换、旋转变换干扰的二维图像中恢复出低秩纹理结构. 它的思想是通过几何变换把数据矩阵所代表的图像区域校正过来, 例如具有横平竖直、对称等特性, 这些特性可以通过低秩性来进行刻画.

基于 TILT, 在文献 [8] 中, Zhang 等提出在不使用先验知识的前提下通过对称几何理论来缓解影射图像匹配中的变换问题. 该方法基于现实世界中物体具有的对称特性, 减小影射图像和原始图像关键点之间的匹配误差, 这种不使用先验知识的方法极大地提升了关键点的匹配精度. 更进一步, Zhang 等在文献 [9] 中提出学习一种结构化的低秩表示, 该方法通过引入一个稀疏的低秩目标函数, 有助于识别图像的分类标签, 从而提高低秩图像在图像分类任务上的性能. Zhang 等还考虑了基于广义柱体表面的变换不变低秩纹理模型^[10], 用于人造建筑物表面的纹理提取. 该方法还被进一步拓展应用到了建筑物几何模型^[11], 相机自动校准和镜头畸变自动校正^[12], 字符校正与识别^[13] 等. 在文献 [14] 中, Lin 等特别研究了 TILT 基于线性自适应惩罚交替方向法 (Linearized alternating direction method, LADM) 的快速算法, 把求解奇异值分解的计算速度提高了 5 倍以上. Zhang 等用它来解决影射变换下图像校正的问题^[15], 取得了一定的效果.

图 1 举例说明了使用 TILT 模型进行图像校正的效果. 其中, 图 1(a)~(d) 为 4 张包含低秩纹理

的原始图像, 图 1(e)~(h) 为经过 TILT 算法转换处理得到的低秩纹理图像. 现存的大多数方法都是将此问题看作是一个“转换问题”, 从图 1(c), (d), (g), (h) 的转换结果可以发现, 由于图像拍摄的背景区域相对复杂, 虽然主体物体具有很好的低秩特性, 但是背景区域像素组成的噪声矩阵占据像素比例较高, 不具有明显的低秩特性. 例如图 1(d), 当背景区域和主体物体是由不同纹理和结构组成时, 换句话说, 当背景和主体物体不能由一个较低的低秩矩阵进行线性表示时 (即图像矩阵是由两种不同特性的矩阵叠加而成), 转换的思想就无法将图像整体转换成较优的低秩图像. 为了跳出图像中不具有明显的低秩特性区域的干扰, 本文转换思路, 尝试直接对图像进行低秩整合, 提出了低秩纹理生成对抗网络 LR-GAN, 利用生成式思想去产生高质量的低秩纹理图像. 本文的主要贡献包括: 1) 对之前提出方法的局限性进行详细研究之后, 尝试使用生成式框架来缓解图像中不具明显低秩特性区域的校正结果不理想的问题, 从而实现了使用一种自动生成方法取代转换方法来解决低秩纹理问题; 2) 提出了低秩纹理生成对抗网络 (Low-rank generative adversarial network, LR-GAN), 使用 TILT 做引导, 利用传统的机器学习方法来辅助 LR-GAN 的学习过程, 整体而言该网络是一种非监督式学习方法, 这种端对端 (由图像生成图像) 的网络减少了整个网络的训练复杂度; 3) 为了得到超越传统方法效果的模型, 经过一定次数 TILT 的引导之后, 设计并加入低秩梯度滤波层, 该层在保证图像质量的前提下, 尽可能地确保低秩约束, 从而使最终生成的图像既具有高质量



图 1 利用 TILT 模型进行图像校正的例子

Fig. 1 Examples of image correction that using TILT

又具有更低的秩.

1 从转换到生成

从视觉任务的结果来看, 压缩感知和深度学习都让我们能够在像素级别上处理图像的全局信息. 无论是来自数学理论的压缩感知低维模型, 还是来自海量数据与训练网络的深度神经网络模型, 两者的结果都可以得到同样强大且可扩展的算法, 并且其流程与结构也有很多相似之处, 甚至可以用低维模型来解释深度神经网络理论的有效性. 由此可以看出, 压缩感知对于深度学习具有一定的理论指导意义. 众所周知, 深度学习对大量标签数据的依赖显而易见, 启发式的有监督学习也成为抑制深度学习纵向发展的一个潜在要素. 机器学习领域一直都在探索使用尽量少的标签数据, 希望实现从监督式学习到半监督式学习, 再到无监督式学习的转化. 如果能将压缩感知算法与神经网络、深度模型以及当今 GPU 强大的并行处理能力结合起来去解决相关的视觉问题, 特别是进行深度生成式模型的研究, 对于提高算法的运行效率至关重要, 计算成本也会大幅降低.

1.1 低秩表示

来自于现实世界中的图像作为像素的集合, 其本身就具有对称性与规律性. 事实上, 对于未加旋转的图像, 可以将其建模成是一个带噪声的低秩矩阵. 当图像由端正发生旋转时, 图像的原始对称性和自相似性就会被破坏, 即各行像素间的线性相关性被这些干扰量破坏, 因此矩阵的秩就会增加.

在 TILT 算法中, 将一个位于平面空间 \mathbf{R}^2 上的二维图像看作是一个矩阵 $\chi(x, y) \in \mathbf{R}^2$, 图像 $\bar{\chi}(x, y)$ 是转变视角后再观察低秩纹理图像 $\chi(x, y)$ 所获取的图像, 用数学公式可以表述为

$$\bar{\chi}(x, y) = \chi \circ \tau^{-1}(x, y) = \chi(\tau^{-1}(x, y)) \quad (1)$$

其中, $\tau: \mathbf{R}^2 \rightarrow \mathbf{R}^2$ 是一个希望得到的旋转变换函数. 显然, 经过 τ 转换过的图像 $\bar{\chi}(x, y)$ 已经不再是一个低秩矩阵. 由于从现实世界中观察得到的原始图像除去空间转换的干扰外, 还会受到许多未知噪声、遮挡与背景环境的影响, 为了能去除它们的影响, 将模型化的噪声变量 E 引入算法.

$$\bar{\chi} = (\chi + E) \circ \tau^{-1} \quad (2)$$

此处的噪声矩阵 E 通常是一个稀疏矩阵. 在低秩表示模型中, 该方法要解决的主要问题可以描述为: 给定一个经过旋转变化和噪声干扰的图像 $\bar{\chi} = (\chi + E) \circ \tau^{-1}$, 从中恢复出低秩纹理图像 χ , 确定其中的转换变量 τ 并且最小化噪声变量的干扰. 这个

过程可以导出如下形式的优化问题:

$$\begin{aligned} & \min_{\chi, E, \tau} (\text{rank}(\chi) + \gamma \|E\|_0) \\ & \text{s.t. } \bar{\chi} \circ \tau = \chi + E \end{aligned} \quad (3)$$

式中, $\text{rank}(\ast)$ 是秩函数, $\|E\|_0$ 表示噪声矩阵 E 的 L_0 范数, 用来表示噪声矩阵的稀疏度. 为了防止过拟合, 引入常量 $\gamma > 0$ 作为平衡纹理矩阵的秩与噪声矩阵稀疏度这两者的权重因子.

从本质上说, TILT 算法是一种无监督的学习算法. 它能够从相关联的高维数据中, 提取到数据间的内在同一性, 从而恢复其低维结构, 具有很高的应用潜力. 所以, 我们试图将低秩模型表示问题与深度学习算法相结合, 让低秩模型为神经网络的有效性提供可能的解释, 同时在实际应用中发挥出出色的效能.

1.2 生成对抗

Goodfellow 等^[16] 在 2014 年提出了 GAN 的模型与思想. 其中, 生成模型一般是指随机生成观测数据的模型, 是所有变量的全概率模型, 判别模型则是在给定观测变量值的前提下, 求目标变量的模型. 该思想将生成器 G 和判别器 D 分别看做两个参加极大-极小游戏的双方玩家. 双方不断相互对抗并且迭代优化的过程使得生成器和判别器的性能均不断提升, 当最终判别器的判别能力提升到一定程度, 并且无法正确判别数据来源时, 可以认为这个生成器已经学到了真实数据的分布^[17].

由于生成对抗网络能够学习图像、声音和数据中的丰富分布, 而这些分布通常因为其具有明确的相似性, 所以很难去建模. 与传统模型相比, GAN 模型生成数据的复杂度与维度线性相关. 因此, 生成对抗模型对于高维结构的表示与处理有很大意义. 而低秩表示模型不仅具有严格的理论基石, 还在数据恢复、特征提取、图像处理和计算机视觉等方面表现出优良特性. 近年来, 基于深度学习和稀疏表示学习的方法相互结合, 各自发挥其相应模块的优势, 形成了各种策略或特性下的深度神经网络模型, 应用于各种视觉任务. 本文将低秩转换理论与深度生成对抗网络模型相融合. 从转换思想过渡到生成思想, 从转换模型过渡到生成模型, 意在找到一种由原始图像直接生成得到低秩纹理图像的机制.

图 2 给出了低秩纹理生成对抗网络的整体结构示意图. LR-GAN 网络由两个主要模块组成: 生成器网络和加入 TILT 的判别器网络. 图 2(a) 是 LR-GAN 网络的整体算法流程; 图 2(b) 生成器网络负责生成原始图像的低秩纹理图像; 图 2(c) 判别器网络将生成器生成的图像和 TILT 算法转换之后的图像进行对抗学习; 图 2(d) 为在训练后期加入的低秩梯度过滤层.

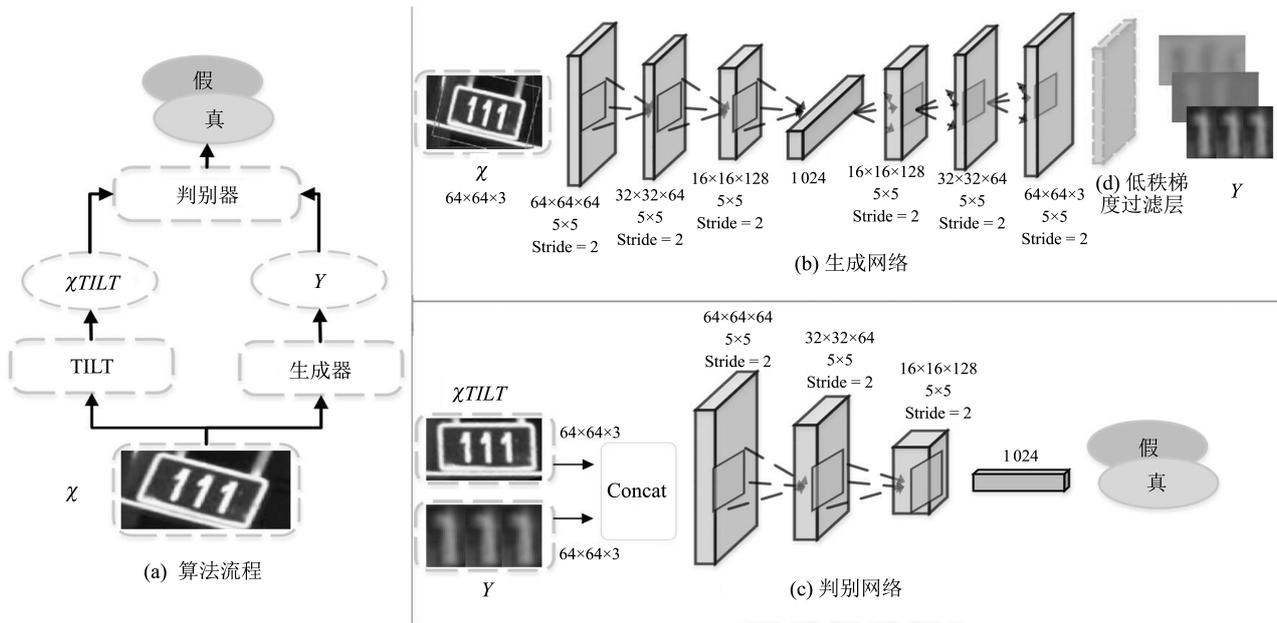


图2 LR-GAN 的网络结构示意图 ((a) LR-GAN 网络的整体算法流程; (b) 生成器网络负责生成原始图像的低秩纹理图像; (c) 判别器网络将生成器生成的图像和 TILT 算法转换之后的图像进行对抗学习; (d) 为在训练后期加入的低秩梯度过滤层)
 Fig. 2 The structure chart of LR-GAN ((a) The general framework of LR-GAN; (b) The Generator generates the low-rank texture image from the original image; (c) The Discriminator distinguishes between the generative image and the TILT image; (d) The layer of the low-rank gradient filter for training.)

假设 $\mathcal{X} = \{\chi^1, \chi^2, \dots, \chi^N\}$, $\chi \in \mathbf{R}^{m \times n \times c}$ 是 N 个 $m \times n$, c 个通道的原始图像. 在训练阶段, 通过 TILT 方法转换求得的样本被用于生成低秩纹理图像 $\chi_{TILT} = TILT(\chi) \in \mathbf{R}^{m \times n \times c}$. 用 G 和 D 分别作为生成器函数和判别器函数. G 的输入为原始图像数据 χ , D 的输入为 TILT 处理后的图像 χ_{TILT} 和生成器生成的图像 Y . $G(\chi)$ 是由 G 生成的逼近原始数据分布 χ_{TILT} 的样本. 那么, LR-GAN 的目标函数可以描述如下:

$$\min_G \max_D \left(\mathbb{E}_{TILT(\chi) \sim TILT(\mathcal{X})} [\log p(y = 1 | TILT(\chi), D)] + \mathbb{E}_{\chi \sim \mathcal{X}} (\chi) [\log (1 - p(y = 1 | G(\chi), D))] \right) \quad (4)$$

LR-GAN 在 TILT 的帮助下将会很容易在最优点上收敛. 经过迭代训练之, 输入原始图像 $\chi \in \mathbf{R}^{m \times n \times c}$, 将会生成低秩纹理图像 $Y = G(\chi)$.

2 低秩纹理生成对抗网络 (LR-GAN)

深度卷积生成对抗网络 (Deep convolutional generative adversarial networks, DCGAN)^[18] 是 GAN 目前一个很大的进步, 它利用深度卷积网络进行生成对抗网络的建模. 对于图像处理问题, DCGAN 相比于其他简单的全连接神经网络更具优势. 并且与其他非监督方法相比较, 其判别网络提取到

的图像特征更有效, 更适合用于图像识别与分类任务. 它的生成网络不但能够保持生成图像之间的连续性, 还可以学习到有意义的过滤层, 并且在多数情况下的训练比较稳定. DCGAN 通过融合全卷积神经网络对之前提到的 GAN 网络进行改进, 从而由输入图像生成输出图像. 受此启发, 本文将 TILT 算法与加入低秩过滤层的 DCGAN 网络结合起来, 将低秩约束用于 GAN 中的端到端的无监督学习框架, 提出了低秩纹理生成对抗网络 LR-GAN, 从原始图像中直接生成低秩纹理图像.

2.1 生成器网络架构

LR-GAN 中使用 $\chi \in \mathbf{R}^{m \times n \times c}$ 作为原始图像 $m = n = 64$, $c = 3$. 生成网络整体上可以被看作是一个由编码到解码的网络, 编码器模块由 4 个卷积层 (G-conv1 ~ G-conv4) 组成, 使用线性单元 $ReLU$ ^[19] 作为每一个卷积层的非线性激活函数. 为了防止采样的波动与模型的不稳定, 除 G-conv4 层之外, 对其余层均使用了批量规范化 (Batch-normalization, BN)^[20]. 思想是归一化当前层输入, 使它们的均值为 0, 方差为 1, 类似于归一化网络输入的方法. 好处在于可以加速收敛, 并且加入 BN 的卷积神经网络模型受权重初始化的影响非常小, 具有非常好的稳定性. 同时还有助于梯度传向更深层的网络, 对于提升卷积性能有很好效果. 解码器使用 3 个反卷积层 (G-deconv1 ~ G-deconv3), 与编码器

类似, $ReLU$ 被用作非线性激活函数. 在训练阶段, 使用 $Tanh^{[21]}$ 激活函数输出生成的图片, 同时也作为关联函数为生成网络与判别网络建立起一个联系.

2.2 判别器网络架构

$Data = \{\chi_i\}_{i=1}^K \in \mathcal{X}$ 是训练集 \mathcal{X} 中的 K 个训练样本, TILT 通过求解公式

$$\begin{aligned} \min_{\chi_{TILT}, E, \tau} & (\text{rank}(\chi_{TILT}) + \gamma \|E\|_0) \\ \text{s.t.} & \chi \circ \tau = \chi_{TILT} + E \end{aligned} \quad (5)$$

得到的低秩图像 χ_{TILT} 用来在起初的训练过程中起到加速收敛的作用.

整体而言, 本方法是一种非监督式的低秩纹理生成方法, 训练时不使用任何标签数据. 判别时将 (χ_{TILT}, Y) 作为一组低秩对抗对, 同时加入到具有 3 个卷积层 (D-conv1 ~ D-conv3) 的判别器中, 并且在判别器网络的每一个卷积层中使用修正线性单元 $ReLU$ 的改进版 $LeakyReLU^{[22]}$ 作为非线性激活函数. 为了保证判别器网络的鲁棒性与非饱和性, 在判别器中使用更为光滑的非饱和梯度的最小二乘损失函数^[23].

$$\begin{aligned} \min_D L(D) = & \\ & \frac{1}{2} E_{TILT(\chi) \sim TILT(\mathcal{X})} \left[(D(\chi_{TILT}) - 1)^2 \right] + \\ & \frac{1}{2} E_{\chi \sim \mathcal{X}} \left[D(Y)^2 \right] \end{aligned} \quad (6)$$

同样, 将此损失函数应用到生成器中

$$\min_G L(G) = \frac{1}{2} E_{\chi \sim \mathcal{X}} \left[(D(Y) - 1)^2 \right] \quad (7)$$

3 训练与优化

网络的训练整体可分为两个阶段: 第一个阶段是使用 TILT 的训练阶段; 第二个阶段是同时使用了 TILT 和低秩梯度滤波层的联合微调阶段.

第一个阶段的学习目标是生成与 χ_{TILT} 具有高相似度图像的同时, 确保该图像具有较低的低秩性. 为了达到这个目标, 生成器应具有低秩约束的损失函数为

$$\min_G L(G) = \frac{1}{2} E_{\chi \sim \mathcal{X}} \left[(D(Y) - 1)^2 \right] + \lambda (\|\chi\|_*) \quad (8)$$

式 (8) 中, 用核范数 $\|\cdot\|_*$ 来逼近秩函数从而保证低秩约束; λ 是用来平衡损失函数与规则项的超参数, 它使整个网络模型既可以充分学习原始图像的分布, 又具有低秩约束特性, 从而得到更好的生成效果.

在第一个阶段, 由于图像的低秩生成方向不明确, 所以网络训练的前期阶段较难收敛, 而 TILT 的加入给前期训练指明了一个低秩收敛方向. 由于式 (8) 很难直接求解 (NP 难问题), 并且经过 TILT 引导生成的低秩图像也并不完美. 为了进一步逼近其最优解, 设计了加入低秩梯度滤波层的生成网络来求解这一问题, 在与原始图像高度相似的同时保证了生成图像的低秩性. 由此, 网络的第二个阶段是同时使用了 TILT 和低秩梯度滤波层的联合微调阶段.

3.1 低秩梯度滤波层的训练

可微的判别器网络在做了判别并得到误差梯度之后, 会将它的梯度回传给生成网络和判别网络. 假设 w_t 是网络权重, g_t 是第 t 次迭代后反向传播的梯度映射, 这里使用一种过滤梯度的策略以确保低秩约束.

$$\|\text{rank}(w_t \chi) - \text{rank}(w_{t+1} \chi)\| < \delta \quad (9)$$

值得注意的是, 为了平衡高相似度和低秩这两个目标, 在最初对 G 和 D 进行网络的训练时, 并没有加入低秩梯度滤波层.

3.2 权重的更新

经过多次迭代之后, 生成的图像与 χ_{TILT} 有极高相似度. 为了达到更低的秩, 低秩过滤层使用特殊的梯度过滤策略来过滤后向传播的梯度. 在随机梯度下降 (Stochastic gradient descent, SGD)^[24] 权值更新策略的基础上, 使用

$$w_{t+1} = \begin{cases} w_t - \alpha g_t, & \|\text{rank}(w_t \chi) - \text{rank}(w_{t+1} \chi)\| < \delta \\ w_t - \beta \alpha g_t, & \|\text{rank}(w_t \chi) - \text{rank}(w_{t+1} \chi)\| \geq \delta \end{cases} \quad (10)$$

来过滤梯度. 式 (10) 中, α 为学习率, β 为用来惩罚 g_t 的惩罚因子. 在式 (9) 中, 如果 g_t 的更新具有较大的波动, δ 可以确保其在可控的范围内小幅波动. 这种策略使生成图像的秩在升高的方向上有明显的衰减, 而在图像秩降低的方向上有更大的梯度秩, 确保了低秩相对于高秩所获得的收益更大, 从而达到低秩约束的效果.

每个原始图像在迭代过程中都具有两个收敛方向. TILT 过程为网络前期的训练选择出了一个方向, 并且使整个网络处在一种非监督学习的状态. 图 3 所示的手风琴图像中, 最低秩有两种情况, 分别是横放与竖放. 在经过更多次迭代之后, 随着低秩梯度滤波层的加入, 使生成图像的秩向着更低的方向进行收敛, 从而能达到更为理想的低秩状态.

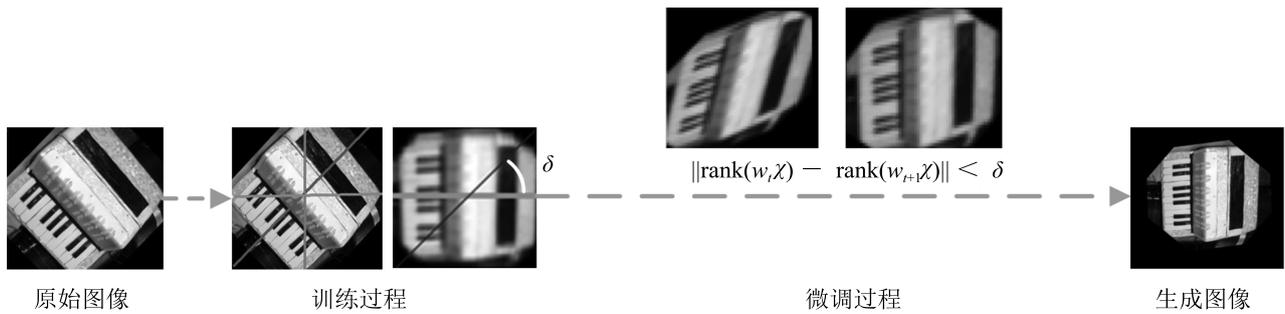


图 3 网络的训练与微调

Fig. 3 Training and fine-tuning

4 实验结果与分析

实验部分将通过定性与定量的结果来验证该模型的性能. 定性实验的第一个目标是研究低秩梯度滤波层的有效性, 第二个目标是评估 LR-GAN 网络从原始图像生成低秩图像的质量. 为了定量评价 LR-GAN 算法的生成质量, 将其应用于有监督的数据集上, 并将生成图像用于分类识别任务, 使用 Alex-net 网络对算法生成的图像进行分类评估. 整个网络框架在 NVIDIA GeForce GTX 1080 × 2 的 Tensorflow 上实现. 本文使用 MNIST, SVHN 和 FG-NET 这三个公开数据集来测试低秩纹理生成对抗网络模型. 在参数设置方面, 所有的实验中均设置 $\beta = 0.1$, $\delta = 4$. 为了避免单个样本在训练时导致较大波动, 生成数据的采样和优化都通过小批量随机梯度下降法 (Mini-batch stochastic gradient descent) 进行. 所有权重的初始化都服从均值为 0, 方差为 0.02 的正态分布. LeakyReLU 中, 模型的 leak 斜率设置为 0.2.

4.1 MNIST 数字校正

为了评估算法的有效性, 首先在较为简单与规则的图像上进行实验. MNIST 数据集^[25] 是用于评估机器学习模型的基准数据集, 共包含 250 个人的 7 万多张手写数字的图像. 这里使用文献 [26] 中的方法, 将 MNIST 数据集中的手写数字图像都进行一定程度的扭曲变形. 用随机的 $-90^\circ \sim +90^\circ$ 之间的角度对数字图像进行旋转变换. 图 4 是 MNIST 数据集上的生成过程. 图 4(a) 是形变的手写数字图像; 图 4(b) 迭代过程中生成模型生成的手写数字; 图 4(c) 最终生成的校正之后的手写数字图像. 在迭代过程中观察生成图像的演变过程, 从图 4(b) 的生成过程可以看出, 这时数字的角度发生初步变化, 数字结构也已经慢慢形成, 但是还有一些数字仍然存在欠缺. 图 4(c) 为完成所有训练之后, 生成模型在最后一轮生成的图像. 这时的生成的数字已经具有较好的质量与较强的辨识度, 秩值更低, 且与原始数

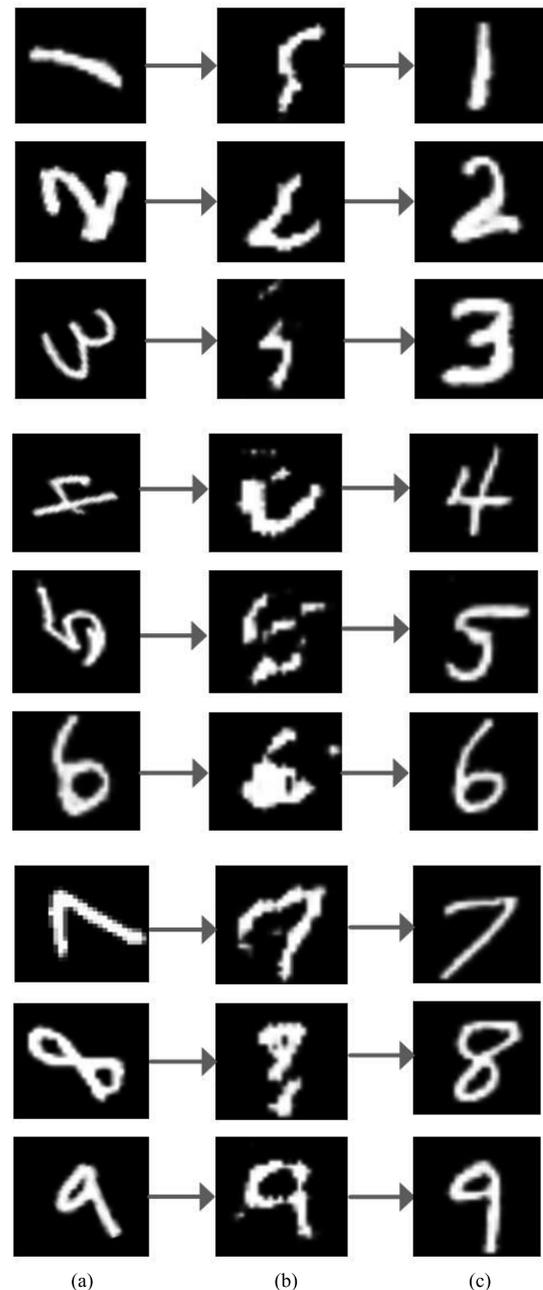


图 4 MNIST 数据集上的生成过程

Fig. 4 The generative process on MNIST dataset

字图像保持着高度的相似。模型在训练过程中的 loss 曲线如图 5 所示, 上方生成模型和下方判别模型的损失函数都在逐渐收敛。尤其是在加入低秩梯度过滤层之后, 损失值的变化幅度很小。为了验证低秩整合层的作用, 并表现出秩的波动, 将图片调整为 70×70 的尺寸。从图 6 可以看出, 经过 500 次迭代, 将低秩梯度过滤层加入到网络中之后, 图像的平均秩有了明显下降。表 1 统计了 MNIST 上的平均秩结果。表 1 中, 在扭曲变形之后的手写数字数据集上, 相比于 TILT 方法, LR-GAN 网络可以达到更低的平均秩值。从模型的整个训练过程可以看出, 一开始生成数字的质量并不理想, 但其扭转方向的趋势比较理想。随着迭代次数的增加, 生成的数字质量越来越好, 模型的低秩特性也充分的体现出来, 直观地反映了两个网络之间的对抗学习关系。

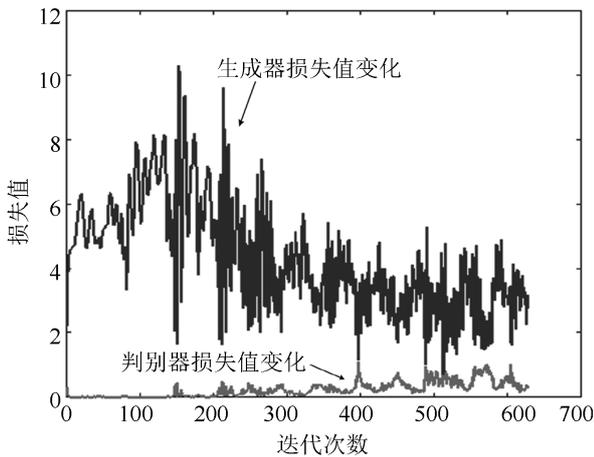


图 5 MNIST 数据集迭代过程中生成器与判别器的损失值变化

Fig. 5 The loss of both the generator and the discriminator on MNIST during the iterations

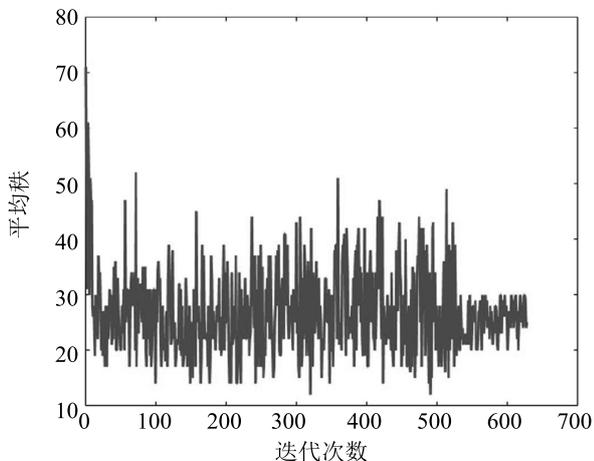


图 6 MNIST 数据集上生成器迭代过程中图像秩的变化
Fig. 6 The changes of the rank during the generator iterations on MNIST

4.2 SVHN 数字标牌校正

实验在街景门牌数据集 (The street view house numbers dataset, SVHN)^[27] 上进行。SVHN 数据集包含 20 多万张由谷歌街景车拍摄的房屋门牌号的 RGB 图像。每幅图像中均包含有 1~3 位数字。由于门牌的形状与风格各异, 再加上门牌上的数字空间排列变化较大, 因此, SVHN 数据集中的图像相比于 MNIST 数据集中的单个数字图像多了一些具有近似低秩特性的区域。实验中使用 48×48 大小的图像, 32 尺寸的 batch, 共进行了 5000 次迭代训练, 每 10 次生成一张图片并测试其秩的大小。图 7 是 SVHN 数据集上的生成过程。图 7(a) 原始的门牌号图像; 图 7(b) 迭代过程中生成的初具辨识度的数字标牌图像; 图 7(c) 最终迭代生成的数字标牌图像。从图 7(b) 所示的生成过程中可以看出, 标牌上数字的大体结构已经形成, 但是能够表征数字标牌细节的特征还没有大量出现。随着迭代次数增长, 图 7(c) 中数字标牌的辨识度越来越高, 与原始图像的细节相似度越来越好。SVHN 数据集迭代过程中生成器与判别器的损失值如图 8 所示。判别器和生成器的损失函数都呈现收敛的趋势, 并且可以看出两个模型的损失函数曲线存在强相关的关系, 在对抗过程中此消彼长。在网络的第 3500 次迭代中加入了低秩梯度过滤层, 图 9 展示出图像秩的变化, 能够很好地反映低秩梯度过滤层对于低秩的约束效果。同时, 从表 1 中 SVHN 上的平均秩结果可以看出, 在 SVHN 数据集上, 本方法得到的图像更具低秩性。本实验说明 LR-GAN 在具有近似低秩特性的图像背景下也可以获得较好的生成效果。

表 1 MNIST 与 SVHN 上的平均秩结果
Table 1 The average rank on MNIST and SVHN datasets

| method | MNIST | SVHN |
|-----------------|-----------|-----------|
| TILT | 31 | 46 |
| LR-GAN | 29 | 40 |
| LR-GAN + Filter | 27 | 37 |

4.3 人脸校正

人脸校正正在人脸检测与识别任务中起到至关重要的作用。为了验证文中的方法在人脸校正问题上的效果, 本部分使用 FG-NET 数据集^[28]。该数据集由 82 个人的 1000 多幅正面人脸图像组成。其中的人脸图像具有较多不规则细节, 纹理相对复杂, 且主体人脸的纹理结构特性与几何对称特性均不明显。实验结果如图 10 所示, 图 10(a) 原始人脸图像; 图

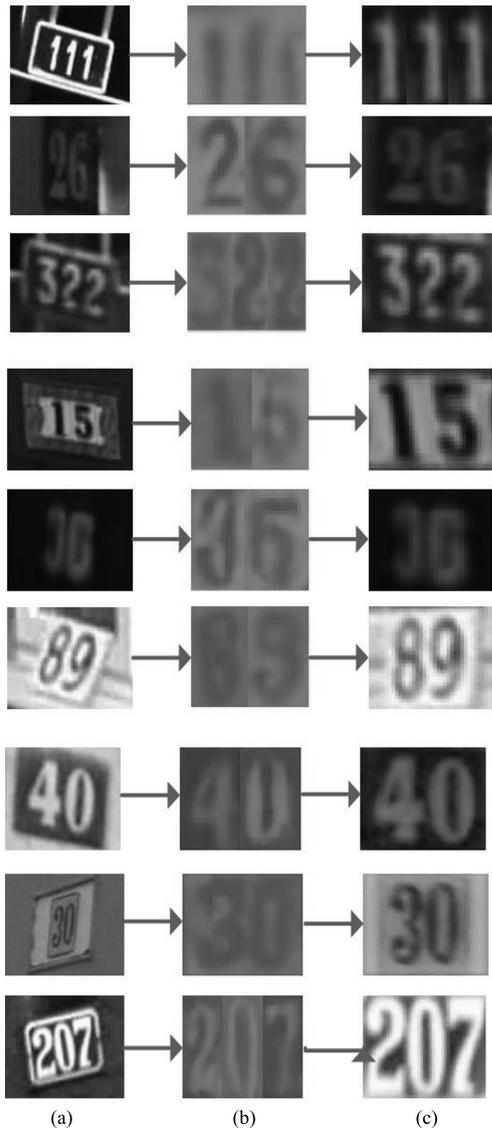


图 7 图为 SVHN 数据集上的生成过程
Fig.7 The generative process on SVHN dataset

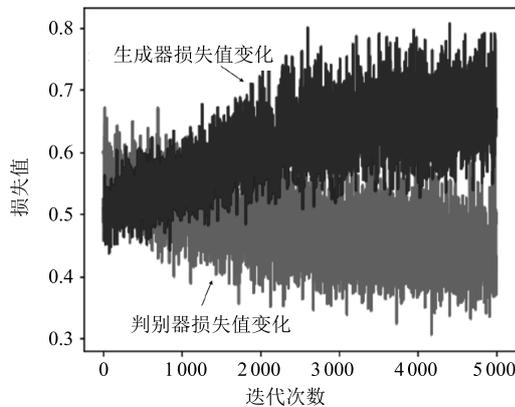


图 8 SVHN 数据集迭代过程中生成器与判别器的损失值变化
Fig.8 The loss of both the generator and the discriminator on SVHN during the iterations

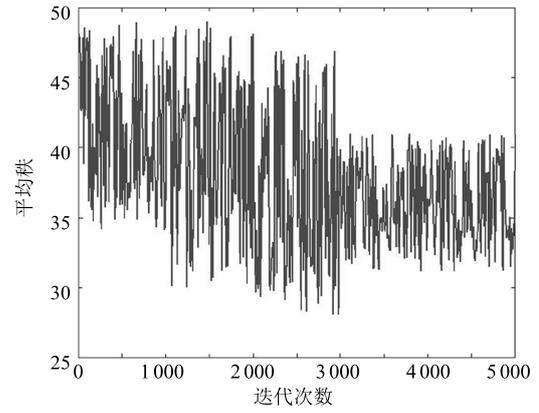


图 9 SVHN 数据集上生成器迭代过程中图像秩的变化
Fig.9 The changes of the rank during the generator iterations on SVHN

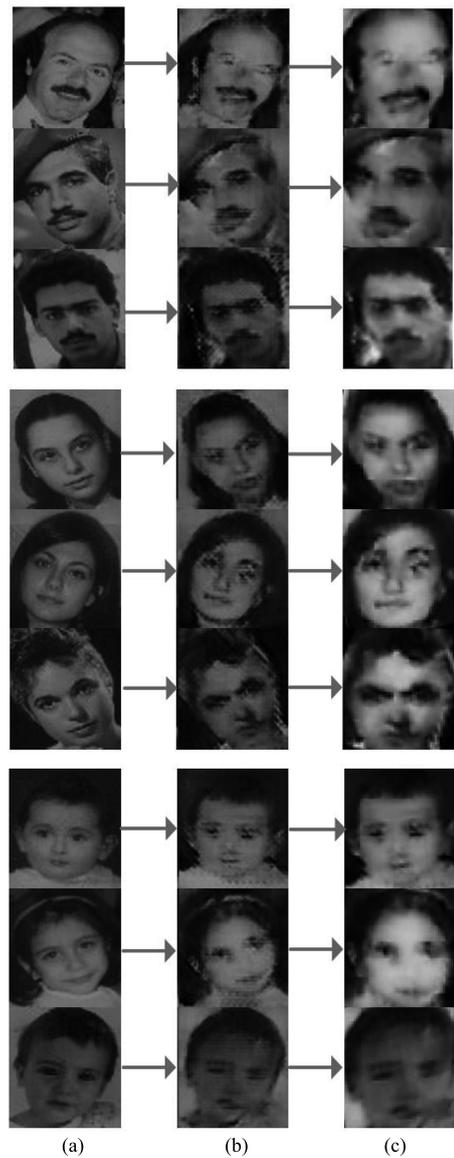


图 10 FG-NET 数据集上的生成过程
Fig.10 The generative process on FG-NET dataset

10(b) 迭代过程中生成的人脸图像, 其视觉质量还有待提高; 图 10(c) 最终生成的校正后的人脸图像. 虽然人脸图像的低秩特性不太明显, 但其整体纹理结构的方向特征比较明确, 因此 LR-GAN 对于倾斜人脸的校正结果比较成功. 生成的人脸图像可以慢慢从歪斜到摆正, 并且校正后人脸图像也高度类似于原始人脸图像. 然而由于生成对抗网络的训练较不稳定, LR-GAN 在人脸图像的生成质量上还有较大的提升空间. 本部分实验将该数据集里的人脸图像大小调整为 64 像素 \times 64 像素, 并且使用 512 尺寸的 batch. LR-GAN 在该数据集上表现出了快速的收敛, 经过 200 次迭代后, 网络整体已经达到收敛状态. 如图 11 所示, 在相对复杂的人脸图像上, 整个网络在 TILT 的引导下快速的收敛. 由于使用了较大的 batch 块, 并且每次迭代都会生成一张图像然后测量其秩的大小, 在 100 次迭代之后加入低秩梯度过滤层, 从图 12 中可以看出, 在前 100 次迭代中人脸图像的生成过程, TILT 在该数据集上的引导作用已经微乎其微, 图像的秩基本没有较大的变化, 而在加入低秩过滤层之后, 经过约 200 次迭代, 人脸图像的秩整体下降了约 0~3 的范围.

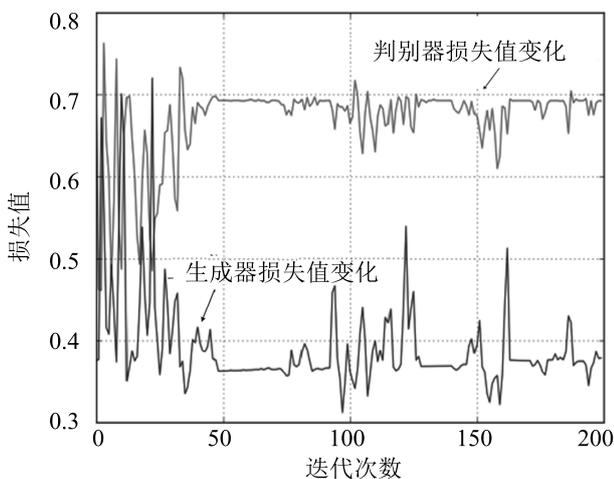


图 11 FG-NET 数据集迭代过程中生成器与判别器的损失值变化

Fig. 11 The loss of both the generator and the discriminator on FG-NET during the iterations

4.4 分类评估

为了定量评估生成图片的质量与生成网络的性能, 使用图像分类模型 AlexNet 网络分别对形变的 MNIST 数据集和 SVHN 数据集上生成得到的图像进行分类识别. 用不做任何处理的数据集 (no) 和使用 TILT 进行低秩转换的图像 (TILT) 同时和文中所提的方法进行对比实验. 识别效果如表 2 和表 3 所示. 从表 2 可以看出, 对于做了形变处理

的 MNIST 数据集, 经过低秩整合后的图像在分类任务上的性能得到大幅的提升, 其中 TILT 提升了 6.02%, 本文的方法提升了 7.96%. 在 SVHN 数据集上, 虽然背景区域的干扰较强, 但是相对于 TILT, 文中所提的方法依然得到了 97.56% 的效果.

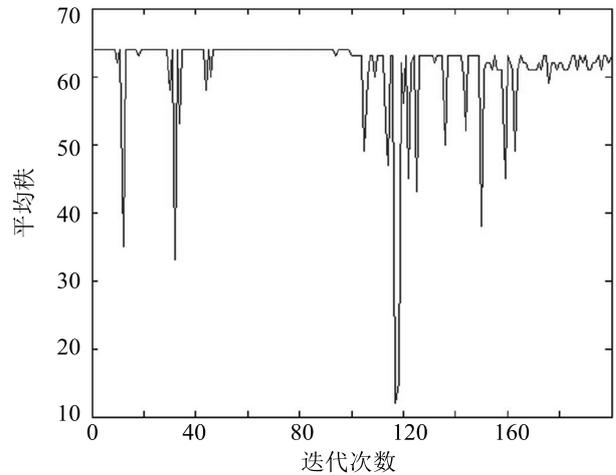


图 12 FG-NET 数据集上生成器迭代过程中图像秩的变化
Fig. 12 The changes of the rank during the generator iterations on FG-NET

表 2 在形变的 MNIST 上的分类识别效果
Table 2 The classification performance on distorted MNIST

| database | method | mAp |
|----------|--------|---------------|
| MNIST | no | 0.5701 |
| | TILT | 0.6303 |
| | ours | 0.6497 |

表 3 在 SVHN 上的分类识别效果
Table 3 The classification performance on SVHN

| database | method | mAp |
|----------|--------|---------------|
| SVHN | no | 0.9609 |
| | TILT | 0.9701 |
| | ours | 0.9756 |

从以上实验可以看出, 将深度生成式神经网络与稀疏表示相结合得到的 LR-GAN 算法, 可以使学习到的特征有利于提升网络模型的整体泛化能力, 同时丰富了网络训练的各种技巧与策略. 使其能够直接迁移应用于许多其他图像校正的实例中, 并且可以结合问题特性适当地引入低秩约束来增强其低秩性, 从而达到预期的生成效果.

5 结论与展望

本文提出了一种非监督式的由图像生成图像的
低秩纹理图像生成对抗网络 (LR-GAN). 该算法将
低秩表示问题与深度生成式对抗网络相结合, 建立
起一种由原始图像直接生成得到低秩纹理图像的
机制. 在训练早期, TILT 的加入使整个模型成为非
监督模型, 并且达到加速训练与引导低秩生成方
向的效果. 经过一定次数的迭代之后, 低秩梯度过
滤层的加入提高了生成器的低秩生成能力, 而且
在保证生成质量的同时, 增强了低秩约束. 实验
结果表明, LR-GAN 网络在使用公共数据集下的
平均秩变化优于现有的 TILT 方法. 该算法能够
有效处理不具明显低秩纹理特性图像的校正问
题, 校正之后的图像具有较好的识别效果. 在未
来研究中, 可以结合不同图像问题的特性, 适
当地引入低秩约束来增强其低秩性, 从而达到
更理想的生成效果. 另外, 对于该生成式网络
自身的不稳定性以及泛化能力等将做相关的探
讨与研究, 进一步完善算法.

同时, LR-GAN 作为 GAN 的衍生模型, 除了
是一种有效的生成式模型, 也是一个比较简单
的平行系统^[29]. 并且, 其作为一种基于深度学
习的底层视觉问题, 我们有理由认为它将能够
充分融入到平行视觉^[30] 的研究体系中, 助力
于提升视觉模型的泛化能力. 希望本文的工作
有助于启发对深度学习与压缩感知螺旋式发
展的持续探索.

References

- Li Shu-Tao, Wei Dan. A survey on compressive sensing. *Acta Automatica Sinica*, 2009, **35**(11): 1369–1377
(李树涛, 魏丹. 压缩传感综述. 自动化学报, 2009, **35**(11): 1369–1377)
- Peng Yi-Gang, Suo Jin-Li, Dai Qiong-Hai, Xu Wen-Li. From compressed sensing to low-rank matrix recovery: theory and applications. *Acta Automatica Sinica*, 2013, **39**(7): 981–994
(彭义刚, 索津莉, 戴琼海, 徐文立. 从压缩传感到低秩矩阵恢复: 理论与应用. 自动化学报, 2013, **39**(7): 981–994)
- Yang S, Wei E L, Guan R M, Zhang X F, Qin J, Wang Y Y. Triangle chain codes for image matching. *Neurocomputing*, 2013, **120**: 268–276
- Brown M, Lowe D G. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 2007, **74**(1): 59–73
- Han J G, Farin D, de With P. A mixed-reality system for broadcasting sports video to mobile devices. *IEEE Multi-Media*, 2011, **18**(2): 72–84
- Cheng L, Gong J Y, Li M C, Liu Y X. 3D building model reconstruction from multi-view aerial imagery and lidar data. *Photogrammetric Engineering and Remote Sensing*, 2011, **77**(2): 125–139
- Zhang Z D, Liang X, Ganesh A, Ma Y. Tilt: transform invariant low-rank textures. In: Proceedings of the 10th Asian Conference on Computer Vision-ACCV 2010. Berlin Heidelberg, Germany: Springer, 2011. 314–328
- Zhang Z D, Ganesh A, Liang X, Ma Y. Tilt: transform invariant low-rank textures. *International Journal of Computer Vision*, 2012, **99**(1): 1–24
- Zhang Y, Jiang Z L, Davis L S. Learning structured low-rank representations for image classification. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA: IEEE, 2013. 676–683
- Zhang Z D, Liang X, Ma Y. Unwrapping low-rank textures on generalized cylindrical surfaces. In: Proceedings of the 2001 International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 1347–1354
- Mobahi H, Zhou Z H, Yang A Y, Ma Y. Holistic 3D reconstruction of urban structures from low-rank textures. In: Proceedings of the 2011 International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain: IEEE, 2011. 593–600
- Zhang Z D, Matsushita Y, Ma Y. Camera calibration with lens distortion from low-rank textures. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 2321–2328
- Zhang X, Lin Z C, Sun F C, Ma Y. Rectification of optical characters as transform invariant low-rank textures. In: Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR). Washington, DC, USA: IEEE, 2013. 393–397
- Lin Z C, Liu R S, Su Z X. Linearized alternating direction method with adaptive penalty for low-rank representation. In: Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain: ACM, 2011. 612–620
- Zhang Q, Li Y J, Blum R S, Xiang P. Matching of images with projective distortion using transform invariant low-rank textures. *Journal of Visual Communication and Image Representation*, 2016, **38**: 602–613
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: ACM, 2014. 2672–2680
- Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: the state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. 自动化学报, 2017, **43**(3): 321–332)
- Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the 2016 International Conference on Learning Representation (ICLR). San Juan, Puerto Rico: 2016. 3, 5, 6

- 19 Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: ACM, 2010. 807–814
- 20 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015. 448–456
- 21 Fan E G. Extended tanh-function method and its applications to nonlinear equations. *Physics Letters A*, 2000, **277**(4–5): 212–218
- 22 Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA: PMLR, 2013.
- 23 Mao X D, Li Q, Xie H R, Lau R Y K, Wang Z, Smolley S P. Least squares generative adversarial networks. In: Proceedings of the 2017 International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 2813–2821
- 24 Zhao S Y, Li W J. Fast asynchronous parallel stochastic gradient descent: a lock-free approach with convergence guarantee. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, Arizona: AAAI, 2016. 2379–2385
- 25 LeCun Y, Cortes C, Burges C J C. The MNIST database of handwritten digits [Online], available: <http://yann.lecun.com/exdb/mnist/>, July 12, 2016
- 26 Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. Spatial transformer networks. In: Proceedings of the 29th Annual Conference on Neural Information Processing Systems. Montreal, Canada: NIPS, 2015. 2017–2025
- 27 Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng A Y. Reading digits in natural images with unsupervised feature learning. In: Proceedings of the 2011 NIPS Workshop on Deep Learning and Unsupervised Feature Learning. Granada, Spain: NIPS, 2011. 2: 5–13
- 28 Panis G, Lanitis A, Tsapatsoulis N, Cootes T F. Overview of research on facial ageing using the FG-NET ageing database. *IET Biometrics*, 2016, **5**(2): 37–46
- 29 Li Li, Lin Yi-Lun, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel learning — a new framework for machine learning. *Acta Automatica Sinica*, 2017, **43**(1): 1–8 (李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 平行学习 — 机器学习的一个新型理论框架. *自动化学报*, 2017, **43**(1): 1–8)
- 30 Wang Kun-Feng, Gou Chao, Wang Fei-Yue. Parallel vision: an ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 2016, **42**(10): 1490–1500 (王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. *自动化学报*, 2016, **42**(10): 1490–1500)



赵树阳 北京理工大学计算机学院硕士研究生. 主要研究方向为计算机视觉, 图像处理与机器智能.

E-mail: zsyprich@bit.edu.cn

(**ZHAO Shu-Yang** Master student at the School of Computer Science and Technology, Beijing Institute of Technology. Her research interest covers

computer vision, image processing, and machine intelligence.)



李建武 博士, 北京理工大学计算机学院副教授. 主要研究方向为计算机视觉, 图像处理, 超分辨率图像重建技术. 本文通信作者. E-mail: ljw@bit.edu.cn

(**LI Jian-Wu** Ph.D., associate professor at the School of Computer Science and Technology, Beijing Institute of Technology. His research interest

covers computer vision, image processing, and super-resolution image reconstruction. Corresponding author of this paper.)