基于生成对抗网络的多视图学习与重构算法

孙 亮 [1] 韩毓璇 [1] 康文婧 [1] 葛宏伟 [1]

摘 要 同一事物通常需要从不同角度进行表达. 然而, 现实应用经常引出复杂的场景, 导致完整视图数据很难获得. 因此研究如何构建事物的完整视图具有重要意义. 本文提出一种基于生成对抗网络 (Generative adversarial networks, GAN) 的多视图学习与重构算法, 利用已知单一视图, 通过生成式方法构建其他视图. 为构建多视图通用的表征, 提出新型表征学习算法, 使得同一实例的任意视图都能映射至相同的表征向量, 并保证其包含实例的重构信息. 为构建给定事物的多种视图, 提出基于生成对抗网络的重构算法, 在生成模型中加入表征信息, 保证了生成视图数据与源视图相匹配. 所提出的算法的优势在于避免了不同视图间的直接映射, 解决了训练数据视图不完整问题, 以及构造视图与已知视图正确对应问题. 在手写体数字数据集 MNIST, 街景数字数据集 SVHN 和人脸数据集 CelebA 上的模拟实验结果表明, 所提出的算法具有很好的重构性能.

关键词 多视图重构, 条件生成对抗网络, 多视图表征学习, 生成模型

引用格式 孙亮, 韩毓璇, 康文婧, 葛宏伟. 基于生成对抗网络的多视图学习与重构算法. 自动化学报, 2018, **44**(5): 819−828

DOI 10.16383/j.aas.2018.c170496

Multi-view Learning and Reconstruction Algorithms via Generative Adversarial Networks

SUN Liang[1] HAN Yu-Xuan[1] KANG Wen-Jing[1] GE Hong-Wei[1]

Abstract Generally, objects often require to represent in different views. However, real-world applications in complex scenarios can hardly have complete views of a given object. In this paper, we propose generative adversarial network (GAN) based multi-view learning and reconstruction algorithms. A novel representation learning algorithm is proposed, which guarantees different views of the same object are mapped to the same representation. Meanwhile, the algorithm guarantees the representation carries enough reconstructed information. To construct multi-views of a given object, a generative adversarial network based reconstruction algorithm is proposed, which includes the representation information in the generation and discrimination models to guarantee the constructed views perfectly map the source view. The merits of the proposed algorithms lie in the fact that they avoid direct mapping among different views, and can solve the problem of missing views in training data and the problem of mapping between constructed views and the source views. Simulated experiments on handwritten digit dataset (MNIST), street view house numbers dataset (SVHN) and CelebFaces attributes dataset (CelebA) indicate that the proposed algorithms yield satisfactory reconstruction performances.

Key words Multi-view reconstruction, conditional generative adversarial networks (CGAN), multi-view representation learning, generative models

Citation Sun Liang, Han Yu-Xuan, Kang Wen-Jing, Ge Hong-Wei. Multi-view learning and reconstruction algorithms via generative adversarial networks. *Acta Automatica Sinica*, 2018, **44**(5): 819−828

实际应用问题中, 同一事物通常可以通过不同途径从不同角度进行表达. 例如: 多媒体记录可以通过视频描述, 也可以通过音频描述; 网页记录可以通过其本身的信息描述, 也可以通过超链接包含的信息描述; 同一语义对象, 可以用多种语言描述. 此外, 同一事物由于数据采集方法不同, 也可以有不同的表达方法. 例如: 人脸识别问题中, 人脸数据可以采集成二维, 也可以采集成三维; 指纹识别问题中, 同一指纹可以通过不同采集器采集出不同的印痕. 上述每一类型数据称为一个特定视图, 多类型数据的总体称为多视图数据. 针对多视图数据的分析研究, 已经引起机器学习研究者的关注[1−4]. 按不同任务, 已有方法可分为多视图子空间学习[5−6]、多视图字典学习[7−8]、多视图度量学习[9] 等. 完成这些任务的重要工作是获得视图间的匹配关系, 可以通过协

收稿日期 2017-09-08 录用日期 2017-12-23
Manuscript received September 8, 2017; accepted December 23, 2017
国家自然科学基金 (61402076, 61572104, 61103146), 吉林大学符号计算与知识工程教育部重点实验室项目 (93K172017K03), 中央高校基本科研业务项目 (DUT17JC04) 资助
Supported by National Natural Science Foundation of China (61402076, 61572104, 61103146), Project of Key Laboratory of Symbolic Computation and Knowledge Engineering of Jilin University (93K172017K03), and Fundamental Research Funds for Central Universities (DUT17JC04)
本文责任编委 王坤峰
Recommended by Associate Editor WANG Kun-Feng
1. 大连理工大学计算机科学与技术学院 大连 116023
1. College of Computer Science and Technology, Dalian University of Technology, Dalian 116023

同训练[10−11]、协同映射[12−13]、信息传播[14] 等方法实现. 在实现过程中, 通常要求每个实例的所有视图都是完整的. 然而, 现实问题中数据通常独立地收集、处理和存储, 受环境因素的影响, 给定一实例, 通常很难获得其所有视图的数据. 因此, 利用已掌握的单一视图, 通过生成式方法获得其他视图数据, 能够更全面地认识事物, 对其进行更准确的表达 [4], 具有重要的意义.

给定单一视图, 首先需要解决的问题是构建它的恰当表示, 即表征. 传统的手动提取特征方法需要大量的人力并且依赖于专业知识, 同时还不便于推广. 随着深度学习技术的发展, 通过深度神经网络 (Deep neural networks, DNN) 学习事物的表征获得了成功[15−17], 它允许算法使用特征的同时也提取特征, 避免了手动提取特征的繁琐, 能够获得单一视图恰当的表征[18]. 通过单一视图的表征构建完整视图, 表征中不仅需要包含其本身的信息, 而且这些信息能够用来构建其他视图. 为解决该问题, 已有方法主要在表征空间通过最大化不同视图间相互关系[1]、最小化不同视图间差异[19]、为差异添加惩罚因子[20−21]、典型相关分析[22] 等方法实现. 然而, 由于现实世界数据的复杂性, 如何构建适用于多视图的有效表征, 仍然是需要研究和解决的问题.

利用单一视图的表征, 通过生成式方法构建完整视图依赖于生成模型的好坏, 需要根据学习而来的模型生成新样本. 传统的生成式方法包括极大似然估计法[23]、近似法[24]、马尔科夫链法[25] 等. 与此同时, 基于 DNN 构建的生成式模型也获得了成功, 典型的网络结构包括循环神经网络 (Recurrent neural networks, RNN)[26]、卷积神经网络 (Convolutional neural networks, CNN)[27]、变分自编码器 (Variation autoencoders, VAE)[28]、生成对抗网络 (Generative adversarial networks, GAN)[29−30] 等. 这些方法针对已掌握的数据进行分布假设和参数学习. 然而在实际应用过程中, 不同视图 (例如图像、视频、传感器等) 的数据数量巨大, 并且都非常复杂、冗余并且异构[31], 如何在生成模型中融入已有视图的表征信息, 仍然是需要研究和解决的问题.

本文的主要工作集中于利用已知单一视图, 通过生成式方法构建其他视图. 为构建适用于多视图的表征, 提出一种新型表征学习方法, 该方法通过 DNN 来实现. 首先, 对于每一视图, 分别搭建 DNN, 通过逐层转换与表达, 借助 DNN 的无限拟合能力将数据映射至特征空间. 通过构建并优化训练过程中的损失函数, 将同一实例的不同视图映射至相同或相近的表征向量. 在众多生成式模型中, 生成式对抗网络 (GAN) 在结构上受博弈论中二人零和博弈启发, 通过构建生成模型和判别模型捕捉真实数据样本的潜在分布并生成新的数据样本. 与其他生成式模型不同, GAN 避免了马尔科夫链式的学习机制, 使得真实数据样本概率密度不可计算时, 模型依然可以应用. 为在生成模型中融入已有视图的表征信息, 本文提出基于 GAN 的生成式模型. 对于每一视图, 分别搭建 GAN, 在生成模型和判别模型的输入端加入随机变量和原始数据及已有视图生成的表征信息, 使得生成模型能够生成与已有视图相对应的新视图数据. 综上所述, 本文的主要贡献包括: 1) 提出基于 DNN 的多视图表征学习方法, 对于同一实例, 将不同视图数据映射至相同或相近的表征向量, 避免了视图间的直接映射; 2) 对于每一视图, 分别搭建 DNN, 训练过程中将每一对视图的 DNN 组合训练, 不需要训练数据的完整视图, 解决了训练数据不完整问题; 3) 提出基于 GANs 的多视图数据生成方法, 将已知视图的表征向量加入生成模型和判别模型中, 解决了新视图数据与已知视图数据正确对应的问题.

本文章节安排如下: 第 1 节用数学模型描述要解决的多视图重构问题; 第 2 节提出基于 DNN 的多视图表征学习方法; 第 3 节提出基于 GANs 的多视图数据生成方法; 第 4 节通过手写体数字数据集 MNIST, 街景数字数据集 SVHN 和人脸数据集 CelebA 验证提出方法的有效性, 并与其他已有算法进行比较分析; 第 5 节总结全文, 并指出进一步的研究方向.

1 问题描述

假定 χ 为一组包含 n 个实例, v 个视图的实例集, 每一实例表示为 $x_i = (x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(v)})$, 其中 $x_i^{(k)} \in \mathbf{R}^{d_k}$ 表示第 i 个实例的第 k 个视图数据, d_k 为第 k 个视图的维度. 与此同时, 每一实例对应指示向量 $\boldsymbol{q}_i \in \mathbf{Q} = \{0,1\}^v$, $q_i^{(k)} = 1$ 表示视图数据 $x_i^{(k)} \in \mathbf{R}^{d_k}$ 可观测, $q_i^{(k)} = 0$ 表示 $x_i^{(k)} \in \mathbf{R}^{d_k}$ 不可观测.

本文工作的主要目标是通过一组训练实例 χ 构建生成模型, 给定任意测试实例的源视图 $x_j^{(s)}$ 预测其他视图, 使得生成模型获得的视图 $\hat{x}_j^{(t)}$ $(t \neq s)$ 接近真实视图 $x_j^{(t)}$, 即最大化条件概率 $P(\hat{x}_j^{(t)} = x_j^{(t)} | x_j^{(s)})$. 为表述方便, 记可观测的第 k 个视图为 $x^{(k)}$.

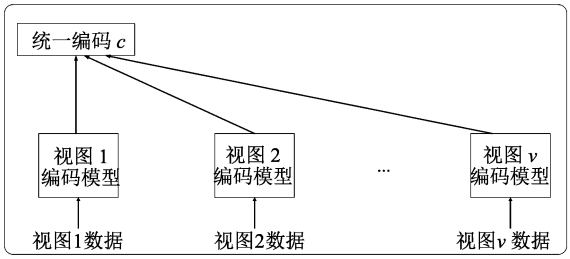
2 基于 DNN 的多视图表征学习

给定第 k 个视图数据 $x^{(k)}$, 通过构造 DNN 编码模型, 可以将其编码成低维向量 $\boldsymbol{c}^{(k)}$, 假设网络的映射函数为 $f^{(k)}(x^{(k)})$, 则 $\boldsymbol{c}^{(k)} = f^{(k)}(x^{(k)})$. 为所有视图分别构造编码模型, 可以得到 v 个 DNN. 这种表示不能获得多视图相同或相近的表征. 因此, 借助 DNN 能够逼近任意函数的能力, 将 $x^{(1)}$, $x^{(2)}$, $\cdots$,

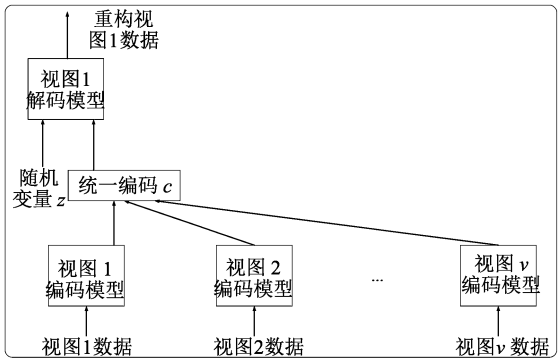
$x^{(v)}$ 映射至相同的表征空间, 如图 1 (a) 所示. 为了保证同一实例的不同视图映射至同一表征向量, 在网络训练过程中, 对任意一对视图 k 和 r, 最小化目标向量间的 JS 散度, 网络优化的目标函数定义为

$$\min_{\theta_1,\cdots,\theta_v}\sum_{k,r}\frac{1}{2}[KL(P(c|x^{(k)})\|\frac{P(c|x^{(k)})+P(c|x^{(r)})}{2})+$$
$$KL(P(c|x^{(r)})\|\frac{P(c|x^{(k)})+P(c|x^{(r)})}{2})] \quad (1)$$

其中, θ_1, θ_2, $\cdots$, θ_v 分别为 v 个 DNN 网络中的所有参数, $KL(P_1\|P_2)$ 表示分布函数 P_1 与 P_2 间的 KL 散度. 实际应用过程中, 为保证表征信息的紧凑性, 将 $\boldsymbol{c}$ 设置为较低维度.



(a) 多视图数据直接映射至同一表征向量
(a) Multi-view data are mapped to the same representative vector directly



(b) 构建解码模型, 使表征向量包含实例的重构信息
(b) Decoding model is built so that representative vector carries reconstruction information of the instance

图 1　多视图表征向量映射

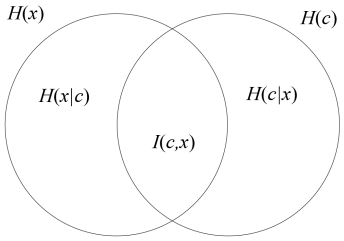
Fig. 1　Multi-view representative vector mapping

图 1 (a) 的网络结构保证了对于任意实例 x_i 的所有视图能够通过相应的神经网络映射至相同的表征向量 $\boldsymbol{c}_i$, 但不能保证表征向量 $\boldsymbol{c}_i$ 中包含实例 x_i 中的重构信息. 根据信息理论, 给定随机变量 x 包含的信息可以通过下式计算:

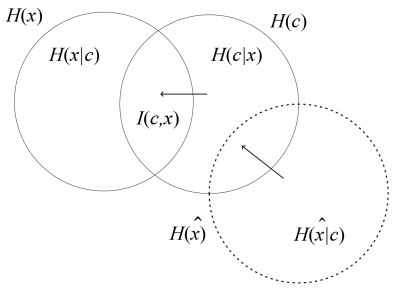
$$H(x)=\sum -P(x)\log P(x)=\mathrm{E}[-\log P(x)] \quad (2)$$

随机变量 x 与随机变量 y 之间的互信息 $I(x;y)$ 可以定义为随机变量 x 中包含随机变量 y 的信息量, 如图 2 (a) 所示, 可以通过下式计算:

$$I(x;c)=H(x)-H(x|c)=H(c)-H(c|x) \quad (3)$$



(a) 原始数视图数据 x 与表征向量 $\boldsymbol{c}$ 间的互信息
(a) Mutual information between original view data $\boldsymbol{x}$ and representative vector $\boldsymbol{c}$



(b) 通过表征向量 $\boldsymbol{c}$ 重构 $\hat{x}$ 使互信息 $I(x;c)$ 最大化
(b) Construct $\hat{x}$ using representative vector $\boldsymbol{c}$ so that mutual information $I(x;c)$ can be maximized

图 2　原始视图数据 x, 表征向量 $\boldsymbol{c}$, 重构视图数据 $\hat{x}$ 间的互信息示意图

Fig. 2　Schematic diagram of mutual information among original view data x, representative vector $\boldsymbol{c}$, reconstructed data $\hat{x}$

从图 2 (a) 可以看出, 为最大化 x 与 $\boldsymbol{c}$ 之间的互信息 $I(x;c)$, 可以拟合 $H(x|c)$ 与 $H(c|x)$, 其中 $H(c|x)$ 可以通过视图的 DNN 编码模型进行优化调整. 然而, $H(x|c)$ 很难直接计算. 为此本文提出以 $\boldsymbol{c}$ 为约束条件, 构建基于 DNN 的解码模型重构 $\hat{x}$, 网络结构如图 1 (b) 所示. x, $\hat{x}$, $\boldsymbol{c}$ 之间的互信息关系如图 2 (b) 所示. $H(x|c)$ 与 $H(\hat{x}|c)$ 可以通过比较原始训练数据与重构数据获得. 通过编码模型可以调整 $H(c|x)$, 通过解码模型可以调整 $H(\hat{x}|c)$. 不断调整 $H(c|x)$, $H(\hat{x}|c)$ 可以使其逼近 $H(x|c)$, 从而最大化互信息 $I(x;c)$. 具体做法如下: 从 v 个视图中, 任选一个视图, 假定为视图 1, 为视图 1 构建解码模型, 解码模型的输入包括来自正态分布的随机向量 $\boldsymbol{z}$ 和编码模型生成的表征向量 $\boldsymbol{c}$. 解码模型的输出为 $\hat{x}^{(1)}$. 网络优化的目标函数重新定义为

$$\min_{\theta_1,\cdots,\theta_v,\theta_{\rm dec}} \sum_{k,r} \frac{1}{2} \times \Big[KL\Big(P(c|x^{(k)})\|\frac{P(c|x^{(k)})+P(c|x^{(r)})}{2}\Big)+ KL\Big(P(c|x^{(r)})\|\frac{P(c|x^{(k)})+P(c|x^{(r)})}{2}\Big)\Big]+ KL(P(\hat{x}^{(1)}|c^{(1)})\|P(x^{(1)})) \tag{4}$$

其中, $\theta_{\rm dec}$ 为解码模型中的所有参数,

综上所述, 为构建适用于多视图的表征, 本文提出的基于 DNN 的多视图表征学习方法概括为: 1) 为每个视图分别构建 DNN, 将同一实例不同视图的数据映射至相同的表征向量; 2) 搭建条件解码模型, 保证表征向量包含关于实例的重构信息.

3 基于 GANs 的多视图数据生成

给定第 2 节提出的基于 DNN 的多视图表征学习方法, 对于测试实例的任意源视图, 可以获得关于该实例通用的表征向量. 接下来的任务是通过表征向量, 重构其他视图.

生成对抗网络的思想来源于博弈论中的纳什均衡, 它利用 DNN 分别构建生成模型 (G) 和判别模型 (D), 通过生成模型和判别模型之间迭代的对抗学习预测真实数据的潜在分布并生成新的样本. 网络优化的目标定义为生成模型与判别模型的博弈, 目标函数如下:

$$\min_G \max_D \{V(D,G) = {\rm E}_{x\sim p_{data}(x)}[\log D(x)] + {\rm E}_{z\sim p_z(z)}[\log(1-D(G(z)))]\} \tag{5}$$

为生成多视图数据, 可以为所有视图分别构建 GAN 网络, 并生成相应视图的数据. 然而, 由于标准的 GAN 生成模型以随机变量 $\boldsymbol{z}$ 为输入, 因此, 它无法指定生成与表征向量相对应的视图数据. 为解决这一问题, 有效的方法是构建条件生成对抗网络 (Conditional generative adversarial nets, CGAN)[32]. 其基本思想是在生成模型和判别模型中引入条件变量, 利用条件变量指导数据的生成. 因此, 本文提出基于对抗生成网络的多视图数据生成算法. 为每一视图构建条件生成对抗网络. 在生成模型中和判别模型中分别加入表征向量 $\boldsymbol{c}$ 作为约束条件作为输入层的一部分, 从而实现利用表征向量指导新视图数据的生成. 网络结构如图 3 所示. 每个 GAN 网络的优化的目标重新定义为以表征向量为约束条件的生成模型与判别模型的博弈.

$$\min_G \max_D \{V(D,G) = {\rm E}_{x\sim p_{\rm data}(x)}[\log D(x|c)] + {\rm E}_{z\sim p_z(z)}[\log(1-D(G(z|c)))]\} \tag{6}$$

从图 3 可以看出, 每个视图的 GAN 网络在训练开始前, 由编码模型生成表征向量 $\boldsymbol{c}$. 训练过程中, 生成模型 G 以采样自正态分布的随机变量 $\boldsymbol{z}$ 作为输入, 同时以表征向量 $\boldsymbol{c}$ 为约束条件. 判别模型 D 以真

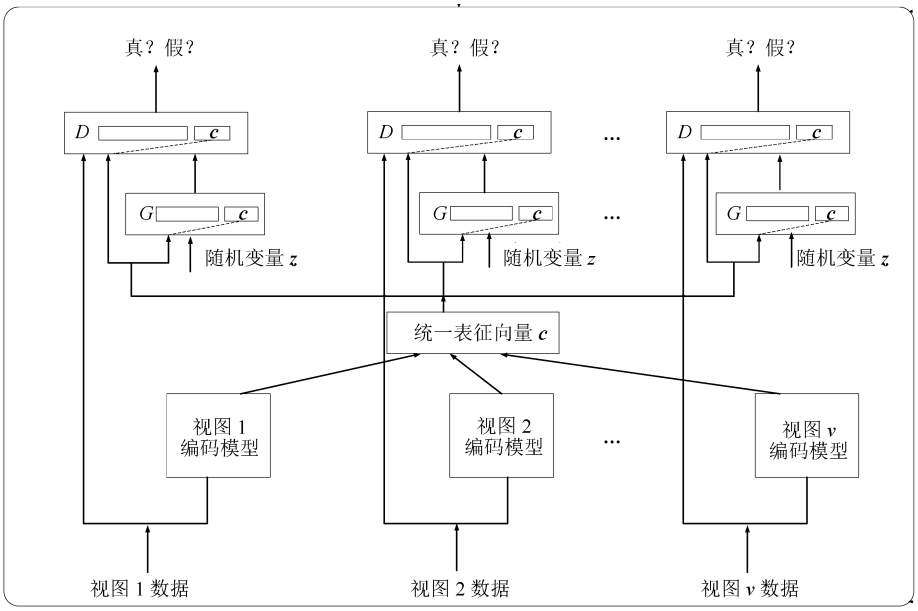


图 3 基于生成对抗网络的多视图数据生成框架

Fig. 3 Framework of the generative adversarial network based multi-view data generation

实训练数据, 或生成模型生成的数据为输入, 同时以表征向量 $\boldsymbol{c}$ 为约束条件. 生成模型和判别模型通过式 (6) 中的对抗训练不断逼近约束条件 $\boldsymbol{c}$ 下真实数据的潜在分布, 并生成新样本. 测试过程中, 由源视图通过编码模型生成表征向量 $\boldsymbol{c}$, 由于式 (4) 中优化目标条件的限制, 向量 $\boldsymbol{c}$ 将包含实例完整的重构信息, 并且可以将其做为约束条件传递至任意其他视图的生成模型. 对应视图的生成模型将以随机变量 $\boldsymbol{z}$ 为输入, 表征向量 $\boldsymbol{c}$ 为约束条件, 生成与源视图相匹配的数据.

4　实验结果与分析

4.1　多视图测试数据集

为验证本文所提算法的有效性, 在如下数据集合上展开实验.

1) 手写数据集合 (MNIST dataset of handwritten digits). MNIST 包含约 7 万幅图像, 每幅图像对应一个手写体数字, 大小为 28 像素 × 28 像素[33];

2) 街景数字集合 (Street view house numbers, SVHN). SVHN 包含约 8.9 万幅图像, 每幅图像对应一个真实世界的街道门牌号, 并且以门牌号的数字为中心, 大小为 32 像素 × 32 像素[34];

3) 人脸数据集合 (CelebFaces attributes, CelebA). CelebA 包含约 20 万幅图像, 每幅图像对应一个真实世界的人脸, 大小裁剪为 64 像素 × 64 像素[35].

4.2　模型评价标准

为了定量地衡量所提算法, 采用结构相似性 (Structural similarity index, SSIM)[36] 和峰值信噪比 (Peak signal to noise ratio, PSNR)[37] 作为评价指标衡量真实图像数据与模型生成的图像数据之间的相似度以及生成图片的质量.

SSIM 作为一种衡量两幅图像相似度的指标, 能够反映图像间的结构相似性. 假定 I_x 为模型生成的图像, I_y 为真实图像 (Ground truth), I_x 与 I_y 之间的 SSIM 定义为

$$SSIM(I_x, I_y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7}$$

其中, μ_x, μ_y 是 I_x 和 I_y 的像素均值, σ_x^2, σ_y^2 分别是 I_x 和 I_y 的方差, σ_{xy} 是 I_x 与 I_y 之间的协方差. 式 (7) 表明 SSIM 值越高, I_x 与 I_y 之间的相似性就越高, 生成的图像越接近真实图像.

PSNR 是一种评价图像的客观标准. 图像经过处理之后, 输出的图像都会在某种程度与原始图像不同. 将真实图像与生成图像对比, 得到生成的图像的 PSNR 值来测试模型的重构效果.

$$MSE = \frac{\sum_{n=1}^{FrameSize} (I_n - P_n)^2}{FrameSize} \tag{8}$$

$$PSNR = 10 \times \log \frac{255^2}{MSE} \tag{9}$$

其中, MSE 代表平均均方误差, I_n 是原始图像第 n 个像素值, P_n 指处理后图像第 n 个像素值, FrameSize 是图像长 × 宽 × 通道数. PSNR 的单位为 dB. PSNR 值越大, 表明图片质量越好, 失真度越小.

4.3　实验设置与结果

4.3.1　MNIST 数据集实验结果

对于 MNIST 数据集, 考虑 3 个视图, 其中原始图像为视图 1, 将图像遮挡 14 像素 × 14 像素的区域作为视图 2, 将图像进行 LBP 特征提取[38], 以特征向量作为视图 3. 对原始图像进行 LBP 特征提取得到了一个 236 维的特征向量, 将特征向量映射到二维空间, 示意图如图 4 (图 4 中, 灰度条展示了 0 ∼ 9 不同数字对应的灰度, 横纵坐标代表降维后二维特征, 共 8 000 张图片) 从图 4 可以看出, 每个类别的特征向量趋向于聚集在一起, 并且类别为 7 的数字与类别为 1 的数字更加接近. 此外, 类别之间出现了轻微重叠现象, 并且有少量数据点分布在坐标系的边缘. 在实验过程中, 首先以训练数据的 3 个视图数据为输入, 训练图 1 (b) 中的编码模型与解码模型. 训练过程采用每一对视图单独训练的方式, 网络训练以式 (4) 为目标函数. 编码模型与解码模型训练完成后, 以表征向量为约束条件, 每一视图训练图 3 中的生成对抗网络. 网络训练以式 (6) 为目标函

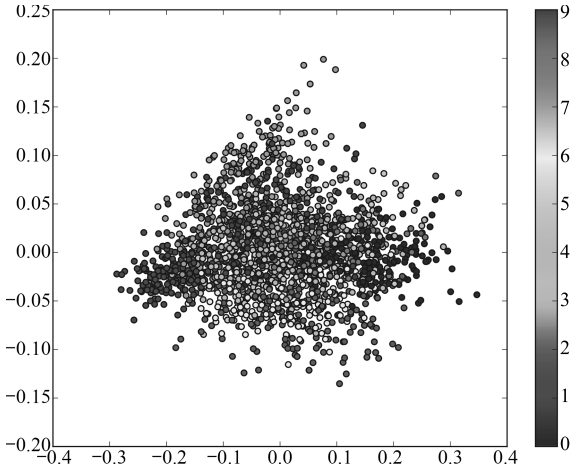


图 4　MNIST 视图 3 数据经过 PCA 后的可视化二维图
Fig. 4　The 2D-visualization of view 3 on MNIST after PCA

数. 测试过程中分别以测试实例的视图 2 和视图 3 作为源视图构建表征向量, 分别以表征向量作为约束条件利用视图 1 的生成模型生成对应的视图 1 数据.

图 5 显示了以视图 2 为源视图在随机挑选的 15 幅测试图像上的实验结果, 第 1 行表示遮挡一部分的源视图, 第 2 行表示源视图对应的真实图像, 第 3 行表示视图 1 生成模型构建的图像. 图 6 显示了以视图 3 为源视图在随机挑选的 15 幅测试图像上的实验结果, 第 1 行表示源视图对应的真实图像, 第 2 行表示视图 1 生成模型构建的图像.

从图 5 和图 6 可以看出, 尽管源视图 2 有较大比例遮挡, 源视图 3 从表达方式方面与原始数据有较大差异, 本文提出的生成算法仍然能够有效重构对应视图 1 数据. 表明第 2 节提出的表征学习方法不仅能获得图像中的语义信息, 而且能够获得包括方向、粗细、倾斜角度等其他信息, 同时表明本文提出的生成模型能够有效根据表征向量重构完整视图.

为进一步表明所提出算法的有效性, 将提的多视图生成对抗网络的实验结果 (Multi-view generative adversarial networks, MVGAN) 与条件生成对抗网络 (Conditional generative adversarial nets, CGAN)[30] 和条件变分自编码模型 (Conditional variational autoencoders, CVAE)[39] 产生的实验结果进行比较. 表 1 给出了三种算法在测试数据上的平均 SSIM 值与平均 PSNR 值. 从表 1 可以看出, 所提的 MVGAN 模型以视图 2 为源数据重构视图 1, SSIM 值和 PSNR 值均高于 CGAN 和 CVAE, 表明 MVGAN 重构的图像更接近真实图像, 并且失真度最小. 在 MVGAN 模型以视图 3 为源数据重构视图 1 上, SSIM 值比 CGAN 和 CVAE 的 SSIM 值低 0.09 和 0.14 左右, PSNR 值比 CGAN 和 CVAE 的 PSNR 值高 0.18 dB 和 0.09 dB 左右, 表明 MVGAN 模型中以视图 3 为源数据重构视图 1 得到的图片比 CGAN 和 CVAE 得到的图片失真度小. 对图片做纹理特征提取并应用数学的统计降维得到的特征向量比原图片损失了部分信息, 由缺失信息的数据重构完整数据时 SSIM 值会相对较低. 与此同时 CGAN 和 CVAE 使用了图片的完整信息, 因此获得了较高的 SSIM 值.

表 1 MNIST 数据集上的 SSIM 和 PSNR 比较结果

Table 1 Comparison results of SSIM and PSNR on MNIST

算法	SSIM 值	PSNR 值 (dB)
MVGAN (视图 2 重构视图 1)	0.8520 ± 0.0001	16.3135 ± 0.0880
MVGAN (视图 3 重构视图 1)	0.6474 ± 0.0013	12.2109 ± 0.1442
CGAN	0.7414 ± 0.0001	12.0301 ± 0.0512
CVAE	0.7912 ± 0.0031	12.1184 ± 0.0013

4.3.2 SVHN 数据集实验结果

对于 SVHN 数据集, 考虑 3 个视图, 其中原始图像为视图 1, 将图像遮挡 16 像素 × 16 像素的区域作为视图 2, 将图像进行 LBP 特征提取, 以特征向量作为视图 3. 在实验过程中, 展开与在 MNIST 数据集上相似的实验. 首先以训练数据的 3 个视图数据为输入, 训练图 1 (b) 中的编码模型与解码模型. 训练过程采用每一对视图单独训练的方式, 网络训练以式 (4) 为目标函数. 编码模型与解码模型训练完

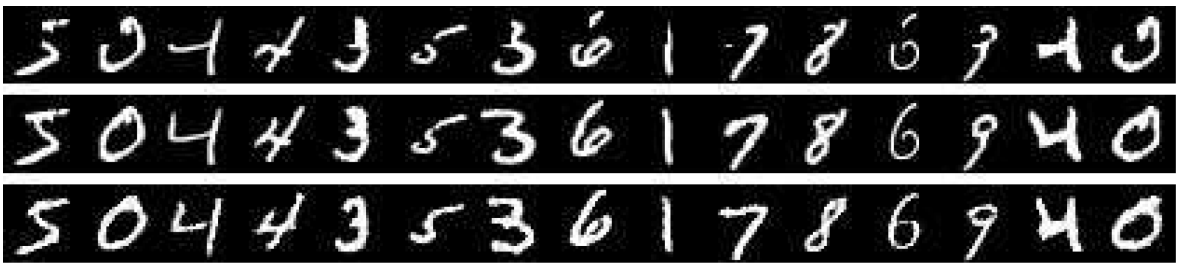


图 5 以视图 2 为源数据在 MNIST 上的重构结果

Fig. 5 Reconstruction results that take view 2 as source data on MNIST

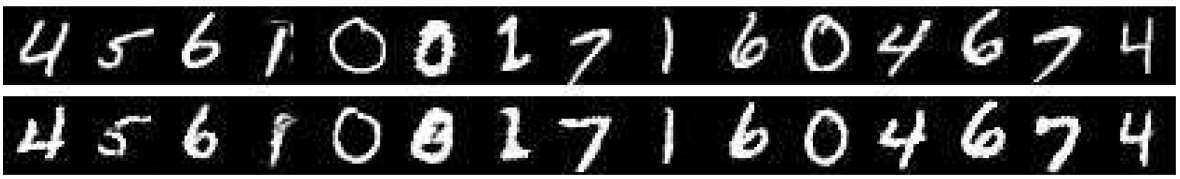


图 6 以视图 3 为源数据在 MNIST 上的重构结果

Fig. 6 Reconstruction results that take view 3 as source data on MNIST

成后, 以表征向量为约束条件, 每一视图训练图 3 中的生成对抗网络. 网络训练以式 (6) 为目标函数. 测试过程中分别以测试实例的视图 2 和视图 3 作为源视图构建表征向量, 分别以表征向量作为约束条件利用视图 1 的生成模型生成对应的视图 1 数据.

图 7 显示了以视图 2 为源视图在随机挑选的 15 幅测试图像上的实验结果, 第 1 行表示遮挡一部分的源视图, 第 2 行表示源视图对应的真实图像, 第 3 行表示视图 1 生成模型构建的图像. 图 8 显示了以视图 3 为源视图在随机挑选的 15 幅测试图像上的实验结果, 第 1 行表示源视图对应的真实图像, 第 2 二行表示视图 1 生成模型构建的图像.

从图 7 和图 8 中可以看出, 尽管源视图 2 有较大比例的遮挡, 源视图 3 从表达方式上与原始数据有较大差异, 但是本文提出的生成式算法仍然可以重构视图 1 的数字类别, 背景以及形状等信息. 表明提出的算法可以通过共同的表征学习达到重构视图数据的目的.

为了进一步说明算法的有效性, 将提出的多视图生成对抗网络 (MVGAN) 的实验结果与 CGAN 和 CVAE 产生的实验结果进行比较.

表 2 给出了这三种算法在测试数据上的平均 SSIM 值与平均 PSNR 值, 从表 2 可以看出, 所提的 MVGAN 模型以视图 2 为源数据重构视图 1, SSIM 值和 PSNR 值均高于 CGAN 和 CVAE, 表明 MVGAN 重构的图像更接近真实图像, 并且失真度最小. 在 MVGAN 模型以视图 3 为源数据重构视图 1 上, SSIM 值比 CGAN 和 CVAE 低 0.15 和 0.16 左右, PSNR 值比 CGAN 和 CVAE 的 PSNR 值高 0.91 dB 和 0.79 dB 左右. 表明 MVGAN 模型中以视图 3 为源数据重构视图 1 得到的图片比 CGAN 和 CVAE 得到的图片失真度小, 同时因为对图片做纹理特征提取并应用数学的统计降维得到的特征向量比原始图片损失了部分信息, 所以由缺失信息的数据重构完整数据得到的 SSIM 值会相对较低, 与此同时 CGAN 和 CVAE 使用了图片的完整信息, 因此获得了较高的 SSIM 值.

表 2　SVHN 数据集上的 SSIM 和 PSNR 比较结果

Table 2　Comparison results of SSIM and PSNR on SVHN

算法	SSIM 值	PSNR 值 (dB)
MVGAN (视图 2 重构视图 1)	0.4140 ± 0.0022	18.7987 ± 0.1475
MVGAN (视图 3 重构视图 1)	0.1848 ± 0.0020	15.8026 ± 0.1306
CGAN	0.3357 ± 0.0017	14.8910 ± 0.0002
CVAE	0.3465 ± 0.0028	15.0137 ± 0.0071

4.3.3　CelebA 数据集实验结果

对于 CelebA 数据集, 考虑 3 个视图, 其中原始图像为视图 1, 将图像遮挡 32 像素 × 32 像素的区域作为视图 2, 选取图像的 10 种属性作为视图 3. 视图 3 包含的图像属性有秃顶 (Bald), 刘海 (Bangs), 黑发 (Black hair), 眼镜 (Eyeglass), 男性 (Male), 嘴微张 (Mouth slightly open), 窄眼 (Narrow eyes), 无胡须 (No beard), 苍白肤色 (Pale skin), 戴帽 (Wearing hat). 表 3 展示了随机选取的 15 幅图片的属性向量的具体取值, 其中 “1” 表示属性为真, “−1” 表示属性为假.

图 7　以视图 2 为源数据在 SVHN 上的重构结果

Fig. 7　Reconstruction results that take view 2 as source data on SVHN

图 8　以视图 3 为源数据在 SVHN 上的重构结果

Fig. 8　Reconstruction results that take view 3 as source data on SVHN

图 9 显示了与表 3 对应的 15 幅测试图像上的实验结果, 第 1 行表示遮挡了一部分数据的视图 2, 第 2 行表示视图对应的真实图像, 第 3 行表示以视图 2 为源数据构建视图 1 的实验结果, 第 4 行表示以视图 3 为源数据构建视图 1 的实验结果.

从图 9 可以看出, 视图 2 虽然有较大比例的遮挡, 但是 MVGAN 能够依据视图 2 对应的 10 维属性信息重构一幅完整的图像, 例如第 1 张图像的人物具有戴眼镜、男性、有胡须的属性, 对应的重构图像同样具有戴眼镜、男性、有胡须的属性. 把原始图像的 10 维属性信息作为视图 3, 可以看出新提出的算法可以根据视图 3 的属性取值重构对应的图像, 例如图 9 第 2 行第 2 张人物具有黑发、男性、有胡须的属性, 对应的第 4 行第 2 张人物也具有黑发、男性、有胡须的属性. 表明提出的表征学习方法隐式地获取了实例中的表征信息, 并且能够通过表征信息重构其他视图的数据.

为进一步说明算法的有效性, 将 MVGAN 的实验结果与 CGAN 和 CVAE 产生的实验结果进行比较. 表 4 给出了三种算法在测试数据上的 SSIM 值与 PSNR 值, 从表 4 可以看出, MVGAN 模型的 SSIM 值和 PSNR 值均高于 CGAN 和 CVAE, 表明 MVGAN 重构的图像比 CGAN 和 CVAE 重构的图像更接近真实图像且失真度最小. 因为 MVGAN 模型在 CelebA 数据集上以重构 10 维属性信息为标准, 且 SSIM 评价指标是一种衡量两张图片相似程度的评价标准, 因此与在 MNIST 与 SVHN 数据集上重构完整视图信息的实验结果相比, 在 CelebA 数据集上得到了较低的 SSIM 值. PSNR 评价指标是一种衡量图片失真度的评价标准, 可以看出 MVGAN 模型重构的图片具有较小的失真度.

5 结论

在多视图学习领域, 研究如何根据已有视图构建完整视图具有重要意义. 其中一个需要解决的问题

图 9 以视图 2 为源数据在 CelebA 上的重构结果

Fig. 9 Reconstruction results that take view 2 and view 3 as source data respectively on CelebA

表 3 CelebA 视图 2 和视图 3 对应选中的 10 维属性

Table 3 The chosen attributes for view 2 and view 3 (10 dimensions)

图片编号	秃顶	刘海	黑发	眼镜	男性	嘴微张	窄眼	无胡须	苍白肤色	戴帽
a	−1	−1	−1	1	1	−1	−1	−1	−1	−1
b	−1	−1	1	−1	1	−1	−1	−1	−1	−1
c	−1	−1	1	−1	1	−1	−1	1	−1	−1
d	−1	−1	1	−1	1	1	−1	1	−1	−1
e	−1	−1	−1	−1	−1	−1	−1	1	−1	−1
f	−1	−1	−1	−1	−1	−1	−1	1	1	−1
g	−1	−1	−1	−1	1	−1	−1	1	−1	−1
h	−1	−1	−1	−1	−1	1	−1	1	−1	−1
i	−1	−1	−1	−1	−1	−1	−1	1	−1	−1
j	−1	−1	−1	−1	1	1	1	1	1	−1
k	−1	1	−1	−1	1	−1	−1	1	−1	−1
l	−1	−1	−1	−1	−1	1	−1	1	−1	−1
m	−1	−1	−1	−1	−1	1	−1	1	−1	−1
n	−1	−1	−1	−1	1	1	−1	−1	−1	−1
o	−1	1	1	−1	1	1	−1	1	−1	−1

表 4 CelebA 数据集上的 SSIM 和 PSNR 比较结果
Table 4 Comparison results of SSIM and PSNR on CelebA

算法	SSIM 值	PSNR 值 (dB)
MVGAN (视图 2 重构视图 1)	0.1143 ± 0.0023	10.0574 ± 0.0605
MVGAN (视图 3 重构视图 1)	0.1132 ± 0.0022	10.0342 ± 0.0587
CGAN	0.0512 ± 0.0036	9.5312 ± 0.0012
CVAE	0.0716 ± 0.0058	9.7881 ± 0.0020

是构建表征向量映射模型, 使得属于同一实例的不同视图数据能够映射至相同的表征向量, 同时表征向量还需包含关于实例的完整重构信息. 针对该问题, 本文提出一种基于 DNN 的多视图表征学习算法, 通过为每一视图构建 DNN, 借助 DNN 能够拟合任何分布的能力将不同视图的数据映射至通用的表征向量, 并且本文提出构建解码模型保证了表征向量中包含关于实例的完整重构信息. 为了依据表征向量信息重构完整视图, 本文提出一种基于生成对抗网络的多视图重构算法. 以表征向量为约束条件, 通过生成器与判别器的对抗训练来生成与源视图匹配的多视图数据. 实验结果表明, 提出的表征向量学习算法不仅得到了实例本身所带有的语义信息, 而且得到了方向、粗细、倾斜角度等其他重构信息. 因此, 提出的生成对抗网络方法能够根据低维的表征信息进行有效的重构.

接下来的研究工作将集中于研究如何获取表征向量的显式含义信息, 并指导多视图数据的生成.

References

1 Chaudhuri K, Kakade S M, Livescu K, Sridharan K. Multi-view clustering via canonical correlation analysis. In: Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada: ACM, 2009. 129−136

2 Kumar A, Daume III H. A co-training approach for multi-view spectral clustering. In: Proceedings of the 28th International Conference on Machine Learning. Washington, USA: Omnipress, 2011. 393−400

3 Wang W R, Arora R, Livescu K, Bilmes J. On deep multi-view representation learning. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: ICML, 2015. 1083−1092

4 Sun S L. A survey of multi-view machine learning. *Neural Computing and Applications*, 2013, **23**(7−8): 2031−2038

5 White M, Yu Y L, Zhang X H, Schuurmans D. Convex multi-view subspace learning. In: Proceedings of the 25th Annual Conference on Neural Information Processing Systems. Lake Tahoe, USA: NIPS, 2012. 1673−1681

6 Guo Y H. Convex subspace representation learning from multi-view data. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence. Washington, USA: AIAA, 2013. 387−393

7 Shekhar S, Patel V M, Nasrabadi N M, Chellappa R. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(1): 113−126

8 Gangeh M J, Fewzee P, Ghodsi A, Kamel M S, Karray F. Multiview supervised dictionary learning in speech emotion recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, **22**(6): 1056−1068

9 Zhai D M, Chang H, Shan S G, Chen X L, Gao W. Multiview metric learning with global consistency and local smoothness. *ACM Transactions on Intelligent Systems and Technology*, 2012, **3**(3): Article No. 53

10 Kumar A, Rai P, Daumé III H. Co-regularized multi-view spectral clustering. In: Proceedings of the 24th Annual Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc., 2011. 1413−1421

11 Chen M M, Weinberger K Q, Blitzer J C. Co-training for domain adaptation. In: Proceedings of the 24th Annual Conference on Neural Information Processing Systems. Granada, Spain: Curran Associates Inc., 2011. 2456−2464

12 Eaton E, desJardins M, Jacob S. Multi-view constrained clustering with an incomplete mapping between views. *Knowledge and Information Systems*, 2014, **38**(1): 231−257

13 Zhang X C, Zong L L, Liu X Y, Yu H. Constrained NMF-based multi-view clustering on unmapped data. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, Texas, USA: AIAA Press, 2015. 3174−3180

14 Yu S, Tranchevent L C, Liu X H, Glanzel W, Suykens J A K, De Moor B, et al. Optimized data fusion for kernel k-means clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(5): 1031−1039

15 Yu Kai, Jia Lei, Chen Yu-Qiang, Xu Wei. Deep learning: yesterday, today, and tomorrow. *Journal of Computer Research and Development*, 2013, **50**(9): 1799−1804
(余凯, 贾磊, 陈雨强, 徐伟. 深度学习的昨天、今天和明天. 计算机研究与发展, 2013, **50**(9): 1799−1804)

16 Guo Li-Li, Ding Shi-Fei. Research progress on deep learning. *Computer Science*, 2015, 42(5): 28−33
(郭丽丽, 丁世飞. 深度学习研究进展. 计算机科学, 2015, **42**(5): 28−33)

17 Hu Chang-Sheng, Zhan Shu, Wu Cong-Zhong. Image super-resolution based on deep learning features. *Acta Automatica Sinica*, 2017, **43**(5): 814−821
(胡长胜, 詹曙, 吴从中. 基于深度特征学习的图像超分辨率重建. 自动化学报, 2017, **43**(5): 814−821)

18 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5876): 504−507

19 Farquhar J D R, Hardoon D R, Meng H Y, Shawe-Taylor J, Szedmak S. Two view learning: SVM-2k, theory and practice. In: Proceedings of the 18th Annual Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2005. 355−362

20 Sindhwani V, Rosenberg D S. An RKHS for multi-view learning and manifold co-regularization. In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: ACM, 2008. 976−983

21 Yu S P, Krishnapuram B, Rosales R, Rao R B. Bayesian co-training. *The Journal of Machine Learning Research*, 2011, **12**: 2649−2680

22 Andrew G, Arora R, Bilmes J, Livescu K. Deep canonical correlation analysis. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, GA, USA: JMLR.org, 2013. 1247−1255

23 Westerveld T, de Vries A, de Jong F. Generative probabilistic models. *Multimedia Retrieval*, Berlin: Springer, 2007. 177−198

24 Rezende D J, Mohamed S, Wierstra D. Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv: 1401.4082, 2014.

25 Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527−1554

26 van den Oord A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. arXiv preprint arXiv: 1601.06759, 2016.

27 van den Oord A, Kalchbrenner N, Vinyals O, Espeholt L, Graves A, Kavukcuoglu K. Conditional image generation with pixelCNN decoders. In: Proceedings of the 30th Annual Conference on Neural Information Processing Systems. Barcelona, Spain: NIPS, 2016. 4790−4798

28 Kingma D P, Welling M. Auto-encoding variational Bayes. In: Proceedings of the 2014 International Conference on Learning Representations. Banff, Canada: ICLR, 2014.

29 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 2672−2680

30 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)

31 Chen Wei-Hong, An Ji-Yao, Li Ren-Fa, Li Wan-Li. Review on deep-learning-based cognitive computing. *Acta Automatica Sinica*, 2017, **43**(11): 1886−1897
(陈伟宏, 安吉尧, 李仁发, 李万里. 深度学习认知计算综述. 自动化学报, 2017, **43**(11): 1886−1897)

32 Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.

33 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278−2324

34 Sermanet P, Chintala S, LeCun Y. Convolutional neural networks applied to house numbers digit classification. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR). Tsukuba, Japan: IEEE, 2012. 3288−3291

35 Liu Z W, Luo P, Wang X G, Tang X O. Deep learning face attributes in the wild. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3730−3738

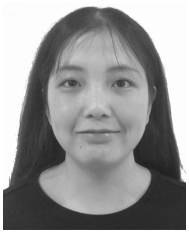
36 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600−612

37 Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 2008, **44**(13): 800−801

38 Xiang Zheng, Tan Heng-Liang, Ma Zheng-Ming. Performance comparison of improved HoG, Gabor and LBP. *Journal of Computer-Aided Design and Computer Graphics*, 2012, **24**(6): 787−792
(向征, 谭恒良, 马争鸣. 改进的 HOG 和 Gabor, LBP 性能比较. 计算机辅助设计与图形学学报, 2012, **24**(6): 787−792)

39 Kingma D P, Rezende D J, Mohamed S, Weling M. Semi-supervised learning with deep generative models. In: Proceedings of the 27th Annual Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2014. 3581−3589

孙 亮 大连理工大学计算机科学与技术学院讲师. 2012 年获得吉林大学计算机应用技术博士学位和高知工科大学信息科学博士学位. 主要研究方向为机器学习, 计算智能, 群智计算理论与应用.
E-mail: liangsun@dlut.edu.cn
(**SUN Liang** Lecturer at the College of Computer Science and Technology, Dalian University of Technology. He received his Ph. D. degree in computer application technology and information science from Jilin University and Kochi University of Technology in 2012. His research interest covers machine learning, computational intelligence, theory and application of swarm based intelligent computing.)



韩毓璇 大连理工大学计算机科学与技术学院硕士研究生. 主要研究方向为智能计算与机器学习方法.
E-mail: yuxuanhan@mail.dlut.edu.cn
(**HAN Yu-Xuan** Master student at the College of Computer Science and Technology, Dalian University of Technology. Her research interest covers computational intelligence and machine learning methods.)

康文婧 大连理工大学计算机科学与技术学院硕士研究生. 主要研究方向为智能计算与机器学习方法.
E-mail: wjkang@mail.dlut.edu.cn
(**KANG Wen-Jing** Master student at the College of Computer Science and Technology, Dalian University of Technology. Her research interest covers computational intelligence and machine learning methods.)

葛宏伟 大连理工大学计算机科学与技术学院副教授. 2006 年获得吉林大学计算机应用技术博士学位. 主要研究方向为计算智能, 机器学习, 系统建模与优化. 本文通信作者.
E-mail: hwge@dlut.edu.cn
(**GE Hong-Wei** Associate professor at the College of Computer Science and Technology, Dalian University of Technology. He received his Ph. D. degree in computer application technology from Jilin University in 2006. His research interest covers computational intelligence, machine learning, system modeling and optimization. Corresponding author of this paper.)