生成式对抗网络年度进展概述

哈尔滨工业大学 刘铭 颜肇义 左旺孟

一、 引言

生成式对抗网络(Generative Adversarial Network，GAN)[1]是近年来深度学习研究的热点之一，在计算机视觉领域中的底层视觉、图像生成和迁移学习等任务中获得了广泛的关注与应用。GAN 最初关注的是图像生成，采用对抗的方式同时学习一个生成网络 G 和一个判别网络 D。G 的输入是零均值标准方差的噪声向量 $\mathbf{z}\in\mathbb{R}^{n\times 1}$ ($z_i\sim\mathcal{N}(0,1)$)，旨在输出高质量的生成图像 $G(\mathbf{z})$。D 是一个二值判别器，用于判断输入的是真实图像 $\mathbf{x}$ 还是生成图像 $G(\mathbf{z})$。生成式对抗网络优化的目标函数为，

$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1-D(G(\mathbf{z})))] \quad (1)$$

由上式可见，判别器的目标在于学习一个最优的二值分类器 D 以尽可能区分真实图像和生成图像，而生成器 G 的目标在于学习一个最优的生成器以尽可能欺骗判别器 D。通过以对抗学习的方式交替训练 D 和 G，模型在理想情况下最终将达到纳什均衡，此时 $\mathbf{x}$ 和 $G(\mathbf{z})$ 的分布将完全相同，$G(\mathbf{z})$ 因而可以产生与真实样本难以区分的生成样本。

在 GAN 模型的各种推广形式中，条件生成式对抗网络(cGAN)是最有价值和最具代表性的方式之一。不同于标准的 GAN，cGAN 的输入在随机变量 $\mathbf{z}$ 之外，还包括一个条件变量 $\mathbf{c}$。CGAN 的生成器因而可以写为 $G(\mathbf{z},\ \mathbf{c})$。条件变量 $\mathbf{c}$ 的引入不但有助于改善 GAN 训练的稳定性(如：InfoGAN、ACGAN)，还可进一步拓展 GAN 的应用范围。如图像翻译、智能图像填充、人脸属性编辑等任务都可以表示为条件图像生成问题。

然而，基于公式(1)的 GAN 或 cGAN 训练往往会遇到梯度不稳定以及等问题而容易致使模式坍塌(Mode Collapse)，导致生成结果的多样性有限。针对这一问题，国内外学者近年来开展了大量的研究工作并提出了一些有效的改进方法，并推动了 GAN 在图像生成与翻译等任务中的广泛应用。

二、GAN 研究进展

1. 分布差异度量

直观地说，GAN 网络通过生成器与判别器的对抗学习，使生成器能够产生判别器难以区分真假的数据样本，这个过程的本质是不断拉近两个(或多个)数据分布之间的距离。因此，分布之间的差异度量是生成式对抗网络的学习中的一个核心问题。

如公式(1)，Goodfellow 等[1]利用 KL 散度(Kullback-Leibler divergence)与 JS 散度(Jensen-Shannon divergence)衡量真实数据分布与生成数据分布之间的差异。然而，生成式对抗网络将低维噪声映射到图像空间，其实质是图像空间中的低维流形，因此，真实数据分布与生成数据分布之间的重叠往往可以忽略。此时，KL 散度与 JS 散度趋近于正无穷或一个常数，无法准确反映两个分布之间的差异程度(距离)，进而导致生成式对抗网络存在训练不稳定、梯度消失、模式坍塌等问题。

Arjovsky 等[2]对原始生成式对抗网络的不足进行了理论分析，据此针对性地提出使用 Wasserstein 距离(又称 Earth-Mover 距离)对数据分布的差异进行度量，并在此基础上提出 Wasserstein GAN(WGAN)[3]，理论上解决了生成式对抗网络梯度消失等问题。Wasserstein 距离可以转换为基于 Lipschitz 约束的形式进行求解，WGAN 通过将判别器参数截断到一定范围内保证 Lipschitz 约束。然而，该方法会导致参数二值化的现象，使得判别器的判别能力大大下降。由

于 Lipschitz 约束$\|T\|_L \le 1$可以由$\|\nabla T\| \le 1$保证，Gulrajani 等[4]进一步提出了改进的 WGAN-GP 模型，通过梯度惩罚代替截断操作保证 Lipschitz 约束，同时采取真伪样本随机插值的方法尽量保证 Lipschitz 约束在整个样本空间的成立。Mescheder 等[5]通过对生成式对抗网络训练稳定性的数学推导，提出了相比于 WGAN-GP 更为简单有效的梯度惩罚项。Wu 等[6]提出了满足对称性约束的 Wasserstein 散度的 WGAN-div，避免了对 Lipschitz 约束的要求。

2. 特征规范化

深度网络的规范化是指对于某层的特征或者权重进行某种形式的规范化，从而改善网络学习的稳定性和训练效率。目前常见的有批规范化(Batch Normalization, BN)、正交正则(Orthogonal Regularization)以及谱规范化(Spectral Normalization)等。其中批规范化[7]在网络训练过程中，对每个 mini-batch 的网络特征首先进行规范化然后进行线性变换，通常能有效改善学习算法的收敛和稳定性。

除了对网络学习的特征进行规范化，我们也可以对网络的权重进行约束。在生成式对抗网络中，由于网络训练的不稳定性，一种较好的初始化网络权重的方法显得尤为重要。为此，[8]提出用正交初始化方法。在生成网络中，如果卷积的权重是正交的，那么特征经过正交的矩阵变换之后可以保持秩不变，从而在较大程度上避免出现梯度消失或爆炸的情况。正交初始化方法如下：

$$L_{ortho} = \sum\left(\left|WW^T - I\right|\right), \tag{2}$$

其中Σ表示对卷积求和，W是卷积权重，I是单位阵。然而，论文[9]指出这个初始化约束太强，所以在[10]中提出一种新的正交规范化变体来最小化卷积核内部的余弦相似度以提升稳定性，

$$\boldsymbol{R}_{\boldsymbol{\beta}}(\boldsymbol{W}) = \boldsymbol{\beta}\left\|\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{W} \otimes (\boldsymbol{1} - \boldsymbol{I})\right\|_{\boldsymbol{F}}^2. \tag{3}$$

正交正则通过将所有的权重的奇异值设置为 1，从而破坏了权重的频谱信息。为此人们提出谱归一化方法，只对权重的谱进行了放缩，保证其最大奇异值为 1，因此并不会破坏网络的矩阵结构。此外，谱归一化能够保证网络的 Lipschitz 连续性，从而稳定网络的训练。对于每层变换$g: h_{in} \mapsto h_{out}$，Lipschitz 范数的定义是$\sup_h \sigma\left(\nabla g(h)\right)$，其中$\sigma(A)$是权重矩阵$A$的谱范数，进而有：

$$\sigma(A) = \max_{h:h\neq \mathbf{0}} \frac{\|Ah\|}{\|h\|} = \max_{\|h\|_2 \le 1} \|Ah\|_2. \tag{4}$$

由上式可见，每层网络的 Lipschitz 范数等价于其权重矩阵A的最大奇异值。一方面，对于线性层$g(h) = Wh$，其 Lipschitz 范数为$\|g\|_{Lip} = \sup_h \sigma\left(\nabla g(h)\right) = \sup_h \sigma(W) = \sigma(W)$。另一方面，对于激活函数也同样符合 Lipschitz 连续性。比如，对于 Lipschitz 范数$\|a_l\|_{Lip} = 1$的激活函数a_l（如 ReLU），可以利用不等式$\|g_1 \circ g_2\|_{\text{Lip}} \le \|g_1\|_{\text{Lip}} \cdot \|g_2\|_{\text{Lip}}$。因此，约束网络每层权重$W^l$，使得其谱范数不大于 1，即可保证整个网络的 Lipschitz 连续性。通过对网络权重W进行谱规范化，即有：

$$\bar{W}_{SN}(W) := W/\sigma(W), \tag{5}$$

则可以保证$\|f\|_{\text{Lip}}$的上界为 1.

然而，谱规范化操作$\sigma(W)$如果直接进行奇异值分解，那么计算的开销是巨大的。作者提出用幂迭代[11]方法来减少计算开销。首先，对于每个权重进行随机初始化向量$\tilde{u}$，迭代构造左奇异向量$\tilde{v}$和右奇异向量$\tilde{u}$：

$$\tilde{v} \leftarrow W^T\tilde{u}/\left\|W^T\tilde{u}\right\|_2, \tilde{u} \leftarrow W\tilde{v}/\|W\tilde{v}\|_2, \tag{6}$$

最终估计出$\sigma(W) \approx \tilde{u}^T W \tilde{v}$。

对于 WGAN-GP 来说，其惩罚采样点$\tilde{x}$的梯度只能是对于当前生成网络的分布进行正则，并不能对生成网络分布的支集也进行正则。另一方面，生成网络的分布及其支集会随着训练不断改变，从而这种方式会造成正则本身的不稳定性。谱规范化对于 WGAN-GP 来说，对于更高的学习率具有更强的稳定性。实际情况下，谱规范化可以和 WGAN-GP 联合使用以获得更好的训练效果。

3. 多尺度生成

随着生成图像空间分辨率的提升，图像生成

任务的问题维度与难度也随之上升。多尺度生成作为降低高分辨率图像生成问题难度的一种有效策略，被广泛应用于生成式对抗模型。

Denton 等[12]提出 LapGAN 模型，通过在拉普拉斯金字塔框架内将多个 GAN 模型级联的策略，由粗到精地生成目标图像。其中，最低分辨率的 GAN 模型以随机噪声作为输入，其后每个 GAN 模型都是以更低分辨率模型输出和随机噪声作为输入的条件 GAN 模型，通过残差学习生成更高分辨率的目标图像。Wang 等[13]在图像翻译任务中采用类似的思想，将多个生成器进行嵌套，提出了 Pix2pixHD 模型，生成2048×1024分辨率的目标图像。同时，该模型采用嵌入物体边缘信息、多尺度判别器等方法进一步提升生成图像质量。

然而，上述方法采用多个生成器与判别器，导致网络结构较为复杂，需要消耗大量的计算资源。Karras 等[14]提出 Progressive GAN(PGGAN 或 ProGAN)模型，通过模型结构的动态调整实现多尺度生成。具体来说，模型从4×4像素的生成器与判别器出发，当前分辨率达到稳定状态时，在生成器末端与判别器前段分别增加新层以提升二者的分辨率，同时采用逐步提升新增层权重的方式进行平滑过渡，减少增加新层带来的不稳定性。该模型一方面充分利用不同分辨率生成难度的差异，降低了高分辨率模型训练的难度，另一方面由于模型的动态结构调整，显著减少了模型训练所需要的计算资源和时间。

Brock 等[10]在 BigGAN 中进一步探究了通道数、批样本数量等对生成式对抗网络生成质量的影响，提出增加通道数、扩大批样本数量可以明显提高生成图像的质量，但同时，二者的增加带来了稳定性问题。因此，BigGAN 采用正交正则化等多种策略提升模型训练的稳定性，并使用噪声截断技巧对生成结果的真实性和多样性进行平衡。此外，BigGAN 采用共享嵌入的方式，将噪声向量嵌入到生成器的多个层，使得不同分辨率的特征直接受到噪声向量的影响。这些策略的应用使得 BigGAN 的图像生成质量得到极大的提升，同时基于共享嵌入的多尺度生成策略也对模型的收敛起到了加速作用。

4. 应用：多样化图像翻译

多样化图像翻译指的是，给定一张图像，可以生成多张互异的目标域的图像。比如，给定一张夏天的图像，网络可以生成多张不同的冬天图像。一种直接的解决方案是在网络(如 Pix2Pix)的输入加入噪声，然而这样做往往会让网络在训练过程中直接忽略所加入的噪声，进而在测试时无法获得多样的输出结果。

为了解决上述问题，[15]提出了一种条件变分自动编码对抗网络(cVAE-GAN)与条件隐变量回归对抗网络(cLR-GAN)相结合的多样化图像翻译解决方法。对于成对的图像(A,B)，cVAE-GAN 将目标图像B先送入编码网络E中进行编码，得到编码向量$E(B)$，再将$E(B)$与输入图一起输出到生成网络G中，并最终获得重构图像$\hat{B}$以定义重建损失和对抗损失。进而，引入 KL 散度损失，使得$E(B)$近似服从高斯分布。

不同于 cVAE-GAN，cLR-GAN 则直接将输入图A与随机噪声z一起输入到生成网络G中，得到$\hat{B}$，再将$\hat{B}$输入E中以还原隐变量z。在此基础上，定义L_{GAN}以保证$\hat{B}$的无法与目标图B无法区分，L_1^{latent}以保证E可以重建隐变量z。由于E需要重建z，因而可以保证$\hat{B}$的生成不会直接忽略z，而是将z看成生成$\hat{B}$的关键条件。通过将 cVAE-GAN 和 cLR-GAN 进行结合，得到最终 Bicycle-GAN 的模型。

虽然 Bicycle-GAN 可以生成多样化的图像，然而需要大量的样本对数据。最近，MUNIT[16]和 DRIT[17]通过将不同域的图像可以分解为域不变内容空间(domain-invariant content space)以及域相关属性空间(domain-specific attribute space)两个部分，研究了基于非样本对数据的多样化图像翻译问题。由于 MUNIT 与 DRIT 的思想以及做法均相似，这里将以 DRIT 为例加以介绍。

DRIT 网络由内容编码器$\{E_X^c,E_Y^c\}$，属性编码器$\{E_X^a,E_Y^a\}$，生成器$\{G_X,G_Y\}$以及域判别器$\{D_X,D_Y\}$以及内容判别器D_{adv}^c。以源域为例，

DRIT 会首先将 X 映射到两个空间，即域不变内容空间（$E_X^c: X \to C$）以及域相关属性空间（$E_X^a: X \to A_X$）。G_X 会根据送入的内容以及属性（$G_X: \{C, A_X\} \to X$），生成相应的图像 u，此时域判别器 D_X 判别 u 和原始图像 X。另一方面，D_{adv}^c 用于判别分别从 X 和 Y 提取的内容，使其无法分辨，即有内容对抗损失 $L_{adv}^{content}$。为实现基于非样本对的图像翻译模型学习，进一步引入循环一致性损失 $L_1^{cc}\left(G_X, G_Y, E_X^c, E_X^a, E_Y^a\right)$。此外，网络还加入了自身重建损失 L_1^{rec}、属性编码器输出 的 KL 损失 L_{KL}、以及隐变量重建损失 L_1^{latent}。

5. 应用：多域图像翻译

相比于双域图像翻译工作[18, 19]，多域图像翻译利用单个模型完成多个领域间的相互映射，一方面避免了双域图像翻译存在的所需模型数量随领域个数的增加指数增长的问题，另一方面可以充分利用不同领域图像数据进行模型学习。如图 1 所示，Perarnau 等[20]提出 IcGAN 模型，利用单个模型生成多个不同领域（属性）的目标图像，但由于生成网络中解码器与编码器分别进行训练，使得 IcGAN 在人脸属性编辑任务中难以保留源人脸图像的内容和身份。

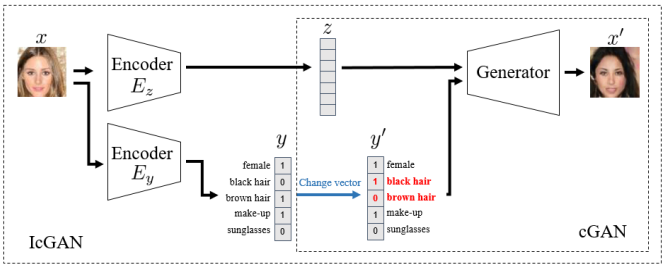


图 1 IcGAN 模型[20]

Choi 等[21]与 He 等[22]分别提出 StarGAN（图 2）与 AttGAN 模型（图 3），将属性识别网络与判别网络进行融合，通过端到端学习约束生成网络产生具有特定属性的目标图像。其中，StarGAN 将目标属性与源图像同时输入生成网络，通过循环一致性损失约束网络保留源图像内容，而 AttGAN 将目标属性与隐层特征结合，通过重建损失约束源图像信息的保留。

如图 4 所示，Liu 等[23]提出使用差值属性向量（即目标属性向量域源属性向量的差值）代替目标属性向量嵌入生成网络，显式地区分需要修改的属性与无需修改的属性，同时引入选择性转移单元，将源图像特征选择性地传递到目标图像。相比于 StarGAN 与 AttGAN，STGAN 有助于减少属性的误修改，有效提升了生成图像的属性编辑准确率和视觉质量（图 5）。

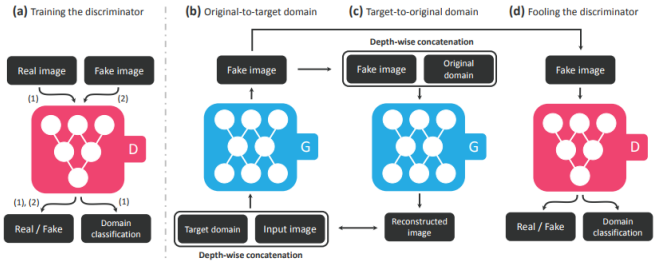


图 2 StarGAN 模型[21]

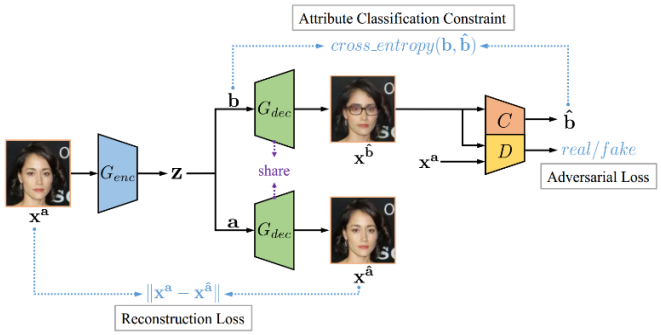


图 3 AttGAN 模型[22]

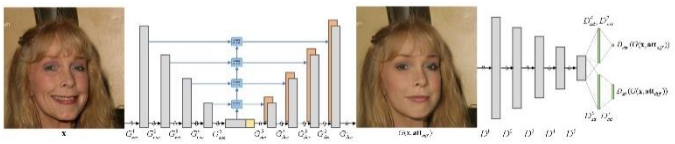


图 4 STGAN 模型[23]

图 5 基于 STGAN 的人脸属性编辑

三、未来发展方向

本文回顾了生成式对抗网络学习与训练方面的一些主要进展，包括从稳健损失出发的 WGAN 和 WGAN-GP，以及从网络参数或特征规格化角度出发的正交规格化和谱规格化。此外，多尺度学

习作为一种常用的改善生成质量策略，也在 Pix2pixHD、Progressive GAN 和 BigGAN 等得以使用与进一步推广。针对 GAN 的应用，本文以图像翻译为例，回顾了多样化和多域图像翻译方面的主要进展。

虽然生成式对抗网络近年来取得了较大的进展，但在学习的收敛性和稳定性等方面仍然缺乏足够的理论与技术支持。最近的 BigGAN 和 StyleGAN [24]虽然能够生成包含丰富纹理细节的高清图像，但往往伴随着一定程度的伪影和瑕疵，以及难以克服全局不一致性问题。此外，相对于训练图像，基于 GAN 生成的图像往往并不能带来额外的多样性，也是许多 GAN 应用中需要解决的一个重要问题。在应用方面，除图像翻译外，如何将生成式对抗网络更好地应用于迁移学习、视频生成以及其它低层视觉任务也是非常值得关注的研究方向。

（责任编辑：苏航）

参考文献

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, 2014.

[2] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. In *ICLR*, 2017.

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *ICML*, 2017.

[4] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *NIPS*, 2017.

[5] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge? In *ICML*, 2018.

[6] Jiqing Wu, Zhiwu Huang, Janine Thoma, Dinesh Acharya, and Luc Van Gool. Wasserstein divergence for gans. In *ECCV*, 2018.

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, 2015.

[8] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Neural photo editing with introspective adversarial networks. *arXiv preprint arXiv:1609.07093*, 2016.

[9] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *ICLR*, 2018.

[10] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. In *ICLR*, 2019.

[11] Frank Lin and William W Cohen. Power iteration clustering. In *ICML*, 2010.

[12] Emily L Denton, Soumith Chintala, Arthur Szlam, and Rob Fergus. Deep generative image models using a laplacian pyramid of adversarial networks. In *NIPS*, 2015.

[13] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *CVPR*, 2018.

[14] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *ICLR*, 2018.

[15] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman. Toward multimodal image-to-image translation. In *NIPS*, 2017.

[16] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018.

[17] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *ECCV*, 2018.

[18] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

[19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.

[20] Guim Perarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M Álvarez. Invertible conditional gans for image editing. 2016.

[21] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *CVPR*, 2018.

[22] Zhenliang He, Wangmeng Zuo, Meina Kan, Shiguang Shan, and Xilin Chen. Arbitrary facial attribute editing: Only change what you want. *arXiv preprint arXiv:1711.10678*, 2017.

[23] Ming Liu, Yukang Ding, Min Xia, Xiao Liu, Errui Ding, Wangmeng Zuo, and Shilei Wen. Stgan: A unified selective transfer network for arbitrary image attribute editing. In *CVPR*, 2019.

[24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

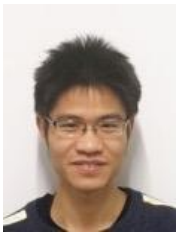
左旺孟

哈尔滨工业大学教授，博导。主要研究方向为的图像增强与复原、图像生成与编辑、视觉跟踪、物体检测与图像分类等计算机视觉任务。
Email: wmzuo@hit.edu.cn

刘铭

哈尔滨工业大学博士生，主要研究方向为图像生成与编辑。

Email: csmliu@outlook.com



颜肇义

哈尔滨工业大学博士生，主要研究方向为智能图像填充、人流密度估计。

Email: yanzhaoyi@outlook.com

基于情感区域自动挖掘的视觉情感预测

南开大学 杨巨峰 折栋宇

随着社交网络的不断普及，越来越多的网络用户倾向于用不同的媒介去表达他们的观点，识别其中所蕴含情感对于理解这些用户行为很有帮助，尤其是理解图片视频等视觉媒体内容中的情感已经在相关领域引起了越来越多的注意。该种分析算法的潜在用途是非常广泛的，主要包括情感图片检索、美学质量分类、意见挖掘、评论助手等应用。

我们研究涉及高度抽象概念的视觉情感分析问题。现有的大多数方法都只关注于提升从图像全局角度捕获特征的表达能力，但通过观察可以发现图像的整体和局部区域都可以传达出重要的情感信息，受到该启我们提出了一个框架来有效利用图片中表达情感的区域。我们首先利用现有的目标检测工具生成候选区域，再使用候选区筛选算法去除冗余和具有干扰的目标区域，然后利用卷积神经网络计算每一个候选区域的情感得分，通过同时考虑物体得分和情感得分自动发掘表达情感能力较强的区域。最后，将整张图片和局部区域的卷积神经网络输出相结合，产生最后的预测结果。由于标注情感区域非常主观并且费力，我们的框架只需要图像级别的标签，可以显著地减少在训练中需要的标注负担，这对于情感分析尤其重要。实验表明我们提出的算法在八个主流的基准数据集上的结果均超过了目前最先进的方法。

以上系列工作发表于 IEEE TMM 2018、IEEE CVPR 2019、AAAI 2017、IJCAI 2017。

（责任编辑：王金甲）

杨巨峰

南开大学副教授，主要研究方向为计算机视觉、机器学习、多媒体计算。

Email: yangjufeng@nankai.edu.cn

折栋宇

南开大学计算机视觉实验室硕士研究生，主要研究方向为计算机视觉、情感计算。

Email: sherry6656@163.com

多长度哈希联合学习方法

山东财经大学 聂秀山 山东大学 刘兴波

随着互联网和信息技术的发展，网络上数据呈爆炸性增长，如何高效的存储、管理、检索、分析大数据成为学术界和产业界关注的热点。哈希学习(Hash Learning)正是解决上述问题的可行方案之一。哈希学习结合数据自身分布和内在特性，利用机器学习方法，把高维的数据转化为二值码表示，同时尽可能保持数据在原特征空间的相似性，哈希学习对数据的二值表示形式，可以显著的节省存储空间，提升数据处理速度，因此在大数据学习中占有重要的地位。另一方面，哈希码也可以看作是一种重要的二值特征表示形式，而二值特征表示在识别、分类、匹配、搜索等计算机视觉和多媒体信息处理领域也取得了较好的应用效果。因此哈希学习越来越受到研究者的重视。

现有的哈希学习算法中，大多在模型训练之前，需要预先设定一个哈希码长度(例如 48 或 128 比特等)，然后利用相关模型训练参数，当哈希长度发生变化时，需要重新运行模型来得到新长度的哈希码。因此，如何在一个框架下同时得到多个不同长度的哈希码，供用户在不同情境下选择使用是一个值得研究的问题。

另一方面，如果把哈希码看作数据的二值特征表示的话，不同长度的哈希码本质上也是数据的多种特征表示。模式识别和计算机视觉领域相关研究表明，充分利用多特征之间的关联关系对特征表示和数据分析性能的提升具有较好的促进作用。

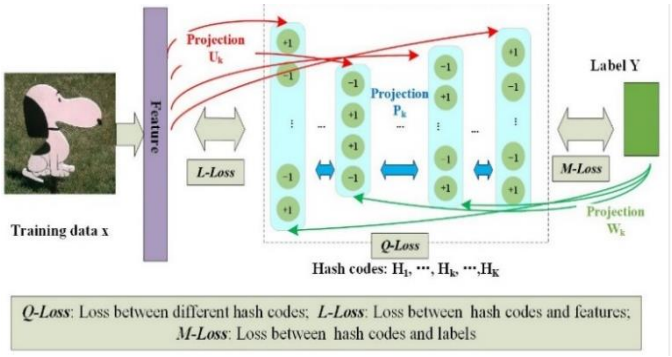


图 1 多长度哈希联合学习框架图

基于以上动机，我们提出了一种多长度哈希联合学习方法。如图 1 所示，该方法把不同长度哈希放在一个模型中，通过联合优化和学习，同时得到不同长度的哈希码表示。该方法的模型共分为三个部分：第一部分(Q-loss)用于刻画多长度哈希之间的关联关系；第二部分(L-loss)用于表达数据原始特征和哈希表示之间的信息损失；第三部分(M-loss)的作用是利用样本的标记信息。以上三部分共同构成模型的目标函数，该目标函数可以通过交替优化的方式求解。实验结果表明，该方法因有效的利用了同一样本不同长度哈希码的关联关系，比直接学习同样长度的哈希码，在性能上有较为明显的提升。

以上部分工作已发表于国际会议 AAAI2019。

(责任编辑：任桐炜)

聂秀山

山东财经大学计算机科学与技术学院教授、博士生导师，主要研究方向为计算机视觉、机器学习、多媒体检索等。

Email: niexsh@sdufe.edu.cn

刘兴波

山东大学计算机科学与技术学院博士生，主要研究方向为哈希学习、多媒体信息处理等。

Email: geshiming@iie.ac.cn