文章编号:2096 - 1383(2024)01 - 0068 - 05

基于改进 InfoGAN 的字体多风格融合模型

陈芯芯[a,b],王江江[a,b]
(大连民族大学 a. 计算机科学与工程学院; b. 大连市汉字计算机字库设计技术创新中心 辽宁 大连 116650)

摘 要:为解决汉字字体结构复杂、多风格融合特征难度大的问题,提出了一种基于改进 InfoGAN 的字体多风格融合方法。对 InfoGAN 特征多风格融合模型进行改进,调整输入向量的维度,添加了通道注意力模块。InfoGAN 可以将较难提取的风格特征清晰化、规律化,通过改进 InfoGAN,实现了对汉字字体图像风格特征的多风格融合,得到了能够控制汉字字体风格的特征向量。把改进的 InfoGAN 模型和 VAE、Beta - VAE、AAE 进行对比实验,再通过模型消融实验证明通道注意力的有效性。实验结果表明:该模型能够更好地将不同风格的特征进行分离,避免了信息重叠和冲突,可以有效准确地完成字体多风格融合任务。

关键词:特征多风格融合;通道注意力;InfoGAN
中图分类号:TP391 **文献标志码**:A

A Font Multi - style Fusion Model Based on Improved InfoGAN

CHEN Xinxin[a,b], WANG Jiangjiang[a,b]
(a. School of Computer Science and Engineering; b. Dalian Chinese Font Design Technology Innovation Center, Dalian Minzu University, Dalian Liaoning 116650, China)

Abstract: To solve the problem of complex font structure and difficulty in fusing multi - style features in Chinese characters, a font multi - style fusion method based on improved InfoGAN is proposed. The InfoGAN feature multi - style fusion model has been improved by adjusting the dimension of the input vector and adding a channel attention module. InfoGAN can clarify and regularize style features that are difficult to extract. By improving InfoGAN, multi - style fusion of Chinese font image style features is achieved, and feature vectors that can control Chinese font style are obtained. The improved InfoGAN model is compared with VAE, Beta - VAE and AAE in experiments, and then the effectiveness of channel attention is demonstrated through model ablation experiments. The experimental results show that the model can better separate features of different styles, avoid information overlapping and conflicting, and complete the task of font multi - style fusion effectively and accurately.

Key words: integration of multiple features and styles; channel attention; InfoGAN

汉字是一种内涵丰富、外观多样化的字符系统,由于汉字的形状与书写风格多样,对汉字多风格特征抽取后融合难度较大。在完成字体图像多风格迁移融合学习过程中,并不需要关注图像的所有特征,只需要关注字体图像的风格特征,此时如果可以把字体的风格特征从整体维度中解耦融合出来,就能更好地控制生成效果。InfoGAN[1]的设计目的就是将这些杂乱无章的特征清晰化规律化,通过解耦方式将多风格特征融合出来以控制图像的生成。

收稿日期:2023 - 08 - 07;**最后修回日期**:2023 - 10 - 13
作者简介:陈芯芯(1997 -),女,山东济南人,大连民族大学计算机科学与工程学院硕士研究生,主要从事计算机技术研究。

多风格字体融合生成技术研究意义重大，Yunjey Choi 等人[2]提出了多个域之间图像转换的 StarGAN 模型，以解决生成过程中图像域的可拓展性和鲁棒性较差问题；李金金等人[3]提出了 IBN－Net，将实例标准化和批标准化结合形成残差网络的基础模块，实现了多种风格域的无监督字体风格迁移；Xiaoxue Zhou 等人[4]在多尺度内容和风格特征融合的基础上，构建了一个风格传输网络，可以同时完成多个字体的风格迁移；陈丹妮等[5]提出了一种 SSNet 网络，结构网络和语义网络负责源域字体特征的提取，可提高生成多风格汉字图像的质量；Jianwei Zhang 等人[6]提出了一种结构语义网（SSNet）的汉字排版生成方法，该方法利用结构模块中解开的笔画特征，语义模块中预先训练的语义特征来生成多风格目标图像；Licheng Tang 等[7]提出了一种通过学习细粒度的局部样式和空间对应关系的内容和参考字形，使用交叉关注机制关注参考字形中局部样式和细粒度样式表示的方法，生成多风格样式融合的字体图像。

本文通过改进深度学习的现有模型多风格融合出生成字体图像的关键风格特征，通过控制多风格融合出的关键风格特征，进行不同字体风格特征的融合。

1　改进的 InfoGAN 特征多风格融合模型

本文 InfoGAN[1]是在传统的生成对抗网络（GAN）基础上加入了一个额外的噪声向量，可以用来融合输入图片的多风格特征。通过将噪声向量分解为多个部分，每个部分对应于输入图片的不同特征，可以使生成器生成更多样化的图片。

具体来说，InfoGAN 的生成器输入由三部分组成：噪声向量（Latent Code）、条件向量（Conditional Code）和分类向量（Class Code）。其中，噪声向量和条件向量是原始 GAN 中的输入，分类向量是额外加入的向量。在训练过程中，分类向量是由判别器预测得到的，用于指导生成器生成具有特定类别的图片。通过对噪声向量分解，可以将其分成两部分：一个部分用于控制输入图片的全局特征，例如位置、角度等；另一个部分用于控制输入图片的局部特征，例如颜色、纹理等。这样，生成器就可以根据噪声向量的不同部分生成具有不同全局和局部特征的图片。

本文在 InfoGAN 原有基础上，修改了输入向量的维度，增加了通道注意力模块，使其可以融合出更多字体图像的相关特征，针对大小写英文字母的网络模型如图 1。

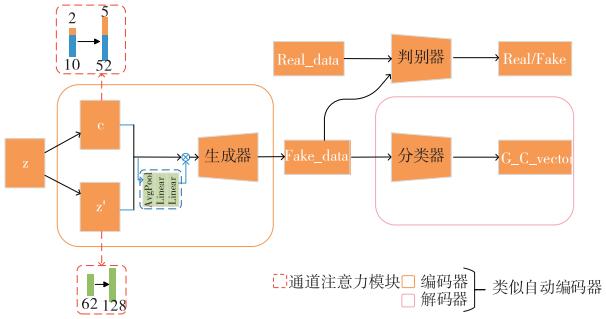

图 1　改进的 InfoGAN 特征多风格融合网络模型

输入向量的“固定”部分包含离散和连续两部分，针对 52 位大小写英文字母数据集，本文将离散的潜码数保持原模型的 1 位不变，离散潜码的维度由 10 维改为 52 维，连续潜码数由 2 位变为 5 位；针对中文数据集，选取常用的 500 个汉字图像，将离散的潜码数保持原模型的 1 位不变，离散潜码的维度由 10 维改为 500 维，连续潜码数由 2 位变为 5 位。添加的通道注意力模块由一个自适应的平均池化层和两个全连接层组成。具体的网络结构见表 1。

表 1　改进的 InfoGAN 网络结构

	核大小	特征图个数	归一化	激活函数
生成器	1	1 024	BN	ReLU
	7	128	BN	ReLU
	4	64	BN	ReLU
	4	3	–	Sigmoid
判别器	4	64	–	LeakyReLU
	4	128	BN	LeakyReLU
	7	1 024	BN	LeakyReLU
类别	3	128	BN	LeakyReLU
	3	64	–	–
	3	32	–	–
	3	32	–	–

1.1　字体图像特征多风格融合

基于 InfoGAN 的多风格融合的核心是分离特征或者提取特征。神经网络中的神经元以某种方式单独学习完整的概念，一个神经元可能学会特定的物体，而不明显依赖于其他神经元。通常，学习到的特征往往是混杂的，它们在数据空间中以一种无序而复杂的方式被编码。如果这些特征是

可以分解的，那么这些特征就更容易理解，就可以更方便的使用这些特征进行编码如图 2。

改进的 InfoGAN 可以最大化相互信息以用来可学习表示，分离离散的和连续的潜在因素从而扩展到复杂的数据集，并且不需要太多训练时间。

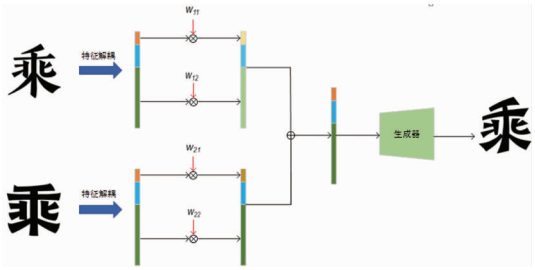

图 2 InfoGAN 的特征解耦示意图

InfoGAN 可以通过无监督的方式学习到数据的高层语义特征，并且可以通过控制隐变量来生成具有不同特征的数据。其中生成器的输入被分成了两部分：随机噪声 Z 和由多个隐变量构成的 Latent Code、c，即可解释的隐变量。其中，c 有先验的概率分布，可以是离散数据，也可以是连续数据，用来表示生成数据的不同特征。通过改变离散特征表示（Categorial Latent Code），可以生成不同种类的字母或汉字。通过改变连续特征表示（Continous Latent Code），可以生成不同风格的字母或汉字。

1.2 通道注意力模块

基于 CNN 的通道注意力[8]是一种注意力机制，它可以根据每个通道的重要程度，来增强有用的特征，减弱无用的特征。

本文的通道注意力模块如图 3。具体来说，输入一个 $H \times W \times C$ 的特征 F，首先，对每个通道的特征图分别进行全局最大池化和平均池化，得到两个 $1 \times 1 \times C$ 的向量。然后，用一个共享的两层全连接网络把这两个向量非线性变换成另外两个 $1 \times 1 \times C$ 的向量。其次，把这两个向量加起来并用一个 Sigmoid 激活函数，得到一个 $1 \times 1 \times C$ 的权重系数向量 Mc。最后，将 Mc 乘以输入的特征 F，得到一个新的特征。

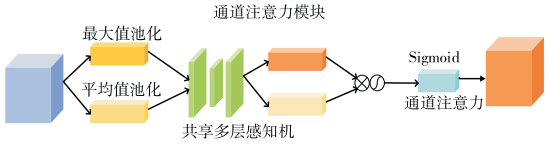

图 3 通道注意力模块

1.3 损失函数

为了保证生成图像的生成质量和细节特征，本文设计了五种损失函数，模型训练时使用不同损失函数的线性组合。

Dversarial Loss[9]：最终的优化目标是让生成器 L_{adv} 最小，判别器 L_{adv} 最大，对抗损失 L_{adv} 如公式（1）所示：

$$L_{adv} = E_{x \in p_{data}}[logD(x,y)] + E_{z \in p_{input}}[log(1 - D(G(x)))]。 \quad (1)$$

Self Rebuilding Loss[10]：为了保证字体图像编码和解码过程中没有信息丢失或模型引起失真，改进的 InfoGAN 模型应具有自我重建的能力，即能够根据编码器提取的结构和风格信息重新生成原图像。自重建损失 L_{rec} 如公式（2）所示：

$$L_{rec} = [\| \hat{x} - x \|_1 + \| \hat{y} - y \|_1]。 \quad (2)$$

Consistency Cycle loss[11]：由于 InfoGAN 采用无监督训练方式，生成的增强图像可能会丢失原始图像的结构信息。因此，采用了循环一致性损失，以保证增强前后图像的结构尽量相似，具体如公式（3）所示：

$$L_{cyc} = E_{x-p_{data}}[|| \tilde{x} - x ||_1] + E_{y-p(y)}[|| \tilde{y} - y ||_1]。 \quad (3)$$

Consistent Style Loss[12]：为了生成的清晰和失真图像与原图保持了相同的风格，以便风格特征提取更一致，引入了风格一致性损失，具体公式如（4）所示：

$$L_{sty} = \| E_x^S(x) - E_y^S(y^x) \|_1 + \| E_y^S - E_x^S(x^y) \|_1。 \quad (4)$$

Consistency Structure Loss[13]：为了让生成图像与原图像的结构更一致，还加入结构一致损失来强化生成前后的结构特征一致性，其具体公式如（5）所示：

$$L_{cont} = \| E_x^C(x) - E_y^C(y^x) \|_1 + \| E_y^C - E_x^C(x^y) \|_1。 \quad (5)$$

1.4 评价指标

为了验证本文提出的模型的性能，通过以下三个评价指标进行定量分析。

（1）均方根误差是衡量预测值与真实值之间的偏差差距。RMSE 的值越小，表示预测模型的精度越高。其中 x 表示汉字字体生成网络生成的融合图像，y 表示期望目标汉字字体图像。其计算公式如（6）所示：

$$RMSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \| x(i,j) - y(i,j) \|^2。 \quad (6)$$

（2）内容相似度是通过 VGG19 网络的分类器提取两张图像的高级特征，这些特征之间的欧氏

距离越小,两张图像的 SIOC 越高。

两个高级特征向量 $F^1 = (F_1^1, F_2^1, F_3^1, \cdots, F_n^1)$ 和 $F^2 = (F_1^2, F_2^2, F_3^2, \cdots, F_n^2)$ 的欧氏距离公式如公式(7)所示:

$$d^F(F^1, F^2) = \sqrt{(F_1^1 - F_1^2)^2 + (F_2^1 - F_2^2)^2 + \cdots + (F_n^1 - F_n^2)^2}\ 。 \tag{7}$$

根据上述高级特征向量之间的欧氏距离可得内容相似度如公式(8)所示:

$$SIOC = \frac{1}{1 + d^F(F^1, F^2)}\ 。 \tag{8}$$

(3)风格相似度是通过卷积层得到特征图组成的 Gram 矩阵,Gram 矩阵之间的欧氏距离越小,两张图像的风格越相似,风格相似度评价标准可根据计算特征之间的相关性 构建 Gram 矩阵,比较两张图像的 Gram 矩阵可以体现风格损失的情况。

特征图组成的两个 Gram 矩阵 $G^1 = (G_1^1, G_2^1, G_3^1, \cdots, G_n^1)$ 和 $G^2 = (G_1^2, G_2^2, G_3^2, \cdots, G_n^2)$ 的欧氏距离公式如公式(9)所示:

$$d^G(G^1, G^2) = \sqrt{(G_1^1 - G_1^2)^2 + (G_2^1 - G_2^2)^2 + \cdots + (G_n^1 - G_n^2)^2}\ 。 \tag{9}$$

根据上述 Gram 之间的欧氏距离可得风格相似度如公式(10)所示:

$$SIOS = \frac{1}{1 + d^G(G^1, G^2)}\ 。 \tag{10}$$

2　实验及实验结果分析

2.1　实验数据集与实验环境

本文使用 Adam 优化算法来训练生成器和判别器,设置批处理大小为 4,初始学习率为 0.000 2,从 GB2312 字符集中选取了 10 个数字和 52 个大小写英文字符,构成了西文数据集,从 GB2312 字符集中选取了常用的 9 169 个中文汉字构成了汉字数据集。

本文的实验环境采用 Intel Xeon(R)CPU E5 - 2620 v4@2.10 GHz×32 处理器,64 GB 内存,配备 2 张 Tesla K40C 显卡,在 Ubuntu 18.04 操作系统下,配置 Python 3.6.5、深度学习框架 PyTorch、Keras,使用 CUDA9.0 和 CuDNN 实现 GPU 加速。

2.2　实验结果分析

本文将训练数据分为西文数据集和中文数据集两部分,分别进行训练和测试。西文数据集包括 41 600 张训练集和 10 400 张验证集,中文数据集包括 24 000 张训练集和 6 000 张验证集。

中文数据集多风格融合的汉字的粗细如图 4。InfoGAN 的优点是它不需要任何监督信息,而是自动地发现数据中的潜在结构和变化因素。

图 4　中文数据集多风格融合出汉字的粗细

本文在 InfoGAN 模型中加入条件变量,使得生成器模型可以针对不同的条件生成不同的风格图片。部分不同西文和中文风格融合后的效果图如图 5。中文融合部分的实验效果是用不同的字体从不同的方面进行融合。

图 5　中西文多风格融合示例图

2.3　实验方法对比

本文对比了 VAE[14]、Beta - VAE[15] 和 InfoGAN 等特征多风格融合方法生成字体图像的质量和多样性如图 6。从图中可以看出本文提出的方法在字形、纹理、风格等融合方面均略好于其他方法。

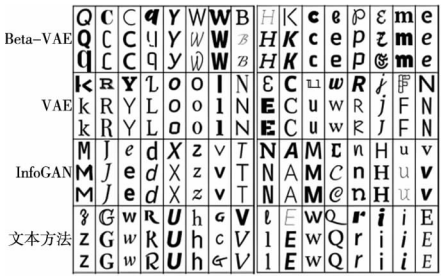

图6 不同方法对比实验示例

此外,通过定量指标或定性指标来评估不同模型的表现见表2。在两组字体图像中,RMSE指标比VAE平均低1.00个百分点,比Beta-VAE平均低1.02个百分点,比InfoGAN平均低1.34个百分点;SIOC指标比VAE平均高0.76个百分点,比Beta- VAE平均高1.12个百分点,比InfoGAN平均高0.48个百分点;SIOS指标比VAE平均高0.83个百分点,比Beta-VAE平均高1.01个百分点,比InfoGAN平均高1.07个百分点。说明本文方法生成的字体融合图像与参考图像的均方根误差最小,模型预测精度最高;内容相似度和风格相似度最高。通过这些实验结果,可以证明本文提出的方法具有较强的特征解耦能力和图像多风格融合能力。

表2 定量评价

字体	汉仪铸字美心体-汉仪元隆黑		
评价指标	*RMSE*	*SIOC*	*SIOS*
VAE	0.783	0.771	0.797
Beta-VAE	0.765	0.779	0.751
InfoGAN	0.796	0.804	0.717
本文方法	0.698	0.862	0.871

2.4 消融实验

为了评估该模型的效果和探究其内部机制,进行了消融实验。使用改进的InfoGAN模型,在200种字库和9 000张字体图像的训练集下进行训练。对比了使用CAM融合和不使用CAM融合两种方式的效果差异。结果表明,改进后的模型能够更好地保留不同汉字书法风格的特征,并生成更加自然、流畅的汉字书法风格。比较原始的InfoGAN模型,具有更好的多风格融合效果。模型改进前后的部分对比示例图如图7。

加入CAM模块的消融实验定量分析见表3。从表中数据可以看出加入CAM模块的改进是有效的。

图7 模型消融实验示例图

表3 模型消融实验定量评价

字体	汉仪铸字美心体-汉仪元隆黑		
评价指标	*RMSE*	*SIOC*	*SIOS*
不使用CAM	0.728	0.834	0.749
使用CAM	0.574	0.886	0.856

3 结 语

通过改进的InfoGAN模型。对不同字体的特征进行提取,将这些特征分解为不同的子空间,即多风格融合。然后再将这些特征进行融合,生成新的字体。因此,可以生成大量具有不同风格的字体,且生成的字体与原始字体的特征关系可以被保留,因此生成的字体质量较高。

参考文献:

[1] CHEN X,DUAN Y,HOUTHOOFT R,et al. InfoGAN:Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[J]. Advances in neural information processing systems,2016,29:2180-2188.

[2] CHOI Y, CHOI M, KIM M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, June 18-23. Salt Lake City: CVPR, 2018: 8789-8797.

[3] 李金金, 徐向纮, 龚心满. 基于StarGANv2的多风格字体生成研究[J]. 中国计量大学学报, 2022, 33(1):73-82.

[4] ZHOU X,ZHANG Z,CHEN X,et al. Chinese character style transfer based on the fusion of multiscale content and style features[C]//2021 40th Chinese Control Conference,July 26-28. Shanghai:IEEE,2021:8247-8252.

[5] 陈丹妮. 基于SSNet的汉字印刷字体生成[D]. 广州: 华南理工大学, 2019.

[6] ZHANG J, CHEN D, HAN G, et al. SSNet: Structure-Semantic Net for Chinese typography generation based on image translation[J]. Neurocomputing, 2020, 371: 15-26.

(下转第79页)

[12] LI R,CHEN H,FENG F,et al. Dual graph convolutional networks for aspect - based sentiment analysis[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online, ACL, 2021: 6319 - 6329.

[13] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. Semeval - 2016 task 5: Aspect based sentiment analysis[C]//ProWorkshop on Semantic Evaluation (SemEval - 2016). Association for Computational Linguistics. San Diego, California: ACL, 2016: 19 - 30.

[14] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha, Qatar: ACL, 2014: 1532 - 1543.

[15] CHEN P, SUN Z, BING L, et al. Recurrent attention network on memory for aspect sentiment analysis[C]// Proceedings of the 2017 conference on empirical methods in natural language processing. Copenhagen, Denmark: ACL, 2017: 452 - 461.

[16] LI X, BING L, LAM W, et al. Transformation networks for target - oriented sentiment classification[J]. arXiv preprint arXiv:1805.01086, 2018.

[17] ZHANG C, LI Q, SONG D. Aspect - based sentiment classification with aspect - specific graph convolutional networks[J]. arXiv preprint arXiv:1909.03477, 2019.

[18] SUN K, Zhang R, MENSAH S, et al. Aspect - level sentiment analysis via convolution over dependency tree [C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP - IJCNLP). Hong Kong: ACL, 2019: 5679 - 5688.

[19] ZHANG M, QIAN T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis[C]//Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP). Online, ACL, 2020: 3540 - 3549.

[20] WU H , ZHANG Z , SONG H, et al. Phrase dependency relational graph attention network for aspect - based sentiment analysis [J]. Knowledge - Based Systems, 2022, 236: 107736.

(责任编辑　王楠楠)

(上接第 72 页)

[7] TANG L, CAI Y, LIU J, et al. Few - shot font generation by learning fine - grained local styles[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18 - 24. New Orleans: CVPR, 2022: 7895 - 7904.

[8] SONG C H, HAN H J, AVRITHIS Y. All theattention you need: Global - local, spatial - channel attention for image retrieval [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, January 4 - 8. Hawall: IEEE, 2022: 2754 - 2763.

[9] OIKARINEN T, ZHANG W, MEGRETSKI A, et al. Robust deep reinforcement learning through adversarial loss[J]. Advances in Neural Information Processing Systems, 2021, 34:26156 - 26167. [10] DAVIS C G, WOHL M J, VERBERG N. Profiles of posttraumatic growth following an unjust loss [J]. Death Studies, 2007, 31(8):693 - 712.

[11] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image - to - image translation using cycle - consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision, October 22 - 29. Venice, Italy: ICCV, 2017: 2223 - 2232.

[12] LUO X, HAN Z, YANG L, et al. Consistent style transfer[J]. arXiv preprint arXiv:2201.02233, 2022.

[13] MCALLESTER D, KESHET J. Generalization Bounds and Consistency for Latent - Structural Probit and Ramp Loss[C]//International Conference on Neural Information Processing Systems, December 12 - 17. Granada Spain:NIPS, 2011: 2205 - 2212.

[14] TOMCZAK J, WELLING M. Vae with a vampprior [C]//International Conference on Artificial Intelligence and Statistics, April 9 - 11. Playa Blanca: PMLR, 2018: 1214 - 1223.

[15] HIGGINS I, MATTHEY L, PAL A, et al. beta - vae: Learning basic visual concepts with a constrained variational framework[C]//International conference on learning representations, April 24 - 26. Toulon: ICLR, 2017: 1342 - 1353.

(责任编辑　王楠楠)