文章编号:2096-1472(2023)-07-29-04
DOI:10.19644/j.cnki.issn2096-1472.2023.007.007

基于姿态交换图像生成的行人重识别

沈江霖,魏 丹,罗一平

(上海工程技术大学机械与汽车工程学院,上海 201620)
✉ goatlmljrf@163.com; weidan@sues.edu.cn; lyp777@sina.com

摘 要:不同摄像设备之间存在角度、分辨率等差异,同时行人兼具刚性和柔性的特性且外观易受穿着、姿态、遮挡物和视角等因素的影响。基于此,文章从生成对抗网络与姿态特征等方面对行人重识别问题展开深入研究,提出了一种姿态可交换行人重识别框架(PSG-Net)。该框架将样本中的每个行人编码为姿态代码,视觉代码,通过切换姿态代码,生成高质量的姿态合成图像。在 Market-1501、DukeMTMC-reID 和 CUHK03 数据集上的实验结果表明,该方法实现了识别性能改进,并在 Market-1501 数据集上的排序第一(*rank*-1),结果能达到 95.1%,优于大多数先进的方法。

关键词:行人重识别;生成对抗网络;图像生成
中图分类号:TP181 **文献标识码**:A

Pedestrian Re-identification Based on Pose-switched Image Generation

SHEN Jianglin, WEI Dan, LUO Yiping

(*School of Mechanical and Automobile Engineering, Shanghai University of Engineering Science, Shanghai* 201620, *China*)
✉ goatlmljrf@163.com; weidan@sues.edu.cn; lyp777@sina.com

Abstract: There are differences in angle and resolution between different camera devices, and pedestrians have both rigid and flexible characteristics, and their appearance is easily affected by wearing, posture, obstruction, and perspective. Based on this, this paper conducts in-depth research on pedestrian recognition from the aspects of Generative adversarial network and attitude characteristics, and proposes a gesture exchangeable pedestrian recognition framework (PSG-Net). This framework encodes each pedestrian in the sample as pose code and visual code, generate high-quality pose synthesis images by switching pose codes. The experimental results on Market-1501, DukeMTMC reID, and CUHK03 datasets show that the proposed method has achieved recognition performance improvement, with ranking first (*rank*-1) results reaching 95.1% on Market-1501 dataset, which is superior to most advanced methods.

Keywords: pedestrian re-identification; generative adversarial network; image generation

0 引言(Introduction)

行人重识别是指利用计算机视觉技术判断图像或者视频序列中是否存在特定行人的技术[1]。行人姿态训练鲁棒性是行人重识别模型中的关键问题之一[2]。现有方法仅包含有限数量的姿态变化,因此在训练过程中 ReID 模型容易出现过拟合的情况。与此同时,生成对抗网络在图像生成、图像编辑方面都取得了令人印象深刻的成果。在文献[3]中,生成对抗网络用于生成具有不同背景的样本以增强 ReID 模型,但该工作未考虑各种行人姿态。ZHENG 等[4]使用生成的未标记样本改进生成对抗网络的性能,但是生成样本的严重失真限制了性能改进效果。本文提出了一种姿态可交换行人重识别框架(PSG-Net),该框架将样本中的每一个人编码为姿态代码和视觉代码;通过切换姿态代码,生成高质量的姿态合成图像。在相关数据集上的实验结果表明,本文方法实现了性能改进,并优于大多数

基金项目:国家自然科学基金项目(62101314).

先进的方法。

1 姿态交换图像生成模型(Pose-switched image generation model)

姿态交换图像生成模型主要由生成模块、增强模块、判别模块三个部分组成,如图1所示。

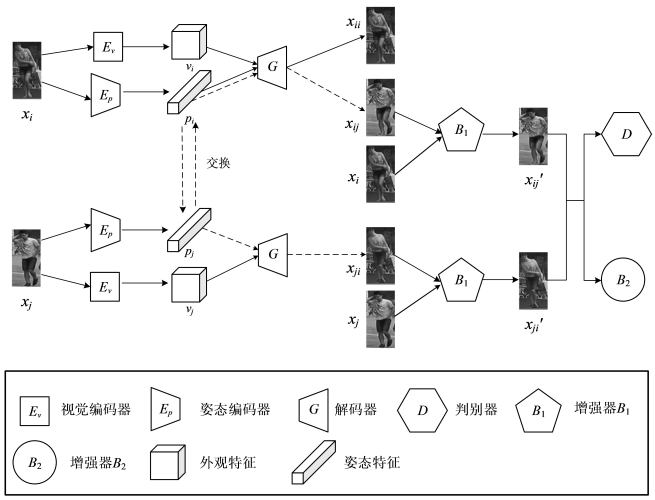

图1 姿态交换图像生成模型

Fig. 1 Pose-switched image generation model

将真实图像表示为 $X=\{x_l\}_{l=1}^{k}$,其中 k 表示图像的数量。给定训练集中的两个真实图像 x_i 和 x_j,生成模块通过交换两个图像的姿态代码生成新的行人图像。如图1所示,生成模块由视觉编码器 $E_v:x_i\rightarrow v_i$;姿态编码器 $E_p:x_j\rightarrow p_j$;解码器 $G:(v_i,p_j)\rightarrow x_{ij}$ 组成。在 $i=j$ 的情况下,生成器可以被视为自动编码器,即 $x_{ii}\approx x_i$。对于生成的图像,本文使用前标表示提供视觉代码的真实图像,后标表示提供姿态代码的真实图像。

1.1 生成模块

生成模块包括两个部分:自我ID生成和交叉ID生成。自我ID生成表示生成模块学习如何从自身重构图像。不同于以相同身份进行图像重建的自我ID生成,交叉ID生成侧重于以不同身份进行图像生成。

1.1.1 自我ID生成

输入两幅不同身份的图像 x_i 和 x_j,基于生成模块中的编码器将每个行人图像分解成两个潜在空间:姿态空间和视觉空间。前者编码姿态(骨架)和骨架关节点位置相关结构信息,后者编码除姿态信息之外的其他身份相关语义信息。由此,行人图像被编码为姿态掩码 p_i、p_j 和视觉掩码 v_i、v_j,通过交换姿态掩码 p_i 和 p_j,利用解码器将视觉掩码和交换后的姿态掩码生成高质量的姿态合成图像 x_{ij} 和 x_{ji}。采用 L_{rec} 表示自我重建图像损失:

$$L_{rec}=E[\|x_i-G(v_i,p_i)\|_1] \tag{1}$$

其中,E 表示期望,G 表示生成器,v_i 表示视觉空间编码得到的视觉特征,p_i 表示姿态空间编码得到的姿态特征。

1.1.2 交叉ID生成

自我身份图像生成以同一身份编码 v_i、p_i 进行图像重建,交叉身份图像生成侧重于以不同身份编码 v_i、p_j 进行图像生成。学习过程中姿态编码 p_i 和 p_j 可以交换信息。采用 L_{cr-id} 表示交叉生成图像损失:

$$L_{cr-id}=E[\|v_i-E_v(G(v_i,p_j))\|_1] \tag{2}$$

其中,E 表示期望,G 表示生成器,E_v 是视觉特征的解码器,v_i 是视觉空间编码 x_i 得到的视觉特征,p_j 是姿态空间编码 x_j 得到的姿态特征。利用解码器将视觉编码和交换后的姿态编码生成姿态合成图像 x_{ij} 和 x_{ji}。

1.2 增强模块

由于生成模块已经生成一幅图像,虽然该图像比较粗糙,但是在姿态和基本颜色上与目标图像接近,因此在增强阶段,模型将通过纠正初始结果中的错误或缺失,专注于生成更多的细节,并且更好地引导图像的生成。增强模块包括图像的细化部分(增强器 B_1)和引导部分(增强器 B_2)。

1.2.1 图像细化(增强器 B_1)

第一阶段对生成具有交叉姿态的行人图像进行外观细节的填充和细化,其输入是生成模块中合成的粗糙图像 x_{ij} 和 x_{ji}。考虑到粗糙图像 x_{ij}、x_{ji} 和目标图像在结构上相似,使用条件DCGAN的衍生模型作为基线。针对全连接层压缩输入中包含的大量信息,移除U-Net(U-网络)中的全连接层,使用U-Net生成一个外观差异映射,保留输入图像中更多的细节,使细化结果更接近目标图像[5]。

在传统的生成对抗网络(GAN)中,判别器负责区分真实图像和生成图像(由随机噪声生成)。然而,在本文的条件网络中,B_1 的输入不是随机噪声而是条件图像 x_{ij}、x_i。因此,真实图像不仅是自然的,而且满足特定的要求。否则,B_1 将被误导为直接输出 x_i,x_i 本身是自然的,而不是细化第一阶段 x_{ij} 的粗略结果。

与传统GAN的另一个不同之处在于,噪声不再是必要的。因此,增强器 B_1 具有以下损失函数:

$$L_{B_1}=L_{bce}(D(x_i,B_1(x_i,x_{ij})),1)+\lambda(L_{rec}+L_{cr-id}) \tag{3}$$

其中,L_{bce} 表示二进制交叉熵损失,D 表示判别器,λ 是生成器损失的权重。

1.2.2 图像引导(增强器 B_2)

针对第一阶段只考虑生成行人样本的视觉真实性,无法保证生成样本能够增强行人重识别模型训练。为此,引出增强模块的第二阶段,即引导生成样本(具有交叉姿态的样本),使经过训练的生成模型更适应行人重识别问题,提高行人重识别的判别能力。增强模块中的引导模块是一个分类(即交叉熵损失)的子网络。将第一阶段生成的图像 x_{ij}' 输入引导模块 B_2 中进行训练。引导模块在目标行人重识别数据集上进行预训练,并进行监督和识别。在生成模块的训练过程中,引导模块传递有判别性的身份信息,并将监督信号从引导模块传递到生成模块。增强器第二部分利用监督信息使得细化后的图像 x_{ij}' 接近生成模块生成的图像 x_{ij}。

用 L_{B_2-ce} 表示交叉熵损失。因此,L_{B_2-ce} 的训练目标可以表示如下:

$$L_{B_2-ce}=E[-\sum d_t \lg q_{B_2}(B_2(v_t,p_t))] \tag{4}$$

其中,d_t 表示类 t 的标签,v_t 表示类 t 图像的视觉特征,p_t 表示 t 类图像的姿态,q_{B_2} 表示增强器 B_2 的输出概率分布。经过细化和引导的生成图像是适应与行人重识别的具有辨识力的各种姿态的标签图像。

1.3 判别模块

通过交换姿态代码生成的图像,将生成的图像视为与现有工作类似的训练样本。为了更好地利用这些生成的图像,可以进行主要特征学习。由于生成模块交叉 ID 合成图像中的类间差异,因此本文采用师生式监督。其中,教师模型只是一个基线卷积神经网络(CNN),在原始训练集上进行识别丢失训练。为了训练用于主要特征学习的判别模块,将判别模块预测的概率分布 $l(x_{ij})$ 和教师模型预测的概率分布 $k(x_{ij})$ 之间的 KL 散度最小化:

$$L_{dis}=E\left[-\sum_{n=1}^{N}k(n\mid x_{ij})\lg\left(\frac{l(n\mid x_{ij})}{k(n\mid x_{ij})}\right)\right] \tag{5}$$

其中,N 表示身份的数量。

因为生成器基于图像 x_i,这同文献[6]的研究结果类似,所以本文对判别器 D 提出以下损失函数:

$$L_D=L_{bce}(D(x_i,B_1(x_i,x_{ij})),0) \tag{6}$$

1.4 优化

整个行人样本生成网络包含三个组件,即生成器、增强器和判别器,本文训练姿态和视觉编码器、解码器、判别器和增强器,用于训练该生成网络的综合损失函数是上述所有损失的加权和:

$$L(E,G,B,D)=L_{B_1}+\alpha L_{B_2-ce}+\beta L_{dis}+L_D \tag{7}$$

其中,α 和 β 是控制相关损失项重要性的权重。在模型的训练过程中,增强器传递鉴别身份信息,并将该监督信号从增强器传播到生成器,从而形成更容易被分类到正确人物类别的行人样本。

2 实验与结果分析(Experiments and analysis of results)

为了验证模型的有效性,本文分别在三个公共行人重识别数据集上进行了实验,其中包括 DukeMTMC-reID[4]、CUHK03[7]和 Market1501[8]数据集。实验表明模型生成的图像更加逼真和多样,并且在所有基准测试中,行人重识别准确度优于大多数现有新算法。

2.1 数据集

DukeMTMC-reID 数据集是 DukeMTMC 数据集的一个子集,用于图像的重识别,它的训练组包含 702 个身份的 16 522 张图像。CUHK03 数据集包含 1 467 个身份的 14 096 张照片,这些照片是由香港中文大学的两台摄像机拍摄的。Market1501 是一个基于图像的 ReID 数据集,它由 12 936 张用于训练的图像组成,每个人在训练集中平均有 17.2 张图像。本文使用两个评估指标评估 ReID 算法的性能,即 *rank*-1 识别率和均值平均精度(*mAP*)。

2.2 实施细节

本文使用通道×高度×宽度表示特征图的大小。编码器 E_p 是一个由 4 个卷积层和 4 个残差块组成浅层网络,输出的是 128×64×32 的姿态代码 p。编码器 E_v 是基于 ImageNet 上预训练的 ResNet-50,移除其全局平均池化层和全连接层,然后附加自适应最大池化层以输出 2 048×4×1 的视觉代码 v。解码器 G 由 4 个残差块和 4 个卷积层组成,每个残差块包含两个自适应实例归一化层,它们集成在一个尺度和偏差参数中。增强器 B_1 包括 $N-2$ 个卷积块的全卷积架构,其中 N 取决于输入的大小。每个残差块由两个步幅为 1 的卷积层和 1 个步幅为 2 的子采样卷积层组成。所有卷积层由 3×3 个滤波器组成,滤波器的数量随每个块线性增加。本文将线性修正单元激活函数(*ReLU*)应用于除全连接层和输出卷积层之外的每一层。增强器 B_2 采用与文献[9]相同的网络架构,鉴别器 D 与文献[10]相同,鉴别器具有简单的堆叠结构。

对于 DukeMTMC-reID 和 Market1501 数据集,使用 Adam 优化器,$\beta_1=0.4$,$\beta_2=0.999$。初始学习率设置为 $e-2$。在 DukeMTMC-reID 上,将卷积块的数量设置为 $N=4$,分别用 8 个小批量的模型训练 10k 次迭代。在 Market-1501 数据集上,将卷积块的数量设为 $N=4$,用 14 个小批量进行 12k 次迭代训练。对于 CUHK03 数据集,使用交叉熵损失训练 ResNet-50。

生成器的输入大小调整为 256×256,并重新缩放为[-1,1],它们来自目标数据集。生成器的输出被发送到鉴别器和引导器。在本文所有实验中,α 和 β 分别设置为 3.0 和 5.0。

2.3 比较结果和讨论

2.3.1 消融研究

首先研究增强器 B_1 和增强器 B_2 的贡献,将提出的方法与 ResNet-50 基线进行比较,结果如表 1 所示。可以观察到,在基线上的性能得到显著改进,主要特征在基线上有很大的改善。除此之外,增强器 B_2 在基线性能上的提升比增强器 B_1 显著,三个数据集上的 *rank*-1 平均提升 11.9%,*mAP* 平均提升 14.9%,结果详见表 1 和表 2。

表1 基线、增强器在 Market1501 与 DukeMTMC-reID 数据集上的组合的比较

Tab. 1 Comparison of baseline and booster on Market1501 and DukeMTMC-reID datasets

方法	Market-1501		DukeMTMC-reID	
	rank-1/%	*mAP*/%	*rank*-1/%	*mAP*/%
Baseline	89.6	74.5	82.0	65.3
Baseline+B_1	90.9	76.2	84.7	67.9
Baseline+B_2	92.7	79.9	85.6	68.6
Baseline+B_1+B_2	**95.1**	**85.4**	**87.3**	**75.2**

表2 基线、增强器在 CUHK03 数据集上的组合的比较

Tab. 2 Comparison of baseline and booster on CUHK03 dataset

方法	CUHK03	
	rank-1/%	*mAP*/%
Baseline	22.2	21.0
Baseline+B_1	28.1	27.0
Baseline+B_2	37.8	36.5
Baseline+B_1+B_2	**47.1**	**45.0**

2.3.2 与先进的方法进行比较

表 3 和表 4 中列出了姿态可交换行人重识别方法(PSG-Net)与其他先进方法的比较结果。与使用单独生成的图像的方法相比,本文方法在 Market-1501 和 DukeMTMC-reID 数据集上的 *rank*-1 实现了明显增益,结果详见表 3。

表3 将所提方法与 Market1501 和 DukeMTMC-reID 数据集上的最新技术进行比较

Tab. 3 Comparison of the proposed method with the state-of-the-art technology on Market1501 and DukeMTMC-reID datasets

方法	Market-1501		DukeMTMC-reID	
	rank-1/%	*mAP*/%	*rank*-1/%	*mAP*/%
Bow+kissme	44.4	20.8	25.1	12.1
SVDNet	82.3	62.1	76.7	56.8
HA-CNN	91.2	75.7	80.5	63.8
APR	84.3	64.7	70.7	51.2
MLFN	90.0	74.3	81.0	62.8
PAN	82.8	63.4	71.6	51.5
Verif-Identif	79.5	59.9	68.9	49.3
Mancs	93.1	82.3	84.9	71.8
Part-aligned	91.7	79.6	84.4	69.3
PCB	93.8	81.6	83.3	69.2
Pose-Transfer(D,Tri)	87.7	68.9	78.5	56.9
DG-Net	94.8	**86.0**	86.6	74.8
本文方法	**95.1**	85.4	**87.3**	**75.2**

基于 ResNet-50 和交叉熵损失，PSG-Net 优于大多数先进方法。对于数据集 CUHK03，PSG-Net 的性能在 *rank*-1 和 *mAP* 两项指标上分别优于排第二的 Pose-Transfer 方法 2.0%、3.0%，结果详见表 4。

表4 将所提方法与 CUHK03 上的最新技术进行比较

Tab. 4 Comparison of the proposed method with the state-of-the-art technology on CUHK03

方法	CUHK03	
	rank-1/%	*mAP*/%
BoW-XQDA	7.9	7.3
LOMO+XQDA	14.8	13.6
PAN	36.9	35.0
DPFL	43.0	40.5
SVDNet	40.9	37.8
Pose-Transfer(R,Tri)	45.1	42.0
本文方法	**47.1**	**45.0**

2.4 参数分析

2.4.1 姿势交换样本数 *N* 的分析

本文分析目标数据集中每个图像的生成样本数如何影响 ReID 模型的性能。使用经过交叉熵损失训练的 ResNet-50 作为增强器，并改进 ReID 模型。对于每个图像，PSG-Net 分别测试 1～10 个姿势交换样本对性能的影响。三个数据集的实验结果如图 2 所示，可以观察到当 $N=4$ 时，验证准确性最高。随着扩展样本的数量进一步增加，性能略有下降。

2.4.2 超参数 μ 的分析

这里的超参数 μ，即 α 和 β 之间的比率，用来控制 L_{B_2-ce} 和 L_{dis} 在训练中的重要性。从 DukeMTMC-reID 数据集的原始训练集中分离出来的验证集上验证参数 μ。根据图 3 中的验证结果，本文在所有实验中选择 $\mu=0.6$。

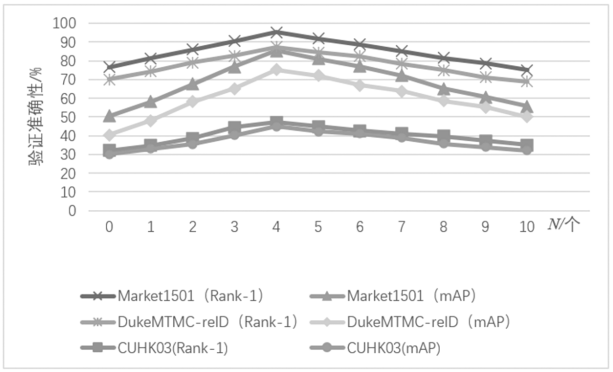

图 2 参数 *N* 对行人重识别模型性能的影响

Fig. 2 The impact of the parameter *N* on the performance of pedestrian re-identification models

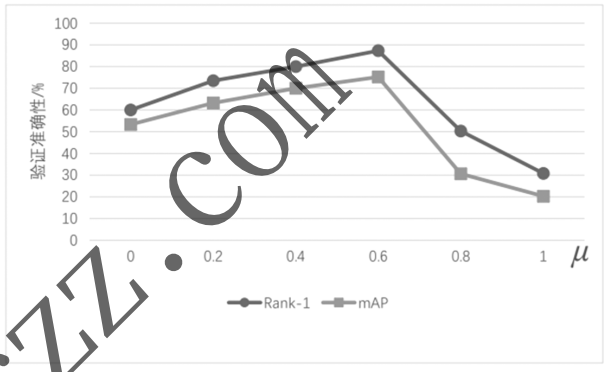

图 3 重识别学习相关超参数 μ 的分析

Fig. 3 Analysis of hyper-parameters μ related to re-identification learning

2.5 可视化结果

本文在图 4 中演示了 PSG-Net 的生成结果，发现 PSG-Net 能够在 Market-1501 数据集中生成逼真和多样的图像。

图 4 通过交换 Market-1501 数据集上的姿态代码生成的图像示例

Fig. 4 Examples of generated images by switching pose codes on the Market-1501 datasets

3　结论(Conclusion)

本文提出了一个姿态可交换行人重识别框架(PSG-Net),解决了现有基准不能提供足够的姿态覆盖训练鲁棒性行人重识别系统的问题。该框架将样本中的每个行人编码为姿态代码和视觉代码,通过切换姿态代码,生成高质量的姿态合成图像。在三个基准上的实验表明,本文提出的方法在图像生成质量和行人重识别精度方面有实质性的改进。

参考文献(References)

[1] 李幼姣,卓力,张菁,等. 行人再识别技术综述[J]. 自动化学报,2018,44(9):1554-1568.

[2] ZHANG Y, JIN Y, CHEN J, et al. PGAN: Part-based nondirect coupling embedded GAN for person reidentification[J]. IEEE Multim, 2020, 27(3): 23-33.

[3] CHEN L, YANG H, WU S, et al. Data generation for improving person re-identification[C]//Association for Computing Machinery. Proceedings of the 2017 ACM on Multimedia Conference. New York: ACM, 2017: 609-617.

[4] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//Institute of Electrical and Electronics Engineers. IEEE International Conference on Computer Vision. Los Alamitos: IEEE, 2017: 3774-3782.

[5] MA L Q, JIA X, SUN Q R, et al. Pose guided person image generation[C]//Institute of Electrical and Electronics Engineers. Annual Conference on Neural Information Processing Systems. Cambridge: MIT press, 2017: 406-416.

[6] MATHIEU M, COUPRIE C, LECUN Y. Deep multi-scale video prediction beyond mean square error[C]//Institute of Electrical and Electronics Engineers. International Conference on Learning Representations. Los Alamitos: IEEE, 2016: 1-14.

[7] LI W, ZHAO R, XIAO T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Institute of Electrical and Electronics Engineers. IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE, 2014: 152-159.

[8] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark[C]//Institute of Electrical and Electronics Engineers. IEEE International Conference on Computer Vision. Los Alamitos: IEEE, 2015: 1116-1124.

[9] LIU J X, NI B B, YAN Y C, et al. Pose transferrable person re-identification[C]//Institute of Electrical and Electronics Engineers. IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE, 2018: 4099-4108.

[10] YAN Y, XU J, NI B, et al. Skeleton-aided articulated motion generation[C]//Association for Computing Machinery. Proceedings of the 2017 ACM on Multimedia Conference. New York: ACM, 2017: 199-207.

作者简介:

沈江霖(1997-),男,硕士生. 研究领域:行人重识别,图像处理.

魏　丹(1982-),女,博士,副教授. 研究领域:行人重识别,图像处理.

罗一平(1965-),男,博士,高级工程师. 研究领域:智能材料.

(上接第 24 页)

[5] ZHAO M, YIN F, LI X, et al. Experimental study of the phase relations in the Co-Si-Zn ternary system at 723 and 873 K [J]. Journal of Alloys and Compounds, 2012, 540(1): 215-221.

[6] LIU Y X, YIN F C, ZHI L I, et al. Experimental determination of 800 ℃ isothermal section in Al-Zn-Zr ternary system[J]. Transactions of Nonferrous Metals Society of China, 2019, 29(1): 25-33.

[7] 胡德林,张帆. 三元合金相图[M]. 西安:西北工业大学出版社,1995:21-36.

[8] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[9] REN X, MALIK J. Learning a classification model for segmentation[C]//IEEE. The Ninth IEEE International Conference on Computer Vision, Nice: IEEE, 2003: 10-17.

[10] KANEZAKIA. Unsupervised image segmentation by backpropagation[C]//IEEE. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary: IEEE, 2018: 1543-1547.

[11] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Springer. The European Conference on Computer Vision, Munich: Springer, 2018: 801-818.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE. The IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City: IEEE, 2018: 7132-7141.

[13] ACHANTA R, SHAJI A, SMITH K, et al. Slic superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[14] LI Z, CHEN J. Superpixel segmentation using linear spectral clustering[C]//IEEE. The IEEE Conference on Computer Vision and Pattern Recognition, Boston: IEEE, 2015: 1356-1363.

作者简介:

刘　玄(1998-),男,硕士生. 研究领域:图像处理与应用,深度学习.

文　勇(1969-),男,硕士,高级工程师. 研究领域:自然语言处,深度学习. 本文通信作者.

梁建烈(1971-),男,博士,教授. 研究领域:核能结构材料,相图相变.

马　坤(1997-),男,硕士生. 研究领域:相图相变.