文章编号:2096-1472(2023)08-0053-05
DOI:10.19644/j.cnki.issn2096-1472.2023.008.012

基于半监督编码的深度卷积对抗网络模型研究

欧莉莉，杜芳芳

(黄河交通学院智能工程学院，河南 焦作 454950)
✉ 1270370009@qq.com; 532637120@qq.com

摘 要:生成对抗网络因为在训练时具有能快速获取真实感、产生大量特征等优点，所以该方法在有监督和半监督的图像识别中逐渐得到广泛应用。文章以提高GAN(Generative Adversarial Networks)模型下的图像识别准确率为目标，基于已有的GAN模型，提出一种基于GAN模型的半监督深度学习模型，并将所建模型放入MNIST、CIFAR-10和Fashion-MNIST三种不同的数据集进行测试，结果显示，SSE-DCGAN模型在三种数据集标签数据较少时，能够很好地识别图像，在三个数据集上识别精度分别达到99.04%、83.66%、89.64%，进行消融实验的结果也表明，在模型中加入编码器后，准确率分别达到0.43%、2.55%、4.44%的提升。

关键词:生成对抗网络;半监督;图像识别;特征匹配
中图分类号:TP391.41 **文献标志码**:A

Research on Deep Convolutional Adversarial Network Model based on Semi-supervised Encoding

OU Lili, DU Fangfang

(*School of Intelligent Systems Engineering, Huanghe Jiaotong University, Jiaozuo* 454950, *China*)
✉ 1270370009@qq.com; 532637120@qq.com

Abstract: Generative adversarial network has been widely used in supervised and semi-supervised image recognition due to its advantages of quickly acquiring realism and generating a large number of features during training. Aiming at improving the accuracy of image recognition under the General Adversarial Networks (GAN) model, this paper proposes a semi-supervised deep learning model based on the existing GAN model, and the proposed model is tested on three different datasets: MNIST, CIFAR-10, and Fashion-MNIST. The results show that the SSE-DCGAN model can effectively recognize images when there is less label data in the three datasets. The recognition accuracy reaches 99.04%, 83.66%, and 89.64% on the three datasets, respectively. The results of ablation experiments also show that after an encoder is added to the model, the accuracy improves by 0.43%, 2.55%, and 4.44%, respectively.

Key words: generative adversarial networks; semi-supervision; image recognition; feature matching

0 引言(Introduction)

图像识别技术是深度学习领域的一项重要内容，在识别图像时，深度学习方法研究的基础是对图像进行处理、分析、理解得出结果，图像的预处理、特征的提取与匹配是图像识别技术非常重要的环节。用计算机进行图像识别，不仅能够提高图像数据的处理效率，而且能对人眼很难观察、提取到的图像信息进行更多细节的辨识[1]。图像识别的方法主要有传统图像识别和深度学习图像识别。传统图像识别方法在识别的过程中，只能对图像的低级边缘信息进行处理，无法获得图像的深层信息，并且需要人工进行预处理，导致图像处理的效率和准确率

收稿日期:2023-02-02

不高。而深度学习图像识别方法能够通过构建多层隐藏层网络，利用计算机自适应地学习图像中的局部细节及图像的空间全局特征，具有很强的识别能力，而且识别图像的准确率也很高。

1 相关理论(Related theories)

2014年，GOODFELLOW等[2]提出了生成对抗网络(Generative Adversarial Networks，GAN)，GAN模型的基本框架由生成器和判别器两个部分组成。在生成器中，主要通过对随机噪音的输入，使得产生的样本更接近实际数据的分布。在此基础上，将数据和产生的样本分别输入判别器，由判别器区分两者，最后得到样本为真值的概率。在2015年提出的深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Networks，DCGAN)模型，是对GAN模型的改进，并首次使用卷积神经网络进行特征提取，从而改善了GAN模型学习的稳定性[3]。2016年，OpenAI发布了一种改进的GAN模型，称为半监督生成对抗网络(Semi-Supervised Generative Adversarial Networks，SSGAN)[4]。在训练SSGAN模型时，它的损失函数是通过有监督和无监督的混合学习实现的，可以提高半监督分类的精度[5-6]。

2 半监督编码深度卷积对抗网络模型研究(Research on semi-supervised coding deep convolutional adversarial network model)

2.1 SSE-DCGAN模型

为了实现对图像更高效和精确的识别，文章提出了半监督编码深度卷积生成对抗网络(Semi-Supervised Encoder Deep Convolutional Generative Adversarial Networks，SSE-DCGAN)模型，该模型将GAN、DCGAN、SSGAN和编码器提取特征等方法相结合，最大限度地发挥每一种方法的优势，建立更准确的图像识别模型。图1为SSE-DCGAN模型的基本架构。

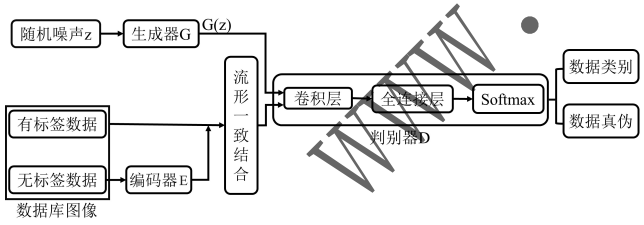


图1 SSE-DCGAN模型基本框架

Fig. 1 Basic framework of SSE-DCGAN model

在SSE-DCGAN模型中使用流形相一致的合并模式，其过程如图2所示，通过将图像数据和特征进行融合，能有效地处理数据间的不匹配问题，降低网络的计算量。

为了避免因批量初始化(Batch Normalization，BN)操作不当造成的训练过程异常问题，对图片进行处理时采用L2范数归一化[7]。在L2范数处理过程中，第 l 层的神经元输出如下：

$$X_{l+1}=f\left\{W_l BN\left[\mathrm{Con}\left(\frac{X_l}{M_1},\frac{z}{M_2}\right)\right]+b_l\right\} \tag{1}$$

将L2范数用于隐变量和特征相结合的归一化操作，在对 $l-1$ 层参数进行逆向求导时，要添加一个导数除以特征模的运算，也就是当SSE-DCGAN模型使用流形一致结合的方式时，可以缩短模型的预计计算时间，进而加快模型的收敛。

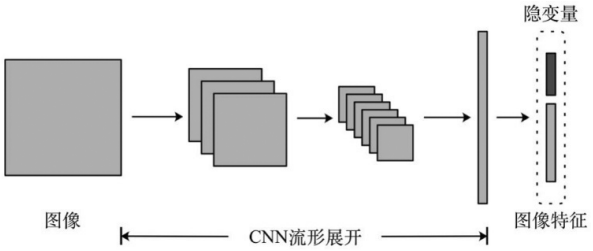


图2 SSE-DCGAN模型中流形一致结合方式

Fig. 2 Consistent binding of manifolds in the SSE-DCGAN model

2.2 SSE-DCGAN模型训练

SSE-DCGAN模型的学习过程就是一个不断优化损失函数的过程。SSE-DCGAN模型中的损失函数包括判别器的损失 L_D、生成器的损失 L_G、编码器的损失 L_E。将三种类型的数据(有标记数据、无标记数据、生成数据)和数据相应的特征输入判别器，并且都有各自对应的损失函数，即有标签数据损失 L_{label}、无标签数据损失 $L_{unlabel}$、生成样本损失 L_{gen}、数据特征(编码器)损失 L_E。如公式(2)所示，有标签数据损失是真实类标签分布和预测类标签的交叉熵损失；无标签数据损失是无标签数据来自真实数据，此时 $y\neq K+1$；生成样本损失是生成器生成的伪样本被判别器判断为假样本的损失，此时 $y=K+1$；数据特征损失编码器用来提取真实数据的特征。

在对判别器损失函数进行优化时，有标记的数据需要被监督的学习，无标记数据、伪样本用于无监督学习。即

$$L_D=L_E+L_{label}+L_{unlabel}+L_{gen}=L_E+L_{supervised}+L_{unsupervised} \tag{2}$$

公式(2)中：

$$\begin{cases}L_{supervised}=L_{label}\\ L_{unsupervised}=L_{unlabel}+L_{gen}\end{cases} \tag{3}$$

编码器损失 L_E 用于解决数据之间的偏移问题，它的损失函数是交叉熵损失，如公式(4)所示：

$$L_E=H(p,q)=-\sum_i^m p(x_i)\lg q(x_i)=-E_{x\sim p}\lg q(x) \tag{4}$$

监督损失 $L_{supervised}$ 主要由 L_{label} 组成，相当于一种有监督的分类工作，对于一个 K 分类问题，网络参数的优化要求使标记数据的样本和模型预测分布 $P_{model}(y|x)$ 之间的交叉熵最小，其表达式如公式(5)所示：

$$L_{label}=-E_{x,y\sim P_{data}}\lg P_{model}(y|x,y<K+1) \tag{5}$$

无监督损失 $L_{unsupervised}$ 主要由 $L_{unlabel}$ 和 L_{gen} 组成。其中，$L_{unlabel}$ 在训练时尽可能最大化无标签数据来自真实数据的概率如公式(6)；以及 L_{label} 尽可能最大化样本来自生成样本的概率如公式(7)：

$$L_{unlabel}=-E_{x\sim P_{data}}\lg[1-P_{model}(y=K+1|x)] \tag{6}$$

$$L_{gen}=-E_{x\sim G}\lg[P_{model}(y=K+1)] \tag{7}$$

对于生成器的损失函数 L_G 来说，为了让生成器对实际数据分布有更好的学习能力，本文使用特征匹配法定义 L_G[8]。在训练的时候，用伪样本和真样本特征匹配的结果作为生成器的损失函数，使损失函数达到最小值，它的定义在公式(8)中已给出。

$$L_G=\| E_{x\sim P_{data}}f(x)-E_{z\sim P_z(z)}f(G(z))\|_2^2 \tag{8}$$

3　实验结果与分析(Experimental results and analysis)

3.1　数据集

本文将 SSE-DCGAN 模型应用于 MNIST、CIFAR-10 和 Fashion-MNIST 数据集进行图像识别研究。三种数据集中的图片类型如图 3 所示。

(a)MNIST　(b)CIFAR-10　(c)Fashion-MNIST

图 3　三种数据集中的图片

Fig. 3 Images of the three datasets

3.2　实验环境

本文的所有实验都是以 Tensorflow 为基础,用 Python 语言完成编程。硬件为通用 PC 机(CPU 3.60 GHz、RAM 32.0 GB);操作系统为 Windows 10 专业版(64 位)。

3.3 SSE-DCGAN 模型验证实验

在 MNIST、CIFAR-10、Fashion-MNIST 三种数据集上进行实验时,实验参数的设置如表 1 所示。

表1　参数设置

Tab.1 Parameter settings

数据集	MNIST	CIFAR-10	Fashion-MNIST
总迭代次数/轮	1 500	1 500	1 500
初始学习率/%	3×10^{-3}	3×10^{-3}	3×10^{-3}
批大小/个	64	64	64
优化器	Adam,动量大小为 0.5	Adam,动量大小为 0.5	Adam,动量大小为 0.5

本文对 SSE-DCGAN 模型的有效性进行分析和验证,具体步骤如下。

3.3.1 MNIST 数据集的实验

在 MNIST 数据集中,为了减少模型占用的运算时间,在训练开始时,先对数据进行归一化处理。为防止模型过拟合,将 Dropout 参数设定为 0.5。有标记样本的数被设定为 100 个、1 000 个和 10 000 个,一共做了三个对照实验,实验对比的结果列于表 2 中。从表 2 中可以看出,当用相同的有标记数据进行测试时,在半监督对抗训练下,SSE-DCGAN 模型能提高图像识别的精度。

表2　MNIST 数据集上测试精度的对比结果

Tab.2 Comparison of test accuracy on the MNIST dataset

模型	ACC/%		
	100 个	1 000 个	10 000 个
CatGAN	98.09	99.10	99.34
Ladder Networks	98.14	99.13	99.39
SSGAN[6]	98.58	99.15	99.45
SSE-DCGAN	99.04	99.23	99.58

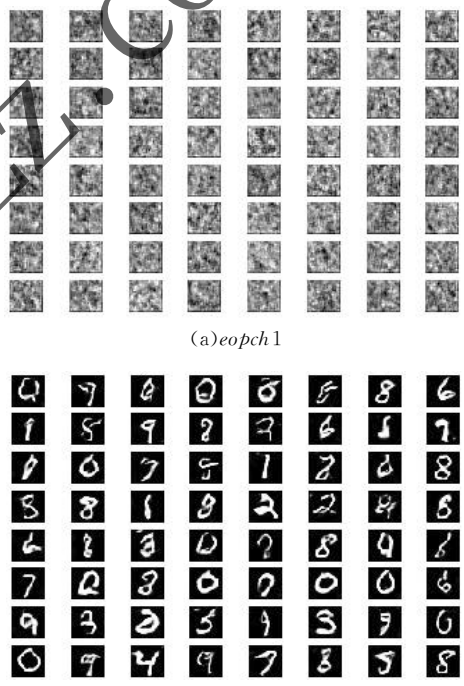
从表 3 中的实验结果可以看出,在模型中加入编码器后,识别精度提高 0.43%,这表明在 SSE-DCGAN 模型中添加编码器对于提高模型的性能是可行的,编码器的引入还可以帮助模型对图像进行更深层次的学习。

表3　MNIST 数据集上的消融实验

Tab.3 Ablation experiments on the MNIST dataset

方法	GAN	GAN+深度卷积网络	GAN+深度卷积网络+半监督方法	GAN+深度卷积网络+半监督方法+编码器
平均准确率/%	99.15	99.24	99.45	99.58

图 4 显示了在训练期间由生成器生成的局部图片的视觉效果:当 $epoch=1$(刚刚开始训练)时,仅能获得一幅模糊的灰度图,但是随着训练次数的增多,各种数值的特征会逐渐呈现出来,其中训练到 1 500 轮时,各种数值的特征呈现更为显著,说明该生成器对实际数据的分布有着较好的拟合效果。



(a)*eopch*1

(b)*epoch*1 500

图 4　部分 MNIST 数据集生成样本图像

Fig. 4 Sample images generated by part of MNIST dataset

3.3.2 CIFAR-10 数据集的实验

CIFAR-10 是一种自然图像数据集,它的图像中包含非常复杂和丰富的细节,生成器中 Dropout 系数的设定可能会在训练时产生误差,因此本文对 SSE-DCGAN 模型进行了大量的实验,最终将 Dropout 系数设定为 0.3。根据这个系数,本文进行了 3 组不同的实验,有标记的样本数为 1 000 个、2 000 个和 4 000 个,实验结果列于表 4 中。由表 4 可知,相比较于其他模型,SSE-DCGAN 模型在半监督对抗学习条件下具有更高的识别率。

表4 CIFAR-10 数据集上测试精度的对比结果

Tab.4 Comparison of test accuracy on the CIFAR-10 dataset

模型	ACC/%		
	1 000 个	2 000 个	4 000 个
Ladder Networks	76.52	79.31	83.06
CatGAN	78.83	80.42	88.90
Bayesian GAN	81.25	82.88	89.94
SSE-DCGAN	83.66	85.14	91.78

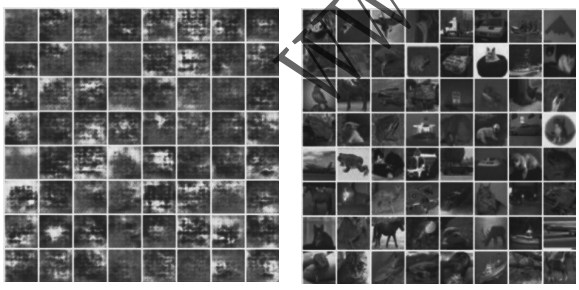
为了测试模型在增加了一个编码器之前和之后的辨识能力，消融实验选择在 CIFAR-10 数据集有标记数据为 4 000 时进行，对比各种模型的实验结果见表 5：在 GAN 模型中添加编码器后，模型的预测准确率从 89.23%上升到 91.78%，提升了 2.55%。实验结果显示，编码器加入 SSE-DCGAN 模型中，能够改善实际数据中的图像特征，从而使 SSE-DCGAN 模型不但能对更复杂的图像进行有效的处理，还能在一定程度上提高图像的识别精度。

表5 CIFAR-10 数据集上的消融实验

Tab.5 Ablation experiments on the CIFAR-10 dataset

方法	GAN	GAN+深度卷积网络	GAN+深度卷积网络+半监督方法	GAN+深度卷积网络+半监督方法+编码器
平均准确率/%	89.23	89.78	90.85	91.78

图 5 是生成器在不同的训练次数时产生的局部图像。可以从图中看到，每一幅图像的特征都越来越清楚，特别是模型训练到 1 500 轮时，每一幅画的特征都很明显。此外，随着训练次数的增加，模型趋于稳定，生成器的性能增强，所生成的虚假样本可以骗过判别器，与实际数据一起训练判别器，最后得到准确的判别表面。



(a)*epoch* 50 (b)*epoch* 1 500

图 5 部分 CIFAR-10 数据集生成样本图像

Fig. 5 Sample images generated by part of CIFAR-10 dataset

3.3.3 Fashion-MNIST 数据集实验

在 Fashion-MNIST 数据集上，与 MNIST 数据集上采用相同的参数设置。同时，在此数据集中，有标记的样本数为 1 000 个和 10 000 个。表 6 中列出了实验结果：在使用相同数据和同样带标记的数据时，SSE-DCGAN 模型有效地提升图像识别准确率，获得优于对比模型的识别准确率。

表6 Fashion-MNIST 数据集上测试精度的对比结果

Tab.6 Comparison of test accuracy on the Fashion-MNIST dataset

模型	ACC/%	
	1 000 个	10 000 个
CatGAN	84.82	89.41
AAE	86.34	90.8
SSGAN	88.6	91.67
SSE-DCGAN	89.64	93.46

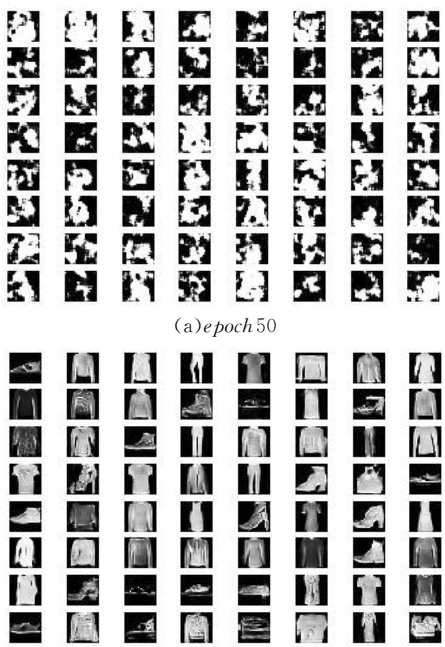
为了检验 SSE-DCGAN 模型对图像的处理效果，消融实验是在标记数据为 10 000 个时进行的，将使用编码器之前和之后的模型识别结果进行比较，实验结果见表 7。四种识别 Fashion-MNIST 数据集模型的准确率不断提高，特别是加入编码器之后，SSE-DCGAN 模型的精度从 89.02%到 93.46%提高了 4.44%。

表7 Fashion-MNIST 数据集上的消融实验

Tab.7 Ablation experiments on the Fashion-MNIST dataset

方法	GAN	GAN+深度卷积网络	GAN+深度卷积网络+半监督方法	GAN+深度卷积网络+半监督方法+编码器
平均准确率/%	89.02	89.71	91.67	93.46

图 6 显示了该生成器在训练期间生成的一部分图像。从图中我们可以看到，训练到 50 轮次时，生成的图片都比较模糊，没有 T 恤和裙子之类的特征。然而，随着训练轮次不断增加，图像中的特征会越来越显著，当 *epoch* 为 1 500 时，这些特征在图片上更明显，这也说明生成器可以很好地对真实数据进行仿真。



(a)*epoch* 50

(b)*epoch* 1 500

图 6 部分 Fashion-MNIST 数据集生成样本图像

Fig. 6 Sample images generated by part of Fashion-MNIST dataset

4 结论(Conclusion)

本文将半监督学习引入到生成对抗网络，并且针对生成对抗网络的不足之处，提出了一种基于生成对抗网络的新算法，并通过在三种不同类型的数据集上进行试验，证明了新模型的优

势。尽管文中提出的模型可以有效地提高图像识别的准确率,但是仍然存在不足之处:(1)增加了训练模型的参数。采用编码器结构后,虽然提高了图像识别的准确率,但是学习过程中存在大量的参数,使得模型耗时较长。在以后的研究中,可以对该模型进行改进,从而提高其计算效率和识别结果的准确性。(2)近几年来,针对生成对抗网络的研究多集中于对生成器或判别器的改进,该领域尚无新的研究方法,今后可考虑将生成对抗网络与传统的机器学习方法相结合,以达到半监督对抗训练的目的。

参考文献(References)

[1] 欧莉莉. 基于半监督编码的深度卷积对抗网络模型研究及应用[D]. 青岛:青岛大学,2021.

[2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems. Proceedings of the 27th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2014: 2672-2680.

[3] 常亮,邓小明,周明全,等. 图像理解中的卷积神经网络[J]. 自动化学报,2016,42(9):1300-1312.

[4] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training gans[C]//Conference on Neural Information Processing Systems. In Proceedings of Advances in Neural Information Processing Systems. Massachusetts: MIT Press, 2016: 2234-2242.

[5] 欧莉莉,邵峰晶,孙仁诚,等. 基于半监督方法的脑梗死图像识别[J]. 计算机应用,2021,41(4):1221-1226.

[6] 张营营. 生成对抗网络模型综述[J]. 电子设计工程,2018,26(5):34-37,43.

[7] ZHENG L, WANG S J, TIAN L, et al. Query-adaptive late fusion for image search and person re-identification[C]// IEEE Conference on Computer Vision and Pattern Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1741-1750.

[8] 唐贤伦,杜一铭,刘雨微,等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. 自动化学报,2018,44(5):855-864.

作者简介:

欧莉莉(1993-),女,硕士,助教。研究领域:图像处理,大数据分析。

杜芳芳(1988-),女,硕士,讲师。研究领域:图像处理,模式识别。

(上接第 39 页)

[5] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2016: 770-778.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2014: 580-587.

[7] GIRSHICK R. Fast R-CNN[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Computer Society, 2015: 1440-1448.

[8] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// Institute of Electrical and Electronics Engineers. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Computer Society, 2020: 390-391.

[9] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2017: 2117-2125.

[10] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2018: 8759-8768.

[11] 蓝金辉,王迪,申小盼. 卷积神经网络在视觉图像检测的研究进展[J]. 仪器仪表学报,2020,41(4):167-182.

[12] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[13] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[DB/OL]. (2022-05-25)[2023-03-09]. https://arxiv.org/abs/2205.12740.

[14] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. (2015-04-10)[2023-03-09]. https://arxiv.org/abs/1409.1556.

[16] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Springer Science. Proceedings of the 15th European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

作者简介:

王耀宗(1998-),男,硕士生。研究领域:软件开发,计算机视觉。

张易诚(2001-),男,本科生。研究领域:计算机视觉。

康宇哲(1998-),男,硕士生。研究领域:计算机视觉。

沈　炜(1973-),男,博士,教授。研究领域:人工智能,计算理论。