

引用格式:赖远哲,陈向阳,李旭东,等. 基于残差结构的 GAN 网络的显著性预测研究[J]. 微电子学与计算机, 2021, 38(8): 95-100. [LAI Y Z, CHEN X Y, LI X D, et al. Research on saliency prediction of GAN network based on residual structure[J]. Microelectronics & Computer, 2021, 38(8): 95-100.]

基于残差结构的 GAN 网络的显著性预测研究

赖远哲¹, 陈向阳^{1,2}, 李旭东¹, 付星堡¹, 曹倩倩¹

(1 武汉工程大学 计算机科学与工程学院, 湖北 武汉 430205;

2 武汉工程大学 智能机器人湖北省重点实验室, 湖北 武汉 430205)

摘 要: 优化了简单生成对抗网络结构, 用于更有效的通过对抗性实例训练得到视觉显著性图, 减少假阳性产生和提高显著性. 网络模型仍遵循传统生成对抗网络结构, 第一阶段是由一个使用残差结构构建的生成器组成, 该模型的权值由显著图的下采样版本的二分类交叉熵损失(BCE)的反向传播计算得到的, 训练得到更有效的显著图. 预测结果由受训练的判别器网络进行生成阶段生成的显著图与真值图之间的二值分类处理. 实验展示了改进生成对抗网络中的生成器的预测显著图的能力对整个网络性能提升, 相较于其他显著图预测模型也有一定领先.

关键词: 视觉显著性; 生成对抗网络; 残差结构; 交叉熵

中图分类号: TP183

文献标识码: A

文章编号: 1000-7180(2021)08-0095-06

Research on saliency prediction of GAN network based on residual structure

LAI Yuanzhe¹, CHEN Xiangyang^{1,2}, LI Xudong¹, FU Xinbao¹, CAO Qianqian¹

(1 School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, Hubei, China;

2 Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, Hubei, China)

Abstract: The structure of simple generated countermeasure network is optimized, which is used to obtain visual saliency map more effectively through antagonistic case training, so as to reduce false positive and improve saliency. The network model still follows the traditional generation countermeasure network structure. In the first stage, it is composed of a generator built with residual structure. The weight of the model is calculated by the back-propagation of the binary cross entropy loss (BCE) of the down sampling version of saliency map, and the more effective saliency map is obtained by training. The prediction results are classified by the trained discriminator network between saliency map and truth graph. The experiment shows that the improved ability to generate predictive saliency map of generators in the countermeasure network can improve the performance of the whole network, and it is also ahead of other saliency map prediction models.

Key words: Visual salience; generating adversarial network; residual structure; cross entropy

1 引言

显著性预测是由 Treisman 和 Gelade 提出的特征整合理论^[1]发展而来. 随着计算机的发展, 90 年

代, Itti 等人定义了第一个预测图像显著性的计算模型^[2], 该模型基于多尺度低层次特征的显著性图预测方法形成显著性图. Harel 等人从低层特征映射出发, 引入了一种基于图的显著性模型^[3] Judd 等

人^[4]在每个图像位置提取低水平和自上而下的特征,并训练分类器来确定每个位置是否显著.上述机器学习方法的局限性在于,它们在很大程度上依赖于学习的特性.这可以称为浅层学习.相比之下,深度学习方法能够从原始像素数据中学习丰富的特征层次结构. Vig 等人提出的 Deep Networks(eDN)集成模型^[5],采用 3 层卷积稀疏编码模型和分层神经网络来学习有效的图像特征.此后, Kümmerer 等人提出了两种深度显著性预测网络^[6]:第一种是基于 AlexNet 模型的 DeepGazeI,第二种是基于 19 层 VGG 网络^[7]的 DeepGaze II^[8],并用 SALICON^[9]数据集进行训练取得了较好的结果. Liu 等人提出一种端到端显著性检测模型模型 DSCLSTM^[10],结合全局上下文和场景上下文来推断显著性. Cornia 等人提出的显著性预测架构包含一个专注的卷积长短期记忆网路(Attentive ConvLSTM)^[11],并探讨了两种 CNN 网络对模型的影响.此后, Pan 等人提出了一种通过对抗性训练的用于显著性预测的深层网络(SalGAN)^[12]. 探讨了对抗网络生成器与鉴别器之间的对抗损失,使用广泛使用的损失函数(如 BCE)相结合时,对抗性训练如何在不同的度量标准下实现最先进的性能.

随着深度卷积神经网络(DCNN)的发展和大规模显著性数据集出现,基于 DCNN 的显著性模型在 MIT 显著性基准上取得了优异的性能. 为了充分利用优秀二维模型的优点,本文提出了基于 ResNet 残差结构的的生成对抗网络视觉显著性预测模型,本文的主要工作是:1)将传统生成对抗网络结构进行改进,融入 ResNet50 残差网络构建生成器;2)对 ResNet 结构进行调整以使得生成器更有效地通过训练输出显著图;3)调整的生成器与对抗训练损失函数,考虑了更有效的评价指标,对综合损失函数中的超参进行调整优化其在特定图像上的性能.

2 模型结构

网络模型的第一阶段是由一个生成器模型组成,其权值先用随机数函初始化再由显著图的下采样图上计算二进制交叉熵 BCE 损失的反向传播来学习. 所得到的显著图再由受训练的判别器网络与真值图之间进行二值分类. 整个模型结构如图 1.

2.1 生成器

生成器遵循经典的编解码结构,编码器包括卷积层和对特征进行压缩、简化网络复杂度的池化层,解码器部分使用上采样层,然后使用卷积操作来构

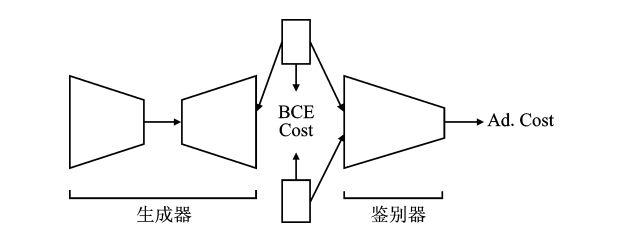


图 1 整体模型结构图

造与输入具有相同分辨率的输出. 调整的编码器部分在结构上与 Resnet50 残差网络相似. ResNet-50 网络由几个堆叠层组成的块之间执行一系列残差映射,即将块的输入添加到它的输出中^[13]. 这有助于避免随着网络深度的增加而出现的精度下降问题,并且在显著性预测的情况下能够提高网络的特征提取能力. 残差模块由 identity_block 和 conv_block 组成,省略了最后的池化层和全连接层. 使用在 ImageNet 数据集上训练的 Resnet-50 模型的权重对网络进行初始化. 在训练期间固定权重以节省计算资源. 为适应显著性预测,对 Resnet 网络参数进行调整. 编码器结构图如图 2.

而解码器的结构和编码器相同但是层的顺序是相反的,平均池化层改成上采样层,所有卷积层的激活函数改为 Leaky ReLU,最后一层卷积层之后添加一个 1x1 的卷积层激活函数设为 sigmoid,以生成显著图. 解码器的权重直接随机初始化.

2.2 鉴别器

鉴别器网络有三个块组成,每个块由两个 3x3 内核卷积层和一个最大池化层组成. 三个块之后是三个完全连接层,卷积层都使用 ReLU 函数激活,前两个完全连接层使用 tanh 函数激活,最后一个采用 sigmoid 激活函数. 具体配置如表 1.

表 1 鉴别器网络结构

layer	depth	kernel	activation
conv1_1	3	1x1	Relu
conv1_2	32	3x3	Relu
pool1		2x2	
conv2_1	64	3x3	Relu
conv2_2	64	3x3	Relu
pool2		2x2	
conv3_1	64	3x3	Relu
conv3_2	64	3x3	Relu
pool3		2x2	
fc4	100		tanh
fc5	2		tanh
fc6	1		sigmoid

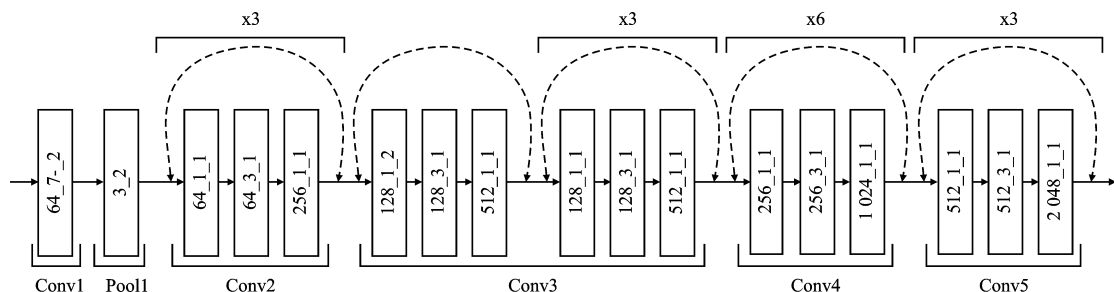


图 2 生成器编码器结构图

3 训练

训练过程中模型的滤波器权重是通过生成器的生成图片的内容损失和与鉴别器对抗训练而产生对抗性损失结合的综合损失进行训练^[14]。其内容损失由预测显著性图与对应的真实图像进行逐像素比较得到。对抗性损失取决于判别器对生成的显著性图的真实和综合预测。

3.1 内容损失

内容损失是以预测显著图的每个值与正确打标签的训练数据对应的点逐像素比较计算得出的。给定图片 I , 显著图 S 表示为概率向量, 其中 S_j 是像素 I_j 被固定的概率。内容损失函数 $L(S, S^*)$, 该函数位于预测的显著图 S^* 与其对应的真实训练数据 S 之间。标记的真实数据显著图每个值归一化到 $[0, 1]$ 之间。每个像素点预测值视为独立于其他点, 在最后一层的每个输出中每个元素用 Sigmoid 函数激活, 这样逐像素预测就可以被认为是独立的二分类随机变量的概率。使用二分类交叉熵作为损失函数, 即所有像素之间单个二分类交叉熵(BCE)的平均值:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{j=1}^N S_j \log(\hat{S}_j) + (1 - S_j) \log(1 - \hat{S}_j) \log(1 - \hat{S}_j) \quad (1)$$

3.2 对抗损失

对于显著性预测, 生成器输入普通图像生成逼真的显著性图, 而不是通过随机噪声生成真实图片, 显著性图所对应的图像评估生成器预测质量, 需要将对应的图片和显著性图同时作为鉴别器的输入。模型使用判别器的误差与交叉熵相结合的生成器的损失函数, 以便提高对抗训练的稳定性 and 收敛速度, 表示为:

$$L_{\text{GAN}} = \alpha \cdot L_{\text{BCE}} - \beta \cdot L(D(I, \hat{S}), 0) \quad (2)$$

式中, $D(I, \hat{S})$ 是欺骗鉴别器的概率, 对抗训练中欺骗鉴别器的概率越低, 生成器的损失越高。在训练判

别器时, 损失函数为:

$$L_D = L(D(I, S), 1) + L(D(I, \hat{S}), 0) \quad (3)$$

在训练期间, 首先用 BCE 训练 15 个周期来引导生成器函数。在此之后, 并入鉴别器, 并开始对抗性训练。

通过 SALICON 训练集总共 10 000 幅图像设置 batch_size 设置为 32 训练网络。在对抗性训练中, 在每次迭代后交替训练生成器和判别器。在训练生成器和鉴别器时, 为防止数据过拟合, 采用 L2 权值正则化(即权衰减), 设置正则项系数 $\lambda = 1 \times 10^{-4}$ 。采用 Adadelta 进行优化。

4 实验分析

4.1 数据集

本文在两个数据集上进行评估, 第一个数据集是 SALICON, 包含 10,000 张训练图、5,000 张验证图和 5,000 张测试图。所有图片微软 CoCo 数据集中选择^[15], 像素为 480×640 。该数据集作为最大的眼睛注视数据集, 通过收集鼠标点击位置来模拟人眼注释。

第二个用来评估的数据集是 MIT300。它包含 300 幅具有自然室外或室内场景的图像。将 MIT 1003 数据集用作 MIT 300 的训练和验证集。这个数据集包含 779 幅风景图片和 228 幅肖像图片。

4.2 评价指标

目前有多种眼睛注视预测的评价指标, 包括陆地移动距离(EMD), 归一化扫描路径显著性(NSS), 皮尔逊相关系数(CC), 相似度(SIM), 曲线下面积(AUC), shuffled 曲线下面积(sAUC), 相对熵(KL), 信息增益(IG)等等。这些指标中, KL、IG 和 SIM 对假阴性最敏感, AUC 指标忽略了低值的假阳性, EMD 的性能取决于空间距离, NSS 和 CC 受假阳性和假阴性的影响相同。本文使用 CC, NSS, AUC, sAUC 来评估提出的方法性能并和其他方法

比较.

4.3 对抗训练

显著性预测可以视为一个多重二元分类问题. 如果将目标和预测均解释为概率,那么最小化交叉熵等效于最小化预测分布与目标分布之间的 KL 散度,因此在非对抗训练时直接使用 BCE 作为内容损失.

本文在预测显著图上进行下采样来计算内容损失,以减少在训练和测试时所需要的计算机资源. 在详细对比下采样因子从 1/2 到 1/16,如表 2,在 1/4 时改进效果最佳.

对于 4.2 节中提到的对抗训练引入的生成器损失,超参数 α 对模型信息增益有一定影响,设置 $\alpha=0.005, \beta=0.5$,调整超参数 α 和 β 的值对抗训练预期有影响.

表 2 验证评估了(15 个时期)下采样显著性图的影响. BCE/ x 为 256×192 的显著性图上 $1/x$ 的下采样因子

	sAUC	AUC-B	NSS	CC
BCE	0.756	0.827	2.476	0.763
BCE/2	0.754	0.822	2.528	0.766
BCE/4	0.762	0.834	2.517	0.765
BCE/8	0.760	0.829	2.509	0.762
BCE/16	0.758	0.828	2.506	0.758

对于提出的模型,直接以二进制交叉熵(BCE)为内容损失函数,对预测的显著图下采样到 64×48 ,评估将加入鉴别器后引入的对抗损失所获得的综合增益. 表 3 中的第三行结果指的是仅在 20 个周期内仅用 BCE 作为内容损失训练 Res-GAN 确定的基线. 随后,考虑引入对抗性损失(第 4 排).

表 3 通过非对抗性训练和对抗性训练获得的最佳成绩在 SALICON 数据集比较基于 resnet50 结构的生成对抗网络模型与原有模型

	sAUC	AUC	NSS	CC
BCE	0.753	0.825	2.562	0.772
SalGAN	0.773	0.859	2.560	0.786
BCE Res	0.756	0.857	2.654	0.791
Res-GAN	0.779	0.864	2.675	0.816

随着迭代周期的增加,图 3 比较了使用改进后的模型结合 ResGAN 和 BCE 的综合损失与单独使用 BCE 进行训练的验证集准确性如何变化,并整体与之前的模型进行对比. 在 AUC Judd 和 shuffled AUC 评价指标下,单独使用 BCE 作为内容损失时,增加迭代周期的数量将会得到一个稳健的度量的改进,相比原有基于 VGG-16 的生对抗网络模型在这

一指标上的改变不明显,有着基准上的提升. 而使用 BCE/Res-GAN 复合损失随着进一步的训练在前期有一个下降后持续改善性能,这是因为 sAUC 旨在对考虑到眼球注视中存在的中心偏差的模型进行惩罚,但所有度量均领先与原有的 VGG 结构的模型结果.

改进后的生成对抗网络模型相对提高性能的指标是标准化扫描路径显著性(NSS). 尽管相对于 VGG16 模型结构,改进的 Resnet50 结构生成器在预测显著性图上有更好的效果,但仍然遵循一个生成器-鉴别器结构,使得 GAN^[16] 训练过程中趋向于产生更平滑,更分散的显着性估计,以更好地匹配真实显著图的统计性质,一定程度上能减少假阳性的出现,但仍不能完全避免. 假阳性率高的影响取决于最终应用,在显著性图被用作多重注意力模型的应用中,例如,在检索应用中,空间特征需要进行加权,误报往往不如假阴性重要,前者包含更多的干扰因素,而后者会去除潜在的有用特征.

在 100 至 120 个迭代周期之后,Res-GAN/BCE 的综合损失函数训练的网络在 4 个指标中比单纯使用 BCE 有实质性的改善.

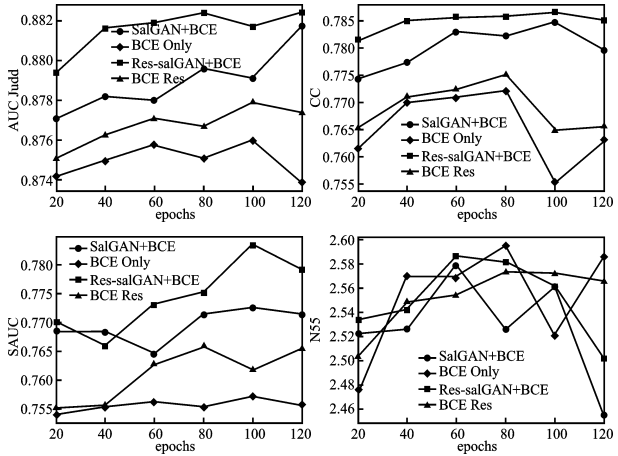


图 3 SALICON 验证集上 Res-GAN+BCE 和单纯 BCE 在不同迭代周期下的精度指标,与 SalGAN 相对应的对比

4.4 与其他模型对比

在本小节中,将 SALICON 测试数据集和 MIT300 数据集上将提出的方法模型与其他几个最先进的显著性检测模型进行了比较,这两个数据集都旨在捕获视觉显著性,但是数据的获取方式却不同, SALICON 的标记真实数据是的鼠标点击产生的,而 MIT300 则是通过眼动追踪器在有限的用户群上构建的,因此,相同指标上一个模型的优势会有差异. 本文在 CC、sAUC、AUC

和 NSS 这几个指标上与 ML-Net^[17], DSCLRCN, DeepGaze II, SalGAN, SAM-VGG 和 SAM-Res-Net^[11]等先进模型对比. 如表 4 中所示,在 SALICON 数据集上,与原有使用 VGG16 结构构建生成器的 SalGAN 模型相比,CC, sAUC, AUC, NSS 四个指标分别提升了 7%, 1. 1%, 2. 2%, 5. 1%. 其中 CC 和 AUC 两个指标上的表现优于其它竞争对手.

表 4 SALICON TEST 数据集上的比较结果.

以粗体显示的结果是表明此评估指标的最佳方法

	CC	sAUC	AUC	NSS
ML-Net	0. 743	0. 768	0. 866	2. 789
DeepGaze II	0. 509	0. 787	0. 885	1. 336
DSCLRCN	0. 831	0. 776	0. 884	3. 157
SalGAN	0. 781	0. 772	0. 884	2. 459
SAM-VGG	0. 825	0. 774	0. 881	3. 143
Propoesd	0. 836	0. 781	0. 886	3. 044

如表 5 在 MIT 数据集上,提出的方法也有着优异的表现,至少一个度量范围上超出或持平所有其他模型的性能.

表 5 MIT300 数据集上的比较结果

	CC	sAUC	AUC	NSS
GBVS	0. 48	0. 63	0. 81	1. 24
BMS	0. 55	0. 65	0. 83	1. 41
PDP	0. 70	0. 73	0. 85	2. 05
DSCLRCN	0. 80	0. 72	0. 87	2. 35
SalGAN	0. 73	0. 72	0. 86	2. 04
SAM-VGG	0. 77	0. 71	0. 87	2. 30
SAM-ResNet	0. 78	0. 70	0. 87	2. 34
Proposed	0. 75	0. 73	0. 87	2. 23

4. 5 定量分析

通过观察模型产生的显著性图,从定性的角度探讨了生成器结构调整对抗网络带来的影响. 图 4 第一张图中显示了将生成器结构改为残差网络后,对轮廓特征明显区域标记显著更加明显. 第二到第四张图显示了改进后的模型相较于前作对于假阳性上的改进. 最后一张图显示预测显著图仍存在一定假阳性比例,这是因为数据集的一个限制.

5 结束语

本文通过实验证明生成对抗网络模型结构对视觉显著性预测的有效性,可以预测自然图像上的人眼注视. 深度卷积神经网络上进行对抗性训练增强

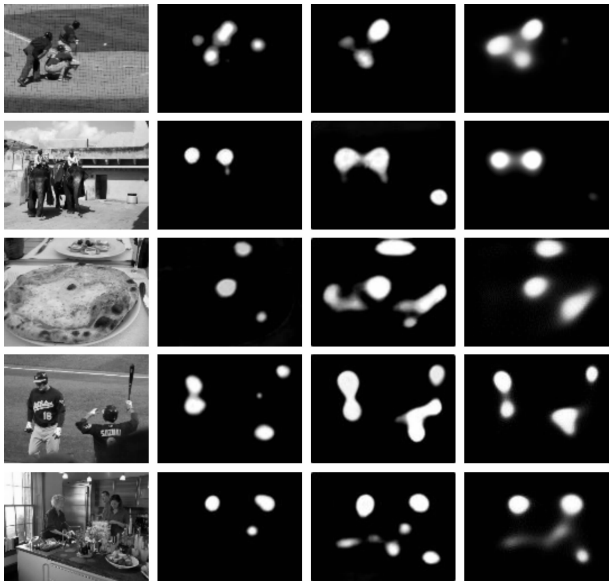


图 4 SALICON 验证集上 SalGAN 与 Res-GAN 的

定性结果, 图片经过一定处理以对比明显

了显著性预测. 生成对抗网络的训练方法是通用的, 因此改进了生成器的结构, 使用相比 VGG16 特征提取能力更强的 ResNet 残差网络结构构建编码器解码器, 获得了比原有结构更好的结果, 同时, 本文通过优化残差网络结构以更适应于显著性预测. 通过广泛的评价, 验证了改进模型的有效性, 提出模型在两个最重要的显著性预测数据集上取得了不错的结果. 由于 Resnet 残差结构能够改进模型随着深度加深而精度下降的问题, 所以加深层数可能会获得更好的效果但也会付出更多的性能. 最后, 近年来 CycleGAN 的提出, 使得生成对抗网络在很多应用上取得更进一步的效果, 由于其需要双输入, 在视觉显著性预测上过程较为复杂, 未来通过调整模型结构和训练方法, 可能会获得更好的效果, 这也是发展方向.

参考文献:

[1] TREISMAN A M, GELADE G. A feature-integration theory of attention[J]. Cognitive Psychology, 1980, 12 (1): 97-136. DOI: 10. 1016/0010-0285 (80) 90005-5.

[2] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (11): 1254-1259. DOI: 10. 1109/34. 730558.

[3] HAREL J, KOCH C, PERONA P. Graph-based visual saliency[C]//Proceedings of the 19th International Conference on Neural Information Processing Systems.

- Cambridge: MIT Press, 2006: 545-552.
- [4] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 2106-2113. DOI: 10.1109/ICCV.2009.5459462.
- [5] VIG E, DORR M, COX D. Large-scale optimization of hierarchical features for saliency prediction in natural images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 2798-2805. DOI: 10.1109/CVPR.2014.358.
- [6] KÜMMERER M, THEIS L, BETHGE M. Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet[J]. arXiv preprint arXiv:1411.1045, 2014.
- [7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [8] JIANG M, HUANG S S, DUAN J Y, et al. SALICON: Saliency in context[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston: IEEE, 2015. DOI: 10.1109/CVPR.2015.7298710.
- [9] LIU N, HAN J W. A deep spatial contextual long-term recurrent convolutional network for saliency detection[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3264-3274. DOI: 10.1109/TIP.2018.2817047.
- [10] CORNIA M, BARALDI L, SERRA G, et al. Predicting human eye fixations via an LSTM-based saliency attentive model[J]. IEEE Transactions on Image Processing, 2018, 27(10): 5142-5154. DOI: 10.1109/TIP.2018.2851672.
- [11] PAN J T, SAYROL E, GIR? -I-NIETO X, et al. SalGAN: Visual saliency prediction with generative adversarial networks[C]//CVPR SUNw: Scene Understanding Workshop 2017. Hawaii: IEEE, 2017.
- [12] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]//European Conference on Computer Vision. Cham: Springer, 2016: 694-711. DOI: 10.1007/978-3-319-46475-6_43.
- [13] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755. DOI: 10.1007/978-3-319-10602-1_48.
- [14] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014: 2672-2680.
- [15] CORNIA M, BARALDI L, SERRA G, et al. A deep multi-level network for saliency prediction[C]//2016 23rd International Conference on Pattern Recognition (ICPR). Cancun, IEEE, 2016: 3488-3493. DOI: 10.1109/ICPR.2016.7900174.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.
- [17] CORNIA M, BARALDI L, SERRA G, et al. A deep multi-level network for saliency prediction[C]//2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016: 3488-3493.

作者简介:

赖远哲 男,(1996-),硕士研究生.研究方向为计算机视觉. Email:517036791@qq.com.

陈向阳 男,(1969-),硕士,副教授.研究方向为网络科学与工程.

李旭东 男,(1995-),硕士研究生.研究方向为自然语言处理.

付星堡 男,(1996-),硕士研究生.研究方向为图像处理.

曹倩倩 女,(1993-),硕士研究生.研究方向为自然语言处理.