基于序列对抗域适应的智能裁剪算法

王皓文 桑农

摘 要 智能裁剪任务一直受到缺乏训练数据的困扰,目前还局限于公开数据集中.因为实际应用场景与训练场 景之间存在域迁移,文中提出基于序列对抗域适应的智能裁剪算法.首先,通过实验证实裁剪数据集 GAICD 和 CPC 之间存在域迁移问题.然后,构造由美学评分模块和对抗域适应模块组成的算法.美学评分模块用于预测图像的美 学评分,并辅助提取面向裁剪任务的不变特征.对抗域适应模块实现基于对抗的域适应学习.不同裁剪数据集之间 的域迁移实验及室内/室外场景之间的域迁移实验均验证文中算法的有效性.

关键词 图像智能裁剪,域迁移,域适应,对抗学习,不变特征
引用格式 王皓文,桑 农.基于序列对抗域适应的智能裁剪算法.模式识别与人工智能,2021,34(8):677-688.
DOI 10.16451/j. cnki. issn1003-6059.202108001
中图法分类号 TP 391.41

Listwise Adversarial Domain Adaption Algorithm for Image Cropping

WANG Haowen¹, SANG Nong¹

ABSTRACT Image cropping is short of training data for its high threshold for annotation. Current research on image cropping is confined on public datasets. Grounded on domain shift between training domain and practical application scene, a listwise adversarial domain adaption algorithm for image cropping is proposed in this paper. Firstly, the domain shift between two image cropping datasets, GAICD and CPC, is proved. Then, an image cropping model composed of an aesthetic evaluation module and an adversarial domain adaptation module is constructed. Aesthetic evaluation module is employed to predict the aesthetic score of current image and assist the model to extract the invariant features for cropping task. Adversarial domain adaptation module is exploited to realize adversarial based domain adaptation learning. Domain migration experiments between different cropping datasets and between different scene domains verify the effectiveness of proposed algorithm.

- Key Words Image Cropping, Domain Shift, Domain Adaption, Adversarial Learning, Invariant Feature
- **Citation** WANG H W, SANG N. Listwise Adversarial Domain Adaption Algorithm for Image Cropping. Pattern Recognition and Artificial Intelligence, 2021, 34(8): 677–688.

accepted July 5, 2021

华为-华中科技大学 DigiX 智慧体验联合创新中心项目资助 Supported by Project of the DigiX Joint Innovation Center of Huawei-HUST

本文责任编委 吴飞

Recommended by Associate Editor WU Fei

- 华中科技大学人工智能与自动化学院图像信息处理与智能控制教育部重点实验室 武汉 430074
- Key Laboratory on Image Information Processing and Intelligent Control of Ministry of Education, School of Artificial In-

图像裁剪任务按照美学标准裁剪图像,将裁剪 后的图像作为生成图像,使生成图像具有更高的美 学质量.图像裁剪操作简便,只涉及空间操作,却显 著提升图像的美学质量,被广泛应用于图像编辑领 域.然而人工裁剪存在效率较低、门槛较高的问题, 裁剪者需要花费大量时间学习摄影知识并逐一裁剪 图像.因此,研究利用智能算法自动裁剪图像是一件 有价值的工作.相比人工裁剪,智能裁剪算法不需要 用户掌握摄影知识,任何人都能在短时间内得到大

收稿日期:2021-02-07;录用日期:2021-07-05

Manuscript received February 7, 2021;

telligence and Automation, Huazhong University of Science and Technology, Wuhan 430074

批图像的优质裁剪结果.

在智能裁剪从理论走向实用的过程中,有两个问题需要解决:构建有监督情况下的智能裁剪算法 和提升智能裁剪算法在无标签域上的泛化能力.具 体来说,研究者首先要构建合适的裁剪算法,用于模 拟人类美学鉴赏的过程,描述当前裁剪图像的美学 质量,筛选较好的裁剪方式,使用美学标签加以约 束,增强裁剪算法在特定数据集上的裁剪效果.然 后,裁剪算法需要具有一定的泛化能力,适应于不同 数据集和不同场景.

智能裁剪任务从算法上分为3类:美学驱动算法、注意力驱动算法、数据驱动算法.数据驱动算法.数据驱动算法.又包含全监督算法和弱监督算法.

美学驱动的智能裁剪算法按照传统美学规则裁 剪图像,会根据构图规则设计手工特征,提取图像的 美学信息并筛选裁剪方式^[1-4].虽然美学驱动算法 具有可解释性,但遵循固定构图规则,不能适应其它 构图模式.并且该类算法采用的原始数据来源于 AVA^[5]、AADB^[6]等美学数据集,缺乏针对裁剪任务 的标注.

注意力驱动的智能裁剪算法从图像中的显著性 区域生成裁剪图像.从经验可知,人物、动物、建筑等 关注度更高的区域显然包含原始图像中的主要信 息,通常应被保留.因此,注意力驱动算法考虑到显 著目标对裁剪图像的影响,取得较优效果^[4,7-13].Lu 等^[7]先通过显著性检测网络得到显著性最高的区 域,再基于该区域生成合适的裁剪区域.Tu 等^[8]和 Lu 等^[14]把显著性检测结果用于监督生成的裁剪结 果.注意力驱动算法往往会将最显著的区域放在裁 剪图正中心,更适用于提取缩略图而不是智能裁剪.

近年来数据驱动的智能裁剪算法得到快速发展. AVA 数据集^[5] 在早期被作为训练数据,用于训练裁剪算法. 但 AVA 数据集只包含原图的美学标注,难以满足裁剪任务的要求. 所以,学者们提出ICDB^[1]、FLMS^[2]、FCDB^[15]数据集,将原图裁剪成更好的区域,并保留裁剪数据作为标签. 学者们进而提出 CPC^[16]、XPView^[16]、GAICD^[17]数据集,每幅图像都包含大量相似的图像对. 在这些数据集的支撑下,数据驱动算法也取得较优效果. 按照标签的类型,现有数据驱动的智能裁剪算法可分为全监督算法和弱监督算法.

全监督数据驱动算法在训练时使用训练集标 签,通常包含裁剪标签(裁剪方式的坐标)和美学标 签(裁剪方式对应的美学分数标签). Wang 等^[11]、

Wei 等^[16]和 Deng 等^[18]使用 AVA、AADB 等传统美 学数据集,训练得到美学分类器,并生成裁剪区域. Zeng 等^[17]摒弃传统的美学分类+裁剪两阶段方法, 定义预选裁剪框,将预选图像区域序列排序.Tu 等^[8]将原图划分成若干块,预测不同块的像素评 分,再按照设定的构图模式得到裁剪图.虽然上述方 法已取得较好效果,但未考虑训练样本分布和实际 测试样本分布之间的差异,在实际使用中会因为域 偏移问题导致性能下降.

弱监督数据驱动算法在训练时只会使用部分标签,如训练数据中只包含当前原始图像的质量,而不包含完整的美学标签^[14,19].而 Li 等^[20]将 Chen 等^[19]的美学分类器作为软标签,通过强化学习的方法找到最优裁剪模式.

研究者们利用有标签的训练集样本训练模型, 再将训练好的模型用于推理其它缺少标签的测试样 本.在这个过程中,训练集样本分布与测试样本分布 之间往往存在差异,这个差异称为域偏移.域适应就 是为了解决域偏移问题.在域适应问题的研究中,有 标签的用于训练的数据域称为源域,缺少标签的数 据域称为目标域,其中,目标域中只有图像而没有任 何标签的情况称为无监督域适应.无监督域适应算 法旨在对齐源域和目标域的特征,在目标域无标签 的情况下,将从源域学到的模型应用于目标域.

无监督域适应算法已大量应用于跨域分类、检测、分割任务.有些方法^[21-25]引入损失函数,衡量源 域和目标域分布的差异,并加以约束.有些方 法^[26-29]采用基于对抗的域适应策略,将特征提取器 作为生成器,混淆对齐两个域的特征.还有些方 法^[30-31]采用梯度反转层^[31]对齐不同域,在生成器 和域判别器之间添加梯度反转层,使生成器和域判 别器按照相反的方向优化.

现有的智能裁剪算法未考虑无监督域迁移的问题,有些算法训练样本和测试样本均来自同一数据 集 GAICD,不涉及域迁移的问题^[8,17,32].还有些算法 的训练集和测试集虽然完全不同,但默认算法能从 单一训练集学到足够的泛化能力,不需要采用域迁 移策略^[7,15-16,33].Lu 等^[14]和 Chen 等^[19]的方法尽管 实现从源域到目标域的迁移,但需要大量目标域高 质量图像作为先验信息,不属于无监督域适应.用户 也很难在实际使用中获得大量目标域高质量图像用 于训练算法.

因此,本文提出基于序列对抗域适应的智能裁 剪算法(Listwise Adversarial Domain Adaption Algori-

679

thm for Image Cropping, LDAIC),提升裁剪算法对于 不同场景的适应能力.在训练时.本文算法依赖的美 学标签只来自源域样本,针对于不同的目标域,选择 对应的无标签图像作为目标域样本,用于训练网络 的域适应能力.具体来说,针对域迁移问题,设计基 于对抗损失的域适应算法,用于对齐源域数据和目 标域数据.当输入样本来自源域/目标域时,域判别 器分辨当前样本的所属域,而特征提取器会尽量混 淆两个域的数据,对齐目标域样本特征与源域样本 特征.进一步,本文设计2个不同的美学分类器,用 于评价当前区域的美学质量,通过约束预测值的相 似度和权重参数的差异度,辅助算法提取到与裁剪 任务相关的不变特征.通过这种策略,提升裁剪算法 在跨域和跨数据集时的泛化能力,能在完全不依赖 目标域美学标注的情况下提升裁剪算法在目标域上 的裁剪性能.

1 智能裁剪中的域适应问题

首先,本文使用2个实验证实裁剪任务中存在 域迁移问题.

实验选用 CPC^[16]、GAICD^[17]数据集. CPC 数据 集包含 10 800 幅图像,每幅图像有 24 种裁剪方式, 每种裁剪方式包含 6 位不同标注者的美学评价. 实 验中将 6 位标注者美学评价的平均值作为当前区域 的美学评分.本文将 CPC 数据集也分为包含7 559 幅图像的训练集和 3 238 幅图像的测试集. GAICD 数据集包含 1 236 幅原始图像,每幅原始图像平均 包含 86 种不同的裁剪方式,共 106 860 种. 每种裁 剪方式都具有对应的美学标签,标签分数取值范围 为1~5,分数值连续分布.数据集上 1 036 幅图像作 为训练集,200 幅图像用于测试和验证. GAICD 数据 集上预选裁剪区域的相对位置按照固定网格(Grid Anchor)规则^[17]确定,因此预选框的相对位置是研 究者已知的先验信息.

将 CPC 数据集和 GAICD 数据集分成的训练集 和测试集交替作为源域和目标域,训练裁剪算法并 测试其在训练域和域迁移情况下的性能.

为了衡量裁剪算法的性能,本文采用如下评价 指标:斯皮尔曼等级相关系数(Spearman's Correlation Coefficient for Ranked Data, SRCC),预测结果 中前 k 种裁剪方式属于美学标签中前 N 种裁剪方 式的准确率($Acc_{K/N}$). SRCC 反映 2 组变量之间关系 的密切程度.一幅原始图像的裁剪子图构成一个图 像序列,对应一个美学标签序列和一个美学预测序 列,美学标签评分序列与美学预测序列相似度越高, SRCC 值越大.按照文献[17], Acc_{K/N}中取 Acc₅和 Acc₁₀作为对比项:

$$Acc_5 = \sum_{n=1}^{4} Acc_{n/5}, \ Acc_{10} = \sum_{n=1}^{4} Acc_{n/10}$$

在实际应用中,用户只关心高质量裁剪结果,不 在意裁剪算法对低质量裁剪结果预测的精确度.虽 然 SRCC 考虑到这一点,赋予高质量裁剪区域更大 的判别权重,但仍会受到低质量裁剪区域的影响. *Acc_{K/N}* 只包含前 *N* 个高质量裁剪区域,是本文首选 的评估指标.

实验1旨在验证使用源域样本训练好的裁剪算法,在迁移到目标域后,性能会显著下降.实验2旨在验证 CPC 数据集和 GAICD 数据集之间的数据分布存在差异,本文构造域判别器,赋予2个域数据各自的域标签进行训练,最后测试域判别器的二分类准确率.

实验1 跨数据集模型性能验证

实验 1 共包含 4 组对照实验,分别将 CPC/ GAICD 作为源域和目标域.这里约定表达方式:CPC → GAICD 表示训练样本来自源域 CPC,测试样本来 自目标域 GAICD 时,算法从 CPC 迁移到 GAICD 上 的实验结果.而 CPC → CPC 表示训练样本和测试样 本均来自 CPC,不存在域迁移时的实验结果.

不同源域数据在不同目标域上的结果如表1所示.当源域样本和目标域样本来自一个数据集时,即 CPC→CPC和GAICD→GAICD,裁剪算法能生成较 优结果.而当CPC→GAICD和GAICD→CPC时,算 法性能会受到严重影响.在固定训练集的情况下,算 法在同分布的测试样本上表现更优.在固定测试集 时,使用同分布的样本训练的算法会取得更优效果.

表1 不同源域数据在不同目标域上的结果

Table 1	Results of different source domains data on different
	target domains

源域 → 目标域	Acc_5	Acc_{10}	SRCC
$GAICD \rightarrow GAICD$	0.540	0.723	0.752
$\mathrm{GAICD} \rightarrow \mathrm{CPC}$	0.451	0.669	0.274
$CPC \rightarrow CPC$	0.742	0.886	0.475
$CPC \rightarrow GAICD$	0.249	0.417	0.360

实验1说明,在裁剪任务中进行数据集迁移会 影响实验效果.为了进一步证实这种性能下降是由 CPC、GAICD数据集上数据分布不同引起的,设置实 验2.

680

实验2 GAICD 和 CPC 的可分性实验

实验2共包含3组对照实验,目的是测试域判别 器能否学习到判别源域图像和目标域图像的能力.

实验模型包含一个计算机视觉组(Visual Geometry Group, VGG)结构的特征提取器、一个由全 连接层和 softmax 组成的判别器,输入为两个域的图 像,输出为当前图像隶属于两个域的置信度.

第1组实验.从 KAIST 数据集^[34]上随机选取 1200幅白天图像和1200幅黑夜图像.从白天数据 和黑夜数据中各选取1000幅图像作为训练集,200 幅图像作为测试集,以0.5作为置信度的阈值,算法 收敛以后的二分类准确率为93.4%.

第2组实验.从GAICD数据集和CPC数据集上 各选取1000幅图像作为训练集,200幅图像作为测 试集,阈值为0.5时的二分类准确率为74.3%.

第3组实验.从 GAICD、CPC 数据集上选取图像,随机从中各抽取 600 幅并两两组合,形成混合 A 数据集和混合 B 数据集.将混合 A、混合 B 数据集各自分成1000 幅训练集和 200 幅测试集,阈值为 0.5 时的二分类准确率为 48%~52%.

在第1组实验中,源域样本和目标域样本在光 照强度上有显著差异,二分类准确率在90%以上, 域判别器可轻松分辨两个类别的图像,说明实验中 的域判别器是有效的.第3组实验中GAICD、CPC数 据集上的数据被完全打乱分到两个数据集,域判别 器无法区分当前样本来自混合A数据集还是混合B 数据集.在第2组实验中,域判别器能取得75%左右 的准确率,与第3组实验中的50%准确率相比有明 显增长,说明GAICD数据集和CPC数据集之间的数 据分布存在显著差异,并且体现在图像而不是标签 上,这也解释了实验1进行数据集迁移之后的模型 性能下降原因.

上述 2 个实验说明 GAICD 数据集和 CPC 数据 集之间的数据分布存在差异,即域偏移,并且这种域 偏移会导致算法在域迁移的过程中性能下降.

2 基于序列对抗域适应的智能裁 剪算法

为了解决裁剪域偏移问题,本文提出基于序列 对抗域适应的智能裁剪算法(LDAIC).

2.1 问题描述

将有标签的 A 数据集作为源域(Source),无标 签的 B 数据集作为目标域(Target).一幅原始图像 对应若干种裁剪方式,因此将按照规定方式裁剪后 的图像作为基本样本单位,命名为裁剪子图.设源域 原始图像集合

 $I_{s} = \{I_{s}(0), I_{s}(1), \dots, I_{s}(k)\},\$ 每幅原始图像对应若干个裁剪子图,裁剪子图集合 $B_{s} = \{B_{s}(0), B_{s}(1), \dots, B_{s}(n)\},\$

源域裁剪子图对应的美学标签集合

 $a_s = \{a_s(0), a_s(1), \dots, a_s(n)\}.$ 设目标域原图集合

$$I_{i} = \{I_{i}(0), I_{i}(1), \cdots, I_{i}(k)\},\$$

裁剪子图集合

 $B_{i} = \{B_{i}(0), B_{i}(1), \dots, B_{i}(n)\},$ 目标域裁剪子图对应的美学标签集合

 $a_{i} = \{a_{i}(0), a_{i}(1), \cdots, a_{i}(n)\}.$

在本任务中,研究者设计算法预测目标域子图 的美学分数,并从中筛选更优的裁剪方式.问题在 于:目标域标签 *a*_i 是缺失的,只能利用源域裁剪子 图 *B*_s、源域美学标注 *a*_s和目标域裁剪子图 *B*_i训练算 法,再通过算法预测 *P*_i 作为目标域子图的美学预 测.另外,设 *P*_s 作为源域子图的美学预测.

本任务的目的是减小算法在目标域上的预测误 差,即减小

$$E_{i}(B_{i},a_{i}) = \sum_{i=1}^{N} D(P_{i}(i),a_{i}(i)).$$

由于目标域标签 a_t 缺失,无法直接约束 $E_t(B_t, a_t)$ 最小,但可使用 $E_s(B_s, a_s)$ 和 $dis(I_s, I_t)$ 衡量 $E_t(B_t, a_t)$. $E_s(B_s, a_s)$ 表示裁剪算法在源域上的预测误差, $dis(I_s, I_t)$ 表示源域和目标域数据分布的差异,即域 偏移量.为了减小 $E_t(B_t, a_t)$,从减小 $E_s(B_s, a_s)$ 和 $dis(I_s, I_t)$ 入手,构建LDAIC.

2.2 网络结构

本文算法网络结构如图 1 所示,主要包含特征 提取器、域判别器、2 个美学分类器.本文采用 VGG 作为特征提取器,用于从原始图像 $I \in \mathbb{R}^{H \times W \times 3}$ 中提 取全局特征 $F_1 \in \mathbb{R}^{H \times W \times D_1}$. F_1 经过自适应梯度反转 层后送入域判别器,而域判别器的作用是将当前样 本按照源域和目标域分成两类. 2 个美学分类器的 结构相同但参数不同,输入同样为 F_1 ,按照 Grid Anchor 规则^[17] 生成的 N种裁剪方式对 F_1 进行重采 样并输出这 N 个裁剪子图的美学预测分数.





2个美学分类器构成本文算法美学评分模块, 用于预测裁剪子图的美学评分并辅助算法提取不变 特征.自适应梯度反转层和域判别器构成对抗域适 应模块.对抗域适应模块通过对抗域适应方法对齐 源域和目标域样本的分布.

2.3 美学评分模块

为了使算法具备裁剪任务相关的美学鉴赏能力,减小在源域的预测误差 E_s(I_s, a_s),本文构建美学评分模块.2个不同参数的美学分类器 A₁, A₂ 组成 美学评分模块.本文首先根据 Grid Anchor 规则^[17] 生成预选裁剪方式并对全局特征进行重采样,将原 始图像中的裁剪区域作为前景重采样.

 F_1 和 N 种裁剪方式输入 ROIAlign^[35],得到前 景特征 $F_f \in \mathbb{R}^{N \times 8 \times 8 \times D_1}$. 对于背景部分,将待裁剪区 域置 0 并输入 ROIAlign 中,得到背景特征 $F_b \in \mathbb{R}^{N \times 8 \times 8 \times D_1}$.将前景特征和背景特征从通道维度叠加, 得到重采样特征 $F_2 \in \mathbb{R}^{N \times 8 \times 8 \times 2D_1}$, F_2 被分别送入 A_1 和 A_2 . A_1 主要由 3 个卷积层和1 个失活层组成. F_2 首 先送入 kernel_size = (8,8) 的卷积层,得到 $F_3 \in \mathbb{R}^{N \times 1 \times 1 \times D_2}$,本文中 $D_2 = 1$ 024. F_4 经过 kernel_size = (1,1)卷积层和1 个失活层,最后被送入输出维度为 1,kernel_size = (1,1) 的卷积层,得到预测结果 $P \in \mathbb{R}^{N \times 1 \times 1 \times 1}$. P中第1 个维度表示 N种裁剪方式,对应每 种裁剪方式的预测结果. A_2 与 A_1 结构相同,此处不 赘述.

将 A₁ 和 A₂ 预测结果的均值作为每种裁剪方式 的美学预测分数,并选择分数最高的前 K 个裁剪子 图作为算法输出.本文采用源域数据集上的裁剪方 式和美学标签作为监督训练模型的美学鉴赏能力. 对于 A₁ 和 A₂,约束其预测结果趋于一致,并保持它 们参数的独立性,辅助提取与美学相关的不变信息.

原始图像 *I_s(i)* 对应1组共 *N* 种裁剪方式,裁剪 后生成的新图像

 $B(i) = \{B_1(i), B_2(i), \dots, B_N(i)\},$ 每个裁剪区域的美学分数标签

 $Sc(i) = \{Sc_1(i), Sc_2(i), \dots, Sc_N(i)\}.$

根据图1,输入原始图像提取特征并使用 ROIAlign 对预设裁剪区域进行重采样,得到每个裁剪子图的 区域特征

 $F_{s}^{j}(i) = G(I_{s}^{j}(i)), j = 1, 2, \dots, N,$

其中, $I_{s}^{i}(i)$ 表示图 $I_{s}(i)$ 对应裁剪区域为 B_{j} 时的裁 剪子图, $F_{s}^{j}(i)$ 表示对应的区域特征,G 表示特征提 取器.

得到区域特征后,输入A₁、A₂,用于预测当前裁 剪子图的美学分数,将美学分数的平均值作为裁剪 子图的美学质量预测结果,即

$$p_s(i) = \{p^1(i), p^2(i), \dots, p^N(i)\}$$

GAIC(Grid Anchor Based Cropping)^[17] 采用SmoothL1 作为美学损失 L_A ,用于约束 $p_s(i)$ 和 Sc(i) 趋于一致,SmoothL1 如下所示:

$$L_{\text{smoothL1}} = \begin{cases} \frac{1}{2}e(i)^2, & |e(i)| \leq \delta\\ \delta |e(i)| - \frac{1}{2}\delta^2, & |e(i)| > \delta \end{cases}$$

其中, $e(i) = p_s(i) - sc(i)$ 表示预测值与美学标签之间的误差, $\delta = 0.5$.

但在 SmoothL1 的计算过程里,所有裁剪子图都 具有相同权重,这与任务目标不符.在智能裁剪任务 中,用户只需要得到高质量的裁剪结果,不关心算法 对低质量裁剪结果预测的精确度.因此,Listwise-Loss^[36]通过非线性归一化的策略放大高质量裁剪 子图的权重.

ListwiseLoss 计算如下:

$$Sc'(j) = \prod_{j=1}^{n} \frac{\Phi(sc_{j})}{\sum_{k=1}^{n} \Phi(sc_{k})}, \quad \Phi(x) = e^{x};$$

$$P'(j) = \prod_{j=1}^{n} \frac{\Phi(f(I^{j}))}{\sum_{k=1}^{n} \Phi(f(I^{k}))}, \quad \Phi(x) = e^{x}; \quad (1)$$

$$L_{\text{listwise}} = -\sum_{i=1}^{m} \sum_{j=1}^{n} Sc'(j)\ln(P'(j)).$$

其中,Se 表示原始标签,Se'表示归一化后的标签,f 表示预测函数,P'表示归一化后的预测值.

本文结合 SmoothL1 和 ListwiseLoss,作为美学 损失函数,命名为 ScalingSmoothL1. 该损失按照式 (1)中的归一化方法处理标签 Sc 和预测值 P,将归 一化后的 Sc'和 P'分别作为标签和预测值代入 SmoothL1 中计算损失,即

$$L_{A} = \begin{cases} \frac{1}{2} (X_{\text{dis}})^{2}, & |X_{\text{dis}}| \leq 1 \\ |X_{\text{dis}}| - \frac{1}{2}, & |X_{\text{dis}}| > 1 \end{cases}$$

其中

$$X_{\rm dis} = \frac{e^{Sc(i)}}{\sum_{i=1}^{n} e^{Sc(i)}} - \frac{e^{p(i)}}{\sum_{i=1}^{n} e^{p(i)}},$$

表示归一化后的预测值与美学标签之间的误差.

通过上述方法,可让*p*(*i*) 趋近真实值 *Sc*(*i*),使 算法学习到美学鉴赏能力.而通过上述的非线性映 射方法,也可将网络的关注点向高质量裁剪子图倾 斜,避免算法耗费过多资源拟合低质量裁剪子图.在 测试阶段,裁剪算法根据 *p*(*i*) 筛选裁剪子图,并输 出裁剪结果.

除此之外,本文算法采用2个结构相同参数不同的美学分类器构建美学评分模块.相比单美学分 类器的方法,2个参数不同的美学分类器有利于算 法提取美学不变特征.参数不同的美学分类器是从 不同角度分析当前子图,如果预测结果趋于一致并 与美学标签吻合,说明算法提取到的特征含有更多 的美学信息,不易受到分类器的影响.这种不受分类 器限制、被输入到不同美学分类器都能得到正确且 相似预测结果的特征,称为美学不变特征.美学不变 特征更关注图像的美学内容,提高算法的美学鉴赏 能力,增强算法的鲁棒性.

为了制造2个有差异的美学分类器,本文通过 权重损失约束分类器参数,降低2个美学分类器参 数的相似度.首先将每个分类器的参数都编码为一 个参数矩阵,再最大化分类器参数之间的余弦距离, 避免分类器在训练过程中趋同,称为权重损失 L_{w} . 设2个分类器 A_1 、 A_2 , W_1 为 A_1 参数矩阵, W_2 为 A_2 参 数矩阵,权重损失

$$L_{w}(W_{1}, W_{2}) = -(1 - \cos\langle W_{1}, W_{2} \rangle) = \frac{\sum_{i=1}^{n} W_{1i} W_{2i}}{\sqrt{\sum_{i=1}^{n} W_{1i}^{2}} \sqrt{\sum_{i=1}^{n} W_{2i}^{2}}} - 1.$$
(2)

当 A_1 、 A_2 分类面相近时, L_w (W_1 , W_2)偏大,最小化 L_w (W_1 , W_2)可使 W_1 与 W_2 相似度减小.由于 A_1 、 A_2 只是具体参数不一样,结构和参数规模都是相同的, 因此可直接代入式(2)计算.

同时,本文约束 A₁、A₂ 对相同数据预测相同的 结果. 输入一幅图像 I(i) 和 k 个预设的裁剪方式

$$B(i) = \{B_1(i), B_2(i), \dots, B_k(i)\}$$

 A_1 预测结果向量 $P_1(i) \in \mathbb{R}^{k \times 1}, A_2$ 同样预测结果向 量 $P_2(i) \in \mathbb{R}^{k \times 1},$ 采用一致性损失 L_c 优化算法, 使 2 个分类器预测结果趋于一致, 即

$$L_{c} = \frac{1}{n} \sum_{i=1}^{n} | \boldsymbol{P}_{1}(i) - \boldsymbol{P}_{2}(i) |.$$

2.4 对抗域适应模块

为了减少源域和目标域之间的数据分布差异 dis(I_s , I_i),本文设计对抗域适应模块,用于解决这 一问题.对抗域适应模块包含一个自适应梯度反转 层和一个域判别器.全局特征 F_1 作为该模块的输入 被送入自适应梯度反转层,在正向传播中,自适应梯 度反转层为恒等变换,输出 F_1 并作为域判别器的输 入.域判别器首先通过一个全图 ROIAlign 对 F_1 进行 重采样,得到 $F_4 \in \mathbb{R}^{8\times 8\times D_1}$. F_4 经过一个 kernel_size = (8,8)的卷积层、1 个全连接层和1 个 softmax,得到 域判别器预测结果 $P_{domian} \in \mathbb{R}^{1\times 2}$, P_{domian} 表示当前图 像隶属于某个域的置信度.

在反向传播的过程中,自适应梯度反转层会将 目标域样本的梯度反转,域判别器会判别当前样本 隶属的域.当输入样本来自目标域时,自适应梯度反 转层会将梯度反转,导致域判别器与特征提取器接 收到相反方向的梯度.因此,在域判别器学习判别能 力的同时,特征提取器也在强化混淆两个域特征的 能力,两者形成对抗,让目标域样本的特征分布与源 域样本的特征分布对齐.

2 个不同分布之间的差异可采用散度衡量. 这 里使用 **F** 表示当前图像提取的特征向量, **F**_s 表示源 域特征, **F**_t 表示目标域特征. **D**_{domain} 表示域判别器. 设源域样本标签为0,目标域样本标签为1,设 H 表 示所有可行的 D_{domain} 参数构成的集合,源域和目标 域之间分布的差异表示如下:

$$d_{H} = 2 \left(1 - \min_{D_{\text{domain}} \in H} (err(D_{\text{domain}}(F_{s})) + err(D_{\text{domain}}(F_{t})) \right),$$

其中, $err(D_{domain}(F_s))$)表示源域样本的误分率, $err(D_{domain}(F_t))$ 表示目标域样本的误分率, d_H 表 示两个域分布的差异.随着样本误分率的提高, d_H 变小,因为域判别器更难区分这两个域,说明它们之 间的特征相似度更高.

本文将全局特征 $F_1 = G(I)$ 输入域判别器中,约 束 d_H ,使两个分布之间的差异 d_H 最小,即

$$\min_{G} d_{H} \leftrightarrow \max_{G} \min_{D_{\text{domain}} \in H} (err(D_{\text{domain}}(F_{s})) + err(D_{\text{domain}}(F_{s}))).$$

具体到算法结构,本文用自适应梯度反转层优 化上式.自适应梯度反转层是梯度反转层(Gradient Reversal Layer, GRL)^[31]的变形,梯度反转层会在 算法反向传播梯度时,将梯度反转,前向传播和反向 传播

$$R_{\lambda}(x) = x,$$
$$\frac{\partial R_{\lambda}}{\partial x} = -\lambda \frac{\partial L_d}{\partial x}$$

其中, $R_{\lambda}(x)$ 表示输入为 x 时正向传播的结果, $\frac{\partial L_{a}}{\partial x}$ 表示反向传播计算出的梯度. GRL 在正向传播时为 恒等变换,反向传播时,GRL将计算的梯度乘 – λ 后 传回上一层网络,本文设置 λ = 0.3.

本文采用二分类交叉熵作为判别器损失:

 $L_d = - [y' \ln(D_{\text{domain}}(\mathbf{F})) +$

$$(1 - y')\ln(1 - (D_{\text{domain}}(F)))].$$

其中:y'表示输入图像 I 对应的域标签,当 I 来自源 域时 y' = 0,当 I 来自目标域时 y' = 1; $D_{\text{domain}}(F)$ 表示 域判别器的输出结果,在经过 softmax 归一化后数值 介于 0 和 1 之间.

采用 GRL 可初步对齐源域和目标域样本特征, 但 GRL 会同时反转源域样本和目标域样本的梯度, 导致特征提取器实际按照

$$D_{\text{domain}}(F_s) = 1, D_{\text{domain}}(F_t) = 0$$
的方向优化,这与本文算法混淆源域和目标域样本

特征的策略不符. 而本文采用自适应梯度反转层只 反转目标域样本的梯度,不改变源域样本的梯度,反 向传播公式如下:

$$\frac{\partial R_s}{\partial x} = \frac{\partial L_s}{\partial x}, \ \frac{\partial R_t}{\partial x} = -\lambda \ \frac{\partial L_t}{\partial x}.$$

当输入样本来自目标域时,特征提取器 G 和域 判别器 D_{domain} 会按照相反的方向优化,在域判别器 学习判别能力的同时,特征提取器会生成混淆 2 个 域数据的样本特征,特征提取器和域判别器形成对 抗. 当输入样本来自源域时,特征提取器 G 和域判别 器 D_{domain} 会按照相同的方向优化,特征提取器与域 判别器不会产生对抗. 按照这种方式,域判别器的优 化方向为

 $D_{\text{domain}}(\mathbf{F}_s) = 0, D_{\text{domain}}(\mathbf{F}_t) = 1,$ 而特征提取器的优化方向为

 $D_{\text{domain}}(F_s) = 0, D_{\text{domain}}(F_t) = 0,$ 此时特征提取器将两个域的样本视为相同类别进行 优化,与域判别器形成对抗,克服 GRL 的缺陷,对齐 两个域的特征分布.

2.5 损失函数

综合上述内容,本文算法是一个端到端模型,并 采用多任务损失综合优化. 在训练阶段,一组源域-目标域图像对,包含一幅源域原始图像和一幅目标 域原始图像,作为单次迭代的样本送入算法. 源域样 本对应的美学标签和域标签会同时计算美学损失 $L_A(采用 ScalingSmoothL1)、一致性损失 <math>L_c$ 、权重损 失 L_w 和判别器损失 L_d . 而目标域样本只有域标签, $L_A = 0$,计算 L_d 、 L_e 、 L_w . 计算一组源域-目标域图像对 的总损失:

 $L = L_A + \lambda_1 L_d + \lambda_2 (L_e + \mu L_w)$, (11) 其中, λ_1 , λ_2 , μ 均为折衷参数,用于平衡4种损失. 本文设定 $\lambda_1 = 0.2$, $\lambda_2 = 0.5$, $\mu = 0.2$.在训练阶段, 特征提取器、美学评分及域判别器都会参与训练.而 在推理阶段,输入图像经过特征提取和美学评分,预 测当前裁剪区域的美学分数,而后根据分数筛选裁 剪结果.

本文采用在 ImageNet 上预训练过的参数来初 始化算法. 整个算法通过自适应矩估计(Adaptive Moment Estimation, Adam)优化器学习参数. 在最 初的30个迭代周期里,采用 $L = L_A$ 优化网络,使网络 具有基本的美学鉴赏能力. 在之后的10个迭代周期 里,固定域判别器之外的所有参数,使用 $L = \lambda_1 L_d$ 训 练域判别器参数,使域判别器能判别源域和目标域 样本. 之后解除所有参数的固定,使用

 $L = L_A + \lambda_1 L_d + \lambda_2 (L_c + \mu L_w)$

优化整个网络,整个过程包括 60 个迭代周期,学习 率为 10⁻⁴,输入原图的批尺寸大小设为 1.

最后得到本文算法流程图如图2所示.



图 2 本文算法流程图 Fig. 2 Flow chart of the proposed algorithm

3 实验及结果分析

3.1 实验数据集和评估指标

本文在实验过程中需要变换源域和目标域,训 练集和测试集也会随之变化,涉及到的数据集主要 有 CPC^[16]、GAICD^[17]、INDOOR、OUTDOOR.

CPC、GAICD数据集具体信息见第2节. INDOOR和OUTDOOR是本文在GAICD、CPC数据 集基础上建立的跨域数据集.INDOOR包含4867幅 室内场景图,OUTDOOR包含7172幅室外场景图, INDOOR、OUTDOOR数据集上的12039幅图像和美 学标注均来自GAICD、CPC数据集.INDOOR、 OUTDOOR数据集按照原始图像的场景域分类,本 文通过文献[37]的场景识别方法将其分为室内/ 室外两类,分别归纳到INDOOR、OUTDOOR数据 集.INDOOR训练集包含3936幅图像,测试集包含 931幅图像.OUTDOOR训练集包含5703幅图像,测 试集包含1469幅图像,比例约为4:1.

评估指标仍采用 SRCC 和 *Acc*₅、*Acc*₁₀,具体见第 2 节.

为了验证算法的跨域自动裁剪能力,实验包含 源域、目标域两类数据集.源域训练集的样本既包含 原始图像也包含美学标注,而目标域训练集只包含 原始图像不包含美学标注.测试集同时具有原始图 像和美学标注.因此,本文主要设计4个实验,分别 是:基准方案选取实验、CPC↔GAICD 域迁移实验、 INDOOR↔OUTDOOR 域迁移实验、消融实验.最后 给出可视化结果.

3.2 基准方案选取实验

首先确定基准方案.将CPC→GAICD的裁剪性

能作为基准方案的筛选条件.这里将源域为CPC、目标域为GAICD的域迁移实验简写为CPC→GAICD, 后文沿用该表达方式.针对算法的特征提取器结构 和损失函数,设置如下两组实验,对特征提取器结构 和美学损失函数进行优选.

实验3 特征提取器结构优选

本文按照文献[17] 将 GAICD 数据集上 1 036 幅图像作为训练集,200 幅图像作为测试集. CPC 数 据集上随机选取 7 559 幅图像作为训练集,剩下的 3 238 幅图像作为测试集.

实验中需要使用 CPC 训练集和 GAICD 测试集. 对比 4 种现有智能裁剪算法用到的特征提取器结构:VGG、ResNet50(Residual Network 50)、Mobile-Net、shuffleNet.结果如表 2 所示.由表可见,VGG 的效果略优于 ResNet50.其实学者们已验证 VGG 在同域有监督的智能裁剪任务中是效果较优的特征提取器结构^[8,17,32].实验 3 进一步验证在 CPC → GAICD 域迁移的过程中 VGG 也能取得较优效果.VGG 能在 智能裁剪任务中取得较优效果可能得益于较少的数 据量.相比 ResNet 这种深层网络,VGG 的层数较少, 裁剪任务的训练集,如 GAICD 数据集,只包含约千 幅原始图像,深层网络极易过拟合于训练集,影响裁 剪算法在其它数据域上的泛化能力.因此,本文算法

表 2 不同特征提取器对 $CPC \rightarrow GAICD$ 的影响

Table 2	Effect of	different	feature	extractors	on	CPC -	\rightarrow

GAICD			
特征提取器	Acc_5	Acc_{10}	SRCC
ResNet50	0.251	0.410	0.324
VGG	0.249	0.417	0.360
MobileNet	0.210	0.365	0.213
shuffleNet	0.224	0.371	0.224

实验4 美学损失函数优选

在美学损失函数方面,对比文献[17]采用的 smoothL1、文献[36]采用的Listwise Loss 和本文采 用的ScalingSmoothL1,结果如表3所示.

表 3 不同损失函数对 CPC \rightarrow GAICD 的影响

Table 3 Effect of different loss functions on CPC \rightarrow GAICD

损失函数	Acc_5	Acc_{10}	SRCC
smoothL1	0.242	0.367	0.244
Listwise Loss	0.237	0.391	0.251
ScalingSmoothL1	0.249	0.417	0.360

由表3可见,ScalingSmoothL1在SRCC指标上 具有显著优势,通过非线性函数适当放大高质量区 域之间的评分差异,压缩低质量区域之间的评分差 异,这有助于提升算法的裁剪能力.

3.3 CPC↔GAICD 域迁移实验

在本次实验中,对比 CPC → GAICD 和 GAICD → CPC 时 LDAIC、基准方案及其它现有跨域裁剪算 法的性能.

实验5 CPC↔GAICD 域迁移实验

按照3.1节的数据集划分,在CPC→GAICD 域 迁移实验中,涉及的数据包括:有标签的CPC训练 集样本、无标签的GAICD训练集样本和有标签的 GAICD测试集样本.在训练LDAIC时,批尺寸大小 为1,一个迭代周期包括1036次迭代.按照设置,一 个迭代周期只包含1036个源域/目标域图像对,但 实验中设置 *shuffle* = True,随着迭代的进行,CPC训 练集上7559 幅图像样本都会参与训练.

按照 3.1 节的数据集划分, GAICD \rightarrow CPC 域迁 移实验中涉及的数据包括: 有标签的 GAICD 训练集 样本、无标签的 CPC 训练集样本和有标签的 CPC 测 试集样本. 同样设置 *shuffle* = True, 一个迭代周期迭 代 1 036 次.

选择本文能复现结果或能找到跨域情况下测试 结果的现有智能裁剪算法:GAIC^[17]、VFN(View Finding Network)^[19]、A2-RL(Aesthetics Aware Reinforcement Learning)^[20],结果如表4所示,表中黑体 数字表示最优结果.

表 4 各方法在 GAICD↔CPC 上的结果

Table 4	Results of	different	methods	on	GAICD↔CPC

<u>→</u> .»+	$CPC \rightarrow GAICD$			$GAICD \rightarrow CPC$		CPC
刀伝	Acc_5	Acc_{10}	SRCC	Acc_5	Acc_{10}	SRCC
基准方案	0.249	0.417	0.360	0.451	0.669	0.274
GAIC	0.229	0.385	0.359	0.334	0.546	0.124
VFN	0.267	0.387	0.450	0.460	0.639	0.257
A2-RL	< 0.230	< 0.385	-	-	-	-
LDAIC	0.276	0.431	0.401	0.541	0.728	0.305

由表4可见,本文算法在Acc,和Acc10上有显著 优势.VFN在SRCC指标上表现更优,但相对Acc指 标而言,SRCC指标考虑低质量区域之间的关系,对 裁剪结果的影响较小.这说明本文算法在跨域情况 下更容易找到高质量的裁剪区域.另外,由于A2-RL 只会生成一个最好的裁剪子图,只能计算Acc15 和 $Acc_{1/10}$, 而

 $Acc_{1/5} > Acc_{5}, Acc_{1/10} > Acc_{10},$ 本文将 $Acc_{1/5}$ 和 $Acc_{1/10}$ 作为 Acc_{5} 和 Acc_{10} 的上界.

3.4 INDOOR↔OUTDOOR 域迁移实验

为了进一步探究 GAICD、CPC 数据集之间分布 的差异,采用文献[37]的场景识别方法对 GAICD、 CPC 数据集进行分析,将所有数据分为室内/室外两 类.实验结果显示:GAICD 数据集上图像室外场景 占 89%,室内场景占11%;CPC 数据集图像室外场 景占53%,室内场景占47%.这或许能从图像层面 上解释2个数据集之间存在的图像分布差异.于是 本文将室内场景(INDOOR)和室外场景 (OUTDOOR)交替作为源域和目标域,验证算法 性能.

实验6 INDOOR↔OUTDOOR 域迁移实验

按照3.1节描述,本文将GAICD、CPC数据集上的数据分成室内/室外两类,分别命名为INDOOR、 OUTDOOR数据集,并在其上进行基准方案和本文 算法的对比实验.参照CPC→GAICD的实验设置, 每个批次的迭代次数由2个数据集上训练集较小的 那个决定,本次实验设置为3936次.为了避免随机 误差造成的影响,选取随机种子{0,10,100,1000, 10000}进行实验,并取平均值作为实验结果,具体 如表5所示.由表可见,相比基准方案,本文算法在 各指标上均有小幅提升,Acc指标提升较显著.

表 5 各方法在 INDOOR↔OUTDOOR 上的结果

Table 5 Results of different methods on INDOOR↔OUTDOOR

- 古注	INDOO	$R \rightarrow OI$	JTDOOR	OUTDO	$OOR \rightarrow 1$	INDOOR
	Acc_5	Acc_{10}	SRCC	Acc_5	Acc_{10}	SRCC
基准方案	0.637	0.791	0.437	0.750	0.894	0.503
LDAIC	0.644	0.800	0.449	0.758	0.903	0.505

3.5 消融实验

为了验证本文算法中两个域适应策略的有效 性,实验设置 CPC → GAICD 作为迁移域,在固定算 法结构的情况下,针对损失函数设计一组消融实验.

本次实验旨在验证两个域适应策略的有效性, 包括4种方案.4种方案均采用2.2节中的结构.方 案1为本文基准方案,将美学损失 L_A 作为损失函数 优化算法,不采用任何域适应策略约束模型,是实验 的对照组.方案2将 L_A + $\lambda_1 L_d$ 作为损失函数优化算 法,利用域判别器,验证对抗域适应策略的有效性. 方案3将 L_A + $\lambda_2 (L_c + \mu L_w)$ 作为损失函数优化算 法,验证本文美学不变特征的有效性. 方案4将L₃作 为损失函数,为本文算法.

各方案的消融实验结果如表 6 所示. 由表可知, 相比方案 1,方案 2 性能有一定提升,对抗域适应策 略即使被单独添加到算法中也能提升裁剪算法在目 标域的泛化能力,而方案 3 的美学不变特征策略在 单独作用的情况下对算法性能提升较小. 但方案 4(本文算法)将对抗域适应策略和美学不变特征策 略结合后,性能得到显著提升,这验证域适应策略的 有效性.

表6 不同方案的消融实验结果

Table 6 Results of different ablation experiment schemes

方案	Acc_5	Acc_{10}	SRCC
方案1	0.249	0.417	0.360
方案2	0.264	0.422	0.377
方案3	0.251	0.420	0.365
方案 4	0.276	0.431	0.401

对于方案 3,本文认为源域和目标域的图像数 据差异过大,仅设置 2 个参数不同的美学分类器很 难约束算法提取到适用于目标域样本的美学不变特 征,算法可能被源域数据过拟合.但结合对抗域适应 策略后,2 个域样本特征的边缘分布被对齐,2 个参 数不同的美学分类器能更好地利用无标签的目标域 图像,提取美学不变特征.此时的特征提取器一边学 习混淆域判别器的能力,一边从源域向目标域传递 其学到的美学鉴赏能力,最终在跨域的智能裁剪任 务上达到更优效果.

3.6 可视化结果

为了更直观地体现本文算法效果,展示本文算 法的裁剪结果和特征分布,各对比算法的裁剪结果 由图 3 给出.





(a)原始图像(a)Original images







(b)基准方案(b)Baseline





(c)LDAIC







(d) A2-RL







(e)VFN

图 3 各算法的裁剪结果对比 Fig. 3 Comparison of cropping results of different algorithms

由图3可看出,本文算法能有效适应各种不同 场合,裁剪后的图像包含原始图像的显著区域,并且 显著目标处在裁剪图约三分之一处,满足摄像构图 中的三分法原理.而其它算法的裁剪结果虽然也能 将原图显著区域包含在内,但显著目标所在的位置 会导致裁剪图不够美观.

4 结束语

本文提出基于序列对抗域适应的智能裁剪算 法.在不需要目标域美学标签的情况下,仅依靠源域 数据和部分无标注的目标域图像实现跨域的智能裁 剪学习.建立一个端到端的卷积神经网络模型,主要 包含美学评分模块和对抗域适应模块.最后通过实 验验证本文算法在跨数据集、跨场景域时的有效性. 值得关注的是,本文结果依然与有监督情况下的智 能裁剪结果有较大差距,今后将考虑进一步挖掘与 图像裁剪相关的不变特征,提升无监督情况下的智 能裁剪算法的性能.

参考文献

- YAN J Z, LIN S, KANG S B, et al. Learning the Change for Automatic Image Cropping // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2013: 971–978.
- [2] CHEN F, LIN Z, MĚCH R, et al. Automatic Image Cropping Using Visual Composition, Boundary Simplicity And Content Preservation Models // Proc of the 22nd ACM International Conference on Multimedia. New York, USA; ACM, 2014; 1105-1108.
- [3] NISHIYAMA M, OKABE T, SATO Y, et al. Sensation-Based Photo Cropping // Proc of the 17th ACM International Conference on Multimedia. New York, USA: ACM, 2009: 669-672.
- [4] PARK J, LEE J Y, TAI Y W, et al. Modeling Photo Composition and Its Application to Photo Re-arrangement // Proc of the 19th IEEE International Conference on Image Processing. Washington, USA: IEEE, 2012: 2741-2744.
- [5] MURRAY N, MARCHESOTTI L, PERRONNIN F. AVA: A Large-Scale Database for Aesthetic Visual Analysis // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2012: 2408-2415.
- [6] KONG S, SHEN X H, LIN Z, et al. Photo Aesthetics Ranking Network with Attributes and Content Adaptation // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 662-679.
- [7] LU P, ZHANG H, PENG X J, et al. An End-to-End Neural Network for Image Cropping by Learning Composition from Aesthetic Photos[C/OL]. [2021-01-25]. https://arxiv.org/pdf/1907. 01432.pdf.
- [8] TU Y, NIU L, ZHAO W J, et al. Image Cropping with Composition and Saliency Aware Aesthetic Score Map // Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 12104-12111.
- [9] CHEN J S, BAI G C, LIANG S H, et al. Automatic Image Cropping: A Computational Complexity Study // Proc of the IEEE Con-

ference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 507-515.

- [10] LI Z P, ZHANG X Y. Collaborative Deep Reinforcement Learning for Image Cropping // Proc of the IEEE International Conference on Multimedia and Expo. Washington, USA: IEEE, 2019: 254– 259.
- [11] WANG W G, SHEN J B, LING H B. A Deep Network Solution for Attention and Aesthetics aware Photo Cropping. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(7): 1531– 1544.
- [12] CHEN L Q, XIE X, FAN X, et al. A Visual Attention Model for Adapting Images on Small Displays. Multimedia Systems, 2003, 9(4): 353-364.
- [13] MARCHESOTTI L, CIFARELLI C, CSURKA G. A Framework for Visual Saliency Detection with Applications to Image Thumbnailing // Proc of the 12th IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2009, 1: 2232-2239.
- [14] LU P, LIU J H, PENG X J, et al. Weakly Supervised Real-Time Image Cropping Based on Aesthetic Distributions // Proc of the 28th ACM International Conference on Multimedia. New York, USA: ACM, 2020: 710-731.
- [15] CHEN Y L, HUANG T W, CHANG K H, et al. Quantitative Analysis of Automatic Image Cropping Algorithms: A Dataset and Comparative Study // Proc of the IEEE Winter Conference on Applications of Computer Vision. Washington, USA: IEEE, 2017: 226-234.
- [16] WEI Z J, ZHANG J M, SHEN X H, et al. Good View Hunting: Learning Photo Composition from Dense View Pairs // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2018: 5437-5446.
- [17] ZENG H, LI L D, CAO Z S, et al. Reliable and Efficient Image Cropping: A Grid Anchor Based Approach // Proc of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2019: 5942-5950.
- [18] DENG Y B, CHEN C L, TANG X O. Image Aesthetic Assessment: An Experimental Survey. IEEE Signal Processing Magazine, 2017, 34(4): 80-106.
- [19] CHEN Y L, KLOPP J, SUN M, et al. Learning to Compose with Professional Photographs on the Web // Proc of the 25th ACM International Conference on Multimedia. New York, USA: ACM, 2017: 37-45.
- [20] LI D B, WU H K, ZHANG J G, et al. A2-RL: Aesthetics Aware Reinforcement Learning for Image Cropping // Proc of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2018: 8193-8201.
- [21] TZENG E, HOFFMAN J, ZHANG N, et al. Deep Domain Confusion: Maximizing for Domain Invariance [C/OL]. [2021-01-25]. https://arxiv.org/pdf/1412.3474v1.pdf.
- [22] GHIFARY M, KLEIJN W B, ZHANG M J. Domain Adaptive Neural Networks for Object Recognition // Proc of the Pacific Rim International Conference on Artificial Intelligence. Berlin, Germany: Springer, 2014: 898–904.
- [23] LONG M S, CAO Y, WANG J M, et al. Learning Transferable

Features with Deep Adaptation Networks // Proc of the 32nd International Conference on Machine Learning. New York, USA: ACM, 2015: 97-105.

- [24] GONG B Q, SHI Y, SHA F, et al. Geodesic Flow Kernel for Unsupervised Domain Adaptation // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2012: 2066-2073.
- [25] LONG M S, ZHU H, WANG J M, et al. Unsupervised Domain Adaptation with Residual Transfer Networks // Proc of the 30th International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2016: 136-144.
- [26] ZHANG Y B, TANG H, JIA K, et al. Domain-Symmetric Networks for Adversarial Domain Adaptation // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2019: 5031-5040.
- [27] CICEK S, SOATTO S. Unsupervised Domain Adaptation via Regularized Conditional Alignment // Proc of the IEEE/CVF International Conference on Computer Vision. Washington, USA: IEEE, 2019: 1416-1425.
- [28] LONG M S, CAO Z J, WANG J M, et al. Conditional Adversarial Domain Adaptation // Proc of the 32nd International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2018: 1640–1650.
- [29] TSAI Y H, HUNG W C, SCHULTER S, et al. Learning to Adapt Structured Output Space for Semantic Segmentation // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2018: 7472-7481.
- [30] CHEN Y H, LI W, SAKARIDIS C, et al. Domain Adaptive Faster R-CNN for Object Detection in the Wild // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2018: 3339-3348.
- [31] GANIN Y, LEMPITSKY V. Unsupervised Domain Adaptation by Backpropagation // Proc of the 32nd International Conference on Machine Learning. New York, USA: ACM, 2015: 1180-1189.
- [32] LI D B, ZHANG J G, HUANG K Q, et al. Composing Good Shots by Exploiting Mutual Relations // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington,

USA: IEEE, 2020: 4212-4221.

- [33] LI D B, ZHANG J G, HUANG K Q, et al. Learning to Learn Cropping Models for Different Aspect Ratio Requirements // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2020: 12682-12691.
- [34] HWANG S, PARK J, KIM N, et al. Multispectral Pedestrian Detection: Benchmark Dataset and Baseline // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 1037–1045.
- [35] HE K M, CKIOXARI G, DOLLAR P, et al. Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [36] LU W R, XING X F, CAI B L, et al. Listwise View Ranking for Image Cropping. IEEE Access, 2019, 7: 91904–91911.
- [37] ZHOU B L, LAPEDRIZA A, KHOSLA A, et al. Places: A 10 Million Image Database for Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (6): 1452-1464.

作者简介



王皓文,硕士研究生,主要研究方向为计算 机视觉、图像处理、模式识别. E-mail:whw95 @ hust. edu. cn.

(**WANG Haowen**, master student. His research interests include computer vision, image processing and pattern recognition.)



桑农(通信作者),博士,教授,主要研究方向为目标检测与识别、目标跟踪、图像/视频语义分割、监控视频智能处理与分析等.
E-mail:Nsang@hust.edu.cn.

(SANG Nong(Corresponding author), Ph. D., professor. His research interests include target

detection and recognition, target tracking, image/video semantic segmentation, surveillance video intelligent processing and analysis.)