

一种基于区分区域定位的细粒度图像识别方法

杨虹, 范勇

(西南科技大学 计算机科学与技术学院, 四川 绵阳 621010)

摘要: 细粒度图像识别的目标为区分大类对象中的子类对象, 由于子类对象间差别细微, 使得细粒度图像识别较为困难。为此, 提出一种基于区分区域定位的细粒度图像识别方法。首先由贝叶斯个性化排序损失 (Bayesian Personalized Ranking Loss, BPRLoss) 监督区域提议网络提议一些重要的局部区域, 随后采用引入高效通道注意力模块的特征提取器提取局部区域的细粒度特征进行识别。同时采用标签平滑策略使同类靠近, 不同类远离以监督网络学习对象有区别的特征, 进一步促进网络定位区分区域。实验结果表明, 所提方法在三种通用的细粒度图像识别数据集 CUB-200-2011、FGVC Aircraft、Stanford Cars 上取得了较高的识别准确率, 分别为 89.0%、93.9%、94.3%, 相比导航网络 (NTS-Net) 有显著提升, 分别提升 1.5 百分点、2.5 百分点和 0.4 百分点。同时, 所提方法较 NTS-Net 能够更为有效地定位区分区域和提取图像的细粒度特征。

关键词: 细粒度图像识别; 通道注意力; 标签平滑; 区域定位; 特征提取

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2023)11-0169-06

doi: 10.3969/j.issn.1673-629X.2023.11.025

A Fine Grain Image Recognition Method Based on Distinguishable Region Location

YANG Hong, FAN Yong

(School of Computer Science and Technology, Southwest University of Science and Technology,
Mianyang 621010, China)

Abstract: The goal of fine-grained image recognition is to distinguish sub-class objects in large class objects. Because of the subtle differences between sub-class objects, fine-grained image recognition is more difficult. For this reason, a fine-grained image recognition method based on differentiated region location is proposed. Firstly, the Bayesian Personalized Ranking Loss (BPRLoss) supervised region proposes that the network proposes some important local regions, and then uses the feature extractor introducing the efficient channel attention module to extract the fine-grained features of the local regions for recognition. At the same time, the tag smoothing strategy is used to make the same class close and different classes far away to monitor the different characteristics of the network learning objects, and further promote the network location to distinguish regions. The experimental results show that the proposed method has achieved high recognition accuracy on three common fine-grained image recognition data sets CUB-200-2011, FGVC Aircraft and Stanford Cars, which are 89.0%, 93.9% and 94.3%, respectively. Compared with the navigation network (NTS-Net), it has significantly improved by 1.5 percentage points, 2.5 percentage points and 0.4 percentage points respectively. At the same time, the proposed method is more effective than NTS-Net in locating and distinguishing regions and extracting fine-grained features of images.

Key words: fine-grained image recognition; channel attention; label smoothing; region location; feature extraction

0 引言

近年来, 细粒度图像识别因其广泛的现实应用得到了较高的关注, 成为计算机视觉领域的热门话题。细粒度图像识别的目标是区分一个大类中不同子类对象间的细微差异。如不同种类的鸟, 不同型号的汽车

等。因不同子类对象间细微的类间差异和较大的类内差异^[1]导致细粒度图像识别任务较一般的图像识别任务更具挑战。因此, 如何定位图像的区分区域和高效提取区分区域的细粒度特征是细粒度图像识别任务的主要难点。一些研究^[2-4]通过使用人工标注的图像

收稿日期: 2023-01-04

修回日期: 2023-05-05

基金项目: 四川省科技重点研发项目 (2021YFG0031)

作者简介: 杨虹 (1997-), 男, 硕士生, 研究方向为计算机视觉、细粒度图像识别; 通信作者: 范勇 (1972-), 男, 教授, 博士, 研究方向为视觉测量技术、图像分析与理解。

部位边界框监督网络检测关键部位区域并进行识别,取得了一定的效果。然而,密集的人工标注耗费大量的人力和物力,使得基于人工标注的识别方法难以应用和扩展。为此,一些研究^[5-8]开始尝试在仅使用图像类别标签的情况下直接对图像区分区域进行弱监督定位和识别。其中,Yang 等^[8]提出 NTS-Net,采用合页损失(Hinge Loss)作为排序损失监督区域提议网络(RPN)^[9]提议出图像的重要区域进行识别。NTS-Net^[8]在一定程度上解决了以往定位识别需要部位标签的问题。

但是,NTS-Net^[8]采用的 Hinge Loss 难以使 RPN^[9]分值列表与对应的置信度列表尽可能保持同序,从而减弱了网络挖掘重要区域的能力。其次,NTS-Net^[8]没有对网络的不同特征通道进行关联学习,难以挖掘对象的细粒度特征。并且,NTS-Net^[8]仅采用排序函数监督往往会定位信息量大而非区分区域,从而提议出较大的框^[10]。为此,该文提出一种基于区分区域定位的细粒度图像识别方法。首先,使用贝叶斯个性化排序损失(BPRLoss)^[11]监督网络以更有效地挖掘重要区域。随后,为加强细粒度特征的提取,采用引入高效通道注意力(ECA)^[12]模块的特征提取器提取区域特征。同时采用标签平滑^[13]策略使同类靠近,不同类远离以监督网络学习有区别的特征,进一步定位区分区域。

1 相关工作

本节将介绍与所提方法相关的工作,包括区分区域定位、通道注意力机制和标签平滑方法。

1.1 区分区域定位

细粒度图像识别中待识别对象之间往往仅存在一些局部,细微的差别。因此,如何有效学习定位出对象具有区分性的局部区域对细粒度图像识别起着重要作用。深度卷积网络滤波器(CNN filter)的响应往往对应于图像的局部区域。因此,一些研究^[14-16]通过使用深度卷积网络滤波器(CNN filter)作为对象部位检测器,从而定位出一些具有区分性的区域。其中,Wang 等^[14]使用了额外的可学习 1×1 卷积滤波器作为对象部位检测器。随后通过全局最大池化层得到滤波器的最高响应用于对象识别,取得了较好的识别效果。注意力机制可以帮助人聚焦于对象的突出部分。受此启发,一些研究^[5-7]使用注意力机制来学习定位对象可区分区域。Fu 等^[5]使用循环注意力模型来定位对象的区分区域。Zheng 等^[6]通过聚合响应同一区域的特征通道来定位关键区域。Zheng 等^[7]设计了一个三线性注意力采样网络,从数百个部位区域提议中学习细粒度细节,并将学习到的特征有效地提取到单个深度

卷积网络中。但是,上述方法都难以对所提议区域的重要性程度进行量化。为此,一些研究^[8,10,17-18]通过借鉴目标检测的方法,使用合页损失(Hinge Loss)作为排序损失监督区域提议网络(RPN)^[9]提议图像的重要区域。然而,如式(1)所示,当正负样本距离大于阈值 1 时,Hinge Loss 值为 0,难以监督目标尽可能地保持同序,从而减弱了网络挖掘图像重要区域的能力。为此,在 Yang 等^[8]方法的基础上,该文提出使用贝叶斯个性化排序损失(BPRLoss)^[11]监督 RPN 网络以更有效地定位图像的重要区域。

$$\max(0, 1 - (s^+ - s^-)) \quad (1)$$

1.2 通道注意力机制

注意力机制在人类感知中起着重要的作用。人们可以通过注意力机制的帮助来筛选出对象的一些重点信息。受此启发,许多研究尝试在深度神经网络中引入不同的注意力机制以促使网络能够有效关注对象的重要特征。通道注意力机制往往被用于挖掘网络不同特征通道之间的关联性。其主要作用是分配每个通道不同的权重,促使网络关注重要的特征,同时抑制不重要的特征。挤压和激励(SE)模块^[19]首次提出一种有效的机制来学习通道注意,并取得了优异的性能。通过对 SE 模块^[19]的深入研究,Wang 等^[12]认为 SE 模块^[19]采用降维的方式降低模块复杂度会给通道的注意学习带来副作用,因此提出高效通道注意力(ECA)模块,采用无降维的通道交互策略,通过一维卷积来使每个通道与相邻的 K 个通道进行交互学习,进一步提高了网络性能。为加强网络对细粒度特征的提取,文中方法使用引入 ECA 模块^[12]的特征提取器提取区域的细粒度特征。通过关联不同的特征通道,促使网络关注对象更具辨别性的细粒度特征。

1.3 标签平滑

Szegedy 等^[13]首次提出一种标签平滑方法,通过利用 one-hot 标签和标签上均匀分布之间的平均值来生成软标签,从而对网络正则化。通过深入研究标签平滑,Zhang 等^[20]认为非目标类不能简单的同等对待,与目标类别相似类别应当被分配更高的标签值,因此提出一种在线标签平滑方法。通过利用网络中的统计信息指导生成软标签,更合理地对待非目标类别。为探索标签平滑^[13]的作用原理,Müller 等^[21]通过可视化方法观察使用标签平滑后分类器的倒数第二层,发现标签平滑可以使同类靠近,不同类远离。针对仅使用排序函数监督网络难以定位可区分区域的问题,该文使用 Szegedy 等^[13]提出的标签平滑方法生成软标签监督网络,通过使同类靠近,不同类远离使网络学习有区别的特征,从而促进网络进一步定位区分区域。

2 文中方法

本节先介绍了文中方法的网络结构与整体流程,随后分别介绍了区域定位方法 bprRPN,引入高效通道注意力机制的特征提取器网络和标签平滑方法。

2.1 网络结构与整体流程

网络结构如图 1 所示。图片首先由引入高效通道注意力(ECA)模块^[12]的特征提取器提取特征,随后分别经过线性层和区域提议网络(RPN)^[9]。由线性层得到对象特征分类向量,并与标签平滑后的图像标签使用交叉熵损失函数计算得到对象特征损失(object loss);由 RPN 提议 M 个局部区域及对应分值,并在原图裁剪这 M 个局部区域并上采样至 224×224 大小。再由引入 ECA 模块^[12]的特征提取器得到 M 个局部区域特征,经线性层和交叉熵损失函数计算得到局部区域置信度的列表。随后通过 BPRLoss 损失函数^[11]监督 M 个局部区域分值的列表和置信度的列表同序,使 RPN^[9]总能够提议重要的区域。再在这 M 个局部区域特征中选择前 N 个区域特征作为部位特征。将部位特征与对象特征拼接(concat)后得到拼接特征。部位特征与拼接特征由线性层得到相应的分类向量,随后分别与标签平滑^[13]后的图像标签使用交叉熵损失函数计算得到部位损失(part loss)和拼接损失(concat loss)。最后,由拼接特征分类向量得出分类结果。

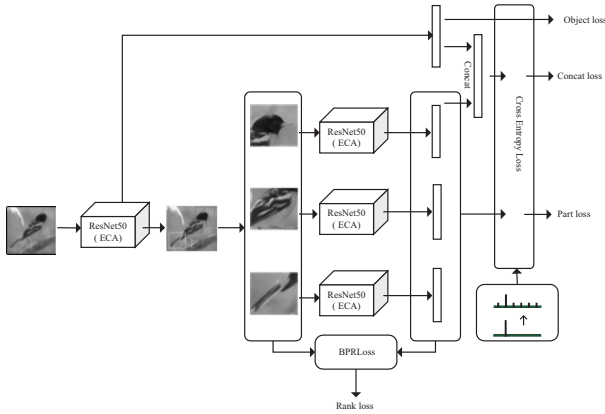


图 1 网络结构

2.2 区域定位方法 bprRPN

为有效定位图像的重要区域,该文提出一种 BPRLoss^[11]监督区域提议网络(RPN)^[9]挖掘图像重要区域的方法。首先将原图调整为 448 大小并为其预设一系列不同大小(实验设置为 48、96、192)和不同比例(实验设置为 1:1、2:3、3:2)的锚框。通过特征提取器得到特征图,再经 RPN^[9]得到预设锚框区域的分值。因此可以得到一个初始的区域分值的列表 $\{S(R_1), S(R_2), \dots, S(R_M)\}$ 。将列表按分值从高到低排序并通过非极大值抑制(NMS)去除重合大且分值小的框。再在剩下的框中选取分值最高的前 M 个

区域,得到 M 个区域分值的列表 $\{S(R_1), S(R_2), \dots, S(R_M)\}$ 。随后, M 个区域由特征提取器提取特征,经线性层和交叉熵损失函数后得到 M 个区域置信度的列表 $\{C(R_1), C(R_2), \dots, C(R_M)\}$ 。Yang 等^[8]认为一个区域的置信度越高,则它对这个图像的识别越重要。基于这一思想,文中方法使用 BPRLoss 损失函数^[11]监督分值列表与置信度列表同序以使网络总能够定位重要的区域。具体来说,文中方法通过 BPRLoss^[11]鼓励置信度越高的区域其在分值列表中有更高的排名,即期望当区域置信度 $C(R_j) > C(R_i)$ 时,总有区域分值 $S(R_j) > S(R_i)$,即有:

$$L_{\text{rank}} = - \sum_{C(R_i) < C(R_j)} \log(\text{sigmoid}(S(R_j) - S(R_i))) \quad (2)$$

2.3 引入高效通道注意力的特征提取器

如何有效提取细粒度特征是细粒度图像识别的一个关键问题。NTS-Net^[8]没有关联不同的特征通道,不能有效提取图像的细粒度特征。为此,该文通过在特征提取器中引入高效通道注意力(ECA)模块^[12]以加强对图像细粒度特征的提取。

以 ResNet50 为例,在网络的每一个残差块中添加 ECA 模块^[12],从而构成 ECA-Res-block,如图 2 所示,其中 BN 为批归一化,Conv2d 为二维卷积,Conv1d 为一维卷积,GAP 为全局平均池化。ECA 主要通过建模相邻 K 个通道之间的关系来生成通道注意力权重。

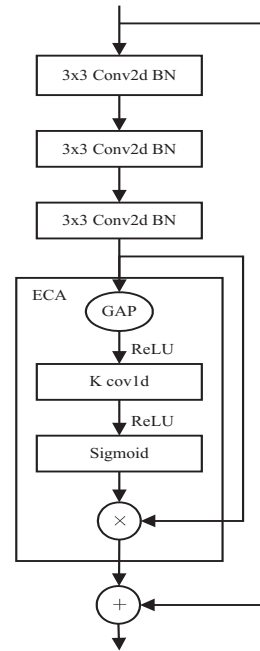


图 2 ECA-Res block 结构

首先,特征图 $x \in R^{W \times H \times C}$ 作为 ECA 模块的输入,经过一个全局平均池化层得到特征向量 $y \in R^C$,再通过一个卷积核大小为 k 的一维卷积来计算特征图相邻 k 个通道的跨通道交互。其中 k 可以根据通道数自适

应计算得到。然后,再由 sigmoid 函数计算得到注意权重 ω 。最后,将特征图与注意权重逐点相乘得到注意后的特征图。注意权重 ω 的计算过程如式(3)所示。

$$\omega = \sigma(\text{C1D}_k(\text{GAP}(x))) \quad (3)$$

其中, σ 表示 sigmoid 激活函数, C1D_k 表示一维卷积操作, GAP 表示全局平均池化操作。

2.4 标签平滑方法

Yang 等^[10]认为,NTS-Net^[8]仅使用排序函数监督网络往往会定位信息量大而非可区分的区域,从而提出较大的框。Schroff^[22]、Yang 等^[10]认为通过最小化同类距离和最大化不同类距离有利于网络学习具有区别的特征。标签平滑^[13]被证实能够使相同类靠近,不同类远离^[21]。针对使用排序函数监督网络会引入大量背景信息,难以定位图像的区分区域的问题,该文使用标签平滑^[13]生成软标签以促使网络学习有区别的特征,进一步定位区分区域,从而提议更小且更准确的区域框。具体来说,设原始标签为 p_i ,即:

$$p_i = \begin{cases} 1, & \text{if}(i = y) \\ 0, & \text{if}(i \neq y) \end{cases} \quad (4)$$

标签平滑后有:

$$p_i = \begin{cases} 1 - \varepsilon, & \text{if}(i = y) \\ \frac{\varepsilon}{k - 1}, & \text{if}(i \neq y) \end{cases} \quad (5)$$

其中, k 为类别数, ε 为超参数。通过平滑标签操作,网络能够更有效地定位区分区域,进一步提高识别准确率。

2.5 文中方法损失函数

文中方法的损失函数包括 4 个部分,即:

$$L = \alpha L_{\text{rank}} + \beta L_{\text{obj}} + \lambda L_{\text{part}} + \gamma L_{\text{concat}} \quad (6)$$

其中, L_{rank} 为排序损失, L_{obj} 为对象损失, L_{part} 为部位损失, L_{concat} 为拼接损失。 α 、 β 、 λ 、 γ 为损失权重。

3 实验

3.1 数据集

为验证文中方法的有效性,将文中方法在三种数据集 CUB-200-2011 (CUB)、FGVC Aircraft (Air)、Stanford Cars (Car) 上分别进行实验。三种数据集的划分如表 1 所示。

表 1 数据集划分

数据集	类别数	训练集	测试集
CUB-200-2011	200	5 994	5 794
FGVC Aircraft	100	6 667	3 333
Stanford Cars	196	8 814	8 041

3.2 实验细节

实验使用在 ImageNet 上预训练好的 ResNet-50

作为网络的 Backbone。对训练图片上采样至 600×600 大小后经中心裁剪为 448×448 大小,随后对其进行随机水平翻转。对测试图片不采用随机水平翻转。网络定位的重要区域数 N 通过定位区域数实验得出。区域定位方法 bprRPN 监督的区域数 M 设置为 6。NMS threshold 设置为 0.25。对于标签平滑操作,超参数 ε 设置为 0.1。网络使用 Momentum SGD 作为优化器,初始学习率设置为 0.001,权重衰减设置为 0.000 1。学习率每 60 epoch 乘以 0.1。batchsize 大小设置为 12。损失函数权重设置为 $\alpha = \beta = \lambda = \gamma = 1$ 。实验中评价指标采用 Top-1 准确率。

3.3 实验结果分析

表 2 展示了文中方法与一些主流方法在 CUB-200-2011、FGVC Aircraft、Stanford Cars 三种数据集上的识别准确率的对比。可以看出,文中方法在三种数据集上都取得了最好结果,分别为 89.0%、93.9% 和 94.3%。同时,文中方法相比 NTS-Net 在三种数据集上的识别准确率分别提升了 1.5 百分点、2.5 百分点和 0.4 百分点,表明了文中方法的有效性。此外,文中方法相比 CIN 在三种数据集上的识别准确率分别提升了 1.5 百分点、1.3 百分点和 0.2 百分点,相比 FDL 在三种数据集上的识别准确率分别提升了 0.5 百分点、0.4 百分点和 0.1 百分点,相比 PCA-Net 则在 CUB、Air 二种数据集上有显著提升,分别提升 0.6 百分点和 1.5 百分点。

表 2 不同方法实验结果

Methods	Backbone	Top-1 Accuracy/%		
		CUB	Air	Car
B-CNN ^[23]	VGG	85.1	84.1	91.3
RA-CNN ^[5]	VGG	85.3	88.5	92.5
MA-CNN ^[6]	VGG	86.5	89.9	92.8
DFB ^[14]	VGG	87.4	91.7	93.1
NTS-Net ^[8]	ResNet-50	87.5	91.4	93.9
TASN ^[7]	ResNet-50	87.9	-	93.8
CIN ^[24]	ResNet-50	87.5	92.6	94.1
FDL ^[17]	ResNet-50	88.5	93.3	94.2
PCA-Net ^[25]	ResNet-50	88.3	92.4	94.3
Ours	ResNet-50	89.0	93.9	94.3

3.4 定位区域数量实验

为探索定位区分区域数量 N 的取值对文中方法识别准确率的影响,记录定位区域数 N 为 3~5 时,文中方法在三种数据集上识别准确率的变化。

实验结果如表 3 所示,在 CUB-200-2011、FGVC Aircraft、Stanford Cars 三种数据集上,定位区域数 N 取值为 4 时,算法有最高的识别准确率,分别为 89.0% 和 93.9% 以及 94.3%。

表 3 定位区域数量实验

N	Top-1 Accuracy/%		
	CUB	Air	Car
3	88.1	92.3	93.9
4	89.0	93.9	94.3
5	88.7	92.0	94.1

3.5 消融实验

为验证文中方法各模块的有效性,将文中方法在 CUB-200-2011、FGVC Aircraft、Stanford Cars 三种数据集上进行了消融实验。如表 4 所示,通过在 ResNet-50 网络中引入提出的区域定位方法 bprRPN 后,网络在三种数据集上的识别准确率分别提升 2.3 百分点、2.3 百分点和 4.0 百分点,表明区域定位方法 bprRPN 能够提议图像的重要区域以促进图像的识别。继续引入 ECA 模块^[12]后,网络在三种数据集上的识别准确率分别能够继续提升 0.3 百分点、0.6 百分点和 0.2 百分点,表明通过引入 ECA 模块^[12]能使网络提取区分性更强的细粒度特征。进一步,采用标签平滑策略^[13]监督网络,网络在三种数据集上的识别准确率分别能够继续提升 0.9 百分点、0.7 百分点和 0.3 百分点,表明通过采用标签平滑策略^[13]能使网络进一步定位区分区域从而有效提高识别效果。此外,如表 5 所示,区域定位方法 bprRPN 相较 NTS-Net^[8]在 CUB-200-2011 和 FGVC Aircraft 数据集上有显著提升,分别提升 0.3 百分点和 1.2 百分点,仅在 Stanford Cars 数据集上有 0.1 百分点的下降。

表 4 消融实验

Methods	Top-1 Accuracy/%		
	CUB	Air	Car
ResNet50	85.5	90.3	89.8
ResNet50+bprRPN	87.8	92.6	93.8
ResNet+bprRPN+ECA	88.1	93.2	94.0
Ours	89.0	93.9	94.3

表 5 区域定位方法比较

Methods	Backbone	Top-1 Accuracy/%		
		CUB	Air	Car
NTS-Net	ResNet50	87.5	91.4	93.9
bprRPN	ResNet50	87.8	92.6	93.8

3.6 可视化实验

图 3 为网络注意图的可视化实验结果。如图 3 (b)(c)所示,区域定位方法 bprRPN 提取对象细粒度特征的能力较 NTS-Net^[8]有一定的提升,但仍不能很好地提取到对象更具区分性的细粒度特征。如图 3 (d)所示,引入高效通道注意力(ECA)^[12]模块后,通过关联不同的特征通道,文中方法能够更有效地提取

对象区分性的细粒度特征。

图 4 为网络定位区域的可视化结果。如图 4 (a)(b)所示,仅通过排序损失监督网络往往会提议较大的框,引入了大量的背景信息。该文通过使用标签平滑^[13]进一步监督网络,促使网络学习有区别的特征,从而定位对象的区分区域。如图 4 (c)所示,网络定位的局部区域更聚焦于对象的关键部位,且不会引入大量的背景信息。同时,如图 4 (a)(c)所示,相比 NTS-Net^[8],文中方法更能有效定位区分区域。

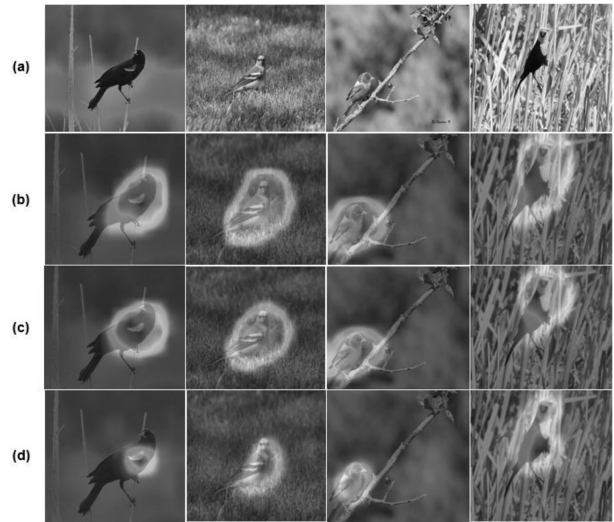


图 3 注意图可视化

(a) 原图 (b) NTS-Net (c) bprRPN (d) bprRPN+ECA

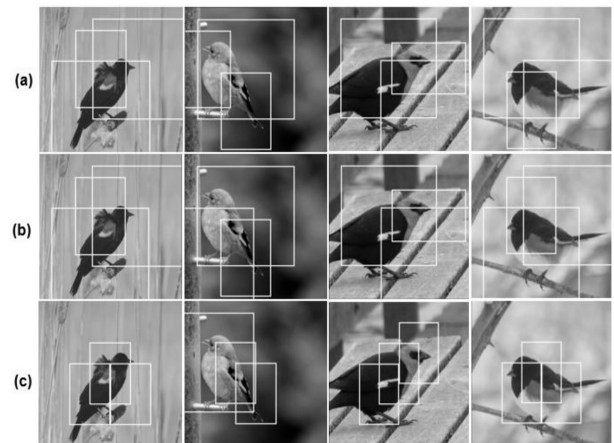


图 4 定位区域可视化

(a) NTS-Net (b) Ours (w/o LS) (c) Ours

4 结束语

针对细粒度图像识别对象往往仅存一些局部、细微的区别,该文提出了一种基于区分区域定位的细粒度图像识别方法。首先,使用提出的区域定位方法 bprRPN 定位一些重要区域。随后,通过引入高效通道注意力(ECA)^[12]模块的特征提取器提取特征进行识别。同时使用标签平滑策略监督网络,促进网络定位区分区域。最终,文中方法在三种数据集上取得了较

高的识别准确率。同时相比 NTS-Net^[8] 能够更为准确地定位区分区域。尽管文中方法能够有效地定位区分区域,但并未考虑到不同区域之间的关联性。因此,后续工作将探索如何关联不同的局部区域以提升网络的识别性能。

参考文献:

- [1] 罗建豪,吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报,2017,43(8):1306-1318.
- [2] ZHANG N,DONAHUE J,GIRSHICK R,et al. Part-based R-CNNs for fine-grained category detection[C]//Computer vision - ECCV 2014. Zurich:Springer,2014:834-849.
- [3] WEI X S,XIE C W,WU J,et al. Mask-CNN:localizing parts and selecting descriptors for fine-grained bird species categorization[J]. Pattern Recognition,2018,76:704-714.
- [4] ZHANG H,XU T,ELHOSEINY M,et al. Spda-cnn:unifying semantic part detection and abstraction for fine-grained recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE,2016:1143-1152.
- [5] FU J,ZHENG H,MEI T. Look closer to see better:recurrent attention convolutional neural network for fine-grained image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu:IEEE,2017:4438-4446.
- [6] ZHENG H,FU J,MEI T,et al. Learning multi-attention convolutional neural network for fine-grained image recognition[C]//Proceedings of the IEEE international conference on computer vision. Venice:IEEE,2017:5209-5217.
- [7] ZHENG H,FU J,ZHA Z J,et al. Looking for the devil in the details:learning trilinear attention sampling network for fine-grained image recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach:IEEE,2019:5012-5021.
- [8] YANG Z,LUO T,WANG D,et al. Learning to navigate for fine-grained classification[C]//Proceedings of the European conference on computer vision (ECCV). Munich:Springer,2018:420-435.
- [9] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2015,39(6):1137-1149.
- [10] YANG S,LIU S,YANG C,et al. Re-rank coarse classification with local region enhanced features for fine-grained image recognition[J]. arXiv:2102.09875,2021.
- [11] RENDLE S,FREUDENTHALER C,GANTNER Z,et al. BPR:Bayesian personalized ranking from implicit feedback[J]. arXiv:1205.2618,2012.
- [12] WANG Q,WU B,ZHU P F,et al. ECA-Net:efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle:IEEE,2020:11531-11539.
- [13] SZEGEDY C,VANHOUCKE V,IOFFE S,et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas:IEEE,2016:2818-2826.
- [14] WANG Y,MORARIU V I,DAVIS L S. Learning a discriminative filter bank within a CNN for fine-grained recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:4148-4157.
- [15] DING Y,ZHOU Y,ZHU Y,et al. Selective sparse sampling for fine-grained image recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul:IEEE,2019:6599-6608.
- [16] HUANG Z,LI Y. Interpretable and accurate fine-grained recognition via region grouping[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle:IEEE,2020:8662-8672.
- [17] LIU C,XIE H,ZHA Z J,et al. Filtration and distillation: Enhancing region attention for fine-grained visual categorization[C]//Proceedings of the AAAI conference on artificial intelligence. New York:AAAI,2020:11555-11562.
- [18] 刘洋,金忠. 一种结合非局部和多区域注意力机制的细粒度图像识别方法[J]. 计算机科学,2021,48(1):197-203.
- [19] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:7132-7141.
- [20] ZHANG C B,JIANG P T,HOU Q,et al. Delving deep into label smoothing[J]. IEEE Transactions on Image Processing,2021,30:5984-5996.
- [21] MÜLLER R,KORNBLITH S,HINTON G E. When does label smoothing help? [J]. arXiv:1906.02629,2019.
- [22] SCHROFF F,KALENICHENKO D,PHILBIN J. Facenet: a unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston:IEEE,2015:815-823.
- [23] LIN T Y,ROYCHOWDHURY A,MAJI S. Bilinear CNN models for fine-grained visual recognition[C]//Proceedings of the IEEE international conference on computer vision. Santiago:IEEE,2015:1449-1457.
- [24] GAO Y,HAN X,WANG X,et al. Channel interaction networks for fine-grained image categorization[C]//Proceedings of the AAAI conference on artificial intelligence. New York:AAAI,2020:10818-10825.
- [25] ZHANG T,CHANG D,MA Z,et al. Progressive co-attention network for fine-grained visual classification[C]//2021 international conference on visual communications and image processing (VCIP). Munich:IEEE,2021:1-5.