



(12)发明专利申请

(10)申请公布号 CN 110751195 A

(43)申请公布日 2020.02.04

(21)申请号 201910965021.X

(22)申请日 2019.10.12

(71)申请人 西南交通大学

地址 610000 四川省成都市二环路北一段

(72)发明人 黄进 李啸天 李剑波 朱明仓

李继秀 杨旭 付国栋 秦泽宇

郑思宇 王敏

(74)专利代理机构 成都东恒知盛知识产权代理

事务所(特殊普通合伙)

51304

代理人 何健雄 廖祥文

(51)Int.Cl.

G06K 9/62(2006.01)

G06N 3/04(2006.01)

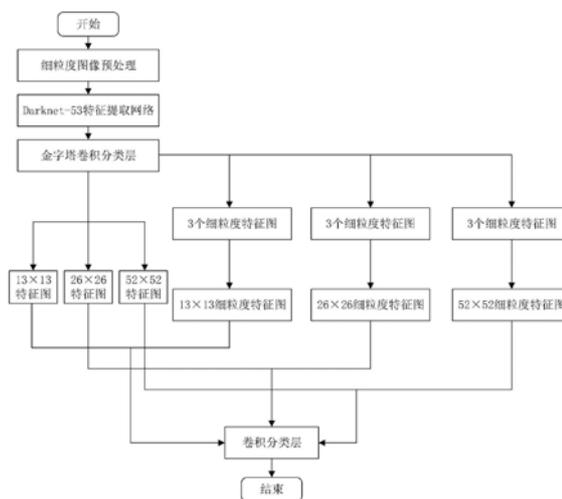
权利要求书3页 说明书5页 附图3页

(54)发明名称

一种基于改进YOLOv3的细粒度图像分类方法

(57)摘要

一种基于改进YOLOv3的细粒度图像分类方法,该方法包括下述步骤:一、对细粒度图像进行预处理;二、通过Darknet-53特征提取网络对输入图像提取特征;三、多层卷积网络得到三种大小的分类特征图,分别为13×13,26×26,52×52;四、提取细粒度特征图。五、将原分类特征图与细粒度特征图融合,改进损失函数,训练网络直到达到预先设定的迭代次数或准确率不再提升;六、细粒度图像类别识别。本发明在不增加图像检测时间的前提下,通过加入细粒度分类层,提高YOLOv3原网络对细粒度图像分类的准确度。



1. 一种基于改进YOLOv3的细粒度图像分类方法,其特征在于,包括如下步骤:

步骤1:把细粒度图像数据集随机分为训练集、验证集、测试集,通过图像预处理将图像缩放至 416×416 ,采用数据增强方法扩充数据集;

步骤2:通过Darknet-53网络提取输入图像特征;

步骤3:经过三个(1×1 卷积+Batch Normalization+LeakyRelu+ 3×3 卷积+Batch Normalization+LeakyRelu)层后进入 1×1 分类卷积层得到尺寸为 $13 \times 13 \times (4+1+类别) \times 3$ 的输出特征图;通过上采样,重复上述操作得到尺寸为 $26 \times 26 \times (4+1+类别) \times 3$ 的输出特征图。再次上采样,重复上述操作得到尺寸为 $52 \times 52 \times (4+1+类别) \times 3$ 的输出特征图;

步骤4:在输出特征图尺寸为 $13 \times 13 \times (4+1+类别) \times 3$ 的网络中抽取第76、78、80层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行 1×1 分类卷积得到尺寸为 $13 \times 13 \times (4+1+类别) \times 3$ 的细粒度分类特征图;在输出特征图尺寸为 $26 \times 26 \times (4+1+类别) \times 3$ 的网络中抽取第80、82、84层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行 1×1 分类卷积得到尺寸为 $26 \times 26 \times (4+1+类别) \times 3$ 的细粒度分类特征图;在输出特征图尺寸为 $56 \times 56 \times (4+1+类别) \times 3$ 的网络中抽取第84、86、88层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换,拼接后进行 1×1 分类卷积得到尺寸为 $56 \times 56 \times (4+1+类别) \times 3$ 细粒度分类特征图;

步骤5:将原YOLOv3网络得到的三个分类特征图和三个细粒度分类特征图拼接得到 $13 \times 13 \times (4+1+类别) \times 3 \times 2$, $26 \times 26 \times (4+1+类别) \times 3 \times 2$, $56 \times 56 \times (4+1+类别) \times 3 \times 2$ 三个尺寸的分类特征图,修改损失函数,将原特征图和细粒度特征图乘以不同的权重系数;

步骤6:训练网络得到权重参数,对细粒度图像分类。

2. 根据权利要求1所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,步骤2中Darknet-53特征提取网络,网络共75层,其框架由Convolutional层和Residual残差连接层组成;每个Convolutional层包含卷积、归一化、LeakyReLU三个操作;Residual残差连接层结构中令输入特征向量为 X ,经过两层Convolutional操作后的输出向量为 X' ,再将输入特征向量矩阵与输出特征向量矩阵逐元素相加得到输出。

3. 根据权利要求1所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,在每个尺度的YOLOv3原网络金字塔卷积分类层中抽取3个 1×1 Convolutional层的输出作为细粒度特征图;在每个尺度上将得到的三个细粒度特征图相互作外积(matrix outer product),经过 $y = \text{Sign}(x) \times \sqrt{|x|+b}$ 函数非线性激活,得到三个大小与输入相同的细粒度特征图,最后将三个特征图拼接成一个特征图。最终得到三个不同尺度的细粒度特征图为 $13 \times 13 \times (1024 \times 3)$ 、 $26 \times 26 \times (512 \times 3)$ 、 $52 \times 52 \times (256 \times 3)$ 。

4. 根据权利要求1所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,步骤3原YOLOv3输出特征图进行 1×1 卷积分类得到三个尺度的输出,即 $13 \times 13 \times (4+1+类别) \times 3$, $26 \times 26 \times (4+1+类别) \times 3$, $52 \times 52 \times (4+1+类别) \times 3$;在每个尺度上预测3个不同大小的预测框,每种大小的预测框值包括4个边框预测值(x, y, w, h), (x, y)对应的是预测目标框的几何中心坐标, (w, h)对应的是预测目标框的宽度与高度;1为该像素点的目标置信度值以及预测的每个类别的得分值;步骤4将三个不同尺度的细粒度特征图进行 1×1 卷积分类得到13

$\times 13 \times (4+1+\text{类别}) \times 3, 26 \times 26 \times (4+1+\text{类别}) \times 3, 52 \times 52 \times (4+1+\text{类别}) \times 3$ 三个输出,并与步骤3输出的结果图进行拼接得到 $13 \times 13 \times (4+1+\text{类别}) \times 3 \times 2, 26 \times 26 \times (4+1+\text{类别}) \times 3 \times 2, 52 \times 52 \times (4+1+\text{类别}) \times 3 \times 2$ 三个输出。

5. 根据权利要求4所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,所述边框预测公式为:

$$b_x = \frac{1}{1 + e^{-(\alpha t_{x1} + (1-\alpha)t_{x2})}} + c_x$$

$$b_y = \frac{1}{1 + e^{-(\alpha t_{y1} + (1-\alpha)t_{y2})}} + c_y$$

$$b_w = p_w e^{(\alpha t_{w1} + (1-\alpha)t_{w2})}$$

$$b_h = p_h e^{(\alpha t_{h1} + (1-\alpha)t_{h2})}$$

其中 c_x, c_y 是特征图中grid cell的左上角坐标,在yolov3中每个grid cell在feature map中的宽和高均为1; $t_{x1}, t_{y1}, t_{h1}, t_{w1}$ 分别是原分类网络预测的边框值, $t_{x2}, t_{y2}, t_{h2}, t_{w2}$ 分别为细粒度分类预测边框值, α 为调解权重的参数,这里取 $\alpha=0.6$ 。

6. 根据权利要求4所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,所述置信度计算公式为:

$$y = \frac{1}{1 + e^{-(\beta d_1 + (1-\beta)d_2)}}$$

其中 d_1, d_2 分别为原网络分类和细粒度分类置信度值, β 为调解权重的参数,这里取 $\beta=0.5$ 。

7. 根据权利要求4所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,所述类别得分计算公式为:

$$y = \frac{1}{1 + e^{-(\gamma p_1 + (1-\gamma)p_2)}}$$

其中 p_1, p_2 分别为原网络分类和细粒度分类预测每个类别的值, γ 为调解权重的参数,这里取 $\gamma=0.6$ 。

8. 根据权利要求1所述的基于改进YOLOv3的细粒度图像分类方法,其特征在于,步骤5所述损失函数计算公式为:

$$\begin{aligned}
\text{loss(object)} = & \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2^{-w_{\text{trath}}} \times h_{\text{trath}}) \left\{ \alpha [(\hat{x}_{1i} \log(x_{\text{trath}}) + (1 - \hat{x}_{1i}) \log(1 - x_{\text{trath}}))] + (1 - \alpha) [\hat{x}_{2i} \log(x_{\text{trath}}) + (1 - \hat{x}_{2i}) \log(1 - x_{\text{trath}})] \right. \\
& \left. + \alpha [(\hat{y}_{1i} \log(y_{\text{trath}}) + (1 - \hat{y}_{1i}) \log(1 - y_{\text{trath}}))] + (1 - \alpha) [\hat{y}_{2i} \log(y_{\text{trath}}) + (1 - \hat{y}_{2i}) \log(1 - y_{\text{trath}})] \right\} + \\
& \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2^{-w_{\text{trath}}} \times h_{\text{trath}}) \left\{ \alpha [(w_{\text{trath}} - \hat{w}_{1i})^2 + (h_{\text{trath}} - \hat{h}_{1i})^2] + (1 - \alpha) [(w_{\text{trath}} - \hat{w}_{2i})^2 + (h_{\text{trath}} - \hat{h}_{2i})^2] \right\} + \\
& \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \left\{ \beta [\hat{C}_{1i} \log(C_{\text{trath}}) + (1 - \hat{C}_{1i}) \log(1 - C_{\text{trath}})] + (1 - \beta) [\hat{C}_{2i} \log(C_{\text{trath}}) + (1 - \hat{C}_{2i}) \log(1 - C_{\text{trath}})] \right\} + \\
& \lambda_{\text{noobj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} \left\{ \beta [\hat{C}_{1i} \log(C_{\text{trath}}) + (1 - \hat{C}_{1i}) \log(1 - C_{\text{trath}})] + (1 - \beta) [\hat{C}_{2i} \log(C_{\text{trath}}) + (1 - \hat{C}_{2i}) \log(1 - C_{\text{trath}})] \right\} + \\
& \sum_{i=0}^{K \times K} \sum_{c \in \text{class}} I_{ij}^{\text{obj}} \left\{ \alpha [\hat{p}_{1i}(c) \log(p_{\text{trath}}(c)) + (1 - \hat{p}_{1i}(c)) \log(1 - p_{\text{trath}}(c))] + (1 - \alpha) [\hat{p}_{2i}(c) \log(p_{\text{trath}}(c)) + (1 - \hat{p}_{2i}(c)) \log(1 - p_{\text{trath}}(c))] \right\}
\end{aligned}$$

其中 λ_{coord} 、 λ_{noobj} 、 α 、 β 为权重调节参数分别取值为0.5、0.5、0.6、0.5； I_{ij}^{obj} 表示该cell中存在物体为1，否则为0； x_{trath} 、 y_{trath} 、 C_{trath} 、 $p_{\text{trath}}(c)$ 为图像中框的真实值， x_{1i} 、 y_{1i} 、 C_{1i} 、 $p_{1i}(c)$ 为原网络预测框的值， x_{2i} 、 y_{2i} 、 C_{2i} 、 $p_{2i}(c)$ 为细粒度分类网络预测框的值。

一种基于改进YOLOv3的细粒度图像分类方法

技术领域

[0001] 本发明涉及目标检测领域,具体为一种基于改进YOLOv3的细粒度图像分类方法。

背景技术

[0002] 目前基于深度学习的目标检测算法主要分为两类:非端到端检测和端到端检测。非端到端检测算法检测效果好,但是耗时高,一般不能实时检测,典型的网络模型有RCNN、faster-RCNN、mask-RCNN。这类算法先在图像中提取可能含有目标的候选框(region proposal),然后将候选框输入到CNN模型,让CNN判断候选框中是否真的有目标,以及目标的类别是什么。端到端检测算法可以实时检测,精度没有非端到端检测算法高,典型网络模型如yolo。这两种检测算法在通用图像(General/Generic Images)分类任务中都有很好的表现。

[0003] 细粒度图像分类问题是对大类下的子类进行识别。细粒度图像分类任务相对通用图像(General/Generic Images)任务的区别和难点在于其图像所属类别的粒度更为精细。常用的方法有基于定位-分类的方法、基于网络集成的方法、卷积特征的高阶编码方法等。基于定位-分类的方法采用强监督学习,则需要大量的人工来标注图像的关键区域,采用弱监督学习则准确率有待提高;基于网络集成的方法干预的因素较多,不能采用端到端的学习方式;卷积特征的高阶编码方法取得了和使用部位(parts)信息相当或甚至更高的性能,其缺点在于外积导致特征的维度增大为原来的平方。

[0004] YOLOv3是一种端到端的可实时目标检测网络,在Titan X GPU上,它以30FPS的速度处理图像,在COCO test-dev上不同种类的mAP为57.9%。经在TeslaK80 GPU上测试YOLOv3在CUB200-2011数据集上平均识别精度只有84.20%。

发明内容

[0005] 针对现有技术中存在的难题,本发明的目的是提供一种基于改进YOLOv3的细粒度图像分类方法,在YOLOv3原网络基础上,采用卷积特征的高阶编码方法,对原网络中的一些特定层通过外积捕获了特征通道之间成对的相关关系,提取检测目标的局部特征。把细粒度特征分类结果与原网络分类结果融合,修改原网络损失计算公式,在保证检测速度的同时提高识别精度。

[0006] 为达到上述目的,本发明的技术方案为:

[0007] 一种基于改进YOLOv3的细粒度图像分类方法,包括如下步骤:

[0008] 步骤1:把细粒度图像数据集随机分为训练集、验证集、测试集,通过图像预处理将图像缩放至 416×416 ,采用数据增强方法扩充数据集;

[0009] 步骤2:通过Darknet-53网络提取输入图像特征;

[0010] 步骤3:经过三个(1×1 卷积+BatchNormalization+LeakyRelu+ 3×3 卷积+BatchNormalization+LeakyRelu)层后进入 1×1 分类卷积层得到尺寸为 $13 \times 13 \times (4+1+类别) \times 3$ 的输出特征图;通过上采样,重复上述操作得到尺寸为 $26 \times 26 \times (4+1+类别) \times 3$ 的输出特征

图。再次上采样,重复上述操作得到尺寸为 $52 \times 52 \times (4+1+\text{类别}) \times 3$ 的输出特征图;

[0011] 步骤4:在输出特征图尺寸为 $13 \times 13 \times (4+1+\text{类别}) \times 3$ 的网络中抽取第76、78、80层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行 1×1 分类卷积得到尺寸为 $13 \times 13 \times (4+1+\text{类别}) \times 3$ 的细粒度分类特征图;在输出特征图尺寸为 $26 \times 26 \times (4+1+\text{类别}) \times 3$ 的网络中抽取第80、82、84层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行 1×1 分类卷积得到尺寸为 $26 \times 26 \times (4+1+\text{类别}) \times 3$ 的细粒度分类特征图;在输出特征图尺寸为 $56 \times 56 \times (4+1+\text{类别}) \times 3$ 的网络中抽取第84、86、88层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换,拼接后进行 1×1 分类卷积得到尺寸为 $56 \times 56 \times (4+1+\text{类别}) \times 3$ 细粒度分类特征图;

[0012] 步骤5:将原YOLOv3网络得到的三个分类特征图和三个细粒度分类特征图拼接得到 $13 \times 13 \times (4+1+\text{类别}) \times 3 \times 2$, $26 \times 26 \times (4+1+\text{类别}) \times 3 \times 2$, $56 \times 56 \times (4+1+\text{类别}) \times 3 \times 2$ 三个尺寸的分类特征图,修改损失函数,将原特征图和细粒度特征图乘以不同的权重系数。

[0013] 步骤6:训练网络得到权重参数,对细粒度图像分类。

[0014] 优选的,步骤2中Darknet-53特征提取网络,网络共75层,其框架由Convolutional层和Residual残差连接层组成;每个Convolutional层包含卷积、归一化、LeakyReLU三个操作;Residual残差连接层结构中令输入特征向量为 X ,经过两层Convolutional操作后的输出向量为 X' ,再将输入特征向量矩阵与输出特征向量矩阵逐元素相加得到输出。

[0015] 优选的,在每个尺度的YOLOv3原网络金字塔卷积分类层中抽取3个 1×1 Convolutional层的输出作为细粒度特征图。在每个尺度上将得到的三个细粒度特征图相互作外积(matrix outerproduct),经过 $y = \text{Sign}(x) \times \sqrt{|x|+b}$ 函数非线性激活,得到三个大小与输入相同的细粒度特征图,最后将三个特征图拼接成一个特征图。最终得到三个不同尺度的细粒度特征图为 $13 \times 13 \times (1024 \times 3)$ 、 $26 \times 26 \times (512 \times 3)$ 、 $52 \times 52 \times (256 \times 3)$ 。

[0016] 优选的,步骤3原YOLOv3输出特征图进行 1×1 卷积分类得到三个尺度的输出,即 $13 \times 13 \times (4+1+\text{类别}) \times 3$, $26 \times 26 \times (4+1+\text{类别}) \times 3$, $52 \times 52 \times (4+1+\text{类别}) \times 3$;在每个尺度上预测3个不同大小的预测框,每种大小的预测框值包括4个边框预测值(x 、 y 、 w 、 h), (x , y)对应的是预测目标框的几何中心坐标, (w , h)对应的是预测目标框的宽度与高度;1为该像素点的目标置信度值以及预测的每个类别的得分值;步骤4将三个不同尺度的细粒度特征图进行 1×1 卷积分类得到 $13 \times 13 \times (4+1+\text{类别}) \times 3$, $26 \times 26 \times (4+1+\text{类别}) \times 3$, $52 \times 52 \times (4+1+\text{类别}) \times 3$ 三个输出,并与步骤3输出的结果图进行拼接得到 $13 \times 13 \times (4+1+\text{类别}) \times 3 \times 2$, $26 \times 26 \times (4+1+\text{类别}) \times 3 \times 2$, $52 \times 52 \times (4+1+\text{类别}) \times 3 \times 2$ 三个输出。

[0017] 优选的,边框预测公式为:

$$[0018] \quad b_x = \frac{1}{1 + e^{-(\alpha t_{x1} + (1-\alpha)t_{x2})}} + c_x$$

$$[0019] \quad b_y = \frac{1}{1 + e^{-(\alpha t_{y1} + (1-\alpha)t_{y2})}} + c_y$$

$$[0020] \quad b_w = p_w e^{(\alpha t_{w1} + (1-\alpha)t_{w2})}$$

$$[0021] \quad b_h = p_h e^{(\alpha t_{h1} + (1-\alpha)t_{h2})}$$

[0022] 其中 C_x, C_y 是特征图中grid cell的左上角坐标,在yolov3中每个gridcell在featuremap中的宽和高均为1; $t_{x1}, t_{y1}, t_{h1}, t_{w1}$ 分别是原分类网络预测的边框值, $t_{x2}, t_{y2}, t_{h2}, t_{w2}$ 分别为细粒度分类预测边框值, α 为调解权重的参数,这里取 $\alpha=0.6$ 。

[0023] 优选的,置信度计算公式为:

$$[0024] \quad y = \frac{1}{1 + e^{-(\beta d_1 + (1-\beta)d_2)}}$$

[0025] 其中 d_1, d_2 分别为原网络分类和细粒度分类置信度值, β 为调解权重的参数,这里取 $\beta=0.5$ 。

[0026] 优选的,类别得分计算公式为:

$$[0027] \quad y = \frac{1}{1 + e^{-(\gamma p_1 + (1-\gamma)p_2)}}$$

[0028] 其中 p_1, p_2 分别为原网络分类和细粒度分类预测每个类别的值, γ 为调解权重的参数,这里取 $\gamma=0.6$ 。

[0029] 优选的,网络损失函数计算公式为:

$$[0030] \quad \begin{aligned} \text{loss(object)} = & \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_{\text{truth}} \times h_{\text{truth}}) \left\{ \alpha [(\hat{x}_{1i} \log(x_{\text{truth}}) + (1 - \hat{x}_{1i}) \log(1 - x_{\text{truth}}))] + (1 - \alpha) [\hat{x}_{2i} \log(x_{\text{truth}}) + (1 - \hat{x}_{2i}) \log(1 - x_{\text{truth}})] \right\} + \\ & \left\{ \alpha [(\hat{y}_{1i} \log(y_{\text{truth}}) + (1 - \hat{y}_{1i}) \log(1 - y_{\text{truth}}))] + (1 - \alpha) [\hat{y}_{2i} \log(y_{\text{truth}}) + (1 - \hat{y}_{2i}) \log(1 - y_{\text{truth}})] \right\} + \\ & \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_{\text{truth}} \times h_{\text{truth}}) \left\{ \alpha [(w_{\text{truth}} - \hat{w}_{1i})^2 + (h_{\text{truth}} - \hat{h}_{1i})^2] + (1 - \alpha) [(w_{\text{truth}} - \hat{w}_{2i})^2 + (h_{\text{truth}} - \hat{h}_{2i})^2] \right\} + \\ & \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} \left\{ \beta [\hat{C}_{1i} \log(C_{\text{truth}}) + (1 - \hat{C}_{1i}) \log(1 - C_{\text{truth}})] + (1 - \beta) [\hat{C}_{2i} \log(C_{\text{truth}}) + (1 - \hat{C}_{2i}) \log(1 - C_{\text{truth}})] \right\} + \\ & \lambda_{\text{noobj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} \left\{ \beta [\hat{C}_{1i} \log(C_{\text{truth}}) + (1 - \hat{C}_{1i}) \log(1 - C_{\text{truth}})] + (1 - \beta) [\hat{C}_{2i} \log(C_{\text{truth}}) + (1 - \hat{C}_{2i}) \log(1 - C_{\text{truth}})] \right\} + \\ & \sum_{i=0}^{K \times K} \sum_{c \in \text{class}} I_{ij}^{\text{obj}} \left\{ \alpha [\hat{p}_{1i}(c) \log(p_{\text{truth}}(c)) + (1 - \hat{p}_{1i}(c)) \log(1 - p_{\text{truth}}(c))] + (1 - \alpha) [\hat{p}_{2i}(c) \log(p_{\text{truth}}(c)) + (1 - \hat{p}_{2i}(c)) \log(1 - p_{\text{truth}}(c))] \right\} \end{aligned}$$

[0031] 其中 $\lambda_{\text{coord}}, \lambda_{\text{noobj}}, \alpha, \beta$ 为权重调节参数分别取值为0.5、0.5、0.6、0.5。 I_{ij}^{obj} 表示该cell中存在物体为1,否则为0。 $x_{\text{truth}}, y_{\text{truth}}, C_{\text{truth}}, p_{\text{truth}}(c)$ 为图像中框的真实值。 $x_{1i}, y_{1i}, C_{1i}, p_{1i}(c)$ 为原网络预测框的值。 $x_{2i}, y_{2i}, C_{2i}, p_{2i}(c)$ 为细粒度分类网络预测框的值。

[0032] 相对于现有技术,本发明的有益效果为:

[0033] 1、本发明所述的所述的一种基于改进YOLOv3的细粒度图像分类方法,在原网络模型基础上添加细粒度分类层。

[0034] 2、本发明所述的一种基于改进YOLOv3的细粒度图像分类方法,在不增加检测时间的同时,CUB200-2011数据集上的测试结果显示该发明的mAP为95.25%,高于原网络识别精度。

附图说明

[0035] 图1为本发明实例中所述算法的流程框图。

[0036] 图2为YOLOv3原网络结构图。

[0037] 图3为本发明结构图

具体实施方式

[0038] 下面结合附图对本发明的实施方案进行详细阐述,以使本发明的优点和特征能易被本领域的技术人员理解,所述是对本发明的解释而不是限定。

[0039] 本发明一种基于改进YOLOv3的细粒度图像分类方法,如图1所示,具体包括以下步骤:

[0040] 步骤1:数据集选用CUB200-2011,随机选取其中的20个不同的鸟类,共1195张图片,其中训练集959张,验证集60张,测试集176张;

[0041] 步骤2:通过Darknet-53网络提取输入图像特征。作为全卷积网络,darknet53主要由卷积层、BatchNormalization及跨层连接组成,激活函数采用LeakyRelu;

[0042] 步骤3:经过三个(1×1卷积+BatchNormalization+LeakyRelu+3×3卷积+BatchNormalization+LeakyRelu)层后进入1×1分类卷积层得到尺寸为13×13×(4+1+类别)×3的输出特征图;通过上采样,重复上述操作得到尺寸为26×26×(4+1+类别)×3的输出特征图。再次上采样,重复上述操作得到尺寸为52×52×(4+1+类别)×3的输出特征图;

[0043] 步骤4:在输出特征图尺寸为13×13×(4+1+类别)×3的网络中抽取第76、78、80层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行1×1分类卷积得到尺寸为13×13×(4+1+类别)×3的细粒度分类特征图;在输出特征图尺寸为26×26×(4+1+类别)×3的网络中抽取第80、82、84层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换、拼接后进行1×1分类卷积得到尺寸为26×26×(4+1+类别)×3的细粒度分类特征图;在输出特征图尺寸为56×56×(4+1+类别)×3的网络中抽取第84、86、88层的输出作为细粒度特征图,然后将这三个细粒度特征图相互做外积,经过符号平方根变换,拼接后进行1×1分类卷积得到尺寸为56×56×(4+1+类别)×3细粒度分类特征图;

[0044] 步骤5:将得到的三个分类特征图和三个细粒度分类特征图拼接得到13×13×(4+1+类别)×3×2,26×26×(4+1+类别)×3×2,56×56×(4+1+类别)×3×2三个尺寸的分类特征图,修改损失函数,将原特征图和细粒度特征图乘以不同的权重系数。

[0045] 5.1) 改进边框预测公式、置信度公式、类别得分公式,如下:

$$[0046] \quad b_x = \frac{1}{1 + e^{-(\alpha t_{x1} + (1-\alpha)t_{x2})}} + C_x$$

$$[0047] \quad b_y = \frac{1}{1 + e^{-(\alpha t_{y1} + (1-\alpha)t_{y2})}} + C_y$$

$$[0048] \quad b_w = p_w e^{(\alpha t_{w1} + (1-\alpha)t_{w2})}$$

$$[0049] \quad b_h = p_h e^{(\alpha t_{h1} + (1-\alpha)t_{h2})}$$

$$[0050] \quad Con = \frac{1}{1 + e^{-(\beta d_1 + (1-\beta)d_2)}}$$

[0051]
$$P = \frac{1}{1 + e^{-(\gamma P_1 + (1-\gamma) P_2)}}$$

[0052] 其中Cx,Cy是特征图中grid cell的左上角坐标,在yolov3中每个grid cell在feature map中的宽和高均为1。tx1、ty1、th1、tw1分别是原分类网络预测的边框值,tx2、ty2、th2、tw2分别为细粒度分类预测边框值,α为调解权重的参数,这里取α=0.6。d1、d2分别为原网络分类和细粒度分类置信度值,β为调解权重的参数,这里取β=0.5。p1、p2分别为原网络分类和细粒度分类预测每个类别的值,γ为调解权重的参数,这里取γ=0.6。

[0053] 5.2) 修改损失函数,将原特征图和细粒度特征图乘以不同的权重系数,如下:

$$loss(object) = \lambda_{coord} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} (2-w_{truth} \times h_{truth}) \left\{ \alpha [(\hat{x}_{1i} \log(x_{truth}) + (1-\hat{x}_{1i}) \log(1-x_{truth})) + (1-\alpha) [\hat{x}_{2i} \log(x_{truth}) + (1-\hat{x}_{2i}) \log(1-x_{truth})]] \right. \\ \left. + \alpha [(\hat{y}_{1i} \log(y_{truth}) + (1-\hat{y}_{1i}) \log(1-y_{truth})) + (1-\alpha) [\hat{y}_{2i} \log(y_{truth}) + (1-\hat{y}_{2i}) \log(1-y_{truth})]] \right\} + \\ \lambda_{coord} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} (2-w_{truth} \times h_{truth}) \left\{ \alpha [(w_{truth} - \hat{w}_{1i})^2 + (h_{truth} - \hat{h}_{1i})^2] + (1-\alpha) [(w_{truth} - \hat{w}_{2i})^2 + (h_{truth} - \hat{h}_{2i})^2] \right\} + \\ [0054] \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} \left\{ \beta [\hat{C}_{1i} \log(C_{truth}) + (1-\hat{C}_{1i}) \log(1-C_{truth})] + (1-\beta) [\hat{C}_{2i} \log(C_{truth}) + (1-\hat{C}_{2i}) \log(1-C_{truth})] \right\} + \\ \lambda_{noobj} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{noobj} \left\{ \beta [\hat{C}_{1i} \log(C_{truth}) + (1-\hat{C}_{1i}) \log(1-C_{truth})] + (1-\beta) [\hat{C}_{2i} \log(C_{truth}) + (1-\hat{C}_{2i}) \log(1-C_{truth})] \right\} + \\ \sum_{i=0}^{K \times K} \sum_{c \in class} I_{ij}^{obj} \left\{ \alpha [\hat{p}_{1i}(c) \log(p_{truth}(c)) + (1-\hat{p}_{1i}(c)) \log(1-p_{truth}(c))] + (1-\alpha) [\hat{p}_{2i}(c) \log(p_{truth}(c)) + (1-\hat{p}_{2i}(c)) \log(1-p_{truth}(c))] \right\}$$

[0055] 其中λcoord、λnoobj、α、β为权重调节参数分别取值为0.5、0.5、0.6、0.5。Iij^{obj}表示该cell中存在物体为1,否则为0。xtruth、ytruth、Ctruth、ptruth(c)为图像中框的真实值。x1i、y1i、C1i、p1i(c)为原网络预测框的值。x2i、y2i、C2i、p2i(c)为细粒度分类网络预测框的值。

[0056] 步骤6:训练网络得到权重参数,对细粒度图像分类,计算平均正确率mAP。对比结果如表1所示。

[0057] 原网络与本发明在CUB200-2011数据集上检测时间和mAP结果对比

	YOLOv3	改进的 YOLOv3
[0058] 单张图片检测时间 (ms) Tesla K80 GPU	50	53
[0059] 平均正确率 mAP	84.20%	95.25%

[0060] 从上表可以看出,改进后的YOLOv3网络在不显著增加检测时间的同时提高了细粒度图像的检测精度。

[0061] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何不经过创造性劳动想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求书所限定的保护范围为准。

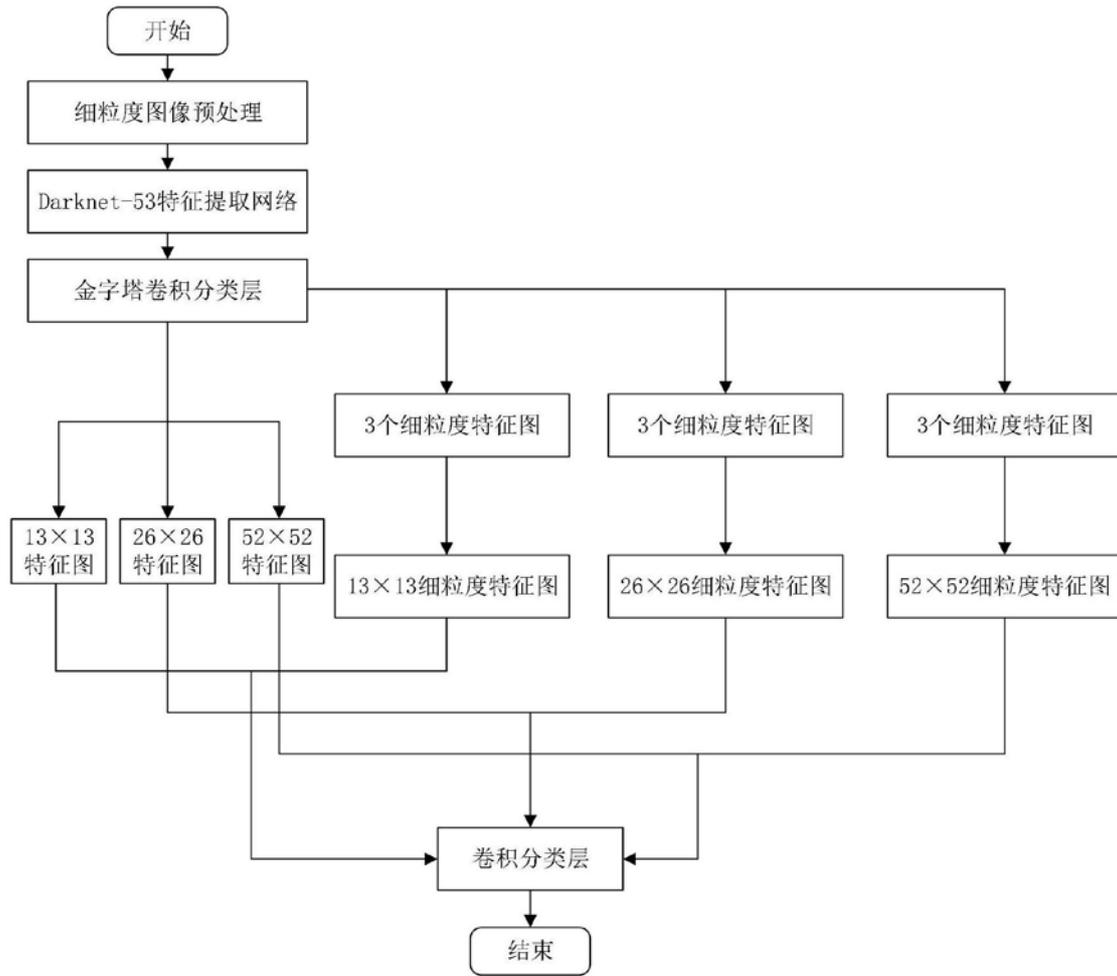


图1

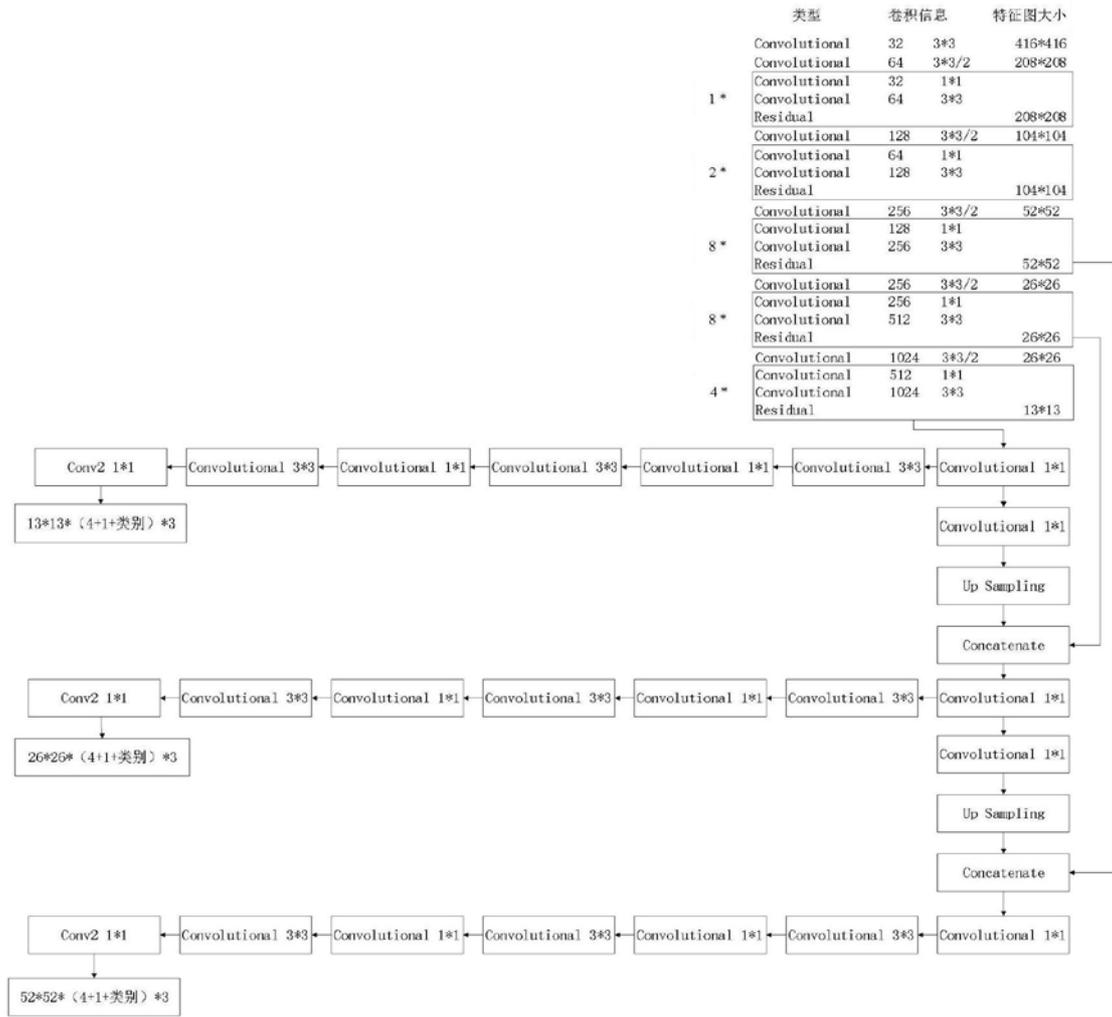


图2

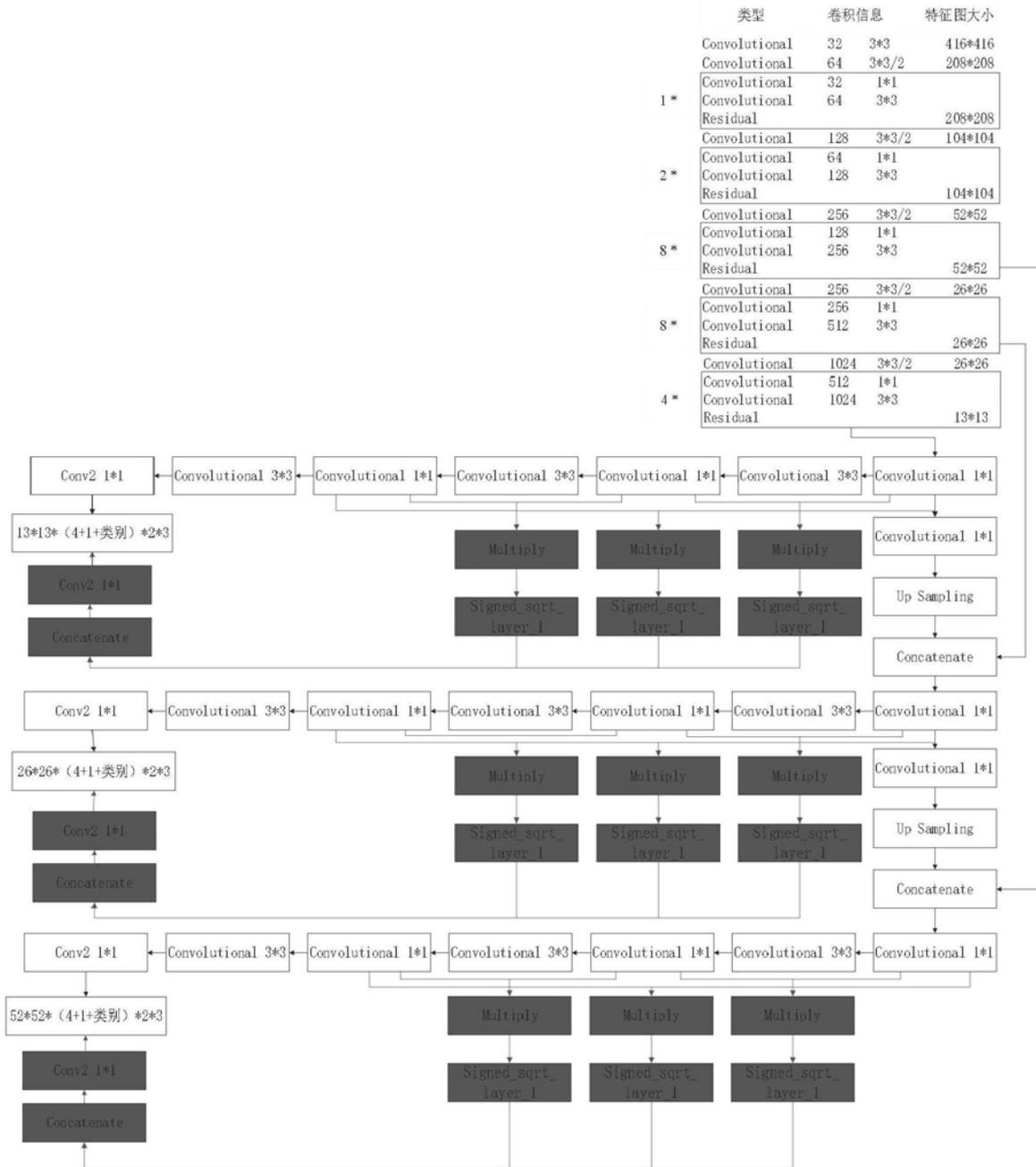


图3