



(12) 发明专利

(10) 授权公告号 CN 109523015 B

(45) 授权公告日 2021.10.22

(21) 申请号 201811328273.3

US 2017124415 A1,2017.05.04

(22) 申请日 2018.11.09

Jianlong Fu 等.Look Closer to See

(65) 同一申请的已公布的文献号

Better: Recurrent Attention Convolutional Neural Network for Fine-grained Image Recognition.《2017 IEEE Conference on Computer Vision and Pattern Recognition》.2017,

申请公布号 CN 109523015 A

(43) 申请公布日 2019.03.26

(73) 专利权人 上海海事大学

地址 201306 上海市浦东新区临港新城海港大道1550号

Miao Kang 等.A modified faster R-CNN based on CFAR algorithm for SAR ship detection.《2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)》.2017,

(72) 发明人 霍煜豪 徐志京

(74) 专利代理机构 上海互顺专利代理事务所

(普通合伙) 31332

代理人 成秋丽

Fan Yang 等.Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers.《2016 IEEE Conference on Computer Vision and Pattern Recognition》.2016, (续)

(51) Int. Cl.

G06N 3/04 (2006.01)

(续)

(56) 对比文件

CN 108416394 A,2018.08.17

CN 106683091 A,2017.05.17

WO 2007120585 A2,2007.10.25

CN 106250812 A,2016.12.21

审查员 姜蝶

权利要求书2页 说明书8页 附图3页

(54) 发明名称

一种神经网络中图像处理方法

(57) 摘要

本发明提供一种神经网络中图像处理方法,方法包括步骤:第一分类网络对所接收到的目标图像进行特征提取;第一定位网络将所得到的第一目标像素个数和第一目标位置发送至第一分类网;第一分类网根据第一目标像素个数确定所选择的第一目标池化层的池化结果;第二分类网络对第一数量个特征图中每一个特征图进行特征提取;第二分类网络确定所选择的第二目标池化层的池化结果,得到第二个尺度层的第二预测标签;确定目标类别标签,以及确定最终的目标定位.应用本发明实施例,通过增加了复数个先验矩形框来描述特征区域,最后通过将所有的特征区域分别分类判定,对每个预测概率进行加权平均强化矩形框在最终判决的影响力。

CN 109523015 B



[接上页]

(51) Int.Cl.

G06K 9/46 (2006.01)

(56) 对比文件

王新宇 等. 基于目标检测及高密度轨迹的动作识别.《复旦学报(自然科学版)》.2016,第55卷(第4期),

1. 一种神经网络中图像处理方法,其特征在于,所述神经网络至少包括第一尺度层和第二尺度层,所述第一尺度层至少包括:第一分类网络、第一定位网络,所述第二尺度层至少包括:第二分类网络、第二定位网络;所述方法包括步骤:

所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络;

所述第一定位网络得到第一目标像素个数和第一目标位置,并将所得到的第一目标像素个数和第一目标位置发送至所述第一分类网络,以及根据所述第一定位网络输出截取出第一数量个特征图作为所述第二尺度层的输入;

所述第一分类网络根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果并反馈至所述第一定位网络,以及得到所述第一个尺度层的第一预测标签;

所述第二分类网络对所述第一数量个特征图中每一个特征图进行特征提取,并将最后一个池化层的输出结果送入所述第二定位网络;

所述第二定位网络获得第二目标像素个数和第二目标位置,并将所得到的第二目标像素个数和第二目标位置发送至所述第二分类网络;

所述第二分类网络根据所述第二目标像素个数确定所选择的第二目标池化层的池化结果并反馈至所述第二定位网络,以及得到所述第二个尺度层的第二预测标签;

根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签,以及确定最终的目标定位;

第一定位网络输出截取出第一数量个特征图的步骤,包括:

获取目标区域的中心点坐标值、任意一个正方形框的长度的二分之一值、宽度的二分之一值,定义比例系数;

规定先验矩形框的面积等于输出的正方形框面积;

得到新的先验矩形框长度的二分之一值、宽度的二分之一值;

根据所得到的新的先验矩形框长度的二分之一值、宽度的二分之一值、目标区域的中心点坐标值,定义先验矩形框的左上角的坐标值和右下角的坐标值;

采用可导的截取函数进行截取,获得截取后的目标区域。

2. 根据权利要求1所述的一种神经网络中图像处理方法,其特征在于,所述神经网络还包括第三尺度层;

所述方法还包括:

所述第二定位网络将其输出的第一数量个特征图作为所述第三尺度层的输入;

所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别,以及确定最终的目标定位的步骤,包括:

根据所述第一预测标签、所述第二预测标签、第三预测标签的融合确定目标类别,以及确定最终的目标定位,其中,所述第三预测标签为所述第三尺度层的预测标签。

3. 根据权利要求1所述的一种神经网络中图像处理方法,其特征在于,所述得到所述第一个尺度层的第一预测标签的步骤,包括:

采用全连接操作后使用softmax函数得到所述第一个尺度层的第一预测标签。

4. 根据权利要求1所述的一种神经网络中图像处理方法,其特征在于,所述得到所述第二个尺度层的第二预测标签的步骤,包括:

将所述第一数量个特征图的预测概率融合成所述第二尺度层的第二预测标签。

5. 根据权利要求1所述的一种神经网络中图像处理方法,其特征在於,所述第一分类网络根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果所采用的公式,具体表达包括:

$$f(I) = \begin{cases} P_3(I) & ,0 < N \leq 64 \\ P_4(I) & ,64 < N \leq 128 \\ P_5(I) & ,128 < N \leq 224 \end{cases}$$

其中,N是第一目标像素个数, P_3 第三池化层, P_4 第四池化层, P_5 是第五池化层,f根据MF-APN网络返回的目标区域像素个数N选择使用最佳的池化输出函数,I为输入图像;

所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络的步骤,包括:

所述第一分类网络对所接收到的目标图像进行特征提取,并将第五池化层的输出结果送入所述第一定位网络。

6. 根据权利要求1-5任一项所述的一种神经网络中图像处理方法,其特征在於,所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签的步骤,包括:

将每所述第一预测标签和所述第二预测标签进行归一化,放入一个全连接层,然后使用softmax函数得到目标类别标签。

一种神经网络中图像处理方法

技术领域

[0001] 本发明涉及运用神经网络进行图像处理的技术领域,特别是涉及一种神经网络中图像处理方法。

背景技术

[0002] 舰船光电图像具有便于获取和高时效性等特点,在军事侦察与预警等领域具有重要的应用价值。但是由于舰船种类繁多,型号复杂多样,所处背景环境复杂多变等原因,快速精准分类舰船类别,及时预警出警成为一个研究难题。

[0003] 针对这个问题,国内外处理方法可以大致分为两个方向。其中一个方向是对图像直接处理,先定位目标位置后分类的识别策略。根据目标边缘信息定位分类,通过分割目标使用支持向量机SVM完成分类。上述研究的算法模型都依赖于先验的图像标注信息训练并且对于细粒度分类效果较差。另一个方向是通过提取舰船尾迹特征进行分类,然而图像中尾迹信息易受天气与环境干扰,影响最终图像的分类而造成漏判,误判。同时,上述方法需要对图像进行复杂的预处理,极易损失图像的信息进而影响最终判别。

[0004] 近年来新兴的深度学习技术作为一种智能信息处理方式,通过训练可以使模型更鲁棒地识别图像中的目标。目前主流的检测方法多采用卷积神经网络(Convolutional Neural Network,简称CNN)提取图像特征,定位目标位置并判别舰船种类。然而采用的神经网络层数较浅,仅通过卷积提取特征,忽视图像特征深层之间联系,则会导致检测效果不理想。

发明内容

[0005] 鉴于以上所述现有技术的缺点,本发明的目的在于提供一种神经网络中图像处理方法,旨在通过增加了复数个先验矩形框来描述特征区域,最后通过将所有的特征区域分别分类判定,对每个预测概率进行加权平均强化矩形框在最终判决的影响力。

[0006] 为实现上述目的及其他相关目的,本发明提供一种神经网络中图像处理方法,所述神经网络至少包括第一尺度层和第二尺度层,所述第一尺度层至少包括:第一分类网络、第一定位网络,所述第二尺度层至少包括:第二分类网络、第二定位网络;所述方法包括步骤:

[0007] 所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络;

[0008] 所述第一定位网络得到第一目标像素个数和第一目标位置,并将所得到的第一目标像素个数和第一目标位置发送至所述第一分类网,以及根据所述第一定位网络输出截取出第一数量个特征图作为所述第二尺度层的输入;

[0009] 所述第一分类网根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果并反馈至所述第一定位网络,以及得到所述第一个尺度层的第一预测标签;

[0010] 所述第二分类网络对所述第一数量个特征图中每一个特征图进行特征提取,并将

最后一个池化层的输出结果送入所述第二定位网络；

[0011] 所述第二定位网络获得第二目标像素个数和第二目标位置,并将所得到的第二目标像素个数和第二目标位置发送至所述第二分类网；

[0012] 所述第二分类网根据所述第二目标像素个数确定所选择的第二目标池化层的池化结果并反馈至所述第二定位网络,以及得到所述第二尺度层的第二预测标签；

[0013] 根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签,以及确定最终的目标定位。

[0014] 本发明的一种实现方式中,所述神经网络还包括第三尺度层；

[0015] 所述方法还包括：

[0016] 所述第二定位网络将其输出的第一数量个特征图作为所述第三尺度层的输入；

[0017] 所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别,以及确定最终的目标定位的步骤,包括：

[0018] 根据所述第一预测标签、所述第二预测标签、第三预测标签的融合确定目标类别,以及确定最终的目标定位,其中,所述第三预测标签为所述第三尺度层的预测标签。

[0019] 本发明的一种实现方式中,所述得到所述第一个尺度层的第一预测标签的步骤,包括：

[0020] 采用全连接操作后使用softmax函数得到所述第一个尺度层的第一预测标。

[0021] 本发明的一种实现方式中,所述得到所述第二个尺度层的第二预测标签的步骤,包括：

[0022] 将所述第一数量个特征图的预测概率融合成所述第二尺度层的第二预测标签。

[0023] 本发明的一种实现方式中,所述第一分类网根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果所采用的公式,具体表达包括：

$$[0024] \quad f(I) = \begin{cases} P_3(I) & ,0 < N \leq 64 \\ P_4(I) & ,64 < N \leq 128 \\ P_5(I) & ,128 < N \leq 224 \end{cases}$$

[0025] 其中,N是第一目标像素个数, P_3 第三池化层, P_4 第四池化层, P_5 是第五池化层,f根据MF-APN网络返回的目标区域像素个数N选择使用最佳的池化输出函数；

[0026] 所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络的步骤,包括：

[0027] 所述第一分类网络对所接收到的目标图像进行特征提取,并将第五池化层的输出结果送入所述第一定位网络。

[0028] 本发明的一种实现方式中,第一定位网络输出截取出第一数量个特征图的步骤,包括：

[0029] 获取目标区域的中心点坐标值、任意一个正方形框的长度的二分之一值、宽度的二分之一值,定义比例系数；

[0030] 规定先验矩形框的面积等于输出的正方形框面积；

[0031] 得到新的先验矩形框长度的二分之一值、宽度的二分之一值；

[0032] 根据所得到的新的先验矩形框长度的二分之一值、宽度的二分之一值、目标区域的中心点坐标值,定义先验矩形框的左上角的坐标值和右下角的坐标值；

[0033] 采用可导的截取函数进行截取,获得截取后的目标区域。

[0034] 本发明的一种实现方式中,所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签的步骤,包括:

[0035] 将每所述第一预测标签和所述第二预测标签进行归一化,放入一个全连接层,然后使用softmax函数得到目标类别标签。

[0036] 如上所述,本发明实施例提供一种神经网络中图像处理方法,增加了复数个先验矩形框来描述特征区域,最后通过将所有的特征区域分别分类判定,对每个预测概率进行加权平均强化矩形框在最终判决的影响力。

附图说明

[0037] 图1是本发明实施例的一种神经网络中图像处理方法的一种流程示意图。

[0038] 图2是本发明实施例的一种神经网络中图像处理方法的第一种实施例示意图。

[0039] 图3是本发明实施例的一种神经网络中图像处理方法的第二种实施例示意图。

[0040] 图4是本发明实施例的一种神经网络中图像处理方法的第三种实施例示意图。

具体实施方式

[0041] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。

[0042] 请参阅图1-4。需要说明的是,本实施例中所提供的图示仅以示意方式说明本发明的基本构想,遂图式中仅显示与本发明中有关的组件而非按照实际实施时的组件数目、形状及尺寸绘制,其实际实施时各组件的型态、数量及比例可为一种随意的改变,且其组件布局型态也可能更为复杂。

[0043] 针对光电图像中舰船分类检测困难的问题,提出了一种基于改进循环注意卷积神经网络(RA-CNN)的舰船目标检测方法。该方法中的VGG19网络采用多个卷积层提取图像特征,注意建议区域网络(APN)通过全连接层的输出定位特征区域,然后采用基于尺度池化(SDP)算法选择VGG19中合适的卷积层输出进行类别判定,最后引入多个先验矩形框(MF)描述特征区域,交叉训练VGG19与APN网络,加速损失收敛提高模型精度。

[0044] 如图1所示,本发明实施例提供一种神经网络中图像处理方法,所述神经网络至少包括第一尺度层和第二尺度层,所述第一尺度层至少包括:第一分类网络、第一定位网络,所述第二尺度层至少包括:第二分类网络、第二定位网络;所述方法包括步骤:

[0045] S101,所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络。

[0046] 需要说明的是,当输入图像I时,图像首先会先经过第一分类网络提取特征,具体的,提取的是特征向量。

[0047] 图2中,10表示原始的输入图像I,11、12、13、14、15分别表示多个卷积层组成的对应卷积块,21表示第一池化层 P_1 、22表示第二池化层 P_2 ,23表示第三池化层 P_3 ,24表示第四池化层 P_4 ,25表示第五池化层 P_5 ,可以理解的是,21、22、23、24、25分别表示对应卷积块中的一

个卷积层。31表示全连接层。示例性的,如图2中所示,采用的是第三池化层 P_3 输出结果,而实际应用中,输出的结果可以是第一池化层 P_1 、第二池化层 P_2 、第三池化层 P_3 、第四池化层 P_4 、第五池化层 P_5 中的任意一个池化层,而实际应用中根据本领域技术人员的经验,第一池化层 P_1 、第二池化层 P_2 、第三池化层 P_3 ,效果较差,往往不做使用。

[0048] 如图2所示,第一分类网络为VGG-SDP(v1),第一定位网络为MF-APN(m1),其中,第一分类网络VGG-SDP(V1)的网络结构如图3所示。

[0049] 如图2可以获得 P_5 是最后一个池化层,将其输出结果发送至第一定位网络。示例性的,第一分类网络VGG-SDP网络将池化结果 P_5 反馈至第一定位网络MF-APN网络,第一定位网络为MF-APN网络。

[0050] S102,所述第一定位网络得到第一目标像素个数和第一目标位置,并将所得到的第一目标像素个数和第一目标位置发送至所述第一分类网,以及根据所述第一定位网络输出截取出第一数量个特征图作为所述第二尺度层的输入。

[0051] 可以理解的是,针对第一分类网络得到的特征区域,第一定位网络MF-APN网络可以计算出特征区域的大小 N ,以及对应的目标位置,并将计算得到的第一目标像素个数和第一目标位置返回至第一分类网络,如图2所示,将计算出来的第一目标像素个数通过第一定位网络MF-APN返回至第一分类网络中。

[0052] 需要说明的是,第一定位网络MF-APN通过使用不同的先验矩形框来框取目标,最后将多个特征区域分类后通过加权平均决策的方式,强化矩形框影响力,弱化原有默认正方形框的影响力,使得网络可以更鲁棒的识别定位目标。

[0053] 具体的,本发明实施例中,本发明的一种实现方式中,第一定位网络输出截取出第一数量个特征图的步骤,包括:获取目标区域的中心点坐标值、任意一个正方形框的长度的二分之一值、宽度的二分之一值,定义比例系数;规定先验矩形框的面积等于输出的正方形框面积;得到新的先验矩形框长度的二分之一值、宽度的二分之一值;根据所得到的新的先验矩形框长度的二分之一值、宽度的二分之一值、目标区域的中心点坐标值,定义先验矩形框的左上角的坐标值和右下角的坐标值;采用可导的截取函数进行截取,获得截取后的目标区域。

[0054] 假设APN网络输出 t_x, t_y 代表目标区域中心点的坐标值, t_1 为正方形框边长的一半, N 为正方形框中的像素个数即目标面积。 W_i, H_i 分别代表第 i 个先验矩形框长宽的一半,定义比例系数 k_i 代表第 i 个矩形框的长宽比值。则有以下关系:

$$[0055] \quad \begin{cases} N = (2t_1)^2 = 4t_1^2 \\ k_i = \frac{W_i}{H_i} \end{cases} \quad (3)$$

[0056] 规定先验矩形框的面积等于输出的正方形框面积,则有:

$$[0057] \quad N = 2W_i \times 2H_i = 4k_i H_i^2 \quad (4)$$

[0058] 将(3)式带入(4)式得到新的 W_i, H_i 表达式:

$$[0059] \quad \begin{cases} W_i = \text{int}\left(\sqrt{k_i} \times t_1\right) \\ H_i = \text{int}\left(\sqrt{\frac{1}{k_i}} \times t_1\right) \end{cases} \quad (5)$$

[0060] (5)式中 $\text{int}(\cdot)$ 函数表示向下取整。使用先验矩形框的左上角和右下角两个顶点来表示矩形框。定义 ul 代表左上角, br 代表右下角,则两点坐标为:

$$[0061] \quad \begin{cases} t_{i,x(ul)} = t_x - W_i & , t_{i,y(ul)} = t_y - H_i \\ t_{i,x(br)} = t_x + W_i & , t_{i,y(br)} = t_y + H_i \end{cases} \quad (6)$$

[0062] 考虑到神经网络反向传播要求可导,不能使用普通的截取方法,所以设计一个可导的截取函数 $M(\cdot)$:

$$[0063] \quad M(\bullet) = \begin{bmatrix} h(x - t_{i,x(ul)}) - h(x - t_{i,x(br)}) \\ \times [h(y - t_{i,y(ul)}) - h(y - t_{i,y(br)})] \end{bmatrix} \quad (7)$$

[0064] $h(\cdot)$ 代表sigmoid函数,公式如下:

$$[0065] \quad h(x) = \frac{1}{1 + e^{-kx}} \quad (8)$$

[0066] 当 k 足够大时,只有在特征区域内的点通过截取函数时 $M(\cdot)$ 的值才为1,在本方案中设置 $k=10$ 。最终截取的目标区域 M_i 可以表示成以下形式:

$$[0067] \quad M_i = I \otimes M(t_x, t_y, t_l, k_i) \quad (9)$$

[0068] 其中, \otimes 运算表示元素点乘。

[0069] 接下来仍使用双线性插值的办法放大目标区域得到下一个尺度的输入。如果让第一个尺度后的每个尺度都选取复数个先验矩形框,最终特征区域数量会成乘性增长,考虑到计算消耗本文仅在第二个尺度层提取 i 个矩形框。此外,新尺度中的 t_1 不能小于上一尺度的 $1/3$,防止特征区域过小不能有效包含特征部分。具体的,第一定位网络MF-APN输出示意图如图4所示。

[0070] S103,所述第一分类网根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果并反馈至所述第一定位网络,以及得到所述第一个尺度层的第一预测标签。

[0071] 本发明的一种实现方式中,所述第一分类网根据所述第一目标像素个数确定所选择的第一目标池化层的池化结果所采用的公式,具体表达包括:

$$[0072] \quad f(I) = \begin{cases} P_3(I) & , 0 < N \leq 64 \\ P_4(I) & , 64 < N \leq 128 \\ P_5(I) & , 128 < N \leq 224 \end{cases}$$

[0073] 其中, N 是第一目标像素个数, P_3 第三池化层, P_4 第四池化层, P_5 是第五池化层, f 根据MF-APN网络返回的目标区域像素个数 N 选择使用最佳的池化输出函数;

[0074] 由于上述最后一个池化层是池化层 P_5 ,所以所述第一分类网络对所接收到的目标图像进行特征提取,并将最后一个池化层的输出结果送入所述第一定位网络的步骤,包括:所述第一分类网络对所接收到的目标图像进行特征提取,并将第五池化层的输出结果送入所述第一定位网络。

[0075] 示例性的,如图3所示,当输入图像 I 时,图像首先会先经过分类网络提取特征。随后MF-APN网络计算出特征区域的大小 N ,VGG-SDP网络将根据 N 选择使用后三个池化层中最佳的池化结果代表图像 I 进行之后的类别判断。选择池化结果的准则如下:

$$[0076] \quad f(I) = \begin{cases} P_3(I) & , 0 < N \leq 64 \\ P_4(I) & , 64 < N \leq 128 \\ P_5(I) & , 128 < N \leq 224 \end{cases}$$

[0077] $Y(I) = F[f(I)]$

[0078] 其中, f 根据MF-APN网络返回的目标区域像素个数 N 选择使用最佳的池化输出。 F 表示最后的全连接与softmax操作。当 N 过大时, 应当选取最终 P_5 的输出, 能更好的描述大目标的特征, 而当 N 较小时, 应选择含有更多信息的 P_3 。然后进行全连接操作后使用softmax函数得到第一个尺度层的预测标签 $Y^{(1)}$ 。

[0079] S104, 所述第二分类网络对所述第一数量个特征图中每一个特征图进行特征提取, 并将最后一个池化层的输出结果送入所述第二定位网络。

[0080] 如图2所示, 第二分类网络为VGG-SDP (v2), 第二定位网络为APN (m2), 其中, 第二分类网络为VGG-SDP (v2) 与第一分类网络VGG-SDP (V1) 的网络结构相同, 如图3所示。

[0081] 将第一个尺度层输出的第一数量个特征图, 例如, 为 i 个特征图, 并将该 i 个特征图输入第二个尺度层的第二分类网络为VGG-SDP (v2) 进行特征提取, 由于第二个尺度层中的定位网络APN (m2) 仅使用普通的APN网络, 因此仍生成 i 个特征图, 当包含第三个尺度层时, 该 i 个特征图进入第三个尺度层。

[0082] S105, 所述第二定位网络获得第二目标像素个数和第二目标位置, 并将所得到的第二目标像素个数和第二目标位置发送至所述第二分类网。

[0083] 该过程与步骤S102的处理方式相同, 本发明实施例在此不做赘述。

[0084] S106, 所述第二分类网根据所述第二目标像素个数确定所选择的第二目标池化层的池化结果并反馈至所述第二定位网络, 以及得到所述第二个尺度层的第二预测标签。

[0085] 该过程与步骤S103的处理方式相同, 本发明实施例在此不做赘述。

[0086] 本发明的一种实现方式中, 所述得到所述第二个尺度层的第二预测标签的步骤, 包括: 将所述第一数量个特征图的预测概率融合成所述第二尺度层的第二预测标签。具体的, 可以通过将 i 个特征图的预测概率融合成第二个尺度层的预测标签 $Y^{(2)}$ 。

[0087] S107, 根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签, 以及确定最终的目标定位。

[0088] 本发明的一种实现方式中, 所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别标签的步骤, 包括: 将每所述第一预测标签和所述第二预测标签进行归一化, 放入一个全连接层, 然后使用softmax函数得到目标类别标签。

[0089] 本发明的一种实现方式中, 所述神经网络还包括第三尺度层; 所述方法还包括: 所述第二定位网络将其输出的第一数量个特征图作为所述第三尺度层的输入; 所述根据所述第一预测标签和所述第二预测标签的融合确定目标类别, 以及确定最终的目标定位的步骤, 包括: 根据所述第一预测标签、所述第二预测标签、第三预测标签的融合确定目标类别, 以及确定最终的目标定位, 其中, 所述第三预测标签为所述第三尺度层的预测标签, 最终将 i 个特征图的预测概率融合成第三个尺度层的预测标签 $Y^{(3)}$ 。

[0090] 示例性的, 最终图像 I 的分类结果是由3个尺度的VGG-SDP网络的预测标签融合而成。首先将每个预测标签 $Y^{(s)}$ 进行归一化后, 放入一个全连接层, 然后使用softmax函数得到最终的预测分类标签。

[0091] 如图2所示,还包括第三尺度层中的第三分类网络为VGG-SDP (v3),第三定位网络为APN (m3),其中,第二分类网络为VGG-SDP (v2)与第三分类网络VGG-SDP (V3)的网络结构相同,如图3所示。

[0092] 示例性的,最终目标定位为第一个尺度层中的正方形特征框,目标类别为三个尺度层的预测标签的融合。

[0093] 针对网络损失部分,如图2所示,采用本发明实施例改进后的RA-CNN整体流程图。图中 p_t 代表真实类别的预测概率; L_{inner} 代表每一个尺度的分类损失,是真实类别标签 Y_{truth} 和预测类别标签 $Y^{(s)}$ 进行交叉熵运算结果; L_{scale} 代表相邻尺度之间的损失。

[0094] 改进后的RA-CNN网络损失函数仍分为尺度内分类损失和尺度间损失两个部分组成,因此网络损失定义如下:

$$[0095] \quad L = \sum_{s=1}^3 \{L_{inner}(Y^{(s)}, Y_{truth})\} + \sum_{s=1}^2 \{L_{scale}(p_t^{(s)}, p_t^{(s+1)})\} \quad (10)$$

[0096] 其中s代表尺度个数, L_{inner} 代表每一个尺度的分类损失,公式如下:

$$[0097] \quad L_{inner}(Y^{(s)}, Y_{truth}) = - \sum \{Y_{truth} \times \log[Y^{(s)}] + (1 - Y_{truth}) \times \log[1 - Y^{(s)}]\} \quad (11)$$

[0098] L_{scale} 代表相邻尺度之间的损失,其计算公式为:

$$[0099] \quad L_{scale}(p_t^{(s)}, p_t^{(s+1)}) = \max\{0, p_t^{(s)} - p_t^{(s+1)} + 0.05\} \quad (12)$$

[0100] 通过取最大值的方法,要求网络更新当前尺度真实类别概率 $p_t^{(s+1)}$ 要比上一尺度真实类别概率 $p_t^{(s)}$ 更小的情况,促使网络在更精细的尺度上预测概率更高。只有当 $p_t^{(s)} - p_t^{(s+1)} + 0.05 > 0$ 时,尺度间损失才会更新。其中附加0.05是为了防止比较双方均为0而导致损失停滞不更新。

[0101] 由于在第二、三尺度有i个特征矩形框,最终的预测类别标签 $Y^{(s)}$ 是这i个特征矩形框的预测概率的加权平均。对于 $Y^{(s)}$ 中第j个类的预测概率 $p_j^{(s)}$,其计算公式为:

$$[0102] \quad p_j^{(s)} = \frac{1}{M} \sum_{i=1}^M a_i \times p_{j,i}^{(s)} \quad (13)$$

[0103] M代表矩形框个数, a_i 代表第i个矩形框的权值且有 $\sum_{i=1}^M a_i = 1$ 。在本文中,设置 $i=3$ 个矩形框,其比例系数 k_i 分别为2,1,0.5,相应的权值 a_i 为0.4,0.2,0.4。通过这样的设置,强化矩形框在预测中的影响,弱化正方形框的影响。

[0104] 此外,本发明实施例还提供了一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现所述神经网络中图像处理方法的步骤。

[0105] 以及,提供了一种终端,包括处理器存储器,存储器存储有程序指令,处理器运行程序指令实现所述神经网络中图像处理方法的步骤。

[0106] 因此,应用本发明提供的实施例,该算法是利用了原RA-CNN网络中每一尺度层都会产生特征区域的特点,通过特征区域像素个数衡量目标区域的大小并融合SDP算法优化分类性能。根据像素个数,选择合适的卷积块输出进行分类判别。根据原有的正方形特征框,增加了复数个先验矩形框来描述特征区域,最后通过将所有的特征区域分别分类判定,

对每个预测概率进行加权平均强化矩形框在最终判决的影响力。

[0107] 上述实施例仅例示性说明本发明的原理及其功效,而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下,对上述实施例进行修饰或改变。因此,举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变,仍应由本发明的权利要求所涵盖。

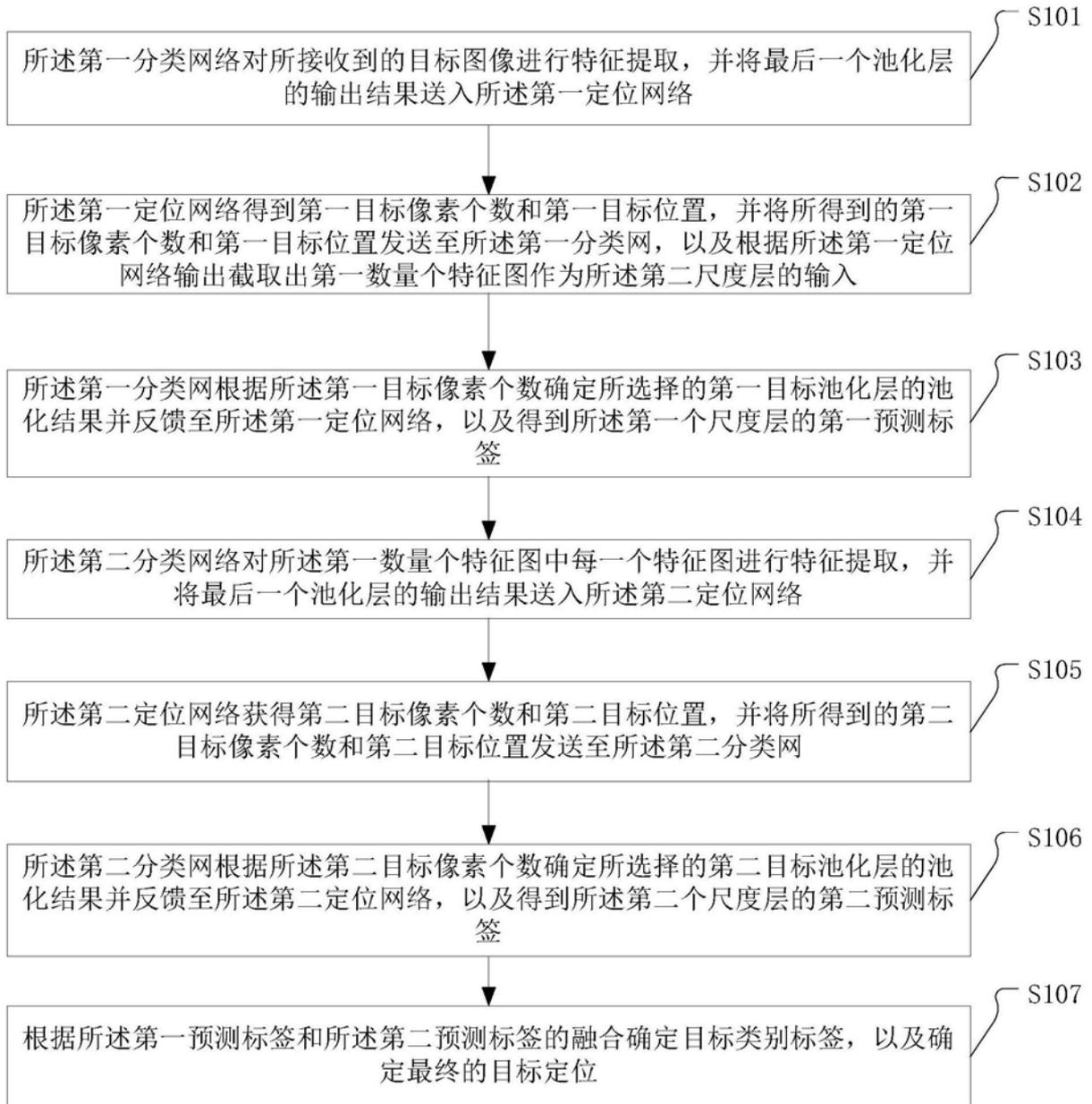


图1

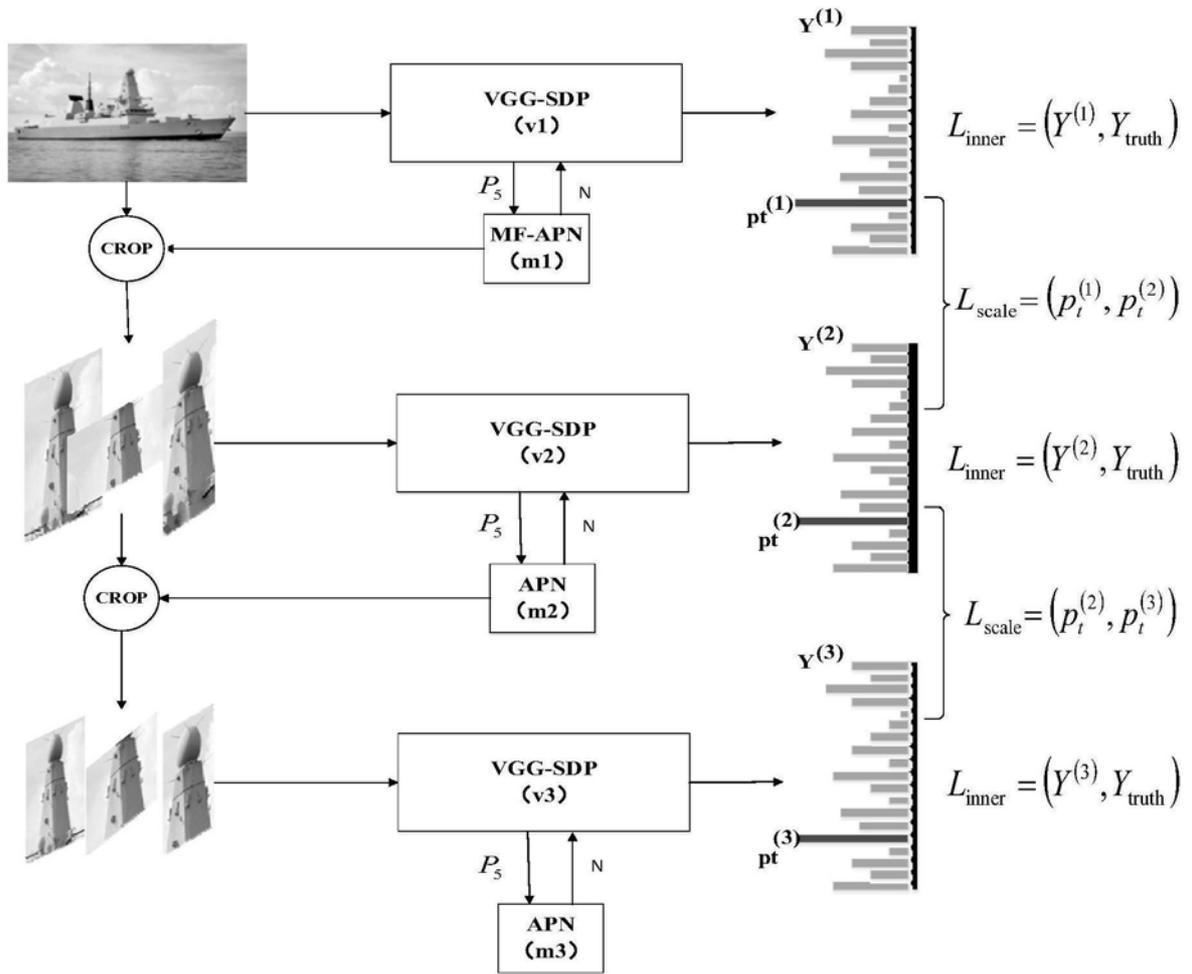


图2

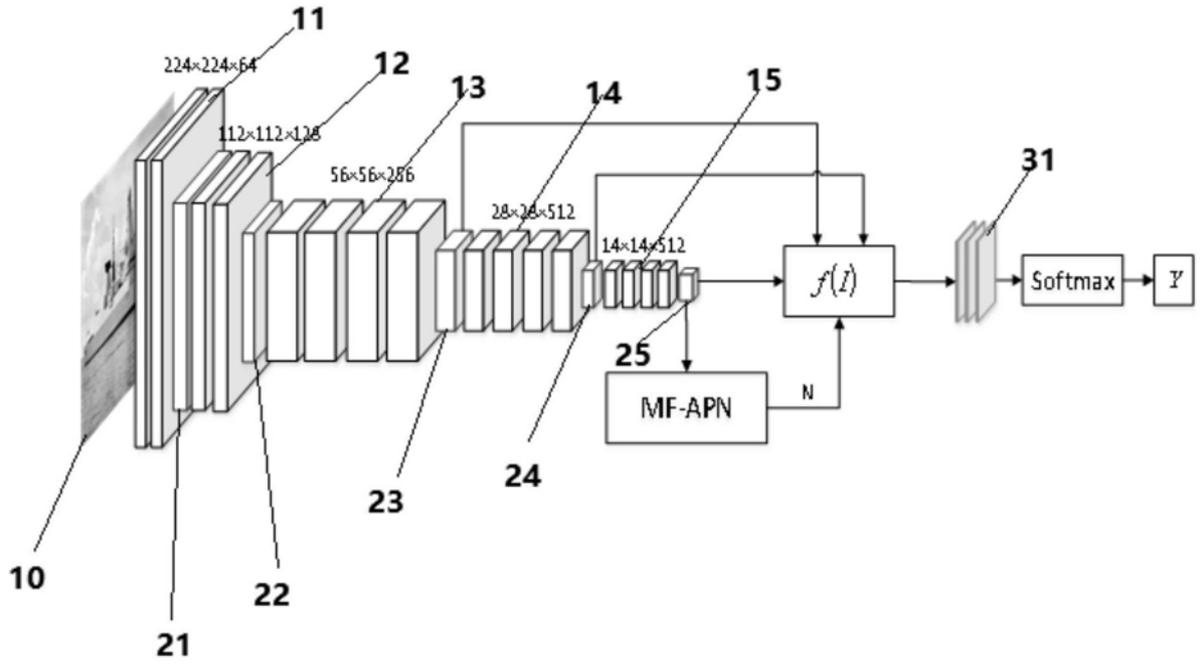


图3

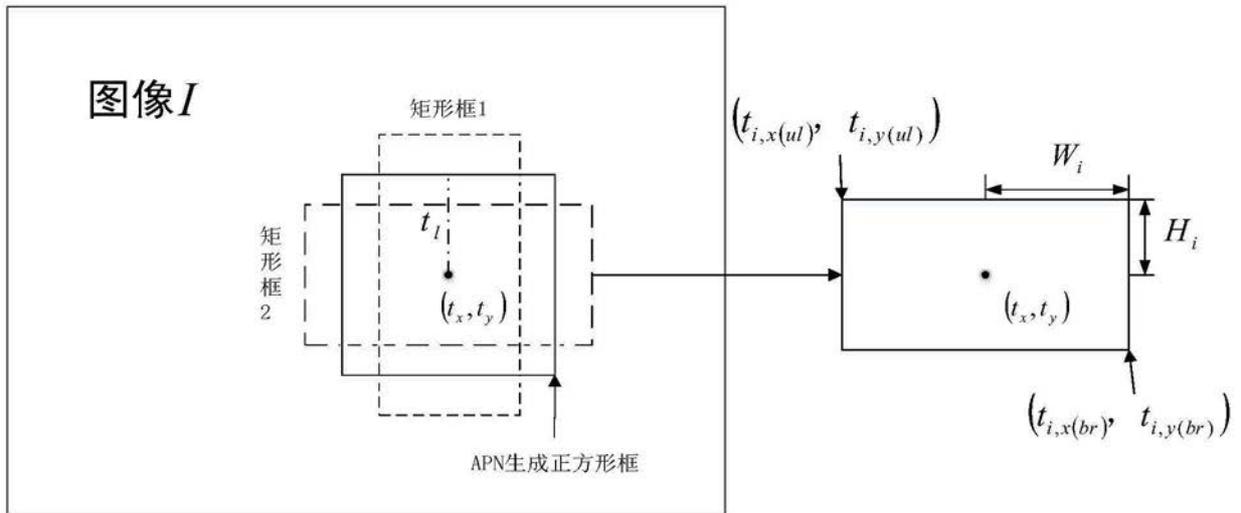


图4