文章编号：2095-2163(2021)08-0023-08 中图分类号：TP391.41 文献标志码：A

基于轻量化 SSD 的菜品识别

姚华莹，彭亚雄，陆安江
（贵州大学 大数据与信息工程学院，贵阳 550025）

摘 要：为了能够在移动设备等计算力弱的平台部署菜品识别系统，帮助人们了解菜品信息，对传统目标检测模型 SSD 做轻量化改进，提高了检测准确率和检测速度。首先使用 MobileNetV2 代替 SSD 模型的 VGG-16，减少模型体积，提升运行速度；使用注意力机制和混洗通道算法，设计新的注意力逆残差块，增强特征提取能力；优化 IOU 计算方式，对回归定位损失函数做改变，加快模型的收敛；最后在自建的中餐菜品数据集 Chinesefood 上进行训练。实验表明，本文提出的 Att_Mobilenetv2_SSDLite 轻量型目标检测模型相比 SSD 和其它目标检测模型效果更佳。

关键词：目标检测；MobileNetV2；SSD；注意力机制；轻量化神经网络

Dishes detection based on lightweight SSD

YAO Huaying, PENG Yaxiong, LU Anjiang
(College of Big Data and Information Engineering, Guizhou University, Guizhou 550025, China)

【Abstract】 In order to be able to deploy a dishes detection system on devices with limited computing power such as mobile devices to help people know the dish information, this paper makes lightweight to the traditional object detection model SSD, which improves the detection accuracy and detection speed. First, use MobileNetV2 to replace the VGG-16 of the SSD model to reduce the size of the model and increase the speed. Then use the attention mechanism and shuffle channel algorithm to design a new attention Inverse Residual block to enhance the feature extraction ability; optimize the IOU calculation method, change the regression positioning loss function, and accelerate the convergence of the model. Finally, trained on the self-built Chinesefood dataset. Experiments show that comprehensive the detection speed, model size and accuracy and other evaluation indicators, the Att_Mobilenetv2_SSDLite lightweight object detection model proposed in this paper is better than SSD and other object detection models.

【Key words】 object detection; MobileNetV2; SSD; attention mechanism; lightweight neural network

0 引 言

近年来，人工智能被研究人员广泛关注并设计研究出大量的人工智能产品[1]。深度学习中的图像识别技术取得的丰硕成果让人们的生活方式更加智能化、高效化。菜品智能识别可以部署在智能餐厅，用于菜品价格自动结算，安装在移动设备实时检测当前食物的卡路里等详细信息等[2-4]，目标检测算法可以实现这一效果。卷积神经网络通过输入的图像数据进行训练，像人类的大脑一样学习图像特征[5]，准确的分类出菜品。目标检测算法分为双阶段检测和单阶段检测。其中，双阶段检测算法中具有代表性的是 R-CNN[6] 和 Faster-R-CNN[7] 算法，这类双阶段算法检测精度高，但检测速度较慢。另一类单阶段检测算法 YOLO[8]（You Only Look Once），是继 R-CNN 之后，为解决目标检测速度问题而提出的另一个框架，在定位边界框的同时获取类别概率，牺牲了一部分检测精度以达到更快的检测速度；由 Liu 在 ECCV 上提出的多尺度单发射击检测算法[9]（Single Shot MultiBox Detector，SSD）在不同尺度的特征图中多步提取特征。与 R-CNN 相比，SSD 速度更快；与 YOLO 相比，SSD 在检测精度 mAP 上有更好的性能，在不降低检测精度的同时保证了检测速度。由于以上目标检测模型包含大量的卷积计算，有大量参数，仅能运行在高性能图像处理器上，不利于移植到数据处理弱的平台。

为此，本文基于优秀的轻量型神经网络 MobileNetV2[10] 对 SSD 网络进行改进，构建轻量型的菜品识别模型。使用注意力机制[11] 和混洗通道[12] 构建注意力逆残差结构提高检测准确率。设计回归定位损失函数，加快模型收敛速度。根据菜品数据集特点设置回归预测层，提升模型检测速度。通过使用数据集 Chinesefood 对模型训练，验证了本文提出的 Att_

基金项目：贵州省科技成果转化项目([2017]4856)。
作者简介：姚华莹(1997-)女，硕士研究生，主要研究方向：深度学习和图像处理；彭亚雄(1963-)男，学士，副教授，主要研究方向：通信系统；陆安江 (1978-) 男，博士，副教授，主要研究方向：嵌入式系统与集成技术、物联网安全、微传感技术。
通讯作者：彭亚雄 Email：515154900@qq.com
收稿日期：2021-04-29

Mobilenetv2_SSDLite 目标检测网络检测效果更佳。

1 SSD 目标检测模型原理

单阶段检测是基于回归的算法(例如 YOLO、SSD 等),把提取候选区域框和特征提取融合在同一网络中,同时实现候选区域选择和分类。检测网络由基础网络层和回归预测层两部分组成。基础网络层用于提取输入图像特征,基础网络很大程度上能影响检测的精度、速度以及检测网络体积;回归预测层用于生成预测框及分类目标。SSD 采用基于金字塔特征层的检测方式,在不同尺度的特征图上进行位置回归和分类。如图 1 所示。SSD 模型结构的前端是用于提取图像特征的基础网络层,后端是用于回归预测的附加层。

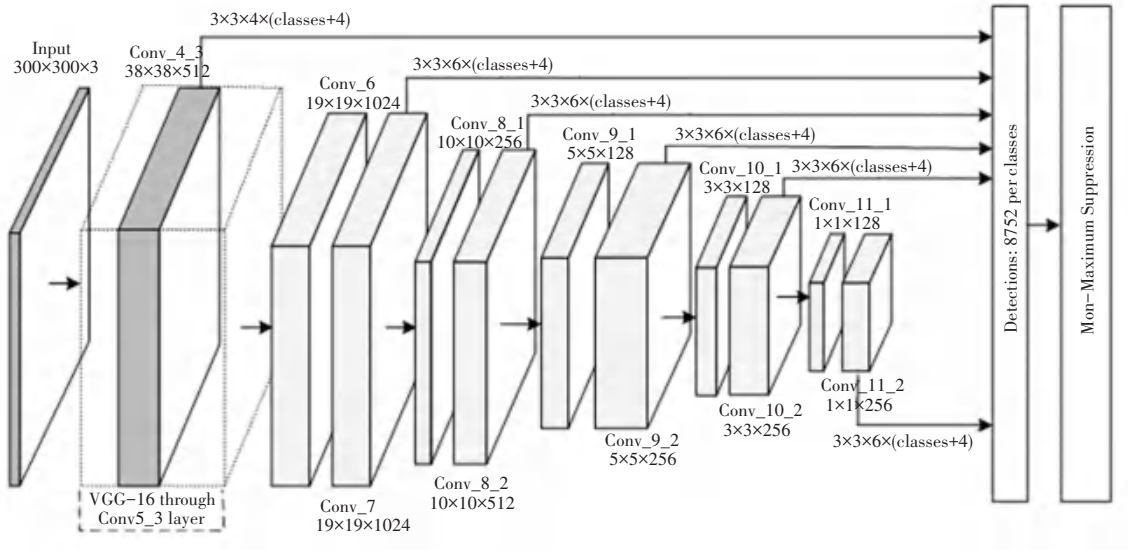


图 1 SSD 网络结构

Fig. 1 Network architecture of SSD

基础网络层使用了 VGG-16[13] 网络的前 5 层卷积层,并将 VGG-16 网络中原有的全连接层 FC6 和 FC7 改为卷积层 Conv6 和 Conv7,提取基本特征;在后端附加 4 个尺度大小不同的卷积层,用于高级特征提取。表 1 为 SSD 网络参数,整个网络包含 6 个特征预测层,前端浅层特征图分辨率高,用于预测小目标;后端深层特征图分辨率低,用来预测较大的目标。

表 1 SSD 网络参数

Tab. 1 SSD network parameters

Conv	Channel	repeat	Stride	Output	Feature
Conv2d_1	64	2	2	150×150	-
Conv2d_2	128	3	1	150×150	-
Conv2d_3	128	3	2	75×75	-
Conv2d_4	512	3	2	38×38	-
Conv2d_5	512	3	1	38×38	4
FC6	1 024	1	1	19×19	-
FC7	1 024	1	1	19×19	6
Conv2d_8	512	1	2	10×10	6
Conv2d_9	256	1	2	5×5	6
Conv2d_10	256	1	1	3×3	4
Conv2d_11	256	1	1	1×1	4

SSD 网络在基础网络中使用了 VGG-16,而 VGG-16 使用了大量的传统卷积,导致模型体积庞大,参数量多,这些弊端导致其难以部署在计算力和存储空间受限的设备中。为了解决这一问题,本文基于 MobileNetV2 轻量型卷积网络替换 VGG-16 网络,对 SSD 进行轻量化改进,针对菜品目标检测的特点,修改 SSD 网络层结构,获得针对菜品识别的目标检测网络 Att_Mobilenetv2_SSDLite。

2 MobileNetV2 模型

MobileNetV2 网络是 Google 团队对 MobileNetV1 的进一步调整,沿用了深度可分离卷积,减少卷积计算量,同时借鉴 ResNet 中的残差结构,设计了逆残差结构,提高网络对低维空间的特征提取能力[14]。

2.1 深度可分离卷积

深度可分离卷积[15](Depthwise sparable convolution),将传统卷积分为一个深度卷积和一个点卷积。首先进行深度卷积,即对每个输入的通道,分别用单个卷积核进行相对应的卷积计算后,用 1×1 的卷积核对深度卷积结果进行线性组合,构建新的特征,这个过程为点卷积。如果不考虑偏置参数,深度

分离后的卷积参数运算量为：

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F, \quad (1)$$

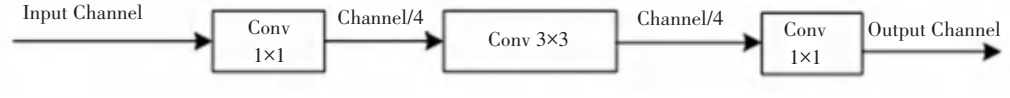
标准卷积计算量为：

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F. \quad (2)$$

其中：$D_K \cdot D_K$ 为卷积核尺寸；$D_F \cdot D_F$ 为输入图像尺寸；M 和 N 分别是输入通道数量和输出通道数量。传统卷积的计算量是深度可分离卷积的 $1/M + 1/K^2$，当卷积核大小为 3×3 时，计算量相比传统卷积减少了 9 倍多，能明显提升运算速度及检测效率。

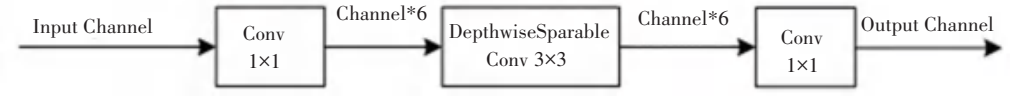
2.2　逆残差结构

残差结构是先对输入图像降维、卷积、再升维，如图 2(a)所示。文献[16]中先使用 1×1 的卷积，将输入通道压缩至原来的 1/4，然后再用 3×3 的卷积进行计算，最后使用 1×1 的卷积对通道数目还原。残差结构通过旁干分支，将前端特征与通过一系列卷积处理后的特征进行融合，可以提高整个网络的特征提取能力。如图 2(b)所示，MobileNetV2 的逆残差结构与残差结构相反，对输入图像升维、通过深度可分离卷积、降维。先使用了 1×1 卷积，将输入通道数目扩大至原来的 6 倍，再经过深度可分离卷积处理，最后通过 1×1 卷积恢复通道数目。对输入通道数目扩展是一个升维的过程，这样做的好处是，可以提取低维特征图上的有效特征，使用了深度可分离卷积后计算量也不会增加[17]。与 ResNet 的残差结构一样，最后会将主干分支与旁干分支特征融合。



(a) ResNet 中使用的残差结构

(a) Residual structure used in ResNet



(b) MobileNetV2 中使用的逆残差结构

(b) Inverted residual structure used in MobileNetV2

图 2　残差结构与逆残差结构对比

Fig. 2　Comparison of residual structure and inverse residual structure

2.3　MobileNetV2 结构

MobileNetV2 的网络结构参数见表 2，共包含 19 个层。其中，在 7 个不同尺度的特征图上会重复逆残差结构(Inverted Residual)1~4 次不等，Conv2 为普通卷积，在通过所有的逆残差块后，特征图经过一个卷积层和平均池化层(Avgpool)可以得到 k 个类别的分类。

表 2　MobileNetV2 网络参数

Tab. 2　MobileNetV2 network parameters

Conv	Channel	repeat	Stride	Output
Conv2d	32	1	2	150×150
Inverted Residual	16	1	1	150×150
Inverted Residual	24	2	2	75×75
Inverted Residual	32	3	2	38×38
Inverted Residual	64	4	2	19×19
Inverted Residual	96	3	1	19×19
Inverted Residual	160	3	2	10×10
Inverted Residual	320	1	1	10×10
Conv2d	1 280	1	1	10×10
Avgpool 7×7	-	1	-	10×10
Conv2d	K-classes	-	1	1×1

3　Att_Mobilenetv2_SSDLite 模型

Att_Mobilenetv2_SSDLite 目标检测算法网络结构如图 3 所示。将 SSD 的基础卷积层替换为改进后的 MobileNetV2，在逆残差结构中增加通道注意力机制和混洗通道加强特征融合，然后对区域候选框进行重构。分别在网络的 Conv11、Conv13_2、Conv14_2、Conv15_2、Conv16_2 这 5 个尺寸不同的特征层中生成预测框，预测框的数量由8 732减少为 2 254；最后通过非极大值抑制，去除置信度低于 0.5 的边框线条，得到检测结果。

Att_Mobilenetv2_SSDLite 使用深度可分离卷积，替换原 SSD 网络中额外增加层的标准卷积层，极大减少了运算量。基础网络使用 MobileNetV2，轻量型的基础网络对 SSD 网络的检测精度、检测速度和模型体积上都有良好的影响。

Att_Mobilenetv2_SSDLite 网络模型参数见表 3。基础网络层保留了 MobileNetV2 的 Conv1~Conv11，使用 Att-Inverted Residual 逆残差结构，缩减残差结构的重复次数，去掉最后用于分类的全连接层和池化层；额外添加 3 个卷积层作回归预测层。

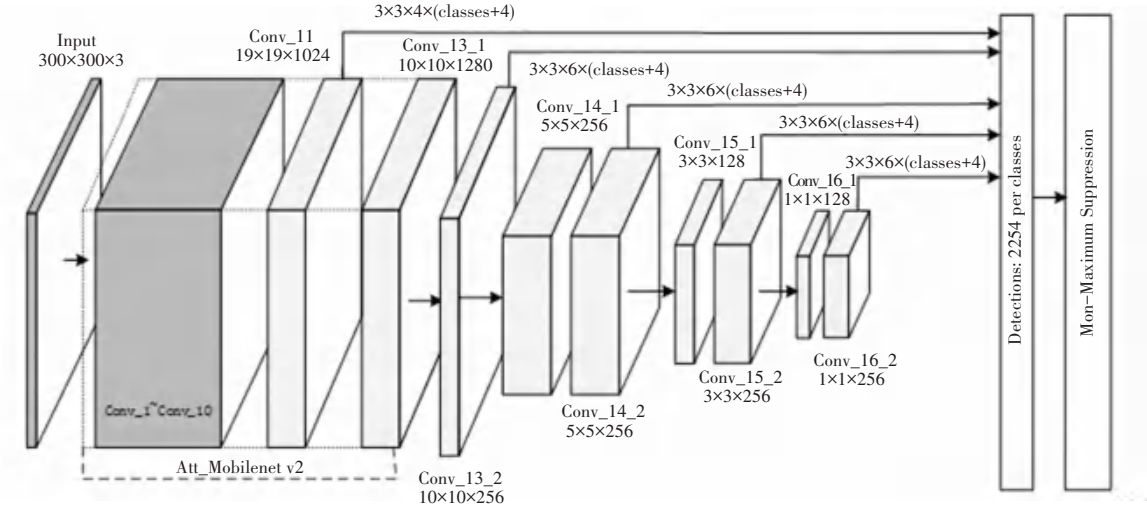


图 3 Att_Mobilenetv2_SSDLite 网络结构

Fig. 3 Network architecture of Att_Mobilenetv2_SSDLite

表 3 Att_Mobilenetv2_SSDLite 网络参数

Tab.3 network parameters of Att_Mobilenetv2_SSDLite

Conv	Channel	repeat	Stride	Output	predict
Conv2d	32	1	1	150×150	-
Att-Inverted Residual	16	1	1	150×150	-
Att-Inverted Residual	24	2	2	75×75	-
Att-Inverted Residual	32	2	2	38×38	-
Att-Inverted Residual	64	2	2	19×19	-
Att-Inverted Residual	96	2	1	19×19	√
Att-Inverted Residual	1 024	2	2	10×10	√
SperableConv2d	512	1	2	5×5	√
SperableConv2d	256	1	2	3×3	√
SperableConv2d	256	1	1	1×1	√

3.1 逆残差结构优化

如图 4 所示,本文采用的注意力逆残差结构(Attention Inverted Residual)与 ResNet 相反,在主干分支先对输入通道使用 1×1 的卷积进行扩张,在高维度上再使用深度卷积,而后再次经过 1×1 的卷积;在残差结构旁干分支中使用注意力算法,融合主干分支和旁干分支两类特征,最后将学习到的特征,通过混洗通道打乱重组,破除固定通道间特征无法学习的障碍,进一步提高网络的特征提取能力。

旁干分支中,先对输入通道进行全局池化压缩操作(Squeeze),经过压缩处理后的二位特征通道将变成不同值的实数,表示其在特征通道中响应的全局分布;然后经过 FC 全连接层、ReLU、FC、Sigmoid 的激活操作(Excitation),得到每个实数对应的权重;最后再经过加权操作(Reweight),分别给每个通道特征乘上权重,将权值赋给对应通道。通过注意力加权操作后,网络将重点关注权值更大的通道特征,有利于提取特征。

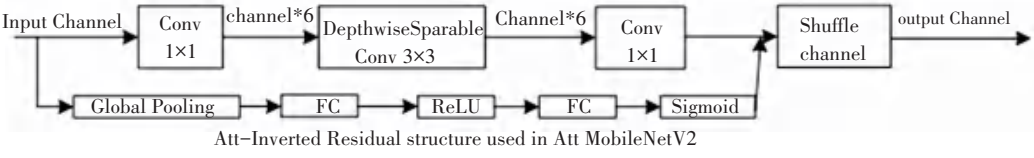


图 4 Att_InvertedResidual 结构

Fig. 4 Att_InvertedResidual structure

在逆残差结构的最后,用混洗通道操作(shuffle channel)打乱重组通道特征,是基于通道分组卷积实现的混洗通道卷积,将输入通道特征分为 g 组,每组分别与对应的卷积核卷积,卷积计算量降为原有的 $1/g$,然后对 g 组通道打乱重组。原本封闭固定的通道在打乱重组后特征得到了交流,解决了由固定分组导致特征融合效果差的问题。

在本文提出的 Att_MobileNetV2_SSDLite 模型中,将基础网络 MobileNetV2 的逆残差结构替换为 Att-InvertedResidual 结构。实验证明,本文的残差

结构可以在减少卷积层数量的同时保证特征提取能力。

3.2　回归定位损失函数优化

特征层中提取到的预测框，通过最后的非极大值抑制，得到置信度较高的预测框，再经过损失函数处理后，最终输出检测结果。SSD 的损失函数包含预测框定位回归损失（*SmoothL*1 损失函数[18]）和类别置信度分类损失（*Softmax* 损失函数）。

在回归过程中，对预测框和默认框重合程度的度量标准是交并比[9]（*IOU*），表示预测框和默认框重合部分占预测框和默认框总面积的比例，用公式表示为：

$$IOU = \frac{A \cap B}{A \cup B}, \tag{3}$$

但这种简单的面积交并比，并不能很好的反应预测框和默认框的重叠情况。例如：$A \cap B = 0$ 时 $IOU = 0$，这时 IOU 无法反向传递梯度，不能指导训练参数改变；并且当 $A \cap B \neq 0$ 时，预测框和默认框有多种重叠形式，同样不能指导梯度传递。为此，本文考虑预测框与默认框的重叠程度，对 IOU 的计算方式做部分改进。将预测框与默认框两中心点欧式距离作为惩罚项，与预测框和默认框重叠面积的最小矩形对角线做比，用公式表示为：

$$M_{IOU} = IOU - \frac{d^2(A,B)}{c^2}, \tag{4}$$

其中，$d^2(A,B)$ 表示预测框 A 中心点与默认框 B 中心点的欧式距离，c^2 表示预测框 A 与默认框 B 最小覆盖的矩形面积的对角线距离。用左上和右下两点坐标表示定位框位置，预测框记：$A = (x_1, y_1, x_2, y_2)$；默认框记：$B = (x_1', y_1', x_2', y_2')$；预测框 A 与默认框 B 最小覆盖的矩形 C 记：$C = (x_1^c, y_1^c, x_2^c, y_2^c)$。

预测框面积 S 与默认框面积 S' 分别记为：

$$S = |(x_2 - x_1) * (y_2 - y_1)|, \tag{5}$$

$$S' = |(x_2' - x_1') * (y_2' - y_1')|, \tag{6}$$

重叠面积 S^O 为：

$$S^o = \begin{cases} |(x_2^o - x_1^o) * (y_2^o - y_1^o)|, x_2^o > x_1^o, y_2^o - y_1^o; \\ 0, otherwise, \end{cases} \tag{7}$$

其中：

$$\begin{cases} x_1^o, y_1^o = \max(x_1, x_1'), \max(y_1, y_1'), \\ x_2^o, y_2^o = \min(x_2, x_2'), \min(y_2, y_2'). \end{cases} \tag{8}$$

得到 IOU 为：

$$IOU = \frac{S^O}{S + S' - S^O}, \tag{9}$$

预测框与默认框最小覆盖矩形对角线 C^l 为：

$$C^l = \sqrt{|(x_2^c - x_1^c)^2 + (y_2^c - y_1^c)^2|}, \tag{10}$$

其中：

$$\begin{cases} x_1^c, y_1^c = \min(x_1, x_1'), \min(y_1, y_1'), \\ x_2^c, y_2^c = \max(x_2, x_2'), \max(y_2, y_2'). \end{cases} \tag{11}$$

预测框 A 与默认框 B 中心点分别为：

$$A^m = \left(\frac{x_2 - x_1}{2} + x_1, \frac{y_2 - y_1}{2} + y_1\right), x_2 > x_1, y_2 > y_1, \tag{12}$$

$$B^m = \left(\frac{x_2' - x_1'}{2} + x_1', \frac{y_2' - y_1'}{2} + y_1'\right), x_2' > x_1', y_2' > y_1'. \tag{13}$$

欧式距离为：

$$d^2 = \sqrt{\left(\frac{x_2' - x_1'}{2} + x_1'\right) - \left(\frac{x_2 - x_1}{2} + x_1\right)^2 + \left(\frac{y_2' - y_1'}{2} + y_1'\right) - \left(\frac{y_2 - y_1}{2} + y_1\right)^2}. \tag{14}$$

最后可以得到预测框回归损失函数为：

$$L_{M_{IOU}} = 1 - IOU + \frac{d^2(A,B)}{c^2}. \tag{15}$$

3.3　预测候选框

预测候选框[9]是在特征图上，按一定纵横比（aspect retios）生成的可能包含预测目标的预测候选框，浅层分辨率高的特征图主要检测小目标，深层分辨率小的特征图检测大目标。预测候选框通过非极大值抑制算法后，输出与目标最为匹配的检测框。表 4 与表 5 分别列出了 SSD 与本文算法 Att_Mobilenetv2_SSDLite 在不同尺度特征图上对应的预测候选框数量。本文主要检测目标为菜品，菜品目标通常是图片中占比最大的目标，故本文在设置默认候选框时，关注更深层的特征图的默认候选框。

第 k 个默认候选框的大小 S_k 计算方式为：

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m - 1}(k - 1), k \in [1, m], \tag{16}$$

$$w_k^a = S_k \sqrt{a_r}, h_k^a = S_k / \sqrt{a_r}, a_r \in \{1, 2, 3, 1/2, 1/3\}. \tag{17}$$

其中，$S_{\max} = 0.9$、$S_{\min} = 0.2$，分别表示默认候选框最大归一化尺寸比例和最小归一化尺寸比例；m 是该特征图的默认候选框个数；w_k^a 为候选框宽度；h_k^a 为候选框高度。除设置的纵横比以外，当 *aspect*

$retios = 1$ 时,额外添加候选框 $S_k' = \sqrt{S_k S_k + 1}$。

表 4 SSD 与 Att_Mobilenetv2_SSDLite 网络参数

Tab. 4 SSD and Att_Mobilenetv2_SSDLite network parameters

SSD			
Size	*aspect retios*	*area candidate box*	*boxes*
38×38	{1,2,1/2}	4	5 776
19×19	{1,2,3,1/2,1/3}	6	2 166
10×10	{1,2,3,1/2,1/3}	6	600
5×5	{1,2,3,1/2,1/3}	6	150
3×3	{1,2,1/2}	4	36
1×1	{1,2,1/2}	4	4

表 5 Att_Mobilenetv2_SSDLite 网络参数

Tab. 5 Att_Mobilenetv2_SSDLite network parameters

Att_Mobilenetv2_SSDLite			
Size	*aspect retios*	*area candidate box*	*boxes*
19×19	{1,2,1/2}	4	1 444
10×10	{1,2,3,1/2,1/3}	6	600
5×5	{1,2,3,1/2,1/3}	6	150
3×3	{1,2,3,1/2,1/3}	6	54
1×1	{1,2,3,1/2,1/3}	6	6

特征图上的每个像素点都会生成 4 个或 6 个默认候选框,本文提出的 Att_Mobilenetv2_SSDLite 网络生成 2 254 个默认候选框,主要舍弃了用于检测小目标的预测框。

4 实验结果与分析

本文针对菜品识别设计了 Att_Mobilenetv2_SSDLite 网络,使用改进后的 Att_MobileNetV2 作基础网络提取特征,额外增加 3 个卷积层作为回归预测层。为了减小模型体积和提高检测速度,类别置信度与定位回归预测中均使用可分离卷积代替传统卷积。针对 *IOU* 不能很好反应预测框与默认框之间的重叠效果,设计了 $L_{M_{IOU}}$ 回归损失函数,替换 *SmoothL*1 损失函数。

4.1 实验环境与数据

实验环境为 pytorch1.7、python3.7,AMD Ryzen 7 4800H 处理器,NVIDIA RTX 2060 6G,在 windows 环境下运行,使用数据集格式为 PASCAL VOC。数据集为自建中餐菜品数据集 Chinesefood,选取中餐菜品数据集 release_data 中的 20 类菜品,每个菜品 1 067张图片,数据集共包含21 340张图片;使用 LabelImg 标准软件生成 VOC 格式数据集,按照 80%、10%、10%的比例划分训练数据、验证数据和测试数据。*Batchsize* 设为 32,初始学习率设为 0.001,动量为 0.9,每迭代 50 次降低学习率为上一阶段 0.5。

4.2 评价指标

实验评价指标为检测速度(*fps*)、平均准确率(*Average Precision*,*mAP*)和平均召回率(*Average Recall*,*AR*)。*fps* 表示每秒检测图片数量,值越高表示检测速度越快;*mAP* 表示不同阈值条件下的平均准确率,值越高表示检测效果越好;*AR* 表示平均召回率,值越高表示漏检率越低。

$$\begin{cases} mAP = mean\dfrac{TP}{TP + FP}, \\ AR = \dfrac{TP}{TP + FN}. \end{cases} \tag{18}$$

式中,*TP* 表示正样本中预测正确部分;*FP* 表示正样本中预测错误部分;*FN* 表示负样本中预测错误部分。

4.3 实验结果

对比实验设计,使用不同数量回归预测层的网络进行对比,寻找最合适数量的回归预测层;针对本文提出的改进点进行控制变量组合训练,探索改进点对网络整体提升能力。训练过程使用自建的中餐菜品数据集 Chinesefood,并对输入的图像随机翻转、裁剪以及颜色随机调整。

4.3.1 不同数量回归预测层对比

本文设计了分别有 4、5、6 个回归预测层的 SSD 模型。表 6 列出了不同层结构的网络在菜品数据集上的表现,其中 *features_number* 表示该模型包含回归预测层的个数。

表 6 不同数量预测层模型检测对比

Tab. 6 Comparison of model detection with different numbers of prediction layers

Model	*Features number*	*fps*	*mAP*	*AR*	*Model size*
Att_Mobilenetv2_SSDLite	4	114.40	94.8%	84.1%	10.4M
Att_Mobilenetv2_SSDLite	5	105.45	96.3%	84.2%	11.5M
Att_Mobilenetv2_SSDLite	6	44.12	97.3%	86.1%	14.1M

回归预测层数量越多,模型检测准确率越高,但是模型计算量随之增加,导致检测速度下降,每增加一层,准确率大约提升 1%,模型体积增大 1~2 M,见表 5。回归预测层数量为 4 时,检测速度最快,模型体积也最小,但检测准确率与召回率最低;回归预测层数量为 6 时,准确率和召回率最高,但检测速度大幅下降,模型体积也增大。图 5 为不同数量特征

层的检测效果。本文使用 5 个回归预测层的模型表现最好，保证了速度和准确率，模型体积也控制在合适范围。

（a）4 层回归预测层　（b）5 层回归预测层　（c）6 层回归预测层

（a）4 feature maps　（b）5 feature maps　（c）6 feature maps

图 5　不同数量预测层检测效果

Fig. 5　The detection effect of different numbers of prediction layers

4.3.2　模型优化效果

实验中分别比较了注意力逆残差块（Att－InvertedResidual）、深度可分离卷积（SperableConv2d）和回归定位损失函数（$L_{M_{IOU}}$）对模型的增益效果。从表 7 结果可见，注意力逆残差块提升准确率约 0.9%，深度可分离卷积可减小模型体积约 24.6 M，回归定位损失可提高准确率 1.6%，召回率 1.2%。

图 6 对比了使用 $L_{M_{IOU}}$ 和 *SmoothL*1 分别作为回归定位损失的损失变化图。实验表明，$L_{M_{IOU}}$ 损失函数对 IOU 计算方式的调整，能更全面的了解预测框与默认框的重叠关系，加快模型收敛，损失值下降更快，提升网络识别准确率。

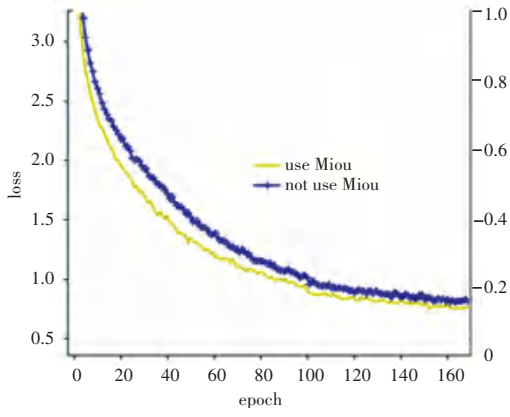


图 6　不同损失函数对比

Fig. 6　Comparison of different loss functions

表 7　不同结构对比

Tab. 7　Comparison of different structural

Model	*Features number*	Att−Inverted−Residual	Sperable−Conv2d	$L_{M_{IOU}}$	*fps*	*mAP*	*AR*	*Model size*
Att_Mobilenetv2_SSDLite	5	√	√	√	105.45	96.3%	84.2%	11.5M
Mobilenetv2_SSDLite	5	−	√	√	109.73	95.4%	84.1%	11.4M
Att_Mobilenetv2_SSD	5	√	−	√	107.54	96.1%	85.5%	36.1M
Att_Mobilenetv2_SSDLite	5	√	√	−	96.02	94.7%	83.0%	11.5M

4.3.3　不同模型效果比较

从表 8 可以看出，对比其它的目标检测模型，本文设计的模型相比 SSD300、Tiny YOLO 和 Mobilenetv1_SSD，准确率相似，但模型体积最小，检测速度也最优越。对比 Tiny SSD300，本文的模型虽然体积略大，但准确率和速度都远远大于 Tiny SSD300。图 7 对比了不同模型对菜品的检测效果，综合模型体积、准确率和速度，本文的模型相比 SSD 和 YOLO 具有更好的效果。

表 8　不同模型检测对比

Tab. 8　Comparison of different models

Model	*fps*	*mAP*	*AR*	*Model size*
Att_Mobilenetv2_SSDLite	105.45	96.3%	84.2%	11.5 M
Mobilenetv1_SSD	90.55	95.3%	85.2%	52.1 M
SSD300	45.94	97.6%	88.5%	104.5 M
Tiny SSD300	39.61	94.3%	82.1%	2.3 M
Tiny YOLO	35.22	96.6%	85.7%	60.5 M

(a) ours (b) Mobilenetv1 ssd (c) ssd300 (d) tinyssd300 (e) tiny YOLO

图 7 不同目标检测模型检测效果

Fig. 7 The detection effect of different target detection models

5 结束语

本文针对菜品识别设计了基于 SSD 的轻量型目标检测网络 Att_Mobilenetv2_SSDLite 模型,该模型采用 MobileNetV2 作为基础网络,缩小模型体积提升检测速度。对原 MobileNetV2 模型中的逆残差结构做调整,使用注意力机制和混洗通道增强特征提取能力。考虑预测框与默认框位置重叠因素设计回归损失函数,加快模型收敛。重新规划回归预测层数量,重点关注图片中大目标,提高检测准确率。通过与其他目标检测网络对比,本文提出的模型更加适合识别菜品,并且适合部署在存储能力弱的其他平台。

参考文献

[1] 梁子鑫. 探讨新时代背景下新兴技术在人工智能中的应用[J]. 软件, 2018, 39(7):166-169.

[2] 汪聪. 基于机器视觉的菜品智能识别技术研究[D]. 广州: 华南理工大学,2019.

[3] 董天骄. 基于卷积神经网络的饮食分类与识别[D]. 杭州: 杭州电子科技大学,2018.

[4] 吴正东. 基于深度学习的中餐菜品图像分类算法研究[D]. 成都: 电子科技大学,2020.

[5] 王瑞. 基于卷积神经网络的图像识别[D]. 开封: 河南大学, 2015.

[6] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph-based image segmentation[J]. International journal of computer vision, 2004, 59(2): 167-181.

[7] 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, 2017, 39(6):1137-1149.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[9] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[10] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510-4520.

[11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:7232-7141.

[12] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848-6856.

[13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations, 2015:1-14.

[14] 毕鹏程,罗健欣,陈卫卫. 轻量化卷积神经网络技术研究[J]. 计算机工程与应用,2019,55(16):25-35.

[15] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[17] 吴天舒,张志佳,刘云鹏,等. 基于改进 SSD 的轻量化小目标检测算法[J]. 红外与激光工程,2018,47(7):47-53.

[18] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.