中国光学 Chinese Optics 核心期刊

八度卷积和双向门控循环单元结合的X光安检图像分类

吴海滨 魏喜盈 王爱丽 岩堀祐之

X-ray security inspection images classification combined octave convolution and bidirectional GRU

WU Hai-bin, WEI Xi-ying, WANG Ai-li, YUJI Iwahori

引用本文:
吴海滨, 魏喜盈, 王爱丽, 岩堀祐之. 八度卷积和双向门控循环单元结合的X光安检图像分类[J]. 中国光学, 2020, 13(5): 1138-1146. doi: 10.37188/CO.2020-0073
WU Hai-bin, WEI Xi-ying, WANG Ai-li, YUJI Iwahori. X-ray security inspection images classification combined octave convolution and bidirectional GRU[J]. *Chinese Optics*, 2020, 13(5): 1138-1146. doi: 10.37188/CO.2020-0073

在线阅读 View online: https://doi.org/10.37188/CO.2020-0073

文章编号 2095-1531(2020)05-1138-09

八度卷积和双向门控循环单元结合的X光安检图像分类

吴海滨[1], 魏喜盈[1], 王爱丽[1*], 岩堀祐之[2]
(1. 哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080;
2. 中部大学 计算机科学学院, 日本 爱知 487-8501)

摘要: 针对主动视觉安检方法准确率低、速度慢, 不适用于实时交通安检的问题, 提出了八度卷积(OctConv)和注意力机制双向门控循环单元(GRU)神经网络相结合的X光安检图像分类方法。首先, 利用八度卷积代替传统卷积, 对输入的特征向量进行高低分频, 并降低低频特征的分辨率, 在有效提取X光安检图像特征的同时, 减少了空间冗余。其次, 通过注意力机制双向GRU, 动态学习调整特征权重, 提高危险品分类准确率。最后, 在通用SIXRay数据集上的实验表明, 对8000幅测试样本的整体分类准确率(ACC)、特征曲线下方面积(AUC)、正类分类准确率(PRE)分别为98.73%、91.39%、85.44%, 检测时间为36.80 s。相对于目前主流模型, 本文方法有效提高了X光安检图像危险品分类的准确率和速度。

关 键 词: X光安检图像; 八度卷积; 双向门控循环单元; 注意力机制
中图分类号: TP391.4 **文献标志码:** A **doi:** 10.37188/CO.2020-0073

X-ray security inspection images classification combined octave convolution and bidirectional GRU

WU Hai-bin[1], WEI Xi-ying[1], WANG Ai-li[1*], YUJI Iwahori[2]
(1. *College of Measurement–Control Technology and Communication Engineering, Harbin University of Science and Technology, Harbin* 150080, *China*;
2. *Computer Science, Chubu University, Aichi* 487-8501, *Japan*)
* *Corresponding author, E-mail: aili*925@*hrbust.edu.cn*

Abstract: Due to the disadvantages of low accuracy and slow speed in the active vision security inspection method, it is not suitable for real-time security inspection. Aiming at this problem, we propose an x-ray inspection image classification algorithm combining octave convolution (OctConv) with attention-based bidirectional Gate Recurrent Unit (GRU). Firstly, OctConv is introduced to replace the traditional convolution operation to divide the input feature vector into high and low frequency, and reduce the resolution of low fre-

收稿日期: 2020-04-23; **修订日期:** 2020-06-15
基金项目: 国家自然科学基金(No. 61671190)
Supported by National Natural Science Foundation of China (No. 61671190)

quency features, effectively extracting the features of security image and reducing the spatial redundancy. Then, the feature weight can be adjusted by dynamic learning through attention-based bidirectional GRU to improve the classification accuracy of threat objects. Finally, a lot of experimental results on SIXRay dataset show that the classification accuracy, AUC value and PRE of 8 000 test samples are 98.73%, 91.39% and 85.44%, respectively, with a time of 36.80 seconds. Compared with the current mainstream model, the proposed algorithm can improve the performance and speed of threat objects recognition in X-ray security images.

Key words: X-ray inspection images; octave convolution; bidirectional GRU; attention mechanism

1　引　言

X 光安检系统广泛应用于交通安全出入控制[1], 可以有效降低犯罪、恐怖袭击的风险[2]。目前 X 光安检工作基本为人工操作, 操作人员必须专注于每个行李包裹的筛选, 即使经过专业培训, 人为错误的可能性仍然较大。随着 5G 和 AI 技术的发展, 探索基于机器学习的高准确率、快速、容错 X 光安检图像危险品分类方法具有实际意义。

基于传统特征提取的 X 光安检图像分类方法存在检测速度慢、泛化能力弱等不足[3-4]。随着深度学习的高速发展, 使其在图像分类[5-6]、目标检测[7-8]、特征融合[9-10] 等任务中发挥了至关重要的作用。近年来, 深度神经网络能够有效学习数据本身的特征信息, 已被用于提高安检图像的分类准确率[11-15]。Mery 等在 GDXray 数据集上分别对基于词典、稀疏表示、深度学习等 10 余种分类方法进行了测试讨论[16]。Galvez 等采用 VGG-16 网络在 ImageNet 数据集上进行预训练, 再将该模型作为特征提取器迁移到危险品分类中[17]。

以上深度神经网络模型需要大量的数据和强大的计算能力以保证训练的有效性。为此, Howard 等使用深度可分离卷积建立了轻量级神经网络 MobileNets, 其具有规模小、延迟少、功耗低等特点[18]。Forrest 等提出 Fire 模块, 使用 1×1 卷积代替部分 3×3 卷积, 降低了模型的存储空间[19]。Chen 等提出一种即插即用 OctConv, 将高频特征和低频特征映射到不同组中, 通过相邻位置间的信息共享, 可以优化缓存和计算量[20]。

门控循环单元神经网络(GRU)[21] 用于处理序列数据, 与前馈神经网络不同, 其会记忆前序信息并用于当前输出计算, 即隐藏层的输入不仅包括输入层的输出, 还包括上一时刻隐藏层的输出。本文将八度卷积与引入注意力机制的双向 GRU 相结合(OctConv-ABiGRU), 提出了 X 光安检图像分类方法, 该方法在提高分类准确率的同时, 还满足实时性要求。

2　OctConv-ABiGRU 模型

对于给定的一组 X 光安检图像, 本文利用判别函数 $y = f(x_n;\theta)$ 预测其中所有可能包含的标签。每幅安检图像可表示为:

$$x_n = \sum\nolimits_c y^*_{n,c} \cdot x_{n,c}\ , \qquad (1)$$

其中, $x_{n,c}$ 表示从原始图像中采样得到的子图像, 每一幅子图像对应一个类别 c。每个维度的 $y^*_{n,c}$ 为 0 或者 1。0 表示不包含该类别, 1 表示包含该类别。

本文方法主要包括 3 个部分, 如图 1 所示。

(1)为了减少类别的不平衡性对分类器的影响, 首先进行数据增强, 通过对少数类样本(至少包含一种危险品的样本)进行随机过采样, 以增加少数类样本的代表性。

(2)利用 OctConv 将从 X 光安检图像中提取的特征向量映射为高频特征和低频特征, 分别进行卷积、池化等操作, 然后利用 BiGRU 构建特征序列。

(3)采用注意机制匹配特征图通道与样本类别, 进一步确定需要注意的特征, 得到不同的特征图。最后采用 sigmoid 分类器进行分类。

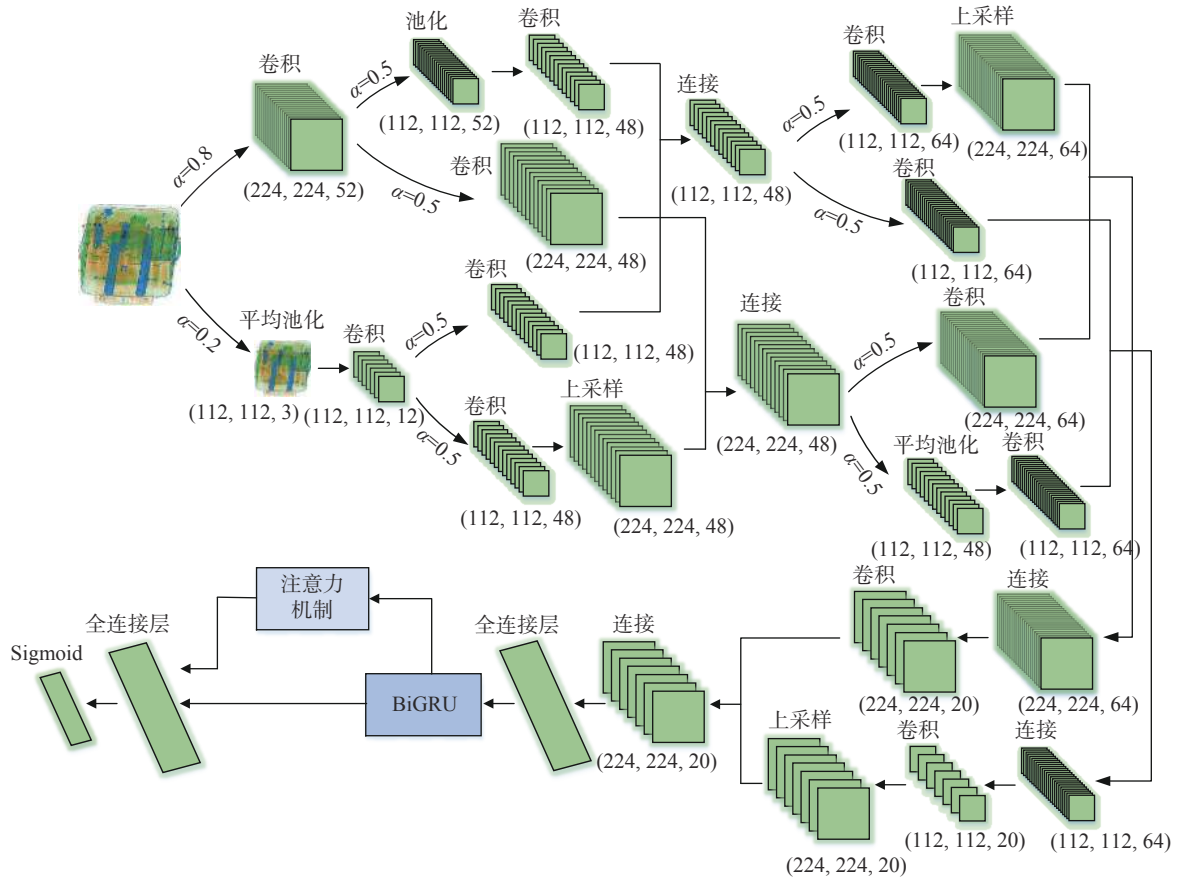

图 1 X 光安检图像分类算法框图

Fig. 1 Block diagram of X-ray security image classification algorithm

2.1 数据增强

X 光安检图像中包含危险品的少数类样本过少，会造成分类面偏倚，导致少数类样本的分类准确率低。因此，本文利用随机过采样方法进行数据增强，即随机复制少数类样本，扩充数据集。为了防止过拟合，在每次生成新数据点时加入轻微随机扰动。

2.2 八度卷积

频域图像包含高频部分和低频部分，同样，卷积层的输出特征映射也可分解为不同频率成分。OctConv 是一种多频特征表示方法，将高频和低频映射存储到不同组中，并且使用低维度向量储存和处理特征图中的低频部分，由于低频分量是冗余的，可以通过降低低频特征的分辨率来降低冗余和计算成本。

在 OctConv 中，输入和输出的特征向量及卷积核都被分解为高频[$\boldsymbol{X}^{\mathrm{H}}$,$\boldsymbol{Y}^{\mathrm{H}}$,$\boldsymbol{W}^{\mathrm{H}}$]和低频[$\boldsymbol{X}^{\mathrm{L}}$,$\boldsymbol{Y}^{\mathrm{L}}$,$\boldsymbol{W}^{\mathrm{L}}$]两个部分。低频向高频进行信息更新时，即$\boldsymbol{W}^{\mathrm{L}\to\mathrm{H}}$，需要对低频分量上采样再卷积；同理，在$\boldsymbol{W}^{\mathrm{H}\to\mathrm{L}}$过程中，需要对高频分量下采样。具体过程如式 (2) 和式 (3) 所示：

$$\boldsymbol{Y}^{\mathrm{H}}_{p,q}=\sum_{i,j\in N_k}\boldsymbol{W}^{\mathrm{T}}_{i+\frac{k-1}{2},j+\frac{k-1}{2}}\boldsymbol{X}^{\mathrm{H}}_{p+i,q+j}+\sum_{i,j\in N_k}\boldsymbol{W}^{\mathrm{T}}_{i+\frac{k-1}{2},j+\frac{k-1}{2}}\boldsymbol{X}^{\mathrm{L}}_{\left(\left\lfloor\frac{p}{2}\right\rfloor+i\right)\left(\left\lfloor\frac{q}{2}\right\rfloor+j\right)},\tag{2}$$

$$\boldsymbol{Y}^{\mathrm{L}}_{p,q}=\sum_{i,j\in N_k}\boldsymbol{W}^{\mathrm{T}}_{i+\frac{k-1}{2},j+\frac{k-1}{2}}\boldsymbol{X}^{\mathrm{L}}_{p+i,q+j}+\sum_{i,j\in N_k}\boldsymbol{W}^{\mathrm{T}}_{i+\frac{k-1}{2},j+\frac{k-1}{2}}\boldsymbol{X}^{\mathrm{L}}_{(2p+0.5+i),(2q+0.5+j)},\tag{3}$$

其中，(p,q)为位置坐标，N_k定义了一个局部领域。

八度卷积结构如图 2 所示，采用平均池化进行下采样操作。其中 α 表示低频特征通道比例，低频特征定义为比高频特征低一个八度，即低频特征图的分辨率仅为高频特征图的一半。

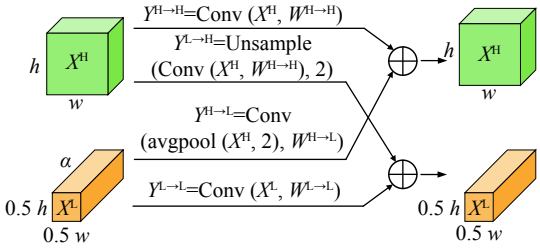

图 2　八度卷积结构

Fig. 2　The structure of octave convolution

2.3　双向门控循环单元神经网络

长短时记忆网络(LSTM)通过在每个单元中引入门控状态来控制传输状态, 记住需要长时间记忆的信息, 忘记不重要信息。

设$\boldsymbol{W}$是门的权重向量, b 是偏置项, σ是 sigmoid 函数, 输入均为$[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t]$, 表示把两个向量连接成更长的向量。LSTM 主要包含 3 个阶段:

(1)忘记阶段。通过计算得到的z^f作为忘记门控, 控制上一状态的哪些信息需要记住, 哪些信息需要忘记。

$$z^f = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_f). \tag{4}$$

(2)选择记忆阶段。对输入的$\boldsymbol{x}_t$进行选择记忆。当前的输入内容由前面计算得到的z^f表示, 选择的门控信号由z^i控制。

$$z^i = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_i), \tag{5}$$

$$\tilde{C}_t = \tanh(\boldsymbol{W}_c \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_c). \tag{6}$$

(3)输出阶段。通过 tanh 激活函数对上一阶段得到的$\boldsymbol{C}_t$进行缩放, 决定哪些将会作为当前状态的输出。

$$C_t = z^f C_{t-1} + z^i \tilde{C}_t, \tag{7}$$

$$o^t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_o), \tag{8}$$

$$h_t = o_t \tanh(c_t). \tag{9}$$

GRU 是 LSTM 的重要变体, 它将 LSTM 中的 3 个门控单元简化为两个: 更新门和重置门, 并且将单元状态和输出合而为一。这样在保证网络性能的同时, 减少了网络参数, 提高了训练效率, 更容易收敛。为了充分利用前后图像特征序列信息, 本文设计了一种双层双向 GRU 网络。相比于传统的单向神经网络, BiGRU 由前后两种状态共同决定。双层网络增强了网络的学习能力, 信息利用的更加充分。在每个时刻, 输入端都提供两个相反方向的 GRU, 输出端由两个 GRU 共同决定。双层 BiGRU 结构如图 3 所示。

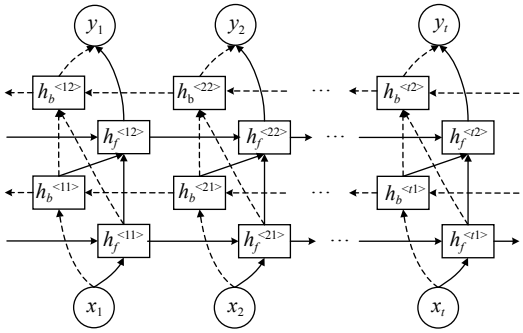

图 3　双层 BiGRU 结构

Fig. 3　The structure of double-layer BiGRU

2.4　注意力机制

注意力机制是模仿人类的生物系统, 只注意输入的相关元素, 而忽略其他不相关信息。如人类在观察图像时, 并非一次性观察整幅图像的每个位置, 而是选择性地观察特定部分, 将注意力集中在含有图像特征的区域[22]。人脑会根据经验, 学习图像注意力集中的目标, 从众多信息中选择出对当前任务更关键的信息, 其核心思想是对不同的信息赋予不同的权重。

一幅 X 光安检图像中通常包含多个对象, 并且背景杂乱, 毫无规律; 每个对象可能出现在图片中的任意位置, 这意味着图像的不同区域的重要性不同, 并且两幅图像之间关注区域可能完全不同; 对于特定类(例如枪支), 它们在不同的图像中的尺寸和形状不同。

本文利用自上而下的残差注意力模型, 为 BiGRU 加入注意力机制, 重点学习图像中待处理要素。当前状态根据前序状态学习得到的关注特征和当前输入特征, 进而学习调整特征权重, 使输出的特征有更强的区分度。具体过程如下:

(1)首先, BiGRU 输出隐藏层向量$\boldsymbol{x}_{i,c}$, 并将其作为注意力模块的输入。通过由注意力机制随机初始化得到的概率权重$\boldsymbol{w}^{\mathrm{T}}$与$\boldsymbol{x}_{i,e}$的乘积经过 Softmax 函数计算得到注意力参数$W_{i,c}(x)$。

(2)其次, 利用残差学习的方式, 将主干特征图与注意力特征图相乘, 再进行规范化, 从而得到注意力层的输出。最终可以让网络集中在需要注

意的特征上，抑制不需要的特征。如式 (10) 和式 (11) 所示：

$$\boldsymbol{y}_{i,c}(x)=(1+W_{i,c}(x))f_{i,c}(x), \tag{10}$$

$$f_{i,c}(x)=\frac{1}{1+\exp\left(-\dfrac{\boldsymbol{x}_{i,c}-mean_c}{std_c}\right)}, \tag{11}$$

其中，$\boldsymbol{y}_{i,c}$为注意力模块的输出，$f_{i,c}$为图片特征张量在通道域上的平均值的激活函数，$mean_c$为均值，std_c为标准差。

3 X 光安检图像分类实验

3.1 SIXray 数据集

本文采用中国科学院在 2019 年公布的 Security Inspection X-ray (SIXray) 数据集[23]，它包含 100 万余幅来自于真实场景的 X 光安检图像，其中正类样本少于 1%，如表 1 所示。

表 1 SIXray 数据集样本分布

Tab. 1 Sample distribution in SIXray dataset

正类样本 (8929)					负类样本
枪支	刀具	扳手	钳子	剪子	
3131	1943	2199	3961	983	1050302

其中一个正类样本中可能包含多种危险品（例如：枪支、刀具以及剪子）。SIXray 数据集中的样本为不同材质的物品分配了不同的颜色，物品的尺度、视角、样式也存在较大差异，并且危险品一般都与其他安全品交叠，如图 4 所示。

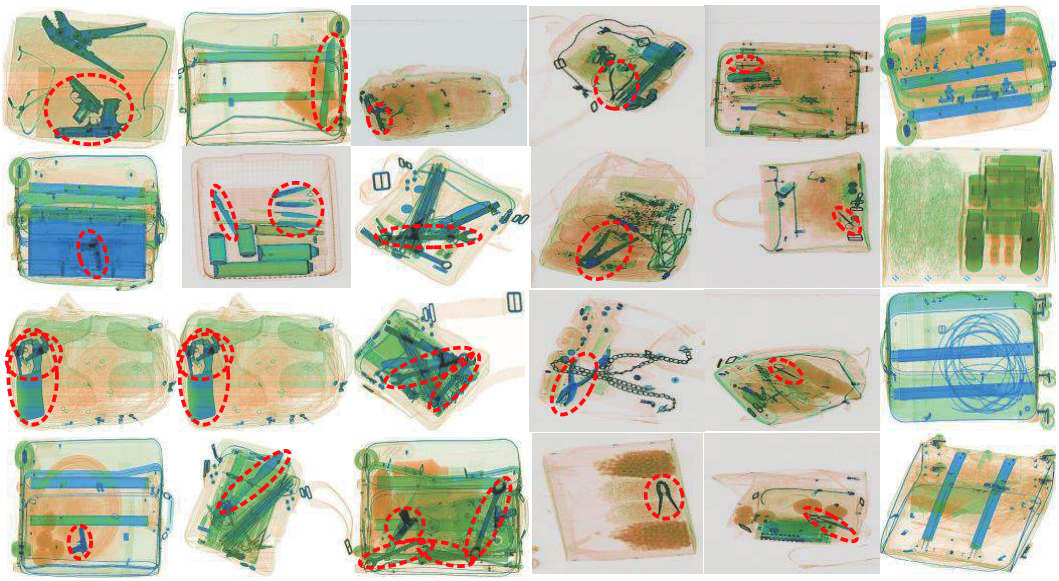

图 4 SIXray 数据集

Fig. 4 SIXRay dataset

3.2 评价指标

为了更好地评估类别不平衡对实验的影响，本文通过整体分类准确率(ACC)，接受者操作特征曲线下方面积(AUC)两个指标评估本文方法性能。

为计算 ACC 定义 4 种情况：TP(True Positives)、FN(False Negatives)、FP(False Positives)、TN(True Negatives)，对于正类样本，如果预测结果中包含危险品标记，则将其定义为 TP，否则定义为 FP；对于负类样本，如果预测结果中包含危险品标记，则将其定义为 FN，否则定义为 TN。

ACC 和 PRE 的定义如下：

$$ACC=\frac{TP+TN}{TP+FN+FP+TN}, \tag{12}$$

$$PRE=\frac{TP}{TP+FP}. \tag{13}$$

AUC 定义为 ROC 曲线下与坐标轴围成的面

积, 其中 ROC 是由 TP、FN、FP、TN 计算得到的曲线。AUC 的值越大, 代表分类器性能越好。

3.3　实验结果分析

本文实验均在 Windows 操作系统下实现, 采用 python 语言和 Keras 库编码, NVIDIA RTX2070 显卡 GPU 加速, CPU 为 32GB Intel® Core™ i7-9750H CPU @ 2.60 GHz。实验采用 SIXRay 数据集中的子集 SIXRay10, 其中包含 8 929 个正类样本, 10 倍的负类样本, 经数据增强后, 将总样本数扩充为 102 331 个。其中训练集为图像总数的 80%, 测试集为图像总数的 20%。

经过数据增强前后的每类危险品正负样本数及不平衡率对比如表 2 所示。由表 2 可知, 经过数据增强后, 各类危险品的类别不平衡比率均有所减少。

表 2　不同类别数据增强前后对比结果

Tab. 2　Comparison results of different types of data before and after data augmentation

种类	增强前后	负类样本数	正类样本数	不平衡比率
枪支	增强前	72 255	2 705	26.27
	增强后	89 672	12 659	7.08
刀具	增强前	73 212	1 748	41.88
	增强后	93 723	8 608	10.89
扳手	增强前	72 948	2 012	36.26
	增强后	92 380	9 951	9.28
钳子	增强前	71 524	3 436	20.82
	增强后	85 574	16 757	5.10
剪子	增强前	74 153	807	91.89
	增强后	99 760	2 571	38.80

每幅图像输入尺寸为 224×224, 并归一化至 (−1,1), 选择 Adam 作为梯度优化算法。训练的学习率为 10^{-4}, 且每 5 个 epoch 后, 学习率衰减为原来 1/4, 共计进行 40 000 次迭代。

在八度卷积部分, 首先设置 α=0.2, 随机选择 20% 的输入特征图, 并对其进行平均池化操作, 输出特征图大小为 112×112, 其余特征图保持原尺寸 224×224; 然后设置 α=0.5, 重复上述操作。连接相同尺寸的特征图, 最后将生成的 224×224 大小的特征图输入到全连接层。

为验证本文分类算法的有效性, 实验对比了目前的主流分类模型, 包括 InceptionV3[24]、VGG19[25]、ResNet[26]、DenseNet[27]、STN-DenseNet[28]。为了验证八度卷积对模型的贡献, 对比了 BiGRU 和 CNN-ABiGRU。

表 3、表 4 分别给出以上 8 种方法对 5 种危险品分类得到的 ACC、AUC 结果, 可知 OctConv-ABiGRU 模型的平均 ACC 和平均 AUC 均为最优, 分别达到 98.73% 和 91.39%。尤其对扳手、钳子、剪子的分类效果提升明显, 与 CNN-ABiGRU 相比, AUC 分别提高了 5.62%、6.06%、7.28%。标准的基线模型, 更关注枪支的分类效果, 而牺牲了其他类别物品判断的正确率。而本模型利用注意力机制将特征图通道和数据集中的类别相对应, 意味着不同的特征图通道能够注意不同的类别, 从而使各个类别的分类准确率相对均衡。

表 3　不同模型的 ACC (%) 比较

Tab. 3　Comparison of ACC (%) for different network modules

方法	枪支	刀具	扳手	钳子	剪子	平均
InceptionV3	94.63	87.52	88.97	80.50	96.95	89.71
VGG19	97.88	98.36	97.48	96.03	97.33	97.42
ResNet	98.36	99.20	98.16	96.10	97.80	97.92
DenseNet	98.69	99.25	98.18	96.16	97.65	97.99
STN-DenseNet	99.15	98.73	97.52	96.32	98.46	98.03
OnlyBiGRU	98.77	99.40	97.73	94.37	99.14	97.88
CNN-ABiGRU	98.89	99.42	98.89	97.07	98.96	98.65
OctConv-ABiGRU	98.60	99.25	99.10	97.50	99.20	98.73

表 4　不同模型的 AUC (%) 比较

Tab. 4　Comparison of AUC (%) for different network modules

方法	枪支	刀具	扳手	钳子	剪子	平均
InceptionV3	63.34	54.57	51.33	52.92	50.74	54.57
VGG19	93.34	89.03	77.49	76.57	71.08	81.50
ResNet	94.06	88.68	76.00	73.92	60.45	78.64
DenseNet	93.91	90.37	72.59	74.65	61.08	78.52
STN-DenseNet	95.69	93.58	75.60	76.98	65.09	81.39
OnlyBiGRU	92.73	93.90	68.03	73.33	89.42	83.48
CNN-ABiGRU	93.96	93.94	82.22	80.09	87.99	87.65
OctConv-ABiGRU	91.53	94.59	87.84	86.15	96.70	91.39

表 5 不同网络用时比较

Tab. 5 Comparison of detection time for different network modules

方法	参数量(百万)	模型大小(MB)	检测时间(s)
VGG19	45.12	344	41.56
DenseNet	57.22	437	24.91
CNN-ABiGRU	14.42	108	75.14
OctConv-ABiGRU	121.47	1382	36.80

表 6 不同方法的分类精度比较

Tab. 6 Comparison of PRE (%) for different network modules

方法	枪支	刀具	扳手	钳子	剪子	平均
VGG19	87.20	86.40	56.60	55.20	46.20	66.32
DenseNet	88.20	82.18	51.25	54.50	38.50	62.93
CNN-ABiGRU	88.50	87.20	63.00	61.20	76.40	75.26
OctConv-ABiGRU	86.78	92.22	77.44	76.22	94.56	85.44

此外，本文还比较了 DenseNet、VGG19、CNN-ABiGRU、OctConv-BiGRU 的参数数量、检测时间和 PRE，如表 5、表 6 所示。由表 5，表 6 可知本模型虽然网络结构较复杂，但是通过降低低频特征的分辨率，降低了计算成本，并且 PRE 大幅提升。检测时间较 DenseNet 增加了 11.89 s，平均 PRE 提高了 22.51%；检测时间较 VGG19、CNN-ABiGRU 分别减少了 4.76 s、38.34 s，平均 PRE 分别提高了 19.12% 和 10.18%。对于扳手，OctConv-ABiGRU 的 PRE 达到 77.44%，相比 DenseNet、VGG19、CNN-ABiGRU 分别提高了 26.19%、20.80%、14.44%；对于钳子，OctConv-ABiGRU 的 PRE 达到 76.22%，相比 DenseNet、VGG19、CNN-ABiGRU 分别提高了 21.72%、21.02%、15.02%。

4 结 论

为了提高 X 光安检图像的分类性能，本文提出八度卷积结合注意力机制双向 GRU 的算法，并对比分析了对于多种不同网络模型危险品的分类性能。对于 SIXRay 数据集，本文算法的平均 ACC、平均 AUC 和平均 PRE 分别达到了 98.73%、91.39% 和 85.44%。实验结果表明，本文模型能够加快学习速度，特定情况下减少了检测时间，提高了危险品分类准确率。未来工作着重对分类算法中加入集成学习，以进一步减少数据不平衡性对分类模型产生的影响。

参考文献：

[1] 陈志强, 张丽, 金鑫. X射线安全检查技术研究新进展[J]. 科学通报, 2017, 62(13): 1350-1365.
CHEN ZH Q, ZHANG L, JIN X. Recent progress on X-ray security inspection technologies[J]. *Chinese Science Bulletin*, 2017, 62(13): 1350-1365. (in Chinese)

[2] CAO S S, LIU Y H, SONG W W, *et al*.. Toward human-in-the-loop prohibited item detection in X-ray baggage images[C]. *Proceedings of 2019 Chinese Automation Congress (CAC)*, IEEE, 2019: 4360-4364.

[3] LYU SH J, TU X, LU Y. X-Ray image classification for parcel inspection in high-speed sorting line[C]. *Proceedings of the 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, IEEE, 2018: 1-5.

[4] 费彬, 孙京阳, 张俊举, 等. 基于稀疏处理的多能X射线分离成像[J]. 光学 精密工程, 2017, 25(4): 1106-1111.
FEI B, SUN J Y, ZHANG J J, *et al*.. Separation of multi-energy X-ray imaging based on sparse processing[J]. *Optics and Precision Engineering*, 2017, 25(4): 1106-1111. (in Chinese)

[5] 王旖旎. 基于Inception V3的图像状态分类技术[J]. 液晶与显示, 2020, 35(4): 389-394.
WANG Y N. Image classification technology based on inception V3[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(4): 389-394. (in Chinese)

[6] CHOUAI M, MERAH M, SANCHO-GOMEZ J L, *et al*.. Supervised feature learning by adversarial autoencoder approach for object classification in dual X-Ray image of luggage[J]. *Journal of Intelligent Manufacturing*, 2020, 31(5): 1101-1112.

[7] 张万征, 胡志坤, 李小龙. 基于LeNet-5的卷积神经图像识别算法[J]. 液晶与显示, 2020, 35(5): 486-490.

ZHANG W ZH, HU ZH K, LI X L. Convolutional neural image recognition algorithm based on LeNet-5[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(5): 486-490. (in Chinese)

[8] 刘恋秋. 基于深度卷积生成对抗网络的图像识别算法[J]. 液晶与显示, 2020, 35(4): 383-388.
LIU L Q. Image recognition algorithms based on deep convolution generative adversarial network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(4): 383-388. (in Chinese)

[9] 龚希, 吴亮, 谢忠, 等. 融合全局和局部深度特征的高分辨率遥感影像场景分类方法[J]. 光学学报, 2019, 39(3): 0301002.
GONG X, WU L, XIE ZH, *et al*.. Classification method of high-resolution remote sensing scenes based on fusion of global and local deep features[J]. *Acta Optica Sinica*, 2019, 39(3): 0301002. (in Chinese)

[10] 贠卫国, 史其琦, 王民. 基于深度卷积神经网络的多特征融合的手势识别[J]. 液晶与显示, 2019, 34(4): 417-422.
YUN W G, SHI Q Q, WANG M. Multi-feature fusion gesture recognition based on deep convolutional neural network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(4): 417-422. (in Chinese)

[11] LIU J Y, LENG X X, LIU Y. Deep convolutional neural network based object detector for X-Ray baggage security imagery[C]. *Proceedings of 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2019: 1757-1761.

[12] AKCAY S, KUNDEGORSKI M E, WILLCOCKS C G, *et al*.. Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(9): 2203-2215.

[13] ZHU Y, ZHANG H G, AN J Y, *et al*.. GAN-based data augmentation of prohibited item X-ray images in security inspection[J]. *Optoelectronics letters*, 2020, 16(3): 225-229.

[14] AKÇAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. Skip-GANomaly: skip connected and adversarially trained encoder-decoder anomaly detection[C]. *Proceedings of 2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2019.

[15] AYDIN I, KARAKOSE M, AKIN E. A new approach for baggage inspection by using deep convolutional neural networks[C]. *Proceedings of 2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, IEEE, 2018: 1-6.

[16] MERY D, SVEC E, ARIAS M, *et al*.. Modern computer vision techniques for X-Ray testing in baggage inspection[J]. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*: *Systems*, 2017, 47(4): 682-692.

[17] GALVEZ R L, DADIOS E P, BANDALA A A, et al.. Threat object classification in X-ray images using transfer learning[C]. *Proceedings of 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, IEEE, 2018: 1-5.

[18] HOWARD A G, ZHU M L, CHEN B, *et al*.. MobileNets: efficient convolutional neural networks for mobile vision applications[C]. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017.

[19] IANDOLA F N, HAN S, MOSKEWICZ M W, et al.. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[C]. *Proceedings of 2017 International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

[20] CHEN Y P, FAN H Q, XU B, *et al*.. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]. *Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, IEEE, 2019: 3434-3443.

[21] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al.. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, *Association for Computational Linguistics*, 2014: 1724-1734.

[22] 董潇潇, 何小海, 吴晓红, 等. 基于注意力掩模融合的目标检测算法[J]. 液晶与显示, 2019, 34(8): 825-833.
DONG X X, HE X H, WU X H, *et al*.. Object detection algorithm based on attention mask fusion[J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(8): 825-833. (in Chinese)

[23] MIAO C J, XIE L X, WAN F, *et al*.. SIXray: a large-scale security inspection X-ray benchmark for prohibited item

discovery in overlapping images[C]. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019: 2119-2128.

[24] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al*.. Rethinking the inception architecture for computer vision[C]. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016: 2818-2826.

[25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. *Proceedings of the 3rd International Conference on Learning Representations*, 2014.

[26] HE K M, ZHANG X Y, REN SH Q, et al.. Deep residual learning for image recognition[C]. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2016.

[27] HUANG G, LIU ZH, VAN DER MAATEN L, et al.. Densely connected convolutional networks[C]. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017.

[28] WANG A L, WANG M H, JIANG K Y, *et al*.. A novel lidar data classification algorithm combined densenet with STN[C]. *Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2019: 2483-2486.

作者简介：

吴海滨(1977—)，男，上海人，博士，教授，2002 年于哈尔滨工业大学获得硕士学位，2008 年于哈尔滨理工大学获得博士学位，现为哈尔滨理工大学测控技术与通信工程学院教授，主要从事机器视觉、医学虚拟现实、深度学习图像分类方面的研究。E-mail: woo@hrbust.edu.cn

王爱丽(1979—)，女，天津人，博士，副教授，2008 年于哈尔滨工业大学获得博士学位，现为哈尔滨理工大学测控技术与通信工程学院副教授，主要从事机器视觉、深度学习图像分类方面的研究。E-mail: aili925@hrbust.edu.cn