

利用卷积神经网络进行毫米波图像违禁物体定位

姚家雄^{1,2,3}, 杨明辉¹, 朱玉琨¹, 吴亮¹, 孙晓玮^{1*}

(1. 上海微系统与信息技术研究所 中科院太赫兹固态技术重点实验室, 上海 200050;

2. 中国科学院大学, 北京 100049;

3. 上海科技大学, 上海 201210)

摘要:随着毫米波器件的成熟,毫米波成像雷达已经应用于人体安检.但毫米波图像中违禁物体的定位仍然是一个艰巨的任务,这极大地限制了毫米波成像雷达的应用.文章将卷积神经网络(Convolutional Neural Network, CNN)应用于毫米波图像,自动定位毫米波图像中的违禁物体,如枪、刀等.利用滑动窗口在输入图像上滑动,并通过CNN得到各个子图块存在违禁物体的概率.图像块是相互重叠的,将各子图块的概率值累积起来,得到概率累积图.概率累积图反映了违禁物体的位置.由于CNN和概率累积图的应用,在实验中,该方法获得了很高的定位准确率,验证了该方法的有效性.

关键词:毫米波图像;卷积神经网络;物体定位;概率累积图

中图分类号:TP751 **文献标识码:**A

Using convolutional neural network to localize forbidden object in millimeter-wave image

YAO Jia-Xiong^{1,2}, YANG Ming-Hui¹, ZHU Yu-Kun¹, WU Liang¹, SUN Xiao-Wei^{1*}

(1. Key Laboratory of Terahertz Technology, Shanghai Institute of Microsystem and Information Technology, Shanghai 200050, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. ShanghaiTech University, Shanghai 201210, China)

Abstract: With the maturity of millimeter-wave devices, millimeter-wave imaging radar has been applied to human security check. However, the localization of forbidden objects in millimeter-wave images is still a difficult task, which greatly limits the application of millimeter-wave imaging radar. This paper adopts convolution neural network (CNN) to automatically localize forbidden objects, such as guns and knives, in millimeter-wave images. A sliding window is applied to slide over the input image. Then the probability of the existence of forbidden object in the image patch can be obtained via CNN. The image patches are overlapped with each other, and the probability values of all image patches are accumulated to obtain the probability accumulation map (PA-map). The PA-map reflects the position of forbidden objects. Due to the application of CNN and PA-map, this method achieves a high accuracy of localization in the experiment, which verifies the effectiveness of this method.

Key words: millimeter-wave image, convolution neural network(CNN), object localization, probability accumulation map

PACS: 84.40.Xb

引言

近年来毫米波技术已被应用于人体安检成像系

统,有逐步取代传统安检技术的趋势.毫米波辐射属于非电离辐射,对人体无害,且具有穿透衣物的能力,因此可以应用于人体成像,检测隐藏在衣服下面

收稿日期:2016-11-17,修回日期:2017-01-08

Received date: 2016-11-17, revised date: 2017-01-08

基金项目:国家自然科学基金(61671439)

Foundation items: Supported by the National Natural Science Foundation of China (61671439)

作者简介(Biography): 姚家雄(1992-),男,广西桂林人,硕士研究生,主要从事毫米波图像自动目标检测的研究. E-mail: yaojx@mail.sim.ac.cn

* 通讯作者(Corresponding author): E-mail: xwsun@mail.sim.ac.cn

的违禁物体.毫米波成像雷达和成像算法已经成熟,并且已经实现商用^[13].但针对毫米波图像的违禁物体定位仍然是一个亟待解决的问题.鉴于毫米波硬件的不稳定性,毫米波图像往往存在较大的本底噪声,物体的纹理非常模糊.而违禁物品种类繁多,各类物体纹理复杂,难以提取有效的特征.这使得毫米波图像的物体定位成为一项艰巨的任务.

PNNL(Pacific Northwest National Laboratory)将图像划分为许多重叠的网格,并在每个网格进行识别,最后合并那些存在物体的网格^[4].该论文依据的是自然图像频谱遵从 $1/f$ 定律,而当某个网格中存在人造物体时,网格频谱不服从 $1/f$ 定律,因此该方法以每个网格的二维傅里叶频谱作为特征,然后使用神经网络进行识别.但是这种方法有两个缺点:(1)对于质量较差的毫米波图像,图像本身就不遵从 $1/f$ 定律,因此二维频谱不是鲁棒的特征;(2)由于最终定位结果是几个网格取并集,因此定位结果不精确.L3公司的毫米波雷达安检系统已经实现商用,其物体定位算法可以针对人体四肢和躯干分别进行识别^[5],并具有较好的物体定位效果.L3的毫米波雷达系统硬件稳定性好,成像质量高,这也降低了物体定位的难度,而本文所用毫米波图像存在较大噪声,定位难度更大.

CNN是一种强大的特征提取方法^[6],并且在许多计算机视觉任务中达到了最好的性能^[7-9].通过CNN,ImageNet数据集^[10]中的分类错误率在2012年降低到15.3%^[11],压倒了基于传统手工特征如SIFT(Scale Invariant Feature Transform)和LBP(Local Binary Pattern)的方法.2015年,ImageNet数据集的识别误差通过CNN降低到了3.57%^[12].考虑到CNN的优异的特征提取能力,本文采用CNN对毫米波图像进行物体定位.我们在实验中达到93.18%的定位准确率,验证了这种方法的有效性.

1 卷积神经网络基本原理

获得高性能CNN的关键步骤是CNN模型的设计和训练.如图1所示,通常CNN是由多个卷积层和池化层顺序堆叠,最后加上全连接层而得.卷积层把输入图像或特征图与卷积核做卷积,得到多个特征图,池化层对特征图做下采样,如此循环往复,直到特征图的尺寸变成 1×1 ,此时便可以得到图像的特征向量.各层特征图对应的特征是不一样的,第一层特征图对应图像的颜色或灰度值特征,第二层对应边沿特征,第三层对应几何纹理特征^[13].深层

特征图是对浅层特征图的组合,可以表达更为复杂和抽象的图像特征,因此网络的深度至关重要^[14].最后得到的特征向量,则包含了图像的颜色/灰度值特征、边沿特征、几何纹理特征等.决定网络性能的除了网络深度,还有网络的宽度,即每层的特征图数量^[15],因此特征图的数量通常会逐层增加.特征图的具体数量需要结合具体应用以及训练样本数量决定,增加特征图的数量可以提高特征提取能力,但也容易引起过拟合.在训练样本足够多的情况下,可以适当增加特征图的数量,若训练样本不足,则应减少特征图数量.卷积层和池化层的作用是提取特征,全连接层则根据具体任务的不同,可以用于分类或回归.全连接层包含了向量和矩阵的乘积运算,会引入大量的可训练参数,在缺乏足够训练样本的前提下,太多的可训练参数容易导致过拟合.为此,省略了全连接层,直接用逻辑回归对样本做分类.

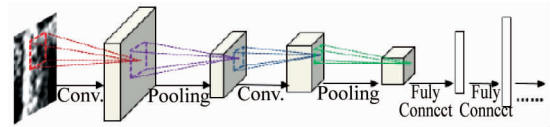


图1 CNN模型一般结构
Fig.1 Common architecture of CNN model

要使CNN具备分类识别能力,必须先使用训练样本进行训练,卷积核等网络参数也需要通过训练获得.在训练前,需要将训练样本正确分类并标记好类别.训练时,将训练样本和类别标签一起输入到网络中,通过梯度下降法不断优化卷积核等网络参数,直到网络的目标函数收敛为止,此时网络就具备了自动提取图像特征并分类识别的能力,从而可以用在新样本的分类识别任务中.

本文的CNN模型由四个卷积层组成,如图2所示.在每个卷积层之后,有一个非线性激活函数.通过CNN提取图像特征,实际上是通过一个函数将图像映射成一个特征向量.因为图像和特征向量之间往往存在非线性映射关系,所以需要在网络中加入非线性激活函数,使网络能够学习到非线性特征变换的能力.由于ReLU(Rectified Linear Unit)具有避免梯度消失的良好特性,本文将其用作非线性激活函数.将卷积核与输入图像或上一层的特征图进行卷积,然后将卷积结果与一个常数偏置相加,便可以得到下一层的特征图.同一个特征图中的所有隐藏节点共享相同的卷积核和偏置,从而使得可训练参数的数量大大减少.池化层中的特征图的数量与卷

积层中的特征图的数量相同,池化层相当于对特征图做下采样.通常使用的池化操作有平均池化和最大值池化.接下来对 CNN 中的操作做详细说明.

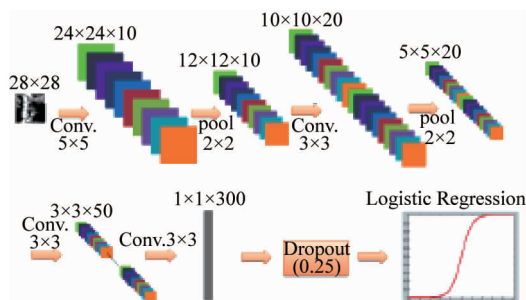


图2 本文的 CNN 模型

Fig. 2 CNN model of this paper

1.1 卷积

在卷积层中,每个输出特征图与上一层的所有特征图连接.令 $\mathbf{u}_i^l (i=1,2,\dots,N_l)$ 表示第 l 层的第 i 个特征图, N_l 表示第 l 层中特征图的数量. $\mathbf{u}_i^l(x,y)$ 表示特征图中坐标位置为 (x,y) 的隐藏节点,那么,每个隐藏节点的响应计算如下:

$$\begin{cases} \mathbf{u}_j^{l+1}(x,y) = f(\mathbf{v}_j^{l+1}(x,y)) \\ \mathbf{v}_j^{l+1}(x,y) = \sum_{i=1}^{N_l} \sum_{p,q=0}^{K-1} \mathbf{k}_{i,j}(p,q) \cdot \mathbf{u}_i^l(xS+p,yS+q) + b_j^{l+1} \end{cases}, \quad (1)$$

$f(x)$ 是非线性激活函数; $\mathbf{k}_{i,j}$ 是卷积核,连接 \mathbf{u}_i^l 和 \mathbf{u}_j^{l+1} 两个特征图.通常来说,浅层的卷积核类似于边缘提取算子,而深层的卷积核则用于提取复杂和抽象的纹理特征^[13].卷积核的大小是 $K \times K$, S 是卷积核的滑动步长,需要人为设定,且小于 K .通常可以设置较小的滑动步长,例如设为 1. b_j^{l+1} 是 \mathbf{u}_j^{l+1} 的偏置,也是通过训练获得.

1.2 非线性激活函数

CNN 提取特征的过程,相当于用一个函数,将输入图像映射成一个高维特征向量.由于图像纹理特征的复杂性,这一映射关系往往是高度非线性的.通过在卷积层之后添加非线性激活函数,可以使 CNN 获得非线性建模能力,从而更好地提取图像特征.传统的非线性激活函数有双曲正切:

$$f(x) = \tanh(x), \quad (2)$$

和 sigmoid 函数:

$$f(x) = 1/(1 + e^{-x}). \quad (3)$$

当 $|x| \gg 1$ 时,这两个函数的梯度非常小.反向传播算法基于链式法则,如果有多个梯度值较小,则总体梯度值就会非常小,从而导致梯度消失效应,导

致网络参数难以训练.本文使用 ReLU 作为非线性激活函数:

$$f(x) = \max(0, x), \quad (4)$$

当 $x < 0$ 时,梯度为 0,网络参数不需要更新.当 $x > 0$ 时,梯度等于 1.因此,ReLU 不会引起梯度消失效应.并且,由于 ReLU 计算非常简单,在 CNN 网络前向计算中具有更快的速度.

1.3 池化

通过卷积获得的特征图总是具有较高的维度,这既增加了计算复杂度,又容易导致过拟合.通过池化操作,我们可以获得主要特征,同时减少特征维度.池化操作单独作用于每个特征图.平均池化和最大值池化的公式如下所示:

$$\mathbf{p}_i^l(x,y) = \text{mean}\{\mathbf{u}_i^l(p,q) \mid xS \leq p \leq xS + K - 1, yS \leq q \leq yS + K - 1\}, \quad (5)$$

$$\mathbf{p}_i^l(x,y) = \max\{\mathbf{u}_i^l(p,q) \mid xS \leq p \leq xS + K - 1, yS \leq q \leq yS + K - 1\}, \quad (5)$$

\mathbf{p}_i^l 是对特征图 \mathbf{u}_i^l 进行池化的结果. K 表示在特征图上 $K \times K$ 邻域内进行池化. S 表示池化窗口的滑动步长,需要人为设定,且小于 K .

1.4 Dropout

Dropout 是一种有效的正则化技术,它是由 Hinton 提出的^[16].在训练过程中,以概率 p 随机地将一些隐藏节点设为零,并保留其他隐藏节点.因此,在每次迭代过程中我们都得到一个简化的模型.由于隐藏节点是随机设置为零的,每次迭代得到的简化模型都是全新的.在测试阶段,我们保留所有隐藏节点,并让所有隐藏节点乘以因子 $1-p$.这相当于获得所有简化模型的平均模型,达到集成学习^[17]的目的,从而提高性能.

1.5 逻辑回归

逻辑回归是一种线性分类器,能够实现二分类.它利用 sigmoid 函数将样本的特征向量 \mathbf{x} 映射成一个概率值 θ .如果 $\theta \geq 0.5$,那么样本属于类别 1,否则样本属于类别 0.假设逻辑回归的权值和偏置分别为 $\bar{\mathbf{w}}$ 和 b ,那么概率计算公式如下:

$$\theta(x) = \frac{1}{1 + e^{-(\bar{\mathbf{w}} \cdot \mathbf{x} + b)}}. \quad (7)$$

一个好的逻辑回归,应该能将尽量多的样本分类正确.也就是说,所有样本的联合概率应该最大.这相当于做极大似然估计.假设特征向量的类别标签表示为 y ,且 $y \in \{0,1\}$,则样本的条件概率为

$$p(y|x) = \begin{cases} \theta(x) & y = 1 \\ 1 - \theta(x) & y = 0 \end{cases}, \quad (8)$$

即 $p(y|x) = \theta(x)^y (1 - \theta(x))^{1-y}$. 假设各样本相互独立, 则联合概率为:

$$P(Y|X) = \prod_{i=1}^N P(y^{(i)} | x^{(i)}) \\ = \prod_{i=1}^N \theta(x^{(i)})^{y^{(i)}} (1 - \theta(x^{(i)}))^{1-y^{(i)}} \quad (9)$$

负对数似然函数为:

$$NLL(\theta(\vec{w}, b)) = -\ln(P(Y|X)) = \\ -\sum_{i=1}^N y^{(i)} \ln(\theta(x^{(i)})) + (1 - y^{(i)}) \ln(1 - \theta(x^{(i)})) \quad (10)$$

式(10)即为逻辑回归的损失函数, 称为交叉熵(cross-entropy)损失函数. 不断更新参数 \vec{w} 和 b , 直到损失函数收敛到极小值, 此时便可得到一个最佳的逻辑回归分类器. 参数更新的方法是梯度下降法(Gradient Descent, GD).

1.6 网络训练

训练的目的在于, 通过不断迭代更新网络参数, 使得损失函数收敛于全局极小值. 参数的更新过程是通过梯度下降法实现的, 即:

$$\vec{w}_{k+1} = \vec{w}_k - \alpha \cdot \nabla L(\vec{w}) \big|_{\vec{w}=\vec{w}_k} \quad (11)$$

这里的 \vec{w} 不仅仅是逻辑回归的权值, 它包括了 CNN 网络中所有的可训练参数. k 表示第 k 次迭代, \vec{w}_0 表示网络参数的初始值. 网络参数可以通过均值为 0 的高斯分布随机初始化^[13]. α 是学习率, 是一个介于 0 和 1 之间的常数. $\nabla L(\vec{w})$ 是损失函数的梯度, 通过反向传播算法(Back-propagation, BP)计算得到.

反向传播算法: 某一层的输出就是下一层的输入, 如图 3 所示. 因此, CNN 的整体结构相当于一系列传递函数的嵌套. 梯度可以通过链式法则求得. 第 l 层网络参数的梯度为:

$$\frac{\partial L(\vec{w})}{\partial \vec{w}} = \frac{\partial L(\vec{w})}{\partial \theta} \cdot \frac{\partial \theta}{\partial x_m} \dots \frac{\partial x_m}{\partial x_{m-1}} \dots \frac{\partial x_{l+1}}{\partial \vec{w}_l} \quad (12)$$

如果等式(12)右边某几项的值很小, 就会导致总体梯度值很小. 根据式(11)可知, 此时权值的更新会变得很慢, 这也就是梯度消失效应. 假如 CNN 网络中使用的非线性激活函数为 tanh 或者 sigmoid, 则必然导致总体梯度非常小, 从而引起梯度消失. 这也就是本文采用 ReLU 而不用 tanh 和 sigmoid 的原因.

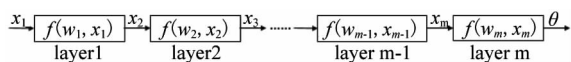


图3 CNN的分层结构

Fig.3 The layer-wise structure of CNN

随机梯度下降法: 对于传统的梯度下降法, 需要输入所有训练样本以完成一次迭代(iteration). 但对于随机梯度下降法(Stochastic Gradient Descent, SGD), 只需要随机选择一个数据批(batch)的样本来完成一次迭代, 即每次迭代时只随机选取一部分样本. 当所有样本都参与了一次迭代, 即完成了一个迭代周期(epoch). 当使用所有训练样本时, 可以计算真实的梯度. 但事实上, 真实的梯度不是必须的. 可以使用数据批来得到梯度的估计值, 然后更新参数. 数据批较大时, 我们可以得到更精确的梯度估计, 但这将降低收敛的速度. 较小的数据批有助于加速收敛, 但是梯度的估计不精确. 如果数据批太小, 梯度的估计值将产生振荡, 从而降低收敛速度. 通常, 数据批的大小可以设置为 100、128、256 等.

2 毫米波图像物体定位方法

本文的物体定位方法包含两个模块. 第一个模块是一个二分类器, 它将毫米波图像的子图块作为输入, 用 CNN 提取特征, 并计算出每个子图块存在物体的概率, 从而将子图块分类为含物体子图和不含物体子图. 第二个模块是将这些子图块的概率值做累积叠加, 获得概率累积图. 一个像素的概率累积值越大, 则该像素越可能包含于一个物体中. 之后我们可以从概率累积值中获得物体的位置和大小. 在本节中, 我们将详细介绍该定位方法.

本文使用的 CNN 模型已经在图 2 中给出. 它包含四个卷积层和两个池化层. 第一个卷积层的卷积核大小为 5×5 , 输出特征图大小为 24×24 , 数量为 10, 之后在特征图 2×2 领域上做最大值池化; 第二个卷积层的卷积核大小为 3×3 , 输出特征图大小为 10×10 , 数量为 20, 之后在特征图 2×2 领域上做最大值池化; 第三个卷积层的卷积核大小为 3×3 , 输出特征图大小为 3×3 , 数量为 50; 最后一个卷积层的卷积核大小为 3×3 , 输出一个 300 维的特征向量, 随后通过 dropout 防止过拟合, 再用逻辑回归计算概率. 网络的输入是尺寸为 28×28 的子图块.

使用滑动窗口从图像中截取训练样本, 滑动窗口尺寸为 28×28 , 滑动步长为 2. 当在图像中进行定位时, 我们仍然将滑动步长设置为 2. 由于滑动步长小于 28, 因此子图块之间相互交叠. 我们将各子图块的概率值累积叠加, 得到概率累积图. 具有较大概率累积值 TH_{prob} 的像素点更可能包含于物体中. 采用硬阈值的方法, 将概率累积值大于阈值的像素点

设为 1,其他像素点设置为 0,所获得的二进制图像代表着潜在物体的位置. 该步骤称为区域提取. 为了进一步减少虚警和提高定位精度,我们进行一个区域裁剪操作. 对于二进制图像中的每个连通区域,我们计算出一个自适应阈值. 计算公式如下:

$$TH_{\text{adap}} = \min(TH_{\text{up}}, a \times PA_{\text{max}} + b \times S), \quad (13)$$

PA_{max} 是连通区域的最大概率累积值, S 是面积. 常数 a 和 b 需要微调,以获得最佳定位准确率. 自适应阈值不得超过上限值 TH_{up} . 概率累积值小于 TH_{adap} 的所有像素点都被去除,剩余的像素点就是物体的准确位置. 最后,通过面积阈值去噪,将面积过小的连通区域去除. 这些区域可以被视为虚警或者噪声.

3 实验结果

实验中的毫米波图像通过毫米波全息成像系统 SimImage 采集,该雷达由上海微系统与信息技术研究所研发^[2]. 我们收集了 440 张图像来验证本文的毫米波图像违禁物体定位方法,其中 40 张图中人体未携带物体,另外 400 张人体各携带一件物体. 图像中包含了三种违禁物体,分别为手枪、陶瓷刀和手机,如图 4 所示. 这些物体以不同的角度放置在衣物里. 对于每个物体,我们用一个尽量小的矩形框将其标出,该数据标记称作地面实况 (ground-truth). 在本节中,我们将首先介绍如何训练 CNN 模型,然后展示定位结果.

3.1 CNN 模型训练

训练样本从毫米波图像中截取. 我们使用滑动窗口从毫米波图像中截取训练所需的子图块. 滑动窗口尺寸为 28×28 ,滑动步长为 2. 对于每个子图块,我们根据以下准则将其分为正样本或者负样本:

正样本: 方差大于阈值 TH_{var} , IoU 大于阈值 TH_{iou} ;

负样本: 方差大于阈值 TH_{var} , IoU 小于阈值 TH_{iou} .

如果子图块的方差太小,它必然是背景,因为背景区域的灰度值比较均匀,而人体区域的灰度值变化剧烈. 背景子图不需要进行识别,所以我们设置一个阈值 TH_{var} 来去除这类子图块. 本文将 TH_{var} 设置为 200. IoU (Intersection over Union) 表明了子图块和 ground-truth 之间的重叠程度. IoU 计算公式如下:

$$IoU = (A \cap B) / (A \cup B), \quad (14)$$

其中 A 为子图块, B 为 ground-truth. 我们尝试了不同的 TH_{iou} 值,从 0.05 到 0.3. 结果表明, TH_{iou} 取 0.05 时具有最佳的定位准确率. 因为某些物体,比如竖直

放置的刀,其 ground-truth 是细长的矩形,所以子图块与 ground-truth 的 IoU 必然很小. 如果 TH_{iou} 设置为一个较大的值,那么我们无法裁剪出刀的正样本,导致逻辑回归分类器失去识别刀的能力.

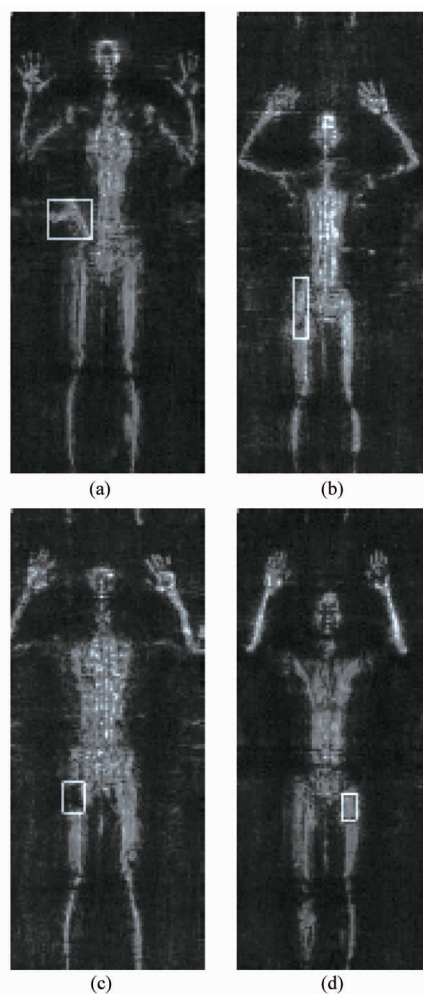


图 4 毫米波图像违禁物体, (a) 手枪、(b) 陶瓷刀、(c) (d) 手机

Fig. 4 Forbidden objects in millimeter-wave images, (a) is gun, (b) is ceramic knife, (c) and (d) are cellphones

为了训练二分类器,我们随机选取 70% 的图像 (308 张) 来截取训练样本和测试样本. 其余 132 幅图像留作测试定位准确率. 数据增强是扩充训练集的常用手段^[11]. 由于物体可以按任意角度放置,所以我们对所有正样本做水平翻转和垂直翻转. 这可以将训练样本增加到原来的 4 倍. 由于截取的负样本子图远远多于正样本子图,我们对负样本子图做随机下采样,使正负样本数量相等. 最后,我们得到 105 万样本子图. 随机选择 80% 样本作为训练集, 10% 作为测试集, 10% 作为验证集.

我们使用 SGD 来训练 CNN 模型,数据批的大

小设置为 100,初始学习率为 0.001,从第 20 个迭代周期开始,每 10 个迭代周期学习速率降为原来的 0.5 倍.我们总共训练了 80 个迭代周期,这时损失函数和分类误差收敛到最小值.收敛后,测试样本中的分类误差仅为 2.03%.

3.2 实验结果分析

在区域提取这一步,经过调试,我们将阈值 TH_{prob} 设置为 100,该值大约相当于概率值 0.51.因为滑动窗口大小是 28×28 ,滑动步长是 2,则每个像素点将被 196 个子图块($14 \times 14 = 196$)覆盖,这意味着每个像素点的最大概率累积值是 196.如果一个像素的概率累积值是 100,这意味着该像素点包含于物体中的概率为 $100/196 \approx 0.51$.在区域裁剪中,常数 a 设置为 1.0, b 设置为 -0.02,上限值 TH_{up} 设置为 150.之后,通过面积阈值去噪,将面积小于 40 个像素的连通区域去除,剩下的连通区域用尽量小的矩形框标出,作为识别结果.整体定位流程如图 5 所示.

本文总共用了 132 张图像测试毫米波图像违禁物体定位方法的有效性,其中.当定位结果满足以下三个判别准则时,则认为该图像被正确定位.

- (1) 图像中没有虚警;
- (2) 图像中没有漏警;
- (3) 定位结果矩形框和 ground-truth 的 IoU 大于

0.3.

总共有 123 张图像满足以上三个判别准则,这意味着定位准确率为 93.18% (在训练图像中定位准确率为 97.73%),这验证了本文毫米波图像违禁物体定位方法的有效性.图 6 展示了更多毫米波图像物体定位的结果.图中物体的放置姿态或放置位置有所不同,但均能定位准确.图 6(g-h) 人体未携带违禁物品,结果也没有产生虚警.表 1 列出了详细的定位结果,其中图像 1 代表人体携带有违禁物品,图像 2 代表人体没有携带违禁物品.表中第二行准确数 + 虚警数 + 漏警数 > 图片数,是因为有一张图片同时出现了虚警和漏警.

两项技术的应用使得本文取得了高定位准确率.第一项是 CNN. CNN 是一种提取图像特征的有力手段,使得本文在第一个模块,即逻辑回归模块获得了高分类准确率.第二项是概率累积图.如果一个像素点包含于一个物体,那么它周围的大多数子图块都具有较大概率值,使得这个像素获得一个较大的概率累积值.概率累积图使得本文获得了像素级

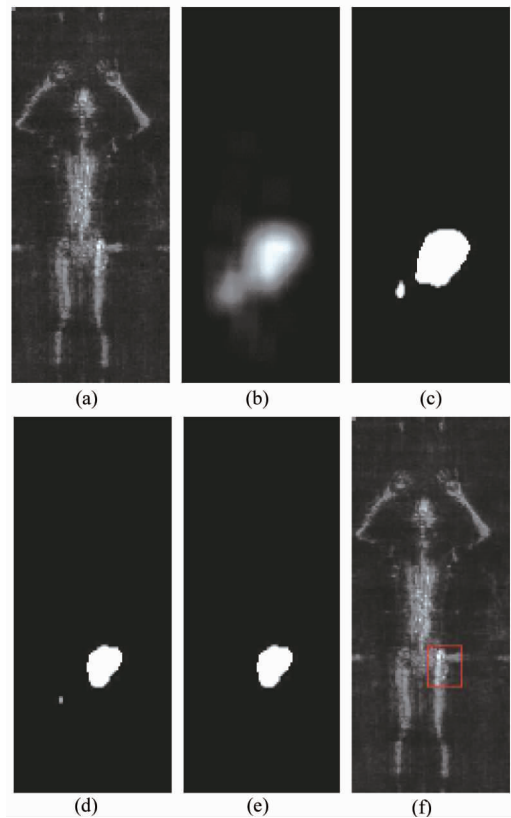


图 5 物体定位流程:(a)输入图像,(b)概率累积图,(c)区域提取,(d)区域裁剪,(e)面积阈值去噪,(f)定位结果

Fig.5 Process of object localization, (a) input image, (b) PA-map, (c) region proposal, (d) region cut, (e) denoise with area opening, (f) localization result

别的定位能力.

表 1 定位结果列表

Table 1 List of localization results

	测试图片数	定位正确	虚警	漏警
图像 1	114	106(92.98%)	4(3.51%)	5(4.39%)
图像 2	18	17(94.44%)	1(5.56%)	0(0%)
总计	132	123(93.18%)	5(3.79%)	5(3.79%)

尽管定位准确率很高,但这是基于一个相对粗略的标准,即上述判别准则(3)中 $IoU > 0.3$.当修改为 $IoU > 0.5$ 时,定位准确率降低到 71.2%.并且,由于滑动窗口的尺寸是固定值,当物体尺寸与滑动窗口尺寸相近时,该方法定位效果较好,当物体尺寸变化较大时,定位效果会下降.在未来的工作中,我们将探索如何更精确地定位不同尺寸的物体.此外,在以后的工作中,我们希望在实现物体定位的基础上,进一步实现物体的分类识别.

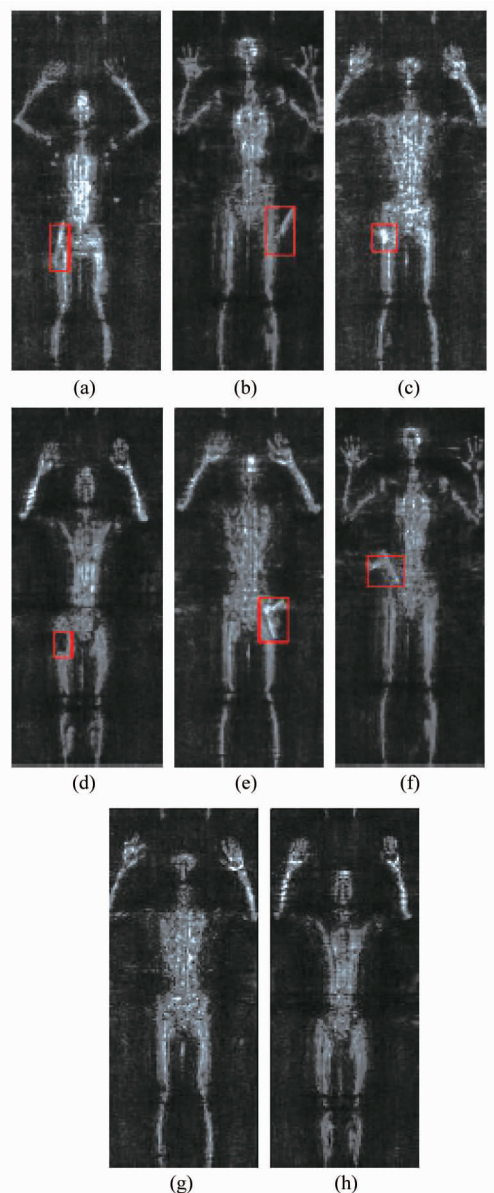


图6 物体定位结果,(a)(b)为陶瓷刀,(c)(d)为手机,(e)(f)为手枪,(g)(h)人体未携带物品

Fig. 6 Object localization results: (a) and (b) are ceramic knives, (c) and (d) are cellphones, (e) and (f) are guns, (g) and (h) no object on human body

4 结论

提出了一种基于 CNN 模型的方法来实现毫米波图像违禁物品定位. 本文使用滑动窗口在毫米波图像上滑动, 并通过 CNN 获得每个子图块存在物体的概率. 将各子图块的概率值累积叠加, 得到概率累积图. 通过硬阈值做区域提取, 自适应阈值法做区域裁剪, 最后用面积阈值去噪法去除虚警, 得到物体最终位置. 实验结果验证了这种方法的有效性和可行

性. 在未来的工作中, 我们希望能够进一步提高定位的精度, 实现对任意尺寸物体的定位, 并且能够通过算法识别出物体的类别信息.

致谢

感谢复旦大学信息科学与工程学院王斌教授、张立明教授和吴晓峰老师的指导.

References

- [1] Sheen D M, McMakin D L, Hall T E. Three-dimensional millimeter-wave imaging for concealed weapon detection [J]. *IEEE Transactions on Microwave Theory & Techniques*, 2001, **49**(9): 1581–1592.
- [2] Yu-Kun Zhu, M-H Y, Liang Wu, et al. Millimeter-wave holographic imaging algorithm with amplitude corrections [J]. *Progress In Electromagnetics Research*, 2016, **49**: 33–39.
- [3] FEI Peng, FANG Wei-hai, WEN Xin, et al. State of the art and future prospect of the active millimeter wave imaging technique for personnel screening [J]. *Journal of Microwaves*, 2015, **31**(2): 91–96.
- [4] Keller P E, McMakin D L, Hall T E, et al. International Joint Conference on Neural Networks, 2006: 2009–2014.
- [5] Communications L.
- [6] LéCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [J]. *Computer Science*, 2013: 580–587.
- [8] 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*, 2016: 1–1.
- [9] Sermanet P, Eigen D, Zhang X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks [J]. *Eprint Arxiv*, 2013.
- [10] Deng J, Dong W, Socher R, et al., 2009: 248–255.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. *Advances in Neural Information Processing Systems*, 2012, **25**(2): 2012.
- [12] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [J]. *Computer Science*, 2015.
- [13] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks: Springer International Publishing, 2013.
- [14] Lin M, Chen Q, Yan S. Network In Network [J]. *Computer Science*, 2014.
- [15] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [J]. 2014: 1–9.
- [16] Nair V, Hinton G E. International Conference on Machine Learning, 2010: 807–814.
- [17] Breiman L. Random Forest [J]. *Machine Learning*, 2001, **45**: 5–32.