改进的 YOLO 特征提取算法及其在服务机器人隐私情境检测中的应用

杨观赐 [1] 杨 静 [1] 苏志东 [1] 陈占杰 [1]

摘 要 为了提高 YOLO 识别较小目标的能力, 解决其在特征提取过程中的信息丢失问题, 提出改进的 YOLO 特征提取算法. 将目标检测方法 DPM 与 R-FCN 融入到 YOLO 中, 设计一种改进的神经网络结构, 包含一个全连接层以及先池化再卷积的特征提取模式以减少特征信息的丢失. 然后, 设计基于 RPN 的滑动窗口合并算法, 进而形成基于改进 YOLO 的特征提取算法. 搭建服务机器人情境检测平台, 给出服务机器人情境检测的总体工作流程. 设计家居环境下的六类情境, 建立训练数据集、验证数据集和 4 类测试数据集. 测试分析训练步骤与预测概率估计值、学习率与识别准确性之间的关系, 找出了适合所提出算法的训练步骤与学习率的经验值. 测试结果表明: 所提出的算法隐私情境检测准确率为 94.48 %, 有较强的识别鲁棒性. 最后, 与 YOLO 算法的比较结果表明, 本文算法在识别准确率方面优于 YOLO 算法.

关键词 YOLO, 特征提取算法, 服务机器人, 隐私情境检测, 智能家居

引用格式 杨观赐, 杨静, 苏志东, 陈占杰. 改进的 YOLO 特征提取算法及其在服务机器人隐私情境检测中的应用. 自动化学报, 2018, **44**(12): 2238−2249

DOI 10.16383/j.aas.2018.c170265

An Improved YOLO Feature Extraction Algorithm and Its Application to Privacy Situation Detection of Social Robots

YANG Guan-Ci[1] YANG Jing[1] SU Zhi-Dong[1] CHEN Zhan-Jie[1]

Abstract To address the limitation of YOLO algorithm in recognizing small objects and information loss during feature extraction, we propose FYOLO, an improved feature extraction algorithm based on YOLO. The algorithm uses a novel neural network structure inspired by the deformable parts model (DPM) and region-based fully convolutional networks (R-FCN). A sliding window merging algorithm based on region proposal networks (RPN) is then combined with the neural network to form the FYOLO algorithm. To evaluate the performance of the proposed algorithm, we develop a social robot platform for privacy situation detection. We consider six types of situations in a smart home and prepare three datasets including training dataset, validation dataset, and test dataset. Experimental parameters such as training step and learning rate are set in terms of their relationships with the prediction accuracy. Extensive privacy situation detection experiments on the social robot show that FYOLO is capable of recognizing privacy situations with an accuracy of 94.48 %, indicating the good robustness of our FYOLO algorithm. Finally, the comparison results between FYOLO and YOLO show that the proposed FYOLO outperforms YOLO in recognition accuracy.

Key words YOLO, feature extraction algorithm, social robot, detection of privacy situations, smart homes

Citation Yang Guan-Ci, Yang Jing, Su Zhi-Dong, Chen Zhan-Jie. An improved YOLO feature extraction algorithm and its application to privacy situation detection of social robots. *Acta Automatica Sinica*, 2018, **44**(12): 2238−2249

收稿日期 2017-05-15 录用日期 2017-08-29
Manuscript received May 15, 2017; accepted August 29, 2017
国家自然科学基金 (61863005, 61640209), 贵州省科技计划项目 (黔科合人字 (2015) 13, 黔科合 JZ 字 [2014] 2004, 黔科合 LH 字 [2016] 7433, 黔科合平台人才 [2018] 5702), 贵州省教育厅研究生教改重点课题 (黔教研合 JG 字 [2015] 002) 资助
Supported by National Natural Science Foundation of China (61863005, 61640209), Science and Technology Foundation of Guizhou Province ((2015) 13, JZ [2014] 2004, LH [2016] 7433, PTRC [2018] 5702), and Graduate Education Reform Fund of Education Bureau of Guizhou Province (JG [2015] 002)
本文责任编委 胡清华
Recommended by Associate Editor HU Qing-Hua
1. 贵州大学现代制造技术教育部重点实验室 贵阳 550025
1. Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University, Guiyang 550025

越来越多的智能家居系统和服务机器人广泛使用摄像头, 这会引入隐私泄漏风险, 是阻碍此类系统推广的最大障碍之一[1]. 前期问卷调查表明[2], 对隐私内容有符合人心理需求反应的系统, 可改善用户体验感受, 如何识别与保护具有视觉设备的服务机器人的隐私数据是值得研究的问题. Arabo 等[3] 设计了一种智能家居环境中隐私与安全框架, Kozlov 等[4] 通过分析智能家居环境下各系统间的安全和隐私与互信风险, 研究了高度依赖法律支持的隐私控

制机制、隐私风险分级方法. 这些研究主要从数据访问控制等角度考虑信息安全, 没有提出数据获取阶段的敏感数据识别与保护方案. Denning 等[5] 指出, 即使智能服务机器人使用加密和认证方式, 网络攻击者也有机会控制机器人或提取敏感数据.

在学术界, 图像特征提取方法是研究的热点[6]. 文献 [7] 通过映射聚合层中各个点的值为 Block 中各个点的激活概率均值, 得到一种均值聚合机制. 虽然此方法的准确率高于基于聚合层进行图像特征提取方法[8], 但是特征提取过程复杂, 模型训练时间较长. 文献 [9] 通过定义新的结构元和自适应向量融合模型, 提出一种加权量化方法自适应融合图像目标和背景. 当图像背景与目标均较大时, 该方法能体现图像全局特征的相关性, 但当目标较小时相关性表征变得困难. 文献 [10] 采用 Gabor 滤波器[11] 和局部模式分析来提取特征, 虽然该方法可以获得较多的灰度图像特征, 但在图像预处理和测试阶段需要将图像归一化大小, 不能检测随机大小的图片. 此类特征提取方法对小规模数据集的特征提取具有很好的表现能力, 但对海量数据, 特别是复杂背景环境下的数据, 其特征提取能力有待进一步提高.

YOLO (You only look once: unified, real-time object detection)[12] 是一种基于卷积神经网络的目标实时检测模型, 因其具有海量数据的学习能力、点对点的特征提取能力以及良好的实时识别效果而备受关注[13−14]. 文献 [15] 通过使用高斯混合模型模拟背景特征, 提出基于高斯混合模型和 YOLO 的行人检测算法, 在检测变电站监控视频中的行人时取得良好效果. 文献 [16] 利用交替方向乘子法[17] 提取灰度图像上下文信息特征, 并将该信息组合成一个 2D 输入通道作为 YOLO 神经网络模型的输入, 形成了基于 YOLO 的实时目标检测算法, 虽然识别精度有所提高, 但是模型的时间开销较大. 文献 [18] 设计了提取图像内文本字符的机制, 并采用 YOLO 进行文本检测和边界框回归. 文献 [19] 评估了目标检测算法 YOLO、Faster-RCNN[20]、霍夫森林[21] 的性能, 并指出 YOLO 在检测速度和识别精度上都要高于两种比较算法. 上述这些研究就提高 YOLO 的性能、拓展其应用等方面做了许多工作, 但是采用 YOLO 神经网络解决图像的特征提取问题时, 存在以下不足[22−23]:

1) 在识别的过程中, YOLO 将需要识别的图像分割为 7×7 的网格, 单元格内用于预测目标的神经元可以属于若干个属于同一类别的滑动窗口, 这使得模型具有很强的空间约束性. 若滑动窗口内涵盖多个不同类别的对象时, 系统无法同时检测出全部的目标对象.

2) 在训练过程中对数据集特征提取, 网络中的单元格最多负责预测一个真实目标, 这导致 YOLO 检测相对靠近且较小的目标时效果欠佳.

3) 在图像预处理阶段, YOLO 将训练数据集的高分辨率图像处理为低分辨率数据并用于最终的分类特征的提取. 经过多次卷积后, 原始图片分布区域中的小目标特征难以保存.

使用服务机器人引起的道德问题没有被充分考虑, 伦理原则应该体现到服务机器人的研发中[24]. 课题组在研发服务机器人时, 采用 YOLO 识别家庭环境中不同情境. 为了提高 YOLO 神经网络对较小目标的识别能力, 解决其在特征提取过程中信息丢失的问题, 本文提出了改进的 YOLO 特征提取算法, 并将其应用于服务机器人隐私情境检测.

1　目标实时检测模型 YOLO

目标实时检测模型 YOLO[12] 包括 18 个卷积层、2 个全连接层和 6 个池化层, 其中卷积层用于提取图像特征, 全连接层预测图像位置与类别估计概率值, 池化层负责缩减图片像素. YOLO 根据输入的图像数据, 运用回归分析法输出图像数据的多个滑动窗口位置及该窗口中检测到的目标类别.

YOLO 将输入图像分成 $S\times S$ 个单元格, 每个单元格的神经元负责检测落入该单元格的对象, 最多可包括两个预测对象的滑动窗口. 滑动窗口的信息采用五元组 $T(x,y,w,h,c)$ 表示, x 与 y 是当前格子神经元预测到的检测对象的置信度中心位置的横坐标与纵坐标. w 和 h 分别是滑动窗口的宽度和高度. c 是置信度, 它反映当前滑动窗口是否包含检测对象及其预测准确性的估计概率, 计算公式为

$$c = P_{\rm o}\times P_{\rm IOU} \tag{1}$$

其中, $P_{\rm o}$ 表示滑动窗口包含检测对象的概率, $P_{\rm IOU}$ 表示滑动窗口与真实检测对象区域的重叠面积 (单位是像素). 若滑动窗口中包含检测对象, 则 $P_{\rm o}=1$, 否则 $P_{\rm o}=0$. 当单元格具有多个滑动窗口时, 它们的最大 $P_{\rm IOU}$ 值将代入式 (1) 计算, 最终只选择重叠面积最大的检测对象输出.

通常, 若 B 为每个单元格可以用于预测对象的滑动窗口数量, C 为类别总数, 则 YOLO 的全连接层的输出维度是: $S\times S\times (B\times 5+C)$.

YOLO 的损失函数计算公式为

$$\lambda_{\rm loss} = E_{\rm c} + E_{\rm IOU} + E_{\rm class} \tag{2}$$

其中, $E_{\rm c}$, $E_{\rm IOU}$ 和 $E_{\rm class}$ 分别表示预测数据与标定数据之间的坐标误差、$P_{\rm IOU}$ 误差与分类误差.

坐标误差 $E_{\rm c}$ 的计算公式为

$$E_{\rm c}=\lambda_{\rm c}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\rm obj}\left[(x_i-\widehat{x}_i)^2+(y_i-\widehat{y}_i)^2\right]+$$
$$\lambda_{\rm c}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\rm obj}\left[\left(\sqrt{w_i}-\sqrt{\widehat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\widehat{h}_i}\right)^2\right] \tag{3}$$

其中, $\lambda_{\rm c}$ 是坐标误差 E_c 的权重系数, YOLO 中的取值为 5. x_i, y_i, w_i 和 h_i 分别表示预测的单元格 i 的滑动窗口的中心点横坐标与纵坐标及其长度和宽度; $\widehat{x}_i$, $\widehat{y}_i$, $\widehat{w}_i$ 和 $\widehat{h}_i$ 分别表示真实的单元格 i 的滑动窗口的中心点横坐标与纵坐标及其长度与宽度; $I_i^{\rm obj}$ 表示单元格 i 包含检测目标对象 (其值归一化为 0 或 1), $I_i^{\rm obj}$ 表示第 j 个滑动窗口中单元格 i 的神经元负责检测目标对象 (其值归一化为 0 或 1).

$P_{\rm IOU}$ 误差 $E_{\rm IOU}$ 的计算公式为

$$E_{\rm IOU}=\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\rm obj}(c_i-\widehat{c}_i)^2+$$
$$\lambda_{\rm nbj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}I_{ij}^{\rm nbj}(c_i-\widehat{c}_i)^2 \tag{4}$$

其中, $\lambda_{\rm nbj}$ 是滑动窗口与真实检测对象区域的重叠面积 $P_{\rm IOU}$ 误差的权重, YOLO 中设置为 0.5. c_i 表示预测的滑动窗口中单元格 i 的置信度值; $\widehat{c}_i$ 表示真实的滑动窗口中的单元格 i 的置信度值. $I_{ij}^{\rm nbj}$ 表示第 j 个滑动窗口内单元格 i 不负责检测目标对象, 即 $I_{ij}^{\rm obj}$ 和 $I_{ij}^{\rm nbj}$ 分别表示检测目标对象是否存在于第 j 个滑动窗口的单元格 i 内. 考虑到单元格包含检测对象与不包含检测对象其 $P_{\rm IOU}$ 误差对整个训练模型的损失函数的贡献不同, 不包含检测对象的单元格的神经元的置信度值趋近于 0, 若采用相同的权重, 则会间接地放大包含有检测对象的单元格的自信度误差. 因此, YOLO 使用 $\lambda_{\rm nbj}=0.5$ 以减小传递误差.

分类误差 $E_{\rm class}$ 计算公式为

$$E_{\rm class}=\sum_{i=0}^{S^2}I_{ij}^{\rm obj}\sum_{k=0}^{C}(p_i(k)-\widehat{p}_i(k))^2 \tag{5}$$

其中, $p_i(k)$ 与 $\widehat{p}_i(k)$ 分别表示预测的与真实的滑动窗口中单元格 i 包含第 k 类对象的条件概率.

2 改进的目标实时检测模型与特征提取算法

2.1 提出改进思想的来源

1) 第 1 类现象: 典型的目标匹配检测方法 DPM (Deformable parts model)[25] 利用梯度信息提取图像的特征, 通过计算梯度方向的直方图获得梯度模型与目标匹配关系, 从而实现目标分类和检测. 对于梯度方向的直方图, 首先将滑动窗口划分为大小相同的细胞单元, 并分别提取相应的梯度信息, 以减少光照或背景因素的影响; 之后, 将相邻细胞单元组合成相互重叠的块以充分利用重叠的单元信息; 然后统计整个块的直方图, 与此同时通过归一化处理每个块内的直方图以减少噪声对图片的影响; 之后, 收集所有直方图特征形成特征向量. 最后, 采用支持向量机分类得到物体的梯度模型. 此方法可以减少背景噪声数据对判定精度的影响, 有利于提高分类和检测的准确性.

2) 第 2 类现象: 最近邻的目标检测方法 RPN (Region proposal networks)[20] 的核心思想是给定输入图像, 经过卷积神经网络对输入的特征图进行卷积和池化, 在最后一个卷积层, 采用滑动窗口进行特征提取操作, 得到相应的特征向量, 再采用 Softmx 分类函数进行分类和边框回归, RPN 方法能够以较低的时间成本获得较高的单一目标识别准确率.

3) 第 3 类现象: 全卷积神经网络 R-FCN (Region-based fully convolutional networks)[26] 只包括卷积层和池化层, 具有实现整个图像信息共享的机制, 在分类准确性方面具有良好表现.

综上可知, 当处理因光照、背景、采集设备等不同而引入的噪声数据时, 可以借鉴 DPM 方法, 通过增加滑动窗口内细胞单元的数量提高复杂背景数据的分类和检测准确性; 与此同时, 对于单一目标的数据, 可以基于 RPN 方法获得较好的识别效果; 而 R-FCN 方法可以保留更多的图像信息, 这有利于特征的提取. 正是基于这些启发, 本文试图通过增加检测窗口内细胞单元的数量, 移除全连接层, 并结合边框回归和滑动窗口, 以提高 YOLO 性能.

2.2 改进的 YOLO 神经网络结构

基于上一节的启发, 本文提出了改进的 YOLO 神经网络结构, 包括 18 个提取图像特征的卷积层、6 个用来缩减图片像素的池化层、1 个 Softmax 输出层和 1 个全连接层, 如图 1 所示. 同时, 采用 Dropout[27] 方法以 0.3 的概率随机将神经元置为零, 从而丢弃神经网络中的部分神经元, 以减少计算成本, 降低节点间耦合性, 缓解过拟合问题. 此结构中, 借鉴 R-FCN 方法采用一个全连接层以减少特征

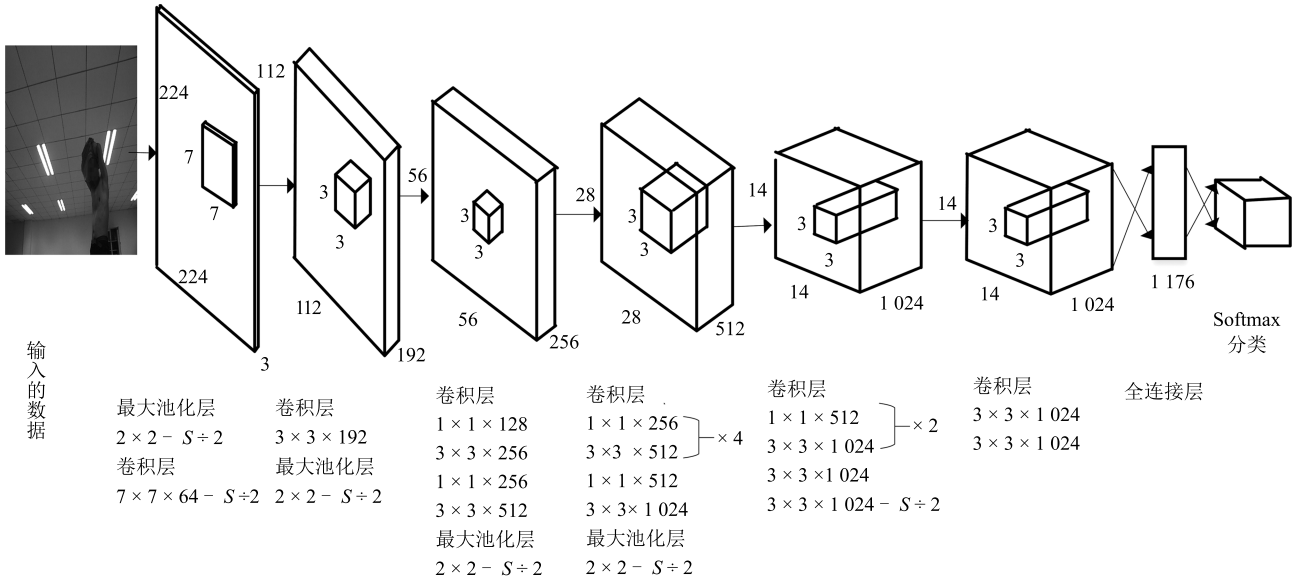


图 1　改进的 YOLO 神经网络结构

Fig. 1　Improved YOLO neural network structure

信息的丢失; 在输入图像后, 借鉴 RPN 方法设置了一个 2×2 的最大池化层以缩小图片尺寸的同时尽可能多的保存原始图片信息. 与此同时, 将多层卷积和池化操作后的网格由原来的 7×7 变为 14×14 以提高网络特征图谱的尺寸. 图 2 是不同网格尺寸下目标识别结果对比图. 由图 2 可知, 在 7×7 网格下, 系统只能预测 2 个目标, 但在 14×14 网格下系统可以识别出 3 个目标, 当图中有多个目标对象, 特别是包括小目标对象时, 这种扩大后的网格尺寸, 可以增加小目标特征的提取能力, 实现对小目标的识别, 从而提高系统的识别准确性. 各种目标是构成不同情境的要素, 通过对目标的识别可以区分不同的情境, 当情境中包括涉及隐私内容的目标时, 即可判定为隐私情境. 因此, 当目标中包括涉及隐私内容的小目标时, 提高小目标识别准确性, 有利于提高隐私情境检测的准确性.

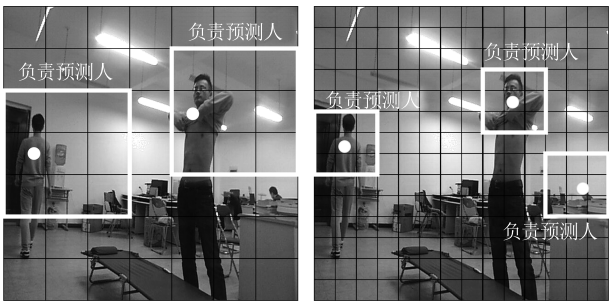


图 2　不同网格规模下的目标识别效果对比图

Fig. 2　Comparison diagram of object recognition with different grid scale

2.3　基于 RPN 的滑动窗口合并算法

YOLO 在检测目标对象时, 一个单元格涉及到多个滑动窗口, 而最终输出的标识目标对象的窗口小于等于图片数据分类数. 当将 YOLO 应用于情境检测时, 不需要标识所有的目标, 而是需要反馈需要检测的对象是否存在于当前视图中. 因此设计了基于 RPN 的滑动窗口合并算法, 具体见算法 1.

算法 1. 基于 RPN 的滑动窗口合并算法

输入. 图片数据 X_{pic}.

输出. 检测目标的滑动窗口位置的集合 L.

步骤 1. 利用网格划分的方法将 X_{pic} 划分为 n 个单元格, 生成集合 $R = \{S_1, S_2, \cdots, S_n\}$;

步骤 2. 初始化单元格 S_i 的相似集合 $m_i = \emptyset$, 并初始化 14×14 规格的滑动窗口;

步骤 3. for 滑动窗口中的邻近区域对 (S_i, S_j) **do**

步骤 3.1 采用 RPN 方法及式 (1) 计算滑动窗口内与 S_i 相邻的所有单元格 S_j 的特征相似度 $F(S_i, S_j)$;

步骤 3.2 找出最大相似度值 $F_{\max}(S_i, S_j)$;

步骤 3.3 更新单元格 S_i 的相似集合 $m_i = m_i \cup \{F_{\max}(S_i, S_j)\}$;

步骤 3.4 while (每一个单元格 S_i 的相似集合 $m_i \neq \emptyset$)

1) 找出集合 m_i 中的元素对应的所有单元格, 并去除不包括检测对象的单元格;

2) 将所获得的单元格与单元格 S_i 合并形成新的 S_i, 并将其作为集合 L 的元素;

步骤 4. 输出目标位置检测滑动窗口集合 i.

运用算法 1 获得的集合 L 可以确定经过卷积和池化操作后的滑动窗口的边框. 通过合并相似区域可以减少冗余和时间开销.

2.4 基于改进 YOLO 的特征提取算法

结合第 2.2 节的结构及第 2.3 节设计的算法, 本节给出基于改进 YOLO 的特征提取算法, 具体见算法 2.

算法 2. 基于改进 YOLO 的特征提取算法

输入. 图片数据集 X.

输出. 图片数据 X 的特征模型 M_{weights}.

步骤 1. 图片数据预处理, 针对图片数据集 X 的每一张图片采用 LabelImg[28] 工具获得真实目标的矩形区域坐标, 生成每张图片中真实目标的坐标信息文件 F_c;

步骤 2. 加载 YOLO 的图片分类训练模型, 同时初始化图片数据 X 的特征模型 M_{weights}, 初始化每张图片的预测矩形区域坐标为空;

步骤 3. 坐标信息文件 F_c, 基于 RPN 方法生成每张图片的若干个目标候选区域矩阵向量;

步骤 4. 将候选区域矩阵向量作为第 1 层的输入, 将其结果作为第二层的输入;

步骤 5. 执行池化操作;

步骤 6. 将步骤 5 中的结果作为输入, 采用一个滑动窗口扫描网格, 进行卷积与池化操作计算出滑动窗口内单元格的特征向量;

步骤 7. 将步骤 6 所得的特征向量作为第 18 个卷积层的输入, 运用 2×2 步幅进行卷积操作;

步骤 8. 将步骤 7 的输出作为全连接层的输入, 采用 1×1 步幅进行卷积操作;

步骤 9. 将步骤 8 的输出作为分类函数 Softmax 的输入, 计算图片数据 X_{pic} 的预测概率估计值 P_{pic}, 并保存运用算法 1 获得的重叠面积最大的 P_{IOU} 对应的目标区域的特征;

步骤 10. 将对应的目标区域的特征保存到特征模型 M_{weights} 中每一个类别相对应的位置;

步骤 11. 输出特征模型 M_{weights}.

算法 2 中, 步骤 1 的 LabelImg 工具用于获得选定区域的坐标信息. 步骤 7 运用 2×2 的最大池化层以缩小图片尺寸的同时尽可能多的保存原始图片的信息, 输出 14×14 的网络特征图谱. 步骤 8 中, 滑动窗口要在 17 个用来提取图像特征的卷积层和 6 个减小图像尺寸的池化层中进行操作. 在这个过程中, 滑动窗口每次进行卷积操作时, 运用算法 1 计算出重叠面积最大的 P_{IOU} 代入式 (2) 计算损失函数的最小值. 在应用系统中, 可以根据步骤 11 中输出的特征模型 M_{weights} 进行应用判定.

3 隐私检测服务机器人硬件平台

图 3 是课题组搭建的服务机器人平台, 包括移动底座、数据处理器、数据采集设备以及机械支架等部分. 图 4 为系统的总体工作流程. 用于输入与显示数据的触摸显示屏是 16 寸的支持 Linux 系统的工业触摸屏; 视觉系统采用 ORBBEC 3D 体感摄像头, 可以采集 RGB 深度图像. 听觉系统是基于科大讯飞语音模块拓展而成, 能够在嘈杂环境中识别语音和定位声音方位. 开发板是拥有 256 核 GPU 的

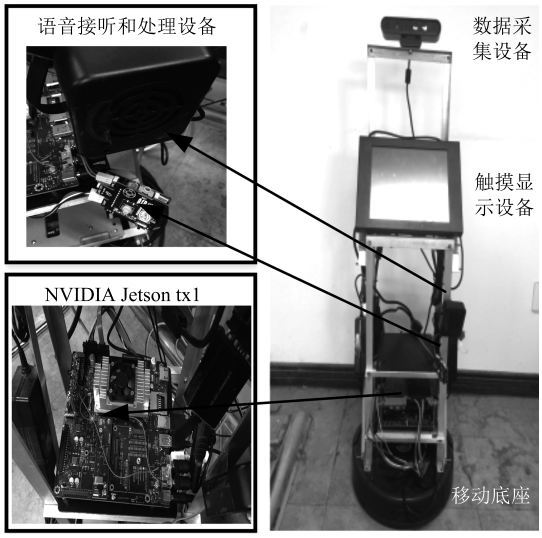


图 3 服务机器人平台

Fig. 3 Social robot platform

图 4 情境检测系统的总体工作流程

Fig. 4 The overall flow chart of the privacy situation detection system

NVIDIA Jetson TX1; 移动底座是 iRobot Create 2. 系统的操作系统是 Ubuntu 16.04, 并安装了 Kinect 版本的 ROS (Robot operation system) 系统. 用于降低服务机器人运算负荷的工作站是 ThinkPad t550 (GPU 为 NVIDA geForce 940 MB), 主要用于数据分析. 同时, 服务机器人与工作站均安装了 OpenCV 3.1 与 TensorFlow 0.9[29], YOLO, ROS 系统. 服务机器人具有无线通讯模块, 可以实现服务机器人与工作站间端到端的通信.

图 4 中, 在收集训练数据集的基础上, 具有 GPU 的工作站运用算法 2 训练数据集以获得特征模型. 然后, 将获得的特征模型传送到服务机器人, 服务机器人接收到模型后开启摄像头, 并按给定频率 (10 秒) 从摄像头读取图片进行情境检测. 最后, 根据检测结果决定机器人动作. 若检测到隐私情境, 机器人调整摄像头的角度, 同时根据识别的隐私内容形成摘要的信息存储在文本文件中, 每隔 30 秒后, 用语音咨询的方式, 询问是否可以将摄像头重新用于观察人的行为. 若回复是否定的, 则系统的摄像头保持不工作的状态, 从而达到保护隐私信息的目的. 例如, 当系统检测到用户在洗澡时, 将摄像头旋转 90 度, 并存储文本信息 "2017 年 3 月 29 日 8:00 用户在洗澡". 同时, 系统开始记时, 30 秒后, 询问是否已经洗澡完成. 若人响应的内容是肯定的, 则摄像头恢复到前一时刻的观察角度继续采集数据, 然后依据识别的数据决定服务机器人的动作.

4　数据集与实验设计

4.1　训练数据集与验证数据集

训练数据集由不同情境下的图片数据组成, 运用提出的算法获得特征模型用于应用系统. 验证数据集用于特征模型提取过程中, 测试特征模型在不同参数下的识别性能以精练特征模型.

本文考虑的家居情境包括 6 类: C1: 洗澡; C2: 裸体或半裸体睡觉; C3: 上厕所; C4: 导致身体裸露的换衣服; C5: 有人但不涉及上述隐私内容; C6: 家居环境中没有人的存在. 数据来源包括两种方式: 1) 在课题组组建的家居环境中, 使用构建的服务机器人平台上的 3D 体感摄像头自主采集的图片, 约占整个数据集的 81 %. 2) 从网络中收集、筛选并进行适当处理后的家居环境中的图片, 它们具有不同的场景、对象、光亮、角度与像素, 以丰富数据集.

训练数据集的 6 类情境共包括 2 580 个样本, 每类包括 430 张样本.

验证数据集的 6 类情境由 360 个样本组成, 每一个类包括 60 个样本.

图 5 是数据集的样本示例.

C1　　C2　　C3

C4　　C5　　C6

图 5　数据集示例

Fig. 5　Samples of the collected dataset

4.2　系统性能测试实验设计与测试数据集

为了测试系统的性能, 设计了 3 个实验.

实验 1. 家居环境包括在训练数据集中的隐私情境检测. 测试数据 a 与 b 的获取方式是: 由训练集中的对象和不在训练集中的对象分别在课题组家居环境中获取的图片. 此实验测试系统对不同检测对象的检测鲁棒性.

实验 2. 检测对象 (人) 相同时, 检测环境不包括在训练数据集中的隐私检测. 这一实验查验训练集中的情境发生变化后, 系统对于隐私检测内容的准确性. 测试数据 c 为: 训练集中的对象在其他家居环境中的图片. 此实验考查系统对不同检测环境的检测表现.

实验 3. 检测对象与家居环境情境均不包括在训练数据集中的隐私检测. 为了体现数据的客观性和多样性, 测试数据 d 从网络上搜集整理而得. 在测试时, 通过模拟系统摄像头实时采集的方式为检测系统提供数据. 该实验检测系统在检测对象与环境均与训练数据完全不同时的性能.

上述 a, b, c, d 四类测试数据, 在每种情境下均测试 40 张图片, 每类数据在 6 种情境下共测试相互各异的 240 张图片. 完成 4 个实验, 共涉及 960 张图片. 测试数据集与训练集无雷同数据.

特别需要说明的是, 考虑到实时采集的数据不方便比较测试与分析, 因此, 后文测试和比较所用的数据集均是提前采集的数据, 测试时模拟摄像头实时工作的机制将数据传送给系统.

5　训练模型参数优化结果与分析

考虑到模型的训练需要花费大量的时间, 不同的训练规模对模型的性能有影响. 为了让提出的训练模型具有较好的性能, 本节研究训练步骤对预测

概率估计值的影响, 从而找出较优 (或者说可行的) 训练步骤规模. 同时, 由于不同的学习率对模型的识别准确性也有影响, 因此通过实验测试, 研究了不同学习率下模型的识别准确性.

5.1 训练步骤规模与预测概率估计值的关系分析

设计了 11 种不同的步骤规模, 并针对上一节给出的验证数据集的 360 个样本, 借鉴 YOLO 的设置给定模型的学习率为 0.001 时, 模型的预测概率估计值、识别准确率及单图识别时间的平均值统计结果如表 1 所示, 变化趋势如图 6 所示, 不同训练步骤下模型的类别估计值统计盒图如图 7 所示.

表 1 不同步骤下的模型性能表现

Table 1 The model performance with different steps

助记符	步骤规模	预测概率估计值均值	识别准确率均值	单图识别时间均值 (ms)
L1	1 000	0.588	0.733	2.46
L2	2 000	0.627	0.750	2.50
L3	3 000	0.629	0.717	2.51
L4	4 000	0.642	0.700	2.53
L5	5 000	0.729	0.800	2.55
L6	6 000	0.731	0.817	2.52
L7	7 000	0.782	0.850	2.45
L8	8 000	0.803	0.883	2.17
L9	9 000	0.830	0.967	2.16
L10	10 000	0.804	0.900	2.21
L11	20 000	0.569	0.417	2.27

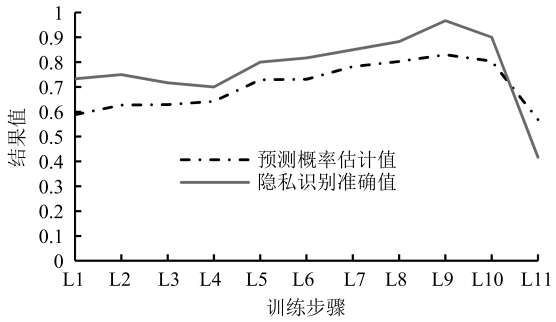


图 6 不同步骤下的模型性能变化趋势

Fig. 6 Variation trends of the proposed model under different steps

从图 6 与表 1 可以看出, 当训练步骤为 1 000 时, 平均预测概率估计值为 0.588, 识别准确率为 0.733; 随着训练步骤的增加, 模型的预测概率估计值和隐私情境识别准确值呈上升趋势, 当训练步骤规模为 9 000 时, 模型的平均预测概率估计值达到最高值 0.830, 同时识别准确率的均值也达到最大 0.967. 当训练步骤继续增大到 20 000 步时, 模型的平均预测概率估计值下降为 0.568, 此时的平均预测准确值为 0.417. 同时, 结合图 7 可知, 在当训练步骤处于 1 000 ∼ 7 000 时, 虽然矩形外的异常值较少, 但是所对应的盒图矩形区域较长, 且中位线较低. 当训练步骤为 8 000 与 10 000 时, 虽然数据的中位线较高, 但是处于矩形框外的异常点也比较多, 而且存在接近 0 的预测估计值奇异点. 当训练步骤为 9 000 时, 盒图矩形区域面积较窄, 并且较其他情况下具有最高的中位线, 虽然存在处于矩形框外的异常点, 但最低的异常点都高于训练步骤为 2 000、3 000 和 4 000 所对应的最低矩形区域; 进一步检查对应的数据发现, 此时的异常点数据仅有 2 个, 均大于 0.450.

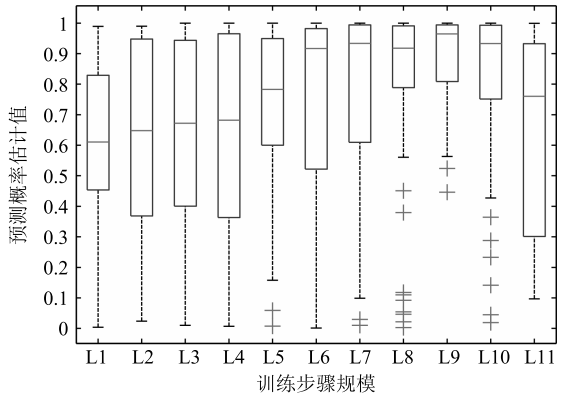


图 7 不同训练步骤下模型的预测概率估计值统计盒图

Fig. 7 Boxplot of prediction accuracy under different training steps

从表 1 中的时间开销统计结果可知, 系统的平均开销时间在 2.1 ∼ 2.6 ms 之间, 模型具有较短的识别时间, 满足实时性要求较低的实时检测应用要求.

综上分析可以得出结论, 当训练步骤设置为 9 000 时, 所提出的模型能够获得最好的预测估计值与识别准确性.

5.2 不同学习率下的识别性能实验结果与分析

为了获得能够让模型发挥最好性能的学习率设置, 结合上一节的结论, 在设置训练步骤为 9 000 时, 考查学习率分别为 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9} 和 10^{-10} 时的模型性能表现. 针对设计的验证数据集的 360 个样本, 模型的预测概率估计值与识别准确率平均值统计结果见表 2、图 8 与图 9.

从表 2 与图 8 可以看出, 当学习率大于 0.100 时, 随着学习率的减小, 模型的平均概率预测估计值与识别准确率均有增大的趋势. 当学习率为 10^{-1}

表 2　不同学习率下的模型性能统计结果

Table 2　The statistical results of model performance with different learning rates

助记符	学习率	预测概率估计值均值	识别准确率均值
R1	1	0.670	0.817
R2	10^{-1}	0.911	1.000
R3	10^{-2}	0.843	0.933
R4	10^{-3}	0.805	0.950
R5	10^{-4}	0.801	0.950
R6	10^{-5}	0.672	0.933
R7	10^{-6}	0.626	0.900
R8	10^{-7}	0.565	0.880
R9	10^{-8}	0.569	0.867
R10	10^{-9}	0.391	0.800
R11	10^{-10}	0.315	0.417

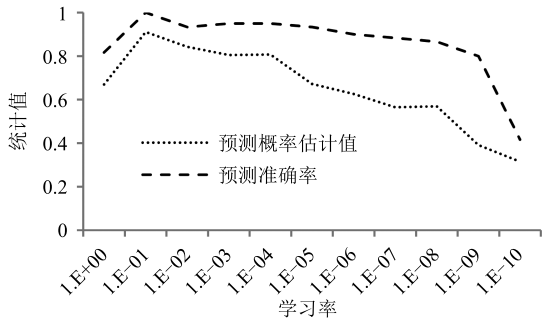


图 8　不同学习率下的模型性能变化趋势

Fig. 8　The trend of model performance under different learning rates

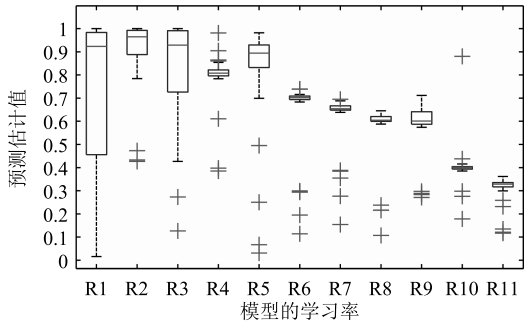


图 9　不同学习率下的预测估计值统计盒图

Fig. 9　Boxplot of prediction accuracy with different learning rates

时, 预测概率估计值达到最大值 0.911, 并且平均识别准确率达到 1. 当学习率从 10^{-1} 减小到 10^{-4} 时, 预测概率估计值处于 0.800 以上, 识别准确率均值处于 0.940 左右, 学习率的变化对上述两性能指标的影响较小. 当学习率从 10^{-4} 减小到 10^{-10} 时, 预测概率估计值与识别准确率的均值随着学习率的变小而呈现出明显下降, 它们的最低平均值分别为 0.315 和 0.417.

从图 9 可进一步发现, 当学习率为 1 时, 对应的矩形框面积最大, 虽然表 2 中对应的平均值只有 0.670, 但其盒图中对应的矩形框延伸到了纵轴上的 0.900 刻度以上, 表明存在一定数量大于 0.900 的预测估计值. 当学习率为 0.100 时, 虽然存在一些异常值, 但其矩形区域较小, 表明系统在大数情况下可以输出较大的预测估计类别值. 在学习率为 $10^{10}\sim 10^{-1}$ 内时, 对应图形存在较多的异常点, 并输出大量较小的预测概率估计值.

综上分析可以得出结论, 当学习率设置为 0.100 时, 所提出的模型具有较好的性能表现, 在应用时可以采用此设置.

6　应用系统性能测试

6.1　系统性能测试结果与分析

在搭建的服务机器人平台上, 部署设计的算法, 同时将学习率与训练步骤分别设置为 0.100 和 9 000, 针对测试数据集中的四类数据进行测试, 系统情境识别准确率、类别估计值及时间开销统计结果见表 3 与表 4, 预测概率估计值统计盒图见图 10.

表 3　系统针对不同测试数据集的隐私识别准确率

Table 3　Privacy situation recognition accuracy of the proposed system for different testing data sets

实验	测试数据集	情境识别准确率					
		C1	C2	C3	C4	C5	C6
实验 1	a 类测试数据	0.900	0.975	0.975	0.975	1.000	0.975
	b 类测试数据	0.850	0.950	0.975	0.925	1.000	0.950
实验 2	c 类测试数据	0.850	0.850	0.950	1.000	1.000	0.925
实验 3	d 类测试数据	0.850	0.850	0.850	0.900	0.975	0.875

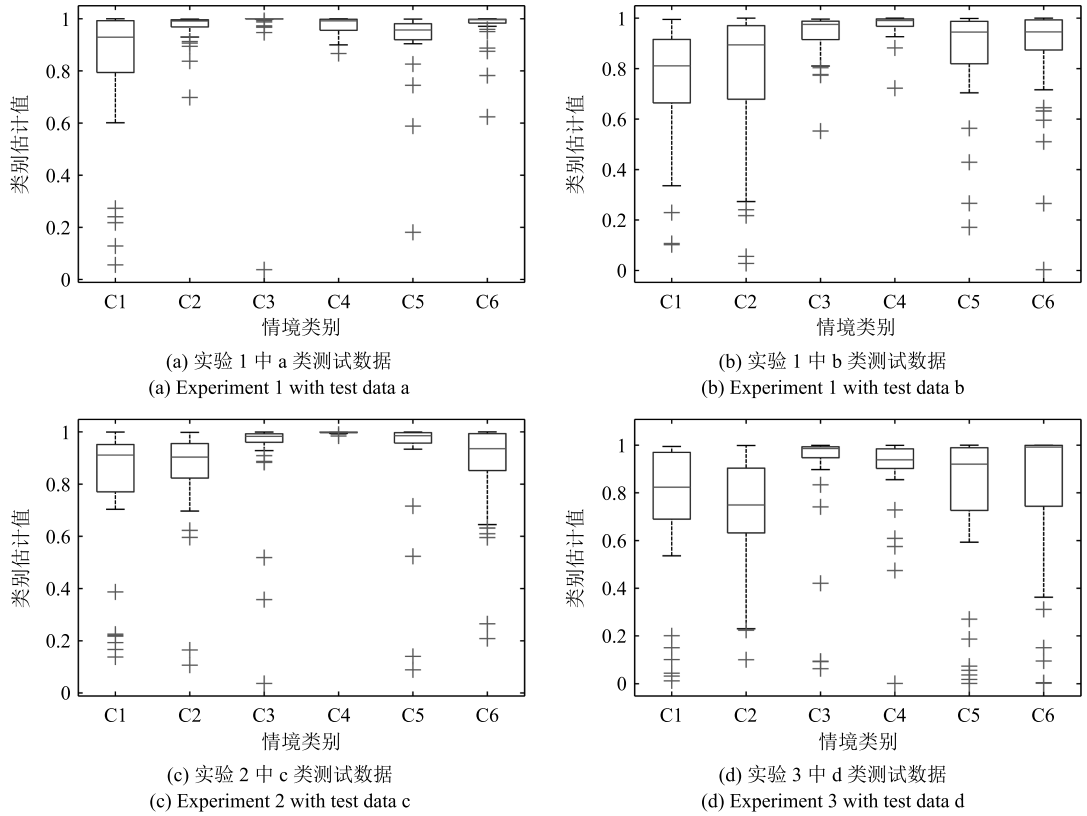
观察这些数据可知:

1) 由实验 1 中的 a 类测试数据可知, 系统的情境识别准确率在情境 C2, C3, C4 和 C6 下为 0.975, 在 C5 情境下为 1, 在 C1 情境下最低为 0.9. 对于实验 1 中的 b 类测试数据, 在 C2, C3, C4 和 C6 情境下分别对应的识别准确率为 0.950, 0.975, 0.925, 0.950, 在 C5 情境下为 1, 在 C1 情境下为 0.850. 表 4 的数据显示, 对于 a 类测试数据, 针对 C1 ∼ C6 情境类别估计值均值分别是: 0.82, 0.968, 0.971, 0.972, 0.920 和 0.972, 与之相对应的标准方差分别为: 0.275, 0.006, 0.168, 0.038, 0.141 和 0.152, 它们

表 4 系统针对不同测试数据的隐私类别估计值统计表

Table 4 Privacy situation recognition accuracy of the proposed system for different testing data sets

测试数据	判别估计值											
	C1		C2		C3		C4		C5		C6	
	均值	方差	均值	方差	均值	方差	均值	方差	均值	方差	均值	方差
a 类测试数据	0.820	0.275	0.968	0.006	0.971	0.168	0.972	0.038	0.920	0.141	0.972	0.152
b 类测试数据	0.789	0.276	0.849	0.192	0.922	0.096	0.997	0.003	0.918	0.216	0.869	0.191
c 类测试数据	0.751	0.359	0.774	0.253	0.937	0.272	0.974	0.047	0.854	0.212	0.864	0.214
d 类测试数据	0.742	0.304	0.713	0.274	0.854	0.292	0.890	0.186	0.768	0.332	0.807	0.311
单图识别时间 (ms)	3.32		1.62		3.13		2.87		2.69		3.15	



(a) 实验 1 中 a 类测试数据
(a) Experiment 1 with test data a

(b) 实验 1 中 b 类测试数据
(b) Experiment 1 with test data b

(c) 实验 2 中 c 类测试数据
(c) Experiment 2 with test data c

(d) 实验 3 中 d 类测试数据
(d) Experiment 3 with test data d

图 10 预测概率估计值统计盒图

Fig. 10 Boxplot of prediction accuracy

的类别估计值较高而方差较小, 表明系统对于测试的数据能够以非常大的概率归到相应的类别中, 对于对象与背景均包括在训练集中的数据, 系统对于不同视角下的对象与背景组成的新情境具有较强的识别能力. b 类测试数据对应的结果, 较 a 类的结果整体稍差一些, 各情境下的类别估计值均值分别是 0.789, 0.849, 0.922, 0.977, 0.918, 0.869, 而识别准确率方面, 情境 C1, C2, C4 与 C6 分别下降了 0.05, 0.025, 0.050 与 0.025, 表明对象的变化对系统的识别性能有一定的影响.

2) 由实验 2 的结果可知, 系统对 C4 和 C5 的情境识别准确率为 1, 对 C1 ∼ C3 和 C6 情境下识别准确率为 0.850, 0.850, 0.950 与 0.925. 对应的预测概率估计值, 较 a 与 b 类测试数据的结果, C1 ∼ C3, C5, C6 情境的均值分别最大下降了 0.069, 0.194, 0.034, 0.066 和 0.108, 表明通过有限的训练集获取的特征, 可以较高的识别准确率预测已在训练集中的对象和未在训练集中的家居环境组成的新情境, 但家居环境的改变会让系统的情境识别性能呈现出下降的趋势.

3) 由实验 3 的数据可知, 虽然系统的识别准确率均值最高为 0.975, 最低为 0.850, 但是其预测估计值均值分布在相对较低的区间 [0.713, 0.890]. 表明当家居环境与对象均发生改变后, 系统的识别准确性与类别估计值均会下降. 但值得注意的是, d 类数据来源于网络, 其背景主题、对象与采集角度均与训练集的数据差异较大, 而系统依然可以获得 0.85 以上的识别准确率, 说明系统对于识别具有较大差异的新样本拥有较强鲁棒性.

4) 观察图 10 可知, 图 10 (a) ∼ (c) 对应的中位线均处于刻度 0.800 之上, 且盒图矩形区域较小, 表明系统对于 a, b 和 c 类数据的识别性能较好. 而由图 10 (d) 可知, 系统输出的预测概率估计值分布区域较大, 表明环境和对象的改变会影响系统的识别性能. 整体上, 虽然系统拥有 94.48 % 的识别准确性, 但是却存在处于矩形外的异常点, 特别是预测估计值非常小的点, 表明系统对于某些情境的识别是在预测概率估计值非常低的情况下做出的判定, 系统对于这一类数据的识别鲁棒性需要改进.

6.2　系统识别有误的数据分析

从上一节的分析可知, 构建的系统有 5.52 % 的情境识别错误, 我们从 960 张测试图片中找出了识别有误的 53 张图片, 分析这些图片可知:

1) 由系统中摄像头采集的数据, 具有光线较暗和存曝光过度的亮区域的特点. 同时我们查验了训练数据, 发现当中不存在此类训练数据.

2) 来自于网络的图片, 具有分辨率低或色彩单一的特点, 这会引入较强的噪声.

因此, 为了提高系统的识别性能, 应扩大训练集的样本多样性, 并将识别错误的样本放置到相应的训练数集中, 以获取更加具有普适性的特征模型.

7　与 YOLO 算法的对比结果与分析

本节给出了本文算法与 YOLO 算法的比较结果. 针对前文设计的情境与实验方案, 部署的 YOLO 算法运行参数与文献 [30] 相同, 情境识别准确率和预测概率估计值统计结果如表 5 与表 6 所示. 结合表 3 与表 4 的数据可知:

1) 对于 a 类测试数据集中的各个情境, 除 C4 情境下 YOLO 的表现优于本文算法, 其他 C1, C2, C3, C5 及 C6 情境, YOLO 的识别准确率均值分别比本文算法低了 0.150, 0.000, 0.025, 0.025, 0.025, 其预测概率估计值分别低了 0.176, 0.029, 0.098, −0.052, 0.036, 且方差分别高 0.091, 0.176, 0.076, −0.008, 0.071.

2) 对于 b 类测试数据集, 除了 C2 情境下 YOLO 的表现优于本文算法, 在 C1, C3, C4, C5 及 C6 情境下, 本文算法的识别准确率均值与预测概率估计值均值均大于 YOLO 算法的均值.

表 5　YOLO 算法的隐私识别准确率统计结果

Table 5　Privacy situation recognition accuracy by applying YOLO

实验	测试数据集	情境识别准确率					
		C1	C2	C3	C4	C5	C6
实验 1	a 类测试数据	0.750	0.975	0.950	1.000	0.975	0.950
	b 类测试数据	0.725	0.975	0.875	0.875	0.825	0.750
实验 2	c 类测试数据	0.625	0.850	0.675	0.675	0.600	0.750
实验 3	d 类测试数据	0.600	0.600	0.600	0.600	0.600	0.725

表 6　YOLO 算法的隐私类别预测概率估计值统计结果

Table 6　Statistical results of privacy situation estimates by applying YOLO

情境类别	判别估计值							
	a 类测试数据		b 类测试数据		c 类测试数据		d 类测试数据	
	均值	方差	均值	方差	均值	方差	均值	方差
C1	0.644	0.366	0.568	0.465	0.540	0.381	0.501	0.413
C2	0.939	0.182	0.923	0.149	0.693	0.317	0.305	0.433
C3	0.873	0.244	0.867	0.302	0.866	0.290	0.851	0.313
C4	0.999	0.001	0.963	0.017	0.647	0.439	0.513	0.399
C5	0.972	0.133	0.815	0.228	0.570	0.381	0.568	0.465
C6	0.936	0.223	0.725	0.339	0.674	0.386	0.622	0.345

3) 对于 c 类测试数据集中的各个情境, YOLO 算法的识别准确率均值比本文算法分别低 0.225, 0.000, 0.275, 0.325, 0.400, 0.175; 而预测概率估计值均值和方差方面, YOLO 算法的表现均差于本文算法.

4) 对于 d 类测试数据的各情境, YOLO 算法的预测概率估计值均值分别为 0.501, 0.305, 0.851, 0.513, 0.568, 0.622; 而本文算法的预测概率估计值分别为 0.742, 0.713, 0.854, 0.890, 0.768, 0.807.

综上可知, 本文提出的改进算法的识别性能优于 YOLO 算法. 导致这种不同的原因正是因为改进后的网络结构可以保留更多的原始图片信息以及增加目标特征的提取能力, 而其中的基于 RPN 的滑动窗口合并算法能够提高具有复杂背景数据的分类和检测的准确性. 正是这些改进使得算法在处理具有不同场景、对象、光亮、角度与像素等的图片时, 能够表现出更好的识别性能.

8 结束语

对隐私内容有符合人心理需求反应的系统, 可以改善用户体验感受, 服务机器人的视觉设备会引入隐私泄漏风险, 因此, 试图通过设计图像特征提取方法及系统以求较好地解决此问题. 本文改进了 YOLO 神经网络的结构、特征提取过程以及图片网格划分大小, 同时设计了基于 RPN 的滑动窗口合并算法, 形成了基于改进 YOLO 的特征提取算法. 通过在课题组建立的隐私情境数据集和搭建的服务机器人平台上进行实验分析, 结果表明, 提出的特征提取算法在服务机器人系统中可以较好地识别智能家居环境中涉及隐私的情境, 算法具有较好的鲁棒性, 可以实时检测家庭环境中的隐私情境. 与 YOLO 的比较结果表明设计的方法具有明显的优势. 下一步工作将丰富涉及隐私信息的情境类别, 丰富隐私图片数据集, 并研究将隐私内容转化为非隐私内容的近似等价方法.

References

1 Shankar K, Camp L J, Connelly K, Huber L L. Aging, privacy, and home-based computing: developing a design framework. *IEEE Pervasive Computing*, 2012, **11**(4): 46−54

2 Fernandes F E, Yang G C, Do H M, Sheng W H. Detection of privacy-sensitive situations for social robots in smart homes. In: Proceedings of the 2016 IEEE International Conference on Automation Science and Engineering (CASE). Fort Worth, TX, USA: IEEE, 2016. 727−732

3 Arabo A, Brown I, El-Moussa F. Privacy in the age of mobility and smart devices in smart homes. In: Proceedings of the 2012 ASE/IEEE International Conference on and 2012 International Conference on Social Computing (SocialCom) Privacy, Security, Risk and Trust. Amsterdam, Netherlands: IEEE, 2012. 819−826

4 Kozlov D, Veijalainen J, Ali Y. Security and privacy threats in IoT architectures. In: Proceedings of the 7th International Conference on Body Area Networks. Brussels, Belgium: ICST, 2012. 256−262

5 Denning T, Matuszek C, Koscher K, Smith J R. A spotlight on security and privacy risks with future household robots: attacks and lessons. In: Proceedings of the 11th International Conference on Ubiquitous Computing. Orlando, USA: ACM, 2009. 105−114

6 Lee A L, Hill C J, McDonald C F, Holland A E. Pulmonary rehabilitation in individuals with non-cystic fibrosis bronchiectasis: a systematic review. *Archives of Physical Medicine and Rehabilitation*, 2017, **98**(4): 774−782

7 Liu Kai, Zhang Li-Min, Fan Xiao-Lei. New image deep feature extraction based on improved CRBM. *Journal of Harbin Institute of Technology*, 2016, **48**(5): 155−159
(刘凯, 张立民, 范晓磊. 改进卷积玻尔兹曼机的图像特征深度提取. 哈尔滨工业大学学报, 2016, **48**(5): 155−159)

8 Lee H, Grosse R, Ranganath R, Ng A Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. New York, USA: ACM, 2009. 609−616

9 Yu Lai-Hang, Feng Lin, Zhang Jing, Liu Sheng-Lan. An image feature extraction method based on adaptive fusion of object and background. *Journal of Computer-Aided Design and Computer Graphics*, 2016, **28**(8): 1250−1259
(于来行, 冯林, 张晶, 刘胜蓝. 自适应融合目标和背景的图像特征提取方法. 计算机辅助设计与图形学学报, 2016, **28**(8): 1250−1259)

10 Ding Y, Zhao Y, Zhao X Y. Image quality assessment based on multi-feature extraction and synthesis with support vector regression. *Signal Processing: Image Communication*, 2017, **54**: 81−92

11 Batool N, Chellappa R. Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints. *Pattern Recognition*, 2015, **48**(3): 642−658

12 Joseph R, Santosh D. YOLO: real-time object detection [Online], available: http://pjreddie.com/darknet, November 3, 2016

13 Liu Y L, Zhang Y M, Zhang X Y, Liu C L. Adaptive spatial pooling for image classification. *Pattern Recognition*, 2016, **55**: 58−67

14 Zhu Yu, Zhao Jiang-Kun, Wang Yi-Ning, Zheng Bing-Bing. A review of human action recognition based on deep learning. *Acta Automatica Sinica*, 2016, **42**(6): 848−857
(朱煜, 赵江坤, 王逸宁, 郑兵兵. 基于深度学习的人体行为识别算法综述. 自动化学报, 2016, **42**(6): 848−857)

15 Peng Q W, Luo W, Hong G Y, Feng M. Pedestrian detection for transformer substation based on Gaussian mixture model and YOLO. In: Proceedings of the 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Hangzhou, China: IEEE, 2016. 562−565

16 Nguyen V T, Nguyen T B, Chung S T. ConvNets and AGMM based real-time human detection under fisheye camera for embedded surveillance. In: Proceedings of the 2016 International Conference on Information and Communication Technology Convergence (ICTC). Jeju, South Korea: IEEE, 2016. 840−845

17 Erseghe T. Distributed optimal power flow using ADMM. *IEEE Transactions on Power Systems*, 2014, **29**(5): 2370−2380

18 Gupta A, Vedaldi A, Zisserman A. Synthetic data for text localisation in natural images. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 2315−2324

19 Parham J, Stewart C. Detecting plains and grevy′s zebras in the realworld. In: Proceedings of the 2016 IEEE Winter Applications of Computer Vision Workshops (WACVW). Lake Placid, USA: IEEE, 2016. 1−9

20 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **39**(6): 1137−1146

21 Gall J, Lempitsky V. Class-specific Hough forests for object detection. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2013. 1022−1029

22 Yeung S, Russakovsky O, Mori G, Li F F. End-to-end learning of action detection from frame glimpses in videos. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 2678−2687

23 Redmon J, Farhadi A. YOLO9000: better, faster, stronger [Online], available: https://arxiv.org/abs/1612.08242, December 30, 2016

24 Körtner T. Ethical challenges in the use of social service robots for elderly people. *Zeitschrift Für Gerontologie Und Geriatrie*, 2016, **49**(4): 303−307

25 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

26 Girshick R, Donahue J, Darrell T, Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(1): 142−158

27 Gao W, Zhou Z H. Dropout rademacher complexity of deep neural networks. *Science China Information Sciences*, 2016, **59**: Article No. 072104

28 Tzutalin. LabelImg [Online], available: https://github.com/tzutalin/labelImg, November 6, 2016

29 Abadi M, Agarwal A, Barham P, Zheng X Q. TensorFlow: large-scale machine learning on heterogeneous distributed systems [Online], available: http://download.tensorflow.org/paper/whitepaper2015.pdf. November 12, 2015

30 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2015. 779−788

杨观赐 贵州大学现代制造技术教育部重点实验室教授. 主要研究方向为智能与自主机器人, 计算智能与智能系统.
E-mail: guanci_yang@163.com
(**YANG Guan-Ci** Professor at the Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University. His research interest covers intelligent autonomous social robots, computational intelligence, and intelligent systems.)

杨 静 贵州大学现代制造技术教育部重点实验室硕士研究生. 主要研究方向为智能视觉计算, 智能与自主服务机器人. 本文通信作者.
E-mail: yang_jing0903@163.com
(**YANG Jing** Master student at the Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University. His research interest covers intelligent vision computing and intelligent autonomous social robots. Corresponding author of this paper.)

苏志东 贵州大学现代制造技术教育部重点实验室硕士研究生. 主要研究方向为自然语言处理, 智能与自主服务机器人. E-mail: suzhidong2016@163.com
(**SU Zhi-Dong** Master student at the Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University. His research interest covers natural language processing and intelligent autonomous social robots.)

陈占杰 贵州大学现代制造技术教育部重点实验室硕士研究生. 主要研究方向为机器人自动建图与导航技术, 智能与自主服务机器人.
E-mail: chenzhanjie0320@163.com
(**CHEN Zhan-Jie** Master student at the Key Laboratory of Advanced Manufacturing technology, Ministry of Education, Guizhou University. His research interest covers simultaneous localization and mapping, and intelligent autonomous social robots.)