

基于图匹配网络的小样本违禁物品分割算法

朱祯悦 吕淑静 吕岳

Few-shot prohibited item segmentation algorithm based on graph matching network

Zhu Zhenyue, Lv Shujing, Lv Yue

在线阅读 View online: <https://doi.org/10.3788/IRLA20210075>

您可能感兴趣的其他文章

Articles you may be interested in

合并分割块的点云语义分割方法

Point cloud semantic segmentation method based on segmented blocks merging

红外与激光工程. 2021, 50(10): 20200482-1-20200482-10 <https://doi.org/10.3788/IRLA20200482>

基于U-Net网络的多主动轮廓细胞分割方法研究

Multi-active contour cell segmentation method based on U-Net network

红外与激光工程. 2020, 49(S1): 20200121-20200121 <https://doi.org/10.3788/IRLA20200121>

基于双模全卷积网络的行人检测算法(特邀)

Pedestrian detection algorithm based on dual-model fused fully convolutional networks(Invited)

红外与激光工程. 2018, 47(2): 203001-0203001(8) <https://doi.org/10.3788/IRLA201847.0203001>

基于C-TOF成像的位姿测量与地物目标识别技术研究

Research on pose measurement and ground object recognition technology based on C-TOF imaging

红外与激光工程. 2020, 49(1): 0113005-0113005(9) <https://doi.org/10.3788/IRLA202049.0113005>

基于图论的单线激光雷达数据匹配方法

Single-line LiDAR data matching method based on graph theory

红外与激光工程. 2018, 47(12): 1226001-1226001(9) <https://doi.org/10.3788/IRLA201847.1226001>

采用卷积神经网络的红外和可见光图像块匹配

Infrared-visible image patches matching via convolutional neural networks

红外与激光工程. 2021, 50(5): 20200364-1-20200364-9 <https://doi.org/10.3788/IRLA20200364>

基于图匹配网络的小样本违禁物品分割算法

朱祯悦¹, 吕淑静^{1,2*}, 吕岳^{1,2}

- (1. 华东师范大学 计算机科学与技术学院, 上海 200062;
2. 上海市多维度信息处理重点实验室, 上海 200241)

摘要: 自动化安检技术是维护公共安全、提升安检效率的一项有效措施。在实际场景中很难获得充足的违禁品标注样本用于神经网络的训练, 并且在不同场景和安全级别下违禁品的类别也有所不同。为解决基于神经网络的违禁品检测方法所面临的样本不均衡问题, 以及避免模型在分割新的违禁品类时重新训练的现象, 文中提出一种基于图匹配网络的小样本违禁物品分割算法。文中模型将测试图像与参考图像并行输入到图匹配网络中, 并根据匹配结果从测试图像中分割出违禁品。所设计的图匹配模块不仅从图间节点的相似性考虑匹配问题, 并利用 DeepEMD 算法建立全局概念, 进一步提高测试图和参考图的匹配结果。在 SIXray 数据集和 Xray-PI 数据集上的实验表明: 本模型在单样本分割任务中得到 36.4% 和 51.2% 的类平均交并比, 分别比目前先进的单样本分割方法提高 2.5% 和 2.3%。由此表明所设计的算法能有效提升小样本 X 光图像分割算法的精确度。

关键词: 语义分割; 小样本学习; 图匹配网络; X 光图像; 违禁品检测

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA20210075

Few-shot prohibited item segmentation algorithm based on graph matching network

Zhu Zhenyue¹, Lv Shujing^{1,2*}, Lv Yue^{1,2}

- (1. Department of Computer Science and Technology, East China Normal University, Shanghai 200062, China;
2. Shanghai Key Laboratory of Multidimensional Information Processing, Shanghai 200241, China)

Abstract: Automated security inspection is an effective measure to maintain public safety and improve the efficiency of security inspection. Usually, it is difficult to obtain enough labelled samples which contain some prohibited items of rarely appearing. Furthermore, the category of prohibited items varies in different scenarios and security levels. A graph matching network algorithm for few-shot prohibited item segmentation was introduced to deal with the imbalance of training samples faced by neural network methods, and to inspect prohibited items of new categories without the requirement of retraining. This model parallelly input a query image and several support images into the graph matching network, and segmented the prohibited items from the query image according to the matching results. The graph matching module not only considered the matching problem from the point of node similarity between two graphs, but also established a global concept to match the graphs with the use of DeepEMD algorithm. Experiments on the SIXray dataset and Xray-PI dataset show that proposed model achieves 36.4% and 51.2% meanIoU for 1-shot tasks and outperforms the state-of-the-art method by 2.5% and 2.3% meanIoU, respectively. The extended experiments demonstrate that proposed algorithm can

收稿日期: 2021-01-27; 修订日期: 2021-04-13

作者简介: 朱祯悦, 女, 硕士生, 主要从事目标检测和计算机视觉方面的研究。

导师简介: 吕岳, 男, 教授, 博士生导师, 博士, 主要从事图像处理、模式识别与机器学习方面的研究。

通讯作者: 吕淑静, 女, 研究员, 博士, 主要从事图像处理、模式识别与机器学习方面的研究。

effectively improve the accuracy of few-shot X-ray image segmentation.

Key words: semantic segmentation; few-shot learning; graph matching network; X-ray image; prohibited item inspection

0 引言

在机场、火车站等重要场所对行李物品等进行安检过程中, 安检人员需要在 X 光图像中识别出违禁物品。对于训练有素的安检人员来说, 判断一张 X 光图像中是否存在违禁物品是件较为容易的事。但随着工作时长的增加, 难免会精神疲倦、注意力下降, 这便为安检带来了隐患。因此, 违禁物品自动识别获得越来越多的关注并已成为一个新热点。

Akay 等人首次对 X 光图像提取深度特征^[1], 并利用迁移学习以帮助模型完成 X 光图像的二分类任务; Jaccard 等人^[2]进一步通过实验论证卷积神经网络 (CNNs) 所提取的 X 光图像特征相较于传统描述算子包含更多信息; Singh 等人提出一种提取图像属性的神经网络框架以优化针对空运行李的 X 光图像分割算法^[3]; Bhowmik 等人对图像中的物体及其组件进行分割, 采用有监督的方式对 X 光图像进行异常检测^[4]; 随着注意机制的发展, An 等人提出一种基于注意力的分割模型用于提取更丰富的 X 光图像特征^[5]。虽然以上的深度学习方法都取得了较好的分割性能, 但它们都依赖于大量的样本进行训练, 并且这些模型只能处理在训练阶段出现过的违禁品类别。而在公共交通背景下, 人们可能无法收集大量的含有违禁品的标注图像。此外, 违禁品的类别也未有一个统一的标准, 在不同的场景或安全级别下, 禁止携带的物品种类也会随之改变。因此, 探索新的 X 光图像分割方法是迫切需要的。

相对于机器学习算法, 人类在见过几个示例后便能准确地图像中识别出之前从未见过的物体。小样本学习的研究便由人类和机器学习算法之间的差距引申而出, 旨在学习出一个仅使用少量标注样本便可预测新类别的模型。目前小样本学习已广泛应用于分类方法中, 包括记忆性循环神经网络^[6]、迁移学习^[7]、网络参数预测^[8]以及度量学习^[9]。与大家工作最息息相关的是关系网络^[10], 它是度量学习的一个分支。关系网络提供了一个可学习的且非线性的分类器用于比较特征间的相似性, 该分类器由多个卷积操

作组成。Shaban 等人首次提出用双分支结构解决小样本语义分割任务^[11], 该方法将标注样本所提取的特征直接作为预测测试图像最后一层输出层的权重; K. Rakelly 等人提出标注样本可生成一个嵌入式特征与测试图像进行特征融合^[12]; CANet 算法^[13]不同于之前的工作, 该网络使用孪生网络提取 CNN 特征, 模型中的稠密比较模块可以被看作是一个以稠密形式扩展的用于解决分割任务的关系网络。

以上所提出的双分支网络与图匹配网络结构相似。已知图匹配网络以一对图作为模型的输入, 利用节点和图相似度等信息进行匹配并保留两图中相似的部分。Li 等人引入基于注意力的图匹配机制^[14], 将两个图之间的节点相似度视为权重, 使匹配网络更专注于其中一个图节点与另一个图中最相似节点之间的差异; Sarlin 等人通过基于注意力的聚合方式^[15]对两组局部特征进行最优匹配。由于关系网络与图匹配网络皆从相似性的角度出发分类特征, 因此 Zhang 等人提出一种金字塔图网络模型^[16], 通过利用小样本学习和图注意力机制在仅含有少量参考图像的情况下对新物体类别进行分割。

不同于金字塔图网络模型^[16]中将测试图和参考图并行输入多个注意力模块对图间节点一一匹配, 文中所提出的算法将测试图和参考图一同输入到单个注意力模块, 以计算测试图节点与参考图节点之间的相似度。另一方面, 文中模型采用 DeepEMD 算法^[17]来计算两个图之间的相似度, 以此代替使用两图平均节点之间的距离表征整体相似度的方法。笔者认为这种做法可以更详细、更精确地表达测试图和参考图之间的全局匹配程度。文中还为图匹配模块设计出一个图相似度损失函数, 训练图相似度的计算结果更贴近真实数值。

文中提出一种用于小样本 X 光图像分割的图匹配网络, 以解决训练样本不均衡的问题, 并在无需重复训练模型的情况下对新加入的违禁品类别进行分割。文中模型加入图匹配模块, 从节点相似度和图相似度两个角度解决图匹配问题。文中设计出一个图相似度损失函数用以纠正测试图和参考图之间的相

似度计算。在 SIXray 数据集和 Xray-PI 数据集上的实验表明:所提模型在单样本分割任务中得到 36.4% 和 51.2% 的类平均交并比, 分别比 SOTA 单样本分割方法提高了 2.5% 和 2.3%。

1 基于图匹配网络的违禁物品分割模型

文中所设计的方法是从局部和全局两个视角比较测试图和参考图之间的相似性。首先, 将测试图和参考图输入节点注意力模块, 从局部的角度出发筛选出与参考图相关性高的测试图节点, 并以此去除测试图中背景节点的噪声干扰。随后, 从全局的角度出发计算测试子图与参考子图之间的图相似度, 通过比较两子图特征的分布规律考察测试子图中是否含有指定的违禁品。

1.1 模型概览

文中算法提出的小样本语义分割模型包含三个阶段: 图嵌入、图匹配和分割。下文将以单样本语义分割为例介绍违禁品检测的算法流程。模型的总体架构如图 1 所示, 文中算法采用孪生网络提取测试图像与参考图像的 CNN 特征并由此建立起图结构。而后将测试图和参考图并行输入计算节点相似度的节点注意力模块。根据求得的节点相似度信息, 在测试图中筛选出可能是违禁品区域的图节点并将筛选出的图节点构建成测试子图, 参考子图则利用参考图的掩模信息构建而成。随后, 模型采用 DeepEMD 算法^[18]计算测试子图与参考子图之间的图相似度。文中算法将图相似度高于一定阈值的测试图预测掩模输入到 ASPP 模块中获取多尺度信息。最后, 该算法将像素级分类后的概率图二值化以获得最终的分割结果。

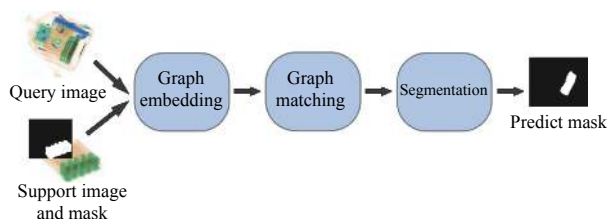


图 1 单样本违禁物品分割模型结构图

Fig.1 Overview of proposed framework in the 1-shot prohibited item segmentation

1.2 图嵌入

在单样本语义分割模型的图嵌入阶段, 模型的输

入是一幅测试图像 $x_q \in R^{H \times W \times 3}$ 和一幅带掩模 $Mask_s \in \{0,1\}^{H \times W \times 1}$ 的参考图像 $x_s \in R^{H \times W \times 3}$ 。所提模型将所输入的一对测试图像和参考图像分别建立各自的全连接图。其中, 每一个图节点皆为卷积神经网络对图像中的某一块感受野所提取的 CNN 特征。

为提取图像特征, 文中模型采用 ResNet-50^[19] 作为特征提取器, 它共包含四个模块。而该模型仅采用前三个模块所提取的特征构建图节点, 并舍弃第三个模块之后所提取的特征, 其原因在于通过 CNN 特征可视化^[18] 可观察到, 越前列的模块所提取的特征更靠近底层, 而越靠后的模块所提取的特征更为抽象。

值得注意的是, 为保持特征图的空间分辨率, 特征提取器的第二个模块中加入了膨胀卷积, 该膨胀卷积的膨胀率为 2。为使特征提取器的三个模块所提取的特征能一一对应, 经这三个模块所提取的特征被重新规整尺寸, 其长和宽都被缩减至原始图像的八分之一, 由此, 模型所构建的图节点便由特征提取器的第一个模块 $f_{block1}(\cdot)$ 、第二个模块 $f_{block2}(\cdot)$ 和第三个模块 $f_{block3}(\cdot)$ 所提取的特征拼接得来。具体公式为:

$$g_q^i = f_{block1}(x_q^i) \parallel f_{block2}(x_q^i) \parallel f_{block3}(x_q^i) \quad i \in [1, H'W'] \quad (1)$$

$$g_s^i = f_{block1}(x_s^i) \parallel f_{block2}(x_s^i) \parallel f_{block3}(x_s^i) \quad i \in [1, H'W'] \quad (2)$$

式中: \parallel 为向量拼接操作; g_q, g_s 分别为测试图像和参考图像所构建的图节点; x_q, x_s 分别为测试图像和参考图像的特征提取感受野; $H'W'$ 分别为原始图像长和宽的八分之一。

经以上步骤便可将测试图像转换为测试图 $G_q = \{g_q^i\}_{i=1}^{H'W'}$, 参考图像转换为参考图 $G_s = \{g_s^i\}_{i=1}^{H'W'}$ 。相对应的参考图像的掩模也被调整至合适大小 $M_s = \{m_s^i\}_{i=1}^{H'W'}$ 作为对应参考图节点的标签。以上表述的是单样本语义分割模型的图嵌入步骤, 当参考图像数量增多时, 由不同参考图像构建而得的图节点可以合并, 从而扩展至更大的参考图。

1.3 图匹配

1.3.1 节点注意力模块

在获得测试图和参考图后, 将其一同输入到节点注意力模块, 如图 2 所示。文中模型根据计算得到的测试图和参考图节点之间的相似性信息筛选出可能含有违禁品区域的测试图节点并建立测试子图。节点相似度矩阵 $S_N = \{s_{ij} \in R\}_{i,j=1}^{H'W'}$ 是由测试图中的每一

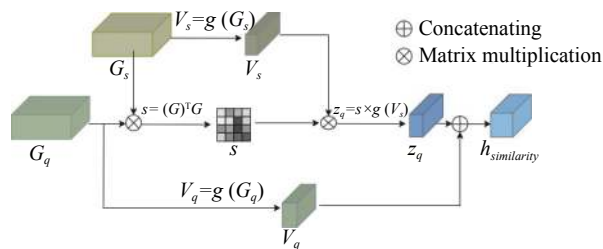


图 2 节点注意力模块结构图

Fig.2 Structure diagram of the node attention module

个节点与参考图节点进行相似度计算得来,其计算公式如下:

$$s_{ij} = (g_q^i)^T g_s^j, \quad i, j \in [1, H'W'] \quad (3)$$

式中: s_{ij} 为第 i 个测试图节点和第 j 个参考图节点间的相似度。

权重矩阵随后与参考图节点的标签 M_s 点乘, 以去除由参考图中的背景节点所引起的噪声。而后利用 softmax 函数对权重矩阵进行规范化处理, 具体公式如下:

$$s_{ij} = \frac{\exp(s_{ij})}{\sum_{j=1}^{H'W'} \exp(s_{ij})} \quad (4)$$

为提炼图节点中的信息, 文中算法采用卷积操作分别对测试图和参考图的特征维度进行压缩。

$$V_q = g(G_q), V_s = g(G_s) \quad (5)$$

式中: G_q, G_s 分别为压缩前的测试图和参考图; V_q, V_s 分别为压缩后的测试图和参考图; $g(\cdot)$ 为卷积操作。

而后, 将计算得到的节点相似度矩阵与带标签的参考图节点相乘获得重建测试图。计算公式如下:

$$z_q^i = \sum_{j=1}^{H'W'} s_{ij} v_s^j, \quad i \in [1, H'W'] \quad (6)$$

式中: s 为某个测试图节点与参考图节点之间的相似度; v_s 为 s 所对应的参考图节点; z_q 为 s 所对应的重建后的测试图节点。

随后通过融合重建前后的两组测试图可以获得采用相似度作为权重进行筛选后的测试节点标签信息。融合公式如下:

$$h_{similarity}^i = \varphi(z_q^i || v_q^i), \quad i \in [1, H'W'] \quad (7)$$

式中: $h_{similarity}$ 为预测的测试节点标签信息; $\varphi(\cdot)$ 为卷积操作。

而后建立测试子图, 子图中将尽可能保留含有违禁品的图节点并将被筛去的图节点数值清零, 参考子

图的建立方式与测试子图相似, 只需依照参考图的标签信息筛选节点即可。为便于后续的图相似度计算, 文算法将进一步对测试子图与参考子图进行滤波操作以提炼信息并调整子图尺寸。滤波公式如下:

$$U_q = F(V_q \cdot h_{similarity}) \quad (8)$$

$$U_s = F(V_s \cdot M_s) \quad (9)$$

式中: U_q 为测试子图; U_s 为参考子图; $F(\cdot)$ 为滤波函数。滤波半径和步长的计算公式为:

$$stride = \text{floor}\left(\frac{input_size}{N}\right) U_q \quad (10)$$

$$kernel_size = input_size - (N - 1) \times stride \quad (11)$$

式中: $input_size$ 为子图的初始边长; N 为预期调整后子图边长; $\text{floor}(\cdot)$ 为向下取整函数; $stride$ 为滤波步长; $kernel_size$ 为滤波半径。

1.3.2 DeepEMD 算法

在获得经过尺寸调整的测试子图与参考子图后, 文中利用 DeepEMD 算法^[17] 计算两子图之间的图相似度。首先 DeepEMD 算法^[17] 是由 EMD 算法^[20] 融入到神经网络中衍化得来, EMD 算法^[20] 测量的是两个概率分布之间的距离。

在 DeepEMD 算法中^[17], 测试图将作为源分布, 测试图中的节点为 i , 其特征向量为 u_i , 权值为 s_i 。同时, 参考图将作为目标分布, 参考图中的节点为 j , 其特征向量为 v_j , 权值为 d_j 。 c_{ij} 表示测试图中的节点 i 与参考图中的节点 j 之间的特征距离。在图匹配算法中, 节点 i 转换为节点 j 的成本由两节点的权重共同决定。

$$c_{ij} = 1 - \frac{u_i^T v_j}{|u_i| |v_j|} \quad (12)$$

结合 DeepEMD 算法^[17], 文中将源分布设置为测试子图 U_q , 目标分布为参考子图 U_s , 根据公式 (12) 计算测试子图与参考子图中各节点之间的距离后, 文中还将计算每个节点对应的权重:

$$w_i = \max\left(u_q^i \cdot \frac{\sum_{j=1}^{N^2} u_s^j}{N^2}, 0\right) \quad (13)$$

式中: u_q^i 为测试子图中的某个节点; u_s^j 为参考子图中的某个节点; $\max(\cdot)$ 为取最大值函数; w_i 为查询子图中节点 i 对应的权重。该公式表明, 测试子图节点的权重与该节点同参考子图的相关性成正比, 即两个子图中

包含的相似区域的权值较高,而背景区域的权值较小。

随后,文中算法对子图中所有节点的权值进行归一化,使测试节点的权值之和等于参考节点的权值之和。

$$\widehat{w}_i = w_i \frac{N^2}{\sum_{j=1}^{N^2} w_j} \quad (14)$$

式中: \widehat{w}_i 为节点 i 的最终权值。而测试子图与参考子图之间的距离为:

$$s(U_q, U_s) = \sum_{i=1}^{N^2} \sum_{j=1}^{N^2} (1 - c_{ij}) |w_i - w_j| \quad (15)$$

式中: c_{ij} 为测试节点 i 到参考节点 j 的距离; w_j 为参考节点 j 的权重; $s(U_q, U_s)$ 为两个子图之间的图相似度。

1.4 分割

在模型的分割阶段,保留高于图相似度阈值所对应的预测测试节点标签信息,并将其输入空间金字塔池模块 (ASPP)^[21]。该模块包含全局平均池化层和 1×1 卷积层用于提取图像级特征,同时 ASPP 模块采用一个 1×1 卷积层和三个不同采样率的放大卷积层来获取响应图的不同尺度信息。将 ASPP 模块生成的上述不同尺度特征用 1×1 卷积层融合后,再通过双线

性上采样便能得到最终的分割结果。

1.5 损失函数

在训练阶段,文中算法采用 Dice Loss^[22] 作为损失函数作用于测试图像的所有像素区域。

$$L_{dice} = 1 - \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (16)$$

式中: X 为所预测的测试图像掩模拉伸至一维后的列向量; Y 为参考图像的真值掩模拉伸至一维后的列向量。

在此基础上,文中提出图相似度损失函数,使预测的测试节点标签信息能更贴近于真值。

$$L_{graph} = -label \times \log(s(U_q, U_s)) - (1 - label) \times (1 - s(U_q, U_s)) \quad (17)$$

式中: $s(U_q, U_s)$ 为两个子图之间的图相似度; $label$ 为图相似度真值。若测试图像中含有违禁品,则其值为 1,反之为 0。以上两个损失函数以相加的方式结合作为文中模型最终的损失函数。

2 实验

2.1 参数设置

文中在表 1 中展示出模型的完整参数配置。其中, #maps 表示卷积核数量, k 表示卷积窗口大小, s 表

表 1 基于图匹配网络违禁物品分割模型的参数设置

Tab.1 Parameter setting of prohibited item segmentation model based on graph matching network

Operational layer	Configuration	
Input image	321×321×3	
Convolution layer	#maps: 64, k : 7×7, s : 2×2	
Maxpool layer	w : 3×3, s : 2×2	
Graph embedding	$\left[\begin{array}{l} \text{\#maps: 64, } k: 1 \times 1 \text{ } s: 1 \times 1 \\ \text{\#maps: 64, } k: 3 \times 3 \text{ } s: 1 \times 1 \\ \text{\#maps: 256, } k: 1 \times 1 \text{ } s: 1 \times 1 \end{array} \right] \times 3$	
	$\left[\begin{array}{l} \text{\#maps: 128, } k: 1 \times 1 \text{ } s: 2 \times 2 \\ \text{\#maps: 128, } k: 3 \times 3 \text{ } s: 2 \times 2 \\ \text{\#maps: 512, } k: 1 \times 1 \text{ } s: 2 \times 2 \end{array} \right] \times 4$	
	$\left[\begin{array}{l} \text{\#maps: 256, } k: 1 \times 1 \text{ } s: 1 \times 1 \\ \text{\#maps: 256, } k: 3 \times 3 \text{ } s: 1 \times 1 \\ \text{\#maps: 1024, } k: 1 \times 1 \text{ } s: 1 \times 1 \end{array} \right] \times 6$	
Convolution layer	#maps: 256, k : 1×1, s : 1×1	
Graph matching	Convolution layer	#maps: 256, k : 1×1, s : 1×1
	Avgpool layer	w : 11×11, s : 1×1
	Convolution layer	#maps: 256, k : 1×1, s : 1×1
	Convolution layer	#maps: 256, k : 1×1, s : 1×1
	Maxpool layer	w : 10×10, s : 1×1
Segmentation	Convolution layer	#maps: 256, k : 1×1, s : 1×1
	Convolution layer	$\left[\begin{array}{l} \text{\#maps: 256, } k: 3 \times 3 \text{ } s: 1 \times 1 \\ \text{\#maps: 256, } k: 3 \times 3 \text{ } s: 1 \times 1 \end{array} \right] \times 3$
	Convolution layer	#maps: 1, k : 1×1, s : 1×1

示步长, w 表示池化窗口大小。在节点注意力模块中, 拼接测试图节点和参考图节点的卷积操作 $\varphi(\cdot)$ 共享参数; 压缩测试图节点和参考图节点特征的卷积操作 $g(\cdot)$ 也同样共享参数。

在训练阶段, 模型通过 SGD 优化器训练网络参数以达到最小化损失函数的目的。文中算法的图相似度阈值设置为 0.5, 模型训练的学习率被设置为 0.0025, 权值衰减设置为 0.0005, batch 的大小设置为 4。在训练阶段该模型的训练时长为 25 h, 平均每轮训练 15 min, 在测试阶段模型处理一张 X 光图像的时长为 10 ms。模型在 Nvidia 2080Ti GPU 上训练与测试。

2.2 实验结果

文中模型分别在 SIXray 数据集和 Xray-PI 数据集上进行实验, 这两个数据集中的 X 光数据集皆为伪彩色图像。由于数据集中的 X 光图像并非由同一款安检机拍摄得来, 因此即使是同一物质的物体, 在不同伪彩色 X 光图像之间也存在色差, 故文中的所有实验在预处理阶段皆引入图像增强的方法^[23] 纠正严重偏移的色差。

以下将分别对这两个数据集进行简要介绍。

(1) SIXray 数据集

SIXray 数据集是由 Miao 等人提供的伪彩色 X 射线图像数据集^[24], 该数据集原是由于目标分类与定位的研究, 因此仅对 X 光图像提供违禁品的边框标注。SIXray 数据集共包含 8929 张含有违禁品的 X 光图像, 该数据集中共含五类违禁品, 各类违禁品的 X 光图像具体数量如下: 手枪 (Gun) 3131 张, 小刀 (Knife) 1943 张, 扳手 (Wrench) 2199 张, 钳子 (Pliers) 3961 张, 剪刀 (Scissors) 983 张。

(2) Xray-PI 数据集

由于目前缺乏像素级标记的 X 光数据集, 因此实验将由 X 光安检机采集得来的伪彩色图像进行像素级标注, 并对所建立的 X 光数据集命名为 Xray-PI。与 SIXray 数据集相比, 该数据集的标注信息更精细, 违禁品类别也更加多样化。Xray-PI 数据集共包含 385 张含有违禁品的 X 光图像, 该数据集中共包含七类违禁品, 各类违禁品的 X 光图像具体数量如下: 扳手 (Wrench) 34 张, 瓶子 (Bottle) 86 张, 烟花 (Firework) 39 张, 枪 (Gun) 70 张, 鞭炮 (Firecracker) 64 张, 钳子 (Pliers) 43 张, 刀片 (Blade) 49 张。

实验的训练策略是在指定的 X 光数据集中选择含有某一类违禁品的 X 光图像作为测试集, 含有其他类别违禁品的 X 光图像作为训练集。模型在训练集上学习网络参数, 而后在测试集中挑选少量标注样本作为分割测试集中 X 光图像的参考图像。此训练策略可以保证模型参数不受测试阶段所分割的违禁品类别的影响。实验将在训练集和测试集中采用随机剪裁、翻转、缩放等方式扩充图像数量, 而后分别在训练集和测试集中随机选择一定数量的图像来验证模型的性能。

实验通过计算测试图像的真值掩模 $pred$ 与预测分割结果 $mask$ 之间的交并比作为该次预测正确与否的衡量指标, 并用类平均交并比体现模型在指定数据集上预测分割的整体性能。交并比的定义公式如下:

$$IoU = \frac{mask \cap pred}{mask \cup pred} \quad (18)$$

2.2.1 模型性能测试

文中算法可将单张测试图像与多张样本图像一同输入模型中进行图匹配。在单样本条件下模型前向传播时所需的计算力为 43.7407 G, 参数总量占 15.6254 M 内存。随着样本图像数量的增加, 模型前向传播时所需的计算力将成倍数增长。

模型在获得测试子图与参考子图后需借助滤波规整子图尺寸, 实验通过控制变量的方式比较在不同滤波处理条件下, 选定不同的子图边长对单样本模型预测的分割结果会产生何种影响, 并根据实验结果选择合适的滤波器及子图边长。

如表 2 中数据所示, 模型采用最大值滤波的分割效果优于采用均值滤波的分割效果, 而在最大值滤波条件下, 当测试子图和参考子图的边长为 10 时, 其分

表 2 模型在不同滤波器和子图边长条件下的分割性能
Tab.2 Segmentation performance of model with different filters and length of subgraphs

Filter	Size	meanIoU	
		Xray-PI	SIXray
Average filter	5	48.8%	34.0%
	10	50.4%	35.8%
	15	49.7%	35.5%
Maximum filter	5	50.1%	34.8%
	10	51.2%	36.4%
	15	50.4%	35.3%

割效果最佳。因此,模型选择使用最大值滤波将测试子图与参考子图规整至边长为 10 的尺寸,以此为后续图计算做准备。

2.2.2 在 SIXray 数据集上的实验结果

图 3 为模型在 SIXray 数据集上获得的实验效果图。模型将与 CANet^[13] 和 PGNet^[16] 进行比较,观察表 3 中的信息可知文中方法在单样本和五样本分割任务中类平均交并比分别高于 CANet^[13]2.5% 和 2.2%, 高于 PGNet^[16]2.6% 和 2.5%。模型在五样本分割任务中类平均交并比比单样本分割任务高 1.3%。

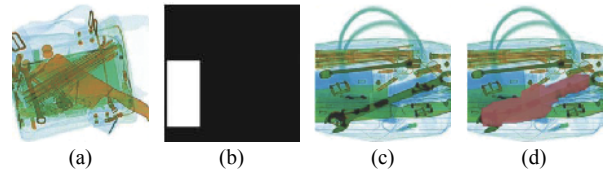


图 3 在 SIXray 数据集上的实验效果图。(a) 参考图像;(b) 参考图像掩模型;(c) 测试图像;(d) 分割结果,其中红色区域为预测违禁品区域

Fig.3 Experimental effect results on SIXray dataset. (a) Support image; (b) Support image mask; (c) Query image; (d) Segmentation result, of which red region is the predicted prohibited item region

表 3 在 SIXray 数据集上进行单样本和五样本分割性能

Tab.3 Segmentation performance of 1-shot task and 5-shot task on SIXray dataset

Methods		Gun	Knife	Wrench	Pliers	Scissors	meanIoU
1-shot	CANet ^[13]	40.3	41.6	35.2	33.9	18.3	33.9%
	PGNet ^[16]	38.9	41.5	37.4	33.2	18.0	33.8%
	Ours	41.4	42.1	35.6	34.0	28.5	36.4%
5-shot	CANet ^[13]	43.0	43.2	36.8	35.4	18.9	35.5%
	PGNet ^[16]	41.1	42.9	37.0	35.7	19.1	35.2%
	Ours	43.7	43.4	36.3	35.9	29.3	37.7%

2.2.3 在 Xray-PI 数据集上的实验结果

在 Xray-PI 数据集上进行的实验效果如图 4 所示,该模型同样与 CANet^[13] 和 PGNet^[16] 这两种先进的小样本语义分割算法进行比较。根据表 4 中的实验结果可证明本方法的类平均交并比得分比 CANet^[13] 在单样本分割任务和五样本分割任务中分别高出 2.3% 和 1.9%。同时,文中方法的类平均交并比得分比 PGNet^[16] 在单样本分割任务中高出 5.7%,在五样本分割任务中高出 5.2%。该模型在五样本分割任务中类平均交并比比单样本分割任务高 1.7%。

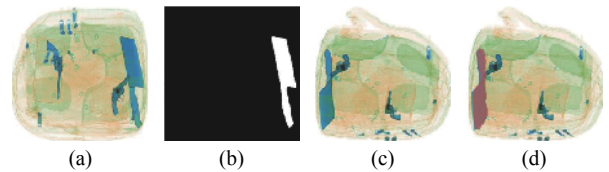


图 4 在 Xray-PI 数据集上的实验效果图。(a) 参考图像;(b) 参考图像掩模型;(c) 测试图像;(d) 分割结果,其中红色区域为预测违禁品区域

Fig.4 Experimental effect results on Xray-PI dataset. (a) Support image; (b) Support image mask; (c) Query image; (d) Segmentation result, of which red region is the predicted prohibited item region

表 4 在 Xray-PI 数据集上进行单样本和五样本分割性能

Tab.4 Segmentation performance of 1-shot task and 5-shot task on Xray-PI dataset

Methods		Firework	Firecracker	Bottle	Gun	Wrench	Pliers	Blade	meanIoU
1-shot	CANet ^[13]	51.4	45.5	42.2	48.1	34.7	53.0	67.9	48.9%
	PGNet ^[16]	44.1	41.9	31.8	47.1	36.3	51.1	66.4	45.5%
	Proposed	52.9	45.7	47.4	51.2	37.5	55.5	68.7	51.2%
5-shot	CANet ^[13]	54.8	47.6	45.2	49.5	35.6	56.1	68.0	51.0%
	PGNet ^[16]	46.0	43.0	35.1	48.5	37.1	55.6	68.3	47.7%
	Proposed	55.4	49.1	48.7	53.6	38.5	56.4	68.9	52.9%

3 结 论

文中提出基于图匹配网络的小样本违禁物品分割算法,该算法的设计初衷是为解决样本不均衡问题,以及避免训练后的模型在分割新的违禁品类时需要重新训练的现象。文中提出从节点与图这两个角度出发匹配图模型,从而找到两图间相似的含有违禁品的区域,同时设计了一个图相似度损失函数以提升图相似度计算的可靠性。在 SIXray 数据集和 Xray-PI 数据集上进行的实验证明该方法在未训练过的测试集 X 光图像中能有效分割出违禁品。

在五样本分割任务中,文中模型前向传播时所需的计算力是单样本分割任务的五倍,但类平均交并比仅比单样本分割任务高出 1.5% 左右。虽然本模型消耗了成倍的计算力,却未能大幅提升类平均交并比,因此该算法在单张测试图像和多张样本图像的匹配方式上仍有待改进。

参考文献:

- [1] Akçay S, Kundegorski M E, Devereux M, et al. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery[C]//Conference on International Conference on Image Processing, 2016: 1057-1061.
- [2] Jaccard N, Rogers T W, Morton E J, et al. Tackling the X-ray cargo inspection challenge using machine learning[C]//Conference on Anomaly Detection and Imaging with X-Rays. International Society for Optics and Photonics, 2016, 9847: 98470N.
- [3] Singh M, Singh S. Image segmentation optimisation for X-ray images of airline luggage[C]//Conference on Computational Intelligence for Homeland Security and Personal Safety, 2004: 10-17.
- [4] Bhowmik N, Gaus Y F A, Akçay S, et al. On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within X-ray security imagery[C]//Conference on Machine Learning and Applications, 2019: 986-991.
- [5] An J, Zhang H, Zhu Y, et al. Semantic segmentation for prohibited items in baggage inspection[C]//Conference on Intelligent Science and Big Data Engineering, 2019: 495-505.
- [6] Yang N, Nan L, Zhang D Y, et al. Research on image interpretation based on deep learning [J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203002. (in Chinese)
- [7] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks [J]. *arXiv preprint*, 2017: arXiv1703.03400.
- [8] Xue S, Zhang Z, Lv Q Y, et al. Image recognition method of anti UAV system based on convolutional neural network [J]. *Infrared and Laser Engineering*, 2020, 49(7): 20200154. (in Chinese)
- [9] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C]//Conference on Advances in Neural Information Processing Systems, 2017: 4077-4087.
- [10] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[C]//Conference on Computer Vision and Pattern Recognition, 2018: 1199-1208.
- [11] Shaban A, Bansal S, Liu Z, et al. One-shot learning for semantic segmentation [J]. *arXiv preprint*, 2017: arXiv1709.03410.
- [12] Rakelly K, Shelhamer E, Darrell T, et al. Conditional networks for few-shot semantic segmentation[C]//Conference on Learning Representations Workshop, 2018.
- [13] Zhang C, Lin G, Liu F, et al. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning[C]//Conference on Computer Vision and Pattern Recognition, 2019: 5217-5226.
- [14] Li Y, Gu C, Dullien T, et al. Graph matching networks for learning the similarity of graph structured objects [J]. *arXiv preprint*, 2019: arXiv1904.12787.
- [15] Sarlin P E, DeTone D, Malisiewicz T, et al. SuperGlue: Learning feature matching with graph neural networks[C]//Conference on Computer Vision and Pattern Recognition, 2020: 4938-4947.
- [16] Zhang C, Lin G, Liu F, et al. Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 9587-9595.
- [17] Zhang C, Cai Y, Lin G, et al. DeepEMD: Few-shot image classification with differentiable earth mover's distance and structured classifiers[C]//Conference on Computer Vision and Pattern Recognition, 2020: 12200-12210.
- [18] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision, 2014: 818-833.

- [19] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [20] Rubner Y, Tomasi C, Guibas L J. A metric for distributions with applications to image databases[C]//Proceedings of the IEEE International Conference on Computer Vision, 1998: 59-66.
- [21] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation [J]. *arXiv preprint*, 2017: arXiv1706.05587.
- [22] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.
- [23] Piao Y, Liu L, Liu X Y. Enhancement technology of video under low illumination [J]. *Infrared and Laser Engineering*, 2014, 43(6): 2021-2026. (in Chinese)
- [24] Miao C, Xie L, Wan F, et al. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images[C]//Conference on Computer Vision and Pattern Recognition, 2019: 2119-2128.