



关键点图对比图像分类方法

卢毅, 陈亚冉, 赵冬斌, 刘暴, 来志超, 王超楠

引用本文:

卢毅,陈亚冉,赵冬斌,刘暴,来志超,王超楠. 关键点图对比图像分类方法[J]. *智能系统学报*, 2023, 18(1): 36–46.

LU Yi, CHEN Yaran, ZHAO Dongbin, LIU Bao, LAI Zhichao, WANG Chaonan. Keypoint-based graph contrastive neural network for image classification[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(1): 36–46.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202112001>

您可能感兴趣的其他文章

融合VAE和StackGAN的零样本图像分类方法

Zero-shot image classification method combining VAE and StackGAN

智能系统学报. 2022, 17(3): 593–601 <https://dx.doi.org/10.11992/tis.202107012>

多尺度特征融合网络的视网膜OCT图像分类

Retinal optical coherence tomography image classification based on multiscale feature fusion

智能系统学报. 2022, 17(2): 360–367 <https://dx.doi.org/10.11992/tis.202111024>

用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction

智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

海底声呐图像智能底质分类技术研究综述

Survey of the intelligent seabed sediment classification technology based on sonar images

智能系统学报. 2020, 15(3): 587–600 <https://dx.doi.org/10.11992/tis.202004026>

基于谱域-空域结合特征和图割原理的高光谱图像分类

Hyperspectral image classification based on spectral-spatial combination features and graph cut

智能系统学报. 2015(2): 201–208 <https://dx.doi.org/10.3969/j.issn.1673-4785.201410040>

DOI: 10.11992/tis.202112001

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20221107.1749.002.html>

关键点图对比图像分类方法

卢毅^{1,2}, 陈亚冉¹, 赵冬斌¹, 刘暴^{3,4}, 来志超^{3,4}, 王超楠^{3,4}

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190; 2. 山东师范大学信息科学与工程学院, 山东 济南 250358; 3. 中国医学科学院北京协和医学院, 北京 100730; 4. 北京协和医院血管外科, 北京 100730)

摘要: 深度学习是目前图像分类的主流方法之一, 其重视感受野内的局部信息, 却忽略了类别的先验拓扑结构信息。本文提出了一种新的图像分类方法, 即 Key-D-Graph, 这是基于关键点的图对比网络方法, 在识别图像类别时可以显式地考虑拓扑先验结构。具体地, 图像分类需要 2 个步骤, 第一步是基于关键点构建图像的图表达, 即采用深度学习识别图像中目标类别的可能关键点, 并采用关键点坐标生成图像的拓扑图表达; 第二步基于关键点的图像图表达建立图对比网络, 以估计待识别图与目标类别之间的结构差异, 实现类别判断, 该步骤利用了物体的拓扑先验结构信息, 实现了基于图像全局结构信息的物体识别。特别的, Key-D-Graph 的中间输出结果为类别关键点, 具有语义可解释性, 便于在实际应用中对算法逐步分析调试。实验结果表明, 提出的方法可在效率和精度上超过主流方法, 且通过消融实验验证了拓扑结构在分类中的作用机制和有效性。

关键词: 关键点识别; 图拓扑结构; 图像分类; 图对比学习; 距离学习; 图神经网络; 暹罗网络; 图分类

中图分类号: TP3 文献标志码: A 文章编号: 1673-4785(2023)01-0036-11

中文引用格式: 卢毅, 陈亚冉, 赵冬斌, 等. 关键点图对比图像分类方法 [J]. 智能系统学报, 2023, 18(1): 36-46.

英文引用格式: LU Yi, CHEN Yaran, ZHAO Dongbin, et al. Keypoint-based graph contrastive neural network for image classification[J]. CAAI transactions on intelligent systems, 2023, 18(1): 36-46.

Keypoint-based graph contrastive neural network for image classification

LU Yi^{1,2}, CHEN Yaran¹, ZHAO Dongbin¹, LIU Bao^{3,4}, LAI Zhichao^{3,4}, WANG Chaonan^{3,4}

(1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 2. School of Information Science and Engineering, Shan Dong Normal University, Ji'nan 250358, China; 3. Chinese Academy of Medical Sciences, Peking Union Medical College Hospital, Beijing 100730, China; 4. Department of Vascular Surgery, Peking Union Medical College Hospital, Beijing 100730, China)

Abstract: At present, deep learning is one of the mainstream methods for image classification. It focuses more on local features in the receptive field than the prior information of topological structure of the category. In this paper, We propose a Keypoint-based Discriminator Graph neural network (Key-D-Graph) for image binary classification method, which is a graph comparison network method based on key points. It explicitly introduces the topology prior structure when identifying image categories. The method contains two main steps. The first step is to build the graph representation of an image with the keypoints, that is, identifying possible key points of the target category in the image by a deep learning method, and then using the coordinates of the key points to generate the topological representation of the image. The second step is to build a graph contrastive network based on the image representation of key points, so as to estimate the structural difference between the graph to be identified and the object graph, realizing object discrimination. In this step, the topological prior structure information of the object is used to realize object recognition based on the global structure information of the image. Especially, the intermediate output results of Key-D-Graph are the key points of categories containing explicit semantic information, which facilitates analysis and debugging of the algorithm step by step in practical application. Contrast experiments show that the proposed method outperforms the mainstream methods both in efficiency and precision. And the mechanism and effectiveness of topological structure in classification are verified by the ablation experiments.

Keywords: keypoint detection; graph topological structure; image classification; graph contrastive learning; metric learning; graph neural network; siamese network; graph classification

收稿日期: 2021-12-01. 网络出版日期: 2022-11-08.

基金项目: 国家重点研发计划项目 (2019YFB1311700); 山东省自然科学基金青年基金项目 (ZR2021QF085); 国家自然科学基金青年基金项目 (62006223, 62006226); 中国科学院战略重点研究项目 (XDA27030400); 中央级公益性科研院所基本科研业务费临床与转化医学研究基金项目 (2019XK320004); 中国医学科学院医学与健康科技创新工程医学人工智能科技先导专项 (2018-12M-AI-004); 中央高校基本科研业务费重点项目 (3332020009).

通信作者: 陈亚冉. E-mail: chenyaran2013@ia.ac.cn.

图像分类是计算机视觉的关键问题之一, 由于图像的复杂背景带来了大量数据噪声, 在几十万维的数据中获取高效的图像特征表达是解决该问题的难点之一。传统的图像分类方法主要利用人工设计的特征, 例如方向梯度直方图、尺度不变特征, 从图像的纹理、形状、颜色角度提取图像

特征, 达到图像降维目的后, 再采用贝叶斯、支持向量机、随机森林、KNN 等分类器实现分类^[1-2]。传统方法的优势在于人工设计的特征表达往往具有语义解释, 因此分类结果具有可解释性, 但是由于真实环境获取的图像背景复杂, 导致提取的特征不能适用于真实环境图片分类任务。随着深度学习方法的发展, 特别是深度卷积神经网络 (deep neural convolutional networks, DCNN) 的出现^[3-4], 例如 Alexnet^[5]、VGG^[6]、Googlenet^[7], 图像特征提取能力进一步提升, 分类精度大幅提高。2016 年, He 等^[8]提出了残差网络, 进一步加深网络提高了 DCNN 提取图像高层特征能力。随后 DenseNet^[9]、Inception 系列网络^[10]等各种深度卷积网络变体陆续提出, 网络加深的同时开始变得更宽, 以便于融合多尺度特征, 从而进一步提升分类精度^[11]。伴随着网络规模快速增长的同时, 降低网络参数量、提高运算效率逐步成为研究者关注的目标。自动机器方法被引入进该领域, 目的是通过算法优化网络结构^[12-14], EfficientNet^[15] 系列网络就是由此提出的兼顾效率和性能的一类网络。虽然 DCNN 方法的效率得以提升, 但网络学习的图像的分类特征仍然难以解释, 端到端网络内部的机制仍不明确。研究人员通过可视化中间层特征, 以分析网络内部工作原理, 发现浅层网络的前几层是一些抽象的图像局部纹理和颜色特征^[16]。2018 年 Gerihos 等^[17]对 DCNN 的工作机制进行了研究, 该研究通过重新组合图像的纹理和轮廓生成新图像, 比较新生成图像与原图像分类结果, 发现 DCNN 决策更偏好于局部纹理信息。

因此本文显式地引入图像中物体拓扑结构并将其与深度学习方法结合, 提出了基于关键点的图对比图像分类方法 (Key-D-Graph), 该方法在有效利用图像局部特征的同时, 结合物体整体拓扑结构进行图像分类。Key-D-Graph 从图像局部特征和全局拓扑特征的利用上进行解耦, 将图像分类任务分解为一个关键点识别问题和一个基于图表征的组合优化问题。通过这种解耦, 可以利用物体结构先验实现图像的高效图表示, 降低了将问题作为一个整体求解的难度。

首先, 关键点识别任务通过识别图像特定的局部特征确定关键点位置。独立的关键点识别研究主要集中在特定的类别, 例如斑点、角点检测, 采用尺度不变特征变换 (scale-invariant feature transform, SIFT) 等局部特征描述符描述图像局部特征并确定关键点位置^[18]。基于深度学习的关键点检测方法常见于特定物体类别的识别任务中, 例如

人脸识别、人体位姿估计等。因为关键点识别无需全局感受野, 常用全卷积网络框架进行稠密估计, 输出整个图像范围的关键点位置响应^[19]。2020 Wang 等^[20]提出的 HRnet (high-resolution representation network) 采用全连接网络融合不同尺度的层级特征, 在图像稠密预测问题中表现出众, 在图像语义分割以及关键点识别等任务中均实现了性能提升。考虑到关键点识别与图像分类相比, 判别对象涉及的图像区域大大减少, 在较小的图像感受野中提取的特征即可实现关键点识别。因此, 本文引入 HRnet 网络实现关键点识别。

其次, 关键点特征特异性差, 容易出现误检, 因此 Key-D-Graph 在关键点识别的基础上结合图像的拓扑信息进行物体类别判断。近几年, 图神经网络是处理图结构信息数据最出色的方法之一, 图卷积网络、图注意力网络、Graph SAGE 网络为代表的各种图网络变体相继被提出, 该方法在分子性质预测、社交网络分析等图结构数据分析任务上取得了令人瞩目的成绩^[21-23]。该类网络通过图节点之间的信息传播、聚合以及节点信息的更新迭代来获取图上的特征。并且图神经网络的出色表现得益于它的拓扑表征能力, 2019 年 Xu 等^[24]提出了图同构网络, 对图神经网络的拓扑结构信息的图表示性能进行了分析, 证明了在单射形式的聚合函数下, 图神经网络可以达到 1-WL (weisfeiler-leman) 测试效果。因此本文建立了基于类别结构先验的图神经网络分类图像结构特征, 最终实现图像分类。

具体地, 通过确定的图像关键点位置, 利用三角剖分构建拓扑结构, 生成基于关键点的图表达, 在该图表达的基础上, 采用暹罗图网络框架对比学习实现类别分类。

暹罗网络最早于 2005 年由 Chopra 等提出, 该框架后来成为对比学习的基本框架之一。对比学习一般通过二元或三元损失以及多元损失, 通过对比相同类别或不同类别之间的相似或不同特征, 实现类别学习^[25-26]。在图像处理领域, 对比学习方法最初主要集中在简单背景下的人脸识别^[27]和手写字体识别任务^[28]中。其主要原因是图像在背景复杂、噪声较多的情况下, 相同类别图像差异过大, 导致网络难以收敛。为实现网络的快速收敛, 许多方法通过人工设计的数据增广规则, 生成正负样本, 在暹罗网络框架下通过自监督学习获得图像类别内的不变特征^[29]。近几年 Hinton 以及何恺明等研究人员陆续提出了 SimCLR^[30]、MOCO (momentum contrast)^[31]等一系列对

比学习方法,方法从正负样本数据增广、批次训练等角度针对对比学习方法进行优化,有效提升了对比学习的分类性能。但是现有用于图像领域的对比方法主干网络仍为 DCNN,训练中仍然受限于精心设计的正负样本对,训练崩溃问题仍然存在^[32]。而基于图神经网络的对比学习框架多用于一些引文、蛋白质、程序图等图结构数据特征提取问题^[33],在图像分类领域研究较少^[34]。

特别地,Key-D-Graph 与已往图像对比学习方法不同在于,本文的暹罗框架输入是图像的图表达而非原始图片,该图表达包含明确对应的图像关键点位置以及拓扑结构,不包含图像颜色纹理等信息,实现了图像局部信息与结构信息的解耦。网络输入两个图像的图表达:一个为已知类别的对照图,另一个为待分类图。对照图的结构作为类别先验引入到了分类中,通过图神经网络单图以及跨图的信息传播,在利用全局的拓扑信息的同时实现类别的区分。本文的主要工作及创新点如下:

- 1) 针对图像建立了基于关键点的图模型,将图像分类任务转化为图分类任务。
- 2) 针对图像分类将局部特征和全局特征进行

分解,两部分特征识别耦合为类别关键点识别和基于拓扑结构先验的图对比分类两个步骤。

3) 在二分类实验中,本文的实验结果验证该方法可以在计算效率和分类精度的双提升,验证了该方法的有效性。超参数分析实验通过调整关键点的阈值分析了中间结果对分类结果的影响。图对比网络消融实验验证了基于关键点的图对比网络在学习图片距离的有效性。仿真数据集上的拓扑结构消融实验则验证了拓扑结构信息在识别结构差别的有效性。

1 Key-D-Graph 图像分类方法

基于关键点的图对比图像分类 Key-D-Graph 方法,如图 1 所示。首先,生成基于关键点的图表达,主要步骤包含图节点生成和图结构的确定,图节点生成采用监督学习,采用一个 HRnet 网络识别图片关键点位置,然后利用关键点在图像中的位置,三角化生成图结构,由此获得图像的图表达;其次,通过图对比方法实现图像类别分类,采用图神经网络结合先验图结构进行类别推理,从而实现基于全局图像拓扑结构的图像分类。

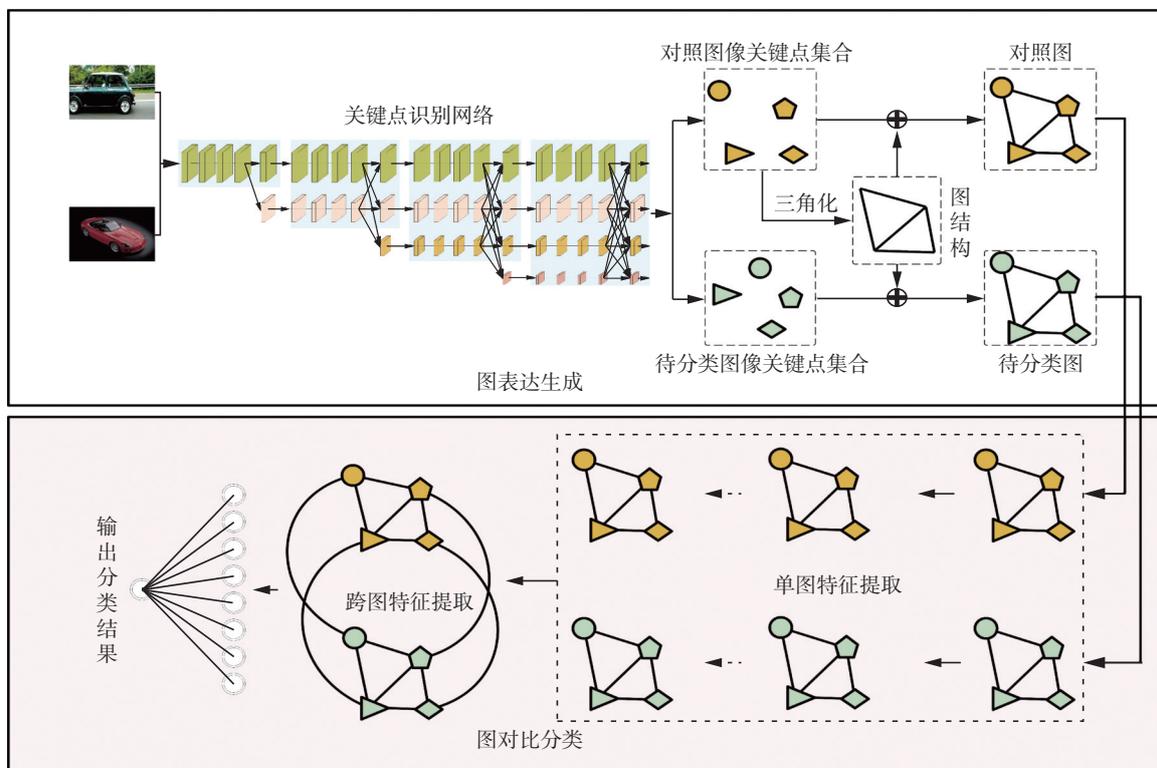


图 1 Key-D-Graph 方法流程

Fig. 1 Flow of the Key-D-Graph

1.1 生成基于关键点的图表达

根据类别的图像的特征,首先确定一组感兴

趣的类别关键点记为 $P = (p_1, p_2, \dots, p_i)$, 一般认为该组特征是对应类别具有语义的一组特征,其中

l 为类别关键点数量, 可以认为关键点集合是类别的属性集合。

1.1.1 图像的关键点识别

首先依据类别知识, 确定类别的关键点集合 P , 关键点集合 P 的定义和识别体现了该类别的局部特征, 包含了该类别在局部图像中与其他类别的不同。考虑到关键点特征可由一个相对较小的图像范围提取获得, 深度卷积神经网络的受限的感受野也能很好的适应该问题, 本文采用了 HRnet 来识别图像中的关键点, 通过监督学习获取关键点位置。

HRnet 网络并行地提取多尺度的特征图, 并将特征图用全连接进行融合, 输出高分辨率的特征。该方法在高分辨率的任务中, 如分割特征点识别等任务中, 用较少的参数获取更好的性能。在识别过程中, 网络的输入是一张图像, 输出的是关键的位置。在监督训练过程中, 监督信号是图像中所有关键点的图像坐标, 实际计算中先将图像的关键点坐标处理成一张热图, 热图的每个红色区域对应一个关键点, 由这张热图作为网络的监督信号, 因此该任务是一个稠密估计任务。

具体地, 网络最后一层输出为 $l \times w \times h$, 其中 l 为最后一层特征的通道数, 与要识别关键点数量相对应; w, h 分别为输出的特征图的宽和高, 这与关键点生成的热图大小一致。生成热图时, 依次在关键点坐标中心位置周围确定一个正方形的网格, 大小为 $s \times s$, 利用二维高斯函数以关键点坐标为重心计算网格概率, 填充到热力图中, 遍历所有关键点即获得一个以关键点坐标为中心的热图。该热图作为网络训练的标签, 每个样本图片 i 标签记为 y_i^{label} , 网络输出记为 y_i^{pred} 。在训练中采用的是均方误差:

$$l_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i^{\text{pred}} - y_i^{\text{label}})^2$$

预测过程中, 网络的输出 $y_i^{\text{pred}} \in \mathbb{R}^{l \times w \times h}$, 每个通道包含一张特征图, 对应一个关键点的输出结果, 特征图的每一个像素值对应了该位置包含关键点的概率, 在大于阈值 θ 的前提下, 一张特征图上最大概率对应的特征图像素位置坐标为对应的关键点位置。HRnet 获取图像关键点坐标的过程, 记为 $N = \text{HRnet}(X)$, 其中 $X \in \mathbb{R}^{3 \times p \times q}$ 表示图像输入, 其中 p 和 q 分别为图像的宽和高, 输出 $N = [p_1, p_2, \dots, p_n]$, $N \in \mathbb{R}^{n \times 2}$ 表示的是识别出的节点构成的节点特征矩阵, 为了便于表示, 本文仍旧用 p_i 表示对应关键点 i 的特征向量, n 为识别出的关键点个数。

1.1.2 图结构的获取

根据获取的关键点坐标, 利用三角化确定关键点之间的连接关系, 本文选择德洛内三角网。该结构为 1934 年德洛内提出的一种具有良好性质的三角剖分结构, 该三角剖分包含一系列相连的但不重叠的三角形的集合, 其中任意两个共边三角形的外接圆不包含另一三角形顶点, 被称为德洛内三角网^[35]。德洛内三角网具有一系列良好的性质, 例如存在唯一性, 且满足最大最小角性质, 德洛内三角网的边都不相交, 新增或删除移动某个顶点只会影响邻近三角形等, 因此被广泛应用于计算机几何计算中。

考虑到德洛内三角网为沃罗诺伊 (Voronoi) 图的对偶图, 因此本文在生成德洛内三角网时, 采用基于 Voronoi 图最小邻近点集的德洛内三角化方法^[35-36] 获得边结构。根据获取的关键点, 生成关键点集合的德洛内三角网结构, 同时对应关键点排列顺序, 唯一确定图的邻接矩阵 A 。

至此本文采用深度卷积网络识别类别关键点, 针对待识别关键点集合 P , 输出获得关键点的位置 N 。通过关键点的图像坐标生成图结构, 该图结构对应节点特征矩阵 N 的邻接矩阵记为 A , 由此得到了一个具有语义含义的图像图表达 $G(N, A)$, 详见图 2。

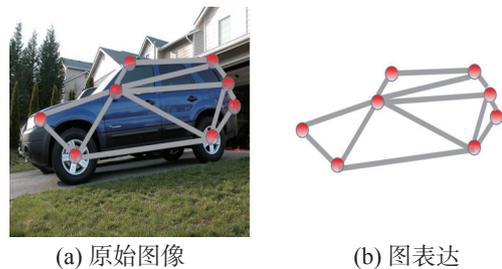


图 2 图像图表征示例

Fig. 2 Graph representation example of an image

1.2 图对比分类

基于上一节生成的图像图表达, 已经将图像分类转化为图分类问题, 特别地, 1.1.2 节获得图表达已经剥离了原图像的纹理颜色等特征, 仅包含关键点位置以及对应的拓扑图结构, 实际上由图 2 可以看出该图表达实际对应了图像物体的拓扑图结构。

考虑到输入图像中物体的位置、大小一般并不相同, 即使是相同的物体在不同尺度、角度下关键点位置之间的距离也会较大。因此图的分类中要考虑不同尺度、位姿差别下的拓扑图结构的一致性特征。而考虑跨图的关键点属性对应, 以及图上的信息传播步骤可以自适应地调整两张图

的隐变量并学习到两者基于类别形状的图距离。

图数据分析问题已经被广泛地研究,特别是图神经网络方法在多个图分类任务获得了令人瞩目的成绩。因此本文采用图神经网络方法提取类别的拓扑特征,同时考虑通过对比学习获取同类比特征,具体地,通过待分类图与对照图的对比区别类别间距离以实现图分类,本节采用对比学习常用的暹罗网络框架,如图 1 的图对比分类部分所示,一共包含 2 个步骤:1)单张图的隐变量获取,该部分通过图网络分别获取对照图和待分类图的隐变量;2)是跨图特征提取,两张图的节点实际对应的是具有实际语义的关键点,在对照图和待分类图的节点之间存在对应关系,跨图特征提取部分在对应关系之间进行图上的信息传播,以获取表达两张图相对关系的隐变量。本文采用图对比网络计算对照图与待分类图的距离,最终实现类别判断。

1.2.1 网络输入

输入的对照图和待分类图分别记为 $G^T = G(A^T, N^T)$, $G^E = G(A^E, N^E)$, 其中 T 标记对照图信息, E 标记待分类图的信息;输入的图信息仅包含图像的关键点位置和关键点生成的全局拓扑结构,不再包含图像局部特征如颜色、纹理等,实现了局部特征与全局特征的解耦。

在计算中,待分类图和对照图的结构一致,采用的都是对照图结构,记为

$$A = A^E = A^T$$

对照图的结构相当于类别结构先验,采用相同图结构目的是使分类图在先验图结构下获得与对照图相似的隐变量。

1.2.2 单图特征提取

图中的节点特征为关键点图像坐标位置,节点特征会依照物体的拓扑结构移动等产生相应的变动,本文目的是通过单图的特征提取,消除物体结构在单图物体的相对移动对节点特征产生的影响。具体地,采用图卷积层提取新的图节点特征:

$$N' = \text{GNN}_1(A, N) = f_{\text{softmax}}(ANW_1) \quad (1)$$

式中: GNN_1 表示单图特征提取的图神经网络, f_{softmax} 为激活函数, W_1 是待学习的参数。在单张图节点特征的提取过程中采用的是共享参数的图神经网络,两张图同时输入,各自输出单图的节点隐变量作为跨图特征提取步骤的输入。

1.2.3 跨图特征提取

获知的单张图特征后,通过两张图 (G^T 和 G^E) 对应节点关系组合而成的图神经网络层学习两图

相对关系隐变量。图中的每个节点属于某类关键点,根据关键点的属性可以得到两张图节点的匹配关系,匹配关系矩阵记为 $A^{(T,E)}$, 两张图像中识别出的关键点的数量分别为 m 和 n 。如果待分类图中的关键点 p_i^T 与对照图中关键点 p_j^E 同属一个类别,则匹配矩阵中元素 $a_{ij}^{(T,E)} = a_{ji}^{(T,E)} = 1$, 否则值为 0。通过这些匹配连接的边对照图和待分类图可以构成一张大图,这张大图的节点,记为 $N_{\text{match}} = [N^T, N^E]$, 邻接矩阵则记为

$$A_{\text{match}} = \begin{bmatrix} I_m & -A^{(T,E)} \\ -A^{(T,E)} & I_n \end{bmatrix}$$

在对照图和待分类图 (G^T 和 G^E) 以及它们节点匹配关系构造生成的大图 $G(N_{\text{match}}, A_{\text{match}})$ 上,通过图上信息传播操作更新全部节点隐变量:

$$N'_{\text{match}} = \text{GNN}_{\text{match}}(A_{\text{match}}, N_{\text{match}}) = f_{\text{softmax}}(A_{\text{match}}N_{\text{match}}W_2) = \left(\begin{bmatrix} I_m & -A^{(T,E)} \\ -A^{(T,E)} & I_n \end{bmatrix} [N^T \quad N^E] W_2 \right) \quad (2)$$

式中: W_2 为该部分待学习参数,在聚合图上所有点的距离特征后,通过一个多层感知机 (multilayer perceptron, MLP), 输出两张图的预测距离, W_3 为该部分参数:

$$g^{\text{pred}} = \text{MLP}_{W_3}(N'_{\text{match}}) \quad (3)$$

最后输出 g^{pred} 表示的是待分类图与对照图的距离,输入相同类别与不同类别的 G^T 和 G^E 对应不同的 g^{pred} 。

1.3 训练与预测过程

图对比方法中的网络参数与关键点识别部分的网络参数是分开训练。图对比网络的训练过程中,训练中任意选取两张图像,类别一致为正样本对标签值为 1, 类别不一致为负样本对标签值为 -1。采用对比学习中常用的软间隔损失函数 (soft margin loss) 作为损失函数:

$$l = \sum_i \frac{\ln(1 + \exp(-g_i \times g_i^{\text{pred}}))}{|g^{\text{pred}}|}$$

式中: g_i^{pred} 为预测值, g_i 为标签值, $|g^{\text{pred}}|$ 表示的输出向量的个数。在预测过程中,对照图为该类别训练集中的一个图表示,待分类图测试图像由 1.1 节获取的图表示,根据输出的 g_i 判断两张图是否来自同类别,实现图像类别判断。

2 实验结果及分析

为了验证提出方法的有效性,本文在真实环境图片以及仿真数据中分别进行了实验。其中在真实环境图片中分别进行对比实验和消融实验,以及关键超参数的分析实验:在对比实验中,将 Key-D- Graph 与主流图像分类深度卷积神经网络

对比, 方法的分类性能和效率进行评估; 在消融实验, 去掉中间的图神经网络层, 与直接利用关键点分类结果进行对比, 以分析图对比网络的作用; 超参数分析实验中设置了多个关键点识别阈值, 以分析第一步关键点识别与图对比分类之间的关系。在仿真数据中设计了拓扑结构信息消融实验, 分析拓扑图结构体现的拓扑信息在距离学习中的作用。

2.1 真实环境数据集上的实验

2.1.1 数据集

针对二分类任务, 本文选择 Pascal VOC2012 数据集^[37] 中的多个类别数据作为训练数据集。该数据集是 Pascal VOC2012 图像分类挑战赛的公开数据集, 共包含 20 类物体, 包含分类、检测、分割等任务标签。例如汽车与非汽车的类别分类, 测试集为原测试集中的全部 153 张汽车图像, 另随机挑选了 153 张非汽车类别的测试集中的图像构成类别评价测试集。伯克利大学发布了 Pascal VOC2012 数据集中 8 702 张图像中的 20 类物体关键点位置, 共标注了 161 种关键点, 其中 7020 张训练图像, 1 682 张测试图像。该数据同时具有图像类别和关键点标注标签, 因此本文选择在该数据集上进行训练测试。

2.1.2 实验设置

在关键点识别网络学习过程中, 首先对输入图像进行预处理, 依据物体所在的检测框进行随机尺度缩放然后进行中心裁切, 然后将裁切后的图像统一缩放为分辨率 256×256 作为网络的输入。在修改尺寸的过程中, 如果随机裁切后的图像分辨率小于 64×64, 则将图像进行尺度缩放后补齐。选取 HRnetv2-18 作为关键点识别网络, 针对实验数据集网络识别 161 类关键点, 输出关键点的坐标作为图对比分类中图节点特征, 如果图中没有此类关键点, 则该关键特征设置为 (-1, -1)。该关键点的位置为训练过程采用 ADAM 优化器。关键点训练过程的超参数设置详见表 1。

表 1 训练超参数

Table 1 Parameters of training in keypoints detection

参数	数值
关键点覆盖网格大小 s /像素	7
标签热图分辨率	64×64
关键点识别阈值 θ	0.01
关键点识别网络初始学习率	0.00001
图对比网络学习率衰减率	0.1
图对比网络训练批次数	300

在图对比分类网络中, 共包含 4 层图结构, 其中暹罗结构下的单图特征提取有 3 层单图信息交换的图卷积层和 1 层跨图的图卷积层。隐变量维度为 20, 对提取的图特征进行距离预测的 MLP (详见式 (1)) 的隐含层数为 1, 隐层节点数与两张图的节点总数相同。训练以及测试过程的对照图都来自于已知类别的训练集, 在训练过程中待分类图也来自于训练集, 测试时待分类图为测试集中的样本。输入过程中通过增加随机生成的关键点进行数据增广, 训练采用 SGD 优化器。

2.1.3 对比指标

实验对比采用分类主要指标准确率 $V_{accuracy}$ 、召回率 V_{recall} 、精确率 $V_{precision}$ 、 F_1 指标 V_{F_1} , 其中准确率计算方式为

$$V_{accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}$$

式中: N_{TP} 正样本中检测正确的样本数, N_{TN} 为负样本中检测正确的样本数, N_{FP} 为正样本中检测错误的样本数, N_{FN} 为负样本中检测错误的样本数。召回率为正样本中识别正确的样本占有所有正样本的比例, 计算公式为

$$V_{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

精确率为预测正确的正样本占有所有正确识别样本的比例, 计算公式为

$$V_{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

精确率与召回率是一组相互矛盾的分类评价指标, 当召回率过高时, 将导致负样本误检为正样本, 分类精确率下降; F_1 指标是一个综合指标, 可以折中评价网络的性能, 计算公式为

$$V_{F_1} = \frac{2 \times V_{precision} \times V_{recall}}{V_{precision} + V_{recall}}$$

2.1.4 对比实验

为了验证方法的分类性能以及计算效率, 在标有关键点的 VOC 数据集中车与其他物体分类任务上将 Key-D-Graph 方法与多个深度学习端到端的分类基准网络进行了对比。

Key-D-Graph 方法的参数为所有待学习参数, 包含了关键点识别和图对比分类网络的参数, 其中 HRnet 网络参数数量为 9.64 MB, 图对比分类网络参数数量为 1.04 MB。本文通过实验发现调整关键点识别的阈值, 网络性能会发生变化, 在对比实验中给出了关键点筛选阈值为 0.1 时以及设置为 0.01 的分类性能。对比实验中选取了 ResNet 和 EfficientNet 两个系列基准网络, 其中 ResNet 是深度卷积神经网络领域最常用方法之一, 对比试

验选取了参数量较少的 ResNet18 和 ResNet34 进行对比; EfficientNet 由谷歌 Tan 等提出, 被认为是分类性能最好的分类器之一^[15], 可以在相同的参数量取得更好的分类效果, EfficientNet-B0 是经过搜索算法获得的分类基线模型, 研究人员在此基础上提出了一系列的 EfficientNet, 在许多数据集上表现出色, 因此除 EfficientNet-B0 外, 本文选择了本系列中与 Key-D-Graph 参数量类似的 EfficientNet-B2 和 EfficientNet-B3 的分类结果进行对比。

实验结果详见表 2, 对比指标为准确率 V_{accuracy} 、参数量 V_{Params} 和乘积累加运算量 (multiply-accumulate operations, MACs) V_{MACs} , 准确率可以衡量网络分类效果, 网络的参数量和 MACs 来对比算法的复杂度。

表 2 对比实验结果
Table 2 Results of contrast experiments

方法	参数量/MB	MACs/GB	准确率/%
EfficientNet-B0	4.01	0.4	96.43
EfficientNet-B2	7.70	1.0	98.21
Resnet18	11.18	2.4	96.18
Resnet34	21.29	4.8	97.45
EfficientNet-B3	10.69	1.8	96.94
Key-D-Graph(0.1)	10.68	4.7	92.48
Key-D-Graph(0.01)	10.68	4.7	98.37

由表 2 可以看出, 与 Resnet 方法相比, 本文方法在相似参数量甚至更少参数量的情况下取得了分类精度的提升, 阈值为 0.1 的 Key-D-Graph 比 21.53 MB 参数的 Resnet34 提高了 1.96%, 阈值为

0.01 时 Key-D-Graph 比 21.53 MB 参数的 Resnet34 提高了 7.85%。同时 MACs 上, HRNet 为 4.659 GB, 而图对比分类部分仅为 1.72 MB, 与深度卷积获取关键点部分相比可以忽略不计。主要是因为图对比分类中的数据采用基于关键点的图表示, 与图像 RGB 表示相比, 维度大大减少。

当阈值为 0.1 时, Key-D-Graph 超过了 EfficientNet-B0, EfficientNet-B2, 仅次于 EfficientNet-B3 的分类结果, 但是 EfficientNet-B3 的参数量比 Key-D-Graph 多 1.55 MB。阈值为 0.01 时 Key-D-Graph 则超过了 EfficientNet-B3 方法 4.91%, 与 EfficientNet 方法相比也能在参数量更少的情况下获得更好的分类结果。同时, 针对对比实验中的关键点阈值对分类结果的影响, 也设计了相关实验以分析该参数的作用。

2.1.5 消融实验

本文设计了消融实验以分析图对比分类网络部分对分类结果的影响。消融实验将本文的分类结果与擦除图分类网络直接对关键点分类的方法进行了对比, 以分析图对比分类网络的分类作用。消融实验将提取的关键点直连一个 MLP 网络进行分类, 该 MLP 网络的设置与公式 (3) 中的设置相同, 记为 HRnet+MLP。HRnet+MLP 与 Key-D-Graph 相比擦除了图对比分类网络, 即图 1 中的图对比分类部分全部采用 MLP 进行替代, 该实验可以分析本文提出的图对比分类网络基于图结构的特征提取能力。两组实验的关键点阈值均为 0.1。除了汽车类别外, 还在行人/其他类别, 牛/其他类别数据上进行了实验对比, 实验结果详见表 3。

表 3 多组消融实验结果
Table 3 Results of ablation experiments

方法 指标	HRnet+MLP				Key-D-Graph			
	准确率	召回率	精确率	F_1	准确率	召回率	精确率	F_1
汽车	89.14	86.45	92.41	89.33	92.48	87.74	97.14	92.20
行人	83.65	74.68	91.47	82.23	84.62	73.42	95.08	82.86
牛	87.04	92.59	83.33	87.72	83.96	92.59	84.74	88.50

该消融实验包含了 3 个分类任务上的对比实验, 分别对比了准确率、召回率、精确率以及 F_1 值。如表 3 所示, 增加图对比神经网络部分后, 3 个分类实验的识别准确率、精确率以及 F_1 指标在均有提升, 该结果验证了图对比学习网络提取分类特征的有效性, 也体现结构特征对分类任务的价值。

2.1.6 超参数分析实验

超参数分析实验主要是针对关键点识别阈值

这一关键超参数进行对比, 在汽车/非汽车训练及测试集上进行了多次实验, 以分析该超参数对实验结果的影响。实验中, 在多个阈值的关键点数据上比较图对比分类网络效果, 结果详见表 4。从表 4 可以看出当阈值减少时, 方法的分类性能有所增加。当阈值减少时, 图像中会识别出更多的关键点, 无论是正确还是错误的关键点都为图分类网络提供更多了关键点以及结构信息。误检的关键点位置并不能与对照图关键点之间的相对

关系保持一致, 增加了与对照图之间的距离, 从而提高了分类性能。

表 4 超参数分析实验结果

Table 4 Results of hyperparameter analysis experiments

阈值	准确率	召回率	精确率	F_1	%
0.1	92.48	87.74	97.14	92.20	
0.05	96.08	93.55	98.63	96.03	
0.01	98.37	98.06	98.70	98.38	

2.2 仿真数据集上的拓扑结构消融实验

除真实环境图像上的分类任务实验外本文仿真数据集上设计了分析测试实验, 以分析 Key-D-Graph 图对比网络中拓扑结构信息度量物体位姿距离的能力, 通过该消融实验的实验结果对比, 可通过分析包含拓扑结构和擦除拓扑结构对位姿距离估计的影响, 以解析图对比网络中的拓扑结构信息在分类中的作用。

仿真数据集为由一个 3D 骨骼数据生成的 2D 图像数据集, 不同的位姿下可以生成不同角度的 2D 图像, 生成图片如图 3 所示, 图中展示了生成的两种图像, 一种是 3D 模型视角图像, 另一种是由 3D 模型投影生成的图像, 两种图片关键点绝对位置不同, 以仿真不同模态的 2D 图像。本测试目的是输入单张图像可以估计当前骨骼的位姿, 即在已知的 3D 位姿的多张 2D 图片中确定最相近的 2D 图像, 该 2D 图像的对应位姿即为估计的 3D 位姿。

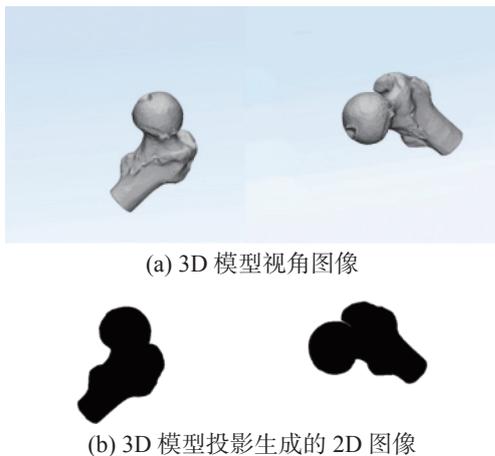


图 3 仿真数据集中的 2D 图片
Fig. 3 2D images in the simulation dataset

数据集将 80 对 2D 图像作为训练集, 3 组位姿估计作为测试集。该消融实验将 Key-D-Graph 和 Key-D-MLP 进行对比, Key-D-MLP 即为擦除了关键点的拓扑图结构, 即图神经网络中基于结构

信息传播步骤采用 MLP 替代, 相当于输入无结构的关键点向量, 与 Key-D-Graph 的不同在于没有图结构以及跨图结构的信息传播, 即式 (1) 及式 (2) 中的 GNN 层均替换为 MLP, 与 3.3 节图对比网络消融实验中的 HRnet+MLP 相比, 保留了暹罗对比学习框架。Key-D-Graph 和 Key-D-MLP 仍通过对比学习不同 2D 图像之间的位姿差异, 监督损失采用的是仿真图像已知对应的 3D 模型位姿距离, 具体的采用 L2 范数作为要估计的 3D 位姿距离。在训练过程中, 通过对比两张图中对应节点之间的相对位置关系, 获得两张图的位姿差异, 这种差异相当于图与图之间的位姿距离。

如表 5 及图 4 实验结果所示, 与擦除掉结构信息的 Key-D-MLP 方法相比, 具有拓扑结构的 Key-D-Graph 可以在测试集的 3 组图像中估计的最小距离均为最佳的位姿匹配图像, 说明 Key-D-Graph 对图像间的 3D 位姿距离估计的更为准确, 该方法与 Key-D-MLP 相比更能有效学习不同图像位姿之间的距离。两个对比方法的差别则在于一个包含先验图结构, 另一个擦除了先验图结构, 实验结果的不同说明了 Key-D-Graph 方法引入的对照图拓扑结构先验能够帮助网络更好地在图像平移变化中获取结构不变量, 可以有效提取物体的结构特征。

表 5 拓扑结构信息消融实验结果

Table 5 Results of topological information ablation experiments

方法	实验结果
Key-D-Graph	3组均成功获取最佳匹配
Key-D-MLP	成功匹配1组

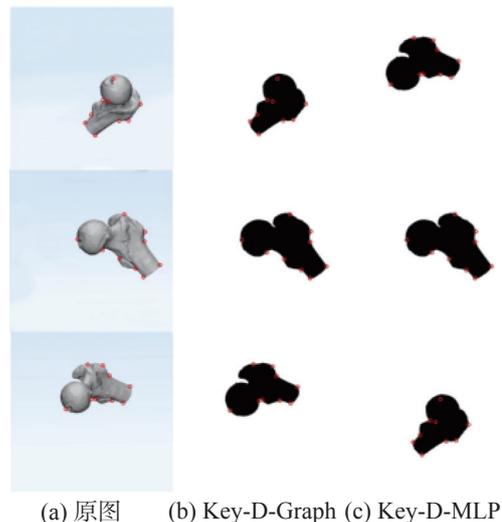


图 4 最佳结果对
Fig. 4 Best image pairs in the results

3 分析讨论

通过以上实验结果说明,本文提出的 Key-D-MLP 分类精度以及计算效率均超过了现有的主流方法且具有如下特点:

1) Key-D-Graph 是中间结果,可解释的两阶段方法。方法的第一阶段输出目标类别关键点,可视化结果见图 5,关键点具有明确语义信息且依据超参数分析实验结果,该阈值与网络识别关键点结果以及识别网络训练相关,该阈值为关键点识别的输出概率,说明可根据阈值调整识别出的关键点数量,从而影响最终的分类结果。而端到端的深度学习方法一般被认为是黑箱结构,难以对网络中间层结果明确地进行语义解释,且难以利用中间输出结果对方法进行有针对性地调整。因此本文提出的 Key-D-Graph 方法依据中间输出的具有可解释性的语义信息且可以对结果进行调控,与端到端的深度学习方法相比更适用于需要部分中间结果分析的应用场景。

但是如何确定最优阈值仍是一个困难的问题。阈值为关键点识别的输出概率,每个阈值可得到一个图分类数据集,而且最优阈值随不同数据集上关键点识别网络的输出范围和图分类网络的训练结果变化。事实上,关键点网络的输出与训练的过程有关,输出结果会因训练不同,导致网络可能在识别相同关键点时对应不同的阈值。因此寻找最优阈值需要在不同数据集和大量关键点识别网络上进行大量训练以归纳总结最优阈值的范围。



图 5 汽车关键点识别可视化

Fig. 5 Detected keypoints visualization of the cars

2) 由对比实验结果可以看出,两阶段的网络将局部信息与全局信息识别进行了解耦操作,与现有的基准模型相比更加简洁、有效。在局部识别过程中,即第一步的关键点的识别过程中,任务所需的感受野小,相比图像分类难度大大降低,同时关键点识别步骤去除了原图像中与分类无关的大量冗余信息,关键点与原图像相比维度

大大降低,也为图表达分类降低了难度,因此在对比实验中可以实现降低计算复杂度的同时提升分类效率。

3) Key-D-Graph 将图像识别解耦为局部信息与全局信息识别过程,这个步骤充分利用了类别先验信息,同时验证了图表达拓扑结构的有效性。首先,通过对类别的认知定义了图像的关键点集合,这部分一是考虑类别特异性,二是在实际应用中,可以依据识别环境对关键点进行选择。其次,在图对比分类步骤中将已知样本结构作为先验知识引入测试样本中,建立了基于全局图像图表达,采用图对比网络实现了全局特征的利用,该步骤基于暹罗结构的图对比方法实现了基于先验图结构的推理。依据拓扑结构消融实验的结果,可以看出先验图拓扑结构在距离学习任务的有效性。已有的深度卷积网络方法完全是通过大量的样本学习提取特征,缺乏先验语义结构的利用,而拓扑结构消融实验说明 Key-D-Graph 引入的对照图结构可以更好地提取图像结构特征,从而弥补了类别先验语义结构利用的不足。

4 结束语

本文提出了一种图像分类方法 Key-D-Graph,通过深度卷积网络获得图像关键点,然后建立基于关键点的图像图表达,结合先验图结构信息利用图对比学习网络实现图像分类。该方法分阶段实现了图像局部特征和全局结构特征识别,在应用中可依据中间输出的关键点结果调试分类结果。对比实验结果表明该方法与已有的深度学习方法相比,在效率和性能上均有提升,同时验证了图拓扑结构信息在分类任务中的重要作用,弥补了现有主流深度学习方法在图像类别先验结构利用上的不足。

消融实验结果说明拓扑图结构在 3D 位姿距离估计中的有效性,因此未来将考虑将不同类别图像的差异转化为不同角度投影图的距离学习,以实现多类别图像分类。确定最优阈值可以更清晰揭示阈值对结果的影响机制,也是未来待研究的工作之一。

参考文献:

- [1] 黄凯奇,任伟强,谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6): 1225-1240.
HUANG Kaiqi, REN Weiqiang, TAN Tieniu. A review on image object classification and detection[J]. Chinese journal of computers, 2014, 37(6): 1225-1240.

- [2] 王亮申, 欧宗璞, 朱玉才, 等. 基于 SVM 的图像分类[J]. *计算机应用与软件*, 2005, 22(5): 98–99,126.
WANG Liangshen, OU Zongying, ZHU Yucui, et al. Classifying images with SVM method[J]. *Computer applications and software*, 2005, 22(5): 98–99,126.
- [3] 任越美, 张艳宁, 李映. 压缩感知及其图像处理应用研究进展与展望[J]. *自动化学报*, 2014, 40(8): 1563–1574.
REN Yuemei, ZHANG Yanning, LI Ying. LI Ying. Advances and perspective on compressed sensing an application on image processing[J]. *Acta automaica sinica*, 2014, 40(8): 1563–1574.
- [4] 马忠丽, 刘权勇, 武凌云, 等. 一种基于联合表示的图像分类方法[J]. *智能系统学报*, 2018, 13(2): 220–226.
MA Zhongli, LIU Quanyong, WU Lingyu, et al. Syncretic representation method for image classification[J]. *CAAI transactions on intelligent systems*, 2018, 13(2): 220–226.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the acm*, 2012, 60: 84–90.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014–09–04)[2021–12–01].<https://arxiv.org/abs/1409.1556>.
- [7] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1–9.
- [8] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [9] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [10] BALDASSARRE F, MORÍN D G, RODÉS-GUIRAO L. Deep koalarization: image colorization using CNNs and inception-ResNet-v2[EB/OL]. (2017–12–09) [2021–12–01].<https://arxiv.org/abs/1712.03400>.
- [11] CHEN Yaran, LI Haoran, GAO Ruiyuan, et al. Boost 3-D object detection via point clouds segmentation and fused 3-D GIoU- L_1 loss[J]. *IEEE transactions on neural networks and learning systems*, 2022, 33(2): 762–773.
- [12] CHEN Y, GAO R, LIU F, et al. ModuleNet: knowledge-inherited neural architecture search[J]. *IEEE transactions on cybernetics*, 2022, 52(11): 11661–11671.
- [13] DING Zixiang, CHEN Yaran, LI Nannan, et al. BNAS: efficient neural architecture search using broad scalable architecture[J]. *IEEE transactions on neural networks and learning systems*, 2022, 33(9): 5004–5018.
- [14] LI Nannan, PAN Yu, CHEN Yaran, et al. Heuristic rank selection with progressively searching tensor ring network[J]. *Complex & intelligent systems*, 2022, 8(2): 771–785.
- [15] TAN Mingxing, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019–05–28)[2021–12–01].<https://arxiv.org/abs/1905.11946>.
- [16] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision. Cham: Springer, 2014: 818–833.
- [17] GEIRHOS R, RUBISCH P, MICHAELIS C, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness[EB/OL]. (2018–11–29)[2021–12–01].<https://arxiv.org/abs/1811.12231>.
- [18] LOWE D G. Object recognition from local scale-invariant features[C]//Proceedings of the Seventh IEEE International Conference on Computer Vision. Kerkyra: IEEE, 1999: 1150–1157.
- [19] WEI Shihen, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 4724–4732.
- [20] WANG Jingdong, SUN Ke, CHENG Tianheng, et al. Deep high-resolution representation learning for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2021, 43(10): 3349–3364.
- [21] LU Yi, CHEN Yaran, ZHAO Dongbin, et al. CNN-G: convolutional neural network combined with graph for image segmentation with theoretical analysis[J]. *IEEE transactions on cognitive and developmental systems*, 2021, 13(3): 631–644.
- [22] LU Yi, CHEN Yaran, ZHAO Dongbin, et al. Graph-FCN for image semantic segmentation[C]//International Symposium on Neural Networks. Cham: Springer, 2019: 97–105.
- [23] 张勇, 高大林, 巩敦卫, 等. 用于关系抽取的注意力图长短时记忆神经网络[J]. *智能系统学报*, 2021, 16(3): 518–527.
ZHANG Yong, GAO Dalin, GONG Dunwei, et al. Attention graph long short-term memory neural network for relation extraction[J]. *CAAI transactions on intelligent systems*, 2021, 16(3): 518–527.
- [24] XU K, HU Weihua, LESKOVEC J, et al. How powerful are graph neural networks? [EB/OL]. (2018–10–01)[2021–12–01].<https://arxiv.org/abs/1810.00826>.

- [25] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005: 539–546.
- [26] SOHN K. Improved deep metric learning with multi-class N-pair loss objective[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 1857–1865.
- [27] TAIGMAN Y, YANG Ming, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1701–1708.
- [28] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[EB/OL]. (2020-06-17)[2021-12-01].<https://arxiv.org/abs/2006.09882>.
- [29] OORD A V D, LI YAZHE, VINYALS O. Representation learning with contrastive predictive coding[EB/OL]. (2018-07-10)[2021-12-01].<https://arxiv.org/abs/1807.03748>.
- [30] CHEN Ting, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[EB/OL]. (2020-02-13)[2021-12-01].<https://www.semanticscholar.org/paper/A-Simple-Framework-for-Contrastive-Learning-of-Chen-Kornblith/34733eaf66007516347a40ad5d9bbe1cc9dadb6b>.
- [31] HE Kaiming, FAN Haoqi, WU Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9726–9735.
- [32] CHEN Xinlei, HE Kaiming. Exploring simple Siamese representation learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 15745–15753.
- [33] ZHU Yanqiao, XU Yichen, YU Feng, et al. Graph contrastive learning with adaptive augmentation[C]//WWW '21: Proceedings of the Web Conference 2021. New York: ACM, 2021: 2069–2080.
- [34] WANG Runzhong, YAN Junchi, YANG Xiaokang. Learning combinatorial embedding networks for deep graph matching[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 3056–3065.
- [35] 张俊, 田慧敏. 一种基于边指针搜索及区域划分的三角剖分算法 [J]. *自动化学报*, 2021, 47(1): 100–107.
- ZHANG Jun, TIAN Huimin. A triangulation algorithm based on edge-pointer search and region-division[J]. *Acta automatica sinica*, 2021, 47(1): 100–107.
- [36] GUIBAS L J, KNUTH D E, SHARIR M. Randomized incremental construction of delaunay and Voronoi diagrams[M]//Automata, Languages and Programming. Berlin/Heidelberg: Springer-Verlag, 2005: 414–431.
- [37] EVERINGHAM M, ESLAMI S, GOOL L, et al. The pascal visual object classes challenge: a retrospective[J]. *International journal of computer vision*, 2014, 111: 98–136.

作者简介:



卢毅, 讲师, 博士, 主要研究方向为计算机视觉、图神经网络。发表学术论文 8 篇。



陈亚冉, 副研究员, 博士, 主要研究方向为计算机视觉、智能驾驶、机器人。发表学术论文 34 篇, 获得包括 IEEE 汇刊 2020 年度唯一优秀论文等多项论文奖励; 获得 2017 年中国智能车未来挑战赛离线测试 2 项第一名, 2020 IEEE ICRA DJI RoboMaster 人工智能挑战赛感知/导航/决策多赛道 3 项第一名。



赵冬斌, 研究员, 博士, IEEE Fellow, 主要研究方向为强化学习、自适应动态规划、智能游戏、智能交通、机器人。发表学术论文 300 余篇。