

doi:10.11835/j.issn.1000-582X.2021.03.013

精细拓扑结构表示与深度特征融合在 多目标图像检索中的应用

刘东^{1,2a}, 王生生^{2a,2b}

(1. 湘南学院 软件与通信工程学院, 郴州 423300; 2. 吉林大学

a. 符号计算与知识工程教育部重点实验室; b. 计算机科学与技术学院, 长春 130012)

摘要: 空间信息表示是增强图像特征表达性能的重要手段, 通过空间关系建模与深度学习方法融合可有效提升深度特征的语义特性, 从而提升图像检索性能。首先, 针对复杂图像的空间关系表示提出了一种新的精细拓扑结构表示模型, 该模型不仅具有完备的拓扑描述性能, 还提供了两种拓扑不变量的推理算法, 使得拓扑不变量可以由表示模型直接推导而不需要繁复的几何计算; 其次, 基于精细拓扑结构表示模型, 提出了有效的拓扑结构相似性度量方法, 为空间关系特征表达奠定了基础; 最后, 进一步结合卷积神经网络, 提出融合复杂空间关系特征与深度特征的多目标图像检索方法。实验结果表明, 所提出的拓扑结构表示模型在空间查询中具有良好的性能; 所提出的图像检索框架取得优于现有方法的精度, 并能够有效地结合手工特征与深度特征各自的优势, 为提升深度学习方法的解释性创造了有利条件。

关键词: 拓扑; 深度学习; 图像检索; 特征提取; 特征融合

中图分类号: TP391

文献标志码: A

文章编号: 1000-582X(2021)03-132-12

On the application of detailed topological structure representation and deep feature fusion to multi-object image retrieval

LIU Dong^{1,2a}, WANG Shengsheng^{2a,2b}

(1. School of Software and Communication Engineering, Xiangnan University, Chenzhou 423300,

Hunan, P. R. China; 2a. Key Laboratory of Symbolic Computation and Knowledge Engineering of the Ministry of Education; 2b. School of Computer Science and Technology, Jilin University, Changchun 130012, P. R. China)

Abstract: Spatial information representation is an important means to improve image visual feature representation performance. The integration of the spatial relation model with deep learning can effectively enhance semantic property of deep features, improving the image retrieval accuracy. In this paper, we proposed a novel detailed topological structure representation model to describe spatial relation of complex images. This model not only had complete topology description performance, but also provided two

收稿日期: 2019-11-21

基金项目: 国家自然科学基金资助项目(61972333); 湖南省教育厅优秀青年项目(18B504); 湖南省自然科学基金项目(2018JJ3479); 郴州市科技局科技发展计划项目(zdyf201906)。

Supported by National Natural Science Foundation of China (61972333), Research Foundation of Education Bureau of Hunan Province(18B504), National Natural Science Foundation of Hunan Province (2018JJ3479) and Science and Technology Planning Project of Chenzhou City(zdyf201906).

作者简介: 刘东(1988—), 男, 博士, 主要从事图像处理方向研究, (E-mail) promisedong@163.com。

通讯作者: 王生生, 男, 教授, 博士生导师, 主要从事计算机视觉方向研究, (E-mail) wss@jlu.edu.com

efficient reasoning algorithms, which made the topological invariants directly deduced from the model without any geometric calculations. Similarity matching approaches based on fine topological structure representation model was proposed for spatial relationship feature representation. Finally, in combination with convolution neural network, a multi-object image retrieval framework was developed by fusing the spatial relation features and deep features. Experimental results demonstrate that the proposed topological model has remarkable performance in spatial query. Moreover, the proposed image retrieval framework outperforms the current methods in terms of precision and with advantages of both the manual and deep features, it provides a superior means to improve interpretability of deep learning methods.

Keywords: topology; deep learning; image retrieval; feature extraction; feature fusion

图像视觉特征提取与表达是图像检索与分类的关键步骤,同时也是计算机视觉领域的重要研究方向^[1]。归纳总结视觉特征研究发展历程,可分为 3 个阶段:底层视觉特征提取、中间层特征表达,以及最近流行的深度学习方法。1)底层视觉特征提取主要针对图像形状、纹理、颜色等信息进行刻画,提取表征图像外观的视觉特征,如常用的尺度不变特征变换^[2]、局部二值模式^[3]等。底层视觉特征提取一般计算较简单,但单独使用时难以胜任复杂的计算机视觉任务。2)中间层特征表达是指对底层视觉特征进一步编码和统计,以挖掘更具判别能力的深层特征表示,如经典的视觉词袋模型(BoVW, bag-of-visual-words)^[4]。中间层特征表达在一定程度上可以弥补底层视觉特征的缺陷。3)深度学习作为一种倍受关注的驱动方法,其不需要手工参与设计,直接以图像像素作为初始数据输入,经过多层网络结构,学习图像的潜在深层特征。以卷积神经网络(CNN, convolution neural network)^[5]为代表的深度学习方法,发展出一系列高效的网络结构,如 AlexNet^[5]、VGGNet^[6]、ResNet^[7]等,被用在高光谱图像分类^[8]、人体行为识别^[9]与表情识别^[10]等多个领域。CNN 虽然在自然场景图像分类中取得巨大成功,但仍然存在一些缺陷,如依赖海量样本数据、可解释性较差、难以推理、较难描述和理解图像内容模式的含义等,这使得 CNN 在处理小样本和对语义特性要求较高的多目标图像时面临许多挑战。

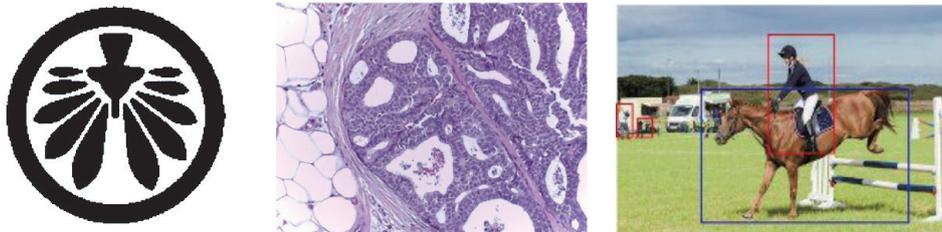


图 1 多目标图像示例

Fig. 1 The examples of multi-object images

空间关系建模作为一种实现图像语义表达的重要手段,可有效增强视觉特征表达性能。相较于纹理、形状、颜色等底层视觉特征,通过刻画图像基元之间的空间关系来识别图像,更符合人类的视觉认知习惯。空间关系刻画可有效跨越和缩小底层视觉特征到高层语义之间的“语义鸿沟”,在计算机视觉与模式识别领域具有重要的地位,尤其是对于医学影像等多目标图像识别具有至关重要的判别能力^[11-12]。然而,现有工作研究的空间关系模型存在诸多缺陷,如描述目标简单、基本关系数量有限、缺少有效的推理与相似性度量方法,难以满足视觉特征表达的现实需求。表 1 总结了目前代表性的空间关系模型,其中 RCC-8^[13]等经典空间关系模型仅有少数种空间关系;CRString^[14]虽然具有完备的空间关系表示,但其目标仅针对凸区域;前期研究成果 DTString^[15]研究了简单区域的完备拓扑空间关系表示,但其不适用于具有复杂区域的多目标图像场景。如图 1 所示,商标、医学和目标分割后的场景图像均为具有复杂区域空间关系的多目标图像。

表 1 代表性空间关系模型总结

Table 1 Summary of representative topological relation models

模型名称	目标	基本关系数量	推理	相似度量
区间代数	单向时间轴区间	13	逆关系,符合关系	—
RCC-8	正则闭集区域	8	逆关系,复合关系	—
DA	圆盘	8	逆关系,复合关系	—
9-IM	简单区域	43	—	—
CRString	凸区域	∞	—	—
DTString	简单区域	∞	逆关系,子区域推理	近似匹配
研究提出的 DTSRM	复杂区域	∞	洞判别推理,子区域推理	图相似性计算

针对上述空间关系描述简单、图像视觉特征表达不充分、语义信息缺失的问题,主要研究复杂图像的空间关系建模与深度特征融合,提出一种能满足视觉特征表达需求的空间关系模型来刻画复杂区域的空间拓扑结构,并与深度特征融合,应用于多目标图像检索,提高检索性能,同时为增强深度学习可解释性创造有利条件。主要贡献包括:1)针对复杂区域表示,提出了一种新的精细拓扑结构表示模型,不仅具有完备的拓扑关系描述性能,还提供了有效的拓扑不变量推理算法和拓扑结构相似性度量方法;2)基于精细拓扑表示模型,提出了融合复杂空间关系特征和深度特征的多目标图像检索算法,取得优于现有方法的性能。

1 精细拓扑结构表示模型

提出一种新的精细拓扑结构表示模型(DTSRM, detailed topological structure representation model)来刻画多目标的复杂区域,主要包括精细拓扑结构的表示、定性推理和相似性计算 3 个部分。这 3 个部分从理论到应用层层递进。为便于介绍,首先给出一些通用符号说明,如表 2 所示。

表 2 通用符号说明

Table 2 Description of universal symbols

符号	说明	符号	说明
A	平面上的有界区域	V, U	成分的集合
A°	A 的内部	E	成分之间连接关系的集合
A^e	A 的外部	L	成分的标签
ω_0	平面上的无界开放区域	S_g	有向子图相似度
a_i	复杂区域中的一个成分(component)	S_u	无向子图相似度
$\partial A, \partial a$	A 或者 a 的边界	S_d	深度特征相似度
G	四元组,提出的精细拓扑结构表示	λ	特征值
G_1, G_2	由 G 拆分成的三元组,分别可看成有向子图与无向子图	T	有向图

1.1 精细拓扑结构表示

首先逐步给出复杂区域的形式化定义。平面上同胚于圆盘的区域称之为简单区域,而带洞的简单区域则定义如下。

定义 1(带洞的简单区域) 给定 c_0, c_1, \dots, c_n 为简单区域,若满足 1) $1 \leq i \leq n$, 有 $c_i \subset c_0$, 且

$|\partial c_i \cap \partial c_0| \leq 1$; (2) 对于 $1 \leq i, j \leq n (i \neq j)$, $|c_i \cap c_j| \leq 1$, 则称 $A = c_0 \setminus \bigcup_{i=1}^n c_i^c$ 为带洞的简单区域, $c_i (i \geq 1)$ 称为 A 的洞。

定义 2(复杂区域) 设 A 为平面上的区域, 若 A 由若干个简单区域或带洞的简单区域组成, 则称 A 为复杂区域。

图 2 展示了简单区域、带洞的简单区域以及复杂区域的区别。给定一个有界的平面区域 A , 称 A° 为正成分, 如图 2(b) 中 c_0 ; A^c 为负成分, 且负成分可进一步划分为两种类型, 一种是有界的, 如图 2(b) 中 c_1, c_2, c_3 ; 一种是无界的, 如 w_0 。容易发现, 一个区域至少包含一个正成分和一个负成分, 且 A^c 有唯一一个无界的开放区域(即 w_0)。

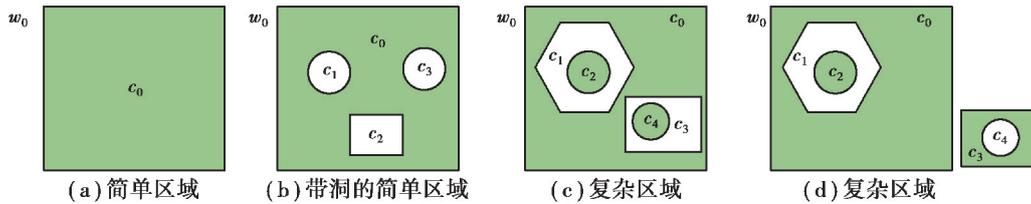


图 2 简单区域与复杂区域示意图

Fig. 2 The examples of simple regions and complex regions

基于上述定义, 针对多目标复杂区域描述提出精细拓扑结构表示模型 DTSRM。如图 3(a) 所示, 在多目标复杂区域中, 对于任意 2 个成分 a_i 和 a_j 之间有 2 种连接关系: 1) 强连接关系, 若 a_i 和 a_j 闭包的交集包含至少一条曲线; 2) 弱连接关系, 若 a_i 和 a_j 闭包的交集仅包含一个或多个离散点。

定义 3(精细拓扑结构表示模型) 复杂区域 A 的精细拓扑结构表示定义为一个四元组 $G = \{V, E_1, E_2, L\}$, 其中 V 是 A 中所有成分的集合; E_1 表示成分之间强连接关系的集合, E_2 表示成分之间弱连接关系的集合, 即 $r_{ij} \in E_1$ 和 $r_{ij} \in E_2$ 分别表示成分 a_i 和 a_j 之间存在强连接关系(即图 3 中的有向边)或者弱连接关系(即图 3 中的无向边); L 则表示每个成分的标签, 可根据需要由视觉特征描述后聚类得到。此外, 在集合 V 中, 唯一的根节点表示开放背景区域(即 w_0), 记根节点 w_0 在第 0 层, w_0 的子节点 c_1, c_2, c_3 在第 1 层, c_1, c_2, c_3 的子节点在第 2 层, 以此类推, 由此得到精细拓扑结构表示模型为图 3 所示。

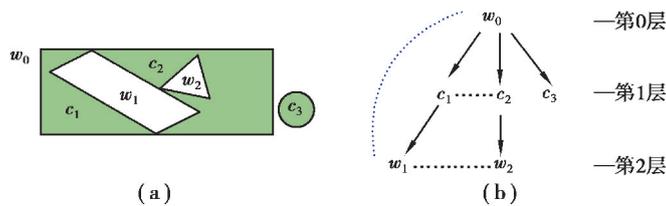


图 3 精细拓扑结构表示模型(DTSRM)

Fig. 3 The detailed topological structure representation model (DTSRM)

下面通过一个直观的例子阐述提出的 DTSRM 模型相比现有方法的优势。图 4 展示了 3 种具有相似拓扑结构的复杂区域, 但是在局部细节上存在较大的差异。使用 CRString^[14] 和 DTString^[15] 模型无法对 3 种拓扑结构进行区分, 因其无法描述复杂区域。使用树表示模型^[16] 无法将点连通区域分开, 即将(a)和(b)视为同一种拓扑结构; 分层图模型^[17] 则无法区分弱连接关系, 即无法区分(a)和(c)2 种情况。而提出的精细拓扑结构表示模型综合考虑了强连接和弱连接关系, 能够有效区分图 4 中 3 种情况, 实现了更精确和完备的拓扑结构表达。值得说明的是, 将 DTSRM 模型应用于多目标图像刻画时, 考虑到像素采样的误差与粘连, 弱连接关系中目标边界交集的离散点可以根据需要抽象为若干个邻域像素集合。

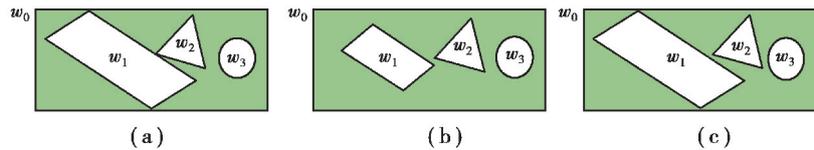


图 4 三种不同的拓扑结构

Fig. 4 Example of three different topological structures

实际上, DTSRM 模型通过有效区分强连接与弱连接关系不仅有助于增强语义特性, 对于描述医学影像等尤其注重细节内容差别的图像具有重要的意义。如图 5 所示的乳腺癌全扫描切片(来自 2018 年 international conference on image analysis and recognition 组织的乳腺癌图像分析大赛^[18]), 其中病理组织(背景区域)、良性(红色区域)、原位癌病变区域(绿色部分)、浸润性癌病变区域(蓝色区域)彼此之间构成复杂区域, 并且区域之间体现出强连接关系和弱连接关系, 对他们之间的空间组织形式与拓扑结构进行表达, 可有效协助医生对癌细胞转移进行分析与判断。

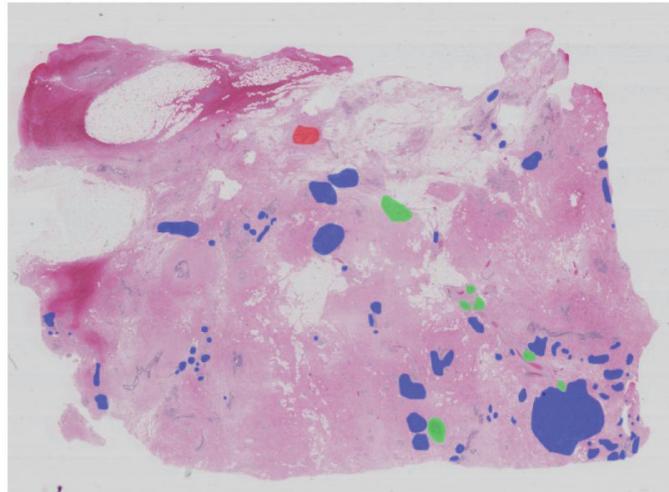


图 5 乳腺癌图像中的复杂区域

Fig. 5 The complex region in breast cancer images

1.2 基于精细拓扑结构表示模型的定性推理

1.2.1 洞判别推理

设复杂区域 A 以及它的精细拓扑结构表示 $G = \{V, E_1, E_2, L\}$ 。实际上, 如图 6 所示, G 可划分为一个有向子图 $G_1 = \{V, E_1, L\}$ 和一个无向子图 $G_2 = \{V, E_2, L\}$ 构成。

定理 1 设 c 为 A 的非根节点, $G'_1 \subset G_1$ 为从 G_1 中移除 c 所形成的子图, 则 c 没有洞当且仅当 G'_1 依然是连通的。

证明: 首先证明必要性。设 c 没有洞, 如果 c 存在子孙节点 b , 假设从根节点到 b 之间的路径均经过 c , 根据模型定义 3, 这意味着 b 的闭包只与 c 的闭包存在交集, 即 b 是 c 的子节点, 假设与前提条件矛盾。因此, 从根节点到 b 之间必然存在一条路径没有经过 c , 即在 G_1 中移除 c 得到的 G'_1 依然是连通的。

接着证明充分性。设 G'_1 是连通的, 假设 c 存在洞 b , 则从根节点 w_0 到 b 的每条路径必然经过 c , 在 G_1 中移除 c 后将造成 G'_1 是不连通。假设与前提条件矛盾, 因此 c 没有洞。综上所述, 即得证。

定理 2 设 c 为 A 的非根节点, 如果从 G_1 中移除 c 将产生一个或多个连通子图, 这些子图可以进一步划分为两种类型: 1) 唯一一个包含根节点的子图 w_0 ; 2) 其他不包含根节点 w_0 的子图 $g_i(c)$, $1 \leq i \leq k$, 则 $g_i(c)$ 即是 c 的洞。

证明: 从定理 1 可知, 如果 c 存在洞 b_i , 若从 G_1 中移除 c 将使得 b_i 与根节点是不连通的, 则 b_i 可看成等

价的子图 $g_i(c)$, 即得证。

下面给出一个直观的例子。在如图 6 所示的复杂区域和它的精细拓扑结构表示, 只有 c_2 节点存在洞 b_2 , 因为在 G_1 中移除 c_2 之后, b_2 单独形成一个不包括根节点 w_0 的子图。

由上述定理 1 和定理 2, 可在 $O(n)$ 时间复杂度内 (n 为节点的数量), 直接从 DTSRM 表示模型中推导出每个子区域包含洞的数量, 不需要额外的图形或者几何计算, 从而大大地降低了计算复杂度, 同时也提供了具有语义特性的拓扑不变量。

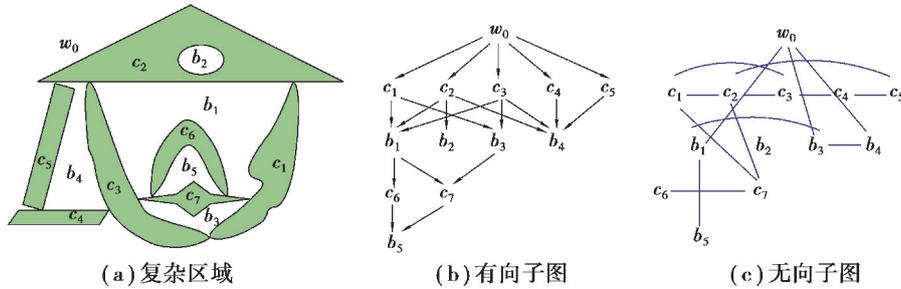


图 6 DTSRM 模型可划分为有向子图和无向子图表示

Fig. 6 DTSRM model divided into a directed sub-graph and a undirected sub-graph

1.2.2 环绕子区域推理

设复杂区域 A 以及它的拓扑结构表示 $G = \{V, E_1, E_2, L\}$, 假定 c 为 A 的任意成分(子区域), 一个切实的问题是能否仅根据 G 就推导出 c 由哪些其他子区域环绕, 以及如何确定环绕子区域的顺序。这些信息不仅是拓扑不变的, 而且对于基于知识表示的图像重构也具有重要意义。下面提出子区域推理来回答上述问题。

定理 3 设 c 为 G 的节点, $S(c)$ 定义为如下集合: $S(c) = \{v_i \subset V \mid (c, v_i) \in E_1 \text{ 或 } (v_i, c) \in E_1, 1 \leq i \leq k\}$, 则 $S(c)$ 为 c 的围绕子区域集合。并且, 如果每个 $v_i \in S(c)$ 恰恰形成 E_2 中的回路 $R = (r_1 r_2 \cdots r_k r_1)$, 则 $(r_1 r_2 \cdots r_k r_1)$ 即为 c 环绕的子区域序列。

证明: 首先, 设任意节点 v_i 满足: $\partial v_i \cap \partial c$ 包含一条简单曲线, 即 v_i 和 c 是强连接的, 则 v_i 是 c 的子节点或者父节点, 即 $(c, v_i) \in E_1$ 或 $(v_i, c) \in E_1$ 。由定义 3 易知, ∂c 的每个线段分量只与其父节点或子节点存在交集, 因此 $S(c)$ 恰为 c 的围绕子区域集合。再者, c 的任意 2 个相邻围绕子区域必然存在弱连接关系(以一个点或多个点相连), 而模型在 E_2 中恰恰描述了这种弱连接关系。因此, 只需要在 E_2 中找到关于 $S(c)$ 中所有节点的回路 $(r_1 r_2 \cdots r_k r_1)$, 即为 c 的环绕子区域的序列。特别的, 如果 $S(c)$ 中仅包括一个节点, 意味着 c 是一个某个子区域的洞, 被其父节点所环绕。

举例如图 6 所示, 其中图 6(b) 和图 6(c) 分别为有向子图 $G_1 = \{V, E_1\}$ 和无向子图 $G_2 = \{V, E_2\}$ 。以计算 b_4 的环绕子区域序列为例, 由图 G_1 易知: $S(b_4) = \{c_2, c_3, c_4, c_5\}$, 而 $S(b_4)$ 中的所有节点在 G_2 中恰恰存在回路 $(c_2 c_3 c_4 c_5 c_2)$, 因此 $(c_2 c_3 c_4 c_5 c_2)$ 即为 b_4 的环绕子区域序列, 在图 6(a) 中可得到验证。同理, 可计算 w_0 和 c_1 分别被 G_2 中回路 $(c_2 c_5 c_4 c_3 c_1 c_2)$, $(b_1 b_3 w_0 b_1)$ 所环绕, 同时上述环路中的节点在 G_1 中也分别是 w_0 和 c_1 的子节点或者父节点。

由上述几个定理可以发现, 所提出的精细拓扑结构表示模型不仅能实现对拓扑结构的精细刻画, 同时提供简易有效的推理, 使得拓扑不变量可以由表示模型直接推导得到, 并可进一步应用于空间查询与图像视觉特征表达。

1.3 基于精细拓扑结构表示模型的相似性度量

相似性度量是空间关系模型应用于视觉特征表达的重要基础。由于精细图表示模型 G 可划分为一个有向图 $G_1 = \{V, E_1, L\}$ 和一个无向图 $G_2 = \{V, E_2, L\}$ 组成, 分别对 2 个子图进行相似性度量。

1.3.1 有向图相似性度量

由于 G_1 是有唯一根节点和层次信息的有向图, 因此 G_1 亦可被看成类似树的结构。采用最大相似子树

来计算 2 个有向图的相似度。首先,假设 T_1 和 T_2 是 2 个有向图,其节点集合分别为 V_1 和 V_2 ,如果存在 $U_1 \subseteq V_1$ 和 $U_2 \subseteq V_2$,对于任意双射 $f:U_1 \rightarrow U_2$,使得子图 U_1 和 U_2 具有相同的邻接关系,则称该双射为一个子图同构。接着,一旦获得 2 个有向图的子图同构 f 之后,即计算子图同构的相似度 W

$$W(f) = \sum_{u \in U_1} \phi(u, f(u)), \quad (1)$$

其中: $\phi(u, v)$ 表示使用模糊形状描述算子^[19]计算 2 个节点(子区域)的形状相似度。表 3 给出了 2 个同构子图相似度的计算方法,则 2 个 T_1 和 T_2 的所有子图同构中具有最大的相似度称之为最大子图同构,其相似度定义为

$$S_g = \max_f W(f). \quad (2)$$

表 3 计算子图相似度的算法

Table 3 The algorithm of computing sub-graph similarity

algorithm: similarity-of-subgraph(u, v)

输入:分别以 u 和 v 为根节点的子图 T_1 和 T_2 ;

输出: T_1 和 T_2 的相似度;

说明: $\phi(u, v)$ 为两个节点(子区域)的形状相似度, $\{w_{ij}\}$ 为相似度矩阵, w_{ij} 对应 i 和 j 的相似度;Assign($\{w_{ij}\}$)为采用匈牙利算法解决 最佳指派问题的函数^[20];

1. 获取 u 的所有子节点,记为 C_u ;
2. 获取 v 的所有子节点,记为 C_v ;
3. for each node c_i in C_u
4. for each node c_j in C_v
5. $w_{ij} = \text{similarity-of-subgraph}(c_i, c_j)$ // 递归调用
6. return $\phi(u, v) + \text{Assign}(\{w_{ij}\})$

1.3.2 无向图相似性度量

考虑到在复杂图像表示的实际应用中,目标区域间的弱连接关系图 G_2 是稀疏的,提取无向图的图谱特征作为相似度量方法。具体的,给定 $G_2 = \{V, E_2\}$,首先计算其邻接矩阵的 n 个特征值,并将其按绝对值大小排序,即得 $t = [\lambda_1, \lambda_2, \dots, \lambda_n]$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$,为表征 G_2 的图谱特征。图谱特征能较好的反应无向图的特性,具有拓扑不变性,且计算较简单。对于任意精细图表示模型 G 和 G' ,获取最大子图同构之后,相应的节点组成的 2 个无向图即可计算其图谱特征的欧氏距离来刻画相似度,记为 S_u 。

2 基于精细拓扑结构表示与深度特征融合的图像检索

深度学习方法因其在场景图像分类等任务中的优异性能受到广泛的关注。深度学习虽然能有效获取深度特征,但其依赖于大量学习样本,并且缺乏语义信息。而空间关系模型蕴含着大量的语义信息,与深度特征融合能够充分利用两者的优势,有望提升视觉特征的表达性能与可解释性。

提出基于复杂空间关系特征与深度特征融合的多目标图像检索框架,如图 7 所示。首先,给定查询图像 I ,使用提出的精细拓扑结构表示模型刻画 I 与检索数据集中的样本,获取最大子图同构,分别计算有向子图相似度 S_g 和无向子图的相似度 S_u 。接着,使用训练好的卷积神经网络提取多目标图像的第一个全连接层特征作为 I 的深度特征 F_d ,使用欧式距离计算任意 2 个图像深度特征相似度,记为 S_d 。最终,任意 2 个多目标图像的相似度定义为

$$S(I_i, I_j) = w_1 s_g + w_2 s_u + w_3 s_d, \quad (3)$$

其中: w_1, w_2, w_3 为权重系数。

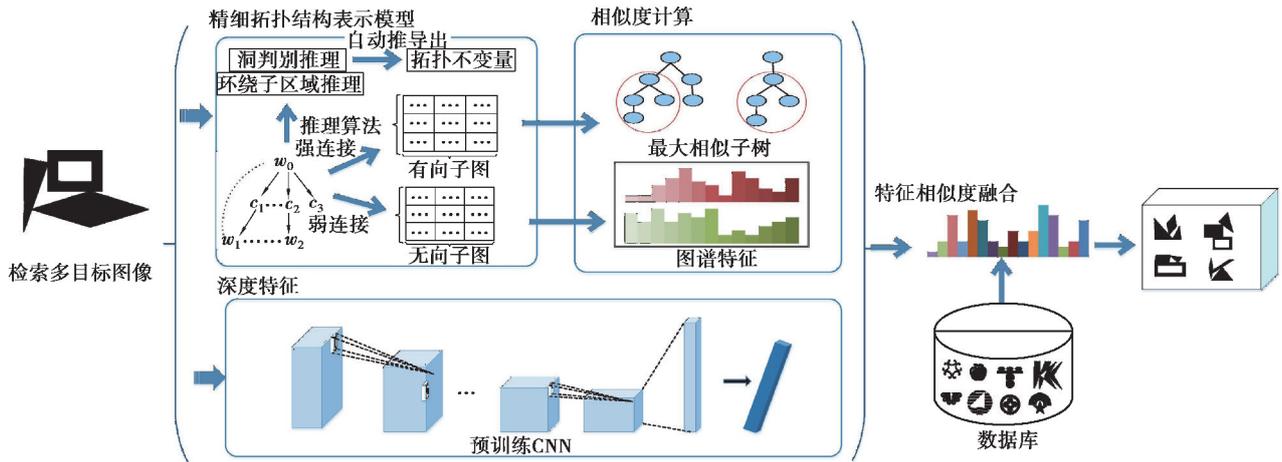


图 7 基于精细拓扑结构表示和深度特征融合的多目标图像检索框架

Fig. 7 Multi-objet image retrieval framework based on DTSRM and deep feature fusion

3 实验与结果分析

实验采用多目标图像数据集,该数据集一部分来自 MPEG-7 形状数据集^[21],一部分来自商标数据集^[22],包括 20 个类别的多目标二值图像,每个类别包含 30~100 不等的的数据样本。使用图像检索领域常用指标查准率(precision)和查全率 (recall)作为评价标准。实验机器配置为: Intel Core i9-7900X CPU, NVIDIA TITAN XP GPU, 16 GB 内存,使用 MATLAB 2019a 及深度学习工具箱作为实验平台。分别构建基于拓扑不变量的空间查询、基于内容的图像检索两组实验。

3.1 基于拓扑不变量的空间查询实验

在该实验中,输入“子区域数量”、“洞的数量”2 组拓扑不变量,基于精细拓扑结构表示模型及拓扑不变量推理算法,在数据集中查询满足条件的图像。图 8 给出了 3 组不同查询条件下的检索示意图。

(子区域, 洞) =(2,0)	(子区域, 洞) =(5,2)	(子区域, 洞) =(12,6)

图 8 多目标图像查询示意图

Fig. 8 Examples of multi-objet image retrieval

从图 8 的查询结果可以发现,当查询拓扑不变量较简单时,检索返回的图像视觉差异较大;但随着拓扑不变量越复杂,检索返回的图像不仅在在拓扑结构上体现出拓扑同构或者相似的结构特性,在视觉上也具有

更大的相似性。

值得说明的是,这种空间查询方式主要为了验证拓扑不变量及相关推理算法的性能,并未使用形状、深度学习等其他特征。实验结果表明,对于多目标图像而言,当限定了多个拓扑不变量同时满足条件查询时,检索出来的图像之间已然具有一定的相似性,并携带着一定的语义特性。随着查询的拓扑不变量越复杂,检索的结果在语义和视觉上越相似,这表明拓扑不变量蕴含着一定的判别能力。研究提出的精细拓扑结构表示模型和推理算法,可以直接在表示模型的基础上推导出拓扑不变量,而不需要繁复的图像计算,这对于空间和语义查询十分便利。

3.2 基于内容的图像检索实验

在该实验中,输入一个查询图像,在数据集中检索相似图像。

首先,对 2 种方法进行评估,一是单独使用所提出的精细拓扑结构表示模型和相似性度量方法,记为 DTSRM,二是融合了深度特征的方法(如公式 3),其中 CNN 网络结构选用 AlexNet^[5]和 VGGNet^[6],分别记为 DTSRM + AlexNet 和 DTSRM + VGGNet。为了评估复杂空间关系特征与深度特征的融合权重系数对检索结果的影响,计算不同权重系数下的平均查准率。即分别从每类数据集中随机抽取 10 副图像作为查询图像进行检索,共执行 200 次查询,抽取前 12 张最相似的图像计算平均查准率,表 4 展示了研究方法在不同权重系数下的平均查准率。

表 4 研究方法在不同权重系数下的平均查准率(%)

Table 4 The average precision of proposed method with different weight

权值(w_1, w_2, w_3)	DTSRM/%	DTSRM+ Alex	DTSRM+ VGGNet
(1,0,0)	80.22	—	—
(0.5,0.5,0)	82.79	—	—
(0.6,0.4,0)	83.17	—	—
(0.4,0.4,0.2)	—	85.38%	85.72%
(0.2,0.4,0.4)	—	85.11%	85.45%
(0.4,0.2,0.4)	—	86.08%	87.31%

由表 3 实验结果表明,1)结合表征弱连接关系相似性权重(即 $w_2 \neq 0$ 时)的检索精度要高于仅考虑强连接关系的检索精度(即 $w_2 = 0$ 时),这说明刻画目标之间的弱连接关系是对精细拓扑结构表达性能的完善。2)无论是在 AlexNet 还是 VGGNet 网络结构下,融合深度特征后的检索精度有了进一步提高,这主要因为复杂空间关系特征与深度特征在表达多目标图像中形成一种良好的互补。3)总体来说,复杂空间关系特征与深度特征的融合体现了一定的鲁棒性,即不同的权重组合方式均较原有基础有了较大提升,其中($w_1 = 0.4, w_2 = 0.2, w_3 = 0.4$)的权重组合取得最好的性能,被应用于后续实验。

接着,分别对比了代表性手工设计特征方法 HOGH^[23], IDSC^[24], TRM^[16], BCF^[25], 以及深度特征学习方法 AlexNet, VGGNet, 它们的查全率和查准率曲线图分别如图 9 和图 10 所示。图 11 展示了 5 种对比方法的检索结果示意图,其中加框的检索结果为返回的不同类别图像。

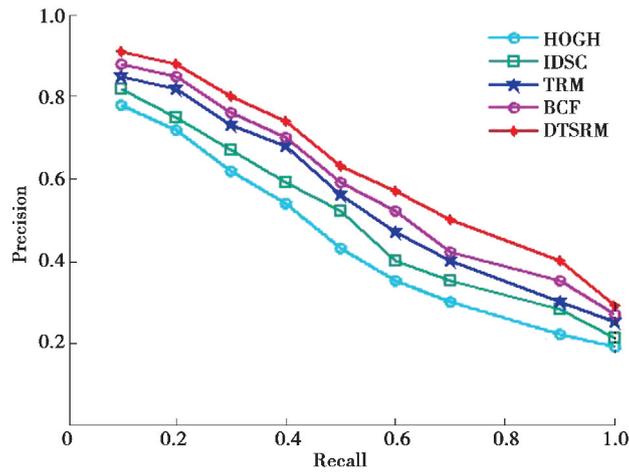


图 9 与手工特征方法的查准率与查全率曲线对比图

Fig. 9 Recall-Precision curves comparison of the proposed method and handcrafted features

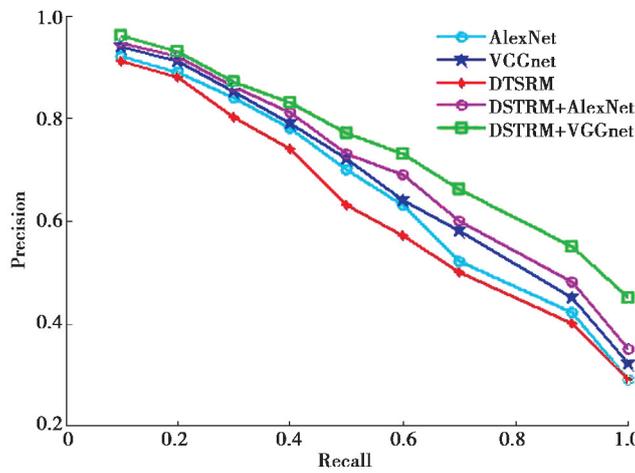


图 10 与深度学习方法的查准率与查全率曲线对比图

Fig. 10 Recall-Precision curves comparison of the proposed method and deep learning features

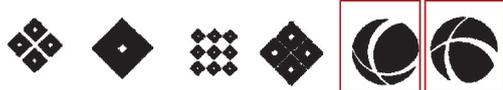
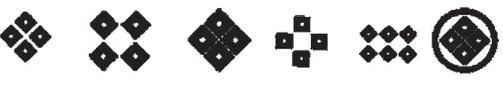
查询图像	对比方法	检索结果
	HOGH	
	BCF	
	AlexNet	
	VGGNet	
	本文方法	

图 11 5 种对比方法进行检索的结果

Fig. 11 The retrieval results of five compared methods

分析上述实验结果,可以得到以下结论:1)由图 9 可知,提出的精细空间表示模型 DTSRM 单独使用时要优于其他手工特征表达方法,如 HOGH、TRM、IDSC 等,这是因为所提出的模型具有完备的拓扑结构描述性能,能有效的刻画多目标的复杂空间关系特征,提供更具判别能力的视觉特征。2)由图 10 可知,无论是在 AlexNet 还是 VGGNet 网络结构下,DTSRM 与深度特征融合后,均相比深度特征单独使用时有了较大提升;此外由图 11 可知,研究方法所检索的结果具有更相似性的拓扑结构,这是主要因为 DTSRM 模型携带了一定的语义信息,与深度特征形成良好的互补,非常适用于多目标图像检索。3)DTSRM+VGGNet 获得了最高的检索精度,这可能得益于 VGGNet 更复杂的网络结构,并与复杂空间关系特征进行了较好地融合。

3.3 时间复杂度分析

图像检索主要包括特征提取、相似度计算两个环节,其中特征提取分为有监督特征提取和无监督特征提取,如 BCF 方法需要训练视觉词袋模型、CNN 方法需要利用训练集进行 fine-tuning,属于有监督特征提取;而其他的手工特征方法直接从单张图像中提取特征,属于无监督特征提取。表 5 展示了不同对比方法在特征提取、相似度计算及模型训练过程花费的平均时间。从中可以看出,DTSRM 方法在特征提取阶段花费时间最少,在相似度计算中涉及最大相似子图计算问题,需要花费 $O(n^2)$ 时间复杂度,其中 n 为图表示模型中的节点数,通常节点数小于 30。因此,在无监督多目标图像检索任务中,DTSRM 单独使用可取得较好的检索效果;如果综合考虑精度与时间复杂度,DTSRM 与深度特征融合则是一种更好的选择。

表 5 对比方法的时间复杂度

Table 5 The temporal complexity of the proposed method

方法名称	单张图片特征提取	相似度计算	训练时间
HOGH	0.133 s	$O(n)$	—
IDSC	0.182 s	$O(n)$	—
TRM	1.012 s	$O(n^2)$	—
BCF	0.212 s	$O(n)$	≈ 0.5 h
DTSRM	0.051 s	$O(n^2)$	—
Alex	0.062 s	$O(n)$	≈ 1.1 h
VGGNet	0.087 s	$O(n)$	≈ 1.3 h

4 结 语

主要研究了复杂多目标图像的空间关系特征表示与深度特征融合问题,为增强图像的视觉特征表达与语义特性提供了新的思路。首先提出了一种新的精细拓扑结构模型,该模型最大创新之处在于它不仅具有完备的拓扑描述性能,还支持两种重要拓扑不变量的推理,使得拓扑不变量可由知识表示直接推理而来。其次,设计了基于精细拓扑表示的相似性度量方法,并提出面向多目标图像检索的空间关系特征与深度特征的融合方法。实验结果表明,所提出的方法在多目标图像检索中取得优于传统手工特征和目前流行的深度学习方法,并且对于不同参数选择具有一定的鲁棒性。在未来研究工作中,进一步考虑将研究方法扩展应用在医学图像分析与场景图像分类中,挖掘具有语义特性的视觉特征。

参考文献:

- [1] Cheng Q M, Zhang Q, Fu P, et al. a survey and analysis on automatic image annotation[J]. Pattern Recognition, 2018, 79: 242-259.
- [2] Zheng L, Yang Y, Tian Q. SIFT meets CNN: a decade survey of instance retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5): 1224-1244.

- [3] Pietikäinen M, Zhao G. Two decades of local binary patterns: A survey[M]. Advances in Independent Component Analysis and Learning Machines. Academic Press, 2015: 175-210.
- [4] Jabeen S, Mehmood Z, Mahmood T, et al. An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model [J]. PloS One, 2018, 13(4): e0194526.
- [5] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. 2014: arXiv: 1409.1556[cs.CV]. <https://arxiv.org/abs/1409.1556>
- [7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// The IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA: IEEE, 2016:770-778.
- [8] Zhang M M, Li W, Du Q. Diverse region-based CNN for hyperspectral image classification[J]. IEEE Transactions on Image Processing, 2018, 27(6): 2623-2634.
- [9] 刘智,冯欣,张杰. 基于深度卷积神经网络和深度视频的人体行为识别[J]. 重庆大学学报, 2017, 40(11):99-106.
Liu Z, Feng X, Zhang J. Action recognition based on deep convolution neural network and depth sequences[J]. Journal of Chongqing University, 2017, 4(11):99-106. (in Chinese)
- [10] Liu Y Y, Yuan X H, Gong X, et al. Conditional convolution neural network enhanced random forest for facial expression recognition[J]. Pattern Recognition, 2018, 84: 251-261.
- [11] Gong C, Anders R A, Zhu Q, et al. Quantitative characterization of CD8+ T cell clustering and spatial heterogeneity in solid tumors[J]. Frontiers in Oncology, 2018, 8: 649-690.
- [12] 于瑞云, 苏展, 谢青, 等. 基于空间分割的人体模型骨骼提取算法[J]. 计算机学报, 2019, 42(9): 2049-2061.
Yu R Y, Su Z, Xie Q, et al. Skeleton extraction of character model based on space segmentation[J]. Chinese Journal of Computers, 2019, 42(9):2049;2061. (in Chinese)
- [13] Chen J, Cohn A G, Liu D, et al. A survey of qualitative spatial representations[J]. The Knowledge Engineering Review, 2015, 30(1): 106-136.
- [14] Li S J, Liu W M. Topological relations between convex regions[C]// The Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, USA: AAAI, 2010: 321-326.
- [15] Wang S S, Liu D, Zhang C, et al. Representation, reasoning and similar matching for detailed topological relations with DTString[J]. Information Sciences, 2014, 276: 255-277.
- [16] Alajlan N, Kamel M S, Freeman G H. Geometry-based image retrieval in binary image databases[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(6): 1003-1013.
- [17] Li S J. A layered graph representation for complex regions[C]// The 12th International Conference on Principles of Knowledge Representation and Reasoning, Toronto, Canada, Morgan Kaufmann, 2010: 581-583.
- [18] Aresta G, Araújo T, Kwok S, et al. Bach: Grand challenge on breast cancer histology images[J]. Medical Image Analysis, 2019: 1-19.
- [19] Escalera S, Alicia Fornés, Pujol O, et al. Blurred shape model for binary and grey-level symbol recognition[J]. Pattern Recognition Letters, 2009, 30(15): 1424-1433.
- [20] Lyzinski V, Fishkind D E, Fiori M, et al. Graph matching: relax at your own risk[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 60-73.
- [21] Abbas S, Farhan S, Fahiem M A, et al. Efficient shape classification using zernike moments and geometrical features on MPEG-7 dataset[J]. Advances in Electrical and Computer Engineering, 2019, 19(1): 45-51.
- [22] Wei C H, Li Y, Chau W Y, et al. Trademark image retrieval using synthetic features for describing global shape and interior structure[J]. Pattern Recognition, 2010, 42(3):386-394.
- [23] Ni H, Guo Z H, Huang B Q. Binary patent image retrieval using the hierarchical oriented gradient histogram [C]// 2015 International Conference on Service Science (ICSS), Weihai, China: IEEE, 2015: 23-27.
- [24] Zhang Z, Tang P, Duan R. Dynamic time warping under pointwise shape context[J]. Information Sciences, 2015, 315: 88-101.
- [25] Wang X G, Feng B, Bai X, et al. Bag of contour fragments for robust shape classification[J], Pattern Recognition, 2014, 47 (6): 2116-2125.