

基于边对注意力的跨图嵌入深度图匹配

杨益纳, 石 林

广东工业大学, 计算机学院, 广东 广州

收稿日期: 2022年3月17日; 录用日期: 2022年4月18日; 发布日期: 2022年4月25日

摘 要

图匹配问题(GM)旨在建立两个图的点集之间的一致性,同时保持它们的边集之间的一致性,在数学上可以被刻画为二次组合优化问题。随着神经网络的兴起和发展,基于深度学习的图匹配方法应用图神经网络技术将每个图的几何结构信息嵌入到结点特征,从而将二阶组合优化的图匹配问题松弛为线性分配问题。然而,这样的图匹配模型在本质上忽略了二阶的边对相似度信息,即没有考虑边集之间的一致性,从而导致这类模型在精度上有所损失。针对上述问题,本文提出了一种新型基于嵌入技术的深度图匹配网络,该网络充分考虑了边对相似性信息:基于二图匹配生成的关联图,将边对相似性信息(等价于关联图上的边)作为跨图嵌入的注意力信息,构造一个跨图注意力模块。本文基于公开数据集进行评估实验,结论表明本文的方法在精度上超越了已有的模型。除此之外,本文的方法相比于目前精度最高的模型,在内存消耗上具有优势,体现了该模型具有处理大规模图匹配任务的潜力。

关键词

组合优化, 图匹配, 图神经网络, 深度学习, 计算机视觉

Cross-Graph Embedding with Edge-Wise Attention for Deep Graph Matching

Yirui Yang, Lin Shi

School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 17th, 2022; accepted: Apr. 18th, 2022; published: Apr. 25th, 2022

Abstract

Graph matching problem (GM) refers to establishing correspondences between two set of points while keeping the consistency between their edge sets and can be mathematically formulated as a quadratic combinatorial optimization problem. With the success of graph neural network (GNN),

recent learning-based graph matching research attempts to solve the problem by way of linear assignment via transferring local structure information into node embedding at each graph. However, such a technique omits second-order pairwise edge similarity in essence causing the decrease of accuracy. Based on this issue, this paper proposes a novel embedding-based graph matching network fully considering the pairwise edge similarity. Edge-to-edge similarity, equivalent to associated edges, is aggregated into nodes in the association graph derived from two matching graphs, to amplify the discriminativeness of cross-graph node embedding. Experimental evaluations are conducted on public datasets, and results show that our method outperforms the state-of-the-art in accuracy. In addition, compared with the state-of-the-art, the proposed method has superiority in space consumption, indicating the potential to solve large-scale graph matching problems.

Keywords

Combinatorial Optimization, Graph Matching, Graph Neural Networks, Deep Learning, Computer Vision

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图匹配旨在解决在两个或多个图上寻找点-点对应的问题。它包含了对的一阶相似度和边对的二阶相似度,使得整体的相似度取得极大值。由于在目标函数中考虑了二阶结构化的边对信息,图匹配往往可以获得比基于点对的方法(如 RANSAC [1]和最近邻迭代 ICP [2])更高的鲁棒性能。尤其在计算机视觉领域,当光照和视角发生大幅变化时,依赖于像素强度的点特征容易产生剧烈变化而几何结构特征则能够保持相对稳定。因此,图匹配已经成为许多计算机视觉应用的核心,如物体识别[3]、动作识别[4]、多物体追踪[5]、立体匹配[6]等。

图匹配问题是满足离散组合约束条件的非凸优化问题,是一个 NP 难的二次组合优化问题[7]。为了降低求解图匹配问题的时间复杂度,传统的优化方法往往需要采用不同程度的松弛策略:1) 谱松弛[8],将离散域下的组合约束条件松弛成连续域下的正交约束,从而将图匹配问题转化成为求解特征向量的问题;2) 半正定松弛[9],将图匹配问题的非凸目标函数松弛成满足半正定性质的凸函数,该问题可以通过随机算法或赢家通吃(winner-take-all)策略求解;3) 双随机松弛[10],将离散域的组合约束条件松弛为连续域的组合约束条件,该二次组合优化问题可以通过多种策略找到局部最优解,如路径跟随策略[11]和基于泰勒展开的线性分配策略[12]等。

然而,传统的图匹配优化算法往往依赖于人工设计的相似度参数作为组合优化问题的输入,极大限制了模型对实际场景中结构匹配问题的泛化性能。针对该问题,基于学习的图匹配算法展现了优势:通过给定图匹配的真值,有监督地学习并构建图匹配的相似度模型。文献[13]提出了基于深度学习的图匹配模型,其中使用卷积神经网络(CNN)提取深度点特征和边特征并使用谱方法[8]作为组合优化问题的求解策略。随着图神经网络(GNN)的兴起,文献[14]提出了基于嵌入技术的深度图匹配模型,其中采用图卷积神经网络(GCN) [15]优化点特征并构建线性匹配模型,最终以 sinkhorn [16]算法求解该问题。虽然该方法将 NP 难的二次组合优化问题松弛成可以被高效求解的线性分配问题,其本质上并没有考虑边到边对的二阶相似度信息,最终导致匹配精度的损失。另外一种解决深度图匹配问题的思路是:将图匹配问题转

化为基于关联图上的结点分类问题, 其中关联图的结点编码了点对相似度信息而关联图的边则编码了边对相似度信息。基于这样的思路, 文献[17] [18]提出了基于关联图的结点二值分类问题, 而文献[19]提出了基于关联图的一阶相似度优化模型。然而, 计算关联图的内存资源消耗是高昂的, 因为它平方了图匹配问题的规模, 内存开销占 $O(n_1^2 n_2^2)$ 。

针对上述提及的问题, 本文从 2 个方面对深度图匹配算法进行提升: 1) 提升基于嵌入方法的深度图匹配模型的精度, 通过引入边边相似度信息作为跨图注意力机制, 既保持线性分配问题的高效求解特性, 又提升图匹配的精度。2) 相比于基于关联图的结点分类问题[17] [18] [19]需要高昂的内存开销, 本文基于关联图的分解模型[20], 隐式地构建基于关联图的边-点嵌入操作, 减少内存资源的消耗。其内存开销约占 $O(n_1^2 + n_2^2 + n_1 n_2)$ 。

本文就基于嵌入技术的深度图匹配网络提出了一种改进算法, 首先引入边边相似度信息, 经过关联图的边-点嵌入操作, 将边对相似度信息转化为二图结点之间的注意力系数, 从而构建跨图注意力机制实现点特征之间的跨图消息传递, 在保持线性分配问题的高效求解特点之外提升了精度, 并且通过隐式构建关联图的边-点嵌入计算, 减小了内存开销。

2. 基础知识

一般情况下, 图匹配问题可以被构造二次组合优化问题(Quadratic Combinatorial Programming), 图匹配示例如图 1 所示。



Figure 1. An example of graph matching
图 1. 图匹配示例

考虑两张有向图 $I_1 = (V_1, E_1)$ 和 $I_2 = (V_2, E_2)$, 其中下标 i 代表第 i 张图, V_i 和 E_i 分别表示图 I_i 的点集和边集。令 $n_1 = |V_1|$, $n_2 = |V_2|$ 且 $n_1 \leq n_2$, 则图匹配问题的一般形式:

$$\begin{aligned} \max \quad & J(X) = \text{vec}(X)^T K \text{vec}(X) \\ \text{s.t.} \quad & X \in \{0, 1\}^{n_1 \times n_2}, X \mathbf{1}_{n_2} = \mathbf{1}_{n_1}, X^T \mathbf{1}_{n_1} \leq \mathbf{1}_{n_2} \end{aligned} \quad (1)$$

其中 X 是点-点分配矩阵, $K \in R^{n_1 n_2 \times n_1 n_2}$ 是全局相似度矩阵(等价于关联图), 其对角线元素和非对角线元素分别由点对相似度元素和边对相似度元素构成。图匹配问题满足离散组合约束条件: 即 I_1 上的一个点至多可以和 I_2 上的一个点建立匹配关系, 反之亦然。

因为显式地构建全局相似度矩阵的内存开销是昂贵的, 文献[20]提出了一种对全局相似度矩阵 K 的分解模型, 该分解模型的定义如下:

$$K = \text{diag}(\text{vec}(K_p)) + (G_2 \otimes G_1) \text{diag}(\text{vec}(K_e))(H_2 \otimes H_1)^T \quad (2)$$

其中, $K_p \in R^{n_1 \times n_2}$ 和 $K_e \in R^{e_1 \times e_2}$ 分别代表点对相似矩阵和边对相似矩阵, \otimes 则是 Kronecker 乘积算子。
 $G_i, H_i \in R^{n_i \times e_i}, i \in \{1, 2\}$ 是图 I_i 的点边对应矩阵, 其详细介绍如图 2(b)和图 2(c)所示。

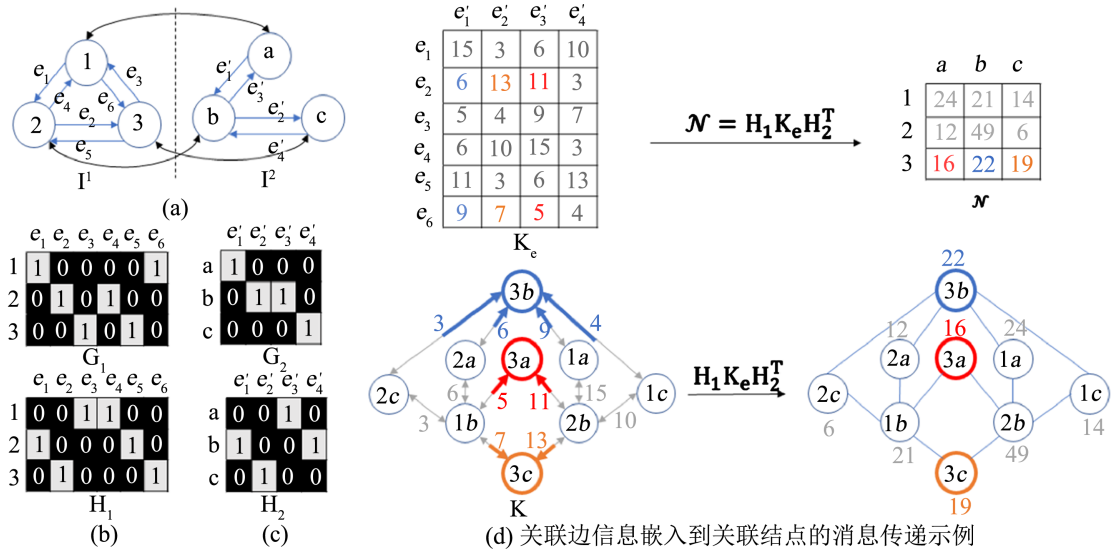


Figure 2. An example of message passing between nodes and edges under an association graph
图 2. 基于关联图的点 - 边消息传递示例

虽然基于嵌入的深度图匹配方法采用图神经网络(GNN)技术将各自图的拓扑连接信息嵌入到点特征中, 从而将图匹配问题松弛至线性分配问题, 但是其本质上没有考虑边对相似信息。从公式(1)和(2)中可以看出, 边对相似信息是图匹配问题中不可或缺的组成部分。针对该问题, 本文提出了基于边对注意力信息的跨图深度图匹配网络, 该方法引入边对相似信息之后, 经过关联图的边-点嵌入操作构建跨图注意力机制, 对二图结点的特征进行优化。

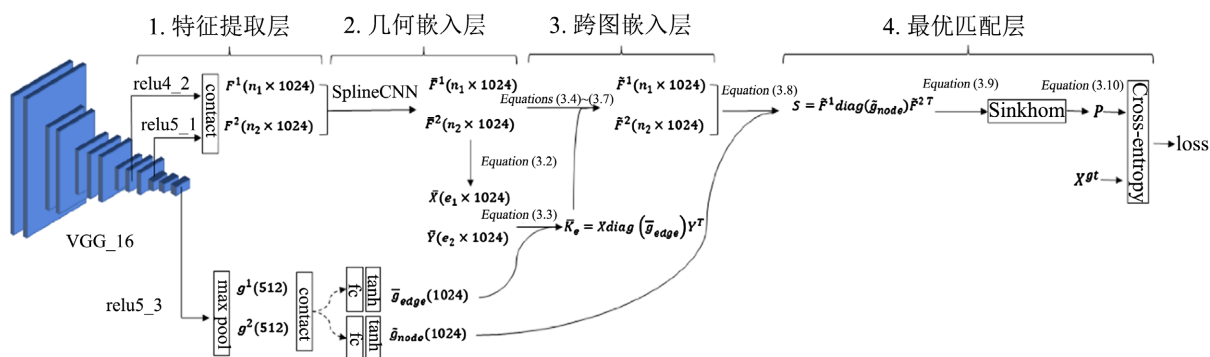
3. 本文方法

图 3 给出了本文方法的概述。本文采用双随机松弛将离散域的组合约束条件松弛至连续域下的组合约束条件, 其目标旨在设计一个神经网络去预测两幅图的软分配矩阵 $P \in [0, 1]^{n_1 \times n_2}$, 该矩阵满足以下组合约束:

$$P1_{n_2} = 1_{n_1}, P^T 1_{n_1} \leq 1_{n_2} \quad (3)$$

3.1. 特征提取层和几何嵌入层

本文采用常规的卷积神经网络 VGG-16 作为初始的特征点特征提取器, 该网络的参数在 ImageNet 数据集上进行了预训练。定义 $F^1 \in R^{n_1 \times d}$ 和 $F^2 \in R^{n_2 \times d}$ 分别表示图 I_1 和 I_2 的结点特征, 其中 d 是特征的维度。 F^1 和 F^2 分别由各自 VGG-16 网络的 relu4_2 和 relu5_1 输出并拼接组成。如图 3 所示, 在特征提取层还有一条分支来自于 relu5_3 的输出, 经过拼接操作得到全局的学习权重 $\bar{g}_{node} \in R^d$ 和 $\tilde{g}_{edge} \in R^d$ 。这种全局学习权重提取技术来自于文献[19], 能够对后端的点对相似性度量和边对相似性度量进行学习和优化。



1. 基于 VGG-16 的特征提取层, 2. 基于 SplineCNN 的几何嵌入层, 3. 基于跨图嵌入模块的跨图嵌入层, 4. 基于 sinkhorn 算法的最优匹配层

Figure 3. An overview of the proposed pipeline: a cross-graph embedding with edge-wise attention for deep graph matching
图 3. 基于边对注意力的跨图嵌入图匹配网络的整体框架图

当采用线性分配模型来求解图匹配问题时, 通过提升特征点之间的差异性, 来提高待匹配特征点之间的相似性得分是常见的。为了提高特征点之间的差异性, 图神经网络可以将结点表示通过聚合来自邻居结点的表示递归计算, 如邻居结点的外观特征和图的拓扑结构信息。为了融合结点的特征信息和图的结构信息, 本文采用 SplineCNN 图卷积神经网络[21]将几何结构信息嵌入到结点特征当中。其中, 两幅图的结构生成自 Delaunay 三角剖分[22]算法, 而边的属性则由两个端点之间的归一化相对坐标定义。然后, 结点特征和边的属性经过 SplineCNN 的最大化聚合, 最终输出嵌入了几何结构信息的点特征 \bar{F}^1 和 \bar{F}^2 。

3.2. 基于跨边注意力得跨图嵌入层

跨图嵌入层旨在建立两幅图之间的特征点消息传递机制。本文引入了边对相似性信息, 作为两幅图结点消息传递机制的桥梁。

如图 2 所示, $G_1, H_1 \in \{0,1\}^{n_1 \times e_1}$ 是图 I_1 的点边对应矩阵, 当 $G_{i,k} = H_{j,k} = 1$ 时, 意味着第 k 条边开始于结点 i 而结束于结点 j , 对于 $G_2, H_2 \in \{0,1\}^{n_2 \times e_2}$ 同样适用。那么边的特征 $\bar{X} \in R^{e_1 \times d}$ 和 $\bar{Y} \in R^{e_2 \times d}$ 分别由各自端点的特征之差来构建:

$$\bar{X} = (G_1^T - H_1^T) \bar{F}^1, \quad \bar{Y} = (G_2^T - H_2^T) \bar{F}^2 \quad (4)$$

边对相似度矩阵 $\bar{K}_e \in R^{e_1 \times e_2}$ 则可以由边的特征内积来测量:

$$\bar{K}_e = \bar{X} \text{diag}(\bar{g}_{edge}) \bar{Y}^T \quad (5)$$

其中 $\bar{g}_{edge} \in R^d$ 是全局可学习权重, $\text{diag}(\cdot)$ 则是构建对角矩阵。

边对相似度矩阵 \bar{K}_e 反映了边与边的匹配信息。为了进一步反映不同边对匹配之间的重要性程度, 本文将非线性归一化操作 softmax 作用于边对相似度矩阵 \bar{K}_e : 增大相似度得分差异的同时强调高相似度得分的重要性程度。经过归一化操作后的相似度矩阵能够反应边对重要性程度, 在本文当中, 也被称之为注意力矩阵 $\varepsilon \in [0,1]^{e_1 \times e_2}$:

$$\varepsilon = \text{softmax}(\bar{K}_e) \quad (6)$$

如图 2(d)所示, 关联图的结点代表了点对相似度信息, 而关联图的边则反映了边对相似度信息, 因此关联图具有跨点匹配和跨边匹配之间的消息传递机制。根据这一消息传递的性质, 将边对的注意力矩阵 $\varepsilon \in R^{e_1 \times e_2}$ 嵌入到点对的跨图转移矩阵 $\tau \in R^{n_1 \times n_2}$:

$$\tau = H_1 \varepsilon H_2^T \quad (7)$$

对于任意两个跨图结点 $i \in V_1$ 和 $j \in V_2$, 给定跨图转移矩阵 τ , 可以构建从结点 j 传递到结点 i 的信息 $m_{j \rightarrow i}$, 该跨图消息传递计算如下:

$$m_{j \rightarrow i} = \sum_{j \in V_2} \tau_{ij} \bar{f}_j^2 \quad (8)$$

由于跨图矩阵 τ 的元素范围是从 0 到 1, 该性质保证了网络的收敛性。最终可以得到更新后的结点特征:

$$\tilde{f}_i^1 = \tilde{f}_i^1 + m_{j \rightarrow i} \quad \forall i \in V_1, j \in V_2 \quad (9)$$

对于第二幅图的结点 $j \in V_2$ 也是相同的跨图操作。

值得注意的是, 基于关联图 K 的结点分类策略[17] [18] [19]也考虑了边对相似度信息, 这一类模型通过对关联图实施图神经网络计算, 从而更新关联图的边和结点表示, 即边对相似度信息和点对相似度信息。然而, 关联图计算的内存开销是高昂的, 占 $O(n_1^2 n_2^2)$ 。与之相比较, 本文的边-点嵌入模块隐式地进行关联图的消息传递计算(公式(6)~(9)), 在内存消耗上具有优势, 约占 $O(n_1^2 + n_2^2 + n_1 n_2)$ 。

3.3. 最优匹配层

最优匹配层作为神经网络的后端, 其目的是在有监督地训练神经网络的情况下, 生成满足约束条件(公式(3))的预测值, 即软分配矩阵 $P \in [0, 1]^{n_1 \times n_2}$ 。最有匹配层通过计算两幅图之间的点对相似度得分 $S \in R^{n_1 \times n_2}$, 寻找令整体相似度得分 $\sum_{i,j} S_{i,j} P_{i,j}$ 最大的软分配矩阵 P 。

经过上述的图嵌入计算, 可以得到两幅图的结点特征 $\tilde{F}^1 \in R^{n_1 \times d}$ 和 $\tilde{F}^2 \in R^{n_2 \times d}$, 点对相似度的得分矩阵 $S \in R^{n_1 \times n_2}$ 的计算如下:

$$S = \tilde{F}^1 \text{diag}(\tilde{g}_{node}) \tilde{F}^{2T} \quad (10)$$

其中, $\tilde{g}_{node} \in R^d$ 是可训练的学习参数, 而 $\text{diag}(\cdot)$ 是建立一个对角矩阵。

图匹配问题的线性分配模型等价于寻找离散分布 a 和 b 之间的最优传输映射(Optimal Transport)。该问题可以采用可微的 sinkhorn 算法[16]求解: 给定点对相似度得分矩阵 S , 经过迭代式地行归一化和列归一化该矩阵地指数映射 $\exp(S)$, 最终得到软分配矩阵 $P \in [0, 1]^{n_1 \times n_2}$:

$$P_{ij} = \text{Sinkhorn}(\exp(S_{ij})) \quad (11)$$

令真值分配矩阵定义为 $X^{gt} \in \{0, 1\}^{n_1 \times n_2}$ 。将经过网络预测的软分配矩阵 $P \in [0, 1]^{n_1 \times n_2}$ 和真值分配矩阵 X^{gt} 作为交叉熵损失函数的输入, 最终输出该模型的误差得分 l :

$$l = -\sum_{i=1}^{n_1} \sum_{a=1}^{n_2} X_{i,a}^{gt} \log P_{i,a} + (1 - X_{i,a}^{gt}) \log(1 - P_{i,a}) \quad (12)$$

4. 实验评估

本文的实验评估是基于两个广泛应用于深度学习图匹配的数据集: Pascal VOC [23]和 Willow ObjectClass [24]。实验采用平均精度来评估不同算法之间的精度表现:

$$acc = \sum_{i \in I_1, j \in I_2} \text{AND}(X_{i,j}^{pred}, X_{i,j}^{gt}) / N \quad (13)$$

其中, $X_{i,j}^{pred} \in \{0, 1\}^{n_1 \times n_2}$ 是软分配矩阵 P 经过匈牙利算法[25]得到的预测分配矩阵, AND 表示与逻辑函数。

实验的比较对象均是基于深度学习的算法, 包含了以下具有代表性的深度学习图匹配模型:

1) GMN [13]: 该模型首次将深度学习技术用于解决图匹配问题, 通过卷积神经网络提取深度点和边特征, 采用谱方法求解图匹配。

2) PCA [14]: 该模型结合了图卷积网络(GCN)将图的邻居结点进行嵌入, 采用 sinkhorn 算法求解图匹配问题。

3) CIE-H [26]: 该模型构建了边特征并将边特征通过图卷积网络嵌入到结点特征, 采用 sinkhorn 算法求解图匹配问题, 并且提出了基于匈牙利的误差设计。

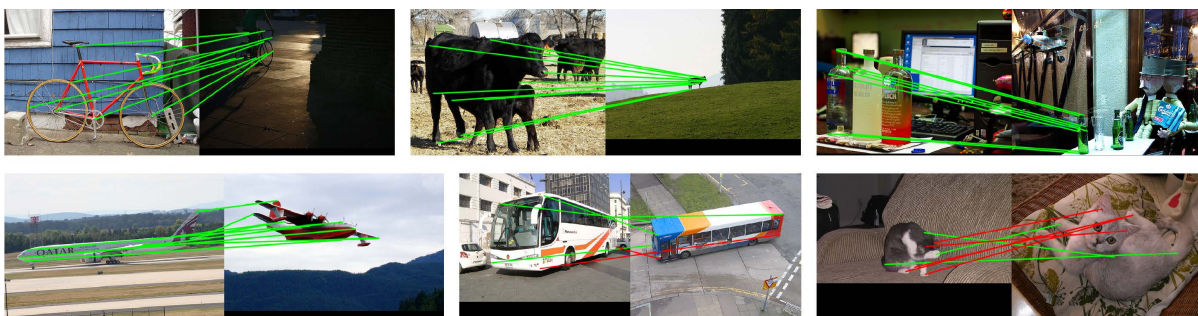
4) LCS [17]: 该模型将图匹配问题转化为基于关联图结点的二值分类问题, 采用卷积神经网络模型对点和边特征进行更新。

5) EAGM [18]: 该模型基于文献[17]扩展了基于像素的边特征生成, 并实现了自适应的边注意力机制, 展现了很好的抗干扰性能。

6) NGM-v2 [19]: 该模型通过将图卷积神经网络的操作扩展至关联图, 实现了对软分配矩阵进行二次优化, 取得了现有方法最优的精度表现。

4.1. PascalVOC 数据集评估

本文的 Pascal VOC [20] 实验设置遵循文献[13]的实验设置, 采用 Berkeley 手工标注特征点[27]作为监督信息。该数据集总共包含 20 类物体, 其中训练集包含 7020 张图片而验证集包含 1682 张图片。每张图片带有 6 至 23 个带有标注信息的特征点。在训练之前, 所有的图片都提取出包含所有标注特征点的方框, 并将该方框调整至 256×256 的像素大小。Pascal VOC 数据集是一个对于图匹配任务极具挑战性的场景, 因为该数据集的一些匹配场景具有大的尺度变形, 姿态变形和光照变形, 如图 4 所示。



绿色线条代表正确匹配而红色线条表示错误匹配。观察可得 Pascal VOC 数据集的一些匹配场景存在较大的尺度和外观变化, 因此该数据集是一个对于图匹配任务具有挑战性的评估平台

Figure 4. Example of graph matching performance of CEGM in Pascal VOC datasets

图 4. 本文提出的 CEGM 模型在 Pascal VOC 数据集下的匹配结果示意图

实验结果如表 1 所示, 本文的 CEGM 神经网络模型在平均精度上超过了已有模型。更具体地, CEGM 在 16 个类别的精度上达到了最优, 而 NGM-v2 在 4 个类别的精度上达到了最优。一方面, CEGM 网络与基于嵌入技术的模型[14] [26]的相同点在于: 都将图匹配问题松弛为线性分配问题, 具有高效求解的特性; 另一方面, 前者在精度上的提升得益于边对相似度信息的引入和计算, 而后者模型忽略了边对相似度信息即两幅图之间的边集一致性。除此之外, 文献[17] [18] [19]也充分考虑了边对相似度信息, 但是这类模型的图更新机制是基于高昂内存开销的关联图, 即全局相似度矩阵 K 。其中, NGM-v2 [19]将图卷积神经网络操作(GCN) [15]作用于关联图, 以此更新关联图的结点表示, 即点对相似度信息, 其内存开销占 $O(n_1^2 n_2^2)$ 。而本文的跨图嵌入模块隐式地构建基于关联图的消息传递机制, 内存开销占 $O(n_1^2 + n_2^2 + n_1 n_2)$ 。因此, CEGM 与 NGM-v2 [19]相比较, 在精度表现上两者接近, 而在内存资源消耗上 CEGM 占优。

Table 1. Average accuracy of each model under Pascal VOC dataset (%)
表 1. 各模型在 Pascal VOC 数据集下的平均精度(%)

模型	飞机	自行车	鸟	船	酒瓶	公交车	汽车	猫	椅子	牛	桌子	狗	马	摩托车	人	盆栽	羊	沙发	火车	电视机	平均精度
GMN [13]	41.6	59.6	60.3	48.0	79.2	70.2	67.4	64.9	39.2	61.3	66.9	59.8	61.1	59.8	37.2	78.2	68.0	49.9	84.2	91.4	62.4
PCA [14]	49.8	61.9	65.3	57.2	78.8	75.6	64.7	69.7	41.6	63.4	50.7	67.1	66.7	61.6	44.5	81.2	67.8	59.2	78.5	90.4	64.8
CIE-H [26]	49.4	63.1	70.7	53.0	82.4	75.4	67.7	72.3	42.4	66.9	69.9	69.5	70.7	62.0	46.7	85.0	70.0	61.8	80.2	91.8	67.6
LCS [17]	46.9	58.0	63.6	69.9	87.8	79.8	71.8	60.3	44.8	64.3	79.4	57.5	64.4	57.6	52.4	96.1	62.9	65.8	94.4	92.0	68.5
EAGM [18]	49.4	62.1	64.6	75.3	90.9	80.9	71.1	61.3	48.7	65.9	87.5	58.4	66.3	60.1	56.3	97.1	64.7	60.6	96.0	93.0	70.5
NGM-v2 [19]	61.8	71.2	77.6	78.8	87.3	93.6	87.7	79.8	55.4	77.8	89.5	78.8	80.1	79.2	62.6	97.7	77.7	75.7	96.7	93.2	80.1
CEGM	60.7	73.9	80.2	80.6	88.5	94.6	88.9	81.6	57.8	79.9	90.5	79.1	79.6	79.6	62.5	98.5	77.7	82.3	98.7	93.2	81.4

4.2. Willow ObjectClass 数据集评估

Willow ObjectClass 数据集[21]总共包含了 256 张图片, 分别来自 5 种类别: 汽车, 鸭子, 脸部, 摩托和酒瓶。每张图片都带有相同的 10 个明显特征点。遵循文献[13]的实验设置, 每张图片都被裁剪至 256 × 256 的方框。对每一个类别进行训练的时候, 前 20 张图片作为训练集, 剩下的图片作为测试集。Willow ObjectClass 数据集相较于 Pascal VOC 数据集, 其尺度和光照变化的程度不如 Pascal VOC 数据集, 且同类物体的姿态保持了对齐, 因此图匹配的难度有所降低。

该数据集的实验结果如表 2 所示, 由于该数据集的难度较之 Pascal VOC 有所下降, 所有的模型在精度上都达到了较好的表现, 其中, 本文的 CEGM 模型在平均精度上取得了最优并略微胜于 NGM-v2 [19]。该实验结果表明: 本文所提出的跨图嵌入图匹配网络 CEGM 在两个公开数据集中均具有精度上的优势。另一方面, CEGM 在内存资源消耗上远远低于文献[17] [18] [19], 即基于关联图更新的深度图匹配网络。因此, CEGM 在精度和内存资源之间的平衡方面达到了目前的最优。

Table 2. Average accuracy of each model under Willow ObjectClass dataset (%)
表 2. 各模型在 Willow ObjectClass 数据集下的平均精度(%)

模型	汽车	鸭子	脸部	摩托	酒瓶	平均精度
GMN [13]	67.9	76.7	99.8	69.2	83.1	79.3
PCA [14]	87.6	83.6	100.0	77.6	88.4	87.4
CIE-H [26]	100.0	90.0	82.2	81.2	97.6	90.2
LCS [17]	91.2	86.2	100.0	99.4	97.7	94.9
EAGM [18]	94.4	89.7	100.0	99.3	99.2	96.5
NGM-v2 [19]	97.4	93.4	100.0	98.6	98.3	97.5
CEGM	97.0	93.3	100.0	99.2	100.0	97.9

4.3. 跨图嵌入模块的比较分析

本文的创新点之一在于提出了一个新的跨图嵌入模块, 与之相似的研究工作则来自于 PCA。在 PCA 模型当中, 其跨图转移矩阵来自于后端的 sinkhorn 匹配层: 首先从 sinkhorn 匹配层计算得到软分配矩阵 $P \in [0,1]^{n_1 \times n_2}$, 经过以下计算更新图 A 和图 B 的结点特征 F^A 和 F^B :

$$\begin{aligned}\hat{F}^A &= [F^A \quad PF^B]W^{cross} \\ \hat{F}^B &= [F^B \quad P^T F^A]W^{cross}\end{aligned}\quad (14)$$

其中, $[\cdot \quad \cdot]$ 表示两个特征向量进行拼接操作, $W^{cross} \in R^{2d \times d}$ 表示跨图嵌入的训练参数, d 表示特征维度, \hat{F}^A 和 \hat{F}^B 表示更新后的结点特征。

从公式(14)可以看出, PCA 的跨图嵌入模块只考虑了点对相似度信息。与之相比较, 本文的跨图嵌入模块(公式(6)~(9))则充分考虑了边对相似度信息, 最终达到精度上的提升。为了进一步验证该假设, 本文设置了一组对照实验: 将 PCA 模型的跨图嵌入模块换成本文的跨图嵌入模块, 保持其它模块不变, 并命名为 PCA-Edge 模型。实验结果如表 3 所示: PCA-Edge 模型在 Pascal VOC 和 Willow ObjectClass 数据集都在平均精度上优于 PCA 模型。该实验表明: 1) 本文提出的跨图嵌入模块有利于现有基于嵌入技术的图匹配模型提高匹配精度; 2) 该跨图嵌入模块能够作为深度图匹配的插件模块, 有效地嵌入到其它深度图匹配模型当中。

Table 3. Average accuracy of cross-graph embedding module comparison experiment (%)

表 3. 跨图嵌入模块的对比实验下的平均精度(%)

模型	Pascal VOC	Willow Object Class
PCA [13]	64.8	87.4
PCA-Edge	65.2	88.3

5. 结论

本文针对现有基于嵌入技术的深度图匹配网络忽略了边对相似度信息的问题, 提出了基于跨图嵌入的深度图匹配模型即充分考虑了图匹配的边集一致性。其中跨图嵌入模块可以作为插件模块, 有利于提高现有图匹配模型的精度。一方面, 本文提出的 CEGM 网络在两个公开数据集上实现了精度上的提升; 另一方面, 相比较于目前最优的 NGM-v2 [19]模型, 本文的跨图嵌入模块隐式地构建基于关联图的边-点嵌入操作, 在内存消耗上优于 NGM-v2 [19]。因此, 本文的 CEGM 模型在精度和内存消耗的平衡上达到了最优, 实验结果证明了该方法的效率和有效性。

参考文献

- [1] Fischler, M.A. and Bolles, R.C. (1981) Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, **24**, 381-395. <https://doi.org/10.1145/358669.358692>
- [2] Zhang, Z. (1994) Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *International Journal of Computer Vision*, **13**, 119-152. <https://doi.org/10.1007/BF01427149>
- [3] 张飞, 黄国兴. 基于 ASIFT 与图匹配的物体识别方法[J]. 计算机工程, 2016, 42(7): 153-158.
- [4] Zhou, F. and De la Torre, F. (2016) Spatio-Temporal Matching for Human Pose Estimation in Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 1492-1504. <https://doi.org/10.1109/TPAMI.2016.2526002>
- [5] He, J., Huang, Z., Wang, N. and Zhang, Z. (2021) Learnable Graph Matching: Incorporating Graph Partitioning with Deep Feature Learning for Multiple Object Tracking. *Proceedings of the IEEE/CVF Conference on Computer Vision*

- and *Pattern Recognition*, Nashville, 20-25 June 2021, 5299-5309. <https://doi.org/10.1109/CVPR46437.2021.00526>
- [6] Chang, J.R. and Chen, Y.S. (2018) Pyramid Stereo Matching Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5410-5418. <https://doi.org/10.1109/CVPR.2018.00567>
- [7] Bai, X., Cheng, J. and Hancock, E. (Eds.) (2012) *Graph-Based Methods in Computer Vision: Developments and Applications: Developments and Applications*. IGI Global, Hershey.
- [8] Leordeanu, M. and Hebert, M. (2005) A Spectral Technique for Correspondence Problems Using Pairwise Constraints. *International Conference of Computer Vision (ICCV)*, Beijing, 17-21 October 2005, 1482-1489.
- [9] Schellewald, C. and Schnörr, C. (2005) Probabilistic Subgraph Matching Based on Convex Relaxation. *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, St. Augustine, 9-11 November 2005, 171-186. https://doi.org/10.1007/11585978_12
- [10] Cho, M., Lee, J. and Lee, K.M. (2010) Reweighted Random Walks for Graph Matching. *European Conference on Computer Vision*, Heraklion, 5-11 September 2010, 492-505. https://doi.org/10.1007/978-3-642-15555-0_36
- [11] Zaslavskiy, M., Bach, F. and Vert, J.P. (2008) A Path Following Algorithm for the Graph Matching Problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, 2227-2242. <https://doi.org/10.1109/TPAMI.2008.245>
- [12] Gold, S. and Rangarajan, A. (1996) A Graduated Assignment Algorithm for Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 377-388. <https://doi.org/10.1109/34.491619>
- [13] Zanfiri, A. and Sminchisescu, C. (2018) Deep Learning of Graph Matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2684-2693. <https://doi.org/10.1109/CVPR.2018.00284>
- [14] Wang, R., Yan, J. and Yang, X. (2020) Combinatorial Learning of Robust Deep Graph Matching: An Embedding Based Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2020.3005590>
- [15] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. arXiv preprint arXiv:1609.02907.
- [16] Cuturi, M. (2013) Sinkhorn Distances: Lightspeed Computation of Optimal Transport. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Vol. 2, Lake Tahoe, 5-10 December 2013, 2292-2300.
- [17] Wang, T., Liu, H., Li, Y., Jin, Y., Hou, X. and Ling, H. (2020) Learning Combinatorial Solver for Graph Matching. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 7565-7574. <https://doi.org/10.1109/CVPR42600.2020.00759>
- [18] Qu, J., Ling, H., Zhang, C., Lyu, X. and Tang, Z. (2021) Adaptive Edge Attention for Graph Matching with Outliers. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, Montreal, 19-27 August 2021, 966-972. <https://doi.org/10.24963/ijcai.2021/134>
- [19] Wang, R., Yan, J. and Yang, X. (2021) Neural Graph Matching Network: Learning Lawler's Quadratic Assignment Problem with Extension to Hypergraph and Multiple-Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2021.3078053>
- [20] Zhou, F. and De la Torre, F. (2015) Factorized Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 1774-1789. <https://doi.org/10.1109/TPAMI.2015.2501802>
- [21] Fey, M., Lenssen, J.E., Morris, C., Masci, J. and Kriege, N.M. (2020) Deep Graph Matching Consensus. *International Conference on Learning Representations*, Addis Ababa, 26-30 April 2020.
- [22] Lee, D.T. and Schachter, B.J. (1980) Two Algorithms for Constructing a Delaunay Triangulation. *International Journal of Computer & Information Sciences*, **9**, 219-242. <https://doi.org/10.1007/BF00977785>
- [23] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J. and Zisserman, A. (2010) The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, **88**, 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [24] Cho, M., Alahari, K. and Ponce, J. (2013) Learning Graphs to Match. *Proceedings of the IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 25-32. <https://doi.org/10.1109/ICCV.2013.11>
- [25] Kuhn, H.W. (1955) The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, **2**, 83-97. <https://doi.org/10.1002/nav.3800020109>
- [26] Yu, T., Wang, R., Yan, J. and Li, B. (2019) Learning Deep Graph Matching with Channel-Independent Embedding and Hungarian Attention. *International Conference on Learning Representations*, New Orleans, 6-9 May 2019.
- [27] Bourdev, L. and Malik, J. (2009) Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations. 2009 *IEEE 12th International Conference on Computer Vision*, Kyoto, 29 September-2 October 2009, 1365-1372. <https://doi.org/10.1109/ICCV.2009.5459303>