

Deep Learning of Graph Matching 论文解读

汪润中



原创技术文章 | 知识库论文讲解 | 精品课程 | 云端实验室



导言

作为一种常用的图数据处理技术,图匹配在计算机视觉中拥有丰富的应用场景和研究价值。CVPR2018 最佳论文提名的工作 Deep Learning of Graph Matching [1]首次将端到端的深度学习技术引入图匹配,提出了全新的深度图匹配框架。本文将首先介绍图匹配问题的背景知识,随后对深度图匹配论文进行深入的解读。

图匹配

图匹配 (Graph Matching) 试图在两个或多个图 (graph) 结构之间,建立节点与节点的对应关系。在计算机视觉领域,图匹配算法通常被用于求解多个图像 (image) 之间,关键点到关键点的匹配关系,如图 1。

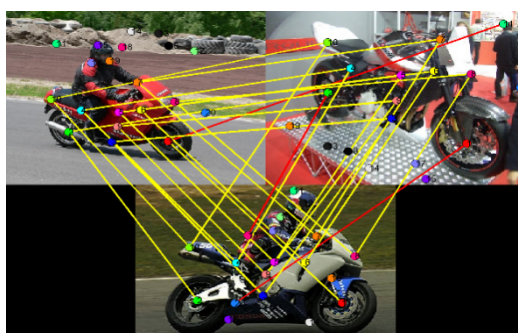


图 1 计算机视觉图匹配示意图[2]0

相比于只考虑节点与节点之间一阶相似度关系的点匹配,图匹配还考虑了图结构中,边到边的二阶相似度,如图 2 所示。图 2 仅展示了图结构中,部分一阶相似度 (蓝色箭头) 与二阶相似度 (红色箭头) 的关系。实际上,在图匹配算法中,任意一对顶点、任意一对边之间,都存在相应的相似度度量。由于额外考虑了图结构中的二阶相似度信息,图匹配通常比简单的一阶点匹配更加精确和鲁棒。

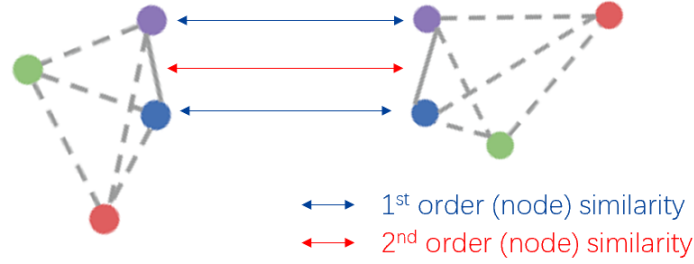


图 2 一阶与二阶相似度示意图 0

考虑两个图结构之间的匹配，我们将一阶、二阶相似度建模为相似度矩阵 M 。假设两图分别有 m, n 个节点，则 $M \in \mathbb{R}^{mn \times mn}$ 。其中 M 的对角线元素代表点对点的一阶相似度，其余非对角元素代表边对边的二阶相似度。图匹配的目标是，找到一个最优的排列矩阵 $X \in \{0,1\}^{m \times n}$ ，最大化如下目标函数

$$J(X) = \max_x x^T M x$$

$$s.t. \quad X\mathbf{1} = \mathbf{1}, X^T\mathbf{1} \leq \mathbf{1}, X \in \{0,1\}^{m \times n}$$

其中， $x = \text{vec}(X)$ 是将矩阵 X 按列向量化的结果， $\mathbf{1}$ 是元素全为 1 的列向量。该问题属于 NP-难的二次指派问题。研究者们已经提出了许多算法，在合理的时间复杂度下尽可能精确地求解该问题。对图匹配算法感兴趣的读者可以参考综述[3]。

除了匹配两个图结构，研究者们还提出了同时匹配多个图结构的多图匹配算法，如[2]。在本文中，我们只考虑两个图结构之间匹配的简单情况，即二图匹配。

深度图匹配简介

论文 Deep Learning of Graph Matching [1]获得了 CVPR2018 最佳论文的提名，这是第一篇结合了端到端深度学习以及图匹配的研究工作。过去的计算机视觉图匹配研究工作，研究者们大多使用 SIFT 等描述子或是手工定义的特征。这些人为构建的特征容易受样本噪声的影响，研究者们往往忽视了机器学习尤其是深度学习在图匹配问题上的巨大潜力。这篇工作将深度学习与图匹配结合，使用卷积神经网络 CNN 提取图像特征，同时学习构建相似度的匹配函数，通过反向梯度传播以及标准的深度学习梯度优化算法，实现端到端的训练。

论文中提出的深度图匹配框架使用现有的 CNN 结构 (VGG16 [4]) 从图像中提取图结构的一阶与二阶特征，基于这些特征构建上文中所述的相似度矩阵 M 。随后，将相似度矩阵 M 输入已有的图匹配求解算法[5]进行求解，得到匹配结果。深度图匹配的前向传播路径如图 3 中黑色实线所示。在训练时，计算匹配结果与监督信息之间的损失函数，通过图 3 中红色虚线所示的路径将梯度反向传播，实现端到端的参数更新。

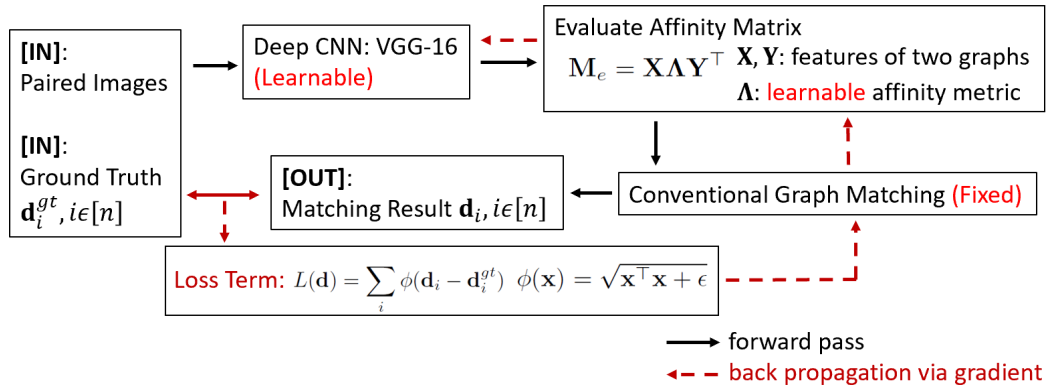


图 3 深度图匹配概览 0

如图 3 所示，在论文中提出的深度图匹配框架下，提取特征的 CNN 网络、衡量相似度的匹配函数具有可学习的参数。论文中采用了已有的图匹配算法[5]，求解相似度矩阵到匹配结果的映射。该图匹配算法具可微分的特性，能够在输入与输出之间传递梯度关系。因此，通过一个完全可导的图匹配框架，作者实现了端到端的深度图匹配。在之后的部分，我们将对深度图匹配论文中的细节进行解读。

细节解读

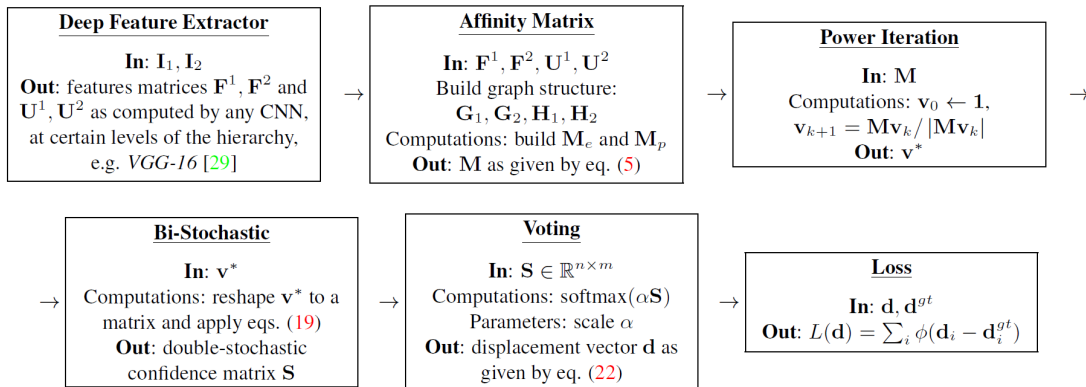


图 4 前向传播框图，摘自论文[1]0

深度图匹配的前向传播框图如图 4 所示，包含了以下步骤：提取深度特征（Deep Feature Extractor）、构建相似度矩阵（Affinity Matrix）、求解图匹配（Power Iteration）、双随机化（Bi-Stochastic）、投票（Voting）以及计算损失（Loss）。在剖析前向传播的细节之前，我们先介绍深度图匹配问题的基本设定。

问题设定 考虑一幅源图片（source image）和一幅目标图片（target image），它们中包含同类别的物体，都具有关键点及其对应关系的标注。在训练和测试时，我们使用标注信息（ground truth）作为源图片的关键点，同时在目标图片中进行网格采样（grid sampling，例如 16×16 采样）作为候选的对应点。这样，就能够模拟实际应用中，已知关键点的源图与未知关键点的目标图之间的匹配关系。对于源图片，论文中使用德罗尼三角剖分（Delaunay Triangulation）构建图结构；对于目标图片，则对所有候选点构建全连接的结构。

提取深度特征 论文中使用了在 ImageNet 分类任务上预训练的 VGG16[4]网络作为深度特征提取器，学习最适合图匹配问题的图像特征。实际上，由于 CNN 网络的中间层能够学习到图像中的共性特征，任何 CNN 模型稍作修改，都可以作为深度图匹配模型的特征提取器。在论文的实现中，作者在不同深度的中间层，分别提取一阶和二阶特征：在 VGG16 网络中，相对浅层的 relu4_2 激活信息被用于构建一阶特征，相对深层的 relu5_1 激活信息被用于构建二阶特征。根据深度学习中感受野（Receptive Field）理论，浅层的神经元拥有更小的感受野，包含更多的局部（local）信息，因此适合建模一阶特征；深层的神经元拥有更大的感受野，包含更多的全局（global）信息，因此适合建模二阶特征。在论文中，提取到的一阶特征由 U^1, U^2 表示，二阶特征由 F^1, F^2 表示，上标的 1,2 代表它属于第一或第二张图。在使用 VGG16 网络时， U, F 都拥有 512 维的特征维度。

构建相似度矩阵 论文中使用了深度学习提取的特征向量，基于人工指定的图结构，构建图匹配的相似度矩阵。其中，一阶相似度矩阵直接由一阶特征点乘得到

$$M_p = U^1 U^{2T}$$

为获得二阶相似度，首先将每条边对应的两个节点的 F 连接，构建两个图的二阶特征矩阵 X, Y （特征维度 1024）

$$X_c = [F_i^1 | F_j^1], Y_c = [F_i^2 | F_j^2]$$

之后构建二阶相似度矩阵

$$M_e = X \Lambda Y^T$$

二阶相似度包含了可学习的参数 $\Lambda \in \mathbb{R}^{1024 \times 1024}$ ，因而论文中的二阶相似度具有一个可学习的匹配函数。获得 M_p, M_e 后，包含一阶、二阶相似度的相似度矩阵 M 可由论文中的式（22）构建，在此不再赘述。

求解图匹配 根据使用的图匹配算法[5]，图匹配的求解可以近似为求解相似度矩阵最大特征值所对应的特征向量（即 leading eigenvector，以下简称最大特征向量）。幂迭代（Power Iteration）即是一种求解最大特征向量的迭代算法。初始化 $v_0 = \mathbf{1}$ ，通过不断迭代， v_k 收敛到矩阵 M 的最大特征向量

$$v_{k+1} = \frac{M v_k}{\|M v_k\|}$$

分母的符号表示二范数。

双随机化 双随机矩阵的定义如下：对于一个方阵 $X \in [0,1]^{n \times n}$ ，若其每行、每列的求和均为 1，则该矩阵称为双随机矩阵。在图匹配问题中，使用双随机矩阵表示匹配结果，可以直观地体现任意一对节点建立匹配关系的可能性，所以图匹配问题经常约束其结果为双随机矩阵。由 Power Iteration 算法求解出的匹配结果不满足双随机性，因此需要将其进行双随机化。我们使用迭代算法将矩阵双随机化：首先将矩阵按列归一化，随后将矩阵按行归一化。通过不断迭代，即可得到双随机矩阵。该步骤的数学表示为

$$S_{k+1} = S_k [1_n^T S_k]^{-1}, S_{k+2} = [S_{k+1} 1_m]^{-1} S_{k+1}$$

投票 在算法实现时，上一步得到的双随机矩阵，同一行、同一列的元素的值相差不大。为

了使差异明显，并为后续计算损失函数提供方便，作者将上一步得到的双随机矩阵乘以一个常数 α （论文中 $\alpha = 200$ ），随后进行 softmax 处理，得到每个候选节点匹配的“可能性”矩阵。投票步骤输出的矩阵，仍然满足双随机的性质。

计算损失 作者使用了一种基于偏移向量的方法计算损失函数，见图 5。源图片中的关键点（第一张图，三角形点）在目标图片中的对应点（第二张图，三角形点）的位置。它们之间的偏移量，就是偏移向量（第三张图，红色箭头）。

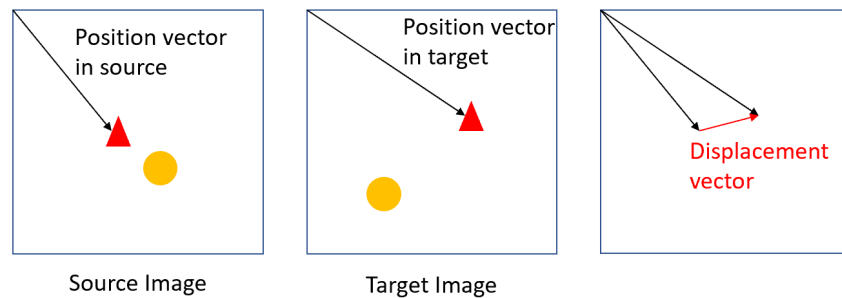


图 5 偏移向量 0

首先，使用数据集的标注信息，可以计算真实的偏移向量 \mathbf{d}^{gt} 。随后，根据投票步骤获得的可能性矩阵，通过加权求和，可以得到模型预测的偏移向量 \mathbf{d} 。论文中采用了如下的损失函数，最小化模型预测的偏移向量与真实的偏移向量之间的差

$$L(\mathbf{d}) = \sum_i \phi(\mathbf{d}_i - \mathbf{d}_i^{gt}) \quad \phi(\mathbf{x}) = \sqrt{\mathbf{x}^\top \mathbf{x} + \epsilon}$$

端到端训练 在模型训练时，得到损失函数后，就可以通过反向梯度传播算法，更新模型各个部分的参数。在论文中，作者详细地推导了提取深度特征、构建相似度矩阵、求解图匹配、双随机化等步骤的反向梯度。反向梯度的数学推导较为复杂，并且当下先进的开源深度学习框架都支持自动计算反向梯度，本文限于篇幅不做讨论，感兴趣的读者可以阅读论文的第三节。

总结

Deep Learning of Graph Matching 这篇工作首次将端到端的深度学习与图匹配问题结合，在学术界已经引起了不小的关注。结合机器学习，尤其是深度学习，提升传统计算机视觉算法的精度，是学术界发展的趋势之一。基于这篇工作，后续可能可以从主干网络、匹配函数、图结构学习、图匹配求解入手，进一步提高深度学习图匹配算法的精度。

参考文献

- [1] Zanfir A, Sminchisescu C. Deep Learning of Graph Matching. Computer Vision and Pattern Recognition, 2018.
- [2] Yan J, Cho M, Zha H, et al. Multi-Graph Matching via Affinity Optimization with Graduated Consistency Regularization. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(6):1228-1242.
- [3] Yan J, Yin X, Lin W, et al. A short survey of recent advances in graph matching. International Conference on Multimedia Retrieval.

- [4] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Representation Learning, 2015.
- [5] Leordeanu M, Hebert M. A spectral technique for correspondence problems using pairwise constraints. International Conference on Computer Vision, 2005



原创技术文章 | 知识库论文讲解 | 精品课程 | 云端实验室

