高阶马尔科夫随机场及其在场景理解中的应用

余 淼 [1, 2] 胡占义 [1]

摘 要 与传统的一阶马尔科夫随机场 (Markov random field, MRF) 相比, 高阶马尔科夫随机场能够表达更加复杂的定性和统计性先验信息, 在模型的表达能力上具有更大的优势. 但高阶马尔科夫随机场对应的能量函数优化问题更为复杂. 同时其模型参数数目的爆炸式增长使得选择合适的模型参数也成为了一个非常困难的问题. 近年来, 学术界在高阶马尔科夫随机场的能量模型的建模、优化和参数学习三个方面进行了深入的探索, 取得了很多有意义的成果. 本文首先从这三个方面总结和介绍了目前在高阶马尔科夫随机场研究上取得的主要成果, 然后介绍了高阶马尔科夫随机场在图像理解和三维场景理解中的应用现状.

关键词 高阶马尔科夫随机场, 能量模型, 能量优化, 参数学习, 场景理解

引用格式 余淼, 胡占义. 高阶马尔科夫随机场及其在场景理解中的应用. 自动化学报, 2015, **41**(7): 1213−1234

DOI 10.16383/j.aas.2015.c140684

Higher-order Markov Random Fields and Their Applications in Scene Understanding

YU Miao[1, 2] HU Zhan-Yi[1]

Abstract Compared with traditional first-order Markov random fields (MRF), higher-order Markov random fields could incorporate more sophisticated qualitative and statistical priors, thus have much more expressive power of modeling. However, it is even harder to minimize their corresponding energy functions. Besides, estimating the value of their parameters becomes much more complex due to the explosive growth of their number. Currently, numerous works have been devoted to solving the modeling, inference and parameter learning problems of higher-order random fields. This paper is a review of the related works as well as a short summary of the applications of higher-order Markov random fields to image understanding and 3D scene understanding.

Key words Higher-order Markov random fields, energy modeling, energy minimization, parameter learning, scene understanding

Citation Yu Miao, Hu Zhan-Yi. Higher-order Markov random fields and their applications in scene understanding. *Acta Automatica Sinica*, 2015, **41**(7): 1213−1234

很多视觉问题, 如图像分割、立体匹配、图像降噪、图像理解等都可以转化成一个标注问题 (Labeling problem), 进而归结为一个马尔科夫随机场 (Markov random fields, MRF) 框架下概率分布函数为指数形式的最大后验概率估计 (Maximum a posteriori, MAP) 问题, 从而可以利用能量函数优化的方法求解[1−3]. 由于存在较为成熟的能量优化算法, 如 Graph cuts[4−5], LBP (Loopy belief propagation)[6−8] 和 TRW (Tree-reweighted message passing)[9−12] 等, 传统的低阶马尔科夫随机场 (First-order Markov random fields) 在很多视觉问题上取得了成功的应用[13]. 但其包含的低阶能量项往往仅能表达邻域平滑等简单的先验知识. 而高阶马尔科夫随机场 (Higher-order Markov random fields) 由于能够表达更加复杂的先验知识及统计信息, 近年来在计算机视觉界得到了广泛的重视并取得了很多有意义的研究成果. 本文将对高阶马尔科夫随机场的研究中取得的主要成果及其在场景理解中的一些典型应用进行总结和介绍.

收稿日期 2014-09-24 录用日期 2015-03-20
Manuscript received September 24, 2014; accepted March 20, 2015
国家高技术研究发展计划 (863 计划) (2013AA122301), 国家自然科学基金 (61273280, 61333015) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2013AA122301) and National Natural Science Foundation of China (61273280, 61333015)
本文责任编委 杨健
Recommended by Associate Editor YANG Jian
1. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190 2. 中原工学院电子信息学院 郑州 450007
1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 2. School of Electric and Information Engineer, Zhongyuan University of Technology, Zhengzhou 450007

1 高阶能量模型的优势和潜力

1.1 高阶马尔科夫随机场及高阶能量函数

对于一个无向图 $G = (\mathcal{V}, \mathcal{E})$, 其中 $\mathcal{V}$ 是节点

的集合, $\mathcal{E}$ 是无向边的集合. 用 X_i 表示节点 $i \in \mathcal{V}$ 对应的随机变量, x_i 表示该随机变量的取值, $\mathcal{X}_i$ 表示该随机变量的状态空间, 即 $x_i \in \mathcal{X}_i$. 所有随机变量的集合记为 $\boldsymbol{X} = \{X_i | i \in \mathcal{V}\}$, $\boldsymbol{x} = \{x_i | i \in \mathcal{V}\}$ 为该随机变量集合的实现 (realization or configuration), 其取值空间 $\mathcal{X}$ 为所有随机变量状态空间的笛卡儿积, 即 $\mathcal{X} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$. 若该随机变量的联合概率分布严格为正 ($p(\boldsymbol{x}) > 0, \forall \boldsymbol{x} \in \mathcal{X}$) 且满足如下式所示的局部马尔科夫性[14−17].

$$X_i \perp\!\!\!\perp \boldsymbol{X}_{\mathcal{V} \setminus \mathrm{cl}(i)} \mid \boldsymbol{X}_{\mathrm{ne}(i)} \tag{1}$$

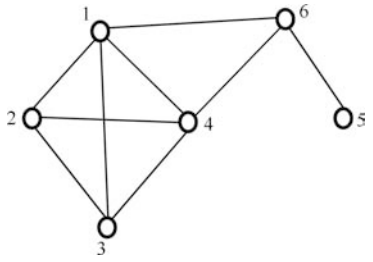
其中, $\mathrm{ne}(i)$ 为节点 i 的邻域集合, $\mathrm{cl}(i) = \{i\} \cup \mathrm{ne}(i)$ 为节点 i 的闭邻域集合 (Closed neighborhood). $X_i \perp\!\!\!\perp X_j \mid X_k$ 表示在给定 X_k 的情况下, X_i 和 X_j 满足条件独立. 则称随机变量的集合 $\boldsymbol{X}$ 是由无向图 G 所描述的马尔科夫随机场. 由 Hammersley-Clifford 定理[18−19], 此时随机变量 $\boldsymbol{X}$ 的联合概率分布为 Gibbs 分布, 可表示为如下的形式:

$$p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{w}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(\boldsymbol{x}_c | \boldsymbol{w}) \tag{2}$$

其中, Z 为归一化因子, c 是图 G 上的团 (Clique), 即全连通子图的节点集合, 若团 c 不被其他任何团所包含, 即 $c \not\subset c', \forall c' \in \mathcal{C} \setminus c$, 则称 c 为极大团 (Maximal clique). $\phi_c(\boldsymbol{x}_c | \boldsymbol{w})$ 为团 c 的取值恒为正的团势函数 (Clique potential), $\boldsymbol{w}$ 为模型的参数. 一般假设 $\mathcal{C}$ 为图 G 上所有极大团的集合. 该假设并非限制性的条件, 因为对任何含有非极大团的表示形式, 总可以通过重新定义极大团的团势函数为其所包含的全部团势函数的乘积, 从而将其转化为仅包含极大团的表示形式. 有时为了计算方便, 也常使用含有非极大团的表示形式.

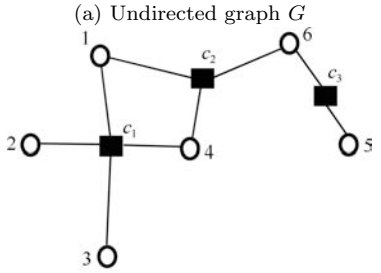
为了更加直观地表示如式 (2) 所示的基于团势函数的联合概率分布的分解形式, 常使用因子图 (Factor graph)[20−21]. 记无向图 $G = (\mathcal{V}, \mathcal{E})$ 的因子图为 $G' = (\mathcal{V}, \mathcal{C}, \mathcal{E}')$, 其中 $\mathcal{V}$ 为图 G 的节点集合, $\mathcal{C}$ 为图 G 上团的集合, $\mathcal{E}'$ 为新的无向边的集合. 因子图 G' 为关于 $\mathcal{V}$ 和 $\mathcal{C}$ 的二分图 (Bipartite graph), 即当且仅当 $i \in c$ 时, 存在边 $(i, c) \in \mathcal{E}'$. 图 1 (a) 和 (b) 为一个无向图及其等价的因子图的表示.

在图 1 (a) 中图 G 的节点集合 $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$, 共有 3 个极大团, 分别为 $c_1 = \{1, 2, 3, 4\}$, $c_2 = \{1, 4, 6\}$ 和 $c_3 = \{5, 6\}$. 其等价的因子图 G', 如图 1 (b) 所示. 因子图 G' 包含两类节点: $\mathcal{V}$ 和 $\mathcal{C} = \{c_1, c_2, c_3\}$, $\mathcal{C}$ 为图 G 的全部极大团的集合. 因子图的节点集合 $\mathcal{C}$ 中的每个元素 c 对应式 (2) 中的团势函数 $\phi_c(\boldsymbol{x}_c | \boldsymbol{w})$, 故因子图非常直观地表示了马尔科夫随机场中随机变量的联合概率分布在团势函数上的分解.



(a) 无向图 G

(a) Undirected graph G



(b) 有向图 G

(b) Directed graph G

图 1 无向图及其等价的因子图表示

Fig. 1 Undirected graph and its corresponding factor graph

因为团势函数恒为正, 可定义团能量函数如下:

$$\psi_c(\boldsymbol{x}_c | \boldsymbol{w}) = -\log \phi_c(\boldsymbol{x}_c | \boldsymbol{w}) \tag{3}$$

此时随机变量的联合概率分布可等价地表示为

$$p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{w}) = \frac{1}{Z} \exp\{-E(\boldsymbol{x} | \boldsymbol{w})\} \tag{4}$$

其中

$$E(\boldsymbol{x} | \boldsymbol{w}) = \sum_{c \in \mathcal{C}} \psi_c(\boldsymbol{x}_c | \boldsymbol{w}) \tag{5}$$

为马尔科夫随机场的能量函数.

因为 "−log" 函数为单调函数, 故在团能量函数, 如式 (3) 所示的定义下, 随机变量 $\boldsymbol{X}$ 的最大后验概率估计问题转化为了如下的能量函数最小化问题,

$$\overset{*}{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathcal{X}} E(\boldsymbol{x} | \boldsymbol{w}) \tag{6}$$

式 (3) 所示的团能量函数 $\psi_c(\boldsymbol{x}_c | \boldsymbol{w})$ 的阶次为团 c 所包含的元素的个数, 记为 $|c|$. 也等于其因子图中节点 c 的度 (Node degree). 能量函数 $E(\boldsymbol{x} | \boldsymbol{w})$ 的阶次为其包含的全部团能量函数 $\psi_c(\boldsymbol{x}_c | \boldsymbol{w})$ 阶次的最大值. 也等于其因子图中 $\mathcal{C}$ 中全部节点的度的最大值. 马尔科夫随机场的阶次为其能量函数的阶次减 1. 图 2 (a) 和 (b) 的能量函数的阶次均为 2, 也称为 Pairwise 能量函数, 其马尔科夫随机场的阶次均为 1, 称为一阶马尔科夫随机场. 当阶次大于 1 时, 马

尔科夫随机场为高阶马尔科夫随机场, 而阶次大于 2 的能量函数称为高阶能量函数. 图 1 是一个高阶马尔科夫随机场的例子, 其阶次为 3, 对应的高阶能量函数的阶次为 4.

由式 (4) ∼ (6) 可知, 高阶马尔科夫随机场的随机变量的分布可完全由其对应的能量函数刻画, 故高阶马尔科夫随机场的研究往往归结于对高阶能量模型的研究. 高阶能量模型包含以下三个主要任务: 建模 (Modeling)、优化 (Inference) 和参数学习 (Parameter learning). 建模的任务主要包含: 1) 选择合适的变量 $\boldsymbol{x}$ 用以表达问题的解; 2) 选择合适的高阶能量项 $\psi_c(\boldsymbol{x}_c|\boldsymbol{w})$ 对变量之间 (先验知识) 及变量和观测之间的关系进行建模. 能量优化的主要任务是由式 (6) 求取高阶能量模型的最优解 $\overset{*}{\boldsymbol{x}}$. 参数学习的主要任务是基于训练数据选择合适的模型参数 $\overset{*}{\boldsymbol{w}}$. 相比于传统的 Pairwise 能量模型, 高阶能量模型具有更好的模型表达能力, 从而在场景理解的应用中具有更大的优势和潜力.

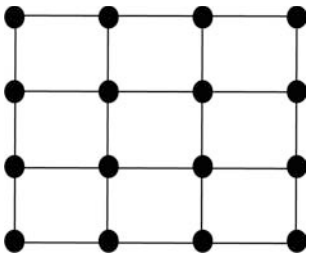
1.2　高阶能量模型在模型表达能力上的优势

传统的低阶 (Pairwise) 能量模型可表示为如下形式:

$$E(\boldsymbol{x}) = \sum_{i\in\mathcal{V}} \psi_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \psi_{i,j}(x_i, x_j) \qquad (7)$$

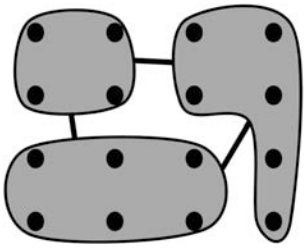
其中, $\psi_i(x_i)$ 为一阶项, 其值为节点 i 取值为 x_i 时的代价, 一般用来衡量随机变量 X_i 的取值与观测数据之间的吻合程度, 故称之为数据项. $\psi_{i,j}(x_i, x_j)$ 为二阶项, 其值为相邻节点 i, j 取值为 x_i, x_j 时的代价, 一般用来约束在空间上相邻节点的取值尽可能地接近, 故称之为平滑项. 图 2 是在视觉问题中常见的两种低阶能量模型. 在图 2 (a) 中, "节点" 为像素, 像素的邻域为四邻域, 使得 "节点" 间的连接为规则网格; 图 2 (b) 中, "节点" 是超像素, 拥有共同边界的超像素相邻, "节点" 间的连接不规则. 由于能够在同一能量优化框架下对观测数据和平滑先验进行建模, 且存在高效、高质量的优化算法, Pairwise 能量函数在计算机视觉领域得到了非常广泛的应用[1, 13, 22].

但由于 Pairwise 能量函数的形式过于简单, 其最大团的阶次为 2, 这意味着只有当任意三个随机变量之间至少存在两个条件独立的随机变量时, 其联合概率分布才能被 Pairwise 能量函数所对应的一阶马尔科夫随机场表达. 如此苛刻的条件独立性要求大大限制了 Pairwise 能量模型的表达能力, 使得其无法表达区域和全局先验, 甚至在其已经取得成功应用的例子上, Pairwise 能量函数也被发现存在建模不够准确的问题. 如早在 1999 年 Veklser 在其博士论文中发现 Greig 等[23] 提出的二值图像去噪的例子中 (Graph cuts 在视觉问题上的第一次应用), 无论 Pairwise 能量函数的一阶和二阶项如何取值, 真值 (Ground truth) 都不可能在该能量函数的最小值或其邻域内得到. 这说明若仅将能量函数限制为二阶, 则不可能对二值图像去噪进行准确的建模. Tappen 等[24] 的进一步研究发现, 无论对于 Graph cuts 或 BP 算法, 其所求取的能量函数的最小值通常比其真值所对应的能量函数值低 40 % 甚至更多, 且随着能量函数值的减小其解的质量往往会变差. 在这种情况下能量函数的进一步优化反而会得到更差的结果. 故 Pairwise 能量函数固有的 "表达能力不足" 所导致的建模不准确性是阻碍其进一步应用的瓶颈.



(a) "节点" 为像素

(a) Node is pixel.



(b) "节点" 为超像素

(b) Node is super-pixel.

图 2　视觉问题中常见两种的低阶能量模型

Fig. 2　Two commonly used pairwise energy models in computer vision problems

高阶能量模型的一般形式可表示如下:

$$E(\boldsymbol{x}) = \sum_{i\in\mathcal{V}} \psi_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \psi_{i,j}(x_i, x_j) + \sum_{\substack{c\in\mathcal{C}\\ |c|\geq 3}} \psi_c(\boldsymbol{x}_c) \qquad (8)$$

其中, $\psi_c(\boldsymbol{x}_c), |c| \geq 3$ 为高阶能量项. 引入高阶能量项可极大地增强能量模型对各种先验知识的表达能力. 具体而言, 高阶能量模型在模型表达能力上的优势主要体现在以下两个方面.

1.2.1　高阶能量模型能够更为有效地描述 "定性先验信息"

通过引入先验知识消除视觉问题中固有的 "不

确定性” 是很多视觉问题获得更好解的关键所在. 先验知识从来源上可分为 “定性先验信息” 和由各种监督、非监督的机器学习算法得到的 “统计性先验信息”.

“定性先验信息” 主要包含对场景结构的定性描述和人的经验性知识. 例如, 在城市场景的三维重建中, 存在大量的人造建筑, 这些建筑主要由平面构成, 即存在 “共面性先验”. 类似的例子也出现在立体匹配中. 在立体匹配问题中, 其能量函数的数据项主要对灰度一致性进行建模, 即同一空间点在不同图像上的观测应具有一致性. 而平滑项主要对场景的结构先验进行建模, 其一般形式可表示如下:

$$E_{\text{smooth}}(\boldsymbol{x}) = \sum_{c\in\mathcal{C}} w_c \rho_c \left(S\left(\boldsymbol{x}_c\right)\right) \tag{9}$$

其中, w_c 为平滑项的系数, ρ_c 常取截断的二阶多项式核函数. x_i 代表像素 i 的深度, $S(\boldsymbol{x}_c)$ 一般定义为深度的导数. 当 $E_{\text{smooth}}(\boldsymbol{x})$ 为 Pairwise 能量项时, 此时 $c = \{p, q\}$ 仅包含两个像素节点, $\mathcal{C}$ 为图像上全部 1×2 和 2×1 的相邻像素的集合, $S(\boldsymbol{x}_{p,q})$ 为深度图的一阶导数,

$$S(\boldsymbol{x}_{p,q}) = x_p - x_q \tag{10}$$

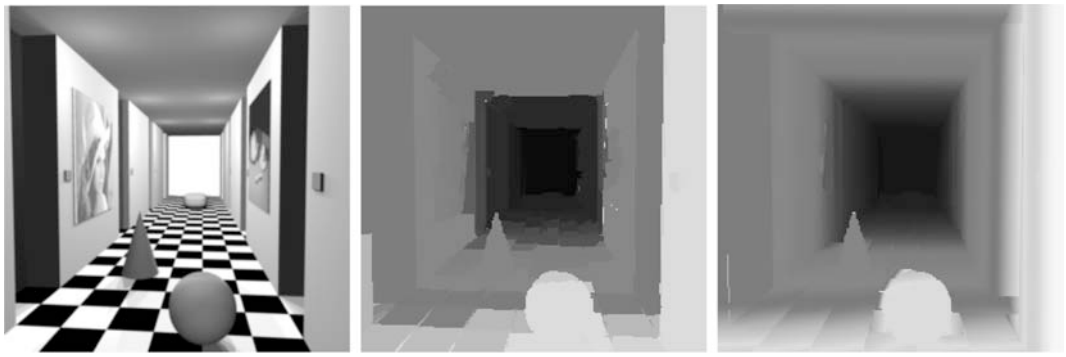
此时, 平滑项 $E_{\text{smooth}}(\boldsymbol{x})$ 为一阶平滑项, 其仅对空间点位于与相机主平面平行 (Fronto-parallel planes) 的同一平面上的图像点 p, q 的惩罚为 0. 即一阶平滑项鼓励相邻像素取相同深度, 这使得其仅能约束与相机主平面平行的同一平面上的点的 “共面性”. 而对大多数场景, 甚至包括人造场景而言, 其大多数空间平面并不与相机的主平面平行. 在这种情况下利用该一阶平滑先验难以得到令人满意的深度图. 图 3 (a) 是一个人造场景的例子. 图 3 (b) 是使用一阶平滑项 (Pairwise 能量函数) 在该参考图像上得到的深度图. 该深度图呈现 “分块常数” 的特点, 无法反映场景中存在较多平面, 且深度连续变化的特点.

为了约束任意方向的同一空间平面上的点的 “共面性”, 至少需要同时考虑三个图像点的深度, 仅利用 Pairwise 能量函数无法对该 “共面性” 先验进行建模. 针对这个问题, Woodford 等[25] 提出了如下的二阶平滑先验:

$$S(\boldsymbol{x}_{p,q,r}) = x_p - 2x_q + x_r \tag{11}$$

其中, $c = \{p, q, r\}$, $\mathcal{C}$ 为图像上所有相邻的 1×3 和 3×1 的像素的集合. 此时, 平滑项 $E_{\text{smooth}}(\boldsymbol{x})$ 为二阶平滑项, 若 $\{p, q, r\}$ 三点共面, 则该平滑项对其的惩罚为 0. 图 3 (c) 为利用该二阶平滑项在参考图像上得到的深度图. 与图 3 (b) 的结果相比, 该二阶平滑项能够较好地保持与相机主平面不平行同一空间平面上的点的 “共面性”. 该论文[25] 获得了 CVPR 2008 年的最佳论文奖. 可见, 仅需将 Pairwise 能量项的阶次加 1, 其得到的最低阶的高阶能量项即可显著地增强对 “空间平面结构” 这类 “定性先验” 的描述能力.

除了 “共面性” 外, 其他关于场景结构的定性先验还包括对称性、结构相似性等. 又如在图像分割中, 传统的基于 Graph cuts 的图像分割算法存在 Shrinking bias, 使得被分割物体的细长部分容易被截掉, 而基于主动轮廓或水平集的方法易陷入局部极值, 求解质量无法得到保证. 而一般来说由于物体是 “连通的”, 通过在全局能量最小化的框架下引入 “连通性先验” 可使得物体分割问题得到更好的解决[26−27]. 物体连通性的先验也被引入到了三维重建中[28]. 类似的例子还出现在利用限界框 (Bounding box) 进行交互式的图像分割中[29]. 其他的 “定性先验” 还包括在场景理解中物体之间关系, 相对的位置、大小、运动等的先验.



(a) 参考图像 (a) Reference image　(b) 一阶平滑项 (b) First-order prior　(c) 二阶平滑项 (c) Second-order prior

图 3 利用一阶和二阶平滑项得到的深度图对比 (实验结果来自文献 [25])

Fig. 3 Comparison of first and second order smoothness priors (The experiment results are from [25].)

1.2.2 高阶能量模型能够更为有效地描述 “统计性先验信息”

“统计性先验信息” 主要指由各种监督和非监督的机器学习算法得到的先验信息. 如区域一致性先验[30−35] 和 Pattern-based 先验[36−39]. 区域一致性先验是指由各种非监督的聚类算法对图像进行过分割得到的每个过分割区域内 (超像素内) 的像素应该具有相同的类别 (或大多数像素取相同的类别). Pattern-based 先验可看作是区域一致性先验的推广, Pattern 指的是区域内部先验的 “模式”. 当这种 “模式” 要求区域内部满足类别一致性时, Pattern-based 先验就变成了区域一致性先验. 如在纹理图像合成、去噪以及图像超分辨率等问题中, 通过在训练样本上对图像片的 Pattern 进行学习, 得到 Pattern-based 先验, 从而最终提高这些问题的求解效果. 在场景理解中, 可以通过发掘标记之间的统计特性从而得到关于标记的先验约束, 如: 1) “共生概率”[40−41]: 通过学习不同类别的物体在同一场景中共同出现的概率, 从而得到 “共生概率” 约束. 这是一种全局性的先验; 2) 关于标记数目的约束[42−43]: 如在三维重建中, 已知物体的大致形状即可获得关于重建物体标记数目的约束; 3) 标记代价约束[44−45]: 该约束可看作是对同一场景中出现的不同类别赋予不同的 “权重” 之后的加权和. 该先验倾向于在加权意义下对场景尽可能简单地表达. 其他的 “统计性先验信息” 还包括物体分割中曲率的先验信息[46] 等.

上述的 “定性先验信息” 和 “统计性先验信息” 从势能团的大小上可分为局部区域先验和全局先验. 局部区域先验主要有共面性先验、区域一致性先验和 Pattern-based 先验及曲率先验等; 全局先验主要有 “连通性先验”、“限界框先验” 及各种关于标记的统计性先验等. 在场景理解中局部区域先验的能量项的阶次通常为几百至几千之间, 而全局先验的能量项的阶次更高, 通常为百万级甚至更大. 无论是区域先验或全局先验, 都远远超过了 Pairwise 能力模型的表达能力, 只有引入高阶能量项才能对其进行有效表达.

1.3 高阶能量模型在场景理解应用中的优势和潜力

在场景理解的应用中, 高阶能量模型的优势和潜力主要体现在如下几个方面:

1.3.1 高阶能量模型能够有效地描述场景(物体)的结构信息

高阶能量模型可以用来描述场景 (物体) 的结构性先验, 如共面性[25, 47]、对称性、结构相似性以及物体的连通性[26−29] 等定性先验信息. 并可引入物体的曲率先验[46] 以及利用统计学习方法得到的区域的 Pattern-based 先验[36−39]. 通过这些高阶能量项可以有效地引入关于场景 (物体) 的结构信息, 有助于对物体进行分割和形状完整化, 并纠正场景重建中出现的错误.

1.3.2 高阶能量模型能够有效地描述场景中物体与物体, 物体与环境之间的关系

高阶能量模型能够用来描述场景中物体之间的关系: 如用来描述场景中不同物体在同一场景共同出现概率的共生概率约束; 使场景尽可能 “简单” 地标记代价约束. 另外, 高阶能量项也可用来描述物体之间的关系, 如物体之间的支撑关系等[48].

1.3.3 高阶能量模型能够在同一能量优化框架下有效地融合多种场景理解手段

高阶能量优化能够对多种场景理解手段之间的 “互信息” 和 “一致性” 进行建模, 从而在同一能量优化框架下对其进行优化求解. 例如, 可利用立体匹配和物体分割之间的 “互信息”: 深度间断的地方可能是物体的边界, 反之亦然[49], 从而在能量优化的框架下同时对立体匹配和物体分割进行求解. 又如可利用高阶能量项对物体检测和语义分割 (Semantic segmentation) 之间的类别一致性进行建模[50−52], 从而同时得到场景中的物体检测和分割的结果, 并能区分同类别的多个相邻物体, 达到整体场景理解的要求.

1.3.4 高阶能量模型能够有效地约束场景的几何和语义信息在多视图之间的一致性

在三维场景理解中, 多视图所能提供的 “互补及冗余信息” 是三维场景理解相对于基于单幅图像的图像理解的主要优势之一. 对于同一物体或空间点, 在该物体 (或空间点) 可见的所有图像上以及图像之间同时考虑几何和语义之间的一致性, 可形成约束几何和语义在多视图之间一致性的高阶能量模型, 从而提高三维场景理解的可靠性. 近几年, 高阶能量模型的这个优势已经引起视觉界的重视, 如 Bleyer 等[28] 在二视图 (Stereo) 情况下的工作.

1.3.5 高阶能量模型能为三维点云和二维图像的融合提供有效的途径

由图像重建的三维点云不仅能够提供几何信息, 点云的分布及空间点的位置关系也能够提供关于物体类别的信息. 当点云稠密时, 仅依赖三维空间点便可对三维场景进行语义分割[53]. 通过对三维空间点及其在图像上对应的所有图像点利用高阶能量项进行建模, 可在同一能量优化框架下对三维和二维信息进行建模求解[54].

2 高阶能量模型

根据高阶能量项作用的区域大小可将高阶能量模型分为基于区域的高阶能量模型和全局高阶能量模型.

2.1 基于区域的高阶能量模型

由于通常意义下高阶能量项的阶次越高, 计算代价越大, 所以最低阶的高阶项 — 二阶平滑项便成了首个得到研究的高阶能量项. Woodford 等[25, 47] 通过构造二阶平滑项从而给出二阶曲率约束 (二阶导数为零, 共面性约束). 该约束能够使得场景中共面的深度信息得到保持, 在很多人造场景的立体匹配问题中取得了显著效果. 基于区域的高阶能量模型根据其约束的性质大致可分为如下两类.

2.1.1 约束区域标记一致性的高阶能量模型

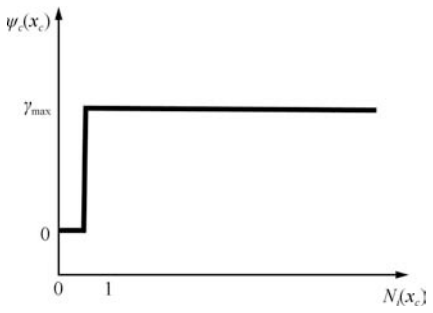
约束区域一致性的高阶能量项可看作是 Pairwise 能量项的 Potts 模型在高阶情况下的一个自然的拓展.

Potts 模型定义如下:

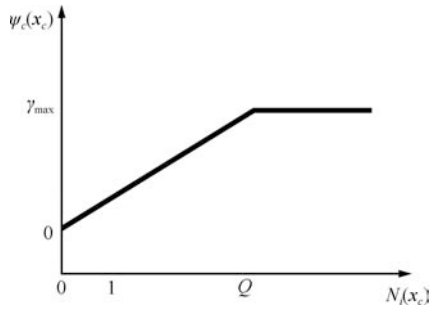
$$\psi_{i,j}(x_i, x_j) = \begin{cases} 0, & \text{若 } x_i = x_j \\ K_{i,j}, & \text{若 } x_i \neq x_j \end{cases} \tag{12}$$

其中, $K_{i,j}$ 是当 $x_i \neq x_j$ 时的惩罚, 其值仅依赖于像素 i, j 位置上的某些属性 (如灰度等). 该模型可看作是一种 "邻域一致性" 的约束, 即约束相邻的随机变量取值相同. 微软剑桥研究院的 Kohli 等[30−31] 提出了 $\mathcal{P}^n$ Potts 模型, 将 Potts 模型的一致性约束的范围由 "邻域" 扩展到 "区域", 如图 4 (a) 所示, 其公式表示如下:

$$\psi_c(\boldsymbol{x}_c) = \begin{cases} 0, & \text{若 } x_i = \ell \in \mathcal{L}, \quad \forall i \in c \\ \theta_1 |c|^{\theta_\alpha}, & \text{其他} \end{cases} \tag{13}$$



(a) $\mathcal{P}^n$ Potts 模型
(a) $\mathcal{P}^n$ Potts model



(b) Robust $\mathcal{P}^n$ 模型
(b) Robust $\mathcal{P}^n$ model

图 4 $\mathcal{P}^n$ Potts 和 Robust $\mathcal{P}^n$ 模型

Fig. 4 $\mathcal{P}^n$ Potts model and Robust $\mathcal{P}^n$ model

其中, θ_1 和 θ_α 是模型的参数, $\theta_1|c|^{\theta_\alpha}$ 是对区域 c 不满足类别一致性时的惩罚. 由图 4 (a) 可知, $\mathcal{P}^n$ Potts 模型仅能提供对区域类别一致性的 "严格" 约束, 只要区域 c 中有任一节点的取值不等于 ℓ, 即添加 $\gamma_{\max} = \theta_1|c|^{\theta_\alpha}$ 的惩罚. 而在实际问题中由过分割算法得到的分割区域往往是不准确的, $\mathcal{P}^n$ Potts 模型严格的类别一致性约束使其不能在这样的过分割区域上使用.

故 Kohli 等于 2008 年[32] 对 $\mathcal{P}^n$ Potts 模型作出如下两点改进: 1) 放宽 "严格" 的类别一致性限制, 允许区域 c 内一定数目 (不超过 Q 个) 的像素的类别取值不等于 ℓ; 2) 在该允许的区域内, 能量项 (惩罚) 的大小随着类别取值不一致的像素的个数的增长而线性增长. 即得到了如图 4 (b) 所示的 Robust $\mathcal{P}^n$ 模型. 该模型定义如下:

$$\psi_c(\boldsymbol{x}_c) = \begin{cases} N_i(\boldsymbol{x}_c)\frac{1}{Q}\gamma_{\max}, & \text{若 } N_i(\boldsymbol{x}_c) \leq Q \\ \gamma_{\max}, & \text{其他} \end{cases} \tag{14}$$

其中, $N_i(\boldsymbol{x}_c)$ 为区域 c 中类别取非 Dominant 类别的像素的个数, 即 $N_i(\boldsymbol{x}_c) = \min_{\ell \in \mathcal{L}}(|c| - n_\ell(\boldsymbol{x}_c))$, $n_\ell(\boldsymbol{x}_c)$ 为区域 c 中类别取值为 ℓ 的像素的个数. $\gamma_{\max} = |c|^{\theta_\alpha}(\theta_1 + \theta_2 G(c))$, $G(c)$ 代表超像素 c 的分割质量. Q 是截断参数. Robust $\mathcal{P}^n$ 模型能够有效地利用由非监督聚类得到的超像素 (区域) 的一致性先验, 即超像素中大部分像素的类别相同. 故该能量模型在很多视觉问题上获得了成功的应用. 如在图像语义分割中[33, 55−57], Robust $\mathcal{P}^n$ 模型能够在同一能量优化框架下有效地利用多种过分割算法所得到的过分割区域所提供类别一致性信息, 可显著提高语义分割的精度, 特别是在物体边界处的分割精度.

式 (14) 定义的 Robust $\mathcal{P}^n$ 模型仍然存在如下两点不足: 1) 其仅利用了过分割区域所提供的类别一致性先验, 而未考虑过分割区域自身的类别先验.

即假设任意类别 ℓ 为超像素 c 的 Dominant 类别的机会均等. 而相对像素而言, 基于超像素的特征 (随着尺度增大) 的类别区分能力更强; 2) 其假定区域 c 中所有位置上的权重相同. 而对超像素而言, 其边界往往不够准确, 故可降低位于超像素边界处的像素的权重. 针对这两点不足, Kohli 等[33] 于 2009 年进一步推广了 Robust $\mathcal{P}^n$ 模型, 得到了如下式所示的 Generalized form of Robust $\mathcal{P}^n$ 模型:

$$\psi_c(\boldsymbol{x}_c)=\min\left\{\min_{\ell\in\mathcal{L}}\left(\Big(P-f_\ell(\boldsymbol{x}_c)\Big)\theta_\ell+\gamma_\ell\right),\gamma_{\max}\right\} \tag{15}$$

其中, $P=\sum_{i\in c}p_i$, 其中 $p_i\geq 0$ 为像素 i 的权重. $f_\ell(\boldsymbol{x}_c)=\sum_{i\in c}p_i\delta_\ell(x_i)$, 其中 $\delta_\ell(x_i)$ 为断言函数, 仅当 $x_i=\ell$ 时, 其值为 1. $\theta_\ell=\frac{\gamma_{\max}-\gamma_\ell}{Q_\ell}$ 且满足 $Q_\ell+Q_{\ell'}<P$, $\forall\ell\neq\ell'\in\mathcal{L}$. 在 Generalized form of robust $\mathcal{P}^n$ 模型下, 像素 i 的权重为 p_i. 超像素的类别信息可由 γ_ℓ 表示, 超像素 c 的类别取 ℓ 的可能性越大, γ_ℓ 越小, 此时对 c 中的像素类别取值偏离 ℓ 的惩罚越大.

由 $Q_\ell+Q_{\ell'}<P$, $\forall\ell\neq\ell'\in\mathcal{L}$ 可知, 至多存在一个类别 ℓ 使得 $(P-f_\ell(\boldsymbol{x}_c))\,\theta_\ell+\gamma_\ell<\gamma_{\max}$. 若存在这样的 ℓ, 则称超像素 c 为同质 (Homogeneous), $\ell\in\mathcal{L}$ 为其 Dominant 类别; 若不存在这样的 ℓ, 则称超像素 c 为异质 (Heterogeneous). 可知, 这 $|\mathcal{L}|+1$ 种情况 ($|\mathcal{L}|$ 种同质加 1 种异质) 互斥, Ladický等[34−35] 通过引入一个具有 $|\mathcal{L}|+1$ 种取值的辅助变量 $x_c^{(1)}\in\mathcal{L}^e=\mathcal{L}\cup\{\mathcal{L}_F\}$ 表示区域 c 的所有 $|\mathcal{L}|+1$ 种可能的情况, 从而将 Generalized form of robust $\mathcal{P}^n$, 如式 (15) 所示, 等价转化为如下含有辅助变量的二阶能量项 $\psi_c^p(\boldsymbol{x}_c,x_c^{(1)})$:

$$\psi_c(\boldsymbol{x}_c)=\min_{x_c^{(1)}}\psi_c^p(\boldsymbol{x}_c,x_c^{(1)})=\min_{x_c^{(1)}}\left(\varphi_c(x_c^{(1)})+\sum_{i\in c}\varphi_{i,c}(x_i,x_c^{(1)})\right) \tag{16}$$

其中

$$\varphi_c(x_c^{(1)})=\begin{cases}\gamma_\ell, & \text{若 } x_c^{(1)}=\ell\in\mathcal{L}\\ \gamma_{\max}, & \text{若 } x_c^{(1)}=\mathcal{L}_F\end{cases} \tag{17}$$

$$\varphi_{i,c}(x_i,x_c^{(1)})=\begin{cases}0, & \text{若 } x_i=x_c^{(1)} \text{ 或 } x_c^{(1)}=\mathcal{L}_F\\ p_i\theta_\ell, & \text{若 } x_i=x_c^{(1)}\neq\mathcal{L}_F\end{cases} \tag{18}$$

对像素层 $\boldsymbol{x}$ 上的每个超像素 c 均可引入一个辅助变量 $x_c^{(1)}$ 代表该超像素, 所有这些辅助变量 $x_c^{(1)},c\in\mathcal{C}$ 构成了超像素层 $\boldsymbol{x}^{(1)}$. 在引入超像素层 $\boldsymbol{x}^{(1)}$ 后, 可使用超像素层上的一阶能量项 $\varphi_c(x_c^{(1)})$ 和像素层与超像素层的层间二阶能量项 $\varphi_{i,c}(x_i,x_c^{(1)})$ 等价表示 Generalized form of robust $\mathcal{P}^n$ 高阶能量项 $\psi_c(\boldsymbol{x})$.

在引入超像素层 $\boldsymbol{x}^{(1)}$ 后, 可在超像素层上定义 Pairwise 能量项, 鼓励相邻超像素类别取值相同. 并可继续在超像素层 $\boldsymbol{x}^{(1)}$ 上进行过分割, 得到 Super-segments, 并在其上再定义 Generalized form of robust $\mathcal{P}^n$ 能量项, 然后再添加 Super-segment 层 $\boldsymbol{x}^{(2)}$ 将定义在超像素层上的高阶能量项用 Super-segment 层的一阶项以及 Super-segment 层和超像素层的层间二阶能量项等价表示. 重复该过程, 即得到了如下递归定义的关联层级网络模型 (Associative hierarchical network)[34−35]:

$$\psi_{\mathcal{H}}=\sum_{i\in\mathcal{V}}\varphi_i^{(0)}(x_i^{(0)})+\sum_{(i,j)\in\mathcal{E}}\varphi_{i,j}^{(0)}(x_i^{(0)},x_j^{(0)})+\min_{\boldsymbol{x}^{(1)}}\psi_{\mathcal{H}}^{(1)}(\boldsymbol{x}^{(0)},\boldsymbol{x}^{(1)}) \tag{19}$$

$\psi_{\mathcal{H}}^{(1)}$ 递归定义如下:

$$\psi_{\mathcal{H}}^{(n)}(\boldsymbol{x}^{(n-1)},\boldsymbol{x}^{(n)})=\sum_{c\in\mathcal{C}^{(n-1)}}\psi_c^p(\boldsymbol{x}_c^{(n-1)},x_c^{(n)})+\sum_{c,d\in\mathcal{C}^{(n-1)}}\varphi_{c,d}^{(n)}(x_c^{(n)},x_d^{(n)})+\min_{\boldsymbol{x}^{(n+1)}}\psi_{\mathcal{H}}^{(n+1)}(\boldsymbol{x}^{(n)},\boldsymbol{x}^{(n+1)}) \tag{20}$$

其中, $\boldsymbol{x}^{(0)}=\boldsymbol{x}$ 为底层像素层, $\mathcal{C}^{(0)}=\mathcal{C}$ 为定义在底层像素层上超像素的集合. $\psi_c^p(\boldsymbol{x}_c^{(n-1)},x_c^{(n)})$ 的定义类似于式 (16). 图 5 给出了像素–区域–场景的多尺度层级网络框架, 其自 Pixel layer 至 Super-segment layer 的部分为关联层级网络模型.

关联层次网络模型是 Generalized form of Robust $\mathcal{P}^n$ 模型的推广, 当关联层次网络模型仅包含像素和超像素两层, 且超像素层上无 Pairwise 能量项时, 两个模型等价. 相比于 Generalized form of robust $\mathcal{P}^n$ 模型, 关联层级网络模型在如下两个方面进行了改进: 1) 其辅助节点层上的 Pairwise 项可直接对区域间的关系进行建模, 而 Generalized form of robust $\mathcal{P}^n$ 模型仅限制为区域内; 2) 关联层级网络模型递归定义, 能够在同一能量模型下融合不同尺度上所提取的特征信息、区域的一致性信息并约束类别在不同尺度上的一致性. 而 Generalized form of robust $\mathcal{P}^n$ 模型仅限于融合像素和超像素两个尺度上的信息. 值得注意的是, 关联层级网络模型和 Generalized form of robust $\mathcal{P}^n$ 模型均存在非常高效且高质量的优化算法[58], 允许该高阶能量项的阶次为数千甚至更高, 从而使得这类模型成为目前

在视觉问题中应用最为广泛的高阶能量模型.

Kohli 等[59] 进一步探索了具有高效能量优化算法的高阶能量项所应具有的一般形式. 对于可表示为如下形式的高阶能量项:

$$\psi_c(\boldsymbol{x}_c) = \otimes_{q\in\mathcal{Q}} f^q(\boldsymbol{x}_c) \tag{21}$$

其中, $\otimes \in \{\min, \max\}$, $f^q(\boldsymbol{x}_c)$ 为如下形式的线性函数:

$$f^q(\boldsymbol{x}_c) = \theta_q + k_q \sum_{i\in c}\sum_{a\in\mathcal{L}} p_{ia}\delta_i(a) \tag{22}$$

其中, $\delta_i(a)$ 为断言函数, 仅当 $X_i = a$ 时其值为 1. p_{ia} 为权重系数. θ_q 和 k_q 是线性函数的参数. 当 $\otimes = \min$ 时, 式 (21) 为线性函数的下包络, 如图 6 (a) 所示, 此时 $\psi_c(\boldsymbol{x}_c)$ 为凹函数; 当 $\otimes = \max$ 时, 式 (21) 为线性函数的上包络, 如图 6 (b) 所示, 此时 $\psi_c(\boldsymbol{x}_c)$ 为凸函数. Kohli 等[59] 指出, 若一个高阶能量项能够用线性函数的下包络表示或近似 ($\otimes = \min$), 则该高阶能量项存在高效且高质量的近似求解算法. Gould[60] 进一步指出, 若限制 $x_i \in \{0, 1\}$, 此时线性函数为伪布尔线性函数 (Pseudo-boolean linear functions), 则其下包络存在高效的精确求解算法. 许多在视觉问题中使用的高阶能量函数 (包括 Robust $\mathcal{P}^n$ 模型等) 都可表示为线性函数下包络的形式. Kohli 等[59] 同时也指出, 若一个高阶能量项仅能够用线性函数的上包络表示或近似, 则对该高阶能量函数的优化是困难的. 探索对具有线性函数上包络表示形式的高阶能量项的高效优化算法目前仍然是一个开放的研究问题.

2.1.2 Pattern-based 高阶能量模型

约束区域一致性的高阶能量模型尽管优化效率很高, 但对于区域先验的表达能力有限. 微软剑桥研究院的 Rother 等[36] 和希腊克里特大学的 Komodakis 等[37] 在 2009 年的 CVPR 上分别独立地提出了基于区域的 Pattern-based 高阶能量模型. 该模型源于如下的直觉: 对于一个区域而言 (如自然图像的一个图像片), 尽管其所有可能的取值个数非常大 (指数规模), 但其实际有可能出现的 (取值有意义) 的取值可能较为稀疏. 图 7 给出了 Pattern-based 高阶能量项的一个例子. 由图 7 可以看出, 随机变量 $\boldsymbol{x}_c$ 仅在其取值空间 $\mathcal{X}_c$ 中的有限个标记上具有较小的能量. 记这些标记的集合为 $\mathcal{X}^0 = \{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_t\}$, 其对应的能量项 $\psi_c(\boldsymbol{x}_c)$ 的值为 $\theta^0 = \{\theta_1, \theta_2, \cdots, \theta_t\}$. 当 $\boldsymbol{x}$ 取其他标记时, 能量项 $\psi_c(\boldsymbol{x}_c)$ 的值均为 $\theta_{\max}$. 则 Pattern-based 高阶能量可写为如下的形式:

$$\psi_c(\boldsymbol{x}_c) = \begin{cases} \theta_q, & \text{若 } \boldsymbol{x}_c = \mathcal{X}_q \in \mathcal{X}^0 \\ \theta_{\max}, & \text{其他} \end{cases} \tag{23}$$

这些在实际中可能出现的标记 (取值) $\mathcal{X}^0$ 称为这个区域的 Pattern. Robust $\mathcal{P}^n$ 模型可看作 Pattern-based 高阶能量模型当其 Pattern 为约束区域内元素类别取值一致时的特例.

值得注意的是, Harmony potentials[38−39] 也可看作是一种 Pattern-based 高阶能量模型. 在像素–超像素–区域–场景的多尺度层级网络模型的框架下, 如图 5 所示, 随着尺度的增大, 有如下两个结论:

1) 区域的异质性越发显著. 在较小的尺度上 (像素、超像素), 往往仅包含单一类别, 较大的区域可能包含具有不同 Dominant 类别的超像素, 而整个场景往往包含多个不同类别的物体. 故随着尺度的增大区域的异质性越发显著.

2) 分类的性能提高, 但分割的精度往往下降. 对分类问题而言, 在较小尺度上的局部分类器的性能往往较弱, 因为其仅依赖于局部的小区域的特征, 难以消除类别的二义性. 而在大的尺度下尽管具有较好的类别区分能力, 但往往难以得到较精确的分割边界.

而由前一节可知, 在关联层级网络模型中, 当区域 c 为异质时, 即 $x_c^{(n)} = \mathcal{L}_F$ 时, 关联层级网络模型对区域 c 中包含的像素的类别取值失去约束. 即关联层级网络模型仅将异质类型 $\mathcal{L}_F$ 看作是一种 "自由" 类别. 这使得关联层级网络模型难以有效地融合较大尺度以及全局尺度下的类别信息. Boix 等[38] 和 Gonfaus 等[39] 认为, 当区域 (或图像) c 为异质, 包含多个类别时, 某些类别组合更有可能出现 (类似于音乐中的和声, Harmony 的由来), 应该有效利用这些类别之间的共生信息. 即在这种情况下, $x_c^{(n)}$ 应从 $\mathcal{L}$ 的幂集合 (Power set)$\mathcal{P}(\mathcal{L})$ 中取那些最有可能出现的类别的组合. Boix 等[38] 和 Gonfaus 等[39] 提出了一种对 $\mathcal{P}(\mathcal{L})$ 中所有元素进行 Ranked subsampling 的方法仅取出区域 c (或图像) 上最有可能出现的一些类别的组合, 将其作为 $x_c^{(n)}$ 取值的候选集合, 记为 $\mathcal{S}$. 此时, 对 $x_c^{(n)} \in \mathcal{S}$ 的任意取值, Harmony potentials 仅惩罚其类别未在 $x_c^{(n)}$ 中出现的像素, 即

$$\varphi_{i,c}^G\left(x_i, x_c^{(n)}\right) = \gamma_i(x_i) T[x_i \notin x_c^{(n)}] \tag{24}$$

其中, $T[x_i \notin x_c^{(n)}]$ 仅当 x_i 的类别未在 $x_c^{(n)}$ 中出现时, 其值为 1. 在图 8 中, $x_c^{(n)} = \{\text{Blue}, \text{Green}\}$, 故仅对其底层两个类别为 Red 的节点施加惩罚. 对比关联层级网络模型的层间能量的定义, 式 (17) 所示, 可知, Harmony potentials 的层间能量将底层节点对上层节点的某一特定类别的一致性 (Consistency) 推广到底层节点对上层节点的某一特定类别组合 (集合) 的一致性. Harmony potentials 定义如

下:

$$\psi_{\mathcal{HP}}=\min_{x_c^{(n)}\in\mathcal{S}}\left(\varphi^G(x_c^{(n)})+\sum_{i\in c}\varphi_{i,c}^G\left(x_i,x_c^{(n)}\right)\right) \quad (25)$$

其一阶项 $\varphi^G(x_c^{(n)})$ 由 Ranking 信息得到, 层间能量 $\varphi_{i,c}^G\left(x_i,x_c^{(n)}\right)$ 由式 (24) 定义.

Harmony potentials 可看作是对关联层级网络模型的推广, 若取 $x_c^{(n)}=\mathcal{L}$ 则等价于关联层级网络中的 "自由" 类别 $\mathcal{L}_F$. 在 Harmony potentials 中, $x_c^{(n)}$ 从幂集合 $\mathcal{P}(\mathcal{L})$ 中取值, 并不仅限于单一类别. 这使得 Harmony potentials 能够有效地融合较大尺度以及全局的类别信息. 而关联层级网络模型更适合融合较小尺度上的类别及一致性信息. 两者结合便可构造像素–区域–场景的多尺度层级网络, 如图 5 所示. 值得注意的是, 从高阶能量项的作用区域来看, Harmony potentials 既可作用于区域上, 也可作用于全局范围, 可看作是从区域到全局过渡的一种高阶能量项.

Pattern-based 高阶能量模型的出现极大地增强了基于区域的高阶能量项的模型表达能力. 原则上, 任何区域先验 (模型) 都可用 Pattern-based 高阶能量项 (包括 Harmony potentials) 表示. 但相比于约束区域一致性的高阶能量项而言, Pattern-based 高阶能量项的能量优化算法还很不成熟, 仅对少数 (足够稀疏的) Pattern-based 高阶能量项具有高质量的优化算法. 这制约了 Pattern-based 高阶能量项在实际视觉问题中的应用.

2.2　全局高阶能量模型

在场景理解中, 基于区域的高阶能量项能够有效地融合场景的局部信息, 对场景理解而言, 是一种自底向上的方式. 而基于全局的高阶能量项能够有效地表达场景的全局性先验, 可看作是一种自顶向下的方式. 现有文献中关于全局高阶能量模型主要有如下两类: 1) 对物体连通性进行约束的高阶能量项; 2) 对场景标记的统计特性进行约束的高阶能量项.

2.2.1　对物体连通性进行约束的高阶能量模型

在图像分割中, 一般认为物体 (前景) 是连通的, 但传统的低阶能量函数无法表述全局性的连通约束, 从而导致物体 (前景) 的细长区域在分割中被 "隔断". 包含连通性约束的高阶能量函数可表示如下:

$$E(\boldsymbol{x})=\sum_{i\in\mathcal{V}}\psi_i(x_i)+\sum_{(i,j)\in\mathcal{E}}\psi_{i,j}(x_i,x_j)$$
$$\text{s.t. } \boldsymbol{x} \text{ being connected} \quad (26)$$

伦敦大学学院的 Vicente 等[26] 指出, 在上述连通性约束下 (记该连通性约束为 C0), 该能量函数, 式 (26) 所示的最小化问题很难求解, 甚至当其仅包含一阶项时, 其在上述连通性约束下的能量函数最小化问题为 NP-hard 问题. 为了能够近似求解该能量优化问题, Vicente 等[26] 对连通性约束进行了松弛, 由约束前景的所有像素组成单一的连通区域松弛为仅约束前景的任意两个点之间的连通性 (记松弛后的连通性约束为 C1). 并针对该松弛的连通性约束提出了一种结合了最短路径算法 Dijkstra 和 Graph cuts 的启发式算法 DijkstraGC. 对于一些实际的二值分割问题, DijkstraGC 算法甚至能求得其全局最优解. 对于连通性约束 C0, 德国马普研究院的 Nowozin 等[27] 在基于利用线性规划松弛近似求解马尔科夫随机场的最大后验概率估计问题的框架下, 利用约束生成技术得到了一种近似求解方法并通过实验验证了该方法在提升物体分割质量上的效果. 在 Nowozin 等[27] 工作的基础上, 微软剑桥研究院的 Lempitsky 等[29] 提出了一种称之为 "限界框先验" 的高阶能量模型作为二维图像连通性先验的推广: 物体在限界框内是连通的且相对于限界框是 "紧" 的. 该高阶能量模型的优化使用了与 Nowozin 等[27] 类似的约束生成技术.

关于空间三维物体的连通性, 维也纳技术大学的 Bleyer 等[28] 在立体匹配问题中给出了一种约束空间三维物体连通性的高阶能量项. 但遗憾的是, Bleyer 等[28] 未能找到该高阶能量项的有效优化算法, 故仅能在每次 Fusion move[61−62] 之后作为验证手段使用.

2.2.2　对场景标记的统计特性进行约束的高阶能量模型

场景标记的统计特性包含三类: 1) 标记出现的数目; 2) 场景中所有出现的标记类别的代价; 3) 场景中出现的标记类别之间的关系.

捷克技术大学的 Werner[42, 63] 引入了一种约束标记出现数目的高阶能量项, 并提出了一种 n-ary maxsum diffusion 算法用以优化该高阶能量. 但该优化算法仅对某些标记数目有效, 不能接受用户的任意输入. Kohli 等[59] 探讨了两类关于标记数目的约束: 1) 非空集合约束, 即集合中至少有一个元素取某一特定标记. 如在多视图重建中, 各视图侧影轮廓上的每一点的反投影线上至少包含一个重建物体的空间点. 2) 规模先验, 即关于某一类别标记的节点数目的先验. Kohli 等[59] 将该高阶能量项表示为线性函数的上包络并设计了一种新的基于消息传递的算法来求解该 Min-max 松弛问题. 但是该算法的求解效率和质量不是很理想.

加拿大西安大略大学的 Delong 等[44−45] 提出了一种表述标记类别代价的高阶能量项. 对每一个

类别标记赋予一定的权重, 该高阶能量项可表示为在场景中出现的类别标记代价的加权和. Delong 等[44−45] 同时提出了一种扩展的 α-expansion 算法用以优化包含该高阶能量项的能量函数.

Ladický等[40−41] 提出的高阶能量项可看作是 Delong 等[44−45] 的高阶能量项的一种推广: 由出现类别标记的加权和的形式推广到出现标记类别的任意单调增函数的形式. 增强了该能量模型的表达能力, 使之不仅能够表达关于标记代价的约束, 而且能够表达关于标记之间的统计特性, 如共生概率等约束. 该高阶能量模型在场景理解中可作为场景模型的 "正则项", 约束场景模型的复杂度并可同时作为场景模型的 "全局性先验".

苏黎世联邦工业大学的 Boix 等和 Gonfans 等[38−39] 提出了一种称之为 Harmony potentials 的高阶能量项. 该高阶能量项可对单一节点同时赋予多个标记, 当作用于区域上时, 该高阶项可看作是一种 Pattern-based 先验, 作为 Robust $\mathcal{P}^n$ 模型的推广; 当作用于全局时, 该高阶项可用来表达如共生概率等场景中出现的标记类别关系的全局先验. 这种可对单一节点同时赋予多个标记的高阶能量项相对于传统的仅能对单一节点赋予一个标记的高阶能量项在模型表达能力上具有很大的优势, 但求解的复杂度也随之大大升高. 所以, 一般只能限定在若干特定的标记组合下使用.

3 高阶能量优化算法

一般来说, 高阶能量模型的阶次越高, 模型的表达能力越强, 但对应的能量优化问题也越困难. 所以, 从某种程度上说, 不考虑对应优化问题的模型设计是没有意义的. 当前文献中关于高阶能量优化算法可大体分为三类: 1) 基于组合优化的高阶能量优化; 2) 基于消息传递的高阶能量优化; 3) 基于松弛及分解的高阶能量优化.

3.1 基于组合优化的高阶能量优化

基于组合优化的高阶能量优化由传统的 Graph cuts 算法[4−5] 发展而来. 这类算法的技术路线一般为: 将多标集的高阶能量项通过 Move-making 转化为二标集的高阶能量项, 然后添加辅助变量进行降阶, 最后对降阶后的二阶二标集能量函数用基于网络流的方法求解. 关于这三个基本步骤的研究在最近几年都取得了一些突破. 对于二阶二标集能量函数优化, 经典的 Graph cuts 算法[5] 仅当其满足 Submodular 条件时才能使用. 这大大限制了 Graph cuts 算法的适用范围. Kolmogorov 等[64] 于 2007 年首次将在组合优化界得到广泛研究的 QPBO (Quadratic pseudo-boolean optimization)[65−66] 引入到视觉界. QPBO 可看作是 Graph cuts 的推广: 当能量函数仅包含 Submodular 项时, QPBO 仅需计算一次 st-mincut 便可对其精确求解 (与传统 Graph cuts 相同); 若能量函数包含 non-submodular 项, 则通过构造与其对应的具有特殊结构 (含约束) 的图, 然后对该图的约束进行松弛, 求其 st-mincut, 从而可得到该能量函数部分节点的标记 (Partial assignment). 同年, Rother 等[67] 给出了一种高效的 QPBO 实现算法, 大大提高了 QPBO 的计算效率, 使之能够适用于很多视觉问题的处理. 同时提出了 QPBO 的一种扩展算法 QPBOI (Quadratic pseudo-boolean optimization improve), QPBOI 是一种近似算法, 可以从任意标记开始, 利用 QPBO 进行迭代, 并保证每次迭代能量值不增. 这些进展大大提高了基于组合优化方式的二阶二标集问题的求解能力, 从仅能求解 Submodular 能量项扩展到可求解任意的二阶二标集问题. 但值得注意的是, 对任意的二阶二标集问题, 其求解质量仍然依赖于问题本身. 当含有较多的 Non-submodular 项, 且其系数较大时, 被标记的节点的数目可能会非常少, 甚至为零, 从而导致求解完全失败.

二阶二标集问题求解上取得的突破激励学术界寻找更好的将高阶二标集问题降阶的方法. 尽管通过添加辅助变量进行降阶的方法在学术界早已为人所熟知[68], 但其所采用的 "替换 (Substitution)" 降阶方法, 对于每一个高阶项, 在很多情况下会添加大量 (阶次的指数倍) 的辅助变量并产生大量的、具有很大系数的 Non-submodular 项, 从而导致该方法在实际问题中不可行[69−70]. Kolmogorov 等[5] 针对满足 Submodular 条件的三阶二标集能量函数提出了一种将其降阶为二阶二标集 Submodular 项的方法. Freedman 等[71] 从代数的角度重新审视并推广了该方法并给出了任意阶次二标集能量函数可降阶为二阶二标集 submodular 项的充分条件. 针对一般形式的高阶二标集问题, 日本早稻田大学的 Ishikawa[69−70] 提出了一种通用的降阶方法 (Minimum-selection), 与传统的替换降阶相比, 仅需添加较少的辅助变量便可降阶. 柯达公司的 Gallagher 等[72] 指出, 应该在各种降阶方法中选择更好的组合策略, 从而降低最终优化的难度. 康奈尔大学的 Fix 等[73] 提出了一种同时考虑含有相同随机变量的多个高阶项进行降阶的方法, 使其降阶后含有的 Non-submodular 项较少. 以上几种方法尽管可对一般形式的高阶二标集问题进行降阶, 但其降阶后常含有 Non-submodular 项, 且在最坏的情况下仍需添加指数倍的辅助变量, 限制了这些方法在更为高阶的能量项上的使用. 针对具有较高阶次的区

域高阶能量项, Kohli 等[30−31] 针对 $\mathcal{P}^n$ potts 模型在 α-expansion 和 $\alpha-\beta$ swap 下的 Move energy 提出了一种仅需添加很少辅助变量且降阶后仅含 Submodular 项的降阶方法, 并将此降阶方法推广到 Robust $\mathcal{P}^n$ 模型[32−33] 的 Move energy 的求解上. 这些降阶方法是 $\mathcal{P}^n$ potts 模型和 Robust $\mathcal{P}^n$ 模型具有高效且高质量优化算法的关键所在. 探求更好的针对高阶二标集能量项的降阶策略是近年来的一个热点研究问题, 也是在基于组合优化的高阶能量优化研究中今后需进一步深入研究的一个方向.

将多标集问题转换为二标集的 Move-making 算法中最具代表性的算法是 α-expansion, $\alpha-\beta$ swap[4]. 这些算法在视觉问题, 特别是低阶问题的优化中取得了很大的成功. Lempitsky 等[61−62] 提出了一种称之为 Fusion move 的新的 Move-making 方法作为 α - expansion, $\alpha-\beta$ swap 的推广. 该方法每次 Move 的空间包含了当前状态和所建议状态之间的所有状态. Ishikawa[69−70] 通过实验对比发现, 对于高阶多标集问题, 采用 Fusion move, 优化效果远好于传统的 α-expansion, $\alpha-\beta$ swap. 值得注意的是, Fusion move 的效果依赖于每次迭代时提出的建议状态的质量, 在最坏的情况下其优化效果与 α-expansion 相同.

3.2　基于消息传递的高阶能量优化

基于消息传递的高阶能量优化由传统的 BP (Belief propagation) 算法[74] 发展而来. 传统的 BP 算法可看作是链图上的动态规划算法在树形图上的扩展. 当图不含回路时 (树), 可精确计算任意变量的边缘概率分布从而得到 MMSE (Minimum mean-squared error) 或最大后验概率估计 (利用 Sum-product 或 Max-product 算法). 当用在含回路的图上时 (不保证收敛性), 就是在视觉问题中经常用到的 LBP 算法[6−7]. LBP 算法本身并未限制能量函数必须为低阶, 但其计算复杂度通常为图中最大团 (能量函数的阶次) 的指数次方, 这使得它在实际的高阶能量优化问题中基本无法使用. 康奈尔大学的 Lan 等[75] 使用了一种自适应状态空间来控制高阶能量项的消息计算和传递复杂度的增长, 从而使得 LBP 算法用来求解高阶能量函数成为可能. 卡内基梅隆大学的 Potetz 等[76−77] 针对当高阶项可表示为其团所包含的变量 (节点) 的线性组合的这一类特殊的高阶能量优化问题, 使用变量替换将高阶项降阶, 并通过自适应直方图约束变量的搜索空间, 提高了 LBP 的计算效率, 使得算法的复杂度由原来随着势能团规模增大呈指数增长降为线性增长. 多伦多大学的 Tarlow 等[78] 构造了两类新的高阶能量函数, 通过精心设计的消息更新策略, 得到了更加高效的 LBP 求解算法. 针对可使用一些仅定义在其子团 (Sub-clique) 上的势能团的和的形式等价表示的高阶能量项, Mcauley 等[79] 提出了一种基于消息传递的加速计算的策略. Felzenszwalb 等[80] 进一步提升了该算法的计算效率. 但总的来说, 这些基于消息传递的高阶能量优化算法目前的研究都还仅限于具有一些特殊结构的高阶能量项, 针对具有更高阶次和更为一般形式的基于消息传递的高阶能量优化算法的研究还很少.

鉴于很多具体视觉应用中的能量函数的高阶项具有所谓的 “稀疏性”, 即在其高阶项的指数次规模的取值空间中, 仅有较少的取值具有低的能量值, 因此充分利用高阶项所具有的 “稀疏性”, 寻求更为有效的 (近似) 表示形式, 再进行降阶求解也是上述两类高阶能量优化算法 (基于组合优化或消息传递) 经常采用的技术思路. 如在基于区域的 Pattern-based 高阶能量模型中, Rother 等[36] 提出了一种更为 “稀疏” 的表示: 用一组标记偏离代价函数作为 Soft-pattern, 对具有 “稀疏性” 的 Pattern-based 高阶能量项进行近似表示. 该近似表示具有更少的参数, 仅需添加较少的辅助变量即可将 Pattern-based 高阶能量项转化为等价的低阶能量项, 从而可使用 BP, TRW 或 Graph cuts 进行优化.

3.3　基于松弛及对偶分解的高阶能量优化

任何马尔科夫框架下的最大后验概率估计问题都可以转化为等价的整数规划问题, 从而使得线性规划松弛及各种凸松弛、拉格朗日松弛及对偶分解[81−83] 等方法都可以用在高阶能量优化中. 该类方法一般先对马尔科夫框架下的最大后验概率估计对应的整数规划问题进行凸松弛 (或线性规划松弛等); 然后, 针对其松弛问题或其松弛问题的对偶问题设计高效的求解方法或者利用对偶分解的方法将其分解为一系列的较易求解的子问题; 最后, 通过合并各子问题的解从而得到原问题的解.

为了求解松弛后的线性规划问题, Nowozin 等[27] 提出了一种利用约束生成技术以减少每次迭代求解线性规划问题时约束的数目的近似求解算法, 并将该算法用于求解对图像连通性约束的高阶能量项. Werner[42, 63] 使用一种称为 n-ary min-sum diffusion 的方法对高阶能量项进行线性规划松弛求解, 该方法可看作是对低阶能量项进行线性规划松弛求解的 Min-sum diffusion 算法[84−85] 在高阶情况下的推广. Komodakis 等[86−87] 提出了一种利用对偶分解进行高阶能量函数优化的框架: 将一个经过松弛的马尔科夫随机场的最大后验概率估计问题的对偶问题分解为一系列相对简单的子问题, 然后通过协调各个子问题从而得到原问题的近似

解. Komodakis 等[37] 在该框架下提出了一种求解 Pattern-based 高阶能量模型的优化方法. 在对图像上物体连通性进行约束的全局高阶能量模型中, Vicente 等[26] 通过将能量模型分解为三个子问题分别求解, 最终得到原能量函数的一个下界及原问题的近似解.

基于松弛及分解的高阶能量优化有如下三个主要优点: 1) 松弛后的问题可使用凸优化中较为成熟的算法求解, 如 Projected subgradient, 各种约束生成技术等; 2) 通过将问题分解进而协调各子问题的解来对原问题求解的框架适合于分布式计算及 CPU-GPU 异构集群的计算架构; 3) 利用对偶间隙 (Duality gap) 可提供对原问题松弛解的质量的定量度量, 从而间接得到对原高阶能量优化的求解质量的度量.

基于松弛及对偶分解的高阶能量优化方法除了需要更加高效的优化算法之外, 也存在如下两个主要困难: 1) 松弛问题的解和原问题的解之间存在何种关系? 2) 如何改进 Relaxation tightness? 针对第一个问题, Swoboda 等[88] 提出了一种通用的可从松弛问题的解恢复部分原问题最优解的方法. 该方法提出了一个可用来测试部分最优性的解的准则, 并从松弛问题的所有整数解开始利用该准则逐步剔除不满足部分最优性的解, 从而尽可能多地恢复原问题的最优解. 该论文获得了 CVPR 2014 年的 Best Student Paper. 针对第二个问题, 学术界已经对该问题进行了一些研究 (如文献 [89−92]). 但总的来说, 这两个方面的研究还比较初步, 需要进一步深入研究才能得到更具有优化质量保证的高阶能量优化算法.

3.4 高阶能量优化算法的竞赛和评测平台

在视觉领域的多个重要的研究方向上, 竞赛和评测平台在激励该方向上产生更好的研究成果和推动其应用方面扮演着重要且不可替代的作用, 能量函数优化领域也不例外. 如 Szeliski 等[13] 在统一的软件架构下集成了多种经典的低阶能量优化算法, 并在平面 4 邻域的马尔科夫随机场模型下对各经典算法进行了对比研究. 该工作对在实际的视觉问题中选取何种能量优化算法具有重要的指导意义, 在视觉界产生了重要的影响. 受该工作启发, Andres 等[93] 在 2010 年对一些基于 Move-making 和消息传递的高阶能量优化算法进行了仿真实验对比研究. 为了对最近几年提出的高阶能量优化算法进行更加深入的对比研究, Andres 等于 2013 年实现了一个包含 24 种能量优化算法以及来源于 20 个不同的视觉问题中的 2 300 个能量模型的评测平台[94] 并公布了其开源软件实现 OpenGM2[95]. 该评测平台具有灵活的软件和数据接口以及较好的扩展性, 方便集成新的能量优化算法并对各优化算法在同一模型下进行对比测试. 但受限于目前高阶能量优化算法的发展现状, 具有较广泛适用性、能在同一能量模型下进行对比测试的高阶能量优化算法数目较少, 该评测平台中包含的高阶能量优化算法的数目还十分有限 (24 种能量优化算法中仅有一部分可用于高阶能量项), 而且高阶能量模型的阶次也相对较低 (最高不超过 300 阶).

机器学习界也组织了一些概率图模型的统计推断问题的竞赛, 如 UAI Approximate Inference Challenge[1], Probabilistic Inference Challenge[2] 等. 与视觉问题主要关注最大后验概率估计问题不同, 这些竞赛关注如下三个统计推断问题: 1) 归一化因子估计; 2) 最大后验概率估计; 3) 单变量或团的边缘概率分布估计. 另外, 值得注意的是, 这些竞赛的最大后验概率估计算法主要针对一般形式的能量优化问题所设计. 目前而言, 在具体视觉问题的应用上, 这些算法的求解效率远低于前述针对视觉问题所设计的高阶能量优化算法[94].

4 高阶能量模型的参数学习

高阶能量模型的参数学习 (估计) 是指在给出 K 个训练样本 $\{\boldsymbol{D}^k, \boldsymbol{x}_G^k\}_{k=1}^K$ 的情况下, 其中 $\boldsymbol{D}^k$ 是第 k 个样本的观测数据, $\boldsymbol{x}_G^k$ 是与第 k 个样本对应的变量的真值, 选择高阶能量模型 $E(\boldsymbol{x}|\boldsymbol{D}, \boldsymbol{w})$ 的最优模型参数 $\boldsymbol{w}^*$ 的过程. 在高阶能量模型的参数学习问题中, 一般均假设高阶能量模型的每一个能量项 $\psi_c(\boldsymbol{x}_c|\boldsymbol{D}, \boldsymbol{w})$ 均可写为关于模型参数 $\boldsymbol{w}$ 的如下形式: $\psi_c(\boldsymbol{x}_c|\boldsymbol{D}, \boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}}\phi_c(\boldsymbol{x}_c|\boldsymbol{D})$, 其中 $\phi_c(\boldsymbol{x}_c|\boldsymbol{D})$ 为关于变量 $\boldsymbol{x}$ 和观测 $\boldsymbol{D}$ 的特征向量. 目前文献中关于参数学习主要有两类方法: 1) 生成式的参数学习方法, 如极大似然估计; 2) 判别式的参数学习方法, 如 Max-margin. 下面分别讨论这两种方法.

4.1 生成式的参数学习方法

生成式的参数学习方法一般为求取训练数据的极大似然估计, 即

$$\overset{*}{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} \prod_{k=1}^{K} p(\boldsymbol{X} = \boldsymbol{x}_G^k|\boldsymbol{D}^k, \boldsymbol{w}) \tag{27}$$

其中, $p(\boldsymbol{X} = \boldsymbol{x}_G^k|\boldsymbol{D}, \boldsymbol{w}) \propto \exp\left(-E\left(\boldsymbol{x}_G^k|\boldsymbol{D}^k, \boldsymbol{w}\right)\right)$ 如式 (4) 所示. 尽管目标函数 (27) 为凸函数且可导, 但计算该目标函数的梯度需要求特征向量 $\phi_c(\boldsymbol{x}_c|\boldsymbol{D})$

[1] http://www.cs.huji.ac.il/project/UAI10
[2] http://www.cs.huji.ac.il/project/PASCAL/index.php

在 $p(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{D},\boldsymbol{w})$ 分布下的期望. 而这在通常情况下是无法求解的. 故在实际中, 经常使用一些近似的方法, 如 Sum-product belief propagation[74] 等, 对马尔科夫随机场的边缘分布进行估计, 从而计算其梯度. Scharstein 等[96] 使用该参数学习的方法训练用于立体匹配的 CRF 能量模型.

4.2 判别式的参数学习方法

判别式的参数学习方法一般采用 Max-margin 的方式[97]. 即寻求参数 $\overset{*}{\boldsymbol{w}}$, 使得在该参数下, 能量函数 $E(\boldsymbol{x}_G^k|\boldsymbol{D},\overset{*}{\boldsymbol{w}})$ 在训练集 $\{\boldsymbol{D}^k,\boldsymbol{x}_G^k\}_{k=1}^K$ 上达到最小值. 该条件可表示为

$$E(\boldsymbol{x}_G^k|\boldsymbol{D}^k,\overset{*}{\boldsymbol{w}})<E(\boldsymbol{x}|\boldsymbol{D}^k,\overset{*}{\boldsymbol{w}}),\quad \forall\boldsymbol{x}\neq\boldsymbol{x}_G^k \tag{28}$$

在实际中, 条件 (28) 难以在全部训练数据上得到满足, 故需增加一个非负的松弛变量 ϵ_k. 另外, 还需增加一个距离函数 $\Delta(\boldsymbol{x},\boldsymbol{x}_G^k)$ 用以衡量 $\boldsymbol{x}$ 与真值 $\boldsymbol{x}_G^k$ 之间的 margin. 此时条件 (28) 变为如下形式:

$$E(\boldsymbol{x}_G^k|\boldsymbol{D}^k,\overset{*}{\boldsymbol{w}})<E(\boldsymbol{x}|\boldsymbol{D}^k,\overset{*}{\boldsymbol{w}})-\Delta(\boldsymbol{x},\boldsymbol{x}_G^k)+\epsilon_k \tag{29}$$

理想情况下, 模型参数 $\overset{*}{\boldsymbol{w}}$ 应使得松弛变量 ϵ_k 尽可能小. 基于 Max-margin 的参数学习问题可表示为如下的带约束优化问题:

$$\min_{\boldsymbol{w},\{\epsilon_k\}}\|\boldsymbol{w}\|^2+C\sum_{k=1}^{K}\epsilon_k\quad \text{s.t. (29)} \tag{30}$$

式中 C 是超参数. 式 (30) 为含有指数规模线性不等式约束的二次规划问题. 文献 [98−99] 使用割平面法求解该二次优化问题: 即从无约束优化问题求解开始, 每次选择最违反该解的约束 (Most violated constraint), 将其加入到二次规划问题的约束集中, 然后再次求解新增约束的二次规划问题. 该算法的一个缺陷是, 每次求解违反当前解的约束问题时, 都需要求解 $E(\boldsymbol{x}|\boldsymbol{D}^k,\boldsymbol{w})-\Delta(\boldsymbol{x},\boldsymbol{x}_G^k)$ 的能量最小化问题, 而在一般情况下, 该问题是一个 NP-hard 问题, 仅在一些特殊情况下, 如满足 Submodular[5] 条件时, 才能得到其精确的最优解.

若使用如下的 Hinge-loss 损失函数 $\xi^k(\boldsymbol{w})$ 代替松弛变量 ϵ^k,

$$\xi^k(\boldsymbol{w})=E(\boldsymbol{x}_G^k|\boldsymbol{D}^k,\boldsymbol{w})-\min_{\boldsymbol{x}}\left(E\left(\boldsymbol{x}|\boldsymbol{D}^k,\boldsymbol{w}\right)-\Delta\left(\boldsymbol{x},\boldsymbol{x}_G^k\right)\right) \tag{31}$$

则可将约束优化问题 (30) 转化为如下的无约束优化问题:

$$\min_{\boldsymbol{w}}\|\boldsymbol{w}\|^2+C\sum_{k=1}^{K}\xi^k(\boldsymbol{w}) \tag{32}$$

式 (32) 与 SSVM (Structured support vector machine) 学习问题[100] 具有相同的形式. 均可使用随机次梯度下降 (Stochastic subgradient descent) 进行求解. 而损失函数 $\xi^k(\boldsymbol{w})$ 的形式决定了其次梯度的计算复杂度, 且每次次梯度的计算需要求解与其原能量函数类似的能量优化问题. Kim 等[50] 以及 Sun 等[51] 使用该方法估计用于图像理解的高阶能量模型的参数.

基于 Max-margin 的两种参数学习方法在很大程度上都依赖于其相应的能量优化算法. 而生成式的参数学习方法的性能在很大程度上依赖于对其边缘概率分布的估计. 故对高阶能量模型而言, 其最大的优势在于模型的表述能力, 而最大的挑战在于对应能量函数的优化.

5 高阶能量优化在场景理解中的应用

场景理解是计算机视觉领域最具挑战性的研究方向之一. 下面将分别讨论高阶能量优化在图像理解和三维场景理解中的应用现状:

5.1 高阶能量优化在图像理解中的应用现状

高阶能量优化在语义分割和整体 (图像) 场景理解中取得了较为成功的应用, 现分别对其进行阐述.

5.1.1 高阶能量优化在语义分割中的应用

图像语义分割 (多类别物体分割) 能够对图像上的所有区域 (像素或超像素) 赋以相应的语义类别, 是图像理解中最重要的研究领域之一. 为了达到更为准确的语义分割结果, 需要解决好如下三个关键问题: 1) 如何有效地利用不同尺度下的特征得到的分类结果; 2) 如何约束图像不同尺度下的 (类别) 一致性; 3) 如何利用全局的先验信息. 传统的基于低阶能量优化的语义分割算法, 仅能在单一尺度上对图像进行语义分割. 这些算法或者基于像素[55−57], 或者基于超像素[101−103]. 基于像素的语义分割算法由于特征的尺度较小, 其分类器的精度较低, 往往导致最终的语义分割结果不甚理想. 而基于超像素的语义分割算法由于存在超像素的初始分割错误无法纠正的问题, 也难以达到较理想的语义分割结果. 而利用关联层级网络模型[34−35, 58] 以及对图像的全局特性进行约束的高阶能量项 (如基于共生概率的高阶能量项[40]), 即可构造像素 – 区域 – 场景的多尺度层级网络, 从而有效地融合图像在不同尺度上的特征, 约束语义类别在各尺度之间的一致性, 进而有效地利用图像的全局先验.

为了更好地说明高阶能量优化在语义分割中的作用. 我们在 MSRC-21[56] 数据集上对如下 4 种算法进行了对比实验: 1) 基于像素的 TextonBoost 算法[56] (记为 Pixel-based CRF); 2) 基于超像素的语

义分割算法[101] (记为 Segment-based CRF); 3) 基于关联层级网络模型的语义分割算法 (记为 Hierarchical CRF); 4) 结合了全局共生概率[40] 和关联层级网络语义分割算法 (记为 Hierarchical CRF with CO). 关联层级网络模型包含像素层和三个嵌套的超像素层, 超像素由均值漂移算法生成[104]. 像素上的分类器基于 TextonBoost[56] 构造, 并利用 Joint boosting[105] 算法找出其判别性能最好的弱分类器的组合. 像素层的一阶项取其分类器输出值的负对数. 二阶项为在图像分割中常用的 Contrast sensitive Potts 模型[106]. 超像素层上的分类器为使用 Joint Boosting[105] 训练得到的基于区域直方图特征的分类器, 超像素之间的二阶项同样为 Contrast sensitive Potts 模型[106]. 4 种算法在 MSRC-21 数据集上得到的分类精度如表 1 所示.

表 1 中的数值代表按百分比显示的召回率, 召回率为 $N_{ii}/\sum_j N_{ij}$, 其中 N_{ij} 表示真值为 i 的类别被标注为类别 j 的像素总数. Global 指标定义为 $\sum_{i\in\mathcal{L}} N_{ii}/\sum_{i,j\in\mathcal{L}} N_{ij}$, 为全部像素中得到正确分类的像素比例, average 指标定义为 $\sum_{i\in\mathcal{L}} \frac{N_{ii}}{\sum_{j\in\mathcal{L}} N_{ij}}$ 为按类别平均的召回率. 由表 1 可看出, 与基于单尺度 (像素或超像素) 的语义分割算法相比, 基于关联层级网络模型的语义分割算法显著提高了像素分类的精度. 而在增加共生概率的高阶能量项[40] 后, 像素分类精度有所提高.

5.1.2 高阶能量优化在整体场景理解中的应用

完整的图像 (场景) 理解至少应包含如下三个主要任务: 1) 检测不同类别的物体并能区分同类别的不同物体; 2) 对物体进行定位; 3) 分割并描述物体的形状. 物体检测[107−108] 通过对检测到的物体给出限界框的形式, 能够完成图像理解的任务 1 和 2. 但无法给出精确的物体形状的表示. 而语义分割[34−35, 55, 57] 通过对每个像素 (或区域) 进行类别标注, 可完成对物体形状的分割, 但仅能区分不同的类别而无法区分同类别的不同物体, 如多辆汽车停靠在一起的情况, 通过融合物体检测和语义分割的结果可对每个像素同时赋以类别 (Class) 和物体 (Instance) 的标注从而完成图像理解的三个任务. 故在完整的场景理解中, 融合物体检测和语义分割是至关重要的步骤. 在融合物体检测和语义分割的算法中, 传统的一些交替优化的方法, 如文献 [109−111] 等, 通过将物体检测 (或语义分割) 的输出作为语义分割的 (或物体检测) 输入, 然后进行优化. 这种交替优化的方法在不同的优化阶段需优化不同的目标函数, 存在以下三个方面的不足: 1) 无法纠正初始错误; 2) 无收敛性保证; 3) 难以处理多个重叠的物体检测限界框的情况.

表 1 MSRC-21 数据集上的分类精度对比

Table 1 Quantitative results on the MSRC-21 dataset

	Pixel-based CRF	Segment-based CRF	Hierarchical CRF	Hierarchical CRF with CO
Global	75	84	87	**89**
Average	62	76	78	**80**
Building	63	73	81	**83**
Grass	**98**	93	96	96
Tree	**89**	84	**89**	**89**
Cow	66	77	74	**75**
Sheep	54	**84**	**84**	**84**
Sky	86	96	**99**	**99**
Aeroplane	63	**85**	84	84
Water	71	91	92	**94**
Face	83	**90**	**90**	**90**
Car	71	86	86	**87**
Bicycle	79	91	**92**	**92**
Flower	71	95	**98**	**98**
Sign	38	91	91	**92**
Bird	23	**41**	35	35
Book	88	92	**95**	**95**
Chair	23	53	53	**55**
Road	88	87	90	**91**
Cat	33	**65**	62	64
Dog	34	**77**	**77**	**77**
Body	43	**70**	**70**	**70**
Boat	**32**	17	12	11

基于全局能量优化的方法通过设计合理的高阶能量项对物体检测和语义分割之间的关系进行建模并联合求解, 能够完成整体场景理解的任务并能同时提高物体检测和分类的精度, 在最近几年得到了广泛重视. 这类方法成功的关键在于设计合适的高阶能量项, 使其既能够对物体检测和语义分割之间的关系进行有效建模, 同时又具有高效的能量优化算法. Gould 等[112] 以及 Wojek 等[113] 提出的高阶能量项由于缺乏有效的能量优化算法, 限制了这些能量模型的进一步应用. 首个同时满足建模有效性和计算有效性的高阶能量模型由 Ladický等[52] 于 2010 年提出, 其能量函数 $\psi_{q_j}^L$ 如图 9 (a) 所示. 在图 9 (a) 中, 横轴 $(N_{\ell_j}(\boldsymbol{x}_{bnd_j}))$ 代表候选物体检测器的限界框中与该候选物体检测器类别不一致的像素所占的比例, 纵轴代表归一化的能量函数值. 由图 9 (a) 可知, 当候选物体检测器中类别不一致的比例 $(N_{\ell_j}(\boldsymbol{x}_{bnd_j}))$ 小于该类别的最大允许不一致比例 $\rho(\ell_j)$ 时, 接受该候选物体检测器 $(y_j = 1,$ true positive), 此时能量函数的值随 $N_{\ell_j}(\boldsymbol{x}_{bnd_j})$ 线性增长; 当 $N_{\ell_j}(\boldsymbol{x}_{bnd_j})$ 大于 $\rho(\ell_j)$ 时, 拒绝该候选物体检测器 $(y_j = 0,$ false positive), 此时能量函数的值保持不变. 故能量项 $\psi_{q_j}^L$ 在接受候选物体检测器时鼓励其限界框内的像素类别与该候选物体检测器的类别一致, 而当拒绝该物体检测器时, 放弃对其限界框内像素类别的进一步约束. 同时, 该能量项属于 Robust $\mathcal{P}^n$ 模型, 具有高效的能量优化算法. 这些优点使得该能量项在图像理解中获得了广泛的应用.

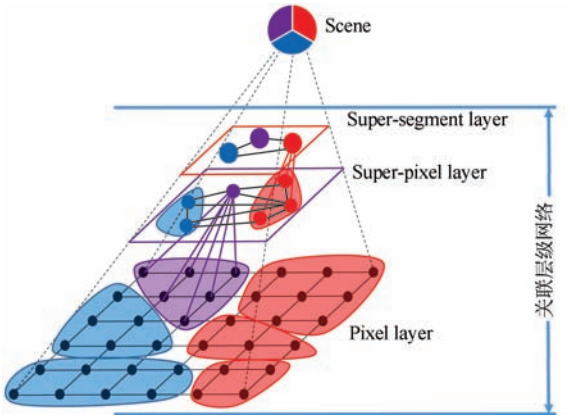
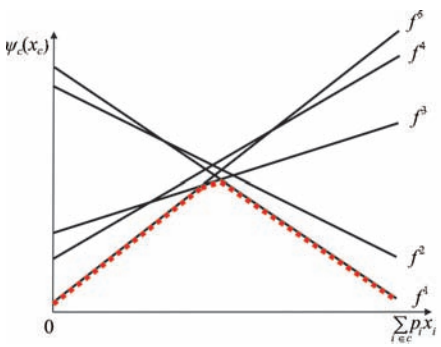


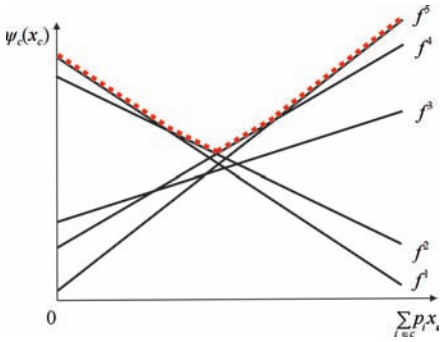
图 5　像素 – 区域 – 场景的多尺度层级网络

Fig. 5　Pixel-region-scene multi-scale hierarchical network



(a) 线性函数的下包络

(a) Lower envelope of linear functions



(b) 线性函数的上包络

(b) Upper envelope of linear functions

图 6　线性函数的上下包络

Fig. 6　Lower and upper envelopes of linear functions

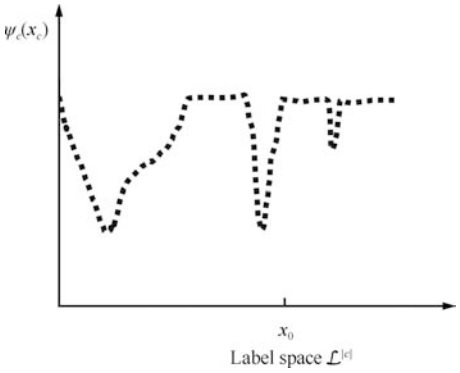


图 7　Pattern-based 能量项

Fig. 7　Pattern-based potential

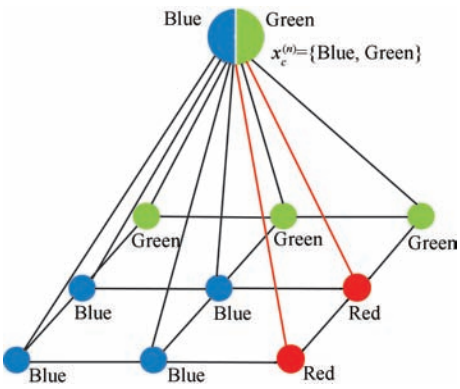


图 8　Harmony 能量项

Fig. 8　Harmony potential

但该能量项仍然存在如下两个不足: 1) 仅在接受候选物体检测器时才约束物体检测和语义分割之间的类别一致性; 2) 仅选取具有较高得分的物体检测器作为候选物体检测器, 导致物体检测的召回率较低.

为了克服这两个不足, Kim 等[50] 以及 Sun 等[51] 提出了一种新的高阶能量项 $\psi_{q_j}^{AKS}$, 如图 9 (b) 所示. 与图 9 (a) 相比, 当拒绝候选物体检测器时 $(N_{\ell_j}(\boldsymbol{x}_{bnd_j}) > \rho(\ell_j))$, $\psi_{q_j}^{AKS}$ 抑制其限界框内的像素取该候选物体检测器的类别, 且 $\psi_{q_j}^{AKS}$ 通过设置尽可能低的可接受的物体检测的得分的阈值从而包含尽可能多的候选物体检测器. 这使得 $\psi_{q_j}^{AKS}$ 能够利用尽可能多的物体检测器的信息从而提高物体检测的召回率, 并能在接受和拒绝候选物体检测器时, 均约束物体检测和语义分割之间的类别一致性.

我们在 PASCAL VOC 2010 数据集[114] 上对高阶能量项 $\psi_{q_j}^{L}$ 和 $\psi_{q_j}^{AKS}$ 进行了对比实验, 并使用关联层级网络模型作为对比的基线算法 (Baseline). 部分典型的实验结果如图 10 所示. 图 10 中共有三行, 每一行为一个典型的实验结果. 图 10 (a) 这一列为原始图像及其 True positive 的候选物体检测器, 图 10 (c) 这一列为原始图像及其 False positive 的候选物体检测器, 图 10 (d) 为语义分割和物体检测的真值, 图 10 (e) 为关联层级网络模型得到的语义分割结果, 图 10 (f) 为 $\psi_{q_j}^{L}$ 得到的语义分割的结果, 图 10 (f) 为 $\psi_{q_j}^{AKS}$ 得到的语义分割和物体检测的结果. 在图 10 (a) 和 (b) 中候选物体检测器的限界框的颜色代表了其得分, 颜色越深候选物体检测器的得分越低. 在第一个例子中, 语义分割的结果将摩托车的一部分错误地分为了人的类别. 高阶能量项 $\psi_{q_j}^{L}$ 和 $\psi_{q_j}^{AKS}$ 均能有效地利用得分较高的 Motorbike 候选物体检测器的输出, 从而纠正语义分割的错误. 在第二个例子中, 两只绵羊均未在语义分割中被检测出, 但高阶能量项 $\psi_{q_j}^{L}$ 仅选用具有较高得分的候选物体检测器构成其候选物体检测器集合. 故其未能有效利用两个得分较低的 True positive 候选物体检测器的输出. 而高阶能量项 $\psi_{q_j}^{AKS}$ 能够有效利用

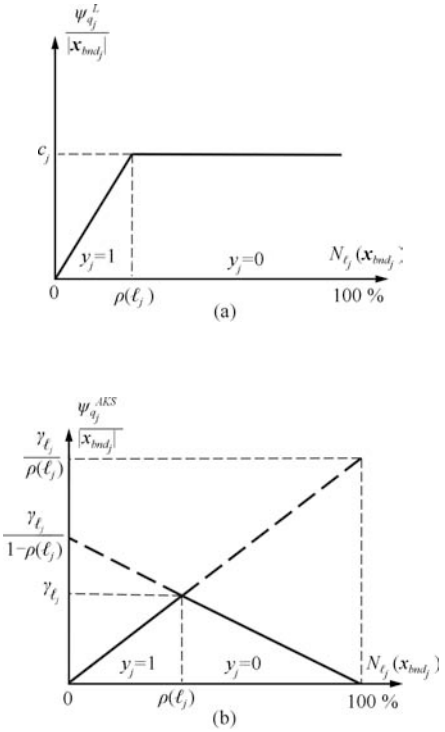


图 9 两种对物体检测和语义分割建模的高阶能量模型

Fig. 9 Two different higher-order energy models for relating semantic segmentation and object detection

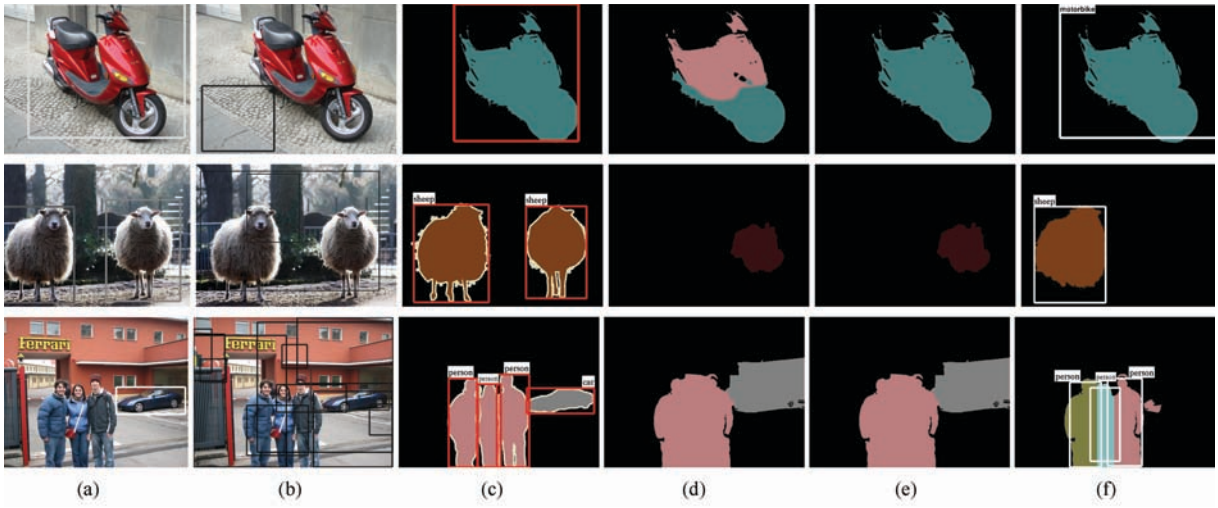


图 10 PASCAL VOC 2010 数据集上的一些典型结果

Fig. 10 Typical results on the PASCAL VOC 2010 data set

全部候选物体检测器的输出, 从而正确地检测和分割出了其中的一只绵羊. 在第三个例子中, 语义分割的结果使得前景 (汽车和人) 过于膨胀, 而所有的前景物体均具有较高得分的 True positive 和多个较低得分的 False positive 候选物体检测器. 高阶能量项 $\psi_{q_j}^{L}$ 仅能鼓励得分较高的候选物体检测器中的像素取与该候选物体检测器一致的类别, 故其未能使过于膨胀的前景物体 "收缩". 而高阶能量项 $\psi_{q_j}^{AKS}$ 能够有效地抑制在得分较低的候选物体检测器的限界框内的像素取与该限界框一致的类别, 从而使得前景物体 "收缩". 值得注意的是, 在该例中, 由于出现了较多得分较低且其限界框覆盖了真实物体较大区域的候选物体检测器, 使得前景 "收缩过度". 在 $\psi_{q_j}^{AKS}$ 的结果中, 汽车的类别输出完全被抑制 (甚至被错分为人的类别), 仅在人的类别上得到了较为精确的输出. 这也是高阶能量项 $\psi_{q_j}^{AKS}$ 的一个主要缺点: 当出现较多与物体有明显重叠且得分较低的候选物体检测器时, 高阶能量项 $\psi_{q_j}^{AKS}$ 不够鲁棒.

这三个算法在 PASCAL VOC 2010 数据集上的语义分割的精度对比如表 2 所示, 由表 2 可以看出, 通过引入对物件检测和语义分割之间的类别一致性进行建模的高阶能量项能够显著地提高多类别的物体分类精度.

在全局能量优化的框架下, 除了融合物体检测和语义分割之外, 还可进一步融合其他场景理解的手段从而提高整体场景 (图像) 理解的质量. 如 Yao 等[115] 提出了包含多个高阶能量项的高阶能量模型, 其中除了对物体检测和语义分割之间的类别一致性进行建模的高阶能量项外, 还包含对图像分类和语义分割之间的类别相容性进行建模的高阶能量项, 使得能够在同一能量优化框架下融合图像分类、物体检测和语义分割等问题, 从而达到更好的整体场景 (图像) 理解的效果. 其他的一些工作还包括纽约大学柯朗数学研究所的 Silberman 等[48] 提出的一种利用高阶能量项对室内场景中的结构类别和支撑关系进行建模和求解的方法, 该方法可同时得到 (图像) 场景中物体之间的支撑关系.

表 2　PASCAL VOC 2010 数据集上的分类精度对比
Table 2　Quantitative results on the PASCAL VOC 2010 dataset

	Baseline	$\psi_{q_j}^{L}$	$\psi_{q_j}^{AKS}$
Average	24.1	26.9	**27.5**
Background	78.5	**80.2**	77.6
Aeroplane	33.2	33.0	**35.2**
Bicycle	6.9	10.9	**14.5**
Bird	19.8	21.6	**24.8**
Boat	18.7	20.9	**21.2**
Bottle	11.7	12.1	**19.2**
Bus	38.5	39.3	**41.1**
Car	32.9	35.1	**36.2**
Cat	26.0	**27.0**	24.7
Chair	10.4	11.3	**12.5**
Cow	12.0	13.1	**14.2**
Dining table	23.1	**25.5**	23.1
Dog	11.1	13.5	**15.1**
Horse	11.2	13.8	**14.2**
Motorbike	**39.0**	36.3	32.1
Person	34.8	34.1	**34.9**
Potted plant	8.5	**18.8**	16.3
Sheep	24.4	**27.1**	26.3
Sofa	12.5	**18.5**	18.3
Train	31.2	**38.9**	35.3
TV monitor	20.8	33.7	**40.6**

5.2　高阶能量优化在三维场景理解中的应用现状

三维场景理解的主要目标主要有: 1) 在空间中分割并识别物体; 2) 完整化物体的几何形状; 3) 确定物体在空间中的位置; 4) 确定物体与物体以及物体与环境之间的关系. 本节主要从三维场景理解的两种主要形式: 1) 多视图 (或视频) 重建的三维场景; 2) 基于立体像对的三维场景, 对相关工作进行简要介绍.

5.2.1　多视图 (或视频) 重建的三维场景中高阶能量优化的应用现状

Sturgess 等[116] 在基于视频重建的三维道路场景理解中使用了与 Brostow 等[53] 相同的由 SFM (Structure from motion) 得到的空间点云提取的特征, 并将这些三维空间点的特征投影到图像上, 然后利用 Joint boosting[105] 对这些特征进行选择并将分类器的输出作为图像像素所属类别的一阶项. 其高阶项仅利用了 Robust $\mathcal{P}^n$ 模型对非监督聚类得到的超像素提供的一致性的约束, 从而在物体边界处取得了更精确的分割 (分类) 结果. 该算法存在如下的两个不足: 1) 高阶项的使用仅局限在二维图像上, 其高阶项的形式和所表达的约束与在图像分割中应用的 Robust $\mathcal{P}^n$ 模型无任何本质区别; 2) 该算法仅利用了三维 (点云) 特征进行图像上的二维语义分割而未对三维点云赋予语义信息. 针对以上两个不足, Floros 等[54] 在 Robust $\mathcal{P}^n$ 模型的基础上, 增加了一个约束三维空间点的类别与其在各视图上所有对应的二维图像点的类别一致性的高阶能量项. 该能量项的引入有效地整合了三维和二维信息, 增强了多视图之间语义一致性的约束. 但是对于三维场景理解而言, 该算法仍然存在以下不足: 由于三维场景中仅存在赋予了语义信息的空间点, 没有物体的概念, 所以无法在三维空间中对物体进行分割、识别和形状完整化等.

Roig 等[117] 提出了一种在条件随机场框架下利用高阶能量项进行多视图多类别的物体检测算法.

有别于传统的在多个视图上分别进行物体检测的算法, 该算法使用高阶能量项约束物体在多视图之间一致性的同时考虑了物体之间的遮挡关系, 并对遮挡关系进行了建模. 但该算法存在以下两个不足: 1) 该算法中的 "物体" 用预定义好的地平面上的离散化的格点 (Cell) 表示, 除粗略的位置信息外, 无任何其他几何信息 (姿态、形状、大小等); 2) 该高阶能量项仅使用类似于 "交替迭代" 的优化策略进行近似计算, 优化效果不是很理想.

5.2.2 基于立体像对的三维场景中高阶能量优化的应用现状

Ladický 等[49] 提出了一种利用立体匹配和物体分割之间的 "互信息" (即立体匹配中深度间断的地方可能为物体的边界, 反之亦然) 的高阶能量项. 利用该高阶能量项能够同时提高图像上物体分割和三维空间中深度估计的精度. 该算法揭示了高阶能量优化在融合多种场景理解手段中的优势: 即能在同一能量优化框架下, 通过联合优化二维和三维结构来提升整体场景理解的水平. 但该方法也存在如下两个不足: 1) 该高阶能量项在使用 Move-making 方法优化时, 其 Move space 是原高阶能量项所对应的 Move space 的投影空间 (Projected move space), 这使得 move space 大为减小, 从而影响最终的优化效果; 2) 该算法仅适用于单一立体像对的情景, 无法直接推广到含多个立体像对的三维场景理解中.

Bleyer 等[28] 提出了一种在立体匹配中同时进行物体分割的算法. 该算法中使用了两个高阶能量项: 1) Object-MDL 项; 2) 对三维空间物体连通性进行约束的高阶能量项. Object-MDL 项与 Delong 等[44−45] 提出的标记代价约束相同, 用于约束场景的复杂度, 使场景中包含尽可能少的物体. 三维空间中物体连通性约束在该算法中作为一个 "硬约束": 即若物体在二维图像上连通或者被深度更小的物体遮挡, 则认为物体在三维空间中连通, 此时惩罚为 0; 否则, 其惩罚为无穷大. 由于该高阶能量目前仍无有效的优化算法, 导致该高阶能量项不能在能量优化的过程中直接使用, 仅能在每次 Fusion move 之后作为验证使用.

另一类在基于立体像对的三维场景中应用的高阶能量项, 是所谓的 "高阶平滑先验", 如 Woodford 等[25, 47] 的工作. 这类算法均具有如下特点: 1) 表达先验的类型单一, 均为 "平滑" 先验; 2) 应用领域受限, 仅能用来提高立体匹配算法的匹配精度, 与 "场景理解" 关联不大.

总体而言, 当前文献中关于高阶能量优化在三维场景理解中的应用还较为初步, 离三维场景理解的主要目标还有一定距离.

6 结论与展望

由于能够有效地描述场景的结构先验, 表达场景的局部和全局性先验并能在同一能量优化框架下有效地融合多种场景理解手段, 高阶能量模型在场景理解的应用中展现出了巨大的优势和潜力. 但高阶能量模型也存在很多固有的不足. 首先是表达能力和可求解性的矛盾. 例如文献中很多复杂的高阶能量模型, 尽管对场景理解提供了丰富的约束, 但由于缺乏对应的优化方法, 只能使用并不适合该模型的通用优化方法求解. 在这种情况下, 实验报道的 "好结果" 是否有具有一般性的意义值得怀疑, 所提出的高阶能量模型到底起了多少作用缺乏理论依据. 所以, 我们觉得, 不考虑可求解性的 "高阶能量模型的建模" 的意义并不大. 另一个问题是, 目前的高阶能量模型或求解方法, 都过于有点 "特定问题, 特定处理", 方法缺乏足够的通用性, 这无疑限制了它的应用范围和科学价值. 本质上, 科学研究的意义在于从 "特殊中发现普遍规律", 一个领域如果长期缺乏一般性指导理论, 是不会有生命力的. 所以, 我们觉得高阶能量优化今后的研究, 以下 4 个方面特别值得关注:

1) 高阶能量模型的表达问题和优化问题应作为一个整体进行研究, 应该寻求同时满足 "约束有效性" 和 "表达稀疏性" 的高阶能量模型.

2) 探求更为通用的高效能量优化方法. 我们觉得, 关联层级条件随机场模型和基于对偶分解的优化方法, 在表达和优化方面具有相当大的潜力, 是值得进一步深入探索的方向.

3) 从目前的报道看, 还鲜有高阶能量优化在 CPU-GPU 异构集群上的高效实现, 估计很快会有这方面的进展和相关报道.

4) 深度神经网络和深度学习, 从某种程度上说, 也是一种层次化的知识表示. 关联层级条件随机场模型与深度神经网络在结构上有一定的相似性, 更一般地, 它们都是层次化结构 (Hierarchical architecture) 下的特殊网络. 关联层级条件随机场模型与目前的卷积神经网络等深度网络有一定的内在联系吗? 目前还没有相关报道. 我们觉得, 探索这种联系不论在理论上还是在指导具体应用网络的构建上均具有重要的意义.

总之, 我们觉得一种理论或方法不能仅仅局限于 "各式各样的应用" 和 "技巧式的特定问题处理" 上, 更需要探索具有普适性的方法和揭示深层次的机理, 高阶能量优化亦是如此.

References

1 Li S Z. *Markov Random Field Modeling in Image Analysis*. London: Springer, 2009.

2 Blake A, Kohli P, Rother C. *Markov Random Fields for Vision and Image Processing*. Cambridge: MIT Press, 2011.

3 Blake A, Kohli P, Rother C [Author], Xie Zhao [Translator]. *Markov Random Fields for Vision and Image Processing*. Beijing: Science Press, 2014.
(Blake A, Kohli P, Rother C [著], 谢昭 [译]. Markov 随机场在视觉和图像处理中的应用. 北京: 科学出版社, 2014.)

4 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(11): 1222−1239

5 Kolmogorov V, Zabin R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(2): 147−159

6 Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 2006, **70**(1): 41−54

7 Weiss Y, Freeman W T. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *IEEE Transactions on Information Theory*, 2001, **47**(2): 736−744

8 Murphy K P, Weiss Y, Jordan M I. Loopy belief propagation for approximate inference: an empirical study. In: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc., 1999. 467−475

9 Wainwright M J, Jaakkola T S, Willsky A S. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 2005, **51**(11): 3697−3717

10 Kolmogorov V. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(10): 1568−1583

11 Kolmogorov V, Wainwright M J. On the optimality of tree-reweighted max-product message-passing. In: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence. 2012.

12 Wainwright M J, Jordan M I. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 2008, **1**(1−2): 1−305

13 Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(6): 1068−1080

14 Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge: MIT Press, 2009.

15 Bishop C. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

16 Jordan M I, Ghahramani Z, Jaakkola T S, Saul L K. An introduction to variational methods for graphical models. *Machine Learning*, 1999, **37**(2): 183−233

17 Lauritzen S L. *Graphical Models*. Oxford: Oxford University Press, 1996.

18 Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1974, **36**(2): 192−236

19 Hammersley J M, Clifford P. Markov fields on finite graphs and lattices. 1971, unpublished. http://www.statslab.cam.ac.uk/∼grg/books/hammfest/hamm-cliff.pdf

20 Loeliger H A. An introduction to factor graphs. *IEEE Signal Processing Magazine*, 2004, **21**(1): 28−41

21 Kschischang F R, Frey B J, Loeliger H A. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 2001, **47**(2): 498−519

22 Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for Markov random fields. In: Proceedings of the 9th European Conference on Computer Vision, Computer Vision-ECCV 2006. Graz, Austria: Springer, 2006. 16−29

23 Greig D M, Porteous B T, Seheult A H. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1989, **51**(2): 271−279

24 Tappen M F, Freeman W T. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In: Proceedings of the 9th IEEE International Conference on Computer Vision, 2003. Nice, France: IEEE, 2003. 900−906

25 Woodford O J, Torr P H S, Reid I D, Fitzgibbon A W. Global stereo reconstruction under second order smoothness priors. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

26 Vicente S, Kolmogorov V, Rother C. Graph cut based image segmentation with connectivity priors. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

27 Nowozin S, Lampert C H. Global connectivity potentials for random field models. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2009). Miami, FL: IEEE, 2009. 818−825

28 Bleyer M, Rother C, Kohli P, Scharstein D, Sinha S. Object stereo — joint stereo matching and object segmentation. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI: IEEE, 2011. 3081−3088

29 Lempitsky V, Kohli P, Rother C, Sharp T. Image segmentation with a bounding box prior. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto: IEEE, 2009. 277−284

30 Kohli P, Kumar M P, Torr P H S. P^3 & beyond: move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(9): 1645−1656

31 Kohli P, Kumar M P, Torr P H S. P^3 & beyond: solving energies with higher order cliques. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

32 Kohli P, Ladickỳ L, Torr P H S. Robust higher order potentials for enforcing label consistency. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

33 Kohli P, Ladický L, Torr P H S. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 2009, **82**(3): 302−324

34 Ladický L, Russell C, Kohli P, Torr P H S. Associative hierarchical random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(6): 1056−1077

35 Ladický L, Russell C, Kohli P, Torr P H S. Associative hierarchical CRFs for object class image segmentation. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto: IEEE, 2009. 739−746

36 Rother C, Kohli P, Feng W, Jia J Y. Minimizing sparse higher order energy functions of discrete variables. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2009). Miami, FL: IEEE, 2009. 1382−1389

37 Komodakis N, Paragios N. Beyond pairwise energies: efficient optimization for higher-order MRFs. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). Miami, FL: IEEE, 2009. 2985−2992

38 Boix X, Gonfaus J M, van de Weijer J, Bagdanov A D, Serrat J, González J. Harmony potentials. *International Journal of Computer Vision*, 2012, **96**(1): 83−102

39 Gonfaus J M, Boix X, Van de Weijer J, Bagdanov A D, Serrat J, Gonzalez J. Harmony potentials for joint classification and segmentation. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 3280−3287

40 Ladický L, Russell C, Kohli P, Torr P H S. Graph cut based inference with co-occurrence statistics. In: Proceedings of the 11th European Conference on Computer Vision, Computer Vision-ECCV 2010. Heraklion, Crete, Greece: Springer, 2010. 239−253

41 Ladický L, Russell C, Kohli P, Torr P H S. Inference methods for CRFs with co-occurrence statistics. *International Journal of Computer Vision*, 2013, **103**(2): 213−225

42 Werner T. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

43 Lim Y, Jung K, Kohli P. Energy minimization under constraints on label counts. In: Proceedings of the 11th European Conference on Computer Vision, Computer Vision-ECCV 2010. Heraklion, Crete, Greece: Springer, 2010. 535−551

44 Delong A, Osokin A, Isack H N, Boykov Y. Fast approximate energy minimization with label costs. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 2173−2180

45 Delong A, Osokin A, Isack H N, Boykov Y. Fast approximate energy minimization with label costs. *International Journal of Computer Vision*, 2012, **96**(1): 1−27

46 Shekhovtsov, Kohli P, Rother C. Curvature prior for mrf-based segmentation and shape inpainting. In: Proceedings of the Joint 34th DAGM and 36th OAGM, Pattern Recognition, Lecture Notes in Computer Science Volume 7476. Berlin Heidelberg: Springer, 2012. 41−51

47 Woodford O, Torr P, Reid I, Fitzgibbon A. Global stereo reconstruction under second-order smoothness priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(12): 2115−2128

48 Silberman N, Hoiem D, Kohli P, Fergus R. Indoor segmentation and support inference from RGBD images. In: Proceedings of the 12th European Conference on Computer Vision, Computer Vision-ECCV 2012. Florence, Italy: Springer, 2012. 746−760

49 Ladický L, Sturgess P, Russell C, Sengupta S, Bastanlar Y, Clocksin W, Torr P H S. Joint optimization for object class segmentation and dense stereo reconstruction. *International Journal of Computer Vision*, 2012, **100**(2): 122−133

50 Kim B S, Sun M, Kohli P, Savarese S. Relating things and stuff by high-order potential modeling. In: Proceedings of the 2012 Computer Vision-ECCV. Workshops and Demonstrations. Berlin, Heidelberg: Springer, 2012. 293−304

51 Sun M, Kim B S, Kohli P, Savarese S. Relating things and stuff via object property interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(7): 1370−1383

52 Ladický L, Sturgess P, Alahari K, Russell C, Torr P H S. What, where and how many? Combining object detectors and CRFs. In: Proceedings of the 11th European Conference on Computer Vision, Computer Vision-ECCV 2010. Heraklion, Crete, Greece: Springer, 2010. 424−437

53 Brostow G J, Shotton J, Fauqueur J, Cipolla R. Segmentation and recognition using structure from motion point clouds. In: Proceedings of the 10th European Conference on Computer Vision, Computer Vision-ECCV 2008. Marseille, France: Springer, 2008. 44−57

54 Floros G, Leibe B. Joint 2d-3d temporally consistent semantic segmentation of street scenes. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI: IEEE, 2012. 2823−2830

55 Shotton J, Winn J, Rother C, Criminisi A. Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 2009, **81**(1): 2−23

56 Shotton J, Winn J, Rother C, Criminisi A. Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Proceedings of the 9th European Conference on Computer Vision, Computer Vision-ECCV 2006. Graz, Austria: Springer, 2006. 1−15

57 Chris R, L′ubor L, Pushmeet K, Philip HS T. Exact and approximate inference in associative hierarchical networks using graph cuts. arXiv preprint arXiv: 1203.3512, 2012.

58 Russell C, Ladickỳ L, Kohli P, Torr P H S. Exact and approximate inference in associative hierarchical networks using graph cuts. In: UAI. AUAI Press, 2010. 501−508

59 Kohli P, Kumar M P. Energy minimization for linear envelope MRFs. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 1863−1870

60 Gould S. Max-margin learning for lower linear envelope potentials in binary Markov random fields. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). Omnipress, 2011. 193−200

61 Lempitsky V, Rother C, Blake A. LogCut-efficient graph cut optimization for Markov random fields. In: Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV 2007). Rio de Janeiro: IEEE, 2007. 1−8

62 Lempitsky V, Rother C, Roth S, Blake A. Fusion moves for Markov random field optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(8): 1392−1405

63 Werner T. Revisiting the linear programming relaxation approach to gibbs energy minimization and weighted constraint satisfaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(8): 1474−1488

64 Kolmogorov V, Rother C. Minimizing nonsubmodular functions with graph cuts — a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(7): 1274−1279

65 Boros E, Hammer P L. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 2002, **123**(1−3): 155−225

66 Boros E, Hammer P L, Tavares G. Preprocessing of Unconstrained Quadratic Binary Optimization. Technical Report RRR 10-2006, RUTCOR, 2006.

67 Rother C, Kolmogorov V, Lempitsky V, Szummer M. Optimizing binary MRFs via extended roof duality. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

68 Rosenberg I G. Reduction of bivalent maximization to the quadratic case. *Cahiers du Centre d′Etudes de Recherche Opérationnelle*, 1975, **17**: 71−74

69 Ishikawa H. Higher-order clique reduction in binary graph cut. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). Miami, FL: IEEE, 2009. 2993−3000

70 Ishikawa H. Transformation of general binary MRF minimization to the first-order case. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(6): 1234−1249

71 Freedman D, Drineas P. Energy minimization via graph cuts: settling what is possible. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005). San Diego, CA, USA: IEEE, 2005. 939−946

72 Gallagher A C, Batra D, Parikh D. Inference for order reduction in Markov random fields. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI: IEEE, 2011. 1857−1864

73 Fix A, Gruber A, Boros E, Zabih R. A graph cut algorithm for higher-order Markov random fields. In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona: IEEE, 2011. 1020−1027

74 Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann, 1988.

75 Lan X Y, Roth S, Huttenlocher D, Black M J. Efficient belief propagation with learned higher-order Markov random fields. In: Proceedings of the 9th European Conference on Computer Vision, Computer Vision-ECCV 2006. Graz, Austria: Springer, 2006. 269−282

76 Potetz B. Efficient belief propagation for vision using linear constraint nodes. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

77 Potetz B, Lee T S. Efficient belief propagation for higher-order cliques using linear constraint nodes. *Computer Vision and Image Understanding*, 2008, **112**(1): 39−54

78 Tarlow D, Givoni I E, Zemel R S. Hop-map: efficient message passing with high order potentials. In: Proceedings of the 13th Conference on Artificial Intelligence and Statistics. 2010. 812−819

79 McAuley J J, Caetano T S. Faster algorithms for max-product message-passing. *The Journal of Machine Learning Research*, 2011, **12**: 1349−1388

80 Felzenszwalb P F, McAuley J J. Fast inference with min-sum matrix product. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(12): 2549−2554

81 Komodakis N, Tziritas G, Paragios N. Fast, approximately optimal solutions for single and dynamic MRFs. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

82 Bertsekas D P. *Nonlinear Programming* (2nd Edition). Belmont, Mass: Athena Scientific, 1999.

83 Vazirani V V. *Approximation Algorithms*. Berlin, Heidelberg: Springer, 2001.

84 Kovalevsky V A, Koval V K. A diffusion algorithm for decreasing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR, 1975.

85 Werner T. A linear programming approach to max-sum problem: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(7): 1165−1179

86 Komodakis N, Paragios N, Tziritas G. MRF optimization via dual decomposition: message-passing revisited. In: Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV 2007). Rio de Janeiro: IEEE, 2007. 1−8

87 Komodakis N, Paragios N, Tziritas G. MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(3): 531−552

88 Swoboda P, Savchynskyy B, Kappes J H, Schnörr C. Partial optimality by pruning for map-inference with general graphical models. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR′14. Washington D. C., USA: IEEE Computer Society, 2014. 1170−1177

89 Komodakis N, Paragios N. Beyond loose *Lp*-relaxations: optimizing MRFs by repairing cycles. In: Proceedings of the 10th European Conference on Computer Vision, Computer Vision-ECCV 2008. Marseille, France: Springer, 2008. 806−820

90 Kumar M P, Torr P H S. Efficiently solving convex relaxations for map estimation. In: Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008. 680−687

91 Sontag D, Jaakkola Y S. New outer bounds on the marginal polytope. In: Proceedings of the 2007 Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2007. 1393−1400

92 Sontag D, Meltzer T, Globerson A, Jaakkola T S, Weiss Y. Tightening LP relaxations for MAP using message passing. In: Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence. 2012.

93 Andres B, Kappes J H, Köthe U, Schnörr C, Hamprecht F A. An empirical comparison of inference algorithms for graphical models with higher order factors using openGM. In: Proceedings of the 32nd DAGM Symposium, Pattern Recognition. Darmstadt, Germany: Springer, 2010. 353−362

94 Kappes J H, Andres B, Hamprecht F A, Schnorr C, Nowozin S, Batra D, Kim S, Kausler B X, Lellmann J, Komodakis N, Rother C. A comparative study of modern inference techniques for discrete energy minimization problems. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR: IEEE, 2013. 1328−1335

95 Andres B, Beier T, Kappes J H. Opengm: A C++ library for discrete graphical models. *arXiv Preprint arXiv: 1206.0111*, 2012.

96 Scharstein D, Chris P. Learning conditional random fields for stereo. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

97 Taskar B, Guestrin C, Roller D. Max-margin Markov networks. *Advances in Neural Information Processing Systems*, 2004, **16**: 25

98 Finley T, Joachims T. Training structural SVMs when exact inference is intractable. In: Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008. 304−311

99 Li Y P, Huttenlocher D P. Learning for stereo vision using the structured support vector machine. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

100 Tsochantaridis I, Hofmann T, Joachims T, Altun Y. Support vector machine learning for interdependent and structured output spaces. In: Proceedings of the 21st International Conference on Machine Learning. New York: ACM, 2004. 104

101 Yang L, Meer P, Foran D J. Multiple class segmentation using a unified framework over mean-shift patches. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR′07). Minneapolis, MN: IEEE, 2007. 1−8

102 Pantofaru C, Schmid C, Hebert M. Object recognition by integrating multiple image segmentations. In: Proceedings of the 10th European Conference on Computer Vision, Computer Vision-ECCV 2008. Marseille, France: Springer, 2008. 481−494

103 Russell B C, Freeman W T, Efros A A, Sivic J, Zisserman A. Using multiple segmentations to discover objects and their extent in image collections. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 1605−1614

104 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 603−619

105 Torralba A, Murphy K P, Freeman W T. Sharing features: efficient boosting procedures for multiclass object detection. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004). Washington D. C., USA: IEEE, 2004. II-762− II-769

106 Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV 2001). Vancouver, BC: IEEE, 2001. 105−112

107 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

108 Maji S, Malik J. Object detection using a max-margin Hough transform. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). Miami, FL: IEEE, 2009. 1038−1045

109 Larlus D, Jurie F. Combining appearance models and Markov random fields for category level object segmentation. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−7

110 Hoiem D, Efros A A, Hebert M. Closing the loop in scene interpretation. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, AK: IEEE, 2008. 1−8

111 Li C C, Kowdle A, Saxena A, Chen T. Towards holistic scene understanding: feedback enabled cascaded classification models. In: Proceedings of the 2010 Advances in Neural Information Processing Systems. 2010. 1351−1359

112 Gould S, Gao T S, Koller D. Region-based segmentation and object detection. In: Proceeding of the 2009 Advances in Neural Information Processing Systems. 2009. 655−663

113 Wojek C, Schiele B. A dynamic conditional random field model for joint labeling of object and scene classes. In: Proceedings of the 10th European Conference on Computer Vision, Computer Vision-ECCV 2008. Marseille, France: Springer, 2008. 733−747

114 Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, **88**(2): 303−338

115 Yao J, Fidler S, Urtasun R. Describing the scene as a whole: joint object detection, scene classification and semantic segmentation. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI: IEEE, 2012. 702−709

116 Sturgess P, Alahari K, Ladický L, Torr P H S. Combining appearance and structure from motion features for road scene understanding. In: Proceedings of the 2009 British Machine Vision Association (BMVC 2009).

117 Roig G, Boix X, Ben Shitrit H, Fua P. Conditional random fields for multi-camera object detection. In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona: IEEE, 2011. 563−570

余 淼 中原工学院讲师, 中国科学院自动化研究所博士研究生. 分别于 2004 年和 2007 获得西南交通大学管理学学士和工学硕士学位. 主要研究方向为场景理解和三维重建.
E-mail: myu@nlpr.ia.ac.cn
(**YU Miao** Lecturer at Zhongyuan University of Technology, and Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree in 2004 and master degree in 2007 from Southwest Jiaotong University, respectively. His research interest covers scene understanding and 3D reconstruction.)

胡占义 中国科学院自动化研究所研究员. 主要研究方向为摄像机标定, 三维重建, 视觉机器人导航. 本文通信作者.
E-mail: huzy@nlpr.ia.ac.cn
(**HU Zhan-Yi** Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers camera calibration, 3D reconstruction, and vision guided robot navigation. Corresponding author of this paper.)