

指导教师： 杨涛

提交时间： 2015/3/29

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 1

姓名： 茹梦

学号： 2012302453

班号： 10011203



基于互斥约束的利用束搜索的场景标记

Anirban Roy
Oregon State University
Corvallis, OR 97330
royan@eecs.oregonstate.edu

摘要

本篇论文主要解决了关于分配类对象标签为图像像素的问题。紧跟最新的整体规划研究，我们将场景标记作为以超像素为基础的条件随机场（CRF）的推论来计算。CRF 推论被指定为类标签分配时带有互斥变量的二次规划（QP），QP 可以用一种能够很好地适配场景标记的束搜索算法来解决，因为这种束搜索可以明确的计算对象的空间扩展，同时符合域知识的不一致约束，并且计算量较小。BS 最终可以得到一个所有节点恰好是候选标记场景的搜索树，其后继节点是重复产生于父节点直至收敛。我们能够证明我们的束搜索算法能够有效的最大化条件随机场的二次子规划目标。最后关于我们针对场景标记的束搜索的高效性已经在斯坦福 MSRC 的标准程序上进行了评估。

1. 简介

本片论文强调了场景标记的问

题，该问题的目标是在一系列的类集

Sinisa Todorovic
Oregon State University
Corvallis, OR 97330
sinisa@eecs.oregonstate.edu

中为每一位图像像素标记一个类标号。目标类集应该包括所有的对象和场景表面（例如：草地，天空等）。由于场景中呈现的物体有可能是被遮挡，被缩放，其空间形态范围很广，实景图像在场景标识中是一个巨大的挑战。

前期工作已经论证了整体可行的会出现和重复出现的所有类集（31, 30, 32, 8, 9, 28, 13, 18, 34, 15, 29），受空间限制的布局为场景标识提供了可行的框架。这些途径通过基于超像素（或图像补丁）的条件随机场来模拟现实场景。这样，他们采取常见的识别策略：a) 平滑相邻的图像区域将被相同的对象所占据；b) 的图像区域将被频繁重现的对象所占据

受到这些方法成功的激励，我们将图像作为全连接的基于超像素的 CRF 进行描绘，并把场景标识作为对 CRF 推论中超像素的类标签分配来规

划。跟随一个固定确立的研究线，我们将 CRF 推论作为二次子规划来计算。相较于现有的线性规划，QP 涉及计算相对较少的变量，提供了一种可分离的约束在优化设置，并允许可微的 large-margin 参数估计。

我们的主要贡献是为了计算基于二次规划的 CRF 推断的束搜索算法 (BS)。BS 非常适合场景标记，因为以下几点：

1. 明确计算了对象的空间扩展；
2. 完全解决了 QP 在离散域符合有用域限制；
3. 不需要普通的 QP 凸化和松弛算法，还有最终解决方案的离散化；
4. 它计算量很小，允许有大量的 CRF 结点，包括全节点连通图，就想为场景对象间建立远距离依赖性模型通常需要的那样。

BS 开始于对应于初始状态的一个初始超像素标签，如图所示. 1。

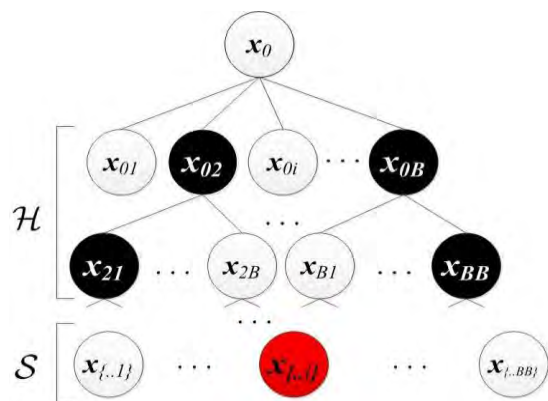


图 1. BS: 搜索树中的状态对应于候选场景标记。树是通过从 B 的启发式函数估计的一个子集的最佳状态 (黑色) 产生后继状态逐步扩大。评分函

数选择最优叶节点 (红色) 作为 CRF 推论的解决方案

然后它逐渐增长为一棵搜索树，其中树节点对应于搜索状态，即候选场景标记。树的深度是增加的就像新的后继状态是继承于当前深度状态的一个子集。该搜索将一直持续到收敛，也就是不能够产生一个“更好”的后继状态时。BS 是由以下三个功能定义的：继承者——通过随机生成后继状态以达到搜索空间的随机探索；启发式——用于选择一系列 B 搜索中“最好”的当前状态，其中 B 是输入的树搜索的宽度参数；得分——用于选择“最佳”状态作为解决方案。

继承者的功能明确解释了对象的空间扩展，通过产生新状态时共同翻转超像素连接组件的一个类标签；启发式功能可以有效的计算父状态和后继状态之间 CRF 的条件对数似然函数的差异。它是有效的，因为它只计算状态之间发生改变的超像素的标签，而不是所有超像素。最后，该分数的功能是有有效评估叶状态的 CRF 条件对数似然函数，并且选择得分最高的状态为最优解。需要注意的是，通过构造，叶子状态可以保证在搜索树的所有状态中拥有最高分数值，从而满足仅在叶子状态中就可以找到解决方案。

我们将我这个针对场景标记的 BS 算法 的有效性呈现在 MSRC [28], Stanford Backgroud [8] , PASCAL VOC2009 and 2010 [3]等平台的数据集上。

后续文章中, 2. 指出我们相对于以前的工作的贡献; 3. 构造 CRF; 4. 用 QP 详细说明 CRF 推理; 5. 构造 BS; 6. 定义 CRF 潜力, 并介绍了如何从训练数据学习潜力和互斥的限制; 7. 展示我们的结果。

2. 前期工作和们的贡献

本节回顾相关工作并且指出我们以下几点贡献: 1. 加强 CRF 推论的硬约束; 2 直接在离散域中求解 QP; 3. 说明我们的 BS 算法的低复杂性。

解释了场景中对象之间的与约束是非常重要的, 因为它们可以帮助解决推论中的矛盾假设问题。在本篇论文中, 我们将重点关注阻止确定标签分配的互斥约束 (例如, 天空不可能出现在草地下面)。CRF 用特征来描绘域约束。为了加强 CRF 推论的约束, 特征加权应该足够大以便惩罚违反约束的场景标签。然而, 由于权重不能无限大, 在某些情况下, CRF 的推论可以推翻约束, 屈服无意义的结果。

我们通过保持CRF针对编码平滑

和上下文的其他潜在功能独立的域约束来解决这个问题。我们将CRF推断作为带有二次约束的QP来计算, 具体地讲, 我们使用CRF平滑和上下文的潜在功能来表达的QP目标, 并分别使用域的互斥约束来表达的QP的二次约束。

作为QP的CRF推断通常需要凸化QP目的来允许使用标准的凸优化算法。凸化可以避免。例如, 通过利用消息传递[16], 或线性规划和QP之间级数递增。然而, 由于它们的复杂度依赖于CRF边缘的数, 这些方法都不是用于全连通的CRF。现有的QP的近似半定规划也不合适, 因为QP的二次目标矩阵不能被假定为是 (“接近”) 半正定。

重要的是, 大部分的QP求解放宽优化约束到连续域。这将是合适的, 因为域约束可能违反连续松弛。相比之下, 我们的BS不使用连续松弛, 而是直接在离散域中解决QP, 严格执行域约束。

我们的方法Swendsen-Wang(SW)相关[1], 都是迭代Metropolis-Hastings算法可逆跳转。每跳随机切割图形边缘并且翻转一整个连接组的的节点的标签, 以便比其它的MCMC算法更快的勘探搜索空间。然而, SW对每个可访问的状态进行评

估，对于较大的图像，他比我们的算法更昂贵一些。另外在实践中，由于运行时间过长，SW迭代常常在收敛之前就中断。相比之下，我们的BS算法能够被heuristi函数有效的引导至“好”的候选解决方案，并保证快速收敛到局部最大值。

基于搜索的结构预测方法是越来越多的计算机视觉研究[10, 5, 14, 24]的趋势，但他们从未使用过的场景标签。其关键限制是对近似损耗函数的要求，并且因此引导着搜索。收到HC搜索[25, 2]的启发，我们改用排序搜索策略，使得搜索决定是通过比较启发函数分配的搜索状态的相对值来做的。如在[25, 2]，我们使用启发式函数引导波束搜索并且用分数函数确定解决方案。关键的区别是，我们从原来的(CRF-based)优化目标得到了启发和分数值，而这两个函数将分别在一个分布式学习架构[25, 2]中学习。

3. CRF 模型

图像被划分成超像素，它们被用来连接我们现场的CRF。特别是，超像素被组织在一个图， $G=(V,E)$ ， V 是一组节点， $i=1,\dots,n, |V|=n$ ，对应于所有的超像素。 E 是边集 $(i,j) \in E$ 是对超像素 i 和 j 之间捕获的依赖。在本文中，我们考虑一个完全连通图，即边能够连接所有节点对 $E=V \times V, |E|=n^2$ 。

CRF 将一个示性随机变量 X_i 与每

一个结点 $i \in V$ 相关联。每个 X_i 的取值从一组对象类的标签 $L = \{1,2,\dots,k\}$ 得到，其中 $|L| = k$ 。当 $X_i = i' \in L$, CRF 分配类标记 i' 给超像素 i 。该集中的所有随机变量记为 $\mathbf{X} = \{X_i : i \in V\}$ 。CRF 的条件对数似然函数被定义为

$$\log P(\mathbf{X} | G) = \sum_{i \in V} \phi_i(X_i = i') + \sum_{(i,j) \in E} \phi_{ij}(X_i = i', X_j = j') - \log Z$$

$\{i',j'\} \in L$, Z 是一个配分函数。单元势能 $\phi_i(X_i = i')$ 被定义为一个其标签 $X_i \in L$ 的对数似然值 X_i 。邻域势能 $\phi_{ij}(X_i = i', X_j = j')$ 代表联合对数似然值 X_i 和 X_j 分别拥有标签 i' 和 j' 。在下文中，我们将使用速记符号 $\phi_i(i') = \phi_i(X_i = i')$ ，和 $\phi_{ij}(i',j') = \phi_{ij}(X_i = i', X_j = j')$ 标识指定的单元和邻域势能。

我们将寻找 MAP 任务 $\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} P(\mathbf{X} | G)$ 指定为我们的场景标记问题。下面的章节中，我们将讨论如何进行这一推论。

4. 作为二次规划的 CRF 推论

本节制定了 MAP 的分配问题 QP。我们首先推导 QP 的二次目标，然后扩展该构想以包含域约束。公主，我想知道不等于好了再说，二，

由(1)可以很方便的得到 $\log(X|G)$ ，由二进制随机变量 $x_{ii'} \in \{0, 1\}$ 来表达每一个超像素标签对。当 $X_i = i'$ ， $x_{ii'} = 1$ ；当 $X_i = i''$ ， $x_{ii'} = 0$ 。所有二进制

随机变量 (n, k) 的列向量用 $x = [\dots x_{ii'} \dots]^T$ 来表示。因此，MAP 分配问题可以表示为

$$\begin{aligned} \max. & \sum_{i \in V, i' \in L} \phi_{ii'} x_{ii'} + \sum_{(i,j) \in E, i', j' \in L} \phi_{ii'jj'} x_{ii'} x_{jj'} \\ \text{s.t.} & \text{for all } i \in V, \sum_{i' \in L} x_{ii'} = 1, \quad x \in \{0, 1\}^{n \cdot k} \end{aligned} \quad (2)$$

(2) 式的二次目标可以紧凑地表示为 $x^T \Phi x$ ，其中 Φ 是 $(n \cdot k) \times (n \cdot k)$ 的关联矩阵，其元素为单元和邻接势能。 Φ 的非对角线元素被定义为 $\Phi_{(ii'),(jj')} = \phi_{ii'jj'}$ ，主对角线元素被定义为 $\Phi_{(ii'),(ii')} = \phi_{ii'}$ 。

正如第 1 和第 2 小节提到的，我们的下一个目标是将域约束包含进 QP，预计通过消除计算中的非法状态来提高解决方案的质量。在本文中，我们专注于互斥约束，禁止某些无意义的标签分配。例如，假设一个 QP 求解认为，一个超像素 i 和 j 获得分配的候选类的标签 $i' = \text{“草”}$ 和 $j' = \text{“天空”}$ ，其中 i 位于图像的顶部， j 在一个底部。由于常识性的知识排除了草地出现在天空上方的自然场景，如果 i 被分配的标签为 $i' = \text{“草”}$ ，即 $x_{ii'} = 1$ ，则 j 不能被指定为标签 $j' = \text{“天空”}$ ，即 QP 求解器必须设置 $x_{jj'} = 0$ 。这种类型的推理可以作为正式的等式约束： $x_{ii'} \cdot x_{jj'} = 0$ 。直观地看，这种等式约束严格迫使两个标签只有其中之一被允许用于两个超像素。

继[21]的方法，所有的互斥约束可以紧凑地表示为

$$x^T M x = 0, \quad (3)$$

其中 M 是第 $(n \cdot k) \times (n \cdot k)$ 的 二元约束矩阵。当它的元素被设置为 1，

$M_{(ii'),(jj')} = 1$ ，则对应标签的分配被禁止为 $x_{ii'} \cdot 1 \cdot x_{jj'} = 0$ 。相反，当 $M_{(ii'),(jj')} = 0$ ，那么超像素 i 和 j 可被指派给 L 任意类的标签，因为二次等式约束仍然满足， $x_{ii'} \cdot 0 \cdot x_{jj'} = 0$ 。注意 M 是典型的稀疏矩阵。SEC. 6 指定每个图像的 M 。

另外，可以很方便地合并该组 (2) 中的问题的线性约束 - 即，对于所有的 $i \in V$ ， $\sum_{i' \in L} x_{ii'} = 1$ - 包含 (3) 式的二次等式约束。对于每一个超像素 i ，我们设置矩阵 M 的所有相应元素为一个值，即 $M_{(ii'),(ij')} = 1$ ，如果 $i' \neq j'$ 。这样做禁止了多个不同的标签非法分配到一个超像素 i 的行为，因为对于所有的 i ，我们有 $x_{ii'} \cdot 1 \cdot x_{ij'} = 0$ ，如果 $i' \neq j'$ 。通过使用 (2) 中的关联矩阵 Φ 和约束矩阵 M ，我们终于得出以下 QP：

$$\begin{aligned} \max. & x^T \Phi x \\ \text{s.t.} & x^T M x = 0, \quad x \in \{0, 1\}^{n \cdot k} \end{aligned} \quad (4)$$

注意，(4) 式不放松原来的问题 (2) 中。而 (2) 和 (4) 中的约束条件是不等价的，(4) 式中的约束和目标导

致了 (2) 中目标和约束的等价问题。

(2) 式中的约束条件使得每一个超像素仅分配一个标签，而 (4) 式中的约束条件只保证每一个超像素不会被分配多个标签。但是，如果一个超像素 1 是未标记的目标 (4) 式的目标将不会式最大值。

在下面的部分，我们将提到我们求解 (4) 式中的 QP 问题的新算法。

5. 束搜索

给定一幅图像和它的超像素，BS 从它们的初始标记 X_0 开始，用新的状态 X 逐步构造搜索树。在每一个树的深度处，根据这些状态的启发式值，BS 只考虑对最佳状态作进一步的探索，探索由从选定的父状态随机抽样得到的后继状态组成，逐渐增加当前的树深度。当没有后继状态给予积极的启发式值时，连续的树扩展停止。下面，我们正式确定我们的搜索架构的元素。

状态空间： 该状态空间被定义为 $\Omega = \{X: X \in \{0, 1\}^{nk}, x^T M X = 0\}$ 。其状态对应于遵守互斥约束的候选场景标记。

后继函数： $\Gamma: X \rightarrow X'$ ，从 x 状态生成新的 X' 状态。 Γ 通过共同改变 X 的一组超像素的标签修改一个给定的状态 x ，使得 X' 严格满足互斥约束。因此，

Γ 定义: i) 如何选择一组要被重新标记的超像素; ii) 如何确定它们的新标签，如下面所解释，图 2 演示了产生一个新的状态的例子。

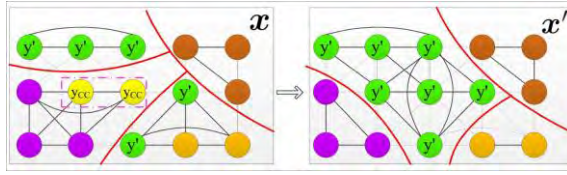
为选择一组超像素，我们首先削减 E 中邻接势能低于某个随机阈值的边。具体地讲，利用邻接势能 $\phi_{ii'jj'}$ 对边 $(i,j) \in E$ 进行表征，其中 i 和 j 被分配为状态 x 的标记 i' 和 j' 。

阈值是在 $\phi_{ii'jj'}$ 的最小值和最大值之间随机选择的值，用以剪掉所有邻接势能小于阈值的边。这样划分 G 成形一系列非连通子图。然后，我们随机选择的子图中的一个，在子图内，再次随机选择那些相邻并且具有相同的标签的超像素的一个连通分量 (CC)。

遵守对象的空间范围，我们共同重新标记选定 CC 中的每个超像素为选定的子图的相邻连通分量的标签之一。这促进了空间平滑度，并去除物体内的孔中得到的场景的标签。

后继状态 X' 将被接受仅当互斥约束能够被满足，即， $x'^T M x' = 0$ 。我们有效地计算这个二次方程如下：对选定的 CC 节点的重标记并不改变 x 整体，仅改变它的一部分向量。

我们用 $\delta = x' - x$ 表示差值，只有对属于 CC 的一部分节点 δ 非零。



图二。考虑状态 x 的显示标识。经过随机切割边，我们有一系列由红色的曲线分割的子图集。我们随机子图中随机选择一个带有标记 y_{cc} 的 CC 更新标签是随机选择作为相邻的 CC (Y' 在这种情况下) 的标签从而导致新的状态 x' 。需要注意的是 CRF 边由于较小的邻接势能，所以子边界并不意味着对象类的边界。

然后，该互斥约束对于 x' 可以被表示为

$$x'^T M x' = (x + \delta)^T M (x + \delta) = 2\delta^T M x + \delta^T M \delta = 0, \quad (5)$$

因为它已经保证了父状态满足互斥约束， $x^T M x = 0$ 。验证 (5) 式的复杂度低，因为 CC 将通常包括仅几个节点， M 是稀疏的。这一步可以保证通过 BS 找到的最终的解决方案严格满足所有互斥约束。随机生成后继状态有助于避免局部最优解。

启发函数: $H(x', x)$ ，给定它们的父状态 x 计算新的状态 x' ，并且通过选择 x 最好的后继者引导搜索树的扩张。理想的情况下，新的状态应该使用 QP 的目标， $x'^T \Phi x'$ ，如 (4) 式所述，进行评估。这将确保 BS 被引导向 QP 的最佳解决方案。但是，对每一个候选状态计算较大的 CRF 二次目标将是非常昂贵的。为了解决复杂度问题，我们重复使用 x' 和 x 的差值向量 δ 表

示 QP 的目标如下：

$$x'^T \Phi x' = x^T \Phi x + 2\delta^T \Phi x + \delta^T \Phi \delta. \quad (6)$$

对于所有新的状态 x' ，我们注意到 (6) 式具有相同的第一项， $x^T \Phi x$ 。幸运的是，计算 (6) 式中的其他两项计算量都不大，因为它们仅占 CC 的几个节点。着激励我们将启发函数定义为：

$$H(x', x) = 2\delta^T \Phi x + \delta^T \Phi \delta \quad (7)$$

一个全局启发式函数可以更好地评估候选状态，相对于我们的算法增加了计算复杂度。

选择 B 的最佳后继的策略就是持续生成 $x' = \Gamma(x)$ ，直到我们得到新状态 B，能够同时满足互斥约束并产生积极的启发式， $H(x', x) > 0$ ，正如我们下面证明，后者要求确保后继必须使 QP 目标单调增加。当经过足够长的运行时间仍没有能够同时满足上述两个条件的后继者，则 BS 算法停止。

评分函数: $S(x_l)$ ，通过合计连接根节点 x_0 (即初始状态) 到叶节点 x_l 的路径 $\{x_0 \dots x_t \dots x_l\}$ 中可用的启发值 $H(x', x)$ ，来有效计算叶状态 B2：

$$\begin{aligned} \max. & \quad x^T \Phi x \\ \text{s.t.} & \quad x^T M x = 0, \quad x \in \{0, 1\}^{n-k}. \end{aligned} \quad (4)$$

再次注意对所有的叶节点，(8) 式中的第一项只能计算一次。最后的场景标记方案将会从叶子集合 $x^* = \arg \max_{x_l} S(x_l)$ 中选择。

研究明确表明, (8) 式中的最大化的分 S 等价于 (4) 式中指定的最优化 QP 目标。因此, BS 单调增加了 QP 目标受到的由 (4) 带来的互斥约束。(4) 式同时也是对所有的叶状态 $x^* = \arg \max_{x \in S} S(x)$ 进行评分后可得到解决方案的方程。

由 (8) 式可得, 我们的算法复杂度为 $O(n \cdot k) + B \cdot L \cdot O((n \cdot k) + n^2)$ 。第一项来自带有 k 个标记的 n 个超像素的初始标记。第二项来自对 L 搜索水平上的 B 状态的后续产生和评估。为产生新状态择取一个 CC 需要 $O(n^2)$ 的计算量以计算 CRF 的 n^2 条边, 验证候选状态满足互斥约束并具有积极的启发值的复杂度为 $O(N \times K)$ 。值得注意的是, 随着标签的设置 BS 的复杂度呈现性增长。值得一提的是 BS 可以很容易并行化。这种并行处理了提供显著加速致平均收敛时间低于每幅图像一秒 (在带有 8 GB 内存的英特尔酷睿 i7 机器上)。

6. CRF 势能和互斥矩阵

本节首先介绍了超像素分割方法, 然后我们指定 CRF 的单元和邻接潜能。在此之后, 它描述了如何计算潜能, 最后指定如何估计互斥矩阵 M。

超像素分割 为与现有的技术状态公平比较, 我们用同样的方法提取

超像素就如在相关著作 [28, 9, 13, 34, 15] 中的使用, 即低级别的分割算法 [4]。

单元势能 被定义为质地, 颜色和位置势能的总和。 $\phi_{ii'} = \phi_{ii'}^{\text{tex}} + \phi_{ii'}^{\text{col}} + \phi_{ii'}^{\text{loc}}$ 。 $\phi_{ii'}^{\text{tex}}$ 被指定为增强分类的置信度, 其中每个弱分类器是基于一个多类逻辑回归的纹理特征的决断类型。对于每一个超像素 i 的纹理特征, 我们使用的高斯的 17 维滤波器组和

Laplacians-of-Gaussian 滤波器响应就像 [28]。 $\phi_{ii'}^{\text{col}}$ 被计算为超像素 i 的 16×3 色彩直方图的负对数混合高斯。被定义为类 i' 的负对数 (即, 频率) 出现在 i 的归一化的位置。我们使用分段训练方法 [28] 分别学习每个势能。

邻接势能 被定义为色彩平滑度和距离势能的总和, 就如 [34]。两个超像素 i 和 j 之间的色彩邻接势能这样计算: 当 $i' = j'$, $\phi_{ii'jj'}^{\text{col}} = g(I_i - I_j)$; 否则, 为 0; g 是一个带有恒等协方差矩阵的负高斯对数。 I_i 和 I_j 是超像素 i 和 j 的色彩直方图。距离势能被定义为:

$\phi_{ii'jj'}^{\text{dis}} = g(\mathbf{s}_i - \mathbf{s}_j)$, if $i' = j'$; 否则为 0,

\mathbf{s}_i 和 \mathbf{s}_j 是超像素 i 和 j 的位置。

互斥估计. 为了确定互斥矩阵 M, 我们假定训练数据集非常大, 用对象类在特定空间布局共同出现的频率, 则该频率可利用训练数据直接估计得来。

具体来说，我们定义一个增强标签集合{(对象, 对象, 结构)}。对于结构的标签，我们使用四种定性空间关系：“左”，“右”，“上面”，“下面”。训练数据中的每对超像素的(I, J)，我们确定它们的结构标签，即，估计相对于 i 估计四个空间关系之一。然后，我们开始计算每对超像素(i,j)出现在四个结构中的真类标签 i^* 和 j^* 的个数。当一个新的图像被遇到，它的每一对超像素对(i,j)都是首次被分配得四个结构标签之一，然后矩阵 M 的所有相关元素将被设置为以下之一。如果对象类对(i',j')没有在超像素对(i,j)的空间结构的训练数据中被观测到， $M_{(i'),(j')}=1$ ；否则，设置为 0，即 $M_{(i'),(j')}=0$ 。

7. 结果

数据集名：我们基于四个基本数据集
评估 BC：MSRC 数据集[28]，斯坦福背景数据集(SBD)[8]和2009, 2010年的国际视觉竞赛[3]数据集，MSRC数据集由21个对象类的591个图像构成我们复制[28]的测评设置来得到训练的标准分割和测试图像，SBD数据集具有7个背景类和1个通用前景类的715幅图像我们遵循了[8]的5倍交叉验证实验装置，2009, 2010年的国际视觉竞赛数据包包含20个对象类的图

片。在这里，我们对训练图像进行训练，测试图像进行测试。如[20]和[15]所做，精确度如标准 VOC 测量[33]那样测量。

对于训练，我们计算一个超像素的先验标签作为其像素的主要先验类标签。对于测试，我们在像素级计算标签分配精确度。因为 BC 和最终结果的收敛时间取决于初始状态，我们初始化搜索到一个逻辑回归的结构化预测。我们用排名前50的逻辑回归预测作为多重初始状态，然后运行每个 BS，最后选择最佳解决方案。请注意，为了提高效率可以并行这些多重搜索。

在 BS 的每一个步骤中，当搜索树深度是增长的，我们从父状态的 2B 继承状态中选择最多的 B 作为最佳候选。总的来说，新状态的接受率严重偏离了搜索步骤，因为当产生新状态时，我们摒弃了那些不满足互斥约束或者没有一个积极的启发式数值的状态。

评价输入参数：我们评价如下输入参数：波束宽度 B 和初始状态数量。图 5 展示出起初精确度随着 B 增大而增大，但在 B=10 后饱和。我们注意到相同的效果可以用来改变初始状态数目。当我们增加 B，BS 保持较多有希望的候选状态，但在一定限度后（我们的情况是 10），光束被伪随体充

满，这不影响准确性但增加了运行时间。在我们的实验中，我们令 $B=10$ ，并设 BS 的初始数目为 50，有图 4, 5。当 BC 平行于光束，我们运行时间关于 B 和初始点有一个线形轮廓基线。

基准线:我们将 BS 和以下四个基线 B1-B4 比较，在 MSRC 数据集中完成和基线得比较。

B1. Swendsen-Wang cut (SW-cut): SW-cut 用 Metropolis-Hastings(MH)抽样算法解决了棘手的 CRF 推断问题。MH 从 CRF 的尾部 $P(X|G)$ 提取样本 x 来产生新状态，各状态之间的跳跃是可逆的，并且被一个建议分布 $Q(x \rightarrow x')$ 控制。这也为选择一个连通分量 (CC) 切断了 CRF 边缘，并且更新到一个随即标签。如果从均匀分布 $u(0, 1)$ 得到的接受率 α 满足

$$\alpha < \min\left\{1, \frac{Q(x' \rightarrow x) P(x'|G)}{Q(x \rightarrow x') P(x|G)}\right\}$$

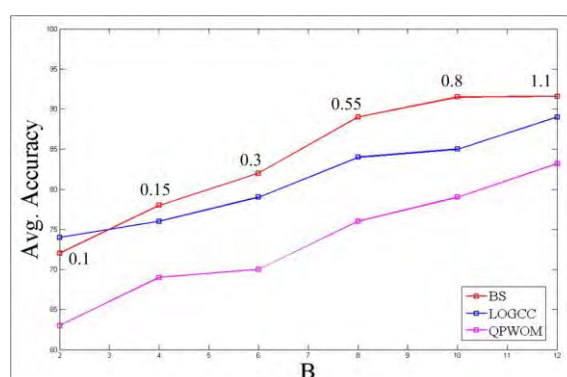
。这个建议将被接受，CRF 之后， $P(X|G)$ 如 (1) 所示被计算，并且该建议分布正比于 CC 选择期间被切割的边的数量，如 [37] 所示。公平起见，我们保持 CC 选择方法和我们的方法一样，81.5% 的精度比我们高了 30-32 秒运行时间的精确度还低了 10%。增加的精确度来源于搜索中的每一步都探索 B 状态而不是在 SW-cut 中只搜索一个状态。

B2. 没有互斥约束的 QP: 在这个底

线中，当我们执行推断时我们排除硬性互斥约束，只保留不允许像素有多个标签的约束。这是为有互斥约束来引导搜索的重要性辩护 (如图 4)。图 3 展示了一个例子，其中 QpwOM 造成了一个不可行的标签：标有“天空”的超像素在标有船的超像素底下，在 BS 中由于硬性互斥约束，这样的标签被严格禁止。



图三，从 MSRC 数据集上提取的图像与 QPM 图像和 BS 图像做对比。标签行不通是因为 QPM 缺少互斥限制。



图四，与 MSRC 数据集的基准作对比。BS 搜索是我们提出的方法，LOGCC 是 B4，这里的 CC 节点是经过逻辑学习更新后的方法，而且 QPWOM 是我们做的没有互斥限制的方法。我们还表示出了 BS 中每个 B 的运行时间。

B3. 有不严格的互斥约束的 QP. 这里，我们用一个标准的 QP 求解器 (IBM CPLEX 优化器)，旨在解决这个最优化问题: $x^* = \arg \max_x T(\Phi - \gamma M)x (\gamma \geq \max_i P_i \Phi_{ij})$, 此处放宽整数限制，只要 $x \in [0, 1]^{nk}$, 这使得互斥作为一个软

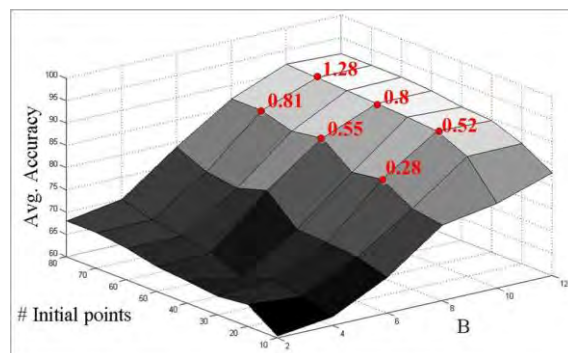
性约束，精度为 85.4%，运行时间 22 秒，大约比我们最高运行时间的精度低了 6%。和 B3 的比较展现了在离散域中解决优化问题比在宽松的相应域中更有效。

B4. 用逻辑回归更新 CC (LOGCC) :

这里不是从邻近超像素中选择一个 CC 的更新标签，我们用多类逻辑回归学习更新了 CC 的标签，因此，每个 CC 的节点被分配具有最高等级可能性的标签。可能性由逻辑回归分类测量。我们注意到，性能不与追加学习一同提高。

现有科技比较 Tab1. 展示了在 MSRC 数据集中与现有科学方法得比较。其中，我们的精确度比之前最好方法提高了 4.5%，那个方法使用了 QP 松弛作为全连通 CRF 模型像素的推断。SBD 和 09, 10PASCAL 数据集的比较在 Tab2 中显示：在 SBD 中，我们地方法比两个现有方法略逊色，不同于我们，他使用了多尺度段和高阶势能。在 2009 年 PASCAL 数据集，这个方法在 [7] 和 [20] 中展现的两个修正版本达到了由于我们的性能。这些方法用目标分割或前景分割作为附加的线索，但我们没有用这种线索。图 6 提出了我们的方法在 4 个数据集上的定性结果，图 7 展示了 BS 的失败案例，

这是一个来自 PASCAL09 数据集的一个图像对象类” 瓦斯瓶” 和对象类 “瓶子” 产生了混淆，并且由于阴影的存在人身体的背影部分也没被检测到。



图五，在 MSRC 数据集上输入评价参数 (B 和初始启动点的数字)。我们在 X 轴改变 B，而且令初始启动点的在 Y 轴上。为了特殊化 B 和初始启动点的数字还展示了运行时间。

Method	MSRC	Test time
[6]	70.0	N/A
[8]	76.4	N/A
[23]	82.9	30-32s
[15]	86.0	0.2s
[33]	86.5	N/A
[34]	87.0	N/A
Ours	91.5	0.8 s

表一。在像素分类精确度 (%) 和每幅图的计算时间 (s) 方面当前先进技术和在 MSRC 数据集上的比较

Method	SBD	Method	P '09	Method	P '10
[8]	76.4	[20]	37.2	[33]	31.2
[22]	76.9	[7]	34.1	[15]	30.2
[29]	74.1	[23]	35.7	[7]	40.1
[19]	81.9	Ours	40.1	[20]	39.7
[27]	82.9			Ours	34.2
Ours	81.1				

表二。在 SBD (左)，PASCAL VOC 2009 (中)，2010

(right) 数据集中最先进的分割精度准确性的对比 (%)。



图七. 来自 PASCAL 09 数据集的一个失败案例。GT 表示真实数据。

8. 结论

我们为场景标记提出了一种新的方法。场景标签作为一个全连通 CRF 的 MAP 分配被提出。MAP 分配制定二次规划，并使用我们的新束搜索 (BS) 算法求解。BS 用以下三个函数来构造搜索树，其搜索状态对应于候选的场景标记。后继函数负责从父状态的子集中产生后继状态；启发函数负责评估并从探索中选择最好的状态 B。评分函数寻找可证我们的 CRF 推论的 QP 目标是最大化的叶节点。BS 是一种能够很好的适配场景标记的算法，因为：解决了 QP 在离散域严格遵守有用域约束，并且具有较低的计算成本，允许有大量节点或全连通的 CRF。

我们的实验评价表明，BS 在某些标准数据集中（例如，MSRC）优于现有技术，同时在其他数据集中表现出竞争性能（例如，斯坦福背景）。此外，当我们从领域知识的不一致性

考虑，相对于忽略这种约束的方法，这样性能提高了 9%。在 MSRC 数据集上。有趣的是，用类标签的预测初始化 BS 比使用随机类标签初始化 BS 并没有显著的性能提升。BS 在计算上是高效的，并且也可以很容易地实施并行。

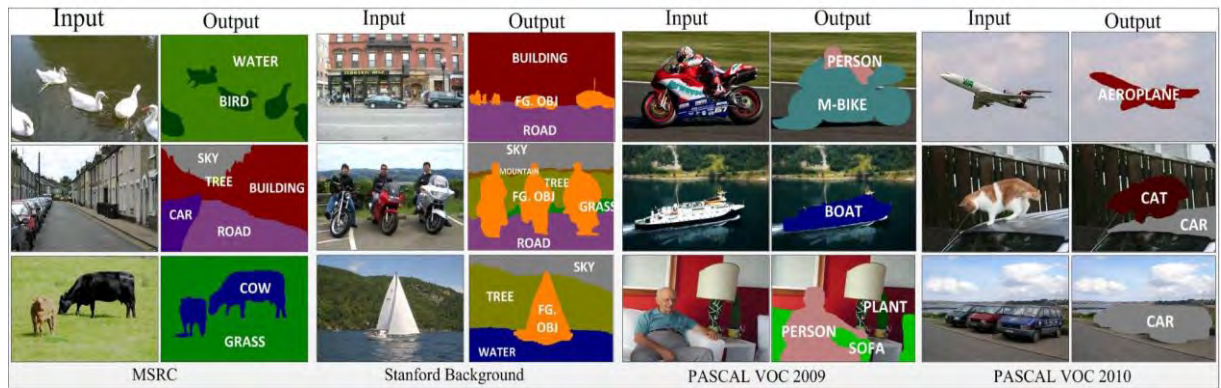
声明

本项工作由美国国家科学基金会资金 RI1302700 赞助支持。

参考

- [1] A. Barbu and S.-C. Zhu. Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities. *PAMI*, 27(8):1239–1253, 2005. 2
- [2] J. R. Doppa, A. Fern, and P. Tadepalli. HC-search: A learning framework for search-based structured prediction. *JAIR*, 2014. 2, 3
- [3] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *IJCV*, 88(2):303–338, 2010. 2, 5
- [4] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graphbased image segmentation. *IJCV*, 59(2):167–181, 2004. 5
- [5] P. F. Felzenszwalb and D. Mcallester. The generalized A* architecture. *JAIR*, 29, 2007. 2
- [6] C. Galleguillos, B. McFee, S. Belongie, and G. Lanckriet. Multi-class object localization by combining local contextual interactions. In *CVPR*, 2010. 7
- [7] J. M. Gonfaus, X. Boix, J. Van De Weijer, A. D. Bagdanov, J. Serrat, and J. Gonzalez. Harmony potentials for joint classification and segmentation. In *CVPR*, 2010. 6, 7

- [8] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and
- [17] A. Kumar, S. Zilberstein, and M. Toussaint. Messagepassing algorithms for MAP estimation



图六 6. 在 MSRC 上的定性结果, 斯坦福背景和 PASCAL 09, 10 数据集.

- [9] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. *IJCV*, 80(3):300–316, 2008. 1, 5
- [10] P. Gupta, D. Doermann, and D. DeMenthon. Beam search for feature selection in automatic SVM defect classification. In *ICPR*, 2002. 2
- [11] J. Jancsary, S. Nowozin, and C. Rother. Learning convex QP relaxations for structured prediction. In *ICML*, 2013. 1
- [12] J. Kappes and C. Schnörr. MAP-inference for highlyconnected graphs with DC-programming. In *Pattern Recognition*, pages 1–10. Springer, 2008. 1, 2
- [13] P. Kohli, L. Ladicky, and P. H. Torr. Robust higher order potentials for enforcing label consistency. *IJCV*, 82(3):302–324, 2009. 1, 5
- [14] I. Kokkinos. Rapid deformable object detection using dualtree branch-and-bound. In *NIPS*, 2011. 2
- [15] P. Krahenbühl and V. Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials. In *NIPS*, 2012. 1, 5, 7
- [16] A. Kumar and S. Zilberstein. Message-passing algorithms for quadratic programming formulations of MAP estimation. In *UAI*, 2011. 2
- using DC programming. In *AISTAT*, 2012. 1, 2
- [18] M. P. Kumar and D. Koller. Efficiently selecting regions for scene understanding. In *CVPR*, 2010. 1
- [19] V. S. Lempitsky, A. Vedaldi, and A. Zisserman. Pylon model for semantic segmentation. In *NIPS*, 2011. 6, 7
- [20] F. Li, J. Carreira, and C. Sminchisescu. Object recognition as ranking holistic figure-ground hypotheses. In *CVPR*, 2010. 5, 6, 7
- [21] T. Ma and L. J. Latecki. Maximum weight cliques with mutex constraints for video object segmentation. In *CVPR*, 2012. 3
- [22] D. Munoz, J. A. Bagnell, and M. Hebert. Stacked hierarchical labeling. In *ECCV*, 2010. 7
- [23] N. Payet and S. Todorovic. Hough forest random field for object recognition and segmentation. *PAMI*, 2012. 6, 7
- [24] N. Payet and S. Todorovic. SLEDGE: sequential labeling of image edges for boundary detection. *IJCV*, 104(1):15–37, 2013. 2
- [25] J. Rao Doppa, A. Fern, and P. Tadepalli. Structured prediction via output space search. *JMLR*, 15, 2014. 2, 3
- [26] P. Ravikumar and J. Lafferty. Quadratic programming relaxations for metric labeling and

- markov random field map estimation. In *ICML*, 2006. 1, 2
- [27] X. Ren, L. Bo, and D. Fox. RGB-(D) scene labeling: Features and algorithms. In *CVPR*, 2012. 6, 7
- [28] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *IJCV*, 81(1):2–23, 2009. 1, 2, 5
- [29] G. Singh and J. Kos̆ecka'. Nonparametric scene parsing with adaptive feature relevance and semantic context. In *CVPR*, 2013. 1, 7
- [30] A. Torralba, K. P. Murphy, and W. T. Freeman. Contextual models for object detection using boosted random fields. In *NIPS*, 2004. 1
- [31] J. Verbeek and W. Triggs. Scene segmentation with CRFs learned from partially labeled images. In *NIPS*, 2007. 1
- [32] L. Yang, P. Meer, and D. J. Foran. Multiple class segmentation using a unified framework over mean-shift patches. In *CVPR*, 2007. 1
- [33] J. Yao, S. Fidler, and R. Urtasun. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. In *CVPR*, 2012. 7
- [34] Y. Zhang and T. Chen. Efficient inference for fullyconnected CRFs with stationarity. In *CVPR*, 2012. 1, 5, 6, 7