

指导教师： 杨涛

提交时间： 2015.3.28

The task of  
**D**igital Image Processing

数字图像处理

School of Computer Science

No: \_\_\_\_\_

姓名： 朱皓晨

学号： 2012302475

班号： 10011204



# 最优决策的概率模型： intersection-over-union 例

Sebastian Nowozin

Microsoft Research

[Sebastian.Nowozin@microsoft.com](mailto:Sebastian.Nowozin@microsoft.com)

## 摘要

一个概率模型使我们有能力将世界合理化，并且能够根据统计理论，利用贝叶斯决策理论，做出最优决策。然而，在实践中的困难的决策问题迫使我们采取简单的损失函数为 0 / 1 的损失或汉明损失，因此结果就是，我们通过 MAP 估计或通过低阶的边际统计会做出错误的决定。在这次研究工作中，我们调查的最优决策是为了更现实损失函数。具体来说，我们认为当前流行的用于图像分割基准的交工会（IOU）评分测试表明，它的结果是硬盘的组合决策问题。为了使这一问题变得可解，我们提出了一个近似的统计目标函数，以及一个近似算法基于参数线性规划。我们应用三个基准数据集上的算法并且获得已改进的，与交叉口最大后验边缘的决定相比较的 intersection-over-union 得分，。我们的工作指出了将现实的损失函数与概率应用于计算机视觉模型的困难。

## 1. 简介

当前在计算机视觉任务一个普遍的观点是假定他们为一个概率推理任务。经典方法如下 [ 28, 15, 25 ]，如图 1 所示。指定一个概率模型  $P(X; Z)$  对  $Z$  和数量观测信号  $X$ ；2. 观察  $X$  和获得的条件后验分布  $P(Z|X)$  关注量的 3. 使用后验推断统计总结或决定。

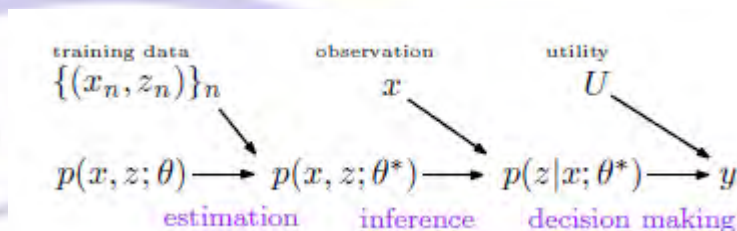


图 1。典型的工作流工作的概率时型号：1. 估计：使用训练数据来找到一个很好的模型在一个特定的模型类，2. 推理：推理后给定一个观察  $X$  的分布，3. 决策：使用后的信念和一个给定的效用做出最优决策。

在过去的十年中，这个方法已成功应用，甚至已经被优化。特别是，当前语义图像分割模型和人体姿态估计往往相互区别的有条件的，在一个高维的参数向量 [ 49, 30 ] 被指定为  $p(z|x; \theta)$ 。这种做法改变了之前方法的第一步，但第二、第三步骤保持不变。

大量的工作用于描述概率模型和复杂的推理过程，但决定性任务（步骤 3.）却是经常被忽视的。例如， [ 49 ] 简单的输出最有可能的状态  $\operatorname{argmax}_z P(Z|X)$  作为推理结果。确定最可能的状态取决于特定的损失函数的选择，即 0-1 损失，  $\ell(z, y) = 1_{\{z \neq y\}}$ 。这种损失是不现实的，并且不符合我们的预测质量评估计算机视觉的基准。我们的工作就是要解决这种不匹配；我们开始从决策理论 [ 2,

32 ]: 给我们的信念  $P(Z, X)$  和损失或效用函数决策效益的措施当真实的世界状态是  $Z$ , 我们倾向于要优化的决定, 那就是尽量减少预期损失或等价的预期效用最大化。在这项工作中, 我们考虑在当使用概率模型时, 也运用 intersection-over-union 效力进行最优决策实用 (IOU) 【11】。

另一种方法是直接学习一个决策函数的使用经验风险最小化方法 (ERM) [50], 这也是在计算机视觉结构 SVM 比较普遍的做法。然而, 保持一个概率模型有它自己的优势: 首先, 我们可以在测试时间时, 使用不同的损失函数; 第二, 概率方便结合独立的子模型, 每一个都可以分别设计和训练; 第三, 目前还没有一致的风险管理方法的结构预测。

### 1.1. 前人成就

以前的工作已经研究了经验风险最小化 (ERM) [50] 和高阶损失函数; 计算机视觉的第一份工作是 [4] 者如何学习在 intersection-over-union 损失限制在边界的交点盒。对于分割 Ranjbar 等。[35] 近似得出经验风险最小化 intersection-over-union 得分, 并为二进制的情况下, Tarlow 和泽梅尔 [44, 45] 提出有效的推理过程的损失-augmented inference 问题。雷克南 "ahenb" 乌尔和 Koltun 延伸 [10] [21] 处理高阶损失函数。Pletscher 和科利 [33] 和 Kuttel 和法拉利 [23] 研究发现, 除了高阶 IoU score, 其他具有功能高阶损失欠条评分功能的模型, 例如标签计数的损失及其逼近简单模型。这以前的工作着重与高阶学习的损失; 但是, 它们都使用经验风险的学习目标来进行学习。我们的工作不同之处就在于, 我们遵循“经典”的方法和第一个构造出一个概率模型, 然后利用任务高阶损失方式求解最优决策。在之前工作的研究中, Tarlow 和亚当斯 [43] 是唯一已经考虑过这个最优决策任务, 并且提出了一个贪心算法的研究者。

#### 贡献

我们的工作提出了以下的贡献。1. 我们推导出一个封闭形式的统计逼近方法, (8), 以 intersection-over-union 得分的情况下有条件独立的信念。2. 我们提出了一种算法, ALG。2, 在 intersection-over-union 得分下, 用于制造近似最优决策。3. 我们的实验评价对多个数据集的算法和展示增加了 intersection-over-union 分数。

## 2. 存在的问题及统计逼近

我们首先在考虑在不确定性 [2] 做出最优决策意味着什么。

**问题 1 (最优决策)**。给定一组  $Y = \{1, \dots, K\}$ , 一个指标集  $\mathcal{V} = \{1, \dots, n\}$ ,

一个效用函数  $U: \mathcal{Y}^{\mathcal{V}} \times \mathcal{Y}^{\mathcal{V}} \rightarrow \mathbb{R}$  和一个概率分布  $P$  超过设定  $\mathcal{Y}^{\mathcal{V}}$ , 找到最优决策

$$y^* = \operatorname{argmax}_{y \in \mathcal{Y}^{\mathcal{V}}} \mathbb{E}_{z \sim P} [U(z, y)].$$

直观地看, (1) 优化下的每一个我们的期望效用可能性  $Z$ , 加权我们的信念的状态世界编码页  $p$ 。这个问题可以推广通过允许更广泛的决策领域, 但对于大多数在计算机视觉 (1) 的应用是足够的。

解决 (1) 是困难的三个原因。首先, 它是一个在大集的大小呈指数增长的

优化问题在  $jV_j$ 。其次，期望的  $E_z P[?]$  需要计算的平均在同一大集。第三，的函数  $U$  可能没有足够的结构进行有效计算。在实践中的易处理性 (1) 依赖于我们是否可以替换为期望的表达简单的封闭形式的解。例如，在简单情况下，当  $U$  是它分解负海明损失相加过  $V$ ，以使期望的通勤和各变量可以单独优化，以获得最大后缘 (MPM) 解决方案 [25]。正如我们现在说明的一样，这不具有更复杂的实用功能。

### 2.1. Intersection-over-Union 效用

intersection-over-union 得分是流行的基准得分语义分割。由于 PASCAL 流行 VOC 的分割挑战 [11]，它在计算机视觉界已成为流行。它被定义为如下。

**定义 2** (Intersection-over-Union 效用) 给定一个地面实况分配

$y \in \mathcal{Y}^V$  和一个估计  $z \in \mathcal{Y}^V$ ，那么 Intersection-over-Union 效用就是

$$U_{iou} = \frac{1}{K} \sum_{k=1}^K U_{iou}^{(k)}(z, y)$$

$$U_{iou}^{(k)}(z, y) = \frac{\sum_{i \in \mathcal{V}} 1_{\{z_i=k \wedge y_i=k\}}}{\sum_{i \in \mathcal{V}} 1_{\{z_i=k \vee y_i=k\}}}$$

当 是 per-class intersection-over-union 效用，并且  $1_{\{\text{predicate}\}}$  是指示符函数是 1 的情况下的谓词是真实的，否则为 0。

对于 PASCAL VOC 分割基准设定， $V$  是集所有测试集图像的所有像素。因此 (2) 包含求和以上的所有像素的比率和不分解。这种特性使 (1) 难以解决。特别是，我们得到 (1) 为以下专业化 intersection-over-union 效用。

对于 PASCAL VOC 分割基准设定， $V$  是集所有测试集图像的所有像素。因此 (2) 包含求和以上的所有像素的比率和不分解。这种特性使 (1) 难以解决。特别是，我们得到 (1) 为以下专业化 intersection-over-union 效用。

$$\mathbb{E}_{z \sim p}[U_{iou}(z, y)] = \frac{1}{K} \sum_{k=1}^K \mathbb{E}_z \left[ \frac{\sum_{i \in \mathcal{V}} 1_{\{z_i=k \wedge y_i=k\}}}{\sum_{i \in \mathcal{V}} 1_{\{z_i=k \vee y_i=k\}}} \right]. \quad (3)$$

方程式 (3) 中含有的比率，期望这不具有闭合形式解。我们现在展示如何这种期望可以使用的技术从近似渐近的统计数据。特别是，根据 Rice，我们有以下结果 [36, 37]。

**主张 3** 令  $S$  和  $T$  两个实值随机变量的任意时刻排序，并与有限时刻  $P(T=0) = 0$  并且  $ET$  不为 0，那么当

$$\Psi_j = (-1)^j \frac{(\mathbb{E}S) \langle^j T \rangle + \langle S, ^j T \rangle}{(\mathbb{E}T)^{j+1}}$$

就有

$$\mathbb{E} \left[ \frac{S}{T} \right] = \frac{\mathbb{E}S}{\mathbb{E}T} + \sum_{j=1}^{\infty} \Psi_j$$

在此结果的应用的无限总和常常截断。由 HEIJMANS [17] 给出，类似但不那么一般的结果中讨论 [39] 和 [42, 第 4.10 节]，和预期的讨论比率。为适用命题 3

我们的期望 (3) 我们定义, 对于每一个  $k \in \mathcal{Y}$ , 有  $S_k = \sum_{i \in \mathcal{V}} 1_{\{z_i=k \wedge y_i=k\}}$  及  $T_k = \sum_{i \in \mathcal{V}} 1_{\{z_i=k \vee y_i=k\}}$ , 那么  $\mathbb{E}[U_{\text{iou}}^{(k)}(z, y)] = \mathbb{E}[S_k/T_k]$ . 我们更愿意将这个式子扩展为  $\mathbb{E}[S_k/T_k] = \mathbb{E}[S_k]/\mathbb{E}[T_k] + \sum_{j=1}^{\infty} \Psi_j$ , 因此我们需要计算  $\mathbb{E}[S_k]$ ,  $\mathbb{E}[T_k]$  和  $\Psi_j$ .

对于一般的 P 这些预期没有简单封闭形式的解决方案, 但如果我们假设条件 *i.e.*  $p(z|x) = \prod_{i \in \mathcal{V}} p_i(z_i|x)$  独立, 就可以取得进展, 这里  $x$  是一个带观测的量. 假设有条件的独立性在许多计算机视觉概率模型是一个自然的建模假设, 并不意味着无条件之间的独立性不同的. 例如, 有条件的独立性摆出使用随机森林时, [40] 预测像素边缘人的语义分割. 在这种假设下我们有以下的结果, 同样也来自 [43, 21].

**命题 4.** 对于 SK 和传统知识的上述定义  $S_k$  和  $T_k$  的条件独立分布  $p(z) = \prod_{i \in \mathcal{V}} p_i(z_i)$  圆周率 (ZI), 我们有, 作为判定  $y$  的函数,

$$\begin{aligned} \mathbb{E}[S_k] &= \sum_{i \in \mathcal{V}} p_i(k) 1_{\{y_i=k\}}, \\ \mathbb{E}[T_k] &= \sum_{i \in \mathcal{V}} (1_{\{y_i=k\}} + p_i(k) 1_{\{y_i \neq k\}}). \end{aligned}$$

我们给了补充材料证明. 我们现在有简单的封闭形式表达  $\mathbb{E}[S_k]$  和  $\mathbb{E}[T_k]$ . 让我们来看看额外扩展条款  $\Psi_j$  对于第一项  $\Psi_1$ , 我们有以下结果

**命题 5.** 对于 intersection-over-union utility  $\mathbb{E}[S_k/T_k]$  我们有  $\Psi_1 = 0$  的扩展 (4). 我们给了补充材料证明. 这是能够分析高阶项  $\Psi_j, j \geq 2$ , 但我们在这里停止, 因为命题 5 已经蕴涵着强烈的保证在近似的质量. 特别是通过标准的结果三角洲方法, [42, 第 4.3 节] 我们有渐近的  $n \rightarrow \infty$ , 那么

$$\mathbb{E} \left[ \frac{S_k}{T_k} \right] = \frac{\mathbb{E}S_k}{\mathbb{E}T_k} + O(n^{-1}).$$

连同 (6) 和 (7) 这会产生一个封闭形式的近似到  $\mathbb{E}[S_k/T_k]$ . 注意,

在一个典型的计算机视觉应用  $n$  将被大 ( $n \gg 10^3$ ), 因此该逼近误差会很小。这种保证也解释这种近似的所用的经验成功按[43, 21], 其中近似推导启发式并描述为“代孕”和“放松”。我们现在检查的保证 (8) 实验。

### 2.2. 逼近实验验证

我们执行下列模拟实验。为每个  $N = 5000$  二元变量, 我们品尝概率与之前一致的参数从狄氏分布  $P_i$ ,  $\alpha = 0.2$ 。然后, 我们选择  $n \in \{100, 200, \dots, N\}$  和执行  $E[S_1 / T_1]$ ,  $E[S_1]$  和  $E[T_1]$  的蒙特卡罗评估, 其中期望是通过模拟评价  $y_i \sim P_i$  和  $z_i \sim P_i$  十万次得到的。然后, 我们为  $E$  比较  $E[S_1 / T_1]$  和  $[S_1] / E[T_1]$  和情节  $n$  反对在对数 - 对数曲线图的错误, 请参阅图 2。

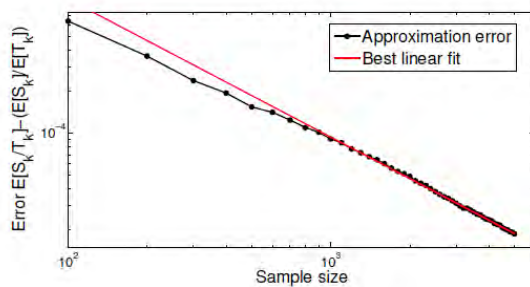


图 2. 逼近正 (8) 作为函数的错误。该结果是使用蒙特卡洛模拟得到的。误差的预测为  $O(n^{-1})$  显然可见与嵌合斜率系数是  $-0.9954 \approx -1$ 。

因为理论预测的误差  $O(n^{-1})$  行为我们拟合线到误差观测的后半部分。该直线的斜率同  $O(n^{-1})$  错误一致

### 3. 方法

我们现在发展作出最优决策的方法 intersection-over-union 的效用函数。通过专业 (1) 以交叉点切换愈合效用, 并通过使用近似 (8) 与封闭形式解 (6) 及 (7), 我们得到以下的问题, 如在[43]。

问题 6 (近似最优决策下, 该法 intersection-over-union 效用)。鉴于边际信仰  $PI$  用  $K$  类, 找到预测通过求解

$$\begin{aligned} \max_{\lambda} \quad & \frac{1}{K} \sum_{k=1}^K \frac{\sum_{i \in \mathcal{V}} p_i(k) \lambda_{i,k}}{\sum_{i \in \mathcal{V}} [p_i(k) + (1 - p_i(k)) \lambda_{i,k}]} \\ \text{sb.t.} \quad & \sum_{k=1}^K \lambda_{i,k} = 1, \quad i \in \mathcal{V}, \\ & \lambda_{i,k} \in \{0, 1\}, \quad i \in \mathcal{V}, \quad k \in \mathcal{Y}. \end{aligned}$$

在这里,  $\lambda_{i,k}$  为指针变量, 选择每个  $y_i$  为一个  $\lambda_{i,k} = 1$  使

$y_i = k$ 。问题 6 的仿射函数比的总和。比的数等于类的问题数  $K$ 。此的优化问题类型是已知的, 在优化社区由不同的名称: 总和-比的线性分式程序[38], 多比双曲线 0-1 编程问题[46], 和广义线性分式计划[6]。对于二元变量这些问题被称为是 NP-hard 的一般即使对于一个单一比例。但是, 如果我们放宽约束,  $\lambda_{i,k} \in \{0, 1\}$  向区间  $\lambda_{i,k} \in [0, 1]$  已知的结果令人惊讶: 只是一个包含总和单比例在多项式时间的 Charnes-库珀变换[7]中, 两个比率的总和也可解在多项式时间[20]和已知是拟凸限制性条件[6, 第 7.5 节]说下做并不适用于我们的情况。对于三套及以上的比例问题, 即使是在更广泛的领域[38]仍然很难。

一个明智的方法来解决 (9) 中提出的问题, [43]其中, 作者运用了大量的本地搜索, 反复一次变化一个变量, 以改善目标。该方法实现起来很简单, 高效, 并保持一个可行的解决方案在所有时间; 我们形容它在补充材料。然而, 在许多优化问题 - 例如离散能量最小化问题[18] - such 一个简单的贪婪方法可以跑赢通过该方法利用问题的结构。这是一般的情况下, 当该问题的特性复杂, 从而导致局部最优, 这时贪婪方法可能会滞留不前, 在这种情况下, 我们的建议的方法是使用问题的结构更加全球化, 但立足于解决放宽尺度, 因此, 可能会导致非整数解。问题是哪种方法是最好的, 实际上是内在复杂的客观的经验问题 (9)。如果它足够简单的贪婪方法可以更好地工作, 相反, 如果它是复杂的, 我们可以看到更多的全球性的方法来更好地工作。

我们的方法来解决 (9) 将是反复改进一个解  $y$  通过优化过的相应的变量块进行标记, 保持其他变量固定不变。这样产生了一个易于解决的子问题, 其中只有两个分数出现在 (9), 我们使用 Konno 的算法[20], 这样就解决了相应的宽松问题。

### 3.1. 优化两种类型

策略优化过的变数相当大的一部分根据当前的候选解决方案已经在使用过 - ? - swap graphcut 算法[5]和也是用于解决一些难的组合问题在超大规模邻域搜索的框架 (大规模邻域) [1], [30, 第 4.5.1 节。它适用于 (8), 我们得到下面的子问题仅限于两种。

**问题 7**(两类问题) 对于两个不同的  $k_1, k_2 \in \mathcal{Y}$ , 和一个基准标记

$y \in \mathcal{Y}^{\mathcal{V}}$ , 令

$\mathcal{W} \subseteq \mathcal{W}_{k_1 k_2} = \{i \in \mathcal{V} : y_i = k_1 \vee y_i = k_2\}, \mathcal{W} \neq \emptyset$ , 就是

说, 该变量的当前的任意非空子集标  $K_1$  或  $K_2$ 。找到限制在最佳标记集  $\mathcal{W}$ , 我们

令  $\lambda_{i,k_1} = \mu_i, \lambda_{i,k_2} = 1 - \mu_i$ ; 并解决

$$\begin{aligned} \max_{\mu} \quad & \frac{a_0 + \sum_{i \in \mathcal{W}} a_i \mu_i}{b_0 + \sum_{i \in \mathcal{W}} b_i \mu_i} + \frac{c_0 + \sum_{i \in \mathcal{W}} c_i \mu_i}{d_0 + \sum_{i \in \mathcal{W}} d_i \mu_i}, \\ \text{sb.t.} \quad & \mu_i \in \{0, 1\}, \quad i \in \mathcal{W}, \end{aligned}$$

，其中涉及到的  $k_1$  和

$$\begin{aligned} a_0 &= \sum_{i \in \mathcal{V} \setminus \mathcal{W}} p_i(k_1) 1_{\{y_i=k_1\}}, \\ a_i &= p_i(k_1), \quad i \in \mathcal{W}, \\ b_0 &= \sum_{i \in \mathcal{V}} p_i(k_1) + \sum_{i \in \mathcal{V} \setminus \mathcal{W}} (1 - p_i(k_1)) 1_{\{y_i=k_1\}}, \\ b_i &= 1 - p_i(k_1), \quad i \in \mathcal{W}, \\ c_0 &= \sum_{i \in \mathcal{V} \setminus \mathcal{W}} p_i(k_2) 1_{\{y_i=k_2\}} + \sum_{i \in \mathcal{W}} p_i(k_2), \end{aligned}$$

$k_2$ ，满足

$$\begin{aligned} c_i &= -p_i(k_2), \quad i \in \mathcal{W}, \\ d_0 &= |\mathcal{W}| + \sum_{i \in \mathcal{V} \setminus \mathcal{W}} (p_i(k_2) + (1 - p_i(k_2)) 1_{\{y_i=k_2\}}), \\ d_i &= p_i(k_2) - 1, \quad i \in \mathcal{W}. \end{aligned}$$

为了解决这个问题，我们调整相适应的算法。[20]。这个算法是基于第一个  $\mu_i \in [0, 1]$ ，然后进行变换 (10)，由 Charnes-的装置库珀变换[7]后跟一个参数单纯方法为线性规划[8]。该 Charnes - 库珀改造定义了一组 UI 变量和一个额外的变量  $u_0 = 1/(d_0 + \sum_{i \in \mathcal{W}} d_i \mu_i)$  和  $\mu_i u_0$ 。问题 7 现在可以在新的变量被改写作为一个比率的总和和一个线性函数如下。

$$\begin{aligned} \max_{u_0, u} \quad & \frac{a_0 u_0 + \sum_{i \in \mathcal{W}} a_i u_i}{b_0 u_0 + \sum_{i \in \mathcal{W}} b_i u_i} + c_0 u_0 + \sum_{i \in \mathcal{W}} c_i u_i, \\ \text{sb.t.} \quad & d_0 u_0 + \sum_{i \in \mathcal{W}} d_i u_i = 1, \\ & u_i \leq u_0, \quad i \in \mathcal{W}, \\ & u_0 \geq 0, \quad u_i \geq 0, \quad i \in \mathcal{W}. \end{aligned}$$

为了摆脱最终的比例，我们引入一个辅助参数  $\xi = b_0 u_0 + \sum_{i \in \mathcal{W}} b_i u_i$ ，并获得下列一系列参数的线性程序。



$$\begin{aligned}
 \mathcal{P}(\xi) = \frac{1}{\xi} \left\{ \begin{array}{l} \max_{u_0, u} \quad \xi c_0 u_0 + \xi \sum_{i \in W} c_i u_i + a_0 u_0 + \sum_{i \in W} a_i u_i, \\ \text{sb.t.} \quad b_0 u_0 + \sum_{i \in W} b_i u_i = \xi, \\ \quad \quad \quad d_0 u_0 + \sum_{i \in W} d_i u_i = 1, \\ \quad \quad \quad u_i \leq u_0, \quad i \in W, \\ \quad \quad \quad u_0 \geq 0, \quad u_i \geq 0, \quad i \in W \end{array} \right\}.
 \end{aligned}
 \tag{11}$$

Konno 等人的算法 [20] 是一个高效的参线性规划方法在全球范围内被最大化  $P$  在  $\xi$  取遍从最小可能值到最大可能值, 跟踪每个最优解  $\xi$  的价值。该解决方案的路径是二次序列在功能上对  $\xi$  的估计, 相比较于我们想象中的图一个典型的例子 3。相对于标准的参数编程在计算机视觉中使用线性规划[3, 第 5.5 节][19], 问题的  $\mathcal{P}(\xi)$  是因为在更具挑战性的的原因在于参数  $\xi$  无论是在目标函数, 还是如在约束 (11) 中出现, 都是一样的。因此, 如果我们增加  $\xi$ , 同时保持一个最佳的解决方案, 我们需要做的是对是原始的单纯更新或对偶单纯形更新[8]。

### 3.2. 方法执行

算法 1 描述的算法的主循环。目标函数 - 我们在图 3 取定一个值作为例子 - 其取值范围被定义在  $\xi_{\min}$  和  $\xi_{\max}$  之间, 并且由在二次函数  $\xi$  分段取值组成。该算法开始于图 3. 函数  $\mathcal{P}(\xi)$  典型例子, 由参数单纯形法在  $\xi \in [\xi_{\min}, \xi_{\max}]$  最大化。在这种情况下,  $\mathcal{P}(\xi)$  是分段二次, 由 441 段组成。

保证  $\xi = \xi_{\min}$ , 同时在每次迭代时增加  $\xi$  使得下一个时间间隔限定另一个二次函数达到 (行 12)。在每个时间间隔内, 我们得到了最接近的解决方案 (行 10) 和跟踪全局最优解的  $\xi^*$ 。我们将 (12) 变换到标准的广泛形式

$$u_i - u_0 + s_i = 0, \text{ 其中 } s_i \geq 0 \text{ 是一个变量。}$$

实现算法 1 是因为这两个充满挑战, 因为原始和对偶单纯形法需要实施[8, 3]。在考虑一个矩阵自由执行时, 可通过得到闭合形式来解决有关两个线性系统到限定的  $\mathcal{P}(\xi)$  的约束系统, 并通过使用非标准更新策略, 以降低成本[48]。我们的优化的 C++源代码可以从作者的主页获得。尽管用我们的优化的算法, 整

体运行复杂度依然为  $O(n^2)$ 。我们展示于图 4 典型的运行时间，并观察它们是经验独立系数。

### 3.3. 通过本地搜索优化 (9)

我们提出的方法是足够快用以优化预测为单个图像。不幸的是，因为当前在分割基准使用条件适用于整个测试数据集，而二次运行时禁止我们优化同时以上所有图像的决策变量。因此，我们把集中的所有变量为随机一个指定大小的子集。每个子集被优化顺序，并且，由于每个问题对应于一个围绕当前的解决方案附近，因此我们可以保证物镜的单调改善 (9)。整体过程示于算法 2，我们使用一个子问题大小  $M=2000$  的实验。对于一个子问题大小  $OFM=1$  算法 2 变成了 [43] 贪心方法。

#### Algorithm 2 Large Neighborhood Local Search for (9)

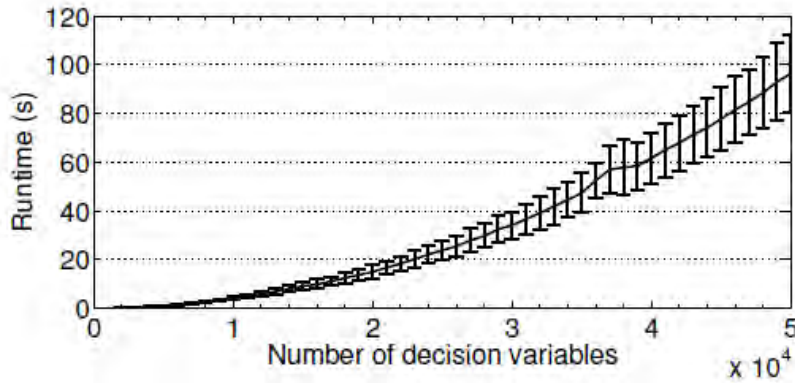
```

1: function OPTIMIZEIOU( $p, \mathcal{V}, K, \text{iters}, M$ )
2:    $y_i \leftarrow \operatorname{argmax}_{k=1, \dots, K} p_i(k)$  ▷ initialize with MAP
3:   for  $t = 1, \dots, \text{iters}$  do
4:      $S \leftarrow \text{RANDOMSHUFFLE}((1, 2, \dots, K))$ 
5:     for  $j = 1, \dots, \lfloor K/2 \rfloor$  do ▷ in parallel
6:        $k_1 \leftarrow S(2j-1), k_2 \leftarrow S(2j)$ 
7:        $\mathcal{W}_{k_1 k_2} \leftarrow \{i \in \mathcal{V} : y_i = k_1 \vee y_i = k_2\}$ 
8:        $\mathcal{W}_{k_1 k_2} \leftarrow \text{RANDOMSHUFFLE}(\mathcal{W}_{k_1 k_2})$ 
9:       for  $r = 0, 1, \dots, \lfloor |\mathcal{W}_{k_1 k_2}|/M \rfloor$  do
10:         $\mathcal{W} \leftarrow \mathcal{W}_{k_1 k_2}(rM+1, \dots, (r+1)M)$ 
11:        Compute  $a_0, a_i, b_0, b_i, c_0, c_i, d_0, d_i$ 
12:         $\mu \leftarrow \text{PARASIMPLEX}(a, b, c, d)$ 
13:        for  $i \in \mathcal{W}$  do
14:           $y_i \leftarrow k_1$  if  $\mu_i \geq \frac{1}{2}$ ,  $k_2$  otherwise
15:        end for
16:      end for
17:    end for
18:  end for
19:  return approximately optimal decision  $y$ 
20: end function

```

## 4. 实验及结果

现在，我们验证了重要贡献；也就是说，我们想要说明，给予相同的边缘后验我们的方法可以优化 intersection-over-union 得分。我们使用三个基准：贪心方法 [43]，简单图 4. 算法 1  $O(N^2)$ ，运行时间为  $n$  个与功能超过 10 重复与随机一个单位标准差选择问题系数。对于  $n=2000$ ，我们采用在算法 2 中，我们有 156ms 的运行时的单核英特尔至强 E5-16503. 20GHz CPU。



在我们当前情况下，MAP 决策 (RF-MAP / MPM)，自相同到最大后缘 (MPM) 的 [25]，和一个简单的“逆权”MAP 法 (MAP)。为了这方法，我们总结为累加某类所有变量  $k$ ，如  $\gamma_k = \sum_{i \in V} p_i(k)$ ，圆周率 ( $k$ ) 时，然后计算加权

$\tilde{p}_i(k) = (p_i(k)/\gamma_k)/(\sum_{\ell} p_i(\ell)/\gamma_{\ell})$ 。我们然后使用  $\sim P$  的 MAP 决策计算。根据我们的直觉是该重加权使我们的新的总信念在所有类均匀分布，因此可以给予更多的概率质量以先验的可能性较小的一类。对于我们的方法，我们还报告该信准确性和 iou 得分。该信 IOU 分数正是我们的目标 (9)，以及可信精度是

$(\sum_{i \in V} p_i(y_i))/|V|$ 。我们运行算法 230 和 60 次迭代获得近似最优决策 RF-T30 欠条和 RF-IOU-opt60 欠条效用下。

对于所有的实验，我们都随机获得边缘，施加到每个象素 [41]。我们使用的决定树实施的作者发布了 [31] 并扩大与简单的直方图 OF-使用的功能渐变，积分图像和图像位置的功能，否则，使用默认设置。我们使用三种语义场景分割数据集：LabelMeFacade [13]，斯坦福背景 [16]，和 PASCAL VOC2012 [11]。虽然我们降尺度图像尽可能使我们的决策是客观的，但是我们总是只能维持原始全分辨率图像评估 intersection-overunion 得分上准确性。

#### 4.1. LabelMeFacade 数据集

该 LabelMeFacade 数据集 [13] 采用 9 语义标签描述表面元素。有我们通过降尺度的 0.125 倍的 845 个测试图像，共 4429952 决策变量。随机基线结果表 1 可比的精度 (71: 28%) 同国家的的现有技术 (67: 33%)，但得到的改进的借条得分后优化 (IoU-opt60, IoU-greedy)。

Method	Acc. belief/actual	IoU belief/actual
RF [13]	- / 49.06	- / -
ICF [14]	- / 60.68	- / -
ICFHGWS+ [14]	- / 67.33	- / -
RF-MAP/MPM	78.69 / <b>71.28</b>	34.79 / 31.74
RF-iwMAP	55.30 / 53.76	32.35 / 32.01
RF-IoU-greedy [43]	75.75 / 69.60	<b>38.75 / 35.96</b>
RF-IoU-opt30	75.60 / 69.47	38.67 / 35.91
RF-IoU-opt60	75.72 / 69.56	38.70 / 35.91

Table 1. LabelMeFacade results (845 test images) [13].

## 4.2. 斯坦福大学的背景数据集

斯坦福大学的背景数据集[16]采用 8 语义类和 715 的图像;标准设置是 5 组交叉验证。我们生产的全分辨率边缘后验对每个五个测试数据为一组,然后按比例缩小的图像由 0.25 总计 3377600 决策变量的一个因素。结果列于表 2 中我们多类精度 (74: 7%) 与国家 of the-大致相当技术 (81: 9% [24])。在我们的结果 MAP / MPM 决定具有最高的准确度,并且 IoU-opt60/IoUgreedy 决定具有最高的 IOU 得分。

Method	Acc. belief/actual	IoU belief/actual
Gould et al. [16]	- / 76.4	- / -
Kumar/Koller [22]	- / 79.42	- / -
Lempitsky et al. [24]	- / <b>81.90</b>	- / -
Tighe/Lazebnik [47]	- / 77.5	- / -
Farabet et al. [12]	- / 81.4	- / -
RF-MAP/MPM	75.75 / <b>74.70</b>	50.42 / 50.51
RF-iwMAP	69.21 / 69.90	47.67 / 49.59
RF-IoU-greedy [43]	75.15 / 74.58	51.79 / <b>52.36</b>
RF-IoU-opt30	75.14 / 74.56	51.78 / 52.34
RF-IoU-opt60	75.14 / 74.58	51.79 / <b>52.36</b>

表 2 五组交叉验证结果 (715 图像) 的斯坦福大学的背景数据集[16]。

## 4.3. PASCAL VOC 2012 数据集

在 PASCAL VOC 语义分割基准[11]是一个包含 21 类数据集,具有挑战性的分割。我们培养了随机森林的“车”子集,产生的后为“VAL”,并且也培养在“trainval”子集并产生后的“测试”。我们的降尺度后路用因子或 0.125,得到 4104672 号决定对于“值”(1449 图像)和 4123073 决策变量变量“测试”(1456 图像)。其结果示于表 3 和表 4。对于这一具有挑战性的数据集,我们不能达到国家的最先进的性能;然而,上“VAL”借据分的 MAP / MPM 决策实现从 3 提高: 51%至 11: 08%。因为这两个决定从同一后部边缘人获得本增加的直接原因是我们的算法。与先前的数据集的最佳精度由实现 MAP / MPM 决定 (RF-MAP / MPM), 最好的 IoU 得分由 IOU 决策 (RF-IOU 贪婪算法)。

## 5. 讨论

结果明确表明,优化 iou 表现强于不清楚的效用函数大盘的方法。然而,简单的贪婪的本地搜索方法[43]是出奇的好,略优于我们的全球的方法。这是一个指示该物镜 (9) 是简单和不具有许多局部最优解。对于实际的目的之一,因此可以使用更有效的贪婪方法。注意,这行为是不同的东西在离散能量观察最小化问题[18]其中贪婪的方法,如迭代条件模式 (ICM) 定期优于通过更多的全球性的方法。要解释这一观察,需要进一步分析 (9)。

简单来说,我们的工作提出了一些问题的时候利用概率模型在计算机视觉和一些解决 iou 问题的具体点。

Method	Acc. belief/actual	IoU belief/actual
RF-MAP/MPM	80.47 / <b>73.33</b>	3.82 / 3.51
RF-iwMAP	21.00 / 24.61	4.09 / 6.68
RF-IoU-greedy [43]	68.43 / 69.17	7.60 / <b>11.65</b>
RF-IoU-opt30	65.98 / 66.95	7.20 / 10.68
RF-IoU-opt60	68.03 / 68.63	7.38 / 11.08

Table 3. PASCAL VOC 2012 validation set results (1449 images).

Method	Acc. belief/actual	IoU belief/actual
BONN O2PCPMC	— / —	— / 47.0
NUS	— / —	— / <b>47.3</b>
UVA CRF	— / —	— / 11.3
RF-MAP/MPM	84.25 / —	4.01 / 3.61
RF-iwMAP	23.35 / —	4.02 / 7.53
RF-IoU-greedy [43]	73.63 / —	7.67 / <b>12.47</b>
RF-IoU-opt30	71.27 / —	7.27 / 11.49
RF-IoU-opt60	72.92 / —	7.58 / 12.27

Table 4. PASCAL VOC 2012 test set results (1456 images).

高阶工具/损失的计算可追踪性功能。

我们假设条件独立信仰  $PI$  为每个变量  $i \in \mathcal{V}$ 。可以说这是一个强大的简化的假设。然而，即使有这样的假设的问题（问题 6）的决定仍然是一个很难组合优化问题。这是合理的假设对于更复杂的模型的任务将仍然难解。鉴于这种困难性，请注意，在最近的“边缘性学习”的框架[10]这是可以做到的基于梯度的参数优化 intersection-over-union 的效用就证明在[21]。虽然框架是基于经验风险最小化（ERM），它保留了概率推理方法作为内部组件。它可能是这种情况企业风险管理是更适合与高阶 tractably 学习丧失功能，因为最困难的部分处理的高阶的损失，可以在训练中进行处理，而测试时间预测仍然有效[45, 33]。

**校准概率的重要性。**

我们可以从实验的结果看出，我们认定的 intersection-over-union 分数偏离实际的分数。难道说我们我们通过优化决定（1）已经“过度拟合”？不，这是不可能的：我们已经有证实  $P$ ，我们只是根据我们的证实做出最好的决定。对于所观察到的说明偏差，我们的概率  $p$  的不完全校准。也就是说，我们可以相信，一个变量  $y_i$  是在一定的某些说明有一定的概率，但我们有系统高估或低估这种可能性。作为结果期望（1）不正确估计我们的决定的后果我们的决定，也同时能够观察到[43, 34]。

我们如何才能提高我们的校准概率？如果我们的概率模型类是足够准确的话，已知的结果保证我们最终会很好地校准[9]。然而，在现实的计算机视觉应用中，我们的模型是错误是显而易见的，甚至使用贝叶斯推理都是有系统性的失准[27]存在的。然后，更务实的方法来校准可以使用装袋和重新校准的方法，[29]。

### 抽样理论和分解。

在 VOC 分割挑战[11] iou 实用程序中，评估整个测试数据集的是混淆矩阵。我们不同意这种观点，该数据组是随机在互联网上取样，从一个人口众多的图像所创建的基础上设置。因此，观察一个单元是一个单一图像和效用函数应该沿着分解这些独立的单位。应用程序 iou 上的整个数据集是没有意义的：如果你这样做，你的决定特定图像可以潜在通过考虑改进未来的独立观测。另一方面将，应用程序 iou 推广至个人是意义重大的。

## 6. 结论

在这项工作中，我们研究了 intersection-over-union 得分，在图像中使用流行的高阶效用函数分割基准。与决策理论出发，我们提出了一个统计近似和算法使 iou 工具下近似最优决策。该实验证实并改善了之前的结果。我们希望今后能够进一步研究学习和最优决策正与高阶损耗函数相关问题。

### 致谢。

作者对丹尼 Tarlow 和克里斯托夫兰伯特进行的讨论和反馈表示由衷的感谢。

### 参考

- [1] R. K. Ahuja, O. Ergun, J. B. Orlin, and A. P. Punnen. A survey of very large-scale neighborhood search techniques. In Proc. Workshop on Discrete Opt. Elsevier, 2002.
- [2] J. O. Berger. Statistical Decision Theory and Bayesian Analysis. Springer, 1985.
- [3] D. Bertsimas and J. N. Tsitsiklis. Introduction to Linear Optimization. Athena Scientific, 1997.
- [4] M. B. Blaschko and C. H. Lampert. Learning to localize objects with structured output regression. In ECCV, 2008.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell., 23(11):1222 - 1239, 2001.
- [6] A. Cambini and L. Martein. Generalized Convexity and Optimization: Theory and Applications. Springer, 2008.
- [7] A. Charnes and W. W. Cooper. Programming with linear fractional functionals. Naval Research Logistics Quarterly, 9:181 - 186, 1962.
- [8] V. Chvátal. Linear Programming. WH Freeman, 1983.
- [9] A. P. Dawid. The well-calibrated Bayesian. Journal of the American Statistical Association, 77:605 - 613, 1982.
- [10] J. Domke. Learning graphical model parameters with approximate marginal inference. IEEE Trans. Pattern Anal. Mach. Intell., 35(10):2454 - 2467, 2013.
- [11] M. Everingham, L. V. Gool, C. K. Williams, J. Winn, and A. Zisserman. The PASCAL voc2012 challenge results.
- [12] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning

- hierarchical features for scene labeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1915 - 1929, 2013.
- [13] B. Frohlich, E. Rodner, and J. Denzler. A fast approach for pixelwise labeling of facade images. In *ICPR*, 2010.
- [14] B. Frohlich, E. Rodner, and J. Denzler. Semantic segmentation with millions of features: Integrating multiple cues in a combined random forest approach. In *ACCV*, 2012.
- [15] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721 - 741, 1984.
- [16] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and semantically consistent regions. In *ICCV*, 2009.
- [17] R. Heijmans. When does the expectation of a ratio equal the ratio of expectations? *Statistical Papers*, 40, 1999.
- [18] J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, and C. Rother. A comparative study of modern inference techniques for discrete energy minimization problems. In *CVPR*. IEEE, 2013.
- [19] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. In *ICCV*, 2007.
- [20] H. Konno, Y. Yajima, and T. Matsui. Parametric simplex algorithm for solving a special class of nonconvex minimization problems. *J. Glob. Opt.*, 1:65 - 81, 1991.
- [21] P. Krahenbühl and V. Koltun. Parameter learning and convergent inference for dense random fields. In *ICML*, 2013.
- [22] P. M. Kumar and D. Koller. Efficiently selecting regions for scene understanding. In *CVPR*, 2010.
- [23] D. Kuttel and V. Ferrari. Learning to approximate global shape priors for figure-ground segmentation. In *British Machine Vision Conference (BMVC)*, 2013.
- [24] V. S. Lempitsky, A. Vedaldi, and A. Zisserman. Pylon model for semantic segmentation. In *NIPS*, 2011.
- [25] J. L. Marroquin, S. K. Mitter, and T. A. Poggio. Probabilistic solutions of ill-posed problems in computational vision. *J. of the Am. Stat. Assoc.*, 82(397):293, 1987.
- [26] D. McAllester. Generalization bounds and consistency for structured labeling. In *Predicting Structured Data*. MIT Press, 2007.
- [27] U. K. Müller. Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5):1805 - 1849, 2013.
- [28] D. Mumford and A. Desolneux. *Pattern Theory: The*

- Stochastic Analysis of Real-World Signals. CRC, 2010.
- [29] A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. In ICML, 2005.
- [30] S. Nowozin and C. H. Lampert. Structured learning and prediction in computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(3-4):185 - 365, 2011.
- [31] S. Nowozin, C. Rother, S. Bagon, T. Sharp, B. Yao, and P. Kohli. Decision tree fields. In ICCV, 2011.
- [32] G. Parmigiani and L. Inoue. *Decision Theory: Principles and Approaches*. John Wiley and Sons, Ltd., 2009.
- [33] P. Pletscher and P. Kohli. Learning low-order models for enforcing high-order statistics. In AISTATS, 2012.
- [34] P. Pletscher, S. Nowozin, P. Kohli, and C. Rother. Putting MAP back on the map. In DAGM, 2011.
- [35] M. Ranjbar, G. Mori, and Y. Wang. Optimizing complex loss functions in structured prediction. In ECCV, 2010.
- [36] S. H. Rice. The expected value of the ratio of correlated random variables. (unpublished note).
- [37] S. H. Rice. A stochastic version of the price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evolutionary Biology*, 8(1), 2008.
- [38] S. Schaible and J. Shi. Fractional programming: the sum-of-ratios case. In M. Fukushima and Y. Yuan, editors, *Proc. 2nd Japanese-Sino Opt. Meeting*, pages 219 - 230, 2003.
- [39] H. Seltman. Approximations for mean and variance of a ratio. (unpublished note).
- [40] J. Shotton, A. W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In CVPR, 2011.
- [41] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In CVPR, 2008.
- [42] C. G. Small. *Expansions and Asymptotics for Statistics*. CRC Press, 2010.
- [43] D. Tarlow and R. P. Adams. Revisiting uncertainty in graph cut solutions. In CVPR, 2012.
- [44] D. Tarlow and R. S. Zemel. Big and tall: Large margin learning with high order loss. In CVPR Workshop on Inference in Graphical Models with Structured Potentials, 2011.
- [45] D. Tarlow and R. S. Zemel. Structured output learning with high order loss functions. In AISTATS, 2012.
- [46] M. Tawarmalani, S. Ahmed, and N. V. Sahinidis. Global optimization of 0 - 1 hyperbolic programs. *Journal of Global*



Optimization, 24:385 - 416, 2002.

[47] J. Tighe and S. Lazebnik. Superparsing - scalable nonparametric image parsing with superpixels. International Journal of Computer Vision, 101(2):329 - 349, 2013.

[48] J. A. Tomlin. On pricing and backward transformation in linear programming. Math. Prog., 6:42 - 47, 1974.

[49] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. International Journal of Computer Vision, 63(2):113 - 140, 2005.

[50] V. N. Vapnik. The Nature of Statistical Learning Theory. Springer, 2nd edition, 2000.

