

指导教师： 杨涛

提交时间： 2016.3.13

# CVPR2015 Paper Translation

No: 48

姓名： 余宏峰

学号： 2013302486

班号： 10011301

# 学习提出对象

Philipp Krahenbuhl  
UC Berkeley

Vladlen Koltun  
Intel Labs

## 摘要

我们提出了一个高度精确的自底向上的对象分割方法。给定一个图像,这种方法就能迅速生成一组区域划定的候选图像中的对象。关键理念是培养了一个整体图形-背景的分割模型。共同训练,使个体模型专业和互补。我们减少共同训练的序列,约束设施选址问题并显示高度精确分割集合体由组合优化训练。培训过程联合优化整体的大小,它的成分,和合并模型的参数相同的目标。基本的乐团操作图像特性,使快速的图像分析。广泛的实验表明,提出的方法优于前一个对象提议由一个重要算法保证,而拥有最低的运行时间。经过训练的概括整个数据集,表明了提出的方法可以学习一种普遍适用的自底向上的分割模型。

## 1. 简介

对象提出算法旨在确定一组小图像区域,每个对象的轮廓并由至少一个不同的地区组成。对象都确定可以自底向上计算,仅基于低阶的边界检测和独立类别分组[7、12、32]。他们都作为起点对象检测和语义分割,并已经成为一个标准的先进的图像分析的第一步管道[5、6、16、17、32]。

支持不同的图像解析任务、对象的建议算法必须有一个数量的特征。他们需要提供区域有用的建议分割形状语义分割和实例[5、6、16、17、23]。他们必须有很高的回忆,产生相应的地区为尽可能多的真正对象。他们必须生成一个可管理的建议限制不必要的工作量。他们必须迅速支持高性能图像解析(16、32)。



图1

图1: 对象的三个照片来着微软 COCO的数据集。从左到右:输入图像,真实实例生成的分割、地区的建议的方法。注意准确实例提议在顶部和中部行,尽管颜色和纹理相似实例。在最后一行,合奏训练正确指出了白色冲浪板作为一个单独的对象有三个连接组件。

在本文中,我们提出一个关注所有特征的对象提议算法。关键的想法是优化图形-背景分割的整体模型。给定一个新的图像,该算法简单适用模型和输出所有的生产前景。该算法快速高效是因为每个模型都能操作基本的图像特征。该算法由一个训练有素的模拟来实现,具体如图1所示。

许多对象之前提议的技术被视为二元分割模型的集合体[7、12、19]。然而,在每种情况下都使用相同的模型只在一个或两个对象,根据预定义的安排是不同的。在某些情况下,多样性是确定分类器[7、12]是否有效且用于排名的重要依据。

相比之下,提出方法优化多样化培训期间合奏全球细分模型。培训的目标是生成具有准确性的对象,且设置平衡它的大小。我们表明,培训目标可以表示而言设施选址问题和优化组合技术。培训共同优化的大小合奏,它的成分,包含的参数模型,为同一目标。该算法对对象的数量在培训时间和建议可以控制的,生成不需要测试时间。

我们在逾越时间节点中进行大量的实验，通过VOC2012数据集和最近微软COCO数据集，提出了方法的性能进行比较最先进的对象提议算法有限很多。我们评估这两个模式的差别的准确性和边界框的建议准确性。在模式差别的准确性，我们的方法优于之前的方法，同时拥有最低的运行时间。例如，实现的方法94%的召回2012数据集以详细的VOC形状重叠：有史以来最高的报道。我们的方法还的收益率最高的，边界框模式的准确性通过提出的边框模式来确定的。我们还表明，分割算法的训练，我们的方法推广到了数据集中。这表明，提出的方法是有能力的学习准确的普遍适用的模型自底向上的对象分割。

## 2. 相关研究

基于滑动窗口的对象检测管道普遍遵循中提琴和琼斯的作品[9, 14, 33]。因为执行的详细分类所有候选人窗口引起不必要的计算成本，许多方法已经开发修剪和等级矩形窗口，从而分配计算预算到最有前途的候选人[1, 8, 22, 35]。最近的排名Zitnick和美元的方法演示了高效率和高召回[35]。不像这些作品，我们致力于促进地区的建议详细的形状暗示，为了支持多样化的图像分析细分任务，包括语义分割和实例[5, 6, 16, 17, 23]。虽然不是主要的焦点我们的工作，简单的边框区域通过我们的模型识别收益率最先进的成果在边界框的提议。

使用自底向上的分割产生的候选人区域对象检测被Malisiewicz主张和埃[25、26]，获得的候选人地区通过应用多种分割算法不同参数，收集得到的片段，和添加区域通过合并相邻段。这建立在罗素等。[29]，他使用了无人监督的对象相似的方法发现在图像集合。使用多个分段和分组相邻段已成为常见的成分后续建议算法[2、3、27、32]。

另一种方法地区候选人的一代计算许多图形-背景分割和每个计算前景区域添加到候选集[7、12、19、21]。通过应用生成指定的建议组分割模型不同的位置图像。最近算法的是基米-雷克南使用这种方法[21]达到最先进的准确性。我们的方法还使用图形-背景分割，但是建议是由多元化，包括生成的多个模型类型。的大小和组成最大化的合奏训练期间进行了优化候选集的精度相对于它的大小。

## 3. 研究模型

我们的方法优化了一个整体  $\mathcal{M} = \{M_1, \dots, M_K\}$  二元分割模型。我们主要使用两种类型的模型。第一个类型是全球CRF产生一个分割对于一个给定的图像。第二种类型是一个本地化CRF，鉴于图像位置考虑。指定的位置作为一个可选的注意机制。应用程序训练有素的合奏的图像如图2所示。

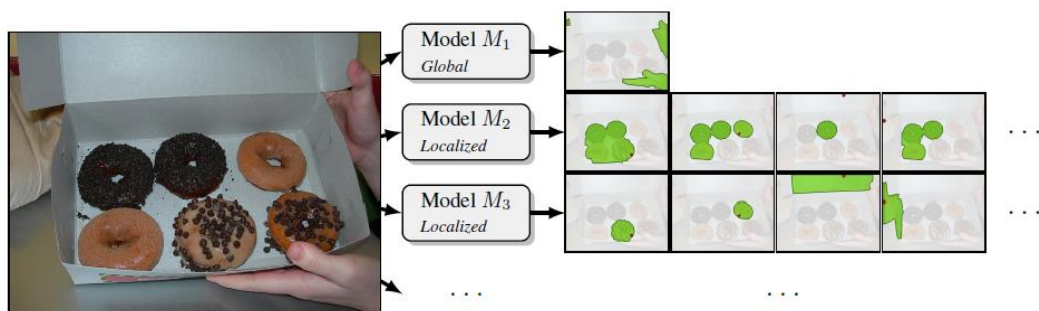


图2：我们的方法共同列出二元分割的整体模型。不同类型的模型产生一个不同的数字地区的建议。在测试时，算法简单的合奏中所有模型适用于一个给定的图像和收集结果建议。培训过程联合优化的规模和组成乐团和所有的参数模型。

让  $k$  模式的  $g$  类型  $M_k$  (e.g. global or localized) 让  $k$  参数向量。让模式  $k$  的集合建议由可视化图像代替。如果是全球性的CRF,  $|\mathbf{X}_k^i| = 1$ 。根据本地化CRF, 提案的数量等于指定的位置在图像的数量。

这两种类型的模型操作超级像素分割一个给定的图像。每个CRF可参数化概率分布的二进制分区本地连接超级像素的图  $(\mathcal{V}, \mathcal{E})$ 。

吉布斯能量的分区  $\mathbf{x} \in \{0, 1\}^n$  是

$$E(\mathbf{x}|I^i; \theta_k) = \sum_{u \in \mathcal{V}} \psi_u(x_u|I^i; \theta_k) + \sum_{(u,v) \in \mathcal{E}} \psi_{u,v}(x_u, x_v|I^i; \theta_k),$$

所以  $\psi_u$  和  $\psi_{u,v}$  分别是一元, 成对势。

我们使用二进制对数线性控与子模块成对潜力。子模块二进制控可以有效地生成提议通过的最大后验 (MAP) 推理 [4]。使参数的对数线性结构估计通过大幅度学习 [31]。

**全球CRF。**  $M_k$  为全球CRF模型, 每个一元潜在的形式为

$$\psi_u(x_u|I; \theta_k) = 1_{[x_u]} f_u^\top \theta_k,$$

当一元特征向量  $f_u$  在超级像素的图评估成对的术语是

$$\psi_{u,v}(x_u, x_v|I; \theta_k) = 1_{[x_u \neq x_v]} f_{u,v}^\top \theta_k,$$

当  $f_{u,v}$  是两两特征向量计算的边缘  $(u, v)$ 。

所有严格积极和成对特性参数  $\theta_k$  非负限制, 在训练以保证子模性。第五节中描述的特性。这个模型产生一个单一的提议对于一个给定的图片:  $\mathbf{X}_k^i = \{\mathbf{x}_k^i\}$ , 所以

$$\mathbf{x}_k^i = \arg \min_{\mathbf{x}} E(\mathbf{x}|I^i; \theta_k).$$

该提议  $\mathbf{x}_k^i$  可能显著不同的函数参数向量  $\theta_k$ 。我们训练一个合并的合奏多个全球模型。共享相同的模型特征向量, 从而掩盖特性计算。不同的培训期间参数向量的联合优化最大化的表现。

全球CRFs有效识别不同的物体全球特征特性, 在只有一个的成本建议每个模型。在培训期间, 不同的全球模型可以专注于不同的常见的对象露面。然而, 全球功能一般

不精确地确定小, 充分表达不突出的对象, 不能有效区分多个实例以相似的外观。由于这些原因, 我们也使用局部模型。

**本地化CRF。** 局部模型有相同的形式全球控, 但他们一元特征向量  $f_u(s)$  包含了功能定义的种子超级像素  $s \in \mathcal{V}$  种子作为轨迹的关注。向量  $f_u(s)$  包括全球模型所使用的所有功能加上简单的特性, 总结之间的距离  $u$  和  $s$  在超级像素图。的距离特性和种子的分布在一个图像中描述的部分5。注意局部模型并不限制任何硬编码依赖种子, 相比之前硬约束和解释工作, 种子位置生成区域封闭的种子 [7、12、21]。我们的培训过程使种子距离特性局部模型与全球功能。的距离特性以不同的方式可以利用不同的模型。例如, 我们观察到本地化模型, 专门描述对象的谎言从给定的种子。

局部模型  $|\mathbf{X}_k^i| = |\mathcal{S}^i|$  产生的建议鉴于图像  $I^i, \mathcal{S}^i$  是种子的集合:

$$\begin{aligned} \mathbf{X}_k^i &= \{\mathbf{x}_k^i(s) : s \in \mathcal{S}^i\}, \\ \mathbf{x}_k^i(s) &= \arg \min_{\mathbf{x}} E(\mathbf{x}|I^i, s; \theta_k). \end{aligned}$$

本地化 crf 的好处之一是 shift-invariance: 一个模型可以专注于外观和类型的对象依靠种子指出个人的实例外观。

## 5. 模型训练

我们  $\mathcal{O} = \{O^1, \dots, O^N\}$  是一组真实对象。对象  $O^i$  表示为一个二进制掩模图像  $I^i$ 。为便于论述, 假设每个图像都包含一个真实对象。如果图像包含的数据集对应相应的多个对象, 那就可以使用复制的方法。

考虑一个候选模型  $M_k$  这个模型  $I^i$  的损失在一个图像的最小 Jaccard 距离定义为对象  $O^i$  和设置的建议  $\mathbf{X}_k^i$ :

$$\Delta(O^i, \mathbf{X}_k^i) = \min_{\mathbf{x} \in \mathbf{X}_k^i} \left( 1 - \frac{|O^i \cap \mathbf{x}|}{|O^i \cup \mathbf{x}|} \right)$$

给定一个乐团  $\mathcal{M} = \{M_1, \dots, M_K\}$ , 对象  $O^i$  上的损失  $\mathcal{M}$  被定义为

$$\min_{k \in \{1, \dots, K\}} \Delta(O^i, \mathbf{X}_k^i).$$

因此失去一个最准确的损失建议。我们的培训目标最小化这一损失训练的例子,平衡的建议:

$$\underset{\mathcal{M}}{\text{minimize}} \sum_i \min_{M_k \in \mathcal{M}} \Delta(O^i, X_k^i) + \lambda \sum_{M_k \in \mathcal{M}} |X_k|, \quad (1)$$

不同的方案生成的总数码  $|X_k|$  通过模型  $M_k$  训练集的图像。的第一项目标最小化 Jaccard 距离建议设置和地面之间的事实对象。的第二个任期惩罚方案集的大小。参数  $\lambda$  平衡的两个目标。一个小值  $\lambda$  生成大型提案集的收益率集合体, 而设定  $\lambda \rightarrow \infty$  收益率模型, 产生一个提议。

目标1优化设置  $\mathcal{M}$  的大小和组成这组的优化参数向量。这一目标是不容易适合 latent variable 期望最大化等训练方法, 但不是的结构优化参数模型 [34]。优化全球完成目标, 我们引入了一种不同的方法。设施的位置。我们的方法减少了培训序列的组合优化问题。具体地说, 假设我们已经生成的潜在的超集模型  $\mathcal{U}$  和整体  $\mathcal{M}$  是来自这候选人设置

$$\underset{\mathcal{M} \subseteq \mathcal{U}}{\text{minimize}} \sum_i \min_{M_k \in \mathcal{M}} \Delta(O^i, X_k^i) + \lambda \sum_{M_k \in \mathcal{M}} |X_k|. \quad (2)$$

这是一个实例 uncapacitated 设施的位置问题 (UFLP) [30]。UFLP 配方问题套设施  $F$  和  $C$  组客户。对于每一个设施  $k \in F$ , 这个设施是开放的成本  $f_k \in \mathbb{R}^+$ 。为每个工厂  $k \in F$  和客户  $i \in C$ , 服务的成本客户  $i$  的设备  $k$  为  $c_{ki}$ 。问题是开立一个设施的子集, 并将每个客户分配给一个开放设备的总成本最小化:

$$\underset{Y \subseteq F}{\text{minimize}} \sum_i \min_{k \in Y} c_{ki} + \sum_{k \in Y} f_k. \quad (3)$$

目标2是 UFLP 与服务成本  $c_{ki} = \Delta(O^i, X_k^i)$

和设施成本  $f_k = \lambda |X_k|$ 。

虽然 uncapacitated 设施选址是 np 难的, 巨大的实际利益, 一直得到广泛的研究。一些已知的算法执行非常嗯, 接近基准问题的精确解在 [28] 百分之一的一小部分。注意, 规的变体 UFLP 与非常近似算法强近似担保 [20, 24]。虽然 Jaccard 距离是一个度量 [10], 目标2不是一个指标 UFLP。然而, 近似算法本身 [20, 24] 有很好的实验在一般情况下性能

甚至 [28]。我们的实现优化目标2使用三个算法 [28、20、24] 并选择成本最低的解决方案。

候选人的一代。目的1是优化解决对自适应序列的设施选址问题生成候选集  $\{\mathcal{U}_1, \dots, \mathcal{U}_T\}$ 。生成一个初始候选集  $\mathcal{U}_1$ 。我们样品的一个子集训练数据随机均匀。一个明显的 CRF 是使用 largemargin 优化为每个样本训练的例子学习 [31]。全局或局部, CRF 类型是随机选择的统一为每个训练的例子。让  $\Gamma_1$  表示这组模型。在这第一次迭代中, 我们设置  $\mathcal{U}_1 = \Gamma_1$ 。然后我们优化目标2的候选人集  $\mathcal{U}_1$  获得一个  $\mathcal{M}_1$ 。培训收益迭代。在迭代  $t$ , 我们样品一个子集  $S_t \subseteq \mathcal{O}$  均匀随机。一套新的模型  $\Gamma_t$  由拟合优化的 CRF 每个采样对象。模型类型是随机选择为每个新 CRF。我们也重新培训每个模型  $M_k$  从系综  $\mathcal{M}_{t-1}$  在训练的例子  $\hat{\mathcal{O}}_k \subset \mathcal{O}$  这是最适合的模型:

$$\hat{\mathcal{O}}_k = \{O^i : \forall l \neq k. \Delta(O^i, X_k^i) \leq \Delta(O^i, X_l^i)\}.$$

这是一个 EM-style 步骤类似于结构化的潜变量培训 [34]。然而, 而不是取代  $\mathcal{M}_{t-1}$  与重新训练模型, 我们添加一组  $\mathcal{M}'_{t-1}$  重新训练模型候选集  $\mathcal{U}_t$ 。随着新车型  $\Gamma_t$  和前面的合奏:  $\mathcal{U}_t = \mathcal{M}_{t-1} \cup \mathcal{M}'_{t-1} \cup \Gamma_t$ 。客观的2然后优化候选集  $\mathcal{U}_t$ 。这产生一个乐团  $\mathcal{M}_t$ 。过程是迭代, 直到最后试装  $\mathcal{M}_T$  生产。我们使用  $T = 10$  在所有实验。该算法算法1进行了总结。

Algorithm 1: Ensemble training

```

 $\mathcal{M}_0 := \emptyset;$ 
for  $t = 1 \dots T$  do
     $S_t :=$  new training examples;
     $\Gamma_t :=$  new models optimized for  $S_t$ ;
     $\mathcal{M}'_{t-1} :=$  reoptimized models from  $\mathcal{M}_{t-1}$ ;
     $\mathcal{U}_t := \mathcal{M}_{t-1} \cup \mathcal{M}'_{t-1} \cup \Gamma_t$ ;
     $\mathcal{M}_t :=$  UFLP( $\mathcal{U}_t$ );
end
Return  $\mathcal{M}_T$ ;
    
```

## 5. 具体实现

我们使用 superpixel 分割 Krahenbuhl 和 Vladlen Koltun [21], 这是基于 ar 和 Zitnick [11] 的边界检测算法。边界检测和 superpixel 分割提供了一种加权 superpixel 图, 重量  $w_{u,v}$  显示的边界强度

相邻superpixels  $u$  和  $v$ . 全球一元特征向量  $f_u$  有18个元素。我们用九RGB颜色特征: 平均颜色 superpixel  $u$ , 整个图像的平均颜色, element-wise平方差异在两个之间。我们也使用五个位置特点: 质量中心  $(x, y)$  的 superpixel  $u$  归一化到  $[-1, 1]$ , 以及  $x^2, y^2$ , 和  $xy$ . 最后, 我们添加四个边界距离特性, 使用的测地距离  $u$  从图像边界 superpixel 图  $(V, E)$  与每个边缘  $(i, j) \in E$  再加权的  $w_{i,j}^\alpha$  到  $\alpha = 0, 1, 2, 3$ . 请注意我们的基本特性: 我们依靠学习算法找到好的参数集, 利用这些简单的根据需要的特性。结果是快速提议的一代不受昂贵功能评估。

局部模型我们添加四个额外的元素, 总结之间的距离  $u$  和一个种子 superpixel  $s$ . 我们使用的测地线距离  $u$  和  $s$  用相同的四套边的权值。成对的特征向量  $f_{u,v}$  有五个要素:

$\exp(-\beta w_{u,v})$  到  $\beta = 0, 1, 2, 3, 4$ . 指数可以确保成对的特点是积极和成对势子模块。

我们训练三种类型的局部模型在三个种子分布。使用种子种子分布位置模型的基米-雷克南 "ahenb" uhl 和 Koltun [21]。训练种子放置模型, 我们分区中的对象的集合 2012年帕斯卡VOC在第三节最大的训练集的大小, 中第三, 和最小的第三。三个种子位置模型分别训练集。位置模型分配平均15, 70年和200年每种子图像, 分别。是过滤掉非常空的建议。Near-duplicate建议使用快速过滤掉重复检测基米-雷克南的 "ahenb" uhl 和 Koltun [21]。

### 5.1 小对象

到目前为止所描述的模型类型——全球 CRF 和本地化 CRFS ——在同一 superpixel 分割操作。这使得快速特性计算和推理, 但图像的量化域是有成本的。任何分区 superpixel 层面都表现不佳为对象的单个 superpixel 差不多大小或更小, 不与 superpixel 边界对齐。这特别适用于微软的 COCO 数据集 [23], 33% 的有一个带注释的对象面积 25 25 像素或更少。在这个数据集的

一部分, 基于我们的 superpixel 分割不能任何建议实现最佳重叠 (ABO 血型) 平均在 45% 以上。这与剩余的数据集, 在吗 superpixel 分割限制最高可实现的 ABO 血型到 90%。

提出了合奏训练方法容易适应额外的模型类型。我们添加一个模型类型专门针对小对象。这个模型 oversegments Felzenszwalb 和 Huttenlocher 的图像使用算法 [15], 提出所有段小于 1000 像素。该模型有两个参数: 颜色空间 (实验室或 HSV) 和最小内部不同参数使用由 Felzenszwalb-Huttenlocher 算法。在培训期间, 这个模型只是抽样与其他类型当一个候选人模型生成。的参数这种模型是随机取样均匀的。自训练过程是完全通用的, 它不需要修改。选择有利的小对象模型自动如果模型中包括他们改善目标 2。

## 6. 研究结果

我们评估方法在帕斯卡 VOC2012 数据集 [13] 数据集和微软 COCO [23]。为帕斯卡 VOC2012 数据集, 我们的火车段注释在训练集 (1464 图片, 1464 分割对象) 和评估验证集 (1449 图片、3422 分段对象)。边界框的建议在较大的检测精度评估数据集 (5823 图片, 13841 边界框)。微软的 COCO 更大的数据集, 以 82783 图像和训练 40504 年验证图像。我们在 8000 年的一个子集训练训练图片与 62135 分段对象和评估完整的验证 296492 分段对象集。所有实验进行一个英特尔酷睿 i7 - 3770 k 在 3.5 GHz 处理器时钟。运行时方法报道单线程执行和覆盖所有业务, 包括边界检测和过度分割。

评估对象的建议我们的质量使用的平均最佳重叠 (ABO 血型) 和召回措施 [7, 21]。地面真理对象集之间的 ABO 血型

$O = \{O^1, \dots, O^N\}$  和一组建议  $X$  计算使用每个地面真理之间的重叠区域  $O^i \in O$  和最亲密的对象的建议  $x \in X$ :

$$ABO = \frac{1}{|O|} \sum_{O^i \in O} \max_{x \in X} \mathcal{J}(O^i, x).$$

这  $\mathcal{J}$  是 Jaccard 系数:  $\mathcal{J}(O^i, x) = \frac{|O^i \cap x|}{|O^i \cup x|} \cdot \alpha$ .

到  $\mathbf{X}$  是部分的分数  $O^i$  在  $\hat{O}$  取最大值  $x \in \mathbf{X} \mathcal{J}(O^i, x) \geq \alpha$

我们先评估我们的培训的不同组件程序, 然后前一组比较工作。训练过程。评估不同的组件我们的培训过程使用不同的变体的过程训练一个大约2200的建议VOC 2012数据集。首先, 我们只使用的初始阶段的过程中, 随机采样训练示例和模型优化个人的例子。我们优化的组合使用一个合奏一轮UFLP。第二, 我们添加T迭代的培训重新训练, 每个模型的最适合的对象吗模型。这是一个EM-style结构化的潜变量的培训由UFLP过程[34], 初始化。第三个变量添加组合(UFLP)每次迭代优化。第四个变体是完整的程序, 注入新车型培训随机取样的例子在每个迭代中。结果被发表在表1。每一个组件的过程改进的性能训练有素的合奏, 并有很强的统计学意义。

Training procedure	ABO	p-value
Initial stage	0.749	-
EM	0.777	<0.01
EM + UFLP	0.781	<0.01
EM + New models + UFLP	0.785	<0.01

表1

评估不同组件的培训过程中, 用我们的完整模型  $\lambda = 0:03$  VOC 2012测试集(约2200提案)。每一行的假定值的措施统计学意义的改进之前的行。

### 6.1 VOC 2012地区的准确性

现在我们评估该地区建议的准确性由我们的方法对VOC 2012数据集。结果被发表在表3。我们比较了六个最先进的对象提议算法。参数  $\lambda$  将几种不同的价值观匹配由不同数量的建议每个方法之前, 如表3所示。为每个方法我们测量的ABO血型, 50%的召回, 召回70%, p -使用学生的学习任务价值计算。研究措施我们的方法的统计显著性优于每个竞争对手的方法。每个对象是地面真理作为一个独立的观察。为每个对象和每个竞争方法, 测试评估是否设置建议由我们的方法有较低或相等与这个对象重叠产生的建议竞争的方法。

我们的方法优于先前的所有的方法统计学意义 ( $p < 0:01$ ), 除了MCG[3]这结果并不显著。我们的方法也有最低的运行时间为所有提案集大小。详情见表3。

建议设置大小的第一层(大约650的建议), 我们的方法具有最高的ABO血型。对于第二个层(1000 - 1600)提议, 我们的方法具有最高的ABO血型和优于之前最近的方法(共和党)2百分点。更高的层(2000以上的建议), 我们的方法的ABO血型3 - 8高出比之前所有的方法除了微克。注意, 回忆我们的方法是一直高于措施MCG, 我们的算法是一个数量级以上得更快。

我们的方法包括0:5的运行时间边界检测和超级像素分割。其余的运行时间是几乎同样分成特性计算, 乘法的特性和参数向量, 通过图削减使能量最小化, 能消除 近两倍的能量。

表2显示了组成一个完整的整体, 训练与  $\lambda = 0:01$  VOC 2012数据集。建议数字公布之前near-duplicate和空的建议去除。全球CRF生产为主大建议, 最适合大约8%的对象。大部分的提议是由局部控生成的, 全球CRF生产为主大建议, 最适合大约8%的对象。大部分的提议是由局部控生成的, 这比大多数其他类型的模型吗的对象。小对象模型的可用性在培训期间对整体精度没有影响大约2000的提议。更高的预算提案, 小对象模型提高70% 1%的召回。没有小对象的ABO血型和50%召回乐团训练模型相差不到0.005。的结果在我们的实验几乎完全呈现方法由于全球和局部CRF模型。

Models	# prop.	% best	$\sqrt{\text{med. area}}$	time
Global	214	8.0	201	0.05s
Localized (l)	1514	27.8	171	0.34s
Localized (m)	1357	23.5	116	0.31s
Localized (s)	2846	48.8	89	0.56s
Small objects	1378	8.6	14	0.16s
All	7309	100.0	115	1.43s

表2: 组成一个训练有素的2012年VOC与  $\lambda = 0:01$  数据集。乐团由全球控, 局部控(小型, 中型, 大型), 和小对象的建议。对于每个模型类型我们报告建议的数量由给定的模型类型(near-duplicate之前空建议删除), 比例的最适合的对象模型给定类型的中间区域的建议和运行时

间。总结百分比超过100%，因为一些的对象是由多个模型同样适用。

建议预算(5000以上)，我们的方法有50%的召回94%：只有6%的挥发性有机化合物的对象2012数据集是错过了。这些对象被显

示在图3中，随着随机取样的照片数据集。错过的对象是部分小段，非常遥远的动物，真实部分注释可怜的自下而上的证据，比如人背后反光的车窗。正如所料，为图像图3中随机抽样，50%的回忆我们的方法是100%。

Method	# prop.	ABO	50%-recall	70%-recall	time	p-value
<i>Our approach</i> , $\lambda = 0.2$	635	<b>0.732</b>	<b>0.861</b>	<b>0.634</b>	<b>0.8s</b>	-
CPMC [7]	646	0.704	0.785	0.609	252s	<0.01
GOP [21]	652	0.720	0.844	0.632	1.0s	<0.01
Global/Local [27]	1056	0.689	0.780	0.579	8s	<0.01
GOP [21]	1199	0.741	0.865	0.673	1.1s	<0.01
<i>Our approach</i> , $\lambda = 0.1$	1236	<b>0.759</b>	<b>0.890</b>	<b>0.685</b>	<b>0.9s</b>	-
RIGOR [19]	1299	0.735	0.832	0.657	5s	<0.01
Cat-Ind OP [12]	1536	0.718	0.821	0.624	119s	<0.01
SCG [3]	2125	0.755	0.871	0.664	5s	<0.01
<i>Our approach</i> , $\lambda = 0.03$	2133	<b>0.785</b>	<b>0.924</b>	<b>0.733</b>	1.1s	-
MCG ranked [3]	2199	<b>0.785</b>	0.897	0.721	30s	0.58
GOP [21]	2286	0.756	0.877	0.699	1.4s	<0.01
<i>Our approach</i> , $\lambda = 0.02$	2707	<b>0.793</b>	<b>0.930</b>	<b>0.762</b>	<b>1.4s</b>	-
GOP [21]	4186	0.766	0.889	0.715	1.7s	<0.01
Selective Search [32]	4374	0.735	0.891	0.597	2.6s	<0.01
<i>Our approach</i> , $\lambda = 0.01$	5144	<b>0.810</b>	<b>0.943</b>	<b>0.785</b>	1.9s	-
MCG [3]	5158	0.808	0.922	0.772	30s	0.14

表3: 定量结果帕斯卡VOC2012数据集。六个建议方法相比，提出了先进的对象的方法。方法是由数量的建议。表分为四层类似的建议设置大小在每个层。我们的方法优于最大幅度的方法，具有较强的统计意义，以最低的运行时间。

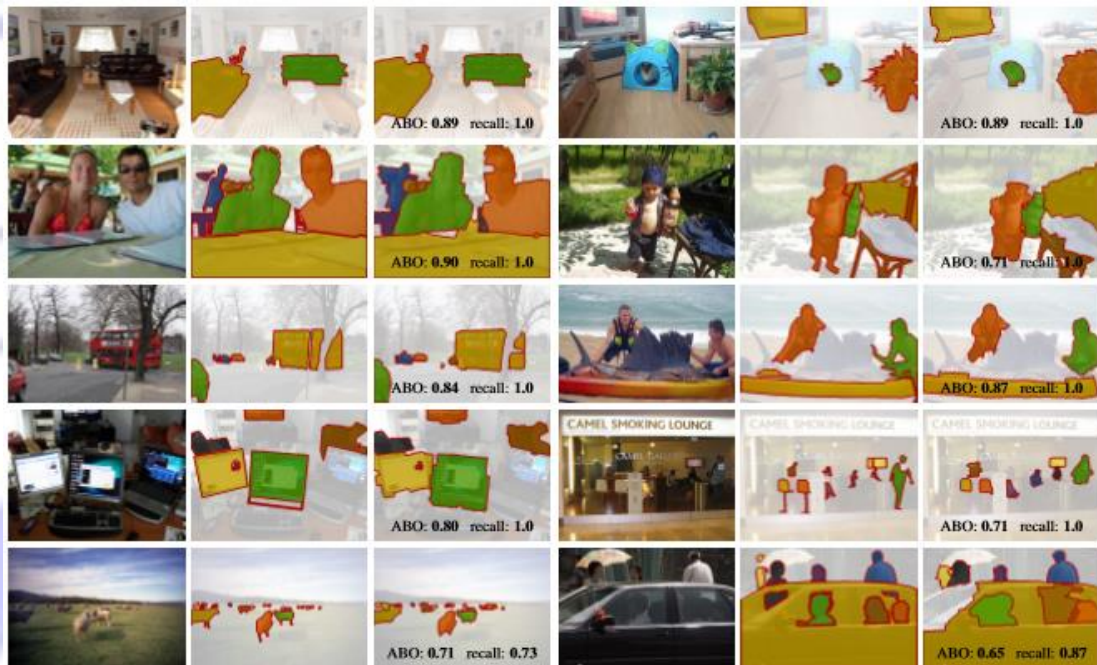


图3: 定性结果2012年VOC的数据集。四大行显示随机图像包含三个或更多的对象。的底下—行显示图像的真实对象，是由我们的算法预测不好。对于每一个图像，图报告ABO血型和50%的回忆我们的算法。有关详细信息，请参阅文本。



## 6.2 VOC 2012边界框的准确性

我们评估的准确性边界框的建议与我们的方法可以获得通过边界箱我们地区的建议。我们遵循耦合方法的 Philipp Krahenbuhl和Kolttun [21]。结果图4所示。客体性 [1] 和必应 [8] 执行召回在50%, 但他们的性能就会降低迅速提高召回阈值。边框 [35] 执行最好的回忆, 但他们的表现也下降70%更严格的精度水平。我们的方法优于所有选择精度高的水平 ( $\mathcal{J} > 0.8$ )。

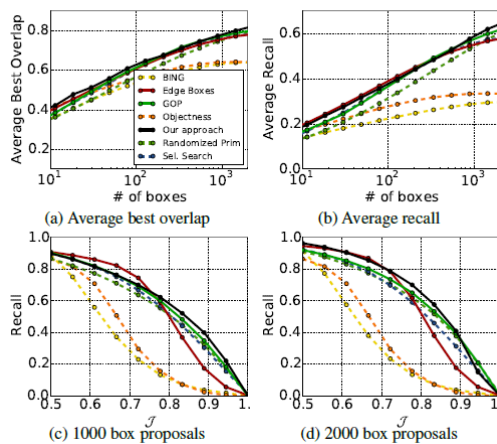


图4: 召回的边界框的建议。(a, b) 平均最好重叠和平均召回不同方案预算。(c, d) 召回1000建议和不同精度的阈值2000年的建议。

我们进一步计算表面下的体积 (VUS 开头) [21], 平均最佳重叠 (ABO 血型), 平均水平回忆 (AR) [18] 2000年边界框的建议。请注意基于“增大化现实”技术是一个特别好的预测检测性能 [18]。结果提供了表4。提出的方法优于所有之前在所有工作精度的措施。

Method	VUS	ABO	AR	Time
BING [8]	0.278	0.640	0.296	0.003s
Objectness [1]	0.324	0.643	0.335	2.2s
Edge boxes [35]	0.527	0.800	0.577	0.3s
Sel. search [32]	0.528	0.781	0.580	2.2s
GOP [21]	0.546	0.797	0.615	0.9s
<b>Our approach</b>	<b>0.558</b>	<b>0.819</b>	<b>0.644</b>	<b>1.1s</b>

表4: 边界框建议精度为2000

## 6.3 微软 COCO

我们评估算法对最近的微软可可数据集 [23]。地面真理分割注释在这个数据集是非常粗糙的。来处理不精确的注解, 我们无视3像素带注释边界的评估。表5报告的准确性我们的方法和先进的提议算法这下有可能运行在大型数据集。我们的方法达到最高70%的召回。在ABO血型我们的方法优于前工作2到4百分点, 在50%召回4到6个百分点。

数据集的泛化。我们也训练模型然后整个VOC 2012细分数据集对其进行评估可可。结果发表在表5。模型训练的COCO和模型训练VOC执行类似的。这强烈表明, 我们的方法能够学习的一般模型自下而上的对象分割, 偏见也具体到一个特定的数据集的对象类。

Method	# prop.	ABO	50%-rec.	70%-rec.
GOP [21]	5501	0.649	0.749	0.527
Sel. search [32]	6504	0.654	0.770	0.471
MCG [3]	5377	0.669	0.759	0.563
Ours, $\lambda = 0.3$	1920	0.626	0.717	0.437
Ours, $\lambda = 0.2$	4078	0.674	0.791	0.526
Ours, $\lambda = 0.1$	5175	0.689	0.809	0.565
Ours (VOC)	2027	0.628	0.707	0.462
Ours (VOC)	4331	0.676	0.781	0.558
Ours (VOC)	5480	0.690	0.802	0.573

表5: 在数据集的泛化。我们训练三个乐团在微软 COCO 数据集和三个帕斯卡在可可VOC2012数据集, 然后测试所有六个。乐团训练在微软 COCO对VOC概括。

## 7. 结论

我们提出了一种新的自底向上的方法对象分割。我们的方法训练破碎形状的合奏分割模型。当应用到一个图像上时, 每个模型都能独立识别候选对象。所有乐团共同训练, 使不同的模型都到不同的结果。我们表明, 共同训练可以减少一个序列的组合优化问题。训练过程是通用和适应不同模型类型。对象的大小和构成优化的参数模型, 为的是同一目标。实验结果证明外面提出的关于详细的形状重叠以及边界框重叠的方法明显优于原先的对象算法方面的建议。结果还表明, 训练有素的对象概括在数据集中,

这同时表明提出的方法能产生普遍适用的自底向上的对象的分割模型。

## 参考文献

- [1] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *PAMI*, 34(11), 2012. 2, 8
- [2] P. Arbeláez, B. Hariharan, C. Gu, S. Gupta, L. D. Bourdev, and J. Malik. Semantic segmentation using regions and parts. In *CVPR*, 2012. 2
- [3] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marqués, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014. 2, 6, 7, 8
- [4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11), 2001. 2
- [5] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Free-form region description with second-order pooling. *PAMI*, 2015. To appear. 1, 2
- [6] J. Carreira, F. Li, and C. Sminchisescu. Object recognition by sequential figure-ground ranking. *IJCV*, 98(3), 2012. 1, 2
- [7] J. Carreira and C. Sminchisescu. CPMC: automatic object segmentation using constrained parametric min-cuts. *PAMI*, 34(7), 2012. 1, 2, 3, 5, 7
- [8] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. H. S. Torr. BING: Binarized normed gradients for objectness estimation at 300fps. In *CVPR*, 2014. 2, 8
- [9] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 2
- [10] M. Deza and M. Laurent. *Geometry of cuts and metrics*. Springer, 1997. 4
- [11] P. Dollár and C. L. Zitnick. Structured forests for fast edge detection. In *ICCV*, 2013. 4
- [12] I. Endres and D. Hoiem. Category-independent object proposals with diverse ranking. *PAMI*, 36(2), 2014. 1, 2, 3, 7
- [13] M. Everingham, L. J. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman. The PASCAL visual object classes (VOC) challenge. *IJCV*, 88(2), 2010. 5
- [14] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI*, 32(9), 2010. 2
- [15] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2), 2004. 5
- [16] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014. 1, 2
- [17] B. Hariharan, P. Arbeláez, R. B. Girshick, and J. Malik. Simultaneous detection and segmentation. In *ECCV*, 2014. 1, 2
- [18] J. Hosang, R. Benenson, P. Dollár, and B. Schiele. What makes for effective detection proposals? In *PAMI*, 2015. 8
- [19] A. Humayun, F. Li, and J. M. Rehg. RIGOR: Reusing inference in graph cuts for generating object regions. In *CVPR*, 2014. 1, 2, 7
- [20] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6), 2003. 4
- [21] P. Krähenbühl and V. Koltun. Geodesic object proposals. In *ECCV*, 2014. 2, 3, 4, 5, 7, 8
- [22] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Efficient subwindow search: A branch and bound framework for object localization. *PAMI*, 31(12), 2009. 2
- [23] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 1, 2, 5, 8
- [24] M. Mahdian, Y. Ye, and J. Zhang. Approximation algorithms for metric facility location problems. *SIAM Journal on Computing*, 36(2), 2006. 4
- [25] T. Malisiewicz and A. A. Efros. Improving spatial support for objects via multiple segmentations. In *BMVC*, 2007. 2
- [26] T. Malisiewicz and A. A. Efros. Recognition by association via learning per-exemplar distances. In *CVPR*, 2008. 2
- [27] P. Rantalankila, J. Kannala, and E. Rahtu. Generating object segmentation proposals using global and local search. In *CVPR*, 2014. 2, 7
- [28] M. Resende and R. Werneck. A hybrid multistart heuristic for the uncapacitated facility location problem. *European Journal of Operational Research*, 174(1), 2006. 4
- [29] B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *CVPR*, 2006. 2
- [30] D. B. Shmoys. Approximation algorithms for facility location problems. In *Approximation Algorithms for Combinatorial Optimization*, 2000. 4
- [31] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *JMLR*, 6, 2005. 2, 4
- [32] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *IJCV*, 104(2), 2013. 1, 2, 7, 8
- [33] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, 2001. 2
- [34] C. Yu and T. Joachims. Learning structural SVMs with latent variables. In *ICML*, 2009. 4, 6
- [35] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014. 2, 8