

指导教师： 杨涛

提交时间： 2016/3/21

CVPR2015 Paper Translation

No: 01

姓名： 马秀

学号： 2013302516

班号： 10011303



PatchCut: 由局部形状转换而来的数据处理的对象分割

作者: Jimei Yang¹, Brian Price², Scott Cohen², Zhe Lin², and Ming-Hsuan Yang¹

¹加州大学默塞德分校 ²Adobe 研究室

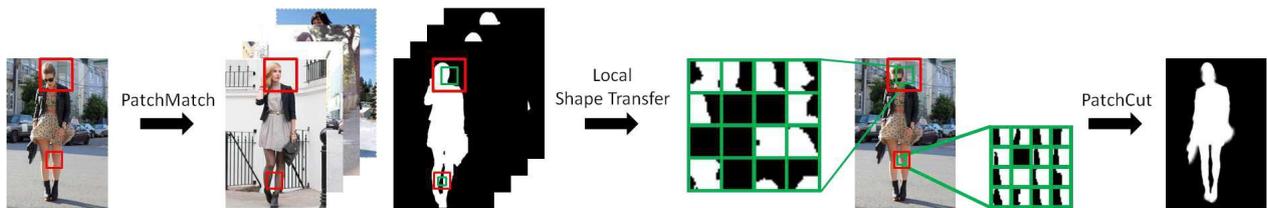


图 1: 用例子来概述所提出的对象分割算法。给定一个测试图像和一组分割的例子, 我们的算法首先展示多尺度图像通过 PatchMatch 与补丁匹配。匹配的补丁内当地的形状面具之后被转移, 来为了测试图像代表 patch-wise 分割候选者。最后, 本地面具候选人基于 MRF 能量函数而被选择, 以一种由粗到细的方式来产生分割。

摘要

对象分割在图像理解和编辑中是非常可取的。当前交互式工具需要大量用户的努力, 而自动方式则被局限于特殊的对象类别或高色彩的对比这种印象中。在本文中, 我们提出了一个数据驱动算法, 使用例子来突破这些限制。由于类似的对象倾向于分享相似的当地形状, 我们将查询图像块与多尺度示例图像匹配以使得本地形状转移。转移的本地形状面具构成补丁级解空间, 因此, 我们开发了一种新型的级联算法 PatchCut, 用来实现由粗到细的对象分割。在级联的每个阶段, 当地形状面具候选人被选择用颜色模型去迭代式改善前一个阶段的估计分割。不同的数据集(魏茨

曼马、时尚达人、对象发现和帕斯卡)的实验结果证明了算法的有效性和鲁棒性。

1. 介绍

对象分割, 从一个有清晰背景中分离前景对象, 一直是计算机视觉的一个重要挑战。它不仅为高端识别面具[7]提供了中端代表, 也对图像编辑[1]有直接应用。在有着前景/背景外观模型[6]的马尔可夫随机域(mrf)中, 对象分割被典型地表述为二进制标记问题。

最近理论[29, 20]显示, 对象分割方法可以用精心准备的目标周围的边界框有效地解决, 以及由用户输入进一步完善。在这些交互式算法

[6, 29, 20]中, 颜色通常被用来前景与背景的分隔。虽然更复杂的图像线索如纹理, 被证明在改进分割性能[34]时是有用的, 而信息的关键来源、对象形状, 在这些算法中显然是不存在的。类似情况存在于突出对象分割[28, 25, 8]中, 在图像有高前后景的色彩对比时, 大多数算法是有效的, 但在处理杂乱的图像时效果很差。无表异常是 Li et al. 的最新成果[22], 通过用经典的显著图整合形状敏感的对象建议[7]在帕斯卡图像[10]中实现了令人印象深刻的对象分割结果。另一方面, 基于模型的算法[5, 19, 18, 4, 36]中, 形状总是分割的主要驱动力。范畴特定的形状模型通常是基于先验知识设计[18]或训练数据[21, 36]离线学习。这个范畴特定的特性限制了基于算法的模型来处理看不见的对象的普遍性。

在这项工作中, 我们提出了一个数据驱动的对象分割算法, 通过使用一个现有的分割例子的数据库来解决上面提到的问题。我们的算法不需要特定范畴的形状模型的离线训练, 也不需要对象形状的先验知识。相反, 我们通过图像检索和匹配把形状面具从类似的分割例子转换成测试图像。与用户和显著驱动算法相比, 转移形状线索帮助解决了从外观模型分割时出现的模棱两可。

现有的数据驱动的对象分割算法[17, 2, 13, 32]主要关注通过基于窗口或基于当地特征的图像匹配转

移整个形状面具。然而, 在本文中, 我们调查一种补丁水平的当地形状转移方案, 发现候选当地形状面具是由补丁匹配算法[3]建立的, 它是通过在查询和示例图像之间有密集对应关系的每块多尺度测试图像。这些候选当地形状面具的确构成一种在线结构化标签空间, 从中可以找到对象分割的解决方案。我们从而开发一种由粗到细的对象分割的新型级联算法。在每个阶段, 我们用在前一阶段估计的粗形状面具定义了一个基于颜色的 MRF 能量函数, 并独立使用最低 MRF 能源, 为每个补丁选择当地形状面具, 从而估计一种具备更多细节的新的形状面具。这种补丁式分割提供了一种全球能源最小化的近似解, 但这个方案更容易并行解决。我们迭代地进行局部形状面具选择, 同时更新前景/背景颜色模型。这个迭代过程与 GrabCut[29]共享一个相似的想法, 但是它在一个结构化的标签空间运行。因此我们命名我们的方法为 PatchCut 方法。我们通过引导以实例、学习、突出为基础的算法的比较, 进行各种对象分割基准数据集的实验。

本文的贡献总结如下:

一种通过补丁级标志转换的新型非参数高阶 MRF 模型的对象分割;

一种有效的迭代算法 (PatchCut), 不使用图像分割而在补丁级解决了提出的 MRF 能量函数;

在各种对象分割基准数据集方面

的艺术层面的性能。

2. 相关工作

基于实例的对象分割。我们的工作与[17, 2]密切相关, 为了使用例子的对象分割。在[17]中, 通过窗口建议, 测试图像与示例图像匹配。通过添加匹配的窗口面具, 估计分割之前包含更多关于对象位置的信息, 而较少对象形状的信息。因此, 其分割性能很大程度上取决于最终的迭代式图像分割的细化步骤。[2]的算法包括两步图像匹配。测试图像的窗口建议首先被定位在示例图片上, 然后每个局部图像被对齐, 使用了尺度不变特征转换流[24]与其相应的测试窗口建议去实现变形面具转移。尽管一个更好的形状先验可以通过这种方式获得, 成千上万的窗口运行筛选流程建议的例子不可避免地导致大量的计算成本。与[17, 2]相比, 我们的算法使用图像补丁来执行多尺度密集匹配, 这个可以通过补丁匹配有效地解决。我们的算法通过由粗到细的方式选择局部面具候选人的途径来估计高质量的形状预设, 而不是增加传输窗口面具, 因此分割并不完全依赖于优化步骤。

结构化的标签空间。我们提出的使用局部面具作为解空间的想法是受到最近的基于结构化森林的图像标注算法[15, 9]的激发。在训练阶段, 标签补丁的聚类结构被利用在分支函数中, 这样每个叶节点存储一个例

子标签补丁。所有叶节点的标签补丁构成结构化标签空间, 为了边缘检测[9]和语义标记[15]。在我们的算法中, 从例子转变的局部形状面具构成另一种为了对象分割的结构化标签空间。

在精神上, 在作预测时, 结构化森林和我们的算法都是致力于保存输出结构(局部文本和形状)。然而, 一个重要的区别是在我们的算法中结构化标签空间是在线算法构造匹配的例子, 这比结构化森林[15, 9]中离线训练推广起来更加灵活和容易。

3. 我们的数据驱动算法

给定一个测试图像 I , 我们的目标是通过使用示例图像 $\{I_m, m = 1, 2, \dots, M\}$ 和分割地面实况 $\{Y_m, m = 1, 2, \dots, M\}$ 估计其分割 Y 。图 1 给出了该提议算法的概述。

3.1. 局部形状转移

我们的算法进行图像匹配来实现从像大多数数据驱动算法的示例的形状转移。然而, 将在大型图像窗口中的整个面具转换可能导致边界质量差[17], 而对齐, 尽管改善边界质量, 但显著增加计算成本[2]。在这项工作中, 我们提出将从多尺度转换局部形状面具。我们通过采样测试 $\{I_s, s = 1, 2, 3\}$ 和示例图像 $\{I_m^s, Y_m^s, s = 1, 2, 3\}$ 构建三层图像金字塔。如果图像的大小是 $[h, w]$, 在 s^{th} 物层取样图像的大小是 $[2^s h_s, 2^s w_s]$ 。对于所有三

个尺度, 我们使用相同大小的图像块来执行匹配和面具转移, 如图 2 所示。在每个规模 ($s = 1, 2, 3$), 我们在每个 2 个照片中取样 16×16 的图像补丁。

对每一块都要确保是可分割的。

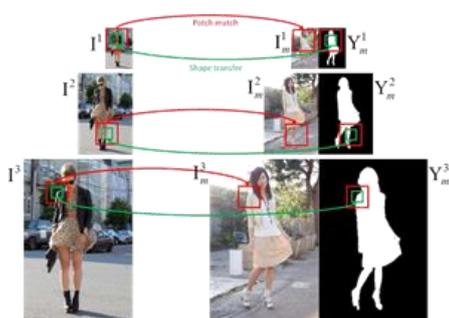


图 2: 多尺度补丁匹配进行局部的形状转换。左边部分显示从时尚达人的一个在三个尺度上的查询图像, 右边部分显示了一个示例图像和分割图像金字塔。对于一个在红色的盒子里的查询图像补丁, 我们找到最佳匹配的示例图像 (红框) 和在绿色盒子里转换其中心分割面具。补丁测试图像 (如图 2 绿框所示), 我们从扩展 32×32 补丁里摘取 SIFT 描述符 x (如图 2 红色框所示)。因此, 在测试 I 和 m^m 例子之间的匹配问题, 可以通

过参数 $\forall k = 1, 2, \dots, K$ 描述, $x_k' m$

是 SIFT 描述符提取 m^m 例子的 K 补丁。这最近邻域问题被 PatchMatch 算法有效地解决了 [3]。因此, 测试补丁以

代价 $d_s = \|x - X_s * m\|_1$ 从 m^m 例子中找到其匹配。在 m^m 例子的匹配的补丁的局部分割面具表示为也提供了分割测试图像的位置和形状信息。我们认为这些局部面具 Z_{skm} 构成补丁式分割测试图像解决空间; 换句话说, 测试图像的分割面具 Y 可以通过 Z_{skm} 近似。注意, 不同的对图像密集通讯方法一直在探索 [24, 14], 使像素式标签转换。但他们的结果是受到当地流 [24] 干扰或被放松的噪音污染 [14]。和 [24, 14] 比较, 我们的方法通过一个更加灵活的匹配方法实现结构化标签转换 (本地面具)。

检测局部形状面具 Z^s 的质量, 对每个补丁来说, 我们预测局部面具的意义, 并找到最有可能的地面真值作为参考。注意, Z^s 事实上限制了局部性状的转移的上界。很明显, 早期形状面具通过添加和 [17] 相似的 Z^{-s} 可被立即估计。

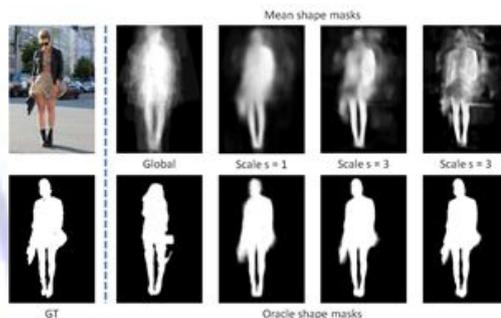


图 3: 在不同的尺度由劣质面具 (上面一行) 和最佳面具 (尾行) 估计的形状前面面具。为了更好的可视化, 面具被取样为原始图像的大小。

图 3 展示了不同尺度的均值和先前形状面具。在粗尺度, 对象是位于

平均的，但它的边界是模糊的面具。朝着更好的尺度，尽管部分粗糙形状（腿）明朗化，其他部分（头和肩膀）是非常乱的。这就是因为局部结构的图像补丁在细尺度保留的好，在边缘模式匹配，但局部面具可能有不一致的前景/背景关系。同时，形状面具的位置和边界质量保持从粗到细的尺度越来越好。这发散的结果表明，如果我们在合适的尺寸下找到正确的标签补丁，就可以获得良好的分割方案。但没有那方面的知识，平均结果远不能令人满意。我们开始了一个好的近似粗尺度，上述观察激励了由粗到细的战略，然后引导我们在合适的尺度选择正确的标签补丁。

3.2. PatchCut

在本节中，我们介绍一种新颖的算法以由粗到细的方式大体估计先验形状。特别是，在 s^m 规模，考虑到形状在以前的规模 $s-1$ ，前细形状被估计使用当地候选形状面具 Z_{skm} 。最后，双分割可以在最细尺度上通过阈值计算。

与先前形状有关的媒体资源功能。我们首先回顾一个基于之前形状的典型对象分割方法，对我们的算法提供了基础。注意，我们暂时忽略规模指数 s ，以保持描述简洁。先前形状的对象分割一般被称为MRF能量方程 [17, 2]，

$$E(Y) = \sum_{i \in V} U(y_i) + \gamma \sum_{i,j \in E} V(y_i, y_j) + \lambda \sum_{i \in V} S(y_i, q_i) \quad (1)$$

Y_i 是像素 i 的双标签， q_j 是先前 Q 形状的可能像素。一元项为每个像

素 $U(y_i)$ 标签的负对数，给定像素的颜色 C_i 和高斯混合模型 (gmm) A_1 和 A_0 ，为了前景和背景的颜色，

$$U(y_i) = -\log P(y_i | c_i, A_1, A_0) \quad (2)$$

二元项 $V(y_i, y_j)$ 衡量不同标签对于两个相邻像素的影响，一般建立在他们的颜色差别上。

$$V(y_i, y_j) = \exp(-\alpha \|c_i - c_j\|^2) I(y_i \neq y_j) \quad (3)$$

当参数被通过图象模糊的色差估计时， I 是指数函数。形状元素衡量先前形状 Q 的不一致性。

$$S(y_i, q_i) = -\log q_i^{y_i} (1 - q_i)^{1 - y_i} \quad (4)$$

该能量函数可以通过交替的两个步骤来解决，以类似于 GrabCut 算法 [29] 的方法：1) 更新 (2) 中当前分段 $\{A_1, A_0\} \leftarrow \tilde{y}$ 的 GMM 色款

2) 解决在 (1) 的 MRF 中的能量函数，由图像分割更新颜色模式： $Y \leftarrow \{A_1, A_0\}$ 。然而，这种方法参数 λ 过于敏感。一方面，如果 λ 为大，颜色模型可能不会改正先前形状的错误；另一方面，如果该 λ 小，分割可能会偏离好的先前形状。

局部形状传递的高阶 MRF

要使用局部候选形状的面具，以解决分割歧义，我们可以自然延伸 (1) 去包括补丁的可能性 PC 和 $(Y(\Delta K))$ 为图像块 $I(\Delta K)$ 激发标签补丁 $Y(\Delta K)$ ，使得与一些局部候选

人造型面具 ZKM= YM (Δ km) 数据库
图像补丁 IM (Δ km) 相似:

$$E'(Y) = E(Y) - \sum_k \log(P_{cand}(Y(\Delta_k))). \quad (5)$$

最后一项是负预期补丁数 (EPLL), Zoran 和 Weiss[37] 使用的一个公式, 为了图像修补以产生先进成果如去模糊。在这里我们通过超过边缘化了的局部形状面具的可能性 MK* 定义补丁, 显示哪个数据库补丁是被选择作为输出补丁的转换的。

$$\begin{aligned} P_{cand}(Y(\Delta_k)) &= \sum_{m=1}^M P(Y(\Delta_k), m_k^* = m) \\ &= \sum_{m=1}^M P(Y(\Delta_k) | m_k^* = m) P(m_k^* = m) \\ &= \sum_{m=1}^M \frac{\exp(-\eta \|Y(\Delta_k) - z_{km}\|_2^2)}{Z_1} \frac{\exp(-\tau d_{km})}{Z_2}, \end{aligned}$$

其中第二项表示我们要转移的 MTH 候选人标签补丁图像的外观的可能性, 与第一个输出标签补丁像素的可能性表示应该是类似的传输补丁。需要注意的是 Z1, Z2 是归一化元素, DKM 是上一节中所介绍的匹配成本。我们假设 η 为大, 以鼓励输出标签补丁 Y(Δk) 作为类似于所选择的候选补丁 ZKM* 越好。对于大型 η 和独特的 ZKM, 我们有

$$P_{cand}(Y(\Delta_k)) \approx \begin{cases} \exp(-\tau d_{km}) / Z_2 & \text{if } Y(\Delta_k) = z_{km} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

and

$$E'(Y) \approx E(Y) + \tau \sum_k H(Y(\Delta_k)), \quad (7)$$

where

$$H(Y(\Delta_k)) = \begin{cases} d_{km} & \text{if } Y(\Delta_k) = z_{km} \\ \infty & \text{otherwise} \end{cases} \quad (8)$$

请注意, 此近似能量 (7) 涉及在引入的非参数高阶随机场 (NHRF) [26], 即认为是自上而下的局部外观匹配 (8) 而不是自下而上的全局图像线索 (1)。

补丁的近似优化。高次项的 H (Y (Δ k) 的) 实际上强制标签补丁从所有的局部形状 ZKM 中筛选, 因而, 当选择的标签补丁在任何重叠区域不匹配时, 能量函数 E' (Y) 不存在。为了解决这个问题, 我们引入一个辅助变量 ZK, 表示第 k 个选定的标签补丁 $ZK \in \{ZK1, \text{转 } K2, \dots, ZKM\}$, 因此通过重写能量 (7)

$$E'(Y, \{z_k\}) \approx E(Y) + \tau \sum_k H(z_k), \text{ s.t. } Y(\Delta_k) = z_k.$$

我们注意到, 能量 E (Y) 可以在 Y (Δ k) 的 = ZK 被进一步分解为局部的能量的总和

$$E'(Y, \{z_k\}) \approx \kappa \sum_k E(z_k) + \tau \sum_k H(z_k), \text{ s.t. } Y(\Delta_k) = z_k.$$

其中常数 κ 与补丁叠加在单个像素的数量成反比。

为了避免 Y(Δk) 和 Zk 之间不一致, 我们通过在每个补丁上引入二次修正将它转化为一个无约束的问题。

$$E'(Y, \{z_k\}) \approx \sum_k (\kappa E(z_k) + \tau H(z_k) + \frac{\beta}{2} \|Y(\Delta_k) - z_k\|^2). \quad (9)$$

在与双分解法 [33] 类似的精神, 这种二次修正能量函数 (9) 可以通过交替解决一系列的在补丁 ZK 上的独立子问题和一个在 Y 的主问题被最小化。但是, 对于足够大的 β, 我们可以通过一个简单的两步骤最小化来近似求解 (9):

$$\hat{z}_k = \arg \min_{z_k} \kappa E(z_k) + \tau H(z_k), \forall k \quad (10)$$

$$\hat{Y} = \arg \min_Y \sum_k \frac{1}{2} \|Y(\Delta_k) - \hat{z}_k\|^2. \quad (11)$$

注意, (10) 可以立即通过评估当地所

有的面具的能量来解决并行的候选人。为了解决(11),我们需要考虑重叠 Z^k 的不一致。通过引入软分割面具(先前形状) $0 \leq Q \leq 1$ 为了 Y , 我们首先通过平均所有选中的候选人解决 $Q = P_k \arg(\Delta k) - Z \min_Q \|k - Q^k\|_2$ Z , 然后以阈值为 0.5 计算 Y 。

算法1 单一尺度PatchCut算法

- 1: 当不是聚合时
- 2: 对于每个补丁 k , 通过(10)选择候选局部形状面具 Z
- 3: 通过平均 Z^k 估计之前形状 Q , 通过(11)细分 Y
- 4: 通过(2)更新前景和背景的 GMM 颜色模型 $\{A_1, A_0\}$
- 5: 结束

鉴于目前的分割 Y , 我们进一步更新在 (2) 中它们的颜色模型 $\{A_1, A_0\}$ 。迭代地, 高阶 MRF 能量 (7) 是利用局部形状掩模候选被最小化。在算法 1 中, 我们总结此过程时, 戏称其为补丁剪切。

级联。在一个单一的尺度使用在图 4 的补丁剪切算法, 装配的级联对象分割算法通过从平均在粗尺度实施例传送的全局形状掩模初始化。注意, 其他软分割方法也可用于初始化 [11, 22]。在每一个规模, 我们用先前估计的 Q_{s-1} 运行算法 1, 颜色模型 A_1, A_0 和候选局部先前形状 $Z(s/km)$ 。该算法直到规模 $S=3$ 为止都可达到。最终的目标对象分割被推断由阈值的先前形状 Q , 表示为 Y^T , 或进一步通过在 (1) 的记为 r 的迭代图切割细化。

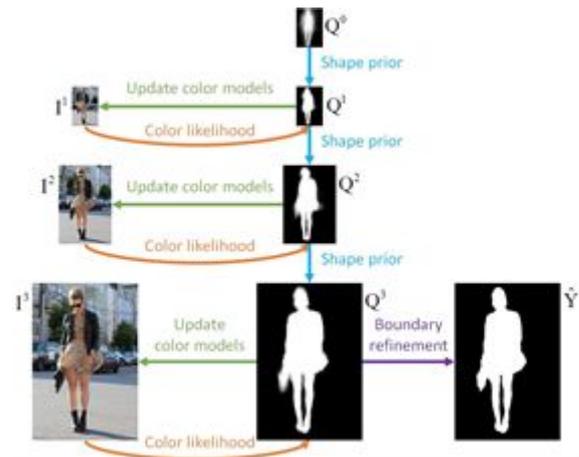


图 4: 对于从粗到细的对象分割的补丁剪切级联

4. 实验结果

对各种对象分割数据集(时尚达人 [35], 魏茨曼马 [5], 对象发现 [30], 和帕斯卡 [10]) 给出实验结果。在 <https://eng.ucmerced.edu/people/jyang44> 可以找到更多的结果。术语 PatchCut-soft 显示形状面具 Q^3 , PatchCut 刺表示二进制分割阈值, PatchCut 表示 Y 细化后的二元分割 Y^r 。

4.1. 时尚达人

此数据集 [35] 由时尚机型 700 街拍摄各种姿势, 杂乱的背景和复杂的外观组成。所有的图像具有相同大小的 600x400 像素。我们在此数据集测试运行 leave-one-out, 这意味着对于每一个图像, 我们通过使用剩余 699 张图像作为数据库运行对象分割 leave-one-out 的测试。如图 5 所示, 我们提出了一些分割结果。在这个实验中的一些分割结果, 我们把我们的算法与目前广泛使用的进行了比较。

多项网络是在图象网络数据集[16]中预先训练的。



图 5: 时尚达人定性结果。

GrabCut 基线[29]。我们通过在每一侧图像边界提供具有 8 个像素的边界框用于为了 GrabCut 的 OpenCV 的执行情况。

表 1: 时尚达人分割性能。

	Jaccard (%)
GrabCut	64.23
PatchCut_thres	86.25
PatchCut	88.33
PatchCut_thres upper bound	95.72
PatchCut upper bound	95.20

我们评估对象分割，是通过平均杰卡德性能（路口过联盟）评分 ($\frac{|Y \cap Y|}{|Y \cup Y|}$)，见表 1。通过简单的阈值的推定形状的面具，贴片切割算法显著优于（超过 20%）GrabCut 的基线，其结果可进一步通过 GrabCut 细化，从 86.25% 提高到 88.33%。图 5 示出的精细化主要发生在对象周围的轮廓上。拿来在 PatchCut 性能上进一步观察，我们计算分割成功率测试的百分比达到上述某种水

平。Figure6 杰卡德分数显示，测试中约 58% 的占比达到 90% 以上杰卡德得分，而约 22% 的测试达到 95% 以上的 Jaccard 得分。

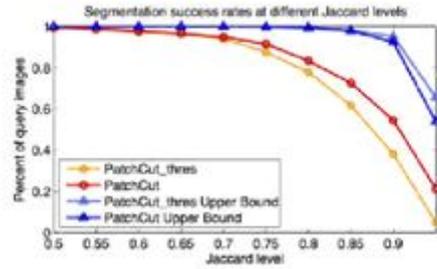


图 6: 时尚达人的分割成功率

上界的表现。我们还评估了 PatchCut 算法的上限的性能。在每个测试图像中，我们估计骨形前面罩 Q 使用地面实况分割 Y，从而产生分割结果。上限分割结果的平均值的 Jaccard 分数在没有细化（表 1）的情况下高达 95.72%，并且，在杰卡德得分水平的 95% 的是分割成功率接近 70% 的（图 6）。这些上限结果证明，从实例转换的局部形状的面具构成一个为了对象分割的有效结构化标签空间。

4.2. 魏兹曼马

魏兹曼马数据集[5]被广泛用于基准对象分割算法。此数据集包括了 328 个马的侧视形象。我们遵循使用 200 图像用于数据库以及其余 128 的用于测试集的典型评价协议。我们在图 7 中展示了一些定性结果。

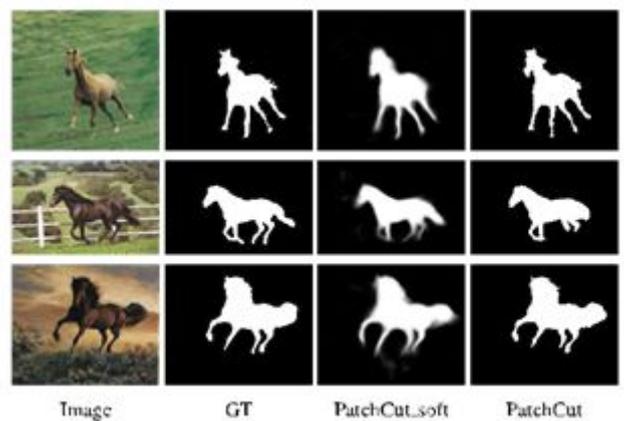


图 7: 魏兹曼马的定性结果

我们用平均杰卡德得分和整体逐像素的分类精度 ($ACC = \frac{|Y == Y'|}{|Y|}$) 评估了对象分割性能。在表 2 中, 我们比较了 PatchCut 算法, 使用的是 Window 面膜转移 [17] 下的领先实例算法, 各种领先的核心化结构 SVM [4], 条件随机场 [23, 21] 基于学习的算法, 以及 Margin 玻尔兹曼机 (BMS) [36]。我们的算法在平均杰卡德评分和精确度两方面都表现的更好。

表 2: 魏兹曼马性能评估

	Jaccard (%)	Acc (%)
PatchCut_thres	80.33	94.7
PatchCut	84.03	95.1
Kernelized Structured SVM [4]	80.10	94.0
Fragment-based CRFs [21]	N/A	95.0
High-Order CRFs [23]	69.90	N/A
Max-Margin BMs [36]	75.78	90.0
Window Mask Transfer [17]	N/A	94.0

特别是, 我们的算法对平均得分杰卡德改进了约 4%。对于马的图像, 我们的算法通常在腿 (图 7) 之前产生高质量形状, 但是由于收缩的腿部偏向较短的边界, 迭代图切割细化趋于切断腿部。

4.3. 对象发现

此数据集包括三个对象类: 飞机、汽车和马和其均由网上收集而来的图像。它最初被设计用来评估对象再整合 [30], 最近由艾哈迈德等人用于对象分割 [2]。此数据集更具挑战性, 因为图像一般具有更复杂的外观。有些图像包含多个小目标, 有些图像是异常值。对于每个类别, 我们使用相同的 100 个测试图像如 [30, 2] 中, 其余为数据库中。图 8 示出了一些定性结果。

我们把我们的与 GrabCut 基线算法 (作为时尚达人有相同的实现) 比较, 一个国家的最先进的再整合算法 [30] 和最新的基于实例的方法 [2]。

表 3: 杰卡德分数上的对象发现。

Jaccard (%)	Airplane	Car	Horse
GrabCut	63.29	67.63	50.32
Co-segmentation [30]	55.81	64.42	51.65
Ahmed et al. [2]	64.27	71.84	55.08
PatchCut_thres	70.44	86.40	63.19
PatchCut	70.49	84.52	64.80

在飞机和马实验中, 精制工序改善的效果微弱, 而在汽车试验中我们的算法达到了更好的效果, 而无需使用细化。可能的原因是, 逐像素颜色的车型可能会混淆阴影与汽车的底部, 而由当地面具估计的先前形状更好得保持了高层结构。

4.4. PASCAL

在这个实验中, 我们展示了使用 PASCAL VOC2010 数据集 [10] 的对于显著的对象分割结果。此数据集更具挑战性, 因为图像是来自 20 个有着大姿势、体型、外观的变化和闭塞的对象阶层, 造型和。Li [22] 等人收集在验证组的 850 个人类受试者的图像的显著目标分割面具。我们使用这些图像作为测试集合。注意, 显著对象分割掩码可能不是二进制, 因为图 10 中作为受试者可能与显著对象的选择不匹配。

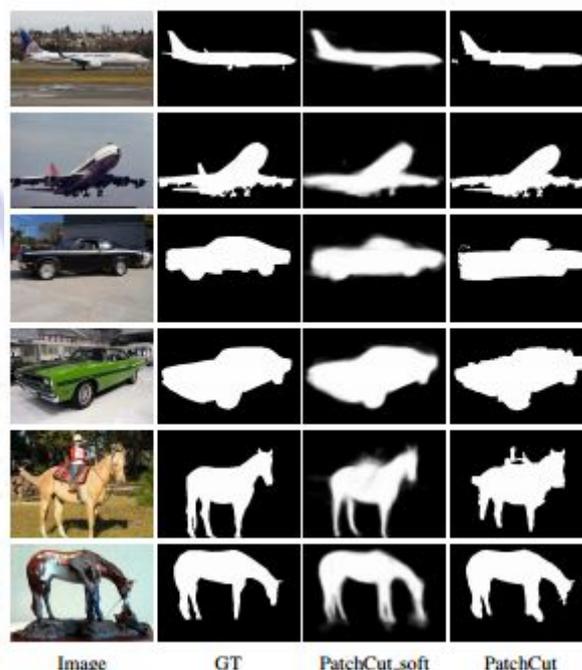


图 8: 对象发现中的定性结果。

另一方面，我们使用所有训练集中的图像来构建我们的例子数据库，并以与[22]中类似的方式收集显著对象分割地面实况。基本上，对于每一个图像，我们使用[27]提供的语义标注作为全分割，并要求6个被试者通过点击选择显著目标区域，所以每段的显著性值定义为接收点击次数除以对象数。不同于以往的实验，我们初始化带有由GBVS算法[11]中产生的突出图的PatchCut算法。映射由其结果表示为GBVS补丁软切割，GBVS PatchCut THRES和GBVS PatchCut。我们主要比较最新的算法CPMC GBVS，在文献[22]中提出并使用GBVS显著性图。图10显示了PatchCut和CPMC GBVS中的一些定性的结果的比较。

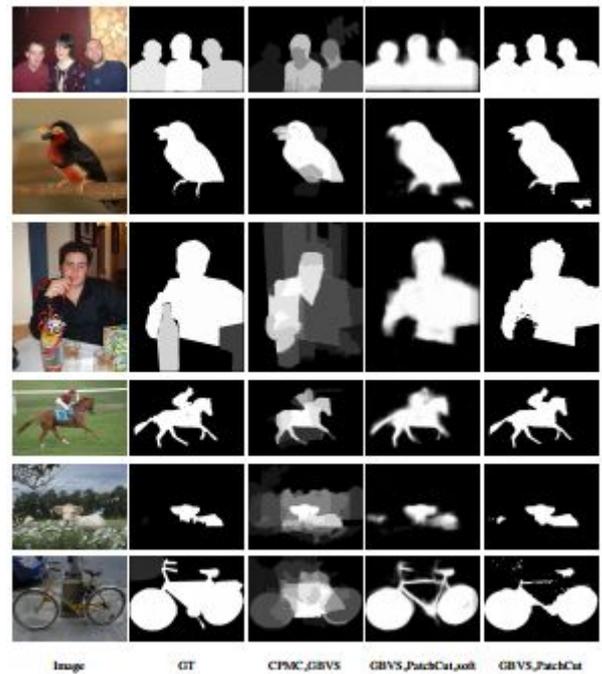


图 10: 在 PASCAL 上比较显著对象分割的结果

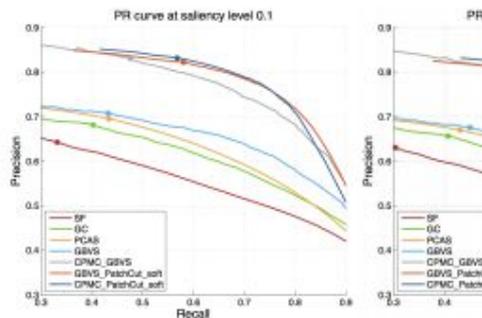


图 9: 比较在精密召回曲线而言的在不同显著性水平上的软分割结果。每条曲线上的点表示的操作点，给出了最好的 F-成绩。

表 4: 在 PASCAL 的 Jaccard 得分

Saliency level	0.1	0.3	0.5
GBVS_GrabCut	45.84	45.25	44.90
CPMC_GBVS [22]	59.43	60.58	60.75
GBVS_PatchCut_thres	60.08	60.22	59.27
GBVS_PatchCut	62.02	62.15	61.14
CPMC_PatchCut.thres	61.37	62.64	62.76
CPMC_PatchCut	63.74	64.92	64.97

定量评价。 我们使用三个阈值：0.1, 0.3, 0.5 把地面实况分割显著性图转换成二进制面具。更大的阈值意味着，具有较高显著度值的物体较少地在地面真相中被选中。我们首先评估精确召回曲线方面的软分割掩码（显著性图）。在图 9 中，我们把 GBVS PatchCut soft 的结果和 CPMC GBVS 以及三个最近的显著性的算法 SF[28], GC [8]和 PCAS[25]) 进行了比较。我们的算法（GBVS PatchCut soft）毫不逊色于 CPMC GBVS 以及明显高于其它显著的算法。其次，我们评估了平均杰卡德得分方

面的二元分割结果。我们通过调整门槛把 CPMC GBVS 的结果转换成二元分割，并找到了最佳的平均杰卡德分数是在阈值 0.3 得到的。表 4 显示，我们的算法（GBVS PatchCut soft）的性能比 CPMC GBVS 稍微好一点，尤其是在低的显著水平时。这个结果也意味着，我们的算法比起 CPMC GBVS 往往倾向于选择更多对象（见在图 10 的例子）。

我们还用 CPMC GBVS 产生的软分割掩码初始化了我们的补丁切割算法，其结果表示为

CPMC PatchCut soft, CPMC PatchCut thres 和 CPMC PatchCut.。凭借这种高品质的初始化，PatchCut 明显优于国有的最先进的两个精确召回曲线（图 9 中的 CPMC PatchCut soft）和平均杰卡德得分（表 4 中的 CPMC PatchCut）。

5. 结论

在本文中，我们提出用实例数据驱动的对象分割算法，它不需要特定类别模型的离线训练，而且可以很好地推广新型的对象。我们的算法通过将例子向局部形状面具候选人转移，为目标分割构建了线上的结构化的标签空间。MRF 标签问题被分解成一系列的独立标签补丁选择子问题，更易于并行解决。我们的算法以一个粗到细的方式操作，并实现在许多领先的对象分割基准的结果，同时具有低的计算成本（大约 10 秒就可以在一典型的台式机上，用未优化的 MATLAB 代码将 200x200 的图像分割）。

致谢

当第一作者是 Adobe 公司的实习生时，这项工作被部分完成。这项成果一部分是获得了由国家科学基金会

职业格兰特 #1149783 和 NSF IIS 格兰特 #1152576 的部分支持，且作为来自 Adobe 公司的馈赠。

参考文献

- [1] Adobe Systems Inc. Photoshop. Creative Cloud, 2014.
- [2] E. Ahmed, S. Cohen, and B. Price. Semantic object selection. In CVPR, 2014.
- [3] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized PatchMatch correspondence algorithm. In ECCV, 2010.
- [4] L. Bertelli, T. Yu, D. Vu, and B. Gokturk. Kernelized structural svm learning for supervised object segmentation. In CVPR, 2011.
- [5] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In ECCV, 2002.
- [6] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In ICCV, 2001.
- [7] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In CVPR, 2010.
- [8] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu. Global contrast based salient region detection. PAMI, 2014.
- [9] P. Dollar and C. Zitnick. Structured forests for fast edge detection. In ICCV, 2013.
- [10] M. Everingham, L. Van Gool,

- C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. ” <http://www.pascalnetwork.org/challenges/VOC/voc2010/workshop/index.html>” .
- [11] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In NIPS, 2006.
- [12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093, 2014.
- [13] J. Kim and K. Grauman. Shape sharing for object segmentation. In ECCV, 2012.
- [14] J. Kim, C. Liu, F. Sha, and K. Grauman. Deformable spatial pyramid matching for fast dense correspondences. In CVPR, 2013.
- [15] P. Kotschieder, S. R. Buló, H. Bischof, and M. Pelillo. Structured class-labels in random forests for semantic image labelling. In ICCV, 2011.
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- [17] D. Kuettel and V. Ferrari. Figure-ground segmentation by transferring window masks. In CVPR, 2012.
- [18] M. P. Kumar, P. Torr, and A. Zisserman. Object. In CVPR, 2005.
- [19] D. Larlus and F. Jurie. Combining appearance models and markov random fields for category level object segmentation. In CVPR, 2008.
- [20] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In ICCV, 2009.
- [21] A. Levin and Y. Weiss. Learning to combine bottom-up and top-down segmentation. In ECCV, 2006.
- [22] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In CVPR, 2014.
- [23] Y. Li, D. Tarlow, and R. Zemel. Exploring compositional high order pattern potentials for structured output learning. In CVPR, 2013.
- [24] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing via label transfer. PAMI, 33(12):2368 - 2382, 2011.
- [25] R. Margolin, A. Tal, and L. Zelnik-Manor. What makes a patch distinct? In CVPR, 2013.
- [26] P. Marquez-Neila, P. Kohli, C. Rother, and L. Baumela. Nonparametric higher-order random fields for image segmentation. In ECCV, 2014.
- [27] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille. The role of context for object detection and semantic segmentation in the wild. In CVPR, 2014.
- [28] F. Perazzi, P. Krahenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In CVPR, 2012.

- [29] C. Rother, V. Kolmogorov, and A. Blake. Grabcut - interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)*, 2004.
- [30] M. Rubinstein, A. Joulin, J. Kopf, and C. Liu. Unsupervised joint object discovery and segmentation in internet images. In *CVPR*, 2013.
- [31] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn reranking. In *CVPR*, 2012.
- [32] J. Tighe and S. Lazebnik. Finding things: Image parsing with regions and per-exemplar detectors. In *CVPR*, 2013.
- [33] H. Wang and D. Koller. Multi-level inference by relaxed dual decomposition for human pose segmentation. In *CVPR*, 2011.
- [34] J. Wu, Y. Zhao, J.-Y. Zhu, S. Luo, and Z. Tu. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. In *CVPR*, 2014.
- [35] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing clothing in fashion photographs. In *CVPR*, 2012.
- [36] J. Yang, S. Safar, and M.-H. Yang. Max-margin boltzmann machines for object segmentation. In *CVPR*, 2014.
- [37] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *ICCV*, 2011.