

基于分层特征选择的高效图像分割

程明明^{1*}, 刘云^{1*}, 侯淇彬¹, 边佳旺¹, Philip Torr², 胡事民³, and 屠卓文⁴

¹ 南开大学, ² 牛津大学, ³ 清华大学, ⁴ 加州大学圣迭戈分校

cmm@nankai.edu.cn

摘要 在本文中, 我们提出了一种实时系统, 即分层特征选择 (Hierarchical Feature Selection, HFS), 该系统以每秒 50 帧的速度进行图像分割。我们通过着重于两个方面来尝试改善以前的图像分割系统的性能: (1) 在现代 GPU 上精心地实现系统以进行有效的特征计算; (2) 一种有效的分层特征选择与学习融合策略。与经典的分割算法相比, 我们的系统展示了其在速度方面的特殊优势, 并且在分割质量上具有可比的结果。在诸如显著性物体检测和拟物性采样之类的应用程序中采用 HFS 可以显著提高性能。我们提出的 HFS 系统 (将是开源的) 可以用于基于图像分割和超像素提取的各种计算机视觉任务。

关键词: 图像分割, 超像素, 聚类

1 引言

图像分割被认为是计算机视觉的主要挑战之一, 过去已经对其进行了广泛的研究。经过数十年的研究, 本领域的研究人员之间达成了共识, 即准确的片段 (无论是大区域还是小超像素) 可作为中级和高级视觉任务的有效输入表示, 尽管本质上是模棱两可的。建立在良好的分割的基础之上的一些典型任务包括: 物体检测/识别 [25, 26]、跟踪 [45]、显著性估算 [11, 23]、拟物性采样 [2, 8, 44] 和 3D 推理 [21] 等。发生这种情况的原因有三点: i) 提取的分割是有意义的单元, 具有诸如形状、纹理等信息特征 [22, 27, 40]; ii) 分割的数量通常大大低于原始图像中的像素数量, 从而产生了更紧凑的表示形式, 并带来了极大的速度优势 [11]; iii) 与原始像素相比, 超像素表示通常具有更好的相干性和鲁棒性 [38]。

过去, 已经出现了几项具有开创性的工作, 它们是在本领域中被广泛采用的最新系统: 基于频谱聚类的归一化切割 (Normalized Cuts) 方法 [39]; 有

* 共同一作。本文为 ECCV 论文 [10] 的中译版。

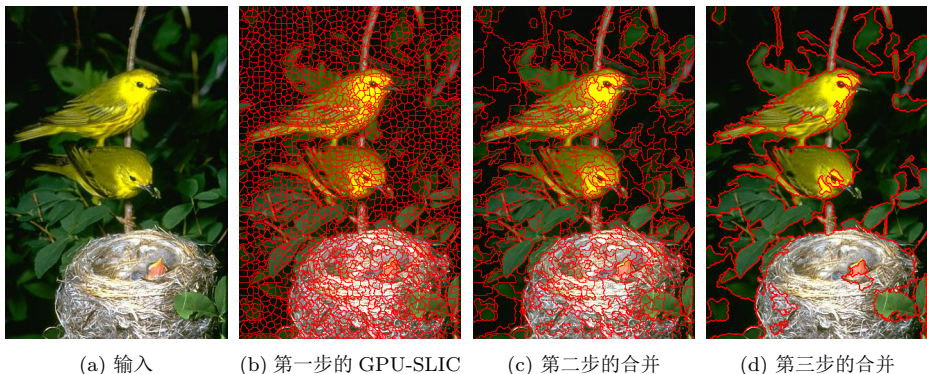


图 1: 来自我们方法不同步骤的样本结果。原始图片 (481×321) 来自 BSDS500 数据集 [4]。这些图像分割结果在 GPU 上以 50fps 生成。

效的特征（颜色）空间模式搜索方法，均值漂移（Mean Shift）算法 [14]；基于图的高效图像分割方法 [18]；基于轮廓转换的分层区域树 [4]；以及多尺度归一化切割算法 [5]。在这些选择中，基于图的高效图像分割（Efficient Graph-Based Image Segmentation, EGB）和 SLIC [1] 方法由于其巨大的速度优势而在计算机视觉和计算机图形学中特别受欢迎 [2, 7, 8, 11, 21, 23, 25, 26, 44, 45]。

在本文中，我们旨在开发一种快速图像分割系统，该系统可以为实时的计算机视觉任务生成高质量的图像分割。我们提出了一个分层的特征选择框架，该框架可学习分层结构各个阶段中的特征组合。我们的工作始于 SLIC 方法的 GPU 版本 [1, 35]，以通过执行过分割来快速获得初始种子区域（超像素）。然后从各个种子区域中提取图像特征，随后进行特征组合过程，并结合从训练数据中学习到的距离度量。请注意，为保持系统的效率，我们仅考虑适合并行计算（即通过 GPU）的那些图像特征。然后基于学习的距离度量执行区域合并过程，以输出用于分层结构中下一级的一组新区域。然后，我们的系统重复进行几次迭代。

本文开发的方法通过以 50fps 生成高质量的图像分割（另请参见图1），对各种实时应用具有实际意义。我们的方法的性能在著名的 BSDS500 数据集 [4] 上进行了定量评估（另请参见第4节）。如评估结果所示（另请参见表2），与其他方法 [1, 4, 5, 18, 41] 相比，我们的方法在分割质量和计算效率之间取得了良好的平衡。我们将开源系统以使其公开可用。

2 相关工作

图像分割是计算机视觉中的一个基本问题 [31]。我们推荐读者参考流行的 BSDS500 基准 [4] 和其他的近期研究 [3, 5, 29, 43] 来获得全面的关于背景的讨论。接下来，我们重点介绍一些与此处提出的方法相关且重要的代表性方法。

过去，人们对基于有效计算和实现归一化切割算法 [39] 的聚类算法给予了一定的关注。[29] 设计了一种多网格特征向量生成器，可以大大提高特征向量的计算速度。在 [43] 中，Taylor 等人尝试使用分水岭过分割来减小特征向量的大小，以实现在不到半秒的时间内计算特征向量的速度。Pont-Tuset 等人 [5] 提出了一种先对特征向量进行下采样的方法，以在减小的尺寸下对其进行求解，然后对结果进行上采样以找回图像的结构。尽管可以通过上述方法获得令人满意的分割结果，但是这些基于光谱聚类的方法的计算时间仍然是瓶颈。

沿着不同的方向，SLIC [1] 成为最著名的方法之一，它在准确性和速度之间取得了很好的平衡，并已在许多应用中被采用 [6, 9, 24, 42, 50]。在 [1] 中，提出了一种 k 均值聚类方法，该方法是通过以规则的网格步长采样像素来初始化聚类中心，然后执行标记过程，其中每个像素都用聚类中心的索引来标记，该聚类中心的搜索区域与它的位置重合。

Felzenszwalb 和 Huttenlocher [18] 提出的基于图的聚类方法也已被广泛使用。对于用边来测量相邻像素之间差异的无向图，[18] 的目标是执行聚类操作，以使每个区域都是所涉及像素的最小生成树。由于它直接从具有简单色彩信息的单个像素开始合并过程，因此 [18] 算法容易产生噪音。与 [18] 相比，我们从比单个像素包含更多信息特征的过分割开始我们的聚类方法。我们将在第3节中详细讨论我们的过程。

其他流行的图像分割方法包括基于特征学习 [36] 的方法。这些方法通过使用具有区分性的分类器将诸如亮度、颜色和纹理属性之类的特征融合在一起，证明了其良好的表示能力。Ren 等人 [36] 提出了一种分层分割方法，其中应用了一系列边界分类器，以从初始过分割开始递归地组合区域。本着这种精神，我们的工作与 [36] 有一定的相似之处，其中使用了一系列的分类器进行区域聚类。在此，通过与 [18] 结合使用时，仔细研究适用于 GPU 实现的区域特征，我们关注实时执行的重要性。我们工作的主要贡献是开发了一个实时的图像分割系统，该系统在许多高层计算机视觉任务中都具有实际意义。

3 方法

在本节中，我们首先介绍问题表述和分层合并算法。然后，我们解释参数学习和特征提取过程，最后讨论关于我们方法背后的设计选择。

3.1 问题表述

给定图像 I ，我们将其划分为 L 级分割 $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_L\}$ 。每个分割 \mathcal{S}_l 是具有 K_l 个区域的图像 I 的分解，

$$\mathcal{S}_l = \{R_1^{(l)}, R_2^{(l)}, \dots, R_{K_l}^{(l)}\}, \quad (1)$$

其中 l 表示分层结构中的层次索引。我们从由大量区域组成的最细分割 \mathcal{S}_1 开始，逐步将区域从 \mathcal{S}_l 合并到较粗略的 \mathcal{S}_{l+1} 。因此，最粗糙层次的分割由最少的区域组成。

在每个步骤中，我们采用基于图的方法 [18] 来实现区域合并过程 $\mathcal{S}_l \Rightarrow \mathcal{S}_{l+1}$ 。令

$$G_l = (\mathcal{S}_l, \mathcal{A}_l) \quad (2)$$

为无向图，顶点为如上定义的一组区域 \mathcal{S}_l ，而边 $(R_i^{(l)}, R_j^{(l)}) \in \mathcal{A}_l$ 对应于成对的可相邻顶点。每条边 $(R_i^{(l)}, R_j^{(l)}) \in \mathcal{A}_l$ 具有特征向量 $\vec{T}_{i,j}^{(l)}$ （另请参见第3.4节）和相应的预测分数 $s_{i,j}^{(l)}$ ，这是区域 $R_i^{(l)}$ 和 $R_j^{(l)}$ 之间距离的非负度量。

根据上述问题定义，我们的任务是快速合并区域，以产生与人类注释最匹配的连贯分割，例如 BSDS500 基准 [4] 中的分割。

3.2 分层合并

为了获得高质量并保持高效率，我们提出：i) 在每个层次进行区域合并后，迭代地学习如何组合特征和更新图像特征；ii) 在合并之前，使用快速的并行的超像素生成方法 [1, 35] 将图像像素分组到初始区域。

我们方法的流程显示在图2中，示例结果显示在图1中，算法显示在算法1中。第一步，利用 GPU-SLIC 方法 [1, 35] 将输入图像过度分割为超像素，这些超像素用作第一层次中的种子区域 $\mathcal{S}_1 = \{R_1^{(1)}, R_2^{(1)}, \dots, R_{K_1}^{(1)}\}$ 。在随后的步骤中，将提取内部和边缘处的特征（另请参见第3.4节）。使用支持向量机（Support Vector Machine, SVM）回归器，我们从训练数据中学习（另请参见第3.3节）如何将特征向量 $\vec{T}_{i,j}^{(l)}$ 映射到区域 $R_i^{(l)}$ 和 $R_j^{(l)}$ 之间的合适的距

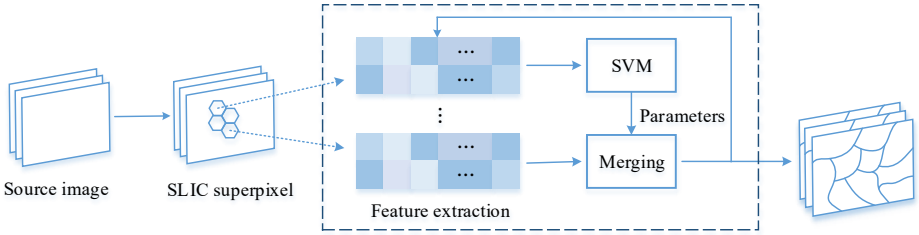


图 2: 所提出方法的流程。

离度量。遵循基于图的高效图像分割（EGB）的框架 [18]，我们用公式 (2) 中定义的图来逐步合并 \mathcal{S}_l 中的区域，以获得较粗略的分割 \mathcal{S}_{l+1} 。

我们的设计原则是受到最近使用判别式学习方法为各种视觉任务找到合适的特征组合的趋势所激发的 [4, 23, 31]。许多心理物理学研究 [37] 表明，人类使用多种线索来分离自然场景中的物体。与临时设计相比，事实证明提取图像特征并允许数据为自己说话是学习如何组合不同视觉线索的合适方法 [4, 31]。在确定是否应合并两个区域时，图像特征在不同层次中扮演着不同的角色，这也促成了我们的系统设计。在精细的层次，例如像素级别，颜色相似性和空间距离很重要，这在许多最新的图像分割方法 [1, 18] 中都可以看到。随着区域合并/分组为粗糙的层次，纹理相似性、区域之间的边缘以及其他线索变得更加重要，成为判断是否应该合并两个区域的决定因素。

与其为所有层次学习线索/特征组合的单一规则 [4, 31]，我们尝试了一种替代方法，该方法可以迭代地更新区域特征及其组合权重（另请参见图3）。我们在实验中看到了我们方法的有利方面。基于这一点，我们设计了一个分层结构，其中涉及多个层次，然后是递归的区域合并 [18] 和特征更新。值

算法 1 HFS 的区域合并

输入：图像 I ，权重 \mathbf{w} ，迭代次数 L

输出：分割 \mathcal{S}_L

初始化： $\mathcal{S}_l = \{R_1^{(l)}, R_2^{(l)}, \dots, R_{K_l}^{(l)}\} \leftarrow \text{GPU-SLIC}(I)$ [1, 35]

for $l = \{1, 2, \dots, L-1\}$ do

 for each $(R_i^{(l)}, R_j^{(l)}) \in \mathcal{A}_l$ do

$s_{i,j}^{(l)} \leftarrow (\vec{T}_{i,j}^{(l)})^T \cdot \mathbf{w}^{(l)}$ ，请分别参见第3.4节和3.3来了解 $\vec{T}_{i,j}^{(l)}$ 和 $\mathbf{w}^{(l)}$

 end for

$\mathcal{S}_{l+1} = \text{EGB}(\mathcal{A}_l, s_{i,j}^{(l)})$ [18]

end for

得注意的一件事是，我们仅提取出简单且适合于在现代 GPU 上并行计算的特征。可以提供更多信息的深度卷积神经网络特征，例如端到端的边缘检测系统 HED [47]，以提高分割效果。我们的实验结果表明，使用 HED 时，F-measure [4]

$$\text{F-measure} = \frac{2 \cdot \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (3)$$

可以增加 6%。但是，由于 HED 的开销（每张图片 0.4 秒），而我们的原始版本为 0.02 秒，因此我们将使用 HED 作为一个可选的选项。

3.3 参数学习

如上所述，给定一组初始区域，我们在每个区域对 $(R_i^{(l)}, R_j^{(l)}) \in \mathcal{A}_l$ 之间学习边的权重 $\mathbf{w}_{i,j}^{(l)} \in \mathbf{w}^{(l)}$ 。由于每个区域对都与一个特征向量 $\vec{T}_{i,j}^{(l)}$ 相关联，因此我们的下一步是为层次 l 中的每个区域对提供一个标签。由于每个层次的初始区域可能具有不规则的形状，因此我们使用 F-measure 来帮助确定 \mathcal{A}_l 中每个区域对的真值标签。

我们首先计算层次为 l 的初始分割的 F-measure，表示为 $F_{init}^{(l)}$ 。然后，对于 \mathcal{A}_l 中的每个区域对 $(R_i^{(l)}, R_j^{(l)})$ ，我们计算在合并 $(R_i^{(l)}, R_j^{(l)})$ 之后的 F-measure。如果合并 $(R_i^{(l)}, R_j^{(l)})$ 之后的 F-measure 大于 $F_{init}^{(l)}$ ，则 $(R_i^{(l)}, R_j^{(l)})$ 的对应标签 $y_{i,j}^{(l)}$ 将被分配为 0。否则， $y_{i,j}^{(l)}$ 将分配为 1。我们采用支持向量机 (SVM) 回归器来学习特征权重 $\mathbf{w}^{(l)}$ 。

3.4 特征提取

我们的系统探索了可以在现代 GPU 上高效计算的一组简单特征。内部和边缘处的特征都在这里被考虑了。表 1 列出了我们考虑的特征。下面我们讨论系统中使用的这些特征的详细信息。

亮度与颜色. CIELAB 颜色空间中的亮度和颜色特征已被证明非常有用 [4, 31]。我们使用平均 $L^*a^*b^*$ 值来表示一个分割区域的颜色。为了容忍亮度和颜色的相对权重的变化，我们将两个相邻区域的欧拉距离 (d_c) 和每个通道的距离 (d_l, d_a, d_b) 同时使用。

沿边界平均的最大梯度. 先前的工作表明，梯度信息是边界检测中的重要线索。与其直接使用梯度，我们使用非极大值抑制之后的梯度。对于相邻的区域 R_i 和 R_j ，计算开始于在像素 $p_k \in \Gamma$ 处放置一个小圆盘，其中 Γ 表

表 1: 相邻区域的特征

特征	维度	符号
CIELAB 各通道的差异	3	d_l, d_a, d_b
CIELAB 值的欧拉距离	1	d_c
沿边界平均的最大梯度	1	d_g
RGB 直方图之间的 χ^2 距离	1	χ_h^2
梯度直方图之间的 χ^2 距离	1	χ_H^2
RGB 值的方差	3	s_r, s_g, s_b
CIELAB 值的方差	3	s'_l, s'_a, s'_b
沿边界平均的 HED 最大值	1	d_h

示其边界。然后，我们计算圆盘中的最大梯度 $\delta'(p_k)$ ，从而计算平均 $\delta'(p_k)$ 作为梯度差异 $d_g(R_i, R_j)$ 。

RGB 直方图之间的 χ^2 距离. 为了利用颜色信息的详细信息，我们采用了颜色直方图，该颜色直方图在 RGB 颜色空间中具有 $8 \times 8 \times 8$ 个维度。对于属于相邻区域的直方图，我们使用 χ^2 距离来衡量它们的差异。

梯度直方图之间的 χ^2 距离. 在为每个区域计算方向梯度直方图时，两个区域的 χ^2 距离也是一个有吸引力的选择。

方差. 方差是衡量一组数据波动的好方法。我们将 RGB (s_r, s_g, s_b) 和 CIELAB (s'_l, s'_a, s'_b) 颜色空间的方差应用于 $R_i \cup R_j$ ，其中 R_i 和 R_j 是相邻的区域。方差的大小反映了两个区域之间的相似性。

沿边界平均的 HED 最大值. HED 特征的计算与上面的梯度特征相似。但是，由于 HED 的额外开销，我们将此选择设为可选。

以上特征在不同层次上扮演着不同的角色。第一层和第二层的特征的权重比较显示在图3中（HED 特征除外）。考虑到计算复杂性，我们只选择一小部分易于计算的特征，而不是全部使用。前五个特征是 d_l, d_a, d_b, d_c 和 d_g 。本文中报告的所有实验结果均基于这些特征。

3.5 实现细节

为了设计一个实用的系统，本文选择 $L = 3$ 作为默认值。我们使用配备有 Intel Xeon CPU E5-2676 v3 @ 2.40GHz 和 NVIDIA GeForce GTX 980 Ti 的计算机进行所有实验。报告的所有运行时间都没有数据并行，除了旨在符合通用规范的拟物性采样的应用部分（第4.2节）。

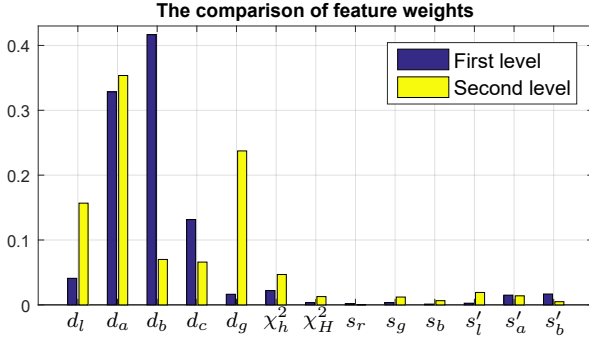


图 3: 在第一层和第二层学习的特征的权重比较。

4 实验

4.1 评测

在这一部分中，我们在 BSDS500 基准 [4] 上评估我们的方法，该基准被广泛用于评估分割和聚类方法。度量有两种选择：最优数据集尺度 (Optimal Dataset Scale, ODS) 是对整个数据集的优化，而最优图像尺度 (Optimal Image Scale, OIS) 是对每张测试图像的优化。对于边界评估，我们使用 ODS 下的基于准确率和召回率的 F-measure 度量。基于区域的度量包括：

- 信息变化 (Variation of Information, VI)，用于测量真值 (Ground Truth, GT) 与所提出的分割之间的距离；
- 概率兰德指数 (Probabilistic Rand Index, PRI)，用于测量 GT 和所提出的分割之间的成对元素分配的兼容性；
- 分割覆盖率 (Segmentation Covering, Covering)，用于衡量 GT 与所提出的分割之间的平均重叠程度。

有关更多详细信息，请参见 [4]。图3展示了所选特征的权重比较。我们可以清楚地看到权重的重要性在不同的层次上是不同的。为了使我们的结果更具说服力，我们将我们的方法与方法 [1, 4, 14, 15, 18, 32, 41]、MCG 以及 SCG 方法 [5] 进行了比较。所有实验均使用公开的源代码完成。

我们在图4中展示了在边界基准上的评测结果，其中所有的执行时间都是在没有数据并行的情况下进行测试的。我们可以看到 [5] 与其他方法相比达到了最佳效果。但是，其简化版本 SCG 仍需要大约 2 秒钟来处理一张图像。因此，尽管具有天赋，但它仍不能在当今的应用中使用。同样，与其他

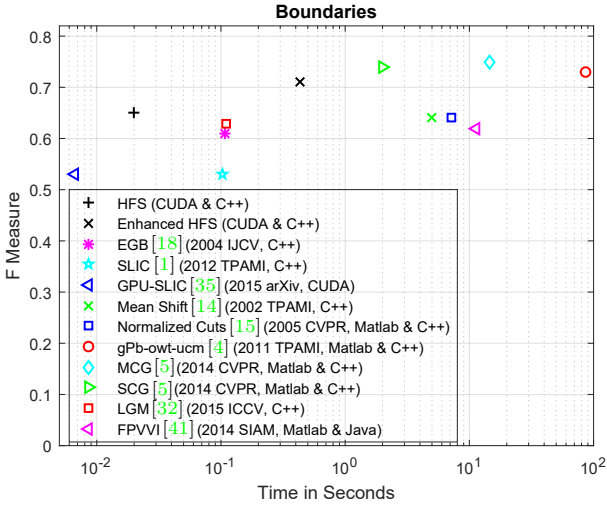


图 4: BSDS500 测试集上的边界的实验评估 [4]。F-measure 是通过 ODS 下的准确率和召回率来计算的，并且在没有数据并行的情况下测试执行时间。

方法相比，[4] 的准确性非常具有竞争力。但是，此方法的速度非常慢，每张图像耗时约 86 秒。我们的方法比 [4,5] 快数百倍，可达到 50fps。启用数据并行性后，速度最高可以达到 200+fps。因此，我们的方法可以轻松用于当今几乎所有的应用程序中，包括一些实时系统。与某些超像素提取方法相比，例如 [1,18,32]，图 4 证明了我们的方法要快得多，更重要的是，准确性也有了显著提高。与其余三种方法 [14,15,41] 进行比较时，我们的速度优势显而易见，尽管 F-measure 仅比它们高一点。其他方面，我们方法的增强版的 F-measure 非常接近最佳性能，并且速度非常快。

表 2 展示了在 BSDS500 数据集 [4] 上基于区域的评测结果。在表 2 中，MCG 在所有指标上均表现最佳，但每张图像大约需要 15 秒。尽管 SLIC 的 GPU 版本可能很快，但 SLIC 的结果却最差。不难发现，在所有这些条件下，我们的方法都接近最佳性能，尤其是增强版本。因此，我们可以得出结论，我们的方法在效率和质量上取得了更好的权衡。图 5 展示了我们的方法与其他两个快速算法 [1,18] 进行比较的示例。

我们的方法没有在每个标准上获得最佳结果的原因有两个。首先，由 SLIC [1] 产生的初始超像素不是很理想。例如，当网格间隔 S 被设置为 8 个像素时 [1]，每张图像将产生大约 2200 个超像素。但是，SLIC 的边界召回率仅为 73%，该指标可以测量落入超像素边界的 8 个邻域的真值轮廓的

表 2: 在 BSDS500 数据集 [4] 上基于区域的评测结果

Methods	GPU	Time(s)	Covering			PRI		VI	
			ODS	OIS	Best	ODS	OIS	ODS	OIS
Human	–	–	0.73	0.73	–	0.87	0.87	1.16	1.16
HFS	Y	0.02	0.56	0.61	0.70	0.81	0.84	1.87	1.68
Enhanced HFS	Y	0.43	0.58	0.65	0.72	0.82	0.86	1.80	1.64
EGB [18]	N	0.11	0.52	0.57	0.69	0.80	0.82	2.21	1.87
SLIC [1]	N	0.10	0.37	0.38	0.48	0.74	0.75	2.56	2.50
GPU-SLIC [35]	Y	0.007	0.34	0.37	0.47	0.73	0.75	2.95	2.81
Mean Shift [14]	N	4.95	0.54	0.58	0.66	0.79	0.81	1.85	1.64
Normalized Cuts [15]	N	7.15	0.45	0.53	0.67	0.78	0.80	2.23	1.89
gPb-owt-ucm [4]	N	86.4	0.59	0.65	0.74	0.83	0.86	1.69	1.48
MCG [5]	N	14.5	0.61	0.66	0.76	0.83	0.86	1.57	1.40
SCG [5]	N	1.98	0.60	0.65	0.74	0.83	0.86	1.63	1.43
LGM [32]	N	0.11	0.52	0.56	0.63	0.78	0.81	1.93	1.79
FPVVI [41]	N	11.3	0.47	0.53	0.62	0.77	0.80	2.10	1.92

比例。这一事实可能会严重影响我们合并策略的第一层的结果。其次，由于 EGB [18] 无法控制生成的超像素的紧凑性，因此我们的合并策略无法获得所需的区域。更具体地说，在 EGB 中，仅使用一个常量参数来防止每个区域变得过大。实际上，由于输入图片的多样性，该标准有时是不合理的。

尽管如此，由于我们有效的架构，我们的结果仍然优于大多数现有的分割方法。以下部分描述了我们的方法在拟物性采样和显著性检测中的应用，从中可以发现我们方法的实用性。

4.2 拟物性采样

拟物性采样已成为近年来的热门话题。作为许多应用（例如物体识别和检测）中的预处理步骤，它会生成许多可能包含物体的边界框。这种类型的算法已在许多现有的物体检测方法中使用 [44, 46]。已经证明，这些物体检测方法的性能要优于经典的基于滑动窗口的范例 [16, 17, 19]。

至于衡量拟物性采样方法的指标，我们用所有类别的平均最佳重叠率（Mean Average Best Overlap, MABO）[44] 和计算效率。Cheng 等人 [13] 最近提出了一种非常快速的拟物性采样方法（BING），该方法可以以 300fps 的速度生成边界框推荐，但是这种方法在 MABO 指标上效果不佳。Chen 等



图 5: EGB、SLIC 和我们的方法的一些示例。我们之所以仅与这两种算法进行比较，是因为它们是仅有的两种足以在应用程序中使用的高效算法。左：图像；左中：SLIC；右中：EGB；右：我们的。区域由其平均颜色表示。所有图片均来自 BSDS500 测试集 [4]。

人 [8] 提出了一种后处理方法 (MTSE)，以优化由拟物性采样方法生成的边界框。在他们的算法中，他们使用 EGB [18] 生成图像分割区域。为了显示

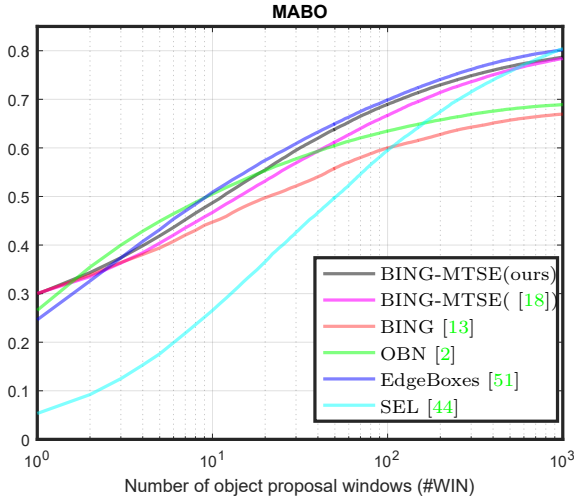


图 6: 不同方法在 VOC2007 测试集上的 MABO 和推荐数量之间的权衡。

我们的分割方法的优势，我们选择 MTSE [8] 作为 BING [13] 的后处理步骤，并用我们的方法替换 EGB [18]。

我们在具有挑战性的 PASCAL VOC2007 数据集 [17] 上对这个新系统进行了广泛的评测。为了展示我们系统的优势，我们将我们的结果与一些当前具有影响力的方法进行了比较，包括 [2, 13, 44, 51]。从图6可以发现，我们修改后的版本比原始版本的 MTSE [8] 的性能更好。使用我们的分割方法，其速度得到了显著提高。我们可以以超过 100fps 的速度获得优秀的推荐框，相比之下，[8] 中报告每张图片需要 0.25 秒。不仅如此，图6显示，使用我们的分割方法的新系统在质量方面是最好的拟物性采样方法之一。因此，我们的新系统在效率和质量之间取得了较好的平衡。

4.3 显著性检测

在这一部分中，我们报告了所提出的方法在计算机视觉的另一个领域中的优越性。长期以来，视觉显著性一直是神经科学、心理学、神经系统和计算机视觉中的一个基本问题。在计算机视觉中，对自然场景中的显著性物体进行检测和分割，也称为显著性物体检测，已经引起了广泛的关注，并产生了许多应用。但是，由于大多数显著性检测方法都是基于区域的，因此有两件事在很长一段时间内都是显著性物体检测的瓶颈。第一个是分割质量 [34]，第二个是计算效率。最近，Jiang 等人 [23] 提出了一种监督学习方

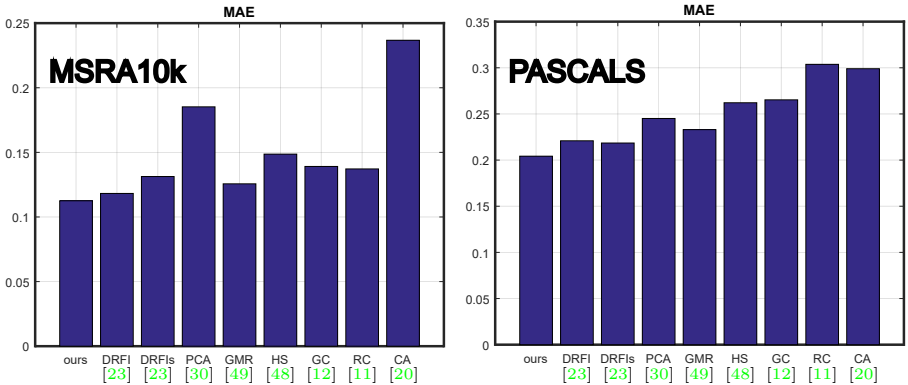


图 7: 最新的方法在 MSRA10K [11] 和 PASCAL-S [28] 上的平均绝对误差。DRFIs 是 DRFI 的单层次版本, 请注意, 我们的方法是单层次的。所提出的方法始终在所有数据集上实现最低的错误率。

法 (DRFI), 以预测由流行的分割方法 EGB [18] 生成的区域的显著性分数, 该方法在多个广泛使用的数据集 (例如 MSRA10K [11] 和 PASCAL-S [28]) 上表现良好。在这里, 我们将 EGB [18] 替换为我们所提出的分割方法 (具有两个层次)。

为了进行公平的比较, 我们使用平均绝对误差 (Mean Absolute Error, MAE) [33] 在上述两个数据集上评测当前流行的显著性检测方法 [11, 12, 20, 23, 30, 48, 49], 平均绝对误差可以反应错误的显著性预测程度。它在显著性图 \mathbf{S} 和二进制真值 \mathbf{GT} 之间定义为

$$\text{MAE} = \frac{1}{|\mathbf{I}|} \sum_x |\mathbf{S}(\mathbf{I}_x) - \mathbf{GT}(\mathbf{I}_x)|, \quad (4)$$

其中 $|\mathbf{I}|$ 是像素总数。这两个数据集上的 MAE 结果显示在图7中。我们的方法在所有数据集上都实现了最低的 MAE 值。具体而言, 相对于次优算法, MAE 得分在 MSRA10K 和 PASCAL-S 数据集上分别降低了 0.57% 和 1.43%。这意味着其预测的显著性图最接近真值。

5 总结

在本文中, 我们提出了一种用于图像分割的分层方法。我们设计了一个分层体系结构, 以享受在不同尺度层次上使用不同特征设置的好处。此外, 我们利用现代 GPU 的能力, 以高效地计算一组简单但有用的特征。与以前

的最新技术相比, 我们的方法可产生高质量的分层区域并且显著提高了速度。在标准的基准 (BSDS500 数据集 [4]) 上的评测结果表明, 我们的方法在效率和质量之间取得了良好的折衷。当将其插入其他计算机视觉任务 (如拟物性采样和显著性检测) 时, 我们的方法可以提高其性能。为了鼓励将来的工作, 我们在<http://mmcheng.net/hfs/>上公开了本文的源代码。

致谢

我们要感谢匿名审稿人的有用反馈。这项研究是由国家自然科学基金委员会 (项目编号 61572264)、华为创新研究计划 (HIRP) 和 CAST 青年人才计划资助的。

参考文献

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI* 34(11), 2274–2282 (2012)
2. Alexe, B., Deselaers, T., Ferrari, V.: Measuring the objectness of image windows. *IEEE TPAMI* 34(11), 2189–2202 (2012)
3. Alpert, S., Galun, M., Brandt, A., Basri, R.: Image segmentation by probabilistic bottom-up aggregation and cue integration. *IEEE TPAMI* 34(2), 315–327 (2012)
4. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE TPAMI* 33(5), 898–916 (2011)
5. Arbelaez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J.: Multiscale combinatorial grouping. In: *IEEE CVPR*. pp. 328–335 (2014)
6. Chang, J., Wei, D., Fisher, J.W.: A video representation using temporal superpixels. In: *IEEE CVPR*. pp. 2051–2058 (2013)
7. Chen, T., Cheng, M.M., Tan, P., Shamir, A., Hu, S.M.: Sketch2photo: internet image montage. *ACM TOG* 28(5), 124 (2009)
8. Chen, X., Ma, H., Wang, X., Zhao, Z.: Improving object proposals with multi-thresholding straddling expansion. In: *IEEE CVPR* (2015)
9. Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D.W.K., Tan, N.M., Tao, D., Cheng, C.Y., Aung, T., Wong, T.Y.: Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE Transactions on Medical Imaging* 32(6), 1019–1032 (2013)

10. Cheng, M.M., Liu, Y., Hou, Q., Bian, J., Torr, P., Hu, S.M., Tu, Z.: HFS: Hierarchical Feature Selection for Efficient Image Segmentation, pp. 867–882 (2016)
11. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.: Global contrast based salient region detection. *IEEE TPAMI* 37(3), 569–582 (2015)
12. Cheng, M.M., Warrell, J., Lin, W.Y., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: *IEEE ICCV*. pp. 1529–1536 (2013)
13. Cheng, M.M., Zhang, Z., Lin, W.Y., Torr, P.: Bing: Binarized normed gradients for objectness estimation at 300fps. In: *IEEE CVPR*. pp. 3286–3293 (2014)
14. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE TPAMI* 24(5), 603–619 (2002)
15. Cour, T., Benezit, F., Shi, J.: Spectral segmentation with multiscale graph decomposition. In: *IEEE CVPR*. vol. 2, pp. 1124–1131 (2005)
16. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: *IEEE CVPR*. pp. 248–255 (2009)
17. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. *IEEE ICCV* 88(2), 303–338 (2010)
18. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *IJCV* 59(2), 167–181 (2004)
19. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE CVPR*. pp. 580–587 (2014)
20. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE TPAMI* 34(10), 1915–1926 (2012)
21. Hoiem, D., Efros, A., Hebert, M., et al.: Geometric context from a single image. In: *IEEE ICCV*. vol. 1, pp. 654–661 (2005)
22. Hu, S.M., Zhang, F.L., Wang, M., Martin, R.R., Wang, J.: Patchnet: a patch-based image representation for interactive library-driven image editing. *ACM TOG* 32(6), 196 (2013)
23. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: A discriminative regional feature integration approach. In: *IEEE CVPR*. pp. 2083–2090 (2013)
24. Jiang, Z., Davis, L.S.: Submodular salient region detection. In: *IEEE CVPR*. pp. 2043–2050 (2013)
25. Juneja, M., Vedaldi, A., Jawahar, C., Zisserman, A.: Blocks that shout: Distinctive parts for scene classification. In: *IEEE CVPR*. pp. 923–930 (2013)

26. Kohli, P., Torr, P.H., et al.: Robust higher order potentials for enforcing label consistency. *IJCV* 82(3), 302–324 (2009)
27. Li, K., Zhu, Y., Yang, J., Jiang, J.: Video super-resolution using an adaptive superpixel-guided auto-regressive model. *Pattern Recognition* 51, 59–71 (2016)
28. Li, Y., Hou, X., Koch, C., Rehg, J., Yuille, A.: The secrets of salient object segmentation. In: *IEEE CVPR*. pp. 280–287 (2014)
29. Maire, M., Yu, S.X.: Progressive multigrid eigensolvers for multiscale spectral segmentation. In: *IEEE ICCV*. pp. 2184–2191 (2013)
30. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: *IEEE CVPR*. pp. 1139–1146 (2013)
31. Martin, D.R., Fowlkes, C.C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE TPAMI* 26(5), 530–549 (2004)
32. Nguyen, R.M., Brown, M.S.: Fast and effective l0 gradient minimization by region fusion. In: *IEEE ICCV*. pp. 208–216 (2015)
33. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: *IEEE CVPR*. pp. 733–740 (2012)
34. Qi, W., Cheng, M.M., Borji, A., Lu, H., Bai, L.F.: Saliencyrank: Two-stage manifold ranking for salient object detection. *Computational Visual Media* 1(4), 309–320 (2015)
35. Ren, C.Y., Prisacariu, V.A., Reid, I.D.: gslicr: Slic superpixels at over 250hz. *arXiv preprint arXiv:1509.04232* (2015)
36. Ren, Z., Shakhnarovich, G.: Image segmentation by cascaded region agglomeration. In: *IEEE CVPR*. pp. 2011–2018 (2013)
37. Rivest, J., Cabanagh, P.: Localizing contours defined by more than one attribute. *Vision research* 36(1), 53–66 (1996)
38. Russell, C., Kohli, P., Torr, P.H., et al.: Associative hierarchical crfs for object class image segmentation. In: *IEEE ICCV*. pp. 739–746 (2009)
39. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE TPAMI* 22(8), 888–905 (2000)
40. Song, X., Zhang, J., Han, Y., Jiang, J.: Semi-supervised feature selection via hierarchical regression for web image classification. *Multimedia Systems* 22(1), 41–49 (2016)
41. Storath, M., Weinmann, A.: Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences* 7(3), 1826–1852 (2014)

42. Sun, J., Ponce, J.: Learning discriminative part detectors for image classification and cosegmentation. In: IEEE ICCV. pp. 3400–3407 (2013)
43. Taylor, C.J.: Towards fast and accurate segmentation. In: IEEE CVPR. pp. 1916–1922 (2013)
44. Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. *IJCV* 104(2), 154–171 (2013)
45. Wang, S., Lu, H., Yang, F., Yang, M.H.: Superpixel tracking. In: IEEE ICCV. pp. 1323–1330 (2011)
46. Wang, X., Yang, M., Zhu, S., Lin, Y.: Regionlets for generic object detection. In: IEEE ICCV. pp. 17–24 (2013)
47. Xie, S., Tu, Z.: Holistically-nested edge detection. In: IEEE ICCV. pp. 1395–1403 (2015)
48. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: IEEE CVPR. pp. 1155–1162 (2013)
49. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: IEEE CVPR. pp. 3166–3173 (2013)
50. Zhang, L., Gao, Y., Xia, Y., Lu, K., Shen, J., Ji, R.: Representative discovery of structure cues for weakly-supervised image segmentation. *IEEE Transactions on Multimedia* 16(2), 470–479 (2014)
51. Zitnick, C.L., Dollár, P.: Edge boxes: Locating object proposals from edges. In: ECCV. pp. 391–405 (2014)