Semantic Soft Segmentation

YAĞIZ AKSOY, MIT CSAIL, USA and ETH Zürich, Switzerland

TAE-HYUN OH, MIT CSAIL, USA

SYLVAIN PARIS, Adobe Research, USA

MARC POLLEFEYS, ETH Zürich, Switzerland and Microsoft, USA

WOJCIECH MATUSIK, MIT CSAIL, USA

语义软分割

原文链接：http://people.inf.ethz.ch/aksoyy/papers/TOG18-sss.pdf

译文出处：https://github.com/hyifan/TranslatePapers

图 1

我们提出了一种方法，通过在单个图形结构中融合高级和低级图像特征，可以自动生成软分割（soft segments），即表示语义上有意义的区域的层以及它们之间的软过渡（soft transitions）。语义软分割，通过为每个片段（segment）指定一个纯色来可视化，可以用作目标图像编辑任务的掩码（mask），或者所选择的层可以在层颜色估计后用于合成。原始图像来自[Lin 等人 2014]（左上角，右下角）、Death to the Stock Photo（右上角）和 Y. Aksoy（左下角）。

精确表示图像区域之间的软过渡（soft transitions）对于高质量图像编辑和合成至关重要。当前用于生成这种表示的技术在很大程度上取决于熟练的视觉艺术家的交互，因为创建这样的精确对象选择是繁琐的任务。在这项工作中，我们引入语义软分割（semantic soft sogments），一组层（layer）对应于图像中具有语义意义的区域，在不同对象之间具有精确的软过渡。我们从光谱分割角度处理这个问题，并提出一种图结构（graph structure），其嵌入来自图像的纹理和颜色特征以及由神经网络生成的更高级别的语义信息。通过精心构造的拉普拉斯矩阵（Laplacian matrix）的特征分解（eigendecomposition），完全自动地生成软分割（soft segments）。我们证明了使用语义软分割可以轻松完成复杂的图像编辑任务。

CCS 概念：**计算方法→图像分割**

附加关键词和短语：软分割（soft segmentation），语义分割（semantic segmentation），自然图像抠图（natural image matting），图像编辑（image editing），光谱分割（spectral segmentation）

1.引言

选择和合成是图像编辑过程的核心。例如，局部调整通常从选择开始，组合来自不同图像的元素是产生新内容的有效方式。但是，创建准确的选择是一项繁琐的工作，尤其是涉及模糊边界和透明度时。诸如磁性套索（magnetic lasso）和魔杖（magic wand）之类的工具可以帮助用户，但他们只利用低级别线索，并且严重依赖用户的技能和对图像内容的解释来产生良好的结果。此外，他们只生产二元选择，需要进一步细化以解释软边界，如毛茸茸的狗的轮廓。还有用于帮助用户完成此任务的修边工具（matting tools），但它们只会增加整个编辑过程的繁琐工作。

如果图像满足若干标准，则通过提供中间图像表示（intermediate image representation），对图像进行精确的预分割可以加快编辑过程。首先，这种分割应该提供图像的不同部分，同时还准确地表示它们之间的软过渡。为了允许有针对性的编辑，每个片段应限于图像中语义上有意义的区域的范围，例如，它不应跨越两个对象之间的边界延伸。最后，分割应该完全自动完成，而不是添加交互点或需要艺术家的专业知识。先前用于语义分割，图像抠图或软颜色分割的方法不能满足这些特性中的至少一个。在本文中，我们引入语义软分割，将输入图像全自动分解为一组覆盖场景对象的层，由软过渡分隔。

我们从光谱分解的角度来处理语义软分割问题。我们将来自输入图像的纹理和颜色信息与我们使用经过场景分析训练的卷积神经网络生成的高级语义线索相结合。我们设计了一个图结构，它揭示了语义对象以及它们在相应拉普拉斯矩阵特征向量中的软过渡。我们引入了一个层稀疏度（layer sparsity）的空间变化模型，该模型利用特征向量生成高质量的层，可用于图像编辑。

我们证明我们的算法成功地将图像分解为少量的图层，紧凑而准确地表示场景对象，如图 1 所示。稍后，我们证明了我们的算法可以成功处理对其他技术具有挑战性的图像，并提供示例编辑操作，例如局部颜色调整或背景替换，这些操作受益于我们的图层表示。

2.相关工作

软分割（soft segmentation）

软分割是将图像分解为两个或多个片段，其中每个像素可能部分属于多个片段。层内容根据相应方法的特定目标而变化。例如，软颜色分割方法通过全局优化或逐像素颜色混合提取均匀颜色的软层。虽然软颜色片段被证明对多个图像编辑应用程序有用，如图像重新着色，但它们的内容通常不涉及对象边界，不允许进行有针对性编辑。为了生成空间连通的软分割，Singaraju 和 Vidal 从一组用户定义的区域开始，多次解决两层软分割问题以生成多个层。另一方面，Levin 等人提出光谱抠图，通过光谱分解自动估计一

组空间连接的软分割。Singaraju、Vidal 和 Levin 等人围绕抠图拉普拉斯构建他们的算法，它为图像中的局部软过渡提供了强大的表示。遵循光谱抠图的想法，我们还使用了抠图拉普拉斯和光谱分解。然而，与以前的工作不同，我们构建了一个图，该图将来自深度网络的高级信息与本地纹理信息融合，以便生成对应于图像中具有语义意义的区域的软分割。

自然图像抠图（natural image matting）

自然图像抠图是用户定义的前景区域的每像素不透明度的估计。自然抠图算法的典型输入是三值图（trimap），它定义了不透明前景，透明背景和未知不透明区域。虽然这个问题有不同的方法，所有这些方法都利用了定义的前景和背景区域的颜色特征，但与我们最密切相关的方法是被归类为基于亲和性（affinity-based）的方法。基于亲和性的方法，如闭环抠图（closed-form matting），KNN 抠图（KNN matting）和信息流抠图（information-flow matting），定义像素间亲和性以构建反映图像中不透明度过渡的图。与自然图像抠图方法相反，我们依靠自动生成的语义特征来定义软分割而不是三值图，并生成多个软分割而不是前景分割。尽管它们看似相似，但自然抠图和软分割具有根本差异。使用三值图作为输入的自然抠图成为前景和背景颜色建模的问题，可能是通过选择颜色样本或颜色信息的传播。同时，软分割侧重于检测最适合目标应用的软过渡，在我们的例子中是与语义边界相对应的软过渡。

有针对性的编辑传播（targeted edit propagation）

一些图像编辑方法依赖于图像上用户定义的稀疏（sparse）编辑并将它们传播到整个图像。ScribbleBoost 提出了一个管道概念，在该管道中，他们对用户通过涂鸦指定的对象进行分类，以允许针对图像中的特定对象类进行编辑，然后 DeepProp 等人利用深度网络传播依赖于类的颜色编辑。Eynard 等人构造了一个与我们的方法平行的图表，该图表分析相应拉普拉斯矩阵的特征分解，以产生连贯的着色结果。 An、Pellacini 和 Chen 等人还定义了像素间的亲和性，并利用拉普拉斯矩阵的性质来求解用户定义的编辑的合理传播。虽然我们的结果也可以用于目标编辑，而不是使用预先定义的编辑，但我们直接将图像分解为软分割，并让艺术家在各种场景中使用它们作为中间图像表示，且使用外部图像编辑工具。

语义分割（semantic segmentation）

随着深度神经网络的引入，语义分割得到了显着改善。虽然关于语义分割的详细报告超出了我们的范围，但语义分割的最新进展包括 Zhao 等人的场景分析工作、He 等人和 Fathi 等人的实例分割方法以及 Bertasius 的工作，后者通过颜色边界线索增强了语义分割。我们也利用深层网络进行语义特征分割，但是我们的软分割方法是类不可知论的，也就是说，我们对涉及语义边界的图像的精确分割感兴趣，但我们并不打算对选定的一组类进行分类或检测。其他人还利用类不可知语义信息来提高视频去模糊或电影图像生成的性能。

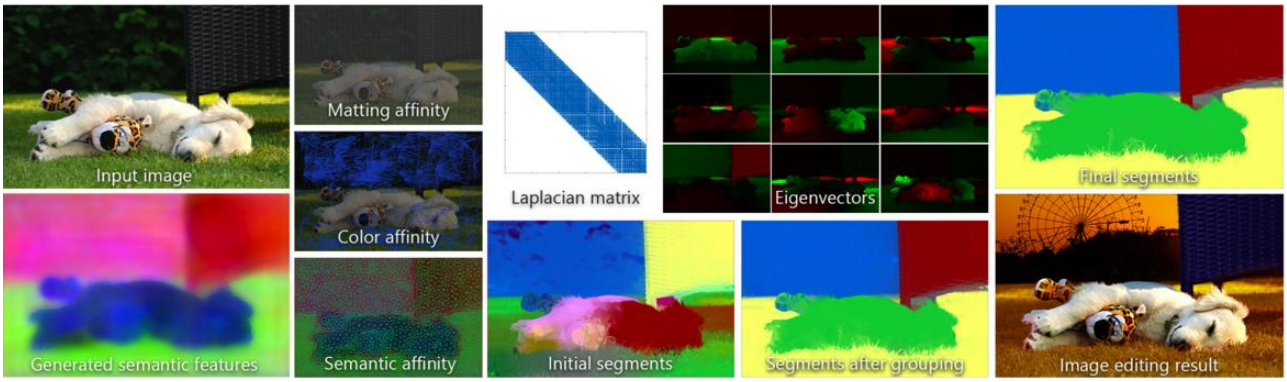

图 2

对于输入图像，我们生成每像素超维度语义特征向量，并使用纹理和语义信息定义图。该图的构造使对应的拉普拉斯矩阵及其特征向量揭示了语义对象和它们之间的软过渡。我们使用特征向量来创建一组初步软分割，并将它们组合起来以获得具有语义意义的片段。最后，我们优化软分割，以便它们可用于目标图像编辑任务。图片来自[Lin 等人 2014]，编辑结果的背景来自 Flickr 用户 rumpleteaser。

3.方法

我们寻求自动生成一个输入图像的软分割，即将图像分解成表示场景中的对象的层，包括透明和软过渡（如果它们存在的话）。每层的每个像素都增加了不透明度值α∈[0,1]，其中α=0 表示完全透明，α=1 表示完全不透明，中间值表示部分不透明度。正如该领域的如 Aksoy、Singaraju、Vidal 等人的其他研究，我们使用了一个附加图像形成模型：

$$(\mathrm{R},\mathrm{G},\mathrm{B})_{\mathrm{input}} = \sum_{\mathrm{i}} \alpha_{\mathrm{i}} (\mathrm{R},\mathrm{G},\mathrm{B})_{\mathrm{i}} \qquad (1a)$$

$$\sum_{\mathrm{i}} \alpha_{\mathrm{i}} = 1 \qquad (1b)$$

即，我们将输入的 RGB 像素表示为每个层 i 中的像素之和，其由相应的α值加权。我们还约束α值在每个像素处总计为 1，表示完全不透明的输入图像。

我们的方法使用与光谱抠图相同的形式，将软分割任务作为特征向量估计问题来制定。这种方法的核心部分是创建一个拉普拉斯矩阵 L，其表示图像中的每对像素属于同一段的可能性。虽然光谱抠图仅使用低级局部颜色分布来构建此矩阵，但是我们描述了如何用非局部提示和高水平语义信息来增强这种方法。最初的方法还描述了如何使用稀疏化从 L 的特征向量创建层。我们将展示这种原始技术的简单版本如何实际产生更好的结果。图 2 显示了我们方法的概述。

3.1 背景

光谱抠图（spectral matting）

我们的方法建立在 Levin 等人的工作基础之上。他们首先介绍了使用局部颜色分布来定义矩阵 L 的抠图拉普拉斯，该矩阵 L 捕获局部补丁（通常为 5×5 像素）中每对像素之间的亲和性。使用该矩阵，他们在用户提供的约束条件下最小化二次函数$\alpha^{\mathrm{T}}\mathrm{L}\alpha$，其中α表示由层的所有α值构成的矢量。该公式表明，与 L

的小特征值相关联的特征向量在创建高质量遮罩（matte）中起着重要作用。受此观察的启发，他们随后对光谱抠图的研究使用了 L 的特征向量来构建软分割。每个软分割是 K 个特征向量的线性组合，对应于 L 的最小特征值，它使得稀疏度最大化，即最小化部分不透明度的发生。通过最小化有利于α = 0 和α = 1 的能量函数来创建片段：

$$\arg\min_{\{y_i\}} \sum_{i,p} |\alpha_{ip}|^{\gamma} + |1-\alpha_{ip}|^{\gamma} \quad \text{其中} \alpha_i = Ey_i \tag{2a}$$

$$\text{s.t.} \quad \sum_i \alpha_{ip} = 1 \tag{2b}$$

其中α_{ip}是第 i 个片段的第 p 个像素的α值，E 是包含具有 L 的最小特征值的 K 个特征向量的矩阵，y_i是定义软分割的特征向量上的线性权重，并且γ <1 是控制稀疏先验强度的参数。

当图像包含具有不同颜色的单个识别良好的对象时，光谱抠图产生令人满意的结果，但它仍与更复杂的对象和场景斗争。由于它仅仅基于抠图拉普拉斯，只考虑小斑块的低级统计，所以它识别对象的能力有限。在我们的工作中 ,我们扩展这种方法以融合相同拉普拉斯公式中的语义特征 ,并捕获更高级别的概念，如场景对象，并对图像数据进行更广泛的查看。

亲和性和拉普拉斯矩阵（affinity and Laplacian matrices）

Levin 等人将他们的方法表示为直接推导出拉普拉斯矩阵的最小二乘优化问题。另一种方法是表达像素对之间的亲和性。具有正亲和性的对更可能具有相似的值，零亲和性对是独立的，具有负亲和性的对可能具有不同的值。在这项工作中，我们将使用相似性方法并使用众所周知的公式构建相应的归一化拉普拉斯矩阵：

$$L = D^{-\frac{1}{2}}(D-W)D^{-\frac{1}{2}} \tag{3}$$

其中 W 是包含所有像素对之间的亲和性的方阵，D 是对应的度矩阵（degree matrix），即具有元素 W 1 的对角矩阵，1 是全为 1 的行向量。正如 Levin 等人所指出的，由于存在负相关性，L 可能并不总是真正的拉普拉斯图，但仍然具有相似的性质，如半正定。

3.2 非局部颜色亲和性 （Nonlocal Color Affinity）

我们定义了一个基于颜色的长范围（longer-range）交互的附加的低级别关联项。一种简单的方法是在抠图拉普拉斯的定义中使用更大的补丁。然而，这个选项很快变得不切实际，因为它使拉普拉斯矩阵变得更密集。另一种选择是从非局部邻域采样像素以插入连接，同时保持矩阵中的一些稀疏度。 KNN 抠图和信息流抠图已经显示出对这种采样的中范围（medium-range）交互的良好结果。然而，这种策略面临稀疏性和鲁棒性之间的权衡：较少的样本可能会错过重要的图像特征，而更多的样本会使计算变得不易处理。

我们提出了基于图像的过度分割的引导式采样。我们使用 SLIC 生成 2500 个超像素并且估计每个超

像素与半径内的所有超像素之间的亲和性，该半径对应于图像尺寸的 20%。这种方法的优点是，每个足够大的特征都被表示为超级像素，稀疏度仍然很高，因为我们对每个超级像素使用一个样本，并且它通过使用大半径链接可能断开的区域，如当背景是通过一个物体的一个洞看到的时候。形式上，我们通过小于图像尺寸的 20%的距离定义两个超像素 s 和 t 的质心之间的颜色亲和性$W_{s,t}^{c}$：

$$W_{s,t}^{c} = \left(\mathrm{erf}\left(a_c\left(b_c - \|c_s - c_t\|\right)\right) + 1\right)/2 \qquad (4)$$

其中c_s和c_t是位于[0,1]的 s 和 t 的超像素的平均颜色，erf 是高斯误差函数，a_c和b_c是控制亲和性降低的速度和在哪里变为零的阈值的参数。erf 取[-1,1]中的值，这里使用的主要是它的 sigmoidal 形状。我们在所有结果中使用$a_c = 50$和$b_c = 0.05$。这种亲和性基本上确保具有非常相似颜色的区域在具有挑战性的场景结构中保持连接，其效果如图 3 所示。

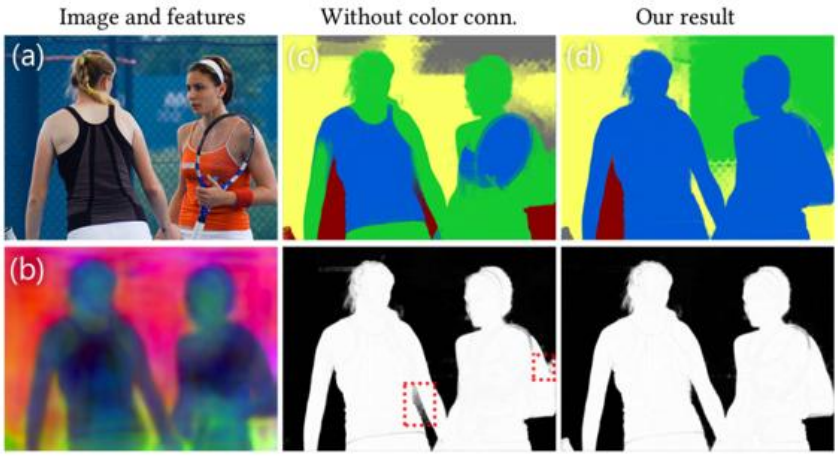

图 3

重点指出，我们所包含的基于颜色的非局部亲和性有助于分解恢复隔离区域，如断开的背景区域或长扩展。图片来自[Lin 等人 2014]。

3.3 高级语义亲和性

虽然非局部颜色亲和性为分割过程增加了长范围的交互，但它仍然是一个低级别的特征。我们的实验表明，在没有附加信息的情况下，分割仍然经常合并属于不同对象的相似颜色的图像区域。为了创建局限于语义相似区域的片段，我们添加了一个语义关联术语，即一个鼓励对属于同一场景对象的像素进行分组，而不鼓励对来自不同对象的像素进行分组的术语。我们在对象识别领域的前期工作的基础上，计算与底层对象相关的每个像素的特征向量。我们通过神经网络计算特征向量，如第 3.5 节所述。生成的特征向量使得属于同一对象f_p和f_q的两个像素 p 和 q 相似，即$\|f_p - f_q\| \equiv 0$，对于不同语义区域中的第三个像素 r，f_r是远离的， 即$\|f_p - f_q\| \ll \|f_p - f_r\|$。

我们还定义了超像素上的语义亲和性。除了增加线性系统的稀疏度之外，超像素的使用还减少了过渡区域中不可靠特征向量的负面影响，从图 4 中模糊的外观可以明显看出这一点。超像素边缘不直接用于线性系统，图中的连接是超像素质心之间的连接。然后，来自质心的该信息扩展到附近的像素，同时使用抠

图亲和项来维护图像边缘。对于每一个超级像素，利用这些向量和上一节（第 3.2 节）中相同的超距，我们将其平均特征向量$\tilde{f}_s$与其质心p_s相关联。我们使用这些向量来定义每个相邻超像素 s 和 t 之间的亲和项：

$$W^S_{s,t} = \mathrm{erf}\left(a_s\left(b_s - \|\tilde{f}_s - \tilde{f}_t\|\right)\right) \quad (5)$$

使用a_s和b_s参数控制亲和函数的陡度以及何时变为负值。我们将在 3.5 节讨论如何设置它们。定义负亲和性有助于图形断开不同对象，而正值连接属于同一对象的区域。

与颜色亲和性不同，语义亲和性只与附近的超像素相关，以有利于创建连接的对象。这种非局部颜色亲和性与局部语义亲和性的选择，允许创建可以覆盖同一语义连贯区域的空间中不连贯区域的层。这通常适用于通常出现在背景中的绿色和天空等元素，这使得它们可能由于遮挡而分裂为多个断开连接的组件。由于包含局部语义亲和性，L 的特征向量揭示了对象边界，如图 4 和图 5 所示。

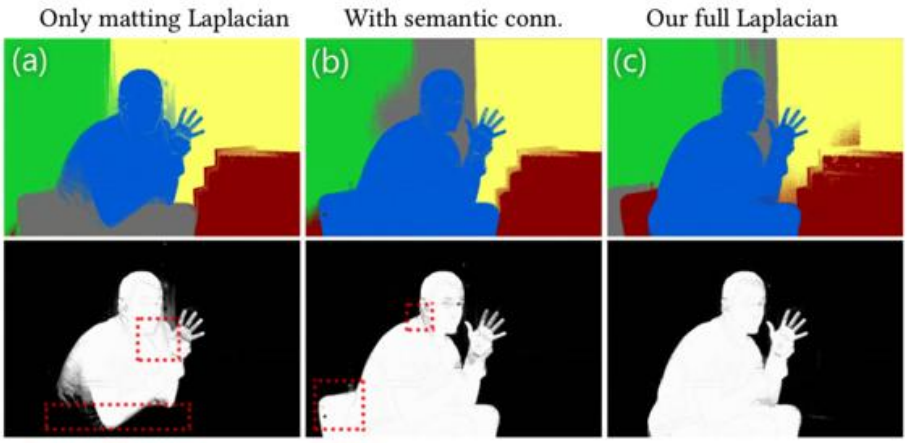

图 4

我们整个管道的结果只使用了抠图拉普拉斯(a)、抠图和语义拉普拉斯(b)以及两者加上稀疏颜色连接(c)，如图 5 所示。顶行显示每个生成的软分割的不同颜色，底行显示所提取的与人物对应的遮罩。由于特征向量无法表示人与背景之间的语义切割，因此仅使用抠图拉普拉斯会导致人的软分割包括大部分背景，如突出部分显示的那样。添加稀疏颜色连接可以提供更清晰的前景遮罩。

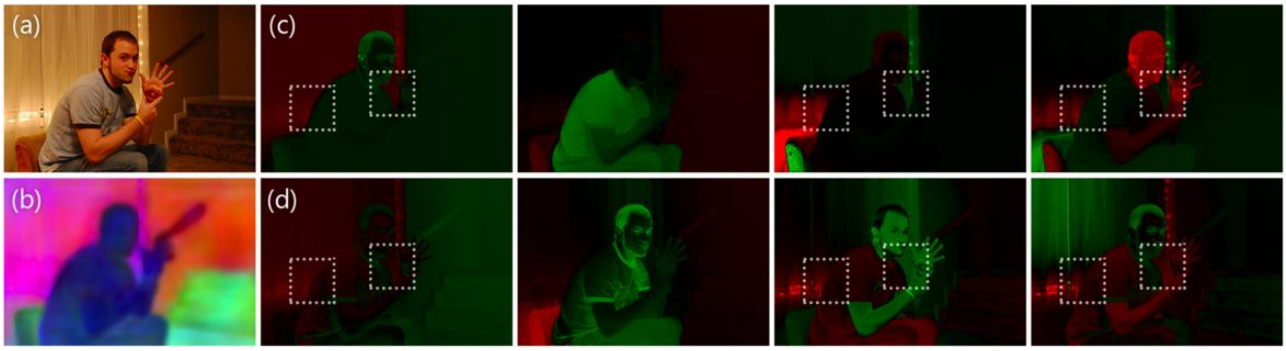

图 5

图像(a)和语义特征(b)用几个特征向量表示，对应于所提出的拉普拉斯矩阵的最小特征值（c，顶行），以及用于光谱抠图的抠图拉普拉斯的最小特征值（d，底行）。绿色表示特征向量的正值，而红色表示负值。我们的拉普拉斯矩阵强烈地揭示了特征向量中的语义切割，而抠图拉普拉斯的特征向量超出了语义边缘，如突出部分显示的那样。图片来自[Lin 等人 2014]。

3.4 创建图层

我们通过使用本节前面描述的亲和性来创建层，以形成拉普拉斯矩阵 L。我们从该矩阵中提取特征向量，并使用两步稀疏过程从这些特征向量创建层。

形成拉普拉斯矩阵（Forming the Laplacian matrix）

我们通过将亲和性矩阵加在一起并使用（3）形成拉普拉斯矩阵 L：

$$L = D^{-\frac{1}{2}}\left(D - (W_L + \sigma_S W_S + \sigma_C W_C)\right)D^{-\frac{1}{2}} \quad (6)$$

其中W_L是包含抠图亲和性的矩阵，W_C是包含非局部颜色亲和性的矩阵（第 3.2 节），W_S是具有语义亲和性的矩阵（第 3.3 节），以及控制每个术语影响程度的σ_S和σ_C参数，两者都是设为 0.01。

受约束的稀疏化（constrained sparsification）

我们提取对应于 L 的 100 个最小特征值的特征向量。我们使用 Levin 等人的优化程序形成一组中间层。在（2）中，γ=0.8。与在特征向量上使用 k 均值聚类来初始化优化的光谱抠图不同，我们对由其特征向量 f 表示的像素使用 k 均值聚类。这个初始猜测与场景语义更加一致，并产生更好的软分割。我们用这种方法生成了 40 个层，实际上，它们中的一些都是零，留下 15 到 25 个非平凡层（nontrivial layer）。我们通过在由它们的平均特征向量表示的这些非平凡层上运行 k = 5 的 k 均值算法来进一步减少层数。这种方法比将 100 个特征向量直接稀疏化为 5 层更有效，因为这种急剧减少会使问题过度约束并且不能产生足够好的结果，特别是在遮罩稀疏性方面。分组前后的初始估计软分割如图 7 所示。不失一般化地我们将段数设置为 5;虽然这个数字可以由用户根据场景结构设置，但我们观察到它对于大多数图像来说是合理的数字。因为这 5 个层被约束在有限数量的特征向量的子空间内，所以实现的稀疏性是次优的，在层中留下许多半透明区域，这在普通场景中是不可能的。接下来，我们介绍一个稀疏化过程的放宽的版本来解决此问题。

放宽的稀疏化（relaxed sparsification）

为了改善层的稀疏性，我们放宽了它们是特征向量的线性组合的约束。不使用线性组合的系数y_i（2），在该步骤中，每个单独的α值是未知的。我们定义了一个能量函数，它可以在像素级别上提升遮罩稀疏度，同时考虑受约束的稀疏化和图像结构的初始软分割估计。我们现在一个术语一个术语地定义我们的能量项。

第一项术语放宽了子空间约束，并且仅确保生成的层保持靠近使用约束稀疏化过程创建的层$\hat{\alpha}$：

$$E_F = \sum_{ip} \left(\alpha_{ip} - \hat{\alpha}_{ip}\right)^2 \quad (7)$$

我们还将放宽到合计为一个要求（1b），将其作为软约束集成到线性系统中：

$$E_C = \sum_{p} \left(1 - \sum_{i} \alpha_{ip}\right)^2 \quad (8)$$

其中α_{ip}是第 i 层中第 p 个像素的α值。下一个术语是拉普拉斯 L 定义的能量，定义了（6）中定义的信息

的空间传播：

$$E_L = \sum_i \alpha_i{}^T L \alpha_i \quad (9)$$

最后，我们制定一个适应图像内容的稀疏项。直观地，部分不透明度来自图像中的颜色过渡，因为在许多情况下，它对应于两个场景元素之间的过渡，例如，泰迪熊和背景之间的模糊过渡。我们使用这种观察来建立一个空间变化的稀疏能量：

$$E_S = \sum_{i,p} |\alpha_{ip}|^{\tilde{\gamma}_p} + |1-\alpha_{ip}|^{\tilde{\gamma}_p} \quad (10a)$$

$$\text{其中} \quad \tilde{\gamma}_p = \min\left(0.9 + \|\nabla_{Cp}\|, 1\right) \quad (10b)$$

其中∇_{Cp}是使用 Farid 和 Simoncelli 的可分离内核计算的像素 p 处的图像中的颜色梯度。我们设计了这样一个术语，使得当梯度足够大的图像区域上的$\tilde{\gamma}p = 1$时，$\alpha_{ip} \in [0,1]$的能量分布是平坦的，即能量仅对有效范围之外的值起惩罚作用，α_{ip}取 0 到 1 之间的任何值。相比之下，在$\nabla_{Cp} \approx 0$的均匀区域中，它鼓励α_{ip}为 0 或 1。这两种效果相结合，有利于更高水平的稀疏度和不透明度过渡的柔和度。我们的空间变化稀疏能量对保持精确软过渡的影响可以在图 6(c,d)中看到。

把这些术语放在一起，我们得到了能量函数

$$E = E_L + E_S + E_F + \lambda E_C \quad (11)$$

每个项的单位权重都是有效的，除了表示具有较高权重λ=10 的软约束的总和项 E_C。没有稀疏项E_S，E 将是标准的最小二乘能量函数，可以通过求解线性系统使其最小化。为了处理E_S，我们采用迭代重加权最小二乘求解器，通过求解一系列线性系统来估计解。我们将在本节的其余部分介绍此方法的详细信息。

我们将层数命名为$N_i = 5$，像素数命名为N_p，将所有α_i的矢量命名为 a，所有$\hat{\alpha}_i$的矢量命名为$\hat{a}$。a 和$\hat{a}$的维数是$N_{ip} = N_i N_p$。为清楚起见，我们还引入了$N_{ip} \times N_{ip}$的单位矩阵 I。利用这个符号，我们以矩阵形式重写E_F（7）：

$$E_F = (a-\hat{a})^T I (a-\hat{a}) \quad (12)$$

我们在这个方程中加入了多余的 I，以便在推导式 17 时得到更清晰的过渡。为了重写E_C（8），我们引入了通过水平连接N_i个单位矩阵得到的$N_i \times N_{ip}$矩阵 C、由N_i个 1 得到的向量1_i和由N_{ip}个 1 得到的向量1_{ip}：

$$E_C = (1_i - Ca)^2 = a^T C^T C a - a^T C^T 1_i - 1_i{}^T C a + 1_i{}^T 1_i \quad (13a)$$

$$= a^T C^T C a - 2 a^T 1_{ip} + N_i \quad (13b)$$

我们使用$a^T C^T 1_i = 1_i{}^T C a$，$C^T 1_i = 1_{ip}$和$1_i{}^T 1_i = N_i$。然后我们重写$E_L$：

$$E_L = a^T \tilde{L} a \quad (14a)$$

$$\tilde{L} = \begin{bmatrix} L & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & L \end{bmatrix} \quad (14b)$$

对于稀疏项E_S，我们引入近似能量：

$$\tilde{E}_S = \sum_{i,p} u_{ip}(\alpha_{ip})^2 + v_{ip}(1-\alpha_{ip})^2 \quad (15a)$$

$$\text{其中：}\ u_{ip} = \left|\alpha^{\circ}_{ip}\right|^{\tilde{\gamma}p-2} \text{和}\ v_{ip} = \left|1-\alpha^{\circ}_{ip}\right|^{\tilde{\gamma}p-2} \qquad (15b)$$

其中α°等于第一次迭代时的约束稀疏化结果，并且等于之前迭代的解。我们使用u_{ip}值构建的对角矩阵D_u、v 和D_v向量、以及使用v_{ip}值构建的对角矩阵，重写了矩阵公式：

$$\tilde{E}_S = a^T D_u a + (1_{ip} - a)^T D_v (1_{ip} - a) \qquad (16a)$$

$$= a^T (D_u + D_v) a - 2a^T v + {1_{ip}}^T v \qquad (16b)$$

我们使用$D_v 1_{ip} = v$和$v^T a = a^T v$。

为了得到一个线性系统，我们把所有能量项以矩阵形式求和，并对 a 的导数最小化为零。这导致：

$$(\tilde{L} + D_u + D_v + I + \lambda C^T C)a = v + \hat{a} + \lambda 1_{ip} \qquad (17)$$

我们使用预条件共轭梯度优化来解决这个方程[Barrett 等人 1994]。在我们的实验中，20 次迭代产生的结果具有令人满意的稀疏性。图 6 说明了我们的方法的好处。

线性系统的尺寸是$N_i N_p$。虽然这很大，但它仍然易于处理，因为软层N_i的数量设置为 5 并且接近于块对角线，对角线以外的唯一系数来自于贡献$C^T C$到系统的总和一项E_C。由于 C 由 5 个并置的$N_p \times N_p$单位矩阵构成，因此$C^T C$由 5×5 布局的 25 个$N_p \times N_p$个单位矩阵构成，即它非常稀疏并且易于由求解器处理。

> 译者注：
>
> 我计算的E_C为：
>
> 通过水平连接N_i个$N_p \times N_p$的单位矩阵得到$N_i \times N_{ip}$矩阵 C、由N_p个 1 得到的向量1_p和由N_{ip}个 1 得到的向量1_{ip}
>
> $$E_C = (1_p - Ca)^2 = a^T C^T Ca - a^T C^T 1_p - {1_p}^T Ca + {1_p}^T 1_p = a^T C^T Ca - 2a^T 1_{ip} + N_p$$
>
> 我们使用$a^T C^T 1_p = {1_p}^T Ca$，$C^T 1_p = 1_{ip}$和${1_p}^T 1_p = N_p$。
>
> 与论文不同，计算结果与论文相同的小伙伴请赐教

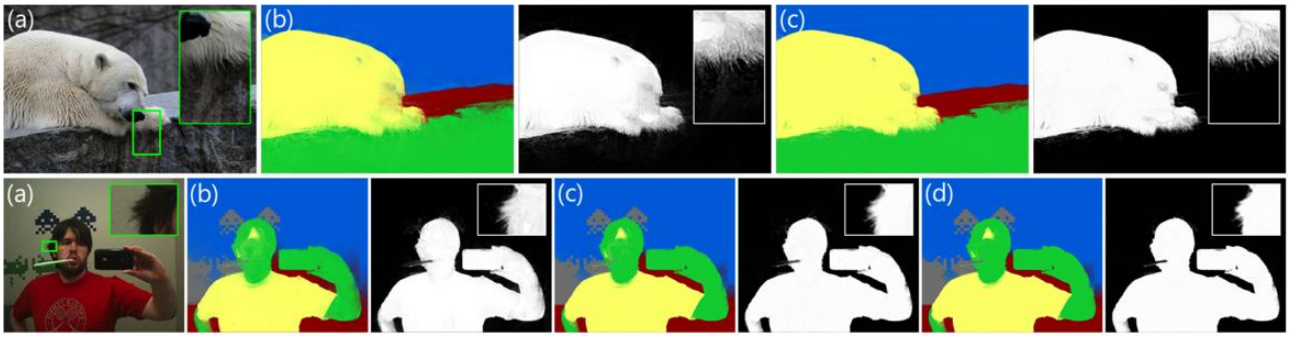

图 6

图像(a)与像素级稀疏化前(b)和后(c)的结果一起显示。颜色编码的段显示为与前景对象对应的单个 alpha 通道。最后一步是清除由于特征向量的表达能力有限当保留软过渡有限而产生的虚假 alpha 值。底部示例还具有使用常数 0.9 作为稀疏度参数 γ(d)的稀疏化结果，而我们使用空间变化的γp 来放宽过渡区域的稀疏度约束。这一效果可以在插图中看到，因为我们的结果(c)保留了头发周围的软过渡，而常量(d)导致结果过于稀疏。图片来自 Lin 等人。

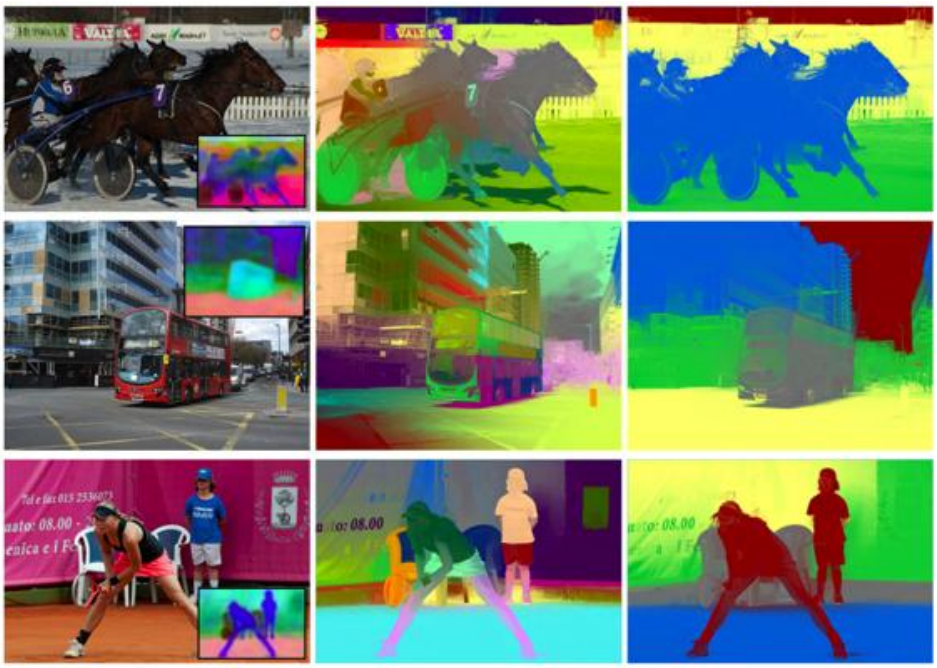

图 7

输入图像和计算出的语义特征显示为初始估计的多层软分割（中间）和分组后的中间软分割（右侧）。通过为每个段指定纯色，可以可视化软分割。请注意，这些是通过进一步放宽稀疏度而完善过的结果。图片来自[Lin 等人 2014]。

3.5 语义特征向量

我们定义了语义关联项（第 3.3 节），其中特征向量 f 对于同一对象上的像素是相似的，而对于不同对象上的像素则是不相似的。可以使用针对语义分段训练的不同网络架构来生成这样的向量。在我们的实现中，我们结合了一个语义分割方法和一个用于度量学习的网络。应该注意的是，我们并不声称特征生成是一种贡献，我们只总结了我们在本节中使用的解决方案。补充材料中提供了详细说明。

我们的特征提取器的基础网络基于 DeepLab-ResNet-101，但它采用度量学习方法[Hoffer 和 Ailon 2015]进行训练，以最大化不同对象的特征之间的 L2 距离。在 Hariharan 和 Bertasius 等人的激励下，我们结合了网络的多个阶段的特性，基本上结合了中高级功能。我们不是在训练时使用图像的所有像素，而是为所有像素生成特征，但只使用一组随机采样的特征来更新网络。网络最小化具有相同真值类别的样本的特征之间的距离，并且最大化其他距离。由于我们仅使用该提示，即两个像素是否属于同一类别，因此在训练期间不使用特定对象类别信息。因此，我们的方法是一种类不可知的方法。这适用于我们语义软分段的总体目标，因为我们的目标是创建覆盖语义对象的软分割，而不是图像中对象的分类。为了利用更多具有计算效率的数据，我们使用了稍微修改过的 N 对损失版本[Sohn 2016]。

我们在 COCO-Stuff 数据集的语义分割任务上训练这个网络。我们使用引导滤波器优化由该网络生成的特征映射使其与图像边缘良好对齐。然后，我们使用主成分分析（PCA）将维数降低到三。这些预处理步骤在图 8 中可视化。虽然原始的 128 维向量可以很好地覆盖我们可能遇到的所有内容，但是每个图像仅展示其中的一小部分，因此降低维数可以提高每个维度更好的精度。最后，我们对向量进行规范化，以获

取[0,1]中的值。这使得设置参数更容易，尤其是在改变特征向量定义的情况下。对于我们给出的所有结果，我们将等式 5 中的 a_s 和 b_s 分别设置为 20 和 0.2。

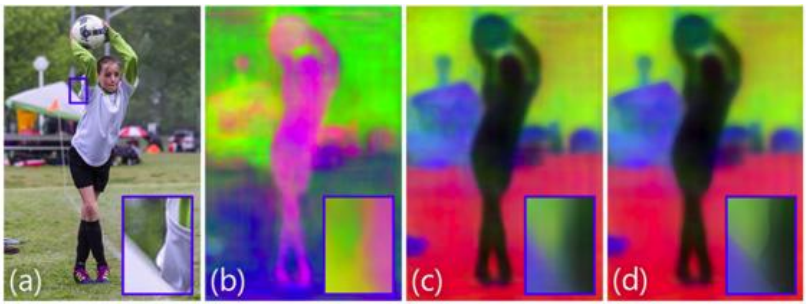

图 8

我们首先为给定图像生成每像素 128 维特征向量(a)。在(b)中示出了 128 维到 3 维的随机投影。我们使用每个图像的主成分分析将特征的维数降低到 3(c)。在降维之前，我们使用引导过滤器对特征进行边缘对齐。图片来自[Lin 等人 2014]。

3.6 实施细节

我们使用 MATLAB 中提供的稀疏特征分解和直接求解器来实现我们算法的约束稀疏化阶段的概念验证实现。对于 640×480 图像，此步骤大约需要 3 分钟。松弛稀疏步骤采用 MATLAB 的预处理共轭梯度优化实现。每次迭代通常在 50 到 80 次迭代中收敛，并且该过程大约需要 30 秒。我们算法的运行时间随着像素数的增加呈线性增长。

4.实验分析

语义软分割是语义分割、自然图像抠图和软分割的交叉点，其数值评价具有一定的挑战性。语义分割数据集提供的二进制标记并不总是像素精确，这使得它们不适合于基准语义软分割。自然图像抠图方法通常在专用基准和数据集上进行评估。这些基准被设计用来评估使用辅助输入的方法，称为三值图，它定义了预期的前景和背景，以及不确定的区域。

此外，我们工作的语义方面超出了这些基准的范围。另一方面，软色分割是一个缺乏对基本事实的可靠定义的问题。虽然 Aksoy 等人提出了几个用于评估的盲指标，它们专门用于软色分割，也忽略了语义方面。因此，我们采用与相关方法的定性比较，并讨论各种方法之间的特征差异。

4.1 光谱抠图和语义分割

在图 9 和图 10 中，我们将我们的结果与作为我们最相关的软分割方法的光谱抠图以及两种最先进的语义分割方法一起显示出来：Zhao 等人的场景分析方法（pspnet）和 He 等人的实例分割方法（mask r-cnn）。更多这些比较可在补充材料中找到。光谱抠图每张图像生成大约 20 个软分割，并通过组合软分割以最大化物体分数提供几种可选前景遮罩。这些遮罩不是准确的结果，而是作为选项提供给用户，并且

显示所有 20 个分段将使得比较更难以评估。相反，我们应用我们的软分割分组方法，该方法将语义特征用于光谱抠图的结果。

实例表明，语义分割方法虽然在识别和定位图像中的目标方面取得了成功，但在目标边缘的定位精度较低。虽然它们的精度可以满足语义分割任务的要求，但对于图像编辑或合成应用程序来说，对象边缘的误差是个问题。在光谱的另一端，光谱抠图能够成功地捕捉到物体周围的大多数软过渡。然而，由于缺乏语义信息，它们的片段通常同时覆盖多个对象，并且对于任何给定对象，alpha 值通常不稀疏。相比之下，我们的方法捕获对象的整体或子对象而不对不相关的对象进行分组，并在边缘处实现高精度，包括适当时的软过渡。

应该注意的是，我们的方法在多个段中表示相同的对象并不少见，例如图 9(2)中的马车或图 9(4)中的背景栏。这主要是由于预设的层数为五，有时超过图像中有意义的区域的数量。尽管被语义特征检测到一些小对象，但是在最终片段中可能遗会丢失，例如图 10(5)中的背景中的人。这是因为，特别是当物体的颜色与周围环境相似时，物体在特征向量中看起来没有很好地定义，并且它们最终被合并为近似段。我们的语义特征不是实例感知的，即同一类的两个不同对象的特征是相似的。这导致多个对象在同一层中表示，例如图 9(1)中的奶牛，图 9(5)中的人或图 10(3)中的长颈鹿。但是，使用实例感知功能，我们的方法将能够为不同的对象实例生成单独的软分割。

对于软分割和图像抠图方法而言，灰度图像尤其具有挑战性，因为缺乏这种方法通常依赖的颜色提示。另一方面，语义分割方法的性能在处理灰度图像时不会显着降低。图 10(5)表明我们的方法可以成功地利用语义信息进行灰度图像的软分割。

4.2 自然图像抠图

原则上，语义软分割可以通过级联语义分割和自然图像抠图来生成。定义前景，背景和软过渡区域的三值图可以从语义硬片段生成，并将其输入到自然抠图方法。Shen 等人和 Qin 等人对类特定问题使用类似的方法。我们在图 11 中展示了这种情景的两个例子，通过使用 Mask R-CNN 和 PSPNet 的结果生成三值图并使用最先进的抠图方法信息流抠图，估计遮罩来证明这种方法的缺点。通过自然图像抠图方法做出的强烈假设是所提供的三值图是正确的，即，定义的前景和背景区域被用作硬约束以指导方法来对层颜色进行建模。然而，估计的语义边界的不准确性通常无法提供可靠的三值图，即使具有大的未知区域宽度。如图中突出显示的，这导致抠图结果中出现的严重伪影。我们展示了使用我们的演示结果生成的准确三值图，使得自然抠图方法成功。

虽然一般的自然图像抠图超出了我们方法的范围，但图 12 显示了几个示例，其中我们的方法能够在自然图像抠图数据集的图像上生成令人满意的结果而无需三值图。

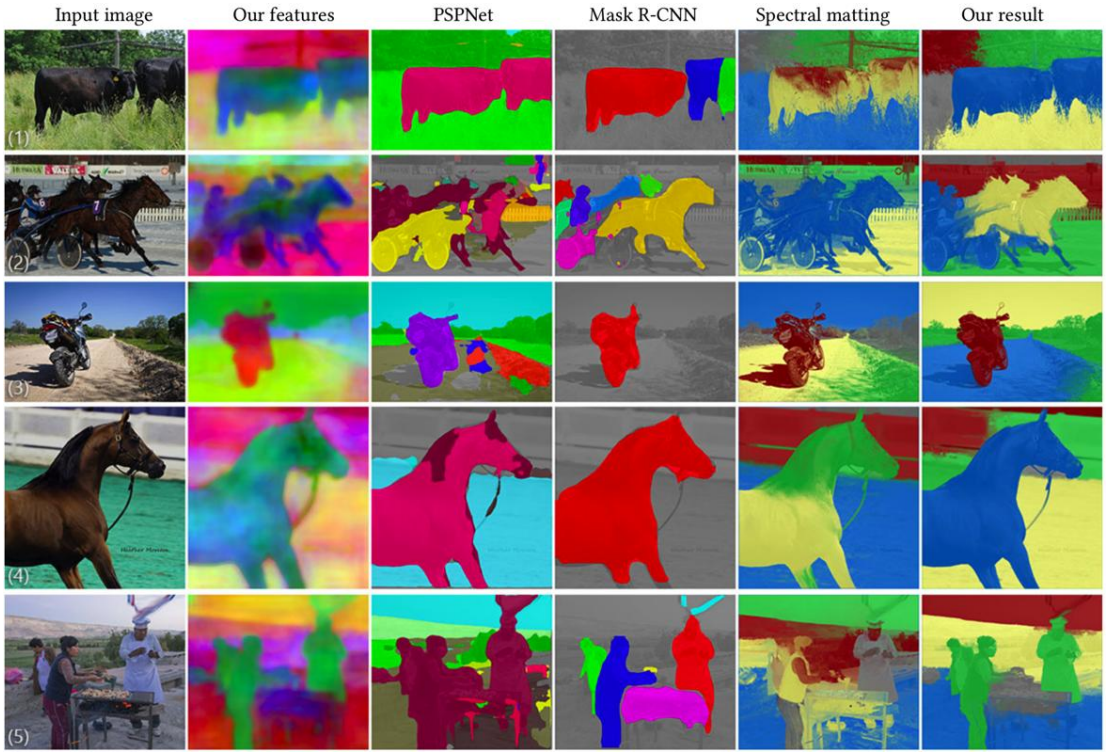

图 9

我们将结果与 Zhao (PSPNet)、He(Mask R-CNN)等人和光谱抠图的结果一起显示。分段被覆盖到图像的灰度版本上，以便更好地评估分段边界。注意 PSPNet 和 Mask R-CNN 在对象边界的不准确性，以及超出物体边界的光谱铺垫的软分割。图片来自[Lin 等人 2014]。

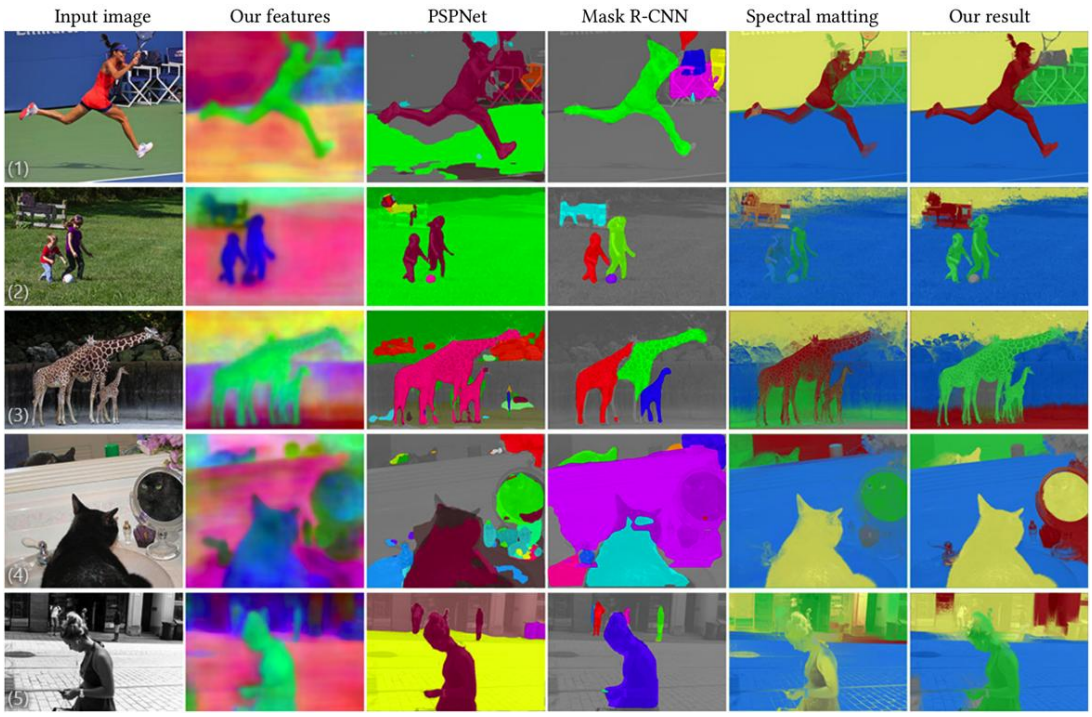

图 10

我们将结果与 Zhao (PSPNet)、He(Mask R-CNN)等人和光谱抠图的结果一起显示。分段被覆盖到图像的灰度版本上，以便更好地评估分段边界。注意 PSPNet 和 Mask R-CNN 在对象边界的不准确性，以及超出物体边界的光谱铺垫的软分割。图片来自[Lin 等人 2014]。

4.3 软色分割

软色分割，最初由 Tai 等人提出的概念，将输入图像分解为均匀颜色的软层，并且已被证明对图像编辑和重新着色应用有用。作为语义软分隔和软色段之间的概念比较，图 13 显示了基于解混的软颜色分割的段。为了更方便的定性比较，我们使用封闭形式的颜色估计方法估算软分割的图层颜色。

可以立即看到软颜色段的内容超出对象边界，而我们的结果显示同一段中具有语义意义的对象，无论其颜色内容如何。由于这些表示彼此正交，因此可以在编排中使用它们来生成目标重新着色结果。

4.4 使用语义软分割进行图像编辑

我们在图 14 中展示了用于目标图像编辑和合成的软分割的几个用例。图 14(1,3,4,7)显示了合成结果，我们使用闭合层颜色估算了我们的段的层颜色估计。注意所选前景层和新背景之间的自然柔和过渡。软分割还可用于目标图像编辑，用于定义特定调整层的遮罩，例如在图 14(2)中为火车添加运动模糊，在图 14(5,6)中分别为人员和背景进行颜色分级，以及在图 14(8)中单独设置热气球、天空、地形和人员的样式。虽然这些编辑可以通过用户绘制的蒙版或自然抠图算法完成，但我们的表示提供了方便的中间图像表示，使艺术家可以毫不费力地进行目标编辑。

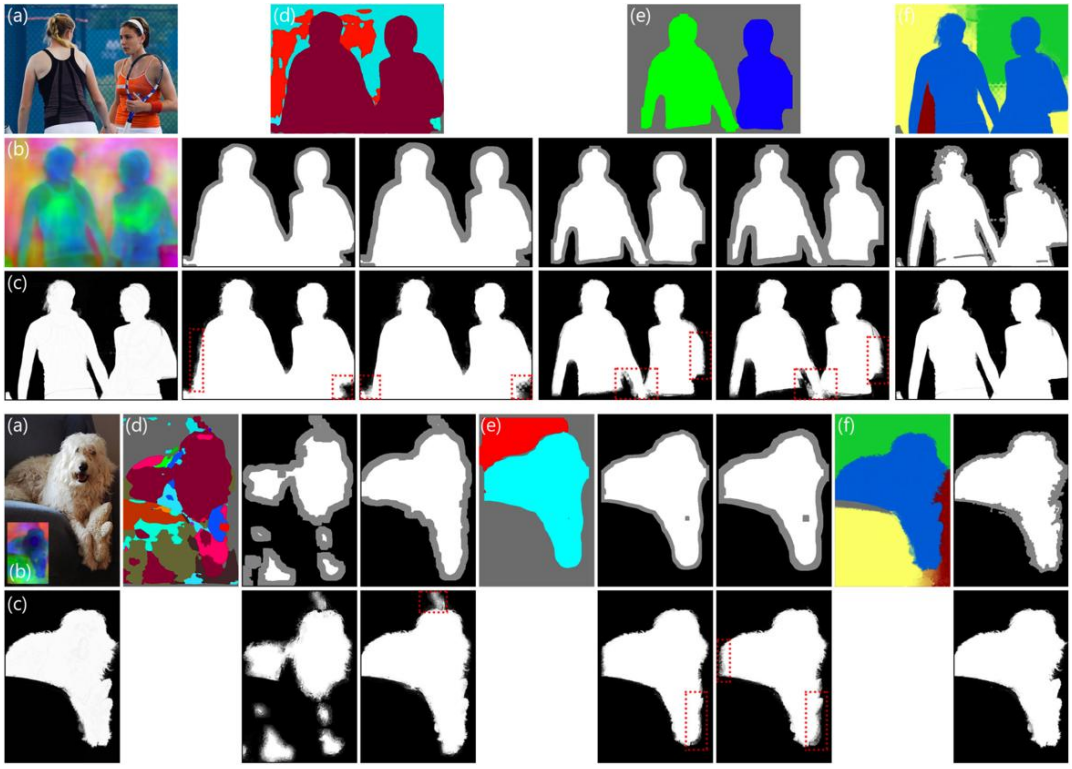

图 11

从输入图像(a)和我们的特征向量(b)，我们的方法生成(c)中所示的遮罩。我们发现，由 PSPNet(d)或 Mask(e)使用语义段生成的具有不同未知区域宽度的三值图，未能可靠地提供前景和背景区域，这会对信息流抠图生成的抠图结果产生负面影响。在底部示例中，通过选择单个类（左）或与该对象对应的所有类来生成 PSPNettrimap。我们还提供适用由我们的结果(f)生成的三值图而产生的抠图结果，其中三值图证明了给定精确三值图下抠图算法的性能。

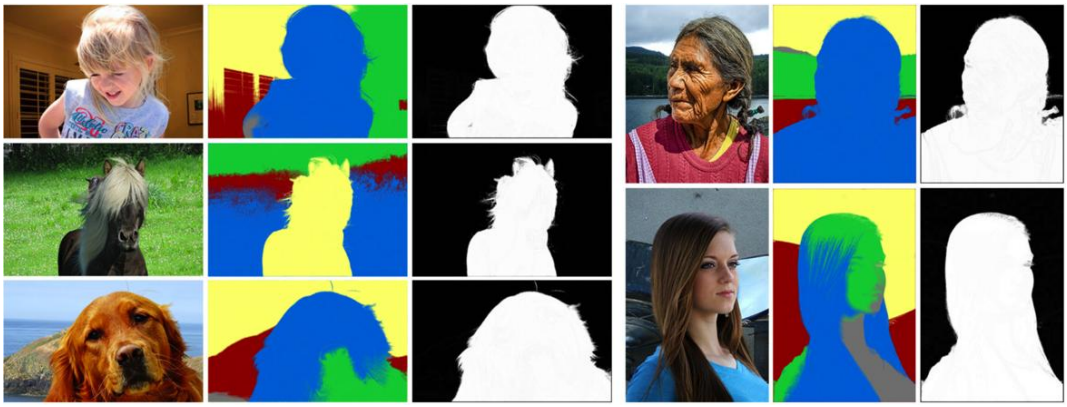

图 12

我们的软分割和前景对象的相应遮罩。请注意 ,通常为自然抠图提供的三值图不会用于产生这些结果。来自[Xu 等人 2017]。

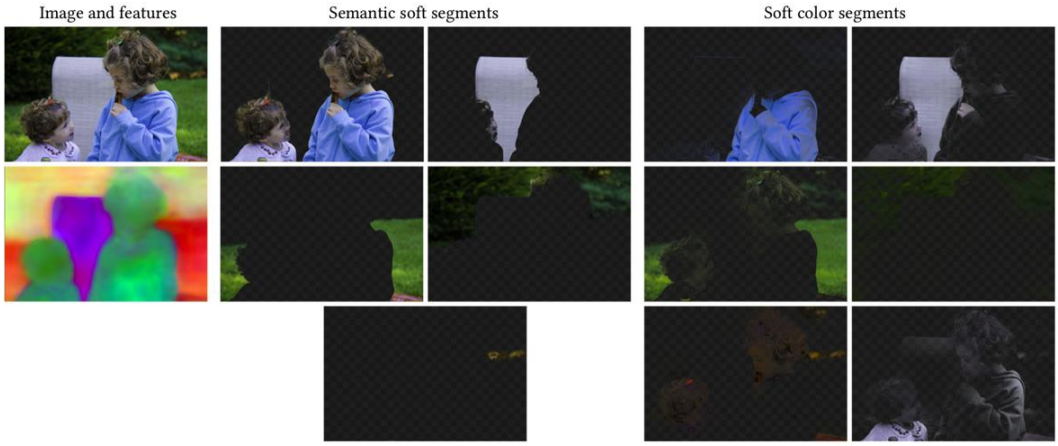

图 13

提出的方法将语义软分割与 Aksoy 等人的软色段结合起来进行概念比较。。这两种方法都是完全自动化的，只需要输入图像进行软分割。来自[Bychkovsky 等人 2011]。

图 14

我们展示了软分割结果以及使用每层操作或简单组合生成的图像编辑结果，以演示在目标图像编辑任务中使用我们的分割。图片来自[Bychkovsky 等人](1)和 Death to the Stock Photo。

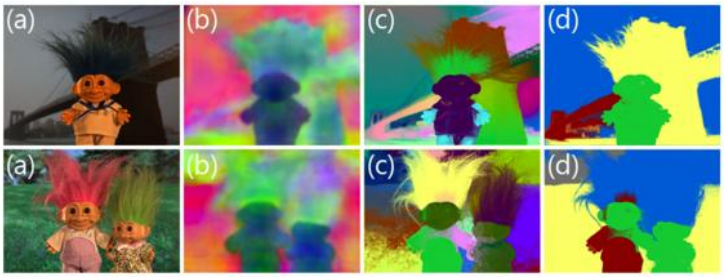

图 15

显示了两种故障情况。顶部示例：如果大区域覆盖具有非常相似颜色的不同对象(a)我们的特征向量(b)和分组前的分段(c)无法识别图像中的单独对象并导致不准确的分割(d)。下面的示例：当我们的特征向量无法表示对象时，即使初始层能够生成准确的软过渡(c)，软分割(d)的分组也可能失败。图片来自[Rhemann]。

5.限制和未来的工作

虽然我们能够生成精确的图像软分割，但在我们的原型实现中，我们的求解器并未针对速度进行优化。因此，我们在一张 640×480 图像上的运行时间在 3 到 4 分钟之间。我们的方法的效率可以通过多种方式进行优化，例如多尺度求解器，但线性求解器和特征分解的有效实现超出了本文的范围。

在约束稀疏步骤中，我们生成大约 15-25 个片段，然后使用特征向量将这些片段分组为 5 个片段。层数是通过经验观测确定的，在某些情况下，一个物体可以分为若干层。虽然这不会影响我们方法的适用性，因为在编辑中合并这些层是很简单的，但是可以设计更复杂的方法来对层进行分组，例如通过识别和分类。

我们的方法不会为同一类对象的不同实例生成单独的图层。这是由于我们的特征向量，它不提供实例感知语义信息。然而，我们的软分割公式对语义特征是不可知的。因此，更高级的特征生成器可以在与更合适的分段分组策略相结合，生成实例级软分割结果。

我们已经从自然抠图数据集中显示了几个结果。但是，应该指出的是，我们的目的并不是解决一般的自然抠图问题。自然抠图是一个成熟的领域，具有许多特定的挑战，例如在非常相似的前景和背景区域周围生成精确的遮罩，最先进的方法取决于两个区域的颜色分布，以提高这些区域周围的性能。如图 15 所示，当对象颜色非常相似时，我们的方法可能在初始约束稀疏步骤中失败，或者由于大型过渡区域的不可靠的语义特征向量，软分割的分组可能会失败。

6.结论

我们提出了一种方法，通过将神经网络中的高级信息与低级图像特征完全自动融合，生成与图像中有语义区域相对应的软分割。我们已经表明，通过仔细定义图像中不同区域之间的亲和性，可以通过对构造的拉普拉斯矩阵的谱分析来揭示具有语义边界的软分割。所提出的用于软分割的松弛稀疏化方法可以产生精确的软过渡，同时还提供稀疏的层组。我们已经证明，虽然语义分割和光谱软分割方法无法提供足够精确的层来执行图像编辑任务，但我们的软分割提供了一种方便的中间图像表示，使多个目标图像编辑任务变得微不足道，否则需要熟练的艺术家人工操作。