符号图推理遇到卷积

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada

Xiaodan Liang[1], **Zhiting hu**[2] , **Hao Zhang**[2] , **Liang Lin**[3] , **Eric P. Xing**[4]
[1] School of Intelligent Systems Engineering, Sun Yat-sen University
[2]Carnegie Mellon University
[3] School of Data and Computer Science, Sun Yat-sen University
[4]Petuum Inc.
xdliang328@gmail.com, {zhitingh,hao, epxing}@cs.cmu.edu, linliang@ieee.org

摘要

除了局部卷积网络，我们还将探索如何利用各种外部人类知识为网络赋予语义全局推理能力。我们提议使用新的符号图推理（SGR）层，而不是使用单独的图形模型（例如 CRF）或约束来为更广泛的依赖关系建模，该层对一组符号节点执行推理，这些符号节点的输出显式表示先前知识图中每个语义的不同属性。为了与局部卷积配合，每个 SGR 由三个模块组成：a）一个原始的局部语义投票模块，其中所有符号节点的特征都是通过从局部表示中投票产生的； b）图推理模块在知识图上传播信息以实现全局语义一致性； c）语义到本地的双重映射模块学习了进化的符号节点与本地表示的新关联，并因此增强了本地特征。可以在任何卷积层之间注入 SGR 层，并使用不同的先验图进行实例化。大量实验表明，结合 SGR，可以在三个语义分割任务和一个图像分类任务上显著改善普通的 ConvNets。更多分析表明，SGR 层在通用知识图的情况下使用不同的标签集学习域/数据集的共享符号表示，这证明了其出色的泛化能力。

1 引言

尽管通过卷积网络实现了标准识别任务（例如图像分类[12]和分割[6]）方面的重大进步，但主要的范式在于更深、更复杂的局部卷积的堆栈中，我们希望它能捕获有关输入和输出之间关系的所有信息。但是，这样的网络损害了功能的可解释性，并且也缺乏对于复杂的现实世界任务至关重要的全局推理能力。因此，一些工作[51，41，5]制定了图形模型和结构约束（例如 CRF [22，19]），作为影响最终卷积预测的循环工作。但是，它们不能显式增强特征表示，从而导致泛化能力有限。最近的胶囊网络[39，14]扩展为学习跨位置的知识共享以查找特征簇，但它只能利用隐式且不可控制的特征层次。正如[3]中强调的那样，对外部知识的视觉推理对于人类的决策至关重要。缺少对上下文和高级语义的显式推理将阻碍卷积网络大型概念词汇中识别对象方面（探索概念相关性和约束性在其中起重要作用）的进步。另一方面，结构化知识提供了丰富的线索来使用符号词（例如名词或谓词）记录人类的观察和常识。因此，希望将符号语义与学习到的局部特征表示进行桥接，以实现更好的图形推理。

在本文中，我们探索了如何将丰富的常识性人类知识[33，53]纳入到局部卷积之外的中间特征表示学习中，并进一步实现全局语义一致性。常识性人类知识可以形成为各种无向图，这些图由概念之间的丰富关系（例如，语义层次，空间/动作交互和属性，并发）组成。例如，“设得兰群岛牧羊犬”和“哈士奇犬”由于某些共同特征而共享一个超类“狗”；人们戴着帽子弹吉他而不是相反；橙色是黄色。将结构化知识与视觉领域相关联后，这些符号实体（例如狗）可以与图像中的视觉证据相关联，因此人类可以整合视觉外观和常识来帮助识

别。

我们试图模仿这种推理过程，并将其集成到卷积网络中，也就是说，首先通过对局部特征进行投票来表征不同符号节点的表示；然后通过图传播进行图推理以增强这些符号节点的视觉证据，从而实现语义一致性。最终将符号节点的演化特征映射回以方便每个局部表示。我们的工作将超越现有方法，迈出了重要的下一步，因为它将对外部知识图的推理直接整合到局部特征学习中，称为符号图推理（SGR）层。请注意，这里我们使用“符号”来表示具有明确语言学意义的节点，而不是用于图形模型或图形神经网络的常规/隐藏图节点[40，18]。

SGR 层的核心包括三个模块，如图 1 所示。首先，每个符号节点的个性化可视化证据可以通过从所有本地表示形式进行投票来产生，称为本地到语义投票模块。投票权重代表每个局部特征对某个节点的语义一致性。其次，给定一个先验知识图，图推理模块被实例化以在该图上传播信息，以演化所有符号节点的视觉特征。最后，双重语义到局部模块学习了进化的符号节点与局部特征之间的适当关联，以结合局部和全局推理的力量。因此，它使特定符号节点的发展知识仅能借助全局推理来驱动对语义兼容的局部特征的识别。

我们的 SGR 层的主要优点在于三个方面：a）通过常识知识促进的局部卷积和全局推理可以通过学习特定于图像的观测值与先验知识图之间的关联来进行协作； b）每个局部特征都通过其相关的传入局部特征得到增强，而在标准局部卷积中，它仅基于其自身传入特征与学习的权重向量之间的比较； c）受益于所学的通用符号节点表示，所学的 SGR 层可以容易地转移到具有不同概念集的其他数据集域。 SGR 层可以插入任何卷积层之间，并根据不同的知识图进行个性化设置。

广泛的实验通过结合我们的 SGR 层，显示出优于普通 ConvNets 的性能，尤其是在三个语义分割数据集（COCO-Stuff，ADE20K，PASCAL-Context）和图像分类数据集（CIFAR100）中识别大型概念词汇时。当将 SGR 层训练的一个域转移到其他域时，我们进一步证明了其有希望的泛化能力。

2 相关工作

探索卷积网络上下文建模的最新研究可以分为两类流派。一个流派使用一系列基于图的 CNN [36，40]和 RNN [25，26]或高级卷积滤波器[43]来利用网络进行图结构化数据的发现，以发现更复杂的特征依存关系。在卷积网络的情况下，可以通过对基本卷积的最终预测起作用[51、41、5]，将诸如条件随机场（CRF）[22、19]之类的图形模型表达为递归网络。相比之下，建议的 SGR 层可以视为简单的前馈层，可以将其注入到任何卷积层之间，并通用于任何网络，以进行大规模和语义相关的识别。我们的工作不同之处在于，将局部特征映射到有意义的符号节点中。位置上的全局推理直接与外部知识对齐，而不是与隐式特征簇对齐，这是引入结构约束的更有效且可解释的方式。

另一个流派将外部知识库探索为便利网络。例如，邓等人[9]使用标签关系图来指导网络学习，而 Ordonez 等人学习了从通用概念到入门级概念的映射。一些工作通过求助于在最终预测分数上的复杂图形推断[9]，层次损失[38]或词嵌入先验[49]来规范化网络的输出。但是，它们的损失约束只能在最终预测层上起作用，并且间接地引导视觉特征具有层次意识，这很难得到保证。最近，Marino 等人[32]使用结构先验知识来增强对多标签分类的预测，而我们的 SGR 提出了一种通用神经层，可以将其注入到任何卷积层中，并允许神经网络利用从各

种人类知识中得出的语义约束。Chen 等[7]利用基于局部区域的推理和全局推理来促进对象检测。相比之下，我们的 SGR 层直接在符号节点上执行推理，并与局部卷积层进行无缝交互以提高灵活性。值得注意的是，人工智能推理的最早努力可以追溯到符号方法[35]，即通过使用数学和逻辑语言对抽象符号进行推理。在将这些符号接地之后，使用统计学习算法[23]提取有用的模式，以在知识库上执行关系推理。对于高级任务来说足够实用的有效推理程序，应该加入局部视觉表示学习和全局语义图推理的力量。我们的推理层与该研究领域有关，它是通过从本地表示进行投票的方式来对语言实体的视觉证据进行显式推理的。

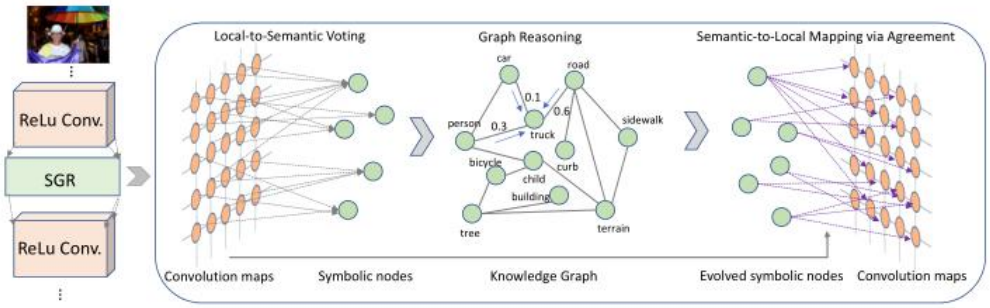

图 1：建议的 SGR 层的概览。每个符号节点都通过本地到语义的投票模块（长灰色箭头）从所有本地特征中接收投票，然后通过语义到本地的映射模块（长紫色）将其在图形推理后的演变特征映射回每个位置。为简单起见，我们在知识图中省略了更多的边和符号节点。

3 符号图推理

3.1 通用图构造

常识图通常用于描述实体之间（例如类，属性和关系）的不同关联，关联可以是任何形式。为了支持通用图推理，知识图可以表示为 G =（N，E），其中 N 和 E 分别表示符号集和边集。这里我们给出三个示例：a）类层次图是由一系列实体类（例如人，摩托车手）构成的，其图边缘承担着概念所有物的责任（例如“是某种”或“是其一部分”）。通过将父类的共享表示传递到子节点中，具有这种层次知识的网络可以鼓励学习特征层次。 b）类共现图将边缘定义为图像中两个类的共现，表征了预测的合理性； c）作为高级语义抽象，语义关系图可以扩展符号节点以包括更多动作（例如“骑”，“玩”），布局（例如“在顶部”）和属性（例如颜色或形状）虽然从语言描述中统计地收集了图的边缘，但结合这些高级常识知识可以帮助网络在了解每个实体对的关系之后修剪虚假的解释，从而实现良好的语义一致性。

基于该通用公式，图推理需要兼容且足够通用，以适用于软图边缘（例如出现概率）和硬图边缘（例如所有物）以及各种符号节点。因此，可以将各种结构约束建模为符号节点上的边缘连接，就像人类使用的语言工具一样。我们的 SGR 层旨在实现适用于编码各种知识图形式的通用图推理。如图 1 所示，它由一个本地到语义的投票模块，一个图形推理模块和一个语义到本地的映射模块组成，如以下各节所述。

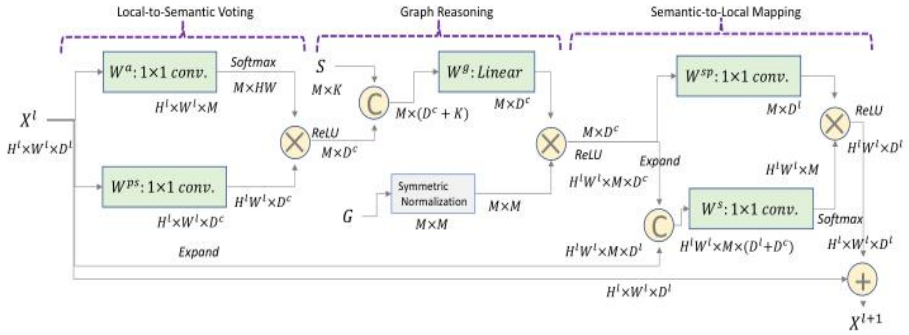

图 2：采用 $H^{l} \times W^{l} \times D^{l}$ 维卷积特征张量作为输入的一个 SGR 层的实现细节。 ⊗ 表示矩阵相乘， ⊕ 表示按元素求和，带 C 的圆表示串联。softmax 操作，张量扩展， ReLU 操作当被标注时会被执行。绿色框表示 1×1 卷积或线性层。

3.2 本地到语义投票模块

给定来自卷积层的局部特征张量，我们的目标是利用全局图推理利用外部结构化知识来增强局部特征。因此，我们首先将以局部特征编码的全局信息汇总为符号节点的表示形式，即，将与特定语义（例如 cat）相关的局部特征汇总在一起，以描述其相应符号节点的特征。形式上，我们使用在 l 层卷积层后的特征张量 $X^{l} \in R^{H^{l} \times W^{l} \times D^{l}}$ 来作为模块的输入，其中 H^{l} 和 W^{l} 分别是特征图的高度和权重， D^{l} 是通道数。该模块旨在使用 X^{l} 来产生所有 $M = |N|$ 的符号节点的视觉表示 $H^{ps} \in R^{M \times D^{C}}$ ，其中 D^{c} 是每个结点的期望特征维度，这可以被公式化为函数 ϕ ：

$$H^{ps} = \phi（A^{ps}，X^{l}，W^{ps}） \qquad (1)$$

其中 $W^{ps} \in R^{D^{l} \times D^{c}}$ 是可训练的转换矩阵，即将每个局部特征 $x_i \in X^{l}$ 转化为维度 D^{C} ，而且 $A^{ps} \in R^{H^{l} \times W^{l} \times M}$ 代表每个符号节点其所有局部特征的投票权重。具体来说，每个节点 n 的视觉特征 $H^{ps}{}_{n} \in H^{ps}$ 是利用代表局部特征 x_i 赋值给节点 n 可能性大小的投票权重 $a_{x_i \to n} \in A^{ps}$ 来对所有经过权重转换的局部特征进行求和，从而计算出来的。更具体而言，函数 ϕ 被计算为：

$$H^{ps}{}_{n} = \sum_{x_i} a_{x_i \to n} x_i W^{ps}, \quad a_{x_i \to n} = \frac{\exp(W_n^{aT} x_i)}{\sum_{n \in N} \exp(W_n^{aT} x_i)} \qquad (2)$$

其中 $W^{a} = \{W^{a}{}_{n} \in R^{D^{l} \times M}\}$ 是一个用于计算投票权重的可训练的权重矩阵。 A^{ps} 通过在

每个位置处使用 softmax 来进行正则化。通过这种方式，不同的局部特征可以自适应地投票给不同的符号节点表示。

3.3 图推理模块

基于符号节点的视觉证据，以结构化知识为指导的推理被用来利用人类常识的语义约束来发展符号节点的全局表示。在这里，我们结合了每个符号节点的语言嵌入和知识连接（即节点边缘）来执行图推理。形式上，对于每个符号节点 n $\in N$，我们使用现成的词向量[17]作为其语言嵌入，表示为 $S=\{s_n\}, s_n \in R^K$。图推理模块通过矩阵乘法形式在所有符号节点的表示 H^{ps} 上执行图传播，从而得出进化特征 H^g：

$$H^g = \partial(A^g B W^g) \qquad (3)$$

$B=[\sigma(H^{ps}), \mathcal{S}] \in \mathbf{R}^{M\times(D^c+K)}$ 通过激活函数 σ （ • ）和语言嵌入 S 连接变换后的 H^{ps} 特征。$W^g \in \mathbf{R}^{(D^c+K)\times(D^c)}$ 是可训练的权重矩阵。节点邻接权重 $a_{n\to n'} \in A^g$ 是根据 $(n,n') \in \mathcal{E}$ 中的边连接来定义的。如第 3.1 节所述，根据不同的知识图资源，边缘连接可以是软权重（例如 0.8）或硬权重（即{0,1}）。使用 $\bar{A}^g$ 的简单乘法会完全改变特征向量的尺度。受图卷积网络的启发[18]，我们可以对 $\bar{A}^g$ 进行归一化，使所有行加总为 1 来解决这个问题，即 $Q^{-\frac{1}{2}} A^g Q^{-\frac{1}{2}}$，其中 Q 是 $\bar{A}^g$ 的对角节点度矩阵。该对称归一化对应于取相邻节点特征的平均值。该公式得出新的传播规则：

$$H^g = \sigma(\hat{Q}^{-\frac{1}{2}} \hat{A}^g \hat{Q}^{-\frac{1}{2}} B W^g), \qquad (4)$$

$\hat{A}^g = A^g + I$ 是图 $\mathcal{G}$ 的邻接矩阵，其附带了考虑节点自身表示的额外连接信息，I 是单位矩阵。$\hat{Q}_{ii} = \sum_j \hat{A}^g_{ij}$

3.4 语义到本地的映射模块

最后，演化出的所有符号节点的全局表示 $H^g \in \mathbf{R}^{M\times D^c}$ 可被用于增强每个局部特征表示的能力。由于图推理后每个符号节点的特征分布已更改，一个关键问题是如何找到从每个符号节点的表示 $h^g \in H^g$ 到所有 x_i 的一个最合适的映射。这对于学习局部特征和符号节点之间的兼容性矩阵可能是不可知的。受消息传递算法的启发[11]，我们通过评估每个符号节点 h^g 与每个局部特征 x_i 的兼容性来计算映射权重 $a_{h^g\to x_i} \in A^{sp}$：

$$a_{h^g \to x_i} = \frac{\exp(W^{sT}[h^g, x_i])}{\sum_{x_i} \exp(W^{sT}[h^g, x_i])}, \tag{5}$$

$W^s \in \mathbf{R}^{D^l + D^c}$是可训练的权重矩阵。兼容性矩阵$A^{sp} \in \mathbf{R}^{H \times W \times M}$则再次被行正则化。通过图推理、在 l+1 层卷积中作为输入的演化特征X^{l+1}可以按照如下方式更新：

$$X^{l+1} = \sigma(A^{sp} H^g W^{sp}) + X^l, \tag{6}$$

$W^{sp} \in \mathbf{R}^{D^c \times D^l}$是用于将符号节点表示的维转换回$D^l$的可训练矩阵，我们使用残差连接[12]进一步增强具有原始局部特征张量X^l的局部表示。每个局部特征通过来自每个符号节点的加权映射来更新，这些加权映射表示语义的不同特征。

3.5 符号图推理层

每个符号图推理层由本地到语义投票模块，图推理模块和语义到本地映射模块的堆栈构成。 SGR 层由具有不同数量的符号节点和不同节点连接的特定知识图实例化。将具有不同知识图的多个 SGR 层组合到卷积网络中可以导致混合图推理行为。我们通过 1×1 卷积运算和非线性函数的组合来实现每个 SGR 的模块，详细信息如图 2 所示。我们的 SGR 灵活且通用，足以将其注入到任何局部卷积之间。但是，由于 SGR 被指定为包含高级语义推理，因此，如我们的实验所示，在以后的卷积层中使用 SGR 更可取。

4 实验

由于我们将建议的 SGR 层作为适用于任何卷积网络的常规模块进行展示，我们因此将它与 CocoStuff [4]，Pascal-Context [34]和 ADE20K [52]上的像素级预测任务（即语义分割）和 CIFAR-100 上的图像分类任务进行了比较[21]。对 Coco-Stuff 数据集进行了广泛的消融研究[4]。

4.1 语义分割

数据集。 我们评估了三个针对大型类别进行细分的公开基准，与其他小型细分数据集（例如 PASCAL-VOC）相比，这构成了更为现实的挑战，并且可以更好地验证全局符号推理的必要性。具体来说，Coco-Stuff [4]包含 10,000 张图片，其中包含 91 种事物（例如书籍，时钟）和 91 种东西类别（像花、木头）的密集注释，其中 9,000 用于训练，1000 用于测试。ADE20k [52]由 20,210 张用于训练的图像和 2,000 张用于验证的图像组成，并带有 150 个语义概念（例如绘画，灯）。 PASCAL-Context [34]包括用于训练的 4,998 张图像和用于测试的 5105 张图像，并带有 59 个对象类别和一个背景。我们使用像素精度（pixAcc）和平均交并比（mIoU）的标准评估指标。

实现。我们在单个服务器上使用 2 个 GTX TITAN X 12GB 卡以及 Pytorch 来进行所有实验。我们按照[6]的程序，使用 Imagenet 预训练的 ResNet-101 [12]作为基本的 ConvNet，采用输出步幅= 8 并将 SGR 层合并到其中。一个 SGR 层的详细实现在图 2 中。我们的最终模型模块使

用{6,12,18,24}的金字塔来创建空间金字塔池（ASSP）[6]模块，以将 ResNet-101 的最终 ResBlock 中的 2,048-d 特征减少为 256-d 特征。在此之后，我们堆叠一个 SGR 层以增强本地特征，然后是最终的 1×1 卷积产生最终的像素预测。因此，本地到语义投票模块和图推理模块中的特征维度的 D1 和 Dc 被设置为 256，并且我们对σ（·）使用 ReLU 激活函数。来自 fastText [17]的字嵌入用于表示每个类，其提取子字信息并很好地概括为词汇外单词，从而为每个节点产生 K = 100-d 向量。

我们对所有数据集使用通用概念层次结构。在[27]之后，从包含 182 个概念和 27 个超类的 COCO-Stuff [4]的标签层次结构开始，我们使用 WordTree 手动将其余两个数据集中的概念合并为[27]。它在最终概念图中产生了 340 个概念。因此，该概念图使得符号图推理层在所有三个数据集中可以是相同的，并且其权重可以容易地与彼此的数据集共享。我们在精确调整期间确定了 ResNet-101 批量标准化的移动方式和变化。我们采用标准的 SGD 优化。受[6]的启发，我们使用“poly”学习率策略，为新初始化的层设置基本学习率为 2.5e-3，为预训练层设置 2.5e-4。我们为 Coco-Stuff 和 PASCAL-Context 训练 64 个时期 - ADE20K 数据集则是 120 个时期。对于数据增加，我们采用随机浮点，随机裁剪和 0.5 到 2 之间的随机调整所有数据集。由于 GPU 内存限制，批量大小用作 6.输入裁剪大小设置为 513×513。

4.1.1 与最先进的技术进行比较

表 1,2,3 分别报告了与 Coco-Stuff，PascalContext 和 ADE20K 数据集上最近的最先进方法的比较。结合我们的 SGR 层显著优于所有三个数据集上的现有方法，证明了在大规模像素级识别之外执行超出局部卷积的显式图推理的有效性。图 3 显示了与基线“Deeplabv2 [6]”的定性比较。我们的 SGR 获得了更好的分割性能，特别是对于一些罕见的类别（例如伞，泰迪熊），从关于概念层次图的频繁概念的联合推理中获益。特别是，将用于分类任务的高级语义约束结合到像素识别中的技术并不是微不足道的，因为将先验知识与密集像素本身相关联是很困难的。先前的工作[38,10,49]也试图隐含地促进具有分级分类目标的网络学习。最近的 DSSPN [27]直接为每个父概念设计了一个网络层，然而，这种方法难以扩展到大规模的概念集，导致对不太可能属于特定概念的像素产生冗余预测。与先前的方法不同，通过仅添加一个推理层，所提出的 SGR 层可以实现更好的结果，同时保留良好的计算和内存效率。

Method	mean IoU	pixel acc.
FCN [31]	29.39	71.32
SegNet [2]	21.64	71.00
DilatedNet [47]	32.31	73.55
CascadeNet [52]	34.90	74.52
ResNet-101, 2 conv [45]	39.40	79.07
PSPNet (ResNet-101)DA_AL [50]	41.96	80.64
Conditional Softmax [38]	31.27	72.23
Word2Vec [10]	29.18	71.31
Joint-Cosine [49]	31.52	73.15
DeepLabv2 (ResNet-101) [6]	38.97	79.01
DSSPN (ResNet-101) [27]	42.03	81.21
Our SGR (ResNet-101)	**44.32**	**81.43**

Table 3: Comparison on the ADE20K val set [52] (%). “Conditional Softmax [38]", “Word2Vec [10]" and “Joint-Cosine [49]" use VGG as backbone. We use “DeepLabv2 (ResNet-101) [6]" as baseline.

Table 4: Curves of the training losses on Coco-Stuff for the Deeplabv2 (Baseline) [6] and our three variants. Following [6], the loss is the summations of losses for inputs of three scales (i.e. 1, 0.75, 0.5).

4.1.2 消融研究

哪个 ConvBlock 添加 SGR 层？表 1 和表 4 比较了将单个 SGR 层添加到 ResNet-101 的不

同阶段的变体。 “SGR ConvBlock4”表示 SGR 层被添加到 res4 的最后一个残余块之前，而所有其他变体是在 res5 的最后一个残余块（最终残余块）之前添加 SGR 层。“SGR ConvBlock4”的性能比“我们的 SGR（ResNet-101）”要差，而对 res4 和 res5（“我们的 SGR（ResNet-101 2 层）”）都使用 SGR 层可以略微改善结果。请注意，为了使用 ResNet-101 中的预训练权重，“我们的 SGR（ResNet-101 2 层）”通过求和直接将来自 res4 和 res5 之后的两个 SGR 层的预测结果融合在一起，以获得最终的预测。这一观察的一个可能的解释是最终 res5 可以编码更多语义抽象的特征，这更适合于进行符号图推理。此外，通过比较“SGR（无余残差）”与我们的完整 SGR，我们发现去除方程 6 中的残差连接会降低最终性能，但仍然优于其他基线。原因在于 SGR 层通过全局推理引起更加平滑的局部特征，因此可能降低边界中的一些判别能力。

语义到本地映射的效果。 请注意，我们的 SGR 分别在本地到语义模块和语义到本地模块中学习不同的投票权重和映射权重。 通过在表 1 和表 4 中的“测试性能和训练收敛”中比较“我们的 SGR（ResNet-101）”和“SGR（w / o 映射）”，可以看出重新评估映射权重的优势。这种对语义到局部映射权重的新一轮评估可以更好地适应图形推理后的演化特征分布，否则演化的符号节点将与局部特征不对齐。

不同的先验知识图。如 3.1 节所述，我们的 SGR 层适用于任何形式的具有软边或硬边权重的知识图。因此，我们评估了利用表 1 中的不同知识图的结果。首先，类并发图通常用于表示出现在一个图像中的任何两个概念的频率，其描绘了统计视图中的类间合理性。我们计算了 Coco-Stuff 上所有训练图像的类并发图，并将其作为 SGR 层的输入，作为“SGR（并发图）”。我们可以看到，并入一个并发驱动的 SGR 层也可以提高分割性能但是它略逊于带有概念层次结构的。其次，我们还依次将一个 SGR 层与层次结构图堆叠在一起，并将一个层与并发图堆叠在一起，从而形成混合版本“Our SGR（ResNet-101 Hybrid）”。该变体在所有模型中实现了最佳性能，验证了利用知识约束的混合来提高语义推理能力的好处。最后，我们进一步探索了一个丰富的场景图，其中包含用于编码高级语义的概念，属性和关系，如“SGR（场景图）”变体。在[24]之后，场景图是从 Visual Genome[20]构建的。为简单起见，我们只选择对象类别，属性和谓词，这些对象类别，属性和谓词至少出现 30 次并且与我们在 Coco-Stuff 中的 182 个目标概念相关联。它导致一个带有 312 个对象节点、160 个属性节点和 68 个谓词节点的无向图。“SGR（场景图）”比“我们的 SGR（ResNet-101）”略差，但优于“SGR（并发图）”。从所有这些研究中观察，我们因此通过平衡效率和有效性，将概念层次图用于所有其余实验。

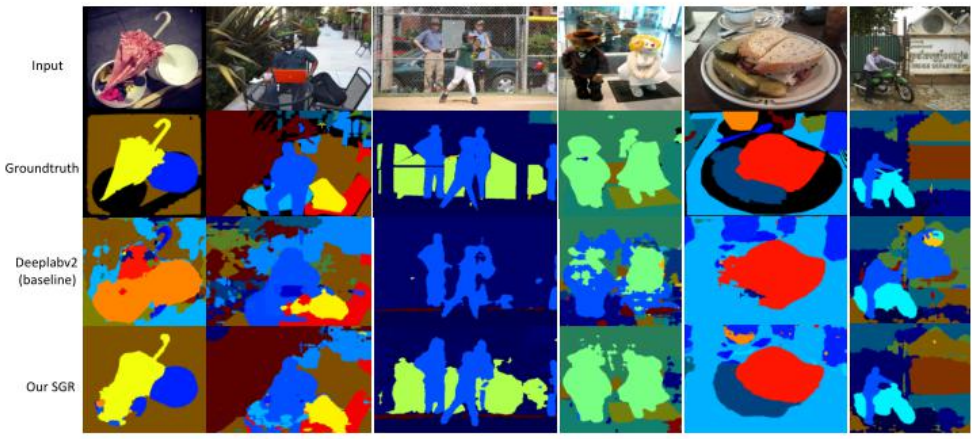

Figure 3: Qualitative comparison results on Coco-stuff dataset.

Method	ResNet [13]	Wide [48]	ResNeXt-29 [46]	DenseNet [16]	DenseNet-100 [16] (baseline)	SGR	SGR 2-layer
Depth	1001	28	29	190	100	100+1*	100+2*
Params	16.1M	36.5M	68.1M	25.6M	7.0M	7.5M	8.1M
Error	22.71	20.50	17.31	17.18	22.19	17.68	17.29

Table 5: Comparison of model depth, number of parameters (M), test errors (%) on CIFAR-100. "SGR" and "SGR 2-layer" indicate the results of appending one or two SGR layer on the final denseblock of the baseline network (DenseNet-100), respectively.

将从一个域学习到的 SGR 转移到其他域。 在从局部特征投票之后，我们的 SGR 层自然地学习编码一般符号节点的显式语义含义，只有当这些域共享一个先行图时，其权重才能容易地从一个域转移到其他域。由于 Coco-Stuff 和 PASCALContext 数据集都使用单个层次图，如图 2 所示我们可以使用在 Coco-Stuff 上预训练的 SGR 模型来初始化 PASCAL-Context 数据集上的训练。"我们的 SGR（传输 convs）"仅表示使用预训练的残差块权重而"我们的 SGR（转移 SGR）"是进一步使用 SGR 层参数的变体。 我们可以看到，传递 SGR 层的参数可以比单独传输卷积块提供更多的改进。

4.2 图像分类结果

我们进一步对 CIFAR-100 [21]上的图像分类任务进行研究，其中包括 100K 类的 50K 训练图像和 10K 测试图像。我们首先探索了 SGR 对于基线网络，即 DenseNet-100 [16]性能提升的程度。我们在最终密集块上附加上 SGR 层，这将生成 342 个 8×8 尺寸的特征图。我们首先使用 1×1 卷积层将 342-d 特征减少到 128-d，然后依次使用一个 SGR 层，全局平均池和线性层来产生最终分类。具有 148 个符号节点的概念层次图是通过将 100 个类映射到 WordTree 生成的，类似于分段实验中使用的策略，包括在补充材料中。我们将D^l和D^c设置为 128. 在训练期间，我们在两个 GPU 上使用 64 的小批量大小，，并使用余弦学习速率调度[16]表示 600 个时期。表 5 中的更多比较表明，我们的 SGR 可以改善基线网络的性能，通过全局推理从增强功能中获益。它获得了可与最先进的方法相比的结果，而模型复杂度却大大降低。

5 结论

为了使局部卷积网络具有全局图推理的能力，我们引入了一个符号图推理（SGR）层，该层利用外部人类知识来增强局部特征表示。提出的 SGR 层是通用的，轻量级的，并且与现有的卷积网络兼容，卷积网络由本地到语义的投票模块，图推理模块和语义到本地的映射模块组成。在关于语义分割的三个公开基准和一个图像分类数据集上的大量实验证明了其优越的性能。我们希望我们的 SGR 设计能够帮助推动研究卷积网络的全局推理特性的研究，并有益于社区中的各种应用。