

数字图像处理课程设计

赵轩磊¹ 严茹丹¹ 苏秦¹ 董思琪¹

摘要 特征上采样是许多现代卷积网络架构中的关键操作，例如特征金字塔。它的设计对于密集预测任务至关重要，例如对象检测和语义/实例分割。在这项工作中，我们提出了内容感知特征重新组装（CARAFE），这是一种通用、轻量级且高效的运算符来实现这一目标。CARAFE 有几个吸引人的特性：(1) 大视野。与之前仅利用子像素邻域的工作（例如双线性插值）不同，CARAFE 可以在大的感受野内聚合上下文信息。(2) 内容感知处理。CARAFE 不是对所有样本使用固定内核（例如反卷积），而是启用特定于实例的内容感知处理，从而即时生成自适应内核。(3) 轻量级且计算速度快。CARAFE 引入的计算开销很小，可以很容易地集成到现代网络架构中。我们对目标检测实例/语义分割和绘画的标准基准进行综合评估。

关键词 关键词 1, 关键词 2, 关键词 3

DIP Course Project

Xuanlei Zhao¹ Rudan Yan¹ Qin Su¹ Siqi Dong¹

Abstract An abstract should be a concise summary of the significant items in the paper, including the results and conclusions. It should be about 5% of the length of the article, but not more than about 200 words. Define all nonstandard symbols, abbreviations and acronyms used in the abstract. Do not cite references in the abstract.

Key words Keyword 1, keyword 2, keyword 3

特征上采样是深度神经网络中最基本的操作之一。一方面，对于密集预测任务中的解码器（例如超分辨率 [7, 20]，图像修复 [13, 32] 和语义分割 [43, 5]），特征图通常会被上采样以匹配高分辨率监督。在另一方面，特征上采样还涉及将高、低分辨率的特征图相互融合，这在许多最先进的架构中被广泛采用，例如，特征金字塔网络[21]、U-Net [34]、堆积沙漏[29]和 HRNet。因此，如何设计有效的特征上采样算子成为一个关键问题。

最广泛使用的特征上采样算子是最近邻和双线性插值，它们采用像素之间的空间距离来指导上采样过程。然而，最近邻和双线性插值只考虑子像素邻域，未能捕捉到密集预测任务所需的丰富语义信息。自适应上采样的另一条途径是反卷积[30]。反卷积层作为卷积层的逆算子，它学习一组与实例无关的上采样内核。但是，它有两个主要缺点。

首先，反卷积算子在图像所有位置上应用相同的卷积核，忽视了图像中重要的内容，这限制了它处理局部变化的能力。其次，使用大尺寸卷积核时，参数数量与计算量会大量增加。这使得其难以覆盖小邻域外的更大区域，从而限制了其表达能力和性能。

最近，一些工作提出了生成式的上采样算法，例如 xx。虽然这样的算法在性能上取得了一定的提升，但是也受到了一些缺点的限制。第一，在生成图片时，自生成算法将原图生成的特征重新排列为空间块来完成特征上采样，而每个空间块内忽略了局部像素的位置关系，这会造成上采样图片的空间关系不匹配。第二，此类算法的计算量与感受野成平方关系，这导致感受野依旧被限制在一小块领域范围内，对性能提升造成了阻碍，在处理高分辨率图片时影响尤为明显。第三，在进行高比例上采样时，算法中需要生成的参数也会激增，导致对计算能力的要求攀升，影响了实际的应用。

在这项工作中，为了克服这些限制，需要寻找一种全新的特征上采样算子，能够满

1. 华中科技大学电子信息与通信学院 湖北 430074

1. School of Electronic Information and Communications,
Huazhong University of Science and Technology Hubei 430074

足以下的要求：1) 兼顾全局与局部信息，2) 根据当前的数据自动生成，3) 建立通道信息相关性，4) 保持计算效率。为此，我们提出了一种轻量但高效的全新方法，称为 **Rezolution**。不同与上述的生成式算法，该方法考虑的是如何对双线性插值上采样的图片进行后处理精调，从而修正需要关注的重点。

在我们看来，我们进行的后处理本质上是对上采样后的图片进行特征锐化操作。对于传统的双线性插值方法，它被广泛使用的原因在于能够高效完成上采样计算，并且保留图像的空间关系。但是就像上述提到的，双线性插值只考虑子像素邻域，未能捕捉到密集预测任务所需的丰富语义信息，对画面中所有信息进行等比例的放大，失去了关注的重点。因此，要解决的问题就是如何对双线性插值放大后的图片进行图像特征锐化，突出对于画面真正重要的内容，是该方法关注的重点。

具体来说，**Rezolution** 对原图与上采样后的新图片，分别提取沿着两个不同方向聚合的特征，再通过原图特征修正新图特征，最后将新特征映射回新图，完成上采样图片修正的过程，得到最终的上采样图片。

为了证明 **Rezolution** 的普遍有效性，我们对语义分割任务进行了综合评估。**Rezolution** 改进了 **HRNet** 的特征图融合部分，在语义分割中的 **VOC12 val** 上提高了 1.8% **mIoU**。当对具有 256 个通道的特征图进行 2 倍上采样时，**Rezolution** 每个像素引入的计算开销仅为 199k FLOPs，是 **xx** 的 **xx**。所有任务的实质性收益表明，**Rezolution** 是一种有效且高效的特征上采样算子，具有作为未来研究的强大基石的巨大潜力。

1 相关算法概述

1.1 上采样运算符

最常用的上采样方法是最近邻和双线性插值。这些插值利用距离来测量像素之间的相关性，并在其中使用手工制作的上采样内核。在深度学习时代，提出了几种使用可

学习算子对特征图进行上采样的方法。例如，反卷积 [30] 是卷积的逆算子，是那些可学习的上采样器中最著名的。**Pixel Shuffle** [35] 提出了一种不同的上采样器，它将通道空间上的深度重塑为空间空间上的宽度和高度。最近，[26] 提出了引导上采样 (**GUM**)，它通过对具有可学习偏移量的像素进行采样来执行插值。然而，这些方法要么利用小邻域中的上下文信息，要么需要昂贵的计算来执行自适应插值。在超分辨率和去噪领域，其他一些作品 [27, 16, 11] 也探索在低级视觉中空间使用可学习内核。本着类似的设计精神，我们在这里展示了内容感知特征重组在多个视觉感知任务中进行上采样的有效性和工作机制，并提供了一个轻量级的解决方案。

1.2 密集预测任务

对象检测是用边界框定位对象的任务，实例分割进一步需要对实例掩码的预测。**Faster-RCNN** [33] 引入了用于端到端训练的区域提议网络 (**RPN**)，通过引导锚定方案 [37] 进一步改进了这一点。[21, 24, 17, 45, 31] 利用多尺度特征金字塔来处理不同尺度的对象。通过添加额外的掩码预测分支，**Mask-RCNN** [9] 及其变体 [1, 12] 产生了有希望的像素级结果。语义分割 [25, 19] 需要对给定图像进行像素级语义预测。**PSP Net** [43] 在多个网格尺度上引入了空间池化。**UperNet** [38] 设计了一个基于 **PSPNet** 的更通用的框架。图像或视频修复 [42, 40, 39] 是填补输入图片缺失区域的经典问题。**Unet** [34] 在最近的作品 [13, 36] 中很受欢迎，并且采用了多个上采样算子。刘等人。[23] 引入部分卷积层来减轻缺失区域对卷积层的影响。我们的 **CARAFE** 在广泛的密集预测任务中展示了普遍的有效性。

2 Rezolution

在许多现代卷积网络架构中，特征上采样是一个关键的算子，这些架构是为目标检测、实例分割和场景解析等任务而开发的。在这项工作中，我们提出了 **Rezolution** 来

对特征图进行上采样, 并获得比主流上采样算子更好的性能, 例如插值或生成式算法。

2.1 公式

Resolution 由三个步骤组成。第一步是对原图与上采样后的图片分别提取特征向量, 第二步是用原图的特征向量更新上采样后的特征向量, 再将新的特征向量映射回图片。给定大小为 $C \times H \times W$ 的特征图 X 和上采样率 σ (假设 σ 是整数), 首先会通过双线性插值生成大小为 $C \times \sigma H \times \sigma W$ 的特征图 X' , 之后 Resolution 将调整特征图 X' , 最终生成大小为 $C \times \sigma H \times \sigma W$ 的新特征图 Y 。

在第一步中, 特征提取模块 ξ 将通过池化分别对上采样前与上采样后特征图的特征图提取关键特征, 如公式 (1) 所示。在第二步中, 利用 ϕ 模块将原图的特征向量拓展, 并使用 μ 更新上采样图的特征向量, 最后将特征向量映射回原图, 如公式 (2) (3) 所示。我们将在本节后续部分中说明 ξ 、 ϕ 、 μ 的详细信息。

$$W_f, O_f' = \xi(\text{Pool}(X), \text{Pool}(X')), \quad (1)$$

$$W = \phi(G(W_f), O_f'), \quad (2)$$

$$Y = \mu(X', W). \quad (3)$$

2.2 特征提取

在特征提取部分, 我们参考了 CANet 的方法。全局池化通常用于通道注意力以全局编码空间信息, 但它将全局空间信息压缩到通道描述符中, 因此难以保留位置信息, 这对于在视觉任务中捕获空间结构至关重要。为了鼓励注意力块以精确的位置信息在空间上捕获远程交互, 我们将全局池化分解为为一对一维特征编码操作, 如公式 (4) (5)。具体来说, 给定输入 X , 我们使用池化内核的两个空间范围 $(H, 1)$ 或 $(1, W)$, 分别沿水平坐标和垂直坐标对每个通道进行编码。因此, 高度为 h 的第 c 个通道的输出可以表示为:

$$z_c^h(h) = \frac{1}{w} \sum_{i=0}^w x_c(h, i). \quad (4)$$

类似地, 宽度为 w 的第 c 个通道的输出可以写为:

$$z_c^w(w) = \frac{1}{H} \sum_{i=0}^H x_c(i, w). \quad (5)$$

上述两种变换分别沿两个空间方向聚合特征, 产生一对方向感知特征图。这两种转换还允许我们的注意力块沿一个空间方向捕获远程依赖关系, 并沿另一个空间方向保留精确的位置信息, 这有助于网络更准确地定位感兴趣的对象。

2.3 特征更新

如上所述, 方程 (4) 和方程 (5) 启用全局感受野并编码精确的位置信息。为了更新上采样后的特征向量, 我们必须先对原图的特征向量进行扩展, 使得原图特征向量的长度与上采样后的特征长度保持一致。对于, 特征向量的扩张, 我们使用了经典方法 PixelShuffle, 即先用卷积增加通道深度, 再利用通道信息将每个位置的向量空间上扩大相应的倍数, 如公式 (6) 所示, 其中 N 代表卷积函数。

$$W'_f = \text{PixelShuffle}(N(W_f)). \quad (6)$$

获得了扩展后的原图特征后, 我们希望使用原图特征去指导上采样后的特征, 通过特征向量中捕获的空间和通道信息, 将被双线性插值稀释的重点信息重新标记出来。为了完成这个目标, 我们提出了一个简单但是有效的方法来更新上采样特征——门控机制。具体来说, 对于扩展后的原图特征 W'_f 与上采样后的特征 O_f , 我们通过计算 W'_f 与 O_f 的内积, 获取这两个向量在高维空间中的几何关系, 将内积后的结果作为门控, 如公式 (7) 所示, 再利用门控机制对上采样后的特征进行更新, 如公式 (8) 所示。

$$G = \text{Sigmoid}(F(W'_f * O_f)), \quad (7)$$

$$W = G * W'_f + (1 - G)O_f. \quad (8)$$

3 实现

语义分割要求模型在整个图像上输出每像素级别的预测, 因此通常使用高分辨率的特征图输出结果。上采样被广泛用于放大特征图, 并在自上而下的路径中以一定比例将特征图融合在一起。我们选择了 3 个目标检测中具有代表性的模型: PSPNet, UperNet, OCRNet, 实现了我们的上采样方法。

表 1 不同上采样方法在语义分割数据集 VOC12-val 上的结果

| 算 法 | 骨干网络 | 上采样方法 | mIoU | mAcc | aAcc |
|---------|----------|------------|-------------|-------------|-------------|
| PSPNet | ResNet50 | Bilinear | 75.4 | 86.1 | 94.4 |
| | | CARAFE | 75.6 | 86.3 | 94.3 |
| | | Rezolution | 76.1 | 87.1 | 94.4 |
| UperNet | ResNet50 | Bilinear | 74.0 | 84.6 | 93.9 |
| | | CARAFE | 74.3 | 84.9 | 94.1 |
| | | Rezolution | 74.7 | 85.4 | 94.2 |
| OCRNet | HRNet18 | Bilinear | 73.4 | 85.0 | 93.8 |
| | | CARAFE | 73.9 | 85.4 | 94.0 |
| | | Rezolution | 74.8 | 85.5 | 94.1 |

3.1 PSPNet

PSPNet 中 PPM 将输入特征图分层下采样到多个尺度 $\{1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6\}$ ，然后使用双线性将它们上采样回原始大小。这些特征最终通过串联与原始特征融合。由于上采样率非常大，我们采用了两步策略作为性能和效率之间的权衡。首先我们使用双线性插值将 $\{1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6\}$ 特征上采样到原始特征图大小的一半，然后使用 Rezolution 将它们进一步上采样 2 倍。

3.2 UperNet

在 UperNet 中，我们修改了 PPM 和 FPN 两个模块中的上采样。PPM 的修改方法与 3.1 节中提到的相同。对于 FPN，UperNet 采用它来丰富语义特征。它只有四个不同的特征级别 $\{P2, P3, P4, P5\}$ ，步幅为 $\{4, 8, 16, 32\}$ 。我们依次从高级特征到低级特征上采样两倍。

3.3 OCRNet

OCRNet 使用 HRNet 作为 backbone，在 backbone 的最后分别将 4 个不同尺度的特征图分别放大 $\{1, 2, 4, 8\}$ 倍，最后将他们拼接在一起获得最终的特征图，在此处我们直接用 Rezolution 上采样对应的倍数。

3.4 CARAFE

我们将目前 SOTA 的方法 CARAFE 作为对比之一，直接使用 mmdetection 中提供的代码。CARAFE 使用生成式的方法，对图片中每一个位置按照上采样比例生成代码块，最终将代码块合成新的图片。在上采样时，我们直接用 CARAFE 替代双线性插值。

4 实验

在本节中，我们首先描述我们的实验设置，然后进行一系列对比实验证明我们算法的有效性，我们报告了与其他上采样方法在语义分割上的结果的对比。之后，我们对不同上采样算法的参数、计算量与 mIoU 进行了对比。接着我们对比了不同算法对于不同比例缩放的性能，证明我们的算法在进行大比例上采样时明显优于以往方法。最后我们进行了一系列消融实验以证明我们的门控方法对性能的贡献。

4.1 实验设置

我们使用基于 PyTorch 的 mmsegmentation 工具来实现我们所有的实验。在训练期间，我们使用衰减和动量为 0.9 的标准 SGD 优化器来训练所有模型。权重衰减始终设置为 5×10^{-4} 。我们使用两 NVIDIA 2080Ti GPU 进行训练，批量大小设置为 8。在没有额外声明的情况下，我们训练所有模型 200 个 epoch。对于数据增强，我们使用与 mmsegmentation 相同的方法。我们报告了 VOC12-val 中语义分割的结果。

4.2 基线结果

在基线结果对比中，我们挑选了 3 个语义分割的主流算法：PSPNet，UperNet 与 OCRNet，前两者的骨干网络为 ResNet50，OCRNet 的骨干网络为 HRNet。对于每种算法，我们分别测试了双线性插值，CARAFE 和 Rezolution 的效果，结果如表 1 所示。可以从表中看出，对于三种算法，Rezolution 方

法都取得了明显的提升。

4.3 性能分析

在性能分析中，我们对比了在不同放大倍率下，不同采样方法的区别。

首先，在大比例上采样的情况下，我们对比了 CARAFE 与 Rezolution 的性能提升。如表 2 所示，我们能够发现在上采样倍率为 2 时，Rezolution 相比于 CARAFE 能够提高 0.4% 的 mIoU。而当上采样倍率变大时，Rezolution 的提升变得更为明显，能够达到 0.9%。证明我们的方法无论是何种放大比例，都能够取得更优的效果。造成的稿件如果第一作者非文章通信作者或通信作者地址变动，务请及时通知编辑部。

表 2 不同上采样倍率下的提升

| 算 法 | 倍率 | 采样方法 | mIoU |
|---------|------|------------|-------------|
| PSPNet | 2x | Bilinear | 75.4 |
| | | CARAFE | 75.6 |
| | | Rezolution | 76.1 |
| UperNet | 2x | Bilinear | 74.0 |
| | | CARAFE | 74.3 |
| | | Rezolution | 74.7 |
| OCRNet | 2-8x | Bilinear | 73.4 |
| | | CARAFE | 73.9 |
| | | Rezolution | 74.8 |

最终指标以外，我们也对三种不同上采样方法的计算性能进行了评估，主要考虑计算量与参数量。我们使用了 OCRNet 作为模型，HRNet 作为骨干模型，评估三种上采样算法的计算量与参数量，如表 3 所示。可以发现，相比于轻量型算法 CARAFE，在取得相当的性能提升以外，我们将计算量缩减了 96%，参数量缩减了 80%。

表 3 不同算法的计算量与参数量

| 上采样方法 | mIoU | Flops | Params |
|------------|-------------|-------|--------|
| Bilinear | 73.4 | 1M | 0 |
| CARAFE | 73.9 | 981M | 1.25M |
| Rezolution | 74.8 | 32M | 0.26M |

4.4 消融实验

在消融实验中，我们评估了特征更新算法的性能，我们使用三种方法进行对比：特

征拼接，GRU 更新，门控更新。如表 4 所示，门控方法虽然很简单，但是相比于其他两种方法取得了显著的提升，证明了我们特征融合方法的高效性。

表 4 不同上采样倍率下的提升

| 融合方法 | mIoU | mAcc | aAcc |
|--------|-------------|-------------|-------------|
| Concat | 74.0 | 85.1 | 94.0 |
| GRU | 73.5 | 84.9 | 93.8 |
| Gate | 74.8 | 85.4 | 94.1 |

5 结论

本文中，我们提出了一种全新的上采样方法 Rezolution。不同于当前流行的生成式算法它将上采样过程定义为一种图片特征锐化的过程，对双线性插值后的结果重新赋予重点，将语义分割的标准基准性能提高了 1.4% mIoU。更重要的是，Rezolution 引入的计算开销非常小，并且可以很容易地集成到各种网络架构中。

第 I 条 References

1. Ran B, Boyce D E. *Modeling Dynamic Transportation Network* (Book style). Berlin: Springer-Verlag, 1996, 69--83
2. Wang F Y, Fundamental Issues in Research of Computing with Words and Linguistic Dynamic Systems. *Acta Automatica Sinica* (Periodical style), 2005, **31**(6): 844--852
3. Roychoudhury R, Bandyopadhyay S, Paul K. A distributed mechanism for topology discovery in ad hoc wireless networks using mobile agents. In: *Proceeding of IEEE First Annual Workshop on Mobile and Ad hoc Networking and Computing* (Conference Proceedings style), Piscataway, USA: IEEE Press, 2000. 145--146
4. Hryniewicz O. An evaluation of the reliability of complex systems using shadowed sets and fuzzy lifetime data. *International Journal of Automation and Computing* (Periodical style—Accepted for publication), to be published.
5. Zhang W. *Reinforcement Learning for Job-Shop Scheduling*. [Ph.D.Dissertation], Oregon State University, 1996
6. *IEEE Criteria for Class IE Electric Systems* (Standards style), IEEE Standard 308, 1969.
7. Jones J. *Networks*, 2nd ed. (Online Sources style). [Online], available: <http://www.atm.com> , May 10, 1991.