

感受野（Receptive Field）

在深度学习中，“感受野”指的是网络中某个神经元（通常是卷积层输出中的一个位置）所“看到”的、在输入空间（如一张图像）上对应的区域大小。它决定了该神经元的输出对输入的多大范围内的像素变化敏感。

一、直观理解

1. 生物学起源

在视觉皮层中，一个神经元的感受野是指该神经元对视野中某个子区域刺激的响应范围。类似地，在卷积神经网络（CNN）里，每个输出特征图位置，也只“看到”输入图像的一个局部区域——这就是它的感受野。

2. 层级叠加

- 第一层卷积：**假设使用 3×3 卷积核，步幅（stride）为 1，则第一层每个输出像素对应输入的 3×3 区域，感受野大小是 3。
- 第二层卷积：**再用一个 3×3 核，对第一层的输出做卷积。此时，第二层的一个位置其实“看到”的是第一层输出的 3×3 区域，而每个第一层位置又对应输入的 3×3 区域，所以整体感受野变成了 5×5 。

二、数学计算

对于第 (i) 层，用以下符号：

- k_i ：卷积核大小（假设正方形，则核宽 = 核高）
- s_i ：该层的步幅（stride）
- j_i ：第 i 层输出相邻两点在输入图像上的间隔（也称“跳跃”或“累积步距”）
- r_i ：第 i 层输出的感受野大小

递推公式为：

$$\begin{cases} j_0 = 1, & r_0 = 1, \\ j_i = j_{i-1} \times s_i, \\ r_i = r_{i-1} + (k_i - 1) \times j_{i-1}. \end{cases}$$

- 初始条件：**把“输入像素”看作第 0 层，每个位置的感受野就是自身，($r_0=1$)，跳距 ($j_0=1$)。
- 跳距 (j_i)：**表示第 i 层输出相邻像素在原图上的距离。
 - 小跳距**（如 $j = 1$ ）：输出网格“密集”，能捕捉细节。
 - 大跳距**（如 $j > 1$ ）：输出网格“稀疏”，更偏重全局视野。
- 感受野 (r_i)：**在原图上，决定第 i 层某个神经元输出的输入区域大小。

示例：两层 3×3 卷积，步幅都为 1

层数 i	核大小 k_i	步幅 s_i	跳距 j_i	感受野 r_i
0（输入）	—	—	1	1
1	3	1	$1 \times 1 = 1$	$1 + (3 - 1) \times 1 = 3$
2	3	1	$1 \times 1 = 1$	$3 + (3 - 1) \times 1 = 5$

所以第二层输出每个点的感受野是 5×5 。

三、拓展点

1. 增加感受野的方法

- **加深网络**：堆叠更多卷积层。
- **增大卷积核**：如从 3×3 换成 5×5 。
- **下采样/池化**：步幅大于 1 的卷积或最大/平均池化都能快速扩大感受野，但同时降低空间分辨率。
- **空洞卷积 (Dilated Convolution)**：在卷积核内部插入“空洞”，以不增加参数的方式扩大感受野。
 - **普通卷积**：一个 $(k \times k)$ 卷积核在输入上按步幅 (stride) 滑动，每次覆盖连续的 $(k \times k)$ 区域。
 - **空洞卷积**：在卷积核元素之间插入 $(d - 1)$ 个“空洞” (dilation rate (d))，使得卷积核在输入上覆盖的区域变大，但只对“非空洞”位置做加权求和。又称“扩张卷积”或“膨胀卷积”，是在普通卷积核中“插入空洞”来增大感受野的一种技术。它能在不增加参数量和计算量的前提下，让网络“看到”更大范围的输入。

优点

- **在相同参数量下增大感受野**
理论感受野：
$$r_i = r_{i-1} + (k - 1) \times d \times j_{i-1}$$
通过空洞率 (d) 在不增加参数的前提下，显著扩大卷积“视野”。
- **保持特征图分辨率**
由于不使用下采样 (stride>1 或 pooling)，输出尺寸与输入相同 (配合合适的 padding)，可保留空间细节。
- **计算量与参数量不变**
空洞卷积的参数量与普通 $(k \times k)$ 卷积相同，FLOPs 基本不增加。
- **多尺度上下文捕获**
可并行使用不同空洞率 (如 ASPP 模块)，一次性获取多尺度信息。

缺点

- **棋盘效应 (Gridding Effect)**
当空洞率较大时，采样位置稀疏，输入信息可能“跳过”某些区域，导致网络难以捕捉连续特征。
- **对局部细节不敏感**
空洞采样间隔大，会降低对边缘和纹理等细微特征的响应。
- **空洞率设计复杂**
不同层或不同分支间的空洞率若搭配不当，易产生信息“盲区”，需要反复调优。
- **增大 Padding 可能引入伪影**
为保持尺寸，padding 通常设为 $(padding = d \cdot \frac{k-1}{2})$ ，过大时边缘可能出现填充伪影。

2. 理论 vs. 有效感受野

- **理论感受野**：如上公式计算出的大小。
- **有效感受野**：实验证明，网络的“真正”影响往往集中在理论感受野的中心区域，呈高斯分布，边缘权重趋近于零。要想让网络更好地利用大感受野，可能需要针对性设计 (如空洞卷积、注意力机制等)。

3. 应用场景

- **语义分割、目标检测**：需要足够大的感受野来捕获全局上下文，否则对大目标或场景理解能力受限。
- **风格迁移、超分辨率**：过大的感受野可能引入过多全局信息，反而影响细节还原。

小结

- **感受野** 衡量了网络某层输出对输入的“视野”范围。
- 通过简单的递推公式可以精确计算不同层的理论感受野。
- 在设计网络结构时，需要在感受野大小、参数量、计算量和空间分辨率之间权衡，以满足具体任务对“局部”与“全局”信息的需求。

1. 图像分类

- **小感受野**：网络只能聚焦于局部纹理和边缘信息，适合区分细粒度的局部特征，但容易忽略物体的整体形状与全局语义。
- **大感受野**：能够捕获整张图的布局和大尺度结构，有助于理解物体的全局轮廓与上下文，但若过大又可能引入多余背景噪声，降低对细节的敏感度。

实践中，大多数分类网络（如 ResNet、EfficientNet）在前几层用小卷积核累积逐步扩大感受野，最后通过全局平均池化（global pooling）实现全局感受野，再接全连接层完成分类。

2. 目标检测

- **局部特征**（小感受野）有助于精确定位小目标的边界和细节。
- **全局上下文**（大感受野）有助于抑制背景误检，并理解物体间的空间关系。

设计策略：

- **多尺度特征融合**（FPN、PANet）：在不同层（不同感受野）提取特征，再进行融合，兼顾小目标的细节与大目标的上下文。
- **可变形卷积**：在保持理论感受野的同时，自适应地调整采样位置，以更灵活地聚焦目标区域。