

Detail-Preserving Pooling in Deep Networks

2018/9/19 王诺

大多数卷积神经网络都是采用某种方法不断地减小隐含层的大小。这通常被称为池化，可用于减少参数的数量，改善某些失真的不变性，并增加感受野的大小。池化操作是一个有损失的过程，保持激励输出的特征对网络的鉴别能力是非常重要的。基于块的最大值和平均值、最大值池化、平均值池化、卷积下采样等为常见的池化标准操作。受专注于局部空间变化的人类视觉系统的启发，本文作者提出了一种细节保留池化操作，一种自适应池化方法，可以放大空间变化并保留重要的结构细节。同时它的参数可以与网络的其余部分一起学习。本文给出了在几个数据集上的实验结果，证明了该池化方法的优越性。

$$O[p] = \frac{1}{k_p} \sum_{q \in \Omega_p} I[q] \cdot \|I[q] - \tilde{I}[p]\|^\lambda$$

图 1

图 1 展示了细节保留的缩小图像方法（DPID）的计算公式，其中 I 代表输入图像， O 代表输出图像， $I[p]$ 代表 I 中 p 位置的像素值， \tilde{I} 代表输入图像经过盒式滤波和高斯平滑滤波后的结果。总地来说，DPID 计算地是输入图像 I 加权平均后的结果，距 p 越近的像素或与 p 处的值相差大的像素，对 p 对应的输出有更大的影响。具体流程如图 2 所示。

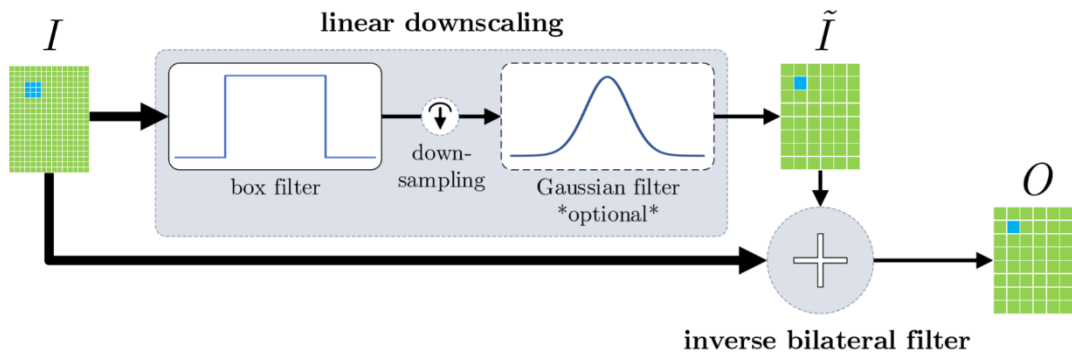


图 2

图 2 为细节保留的图像缩小方法对应的流程图。

$$\mathcal{D}_{\alpha,\lambda}(I)[p] = \frac{1}{\sum_{q' \in \Omega_p} w_{\alpha,\lambda}[p, q']} \sum_{q \in \Omega_p} w_{\alpha,\lambda}[p, q] I[q].$$

图 3

图 3 展示了本文提出的细节保留池化操作（DDP）相关的计算公式。即用一个可学习的回馈函数（reward function）替换 DPID 计算公式中 L_2 范数。其中可学习的权重 $w_{\alpha,\lambda}$ 的定义如图 4 所示。

$$w_{\alpha,\lambda}[p, q] = \alpha + \rho_{\lambda} \left(I[q] - \tilde{I}[p] \right)$$

图 4

图 4 为权重 $w_{\alpha,\lambda}$ 的定义，文中提供了两种 ρ_{λ} 函数，一种为对称型，另一种为非对称型。其中对称型的 ρ_{λ} 强化所有细节，而非对称性的 ρ_{λ} 只强调较为突出的细节。具体的函数定义如图 5 所示。

$$\rho_{\text{Sym}}(x) = \left(\sqrt{x^2 + \epsilon^2} \right)^{\lambda}$$

$$\rho_{\text{Asym}}(x) = \left(\sqrt{\max(0, x)^2 + \epsilon^2} \right)^{\lambda}$$

图 5

图 5 展示了对称型的 ρ_{λ} 和非对称性的 ρ_{λ} 的定义，其中 ϵ 为一个小常数。

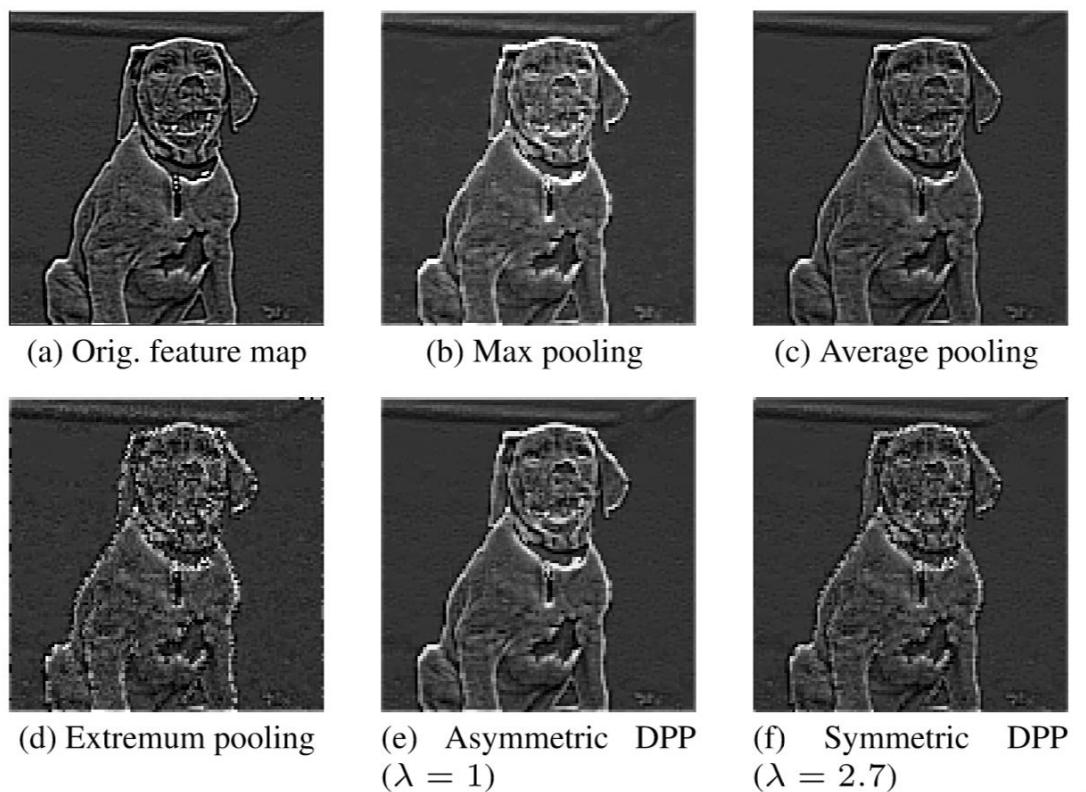


图 6

图 6 展示了池化操作后的数据可视化后的结果。

Network	Error [%]
ResNet-110 (standard, $2 \times$ strided convolution)	6.89
ResNet-110 ($1 \times$ strided convolution + $1 \times$ DPP)	6.68
ResNet-110 ($2 \times$ DPP)	6.59

图 7

图 7 通过错误率对比，展示了本文提出的池化方法相比于 strided convolution 方法的优越性。