4-2 卷积神经网络 I

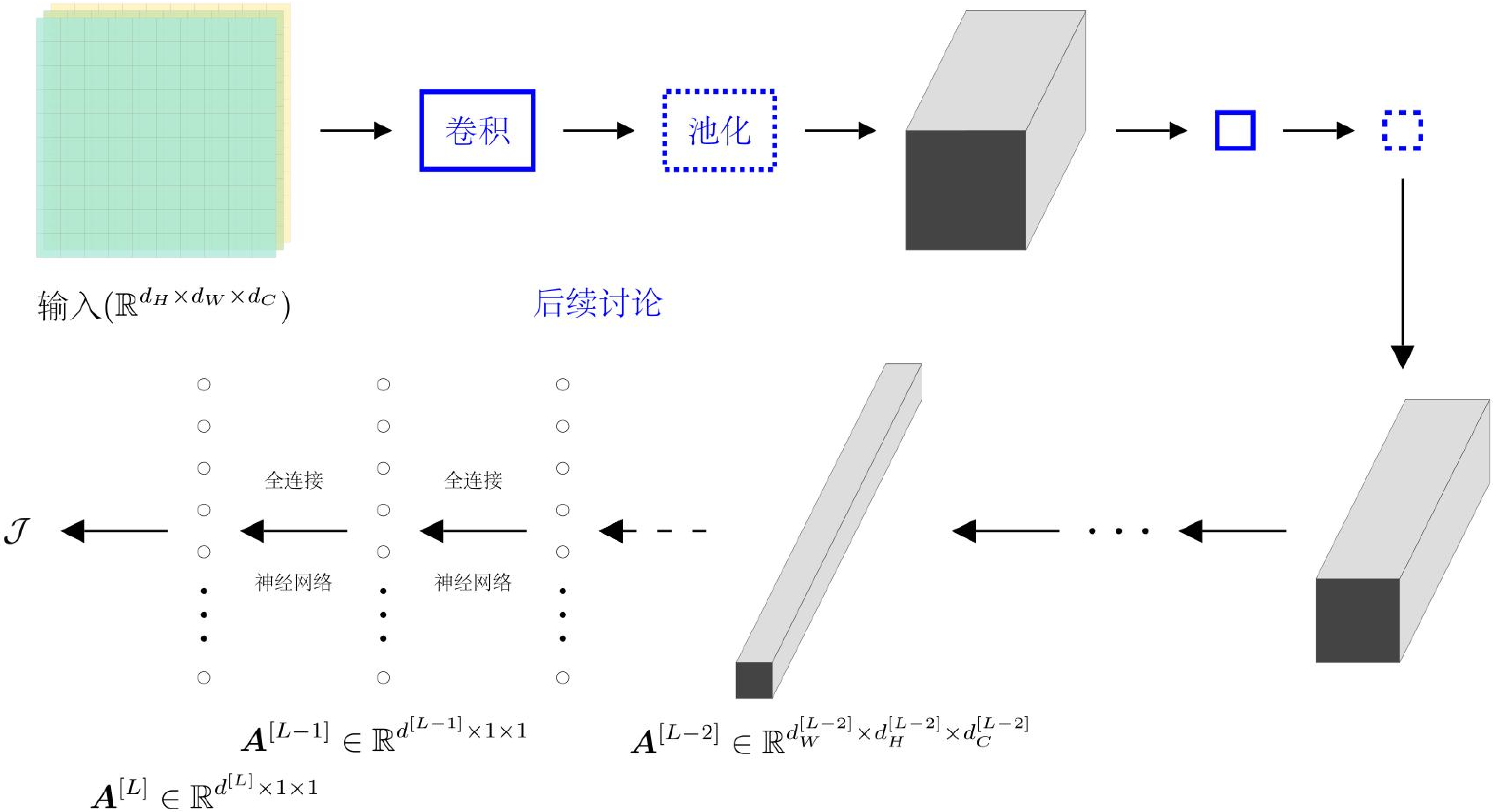
王中雷

厦门大学王亚南经济研究院和经济学院, 2025

内容摘要

简介

1. 当输入特征是图片时，为什么不考虑全连接神经网络
   - 模型参数将变得非常巨大，容易造成过拟合
     - ▷ 为了解决这个问题，我们需要更多的训练样本、内存以及更快的算法
     - ▷ 实际上几乎不可能做到
   - 忽略了图片中的空间结构
   - 不具有空间平移不变性
2. 针对图片处理，我们需要考虑一个新的模型框架（卷积神经网络）


卷积
池化
输入($\mathbb{R}^{d_H \times d_W \times d_C}$)
后续讨论
全连接
神经网络
全连接
神经网络
$\mathcal{J}$
$\boldsymbol{A}^{[L-1]} \in \mathbb{R}^{d^{[L-1]} \times 1 \times 1}$
$\boldsymbol{A}^{[L-2]} \in \mathbb{R}^{d_W^{[L-2]} \times d_H^{[L-2]} \times d_C^{[L-2]}}$
$\boldsymbol{A}^{[L]} \in \mathbb{R}^{d^{[L]} \times 1 \times 1}$

总结

1. 一般来讲，$d_C^{[l]}$随着层数l的增加而增加，但$d_W^{[l]}$ 和$d_H^{[l]}$ 降低
2. 向量化不涉及模型参数的估计
3. 通常来讲，卷积神经网络有如下三个“层”：
   - 卷积（Convolution）层（CONV）
   - 池化（Pooling）层 (POOL)
   - 全连接（FC）层

符号

1. l：层指标
2. $d_H^{[l]}$：第l 层对应“图片”的“高”
3. $d_W^{[l]}$：第l 层对应“图片”的“宽”
4. $d_C^{[l]}$：第l 层对应“图片”的“通道”数
5. $f^{[l]}$：第l 层对应卷积核的尺寸
6. $p^{[l]}$：第l 层对应填充（padding）数
7. $s^{[l]}$：第l 层对应步幅（stride）数

向量化

1. $\boldsymbol{A}^{[l-1]} \in \mathbb{R}^{n \times d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$：第$l$层的输入“图片”
2. $\boldsymbol{Z}^{[l]} \in \mathbb{R}^{n \times d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$：第$l$层的线性变换后的结果

$$\boldsymbol{Z}^{[l]} = \boldsymbol{A}^{[l-1]} \text{“} * \text{”} \boldsymbol{W}^{[l]} \text{“} + \text{”} \boldsymbol{b}^{[l]}$$
$$\boldsymbol{A}^{[l]} = \sigma^{[l]}(\boldsymbol{Z}^{[l]})$$

- $\boldsymbol{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times f^{[l]} \times f^{[l]} \times d_C^{[l-1]}}$：第$l$层的卷积核
- $\boldsymbol{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1 \times 1}$：第$l$层的偏置项
- “$*$”: 卷积运算
- “$+$”: 每个通道与特定的偏置项相加
- $\sigma^{[l]}(\cdot)$：第l层的激活函数

向量化

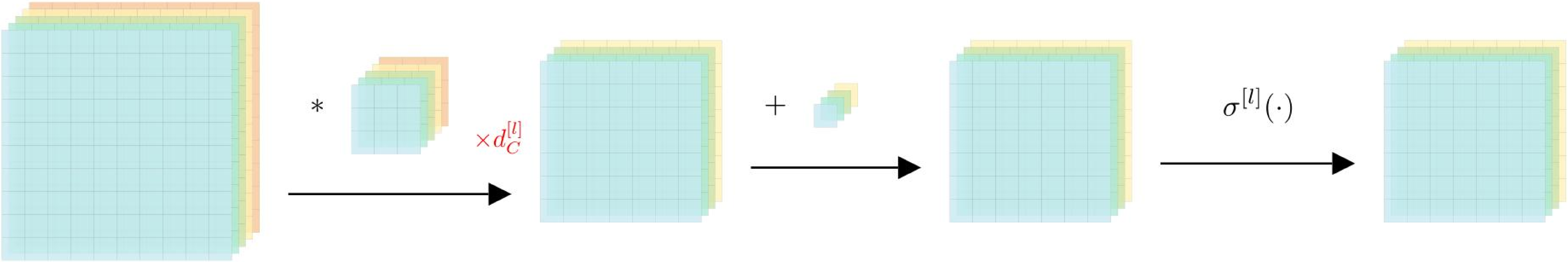
1. 简单起见，我们针对单一训练样本介绍前向和后向传播过程
   - $\boldsymbol{A}^{[l-1]} \in \mathbb{R}^{d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$
   - $\boldsymbol{Z}^{[l]} \in \mathbb{R}^{d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$
2. 前向传播

$$\boldsymbol{Z}^{[l]} = \boldsymbol{A}^{[l-1]} \text{“} * \text{”} \boldsymbol{W}^{[l]} \text{“} + \text{”} \boldsymbol{b}^{[l]}$$
$$\boldsymbol{A}^{[l]} = \sigma^{[l]}(\boldsymbol{Z}^{[l]})$$

卷积运算

1. 考虑:
   - $p^{[l]} = 0$
   - $s^{[l]} = 1$

卷积运算



$\boldsymbol{A}^{[l-1]} \in \mathbb{R}^{d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$

$\boldsymbol{Z}_0^{[l]} = \mathbb{R}^{d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$

$\boldsymbol{Z}^{[l]} \in \mathbb{R}^{d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$

$\boldsymbol{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times f^{[l]} \times f^{[l]} \times d_C^{[l-1]}}$

$\boldsymbol{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1 \times 1}$

$\boldsymbol{A}^{[l]} \in \mathbb{R}^{d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$

池化（Pooling）

1. 池化（Pooling）运算往往用于卷积运算之后，
   - 在每个通道上，降低“图片”尺寸
   - 通过忽略重复或者无用信息，提高模型稳健性
   - 增加感受野（receptive field）
   - 对输入图片的小规模扰动、旋转以及其他不利变换达到稳健性
   - 防止过拟合（尤其与 dropout 同时使用时）

池化（Pooling）

1. 常见的两种池化运算：
   - 平均池化（Average pooling）
   - 最大池化（Max pooling）
2. 注意：
   - 池化运算针对每个通道单独进行
   - 池化运算并不引入新的模型参数

平均池化（Average pooling）

1. 输入尺寸: 6×6
2. （池化）核尺寸: 2×2
3. 步幅大小（Stride）：2

平均池化（Average pooling）

输入

3	6	5	4	8	9
1	7	9	6	8	0
5	0	9	6	2	0
5	2	6	3	7	0
9	0	3	2	3	1
3	1	3	7	1	7

（池化）核

0.25	0.25
0.25	0.25

结果

4.25	6.0	6.25
3.0	6.0	2.25
3.25	3.75	3.0

最大池化（Max pooling）

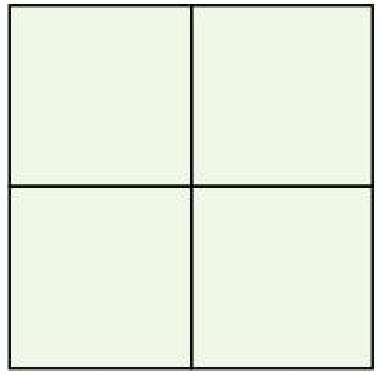
1. 输入尺寸: 6×6
2. （池化）核尺寸: 2×2
3. 步幅大小（Stride）：2

最大池化（Max pooling）

输入

3	6	5	4	8	9
1	7	9	6	8	0
5	0	9	6	2	0
5	2	6	3	7	0
9	0	3	2	3	1
3	1	3	7	1	7

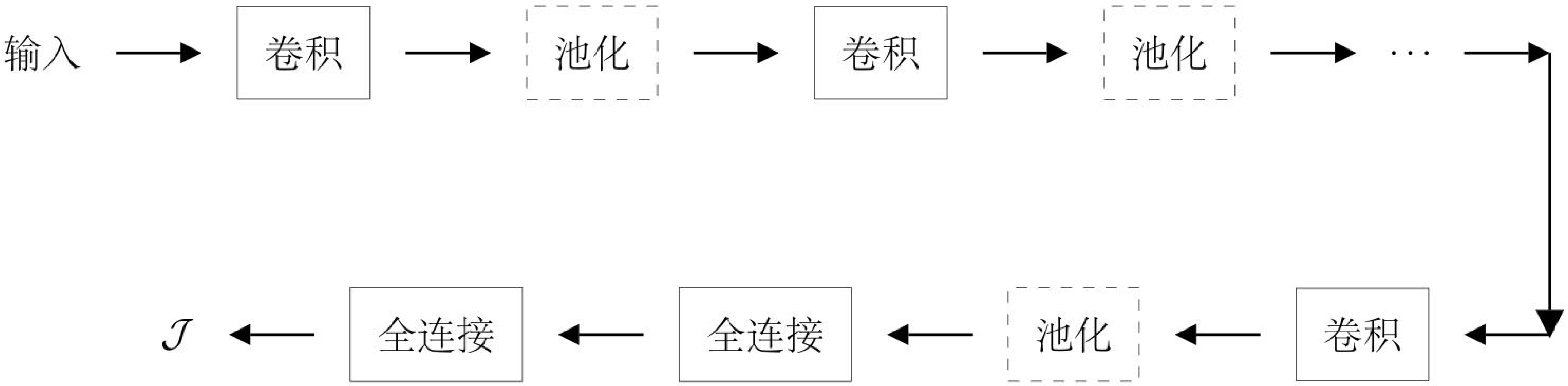
（池化）核



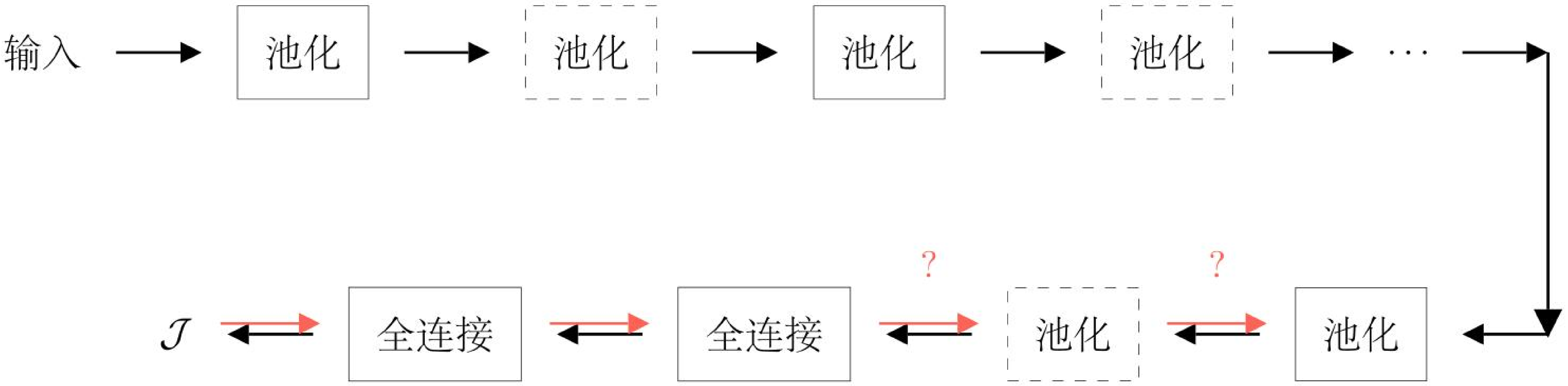
结果

7	9	9
5	9	7
9	7	7

前向传播



后向传播



后向传播

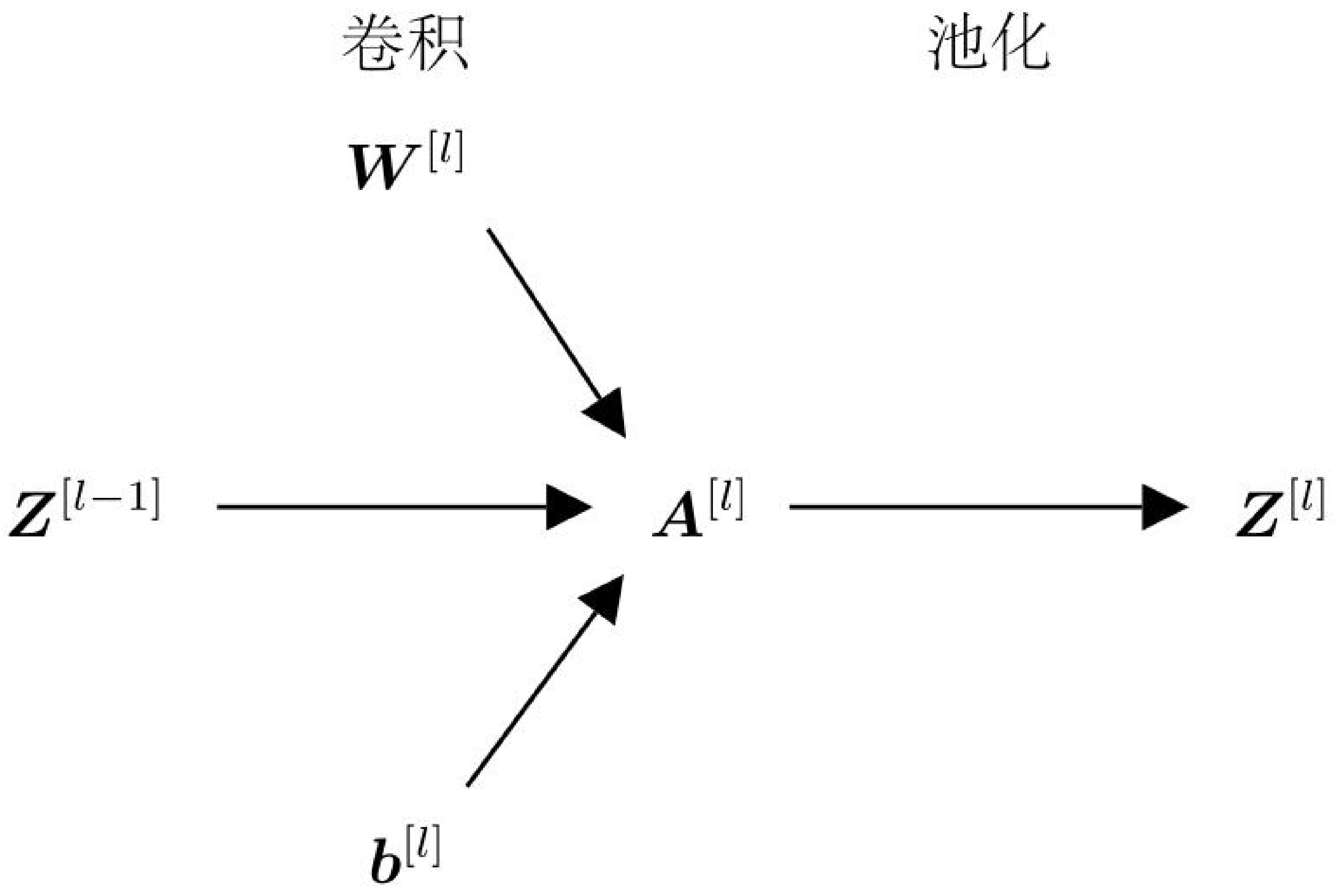
1. 我们已经学习过全连接神经网络的反向传播过程
2. 因此，我们只需讨论下列运算的反向传播过程
    - 池化层（POOL）
    - 卷积层 (CONV)
3. 在下面的分析中，我们假设卷积运算后立即进行池化运算

后向传播

1. 记

- $\boldsymbol{A}^{[l-1]} \in \mathbb{R}^{? \times ? \times d_C^{[l-1]}}$：第$(l-1)$层卷积运算输出
- $\boldsymbol{Z}^{[l-1]} \in \mathbb{R}^{d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$：第$(l-1)$层池化运算输出，其输入为$\boldsymbol{A}^{[l-1]}$
- $\boldsymbol{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times f^{[l]} \times f^{[l]} \times d_C^{[l-1]}}$：第$l$层卷积运算的卷积核，其输入为$\boldsymbol{Z}^{[l-1]}$
- $\boldsymbol{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1 \times 1}$：第$l$层卷积运算的偏置项，其输入为$\boldsymbol{Z}^{[l-1]}$
- $\boldsymbol{A}^{[l]} \in \mathbb{R}^{? \times ? \times d_C^{[l]}}$：第$l$层卷积运算输出
- $\boldsymbol{Z}^{[l]} \in \mathbb{R}^{d_H^{[l]} \times d_W^{[l]} \times d_C^{[l]}}$：第$l$层池化运算输出，其输入为$\boldsymbol{A}^{[l]}$

结构



1. 假设 $\mathrm{d}\boldsymbol{Z}^{[l]}$ 是已知的

平均池化的反向传播

$\boldsymbol{A}^{[l]}$

3	6	5	4	8	9
1	7	9	6	8	0
5	0	9	6	2	0
5	2	6	3	7	0
9	0	3	2	3	1
3	1	3	7	1	7

（池化）核$\boldsymbol{M}$

0.25	0.25
0.25	0.25

$\boldsymbol{Z}^{[l]}$

4.25	6.0	6.25
3.0	6.0	2.25
3.25	3.75	3.0

平均池化的反向传播

1. 对于平均池化，我们有

$$\mathrm{d}\boldsymbol{A}^{[l]} = \mathrm{d}\boldsymbol{Z}^{[l]} \otimes \boldsymbol{M}$$

- $\boldsymbol{A} \otimes \boldsymbol{B}$：两个矩阵$\boldsymbol{A}$ 以及$\boldsymbol{B}$ 的 Kronecker 乘积

最大池化的反向传播

$\boldsymbol{A}^{[l]}$

3	6	5	4	8	9
1	7	9	6	8	0
5	0	9	6	2	0
5	2	6	3	7	0
9	0	3	2	3	1
3	1	3	7	1	7

（池化）核

$\boldsymbol{Z}^{[l]}$

7	9	9
5	9	7
9	7	7

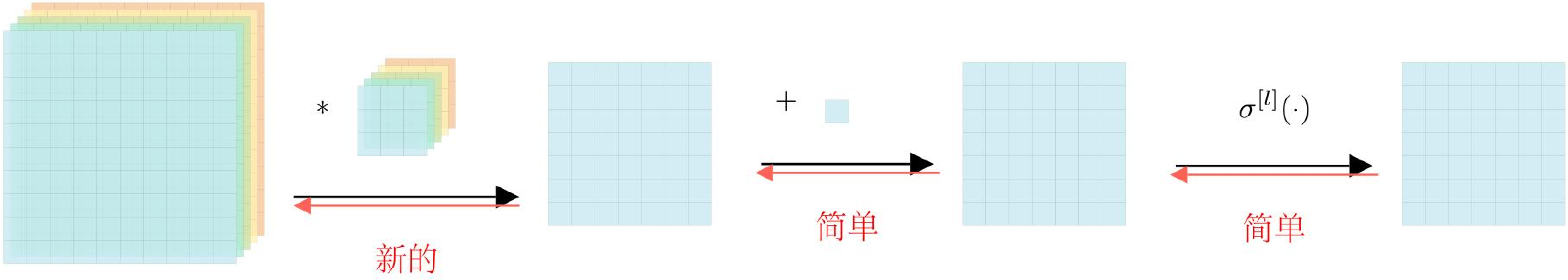
最大池化的反向传播

1. 对于最大池化而言，$\mathrm{d}\boldsymbol{A}^{[l]}$ 是通过如下过程得到的
   - 将矩阵$\boldsymbol{Z}^{[l]}$ 中每个元素乘上一个对应的2×2 掩码矩阵，再相加
   - 这些2×2 的掩码矩阵中，只有一个元素为 1，其他为 0
2. 我们已经介绍完了从$\mathrm{d}\boldsymbol{Z}^{[l]}$ 到$\mathrm{d}\boldsymbol{A}^{[l]}$ 的反向传播过程
   - 池化运算不涉及任何额外的模型参数
3. 在后续分析中，我们假设$\mathrm{d}\boldsymbol{A}^{[l]}$ 已知

卷积运算的后向传播

1. 简便起见，假设卷积运算结果的通道数为$d_C^{[l]} = 1.$
2. 假设$\mathrm{d}\boldsymbol{A}^{[l]}$ 已知

卷积运算的后向传播



$\boldsymbol{Z}^{[l-1]} \in \mathbb{R}^{d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$ $\quad$ $\boldsymbol{Z}_0^{[l]} \in \mathbb{R}^{? \times ? \times 1}$ $\quad$ $\boldsymbol{Z}_1^{[l]} \in \mathbb{R}^{? \times ? \times 1}$

$\boldsymbol{W}^{[l]} \in \mathbb{R}^{f^{[l]} \times f^{[l]} \times d_C^{[l-1]}}$ $\quad$ $\boldsymbol{b}^{[l]} \in \mathbb{R}^{1 \times 1 \times 1}$ $\quad$ $\boldsymbol{A}^{[l]} \in \mathbb{R}^{? \times ? \times 1}$

$\mathrm{d}\boldsymbol{W}^{[l]} = ?$ $\quad$ $\mathrm{d}\boldsymbol{b}^{[l]} = \sum_{ij} \mathrm{d}Z_{1,ij}^{[l]}$

$\mathrm{d}\boldsymbol{Z}^{[l-1]} = ?$ $\quad$ $\mathrm{d}\boldsymbol{Z}_0^{[l]} = \mathrm{d}\boldsymbol{Z}_1^{[l]}$ $\quad$ $\mathrm{d}\boldsymbol{Z}_1^{[l]} = \sigma^{[l]\prime}(\boldsymbol{Z}^{[l]}) \circ \mathrm{d}\boldsymbol{A}^{[l]}$

卷积运算的后向传播

1. 为了求函数对某自变量的导数，我们需要知道该自变量被用于哪些计算
2. 初始化$\mathrm{d}\boldsymbol{Z}^{[l-1]}$为一个 0 矩阵，其规模与$\boldsymbol{Z}^{[l-1]}$相同
3. 符号不那么严格地，我们可以得到

$$\mathrm{d}\boldsymbol{Z}^{[l-1]}_{\text{slice},ij} += \mathrm{d}Z^{[l]}_{0,ij} \times \boldsymbol{W}^{[l]}$$

- $\boldsymbol{Z}^{[l-1]}_{\text{slice},ij}$：$\boldsymbol{Z}^{[l-1]}$中被用于得到$Z^{[l]}_{0,ij}$的部分

卷积运算的后向传播

1. 很明显，$\boldsymbol{Z}_0^{[l]}$ 中的每个元素均包含$\boldsymbol{W}^{[l]}$ 的信息
2. 因此，我们有

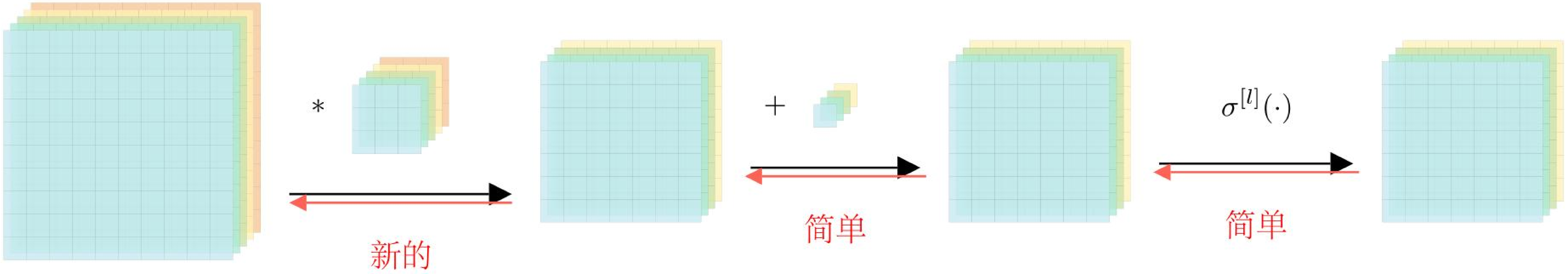
$$\mathrm{d}\boldsymbol{W}^{[l]}=\sum_{ij}\mathrm{d}Z_{0,ij}^{[l]}\times\boldsymbol{B}_{ij}^{[l-1]}$$

- $\boldsymbol{B}_{ij}^{[l-1]}$: $\boldsymbol{Z}^{[l-1]}$ 中被用于得到$Z_{0,ij}^{[l]}$ 的部分

卷积运算的后向传播

1. 更一般地，考虑输出有$d_C^{[l]}$ 个通道的情况

卷积运算的后向传播



$\boldsymbol{Z}^{[l-1]} \in \mathbb{R}^{d_H^{[l-1]} \times d_W^{[l-1]} \times d_C^{[l-1]}}$ $\quad$ $\boldsymbol{Z}_0^{[l]} \in \mathbb{R}^{? \times ? \times d_C^{[l]}}$ $\quad$ $\boldsymbol{Z}_1^{[l]} \in \mathbb{R}^{? \times ? \times d_C^{[l]}}$

$\boldsymbol{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times f^{[l]} \times f^{[l]} \times d_C^{[l-1]}}$ $\quad$ $\boldsymbol{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1 \times 1}$ $\quad$ $\boldsymbol{A}^{[l]} \in \mathbb{R}^{? \times ? \times d_C^{[l]}}$

$\mathrm{d}\boldsymbol{W}^{[l]} = ?$

$$\mathrm{d}\boldsymbol{b}_c^{[l]} = \sum_{ij} \mathrm{d}Z_{1,ijc}^{[l]} \quad (c = 1, \ldots, d_C^{[l]})$$

$\mathrm{d}\boldsymbol{Z}^{[l-1]} = ?$ $\quad$ $\mathrm{d}\boldsymbol{Z}_0^{[l]} = \mathrm{d}\boldsymbol{Z}_1^{[l]}$ $\quad$ $\mathrm{d}\boldsymbol{Z}_1^{[l]} = \sigma^{[l]\prime}(\boldsymbol{Z}^{[l]}) \circ \mathrm{d}\boldsymbol{A}^{[l]}$

卷积运算的后向传播

1. 在计算导数时，我们只需要找到对应自变量在哪些计算被用到即可
2. 矩阵$\boldsymbol{Z}^{[l-1]}$ 的信息存在于$\boldsymbol{Z}_0^{[l]}$ 的每个元素中
3. 将$\mathrm{d}\boldsymbol{Z}^{[l-1]}$ 初始化为一个 0 矩阵，其规模与$\boldsymbol{Z}^{[l-1]}$ 相同
4. 符号不是那么严谨地，我们有

$$\mathrm{d}\boldsymbol{Z}_{\text{slice},ij}^{[l-1]} \mathrel{+}= \sum_{c=1}^{d_C^{[l]}} \mathrm{d}Z_{0,ij\color{red}{c}}^{[l]} \times \boldsymbol{W}_{\color{red}{c}}^{[l]}$$

- $\boldsymbol{Z}_{\text{slice},ij}^{[l-1]}$: 矩阵$\boldsymbol{Z}^{[l-1]}$ 中被用于得到$Z_{0,ijc}^{[l]}$ 的部分，其中$c = 1, \ldots, d_C^{[l]}$
- $\boldsymbol{W}_{\color{red}{c}}^{[l]}$: 第 l 层中第 c 个卷积核

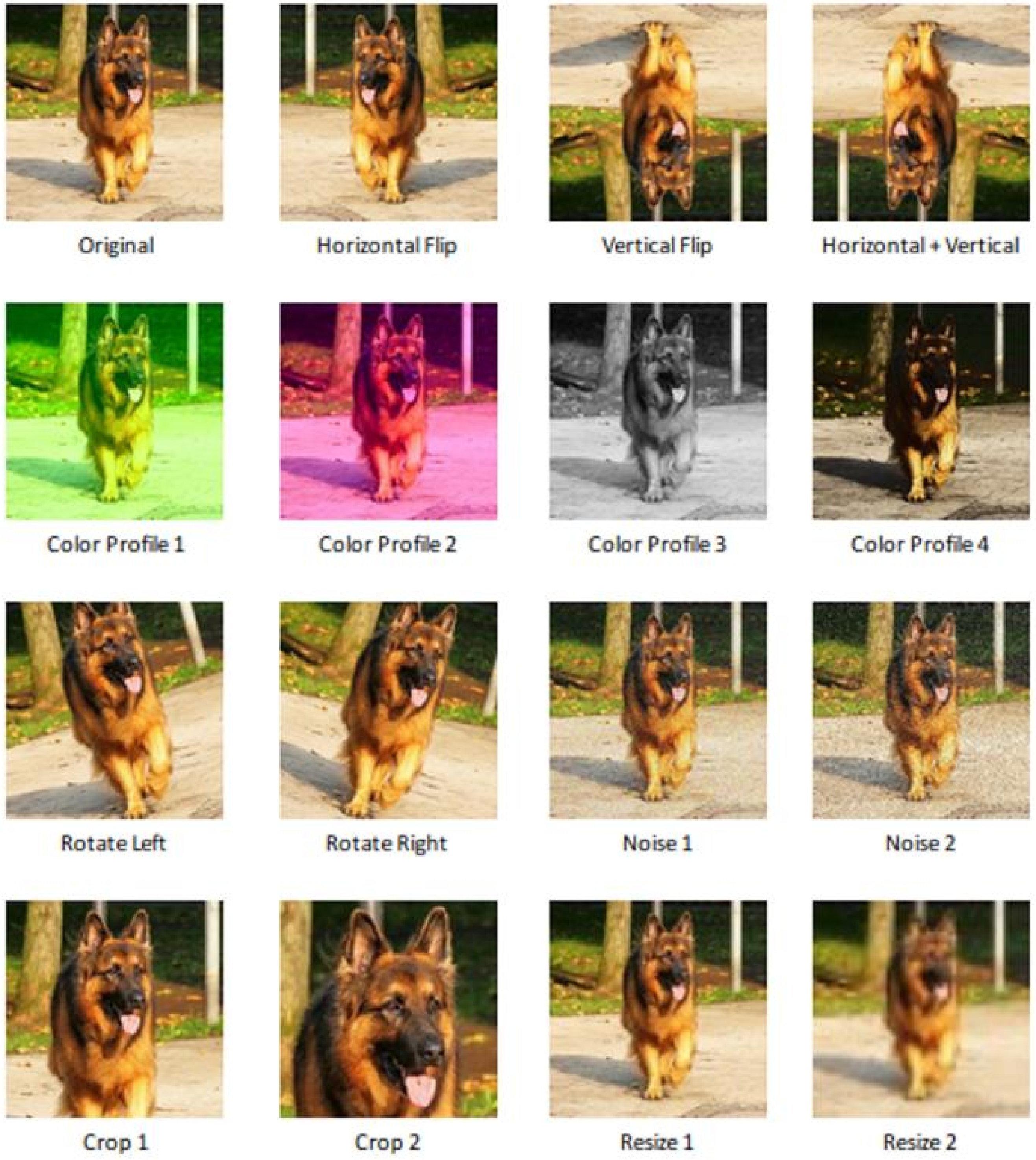
卷积运算的后向传播

1. 很明显，$\boldsymbol{Z}^{[l]}_{0,c}$ 中的每个元素均包含$\boldsymbol{W}^{[l]}_{c}$ 的信息，其中$c = 1, \ldots, d^{[l]}_{C}$
2. 因此，我们有

$$\mathrm{d}\boldsymbol{W}^{[l]}_{c} = \sum_{ij} \mathrm{d}Z^{[l]}_{0,ijc} \times \boldsymbol{B}^{[l-1]}_{ij} \quad (c = 1, \ldots, d^{[l]}_{C})$$

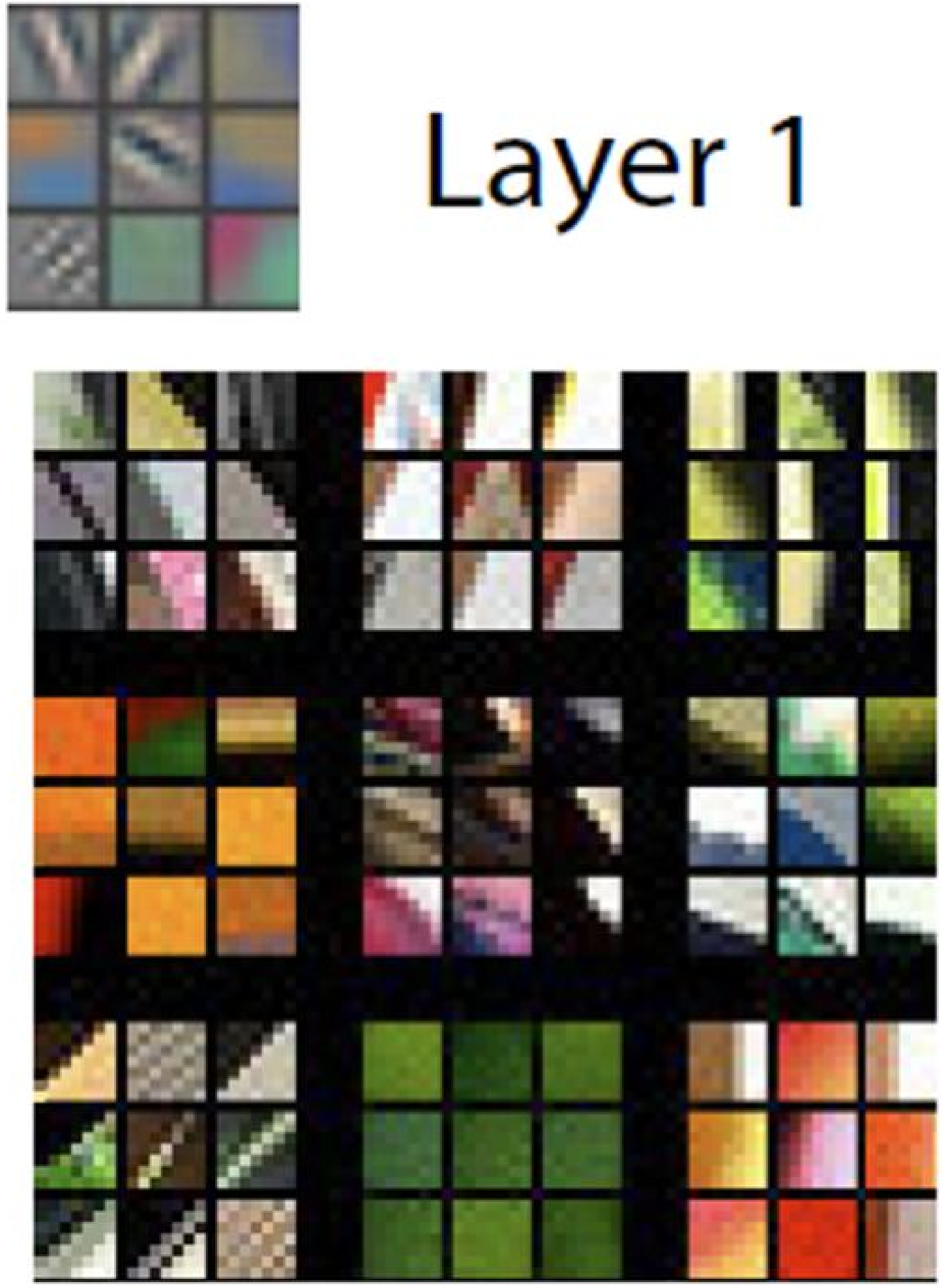
- $\boldsymbol{B}^{[l-1]}_{ij}$: 矩阵$\boldsymbol{Z}^{[l-1]}$ 中被用于计算$Z^{[l]}_{0,ijc}$ 的部分

数据增强（Data augmentation）



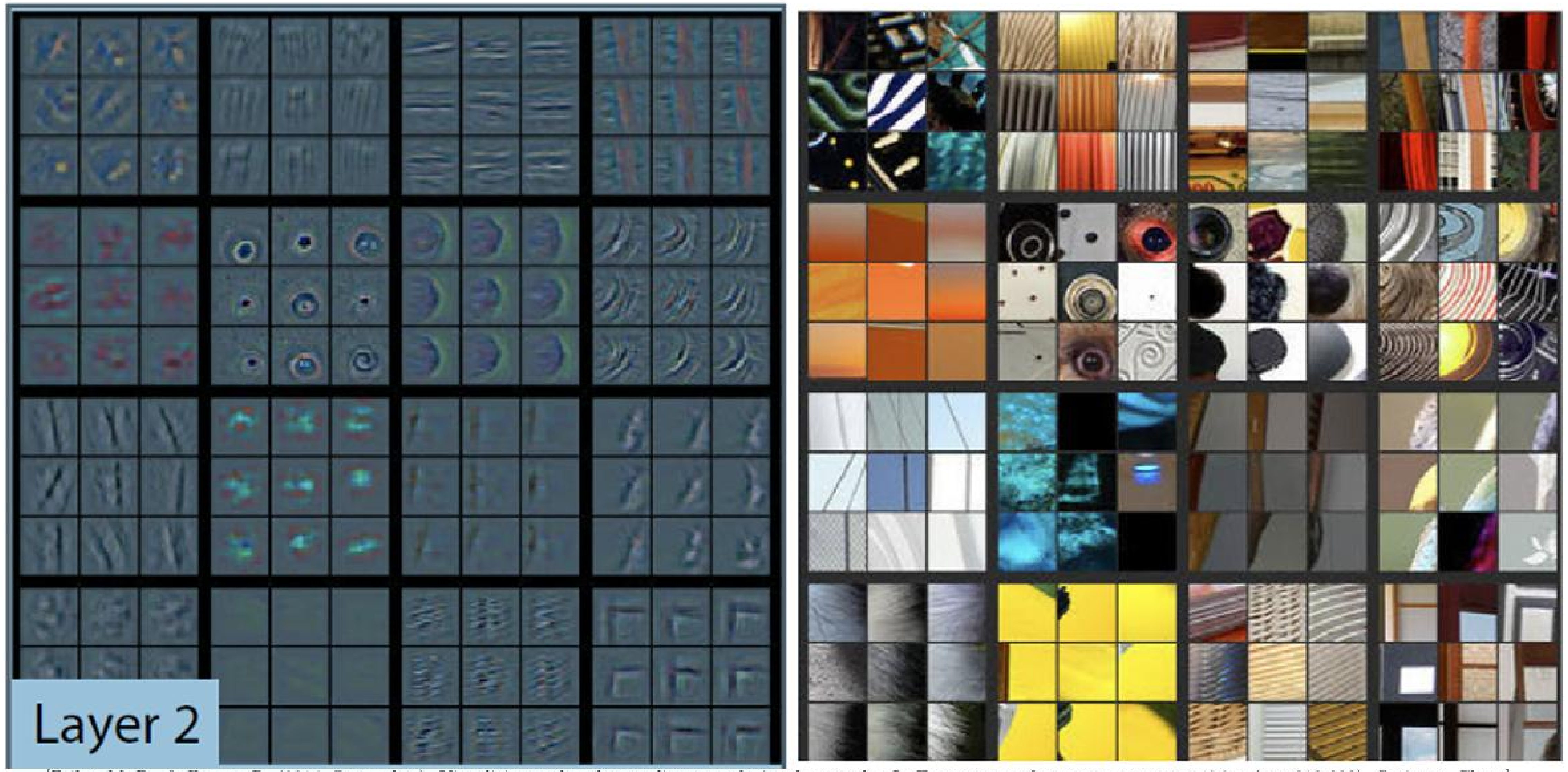
[https://www.volansys.com/blog/data-augmentation-in-ml/]

理解卷积神经网络



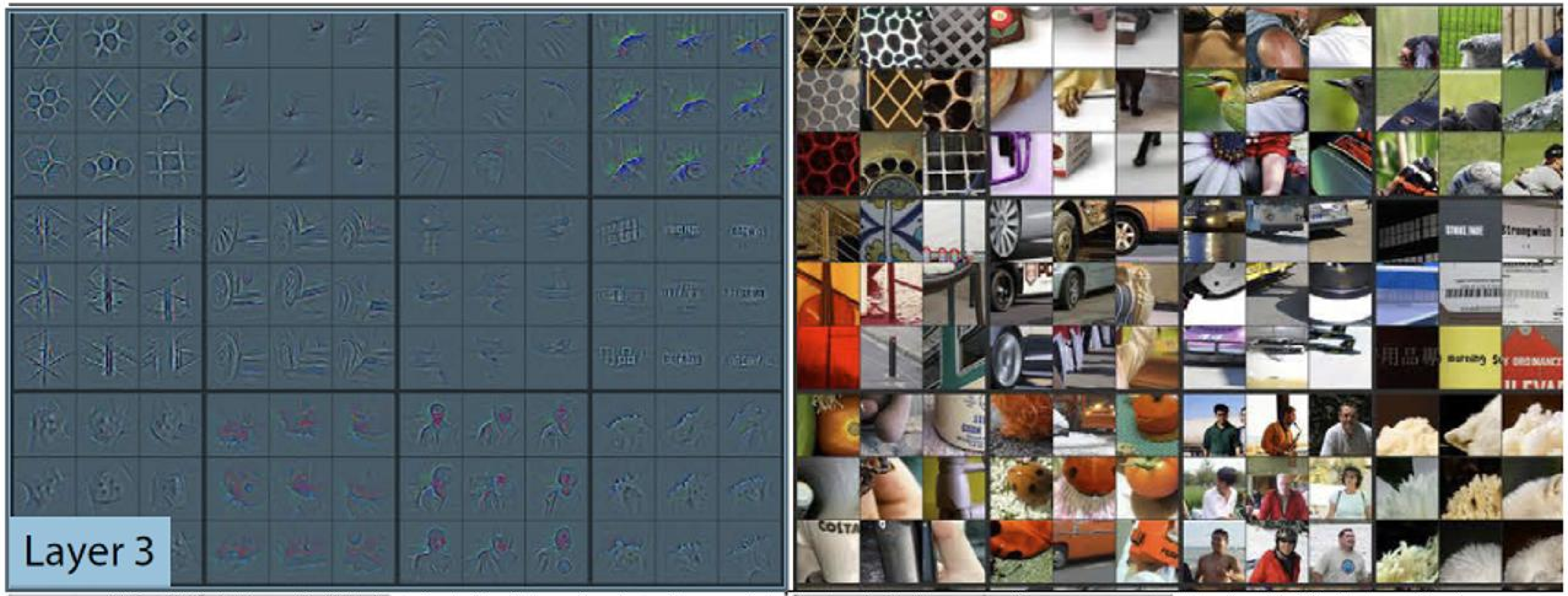
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



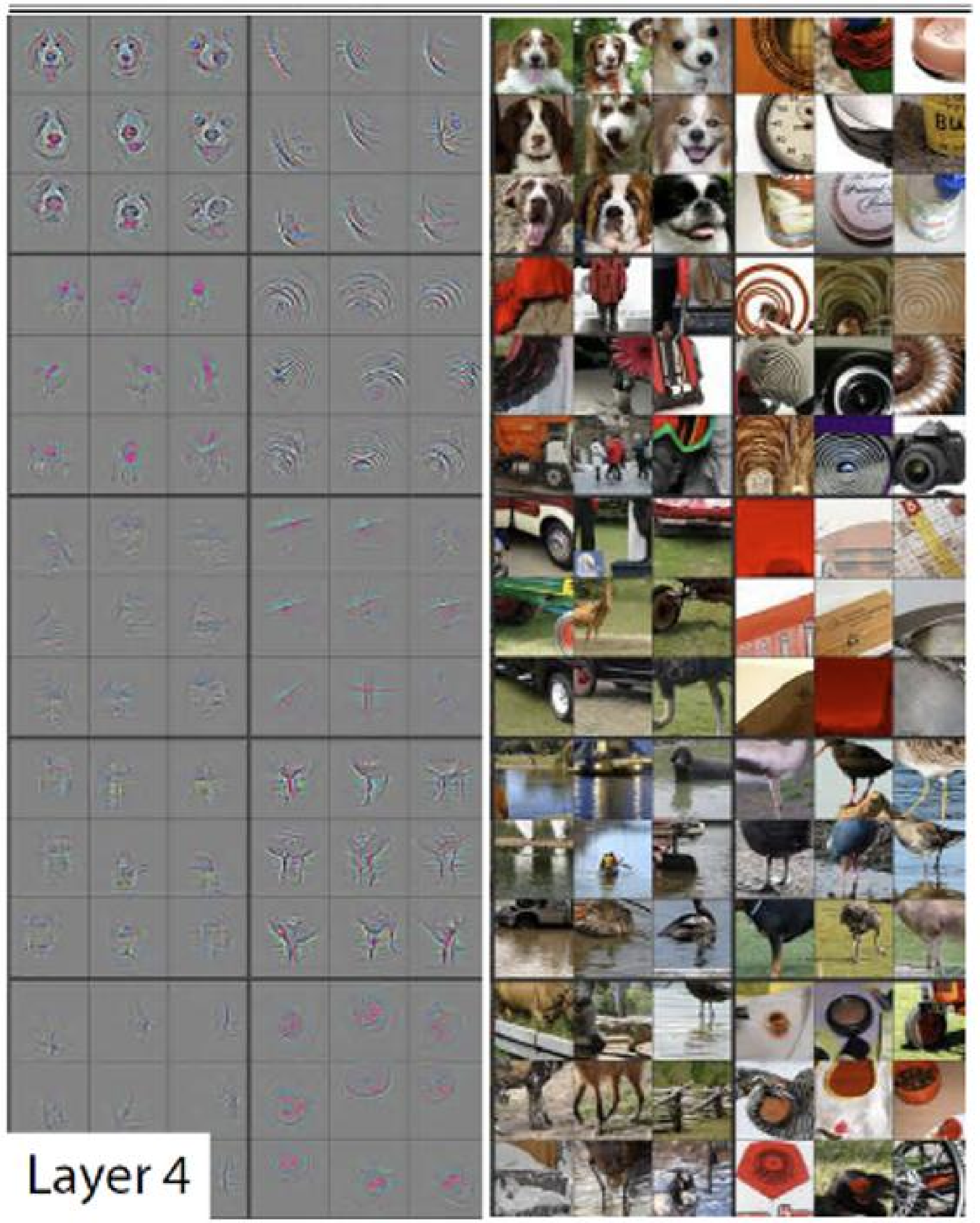
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



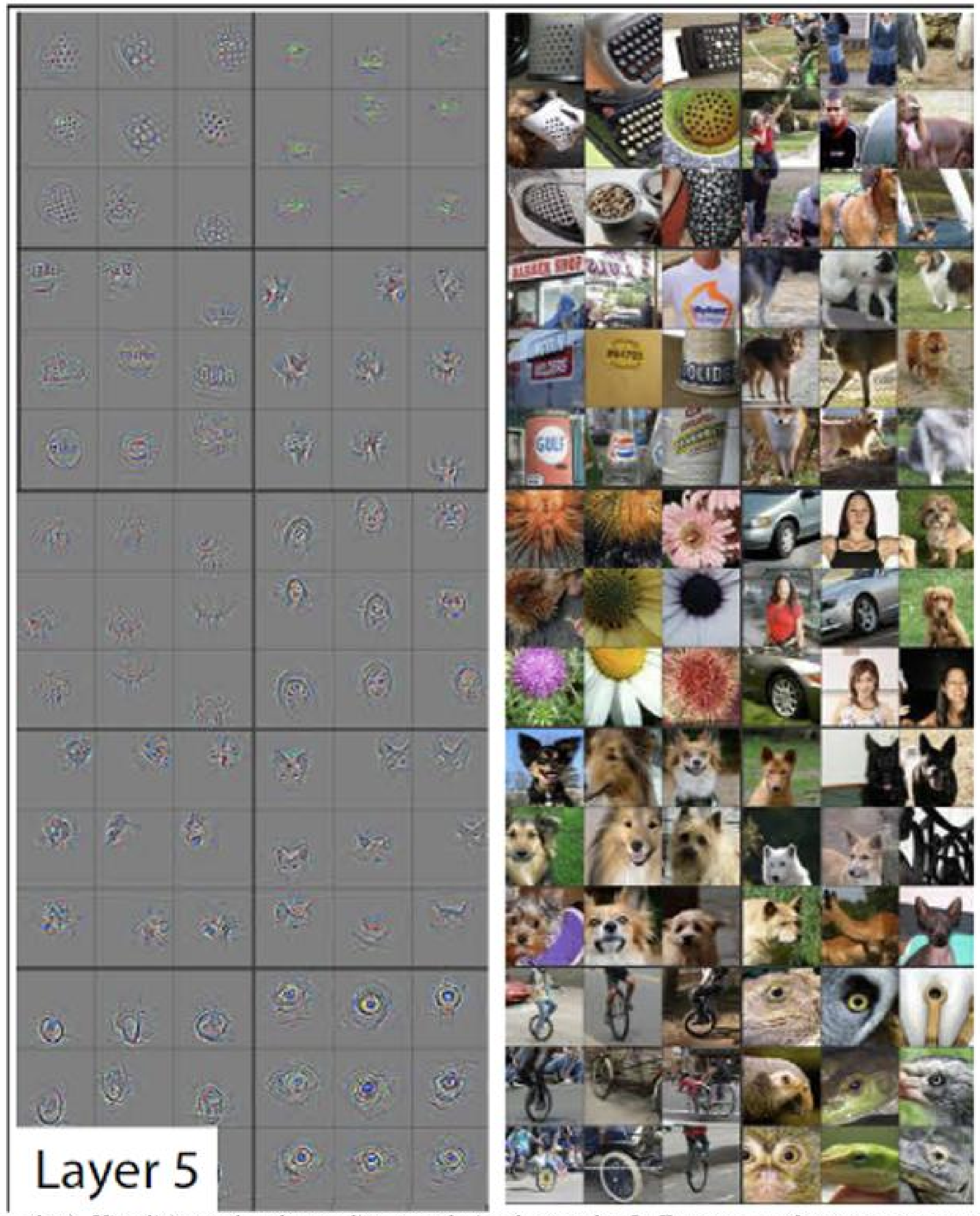
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]