



人工智能与深度学习

4-2 卷积神经网络 I

王中雷

经济学院、王亚南经济研究院、邹至庄经济研究院

厦门大学

(第一版)

内容摘要

1. 基本框架

2. 前向传播

3. 后向传播

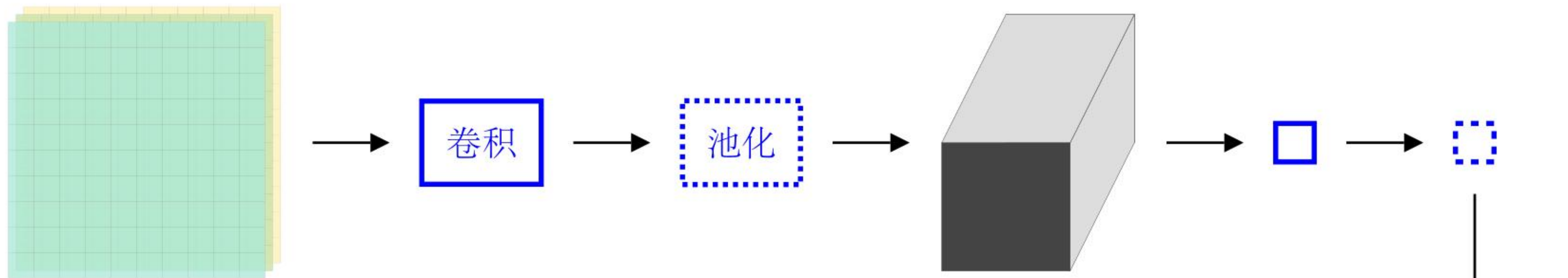
4. 理解卷积神经网络

背景

1. 当输入特征是图片时，为什么不考虑全连接神经网络

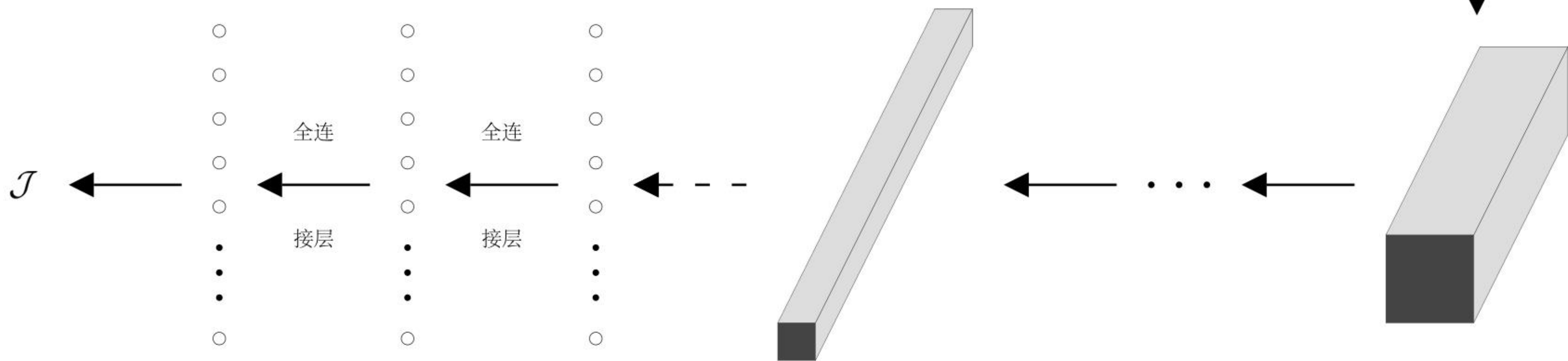
- 模型参数将变得非常巨大，容易造成过拟合
 - ▷ 为了解决这个问题，我们需要更多的训练样本、内存以及更快的算法
 - ▷ 在高维图像上通常会带来很高的存储与计算成本
- 忽略了图片中的空间结构
- 不具有空间平移等变性

2. 针对图片处理，我们需要考虑一个新的模型框架，即卷积神经网络



输入($\mathbb{R}^{d_C \times d_H \times d_W}$)

后续讨论



$$\mathbf{A}^{[L-1]} \in \mathbb{R}^{d^{[L-1]} \times 1 \times 1}$$

$$\mathbf{A}^{[L-2]} \in \mathbb{R}^{d_C^{[L-2]} \times d_H^{[L-2]} \times d_W^{[L-2]}}$$

$$\mathbf{A}^{[L]} \in \mathbb{R}^{d^{[L]} \times 1 \times 1}$$

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. 展平 (flatten) 操作, 不涉及新的模型参数
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 (Fully Connected) 层 (FC)

符号

1. l : 层指标
2. $d_C^{[l]}$: 第 l 层对应“图片”的“通道”数
3. $d_H^{[l]}$: 第 l 层对应“图片”的“高”
4. $d_W^{[l]}$: 第 l 层对应“图片”的“宽”
5. $f^{[l]}$: 第 l 层对应卷积核的尺寸（默认 $f_1^{[l]} = f_2^{[l]}$ ）
6. $p^{[l]}$: 第 l 层对应的填充（padding）（默认 $p_1^{[l]} = p_2^{[l]}$ ）
7. $s^{[l]}$: 第 l 层对应的步幅（stride）（默认 $s_1^{[l]} = s_2^{[l]}$ ）
8. 不考虑空洞卷积，即令空洞率为 1

卷积层的前向传播

1. $\mathbf{Z}^{[l]} \in \mathbb{R}^{n \times d_C^{[l]} \times d_H^{[l]} \times d_W^{[l]}}$: 线性输出

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

- $\mathbf{A}^{[l-1]} \in \mathbb{R}^{n \times d_C^{[l-1]} \times d_H^{[l-1]} \times d_W^{[l-1]}}$: 第 l 层的输入 “图片”
- $\mathbf{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times d_C^{[l-1]} \times f^{[l]} \times f^{[l]}}$: 第 l 层的卷积核
- $\mathbf{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1}$: 第 l 层的偏置项
- “ $*$ ”: 卷积运算
- “ $+$ ”: 每个通道与特定的偏置项相加

2. $\mathbf{A}^{[l]} \in \mathbb{R}^{n \times d_C^{[l]} \times d_H^{[l]} \times d_W^{[l]}}$: 激活输出

$$\mathbf{A}^{[l]} = \sigma^{[l]}(\mathbf{Z}^{[l]})$$

- $\sigma^{[l]}(\cdot)$: 第 l 层的激活函数

卷积层的前向传播

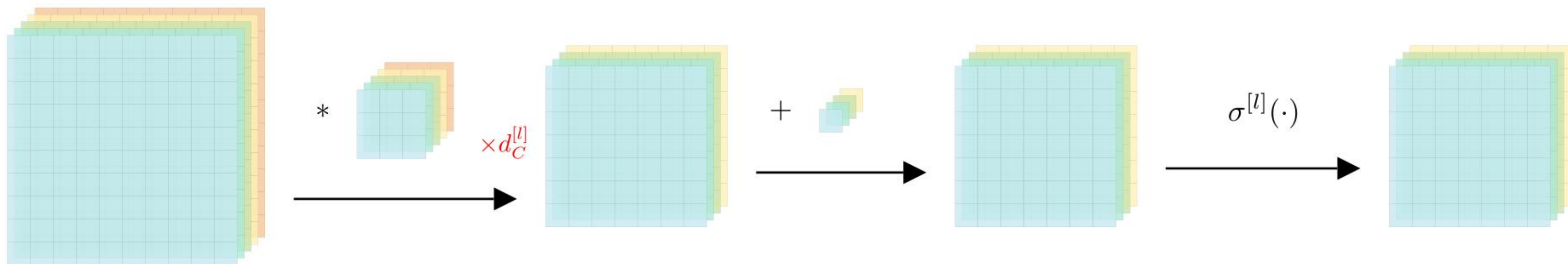
1. 简单起见，我们针对单一训练样本，详细介绍卷积层的前向和后向传播过程

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

2. 前向传播

$$\begin{aligned}\mathbf{Z}^{[l]} &= \mathbf{A}^{[l-1]} \ast \mathbf{W}^{[l]} + \mathbf{b}^{[l]}, \\ \mathbf{A}^{[l]} &= \sigma^{[l]}(\mathbf{Z}^{[l]})\end{aligned}$$

直观解释



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

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

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

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

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

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

池化 (Pooling)

1. 池化 (Pooling) 运算，往往用于卷积层之后，以实现如下作用

- 在每个通道内，进行局部聚合和下采样
- 可降低空间分辨率及后续计算量
- 增加感受野 (receptive field)
- 可能提升模型对部分局部扰动的稳健性
- 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

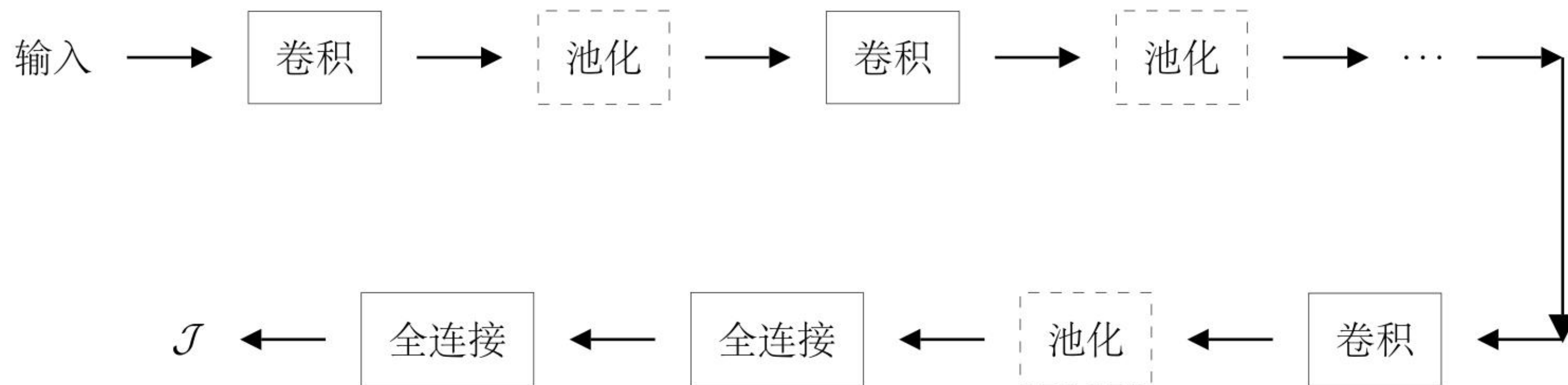
3	6	5	4	8	9
1	7	9	6	8	0
5	0	9	6	2	0
5	2	6	3	7	0
9	0	3	2	3	1
3	1	3	7	1	7

(池化) 核

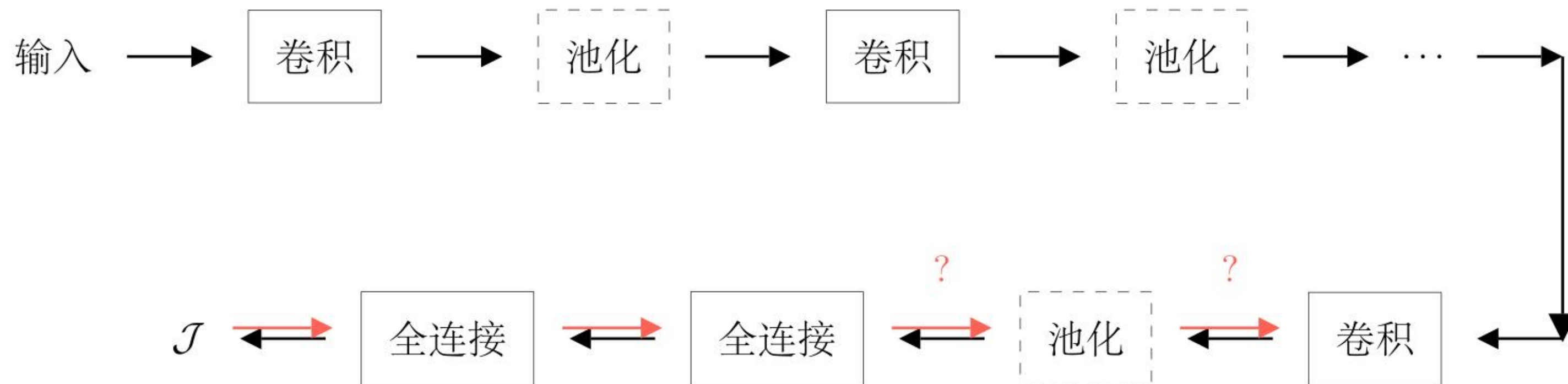
结果

7	9	9
5	9	7
9	7	7

前向传播



后向传播



1. 我们已经学习过全连接神经网络的后向传播过程
2. 因此，我们只需讨论下列运算的后向传播过程
 - 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

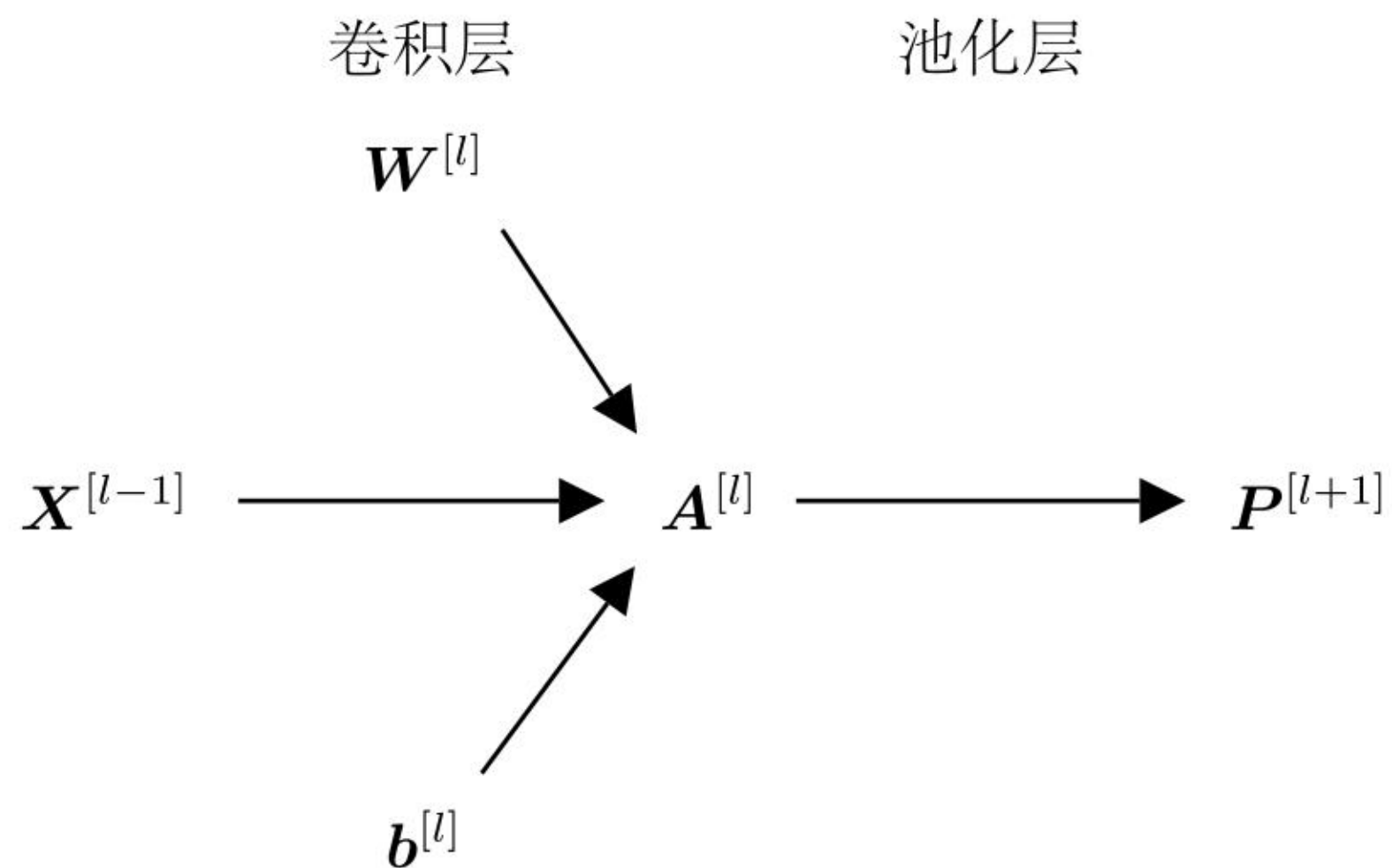
1. 考虑下面的情况：

- 第 l 层为卷积层
- 第 $(l + 1)$ 层为池化层

2. 记

- $\mathbf{W}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times d_C^{[l-1]} \times f^{[l]} \times f^{[l]}}$ ：第 l 层卷积运算的卷积核
- $\mathbf{b}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times 1 \times 1}$ ：第 l 层卷积运算的偏置项，其输入为 $\mathbf{P}^{[l-1]}$
- $\mathbf{A}^{[l]} \in \mathbb{R}^{d_C^{[l]} \times d_H^{[l]} \times d_W^{[l]}}$ ：第 l 层（卷积层）运算输出
- $\mathbf{P}^{[l+1]} \in \mathbb{R}^{d_C^{[l]} \times d_H^{[l+1]} \times d_W^{[l+1]}}$ ：第 $(l + 1)$ 层（池化层）运算输出

结构



1. 假设 $d\mathbf{P}^{[l+1]}$ 是已知的

- 展平运算的后向传播是简单的

平均池化的后向传播

$A^{[l]}$

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 M

0.25	0.25
0.25	0.25

$P^{[l+1]}$

4.25	6.0	6.25
3.0	6.0	2.25
3.25	3.75	3.0

$$P_{11}^{[l+1]} = 0.25 \times (A_{11}^{[l]} + A_{12}^{[l]} + A_{21}^{[l]} + A_{22}^{[l]})$$

$$dA_{11}^{[l]} = \frac{\partial \mathcal{J}}{\partial A_{11}^{[l]}} = \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial \mathcal{J}}{\partial P_{ij}^{[l+1]}} \times \frac{\partial P_{ij}^{[l+1]}}{\partial A_{11}^{[l]}} = \frac{\partial \mathcal{J}}{\partial P_{11}^{[l+1]}} \times \frac{\partial P_{11}^{[l+1]}}{\partial A_{11}^{[l]}} = dP_{11}^{[l+1]} \times 0.25$$

$$dA_{12}^{[l]} = dA_{21}^{[l]} = dA_{22}^{[l]} = 0.25 \times dP_{11}^{[l+1]}$$

平均池化的后向传播

1. 对于平均池化，我们有

$$d\mathbf{A}^{[l]} = d\mathbf{P}^{[l+1]} \otimes \mathbf{M}$$

- $\mathbf{A} \otimes \mathbf{B}$: 两个矩阵 $\mathbf{A}$ 以及 $\mathbf{B}$ 的 Kronecker 乘积
- 只对下面这个特殊情况成立:
 - ▷ 池化核的尺寸“等于”步幅
 - ▷ 图片的尺寸可以被池化核的尺寸“整除”

2. 我们可以用链式法则，得到更一般的结果（请自行验证）

最大池化的后向传播

$A^{[l]}$

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

$P^{[l+1]}$

7	9	9
5	9	7
9	7	7

$$P_{11}^{[l+1]} = \max\{A_{11}^{[l]}, A_{12}^{[l]}, A_{21}^{[l]}, A_{22}^{[l]}\} = A_{22}^{[l]}$$

$$dA_{11}^{[l]} = \frac{\partial \mathcal{J}}{\partial A_{11}^{[l]}} = \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial \mathcal{J}}{\partial P_{ij}^{[l+1]}} \times \frac{\partial P_{ij}^{[l+1]}}{\partial A_{11}^{[l]}} = \frac{\partial \mathcal{J}}{\partial P_{11}^{[l+1]}} \times \frac{\partial P_{11}^{[l+1]}}{\partial A_{11}^{[l]}} = dP_{11}^{[l+1]} \times 0 = 0$$

$$dA_{12}^{[l]} = dA_{21}^{[l]} = 0; \quad dA_{22}^{[l]} = dP_{11}^{[l+1]}$$

最大池化的后向传播

1. 对于最大池化而言， $d\mathbf{A}^{[l]}$ 是通过如下过程得到的

- 将矩阵 $\mathbf{P}^{[l+1]}$ 中每个元素，乘上一个对应的 2×2 掩码矩阵，并将其赋值给 $d\mathbf{A}^{[l]}$ 的相应位置
- 只对下面这个特殊情况成立：
 - ▷ 池化核的尺寸“等于”步幅
 - ▷ 图片的尺寸可以被池化核的尺寸“整除”

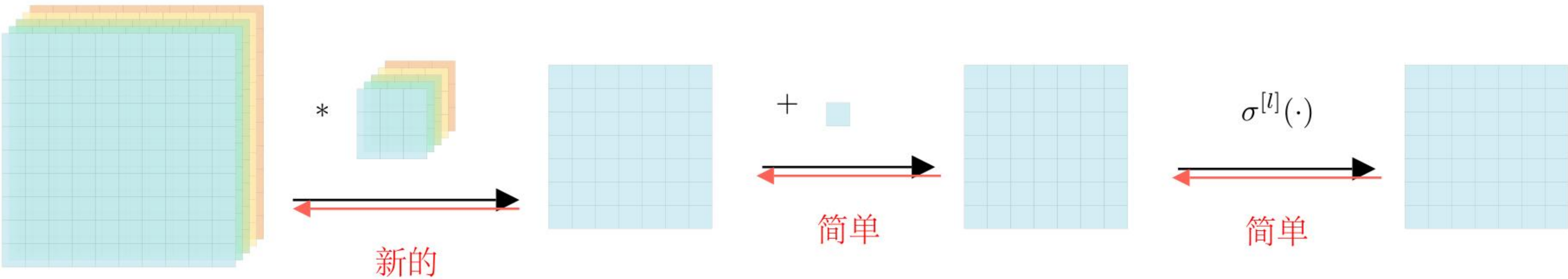
2. 我们可以用链式法则，得到更一般的结果（请自行验证）

备注

1. 我们已经介绍完了，从 $d\mathbf{P}^{[l+1]}$ 到 $d\mathbf{A}^{[l]}$ 的后向传播过程
 - 对于平均池化，在特定条件下，梯度被平均分配
 - 对于最大池化，在特定条件下，梯度按照前向最大值位置传回
 - 池化运算不涉及任何额外的模型参数
2. 在后续分析中，我们假设 $d\mathbf{A}^{[l]}$ 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 $d_C^{[l]} = 1$.



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

$$\mathbf{Z}_0^{[l]} \in \mathbb{R}^{1 \times d_H^{[l]} \times d_W^{[l]}}$$

$$\mathbf{Z}_1^{[l]} \in \mathbb{R}^{1 \times d_H^{[l]} \times d_W^{[l]}}$$

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

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

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

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

$$d\mathbf{b}^{[l]} = \sum_{i=1}^{d_H^{[l]}} \sum_{j=1}^{d_W^{[l]}} dZ_{1,ij}^{[l]}$$

$$d\mathbf{X}^{[l-1]} = ?$$

$$d\mathbf{Z}_0^{[l]} = d\mathbf{Z}_1^{[l]}$$

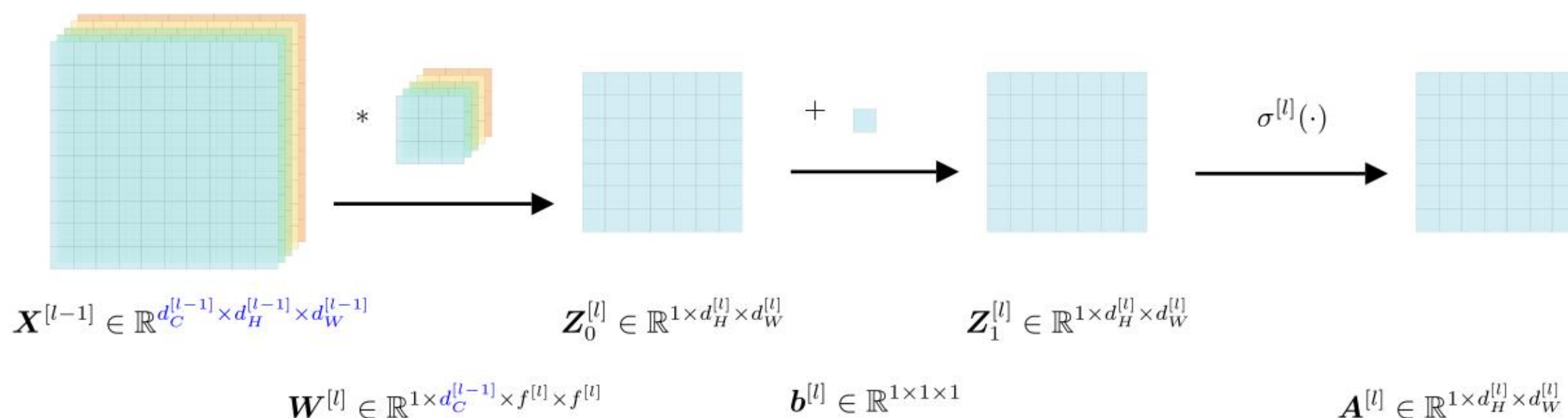
$$d\mathbf{Z}_1^{[l]} = \sigma^{[l]'}(\mathbf{Z}^{[l]}) \circ d\mathbf{A}^{[l]}$$

卷积运算的后向传播

1. 为了求函数对某个自变量的导数，我们需要知道该自变量参与了哪些计算
2. 初始化 $d\mathbf{X}^{[l-1]}$ 为一个所有元素均为 0 的三维张量，其规模与 $\mathbf{X}^{[l-1]}$ 的规模相同
3. 符号不那么严格地，对于 $i = 1, \dots, d_H^{[l]}$ 和 $j = 1, \dots, d_W^{[l]}$ ，我们可以得到

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

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

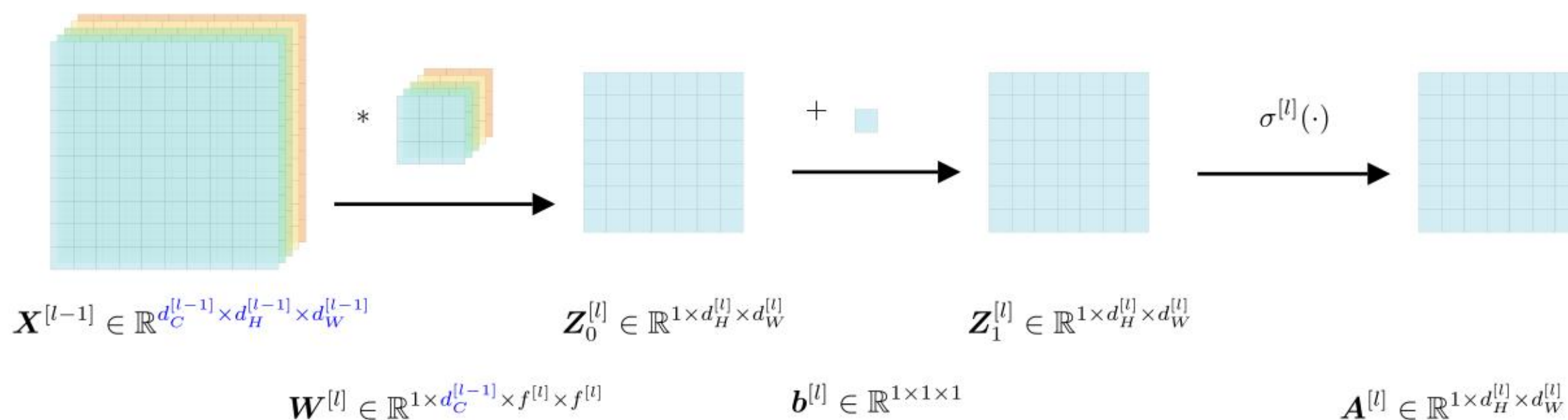


卷积运算的后向传播

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

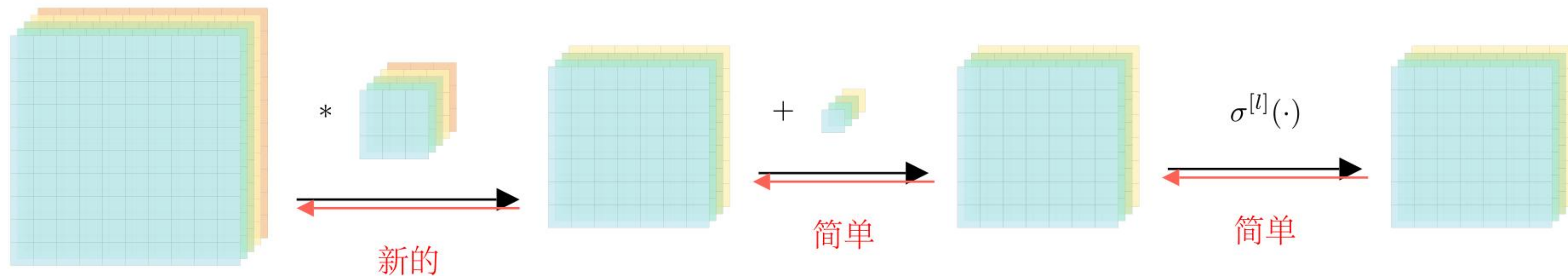
$$d\mathbf{W}^{[l]} = \sum_{i=1}^{d_H^{[l]}} \sum_{j=1}^{d_W^{[l]}} d\mathbf{Z}_{0,ij}^{[l]} \times \mathbf{B}_{ij}^{[l-1]}$$

- $\mathbf{B}_{ij}^{[l-1]}$: $\mathbf{X}^{[l-1]}$ 中被用于得到 $\mathbf{Z}_{0,ij}^{[l]}$ 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



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

$$d\mathbf{b}_c^{[l]} = \sum_{i=1}^{d_H^{[l]}} \sum_{j=1}^{d_W^{[l]}} dZ_{1,cij}^{[l]} \quad (c = 1, \dots, d_C^{[l]})$$

$$d\mathbf{X}^{[l-1]} = ?$$

$$d\mathbf{Z}_0^{[l]} = d\mathbf{Z}_1^{[l]}$$

$$d\mathbf{Z}_1^{[l]} = \sigma^{[l]'}(\mathbf{Z}^{[l]}) \circ d\mathbf{A}^{[l]}$$

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. 输入图片 $\mathbf{X}^{[l-1]}$ 的信息，均被用于计算 $\mathbf{Z}_0^{[l]}$ 每个通道的元素
3. 将 $d\mathbf{X}^{[l-1]}$ 初始化为一个所有元素均为 0 的三维张量，其规模与 $\mathbf{X}^{[l-1]}$ 相同
4. 对于 $i = 1, \dots, d_H^{[l]}$ 和 $j = 1, \dots, d_W^{[l]}$ ，我们有

$$d\mathbf{X}_{\text{slice}, ij}^{[l-1]} + = \sum_{c=1}^{d_C^{[l]}} d\mathbf{Z}_{0, \textcolor{red}{c}ij}^{[l]} \times \mathbf{W}_{\textcolor{red}{c}}^{[l]}$$

- $\mathbf{X}_{\text{slice}, ij}^{[l-1]}$: $\mathbf{X}^{[l-1]}$ 中，被用于得到 $\mathbf{Z}_{0, \textcolor{red}{c}ij}^{[l]}$ 的部分，其中， $c = 1, \dots, d_C^{[l]}$
- $\mathbf{W}_{\textcolor{red}{c}}^{[l]}$: 第 l 层中，第 c 个卷积核

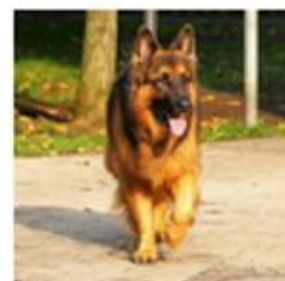
卷积运算的后向传播

1. 很明显, $\mathbf{Z}_{0,c}^{[l]}$ 中的每个元素, 均包含 $\mathbf{W}_c^{[l]}$ 的信息, 其中, $c = 1, \dots, d_C^{[l]}$
2. 但 $\mathbf{W}_c^{[l]}$ 并不用于计算其他通道的元素
3. 因此, 我们有

$$d\mathbf{W}_{\textcolor{red}{c}}^{[l]} = \sum_{i=1}^{d_H^{[l]}} \sum_{j=1}^{d_W^{[l]}} d\mathbf{Z}_{0,\textcolor{red}{c}ij}^{[l]} \times \mathbf{B}_{ij}^{[l-1]} \quad (c = 1, \dots, d_C^{[l]})$$

- $\mathbf{B}_{ij}^{[l-1]}$: $\mathbf{X}^{[l-1]}$ 中, 被用于计算 $\mathbf{Z}_{0,\textcolor{red}{c}ij}^{[l]}$ 的部分

数据增强 (Data augmentation)



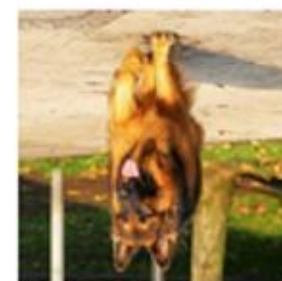
Original



Horizontal Flip



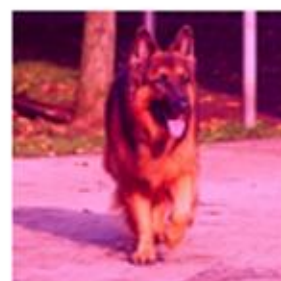
Vertical Flip



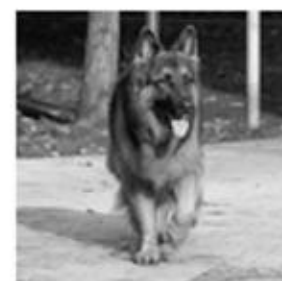
Horizontal + Vertical



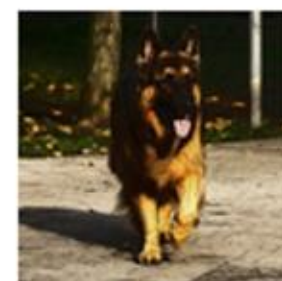
Color Profile 1



Color Profile 2



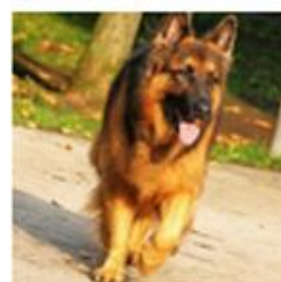
Color Profile 3



Color Profile 4



Rotate Left



Rotate Right



Noise 1



Noise 2



Crop 1



Crop 2

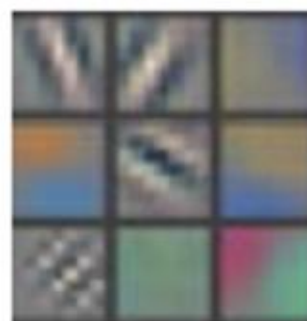


Resize 1

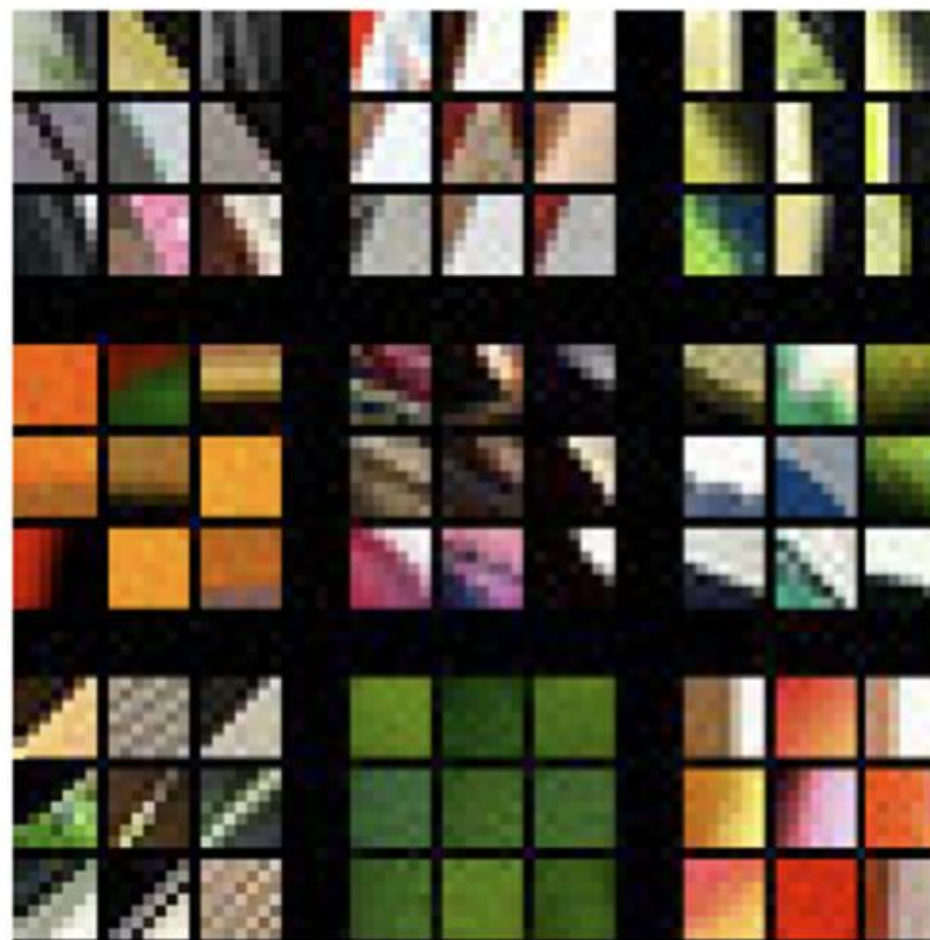


Resize 2

理解卷积神经网络

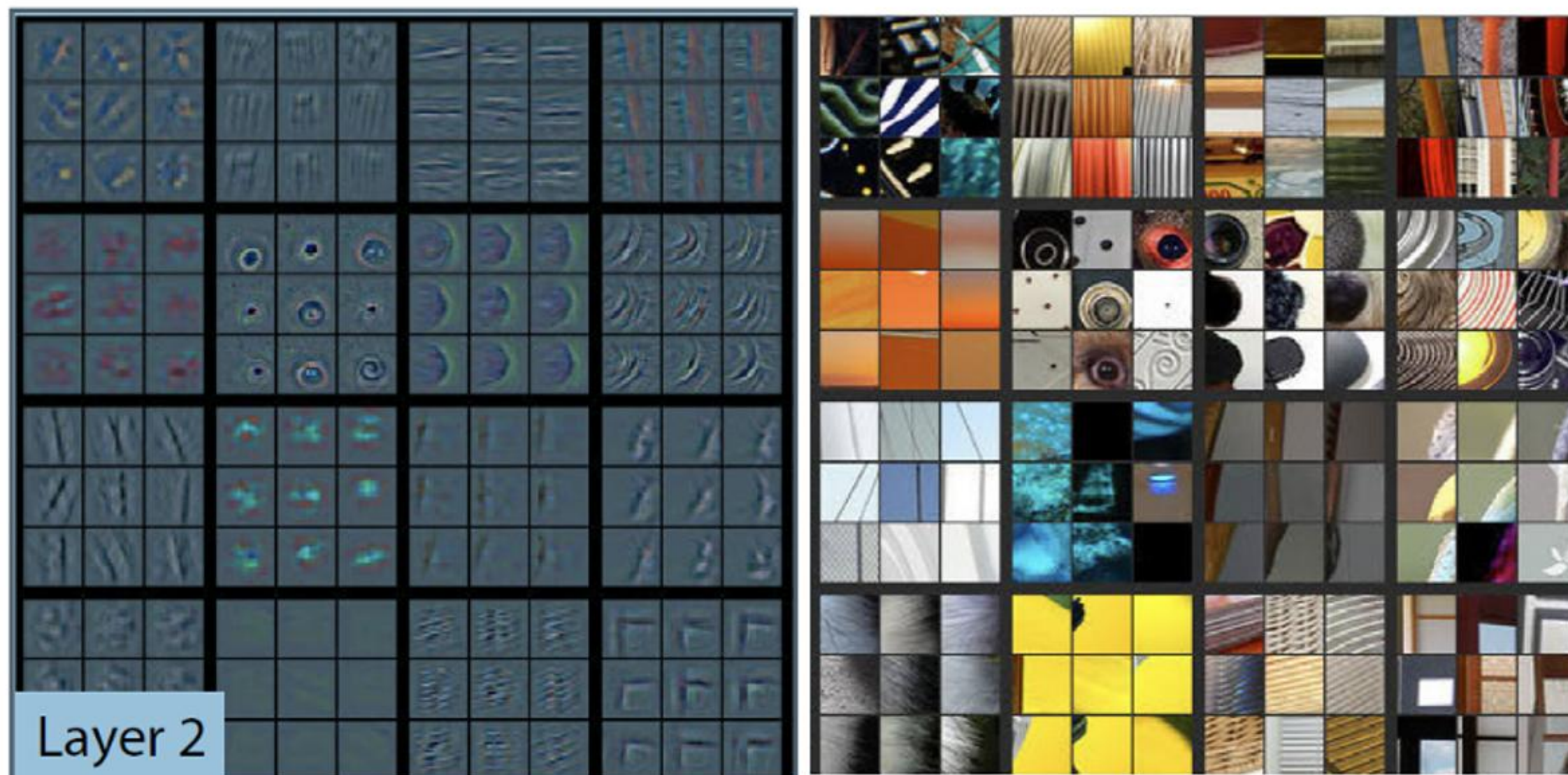


Layer 1



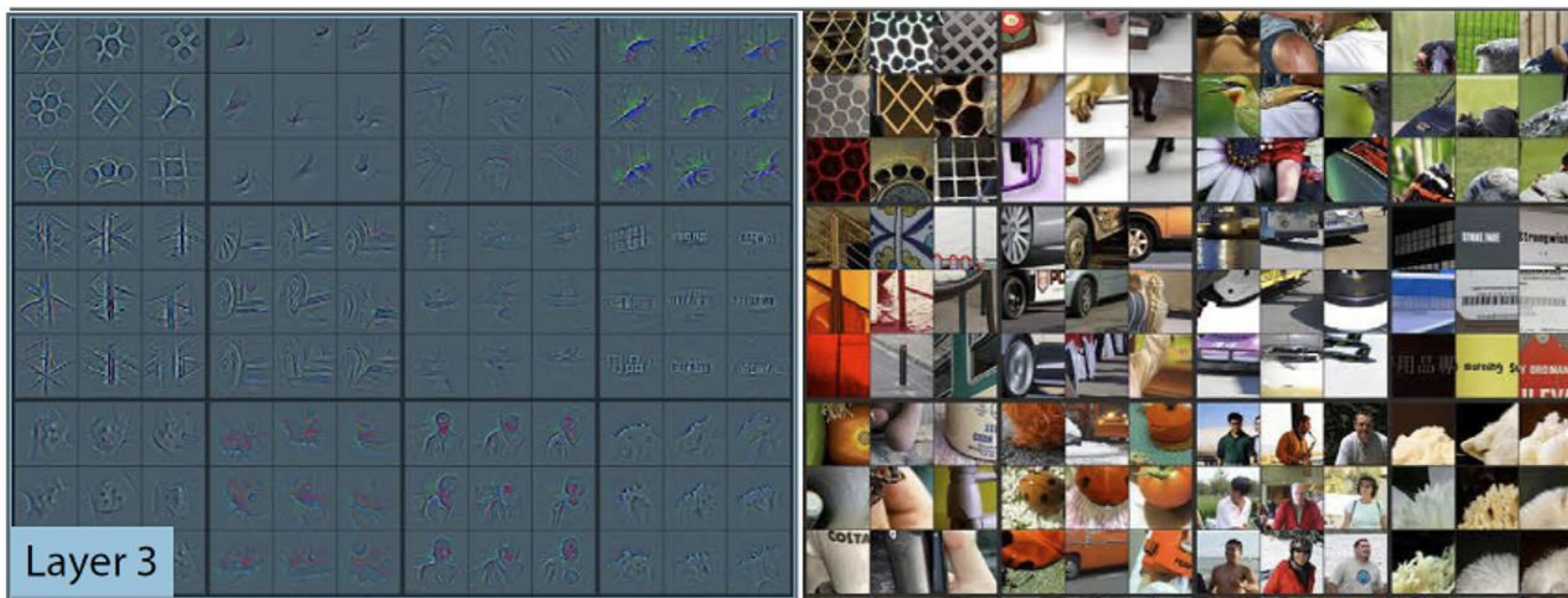
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



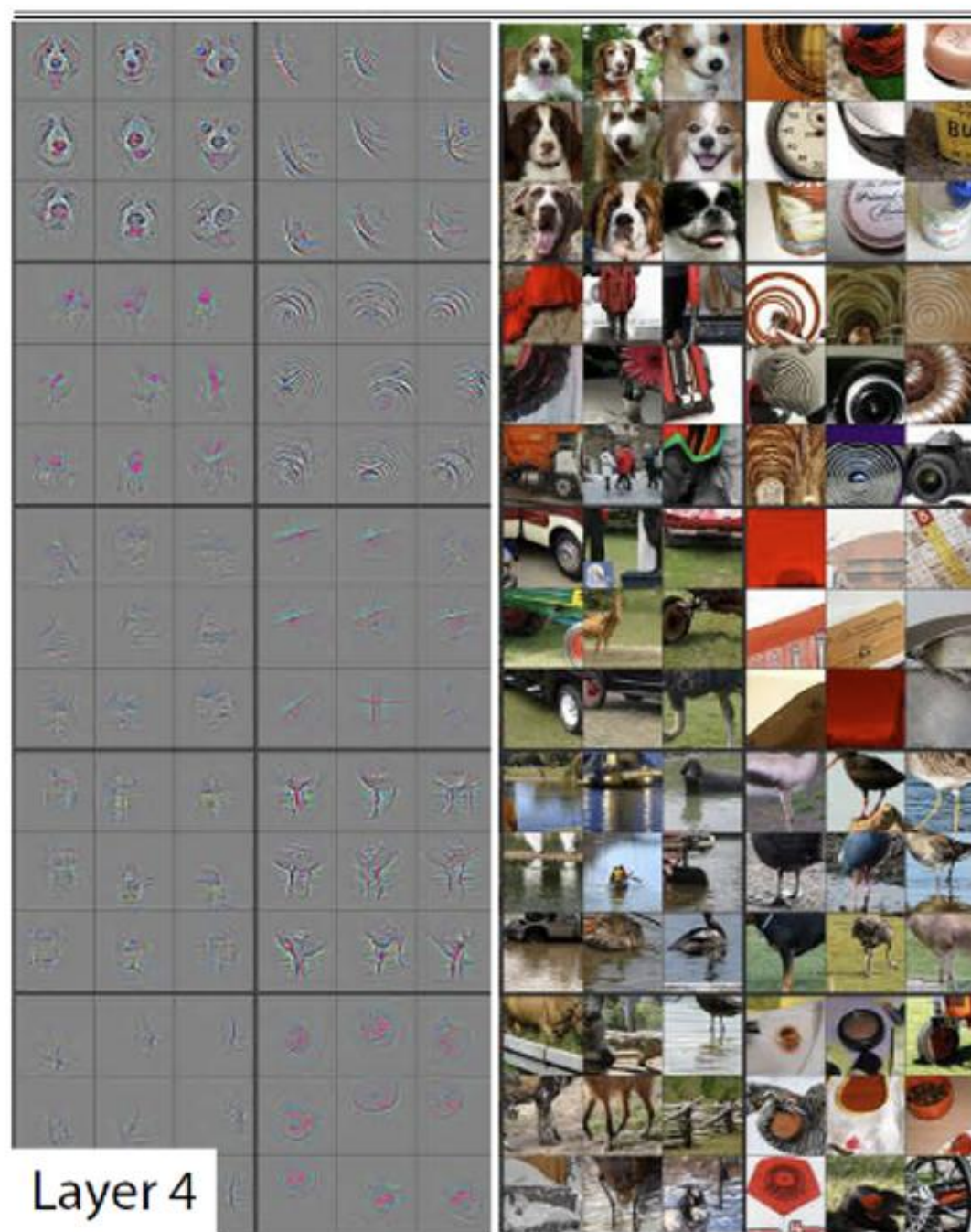
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

理解卷积神经网络



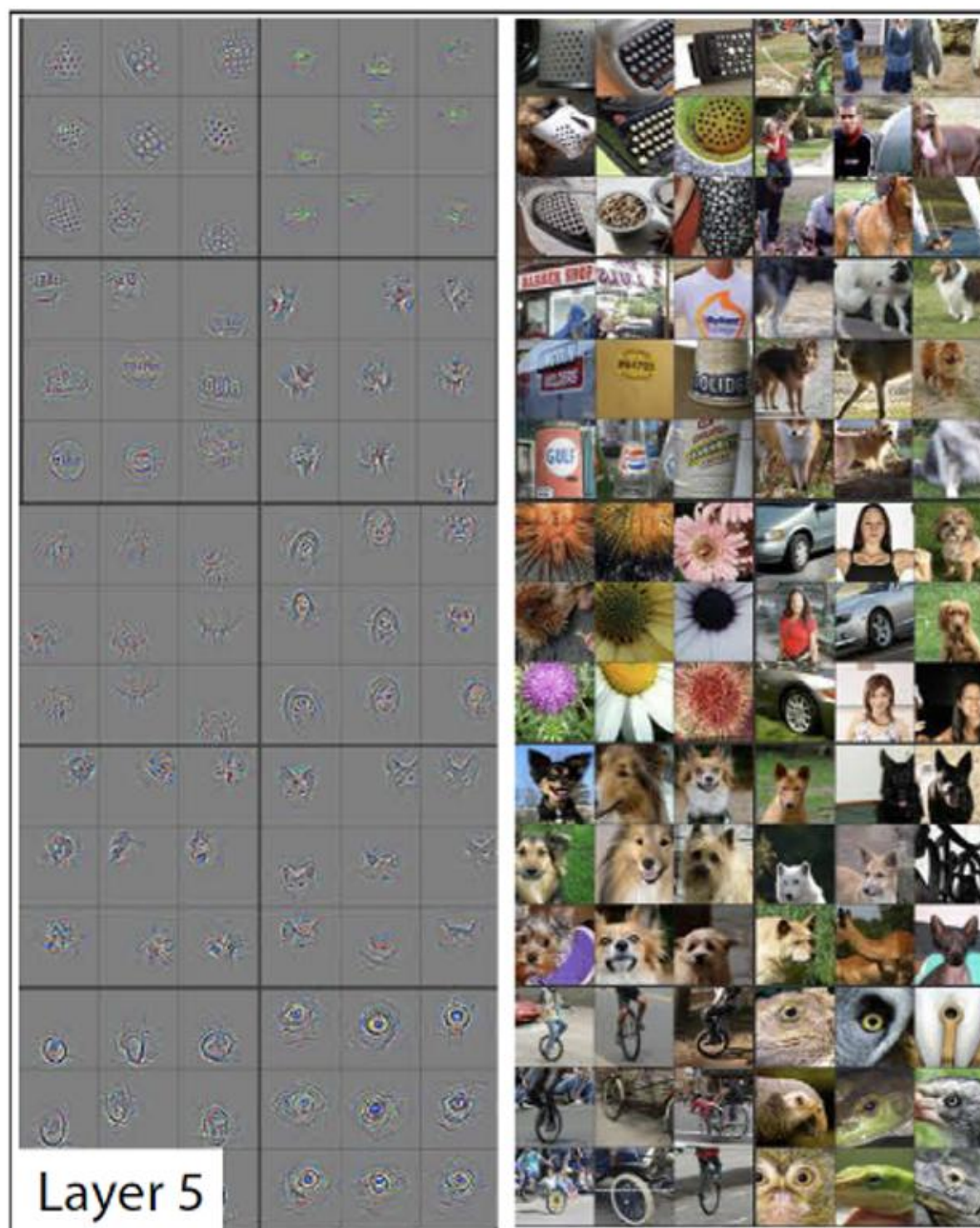
[Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham]

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]

课程总结

1. 动机：图像的空间结构、局部性和参数共享需求
2. 卷积层：提取局部特征
3. 池化层：聚合局部响应并进行下采样
4. 浅层、中层、深层特征表示通常呈现出由简单到复杂的层级结构