



人工智能与深度学习

4-3 卷积神经网络 II

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. LeNet-5

2. AlexNet

3. VGG-16

LeNet-5

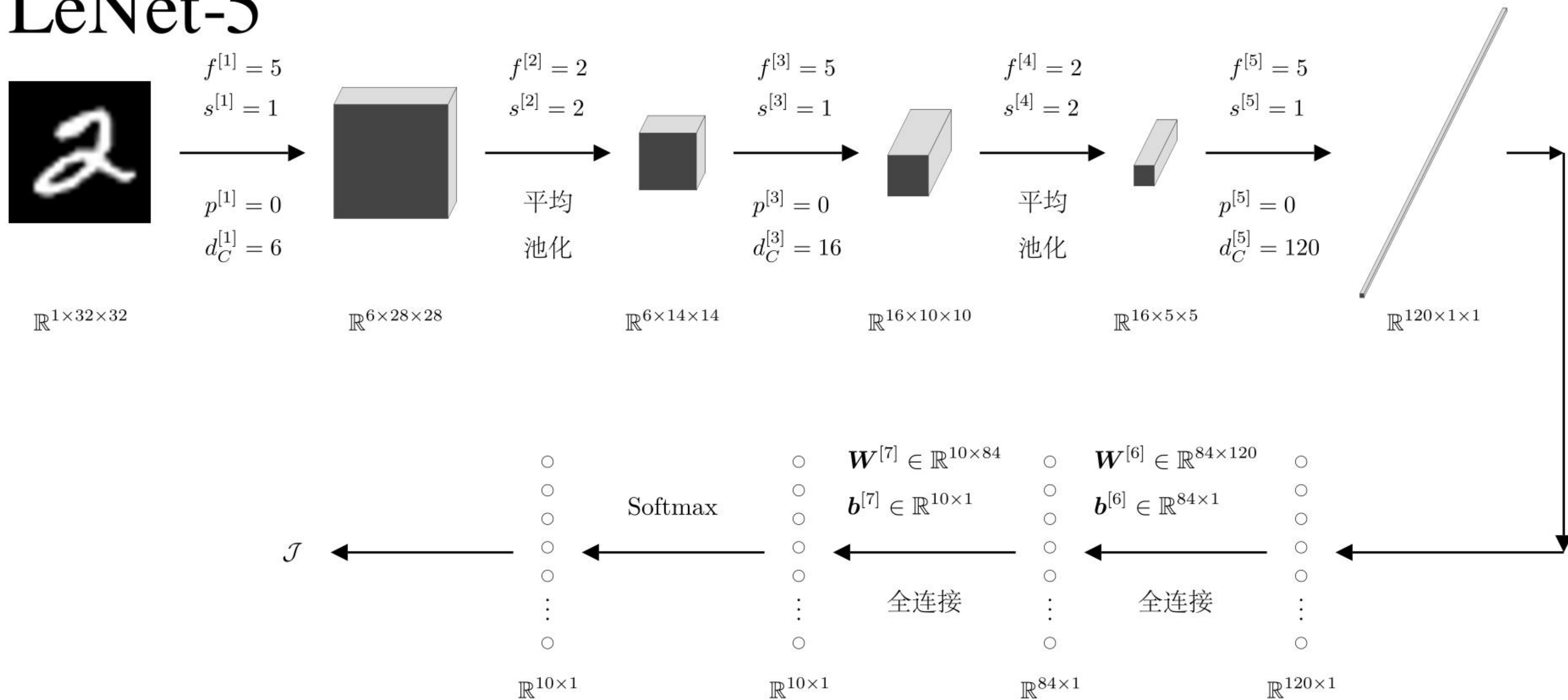
1. LeNet-5 于 1998 年被 LeCun 等提出

- 最早的几个卷积神经网络模型之一
- 在计算机视觉领域，具有里程碑意义
- 被用于识别 32×32 的灰度手写数字图像（10 类）
- 名称中的“5”表示模型有 5（或者 7）层
- $\tanh$ 是隐藏层的激活函数

2. 一定程度上，摆脱了高度依赖人工设计特征的传统路线

3. 将采用现代平均池化和 Softmax 的简化为大家展示 LeNet-5 的实现

LeNet-5



备注

1. 该模型具有早期神经网络的设计风格：

- 用的是带可学习缩放和偏置的平均型下采样，而不是现代模型中更常见的最大池化
- 将 $\tanh$ 作为隐藏层的激活函数

2. 核心价值：

- 清晰展示了卷积神经网络的核心 workflow
- 证明卷积结构可以在真实视觉任务中有效工作

AlexNet

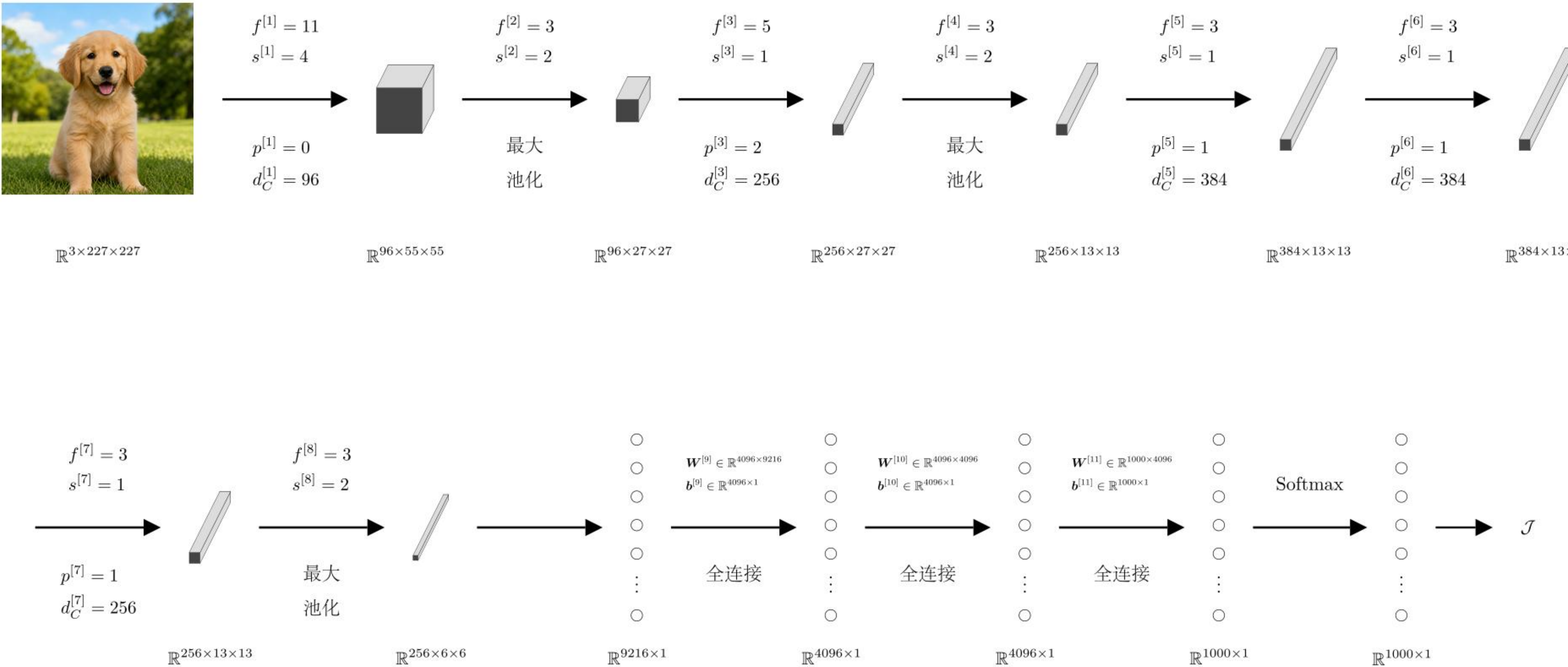
1. 在计算机视觉领域，具有里程碑意义的模型

- 在 ILSVRC 2012 图像分类竞赛中，获得第一名
- 模型性能显著超越其他方法
- 极大促进了深度学习在计算机视觉领域的发展
- 用于识别分辨率为 $224 \times 224 \times 3$ 彩色图片的类别（1000 类分类问题）

2. 贡献：

- 将 ReLU 作为激活函数
- 使用 dropout 和数据增强缓解过拟合，并改善泛化性能

AlexNet



备注

1. 相比于 LeNet-5, AlexNet 具有如下特征

- 输入规模更大：从小型灰度图到大型彩色图片
- 网络规模更大，模型容量更高
- 各卷积层通常具有更多通道
- 池化和中间层组织方式，更适合复杂自然图像

2. LeNet-5：卷积神经网络这一路线的可行性

3. AlexNet：这条路在大规模现代视觉任务上，可以成为主流

VGG-16

1. 计算机视觉领域的经典模型

- ILSVRC 2014 的定位任务中获得第一名，在分类任务中获得第二名
- 用于识别分辨率为 $224 \times 224 \times 3$ 彩色图片的类别（1000 类分类问题）

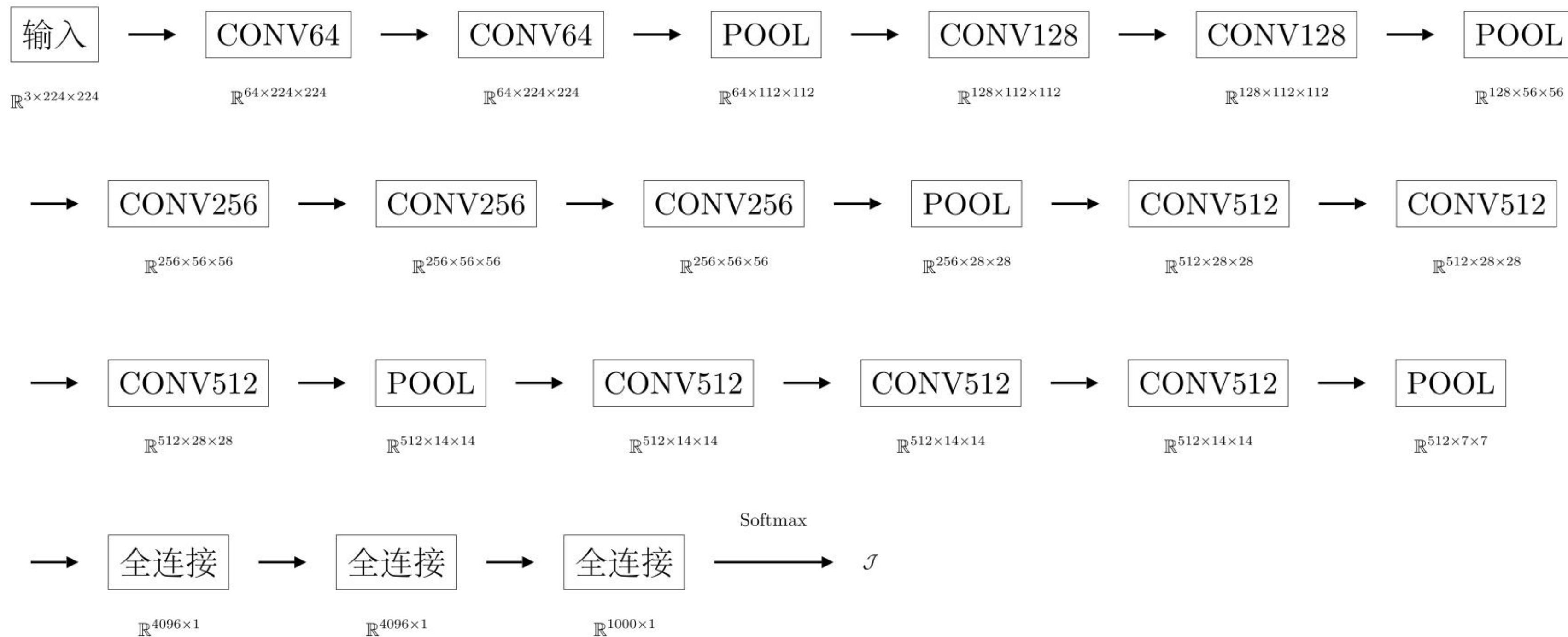
2. 符号

- CONV[num]: 使用 num 个 3×3 卷积核，步幅为 1、填充为 1
- POOL: 进行最大池化，对应的（池化）核尺寸为 2×2 、步幅为 2

3. 注意

- 卷积运算：不改变“输入图片”的尺寸
- 最大池化：将“输入图片”的尺寸缩小一半

VGG-16



备注

1. VGG-16 的构架非常统一

- 卷积层不改变空间分辨率：卷积核尺寸为 3×3 ，填充为 1
- 池化层使用 2×2 池化核和步幅 2，将空间分辨率减半
- 分工明确：卷积层负责特征提取，池化层负责空间缩减

2. LeNet-5：卷积神经网络这一路线的可行性

3. AlexNet：这条路在大规模现代视觉任务上，可以成为主流

4. VGG-16：确立了模型深度和统一设计规则的重要性

课程总结

1. LeNet-5:

- 奠定了卷积神经网络的基本架构
- 层次化特征堆叠的思想

2. AlexNet:

- 打破了深度卷积神经网络的“实用性瓶颈”
- 借使用 ReLU 加快深层网络的优化，并使用 Dropout 和数据增强缓解过拟合

3. VGG-16:

- 在其架构体系中，增加网络深度并统一使用小卷积核能够显著提升识别性能
- 提出统一的设计框架：小卷积核 + 最大池化