

这份复习笔记基于西安电子科技大学《现代机器学习》课程课件，涵盖了卷积神经网络（CNN）的核心原理、数学表达、操作细节及经典应用。

---

## 第一部分：卷积神经网络基础 (Part #1)

### 1. 为什么图像需要专门的架构? (P6-P8)

- **高维性问题**: 典型的 $224 \times 224$  RGB图像有约15万个输入维度。全连接网络（FC）的权重数量会呈平方级增长（如 shallow network 可能需要220亿个参数），导致计算量、内存和数据量不可接受。
- **空间相关性**: 图像中相邻像素在统计上是相关的。FC网络不考虑“邻近”概念，打乱像素顺序对FC网络无差别，但对图像理解是致命的。
- **几何稳定性**: 图像中的物体即使平移几个像素，其本质不变。FC网络必须在每个位置重新学习相同的模式，效率极低。
- **CNN的优势**:
  - **局部处理**: 独立处理局部区域。
  - **参数共享**: 整个图像共用同一组权重（卷积核）。
  - **捕捉空间关系**: 利用邻近像素的关联。

### 2. 不变性 (Invariance) 与 等变性 (Equivariance) (P10-P13)

- **不变性 (Invariance)**: 函数  $f[t[x]] = f[x]$ 。即输入经过变换（平移、旋转、翻转），输出保持不变。
  - **典型应用**: **图像分类**（无论猫在左上角还是右下角，结果都是“猫”）。
- **等变性 (Equivariance)**: 函数  $f[t[x]] = t[f[x]]$ 。即输入的变换会引起输出产生同样的变换。
  - **典型应用**: **图像分割**（物体平移了，分割出的掩码也应随之平移）。

### 3. 1D 卷积数学表达 (P15, P22)

- **卷积核/滤波器 (Kernel/Filter)**:  $\omega = [\omega_1, \omega_2, \omega_3]^T$ 。
- **计算公式**:  $z_i = \omega_1 x_{i-1} + \omega_2 x_i + \omega_3 x_{i+1}$ 。
- **卷积层组成**: 卷积操作 + 偏置  $\beta$  + 激活函数  $a[\cdot]$ 。
  - $h_i = a[\beta + \sum_{j=1}^3 \omega_j x_{i+j-2}]$ 。

## 4. 卷积关键超参数 (P17-P20)

- **填充 (Padding):**
  - **Zero Padding:** 在边界补0, 使输出尺寸与输入一致。
  - **Valid Convolution:** 只在卷积核完全覆盖图像的区域计算, 输出会变小。
- **步幅 (Stride):** 卷积核每次移动的位移  $k$ 。步幅越大, 输出尺寸越小。
- **扩张/空洞卷积 (Dilation):** 在卷积核元素间插入0。
  - **作用:** 不增加参数量的前提下增大**感受野 (Receptive Field)**。

## 5. 通道 (Channels) 与 参数计算 (P29-P33)

- **通道的必要性:** 单个卷积核会丢失信息, 因此并行使用多个卷积核, 产生多个**特征图 (Feature Map)**。
- **参数量计算 (1D) :**
  - 输入通道  $C_i$ , 输出通道  $C_o$ , 卷积核大小  $K$ 。
  - **权重数量:**  $C_i \times C_o \times K$ 。
  - **偏置数量:**  $C_o$ 。

## 6. 感受野 (Receptive Fields) (P35-P38)

- **定义:** 隐藏单元对应的原始输入区域大小。
- **演变:** 感受野随层数加深而线性增加。例如: 连续两层  $K = 3, Stride = 1$  的卷积, 第二层的感受野为 5。

---

## 第二部分: 2D 卷积与高级应用 (Part #2)

### 1. 2D 卷积 (P48-P53)

- **2D计算:** 在  $K \times K$  的方形区域内进行加权求和。
- **RGB 图像处理:** 卷积核也是 3 维的 ( $K \times K \times 3$ ), 分别作用于红、绿、蓝三个通道并求和, 加上偏置后得到输出的一个点。
- **2D参数量:** 权重  $\Omega \in \mathbb{R}^{C_i \times C_o \times K \times K}$ 。

## 2. 下采样与上采样 (P56-P60)

- **下采样 (Downsampling):** 减小特征图尺寸, 增加感受野。
  - **方式:** 步幅为2的卷积、**最大池化 (Max Pooling)** (提取最强特征, 具平移鲁棒性)、**平均池化 (Mean Pooling)**。
- **上采样 (Upsampling):** 恢复图像分辨率。
  - **方式:**
    1. **复制 (Duplicate):** 直接放大。
    2. **Max-upsampling:** 记录池化位置并还原。
    3. **双线性插值 (Bilinear interpolation):** 数学计算填充缺失值。
    4. **转置卷积 (Transposed Convolution):** 可学习的上采样, 矩阵操作是卷积的转置。

## 3. 1x1 卷积 (P61-P62)

- **本质:** 在每个像素位置运行一个全连接网络。
- **核心作用:** 跨通道整合信息、**改变通道数** (升维或降维), 而不改变图像的空间尺寸。

## 4. 经典架构: 图像分类 (P64-P70)

- **ImageNet 数据库:** 120万训练图, 1000个类别。
- **AlexNet (2012):** 深度学习转折点。8层 (5个卷积+3个全连接), ReLU激活, Dropout, 数据增强。
- **VGG (2015):** 强调**深度** (19层), 使用更小的卷积核串联, 性能优于 AlexNet。
- **趋势:** 随着时间推移, ImageNet 的 Top-1 错误率持续下降。

## 5. 目标检测 (Object Detection) - YOLO (P72-P76)

- **任务:** 识别物体并定位 (Bounding Box) 。
- **YOLO (You Only Look Once):**
  - 将图像划分为  $7 \times 7$  网格。
  - 每个网格预测边界框 (x, y, w, h, confidence) 和类别。
  - **迁移学习 (Transfer Learning):** 先在 ImageNet 上预训练分类, 再微调 (Fine-tune) 用于检测。

## 6. 语义分割 (Semantic Segmentation) (P78-P82)

- **任务：**逐像素分类。
  - **架构：**编码器-解码器 (Encoder-Decoder)，也叫沙漏网络 (Hourglass)。
    - **编码器：**类似 VGG，通过下采样提取全局语义信息。
    - **解码器：**通过转置卷积和上采样恢复空间分辨率。
    - **输出：**最终通过  $1 \times 1$  卷积产生对应类别的概率图。
- 

## 第三部分：总结与考点 (Summary)

- **卷积层核心：**权重和偏置在空间位置上共享  $\rightarrow$  参数量极少  $\rightarrow$  不随输入图像增大而爆炸。
  - **CNN 的演进规律：**随着网络深入，**空间维度减少，通道数增加**（信息从局部细节转化为高层语义）。
  - **归纳偏置 (Inductive Bias)：**卷积的平移等变性是一种强大的先验知识，使得 CNN 在图像任务上远优于全连接网络。
  - **深度限制：**网络不能无限加深，否则会难以训练（梯度问题），这引出了**残差连接 (Residual Connections)** 的需求。
- 

**复习建议：**重点掌握 (1) 参数量计算公式，(2) 卷积/池化对特征图尺寸的影响，(3) 不变性与等变性的区别，以及 (4)  $1 \times 1$  卷积的作用。这些是面试和考试中最常见的知识点。