**《深度学习》读书笔记**

版本：MIT Press 2016（中文第 1 版）

21计科1赵晓英 202111680393

重点：与 AIGC 词汇学习课题相关的深度学习基础、模型与研究方法

目录速览

|  |  |  |
| --- | --- | --- |
| Part | 章节 | 核心主题 |
| I. Applied Math & ML Basics | Ch 1–6 | 线性代数、概率论、数值优化、机器学习基础 |
| II. Modern Deep Networks | Ch 7–15 | 前馈网络、正则化、卷积网络、序列模型、实践技巧 |
| III. Deep Learning Research | Ch 16–20 | 线性因子模型、自编码器、表示学习、结构化概率模型、蒙特卡洛方法 |

第一部分：基础

Ch 1 引言

DL 三大驱动：数据规模、计算能力、算法创新。

核心优势：自动学习多层次特征表示；端到端优化。

对毕设启示：图像生成与 CLIP 评分依赖深层神经网络特征 —— 深度学习正是底层方法论。

Ch 2 线性代数

向量/矩阵记号：$\mathbf{x}$, $\mathbf{W}$；广播与张量乘法在 PyTorch/OpenAI 接口中的映射。

本书惯例：列向量默认 shape $(n,1)$，与 NumPy 行向量 $(1,n)$ 区分。

实践：实现 CLIP 余弦相似时，需先 F.normalize → 行向量归一化，避免数值误差。

Ch 3 概率统计

联合分布、边缘分布、条件独立 提供建模语言。

应用：Prompt→Image 的生成概率 $p(\text{image}|\text{prompt})$ 被隐式建模在扩散网络中。

Ch 4 数值计算

梯度消失/爆炸：文中斜体符号 $\nabla\_θ J(θ)$；现代对策有 Residual、LayerNorm、ReLU。

FP16 / bfloat16：与论文实验中 GPU 推理有关（Sora/SDXL 部分支持混合精度）。

Ch 5 机器学习基础

泛化误差 $\mathcal{E}\_{gen}$ 与 训练误差 区别。

偏差-方差分解：解释何以需要正则化（Ch 7）。

Ch 6 深度前馈网络数学基础

多层感知机 (MLP) = 线性映射 + 非线性激活。

链式法则（反向传播）

第二部分：现代深度网络

Ch 7 正则化技巧

|  |  |  |
| --- | --- | --- |
| 方法 | 公式/伪代码 | 毕设相关性 |
| $L^2$ weight decay | $\Omega=\lambda|W|\_2^2$ | 提高 Sora 生成网络泛化 |
| Dropout | 随机掩码 $m \sim Bernoulli(p)$ | CLIP 训练中标准做法 |
| BatchNorm | $\hat{x}=\frac{x-μ}{σ}$ | SDXL/UNet 中常见 |

Ch 8 优化算法

SGD → Adam：实验脚本可采用 torch.optim.AdamW 调整超参。

学习率退火：Cosine Annealing；Sora、SDXL 均使用。

Ch 9 卷积网络

局部感受野 与 权共享 两大特征。

CLIP-Visual Backbone 正是改进的 ResNet / ViT，与本章紧密关联。

Ch 10 序列建模：循环与注意力

RNN → LSTM → Transformer。

GPT‑4o 使用多层 Transformer，Prompt 生成部分可视作序列到序列。

Ch 11–15 实践、面向任务的网络

迁移学习：预训练 + 微调。论文实验直接调用 OpenAI/CLIP 预训练权重，属冻结特征萃取。

深度生成模型（Ch 14）：扩散模型概念，为 Sora/SDXL 的理论基础。

第三部分：研究前沿

Ch 16 线性因子模型 & PCA

为表征学习提供解释性视角；对 Prompt Embedding 分析可借鉴主成分可视化。

Ch 17 自编码器

Denoising / Variational Autoencoder；扩散模型可视为无限深的 VAE。

Ch 18 表示学习

深度学习最大贡献：分层抽象表示。CLIP 通过对比学习同时对齐文本和图像嵌入。

Ch 19–20 近似推断 & 蒙特卡洛

MCMC、重要性采样等方法在现代扩散采样器（DDPM→DDIM→DPM‑Solver）中变体出现。

附录：关键公式 & 代码片段

|  |  |  |
| --- | --- | --- |
| 概念 | 关键公式/伪代码 | 代码示例 |
| 余弦相似度 | $\cos(\theta)=\frac{a\cdot b}{|a|,|b|}$ | F.cosine\_similarity(a,b) |
| 交叉熵损失 | $\mathcal{L}= -\sum y \log(\hat{y})$ | nn.CrossEntropyLoss() |
| Adam 更新 | $m\_t=β\_1 m\_{t-1}+(1-β\_1)g\_t$ | optim.Adam |

# 计算 CLIP 归一化得分（毕设实验对应）

import torch, open\_clip, PIL.Image as Image

model, \_, preprocess = open\_clip.create\_model\_and\_transforms("ViT-B-32", pretrained="openai")

text = preprocess(Image.open("prompt.png")) # 示例

# ...

个人反思与应用思路

Prompt 设计：第 10 章注意力机制表明语言上下文对生成极为敏感；结合语义场景模板能稳定输出。

正则化 vs 多样性：过强正则化会降低生成图像多样性；在毕设实验中可通过调整 cfg.model.steps 与 CLIP 分布评估。

多模态对齐：CLIP 的 contrastive 目标体现了表征学习核心思想（Ch 18）。未来可尝试将词汇的语音特征也融入多模态评分。

后续阅读：NeurIPS 2023《Consistency Models》、CVPR 2024 《LDM‑3D》—— 深度生成模型最新进展。