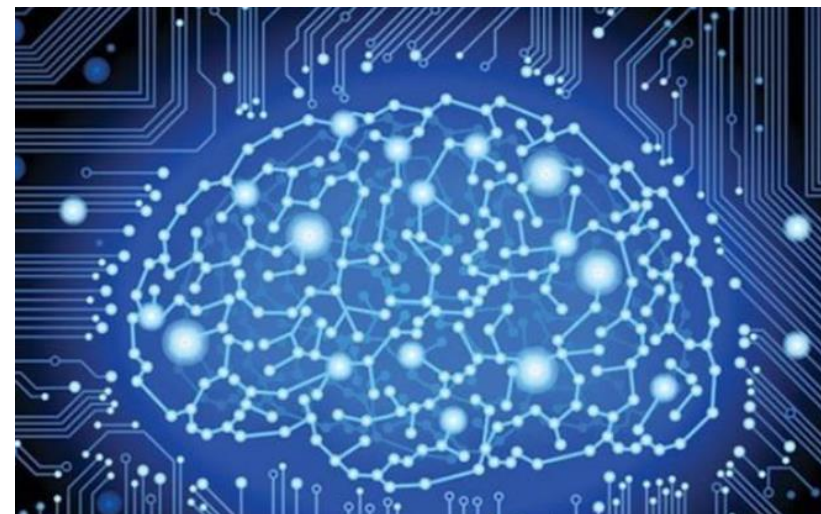


基于模糊视觉的猫狗 二分类

组员：俞一诺，刘一诺 (汇报人)

人工智能与机器学习





Q1: 计算机能否拥有像人一样的**模糊视觉**?

Q2: 机器模型与人类**视觉系统**是否相同?



浙江大学
ZHEJIANG UNIVERSITY

CONTENTS



**研究背景
&
研究意义**



问题描述
&
研究方案



算法描述
&
模型训练



求解结果
&
比较分析

机器学习

近年来，机器学习领域中的**深度学习**越发热门，它是学习样本数据**内在的规律**与层级表示，力图让机器像人一样具有分析学习能力。

模糊视觉

图像识别的数据集大多为清晰图像。在实际生活中，却不免面临**图像模糊失真**问题。这种失真却很可能导致分类器的性能显著下降。



人工神经网络

人工神经网络是机器学习的一个分支，在计算机视觉中应用广泛，包括图像识别、分类、定位、恢复等。

研究意义

理论层面：填补领域内空白
实践层面：失真图像恢复



浙江大學
ZHEJIANG UNIVERSITY

CONTENT



研究背景
&
研究意义



**问题描述
&
研究方案**



算法描述
&
模型训练



求解结果
&
比较分析

问题描述

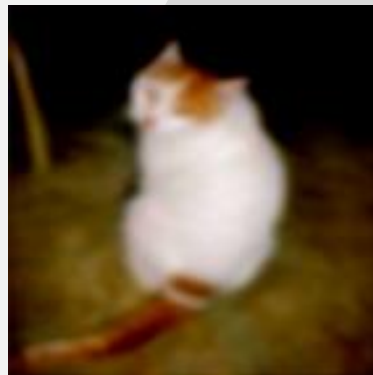
模糊视觉的分类问题

猫狗二分类数据集(Asirra Dataset)

包括清晰照片和模糊照片

不同结构的CNN

不同图片处理方式



研究重点与方案

- i) **对模糊图像的分类准确率**: 比较有/无预训练过程的模型结果
- ii) **不同CNN结构**: AlexNet和VGG-16的结果对比
- iii) **与人体视觉神经系统的异同**: 将卷积层训练到的特征和人体视觉系统的层级结构做比较分析



浙江大學
ZHEJIANG UNIVERSITY

CONTENT



研究背景
&
研究意义



问题描述
&
研究方案

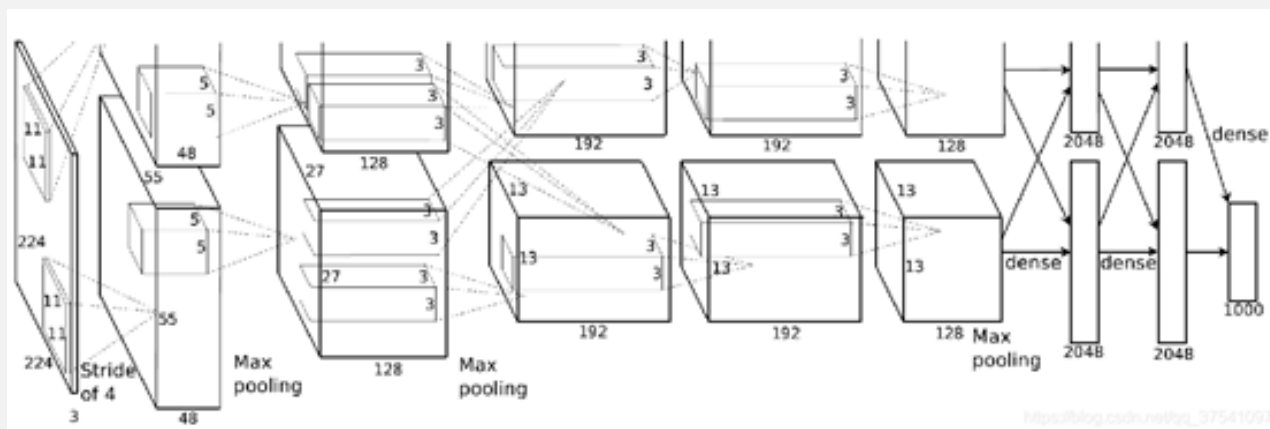


**算法描述
&
模型训练**



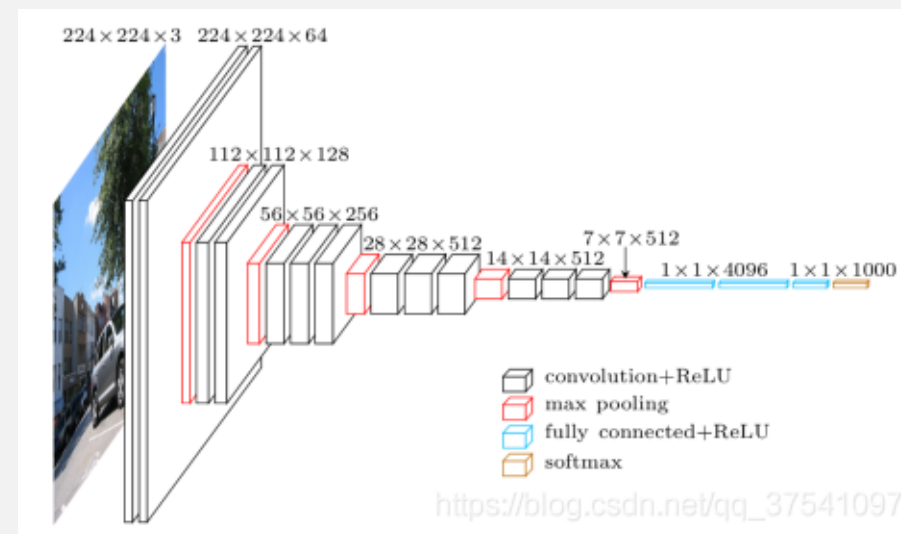
求解结果
&
比较分析

AlexNet



- 首次利用GPU 进行网络加速训练。
- 使用了 **ReLU 激活函数**，而不是传统的 Sigmoid 激活函数以及 Tanh 激活函数。
- 使用了 LRN 局部响应归一化。
- 在全连接层的前两层中使用了 **Dropout方法**按一定比例随机失活神经元，以减少过拟合。

VGG-16



该网络中的亮点：通过**堆叠多个小的卷积核**来替代大尺度卷积核 (在拥有相同感受野的前提下能够减少所需参数，增加网络深度)

Asirra Dataset

含四类数据，即清晰的猫猫、清晰的狗狗、模糊的猫猫和模糊的狗狗。其中，模糊的照片满足参数为5的高斯分布。

Cat (Noise-free)



Dog (Noise-free)



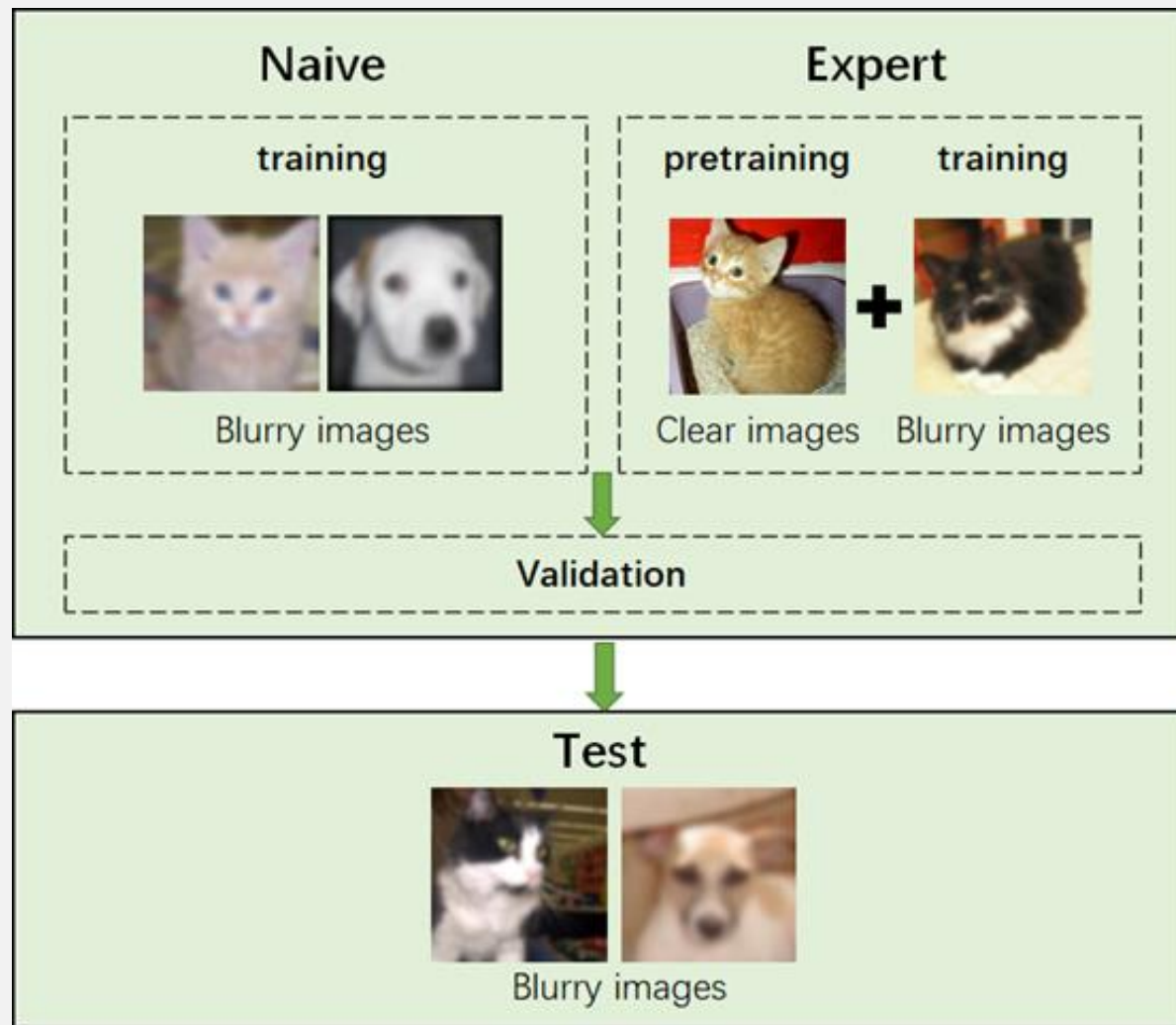
Cat (Noisy)



Dog (Noisy)



- 训练集包括**17953张**清晰图像和17953张模糊图像，其中20%用做验证集；测试集包括**2494张**清晰图像和2494张模糊图像。
- 采用**两种方法**，分别训练了两个模型。
- 分别采用AlexNet和VGG-16**两个网络结构**，最终得到四个结果。





浙江大学
ZHEJIANG UNIVERSITY

CONTENT



研究背景
&
研究意义



问题描述
&
研究方案



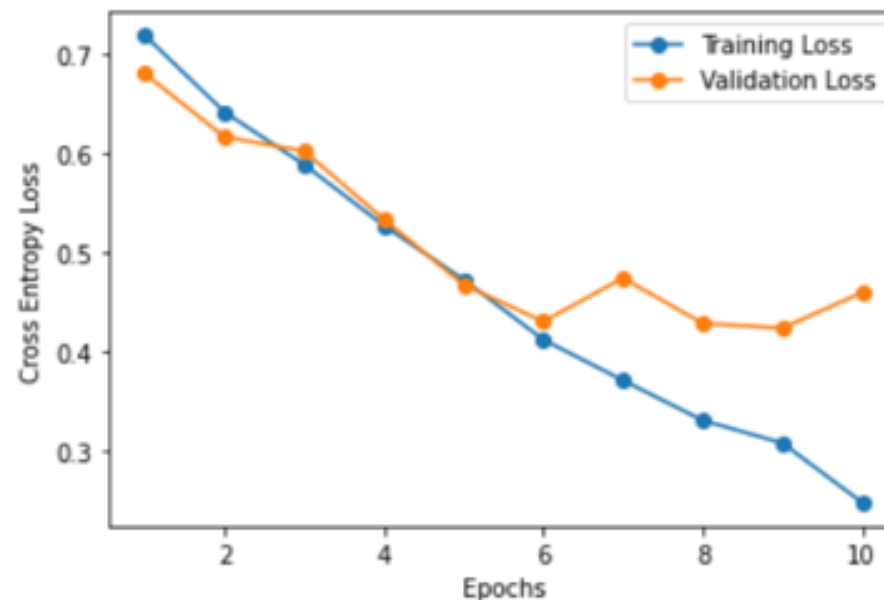
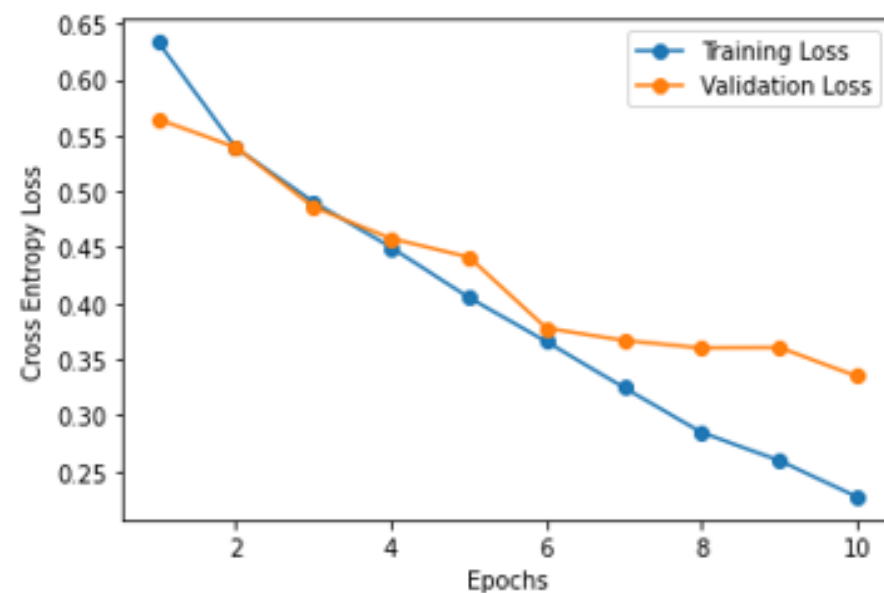
算法描述
&
模型训练



**求解结果
&
比较分析**

Naive Model (epoch=10)

		训练前		训练后	
		清晰图像	模糊图像	清晰图像	模糊图像
AlexNet	训练集	50	50	89.44	93.39
	测试集	50	50	87.81	87.01
VGG16	训练集	50	50	87.98	93.22
	测试集	50	50	81.76	83.48

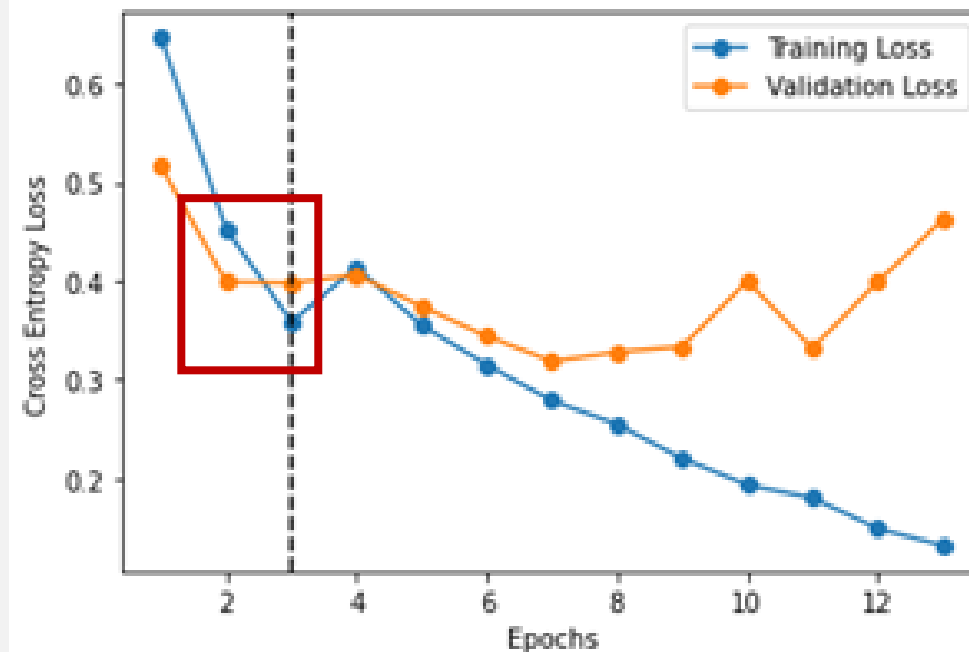


AlexNet Expert Model (epoch = 3+10)

		训练前		预训练后		训练后	
		清晰图像	模糊图像	清晰图像	模糊图像	清晰图像	模糊图像
AlexNet	训练集	50	50	85.26	62.70	87.44	95.73
	测试集	50	50	82.88	61.75	85.28	86.57

- Expert Model在测试集上的准确率结果与 Naive Model没有明显差异
- 预训练损失未接近收敛，可能导致效果不显著

=>增加预训练的epoch (3->10)

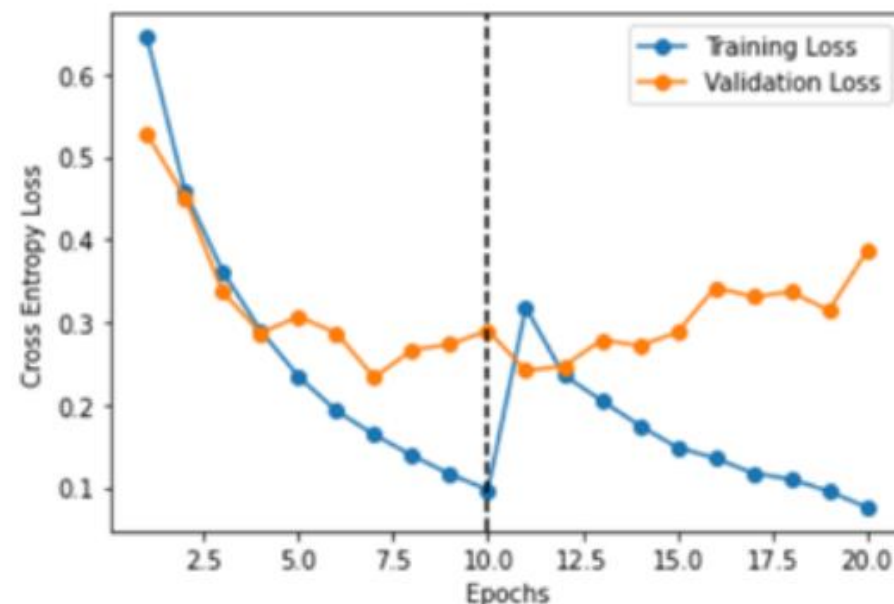


AlexNet Expert Model (epoch = 10+10)

		训练前		预训练后		训练后	
		清晰图像	模糊图像	清晰图像	模糊图像	清晰图像	模糊图像
AlexNet	训练集	50	50	98.16	75.21	88.49	97.57
	测试集	50	50	92.34	73.10	86.09	88.41

- 预训练后，对于清晰图像有非常高的准确率
- 完成训练后，清晰图像准确率下降，但模糊图像准确率得到很大提升

=> 利用清晰图像进行预训练帮助识别模糊图像



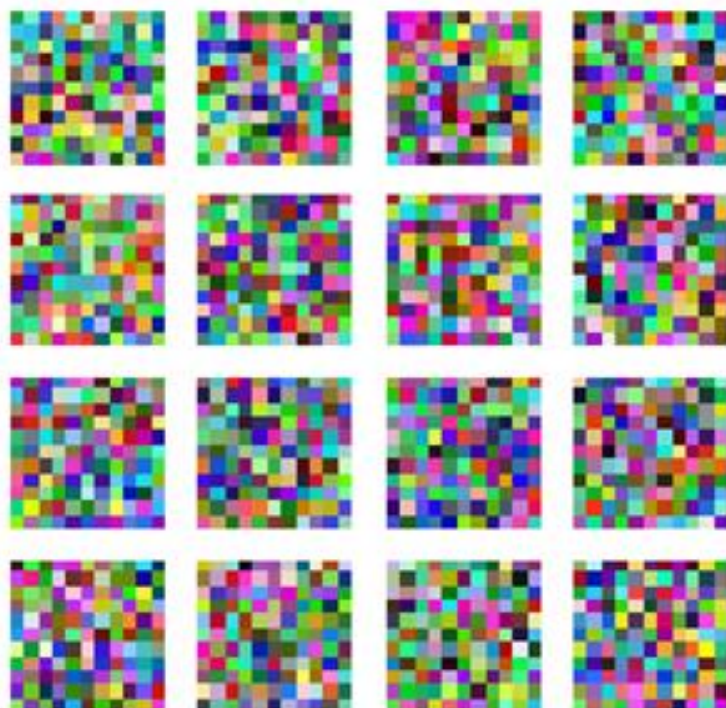
Expert Model (epoch = 10+10)

	训练前	训练后
AlexNet	50	88.41
VGG	50	90.34

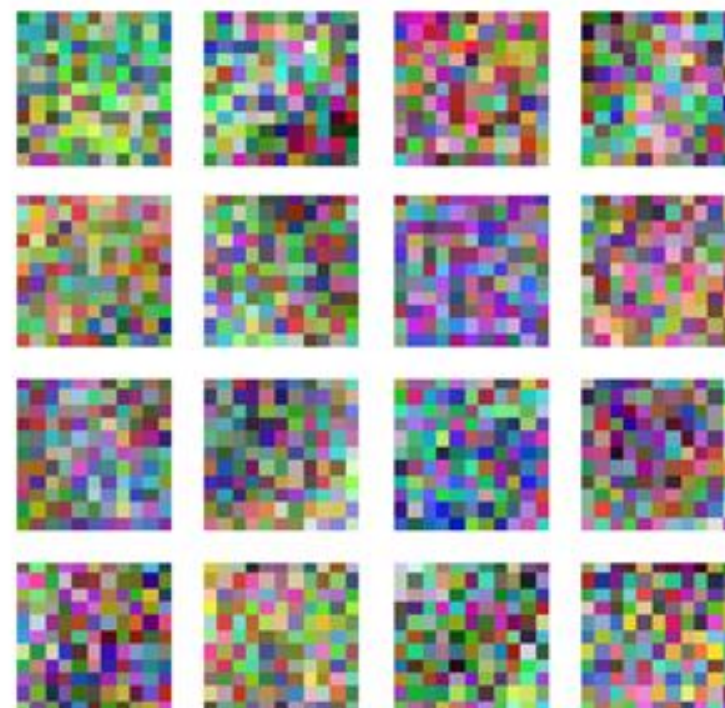
- VGG准确率高于AlexNet
- **网络结构差异**：VGG使用堆叠的小卷积核来代替大卷积核，使网络深度增加。
更深的神经网络在利用更多数据进行多次迭代训练后，能得到更优的模型。

模型特征提取 AlexNet Naive Model

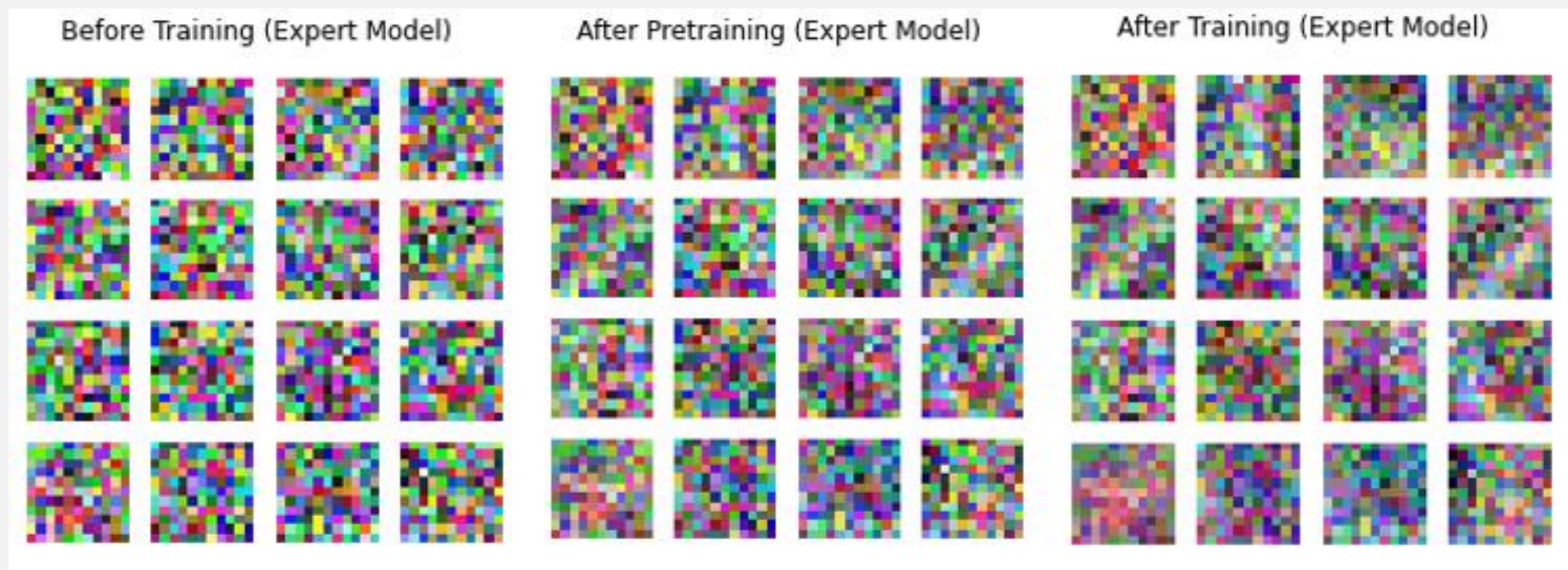
Before Training (Naive Model)



After Training (Naive Model)

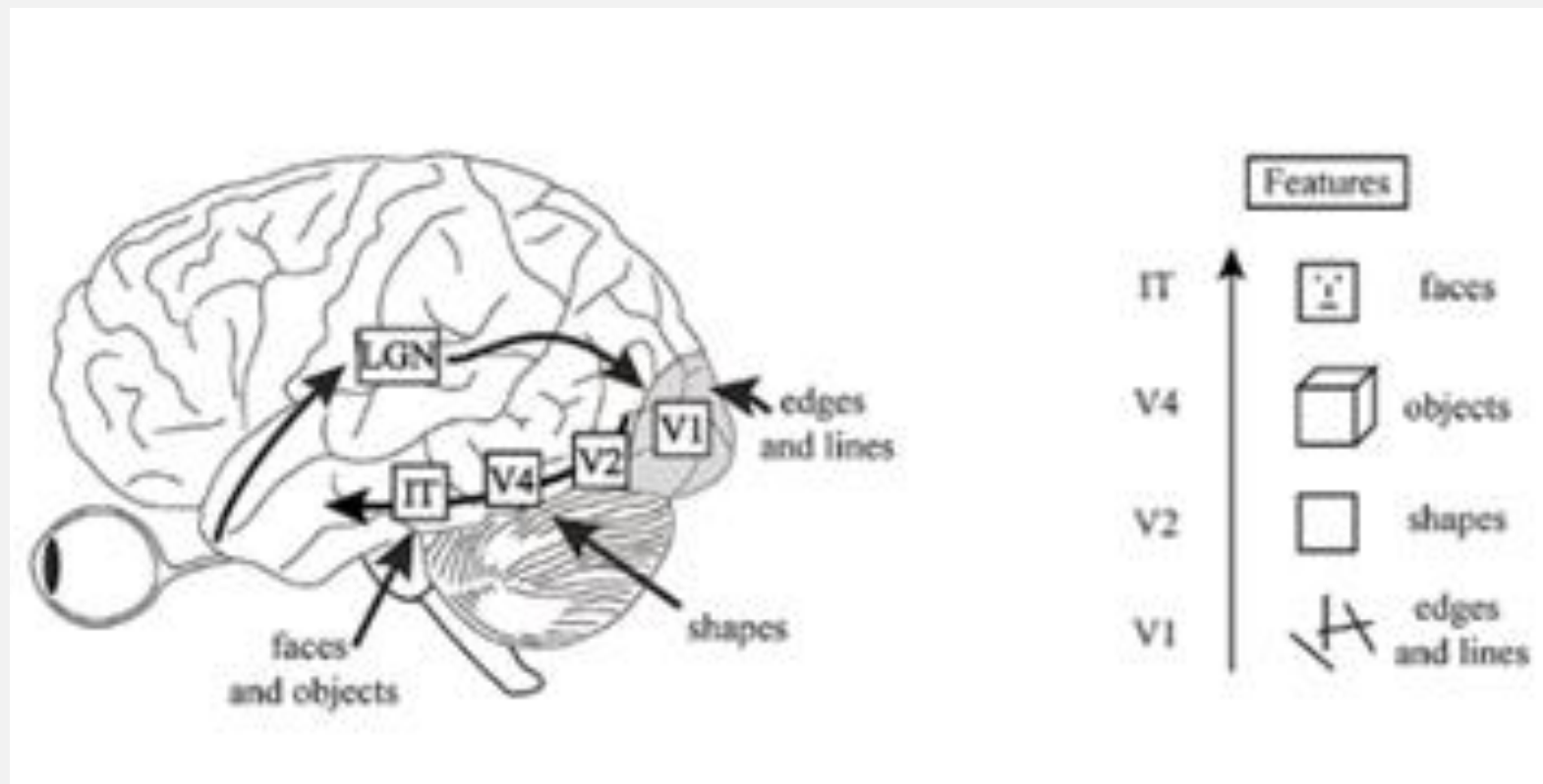


模型特征提取 AlexNet Expert Model

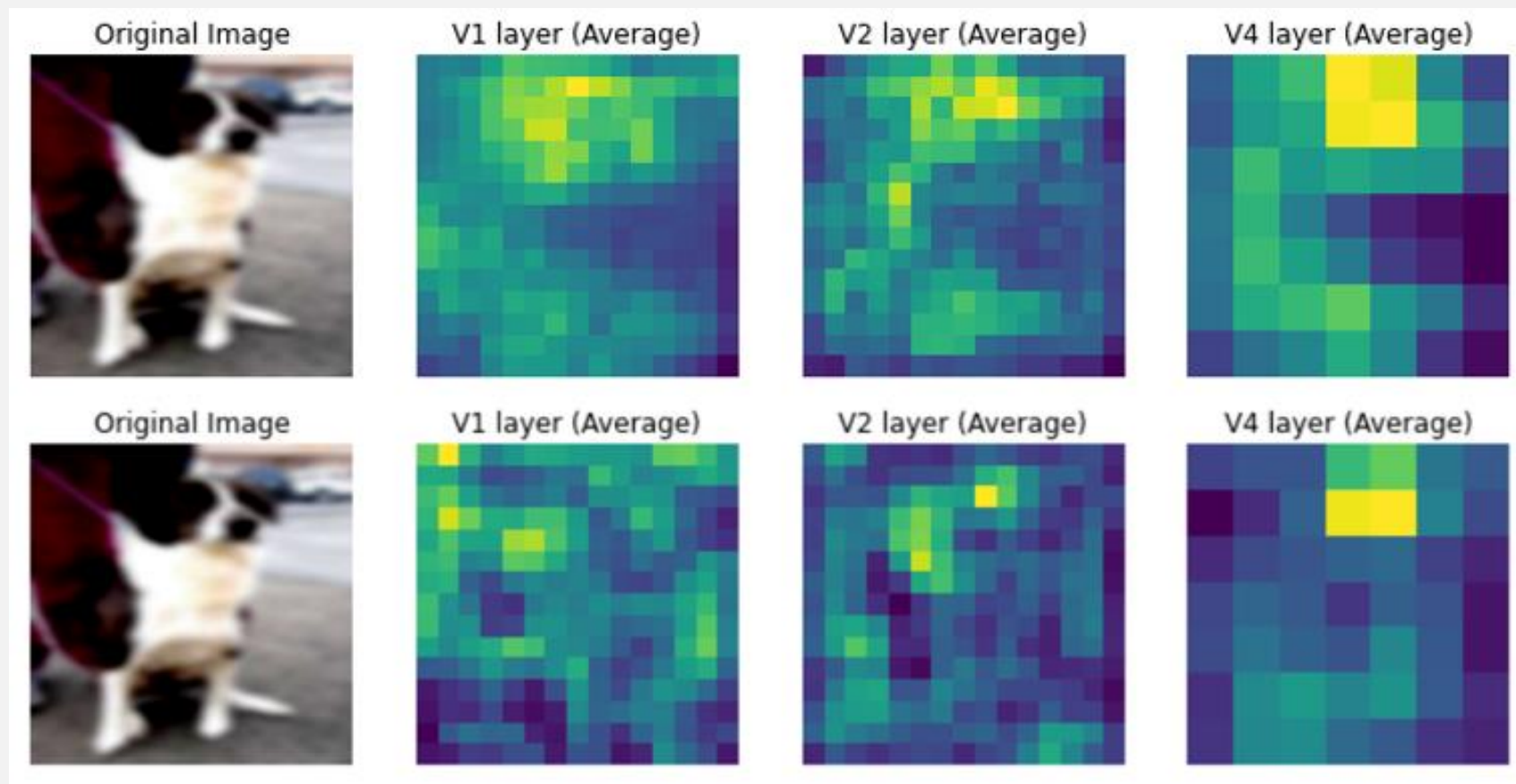


人类视觉系统

- 信息传递的**层级结构**
- AlexNet第三四五层卷积层
对应V1、V2、V4



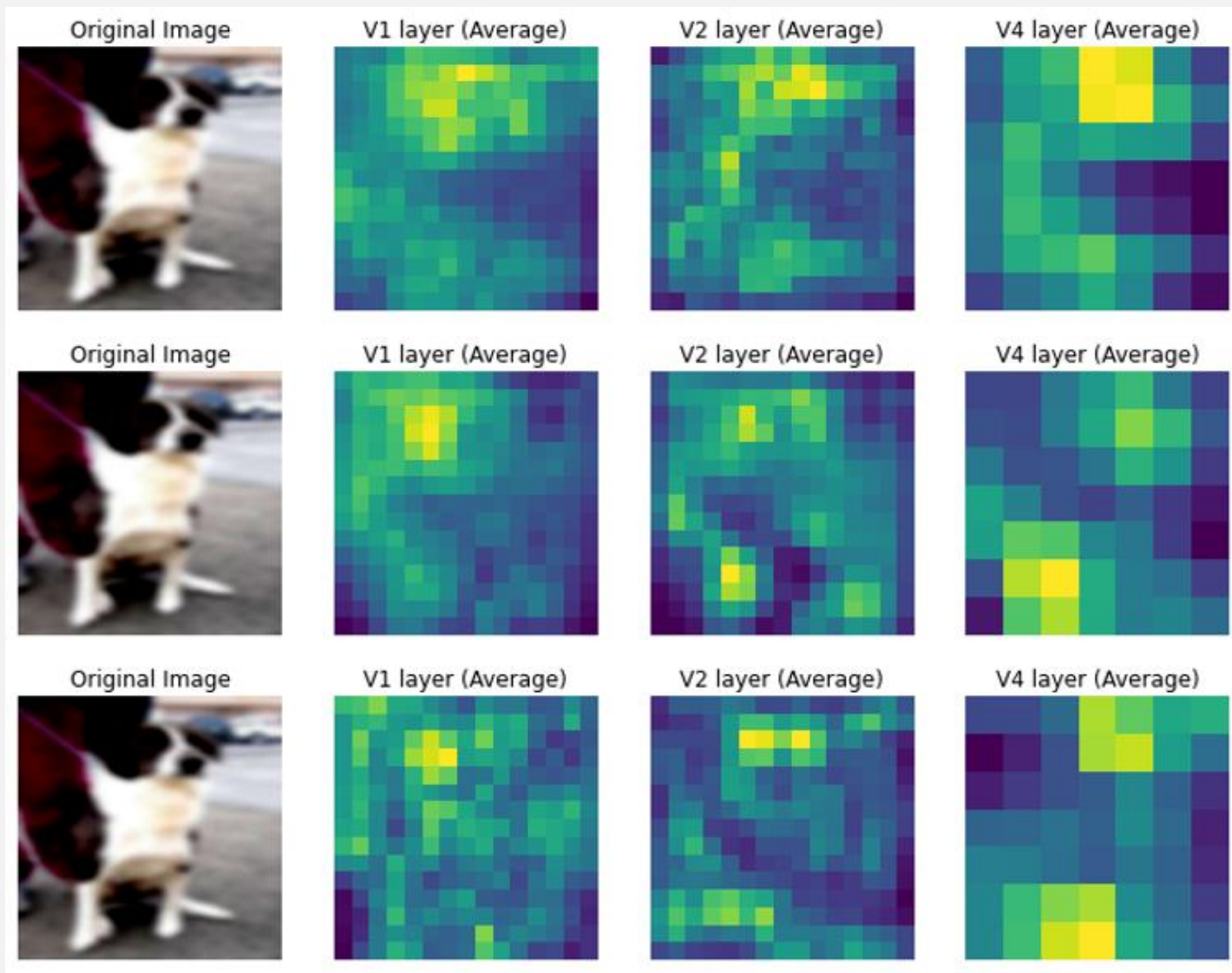
AlexNet Naive Model



- V1: 边缘线条、朝向
 - V2: 轮廓形状
 - V4: 物体、颜色
-
- 上图为训练前
 - 下图为训练后

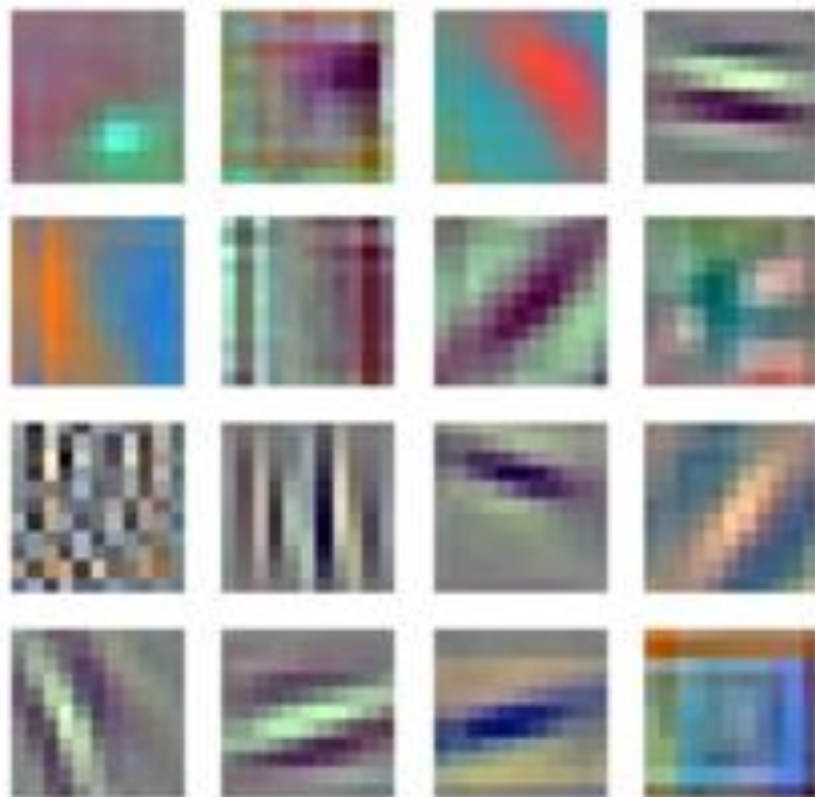
AlexNet Expert Model

- V1: 边缘线条、朝向
 - V2: 轮廓形状
 - V4: 物体、颜色
-
- 上图为训练前
 - 中间为预训练后
 - 下图为训练后

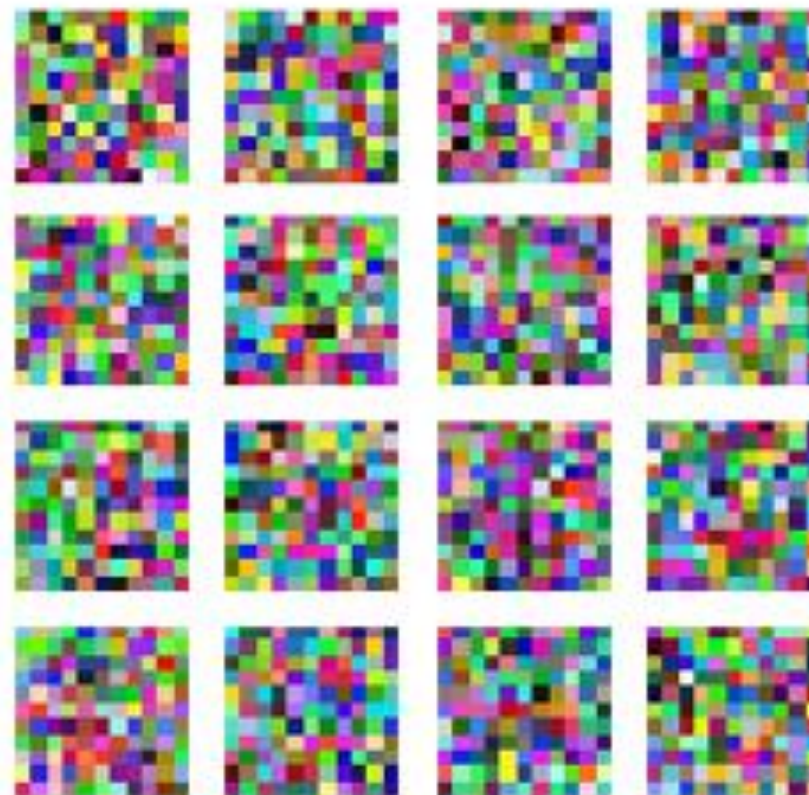


ImageNet Pretrained AlexNet

Before Training (ImageNet Pretrained Model)



Before Training (Expert Model)



总结

- 我们采用了两种不同的训练方式：预训练的模型能够得到更优的绩效表现，这说明 **“经验”** 对模型的重要性
- 通过比较AlexNet和VGG两个模型：网络结构更深的VGG略优于AlexNet，即 **网络结构** 对于模型训练也有一定作用
- 通过对AlexNet进行可视化：只有经过大样本数据集训练后的网络才会表现出与视觉系统类似的特征，证明了 **视觉系统的复杂性**，模型结构可继续优化



勤学 / 修德 / 明辨 / 笃实