

华中科技大学

2024

人工智能导论实验报告

专	业:	计算机科学与技术专业
班	级:	计算机本硕博 2301
学	号:	U202315760
姓	名:	潘裕洲
电	话:	15180972101
邮	件:	3285791820@qq.com
完成日期:		2024 年 1 月 18 日

摘要：在深度学习领域，卷积神经网络（CNN）的发展日新月异，Inception 系列网络架构在其中占据着重要地位。从最初的 GoogLeNet（Inception V1）到后续的 BN-Inception、Inception V3、Inception V4 以及 Xception 等，这些架构的不断演进推动了图像识别等任务性能的持续提升。本报告将详细阐述 Inception 系列网络架构的发展历程、关键技术创新以及各版本之间的差异与联系。

关键词：GoogLeNet，卷积神经网络（CNN），池化量（Pooling）

1 GoogLeNet (Inception V1)

1.1 Motivation

深度神经网络性能提升常依赖增加网络规模，但这易引发过拟合与高计算量问题。传统稀疏连接虽曾用于打破对称性和提升学习能力，但受限于计算机对非统一稀疏数据结构的处理能力。传统的网络使用随机稀疏连接，到了 AlexNet 的时代重新启用了全连接层，目的是为了实现在更好的并行运算加速。为兼顾网络稀疏性与密集矩阵计算性能，Inception 模块应运而生。

1.2 Inception module

核心思想是以密集模块逼近局部最优稀疏结构。

该版本采用 1×1 、 3×3 和 5×5 卷积核及池化层组合，利用 1×1 卷积提取局部信息，不同尺寸卷积核获取不同范围空间信息，池化层增强模型对特征的抽象能力。同时，依据网络深度调整卷积核比例，网络越靠前的层越注重提取局部信息，越靠后的层越注重提取大范围空间信息。并通过 1×1 卷积在 3×3 和 5×5 卷积前降维，有效控制计算复杂度，实现多尺度视觉信息处理与聚合。

总而言之，Inception 模块可以增加模型宽度和深度，堆叠时不会导致无法控制的计算复杂度。视觉信息可以在不同尺度上进行处理，然后进行聚合，以便下一阶段可以同时从不同尺度上提取特征。

1.3 GoogLeNet 网络架构

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

图 1 GoogLeNet 网络架构

整体网络包含多个卷积层、池化层和 Inception 模块。输入为 224×224 的 RGB 图像，经预处理后进入网络。各层参数设置精细，如卷积层的 patch size、stride 及不同卷积核数量等。在训练过程中，使用 ReLu 激活函数提升非线性表达能力，借助平均池化替代 flatten 操作提高 top-1 精确度，引入 dropout 防止过拟合，还在特定模块后设置辅助分类器辅助训练，综合提升网络性能。

2 BN - Inception

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	double #3×3 reduce	double #3×3	Pool +proj
convolution*	7×7/2	112×112×64	1						
max pool	3×3/2	56×56×64	0						
convolution	3×3/1	56×56×192	1		64	192			
max pool	3×3/2	28×28×192	0						
inception (3a)		28×28×256	3	64	64	64	64	96	avg + 32
inception (3b)		28×28×320	3	64	64	96	64	96	avg + 64
inception (3c)	stride 2	28×28×576	3	0	128	160	64	96	max + pass through
inception (4a)		14×14×576	3	224	64	96	96	128	avg + 128
inception (4b)		14×14×576	3	192	96	128	96	128	avg + 128
inception (4c)		14×14×576	3	160	128	160	128	160	avg + 128
inception (4d)		14×14×576	3	96	128	192	160	192	avg + 128
inception (4e)	stride 2	14×14×1024	3	0	128	192	192	256	max + pass through
inception (5a)		7×7×1024	3	352	192	320	160	224	avg + 128
inception (5b)		7×7×1024	3	352	192	320	192	224	max + 128
avg pool	7×7/1	1×1×1024	0						

图 2 BN - Inception 架构

GoogLeNet Inception V1 改进，主要引入批次归一化（Batch Normalization）技术，优化网络训练过程中的梯度传播与数据分布，增强模型稳定性与收敛速度。从上图可以看到，我们将 Inception 模块中的 5×5 卷积核替换为 2 个 3×3 卷积核，减少计算量，并在池化与降维操作上进行优化，进一步提升网络性能。

3 Inception V3

3.1 朴素设计

包括避免浅层网络过度降维，保持特征图尺寸合理缩减；增加特征数量促进收敛；利用 1×1 卷积在大卷积核前降维；均衡网络宽度与深度等，为网络设计提供重要指导。

3.2 优化思路

首先是分解大卷积：用多个小卷积替代大卷积核，如 2 个 3×3 卷积替代 5×5 卷积，或采用非对称卷积（如 3×1 和 1×3 组合）分解，有效降低计算量且提升性能，同时确定各阶段 ReLu 激活函数的优势。

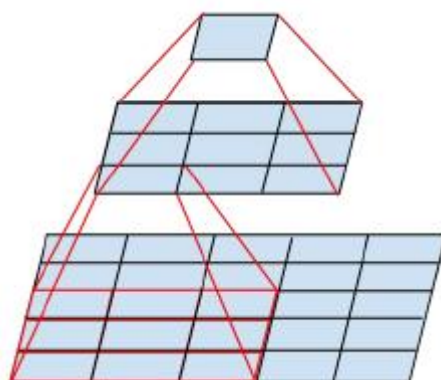


图 3 分解为小卷积

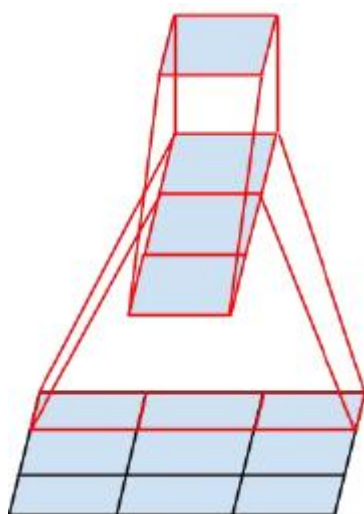


图 4 分解为非对称卷积

然后是辅助分类器：最初目的是将有用的梯度推到较低的层，使其立即有用，以及在非常深的网络中解决消失梯度问题，但从具体表现来看其主要起正则化作用，且在训练不同阶段对收敛性影响各异。

最后还包括高效下采样技巧：创新地在 Inception 模块中采用并行结构实现扩充通道数与下采样同步，避免传统池化升维方法的信息丢失或计算量过大问题。

3.3 Inception V3 网络架构

type	patch size/stride or remarks	input size
conv	$3 \times 3/2$	$299 \times 299 \times 3$
conv	$3 \times 3/1$	$149 \times 149 \times 32$
conv padded	$3 \times 3/1$	$147 \times 147 \times 32$
pool	$3 \times 3/2$	$147 \times 147 \times 64$
conv	$3 \times 3/1$	$73 \times 73 \times 64$
conv	$3 \times 3/2$	$71 \times 71 \times 80$
conv	$3 \times 3/1$	$35 \times 35 \times 192$
3×Inception	As in figure 5	$35 \times 35 \times 288$
5×Inception	As in figure 6	$17 \times 17 \times 768$
2×Inception	As in figure 7	$8 \times 8 \times 1280$
pool	8×8	$8 \times 8 \times 2048$
linear	logits	$1 \times 1 \times 2048$
softmax	classifier	$1 \times 1 \times 1000$

图 5 Inception V3 网络架构

网络架构深度达 42 层，对 V1 架构的卷积层进行分解优化，Inception 模块依特定规则组合，在保证计算效率前提下提升性能，输入感受野可灵活调整，适应不同场景需求。

4 Inception V4

4.1 Motivation

2015 年，何凯明引入残差连接，大大提升了图像识别和目标检测的效果。受残差连接在深度网络训练中显著效果启发，探索其在 Inception 架构中的应用。虽不使用残差连接也能训练深层网络，但残差连接可大幅加速收敛。

4.2 Residual Inception

Inception - ResNet 系列融合残差连接与 Inception 模块。在设计上，Inception 后采用 1×1 卷积升维，合理设置 BN 层位置减少计算量，并针对卷积核过多导致的输出不稳定问题，提出缩放残差分支结果的方法，确保网络性能稳定提升。

具体地，给出了 3 种网络架构：

Inception-ResNet-v1：一种混合的 Inception 版本，其计算成本与 Inception-v3 类似。

Inception-ResNet-v2：成本更高的混合版 Inception，识别性能显著提高。

Inception-v4：一种没有残差连接的纯 Inception 变体，其识别性能与 Inception-ResNet-v2 大致相同。

5 Xception

5.1 Xception 模块

Xception 模块处于普通卷积与深度可分离卷积之间，将通道分组后分别卷积实现空间与跨通道信息解耦。与深度可分离卷积相比，在操作顺序和非线性激活应用上存在差异，且经研究确定了这些差异对网络性能的影响。

5.2 Xception 网络架构

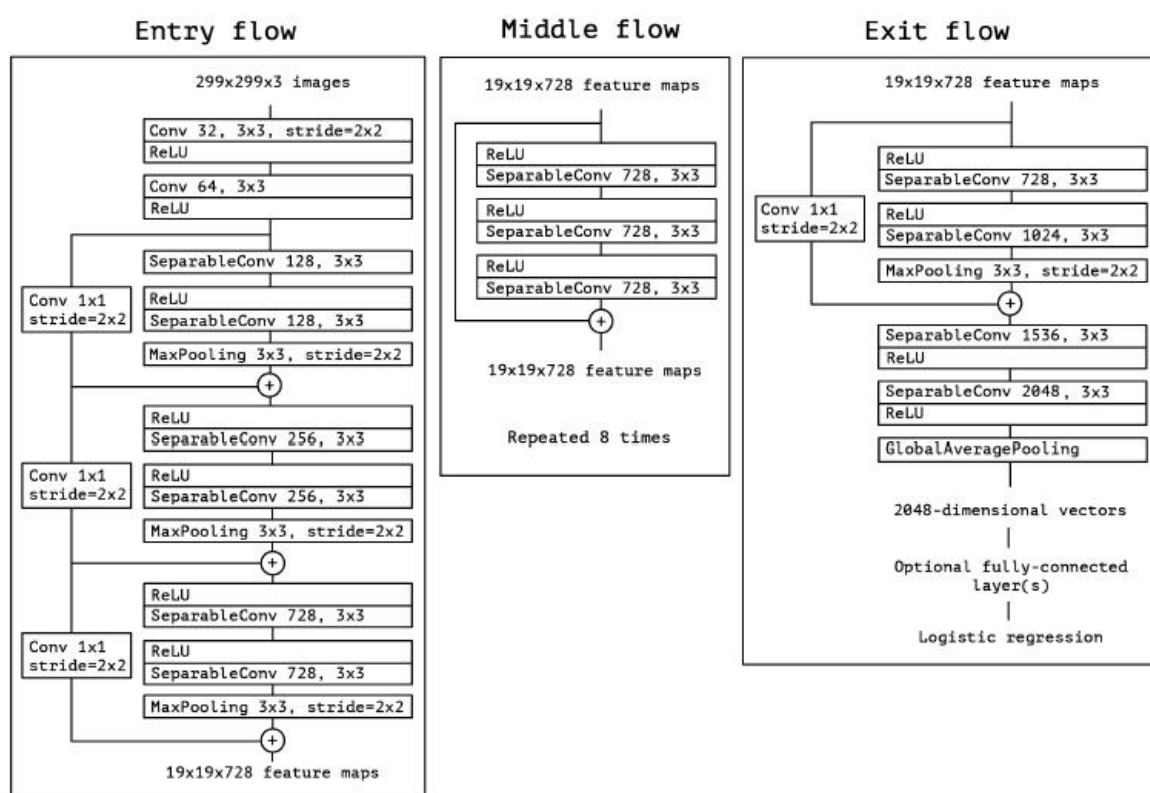


图 6 Xception 网络架构

Xception 网络架构由 Entry flow、Middle flow 和 Exit flow 构成，包含 36 个卷积层分为 14 个模块，除首尾模块外均采用线性残差连接，基于扩大数据集的正则化效果未使用 dropout 和辅助分类器，整体架构简洁高效。

6 总结

Inception 系列网络架构通过不断创新与优化，在网络结构设计、模块组合、训练技巧等方面取得显著进展，持续推动图像识别等领域性能提升。未来研究可聚焦于进一步融合新兴技术，如自适应计算、新型激活函数等，探索更高效的网络架构，拓展其在复杂场景与多模态数据处理中的应用，有望在深度学习领域持续发挥重要作用并取得新突破。