

基于深度学习的数学模型在计算机视觉中的应用

梁家德

中国人民大学, 数学学院, 北京 100872

摘要: 随着大数据与人工智能技术的快速发展, 深度学习在计算机视觉领域得到了极大关注和广泛应用。本文从数学模型的角度出发, 系统回顾了深度神经网络的研究历程与关键方法, 介绍了基于深度卷积的数学模型、对抗生成模型以及小样本与无监督学习等领域的最新进展。最后, 针对如何进一步摆脱对海量数据的依赖、提升模型的泛化能力以及增强系统的可解释性等方面展开探讨。本文旨在为相关研究人员与应用开发者提供系统的理论指导与思路借鉴。

关键词: 深度学习; 数学模型; 计算机视觉

Application of Mathematical Model Based on Deep Learning in Computer Vision

Liang, Jiade

School of Mathematics, Renmin University of China, Beijing, 100872, China

Abstract: With the rapid development of big data and artificial intelligence technology, deep learning has attracted great attention and been widely used in the field of computer vision. From the perspective of mathematical model, this paper systematically reviews the research history and key methods of deep neural network, and introduces the latest progress in the fields of mathematical model based on deep convolution, confrontation generation model, small sample and unsupervised learning. Finally, how to get rid of the dependence on massive data, improve the generalization ability of the model and enhance the interpretability of the system are discussed. This paper aims to provide systematic theoretical guidance and ideas for relevant researchers and application developers.

Keywords: Deep learning; Mathematical model; Computer vision

DOI: 10.62639/sspis37.20250203

在计算机视觉领域, 研究者致力于让计算机具备类人视觉感知能力, 但因视觉系统复杂性面临诸多挑战。早期的模式识别依赖人工设计特征, 虽实现一定目标识别, 但对高维图像和复杂场景鲁棒性差。统计学习方法如 Viola 的 AdaBoost 算法结合 Haar-Like 特征, 提高了实时人脸检测的效率, 但仍受限于特征选择的耗时和泛化能力不足。随着硬件和数据的发展, 深度学习依靠海量数据训练展现出卓越的特征学习能力, 尤其是 2012 年 AlexNet 模型的成功引发了“深度学习革命”。尽管其推动了计算机视觉的突破性进展, 对标注数据的依赖也促生了小样本、无监督学习等新研究方向, 揭示了数学模型优化的研究价值^[1]。

一、数学模型与深度神经网络

(一) 人工神经元与前向传播

人工神经网络 (Artificial Neural Network, ANN) 最初源于对生物神经系统的模拟。在最基础的形式中, 单个神经元可视为一个从输入到输出的非线性映射: $a = f(\sum_{i=1}^n w_i x_i + b)$ 。其中, x_i 为输入特征, w_i 为连接权重, $f(\cdot)$ 为偏置, 为 Sigmoid、Tanh 或 ReLU 等非线性激活函数。若将上述神经元堆叠为多层, 则构成了多层感知器 (Multi-Layer Perceptron, MLP), 其本质是对输入向量进行多次线性变换与非线性映射, 最终

输出某种分类或回归结果。该过程可称为前向传播 (Forward Propagation)。

(二) 反向传播与梯度下降

反向传播 (BP) 通过定义损失函数计算梯度, 结合梯度下降优化网络权重和偏置。常用变形算法包括随机梯度下降、Adam 等。由于高维参数空间中损失函数复杂, 可能出现局部极小点和鞍点, 影响收敛速度与效果。为此, 改进算法 (Adam、RMSProp)、正则化 (L2 正则、Dropout) 及良好的初始化方式被广泛采用, 显著地提高了模型的收敛稳定性和性能。

二、卷积神经网络与视觉应用

(一) 卷积操作与局部感受野

对图像任务而言, 若直接将所有像素展开为一维输入, 参数量将十分庞大且易破坏图像的空间关联性。为此, 卷积神经网络 (Convolutional Neural Network, CNN) 在输入端引入了局部感受野与参数共享机制: 通过卷积核 (即可学习的滤波器) 在图像局部滑动, 提取特征并生成相应的子特征图 (Feature Map)。这种设计既能保留图像的局部结构信息, 又显著减少了网络参数, 避免过拟合。

假设输入图像为二维矩阵 $I \in \mathbb{R}^{H \times W}$, 卷积核 (权重矩阵) 为 $K \in \mathbb{R}^{k \times k}$, 则卷积操作可以表示为:

(稿件编号: IS-25-3-72001)

作者简介: 梁家德 (1994-), 男, 汉族, 硕士在读, 河南省商丘市人, 内容单位标注: 中国人民大学, 从事计算机和数学相关研究。

$$F(x, y) = \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} K(i, j) \cdot I(x+i, y+j) + b$$

其中: $F(x, y)$ 是输出特征图的值; (x, y) 是特征图中的位置; k 是卷积核的大小 (如 3×3 或 5×5); b 是偏置项。

在上述公式中, 卷积核 K 在整个输入图像上滑动, 所有位置共享相同的权重参数 $K(i, j)$ 。这大大降低了参数数量, 相较于全连接网络需要独立学习每个权重的情况, 卷积层学习到的参数仅为 k^2+1 (包括偏置)。

局部感受野指的是每个输出位置 $F(x, y)$ 仅依赖于输入图像中 $k \times k$ 的局部区域, 这种机制保证了特征提取的空间局部性, 同时保留了图像的空间结构。

若输入图像有 C 个通道 (如 RGB 图像的 $C=3$), 则卷积核会为每个通道分别定义一个权

重矩阵 K_C 。多通道卷积的公式为:

$$F(x, y) = \sum_{c=1}^C \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} K_c(i, j) \cdot I_c(x+i, y+j) + b$$

其中: C 是输入通道数; 输出特征图的通道数与卷积核的数量相等。

在实际操作中, 卷积的步幅 s 和填充 p 也会影响输出特征图的尺寸:

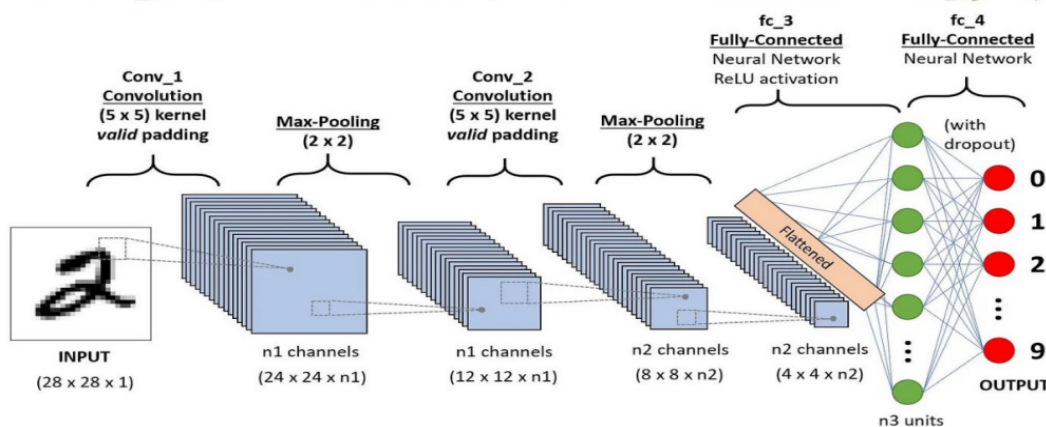
输出特征图的尺寸为:

$$H_{out} = \frac{H_{in}-k+2p}{s} + 1, W_{out} = \frac{W_{in}-k+2p}{s} + 1$$

当 $s=1$ 且 $p=(k-1)/2$ 时, 称为“填充卷积”, 可以保持输入输出的空间尺寸相同。

在卷积层的后面通常会配合池化 (Pooling) 操作 (如最大池化、平均池化), 以实现特征图的下采样, 从而缩小特征图空间尺寸、减少计算量并增强网络的平移与尺度不变性。

卷积 池化 全连接



卷积图

(二) 深度卷积网络的演进

自 AlexNet 成功以来, VGGNet、GoogLeNet、ResNet 和 DenseNet 等深度卷积网络显著提升了图像分类、目标检测和语义分割精度。VGGNet 用 3×3 小卷积核堆叠提高特征提取细腻度; GoogLeNet 的 Inception 模块结合 1×1 卷积实现降维和网络宽度扩展; ResNet 引入残差连接, 解决深度网络中的梯度消失问题; DenseNet 通过密集连接复用低层特征。这些设计基于对卷积运算和激活函数的数学思考, 成为许多视觉任务的基础模型。

(三) 胶囊网络 (CapsNet) 与同变性

传统的 CNN 通过池化层实现平移等变化的不变性, 但也意味着网络对位置、角度等关键信息的忽略。一旦物体发生大幅旋转或视角改变, 网络识别率急剧下降。为解决此问题, Hinton 等人提出了胶囊网络 (Capsule Network, CapsNet), 以向量 (称为 Capsule) 来表示图像中更高级的“实体特征”, 对朝向、大小等属性进行建模。CapsNet 在多数数字重叠任务和重构任务上取得了有别于传统 CNN 的表现, 但其核心的“动态路由”机制训练代价大, 迭代不稳定,

目前仍是研究热点。

三、深度学习在计算机视觉中的主要应用

(一) 深度学习在图像分类与小样中的应用

图像分类是计算机视觉的基础任务之一, 深度卷积网络在大型数据集上不断刷新准确率。从 AlexNet 展示深度学习潜力到 ResNet 通过残差连接解决梯度消失问题, 模型性能大幅提升。然而, 医疗影像等领域因标注数据不足, 小样本学习成为研究热点。小样本学习方法包括: (1) 迁移学习, 利用在大规模数据集上预训练的模型, 在目标数据集上微调, 减少标注数据需求; (2) 记忆组件, 如 MANN, 引入记忆模块存储和检索知识, 提升识别能力。

(二) 深度学习在语义分割与场景理解中的应用

在语义分割任务中, 需要对每个像素进行分类, 这对模型的空间定位能力提出了更高要求。Fully Convolutional Network (FCN) 首先将卷积网络扩展为“全卷积”形式, 实现了端到端的像素级预测。随后, DeepLab、SegNet、PSPNet

等模型引入了空洞卷积、编码器—解码器结构及多尺度特征池化等策略, 不断提高分割精度。

(三) 深度学习在生成对抗网络 (GAN) 与图像生成中的应用

生成对抗网络 (GAN) 通过生成器和判别器的对抗训练, 在无监督情境下展现出卓越的图像生成能力。它广泛应用于图像修复、风格迁移和语义合成等领域。GAN 可智能填补图像缺失部分, 还原原貌; 通过风格迁移, 将一种艺术风格应用于另一图像; 在语义合成中, 根据语义信息生成对应场景。改进模型如 DCGAN 稳定了训练过程, 提升生成质量; WGAN 引入 Wasserstein 距离及梯度惩罚, 缓解模式崩溃问题; CycleGAN 和 Pix2Pix 则实现了图像到图像的惊人转换, 可以将普通马匹变为带有斑马条纹的模样。

四、数学模型在视觉前沿研究中的新趋势

(一) 无监督学习与自监督学习

深度学习在图像识别中通常依赖大量标注数据, 而标注工作昂贵且耗时。为此, 数学模型被用于突破大数据依赖的限制, 支持无监督学习和自监督学习的研究。

无监督学习的核心在于通过优化某种目标函数, 从无标签数据中提取有意义的潜在表示。例如, 在生成对抗网络 (GAN) 和自编码器 (Autoencoder) 中, 数学建模主要依赖于损失函数的设计:

GAN 通过生成器 G 和判别器 D 的对抗训练生成逼真的数据, 其损失函数为: $\min_G \max_D \mathcal{L}(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))]$ 。

其中, $p_{\text{data}}(x)$ 是真实数据分布, $p_z(z)$ 是生成器的输入噪声分布。

自编码器的目标是通过编码器 E 和解码器 D 学习数据的有效表示 z , 其优化目标为: $\mathcal{L}(x, \hat{x}) = \|x - D(E(x))\|_2^2$ 。

自监督学习通过设计辅助任务, 利用数据内在的结构化信息生成伪标签, 数学建模常以辅助任务的优化目标为核心, 例如:

将图像按 θ 度旋转 ($\theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$), 训练网络预测旋转角度的交叉熵损失:

$$\mathcal{L}_{\text{rotation}} = -\sum \theta p(\theta) \log q(\theta),$$

其中 $p(\theta)$ 是真实旋转标签, $q(\theta)$ 是模型预测概率。

对比学习 (Contrastive Learning), 即学习样本间的相似性表示, 优化对比损失 (如 InfoNCE): $\mathcal{L}_{\text{contrastive}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(z_i, z_k)/\tau)}$, 其中 z_i, z_j 是正样本对的嵌入, τ 是温度参数。

(二) 更紧凑高效的模型

随着深度模型参数数量的增加, 计算和存储成本呈指数增长, 数学模型为优化网络结构提供了有效途径。

权重剪枝: 通过稀疏化权重矩阵 W , 剔除低重要性的参数。例如, 将权重小于阈值 ϵ 的部分置零:

$$W_{ij} = \begin{cases} W_{ij}, & \text{if } |W_{ij}| > \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

量化: 将浮点参数离散化为低精度表示 (如 8-bit 整数), 显著降低存储需求:

$$W_q = \text{round}\left(\frac{W}{\Delta}\right) \times \Delta,$$

其中 Δ 是量化步长。

神经架构搜索通过数学优化方法自动发现最优模型结构。例如, 基于强化学习的目标函数为: $\mathcal{L}_{\text{NAS}} = \mathbb{E}_{m \sim p(m|\theta)} [\mathcal{L}_{\text{task}}(m)]$,

其中 $p(m|\theta)$ 是模型结构的概率分布, $\mathcal{L}_{\text{task}}(m)$ 是候选模型在特定任务上的损失。

(三) 对抗鲁棒性与可解释性

深度模型的对抗鲁棒性和可解释性直接影响其在高风险场景 (如医疗、自动驾驶) 中的应用。

对抗攻击通过添加微小扰动 δ 生成对抗样本 $x' = x + \delta$, 导致模型预测错误。防御方法可建模为对抗训练的优化问题: $\min_{\theta} \mathbb{E}_{(x, y) \sim \mathcal{D}} [\max_{\delta} \|\delta\|_p \leq \epsilon \mathcal{L}(f_{\theta}(x + \delta), y)]$,

其中 ϵ 是扰动范围, $\mathcal{L}$ 是损失函数。

可解释性常基于模型输出对输入的梯度分析:

$$\text{Saliency}(x) = \left| \frac{\partial f_{\theta}(x)}{\partial x} \right|,$$

该方法直观地反映了模型预测对输入像素的重要性。

五、结论与展望

深度学习在计算机视觉中的成功得益于神经网络强大的特征学习能力和大规模数据支撑。数学上, 深度神经网络通过非线性映射拟合了输入到输出的复杂分布关系。然而, 当前模型仍面临依赖大数据、易受对抗攻击、缺乏可解释性以及训练高成本等问题。未来研究应关注以下方向: 一是探索新型网络结构、优化策略, 提升模型鲁棒性与效率; 二是摆脱对大规模标注数据的依赖, 发展无监督、小样本和迁移学习; 三是增强模型可解释性与可靠性, 应对高风险应用中的安全和隐私挑战; 四是推进多模态融合, 将视觉与语音、文本等模态协同, 构建智能感知系统。随着新数学模型和硬件算力的发展, 深度学习将推动视觉系统在更多领域实现强大的自动化理解与决策能力, 为社会变革带来新机遇。

参考文献:

- [1] 宋宗昊, 闫伸, 等. 基于深度学习的箱式烘烤内部温湿度预测 [J]. 中国烟草科学, 1-8.
- [2] 黎清万. 深度学习视域下的信息科技项目教学三维三阶模型 [J]. 计算机教育, 2024, (12): 3-6+11.
- [3] 陈武超, 俞翔栋, 陈洪宇, 柯瑞庭, 陶建峰. 基于深度学习的机理模型与数据混合驱动的视觉转角测量方法 [J]. 仪表技术与传感器, 2024, (06): 121-126.
- [4] 李仁智. 基于深度学习的番茄成熟度识别及成熟期预测模型 [D]. 云南农业大学, 2022.
- [5] 潘敏婷, 王韞博, 朱祥明, 高思宇, 龙明盛, 杨小康. 基于无标签视频数据的深度预测学习方法综述 [J]. 电子学报, 2022, 50 (04): 869-886.