

CNN 在计算机视觉中的应用：从空间域到频域

李孟霖 2311067

Abstract

本文综述了卷积神经网络（CNN）在计算机视觉领域的应用演进，重点探讨其从空间域建模到频域分析的转变过程。通过分析代表性论文，我们归纳了 CNN 在图像识别、纹理分类等任务中的发展轨迹，并展望了未来结合频域表示的新方向。

1 引言

计算机视觉旨在使计算机能够“理解”图像和视频中的信息。卷积神经网络（CNN）因其对空间结构建模的能力，成为了该领域的核心方法之一。近年来，研究者逐步探索将频域信息引入 CNN 架构，进一步增强模型的表达力与鲁棒性。本文将从五篇代表性论文出发，梳理 CNN 的演化路径。

2 空间域 CNN 的奠基与发展

2.1 Gradient-Based Learning and LeNet

LeCun 等人提出的 LeNet 网络体系结构 [1] 被广泛认为是卷积神经网络（CNN）在计算机视觉领域的奠基之作。该工作展示了如何利用反向传播算法对多层神经网络进行有效训练，使其能够在几乎无需特征工程的前提下，学习到复杂的非线性决策边界，从而实现对高维图像数据（如手写数字）的高精度分类。

LeNet 所代表的 CNN 架构专为处理二维图像形状的多样性而设计，核心结构包括局部感受野、共享权重的卷积层、下采样层（池化）以及全连接分类器。这一结构首次系统性地将空间结构建模与参数效率结合起来，显著提升了模型的泛化能力和训练效率。在标准手写数字识别任务中的实验证明，CNN 的表现优于其他传统机器学习方法。

该论文还提出了图变换网络（Graph Transformer Networks, GTNs）这一新颖的全局训练范式，用于将文档识别流程中的多个模块（如字段提取、字符分割与识别、语言建模）统一为一个可微结构，使整个系统能够通过梯度下降进行端到端优化。这种思想在今天的深度学习系统中依然具有重要影响。

LeNet 及其相关系统最终被部署在实际的支票读取系统中，每天处理数百万张业务与个人支票，标志着深度学习模型在工业级图像识别任务中的首次大规模应用。这不仅验证了 CNN 模型在实际场景下的可行性，也为之后一系列视觉识别任务中的深度学习方法奠定了基础。

2.2 VGGNet: Very Deep Convolutional Networks

VGGNet 是由 Simonyan 和 Zisserman 在 2014 年提出的一种深度卷积神经网络架构 [2]，其核心创新在于通过增加网络深度显著提升图像识别的准确率。该工作系统地考察了卷积神经网络深度对大规模图像分类任务性能的影响，采用了统一且极简的设

计——所有卷积层均使用尺寸为 3×3 的小卷积核，这一设计既保证了模型的表达能力，又有效控制了参数规模。

通过将网络深度推至 16 到 19 层，VGGNet 在当时的 ImageNet 图像识别挑战赛中取得了突破性成绩，分别获得了定位任务和分类任务的第一与第二名。该网络架构不仅在 ImageNet 数据集上表现出色，还展现了良好的泛化能力，在其他图像识别数据集上同样达到了当时的最先进水平。

VGGNet 的成功表明，深度是提升卷积神经网络性能的关键因素之一，而小卷积核的堆叠方式有效地增强了模型的非线性表达能力和细粒度特征捕捉能力。这一设计理念对后续深度学习模型的发展产生了深远影响，并且 VGGNet 预训练模型的公开发布，为后续计算机视觉任务的迁移学习提供了宝贵资源。

2.3 ResNet: Deep Residual Learning

随着神经网络层数的增加，训练深层网络面临梯度消失和优化困难等挑战。He 等人于 2015 年提出的深度残差网络 (ResNet) 创新性地引入了残差学习框架，通过在网络层之间引入“跳跃连接” (skip connections)，使得每一层学习的是相对于输入的残差函数，而非直接学习未参照的映射函数。这种设计极大地缓解了深层网络训练中的退化问题，使得训练极深的网络成为可能。

ResNet 证明了网络深度的提升对性能有显著的正向影响，其设计使得网络在加深至超过 150 层时，依然能够稳定优化且有效提升分类准确率。在 ImageNet 数据集上的实验证明，残差网络不仅显著降低了训练误差，还达到了 3.57

ResNet 的残差结构成为深度学习领域的基础性架构，对后续各种变体和改进模型产生了深远影响，推动了视觉识别任务中网络设计的深度和复杂度不断突破。

3 频域 CNN 的探索：从图像结构到纹理特征

3.1 Wavelet CNN

纹理分类是图像处理领域中的一个重要且具有挑战性的问题。尽管卷积神经网络 (CNN) 在图像分类任务中取得了显著的成功，但纹理分类依然较为困难，因为纹理往往缺乏足够的对象形状信息。传统的纹理分类方法多采用频域的谱分析技术，利用纹理中重复结构的频率特征进行识别。相比之下，传统 CNN 直接在空间域处理图像，忽略了频域中潜在的重要信息，这使得两者在性能表现上存在差异。

针对这一问题，Fujieda 等人提出了一种新颖的网络架构——小波卷积神经网络 (Wavelet CNN) [4]，将频域的谱分析整合到 CNN 结构中。其核心思想是将传统 CNN 中的卷积层和池化层视为频谱分析的有限形式，进而通过小波变换对这两种操作进行推广，从而实现更为全面的频域特征提取。通过引入小波变换，Wavelet CNN 能够有效利用传统 CNN 中丢失的频谱信息，显著提升纹理分类的准确率。

实验结果表明，Wavelet CNN 在多个纹理分类任务中优于现有的 CNN 模型，同时由于参数量显著减少，该模型更易训练且对硬件资源的需求更低。这一创新性的融合方法为纹理识别任务提供了新的思路，展示了频域信息在深度学习中的潜力和价值。

3.2 Spectral Representations for Convolutional Neural Networks

传统的卷积神经网络 (CNN) 主要在空间域中进行操作，但 Rippel 等人提出的研究表明 [5]，频域不仅可以提高计算效率，还能增强模型的表达能力。通过引入频域表示，该研究提出了几项创新：

首先，作者提出了“频谱池化”（spectral pooling）方法，该方法通过在频域中截断表示来实现降维。与传统的最大池化和平均池化相比，频谱池化在保留信息的同时提供了更大的灵活性。

其次，研究引入了一种新的随机正则化形式，即通过随机修改频域分辨率来增强模型的泛化能力。这种方法在不使用传统正则化技术（如 dropout 或最大池化）的情况下，仍能在分类和逼近任务中取得竞争性的结果。

最后，作者展示了使用复系数频谱参数化卷积滤波器的有效性。虽然这种方法并不改变模型的基础结构，但它显著加快了训练过程中的收敛速度。

综上所述，该研究表明，将频域表示引入 CNN 架构，不仅可以提高计算效率，还能增强模型的表达能力和训练效率，为未来的深度学习模型设计提供了新的方向。

4 展望与未来方向

随着频域分析在卷积神经网络（CNN）中的应用逐步深入，其在计算效率、信息保留和模型泛化能力等方面展现出诸多优势。未来研究可在以下几个方向进一步拓展与深化：

- **融合空间与频域的混合网络结构**：传统 CNN 主要在空间域中操作，而频域分析擅长捕捉全局特征。将两者优势结合，可构建具备更强表达能力的混合网络架构。例如，通过在不同网络层之间引入频域变换模块，使模型在捕捉局部细节的同时兼顾全局频率分布，从而提升识别精度与鲁棒性。
- **Transformer 和频域卷积的结合**：近年来 Transformer 在计算机视觉中的表现日益突出，其全局注意力机制天然适配频域特征的处理。未来研究可探索将频域表示与视觉 Transformer（如 ViT、Swin Transformer）相结合，例如在注意力机制中引入频谱信息，或以频域为输入对图像进行建模，以提升对纹理、重复结构等频率敏感特征的建模能力。
- **更高效的频域 CNN 部署策略**：尽管频域 CNN 在训练阶段表现出较高效能，实际部署仍存在计算资源、平台支持等方面的挑战。未来可从模型压缩、稀疏频谱建模、低比特量化、FFT/IFFT 优化等角度出发，设计适用于嵌入式设备与边缘计算平台的频域网络结构，以促进其在实际场景中的落地应用。
- **在生成建模、跨模态视觉理解中的拓展应用**：频域特征对图像的全局结构与周期性信息具有较强表达能力，适合用于图像生成、风格迁移、图像复原等任务。同时，结合频域分析的模型在跨模态任务（如图文匹配、视觉问答）中，也有望作为中间表示桥梁，有效增强不同模态之间的信息对齐与关联建模能力。

总之，频域表示为 CNN 架构提供了新的视角与工具，其研究不仅有助于理论层面的模型理解，还将为实际应用中的高效、鲁棒模型设计提供坚实基础

5 总结

卷积神经网络（CNN）作为计算机视觉领域的核心方法，近年来在图像分类、目标检测、语义分割等任务中取得了显著成就。尽管传统 CNN 主要在空间域进行特征提取和模式识别，但随着理论研究的不断深入，频域特征的引入逐渐成为提升模型表达能力和泛化性能的重要方向。

本文系统回顾了将频域信息融入 CNN 的关键进展，包括将小波变换作为预处理或网络模块引入以增强纹理建模能力的 Wavelet CNN，以及利用傅里叶变换进行谱域特征学习与降维的 Spectral CNN 等代表性工作。这些研究不仅展示了频域表示在保留信息、抑制噪声和加速训练等方面的潜力，也从模型架构层面为卷积操作赋予了新的物理和数学意义。

同时，我们也讨论了深度神经网络在频率维度上的训练行为，如低频优先学习现象对早停、泛化性能的解释作用，以及频域正则化、参数重构等新型策略对模型训练的积极影响。

展望未来，空间-频域混合网络、Transformer 与频域操作的集成、面向资源受限平台的高效部署方法，以及频域技术在生成模型与跨模态理解中的扩展应用，均为该方向提供了广阔的发展前景。

综上所述，将频域分析纳入卷积神经网络设计之中，不仅是一种特征建模方式的延伸，更是理解深度学习内部机制和拓展其应用边界的重要途径。随着计算资源的持续提升和理论框架的不断完善，频域 CNN 有望在未来计算机视觉乃至更广泛的人工智能领域中发挥更大作用

References

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [4] S. Fujieda and K. Takayama, “Wavelet Convolutional Neural Networks for Texture Classification,” *The University of Tokyo, Digital Frontier Inc.*, 2020.
- [5] O. Rippel, J. Snoek, and R. P. Adams, “Spectral representations for convolutional neural networks,” *arXiv preprint arXiv:1506.03767*, 2015.