arXiv:2304.01227v1 [cs.CV] 2 Apr 2023

基于傅里叶神经算子的分辨率不变图像分类*

Samira Kabri[1(✉)]，Tim Roith[1]，Daniel Tenbrinck[1]，和 Martin Burger[2,3]

[1] 埃尔朗根-纽伦堡弗里德里希-亚历山大大学，德国埃尔朗根 91058 [2] 德国电子同步加速器，德国汉堡 22607 [3] 汉堡大学数学系，德国汉堡 20146✉ samira.kabri@fau.de

摘要。 本文研究了傅里叶神经算子（FNO）用于图像分类的效果，并将其与标准卷积神经网络（CNN）进行比较。神经算子是对无限维函数空间之间算子进行近似的一种离散化不变泛化形式，推广自神经网络。FNO——采用特定参数化的神经算子——已成功应用于参数化偏微分方程领域。我们以勒贝格空间为例，推导出连续且弗雷歇可微的神经算子的 FNO 架构。此外，我们还展示了如何将 CNN 转换为 FNO，反之亦然，并提出了该架构的一种插值等变改编方案。

关键词： 神经算子 · 三角插值 · 傅里叶神经算子 · 卷积神经网络 · 分辨率不变性

1 引言

神经网络，尤其是卷积神经网络，是图像分类任务中非常有效的工具。以卷积层取代全连接层，可以在不同细节层次上高效提取局部特征，同时保持较低的复杂度。然而，神经网络通常不具备分辨率不变性，这意味着它们无法很好地泛化到未见过的输入分辨率。除了将输入插值到训练分辨率之外，人们还提出了多种其他方法来解决这一问题，参见例如 [3,14,17]。在本工作中，我们着重将数字图像解释为函数的离散化。这使我们能够借助所谓的神经算子，将特征提取器建模为无限维空间之间的映射，参见 [13]。在第 2 节中，我们利用关于 Nemytskii 算子的已有结果，推导神经算子良定义性、连续性和 Fréchet 可微性的条件

* 本研究得到欧盟“地平线2020”计划玛丽·斯克沃多夫斯卡-居里资助协议（编号：777826）的支持。TR 和 MB 感谢德国联邦教育与研究部的支持，资助协议编号为 05M2020。SK 和 MB 感谢德国研究联合会项目 BU 2327/19-1 的支持。本研究是在 MB 任职于德国埃尔朗根-纽伦堡大学（FAU）期间完成的。

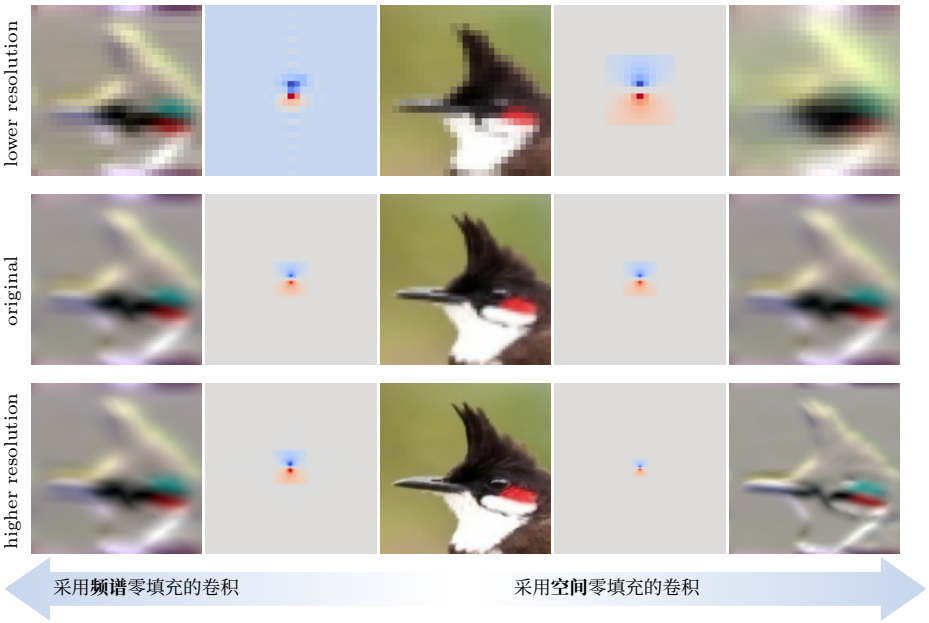


图 1：对不同分辨率的同一图像应用卷积滤波器的效果[4]。空间零填充（标准卷积神经网络实现）会改变核支撑集与图像定义域之间的关系，而频谱零填充（傅里叶神经算子实现）则能在所有分辨率下提取相当的特征。

在勒贝格空间上。我们具体证明了 [15] 中提出的傅里叶神经算子类具有这些性质；该类是卷积神经网络的一种离散化不变推广。

傅里叶神经算子的核心思想是通过傅里叶系数（即在谱域中）对卷积核进行参数化。近期研究探讨了在神经网络图像处理的背景下，使用傅里叶域中的可训练滤波器表示卷积核在性能和鲁棒性方面的表现，例如 [4,19,25]。在第3节中，我们将分析卷积神经网络与傅里叶神经算子在优化、参数复杂度以及对不同输入分辨率的泛化方面的互换性。虽然我们的理论推导仅限于实值函数，但需要指出的是，这些推导也可以自然地扩展到向量值函数。第 [24]节中的 FashionMNIST 和 Birds500 [18] 数据集4中的数值实验支持了我们的研究结果。

2 勒贝格空间上的神经算子构造

2.1 良定义性与连续性

如 [13] 中所定义，神经算子是由有限但数量任意的所谓算子层复合而成的。本节我们推导算子层各

[4] 这张图片展示了一只来自 Birds500 数据集的红耳鹎 [18]。

组成部分所需满足的条件，使其成为两个勒贝格空间之间定义良好且连续的算子。更具体地说，对于有界区域 $\Omega \subset \mathbb{R}^d$ 和 $1 \leq p,\ q \leq +\infty$，我们的目标是构造一个连续算子 $\mathcal{L}: L^p(\Omega) \to L^q(\Omega)$，使输入函数 $u \in L^p(\Omega)$ 映射为

$$\mathcal{L}(u)(x) = \sigma\left(\Psi(u)(x)\right) \qquad \text{for a.e. } x \in \Omega, \tag{1}$$

其中，我们用算子 Ψ 汇总所有仿射运算，使得

$$\Psi(u) = Wu + \mathcal{K}u + b. \tag{2}$$

这里，权重 $W \in \mathbb{R}$ 实现了一个残差分量，而核积分算子 $\mathcal{K}: \ u \mapsto \int_\Omega \kappa(\cdot, y)\, u(y)\, dy$由核函数$\kappa: \ \Omega \times \Omega \to \mathbb{R}$ 确定，将神经网络中的离散加权推广到了连续情形。类似地，偏置函数 $b: \ \Omega \to \mathbb{R}$ 是偏置向量的连续对应形式。（非线性）激活函数 $\sigma: \ \mathbb{R} \to \mathbb{R}$ 逐点作用，因此是一个 Nemytskii 算子（例如参见 [6]）。因此，稍微滥用一下记号，相关的 Nemytskii 算子可写为

$$\sigma : v \mapsto \sigma(v(\cdot)), \tag{3}$$

其中，我们假设 σ 是一个可测函数。为了确保相关的 Nemytskii 算子定义了一个映射 $\sigma: L^p(\Omega) \to L^q(\Omega)$，其中 $1 \leq p,\ q \leq \infty$ 我们要求满足以下条件:

$$\begin{aligned} \underline{p, q < \infty}: &|\sigma(x)| \leq K + \beta |x|^{\frac{p}{q}} \text{ for all } x \in \mathbb{R} \text{ and constants } \beta, K \in \mathbb{R}, \\ \underline{p = \infty}: &|\sigma(x)| \leq K(c) \text{ for every } c > 0 \text{ for all } x, |x| < c \\ &\text{and a constant } K(c) \in \mathbb{R} \text{ depending on } c, \\ \underline{p < \infty,\ \ q = \infty}: &|\sigma(x)| \leq K \text{ 对所有 } x \in \mathbb{R} \text{ 及常数 } K \in \mathbb{R}, \end{aligned} \tag{4}$$

这些在 [6]中用到。

引理 1。对于 $1 \leq p,\ q \leq \infty$ 假设 σ 满足 (4)。则对应的 *Nemytskii*算子是一个映射 $\sigma: L^p(\Omega) \to L^q(\Omega)$。

证明。与 [6, 定理 1], 类似，利用（4）中的估计可直接得出结论。

由于我们关注的是（1）中层的连续性性质，因此考虑以下关于 Nemytskii算子的连续性结果。

引理 2。对于 $1 \leq p \leq \infty, 1 \leq q < \infty$ 假设函数 $\sigma: \ \mathbb{R} \to \mathbb{R}$连续，且在 $q = \infty$的情况下均匀连续。若对应的 *Nemytskii*算子是一个映射 $\sigma: L^p(\Omega) \to L^q(\Omega)$，则它是连续的。

证明。对于 $q < \infty$，证明可由 [23, 第 155-158] 页改编。对于情形$q = \infty$，请参见 [6, 定理 5]。

备注 *1.* 对于 $1 \leq p \leq q < \infty$，只需使 σ 对 p满足 p/q-Hölder连续或局部Lipschitz连续，$q = \infty$。在这种情况下，Hölder连续性和Lipschitz连续性分别传递到 Nemytskii算子，见 [22]。

例 *1.* ReLU（修正线性单元，见 [5]）函数 $\sigma(x) = \max(0, x)$对任意 $p \geq q$都会生成连续的 Nemytskii算子 $\sigma: L^p(\Omega) \to L^q(\Omega)$ 。为证明这一点，我们注意到函数 σ 满足 Lipschitz连续性，并且借助 $p \geq q$，对所有 $x \in \mathbb{R}$，有 $|\sigma(x)| \leq |x| \leq 1 + |x|^{\frac{p}{q}}$.

命题 1. 对于 $1 \leq p, q \leq \infty$ 设 $\mathcal{L}$ 是由以下形式给出的算子层 (1)，其激活函数为 $\sigma: \mathbb{R} \to \mathbb{R}$。若存在 $r \geq 1$ 使得

(i) 仿射部分定义映射 $\Psi: L^p(\Omega) \to L^r(\Omega)$, *(ii)* 激活函数 σ 生成 *Nemytskii* 算子 $\sigma: L^r(\Omega) \to L^q(\Omega)$,

则有 $\mathcal{L}: L^p(\Omega) \to L^q(\Omega)$。此外，若 Ψ 在指定空间上是连续算子，且函数 σ 连续，或者在 $q = \infty$的情形下是一致连续的，则算子 $\mathcal{L}: L^p(\Omega) \to L^q(\Omega)$ 也是连续的。

证明。 根据关于 σ的假设，我们直接得到 $\mathcal{L} = \sigma \circ \Psi: L^p(\Omega) \to L^q(\Omega)$。 $\mathcal{L}$ 的连续性由引理2得出。

例 2. 在周期域 $\Omega = \mathbb{R}/\mathbb{Z}$ 上，考虑由式 (2) 定义的仿射算子 Ψ，其中积分算子是卷积算子，即 $\kappa(x, y) = \kappa(x - y)$，这里略微滥用了记号。若对于 $1 \leq p, r, s \leq \infty$，有 $\kappa \in L^s(\Omega)$，且 $1/r + 1 = 1/p + 1/s$，则由 Young卷积不等式（参见例如 [8, 定理 1.2.12]）可知， $\mathcal{K}: L^p(\Omega) \to L^r(\Omega)$ 是连续的。若进一步有 $b \in L^r(\Omega)$, 且在 $r > p$ 的情形下 $W = 0$，则可直接得出 $\Psi: L^p(\Omega) \to L^r(\Omega)$ 是连续的。

2.2 可微性

为分析神经算子层的可微性，我们首先将一般 Nemytskii算子的结果从 [6,定理 7] 转用于我们的情形。

定理 *1.* 设 $1 \leq q < p < \infty$ 或 $q = p = \infty$ 和 $\sigma: \mathbb{R} \to \mathbb{R}$ 是一个连续可微函数。此外，设与导数 σ' 相关联的 *Nemytskii*算子是一个连续算子 $\sigma': L^p(\Omega) \to L^s(\Omega)$，其系数为$s = pq/(p - q)$，适用于 $q < p$，并且 $s = \infty$ 适用于 $q = p = \infty$。那么，与 σ 相关联的 *Nemytskii*算子是弗雷歇可微的，其弗雷歇导数$D\sigma(v): L^p(\Omega) \to L^q(\Omega)$ 在 $v \in L^p(\Omega)$ 中定义为

$$D\sigma(v)(h) = \sigma'(v) \cdot h, \qquad for\ all\ h \in L^p(\Omega).$$

由于示例中的 ReLU激活函数1 不可微，因此不满足定理1的要求。一种替代方案是所谓的高斯误差线性单元（GELU），由 提出。

例 *3.* GELU 函数 $\sigma(x) = x\Phi(x)$（其中 Φ 表示标准正态分布的累积分布函数）生成一个 Fréchet 可微的 Nemytskii算子，其导数为 $D\sigma(v): L^p(\Omega) \to L^q(\Omega)$, 适用于任意 $p \geq q$ 和 $v \in L^p(\Omega)$。为证明这一点，我们计算 $\sigma'(x) = \Phi(x) + x\phi(x)$，其中$\phi(x) = \Phi'(x)$是标准正态分布。我们看到， σ' 是连续的，并且对所有 $p \geq q$都有 $|\sigma'(x)| \leq 1 + |x|/\sqrt{2\pi} \leq 1 + 1/\sqrt{2\pi} + |x|^{\frac{p}{q}}/\sqrt{2\pi}$。

命题 2. 对于 $1 \le p$， $q \le \infty$，设 $\mathcal{L}$ 为由 (1) 给出的算子层，其仿射部分为 Ψ，如 (2)所示。若存在 $r > q$，或 $r = q = \infty$，使得

(i) 仿射部分是连续算子 Ψ： $L^p(\Omega) \to L^r(\Omega)$，*(ii)* 激活函数 σ： $\mathbb{R} \to \mathbb{R}$ 是连续可微的，*(iii)* 激活函数的导数生成一个 *Nemytskii*算子σ'： $L^r(\Omega) \to [L^r(\Omega) \to L^s(\Omega)]$，其中 $s = rq/(r-q)$，

那么，对于任意 $v \in L^p(\Omega)$， $\mathcal{L}$： $L^p(\Omega) \to L^q(\Omega)$ 均为弗雷歇可微，其弗雷歇导数为 $D\mathcal{L}(v)$： $L^p(\Omega) \to L^q(\Omega)$

$$D\mathcal{L}(v)(h) = \sigma'(\Psi(v)) \cdot \tilde{\Psi}(h),$$

其中， $\tilde{\Psi}$表示 Ψ的线性部分，即 $\tilde{\Psi} = \Psi - b$。

证明。由定理 1可知， $D\sigma(v)$： $L^r(\Omega) \to L^q(\Omega)$ 定义良好且在 $v \in L^r(\Omega)$ 上连续。根据巴拿赫空间上线性连续算子的弗雷歇可微性（例如，参见 [1, 例 1.3]），可知 $\tilde{\Psi}$： $L^p(\Omega) \to L^r(\Omega)$ 在所有满足 $D\tilde{\Psi}(v)(h) = \tilde{\Psi}(h)$ 的 $v \in L^p(\Omega)$ 上连续。根据弗雷歇可微算子的链式法则即可得出结论，参见 [1,命题 1.4 (ii)]。

对于 $p < q$ Nemytskii算子的弗雷歇可微性意味着其生成函数为常数；对于 $p = q < \infty$，则意味着其为仿射线性函数，参见[6, 第 3 章1]。因此，除非 $p = \infty$，具有非仿射线性激活函数的神经算子若要实现弗雷歇可微，就只能将仿射部分的输出映射到正则性较低的空间。

示例 *4.* 对于示例 2中构造的连续卷积神经算子层，我们考虑将核函数参数化为一组$\hat{}$ $\hat{}$参数 $\theta = \{\theta_k\}_{k \in I} \subset \mathbb{C}$，其中 I 是一个有限指标集，满足

$$\kappa_{\hat{\theta}}(x) = \sum_{k \in I} \hat{\theta}_k \, b_k(x), \tag{5}$$

其傅里叶基函数为 $b_k(x) = \exp(2\pi i \, kx)$，其中 $x \in \Omega$。实际上，这相当于用有限个傅里叶系数对核函数进行参数化。所得线性算子和算子层分别记为$\Psi_{\hat{\theta}}$ 和 $\mathcal{L}_{\hat{\theta}}$。我们注意到， [15] 提出的傅里叶神经算子由此类算子层构成。容易看出，由 (5) 定义的核函数有界，因此 $\kappa_{\hat{\theta}} \in L^\infty(\Omega)$。于是，对于适当的激活函数，命题 2 推出 $\mathcal{L}_{\hat{\theta}}$ 关于其输入函数 v是弗雷歇可微的，这一点在 [16]中也有类似观察。此外，对于固定的 v，我们考虑算子 $\hat{\mathcal{L}}_{(\cdot)}(v)$： $\theta \mapsto \mathcal{L}_{\hat{\theta}}(v)$，它将一组参数映射为一个函数。利用命题 2 中的论证，我们推导出傅里叶神经算子层关于其参数的偏弗雷歇导数，其中 $h_k = (1+i)\, e_k$ 为

$$D_{\hat{\theta}_k} \mathcal{L}_{\hat{\theta}}(v) := D\mathcal{L}_{\hat{\theta}}(v)(h_k) = \sigma'(\Psi_{\hat{\theta}}(v)) \, D\Psi_{\hat{\theta}}(v)(h_k),$$

其中， e_k 表示第 k个典范基向量。按照维廷格微积分的意义计算 Ψ 的弗雷歇导数（[20,第 1]章），可将其改写为$D_{\hat{\theta}_k} \mathcal{L}_{\hat{\theta}}(v) = \sigma'(\Psi_{\hat{\theta}}(v)) \hat{} v_k \, b_k$，其中$\hat{} v_k$ 表示 v 的第 k个傅里叶系数。这里，对于复数 $z \in \mathbb{C}$，我们用 $\bar{z}$ 表示其复共轭。

3 与卷积神经网络的联系

本节分析傅里叶神经算子与卷积神经网络之间的联系。因此，在本文余下部分中，我们将定义域设为 d维环面，即$\Omega = \mathbb{R}^d/\mathbb{Z}^d$。如例4所述，傅里叶神经算子的核心思想是用有限个傅里叶系数对ˆ ˆ卷积核进行参数化 $\theta = \{\theta_k\}_{k\in I} \subset \mathbb{C}$, 其中 $I \subset \mathbb{Z}^d$ 是一个有限指标集。利用卷积定理（例如参见 [8, 命题 3.1.2（9）], ，核积分算子可写为

$$\mathcal{K}_{\hat{\theta}} v = \mathcal{F}^{-1}\left(\hat{\theta} \cdot \mathcal{F} v\right), \tag{6}$$

其中， $\mathcal{F}\colon\ [\Omega \to \mathbb{C}] \to \left[\mathbb{Z}^d \to \mathbb{C}\right]$ 表示环面上的傅里叶变换（例如参见 [8, 第 3] 章），而 $\cdot$ 表示逐元素乘法，具体而言：

$$\left(\hat{\theta} \cdot \mathcal{F} v\right)_k = \begin{cases} \hat{\theta}_k \ (\mathcal{F} v)_k & \text{for } k \in I, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

我们只考虑满足以下条件的参数： $\mathcal{K}$ 将实值函数映射为实值函数。这等价于厄米对称性，即 $\theta_k = \theta_{-k}$ ，特别是 $\theta_0 \in \mathbb{R}$。按照 [15], 针对 $N \in \mathbb{N}$ 提出的方法，我们选择多重指标集合 $I_N\colon = \{-\lceil (N-1)/2 \rceil,\ \ldots,\ 0,\ \ldots,\ \lfloor (N-1)/2 \rfloor\}^d$，这对应于对每个维度中最低的 N 个频率进行参数化。这与 [12]中推导出的傅里叶神经算子的通用逼近结果一致。此处，我们假设 N 为奇数，以避免所需对称性方面的问题；并在第 3.3 节中将此方法扩展到 N 为偶数的情况。尽管傅里叶神经算子由有限个参数表示，但要处理离散数据（例如数字图像），仍需对式（6）进行离散化。因此，我们定义空间多重指标集合 $J_N\colon = \{0, \ldots, N-1\}^d$，并将映射$v\colon\ J_N \to \mathbb{R}$记为 $v \in \mathbb{R}^{J_N}$。此外，我们对 $v \in \mathbb{R}^{J_N}$ 的傅里叶变换进行离散化，得到：

$$(Fv)_k = \frac{1}{\lambda} \sum_{j \in J_N} v_j \, e^{-2\pi i \left\langle k, \frac{j}{N} \right\rangle} \qquad \text{for all } k \in I_N$$

以及其逆变换为 $\hat{v} \in \mathbb{C}^{I_N}$

$$\left(F^{-1}\hat{v}\right)_j = \frac{\lambda}{|J_N|} \sum_{k \in I_N} \hat{v}_k \, e^{2\pi i \left\langle k, \frac{j}{N} \right\rangle} \qquad \text{for all } j \in J_N,$$

其中 $\lambda \in \{1, \sqrt{|J_N|}, |J_N|\}$决定归一化因子。由 $\hat{\theta} \in \mathbb{C}^{I_N}{}_{\text{对称}} := F(\mathbb{R}^{J_N})$ 参数化的离散卷积算子定义为

$$K(\hat{\theta})(v) = F^{-1}\left(\hat{\theta} \cdot Fv\right) \qquad \text{for } v \in \mathbb{R}^{J_N}.$$

在本文余下部分，我们将上述卷积实现称为 FNO 实现。下文中，我们将 FNO 实现与卷积的标准实现进行比较，后者作用于 θ 和

$v \in \mathbb{R}^{J_N}$，并可在传统卷积神经网络中表示为

$$C(\theta)(v)_j = \sum_{\tilde{\jmath} \in J_N} \theta_{j-\tilde{\jmath}} v_{\tilde{\jmath}} \qquad \text{for all } j \in J_N.$$

为简单起见，我们假设这些值可以周期性延拓，以此处理负索引，尽管在实践中通常不会这样做。

3.1 通过零填充扩展到更高输入维度

迄今为止，所介绍的卷积实现要求参数的维度$\hat{}$ θ与 θ 以及输入 v 的维度相同。根据（7）， [15]的作者提出通过对谱参数进行零填充来处理维度不匹配。更确切地说，将低维参数集$\hat{}\theta \in \mathbb{C}^{I_M}$适配到具有奇数 $v \in \mathbb{R}^{J_N}$ 的输入 $N \in \mathbb{N}$，方法是令

$$\hat{\theta}_k^{M \to N} = \begin{cases} \hat{\theta}_k & \text{for } k \in I_N \cap I_M, \\ 0 & \text{for } k \in I_N \backslash I_M. \end{cases}$$

由于我们选择 N 为奇数，上述操作不会破坏所需的对称性。于是，卷积的扩展 FNO 实现对于 $\hat{}\theta \in \mathbb{C}^{I_M}$ 和 $v \in \mathbb{R}^{J_N}$ 可表示为

$$K(\hat{\theta})(v) := K(\hat{\theta}^{M \to N})(v).$$

类似地，在传统的卷积神经网络实现中，参数 $\theta \in \mathbb{R}^{J_M}$ 和 $v \in \mathbb{R}^{J_N}$ 与 $N \geq M$ 的卷积计算为

$$C(\theta)(v) := C(\theta^{M \to N})(v),$$

其中再次， $\theta^{M \to N} \in \mathbb{R}^{J_M}$ 表示 θ的零填充版本。我们强调，尽管将这些实现推广到更高输入维度时所用的技术相同，但得到的结果却有显著差异。这一点已在 [13, 第 4]节中提及，并将在第3.4节中进一步讨论。

3.2 可转换性与复杂度

利用卷积定理从卷积神经算子推导傅里叶神经算子，表明只要输入维度固定，就可以将一种卷积实现转换为另一种。下面的引理证明了这一点确实可行。

引理3。设 $M \leq N$ 均为奇数，且 $T: \mathbb{R}^{J_N} \to \mathbb{C}^{I_N}$ 对$\theta \in \mathbb{R}^{J_N}$定义为 $T(\theta) = \lambda F(\theta)$。对于任意 $\theta \in \mathbb{R}^{J_M}$和 $v \in \mathbb{R}^{J_N}$，均有

$$C(\theta)(v) = K(T(\theta^{M \to N}))(v)$$

且对于任意 $\hat{}\theta \in \mathbb{C}^{I_M}{}_{\text{对称}}$ 以及 $v \in \mathbb{R}^{J_N}$，该关系成立

$$K(\hat{\theta})(v) = C(T^{-1}(\hat{\theta}^{M \to N}))(v).$$

证明。 根据向更高输入维度扩展的定义，我们可以假设 $M = N$。对于 θ, $v \in \mathbb{R}^{J_N}$，将离散傅里叶变换的定义代入，可得卷积定理的离散形式:

$$F(C(\theta)(v)) = \lambda\, F(\theta) \cdot F(v).$$

利用 $F \colon \mathbb{R}^{J_N} \to \mathbb{C}^{I_N\text{对称}}$ 是一个双射，可得

$$C(\theta)(v) = F^{-1}(\lambda\, F(\theta)\, F(v)) = K(T(\theta))(v).$$

第二个结论可以类似地证明。注意，对于奇数 N， T^{-1} 是良定义的，因为 $\lambda \geq 1$ 。

尽管上述引理证明了在固定参数集和固定输入维度下的可转换性，但转换可能会增加所需参数的数量，因为通常转换后参数的维度必须与输入维度相匹配。显然，采用所提出的傅里叶神经算子参数化无法强制实现空间局部性，而采用卷积神经网络参数化则无法强制实现谱局部性。因此，如果参数规模与输入规模不匹配，训练过程中的行为预计会有所不同。此外，以下引理表明，即使维度匹配，要使基于梯度的优化（如最速下降法）表现等效，也需要仔细调整学习率，因为梯度的计算并不具有关于函数 T 的等变性。

引理4。 对于奇数 $N \in \mathbb{N}$ 和 v， $\theta \in \mathbb{R}^{J_N}$ 和 $\hat{}\theta = T(\theta)$，有

$$\nabla_{\hat{\theta}} K(\hat{\theta})(v) = \frac{1}{|J_N|}\, T\left(\nabla_\theta C(\theta)(v)\right).$$

$\hat{}$证明。 代入 $\theta = T(\theta)$ 根据引理 3，由链式法则可得

$$\frac{\partial K(\hat{\theta})(v)_l}{\partial \hat{\theta}_k} = \sum_{j \in J_N} \frac{\partial C(\theta)(v)_l}{\partial \theta_j} \frac{\partial T^{-1}(\hat{\theta})_j}{\partial \hat{\theta}_k} = \frac{1}{|J_N|} \sum_{j \in J_N} \frac{\partial C(\theta)(v)_l}{\partial \theta_j}\, e^{-i2\pi\langle k, \frac{j}{N}\rangle}$$

对于 $k \in I_N$。现在代入 T 的定义即可得到结论。

3.3 偶数维度的适配

本文余下部分考虑特殊情况 $\Omega = \mathbb{R}^2/\mathbb{Z}^2$ ，并将 FNO 实现调整为偶数维度。对于奇数维度 M， N，对一组谱系数进行零填充不会违反条件 $\hat{}\theta^{M\to N} \in \mathbb{C}^{I_{N_{\text{s}}}}$ ym。对于偶数维度，这一性质通常不再成立。由于在奇数维度下，谱域中的零填充等价于三角插值，因此我们对维度进行调整，使 $\hat{}_\theta{}^{M\to N}$成为实值函数的三角插值器（有关此主题的详尽研究，请参见 [2]）。实际操作中，这意味着拆分对应于奈奎斯特频率的系数，以便从偶数维度插值到

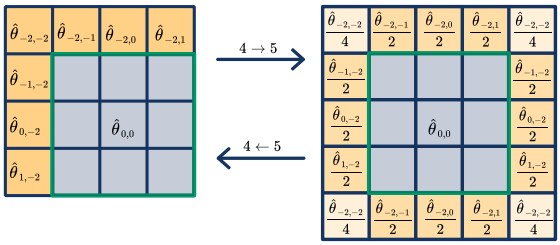


图 2：对谱参数进行奈奎斯特拆分，以将实值三角插值扩展到偶数维度。

下一个更高的奇数维度；或者反向执行此拆分，以便从奇数维度插值到下一个更低的偶数维度（见图 2）。将 $v \in \mathbb{R}^{J_M}$ 实值三角插值到维度 N 的结果为

$$v^{M \xrightarrow{\Delta} N} := F^{-1}\left((Fv)^{M \to N}\right).$$

我们通过定义，将 FNO 实现扩展到参数 $\hat{\theta} \in \mathbb{C}^{I_N}$对称 以及具有偶数 $N \in \mathbb{N}$ 的输入 $v \in \mathbb{R}^{J_N}$

$$K(\hat{\theta})(v) := \left(K(\hat{\theta}^{N \to \tilde{N}})(v^{N \xrightarrow{\Delta} \tilde{N}})\right)^{\tilde{N} \xrightarrow{\Delta} N}, \tag{8}$$

其中 $\tilde{N} = N + 1$。我们注意到，采用这一选择后，我们将无法直接转换为卷积神经网络实现，因为通常对于偶数维度

$$K(\hat{\theta})(v) \neq F^{-1}(T^{-1}(\hat{\theta})\, Fv),$$

右侧相当于对谱系数进行零填充，而不论维度是奇是偶。不过，我们仍然可以通过调整系数的幅值以适应奈奎斯特拆分的影响，在 FNO 实现与卷积神经网络实现之间进行相互转换。

3.4 插值等变性

我们按照上一节中的提议将其适配到偶数维度，动机在于，由此得到的卷积实现对于（实值）三角插值具有等变性。

推论 1. 对于 $\hat{\theta} \in \mathbb{C}^{I_M}$对称的 $v \in \mathbb{R}^{J_N}, M \leq N$，对任意 $L \geq M$ 均成立

$$K(\hat{\theta})(v^{N \xrightarrow{\Delta} L}) = \left(K(\hat{\theta})(v)\right)^{N \xrightarrow{\Delta} L}.$$

证明. 我们首先注意到，对于任意选择的 $M \leq N, L$ ，以下结论均成立

$$K(\hat{\theta})(v^{N \xrightarrow{\Delta} L}) = \left(K(\hat{\theta}^{M \to \tilde{M} \to \tilde{L}})(v^{N \xrightarrow{\Delta} \tilde{N} \xrightarrow{\Delta} \tilde{L}}) \right)^{\tilde{L} \xrightarrow{\Delta} L},$$

其中 $\tilde{L} = L + (1 - L\%2)$、 $\tilde{M} = M + (1 - M\%2)$、 $\tilde{N} = N + (1 - N\%2)$，% 表示取模运算。因此，不失一般性，我们可以假设 L、M 和 N 均为奇数，从而有 $\tilde{L} = L$、 $\tilde{M} = M$ 和 $\tilde{N} = N$。考察离散傅里叶系数可得

$$F(K(\hat{\theta})(v^{N \xrightarrow{\Delta} L}))_k = \begin{Bmatrix} \hat{\theta}_k \, F(v)_k & \text{for } k \in I_M, \\ 0 & \text{otherwise} \end{Bmatrix} = F((K(\hat{\theta})(v))^{N \xrightarrow{\Delta} L})_k.$$

应用傅里叶逆变换即可完成证明。

4 数值示例

本节将在图像分类背景下，对所讨论的卷积实现进行数值比较。[5] 此处，任务是为给定图像 $s \in \mathbb{N}$ 分配一个类别（共 $v : [0,1]^2 \to \mathbb{R}^{n_c}$），其中 $n_c \in \mathbb{N}$ 表示颜色通道数。数值求解该任务需要形如 $v^N = v|_{J_N/N} \in \mathbb{R}^{J_N \times n_c}$ 的离散输入图像，其中 $N \in \mathbb{N}$ 表示维度。由于我们考虑的是固定函数域，因此维度与分辨率成正比。若假设 N 固定，则网络是一个函数 f_θ： $\mathbb{R}^{J_N \times n_c} \to \mathbb{R}^s$。给定有限训练集 $D \subset \mathbb{R}^{J_N \times nc} \times \mathbb{R}^s$，我们通过最小化基于交叉熵的经验损失 θ 来优化参数 [7,第 3]。实验所用的网络由若干用于特征提取的卷积层和一个全连接分类层组成。为使所有架构均可处理任意分辨率的输入，我们在特征提取器与分类器之间插入一个自适应平均池化层。

4.1 不同卷积核大小下的表达能力

在第一个实验中（见图 3），我们训练了一个不含任何残差组件的卷积神经网络，在 FashionMNIST[6] 数据集上进行实验。该网络包含两个采用周期填充且不进行步幅采样的卷积层，后接一个自适应池化层和一个线性分类器。由于在测试集上使用不同卷积核大小时，性能并未出现明显变化，我们因此认为，在这个数据集上，小卷积核架构的表达能力与大卷积核架构相当。随后，我们将卷积神经网络中采用 3×3-28×28 卷积核和卷积核的卷积层替换为傅里叶神经算子层，并采用不同数量的频谱参数。这里我们观察到，频谱卷积核越小，性能越差，

[5] 我们的代码已在网上公开：github.com/samirak98/FourierImaging。[6] 该数据集包含 60，000 个训练样本和 10，000 个测试 28×28 图像（灰度图）。

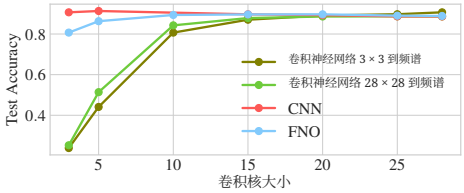


图 3：不同卷积核大小下卷积神经网络和傅里叶神经算子的测试准确率。

这表明，学习到的空间卷积核无法用较少的频率很好地表示。不过，在这个例子中，采用相同结构训练傅里叶神经算子，几乎弥合了这一性能差距。这意味着存在表达能力足够的低频卷积核。关于采用频谱参数化进行训练的研究，请参见 [11]。

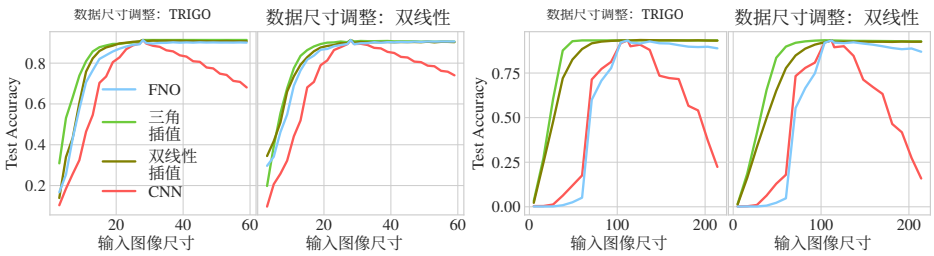
4.2 分辨率不变性

在第二个实验中，我们研究了不同卷积实现的分辨率不变性。在图 4a中，我们分别比较了采用三角插值或双线性插值将测试数据调整至不同分辨率后的准确率。这里，CNN 指采用 5×5 卷积核的传统卷积神经网络实现，其中通过对卷积核进行空间零填充来补偿维度不匹配。FNO 指 FNO 实现，其中通过三角插值使卷积核适应输入维度。此外，我们还展示了将输入重新缩放至训练分辨率时 CNN 的表现。在卷积层之前应用三角插值，可以视为输出维度预先确定的傅里叶神经算子层。

卷积神经网络的性能会随输入维度发生显著变化，并在其训练所用的分辨率下达到峰值。这一结果与图 1所展示的效应一致：通过空间零填充调整维度会改变卷积核的局部性，因此在不同分辨率下提取到不同的特征。尽管三角插值的表现最佳，我们发现傅里叶神经算子具有很强的适应能力。尤其是输入分辨率较高时，其性能仅略有下降，而标准卷积神经网络并非如此。

此外（见图 4b），我们在 Birds500 18 [9] 数据集上训练了一个 ResNet[7]，训练尺寸缩减为 112×112。为了使模型能够泛化到不同分辨率，尤其是针对 FNO 实现，我们用三角下采样替代标准步幅操作。与第一个实验相比，值得注意的是，输入分辨率低于 112×112时，FNO 的表现较差，但在更高分辨率下性能仅略有下降。我们将这一现象归因于架构中的降维操作。

[7] 我们使用该数据集的早期版本，其中包含 76,262 张 RGB图像用于训练，以及 2,250 张尺寸为 224×224 的图像用于测试；任务是从 450 个可能的类别中对鸟类进行分类。



(a) FMNIST (b) Birds500

图 4：在使用图上方标注的插值方法调整尺寸后的测试数据上，不同插值方法的性能。

5 结论与展望

在本研究中，我们探讨了勒贝格空间上神经算子的正则性，并研究了按照傅里叶神经算子的方式实现卷积层所产生的影响。基于从标准卷积神经网络到傅里叶神经算子的可转换性的理论推导，我们的数值实验表明，可以将采用标准卷积神经网络架构训练的网络转换为傅里叶神经算子。由此，我们能够结合两种方法的优势：训练期间强制实现空间局部性，同时使用较少的参数；评估期间采用一种能够很好地泛化到更高输入维度的实现方式。然而，我们发现，输入的三角插值优于所有其他受考察的方法。未来，我们希望研究如何将傅里叶神经算子和三角插值的思想融入 U-Net 等图像到图像架构，正如 [21]中所提出的那样。此外，我们还希望进一步探索在谱域中训练的影响，例如其对抗鲁棒性方面的影响。

参考文献

1. Ambrosetti, A., Prodi, G.：《非线性分析入门》。剑桥大学出版社（1993） 2. Briand, T.：图像的三角多项式插值。《在线图像处理》**9**，291–316（2019年10月） 3. Cai, D.、陈 K.、Qian, Y.、Kämäräinen, J.K.：卷积低分辨率细粒度分类。《模式识别快报》**119**，166–171（2019） 4. Chi, L.、Jiang, B.、Mu, Y.：快速傅里叶卷积。《神经信息处理系统进展》**33**，4479–4488（2020） 5. Fukushima, K.C.：认知机：一种自组织多层神经网络。《生物控制论》**20**，121–136（1975） 6. Goldberg, H.、Kampowsky, W.、Tröltzsch, F.：抽象函数 Lp 空间中的 Nemytskij 算子。《数学新闻》**155(1)**，127–140（1992） 7. Goodfellow, I.、Bengio, Y.、Courville, A.：《深度学习》。麻省理工学院出版社（2016）

8. Grafakos, L.：《经典傅里叶分析》。数学研究生教材，Springer，纽约州纽约市，第3版（2014） 9. 何 K.、张 X.、任 S.、孙 J.：用于图像识别的深度残差学习。载于：IEEE计算机视觉与模式识别会议论文集。第770–778页（2016） 10. Hendrycks, D.、Gimpel, K.：高斯误差线性单元（高斯误差线性单元）。arXiv:1606.08415（2016） 11. Johnny, W.、Brigido, H.、Ladeira, M.、Souza, J.C.F.：用于图像分类的傅里叶神经算子。载于：2022年第17届伊比利亚信息系统与技术会议（CISTI）。第1–6页（2022） 12. Kovachki, N.B.、Lanthaler, S.、Mishra, S.：关于傅里叶神经算子的通用逼近与误差界。《机器学习研究期刊》（2022） 13. Kovachki, N.B.、李 Z.、刘 B.、Azizzadenesheli, K.、Bhattacharya, K.、Stuart, A.M.、Anandkumar, A.：神经算子：学习函数空间之间的映射。arXiv:2108.08481（2021） 14. Koziarski, M.、Cyganek, B.：低分辨率对深度神经网络图像识别的影响：一项实验研究。《应用数学与计算机科学国际期刊》**28**(4)，735–744（2018） 15. 李 Z.、Kovachki, N.B.、Azizzadenesheli, K.、刘 B.、Bhattacharya, K.、Stuart, A.M.、Anandkumar, A.：用于参数化偏微分方程的傅里叶神经算子。载于：第9届国际学习表征会议（ICLR）（2021） 16. 李 Z.、郑 H.、Kovachki, N.、Jin, D.、陈 H.、刘 B.、Azizzadenesheli, K.、Anandkumar, A.：用于学习偏微分方程的物理信息神经算子。arXiv预印本 arXiv:2111.03794（2021） 17. Peng, X.、Hoffman, J.、Stella, X.Y.、Saenko, K.：低分辨率图像分类中的由细到粗的知识迁移。载于：2016 IEEE国际图像处理会议（ICIP）。第3683–3687页。IEEE（2016） 18. Piosenka, G.：Birds 500——物种图像分类（2021），https://www.kaggle.com/datasets/gpiosenka/100-bird-species 19. Rao, Y.、赵 W.、朱 Z.、陆 J.、周 J.：用于图像分类的全局滤波网络。《神经信息处理系统进展》**34**，980–993（2021） 20. Remmert, R.：《复函数理论》。Springer，纽约，纽约州纽约市（1991） 21. Ronneberger, O.、Fischer, P.、Brox, T.：U-Net：用于生物医学图像分割的卷积网络。载于：Navab, N.、Hornegger, J.、Wells, W.M.、Frangi, A.F.（编）：医学图像计算与计算机辅助干预——MICCAI 2015。第234–241页。Springer国际出版，Cham（2015） 22. Tröltzsch, F.：《偏微分方程最优控制：理论、方法与应用》，数学研究生学习丛书，第112卷。美国数学学会，普罗维登斯，罗得岛州（2010） 23. Vaˇınberg, M.M.：《非线性方程理论中的变分法与单调算子法》。第22090号，John Wiley & Sons（1974） 24. 肖 H.、Rasul, K.、Vollgraf, R.：Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集。arXiv:1708.07747（2017） 25. 周 M.、于 H.、黄 J.、赵 F.、顾 J.、Loy, C.C.、孟 D.、李 C.：深度傅里叶上采样。arxiv:2210.05171（2022）