

证券研究报告 / 金融工程研究报告

基于 CNN-Transformer 的深度学习模型探究

---人工智能系列之二

报告摘要：

金融数据的特点是高维度、高频率和非线性，这使得传统分析方法往往束手无策，难以捕捉其中的深层次模式和关系。CNN 以其卓越的特征提取能力，特别适合于处理图像和时间序列数据；而 Transformer，凭借其独特的自注意力机制，能够有效处理长距离序列依赖问题。这两种模型的结合，为金融市场分析提供了一种全新的视角和方法。

本篇报告在第二章详细介绍了 CNN 和 Transformer 的基本原理、关键特性及其在金融数据分析中的应用。通过研究这些深度学习技术方法和原理，总结它们在各个应用领域取得成功的原因，总结不同领域之间的共性，以探索它们在金融时间序列数据分析中的潜在价值。

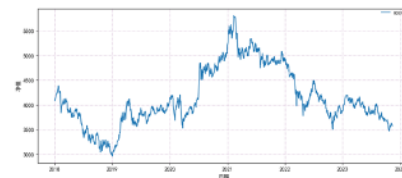
在报告第三章介绍了基于 CNN 和 Transformer 的时间序列模型（CTTS）的模型构造和训练方案以及实验配置和结果。CTTS 将日内股价数据序列通过一维卷积以及多层 Transformer，最后经过 MLP 得出“上涨”、“下跌”、“持平”的预测概率。在原文实验的所有统计中，CTTS 的预测准确率都高于三种基准策略，这表明了结合 CNN 和 Transformer 用于时间序列预测的有效性。

最后总结 CNN 与 Transformer 两种模型的优势与特点，结合 CTTS 模型的启发，提出了将 CNN 与 Transformer 结合应用于量化分析的研究方案：

- 采用类 CTTS 的模型结构进行基于分钟频率的高频数据选股因子挖掘。从当前的研究情况来看，采用类似的结构基于日频数据进行选股因子挖掘的效果并不理想。要想增加模型的拟合能力，需要提升自注意力头数以及 Transformer block 层数，相应的也需要更大规模的训练数据。
- 与 CTTS 相同的，尝试根据预测涨跌平分类训练模型，挖掘有效的股票择时策略。
- 根据两种模型不同的特性，将两种模型在不同频率的数据上进行训练，得到选股因子并合成。

风险提示：以上分析基于模型结果和历史测算，存在模型失效风险。

沪深 300 走势图



相关报告

《上月红利、Beta、价值因子表现较优》

--20240203

《可转债风险模型构建与应用》

--20240129

《雪球产品敲入规模分布估算和市场影响点评》

--20240124

《债基回报改善，久期杠杆回升》

--20240123

《2023 年四季度公募主动权益基金持仓解析》

--20240123

证券分析师：王琦

执业证书编号：S0550521100001

021-61002390 wangqi_5636@nesc.cn

目 录

1.	引言.....	3
2.	卷积神经网络-CNN 与 Transformer	4
2.1.	卷积神经网络-CNN.....	4
2.1.1.	卷积神经网络的基本架构.....	5
2.1.2.	卷积神经网络的发展和应用.....	13
2.2.	Transformer 模型结构	17
2.2.1.	Transformer 简介.....	17
2.2.2.	Transformer 在金融数据分析领域的应用.....	19
3.	基于 CNN-Transformer 的时间序列模型	21
3.1.	基于 CNN 和 Transformer 的时间序列模型 (CTTS)	21
3.2.	CTTS 的训练设置.....	23
3.3.	CTTS 的实验结果与结论.....	24
4.	基于 CNN-Transformer 的模型探索	26
5.	参考文献	27
6.	风险提示	28

图表目录

图 1:	卷积神经网络的基本架构图	5
图 2:	卷积神经网络的卷积核计算示意图	7
图 3:	卷积神经网络的填充示意图	8
图 4:	多输入通道的卷积神经网络示意图	10
图 5:	多输出通道的卷积神经网络示意图	10
图 6:	卷积神经网络的池化层计算示意图	12
图 7:	卷积神经网络的池化层计算示意图	12
图 8:	transformer 架构	17
图 9:	GoogLeNet Inception 块结构与 Transformer 多头注意力机制对比图	18
图 10:	残差连接概念示意图	18
图 11:	基于 CNN-Transformer 的模型示意图	22
表 1:	CTTS 模型配置与超参数.....	24
表 2:	CTTS 模型及其三种基准策略在测试集上的预测准确率	24

1. 引言

在金融市场分析领域，量价数据的深度解析一直是揭示市场运动规律和挖掘有效选股因子的关键。然而，面对海量的日频数据，传统的分析方法受限于其处理高维数据和识别复杂时间序列依赖关系的能力。随着深度学习技术的进步，特别是卷积神经网络（CNN）和 Transformer 模型的发展，为解决这些挑战提供了新的可能性。这两种技术已经在图像处理、自然语言处理等领域显示出了卓越的性能，但在金融数据分析上的应用仍有待深入探索。

本报告尝试探究了 CNN 和 Transformer 结合的深度学习模型，旨在更有效地分析股票日频量价数据。CNN 以其出色的特征提取能力，能够识别出时间序列数据中的局部模式和趋势，而 Transformer 则通过其自注意力机制，优异地处理长距离依赖问题。将两者结合，旨在充分利用 CNN 在时间序列分析上的局部特征提取优势和 Transformer 在捕捉全局依赖关系上的能力。这种结合不仅增强了模型对于复杂金融数据的理解能力，也提高了在挖掘深层次、非线性选股因子中的效率和准确性。

本报告尝试使用日频量价数据进行选股因子挖掘，模型的潜力远未被完全释放。后续的研究将考虑使用更高频率的分钟级数据，这可能挖掘出更为精细的市场动态和交易机会。更高频的数据能够提供更丰富的信息，有助于模型捕捉到更加微妙的市场变化规律，进而可能获得更好的预测效果和选股性能。

基于 CNN + Transformer 的深度学习模型不仅可用于选股因子挖掘，其应用潜力还远远不止于此。例如，在择时研究方面，通过更高频率的量价数据的学习，模型能够通过分析短期内的量价变化，进行买入或卖出时机的预测。

2. 卷积神经网络-CNN 与 Transformer

随着大数据时代的到来，深度学习技术在众多领域展现出了前所未有的潜力和成效，尤其是卷积神经网络（CNN）和 Transformer 这两种模型，在图像识别、自然语言处理等领域取得了革命性的进步。这些进展不仅推动了人工智能技术的发展，也为解决传统分析方法难以克服的复杂问题提供了新的思路。金融市场分析，特别是股票日频量价数据的挖掘，是这些技术潜能可以大显身手的领域之一。

金融数据的特点是高维度、高频率和非线性，这使得传统分析方法往往束手无策，难以捕捉其中的深层次模式和关系。CNN 以其卓越的特征提取能力，特别适合于处理图像和时间序列数据；而 Transformer，凭借其独特的自注意力机制，能够有效处理长距离序列依赖问题。这两种模型的结合，为金融市场分析提供了一种全新的视角和方法。

本章节详细介绍 CNN 和 Transformer 的基本原理、关键特性及其在金融数据分析中的应用。通过对这些深度学习技术的探讨，我们不仅能够更好地理解它们在其他领域取得成功的原因，也能够探索它们在金融市场分析中的潜在价值。我们将探讨为什么结合使用这两种模型能够为挖掘股票市场中的量价信息提供更加强大的工具。

2.1. 卷积神经网络-CNN

卷积神经网络（Convolutional Neural Networks, CNN）是一种深度学习模型，特别适合于处理具有网格结构的数据，如图像（2D 网格）和时间序列数据（1D 网格）。

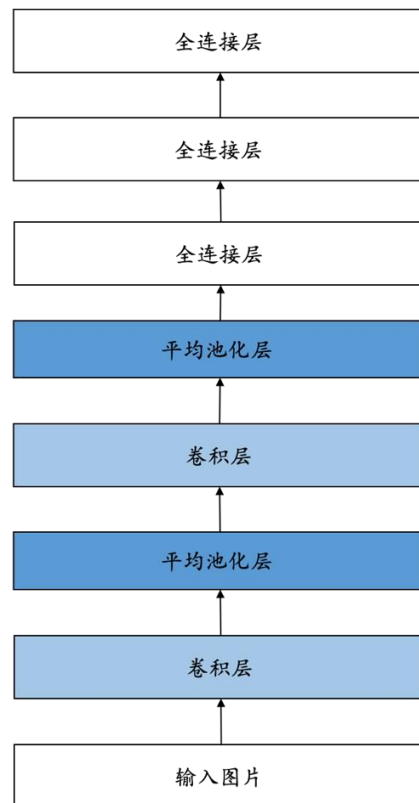
卷积神经网络是一类强大的、为处理图像数据而设计的神经网络。基于卷积神经网络架构的模型在计算机视觉领域中已经占主导地位，当今几乎所有的图像识别、目标检测或语义分割相关的学术竞赛和商业应用都以这种方法为基础。CNN 通过模拟人类视觉系统的工作机制来识别图像中的模式，如边缘、角点、纹理等。它利用卷积层来自动学习和提取特征，无需手动特征工程。这种学习过程通过在训练数据上反复应用卷积、激活和池化操作来完成。

卷积神经网络需要的参数少于全连接架构的网络，而且卷积也很容易用 GPU 并行计算。因此卷积神经网络除了能够高效地采样从而获得精确的模型，还能够高效地计算。久而久之，从业人员越来越多地使用卷积神经网络。即使在通常使用循环神经网络的一维序列结构任务上（例如音频、文本和时间序列分析），卷积神经网络也越来越受欢迎。自从在图像识别领域取得突破性进展以来，CNN 已经广泛应用于多个领域，包括视频分析、自然语言处理、医学图像分析等。下面是对 CNN 的基本原理、架构和关键特性的介绍。

2.1.1. 卷积神经网络的基本架构

一个卷积神经网络的典型架构包括以下几个主要部分：卷积层 (Convolutional Layer)、激活函数、池化层 (Pooling Layer, 也被称为汇聚层)、全连接层。每个部分都有其独特的功能和作用。这些组成部分协同工作, 使 CNN 能够有效地处理图像、视频、时间序列数据等。图 1 是一个简化的 CNN 模型, 也是最早发布的卷积神经网络模型 LeNet-5^[1]的组成结构。LeNet-5 包含两个卷积层, 2 个全连接层, 共计 6 万个学习参数, 在结构上与现代的卷积神经网络十分接近, 此外, LeNet-5 是首个采用“权重随机初始化+随机梯度下降 (Stochastic Gradient Descent, SGD)”方式进行训练的模型, 这一策略也被后续的深度学习方法沿用, 具有深远的影响。

图 1: 卷积神经网络的基本架构图



数据来源：东北证券、[1]

卷积层 (Convolutional Layer) 是 CNN 的基础, 负责从输入数据中提取特征。它通过卷积操作将多个小的、可学习的滤波器 (也被称为卷积核) 应用于输入数据。这些滤波器在空间维度 (对于图像数据) 或时间维度 (对于时间序列数据) 上移动, 计算滤波器与数据的局部区域之间的点积, 以生成特征图 (feature maps)。这一过程涉及多个关键概念, 包括卷积核、填充、步幅、通道等。下面详细介绍这些概念和原理, 以及卷积层的计算方法。

2.1.1.1. 卷积层

卷积神经网络的设计初衷是更好地处理自然图像信息。对于多层感知机来说，图像的输入是不带有显式的位置信息的，而是一个按像素点顺序排列的像素值列表。一张图片的像素级别越高，这个输入列表就越长，网络的每个输入的维度甚至能够达到百万、千万的级别。在这种输入维度下，仅仅需要个位数的隐藏层层数，以及不多的隐藏层单元，网络的参数量级就会膨胀到无法承受的地步。隐藏单元数量不足则不能够学习到良好的图像特征，高数量级的参数更需要收集海量的数据进行训练才能够拟合。

2.1.1.1.1. 卷积核 (Convolution Kernel)

CNN 的设计创造性地利用了自然图像的普遍特性，以较少的参数解决了图像处理中的许多问题。CNN 的设计包括两个特性：“平移不变性 (Translation Invariance)”和“局部性 (Locality)”。这两个特性也是“卷积”在神经网络中的表现。

假设神经网络的输入是二维图像 \mathbf{X} ， \mathbf{H} 则是该图像在神经网络中的隐藏表示，两者具有相同的维度。使用 $[\mathbf{X}]_{i,j}$ 和 $[\mathbf{H}]_{i,j}$ 表示图像和隐藏表示中位置 (i,j) 的像素，则对于多层感知机的全连接层，可以作如下的表示：

$$[\mathbf{H}]_{i,j} = [\mathbf{U}]_{i,j} + \sum_a \sum_b [\mathbf{V}]_{i,j,a,b} [\mathbf{X}]_{i+a,j+b}$$

其中，索引 a,b 是位置偏移，对隐藏表示中任意给定位置 (i,j) 处的像素值 $[\mathbf{H}]_{i,j}$ ，都可以通过在 \mathbf{X} 中以 (i,j) 为中心对像素进行加权求和得到， $[\mathbf{V}]_{i,j,a,b}$ 即为该加权求和使用的权重， $[\mathbf{U}]_{i,j}$ 为偏置参数。

而满足平移不变性的情况下，图像中的元素在输入 \mathbf{X} 中的平移将仅导致隐藏表示 \mathbf{H} 中的平移，即上式中加权矩阵仅与索引偏移 a,b 有关，和像素的位置 (i,j) 无关： $[\mathbf{V}]_{i,j,a,b} = [\mathbf{V}]_{a,b}$ ，偏置参数 $[\mathbf{U}]_{i,j}$ 也成为常数 u ，由此可以简化得到隐藏表示的公式为：

$$[\mathbf{H}]_{i,j} = u + \sum_a \sum_b [\mathbf{V}]_{a,b} [\mathbf{X}]_{i+a,j+b}$$

或许是因为该式与数学定义上的卷积形式上非常相似，这个使用权重系数 $[\mathbf{V}]_{a,b}$ 对位置 (i,j) 附近的像素 $(i+a,j+b)$ 进行加权得到 $[\mathbf{H}]_{i,j}$ 的过程被提出该方法的作者称为卷积 (Convolution)。相比于多层感知机的权重参数 $[\mathbf{V}]_{i,j,a,b}$ ，CNN 的权重参数 $[\mathbf{V}]_{a,b}$ 的系数减少了两个指数级。

进一步地，为了满足局部性，用来训练参数 $[\mathbf{H}]_{i,j}$ 的信息应该在位置 (i,j) 就近收集。用数学语言描述则是，当 $|a| > \Delta$ or $|b| > \Delta$ 时 $[\mathbf{V}]_{a,b} = 0$ 。那么 $[\mathbf{H}]_{i,j}$ 可以进一步地减少运算复杂度至下式：

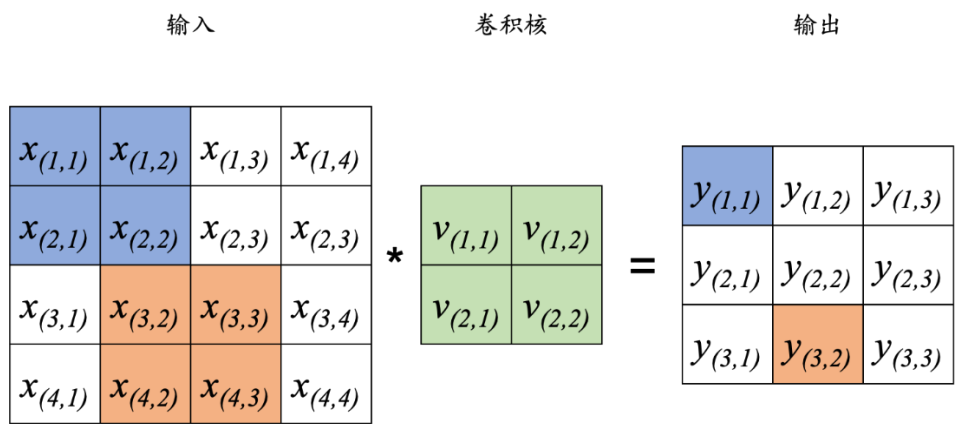
$$[H]_{i,j} = u + \sum_{-\Delta}^{\Delta} \sum_{-\Delta}^{\Delta} [V]_{a,b} [X]_{i+a,j+b}$$

上式即为一个卷积层 (Convolution Layer)。而该卷积层的权重 $[V]_{a,b}$ 则被称为卷积核 (Convolution Kernel) 或滤波器 (Filter)，是 CNN 中最重要的可学习参数。

相比于多层感知机的隐藏层，CNN 的卷积核的参数量级大幅度减少，并且不需要改变输入的形状。例如，一张百万像素级别的图像，一层包含 1000 个隐藏单元的感知机就需要十亿个参数，而一层卷积核的尺寸为 $[100 \times 100]$ 的卷积层，则只需要一万个参数，这在参数量级的尺度上是质的减少。

这种参数的大幅减少并不是没有代价的，而是将特征提取的尺度由图像或序列的全局缩小到了局部，每一层卷积层中都只包含局部的信息。这种代价可以用层次化的方法来改善，CNN 通过堆叠多个卷积和池化层，能够从浅层（距离输入端较近的卷积层）简单的边缘和纹理特征逐渐学习到深层（远输入端的卷积层）复杂的对象结构，实现从低级到高级特征的自动学习。

图 2：卷积神经网络的卷积核计算示意图



数据来源：东北证券

用数学语言和公式描述卷积过于抽象，卷积层中的卷积核的运行规律如图 2 所示。将二维的输入视作一张图像，则经过卷积的输出可以被视作由输入图像的局部特征拼接而成的“特征图”。在计算输出的过程中，卷积核会在输入图像上“移动”。每移动到一个位置，将被卷积核覆盖的输入图像的像素，以卷积核作为权重做加权求和得到一个特征值，即为特征图对应位置上的输出，如图 2 所示，输入的局部像素与卷积核加权求和得到的结果，在特征图上对应颜色相同的位置。特征图的尺寸由输入图像的尺寸与卷积核的尺寸共同决定，尺寸为 $[x_m \times x_n]$ 的输入图像经过卷积核尺寸为 $[k_m \times k_n]$ 的卷积层后，特征图的尺寸为：

$$(x_m - k_m + 1) \times (x_n - k_n + 1)$$

2.1.1.1.2. 步幅 (Stride) 和填充 (Padding)

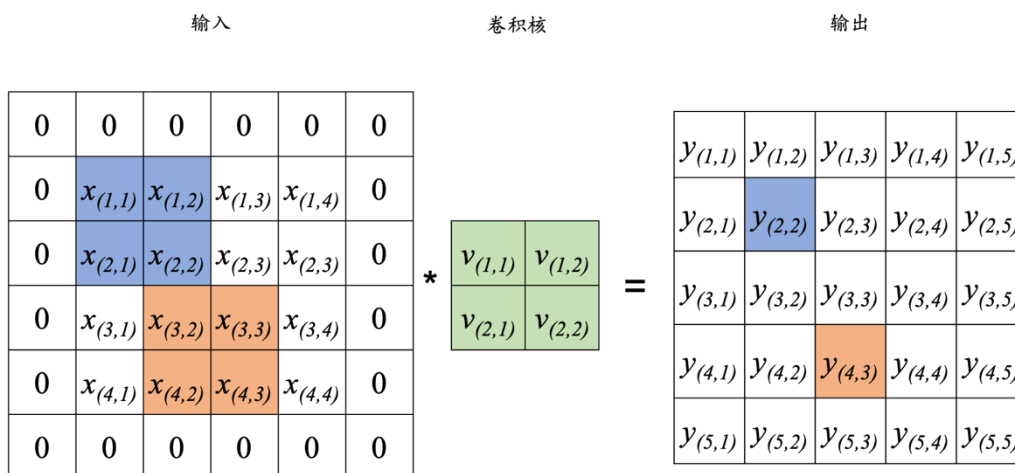
除了输入尺寸和卷积核尺寸之外，卷积层特征图的尺寸还受到步幅和填充的影响，下面介绍步幅与填充的概念。

当卷积核与输入图像的尺寸都比较大时，卷积核每次“移动”一格得到的特征图将包含较多的重复特征信息，造成计算资源的浪费，加大步幅可以避免这种问题。卷积核每次在输入数据上“移动”的距离即为步幅 (Stride)。步幅为 1 意味着卷积核每次“移动”一个像素，步幅为 2 则每次“移动”两个像素。实际上，每一个位置上的卷积运算是并行的，“移动”仅为一种比喻，步幅的概念实际上描述的是卷积核抽取输入图像的局部信息之间的间隔。步幅的大小会直接影响特征图的大小，对于同一张输入图像，步幅越大，抽取局部特征的间隔越大，得到的特征数量越少，特征图也越小。另外，卷积核在水平方向上与垂直方向上的步幅可以不相同，这是由输入图像和卷积核的形状，以及希望得到特征图的形状决定的。假设卷积核在水平方向上的步幅为 s_m ，在垂直方向上为 s_n ，则特征图的输出形状变为：

$$\left\lfloor \frac{x_m - k_m + s_m}{s_m} \right\rfloor \times \left\lfloor \frac{x_n - k_n + s_n}{s_n} \right\rfloor$$

填充 (Padding) 则用于解决卷积层的另一个问题：在应用多层卷积时，图像边缘的像素会随着卷积次数的增加而丢失，深层次的卷积会累计丢失许多像素，导致在最终的输出结果中，卷积神经网络提取图像中间的像素与图像边缘的像素的信息比例不同。换句话说，深层次的卷积网络会更加“关注”图像中间的信息，而忽略图像的边缘。填充就是在输入图像的边界增加一圈元素 (通常为 0)。

图 3：卷积神经网络的填充示意图



数据来源：东北证券

假设在输入图像的水平方向的填充行数为 p_m （左右各一半），在垂直方向上的填充行数为 p_n （上下各一半），则特征图的输出形状变为：

$$\left\lfloor \frac{x_m - k_m + p_m + s_m}{s_m} \right\rfloor \times \left\lfloor \frac{x_n - k_n + p_n + s_n}{s_n} \right\rfloor$$

通常情况下，为了可以在构建网络时更容易地预测每个图层的输出形状，人们会将卷积层的步幅设为 1，而填充则设置为 $p_m = k_m - 1$ 和 $p_n = k_n - 1$ ，则能够保持经过该卷积层输出的特征图的形状与输入图像相同。在此基础上，为了保证输入图像上下、左右具有相同行数的填充，会将卷积核的尺寸设置成奇数，保证填充行数为偶数。在实践中，为了避免增加不可知的问题，一般不会将水平和垂直方向上的步幅或填充设置为不同的值。

2.1.1.1.3. 通道

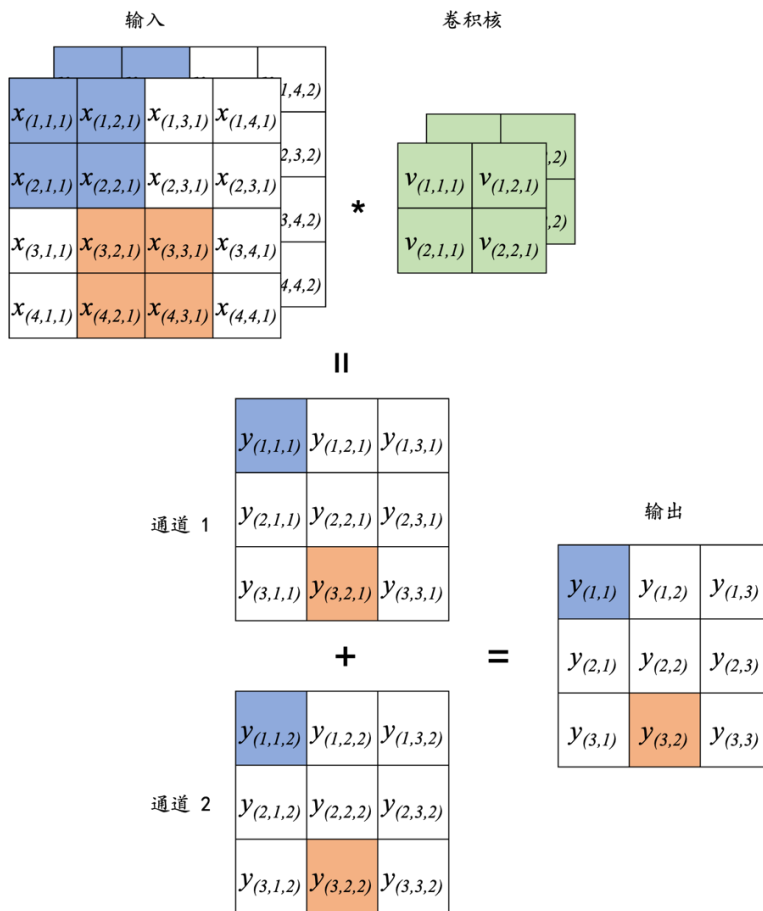
对于一张彩色图片来说，每一个像素点都是由三种原色的值所组成的。因此，一个像素点包含了 3 个值，一张彩色图像实际上是一个由高度、宽度和颜色组成的三维张量，高度和宽度的坐标表示像素点的空间位置，颜色轴则可以看作是每个像素的多维表示。因此，对于一个卷积层，输入 \mathbf{X} 的索引增加一个维度 k ，变为 $[\mathbf{X}]_{i,j,k}$ ，卷积核的权重也相应地从 $[\mathbf{V}]_{a,b}$ 变为 $[\mathbf{V}]_{a,b,c}$ 。

对于隐藏表示 \mathbf{H} ，也采用三维张量 $[\mathbf{H}]_{i,j,d}$ 来表示，其中 d 表示输出通道，以便于随后的输出 \mathbf{H} 将继续以三维张量作为输入进入下一个卷积层。因此，卷积核的权重索引应再次增加一个维度，变为 $[\mathbf{V}]_{a,b,c,d}$ 。

将在卷积层之中传递的三维张量视作一组相互堆叠的二维网格，每一个二维网格的则被称为一条通道。输入 \mathbf{X} 的索引 k 用来索引输入通道，而输出张量 \mathbf{H} 的索引 d 用来索引输出通道，两者可以不相同。

多输入通道与多输出通道的卷积层计算如图 4、5 所示。

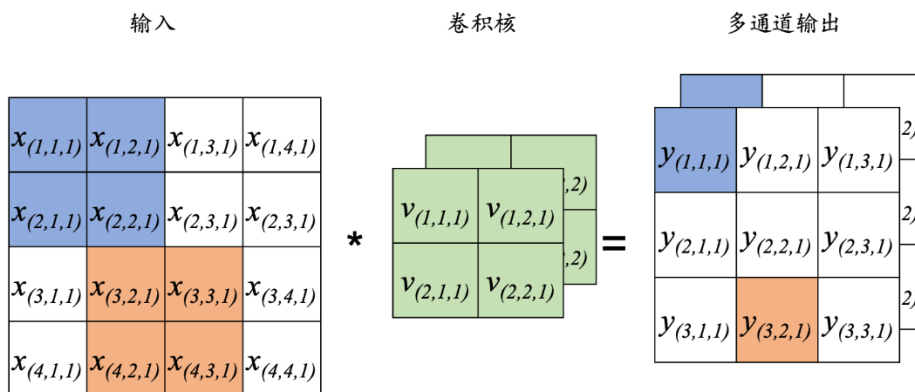
图 4：多输入通道的卷积神经网络示意图



数据来源：东北证券

当输入数据包含多个通道时，在卷积操作中，每个卷积核会对输入数据的每个通道进行卷积，然后将结果求和得到最终的特征图。

图 5：多输出通道的卷积神经网络示意图



数据来源：东北证券

多输入多输出的卷积层，其卷积核为四维张量，难以用示意图表示，若输入通道数为 k ，输出通道数为 d ，则可将其视为 $k \times d$ 个堆叠的二维核张量。每一组输入、输出通道组合都对应一个二维核张量。在互相关运算中，每个输出通道先获取所有输入通道，再对应该输出通道的卷积核计算出结果。

在最流行的神经网络架构中，随着神经网络层数的加深，人们常会增加输出通道的维数，通过减少空间分辨率以获得更大的通道深度。直观地说，可以将每个通道看作对不同特征的响应。而现实可能更为复杂一些，因为每个通道不是独立学习的，而是为了共同使用而优化的。

2.1.1.2. 池化层 (Pooling Layer)

池化层也称为下采样 (Subsampling) 或汇聚层，是卷积神经网络 (CNN) 中的一个关键组成部分，紧跟在卷积层之后。池化层的主要作用是逐步降低数据的空间大小 (即高度和宽度)，从而减少网络中的参数数量和计算量，同时帮助提高特征的空间不变性。

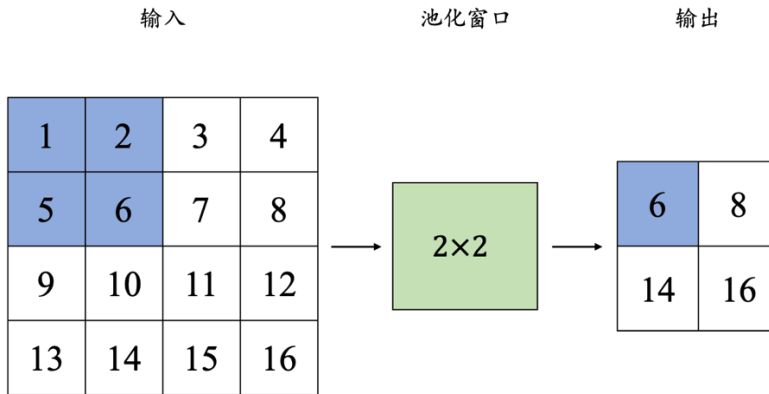
与卷积操作相类似，池化操作通过在输入特征图上滑动一个固定大小的窗口 (也称为“池化核”或“滤波器”)，并在每个窗口上应用最大化或平均化操作，来降低特征图的空间维度。与卷积操作不同的是，池化操作的运算是固定的 (最大化或平均化)，因此不包括需要学习的参数。池化层的主要参数包括：

- (1) 池化窗口大小：池化操作覆盖的区域大小，通常是 $[2 \times 2]$ 或者 $[3 \times 3]$ 。
- (2) 步幅：池化窗口移动的步长。一般情况下，步幅等于池化窗口的大小，以避免重叠。
- (3) 填充：与卷积层类似，池化层也可以使用填充来控制输出特征图的大小。但在实践中，由于与池化层缩小数据尺寸的目的向背，一般情况下池化层不使用填充。

池化层主要有两种类型：最大池化 (Max Pooling) 和平均池化 (Average Pooling)。在最大池化操作中，池化窗口内的最大值被选取为该窗口的代表值。最大池化有助于提取图像中的显著特征，如边缘和纹理。在平均池化操作中，池化窗口内所有值的平均值被计算出来作为该窗口的代表值。平均池化有助于平滑图像特征。

在实践中，最大池化因其能够更好地保留特征的显著性而比平均池化更为常用。图 6 为一个池化窗口为 $[2 \times 2]$ ，步幅为 2 的最大池化层的计算示意图。

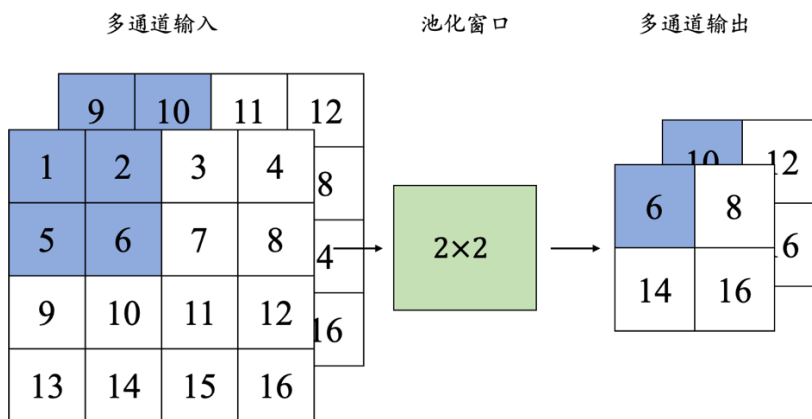
图 6：卷积神经网络的池化层计算示意图



数据来源：东北证券

在处理多通道输入数据时，池化层在每个通道上单独运算，并直接分通道输出，而不是像卷积层一样在通道上对输入进行汇总。因此，池化层的输出通道数与输入通道数相同，计算过程如图 7。

图 7：卷积神经网络的池化层计算示意图



数据来源：东北证券

而当输出数据包含多个通道时，卷积核的深度需要与输出数据的通道数相匹配，以确保能够有效地提取跨通道的特征信息。

池化层的作用主要包括以下几个方面：

- (1) 降低维度：池化层通过减小特征图的高度和宽度，有效减少了后续层的参数数量和计算量，降低了模型的复杂度。通过减少参数数量和计算量，池化层也有助于防止模型过拟合。
- (2) 提高不变性：池化操作有助于使网络对输入图像的小变化（如平移、旋转和缩放）更加鲁棒。最大池化特别有效于保持纹理信息，而平均池化则有助于保持背景信息。
- (3) 抽象和压缩特征表示：池化层通常在连续几个卷积层之后使用，以逐步抽象和压缩特征表示，为高级任务（如分类）提供更加简洁和强大的特征描述。

2.1.2. 卷积神经网络的发展和应用

卷积神经网络(CNN)的发展历史可以追溯到 20 世纪 60 年代，但直到近几十年内，随着计算能力的提升和大数据的可用性，CNN 才真正实现了技术上的突破和广泛应用。

2.1.2.1. 卷积神经网络的发展

1962 年，Hubel 和 Wiesel 的猫脑视觉皮层细胞的研究揭示了视觉感知的层次结构原理，为后来的 CNN 提供了生物学上的灵感。1980 年，日本学者福岛邦彦提出了神经认知机 (Neocognitron) 模型，这是一个受生物视觉启发的分层神经网络，能够识别简单的图案。Neocognitron 是 CNN 早期的雏形，但还并不包含现代 CNN 中的反向传播算法。

1989 年，LeCun 等人将反向传播算法应用于神经网络，创建了用于手写数字识别的 LeNet-5 网络。这是第一个成功的 CNN 实例，它在手写数字识别任务上取得了很好的效果。随后受限于计算能力与数据规模，CNN 与深度学习一同经历了近 20 年的沉寂。

时间进入 21 世纪之后，随着计算能力的增强（尤其是 GPU 的应用）和大规模数据集（如 ImageNet）的创建，CNN 的研究和应用开始快速发展。2012 年，AlexNet 在 ImageNet 挑战赛上取得压倒性胜利，标志着深度学习时代的到来。AlexNet 和 LeNet 的设计理念非常相似，但要比 LeNet-5 要深得多，包含了五个卷积层、三个池化层、两个全连接隐藏层和一个全连接输出层。AlexNet 的成功展示了深度 CNN 在处理图像识别任务上的巨大潜力。

此后，深度 CNN 架构如 VGG(2014)^[2]、GoogLeNet(Inception,2015)^[3]、ResNet(2016)^[4]、DenseNet (2017)^[5]等相继出现，不断刷新图像识别等任务的性能记录。其中的许多工作都为后续的深度技术学习发展产生了深远的影响。

VGG 模型来自牛津大学视觉几何组（Oxford Visual Geometry Group）的工作，其命名也来源于此。VGG 是首个使用块（block）结构进行设计的模型。VGG 块包含了经典卷积神经网络的基本组成部分，包括数个卷积层、非线性激活函数以及池化层。块的思想简化了神经网络的构造方式，对深度学习各领域的研究都具有深远的影响。

GoogLeNet 模型提出了并行连接的概念，其基本的卷积块命名为 Inception 块。Inception 块由四条并行路径组成，每条路径都是卷积核尺寸不同的卷积层，最后在通道的维度上将四条路径上的输出合并。不同尺寸的卷积核能够拟合出不同尺度的图像特征信息，大幅提升了神经网络的特征提取能力。同样的，并行连接的思想也对后续研究产生了深远的影响。

ResNet 模型提出了残差连接的思想，深刻地影响了后来的深度神经网络的设计。ResNet 的基础架构残差块（Residual block），采用的残差映射能够更容易地学习同一个函数。另外，残差块的输入可以通过层间的残差连接更快地向前传播，采用残差连接可以将深度神经网络设计地更深。

随着 CNN 在图像识别和其他计算机视觉任务中的性能大幅提升，其应用也逐渐从图像处理扩展到视频分析、自然语言处理、医学图像分析等多个领域。不仅在学术界取得了巨大成功，CNN 在工业界如自动驾驶、医疗诊断、安全监控等领域也得到了广泛应用。

2.1.2.2. 卷积神经网络在各领域的应用

卷积神经网络（CNN）虽然最初是为图像处理任务设计的，但其能力不限于此。在语音、视频识别、自然语言处理，医学及自然科学研究中都有其应用场景。以下是 CNN 在各个领域的一些关键应用：

（1） 图像和视频识别：

- a) 图像分类：CNN 能够识别和分类图像中的对象，是进行图像分类任务的首选方法。例如，识别照片中的动物、车辆或其他物体。
- b) 物体检测和识别：不仅能够识别图像中的对象，CNN 还能确定对象的位置（通常以边界框的形式表示）。这在安防监控、自动驾驶车辆等领域中非常有用。
- c) 人脸识别和验证：CNN 在人脸识别技术中的应用包括自动标记社交媒体图片、安全认证等。
- d) 视频分析：CNN 用于分析视频数据，包括行为识别、事件检测以及实时体育赛事分析。

（2） 自然语言处理：尽管传统的 NLP 任务通常使用循环神经网络（RNN）和 Transformer 模型，CNN 也在一些特定的 NLP 任务中显示出了其效果，如：

- a) 文本分类和情感分析：通过对文本进行卷积操作，CNN 能够捕捉局部依赖特征，用于新闻分类、情感分析等任务。

- b) 句子建模和语义分析: CNN 能够捕捉句子中的关键词组合, 用于语义角色标注和句子相似度评估。
- (3) 语音识别: 虽然语音识别通常利用 RNN 等模型, CNN 也可应用于提取语音信号的特征, 主要用于:
 - a) 语音到文本转换: 将人类的语音信息转换为书面文本
 - b) 语音命令识别: 识别和执行语音命令, 应用于虚拟助手和智能家居设备。
- (4) 自然科学研究和应用: CNN 在自然科学研究领域的应用日益增多, 包括但不限于:
 - a) 医学图像分析: 自动识别医学影像 (如 X 光片、MRI 和 CT 扫描) 中的病理特征, 辅助医生进行诊断; 在细胞水平上对图像进行分割, 用于量化分析如肿瘤的大小和形状; 分析化合物的结构图像, 预测其潜在的药理活性等应用。
 - b) 物理学研究: 在高能物理学、天体物理学等领域对于特征的分析 and 识别。
 - c) 遥感科学: CNN 在遥感科学, 尤其是卫星遥感中, 用于解析遥感图像的几何、纹理和空间分布特征等。
 - d) 大气科学: CNN 在大气科学中被用于数值模式格点输出的后处理问题, 包括统计降尺度 (Statistical Downscaling, SD)、预报校准、极端天气检测等。

在金融数据分析领域, CNN 通过识别和利用金融时间序列数据中的模式和特征, 同样显示出了巨大的潜力。以下是 CNN 在金融数据分析中的几种应用:

- (1) 股票市场预测: CNN 可以用于分析股票价格和交易量的历史数据, 从而预测未来的股票价格走势。通过将时间序列数据转换为图像 (例如, 将价格和交易量的变化映射为像素强度), CNN 可以像处理图像一样处理这些数据, 识别出潜在的趋势和模式。
- (2) 金融时间识别: CNN 能够从复杂的金融数据中识别特定的事件和模式, 如市场波动、异常交易行为等。通过分析股票市场的交易数据或新闻报道, CNN 可以帮助投资者捕捉到可能影响市场的重大事件。
- (3) 信用评分和风险管理: CNN 可以用于分析客户的财务记录、交易历史和社交媒体活动等数据, 以评估信用风险和制定个性化的信用评分模型。这种方法可以帮助金融机构在提供贷款、信用卡等金融产品时做出更准确的风险评估。
- (4) 异常行为识别: CNN 也可以用于识别异常交易行为和其他指标, 这可能表明欺诈活动。通过分析交易模式和账户行为, CNN 可以帮助金融机构及时发现和防止欺诈。

总结 CNN 模型的原理和特性, 能够得到其具有以下一些特点和优势, 使其适合应用于量化投资:

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/007036116003006043>