卷积神经网络：从图像结构到分类计算

机器学习 · 深度学习基础

学习目标

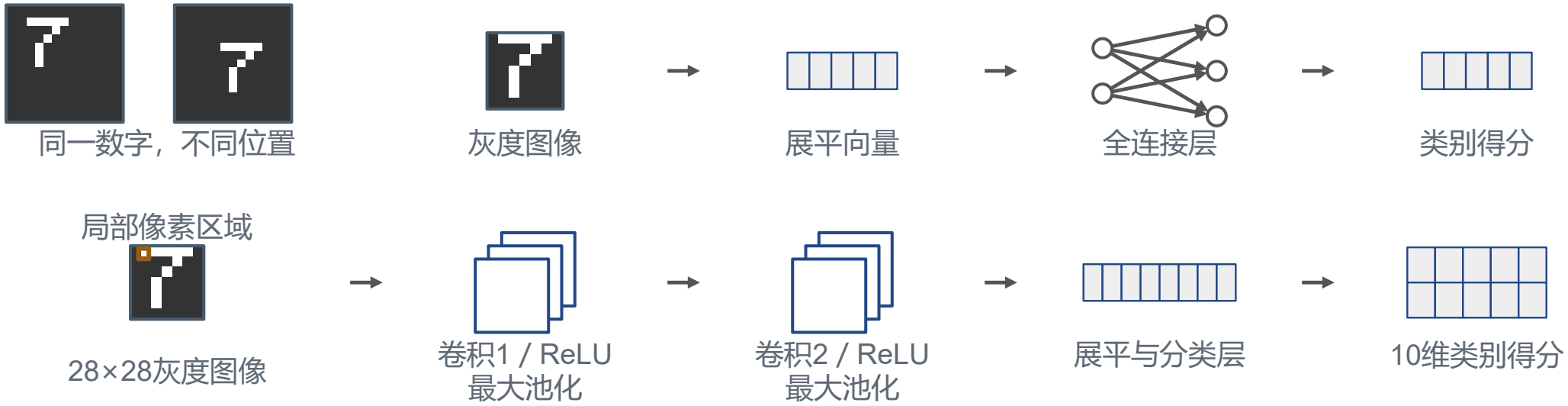
- 解释局部连接和参数共享怎样利用图像结构，并说明适用条件
- 在给定条件下完成基本卷积计算，理解共享参数的梯度怎样累加
- 解释输入 / 输出通道及感受野，跟踪空间尺寸、参数数量与原图依赖范围
- 跟踪一个小型CNN从图像、局部表示到分类得分与损失的过程

1 结构约束与共享学习

从数字分类进入卷积计算、特征图与共享学习。

手写数字分类与网络结构

手写数字分类：输入28×28灰度图像，判断数字0—9。多层感知机（MLP）可对展平图像分类；**卷积神经网络**（CNN）利用图像的空间结构。



多层感知机（上）与卷积神经网络（下）的图像分类结构

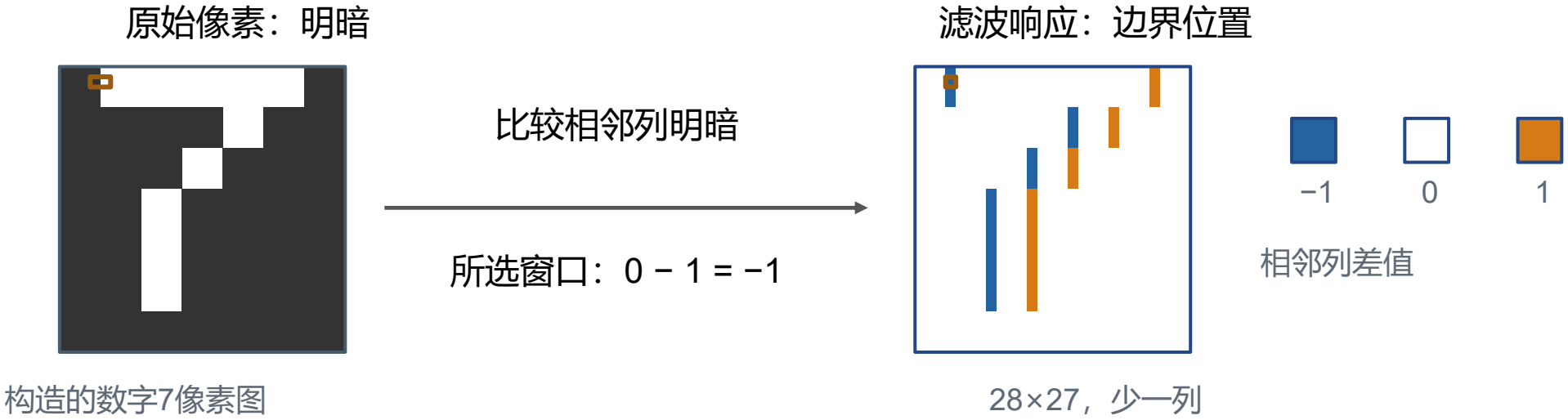
- 每个输入**通道**对应一张空间数值平面；灰度1张、RGB3张，单样本形状为$C \times H \times W$
- **展平**按固定顺序将像素值重排为向量；普通全连接不强制邻域或跨位置共享

数字的位置改变了，类别通常不变。识别笔画的局部规则能否跨位置复用?

从像素图像到特征表示

分类需要能够帮助区分类别的表示。**特征**描述输入的某种性质，例如数字笔画的边界；提取这些特征，为后续分类提供可组合的信息。

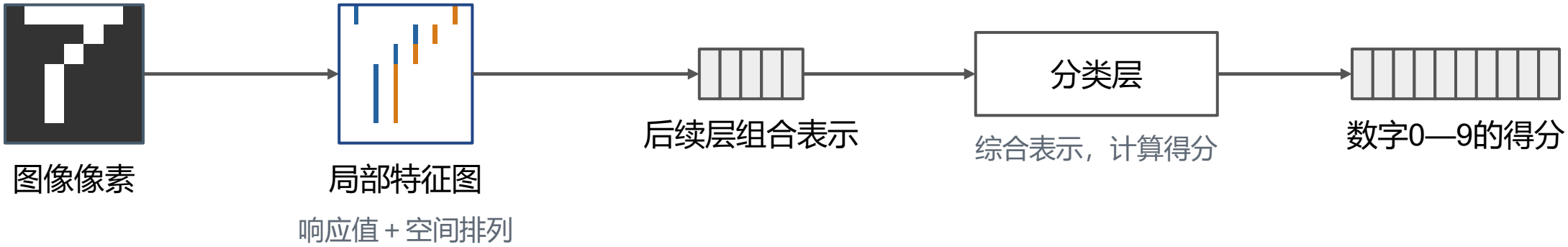
滤波按选定的局部规则变换图像。例如，比较相邻列的明暗，可以突出明暗沿水平方向变化的边界。



- 局部计算得到的数值称为**响应**；按空间位置排列的响应图，可作为表示某种输入性质的**特征图**。本例的非零响应突出相邻列明暗变化的位置

特征表示怎样用于分类

特征图同时保留响应值与空间排列，供后续层继续组合；分类层综合得到的表示，输出各类别的得分。



局部表示经后续组合支持分类

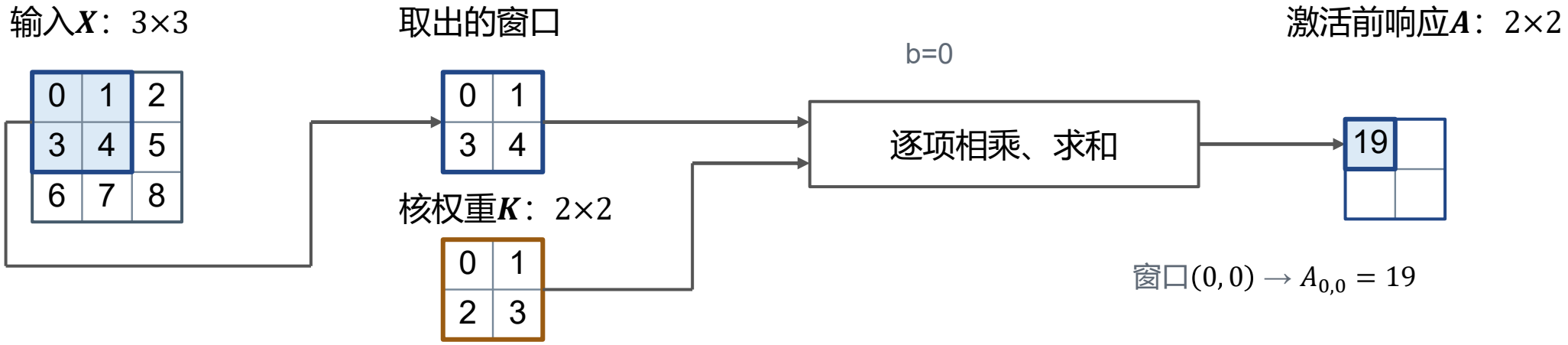
经典图像处理中可以人工设计提取特征的规则；CNN根据任务损失学习卷积核，使得到的表示服务于分类任务。

- 前页的人工边界图说明局部表示的作用；单一边界图不足以直接确定数字类别
- 后续层组合空间邻域和不同通道的响应，分类层据此计算整图得分；具体过程在本讲后半部分展开

卷积核与局部响应

前面的相邻像素求差使用权重$(1,-1)$。把权重按局部窗口的位置排列，就得到**卷积核**。

窗口中的输入与对应权重相乘、求和，再加偏置，得到局部响应。以下列矩阵为例：窗口为2×2，单输入通道、偏置$b=0$、无填充，窗口起点每次移动一格；先计算左上窗口。



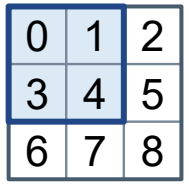
左上窗口与核逐项相乘并求和：

$$A_{0,0} = 0\times0 + 1\times1 + 3\times2 + 4\times3 = 19$$

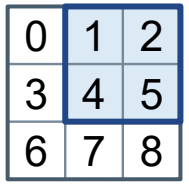
卷积计算与特征图

沿用局部算例的输入、核及$b = 0$，在各合法位置重复局部乘加。窗口每次移动的距离称为**步幅**，本例为1。

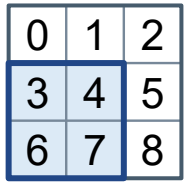
同一个$\boldsymbol{K}$：四个位置复用



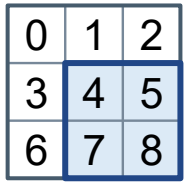
窗口$(0,0) \rightarrow 19$



窗口$(0,1) \rightarrow 25$

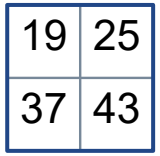


窗口$(1,0) \rightarrow 37$



窗口$(1,1) \rightarrow 43$

特征图$\boldsymbol{A}$（激活前）



按窗口起点编号

- 四个窗口各产生一个输出，输出坐标对应窗口起点
- 同一核的响应按位置组成二维特征图；本例显示激活前值，激活后应另行标明
- 核不翻转的滑动乘加称为**互相关**，深度学习框架通常称其为**卷积**；严格数学卷积需要翻转核

单通道卷积的计算表达

前两页先算出一个响应，再滑动得到特征图。下面用通式概括：单输入通道、步幅1、无填充，核大小为$k_h \times k_w$，偏置为b。

$$A_{i,j} = b + \sum_{u=0}^{k_h-1} \sum_{v=0}^{k_w-1} K_{u,v} X_{i+u,j+v}$$

输入中的窗口

0	1	2
3	4	5
6	7	8

起点$(i,j) = (0,1)$

$X_{i+u,j+v} = X_{1,1} = 4$

核内的位置

0	1
2	3

$(u,v) = (1,0)$

$K_{1,0} = 2$

特征图中的位置

19	25
37	43

$A_{0,1} = 25$

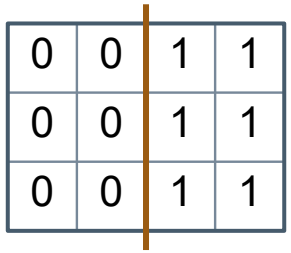
- i,j：窗口起点，也是输出位置；u,v：核内行、列偏移，均从0编号
- 固定i,j，遍历核内u,v并求和，得到一个响应；遍历合法窗口起点，得到整张特征图

A表示激活前响应；前例取$b = 0$。响应值取决于输入与权重，不是类别概率。

差分核的边界响应

横向差分比较相邻列，两值相同则响应为0，明暗变化则响应非零。

二值图：暗0，亮1



明暗交界

同一差分核$(1,-1)$

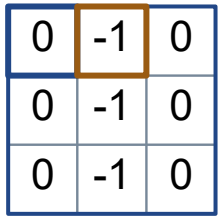


平坦区域：位置(0,0)



跨越边界：位置(0,1)

特征图（激活前）

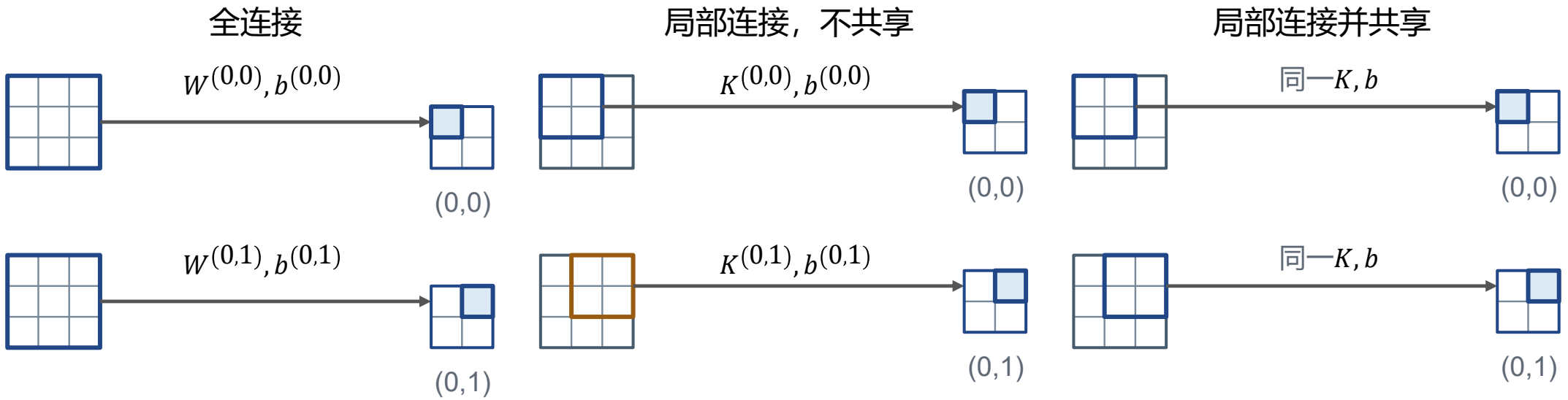


其余位置按同一规则计算

- 此核计算 $A_{i,j} = X_{i,j} - X_{i,j+1}$；0→1得到−1，反方向变化会改变符号
- 这是带符号的激活前响应；ReLU会把负响应置零，两种图不能混称
- 手工边缘核仅用于说明局部规则；实际CNN核通常由训练学习，隐藏通道未必有可命名的单一含义

从卷积计算看局部连接与参数共享

沿用前例，比较输出位置$(0,0)$、$(0,1)$的输入范围与参数。



上下行分别对应蓝框位置；水平箭头连接输出图，旁标所用参数

- **局部连接**：每个输出只读取对应窗口中的输入
- **参数共享**：各位置使用同一组权重和偏置，更新后共同使用新值
- 不共享时，各位置的参数独立学习；即使初始数值相同，也不是同一组参数

局部连接与参数共享减少多少参数

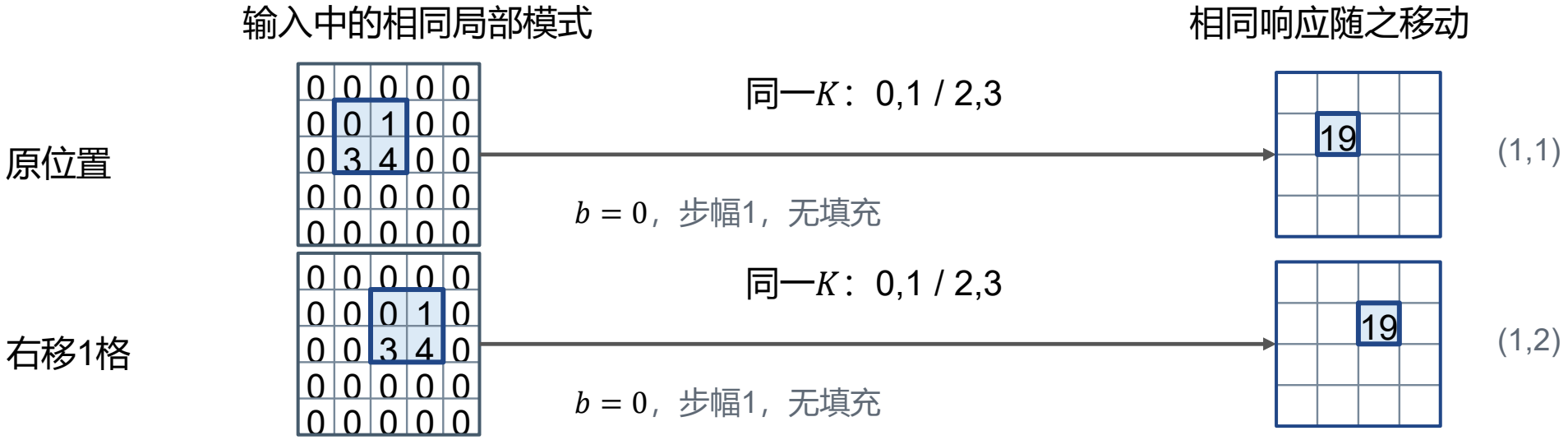
沿用3×3单通道输入和2×2输出，共9个输入、4个输出。局部窗口为2×2，步幅1、无填充；三种结构都计入偏置。

全连接	局部连接，不共享	局部连接并共享
		
4个输出位置	4个输出位置	4个输出位置
每位置各有一组	每位置各有一组	4个位置共用一组
每组9权重 + 1偏置	每组4权重 + 1偏置	每组4权重 + 1偏置
$4\times(9+1)=40$	$4\times(4+1)=20$	$1\times(4+1)=5$

- 局部连接减少每组权重：9个变为4个；参数共享减少独立参数组：4组变为1组
- 三种情况仍产生4个输出值；重复使用参数不增加参数数量，也不保证预测更好

共享规则与平移等变性

共享核跨位置复用同一规则；步幅和边界满足条件时，相同响应随局部模式移动。

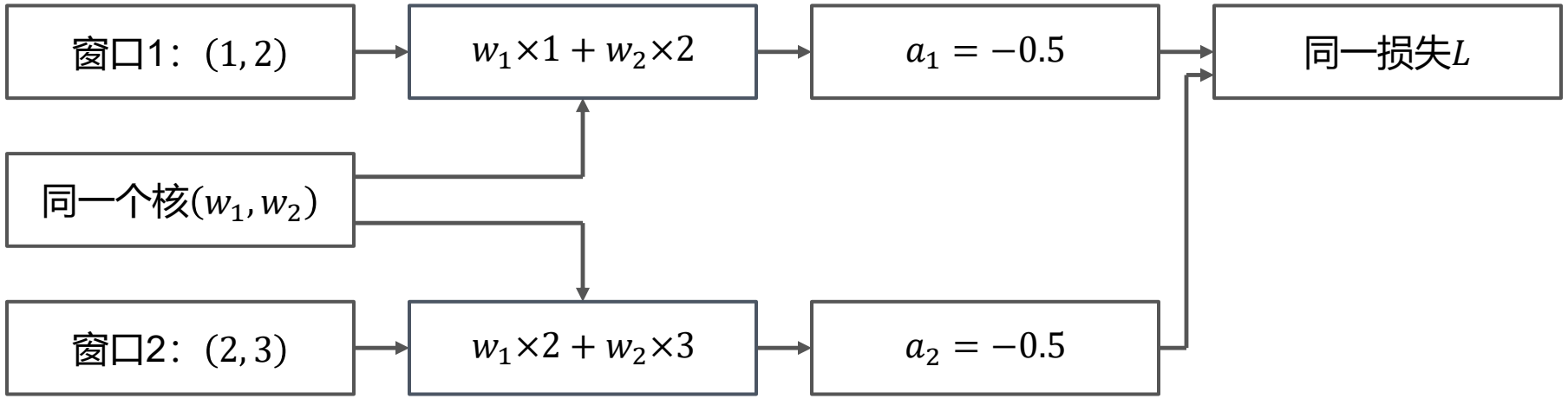


- 步幅1且不受有限边界影响时，输入移动、响应对应移动，称为**平移等变性**；响应位置并非不变
- 普通全连接层不约束不同位置使用同一组参数，因此**不保证**平移等变性
- 等变性不保证分类结果不变；后续汇总与分类仍会影响结果

共享卷积核的学习

前面用同一个核计算不同位置的响应。训练时，这些位置的误差怎样共同更新这个核?

用一维小例展示：输入$(1,2,3)$，当前核$(w_1,w_2)=(0.5,-0.5)$，步幅1、无填充、无偏置。



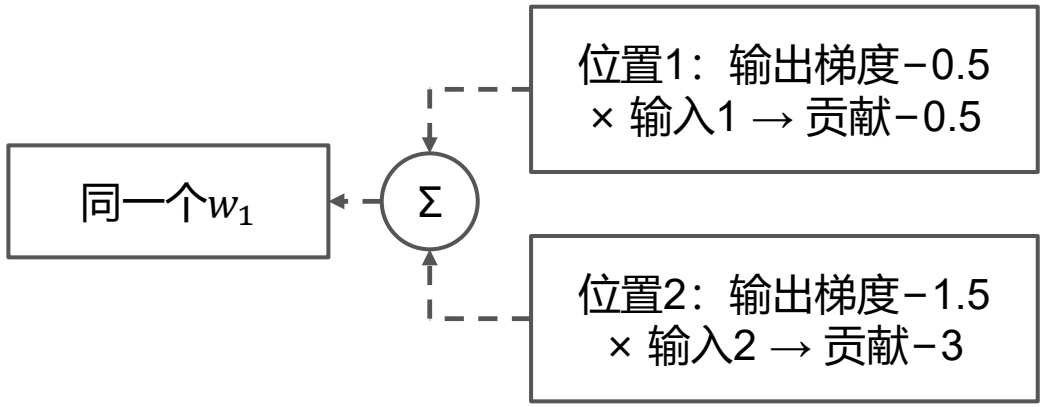
在这个简化拟合任务中，直接用两位置响应a_1,a_2预测给定目标$(t_1,t_2)=(0,1)$，采用平方误差和：

$$L=\frac{1}{2}[(a_1-t_1)^2+(a_2-t_2)^2]=1.25$$

两个位置的误差都依赖同一组权重。下一页用链式法则计算它们如何共同更新w_1。

共享参数的梯度汇总

共享的是同一个被学习的参数；总梯度要加上所有使用位置的链式法则贡献。



汇总使用路径

只核对w_1的两项贡献：

$$\frac{\partial L}{\partial w_1} = (-0.5)\times 1 + (-1.5)\times 2 = -3.5$$

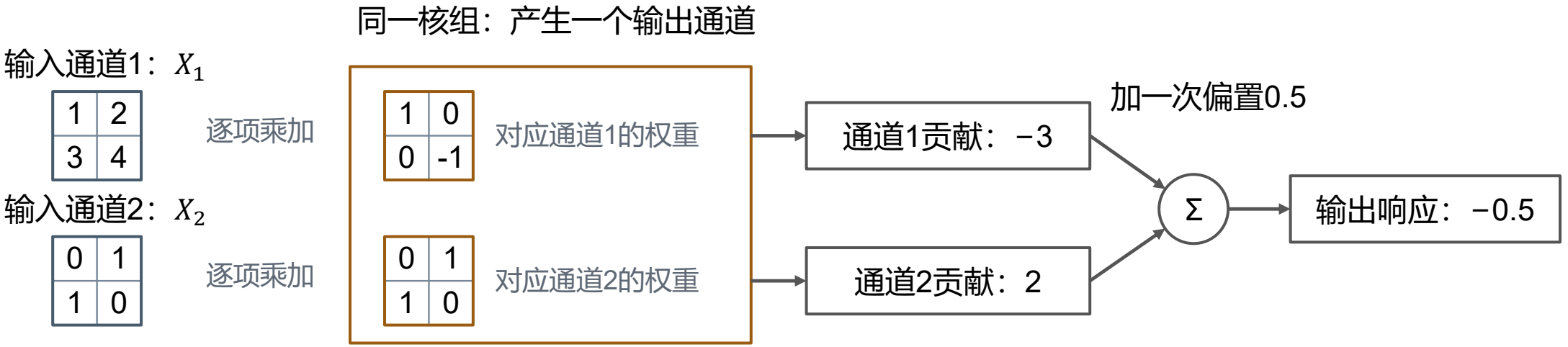
- 学习率 $\eta = 0.1$ 时，同一参数更新为 $w_1^{\text{new}} = 0.5 - 0.1(-3.5) = 0.85$；不是分别更新两个位置的权重
- 多个位置提供学习信号，更新仍作用于同一组核；忽略一个使用位置会漏掉路径
- 本例L是误差和；若L定义为平均损失，平均因子应进入导数，不因共享而额外默认平均

2 通道、尺寸与空间范围

区分通道、空间尺寸及原始输入依赖，说明局部汇总的作用。

多通道信息的组合

一个输出位置需要组合各输入通道的局部信息。固定一个输出通道和一个输入通道，对应的二维权重阵列称为**核平面**；该输出通道的全部核平面构成一个**核组**。



- 各核平面分别调节对应通道、各位置的贡献；把多通道窗口的加权求和按通道拆开计算，最后相加：$A = (-3) + 2 + 0.5 = -0.5$
- 同一核组在不同空间位置复用，得到一张特征图；此处只计算其中一个位置

多组核形成多种特征响应

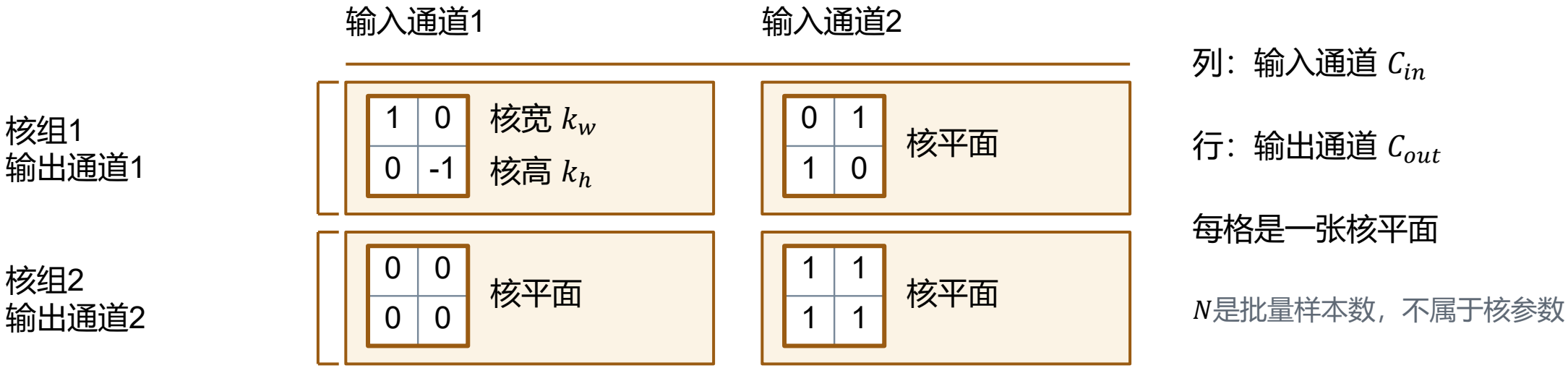
权重决定各位置、各通道的贡献；同一输入经不同核组加权，可产生不同响应。每组沿空间复用，形成一张特征图。



- 核组1组合两通道；核组2只累加通道2。各输出表示响应，不是类别
- 输入更大时，各组滑动形成二维图，供后层组合。训练按任务损失调整权重，不保证每个通道有可命名的特征含义

卷积核张量与参数数量

固定输出通道与输入通道，得到一张二维**核平面**；每个输出通道对应一个**核组**。图中每格是核平面，每行是核组；所有核组构成四维权重张量。



$\boldsymbol{K} \in \mathbb{R}^{C_{out}\times C_{in}\times k_h\times k_w}$

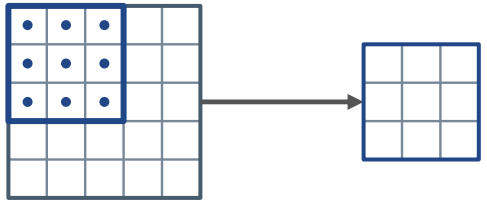
- 参数数目为 $C_{out}(C_{in}k_hk_w+1)$；上例 $2(2\times2\times2+1)=18$，每输出通道一个偏置
- 公式限普通有偏置、非分组卷积；固定通道与核大小，图像高宽主要改变空间计算量和输出尺寸

零填充与输出尺寸

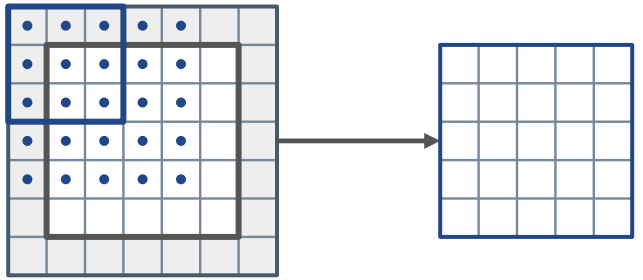
不填充时，只计算能完整放入的窗口。步幅1、核高宽为k_h, k_w时，输出高宽各减少$k_h - 1, k_w - 1$。

原输入5×5，核3×3；步幅1

不填充：输出3×3



每侧补1：输出5×5



蓝点：合法窗口起点

灰边：补入的零，不是观测像素

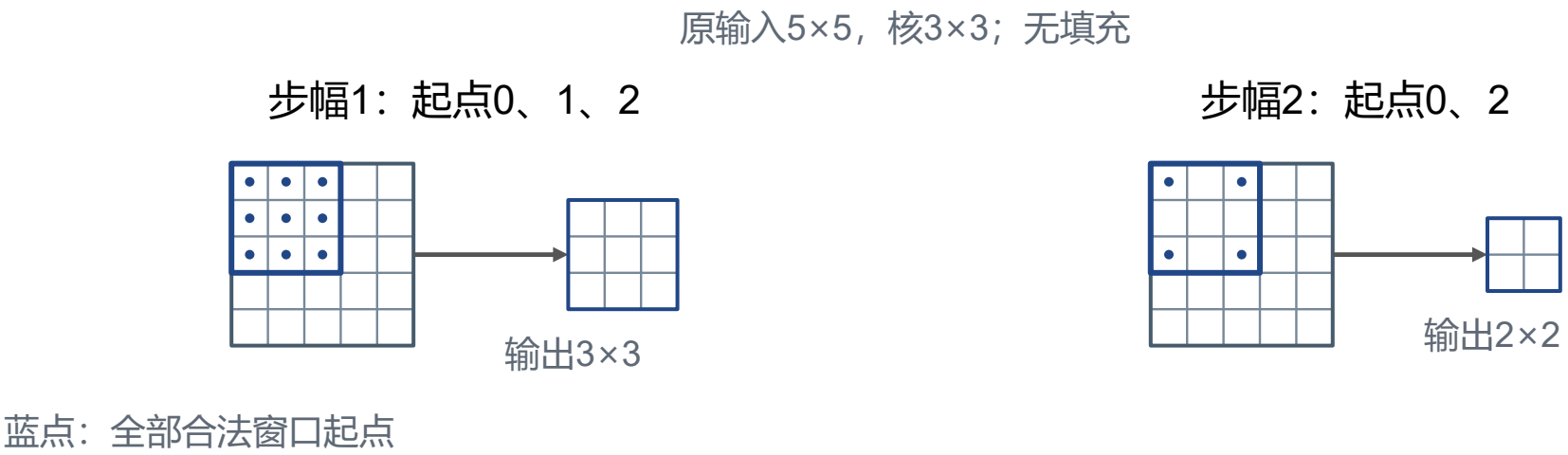
零填充在输入边界外补零，p表示每侧补入数量。本例每侧补1：

$$H_{out} = W_{out} = 5 + 2\times1 - 3 + 1 = 5$$

- 填充可控制连续卷积造成的尺寸缩小，让窗口覆盖原图边界附近；补零没有增加观测信息
- 前面的2×2核使高宽各少1；本页3×3核使高宽各少2

步幅与采样密度

步幅s是相邻窗口起点的移动距离。同一个局部运算可在选定起点执行：步幅1逐格计算，步幅2每隔一格计算。



较大步幅主动降低空间分辨率和计算量，也可能遗漏细节；步幅1保留更密的空间响应。

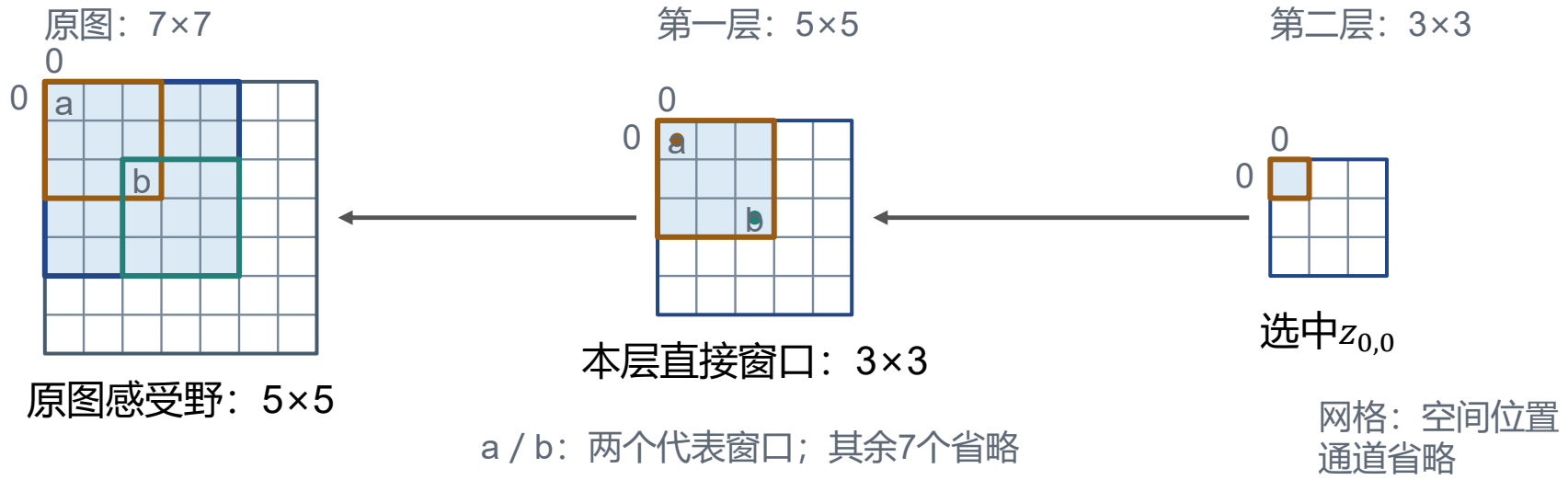
普通、无膨胀卷积，每个方向分别按下式计数：

$$n_{out} = \left\lfloor \frac{n+2p-k}{s} \right\rfloor + 1$$

n, k, p, s分别是该方向的输入尺寸、核尺寸、每侧填充与步幅；向下取整保证只计完整窗口。

局部窗口与原图感受野

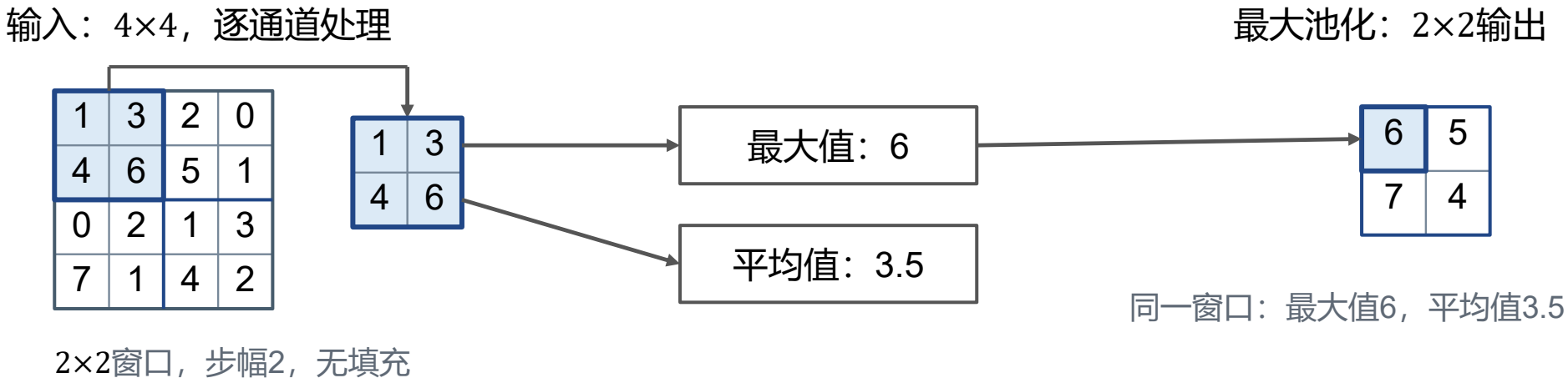
组合邻近表示，可以利用更大范围的形状信息。**感受野**是一个输出可能依赖的原图范围。本例两层3×3卷积，步幅1、无填充、无膨胀。



- $z_{0,0}$使用第一层9个位置；各位置分别依赖原图中错开的3×3窗口，9个窗口的**并集**为5×5
- 感受野表示可能依赖，影响强弱由参数和激活状态决定；有填充时还须区分真实像素与补入值

池化与局部汇总

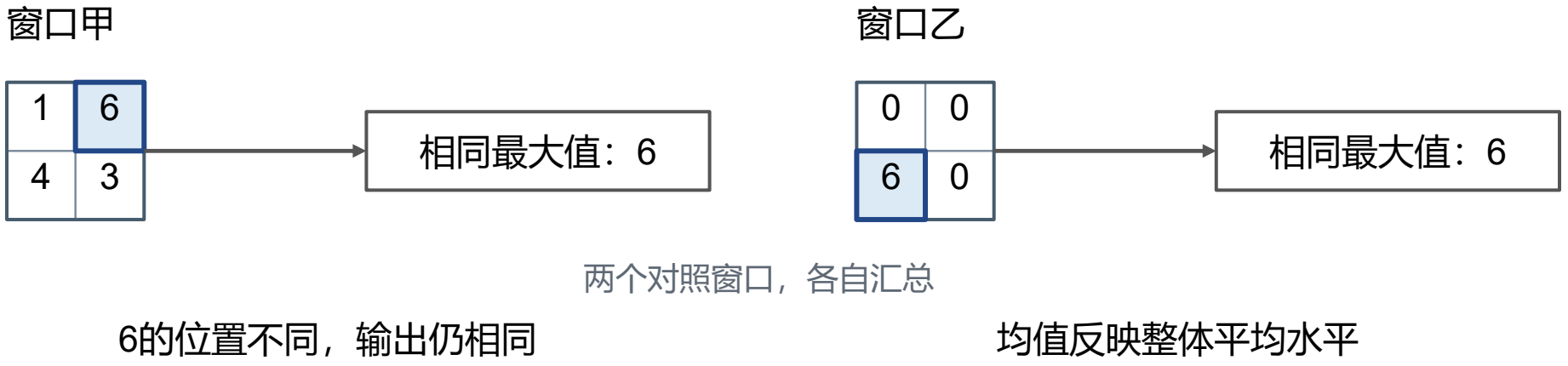
池化在每个通道内把局部响应汇总为一个值；配合较大步幅，形成更粗的空间表示，减少后续处理的位置数。



- **最大池化**保留窗口最大响应；**平均池化**保留平均水平。左上窗口分别为6和$(1+3+4+6)/4=3.5$
- 普通池化无可学习核权重、通道数不变；本例步幅2使高宽各减半，同时丢失部分数值和精确位置

池化保留的信息与丢失的细节

局部模式移动后，汇总值可能保持；下采样会丢失空间细节。



- 最大值概括最强响应，不能恢复全部原值，也未必保留最大值的精确位置；均值概括整体平均水平
- 窗口内部分移动可保持最大值，跨窗口移动、步幅和边界仍会改变结果，不保证任意平移下分类不变
- 普通池化直接汇总且无可学习权重；步幅卷积使用可学习的加权规则，两者计算方式不同

3 从局部表示到整图分类

在同一LeNet类小型网络中，连接局部表示、空间汇总、分类得分与训练。

CNN的整体分类计算

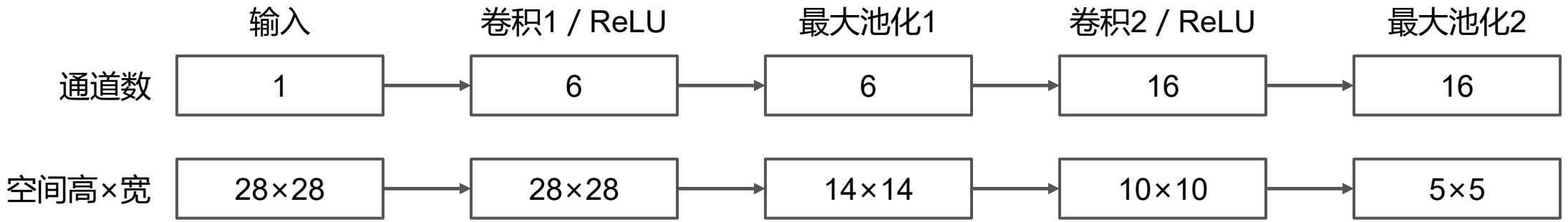
回到开头定位的同一LeNet类教学变体：28×28灰度输入，ReLU与最大池化，用于数字0—9分类，输出10维logits，即各数字类别的未归一化得分。



- 每组覆盖全部输入通道：卷积1每组覆盖1通道；卷积2每组覆盖6通道
- 展平只重排每样本的数值；分类层由整图表示给出10个得分，局部响应不是类别
- 局部响应的平移等变不保证分类不变；后续汇总与分类层仍会影响位置敏感性

卷积块的通道与空间变化

两卷积均5×5、步幅1；卷积1每侧填充2，卷积2无填充；两次池化均2×2、步幅2、无填充。

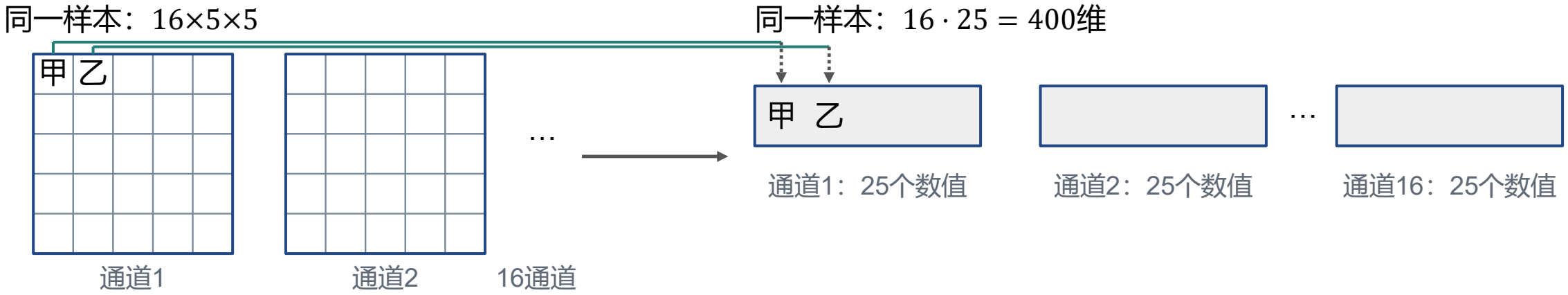


ReLU改变数值，不改变形状

- 核组数决定卷积输出通道$1 \to 6 \to 16$；卷积配置与池化改变空间尺寸，池化保留通道数
- 卷积1有156个参数，卷积2有2416个参数；ReLU与普通池化无参数
- 图中省略批量轴N，各阶段均保留N，不把不同样本合并

展平表示与分类层

第二个池化块每样本输出16张5×5特征图；展平按固定顺序将数值重排为400维向量。



标记格在重排前后表示同一数值

每个样本分别重排，批量轴N保持

环节	表示变化与职责
隐藏全连接与ReLU	$400 \to 120 \to 84$，组合输入表示并作非线性变换
最后全连接	$84 \to 10$，直接输出类别logits

- 全网络共61706个参数，主要来自分类层；卷积共享不会自动使整个CNN参数量很小
- 最后一层的得分供分类与损失使用；预测性能仍需在相应任务与数据上评价

类别得分与交叉熵损失

每个样本的10维logits $\boldsymbol{o}$不是概率。最大得分对应预测类别；Softmax将得分映为概率，交叉熵取决于真实类别t的概率。

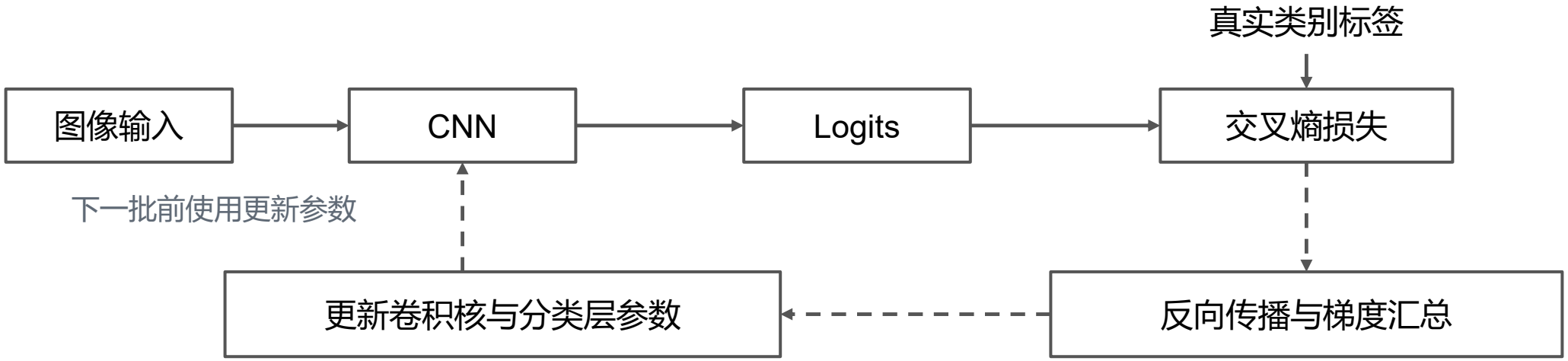
$$p_j = \frac{\exp(o_j)}{\sum_{r=0}^{9} \exp(o_r)}, L = -\ln p_t$$

- t是0到9的类别索引，exp是指数函数；提高真实类别概率会减小该单样本损失
- 给定微例 $\boldsymbol{o} = (\ln 9, 0, \ldots, 0), t = 0$：预测0，$p_0 = 0.5, L = \ln 2 \approx 0.6931$；选对类别不等于高置信度
- CrossEntropyLoss直接接收logits与整数类别索引，接口前不再手动Softmax

本例得分由人为设定，用于核对概率与损失的计算，不代表参考CNN训练后的预测结果或准确率。

卷积网络的训练与评价

CNN复用前几讲的训练框架：任务损失同时向卷积层与分类层提供学习信号。



- 每批先清零梯度；共享卷积核汇总各样本、各空间位置的梯度贡献，分类层从同一损失获得梯度
- 评价使用独立划分、固定参数；本模型无BN / Dropout，含这些模块的网络还需按前述要求切换到评价模式

配套代码只运行合成输入的形状、更新和固定参数评价检查；未做数据集训练，不报告任务成绩。

本讲总结

- 结构层：局部连接限定空间邻域，参数共享让同一局部规则跨位置复用；两种约束承担不同职责
- 计算层：卷积形成局部响应，输入通道在核组内求和，多核组产生输出通道；层级堆叠扩大原图感受野，空间汇总改变分辨率与保留的信息
- 任务层：局部表示经重排和分类层形成整图类别得分；根据任务损失反向计算梯度，并更新卷积层与分类层的参数

等变不等于分类不变；参数更少、感受野更大或微例正确，均不单独保证预测更好。实际表现需要相应任务与数据评价。