wupan翻译

2018年4月07日 12:01

94页论文综述卷积神经网络：从基础技术到研究前景

卷积神经网络（CNN）在计算机视觉领域已经取得了前所未有的巨大成功，但我们目前对其效果显著的原因还没有全面的理解。近日，约克大学电气工程与计算机科学系的 Isma Hadji 和 Richard P. Wildes 发表了论文《What Do We Understand About Convolutional Networks?》，对卷积网络的技术基础、组成模块、当前现状和研究前景进行了梳理，介绍了我们当前对 CNN 的理解。机器之心对本论文进行了摘要式的编译，更详细的信息请参阅原论文及其中索引的相关文献。

论文地址：https://arxiv.org/abs/1803.08834

1 引言

1.1动机

过去几年来，计算机视觉研究主要集中在卷积神经网络（常简称为 ConvNet 或 CNN）上。这些工作已经在广泛的分类和回归任务上实现了新的当前最佳表现。相对而言，尽管这些方法的历史可以追溯到多年前，但对这些系统得到出色结果的方式的理论理解还很滞后。事实上，当前计算机视觉领域的很多成果都是将 CNN 当作黑箱使用，这种做法是有效的，但其有效的原因却非常模糊不清，这严重满足不了科学研究的要求。尤其是这两个可以互补的问题：（1）在被学习的方面（比如卷积核），究竟被学习的是什么？（2）在架构设计方面（比如层的数量、核的数量、池化策略、非线性的选择），为什么某些选择优于另一些选择？这些问题的答案不仅有利于提升我们对 CNN 的科学理解，而且还能提升它们的实用性。

此外，目前实现 CNN 的方法需要大量训练数据，而且设计决策对结果表现有很大的影响。更深度的理论理解应该能减轻对数据驱动的设计的依赖。尽管已有实证研究调查了所实现的网络的运行方式，但到目前为止，这些结果很大程度上还局限在内部处理过程的可视化上，目的是为了理解 CNN 中不同层中发生的情况。

1.2目标

针对上述情况，本报告将概述研究者提出的最突出的使用多层卷积架构的方法。要重点指出的是，本报告将通过概述不同的方法来讨论典型卷积网络的各种组件，并将介绍它们的设计决策所基于的生物学发现和/或合理的理论基础。此外，本报告还将概述通过可视化和实证研究来理解 CNN 的不同尝试。本报告的最终目标是阐释 CNN 架构中涉及的每一个处理层的作用，汇集我们当前对 CNN 的理解以及说明仍待解决的问题。

1.3 报告提纲

本报告的结构如下：本章给出了回顾我们对卷积网络的理解的动机。第 2 章将描述各种多层网络并给出计算机视觉应用中使用的最成功的架构。第 3 章将更具体地关注典型卷积网络的每种构造模块，并将从生物学和理论两个角度讨论不同组件的设计。最后，第 4 章将会讨论 CNN 设计的当前趋势以及理解 CNN 的工作，并且还将重点说明仍然存在的一些关键短板。

2 多层网络

总的来说，本章将简要概述计算机视觉领域中所用的最突出的多层架构。需要指出，尽管本章涵盖了文献中最重要的贡献，但却不会对这些架构进行全面概述，因为其它地方已经存在这样的概述了（比如 [17, 56, 90]）。相反，本章的目的是为本报告的剩余部分设定讨论基础，以便我们详细展示和讨论当前对用于视觉信息处理的卷积网络的理解。

2.1 多层架构

在近来基于深度学习的网络取得成功之前，最先进的用于识别的计算机视觉系统依赖于两个分离但又互补步骤。第一步是通过一组人工设计的操作（比如与基本集的卷积、局部或全局编码方法）将输入数据变换成合适的形式。对输入的变换通常需要找到输入数据的一种紧凑和/或抽象的表征，同时还要根据当前任务注入一些不变量。这种变换的目标是以一种更容易被分类器分离的方式改变数据。其次，被变换的数据通常用于训练某些类型的分类器（比如支持向量机）来识别输入信号的内容。通常而言，任何分类器的表现都会受到所使用的变换方法的严重影响。

多层学习架构为这一问题带来了不同的前景，这种架构提出不仅要学习分类器，而且要从数据中直接学习所需的变换操作。这种形式的学习通常被称为「表征学习」，当应用在深度多层架构中时即被称为「深度学习」。

多层架构可以定义为允许从输入数据的多层抽象中提取有用信息的计算模型。一般而言，多层架构的设计目标是在更高层凸显输入中的重要方面，同时能在遇到更不重要的变化时变得越来越稳健。大多数多层架构都是将带有交替的线性和非线性函数的简单构建模块堆叠在一起。多年以来，研究者已经提出了很多不同类型的多层架构，本章将会覆盖计算机视觉应用中所采用的最为突出的此类架构。人工神经网络是其中的关注重点，因为这种架构的表现非常突出。为了简单起见，后面会直接将这类网络称为「神经网络」。

2.1.1 神经网络

典型的神经网络由一个输入层、一个输出层和多个隐藏层构成，其中每一层都包含多个单元。

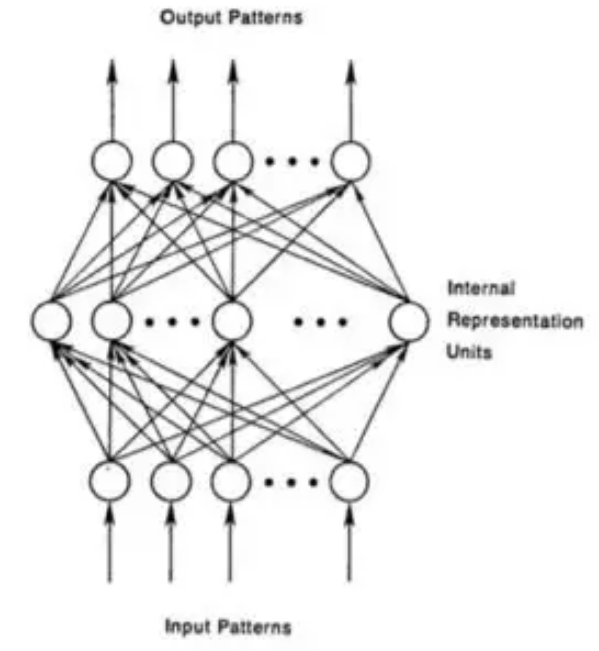


图 2.1：典型神经网络架构示意图，图来自 [17]

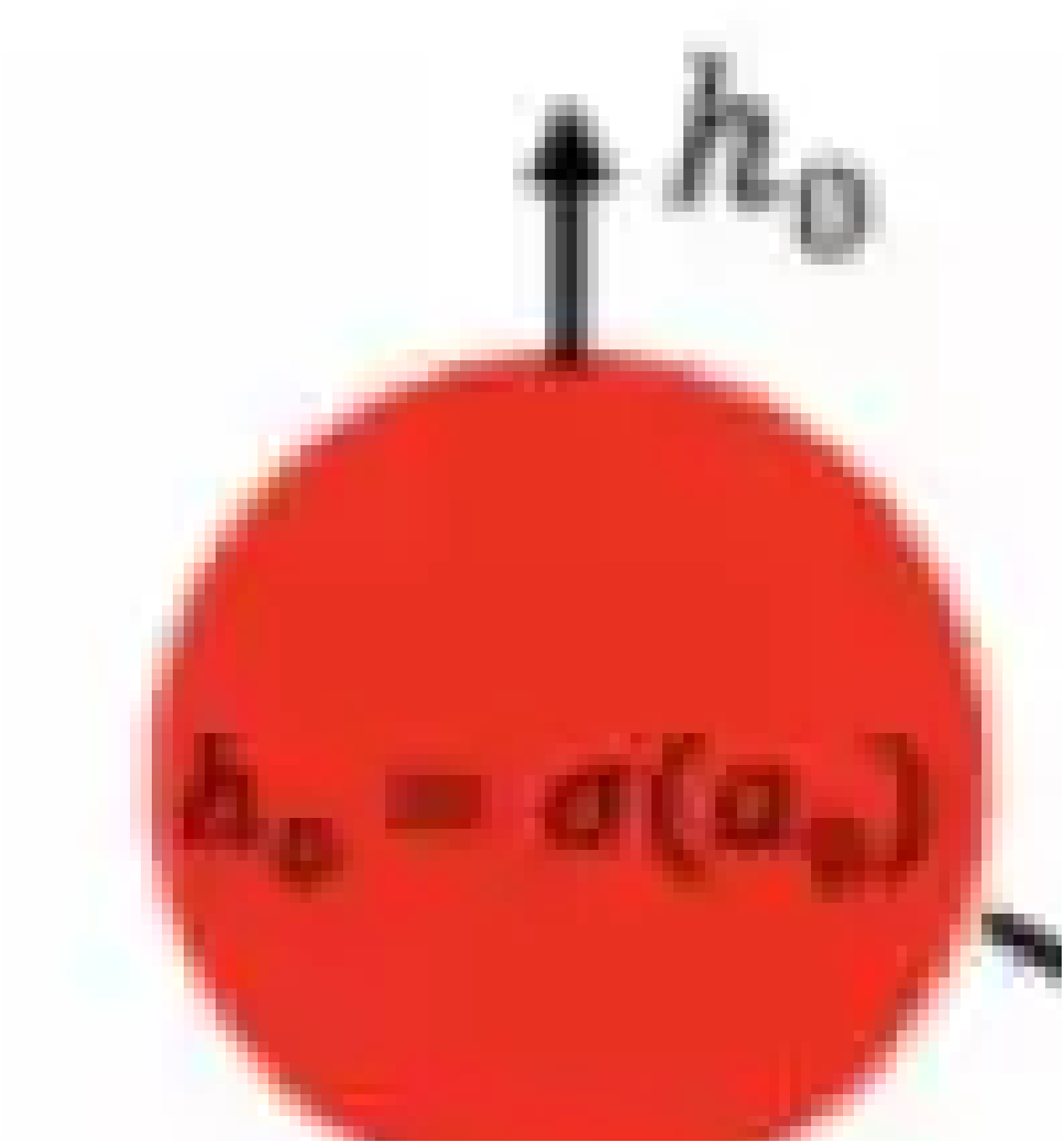
自动编码器可以定义为由两个主要部分构成的多层神经网络。第一个部分是编码器，可以将输入数据变换成特征向量；第二个部分是解码器，可将生成的特征向量映射回输入空间。


Input

图 2.2：典型自动编码器网络的结构，图来自 [17]

2.1.2 循环神经网络

当谈到依赖于序列输入的任务时，循环神经网络（RNN）是最成功的多层架构之一。RNN 可被视为一种特殊类型的神经网络，其中每个隐藏单元的输入时其当前时间步骤观察到的数据和其前一个时间步骤的状态。



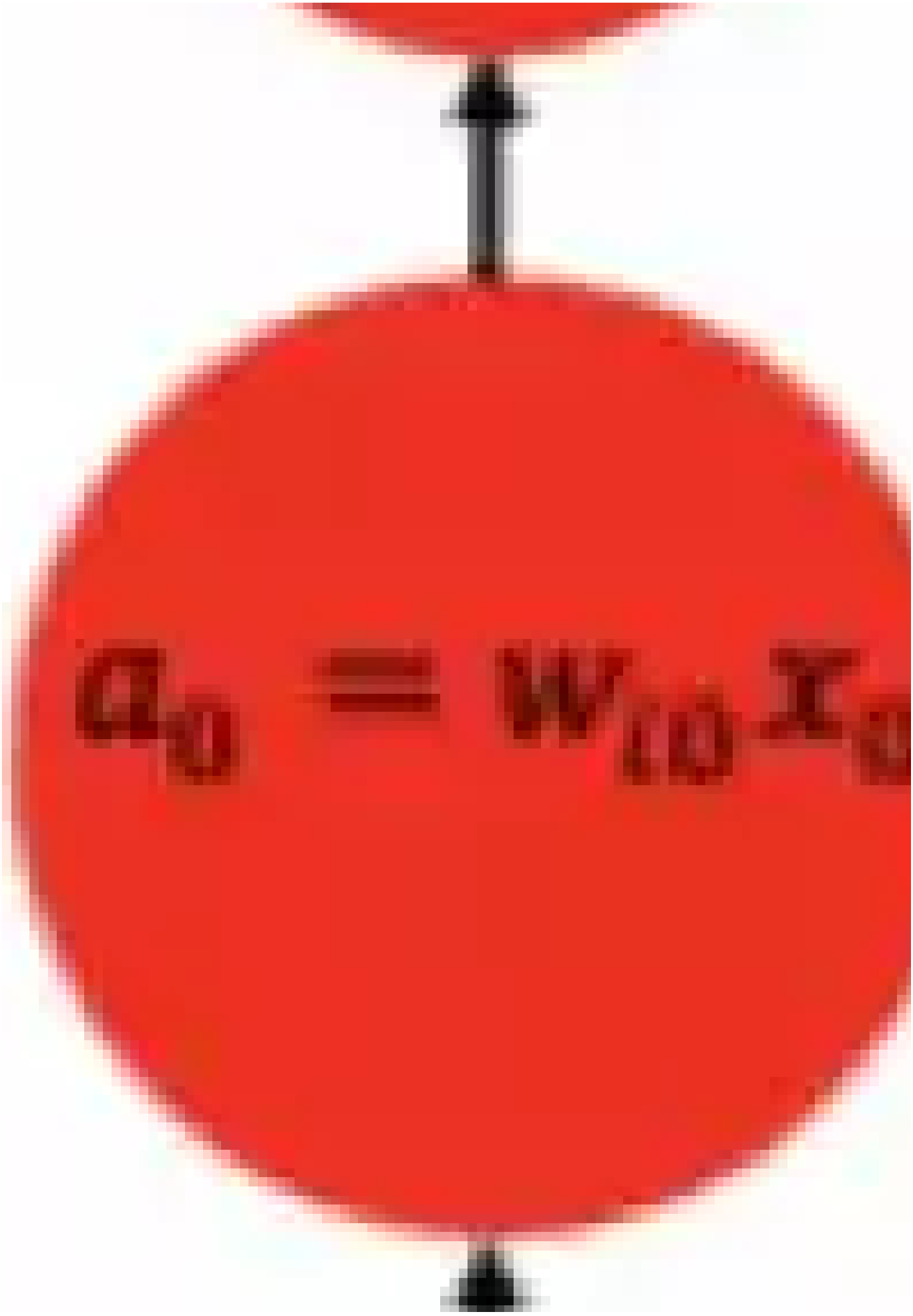
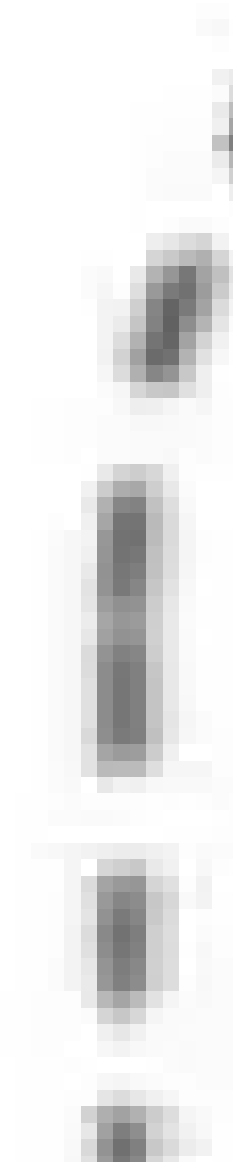
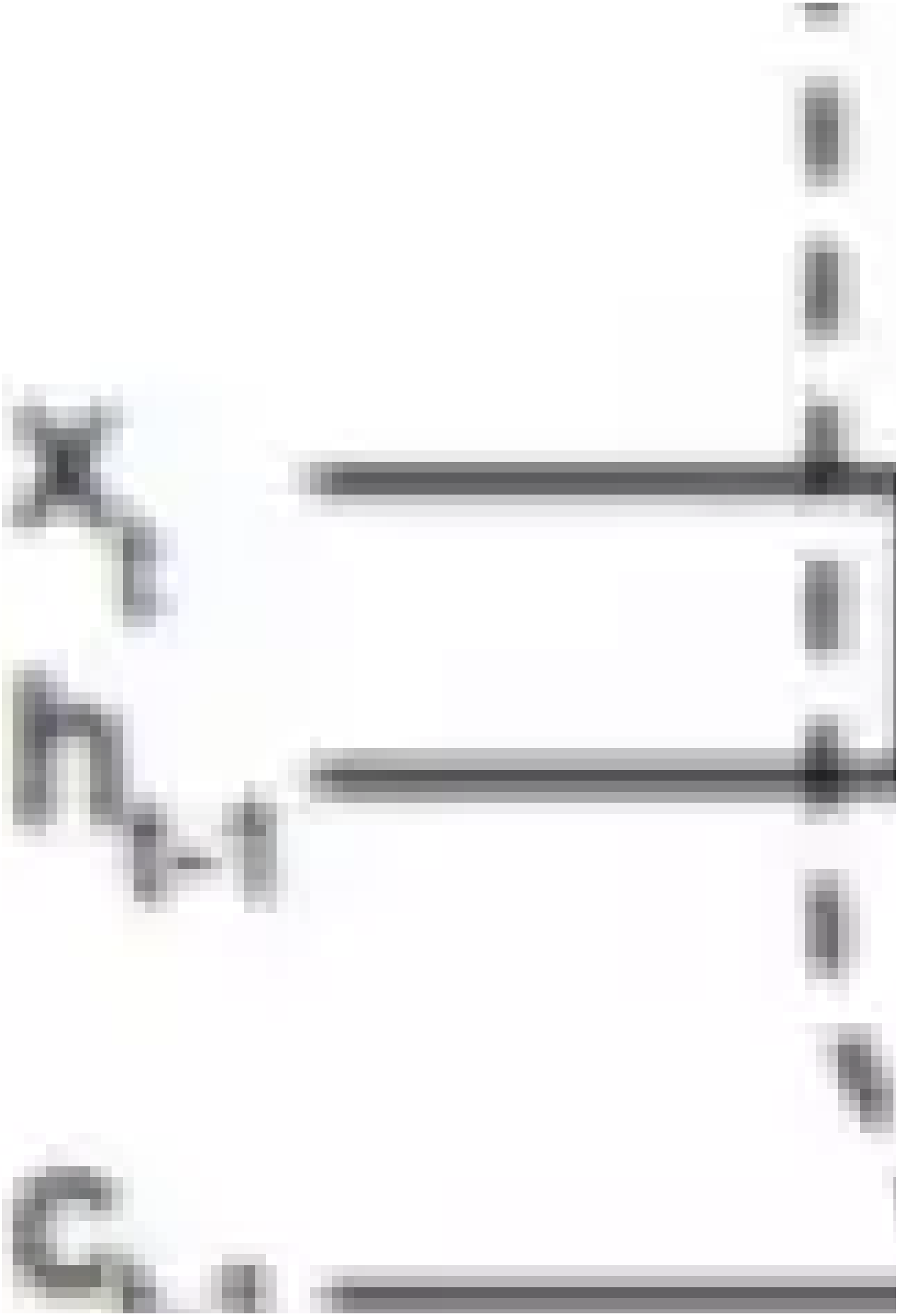


图 2.3：标准循环神经网络的运算的示意图。每个 RNN 单元的输入都是当前时间步骤的新输入和前一个时间步骤的状态；然后根据 $h_t = \sigma(w_i x_t + u_i h_{t-1})$ *计算得到新输出，这个输出又可被馈送到多层 RNN 的下一层进行处理。*





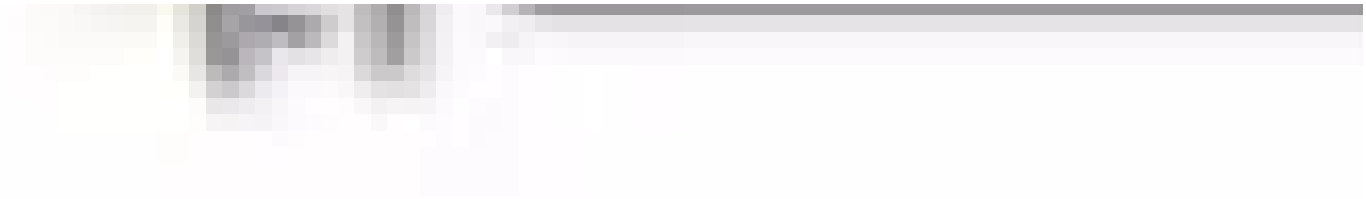
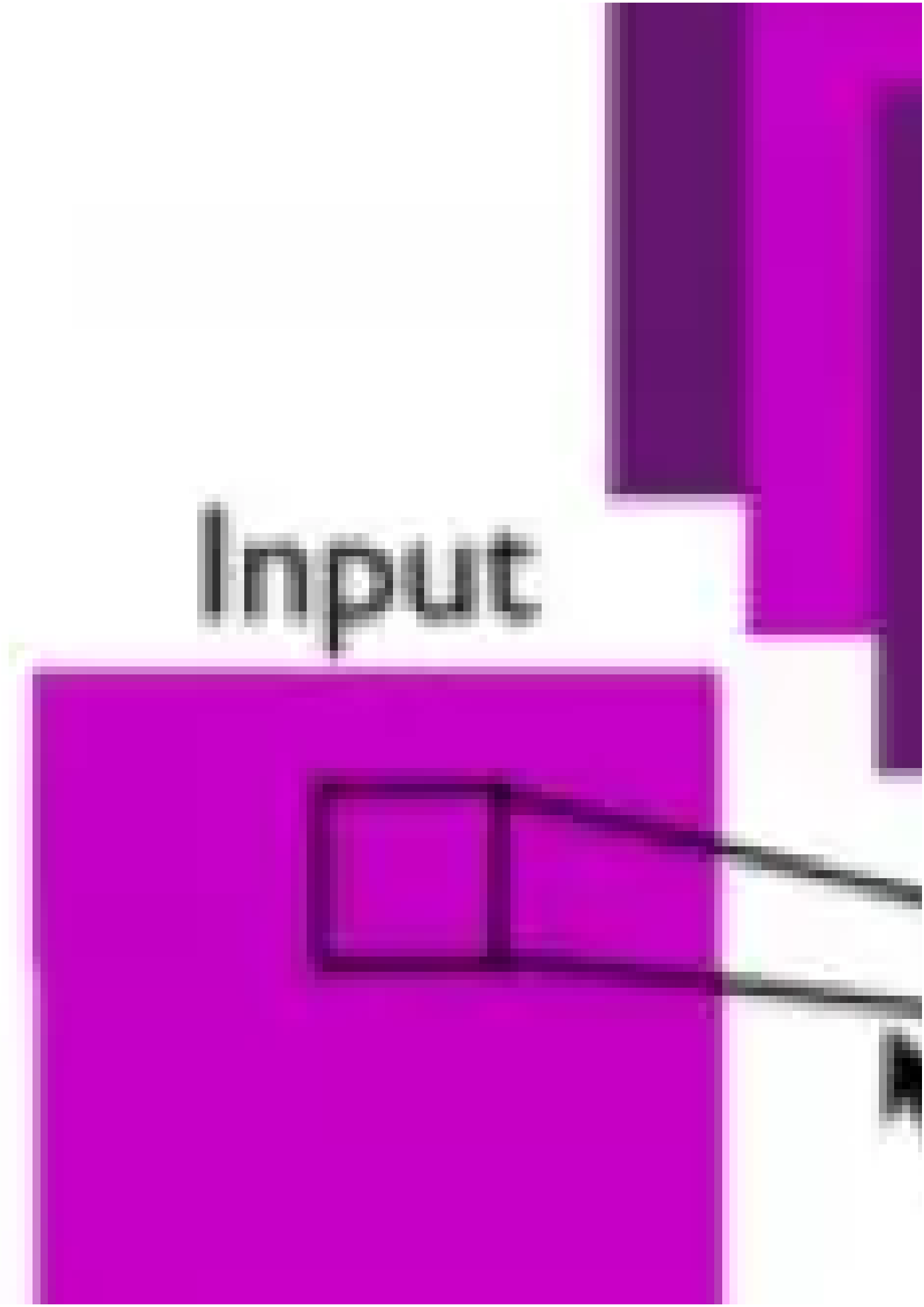


图 2.4：典型 LSTM 单元示意图。该单元的输入是当前时间的输入和前一时间的输入，然后它会返回一个输出并将其馈送给下一时间。LSTM 单元的最终输出由输入门、输出门和记忆单元状态控制。图来自 [33]

2.1.3 卷积网络

卷积网络（CNN）是一类尤其适合计算机视觉应用的神经网络，因为它们能使用局部操作对表征进行分层抽象。有两大关键的设计思想推动了卷积架构在计算机视觉领域的成功。第一，CNN 利用了图像的 2D 结构，并且相邻区域内的像素通常是高度相关的。因此，CNN 就无需使用所有像素单元之间的一对一连接（大多数神经网络都会这么做），而可以使用分组的局部连接。第二，CNN 架构依赖于特征共享，因此每个通道（即输出特征图）是在所有位置使用同一个过滤器进行卷积而生成的。


Input

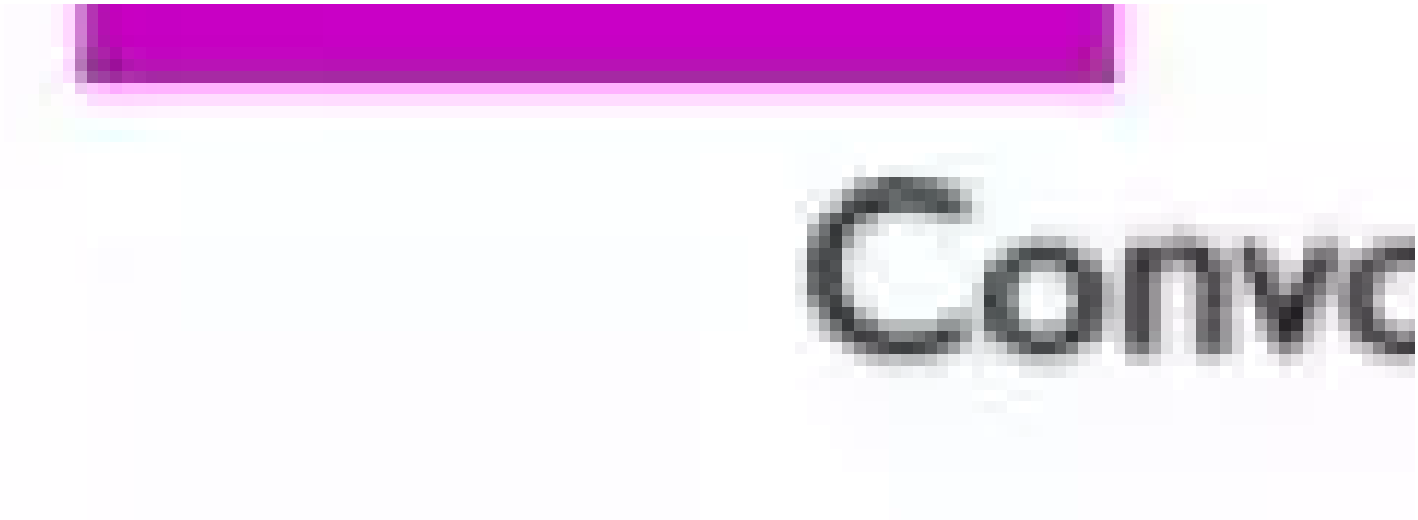


图 2.5：标准卷积网络的结构的示意图，图来自 [93]

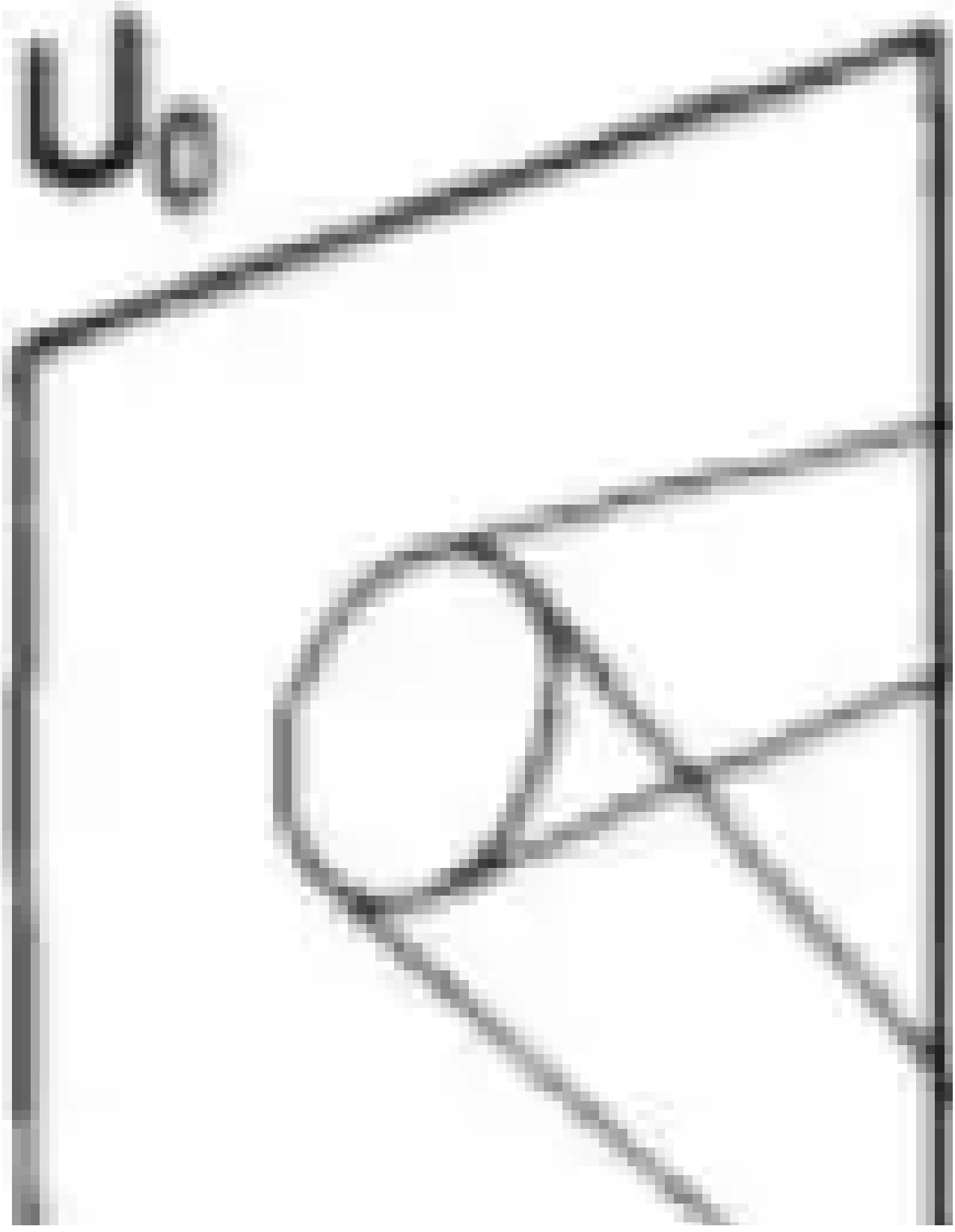
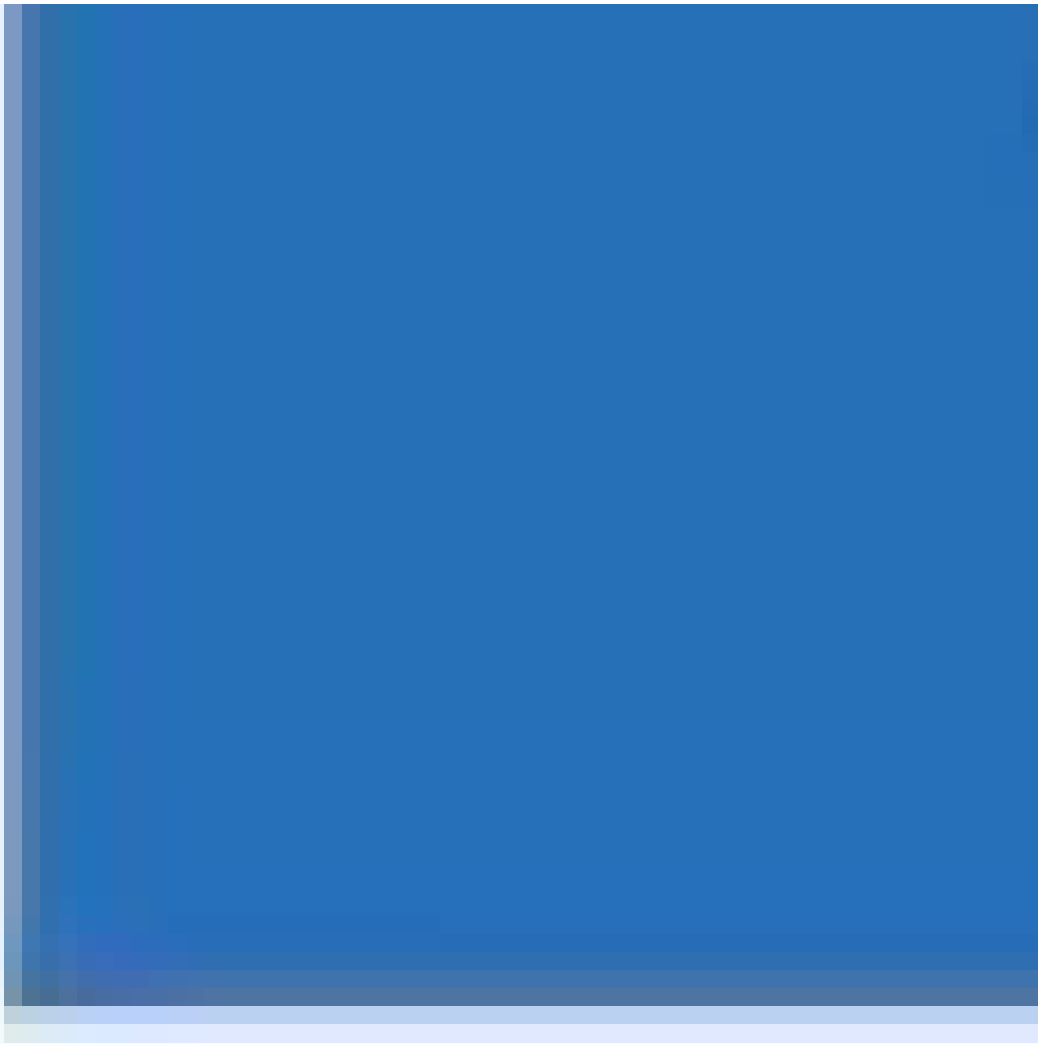
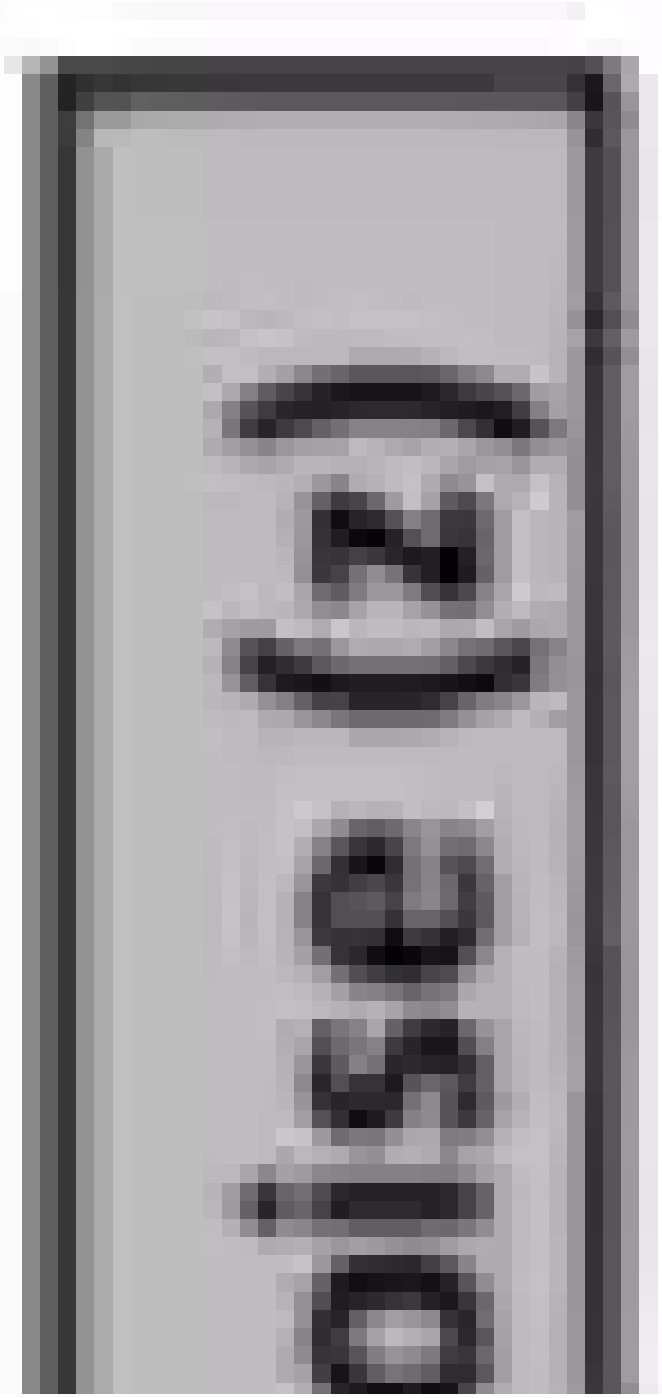


图 2.6：Neocognitron 的结构示意图，图来自 [49]

2.1.4 生成对抗网络

典型的生成对抗网络（GAN）由两个互相竞争的模块或子网络构成，即：生成器网络和鉴别器网络。





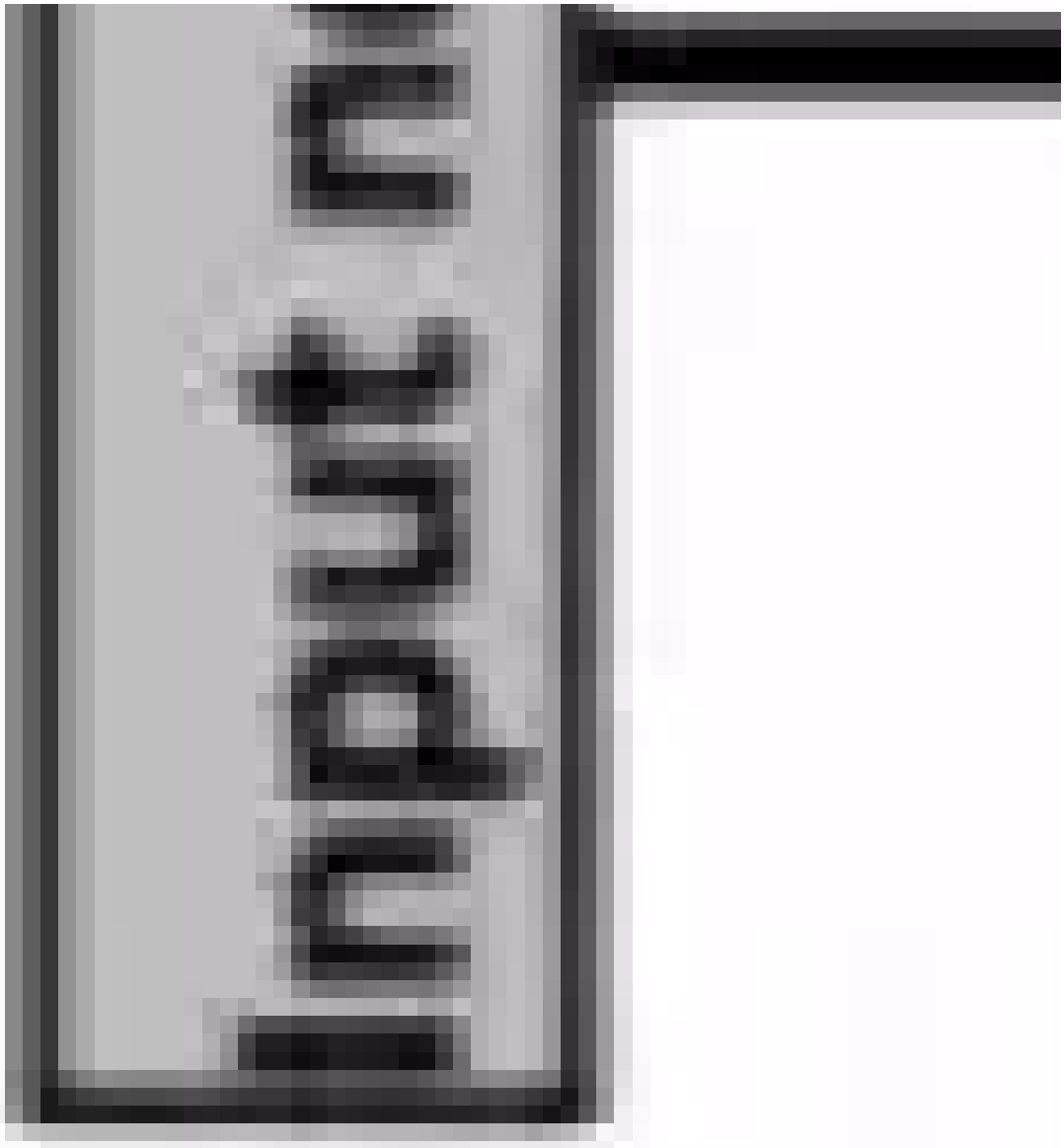


图 2.7：生成对抗网络的一般结构的示意图

2.1.5 多层网络的训练

如前面讨论的一样，多种多层架构的成功都很大程度上取决于它们的学习过程的成功。其训练过程通常都基于使用梯度下降的误差的反向传播。由于使用简单，梯度下降在训练多层架构上有广泛的应用。

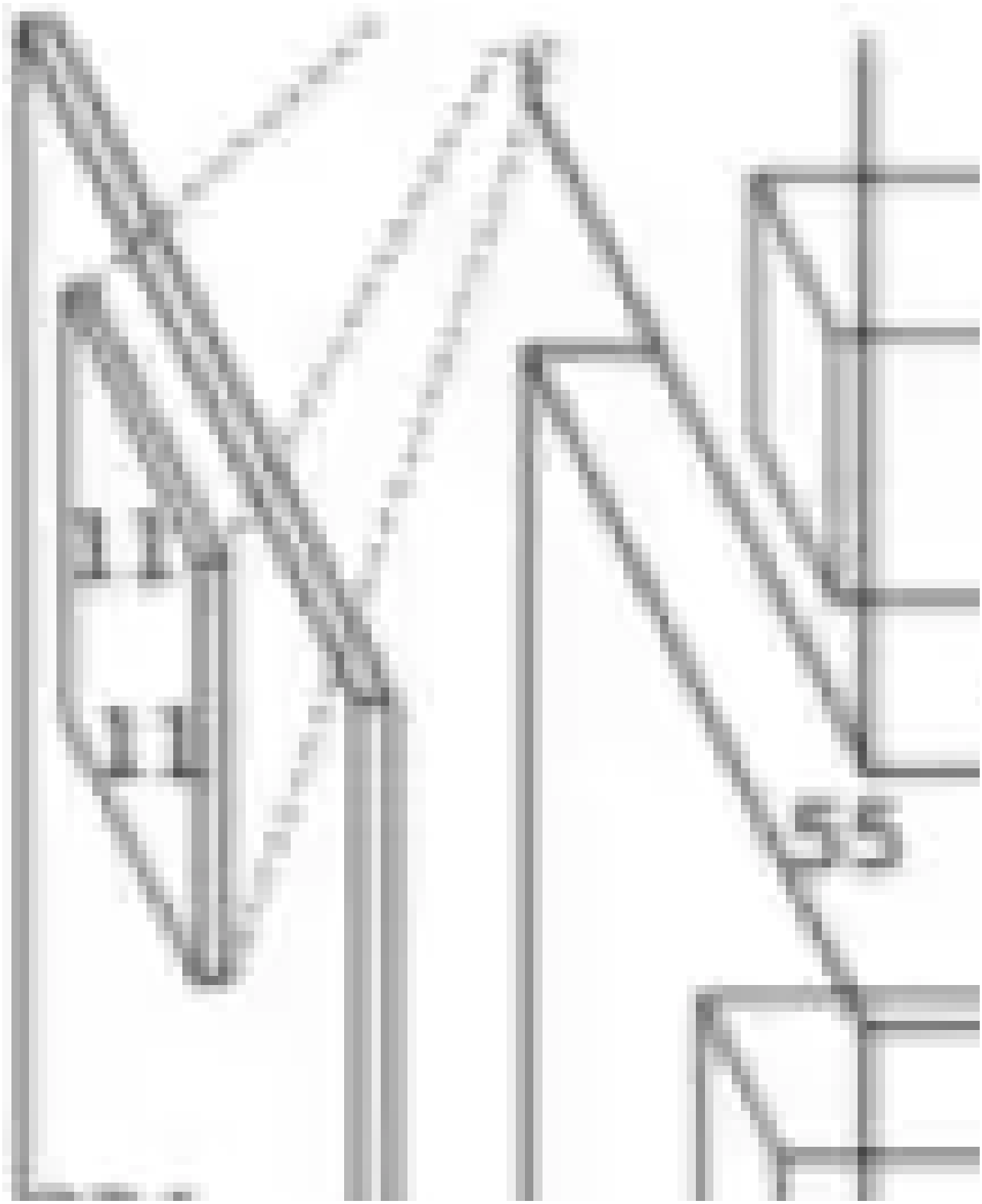
2.1.6 简单说说迁移学习

使用多层架构提取的特征在多种不同数据集和任务上的适用性可以归功于它们的分层性质，表征会在这样的结构中从简单和局部向抽象和全局发展。因此，在其层次结构中的低层级提取的特征往往是多种不同任务共有的特征，因此使得多层结构更容易实现迁移学习。

2.2 空间卷积网络

理论上而言，卷积网络可以应用于任意维度的数据。它们的二维实例非常适用于单张图像的结构，因此在计算机视觉领域得到了相当大的关注。有了大规模数据集和强大的计算机来进行训练之后，CNN 近来在多种不同任务上的应用都出现了迅猛增长。本节将介绍为原来的 LeNet 引入了相对新颖的组件的比较突出的 2D CNN 架构。

2.2.1 CNN 近期发展中的关键架构



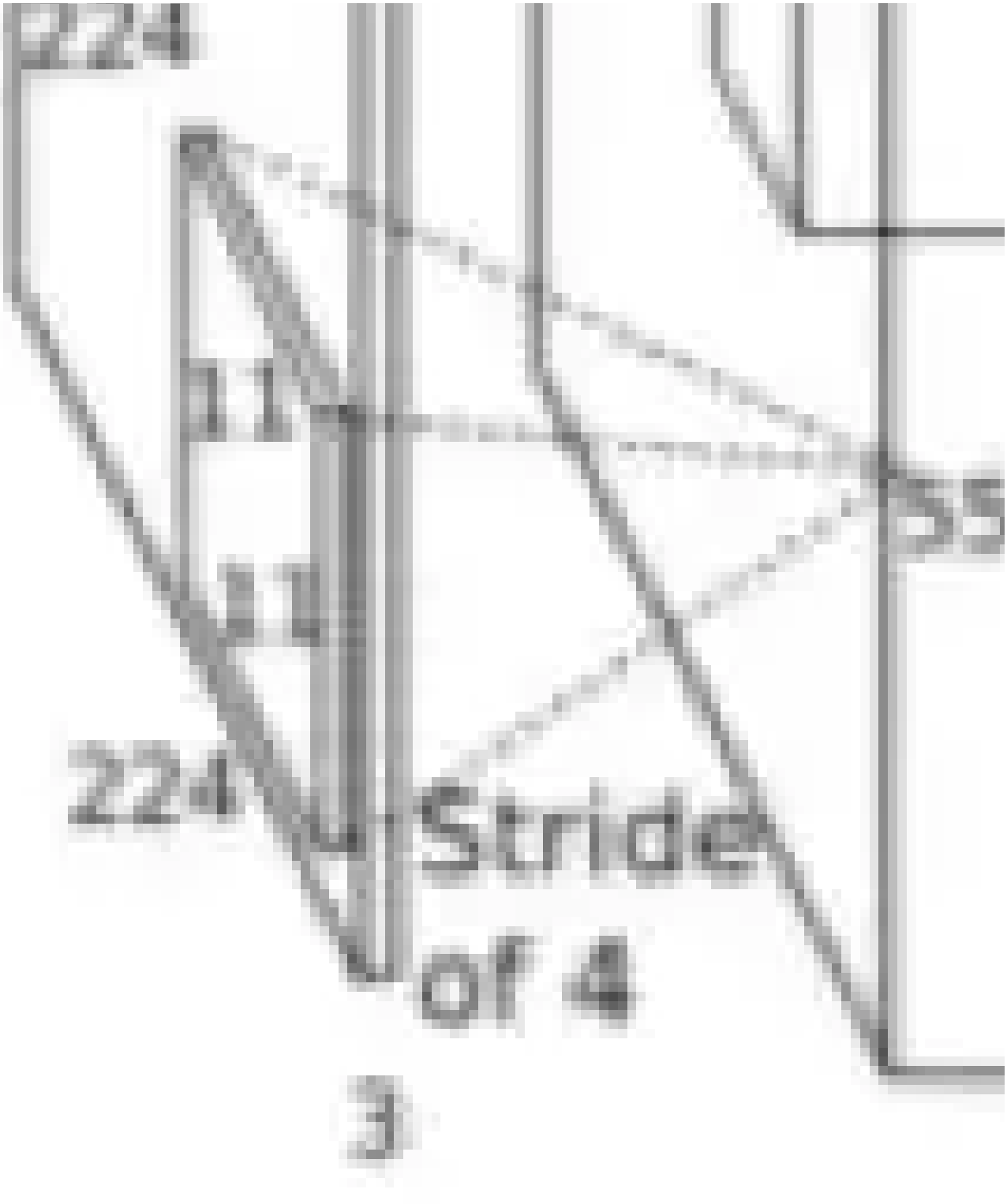


图 2.8：AlexNet 架构。需要指出，虽然从图上看这是一种有两个流的架构，但实际上这是一种单流的架构，这张图只是说明 AlexNet 在 2 个不同 GPU 上并行训练的情况。图来自 [88]

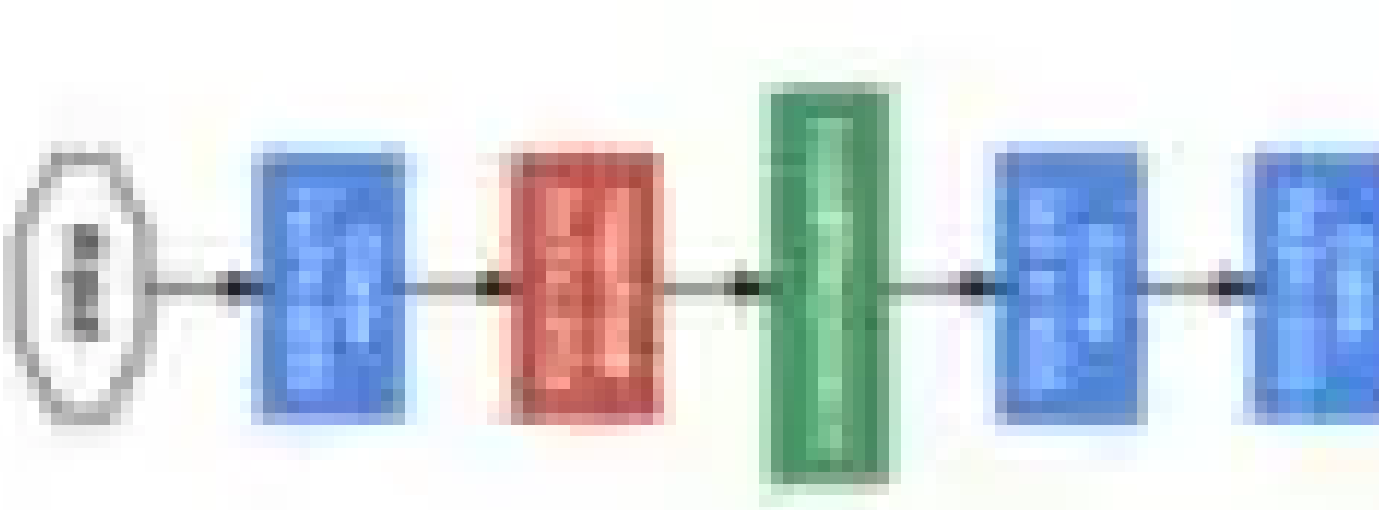
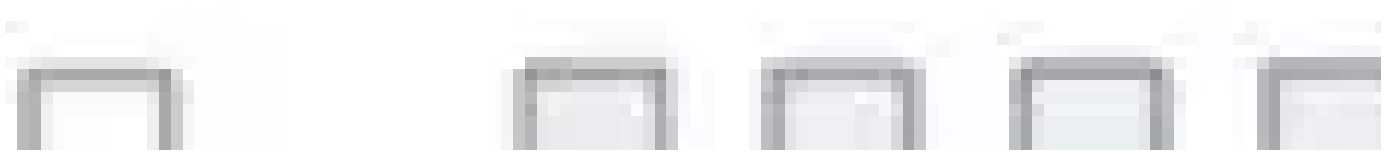


图 2.9：GoogLeNet 架构。（a）典型的 inception 模块，展示了顺序和并行执行的操作。（b）由层叠的许多 inception 模块构成的典型 inception 架构的示意图。图来自 [138]



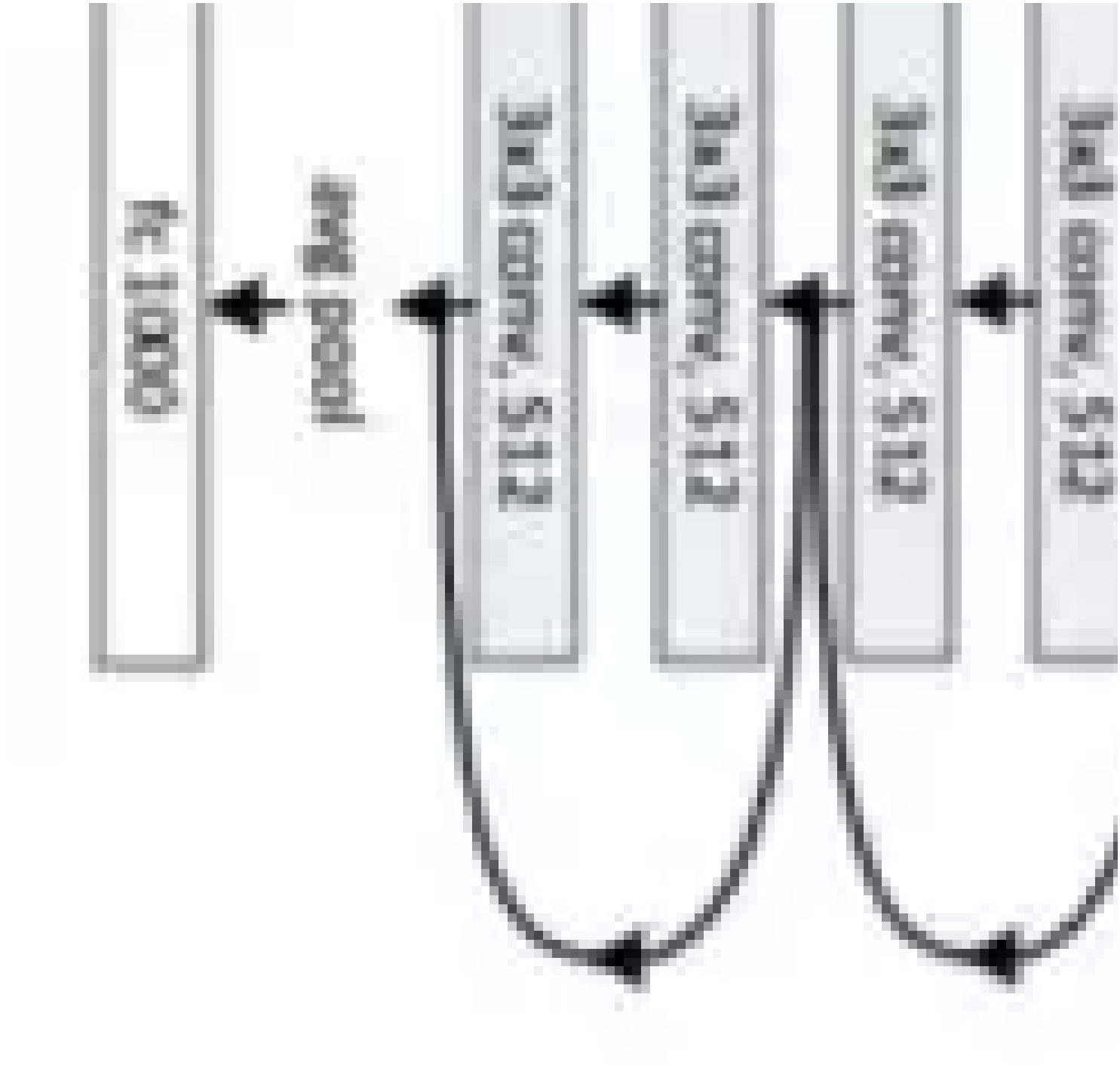


图 2.10：ResNet 架构。（a）残差模块。（b）由层叠的许多残差模块构成的典型 ResNet 架构示意图。图来自 [64]

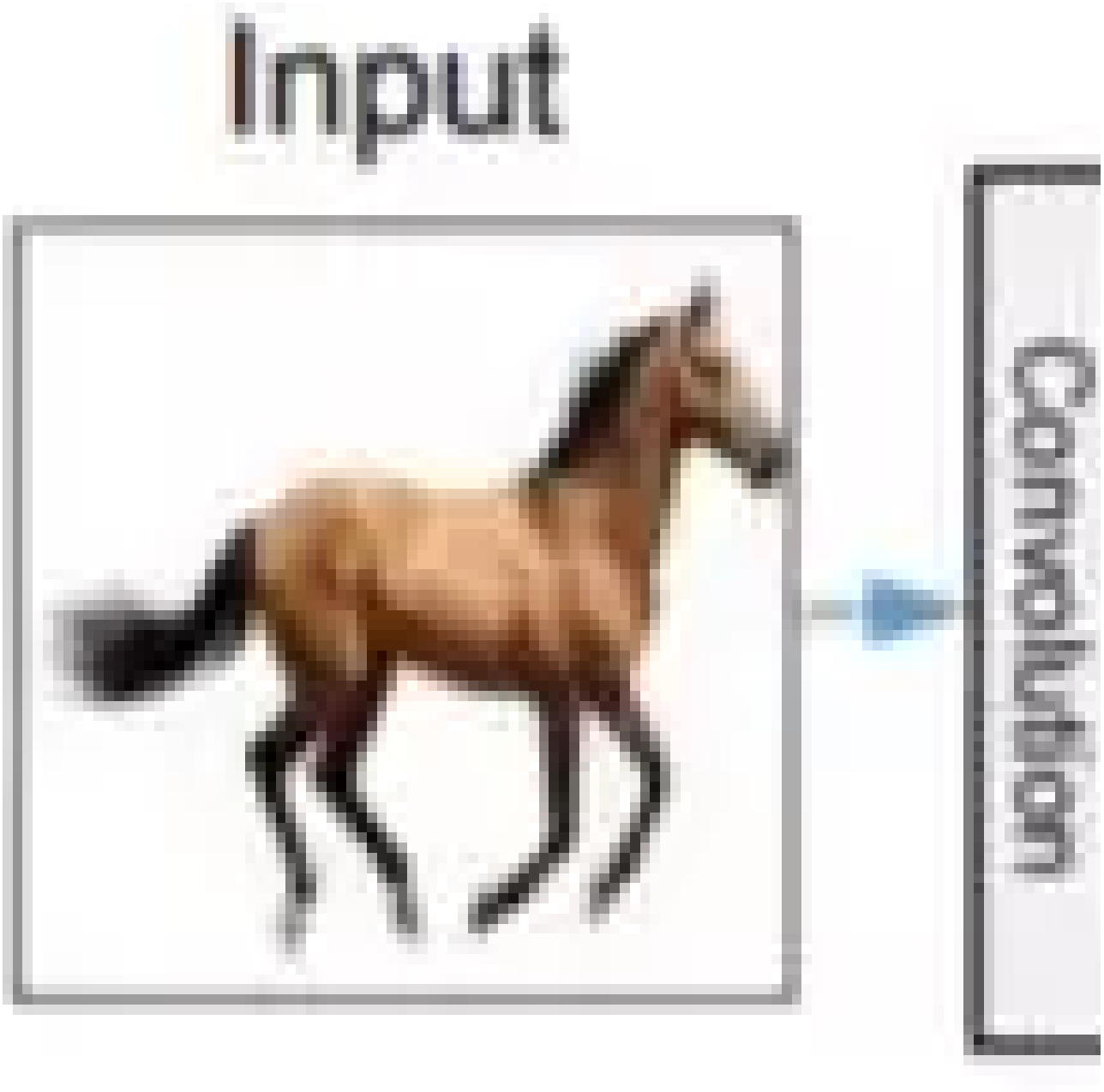

Input
Convolution

图 2.11：DenseNet 架构。（a）dense 模块。（b）（b）由层叠的许多 dense 模块构成的典型 DenseNet 架构的示意图。图来自 [72]

2.2.2 实现 CNN 的不变性

使用 CNN 的一大难题是需要非常大的数据集来学习所有的基本参数。甚至拥有超过 100 万张图像的 ImageNet 等大规模数据集在训练特定的深度架构时仍然被认为太小。满足这种大数据集要求的一种方法是人工增强数据集，具体做法包括对图像进行随机翻转、旋转和抖动（jittering）等。这些增强方法的一大优势是能让所得到的网络在面对各种变换时能更好地保持不变。

2.2.3 实现 CNN 的定位

除了识别物体等简单的分类任务，CNN 近来也在需要精准定位的任务上表现出色，比如形义分割和目标检测。

2.3 时空卷积网络

使用 CNN 为各种基于图像的应用带来了显著的性能提升，也催生了研究者将 2D 空间 CNN 扩展到视频分析的 3D 时空 CNN 上的兴趣。一般而言，文献中提出的各种时空架构都只是试图将空间域 (x,y) 的 2D 架构扩展到时间域 (x, y, t) 中。在基于训练的时空 CNN 领域存在 3 种比较突出的不同架构设计决策：基于 LSTM 的 CNN、3D CNN 和 Two-Stream CNN。

2.3.1 基于 LSTM 的时空 CNN

基于 LSTM 的时空 CNN 是将 2D 网络扩展成能处理时空数据的一些早期尝试。它们的操作可以总结成图 2.16 所示的三个步骤。第一步，使用一个 2D 网络处理每一帧，并从这些 2D 网络的最后一层提取出特征向量。第二步，将这些来自不同时间步骤的特征用作 LSTM 的输入，得到时间上的结果。第三步，再对这些结果求平均或线性组合，然后再传递给一个 softmax 分类器以得到最终预测。

2.3.2 3D CNN

这种突出的时空网络是将 2D CNN 最直接地泛化到图像时空域中。它直接处理 RGB 图像的时间流，并通过应用所学习到的 3D 卷积过滤器来处理这些图像。

2.3.3 Two-Stream CNN

这种类型的时空架构依赖于一种双流式（two-stream）的设计。标准的双流式架构是采用两个并行通路——一个用于处理外观，另一个用于处理运动；这种方法类似于生物视觉系统研究中的双流式假设。

2.4 整体讨论

需要重点指出的是，尽管这些网络在很多计算机视觉应用上都实现了很有竞争力的结果，但它们的主要缺点仍然存在：对所学习到的表征的确切本质的理解很有限、依赖于大规模数据训练集、缺乏支持准确的表现边界的能力、网络超参数选择不清晰。

3 理解 CNN 的构建模块

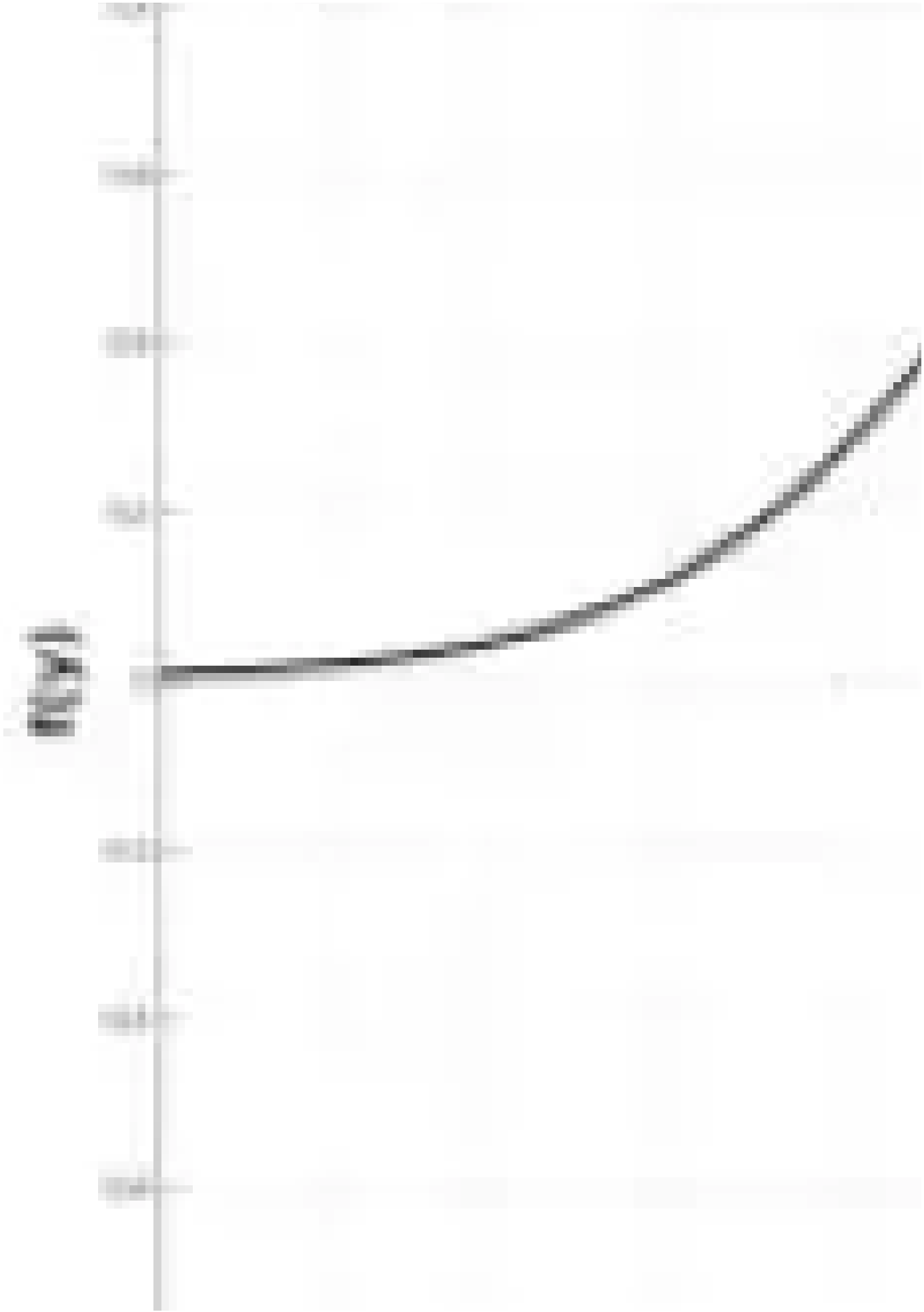
鉴于 CNN 领域存在大量悬而未决的问题，本章将介绍典型卷积网络中每种处理层的作用和意义。为此本章将概述在解决这些问题上最突出的工作。尤其值得一提的是，我们将从理论和生物学两个角度来展示 CNN 组件的建模方式。每种组件的介绍后面都总结了我们当前的理解水平。

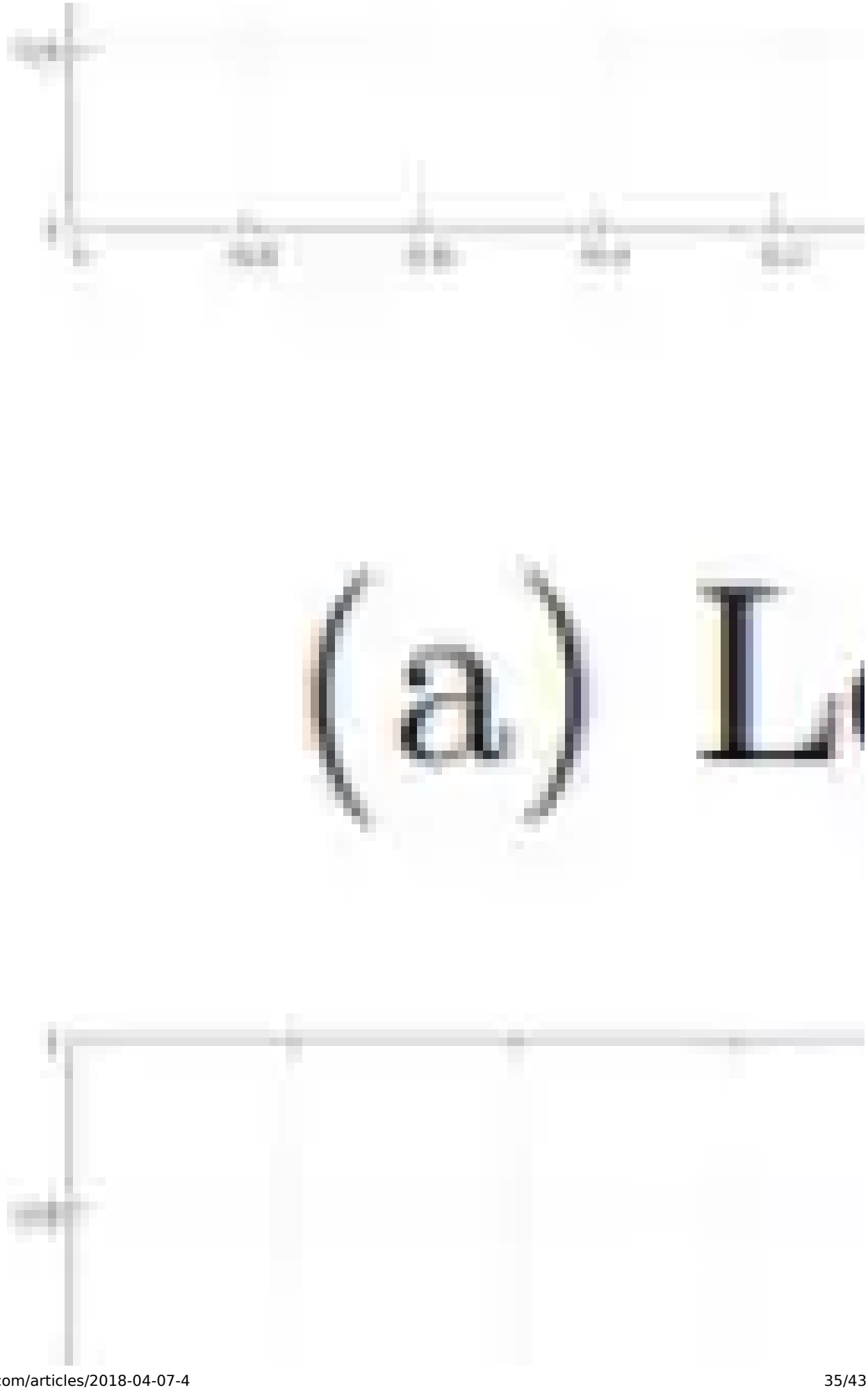
3.1 卷积层

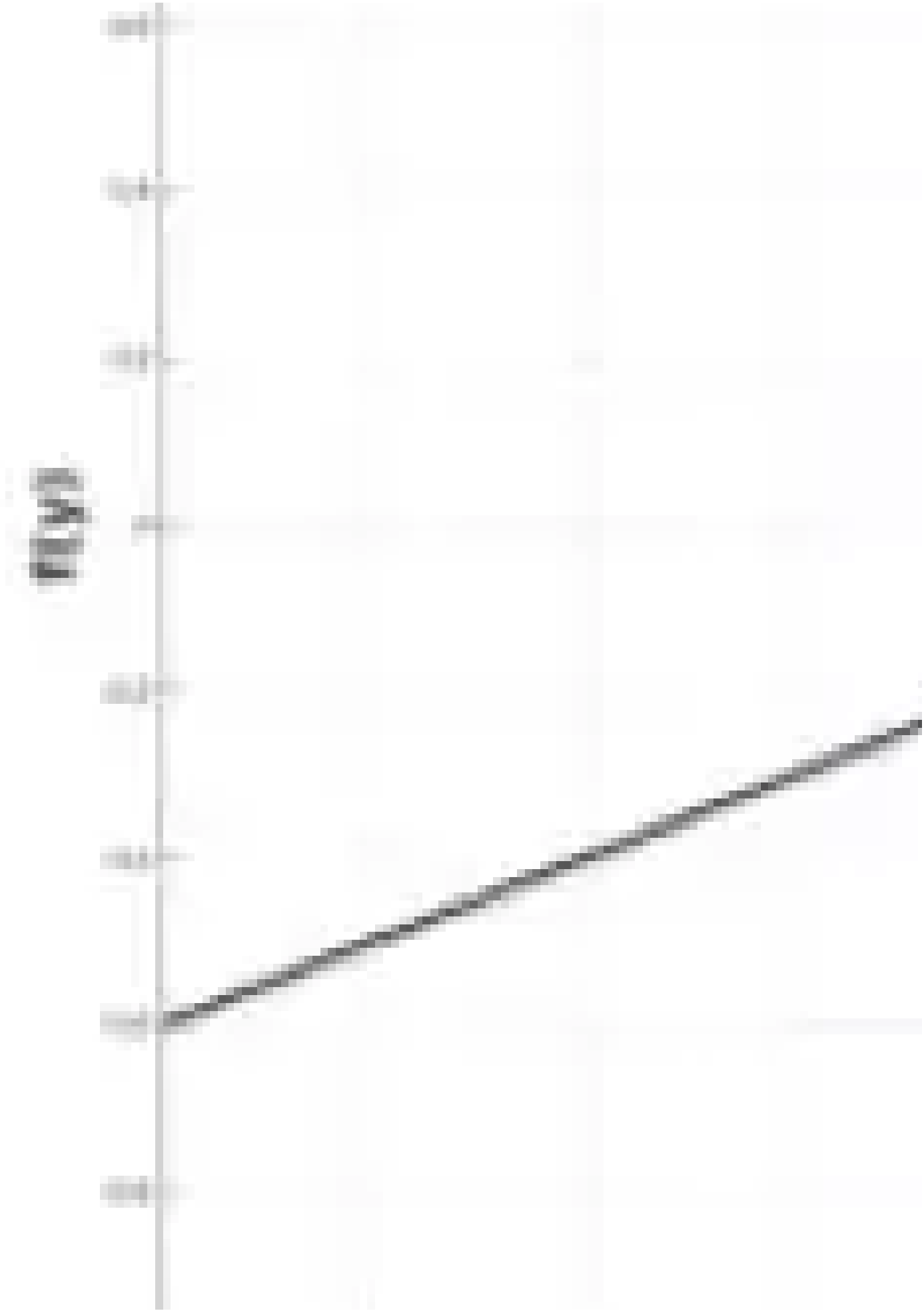
卷积层可以说是 CNN 架构中最重要的步骤之一。基本而言，卷积是一种线性的、平移不变性的运算，其由在输入信号上执行局部加权的组合构成。根据所选择的权重集合（即所选择的点扩散函数（point spread function））的不同，也将揭示出输入信号的不同性质。在频率域中，与点扩散函数关联的是调制函数——说明了输入的频率组分通过缩放和相移进行调制的方式。因此，选择合适的核（kernel）对获取输入信号中所包含的最显著和最重要的信息而言至关重要，这能让模型对该信号的内容做出更好的推断。本节将讨论一些实现这个核选择步骤的不同方法。

3.2 整流

多层网络通常是高度非线性的，而整流（rectification）则通常是将非线性引入模型的第一个处理阶段。整流是指将点方面的非线性（也被称为激活函数）应用到卷积层的输出上。这一术语借用自信号处理领域，其中整流是指将交流变成直流。这也是一个能从生物学和理论两方面都找到起因的处理步骤。计算神经科学家引入整流步骤的目的是寻找能最好地解释当前神经科学数据的合适模型。另一方面，机器学习研究者使用整流的目的是为了让模型能更快和更好地学习。有趣的是，这两个方面的研究者往往都认同这一点：他们不仅需要整流，而且还会殊途同归到同一种整流上。




(a)



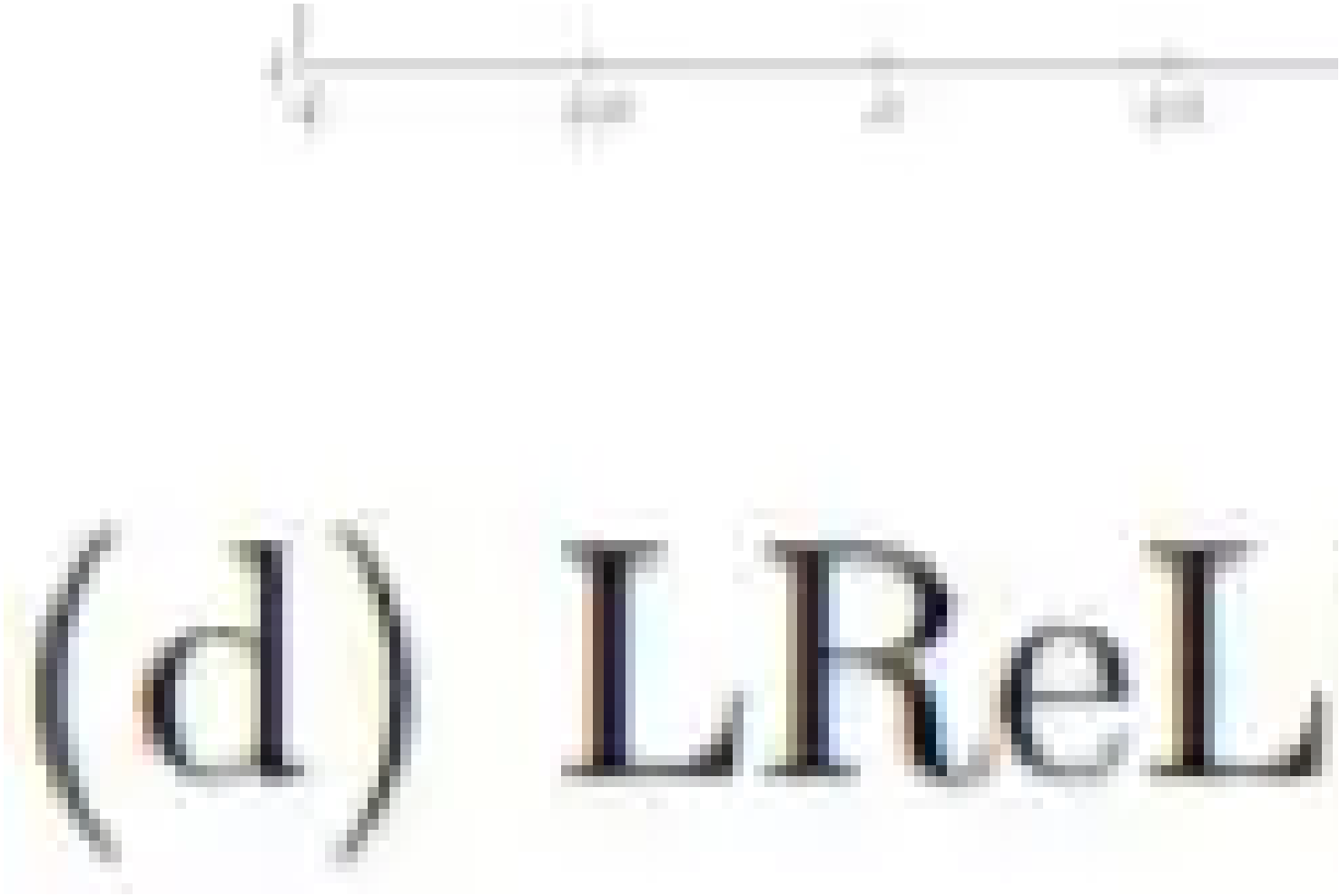


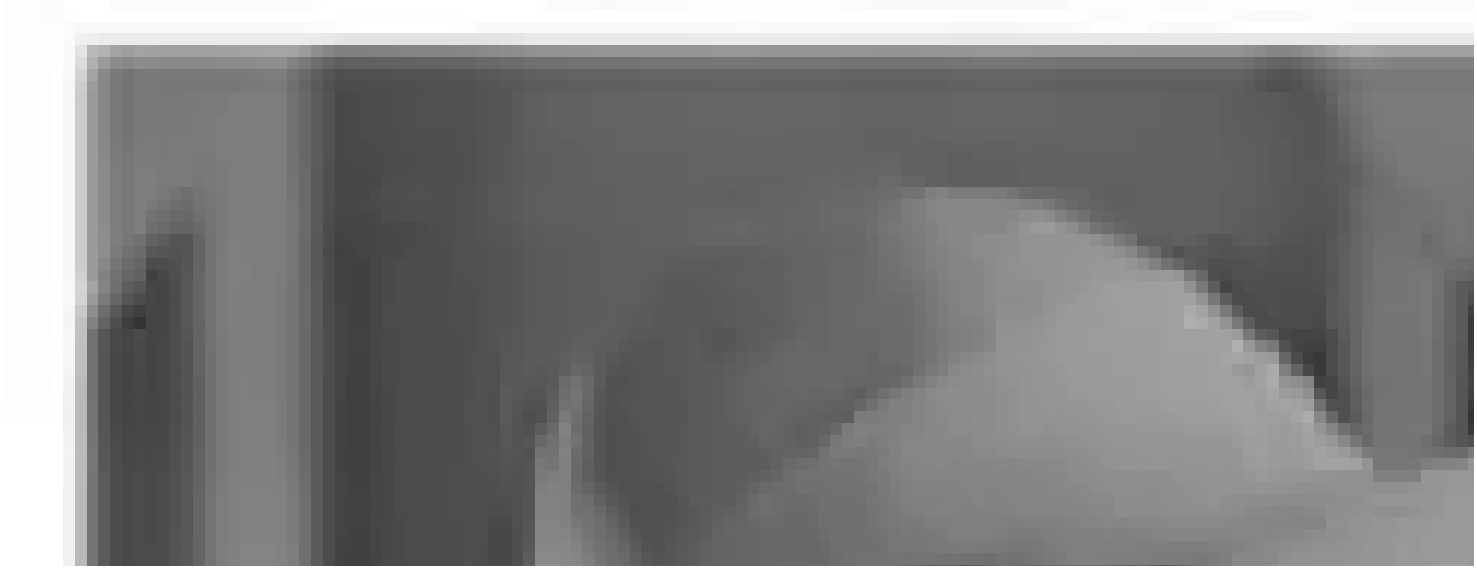
图 3.7：多层网络的文献中所使用的非线性整流函数

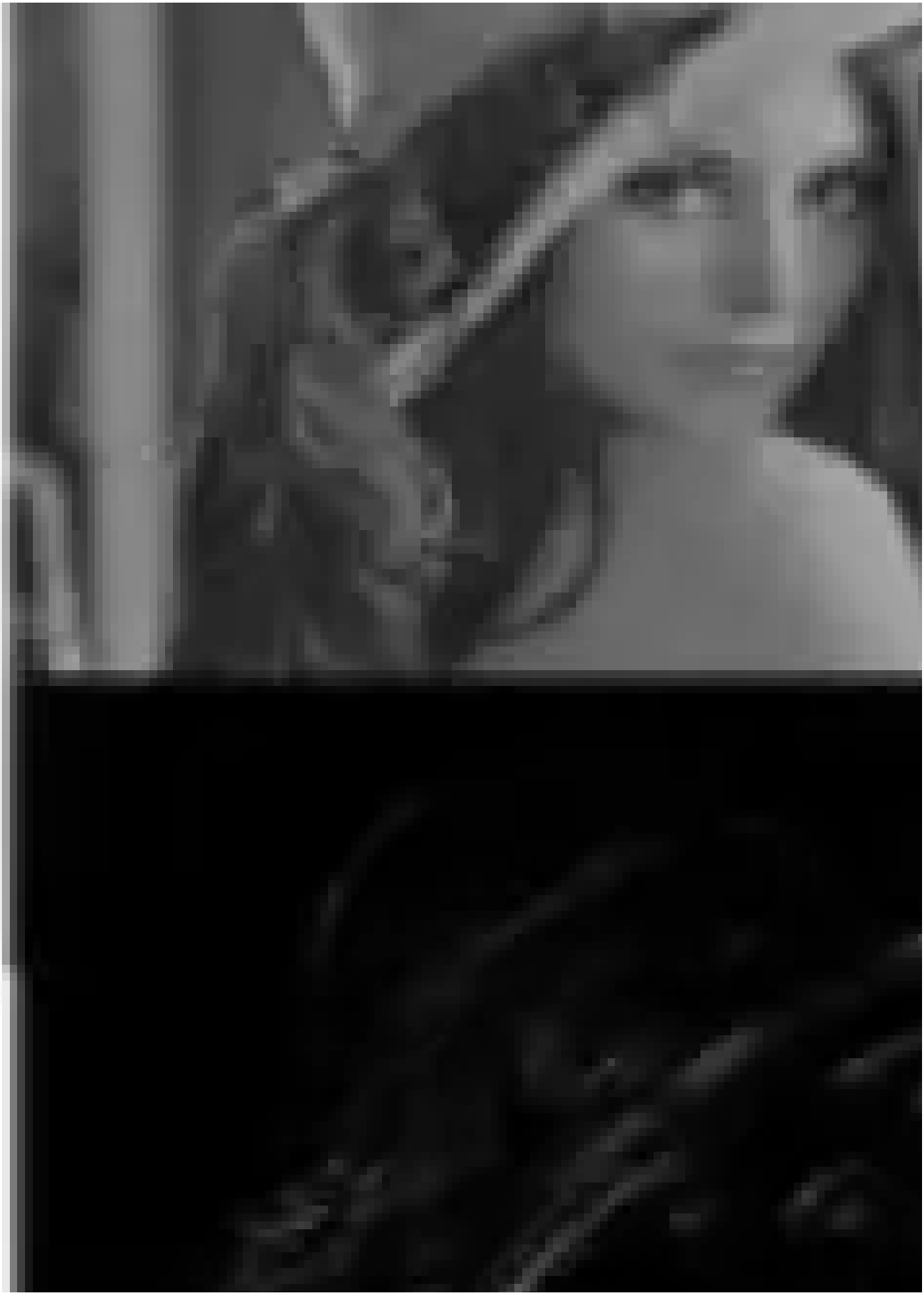
3.3 归一化

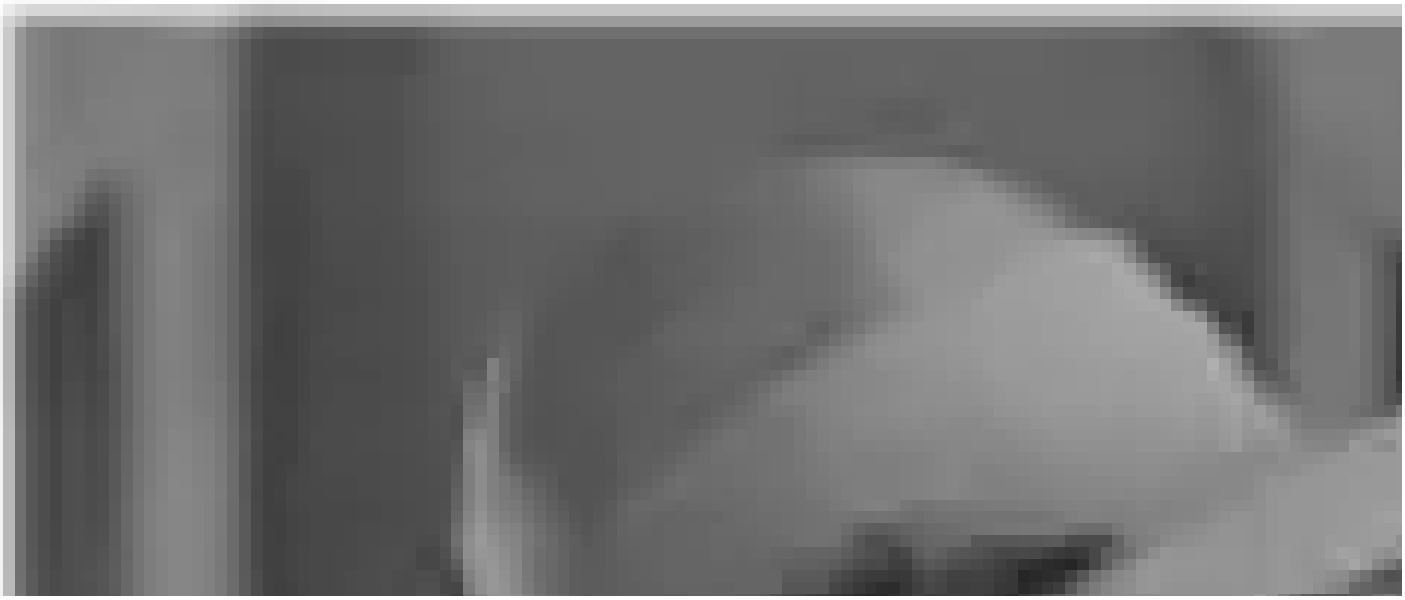
正如前面提到的，由于这些网络中存在级联的非线性运算，所以多层架构是高度非线性的。除了前一节讨论的整流非线性，归一化（normalization）是 CNN 架构中有重要作用的又一种非线性处理模块。CNN 中最广泛使用的归一化形式是所谓的 Divisive Normalization（DN，也被称为局部响应归一化）。本节将介绍归一化的作用并描述其纠正前两个处理模块（卷积和整流）的缺点的方式。同样，我们会从生物学和理论两个方面讨论归一化。

3.4 池化

不管是生物学启发的，还是纯粹基于学习的或完全人工设计的，几乎所有 CNN 模型都包含池化步骤。池化运算的目标是为位置和尺寸的改变带来一定程度的不变性以及在特征图内部和跨特征图聚合响应。与之前几节讨论的三种 CNN 模块类似，池化在生物学和理论研究上都具有支持。在 CNN 网络的这个处理层上，主要的争论点是池化函数的选择。使用最广泛的两种池化函数分别是平均池化和最大池化。本节将探索相关文献中描述的各种池化函数的优点和缺点。







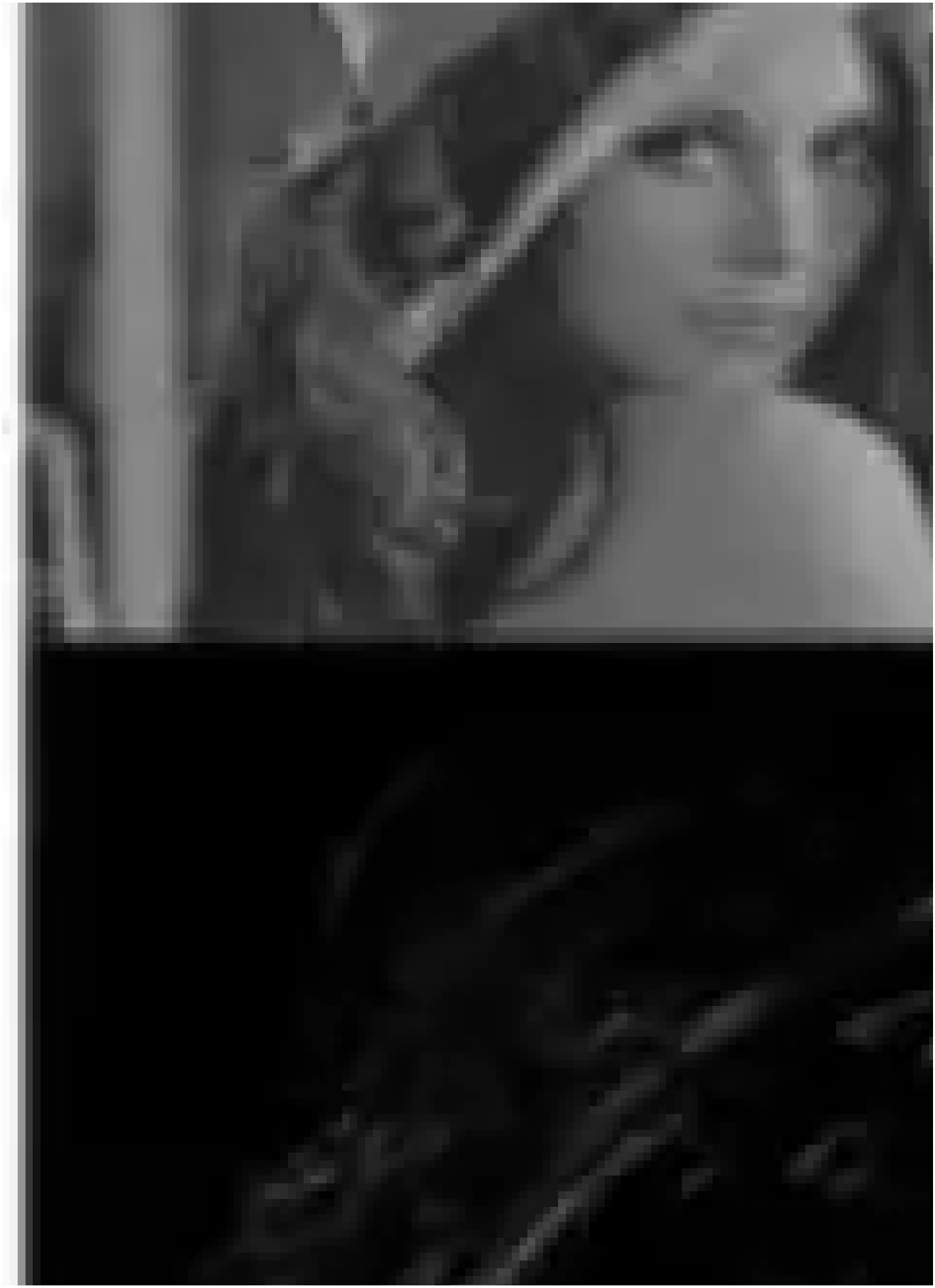


图 3.10：平均池化和最大池化在 Gabor 滤波后的图像上的比较。（a）展示了不同尺度的平均池化的效果，其中（a）中上面一行是应用于原始灰度值图像的结果，（a）中下面一行是应用于 Gabor 滤波后的图像上的结果。平均池化能得到灰度值图像的更平滑的版本，而稀疏的 Gabor 滤波后的图像则会褪色消散。相对而言，（b）给出了不同尺度的最大池化的效果，其中（b）中上面一行是应用于原始灰度值图像的结果，（b）中下面一行是应用于 Gabor 滤波后的图像上的结果。这里可以看到，最大池化会导致灰度值图像质量下降，而 Gabor 滤波后的图像中的稀疏边则会得到增强。图来自 [131]

4 当前状态

对 CNN 架构中各种组件的作用的论述凸显了卷积模块的重要性，这个模块很大程度上负责了在网络中获取最抽象的信息。相对而言，我们对这个处理模块的理解却最少，因为这需要最繁重的计算。本章将介绍在尝试理解不同的 CNN 层所学习的内容上的当前趋势。同时，我们还将重点说明这些趋势方面仍有待解决的问题。

4.1 当前趋势

尽管各种 CNN 模型仍继续在多种计算机视觉应用中进一步推进当前最佳的表现，但在理解这些系统的工作方式和如此有效的原因上的进展仍还有限。这个问题已经引起了很多研究者的兴趣，为此也涌现出了很多用于理解 CNN 的方法。一般而言，这些方法可以分成三个方向：对所学习到的过滤器和提取出的特征图进行可视化、受理解视觉皮层的生物学方法启发的 ablation study、通过向网络设计中引入分析原理来最小化学习过程。本节将简要概述其中每种方法。

4.2 仍待解决的问题

基于上述讨论，基于可视化的方法存在以下关键研究方向：

- 首要的一点：开发使可视化评估更为客观的方法是非常重要的，可以通过引入评估所生成的可视化图像的质量和/或含义的指标来实现。

- 另外，尽管看起来以网络为中心的可视化方法更有前景（因为它们在生成可视化结果上不依赖网络自身），但似乎也有必要标准化它们的评估流程。一种可能的解决方案是使用一个基准来为同样条件下训练的网络生成可视化结果。这样的标准化方法反过来也能实现基于指标的评估，而不是当前的解释性的分析。

- 另一个发展方向是同时可视化多个单元以更好地理解处于研究中的表征的分布式方面，甚至同时还能遵循一种受控式方法。

以下是基于 ablation study 的方法的潜在研究方向：

- 使用共同的系统性组织的数据集，其中带有计算机视觉领域常见的不同难题（比如视角和光照变化），并且还必需有复杂度更大的类别（比如纹理、部件和目标上的复杂度）。事实上，近期已经出现了这样的数据集 [6]。在这样的数据集上使用 ablation study，加上对所得到的混淆矩阵的分析，可以确定 CNN 架构出错的模式，进而实现更好的理解。

- 此外，对多个协同的 ablation 对模型表现的影响方式的系统性研究是很受关注的。这样的研究应该能延伸我们对独立单元的工作方式的理解。

最后，这些受控方法是很有前景的未来研究方向；因为相比于完全基于学习的方法，这些方法能让我们对这些系统的运算和表征有更深入的理解。这些有趣的研究方向包括：

- 逐步固定网络参数和分析对网络行为的影响。比如，一次固定一层的卷积核参数（基于当前已有的对该任务的先验知识），以分析所采用的核在每一层的适用性。这个渐进式的方法有望揭示学习的作用，而且也可用作最小化训练时间的初始化方法。

- 类似地，可以通过分析输入信号的性质（比如信号中的常见内容）来研究网络架构本身的设计（比如层的数量或每层中过滤器的数量）。这种方法有助于让架构达到适宜应用的复杂度。

- 最后，将受控方法用在网络实现上的同时可以对 CNN 的其它方面的作用进行系统性的研究，由于人们重点关注的所学习的参数，所以这方面得到的关注较少。比如，可以在大多数所学习的参数固定时，研究各种池化策略和残差连接的作用。 SYNCED

本文由机器之心编译出品，原文来自arXiv，作者Panda，转载请查看要求，机器之心对于违规侵权者保有法律追诉权。

理论
4

wupan

机器之心编辑

登录后评论

暂无评论~

联系电话：+86 010-57150141

联系邮箱：contact@jiqizhixin.com

返回顶部