基于脉冲神经网络的骨骼检测方法综述

摘要：骨骼检测通过学习骨架关键点的位置和连接方式来实现画面中特定生物骨骼对象的结构化提取。随着深度学习方法和算力的提升，基于卷积神经网络的方法在骨骼检测中取得巨大成功。作为“第三代神经网络”的脉冲神经网络 (SNN) 因其神经元模拟生物神经元的脉冲放电过程，能实现更加高效的类脑计算，在目标检测等多个方面逐渐得到有效应用。因此，有理由认为 SNN 在骨骼检测方面也将有所作为，尤其在节能和高效计算方面。本文对目前脉冲神经网络及其在骨骼检测以及一般的目标检测领域的应用现状，试对其潜在优势进行总结，并对其局限性进行探析和展望。
关键词：骨骼检测；目标检测；计算机视觉；深度学习；脉冲神经网络

Abstract： Bone detection realizes the structured extraction of specific biological bone objects in the picture by learning the position and connection of key points of the skeleton. With the improvement of deep learning methods and computing power, convolutional neural network-based methods have achieved great success in bone detection. As a "third-generation neural network", spiking neural network (SNN) has been gradually effectively applied in many aspects such as object detection because its neuron simulates the pulse discharge process of biological neurons, which can achieve more efficient brain-like computing. Therefore, it is reasonable to assume that SNNs will also make a difference in bone detection, especially in terms of energy saving and efficient computing. In this paper, the current status of spiking neural networks and their application in the field of bone detection and general object detection is summarized, and its potential advantages are summarized, and its limitations are analyzed and prospected.

Keywords： Skeleton Detection; Object Detection; Computer Vision; Deep learning; SNN

目录

1. 引言

计算机视觉是人工智能领域中的重要研究方向之一，其目的是让计算机模拟人类视觉系统，实现对图像和视频的感知、理解和分析。其中，骨骼检测是计算机视觉中的重要问题之一，其目的是在给定的图像或视频中准确地识别和定位人体的关键部位，如骨骼关节和身体部位等。骨骼检测在很多应用场景中都具有重要的应用价值，例如人体姿态估计、动作识别、医学影像分析等领域。

近年来，随着深度学习技术的发展和应用，基于神经网络的骨骼检测方法取得了重大进展。这些方法通常基于深度卷积神经网络(CNN)或循环神经网络(RNN)等模型，结合一些优秀的计算机视觉技术，如图像分割、物体检测等，能够实现更加准确和鲁棒的骨骼检测。此外，随着神经网络计算能力的提高，新兴的脉冲神经网络(SNN)在骨骼检测领域也逐渐得到了应用。

尽管基于神经网络的骨骼检测方法取得了很大的成功，但是其仍然存在一些问题和挑战，如鲁棒性不足、需要大量的数据和计算资源等。因此，如何进一步提高基于人工智能的计算机视觉骨骼检测的准确性和效率，仍然是当前研究的重要课题之一。

脉冲神经网络(SNN)是一种新兴的神经网络模型，其神经元模拟生物神经元的脉冲放电过程，能够实现更加节能和高效的计算。在骨骼检测方面，SNN 也逐渐得到了应用。虽然 SNN 在骨骼检测方面的应用仍处于初级阶段，但是其具有很大的潜力，尤其在节能和高效计算方面，有望在未来得到更广泛的应用。

本文对目前脉冲神经网络及其在骨骼检测以及一般的目标检测领域的应用现状，并试对其潜在优势进行总结，对其局限性进行探析和展望。

2. 计算机视觉骨骼检测的神经网络

2.1 神经网络介绍

作为当前最热门的计算机科学技术之一，人工神经网络(ANN, Artificial Neural Network)的最初的想法和模型可以追溯到上世纪 40 年代中期，最初是由人类神经元系统中信号传递模式的启发，并结合计算机技术而发展出的一系列人工智能算法。[1]而在此之前，就已经产生了所谓人工神经元和类神经网络的理念，即通过从信息处理角度对人脑神经元网络的激活与抑制两种状态的输入和输出过程进行抽象，获得对单个神经元的工作状况进行抽象的模拟。而将这种人工神经元按一定的组织方式连接成相应的网状结构，即人工神经网络。一般而言，人工神经网络具有较强大的学习能力，这主要指的是它能通过利用外界有监督信息来实现内部结构中参数的调整，是一种自适应系统。而这个利用有监督的外界信息来调整内部参数的过程，就被称为“学习”。

在现代计算机被发明后，随着计算机算力的不断提高，通过模仿人脑运作方式的过程来挖掘数据中的潜在关系的理念逐渐由粗略的仿生学想法转化为现实的计算模型。神经网络基于对生物系统中信息处理进行建模，试图建立一种人工的“智能”。然而，与冯诺依曼体系不同，神经网络计算没有将内存和处理分开。

迄今为止，人工神经网络的发展已经经历了三代沿革[2]。起初，由于算力的限制和算法设计的局限性，神经网络仅仅局限于单个神经元层次。其本身的局限性也十分明显，这是因为，由于不涉及非线性层，初代神经元只有简单的输入、

线性函数和输出，而不包含非线性层，这意味着它无法处理非线性函数的拟合问题。严格来说，它只能被称为“神经元”，而并未形成真正的“网络”。随着算力的增进和算法的创新，在第一代神经网络的基础上，又涌现出了第二代神经网络，它通过引入隐藏层来实现多层网络结构，并引入非线性激活层来解决非线性拟合问题。随后，大量的研究关注了加深神经网络层数这一问题，并最终形成了所谓“深度神经网络”为核心的深度学习理论和方法，使神经网络方法在多个领域中展现出强大能力并真正发扬光大。但总得来说，前两代神经网络都停留在感知机层次。

2.1.1 第一代神经网络

第一代人工神经网络是以 McCulloch-Pitts 神经元为代表的阈值神经元。这是一个概念上很简单的模型，它的模型仅包含两层神经元，即输入层输出层。它的层间连接完全采取线性结构。因此，它不能解决非线性的问题。最典型地，初代神经元无法拟合异或函数。

以基于 McCulloch-Pitt 神经元的感知机为例，它仅由输入和输出组成，该个神经元的计算模型是执行阈值运算并输出数字 0 或 1。它具有一个二进制的输出：如果神经元的加权输入总和超过阈值，神经元就会输出激活信号。但是，由于其简单的结构，导致其无法解决“异或”问题等非线性函数的拟合问题。因此，第一代神经网络又称为感知器。

尽管第一代神经元只产生二进制的输出，但它们已逐渐被用于组成稍复杂的人工神经网络系统，这被称为“多层感知机”。例如，具有单个隐藏层的 Hopfield 网络就可以计算具有布尔输出的任何函数。这些网络又被称为数字计算通用网络。这类多层感知机即成为第二代神经网络的前身。例如，20 世纪 60 年代，就发展出了包含 3 层结构(即输入层、隐藏层和输出层)的神经网络[3]。第一代神经网络的单个神经元示意图如图 1。

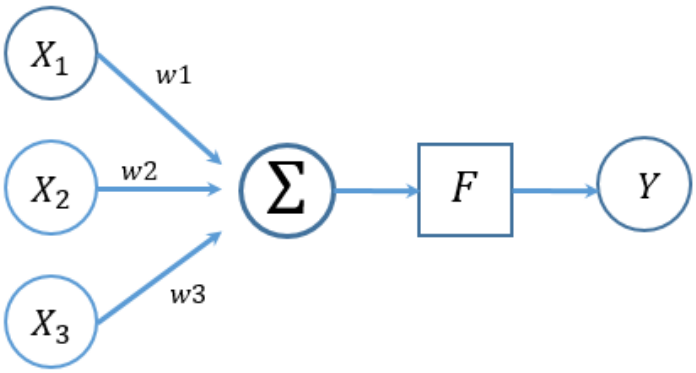


图 1 第一代神经网络：单个神经元的示意图

2.1.2 第二代神经网络

组成第二代神经网络中的基本单位神经元已经不再是阶跃函数或阈值函数，而是使用连续激活函数(例如 Sigmoid 激活函数、双曲正切激活函数等)。连续激活函数的使用，使第二代神经网络具有新的特征，即可以适用于模拟输入和输出。与第一代神经网络相比，该代神经网络在原感知机的输入层和输出层之间添加了至少一层的隐藏层，并在其中引入一些非线性的结构，从而解决了之前无法模拟异或逻辑的缺陷[4]。

通过在线性的全连接层之间加入非线性的激活函数(如 sigmoid、双曲正切函数、ReLU 等)，以及添加了隐藏层第二代神经网络首次实现了网络中的非线性，成功地解决了非线性的问题。而连续激活函数的使用本身也很好地解决了网络中参数不易自动更新的问题。基于反向传播(back propagation)机制，其网络能够根据损失函数的梯度方向计算一组连续的权重更新值。也正是基于此，神经网络实现了自动的参数更新。例如，20 世纪 80 年代，由 Rumelhart、Williams 等提出的多层感知机(MLP)，即是第二代神经网络的典型代表。第二代神经网络因此又被称为多层感知机[5][6]。

多层感知机的想法很自然地引出人们关于继续加深网络层数的广泛思考——毕竟，既然多层网络好于单层网络，那么，随着层数的加深，网络应该具有更复杂的参数结构，进而也更可能拥有更强大的学习能力。然而，研究人员发现，随着多层感知机中层数的增加，起初，神经网络的学习能力逐渐提升，例如，三层网络就可以很好地学习到手写阿拉伯数字的特征[7]。但是，随着层数的继续增加，其学习表现却逐渐下降，甚至不及浅层网络[8]。第二代神经网络的示意图如图 2。

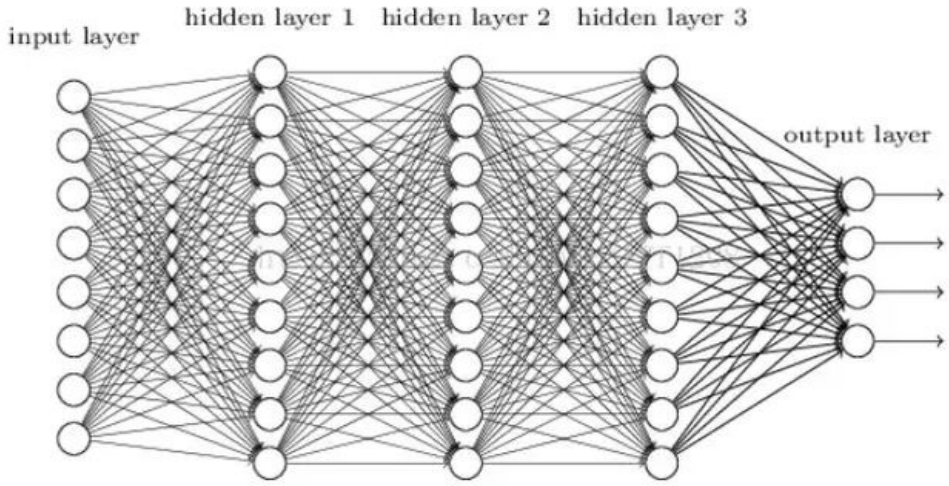


图 2 典型的二代神经网络：多层感知机 MLP(Multi-layer Perceptron)

随着研究的进一步深入，人们发现，制约神经网络进一步加深的主要原因在于，反向传播过程中，随着网络层数加深，各隐藏层中，越接近输入层的学习速率往往越低，这导致靠近输入层的隐藏层权值更新缓慢，甚至更新停滞。即随着隐藏层数目的增加，分类准确率反而下降。这种现象被称为梯度消失。另一方面，优化函数愈发容易出现局部最优解的现象，这也导致了深层网络往往难以训练。

直到 2006 年，Hinton 采取了一种被称为无监督预训练的方法，成功克服了梯度消失问题，使深度神经网络变得可训练将隐含层发展到 7 层 [9]。从此，神经网络真正意义上有了“深度”，由此揭开了深度学习的浪潮，神经网络方法作为机器学习的主流范式开始正式兴起。2014 年，何凯明等人采取被称为残差网络的新型并联网络结构来构建超深网络的努力获得了巨大成功。由此，第二代神经网络中，多层感知机的层数不断加深。这种依赖加深网络层数来获取更强学习能力的范式被称为深度学习。

当前，第二代神经网络业已成为最广为使用的神经网络算法。几乎所有的主流深度神经网络(Deep Neural Networks，DNN)，都可以被包括于第二代神经网络

的范畴[10]。一般而言，在不考虑梯度消失问题的情况下，深度神经网络的层数很大程度上与其学习能力呈正相关。这也是深度学习在近些年广为流传的根本原因。本质上

也有一些观点倾向于把深度学习为代表的这一代新的神经网络视作第三代神经网络[11]。但实际上，更主流的观点认为，深度学习所代表的这一类神经网络方法仍未超出第二代多层感知机神经网络的范畴[12][13][14]。尤其是，随着深度的增加，深度学习本身也存在着参数量过大、难以拟合、能耗过高、硬件算力瓶颈等诸多棘手问题。这些问题已经逐渐成为阻碍深度学习进一步发展的瓶颈。而这些问题，究其原因，与第二代神经网络，即感知机的定义方式是直接相关的。因此说，深度学习本质上仍然处于第二代神经网络的范畴之内，而前两代神经网络，则又都停留在感知机层次。

2.1.3 第三代神经网络——脉冲神经网络

2.1.3.1 脉冲神经网络介绍

由前所述，第一代神经网络，即感知机本身存在着无法拟合复杂非线性函数(例如异或运算)的问题。第二代神经网络模型虽然解决了这一问题，但其中难以加入时间信息，并且存在着许多制约神经网络方法进一步发展的问题。这些都为新一代神经网络的发展提供了客观需求。同时，前两代神经网络模型中的问题被寄希望于在第三代神经网络中得到解决。

随着神经网络的进一步发展，脉冲神经网络(Spiking Neural Network，SNN)被提出。1952 年，Alan Lloyd Hodgkin 和 Andrew Huxley 提出了第一个脉冲神经网络模型，这个模型的主要贡献在于创新性地关注了动作电位如何产生和传播的重要问题，这为相关研究开辟了全新的思路[15]。由于该类神经网络具有以“整合放电”(integrate-and-fire)方式传递信息的脉冲神经元，并通过这种脉冲交换信息，因此得名脉冲神经网络。这是一种在模拟神经元方面更加接近实际情况的新式神经网络，且能把时间信息的影响包含在内。脉冲神经网络具有前两代神经网络所不具备的许多优良特点，因为经常被誉为第三代人工神经网络[12]。

具体而言，其与前两代神经网络(统称为 ANN，即 Artificial Neural Network，人工神经网络，下同)的差别主要在于：不同于第一代神经网络(即感知机)和第二代神经网络(即多层感知机和各种深度网路)，脉冲神经网络直接从思路上进行了较为彻底的变革。在脉冲神经网络中，网络中的神经元的膜电位达到某一个特定值的情况下才被激活。通常用微分方程模拟神经元激活水平的动力学。一个输入脉冲会使当前值升高。然后，作用持续一段时间，再逐渐衰退。

SNN 与前两代神经网络(ANN)的区别还体现在信息传递方式。ANN 的输入都是静态的，即时不变的；而 SNN 则基于动态二进制尖峰(作为时间的函数)输入进行操作。从这个角度来看，SNN 从设计上更接近真正的生物神经网络：SNN 不像 ANN 那样处理不断变化的时间值，而是处理在定义时间发生的离散事件。SNN 将一组尖峰作为输入并产生一组尖峰作为输出(一系列尖峰通常称为尖峰序列)。典型的三代神经网络如图 3 所示。

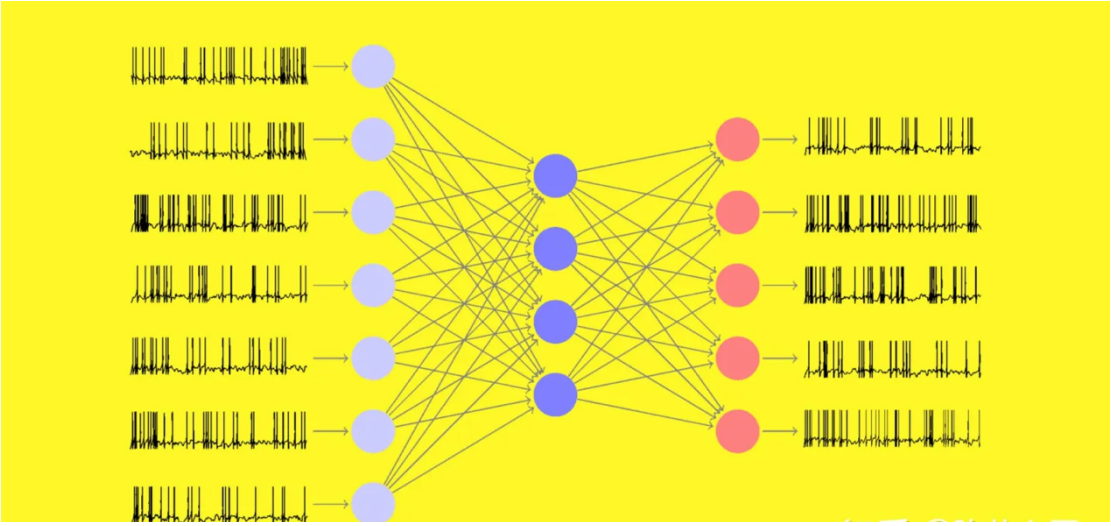


图 3 典型的三代神经网络：脉冲神经网络 SNN

与典型的多层感知机网络不同，动态神经网络中的神经元不是在每一次迭代传播中都被激活，而是在它的膜电位达到某一个特定值才被激活。而当一个神经元被激活，它会产生一个信号传递给其他神经元，提高或降低其膜电位。但是从本质来讲，这些神经网络都是基于神经脉冲的频率进行编码(rate coded)。

一般地，脉冲神经网络的这些特点使其很好地包含了时变信息，从而容易观察其在时间上的动力学特征(激活水平对于时间的偏导数)。

2.1.3.2 脉冲神经网络的工作原理

脉冲神经网络与自然神经网络非常相似。就其工作原理而言，除了神经元和突触状态之外，SNN 还将时间纳入其工作模型。这个想法是，SNN 中的神经元不会在每个传播周期结束时传输信息(就像它们在传统的多层感知器网络中所做的那样)，而只会在膜电位(神经元与其膜电荷相关的内在质量)时传输信息。达到一定的值，称为阈值[16]。

当膜电位达到阈值时，神经元会发射信号，向邻近的神经元发送信号，这些神经元会根据信号增加或减少它们的电位。尖峰神经元模型是在阈值交叉时刻激发的神经元模型。脉冲神经网络中的电位激活动力学例如图 4。

具体地，脉冲神经网络的总体工作原理可以被概括为如下几个方面：

1) 脉冲神经网络中的每个神经元都被赋予一个值，这个值用于表征任何给定时间t上的神经元的生物电势。
2) 神经元的生物电势值可根据其数学模型(被先验地赋予，用微分方程来表示)而变化。例如，如果一个神经元从上游神经元得到一个脉冲，它的值可能会上升或下降。
3) 如果神经元的值超过某个阈值，神经元将向连接到第一个神经元的每个下游神经元发送单个脉冲，神经元的值将立即降至其平均值以下。
4) 结果，神经元将经历类似于生物神经元的不应期。随着时间的推移，神经元的值将逐渐恢复到其平均值。

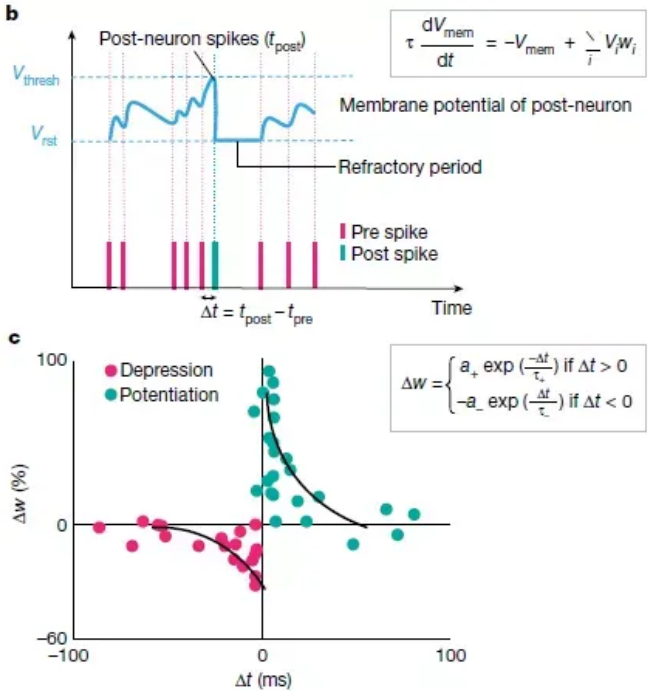


图 4 脉冲神经网络中的脉冲动力学函数

2.1.3.3 脉冲神经网络的优势

综上，脉冲神经网络的优势是显而易见的，可以归纳为以下诸方面：

节省计算资源：时间编码表明，单一的神经元即可取代上百个 Sigmoid 型隐藏层节点。相比于传统的神经网络，SNN 可以使用比较少的资源进行计算。SNN 通常需要更少的资源进行运算，其原因主要可以被归结为两大方面：一方面，因为它们不需要额外的计算或存储开销来维护激活函数，因此可以节省大量的计算资源。另一方面，因为 SNN 中的神经元通过脉冲来表示信息，而不是通过浮点数值，从而减少了存储和计算开销。

可解释性和可靠性：由于 SNN 通过脉冲表示信息，因此可以通过观察脉冲活动来理解神经网络的决策过程。这可以帮助诊断错误并对模型进行优化。同时，SNN 模型的可靠性高，因为它们不依赖于激活函数，并且可以通过硬件加速器实现。

支持时序数据：SNN 具有动态特征表示能力，即通过记录神经元的激活时间来捕捉动态特征，这对于处理时序数据和动态信号非常有效。SNN 可以很好地处理时序数据，SNN 可以在记录每个神经元的激活时间的同时，捕捉时间相关性，并通过记录脉冲活动来捕捉动态特征。这使得 SNN 适用于处理具有时间相关性的数据，例如语音信号和视频序列。因此，相比传统 ANN，在处理时间序列等问题上，具备与生俱来的强大优势。

高效节能：SNN 的算法简单，易于实现，可以通过硬件加速器进一步优化。首先，SNN 中的神经元单元仅当接收或发出尖峰信号时才处于兴奋状态，故其是事件驱动的，因而可大大节省能量功耗。而且，SNN 在硬件上的实现效率通常比传统的神经网络高。例如，在 FPGA(可编程逻辑器件)和 ASIC(专用集成电路)设备上实现 SNN 可以显著提高计算效率。例言之，SNN 在处理计算机视觉问题时，通常将图片转化为脉冲信号。由于每幅图像往往只有若干个峰值作为特征，

这也使得它更适合于神经形态硬件的实现。例如，SNN 可以方便地通过硬件加速器(如 FPGA 和 ASIC)来实现高效的计算，因为它们的算法可以与硬件的结构很好地匹配。因此，由 SNN 构成深度网络，有利于实现神经网络的高效节能化。

更复杂和真实的仿生学特征：这种新型的神经网络更加接近真实的神经网络构造，它更直接地以脑科学和认知神经科学的研究成果为基础。脉冲神经网络的模拟脉冲直接模仿生理学的电脉冲，使得其可以被用于模拟真实生物神经系统的工作。由于采用脉冲编码(spike coding)，脉冲神经网络可以通过获得脉冲发生的精确时间，这种新型的神经网络可以进行获得更多的信息和更强的计算能力。目前，已经出现了很多编码方式把这些输出脉冲序列解释为一个实际的数字，这些编码方式会同时考虑到脉冲频率和脉冲间隔时间。借助于神经科学的研究，人们可以精确的建立基于脉冲产生时间神经网络模型。

这些都是 SNN 的显著优势，并且几乎都直击第二代神经网络的瓶颈与痛点。也正是因此，作为第三代神经网络的脉冲神经网络被寄予了很大的期望。但是，我们也需要考虑到 SNN 的局限性，例如训练困难等。因此，在使用 SNN 时，研究人员提出了 ANN to SNN 的转编码方案来解决这一问题。历代神经网络发展严格的总结和对比如图 5。

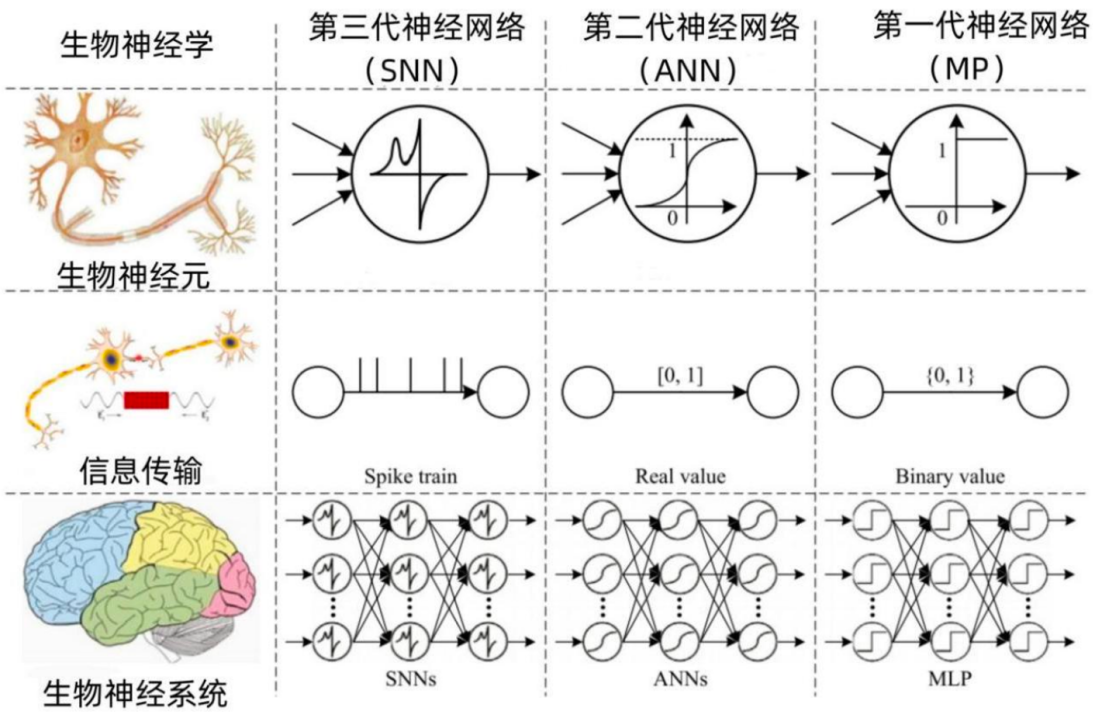


图 5 历代神经网络及其对比

2.1.3.3 脉冲神经网络的应用

脉冲神经网络的应用大体可以被归纳为以下几个方面：

首先，脉冲神经网络基本可以实现传统的人工神经网络的各种信息处理功能。例言之，这写功能包括而不限于自然语言处理、图像识别、机器人控制、智能家居、自动驾驶等基本领域。在计算机视觉领域，SNN 可以用于图像分类、目标检测、图像识别等任务。SNN 可以高效地编码图像特征，并基于这些稀疏特征进行高效分类。SNN 还可以用于识别物体的位置、大小和形状等信息。在语音

识别领域，SNN 可以用于语音识别、语音合成等任务。SNN 可以处理语音信号的频谱特征，并通过语音识别技术将语音转化为文本。在机器人控制领域，SNN 有望用于机器人的状态感知、决策和控制等。SNN 可以识别机器人周围环境的特征，并基于这些特征决策如何控制机器人。在在神经计算领域，SNN 可以用于各种神经计算任务，例如计算机视觉中的图像识别。SNN 可以通过模拟人脑神经元的活动来处理信息，从而获得更高的计算效率。

其次，SNN 也被证明在神经系统科学的研究中有重要作用。SNN 可以帮助我们更好地理解人脑的工作原理，并为解决神经疾病提供新的思路和方法。神经科学研究表明，人脑中的神经元通过发送脉冲来传递信息。SNN 模拟了这种脉冲信息传递的机制，故有望用于辅助关于人脑的工作原理的研究。此外，SNN 可以用于模拟大脑的信息处理过程，例如视觉、听觉和认知等，因而有望用于研究神经疾病，例如癫痫、抑郁症和认知障碍等机制的研究。

2.2 骨骼检测介绍

2.2.1 任务目标与应用场景

骨骼检测，又被称作人体骨骼关键点检测(Pose Estimation)，主要检测人体的一些关键点，如关节，五官等，通过关键点描述人体骨骼信息，是一项重要的图像分析技术[17]。它被广泛应用于生物医学、运动学和计算机视觉等领域。

骨骼检测的基本目标是通过对人体动态影像的分析，识别出人体的骨骼结构，并对骨骼的运动进行跟踪。主要用于生物医学诊断、运动学研究和人机交互等方面。骨骼检测需要解决的主要技术问题包括人体骨骼识别、骨骼跟踪和姿态估计等。

2.2.2 骨骼检测的研究现状

目前，骨骼检测的主要研究方向有以下几个方面[18]：

1) 单人体骨骼检测：通过利用单个图像或视频序列中的人体姿势信息，来实现单人骨骼的检测[19]。
2) 多人体骨骼检测：实现对多个人体在同一图像或视频序列中的骨骼信息的检测[20]。
3) 实时骨骼检测：通过实时处理视频帧以获得人体骨骼信息，以满足实时应用的需求。
4) 骨骼检测的实际应用：骨骼检测技术在生物医学诊断、运动分析、人机交互等领域的应用[21]。

图 6 骨骼检测示意

目前，骨骼检测的研究仍在不断深入和拓展，希望能够在准确性和效率方面取得更大的提高，并在更多的实际场景中得到广泛应用[22]。一般地，骨骼检测的效果如图 6 所示。

2.2.2 骨骼检测的方法

骨骼检测的方法主要有分基于模板匹配和基于深度学习两类方法：基于模板匹配的方法通常采用图像处理技术，如阈值分割、轮廓提取等，以实现人体骨骼信息的检测[23]。然而，这类方法通常存在误识别率较高、对不同姿势的适应性较差等问题。基于深度学习的骨骼检测方法则采用了深度学习技术，如卷积神经网络注意力网络等，以实现对人体骨骼信息的准确检测。相比之下，基于深度学习的方法有更高的识别率、对姿势变化敏感性更低[24]。

尽管骨骼检测技术取得了显著进展，但仍存在一些挑战。例如，骨骼检测的准确性仍然需要提高，特别是对于复杂姿势和遮挡情况下的检测效果。另外，随着人体骨骼检测技术的普及，如何保护用户隐私也成为了一个重要问题[25]。

2.2.3 骨骼检测中的传统算法

传统的人体骨骼检测主要是采用基于模版的匹配算法[26]。简而言之，即先从经验中总结出人体骨架结构的几何先验，即所谓的模板。然后则用算法来计算待检验目标与人体骨骼模板之间的相似度。如果相似度高于某种阈值，则认为检测出了人体骨架。显然，传统算法的关键在于模版表示问题，即关键点、肢体、及其连接方式等如何表示的问题[27]。另外，人体的姿态是十分复杂多变的，模版既要做到能够尽可能地匹配更多人体姿态，又要避免将非人体的部分误判为人体，这成为骨骼检测算法中的主要矛盾。

最具代表性的模板匹配算法是 Pictorial Structure[28]。它主要包含两个部分，即单元模版和模版关系。其中，单元模板意味着把人体各部分表示为诸个基本部件。而模版关系方面的创新则至为关键，即弹簧形变模型。该模型用弹簧作为维系并约束各个相邻部件之间相对关系的机构。该模型成功解决了前述的矛盾，在约束部件的合理相对关系的同时保持了较好的灵活性，如图 7。

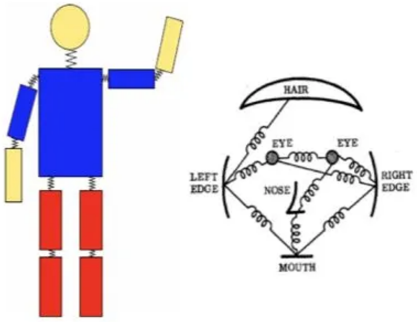


图 7 弹簧形变模型

此后，Ramanan 等提出了“mini parts”的概念来实现更大的姿态匹配范围[29]。该方法将肢体的每个部件切分成更小的部分，以模拟更多的姿态变化，从而提高模版匹配的效果。其概念要点如图 8 所示。

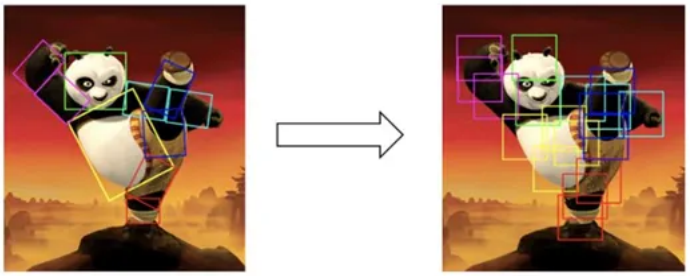


图 8 Mini Parts 方法

在计算机视觉领域中，传统的算法被用于骨骼检测的例子还有很多。其中一些较有代表性的传统算法还包括：

关节姿态估计(APE)：APE 是一种基于模型的方法，它使用一组预定义的模型来估计图像中人物的姿态。模型代表了人物的各种可能姿态，算法在观察到的图像和预定义模型之间找到最佳匹配[30]。

基于部件的模型(PBM)：PBM 是一种可变形部件模型，它将身体划分为多个部分，并分别对每个部分进行建模。然后将部件组合在一起形成图像中人物的最终姿态[31]。

分层姿态估计(HPE)：HPE 是一种多阶段方法，它使用人体的分层表示来估计图像中人物的姿态。算法从大致的姿态估计开始，在每个阶段都细化估计，直到得到最终姿态[32]。

基于关节和骨骼的模型(JBM)：JBM 是一种基于模型的方法，它将人体表示为一系列关节和骨骼。算法估计关节和骨骼的位置，然后计算图像中人物的最终姿态[33]。

2.3 骨骼检测中的神经网络

2.3.1 基于深度学习的骨骼检测

随着技术的发展，骨骼检测的研究也不断发展，许多新的算法也被引入。最近，基于深度学习的骨骼检测方法已经引起了很多关注，因为它们可以在不依赖预定义模型的情况下生成高质量的结果[35]。

近年来，随着深度学习技术的不断发展，基于深度学习的骨骼检测方法已经成为了计算机视觉领域中一个研究热点。骨骼检测是指通过图像处理技术来识别人体骨骼关键点的过程，它在很多领域中具有重要的应用价值，例如运动分析、人机交互、动作识别等[36]。

基于深度学习的骨骼检测方法主要分为两类：基于单帧图像的方法和基于多帧图像的方法。基于单帧图像的方法通常采用卷积神经网络(Convolutional Neural Network，CNN)和递归神经网络(Recurrent Neural Network，RNN)等深度学习模型来识别人体骨骼关键点[37][38]。例如，Luvizon 等人提出了一种基于单帧图像

的骨骼检测方法，该方法使用了卷积神经网络和递归神经网络来识别人体骨骼关键点，并且在多个数据集上得到了良好的实验结果[39]。

基于多帧图像的方法则通常采用长短时记忆网络(Long-Short Term Memory，LSTM)和光流法等技术来识别人体骨骼关键点。例如，Lee 等人提出了一种基于多帧图像的骨骼检测方法，该方法使用了长短时记忆网络和光流法来识别人体骨骼关键点，并且在多个数据集上取得了优秀的实验结果[40]。

除了上述两类基于深度学习的骨骼检测方法外，还有一些结合了多种技术的方法。例如，Chen 等人提出了一种结合了卷积神经网络、长短时记忆网络和光流法的骨骼检测方法，该方法在多个数据集上均取得了很高的准确率。

总的来说，基于深度学习的骨骼检测方法具有较高的准确性和实时性，并且在多个数据集上均取得了良好的实验结果。然而，目前的骨骼检测方法仍存在一些不足，例如对于骨骼外形变化的敏感性较弱、对于不同姿态的识别能力不一致等问题。因此，未来的研究任务应该着重于解决这些问题，以进一步提高骨骼检测方法的准确性和实用性。

在骨骼检测中，进一步又可分为单人人体骨骼检测和多人人体骨骼检测主要有两个方向：即自上而下和自下而上两种。

2.3.2 自上而下的人体关键点检测算法

自上而下的人体关键点检测算法是一种基于先验模型的方法，即先预测整个人体的姿态，然后再逐步预测每个关键点的坐标位置。该方法的主要思路是将整个人体作为一个整体进行处理，将人体的姿态信息作为先验模型来指导每个关键点的检测[41]。

自上而下的人体关键点检测算法主要包括以下几个步骤：

1) 预测人体的姿态：使用一些预训练的模型，如基于深度学习的卷积神经网络(CNN)，预测整个人体的姿态，包括关键点的位置、角度、旋转矩阵等信息。该步骤可以使用现有的人体姿态估计算法进行预测。
2) 候选框生成：根据预测的人体姿态，对每个关键点的可能位置生成候选框，即在每个可能位置周围生成一个固定大小的矩形区域，作为关键点检测的候选区域。
3) 关键点检测：在每个候选框中使用 CNN 或其他检测算法来检测关键点。这些算法通常是由多个卷积层、池化层、全连接层等构成，可以从输入图像中提取特征，并预测每个关键点的坐标位置。
4) 关键点筛选：通过一些筛选方法，如非极大值抑制(NMS)、置信度评估等，对检测到的关键点进行筛选和修正，去除误检和漏检的关键点，从而得到最终的关键点检测结果。

自上而下的人体关键点检测算法相较于自下而上的算法，具有更高的精度和稳定性，但是计算量也更大。该方法主要应用于人体姿态估计、人体行为分析、动作识别等领域，并在一些比赛中取得了较好的成绩。

对于自上而下的人体关键点检测算法中的目标检测部分，将在以下的第四章进行展开讨论和评价，这里主要讨论关键点检测算法部分。这部分算法的主要挑战来自三大方面：

1) 关键点局部信息的区分性很弱，即背景中很容易会出现同样的局部区域造成混淆，所以需要考虑较大的感受野区域；

2) 人体不同关键点的检测的难易程度是不一样的，对于腰部、腿部这类关键点的检测要明显难于头部附近关键点的检测，所以不同的关键点可能需要区别对待；
3) 自上而下的人体关键点定位依赖于检测算法的提出的 Proposals，会出现检测不准和重复检测等现象。

2.3.2 自下而上的人体关键点检测算法

自下而上的人体关键点检测算法是一种基于图像中的低级别特征进行多阶段聚类的方法，从而实现关键点的检测。该算法的基本思想是，从图像中检测出局部的特征点，并通过分析这些特征点之间的关系，进而识别出全局的关键点。

自下而上的人体关键点检测算法通常分为三个主要阶段。第一阶段是特征提取，通过使用滤波器或卷积神经网络等技术从图像中提取特征点。第二阶段是局部极值检测，这一阶段旨在检测出特征点中的局部极值。第三阶段是关键点连接，通过将不同的局部特征点进行连接，并对它们进行聚类，识别出全局的关键点。

自下而上的人体关键点检测算法相对于自上而下的算法，其优势在于不需要预先定义人体部位的位置和形状，而是通过对图像中的特征点进行聚类，识别出人体部位的位置和形状。同时，该算法还可以自适应地处理不同姿势、尺度和形变的人体图像，从而提高了算法的鲁棒性。

然而，自下而上的算法也存在一些缺点。首先，该算法需要在处理大量特征点时进行大量计算，导致计算时间较长。其次，算法对噪声和遮挡等干扰因素比较敏感，容易产生误检测和漏检测。

近年来，一些研究者通过将自上而下和自下而上两种方法进行结合，提出了一些混合型的关键点检测算法，通过充分利用两种算法的优势，提高了检测的准确性和鲁棒性。

2.3.2 相关数据集

深度学习算法的准确性很大程度上不仅取决于预定义的模型的正确性和匹配的效果，而且取决于所使用的数据集的质量。有几个公开可用的骨架检测数据集，通常用于训练和评估该领域的算法。一些使用最广泛的数据集包括：

MPII 人体姿势数据集[42]：该数据集包含超过 25，000 张不同姿势的人的图像，并用代表人体关节的 14 个关键点进行注释。

MS COCO 关键点数据集[43]：该数据集包含超过 330，000 张图像，注释有 80 个代表人体和其他物体的关键点。

LSP 数据集[44]：此数据集包含各种姿势的人的图像，并用代表人体关节的 14 个关键点进行注释。

FLIC 数据集[45]：此数据集包含各种姿势的人的图像，并带有代表人体关节的 14 个关键点注释。

Human3.6M 数据集[46]：此数据集包含执行日常动作的人的动作捕捉数据，并用代表人体关节的 3D 关键点进行注释。

这些数据集为骨骼检测领域的研究人员和从业者提供了宝贵的资源，并已在许多研究中用于训练和评估人体姿势估计的算法。这些数据集的可用性也推动了该领域的进步，因为研究人员正在努力开发更准确，更有效的骨骼检测算法。

3. 讨论

3.1 第二代神经网络的局限性及其讨论

由前所述，不难总结出，第二代神经网络虽然拥有许多优势，但也有很大的局限性。而其局限性主要有以下几点：

1) 需要大量的手动特征工程：在第二代神经网络中，每一层神经元都需要对输入的图像进行特征提取。这需要大量的手动特征工程，如选择适当的滤波器、阈值等。因此，对于不同的数据集，可能需要不同的特征工程。
2) 计算效率较低：第二代神经网络的计算方法是基于前馈网络的，它需要对图像进行多次卷积和池化操作，以计算出图像的特征。这需要大量的计算，导致计算效率较低。
3) 难以处理高维数据：第二代神经网络的网络架构是基于低维数据的，难以处理高维数据。因此，对于高维数据，需要进行降维处理，以使其能够被第二代神经网络处理。
4) 模型的不可解释性：第二代神经网络的模型是不可解释的，即很难确定生成结果的原因。这对于研究人员来说是一个问题，因为他们需要确定模型的工作原理。

通过以上四点，我们对第二代神经网络的局限性做进一步讨论和引申：

首先，由于第二代神经网络通常是以每层为单位的，没有明确的网络结构，这使得其难以捕捉复杂的模式。

其次，第二代神经网络极为依赖数据预处理，它需要进行大量的数据预处理，以确保输入数据的质量和有效性。

第三，第二代神经网络缺乏通用性，它的性能取决于训练数据的质量和特征的选择，因此它们缺乏通用性。

再者，第二代神经网络的训练时间通常会比较长，这限制了它们的实际应用。

最后，第二代神经网络对特征的依赖性高，这导致对特征的选择非常敏感，它们的性能取决于特征的选择。

3.2 关于第三代神经网络与第二代神经网络之比较的个人思考

通过对上述诸问题的分析，我们可以看到第三代与第二代神经网络之间多个方面的差异。在此，我以个人的有限理解来对这些比较作粗浅讨论：

近年来，以第二代神经网络为核心的深度学习几乎全面地革新了机器学习领域，尤其是在计算机视觉领域。在该方法中，使用反向传播以监督方式训练深度(多层)人工神经网络(ANN)。需要大量的标记训练示例，但由此产生的分类精度确实令人印象深刻，有时甚至超过人类。神经网络中的神经元以单个、静态、连续值激活为特征。

然而，生物神经元使用离散的尖峰来计算和传输信息，除了尖峰频率之外，尖峰时间也很重要。因此，Spiking neural networks(SNN)在生物学上比 ANN 更为现实，如果人们想了解大脑是如何计算的，这可能是唯一可行的选择。SNN 也比 ANN 更为硬件友好和节能，因此对技术，尤其是便携式设备具有吸引力。

这意味着，对于相同的任务，而 SNN 通常需要更少的操作然而，训练深度 SNN 仍然是一个挑战。Spiking 神经元的传递函数通常是不可微的，这阻止了使用反向传播。

在这里，最近用于训练深度 SNN 的有监督和无监督方法，并在准确性、计算成本和硬件友好性方面对它们进行了比较。新出现的情况是，SNN 在准确性方面仍然落后于 ANN，但差距正在缩小，甚至可能在某些情况下消失。

3.3 试论骨骼检测任务中第三代神经网络的前景

目前，直接基于第三代神经网络的骨骼检测应用研究还相对较少。大多数研究还是基于深度学习的骨骼检测方法。深度学习对于推动人体运动分析、人机交互和动作识别等领域的发展具有重要意义。基于深度学习，骨骼检测技术可以提供对人体姿态、动作和运动的准确识别，为进一步研究人体运动学和生物力学提供基础数据。此外，骨骼检测技术还可以应用于许多其他领域，例如医学影像分析、安全监控和体育等。

然而，未来的研究应该着重于开发更准确、更实时的骨骼检测技术，以满足不同领域的需求。此外，还需要对基于深度学习的骨骼检测方法进行进一步的改进，以提高其在实际应用中的效率和稳定性。

而第三代神经网络恰恰具备捕捉时变特征的本质特性，这为进一步的相关研究工作提供了极其广阔的想象空间。因此，总的来说，基于第三代神经网络的骨骼检测方法是一个具有广阔前景和巨大潜力的研究方向，值得我们进一步探究和开发。

以下对脉冲神经网络在骨骼检测方面的应用做一展望，可能的应用方向主要在以下几个方面：

1) 骨骼关键点检测：SNN 可以用于检测骨骼关键点，其原理是将输入的图像转换为脉冲信号，通过神经元之间的突触连接进行信息传递，最终输出预测的骨骼关键点坐标。相较于传统的基于深度学习的方法，SNN 可以减少计算量，实现更加高效的骨骼关键点检测。
2) 姿态估计：SNN 也可以应用于人体姿态估计，其原理是在骨骼关键点检测的基础上，通过对骨骼关键点的分析和处理，得到人体的姿态信息。在实际应用中，SNN 的姿态估计能够实现更加准确和稳定的结果。
3) 动作识别：SNN 还可以应用于人体动作识别，其原理是在骨骼关键点检测和姿态估计的基础上，通过对人体动作序列的分析和处理，识别出不同的动作。相较于传统的基于深度学习的方法，SNN 可以实现更加高效的动作识别。

4. 神经网络在自己研究领域(目标检测)中的应用

4.1 目标检测的定义、方法和应用

4.1.1 目标检测的定义和目标

目标检测是一种图像分析技术，它将计算机视觉技术应用于图像处理，以便识别和定位特定的对象[47]。目标检测的目的是检测出图像中的所有物体，并在

图像中为它们定位。它是一种基于深度学习的计算机视觉技术，通过分析图像来检测对象的位置。

目标检测的主要目标是识别和定位图像中的目标物体，这些物体可以是任何形状或大小的目标，如人、动物、植物、车辆、机器人等。在识别和定位目标时，目标检测算法可以使用卷积神经网络(CNN)，支持向量机(SVM)，聚类，图像识别和其他机器学习算法[47]。

4.1.2 目标检测的应用场景

目标检测的最终目的是帮助机器人和无人机识别和定位目标，以便能够发挥其最大的效能。它还可用于以下用途[47]：

1) 安全监控：可以用来检测和识别人群中的可疑活动，以帮助识别和抓住犯罪分子。
2) 自动驾驶：可以用来检测道路上的障碍物，以帮助自动驾驶车辆避免碰撞。
3) 农业：可以用来检测农田中的植物和动物，以帮助农民更好地管理农业。
4) 医疗：可以用来检测医院的病人，以及 CT 和 X 射线扫描中的病灶，以帮助医生更准确地诊断病情。
5) 图像搜索：可以用来检测图像中的特定对象，以帮助用户更快地搜索图像。
6) 交通管理：可以用来检测道路上的车辆、行人、动物等，以帮助更有效地管理交通流量。

因此，目标检测是一项重要的机器视觉技术，它可以被进一步嵌入于各种应用程序中，以提高机器性能和准确性。它能够检测出图像中的所有物体，从而有助于进行更准确的机器行为模拟和控制。

4.1.3 目标检测的传统算法

4.1.3.1 Viola Jones Detector

Viola-Jones 算法的关键是 Viola-Jones 检测器，它基于滑动窗口的方式逐一检查目标是否存在窗口之中。因此，这些特征是简单的矩形特征，可用于捕获图像中的各种图案[48]。

Viola-Jones 算子通过估计一组类似 Haar 的特征及其相应的阈值，快速排除掉大量不含特征的窗口。如果窗口在某个阶段通过了所有测试，则会将其传递到下一阶段进行进一步评估。

该检测器算法设计非常简单，但受限于过高的时间复杂度，难以应用于实时场景。改进的思路主要有三种：

1) 特征的快速计算方法-积分图；
2) 有效的分类器学习方法-AdaBoost；
3) 高效的分类策略-级联结构的设计。

Viola-Jones 检测器广泛用于计算机视觉应用中的人脸检测，例如安全系统，视频会议和人机交互。它高效快速，可以在标准个人计算机上实时运行。但是，它有一些限制，例如对比例、方向和照明条件的敏感性，这使得它在某些情况下效果较差。

4.1.3.2 HOG Detector

HOG(Histogram of Oriented Gradients，梯度方向直方图)检测器是一种基于梯度方向直方图概念的计算机视觉目标检测算法[49]。

在 HOG 检测器中，图像被划分为小单元格，并计算每个单元格内像素的梯度方向，然后将其直方图化。这导致了一个梯度方向的直方图，描述了每个单元格中的主导边缘方向。然后将这些直方图连接在一起，形成一个代表整个图像或感兴趣区域的特征向量。

然后将特征向量传递给分类器，如支持向量机(SVM)，以做出图像或区域是否包含感兴趣对象的最终决策。HOG 检测器已被广泛用于计算机视觉应用中的行人检测，例如自动驾驶汽车、视频监控和行人安全系统。

HOG 检测器的优势之一是它的不变性，比如对平移、尺度、光照等变化的相对不敏感，因此可以处理尺度和方向的变化。HOG 通过在均匀间隔单元的密集网格上计算重叠的局部对比度归一化来提高检测准确性，因此 HOG 检测器是基于本地像素块进行特征直方图提取的一种算法，它在目标局部变形和受光照影响下都有很好的稳定性。但另一方面，HOG 对超参数的选择(如单元格的大小和直方图中的箱数)可能敏感。

HOG 为此后的很多检测方法奠定了重要基础，相关技术也一度被广泛应用于计算机视觉各大应用。

4.1.3.3 DPM Detector

DPM(可变形部件模型)检测器是一种计算机视觉目标检测算法，使用滑动窗口方法与概率模型相结合，以检测感兴趣的对象[50]。

在 DPM 检测器中，使用可变形部件的模型来表示感兴趣的对象，其中每个部件都由矩形区域和一组可变形模板表示。该模型是在感兴趣的对象的大量正例上训练的，可变形模板学习成为对正例最大响应的同时对负例最小响应。

DPM 检测器通过将窗口滑动到图像上，并计算在给定可变形部件的模型的情况下窗口内对象存在的可能性。如果可能性超过阈值，窗口将被分类为正例。DPM 检测器可以处理比例和宽高比的大变化，并且可以同时检测图像中的多个对象。

DPM 检测器的优点之一是它可以处理对象的变形，例如位置、方向和形状的变化。然而，它在计算上是昂贵的，对超参数的选择(如滑动窗口的大小和位置，以及确定正例的阈值)很敏感。因此，使用 DPM 检测器时需要认真考虑这些因素，以确保获得准确和可靠的结果。

4.1.3.4 局限性

目标检测的传统算法，如 HOG、Viola-Jones 等，存在一些明显的局限性，主要是以下几点：

1) 计算代价高：传统算法通常需要大量的计算资源，特别是在处理高分辨率图像时，导致实时性能差。
2) 对尺寸和形状变化的敏感性：在目标大小和形状有很大变化的情况下，传统算法的准确性下降。
3) 需要先验知识：传统算法通常需要关于目标的先验知识，如其大小和形状，这限制了其适用性。
4) 对背景影响大：如果目标与背景相似，或者目标周围有明显的噪声，传统算法的准确性会受到影响。

尽管这些局限性，传统算法仍然是目标检测领域的重要贡献，并且仍然可以在特定的场景中获得良好的效果。

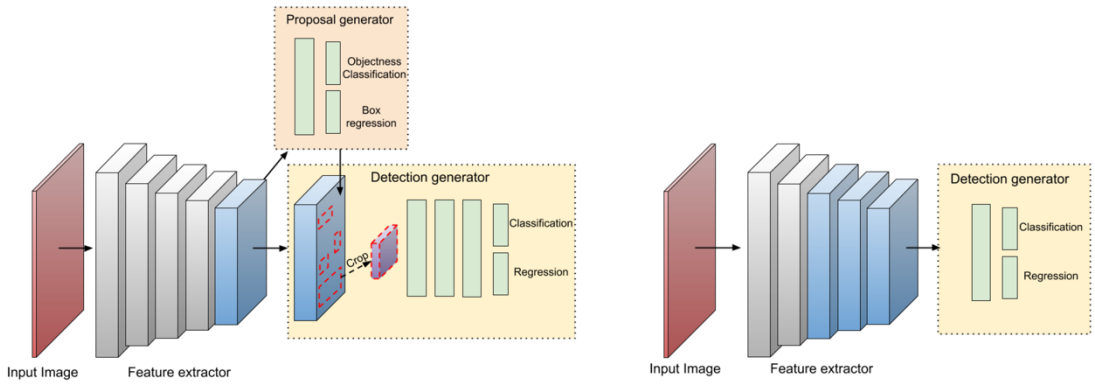
4.2 基于深度学习的目标检测算法

基于手工提取特征的传统目标检测算法进展缓慢，性能低下。随着卷积神经网络的提出，目标检测的相关研究打开了新的格局。这些算法在大量的图像数据上进行训练，以学习目标的语义特征和位置特征。基于 CNNs 的目标检测算法主要有两大方法：anchor-based 和 anchor-free[51]。

Anchor-based 方法是目前主流的目标检测算法，它们使用预定义的一组目标框(称为锚框或先验框)，并通过卷积神经网络来预测这些框的偏移量和类别概率。在这个方法中，一阶段检测算法(如 YOLO 和 SSD)直接从输入图像中预测目标框，而二阶段检测算法(如 Faster R-CNN 和 Mask R-CNN)则使用区域提议网络(RPN)来生成一组候选框，然后对这些框进行分类和回归，最终输出检测结果[51]。

相比之下，Anchor-free 方法没有预定义的锚框，而是通过网络自适应地学习目标的位置和形状。这些方法通常使用类似于关键点检测的技术来检测目标，比如 CornerNet、CenterNet 和 FCOS。这些方法在一些情况下表现出了更好的性能，特别是在小目标检测和密集目标检测方面[52]。

总的来说，anchor-based 方法是目前主流的目标检测算法[53]，具有高准确性和较低的计算成本。相比之下，anchor-free 方法则在某些情况下表现出了更好的性能，但需要更高的计算成本。而 anchor-based 方法则可被进一步区分为一阶段和二阶段检测算法两种模式。一般而言，其中二阶段目标检测算法比一阶段精度高，但比一阶段检测算法速度慢。二者的网络架构和原理如图 9 所示。



(a) Two-stage Faster R-CNN　　(b) One-stage RetinaNet

图 9 一阶段算法和两阶段算法对比

基于深度学习的目标检测算法的优势在于它们能够以端到端的方式学习目标的复杂特征，而且不需要人工设计特征，因此具有更高的准确性。此外，它们还能适应多种不同的目标形状和大小，在图像中识别复杂背景和遮挡等情况。

当然，基于深度学习也存在一些局限性。首先，它们需要大量的训练数据，否则容易出现过拟合的情况。此外，它们还需要高性能的计算机资源来进行训练和推理，因此可能不适用于低端设备或实时应用。此外，由于模型复杂度较高，训练时间也很长，因此可能不利于快速实现和部署[53]。

4.2.1 两阶段算法

就 anchor-based 的方法而言，基于深度学习的目标检测算法又可被进一步归为两类：一阶段目标检测和两阶段目标检测。一阶段目标检测算法将目标检测任务作为一个单一的端到端学习任务，直接在整个图像上进行预测。两阶段目标检测算法则先通过一个预处理阶段提取候选目标，再在这些候选目标上进行详细的目标分类和定位。

两阶段算法是一种基于深度学习的目标检测算法，它通常包括两个阶段：候选区域生成和目标检测。在第一个阶段中，算法使用区域生成网络(Region Proposal Network，RPN)或其他类似的方法来生成潜在的目标区域。在第二个阶段中，对这些区域进行分类和回归，以确定是否存在目标以及目标的位置和大小。

两阶段算法的典型代表是 Faster R-CNN 算法[54]。在 Faster R-CNN 中，RPN 网络首先在输入图像上生成一组候选区域，然后这些区域经过 ROI Pooling 层提取特征，并输入到全连接层进行分类和回归。整个过程可以通过端到端的方式进行训练，从而使得算法能够自动学习目标的特征和位置信息[54]。

相比于一阶段算法，两阶段算法的主要优势在于精度更高，能够检测出更多的目标和更准确的位置信息。这是因为两阶段算法能够生成更多的候选区域，从而能够更全面地探测图像中的目标。此外，两阶段算法能够对候选区域进行更细致的分类和回归，从而能够更准确地确定目标的位置和大小。

然而，两阶段算法也存在一些缺点。首先，它们通常比一阶段算法更复杂，需要更多的计算资源和更长的训练时间。此外，由于需要生成大量的候选区域，它们的速度相对较慢，不适用于一些实时应用场景。

4.2.2 一阶段算法

一阶段算法是一种基于深度学习的目标检测算法，与两阶段算法相比，它只有一个阶段，直接从输入图像中生成目标框和类别标签。典型的一阶段算法包括 YOLO[55]和 SSD[56]。

在 YOLO 中，输入图像被分成网格，每个网格预测固定数量的目标框和对应的类别概率。对于每个目标框，算法同时预测其边界框偏移量和置信度，其中置信度表示该目标框包含物体的概率，边界框偏移量用于修正生成的目标框。最终，算法通过 NMS(非极大值抑制)来过滤出置信度较低的目标框，得到最终的检测结果。

在 SSD 中，算法通过一系列卷积层来提取特征，并将不同层的特征进行组合，生成一组固定数量的目标框和类别概率。与 YOLO 不同的是，SSD 采用了多尺度特征融合的策略，使得算法能够对不同大小的物体进行检测。此外，SSD 还引入了多个比例和宽高比的先验框，使得算法能够更好地适应不同形状和大小的目标。

相比于两阶段算法，一阶段算法具有更快的检测速度和更低的计算复杂度，因此适用于一些实时应用场景。此外，一阶段算法还可以直接从输入图像中生成目标框和类别标签，因此可以避免候选区域生成的过程，从而能够更准确地检测小目标。

然而，一阶段算法也存在一些缺点。首先，它们通常比两阶段算法的精度稍低。此外，由于直接从输入图像中生成目标框和类别标签，一阶段算法容易出现重复检测的问题，导致结果不够准确。

总之，一阶段算法是一种快速和高效的目标检测方法，适用于一些实时应用场景。随着计算资源和算法的不断优化，一阶段算法的应用范围将会越来越广泛。

5. 脉冲神经网络在的潜在目标检测中的应用与自己的思考

5.1 基于脉冲神经网络的目标检测算法

目标检测是计算机视觉中的重要问题，它旨在识别图像中存在的对象，并定位它们的位置。传统的目标检测算法主要基于卷积神经网络。但这些算法存在一些问题，例如计算复杂度高、能效性差等。因此，研究人员开始探索基于脉冲神经网络的目标检测算法，这些算法通过使用离散化的脉冲信号来代替连续的激活值，从而降低了算法的计算复杂度和能耗。

基于脉冲神经网络的目标检测算法主要包括以下几个步骤：

1) 数据准备：和传统算法一样，数据准备也是基于脉冲神经网络的目标检测算法的第一步。数据准备包括数据集的构建、图像的预处理等。
2) 神经网络模型构建：脉冲神经网络模型包括脉冲卷积层、脉冲池化层、脉冲积分激活函数等组件。这些组件的设计旨在实现离散化的脉冲信号的传递和处理。
3) 训练：脉冲神经网络的训练过程包括两个阶段。首先，需要将传统的连续值的卷积核和权重转化为脉冲形式。然后，利用训练数据集进行网络的优化和训练，获得脉冲卷积核和权重，使得网络能够更好地实现目标检测任务。
4) 目标检测：在目标检测阶段，输入的图像首先被转化为脉冲信号，然后被送入脉冲神经网络中进行处理。最后，输出的脉冲信号被转化为连续值，以获得目标检测结果。

目前，基于脉冲神经网络的目标检测算法主要有 Spiking-YOLO[57]、Spiking-RCNN[58]和 Spiking-SSD[59]。

目前，基于脉冲神经网络的目标检测算法主要有以下几种：

1) Spiking-YOLO：这是一种基于 SNN 的目标检测算法，它通过将 YOLO 网络中的卷积层替换为脉冲卷积层来实现。此外，Spiking-YOLO 还使用了一种新的激活函数，称为“脉冲积分激活函数”，以将脉冲信号转换为实值输出。实验结果表明，Spiking-YOLO 在目标检测方面的性能与传统 YOLO 算法相当。
2) Spiking-RCNN：这是一种基于 SNN 的目标检测算法，它通过将 RCNN 网络中的卷积层和池化层替换为脉冲卷积层和脉冲池化层来实现。此外，Spiking-RCNN 还使用了一种新的损失函数，称为“脉冲率损失函数”，以计算目标检测结果中的脉冲率。实验结果表明，Spiking-RCNN 在目标检测方面的性能优于传统 RCNN 算法。
3) Spiking-SSD：这是一种基于 SNN 的目标检测算法，它通过将 SSD 网络中的卷积层和池化层替换为脉冲卷积层和脉冲池化层来实现。此外，Spiking-SSD 还使用了一种新的激活函数，称为“脉冲积分激活函数”，以将脉冲信号转换为实值输出。实验结果表明，Spiking-SSD 在目标检测方面的性能与传统 SSD 算法相当。

总的来说，基于脉冲神经网络的目标检测算法在实现上与传统算法类似，但通过使用脉冲卷积层、脉冲池化层和脉冲积分激活函数等新的组件，可以实现更好的生物合理性和能效性。然而，虽然目前这些算法使用不同的脉冲神经网络模型和优化方法，但它们都具有良好的生物合理性和能效性，并取得了一定的目标检测性能，但许多时候，这些算法的性能还不如传统算法，需要进一步的研究和改进。

5.1.1 Spiking-YOLO

在此方面，Kim S 等深入分析了深度脉冲神经网络中应用于目标检测任务时会出现传统归一化方法的效率低和缺少 leaky-ReLU 在 SNN 神经元中精确转换等问题，并针对以上问题提出了不平衡阈值的符号神经元(IBT)和细致通道归一化两种方法，建立了第一个基于脉冲神经网络的目标检测模型(Spiking-YOLO)[44]。其实验结果表明，前者准确有效地实现了 leaky-ReLU，能够在多个神经元中获得更高更适当的放电率，但后者效果不是很明显。这导致与标准目标检测网络相比相对较低的准确度。此外，该网络不使用 STDP 算法来限制噪声鲁棒性或从少量标记数据中学习。针对以上问题，Chakraborty 等人提出了一种新的基于全脉冲神经网络的目标检测器，该检 测器使用无监督的基于 STDP 的学习和有监督的反向传播学习的混合[45]。此外，Kim S 等人为了完善之前的工作提出了通过贝叶斯优化在每个隐藏层中找到最佳阈值电压，并引入两阶段阈值电压，从而在生物启发的 SNNs 中实现最先进的目标检测精度[46]。为了实现近乎无损的 ANN-to-SNN 转换，大部分研究工作集中在寻找合适的阈值平衡算法和 等效归一化网络不同层的权值。上述相关工作为深度脉冲神经网络的监督学习算法提供 了理论基础。为了减少 ANN 转换的损失，提高 SNN 在目标检测任务中的检测精度。

5.1.2 Spiking-RCNN

Spiking-RCNN 是一种基于脉冲神经网络的目标检测算法，它是 RCNN 系列算法中的一种，由德国特里尔大学的研究团队于 2018 年提出。Spiking-RCNN 采用了脉冲神经网络的思想，利用离散化的脉冲信号来代替传统的连续值，从而降低了算法的计算复杂度和能耗，同时也具有更好的生物合理性。

Spiking-RCNN 的网络结构和传统的 RCNN 类似，包括了一个卷积神经网络(CNN)和一个区域提议网络(RPN)，其中 CNN 用于提取图像特征，RPN 用于生成候选区域。不同的是，在 Spiking-RCNN 中，CNN 被改造成了脉冲卷积神经网络(SCNN)，RPN 则采用了一种基于脉冲神经网络的区域提议模型(SRPN)。

具体来说，Spiking-RCNN 中的 SCNN 采用了脉冲积分神经元(SNN)来替代传统的整流线性单元(ReLU)，并且使用了时空滤波器和最大化池化层来处理离散化的脉冲信号。而 SRPN 则利用了一种名为时间方向区域池化(TDP)的技术，将连续的区域提议结果转化为离散的脉冲信号，以便与 SCNN 配合使用。

Spiking-RCNN 的训练过程分为两个阶段。首先，需要将传统的连续值的权重和卷积核转化为脉冲形式，这需要通过一种名为 STBP(spiking temporal backpropagation)的优化方法来实现。然后，在模型的第二个阶段，利用训练数据集对脉冲神经网络进行优化和训练，得到脉冲卷积核和权重，使得网络能够更好地实现目标检测任务。

实验结果表明，Spiking-RCNN 在准确率和速度方面都比传统的 RCNN 算法有所提升，而且在能耗方面也具有明显的优势。不过，Spiking-RCNN 仍然面临一些挑战，例如训练效率不高、目标检测精度还需要进一步提高等问题。

5.1.3 Spiking-SSD

Spiking-SSD 是一种基于脉冲神经网络的目标检测算法，是 SSD 系列算法中的一种，由韩国庆熙大学的研究团队于 2019 年提出。Spiking-SSD 采用了脉冲神经网络的思想，利用离散化的脉冲信号来代替传统的连续值，从而降低了算法的计算复杂度和能耗，同时也具有更好的生物合理性。

在 Spiking-SSD 中，网络采用了脉冲卷积神经网络(SCNN)来替代传统的卷积神经网络，同时也使用了脉冲神经网络的区域提议模型(SRPN)来生成候选区域。

具体来说，Spiking-SSD 中的 SCNN 采用了脉冲积分神经元(SNN)来替代传统的整流线性单元(ReLU)，并且使用了时空滤波器和最大化池化层来处理离散化的脉冲信号。而 SRPN 则使用了与 Spiking-RCNN 类似的时间方向区域池化(TDP)技术，将连续的区域提议结果转化为离散的脉冲信号，以便与 SCNN 配合使用。

Spiking-SSD 的训练过程与传统的 SSD 类似，也采用了基于梯度下降的反向传播算法进行模型优化。不同的是，在反向传播过程中，需要将连续值的梯度转化为脉冲形式，这需要通过一种名为 STBP(spiking temporal backpropagation)的优化方法来实现。

实验结果表明，Spiking-SSD 在准确率和速度方面都比传统的 SSD 算法有所提升，而且在能耗方面也具有明显的优势。同时，Spiking-SSD 还具有更好的生物合理性，因为它使用了离散化的脉冲信号来代替连续值，这更符合神经元的生物特性。不过，Spiking-SSD 仍然需要解决一些挑战，例如脉冲神经网络的训练效率不高、目标检测精度还需要进一步提高等问题。

5.2 脉冲神经网络的编码方式

人工脉冲神经网络旨在进行神经计算。为了实现目标检测任务，必须将图片信息输入到脉冲神经网络中，这不仅仅是需要对图片进行编码，还需要对图片数据与编码信息进行加工处理，与一般的人工神经网络输入信息不同，脉冲神经网络必须输入脉冲信号，而图片信息采用的是以像素点格式进行输入，因此，所有脉冲神经网络的第一层通常都是编码层，通过不同的编码将不同类 型的数据转换为脉冲信号才能输入到脉冲神经网络中。并且对于数据编码产生的频率有严格的要求，稍有不慎都会导致数据丢失。研究者从脉冲编码方式上取得突破：虽然在生物学上合理建立了脉冲神经元模型，但是由这些神经元所连接形成的神经网络系统与信息编码方式仍然是难以了解的。这需要给神经尖峰赋予意义：对计算重要的变量必须根据尖峰神经元与之通信的尖峰来定义。基于生物学知识，研究者提出了多种神经元信息编码：

5.2.1 二进制编码(Binary Coding)

二进制编码是脉冲神经网络中的一种编码方式，也被称为时间间隔编码(Interval Encoding)[60]。在二进制编码中，信息被编码为一组脉冲的时间间隔，

而不是脉冲的时刻。这些时间间隔由多个脉冲的位置和数量共同决定，通常采用二进制数的形式进行表示。

二进制编码是一种全有或全无的编码，其中神经元在特定时间间隔内处于活动状态或不活动状态，并在整个时间范围内发射一个或多个尖峰。二进制编码中，每个输入特征都对应一个神经元，而每个神经元会输出一组脉冲，表示该特征的二进制编码。例如，如果一个神经元对应的输入特征值为 3，则该神经元会输出一组间隔为 1、2、4 的三个脉冲，表示该特征的二进制编码为 011。

在二进制编码中，输入信息可以同时被多个神经元编码，每个神经元输出的脉冲可以表示不同的特征或特征的组合。这种方式可以提高网络的表达能力和识别能力，同时也可以增强网络的鲁棒性和容错性。

与其他编码方式相比，二进制编码具有一定的优势。首先，二进制编码的脉冲频率比率编码低，因此可以提高能量效率。其次，二进制编码具有良好的噪声容忍性，即使输入信号存在噪声或失真，也可以正确地识别输入特征。最后，由于二进制编码可以通过硬件电路实现，因此可以在低功耗、高速度的嵌入式系统上实现脉冲神经网络的实时计算。

因此，二进制编码在脉冲神经网络中得到了广泛的应用。例如，在基于脉冲神经网络的机器学习算法中，二进制编码已经成功地应用于图像识别、语音识别、模式识别等任务中。

5.2.2 率编码(Rate Coding)

率编码是 SNN 最常用的一种编码方式[61]。它将信息转化为一系列脉冲的发放频率，即脉冲数量与时间的比例。例如，如果一个神经元的输出为 1，则表示该神经元在某个时间段内发放了 n 个脉冲，其中 n 是该时间段的持续时间与脉冲发放频率之积。在这种编码方式下，神经元的输出是一个实数，可以通过类似于传统神经网络的梯度反向传播算法进行优化。

在这种编码方式下，只有一个区间内的尖峰率被用作速率编码中传达的信息的度量，这是对尖峰的定时性质的抽象。生理神经元为更强的(感觉或人工)刺激更频繁地激发这一事实激发了速率编码。

它可以用于单神经元水平或再次解释脉冲序列。在第一种情况下，神经元被直接描述为速率神经元，它在每个时间步将实数值输入“速率”转换为输出“速率”。在技术背景和认知研究中，速率编码一直是传统人工“S 形”神经元背后的概念。

5.2.3 时序编码(Temporal Coding)

时序编码(Temporal Coding)是脉冲神经网络中的一种编码方式，也称为相位编码(Phase Encoding)或阶段编码(Spike-Timing Encoding)[61]。在时序编码中，信息被编码为脉冲的时刻或时间间隔，不同于率编码和延迟编码中使用脉冲的数量或传输延迟来表示信息。

时序编码是一种将信息转化为时间间隔的编码方式。在这种编码方式下，信息被编码为脉冲的时间间隔或发放时间，而不是发放频率。例如，如果一个神经元发放两个脉冲，分别在时间t_1和时间t_2，那么它的输出就是一个包含两个时间点的时间序列$\{t_1, t_2\}$。在时序编码中，脉冲的时间间隔与信息的大小成反比，即脉冲间隔越短，信息量越大，反之亦然。时序编码需要更为复杂的信号处理和学习算法，在训练过程中需要考虑到脉冲的时间关系和神经元之间的相互影响，因此相对于率编码，时序编码的训练和优化更加困难。

在时序编码中，脉冲的发放时刻与输入信息的特征向量相关。通常情况下，如果一个神经元对应一个特征，那么输入的每一个特征就会产生一个脉冲。不同的特征产生的脉冲在时刻上有一定的相位差，也就是所谓的相对时刻(Relative Timing)。因此，输入信息就被编码为一系列脉冲的相对时刻和脉冲的数量。

时序编码中，相邻的脉冲之间的时间间隔通常是固定的，也就是说，脉冲的时刻由脉冲发放的位置和网络的时间基准点(或时钟信号)共同决定。当网络接收到一个输入时，每个特征会产生一个脉冲，而这些脉冲的时刻可以描述输入的特征向量，即通过这种方式对输入信息进行了编码。

时序编码在脉冲神经网络中得到广泛应用，具有很好的鲁棒性和容错性，可以应对输入中可能存在的噪声和失真。同时，由于时序编码的计算效率较高，可以在硬件上实现，因此被广泛应用于各种脉冲神经网络的实现中。

5.2.4 延迟编码(Delay Encoding)

延迟编码是脉冲神经网络中一种特殊的编码方式，也称为多项式编码(Polynomial Encoding)[62]。与常规的率编码和时序编码不同，延迟编码不是将信息表示为脉冲的数量或时间间隔，而是利用脉冲的传播延迟来表示信息。

在延迟编码中，不同的信息通过在网络中不同的路径上传播到达目标神经元。例如，假设输入层有 n 个神经元，分别编号为 1, 2, …, n，目标神经元为 m，如果第 i 个神经元发放脉冲，则它会通过第 i 条路径传输到达目标神经元 m。为了实现延迟编码，每条路径的传输延迟不同，即第 i 条路径的传输延迟为$d(i)$，其中$d(i)$是一个整数，表示脉冲传输到目标神经元需要的时间步数。这样，对于一个输入样本，目标神经元的输出可以表示为：

$$O(m) = \sum w(i) \times S\big(t - d(i)\big)$$

其中，$w(i)$是输入层第 i 个神经元与目标神经元之间的权重，$S(t)$是一个时间函数，表示网络中某个神经元在时刻 t 是否发放脉冲。

延迟编码具有较好的容错性和鲁棒性，因为对于输入中可能存在的噪声或失真，延迟编码可以通过对脉冲传输的路径和时间进行适当的调整来保证输出的正确性。

但是，由于需要设置多条路径和不同的传输延迟，延迟编码的实现比较复杂，且随着网络规模的增大，需要处理的路径数呈指数级增长，对硬件实现也提出了挑战。因此，目前延迟编码在实际应用中的使用相对较少，典型的主要有 SpikeProp 和 Chronotron 等。

5.3 基于脉冲神经网络的目标检测数据集

特别值得一提的是，在数据集方面，除了可以通过转换编码方式继续使用传统 ANN 领域积累的大量的目标检测成熟数据集(如小数据集 MNIST[47]、CIFAR-10[63]，大数据集 MS COCO[43]、CIFAR-100[63]、ImageNet[64]等)以外，脉冲神经网络还支持根据自身能处理包含时空序列的离散事件的优势，开发专用的事件流数据集。

通过 DVS 相机进行拍摄，可以方便地生成带时间流的图像数据，这意味着能够很好地处理高速运动的物体的记录问题。由此，诞生了许多专用的脉冲神经

网络时间流数据集，典型的如 DVS-MNIST[65]、DVS-CIFAR10[66]、DVS-Gesture[67][68]等。

通过定制专用于 SNN 的事件流数据集，SNN 的潜力可以被更加充分地发掘。目前，制作脉冲神经网络的时空事件流数据集已成为 SNN 进一步发展的主要驱动力之一。

5.4 思考和讨论

5.4.1 目标检测领域脉冲神经网络的优势及其评价

脉冲神经网络的优势是显而易见的，可以很容易地被归纳出来：

在目标检测领域，脉冲神经网络具有以下优势：

1) 低功耗和高效率：脉冲神经网络的计算模型基于事件驱动的脉冲信号，具有低功耗、高效率的特点，可以在较低的计算资源下完成目标检测任务。
2) 更高的编码效率：时间编码表明，单一的神经元可以取代上百个传统的 Sigmoid 型隐藏层节点。更高的编码效率使得第三代神经网络可以在节能型和计算效率上有质的提升。从信息论的角度来看，它以脑科学和认知神经科学的研究成果为基础，拓展了智能信息处理的方法，为解决复杂问题和智能控制提供有效的途径。
3) 适应动态环境：由于这种新型的神经网络采用脉冲编码(spike coding)，通过获得脉冲发生的精确时间，这种新型的神经网络可以进行获得更多的信息和更强的计算能力。例如，脉冲神经网络可以动态地响应环境变化和输入数据的变化，对于实时目标检测任务具有很强的应用能力。
4) 增强鲁棒性：脉冲神经网络具有一定的容错能力，可以处理噪声和失真的输入数据，提高了目标检测系统的鲁棒性和稳定性。
5) 并行计算能力：脉冲神经网络可以进行并行计算，将多个神经元同时处理输入信号，提高了计算速度和精度。

近年来，基于脉冲神经网络的目标检测算法也不断涌现，这些算法在目标检测精度和速度方面都具有一定的优势，为实现低功耗、高效率的目标检测系统提供了新的思路和技术手段。

5.4.2 目标检测领域脉冲神经网络的局限性及其认识

如上节所述，SNN 相比传统的 ANN 而言具有多方面的优势。但是，SNN 的应用也有一定的局限性。现实中，由于缺乏统一而有效的训练方法，导致大规模的脉冲神经网络计算的发展还很缓慢。目前的人工智能研究中，大规模 SNN 用于解决复杂子任务的成功案例还相对较少，而这些恰恰又是第二代神经网络所最为擅长的。

SNN 的局限性主要可以被归纳为以下几个方面：

1) 训练困难：SNN 的训练难度比传统的人工神经网络(ANN)更高，因为它的训练方法和 ANN 有很大的不同。与 ANN 使用的误差反向传播算法不同，SNN 使用的是脉冲驱动学习算法，因此训练 SNN 更困难。
2) 数据预处理：SNN 的输入必须是脉冲信号，因此对于不能直接转化为脉冲信号的数据，需要进行预处理。这增加了 SNN 的数据处理难度。

3) 网络结构：SNN 的网络结构设计比较困难，因为它必须考虑脉冲信号的传递和信息处理。如果网络结构设计不当，将影响 SNN 的性能。
4) 计算复杂度：SNN 的计算复杂度比较高，因为它需要同时处理大量的脉冲信号。这降低了 SNN 的实时性。
5) 参数敏感性：SNN 对参数的选择非常敏感，如果参数选择不当，将对 SNN 的性能产生影响。

也正是基于此，具体到目标检测领域，虽然基于脉冲神经网络的目标检测算法具有很好的生物合理性和能效性，但它们目前还面临着一些挑战和限制，包括：

1) 神经网络模型的复杂性：基于脉冲神经网络的目标检测算法需要设计和构建复杂的神经网络模型，这对于算法的实现和优化带来了挑战。
2) 训练效率的问题：目前，基于脉冲神经网络的目标检测算法的训练效率还不如传统的卷积神经网络。这是因为脉冲神经网络需要对连续值进行离散化，这会导致信息的丢失和噪声的引入。
3) 目标检测精度的限制：尽管基于脉冲神经网络的目标检测算法已经取得了一定的成果，但目前的算法性能还无法和传统的目标检测算法相比。

总之，SNN 在目标检测方面应用的局限性主要是在训练方面和计算复杂度方面。尽管如此，随着技术的不断发展，人们正在努力开发新的方法来解决这些问题。例如，一些研究者正在研究使用超级计算机来加速 SNN 的训练，同时其他研究者正在研究使用先进的算法来简化 SNN 的网络结构。另外，在许多情况下，SNN 的优点仍然可以充分发挥，并且其应用前景非常广阔。例如，在语音识别、图像识别和机器人控制等领域，SNN 仍然具有很大的潜力。

因此，虽然 SNN 存在一些局限性，但是它仍然是一个非常有前途的技术，有很大的应用潜力。

5.4.3 展望

总的来说，目前，的确存在着一些制约着基于脉冲神经网络的目标检测算法进一步发展的因素，这一方面是由于当前脉冲神经网络的发展还处于起步阶段，人们对脉冲神经网络的研究还有待进一步深入；另一方面，也是由当前脉冲神经网络本身结构特性所决定的。

因此，基于脉冲神经网络的目标检测算法进一步发展根本地还是依赖于脉冲神经网络相关理论的进一步深入和相关应用的进一步拓展。尤其是对基于脉冲神经网络的基本结构和训练方式的更深刻的、更加结合硬件算力发展方向的变革，是最为关键的。

无论如何，基于脉冲神经网络的目标检测算法是一个非常新颖的研究领域，它具有很大的潜力和发展前景。随着技术的不断进步，我们相信这些算法在未来将会得到更广泛的应用和发展。

6. 结论

随着计算能力的快速增长，以深度神经网络为代表的人工神经网络方法一度大放异彩，并成为我们这个时代的“显学”。深度学习方法一度在模式识别的许多领域都显示出突破性的性能。尽管它们在分层特征提取和分类方面很有效，但这些类型的神经网络计算成本高昂，并且难以在便携式设备的硬件上实现。在神经网络架构的另一项研究中，SNN 被描述为节能模型，因为它们的稀疏，基于峰

值的通信框架。最近的研究试图利用这两个框架(深度学习和 SNN)来开发多层 SNN 架构，以实现最近证明的深度网络的高性能，同时实现生物启发的节能平台。

但是，随着任务场景的日益复杂，模型的参数量和计算量也急剧增长，由此带来更高的存储、计算和功耗需求。而基于神经形态计算的脉冲神经网络(Spiking Neuron Networks，SNN)自身具有计算量小、功耗低、信息传递速度快等优点，为上述问题提供很好的解决方案。然而脉冲神经网络固有的训练困难的特点，导致其难以推广应用于复杂的机器视觉应用中，所以，面向新模型结构和训练策略的探索对脉冲神经网络的研究和工程应用具有重要的意义。

脉冲神经网络是一种基于事件驱动的神经网络，其特点是低功耗、高效率、并行计算、适应动态环境和增强鲁棒性等。在目标检测领域，脉冲神经网络已经被广泛应用，并且一些基于脉冲神经网络的目标检测算法不断涌现。这些算法具有低功耗、高效率、高精度和快速响应的特点，有望为实现低功耗、高效率的目标检测系统提供新的思路和技术手段。因此，脉冲神经网络在目标检测领域具有重要的应用价值。

本文回顾了 SNN 以及基于其的目标检测方法，以解决该领域的一些开放性问题。文献表明，SNN 神经元以脉冲的形式进行通信，不同于传统的神经网络模型。相比于传统神经网络，SNN 具有低能耗、低能耗鲁棒性、可塑性和生物真实性等诸多优势，被相关研究寄予厚望。但同时，也面临了训练困难、模型复杂度偏高、任务适应性偏差、缺乏广泛接受的标准和统一的评估方法等一些弱点。

总之，SNN 是一种有潜力的神经网络模型，具有许多传统神经网络所不具备的优点，但是仍然存在许多挑战和问题需要进一步解决和探索。

7. 参考文献

[1] Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research[J]. Journal of pharmaceutical and biomedical analysis, 2000, 22(5): 717-727.

[2] 陶文涛. BP 神经网络与脉冲神经网络的时空信息结合[D]. 大连理工大学, 2021.

[3] 马婷, 李万杰, 冯佳楠, 等. 光脉冲神经网络研究进展[J]. OPTICS & OPTOELECTRONIC TECHNOLOGY, 2022, 20(4): 96-111.

[4] 毛健, 赵红东, 姚婧婧. 人工神经网络的发展及应用[J]. 电子设计工程, 2011, 19(24): 62-65.

[5] Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation[R]. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[6] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533-536.

[7] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal processing magazine, 2012, 29(6): 82-97.

[8] 周祥全, 张津. 深层网络中的梯度消失现象[J]. 科技展望, 2017, 27(27): 284.

[9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Tavanaei A, Ghodrati M, Kheradpisheh S R, et al. Deep learning in spiking neural networks[J]. Neural networks, 2019, 111: 47-63.

[12] Ghosh-Dastidar S, Adeli H. Spiking neural networks[J]. International journal of neural systems, 2009, 19(04): 295-308.

[13] Ponulak F, Kasinski A. Introduction to spiking neural networks: Information processing, learning and applications[J]. Acta neurobiologiae experimentalis, 2011, 71(4): 409-433.

[14] Vreeken J. Spiking neural networks, an introduction[J]. 2003.

[15] Hodgkin A L, Huxley A F. Propagation of electrical signals along giant nerve fibres[J]. Proceedings of the Royal Society of London. Series B-Biological Sciences, 1952, 140(899): 177-183.

[16] Xin J, Embrechts M J. Supervised learning with spiking neural networks[C]//IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222). IEEE, 2001, 3: 1772-1777.

[17] Iqbal U, Milan A, Gall J. PoseTrack: Joint Multi-person Pose Estimation and Tracking[C]// Computer Vision and Pattern Recognition. IEEE, 2017:4654-4663.

[18] S. Johnson and M. Everingham. Learning effective human pose estimation from inaccurate annotation. In CVPR, pages 1465–1472. IEEE, 2011.

[19] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ra- manan, P. Dollar, and C. L. Zitnick. Microsoft coco: Com- mon objects in context. In ECCV, 2014.

[20] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 3686– 3693. IEEE, 2014.

[21] Fischler M A, Elschlager R A. The Representation and Matching of Pictorial Structures[J]. IEEE Trans Computers C, 1973, 22(1):67-92.
[22] Wei S E, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines[J]. 2016:4724-4732.
[23] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts[J]. 2011, 32(14):1385-1392.
[24] Toshev A, Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks[C]// Computer Vision and Pattern Recognition. IEEE, 2014:1653-1660.
[25] Papandreou G, Zhu T, Kanazawa N, et al. Towards Accurate Multi-person Pose Estimation in the Wild[J]. 2017:3711-3719.
[26] Huang S, Gong M, Tao D. A Coarse-Fine Network for Keypoint Localization[C]// IEEE International Conference on Computer Vision. IEEE, 2017:3047-3056.
[27] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. 2017.
[28] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Pictorial Structures for Object Recognition[J]. International Journal of Computer Vision, 2005, 61(1):55-79.
[29] Ren X, Ramanan D. Histograms of sparse codes for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 3246-3253. Chen Y, Wang Z, Peng Y, et al. Cascaded Pyramid Network for Multi-Person Pose Estimation[J]. 2017.
[30] Fang H S, Xie S, Tai Y W, et al. RMPE: Regional Multi-person Pose Estimation[J]. 2016:2353-2362.
[31] Eichner M, Marin-Jimenez M, Zisserman A, et al. 2d articulated human pose estimation and retrieval in (almost) unconstrained still images[J]. International journal of computer vision, 2012, 99: 190-214.
[32] Josyula R, Ostadabbas S. A Review on Human Pose Estimation[J]. arXiv preprint arXiv:2110.06877, 2021.
[33] Guo Y, Cui H, Li S. Excavator joint node-based pose estimation using lightweight fully convolutional network[J]. Automation in Construction, 2022, 141: 104435.
[34] Needham L, Evans M, Cosker D P, et al. The accuracy of several pose estimation methods for 3D joint centre localisation[J]. Scientific reports, 2021, 11(1): 20673.
[35] Xia F, Wang P, Chen X, et al. Joint Multi-person Pose Estimation and Semantic Part Segmentation[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:6080-6089.
[36] Cao Z, Simon T, Wei S E, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[J]. 2016:1302-1310.
[37] Newell A, Huang Z, Deng J. Associative Embedding: End-to-End Learning for Joint Detection and Grouping[J]. 2016.
[38] Papandreou G, Zhu T, Chen L C, et al. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model[J]. 2018.
[39] Luvizon D C, Picard D, Tabia H. 2d/3d pose estimation and action recognition using multitask deep learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5137-5146.
[40] Lee K, Lee I, Lee S. Propagating lstm: 3d pose estimation based on joint interdependency[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 119-135.
[41] Akhter I, Jalal A, Kim K. Adaptive pose estimation for gait event detection using context-aware

model and hierarchical optimization[J]. Journal of Electrical Engineering & Technology, 2021, 16: 2721-2729.

[42] Rohrbach A, Rohrbach M, Tandon N, et al. A dataset for movie description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3202-3212.

[43] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.

[44] Hagemann S. An improved land surface parameter dataset for global and regional climate models[J]. 2002.

[45] Chen X, Yuille A L. Articulated pose estimation by a graphical model with image dependent pairwise relations[J]. Advances in neural information processing systems, 2014, 27.

[46] Ionescu C, Papava D, Olaru V, et al. Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 36(7): 1325-1339.

[47] Zou Z, Chen K, Shi Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023.

[48] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Ieee, 2001, 1: I-I.

[49] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Ieee, 2005, 1: 886-893.

[50] Felzenszwalb P F, Girshick R B, McAllester D A, et al. Discriminative latent variable models for object detection[C]//Proceedings of the 27th International Conference on Machine Learning (ICML-10). 2010: 11-12.

[51] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9759-9768.

[52] Tian Z, Shen C, Chen H, et al. Fcos: A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 1922-1933.

[53] Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 840-849.

[54] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[55] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[56] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[57] Kim S, Park S, Na B, et al. Spiking-yolo: spiking neural network for energy-efficient object detection[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11270-11277.

[58] Tavanaei A, Kirby Z, Maida A S. Training spiking convnets by stdp and gradient descent[C]//2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018: 1-8.

[59] Mtetwa N, Smith L S. Smoothing and thresholding in neuronal spike detection[J]. Neurocomputing, 2006, 69(10-12): 1366-1370.

[60] Al-Hamid A A, Kim H W. Optimization of Spiking Neural Networks Based on Binary Streamed Rate Coding[J]. Electronics, 2020, 9(10): 1599.

[61] Kiselev M. Rate coding vs. temporal coding-is optimum between?[C]//2016 international joint conference on neural networks (IJCNN). IEEE, 2016: 1355-1359.

[62] Almomani A, Alauthman M, Alweshah M, et al. A comparative study on spiking neural network encoding schema: implemented with cloud computing[J]. Cluster Computing, 2019, 22: 419-433.

[63] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web][J]. IEEE signal processing magazine, 2012, 29(6): 141-142.

[64] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[65] Abouelnaga Y, Ali O S, Rady H, et al. Cifar-10: Knn-based ensemble of classifiers[C]//2016 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2016: 1192-1195.

[66] Henderson J A, Gibson T T A, Wiles J. Spike event based learning in neural networks[J]. arXiv preprint arXiv:1502.05777, 2015.

[67] Vicente-Sola A, Manna D L, Kirkland P, et al. Evaluating the temporal understanding of neural networks on event-based action recognition with DVS-Gesture-Chain[J]. arXiv preprint arXiv:2209.14915, 2022.

[68] Zheng H, Wu Y, Deng L, et al. Going deeper with directly-trained larger spiking neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(12): 11062-11070.