西南科技大学信息工程学院

硕士研究生文献综述报告

年级：2021 级 专业：信息与通信工程

姓名：吴豪　学号：7020210232

光谱重建与自注意力机制文献综述

1. 光谱重建的由来和发展现状

光谱，作为光与物质相互作用后承载物质“指纹”信息的载体，能够有效反映目标物体的理化特性。通过光谱分析设备，将物体的辐射（或者反射信号）分离成不同波长的光谱信息，利用所获取的高分辨率光谱信息结合物体的“指纹效应”，可以实现对目标特性的识别。因此，作为一种非接触式、连续在线、可多组分测量且灵敏度高、检测范围广的检测手段，光谱检测技术在生物医药、食品检验、化学分析，医疗保健，环境监测，远程探测，半导体工业，太阳能，物联网，安全控制，防伪等诸多领域都有着重要应用。

1.1 传统光谱仪到计算光谱仪的发展

光谱探测已经有了悠久的历史。第一次光谱探测是由艾萨克·牛顿在 1665 年进行的（牛顿，1979 年），使用三角形棱镜将阳光分割成彩虹色的图案。然而，直到 1859 年，基尔霍夫和邦森（1861 年）开发了第一个实用的光谱仪，定量光谱测量才有可能实现。20 世纪 60 年代，随着半导体和光电子器件的发展，直读光谱仪诞生，使光谱数据的存储和处理更加方便。自 20 世纪 80 年代以来，随着二维（2D）电荷耦合器件（CCD）阵列的出现，更好的光学设计、改进的电子技术和先进的制造都将性能提高了一个数量级，将频谱采集仪器带入了一个繁荣的时代。光谱成像技术从此得到了广泛的应用（Hagen 和

Kudenov，2013）[1]。然而，传统的光谱仪和光谱成像设备仍然存在成本高、体积大、重量重等诸多缺点。

近年来，随着计算机科学的发展，越来越多的硬件实现可以通过软件编程来实现，为成像仪器带来了紧凑、廉价、快速的组件。由于许多计算方法已被引入光谱检测（班加罗尔等人，1996 年；Vigneau 等人，1997 年；黑川等人，2011 年；Rajwade 等人，2013 年），特别是压缩感知(Candes 等人，2006 年；Donoho，2006；巴拉尼克，2007；蜡烛和 Wakin，2008)[2][3][4][5][6][8][9][10]和计算重建算法，最初的缺点有望被克服。此外，廉价、轻、小频谱采集设备可以广泛应用，降低了频谱检测的阈值。即使在智能手机平台上实现便携式传感系统也已成为可能（Das 等人，2016 年）[11]。

近年来，人们发展出了各种不同的计算光谱检测方法。一般用来区分是否是计算光谱检测系统首先看检测系统数据的收集是否为光谱调制或下采样，并需要变换后才能得到光谱数据。其次，需要采用复杂的算法从原始数据中提取光谱信息。然而，一些非计算实现采用或多或少的计算方法，如去噪算法、超分辨率方法和/或数据映射。

我们主要是关注那些原始数据可以通过只应用计算变换来理解的方法。坦率地说，系统硬件获取的原始数据与软件转换的输出数据完全不一样。这使得计算成为该系统的核心，它将计算实现与传统的光谱信息采集系统区分开来。非计算和计算光谱检测原理图如图 1 所示。

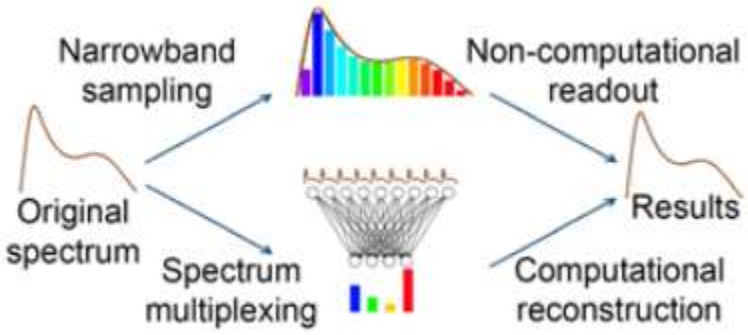

图 1 非计算和计算方法光谱检测的比较

1.2 计算光谱仪

光谱仪是检测光谱的最典型的仪器。它获取一维（一维）光谱信息。根据基本的操作原理，计算光谱仪可以归类为基于光栅的（沃尔芬布特特尔，2004；Chaganti 等人，2006）[11][12]和基于滤波片的光谱仪。

1.2.1 基于光栅的编码光径计算型光谱仪

如图 2 所示，传统的基于光栅的光谱仪由五个部分组成，即狭缝入口、准直光学、色散元件（通常是光栅）、聚焦光学和探测器阵列（Wolffenbettel，2004）[13]。硬件系统基于光栅元件的色散特性，对测试光源的狭缝部分进行采样，并在探测器阵列平面上形成跨越的光谱分布。因此，通过测量散射光的强度分布来获得光谱的形状。然而，在基于狭缝的色散光谱仪的设计中有一个主要的权衡，即光谱分辨率与光的吞吐量。在保持光谱分辨率的同时，增加狭缝光谱仪中的光吞吐量需要更高的狭缝和探测器，从而增加了系统的大小和成本（Cull et al.，2007）[14]。在不牺牲光谱分辨率的情况下最大化光谱仪吞吐量的设计被认为具有一种优势（大面积或吞吐量）（Jacquinot，1960）[15]。利用编码孔径和计算方法，可以打破权衡。

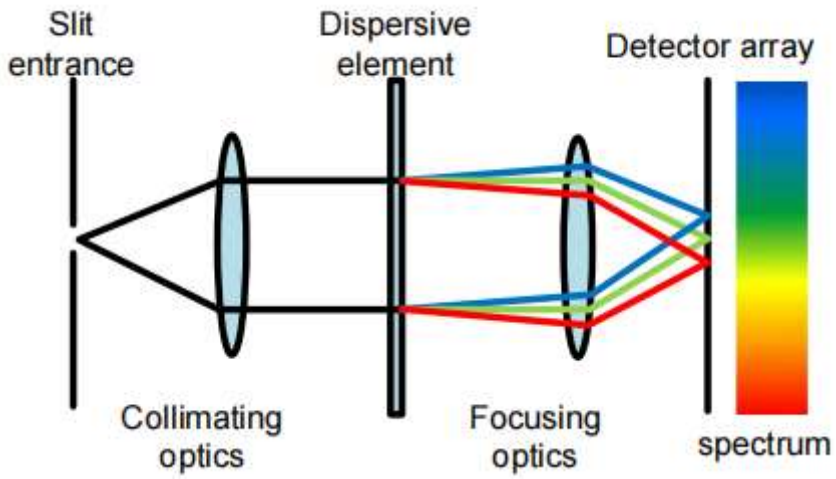

图 2 传统光栅基光谱仪示意图

事实上，实现编码孔径光谱的方法已经进行了很长时间了。在 20 世纪 50 年代早期，Golay（1949,1951）[16]创建了第一个编码孔径光谱，而 Girard（1963）也做了一些改进。随着数学方法的发展，阿达玛变换（HT）光谱成为编码孔径光谱的大多数（Decker，1971；汉森和斯特朗，1972 年；菲利普斯和布里奥塔，1974 年；Swift 等人，1976 年）。然而，这些 HT 光谱只有单勒耳探测器或有限的离散探测器阵列和掩模运动部分。

Gehm 等人（2006）[18]提出了一种二维编码孔径方法，该方法优化用于漫射源的光谱表征，称为静态多模态和多重光谱仪（静态 MMS）。Gehm 等人（2006）从数学上证明了二维编码必须满足正交性约束。可以通过基于任何一个正交函数族的输入孔径模式来满足设计要求。通过用双行阿达玛矩阵的正交列码替换狭缝，系统在不牺牲光谱分辨率的同时提高了光吞吐量和信噪比（信噪比），而不牺牲光谱分辨率。Gehm 等人（2007）[19]通过用全息光栅取代色散元件改进了静态 MMS，并将探测器阵列从单色 CCD 改为彩色阵列（Cull 等人，2007）[20]。全息图光栅被设计为有三个不同的光谱波段，中心波长对应于蓝、绿、红光。使用非负小平方算法对全息光栅和 CCD 拜耳滤波器之间的光谱响应差异引起的数据位移进行了校准（Cull et al.，2007）[21]。

1.2.2 基于滤波片的计算光谱仪

基于滤光片的光谱仪的基本结构与基于光栅的光谱仪基本相同，但有一些微小的不同。硬件主要由三个部分组成，即光收集光学、滤光器和探测器。集收集部分从测试样品中收集光线，制作入射光的光学参数（如入射角、通量和杂散光水平）更适合于滤波和检测。当光通过滤光片传输时，不同的波长分量在顺序或空间上被分离。最后，由检测器或检测器阵列测量每个分量，从而由系统获得测试样本的光谱信息。

通常，选择带通滤波片来执行波长分解任务。为了获得更高的光谱分辨率，使用了更多的通带更窄的滤波器。这种策略增加了整个系统的体积和复杂性。同时，当光谱响应曲线变窄时，光通量下降，导致信噪比降低。在计算实现中，滤波器是宽带的，这使得原始数据看起来与原始频谱完全不同。然而，通过应用计算重建算法，可以解决频谱分辨率。由于宽带滤光片允许更多的光通过，它们允许从较暗的场景中检测光谱。根据压缩传感理论，可以恢复稀疏谱高概率使用适当设计传感滤波器，而过滤器的数量比期望的频谱通道（从低维向量恢复高维向量）。另一方面，通过应用更大的滤波器数，可以使用正则化算法（从高维向量噪低维向量）来降低噪声，这提高了信噪比，使整个系统更加健壮。

Chang 和 Lee（2008）[22]展示了一种基于低性能和低成本滤波器阵列的芯片上的精细光谱仪，将光信号的输出经过 CCD 传感器转换为电信号输入到一个数字信号处理器中（Chang and Lee，2008）[23]。在原型系统中，滤波片数为 40，频谱数据采用非负约束最小二乘（NNLS）[24]算法进行重构，如图 3 所示。

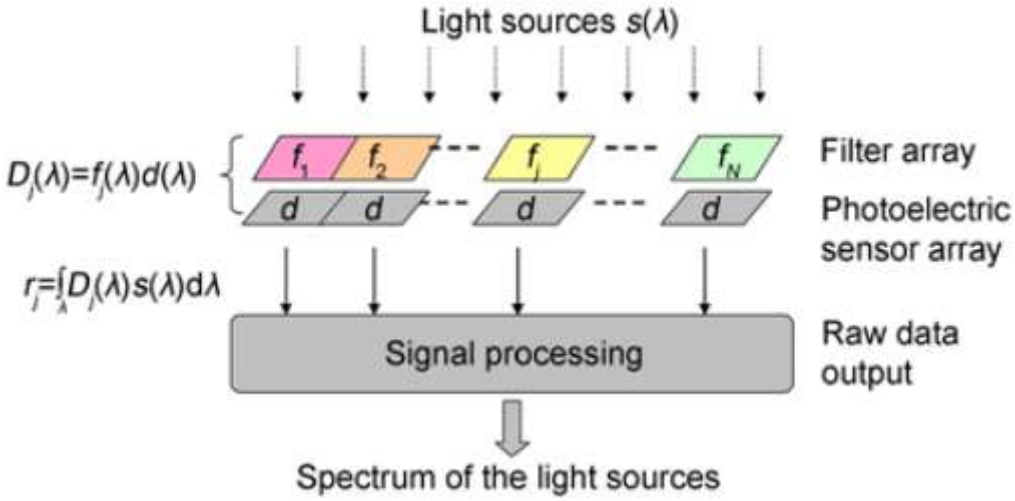

图 3 基于滤波器阵列的光谱仪示意图

Chang 等人（2011b）[25]改进了他们的软件实现，并对重建精度进行了复杂的分析。重构方法采用高斯核模板进行降噪，并将算法改进为 l1-范数最小化方法。在仿真和实际测量中，对 NNLS、Tikhono 夫正则化非负最小二乘（TNNLS）和 l1-范数算法进行了比较。测量结果很大程度上依赖于早期训练和高斯基数的选择，这限制了应用前景。Oliver 等人（2012）（与 Chang 相同的团队）利用了信号频谱的稀疏性，表明分辨率可以进一步提高，模仿压缩感知（CS）的一些工作（Candes 等人，2006；Donoho，2006；巴拉尼克，2007；蜡烛和韦金，2008）[26][27][28]关于使用随机矩阵进行信号采集和恢复，利用 40 个滤波片，当频谱通道数 N 达到 405 时，重构均方误差（MSE）控制在−5dB 以下。值得注意的是，该分辨率来自于一个基于 MSE 的定义，它有效地代表了稀疏信号重建的精灵辅助 MSE 水平（Oliver et al.，2013）[29]。

光谱信息采集技术在过去的几十年里发展迅速。由于新材料和计算算法的应用，许多新的实现已经出现，向我们展示了光谱仪和光谱成像的新可能性。与传统设备相比，大多数计算光谱学设备的工作精度可能仍然较低。然而，从无处不在的应用程序的角度来看，如基于智能手机的检测或基于频谱的识别，

这一缺陷变得不那么关键了。相反，这些方法以其低成本、重量轻、紧凑的外壳而突出。这些优势总是比在这类应用程序中追求完美的性能更有吸引力。近年来，人们一直在努力对消费者使用的光谱检测。我们相信，硬件和计算算法的共同设计将导致光谱的广泛应用随着物联网（IoT）和人工智能（AI）的发展，无处不在的光谱仪和光谱成像将为人类带来好处。

2.注意力机制

人类的注意力是所有知觉和认知操作的一种核心属性。因为我们在处理信息来源的能力有限，注意机制会选择、调节和关注与行为最相关的信息。几十年来，注意力的概念和功能一直在哲学、心理学、神经科学和计算等领域进行研究，而在过去的六年里，这一特性在深度神经网络中得到了广泛的探索与应用。

2.1 介绍

注意是一种行为和认知过程，选择性地关注信息的离散方面，无论是主观的还是客观的，而忽略其他可感知的信息，在人类认知和生物的生存中起着重要作用。[30]对于人类来说注意是必要的，因为在任何时候，环境呈现的感知信息都比有效处理的要多，记忆包含了比记忆更多的竞争特征，可用的选择、任务或运动反应比所能处理的要大得多。[31]

在过去的几十年里，注意力的概念已经渗透到感知和认知研究的大部分方面，被认为是多重和不同的感知和认知操作的一个属性。注意力已经成为定义大脑如何控制其信息处理的一个广义术语，它的影响可以通过有意识的内省、电生理学和大脑成像来衡量，因此长期以来，人们一直从不同的角度来研究和关注注意力机制。注意力机制模仿了生物观察行为的内部过程，即一种将内部

经验和外部感觉对齐从而增加部分区域的观察精细度的机制。注意力机制可以快速提取稀疏数据的重要特征，因而被广泛用于自然语言处理任务，特别是机器翻译。而自注意力机制是注意力机制的改进，其减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性。

2.2 深度学习模型出现前的注意力机制的使用

基于心理物理模型的计算注意系统，由神经生物学证据支持，已经存在了至少 30 年的[39]。特雷斯曼的特征集成理论（FIT）[40]，沃尔夫的指南搜索[41]，三元架构[42]，布罗德本特的模型[43]，诺曼注意模型[44]，闭环注意模[45]型，选择性注意模型[46]，以及其他几个模型，介绍了计算注意系统的理论基础。

最初，注意力主要是通过视觉实验来研究的，其中一个被试看一个随着时间[47]变化的场景。在这些模型中，注意系统仅局限于视觉搜索任务中的选择性注意成分，专注于通过传感器提取多个特征。因此，大多数注意计算模型出现在计算机视觉中，以选择重要的图像区域。Koch 和 Ullman[48]介绍了该地区第一个基于 FIT 的视觉注意架构[39]。它背后的想法是，几个特征是并行计算的，它们的显著性被收集在一个显著性地图上。赢家通吃（WTA）确定了地图上最突出的区域，它最终被路由到中央表示。从那时起，只有感兴趣的区域继续进行更具体的处理。神经形态视觉工具包（NVT）来源于 Koch-Ullman 模型[49]，是多年来发展计算视觉注意力研究的基础。纳瓦尔帕坎和 ITTI 引入了 NVT[50]的衍生物，它可以处理自上而下的线索。其思想是从一个训练图像中学习目标的特征值，其中一个二值掩模表示目标。Hamker 的注意力系统[51][52]会计算出各种特征和对比地图，并将其转化为感知地图。

在深度学习（DL）之前，计算注意系统被成功地应用于目标识别[54]、图像压缩[55]、图像匹配[56]、图像分割[57]、目标跟踪[58]、主动视觉[59]、人机交互[60]、机器人中的目标操作[61]、机器人导航和 SLAM[62]。在 1997 年中期，Scheier 和 Egner 提出了一个利用注意力进行导航的移动机器人。尽管如此，在 90 年代，巴鲁贾和波莫洛使用了一个注意力系统来导航自动驾驶汽车，并跟踪投影地图的相关区域。Walther 将注意系统与基于 SIFT 特征的目标识别器相结合，证明了注意前端增强了识别结果。Salah 等人将注意力与神经网络结合在可观察马尔可夫模型中，用于手写数字识别和人脸识别。乌尔哈尼等人提出了聚焦图像压缩，它根据图像的显著性来决定分配给图像编码区域的位数。高显著性区域对图像的其余部分具有高质量的重建能力。

2.3 注意力开始被应用到深度学习中

到 2014 年，DL 社区注意到关注是推进深度神经网络的一个基本概念。目前，该领域最先进的技术是使用神经注意模型。如图 4 所示，在领先的存储库中，已发布的作品数量每年都在显著增长。在神经网络中，注意机制动态地管理信息流、特征和可用资源，从而改善学习能力[65][66][67]。这些机制过滤掉了任务的无关刺激，帮助网络简单地处理长期依赖。

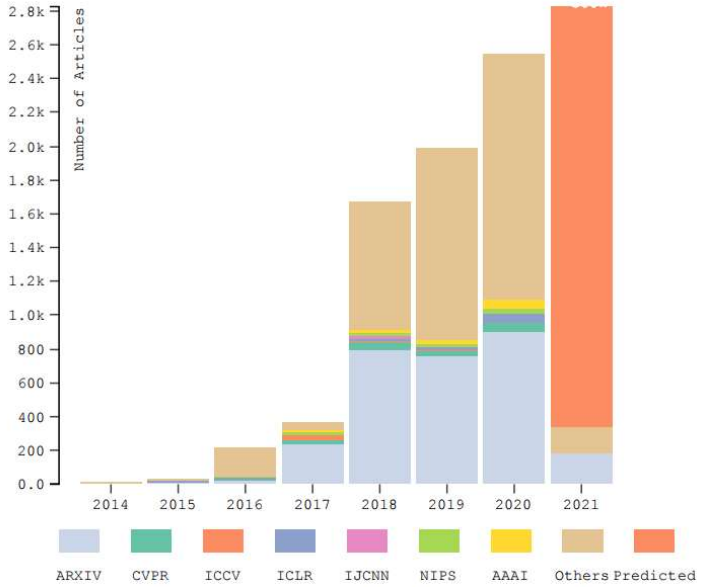

图 4 2014 年 1 月 1 日至 2021 年 2 月 15 日期间全年出版的作品

历史上，计算注意力系统的研究自 20 世纪 80 年代就存在。直到 2014 年年中，神经注意网络（NANs）才出现在自然语言处理（NLP）中，人们的关注提供了重大进展，通过可扩展和直接的网络带来了有希望的结果。注意力让我们转向复杂的任务，如会话机器理解、情绪分析、机器翻译、问题回答和迁移学习，以前具有挑战性。随后，nan 出现在其他对人工智能同样重要的领域，如计算机视觉、强化学习和机器人技术。目前有许多注意架构，但很少有显著的更高的影响，如图 5 所示。在这张图片中，描述了根据引文水平和创新组织的最相关的作品组，其中 RNNSearch[73]、变压器[72]、记忆网络[73]、“展示、参加和讲述”[74]和 RAM[75]是关键的发展。

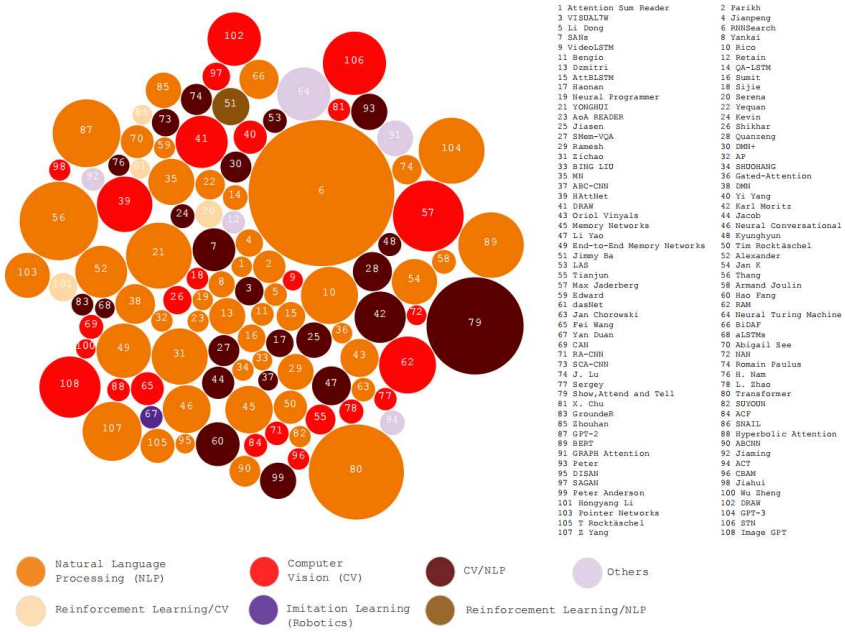

图 5 主要神经注意网络（NAN）。

经典的编解码器框架中的瓶颈问题是深度学习中注意力研究的初始动机。在这个框架中，编码器将一个源句编码为一个固定长度的向量，解码器从其中生成翻译。Cho 等人表明[76]，经典编解码器的性能随着输入句子大小的增加而迅速恶化。为了最小化这一瓶颈，Bahdanau 等人提出了 RNNSearch，这是对编码-解码器模型的扩展，学习对齐和翻译在一起。RNNSearch 在每个时间步长生成一个翻译的单词，在源句中寻找最相关单词的位置。注意机制允许额外的信息通过网络传播，消除了固定大小的上下文向量的信息瓶颈。

2.3.1RNNSearch

RNNSearch 是注意研究的基础，该体系结构的注意模块被广泛地应用。在语音识别[48]中，允许一个 RNN 处理音频[77]，而另一个 RNN 在生成描述时集中检查相关部分。文本内分析[78]，它允许模型在生成分析树时查看单词。在会话建模中，它允许模型在生成响应时关注会话的最后部分。BiDAF 提出了一种多阶段层次的问答过程。

Yang 等人提出了层次注意网络（HAN）[81]来获取关于文档结构的两个基本见解。文档具有层次结构：单词构成句子，句子构成文档。同样，人类通过首先构建句子表示，然后将它们聚合为文档表示来构建文档表示。文档中不同的单词和句子可以提供不同的信息。Xiong 等人[82]创建了一个协同注意编码器，它捕获问题和文档之间的交互，通过一个动态指向解码器在估计回答跨度的开始和结束之间交替。为了学习计算上棘手的问题的近似解，Ptr-Net[83]修改了 RNNSearch 的注意机制来表示可变长度的字典。它使用注意机制作为一个指针。

而 FusionNet[84]采用了一种完全感知的多层次注意机制和一个利用单词历史的注意评分函数。该机制允许模型参与过去的输出向量，解决了 LSTM 的细胞状态瓶颈，让有注意的 LSTM 不需要捕获 LSTM 单元状态中前提的整个语义。相反，注意力在读取前提并积累单元状态的表示时生成输出向量，该表示通知第二个 LSTM 注意哪个前提的输出向量来确定 RTE 类。

人类的视觉注意机制可以在突出图像相关部分的同时探索图像的局部差异。一个人同时将注意力集中在图像的某些部分上，在识别过程中快速扫描整个图像以找到主要区域。在这个过程中，不同区域的内部关系引导眼睛的运动，以找到下一个需要聚焦的区域，忽略那些不相关的部分会使它在存在混乱时更容易学习。卷积神经网络（CNNs）是非常不同的，cnn 是线性的，参数的数量随图像的大小呈线性增长。此外，为了让网络捕获像素之间的长距离依赖关系，体系结构需要有多层，这影响了模型的收敛性。此外，该网络以相同的方式处理所有像素。这个过程不像人类的视觉系统，它包含视觉注意机制和一瞥结构，在物体识别中提供无与伦比的表现。

RAM [85]和 STN 是基于人类视觉注意的注意界面的开创性架构。RAM [86]可以通过自适应地选择一系列区域，从图像或视频中提取信息，并且只处理高分辨率的选定区域。遵循 RAM 方法，深度循环注意作者（DRAW）代表了一种更自然的图像构造方式的改变，其中场景的某些部分是独立于其他部分创建的。这个过程是人类如何通过按顺序重新创建一个视觉场景来绘制一个场景，为多次迭代细化绘制的所有部分，并在每次修改后重新评估他们的工作。

2.3.2 端到端注意模型

2017 年年中，出现了针对端到端关注模型的研究。神经转换器（NT）[66]和图注意网络[96]——纯注意结构——向科学界证明了注意力是深度学习未来发展的关键因素。该变压器的目标是使用自我注意来最小化传统的递归神经网络的困难。神经变压器是第一个只使用注意模块和全连接的神经网络来成功处理序列数据的神经结构。它分配递归和卷积，捕获序列元素之间的关系，而不管它们之间的距离如何。注意允许变压器是简单，并行，低成本的训练。图注意网络（GATs）是 gnn 的端到端注意版本[93]。它们有成堆的注意层，帮助模型关注于非结构化数据中最相关的部分来做出决策。注意的主要目的是通过提高信噪比（SNR），同时降低结构的复杂性，从而避免图中的噪声部分。

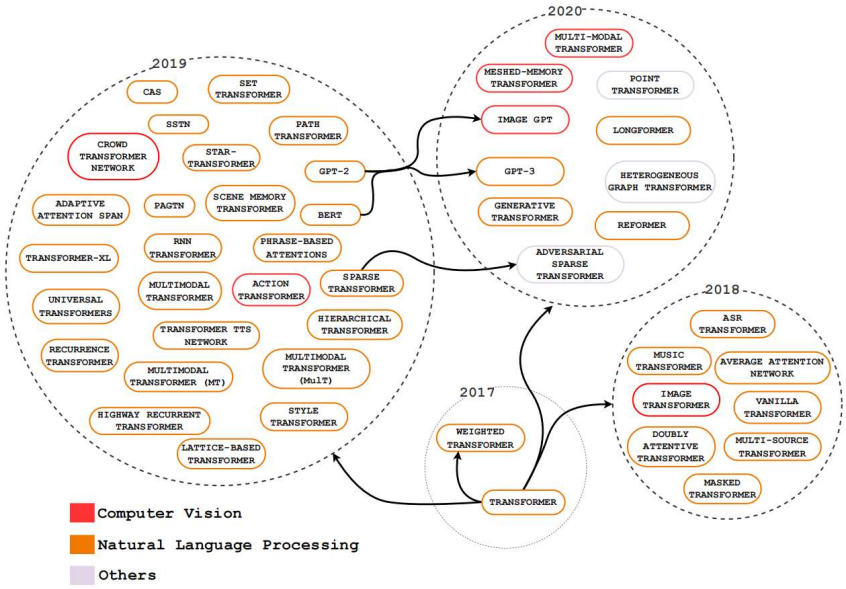

图 6 基于变压器的神经网络

人们对 NT 和 gat 的兴趣越来越大，并提出了一些扩展的，包括许多基于变压器的架构[98][99][100][101]，如图 6 所示。这些结构和所有使用自我注意的结构都属于一种新的神经网络，称为自我注意神经网络。他们的目标是探索自我关注各种任务，改进了以下缺点：1）大量参数，训练迭代收敛；2）每层内存成本高，按序列长度内存二次增长；3）自回归模型；4）解码器层的并行性低。具体来说，加权变压器[102]提出对注意层进行修改，实现 40 %的收敛。多头注意模块被模型在训练过程中学习匹配的分支注意模块所取代。Lee 等人还具有一种注意机制，可以将自我注意从二次减少到线性，允许对高输入和数据集进行缩放。

一些方法使变压器适应于新的应用和领域。在自然语言处理过程中，出现了几种新的架构，主要是在多模态学习中。双注意变压器提出了一种包含视觉信息的多模态机器翻译方法。它修改了注意解码器，允许来自预先训练过的 CNN 编码器的文本特征和视觉特征。多源变压器探索了四种不同的策略，将输入组合到多头注意解码器层进行多模态转换。样式变压器、分层变压器、高速

循环变压器、基于晶格的变压器、变压器 TTS 网络、基于短语的注意[114]是样式转换、文档摘要和机器翻译中的一些重要架构。N 语言 P 的迁移学习是变压器的主要贡献领域之一。基于 BERT[105] 、GPT-2 [106]和基于 GPT-3[107]的 NT 体系结构解决了 NLP 中的迁移学习问题，因为目前的技术限制了预先训练的表示的能力。在计算机视觉中，图像的生成是变压器的一个好消息。图像变压器[108]，SAGAN[109]和 GmageGPT[110]使用自我注意机制来参加当地的社区。尽管每层比典型的卷积神经网络保持了明显更大的接受域，但该模型在实践中可以处理的图像的大小显著增加。最近，在 2021 年初，OpenAi 向科学界介绍了 DALL · E[111]，这是基于变压器和 GPT-3 的最新语言模型，能够从文本中生成图像，扩展 GPT-3 的知识，仅需 120 亿个参数即可查看。

2.3.3 注意在深度学习中的现状

目前，采用了在深度学习中使用注意力的主要关键发展的混合模型（图 7）已经引起了科学界的兴趣。主要是基于变压器、gat 和记忆网络的混合模型出现在多模态学习和其他几个应用领域。双曲注意网络（HAN）[113]、双曲图注意网络（GHN）[114]、时间图网络（TGN）[115]和基于记忆的图网络（MGN）[87]是最有前途的发展方向之一。双曲网络是一类新的体系结构，它结合了自注意、记忆、图和双曲几何在激活神经网络方面的好处，以对深度神经网络产生的高容量嵌入进行推理。自 2019 年以来，这些网络作为一个新的研究分支而脱颖而出，因为它们代表了神经机器翻译、图形学习和视觉问题回答任务的最先进的泛化，同时保持了神经表征的紧凑。自 2019 年以来，盖茨还由于能够学习从生物学、粒子物理学、社会网络到推荐系统的复杂关系或相互作用而受到了广泛关注。为了改进节点的表示，扩大 gat 处理动态性质（即随时

间变化的特征或连接）数据的能力，提出了结合内存模块和时间维度的架构，如 mgn 和 tgn。

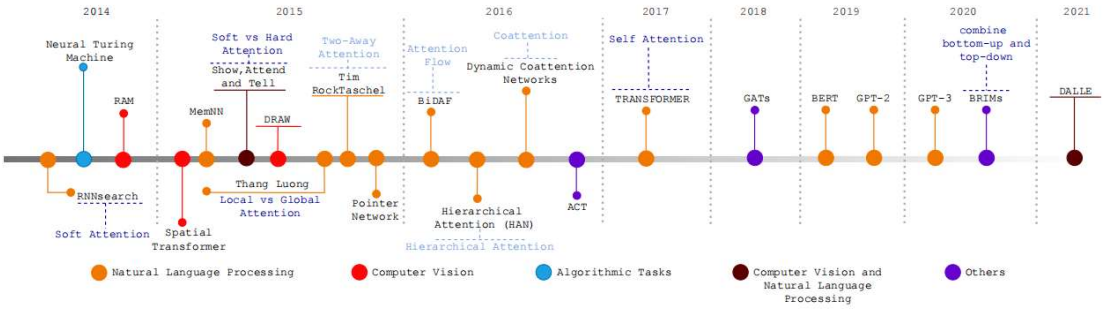

图 7 DL 时间轴中关注的关键发展

到 2020 年底，文献中仍未被探索的两个研究分支得到了加强： 1)双向递归神经网络中自下而上和自上而下刺激的显式组合，2)自适应计算时间。经典的递归神经网络在特定的表示级别内执行重复迭代，而不是使用自上而下的迭代，其中较高的层次在较低的层次上工作。然而，Mittal 等人[125]通过注意机制回顾了双向循环层，明确地引导自下而上和自上而下的信息的流动，促进了两个刺激层次之间的选择迭代。该方法将隐藏状态分为几个模块，使自下而上和自上而下的信号之间的向上迭代能够得到适当的集中。该层结构具有并发模块，因此每个分层层都可以以自下而上和自上而下的方向发送信息。

2.4 注意力机制

深度学习中的注意机制可分为软注意（全局注意）、硬注意（局部注意）和自我注意（内部注意）。

软的注意。软注意为每个输入元素分配了一个 0 到 1 的权重。考虑到深度神经网络的机制和目标之间的相互依赖性，它决定了对每个元素的关注。它使用在注意层中的 softmax 函数来计算权重，使整个注意模型具有确定性和可微性。软注意可以在空间和时间环境中起作用。空间语境主要是提取特征或最相关特征的权重。对于时间上下文，它通过调整滑动时间窗口中所有样本的权重

来工作，因为不同时间的样本有不同的贡献。尽管软机制是确定性的和可微的，但软机制对于大的输入具有很高的计算成本。图 8 显示了一个软注意机制的直观示例。

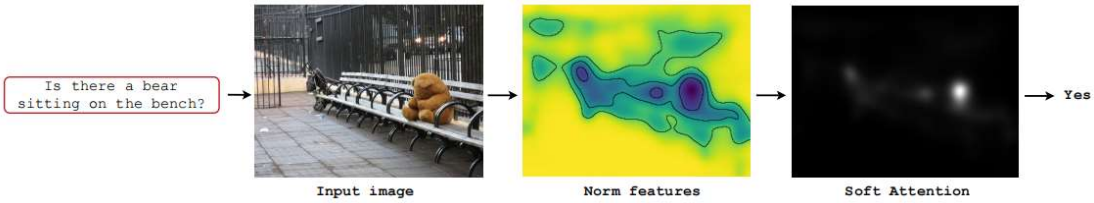

图 8 软注意力的一个直观的例子。

硬性注意力硬注意决定了是否应该考虑机制的一部分输入，反映了机制的输入与深度神经网络的目标之间的相互依赖关系。分配给输入部件的权重为 0 或 1。因此，当可以看到输入元素时，目标是不可微的。这个过程包括对参加哪个部分进行一系列的选择。例如，在时间上下文中，模型关注输入的一部分以获取信息，并根据已知信息决定下一步的位置。一个神经网络可以根据这些信息来做出选择。然而，由于没有基本的事实来表明正确的选择策略，硬注意类型的机制是由随机过程来表示的。由于模型是不可微的，强化学习技术需要集中训练模型。与软机制相比，一旦整个输入没有被存储或处理，推理时间和计算成本就会减少。图 9 显示了一个硬注意机制的直观示例。

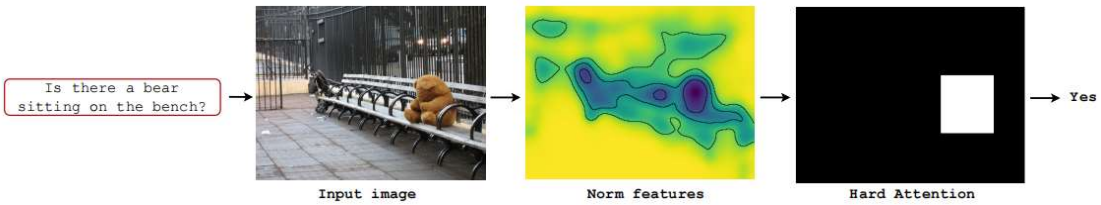

图 9 硬注意力的一个直观的例子

自注意力自我注意量化了该机制的输入元素之间的相互依赖性。这种机制允许输入相互“自我”，并决定他们应该更关注什么。与软机制和硬机制相比，自注意层的主要优点是对长输入的并行计算能力。这个机制层使用简单且

易于并行的矩阵计算，用所有相同的输入元素来检查注意力。图 10 显示了一个自我注意机制的直观示例。

图 10 自我注意的例子

2.5 注意力机制的优势

然而，在过去的十年里，计算能力的急剧提高，以及处理器价格和尺寸的降低，使得人们的注意力转向了基于相对简单且往往是无序的设备框架的光谱重建方案。我们相信这些系统代表了最有前途的范例，因为它们的性能不仅可以通过增强硬件来提高，有时还可以通过更直接地优化支持它们的软件来提高。通过基于机器学习的技术的进一步发展和优化，这一趋势似乎可能会继续下去，其中伴随的处理系统的计算能力可以承担提高光谱分辨率的大部分负担。随着这些光谱重建算法的成熟，它们将越来越能够更多地弥补进一步小型化所需要的探测器性能上的妥协，允许超紧凑但高性能的系统。在这方面仍然存在明显的障碍。例如，深度学习算法通常需要非常大的、有标记的数据集来正确地训练所使用的神经网络，以便在测量值和重建的频谱之间建立准确的关系。而注意力机制有 4 大优势，参数更少、速度更快、效果更好和更易解析。其中特别是速度更快这个优势，因为 Attention 不像 RNN 需要等待上一个单元的预测结果，因此可以并行处理。同时，可解析性也是十分有意义的一个优势，阿里图灵实验室也：一个模型如果它更加透明，能让别人知道它内部的原

理（例如，在情感分析上，这个模型是根据什么来判断文本的情感的），让用户认为它是可靠、安全。

参考文献

[1]Hagen NA, Kudenov MW, 2013. Review of snapshot spectral imaging technologies. *Opt Eng*, 52(9):090901. https://doi.org/10.1117/1.oe.52.9.090901

[2]Bulygin TV, Vishnyakov GN, 1992. Spectrotomography: a new method of obtaining spectrograms of two dimensional objects. Analytical Methods for Optical Tomography, p.315-323. https://doi.org/10.1117/12.131904

[3]Candès EJ, Wakin MB, 2008. An introduction to compressive sampling. *IEEE Signal Process Mag*, 25(2):21-30. https://doi.org/10.1109/msp.2007.914731

[4]Candès EJ, Romberg J, Tao T, 2006. Robust uncertainty prin-ciples: exact signal reconstruction from highly incom-plete frequency information. *IEEE Trans Inform Theory*, 52(2):489-509. https://doi.org/10.1109/tit.2005.862083

[5]Baraniuk RG, 2007. Compressive sensing. *IEEE Signal Process Mag*, 24(4):118-121. https://doi.org/10.1109/msp.2007.4286571

Rajwade A, Kittle D, Tsai TH, et al., 2013. Coded hyperspectral imaging and blind compressive sensing. *SIAM J Imag Sci*, 6(2):782-812. https://doi.org/10.1137/120875302

[6]Vigneau E, Devaux MF, Qannari EM, et al., 1997. Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration. *JChemomet*, 11(3):239-249. https://doi.org/10.1002/(SICI)1099-128X(199705)11:3< 239::AID-CEM470>3.0.co;2-A

[7]Kurokawa U, Choi BI, Chang CC, 2011. Filter-based miniature spectrometers: spectrum reconstruction using adaptive regularization. *IEEE Sens J*, 11(7):1556-1563. https://doi.org/10.1109/jsen.2010.2103054

[8]Das AJ, Wahi A, Kothari I, et al., 2016. Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness. *Sci Rep*, 6:32504. https://doi.org/10.1038/srep32504

[9]Okamoto T, Yamaguchi I, 1991. Simultaneous acquisition of spectral image information. *Opt Lett*, 16(16):1277-1279. https://doi.org/10.1364/ol.16.001277

[10]Oliver J, Lee W, Park S, et al., 2012. Improving resolution of miniature spectrometers by exploiting sparse nature of signals. *Opt Expr*, 20(3):2613-2625.

[11]Oliver J, Lee WB, Lee HN, 2013. Filters with random transmittance for improving resolution in filter-array-based spectrometers. *Opt Expr*, 21(4):3969-3989. https://doi.org/10.1364/oe.21.003969

[12]Hansen P, Strong J, 1972. High resolution Hadamard transform spectrometer. *Appl Opt*, 11(3):502-506. https://doi.org/10.1364/AO.11.000502

[13]Hayes MH, 1996. Statistical Digital Signal Processing and Modeling. John Wiley & Sons, New York, USA.

[14]Golay MJE, 1951. Static multislit spectrometry and its application to the panoramic display of infrared spectra. *J Opt Soc Am*, 41(7):468-472. https://doi.org/10.1364/josa.41.000468
[15]Wang Z, Yi S, Chen A, et al., 2019. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nat Commun*, 10(1):1020. https://doi.org/10.1038/s41467-019-08994-5
[16]Willett RM, Gehm ME, Brady DJ, 2007. Multiscale reconstruction for computational spectral imaging. Computational Imaging V, Article 64980L. https://doi.org/10.1117/12.715711
[17]Wolffenbuttel RF, 2004. State-of-the-art in integrated optical microspectrometers. *IEEE Trans Instrum Meas*, 53(1): 197-202. https://doi.org/10.1109/tim.2003.821490
[18]Soldevila F, Irles E, Durán V, et al., 2013. Single-pixel polarimetric imaging spectrometer by compressive sensing. *Appl Phys B*, 113(4):551-558.
[19]Shaltout A, Liu JJ, Kildishev A, et al., 2015. Photonic spin Hall effect in gap—plasmon metasurfaces for on-chip chiroptical spectroscopy. *Optica*, 2(10):860-863. https://doi.org/10.1364/optica.2.000860
[20]Rueda H, Arguello H, Arce GR, 2015. DMD-based implementation of patterned optical filter arrays for compressive spectral imaging. *J Opt Soc Am A*, 32(1):80-89. https://doi.org/10.1364/JOSAA.32.000080
[21]Sun T, Kelly K, 2009. Compressive sensing hyperspectral imager. Computational Optical Sensing and Imaging,
Article CTuA5. https://doi.org/10.1364/COSI.2009.CTuA5
[22]Swift RD, Wattson RB, Decker JA, et al., 1976. Hadamard transform imager and imaging spectrometer. *Appl Opt*, 15(6):1595-1609. https://doi.org/10.1364/AO.15.001595
[23]Takhar D, Laska JN, Wakin MB, et al., 2006. A new compressive imaging camera architecture using optical
domain compression. Computational Imaging IV, Article 606509. https://doi.org/10.1117/12.659602
[24]Wagadarikar A, John R, Willett R, et al., 2008. Single disperser design for coded aperture snapshot spectral imaging. *Appl Opt*, 47(10):B44-B51. https://doi.org/10.1364/ao.47.000b44
[25]Wagadarikar AA, Pitsianis NP, Sun XB, et al., 2009. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Opt Expr*, 17(8):6368-6388. https://doi.org/10.1364/oe.17.006368
[26]Wang LZ, Xiong ZW, Gao DH, et al., 2015. Dual-camera design for coded aperture snapshot spectral imaging. *Appl Opt*, 54(4):848-858. https://doi.org/10.1364/ao.54.000848
[27]Redding B, Liew SF, Sarma R, et al., 2013. Compact spectrometer based on a disordered photonic chip. *Nat Photon*, 7(9):746-751. https://doi.org/10.1038/nphoton.2013.190
[28]Ren WY, Fu C, Arce GR, 2018. The first result of compressed channeled imaging spectropolarimeter. Imaging and Applied Optics, Article JTu4A.21. https://doi.org/10.1364/3D.2018.JTu4A.21

[29]Huang E, Ma Q, Liu ZW, 2017. Etalon array reconstructive spectrometry. *Sci Rep*, 7:40693. https://doi.org/10.1038/srep40693
[30] Esther Luna Colombini, A da Silva Simoes, and CHC Ribeiro. *An attentional model for intelligent robotics agents.* PhD thesis, Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil, 2014.
[31] Marvin M Chun, Julie D Golomb, and Nicholas B Turk-Browne. A taxonomy of external and internal attention. *Annual review of psychology*, 62:73–101, 2011.
[32] Roger BH Tootell, Nouchine Hadjikhani, E Kevin Hall, Sean Marrett, Wim Vanduffel, J Thomas Vaughan, and Anders M Dale. The retinotopy of visual spatial attention. *Neuron*, 21(6):1409–1422, 1998.
[33] Marty G Woldorff, Christopher C Gallen, Scott A Hampson, Steven A Hillyard, Christo Pantev, David Sobel, and Floyd E Bloom. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, 90(18):8722–8726, 1993.
[34] Heidi Johansen-Berg and Donna M Lloyd. The physiology and psychology of selective attention to touch. *Front Biosci*, 5:D894–D904, 2000.
[35] Christina Zelano, Moustafa Bensafifi, Jess Porter, Joel Mainland, Brad Johnson, Elizabeth Bremner, Christina Telles, Rehan Khan, and Noam Sobel. Attentional modulation in human primary olfactory cortex. *Nature neuroscience*, 8(1):114–120, 2005.
[36] Maria G Veldhuizen, Genevieve Bender, R Todd Constable, and Dana M Small. Trying to detect taste in a tasteless solution: modulation of early gustatory cortex by attention to taste. *Chemical Senses*, 32(6):569–581, 2007.
[37] William James. *The Principles of Psychology*. Dover Publications, 1890.
[38] Simone Frintrop, Erich Rome, and Henrik Christensen. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 7, 01 2010.
[39] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97– 136, 1980.
[40] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3):419, 1989.
[41] Ronald A Rensink. The dynamic representation of scenes. *Visual cognition*, 7(1-3):17–42, 2000.
[42] Donald Eric Broadbent. *Perception and communication*. Elsevier, 2013.
[43] Donald A Norman. Toward a theory of memory and attention. *Psychological review*, 75(6):522, 1968.
[44] Daniel Kahneman. *Attention and effort*, volume 1063. Citeseer, 1973.
[45] Frank Van der Velde, Marc de Kamps, et al. Clam: Closed-loop attention model for visual search. *Neurocomputing*, 58:607–612, 2004.
[46] R Hans Phaf, AHC Van der Heijden, and Patrick TW Hudson. Slam: A connectionist model for attention in visual selection tasks. *Cognitive psychology*, 22(3):273–341, 1990.

[47] Simone Frintrop, Erich Rome, and Henrik I Christensen. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception (TAP)*, 7(1):1–39, 2010. 40A PREPRINT - APRIL 1, 2021
[48] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.
[49] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE PAMI*, 20(11):1254–1259, 1998.
[50] Vidhya Navalpakkam and Laurent Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *2006 IEEE CVPR)*, volume 2, pages 2049–2056. IEEE, 2006.
[51] Fred H Hamker. The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision. *Computer Vision and Image Understanding*, 100(1-2):64–106, 2005.
[52] Fred H Hamker. Modeling feature-based attention as an active top-down inference process. *BioSystems*, 86(1-3):91–99, 2006.
[53] Simone Frintrop. *VOCUS: A visual attention system for object detection and goal-directed search*, volume 3899. Springer, 2006.
[54] Albert Ali Salah, Ethem Alpaydin, and Lale Akarun. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE PAMI*, 24(3):420–425, 2002.
[55] Nabil Ouerhani. *Visual attention: from bio-inspired modeling to real-time implementation*. PhD thesis, Université de Neuchâtel, 2003.
[56] Dirk Walther. *Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics*. PhD thesis, California Institute of Technology, 2006.
[57] Dirk Walther, Duane R Edgington, and Christof Koch. Detection and tracking of objects in underwater video. In *Proc. of the IEEE CVPR*, volume 1, pages I–I. IEEE, 2004.
[58] James J Clark and Nicola J Ferrier. Modal control of an attentive vision system. In *IEEE ICCV*, pages 514–523. IEEE, 1988.
[59] Cynthia Breazeal and Brian Scassellati. A context-dependent attention system for a social robot. *rn*, 255:3, 1999.
[60] A Rotenstein, Alexander Andreopoulos, Ehzan Fazl, David Jacob, Matt Robinson, Ksenia Shubina, Yuliang Zhu, and J Tsotsos. Towards the dream of intelligent, visually-guided wheelchairs. In *Proc. 2nd Int'l Conf. on Technology and Aging*, 2007.
[61] James J Clark and Nicola J Ferrier. Attentive visual servoing. In *Active vision*. Citeseer, 1992.
[62] Simone Frintrop and Patric Jensfelt. Attentional landmarks and active gaze control for visual slam. *IEEE Transactions on Robotics*, 24(5):1054–1065, 2008.
[63] Christian Scheier and Steffen Egner. Visual attention in a mobile robot. In *ISIE'97 Proceeding of the IEEE International Symposium on Industrial Electronics*, volume 1, pages SS48–SS52. IEEE, 1997.

[64] Shumeet Baluja and Dean A Pomerleau. Expectation-based selective attention for visual monitoring and control of a robot vehicle. *Robotics and autonomous systems*, 22(3-4):329–344, 1997.
[65] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*, 2015.
[66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv:1706.03762 [cs]*, June 2017. arXiv: 1706.03762.
[67] Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.
[68] Feng Wang and David MJ Tax. Survey on the attention based rnn model and its applications in computer vision. *arXiv preprint arXiv:1601.06823*, 2016.
[69] Dichao Hu. An introductory survey on attention mechanisms in nlp problems. In *Proceedings of SAI Intelligent Systems Conference*, pages 432–448. Springer, 2019.
[70] Andrea Galassi, Marco Lippi, and Paolo Torroni. Attention in natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
[71] John Boaz Lee, Ryan A. Rossi, Sungchul Kim, Nesreen K. Ahmed, and Eunyee Koh. Attention models in graphs: A survey. *arXiv:1807.07984 [cs]*, July 2018. arXiv: 1807.07984.
[72] Sneha Chaudhari, Gungor Polatkan, Rohan Ramanath, and Varun Mithal. An attentive survey of attention models. *arXiv preprint arXiv:1904.02874*, 2019.
[73] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014. 41A PREPRINT - APRIL 1, 2021
[74] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning*, pages 2048–2057, June 2015.
[75] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. *arXiv preprint arXiv:1406.6247*, 2014.
[76] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
[77] William Chan, Navdeep Jaitly, Quoc V Le, and Oriol Vinyals. Listen, attend and spell. *arXiv preprint arXiv:1508.01211*, 2015.
[78] Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. Grammar as a foreign language. In *Advances in neural information processing systems*, pages 2773–2781, 2015.
[79] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.
[80] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flflow for machine comprehension. *arXiv:1611.01603 [cs]*, November 2016. arXiv: 1611.01603.

[81] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. In *Advances in neural information processing systems*, pages 2440–2448, 2015.
[82] Seoung Wug Oh, Joon-Young Lee, Ning Xu, and Seon Joo Kim. Video object segmentation using space-time memory networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9226–9235, 2019.
[83] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. In *International conference on machine learning*, pages 1378–1387. PMLR, 2016.
[84] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
[85] Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. Key value memory networks for directly reading documents. *arXiv preprint arXiv:1606.03126*, 2016.
[86] Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. Memory graph networks for explainable memory-grounded question answering. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 728–736, 2019.
[87] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
[88] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *The World Wide Web Conference*, pages 2022–2032, 2019.
[89] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10296–10305, 2019.
[90] Sami Abu-El-Haija, Bryan Perozzi, Rami Al-Rfou, and Alexander A Alemi. Watch your step: Learning node embeddings via graph attention. In *Advances in Neural Information Processing Systems*, pages 9180–9190, 2018.
[91] Linjie Li, Zhe Gan, Yu Cheng, and Jingjing Liu. Relation-aware graph attention network for visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10313–10322, 2019.
[92] Karim Ahmed, Nitish Shirish Keskar, and Richard Socher. Weighted transformer network for machine translation. *arXiv preprint arXiv:1711.02132*, 2017.
[93] Qipeng Guo, Xipeng Qiu, Pengfei Liu, Yunfan Shao, Xiangyang Xue, and Zheng Zhang. Star-transformer. *arXiv preprint arXiv:1902.09113*, 2019.
[94] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, Monica Dinculescu, and Douglas Eck. Music transformer. *arXiv preprint arXiv:1809.04281*, 2018.

[95] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers. *arXiv preprint arXiv:1904.10509*, 2019.
[96] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. *arXiv preprint arXiv:1810.00825*, 2018.
[97] Hasan Sait Arslan, Mark Fishel, and Gholamreza Anbarjafari. Doubly attentive transformer machine translation. *arXiv preprint arXiv:1807.11605*, 2018.
[98] Jindrich Libovicky, Jindrich Helcl, and David Marecek. Input combination strategies for multi-source transformer decoder. *arXiv preprint arXiv:1811.04716*, 2018.
[99] Ning Dai, Jianze Liang, Xipeng Qiu, and Xuanjing Huang. Style transformer: Unpaired text style transfer without disentangled latent representation. *arXiv preprint arXiv:1905.05621*, 2019.
[100] Yang Liu and Mirella Lapata. Hierarchical transformers for multi-document summarization. *arXiv preprint arXiv:1905.13164*, 2019. 44A PREPRINT - APRIL 1, 2021
[101] Ting-Rui Chiang, Chao-Wei Huang, Shang-Yu Su, and Yun-Nung Chen. Learning multi-level information for dialogue response selection by highway recurrent transformer. *arXiv preprint arXiv:1903.08953*, 2019.
[102] Fengshun Xiao, Jiangtong Li, Hai Zhao, Rui Wang, and Kehai Chen. Lattice-based transformer encoder for neural machine translation. *arXiv preprint arXiv:1906.01282*, 2019.
[103] Naihan Li, Shujie Liu, Yanqing Liu, Sheng Zhao, and Ming Liu. Neural speech synthesis with transformer network. In *Proceedings of the AAAI Conference on Artifificial Intelligence*, pages 6706–6713, 2019.
[104] Phi Xuan Nguyen and Shafifiq Joty. Phrase-based attentions. *arXiv preprint arXiv:1810.03444*, 2018.
[105] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805 [cs]*, October 2018. arXiv: 1810.04805.
[106] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.
[107] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.
[108] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. Image transformer. In *International Conference on Machine Learning*, pages 4055–4064. PMLR, 2018.
[109] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International conference on machine learning*, pages 7354–7363. PMLR, 2019.

[110] Mark Chen, Alec Radford, Rewon Child, Jeff Wu, Heewoo Jun, Prafulla Dhariwal, David Luan, and Ilya Sutskever. Generative pretraining from pixels. In *Proceedings of the 37th International Conference on Machine Learning*, volume 1, 2020.
[111] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, and Scott Gray. Dall·e: Creating images from text, 2021.
[112] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, et al. Hyperbolic attention networks. *arXiv preprint arXiv:1805.09786*, 2018.
[113] Yiding Zhang, Xiao Wang, Xunqiang Jiang, Chuan Shi, and Yanfang Ye. Hyperbolic graph attention network. *arXiv preprint arXiv:1912.03046*, 2019.
[114] Emanuele Rossi, Ben Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.
[115] Sarthak Mittal, Alex Lamb, Anirudh Goyal, Vikram Voleti, Murray Shanahan, Guillaume Lajoie, Michael Mozer, and Yoshua Bengio. Learning to combine top-down and bottom-up signals in recurrent neural networks with attention over modules. In *International Conference on Machine Learning*, pages 6972–6986. PMLR, 2020.
[116] Alex Graves. Adaptive computation time for recurrent neural networks. *arXiv:1603.08983 [cs]*, March 2016. arXiv: 1603.08983.
[117] Michael Figurnov, Maxwell D Collins, Yukun Zhu, Li Zhang, Jonathan Huang, Dmitry Vetrov, and Ruslan Salakhutdinov. Spatially adaptive computation time for residual networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1039–1048, 2017.
[118] Cristobal Eyzaguirre and Alvaro Soto. Differentiable adaptive computation time for visual reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12817–12825, 2020.
[119] Dongbin Zhao, Yaran Chen, and Le Lv. Deep reinforcement learning with visual attention for vehicle classifification. *IEEE Transactions on Cognitive and Developmental Systems*, 9(4):356–367, 2016. attention. *arXiv preprint arXiv:1511.04119*, 2015.

导师评价：

导师签名：

分数：

年　　月　　日