

# 注意力机制在医学上的应用综述

盛诗梦<sup>1</sup>, 丁皓<sup>2</sup>, 徐欣茹<sup>1</sup>, 丁思吉<sup>1</sup>, 夏冬阳<sup>1</sup>

<sup>1</sup>上海理工大学健康科学与工程学院, 上海

<sup>2</sup>上海健康医学院医疗器械学院, 上海

收稿日期: 2022年10月11日; 录用日期: 2022年11月28日; 发布日期: 2022年12月8日

## 摘要

注意力机制作为一种解决资源分配的手段, 用来解决资源过载问题。最开始仅应用在机器翻译、文本处理等领域, 最近十年也开始被应用于图像处理领域中, 考虑将其与医学领域相结合, 用注意力解决医学相关问题, 有利于提高我国医疗服务行业的发展。文章介绍了注意力机制并总结了注意力机制在医学上的主要应用: 医学图像检测与分割、医学图像分类、疾病预测、行为识别。在医学领域引入注意力机制, 有利于提高医生的诊断效率, 缩短就诊时间, 也为虚拟现实技术的进一步发展带来了可能。

## 关键词

注意力机制, 图像检测, 图像分割, 行为识别, 医学

# A Review of the Application of Attention Mechanism in Medicine

Shimeng Sheng<sup>1</sup>, Hao Ding<sup>2</sup>, Xinru Xu<sup>1</sup>, Siji Ding<sup>1</sup>, Dongyang Xia<sup>1</sup>

<sup>1</sup>College of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

<sup>2</sup>College of medical Devices, Shanghai University of Medicine and Health Sciences, Shanghai

Received: Oct. 11<sup>th</sup>, 2022; accepted: Nov. 28<sup>th</sup>, 2022; published: Dec. 8<sup>th</sup>, 2022

## Abstract

As a means of resource allocation, attention mechanism is used to solve the problem of resource overload. At first, it was only used in machine translation, text processing and other fields. In the last decade, it has also been used in the field of image processing. It is considered to combine it with the medical field to solve medical related problems with attention, which is conducive to improving the development of China's medical service industry. This paper introduces the attention mechanism and summarizes its main applications in medicine: medical image detection and seg-

mentation, medical image classification, disease prediction, and behavior recognition. The introduction of attention mechanism in the medical field is conducive to improving the diagnostic efficiency of doctors, shortening the time for medical treatment, and also brings possibilities for the further development of virtual reality technology.

## Keywords

Attention Mechanism, Image Detection, Image Segmentation, Behavior Recognition, Medical Science

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着科学的迅猛发展及媒体技术的广泛应用, 社会各个领域已全面迈入信息时代, 现有的信息处理技术也面临着革新。而医学作为与人们生活密不可分的重要领域, 其接收和需要处理的信息量更是与日俱增。如仅依靠人类的大脑去进行如此量大、复杂且高难度的信息分析工作必将大大提高人工成本, 且效率得不到保证。此时, 人工智能的出现, 使这些困境得以迎刃而解。人工智能的发展不仅能通过缩短医生的审片、阅片时间等方式整体提高诊断效率, 还能通过远程询问和自主治疗便利了病人的就诊过程。人工智能在推动医疗行业进步的同时, 也加大了医疗行业对人工智能的需求。如今, 神经网络在医学领域上有着广泛的应用, 涉及基础医学到临床医学的各个方面[1], 其中在医学影像识别、健康监测、疾病风险预测、药物研发和辅助疾病诊断这五个方面最为普遍[2]。而注意力机制作为目前深度学习领域的主流方法和研究热点之一, 它使得传统深度学习框架中存在的不足与缺陷得以修正, 性能得以提升。注意力机制通过对模型中输入信息需要关注的不同部分来赋予不同的权重系数, 从中抽取更加集中和重要的关键信息, 辅助模型进行更为准确的判断, 在医学上逐渐成为辅助医生工作的手段之一, 为医生的工作提供了便利。

本文基于此, 以注意力的框架为最初起点, 介绍了注意力机制的分类以及目前注意力机制在医学领域上的应用现状, 并在最后对注意力机制的未来发展做出展望, 使其能够更好地作为一种辅助手段, 有效地帮助医生解决问题与提高工作效率。

## 2. 注意力机制

### 2.1. 原理及定义

注意是心理活动对一定事物的指向和集中, 而注意力是衡量注意程度的指标, 它由注意的广度、稳定性、分配和转移的力度组成[3]。注意力的思想首先是在图像识别领域, 由 Mnih 等人[4]于 2014 年发表中提出, 它是一种模拟人脑注意力机制的模型。

人们在观察或倾听一件事情上, 往往不会一次性将所有内容吸收, 反而选择性地关注他们所认为重要的一部分, 而忽略掉他所认为的不重要的部分, 并且随着时间的推移, 人的视线注意力会根据想要关注部分的变化而转移[5], 这就是注意力。而在无意识地关注某个个体或部分、忽略其他个体或部分的过程就是权重分配过程。高权重意味着把更多的注意力聚焦在重要的事件上, 而低权重则是弱化不重要的信息, 在这一过程中, 会不断调整权重分配。不同的权重分配会直接影响到结果的输出, 合理的权重分

配则会提高结果的处理效率,以便更快速、更高效地达到目的。目前,Guo Jiabao 和 Liu Gang [6]将注意力机制与 BiLSTM (Bi-directional Long Short-Term Memory)相结合用于文本分类;Li Yong 等人[7]使用带有注意力机制的 CNN,可感知人脸的遮挡区域并专注于最具辨别力的未遮挡区域,以提高人脸面部表情的识别精度;Chen Yuantao 等人[8]利用注意力机制进行了图像超分辨率重建,有效改善了图像的视觉效果。注意力机制已成为目前神经网络架构中越来越普遍的成分,应用于各种任务中[9]。

目前,有三种与注意力机制结合使用的网络框架[9]。其一是基于经典的 Encoder-Decoder 框架,也是大多数模型使用的框架;其二是与记忆网络结合使用,将一些与任务相关的信息保存在辅助记忆中,然后在需要时读取;最后一种则是特殊的神经网络结构,其在捕获长距离依赖关系的过程中可以不使用 RNN。在这三种框架中,主流还是基于 Encoder-Decoder 框架。

## 2.2. 分类

### 2.2.1. 软注意力与硬注意力

2014年,Xu 等人[10]根据注意力模型在同一时间步长内所关注的区域是指定的还是整体的,将注意力机制分为了软注意力机制与硬注意力机制。

软注意力机制:它的分类分布是在一系列元素上计算的,同时考虑所有输入,不同的概率反映了不同元素的重要性。

硬注意力机制:它是一个随机过程。在某一时刻只关注一个元素,摒弃其余元素,注意力更加集中。

软注意力模型和硬注意力模型都是解决神经网络问题中的重要手段。软注意力模型在自然语言处理中比较受欢迎[11],被用于机器翻译[12],图像字幕[13]等,最近也开始用于图像分类[14],但并不常见。与软注意力模型相比,硬注意力模型的计算成本更低,但由于不可微且无法进行反向传播[4],依赖于强化学习,且难以扩展到复杂的数据集,所以训练起来效率很低。Shen Tao 等人[15]将硬注意力模型和软注意力模型整合为“增强自我注意力模型(ReSA)”,利用两种模型各自的优势运用于自然语言处理任务中。后将这两种优势上的混合体看作“局部注意力”。

Item-wise 和 location-wise 是按照输入特征序列划分的两种方式。前者的输入特征是一个项目、单个特征图或者单个词,后者则针对难以获得不同输入项的序列。它们可以分别与软、硬注意力模型相结合解决问题,共有四种组合方式。

### 2.2.2. 共同注意力

2016年,Lu Jiasen 等人[16]在改进视觉问答系统时,基于现有的注意力模型,提出了一种新颖的“共同注意力模型”,即将问题和图像联结起来,共同推理视觉注意和问题注意。目前,共同注意模型应用于情绪分析[17]、诊断预测[18]、图像分割[19]等,但主要还是以视觉问答为主[20] [21] [22]。而根据生成图像和问题注意的顺序,共同注意力又可分为平行共同注意力和交替共同注意力。

平行共同注意力:同时生成图像和问题注意,通过计算所有图像位置和问题位置中的图像、问题特征的相似度来连接图像和问题。

交替共同注意力:在图像和问题注意中交替生成,它根据问题关注图像,又根据图像特征再次捕捉问题。

从结构上看,在基于 Question 与 Image 的两个输入集合中,交替的方式更能节省计算资源,因为这两个集合本来就包含了逻辑关系。从广义上讲,交替协同注意力的方式虽然节省了计算资源,但相比平行结构,可能会丢失一些特征信息进而影响最终输出集合的精确度。

### 2.2.3. 单输出注意力和多输出注意力

单输出注意力:在每一个时间步长内,只表示单独一个特征。

多输出注意力：以多头注意力模型更为常见。通过连接输入序列线性投影到每个子空间后缩放的点积，模型不仅只关注一个部分的子空间信息，而是同时关注多个[23]。

相比较多头注意力，单输出注意力更为常见。但在某些情况下，一些下游任务无法仅仅依靠单一的注意力分布解决，在表达输入数据上，多输出通道要优于单输出通道。比如，吴军等人[24]用多头注意力机制预测磷酸化位点；赵丹丹等人[25]将多头注意力模型与字词融合进行中文命名实体识别；石磊等人[26]在文章中总结到，LSTM (Long Short-Term Memory)或 CNN 的解码和编码可被多头自注意力网络替代。

#### 2.2.4. 分层注意力

分层注意力模型是 Yang Zichao 等人[27]在 2016 年解决文档分类问题时提出的注意力模型，它具有层次结构，可用于解决多层次问题。故而主要运用于文档分类的任务中。在文档分类时，注意力权重由低级部分往高级部分学习，先用注意力机制提取词语，将词语整合到句子中，再将类似的句子整合到文档中。网络结构的分层，使得它的效率比其他的方法更快，也更加便于训练。

目前，分层网络也逐渐应用于情感分类[28]、顺序推荐系统[29]、机器翻译[30]、阅读理解[31]和动作识别[32]中。

### 2.3. 当前研究进展

除了上述提到的分类，还有一种自我注意力模型，它是由 Wang Bingning 等人[33]于 2016 年在研究“答案选择”时，针对“RNN 对词的选择有权重偏差问题”提出的。自我注意力模型不使用额外信息，仅仅关注自身，注意力机制得以完全应用也是基于此。

2017 年，Vaswani 等人[23]提出了简单网络架构 Transformer，它的结构也是由 Encoder-Decoder 组成，但它不需要使用 RNN 或卷积，而是依靠自注意力进行计算输入和输出。与之前相比，Transformer 允许并行计算，减少了训练的时间。在 Vaswani 等人的这篇文章中，实验证明了 Transformer 模型明显提高了机器翻译的速度，后续 Transformer 模型也被广泛用于多个任务中。后来在此基础上，又衍生出来了一系列变体，比如 Ji Jiayi 等人[34]在图像字幕上使用全局增强型的 Transformer (GET)，自适应地引导解码器生成高质量地字幕；Dehghani 等人[35]提出了 UT (Universal Transformer)模型，在标准 Transformer 的基础上，将递归神经网络应用在每个位置向量表示的连续更新过程中，解决了 Transformer 无法实现 while 循环的问题；2020 年，Dosovitskiy 等人[36]将标准 Transformer 应用在图像块序列的纯变换中，提出了 ViT (Vision Transformer)，该模型通过预训练就可以在图像分类任务中达到很好的效果，由此 Transformer 在计算机视觉领域有了新突破。

## 3. 医学应用

注意力机制作为一种新型的研究工具和手段，它的应用囊括了各个领域，如机器翻译、问答系统、文本分类、图像识别等。医学作为人类生活和发展不可或缺的领域，随时处于亟待突破的环境下，因此合理高效地运用注意力机制，利用它缩短医生处理问题的时间、提高医生的工作效率显得更为重要。目前，注意力机制已在医学图像检测与分割、医学图像分类、疾病预测、行为识别这几个方面得到较为广泛的应用。

### 3.1. 医学图像检测与分割

注意力模块并不是首个应用于目标检测与分割的手段。在深度学习的研究任务中，研究人员利用多种模型不断进行改进、融合，以试图不断提高精度。但在识别的过程中，由于面临图像背景复杂度过大、有过冗余杂图像，以及与周围边界模糊导致识别和分割难度大等问题，研究人员在之前的任务基础上引



入了注意力模块。在医学图像中引入注意力模块,自动关注感兴趣的部位,可以准确得到病灶的位置,辅助医生判断和审查病灶,目前已在肺结节[37]、肺部肿瘤[38]、结肠癌[39]、阴道炎致病微生物[40]的检测中得以应用。将该检测算法直接嵌入系统后,后续可实现批量与自动检测。

检测与分割往往是一体存在,但检测到的目标图像可能只是一个大的范围轮廓,且受周围环境因素的影响,无法将它很好地分割出来。有研究证明,在不同焦点下可以得到不同的分割结果[41],引入注意力模块,它会根据权重的分配将注意力更多地放在需要关注的目标特征上,强化所需要关注的目标区域,忽略冗余的区域,提高分割精度。将注意力模型应用于细胞分割上,解决了由于图像亮度不一致以及对比度低对模型的干扰问题,提高了模型的灵敏度[42];应用于肾脏肿瘤分割中,则能增强CT图像中肾脏肿瘤横断面最大直径的分割度,也提升了肾脏小肿瘤的检出率[43]。注意力机制的应用让医学图像在分割上取得了良好的识别精度。目前,结合多分支结构与自注意力机制的U-Net模型通过对来自肺部、眼球血管、肾部3个医学图像数据集上的医学影像进行仿真实验。表明了改进的U-Net模型在DICE相似性系数、准确性、特异度等多项评价指标上均高于已有医学图像分割方法[44]。

### 3.2. 医学图像分类

医学图像分类一直是亟待解决的问题。医学图像中的软体组织较低,且病灶区域与周围区域的界限不明显,如只依靠医生本人区分,不仅耗时耗力,也难以达到高效的工作要求。再加上样本之间区分度不大,而图像之间一小块区域模块的差异都会造成分类结果的不同,所以更需要对区域进行分类关注。引入注意力机制,利用深度网络在分类任务的学习中获得的自注意能力对特征图进行约束,强化重要信息,抑制不重要的信息,增加了模型的可解释性。张建鹏[45]将注意力机制与残差网络相结合,提出注意力残差学习机制,在此模块基础上构建注意力残差卷积神经网络,用大数据进行预训练再进行小参数微调后可对医学图像进行较好的分类,减少了训练时间。除此之外,注意力机制与残差网络的结合还可用于解决乳腺瘤图像分类[46],颅内动脉瘤分类[47]问题等。李宇[48]则选择在胶囊网络中加入注意力胶囊机制,该注意力层带有预测信息,协助增加关键权重比,提高目标分类的准确性。

与通用图像的分类任务不同,医学图像在收集训练数据的成本较大,Ma Congbo等人[49]使用交叉注意力网络,通过图像级注释来关注,再加入一个损失函数,帮助实现交叉注意过程,此方法在胸部疾病分类中实现了当时最先进的结果。

目前,注意力机制已经作为一种辅助手段,应用于多种医学图像的分类任务中。注意力机制不止作为一个模块与其他模型混合输入,也可以直接利用带CNN的VIT模型,通过模型的调参,使胸腔X线影像的分类相比现有的方法得到了进一步提升[50]。除此以外,注意力机制在血细胞疾病的分类[51]和骨肉瘤组织学图像的分类[52]等问题上也做出了贡献。

### 3.3. 疾病预测

目前,病人的患病情况、治疗信息等一系列相关资料都被储存在电子健康记录中,这些诊断信息由医学代码记录,通过这些代码进行数据分析从而进行疾病预测,能够推进医疗的服务质量和效率。但是,在代码分配过程中,会因为算法偏差[53]等原因导致错误分配,如果再加上患者的就医信息数据量大,预测效果便会大大下降。传统算法无法很好地解决这个问题,即使引入神经网络模型也无法达到很好的效果。葛伟[54]为了解决这个问题,将双向循环神经网络与注意力机制模型融合。交叉注意力能对历史疾病信息进行整合,根据患病记录里包含的有关疾病和治疗的信息权重分配,帮助完成疾病预测。在神经网络中引入注意力模块,能有效提高对重要特征的选择能力,在胸部常见疾病的预测[55]、房颤预测[56]等方面都有应用,包括磷酸化位点的预测[24],它的出现经常与疾病风险挂钩。

疾病预测不仅与空间图像有关，还与时间序列挂钩。同样地，注意力模型不仅能在图像上关注重点部位，也可用于时间序列上。王天罡[57]基于此，运用分层注意力机制分离时间和风险因素，使不同检查项享有单独的不定长时序输入序列，便可最大限度还原数据集本身的特征。

### 3.4. 行为识别

相比较静态识别，动态识别才是研究行为识别的重点。人一系列行为的产生是一串连贯的动作，将这些动作拆解成一帧帧图像送给神经网络去训练，根据关键动作进行捕捉最终完成行为识别。引入注意力机制，忽略全局信息，只关注局部，在捕捉特征上速度快、效率高。在医学上，具体以手语识别和情绪识别为典型。

#### 3.4.1. 手语识别

手语识别是连接聋哑人与他人沟通的重要桥梁。并非每个人都能理解聋哑人的行为，进行手语识别，将聋哑人进行的手语动作自动翻译出来，极大地解决了医生和聋哑人之间难以沟通的问题，提高了就诊效率。

但在手语识别的过程中，由于信息冗余、手部遮挡等其他因素的干扰，识别的精度并不高。引入注意力，将关注重点放在手部，降低其余影响因素的权重，强制关注空间区域，提高算法的速度和准确性。陆蒙和李荣辉[58]为了提高识别精度，将图卷积网络和多尺度双语识别网络融合，而多尺度双语网络主要由三个注意力模块组成，利用注意力网络能学习依赖关系的能力，在 CSL-500 的数据集上达到了 98.8% 的准确率，此研究充分证明了注意力机制的优越性。

除了对手部的直接捕捉，注意力机制还能将手语表示为一系列骨架姿势向量。Saunders 等人[59]基于此提出了骨架图自注意力，此研究在 RWTH-PHOENIX-Weather-2014T (PHOENIX14T)数据集上评估了骨架图自注意力架构，实现了最先进的反向翻译性能。

由于手语动作是一连串的视频帧，前后帧之间的依赖关系紧密，注意力模块还能加入到视频帧序列中，忽略最前和最后不重要的视频帧，重点关注中间部分。在用空间注意力网络选取感兴趣的领域后，再用时间注意力来选择重要的运动进行分类，提高手语识别能力。Huang Jie 等人[60]证明了该方法的可行性。

而由于需要考虑序列关系，注意力机制也通常与长短时记忆网络融合使用，经常应用于连续翻译。有研究设计在一方面用引入注意力模块增强视觉注意力，另一方面用长短时记忆网络提取句子嵌入，在这两方面的基础上提出空间和句子的一致性约束，这样的模型在 PHOENIX-2014、PHOENIX-2014-T 和 CSL 这三个基准测试中实现了最先进或具有竞争力的连续手语识别[61]。

#### 3.4.2. 情绪识别

情绪是人表达感知的一种行为，通过对情绪的识别来分析判断他人是否处于焦虑、抑郁、痛苦或其他负面情感状态中，能够帮助治疗师进行心理疏导。也可通过对产生快乐、兴奋、愉悦等其他正面情感状态的人分析其产生情绪的原因，帮助促进正面情感状态的发展。在情绪识别上，主要有四种方式：面部情绪识别、语音情绪识别、文本情绪识别、脑电波情绪识别。

面部情绪识别是视觉应用的一种，如今，识别算法已广泛用于各行各业。然而，在识别的过程中容易受到光照角度等环境因素的影响，再加上拍摄角度等问题，对于精准识别面部情绪增加了难度。利用注意力机制，学习不同人脸的权重，强化明确能表达情绪特征的部位，如眼睛、鼻子、眉毛、嘴巴[62]，能够提高识别的准确性。但在实际情况中，面部易受遮挡，识别准确率也会因此下降，使用带有注意力机制的卷积网络 ACNN，它不仅能专注于未遮挡区域，还能感知人脸的遮挡区域。实验证明，这种网络

不仅提高了对非遮挡区域的识别精度，同时也提高了遮挡区域的识别精度，具有先进性[7]。但由于面部受到遮挡的几率较高，目前也有仅针对特定区域的特征来识别用户情绪状态的研究，以眼动为主[63] [64] [65]。

在语音情绪的识别上，注意力主要与长短时记忆网络 LSTM 相结合使用。仅使用单层的 LSTM 进行识别时，泛化能力不高，但将 LSTM 得到的情感特征输入注意力层，计算每一帧语音信号的权重，根据权重分配的高低来确定情绪分类具有较好的识别能力[66]，后续相关研究人员还提出了将注意力机制与双向长短时记忆网络相结合的研究，进一步提高了准确率[67]。与语音情绪的识别相似，文本情绪的识别也主要与长短时记忆网络相结合使用。LSTM 擅长提取文本上下文的语义特征，注意力机制再针对文本提取关键情感特征，将两者进行融合以提升情感分类效果。

相较于面部或语音此类较依靠人类主观判断的情绪识别，脑电信号情绪识别则更为客观、精准，它是一种非线性、非平稳的随机信号。注意力模块的使用通常被用于分类，通过选取序列中最为重要的一个片段和 LSTM 的输出结果并连作为最终的输出向量。但由于它的信号很微弱，容易受到外界的干扰，加上脑电通道之间有一定的依赖联系，单纯依靠卷积神经网络很难捕捉，所以在识别的任务上效果不佳。有研究利用脑电信号的通道节点上不同的局部结构，选择用图数据表示。用图神经网络进行脑电情绪识别的方法性能已经超过了其他类型[68]。已此为研究基础，李景聪等人[69]提出了一种基于图注意力网络的脑电情绪识别方法，能学习各脑电通道之间的相关性[70]，捕捉不同对象的情绪脑电信号的时空域特征，实现跨被试情绪识别，在 SEED、SEEDIV 和 DREAMER 数据集上测试结果表明，MPGAT 已达到当前最先进模型的性能水平。

#### 4. 总结与展望

随着深度学习越来越多与医学相结合应用，极大促进了医疗服务行业的进步。注意力的引入推动了医疗行业的发展，它的作用就是为想要关注的部位分配更高的权重，忽略全局，只关注局部。进而缓解了医生的压力，也提高了就诊效率。再加上医学图像的数据集并不充足、细微的差距容易导致人为判断错误等原因，则更需要依靠注意力机制解决相关问题。

在此之前，注意力仅仅是单独作为一个模块输入，而如今作为一种应用技术，能够与多领域结合应用。如上文所说，注意力的手段多用于解决自然语言问题，考虑将在其他领域的注意力融合到医学上，有助于实现医学在技术手段方面的突破。例如，可以衍生到捕捉关键词后进行自动分类，生成病历本。根据病理关键词，生成一系列与其相对应的药物信息，必要时可根据更加详细的病理分析，附上以往的治疗手段作为参照，以及问答系统，网络作为更加方便的一种交流手段，需要准确地抓住病人想要询问的核心问题，自动捕捉用户自行检索的关键词，进行一系列比对分析，再提供相应的回答引导。如问答系统能够在心理辅导中得以有效应用，将大大提高诊疗效率，亦能推动医疗服务行业迈入新进程。其实，以上所说的实现路径已经随着人工智能的到来逐渐发展起来，但若在此基础上，引入注意力机制，利用它有效分配权重的能力进行问题处理，效率将会得到更大提高，也是医疗服务行业的一大跨越。

现如今，各行各业对虚拟现实技术的需求日益增强，虚拟现实技术是人机交互的一项巨大突破，它在不断高速发展的同时也越来越紧密地与医学领域相结合。加之当前疫情影响，虚拟现实技术可以隔绝人与人之间的直接接触，减少感染风险，可以预见，这种技术将以巨大的热浪席卷医学领域。在这个过程中，无论是辅助判断手术还是虚拟手术模拟，都需要对目标进行跟踪。而缺少了直接接触，阅片、审片和捕捉细节的难度也会因此提高，所以在这个环节里，凸显特征也就显得尤为重要。为想要关注的重点部分分配合适的权重，增强特征的空间位置，有助于为医护人员提供更好的服务。

医疗服务行业还有很长的一段路要走，尽可能地利用更新更快的手段的方法来减少医生的劳动量，也有利于进一步缩小我国和其他国家在医学领域上的差距。



## 参考文献

- [1] 毛健, 赵红东, 姚婧婧. 神经网络的发展及应用[J]. 电子设计工程, 2011, 19(24): 62-65.
- [2] 王海星, 田雪晴, 游茂, 等. 人工智能在医疗领域应用现状、问题及建议[J]. 卫生软科学, 2018, 32(5): 3-5+9.
- [3] 葛雪松. 影响注意力的十大因素[J]. 教学研究, 2006, 29(6): 555-557.
- [4] Mnih, V., Heess, N. and Graves, A. (2014) Recurrent Models of Visual Attention. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Volume 2, 2204-2212.
- [5] Rensink, R.A. (2000) The Dynamic Representation of Scenes. *Visual Cognition*, **7**, 17-42. <https://doi.org/10.1080/135062800394667>
- [6] Liu, G. and Guo, J. (2019) Bidirectional LSTM with Attention Mechanism and Convolutional Layer for Text Classification. *Neurocomputing*, **337**, 325-338. <https://doi.org/10.1016/j.neucom.2019.01.078>
- [7] Li, Y., Zeng, J., Shan, S., et al. (2019) Occlusion Aware Facial Expression Recognition Using CNN with Attention Mechanism. *IEEE Transactions on Image Processing*, **28**, 2439-2450. <https://doi.org/10.1109/TIP.2018.2886767>
- [8] Chen, Y., Liu, L., Phonevilay, V., et al. (2021) Image Super-Resolution Reconstruction Based on Feature Map Attention Mechanism. *Applied Intelligence*, **51**, 4367-4380. <https://doi.org/10.1007/s10489-020-02116-1>
- [9] Niu, Z., Zhong, G. and Yu, H. (2021) A Review on the Attention Mechanism of Deep Learning. *Neurocomputing*, **452**, 48-62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [10] Xu, K., Ba, J., Kiros, R., et al. (2015) Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. *Proceedings of the International Conference on Machine Learning*, Lille, 7-9 July 2015, 2048-2057.
- [11] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. *International Conference on Learning Representations 2015*, San Diego, 7-9 May 2015, 1-15.
- [12] Luong, M.-T., Pham, H. and Manning, C.D. (2015) Effective Approaches to Attention-Based Neural Machine Translation. <https://doi.org/10.18653/v1/D15-1166>
- [13] Chu, Y., Yue, X., Yu, L., et al. (2020) Automatic Image Captioning Based on ResNet50 and LSTM with Soft Attention. *Wireless Communications and Mobile Computing*, **2020**, Article ID: 8909458. <https://doi.org/10.1155/2020/8909458>
- [14] Wang, F., Jiang, M., Qian, C., et al. (2017) Residual Attention Network for Image Classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6450-6458. <https://doi.org/10.1109/CVPR.2017.683>
- [15] Shen, T., Zhou, T., Long, G., et al. (2018) Reinforced Self-Attention Network: A Hybrid of Hard and Soft Attention for Sequence Modeling. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 4345-4352. <https://doi.org/10.24963/ijcai.2018/604>
- [16] Lu, J., Yang, J., Batra, D., et al. (2016) Hierarchical Question-Image Co-Attention for Visual Question Answering. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 289-297.
- [17] Li, X., Song, K., Feng, S., et al. (2018) A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, October-November 2018, 4752-4757. <https://doi.org/10.18653/v1/D18-1506>
- [18] Gao, J., Wang, X., Wang, Y., et al. (2019) CAMP: Co-Attention Memory Networks for Diagnosis Prediction in Healthcare. *Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM)*, Beijing, 8-11 November 2019, 1036-1041. <https://doi.org/10.1109/ICDM.2019.00120>
- [19] Gong, X., Liu, X., Li, Y., et al. (2020) A Novel Co-Attention Computation Block for Deep Learning Based Image Co-Segmentation. *Image and Vision Computing*, **101**, Article ID: 103973. <https://doi.org/10.1016/j.imavis.2020.103973>
- [20] Yu, Z., Yu, J., Cui, Y., et al. (2019) Deep Modular Co-Attention Networks for Visual Question Answering. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 6274-6283. <https://doi.org/10.1109/CVPR.2019.00644>
- [21] Liu, Y., Zhang, X., Zhang, Q., et al. (2021) Dual Self-Attention with Co-Attention Networks for Visual Question Answering. *Pattern Recognition*, **117**, Article ID: 107956. <https://doi.org/10.1016/j.patcog.2021.107956>
- [22] Zhang, S., Chen, M., Chen, J., et al. (2021) Multimodal Feature-Wise Co-Attention Method for Visual Question Answering. *Information Fusion*, **73**, 1-10. <https://doi.org/10.1016/j.inffus.2021.02.022>
- [23] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 1-11.



- [24] 吴军, 欧阳艾嘉, 张琳. 基于多头注意力机制的磷酸化位点预测模型[J]. 广西师范大学学报(自然科学版), 2022, 40(3): 161-171.
- [25] 赵丹丹, 黄德根, 孟佳娜, 等. 多头注意力与字词融合的中文命名实体识别[J]. 计算机工程与应用, 2022, 58(7): 142-149.
- [26] 石磊, 王毅, 成颖, 等. 自然语言处理中的注意力机制研究综述[J]. 数据分析与知识发现, 2020, 4(5): 1-14.
- [27] Yang, Z., Yang, D., Dyer, C., et al. (2016) Hierarchical Attention Networks for Document Classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, June 2016, 1480-1489. <https://doi.org/10.18653/v1/N16-1174>
- [28] Cheng, J., Zhao, S., Zhang, J., et al. (2017) Aspect-Level Sentiment Classification with Heat (Hierarchical Attention) Network. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore, 6-10 November 2017, 97-106. <https://doi.org/10.1145/3132847.3133037>
- [29] Ying, H., Zhuang, F., Zhang, F., et al. (2018) Sequential Recommender System Based on Hierarchical Attention Network. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 3926-3932.
- [30] Miculicich, L., Ram, D., Pappas, N., et al. (2018) Document-Level Neural Machine Translation with Hierarchical Attention Networks. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, October-November 2018, 2947-2954. <https://doi.org/10.18653/v1/D18-1325>
- [31] Wang, W., Yan, M. and Wu, C. (2018) Multi-Granularity Hierarchical Attention Fusion Networks for Reading Comprehension and Question Answering. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 1705-1714. <https://doi.org/10.18653/v1/P18-1158>
- [32] Wang, Y., Wang, S., Tang, J., et al. (2016) Hierarchical Attention Network for Action Recognition in Videos. *The International Conference on Computer Vision (ICCV)*, 61-70. <https://arxiv.org/abs/1607.06416>
- [33] Wang, B., Liu, K. and Zhao, J. (2016) Inner Attention Based Recurrent Neural Networks for Answer Selection. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 1288-1297. <https://doi.org/10.18653/v1/P16-1122>
- [34] Ji, J., Luo, Y., Sun, X., et al. (2021) Improving Image Captioning by Leveraging Intra- and Inter-Layer Global Representation in Transformer Network. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 1655-1663. <https://doi.org/10.1609/aaai.v35i2.16258>
- [35] Mirsamadi, S., Barsoum, E. and Zhang, C. (2017) Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention. *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, 5-9 March 2017, 2227-2231. <https://doi.org/10.1109/ICASSP.2017.7952552>
- [36] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2010) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [37] 金亚荣. 基于深度学习的 CT 图像病灶检测和识别[D]: [硕士学位论文]. 北京: 北京工业大学, 2020.
- [38] 周涛, 董雅丽, 刘珊, 等. 用于肺部肿瘤图像分割的跨模态多编码混合注意力 U-Net[J]. 光子学报, 2022, 51(4): 368-384.
- [39] 梁美彦, 张倩楠, 任竹云, 等. 基于注意力机制的结肠癌病理学图像识别研究[J]. 测试技术学报, 2022, 36(2): 93-100.
- [40] 郝如茜, 王祥舟, 张静, 等. 采用注意力机制的显微图像智能检测方法[J]. 光电工程, 2022, 49(3): 44-54.
- [41] 侯海燕, 符志鹏. 视觉注意力焦点显著性模型及其医学应用[J]. 计算机工程与应用, 2017, 53(10): 213-217.
- [42] 张文秀, 朱振才, 张永合, 等. 基于残差块和注意力机制的细胞图像分割方法[J]. 光学学报, 2020, 40(17): 70-77.
- [43] 吴倩倩, 周蕾蕾, 陆小妍, 等. 基于多头自注意力机制与 U-Net 的增强 CT 图像肾脏小肿瘤自动分割研究[J]. 中国医学装备, 2022, 19(2): 27-31.
- [44] 闫超, 孙占全, 田恩刚, 等. 结合多分支结构与自注意力机制的医学图像分割模型及其应用[J]. 小型微型计算机系统, 2021, 42(11): 2375-2381.
- [45] 张建鹏. 基于深度学习的医学图像分类技术研究[D]: [硕士学位论文]. 西安: 西北工业大学, 2019.
- [46] Zou, Y., Zhang, J., Huang, S., et al. (2022) Breast Cancer Histopathological Image Classification Using Attention High-Order Deep Network. *International Journal of Imaging Systems and Technology*, **32**, 266-279. <https://doi.org/10.1002/ima.22628>
- [47] Zhao, Y., Wang, S., Ren, Y., et al. (2022) CRANet: A Comprehensive Residual Attention Network for Intracranial

- Aneurysm Image Classification. *BMC Bioinformatics*, **23**, Article No. 322. <https://doi.org/10.1186/s12859-022-04872-y>
- [48] 李宇. 基于胶囊网络的脑肿瘤图像分类研究与实现[D]: [硕士学位论文]. 成都: 电子科技大学, 2021.
- [49] Ma, C., Wang, H. and Hoi, S.C. (2019) Multi-Label Thoracic Disease Image Classification with Cross-Attention Networks. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Shenzhen, 13-17 October 2019, 730-738. [https://doi.org/10.1007/978-3-030-32226-7\\_81](https://doi.org/10.1007/978-3-030-32226-7_81)
- [50] 邢素霞, 鞠子涵, 刘子娇, 等. 视觉 Transformer 预训练模型的胸腔 X 线影像多标签分类[J/OL]. 中国图象图形学报, 2022. [http://www.cjig.cn/jig/ch/reader/view\\_abstract.aspx?flag=2&file\\_no=202203170000002&journal\\_id=jig](http://www.cjig.cn/jig/ch/reader/view_abstract.aspx?flag=2&file_no=202203170000002&journal_id=jig)
- [51] Sadafi, A., Makhro, A., Bogdanova, A., et al. (2020) Attention Based Multiple Instance Learning for Classification of Blood Cell Disorders. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Lima, 4-8 October 2020, 246-256. [https://doi.org/10.1007/978-3-030-59722-1\\_24](https://doi.org/10.1007/978-3-030-59722-1_24)
- [52] Pan, L., Wang, H., Wang, L., et al. (2022) Noise-Reducing Attention Cross Fusion Learning Transformer for Histological Image Classification of Osteosarcoma. *Biomedical Signal Processing and Control*, **77**, Article ID: 103824. <https://doi.org/10.1016/j.bspc.2022.103824>
- [53] Hajian, S., Bonchi, F. and Castillo, C. (2016) Algorithmic Bias. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016.
- [54] 葛伟. 基于注意力机制的疾病预测关键技术研究[D]: [硕士学位论文]. 济南: 山东大学, 2020.
- [55] 王江峰, 刘利军, 黄青松, 等. 基于胸部 X 射线影像的常见疾病预测方法研究[J]. 激光与光电子学进展, 2022, 59(18): 387-394.
- [56] 吴石远. 基于卷积和循环神经网络的房颤预测模型的研究[D]: [硕士学位论文]. 武汉: 武汉科技大学, 2021.
- [57] 王天罡. 基于深度学习的疾病预警及诊断[D]: [硕士学位论文]. 西安: 西安工程大学, 2021.
- [58] Meng, L. and Li, R. (2021) An Attention-Enhanced Multi-Scale and Dual Sign Language Recognition Network Based on a Graph Convolution Network. *Sensors*, **21**, 1120. <https://doi.org/10.3390/s21041120>
- [59] Saunders, B., Camgoz, N.C. and Bowden, R. (2021) Skeletal Graph Self-Attention: Embedding a Skeleton Inductive Bias into Sign Language Production.
- [60] Huang, J., Zhou, W., Li, H., et al. (2019) Attention-Based 3D-CNNs for Large-Vocabulary Sign Language Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, **29**, 2822-2832. <https://doi.org/10.1109/TCSVT.2018.2870740>
- [61] Zuo, R. and Mak, B. (2022) C<sup>2</sup>SLR: Consistency-Enhanced Continuous Sign Language Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 5121-5130. <https://doi.org/10.1109/CVPR52688.2022.00507>
- [62] 李广鹏, 刘波, 李坤, 等. 一种基于机器学习的人脸情绪识别方法研究[J]. 计算机技术与发展, 2019, 29(5): 27-31.
- [63] 李小伟. 脑电, 眼动信息与学习注意力及抑郁的中文相关性研究[D]: [博士学位论文]. 兰州: 兰州大学, 2015.
- [64] 马伟娜, 朱蓓蓓, 谢宇. 孤独症儿童面部表情识别能力的眼动研究[J]. 应用心理学, 2015, 21(1): 76-88.
- [65] 李学聪, 战荫伟, 杨卓, 等. 基于标签分布学习的眼部情绪识别[J]. 计算机科学与应用, 2022, 12(4): 1213-1225. <https://doi.org/10.12677/CSA.2022.124123>
- [66] 刘天宝, 张凌涛, 于文涛, 等. 基于嵌入注意力机制层级 LSTM 的音视频情感识别[J]. 激光与光电子学进展, 2021, 58(2): 175-182.
- [67] 姜特, 陈志刚, 万永菁. 基于注意力机制的多任务 3D CNN-BLSTM 情感语音识别[J]. 华东理工大学学报(自然科学版), 2022, 48(4): 534-542.
- [68] Wu, Z., Pan, S., Chen, F., et al. (2020) A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, **32**, 4-24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [69] 李景聪, 潘伟健, 林镇远, 等. 采用多路图注意力网络的情绪脑电信号识别方法[J]. 智能系统学报, 2022, 17(3): 531-539.
- [70] Wan, S., Pan, S., Yang, J., et al. (2021) Contrastive and Generative Graph Convolutional Networks for Graph-Based Semi-Supervised Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2-9 February 2021, 10049-10057. <https://doi.org/10.1609/aaai.v35i11.17206>