

关于神经网络与深度学习领域论文的阅读摘要

蒋浩天 1120231337

2024 年 5 月 24 日

摘要

本文讨论了在阅读近期一些神经网络与深度学习及其相关领域中发表的研究型学术论文后, 对于相关成果的概括, 总结, 对于相关领域发展方向的观点, 以及对于未来研究方向的分析. 在神经网络与深度学习领域, 近期的研究集中于:

- 利用神经网络对传统计算手段中难以量化的复杂信息进行处理与特征提取, 进而帮助智能设备更好获取环境信息, 从而强化人工智能服务人类需求的能力;
- 利用 FPGA, GPU 等设备在并行计算, 浮点运算上的强大计算能力, 加速复杂的神经网络计算, 实现及时化, 小型化神经网络计算;
- 利用神经网络的参数训练方法模拟人类经验判断, 使其能在一些强经验领域对人类工作者进行辅助判断, 以减轻特定工作者负担.

1 背景

近年来, 随着人工智能的大热, 大量学者开始研究如何利用卷积神经网络 (CNN, Convolution Neural Network), 递归神经网络 (RNN, Recurrent Neural Network), 生成式对抗网络 (GAN, Generative Adversarial Network) 等方法, 更好地解决传统上计算机难以处理的问题, 例如对高复杂度信息进行量化, 对图像与视频等视域空间信息进行处理, 并取得了相当突出的成果.

作为新一代计算机专业学生, 通过阅读文献, 可以了解时下技术热点, 进而通过自身理解把我技术发展方向, 有助于未来研究领域上的个人发展.

阅读的文献有:

- 1) 基于可变形卷积网络的视频去模糊算法^[1]
- 2) 基于移动终端的人体姿态检测技术研究 with 实现^[2]
- 3) 基于轻量化卷积神经网络的表情识别与可视化分析^[3]
- 4) 基于 FPGA 和 Lenet-5 的字符识别系统设计与实现^[4]
- 5) 基于机器学习的低剂量 CT 图像降噪方法研究^[5]
- 6) 基于 FPGA 的神经网络加速器^[6]
- 7) 基于卷积神经网络的多模光纤图像重构研究^[7]
- 8) 毫米波雷达手势识别轻量化网络研究^[8]

9) 基于神经网络的声音事件检测系统研究^[9]

10) 基于时空关系图网络的视频动作识别研究^[10]

2 成果概括

基于可变形卷积网络的视频去模糊算法^[1] 作者介绍了图像视频处理领域常用的 Transformer 神经网络及其改进 Uformer, 可变滤波器, STFAN (Spatio-Temporal Filter Adaptive Network)^[11], EDVR (Video Restoration framework with Enhanced Deformable convolutions)^[12] 等技术, 并指出其不足之处与可改进方向.

作者介绍了可变卷积神经网络 (Variable CNN), 说明了对其改进方案, 指出了在实验中发现的注意区域的增大, 可用采样位置的增多等成效. 作者介绍了加入 DFnet 层后的动态滤波器网络, 指出利用可变滤波器的自适应特性可以很好的解决输入中噪声的干扰. 作者指出, 可变卷积神经网络通过相邻帧中清晰像素分析, 可以在视频的视域空间中获取更多有效信息进行去模糊处理.

作者指出, 在预处理阶段, 输入视频需要进行预处理. 传统的光流法无法达到效果与性能开销之间的平衡, 作者采用多维度可变卷积神经网络进行视频隐藏特征信息对齐, 达到了比较好的效果与较小的计算开销. 在特征融合与重建方面, 作者采用了频道维度的时序特征融合与反卷积层上采样的特征重建, 将其整合进入神经网络中.

作者提出了一种自适应时空卷积神经网络, 采用清晰像素与模糊像素融合的方式来挖掘视频中的时空信息, 利用动态滤波网络来进行预处理以提高效率质量. 其采取的流程如下:

特征提取 → VCNN 特征对齐 → 时序特征融合 → 特征重建

基于移动终端的人体姿态检测技术研究与应用^[2] 作者介绍了在人体姿态检测领域现阶段研究中常用的计算机视觉与传感器技术, 简要介绍了 CNN, CRNN (卷积递归神经网络, Convolution Recurrent Neuro Network), KNN (K-邻近) 等技术与当前应用.

作者提出了在解决该类问题时, 利用多通道卷积神经网络 (Multi-channel CNN), 卷积神经网络与长短期记忆递归神经网络结合 (CNN-LSTM) 可取得的效果, 并提出了使用结合了卷积头与 LSTM 网络的 ConvLSTM 网络来解决人体姿态传感问题.

作者总结表示, 现有应用中通常使用智能手机的加速度传感器与陀螺仪获取数据. 在基于机器学习的应用中, 常使用 SVM (支持向量机, Support Vector Machine), 简单贝叶斯层 (Naive Bayes Layer), 决策树 (Decision Tree) 来进行处理. 在基于深度学习的应用中, 多使用 CNN, RNN, LSTM 来处理.

作者说明, 对于实验中的数据集, 采用 Android 应用访问设备传感器数据进行记录, 并通过采样与滤波器进行预处理. 作者说明了利用 Keras 框架进行一维卷积神经网络的构建, 并通过超参数调整的手段进行网络调优. 作者说明了多通道卷积神经网络分割-合并的工作方式.

作者分析了循环神经网络 (RNN) 的缺点, 指出其准确性低, 梯度爆炸, 梯度消失的问题; 分析了长短期记忆递归神经网络 (LSTM) 的缺点, 指出其较低准确性, 更少的层数需要更多参数等问题.

作者说明了利用卷积层作为前端, LSTM 作为分支预测分类的一维 CNN-LSTM 混合网络, 提出了 ConvLSTM, 表示这一网络同时利用了 LSTM 在时间领域上的较强能力和 CNN 在局部空间上的能力, 在实验中有较好效果.

基于轻量化卷积神经网络的表情识别与可视化分析^[3] 作者介绍了现今研究中常用的技术, 如 CNN, DBN (深度置信网络, Deep Belief Network), DAE (深度自动编码器, Deep Autoencoder), RNN, GAN;

概括了之前研究中提出的模型与网络,如基于 SVM 的 L2-SVM,有更深层数更宽宽度的层关联 CNN,学习能力强于微调的 FaceNet2ExpNet, SBN-CNN.

作者介绍了人脸表情识别的要求,包括预处理中光照条件与人脸位置的归一化,人脸检测,人脸特征提取分类,列举了人脸对齐时常用的模型,如 MTCNN, CenterFace, RetinaFace.

作者介绍了 ResNet,说明其深度卷积和逐点卷积能力,指出模型轻量化的方法有下采样,减小卷积核尺寸,全局池化,减少参数数量,减少计算量.作者介绍了有意义扰动 (Meaningful Perturbation),指出通过 Adam 优化后可以使网络更简单,更高效,减少了网络对超参数调整的依赖.

作者介绍了 Score-CAM,指出该网络具有前向传播的特性,噪声干扰更轻,能在计算量减少的同时保持足够健壮性.

基于 FPGA 和 Lenet-5 的字符识别系统设计与实现^[4] 作者介绍了现今领域内常用的图像预处理算法,包括图像滤波器,图像增强,字符分割.作者介绍了在 FPGA 上运行 CNN 的方法,指出由 HLS 翻译为 HDL (硬件设计语言, Hardware Designing Language) 所带来的代码冗余与浪费, PS + PL 方式的不便,和采用 HDL 直接设计的原因.作者指出,FGPA 运行 CNN 的主要限制是板上 RAM 的缺乏,DDR RAM 读写的不便与缓慢.

作者介绍了常用的几种滤波方式,指出均值滤波会带来的平滑模糊图像,而中值滤波,高斯滤波,小波变换滤波表现更好.在图像提升方面,作者指出直方图均衡算法 (HE) 运行速度快,实现原理简单,但存在图像灰度级均衡化后真题图像亮度不均匀,细节信息丢失等缺点.

作者介绍了字符矫正中所需要的旋转矫正,倾斜矫正,切向矫正,介绍了字符分割中的二值化,投影法等方法.作者说明了字符识别网络 Lenet-5 中卷积工作原理,对单个字符全连接神经元识别,介绍了其层级结构:

卷积层 → 池化层 → 卷积层 → 池化层 → 全连接层

作者介绍了实验中的具体实现方式,介绍了 FPGA 的板上 RAM 大小与寄存器数量的选择原因和效果,说明了实现内存地址生成器对于‘移窗’式计算的实用意义,解释了在储存和流水线设计方面效率和速度上的优势,介绍了为部件间交流问题而设计的信号与使能器系统.在实验中,作者分析得出了不同光照条件下图像预处理的不同方式,对于流水线设计所带来的并行计算优势进行了定量测试.

基于机器学习的低剂量 CT 图像降噪方法研究^[5] 作者介绍了现阶段在 CT 图像分析领域常使用的计算机技术与算法,介绍了 sin 函数滤波器,迭代重构,深度学习后处理等方法;分析说明了深度学习所存在的一些问题,例如黑箱结构难以预测,图像处理后会有一部分关键信息的损失,训练用数据集受限制,无法与后续医疗诊断进行融合.

作者介绍了使用编码器-解码器 (Encoder-Decoder) 结构的 CNN 网络与对抗性网络.在噪声学习方面,作者分析了使用可变自动滤波器 (Variational Autoencoder, VAE),生成性对抗网络 (GAN) 来进行噪声提取并进行噪声学习的方法.

在任务关联的降噪方面,作者提出了一种即插即用 (Plug 'n Play, PnP) 的框架,称为 LIDnet (病灶启发的降噪网络, Lesion-Inspired Denoising Network),将降噪任务与下游检测任务结合起来,进行协同训练.

作者提出了一种由放射科医生诊断行为启发的模型 (RIDnet, Radiologist-Inspired Deep Neural Network),通过构建 3D-GCN (3-Dimensional Graph Convolutional Network, 三维图卷积网络) 通过非局部图像获取非局部信息,利用深度图卷积模块,局部信息提取层,特征融合层的组合,构建整体模型.

作者表示, 这套模型在系统性的双盲实验中表现出了与商业迭代重建方案由竞争力的将扫性能, 并在放射科医生的评价中获得了较高成绩, 被认为有利于医生发现诊断一些无法观察到的病灶, 在临床上有巨大的潜力.

基于 FPGA 的神经网络加速器^[6] 作者介绍了在神经网络快速发展的当下, 网络的参数数量以及其需要的计算量激增, 而 FPGA 调整灵活计算高效的特性, 使其成为低成本神经网络加速器的一大选择. 作者指出, FPGA 上的神经网络加速器普遍无法兼容异构平台, 也无法对不同类型的网络进行适应.

作者简要介绍了神经网络的类型与现状, 介绍了 CNN 从 ReLU 到 ResNet 的前期发展, 与 Transformer^[13] 出现后的巨变; 介绍了 RNN 的发展过程, LSTM 网络的实现结构. 作者介绍了 CNN 的加速方式, 包括参数量化, 剪枝, 搜索, 深度可分离卷积, Winograd 变换, 脉动计算, 分块计算; 介绍了 LSTM 的加速方式, 包括量化, 剪枝, 并行计算.

作者简单介绍了神经网络的工作原理, 对于 CNN, 说明了其层级组成, 包括卷积层, 激活函数层, 池化层, 全连接层; 对于 RNN, 说明了其对内存的依赖, 与其遗传式的工作方式. 作者说明了定点数量化的原理, 指出定点数量化可以减少 DSP (Digital Signal Processor, 数字信号处理器) 的性能占用, 也更有利于 ReLU 激活函数工作.

作者介绍了选用的 FPGA 开发平台, 简单说明了其构成与各项性能. 作者解释了 FPGA 开发的不同路线, 说明了 HLS (High Level Synthesis, 高层次综合) 技术路线存在的效率问题与移植性弱点, 说明了选用 HDL (Hardware Designing Language) 的具体原因.

作者提出了具体优化策略, 在数据传输上, 作者采用 FPGA 开发平台提供的 AXI 总线与 AXI DMA 控制器; 在命令控制方面, 作者指出考虑到灵活性, 采用软件控制命令而非 HDL 设计逻辑, 同时也利用 AXI-lite 总线协议实现了对寄存器的轻量级控制. 在计算方面, 作者实现了循环展开与循环并行处理, 实现了卷积核心并行处理和高维处理, 通过行缓冲提高了处理效率; 在数据分块储存上, 作者实现了由 SD 卡, DDR RAM, 板上 RAM 组成, 通过系统总线连接的的三级储存层级; 在流水线方面, 作者通过双缓冲区实现了乒乓式缓冲区 (Ping-pong Buffers), 将读写需求分离以提高效率, 通过常数优化提高了量化模块运行效率, 通过行缓冲提和可配置偏移寄存器提高了池化时的计算效率.

基于卷积神经网络的多模光纤图像重构研究^[7] 作者简要介绍了多模光纤成像散斑的重建所存在的难点与问题, 介绍了一种利用迁移学习的卷积神经网络, 实现了降低网络复杂度的同时获得图像的高质量重构.

作者介绍了多模光纤的优势以及其工作原理, 说明了其散斑成像的形成原因. 作者简单介绍了现今领域内的散斑图像重建方法, 主要由模拟相位共轭技术, 数字相位共轭技术, 传输矩阵技术. 作者简要介绍了目前研究人员在神经网络重建散斑图像上取得的一些成果和不足处, 如利用 VGG 网络的重建模型对图像光源波长变化的健壮性不足; 需要经过一定调整的训练方式才能避免受到各种扰动的干扰; 利用将二维空间信息转换为一维时间脉冲流进而重建的方式对光纤长度有一定要求.

作者简要介绍了阶跃多模光纤与渐变多模光纤的工作原理, 说明了使用 Python 语言的自动采集程序的实现方式. 作者说明了使用 Python 语言编写散斑图像的批量裁剪程序, 对图像进行裁剪, 尺寸调整, 通道转换, 并用于实验图像的预处理.

作者简要介绍了 LeNet-5 网络模型的工作原理; 介绍了 AlexNet 在图像分类领域的成就与改进, 介绍了其对 ReLU 激活函数的应用, 其实现多 GPU 并行加速计算, 重叠池化 (Overlapping Pooling), 随机失活机制; 介绍了 VGG 网络对于网络深度对准确率影响的研究; 介绍了 ResNet 对于防止网络退化的成果; 介绍了卷积神经网络的基本工作原理; 介绍了激活层的工作原理, 说明了 ReLU 激活函数在防止梯度

消失和提升计算性能方面的优势;介绍了池化层的工作原理。

作者实现了一种基于 VGG-16 散斑重构网络的模型,利用 UpSampling2D 上采样层实现更改输出图尺寸大小,使得网络模型更灵活;通过调整重构训练轮数,实现了训练时间,重构效果的较好平衡;作者实现了一种基于改进的 ResNet 的散斑重构模型,使用更少的参数数量达到了相近的重建效果;作者实现了一种基于 U-net 的散斑重构模型,利用 Dropout 机制避免过拟合后,用较少参数数量和较短训练时间达到了相当的重建效果。

作者提出,利用迁移学习方法,可以将具有泛化能力的模型迁移到目标网络上,从而减少数据集数据量,便于模型的训练。作者实现了在 VGG, ResNet, U-net 网络上运用迁移网络进行微调训练,减少了训练需要的数据量与时间。

毫米波雷达手势识别轻量化网络研究^[8] 作者介绍了在毫米波雷达得到广泛利用的当下,使用毫米波雷达对手势进行识别的一种神经网络手段。作者简单介绍了现今利用视觉和可穿戴设备进行手势识别的技术路线,并指出其不足之处:基于视觉方案存在泄露使用者隐私的问题;基于可穿戴设备方案存在成本与便捷性问题。作者简单介绍了现有方案,如基于传感器的 CyberGlove, MYO 系统;基于视觉的 Kinect 方案;基于雷达的 Wisee 系统, HMM (Hidden Markov Model) 方案, QEA 方案。

作者介绍了毫米波雷达成像的物理原理,简要说明了基于二维快速傅里叶变化 (2D-FFT, 2-Dimensional Fast Fourier Transform) 的多普勒估计,基于多重信号分类 (MUSIC, Multiple Signal Classification) 的角度估计的基本原理。

作者介绍了其对于手势识别的手势设计,讨论了对于杂波的抑制处理方案,采用反馈是四阶对消器,以获得更好的频率特性;作者讨论了手势数据处理和距离信息提取的方式,指出通过脉冲方式获取数据,再进行快时间 FFT,获取每个脉冲对应的距离信息,以慢时间拼接后得到完整的手势距离变化信息。作者讨论了通过 STFT 操作实现时频分析以进行手部多普勒信息的提取。作者提出,使用多通道激光雷达回波数据进行拼接,利用 MUSIC 算法获得角度分布,继而获取手势的角度信息。

作者介绍了基于特征融合的卷积神经网络,在特征提取模块借鉴了 VGG 网络结构,使用更多的小型卷积核替代,在获得同样大小感受野的前提下减少了参数量,同时增强了网络的非线性表达能力。作者介绍了模型压缩的各种算法,如稀疏训练,模型剪枝,知识蒸馏。

作者提出了一种基于模型压缩的深度手势识别网络,通过粗略剪枝和精细剪枝,将模型的参数量降低,可通过微调实现在更少的参数与更少的迭代次数下达到足够的准确度。作者实验了知识蒸馏在该问题上的效果,发现联合剪枝和知识蒸馏的模型,在大幅减少计算量和参数量的同时得到了轻微的准确率提高。

作者介绍了轻量型卷积神经网络的实现方式,包括深度可分离卷积,轻量型神经网络,包括 MobileNet, ShuffleNet, SqueezeNet。作者提出了可瘦身机制这种具有灵活性的神经网络设计方式,介绍了其可切换的批量归一化层设计细节,基于此实现了轻量型手势识别网络。作者介绍,手势识别网络分为特征提取网络与预测输出模块,特征提取模块的主干结构可采用 MobileNetV2,预测输出模块采用两层全连接层和一层 softmax 层。

基于神经网络的声音事件检测系统研究^[9] 作者介绍了现今声音事件领域的常用技术,包括传统的 GMM-HMM (高斯混合模型-隐马尔可夫模型) 模型,基于神经网络的 SED 算法,基于 DNN (Deep Neural Network, 深度神经网络) 的深度学习算法。作者指出,现有声音事件检测系统普遍存在计算量大,识别不准确的问题。

作者介绍了现有的常用神经网络模型训练策略,包括网络继承,迁移学习,注意力机制和可分离卷积,

介绍了一种基于简单对数幂短时傅里叶变换频谱图的将时间注意力和通道注意力结合起来的 CNN 算法,通过互补信息来增强 CNN 的代表性.

作者介绍了一种新的深度声音识别的学习方法,BC 学习 (班间学习),介绍了一种利用两个 LSTM (长短期记忆网络) 的模型,介绍了一种基于 GAN (生成对抗网络) 的训练音频增强方式,介绍了使用分布式特征使联邦学习 (FL) 的方法来缓解隐私问题. 作者指出,这些方法虽然准确性较高,训练难度有所下降,但是计算量仍然较高,难以在物联网领域应用.

作者介绍了声音事件检测算法的基本构成,包括音频输入,声音特征提取,神经网络,系统输出. 作者分析了声音特征提取时原始音频采样率过高,数据量过大的问题,提出使用 STFT 声谱图作为神经网络的输入,进行小波变换后提取信号特征,对不同频率信号进行聚焦,使用梅尔频谱特征和梅尔倒谱系数分析音频对人耳的实际特征.

作者介绍了传统的 GMM-HMM 模型,现今基于 CNN 的 SED 算法 LSED,选择性可分离卷积机制的设计,协调注意力机制的设计. 作者实验了在 FPGA 的 DPU (深度学习处理单元) 上进行声音事件检测系统的实现,实现了在 FPGA 上运行 LSED 算法,以极少的参数量和较少的运算量达到了足够的准确度.

基于时空关系图网络的视频动作识别研究^[10] 作者介绍了视频理解的子任务分割,包括动作识别,动作检测和定位,动作预测,视频描述,视频问答. 作者介绍了目前基于 RGB 帧的有监督动作识别方案中遇到的提取帧间空间信息的难题,指出了提升动作识别模型性能的关键.

作者介绍了传统上的光流法密集轨迹视频动作识别法, SVM 分类器法, SFV 法,深度学习上的并行双 CNN 支路双流网络模型,采用 3D 卷积的 I3D, R3D, Slowfast 等模型,基于长距离关系推理的 Non-local 网络. 作者指出,目前的问题在于模型的轻量化与长距离建模能力的进一步提高.

作者介绍了基于 Transformer^[13] 的自注意力模型,包括 STN 方法, SENet 方法, CBAM 方法, Highway, SKNet 等网络,表明了基于注意力的方法的巨大潜力.

作者介绍了基于 2D 双流卷积网络的 TSN 模型, TSM 模型,指出了其在长距离帧间关系提取能力上的不足. 作者介绍了基于 3D 卷积网络的 I3D 模型,指出其通过迁移学习手段使得训练难度降低,收敛速度加快.

作者分析了在视频序列中构建图的方式,分为时间关系图和时空关系图两种,作者介绍了通过线性变换,嵌入式点积的方法构建时间关系图的方法,说明了图卷积网络长句关系推理的具体原理.

作者介绍了基于时空注意力机制的关键特征选择,分析了时间注意力模块,时空注意力模块的具体设计原理与方式. 作者介绍了基于金字塔时空图网络的视频动作识别,提出利用图像分割网络 Deeplab V3 和目标检测网络 Trident Networks 中使用并行的空洞卷积来控制感受野,将不用感受野得到的特征进行融合的方式来提升多尺度网络的性能. 作者具体说明了时空金字塔和时空金字塔网络结构的原理和实现方式,实现了整套视频动作识别模型框架 STPG-Net 模型,在不同数据集上都表现出了良好的性能与较高的准确性.

3 成果评价

基于可变形卷积网络的视频去模糊算法^[1] 作者的成果展示了可变形卷积网络在视频处理领域的巨大潜力,其成果在挖掘视频帧间潜在信息进行去模糊上取得的成果显著;同时,其提出的自适应时空卷积神经网络中将不同网络得到的结果进行特征融合的方式创新型强,有拓展至图象视频处理领域其他问题的巨大潜力.

基于移动终端的人体姿态检测技术研究实现^[2] 作者的成果展示了利用智能手机传感器数据和轻量化神经网络结合进行本地计算检测人体姿态的方案, 结合了智能设备的集成特性和神经网络轻量化技术, 在充分保护用户隐私的同时也做到了兼顾便捷性和性能. 作者的这套解决方案, 可以通过读取不同的传感器进行针对性训练, 实现对多种信息的移动端本地采集分析, 例如运动数据, 睡眠数据, 光照情况数据, 契合了现今信息科学进一步的数字化智能化的发展趋势.

基于轻量化卷积神经网络的表情识别与可视化分析^[3] 作者的成果创新型使用了 ResNet 与有意义扰动, Adam 优化的结合, 赋予轻量型神经网络前向传播特性, 使得网络更简单高效, 降低了网络对超参数调整的依赖, 降低了噪声对网络的干扰, 在保证健壮性的前提下大幅度减小了网络计算量. 作者的成果对于类似领域的神经网络模型小型化工作具有较强借鉴意义.

基于 FPGA 和 Lenet-5 的字符识别系统设计实现^[4] 作者通过在 FPGA 上设计完整的计算流水线, 实现了使用 FPGA 上设计的并行神经网络计算单元对卷积神经网络 LeNet-5 进行并行加速, 创新性设计了内存地址生成器这一结构以对卷积计算中的移窗操作进行优化加速. 作者的成果对于未来将其他卷积神经网络迁移至 FPGA 平台进行并行加速优化的工作具有较强的参考意义.

基于机器学习的低剂量 CT 图像降噪方法研究^[5] 作者的成果创新性的结合了神经网络图像降噪技术和放射科医生的具体诊断流程, 将放射科图像处理和后续医疗诊断有机结合起来, 在便利了医疗工作者图像处理的繁重工作的同时避免了处理后图像信息部分丢失的情况. 作者创造性地设计出了即插即用的图像处理网络框架, 具有很强的拓展性, 后续可以较为轻松地拓展到更多相关领域. 该研究紧贴应用实际, 创新性值得借鉴, 对于更多与生产应用紧密相关的研究项目有重大参考价值.

基于 FPGA 的神经网络加速器^[6] 作者利用 FPGA 的灵活性特点, 采用 HDL 对神经网络计算采取了针对性的结构设计, 通过量化, 剪枝, 循环展开, 循环并行处理, 缓冲区设计等多种方式提高了 FPGA 上运行神经网络的效率, 设计出了较为通用的 FPGA 神经网络加速器, 为更多未来研究奠定了坚实的基础.

基于卷积神经网络的多模光纤图像重构研究^[7] 作者设计了一种基于 VGG-16 的散斑重构网络, 创新性地使用 UpSampling2D 上采样层使得模型更具灵活性, 提高了模型的工作效率, 在图像重建方面推进了多模光纤成像技术的实际应用.

毫米波雷达手势识别轻量化网络研究^[8] 作者创新性地使用毫米波雷达进行对手势信息的采集, 并使用神经网络对采集到的毫米波雷达波形进行处理从而进行手势的分类识别. 作者采用了可瘦身机制这种具有灵活性的神经网络设计方式, 通过精妙的可切换式模块设计实现了模型的轻量化. 作者选取的研究方向在未来潜力巨大: 手势识别可用在诸多领域, 而轻量化的模型设计降低了使用门槛, 赋予小型物联网设备手势识别能力, 在未来的物联网应用中前景明朗.

基于神经网络的声音事件检测系统研究^[9] 作者设计了一种 LSED 算法, 基于 STFT 声谱图和小波变化优化了神经网络的计算, 并通过部署在 FPGA 上获得了加速计算能力, 实现了模型的小型化和高效化. 该成果可用于智能家居等场景下物联网设备的语音交互和智能感知, 在充分保护用户隐私的前提下进一步实现家庭智能化.

基于时空关系图网络的视频动作识别研究^[10] 作者创新性地构建了金字塔形时空图网络,通过时间注意力和时空注意力相结合的方式对神经网络进行强化训练,使得视频动作识别模型 STPG-Net 表现出良好的性能和较高的准确性.这一模型在未来可广泛应用于监控分析,动作检测,健身指导等多方面领域,具有广阔的应用前景.

4 展望

在这几篇论文中,可以发现近期神经网络领域的热点集中在:

- 对传统上难以量化的复杂信息进行量化,例如人的面部表情,声音,图像视频信息;
- 利用不同手段对神经网络模型进行优化,例如对模型进行剪枝,量化,采用深度可分离卷积,使用 FPGA,多 GPU 设备进行加速;
- 利用神经网络的推理能力,对原先计算机无法解决的复杂物理模型进行推断.

展望未来发展,神经网络的未来发展方向可能将聚焦于进一步对于日常生活中常见复杂信息的信息化采样与智能分析;同时进一步轻量化小型化,使得神经网络模型有能力运行在各种小型设备上,实现 AIoT (Artificial Intelligence Internet of Things),从而进一步实现人们生活的智能化;同时,随着神经网络在大语言模型领域的迅速发展,在未来可能会进一步结合神经网络的推理能力,弥补目前大语言模型在逻辑能力和推理能力上的缺失.

参考文献

- [1] 段风志. 基于可变形卷积网络的视频去模糊算法[EB/OL]. 2022. DOI: 10.27061/d.cnki.ghgdu.2022.002329.
- [2] 袁环宇. 基于移动终端的人体姿态检测技术研究[EB/OL]. 2022. DOI: 10.27005/d.cnki.gdzku.2022.002624.
- [3] 王聪荣. 基于轻量化卷积神经网络的表情识别与可视化分析[EB/OL]. 2022. DOI: 10.27061/d.cnki.ghgdu.2022.002158.
- [4] 程永龙. 基于 FPGA 和 Lenet-5 的字符识别系统设计与实现[EB/OL]. 2022. DOI: 10.27061/d.cnki.ghgdu.2022.001688.
- [5] 陈柯成. 基于机器学习的低剂量 CT 图像降噪方法研究[EB/OL]. 2022. DOI: 10.27005/d.cnki.gdzku.2022.002185.
- [6] 张渤钰. 基于 FPGA 的神经网络加速器[EB/OL]. 2023. DOI: 10.27005/d.cnki.gdzku.2023.000016.
- [7] 龚志宝. 基于卷积神经网络的多模光纤图像重构研究[EB/OL]. 2022. DOI: 10.27061/d.cnki.ghgdu.2022.001240.
- [8] 张丽君. 毫米波雷达手势识别轻量化网络研究[EB/OL]. 2022. DOI: 10.27005/d.cnki.gdzku.2022.002597.

- [9] 王雨江. 基于神经网络的声音事件检测系统研究[EB/OL]. 2022. DOI: 10.27005/d.cnki.gdzku.2022.003667.
- [10] 耿甜甜. 基于时空关系图网络的视频动作识别研究[EB/OL]. 2022. DOI: 10.27005/d.cnki.gdzku.2022.003750.
- [11] ZHOU S, ZHANG J, PAN J, et al. Spatio-temporal filter adaptive network for video deblurring [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019.
- [12] WANG X, CHAN K C, YU K, et al. Edvr: Video restoration with enhanced deformable convolutional networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2019.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C/OL]//GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: Vol. 30. Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.