

Huxianjun

专业技能

工程开发：熟练使用Python、C++等，熟悉安卓开发/数据库，云、边、端模型部署全链路开发

研究生与目标岗位相关专业课程：机器学习&NLP，数字图像处理，软硬件协同开发等

技术领域关键词：机器学习，深度学习，计算机视觉，NLP，多模态，点云，三维重建

个教育经历

2009.9-2013.6	黄冈师范学院	电子信息科学与技术	本科（统考全日制）
2015.9-2018.7	北京大学	电子通信工程	硕士研究生（统考全日制）

实习经历

2016.6-2016.10	中科院信工所	图像视觉算法实习（缺陷检测）
2016.11-2017.9	百度	CV算法实习（风控审核）
2017.10-2018.2	滴滴	计算机视觉算法（证照核验）

工作经历

2018.3.10-2019.10.14	搜狗（北京）	副研究员
2019.10.15-2020.11.24	青岳科技（北京）	视觉算法工程师
2020.11.25-2022.2.25	百度（北京）	计算机视觉算法工程师
2022.3.10-2022.9.30	上海深视（上海）	高级视觉算法工程师
2022.11.28-2023.12.29	阿维塔新能源科技有限公司（上海）	多模感知算法专家

项目经历

项目一：智能座舱视觉算法以及LLM大模型算法（2022.11~至今）

智能座舱视觉算法工作，包括人脸检测、人脸关键点检测、视线估计、表情及其他属性识别等等。主要做了两方面的工作，一个是算法优化，另一个是工程部署相关的工作。首先，算法层面，系统梳理了相关领域学术界和工业界的最新进展，其中关于疲劳检测小眼睛问题，通过调研和反复实验尝试，确立了一套算法逻辑，有效解决了问题，灵敏度比较高，针对鲁棒性问题，设计了前处理算法来得到自适应阈值。通过对眼睛轮廓关键点的反复验证试验，对以往算法做了一些改进。关于点头和摇头，从两个方向做了一些工作，一个是模型END-TO-END,另一个是头部姿态加后处理，两种方法各有优劣。视线估计梳理了一套可视化方案，将变化的视线以图形的方式展示出来。整体开发都是采用安卓JNI联合C++开发编译部署。另外，工程部署这块，核心是安卓端C++开发以及模型转换，另外是解决QNX平台板子部署问题。具体来说，安卓端开发相关算法实现DEMO，然后在QNX上部署。疲劳检测睁眼闭眼在以往算法的基础上做了一些局部优化。安卓端实现了视线估计算法以及可视化安卓端实现。支持组内各部分模型转换，解决高通平台转换不成功问题，最终成功移植到QNX嵌入式平台。其中，针对转换到高通平台推理性能低下的问题，自主探索了一套行之有效的方法，尝试交叉编译QNX嵌入式平台的动态库，同时在QNX平台编译了替代当前推理工具的第三方库，推理性能由原来的16FPS提升到160FPS，接近10倍的加速。针对相关场景感知算法、大语言模型、多模态方向提交专利若干。

大模型部分了解了stable diffusion扩散模型，同时试验了chatglm,chatglm2,baichuan,baichuan2等方法并针对特定场景结合QLARA等finetune model完成特定任务，其中针对chatglm2-6B,baichuan-7B开发了流式输出的逻辑服务。同时开发了结合大模型的百科服务，天气服务等。其中结合LLM的百科做了迭代优化的工作，实现了性能由10s左右到4s左右到2s左右再到不到1s的性能提升。整体完成了所负责模块的算法优化，对外提供服务以及集成等。另外完成了prompt管理平台的开发，我主要负责后端部分，完成前后端联调。

项目二：工业缺陷检测（2016年中科院信工所实习、上海深视科技 2022.3.10~2022.9.30）

很多工业领域对制造的产品外观有严格的要求，有可能每一道工序之后都会有一次检测的工序。而所有的这些工作，如果全部由人工来完成的话，不仅周期长成本高，而且由于人的视觉疲劳，正确率难以保证。关于手机划痕检测和PCB三防漆检测，数据采集采用工业相机，因为缺陷是非刚性的，而且大小不一，所以没法通过捕捉高层语义信息来识别，而低层信息比较有效。所以采用了结合底层信息类似于FPN的结构，单纯的分类损失无法有效监督，最终是联合检测loss和分类loss。

开发集成了缺陷检测统一框架-NDML，当前兼容不同缺陷数据。整体方案集成了度量学习，多分类、基于代理的正负样本的表征学习。重点开发了针对任务需求的特征提取网络，融合了多种特征提取以及对学习的监督方案。所有方案都可以自由配置，跑benchmark修改配置即可。代码总体上包括数据加载，数据采样，度量策略算法集合，backbone特征提取网络（包含多种中间结果输出），训练约束损失，自研多维度推理方案，可视化代码等几大模块。开发了缺陷数据增强方案，包括基于缺陷cell库构建缺陷数据和基于纹理库构建缺陷数据两种情形。基于缺陷cell库要事先建立比较丰富的缺陷cell样本，以及对应的mask，然后后续需要进行随机旋转、缩放、平移等操作，背景采用模板图，通过传统图像融合操作来生成丰富的缺陷数据，目前通过工业界调研发现，该思路已在工业4.0，智能制造领域落地。基于纹理库通过分析数据特点，OK与NG数据严重不平衡。

在实际项目迭代中，为了高效优化模型，数据快速处理至关重要。结合数据特点，存在目标对象分辨率低，缺陷特征聚焦 低语义信息，因此这种场景下的数据增强通过简单的 Cut-Paste 就能有比较好的效果，通过高维空间 投影可视化，验证确实如此。

项目三：人脸检测及关键点检测(搜狗 2018.7~2018.9)

人脸检测和关键点检测属于通用基础研究，目前放出的论文层出不穷，而且不少开源实现。各家的必争之地包括精度和速度，另外在视频中可能会涉及到稳定性的问题。人脸检测以及关键点的检测算法调研及实现，针对抖动现象的优化方案的探索。该方向 当时项目没有考虑优化，但是平时会持续关注该领域热点优化思路，譬如现在处理稳定性是会结合 3D 来确保大角度和遮挡的情况，另外，检测算法的模型压缩与加速，模型 在嵌入式设备上的加速部署。

项目四：NCNN,Tengine 在嵌入式端的模型压缩及加速部署(搜狗 2018.6~2018.8)

熟悉 mobilenet v1/v2/v3,shufflenet v1/v2 等小模型网络结构以及不同版本的 优化思路，熟悉常用的模型压缩方法，比如量化，剪枝，知识蒸馏等。能够使用常见加速框架部署模型并评估 FLOPS 以及 Parameter size 等指标。熟悉 pth-onnx-tensorRT 模型转换流程，分析转换出现问题并逐层定位分析，熟悉 tritonserver 搭建服务，会用 docker flask 以及 shell 一键部署服务。

项目五：人脸属性编辑(搜狗 2018.7~2019.7)

根据任务需求调研并确定了研究方向，主要思路是通过编码器复杂特征映射到隐空间，从而可以用一些编码方式(比如独热编码)来表征不同的属性，这样就将复杂问题 简单化，同时使后面的对抗训练能够更加稳定。本项目面临的主要问题是编辑某一属性 时，其他属性会受到影响，这样造成生成结果不理想。在此基础上的优化方案是引入了 门控机制，主要解决的问题是在属性编辑时对一些属性的重复编辑带来的影响，在编辑 某一属性时，尽量保证其他属性不发生变化，已达到生成结果的稳定性。当图片具有目标属性时，门控关闭，不进行编辑，否则，执行编辑策略。在人脸属性编辑的时候根据 需要定制化个性图像的任务需求，结合论文，自主通过 Keras 和 Tensorflow 复现 SC-FEGAN。在此过程中，需要关注不同区域，因此研究了一下脸部区域分割以及显著性目标检测的内容，显著性目标检测在处理分割边缘有一些比较好的方案，包括扩展的 可以进行比较精确的区域图片编辑，比如唇彩，美瞳，瘦脸(瘦腰)。

项目六：表情包图生文、文生图相关项目(2019年参加搜狗公司级别黑马大赛)

完成了本项目的 pipeline,包括前期的调研，方案确定，数据获取及清洗，数据预处理，训练优化，服务部署。在调研国内外现有方案的基础上，结合中文场景，训练了 可以产生对应表情包文字的模型。并在此基础上，拓展了图片生成诗词的模型。调研了 BigGan、Glow 等目前表现优秀的图像生成技术，目前针对抽象的表情包文本效果有限，从而提出了多模检索+风格迁移的设想。最后结果是图片到文字基本可以生成比较通顺 的段子，后期的优化点是加入语义平滑模块，使句子更通顺。目前整体方案已撰写相关 文档并申请发明专利。另外，通过 flask 搭建相应服务对外提供接口。

项目七：网络结构搜索(搜狗能力提升学习)

调研了网络结构搜索在分类，检测和语义分割方面的应用案例，目前都有比较好的搜索案例，有的精度超越人工，有的在速度占优。该项目做了一些实验，根据论文复现复现了 ProxylessNAS 以及 Mnasnet,搭建多机多卡分布式开发环境进行网络搜索，解决了搭建多机分布式训练的各种坑。熟练使用 horovod 以及 pytorch 的分布式框架。

项目八：人脸融合(搜狗 2018.9~2019.10)

人脸融合目前有比较多的方案，在实践过程中遇到的问题是融合之后出现鬼影和肤色不均的现象，解决方案是提取高层和低层语义信息通过泊松融合的方式来处理。方案详细思路如下，在调研国内外现有方案的基础上，梳理了主要的处理思路。目前比较类似的做法是换脸，采取分别训练预先训练好两个编码器和两个解码器，然后输入的 A 和 B 可以通过编码器分别投影到隐空间，然后将隐空间特征通过解码器映射回图片，中间可以任意添加 mask 来完成局部特征的融合。这样在更高维度的融合可以使最终的效果自然逼真。在最终生成结果时，加入了对抗训练。最终形成了实现功能的demo，并设计成可视化服务部署。

项目九：融合 3D 信息的表情生成技术(搜狗 2019.6~2019.10)

通过调研学术界在表情迁移方面的工作，初步确定了两条路线。一种是 GAN，另一种是结合 3D 信息。最终考虑到可控性，采用了融合 3D 信息的方案。通过 2D 到 3D 的映射，得到一些跟踪信息，这些信息是单纯 2d 无法表达的，进而可以跟踪表情变化，实现表情迁移，另外，考虑到局部细节，采取了后处理局部完善的思路，该方案整体可控性更强，生成的表情图片或视频质量更高。目前阶段成果是完成 3D 信息拟合，以及后处理的工作，后处理主要解决一些细节的填充，目前实验效果很不错，比如生成的牙齿轮廓纹理清晰可见，分辨率高，自然逼真，但是其他环节还有一些问题需要解决。另外本项目解决的另外一系列环境冲突版本不兼容问题，包括所涉及到的工具如 PCL,VXL, boost, cuda 等的版本问题，在处理 python 和 c++相互转换时的库冲突问题，对问题的跟踪定位及有效解决。

项目十：Reid 与目标跟踪(北京青岳科技 2019.10~2022.11)

深入调研了目前行人重识别的主流处理方案，主要有基于表征学习的，基于度量学习的，基于局部特征的，基于视频序列的，基于 GAN 造图的等方案。目前尝试了一些 backbone 以及训练了相应的模型，不断刷新了 baseline,了解 re_ranking 算法机制，同时修改了 baseline 网络架构，融合了更多的信息进来，将原来的 baseline mAP 提升了两个点，同时 rank1 也提升了一个点。目标跟踪也调研了主流的方法，熟悉卡尔曼滤波算法，熟悉匈牙利算法和粒子滤波。了解单目标的 sort 和 deepsort,并且将前面训练好的最优 reid 模型整合进来，通过该模型提取特征。深入分析了 deepMOT 以及 JDE，针对 deepsort ID-SWITCH 提出了一种思路，就是在跟丢更新 ID 前采用局部特征而不是 REID 特种纠错，有待进一步实现及验证。利用 flask 搭建相应的服务提供 API 接口以及可视化的功能，整合了迭代通用性强的 pipeline。整个工作内容涉及 ReID, SOT, MOT, Multi view Multi object Tracking,做了大量的实验以及思路方法的梳理工作。

项目十一：表结构识别（基于图像分割联合 OCR 结果的表结构识别方案）（北京百度 2020.11~2022.2）

电子化制表程序 (例如 Excel) 是现代企业办公经常用到的软件, 其中包含很多有价值的信息。不同表格的版式、布局和文本元素差异较大, 使得自动表格信息面临着很多挑战。分析近几年学术界和工业界的方案, 梳理了一下主要思路。基于传统图像处理 (例如图像分割) 的方法处理全线框表格准确率较高, 但是泛化能力不足, 而且对于非全线框的表格捉襟见肘; 基于深度学习模型抽取特征的方法, 泛化能力不错, 能抽取图片中的语义信息, 但是准确率稍差, 尤其是当表格样式复杂多样、外部环境差异较大时, 准确率无法保障; 图神经网络面临网络深度与计算复杂度的问题以及过度平滑的问题, 另外少线表和错切表效果仍不理想。

另外, 实际应用方面, 基于传统图像处理的方法只能应对全线表格, 但是实际生产中全线表格和非全线表格占比相当。基于图神经网络的方法在工业界大规模落地面临的主要挑战主要来自于大规模图网络的训练以及线上更新预测两方面。我的方案包括 1) 图像矫正; 2) 表格线分割; 3) 后处理几大模块。产出一篇发明专利, 该发明主要用于处理规整或存在倾斜的无畸变的文档, 提取表格结构信息, 转化为 XML 等结构化数据格式。另外, 持续跟踪业界最新文档结构化方案。

项目十二: 文档矫正 (北京百度 2021.6~2022.2)

另外, 持续跟踪业界最新文档结构化方案, 文档矫正主要处理的是自然场景的文档, 这一类文档的显著特点是背景干扰因素, 拍摄镜头方向随机, 人为因素导致的扭曲, 属于文档结构化上游细分领域。目前在此方向的核心工作是开发了基于特定场景数据的包含十种形变特点的数据集, 包含原始图, 光流图, 遮罩图。系统梳理了目前的一些方法, 包括: 1) 基于 CNN 的融合全局信息以及局部信息来拟合矫正光流; 2) 基于时序模型的光流拟合; 3) 基于对抗生成网络的方法。借鉴一些文章的思想, 只要在以下几方面做了探索工作: 1) 矫正光流拟合生成矫正后的图像的优化起到加速的效果。2) 损失约束借鉴一些思想, 包括局部与全局的约束, 基于图像重建的约束, 以及直线度的约束等形成多任务联合约束。3) 整个网络的端到端训练的改进。

业余爱好及特点

读 paper、打羽毛球、象棋

优秀的自学能力, 很好的自驱力, 具有良好的抗压能力和团队协作能力以及爱好钻研