

第三章 产品与技术

3.1 产品介绍

3.1.1 产品概述

本产品是一款针对老年代步电车(如图 3-1)设计的智能辅助刹车系统。该系统采用国际领先的基于机器学习和深度卷积神经网络的动态图像检测技术,可将图像识别准确率提高至 90% 以上。同时,结合云计算技术,可大大缩短识别时间,将反馈时间严格控制在毫秒(ms)级别。预计 5G 技术实现后,这一时间还能进一步压缩。此外,在深度卷积神经网络的训练下,该系统可识别障碍物种类多达 17 种,针对不同障碍物,可提供 5 种刹车方案。系统通过“先识别、再预警、后刹车”的决策逻辑,避免出现错刹、误刹、漏刹,为老年群体提供了代步电车驾驶安全的第二重有力保障。



图 3-1 老年代步车

3.1.2 产品功能及架构

系统技术分为本地和云端两部分,分别承担一定的图像识别计算任务,减少成本的同时,大大降低反馈时间,使系统更加稳定可靠。

本地通过主板 CPU 模块,对图片进行卷积(Convolution & Pooling)过程,在极短时间内,将大体积图像压缩为小体积特征点数据,大大减少图像传输的带宽占用,也降低了硬件

成本。云端将特征点数据放入经过训练的深度学习神经网络 (Deep Neural Network) 中进行识别，借助高性能服务器强大的算力，在极短时间内得到结果，并通过决策模型反馈指令。本地依据指令做出相应决策。整个过程被严格控制在毫秒级别。特别地，为了应对网络缺失状况，系统还集成了红外线距离传感器，在危险情况下给出应急操作。

3.2 技术研发路线

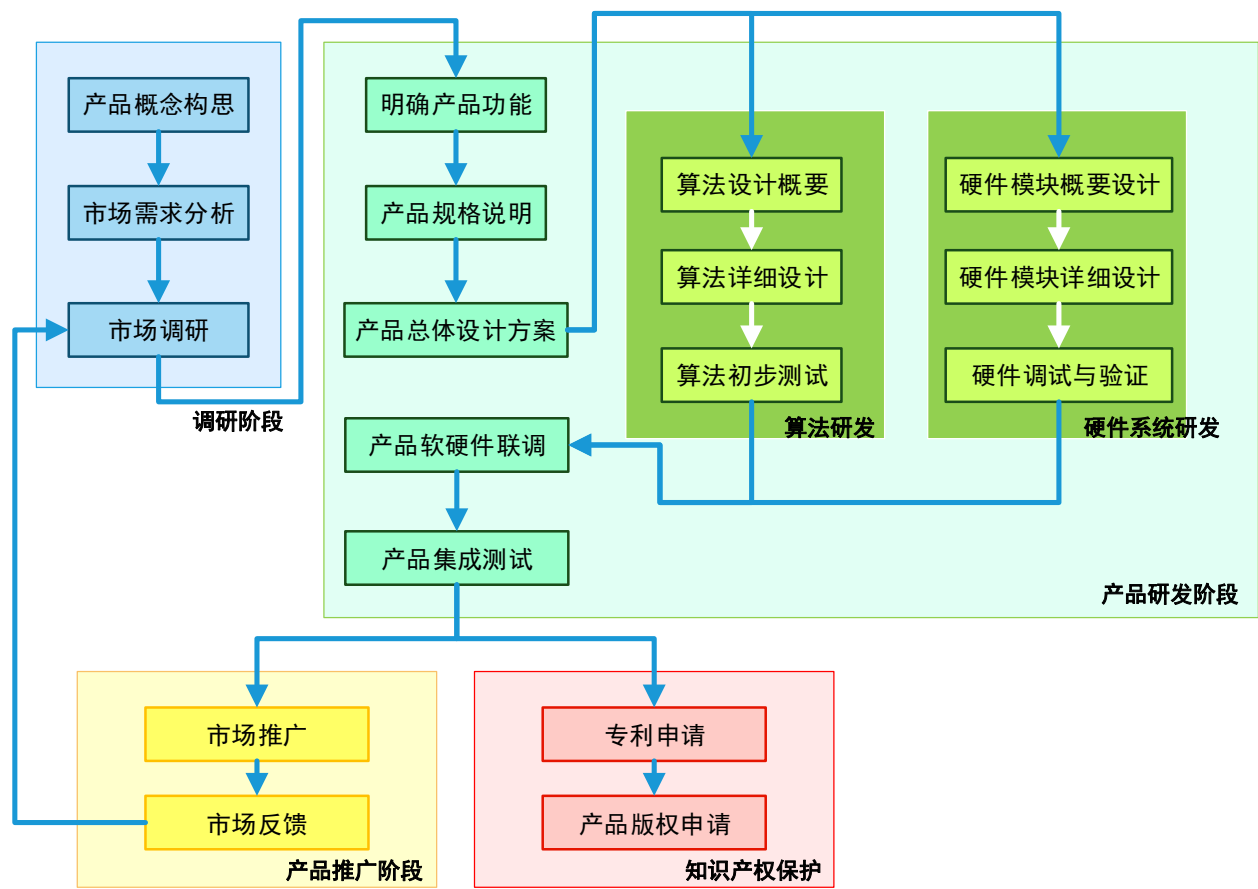


图 3-2 老年代步车智能辅助刹车系统技术研发流程图（待改）

3.2.1 产品需求分析

产品功能由其需求所决定。已有的汽车辅助刹车系统，大部分依靠红外线距离传感器和雷达探测障碍物进行制动，少部分高端车型具备图像识别功能，其实现过程全部依靠本地硬

件。据估算，完全基于离线图像识别的老年代步电车辅助刹车系统，为了满足识别速度和准确度的要求，需借助 CPU 与 GPU 的协同运算，硬件成本将达到 2000 元。

过高的成本将提高电动车购买门槛。为了降低硬件成本，我们借助高性能的云计算，在多辆电车共享同一高性能设备的同时，也保证了识别速度和准确性的要求。此外，考虑到即将到来的 5G 时代，能够将网络延迟控制在 1ms 级别，数据传输的耗时可以忽略不计。

但是实时的图像传输，相当于电车与服务器始终保持视频级别的交流，这一过程将占据大量带宽。因此，我们考虑采用分步执行的方法，先进行本地图像卷积与池化，将提取到的特征点上传至服务器，再由服务器进行识别反馈。这一方案不仅能够大大降低带宽占用，也方便了日后深度神经网络结构的调整与算法的更新，为技术迭代做铺垫。

考虑到网络缺失的特殊情况，我们在本地集成了红外线距离传感器，以应对突发情况。

3.2.2 算法研发

图像识别算法原理

图像识别是该系统的核心部分，对于图像识别我们采用卷积神经网络图像识别的方式。如下图 3-8 所示，卷积神经网络是一种为了处理二维输入数据而特殊设计的多层人工神经网络，网络中的每层都由多个二维平面组成，而每个平面由多个独立的神经元组成，相邻两层的神经元之间互相连接，而处于同一层的神经元之间没有连接。一个简单的卷积神经网络模型的结构示意图如图所示，该网络模型由两个卷积层(C1, C2)和两个子采样层(S1, S2)交替组成。

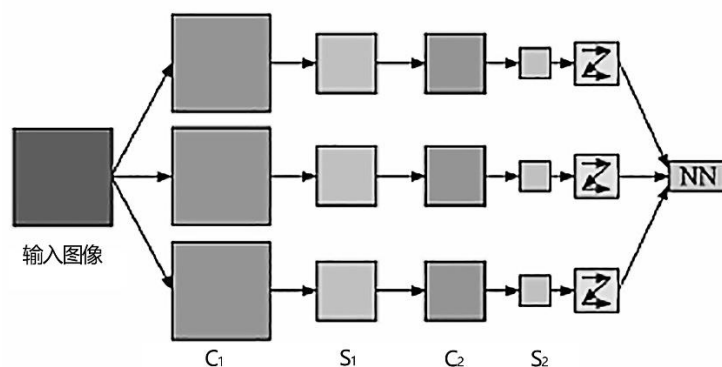


图 3-8 卷积神经网络图像识别原理

图像识别的具体技术细节如下所示：

卷积特征提取：

自然图像有其固有特性，即对于图像的某一部分，其统计特性与其他部分相同。这意味着在这一部分学习到的特征也能用在另一部分上，因此对于图像上的所有位置，可以使用同样的学习特征。换句话说，对于大尺寸的图像识别问题，首先从图像中随机选取一小块局域作为训练样本，从该小块样本中学习一些特征，然后将这些特征作为滤波器，与原始整个图像作卷积运算，从而得到原始图像中任一位置上的不同特征的激活值。如下图 3-9 所示，给定分辨率为 $r \times c$ 的大尺寸图像，将其定义为 x_{large} ，首先从 x_{large} 中抽取 $a \times b$ 的小尺寸图像样本 x_{small} ，通过训练稀疏自编码器得到 k 个特征和激活值 $f(Wx_{small}+b)$ ，其中 W 和 b 是训练得到的参数。然后对于 x_{large} 中每个 $a \times b$ 大小的 x_s ，计算对应的激活值 $f_s(Wx_{small}+b)$ ，进一步使用 x_{small} 的激活值与这些激活值 f_s 作卷积运算，就可以得到 $k \times (r-a+1) \times (c-b+1)$ 个卷积后的特征映射图。如下图 3-9 所示。

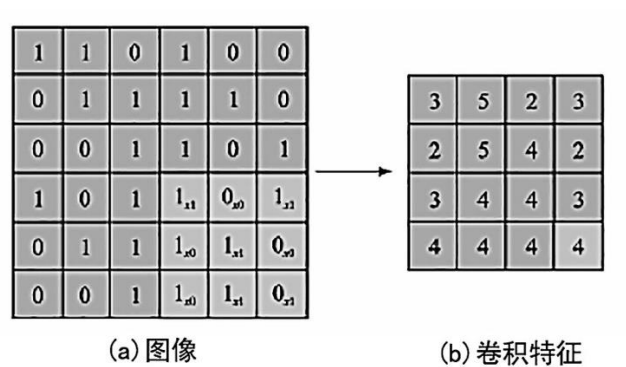


图 3-9 卷积特征提取过程

池化操作

进行卷积特征提取后，将会生成高维度向量，该高维度向量放入分类器中进行训练需要产生非常庞大的计算资源，也会产生过拟合问题。然而，由于图像具有一种“静态性”的属性，在图像的一个局部区域得到的特征极有可能在另一个局部区域同样适用。因此，可以对图像的一个局部区域中不同位置的特征进行聚合统计操作，这种操作称为“池化”。池化之后特征单元不变。与直接使用卷积后的特征相比，这些概要统计特征不仅能够极大地降低特征向量的维度，进一步降低训练分类器所需的计算量，而且能够有效地扩充训练数据，有利于防止过拟合。

图像分类

图像分类问题是通过分析图像，将图像划归为若干个类别中的某一种，主要强调对图像整体的语义进行判定。找出图像的最优局部稀疏结构，并将其近似地用稠密组件替代。这样做一方面可以实现有效的降维，从而能够在计算资源同等的情况下增加网络的宽度与深度;另一方面也可以减少需要训练的参数，从而减轻过拟合问题。运用 ImageNet 所训练的模型具有很强的推广能力，在其他数据集上也能取得良好的分类结果;而如果进一步在目标数据集上进行微调，与只用目标数据集进行训练相比大多能够获得更好的效果。

物体检测

与图像分类比起来，物体检测是计算机视觉领域中一个更加复杂的问题，因为一张图像中可能含有属于不同类别的多个物体，需要对它们均进行定位并识别其种类。因此，在物体检测中要取得好的效果也比物体分类更具有挑战性，运用于物体检测的深度学习模型也会更加复杂。利用 R-CNN 模型，对于每张图像大约产生 2000 个候选区域，而对于每张图像，它的所有候选区域都要分别进行特征提取。该模型可以较为有效的进行物体检测。

图像分割

在进行上述步骤之后，进一步便是对图像上每个像素点的预测，这个任务就是图像分割。图像分割是这样一类问题：对于一张图来说，图上可能有多个物体，多个人物甚至多层背景，希望能做到对于原图上的每个像素点，能预测它是属于哪个部分的（人/动物/背景）。利用全卷积网络来进行图像分割。首先利用一些流行的分类网络在保留一些它们在图像分类方面训练所得参数基础上，进行“修剪”，转变为针对图像分割的模型。然后，他们将一些网络较深的层的所得特征和一些较浅的层所得特征结合起来，最后用一个反卷积层放大到原始图像大小来提供一个更为准确的分割结果,称之为跳跃结构,如下图 3-10 所示。

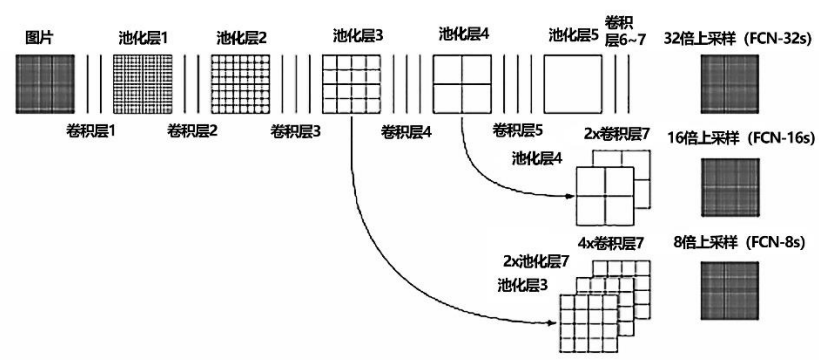


图 3-10 图像分割过程

卷积神经网络每卷积一层，实际上信息量都会丢失一些，所以如果想增加信息量，要做的就是到更浅的网络层获取信息，这是之前所提的跳跃结构的核心思想。对于信息丢失

过多的最后一层卷积层，可以先将它反卷积扩大 1 倍，达到与上一个卷积层 pooling 完了之后的一样的大小，之后，将两者的信息整合(一般是相加)，进而做次反卷积，这样就只需再放大 16 倍，取得的效果也有所提升。可以更进一步，再加入 pool3 的信息，也就是将之前一步的结果先再做一次扩大 2 倍的反卷积(相较于最初实际上相当于扩大了 4 倍)，与 pool3 的结果相加后，再做一次放大 8 倍的反卷积。通过该过程可以更好的对物体进行识别。

物体识别

将物体图像划分成若干个重叠的区域，对每个子区域都训练一个卷积神经网络，并将多个网络的最后一个隐含层的特征进行融合，作为受限玻尔兹曼机的输入，通过反向传播算法进行整体联合优化。最终实现物体的准确识别。

3.2.3 本地技术实现

电动车的本地技术实现，按功能分为如下几个模块：图像捕捉模块、卷积压缩模块、刹车联动模块、语音提示模块、制动执行模块。

图像捕捉模块

图像捕捉的主要功能由硬件负责实现，但不同种类的摄像头录制视频制式与格式不同，因此，需要特殊的格式转换进行数据统一。该系统中，将视频统一以帧为单位进行图像捕获，传递给 CPU。

此外，红外线测距模块的反馈，作为特殊的图像数据，可实现除图像外第三维度的数据支撑，为决策提供帮助。

卷积压缩模块

基于卷积神经网络的图像识别算法，卷积部分借助本地 CPU 进行实现。卷积的主要作用是，将大体积图片压缩为小体积特征点图片 (Feature Map)，方便与服务器的数据交流。其

主要包括卷积（Convolution）与池化（Pooling）两个过程的交替进行。

经测试，在 Raspberry Pi 3 Model B 型号主板，实验环境为 Linux CentOS7 系统，Python3.7 语言，Keras 库（使用 Tensor Flow backend），四核 CPU 频率 1.2GHz 的算力支持下，经过四次的卷积（Convolution）与池化（Max Pooling），对 1000 张图片测试取平均值，每张三维图片的卷积时间保持在 10um 左右，卷积后图片被压缩为 25（Fliter）*18（px）*18（px）的特征值矩阵（Feature Map）。

实验充分验证了本地图片压缩的可行性与优越性。

刹车联动模块

为解决刹车系统与驾驶者的“争权”问题，刹车联动模块应运而生。正常驾驶时，驾驶者的刹车权处于优越状态，当系统检测到危险，且超过一定阈值（障碍物危险级别及相对距离）驾驶员仍未采取刹车动作，则由刹车系统采取相应动作。

语音提示模块

语音提示模块用于实现行驶中的警示与提醒。通过接收云端反馈信息，播报不同语言提示。该模块目前借助百度 AI 语音合成接口，在实验室条件下初步实现。

制动执行模块

不同车型由于其体积、质量的不同，有效制动距离也相异。同时，考虑到老年代步电车可能采用胀闸、碟刹、油刹等不同刹车方式，因此，制动执行算法也需要做出相应的改变。我们参考防抱死刹车系统（ABS），在算法层面上设计了防抱死系统。保证制动时，刹车片与车轮有 20%左右的滑移率，使车轮与地面处在边滚边滑状态，以获得最大摩擦力。

3.2.4 云端技术实现

云端技术与本地技术相辅相成，互为补充与支持。且未来 5G 技术的实现，能够让云计

算的时间成本大大缩短。当前 4G 网络在网络条件良好的情况下，延迟在 43ms 左右，据报道未来的 5G 技术，这一时间将缩短至 1ms，耗费在网络上的时间延迟几乎忽略不计。考虑到刹车系统的整体要求，云端系统功能分为如下两部分：图像识别分类与决策模型。

图像识别及分类

云计算主要用于解决图像识别问题。前期，我们会大量收集高质量道路障碍物图片，用以训练出高准确度的深度神经网络。借助高性能服务器的硬件支持，该网络能够快速得出可靠的识别结论。

在实际测试中，Linux CentOS7 操作系统下，使用 Python 语言 Keras 库（使用 Tensor Flow 作为 backend），硬件条件 CPU Intel i5 9400F，显卡 AMD RX580，通过测试识别 1000 张卷积后的彩色图片（25*18*18），得到每张图片平均识别时间为 143um，识别准确率为 90.2%。实验充分证明高性能服务器能大大缩短图像识别时间，相比于本地计算，云计算能够在相同时间内识别更多图片、服务更多用户，同时拥有更深的网络层次，更高的准确度。

决策模型

决策模型前期通过实验室测试及经验得出，以稳定性为主要追求目标。后期通过收集大量用户数据，分析位置信息及刹车习惯，总结出更多、更全面的影响因素，如天气原因、道路材质等，建立出更规范、可信可靠的决策模型，提供更智能的决策。

3.3 亮点与特色

1. 运用机器学习和卷积神经网络的图像识别技术，大大提高识别准确度，提高决策正确性；
2. 本地与云端的分离计算，能够大大降低硬件成本，提高利润空间；
3. 算法全部在云端实现，算法更可控，技术迭代更方便；

4. 可移植性强，支持针对不同车型进行定制，市场广阔。