**目 录**

引言 1

1 研究内容完成情况 1

1.1 深度学习模型压缩研究 1

1.1.1 Discrimination-aware network pruning for deep model compression[1] 1

1.2 拟人机器人系统研究 3

1.2.1 陪伴机器人 3

1.2.2 语音导购机器人 6

1.2.2.1 技术方案 7

1.2.2.2 算法原理 8

1.2.2.3 数据集 11

1.2.2.4 模型评测 11

1.2.2.5 系统展示 12

1.3 应用示范 14

1.3.1 虚拟换衣系统 14

1.3.1.1 需求分析 14

1.3.1.2 解决方案 15

1.3.1.3 系统详细设计与实现 15

1.3.1.4 结论 20

2 总结 21

参考文献 22

插图清单

图 1 DCP方法的描述 1

图 2 DCP算法总体流程 2

图 3 用于通道选择的贪心算法流程 2

图 4 桌面陪伴机器人电路图 4

图 5 桌面陪伴机器人实际电路连接图 4

图 6 专家系统中控软件界面图 6

图 7 导购机器人架构示意图 8

图 8 领域识别模型 8

图 9 一个序列标注的例子 9

图 10 槽填充模型结构图 10

图 11 状态转移图 11

图 12 换衣系统架构图 15

图 13 换衣系统前端界面图 16

图 14 换衣系统流程图 17

图 15 (a)人物图像 (b)关键骨骼点 18

图 16 (a)人物图像 (b)语义分割结果 19

图 17 (a)试穿衣服 (b)扭曲的试穿衣服 19

图 18 (a)人物图像 (b)试穿衣服 (c)换衣结果 20

附表清单

[表 1 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果 3](#_Toc104474160)

[表 2 在人脸识别任务与不同数据集上的对比实验结果 3](#_Toc104474161)

[表 3 领域识别验证结果 11](#_Toc104474162)

[表 4 意图识别（操作意图）验证结果 11](#_Toc104474163)

[表 5 意图识别（需求意图）验证结果 12](#_Toc104474164)

[表 6 多任务槽值填充模型与单任务槽值填充模型实验结果对比 12](#_Toc104474165)

[表 7 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果 18](#_Toc104474166)

# 引言

随着视觉感知、语音识别与合成技术以及人工智能技术与产业的发展，智能机器人与人们生活具有越来越密切的联系，人机交互已经获得很大的进步。但是对于多模态信息的智能理解与融合仍是在许多人机交互在真实环境场景应用急需解决的问题。本年度主要研究内容如下：（1）深度学习模型压缩研究；（2）对拟人机器人系统进行开发：陪伴机器人，语音导购机器人；（2）在行业进行示范应用：虚拟换衣系统。本年度计划申请软件著作权4项，整理论文集和结题报告。目前已经申请软件著作权5项，完成软件系统开发2项。

# 研究内容完成情况

## 深度学习模型压缩研究

### Discrimination-aware network pruning for deep model compression[1]

过大的模型参数与高昂的计算成本使得如今许多神经网络模型无法放到限制内存与计算资源的设备上推演，而模型压缩是解决这个问题的一个有效方法。近几年在模型压缩上的研究主要可分为模型量化（Model quantization）、稀疏低秩压缩（Sparse or low-rank compression）、网络剪枝（Network pruning）三类，然而前两种方法存在参数离散化导致性能下降[2]、不规则卷积核导致运算困难[3]等问题。

针对上述问题，我们提出了一种新的网络剪枝方法，称为Discrimination-aware channel pruning（DCP）。如图 1所示，该方法融合了目前网络剪枝的两种主要策略：Training from scratch[4]、Reconstruction[5]。

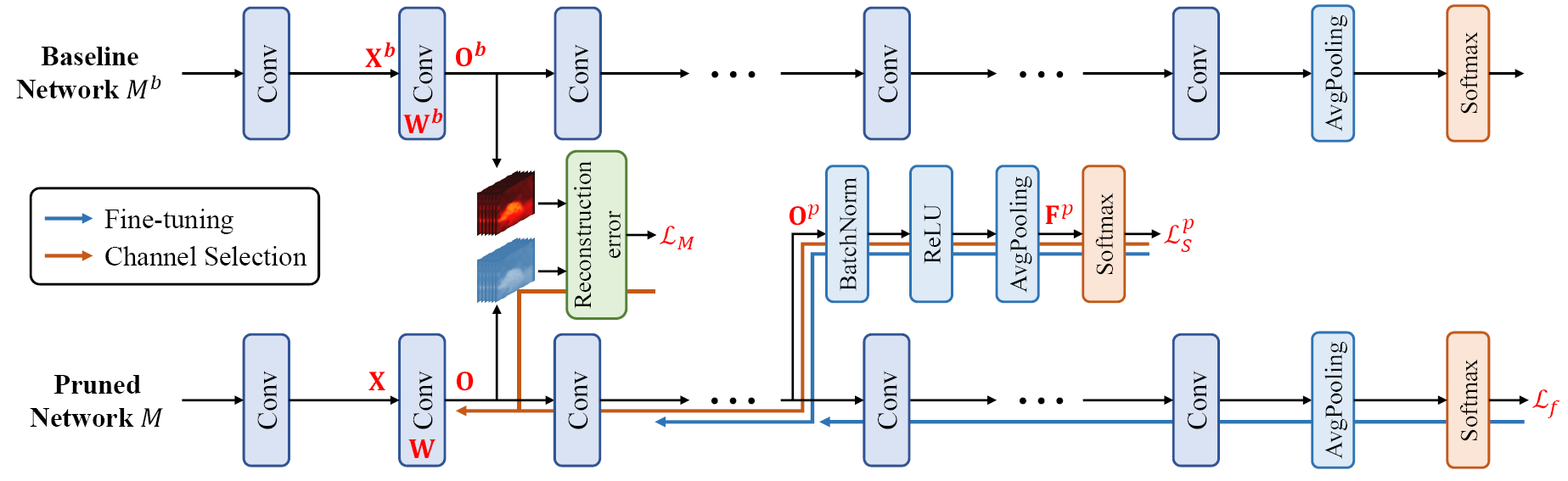


图 1 DCP方法的描述

DCP方法将网络剪枝任务视为模型通道选择任务（Channel selection），认为剪枝后模型每一层的通道都应该对其分类能力有所贡献。鉴于目前主要的剪枝策略，DCP将通道选择分为两个步骤：1）将判别性Loss加入模型并进行微调、2）使用判别性Loss和重构Loss对微调后的模型进行迭代式通道选择。

对网络中的一些卷积层而言，我们构造了能够用于提升网络分类能力的判别性Loss，并在该层卷积运算后叠加BN、ReLU、平均池化与全连接，以衡量该层输出特征图的分类能力，该过程如图 2所示。

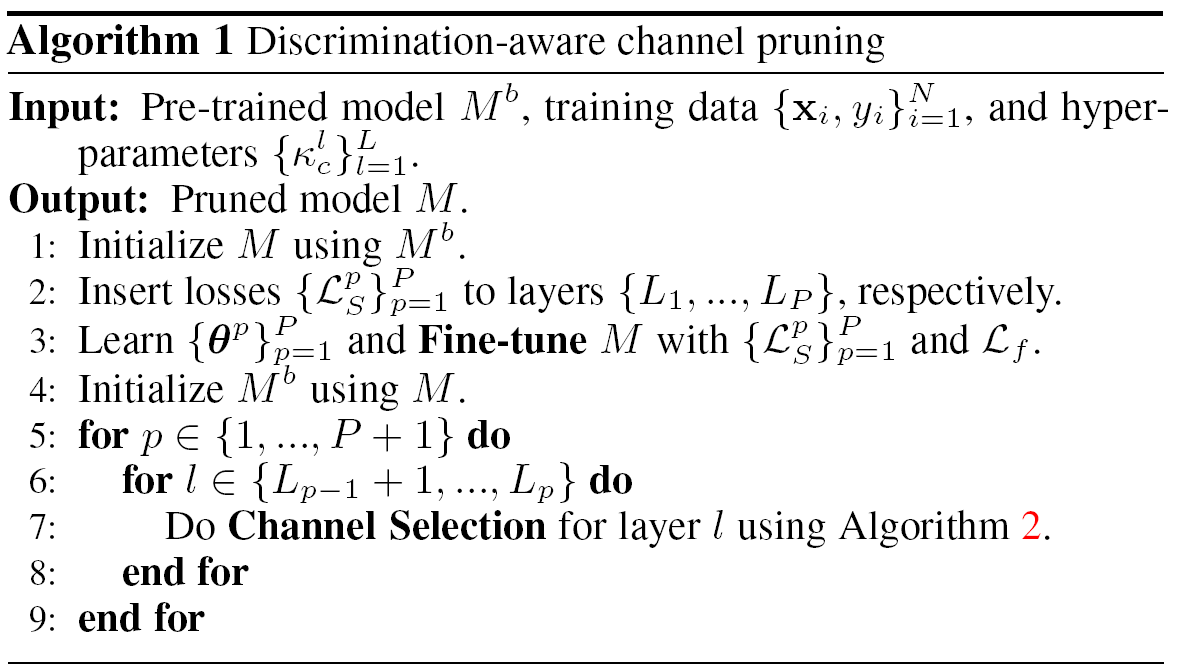


图 2 DCP算法总体流程

在完成模型微调后，仍然针对网络的某一层，我们利用原始模型中该层的输出特征图与微调后该层的输出特征图构造出重构Loss，并利用我们提出的*l*2,0正则化将通道选择问题转化为一个非凸优化问题[6]。我们随后提出了用于通道选择的贪心算法，从而将该非凸优化问题转化为迭代式求解的凸优化问题，该流程如图 3所示。

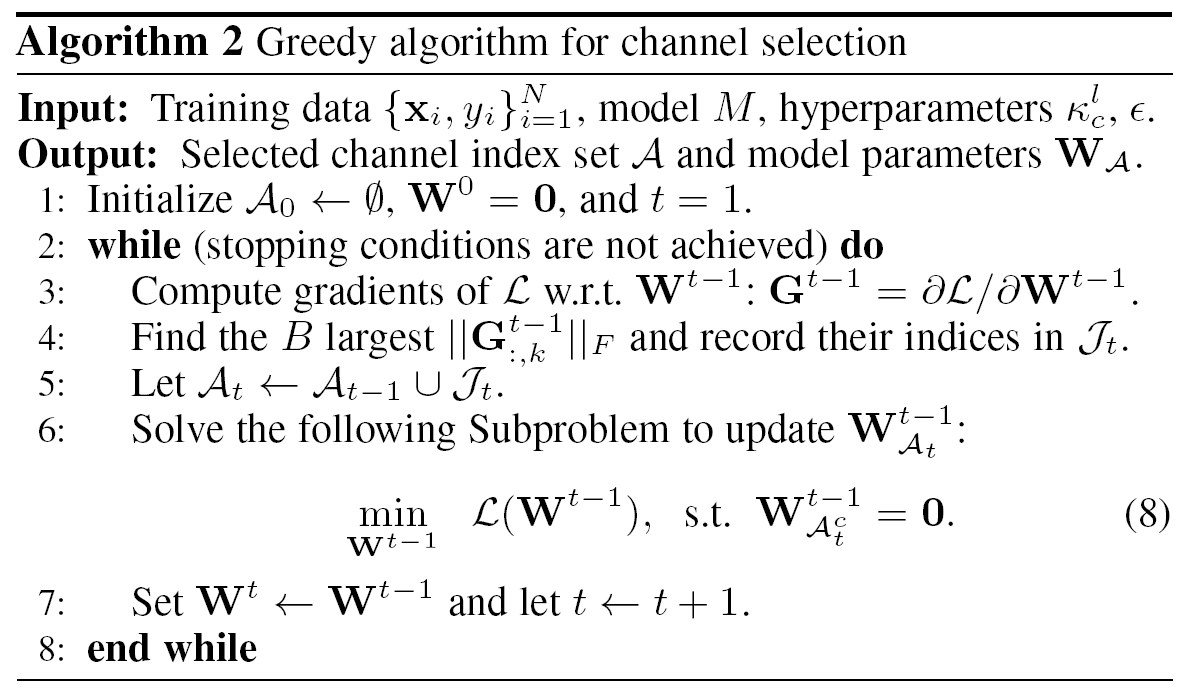


图 3 用于通道选择的贪心算法流程

以上两个步骤涉及到DCP在训练和实现上的许多细节，包括：1）判别性Loss的叠加次数与策略选择、2）求解最优化问题时通道重复选择的避免、3）迭代式求解的自适应终止条件设计、4）提升模型微调速度与迭代优化速度的方法、5）各种超参数的取舍和影响效果等。此外，鉴于DCP存在通道粒度大而导致仍然有多余的卷积核被保留等问题，我们又提出了该方法的变体，称为Discrimination-aware kernel pruning（DKP），DKP方法除了将DCP中的选择对象从模型通道改为通道内的卷积核以外，其他流程基本类似。

在图像分类上，我们针对ResNet-50、MobileNetV1等模型，将DCP与DKP同NISP[7]、ThiNet[8]、CP[9]等SOTA的剪枝方法进行对比，表 1 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果所示的实验结果可看出我们的方法在大多数图像分类模型上都有最好的压缩效果，同时体现出最高的分类准确度。在人脸识别任务上，我们将DCP方法应用到LResNet34E-IR、MobileFaceNet模型上，并与未剪枝的SphereFace和CosFace模型对比，从表 2 在人脸识别任务与不同数据集上的对比实验结果所示的实验结果也可看出我们的方法拥有最高的准确度，在移动设备上的推演时间也大幅降低。本文还针对上述细节做了一些消融实验，从而得出以下结论：1）针对不同模型，适当的剪枝率、Loss叠加次数、迭代终止条件等参数和策略的选择能够使得剪枝后的模型拥有最好的性能，甚至能够超过原模型的效果；2）采用预存模型参数、单轮微调、特征图重用等技术能够在不对模型性能有大影响的前提下大幅加速剪枝过程。

表 1 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果

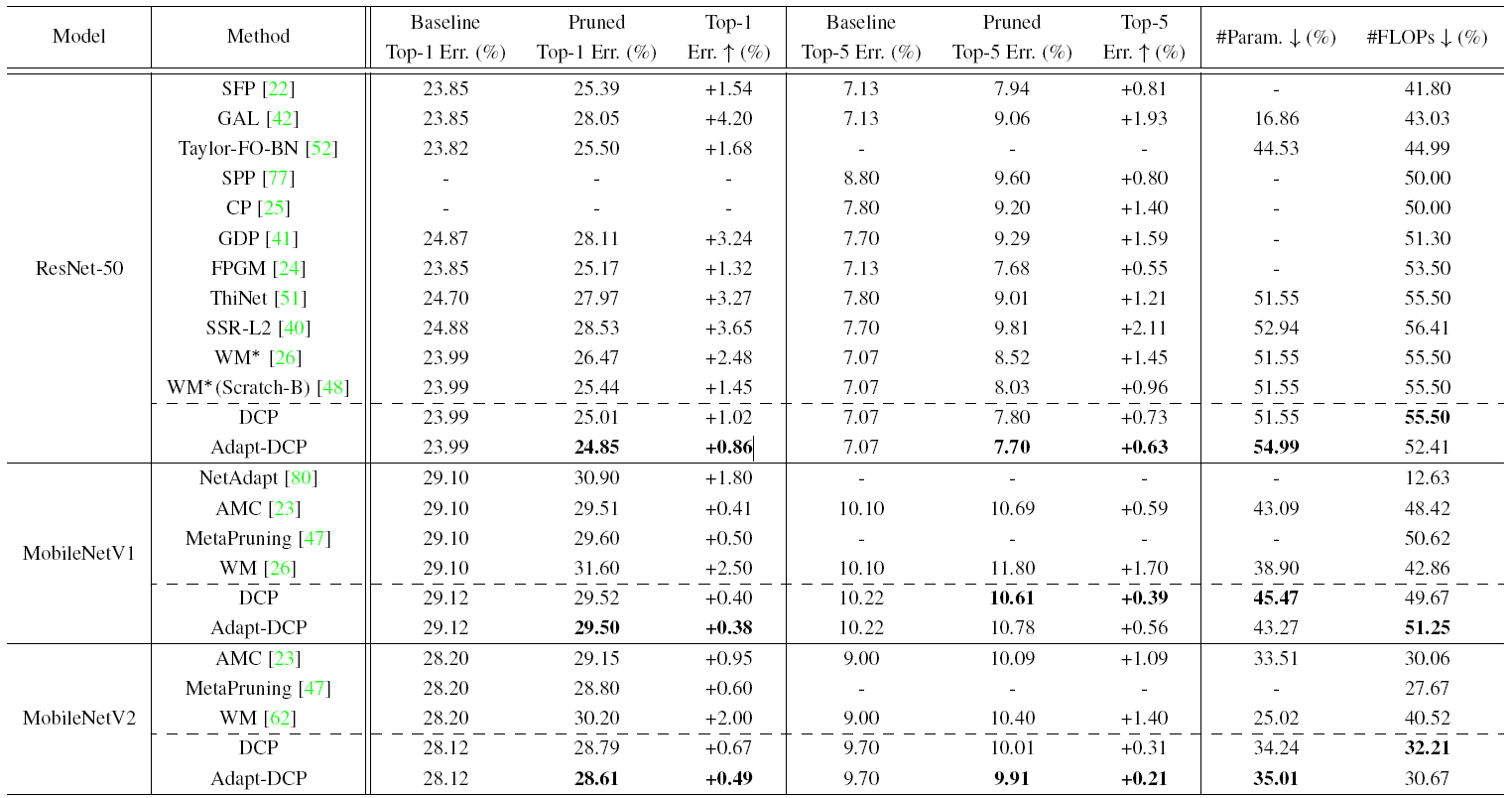
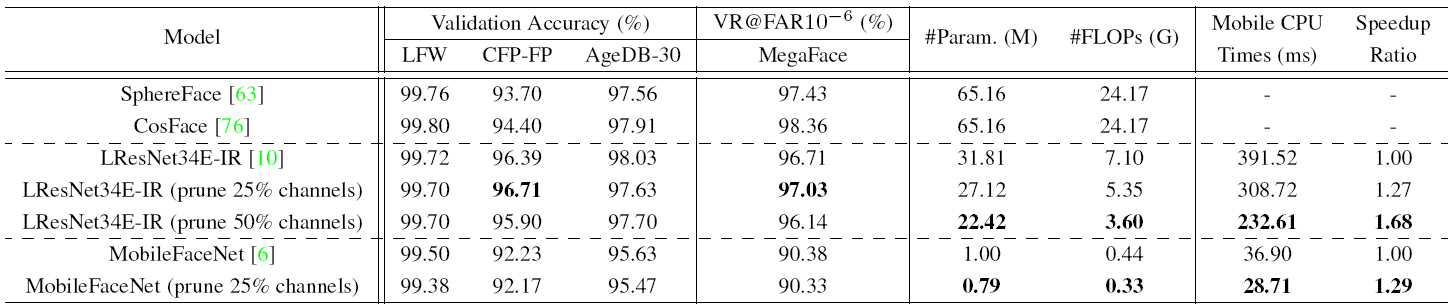


表 2 在人脸识别任务与不同数据集上的对比实验结果



## 拟人机器人系统研究

### 陪伴机器人

陪伴机器人是基于esp8266及一些相关外设开发的桌面陪伴机器人，它包含了硬件系统和软件系统。其中，硬件系统由esp8266主控板、OLED屏幕、LED小灯泡、光敏电阻、蜂鸣器、超声波传感器、触摸传感器、杜邦线、面包板等组成；软件系统由electron +vue+typescript开发。软硬件系统间的通信协议主要为socket，其中socket首先用于在硬件系统的AP模式下，软件系统与其通信并对其进行变量配置；然后用于当软硬件系统连接到同一局域网后，进行指令发送的操作。接下来，具体介绍一下软硬件系统的架构及开发细节。

硬件系统电路图如图 4所示，esp8266控制板的主板管脚数量不够，因此需要外接面包板和电阻来拓展管脚。具体的成品图如图 5所示。硬件设备中，主控板代码由micropython编写，代码主要分为三大功能模块：Ombot单神经元跳跃机器人模块、网络配置模块、专家系统指令接收及操作模块。

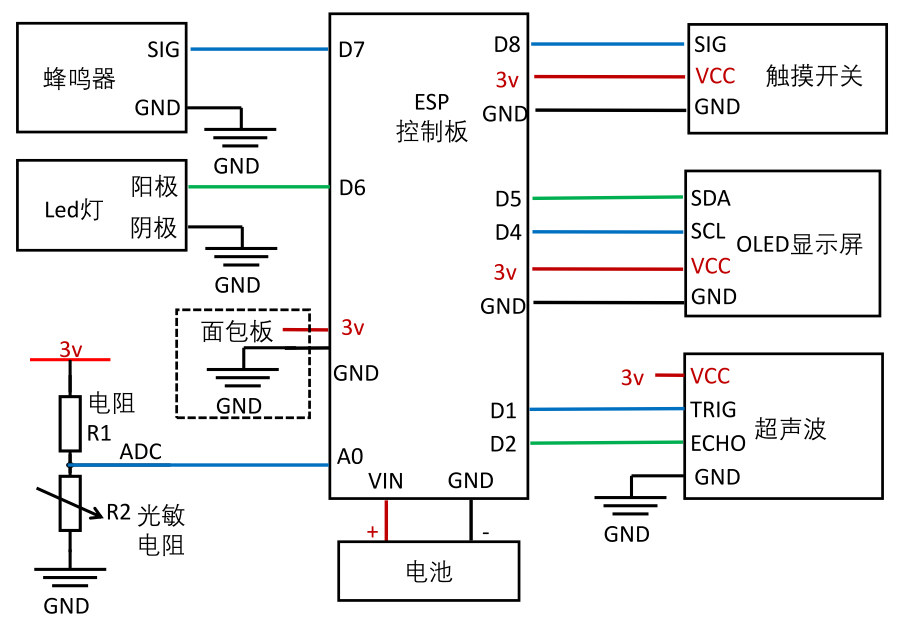


图 4 桌面陪伴机器人电路图

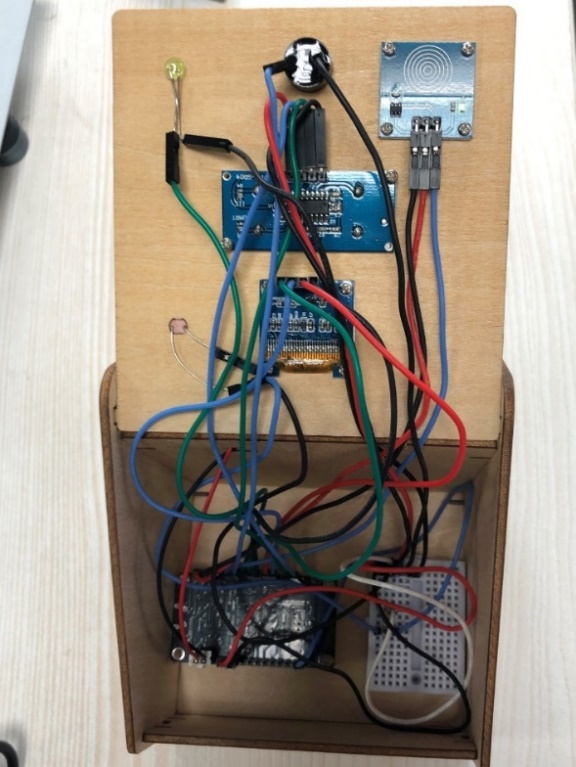


图 5 桌面陪伴机器人实际电路连接图

Ombot单神经元跳跃机器人模块以单个神经元作为机器学习模型，训练特征为机器人起跳时与最近障碍物的距离、最近障碍物的宽度这两个数值，训练标签为“起跳与否”这一决策的判断结果。正负样本分别为用户手动控制的跳跃操作，与根据概率随机采样的不起跳时参数。用户通过触摸按钮控制机器人的跳跃活动，完成一轮游戏后，主控板自动把收集到的参数进行训练，并将训练结果实时展现在OLED屏幕上，最后利用得到的训练结果控制机器人自动跳跃。自动跳跃时用户是无法进行互动的，但是可以观察每轮训练参数的好坏，即自动跳跃时碰撞次数的多少。

网络配置模块为主控板内置的配置功能模块。用户最开始可以操作机器人使其进入esp8266板的AP模式，使得整个机器人变成一个热点，让pc机可以连接这个热点，并将网络配置参数发送给机器人。这样做的优点在于避免了用户通过触摸按钮和OLED屏幕进行输入操作，触摸按钮灵敏度不高，OLED屏幕也非常小，如果需要用它们来输入，则非常麻烦。而通过AP模式和网络通信模块，由专家系统中控软件直接发送局域网的相关配置，例如wifi ssid与wifi密码等，相较而言非常方便。当机器人接收到中控软件发送的网络配置信息后，用户可长按触摸按钮连接wifi，使得软硬件系统接入到同一个局域网内。

专家系统指令接收及操作模块涉及到了软件和硬件。硬件方面，当机器人主动连接到某个局域网后，就成为了一台只能接收指令，而不能发送指令的终端设备了。此时，用户可以操作pc上位机里的专家系统中控软件去编写事实、问题、规则，并推理运行，被出发的规则将会通过socket网络模块直接发送给硬件机器人，并让硬件机器人根据收到的response解析出具体的外设控制指令，例如让led灯亮灭、让蜂鸣器发出声音等。软件方面，成品界面如图 6所示，开发中用了electron框架、vue框架以及typescript语言开发了前端和后端部分，后端中包括了专家系统的推理模块，采用的为最经典的专家系统模型：规则库、事实库、推理引擎。其中规则库和事实库都是由用户输入形成的，推理引擎会在用户启动运行时开始推理，去查询每一条规则是否会被触发，如果被触发则会调用网络通信模块，把规则对应的指令发送给硬件机器人。



图 6 专家系统中控软件界面图

### 语音导购机器人

在互联网的帮助下，人们可以很方便地在网络上购买到自己心仪的产品。自2013年至今，国内网络购物人数与日俱增。根据商务部数据[13], 2013年中国电子商务交易规模突破10万亿元, 网络零售交易额超过1.85万亿元, 占社会消费品零售总额的7.8%, 成为全球第一大网络零售市场。但是在购买某些专业性较强的产品（如：手机、电脑、相机等）时，因为专业知识的缺乏常常会遇到难以选择的情况。这时候如果有导购人员在旁协助用户，不仅能给到用户更好的购买建议，还能为商家尽可能争取到这部分消费者，促成交易。但是，在电商场景下，全天候配备充足的导购员不但会让商家产生巨额的成本还会造成很大的劳动力资源的闲置。

得益于人工智能的发展，许多自然语言处理的任务效果得到了很大的提升。而在自然语言处理的任务中有一个从人工智能诞生之初就被提出了的任务，也就是对话系统[14][15][16][17][18]。在图灵测试中，对话系统被用来作为衡量一个系统是否智能的指标。随着深度学习的普及，对话系统也在得到了一定的发展。对话系统按照功能常常被分为两种类型，一种是和用户闲聊的聊天机器人，另一种则是帮助用户完成某个特定任务的任务型对话系统。比起聊天机器人，任务型对话系统在实际生活中，往往能为用户解决实际问题和需求，从而产生一定的社会价值和经济价值。一些落地的任务型对话系统如Apple Siri、Microsoft Cortana、Google Now、Amazon Echo、Baidu Duer、Facebook M、小米小爱同学、机票预定系统（ATIS、HMIHY）以及出门问问等。

基于此背景，我们提出并设计了专门用于电子产品推荐的导购机器人对话系统，该系统是任务驱动的问答型对话系统，系统能够识别用户的语音输入，并且从多轮对话中提取用户的购买意图，最终从商品知识库中推荐最符合需求的产品给用户。

#### 1.2.2.1 技术方案

本系统使用任务型对话系统的流水线型的框架进行实现，采用前后端分离的技术，前端使用Html实现，后端使用flask框架。本系统采用流水线的框架设计而成，包括自然语言理解模块、对话管理模块和自然语言生成模块三大功能组件。为了识别用户表述中产品的功能属性和体验属性，我们引入了外部知识库，实现功能属性、体验属性到产品配置和产品的映射。本系统的整体架构如图 7所示。在本系统中，自然语言理解模块主要负责解析用户的请求，包括领域识别、意图分类、槽值填充和需求感知四个子任务；对话管理模块主要是对对话状态进行管理并根据当前对话状态采取合适的行为或策略；自然语言生成模块主要是根据上一模块采取的策略进行回复生成。由于对产品的功能属性和体验属性进行识别需要外部知识作为支撑，因此本系统还涉及到知识库的自动构建。主要包括四个重要的模块：

1. 自然语言理解模块： 提取用户输入的语句中与任务相关的关键信息，包括领域识别，意图识别，槽填充任务，以及需求感知。我们会在自然语言理解模块下的领域识别和槽值填充任务中使用深度学习模型。在领域识别任务中，我们使用CNN自动识别用户的意图（购买手机、电脑或者相机）；在槽值填充问题上，我们使用基于BiLSTM+CRF模型的多任务学习框架自动提取句中出现的产品的基本属性。在意图分类任务中，我们使用层级意图控制策略对用户的意图进行有效的管理。
2. 对话管理模块： 在对话管理模块，管理每个回合的输入以及对话历史，我们采用有限状态机管理对话状态。
3. 自然语言生成模块：在自然语言生成模块，我们使用模板生成的方法自动为用户的请求生成回复。
4. 知识库，产品库：在知识库构建过程中，本系统将采用种子词搜索方法分别实现体验属性和功能属性到产品的映射。

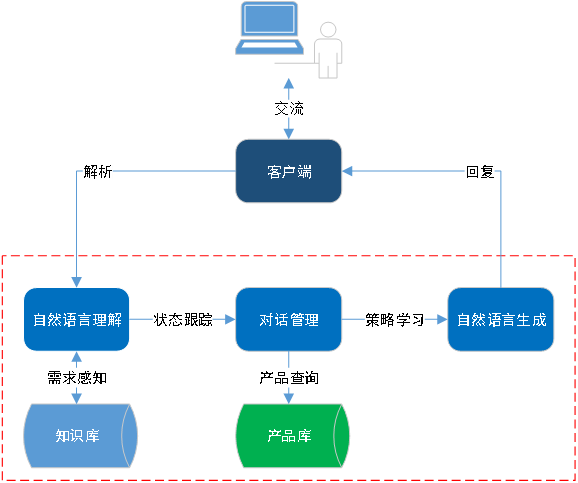


图 7 导购机器人架构示意图

#### 1.2.2.2 算法原理

1. 自然语言理解模块

作为任务型对话系统流水线模型的第一阶段，至关重要。自然语言理解模块的主要功能是从用户的表述中解析出用户意图（如：购买手机）和需求（包括：产品的基本属性，内存大小等；产品的功能属性，能玩游戏；体验属性，散热好），可以细分为4个子任务：1)领域识别；2)意图分类;3) 槽值填充; 4)需求感知。

1. 领域识别

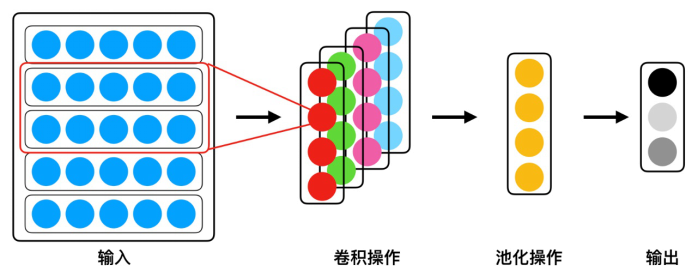


图 8 领域识别模型

在对话交流的开始，机器人需要识别出用户想要咨询的产品类别，这里我们称之为领域识别。领域识别需要识别出用户咨询的是电脑，相机还是手机。在领域识别任务上，我们使用卷积神经网络（Convolutional Neural Networks，CNNs）识别用户当前谈论的产品领域（手机、电脑或相机）, 图 8描述的是自然语言理解模块中的领域识别的模型。例如，“我快开学了，有什么好用的PC推荐下?”,机器人需要识别出该句话是属于电脑领域。卷积神经网络中包括卷积操作以及池化操作，在池化操作之后会通过全连接层以及softmax函数对全连接层的输出进行归一化，从而得到领域标签。

1. 意图识别

在对话过程中，机器人需要识别出用户表达语句的意图，这里我们称之为意图识别。意图识别需要识别出用户的语句的意图是什么，意图包括两部分：1. 操作意图：否定回答，肯定回答，回答槽值，咨询槽值列表；2. 需求意图：需要，不需要，没有要求。在意图识别任务上，我们使用卷积神经网络（Convolutional Neural Networks，CNNs）识别用户当前谈论的语句的意图。

1. 槽填充

对于一个具体的任务，槽位是预先给定的，槽填充任务需要从用户输入的句子中抽取出与具体任务相关的信息填充到槽位中。槽填充任务经常被视为是序列标注问题来解决。序列标注方法对于给定一个句子S = { }，给每个字符 标注一个预定义的标签，最后根据给定的标签提取关键信息到指定的槽位。图 9是一个中文的计算机导购领域下的数据利用序列标注方法实现槽填充任务的例子。

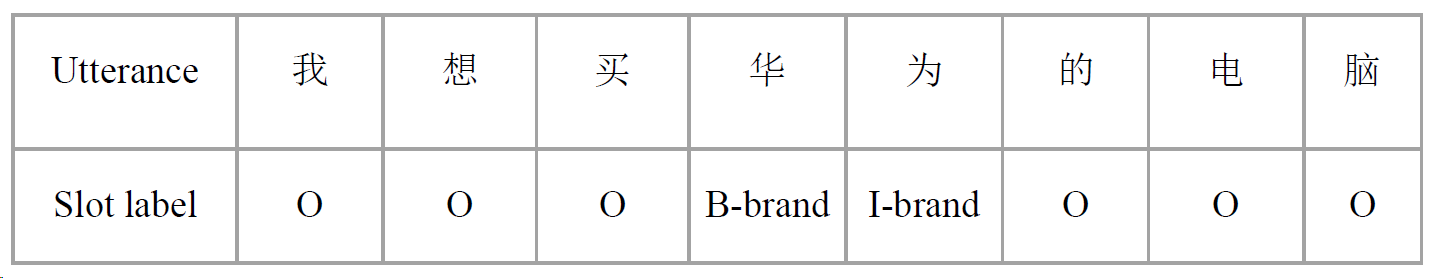


图 9 一个序列标注的例子

这里的“B-brand”表示品牌信息的首字符，“I-brand”表示品牌信息的非首字符，“O”表示非预定义的槽位。根据标签信息，可以提取出电脑的品牌信息“华为”。在本系统中，笔记本类别的有以下槽位标签：品牌（brand），处理器（CPU type），数据接口型号（data interface），硬盘大小（disk size），显卡型号（GPU type），键盘类型（keyboard type），价格上限（price upper），价格中位数（price middle），价格下限（price lower），产品类型（product type），内存大小（rom size），屏幕大小（screen size），材质（shell material），系统（system），重量（weight）。

槽填充的模型整体结构如图 10所示。在领域识别任务中，词序不会对结果造成特别大影响。但是，在自然语言理解中，模型需要识别用户提到的产品的属性，此时，词的顺序会变得十分重要。因此，在这个模型中，我们基于RNNs对模型进行构建。为了缓解梯度消失的问题，我们在RNNs中使用了Long Short-Term Memory（LSTM）。同时，还使用了条件随机场（Conditional Random Field，CRF）进一步处理LSTM的输出，提升识别效果。

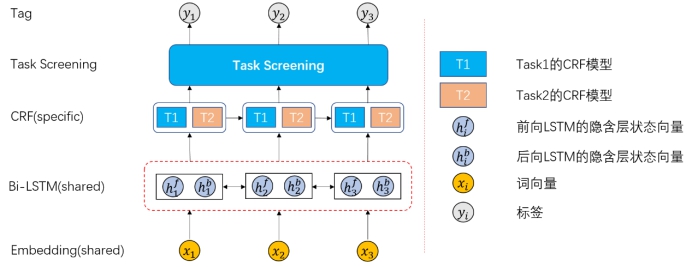


图 10 槽填充模型结构图

1. 需求感知

产品知识库包括产品库和知识库。其中知识库主要用于功能属性的需求感知，产品库主要用于产品查询。在本系统中，我们使用种子词搜索技术对知识库进行构建。产品知识库的主要功能是将经自然语言理解模块解析出的用户表述中的产品功能属性（如：能玩游戏）和体验属性（如：散热好）分别映射到产品的基本属性（显存、内存等属性）和产品上。在本系统中，我们将这一过程称之为需求感知。

1. 对话管理模块

在对话管理模块中，我们使用基于槽特征的有限状态机来对对话状态进行管理。在本系统中，我们将对话过程视为自动机的执行过程。对话策略引起的对话状态的改变，可以看作自动机的状态变化。状态变化如果达到自动机的终止条件，则对话结束。这种方法首先需要设计出自动机的状态，然后根据对话流程确定状态之间的转移关系和转移条件。通过不断调整自动机的结构对对话过程进行控制。图 11本系统中涉及的状态转移。其中init代表系统的初始状态，result是系统的终止状态。

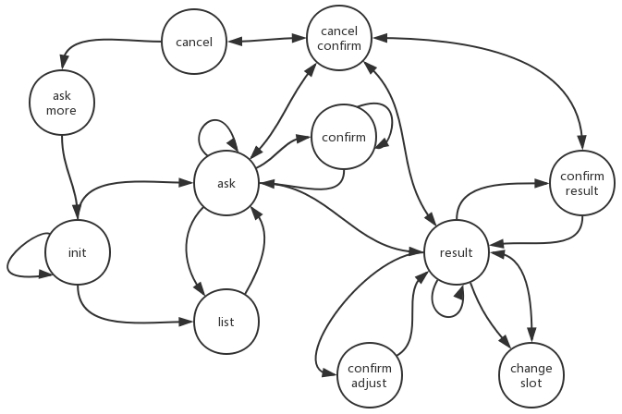


图 11 状态转移图

1. 自然语言生成模块

自然语言生成部件根据接收到的对话指令生成回复。在自然语言生成模块中所使用的回复生成方法，可以是基于模板的生成方法，也可以是语言模型以及最近使用的序列到序列的方法。

#### 1.2.2.3 数据集

我们从领域相关的论坛（电脑、手机）中爬取用户交流的语料，并对这些语料中我们关心的属性（如电脑的内存容量）等信息进行标注。数据分别从相机，计算机，手机领域的论坛网站获取相关的咨询对话，然后再进行人工筛选清洗后标记获得最终的数据集。根据8：1：1的比例将数据集进行划分得到训练集，验证集，测试集。

#### 1.2.2.4 模型评测

自然语言模块的模型识别率很大程度上决定了系统完成导购任务的成功率。本文分别对领域识别，意图识别及槽填充进行评测。评测指标使用广泛应用于信息检索领域的准确度（P），召回率（R）和F1值（F1）。结果如下：

1. 领域识别

表 3 领域识别验证结果

|  |  |  |  |
| --- | --- | --- | --- |
| **label** | **precision** | **recall** | **f1** |
| 整体结果 | 97.57% | 97.46% | 97.51% |

1. 意图识别

表 4 意图识别（操作意图）验证结果

|  |  |  |  |
| --- | --- | --- | --- |
| **label** | **precision** | **recall** | **f1** |
| 整体结果 | 100% | 100% | 100% |

表 5 意图识别（需求意图）验证结果

|  |  |  |  |
| --- | --- | --- | --- |
| **label** | **precision** | **recall** | **f1** |
| 整体结果 | 100% | 100% | 100% |

1. 槽值填充

我们使用多任务槽值填充模型识别用户需求中的语义槽。从结果上看，我们的槽填充模块能取得比单任务槽值填充模型（BiLSTM+CRF）更好的结果。

表 6 多任务槽值填充模型与单任务槽值填充模型实验结果对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 产品领域 | 模型 | P(%) | R(%) | F1(%) |
| Camera | Single-Bi-LSTM-CRF | 86.6 | 96.2 | 91.16 |
| MT-Bi-LSTM-CRF | 91.6 | 97.5 | **94.48** |
| Computer | Single-Bi-LSTM-CRF | 87.2 | 88.7 | 87.93 |
| MT-Bi-LSTM-CRF | 96.5 | 97.3 | **96.90** |
| Phone | Single-Bi-LSTM-CRF | 90.6 | 91.9 | 91.25 |
| MT-Bi-LSTM-CRF | 91.2 | 92.4 | **91.79** |

如上结果所示，我们系统使用的模型的评测指标都达到最高的水平，这保障了我们系统的成功率。

#### 1.2.2.5 系统展示

以下是一个系统的演示，这里使用手机导购作为案例进行演示。可以从以下过程中看出，我们的系统能够流畅地引导整个导购的流程。

1. 初始界面：  
   
2. 手机领域识别成功后，进入手机产品的导购流程：  
   
3. 产品需求识别：  
   
4. 产品推荐：



1. 完成推荐：  
   

## 应用示范

### 虚拟换衣系统

随着互联网的普及,传统的线下购物方式不再是购物的主体,线上购物平台不断发展。近几年来，电子商务的发展势头十分迅猛，服装购物作为电商平台提供的主要服务之一，用户对虚拟试衣的需求越来越迫切。在这个背景下,虚拟换衣技术应运而生,人们足不出户就可以在虚拟试衣平台上体验到不同服装的试穿效果。

现阶段传统的虚拟试衣平台通常存在几个问题： 第一，3D建模成本大，计算复杂，模型推理速度慢。第二，需要用户自己手动输入身高体重等数据，得到的三维人体模型不具有针对性，很难让用户有真实代入感。第三，平台上提供的衣服品类有限，一般不支持用户自己提供的衣服照片。

针对目前传统的虚拟换衣平台所面临的问题和不足，本研究旨在深度学习前沿技术设计和实现一套用户可以独自操作并获得较好的换衣结果的虚拟换衣系统。通过该系统用户可以上传自己想要试穿的衣服照片，利用摄像头捕获人物图像后上传到系统后台进行处理，进行虚拟换衣相关算法操作后返回结果到显示屏幕上，让人们更加便捷的实现虚拟试穿。

#### 1.3.1.1 需求分析

1. 上传想要试穿的衣服。用户可以上传想要试穿的衣服到系统上，或者选择系统提供的参考样例衣服。在用户选择好试穿衣服之后，该衣服会显示在屏幕上。
2. 拍取人像。用户通过本地摄像头进行拍照，将拍取的人像照片上传到系统上，该人像会显示在屏幕上。
3. 虚拟换衣。在获得上述人像和试穿衣服后，输到后台进行算法处理，得到的换衣结果中，人像照片中的人穿着所选择的要试穿的衣服，保留未选择的其余部分，不参与换衣过程。该换衣结果返回到屏幕上。

#### 1.3.1.2 解决方案

本研究的虚拟换衣系统示意图如图 12所示，采集设备主要用于采集人像，显示设备主要用于显示输入和输出结果，客户端用于将采集到的人像上传到系统，以及用户和后台的交互。服务器用于处理模型，包含对获取到的人像基于Openpose[10] 2D 姿态点检测算法进行关键点提取; 对人像利用人体解析算法PGN[11]进行语义分割,得到的分割结果包含手臂、腿、人脸、头发、衣服、裤子等20个分割类别; 根据提取出来的关键点和分割结果对所选择试衣图像进行TPS薄板样条[12]扭曲和人物图像融合。人物图像穿着试穿衣服的结果会被返回于显示屏幕上。



图 12 换衣系统架构图

#### 1.3.1.3 系统详细设计与实现

* 1. web前端

1. 前端实现框架

前端主要使用了react开发框架，同时选用多种工具搭配构建项目，如Vue Router、Vuex和Axios等，主要通过HTML、JavaScript和CSS语言进行开发。前端的组件使用Material-UI组件库。该组件库基于React实现，实现了谷歌Material Design设计规范。Material-UI组件是相互独立的，自支持的，工作时仅注入当前组件所需要的样式。前端页面的跳转与局部刷新完全由路由控制，可以按需加载。多个组件的共享数据由状态管理工具进行同步。前端通过Web API与后端交互，并使用WebSocket实现动态数据交互。前端界面如图 13示。

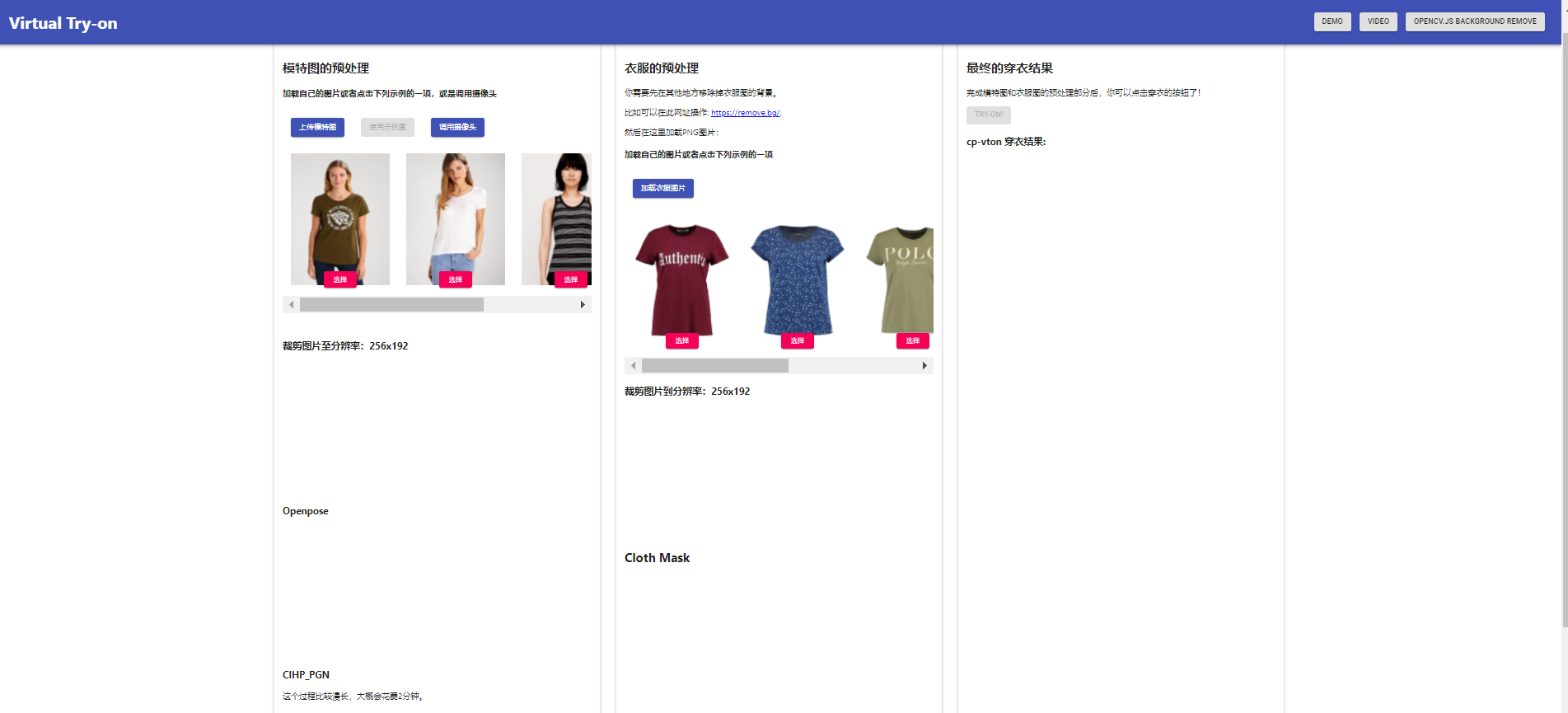


图 13 换衣系统前端界面图

1. 用户操作流程

web前端系统的操作流程如图 14所示，进入页面之后，用户需要选择一种方式提供模特图片，可选的方式包括上传本地模特图片、使用前端系统已有模特图片和使用摄像头拍照。选用摄像头拍照时，需要拍照后确认图片无误后再上传。获得人体照片后，信息会被传递到后台，后台模型进行照片的预处理。第二步中，用户需要选择衣服图片，方式包括上传本地衣服照片和选择系统示例衣服照片。选择完毕后，前端系统可以完成衣服图片的预处理。第三步中，用户点击穿衣后，由后台系统结合前两步的结果，进行虚拟穿衣。前端会随之展示虚拟穿衣结果。

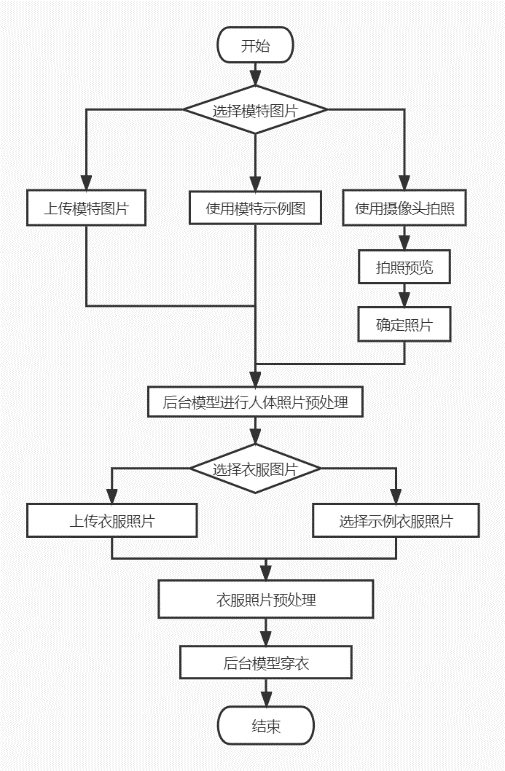


图 14 换衣系统流程图

1. 前端重要功能细节

前端中较为重要的功能包括本地文件上传和调用本地摄像头采集照片。

在本地文件上传中，处理流程为客户端添加事件，创建fileReader实例，用户选择本地图片后，函数获取本地图片路径，根据路径读取文件，读取成功后用axios，在封装的接口中将本地文件上传到后端。

在调用本地摄像头采集照片功能时，首先使用WebRTC API识别设备可用摄像头。WebRTC代表Web Real Time Communication，即网页即时通信的API，这个API能帮助Web应用开发人员通过简单的JavaScript编程就能实现功能丰富的实时多媒体应用。获取用户相机权限后，调用函数进行拍照，采集到拍照结果后，调用canvas标签对应Context的API drawImage将显示摄像头内容绘制到canvas标签页上，并且调用上传接口，将图片上传到后端。

1. 接口设计

本系统的前端通过Web API与后端进行交互，从而实现对后端功能的调用。系统和后台交互的重要接口如表 7 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果所示：

表 7 在图像分类任务与ILSVRC-12数据集上比较不同剪枝方法的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 功能 | 路径 | 参数 | 返回值 | 方法 |
| 上传人体模特图 | /openpose-api | file | rendered\_img,keypoints | post |
| 上传衣服图片 | /cpvton | modelFile  modelParse  cloth  clothMask  pose | warpCloth  warpMask  tryon | post |

* 1. 后台模型处理

1. Openpose提出关键姿态点

为了提取出能代表人体动作的关键骨骼点，本系统利用OpenPose对传输进来的人像进行处理。OpenPose 是由美国卡耐基梅隆大学（CMU）基于卷积神经网络和监督学习算法开发的一个用于检测 2D 图片中人体关键点( 面部，四肢等) 的开源实时系统。OpenPose 区别于传统识别算法的是，研究时使用了自下而上的人体关键点信息识别，即寻找出人体的各个部位，再通过部分亲和域( Part Affinity Fields) 连接各部分。这使得 OpenPose 具有高鲁棒性、可实时输出 的优点。所以本系统选择使用 OpenPose 识别人体关键骨骼点信息。

如图 15所示，12(a)为摄像头拍摄下的人物图像，12(b)为经过OpenPose处理后得到的骨骼节点可视化结果。

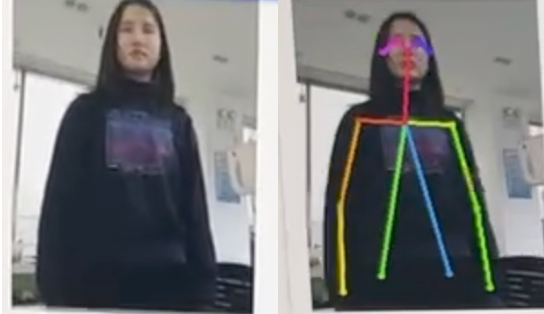


图 15 (a)人物图像 (b)关键骨骼点

1. PGN获得人体语义分割

为了提取出人物图像中不同的区域，本系统采用PGN对传输进来的人像进行分割处理。PGN是中山大学和商汤基于实例级人体解析任务提出的部分分组网络，利用一个统一的网络对两个连续的分割部分进行分组，包括部分级像素分组（part-level pixel-grouping）和实例级部分分组（instance-level part-grouping）。首先，部分级像素分组可以借助以单个像素为一个部分标签的语义部分分割任务得到解决，这个方法学习了类别特征。之后，给定一组独立的语义部分，实例级部分分组可以根据预测的实例感知边缘（instance-aware edge）确定所有部分到底属于哪个实例，其中由实例边缘分隔的部分将被分给不同的人物实例。这种联合优化语义部分分割和实例感知边缘检测的无检测统一网络被称为部分分组网络。由于两个模块通过共享一致的分组目标而保持了相互之间的高度相关性，PGN得以进一步整合了一个细化分支，利用互补的语境信息使两个目标相互受益。所以本系统选择使用 PGN获得人体语义分割结果。如图 16所示，13(a)为摄像头拍摄下的人物图像，13(b)为经过PGN处理后得到的人体语义分割可视化结果。



图 16 (a)人物图像 (b)语义分割结果

1. 虚拟换衣算法
2. 对试穿衣服进行tps扭曲

在得到人物图像中的关键姿态点后，本系统通过将该姿态点输入到网络对姿态进行处理，从而将试穿衣服根据该姿态进行扭曲和变形。本系统采用TPS薄板样条插值算法对试穿衣服进行变换。薄板样条是一种常见的插值方法。因为它一般都是基于2D插值，所以经常用在在图像配准中。在两张图像中找出N个匹配点，应用TPS可以将这N个点形变到对应位置，同时给出整个空间的形变(插值)。薄板样条插值具有数值稳定性好和收敛性好的特性，且能近似曲线拟合复杂的形状，变形精确度和变形效果由特征点选择和特征点的数量决定。所以本系统选择使用 TPS来变换试穿衣服的形态，从而使得试穿衣服和人物图像中人的姿态保持一致，并尽可能的在变换过程中不丢失衣服的纹理色彩信息。如图 17所示， 14(a)为用户所选择的试穿衣服，14(b)为经过TPS插值处理后得到的扭曲衣服的可视化结果。



图 17 (a)试穿衣服 (b)扭曲的试穿衣服

1. 融合图像

通过前面获得的人体语义分割结果，可以去除掉人物图像中原来的衣服区域，保留剩下的外观区域，将扭曲的试穿衣服和人物图像中剩下的区域一起输入到一个U-net网络结构中，分别通过两个分支得到掩膜和换衣的生成图像。此时的生成图像保留了人物图像中不需要参与变换的大部分外观，但是在需要变换的衣服区域得到的纹理会较为粗糙，因此需要生成掩膜进行感兴趣区域提取，根据感兴趣区域替换，将TPS的扭曲试穿衣服与生成图像进行第二次融合，得到最终的结果。如图 18所示，15(a)为人物图像， 15(b)为用户所选择的试穿衣服，5(c)是经过换衣算法生成的最终换衣结果。



图 18 (a)人物图像 (b)试穿衣服 (c)换衣结果

#### 1.3.1.4 结论

本系统利用深度学习技术设计和实现了一套用户可以独自操作并获得较好的换衣结果的虚拟换衣系统，能够满足人们足不出户就可以选择自己想要试穿的衣服进行试穿，得到图像的结果。本系统主要工作与贡献如下：

设计虚拟换衣系统各个功能模块，在此基础上设计系统的总体架构。

设计并实现了web前端。使用WebRTC API识别设备可用摄像头，获取用户相机权限后，调用函数进行拍照，采集到拍照结果后，调用canvas标签对应Context的API drawImage将显示摄像头内容绘制到canvas标签页上。

设计并实现了后台模型。利用OpenPose提出人物图像的关键骨骼点来表示姿态；利用PGN对人物图像进行语义分割来提取出要替换的衣服区域和非替换区域；利用tps 薄板样条插值来扭曲试穿衣服使得变形为和人物一样的姿态；通过图像融合和感兴趣区域提取将扭曲衣服无缝拟合到人物图像身上

本系统仍然存在可以改进和完善的地方：

（1）本系统目前只支持模型处理无背景的模版衣服，这需要用户自行搜索，而网络上更为可见的是穿着衣服的人物图像。

（2）本系统提供的换衣模型对人物图像的尺度和姿势有较为严格的约束，目前无法处理太大或太小的尺度以及十分复杂的人体姿势。

# 总结

本年度主要对已经研究的成果进行实际部署和应用示范，主要从深度学习模型压缩研究进行展开，具有一定的新颖性和理论价值，应用前景十分广阔。同时，对拟人机器人系统进行开发，包括陪伴机器人和语音导购机器人，并落地虚拟换衣系统作为行业应用示范。本年度已经申请软件著作权5项，完成软件系统开发2项，高质量完成了本年度计划。同时整理相关研究资料、已发表成果、文档材料等，完成项目成果鉴定，准备项目验收相关内容。

# 参考文献

1. Liu J, Zhuang B, Zhuang Z, et al. Discrimination-aware network pruning for deep model compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
2. M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi. Xnor-net: Imagenet classification using binary convolutional neural networks. In Proc. Eur. Conf. Comp. Vis., pages 525–542, 2016.
3. C. Tai, T. Xiao, X. Wang, and W. E. Convolutional neural networks with low-rank regularization. In Proc. Int. Conf. Learn. Repren., pages 1–11, 2016.
4. J. M. Alvarez and M. Salzmann. Learning the number of neurons in deep networks. pages 2270–2278, 2016.
5. Y. He, X. Zhang, and J. Sun. Channel pruning for accelerating very deep neural networks. In Proc. IEEE Int. Conf. Comp. Vis., pages 1389–1397, 2017.
6. Z. Zhuang, M. Tan, B. Zhuang, J. Liu, Y. Guo, Q. Wu, J. Huang, and J. Zhu. Discrimination-aware channel pruning for deep neural networks. pages 881–892. 2018.
7. R. Yu, A. Li, C.-F. Chen, J.-H. Lai, V. I. Morariu, X. Han, M. Gao, C.-Y. Lin, and L. S. Davis. Nisp: Pruning networks using neuron importance score propagation. In Proc. IEEE Conf. Comp. Vis. Patt. Recogn., pages 9194–9203, 2018.
8. J.-H. Luo, H. Zhang, H.-Y. Zhou, C.-W. Xie, J. Wu, and W. Lin. Thinet: pruning cnn filters for a thinner net. IEEE Trans. Pattern Anal. Mach. Intell., 2018.
9. Y. He, X. Zhang, and J. Sun. Channel pruning for accelerating very deep neural networks. In Proc. IEEE Int. Conf. Comp. Vis., pages 1389–1397, 2017.
10. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “Open- pose: Realtime multi-person 2d pose estimation using part affin- ity fields,” TPAMI, 2019.
11. Gong K, Liang X, Li Y, et al. Instance-level human parsing via part grouping network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 770-785.
12. Bookstein F L. Principal warps: Thin-plate splines and the decomposition of deformations[J]. IEEE Transactions on pattern analysis and machine intelligence, 1989, 11(6): 567-585.
13. 王宝义.中国电子商务网络零售产业演进、竞争态势及发展趋势[J].中国流通经济,2017,31(04):25-34.
14. 贾熹滨,李让,胡长建,陈军成.智能对话系统研究综述[J].北京工业大学学报,2017,43(09):1344-1356.
15. 陈龙,孙泽健.面向任务的对话系统现状研究[J].电子技术与软件工程,2017(23):172-173.
16. 徐梓翔. 任务型对话系统平台的设计与实现[D].哈尔滨工业大学,2017.
17. POMDP-based statistical spoken dialog systems:A review. Young S,Gasic′M,Thomson B,et al. Proceedings of Tricomm . 2013.
18. Yan Z, Duan N, Chen P, et al. Building Task-Oriented Dialogue Systems for Online Shopping.[C]. national conference on artificial intelligence, 2017: 4618-4626.