基于深度学习的无人机手势控制

王学嘉1，马雨婷2，姜芊鹏3，万瑜丹4

1.哈尔滨工业大学英才学院 2015级 自动化专业；

2.哈尔滨工业大学英才学院 2015级 自动化专业；

3.哈尔滨工业大学英才学院 2015级 自动化专业；

4.哈尔滨工业大学英才学院 2015级 自动化专业；

**摘 要**: 为了更简便的操控无人机，采用手势控制无人机的飞行和操作。针对无人机飞行特点，选取多种手势对应不同的控制命令，采用OpenCV对采集的图像进行预处理，利用Tensorflow深度学习平台搭建神经网络，提取手势特征，并将其系统搭载在无人机上。项目进展至今，完成了对静态手势图像的识别，成功率约为80%。将进一步完善模型，完成对动态手势进行识别，实现在不同环境中手势对无人机的灵敏操控。

**关键词**: 无人机；手势识别；深度学习

Gesture Control of UAV Based on Deep Learning

Wang Xuejia1, Ma Yuting2 , Jiang Qianpeng3, Wan Yudan4

1(Automation, 2015, Honors School, Harbin Institute of Technology)

2(Automation, 2015, Honors School, Harbin Institute of Technology)

3(Automation, 2015, Honors School, Harbin Institute of Technology)

### 4(Automation, 2015, Honors School, Harbin Institute of Technology)

**Abstract**: Using hand gesture to operate UAVs is convenience. For the gesture control of the UAVs, we define several basic gestures first, then we use OpenCV to preprocess the collected images, after that, we build a neural network for feature extraction based on convolutional neural network (CNN) using Tensorflow, and finally we use an on-board computer to carry the whole system. So far, our work on recognition of static hand gesture images has had a high success rate, about 80%. The model will be improved in the future in order that it can do well in different environments, and we will also carry out the experiment for dynamic hand gesture control.

**Key words**: UAV；gesture recognition；deep learning

**引言：**

手势作为一种直观、自然、简洁的肢体语言，是一种很好的人机交互方式，智能机器可以通过识别不同的手势获得不同的指令。一个基于视觉的手势识别系统应该包括图像采集、图像预处理、特征提取和选择、分类器的设计，以及手势识别。

基于单目视觉的手势识别技术主要有三大类：模板匹配技术、统计分析技术和神经网络技术[1]。其中神经网络由于具有自组织和自学习能力，在海量数据中能有效提取特征，成为近几年研究的一大热点。

2012年马克·布朗在Wired英国网站上的一篇文章中提出关于手势引导无人机问题，基于视觉的智能手势识别技术同无人机飞行平台相结合，使无人机可根据人的意图完成不同的飞行。未来，人们将无需进行漫长的专业无人机操控学习，仅仅通过练习几分钟的手势训练就可以控制机器人完成复杂的任务[2]。

无人机行业纷纷展开手势控制研发。如苹果公司在2016年开发的手环，可以通过手势操控无人机。大疆在2016年9月发布的新款无人机Mavic Pro，能够识别拍照手势，有60%以上的成功率。2017年5月中科院沈阳自动化研究所展示成果中，带着橙色手套的操作员可以通过手势指挥无人机飞行，成为国内第一套基于视觉的智能手势控制无人机系统。

基于对深度学习和对无人机的兴趣，我们通过深度学习平台搭建神经网络识别手势，从而做到对无人机高效便捷的操控。

1. 器材选取与工作平台搭建

项目初步阶段采用妙算Manifold作为机载处理器进行图像的处理和识别，搭载在大疆M100飞行器上，利用大疆开发者平台，实现手势识别结果到对无人机飞行姿态及操作命令的映射，从而操控无人机。

1.1Manifold模块

Manifold为 DJI 第一代 On-board SDK 开发平台，配备了高性能嵌入式芯片Tegra K1 作为核心处理器。Manifold自带CAM\_IN/CAM\_OUT 扩展 USB 口，其中CAM\_IN 连接摄像设备。CAM\_OUT 连接经纬 M100 飞控。其预装内置系统为ARM框架下的ubuntu14.04。在此基础上需安装相应框架下的OpenCV、Tensorflow及ROS，其中OpenCV用来处理图像，Tensorflow用来搭建神经网络，ROS用于控制M100无人机及摄像头的运转工作。

1.2 ROS 模块

1.2.1 ROS概述

ROS是实现DJI机载SDK的工具。ROS中开启进程可实现对无人机的控制及通信，可以实现诸如起飞、着陆、姿态控制、拍照、启动/停止录像、本地导航、全局导航、航路点导航等实用性强的功能。

1.2.2 ROS安装

在Ubuntu14.04系统下选择安装ROS Indio版本，此外需要预先安装C、C++编译器及相关开发工具和CMake3.2及更高版本。

1.3具体操作流程

在ROS操作系统下，在M100上连接好摄像头和Manifold，摄像头以每秒10帧采集图像，将所得图像先用OpenCV进行预处理，提取出手势，去除复杂背景，利用在Tensorflow搭建的卷积神经网络，以张量的形式输出手势识别结果，存入文件，再利用ROS读出相应的张量文件，控制飞机的动作，实现通过手势指挥无人机。

1. 图像预处理

由于无人机实际工作环境多变且复杂，为了无人机准确识别手势，需要对拍摄的照片进行预处理。预处理分为以下三个步骤：定位手势位置、识别截取手势图片和减轻背景干扰。

2.1定位手势位置

为了克服人手特征较少和在实际拍摄中手形状轮廓变化大的问题，我们采取一种间接的方法定位。首先进行人脸识别，在拍摄的图片中找到人脸，然后在人脸的特定位置定位手势。

我们的重点在于手势定位，因此不需要通过人脸识别出照片中人物的身份信息，仅需在照片中找到人脸。因此，我们使用OpenCV 提供的物体检测功能。经过训练后即能够检测出任何需要的物体。该库自带了可以相关的检测参数，如人脸、眼睛、嘴。检测引擎由一些非常简单的检测器级联组成。这些检测器被称为 Haar 特征检测器，它们各自具有不同的尺度和权重。在训练阶段，决策树会通过已知的正确或错误的图片进行优化。通过该方法我们可以在照片中找到人的正脸，并用圆圈标记。然后计算圆圈占据的像素点和半径的大小。假设该圆的半径为R，我们在人脸的左侧创建一个边长为2R的正方形区域。只有在这个区域的手势才会被我们截取，即只有在人脸左侧特定的正方形区域内的手势是有效的。这样，我们通过间接的方式模糊的定位了手势[3]。

2.2识别截取手势图片

虽然创建了正方形区域，但是区域内不一定会有手势，因此，我们要对这一区域进行识别，检测这一区域内是否有手势。

我们通过手颜色直方图来匹配，在区域内寻找手势。常用的两种直方图是蓝、绿、红（BGR）直方图和HSV（H：色调，hue；S：饱和度，saturation；V：亮度值，value）直方图。光强不同时，皮肤的颜色差异较大，因此这里BGR直方图不适用。HSV直方图中，色调和饱和度这两个参数几乎不受光强影响，因此我们采用HSV直方图中的H和S这两个参数进行肤色匹配，即H-S肤色直方图。

我们在OpenCV中使用H-S直方图。首先将原图转化为HSV直方图，然后将色调量化为30个等级，饱和度量化为32个等级，再调用calcHist算子进行直方图转化，得到HS直方图。

为了提高识别的准确率，在无人机拍摄前，我们要采集控制无人机人员左右手的照片，通过这些照片获取H-S肤色直方图。然后在手势识别时用这些H-S肤色直方图进行匹配。如果在照片中匹配中率达到50%以上，截取图片。否则判定指定区域内无手势。

2.3减轻背景干扰

经过以上步骤我们截取了含有手势的照片，但手势的背景复杂，不利于我们识别。为克服这一点，我们需要将背景简化，从而使得手势的特征更加明显。我们使用反向投影，即先计算某一特征的直方图模型，然后使用模型去寻找图像中存在的该特征的方法。我们用的模型是提前获取的H-S肤色直方图。用它来检测测试图像中的皮肤区域。。检测步骤如下：

1）对图像中的每个像素（p(i,j)）获取色调数据并找到该色调在直方图中的bin的位置。

2）查询模型直方图中对应bin的数值。

3）将此数值储存在新的反射投影图中。也可以先归一化直方图数值到0-255范围，这样可以直接显示反射投影图像（单一通道图像）。

4）通过对测试图像中的每个像素采用以上步骤，可以得到最终的反射投影图像。

在反向投影中储存的数值代表了测试图像中该像素的属于肤色的概率，亮度高的区域是皮肤区域的概率更大。



图一 手势预处理样图

1. 手势识别

通过建立神经网络，可以识别经过预处理的图像，将手势归类，并输出识别结果。为了取得最好的识别结果，将实验分为三步，分别为：处理静态简单背景的手势图像、处理静态复杂背景的手势图像、处理动态的手势图像。

其中静态手势识别是动态手势识别的基础，比较了基于限制玻尔兹曼机的多层网络和卷积神经网络后，由于卷积神经网络具有权值共享的机制，并且可以将原始图像直接作为网络的输入，大大减少了权值的个数和网络的复杂程度，因此选择了卷积神经网络构造分类算法，下文中将对算法的结构以及项目中的实际训练过程和结果展开介绍。

3.1卷积神经网络结构

利用Google发布的人工智能系统Tensorflow平台搭建完成。共建立一个输入层、两个卷积层、两个池化层、两个全连接层、一个输出层。

3.1.1输入层

卷积神经网络的输入为一组图片***input***，目的为加快训练速度，并且能在一定程度上避免过拟合现象。此处神经网络中的输入为[50,28,28,3]的张量。50为图片数量，28×28为图片大小，3为图片通道数。

3.1.2第一卷积层

卷积层是将图片的张量与一卷积核做卷积操作。第一卷积层的卷积核***weight\_1***为[5，5，3，64]的张量。5×5代表卷积核的大小，3代表卷积核的通道数，64代表卷积核的个数。偏置为一[50，28，28，64]的张量***bias\_1***。经卷积后，得到一[50，28，28，64]的特征图***map\_1***。

3.1.3第一池化层

此处用的是最大池化的方法，池化大小为3×3，每次移动2个像素点，得到一[50,14,14,64]的特征图***map\_pool\_1***。

3.1.4第二卷积层

第二卷积层的卷积核为一[5，5，64，64]的张量***weight\_2***。5×5代表卷积核的大小，第一个64代表卷积核的通道数，第二个64代表卷积核的个数。偏置为一[50，14，14，64]的张量***bias\_2***。经卷积后，得到一[50，14，14，64]的特征图。

3.1.5第二池化层

此处用的是最大池化的方法，池化大小为3×3，每次移动2个像素点，得到一[50,7,7,64]的特征图***map\_pool\_2***。

3.1.6第一全连接层

将[50，7，7，64]的张量变成一个[50，3136]的张量***tensor\_1***。权重张量***weight\_3***大小为[3136,448]。偏置张量***bias\_3***大小为[50，448]，得到一[50,448]的张量***tensor\_2***。

3.1.7第二全连接层

该层权重张量***weight\_4***大小为[448，6]。偏置张量***bias\_4***为[50，6]。得到一[50，6]的张量***X***。

3.1.8输出层（softmax层）

将张量中的三个元素做softmax运算，得一新的[50，6]张量***y*，**此张量中的每个值代表概率。

3.1.9损失计算

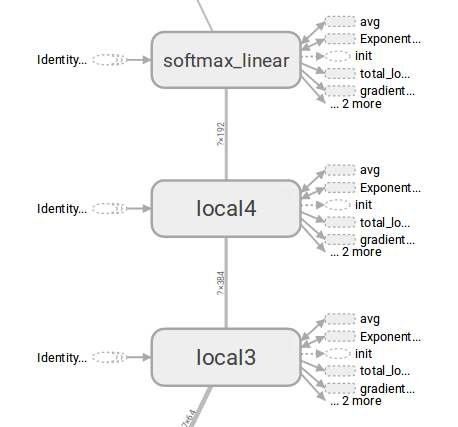
为了训练我们的模型，我们首先需要定义一个指标来评估这个模型是好的。在机器学习中，我们通常定义指标来表示一个模型是坏的，这个指标称为成本（cost）或损失（loss），然后尽量最小化这个指标。此处我们用的是一个成本函数“交叉熵”

***y\_***是我们的实际分布，***y***是神经网络预测的分布。

3.1.10优化算法

反向传播算法可以有效地确定变量如何影响*loss*值，然后用优化算法来不断地修改变量来最小化*loss*。此处采用的是梯度下降算法，一维的梯度下降迭代公式可以写成：

*learn\_rate*为学习率，*learn\_rate*随训练组数*step*增长而下降。



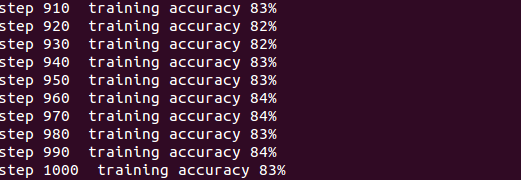
图二 神经网络部分数据流动过程

3.2训练过程

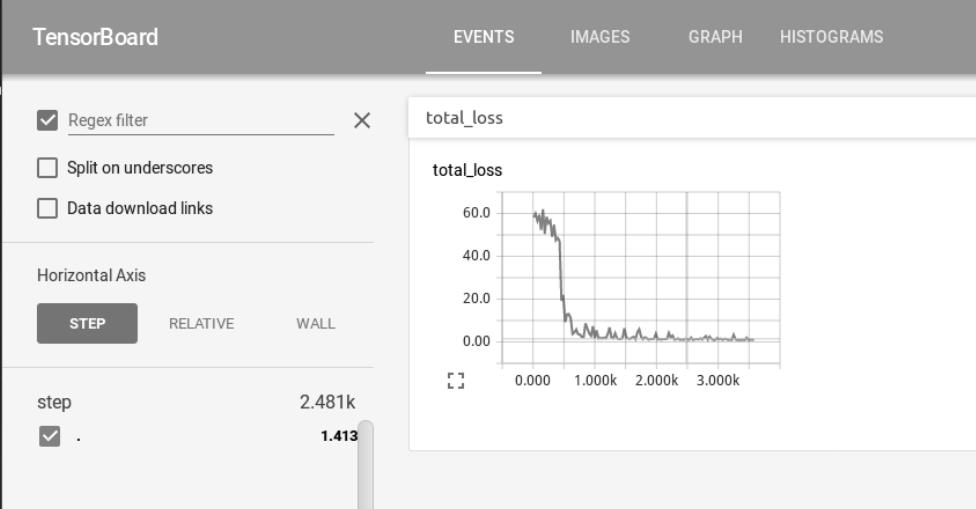
将分好类的图像读入神经网络。训练过程包含3600张图片，会以一个batch（50张图片）为单位提取图片，为降低网络中参数对于图片对比度、拍摄角度、使用左手或右手的敏感度，会在一定范围内随机改变图像的亮度、并对其进行角度变化和翻转。

对于每个batch的图片，以减少loss值为目标，批量化的优化参数。利用Tensorboard显示训练过程中loss变化的折线图，观察loss的变化，调整学习率，保证训练过程中loss的稳步下降，使得图像数据能够被有效利用，且不会出现过拟合。同时利用Tensorboard也可显示数据在网络结构中的完整流动过程。

训练过程完成后，保存训练过程中的参数，即可将其运用在图像的测试过程中，输出手势分类结果。



图三 训练神经网络参数过程



图四 训练初期loss值变化

1. 项目研究成果。

人工控制无人机的位置使其稳定后，可以通过摄像头进行图像截取，当视野中出现静态手势时，能完成识别，并控制无人机飞行，对于摄像头传回的图片，识别正确率接近80%。

**参考文献:**

[1]吴杰. 基于深度学习的手势识别研究[D].电子科技大学,2015.

[2]洪汉梯. 基于深度学习和稀疏表示的手势识别研究[D].西安电子科技大学,2014.

[3]毛星云.OpenCV3编程入门 [M]. 北京：电子工业出版社，2015.