

基于 CNN 的静态手势识别系统

手势识别使用 <https://github.com/asinh33/CNNGestureRecognizer> 的实现。

CNNGestureRecognizer 具有获取摄像头视频流、制作数据集、训练权重和判断手势几个模块，使用 Tensorflow 和 OpenCV。其特点便是允许用户自己快速制作数据集。

1. 图像预处理

为了简化运算，该系统只识别设定 ROI 之中的手势，ROI 的位置可以由用户调整。Binary 处理方法首先将 ROI 数据转换为灰度，高斯模糊之后使用动态阈值进行二值化，适用于空白背景的场所。SkinMask 方法将 ROI 数据转换为 HSV 颜色系统，随后提取类似皮肤颜色的部分作为 mask 以期滤去所有非皮肤部分，适用于光照良好的场景。作者提供了近 4000 张分为五类的图片，用户还可以输入新的手势名称，录制新的手势图片。

数据集是 200×200 的图片，在喂给神经网络之前会被处理为 numpy 数组的形式。

2. 训练 CNN 权重

这里使用的 CNN 模型与用在 MNIST 进行分类的模型差别不大，共 12 层。第一和第三层是卷积层，使用 3×3 的核进行卷积。第二和第四层是 Rectified Linear Unit 函数用来更好更快地提取特征。第五层是 MaxPooling 池化层，使用 2×2 的核对图像进行尺寸和特征上的压缩。第六层和第十层各丢弃 50% 的神经网络单元以加快训练速度，防止过拟合。第七层作为到全连接层的过渡，使用 Flatten 压缩维度。第八层和第十一层分别是有 128 和手势个数个单元的全连接层，最后再使用 softmax 函数将输入分为需要的类型个数。对应 softmax 函数，目标函数设定为 categorical_crossentropy。优化器使用 Adadelta 算法。多分类情形使用 accuracy 指标。

这个静态手势识别相对结构简单，但可以达到可靠的准确度，不过在实际比赛中问题要更为复杂。在比赛中使用的手势识别系统需要具有克服不同光照条件、复杂背景、剧烈抖动和各角度倾斜的健壮性，同时能够准确判断手势位置以确保队员得到安全掩护。由于判断手势需要占用计算能力，该系统也需要合理的暂停和恢复功能。

根据比赛规则。队员应佩戴醒目的手套增强特征，方便手势识别过程的激活与暂停，简化处理量。在掩体后使用手势对无人机下达命令，通过实现动态手势识别传递更多的指令，包括指定无人机、搜索关键部件、提供掩护与治疗等复杂命令。该系统不必安装在所有无人机上，可以在两台无人机中安装进行相互确认。手势识别不像语音识别一样收到现场环境嘈杂的影响，是可靠的信息传递工具。