

# 基于迁移学习的卷积神经网络识别简单人体行为

|  |  |
| --- | --- |
| **姓名** | **王艺博** |
| **院系** | **控制科学与工程学院** |
| **专业班级** | **控制科学与工程硕士2班** |
| **年级** | **2017级** |
| **学号** | **201714403** |

**2017年11月17日**

摘要：许多机器学习和数据挖掘算法都会基于一个假设：训练数据和未来将要处理的数据都处在相同的特征空间，并且有着相同的分布规律。但是，在现实世界的很多应用中，这个假设很可能是不成立的，于是迁移学习应运而生，迁移学习的基本理念是将从一个环境中学到的知识用来帮助新环境中的学习任务，让机器学会“举一反三”。卷积神经网络在图像识别领域占据了极高的地位，取得了非常好的效果，但是在一个数据集上训练的模型却无法直接识别其它数据集的类别。本文先将KTH人体行为数据集的视频进行处理，得到一系列图片构成训练数据和测试数据，然后将在ImageNet训练过并取得不错效果的Google Inception\_v3网络通过迁移学习的方法，使其能够识别KTH人体行为数据集中的简单行为。

Abstract：Many machine learning and data mining algorithms are based on the assumption that training data and future data to be processed are in the same

featurespace and have the same distribution pattern. However, in many real-world applications, this assumption is likely not to be true. Therefore, Migration

Learning emerges as the time requirement. The basic concept of Migration Learning is to use the knowledge learned from one environment to help the learning tasks in a new environment. Machine learning "give top priority." Convolutional neural networks occupy a very high position in the field of image recognition and have achieved very good results, but the models trained on one dataset can not directly identify the other dataset categories. In this paper, the video of KTH human behavior data set is firstly processed to obtain a series of pictures to form training data and test data, and then the Google Inception\_v3 network which has been trained on ImageNet and has achieved good results can be used to identify easy KTH human actions through transter learning.

目录

[摘要 1](#_Toc501389641)

[第一章 绪论 3](#_Toc501389642)

[一、 迁移学习简介 3](#_Toc501389643)

[二、卷积神经网络简介 4](#_Toc501389644)

[第二章 数据集的搭建 4](#_Toc501389645)

[一、 KTH人体行为数据集概览 4](#_Toc501389646)

[二、从视频中提取含有人的图片帧 5](#_Toc501389647)

[第三章 迁移学习的实现 6](#_Toc501389648)

[一、 迁移学习用于卷积神经网络 6](#_Toc501389649)

[二、分析与改进 7](#_Toc501389650)

[三、总结 7](#_Toc501389651)

[参考文献 8](#_Toc501389652)

## 第一章 绪论

无论是传统的机器学习、深度学习还是强化学习，训练一个模型需要的数据是巨大的，在监督学习的背景下，对如此巨大的数据进行标注更是一项巨大的工程，不仅如此，应用如此巨大的数据量从头训练一个机器学习模型不仅是一个花费巨大的问题，而且一旦应用场景发生改变，模型就可能不能泛化。在此背景下，迁移学习应运而生，迁移学习的基本理念是将从一个环境中学到的知识用来帮助新环境中的学习任务，让机器学会“举一反三”。本文将从以下几个方面展开：第一章绪论将会介绍迁移学习和卷积神经网络的发展情况；第二章数据集的搭建将会介绍如何从KTH人体行为数据集中产生我们需要的训练数据和测试数据；第三章将会介绍将迁移学习理念用于卷积神经网络并识别人体行为的方法。

### 迁移学习简介

随着社会发展的信息化和网络化,人们在日常生活和工作中无时无刻不在获取信息,分析信息,并以此作为决策的依据.在一定程度上,信息的拥有量已经成为决定和制约人类社会发展的重要因素.想要高效、准确地寻找到所需的信息,信息分类是必不可少的第一步.通过分类,信息可以得到有效的组织管理,有利于快速、准确地定位信息.分类学习问题,是机器学习中一种重要的学习方法,目前已经得到广泛的研究与发展.

文献[1]指出在传统分类学习中,为了保证训练得到的分类模型具有准确性和高可靠性,都有两个基本的假设:(1) 用于学习的训练样本与新的测试样本满足独立同分布的条件;(2) 必须有足够可利用的训练样本才能学习得到一个好的分类模型.但是,在实际应用中我们发现,这两个条件往往无法满足.首先,随着时间的推移,原先可利用的有标签的样本数据可能变得不可用,与新来的测试样本的分布产生语义、分布上的缺口.比如,股票数据就是很有时效性的数据,利用上月份的训练样本学习得到的模型并不能很好地预测本月份的新样本.另外,有标签的样本数据往往很匮乏,而且很难获得.在 Web 数据挖掘领域,新数据不断涌现,已有的训练样本已经不足以训练得到一个可靠的分类模型,而标注大量的样本又非常费时费力,而且由于人的主观因素容易出错,这就引起了机器学习中另外一个重要问题,如何利用少量的有标签训练样本或者源领域数据,建立一个可靠的模型对目标领域数据进行预测（源领域数据和目标领域数据可以不具有相同的数据分布）.数据分类首先要解决训练集样本抽样问题,如何抽到具有代表性的样本集作为训练集是一个值得研究的重要问题.

Pan 和 Yang[2]针对源领域和目标领域样本是否标注及任务是否相同或是否单一对迁移学习进行了划分。根据源领域和目标领域中是否有标签样本可将迁移学习划分为 3 类:目标领域中有少量标注样本的归纳迁移学习(inductive transfer learning)只有源领域中有标签样本的直推式迁移学习(transductive transfer learning)，以及源领域和目标领域都没有标签样本的无监督迁移学习。另外，还根据源领域中是否有标签样本把归纳迁移学习划分成多任务学习、自学习。Pan 和 Yang[1]还给出了传统机器学习与各种迁移学习情形之间的关系，以及各种情形下，源领域与目标领域是否相同、源领域与目标领域的任务是否相同。迁移学习是和传统学习相对应的一大类学习方式，传统学习处理源领域和目标领域相同且源领域和目标领域的任务是相同的学习，迁移学习处理除此情形之外的学习，包括:源领域和目标领域的任务相关但不同的归纳迁移学习。

### 二、卷积神经网络简介

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。CNNs受到早期的时延神经网络(Time-delay neural networks)的启发，TDNN通过在时间维度上共享权值来降低网络训练过程中的计算复杂度，适用于处理语音信号和时间序列信号。CNNs采用了权值共享网络结构使之更类似于生物神经网络，同时模型的容量可以通过改变网络的深度和广度来调整，对自然图像也具有很强的假设(统计的平稳性和像素的局部相关性)。因此，与每层具有相当大小的全连接网络相比，CNNs能够有效降低网络模型的学习复杂度，具有更少的网络连接数和权值参数，从而更容易训练[3]。

在CNN的一个卷积层中，如图1.1所示，通常包含若干个特征平面(featureMap)，每个特征平面由一些矩形排列的的神经元组成，同一特征平面的神经元共享权值，这里共享的权值就是卷积核。卷积核一般以随机小数矩阵的形式初始化，在网络的训练过程中卷积核将学习得到合理的权值。共享权值（卷积核）带来的直接好处是减少网络各层之间的连接，同时又降低了过拟合的风险。子采样也叫做池化（pooling），通常有均值子采样（mean pooling）和最大值子采样（max pooling）两种形式。子采样可以看作一种特殊的卷积过程。卷积和子采样大大简化了模型复杂度，减少了模型的参数。

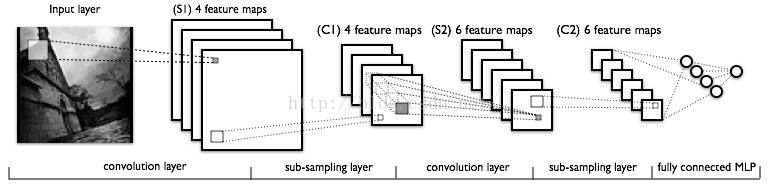


图1.1 卷积神经网络构造[4]

## 第二章 数据集的搭建

### KTH人体行为数据集概览

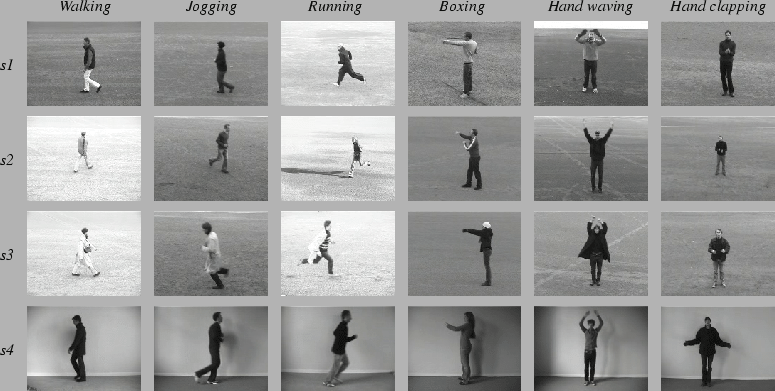
KTH 数据库共包含了 6 种运动，分别是 walk、jogg、run、box、hand-waving、hand-clapping。每 种 动 作 由25 个人完成。数据库的采集是在 4 种不同的场景下进行的， 这 四 个 场 景 分 别 是 outdoors、outdoors with scalevariation、outdoors with different clothes、indoors。 数据库的采集是在静止的背景下进行的。但是，每个动作是在不一样的尺度下和不同的背景下完成的。数据库中所包含的每个视频，完成动作的时间是 2 秒，视频帧的速率是 25 f/s，分辨率为 160×120。KTH 数据库共有 600 个视频片段[5]。数据集动作如图2.1所示。

图2.1 KTH人体行为数据集

### 二、从视频中提取含有人的图片帧

在KTH的视频中，不仅含有人的行为帧，还含有很多无人的背景帧，这就使得我们在搭建数据集的时候不能进行随意抽样，要先提取出含有人的帧图片再进行抽样。具体实现方法如下：

1. 首先对视频进行高斯混合建模提出前景目标，如图2.2所示。

图2.2 高斯混合建模后的帧图片

从图2.2可以看出，高斯混合模型提取出的模型中不仅有我们需要的前景，还包含了很多椒盐噪声，这就使得我们不能够单凭高斯混合建模前景提取后的图片来判断原视频帧是我们需要的含有人的帧还是只有背景的帧。

1. 解决椒盐噪声的问题就要用的开操作，这里我们选用了3\*3的核来对图像帧进行开操作，结果如图2.3所示。

图2.3 开操作后的图像帧

1. 有了前景之后我们就可以提取出只含人的图像帧，这里我用的方法是将图2.3这样的图片进行二值化，然后用二值化后的图片与未经处理的图像帧相乘，得出前景图片，如图2.4所示。

图2.4 提取出的帧图片

最后得到的数据集中训练集每个动作有1400张图片，共六个动作8400张图片，测试集每个动作共有200张图片，共有1200张图片。

## 第三章 迁移学习的实现

众所周知，卷积神经网络随着卷积层数的增加，提取的特征越“高级”，前几层提取的都是“低级”特征，比如小边缘。高级特征是低级特征的组合。所以，已经训练好的卷积神经网络可以根据迁移学习的思想进行“微调（fine-tuning）”来适合我们自己的任务（归纳式迁移学习）。

### 迁移学习用于卷积神经网络

迁移学习在卷积神经网络上的运用是源域和目标域不同，源任务与目标任务也不同的归纳式迁移学习。卷积神经网络的迁移学习方法是将已经训练好的卷积神经网络保留前几个卷积层，使用已经训练好的权重提取特征，最后送入自己定义的分类器进行分类。

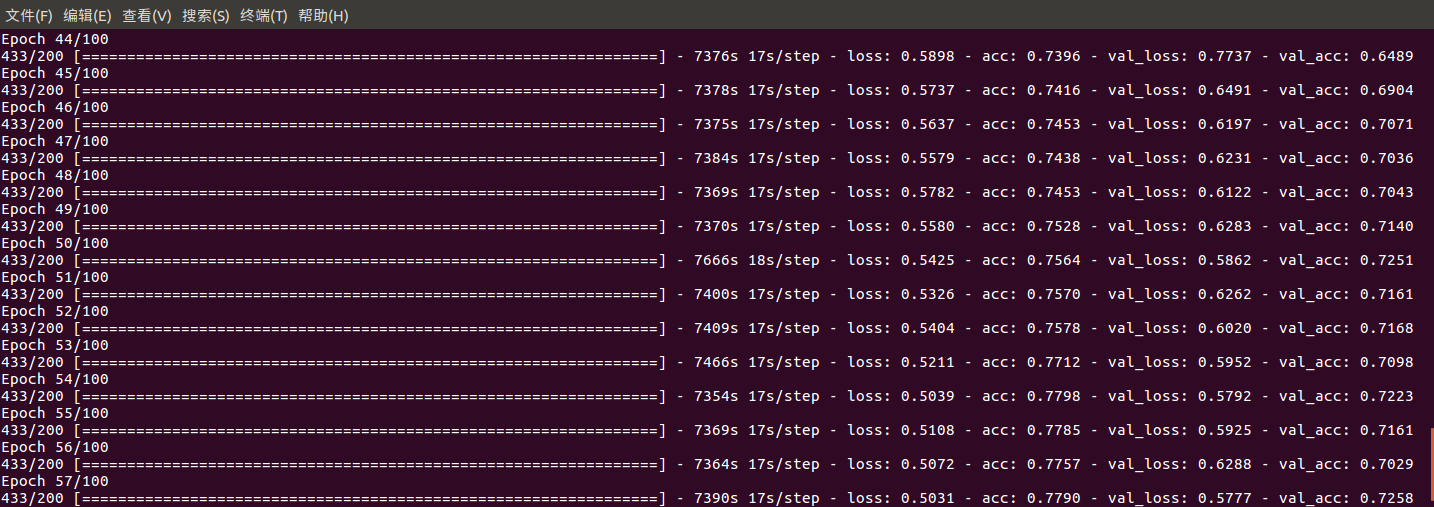
本文采取的预训练模型是Google的Inceotion\_v3网络，本次训练保留网络的前三个卷积块，用自己定义的softmax层进行分类，采用的数据集是上一章中生成的数据集，最后得到的测试结果如图3.1所示。

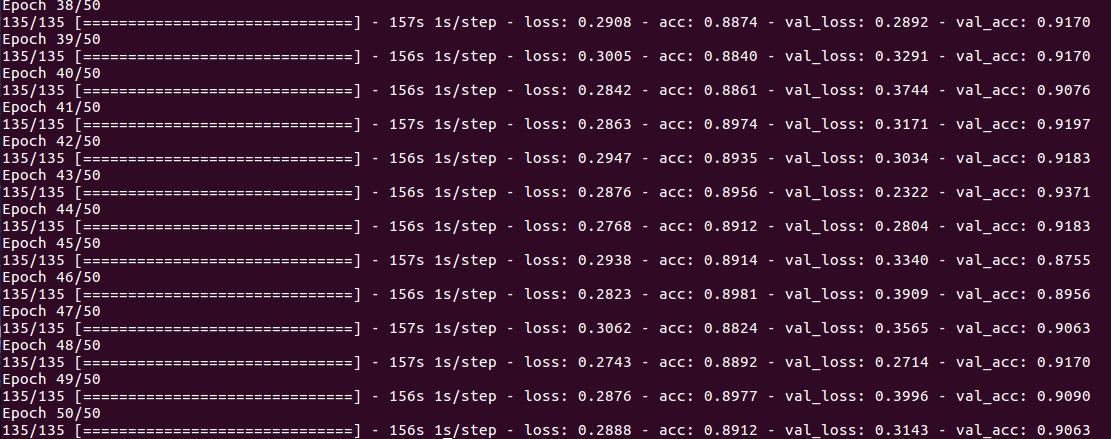
图3.1 基于迁移学习卷积神经网络测试结果

从图3.1看出，在测试集的准确率只有70%左右，是迁移学习的效果不理想吗？下面分析准确率的原因。

### 二、分析与改进

由于动作不仅仅含有静态特征，如图3.2所示，不同的动作从图片中很难分辨，要想准确的识别动作，必须加入动态特征，这也是准确率低的原因，于是，为了验证这一点，我将数据集从六个动作减为四个动作，即跑、挥手、走、挥拳，最终的实验结果如图3.3所示。

图3.2 相似动作的对比

图3.3 改进后的实验结果

### 三、总结

本文尝试了如何从视频中提取需要的图像帧并且应用了迁移学习的方法，虽然最后的实验结果不理想，但是通过总结和改进，实现了对简单动作的识别，并对我下一步的研究提供了经验，希望以后能加深对迁移学习和卷积神经网络的理解，能够实现对复杂动作的识别。

## 参考文献

[1]庄福振, 罗平, 何清,等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1):26-39.

[2]Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359.

[3]卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1):1-17.

[4]图片来源http://blog.csdn.net/yunpiao123456/article/details/52437794

[5]曹博, 刘志刚, 刘文评,等. 人体行为识别的数据库对比研究[J]. 中国科技信息, 2017(17):97-98.