

Deep Judge 项目文档

一、软件概述

现阶段，在竞技体育的各大赛事中，如跳水，体操等，需要有大量的人工裁判对运动员的表现进行打分评判。一方面，人工裁判员不可避免的会出现误判的情况；另一方面，一名富有经验的人工裁判也需要大量的时间，财力去培养。

在大数据时代的背景下，我们可以在各大网站上爬取比赛的录像用于构建数据集。同时，随着深度学习浪潮的来袭，借助 Intel Caffe 我们可以更加容易的训练，测试网络。所以我们打造了一款基于深度学习和 Intel Caffe 的跳水运动打分系统。主要功能有以下三个：

- 一．智能打分：使用 FocNet (深度学习模型) 对用户上传的跳水动作视频给出分数
- 二．躯干识别：使用 PoseNet (深度学习模型) 识别上传视频中的关键点，以便清晰的展示跳水过程，并通过特征可视化来展示打分的可靠性
- 三．动作分析：使用 ActionNet (深度学习模型) 对上传视频进行动作分析，给出整个过程中精彩及不足的地方，以帮助运动员改正

如今，人工智能浪潮汹涌来袭，我们希望将其开创性的引入竞技体育裁判这一领域，因为我们深知人工智能的公正无私正是裁判的最重要的属性。另外，我们也希望有更多的后来者根据我们的思路开发诸如体操等一系列赛事评判系统。

Nowadays, a large number of human judges participate in sports games such as Diving and Gymnastics. However, because of the human's subjectivity caused by intentions or instability, the grading may not correspond to players' performance in one single event. This will unfortunately harm not only players but also countries and nations.

With the help of Intel Caffe, we are committed to building a diving sports scoring evaluation system based on deep learning. Our model is based on 3D convolution and LSTM to achieve the following functions:

- Firstly, we can make a stable and accurate score given a diving video.
- Secondly, we can build a model to recognize the key points of human.
- Thirdly, we will analyze the detail motions of diving events, pointing out the problems of human motions.

Deep Learning has changed our lives. The stability and fairness of machine fit the sports judge well. In addition, we hope that there will be more followers to focus on the sports judging systems, and transfer our idea to other games besides diving.

二、需求分析

（1）背景:跳水赛事的现状

1. 跳水赛事不公平、争议现象

有些不公平现象很明显,在比赛过后都会引起巨大的舆论争论。比如在北京时间 2008 年 8 月 17 日凌晨进行的男子跳水 3 米板比赛中,中国选手彭勃在比赛前四轮大比分领先的情况下进行最后一跳,裁判故意判罚中国队失误,造成金牌流失。又比如在 10 年全运会中某裁判实名揭露跳水金牌全部内定,自己及裁判团被操纵。

2. 培养裁判耗时, 耗费

一名优秀的裁判往往需要数十年时间去培养,在这期间裁判又需要去不停的学习,而且一场比赛往往需要 5-7 名裁判,雇佣裁判的费用往往也是大赛资金中重要的支出。

3. 跳水赛事打分规则

跳水比赛的评判有 7 人及 5 人制,再加 1 个裁判长。但奥运、世界锦标赛和世界杯赛必须有 7 名评判评分。每一个动作的满分为 10 分。评分时,删去最高和最低分,将剩下的分数相加再乘难度分,便得出该动作的分数。

评判会根据运动员的助跑(即行板、跑台)、起跳、空中和入水动作来评定分数。评分主要是由难度分与裁判评分构成,难度分是绝对客观的,但裁判评分完全是由裁判主观打分构成。虽然裁判打分有一定细则,但更多的仍是依赖裁判个人经验。而这种经验不可靠,打分结果能不能服众都有待商榷。

（2）动机:利用深度学习构造打分系统的理由

1. 不公平、有偏差的评分现象

基于上述所说的跳水赛事不公平、评分有偏差、结果难以服众的现象,我们认为基于数据学习的深度学习系统能避免这些弊端。

2. 大数据时代的充沛数据

在大数据时代各种赛事视频可以轻易的被爬取并且标注,我们收集了多达一万个样本用于训练,为深度学习模型的数据来源提供了可靠的支撑。

3. 深度学习的持续发力

在近几年中深度学习持续发力,在计算机视觉的竞赛 ImageNet 中甚至取得了超过人类平均水平的表现。而且深度学习对于视频的处理也是目前 state of the art (最佳) 的表现。我们相信用深度学习来为跳水打分是一条正确的道路。

（3）研究:市场调查和评价结论

我们一方面采访了从事体育方面工作的体育局工作人员和体育老师以及跳水运动员,他们从专业体育的角度向我们介绍了有关情况,并表示很高兴能够有这样一款产品来辅助评判和训练。

同时,我们发放了大约 300 份调查问卷来调查普通观众对跳水赛事的看法,其中 82% 的人听说或了解体育赛事上发生的不公平现象,63% 的人对某些或某一运动员的评分产生过质疑或疑惑,87% 的人对人工智能辅助代替裁判表示欢迎(但也有人质疑人工智能的准确率)。

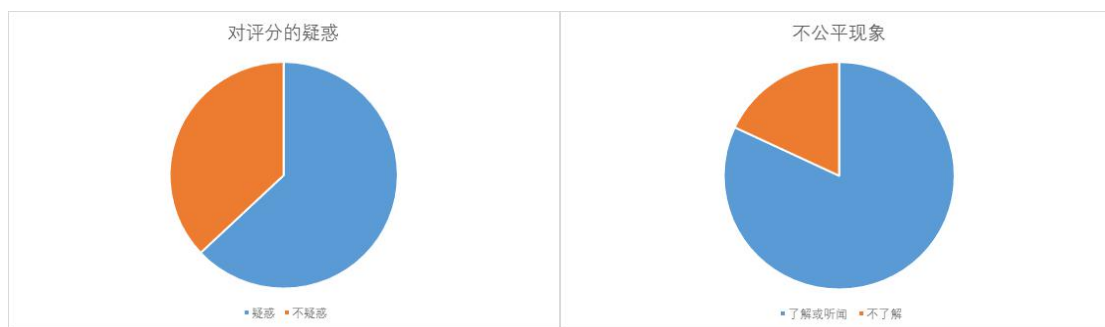


图 1 调查问卷结果统计图

由上述调查结果，我们了解到我们的这款产品是有其实际需求的，不光是对裁判，对运动员，对教练，即便是普通的观众也希望有一款产品来给他们解惑。但是，我们同时注意到了他们对于准确率的要求，这一点我们也会放在第一位考虑。

三、软件功能描述

(1) 呈现形式

目前以网页应用（B/S 架构的 web application）的形式呈现。

界面分为三个主要板块：

- 打分版块：用户给出视频，系统返回分数
- 关键点识别版块：用户给出视频，系统返回识别后的热图视频
- 动作分析版块：用户给出视频，系统以云图方式返回评语。

(2) 主要功能(基于 Intel Caffe)描述

- 打分功能：对上传的跳水视频进行打分。
- 分析动作功能：分析跳水动作的标准与不标准之处。
- 躯干识别：识别出人物的躯干与关键点。

(3) 实用性及可扩展性分析

实用性：

- 可以实际应用在各大大跳水赛事中进行打分。如奥运会、世锦赛等，在这些赛事中可以辅助或代替裁判打分，可以给出打分依据使运动员与观众信服。
- 可以应用在运动员平时的训练中进行。我们的动作分析功能可以向运动员讲述动作的不标准之处，十分适合教练不在身边或者没有教练的情况。

可扩展性：

- 从赛事种类的角度，可以从跳水的某一项目扩展到整个跳水，也可以从跳水扩展到体操，花样游泳等类似的打分项目。
- 从功能的角度，可以扩展出解说功能，为赛事提供完整的一体式服务。
- 从软件开发的角度的度，我们的深度学习模块已经封装成为 docker 镜像，与网页耦合度极低，用户可以自行架构上层封装。

四、技术分析

（1）参赛作品的主要技术路线

我们根据我们的功能，将打造三大深度学习模型：打分系统模型-FocNet、躯干识别系统模型-PoseNet 和动作分析系统-ActionNet。其中模型的构建包括以下几个步骤：构建数据集、使用 Intel Caffe 搭建网络结构、训练模型调整参数、优化网络结构。

- 数据集方面，我们增加保证模型的鲁棒性，用大量而多样的数据训练模型，我们采取数据众包的形式，雇佣 4 名标注采集者来在公网环境下进行数据采集，数据来源是各大视频网站上的赛事视频。共收集 10007 个可用于训练 FocNet 的跳水片段，又对这些样本标注了更加详细的标签来训练 ActionNet。

- 构建模型方面，借助 Intel Caffe 在图像处理方面巨大的优势，以及我们为 Intel Caffe 新增加的用于读取 5 维数组的 python 预处理文件以及用于处理视频的 nd convolutionLayer, nd PoolingLayer 等更加适合处理视频的神经网络层。在构建打分系统模型及动作分析系统模型时节省了大量的时间。

- 训练模型方面，借助 Intel Xeon E5 强大的处理性能以及 Intel Caffe 的优化，我们训练仅仅一周就得到初步结果。这也为优化网络结构奠定了良好的初步结果。

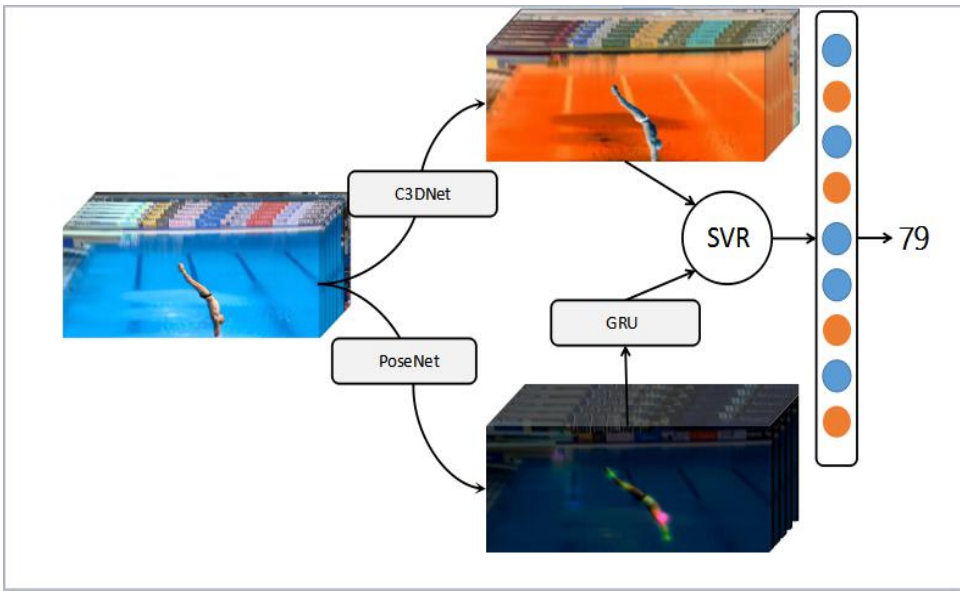


图 2 FocNet 网络结构图

- 优化网络结构方面，由于初步训练的结果已经较为满意，我们选取了 10 组不同的超参数分别训练最后得到了表现最好的一组超参数，最终表现达到人类同等水平 (test 集平均绝对误差上可以达到 0.5，相关系数达到 0.87)。

- 模型表现方面，我们采用了两种指标来衡量网络，分别是平均绝对值误差以及相关系数，平均绝对误差的计算方法是对预测值与真实值之间做差之后求平均，可以刻画打分的精确程度，我们在这项指标上的表现是 0.5，也可以认为对于大规模样本，模型给出的得分与真实值的差距的数学期望不会高于 0.5。另一方面，相关系数是用来衡量排名是否准确的指标，我们在该指标的表现是 0.87。我们选择了 2012 年伦敦奥运会的赛事视频作为测试集，对于奥运会的裁判，我们将用同样的指标对打分的裁判进行了测试，奥运会共有七名裁判，我们对七名裁判分别测试，发现 deepjudge 在平均绝对值误差上的表现超越了大多数单体育裁，位居第二，在相关系数上位居第三。可以认为达到了同等专家的水平。

（2）核心技术关键

核心技术枚举：

1. 跳水打分模型 FocNet
2. 动作分析模型 ActionNet
3. 躯干识别模型 PoseNet

跳水打分模型 FocNet：

整个打分模型是我们产品的关键。打分模型 FocNet 由两部分深度模型后接 SVR 构成。

第一部分模型是 C3DNet，它基于我们加强功能后的 Intel Caffe 实现，在原装 Intel Caffe 上我们编译了新的直接用于读取视频数据以及 3 维的卷积层，这使得我们无需使用分帧和传统的卷积层对帧操作，而是可以用三维卷积直接对四维的数组(blob)进行处理。

第二部分模型是 PoseNet，也即躯干识别模型。由于 PoseNet 可以准确捕捉人体的姿态特征，我们认为特征是对 C3DNet 无法捕捉到的信息的良好弥补。关于 PoseNet 的详细介绍可以详细看本模块第三部分。

对于两部分模型的融合我们采用 SVR，两个模型都是直接使用 ground-truth 训练，输出数值型分数。SVR 的作用在于实现 Ensemble，输入是两个模型的特征层级联后的结果，训练数据就是真实的得分。

动作分析模型 ActionNet：

实现动作分析模型的关键在于数据，我们雇佣了 4 名标注者为 10007 个跳水样本标注了动作分析结果。动作分析结果的标签集是 6 个常用的标签，如：起跳时机过早，水花过大等。每个视频有 4 个标签，所以 ActionNet 转化为一个分类问题，其中我们采用与 FocNet 中一致的 C3DNet 进行特征提取，把最后的 SVR 更换为 SVM 来实现这个模型。

躯干识别模型 PoseNet：

躯干识别模型一方面为打分模型提供了部分特征，另一方面它的特征可视化结果也会在平台上展现。PoseNet 是基于卡内基梅隆大学开源的 OpenPose 进行 fine-tuning(微调)，由于 OpenPose 已经达到了十分精确的结果，我们只标注了少量的跳水视频就取得了非常完美的结果。

（3）实现的难点与解决方案

难点枚举：

1. 数据集数量较少
2. 难以刻画连续跳水特征
3. 跳水动作种类繁多

难点原因及解决方案：

1. 数据集数量问题：

跳水打分问题是一个非常新颖的领域，公网上并没有开源的带标签数据集，因此我们需要自行采集。限于时间，财力，我们只得到了 10007 个良好标注的样本，这对于深度学习来说太过微小，直接训练会导致严重的过拟合，梯度无法正常反向传播，网络无法正常训练。为克服这个问题，我们采用了预训练的方法，引入 sports1M 数据集来预训练网络，该数据集数量超过一百万，是一个运动分类数据集，最终我们成功的利用预训练方法解决了网络难以训练的问题。

2. 难以刻画连续跳水特征：

跳水动作是一个非常复杂的连续动作，运动员处在高速的运动中，CNN, VGG 甚至 ResNet 在捕捉这些特征的时候表现都不尽人意，后来我们采用集成学习的思想，分别用两个网络刻画跳水的特征，C3D 网络刻画时序特征，PoseNet 网络刻画单帧人体信息，最后经 SVR 融合得到最终分数。

3. 跳水种类繁多：

跳水中不同的动作种类有不同的评分标准，即使动作相同得分也可能不同。然而数据本身就十分稀少，如果对每个动作都训练一个模型，将使得数据更加稀疏。因此我们引入多任务学习的方法，在浅层参数上，所有的种类共享参数，在最后的 SVR 上分别训练各个网络，是一种基于参数共享的多任务学习方法，最后有效的提高了网络表现。

五、创新点总结

1. 领域创新：

开创性的将深度学习用于体育赛事打分中来，无论是工业界还是学术界都未有前人做出类似的工作。并且人工收集了第一个大型赛事打分的数据集，为后来的研究者提供了方便和开创性的思路。同时也解决现实中存在的打分不公平这个严重问题。

2. 算法创新：

基于我们对 Intel Caffe 的增强，我们的模型融合了两种主流方法的优点，实现了超越传统方法 (iDT) 及深层方法 (two stream) 的表现，同时模型的速度也得到了大幅度的优化。

五、产品团队组成与角色分工

LYQ（功能实现与测试）

DCX（深度学习相关）

LJY（功能实现与测试）

LMY（深度学习相关）

六、项目时间进度表

项目重要里程碑	完成日期
Intel Caffe 的搭建	2017/5/21
跳水数据集的构建	2017/7/10
网页的 UI 设计	2017/7/15
深度学习模型的初步搭建	2017/8/03

模型调试完成	2017/8/20
网页前端后台整合	2017/8/25
各个服务融合调通	2017/8/29