

# 基于计算机视觉技术的排球综合分析系统的搭建研究

## An Integrated Volleyball Analyzing System Based on Computer Vision

### Abstract

We are planning to develop a volleyball analysis system that integrates action recognition, athlete re-identification, and highlight videoes. After preliminary research, we plan to use the *Segment Anything* model to preprocess the match videos and accurately segment the contours of the athletes. Subsequently, we will use the *Multisports* dataset to train a network that can recognize athlete actions, and then use downstream modules to achieve the above functions. In terms of specific neural network selection, we have focused on researching the Resnet model and the Transformer model. Considering that Transformer can learn global features, we finally chose it as our training neural network. After preprocessing the dataset, we organized it into a specified directory, and then used the MMAction2 network to train it, achieving good results. Our follow-up work includes adapting the input of Multisports to the output of Segment Anything, researching relevant downstream modules, and increasing the amount of dataset training.

### 一、前期调研

#### 1.1 计算机视觉在排球应用现状

不同于体操的AI打分已经用于奥运会这一现状，排球领域中计算机视觉的应用并不明显。市面上的分析公司，如Hudl可以实现位于特定时间下的动作识别，如图1-1。由于其作用不够全面，识别精准度也有待提高。因此，我们希望搭建一个综合的排球分析系统，实现实时对运动员的动作识别、比赛走向的分析、赛后精彩集锦等多项功能合一的排球分析系统搭建，这样可以满足排球运动员、教练、球赛观众等多方需求，实用价值较大。

#### 1.2 Segment Anything与Multisports调研

在计算机视觉应用于比赛领域时，运动员识别精准度是一大难题。若用边界框识别，由于其包含背景内容，结果不够精确。Kirillov等人研发了Segment Anything模型，可以较为精确地切割出图片中物体的边框。其被应用于运动员识别的效果也较好（如图1-2）。因此，我们打算使用该模型进行运动员识别。

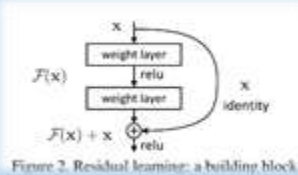


经过初步调研，我们决定将Multisports作为我们的重点调研对象。Multisports是一个多人运动动作识别的数据集，其领域涉及篮球、排球、足球等多个领域。通过调研Multisports的相关数据，我们认为其可以作为我们的训练神经网络的数据集。Multisports为已标注的逐帧数据集，方便训练。同时，Multisports作为较新的数据集，采用MMAction2神经网络进行训练，其准确度也较高，如图1-1。另外，需要注意的一点是，Multisports使用边界框识别运动员，而这这就要求我们在后续需要对其代码进行更改，使其适应Segment Anything的输出。

### 二、神经网络模型的选择

#### 2.1 Resnet模型

该网络的优势主要体现在其采用的残差连接方法可以有效解决大多数网络达到一定深度后的梯度消失问题。而该系统需要实现对排球比赛球员的高精确度跟踪和动作识别，利用resnet可提高图像识别的性能。

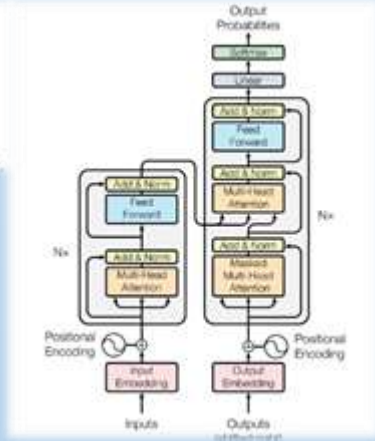


#### 2.2 Transformer模型

另一个同样值得注意的网络是transformer。该网络只采用注意力机制以实现识别局部与全局关系的功能，可以准确提取出输入中的关键内容。且该网络模型可以并行计算，处理长序列基本不会丢失信息，与其他网络模型相比训练时间更短。该网络最初运用于翻译领域，而在2021年Dosovitskiy 和 Beyer 发表论文ViT，将transformer模型首次运用于图像分类，经过训练后在分类任务中同样表现出色。

#### 2.3 优劣对比及模型选择

我们最后选择了Transformer作为排球综合分析系统的主要模型。ResNet的卷积操作只能学习到局部的特征表示，而Transformer的自注意力机制可以学习到全局的特征表示，因而在分析比赛视频的过程中更不容易丢失主要信息，对于动作的打分或是比赛的评估等下游任务的准确度会更高。



### 三、数据集的训练

整理好数据集后，在服务器上创建虚拟环境，并按照如下列表搭建环境：

```
System environment:
python: 3.9.0 (default, Nov 15 2020, 14:28:50) [GCC 7.3.0]
CUDA available: True
numpy.random.seed: 133023599
GPU 0: 0, 1, 2, 3, 4, 5, 6, 7: NVIDIA GeForce RTX 3090
CUDA_HOME: /usr/local/cuda
NVCC: Cuda compilation tools, release 11.3, V11.3.140
GCC: gcc (Ubuntu 9.4.0-2ubuntu1~20.04.1) 9.4.0
PyTorch: 1.10.1+cu113
PyTorch compiling details: PyTorch built with:
  - GCC 9.2
  - C++ Version: 101402
  - Intel(R) Math Kernel Library Version: 2020.0.0 Product Build 20191122 for Intel(R) 64 architecture applications
  - Intel(R) MKL-ML 2020.0.0 (64-bit) Build: 20201103h1099/0aa1b04677304f049810
  - OpenMP 201511, 0.0.0, 0.0.0, 0.0.0
  - LAPACK is enabled (usually provided by MKL)
  - NNPACK is enabled
  - CPU capability usage: AVX2
  - CUDA Runtime 11.3
```

上传数据集，选用Resnet50网络和kinetic400经过多轮训练，得到的最终准确率为。预训练模型，按照训练集：测试集=1:9进行训练，

### 四、后续工作

#### 4.1 修改Multisports的输入数据格式

在利用Multisports训练出可识别运动员动作的网络后，由于其利用边界框来识别运动员，而这与SAM输出的运动员的切割图并不相符。因此，为了实现更加精准的识别，我们将修改Multisports的输入部分，使其适应SAM对运动员的切割结果。

#### 4.2 调研相应的下游模块

我们设想通过下游模块来实现对排球运动员的动作打分，对比赛的实时评估、对选手表现的分析等工作。为此，我们还需要进一步对下游模块进行调研，将Multisports输出的动作结果等各项数据进一步分析，得到最终的输出结果。

#### 4.3 增加数据集训练量

如前文所述，我们采用Transformer作为神经网络模型，需要大量的数据集进行支撑。因此我们应进一步调研排球相关数据集。