**基于BERT-IT的世界杯球队图像识别模型**

马龙 2020110624

摘 要：**在近年来，图像识别技术已经成为计算机视觉领域的一个热门话题，并且在各个领域都有着广泛的应用。图像识别技术可以帮助我们在海量的图像数据中快速找到我们感兴趣的内容，并且可以通过对图像的分析来获取有价值的信息。例如，在医学领域，图像识别技术可以帮助医生快速找到病人的疾病；在交通领域，图像识别技术可以帮助我们快速识别交通标志；在自然语言处理领域，图像识别技术可以帮助我们快速识别文字。基于这些原因，本人开发了一个基于 BERT-IT的世界杯球队图像识别模型，可以快速准确地识别世界杯各球队球员的图像，通过球员的球衣特征来识别具体的球队名称，为球迷提供便捷的服务。并且我希望这个模型可以在进一步提升图像识别技术的同时，能够拓展到其他体育赛事中，进而让图像识别涉及到更多领域。**

关键词：**BEIT、Bert、Transformer、图像识别、世界杯**

# 研究背景及意义

## 1.1研究背景

在过去的几十年中，图像识别技术一直是计算机视觉领域的一个热门话题。早期的图像识别算法主要基于手工设计的特征，例如纹理、形状和边缘信息。随着深度学习技术的发展，图像识别算法也逐渐向基于深度神经网络的方法转变。这些方法能够从大量的训练数据中自动学习有用的特征，并且在各种不同的应用中取得了显著的成功。

在当今社会中，图像识别技术已经广泛应用于各种领域。例如，在自动驾驶领域，图像识别技术被用来识别道路、信号灯和障碍物；在医学领域，图像识别技术被用来诊断疾病；在商业领域，图像识别技术被用来识别产品和品牌。

本研究旨在开发一种基于图像识别技术的世界杯求队图像识别模型。我们将基于著名的BERT-IT模型，通过对大量世界杯求队图像进行训练，来实现对世界杯球队图像的自动识别。我们针对这一研究目标，提出了以下两个研究问题：

问题1：能否利用图像识别技术，准确识别世界杯各队球员的图像？

问题2：在识别准确率方面，哪种图像识别模型表现最优？

## 1.2项目意义

本项目的意义在于，图像识别技术在当今社会已经起到了很大的作用，而在体育界的应用尤为明显，例如在世界杯这种大型赛事中，球队的图像识别可以帮助球迷更好地认识和记住球队和球员，而本项目就是为了更好地解决这个问题而设计的。同时，本项目还可以为图像识别技术的发展做出贡献，为其在体育界的应用提供新的思路和方法。

另外，该模型具有较高的准确率和效率，能够快速识别世界杯的球队和球员，满足社会对于快速获取信息的需求。并且采用了最新的图像识别技术，如Beit模型，对现有图像识别理论进行实践。

## 1.3图像识别在体育领域的应用

在这届世界杯中，国际足联引入了可视化审判系统（VAR）作为判定依据。通过使用VAR，裁判员可以在判决关键点球员越位、犯规、手球等情况时，获得帮助。同时，VAR还可以帮助裁判员在判决关键进球时是否有效、球是否越过线界等情况。通过使用VAR，裁判员可以更准确、更公正地判断比赛中的关键决策。

在本项目中，我们也使用了图像识别技术来帮助我们识别世界杯各队的球员。通过使用这种技术，我们可以在比赛中快速、准确地确定球员的身份，为比赛和球员的管理提供便利。同时，这也为球迷和足球爱好者提供了一种新的、有趣的方式来了解世界杯各队的球员。

# 研究方法与实验设计

## 2.1研究方法

**项目目标**

本项目旨在通过图像识别技术来识别2022年世界杯球队的球员，我们的目标是通过12支队伍的球员及教练图像，球员识别为Argentina, Belgium, Brazil, Croatia, England, France, Germany, Japan, Poland, Portugal, Spain, Uruguay这12只队伍以及Coach（教练组）和Goalkeeper（守门员）共14个类别。通过识别这些球员，我们希望能够更加精确地了解每支队伍的球员组成，并为后续的比赛分析和预测提供参考。

**数据选取**

我们的训练数据集选取自FIFA官网，包含了来自Argentina, Belgium, Brazil, Croatia, England, France, Germany, Japan, Poland, Portugal, Spain, Uruguay这12只队伍的球员图像，以及Coach（教练组）和Goalkeeper（守门员）的图像。接着将这些图像按照所属的队伍划分到不同的文件夹中，并从中选取出Messi、C.Ronaldo、Neymar等球员图像作为测试集（已经提前从训练集中剥离出），剩余的图像则作为训练集。通过这种方式，可以利用训练集来训练模型，并利用测试集来预测模型的效果。

**模型选择**

在图像识别领域，常用的模型包括传统的机器学习方法，如SVM、决策树和KNN，以及基于神经网络的深度学习模型，如CNN、RNN和GAN。近年来，还有一些基于自监督学习的模型，如BERT-IT在图像识别任务中表现出色。Beit是一种基于自监督学习的图像识别模型，它通过自动学习图像的内在特征来识别图像。

在训练过程中，BERT-IT不需要人工标注的数据，而是通过自监督学习的方式自动学习图像的内在特征。这使得Beit+Transformer在图像识别任务中具有更好的泛化能力，能够准确识别出更多的图像。此外，BEIT还具有较快的训练速度和较少的训练数据要求，因此选择它作为我们图像识别的核心模型。

## 2.2实验设计

本实验旨在通过图像识别方法来对世界杯的球员进行分类。首先，选择训练数据集，使用FIFA官网的球员图像。通过将图像与图像所属的文件夹绑定，我们可以自动地提取标签。接着，使用BERT-IT图像识别模型对数据进行训练。最后，我们使用测试集（从训练集中剥离出来的球员图像）来预测模型的效果。我们希望通过这些步骤，能够得到一个很高效的世界杯球队的图像识别模型。

## 2.3模型介绍

BERT Pre-Training of Image Transformers（简称BERT-IT）是一种用于图像预训练的模型，它结合了自然语言处理中的BERT模型和计算机视觉中的transformer模型的思想，旨在对图像进行预训练，以便在更多的视觉任务中进行转移学习。

BERT-IT的结构与BERT模型非常相似，由若干层的双向的多头注意力机制组成。不同的是，BERT-IT在输入层使用了卷积神经网络（CNN）来提取图像的特征，而BERT模型则使用词嵌入提取文本的特征。BERT-IT使用的是图像中的像素值作为输入，而BERT模型使用的是文本中的词作为输入。

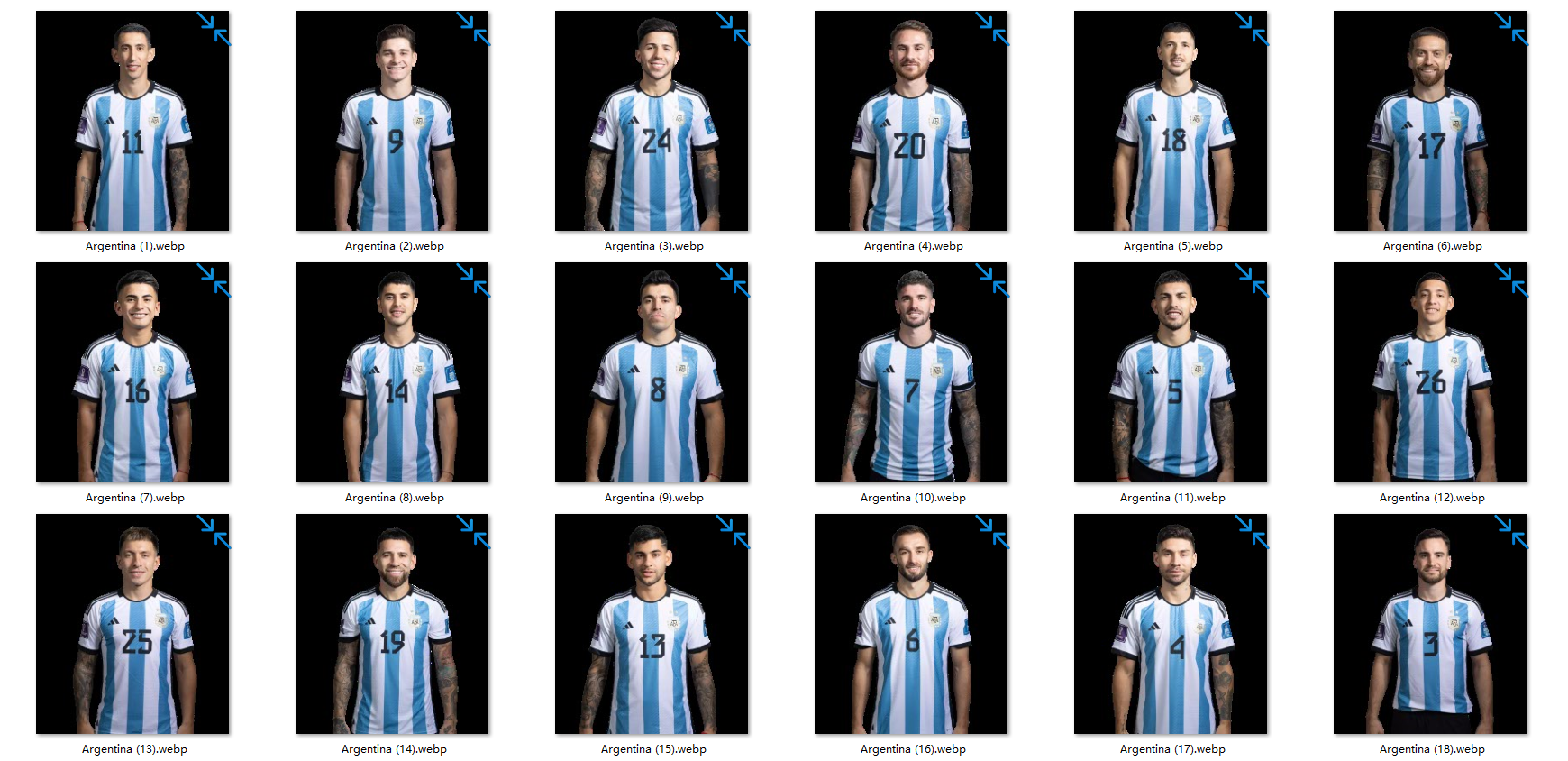
BERT-IT通过使用注意力机制，可以从图像中提取有用的信息，并将其转换为输出。BERT-IT预训练的目的是希望在预训练的过程中学习到一些通用的图像特征，这些特征可以在转移学习的过程中被重用。

# 实验过程

## 3.1前期准备

**数据集介绍**

球员图像数据集源于FIFA官网TEAM板块。训练集位于team文件夹下，该文件夹下拥有Argentina, Belgium, Brazil, Croatia, England, France, Germany, Japan, Poland, Portugal, Spain, Uruguay这12只队伍以及Coach（教练组）和Goalkeeper（守门员）这14个子文件夹，对应着14种分类。也就是说对于每一张球员图像，它所处的文件夹即为训练集label。如下图所示:



**文件介绍**

**code.ipynb**:项目主文件，包括模型、模型参数、预测结果

**team**:训练集文件夹，存放所有预训练球员图像信息

**test**:测试集文件夹，存放待预测球员图像信息

## 3.2实验步骤

**步骤一 导入项目核心依赖**

**torch**:核心依赖，包括nn神经网络模型，求解器，各种损失函数以及数据加载器DataLoader

**cv2**: 处理图像和视频的开源计算机视觉库，包括图像的读取、写入、显示、转换、图像处理等。

**PIL**: 处理图像的Python库，它提供了许多用于读取、写入、显示和处理图像的功能，包括图像缩放、旋转、裁剪、转换格式等。

**transforms**: 图像处理库，本使用中主要运用特征提取、预处理模型加载以及求解器的配置。

**numpy**:科学计算库，本实验中主要用于数据预处理

**tqdb**:深度学习模型中，监测模型训练进度的库

**步骤二 数据预处理**

在数据预处理中，我们首先使用通过CNN（卷积神经网络）来对预训练的模型提取特征，然后通过使用BeitForImageClassification创建模型并将其初始化。其中，预处理模型选自微软的microsoft/beit-base-patch16-224-pt22k-ft22k。接着，我们创建了一个自定义数据集类MyDataset如下：

class MyDataset(torch.utils.data.Dataset):

 def \_\_init\_\_(self, images, labels):

        self.data = feature\_extractor(images=images, return\_tensors="pt")

        self.data['labels'] = torch.LongTensor(labels)

        self.labels = torch.LongTensor(labels)

    def \_\_len\_\_(self):

        return len(self.labels)

    def \_\_getitem\_\_(self, index):

        return {

            'pixel\_values' : torch.FloatTensor(self.data['pixel\_values'][index]),

            'labels' : self.labels[index]

        }

该类继承自torch工具模块Dataset，通过实现getitem和len方法来实现MyDataset模块的封装，便于批量处理。接着将球队以及教练组、守门员进行离散化，采用0-13之间的整数作为离散结果，方便模型计算。

然后读取team中的球员图像及分类加载到images、labels中，接着将它们作为MyDataset参数初始化train\_dataset。最后，我们使用DataLoader创建了训练数据集train\_dataset的加载器train\_loader，该加载器将使用批量大小为16的训练数据集进行随机排列。

**步骤三 模型训练**

开始配置求解器，损失函数、激活函数以及进度查看器。求解器采用transforms模块自带的AdamW优化器来帮助我们训练模型，具体使用方法是将model.parameters作为参数代入优化器作为求解器。同时我们还使用了线性学习率调度器来调整学习率，并对运行环境进行检测，如果运行环境当中具有英伟达GPU资源，则调为cuda模式，将模型的执行以及损失函数的计算都放在GPU当中来进行。

以下是模型的训练过程：

model.train()

for epoch in range(epochs):

    for idx, batch in enumerate(train\_loader):

        outputs = model(\*\*batch)

        loss = outputs.loss

        loss.backward()

optimizer.step()

        lr\_scheduler.step()

        optimizer.zero\_grad()

        progress\_bar.update(1)

首先将模型置于训练状态，模型的参数会随着不断的训练进行更新。接着通过设置的轮数来进行迭代，将每一批数据代入模型开始训练。由于模型的outputs含有loss属性，故可以根据该loss来进行反向传播计算梯度。然后通过optimizer.step()更新参数同时更新学习率，让其随训练批数增加而减少。最后将梯度置零，并在进度条进行更新。

**步骤四 模型预测**

项目的预测部分，我们将使用模型对球员图像进行预测，为了更好地观察模型预测的效果，我们还将预测结果在图像上进行可视化，以便观察。首先，将模型设置为评估模式，模型的参数不会随着模型的执行而进行更新。然后通过matplotlib创建一个图像绘制载体。接着，遍历测试集目录中的所有图像，在测试集上运行模型，并使用输出的logits预测图像类别，可视化测试图像和预测结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **Argentina** | **Portugal** | **Brazil** | **Poland** | **Croatia** |
|  |  |  |  |  |
| **Germany** | **Uruguay** | **France** | **Belgium** | **Goalkeeper** |
|  |  |  |  |  |
| **France** | **England** | **Spain** | **Japan** | **Coach** |

基于以上预测数据我们可以发现，该模型不仅准确预测出所有球员所属的国家队，还对于队服相近的国家队例如葡萄牙、西班牙、比利时都有比较良好的识别效果，特别是Goalkeeper（守门员）在训练数据时就具有不同颜色的球衣，但是仍然能够很高效的被识别出来。我们回到本文最初的问题之一：能否利用图像识别技术，准确识别世界杯各队球员的图像？我想我们的模型已经交上了一份完美的答案。

# 项目分析与展望

## 4.1项目创新点

1)采用BERT-IT模型进行图像识别：BERT-IT是一种非常有效的图像预训练模型，通过结合BERT模型和transformer模型的思想，可以在图像预训练的过程中学习到一些通用的图像特征。使用BERT-IT模型可以提高图像识别的准确率，并且在转移学习的过程中可以很好地重用之前学到的信息。

2)应用于世界杯球队图像识别：你的项目是针对世界杯球队图像识别的应用，这可以作为项目的一个创新点。你的模型将能够识别世界杯各个球队的队徽或球员照片，这可以为足球迷提供方便，并且在相关领域也具有一定的市场潜力。

3)多轮训练和分批加载：在训练过程中，我们采用了多轮训练和分批加载的方式，这可以提高训练的效率，减少训练时间。

4)学习率调整机制：我们在训练过程中使用了学习率调整机制，这可以帮助我们调整模型的学习速率，使得模型能够更快地收敛。

5)可观测模型进度：我们的项目还使用了可观测模型进度的功能，这可以让我们更好地了解模型的训练情况，并及时调整训练策略。

## 4.2项目不足之处

1)缺少GPU资源：我们的项目缺少GPU资源，这可能导致训练时间较长，影响项目的效率。

2)没有进行性能比较：我们没有对多种模型进行性能比较，无法确定我们所使用的模型是否是最优的选择。这可能会导致模型的准确率不够高，影响项目的效果。

## 4.3项目展望

我们的项目是一个基于BERT-IT的世界杯球队图像识别模型，它具有高效的图像识别能力，并且应用于特定领域，但同时也存在缺少GPU资源以及没有进行多种模型性能比较的不足之处。

在未来，我们可以考虑以下方向来改善项目：

1)获取更多的GPU资源：通过获取更多的GPU资源，我们可以缩短训练时间，提高项目的效率。

2)进行多种模型的性能比较：我们可以对多种模型进行性能比较，找出最优的模型，以提高模型的准确率。

3)应用于更多领域：我们可以将项目应用于更多的领域，扩大项目的应用范围。

4)优化训练方式：我们还可以通过优化训练方式，例如采用更复杂的网络结构或者更优秀的优化算法，来进一步提高模型的性能。

# 参考文献

[1] P. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in arXiv:1810.04805, 2018.

[2] J. Vaswani, A. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All You Need," in Advances in Neural Information Processing Systems, 2017, pp. 5998-6008.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818-2826.

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," in International Journal of Computer Vision, vol. 115, no. 3, pp. 211-252, 2015.