Prospectus

Pose Estimation for Weight Lifting Form Analysis
This is now just notes on the topic

Alex Martin, Michael Kingsley, Vashisth Tiwari

CSC 298
University of Rochester
October 2022

UNIVERSITY of ROCHESTER

1 Introduction/Goal

Form is crucial in athletic performance. Minor changes in form can increase power, speed, and reduce the risk of injury for athletes. Form is easy to critique in weight lifting due to the standards from Olympic and competition lifting. There is less variability in form for weight lifting than in other sports where more advanced bio-mechanics are at play.

The goal of this paper is to use 2D human pose estimation to provide ways to score the form of lifters. This can be used in competition to red-light athletes not conforming to regulations or to provide coaching in form changes and differences for training.

2 Background

2.1 Pose Estimation

Pose estimation is an important focus in the computer vision community due to its large range of real-world applications. Pose estimation aims to automatically predict and track human posture by localizing joints and defining limb orientation.

2.2 Pose Datasets

There are many existing datasets for pose estimation. The ones we will explore using for this project are MPII, Penn Action, AI Challenger, and COCO. An example of the annotated data from MPII can be seen in figure 1.

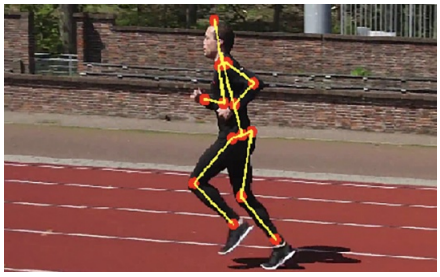


Figure 1: An example annotation from MPII

Max Planck Institute for Informatics (MPII) Human Pose dataset is a standard benchmark for single-person 2D pose estimation. The dataset consists of 25k images with 40k subjects collected from YouTube videos covering 410 daily human activities with complex poses, varying scales, and different appearances [3].

AI Challenger is a much larger dataset than MPII, consisting of 300K images for 2D pose estimation [6]. We include this dataset because ViTPose backbones were trained on a combination of it and COCO.

Penn Action Dataset consists of video sequences instead of images [4]. This will help in validating our process on videos so that real-time training feedback can be provided.

The COCO dataset provides a variety of human pose data with unconstrained environments, occlusion patterns, and different body scales [5]. This will be helpful in training for real-world applied use where the video environment can't be guaranteed to be perfect.

2.3 Models

ViTPose

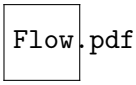


Figure 2: Model Framework

3 Our Data

To critique weight lifting form we need "gold" data. We plan to take videos of Olympic and other competition lifters and use these professionals as our gold forms. From there we will try to find videos of other people lifting, or film ourselves lifting and use these as our other data for critiquing form.

4 Framework

In this paper, we will propose a framework for lift classification, single-person pose estimation, and lift scoring. In this paper, we will make 3 possible contributions to the literature. **(1)** A framework for classifying and scoring exercise forms. **(2)** An additional implementation of ViTPose, either trained on new data or an additional implementation. **(3 maybe)** We will provide hand-annotated data on exercises for pose estimation and form correction labels.

4.1 Lift Classification

The lift classifier will be a Deep CNN. This CNN will be trained on lifting images that we want to classify. The CNN will then take an input image X_{img} that is a photo of someone lifting. The output of the model will be the label for that image corresponding to the lift that it is. Using this label, we will then select the 'gold' pose estimation corresponding to that label for the next part of the framework.

4.2 ViTPose

The input image X_{img} and the gold image Y_{lift_name} will be passed through the ViTPose model. For this task, we will fine-tune ViTPose on pose estimation for exercise through the PennAction dataset or our own data. If this does not provide satisfactory results, we will add to the base of ViTPose using methods similar to ViTPose-B.

4.3 Scoring Lifts

For now, we are only able to think of a simple way to score lifts.

$$score = \alpha * limb\_similarity + \beta * joint\_similarity \tag{1}$$

Where joint similarity is the difference between the joint angles

$$joint\_similarity = \frac{1}{N}\sum^{N} |gold\_degree - degree| \tag{2}$$

and limb similarity is the number of properly oriented limbs.

$$limb\_similarity = \frac{1}{N}\sum^{N} orientation \tag{3}$$

5 EVERYTHING ELSE IS USELESS BUT I DONT WANT TO DELETE IT

5.1 Training Pose Estimation

5.2 Testing Pose Estimation

Test on labeled data from the pose estimation to show valid performance.

5.3 Lifting Pose Estimation

Due to limited time, we will not annotate and test our model's performance on the lifting videos. Penn Action should suffice in the training and testing to prove performance on athletic videos.

5.3.1 Learning Good Form

Good form will be dictated by the competition lifting videos collected.

5.3.2 Evaluating Form

Using the videos we create or videos taken from the internet, we will use pose estimation to score our similarity to competition lifting form.

When comparing a gold form with itself, it will score 1, meaning perfect similarity. Then there will be scores assigned to limb orientation and joint angles, with their weighted effects (α, β) on the lifters similarity score to be decided as edge cases are tested.

$$score = \alpha * limb\_similarity + \beta * joint\_similarity \tag{4}$$

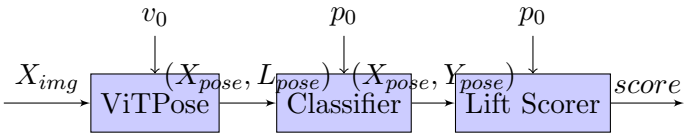
Where joint similarity is the difference between the joint angles

$$joint\_similarity = \frac{1}{N}\sum^{N} |gold\_degree - degree| \tag{5}$$

and limb similarity is the number of properly oriented limbs.

$$limb\_similarity = \frac{1}{N}\sum^{N} orientation \tag{6}$$

5.4 Framework



5.5 ViTPose

Image to Pose Estimation

5.6 Lift Classifier

The lift classifier will use a CNN framework.

5.7 Lift Scorer

$$\frac{1}{N}\sum^{N}(differences) \tag{7}$$

References

[1] Xu et al (2022), *ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation*

[2] Yang et al (2020), *TransPose: Keypoint Localization via Transformer*

[3] Andriluka et al (2014), *2D Human Pose Estimation: New Benchmark and State of the Art Analysis*

[4] Zhang et al (2013), *From actemes to action: A strongly-supervised representation for detailed action understanding*

[5] Lin et al (2014), *Microsoft COCO: Common Objects in Context*

[6] Wu et al (2017), *AI Challenger : A Large-scale Dataset for Going Deeper in Image Understanding*