GyML: Smart Fitness Trainer Using 3D Human Feedback Models

Ishan Khare*, **Anthony Qin***, and **Aditya Tadimeti***

*{iskhare, antqin27, tadimeti}@stanford.edu

CS 229 | Stanford University

Stanford Computer Science

Introduction

- Proper exercise form is crucial to maximize fitness benefits and minimize risk of injury
- We use state-of-the-art vision models to provide immediate feedback on a user's exercise form
- **Input:** video of user performing exercise
- **Output:** exercise classification and feedback

Dataset and Features

- Used the **FLAG3D Dataset**[2]: 7,204 labeled examples of 60 different fitness activities
- For each video, the pose data has dimensions (num. frames × 72)
- Data was flattened and padded with 0s to ensure the same dimensions before training
- Ran **principal components analysis (PCA)**[3] with K = 0.75 to speed up training and improve model generalizability
- We also captured some of our own raw 4K video data on an iPhone 15 Pro Max

Example user input video

Methods

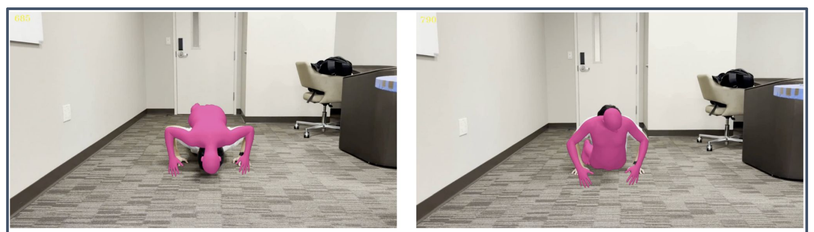
- **Cross-entropy loss** for a single instance:
  - $$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij}), \text{such that } \hat{y}_{ij} = \frac{e^{z_{ij}}}{\sum_{k=1}^{C} e^{z_{ik}}}, \text{where:}$$

$$N = \text{number of instances}, C = \text{number of classes},$$
$$y_{ij} = \text{binary indicator of instance } i \in \text{class } j,$$
$$\hat{y}_{ij} = \text{predicted probability of instance } i \in \text{class } j,$$
$$z_{ij} = \text{instance } i\text{'s raw score (logit) for class } j.$$

- **One-vs-Rest (OvR) Logistic Regression (LR)**: 60 specialized models, each distinguishes a particular exercise from the remaining 59
- **Multinomial LR**: single model that assigns a probability distribution to each exercise
- **Statistical Coach**: identifies video components along with feedback
  - employed L2-norm to develop repetition counting tool
  - Compare with 'gold standard' video for feedback

Experiments and Results

- **Human Mesh Recovery (HMR2.0)**[4] results for pose estimation



(a) HMR2.0 Video Output on Good Pushup Video (b) HMR2.0 Video Output on Bad Pushup Video

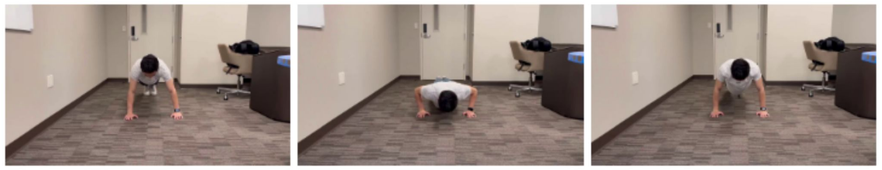
- **Classification Model Results**
  - Given true/false positives (TP/FP), and true/false negatives (TN/FN):

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP},\ precision = \frac{TP}{TP+FP},\ recall = \frac{TP}{TP+FN}$$

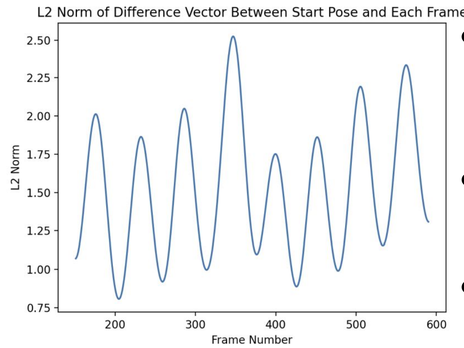
Approach	Training time	Accuracy	Precision	Recall	Num. Features
OvR LR	25 hours	99%	99%	99%	222,408
MLR	26 hours	99%	99%	99%	222,408
OvR LR+PCA	5 mins	81%	83%	81%	59
MLR+PCA	5 mins	96%	96%	96%	59

(7,204 examples split into 70/20/10 ratio for train-dev-test)

Statistical Coach Results



(b) Start Frame (c) First Peak (d) First Valley



(a) L^2 Norm for Start Pose - All Other Poses

- Valleys correspond to frames with similar norms to the start frame; *e.g.,* subplot (b) compared to (d)
- Avg. L2-norm between 'bad' valley and 'gold' start is 4.427; Max diff in 2 'gold' frames is 2.5
- Statistical coach isolates each corresponding rep and location where the diff norm exceeds the max 'gold' norm diff.

Conclusions and Future Work

- Human Mesh Recovery models worked just as expected – very well!
- For classification: models w/o PCA were heavy and likely don't generalize well; MLR+PCA is lightweight with good performance
- Statistical coach performed well on pushup case study due to the repetition in the videos – should work well for rep-based exercises
- In future, we want to improve robustness by curating our own data
- Will use a classifier for rep quality rather than basic norm analysis
- Coach system will support natural language feedback on exercises

References

1) Poster template from https://github.com/njwfish/stanford-poster-template
2) Y. Tang, J. Liu, A. Liu, B. Yang, W. Dai, Y. Rao, J. Lu, J. Zhou, and X. Li. Flag3d: A 3d fitness activity dataset with language instruction. In CVPR, 2023.
3) K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11):559–572, 1901.
4) S. Goel, G. Pavlakos, J. Rajasegaran, A. Kanazawa, and J. Malik. Humans in 4D: Reconstructing and tracking humans with transformers. In ICCV, 2023.

Stanford University