Applying Computer Vision and Deep Learning to the Art of Boxing

Evan Huang (evhuang@stanford.edu)
Cedric Fraces Gasmi (cfraces@stanford.edu)

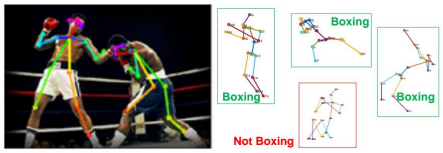
Background

This project focuses on the identification of boxing poses and the recognition of styles of famous boxers. We use a series of deep learning methods to identify, classify and recognize different boxing stances. We use transfer learning at various levels to encode human body poses using pose extraction approach. We then build a labeled dataset by gradually applying classification Neural nets to series of image frames and videos borrowed from boxing and non-boxing footages. We then apply unsupervised learning methods to cluster boxers based on their styles and extract "signature moves" characteristic of some boxers. The classification was done on two famous boxers of the 20th century (Ali and Tyson). This tool can be used to help boxers around the world assess their performances and track their progress

Data

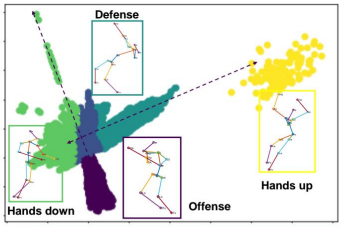
Set of 17 coordinates poses captured from images and videos of boxing/non boxing scenes
Two challenges to address in getting a clean labeled dataset :

- Eliminate poses that did not correspond to boxers
- Recognize the right boxer from a frame generally showing two fighting

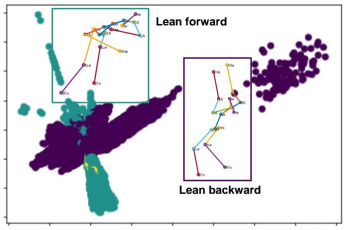


[1CMU OpenPose Project: https://github.com/CMU-Perceptual-Computing-Lab/openpose
ImageNet: http://www.image-net.org/
Data-Driven Ghosting using Deep Imitation Learning (Disney Research) Peter Carr (https://www.disneyresearch.com/publication/data-driven-ghosting/)
Swingbyte Golf Assistant (Modeling and analyzing golf swings using machine learning) (http://www.swingbyte.com/)
Boxing styles: https://en.wikipedia.org/wiki/Boxing_styles_and_technique

5 clusters K-means
Hands position and offense/defense



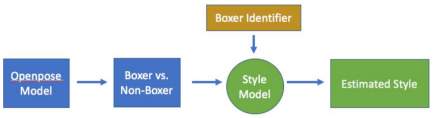
2 clusters EM
Lean forward – Lean backward



Methods

Transfer Learning

One concept taken advantage of multiple times in this project was transfer learning. In order to get pose data, we used OpenPose[1] a deep learning model that converts the initial data type of images into poses. We apply additional layers of learning to output boxer or non-boxer (and styles).
We used the architecture of existing classifiers and adjusted hyperparameters for our application. One such model is pose detection by Srini Aananthakrishnan which used an Adam optimizer, dropout = 0.1, and a relu activation function. The ultimate intended model design is a combination of separate models.



Unsupervised Learning

One of the main challenges in training an algorithm to recognize boxers is that the majority of boxing footages feature two boxers. We use a combination of K-means and Mixture of Gaussians algorithms to cluster the various poses and a PCA transform to represent them in 2D. Signature moves were extracted from this process.

Supervised Learning

We train a neural network on boxing and non-boxing poses in pictures from Imagenet. We used a 6 layers dense network with decreasing number of hidden units. ReLu activation functions on all layers and a sigmoid logistic regression on the last layer. We employed data augmentation techniques to improve our model's performance. We added random noise to the existing data and ran the same training. The results showed a substantial improvement.

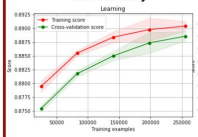
Results

We adopted a sequential approach:

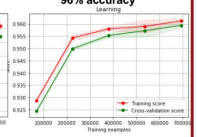
- Human vs Non human: Heuristic (close to 100%)
- Boxers vs non boxers: NN (96% accuracy)
- Tyson vs Ali: Unsupervised learning (~75% accuracy)

Sensitivity of the model to training data. Tripling data through additional poses and augmentation and results in an increase from 89% to 96% accuracy

250,000 poses
89% accuracy



700,000 poses
96% accuracy



Discussions

We experimented various methods and approached the problem in a sequential manner. We first attempted to recognize boxers vs non boxers poses. We then classified various poses based on the classes of positions they represented (offense, defense,…). We finally were able to identify clusters of moves that were representative of two emblematic boxers (Muhammad Ali and Mike Tyson).
This low dimension representation provides a consistent way of comparing boxers. It sets the path for a quantitative assessment of technical skills. The extraction of clean poses was the main challenge. We were able to create classifiers that allowed us to use rawer footages (and hence increase the size of our training dataset) at every step.
We are confident that this work can be re-used to push passed our objective and expand into the area of movement evaluation and correction.