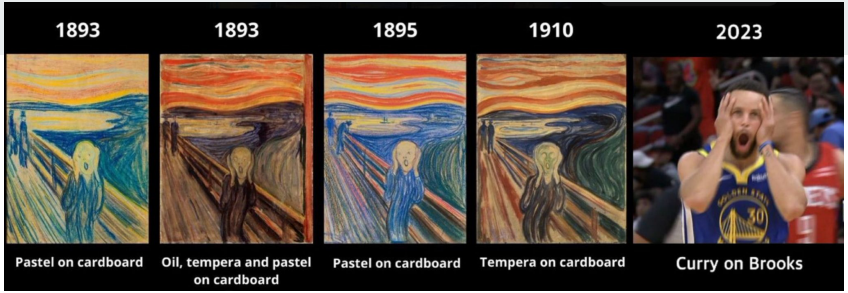
From Field to Canvas: Bridging Sports and Art with Deep Learning Pose Analysis

Patricia Luc, Miira Efrem, Asia Gray, Dina Cazun-Moreno

Motivation



Goal: develop deep learning system that can analyze live sports pictures and match them with existing art pieces

- The system will use pose estimation and posture analysis to identify similarities between the poses depicted in the sports pictures and those in the art pieces.

Expected outcome: model that can accurately identify and match poses from live sports pictures with corresponding poses in an art images dataset

- The system will identify key points in the body postures of athletes in the sports pictures and search for similar poses within the art
- The matching process will involve comparing pose features, body proportions, and overall composition to find the closest matches.

Inspiration: X account @ArtButSports.

- Note that this account does not use AI to generate any of its Sports to Arts matches.

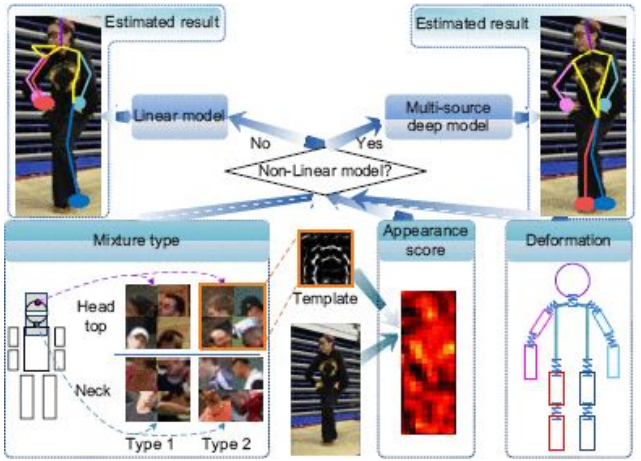
Previous Work

Action recognition using pose estimation with an artificial 3D coordinates and CNN

- by Jisu Kim, Deokwoo Lee
- Explores the use of CNNs to recognize human actions based on pose estimation data, specifically by generating artificial 3D coordinates to enhance the model's understanding of spatial relationships and motion patterns

Multi-source Deep Learning for Human Pose Estimation

- by Wanli Ouyang; Xiao Chu; Xiaogang Wang
- Discusses a deep learning framework that integrates multiple sources of information, such as images and videos, to improve the accuracy and robustness of human pose estimation models.



Overview

We have segmented our implementation plan into specific categories. First, we started off with data collection, which consisted of finding a sports images dataset and an art dataset. Then we worked to find a pretrained model that was able to perform pose recognition from the actions performed in images. The particular model we used was trained on the Stanford40 dataset which contains 40 human actions. We then tested the model on both the datasets, and our project takes in a sports image input and finds the closest art image with the same pose.

Technical Description

Data Collection:

- Gather a diverse dataset of live sports pictures covering various sports and poses.
- Compile/Look for an existing database of paintings and other art pieces with human subjects in various poses.

Model Development:

- Utilizing the pretrained Image Action Recognition Model to identify key points in sports pictures and art pieces.

Matching Algorithm:

- The model calculates the probabilities of the image being each of the 40 different poses
- To compare pose features and compositions between the input sports image and find the art image with the closest pose, our model calculates Euclidean distances between the input probability array and that for each of the art images, the most similar image having the minimum distance

Testing and Evaluation:

- Evaluate the performance of the system using a validation dataset of sports pictures and art images.
- Measure the accuracy of pose matching

Technical Description- Resources

Image Action Recognition with Stanford40 Dataset and GUI: A model for image and action estimation in real-time.
We used this as our training and testing model for developing the matching algorithm and fine tuned it to be adaptable to the oftentimes complex poses seen in paintings and especially in live sports pictures.

OpenCV:
We used OpenCV toward the beginning of our project just to test out our image segmentation of the humans and the painting so we can properly extract figures and then from there we know that we can extract the people and then from there analyze the poses.

TensorFlow Hub: Provides pre-trained models and modules for various deep learning tasks, including pose estimation.
We also checked out the TensorFlow Hub to see their pose estimation models

Art Images: Drawing/Painting/Sculptures/Engravings: Art images API
This is the art API that we used. It has 9000 images containing 5 types of art- drawing, engravings, iconography, painting, and sculpture.

100 Sports Images Classification: Sports images API
This is the sports API that we used. It's a collection of sports images from 100 different sports.

Image Action Recognition with Stanford40 Datasets and GUI.

It allows users to interact with the system through a Graphical User Interface (GUI) to recognize actions performed in images and we use VIT model as the base model. The Stanford40 dataset contains images of 40 different human action classes, such as "applauding", "fishing", "holding an umbrella" etc.

Features

- Recognize actions in images using pre-trained models.
- Interact with the system through a user-friendly GUI.
- Display the recognized action class along with the input image.

Requirements

- Python 3.x
- Tkinter library for GUI (usually pre-installed with Python)

The image action recognition model used in this project is based on the Vision Transformer (VIT) architecture. The VIT model is a state-of-the-art deep learning model for image classification tasks.

The VIT model implementation used in this project is based on the repository ViT-pytorch by jeonsworld. The pre-trained VIT model provided in this repository is fine-tuned on the Stanford40 dataset for action recognition.

Results & Analysis

We were able to compile diverse datasets, not only paintings, but also engravings and sculptures. The model we used was able to come up with some neat results regarding matching the sports pictures to the art images that we had available.

The actions perceivable by the model aren't as vast as a the ones a human can identify, but while it was limited to 40 actions, we were still able to get pretty good results. In the future, we would expand the number of actions that are detectable by our pretrained model and also increase the diversification of the dataset.

input:

output:

input:

output:

Conclusion

Potential future project adjustments

- Expand the number of actions that are detectable by our pretrained model
- This is done by increasing the diversification of both datasets and retraining the model on additional images with new action labels
- Working with more GPUs to speed up the training and validation processes, this model currently runs pretty slow
- Expanding the scope to also do pose detection on videos
    - A lot of the other models we initially considered did this
- Developing a more complex algorithm to compare images pose detection results to better find the best match

Sources

Arhpa. “ARHPA/Image-Action-Recognition: Image Action Recognition System That Uses The Stanford40 Dataset.” GitHub, github.com/ARHPA/Image-Action-Recognition?tab=readme-ov-file#features. Accessed 29 Apr. 2024.

ArtButMakeItSports (@artbutsports) · x, twitter.com/ArtButSports. Accessed 29 Apr. 2024.

Li, S., Liu, Z.-Q., & Chan, A. B. (2014). Heterogeneous Multi-task Learning for Human Pose Estimation with Deep Convolutional Neural Network. *International Journal of Computer Vision*, *113*(1), 19–36. https://doi.org/10.1007/s11263-014-0767-8

Li, S., Zhang, J., & Chan, A. B. (2015). Maximum-Margin Structured Learning with Deep Networks for 3D Human Pose Estimation. *ArXiv (Cornell University)*. https://doi.org/10.1109/iccv.2015.326

Luvizon, D., Picard, D., & Tabia, H. (2020). Multi-task Deep Learning for Real-Time 3D Human Pose Estimation and Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. https://doi.org/10.1109/tpami.2020.2976014

Ouyang, W., Chu, X., & Wang, X. (2014). Multi-source Deep Learning for Human Pose Estimation. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2014.299