AI-Assisted Performance Analysis Deep Learning for Live and Archival Theater

Michael J. Rau,¹ Peter Broadwell,² Simon Wiles,² Vijoy Abraham² ¹ Theater and Performance Studies ² Center for Interdisciplinary Digital Research ^{1/2} Stanford University

This project seeks to build on recent advances in human pose-estimation and object tracking technologies powered by emergent deep-learning methods to develop new tools and techniques for the analysis and augmentation of theatrical performances.





capable of performing inference of body positions and motions from images captured by standard visual-light cameras, unlocking for archival and live footage possibilities that previously only existed in constrained environments with expensive motion-capture suits and equipment.

Leveraging these methods, we are developing software and methodologies that use pose data to aid in the investigation, analysis, elucidation, and quantification of insights into aspects of theatrical performance including directorial style and choreographic patterns.

- A window into a performance is provided by the overview which presents pose-estimation results on a timeline.
- Here it is possible to see how many poses have been identified on a frame-by-frame basis, and how many have been successfully tracked across frames.



Details for frame #53597			
#Poses:	7		
#Tracked Poses:	6		
Time:	37m15s		
Pose #1 Confidence: 0.807 Track 1375		similar	Ø
Pose #2 Confidence: 0.774 Track 1379		similar	Ø
Pose #3 Confidence: 0.744 Track 1370		similar	8

• The Frame Viewer component visualizes an individual frame and associated pose data

similar 🛛 🗞 Pose #4 | Confidence: 0.735 | Track 1383 Pose #5 | Confidence: 0.659 | Track 1382 similar 🛛 🗞 Pose #6 | Confidence: 0.566 similar 🛛 🗞 similar 🛛 🗞 Pose #7 | Confidence: 0.578 | Track 1381

IUDVter video posedata explorer (autosave

+ ≫ 🖓 🖺 ↑ 🔸 ► Run 🔳 C 🕨 Code

ptional: cluster analysis of normalized pos

sorted_bit_counts, normalized_poses, normalized_pose_metadata, pose_data, video_file, clusters_to_draw=10, average_backgrounds=False,

ustering video poses into 100 cluste

Drawing averages of cluster poses (and backgrounds, if requested Cluster: 9 | total poses: 5469

e cell below computes a K-means clustering of the poses based on the similarities of their vectors, then calculates and visualizes the relative sizes of

the cluster representative poses can take quite a bit longer due to the overhead of averaging the source images for the backgroup

of their poses. Note that if the average_backgrounds parameter is set to True rather than False

• Using vector similarity searching, the poses most similar to the target pose can be identified and located within the performance. Work continues to augment simple nearest-neighbors -style similarity across entire poses with other heuristics that help to focus in on the context-specific kinds of similarity that are meaningful to researchers.



Technical Stack

- At the present time the machine-learning stack is centered around pose-estimation provided by **OpenPifPaf**, augmented with object tracking across frames based on work by **ByteTrack**. Vector similarity is performed using algorithms from the **Faiss** library.
- The application stack is oriented around a **PostgreSQL** database augmented with **vector storage and search** capabilities to handle pose/armature vectors and embeddings.
- A **Python**-based backend handles ingestion and analysis of new video files, committing data to the database and surfacing it to the primary user interface via a **FastAPI**-based REST API.
- The web-based frontend is developed in **Astro** and **Svelte**.
- The application stack is developed and deployed as a (currently) **four-container docker orchestration** for convenience and



portability.

• The Python backend also serves a Jupyter notebook server out of the same environment to facilitate further exploration and experimentation.



https://github.com/sul-cidr/mime

