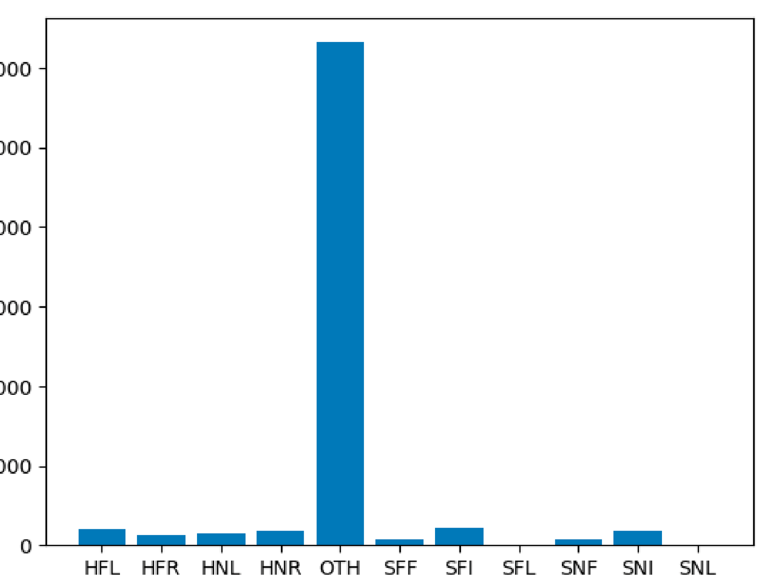
Pose Estimation and LSTM for Tennis Stroke Classification

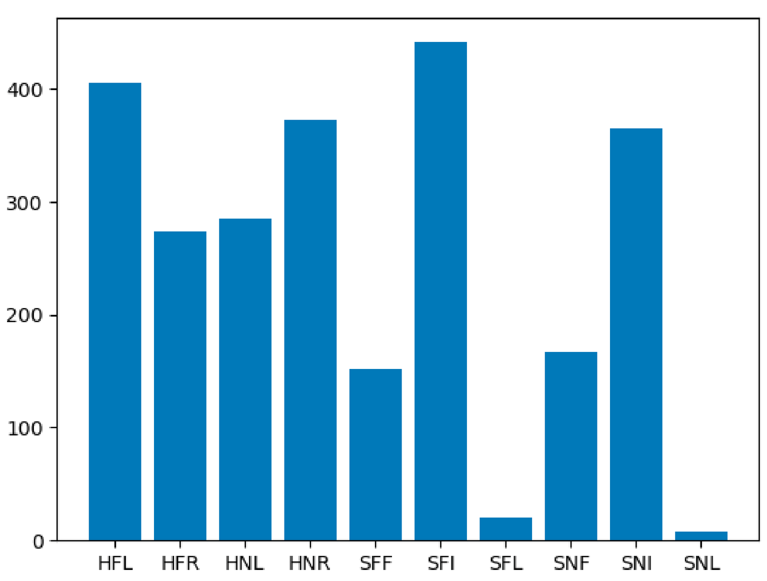
Vanessa Bellotti and Tanay Nistala

Research Question: Can a transfer learned LSTM learn to classify tennis strokes given pose keypoints as input?

Data Understanding

Goal: discern class imbalance in the dataset





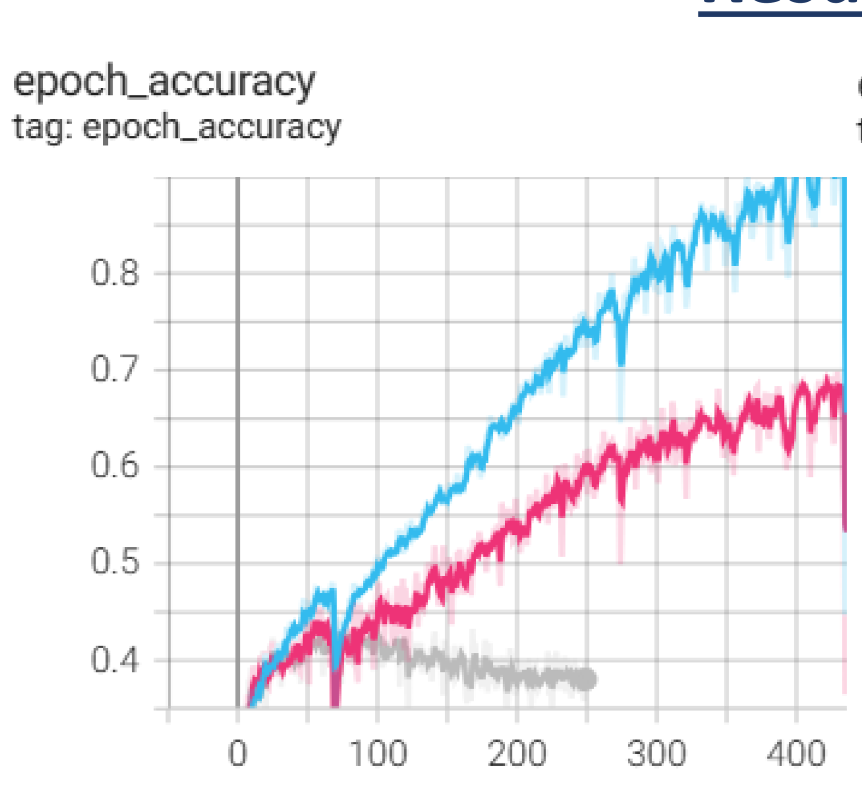
Pose Estimation

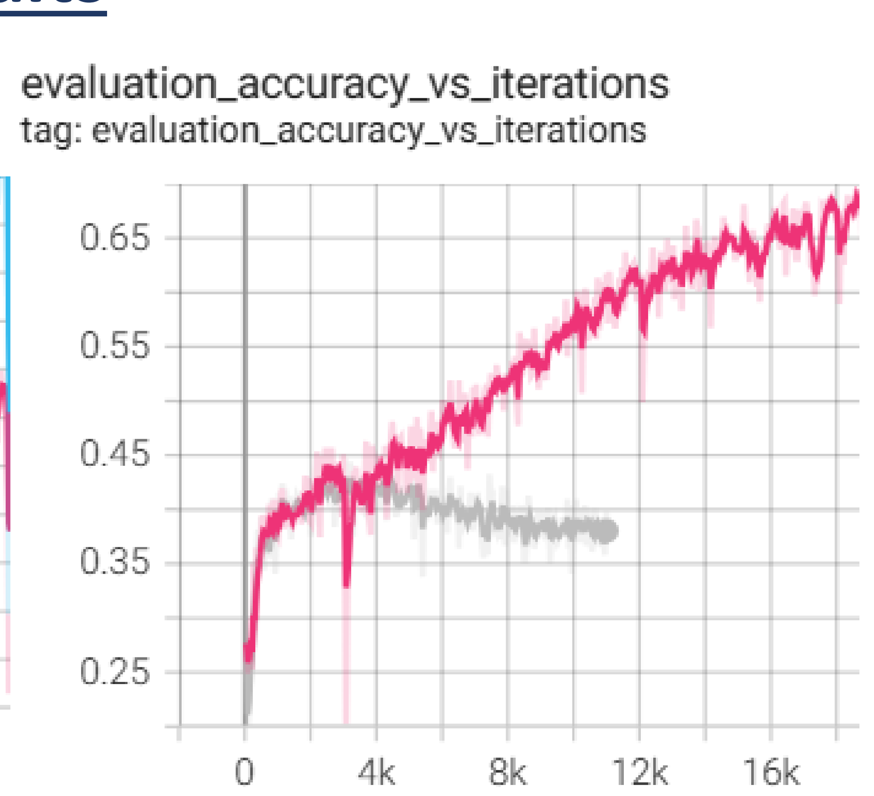
Goal: measure preliminary results from keypoint detection

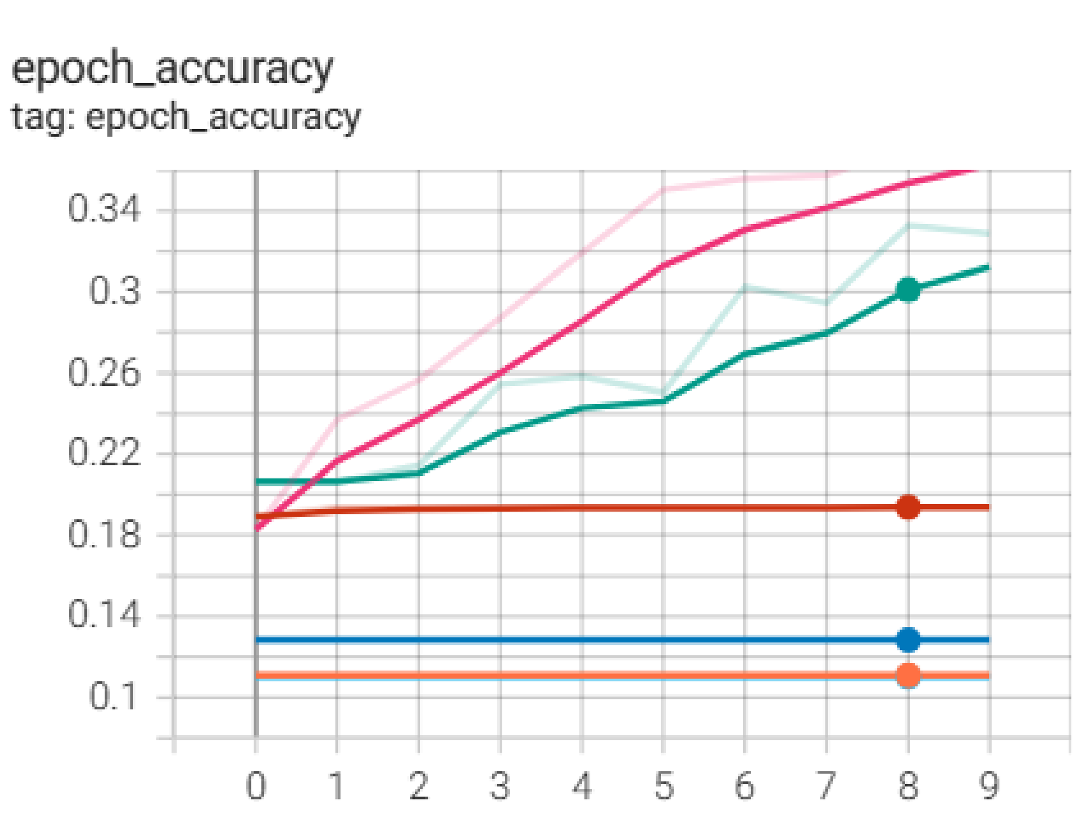
Model Selection

- Long Short-Term Memory networks excel at modeling temporal dependencies and can process variable-length sequences
- LSTMs can retain contextual information over time, enabling the model to understand a tennis match’s evolving dynamics.
- The ability of LSTMs to maintain a memory of previous poses is crucial for understanding transitions between stances and predicting based on player pose history.
- Tennis strokes often follow distinct sequential patterns, such as the build-up to a serve or the sequence of movements in a forehand stroke.
- Combining LSTMs with pose embeddings enhances the model's ability to understand both spatial and temporal aspects of tennis poses.

Results







Conclusion

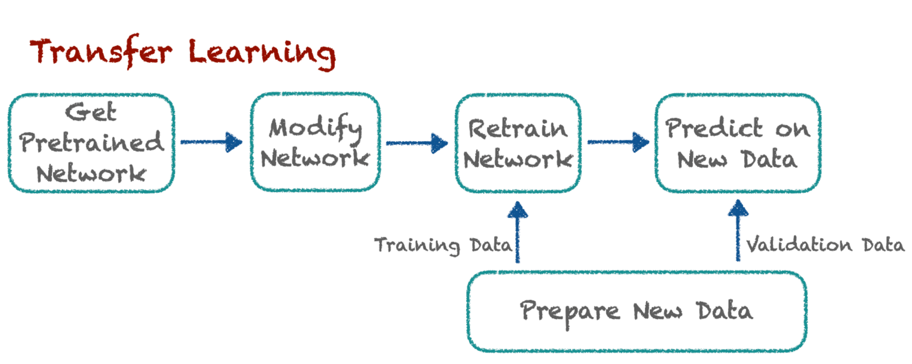
1. Multiclass classification of tennis strokes via keypoints is **possible** but needs to be further ***finetuned***
2. Dataset needs to be **upsampled** to address the “Other” issue

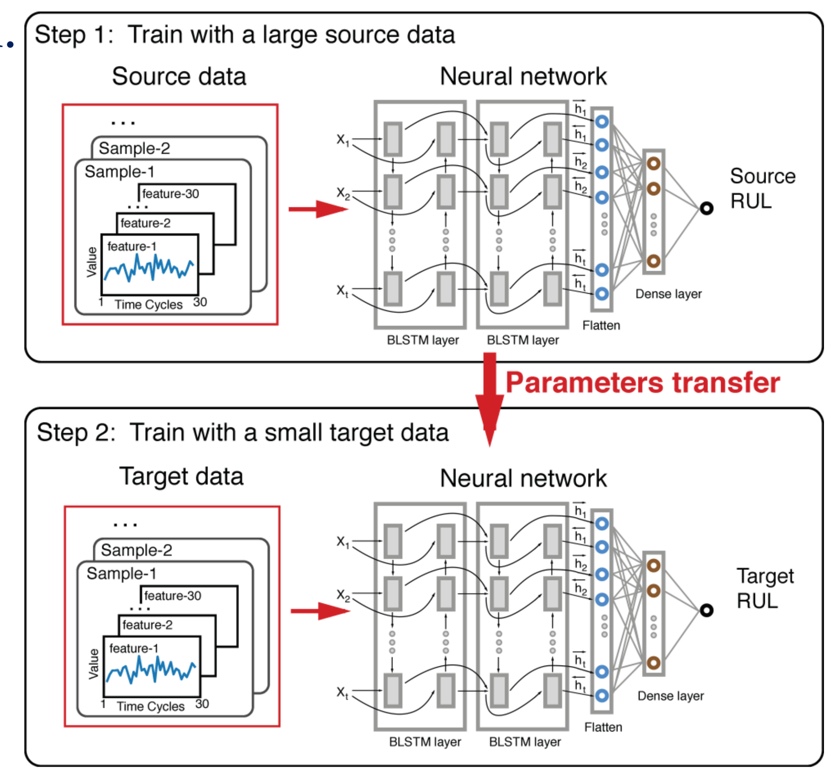
Data Preprocessing

- •JSON Data: Parsed through files to eliminate unnecessary features
- • TXT Data: Implemented output classes parsing; Pruned "Other" set for precision
- •Frame Filtering: Considered removing frames with multiple players. Opted against it to retain diverse interactions
- •Color Space Transformation: BGR to RGB conversion for MoveNet Pose estimator to enhance model robustness

Transfer Learning

Transfer learning let us capitalize on pre-existing knowledge from a model trained on a more extensive dataset, mitigating the challenge of insufficient labeled tennis-specific data.

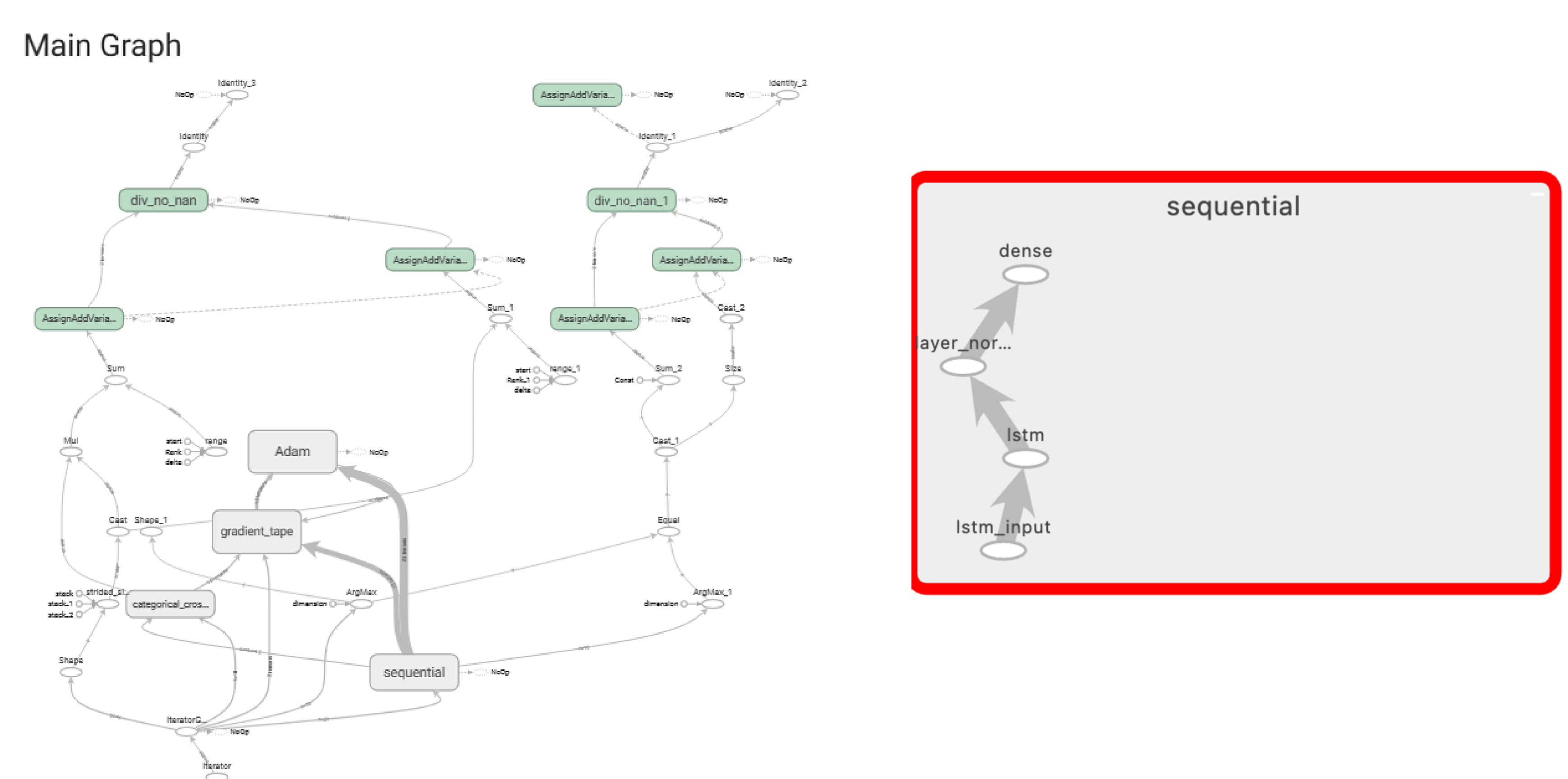




Model Training Pipeline and Process

1. **MoveNet Pose Detection:**
   1. Utilized MoveNet for pose detection on tennis image data.
   2. Extracted key points representing crucial body joints in each image.
2. **One-Hot Encoding Labels:**
   1. Defined output classes representing tennis player types of strokes.
   2. Employed one-hot encoding to transform these categorical labels into a format suitable for neural network training.
3. **Transfer Learning Approach:**
   1. Adopted a transfer learning strategy to capitalize on the pre-trained MoveNet model's understanding of human pose.
   2. Leveraged the learned features and representations from MoveNet's training on a diverse dataset.
4. **Neural Network Training:**
   1. Designed and trained a neural network to predict tennis player stance and strokes.
   2. Transferred the knowledge gained from MoveNet's key point extraction to enhance the learning process of our specific task.
5. **Improved Generalization:**
   1. Transfer learning enhanced the model's ability to generalize well to our tennis-specific task by leveraging knowledge from a more generalized pre-trained model.

Neural Architecture



Contributions

1. This work introduces a transfer learned neural network layer on top of pose estimation keypoints to learn tennis stroke classes
2. This work identifies highly fine-tuned classes, making the results applicable to future work like commentary generation or an automated umpire application

www.PosterPresentations.com