Simultaneous Identification of Who's Doing What in Videos
Nyle Siddiqui PhD, Philomina Ekezie
University of Central Florida CRCV Lab, Mercer University

Project Introduction
- Subject detection and action recognition are two major tasks within computer vision and video processing
- **Problem at Hand:** There exists virtually no models or datasets that aim to complete these tasks using a simultaneous approach

Dataset Retrieval
- Our dataset has ten subject classes and six distinct action classes.
- The dataset was populated by collecting trimmed samples of each action from full-length tennis match videos via YouTube.
- The dataset currently has 3000 samples; this number will increase in the following months.

Project Objectives
- Creating a novel dataset with both subject and action labels for the novel problem of simultaneous subject detection and action recognition
- Training our existing complex model using the novel dataset to successfully accomplish both tasks

Model Architecture
Our model consists of a 3D-CNN encoder that extracts spatio-temporal features from the input. These features are then passed through two transformer decoders, utilizing disentangled feature learning to effectively learn subject and action features separately.

Preliminary Results

<table>
<thead>
<tr>
<th></th>
<th>Subject</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Model</strong></td>
<td>~66%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Our preliminary results were produced on a dataset size of 200 samples. Both percentages show that the model was able to detect the subject and the action performed fairly well.

Future Efforts
As the dataset grows, the model will continuously be retrained to achieve higher results. The utilization of bounding boxes may also be included.

Conclusion
The creation of this novel dataset and our accompanying complex model solves the novel problem of simultaneous subject and action detection. This problem has many real-world applications, e.g., surveillance systems and sports video analytics, where both who and what are important factors.