Motivation

Falls are one of the most dangerous situations for elderly people, resulting in fatal injuries and loss of independence of the elderly. Video-based fall detection has been more and more popular thanks to its broad application.

In this project, we predict the occurrence of fall in video clips with a variety of background settings. Using a two-stream model combining both spatial and temporal information, we achieved an F-1 score of 91.77%.

Data

**Data source**
- 873 videos from three online datasets: UI Fall Detection Dataset, Le2i Dataset, and Multiple Camera Fall Dataset
- Background settings: coffee room, home, lecture room, office.
- Long videos with multiple falls and other activities are cut into shorter clips.

**Pre-processing**
- Extract frames from videos using FFmpeg
- Data augmentation: flipping, rotation, scaling, etc.

![Image 1: No Fall](image1.png)

![Image 2: Fall](image2.png)

Model

![Architecture of two-stream video classification](architecture.png)

1. **Spatial Stream**: Recognize actions from static images
   - Model: ResNet 18, 101, 152
   - Input: static video frames
   - Pre-trained on ImageNet dataset
   - Dropout 0.8
   - Model: ResNet 152

2. **Temporal Stream**: Detect motion with optical flow
   - Model input: Optical flows, which capture the motion between consecutive frames.
   - We use TV-L1 optical flow estimation from OpenCV
   - Transform images to optical flow:
     - Input (a) and (b); two consecutive images $I_0(x)$ and $I_1(x)$, with $x = (i, j)$ the pixel index.
     - Vector field $v$: $v(x) = (v_0(x), v_1(x))$
   - Output: optical flows: horizontal components of vector field transforms to optical flow (d), vertical components to (e).

3. **Merge the two streams**
   - Average the class score of output layer from two streams. Make prediction based on the averaged scores.

![Image 3: No Fall](image3.png)

![Image 4: ResNet 152](image4.png)

![Image 5: Explains optical flow](image5.png)

Results and Discussion

**Results**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>Train</td>
<td>96.29%</td>
<td>94.58%</td>
<td>95.18%</td>
<td>96.26%</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>93.51%</td>
<td>92.56%</td>
<td>91.64%</td>
<td>91.77%</td>
</tr>
<tr>
<td>Temporal</td>
<td>Train</td>
<td>99.99%</td>
<td>100%</td>
<td>99.99%</td>
<td>99.99%</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>87.01%</td>
<td>87.60%</td>
<td>84.11%</td>
<td>85.52%</td>
</tr>
<tr>
<td>Fusion</td>
<td>Test</td>
<td>91.83%</td>
<td>90.76%</td>
<td>86.40%</td>
<td>88.53%</td>
</tr>
</tbody>
</table>

![Image 6: Train set F1 Score](image6.png)

![Image 7: Test set ROC curve](image7.png)

**Discussion**

- Error Analysis: the models perform not as good (1) when people fall on objects that are not on the ground and (2) when the action of falling happens within a very short time in a relatively long video.
- For the protection of the elderly, it is important to detect as many falls as possible, and to keep false-negative rate low. In real world application, prediction threshold can be set lower than 0.5.

**Future work**

- Gather more videos where people fall on objects such as coffee table
- Object detection: detect the person before activity recognition.
- Detect multiple fall activities

References

