Machine Learning in Video Surveillance for Fall Detection

L. Anishchenko
Remote Sensing Laboratory, Bauman Moscow State Technical University, Moscow, Russia

http://www.rslab.ru/english/
Presentation structure

1) Introduction (state of the art)
2) Deep learning in CCTV cameras data analysis
3) Experimental data
4) Results and Discussions
5) Conclusion
State of the art

1 Biometric identification (NeoFace)

2 Recognition of emotions or facial expressions (DeepFace).

Problems with intelligent video analysis application on practice:

• classification algorithms are usually extremely sensitive to lighting conditions, vibrations, etc.
• classification algorithms performance are overestimated as being measured for known conditions.
Problem to solve

Detecting an abnormal situation by means of CCTV cameras intellectual analysis when a person is in danger and there are no passers around, who can help.

In the cold season it is essential to provide first aid as soon as possible in order to prevent the negative consequences associated with the hypothermia.
Limitations of Experimental Datasets

Unchanged laboratory conditions (uniform illumination of the entire analyzed scene).

One and the same person acts as test subjects.

Movement artifacts of the "fall" type are similar in preceding actions of the subject and her/his relative position toward the camera at the time of the fall.

Falls are performed on a cushioning mat with a significant color contrast with respect to the clothing of the subject.
Experimental Datasets
provided by the Laboratory of Electronics and Imaging of the National Center for Scientific Research in Chalon-sur-Saone*

Video details:
• Frame width: 320
• Frame height: 240
• Frame rate: 25 frames/seconds

Video with fall episode
(Surrounding conditions: Home )

Experimental Datasets

provided by the Laboratory of Electronics and Imaging of the National Center for Scientific Research in Chalon-sur-Saone*

Video without fall episode
(Surrounding conditions: Lecture room)

Video signals are recorded for various environmental conditions.

The illumination of the experimental scene is irregular.

4 different subjects (3 men and 1 woman).

Falls were performed at different viewing angles, both from the standing position and from the sitting position.

Volunteers were falling both on a specially prepared cushioning base, and directly on the floor.
AlexNet* available by installing Neural Network Toolbox™ Model for AlexNet Network support package was used as a pre-trained CNN.

Data processing

Dividing record into separate frames $\rightarrow$ Manually classifying $\rightarrow$ Reshaping $\rightarrow$ Split DS into training and testing sets $\rightarrow$ Setting initial learning rate $\rightarrow$ Training CNN

30 records from the dataset were used to train and to test classifier.
Data processing

For each record operator visually detected fall episodes, and provide the number of the frame corresponding to the start and the end of the fall episode. The duration of fall episodes was estimated as 22±9 frames (meanSD) or 0.7±0.3 s.

Fall (class 0) 
Non-fall (class 1)
Each frame was re-sampled to make it compatible with CNN AlexNet (277x277 pixels).

The dataset was randomly divided into training and testing sets in the ratio 80:20%.

Initial learning rate parameter was set to be equal 0.001 to make changes of the initial values of weights small.
Data processing

1. Dividing record into separate frames
2. Manually classifying
3. Reshaping
4. Split DS into training and testing sets
5. Setting initial learning rate
6. Training CNN

Graphs:
- Accuracy (%)
- Loss

Iteration range: 0 to 40
Fall/non-fall classification results
(known conditions)

- Cohen's kappa = 0.93
- Accuracy = 0.99
- Sensitivity = 0.93
- Specificity = 0.99
- Positive predictive value = 0.94
- Negative predictive value = 0.99
Fall/non-fall classification results
(unknown conditions)

- **Cohen’s kappa**: 0.60
- **Accuracy**: 0.99
- **Sensitivity**: 0.61
- **Specificity**: 0.99
- **Positive predictive value**: 0.61
- **Negative predictive value**: 0.99

Confusion Matrix:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>63</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>41</td>
<td>7172</td>
</tr>
</tbody>
</table>

- True Positives: 99.4% (41/7172)
- False Positives: 0.6% (40/63)
- True Negatives: 98.0% (61.2% of 63)
- False Negatives: 0.5% (60.6% of 63)

- Positive predictive value: 0.61
- Negative predictive value: 0.99
The proposed method allows detecting fall events on a video records. We achieved Cohen’s kappa of 0.93 and 0.60 for the fall – non-fall classification for the known and unknown for classifier surrounding conditions, respectively.

Limitations of the study: the dataset is small (4 young examinies).

It is possible to improve the classifier performance by using additional heuristics.
Acknowledgements

The study was supported by the grant of Russian Foundation for Basic Research (#17-20-03034).
Thanks for your attention!

Any Questions?

Lesya Anishchenko, PhD
Senior Researcher
Remote Sensing Laboratory,
Bauman Moscow State Technical University
anishchenko@rslab.ru
www.rslab.ru