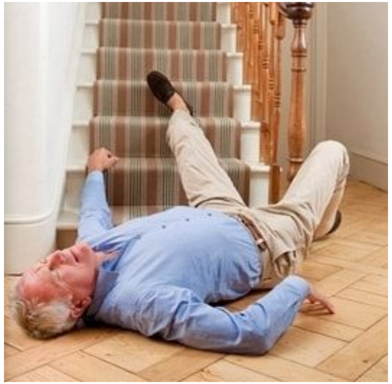
NCSA Fall Hackathon 11/10/2019 Team 1B

Hongyu Shen , William Wei , Asad Khan , Shirui Luo , Madhu Vellakal

Fall Detection

- On an average ~30%, people >65 fall
- Healthcare costs total upto US$55 billion by 2020
- Fall detection systems are need of the hour
- Major types – wearable sensors, vision devices

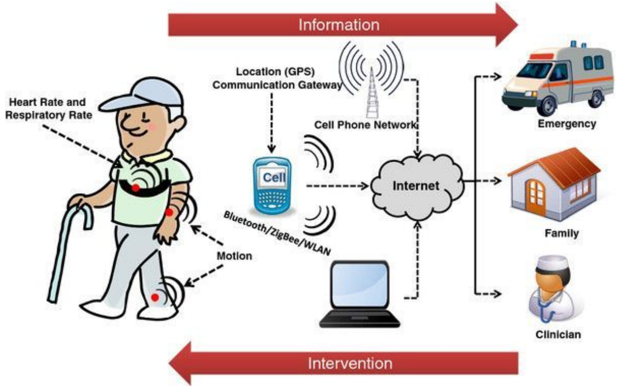


https://medicalalertsystemreviews.net/automatic-fall-detection-alert/

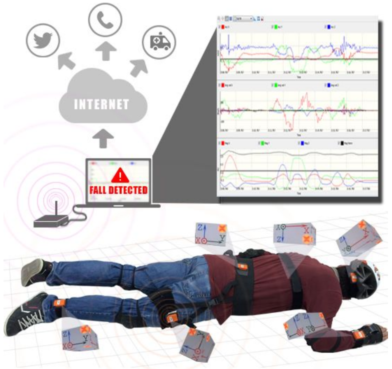
https://happyseniors.care/en/technology/fall-detection-and-prevention-systems/

Wearable Sensors

- Intrusive
- Prediction based on accelerometer and threshold based algorithms
- Poor differentiation between activities of daily living (ADLS) and falls



A review of wearable sensors and systems with application in rehabilitation, Patel et.al



Detecting Falls with Wearable Sensors Using Machine Learning Techniques, Ozdemir et.al

Vision Devices

- Non-intrusive
- Multi-mode data capture – angle, cameras, activities
- Remote verification of events

https://visl.technion.ac.il/projects/2014s08/

Camera-Based Fall Detection on Real World Data, Glen Debard, Peter Karsmakers, Debard et.al

Datasets

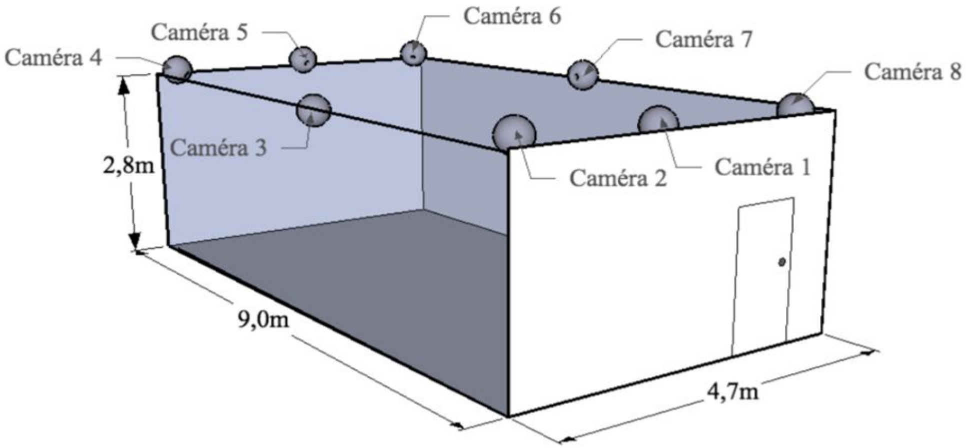
- UR Fall Detection Da
  - 70 videos – 30 fa
- Kinetic Human Actio
  - 400 class of acti
- Multiple Cameras Fall Dataset

(g) riding a bike

http://fenix.univ.rzeszow.pl/~mkepski/ds/uf.html
https://arxiv.org/pdf/1705.06950.pdf

Multiple Cameras Fall Dataset

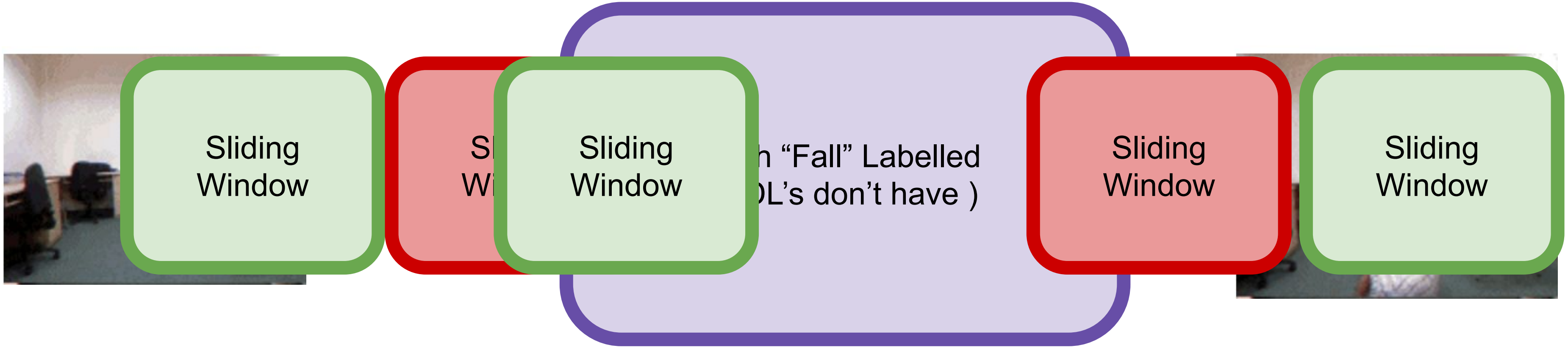
- 24 scenarios, each scenario has 8 cameras from different angles, in total 24 X 8 = 192 videos
- Each videos have 1000 ~ 5000 frames of images, with ~30 images are falling images



Camera configuration

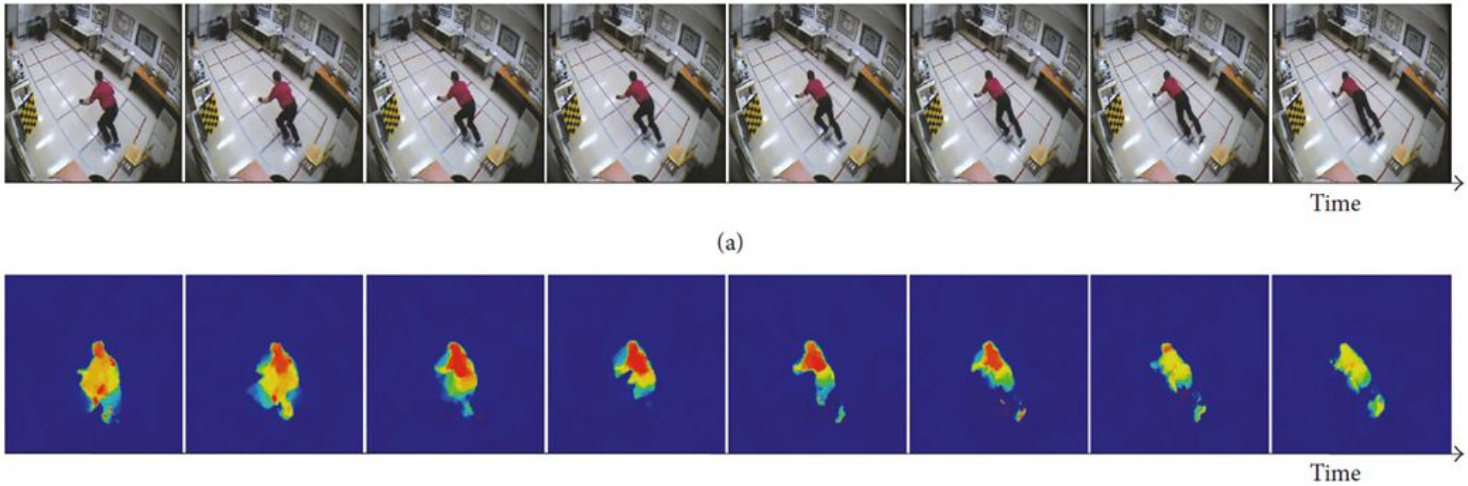
Data Preparation

- Data split (works in RGB videos and optical flow)



Optical flow:

- algorithm represents the patterns of the motion of objects as displacement vector fields between two consecutive images.
- anything static (background) is removed and only motion is taken into account



(b)

(a) Sample of sequential frames of a fall from the Multiple Cameras Fall Dataset and (b) their corresponding optical flow horizontal displacement images

2017, Wireless Communications and Mobile Computing

Optimal Flow Algorithm: Gunnar Farneback

Change RGB color to HSV(hue、saturation、value)

Consider a pixel $I(x, y, t)$ in first frame (Check a new dimension, time, is added here. Earlier we were working with images only, so no need of time). It moves by distance (dx, dy) in next frame taken after dt time. So since those pixels are the same and intensity does not change, we can say,

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$

Then take taylor series approximation of right-hand side, remove common terms and divide by dt to get the following equation:

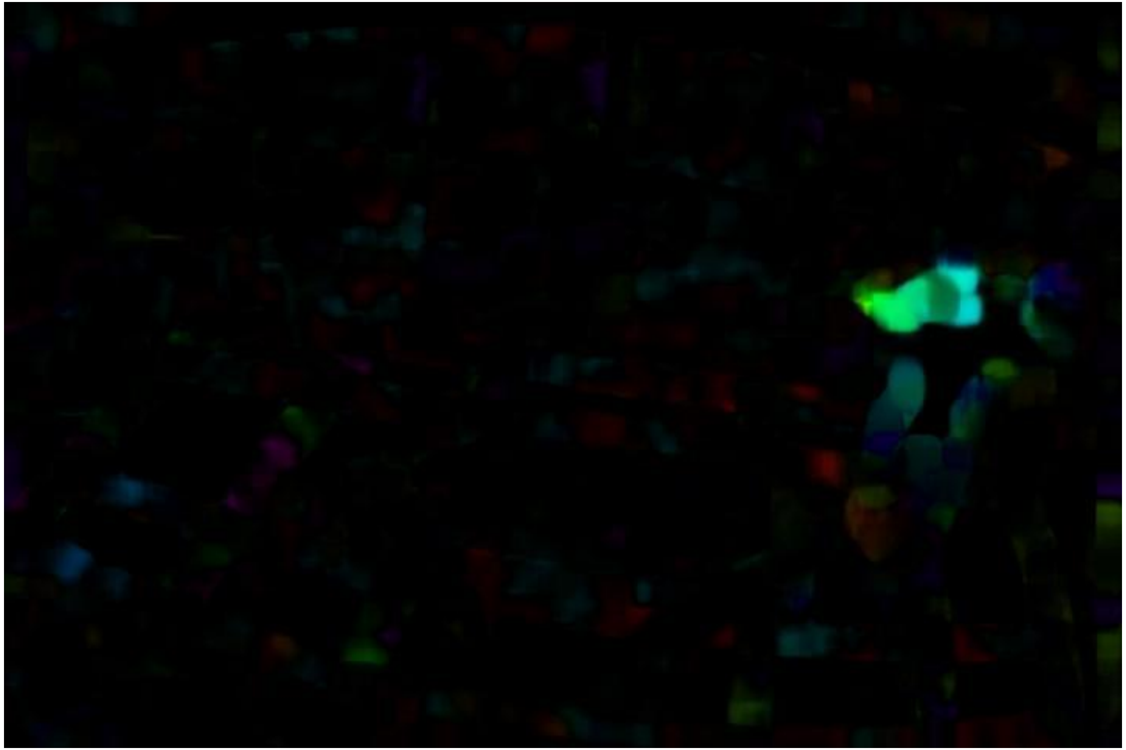
$$f_x u + f_y v + f_t = 0$$

where:

$$f_x = \frac{\partial f}{\partial x} \; ; \; f_y = \frac{\partial f}{\partial x}$$
$$u = \frac{dx}{dt} \; ; \; v = \frac{dy}{dt}$$

Above equation is called Optical Flow equation. In it, we can find f_x and f_y, they are image gradients. Similarly f_t is the gradient along time. But (u, v) is unknown. We cannot solve this one equation with two unknown variables. So several methods are provided to solve this problem and one of them is Lucas-Kanade.

First scenario



Raw video

Dense Optical flow video

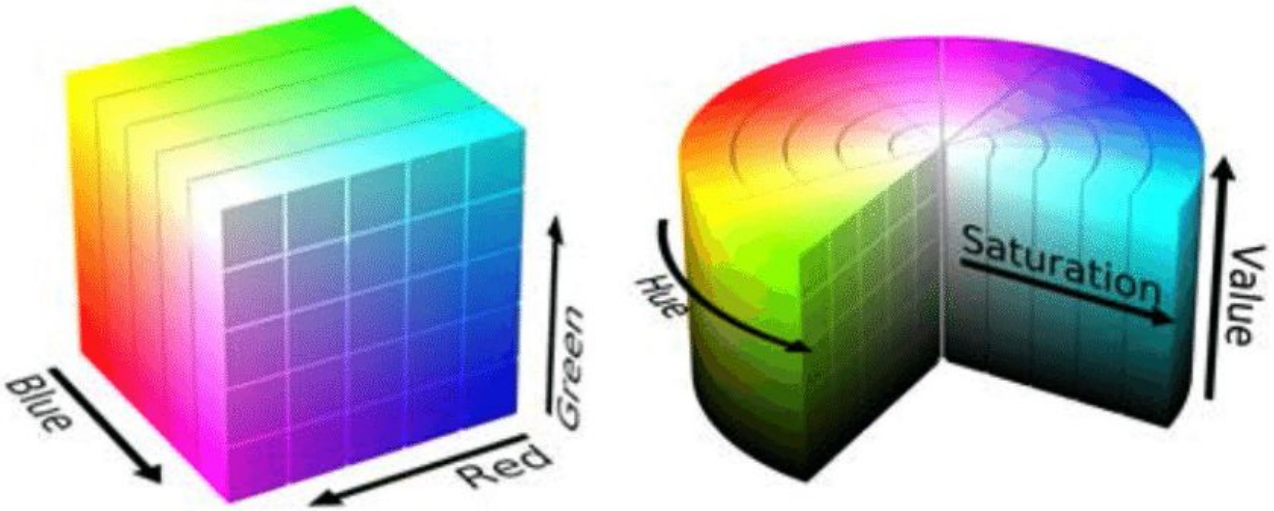
Same scenario but from different angle

Raw video

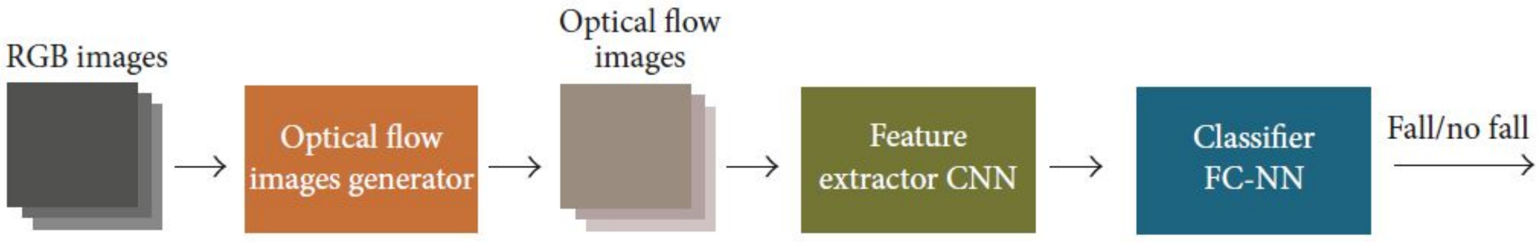
Dense Optical flow video

RGB vs HSV (hue, saturation, value)

- Unlike [RGB](), [HSV]() separates *luma*, or the image intensity, from *chroma* or the color information.
- In computer vision you often want to separate color components from intensity for various reasons, such as robustness to lighting changes, or removing shadows.
- Algorithm less sensitive (if not invariant) to lighting variations



CNN with Optical Flow

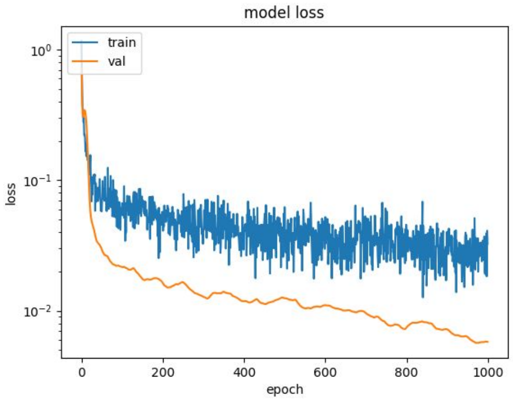


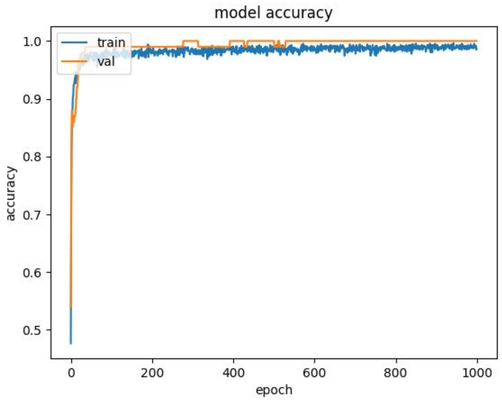
2017 Vision-Based Fall Detection with Convolutional Neural Networks

- Based on VGG-16 network;
- Pretrained with ImageNet;
- Calculate optical flow images of UCF101 action dataset;
- Retrain with optical flow images;
- Data size - 224 X 224 patch, stack size - 20;
- Apply transfer learning on UR Fall Detection Dataset .

Performance Metrics

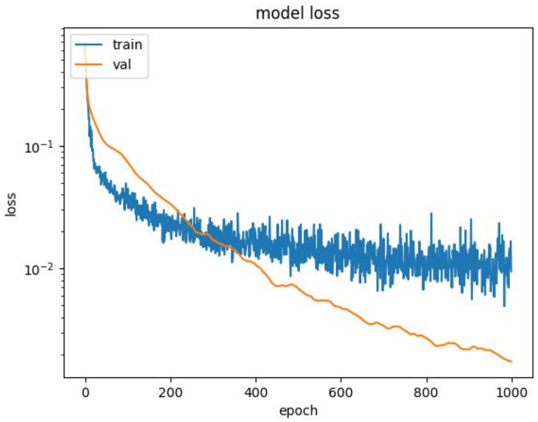
- Accuracy - 99% (+/- 1%)
- TPR - 100% (+/- 1%)
- TNR - 99% (+/ - 1%)
- FPR - 1% (+/ - 1%)
- FNR - 0% (+/ - 1%)

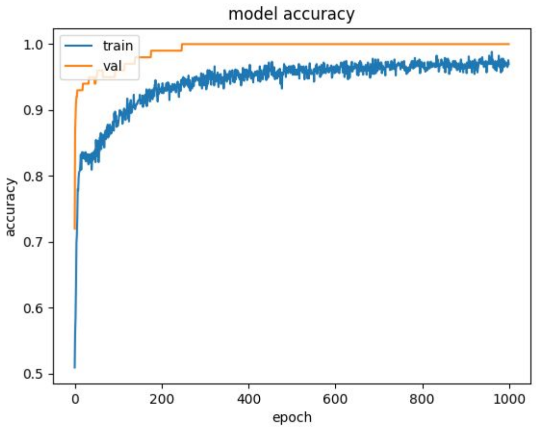


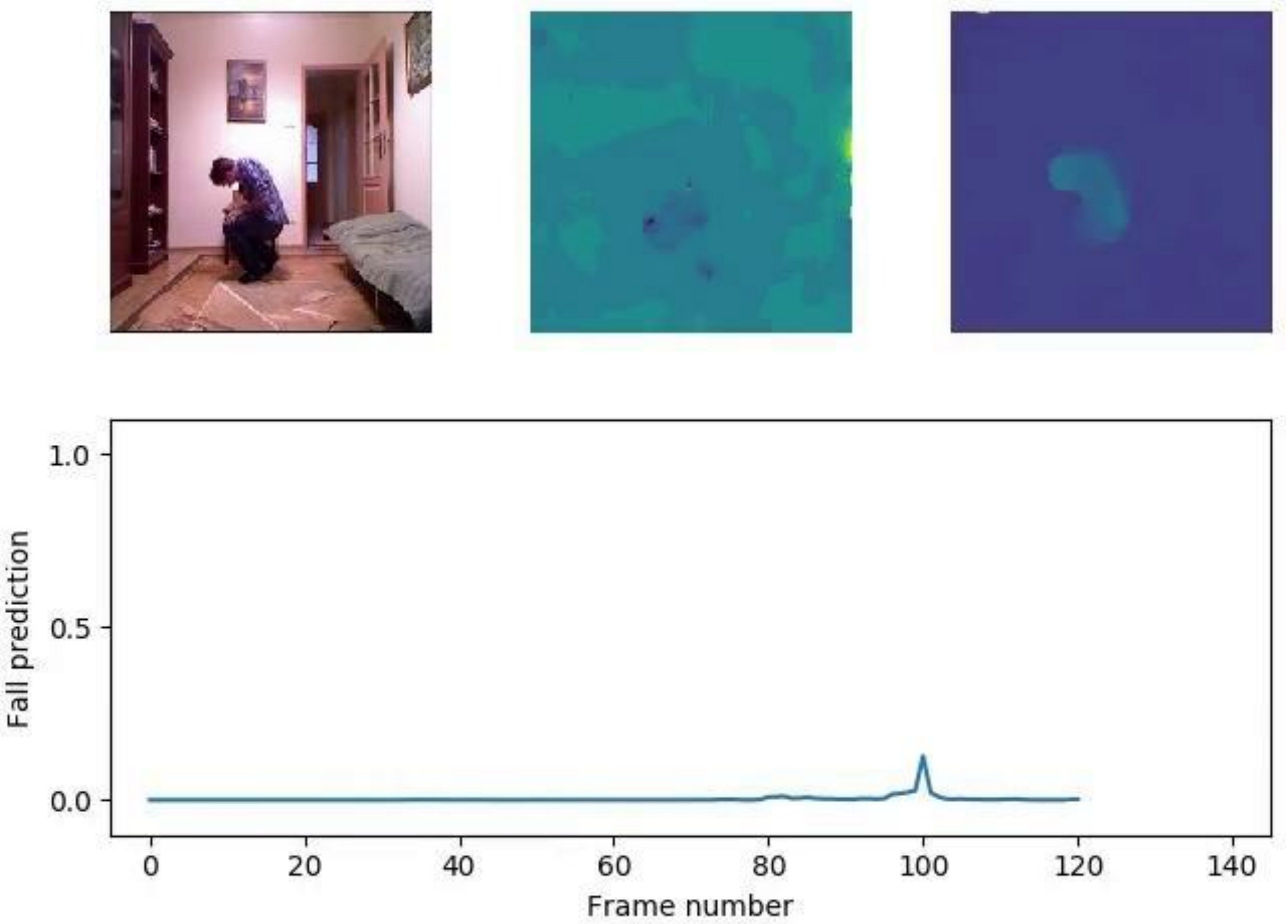


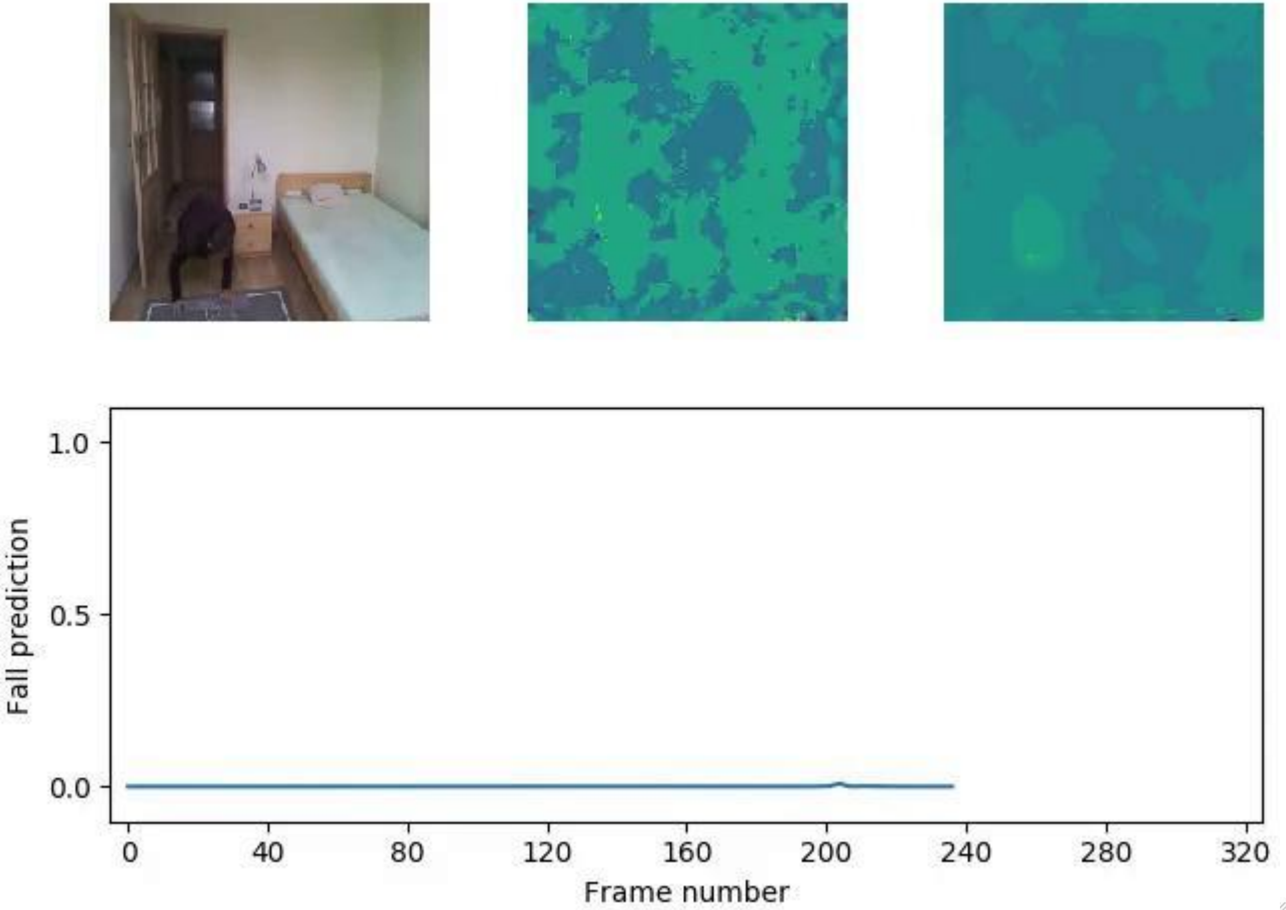
Updated Results on Uncorrelated Train/Test Set (See Appendix: Q and A)

- Accuracy - 99% (+/- 1%)
- TPR - 94% (+/- 3%)
- TNR - 99% (+/- 1%)
- FPR - 1% (+/ - 1%)
- FNR - 6% (+/ - 3%)

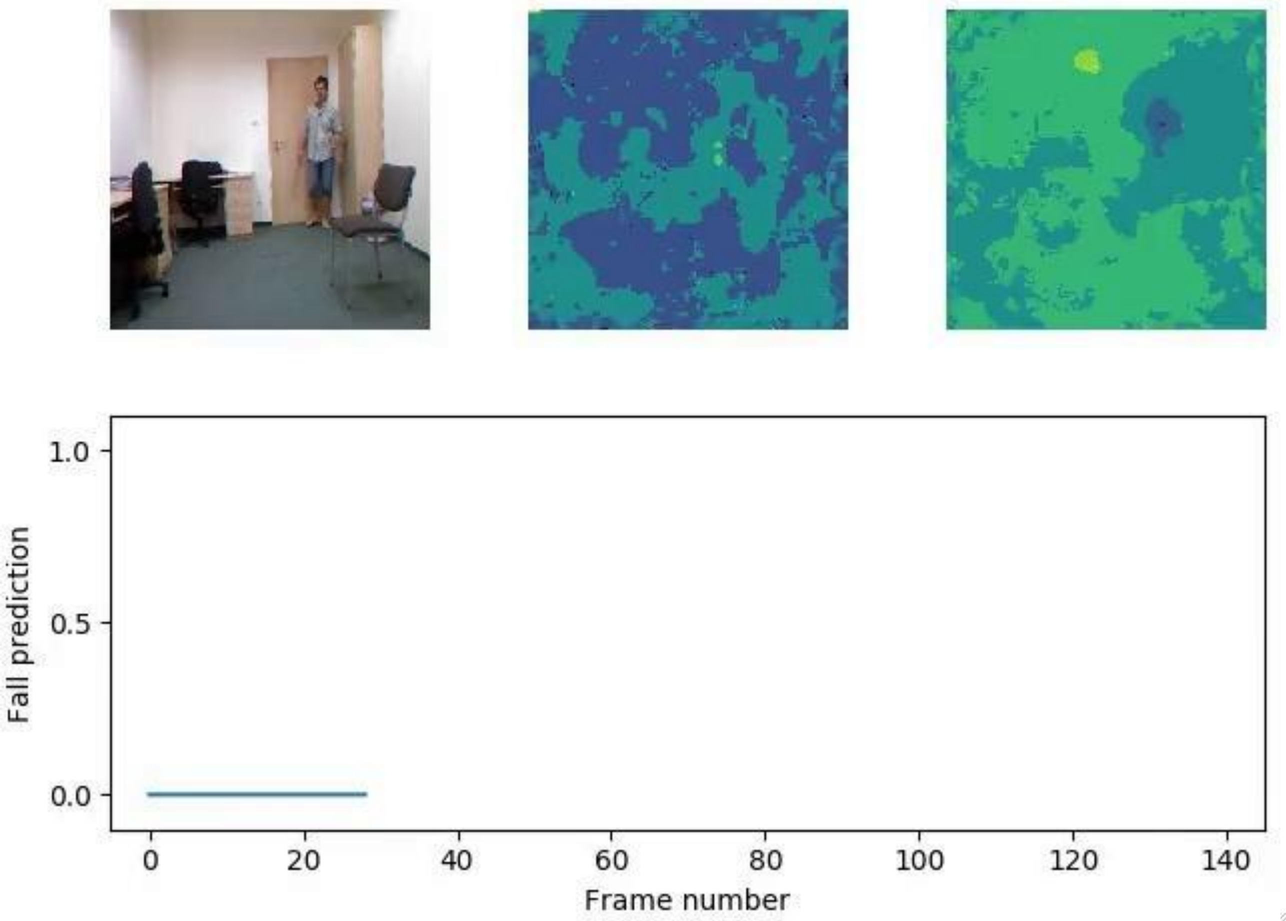





1.0
0.5
0.0
Fall prediction
0
20
40
60
80
100
120
140
Frame number



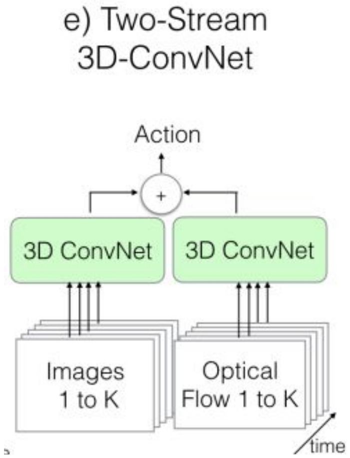
1.0
0.5
0.0
Fall prediction
0
40
80
120
160
200
240
280
320
Frame number

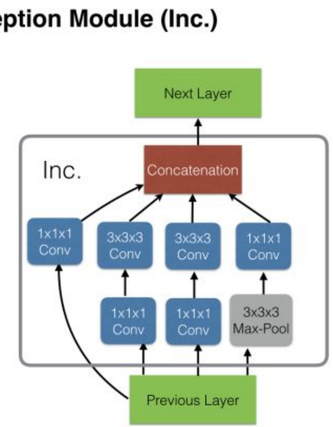

1.0
0.5
0.0
Fall prediction
0
20
40
60
80
100
120
140
Frame number

Two-Stream Inflated 3D ConvNet (I3D)

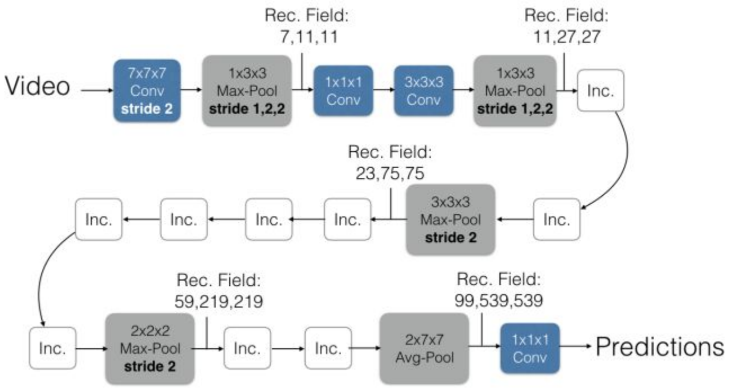
- **Model:**
  - I3D based on Inception v1 [pretrained on ImageNet]
  - filters and pooling kernels of deep image ConvNets are expanded into 3D
  - learns spatio-temporal features extracted from video

- **Pre-training Data:** Kinetics-400. 400 human action classes with more than 400 examples for each class, each from a unique YouTube video.





Inflated Inception-V1

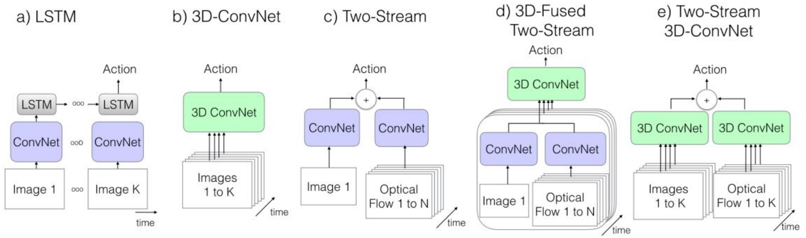


[1] Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset

Two-Stream Inflated 3D ConvNet (I3D)

- **Model:**
  - I3D based on Inception v1 [pretrained on ImageNet]
  - filters and pooling kernels of deep image ConvNets are expanded into 3D
  - learns spatio-temporal features extracted from video

Architecture	UCF-101			HMDB-51			Kinetics		
	RGB	Flow	RGB + Flow	RGB	Flow	RGB + Flow	RGB	Flow	RGB + Flow
(a) LSTM	81.0	–	–	36.0	–	–	63.3	–	–
(b) 3D-ConvNet	51.6	–	–	24.3	–	–	56.1	–	–
(c) Two-Stream	83.6	85.6	91.2	43.2	56.3	58.3	62.2	52.4	65.6
(d) 3D-Fused	83.2	85.8	89.3	49.2	55.5	56.8	–	–	67.2
(e) Two-Stream I3D	**84.5**	**90.6**	**93.4**	**49.8**	**61.9**	**66.4**	**71.1**	**63.4**	**74.2**

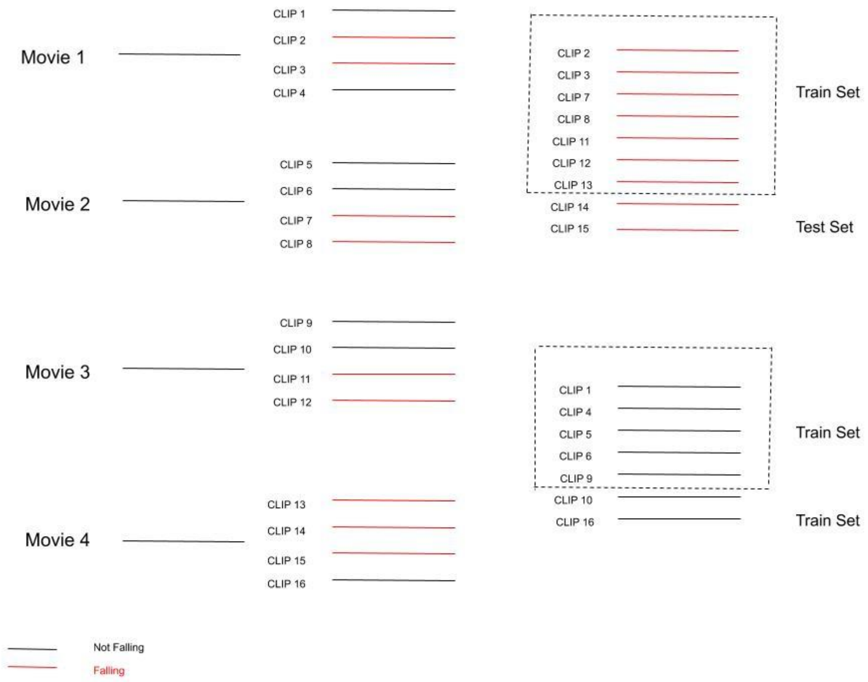


- **Pre-training Data:** Kinetics-400. 400 human action classes with more than 400 examples for each class, each from a unique YouTube video.

[1] Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset

Appendix: Questions and Answers

Q1: For the i3d model, since you split the train/test sets on the clip level (rather than the video level), and a clip in the test set can be a continuation of a clip in train set, aren't there strong correlation between the two data sets?



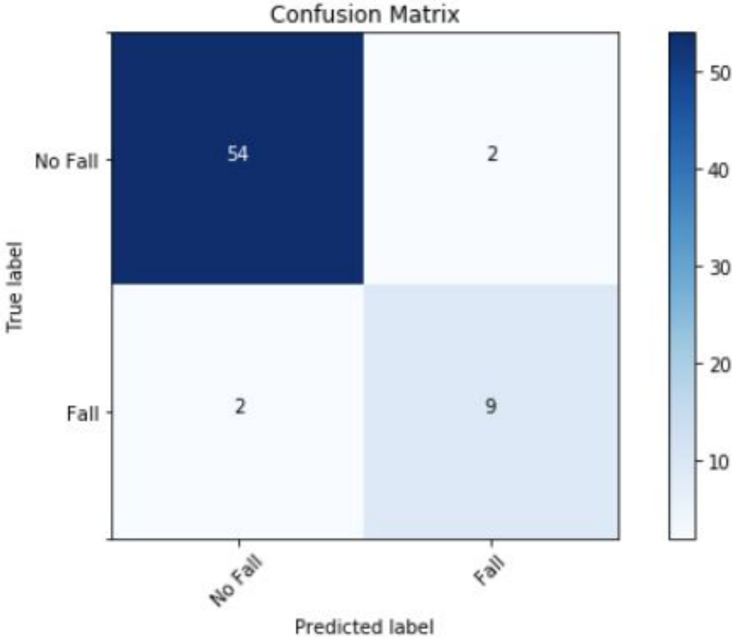
Answer: Yes, a clip in test set can be a continuation of a clip in train set, but this only affects a couple of clips at most. This is illustrated in the diagram above.

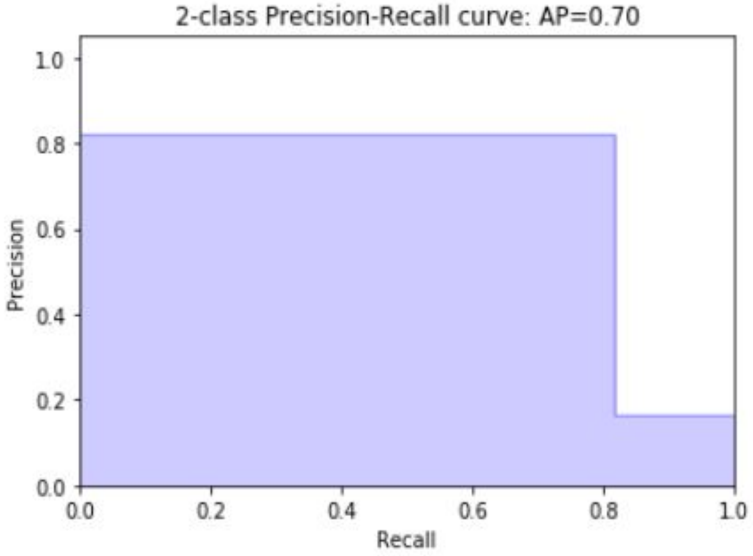
As can be seen in the image above, at most only clips 13,14, and clips 9,10 are continuation of each other in the train/test set respectively. So at most 4 clips (but possibly less) suffer the kind of correlation suggested in the question.

However to fully avoid this, we recreated the train/test sets by first splitting whole movies into two sets and then creating clips. We retrained the model on this non-overlapping sets and got similar results.

Appendix: Updated Results

Accuracy: 94%
Precision: 0.82
Recall: 0.82





Thanks !