INTRUDER DETECTION USING HUMAN BEHAVIOUR ANALYSIS ON FRONT DOOR CAMERA FOOTAGE

Manish Beesetti

DATA PREPROCESSING

- Collected 35 videos of each class (intruder , normal) for training.
- Flipped the videos to increase the dataset to 70 videos in each class.
- **Edited the videos to get the required content for better training.**
- **Converted the videos into frames and used necessary preprocessing techniques to improve training process.**
- **Implement proper frames and folders labeling using scripts.**

Implementation

- Using FFMPEG read all the videos and converted them into frames
- Wrote several scripts to simplify the labeling process and classifying the labeled data.
- Flipped videos horizontally, Increasing training and testing data
- Utilized C3D to classify the given data into labels.

Model

- This is the C3D model that has been used where we can classify multilabel data
- To capture the essence of the scene more deeply every 16 frames of a video are trained as a feature.

Layer (type)	Output Shape	Param #
input_35 (InputLayer)	[(None, 112, 112, 16, 3)]	0
conv3d_170 (Conv3D)	(None, 112, 112, 16, 64)	5248
max_pooling3d_170 (MaxPoolin	(None, 56, 56, 16, 64)	0
conv3d_171 (Conv3D)	(None, 56, 56, 16, 128)	221312
max_pooling3d_171 (MaxPoolin	(None, 28, 28, 8, 128)	0
conv3d_172 (Conv3D)	(None, 28, 28, 8, 128)	442496
max_pooling3d_172 (MaxPoolin	(None, 14, 14, 4, 128)	0
conv3d_173 (Conv3D)	(None, 14, 14, 4, 256)	884992
max_pooling3d_173 (MaxPoolin	(None, 7, 7, 2, 256)	0
conv3d_174 (Conv3D)	(None, 7, 7, 2, 256)	1769728
max_pooling3d_174 (MaxPoolin	(None, 4, 4, 1, 256)	0
flatten_34 (Flatten)	(None, 4096)	0
dense_102 (Dense)	(None, 2048)	8390656
dropout_68 (Dropout)	(None, 2048)	0
dense_103 (Dense)	(None, 2048)	4196352
dropout_69 (Dropout)	(None, 2048)	0
dense_104 (Dense)	(None, 2)	4098
activation_34 (Activation)	(None, 2)	0

Training and Validation Statistics

```
  warnings.warn('`Model.fit_generator` is deprecated and
Epoch 1/16
93/93 [==============================] - 9889s 107s/step - loss: 27.3422 - accuracy: 0.5956 - val_loss: 25.4229 - val_accuracy: 0.9067
Epoch 2/16
93/93 [==============================] - 323s 4s/step - loss: 24.8967 - accuracy: 0.7329 - val_loss: 23.0014 - val_accuracy: 0.8909
Epoch 3/16
93/93 [==============================] - 311s 3s/step - loss: 22.5749 - accuracy: 0.8336 - val_loss: 20.9077 - val_accuracy: 0.9365
Epoch 4/16
93/93 [==============================] - 312s 3s/step - loss: 20.5526 - accuracy: 0.8757 - val_loss: 19.0302 - val_accuracy: 0.9534
Epoch 5/16
93/93 [==============================] - 311s 3s/step - loss: 18.9676 - accuracy: 0.9146 - val_loss: 18.6896 - val_accuracy: 0.9593
Epoch 6/16
93/93 [==============================] - 313s 3s/step - loss: 18.7273 - accuracy: 0.9208 - val_loss: 18.5229 - val_accuracy: 0.9544
Epoch 7/16
93/93 [==============================] - 315s 3s/step - loss: 18.5203 - accuracy: 0.9462 - val_loss: 18.3185 - val_accuracy: 0.9722
Epoch 8/16
93/93 [==============================] - 316s 3s/step - loss: 18.3444 - accuracy: 0.9422 - val_loss: 18.1440 - val_accuracy: 0.9692
Epoch 9/16
93/93 [==============================] - 317s 3s/step - loss: 18.2014 - accuracy: 0.9416 - val_loss: 18.1146 - val_accuracy: 0.9722
Epoch 10/16
93/93 [==============================] - 316s 3s/step - loss: 18.1802 - accuracy: 0.9430 - val_loss: 18.1118 - val_accuracy: 0.9633
Epoch 11/16
93/93 [==============================] - 318s 3s/step - loss: 18.1699 - accuracy: 0.9350 - val_loss: 18.0968 - val_accuracy: 0.9633
Epoch 12/16
93/93 [==============================] - 316s 3s/step - loss: 18.1535 - accuracy: 0.9442 - val_loss: 18.0622 - val_accuracy: 0.9702
Epoch 13/16
93/93 [==============================] - 316s 3s/step - loss: 18.1292 - accuracy: 0.9461 - val_loss: 18.0704 - val_accuracy: 0.9673
Epoch 14/16
93/93 [==============================] - 317s 3s/step - loss: 18.1234 - accuracy: 0.9428 - val_loss: 18.0513 - val_accuracy: 0.9722
Epoch 15/16
93/93 [==============================] - 317s 3s/step - loss: 18.0960 - accuracy: 0.9619 - val_loss: 18.0755 - val_accuracy: 0.9593
Epoch 16/16
93/93 [==============================] - 317s 3s/step - loss: 18.1128 - accuracy: 0.9480 - val_loss: 18.0659 - val_accuracy: 0.9663
```

Results

- Using this model you will be able to detect any signs of intrusion from a video footage.

- There is lot of scope for improvement, given the dataset limitations the model had to be trained with a limited number of footages.

- This model can be implemented even for real time intrusion detection.

Related Works

1. S. Sadanand and J. Corso. Action bank: A high-level representation of activity in video. In CVPR, 2012.
2. P. Scovanner, S. Ali, and M. Shah. A 3-dimensional sift descriptor and its application to action recognition. In ACM MM, 2007.
3. N. Shroff, P. K. Turaga, and R. Chellappa. Moving vistas: Exploiting motion for describing scenes. In CVPR, 2010.
4. K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In NIPS, 2014.
5. K. Soomro, A. R. Zamir, and M. Shah. UCF101: A dataset of 101 human action classes from videos in the wild. In CRCV-TR-12-01, 2012.
6. G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. Convolutional learning of spatio-temporal features. In ECCV, pages 140–153. Springer, 2010.
7. B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places databases. In NIPS, 2014.
8. H. Wang and C. Schmid. Action recognition with improved trajectories. In ICCV, 2013
9. B. Chen, J. A. Ting, B. Marlin, and N. de Freitas. Deep learning of invariant spatio-temporal features from video.
10. H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In Proc. ICCV, pages 1–8, 2007.
11. A Hierarchical Context Model for Event Recognition in Surveillance Video
12. Extensible Hierarchical Method of Detecting Interactive Actions for Video Understanding
13. Action Machine: Rethinking Action Recognition in Trimmed Videos
14. AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions CVPR 2018
15. Large-scale weakly-supervised pre-training for video action recognition CVPR 2019
16. A Closer Look at Spatiotemporal Convolutions for Action Recognition CVPR 2018
17. A Multigrid Method for Efficiently Training Video Models CVPR 2020
18. SlowFast Networks for Video Recognition ICCV 2019
19. Temporal Segment Networks for Action Recognition in Videos 8 May 2017
20. Temporal Action Detection with Structured Segment Networks ICCV 2017

Links

GitHub Link :

https://github.com/ManishBeesetti/Human-Behaviour-Analysis.git

Data Link :

https://drive.google.com/drive/folders/1COd8IOvKmkR27h3WYXX_fRCn_uIzm4Xp?usp=sharing