

DRIVER DROWSINESS DETECTION MINI REPORT 9

BASU VERMA
(142002007)

under the guidance of

Dr Satyajit Das



IIT PALAKKAD

**INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD
PALAKKAD - 678557, KERALA**

Contents

List of Figures	ii
1 Literature review	1
References	5

List of Figures

1.1	sample ARID dataset.	1
1.2	Histogram of ARID dataset (a) and ARID-GIC (b)	2
1.3	RGB and Y histogram of ARID (a) and ARID-GIC(b)	2
1.4	Performance of various 3D-CNN based action recognition models on variants of ARID enhanced by GIC,HE,LIME,BIMEF,KinD	3
1.5	Comparison of the sampled frames and their respective RGB histograms from (a) ARID, (b) ARID-GIC, (c) ARID-HE, (d) ARID-LIME, (e) ARID-BIMEF and (f) ARID-KinD	3
1.6	Comparison of sampled frames and their corresponding CAMs of classes: (a) Jumping and (b) Standing, extracted by utilizing 3D-ResNext-101 model. The sampled frames and their CAMs are from (i) ARID, (ii) ARID-GIC, (iii) ARID-HE, (iv) ARID-LIME, (v) ARID-BIMEF and (vi) ARID-KinD.	4

Part. 1

Literature review

[1] A comprehensive study on recognizing actions in the dark and a new benchmark dataset.

- Datasets used in this paper are Action recognition in the dark (ARID) and HMDB-51-dark datasets.



Fig. 1.1 sample ARID dataset.

- ARID dataset is focused on human actions in dark videos containing a total of 3,784 video clips with each class containing at least 110 clips.
- The video clips are fixed at 30 FPS with 320x240 resolution. The minimum clip length is 1.2 seconds with 36 frames. Videos are in .avi format.

- To obtain action recognition results on ARID dataset, they used two-stream model and 3D-CNN based model. Input to 3D-CNN model are sequence of 16 sampled frames, each of size 224x224. The input to spatial stream of two-stream models are RGB sampled frames resized to 224x224.
- For better recognition of dark videos, some frame enhancement methods are applied. These are:-
 - Histogram Equalization (**HE**) - Used to produce higher contrast images.
 - Gamma Intensity Correction (**GIC**) - Used to adjust the luminance of images.
 - **LIME** - Estimates the illumination map of dark images while imposing a structure prior to the initial illumination map.
 - **BIMEF** - It's a multi-exposure fusion algorithm.
 - **KinD** - Deep neural network based method utilizing a two-stream structure for simultaneous reflectance restoration and illumination adjustment.

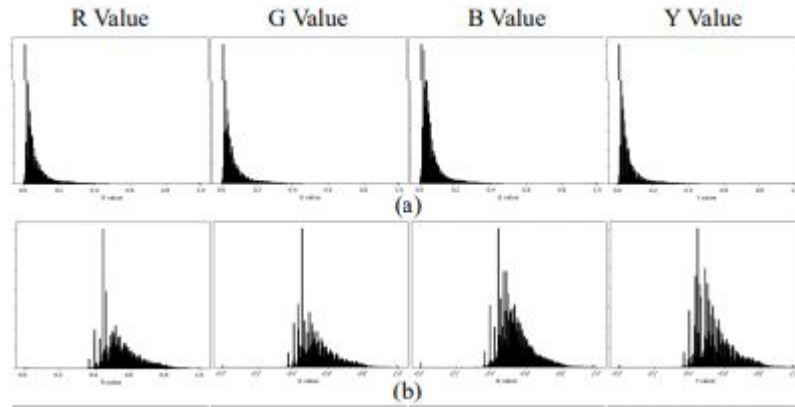


Fig. 1.2 Histogram of ARID dataset (a) and ARID-GIC (b)

- GIC method shows that pixel value would shift towards regions of larger values quite significantly.

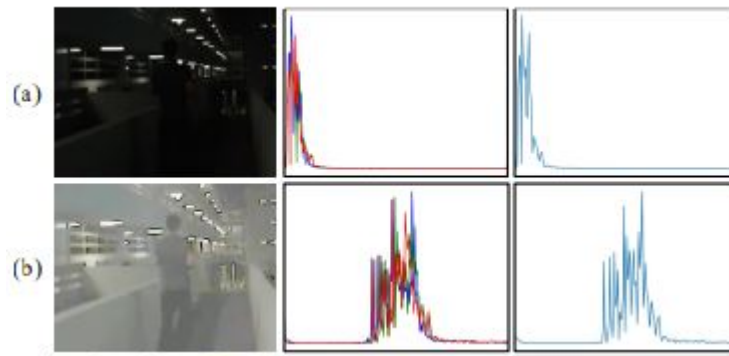


Fig. 1.3 RGB and Y histogram of ARID (a) and ARID-GIC(b)

- Though enhanced frames are clearer visually, some enhancements break the original distribution of videos and introduce noise. The change in distribution and introduction of noise could lead to a decrease in performance for action recognition models.

Datasets	Accuracy	C3D	I3D-RGB	3D-ResNet-101	3D-ResNext-101
ARID-GIC	Top-1	44.09%	69.14%	75.15%	78.06%
	Improv.	3.75%	0.85%	3.58%	3.33%
ARID-HE	Top-1	39.49%	63.67%	65.49%	75.82%
	Improv.	-0.85%	-4.62%	-6.08%	1.09%
ARID-LIME	Top-1	39.61%	73.02%	75.45%	77.40%
	Improv.	-0.73%	4.73%	3.88%	2.67%
ARID-BIMEF	Top-1	45.23%	68.89%	68.28%	73.39%
	Improv.	4.89%	0.60%	-3.29%	-1.34%
ARID-KinD	Top-1	46.64%	67.55%	70.59%	69.62%
	Improv.	6.30%	-0.74%	-0.98%	-5.11%
ARID	Top-1	40.34%	68.29%	71.57%	74.73%

Fig. 1.4 Performance of various 3D-CNN based action recognition models on variants of ARID enhanced by GIC,HE,LIME,BIMEF,KinD

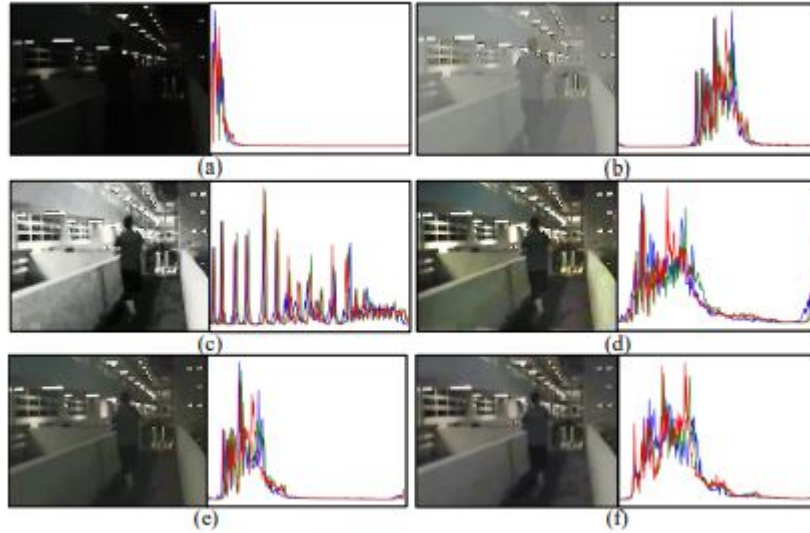


Fig. 1.5 Comparison of the sampled frames and their respective RGB histograms from (a) ARID, (b) ARID-GIC, (c) ARID-HE, (d) ARID-LIME, (e) ARID-BIMEF and (f) ARID-KinD

- The visualizations of features are presented as Class Activation Maps (CAM) which depicts the focus of the model with respect to the given prediction. . Specifically, CAMs are extracted by utilizing the 3D-ResNext-101 model first, due to the best performance achieved by the 3D-ResNext-101 on ARID datasets.

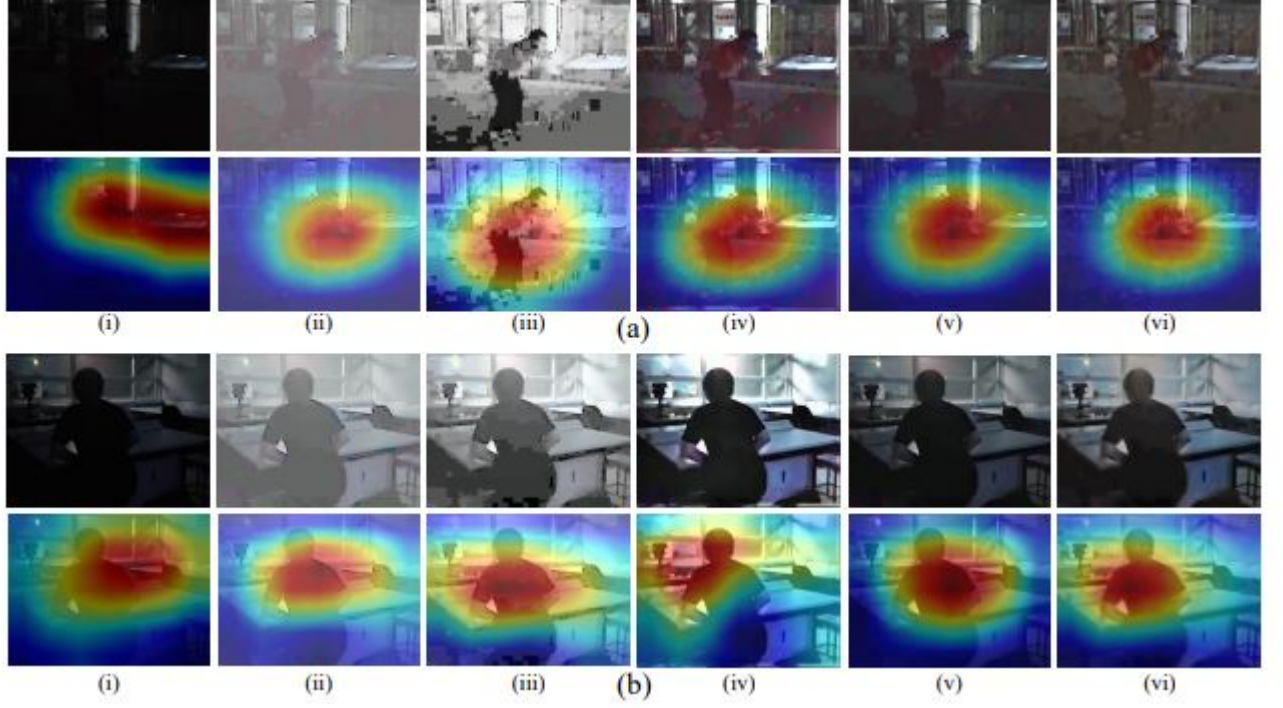


Fig. 1.6 Comparison of sampled frames and their corresponding CAMs of classes: (a) Jumping and (b) Standing, extracted by utilizing 3D-ResNext-101 model. The sampled frames and their CAMs are from (i) ARID, (ii) ARID-GIC, (iii) ARID-HE, (iv) ARID-LIME, (v) ARID-BIMEF and (vi) ARID-KinD.

References

- [1] Y. Xu, J. Yang, H. Cao, K. Mao, J. Yin, and S. See, “Arid: A comprehensive study on recognizing actions in the dark and a new benchmark dataset,” 2021.