DRIVER DROWSINESS DETECTION
MINI REPORT 1

BASU VERMA

(142002007)

under the guidance of

Dr Satyajit Das

IIT PALAKKAD

INDIAN INSTITUTE OF TECHNOLOGY PALAKKAD
PALAKKAD - 678557, KERALA

Contents

Part. 1

Models

Models	**Feature Extraction (CV Technique)**	**Metrics**	**Accracy**
CNN [1]	AdaBoost and PERCLOS	Eye State	95.18%
CNN with softmax [2]	Viola and jones algorithm	Visual feature	78%
CNN [3]	Viola and jones algorithm	Eye gaze	98.32%
SVM [1]	-	Eye State (video with 15fps)	93.5%

Part. 2

Self-Supervised Learning

It is the process of training unlabelled data. Self-supervised learning has gained popularity because of its ability to avoid the cost of annotating large-scale datasets. It is capable of adopting self-defined pseudo labels as supervision and use the learned representations for several downstream tasks. It aims at embedding augmented versions of the same sample close to each other while trying to push away embedding's from different samples.

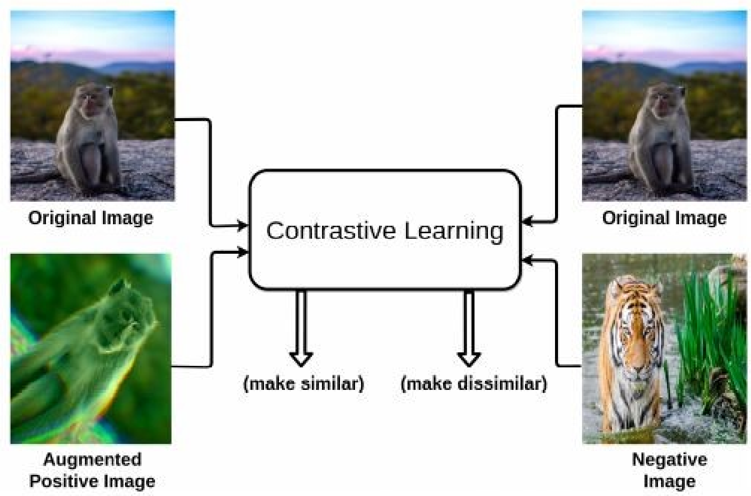

Fig. 2.1 Basic Intution behind contrastive learning paradigm: push original and augmented images closer and push original and negatives images away[4]

Self-supervised tasks / pretext task act as an important strategy to learn representations of the data using pseudo labels. These pseudo labels are generated automatically based on the attributes found in the data. The learned model from the pretext task can be used for any downstream tasks such as classification, segmentation, detection, etc. in computer vision.

For instance, one sample from the training dataset is taken and a transformed version of the sample is retrieved by applying appropriate data augmentation techniques. During training, the augmented version of the original sample is considered as a positive sample, and the rest of the samples in the batch/dataset (depends on the method being used) are considered negative samples. Next, the model is trained in a way that it learns to differentiate positive samples from the negative ones. The differentiation is achieved with the help of some pretext task. In doing so, the model learns quality representations of the samples and is used later for transferring knowledge to downstream tasks.

2.0.1 Pretext task

A video contains a sequence of semantically related frames. This implies that frames that are nearby with respect to time are closely related and the ones that are far away are less likely to be related. Intuitively, the motive for using such an approach is, solving a pretext task that allows the model to learn useful visual representations while trying to recover the temporal coherence of a video. Here, a video with shuffled order in the sequence of its image frames acts as a positive sample while all other videos in the batch/dataset would be negative samples.

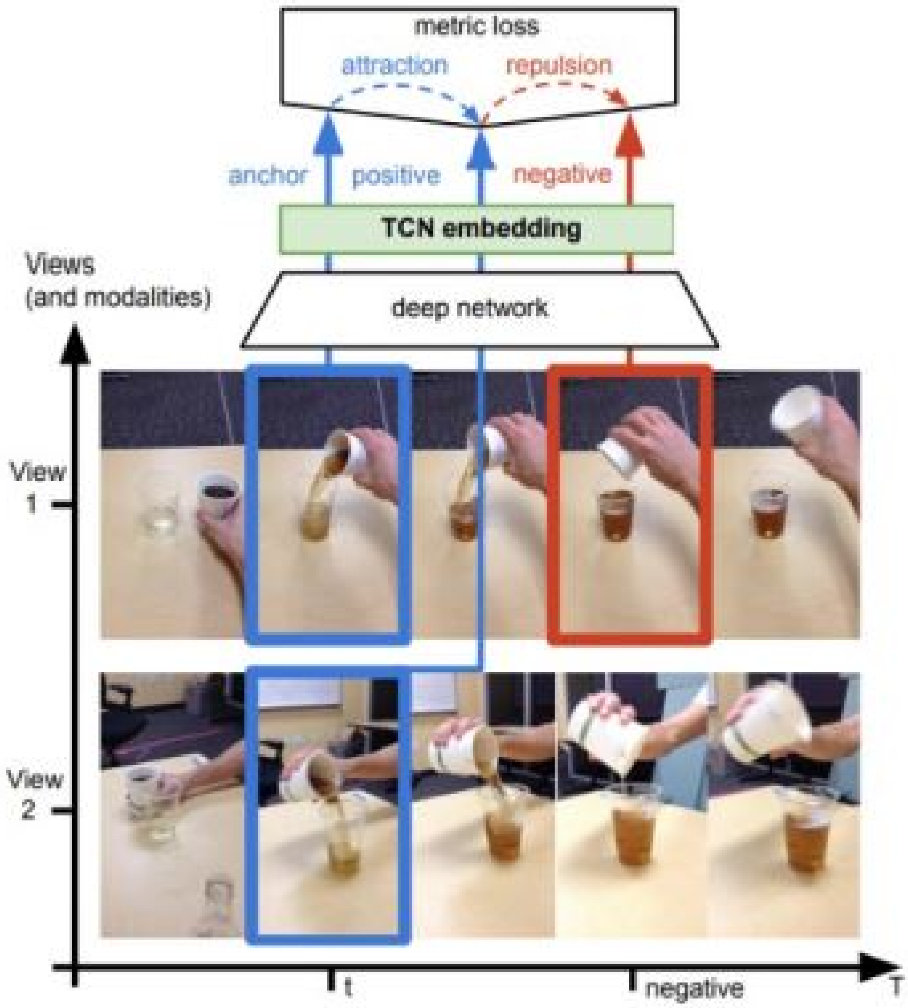

Fig. 2.2 Learning representation from video frame sequence [4]

Part. 3

Datasets

3.0.1 Drozy database

[5] The "ULg Multimodality Drowsiness Database", also called DROZY, is a database containing various types of drowsiness-related data (signals, images, etc.).

3.0.2 Yawn Detection Dataset

[6] Two video datasets of drivers with various facial characteristics, to be used for designing and testing algorithms and models for yawning detection. In the first dataset, the camera is installed under the front mirror of the car. Each participant has three or four videos and each video contains different mouth conditions such as normal, talking/singing, and yawning. In the second dataset, the camera is installed on the dash in front of the driver, and each participant has one video with the above-mentioned different mouth conditions all in the same video

Figure 1. Some Participants and their angle of capture; top: camera under the mirror; bottom: camera on the dash.

3.0.3 Eye-Chemera dataset

[7] Frames extracted are the relevant ones from videos recorded either with Canon 600D (640 x 480 resolution), either with Panasonic HDC-TM 60 (1920 X 1080) cameras on 40 subjects in the 20-30 age range. Illumination is typical-indoor and the light comes from one side.

3.0.4 NTHU drowsey driver detection dataset

The training dataset contains 18 subjects with 5 different scenarios (BareFace,Glasses, Night_BareFace, Night_Glasses, Sunglasses). The video resolution is 640x480 in AVI format. The videos of Night_BareFace, Night_Glasses scenario were captured at 15 frames per second; BareFace, Glasses, Sunglasses scenario were 30 frames per second.

Same scenario, different behavior

Same behavior, different scenarios

References

[1] M. Ngxande, J.-R. Tapamo, and M. Burke, "Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques," 11 2017, pp. 156–161.

[2] K. Dwivedi, K. Biswaranjan, and A. Sethi, "Drowsy driver detection using representation learning," in *2014 IEEE International Advance Computing Conference (IACC)*, 2014, pp. 995–999.

[3] A. George and A. Routray, "Real-time eye gaze direction classification using convolutional neural network," 2016.

[4] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," 2021.

[5] L. T. F. C. V. J. Massoz, Quentin, "The ulg multimodality drowsiness database (called drozy) and examples of use," in *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. New York, NY, USA: IEEE Winter Conference on Applications of Computer Vision (WACV 2016), 2016. [Online]. Available: https://orbi.uliege.be/handle/2268/191620

[6] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "Yawdd: A yawning detection dataset," in *Proceedings of the 5th ACM Multimedia Systems Conference*, ser. MMSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 24–28. [Online]. Available: https://doi.org/10.1145/2557642.2563678

[7] C. F. research, "Eyechemera dataset." [Online]. Available: http://imag.pub.ro/common/staff/cflorea/cf_research.html

[8] CVlab, "Driver driwsiness detection dataset," 2016. [Online]. Available: http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/