**Báo cáo hàng tuần luận văn**

**Đề tài**:

Nghiên cứu các phương pháp học sâu ứng dụng trong phân tích hành vi người lái xe ô tô

**09/04/2023:**

**Công việc đã thực hiện**: em đã tải xong toàn bộ 95 GB dữ liệu và viết tài liệu mô tả, thống kê dữ liệu ở bên dưới.

* Link tải dữ liệu: <https://drive.google.com/drive/folders/1TVzNODlKiRKj60cwadnXofEsBkCDl1DZ>
* Link bài báo pulish dataset: <https://arxiv.org/pdf/2009.14660.pdf>

1. **Dữ liệu**
   1. **Thông tin chung**

* Tên tập dữ liệu: Driver Anomaly Detection Dataset (DAD)
* Đơn vị phát hành: The Institute for Human-Machine Communication (MMK), Technische Universität München (TUM)
* Kích thước bộ dữ liệu: 95 GB
* Loại dữ liệu: ảnh .png . Anh được bóc ra từ video 45fps, độ phân giải 224x171 pixels.
* Góc nhìn: trong cabin, từ trên xuống (top) và phía trước mặt người lái xe (front)
* Gồm có ảnh biểu diễn chiều sâu (depth) và ảnh hồng ngoại (IR)

Graphical user interface, website

Description automatically generated

* Bản quyền: chỉ sử dụng cho mục đích nghiên cứu
  1. **Phân chia bộ dữ liệu:**
* ***Nhãn:*** 
  + Tập trung (Normal)
  + Mất tập trung (Abnormal):
    - Có trong cả tập train và test:
      * Adjusting radio
      * Drinking
      * Messaging left
      * Messaging right
      * Reaching behind
      * Talking with passenger
      * Talking with phone left
      * Talking with phone right
    - Chỉ có trong tập test:
      * Điều chỉnh gương đầu xe
      * Chỉnh quần áo
      * Chỉnh kính đeo mắt
      * Chỉnh gương chiếu hậu
      * Chỉnh cửa sổ trời
      * Lau mũi
      * Gục đầu
      * Đang ăn
      * Đeo kính
      * Cởi kính
      * Lấy đồ vật
      * Lau mồ hôi
      * Sờ tóc, sờ mặt
      * Hắt hơi
      * Ho
      * Đang đọc
* ***Train/Test:***
  + Tập train: chiểm 83% kích thước tập dữ liệu
    - Gồm 25 đối tượng được ghi hình
    - Mỗi đối tượng có 6 videos lái xe trong tình trạng tập trung (normal) và 8 videos lái xe trong tình trạng không tập trung (abnormal).
    - Mỗi video tập trung dài 3.5 phút, mỗi video không tập trung dài 30 giây.
    - Tổng cộng 550 phút tập trung và 100 phút không tập trung
  + Tập test: chiếm 17% kích thước tập dữ liệu
    - Gồm 6 đối tượng
    - Mỗi đối tượng có 6 video, mỗi video dài 3.5 phút
    - Các tình huống mất tập trung xảy ra ngẫu nhiên trong video.
    - Ngoài 8 loại mất tập trung (abnormal) bên trên, còn có thêm 16 loại mất tập trung khác mà không xuất hiện trong tập training
    - Tổng cộng 88 phút lái xe trong tình trạng tập trung và 45 lái xe mất tập trung.

**Chart

Description automatically generated with low confidence**

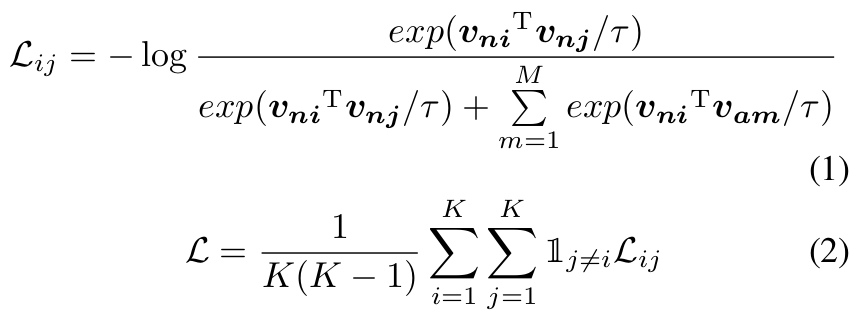
* ***Nhận xét:*** 
  + Dữ liệu khá mất cân bằng giữa 2 nhãn tập trung (normal) và mất tập trung (abnormal).
  + Tỉ lệ phân chia 2 tập train-test là khá phù hợp.
  + Tập test có chứa thêm các tình huống mất tập trung khác mà không có ở tập train, nên khá phù hợp để đánh giá mức độ phổ quát (generalization) của mô hình.
  1. **Các kỹ thuật làm giàu dữ liệu có thể xem xét:**
* Multi-scale random cropping
* Salt and pepper noise
* Random rotation
* Random horizontal flip
  1. **Các kỹ thuật xử lý mất cân bằng dữ liệu có thể xem xét:**
* Dữ liệu: upsampling
* Hàm loss:
  + Focal loss
  + Weighted cross entropy loss

TODO:

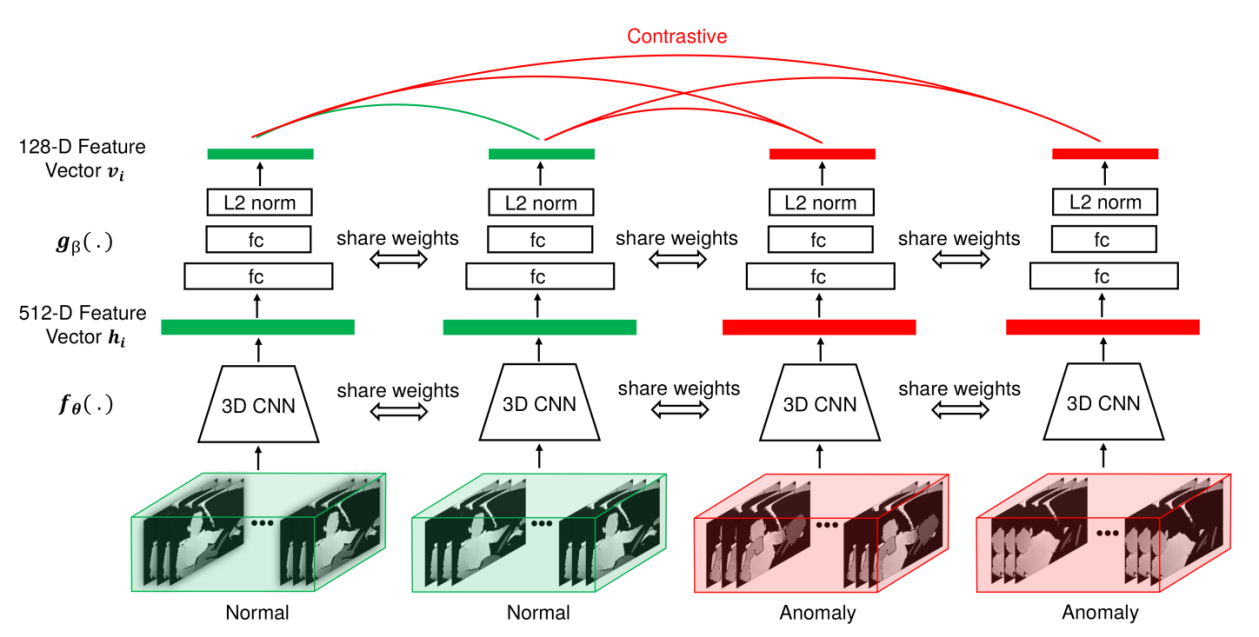
* Hiểu source code/paper
* Chạy
* Problem từ code/paper
* Hiểu figure số 4: how hàm f, g share trọng số , contrastive loss
* Overleaf latex

**04/05/2023:**

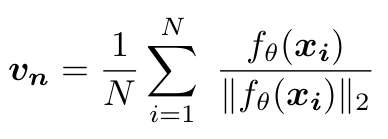
1. **Cài đặt và chạy source code Driver Anomaly Detection:**
   1. **Cài đặt môi trường:**
      * Conda env
      * Các thư viện cần thiết
   2. **Chạy:**
      * Fixed các bugs không tương thích môi trường phần cứng & thư viện.
      * Train from scratch: Done
      * Train from checkpoint: Done
      * Test on val set: Done
2. **Tìm hiểu về kiến trúc mô hình được giới thiệu trong paper:** 
   1. **Contrastive learning framework:**
      * Base encoder f
        + 3D-CNN
        + Resnet, Shufflenet…
      * Projection head g
        + MLP with one hidden layer.
        + Embedding vector được chuẩn hóa L2
      * Contrastive loss
        + K clips tập trung và M clips mất tập trung trong mỗi minibatch
        + Công thức hàm loss:



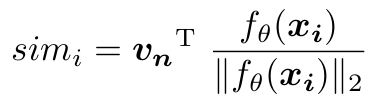
* + - * Noise contrastive estimation được sử dụng thay thế cho hàm non-parametric softmax



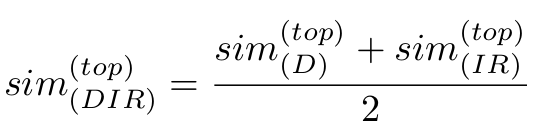
* 1. **Test time recognition**
     + Bỏ projection head chỉ sử dụng 3D-CNN backbone để mã hóa chuỗi frame (clip) thành các vector 512 chiều được chuẩn hóa L2.
     + Tính toán vector đại diện cho normal bằng trung bình cộng của tất các cả các vector normal



* + - Tính cosine distance giữa vector đại diện normal và clip cần phân loại, nếu nhỏ hơn ngưỡng gamma thì là bất thường



* 1. **Dung hợp 4 model trên 4 tổ hợp của views + modalities:**
     + Mỗi view, modality được train với mô hình riêng biệt
     + Sau đó chúng được dung hợp (fused) với nhau ở bước tính toán cosine similarity score bằng cách lấy trung bình ứng với mỗi view



TODO:

* Chọn minibatch (random or not)
* ***Tìm hiểu về tau trong paper face…***
* ***Tính chất của vector vn (sao lại dùng khoảng cách l2) , visualize …***
* ***Hiểu sâu hơn về đặc trưng của vni và vai: sử dung tSNE***
* Reproduce
* Post processing
* Thay bằng focal loss
* Tìm thêm dataset khác

**11/05/2023:**

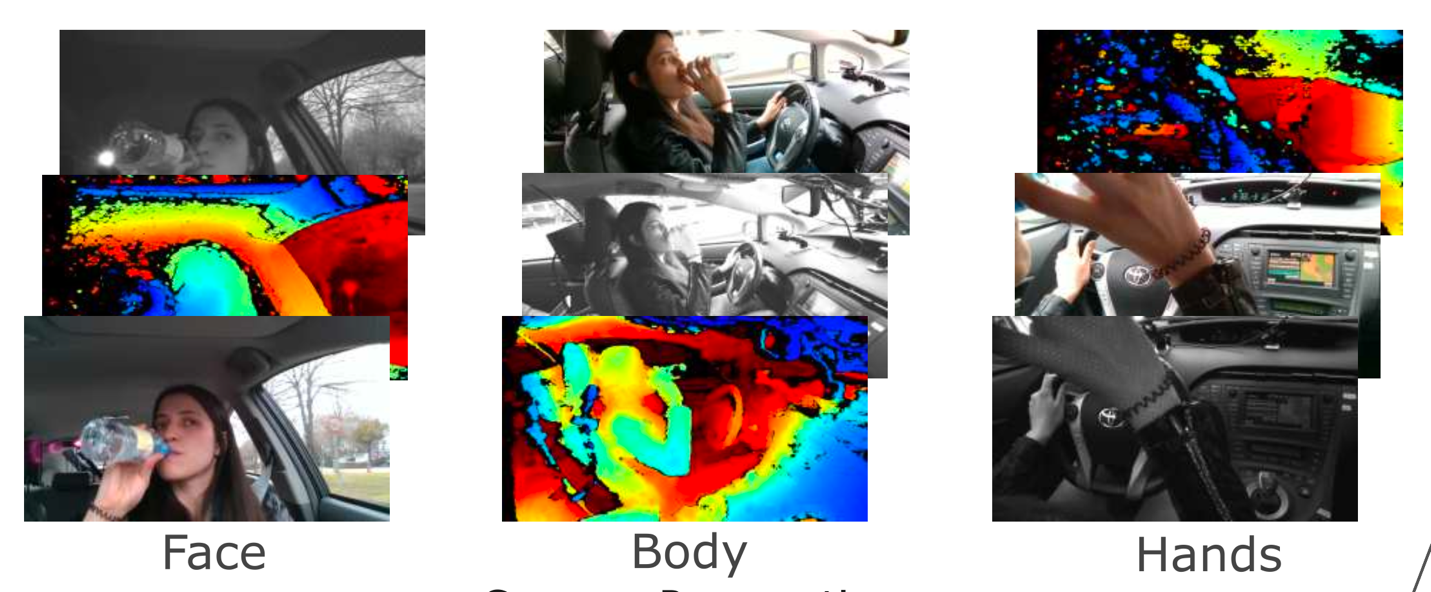
* 1. Một số vấn đề cần làm rõ:
* Minibatch
  + Minibatch đc lấy ngẫu nhiên từ tập dữ liệu
  + Load normal data và abnormal data riêng biệt
  + Từ normal clip data, trộn và lấy ra 1 minibatch = K clips
  + Từ abnormal clip data, trộn và lấy ra 1 minibatch = M clips
  + Minibatch = K + M
* Tau?
* Công cụ visualization:
  + sử dụng tSNE
* Chạy lại Evaluation trên val set:
  + DONE => Kết quả đúng như paper.
* Post processing:
  + Lấy trung bình score dự đoán của score clip hiện tại với k clip liền trước để ra dự đoán cuối cùng cho clip (1 clip = 16 frames)
  + Chọn k=6 cho kết quả tốt nhất trên val set => overfit val set
* Dataset khác:
  + Driver&Act
    - Public to download for researching
    - Views:
      * Steering wheel:A person in a car

        Description automatically generated with low confidence
      * CeilingA person in a car

        Description automatically generated with medium confidence
      * A-column co-driverA person in a car

        Description automatically generated with medium confidence
      * A-column driver
      * Center miror A person sitting in a car

        Description automatically generated with medium confidence
    - Modals:
      * NIR (tất cả các views)
      * IR, Depth, RGB (A-column co-driver)
  + DMD:
    - Contact for download
    - Views:
      * Face, Body, Hand
    - Modals:
      * IR, Depth, RGB



* Thử nghiệm với hàm loss khác:
  + Focal loss