ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA ĐIỆN – ĐIỆN TỬ
BỘ MÔN VIỄN THÔNG

Luận Văn Tốt Nghiệp

Mạng Tích Chập Sâu – Nhận Dạng Hành Động Con Người

Sinh Viên Thực Hiện:
Đặng Lê Anh Khoa – 1511561
Nguyễn Khắc Trung Tín – 1513489

Giảng viên hướng dẫn:
PGS.TS. Hà Hoàng Kha

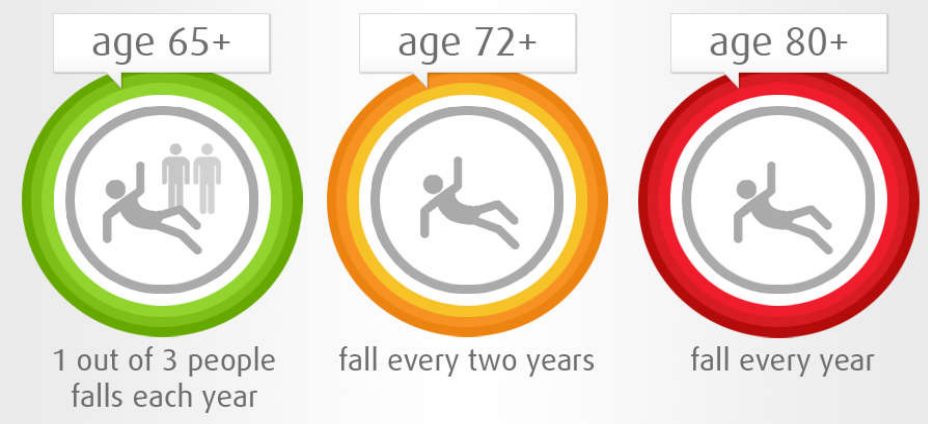
Nội dung trình bày

1. Tổng quan đề tài
2. Lý thuyết học sâu (Deep Learning)
3. Huấn luyện mô hình
4. Nhận dạng hành động con người
5. Kết luận – Hướng phát triển

1 Tổng quan đề tài

Lý do thực hiện đề tài

Sự phát triển vượt bậc của AI



Tỷ lệ té ngã ở người cao tuổi

Thống kê tỷ lệ bạo lực học đường ở Việt Nam

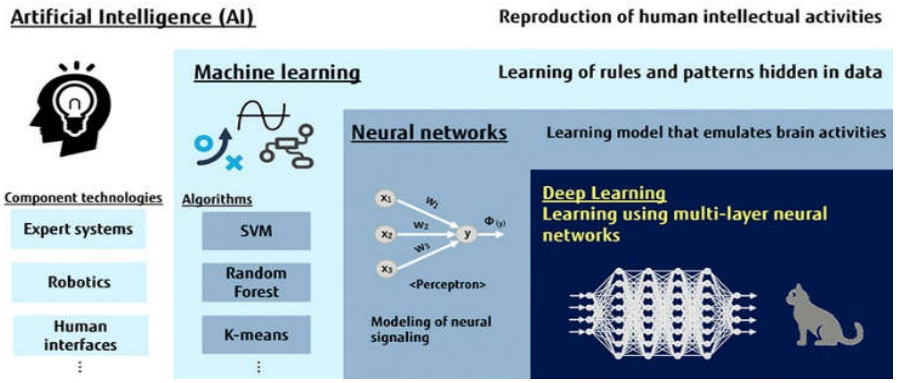
Đề tài nhận dạng hành động con người

Mục tiêu đề tài

- Tìm hiểu lý thuyết về Deep Learning, Deep Convolutional Network.
- Tìm hiểu các thuật toán cho bài toán nhận dạng hành động.
- Tìm hiểu bộ dữ liệu AVA, xây dựng mô hình nhận dạng hành động trên bộ dữ liệu AVA.
- Đánh giá mô hình, hướng phát triển, kết luận.

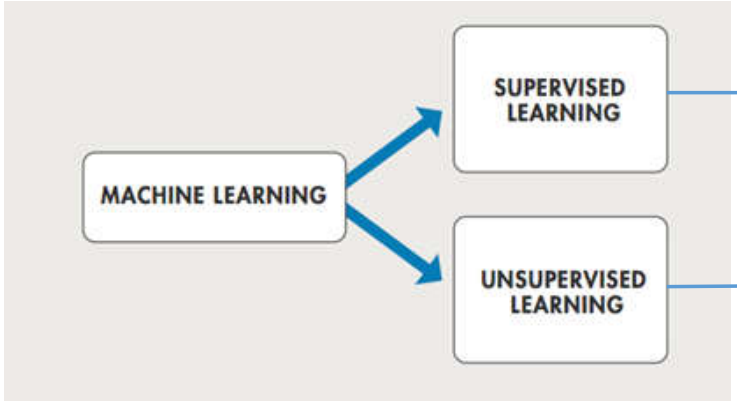
2 Lý thuyết học sâu (Deep Learning)

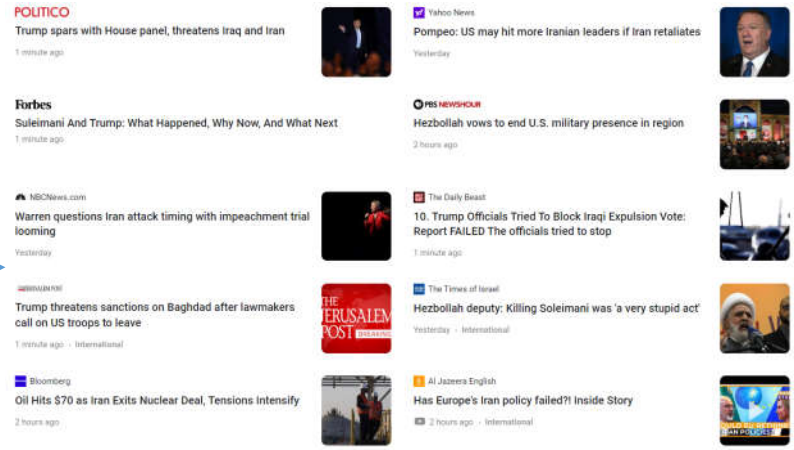
2.1 Học Máy – Machine Learning



Mối quan hệ giữa AI, Machine Learning, Deep Learning

Con người

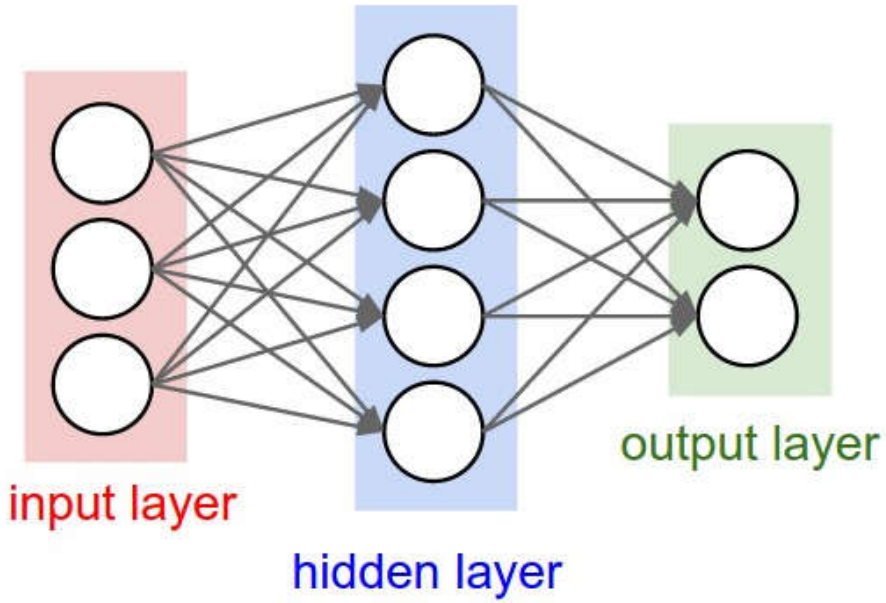




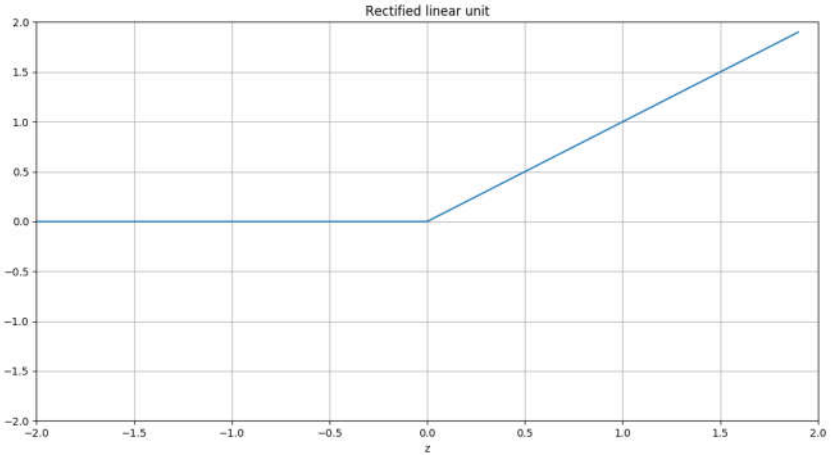
Machine learning được phân loại thành học có giám sát và học không giám sát

2.2 Mạng tích chập sâu

Lớp kết nối đầy đủ

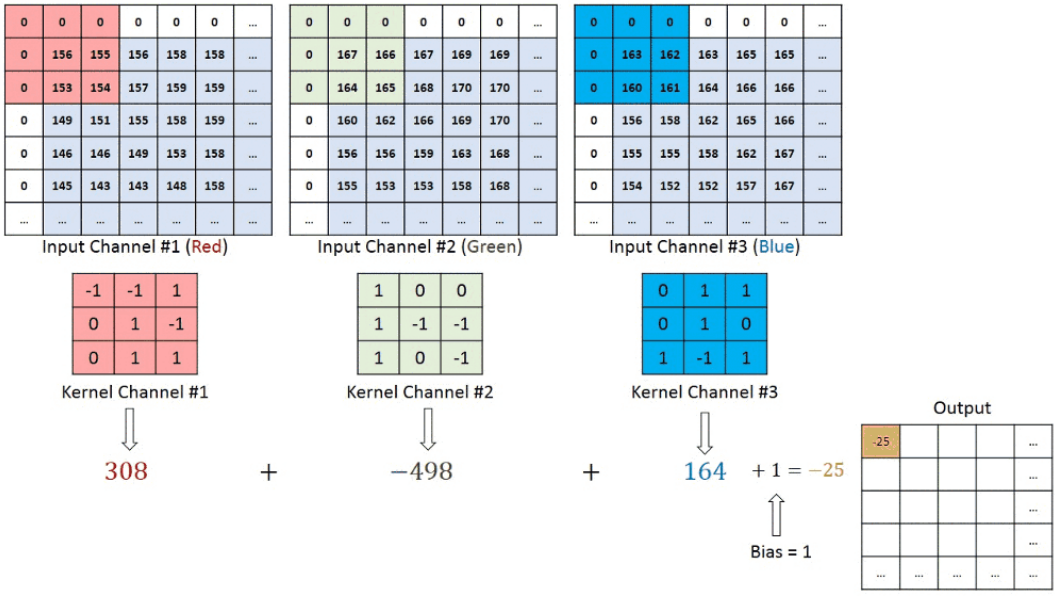


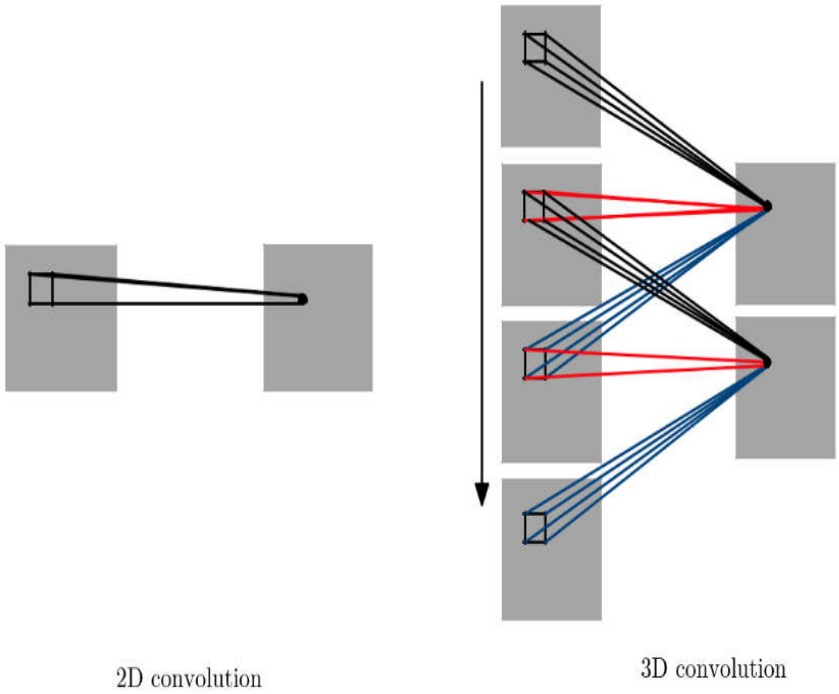
$$\text{ReLU}(x) = \begin{cases} x & if\ x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



2.2 Mạng tích chập sâu

Lớp chập



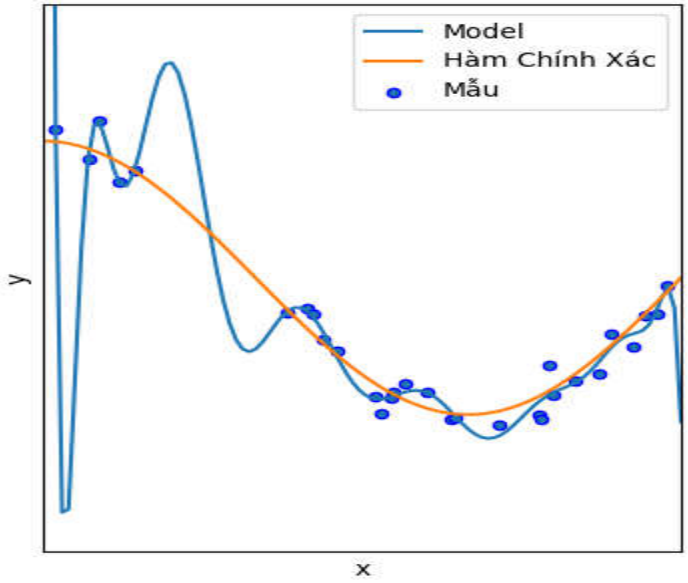


So sánh giữa 2D convolution và 3D convolution

2.2 Mạng tích chập sâu

Cơ chế tắt ngẫu nhiên (Dropout)

Một cách đơn giản để ngăn chặn overfitting là sử dụng lớp dropout. Lớp dropout có một siêu tham số tỉ lệ dropout *p*, là một thành phần xác suất, lựa chọn ngẫu nhiên một số thành phần của đầu vào và loại bỏ nó trong quá trình huấn luyện.



Overfitting

Dropout = 0.5

2.3 Tối ưu hóa tham số

Categorical crossentropy

- Các tham số $\theta_{optimal}$ của mô hình f là tối ưu khi:

$$\theta_{optimal} = \arg\min_{\theta} \sum_{i=0}^{N} \mathcal{L}(f(\mathbf{x}^{(i)};\theta), \mathbf{y}^{(i)})$$

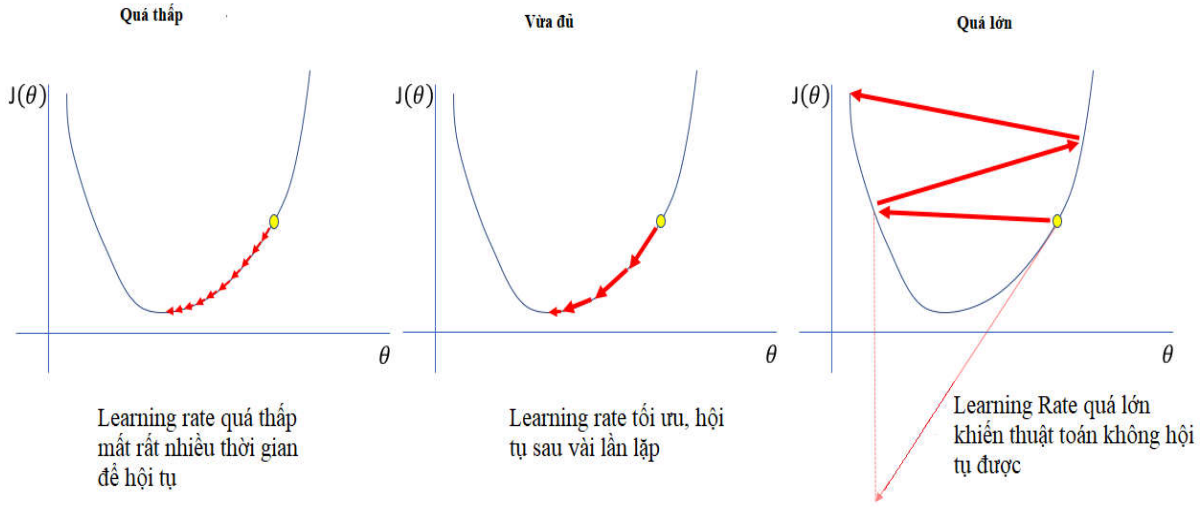
- Categorical crossentropy là một hàm dùng để tính error khi phân lớp, được định nghĩa như sau:

$$\mathcal{L}_{cross}(\tilde{\mathbf{y}},\mathbf{y}) = -\sum_{k=0}^{K} y_k \log \tilde{y}_k$$

2.3 Tối ưu hóa tham số

Gradient Descent

Gradient Descent (GD) là một kỹ thuật đơn giản để tối ưu hóa các mục tiêu phi tuyến. Cập nhật GD được định nghĩa như sau: $$\theta_{t+1} = \theta_t - \gamma_{theta} \nabla_{\theta_t} g(\mathbf{x}_i; \theta_t)$$



Chọn Learning rate = 0.1
Learning rate decay = 3

2.3 Tối ưu hóa tham số

Weight Decay

Weight Decay (Decay có nghĩa tiêu biến) là một cách đơn giản để giảm khả năng bị overfitting trong DCNs bằng cách thêm một đại lượng regularization vào hàm mất mát

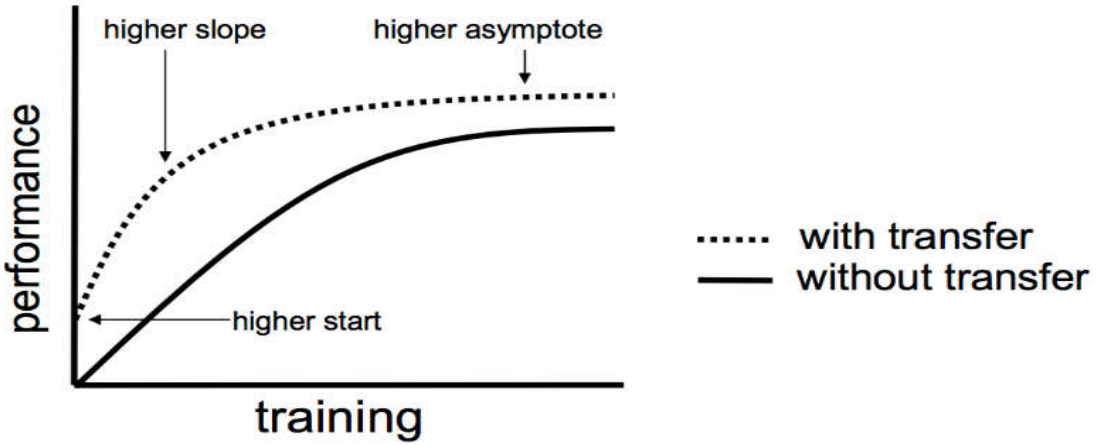
$$\theta_{t+1} = \theta_t - \gamma\left(\frac{1}{N}\sum_{i=0}^{N}\nabla_{\theta_t}\mathcal{L}(f(\mathbf{x}^{(i)};\theta_t),\mathbf{y}^{(i)}) - \alpha\theta_t\right)$$

Chọn weight decay = 0.0000001

2.3 Tối ưu hóa tham số

Học chuyển tiếp

Học chuyển tiếp cho phép thu được các tham số từ một nhiệm vụ đã thực hiện, được chuyển sang một nhiệm vụ mới.

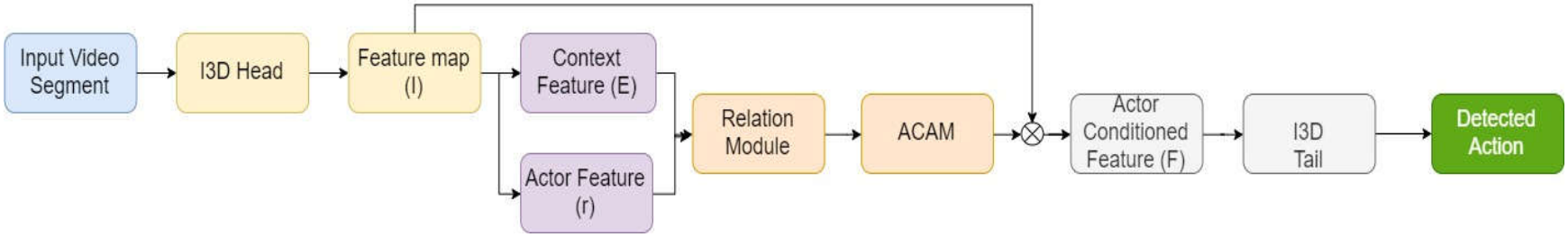


Chúng tôi sử dụng mô hình I3D pre-trained trên ImageNet+Kinetics
https://github.com/piergiaj/pytorch-i3d

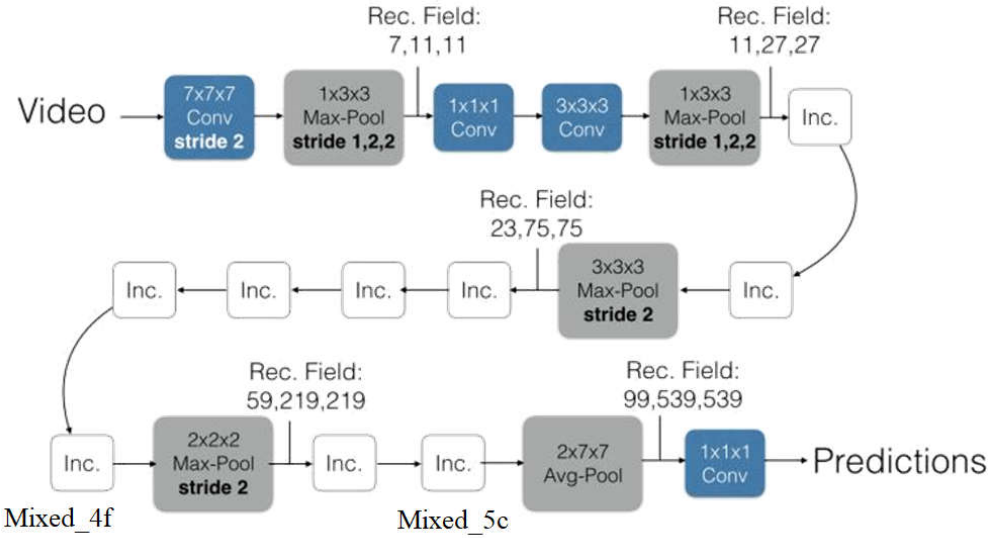
3 Huấn luyện mô hình

3.1 ACAM

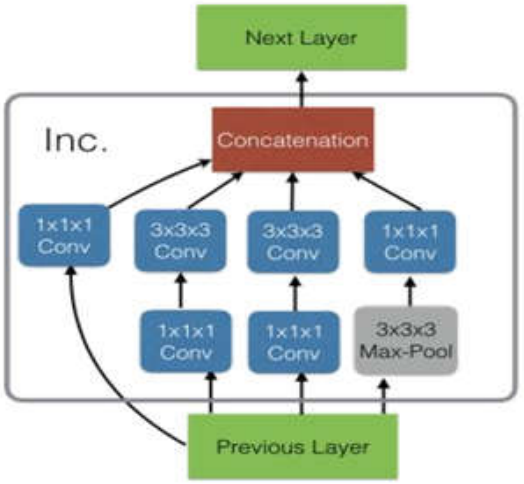
Kiến trúc hệ thống ACAM



Inflated Inception-V1



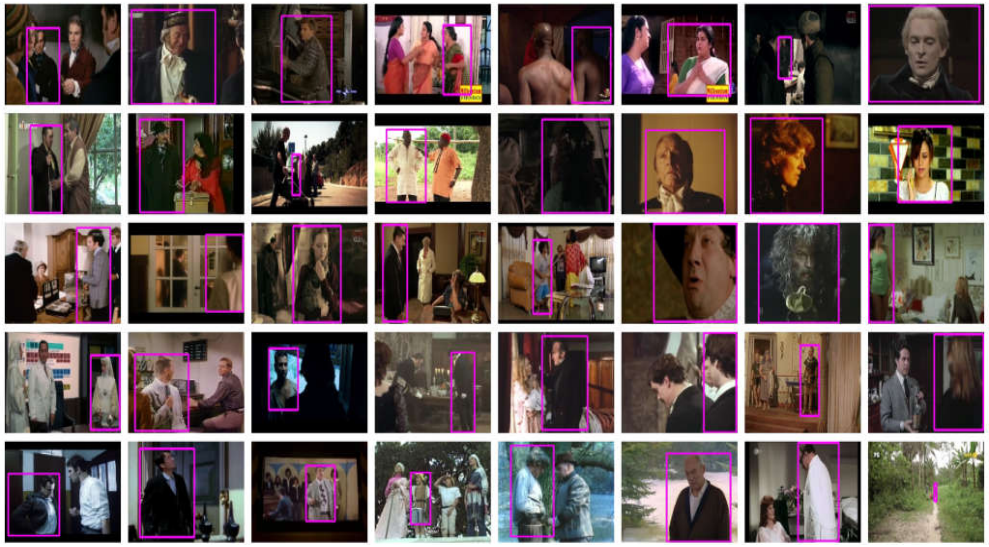
Inception Module (Inc.)



3.2 Bộ dữ liệu AVA

Đặc điểm:

- Nhận dạng được 80 lớp hành động với 1.58 triệu nhãn trong 300 video clip.
- Sử dụng các phân đoạn video 1s cho chú thích.
- Sử dụng phim ảnh để đạt được sự "tự nhiên" cần thiết.
- Sự phân bố không đều các lớp dữ liệu.



Bộ dữ liệu AVA của Google

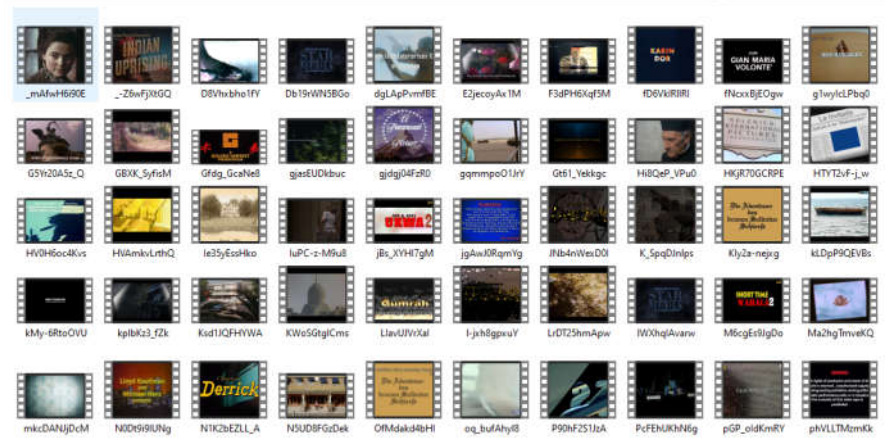
Bộ dữ liệu AVA được Google public tại:
https://research.google.com/ava/

3.3 Huấn luyện mô hình

Tải về bộ dữ liệu AVA

```
['GBXK_SyfisM;1457;0.339;0.257;0.683;0.777;17;493']
[youtube] GBXK_SyfisM: Downloading webpage
[youtube] GBXK_SyfisM: Downloading video info webpage
[youtube] GBXK_SyfisM: Downloading js player vfl7Ksmll
[youtube] GBXK_SyfisM: Downloading js player vfl7Ksmll
WARNING: Requested formats are incompatible for merge and will be merged into mkv.
[download] Destination: outdir/GBXK_SyfisM;1457;0.339;0.257;0.683;0.777;17;493.f135.mp4
[download]   5.2% of 202.30MiB at  1.32MiB/s ETA 02:24
```

Youtube-dl



15	32HR3MnDZ8g
16	3IOE-Q3UWdA
17	3_VjIRdXVdM
18	4Y5qi1gD2Sw
19	4ZpjKfu6Cl8
20	4gVsDd8PV9U
21	4k-rTF3oZKw
22	5LrOQEt_XVM
23	5MxjqHfkWFI
24	5YPjcdLbs5g
25	5milLu-6bWI
26	7YpF6DntOYw
27	7g37N3eoQ9s
28	7nHkh4sP5Ks
29	8JSxLhDMGtE
30	8VZEwOCQ8bc
31	8aMv-ZGD4ic
32	8nO5FFbIAog
33	9HOMUW7QNFc
34	9IF8uTRrWAM

3.3 Huấn luyện mô hình

Xử lý video để chuẩn bị cho quá trình huấn luyện

```
Input #0, matroska,webm, from 'train/rJKeqfTlAeY.mkv':
  Metadata:
    COMPATIBLE_BRANDS: iso6avc1mp41
    MAJOR_BRAND     : dash
    MINOR_VERSION   : 0
    ENCODER         : Lavf58.20.100
  Duration: 01:27:59.98, start: -0.007000, bitrate: 1659 kb/s
    Stream #0:0: Video: h264 (High), yuv420p(tv, bt709, progressive), 1920x1080 [SAR 1:1 DAR
    Metadata:
      HANDLER_NAME    : ISO Media file produced by Google Inc. Created on: 02/16/2019.
      DURATION        : 01:27:59.960000000
    Stream #0:1(eng): Audio: opus, 48000 Hz, stereo, fltp (default)
    Metadata:
      DURATION        : 01:27:59.981000000
Stream mapping:
  Stream #0:0 -> #0:0 (h264 (native) -> h264 (libx264))
  Stream #0:1 -> #0:1 (copy)
Press [q] to stop, [?] for help
[libx264 @ 0x56531fdcd740] using SAR=1/1
[libx264 @ 0x56531fdcd740] using cpu capabilities: MMX2 SSE2Fast SSSE3 SSE4.2 AVX FMA3 BMI2 A
[libx264 @ 0x56531fdcd740] profile High, level 4.0
[libx264 @ 0x56531fdcd740] 264 - core 152 - H.264/MPEG-4 AVC codec - Copyleft 2003-2017 - htt
x113 me=hex subme=7 psy=1 psy_rd=1.00:0.00 mixed_ref=1 me_range=16 chroma_me=1 trellis=1 8x8
eads=2 sliced_threads=0 nr=0 decimate=1 interlaced=0 bluray_compat=0 constrained_intra=0 bfra
=250 keyint_min=25 scenecut=40 intra_refresh=0 rc_lookahead=40 rc=crf mbtree=1 crf=23.0 qcomp
[mp4 @ 0x56531fdcae40] track 1: codec frame size is not set
Output #0, mp4, to 'train_ver2/rJKeqfTlAeY_1586_1589.mp4':
```

ffmpeg

_a9SWtcaNj8_950_953

_a9SWtcaNj8_953_956

_a9SWtcaNj8_956_959

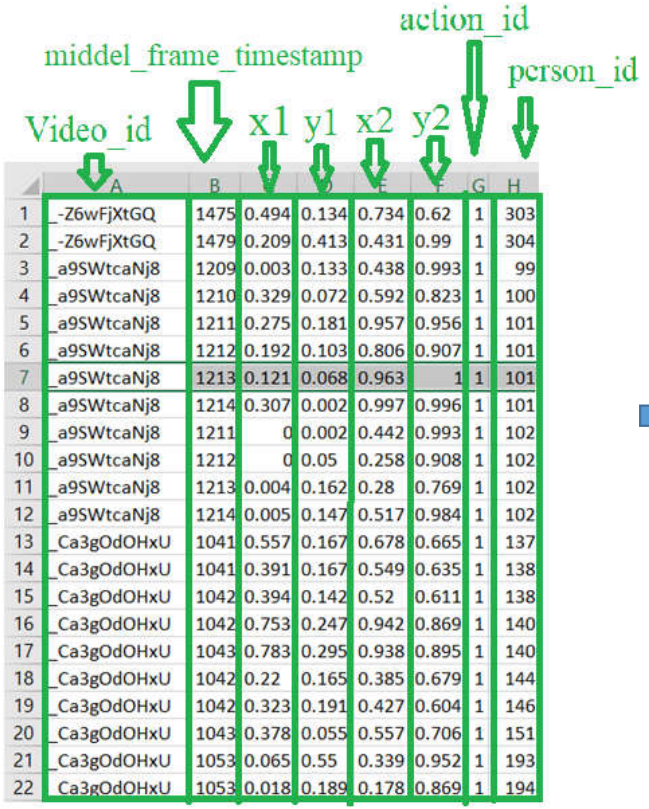
_a9SWtcaNj8_986_989

_a9SWtcaNj8_989_992

_a9SWtcaNj8_992_995

3.3 Huấn luyện mô hình

Chỉnh sửa file chú thích (file csv) của bộ huấn luyện



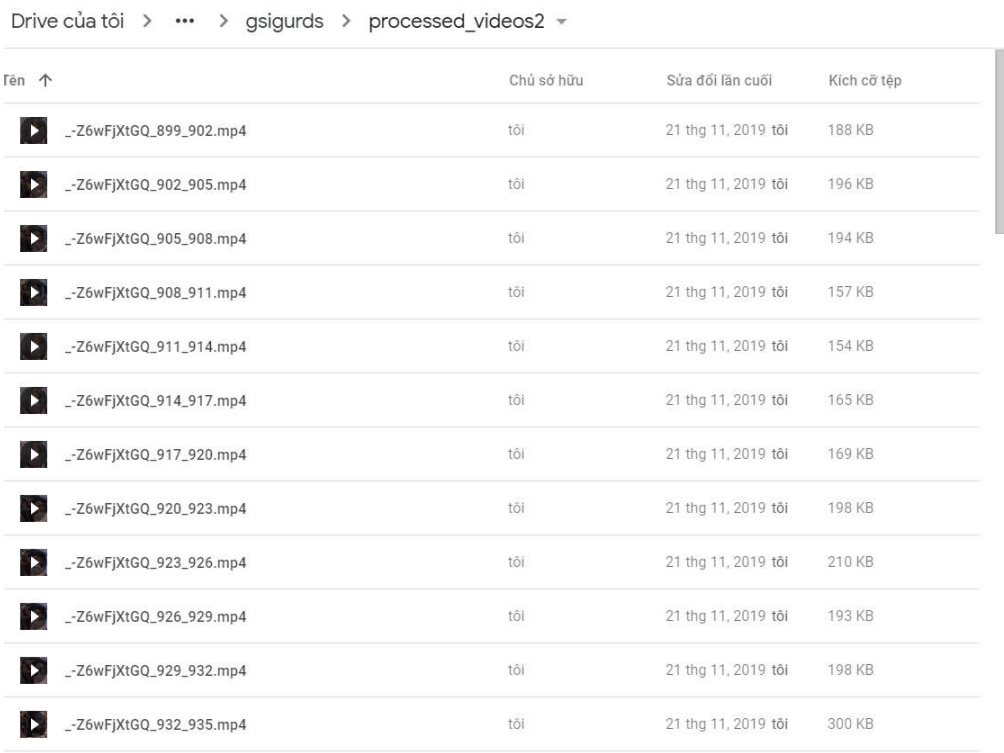
	A	B	C	D	E	F	G	H
1	_-Z6wFjXtGQ	1475	0.494	0.134	0.734	0.62	1	303
2	_-Z6wFjXtGQ	1479	0.209	0.413	0.431	0.99	1	304
3	_a9SWtcaNj8	1209	0.003	0.133	0.438	0.993	1	99
4	_a9SWtcaNj8	1210	0.329	0.072	0.592	0.823	1	100
5	_a9SWtcaNj8	1211	0.275	0.181	0.957	0.956	1	101
6	_a9SWtcaNj8	1212	0.192	0.103	0.806	0.907	1	101
7	_a9SWtcaNj8	1213	0.121	0.068	0.963	1	1	101
8	_a9SWtcaNj8	1214	0.307	0.002	0.997	0.996	1	101
9	_a9SWtcaNj8	1211	0	0.002	0.442	0.993	1	102
10	_a9SWtcaNj8	1212	0	0.05	0.258	0.908	1	102
11	_a9SWtcaNj8	1213	0.004	0.162	0.28	0.769	1	102
12	_a9SWtcaNj8	1214	0.005	0.147	0.517	0.984	1	102
13	_Ca3gOdOHxU	1041	0.557	0.167	0.678	0.665	1	137
14	_Ca3gOdOHxU	1041	0.391	0.167	0.549	0.635	1	138
15	_Ca3gOdOHxU	1042	0.394	0.142	0.52	0.611	1	138
16	_Ca3gOdOHxU	1042	0.753	0.247	0.942	0.869	1	140
17	_Ca3gOdOHxU	1043	0.783	0.295	0.938	0.895	1	140
18	_Ca3gOdOHxU	1042	0.22	0.165	0.385	0.679	1	144
19	_Ca3gOdOHxU	1042	0.323	0.191	0.427	0.604	1	146
20	_Ca3gOdOHxU	1043	0.378	0.055	0.557	0.706	1	151
21	_Ca3gOdOHxU	1053	0.065	0.55	0.339	0.952	1	193
22	_Ca3gOdOHxU	1053	0.018	0.189	0.178	0.869	1	194

3.3 Huấn luyện mô hình

Cài đặt thông số huấn luyện

```
'--dataset', 'ava_mp4',
'--arch', 'aj_i3d',
'--lr', '0.1',
'--lr-decay-rate', '3',
'--wrapper', 'default',
'--criterion', 'background_criterion',
'--epochs', '20',
'--batch-size', '5',
'--train-size', '1.0',
'--dropout' , '0',
'--weight-decay', '0.0000001',
'--val-size', '0.1',
```

Tải dữ liệu đã xử lý lên Google Drive



3.3 Huấn luyện mô hình

Tạo notebook Google Colab, tải dữ liệu lên Google Drive. Kết nối Google Colab với Drive và bắt đầu quá trình huấn luyện mô hình.

```
cd /content/drive/My Drive/Google_Colab/PyVideoResearch
```

```
/content/drive/My Drive/Google_Colab/PyVideoResearch
```

```
!python i3d_ava.py
```

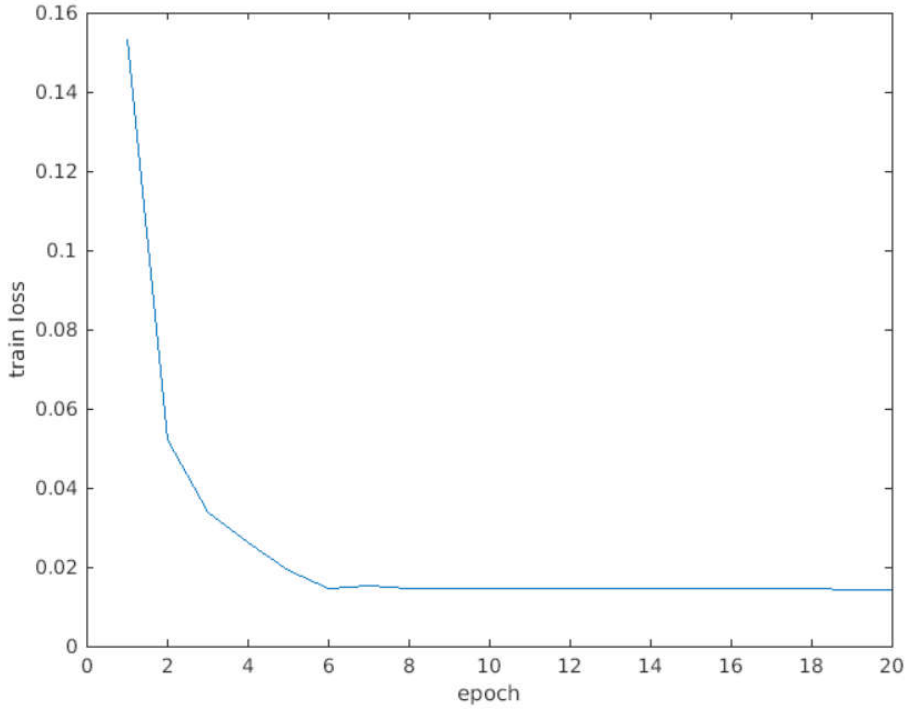
```
parsing arguments
Logging to file ./nfs.yoda/gsigurds/caches/i3d_ava//log.txt
{'name': 'i3d_ava', 'resume': './nfs.yoda/gsigurds/caches/i3d_ava/model.pth.tar;./nfs.y
experiment folder: /content/drive/My Drive/Google_Colab/PyVideoResearch
fatal: not a git repository (or any parent up to mount point /content)
Stopping at filesystem boundary (GIT_DISCOVERY_ACROSS_FILESYSTEM not set).
Command '['git', 'describe', '--always']' returned non-zero exit status 128.
git hash:
setting Dropout p to 0.5
=> loading checkpoint './nfs.yoda/gsigurds/caches/i3d_ava/model.pth.tar'
=> loaded checkpoint './nfs.yoda/gsigurds/caches/i3d_ava/model.pth.tar' (epoch 18)
setting start epoch to model epoch 18
```

Auto-click tránh việc mất kết nối với Colab

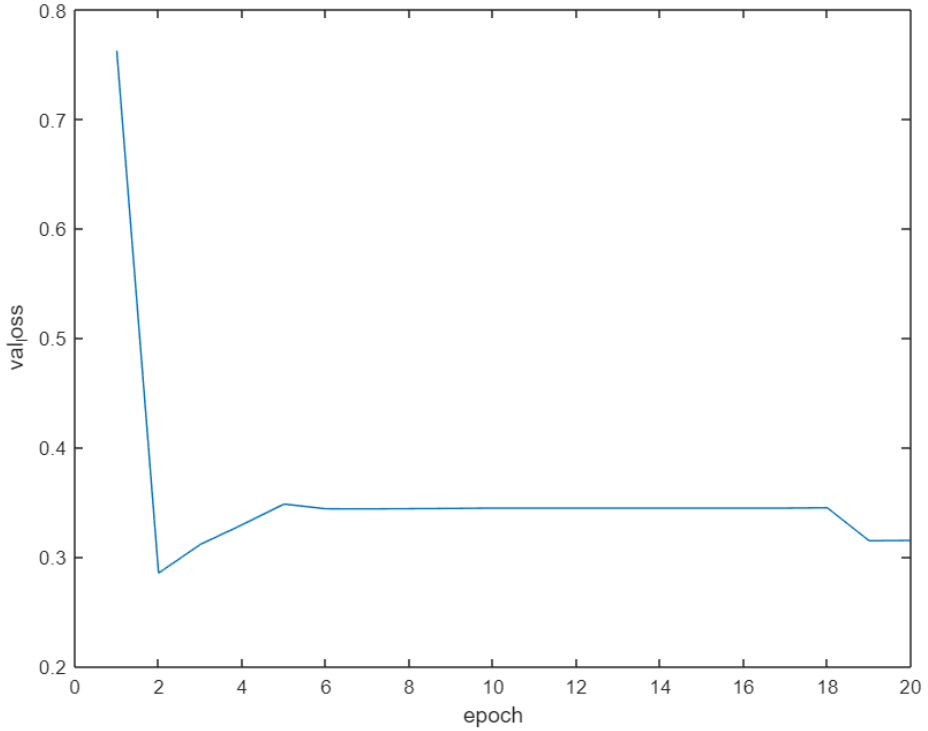
```
function ClickConnect(){
console.log("Working");
document.querySelector("colab-
toolbar-button#connect").click()
}
setInterval(ClickConnect,60000)
```

3.3 Huấn luyện mô hình

Kết quả huấn luyện



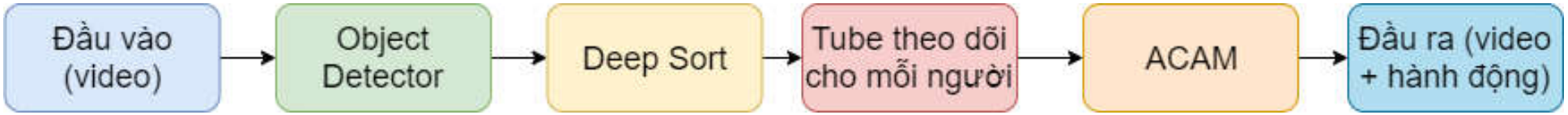
Train loss trong 20 epoch



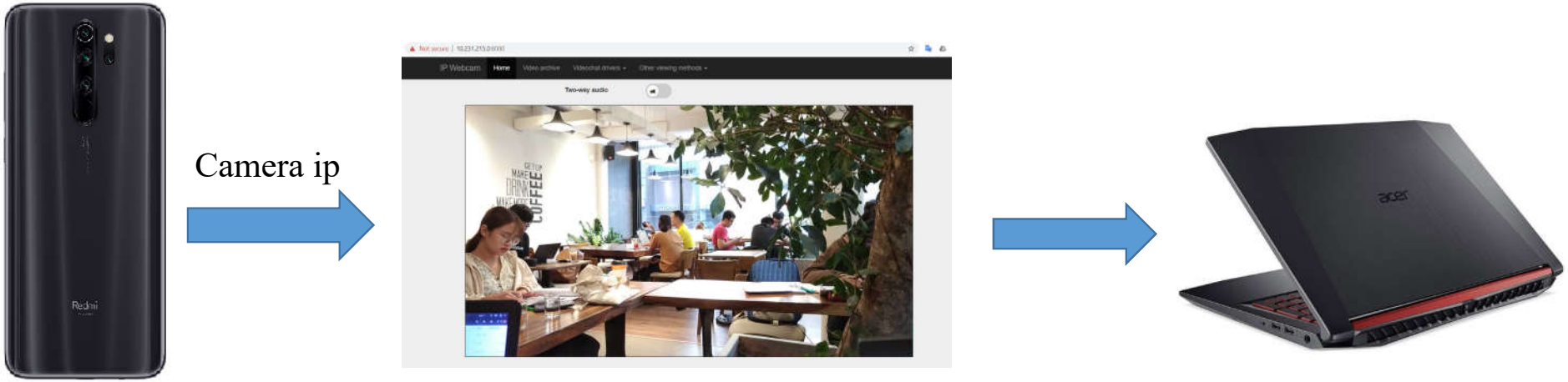
Val loss trong 20 epoch

4 Nhận dạng hành động con người

4.1 Nguyên lý hoạt động chương trình

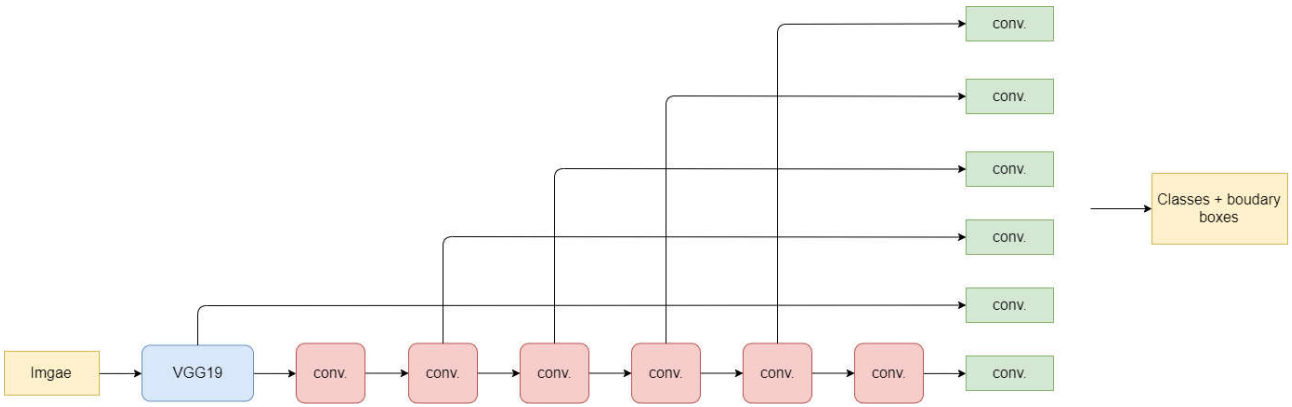


Đầu vào: Video. Có thể là video đã được tải về máy tính hoặc video chạy real-time.

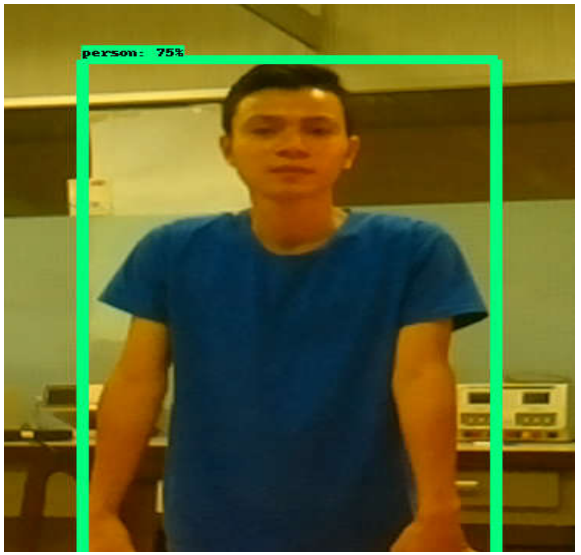


4.2 Nhận dạng đối tượng

Mô hình Sigle Shot Detector VGG19



Nhận dạng đối tượng, ở đề tài này là con người



4.3 Deep SORT

Deep SORT

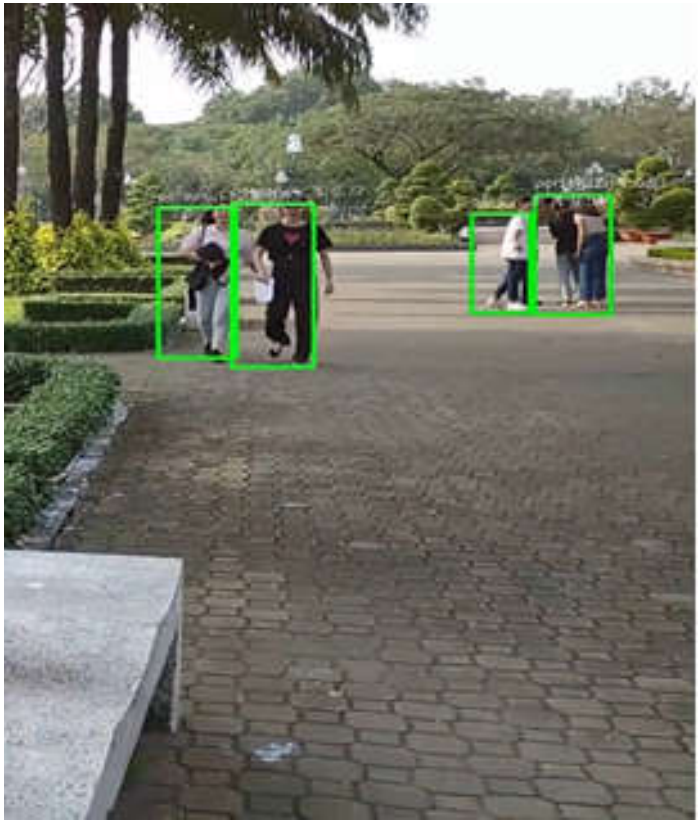
➢Bộ lọc Kalman.

➢Thuật toán Hungrary tiêu chuẩn.

➢Thước đo Mahalanobis bình phương.

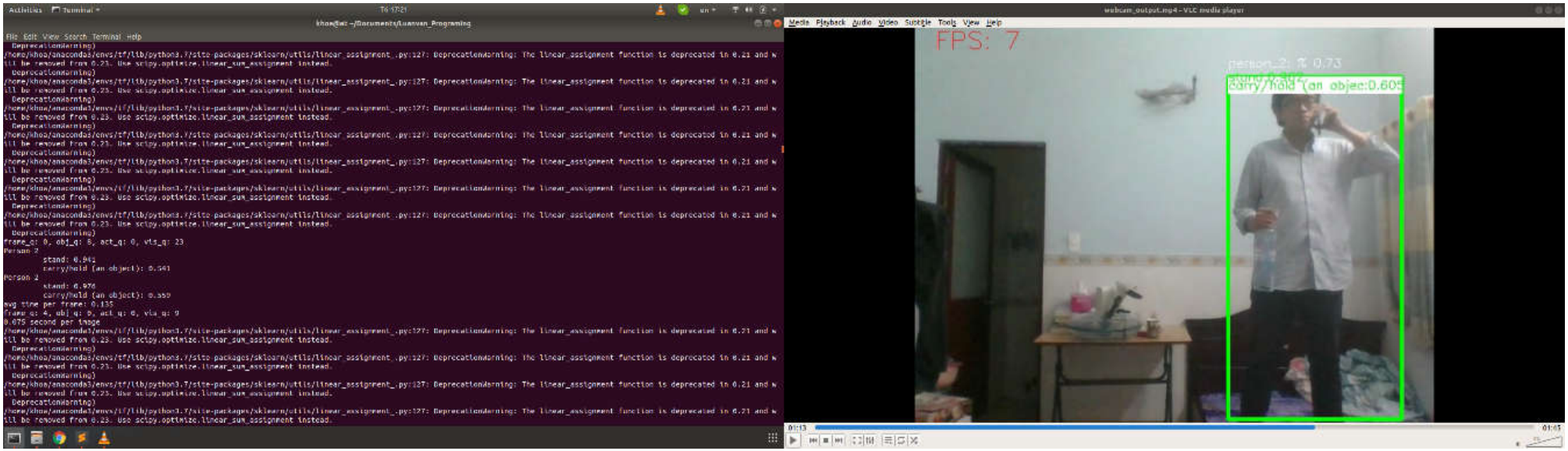
➔Theo dõi đối tượng sau khoảng thời gian T frame.

➔Đánh số thứ tự người xuất hiện trong khung hình.



4.5 Đầu ra

Kết quả của chương trình được thể hiện ở Terminal và hình ảnh được xuất ra bằng OpenCV.

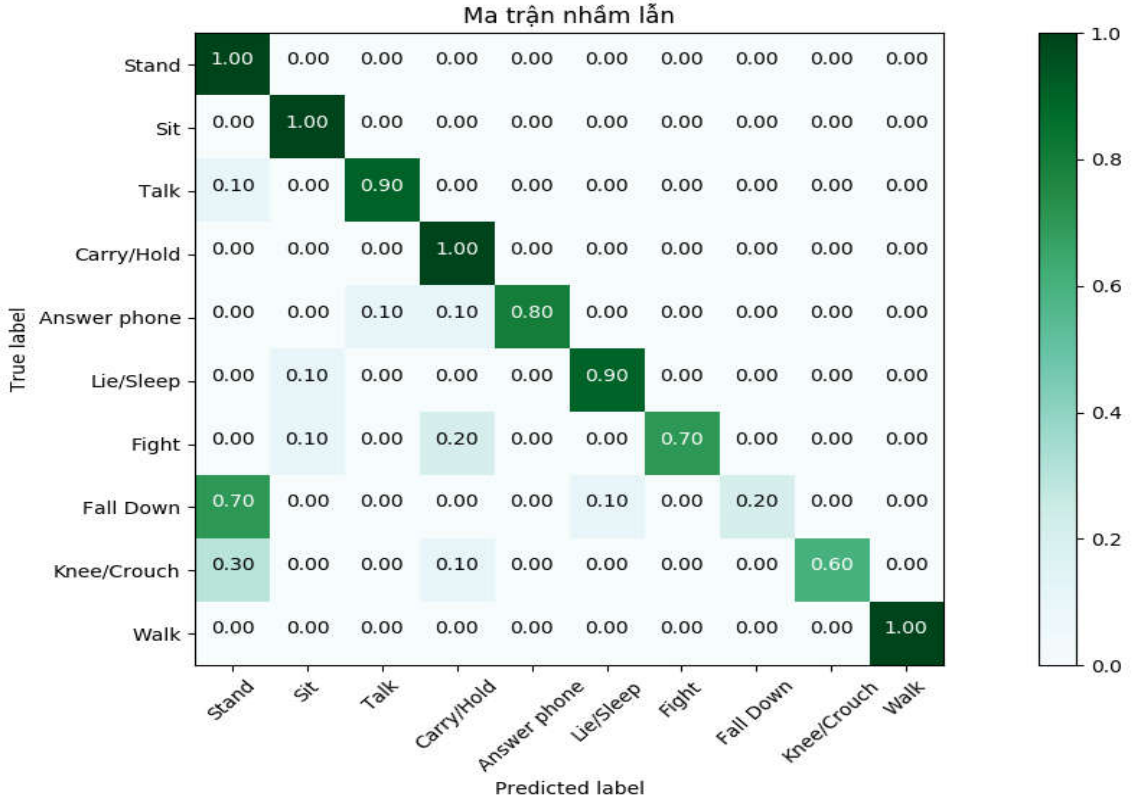


4.6 Kết quả


Luận Văn Tốt Nghiệp
Nhận Dạng Hành Động Con Người
Anh Khoa – Trung Tín

4.7 Đánh giá mô hình

Ma trận nhầm lẫn của mô hình, mỗi lớp được kiểm tra 10 lần bằng các video trong tập validation.



5 Kết luận – Hướng phát triển

5.1 Kết luận

Các mục tiêu đã đạt được:

- Tìm hiểu được các lý thuyết về học sâu, mạng tích chập sâu.
- Tìm hiểu về bộ dữ liệu AVA, huấn luyện một mô hình trên bộ dữ liệu với 4 lớp hành động.

Các mục tiêu chưa đạt được:

- Chưa ứng dụng được mô hình đã huấn luyện vào thực tế.

5.2 Hướng phát triển

- Để có thể xây dựng một mô hình thời gian thực có hiệu suất tương đương, chúng tôi có các đề xuất sau:
  - ➢ Sử dụng cảm biến bổ sung.
  - ➢ Sử dụng các phương pháp trích xuất đặc trưng hiệu quả hơn.
  - ➢ Giảm chất lượng hình ảnh đầu vào, cân bằng thông tin giữa đầu vào và đầu ra.

Cảm ơn thầy và các bạn đã lắng nghe!