2021 年臺灣國際科學展覽會
優勝作品專輯

作品編號 190039

參展科別 電腦科學與資訊工程

作品名稱 A deep learning-based home safety perception system for household service robot

得獎獎項 大會獎 三等獎

國　　家 Macau

就讀學校 Pui Ching Middle School Macau

指導教師 LAO KUN WA

作者姓名 Lei In Hei

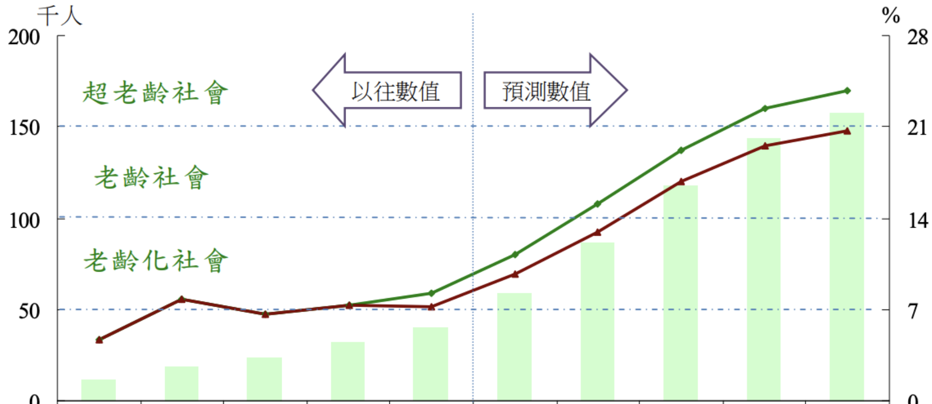
作者照片

Abstract

In 2016, the population of people over the age of 65 in Macau was 11.2%. This means that Macau has already become an aging society. As such, more younger generations are needed to look after the elderly. According to statistics, falls are the primary cause of injury or death for the elderly over 65 years old. About 30% of the elderly over 65 years old fall every year. Along with the increase in the elderly population, it is urgent to find a fast and effective way to ensure the safety of the elderly. As there is a lot more danger besides falling in an elderly life, we aim to build a robot collocated with its danger detection system to ensure the safety of the elderly at home. The reason we decided to use robots is that we want to have larger flexibility and mobility, for example, we can give elderly rescue materials when they need help. Moreover, more home robots will be used in the future, they can just apply our system to theirs and ensure the safety of elderlies. In this research, we mainly used cameras with the Openpose model to detect dangers such as falling, potential human action danger, and environmental danger. Innovative ways are used to detect fall action, collocated with our home robot, it is a foreseeing project that could ensure the safety of the elderly in a home environment.

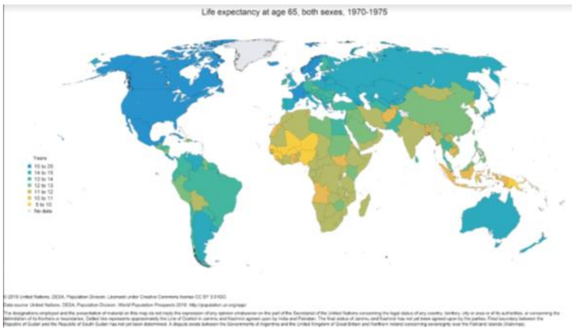
1. Background

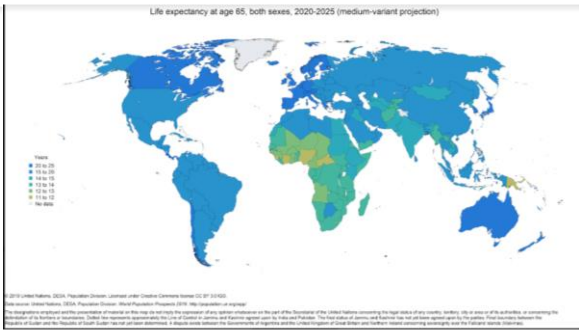
Nowadays, people pay more attention to the aging population. According to the United Nations, in 1970, the population proportion of Japan's elderly was 7.9%, but now the population proportion of Japan's elderly is 28.4%. Here is another example, Italy's population proportion of elderly was 11% in 1970. However, the population of Italy's elderly is 23.4%. Furthermore, China is facing this problem. In 1970, China's population proportion of elderly was 2.9%, yet the percentage raced up to 16% this year. In the future, the percentage of the population proportion of the elderly will be much higher.



	1970	1981	1991	2001	2011	2016	2021	2026	2031	2036
長者人數 (千人)	11.6	18.7	23.5	31.7	40.0	59.1	86.5	117.4	143.4	157.3
本地長者比例 (%)	~	~	~	~	8.2	11.2	15.1	19.2	22.4	23.7
總體長者比例 (%)	4.7	7.7	6.6	7.3	7.2	9.7	12.9	16.8	19.5	20.7

Pic.1 Population aging trend and forecast of Macau





Pic.2 Population proportion of 1970 ***Pic.3 Population proportion of 2020***

The aging population brings a lot of problems such as the number of elderly is increasing, and the labor force is diminishing. That causes workers in places like homes for the aged, the recreation center is not enough, and they can't take good care of the elderly, so the elderly may face more danger than before. As stated by the WHO, falling is the second leading cause of death. The elderly may fall and get hurt when they are alone. Moreover, the complication of falling can influence the elderly's life a lot. To take better care of the elderly, and also to keep them safe, we invented the "Self Supporting Elderly Robot".

1.1 Current Solutions

There are many elderly robots in the market. It is about to be divided into two parts, life assistant and emotional companion. Here are some examples.

1.ElliQ

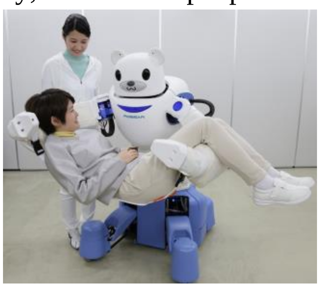
ElliQ is a robot that can communicate with elderlies. ElliQ can help them to interface with their friends and families. In daily life, ElliQ can answer questions from the elderly, and it can give some suggestions to elderlies too. Also, it can remind the elderly something so that they can remember it.

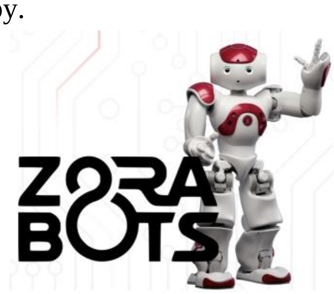
2.ROBEAR

ROBEAR is a nursing robot that can take care of elderlies in their life. Its appearance looks like a giant bear. ROBEAR can support the elderly to stand up. Furthermore, it can lift an elderly from a bed to a wheelchair or other places.

3.Zora

Zora is a small robot that can speak 19 languages. It can accompany the elderly. Zora can do tasks like singing, playing games, reading, etc. Zora is used to keeping elderlies active by dancing and doing exercise. Though it has a small body, it can make people feel happy.





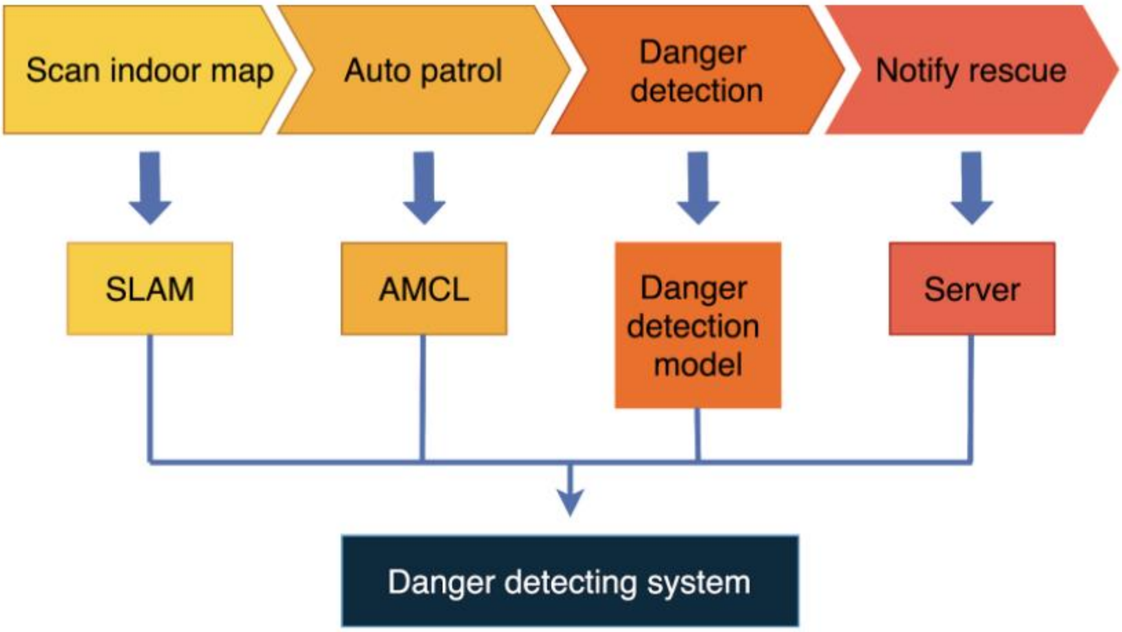
Pic.4 picture of ElliQ ***Pic.5 picture or ROBEAR*** ***Pic.6 picture of Zora***

1.2 Conclusion of solutions

To sum up, the existing elderly robots are sometimes expensive for elderlies and their families. On the other hand, the existing elderly robots can only take care of the elderly or communicate with them. These robots cannot detect if the elderly are in danger or not. If the elderly fall or get hurt, they may not be useful in this case.

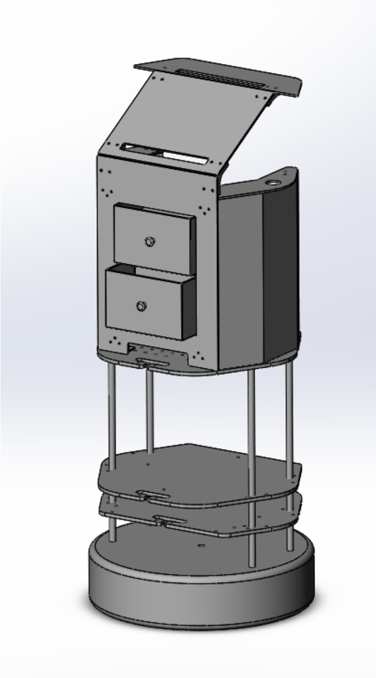
2. Research Goal

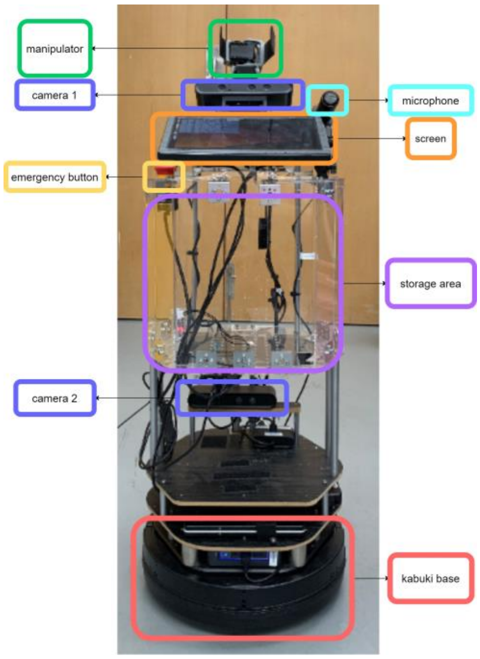
Our research goal is to build a robot that can detect the action/environment danger in the elderly's home. For action danger, it refers to fall and danger behavior detection. On the other hand, examples of environmental danger are fire detection. When the elderly forget to turn off the fire, our robot can remind them to prevent the fire from breaking out. In this report, we will focus more on fall detection as it is the most significant hidden danger for living alone. We aim to ensure the safety of the elderly in their home.



3. Hardware design

First, we used Solidworks modeling software to design our bodies. The body is mainly divided into three layers: the first layer is equipped with a robotic arm, touch screen, microphone, and RGBD lens. The role of this layer is to help and understand the elderly and In real-time communication with the outside world, the elderly can use the microphone and screen to command the robot to give instructions, and the robot can also understand the real-time situation of the elderly through the lens; the second layer is the storage layer, where the elderly can place heavier objects. The current load on this layer is 8 kg; the third layer is the body layer, all computers, batteries, and chassis are also installed on this layer, which promotes the progress and operation of the robot.





3.1 RGB-D Camera

We used an RGB-D camera to collect data for our model training and testing. However, as we have to detect the elderly's different actions, we mounted two servos to adjust the camera angle in the x and y-axis.

Parameter	Sensor			
	ORBBEC ASTRA S	MICROSOFT KINECT II	INTEL SR300	INTEL D435
Range	0.4m-2m	0.5m-4.5m	0.3m-2m	0.2m -10m
RGB FOV	60°(H) × 49.5°(V) × 73°(D)	70.6°(H) × 60°(V)	41.5°(H) × 68°(V) × 75.2°(D)	69.4°(H) × 42.5°(V)) × 77°(D)
Depth FOV	60°(H) × 49.5°(V) × 73°(D)	70.6°(H) × 60°(V)	55°(H) × 71.55°(V) × 88°(D)	91.2°(H) × 65.5°(V)) × 100.6°(D)
Frame rate	30 fps	30 fps	30, 60 fps	30, 60, 90 fps
RGB resolution	640×480 pixel	1920 × 1080 pixel	1920 × 1080 pixel	1920 × 1080 pixel 1280×720 pixel 848×480 pixel 640×480 pixel
Depth resolution	640×480 pixel	512 × 424 pixel	640 × 480 pixel	1280 × 720 pixel 848×480 pixel 640×480 pixel
Weight	300 gr	966gr	300gr	100gr
Size	165mm × 30mm × 40mm	255mm × 66mm × 67mm	110mm × 12.6mm × 4.1mm	90mm × 25mm × 25mm
Power supply	USB 2.0	power supply + USB 3.0	USB 3.0	USB 3.0
Power consumption	< 2.4 W	∼ 15 W	650-1800 mW	618 -1978 mW
Operating system	Android, Linux Windows 7/8/10	Linux Windows 8/10	Linux Windows 8/10	Linux Window 8/10
SDK	Astra SDK OpenNI2 3rd party S	Kinect V2 SDK libfreenect2	Intel©RealSense™ SDK librealsense sdk [6]	Intel©RealSense™ SDK librealsense SDK[5] hand and face tracking

We used the ORBBEC Astra s camera as it has the lowest price of 1,000 RMB and the resolution it has is enough for our detecting system

3.2 Storage area

The storage area is an essential hardware section in our robot design. We could transport safety materials (First aid kit) immediately to the elderly when a fall occurred, this is what current fall detection solutions couldn't obtain. Moreover, we could also help our elderly carry objects as well as trash. When designing this storage area, the aim is to create a large place as possible. As we don't want our robot to be too heavy, we used transparent plastic in our storage area design

3.3 Development cost

In our research, we found out that the average price of a home robot is XXX. As a result, we want as little money as possible to build our robot.

Num	Item	Amount	Price (RMB)	Total price (RMB)
1	Intel NUC Kit BOXNUC6I3SYH Black	1	1,437.00	1,437.00
2	Dynamixel ax-12a manipulator	1	8,000.00	8,000.00
3	Robot battery	1	430.00	430.00
4	Apaxq speaker	1	40.00	40.00
5	ORBBEC Astra	2	1000.00	1000.00
6	ReSpeaker Mic Array v2.0	1	510.00	510.00
7	NEJIFU notebook charger	1	729.00	729.00
8	Touch screen LCD	1	114.00	114.00
9	Logitech USB hub 4 port with USB 2	2	109.00	218.00
10	Safety button	1	26.50	26.50
11	Type-C to DC 5.5mm x 2.5mm	1	11.07	11.07
12	500mm 3 Pin X-Series Compatible Cable	1	25.00	25.00
13	U2D2	1	499.00	499.00
14	U2D2 Power Hub	1	260.00	260.00
15	Turtlebot3 IPL-01	3	160.00	480.00
16	M3 Aluminum pillar - M3 * 6	4	0.70	2.80
17	M3 Aluminum pillar - M3 * 12	8	1.20	9.60
18	M3 Aluminum pillar - M3 * 20	4	1.60	6.40

19	M3 Aluminum pillar - M3 * 30	4	1.90	7.60
20	Turtlebot 2	1	12,280	12,280
			Total	25,635.9

From the above table, we can see that our robot is x times less than an average home robot. This is affordable for a household or the elderly to buy it.

4. Safety detection system

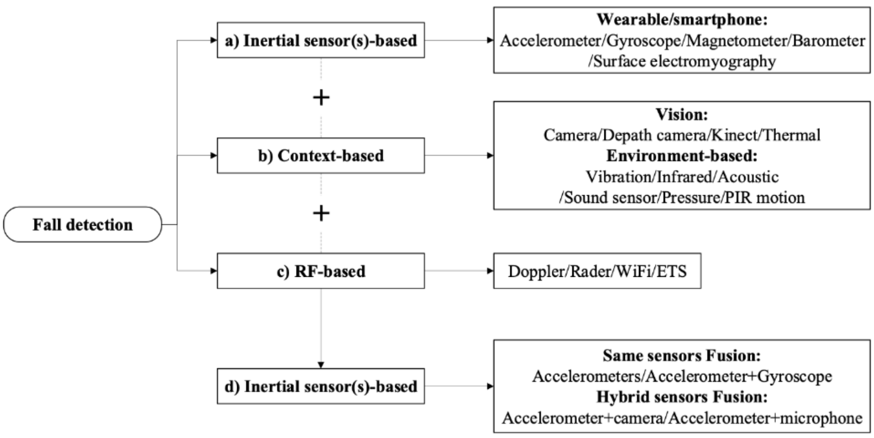
With this method, our robot can detect danger in the elderly's life with its RGB-D camera. The main detection system is divided into these two categories:

1. Danger human activity detection (Fall, elderly potential injured action)
2. Environment danger detection (Burner is not turned off, the floor is slippery)

As falling is the most vital problem in the elderly's daily life, we used it as an exemplification for our method section. All of the danger detection functions are built on the same method as fall detection, which used data from the RGB-D camera.
(add method photo draw.io)

4.1 Current fall detection method

The advantages and disadvantages of various classifications are shown in Table x. The method proposed in this paper belongs to vision-based. It can make full use of the cameras around our lives, which is convenient and has high accuracy, low cost, and is easy to implement.



Pic7. Current solution for fall detection

	Inertial Sensor(s)-Based	Context-Based		RF-Based	Sensors Fusion-Based
		Ambient-Based	Vision-Based		
Advantages	Easy to implement Few privacies issue High accuracy Real time	Least intrusive Few privacy and security issues	Convenient Accurate	Real-time Contactless Low-cost Nonintrusive	Accurate Significant performance
Shortcoming	Intrusive	Limited detection range Easier affected by the external environment	Considerable computing Privacy issue Limited capture space	Coverage issue Limited range	Information redundancy Robust fusion algorithm

Pic8. The pros and cons of current solution

4.2 Our method

The method we proposed used Openpose to get the skeleton information of the human body. We used the 18 feature points on a human skeleton with x, y, and z coordinates to train our model and use it to detect whether our elderly are in danger or not. We followed the 7 steps of machine learning from Yufeng G[1], a google programmer to achieve our fall detection system. Using the following steps, we trained an SVM and a CNN-VGG16 model for comparison.
Steps of how we train our detecting system:

1. Gathering data
2. Data preparation
3. Choosing a model
4. Training
5. Evaluation
6. Hyperparameter tuning
7. Prediction

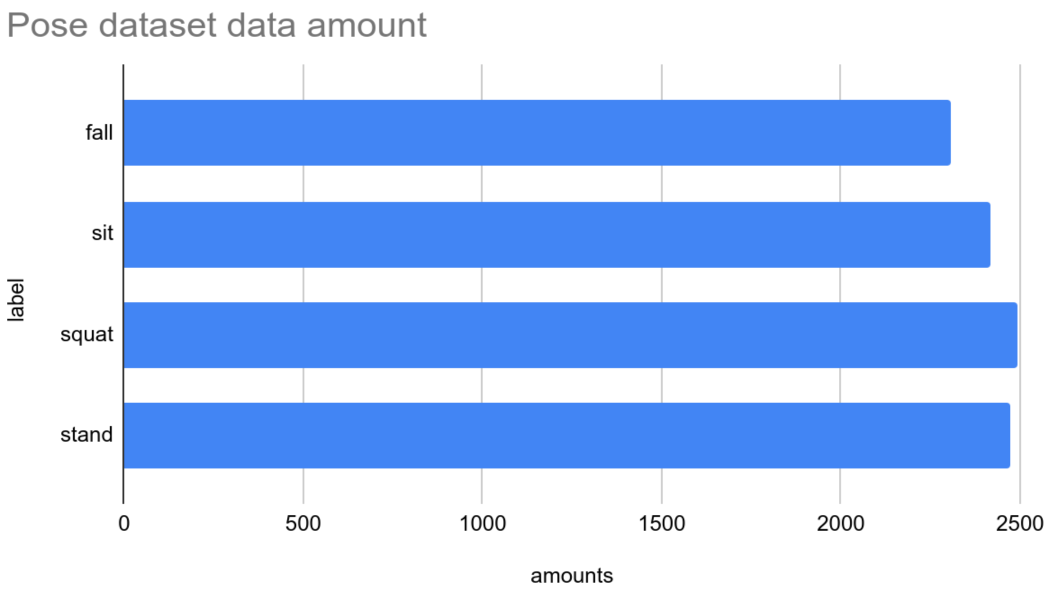
4.2.1 Gathering data

We collected 10,000 photos divided into 4 categories: stand, sit, squat, fall. Using the RGB-D camera on our robot, we get both RGB and depth images. The RGB photo is used to convert the initial image into a limb image to obtain the limb skeleton of the human body. On the other hand, the depth photo is used to get the z value of the image so that we can distinguish the position of where our elderly is falling. In our model, x,y,z coordinates of points will be put into our model training. After the addition, we processed the skeleton in the same proportion, and then put them into the same size In a completely black picture, this can eliminate the error caused by different positions or different sizes of the skeleton in the picture. After finishing processing the image, we begin to train the model.

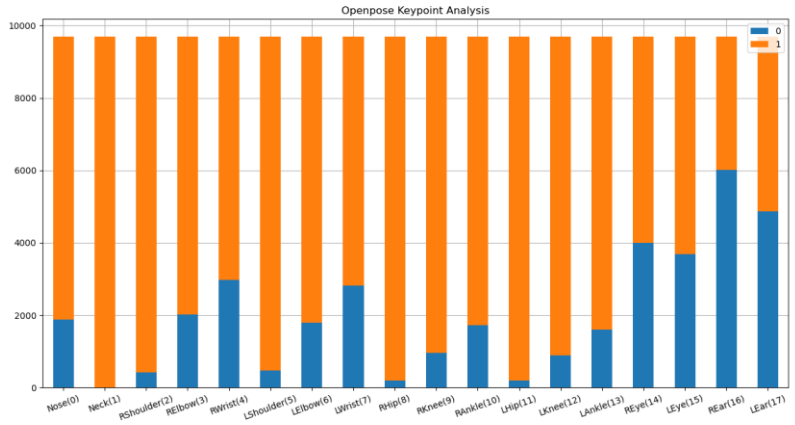
Pic9. Our image processing step when feeding into CNN.
We'll find the skeleton of a person and draw it onto a black board

4.2.2 Data preparation

In this step, we analyze the number of pose images we generate. From the graph above, the photos in the four categories are evenly distributed. There are 2250 more photos in each category, it is sufficient for our model training. On the other hand, we found out that some feature points are missing from the Keypoint analysis graph. Although, the four points on the head (left, right ear, and left, right eye) couldn't be detected much from our dataset. It is not important compared to the main human sculpture feature points. As a result, we focus on the amount of data in keypoint 0 to 14. From all the 14 pose points, an average of 8750 photos could be detected.



Pic10. Our dataset images amount according to each status.
Some of the images are ignored due to quality



Pic11. Skeleton points detected in oPic9. SVM training results in 17 models by changing there learning rate images.
To construct a skeleton, we need 18 points corresponding to the body parts.
Some points may not be detected due to human pose when we capture it.
We've totally about 30000 points of each label of points collected
We can find if that body part had detected very often by seeing the ratio between the orange and blue bar

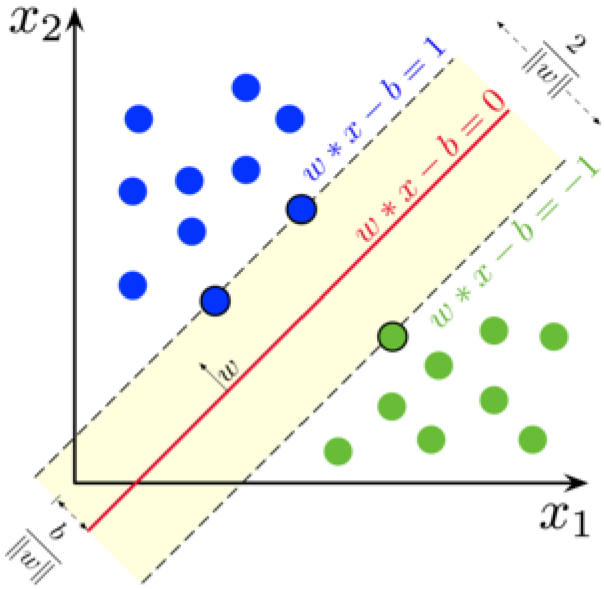
4.2.3 Choosing a model

SVM

The basic concept of SVM is very simple, that is, to find a decision boundary to maximize the boundary between the two categories so that it can be perfectly separated, so as to find the optimal solution.

In the process of training the model, we generated 20 models. Each model has the same kernel: "poly". This kernel has 2 different hyperparameters, one is C and the other is a degree. We haven't changed C, but we have changed degrees. Conceptually, degree refers to the dimension used in data processing, from one dimension to twenty dimensions to generate 20 models. Therefore, each model also has a different accuracy rate, and model 17 has an accuracy rate as high as 74.957.

But in fact, this method also has shortcomings, that is, when dealing with pictures with different backgrounds, the accuracy rate will obviously decrease. The accuracy rate will drop because the coordinates of the human body's state in the image are changeable. In such a complex state, it is impossible to recognize this changeable and complex state with limited data. Therefore, we made the final improvement.



Pic12. An SVM dividing two bunches of data

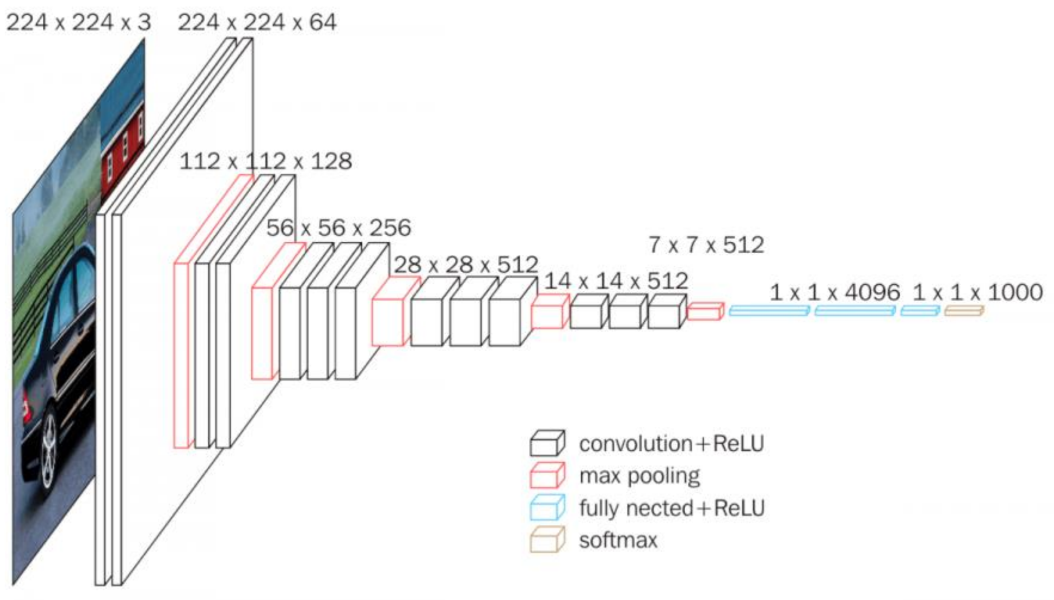
CNN-VGG16

This method is the final improvement of our prototype. Convolutional neural networks have excellent performance for large-scale image processing, so we adopted this method.

A convolutional neural network is composed of one or more convolutional layers and a fully connected layer at the top. It also includes associated weights and pooling layers.

We used the CNN method for CNN is a network structure designed for image processing, the convolution layers can filter noise information from the image and keep the key information.

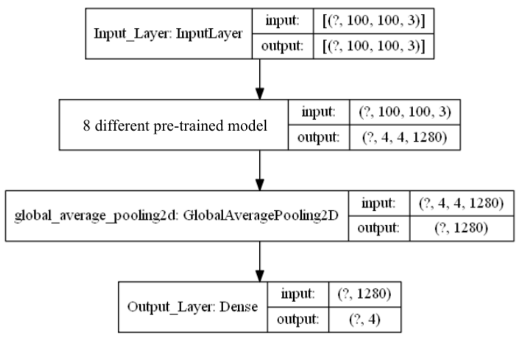
In these algorithms, we used VGG16 model structure for us to train our model, for we;ve a mini-sized dataset so complicated model structures are not useful here, so we pick a simple structure to train our model



Pic13. Model structure of VGG16

4.2.4
Training

CNN model structure



Hyper parameters

LR transfer	1e-5
Epochs transfer	50
Optimizer transfer	Adam
LR fine-tune	1e-6
Epoch fine-tune	50
Optimizer fine-tune	RMSProp
Batch size	32

Training : Validation
8 : 2

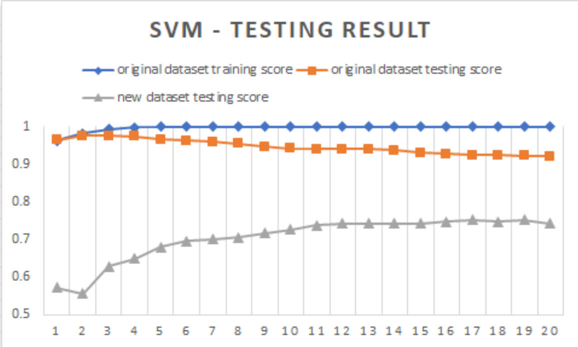
Train with 8 different pre-trained models

4.2.5 Evaluation

We collected 200 photos that are totally different from the datasets. We used this new dataset to test for both our SVM and CNN-VGG16

SVM

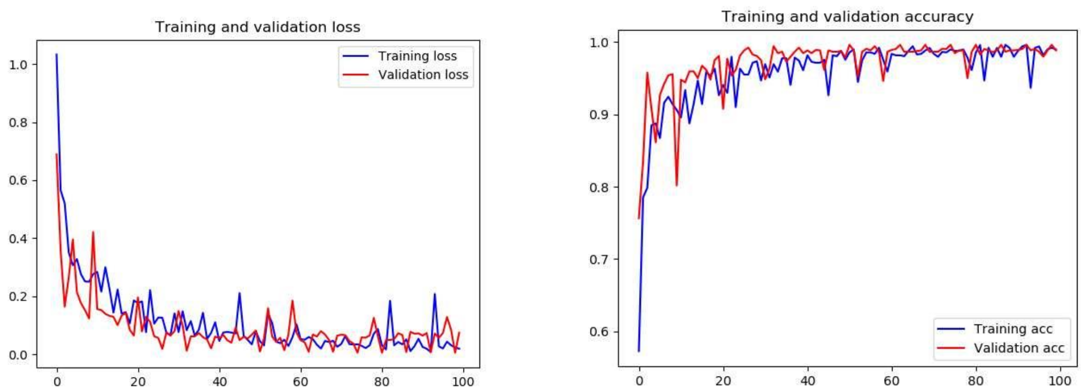
training degree	original dataset training score		original dataset testing score		new dataset testing score	
1	0.962666667		0.964476021	rising trend	0.5721649484536082	drop trend
2	0.981877777		0.975133214		0.5567010309278351	
3	0.993333333		0.975133214		0.6288659793814433	
4	0.998222222	rising trend	0.973357015		0.6494845360824743	
5	0.9991111111		0.966252220		0.6804123711340206	
6	0.999555555		0.962699822		0.6958762886597938	rising trend
7	0.999555555		0.959147424	drop trend	0.7010309278350515	
8	0.999555555		0.953818827		0.7061855670103093	
9	0.999555555		0.946714031		0.7164948453608248	
10	0.999555555		0.941385435		0.7268041237113402	
11	0.999555555		0.939609236		0.7371134020618557	
12	0.999555555	remain the same	0.939609236	remain the same	0.7422680412371134	
13	0.999555555		0.939609236		0.7422680412371134	remain the same
14	0.999555555		0.936056838		0.7422680412371134	
15	0.999555555		0.930728241	drop trend	0.7422680412371134	
16	0.999555555		0.927175843		0.7474226804123711	
17	0.999555555		0.923623445	remain the same	0.7525773195876289	rise and drop
18	0.999555555		0.923623445		0.7474226804123711	
19	0.999555555		0.921847246	drop trend	0.7525773195876289	
20	0.999555555		0.920071047		0.7422680412371134	



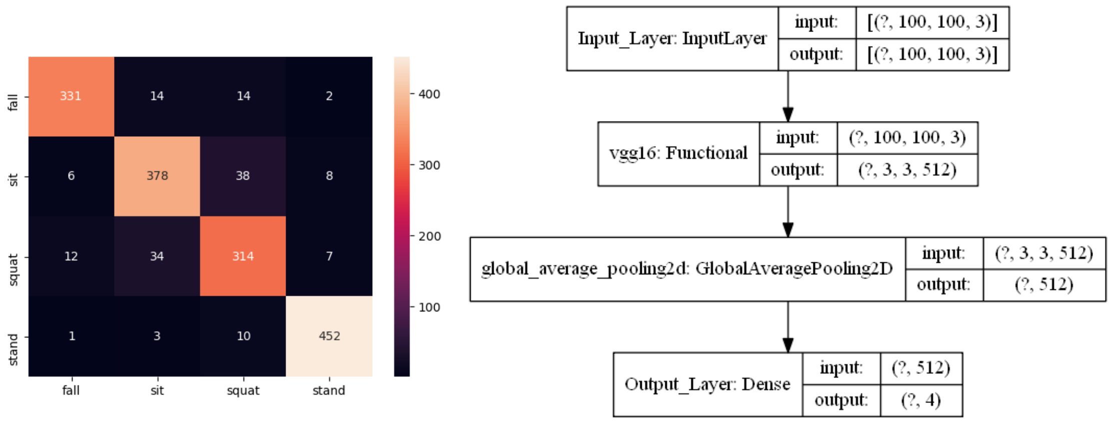
Pic14, 15. SVM training result in 17 models by changing their learning rate

CNN-VGG16

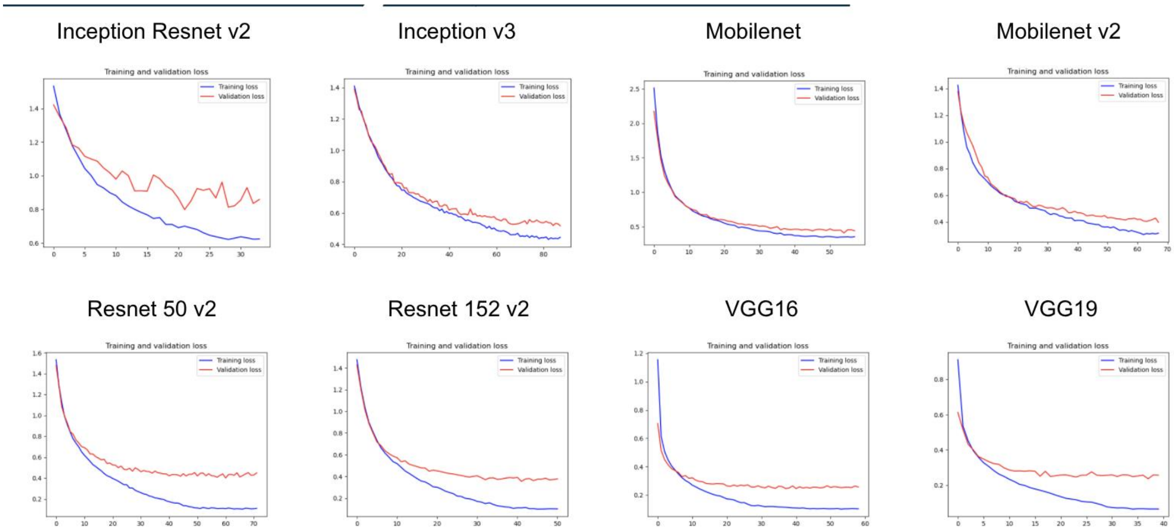
Using this model we collected the lowest loss of 0.15 and the highest accuracy of 0.95. This means in 100 photos, our models can detect 95 right images of different postures.

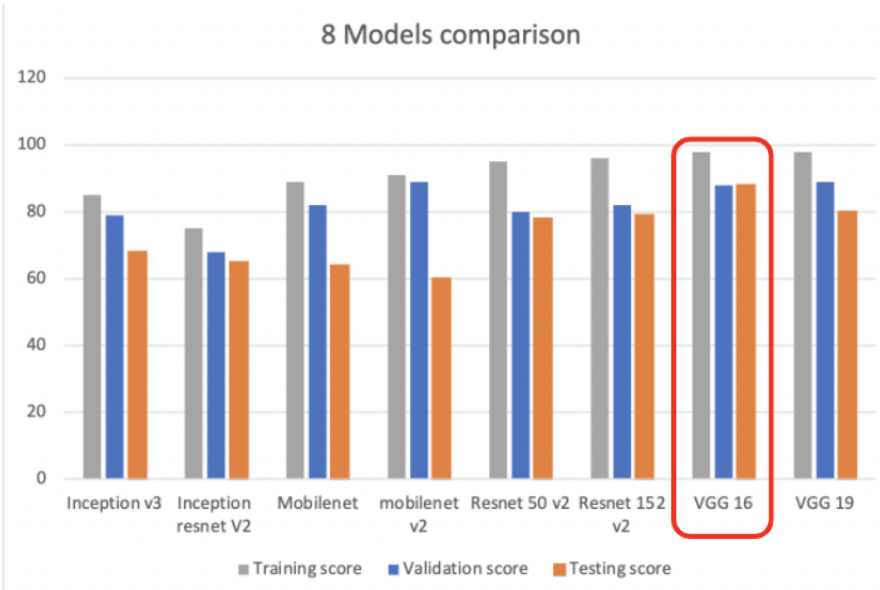


Pic16, 17. The accuracy and loss changes when training the model



Pic18, 19 The Confusion matrix and model structure of our CNN-VGG16 model





4.2.6 Hypereparameter tuning

In the training process, we've tuned the model learning rate (LR), the train and validation split percentage, and the Optimizer. By using Other's model structure, we're also required to tune the fine-tuning epochs and the model structure that we use, which is also a hyperparameter.

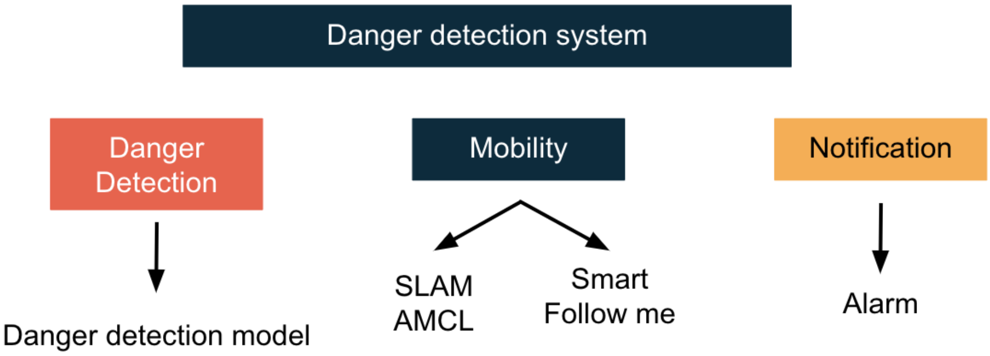
As we've used VGG16, so we've tuned our first hyperparameter. After the first training process, we've found that LR of the fine-tuning epoch must be smaller due to the fine-tuning process being only small changes of the model, so large training steps may make the model overfit. At last, we've tuned our training LR and fine-tuning LR as 1e-5, 1e-8.

At the training process and fine-tuning process, we've used Adam and RMSprop as the optimizer. We've tried to tune the validation split while we've low accuracy of our model training result. We finally set the validation split as 0.2, which means 20% of data will be used as validation

5. Conclusion:

5.1 Project conclusion

In conclusion, the aging population is an assignable question. Elderlies is facing more danger in the future. Besides, robots replacing humans is the trend of the future. So, we invented this robot to serve elderlies, take care of elderlies, and check if they are safe.



5.2 Advantage of our project

The home robot nowadays can already work indoors to take care of the elderly. However, they couldn't monitor the danger to the elderly. This is as such the stand out point of our project. Through the technology of observing the limbs of the elderly, we can better monitor the status of the elderly, and immediately notify rescuer when the elderly faint, fall, or have sudden illnesses and increase the probability of the elderly being rescued, to better protect the elderly. In addition, we can also use robotic arms, storage space, screens, and other components to serve and communicate with the elderly.

6. Future:

In the future, we may connect our robot with a wristband or watch so that the accuracy of fall detection can increase. We are also going to visit some places where elderlies live, such as care and attention homes for the elderly and elderly centers. We will interview these places to test our robot and get more data.

7. Reference

[1]: The 7 steps of machine learning https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e
[2][3]: Population proportion of 1970, Population proportion of 2020
https://population.un.org/wpp/Maps/
[4]: picture of ElliQ
https://www.gadgenda.com/portal-img/default/6/elliq-showing-pictures.jpg
[5]: picture or ROBEAR
https://d.ibtimes.co.uk/en/full/1425745/robear-new-care-support-robot-bear-nurse-that-can-lift-patients-gently-transfer-them-between.jpg
[6]: picture of Zora
https://www.steadytech.com.au/wp-content/uploads/2020/06/img_0833.jpg

【評語】190039

In this project, the author aims to build a robot collocated with its danger detection system to ensure the safety of the elderly at home. The core technology in this project is image recognition. This is a complete project, including the performance evaluation. However, the topic is not a new one. The author may consider to have more survey on the related works.