Dissertation on

Multiple face recognition and gender detection of group images

Thesis submitted towards partial fulfilment of the requirements for the degree of

Master of Technology in IT (Courseware Engineering)

Submitted by

Debarati Raha

EXAMINATION ROLL NO.: M4CWE22003
UNIVERSITY REGISTRATION NO.: 154485

Under the guidance of

Prof. Joydeep Mukherjee

School of Education Technology
Jadavpur University

Course affiliated to

Faculty of Engineering and Technology
Jadavpur University
Kolkata-700032
India

2022

M.Tech. IT (Courseware Engineering)
Course affiliated to
Faculty of Engineering and Technology
Jadavpur University
Kolkata, India

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled **"Multiple face detection and gender detection of group images"** is a bonafide work carried out by **Debarati Raha** under our supervision and guidance for partial fulfilment of the requirements for the degree of Master of Technology in IT (Courseware Engineering) in School of Education Technology, during the academic session 2021-2022.

SUPERVISOR
School of Education Technology
Jadavpur University,
Kolkata-700 032

DIRECTOR
School of Education Technology
Jadavpur University,
Kolkata-700 032

DEAN - FISLM
Jadavpur University,
Kolkata-700 032

M.Tech. IT (Courseware Engineering)
Course affiliated to
Faculty of Engineering and Technology
Jadavpur University
Kolkata, India

CERTIFICATE OF APPROVAL **

This foregoing thesis is hereby approved as a credible study of an engineering subject carried out and presented in a manner satisfactory to warranty its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not endorse or approve any statement made or opinion expressed or conclusion drawn therein but approve the thesis only for purpose for which it has been submitted.

	--
Committee of final examination for evaluation of Thesis	--
	--
	--

** Only in case the thesis is approved.

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis contains literature survey and original research work by the undersigned candidate, as part of her **Master of Technology in IT (Courseware Engineering)** studies.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by this rule and conduct, I have fully cited and referenced all materials and results that are not original to this work.

NAME: DEBARATI RAHA

EXAMINATION ROLL NUMBER: M4CWE22003

THESIS TITLE: Multiple face detection and gender detection of group images

SIGNATURE: DATE:

Acknowledgement

With profound sense of regard and gratitude I would like to first thank my guide Mr. Joydeep Mukherjee for having me as his student and providing me the opportunity to pursue my M.Tech research that led to this dissertation. It has been a privilege to have him as my advisor. His enthusiasm and love for research motivated me to pursue research as a career. He has been extremely helpful, provided me with great insight on various problems and immense encouragement through my stint at masters. I would like to thank him for his immense guidance and support throughout the project. He helped me a lot to understand the Finite Element Method concepts, and supported throughout my project. Above all I would like to thank my parents and all my family members and friends for constantly supporting me in building my career.

Contents

LIST OF FIGURES

0.6 Comparative Analysis

0.7 Conclusions and Future Scopes

0.8 References

0.1 Executive Summary

Foundational facial recognition concepts emerged more than 50 years ago.1 Recent advancements in machine learning, coupled with advancements in camera and computer vision technologies, have accelerated the design, development, testing, deployment, and operation of facial recognition systems. Concerns about systems used to collect, track, or surveil a unique and exposed part of the human body - one that is, for many, directly associated with identity, privacy, safety, democracy, and security - raise important questions about the appropriate role of this technology in society.2 These considerations have prompted calls for policymakers around the world to take immediate steps to determine whether and how facial recognition systems can be used to benefit people without violating human rights and civil liberties. 3 The Partnership on AI (PAI) believes that policymakers must understand how facial recognition systems work in order to craft comprehensive legal and regulatory environments.4 PAI's Facial Recognition Project is intended to demystify facial recognition systems and provide a common language for policymakers and other stakeholders to use when discussing and evaluating their capabilities.5 Explaining these systems can help bridge conversations between those developing and using the technology, policymakers, and those whose faces and names are wittingly or unwittingly included in these systems. This paper is the result of a series of workshops on facial recognition systems convened by PAI between September 2019 and January 2020. The workshop series brought together Partner organizations and communities developing, engaging with, and affected by these systems. Presenters illuminated the state of the art in today's systems, described advancements in research, and provided societal context for the environments where facial recognition technologies are currently being deployed. PAI's Facial Recognition Project also reinforces the importance of increasing transparency and understanding around the design, development, testing, procurement, deployment, and operation of AI systems - especially those deployed in high-stakes domains. To further this objective, we define facial recognition systems, and illustrate how they work. We also include a list of questions for policymakers and other stakeholders to elicit additional information about the technical and related aspects of facial recognition systems. While specific policy recommendations related to the use of facial recognition systems are out of the scope of this paper, it sheds light on common misunderstandings about these systems and is intended to provide useful information to inform the important policy debates unfolding on this topic around the world.

Face recognition has been growing rapidly in the past few years for its multiple uses in the areas of Law Enforcement, Biometrics, Security, and other commercial uses. Face appears to offer several advantages over other biometric methods. Face detection is an important area of research in computer vision, as it serves as a necessary first step, for any face processing system, such as human computer interfacing, crowd surveillance, content based image retrieval, real time surveillance system, secure access control, video conferencing, financial transaction, forensic applications, pedestrian detection, driver alertness monitoring systems etc. Face detection in unconstrained environment is an important first task to be performed in automating these activities. In terms of applications, face detection is quite important for the face recognition problem, since it is the pre-processing step for a face recognition system. The main goal of the face detection is to detect and locate faces in any arbitrary digital image or video sequence. This involves making the machine intelligent enough to acquire the human perception and knowledge to detect, localize and recognize the face in any arbitrary image with the same ease as we humans do it. In the present thesis, a novel approach to perform face detection and face recognition is proposed. The problem of face recognition has been addressed by functionally dividing it into face detection and face recognition. Different approaches to the problems of face detection and face recognition have been evaluated, and implemented using the **yolov5** technical computing language with the help of **roboflow.**

YOLO is **an object detection algorithm that divides images into a grid system**. Each cell in the grid is responsible for detecting objects within itself.

Face detection research has gone through lots of transitions since 1970. However, existing algorithms for face detection are inadequate for the class of images with multiple faces under varying illumination conditions. In this research work face detection algorithms for such a class of images have been proposed.

While it is important to understand how these systems work, PAI also recognizes that facial recognition systems are developed by humans, and their use cannot be separated from existing cultural, social, and economic power dynamics. These systems can make some aspects of life easier, and they can also amplify civil liberties and human rights concerns, including challenges of bias.6 7 PAI believes that meaningfully engaging underrepresented and at-risk communities, including women and gender non-binary people, communities of color, the

LGBTQI community, immigrants, workers, those with disabilities, low-income individuals, and religious minorities, is essential for truly equitable outcomes.

Our work is informed by the following key findings and understandings:

• **Facial Recognition Systems Defined** - Facial recognition systems predict similarity between two faces in order to attempt to verify or determine someone’s identity.

• **How These Systems Work** - A facial recognition system works by first detecting whether an image contains a face. If so, it then tries to recognize the face in one of two ways:

• **During facial verification**, the system attempts to verify the identity of the face. It does so by determining whether the face in the image potentially matches a specific face (identity) previously stored in the system.

• **During facial identification**, the system attempts to predict the identity of the face. It does so by determining whether the face in the image potentially matches any of the faces (identities) previously stored in the system.

• **Each System is Unique** - There is no one standard system design for facial recognition systems. Not only do organizations build their systems differently, and for different environments, but they also use different terms to describe how their systems work. The explanations in this paper, informed by briefings from experts participating in our workshops, aim to provide a consistent set of descriptions to ground future discussions.

• **Design Matters** - The results that facial recognition systems present to users are dependent on how the systems were designed, developed, tested, deployed, and operated. The impact of key aspects of the system such as training datasets, enrolment databases, and match thresholds need to be understood in order to properly evaluate these results.

• **Beyond Facial Recognition** - "Facial recognition" is sometimes described as encompassing facial characterization - also called facial analysis - systems, which detect facial attributes in an image, and then sort the faces by categories such as hair colour, gender, or race. We do not consider such systems to be a part of facial recognition systems because they are not used to predict the identity of a person.

During facial verification, the system attempts to verify the identity of the face. It does so by determining whether the face in the image potentially matches a specific face (identity) previously stored in the system. During facial identification, the system attempts to predict the identity of the face.

The problem of face detection can be viewed as a problem of binary classification of image as either containing or not containing faces. In order to classify it is required to describe an image in terms of features, which would be good indicators of face presence or absence on a given image. This requires robust features which distinguish the face images from other non-face images and hence contributing for good classification. In the current work, in order to speed up the process, a holistic approach is used rather than looking for individual facial features.

Two approaches are proposed to perform Face detection. The first method is performed using **Edge Tracking Algorithm (YOLO Algorithm)** and **Neural Network** . Then, to improve face detection in blurred scenes or low-resolution situations, we integrated image super resolution technology on the detection head. In addition, some representative deep-learning algorithm based on face detection is discussed by grouping them into a few major categories, and the popular face detection benchmarks are enumerated in detail. Finally, the wider face dataset is used to train and test the SR-YOLOv5 model. Compared with multitask convolutional neural network (MTCNN), Contextual Multi-Scale Region-based CNN (CMS-RCNN), Finding Tiny Faces (HR), Single Shot Scale-invariant Face Detector (S3FD), and TinaFace algorithms, it is verified that the proposed model has higher detection precision, which is 0.7%, 0.6%, and 2.9% higher than the top one. **SR-YOLOv5** can effectively use face information to accurately detect hard-to-detect face targets in complex scenes. Detection rate is also compared for all the schemes. The detection rate of this model is higher than the other techniques(USING **MATLAB).**

The algorithms proposed in this work can also detect faces under slight variations in pose, bright sunlight, normal room lighting conditions with slight variations in expressions and with spectacles.

The detection and recognition as male or female, rates are compared with the existing algorithms. In the present research, the performance of the algorithms with the real time database is tested and the results are presented.

0.2 Introduction

2.1 Objective of Face Detection

In the current day scenario where most of the human activities are mechanized and automated, computers play a major role in the activities of human life. Human beings have designed and deployed different kinds of machines possessing the required intelligence to carry out the intended task. There are many such important activities like face detection, recognition, object detection and tracking, gesture recognition, surveillance system where the machine capabilities can be used at its best. One such important activity considered in this work is face detection and localization. The main objective of the face detection is to check whether the given image contains any face and if faces are present finding out the exact location of the faces in any colour image. This involves making the machine intelligent enough to acquire the human perception and knowledge to detect, localize and recognize the face in any arbitrary image with the same ease as humans do it.

The early work on face detection dates back to early 70s, where simple heuristics and knowledge about facial structure were used to convert them in to simple rules. Most of the research work used single upright frontal images with plain or uniform 3 background. Any slight change in these conditions required changing the rules. It is around mid-90s the work on face detection has picked up and different researchers have come up with different ideas to address this problem efficiently.

2.2 Applications of Face Detection

In terms of applications, face detection is quite important for the face recognition problem, since it is the preprocessing step for a face recognition system. Facial detection technology has been integrated into a wide range of products and services, including online social networks, digital billboards, and mobile apps. Face detection also has potential applications in secure access control, video conferencing, financial transaction, forensic applications, pedestrian detection, driver alertness monitoring systems, crowd surveillance etc.

YOLO is an object detection algorithm that divides images into a grid system. Each cell in the grid is responsible for detecting objects within itself. YOLO is one of the most famous object detection algorithms due to its speed and accuracy. Shortly after the release of YOLOv4 Glenn Jocher introduced YOLOv5 using the Pytorch framework.

Author: GlennJocher

Released: 18 May 2020

There are several versions of YOLO. Those are, YOLOv1, YOLOv2, YOLOv3, YOLOv4 & YOLOv5. In this project the most updated version of YOLO is used i.e YOLOv5.

At first we need to create a dataset,i.e the group of images. Here we used the **Fatkun batch image downloader.** After that 128 images are used for both training and validation to verify our training pipeline is capable of overfitting. **data/coco128.yaml**, is the dataset configuration file that defines 1) an optional

download command/URL for auto-downloading, 2) a path to a directory of training images (or path to a *.txt file with a list of training images), 3) the same for our validation images, 4) the number of classes, 5) a list of class names.
After that labels should be created. Then export the labels to **YOLO format.** Images will be objectedon 2D plane(Each row is **class x_center y_center width** height format).
The third step is Organize Directories, in this step we Organize train images and labels according to the **yolov5 directory**.**YOLOv5 locates labels automatically for each image** by replacing the last instance of /images/ in each image path with /labels/. Select a pretrained model to start training from.
Here we select **YOLOv5s**, the smallest and fastest model available. Train the **YOLOv5s** model on **COCO128** by specifying dataset, batch-size, image size and either pretrained --**weights yolov5s**.pt (recommended), or randomly initialized --**weights** '' --**cfg yolov5s**.yaml (not recommended). Pretrained weights are auto-downloaded from the **latest YOLOv5 release.**
All training results are saved to **runs/train/** with incrementing run directories. For more details we can see the Training section of our **Google Colab** Notebook.

Weights & Biases (W&B) is now integrated with **YOLOv5** for real-time visualization and cloud logging of training runs. This allows for better run comparison and introspection, as well improved visibility and collaboration among team members. To enable W&B logging install wandb, and then train normally.
By using **Weights & Biases** we can **integrate quickly**, i.e track, compare, and visualize ML experiments with 5 lines of code. Free for academic and open source projects.

Second one is **visualize seamlessly**. Add W&B's lightweight integration to the existing ML code and quickly get live metrics, terminal logs, and system stats streamed to the centralized dashboard.

Third one is **collaborate in real time**, this explain how our model works, like show graphs of how model versions improved, discuss bugs, and demonstrate progress towards milestones.

Local Logging is another type of visualize. All results are logged by default to runs/train, with a new experiment directory created for each new training as runs/train/exp2, runs/train/exp3, etc. View train and test jpgs to see mosaics, labels, predictions and augmentation effects. Note a **Mosaic Dataloader** is used for training (shown below), a new concept developed by Ultralytics and first featured in **YOLOv4**.

Training losses and performance metrics are also logged to Tensorboard and a custom results.txt logfile which is plotted as results.png after training completes. Here we show **YOLOv5s** trained on **COCO128** to **300 epochs** (An epoch in machine learning means **one complete pass of the training dataset through the algorithm**. This epochs number is an important hyperparameter for the algorithm. It specifies the number of epochs or complete passes of the entire training dataset passing through the training or learning process of the algorithm. With each epoch, the dataset's internal model parameters are updated. Hence, a 1 batch epoch is called the batch gradient descent learning algorithm. Normally the batch size of an epoch is 1 or more and is always an integer value in what is epoch number.

It can also be visualized as a 'for-loop with a specified epoch number with each loop path traversing the entire training dataset. In the for-loop is a nested for-loop that allows the loop to iterate over a specified sample number in a single batch when the samples “batch size” number is specified as one. Typical values of the number of epochs when training algorithms can run into thousands of epochs, and the process is set to continue until the model error is sufficiently minimized. Normally tutorials and examples use values like 10, 500, 100, 1000, or even larger numbers. We used 300 epochs here to obtain the result. Line plots can be created for the training process, with the X-axis having the epoch in machine learning and the Y-axis having the skill or model error. Such line plots are called the learning curve of the algorithm and help diagnose problems such as fitting of the training set being under, over or suitably learned. epoch in the neural network.

Here we want to highlight the difference between Epochs and Batches for reference, model gets updated when a specific number of samples are processed. This is known as the batch size of samples. The number of training dataset’s complete passes is also significant and called the epoch in machine learning number in the training dataset. Batch size is typically equal to 1 and can be equal to or less than the training dataset’s sample number. The epoch in a neural network or epoch number is typically an integer value lying between 1 and infinity. Thus one can run the algorithm for any period of time. To stop the algorithm from running, one can use a fixed epoch number and also use the factor of rate of change of model error being zero over a period of time.

Both batch size and epoch in machine learning of learning algorithms are hyper-parameters with integers as values used by the training model.

These values are not found by a learning process since they are not internal parameters of the model and must be specified for the process when training an algorithm on the training dataset. These numbers are also not fixed values and, depending on the algorithm, may require trying various integer values before finding the most suitable values for the process. In discovering differences in gradient descent that is stochastic in an epoch in machine learning and batches, one can say that the gradient descent stochastic algorithm uses a dataset for training with its learning algorithm that is iterative when updating the model. The batch size is a gradient descent hyperparameter that trains the training samples numbers before the internal parameters of the model are updated to work through the batch. The epoch number is again a gradient descent hyperparameter that defines the numbers of passes that are complete when passing through datasets under training.),

starting from scratch, and from pretrained --weights yolov5s.pt. YOLOv5 may be run in any of the following up-to-date verified environments,

- **Google Colab and Kaggle**
- **Google Cloud**
- **Amazon** Deep Learning AMI.
- **Docker Image**.

2.3 Difficulties

Face detection is difficult mainly due to large variations in shape, colour, size and textural differences among faces. Differences in facial appearance of different people also contribute to the complexity of the problem. Additional features such as spectacles or a moustache, occlusion, lighting conditions and a cluttered image back ground can make face detection a more complex problem. In the past several years, many face detection techniques have been proposed throughout the literature that addresses most of the above problems. Pose variations, hardware characteristics and setting conditions of camera as well as imaging environment add more constraints to feature space used for face detection. However, there is still no single face detection technique available that fully addresses these problems as a whole. Face detection is a difficult pattern recognition problem, because of the high variability of the face appearance. Faces are non-rigid, dynamic objects with a large diversity in shape, colour and texture.

Multiple factors such as head pose, lighting conditions (contrast, shadows), facial expressions, occlusions (spectacles, make-up) and other facial features (beard, tresses and moustache) make face detection a complex problem. Humans can effortlessly detect and recognize faces with all these variations.

The challenge is to make the machine capable of detecting and recognizing faces with all these variations. In general, a good face detection system should handle faces of different shapes, sizes (face size and appearance vary from person to person) as well as different possible poses in which human faces occur in different image clippings. Humans can very easily recognize and distinguish actual faces from cartoons and drawings. Machines should be trained to effectively distinguish the actual faces from the cartoons/drawings and also precisely locating the faces irrespective of the image background and illumination condition. Human beings possess an inherent capability to identify and recognize human faces in any position and even if it is partially visible with any type of occlusion. Algorithms should be devised to enable machines to detect faces with any type of occlusion and position. There is a need for good searching process for locating the faces in images. Good face detection system should search only the regions which contain faces and avoid searching regions which do not contain faces and thus reduce the computational costs and increase the running time efficiency of the algorithm, as algorithm’s running time and space are inversely related. Face detection also provides interesting challenges to the underlying pattern classification and learning techniques. When a raw or unfiltered image is considered as input to a pattern classifier, the dimension of the feature space is extremely large (i.e., the number of pixels in normalized training images). Having a robust feature space with reduced dimension demands a good feature extraction method, and in turn reducing the computational cost is one of the key aspects of an efficient face detection system. Some of the research issues involved in face detection are extraction of strong and robust features representing the face in any type of imaging conditions, locating faces of different sizes, finding the precise locations of faces in images, speeding up the process of detecting and locating faces in camouflaged images. Some of the training procedures are is very time consuming. There is a need for robust feature set which consumes less time for training, so we trained a model with the halp of **YOLOv5** and it's object detection algorithm which will be much more efficient and less time consuming. It has much more efficiency than other models.

2.4 Problem Statement

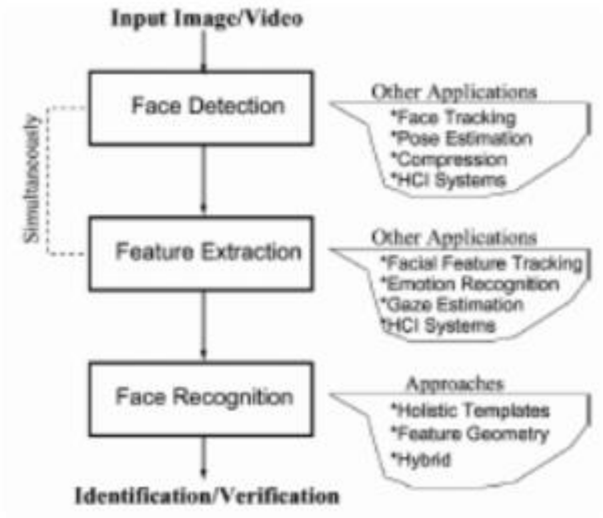
In this research work, the human faces and genders both are detected accurately while wearing the glasses and it detects the gender from any of the side profiles (i.e. left or right). In this proposed approach more storage, many variable, low speed, and lack of face database testing are used for better performance. In this research work the objectives are more about helping & solving the social problems with the appropriate usage of technologies for the help of mankind. The research work has been motivated by the problems like lack of security, frauds, child molestation, rapes, robbery, criminal identification etc. Now-a-days all over the world the crime rates have increased extensively where the ones who are associated with the crimes have also become more and more intelligent to sneak through the eyes of the police, security systems & government which leads the crimes become unbeatable so we need to take the help of the technologies for identification of the criminals and make our security systems more stringent. This research work is just a token to help the mankind for better gender and face detection for better security purpose.

The project can also be used to overcome the frauds that can take place during voting i.e. can be used for voter identification. The old generation has the difficulty to operate computers with ease. This bridge can be lessened by improving Human-Computer Interaction (HCI). The child molestation cases can be tackled at a faster rate by comparing school surveillance camera images to know child molesters and the same can be used for verifying the court records thereby minimizing victim trauma. Similarly, it can also be used for surveillance at banks and residential areas.

0.3 Literature Survey

Approach of face recognition aims to detect faces in still image and sequence image from video have many method such as local, global, and hybrid approach. The main problem of face recognition are intensity, illumination, pose, difficult to controlling and large occlusion. In 3D capture creates larger data files per subject which applies significant storage requirements, slow processing, most new devices can be capture in 3D. This is the problem for our future work that want to solving and create accuracy gain for widely accept in face recognition system.

Face recognition is one of the few biometric methods that possess the merits of both high accuracy and low intrusiveness. It has the accuracy of physiological approach without being intrusive. Over past 30 years, many researchers have been proposed different face recognition techniques, motivated by the increased number of real world applications requiring the recognition of human faces. There are several problems that make automatic face recognition a very difficult task. However, the face image of a person inputs to the database that is usually acquired under different conditions. The important of automatic face recognition is much be cope with numerous variations of images of the same face due to changes in the following parameters such as pose, illumination, expression, motion, facial hair, glasses, and background Face recognition technology is well advance that can applied for many commercial applications such as personal identification, security system, image-film processing, psychology, computer interaction, entertainment system, smart card, law enforcement, surveillance and so on. A general problem of face recognition can be done in two formulate both a still image and video image of a scene. There have divided into two basic applications: identification and verification. In the identification problem, the face to be recognized is unknown and matched against face of a database containing known individuals. In the verification problem the system confirms or rejects the claimed identify of the input face. However, before face recognition is performed. the system should determine whether or not there is a face in a given image or given video, a sequence of images. This process is called face detection. Once a face is detected, face region should be isolated from the scene for the face recognition.



source :https://www.google.com/search?q=literature+review+of+face+detection&rlz=1C1UEAD_enIN996IN996&sxsrf=ALiCzsbzn1t_PreDTaHIDzvAMD0hGwsuaQ:1661235112730&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjl_KCGp9z5AhWbm1YBHYJYD9QQ_AUoAnoECAEQBA&cshid=1661235113882800&biw=1280&bih=609&dpr=1.5#imgrc=jkADq4EH2gjX5M

There are so many techniques and methodologies for face recognition and gender detection. Here we're going to discuss some previous face recognition and gender detection approaches. The field of face detection has achieved considerable progress in the last decade especially, since the seminal work by P. Viola and Jones [36], [53] has made face detection practically feasible in real world applications such as digital cameras. A number of methods have been proposed to detect faces in images. Most of these methods are reviewed in survey papers by M. H. Yang et al. [47], Hjelmas et al. [29], Cha Zhang et al. [75], as well as W. Zhao et al. [89]. Hjelmas et al. [29] has organized them into two categories namely Feature-based approaches and Image-based approaches. Feature based techniques are further classified as low level analysis, feature analysis, active shape model based. Low level analysis deals with edges, height to width ratio, gray information, colour and motion. Feature analysis deals with feature searching, constellation searching where as active shape models are categorized into snakes, deformable templates and point distribution models. In general, face detection methods are broadly classified as holistic methods and featurebased methods. Holistic methods use the whole face region as raw input to the system. In feature-based methods, local features on face such as nose, eyes and mouth are extracted and then used as input data. Different approaches have been used to extract these facial feature points from images or video sequences containing faces. In feature based techniques, the probable face region is checked for the presence of various facial features such as eyes, nose and mouth. Image based approaches are further classified as subspace based, neural network based, statistical approach based. These image based approaches are holistic in nature and use a scanning window which scans across the whole image at different scales and resolutions to locate the faces. Face

detection methods can be broadly classified into four categories; knowledge-based approaches, feature invariant approaches,
template matching approaches and appearance-based approaches. The details of different face detection and classification techniques used by different researchers are available in the survey papers of Face Recognition and Identification using Deep Learning Approach To cite this article: KH Teoh et al 2021 J. Phys.: Conf. Ser. 1755 012006.

3.2 Literature Survey Based on Different Classification Techniques

Face recognition can be done in both a still image and video sequence which has its origin in still-image face recognition. Different approaches of face recognition for still images can be categorized into tree main groups such as holistic approach, feature-based approach, and hybrid approach.

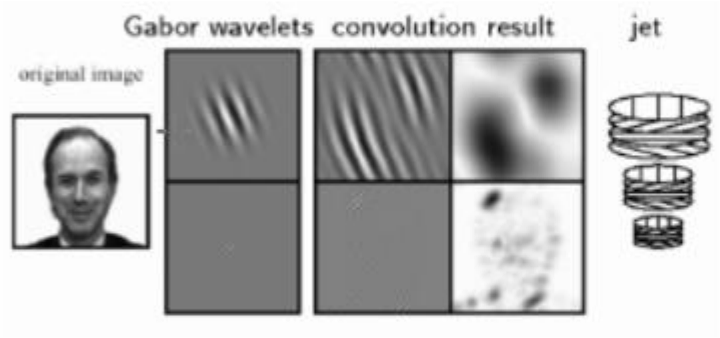
Based on different feature extraction and classification methodologies used by different researchers, the face detection methodology can be grouped as feature based, skin pixel based, statistical measures with neural networks and support vector machines, linear subspace, wavelets and fuzzy approach/genetic algorithms.

3.2.1 Feature Based Approach

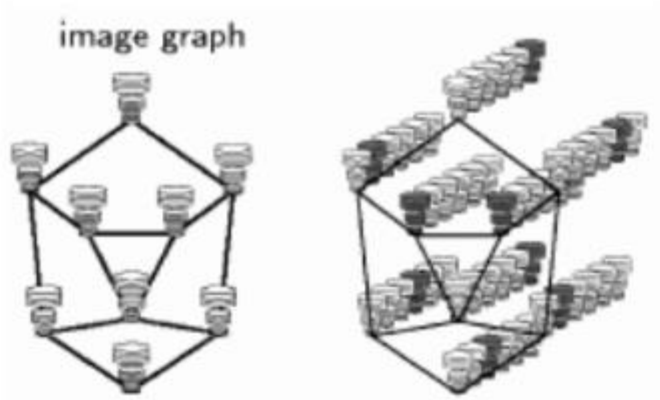
In feature-based approaches or local feature that is the features on face such as nose, and then eyes are segmented and then used as input data for structural classifier. Pure geometry, dynamic link architecture, and hidden Markov model methods belong to this category. One of the most successful of these systems is the Elastic Bunch Graph Matching (EBGM) system, which is based on DLA.

Wavelets, especially Gabor wavelets, play a building block role for facial representation in these graph matching Methods. A typical local feature representation consists of wavelet coefficients for different scales and rotations based on fixed wavelet bases. These locally estimated wavelet coefficients are robust to illumination change, translation, distortion, rotation, and scaling. [42], [43]. The grid is appropriately positioned over the image and is stored with each grid point's locally determined jet in figure 2(a), and serves to represent the pattern classes. Recognition of a new image takes place by transforming the image into the grid of jets, and matching all stored model graphs to the image. Conformation of the DLA is done by establishing and dynamically modifying links between vertices in the model domain.

DLAs attempt to solve some of the conceptual problems of conventional artificial neural networks, the most prominent of these being the representation of syntactical relationships in neural networks.
DLAs use synaptic plasticity and are able to form sets of neurons grouped into structured graphs while maintaining the advantages of neural systems. The DLA architecture was recently extended to Elastic Bunch Graph Matching in figure 10[41] This is similar to the graph described above, but instead of attaching only a single jet to each node, the authors attached a set of jets that can show bunch graph representation in below figure.



(a) Elastic graph representation



(b) Bunch graph

Source:
https://www.google.com/search?q=literature+review+of+face+detection&rlz=1C1UEAD_enIN996IN996&sxsrf=ALiCzsbzn1t_PreDTaHIDzvAMD0hGwsuaQ:1661235112730&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjl_KCGp9z5AhWbm1YBHYJYD9QQ_AUoAnoECAEQBA&cshid=1661235113882800&biw=1280&bih=609&dpr=1.5#imgrc=bxdGV1U1cqYP0M

This is the most old and very commonly used approach though it's very efficient to detect the faces and gender too. In this approach while training the entire dataset we have to add and extracts features. This approach looks for individual facial features such as eyes, ear, mouth based on edge strength and intensity, image texture, shape, colour. Then these features are grouped into candidate regions and classified. Several researchers have used this feature based approach to detect faces in the input images.

In this **Feature Based Approach** papers have located eye analogue segments at a given scale by finding regions which are roughly as large as eyes and darker than their neighborhoods. A pair of eye analogue segments are hypothesized to be eyes in a face and combined into a face candidate if their placement is consistent with the anthropological characteristic of human eyes. This method is robust as it can deal with illumination change and works well for frontal faces with eyes wide open.

3.2.2 Skin Pixel based Approach

Skin region detection is a very important problem in human-computer interfaces. The most practical application of this is skin segmentation based face detection. The main aim is to identify and select skin pixel regions and avoid other non-skin regions from further processing. Irrespective of the different colour spaces used for classification, the main goal is to classify whether any arbitrary pixel of the image under consideration represents skin colour in the chosen colour space. Chrominance based 2D models are preferred as they provide better generalization of skin colour irrespective of the different ethnic groups of human race. The skin colour of different people and of different ethnic origins, e.g. Caucasian, Asian, African, etc., differs mainly in terms of intensity rather than chrominance. Thus chrominance space is preferable taking robustness to illumination component and dimensionality reduction into consideration. Many researchers have exploited this information very effectively. Segmentation of facial regions based on skin colour information requires choosing a suitable colour space to identify skin regions. In [13] Lin introduced an approach for the detection of faces in colour images that composed of two stages.

In [13] Lin introduced an approach for the detection of faces in colour images that composed of two stages. The first stage involves converting the input RGB image into a binary image based on colour segmentation using the relative ratio between the R, G, and B components. This is to eliminate the effect of lighting conditions. In the second stage, 4-connected components are identified and labelled. The centre of each block and the three other blocks are scanned to find the combination that forms an isosceles triangle. These components are potential face regions.

Sanjay Kr. Singh1 et al. [61] have considered three colour spaces, RGB, YCbCr and HSI. They have compared the algorithms based on these colour spaces and have combined them to get a new skin colour based face detection algorithm which gives higher accuracy.

They assert that YCbCr and HSI colour space are more efficient in comparison to RGB to classify the skin region. But still, individually, both these spaces are not able to give very good results. Merging the segmented regions obtained in all the three colour spaces, facial feature (eyes, ear and mouth) are extracted from the merged region using suitable threshold, followed triangle formation and ratio of distance estimation hypothesis which finally detects the presence of face. It is robust and efficient in segmenting the skin colour region followed by face region classification for images with frontal faces.

In [28] Hichem Sahbi et al. present a skin colour approach for fast and accurate face detection which combines skin colour learning and image segmentation. This approach starts from a coarse segmentation which provides regions of homogeneous statistical colour distribution. Some regions represent parts of human skin and are selected by minimizing an error between the colour distribution of each region and the output of a compression decompression neural network, which learns skin colour distribution for several populations of different ethnicity. This **Artificial Neural network(ANN)** is used to find a collection of skin regions which are used to estimate the new parameters of the Gaussian models using a 2-means fuzzy clustering in order to adapt these parameters to the context of the input image. A Bayesian frame work is used to perform a finer classification and makes the skin and face detection process invariant to scale and lighting conditions. Finally, a face shape based model is used to validate the face hypothesis on each skin region.

3.2.3 Statistical Measures with Neural Networks and Support Vector Machines

This method is very popular method for detecting faces and it's use widely. It consumes lesser time than previous approaches. So it has high efficiency to detect faces and also detect the gender with the help of **Neural Networks and Support Vector Machines.**

Support vector machine (ın machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features.

If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three. **Support Vector Machine(SVM)** is a supervised machine learning algorithm used for both classification and regression) is an important part of ML and DL . there are some advantages of **svm** those makes this method more helpful those are:

Advantages of SVM:

- Effective in high dimensional cases
- Its memory efficient as it uses a subset of training points in the decision function called support vectors
- Different kernel functions can be specified for the decision functions and its possible to specify custom kernels

Statistical Measures with Neural Networks and Support Vector Machines related Several researchers have considered statistical information such as two dimensional central geometrical moments [59], Spectral histograms [15], colour histogram [35] and edge information [2], [22], [9], [15], [31], [38], [50], [71], [73], for face detection. The use of statistics and neural networks has also enabled faces to be detected from cluttered scenes at different distances from the camera. According to [16] and Yongmin Lia et al. [72], since only the face and non-face patterns located around the boundary are of prime interest. Estimating a boundary which robustly separates the two classes of patterns is more promising than other methods such as probabilistic modeling of the two classes.

The main challenge in object detection is the amount of variation in visual appearance. H. Schneiderman and T. Kanade [84] describe a statistical method for 3D object detection. They represent the statistics of both object appearance and “non-object” appearance using a product of histograms. Each histogram represents the joint statistics of a subset of wavelet coefficients and their position on the object. They use many such histograms representing a wide variety of visual attributes. Using this method, they have developed the first algorithm that can reliably detect human faces that vary from frontal view to full profile view. In [15], [72] spectral histogram and Support Vector Machines are used for classification. M.Saadia et al. [50] presented a new method to localise a face in an image. The originality of work presented consists on the use of vectors of geometrical moments like entries to a Forward Back-Propagation neural network. Using this approach one can localize individual facial features like eyes, mouth, nose etc in images and video.

Rowley, et al. [80], [82] presented a neural network-based algorithm to detect upright, frontal views of faces in gray-scale images. The algorithm works by applying one or more neural networks directly to portions of the input image, and arbitrating their results. Each network is trained to output the presence or absence of a face. The algorithms and training methods are designed to be general, with little customization for faces. This approach is useful in detecting upright frontal faces facing camera. It could detect 90.5% of the faces over a test set of 130 complex images, with an acceptable number of false positives. The images contain a total of 507 frontal faces, and require the networks to examine 83,099,211 20x20 pixel windows and it is computationally expensive.

Paul Viola, et al. [36], [53] they describe a face detection framework that is capable of processing images extremely rapidly while achieving high detection rates. Instead of directly using pixel information, P. Viola et al. [53] used a set of simple Haar like features. A feature is computed by summing the pixels in the white region and subtracting those in the dark region. Haar-like features can be computed efficiently with the integral image representation or summed area table. There are three key contributions in this approach. The first is the introduction of a new image representation called the "Integral Image" which allows the features used by the detector to be computed very quickly. At a given location (x; y) in an image, the value of the integral image ii(x; y) is the sum of the pixels above and to the left of (x; y): as given by equation

$$\mathbf{ii(x;\ y)} = \sum_{\mathbf{x}' \le \mathbf{x}, \mathbf{y}' \le \mathbf{y}} \mathbf{i(x';\ y')}$$

Here (x' ; y') is the pixel value of the original image at location (x' ; y').

Kevin Curran et al. [39] have worked on the construction of robust real-time face detection systems using neural networks. The Algorithm can detect between 67% and 85% of faces from images of varying size, background and quality with an acceptable number of false detections. It can detect between 'face' and 'non-face'. The normalization routine which makes training data uniform (especially where lighting is involved) – can drastically affect the classifying of images. The neural network approach is known to be highly sensitive to the grey levels in an image but by subjecting each trained and tested image to the routine the system sidestepped the problem. The conclusion was bootstrapping algorithm would have been more beneficial, had the database been continuously updated. At a very high threshold, only nine faces were detected and five of these were faces from the Yale face database and they were a part of the training set. Thus with a larger dataset of faces, the program can have better accuracy. A threshold of between 0.5-0.6 gives the best range of results out of the threshold set tested.
Kevin Curran et al. [39] have worked on the construction of robust real-time face detection systems using neural networks. The Algorithm can detect between 67% and 85% of faces from images of varying size, background and quality with an acceptable number of false detections. It can detect between 'face' and 'non-

face'. The normalization routine which makes training data uniform (especially where lighting is involved) – can drastically affect the classifying of images. The neural network approach is known to be highly sensitive to the grey levels in an image but by subjecting each trained and tested image to the routine the system sidestepped the problem. The conclusion was bootstrapping algorithm would have been more beneficial, had the database been continuously updated. At a very high threshold, only nine faces were detected and five of these were faces from the Yale face database and they were a part of the training set. Thus with a larger dataset of faces, the program can have better accuracy. A threshold of between 0.5-0.6 gives the best range of results out of the threshold set tested.

3.2.4 Holistic Approach

Holistic approaches consider the face as a whole and represent it by a single feature vector. The most popular methods include principal component analysis, and its extensions are independent component analysis (ICA) and linear discriminant analysis (LDA). Kirby and Sirovich [43] demonstrated that images of faces can be linearly encoded using a modest number of basis images. This demonstration is based on the Karhunen-Loeve transform, which also goes by other names, e.g., principal component analysis [30], [44] and the Hotelling transform [25]. Given a collection of n by m pixel training images represented as a vector of size m X n, basis vectors spanning an optimal subspace are determined such that the mean square error between the projection of the training images onto this subspace and the original images is minimized. They call the set of optimal basis vectors Eigen pictures since these are simply the eigenvectors of the covariance matrix computed from the vectorized face images in the training set. Turk and Pentland [44] applied principal component analysis to face recognition and detection. Principal component analysis on a training set of face images is performed to generate the Eigen-pictures (here called Eigenfaces) which span a subspace (called the face space) of the image space. Images of faces are projected onto the subspace and clustered. Similarly, non-face training images are projected onto the same subspace and clustered. Since images of faces do not change radically when projected onto the face space, while the projection of non-face images appear quite different. To detect the presence of a face in a scene, the distance between an image region and the face space is computed for all locations in the image. The distance from face space is used as a measure of “faceness,” and the result of calculating the distance from face space is a “face map.” A face can then be detected from the local minima of the face map.

The method describes an approach to detect faces whose size and position are unknown in an image with a complex background. The candidates of faces are detected by finding out "face like" regions in the input image using the fuzzy pattern matching method. The perceptually uniform colour space is used in their research in order to obtain reliable results. The skin colour is used to detect face like region, is represented by a model called skin colour distribution function. The skin colour regions are then extracted by estimating a measure that describes how well the colour of a pixel looks like the skin colour for each pixel in the input image.

The face like regions are extracted using a fuzzy pattern matching approach using these face models. The face candidates are then verified by estimating how well the extracted facial features fit a face model which describes the geometrical relations among facial features.

3.2.5 Fuzzy Approach/Genetic Algorithms

This approach is one of the preliminary approach among all. At the very beginning of this kind of research i.e feature based **Fuzzy Approach/Genetic Algorithms** was popular. It is one of the easiest approach. Fuzzy Approach/Genetic Algorithms presents a shape comparison approach to achieve fast, accurate face detection that is robust to changes in illumination and background. The proposed method is edge-based and works on gray scale still images. The Hausdorff distance is used as a similarity measure between a general face model and possible instances of the object within the image. A two-step process that allows both coarse detection and exact localization of faces is used. **Genetic Algorithms** are a powerful tool that can help in finding an appropriate model for face localization. The presented framework led to a model that performed considerably better than a simple hand-drawn model. Face localization can be improved by a multi-step detection approach that uses more than one model in different grades of detail. Each of these models can then be optimized separately. This does not only speed up the localization procedure but also produces more exact face coordinates.

All the above mentioned technologies require some voluntary action by the user, namely the user is supposed to place his hand on a handrest for finger printing or hand geometry detection and has to stand in a static position in front of a camera for Iris or Retina identification. Moreover, these technologies require more sophisticated equipments that are more sensitive to any body motion.

3.2.6 Hybrid Approach

The idea of this method comes from how human vision system perceives both holistic and local feature. The key factors that influence the performance of hybrid approach include how to determine which features should be combined and how to combine, so as to preserve their advantages and avert their disadvantages at the same time. These problems have close relationship with the multiple classifier system (MCS) and ensemble learning in the field of machine learning. Unfortunately, even in these fields, these problems remain unsolved. In spite of this, numerous efforts made in these fields indeed provide us some insights into solving these problems, and these lessons can be used as guidelines in designing a hybrid face recognition system. For example, components of a hybrid system, either feature or classifier, should be both accurate and diverse, such that a complementary advantage can be feasible. In fact, local features and global features have quite different properties and can hopefully offer complementary information about the classification task.

Table 1 summarizes qualitatively the difference between the two types of features. We can see from the table that local features and global ones are separately sensitive to different variation factors. For instance, illumination changes may have more influence on local features, while expression changes have more impact on holistic features. For these observations, hybrid approach that use both holistic and local information for recognition may be an effective way to reduce the complexity of classifiers and improve their generalization capability.

Table 1 Comparison of the local features and global features' sensitiveness to variations

Variation factors	Local features	Holistic features
Small Var.	not sensitive	sensitive
Large Var.	sensitive	very sensitive
Illuminations [31]	very sensitive	sensitive
Expressions [32],[33]	not sensitive	sensitive
Pose [34]	sensitive	very sensitive
Noise [35]	very sensitive	Sensitive
Occlusion [32],[33]	not sensitive	very sensitive

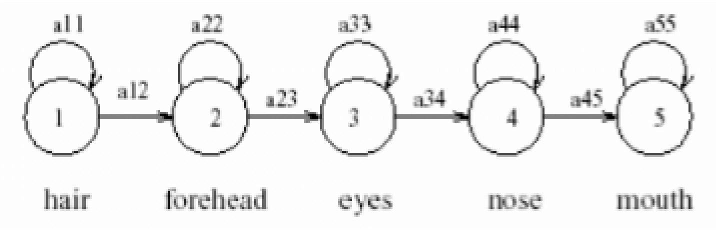
Despite the potential advantages, the work in this the category is still relatively few, possibly due to the difficulties mentioned above, while typical hybrid approach in traditional sense such as flexible appearance models [36], hybrid LFA, are generally not suitable for handling the one sample problem. Hope more researching efforts could be engaged in this approach, and in doing so, believe that the potential power of hybrid approach would put forth sooner or later.

3.2.7 Hidden Markov Models Method

Hidden Markov models (HMM) is another promising method that works well for images with variation in different lighting, facial expression, and orientation. HMM is a set of statistical models used to characterize properties of signals. It has very good performance in speech recognition and character recognition, where the data is 1-dimentional. The system being modelled is assumed to be a **Markov process** with unknown parameters, and the goal is to find hidden parameters from the observable parameters. Each state in HMM has a probability distribution over the possible output whereas each state in a regular Markov model is observable. In Nefian's paper [8], the authors use HMM approach for face recognition based on the extraction of 2- dimensional discrete cosine transformation (DCT) feature vectors. The author takes advantage of DCT compression property for feature extraction. An image is divided by blocks of a sub-image associated with observation vector. More details about HMM method are provided in the following sections.

In HHM, there are unobservable Markov chain with limited number of status in the model, the observation symbol probability matrix B, a state
transition probability matrix A, initial state distribution π, and set of probability density functions. A HMM is defined as the triplets λ = (A, B, π).

For frontal human face images, the important facial components appear in top to bottom order such as hair, forehead, eyes, nose, mouth, and chin. This still holds although the image rotates slightly in the image plane. Each of the facial region is assigned to one state in 1-D continuous HMM. The transition probability a_j^i and structure of face model is illustrated in below Figure.

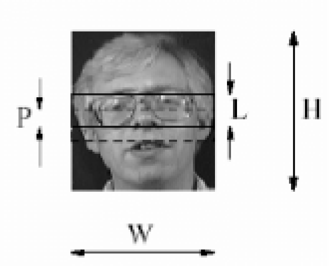


Source:
https://www.google.com/search?q=literature+review+of+face+detection&rlz=1C1UEAD_enIN996IN996&sxsrf=ALiCzsbzn1t_PreDTaHIDzvAMD0hGwsuaQ:1661235112730&source=lnms&tbm=isch&sa=X&ved=2ahUKEwjl_KCGp9z5AhWbm1YBHYJYD9QQ_AUoAnoECAEQBA&cshid=1661235113882800&biw=1280&bih=609&dpr=1.5#imgrc=9ozCaZJoFymwpM

Each face image with W and height H is divided into overlapping blocks of height L and the same width. The block extracting is shown in below Figure. The amount of the overlapping P has a significant effect on recognition rate since features are captured independent of vertical position. The magnitude of L is also important. Small length of L will assign insufficient information to discriminate to the observation vector. On the other hand, large value of L will increase the chances of cutting across the feature. Therefore, it is important to find good value for L. Once blocks are extracted from the image, a set of DCT coefficients are calculated for each block. When each block is transformed with

DCT, the most important coefficients with low frequencies are converged and clustered in small area in the DCT domain. The author from [9] uses 12x3 size window to pick these significant information of signal energy.

In this way, the size of observation vector is reduced significantly, which makes the system very efficient while still retaining good detecting rate. In the training phase, the image is segmented from top to bottom where each segment corresponds to a state, and initial observation probability matrix B is obtained from observation vectors associated with each state. Once B is obtained, the initial value of A and π are set given the left to right structure of the face.

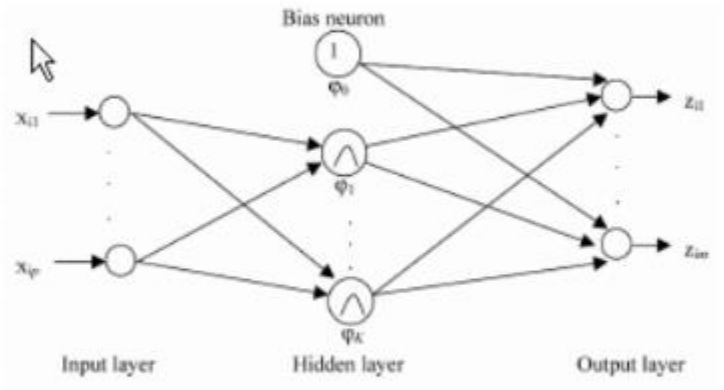


The face image is recognized if given Markov mode, the probability of observation symbols is maximum. For experiment in the paper, 400 images of 40 individuals with 10 face images per individual are used. The image database contains face images with different expressions, hair styles, eye wears and head orientations. The system achieves 84 % correct classification with L =10 and P =9 while eigenfaces approach achieves 73% of correct classification with the same dataset. Considering this fact, HMM approach has a bit better performance than eigenfaces method for images with variations.

3.2.8 Neural Network Method

Neural networks-based approaches are learned from the example-images and rely on the techniques from machine learning to find the relevant characteristics of face images. The learned characteristics, in the form of discriminant functions (i.e. non-linear decision surfaces), are subsequently used for face recognition. Conventionally, face images are projected to a low-dimensional feature space and nonlinear decision surface is formed using multilayer neural networks for classifications and recognition. Neural networks have also been used successfully for face recognition problem.

The advantage of using the neural networks for face recognition is that the networks can be trained to capture more knowledge about the variation of face patterns, and thereby achieving good generalization [13]. The main drawback of this technique is that the networks have to be extensively tuned to get exceptional performance. Among the neural networks approaches for face recognition, **multilayer perceptron** (MLP) with **back propagation (BP) algorithm** has been mostly used. However, the convergence of the MLP networks is slow and the global minima of the error space may not be always achieved [11]. On the other hand, the RBF neural networks have fast learning ability [15] and best approximation property [16]. So, in recent times, many researches have used RBF networks for face recognition and show in below figure.



However, their success rates are not so promising as the error rates vary from 5 to 9% under variation of pose, orientation, scale and light. This may be due to the fact that the selection of the centres of the hidden layer neurons might not have been done by capturing the knowledge about the distribution of training patterns and variations of face pose, orientation and lighting.

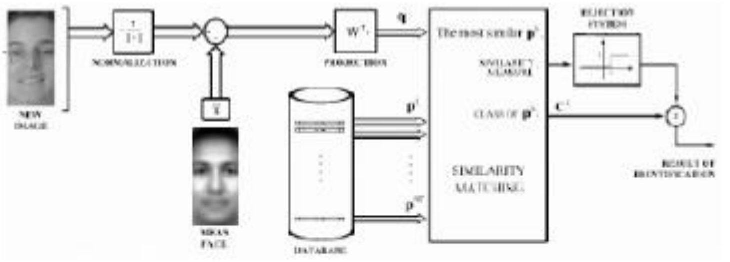
However, their success rates are not so promising as the error rates vary from 5 to 9% under variation of pose, orientation, scale and light [13]. This may be due to the fact that the selection of the centres of the hidden layer neurons might not have been done by capturing the knowledge about the distribution of training patterns and variations of face pose, orientation and lighting.

3.2.9 Eigen face Method

In 1991, Turk and Pentland used PCA projections as the feature vectors to solve the problem of face recognition, using the Euclidean distance as similarity function. This system, later called Eigenfaces, was the first eigenspace-based face recognition approach and, from then on, many eigenspace-based systems have been proposed using different projection methods and similarity functions. In particular, Belhumeur et al. proposed in 1997 the use of FLD as projection algorithm in the so-called Fisherfaces system [21]. In all standard eigenspace-based approaches a similarity function, which works as a nearest-neighbor classifier [22], is employed.

In 1997, Pentland and Moghaddam proposed a differential eigenspace-based approach that allows the application of statistical analysis in the recognition process [23]. The main idea is to work with differences between face images, rather than with single face images. In this way the recognition problem becomes a two-class problem, because the so-called "differential image" contains information of whether the two subtracted images belong to the same class or to different classes. In this case the number of training images per class increases so that statistical information becomes available, and a statistical classifier can be used for performing the recognition. The system proposed in [23] used Dual-PCA projections and a Bayesian classifier.

Eigenspace-based approaches approximate the face vectors (face images) by lower dimensional feature vectors. These approaches consider an off-line phase or training, where the projection matrix (W ∈ RN ×m), the one that achieve the dimensional reduction, is obtained using all the database face images. In the off-line phase, the mean face (x) and the reduced representation of each database image (pk) are also calculated. The recognition process works as follows. A preprocessing module transforms the face image into a unitary vector and then performs a subtraction of the mean face. The resulting vector is projected using the projection matrix that depends on the eigenspace method been used (PCA, FLD, etc.). This projection corresponds to a dimensional reduction of the input, starting with vectors in RN (where N is the dimension of the image vectors) and obtaining projected vectors q in Rm, with m<N (usually m<<N). Then, the similarity of q with each of the reduced vectors € pk (€ pk ∈ Rm) is computed using a certain criterion of similarity (Euclidean distance for example). The class of the most similar vector is the result of the recognition process, i.e. the identity of the face. In addition, a Rejection System for unknown faces is used if the similarity matching measure is not good enough and show in below figure.



3.2.10 Fisherface Method

Fisherface algorithm considers the ratio between the variation of one person and that of another person. It maximizes the determinant of between-class scatter matrix simultaneously, minimizing the determinant of within-class scatter matrix. For a C – class problem, the between -class scatter matrix is defined as follows:

$$S_b = \sum_{i=1}^{c} \Pr(\Omega_i)(\mu_i - \mu)(\mu_i - \mu)^T \qquad (1)$$

where $\Pr(\Omega_i)$is the prior class probability,μ_i is the mean sample of class Ω and μ is the mean sample of all classes. The within-class scatter matrix is defined as follows:

$$S_w = \sum_{i=1}^{c} \Pr(\Omega_i)\Sigma_i \qquad (2)$$

Where, $\Sigma_i = \frac{1}{N_i} \sum_{x_k \in \Omega_i} (x_k - \mu_i)(x_k - \mu_i)^T$

is covariance matrix of within-class sample is the number of samples in class . N_i and Ω_i Fisher criteria function is defined as follows:

$$J(W) = \frac{|W^T S_b W|}{|W^T S_w W|} \qquad (3)$$

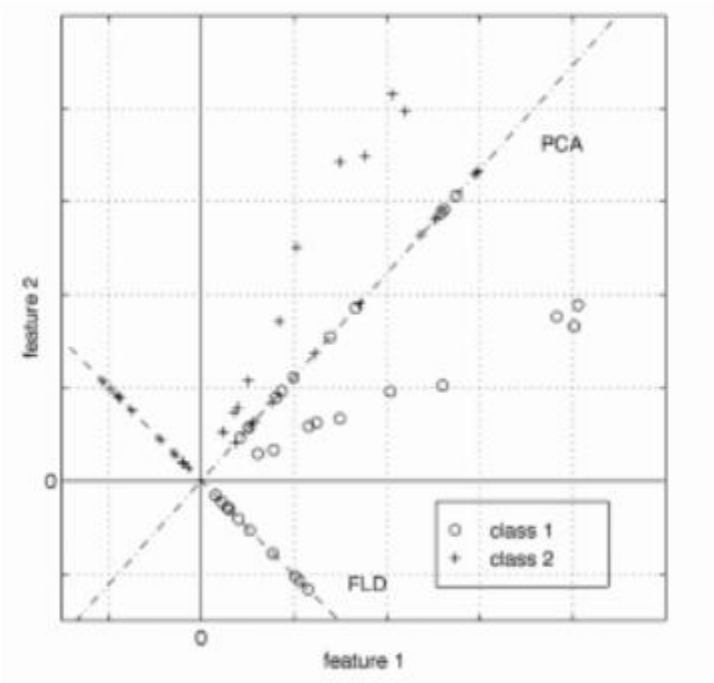
Then the projective matrix W_{fld} i s as follows:

$$W_{fld} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \qquad (4)$$

W_{fld} can be calculated by solving the generalized eigenvalue problem:

$$S_b W = S_w W \Lambda \qquad (5)$$

In face recognition applications, because rank of Sw is at most N- c, where N is the number of images in training set and typically much smaller than , number of pixels in each image, the within-class scatter matrix is always singular. To overcome this problem, PCA is first utilized to reduce the dimension of the images from N to N-c, then recalculated will be non-singular and FLD can be utilized to find the projective matrix S_w , is always singular. To overcome this problem, PCA is first utilized to reduce the dimension of the images from N to N-c, then recalculated will be non-singular and FLD can be utilized to find the projective matrix W_{fld} , which is referred to as Fisherfaces and Figure 7 is a comparison of PCA and FLD for a two-class problem in which the samples from each class are randomly perturbed in a direction perpendicular to a linear subspace.



3.3 Summary of Review

The approach for 3D face recognition involves sensitivity to size variation that can be use a purely curvature-based representation and handle size change between faces, but run into problems with change of facial expression between the enrollment image and the image to be recognized. In facial recognition system should be able to handle variation in expression. The seriousness of the problem of variation in facial expression between the enrollment image and the image to be recognized is illustrated in the results shown in Figure 2. This experiment focuses on the effects of expression change. Recognition was done with PCA-based 2D and 3D algorithm. The upper cumulative match characteristic (CMC) curves represent performance with time lapse only between gallery and probe. A main problem to experimental validation and comparison of 3D face recognition is lack of appropriate datasets. Desirable properties of such a dataset include:

(1) a large number and demographic variety of people represented,

(2) images of a given person taken at repeated intervals of time,

(3) images of a given person that represent substantial variation in facial expression,

(4) high spatial resolution, for example, depth resolution of 0.1 mm or better,

(5) low frequency of sensor-specific artifacts in the data.

0.4 Proposed Approach

Face detection is the first step of many advanced techniques and applications, such as face tracking, face recognition, and video surveillance. It is the method of determining all possible faces at various locations with diverse sizes in a given image. The objective of face detection is to find out whether there are any faces (blur too) in the given image and, if present, detect the face and extent of each face then determine the gender too.

Research work in the field of face detection has gone through lots of transitions. Researchers have tried different techniques to detect the human faces. Most of the researchers have used different standard image databases to check the performance of the algorithms. Here it is noticed that most researchers have used a combination of different classification approaches such as histograms, moments, neural networks, support vector machines etc. while developing a robust face detection system. Comparing these experimental results leads thinking to the fact that this is a relative comparison but no one can come out with an absolute decision that a specific technique is better than the others unless all the experiments have been carried out using the same dataset, using the same evaluation procedure, and under the same conditions such as running on the same machines. This yields to the conclusion that the face detection research needs a standard evaluation procedure including a standard benchmarking image database. In addition, when comparing different face detection techniques, judging using the detection rate is not an enough indication because it is relative to the false positive rate and number of samples present in the database. This is an important criterion to consider when choosing a technique to be used for real life application, because the selection is dependent on the cost or risk function of false alarms within the target application.

Although it appears to be an insignificant task for human beings, it is highly challenging for computers and has been one of the most studied research topics in the past few decades. The difficulty associated with face detection can be attributed to many variations in scale, location, orientation (in-plane rotation), pose (out-of-plane rotation), facial expression, lighting conditions as well as occlusions.

Most of the face detection methods focus on detecting frontal faces with good lighting conditions, but in this research the model with **YOLOv5** is able to detect the faces from any side profiles i.e. right profile or left profile. The feature extraction technique is that much effective so it can detect both the face and determine the gender from side profiles.

There are variety of face detection methods. Those are knowledge based, appearance based, feature invariant and template matching.

In this thesis **YOLOv5** is aimed to achieve better performance in terms of mean average precision (**mAP**) and speed. **YOLOv5 algorithm** is known for its speed and accuracy. Then, to improve face detection in blurred scenes or low-resolution situations, we integrated image super resolution technology on the detection head. In addition, some representative deep-learning algorithm based on face detection is discussed by grouping them into a few major categories, and the popular face detection benchmarks are enumerated in detail. Finally, the wider face dataset is used to train and test the SR-YOLOv5 model. Compared with multitask convolutional neural network (MTCNN), Contextual Multi-Scale Region-based CNN (CMS-RCNN), Finding Tiny Faces (HR), Single Shot Scale-invariant Face Detector (S3FD), and TinaFace algorithms, it is verified that the proposed model has higher detection precision, which is 0.7%, 0.6%, and 2.9% higher than the top one. SR-YOLOv5 can effectively use face information to accurately detect hard-to-detect face targets in complex scenes.

4.1 YOLO

YOLO is state of the art object detection algorithm and it is so fast that it has become almost a standard way of detecting objects in the field of computer vision. Previously **sliding window object detection** then more faster versions were invented, such as **RCNN**, **fast CNN** and **faster RCNN** but in 2015 **YOLO** was invented which outperformed all the previous object detection algorithms. The full form of **YOLO** is **YOU ONLY LOOK ONCE.**

In the last five years, the **YOLO algorithm** has been transformed into the fifth version with many innovative ideas from the object detection community. The first three versions including **YOLOv1 [7]**, **YOLOv2 [8]**, and **YOLOv3 [9]** were all proposed by the author of the original YOLO algorithm, and **YOLOv3 [9]** is recognized as a milestone with big improvements in performance and speed. We can find multiscale features **(FPN) [23]**, a better backbone network (Darknet53), and replacing the soft-max loss with the binary cross-entropy loss in this algorithm.

YOLOv4 [10] was released by a different research team in early 2020. The team explored a lot of options in almost all aspects of the **YOLOv3 [9]** algorithm, including the backbone, and what they call bags of freebies and bags of specials. One month later, the **YOLOv5 [11]** was released by another different research team which significantly reduced size, increased in speed, and had a full implementation in Python **(PyTorch)**. It is welcome by the object detection community until now.

If we're working on an image classification problem where we want to decide if the image of a animal or a person. In this case the output of neural network is pretty simple. Here animal is equal to one and person is equal to zero but when we talk about object localization, we're not only telling which class this, it also describes the bounding box or the position of an object within the image. So here in addition to dog is equal to one and the person is equal to zero.

In terms of neural network output there's a vector with a probability of a class. Since it is a supervised learning problem we need to give the bounding boxes. Neural network only understand numbers so it should be converted into vectors.

There can be 10,000 images and a neural network can be trained in a way that if input is a new image it will tell us if it's a dog or not and also tells the bounding box, basically essentially gives the answer for the object detection or object localization rather.

This works fine for a single object but if there are multiple objects in the image, if a person and dog in a same image so it's hard to determine the neural network output. If there's one object it's pretty fixed but in case of n no of objects then determining the size of the neural network out put is very hard.

If there are two objects in a image so two bounding boxes should be there, what **YOLO algorithm** will do is it will divide the image into small kind of grid cells it can be three by three, four by four. There's no fixed rule that it has to be exact four by four, and for each of the grid cells we can encode the vector which we discussed previously which is pc(probability of occurance) , bounding box c1 and c2. If there's no object in any cell so the probability of class is zero. And the rest of the values of vector don't matter but for particular grid cell, where the dog or person is that is expanding to multiple grid cells. In this case the aim is find the central place of that dog or person and that belongs to that particular grid cell. So now in that particular grid or cell the coordinates of the cell can be (0,0) and (1,1), now the pc(probability of occurance of person) is 0 for the person and pc(probability of occurance of dog) for dog is 1 if the bounding box is containing god not person and vice versa so the vector will reflect the occurance of particular object. If there is no object detected for a particular cell then both the pc's will be 0 in the vector. So for remaining all the cells the vector will be, pc is 0 and the remaining will be don't care. Now it is easy to form training data set the training data set will have so many images. Each image will have bounding rectangle and based on that rectangle we will try to derive. At first the kind of grid will be 3x3 grid or 4x4 grid and the target vector will be for each cell there will be one vector so there will be 16(for 4x4) such vector per training sample or per training image. By using this now we can train the neural network and after the training it can do predictions. So when we give an image as an input it can produces 16 such vectors and as it's 4 by 4 grid which is basically showing the bounding boxes for each objects so this is the **YOLO algorithm** it is called **YOU**

ONLY LOOK ONCE because it’s not repeating. It performs all the predictions in one forward pass that is why it’s called you only look once.

So this is a basic algorithm there can be few issues with this approach, the first issue is the algorithm might detect multiple bouding boxes for a given object. For this type of issue which rectangle has maximum probability rather than the others that bounding box will be the accurate one. If there’s another person in that image so for this issue neural network doesn’t know where it is so it cann’t take the max here we have to use a different approach. Here the concept of IOU will be used. **IOU** is **Intersection Over Union,** which is the intersection of the selected rectangles with highest probability. First we need to take a rectangle with a probability P and then for that same class which is person take the all other rectangles and then find overlapping area and to find overlapping area **IOU** is used. So in this case the overlapping area which is the intersection of rectangles. So the division of these two and if the objects are overlapping this value will be more. If the values are 0.6 and 0.7 it can be said that the rectangles are overlapping, if they are completely overlapping the value will be 1, if they’re not overlapping at all value will be 0. For each and every object the overlapping area will occur and achive the final bounding boxes this technique is also called **no max supression.** So after neural network has detected all the objects and we apply **no max supression** and will get unique bounding boxes.

There could be another issue, what if a single cell contains the center of two objects in this case both the centers of two objects are in the middle .
OLOv5 using CSPDarknet as a network of feature extraction, target information is extracted from the input image. The combination of CSP and Darknet formed the CSPDarknet. For the input tensor, CSP divides it into two parts in the channel,

one part is convoluted once, the other part is convolution-residuals multiple times. The tensor is obtained by multiple convolution-residual operations, and the tensor obtained by one convolution of the previous part is spliced in channel dimensions. CSP makes the output graph retain more network gradient information and maintains the performance of the network while reducing the computational effort.

In the operation, the features of the previous stage can be used as the input of the next stage for up-sampling or down-sampling, and at the same time, the CONCAT with the feature map of the same size in the main part. This pyramid structure makes the high-level feature map integrate the accurate position information of the low level and improves the accuracy of regression.

During detection, the input tensor is divided into **S x S** grids, and any one of the grids will be responsible for detecting the target if the center point of the target is located in it. For each grid, there will be B anchors. Specifically, for each anchor frame, **(5+C)** values are predicted, with the first 5 values used to regress anchor's center point position, the size of the anchor frame, then to determine whether there is a target.**C** is the total number of target categories. If the center of the target is in this grid, then the target will be acquired and judge whether it is a human face. The position of the regression box of the target can be obtained by the following formula:

$$C_i^j = p_{j,i} * IOU_{pred}^{truth} \qquad (1)$$

In the above parameters, i and j represent the jth regression box of the ith grid , C_i^j represents the confidence score of the jth bounding box of the ith grid. $P_{i,j}$ represents whether there is a target, if the target is in the jth box, the value of $P_{i,j}$ =1; otherwise, $P_{i,j}$ = 0. The IOU_{pred}^{truth} is a widely used parameter that represents the intersection over union between the predicted box and ground truth box. The higher the IOU score, the more accurate the position of the predicted box.

4.2 Method

This paper focuses on improving the detection accuracy of small faces in surveillance images. Because of the comparison of the four versions of YOLOv5 including **YOLOv5m**, YOLOv5l, YOLOv5x, and YOLOv5s, the YOLOv5s model is smaller and easier to deploy quickly. Therefore, our research is based on the YOLOv5s model. We optimize the backbone, then integrate image super resolution technology on the head and improve the loss function to ensure efficient detection speed.

4.2.1 SR-YOLOv5

4.2.1.1 Adaptive Anchor

The calculation of adaptive anchor is added in YOLOv5s. Before each training, the -means algorithm is used to cluster the ground truth of all samples in the training set and to find out the optimal group of anchor point frames in the high complexity and high recall rate. The results of anchor boxes clustered by the algorithm in a tabular format (image sizes will be there MxN).

4.2.1.2 Network Architecture

The first one is Backbone. The overall architecture of improved YOLOv5s is consists of the backbone, detection neck, and detection head. Firstly, a newly designed backbone named CSPNet is used. Secondly, a stem block is used to replace the focus layer in YOLOv5s. Thirdly, a C3 block is used to replace the original CSP block with two halves. One is passed through a CBS block, some bottleneck blocks, and a Conv layer, while another consists of a Conv layer. After the two paths with a CONCAT and a CBS block followed, we also change the SPP block [4] to improve the face detection performance. In this block, the size of the three kernels is modified to smaller kernels.

The second one is Detection Neck. The structure of the detection neck is consists of a normal feature pyramid network (FPN) and path aggregation network (PAN). However, we modify the details of some modules, such as the CS block and the CBS block we proposed.

And lastly the Detection Head. Through feature pyramid structure and path aggregation network, the front segment of the network realizes the full fusion of low-level features and high-level features to obtain rich feature maps, which can detect the most high-resolution face samples. However, for low-resolution images, feature fusion cannot enhance the original information of the image, and through layers of iteration, the prior information of small faces is still lacking. To enhance the detection rate of small faces in low-resolution images, SR is fused in the detection head part of the network. For the grid to be determined, the region information is input into SRGAN to carry out superresolution reconstruction and face detection again through its coordinate information.

Finally, the output of the two-stage face detector is integrated and output.All these are important part of the architecture of YOLOv5s.

4.2.1.3 Loss Function

IOU is a frequently used index in target detection. In most anchor-based [34] methods, it is used not only to judge the positive and negative sample but also to assess the distance between the location of the predicted box and the ground truth. The paper proposes that a regression positioning loss [35] should be considered: overlapping area, centre point distance, and aspect ratio, which have aroused wide concern. At present, more and more researchers propose better performance algorithms, such as IOU, GIOU, DIOU, and CIOU.

0.5 Experimentations and Results

Faces play a major role in identifying and recognizing people. The motivation for face detection is capturing people in surveillance video, human computer interfacing, face recognition, facial expression analysis. Face detection is a pattern classification problem. It calls for classifying the selected segmented region as face or non-face. For efficient classification a robust feature set as well as robust classification method is very much essential. This requires an efficient pre-processing, feature extraction followed by good classification method.

The goal of this research is to detect human faces, locate the exact position and detect the gender of all human faces in any arbitrary given colour image. The first stage of the proposed face detection system is gender based segmentation, which removes most of the background area and outputs candidate facial region to speed up the detection process. In this work, the main focus is specifically on the problem of detecting faces in images. After going through some processes the ultimate result will be obtained.

5.1 Dataset and Experimental Environment

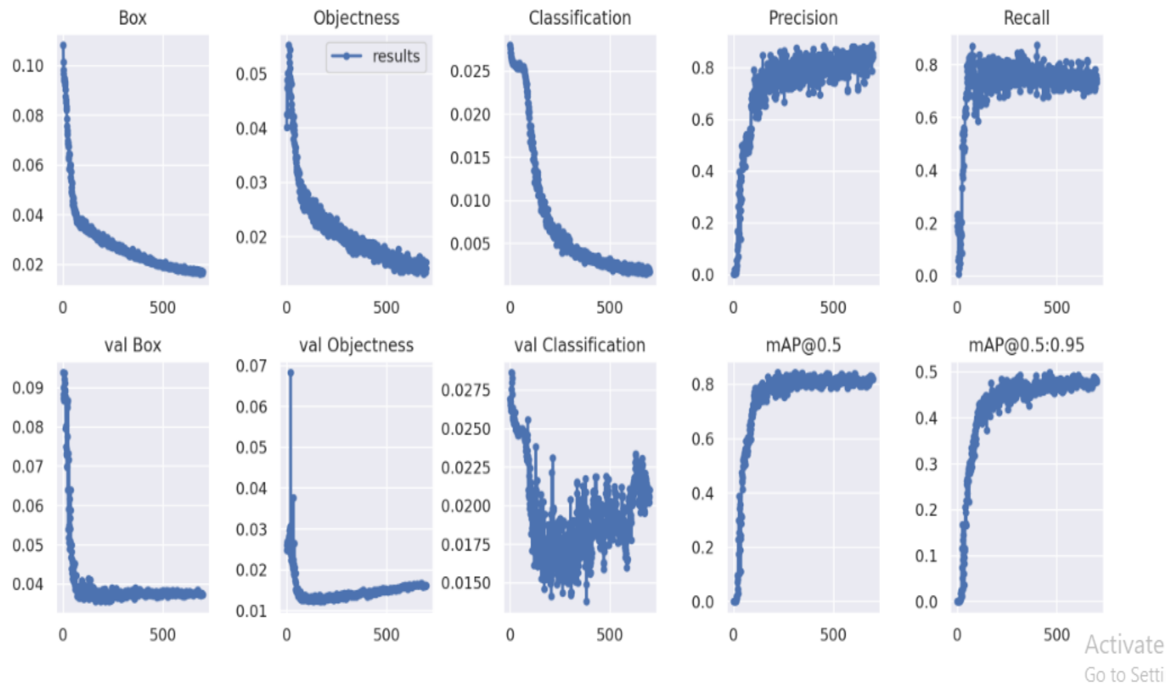
This experiment uses a face detection benchmark called wider face [27], which is recognized as the largest one among public available datasets. The details of publicly available datasets are shown in Table 2. These faces in the wider face dataset have great changes in scale, posture, and occlusion with an average of 12.2 faces per image, and there are many dense small faces. The dataset contains three parts: training set, validation set, and test set, accounting for 40%, 10%, and 50% of the sample number, respectively. This paper focuses on the detection of small faces, which will be more difficult to detect. Therefore, the verification set and test set are divided into three difficulty levels: easy, medium, and hard. There are many small-scale faces in the hard subset, most of which are 10 pixels~50 pixels. Thus, this benchmark is suitable to verify the effectiveness and performance in realistic scenes. The experimental environment configuration is shown in Table below.

Experimental environment	Configuration
Operating system	windows
GPU	Inter(R) HD Graphics 620
CPU	Intel(R) Core(TM) i5-7200U CPU @ 2.50GHz 2.70 GHz
Deep learning framework	PyTorch

5.2 Training and Testing of YOLOv5 Model

5.2.1 Training Model

The YOLOv5s code [11] is used as our basic framework, and we implement all the modifications as described above in PyTorch. We set the initial learning rate at 1E-2, and then we go down to 1E-5 with the decay rate of 5E-3. We set momentum at 0.8 in the first 20 epochs. After that, the momentum is 0.937. The precision-recall (PR) curves of our YOLOv5 detector are shown in Figure below.



5.2.2 Testing Model

The detection effect of our improved algorithm on the wider face dataset is shown in below Figure . It can be seen that this method has good robustness and high accuracy for small faces in various complex scenes. (a) The figure can detect faces with slight occlusion. (b) The figure itself has a low resolution, but the detection result shows that the detection effect is still good. (c) The figure fully shows that numerous small faces can be well detected even in a high-density crowd. (d) if there are multiple faces in a picture some of the faces can be blurred out. This model detects those blurred faces with a quite good accuracy and also determine their genders. Displayed an example of this.

Source:

https://www.google.com/search?q=group+images&rlz=1C1UEAD_enIN996IN996&sxsrf=ALiCzsbD4nMNb4iZvRk-FfE-OqCnJIGdwA:1661235449069&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiBvtGmqNz5AhUxmVYBHXGTBdQQ_AUoAXoECAEQAw&biw=1280&bih=609&dpr=1.5

0.6 Comparative Analysis

In the YOLOv5 network, the rationality and effectiveness of the fused network should be verified first. We select 1000 pictures from the test set for network model test and comparison. Compared with YOLOv3, the speed of the network after the fusion of superpartition reconstruction technology is reduced, because the network depth is increased when the new network is integrated. Compared with the HR using Resnet101 as the backbone network, the average detection accuracy of the improved network has been significantly improved, which is 2.3% higher than HR.

To demonstrate the effectiveness of the algorithm, some excellent face detection algorithms are selected to test on the wider face dataset, and the results are analysed. all existing methods achieve mAP in a range of 85.1-95.6% on the easy subset, 82.0-94.3% on the medium subset, and 62.9-85.3% on the hard subset. The mean average precision of the proposed algorithm on the easy, medium, and hard validation subsets are 96.3%, 94.9%, and 88.2%, respectively, which is 0.7%, 0.6%, and 2.9% higher than the top one.

Face detection algorithms	Easy	Medium	Hard
MTCNN	85.1%	82.0%	62.9%
CMS-RCNN	90.2%	87.4%	64.3%
HR	92.5%	91.0%	81.9%
S3FD	81.9%	92.5%	85.9%
TinaFace	95.6%	94.3%	85.3%
YOLOv5	96.3%	94.9%	88.2%

The SR-YOLOv5 proposed in this paper is improved on the YOLOv5s network, and the image super resolution reconstruction technology is introduced for the secondary detection of small-scale fuzzy faces, deepening the network to make facial features easier to be detected, capturing small target information, and making the network more accurate when processing complex face and nonface classification and detection. Through the comparative experiment on the wider face dataset, it is verified that the method used in this paper has higher detection accuracy and better robustness, especially in the hard subset, it has more outstanding performance.

0.7 Conclusions and Future Scopes

In reviewing last research, it is found that many approach can be used for face recognition that each method have different advantage and disadvantage such as local, global, and hybrid method. There are have 2 type of image for face recognition technique: still image and video image (still image sequence). However, we found some problems in face recognition system such as: (1) pose problem: due to cannot control face image for capturing and have many pose variations will be change every time. (2) illumination problem: due to source image have light condition or different lighting and viewing variations. (3) environment problem: due to in fact, motion and expression cannot controlling that is natural image. (4) 3D problem: due to in 3D image must be used more storage, many variable, low speed, and lack of face database testing. In future work, it is easy to propose a novel method for face recognition by hybrid approach combines 3D face and face expression. (eyes, nose, and mouth are location feature for extraction). Due to our survey found 3D recognition more accurate than 2D recognition, 3D capture creates larger data files per subject which applies significant storage requirements, slow processing, most new devices can be capture in 3D, and cannot control environment from the real world. Lastly, we want to widely accepted in 3D such as 3D Morpheble Model approach can be recognizing both frontal face and non-frontal face image. This is idea to use for improving overall recognition performance.

To improve the face detection rate of security surveillance scenes with diverse scales in dense face images, this paper proposes a small face detection algorithm suitable for complex scenes. We integrate the image superresolution reconstruction technology into the network structure of the target detection algorithm YOLOv5s. YOLOv5s has a fast detection speed, but its detection accuracy is reduced compared with other SOTA detection algorithms. SRGAN is used to improve the performance of the detection head and then improve the detection accuracy of small-scale fuzzy faces in complex scenes. In the same environment with other face detection algorithms, using the same dataset to carry out comparative experiments, the results confirm the feasibility and superiority of the proposed method. With these algorithms the updated model may can detect the skin tones of recognised faces and it will also can detect faces from video i.e. mp4 not only mp3(images) files. Upgradation of YOLO algorithm can make these changes in near future. The outcome will be more impressive. In this chapter, the proposed algorithm performs well under complex background. If the researchers want to improve the efficiency under complex background and blurred images they can use other algorithms than this. Feature extraction is performed using rectangular features. Thus, the backgrounds are eliminated and the face regions localized faster. Non-face and facial regions can be easily obtained from the sub-images. The feature values are fed into a trained BPN Network to classify the block as either face or non-face. For both the complex background and blurred images the Feature extraction algorithm should be upgrated to get better outcome.

And also the Feature extraction will be designed in a way so it can determine the baby faces, as it will be difficult to determine whether it’s a baby girl or baby boy. This task will be much more difficult and challenging too. But in near future it can be done.

0.8 References

1. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios Xingkui Zhu, Shuchang Lyu, Xu Wang, Qi Zhao, 26 Aug 2021.

2. Evolution of YOLO Algorithm and YOLOv5: The State-of-the-art Object Detection Algorithm by Do Thuan, Oulu University of Applied Sciences, Spring 2021

3. Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition
Liu, Yifan, Lu, BingHang, Peng, Jingyu, Zhang, Zihao.
World Scientific Research Journal ; 6(11):276-284, 2020.

4. Application of an Improved YOLOv5 Algorithm in Real-Time Detection of Foreign Objects by Ground Penetrating Radar by Zhi Qiu,Zuoxi Zhao,Shaoji Chen,Junyuan Zeng,Yuan Huang and Borui Xiang, Published: 14 April 2022.

5. CERTAIN INVESTIGATIONS ON FACE DETECTION AND FACE RECOGNITION TECHNIQUES A THESIS Submitted By ANILA S, DECEMBER 2012.

6. Multiple Face Detection and Localization of Skin-tone Group Images by H. C. Vijayalakshmi, September 2014.

7. https://www.hindawi.com/journals/cmmm/2021/7748350/

8.https://www.jigsawacademy.com/blogs/ai-ml/epoch-in-machine-learning#:~:text=An%20epoch%20in%20machine%20learning,learning%20process%20of%20the%20algorithm.

9. https://docs.ultralytics.com/quick-start/

10. https://shodhganga.inflibnet.ac.in/handle/10603/22130

11. Development of automatic gender classification algorithms for person specific identification, by Geetha A, 2020.

12. Genetic based face recognition by Deepa S, 28th july 2021 .

13. A comparative analysis of the automatic human skin detection techniques in color images and their applications to face detection by Ravi, S, 30th Jan-2017.

14. Human skin appearance based face detection from causal photographic scenes by Pasupathi P, 14th March 2019.

15. Design and development of image processing methodologies applied towards human age and gender estimation by Berlin jinu C K, 19th Aug 2019.

16. Development of Image Segmentation Methods for the Extraction of Human Faces by Hemlata, A., 28th Feb 2022.

17. Multi level region growing approach for efficient face tracking in video surveillance with invariant features using tri view binary patterns by Vivek Yoganand, A, 24-Sep-2021.

18. https://blog.roboflow.com/tag/case-studies/

19. Garbage Object-Detection to Identify Disposal Class by Mohamed Attia.

20. https://universe.roboflow.com/browse/object-detection

21. Lu J., Yuan X., and Yahagi T., A method of Face recognition based on Fuzzy c-Means clutering and associated sub-NNs. Proc. IEEE, 18(1), 2007, 150-159.

22. Effective Face Detector Based on YOLOv5 and Superresolution Reconstruction
Qingqing Xu, Zhiyu Zhu , Huilin Ge , Zheqing Zhang, and Xu Zang School of Electronic Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China Correspondence should be addressed to Zhiyu Zhu; zzydzz@163.com Received 3 September 2021; Revised 18 September 2021; Accepted 28 September 2021; Published 16 November 2021.

23. Library Attendance System using YOLOv5 Faces Recognition
October 2021 DOI:10.1109/ICCTEIE54047.2021.9650628
MardianaMeizano Ardhi Muhammad, Lampung University , **Yessi Mulyani,** Lampung University

24. Face Mask Detection using YOLOv5 for COVID-19 In affiliation, with California State University, San Marcos In partial fulfillment of the Requirements for the Degree of Master of Computer Science By Vinay Sharma November 24, 2020.

25. Accurate Real-Time Face Mask Detection Framework Using YOLOv5, **Published in:** 2022 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS), 06-09 June 2022, **DOI:** 10.1109/DTS55284.2022.9809855.

26. Real-Time Surveillance System based on Facial Recognition using YOLOv5, Published in: 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), Date of Conference: 15-17 July 2021, Date Added to IEEE *Xplore*: 06 September 2021, INSPEC Accession Number: 21138733, DOI: 10.1109/MAJICC53071.2021.9526254.

27. A Case Study: Cat-Dog Face Detector Based on YOLOv5, **Published in:** 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Date of Conference: 29-30 September 2021, Date Added to IEEE *Xplore*: 04 November 2021,INSPEC Accession Number: 21415150, DOI: 10.1109/3ICT53449.2021.9581987.

28. Automated Detection of Gender from Face Images, **Published in:** International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072