**Chapter 1**

Introduction

* 1. **Introduction**

Facial key points detection is a computer technology being used in a variety of applications that identifies human faces in digital images. Facial key points detection also refers to the psychological process by which humans locate and attend to faces in a visual scene. Face detection can be regarded as a specific case of object-class detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class.[1] Detect facial keypoints is a critical element in face recognition. However, there is difficulty to catch keypoints on the face due to complex influences from original images, and there is no guidance to suitable algorithms. In this paper, we study different algorithms that can be applied to locate keyponits. Specifically: our framework (1) Prepare the data for further investigation (2) Using PCA and LBP to process the data (3) Apply different algorithms to analysis data, including linear regression models, tree-based model, neural network and convolutional neural network, etc. Finally, we will give our conclusion and further research topic. A comprehensive set of experiments on dataset demonstrates the effectiveness of our framework.[2\*]

* 1. **Motivation**

Detecting facial key points is a very challenging problem.  Facial features vary greatly from one individual to another, and even for a single individual, there is a large amount of variation due to 3D pose, size, position, viewing angle, and illumination conditions. Computer vision research has come a long way in addressing these difficulties, but there remain many opportunities for improvement.[2] Recognizing faces is a very challenging problem in the field of image processing. The techniques presently being used are Biometric recognition (but Iris scanners are far too expensive), Eigenfaces(inaccurate with varying image factors like intensity, camera angles), line edge maps. While facial features in images depend on a lot many conditions, faces can be recognized easily if we use the relevant keypoints and landmarks for identification. The unchanging ratios and distances between these mark the importance of this approach. [4\*]

* 1. **Objective**

The objective of this task is to predict key points positions on face images. Detection of facial keypoints is building block for many applications in computer vision. Research has been done on this part but still there is hope for improvement. What really motivated me was that this problem works as the first steps for several applications, such as:

* Tracking faces in images and video
* Analyzing facial expressions
* Detecting dysmorphic facial signs for medical diagnosis
* Biometrics / face recognition
* Smart banking
* Forensic Investigation
* Security
  1. **Scope of the Work**

Importance of Face Recognition System as a Security Solution Face is considered as the most important part of human body. Research shows that even face can speak and it has different words for different emotions. It plays a very crucial role for interacting with people in the society. It conveys people’s identity, so it can be used as a key for security solutions in many organizations. Nowadays, face recognition system is getting increasing trend across the world for providing extremely safe and reliable security technology. It is gaining significant importance and attention by thousands of corporate and government organizations only because of its high level of security and reliability. Moreover, this system is providing vast benefits when compared to other biometric security solutions like palm print and finger print. As computation processing powers increases and large storage are available to store data, hence demand of it increases as it is used in several real-world applications.[3] Use of deep learning made more interested in the thesis. Convolutional Neural Network is chosen because it is more accurate and allow us to modify at various stages.

* 1. **Outline of the Thesis**

In “Introduction”, the 1stchapterof the book contains the information about project motivation, theobjective of the project and so on.

In “Literature Review”, the 2ndchapter of the book contains the information about Facial key points detection mechanism, Existing algorithms and its pros and cons, and a gentle overview of Machine learning, Deep learning and Convolutional Neural Network.

In “Existing system overview”, the 3rdchapter of the book contains the information about Existing system’s technical details and workflow, its pros and cons and features.

In “Proposed ML based facial key points detection system”, the 4thchapter of the book contains the information about Approach of proposed system, Diagram, Workflow and Merits.

In “Conclusion”, the 5thchapter of the book shows conclusion of this project.

**Chapter 2**

Literature Review

* 1. **Introduction**

Face detection is a computer technology that is being applied for many different applications that require the identification of human faces in digital images or video. It can be regarded as a specific case of object-class detection, where the task is to find the locations and sizes of all objects in an image that belong to a given class. The technology is able to detect frontal or near-frontal faces in a photo, regardless of orientation, lighting conditions or skin color.[1]

* 1. **How Facial Key Points can be Detected?**

Face detection applications use algorithms that decides whether an image is a positive image also called face image or negative image also called non-face image. This is called a classifier. To classify a new image correctly, it is trained on hundreds of thousands of face and non-face images. This feature answers the question “Where are the faces in this picture?”. For each face detected, you get a complete analysis of key points also called landmarks around the eyes, eye brows, jaw, nose and mouth.

* 1. **Existing Key Points Detection Algorithms**

OpenCV is a popular computer vision library started by Intel in 1999. The cross-platform library sets its focus on real-time image processing and includes patent-free implementations of the latest computer vision algorithms. OpenCV 2.3.1 now comes with a programming interface to C, C++, Python and Android. OpenCV 2.4 now comes with the very newFaceRecognizerclass for face recognition, so you can start experimenting with face recognition right away.[4] The currently available algorithms are:

* Eigenfaces
* Fisherfaces
* Local Binary Patterns Histograms

Among above mentioned available algorithms, Local Binary Patterns Histograms is the most widely used and most efficient algorithm.

* + 1. **Advantages**

Some advantages of OpenCV face detection methods are:

* Computationally simple and fast
* Shorter training time
* Low false positive rate
* Better performance in offline learning system
* Vision based system
* Lower configuration PC required
  + 1. **Disadvantages**

Some disadvantages of OpenCV face detection methods are:

* Not a knowledge-based system
* Cannot be used in online learning system
* Difficult to evolve
  1. **Machine Learning, Deep Learning and Convolutional Neural Network**

**2.4.1 Machine Learning**

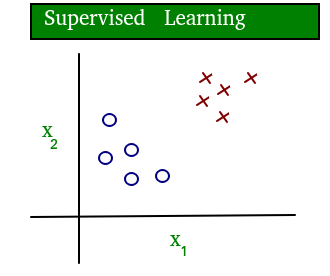
Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.[5] Machine learning is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

**2.4.1.1 Supervised Machine Learning**

Supervised learning: The computer is presented with example inputs and their desired outputs, given by a “teacher”, and the goal is to learn a general rule that maps inputs to outputs. The training process continues until the model achieves the desired level of accuracy on the training data. Some real-life examples are:

**Image Classification:** It is about train images with labels. Then in the future you give a new image expecting that the computer will recognize the new object.

**Market Prediction/Regression:** It is about train the computer with historical market data and ask the computer to predict the new price in the future.



**Fig 2.1: Supervised learning**

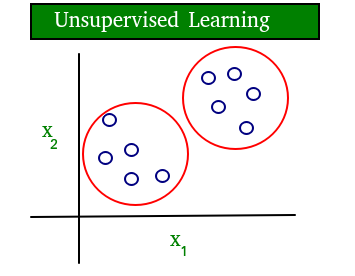
**2.4.1.2 Unsupervised Machine Learning**

Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. It is used for clustering population in different groups. Unsupervised learning can be a goal in itself (discovering hidden patterns in data).

**Clustering:** It is about to ask the computer to separate similar data into clusters, this is essential in research and science.

**High Dimension Visualization:** Use the computer to help us visualize high dimension data.

**Generative Models:** After a model captures the probability distribution of your input data, it will be able to generate more data. This can be very useful to make your classifier more robust.



**Fig 2.2: Unsupervised learning**

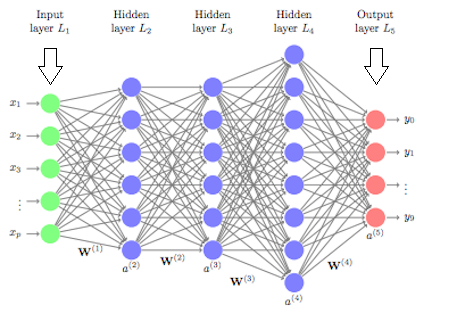
**2.4.2 Deep Learning**

Deep learning is a subset of machine learning in Artificial Intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as Deep Neural Learning or Deep Neural Network.[5]

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It’s achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.



**Fig 2.3: Layers of Deep Neural network**



**Fig 2.4: Performance of Deep Neural network**

**2.4.3 Convolutional Neural Network**

A convolutional neural network (CNN) is a type of artificial [neural network](https://searchenterpriseai.techtarget.com/definition/neural-network) used in [image recognition](https://searchenterpriseai.techtarget.com/definition/image-recognition) and processing that is specifically designed to process pixel data.

CNNs are powerful image processing, artificial intelligence ([AI](https://searchenterpriseai.techtarget.com/definition/image-recognition)) that use deep learning to perform both generative and descriptive tasks, often using computer vision that includes image and video recognition, along with recommender systems and natural language processing.[5]



**Fig 2.4: Working Principle of CNN**

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm a Convolutional Neural Network. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlaps to cover the entire visual area. In deep learning, a convolutional neural network is a class of deep neural networks, most commonly applied to analyzing visual imagery. CNNs are regularized versions of multilayer perceptrons.

* 1. **Summary**

In this chapter it is discussed where the system is being applied for. It is also shown that the system can be regarded as a specific case of object-class detection. Hundreds of thousands of face and non-face images data are fed to the system to classify new image correctly. OpenCV classifiers are existing solution to the system. Its pros and cons are also described in the chapter. Finally, a gentle overview to the Machine Learning, Deep Learning and Convolutional Neural Network is provided.

**Chapter 3**

Existing System Overview

* 1. **Introduction**

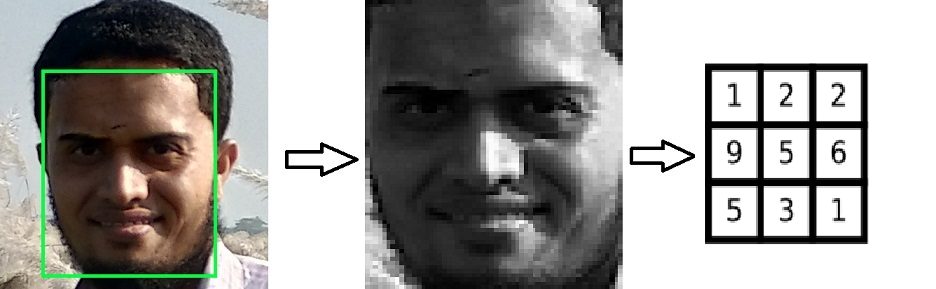
OpenCV is a popular computer vision library started by Intel in 1999. The cross-platform library sets its focus on real-time image processing and includes patent-free implementations of the latest computer vision algorithms. OpenCV 2.3.1 now comes with a programming interface to C, C++, Python and Android. OpenCV 2.4 now comes with the very newFaceRecognizerclass for face recognition, so you can start experimenting with face recognition right away.[4] The currently available algorithms are:

* Eigenfaces
* Fisherfaces
* Local Binary Patterns Histograms

Among above mentioned available algorithms, Local Binary Patterns Histograms is the most widely used and most efficient algorithm.

* 1. **OpenCV**

Local Binary Patterns methodology is most widely used over the world for face detection. It is famous for its high accuracy. Hence it is used in this thesis for its high accuracy. Local Binary Patterns methodology has its roots in 2D texture analysis. The basic idea of Local Binary Patterns is to summarize the local structure in an image by comparing each pixel with its neighbourhood. Take a pixel as centre and threshold its neighbours against. If the intensity of the centre pixel is greater-equal its neighbour, then denote it with 1 and 0 if not. You’ll end up with a binary number for each pixel, just like 11001111. So with 8 surrounding pixels, you’ll end up with 2^8 possible combinations, called *Local Binary Patterns* or sometimes referred to as *LBP codes*. The first LBP operator described in literature actually used a fixed 3 x 3 neighbourhood just like this:



**Fig 3.1: Image conversion**

 **Fig 3.2: LBP method**

A more formal description of the LBP operator can be given as:

LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c)

, with (x_c, y_c) ascentral pixel with intensity i_c; and i_n being the intensity of the neighbor pixel. s is the sign function defined as:

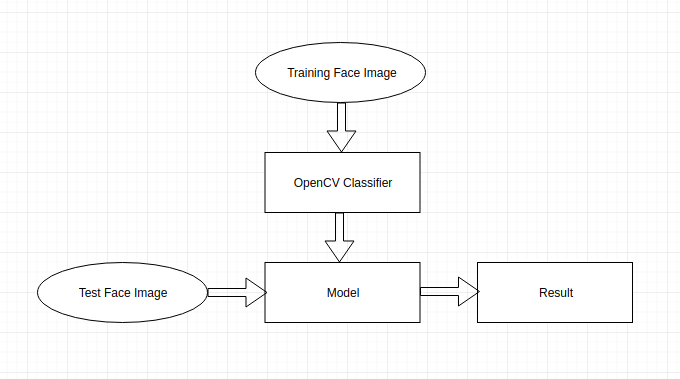
\begin{equation}
s(x) =
\begin{cases}
1 & \text{if $x \geq 0$}\\
0 & \text{else}
\end{cases}
\end{equation}

By definition, the LBP operator is robust against monotonic grey scale transformations. Dividing the LBP image into m local regions and extract a histogram from each. These histograms are called *Local Binary Patterns Histograms*.[4]

* 1. **Workflow**

Face Recognition process is about three steps:

* Prepare Training Data: Read training images for each person/subject along with their labels, detect faces from each image and assign each detected face an integer label of the person it belongs.
* Train Face Recognizer: Train OpenCV's LBPH recognizer by feeding it the data we prepared in step 1.
* Apply Test Data: Image data that will be tested against the model are applied to the model.
* Prediction: Introduce some test images to face recognizer and see if it predicts them correctly.[6]
* Result: Get the result after prediction and it is further processed in the next phase of the system.
* Analysis: After get the result of the system, it is analyzed with respect to some predefined threshold value.

****

**Fig 3.3: Workflow of existing system**

* 1. **Advantages and Disadvantages**
     1. **Advantages**

Some advantages of OpenCV face detection methods are:

* Computationally simple and fast
* Shorter training time
* Low false positive rate
* Better performance in offline learning system
  + 1. **Disadvantages**

Some disadvantages of OpenCV face detection methods are:

* Not a knowledge-based system
* Cannot be used in online learning system
* Difficult to evolve
  1. **Summary**

In this chapter it is discussed what OpenCV is. It is also shown that what methods of face detection are offered by OpenCV and how to use that methods. OpenCV's technical details, to be specific how Local Binary Pattern method works are discussed here. It also depicts workflow of OpenCV face detection and its pros and cons are also described in the chapter.

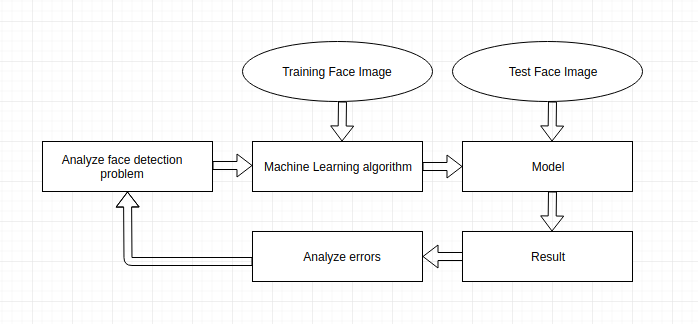
**Chapter 4**

Proposed Machine Learning Based Facial Key Points Detection System

* 1. **Introduction**

In this proposed system, the combined knowledge of computer vision techniques and deep learning architectures will be applied to build a facial key points detection system. Facial key points include points around the eyes, nose, and mouth on a face and are used in many applications. This system has several applications including facial tracking, facial pose recognition, facial filters, and emotion recognition. The final system should be able to look at any image, detect faces, and predict the locations of facial key points on each face.[7]

* 1. **Approach of the Proposed System**

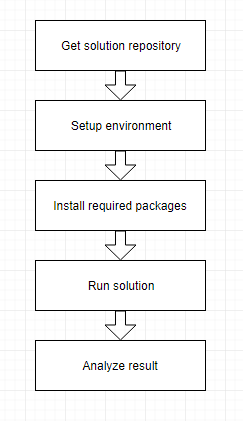
****

**Fig 4.1: Proposed ML based system**

In the proposed system, analysis of the face detection problem is performed at first. It includes data cleaning, data visualization and features extraction. If this phase is completed successfully then the next phase of the system is performed which is applying Machine Learning algorithm. An appropriate Machine Learning algorithm is applied on training data such as Logistic regression, Deep Learning, Convolutional Neural Network etc. After applying the Machine Learning algorithm, a model can be obtained and test data are applied on that model. Result are obtained after scientific computation of the particular algorithms. And then error analysis of the model is performed. If error rate is greater than a threshold value then changes to model are required else the model is deployed. Otherwise, the problem is analyzed again in similar fashion. This process is repetitive.[8]

* 1. **Workflow of the Proposed System**

**4.3.1 User Perspective**

****

**Fig 4.2: User perspective workflow**

User needs to get the solution repository either from git or from other sources. Then he should install Python 3.5, Tensorflow latest version, Keras Latest version and install other packages through pip package manager. These packages are required to perform scientific computation, visualization etc. Then it is required to run the solution along with data. Finally, it needs to collect result and analyze it.

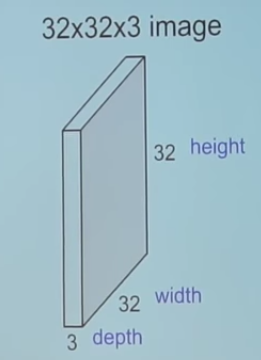
* 1. **How Convolutional Neural Network works?**

**4.4.1 What Convolutional Neural Network is?**

Convolutional Neural Networks (CNNs), like neural networks, are made up of neurons with learnable weights and biases. Each neuron receives several inputs, takes a weighted sum over them, pass it through an activation function and responds with an output. The whole network has a loss function and all the tips and tricks that we developed for neural networks still apply on CNNs.

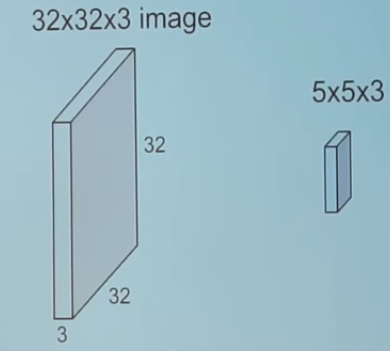
**4.4.2 CNNs operate over Volumes**

Unlike neural networks, where the input is a vector, here the input is a multi-channeled image (3 channeled in this case). There are other differences that we will talk about in a while. Before we go any deeper, we have to understand what convolution means.



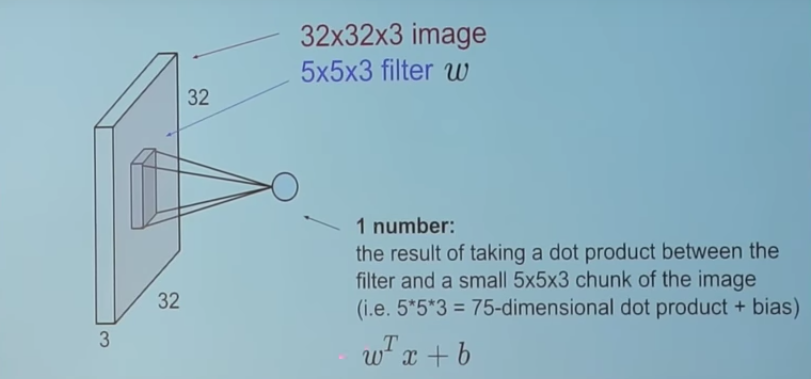
**Fig 4.3: Example of a RGB image**

**4.4.2 Convolution**

****

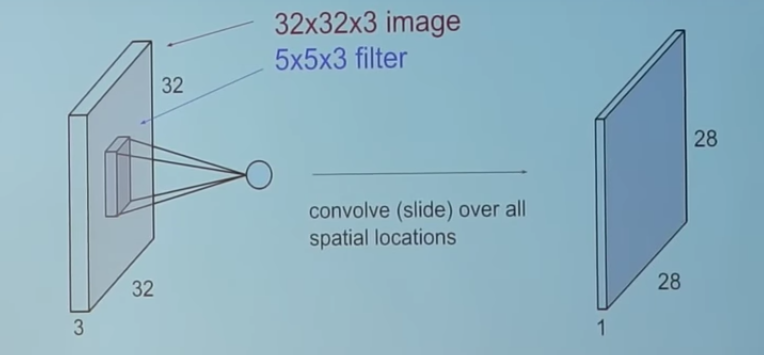
**Fig 4.4: Convolving an image with a filter**

We take the 5\*5\*3 filter and slide it over the complete image and along the way take the dot product between the filter and chunks of the input image.



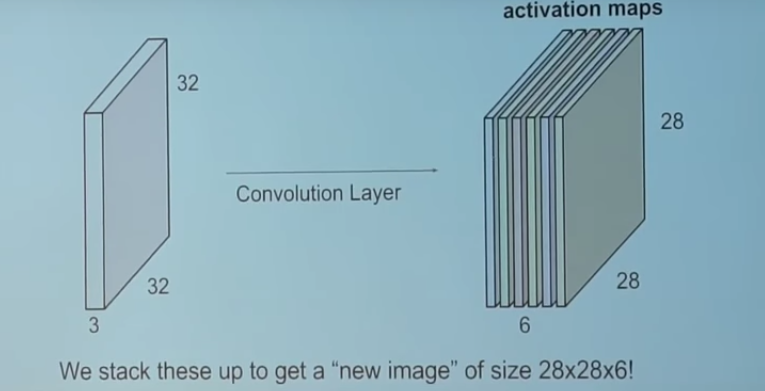
**Fig 4.5: How convolution looks**

For every dot product taken, the result is a scalar. when we convolve the complete image with the filter we will get 28\*28 unique positions where the filter can be put on the image.



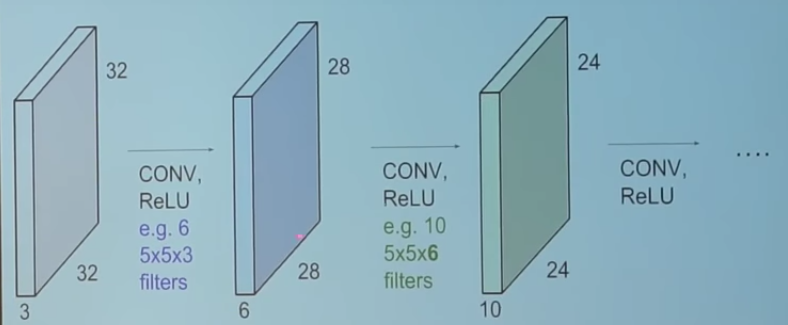
**Fig 4.6: Convolved image**

Now, if we back to CNNs the convolution layer is the main building block of a convolutional neural network.



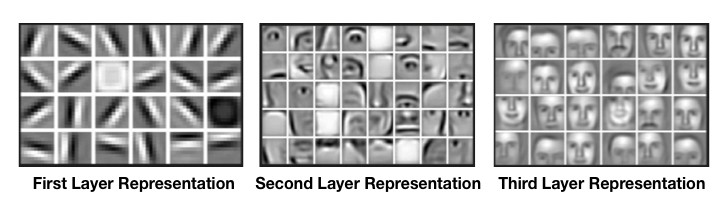
**Fig 4.7: Convolution Layer**

The convolution layer comprises of a set of independent filters. In this example there are 6 independent filters. Each filter is independently convolved with the image and we end up with 6 feature maps of shape 28\*28\*1. Suppose we have a number of convolution layers in sequence. Then we will get convolution layers in sequence like below.



**Fig 4.8: Convolution Layers in sequence**

All these filters are initialized randomly and become our parameters which will be learned by the network subsequently. An example of a trained network of face images is look like below.



**Fig 4.9: Filters in a trained network**

Take a look at the filters in the very first layer. These are our 5\*5\*3 filters. Through back propagation, they have tuned themselves to become blobs of colored pieces and edges. As we go deeper to other convolution layers, the filters are doing dot products to the input of the previous convolution layers. So, they are taking the smaller colored pieces or edges and making larger pieces out of them.

Take a look and imagine the 28\*28\*1 grid as a grid of 28\*28 neurons. For a particular feature map the output received on convolving the image with a particular filter is called a feature map, each neuron is connected only to a small chunk of the input image and all the neurons have the same connection weights. So again coming back to the differences between CNN and a neural network.

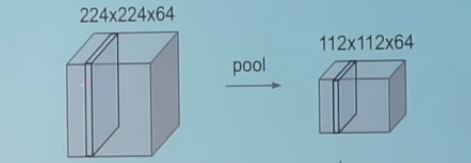
**4.4.3 CNNs parameter and connectivity**

Parameter sharing is sharing of weights by all neurons in a particular feature map. Local connectivity is the concept of each neural connected only to a subset of the input image unlike a neural network where all the neurons are fully connected. This helps to reduce the number of parameters in the whole system and makes the computation more efficient.

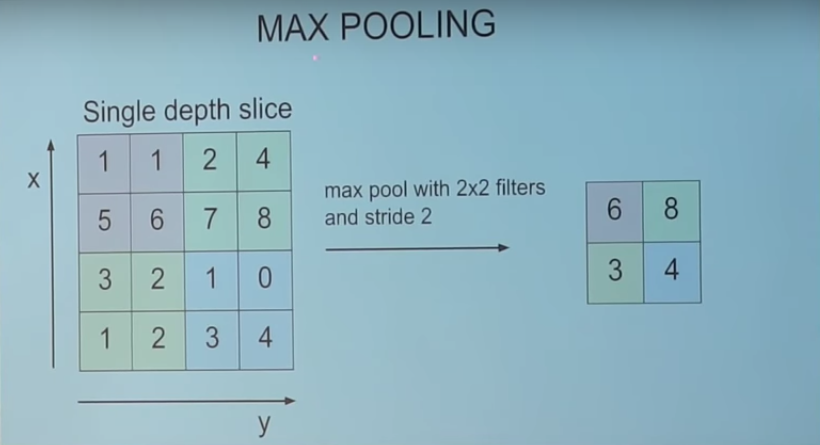
**4.4.4 Pooling Layers**

A pooling layer is another building block of a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the number of parameters and computation in the network. Pooling layer operates on each feature map independently.

The most common approach used in pooling is max pooling. Where the maximum value of a chunk is selected for the next pooling layers. Further all the selected maximum values are together to make the next stage of pooling layer.



**Fig 4.10: Pooling**

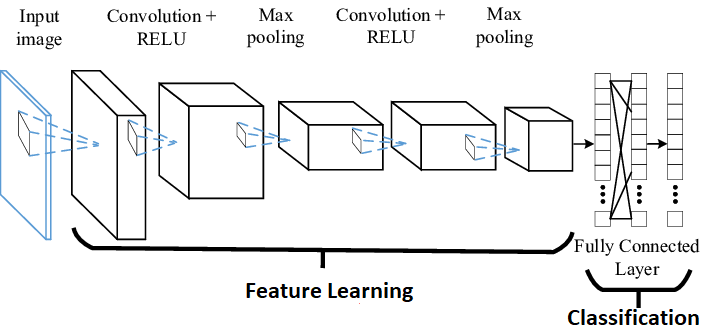
****

**Fig 4.11: Max Pooling**

**4.4.5 Typical architecture of CNN**

We have already discussed about convolution layers denoted by CONV and pooling layers denoted by POOL. Rectified Linear Unit also denoted as RELU is just a non-linearity which is applied similar to neural networks.

The FC is the fully connected layer of neurons at the end of CNN. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks and work in a similar way.



**Fig 4.12: Typical architecture CNN**

* 1. **Summary**

In this chapter proposed system is discussed in details. A diagram is also provided that depicts how the proposed system will work and its description provided as well. A user view workflow, features and merits of the proposed system is also given. Finally, required tools for the proposed system is also mentioned.

**Chapter 5**

Implementation and Result

* 1. **Required Tools**

**5.1.1 Hardware Tools**

Required hardware tools for the proposed system are:

* CPU: Core i7 at least 3.0 GHz and quad core
* RAM: 16 GB
* GPU: Nvidia graphics at least 3 GB
* Auxiliary Storage: 512 GB SSD with 1 TB HDD

**5.1.2 Software Tools**

Required software tools for the proposed system are:

* Python 3.5
* IDE
* Numpy
* Matplotlib
* Pandas
* Seaborn
* Tensorflow
* Keras
  1. **Features**

Some features of the proposed system are:

* This system is a knowledge-based system
* It can be used in online learning system
* It is easy to evolve the model
* It is possible to automate the learning system
  1. **Merits of the System**

Some merits of the proposed system are:

* Detect facial key points with high accuracy
* Easy to use in time series problem
* Low false positive rate
* Improved computing performance with short training time

**Chapter 5**

Conclusion

* 1. **Conclusion**

We conclude that CNN, which is indeed a faster and more accurate method of deep learning, when applied facial keypoints detection gives result far better than simple networks. Also that the approach is not affected by imaging factors. Detecting facial keypoints is a very challenging problem. Facial features vary greatly from one individual to another due to position, viewing angle, and illumination conditions. Computer vision research has come a long way in addressing these difficulties, but there remain many opportunities for improvement. This proposed system will provide high accuracy performance to detect facial key points using latest trending technologies. Adjusting filter sizes, keep larger filters for data input layers and decreasing the size in subsequent layers, produces better result. With further tuning of networks and large data size, the network is bound to improve.

**5.2 Future Work**

Future work of this thesis is to build more accurate machine learning model so that it will be able to determine facial key points better. Suitable, advanced and new machine learning methods will be applied to enhance the system. The portable binary version of the mini project will be made and package will also be available so that other developers can include it easily. Also there is a lot that can decrease the complexity of processing such as: Simply try changing nodes in the hidden layer of the new improved structure to reach an optimal value, With better computing resources trying to train data with more epochs then the root square mean equation will be better, Adding some layers of preprocessing just before of two ending layers can help reduce the complexity of the problem together with gaining accuracy.

**REFERENCES**

[1] Akshay Chandra, 21 July 2018, Towards Data Science magazine

[2\*] Shenghao Shi, 15 Oct 2017, Machine Learning, Cornell University

[4\*] Ajay Sharma, 01 April 2015, Face detection: Artificial Intelligence, IIT Kanpur

[2] https://www.kaggle.com/c/facial-keypoints-detection

[3] Jyoti P Dandale, 01 April 2010, International Journal of Computer Science and Information Security

[4] https://docs.opencv.org/2.4.13.7/modules/contrib/doc/facerec/facerec\_tutorial.html

[5] https://en.wikipedia.org/wiki/Machine\_learning

[6] Lopez & Ruiz, 08 July 2017, Super Data Science magazine

[7] Siraj Raval, 12 July 2017, The School of AI

[8] Aurelien Geron, 13 March 2017, Hands-On Machine Learning with Scikit-Learn and TensorFlow

[9]