# INTRODUCTION

The face plays an important role in the process of Visual Communication. By looking at the face, humans can automatically extract many nonverbal messages, such as human identity, intent, and emotions. In Computer Vision, to automatically extract this kind of information facial keypoints play a crucial role. The locations of these facial keypoints or facial landmarks helps to capture the deformations in the faces due to head movements and facial expressions. Accurate detection of these keypoints is usually very crucial for many facial analysis methods. Some of the applications and analysis include :

- Facial Expression Recognition [1]
- Head Pose Estimation [2]
- Face Recognition and Identification [3]
- Security Surveillance (e.g. Apple Face ID unlock feature)
- Human Computer Interface [4]
- Human Centered Autonomous Vehicle Systems [5]

Facial keypoints detection algorithms aim to automatically identify the locations of the key facial landmarks or keypoints on the facial images or videos. These predicted keypoints are either the dominant keypoints describing or pointing to a unique location of a facial component (e.g. the pupil in the eye, nose, etc) or it could be an interpolation of points connecting those facial keypoints around the facial components.

The problem of facial keypoints detection, though very interesting and has many applications, is a hard problem. This is mainly because of these broad reasons :

1. Appearance of the face changes significantly in different expressions and head pose.
2. Environmental Illumination conditions
3. Facial Occlusions
4. Faces of any two people is very different at the pixel level, though they might look similar in appearance.

# LITERATURE SURVEY

Facial Keypoints detection and prediction is very crucial for many facial analysis methods and applications. This was understood by many scientists in the field of Computer Vision and the work on the detection algorithms dates back to several decades. Since then there has been a lot of developments in the algorithms. In the early days of the field the algorithms focused mostly on easy situations and facial images without a lot of variations in facial expressions and very less occlusion and straightforward head pose. Later on, as the field started to progress and development of more efficient and fast algorithms along with the availability of more compute lead to the work being focused on more harder situations like different facial expressions and different head poses. Nowadays, the focus is mostly in the so called “in the wild” situations and conditions where the face can be in any kind of illumination, head pose, expression, occlusion etc. Even after all this progress there is still no robust algorithm or method which can solve the problem end-to-end without error in all situations and variations.

All the algorithms in this field of facial keypoints can be categorised into the following four categories :

- Holistic Methods
- Constrained Local Method
- Regression Methods

## Holistic Methods

This is one of the earliest methods developed for facial keypoints detection. It is a purely Computer Vision based approach where the algorithm leverages the information from the global facial shape patterns for facial keypoints detection. One of the famous algorithms in this type of methods is the Active Appearance Model (AAM) method developed by Taylor and Cootes [6]. This is a statistical model that controls facial shape variations and appearance using a few coefficients. These coefficients are generalized using a minimalistic dataset. Thus, the algorithm now has a global general representation of the fitted model. This method is not very robust because we use predefined mathematical models for fitting the model like PCA(Principal Component Analysis) for dimensionality reduction and Affine transformations. Therefore, this cannot work in the wild. To make these algorithms more robust advancements have been made to develop models that are adaptive in nature but these are not very robust.

## Constrained Local Methods (CLM)

The Constrained Local Methods [7] are similar to the Holistic Method in the sense that they try to fit the general model based on the global facial patterns of the data it is fitted on. Along with this, CLM also considers the independent information of the local facial components. As this method considers both the global patterns and the local feature information it is more robust to illumination and occlusion. It is therefore a more general algorithm. This is a relatively slow process as there are a lot of computations to perform.

## Regression Methods

The Regression Methods are the most robust and general out the three categories. This is because these methods directly learn the mapping from the image to the facial keypoints. As the mapping is left for the algorithm to discover it can learn the features which it finds important to reduce a prediction cost unlike the previous methods where there was constraint on the features to fit. This category includes the methods like Support Vector Machines (SVM's) [8], deep-learning methods [9]. Therefore, these methods are more robust and better performing and also fast.

# PROBLEM DEFINITION

The problem is to predict the facial keypoints on an image accurately (i.e. as close to where it should originally be if hand annotated by humans). There are a total of 15 keypoints to be predicted. This includes 3 points for each of the eyes, 2 points for each of the eyebrows, 1 point for the nose and 4 points for the mouth. This can later be implemented on a video feed as a video is merely a collection of images or frames in sequence.

# PROPOSED SOLUTION

The solution we propose to the above stated problem is the use of a Deep Learning approach. We will mainly be using the Convolutional Neural Networks (CNN's) [10] as these networks are robust in working with images. There have been many solutions using CNN's but we propose an ensemble architecture. To make the proposed model more robust we also propose the use of Data Augmentation.

## REFERENCES


1. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence
2. Murphy-Chutorian, E., Trivedi, M.: Head pose estima-tion in computer vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence
3. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deep-face: Closing the gap to human-level performance in face verification
4. Jilin Tu, T. Huang, Hai Tao : Face as mouse through visual face tracking. Canadian Conference on Computer and Robot Vision
5. Lex Fridman : Human Centered Autonomous Vehicle Systems: Principles of Effective Shared Autonomy
6. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence
7. Cristinacce, D., Cootes, T.F.: Feature detection and tracking with constrained local models. In: British Machine Vision Conference
8. C Cortes, V Vapnik : Support Vector Machines (1995) : Machine Learning
9. Sun, Yi, Xiaogang Wang, and Xiaoou Tan : Deep convolutional network cascade for facial point detection." Proceedings of the IEEE Con-ference on Computer Vision and Pattern Recognition.
10. Yann LeCunn, Leon Buttou, Yoshua Bengio, Patrick Haffner. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE (1998)
11.