# INTRODUCTION

Object detection is a computer technology in the field of computer vision which is used to identify objects in a digital image. Face detection is a type of object detection which is used to identify the location and size of the frontal human faces within an image.

Some of the leading applications of face detection comprise of

* **Facial motion capture** : converts the movement of a person’s face into a digital database which can be used for animations, movies, games.
* **Facial recognition** : used to identify or verify a person from digital image or a video frame.
* **Photography** : digital cameras in today’s world uses autofocus to detect faces. Additionally, smile detection can be used for taking photo at an appropriate time.
* **Emotional inference**: It is used to help people with autism to understand the feelings of people around them.

The libraries used in this project are:

* **OpenCV** : OpenCV is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.
* **DLIB** : Dlib is a modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems such as face detection, face recognition, etc.

# DATASET

WIDER FACE dataset is a face detection benchmark dataset, of which images are selected from the publicly available WIDERFACE dataset. We choose 32,203 images and label 393,703 faces with a high degree of variability in scale, pose and occlusion as depicted in the sample images. WIDER FACE dataset is organized based on 61 classes. Each class is describing an event. For each event class, we randomly select 40%/10%/50% data as training, validation and testing sets. A face detector is trained using WIDER FACE training/validation partitions and tested on WIDER FACE test partition.

* WIDER Face Training Images (12880 images)
* WIDER Face Validation Images (3226 images)
* WIDER Face Testing Images (16097 images)
* Face Annotations

The dataset is not limited to the fixed size images. It consists of images with different size and scale.

The annotation contains the ground truth values consists of Number of bounding boxes, x-coordinate, y-coordinate, width, height, blur, expression, illumination, invalid, occlusion and pose.

* Blur is categorized into 3 category clear, normal blur and heavy blur.
* Expressions are categorized into typical expression and exaggerate expression.
* For Occlusion, a face is defined as ‘no occlusion’ for 0% as ‘partially occluded’ for 1%-30% and as ‘heavily occluded’ if over 30% of the total face area is occluded.
* Pose is defined as two pose deformation levels, namely typical and atypical. we assign a face pose as atypical if either the roll or pitch degree is larger than 30-degree; or the yaw is larger than 90-degree. Otherwise a face pose is classified as typical.
* Illumination also categorized into normal illumination and extreme illumination.

# MODELS

**Haar Cascade Classifiers using OpenCV**

**Histogram of Oriented Gradients using Dlib**

HOG works like a sliding window. A block is considered as a pixel grid in which gradients are constituted from the magnitude and direction of change in the intensities of the pixel within the block.

The idea behind HOG is to extract features into a vector and feed it into a classification algorithm like a Support Vector Machine for example that will assess whether a face (or any object you train it to recognize actually) is present in a region or not.

In the HOG feature descriptor, the distribution (histograms) of directions of gradients (oriented gradients) are used as features. Gradients (x and y derivatives) are typically large around edges and corners and allow us to detect those regions.

The process was implemented for human body detection, and the detection chain was the following:

* Preprocessing: The input images must be of the same size (crop and rescale images).
* Calculate the Gradient Images: Compute the horizontal and vertical gradients of the image, by applying kernels. The gradient of an image typically removes non-essential information.
* Calculate Histogram of Gradients: The image is then divided into 8x8 cells to offer a compact representation and make our HOG more robust to noise. Then, we compute a HOG for each of those cells.
* Block Normalization: Normalize the image and make it invariant to lighting by constructing 16x16 block. This is simply achieved by dividing each value of the HOG of size 8x8 by the L2-norm of the HOG of the 16x16 block.
* Calculate the HOG feature vector: To calculate the final feature vector for the entire image patch, vectors are concatenated into one giant vector.

**Convolutional Neural Networks using Dlib**

One of the most popular models used in computer vision is CNN which are a feed-forward neural network. CNN’s offer an automated image pre-treatment as well as dense neural network part. CNN are majorly used for processing data with grid-like topology.

With the rise of deep learning and greater computation capabilities, this work of extracting features in order to extract as much information from the image as possible can now be automated. The model makes use of the pre-trained model defined in mmod\_human\_face\_detector.dat. This pre-trained model was created based on a dataset containing face images from ImageNet, AFLW, Pascal VOC, the VGG dataset, WIDER, and face scrub.

# RESULTS

# CONCLUSION

# BIBLIOGRAPHY