Improved Facial Expression Recognition Based on DWT Feature for Deep CNN

Ankush Dey MDS202108
Ritirupa Dey MDS202136
Anjali Pugalia MDS202107

Chennai Mathematical Institute

April 22, 2023

Outline

Abstract

Facial expression recognition (FER) has become one of the most important fields of research in pattern recognition. In this paper, we propose a method for the identification of facial expressions of people through their emotions. It has many different applications in various fields such as security-surveillance, artificial intelligence, military and police services, and psychology, among others. Facial expressions are classified into six basic categories; namely, anger, disgust, fear, sadness, happiness, and surprise —a neutral expression was also added to this group.
Experiment was performed on the CK+ database and JAFFE face database.

Abstract

This paper combines 4 steps to create an efficient emotion detection algorithm. Which are :

1. Viola Jones Algorithm to locate face and facial features
2. Using Contrast limited Adaptive Histogram Equalization (CLAHE) for facial image enhancement.
3. Discrete Wavelet Transformation (DWT) for extracting facial features.
4. Deep CNN which directly uses the extracted features for training.

Viola-Jones Algorithm

About Viola Jones Algorithm:

1. Brings together new algorithms and insights to construct a framework for robust and extremely rapid visual detection.

Viola-Jones Algorithm

About Viola Jones Algorithm:

1. Brings together new algorithms and insights to construct a framework for robust and extremely rapid visual detection.
2. Constructs a frontal face detection system working only with information present in a single greyscale image.

Viola-Jones Algorithm

There are three key contributions.

1. Introduction of a new image representation called the "Integral Image".

Viola-Jones Algorithm

There are three key contributions.

1. Introduction of a new image representation called the "Integral Image".
2. A simple and efficient classifier which is built using the AdaBoost learning.

Viola-Jones Algorithm

There are three key contributions.

1. Introduction of a new image representation called the "Integral Image".
2. A simple and efficient classifier which is built using the AdaBoost learning.
3. A method for combining classifiers in a "cascade" which allows back- ground regions of the image to be quickly discarded

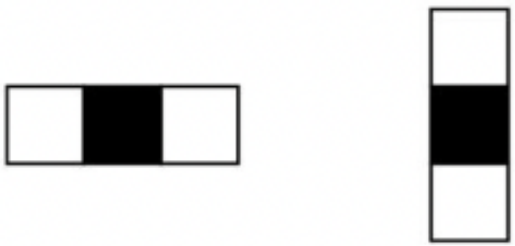
Viola-Jones ALgorithm

Instead of using pixels this paper uses features. It obtains them by applying these Haar like filters to sub-windows of the images. The paper uses three kind of features. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions.

Type 1 Type 2

A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle.

Type 3 Type 4



Viola-Jones Algorithm

Finally a four-rectangle feature computes the difference between diagonal pairs of rectangles.

Type 5

Viola-Jones Algorithm

This is the 1st contribution of the viola-Jones Algorithm.

Integral Image

It allows for very fast feature evaluation.Integral image, also known as a summed area table, is an algorithm for quickly and efficiently computing the sum of values in a rectangle subset of a pixel grid.

viola-Jones Algorithm

The integral image at location x, y contains the sum of the pixels above and to the left of x, y inclusive :

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

where $i(x, y)$ is the pixel value of the original image and $ii(x', y')$ is the corresponding image integral value.

1	1	1
1	1	1
1	1	1

Input image

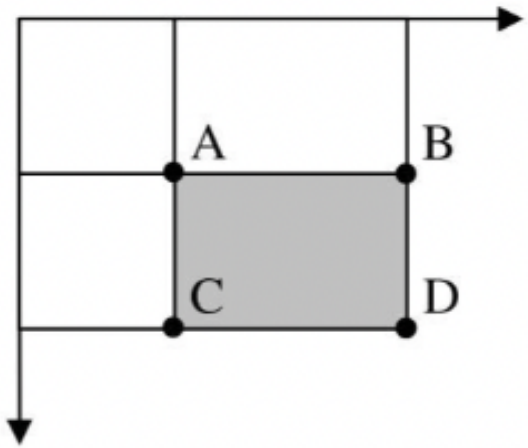
1	2	3
2	4	6
3	6	9

Integral image

Viola-Jones Algorithm

Using the integral image to compute the sum of any rectangular area is extremely efficient. The sum of pixels in rectangle ABCD can be calculated with only four values from integral image:

$$\sum_{(x,y)\in ABCD} i(x,y) = ii(A) + ii(D) - ii(B) - ii(C)$$



Sum of grey rectangle = D - (B + C) + A

Viola-Jones Algorithm

AdaBoost is an aggressive mechanism for selecting a small set of good classification functions which nevertheless have significant variety.
A single weak classifier is defined as:

$$h(f, x, p, \theta) = \begin{cases} 1 & pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases}$$

Where f is the feature θ is the threshold p is the polarity indicating the direction of the inequality and x is a 24×24 pixel sub window of the image.

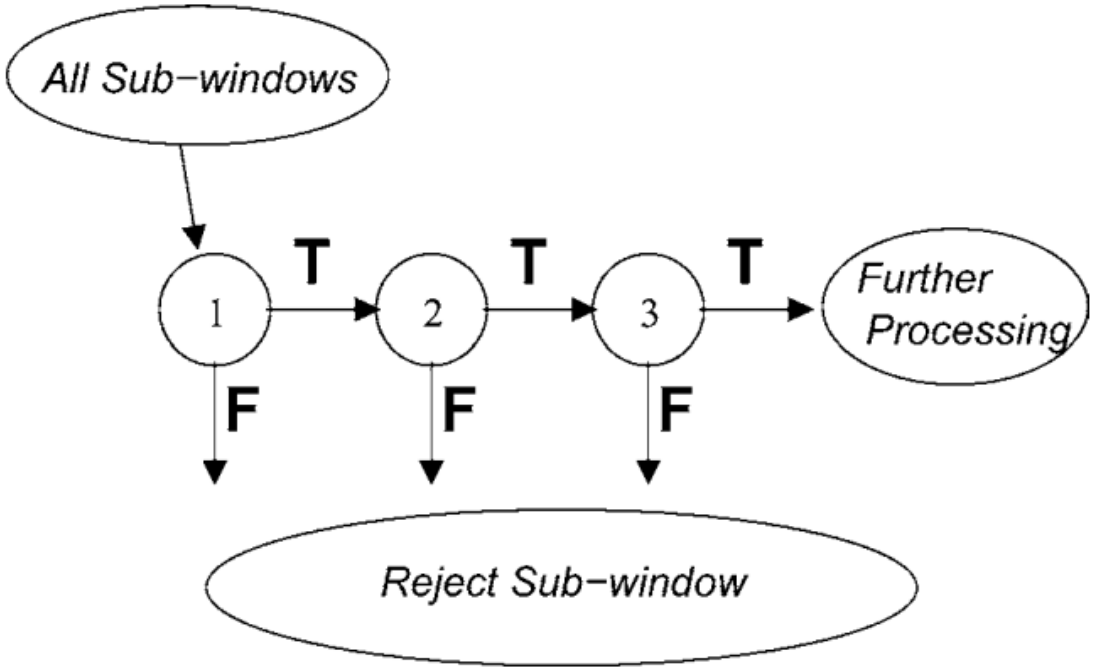
Viola-Jones Algorithm

The Attentional cascade

- Cascading in classifiers is a technique used in machine learning and computer vision to improve the efficiency and accuracy of classification algorithms.
- It involves using a series of classifiers that are arranged in a hierarchical fashion, with each subsequent classifier becoming more specialized and accurate in its classification task.
- The key insight is that smaller, and therefore more efficient, boosted classifiers can be constructed which reject many of the negative sub-windows while detecting almost all positive instances

Viola-Jones Algorithm

- The stages are constructed by training classifiers using Adaboost.
- The Algorithm learns a series of weak classifiers to combine them to make a strong classifier.



Viola-Jones Algorithm

- With only a few actions, the classifier can drastically minimise the number of sub-windows that require additional processing.
  1. Evaluate the rectangle features (requires between 6 and 9 array references per feature)
  2. Compute the weak classifier for each feature (requires one threshold operation per feature).
  3. Combine the weak classifiers (requires one multiply per feature, an addition, and finally a threshold)
- When any of the weak classifiers in the cascade fail in a subwindow, the algorithm moves to the next subwindow.
- Due to the decision tree nature of the process, the 2nd classifier is faced with a more difficult task than the first one.

Viola-Jones Algorithm

Training:-

1. Collecting positive negative samples.
2. Preparing the samples.
3. Creating a positive samples vector.
4. Creating a negative samples vector.
5. Training the classifier.
6. Testing the classifier.
7. Tweaking the classifier.
8. Using the classifier.

Viola-Jones Algorithm

The false positive rate of the cascade classifier is given by:

$$F = \prod_{i=1}^{K} f_i$$

F is the positive rate of the cascaded classifier, K is the number of classifiers, f_i is the false positive rate of the ith classifier.

The detection rate of the cascade is:

$$D = \prod_{i=1}^{K} d_i$$

, where d_i is the detection rate of the ith classifier.

Viola-Jones Algorithm

Example

Suppose we want a cascade with a detection rate of D = 0.90.
let K = 10.
So we need $d_i = 0.99$, [since $0.9 \approx 0.99^{10}$]
it can be shown that if we have a false positive rate of 30% then the cascade of the classifier will have a false positive rate $0.3^{10} \approx 6x10^{-6}$.
which is quite low!

Viola-Jones Algorithm

- We can also evaluate the number of features since it is a probabilistic process.
- Analysis of the image distribution allows us to predict how the process will behave.
- The expected number of features can be given as:

$$N = n_0 + \sum_{i=1}^{K} (n_i \prod_{j<i} p_j)$$

where N = expected number of features evaluated,
n_i = expected number of features of ith classifier.
p_i = positive rate of the ith classifer.

Viola-Jones Algorithm

Algorithm

- User selects values for f and d.
- User selects target overall false positive rate, F_{target}.
- P = set of positive examples; N = set of negative examples.
- $F_0 = 1.0$; $D_0 = 1.0$
- i = 0
- while $F_i > F_{target}$
  - $i \leftarrow i + 1$
  - $n_i = 0$; $F_i = F_{i-1}$
  - while $F_i > f \times F_{i-1}$
    - $n_i \leftarrow n_i + 1$
    - Use P and N to train a classifier with n_i features using AdaBoost.
    - Evaluate current cascaded classifier on validation set to determine F_i and D_i .
    - Decrease threshold for the ith classifier until the current cascaded classifier has a detection rate of at least $d \times D_{i-1}$(this also affects F_i)
  - $N \leftarrow \emptyset$
  - If $F_i > F_{target}$ then evaluate the current cascaded detector on the set of non-face images and put any false detections into the set N

Viola-Jones Algorithm

Benefit

There are several benefits of using a cascade of classifiers for object detection, which operates pointwise by evaluating each object candidate at multiple stages before making a final decision:

1. Reduced computation time
2. Increased accuracy.
3. Robustness to variation

Drawback

1. the training set of negative examples would have to be relatively small.
2. it requires careful tuning of several parameters, including the number of stages, the number of features per stage etc.

Discrete Wavelet Transform

wavelet

A wavelet is a waveform of effectively limited duration that has an average value of zero and nonzero norm.

It is used to used extract information from many different kinds of data, including audio signals and images.

Discrete Wavelet Transform

Why choose DWT

- A discrete wavelet transform (DWT) can locate a signal in both time and frequency resolutions.
- Wavelet transformation can be used to decompose an image into its component parts at different scales and resolutions.
- It can also be used to extract features from images that are relevant to facial expression recognition, such as texture or color information.

Discrete Wavelet Transform

Here we will deal with discrete wavelet transformation:
let f(x) is a function on the spatial domain. $f(x) \in L^3(R)$ $\varphi(x)$ is the scaling function and $\psi(x)$ is the wavelet function.

$$f(x) = \sum_k W_\varphi(j_0, k)\varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_k W_\psi(j, k)\psi_{j,k}(x)$$

where the coefficient $W_\varphi(j_0, k)$ and $W_\psi(j, k)$ can be written as:

$$W_\varphi(j_0, k) = \frac{1}{\sqrt{M}}\sum_x f(x)\varphi_{j_0,k}(x)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}}\sum_k f(x)\psi_{j,k}(x)$$

Discrete Wavelet Transformation

High pass filter

- used to extract high-frequency component
- contains information about the sharp changes and edges in the signal.

Low pass filter

- used to extract low-frequency components.
- contains information about the overall trend and shape of the signal.

Convolutional Neural Networks

Introduction to ConvNets

- A class of neural networks that specializes in processing data that has a grid-like topology, such as an image.
- A digital image is a binary representation of visual data. It contains a series of pixels arranged in a grid-like fashion that contains pixel values to denote how bright and what color each pixel should be.
- Works analogously as the human brain
- CNNs are a type of multi-layer neural network that can discern visual patterns from pixel images.
- Provides vision to computers

Convolutional Neural Networks

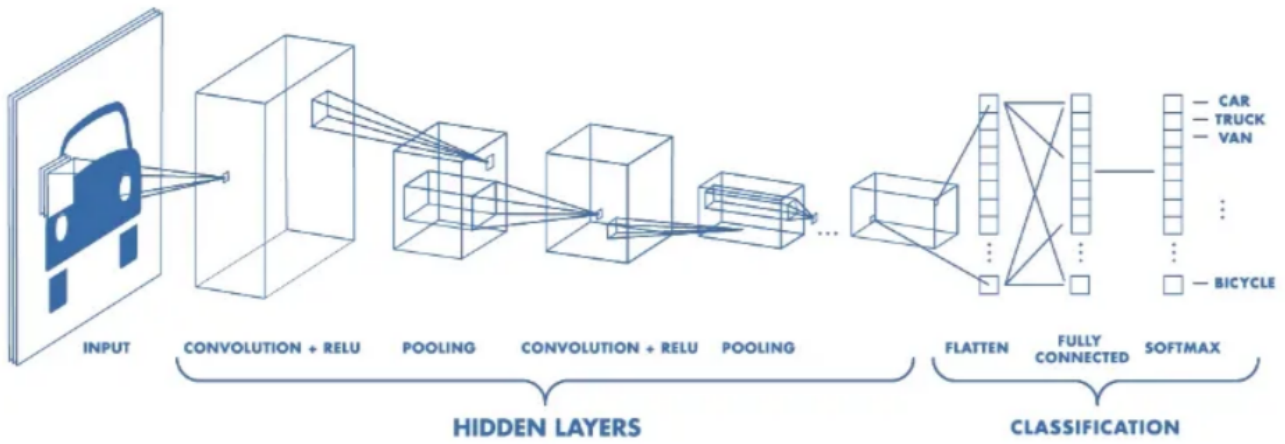


Figure: Architecture of a CNN

A CNN typically has three layers:

- a convolutional layer
- a pooling layer
- fully connected layer

Convolutional Neural Networks

Key components The convolutional layer is the key building block of a Convolutional Neural Network (CNN). It applies a set of filters to the input data to produce a set of feature maps that capture local patterns and structures in the input.
The key components of this convolutional layer are:

- Filters: Small matrices of weights that are learned during training, typically square.
- Sliding Window: Filters are slid over the input image or feature map in a process called convolution. At each position, the filter weights are multiplied by the corresponding input values, and the results are summed to produce a single output value.
- Feature Map: The output of the convolution operation at each position is stored in a new matrix, called a feature map, represents a local feature.

Convolutional Neural Networks

- Padding: To avoid losing information at the edges of the input data, the input data can be padded with zeros before applying the filters. This is called zero-padding.
- Stride: The filters can be applied with a stride, which determines the distance between each position where the filter is applied.
- Number of filters: A hyperparameter that determines the complexity and capacity of the layer.

Convolutional Neural Network

How it Works

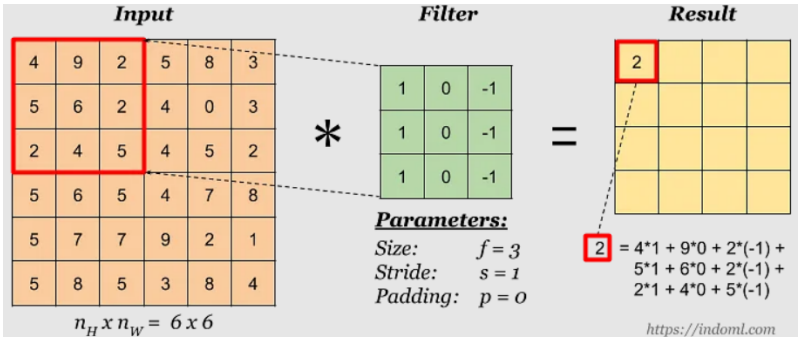


Figure: Output of Convolutional Operation with kernel size=3, stride=1, padding=0

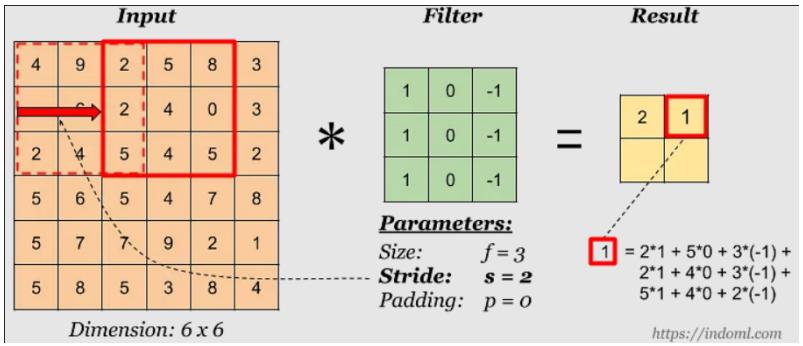


Figure: Output of Convolutional Operation with kernel size=3, stride=2, padding=0

Convolutional Neural network

The pooling layer

- Responsible for dimensionality reduction of the feature map
- Decreases the required amount of computation and weights.
- Pooling can be divided into two main types: maximum pooling and average pooling

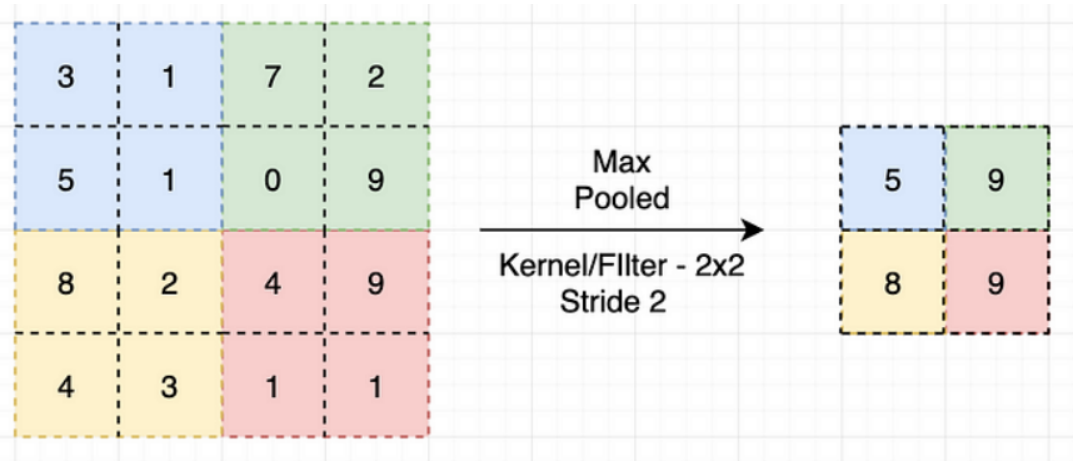


Figure: Max pooling

Convolutional Neural network

Fully connected layer

- frequently found near the end of CNN architectures
- perform classification or regression tasks on the output features learned by the convolutional layers
- They take the learned features from the convolutional layers,flatten them, perform the normal mathematical operations performed in a basic neural net and use them to classify the input data.

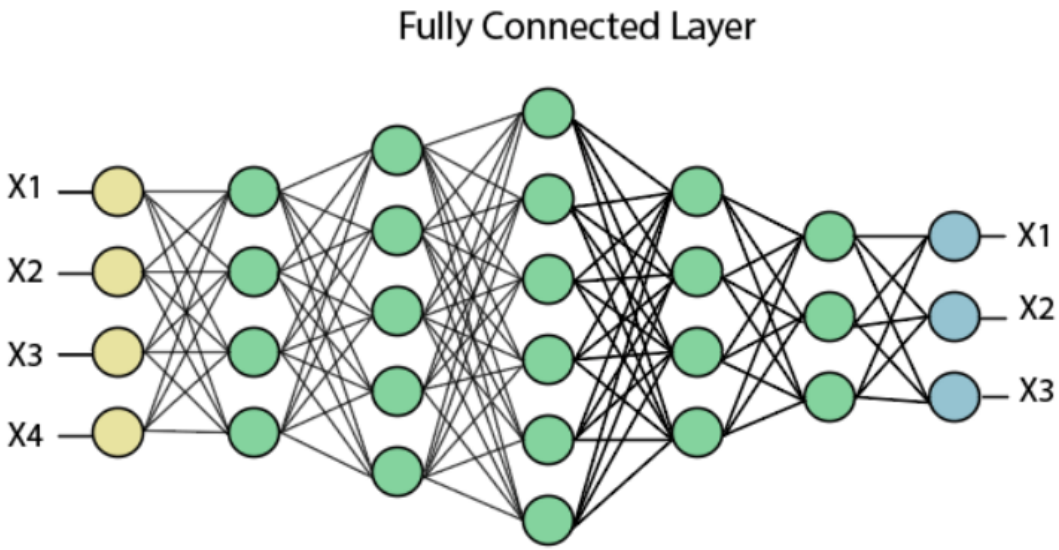


Figure: Fully Connected Layer

Results

Table 8. The comparison between different approaches and our approach for the JAFFE face database. CNN— convolutional neural network.

Approach	Recognition Rate %
SVM [34]	95.60
Gabor [35]	93.30
2-Channel CNN [12]	94.40
Deep CNN [14]	97.71
Normalization+ DL [36]	88.73
Viola-Jones+ CNN	95.30
Proposed Method	98.63

Table 9. The comparison between different approaches and our approach for the CK+ face database.CNN— convolutional neural network.

Approach	Recognition Rate %
SVM [37]	95.10
Gabor [35]	90.62
3D-CNN [38]	95.00
Deep CNN [14]	95.72
Normalization+ DL [36]	93.68
Viola-Jones+ CNN	95.10
Proposed Method	97.05

Shortcomings

- Cannot detect faces with goggles, spectacles or any kind of objects which cover the eyes.
- It can't properly detect side faces.
- It is unable to detect when the eye is partially closed
- Doesn't give proper results when the face is tilted
- Cannot detect images with low resolution
- Unable to detect faces in a blurred image
- cannot detect small faces in a group photograph

References

- Improved Facial Expression Recognition Based on DWT Feature for Deep CNN:- Ridha Ilyas Bendjillali, Mohammed Beladgham, Khaled Merit and Abdelmalik Taleb-Ahmed
- Robust Real-Time Face Detection:- PAUL VIOLA, MICHAEL J. JONES
- Study of Viola-Jones Real Time Face Detector: Kaiqi Cen
- The wavelet transform:- `https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34`
- Hands-On Machine Learning with Scikit-Learn and TensorFlow:- Aurélien Géron