# DETECTING FACIAL RETOUCHING (OR DIGITAL ALTERATION) USING DEEP LEARNING

## TEAM MEMBERS:


1. KAMAL SHARMA (2017CSB1084)
2. SHUBHAM (2017EEB1167)
3. MCMILLON WALTON (2017CHB1047)


## INTRODUCTION:

Detecting facial retouching (or digital altercation) using deep learning- The project aims to implement a supervised algorithm to classify face images as original or retouched and analyze the effect of digital altercations on performance of common face recognition algorithms. We test face recognition algorithms on a database having two sets of face images- unaltered and retouched images for identification accuracy. To detect retouching in face images, a Supervised Deep Boltzmann Machine algorithm will be used which uses facial parts to learn discriminative features to classify the face images as original or retouched.

## MOTIVATION FOR THE PROBLEM:

1. In several photo-identification documents such as driver's license, passport, and other government issued id-cards, hard-copy photographs are required. Due to lack of process where ISO standards are strictly followed, these images are used for creating identification documents including passport. But mostly the user gets the image retouched for better print quality which is later used as a biometric enrollment image to be matched with real-time photographic/untouched images. If retouched images are used in the biometrics application, recognition accuracy can be considerably affected. In the image forensic domain, this problem is known as "detecting photographic doctoring".

2. In social media platforms like Facebook, Instagram, Flickr etc. users often retouch their images for beautification purposes like age spot, wrinkles, pimples removal, change shape of eyebrows and nose, adjust skin tone, make the face thinner or fuller etc. While auto-tagging the face recognition algorithms may not yield good results on these retouched photos.

3. The perfect skin and facial features projected by the fashion industry after retouching photographs of celebrities distorts the mindset of the general public towards weight loss and causes anorexia, inferiority complex with their own skin and body shape, lower self-esteem and other disorders as is clear from the mandate by several countries government to declare if an advertisement photo is digitally retouched or not called the 'photoshop law'.

**Figure 1 :** The left one is the original photo, and the right one is "retouched"

## Related Work:

**1. E. Kee, and H. Farid, A perceptual metric for photo retouching, P.N.A.S., vol. 108, pp. 19907–19912, 2011 [6]**

They also tried on a similar problem. They collected a diverse set of 468 original and retouched photos from a variety of on-line sources. Human observers were then asked to rank the amount of photo alteration done. They then correlated those ratings with summary statistics using non linear Support vector regression (SVR)

**2. Virtually Perfect: Image Retouching and Adolescent Body Image Kristen Harrison & Veronica Hefner [7]**

A similar study on how not the face but body also is retouched to alter and decrease the age of a person.

**3. Detecting Facial Retouching using SDL Technique Y.Arockia Dayana | Dr.K.Mahesh [8]**

They proposed a supervised RBM could be learned from a labelled training base. The supervised RBM can be stacked to form a deep learning framework. Greedy layer-by-layer training is performed to learn the weights and parameters of the supervised RBM. Feature-Based Image Metamorphosis is used to detect the image forensic. They concluded that improvement in classification accuracy can be attributed to the supervised DBM and to the form of the SVM used for classification. It uses facial parts (which they extracted using Voila jones algorithm) to learn discriminative features to classify face images as original or retouched.

**4. Demography-based Facial Retouching Detection using Subclass Supervised Sparse Autoencoder University of Notre Dame, IIIT-Delhi [9]**

They proposed an autoencoder – an unsupervised encoding to learn features that well represent the input data.

# METHODOLOGY:

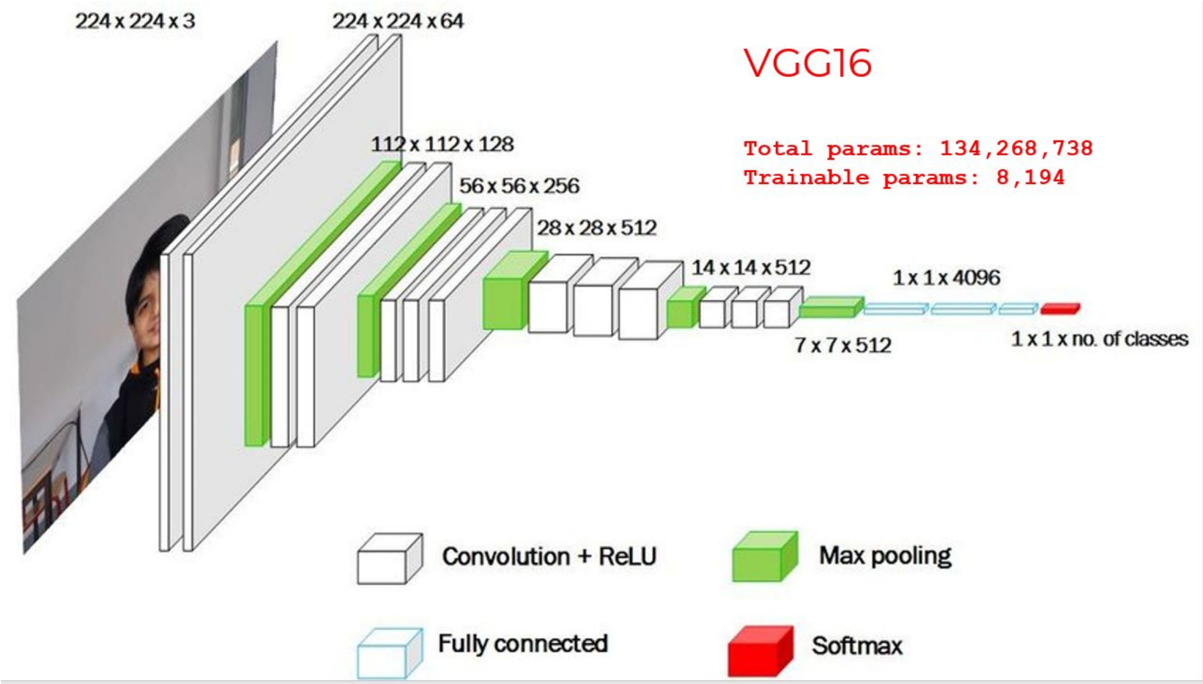
1) Tried standard CNN classification architectures (VGG16, Resnet50, Resnet34)
2) Implemented Unsupervised RBM.
3) Extracted facial features using shape_predictor_68_face_landmarks
4) Tried error level analysis on the whole image.
    a) Accuracy ranges from 0.5 to 0.65 in the tried models.
5) Edited facial patches extraction code to get more wider features
6) Implemented SRM (with loss function similar to logistic classifiers).
7) Implemented SVM for classification.
8) 7 Facial patches are extracted from the image and converted to a fixed size vector
9) 7 SRBMs are trained, each on a particular facial patch.
10) Concatenated output of all RBMs are fed to SVM for classification.

    Test Accuracy: 0.76

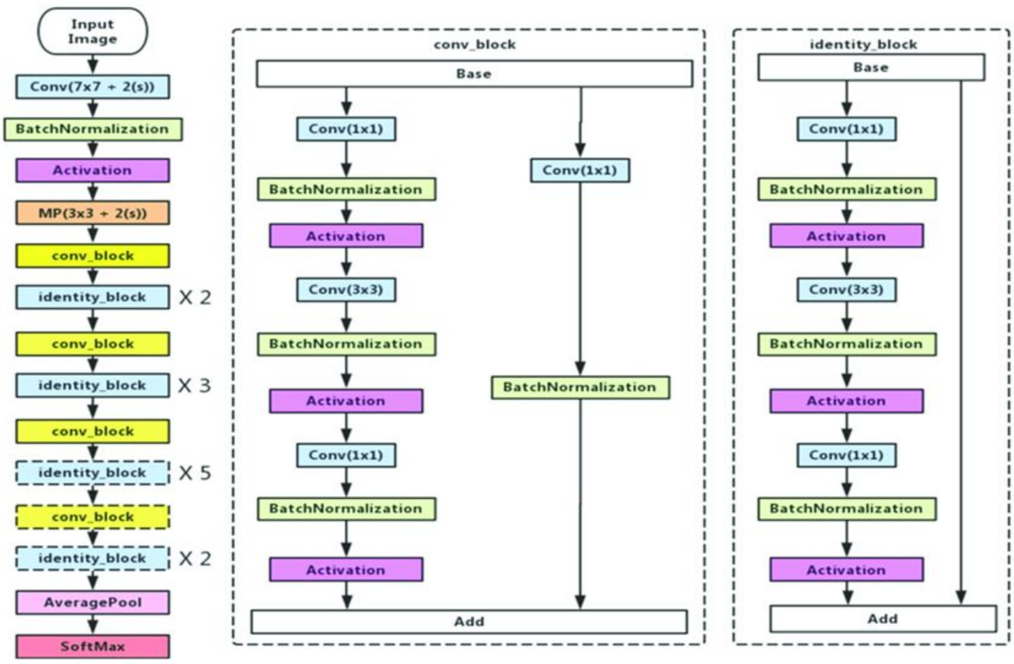
11) Implemented SRBM with different loss function.

# Results of experiment performed using well known Classification CNNs and DBN

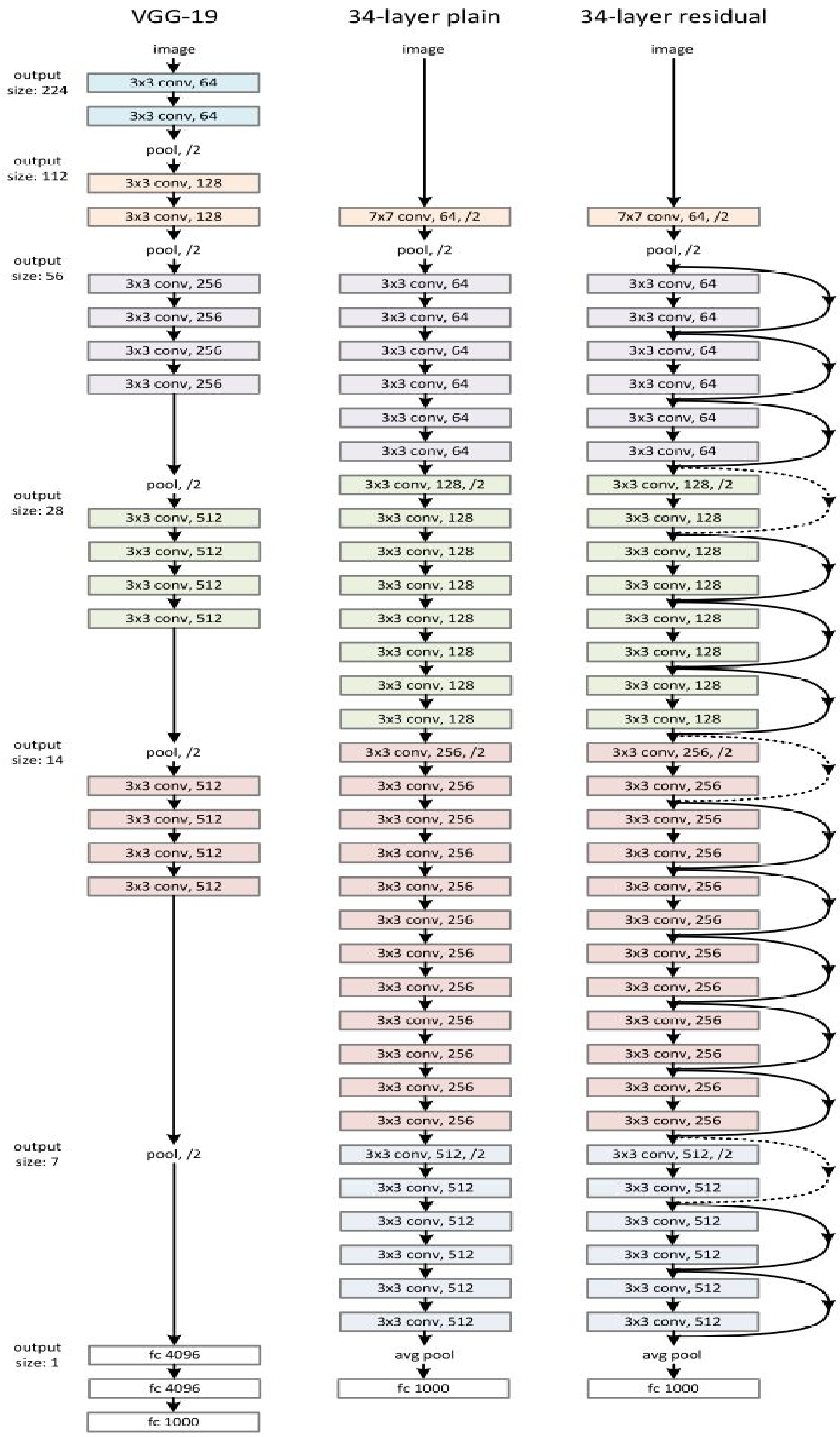
## 1. VGG-16



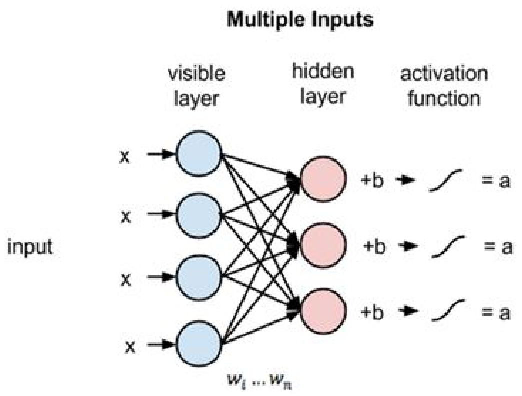

## 2. Resnet 50

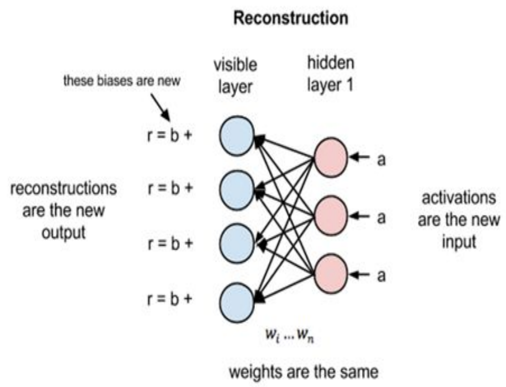


## RESNET-34


VGG-19
34-layer plain
34-layer residual
output
size: 224
output
size: 112
output
size: 56
output
size: 28
output
size: 14
output
size: 7
output
size: 1
image
3x3 conv, 64
3x3 conv, 64
pool, /2
3x3 conv, 128
3x3 conv, 128
pool, /2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
pool, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
pool, /2
fc 4096
fc 4096
fc 1000
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
avg pool
fc 1000
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
avg pool
fc 1000

# Unsupervised Restricted Boltzmann Machine[3]



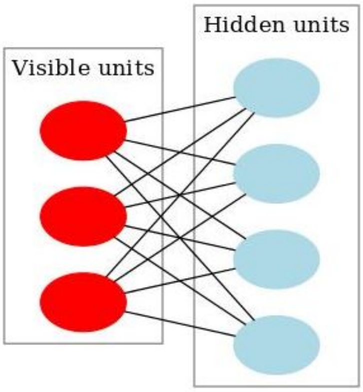




## Boltzmann Machine:

- Boltzmann Machine is a generative unsupervised model, which involves learning a probability distribution from an original dataset and using it to make inferences about never before seen data.
- Boltzmann Machine has an input layer (also referred to as the visible layer) and one or several hidden layers (also referred to as the hidden layer).
- Boltzmann Machine doesn’t expect input data, it generates data. Neurons generate information regardless they are hidden or visible.
- Boltzmann Machines are primarily divided into two categories: Energy-based Models (EBMs) and Restricted Boltzmann Machines (RBM). When these RBMs are stacked on top of each other, they are known as Deep Belief Networks (DBN).

## Restricted Boltzmann Machine:

What makes RBMs[3] different from Boltzmann machines is that visible *nodes aren't connected to each other*, and hidden nodes aren’t connected with each other. Other than that, RBMs are exactly the same as Boltzmann machines.




Representation of RBM

- RBM is the neural network that belongs to the energy-based model
- It is a probabilistic, unsupervised, generative deep machine learning algorithm.
- RBM's objective is to find the joint probability distribution that maximizes the log-likelihood function.
- RBM is undirected and has only two layers, Input layer, and hidden layer
- All visible nodes are connected to all the hidden nodes. RBM has two layers, visible layer or input layer and hidden layer so it is also called an **asymmetrical bipartite graph.**
- No intralayer connection exists between the visible nodes. There is also no intralayer connection between the hidden nodes. There are connections only between input and hidden nodes.
- The original Boltzmann machine had connections between all the nodes. Since RBM *restricts* the intralayer connection, it is called a *Restricted Boltzmann Machine.*

## Error Level Analysis:

In ELA(Error Level Analysis)[2], image is resaved with reduced quality. Then we take the difference between both original and retouched images so that digitally altered features can be enhanced. This technique is generally used in forgery.

# Results on IIT Ropar dataset

1. **RESNET50- Results on IIT Ropar dataset**

   Accuracy : 0.45

   Classification Report

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| retouched | **0.43** | **0.33** | **0.38** | **60** |
| original | **0.46** | **0.57** | **0.51** | **60** |

2. **VGG16 - Results on IIT Ropar dataset**

   Accuracy : 0.53

   Classification Report

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| retouched | **0.52** | **0.83** | **0.64** | **60** |
| original | **0.57** | **0.22** | **0.31** | **60** |

3. **RESNET34- Results on IIT Ropar dataset**

   Accuracy : 0.52

   Classification Report

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| retouched | **0.54** | **0.51** | **0.52** | **60** |
| original | **0.50** | **0.53** | **0.51** | **60** |

4. **URBM+ELA-Results on IIT ROPAR dataset:**

Accuracy : 0.61

Classification Report

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| retouched | **0.61** | **0.64** | **0.62** | **60** |
| original | **0.62** | **0.58** | **0.60** | **60** |

# Viola–Jones Algorithm

The **Viola–Jones** is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. All faces have some similarities which can be matched using **Haar Features**.

# Haar Features

These are rectangular areas which involve the sum of the pixel intensities.

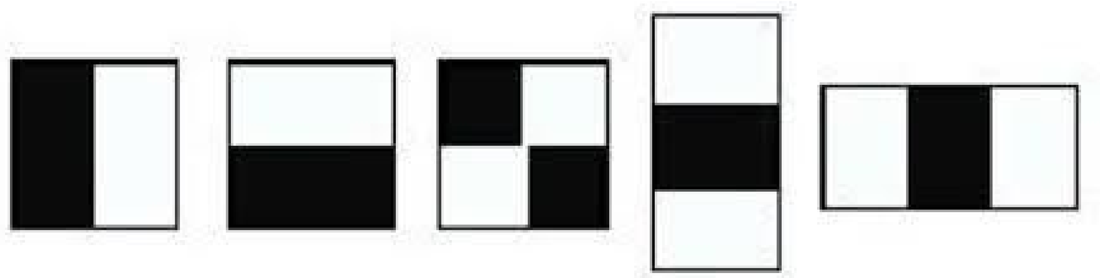


**Figure 2:** A collection of Haar features

The value of any given feature is the sum of the pixels within clear rectangles (white pixels ) subtracted from the sum of the pixels within shaded rectangles (black pixels). For example the eye region is usually darker than the cheek region so this can be matched using the following Haar feature or the nose region in the middle is brighter than the 2 eyes.

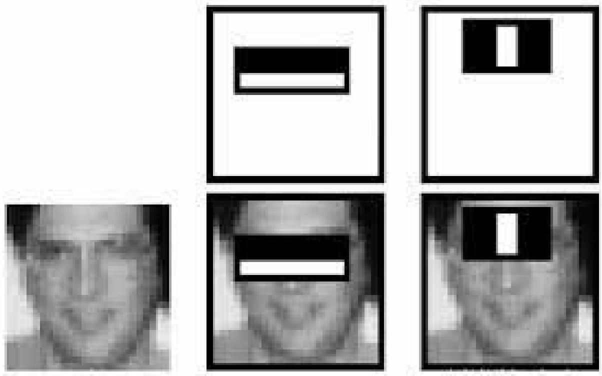


**Figure 3:** Matching Haar features with the facial features

The features are matched after gray scaling the image. But obviously these are ideal cases that the feature consists of only 2 intensities and not a spectrum.

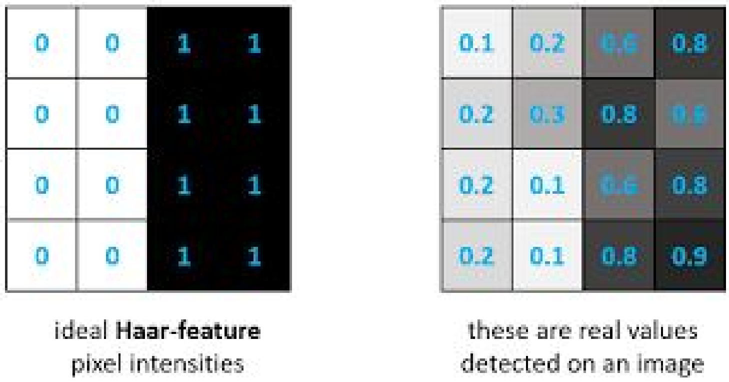



**Figure 4:** The ideal haar feature and non- ideal realistic haar feature

When we are trying to identify a face in an image we will use multiple Haar features. Some of them will have to be assigned more weights than others in order to get a greater accuracy of classifying a face.

$$h(\mathbf{x}) = \mathrm{sgn}\left(\sum_{j=1}^{M} \alpha_j h_j(\mathbf{x})\right)$$

h(x) : Face classifier

$\alpha_i$ : weights

$h_i(x)$ : classifier matched with each of the facial features.

There is also a threshold for the value of pixel intensity to be chosen, to classify which is dark and which is the bright region of the haar feature.

$$h_j(\mathbf{x}) = \begin{cases} -s_j & \text{if } f_j < \theta_j \\ s_j & \text{otherwise} \end{cases}$$

$f_j$ is the haar feature and $\theta_j$ is the threshold value.

$S_j \in -1,1$

After adjusting the parameters of Haar features we can get a good classifier for a face.

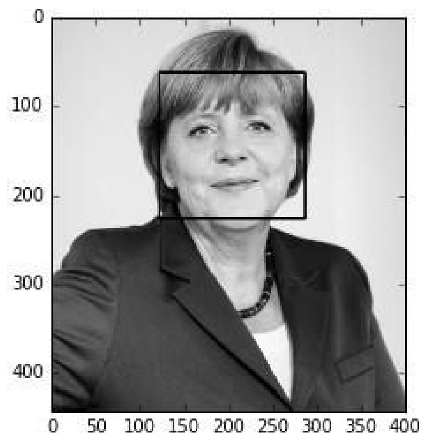
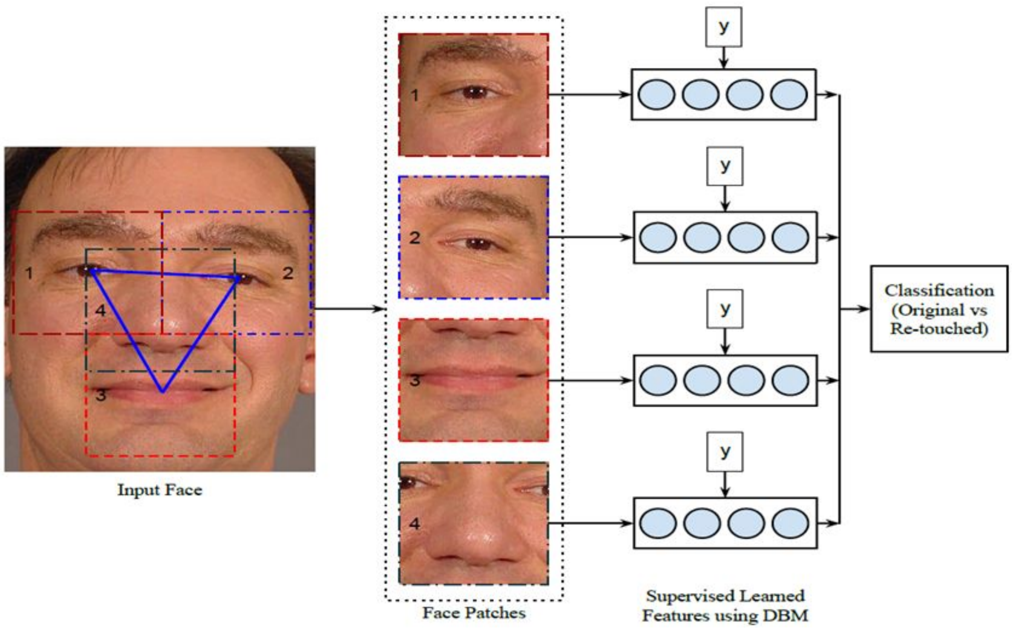



**Figure 5:** Facial detection using Haar cascade classifier




Source: Detecting facial retouching using supervised deep learning paper [1]

1. Extracted facial features using shape_predictor_68_face_landmarks
2. Independently extract vectors from each facial batch

**Loss function for supervision :**

$$E(\mathbf{x}, \mathbf{h}; \theta) = -\sum_i \sum_j \frac{x_i}{\sigma_i} w_{ij} h_j - \sum_i \frac{(x_i - b_i)^2}{2\sigma_i^2} - \sum_i a_j h_j \quad (1)$$

where $\mathbf{x} \in \mathbf{R}^D$ denotes real-valued input vector (visible layer vector), $\mathbf{h}$ is the hidden layer representation, and $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}, \sigma\}$ are the model parameters.

Lsup = Lgen + $\lambda$ ||Zc||2,1

Lgen = − Σi ( logP(xi) )

P(x) = (1/Z) * Σh ( exp(−E(x,h; $\theta$ )) )

Where, Z = Σx,h ( exp(−E(x,h)) )

# Algorithm Restricted Boltzmann Machine weight updation:

---

**Algorithm 1.** $k$-step contrastive divergence

---

**Input**: RBM $(V_1, \ldots, V_m, H_1, \ldots, H_n)$, training batch $S$
**Output**: gradient approximation $\Delta w_{ij}$, $\Delta b_j$ and $\Delta c_i$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$

init $\Delta w_{ij} = \Delta b_j = \Delta c_i = 0$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$
**forall the** $\boldsymbol{v} \in S$ **do**
  $\boldsymbol{v}^{(0)} \leftarrow \boldsymbol{v}$
  **for** $t = 0, \ldots, k-1$ **do**
    **for** $i = 1, \ldots, n$ **do** sample $h_i^{(t)} \sim p(h_i \,|\, \boldsymbol{v}^{(t)})$
    **for** $j = 1, \ldots, m$ **do** sample $v_j^{(t+1)} \sim p(v_j \,|\, \boldsymbol{h}^{(t)})$
  **for** $i = 1, \ldots, n$, $j = 1, \ldots, m$ **do**
    $\Delta w_{ij} \leftarrow \Delta w_{ij} + p(H_i = 1 \,|\, \boldsymbol{v}^{(0)}) \cdot v_j^{(0)} - p(H_i = 1 \,|\, \boldsymbol{v}^{(k)}) \cdot v_j^{(k)}$
    $\Delta b_j \leftarrow \Delta b_j + v_j^{(0)} - v_j^{(k)}$
    $\Delta c_i \leftarrow \Delta c_i + p(H_i = 1 \,|\, \boldsymbol{v}^{(0)}) - p(H_i = 1 \,|\, \boldsymbol{v}^{(k)})$

---

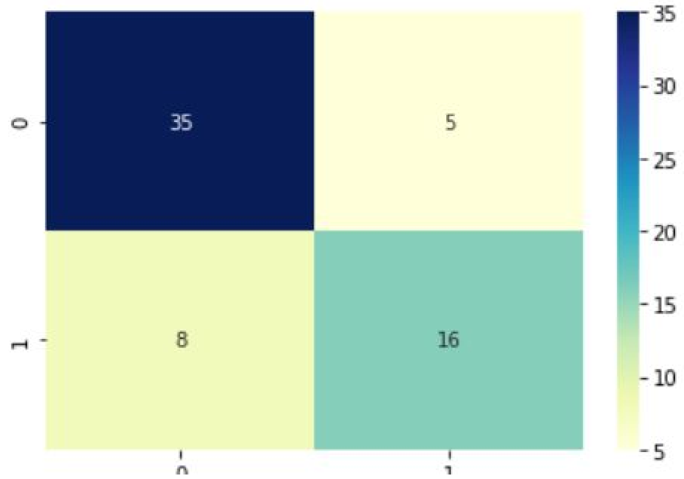
# RESULT AND DISCUSSION:

Final as per the following research paper[1], we implemented the Restricted Boltzmann Machine and used SVM(2c) as a classifier. To improve accuracy and explore more aspects instead of 4 Facial patches, we extracted 7 facial patches(both eyes, nose, lips, cheeks, chin) and fed each facial patch into RBM after converting image to vector. Hidden layer of each RBM is concatenated together and then SVM is used for classification(Original/Retouched)

Here are the results: 0 means original whereas 1 corresponds to retouched class label

classification report

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 0.81 | 0.88 | 0.84 | 40 |
| 1.0 | 0.76 | 0.67 | 0.71 | 24 |
| accuracy | | | 0.80 | 64 |
| macro avg | 0.79 | 0.77 | 0.78 | 64 |
| weighted avg | 0.79 | 0.80 | 0.79 | 64 |




# SUMMARY:

This project is actually the subpart of BTech project(Kamal Sharma). The project gave us immense exposure to new Machine Learning tools and techniques. We learnt about the unsupervised generative technique Restricted Boltzmann Network which is extensively used in recommendation systems, Facial feature extraction and Error Level Analysis methods.

Our method comprises Unsupervised RBM, Facial_features_Extraction, Error_Level_analysis and SVM. Although we are not able to achieve the accuracy more than the research paper[1] but certainly learnt a lot and also got precision, recall and accuracy better than existing methods.

We got 0.8 accuracy and 0.79 f1-score as our best model's result in contrast to ~ 0.5 accuracy of CNN architectures(VGG6,Resnet34,Resnet50) and 0.87 of SRBM accuracy(claimed in paper[1]).

Link to codes:

1. RBM : link
2. Facial_features_extraction : link
3. Facial landmarks : link
4. Error Level Analysis :
5. SRBM : link
6. ELA : link
7. VGG16,Resnet34,Resnet50 : link

   Add codes at same place : github repo

# REFERENCES:


1. A. Bharati, R. Singh, M. Vatsa and K. W. Bowyer, "Detecting Facial Retouching Using Supervised Deep Learning," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 9, pp. 1903-1913, Sept. 2016
2. Error Level Analysis: Daniel CJ, Yuri C. Borges, Leandro D. Coelho, Image forgery detection by semi-automatic wavelet soft-Thresholding with error level analysis, Nov. 2017
3. Unsupervised Restricted Boltzmann Machine: Asja Fischer, C. Ijel, An introduction to Restricted Boltzmann Machine, Jan 2012
4. ND-IIITD Dataset, https://cvrl.nd.edu/projects/data/#nd-iiitd-retouched-face-database
5. IIT Ropar Dataset, 250+ celebs Before-After dataset
6. E. Kee, and H. Farid, A perceptual metric for photo retouching, P.N.A.S., vol. 108, pp. 19907–19912, 2011
7. Virtually Perfect: Image Retouching and Adolescent Body Image Kristen Harrison & Veronica Hefner
8. Detecting Facial Retouching using SDL Technique Y.Arockia Dayana | Dr.K.
9. Demography-based Facial Retouching Detection using Subclass Supervised Sparse Autoencoder University of Notre Dame, IIIT-DelhiMahesh
10. VGG16, Resnet34, Resnet50 : https://keras.io/api/models/sequential/
11. An Analysis of the Viola-Jones Face Detection Algorithm ; Yi-Qing Wang