Computer Vision Project

Facial Emotion Recognition

Authors:

Alexandre Robin
Yanis Chamson

Supervisor:

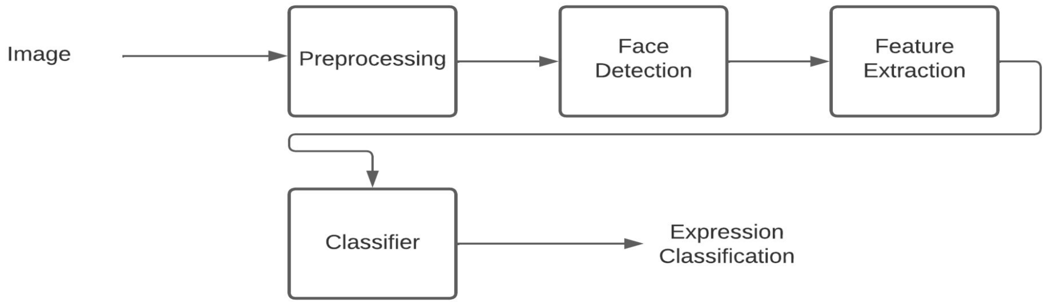
Maria Vakalopoulou

Facial Emotion Recognition (FER)

- FER merges psychological insights with technological advancements, playing a pivotal role in computer vision
- Applications range from UX design enhancement to fortifying security systems
- Our project focuses on refining traditional computer vision techniques for effective FER
- **Key objective**: Accurately classify emotional states from facial expressions to contribute to theoretical and practical domains and comparative analysis of traditional computer methods

Feature Extraction

- Critical stage in Facial Emotion Recognition (FER).
- Extraction of discriminative features from facial expressions.



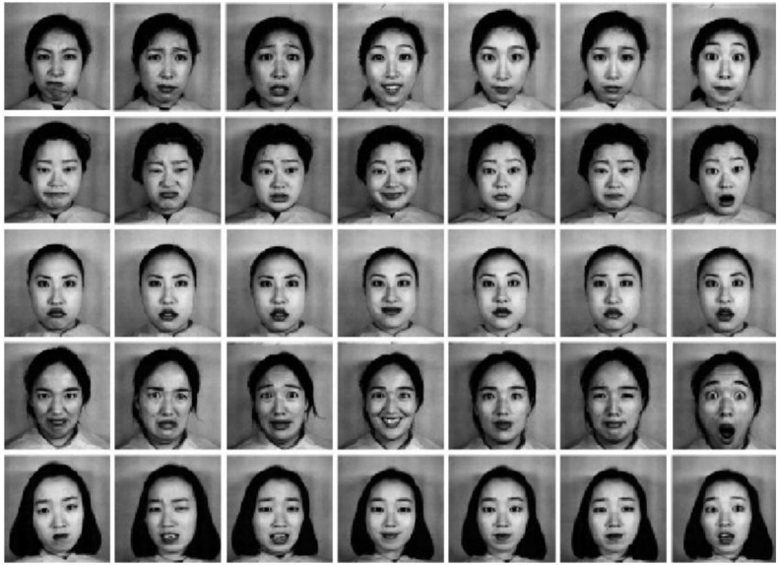
Facial Expression Classification diagram block

CK+ Dataset

- Designed to facilitate research in facial expression analysis, emotion recognition, and related computer vision tasks
- Captures facial expressions associated with basic emotions, including happiness, sadness, surprise, anger, disgust, and fear.
- Includes sequences from multiple subjects, contributing to its diversity.
- Primarily consists of grayscale, moderate resolution images, with variations in lighting and facial poses.

Japanese Female Facial Expression

- Compiled to address the need for cultural diversity in facial expression analysis, specifically focusing on Japanese females
- Captures a range of facial expressions corresponding to culturally relevant emotions, including joy, sadness, surprise, anger, disgust, and fear within the Japanese cultural context.
- Utilizes high-resolution color images to capture facial features with greater detail
- Images are curated to reflect diverse social and cultural situations unique to Japan, enhancing the dataset's cultural relevance.



Evaluation

Metric

- Accuracy

Model

- SVM with rbf Kernel

Parameters are optimized.

Cross Validation is realized on a K-fold with parameter K=4.

Histogram of oriented Gradients

- Local object appearance and shape can often be characterized by the distribution of local intensity gradients or edge directions
- Uses **magnitude** and **direction** of the gradient to compute the features
- Generates histograms using those informations for each region of an image

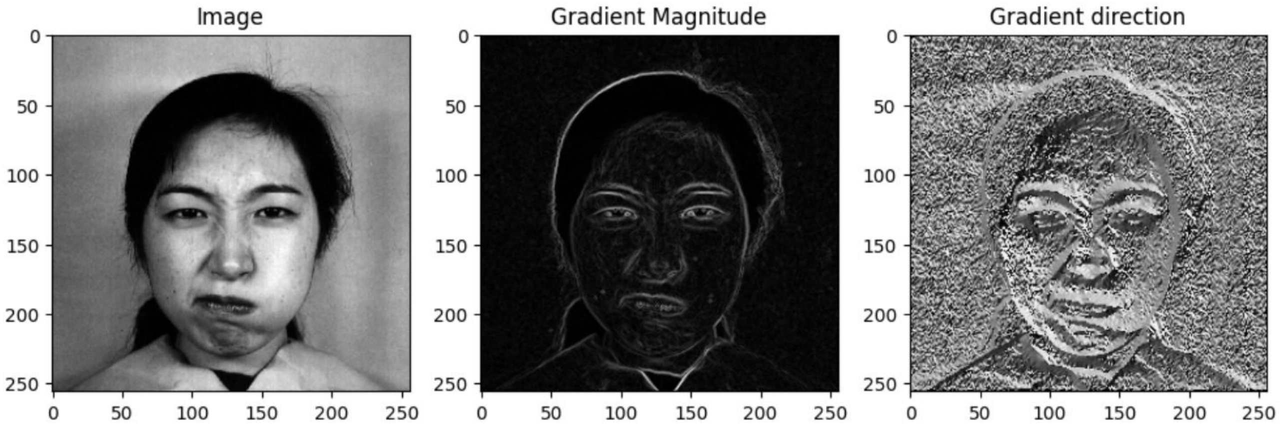
Step 1 : Gradient Calculation

- Calculate gradient magnitude (g) following this formula :

$$g = \sqrt{g_x^2 + g_y^2}$$

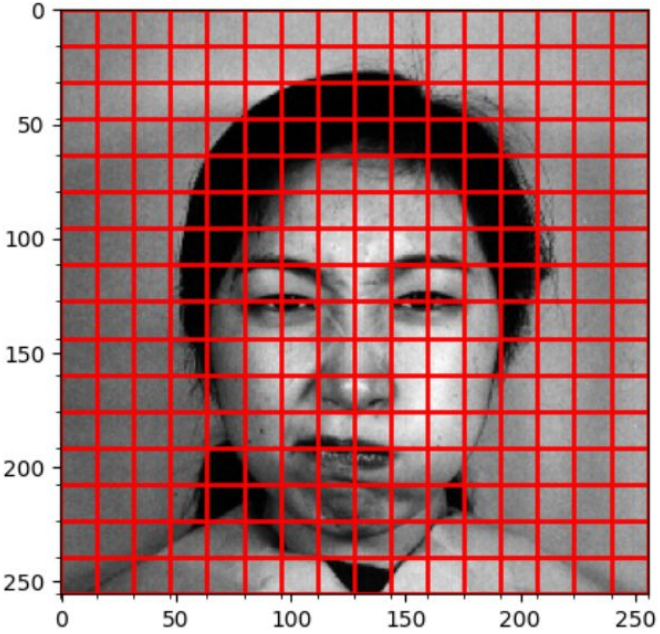
- Calculate Gradient direction (θ) following this formula :

$$\theta = \arctan \frac{g_y}{g_x}$$



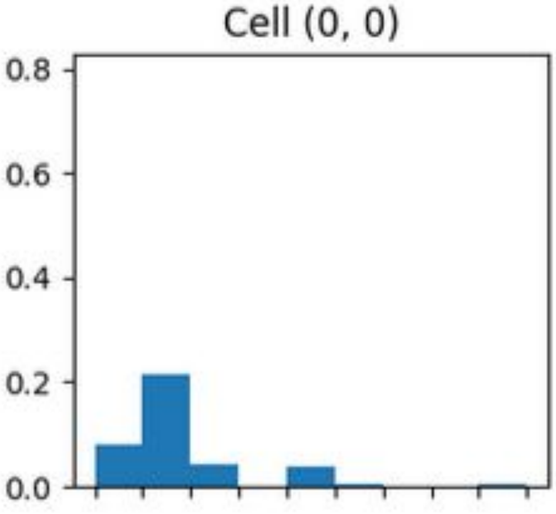
Step 2 : Cell Division

- The image is divided into cells.
- Common cell sizes are 8x8 pixels or 16x16 pixels.
- The gradient magnitude and direction information is then collected for each cell.



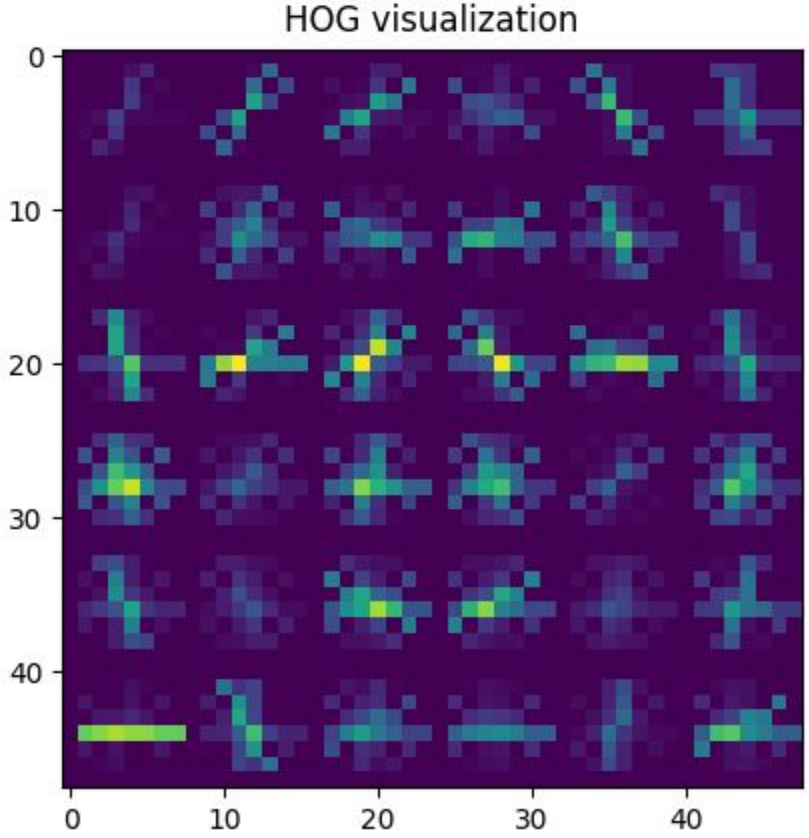
Step 3 : Histogram calculation for each cell

- For each cell, we create a histogram of gradient orientations
- The goal is to represent the dominant directions of gradients within the cell
- Each pixel in the cell has an associated gradient vector, which consists of both magnitude and direction.
- The gradient direction is the angle of the vector in relation to a reference axis
- Divide the gradient direction range (usually 0 to 360 degrees) into a certain number of bins
- Accumulate the gradient magnitudes into the corresponding bins of the histogram based on their orientations.



Step 4 : Block Normalization

- Group adjacent cells into blocks
- Common block sizes are 2x2 cells or 3x3 cells
- Normalize the histograms within each block. This can be done by dividing each histogram by the sum of the histograms within the block.
- Normalization helps to make the descriptor more robust to changes in lighting and contrast.
- Concatenate the normalized histograms of all the blocks to form the final HOG descriptor for the image.


HOG visualization
0
10
20
30
40
0
10
20
30
40

Histogram of Oriented Gradients

Results

HOG + SVM	Mean	Standard Deviation
CK+	99.18%	0.87%
JAFFE	88.25%	1.63%

Linear Binary Patterns

- Represents textures and local patterns
- Initially dedicated to texture classification and segmentation
- Compares intensity of a pixel with the intensity of neighboring pixels

Algorithm

1. Division of image into small regions
2. Computation of LBP patterns and histogram for each region
3. Concatenation of histograms into a feature vector

Local Binary Patterns

Advantages

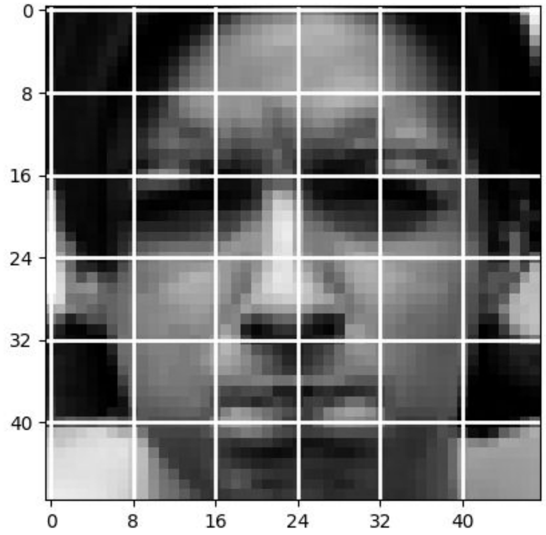
- Rotation invariance
- Gray-scale invariance (robustness to changes in illumination)

Drawbacks

- Focuses on local patterns
- May not capture global spatial information

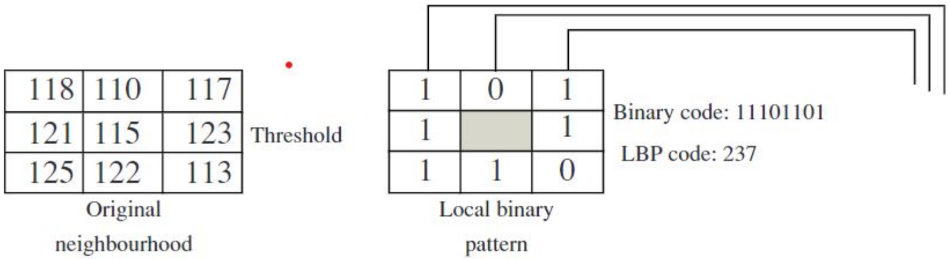
Linear Binary Patterns

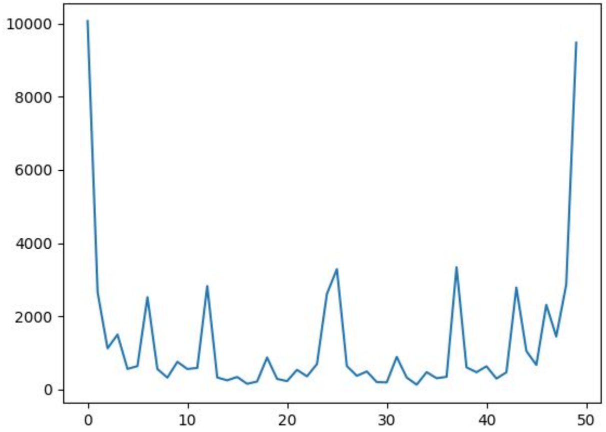
Division of image into small regions



Linear Binary Patterns

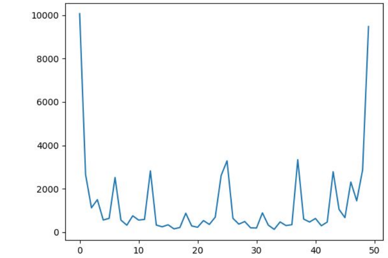
Computation of LBP and histogram for each region

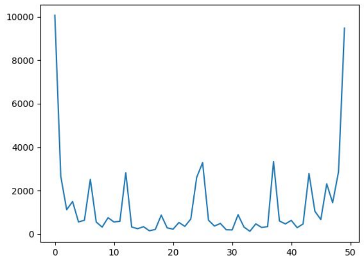


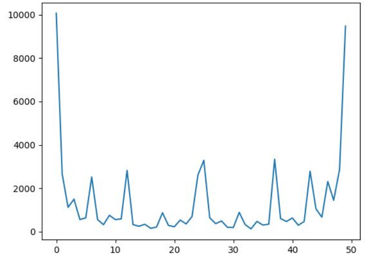


Linear Binary Patterns

Concatenation of histograms into a feature vector







[12, 34, 29, 4, 45, 0 …]

Linear Binary Patterns

Results

LBP + SVM	Mean	Standard Deviation
CK+	95.61%	0.61%
JAFFE	78.41%	2.72%

- Filtering out background might have improved results for JAFFE dataset.

Gabor filters

Gabor filters are commonly recognized as one of the best choices for obtaining localized frequency information.

They are interesting for their ability to capture texture patterns and variations in different facial regions.

A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image.

Algorithm

1. Creation of a filter bank
2. Image filtering
3. Concatenation of filtered images
4. Dimensionality reduction of concatenated filtered images

Gabor filters

Advantages

- Captures texture
- Sensible to orientations - captures facial features that may vary in orientation
- Good spatial localization - captures facial features at specific locations
- Robust to illumination change

Drawbacks

- Many parameters
- Computational complexity

Gabor filters

Gabor wavelets

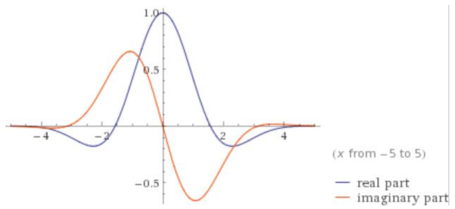
A wavelet is a wave-like oscillation with an amplitude that begins at zero, increases or decreases, and then returns to zero one or more times. Wavelets are termed a "brief oscillation".

The equation of a 1-D Gabor wavelet is a Gaussian modulated by a complex exponential, described as follows:

$$f(x) = e^{-(x-x_0)^2/a^2} e^{-ik_0(x-x_0)}$$

The equation of a 2-D Gabor filter is given by

$$f(x, y; \theta, \lambda, \sigma, \gamma, \psi) = e^{-\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma_x^2}\right)} e^{i\left(\left(2\pi \frac{x'}{\lambda} + \psi\right)\right)}$$



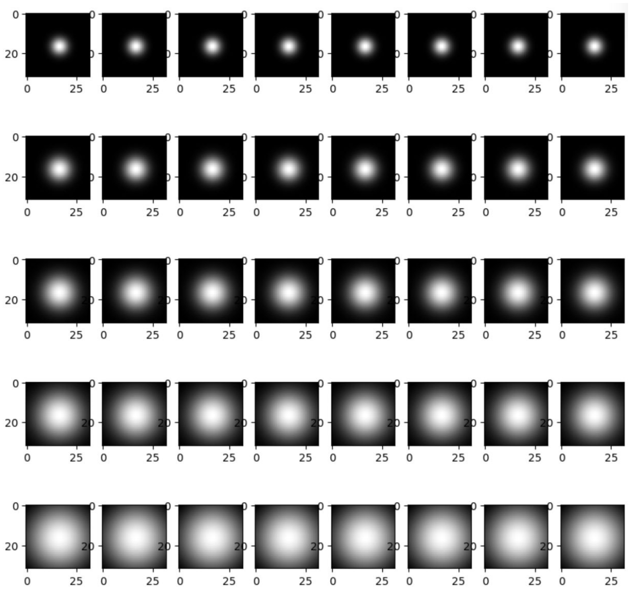
Where

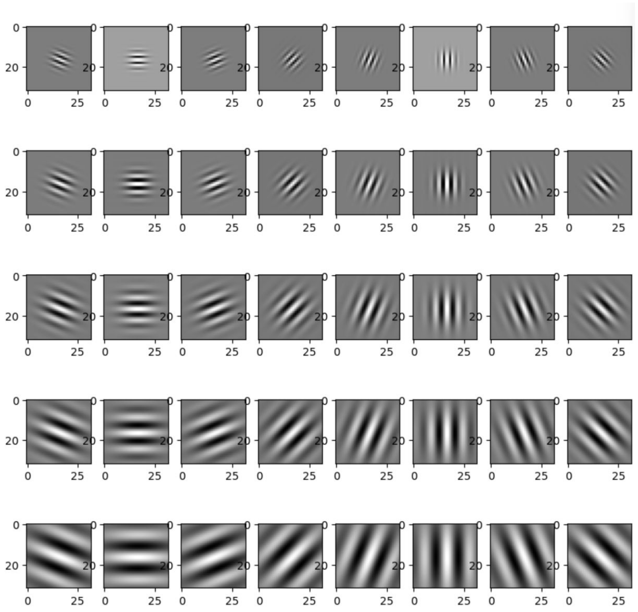
$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

Gabor filters

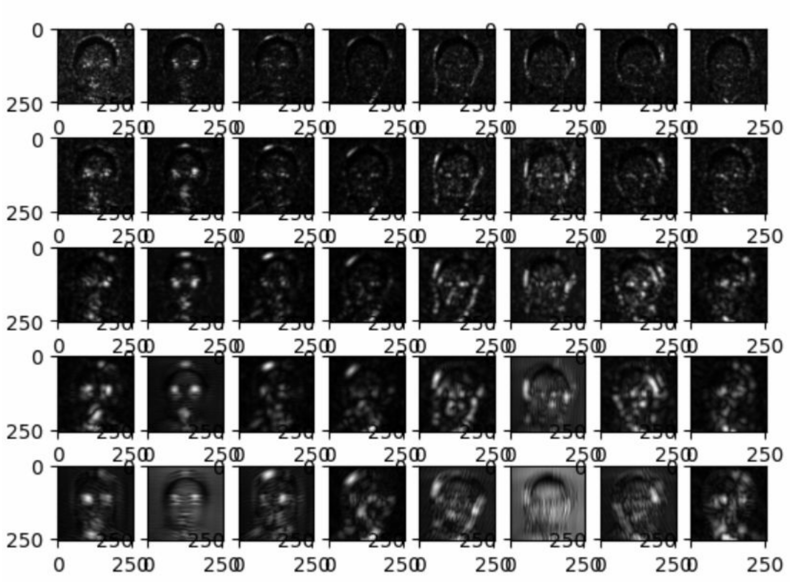
Bank of filters

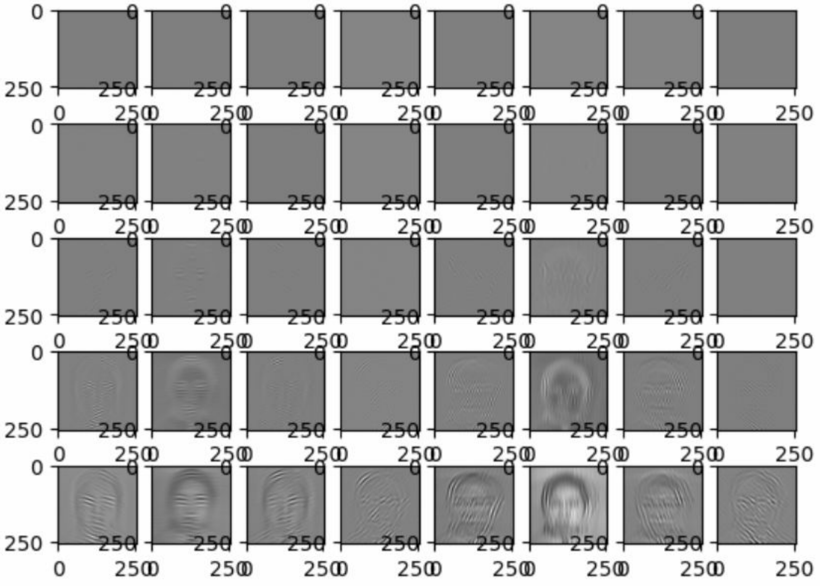




Gabor filters

Bank of filters





Gabor filters

Dimensionality reduction

- PCA on the concatenated filtered images.

In our work instead of doing dimensionality reduction with started by reducing the dimensionality of our images.

Gabor filters

Results

Gabor+ SVM	Mean	Standard Deviation
CK+	98.36%	1.29%
JAFFE	82.13%	3.22%

- Using traditional PCA instead of reducing the input images’ dimension perform better.
- Due to the complexity the best parameter search could not be exhaustive.

Conclusion

HOG + SVM

- HOG (Histogram of Oriented Gradients) focuses on capturing the local intensity gradients in an image, making it effective in representing facial features related to shapes and edges.

  > Good results on both datasets.

LBP + SVM:

- LBP is good at encoding texture information by analyzing the patterns formed by pixel intensities in local neighborhoods. It can perform well in scenarios where expressions involve texture-based variations, providing robustness against lighting changes.

  > Good result on CK+. JAFFE background might have had a negative impact on results.

Gabor + SVM

- Gabor filters are designed to capture both spatial and frequency information in an image, making them effective for analyzing facial features at multiple scales and orientations. Particularly powerful in capturing facial expressions that involve fine details and subtle changes in texture, as Gabor filters are sensitive to variations in both spatial and frequency domains.

  > Good results in both datasets.

Thank you for your “attention” !