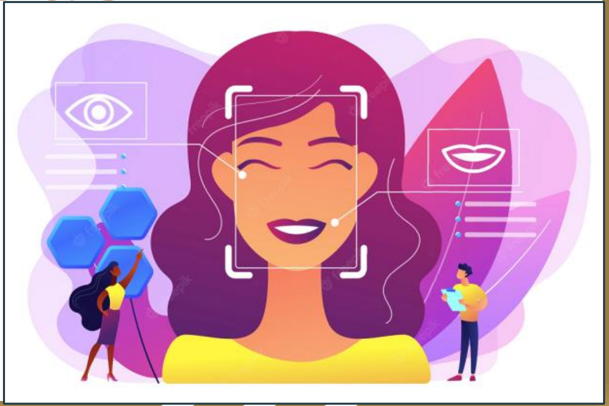
FINAL
Face Emotion Recognition from images - comparing methods.



Group 1: Mateusz Guściora

AGENDA

1. Problem formulation and goal
2. Theory and literature review and selected methods - *brief reminder*
3. Dataset and Environment
4. Algorithm procedure
5. Experiments and Evaluation
   a. what was focused on
   b. experiments
   c. evaluation
6. Conclusion and Future work
7. Literature

PROBLEM AND GOAL

Problem Formulation: *Recognizing Emotion from images and comparing methods.*

Goal: *Recognize and classify emotion from face images using feature extraction methods. Learn and craft those methods for recognizing, classifying emotions and compare them. Implement the system (or part) that will use such algorithms to recognize.*

Face Emotion Recognition

What is face emotion recognition?

★ (FER) - Face Emotion Recognition(or Facial Expression Recognition) - system, process, task of recognizing emotion from facial expressions.

★ Facial expression - a form of nonverbal signaling using the movement of facial muscles.[...] facial expression also reflects an individual’s emotional state. [1]

★ General approach to the problem:

Face emotion recognition can be divided into two types (based on approach) [2]:

- Traditional / Conventional approach
- Deep learning models (e.g. CNN)

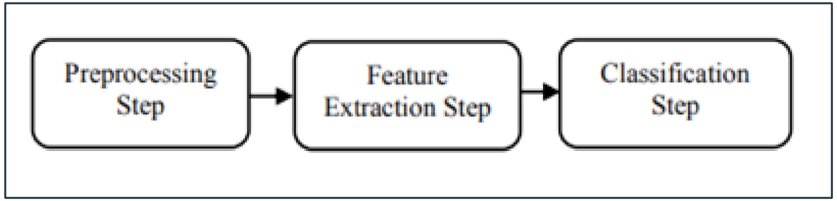


Figure 1: The three steps of facial expression recognition system [2]

FACE EMOTION RECOGNITION

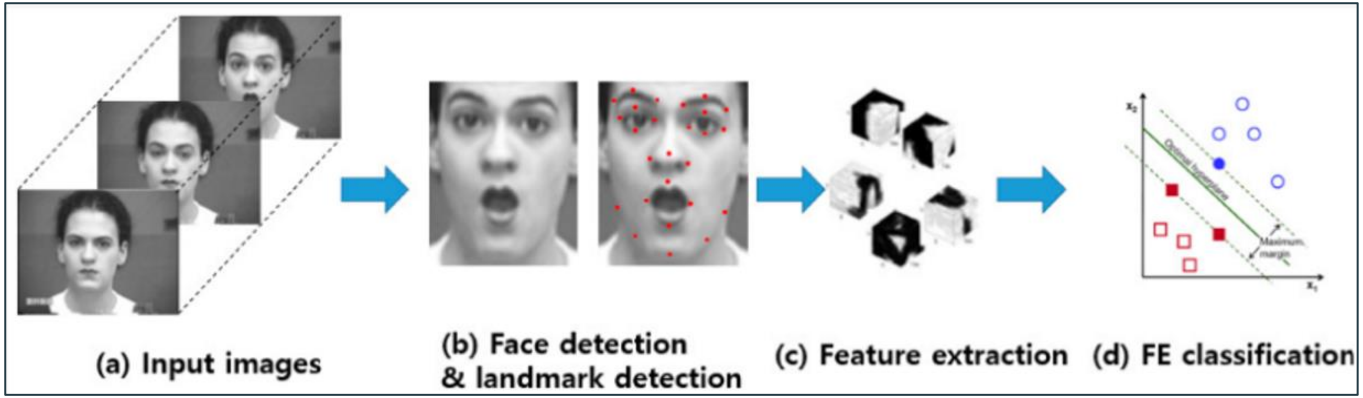


Figure2: Procedure used in conventional FER approaches [2]

← Conventional ML approach

Deep learning → approach (e.g. CNN)

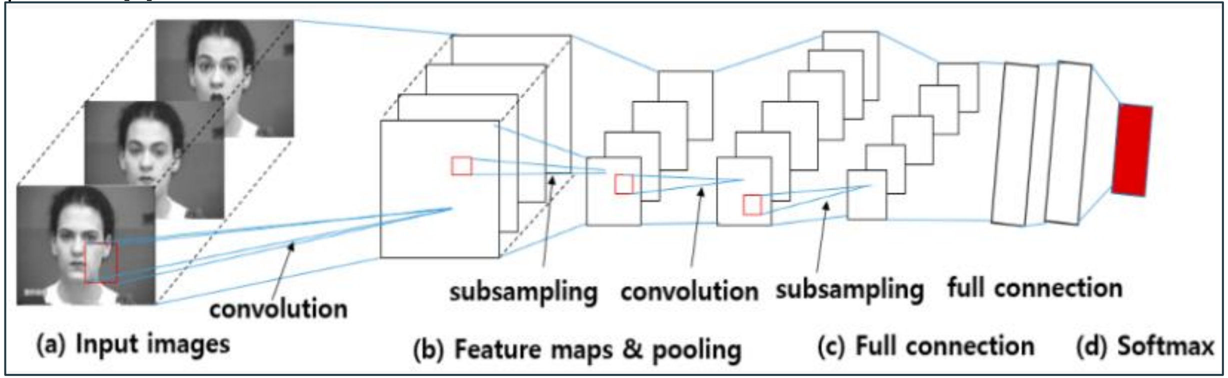
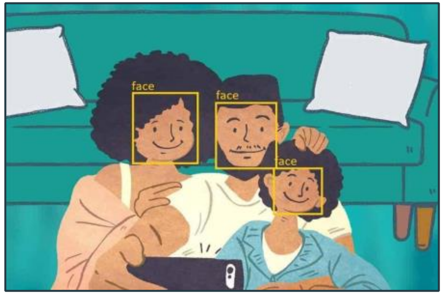


Figure3: Procedure of CNN-based FER approaches[2]

FACE DETECTION

- Is a technology or process that involves detecting and locating the human face.
- Is a necessary starting point for many face-related tasks, like facial landmark detection, gender classification, face tracking, face recognition, and face emotion recognition.



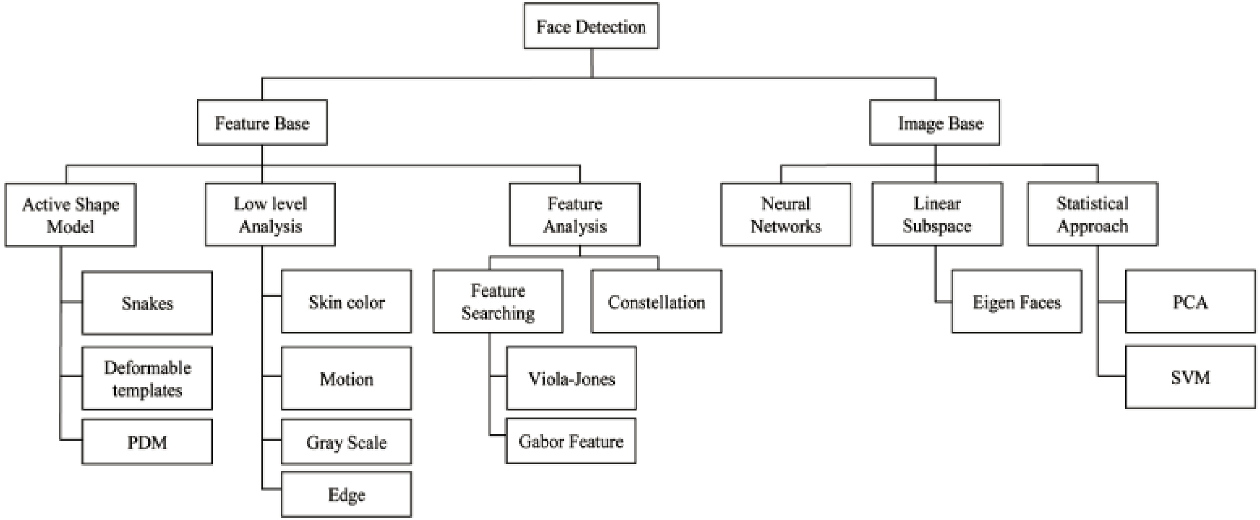


Figure: General face detection methods [3]

- Other challenges in preprocessing step such as: occlusion, complex backgrounds, rotations, illumination and orientation.

LOCAL BINARY PATTERN AND HIST.[4]

- ★ Into grayscale
- ★ Regions (each 3x3 pixels)
- ★ Intensity of each pixel 0-255
- ★ Threshold center value
- ★ Set into binary
- ★ Convert into decimal - center pixel
- ★ Repeat for each pixel in the image
- ★ New image with better characteristic
- ★ Regions - Grids
- ★ Produce histograms and later one concatenated histogram - feature histogram [3]

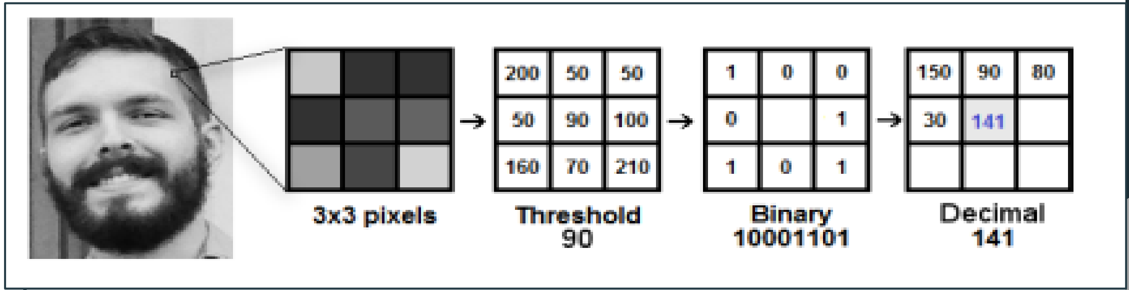


Figure4: LBP operator

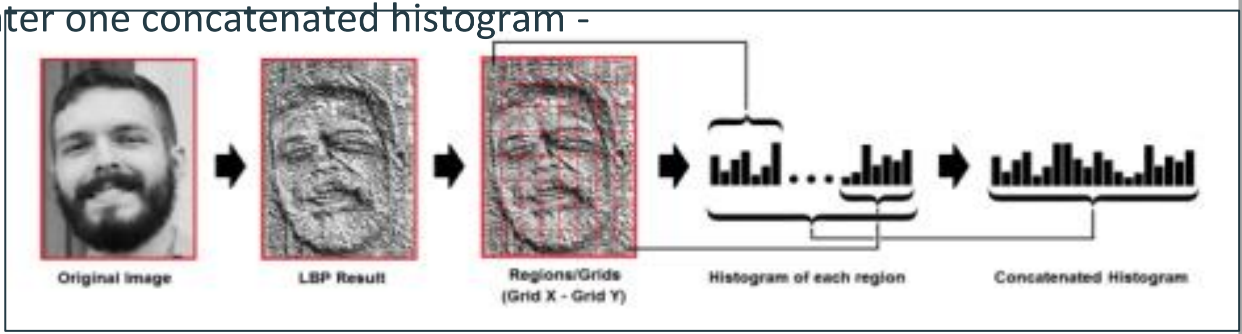


Figure5: LBP procedure

LBP- "*FACIAL EXPRESSION RECOGNITION BASED ON LOCAL BINARY PATTERNS*" *[5]*

Figure6: Face detecte
cropped

Figure7: Image divided into 80 regions

- ★ Preproc.: Cropped elliptical mask to exclude non-face area
- ★ Feature Extraction:
  - a) Divide the face image into 80 small regions
  - b) Calculate the LBP histogram of each region
  - c) Concatenate each region's LBP feature -histograms into a single feature vector (58 x 80)
- ★ Classification: decision tree

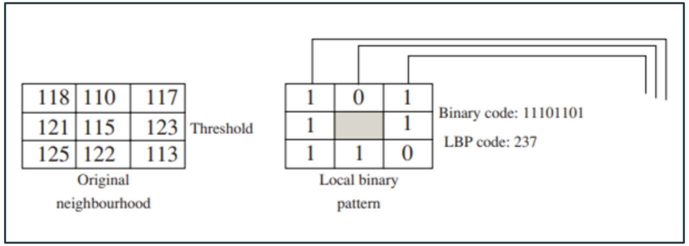


Figure7: LBP basic operator

HOG

The HOG feature descriptor comes from the last step of Lowe's scale invariant feature transformation of the SIFT algorithm. However, because of its effective expressive power, it has received widespread attention and has become an important application is that it is used as a feature descriptor in pedestrian detection. HOG descriptors and LBP operators have some similarities. They both belong to the differential mode information extraction method, and both reduce the influence of the gray changes due to the linear illumination anges.[6]

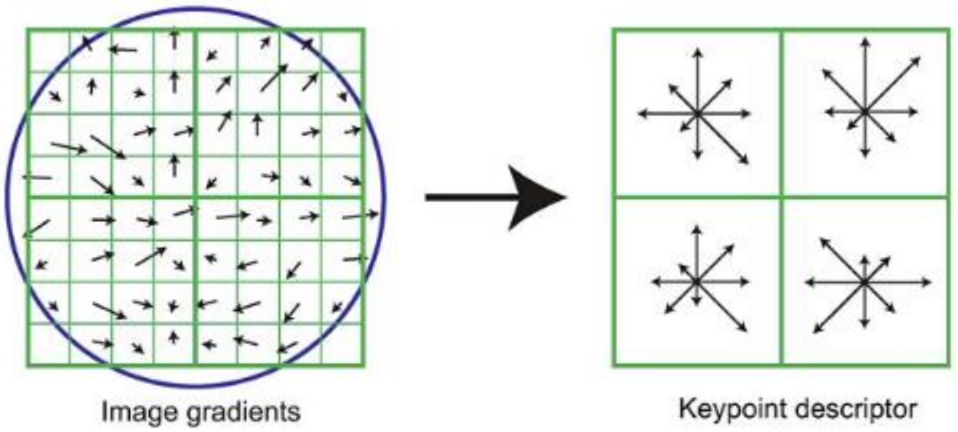


FIGURE 5. SIFT feature descriptor [28].

FIGURE 7. Face block division.

HOG [6]

Since the SIFT feature descriptor has good expressive power,Albiol et al. used it for the HOG-EBGM algorithm. They also used the 4×4=16 cell pattern proposed by Lowe.The difference is that each cell size of their HOG descriptor is 5×5 pixels and a sampling window has a total of 20×20 pixels, as shown in Figure 6.

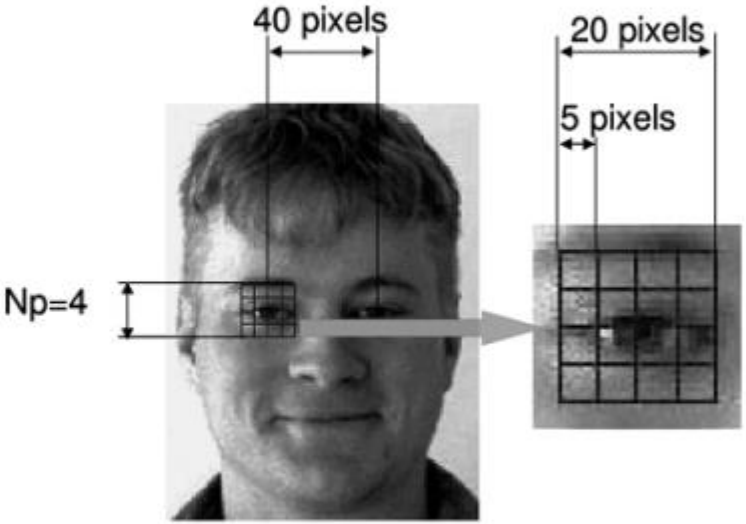


FIGURE 6. **The HOG-EBGM algorithm feature extraction window size**

HOG STEP BY STEP [6]

The face image is taken as the sampling window with 8×8 pixel neighborhood blocks. The blocks are spread all over the human face without overlap. The 80×64 pixel face image is divided into 10×8=80 small block forms, as shown in Figure 7.

In each block, the gradient direction and amplitude are calculated in the cell.

In each block, the histogram of the gradient direction is statistically graded according to the 4×4 pixel size.

the histogram of each cell in the block is connected to a vector.

Then, the histogram vectors in the block are normalized by the L2-norm.

After the L2-norm, the maximum value is limited to the threshold (such as 0.2), and then the L2-norm is standardized again.

inally, all normalized histogram vectors form a n×m matrix for representing the facial HOG features, where n is the histogram vector dimension of the block and m is the number of blocks to be computed for the entire face.

Then, the nearest neighbor classifier is used to recognize faces.

Gabor

The Gabor wavelet is one of the most popular feature description methods, and it is also one of the mainstream methods of facial descriptions. Gabor wavelets canwell simulate mammalian visual neurons and capture salient visual features. The Gabor wavelet can extract spatial and frequency-domain information from multiple scales and multiple directions, which can enlarge the difference between classes.14]

Gabor template

Let vand µ represent the scale and direction of the Gabor filter respectively. It usually has 5 scales ν∈{0,...,4}and 8 directions µ ∈{0,...,7}, as shown in Figure 1

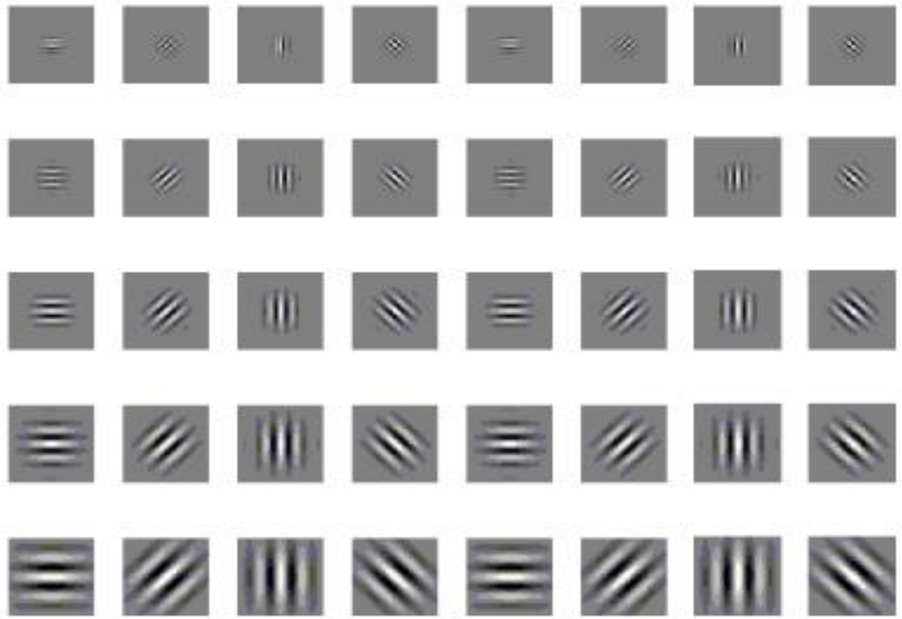


FIGURE 1. Five scales from top to bottom ($\nu \in \{0, \ldots, 4\}$) and 8 directions from left to right ($\mu \in \{0, \ldots, 7\}$) of the Gabor filter.

FEATURE EXTRACTION - “*IMAGE PROCESSING TECHNIQUES TO RECOGNIZE FACIAL EMOTIONS*” *[8]*

- ★ Preprocessing: Median, Wiener and Gaussian filters, Face detection and face cropping
- ★ Feature Extraction:

  Image segmentation, morphological operations

  Spatial filtering - masking

  Mouth area is calculated:

- ★ Classification of emotion: ‘Neutral, ‘Smile’, ‘Cry’

Facial Emotions	Mouth Area in mm	
	Min Value	Max Value
Neutral	3	4
Smile	11	17
Cry	5	10

Figure: Classification of emotions based on Mouth Area [13]

Fig 11. Mask Applied Image

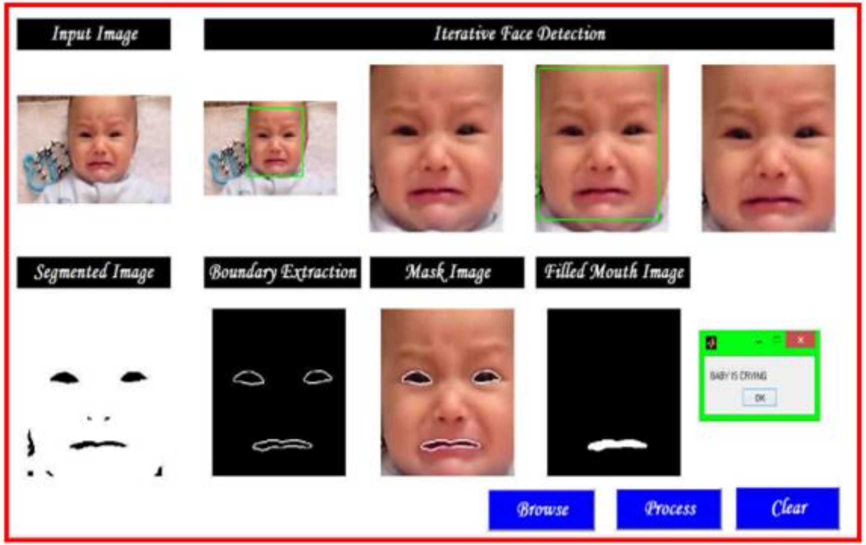


Figure: Procedure of recognizing and classifying emotion - cry[13]

COMPARISON BETWEEN LBP, HOG AND GABOR [9]

TABLE 2. Performance comparison of HOG, LBP and GFC face recognition methods in FERET face database.

Methods	Resolution	dimension	Test set (Recognition rate) (%)			
			fb	fc	dup1	dup2
LBP	80×64	4720	95.5	40.7	61.6	47.4
LBP (nonweighted) [18]	128×128	3304	93	51	61	50
LBP	160×128	18880	95.3	65.5	65.9	57.7
HOG	80×64	2560	92.6	72.2	66.3	61.1
GFC [33]	64×64	250	96.3	81.4	68.8	49.1

Figure8: Performance comparision HOG, LBP, Gabor

TABLE 3. Time comparison of the feature extraction time between the HOG, LBP and Gabor.

Feature extraction descriptor	Image resolution	Feature extraction time (ms)
LBP (8×8)	80×64	9.0
HOG	80×64	14.9
LBP (18×21)	128×128	20.2
LBP (8×8)	160×128	23.6
Gabor	64×64	258.2

Figure9: Time of feature extraction comparision

CNN - “*A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing*”

[10]

- ★ Preprocessing: a) Face detection - HAAR Cascade and Adaboost b) Scale Normalization c) Gray level equalization and Image edge detection
- ★ CNN Layers: Input, 3 convolution, 2 pooling, 1 full connection and 1 Softmax.
- ★ CNN: local perception, weight sharing and down sampling.
- ★ Classification: Softmax classifier, Seven facial expressions

$$y_j^l = \theta\left(\sum_{i=1}^{N_j^{l-1}} w_{i,j} \otimes x_i^{l-1} + b_j^l\right), \quad j = 1, 2, \cdots, M \qquad (16)$$

Figure11: Computational expression of convolution layer

$$\theta(x) = \max(0, x)$$

Figure12: Activation Function

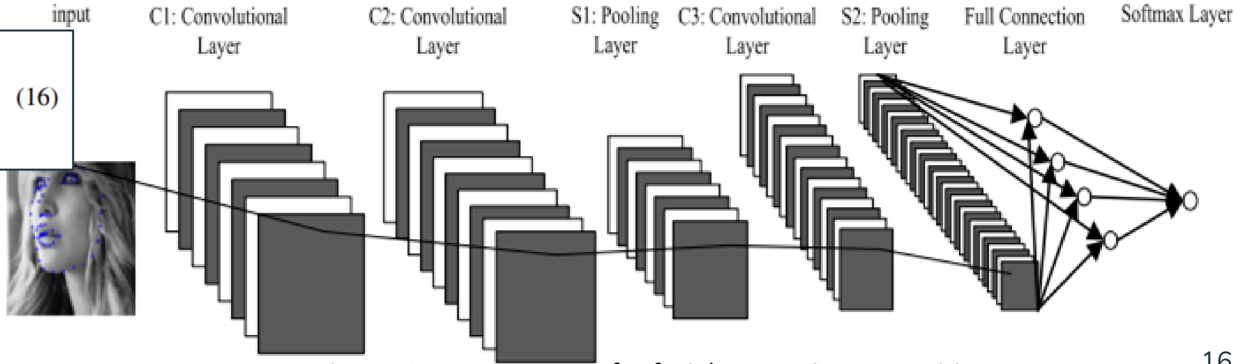


Figure10: CNN structure for facial expression recognition

“A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing” - Results [10]

METHOD	TRAINING TIME(S)	TEST TIME(S)	RECOGNITION RATE(%)
THE PROPOSED ALGORITHM	178	24.89	88.56
R-CNN ALGORITHM	256	33.97	79.34
FRR-CNN ALGORITHM	148	17.92	70.63

Figure13: Performance Comparison of three algorithms

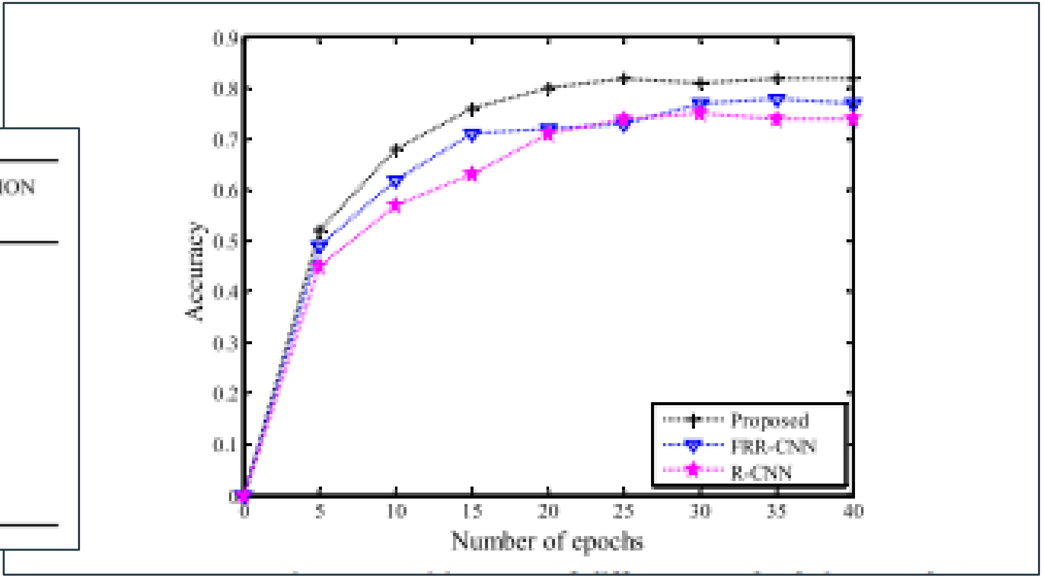
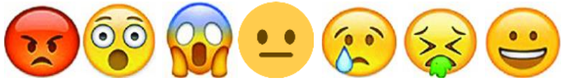


Figure14: Expression recognition rate of different methods in complex background

DATASET [11]



- ★ Train and Test (80%/20%) in each folders - ‘correct’ labels - emotions: angry, suprised, fearful, neutral, sad, disgusted, happy.
- ★ Image: Roughly preproc (cropped, gray, etc...), 48x48 pixel.

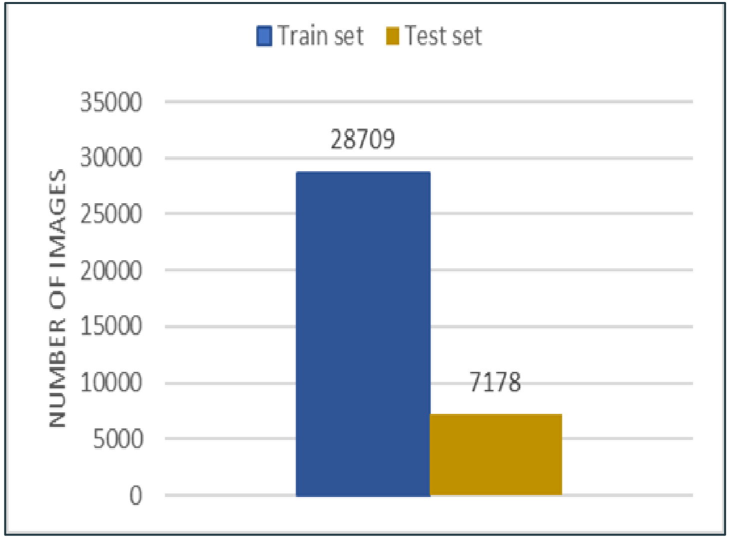


Figure15: Dataset, sets and number of images

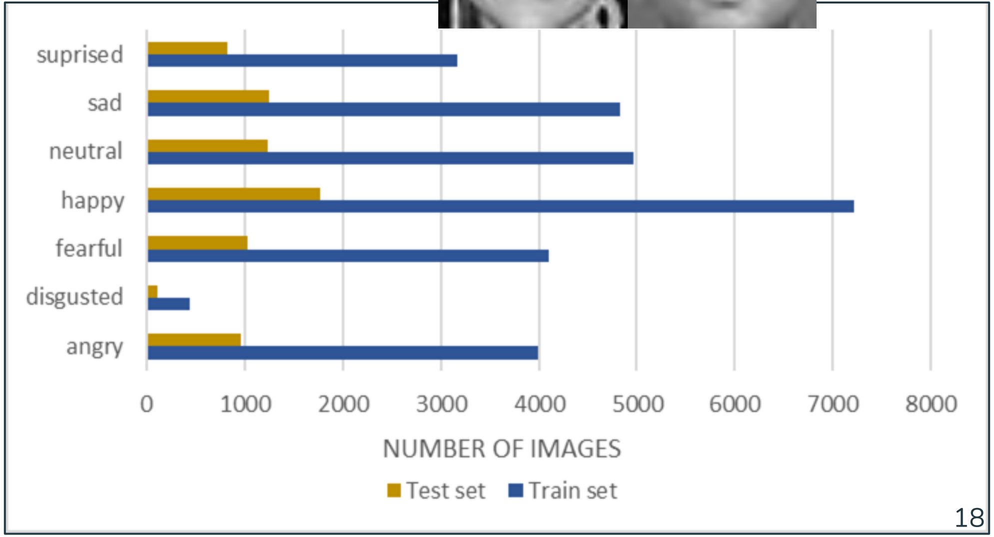


Figure16: Dataset, sets and number of images per label

RESEARCH ENVIRONMENT

Language and Environment : Python 3.9. and *VS*

Libraries: *sklearn, matplotlib, numpy, openCV*

HW configuration: CPU - *AMD Ryzen 7 4700U* with *Radeon Graphics 2.00 GHz*, RAM - *16GB*

Selected methods: handcrafted *LBP, HOG* (library), clf - *SVM.svc, SVM.linear svc, GausianNB*

LOCAL BINARY PATTERN - HANDCRAFTED ALGORITHM

Figure17: Original image

Figure18: LBP image

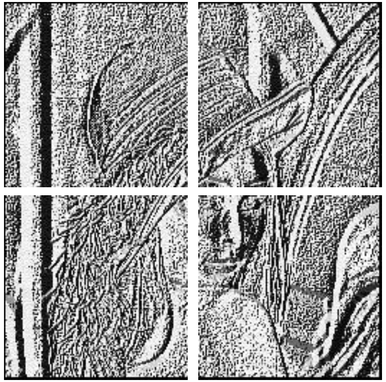


Figure19: grid of lbp image with tilesize parameter (here: 2,2)

[Calculate histograms for each tile and concatenated into one]

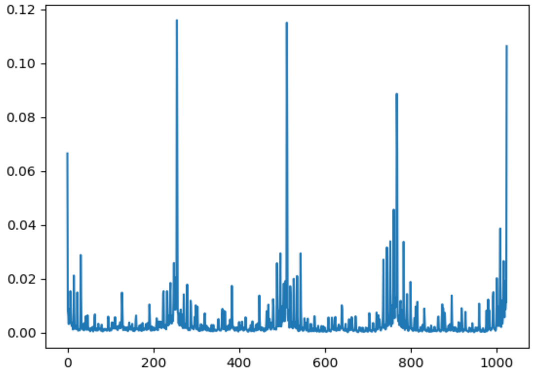


Figure20: Concatenated Histogram

EXPERIMENTS

what was focused on.

★ Measuring Accuracy and Visualize Confusion Matrix

★ Algorithm LBP handcrafted and HOG (scikit library) in practice

★ Parameter(C, random_state, dual,...) of SVM.svc and SVM.linearsvc

★ After normalizing histograms

★ Classifiers: NaiveBayesGaussian, svm.linearSVC, svm.SVC

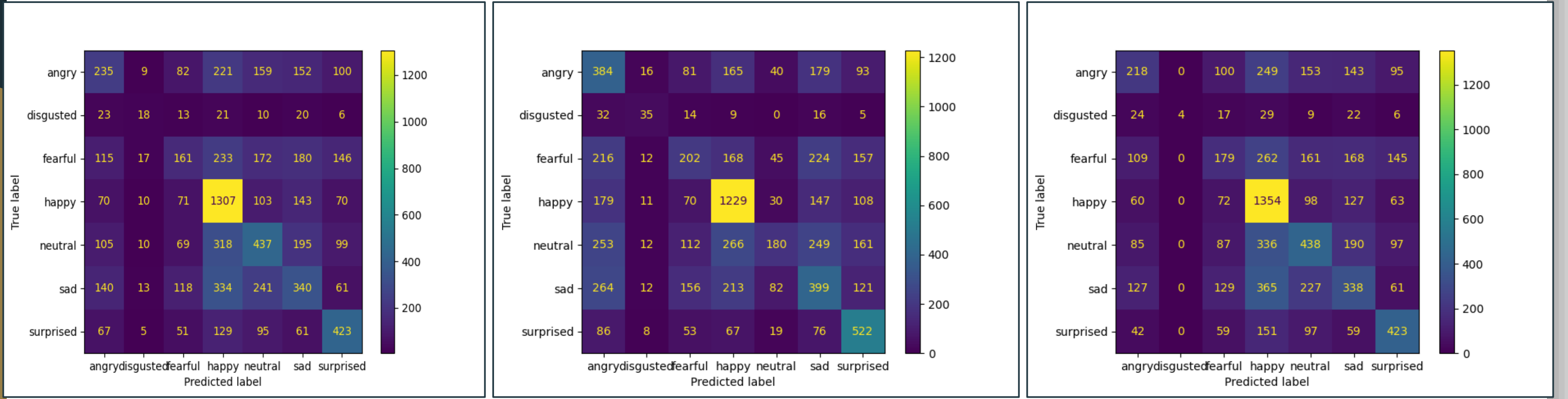
EXPERIMENTS

Algorithm	tileSize	Parameters	Classifier	AfterNorm (Y/N)	Accuracy
LBP handcrafted	(2,2)	C=100.0	SVM.LinearSvC	N	0.406
LBP handcrafted	(4,4)	C=100.0	SVM.LinearSvC	N	0.411
LBP handcrafted	(2,2)	C=100.0	SVM.svc	N	0.489
LBP handcrafted	(2,2)	C=10.0	SVM.svc	N	0.473
LBP handcrafted	(4,4)	C=10.0	SVM.svc	N	0,552
LBP handcrafted	(2,2)	-	NaiveBayes Gausian	N	0.313
LBP handcrafted	(2,2)	C=100.0	SVM.LinearSvC	Y	0.411
LBP handcrafted	(4,4)	C=100.0	SVM.svc	Y	0.547
LBP handcrafted	(4,4)	-	NaiveBayes Gausian	Y	0.304
HOG	(8,8) (2x2)		SVM.svc	-	0.568
HOG	(8,8) (2x2)		NaiveBayes Gausian	-	0.353

Tabel1: Comparision

CONFUSION MATRIX

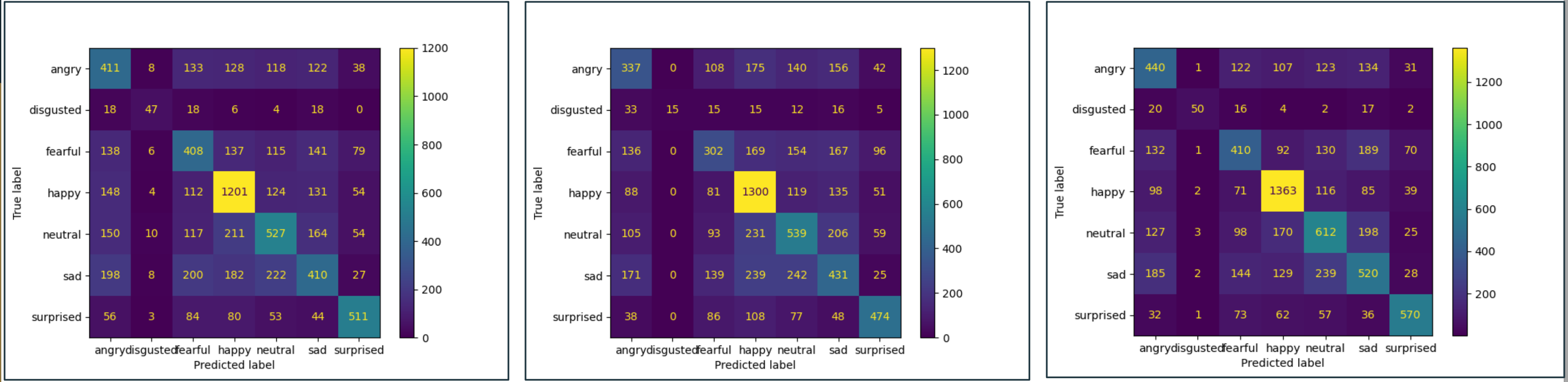
1. LBP handcrafted (2,2) C=100.0 SVM.LinearSvC, (accuracy 0.406)
2. LBP handcrafted (4,4) C=100.0 SVM.LinearSvC, (accuracy 0.411)
3. LBP handcrafted (2,2) C=100.0 SVM.LinearSvC, Norm (accuracy 0.411)



Figures20-23: Confusion map(true label vs predicted label)

CONFUSION MATRIX

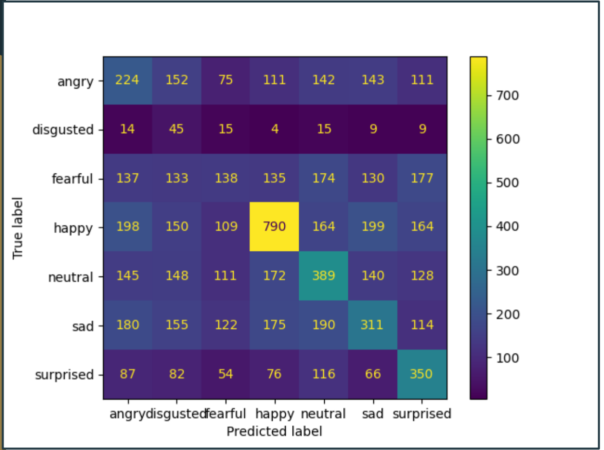
1. LBP handcrafted (2,2) C=100.0 SVM.svc (Accuracy 0.489)
2. LBP handcrafted (2,2) C=10.0 SVM.svc (Accuracy 0.473)
3. LBP handcrafted (4,4) C=10.0 SVM.svc (Accuracy 0,552)



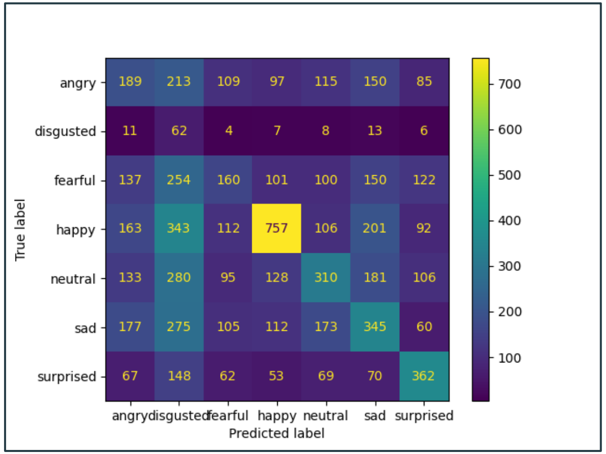
Figures24-27: Confusion map(true label vs predicted label)

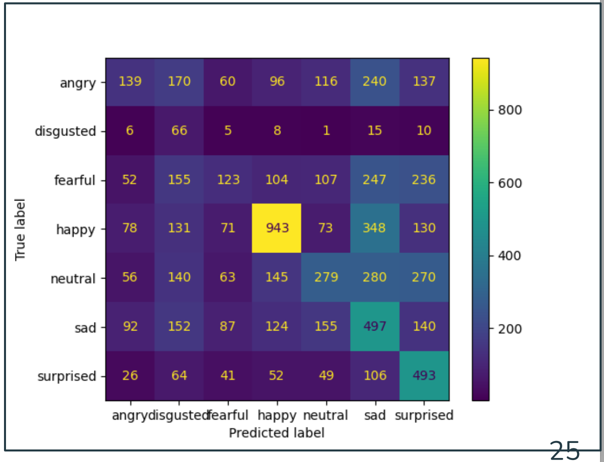
CONFUSION MATRIX

1. LBP handcrafted (2,2) - NaiveBayes Gausian (accuracy 0.313)
2. LBP handcrafted (4,4) - NaiveBayes Gausian norm (accuracy 0.304)
3. HOG (8,8) (2x2) NaiveBayes Gausian- (accuracy 0.353)



Figures24-27: Confusion map(true label vs predicted label)

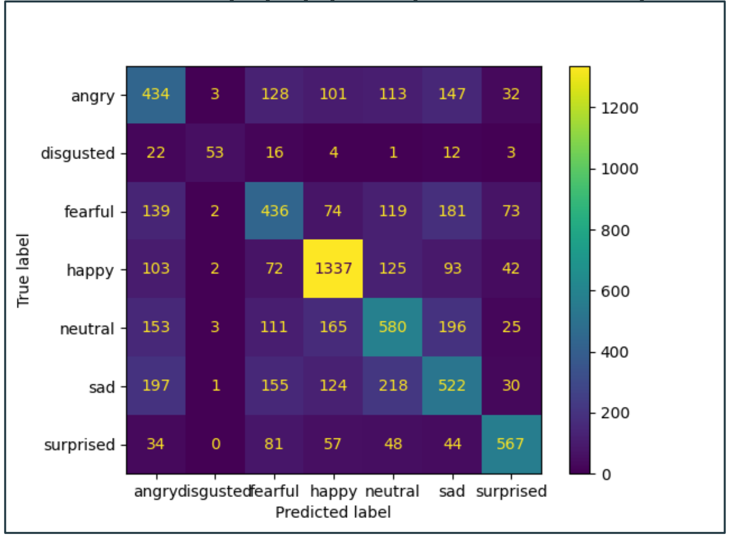


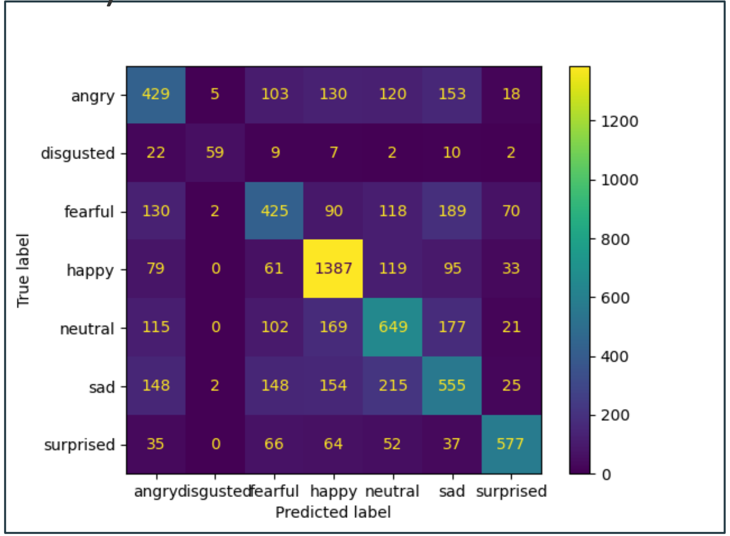


Figures28-30: Confusion map(true label vs predicted label)

CONFUSION MATRIX - BEST

1. LBP handcrafted(4,4) C = 100.0 SVM.svc (Accuracy 0.547)
2. HOG (8,8) (2x2) SVM.svc - (Accuracy 0.568)





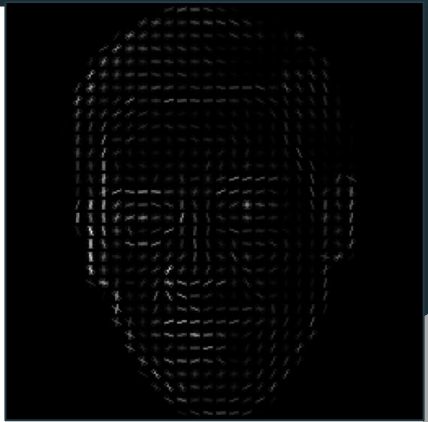
Figures31,32 : Confusion map(true label vs predicted label)

LBP TEST ON SAMPLE IMAGES

- ★ model works and can detect emotions (but not always correct!)
- ★ model tested on sample photos

Figures32-37: Predicted images with puttext function used best model

HOG - Algorithm



Library: *Scikit learn HOG*

Compute a Histogram of Oriented Gradients (HOG) by

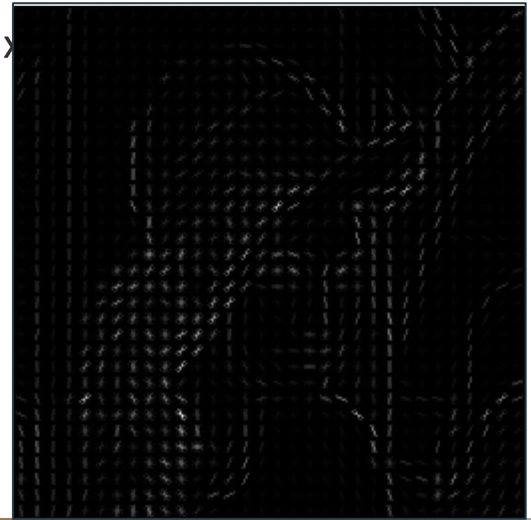
1. (optional) global image normalisation

he gradient image in

gradient histograms

across blocks

to a feature vector



Figures38-40: HOG algorithm in progress (image visualization)

CONCLUSION

- ★ Recognizing emotions is challenging task
- ★ There are plenty methods of feature extraction for image processing?
- ★ Preprocessing and feature extraction steps are very important in face image emotion recognition
- ★ Prepering such feature vector that classifier can best predict class is crucial
- ★ Best models can be used to recognize emotion from images of faces but not with great accuracy (possible improvments)

FUTURE WORK

- ★ improve preprocesing, feature extraction, better adjust input for ml classifiers
- ★ improve of loading dataset and optimizing (e.g. independent from size px)
- ★ different datasets
- ★ improving handcrafted algorithm LBP (e.g. add weighted)
- ★ comparing with LBP from library OpenCV (or others)
- ★ construct handcrafted algorithm HOG
- ★ clean code and code refactoring
- ★ more no. of experiments (e.g different algorithm param., different clf param.)
- ★ construct CNN
- ★ improve evaluation/visualization of future results

LITERATURE

[1] - https://dictionary.apa.org/facial-expression (date: 22.11.2022)

[2] - Byoung Chul Ko, A Brief Review of Facial Emotion Recognition Based on Visual Information, 2018, Sensors https://www.mdpi.com/1424-8220/18/2/401/htm

[3] - A Survey of Feature Base Methods for Human Face Detection Hiyam h, Zou B Raed., 2015

[4] - Human Emotion Detection using Machine Learning Techniques Punidha A, Inba. S, Pavithra. K.S., Ameer Shathali. M, Athibarasakthi. M Coimbatore Institute of Technology, Coimbatore

[5] - Facial Expression Recognition Based on Local Binary Patterns X. Fenga, M. Pietikäinenb, A. Hadidb, 2007

[6] - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8314117 (date: 27.01. 2022)

[] - gabor

[8]- Image Processing Techniques To Recognize Facial Emotions A. Mercy Rani, R. Durgadevi, 2019

[9] - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8489329

[10] - A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing Zhang H, Jolfaei A, Alazab M, 2019

[11] dataset -https://www.kaggle.com/datasets/ananthu017/emotion-detection-fer

LITERATURE

Other:

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=554193

https://www.researchgate.net/file.PostFileLoader.html?id=57ea918ef7b67eadd64fb462&assetKey=AS%3A410954002124806%401474990477987

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8489329

https://reader.elsevier.com/reader/sd/pii/S0167865509003055?token=F64E50AF47EAC25339FCEE486E32AD181F3902AFCD44754C3DADC58EC5045EB49DA0C24162248C63001352328F177E7E&originRegion=eu-west-1&originCreation=20221108142414

https://www.aimspress.com/fileOther/PDF/MBE/mbe-17-02-082.pdf

¿ QUESTIONS?

THANK YOU!