Table of content

1. Topics introduction
2. What was already tried
3. Proposed Architecture
    a. preprocessing
    b. feature extraction
    c. selection
    d. classification
4. Datasets
5. References

Topics Introduction

Emotion detection with super resolution

Goals:

- Low quality images or video frames
  - low resolution
  - blur
  - noise
- Classify each frame
- Ekman emotions (anger, disgust, fear, happiness, sadness, surprise) + contempt
- Measurements

Topic narrowings:

- Single face at a time
- Single modal
  - no audio (speech)
  - no hand gesture
  - no background
- Face focused

What was already tried

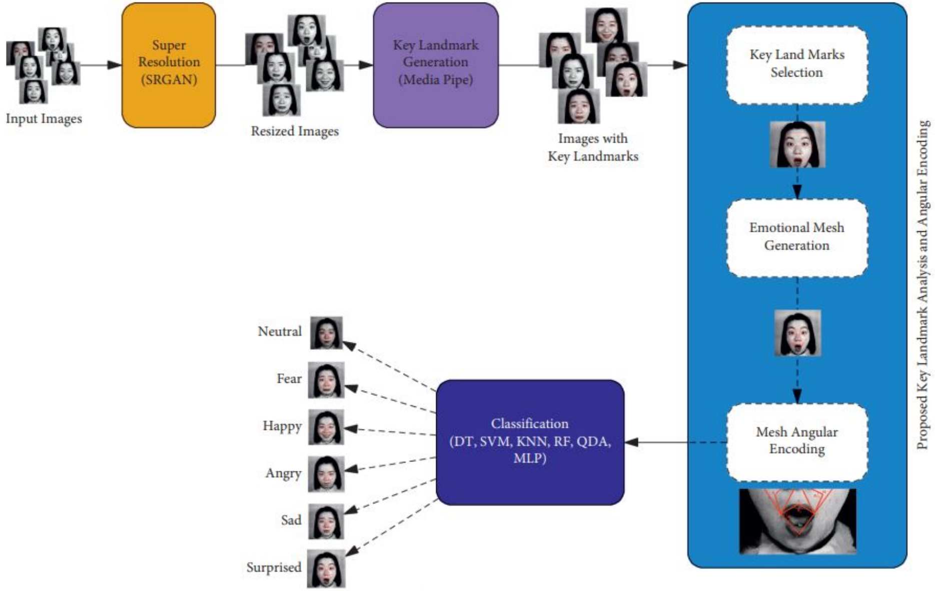
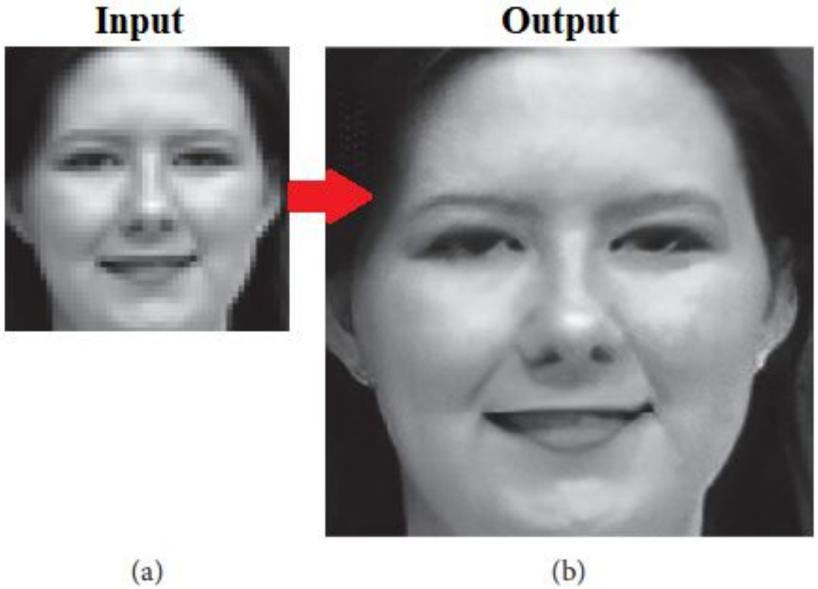

Input Images
Super Resolution (SRGAN)
Resized Images
Key Landmark Generation (Media Pipe)
Images with Key Landmarks
Key Land Marks Selection
Emotional Mesh Generation
Mesh Angular Encoding
Proposed Key Landmark Analysis and Angular Encoding
Classification (DT, SVM, KNN, RF, QDA, MLP)
Neutral
Fear
Happy
Angry
Sad
Surprised

FIGURE 4: The proposed framework.



Preprocessing step (super-resolution): (a) original image [48 × 48]; (b) 4*x* upscaled super-resolved image using SRGAN [192 × 192].

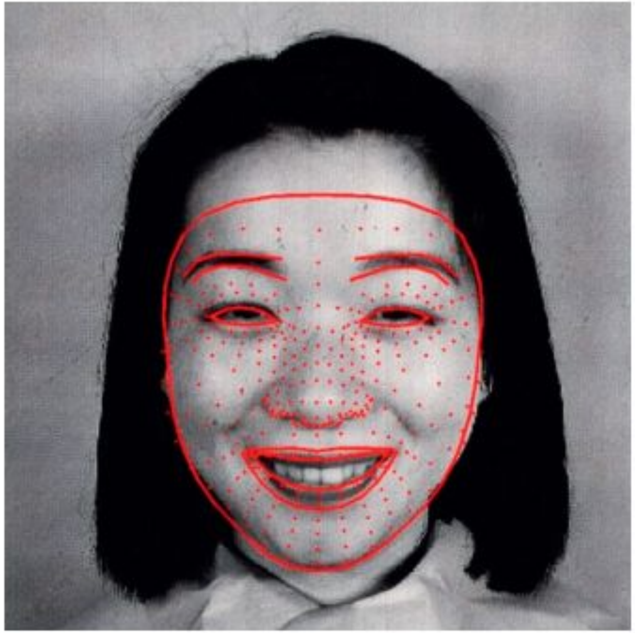


Figure 7: An image with 468 annotated landmarks using MediaPipe face mesh.

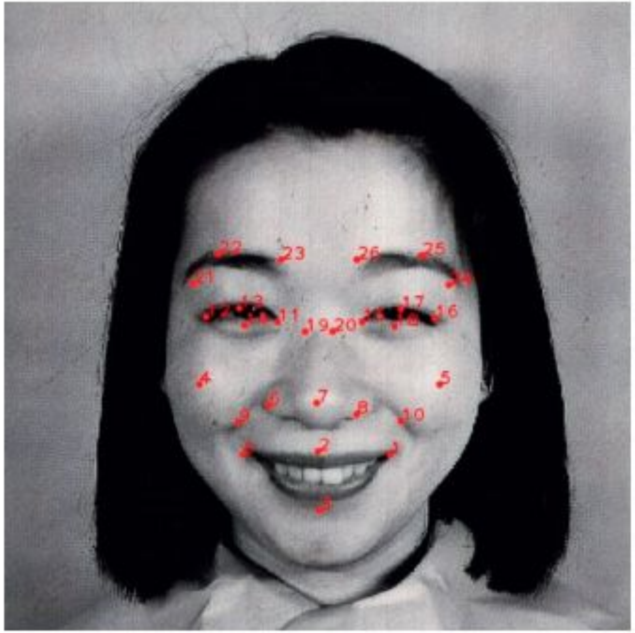


Figure 8: The 27 key landmarks and their locations.

Deformation of emotion face mesh measured by the deviation of angles between edges reflects facial muscle contraction and relaxation, which will be used to identify facial emotions

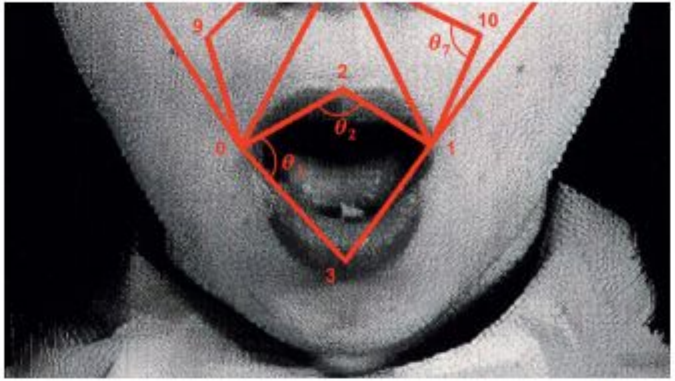


Figure 10: An example showing the features (θ_1, θ_2, and θ_7) and their locations.

Table 7: Comparison of eight ML classifiers on CK+ dataset.

Classifier	Classification accuracy	Precision	Recall	*F*1-score	Training time (seconds)
Gaussian NB	0.84	0.84	0.84	0.84	0.005
QDA	0.86	0.85	0.86	0.85	0.006
DT	0.86	0.85	0.86	0.85	0.018
LR	0.87	0.86	0.87	0.86	0.15
RF	0.89	0.90	0.89	0.88	0.74
MLP	0.94	0.94	0.94	0.94	1.82
SVM	0.94	0.94	0.94	0.94	0.06
KNN	0.97	0.97	0.97	0.97	0.005

Table 8: Comparison of eight ML classifiers on JAFFE dataset.

Classifier	Classification accuracy	Precision	Recall	*F*1-score	Training time (seconds)
Gaussian NB	0.90	0.93	0.90	0.91	0.005
QDA	0.79	0.80	0.79	0.79	0.005
DT	0.90	0.91	0.90	0.90	0.005
LR	0.86	0.90	0.86	0.86	0.09
RF	0.93	0.94	0.93	0.93	0.33
MLP	0.90	0.94	0.90	0.91	0.41
SVM	0.88	0.91	0.88	0.88	0.008
KNN	0.95	0.97	0.95	0.95	0.002

Table 9: Comparison of eight ML classifiers on RAF-DB.

Classifier	Classification accuracy	Precision	Recall	*F*1-score	Training time (seconds)
Gaussian NB	0.65	0.62	0.65	0.61	0.01
QDA	0.64	0.62	0.64	0.61	0.02
DT	0.53	0.52	0.53	0.52	0.26
LR	0.66	0.62	0.66	0.63	0.9
RF	0.65	0.61	0.65	0.62	11
MLP	0.67	0.64	0.67	0.64	20
SVM	0.67	0.64	0.67	0.64	24.5
KNN	0.63	0.60	0.63	0.61	0.053

Table 10: Comparison between the proposed work and the state-of-the-art works.

Work	Year	Method	Accuracy (%)			Time (s)
			JAFFE	CK+	RAF	
[40]	2018	CNN	50.12	93.64		—
[28]	2019	mSVM	88.95	91.98	65.12	—
		LDA	83.45	92.33	56.93	—
[41]	2018	RF	—	93.4	—	—
[42]	2018	SVM	—	95.8	—	—
[43]	2018	AlexNet	93	90.2	—	—
		VGG16	96	92.4	—	0.94
[44]	2018	VGG19	93	93	—	—
[45]	2021	CNN	96.8	86.5	—	62.5
[46]	2018	AlexNet	—	—	55.6	—
		VGG	—	—	58.2	—
Proposed	**2021**	**MLP**	**90**	**94**	**67**	**1.12**
		SVM	**88**	**94**	**67**	**0.034**
		KNN	**95**	**97**	**63**	**0.004**
		LR	**86**	**87**	**66**	**0.12**

Problems with KNN

- Label flickering in video classification
- Mostly good at categorizing emotions at their peak level (not as good at mid and subtle emotion detection)

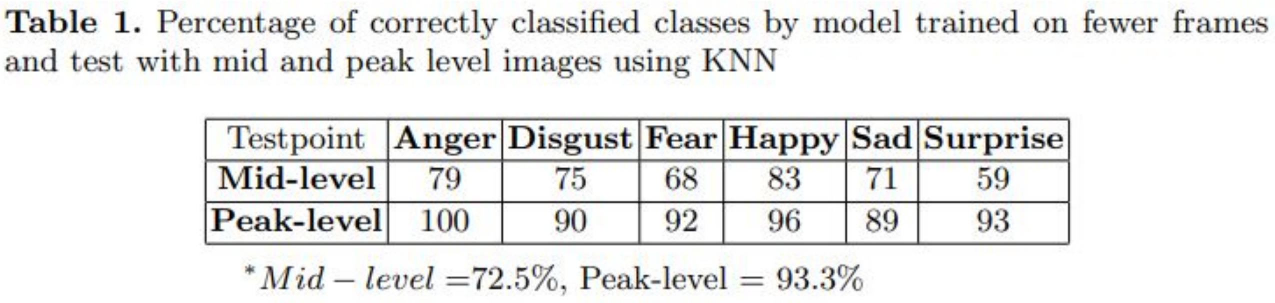


Table 1. Percentage of correctly classified classes by model trained on fewer frames and test with mid and peak level images using KNN

Testpoint	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Mid-level	79	75	68	83	71	59
Peak-level	100	90	92	96	89	93

$^{*}Mid - level$ =72.5%, Peak-level = 93.3%

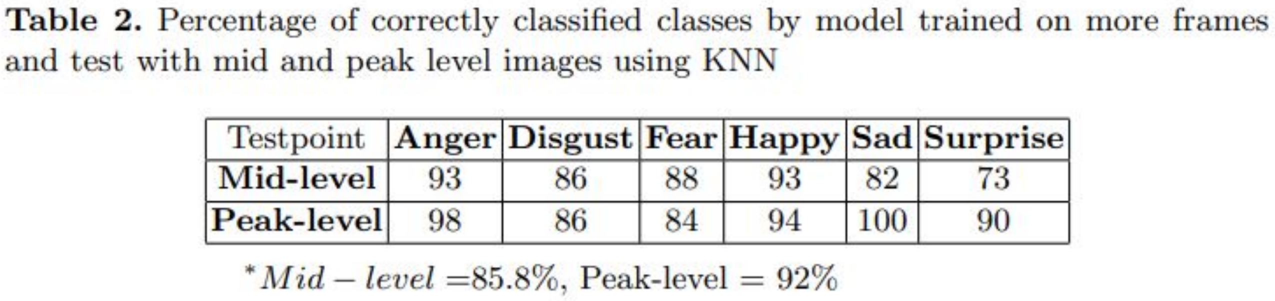
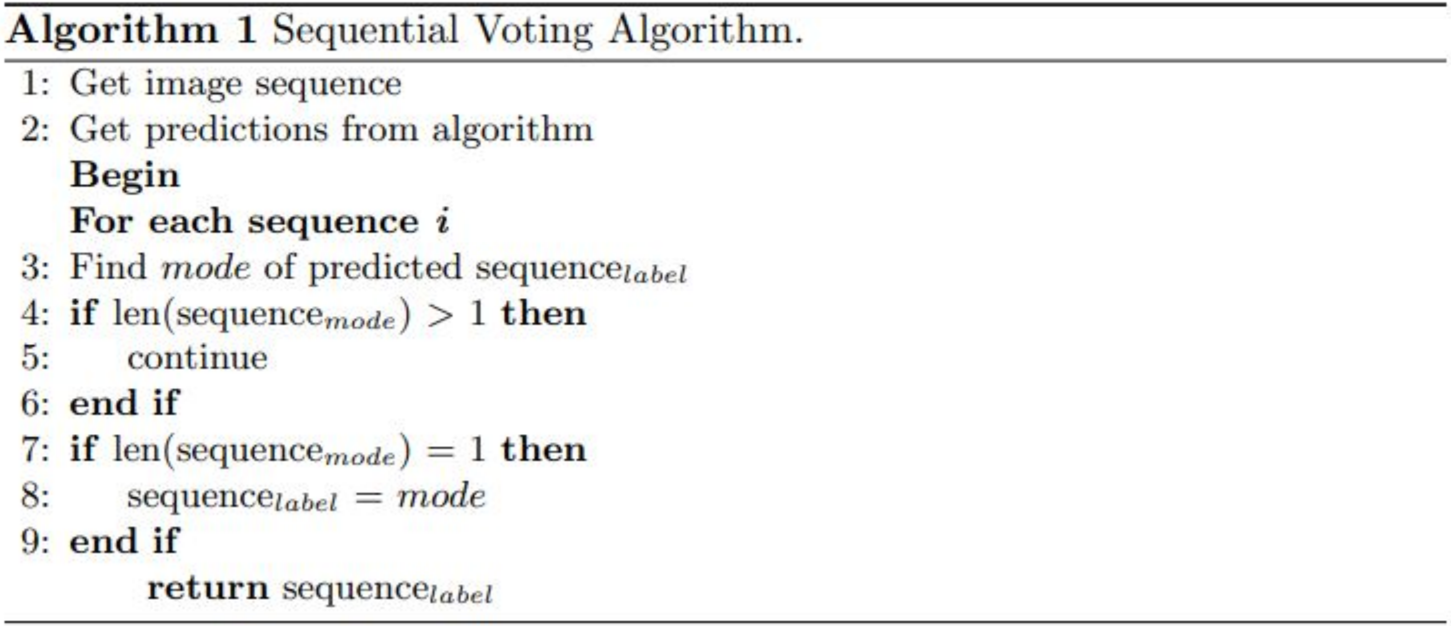


Table 2. Percentage of correctly classified classes by model trained on more frames and test with mid and peak level images using KNN

Testpoint	**Anger**	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**
Mid-level	93	86	88	93	82	73
Peak-level	98	86	84	94	100	90

$^{*}Mid - level$ =85.8%, Peak-level = 92%

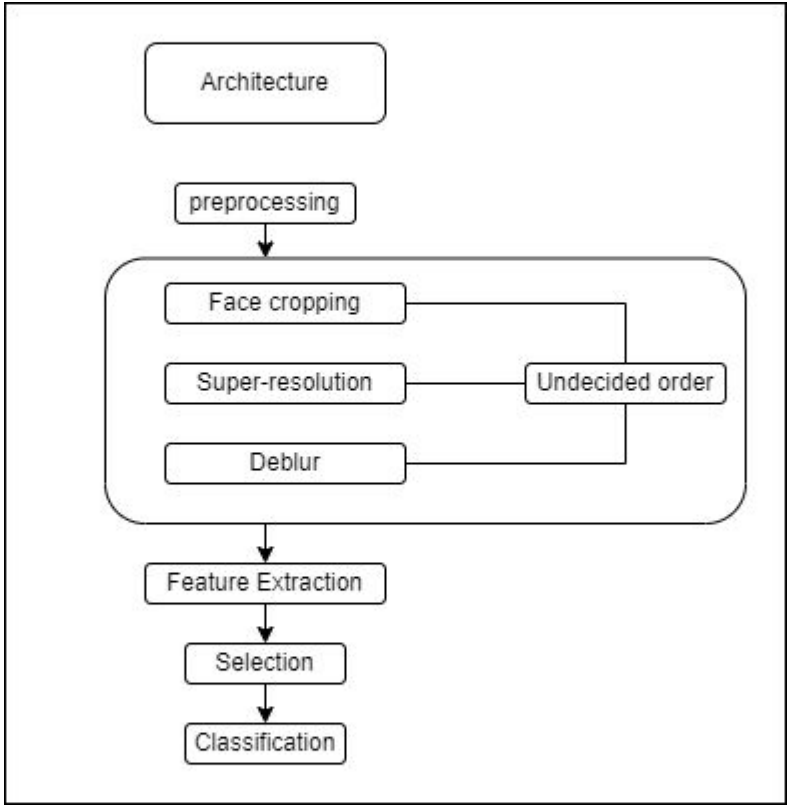


Algorithm 1 Sequential Voting Algorithm.

```
Get image sequence
Get predictions from algorithm
Begin
For each sequence $i$
Find $mode$ of predicted sequence$_{label}$
if len(sequence$_{mode}$) $> 1$ then
    continue
end if
if len(sequence$_{mode}$) $= 1$ then
    sequence$_{label}$ $= mode$
end if
    return sequence$_{label}$
```

Table 5. Prediction accuracy on six basic emotions of the CK+ database before and after sequential voting (+SV)

Expressions	Algorithms			
	RF	RF + SV	KNN	KNN + SV
An	94	99	94	98
Di	88	95	90	93
Fe	90	96	92	95
Ha	95	98	95	98
Sa	92	100	93	99
Su	84	91	85	93
Standard dev.	0.18	0.21	0.0	0.0
Average acc.	90.5±0.1	96.5±0.1	91.5	96.0



Architecture

preprocessing

Face cropping

Super-resolution

Deblur

Undecided order

Feature Extraction

Selection

Classification

Preprocessing

Super resolution:

- SRGAN

The selected super resolution method should minimize the Mean SquareError (MSE) and maximize the Peak Signal-to-Noise Ratio (PSNR)

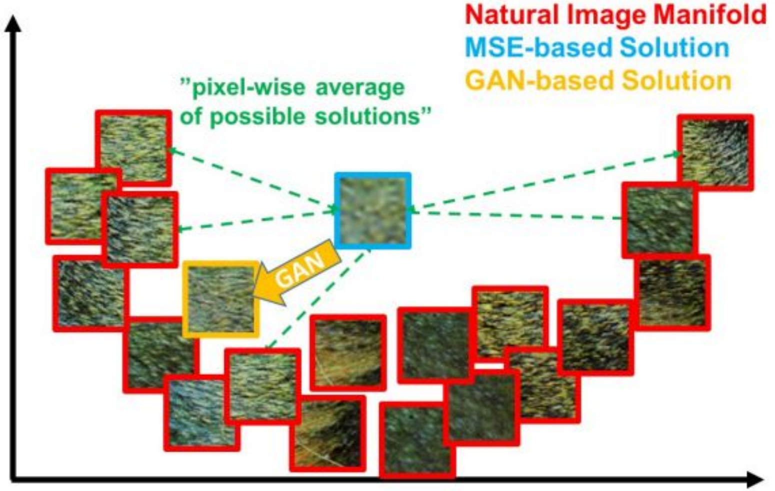


Figure 3: Illustration of patches from the natural image manifold (red) and super-resolved patches obtained with MSE (blue) and GAN (orange). The MSE-based solution appears overly smooth due to the pixel-wise average of possible solutions in the pixel space, while GAN drives the reconstruction towards the natural image manifold producing perceptually more convincing solutions.

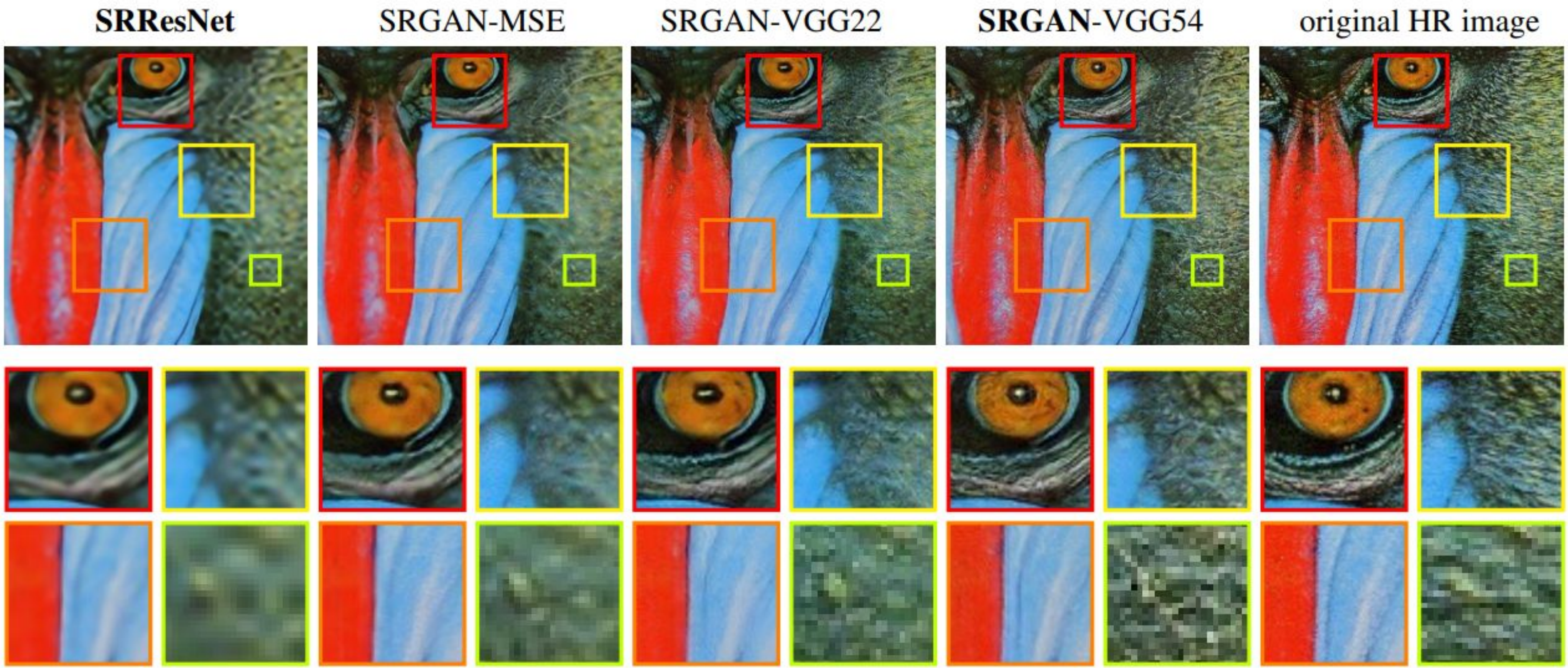


Figure 6: **SRResNet** (left: a,b), SRGAN-MSE (middle left: c,d), SRGAN-VGG2.2 (middle: e,f) and **SRGAN**-VGG54 (middle right: g,h) reconstruction results and corresponding reference HR image (right: i,j). [4× upscaling]

Face cropping (face recognition):

- FaceNet
- OpenFace
- MediaPipe BlazeFace
- VGG-Face
- DeepFace

Feature extraction (mash of the face)

- MediaPipe Face Mesh
- OpenFace
- Dataset contains it

Classification:

- KNN

Dataset

- CMU-MOSEI Dataset (6 base emotions = Ekman emotions)
  - videos
- CK+ (Extended Cohn-Kanade dataset)
  - Ekman emotions (anger, disgust, fear, happiness, sadness, surprise) + contempt
  - 593 video sequences
  - facial shift from the neutral expression to a targeted peak expression
  - resolution of either 640x490 or 640x480 pixels
  - 30 frames per second (FPS)
- DEMoS (Ekman emotions)

proposed timeline, pipeline

- ~2 weeks
  - putting the project together
  - GitHub (README)
  - research diary
- ~1 week
  - testing measurements
  - comparisons
- ~1 week for unexpected complications

References

[1] Ali I. Siam, Naglaa F. Soliman, Abeer D. Algarni, Fathi E. Abd El-Samie, Ahmed Sedik,
"Deploying Machine Learning Techniques for Human Emotion Detection",
Computational Intelligence and Neuroscience,
vol. 2022, Article ID 8032673, 16 pages, 2022.
https://doi.org/10.1155/2022/8032673

[2] Shehu, H. A., Browne, W. & Eisenbarth, H. (2020).
Emotion Categorization from Video-Frame Images Using a Novel Sequential Voting Technique.
Advances in Visual Computing (12510 LNCS, pp. 618-632).
Springer International Publishing. https://doi.org/10.1007/978-3-030-64559-5_49

[3] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network
arXiv:1609.04802
https://arxiv.org/abs/1609.04802

Thank you!