Facial Emotion Recognition in Real-Time

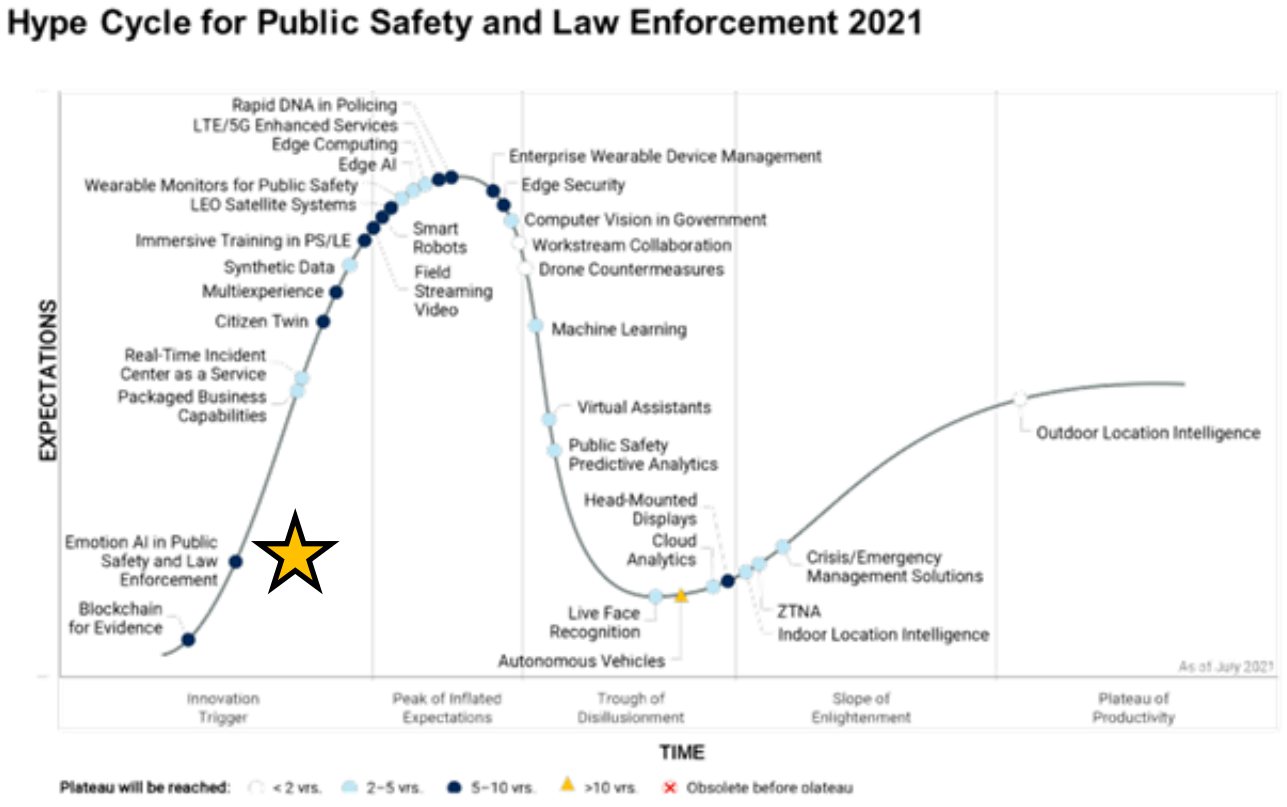
By Amira (DSI-28)
7 Jul 2022

Table of Contents

01

Context & Problem

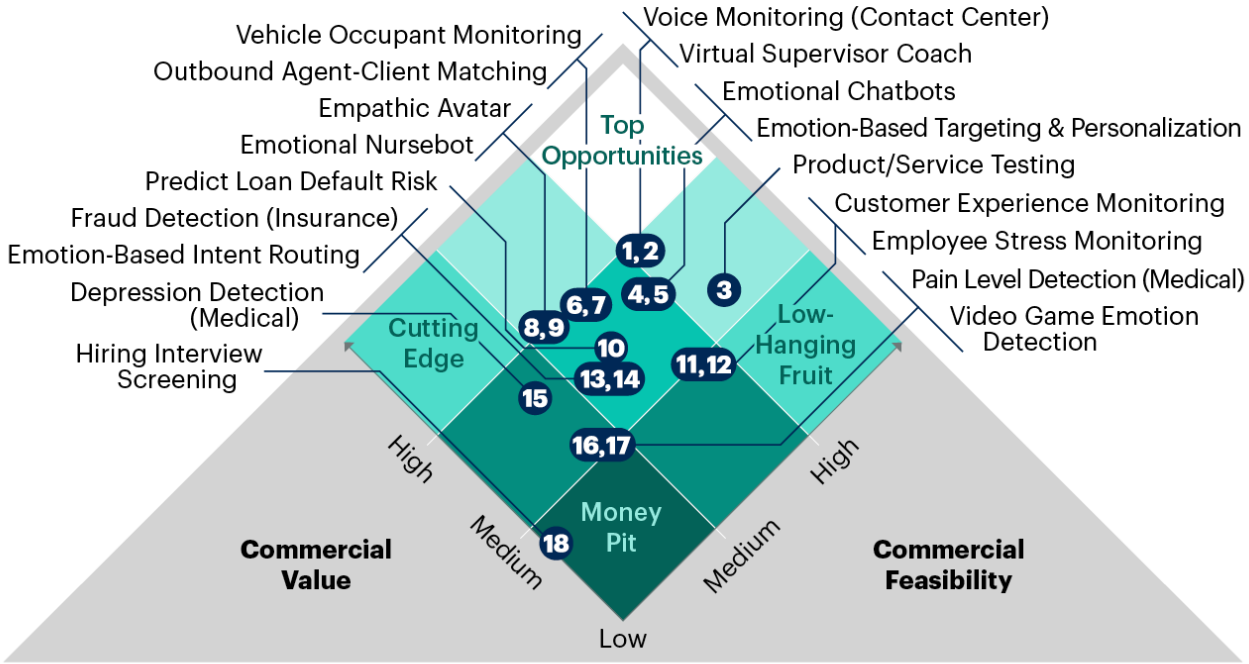
Rising Interest in/Potential of Emotion AI for Law Enforcement



Source: Gartner (July 2021)

In a wider context ...

Emotion AI Technology Opportunity Prism



Source: Gartner

741689_C

Gartner

Why Emotion AI for Law Enforcement?

Human Interaction is Our Core Domain

Digital Interaction is the Present & Future

Keeping Pace with Digital Innovation

Focus on Detecting Emotions Through Facial Expression

Components of Emotion AI

Proposed Solution

- **Test data architectures of pre-trained Convolutional Neural Network models**
  - Baseline model: Custom CNN model
  - Pre-trained models: ResNet-50, MobileNetV2, VGG16, VGGFace
- **Evaluation Criteria:**
  - Accuracy scores on unseen data
  - F1-scores of each emotion class on unseen data

02

Data & Insights

FER-2013 Dataset

- ~ 28k labelled images in train set and ~3.5k in test set
- Posed and unposed headshots in grayscale 48 x 48 pixels
- 7 emotion classes: angry, disgust, fear, happy, neutral, sad, surprise
- Created using Google image search of each emotion and synonyms of the emotions
- Well-explored dataset in the data science community

Class Imbalance Especially with Happy and Disgust

Our Dataset is Not Perfect

Occlusions

Unrealistic

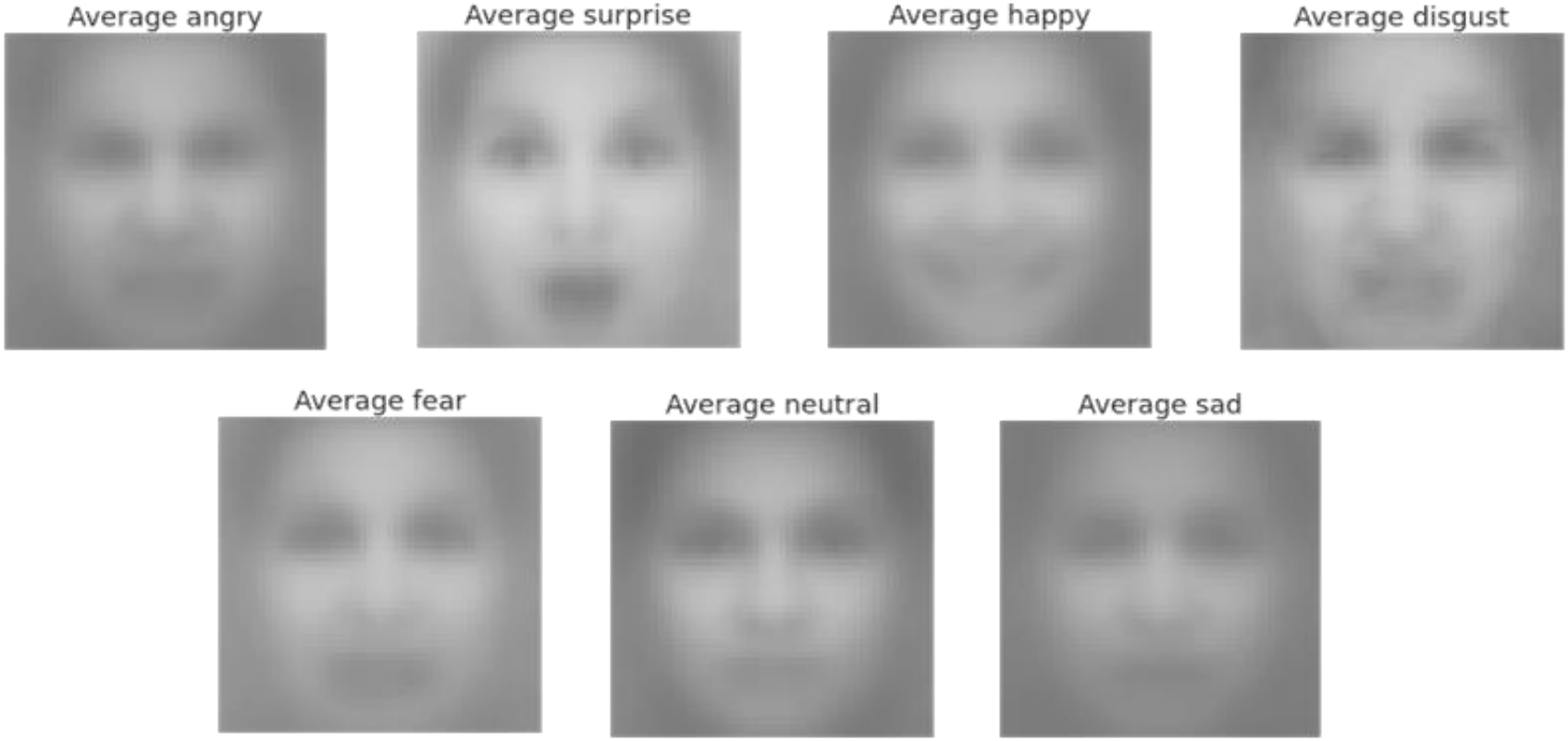
Mis-labelling?

Lighting Variations

Angle Variations

Failed to account for age, gender or cultural differences?

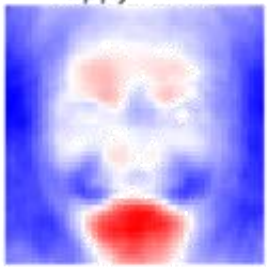
Average Images Distinguish Key Facial Features for Each Emotion



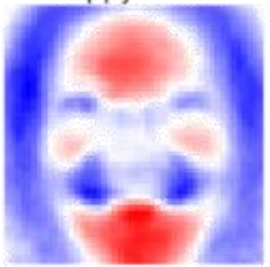
Contrast Images Further Isolate Key Facial Features to Differentiate Emotions

Happy v Angry

Happy v Fear



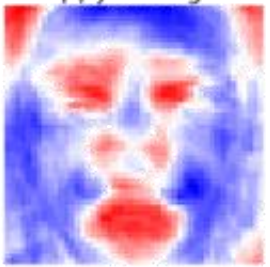
Happy v Sad



Happy v Surprise

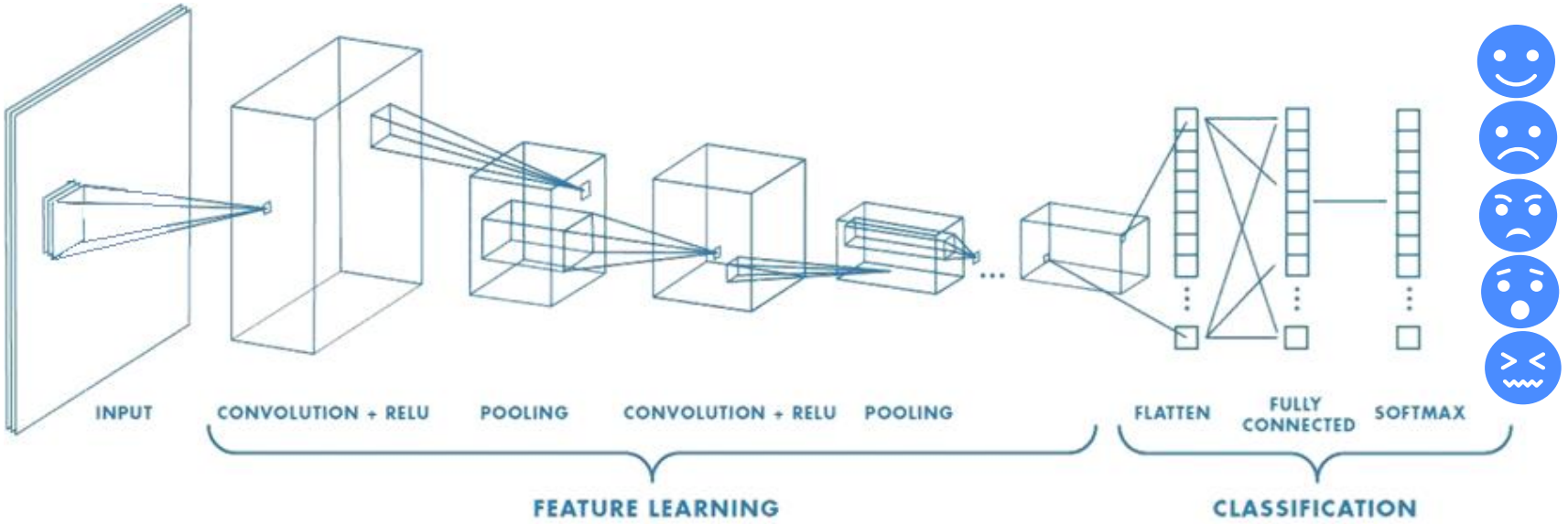
Happy v Neutral

Happy v Disgust



03
Model Workflow & Evaluation

Structure of Convolutional Neural Network



Source: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Baseline Model

- 4 Convolution w Max-Pooling Layers
- 2 Fully Connected Layers

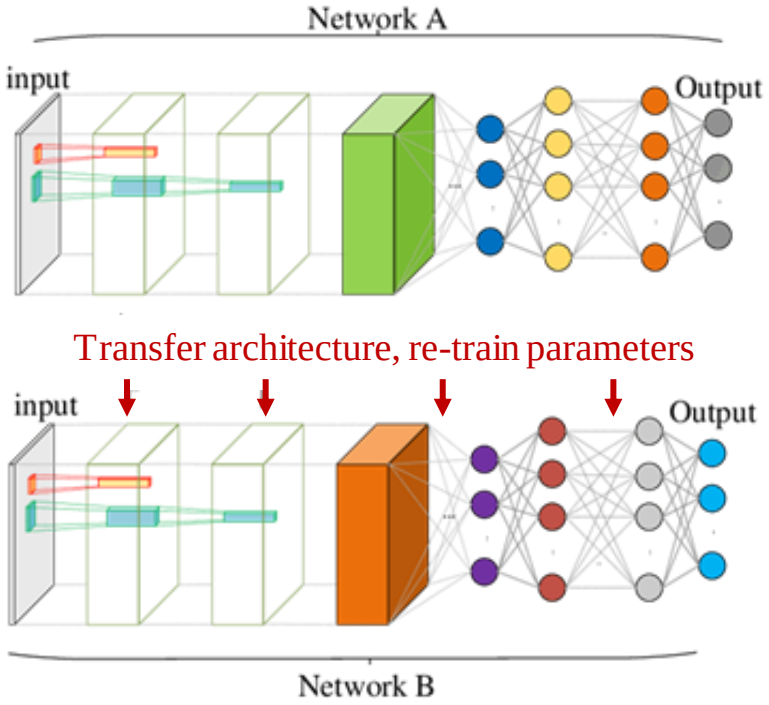
Layer (type)	Output Shape	Param #
conv_layer_1 (Conv2D)	(None, 46, 46, 128)	3584
batch_normalization (BatchNormalization)	(None, 46, 46, 128)	512
pool_layer_1 (MaxPooling2D)	(None, 23, 23, 128)	0
dropout (Dropout)	(None, 23, 23, 128)	0
conv_layer_2 (Conv2D)	(None, 21, 21, 256)	295168
batch_normalization_1 (BatchNormalization)	(None, 21, 21, 256)	1024
pool_layer_2 (MaxPooling2D)	(None, 10, 10, 256)	0
dropout_1 (Dropout)	(None, 10, 10, 256)	0
conv_layer_3 (Conv2D)	(None, 8, 8, 512)	1180160
batch_normalization_2 (BatchNormalization)	(None, 8, 8, 512)	2048
pool_layer_3 (MaxPooling2D)	(None, 4, 4, 512)	0
dropout_2 (Dropout)	(None, 4, 4, 512)	0
conv_layer_4 (Conv2D)	(None, 2, 2, 512)	2359808
batch_normalization_3 (BatchNormalization)	(None, 2, 2, 512)	2048
pool_layer_4 (MaxPooling2D)	(None, 1, 1, 512)	0
dropout_3 (Dropout)	(None, 1, 1, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 512)	262656
dropout_4 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 5)	1285

Total params: 4,239,621
Trainable params: 4,236,805
Non-trainable params: 2,816

If you know how to ride a bicycle, it will be easier to learn to ride a bike

4 Pre-Trained Architectures

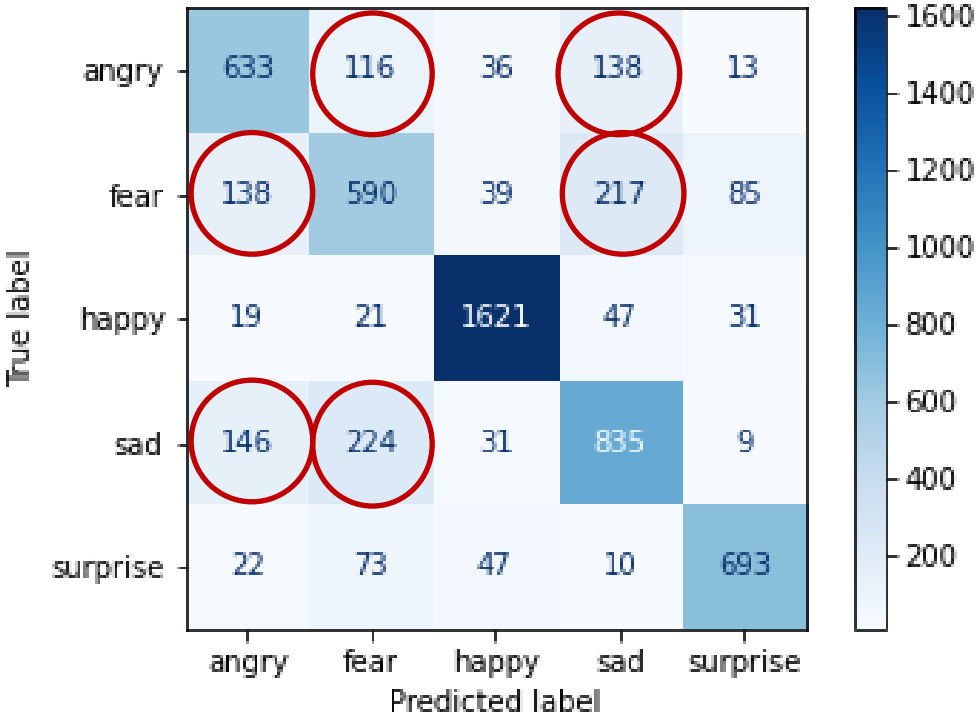
- ResNet-50
- MobileNetV2
- VGG16
- VGGFace



Model Evaluation

	Test Accuracy Score	
Baseline	58%	
VGGFace Architecture	75%	
MobileNetV2 Architecture	68%	
VGG16 Architecture	68%	
ResNet50 Architecture	65%	

Our model is very good at classifying happy, but tend to confuse angry, fear and sadness



Comparison of F1-Scores

	VGGFace	MobileNetV2	VGG16	ResNet-50
angry	0.67	0.57	0.54	0.52
fear	0.56	0.40	0.51	0.47
happy	0.92	0.89	0.87	0.82
sad	0.67	0.62	0.61	0.56
surprise	0.83	0.77	0.78	0.77

Actual Label: Sad

Predictions

VGGFace	Sad
VG166	Happy
MobileNetV2	Angry
ResNet-50	Angry

Actual Label: Angry

Predictions

VGGFace	Angry
VG166	Angry
MobileNetV2	Sad
ResNet-50	Sad

Actual Label: Surprise

Predictions

VGGFace	Fear
VG166	Fear
MobileNetV2	Surprise
ResNet-50	Surprise

Demo:
Model Deployment Using OpenCV

04
Conclusion & Next Steps

Conclusion

Quality of labelled data matters (a lot!)

Cultural differences & bias matter too

Tech adoption will be extra tough for a traditional sector

Next Steps

1. **To explore emotion recognition with facial landmark detection using Computer Vision libraries e.g. dlib**

- Possibly, combining with Paul Ekman's theory on micro expressions

2. **To acquire 'culturally relevant' datasets and re-run modelling steps**

3. **To explore prospects of emotion detection using audio and text analysis too**

Thank you

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**