Facial Emotion Recognition and Interpretation

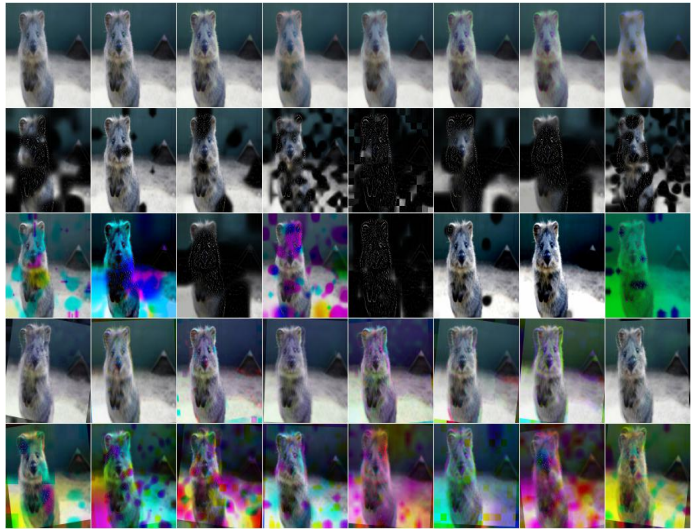
Computer Vision & Deep Learning
SEP Final Project

Mahdi Mohammadi, Leah Kawka, Jiawen Wang

Data Augmentation

- Increase diversity and amount of data

- Helps network to learn about invariances in data domain
- Important for small and unbalanced datasets

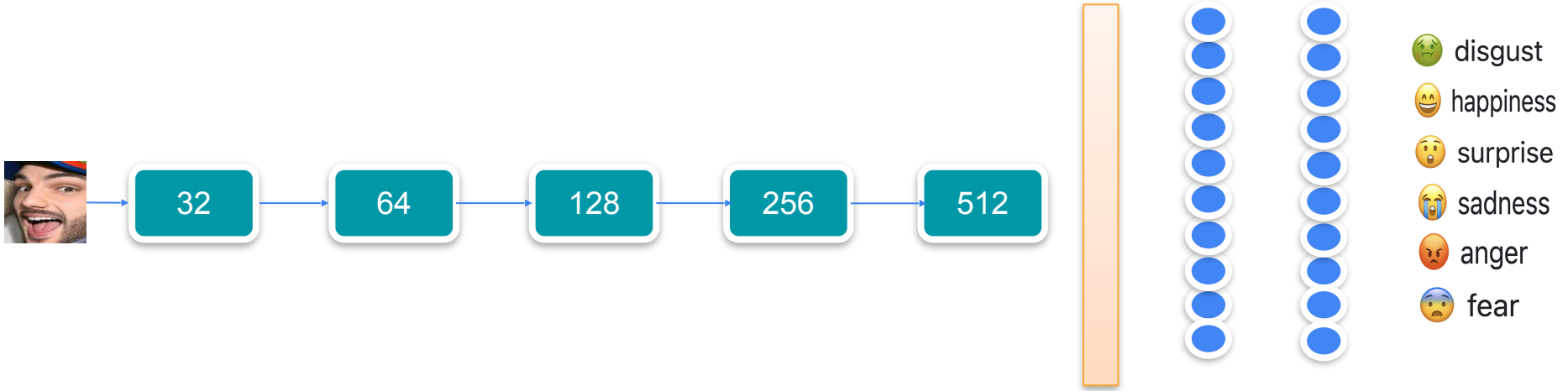


Data
Augmentation

Valid acc : 0725

with DA

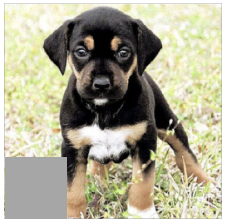
Valid acc : 0653
without DA

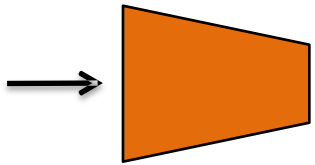


Visualization and Interpretability

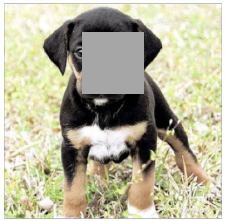
The occlusion experiment

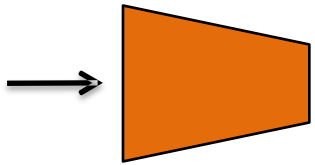
- Block different parts of the image and see how the classification score changes





DOG 0.95



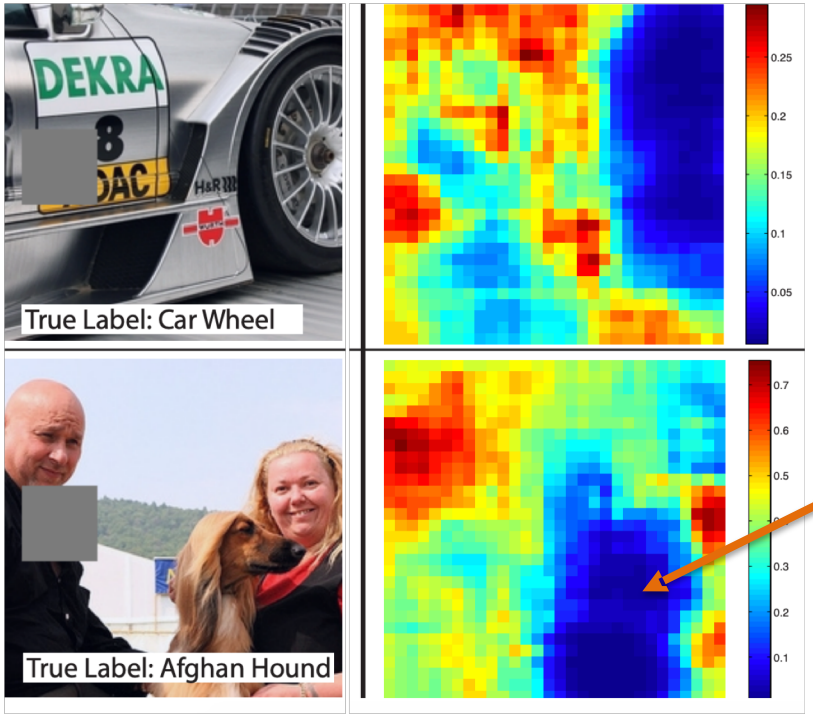


DOG 0.35

The face of the dog is more important for correct classification

Zeiler and Fergus. „Visualizing and understanding convolutional neural networks“. ECCV 2014

The occlusion experiment

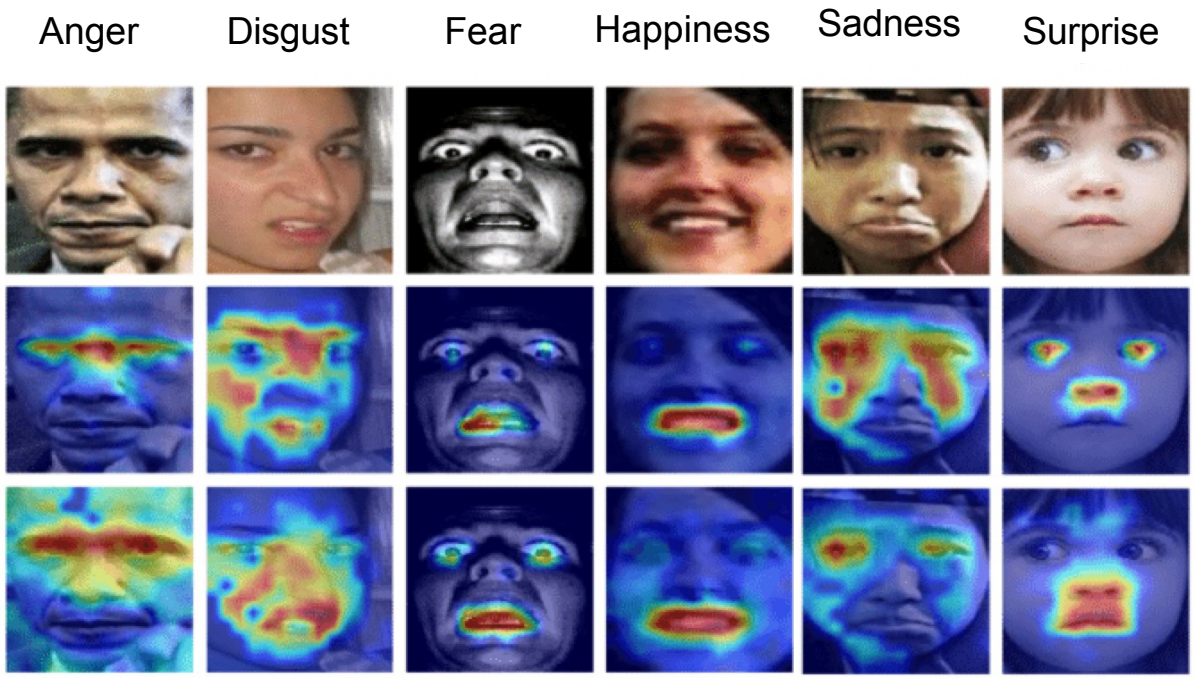


Most important pixels for classification

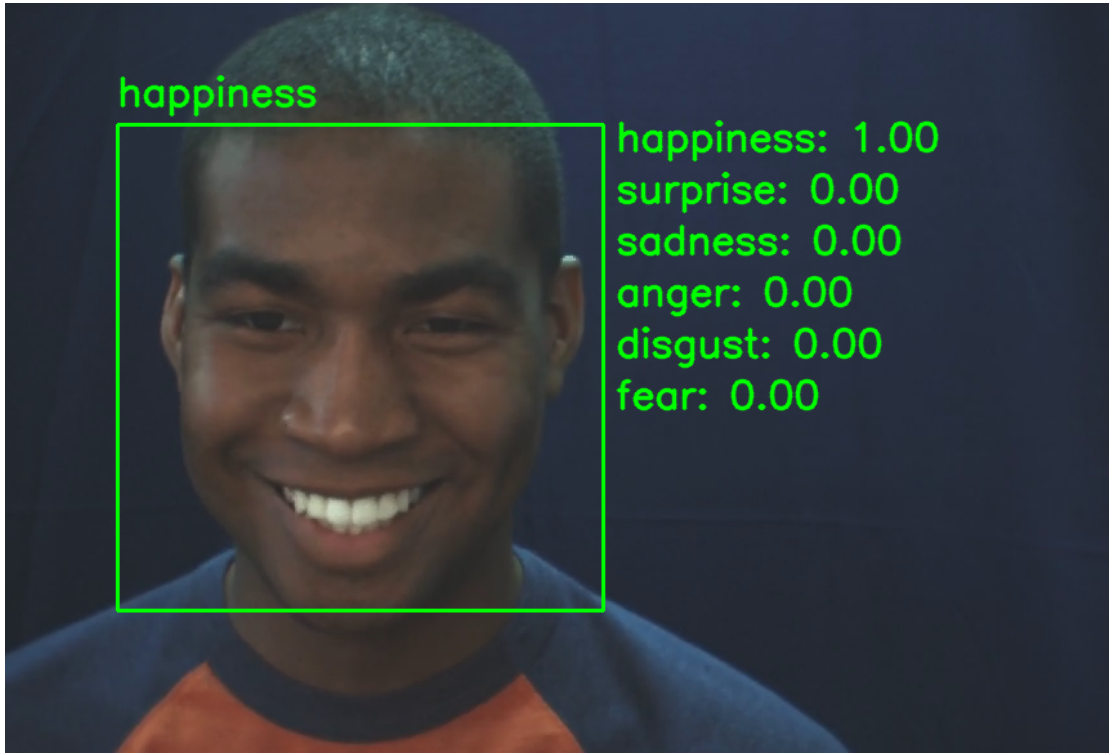
Zeiler and Fergus. „Visualizing and understanding convolutional neural networks“. ECCV 2014

Visualization helps

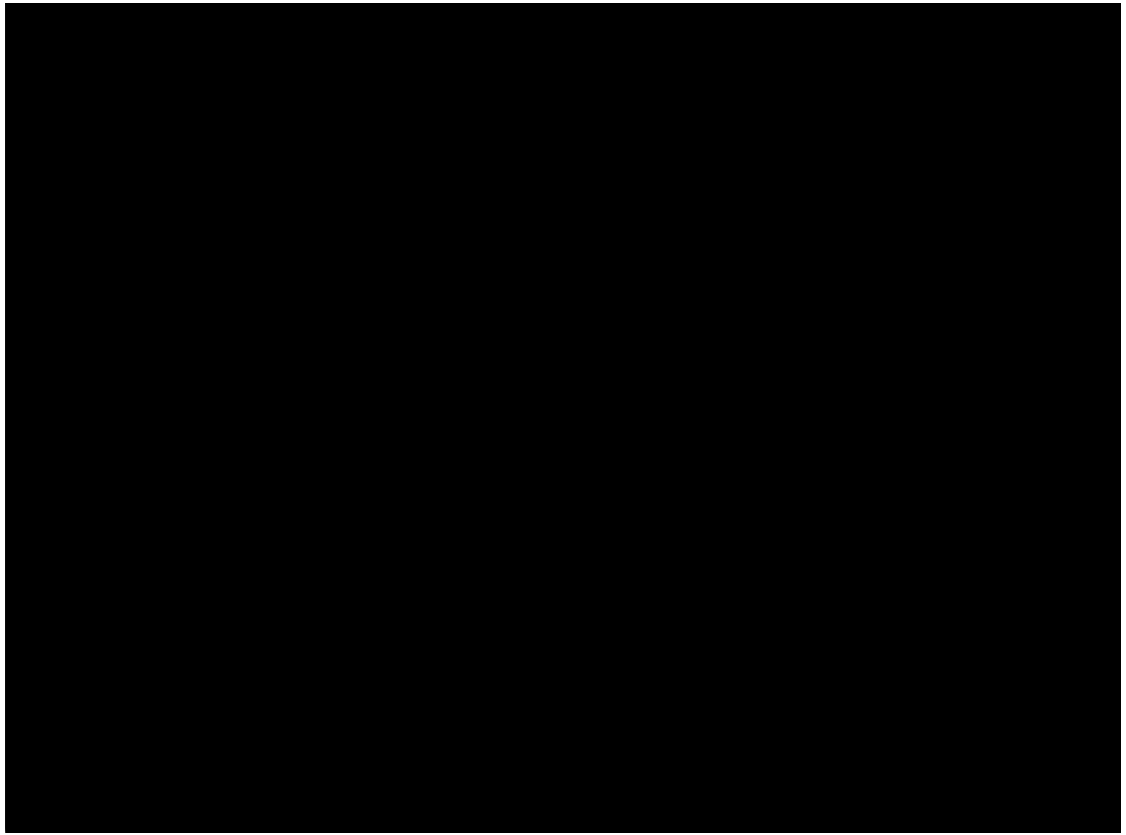
Grad-CAM



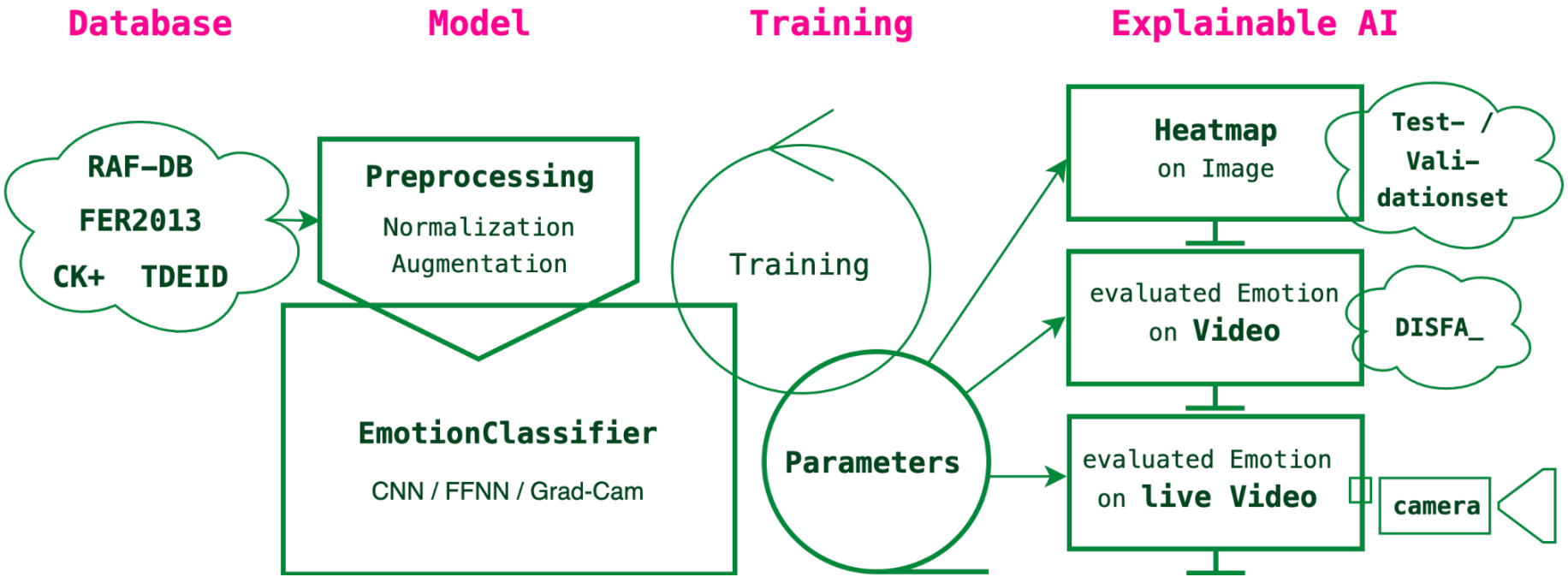
Demo Video


happiness
happiness: 1.00
surprise: 0.00
sadness: 0.00
anger: 0.00
disgust: 0.00
fear: 0.00

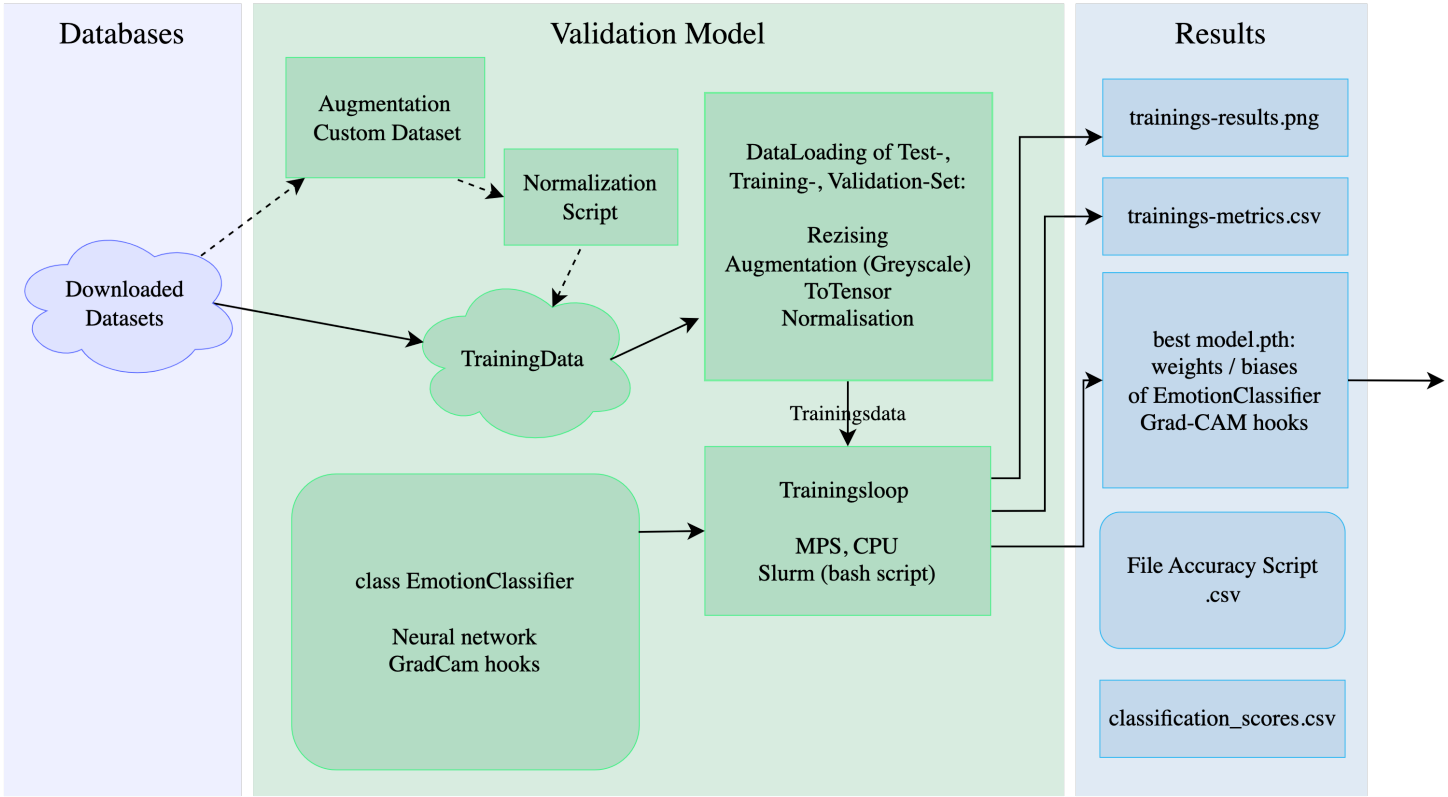
Demo Video



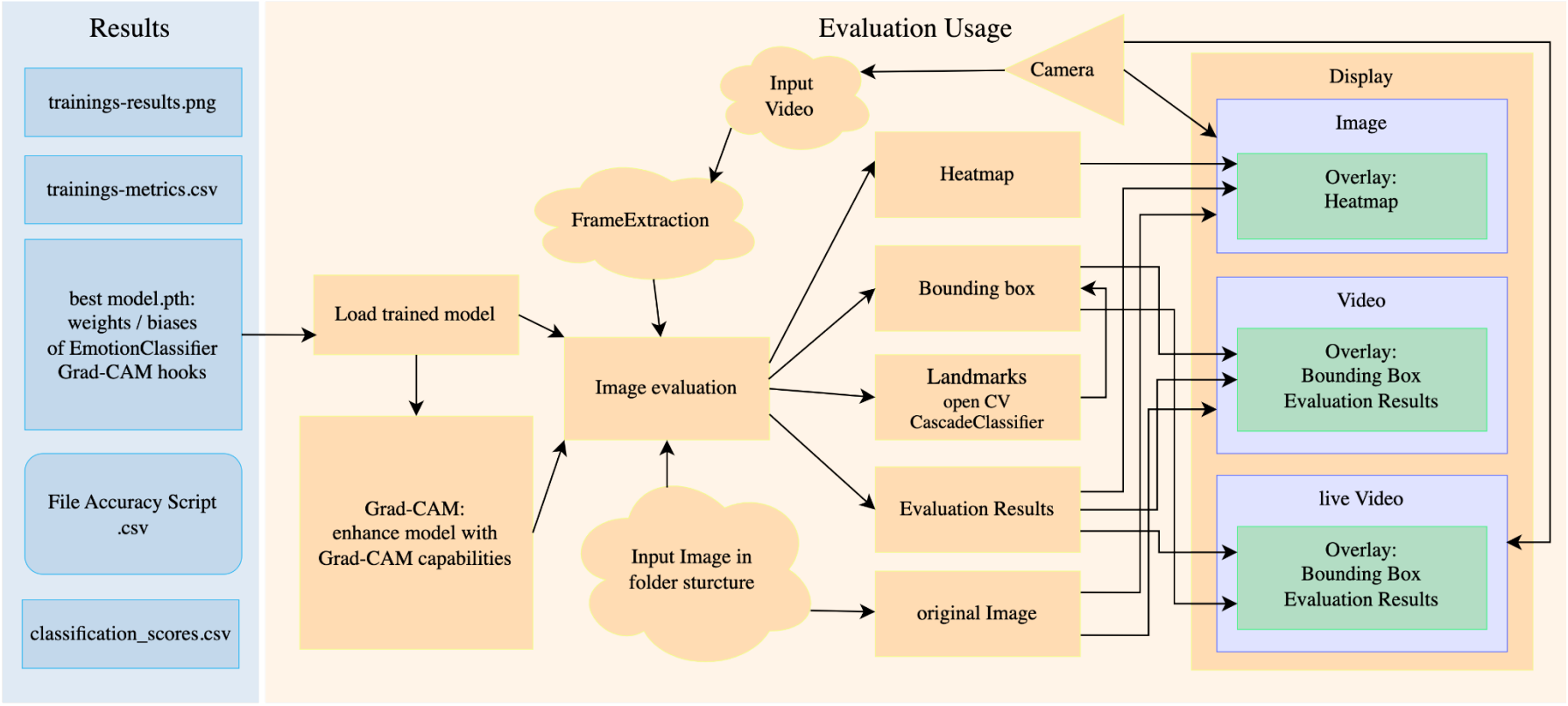
Global Experimental Pipeline



Detailed Experimental Pipeline



Detailed Experimental Pipeline

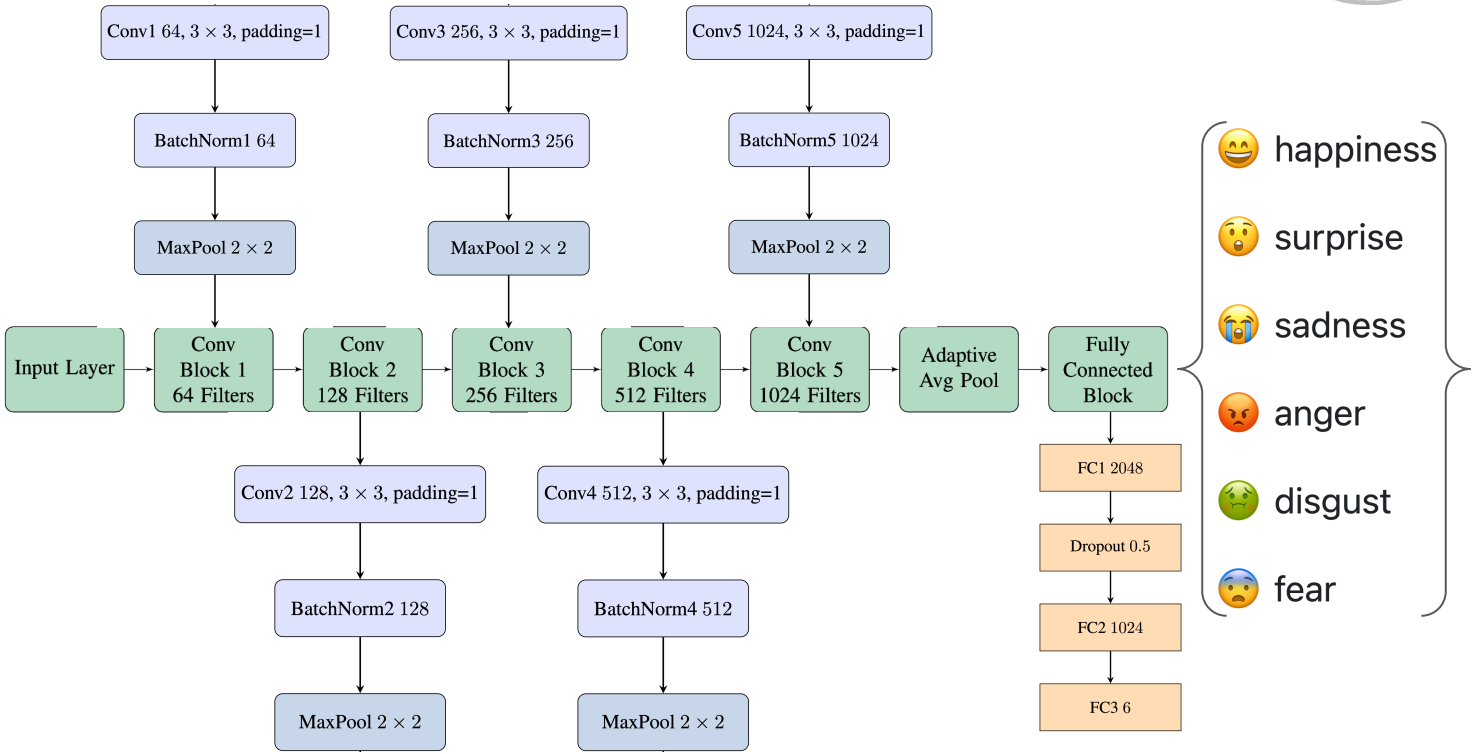


Datasets

Dataset	#Images in training set	#Images in testing set	# Video/Images in validation set
Denver Intensity of Spontaneous Facial Action Database (DISFA [13])	-	-	27
Real-world Affective Faces Database (RAF-DB [8, 9])	9747	2388	600
Facial Expression Recognition 2013 (FER2013 [1])	23743	5945	600
Extended Cohn-Kanade Dataset Plus (CK+ [11])	309	-	600
Taiwanese Facial Expression Image Database (TFEID [3, 10])	229	-	600

Table 3. Overview of the datasets statistics used in our experiment

Model Architecture



Empirical Results

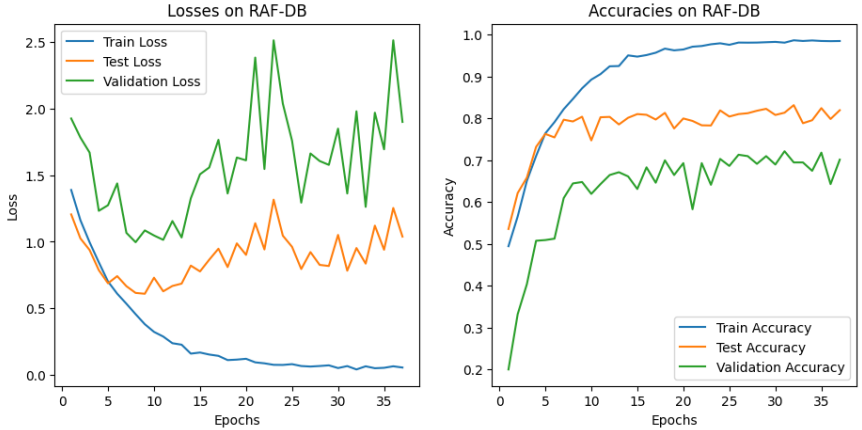
Dataset	Models	Architecture	Accuracies			# Parameters
			Training	Testing	Validation	
RAF-DB [8, 9]	ResNet18 [5]	Residual Block	**98.9**	81.3	67.9	11179590
	Ours (13 layers)	+BN-SE	96.6	80.6	66.8	2606086
	Ours (10 layers)	-BN-SE	96.3	76.9	60.6	10474118
	Ours (16 layers)	+BN+SE	98.4	81.7	71.1	10478598
	Ours (15 layers)	+BN-SE	98.6	**83.1**	**72.1**	10478086
	Ours (17 layers)	+BN-SE	97.5	82.5	70.0	41950726
FER2013 [1]	Ours (13 layers)	+BN-SE	86.6	64.1	40.2	2606086
	Ours (15 layers)	+BN-SE	89.6	**65.6**	40.7	10478086
	Ours (17 layers)	+BN-SE	**96.0**	65.5	**41.6**	41950726

Table 2. Accuracies (%) for different models (with specific architectures and numbers of parameters) in our experiments (Note that SE stands for the squeeze and excitation block and BN for the batch normalization; +/- represent with/without respectively)

Explored Hyper-parameters

Hyperparameter	Value
Learning rate	{0.01, 0.001, 0.0001}
Batch size	{8, 16, 32, 64}
Dropout rate	{0.2, 0.5}
Epoch	{20, 30, 40}
Early stopping	{`True`, `False`}
Patience	{5}

Table 4. Explored hyperparameter space for our models



Why is **interpretability** important?

- Know how models make decisions
- Know what happened in models
- Understand why
- Understand why not
- See when we succeed
- See when we fail
- Further improve the data
- Further improve the model

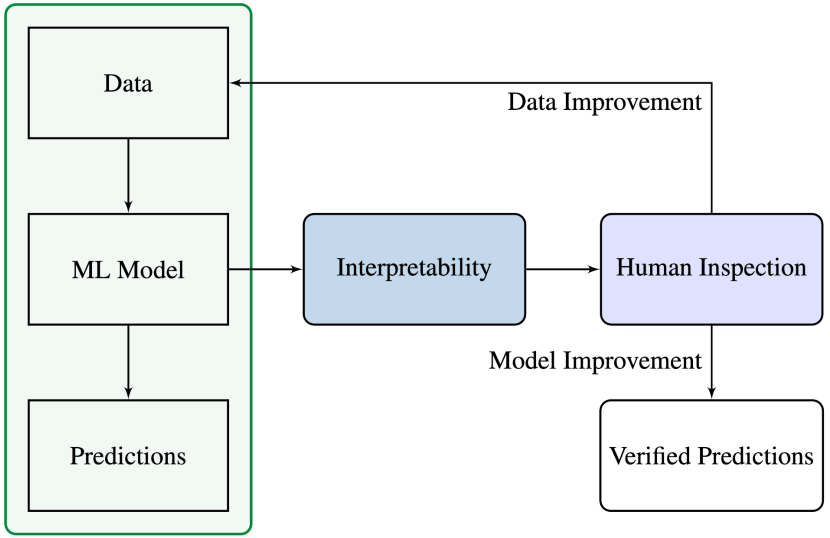


Figure 1. Overview of the traditional standardized ML evaluation process (see the illustration in the large green box) and the explainable AI pipeline

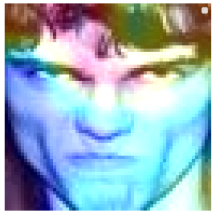
Interpretable Results (XAI)

😄 happiness

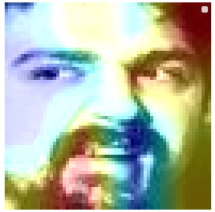

😲 surprise


😭 sadness

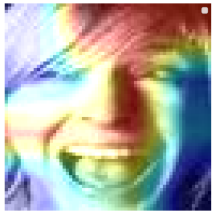




😡 anger

🤢 disgust

😨 fear

Classification Results

filepath	happiness	surprise	sadness	anger	disgust	fear
archive/RAF-DB/test/test_0524_aligned_happiness.jpg	**1.00**	0.00	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_0093_aligned_happiness.jpg	**1.00**	0.00	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_2193_aligned_sadness.jpg	0.03	0.01	**0.90**	0.01	0.02	0.04
archive/RAF-DB/test/test_1214_aligned_happiness.jpg	**1.00**	0.00	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_1816_aligned_surprise.jpg	0.00	**1.00**	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_0294_aligned_surprise.jpg	0.01	**0.99**	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_1128_aligned_happiness.jpg	**1.00**	0.00	0.00	0.00	0.00	0.00
archive/RAF-DB/test/test_1799_aligned_sadness.jpg	0.37	0.02	**0.45**	0.02	0.1	0.03
archive/RAF-DB/test/test_0610_aligned_sadness.jpg	0.02	0.00	**0.74**	0.02	0.21	0.01
archive/RAF-DB/test/test_1373_aligned_anger.jpg	0.00	0.00	0.00	**1.00**	0.00	0.00
archive/RAF-DB/test/test_1788_aligned_fear.jpg	0.00	0.00	0.00	0.00	0.00	**1.00**
archive/RAF-DB/test/test_0007_aligned_disgust.jpg	0.03	0.05	0.03	**0.58**	0.18	0.13
archive/RAF-DB/test/test_0804_aligned_disgust.jpg	0.02	0.00	0.18	0.02	**0.77**	0.00

Table 1. Overview of our random testing results examples extracted from the CSV file

Literature

- [1] Emad Barsoum, Cha Zhang, Cristian Canton-Ferrer, and Zhengyou Zhang. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016, Tokyo, Japan, November 12-16, 2016*, pages 279–283. ACM, 2016.
- [3] L.F. Chen and Y.S. Yen. Taiwanese facial expression image database, 2007.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society, 2016.
- [8] Shan Li and Weihong Deng. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, 28(1):356–370, 2019.
- [9] Shan Li, Weihong Deng, and Junping Du. Reliable crowd-sourcing and deep locality-preserving learning for expression recognition in the wild. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2584–2593. IEEE Computer Society, 2017.
- [10] Shanshan Li, Liang Guo, and Jianya Liu. Towards east asian facial expression recognition in the real world: A new database and deep recognition baseline. *Sensors*, 22(21): 8089, 2022.
- [11] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason M. Saragih, Zara Ambadar, and Iain A. Matthews. The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2010, San Francisco, CA, USA, 13-18 June, 2010*, pages 94–101. IEEE Computer Society, 2010.
- [13] S Mohammad Mavadati, Mohammad H Mahoor, Kevin Bartlett, Philip Trinh, and Jeffrey F Cohn. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.
- https://www.kaggle.com/datasets/shuvoalok/raf-db-dataset
- https://www.kaggle.com/datasets/msambare/fer2013
- http://www.whdeng.cn/raf/model1.html
- http://mohammadmahoor.com/disfa/

Live **Demo** 😄

Thanks for your attention!

Any questions or remarks?

Mahdi Mohammadi, Leah Kawka, Jiawen Wang

Backup Appendix 😄