Facial Emotion Recognition (FER)

Rui Shi, Yuting Ma, Jiayi Wang, Menglei Wang

FER-2013 dataset

happy



neutral



surprise



angry





happy

neutral

angry



happy



neutral



surprise



angry



disgust



fear



sad



disgust



fear



sad



Data Preprocessing and Augmentation

- Standardization: Images are resized to 48x48 pixels and denoised to improve quality.
- Data Augmentation: Augment minority classes with transformation techniques like rotation, flipping, and zooming while downsampling some majority classes.

Feature Extraction: HOG + DCT

Step 1: Extract HOG (Histogram of Oriented Gradients) features

Step 2: Apply DCT (Discrete Cosine Transform) to HOG features



Classification with HOG + DCT

Confusion Matrix (XGBoost)										
True label	angry -	257	34	86	116	142	117	47		
	disgust -	12	22	14	4	12	16	7		
	fear -	106	34	225	85	115	132	123		
	happy -	111	20	65	902	125	162	58		
	neutral -	109	30	89	107	409	170	79		
	sad -	120	33	125	121	179	347	41		
	surprise -	34	13	53	59	52	43	380		
		angry -	disgust -	- Lear	- Addey dicted la	lede neutral -	- sad -	surprise -		

	Confusion Matrix (RF)									
angry -	188	76	36	132	140	145	82			
disgust -	9	16	9	14	17	10	12			
fear -	94	69	111	117	140	144	145			
happy -	131	76	38	793	120	188	97			
neutral -	108	54	51	125	352	159	144			
sad -	123	67	49	153	196	309	69			
surprise -	36	24	17	75	85	56	341			
	angry -	disgust -	fear -	, Addey	neutral -	- sad -	surprise -			
	angry - disgust - fear - happy - neutral - sad - surprise -	angry - 188 disgust - 9 fear - 94 happy - 131 neutral - 108 sad - 123 surprise - 36	angry 188 76 angry 9 16 disgust 9 16 fear 94 69 happy 131 76 neutral 108 54 sad 123 67 surprise 36 24	angry 188 76 36 disgust 9 16 9 fear 94 69 111 happy 131 76 38 neutral 108 54 51 sad 123 67 49 surprise 36 24 17 but but but but but but the but the but the but sad 123 67 49 surprise 36 24 17 but the but the but the but but the but the but the but <	angry 188 76 36 132 disgust 9 16 9 14 fear 94 69 111 117 happy 131 76 38 793 neutral 108 54 51 125 sad 123 67 49 153 surprise 36 24 17 75 been been been been been been been been	angry 188 76 36 132 140 disgust 9 16 9 14 17 fear 94 69 111 117 140 happy 131 76 38 793 120 neutral 108 54 51 125 352 sad 123 67 49 153 196 surprise 36 24 17 75 85 fear ing ing ing ing ing ing sad 123 67 49 153 196 ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing ing <td>angry - 188 76 36 132 140 145 disgust - 9 16 9 14 17 10 fear - 94 69 111 117 140 144 happy - 131 76 38 793 120 188 neutral - 108 54 51 125 352 159 sad - 123 67 49 153 196 309 surprise - 36 24 17 75 85 56 bee isig isig isig isig isig isig isig isig redicted libelibelibelibelibelibelibelibelibelibe</td>	angry - 188 76 36 132 140 145 disgust - 9 16 9 14 17 10 fear - 94 69 111 117 140 144 happy - 131 76 38 793 120 188 neutral - 108 54 51 125 352 159 sad - 123 67 49 153 196 309 surprise - 36 24 17 75 85 56 bee isig isig isig isig isig isig isig isig redicted libelibelibelibelibelibelibelibelibelibe			

Feature Extraction:LBP



Pattern = 00011001 LBP = 1 + 8 + 16 = 25

Feature Extraction: Eigenface Analysis

- Eigenfaces are the "principal components" of a dataset of face images.
- Each eigenface represents a specific direction of variation in facial features, which has a good explanation of the variance.



Feature Extraction:Gabor Filter

- A linear filter used in image processing that's especially good at analyzing textures and edges.
- It essentially looks for specific frequencies and orientations within localized regions of an image for special patterns.





The Performance of Some Feature-based Classifiers

PCA + XGBoost: overall accuracy: 0.48 accuracy: 0.54

HOG & PCA + Gabor Filter & PCA + XGBoost: overall



Feature Extraction:MediaPipe

- Extract 478 3D facial landmark points per image.
- Output: Landmark-based geometric features representing facial structure.
- Remark: PCA with 15 components captures the full variance of the MediaPipe features



BaseLine Model: MediaPipe+PCA+KNN, overall accuracy = 34%

Classification Ensembles for MediaPipe

- PCA of 249 features with Gabor Filter captures 99.5% Variance
- Concatenate total 269 features from Gabor Filter and MediaPipe
- Apply Random-Forest and XGBoost as classifiers



Deep Learning Approach: CNN

- - Training Strategy:
- * Loss function: Categorical cross-entropy.
- * Optimizer: Adam with learning rate scheduling.

Deep Learning Approach: CNN



Transfer Learning Approach: ResNet50

ResNet50: a 50-layer deep convolutional neural network pre-trained on ImageNet Fine-tuning for FER-2013: replace the fully connected layer to output the number of



Future Work

- Consider preprocessing techniques that account for lowresolution images
- Explore alternative methods such as EfficientNet or Vision Transformers
- Explore explainability techniques, such as Grad-CAM or SHAP, to better understand model predictions
- Investigate the integration of emotion recognition into real-world systems, such as interactive VR or mental health monitoring tools