Multimodal Approach for Novelty Aware Emotion Recognition with Situational Knowledge

Presented By

Mijanur R Palash

PhD Candidate

Committee:

Dr. Bharat Bhargava (Advisor)

Dr. Chunyi Peng

Dr. Jianguo Wang

Dr. Vaneet Aggarwal

July 20, 2023

Agenda

Background and Motivation

Contribution 1: SAFER (Emotion Recognition From Face)

Contribution 2: EMERSK (Multimodal Emotion Recognition)

Contribution 3: CoNERS (Novelty Aware Emotion Recognition)

Question & Answer

Background and Motivation

The USA: in a mental health crisis

How can we help??

Mental health: influences gun violence, school shooting, suicide etc.

Emergency responders end search for man whc jumped into Wabash

STAFF REPORTS Jul 3, 2023

1 of 4

Submit a Le

If you're intereste click here.

Submit

Other Headli

- South Bend st at Purdue
- Purdue track
- Northern Ligh

The New York Times

'It's Life or Death': The Mental Health Crisis Among U.S. Teens

What the Florida school shooting reveals about the gaps in our mental health system - Los Angeles Times

Background and Motivation

- **Close relation between emotion and mental health**
- **Changes in emotions over time used for:**
  - Trigger identification
  - Early sign of instability
  - Preventive steps

- **Our idea of help:**
  - Automated emotion recognition which can be used for:
    - Automated monitoring
    - Advance warning
    - Alarm triggering

Emotion Indicators

Emotions can be conveyed through both visual and non-visual indicators.

Visual	Non-visual
❑ Facial expression ❑ Posture ❑ Gait	❑ Speech ❑ Text ❑ Brain scan

Challenges in Emotion Recognition

Accuracy

- ❑ Providing high accuracy in emotion recognition
- ❑ Most focused area

Explainability

- ❑ Giving transparent explanations of the results
- ❑ Lack of focus

Novelty Handling

- ❑ Detecting and adapting to novel situations
- ❑ Lack of focus

Existing Works

Face Based	❑ Heavily focused on facial emotion recognition (FER)
Unimodal/Bimodal	❑ Use only one or two modes
Not Explainable	❑ Not focused on explainable output
Novelty	❑ Not designed to handle novelty

Contributions

SAFER: Improved facial emotion recognition

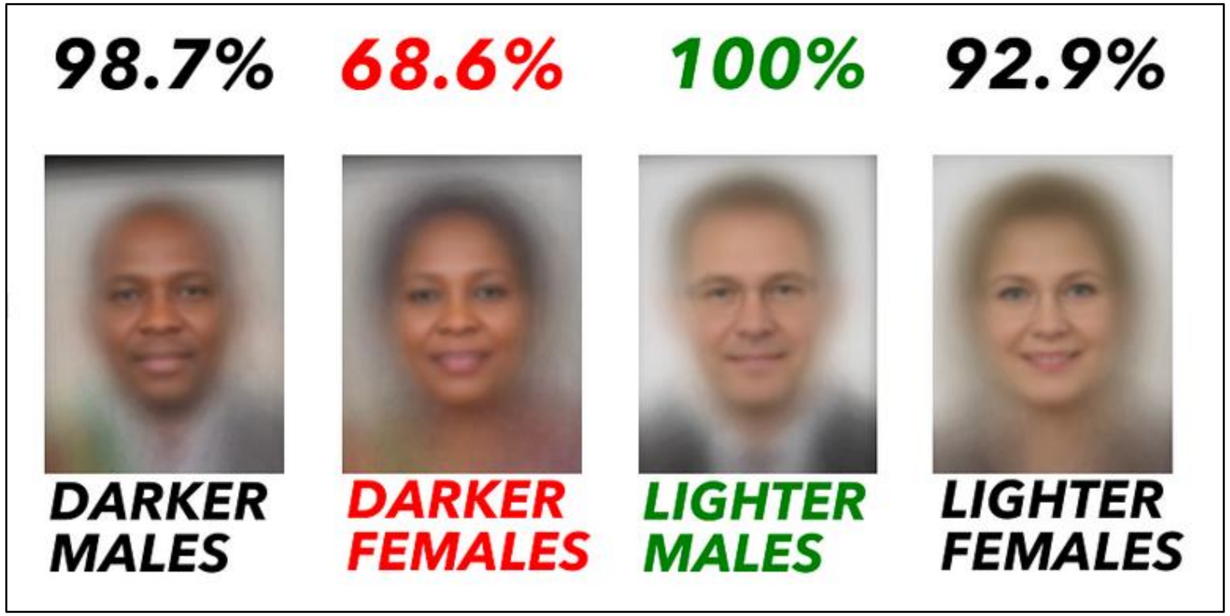
EMERSK: Explainable multimodal emotion recognition

CoNERS: Novelty detection and handling

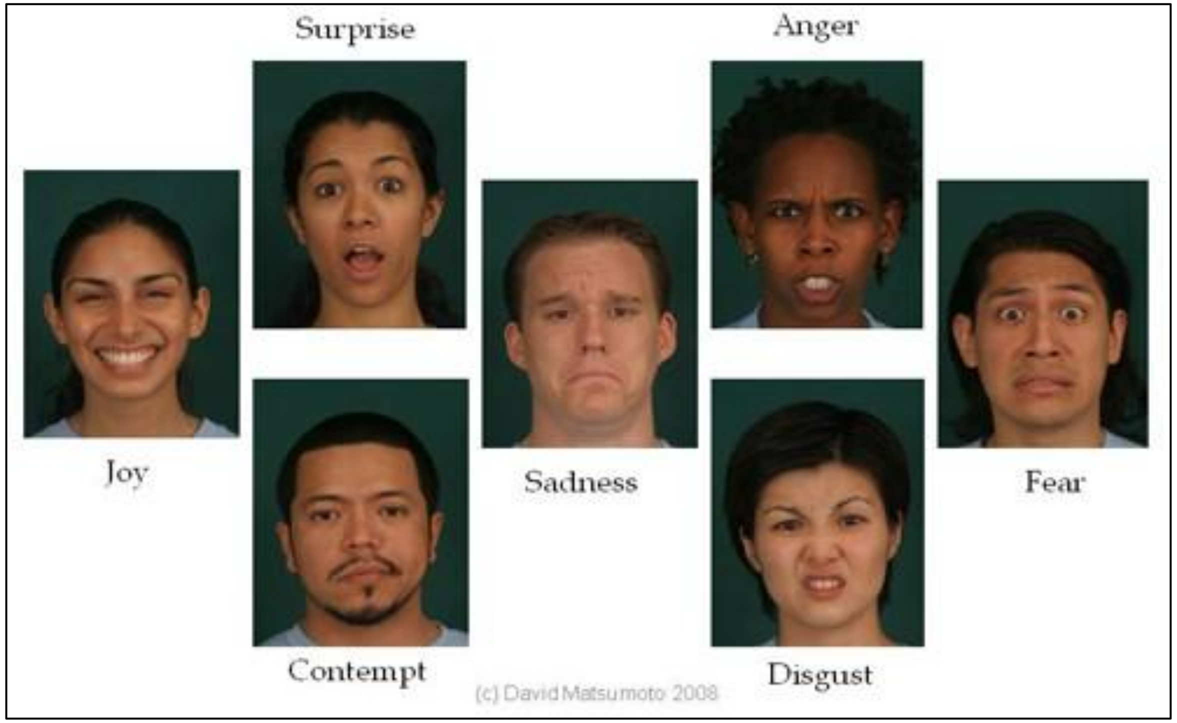
SAFER: Situation Aware Facial Emotion Recognition

Problem Statement

- Face: important medium of emotion
- Subject to bias: need generalization



Bias in Amazon AI gender classification



Facial expression of emotions

Can we improve the facial emotion recognition?

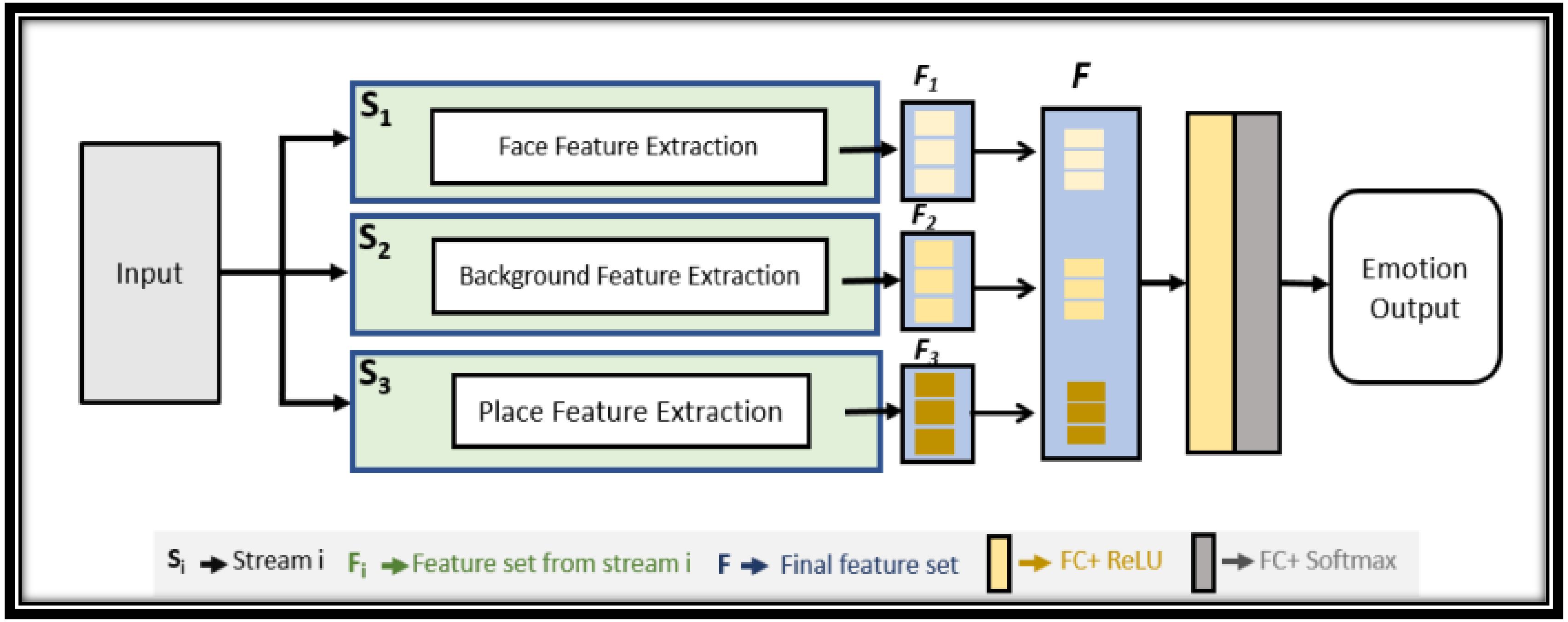
SAFER Architecture

Face feature extraction

Background feature extraction

Place feature extraction

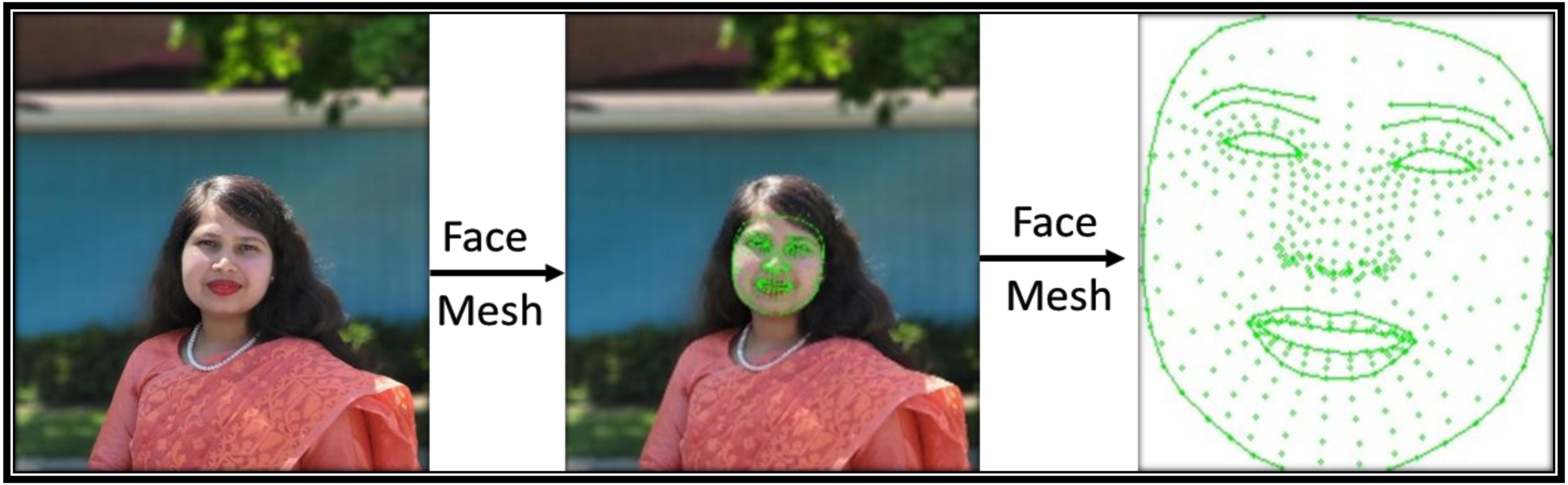
Classification network



Face Feature Extraction: Face Detection

BlazeFace [11] for face detection

- ❑ Identifies key points
- ❑ Generates face mesh



Face detection

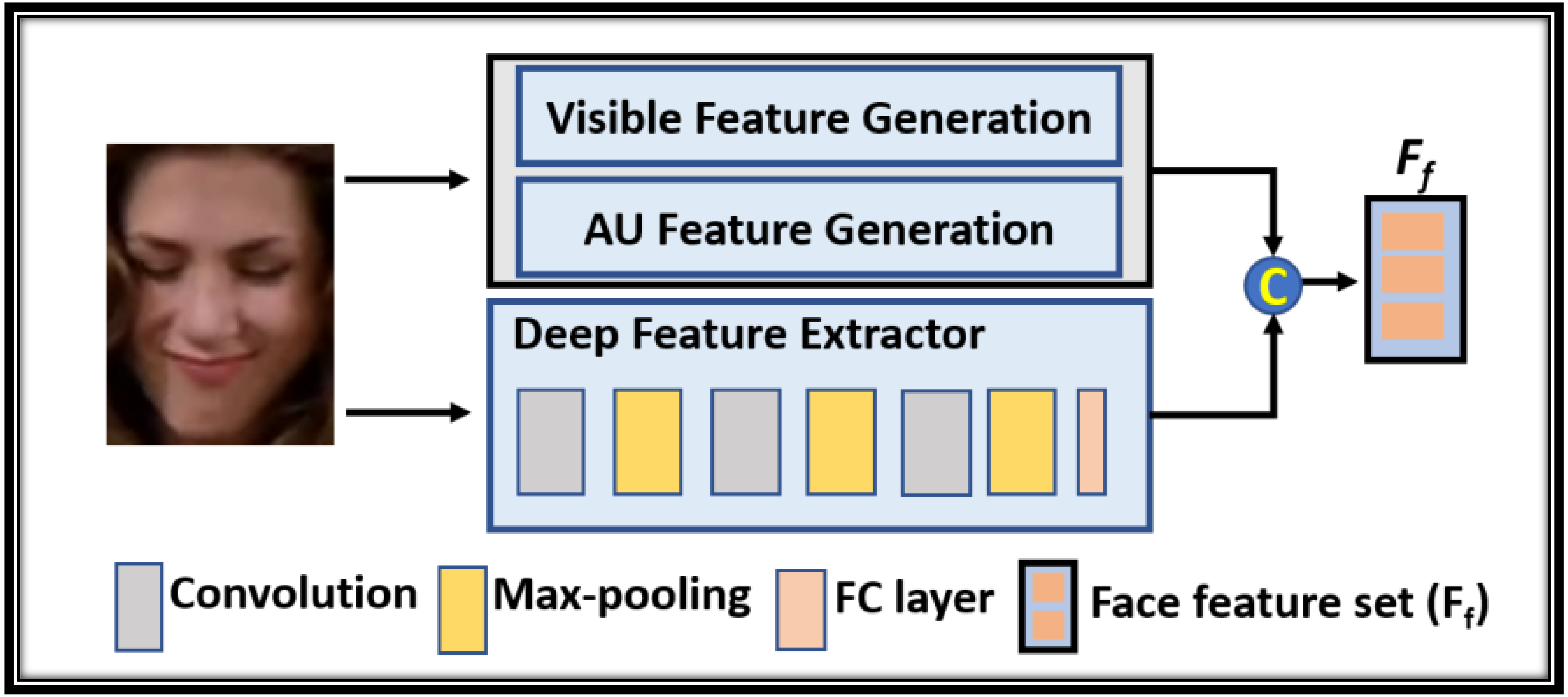
Face Feature Extraction: Feature Types

Face Feature Types

Action unit (AU) features

Visible features

Deep features



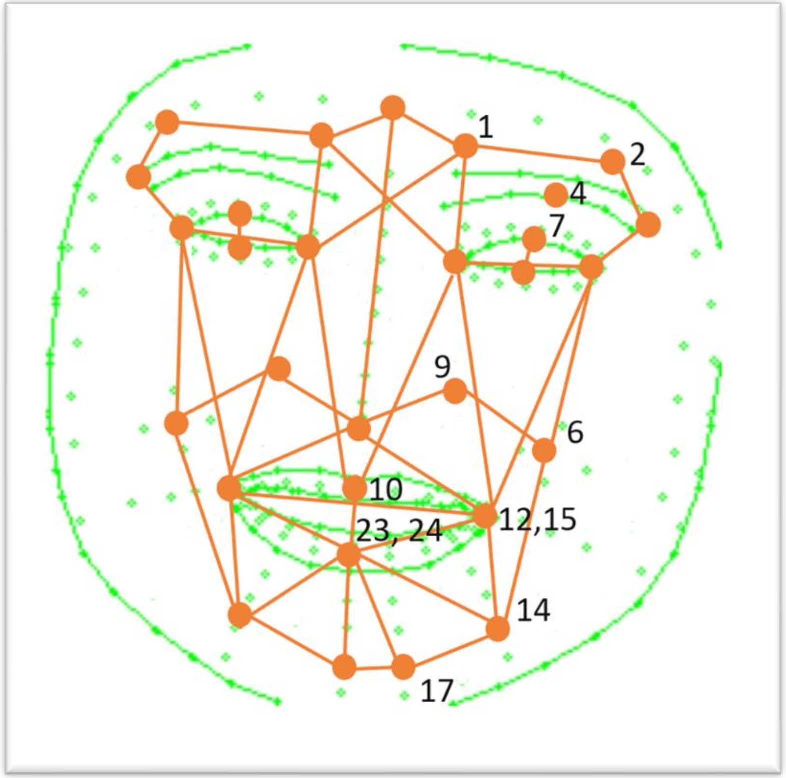
Face feature extraction module

Face Feature Extraction: Action Unit (AU) Features

- ❑ **AUs:** set of face muscles that corresponds to specific expressions
- ❑ **BlazePose:** computer vision model used to detect the centers of the AUs

AU ID	AU Name	Points
1	Inner brow raiser	Above inner brow
6	Cheek raiser	At cheek center
24	Lip pressor	Bottom lip center

Action units for “Sadness”



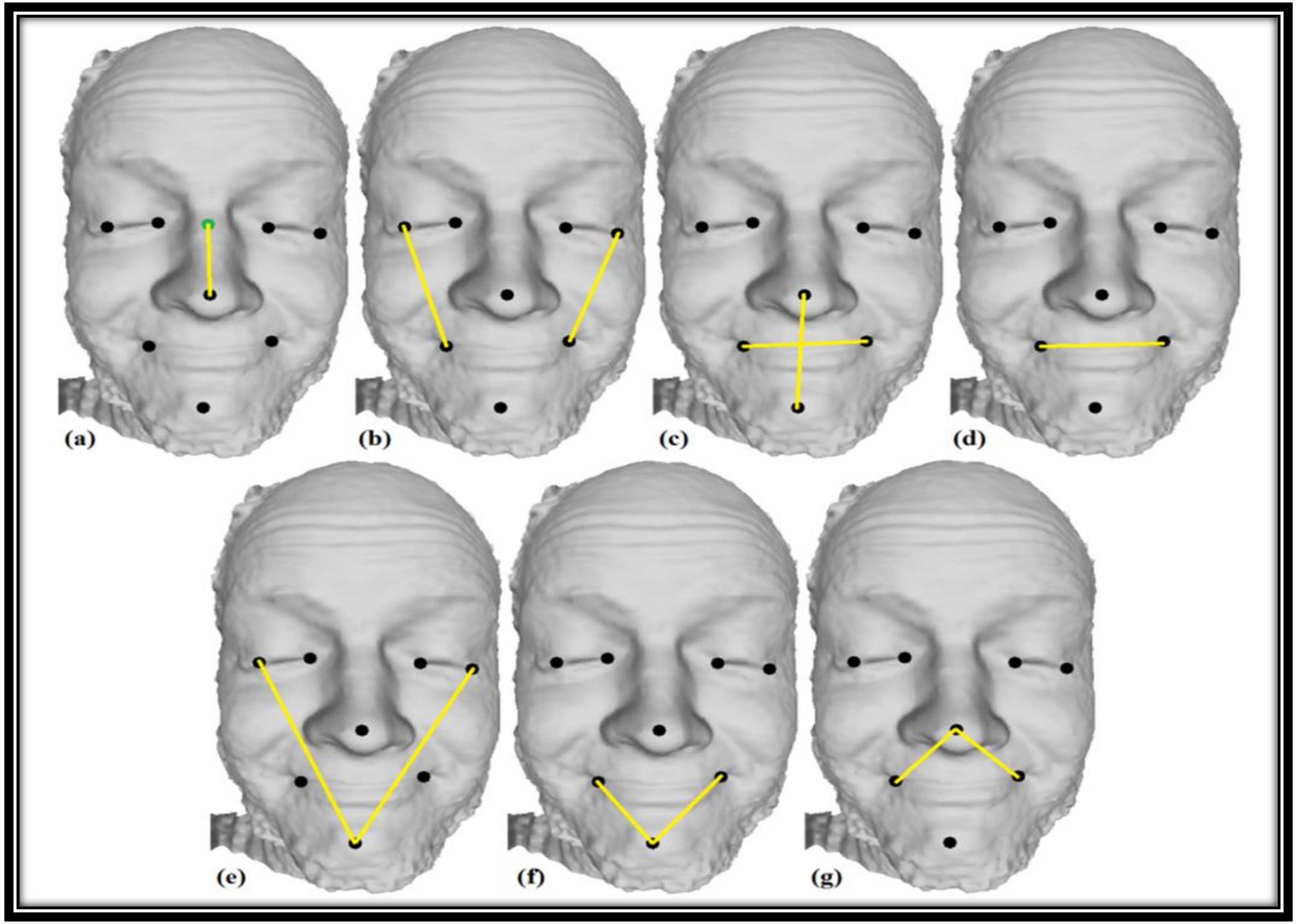
Action unit features generation

Face Feature Extraction: Visible Features

Visible features

- ❑ Reflect physical changes of face parts with emotion
- ❑ Measure as width, distance and angle

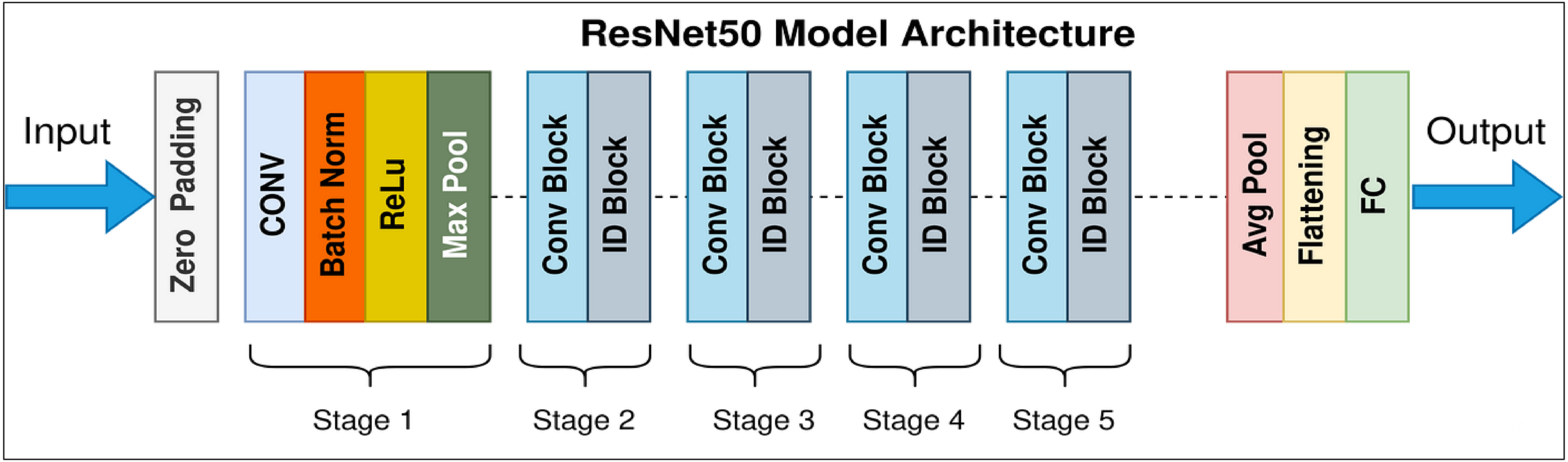
Feature type	**Description**
Width	Left eye
Distance	Left and right eyes
Angle	Left eye with right eye and mouth



Visible features

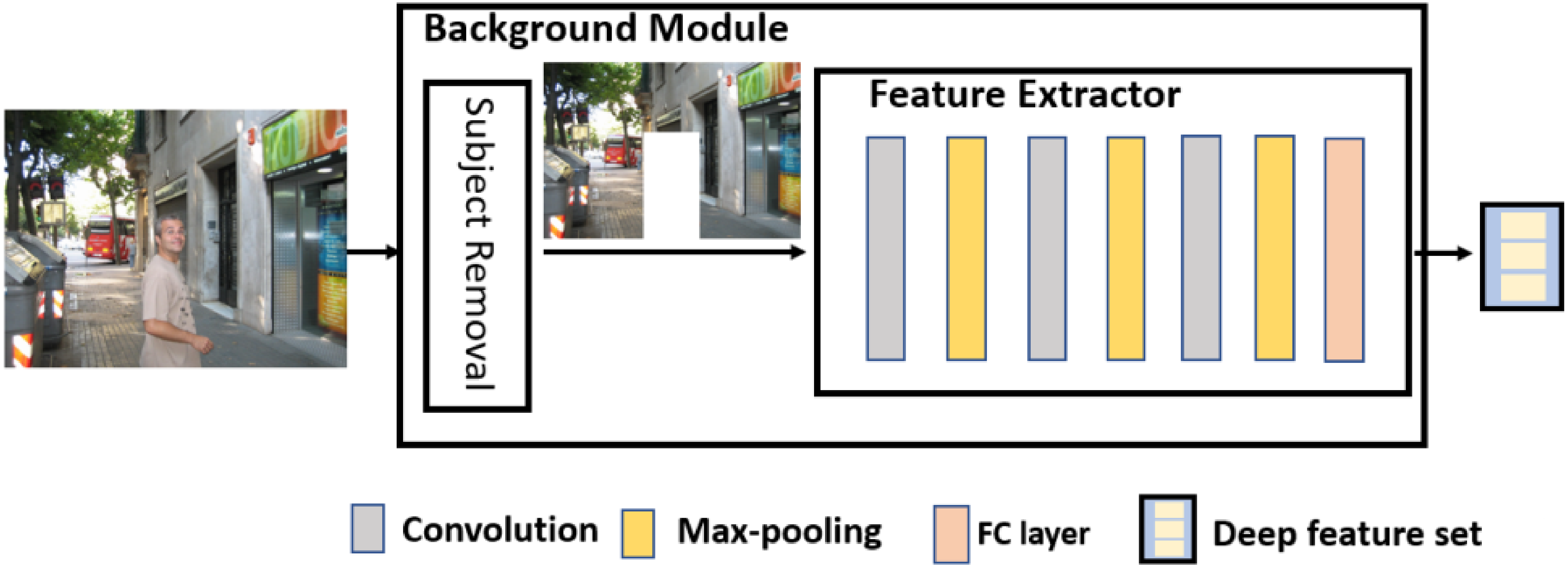
Face Feature Extraction: Deep Features

- **Deep features:** representations from the deeper layers of a CNN
- **Transfer learning:**
  - Knowledge gained in one task applied to improve the performance of a related but different task
  - Resnet-50 pre-trained on ImageNet dataset (14 million samples)



Background Feature Extraction

- **Background:** source of important contextual information
- **Process:**
  - subject removal
  - convolutional feature extractor



Place Feature Extraction

- **Places are associated with emotion:**
  - garden: happiness, cemetery: sadness
- **Provides additional information in emotion recognition and explanation generation**
- **Pre-trained Model**
  - AlexNet
- **Place dataset [23]**
  - 10 million labeled images
  - 205 place categories

Place category: “Bedroom”

Evaluation Setup

- **PC:**
  - 2.6 GHz 20 Cores Intel Xeon CPU
  - 96 GB of RAM
  - 3 NVIDIA TESLA GPUs with 24 GB of memory each
- **Dataset preparation**
  - Split into training, validation and test sets in an 80:10:10 ratio
  - Images resized to 224 × 224 pixels
  - Augmentation: cropping, rotation, brightness, and contrast adjustments
- **Evaluation Metrics**
  - Accuracy (%): $\frac{\text{\#samples correctly predicted}}{\text{\#total samples}} \text{x100}$

Evaluation Setup: Related Works

Name	Method	Limitations
Wen et al. [34]	Ensemble CNN	Low accuracy
Dhankar et al. [17]	ResNet-50	Low accuracy
Renda et al. [35]	Ensemble CNN	Low accuracy
Gan et al. [16]	Soft Label boosting+ ECNN	Not emphasized on all face points
A-C [18]	Adaptive correlation-based loss	Orthogonal work
Lee et al. [6]	Two stream architecture with adaptive fusion	Not well generalized as mainly focused on CAER-S dataset
Kosti et al. [5]	Dual stream CNN	Not well generalized as mainly focused on EMOTIC dataset
Li et al. [33]	Relational region-level analysis with Body-Object and Body-Part attention+ GCN	Accuracy can be improved

Evaluation Setup: Dataset

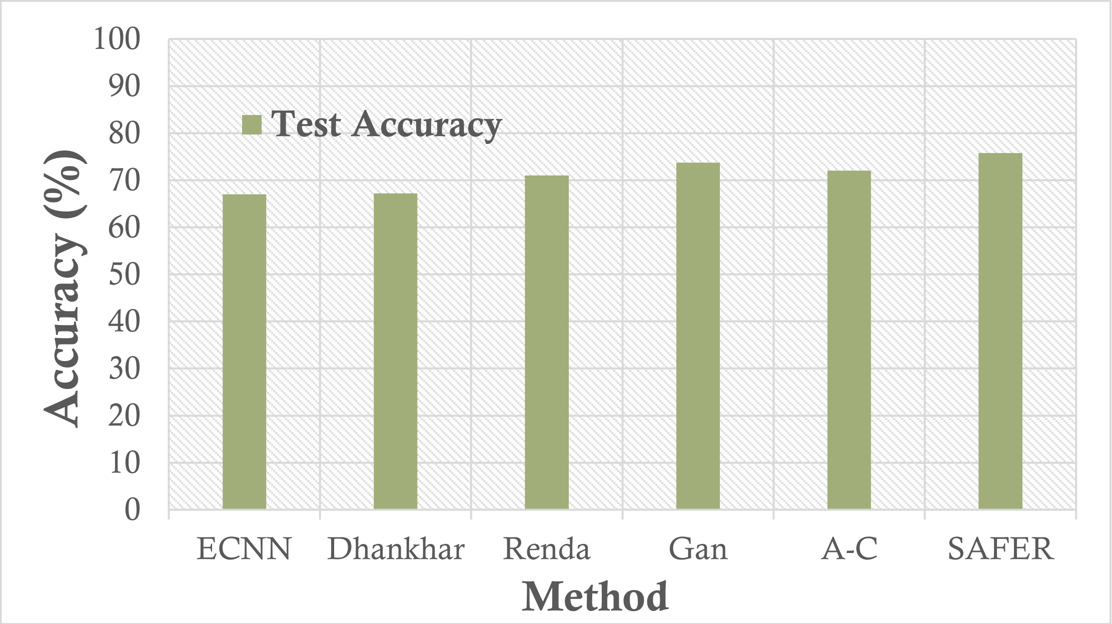
FER-2013: 3.2K posed images	**CK+: 593 posed and spontaneous videos**	**AffectNets: 450K spontaneous image**	**CAER-S: 70K image from TV shows**	**RAF-DB: 30K diverse face images**	**FABO: 206 posed videos**



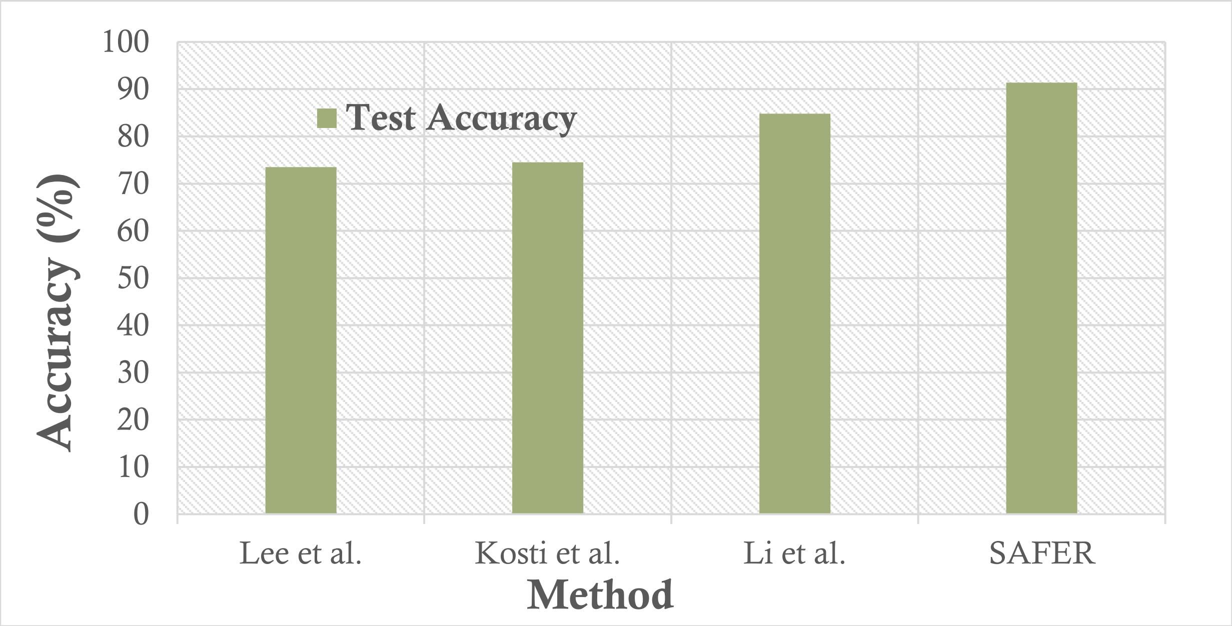

Sample images from the datasets

Experimental Results and Findings

Research question: Does safer improve accuracy?



Results on FER-2013



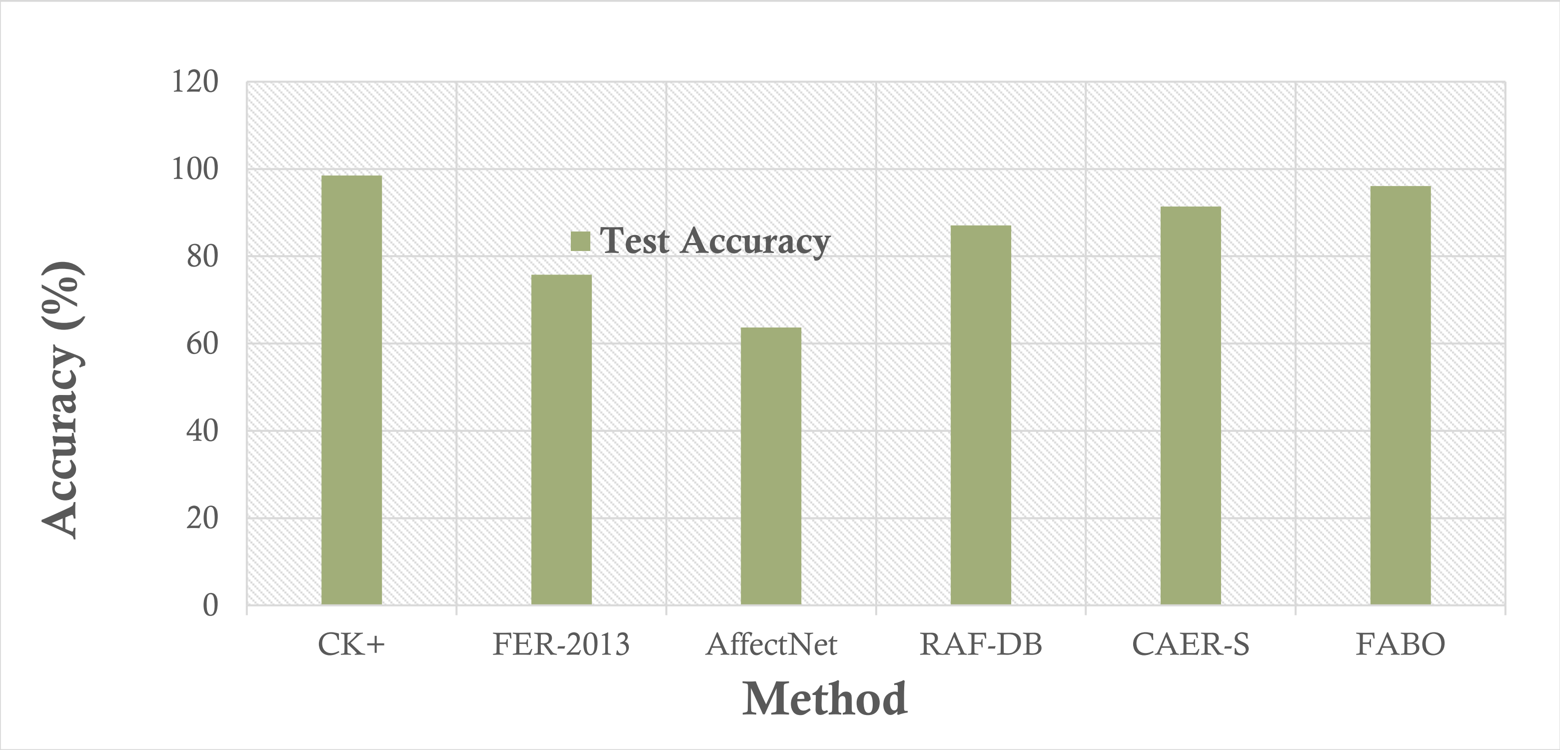
Results on CAER-S

- X axis: Name of the method ; Y axis: Accuracy reported by them in the dataset
- The higher the bar, the better!

Findings: SAFER improves accuracy and outperforms state-of-the-art methods.

Experimental Results and Findings

Research question: Does Safer Generalizes Result?



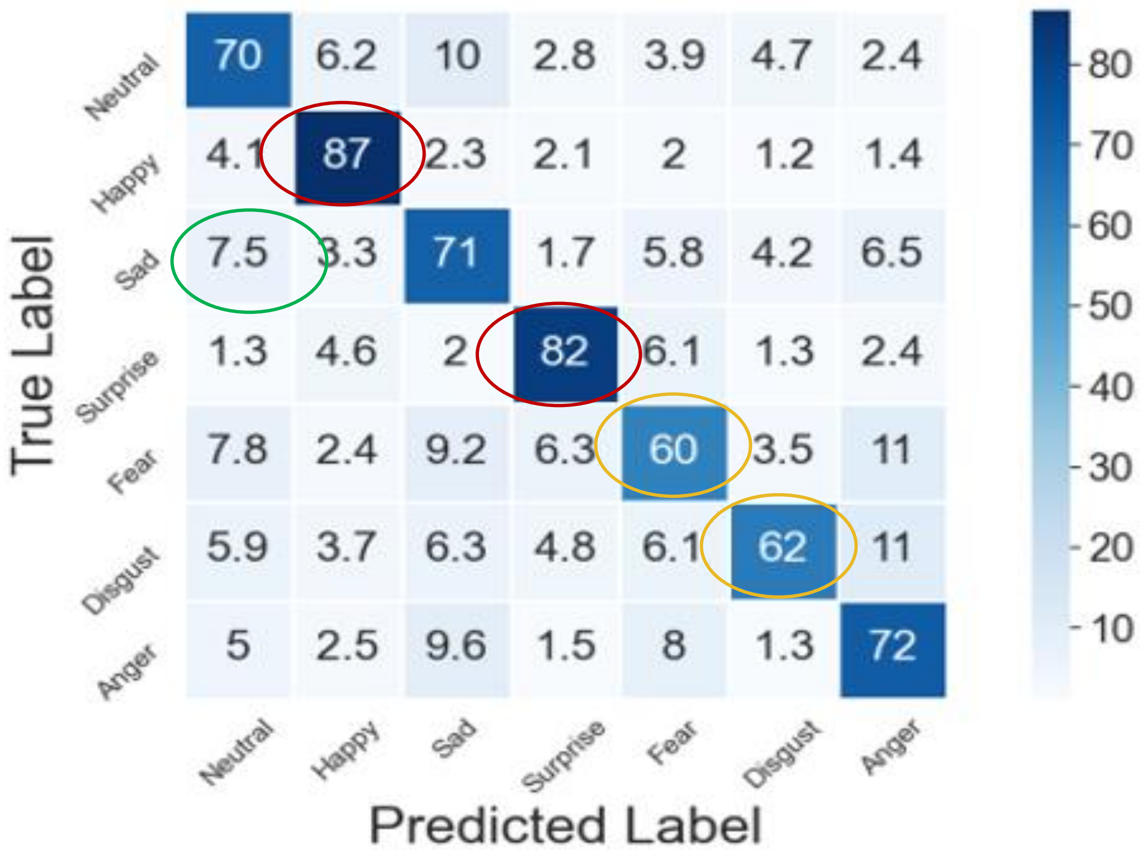
Results on various FER datasets

Findings: SAFER shows high accuracy in all six datasets which proves good generalization.

Experimental Results and Findings

Research question: Which emotions are easy, and which are difficult to identify

Effect of sample numbers	❑ Lower sample number, underfitting ❑ 'Happiness': High accuracy, 'Disgust : Low accuracy ❑ 'Happiness': 7000 samples, 'Disgust': 436 samples
Effect of related classes	❑ Share some of the facial expressions ❑ 'Disgust' and 'Anger', 'Sad' and 'Neutral' ❑ Misclassified for each other



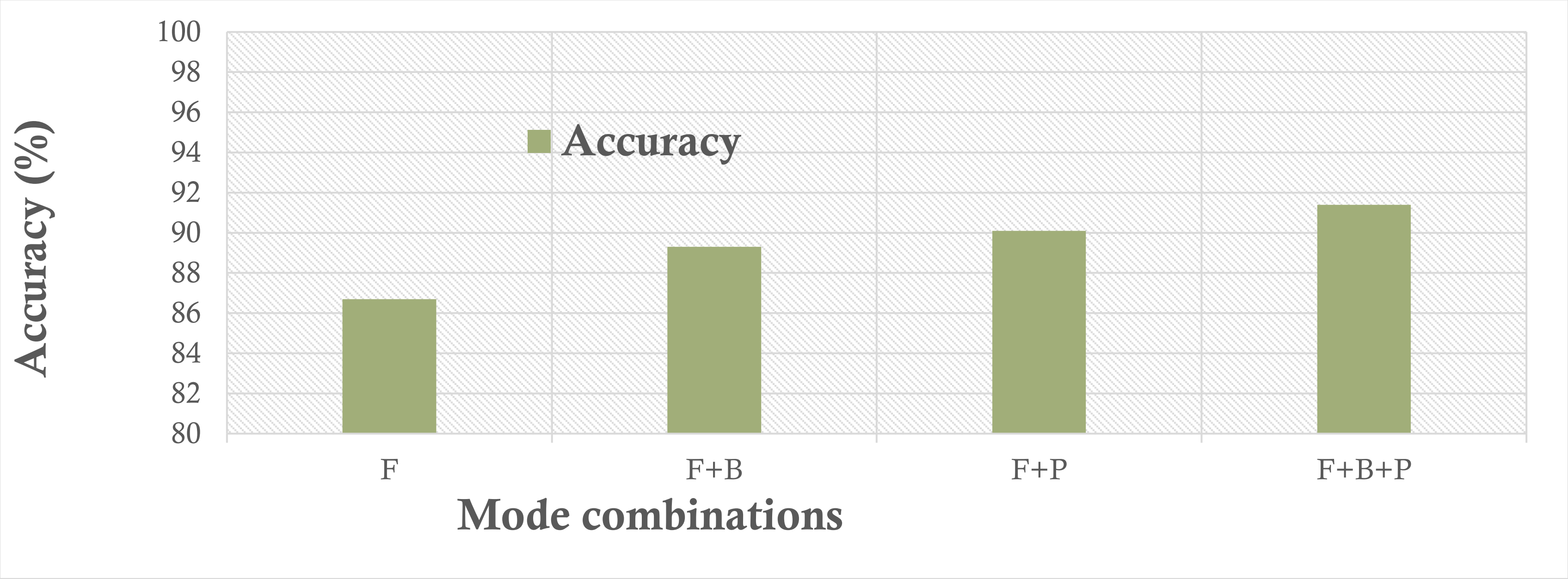
Confusion Matrix on FER-2013

Findings: Unbalanced classes and shared facial expressions degrade performance.

Ablation Study

Research question: Which combination offers best results

F: Face
B: Background
P: Place



Experiments on CAER-S dataset

Findings: Best result is achieved when F,B and P are combined.

Issues with Current Datasets

- **Bias:**
  - gender and racial
- **Quality concern**
- **No face mask**

Gender bias in FER-2013	
Class	**Sample with male subject**
Anger	70%
Happy	60%

Example of bad samples

Proposed Dataset

Research question: How can we improve the FER training?

A new dataset-

- ❑ Seven emotion classes
  - Balanced
  - 3000+ sample each
- ❑ Gender and ethnically diverse
- ❑ Section for masked sample

Research Contribution of SAFER

A novel face feature extraction module

A novel facial emotion recognition system with background and place features

A detailed evaluation framework to prove the high accuracy and generalizability

A novel dataset for FER with masked subjects

Contributions

SAFER: Improved facial emotion recognition

EMERSK: Explainable multimodal emotion recognition

CoNERS: Novelty detection and handling

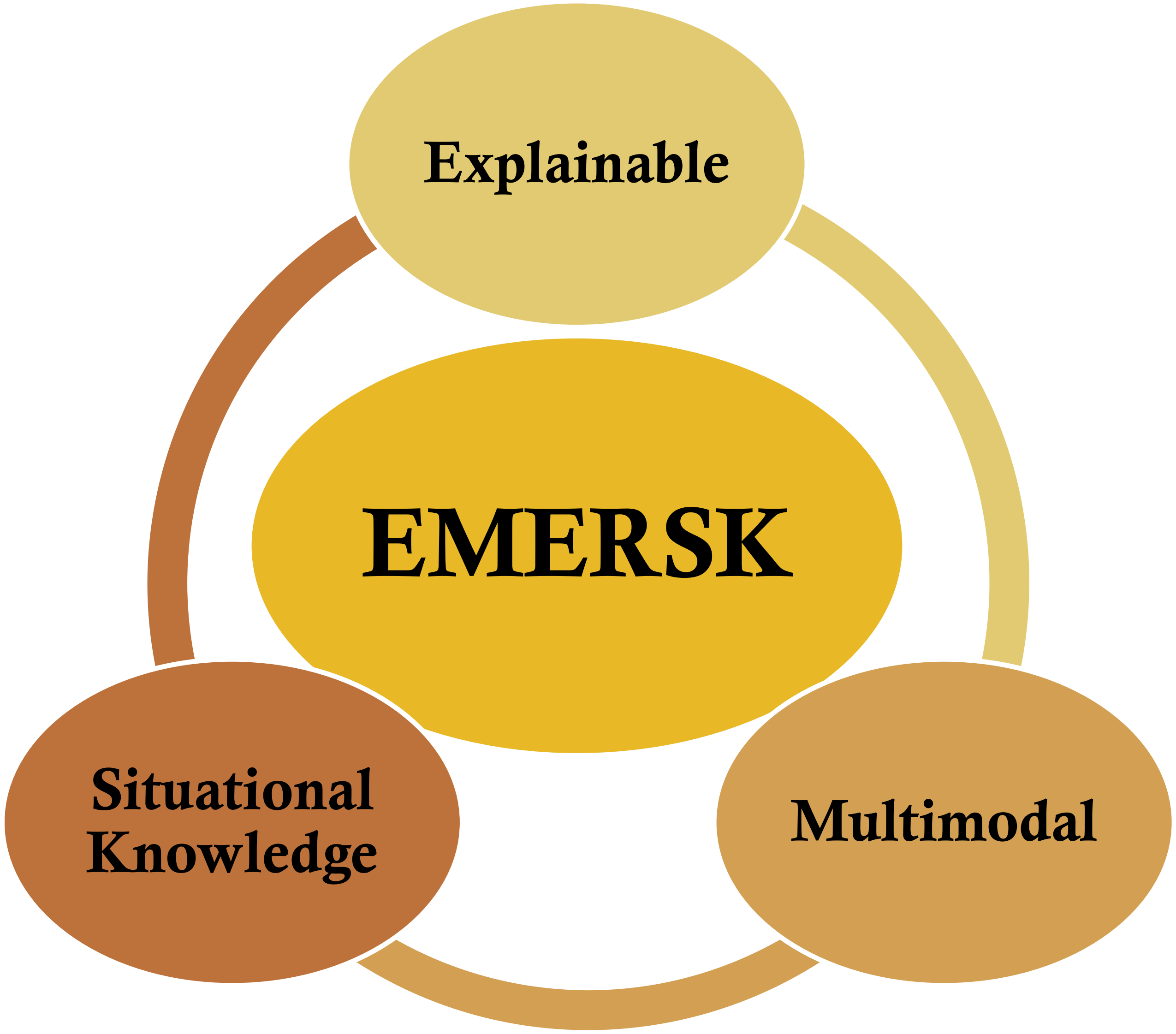
Problem Statement

Issues with the facial expression

- Face covering
- Intentional misleading

Use of multiple modalities can help!!

EMERSK



EMERSK: Architecture

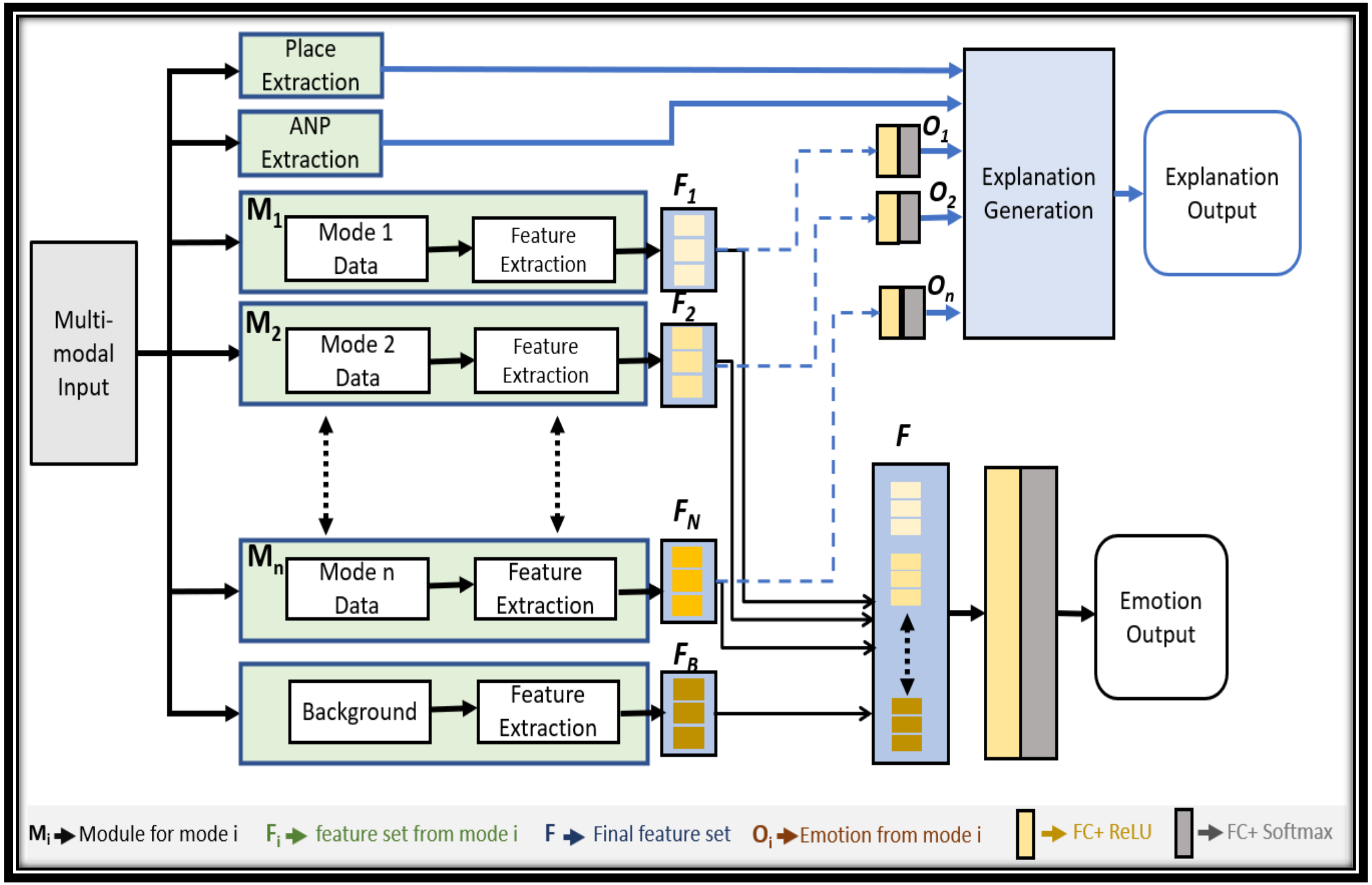
Facial Module

Posture Module

Gait Module

Background Modul

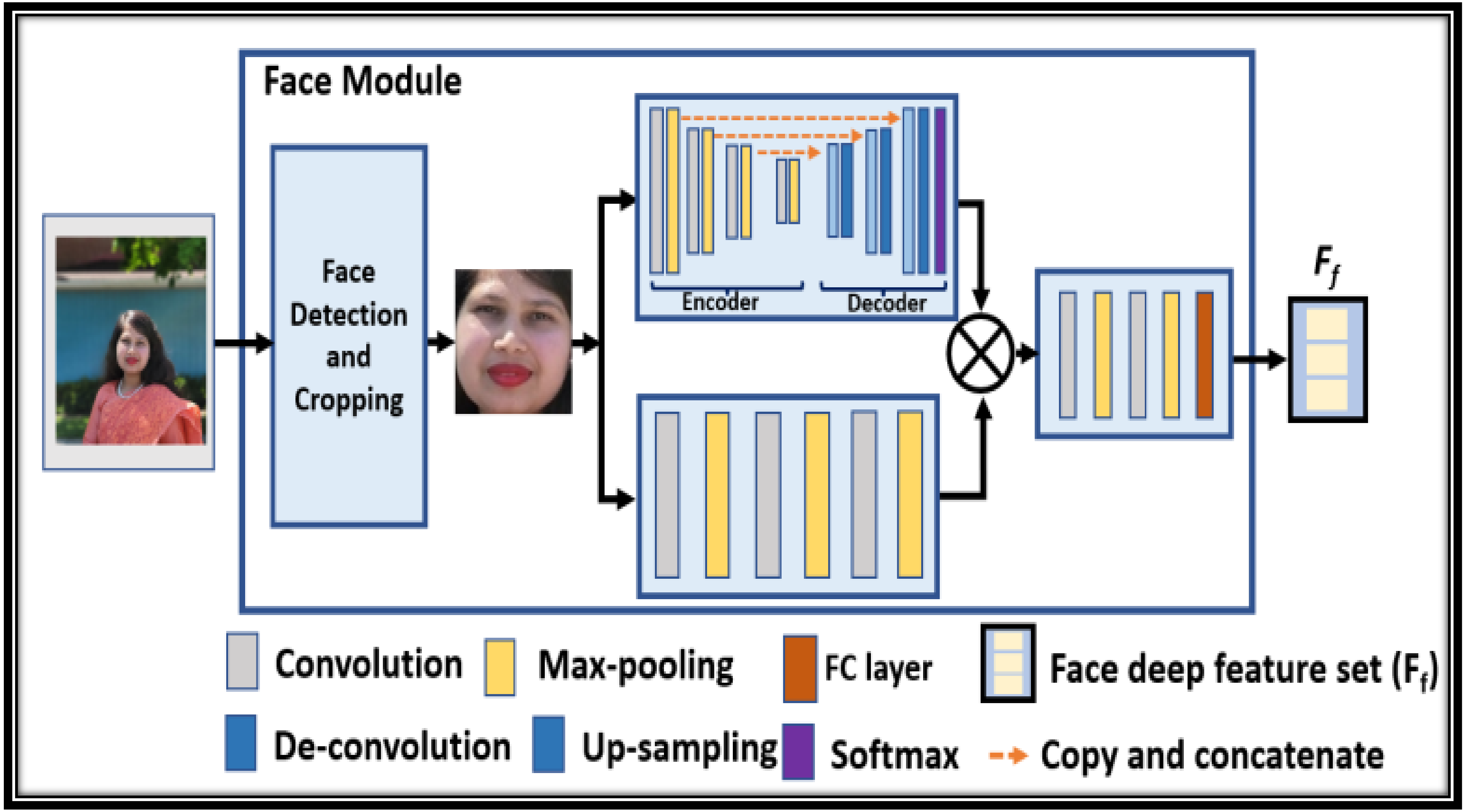
Explanation Module



Face Module

Two-stream architecture

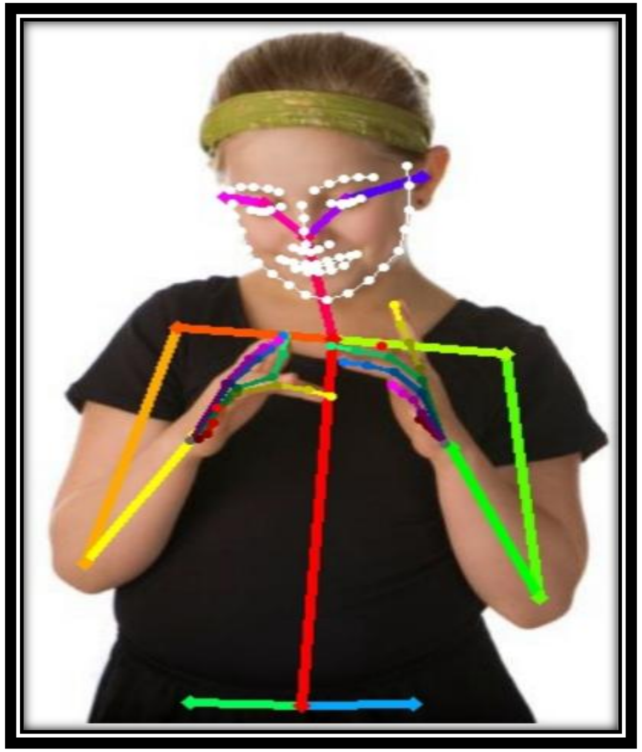
CNN	Attention based encoder-decoder network



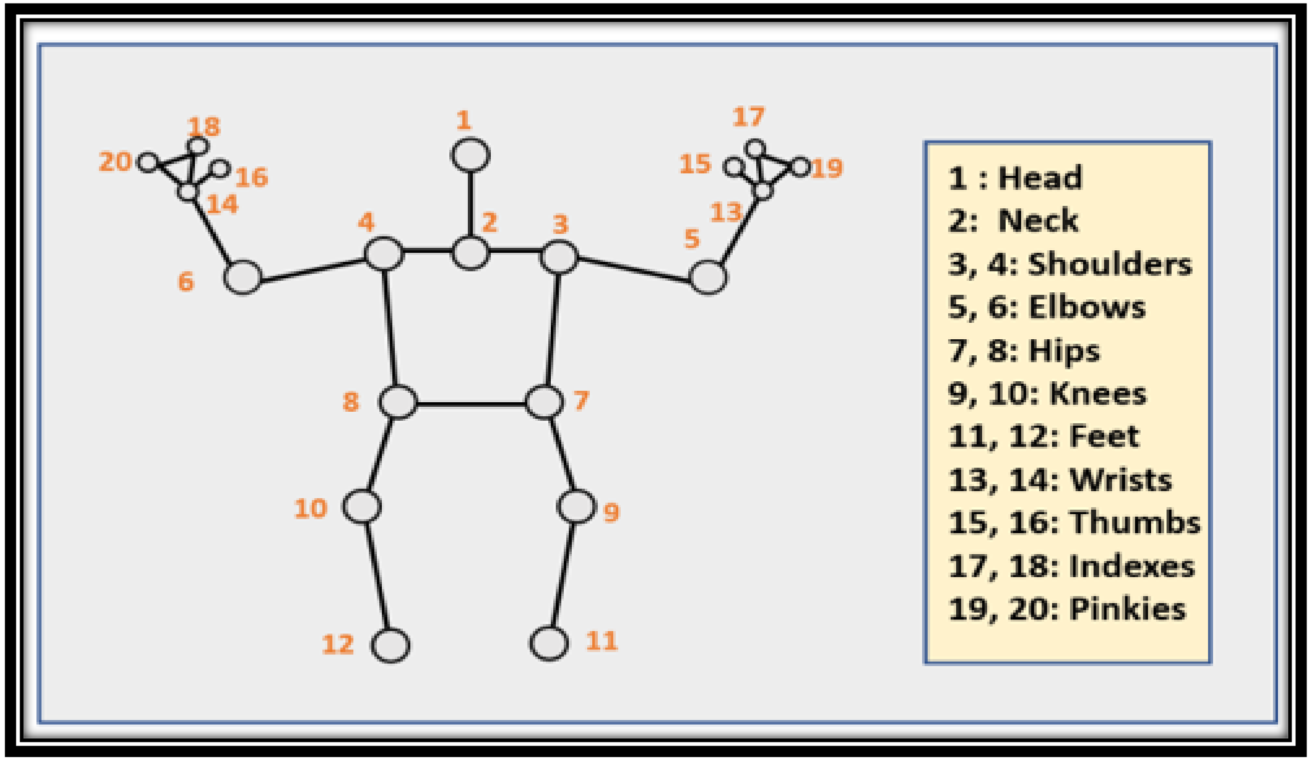
Posture Module

Body modeling and posture detection

- ❑ **Kinematic representation of human body**
  - **Collection of joints**
- ❑ **BlazePose for body point detection**



Posture example



Body points

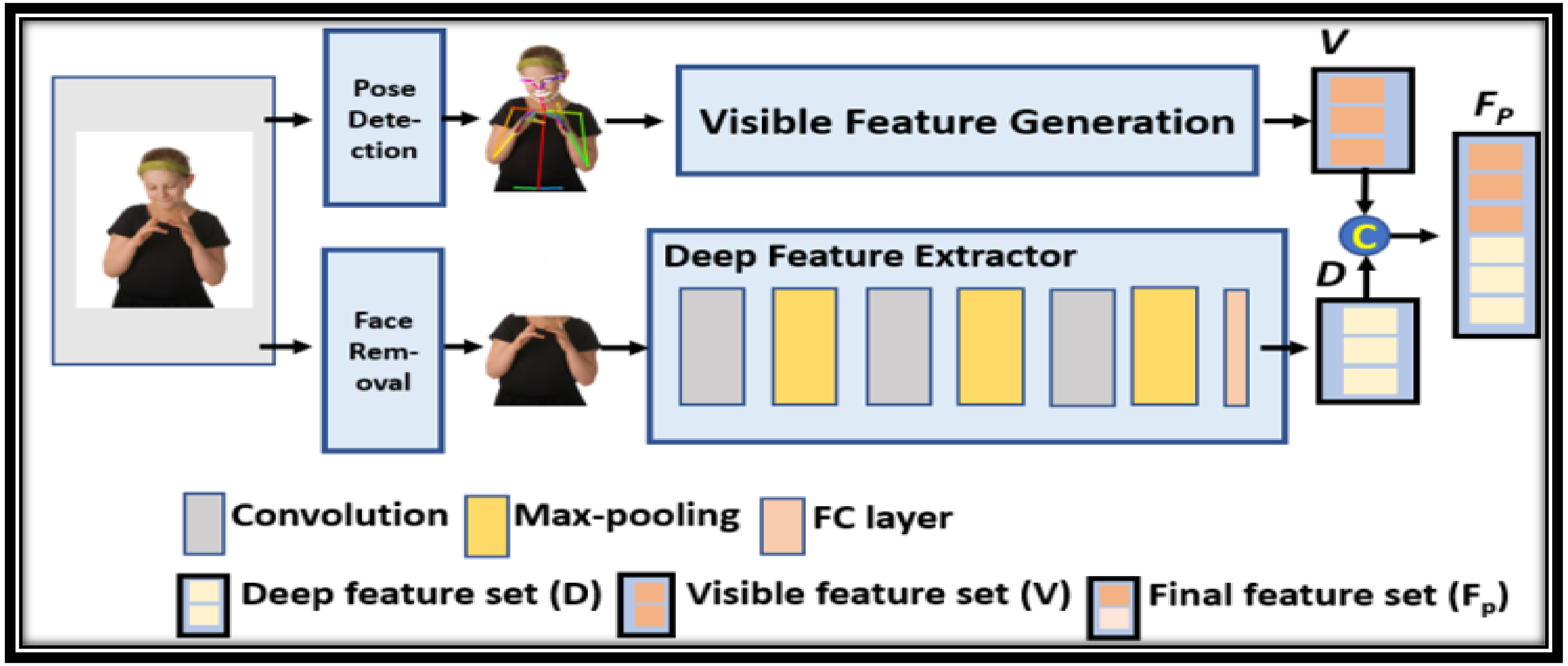
Posture Module

Visible feature generation

- Calculated from the body points in the form of distance, angle, area etc.

Deep feature generation

- Deep representation using convolutional network

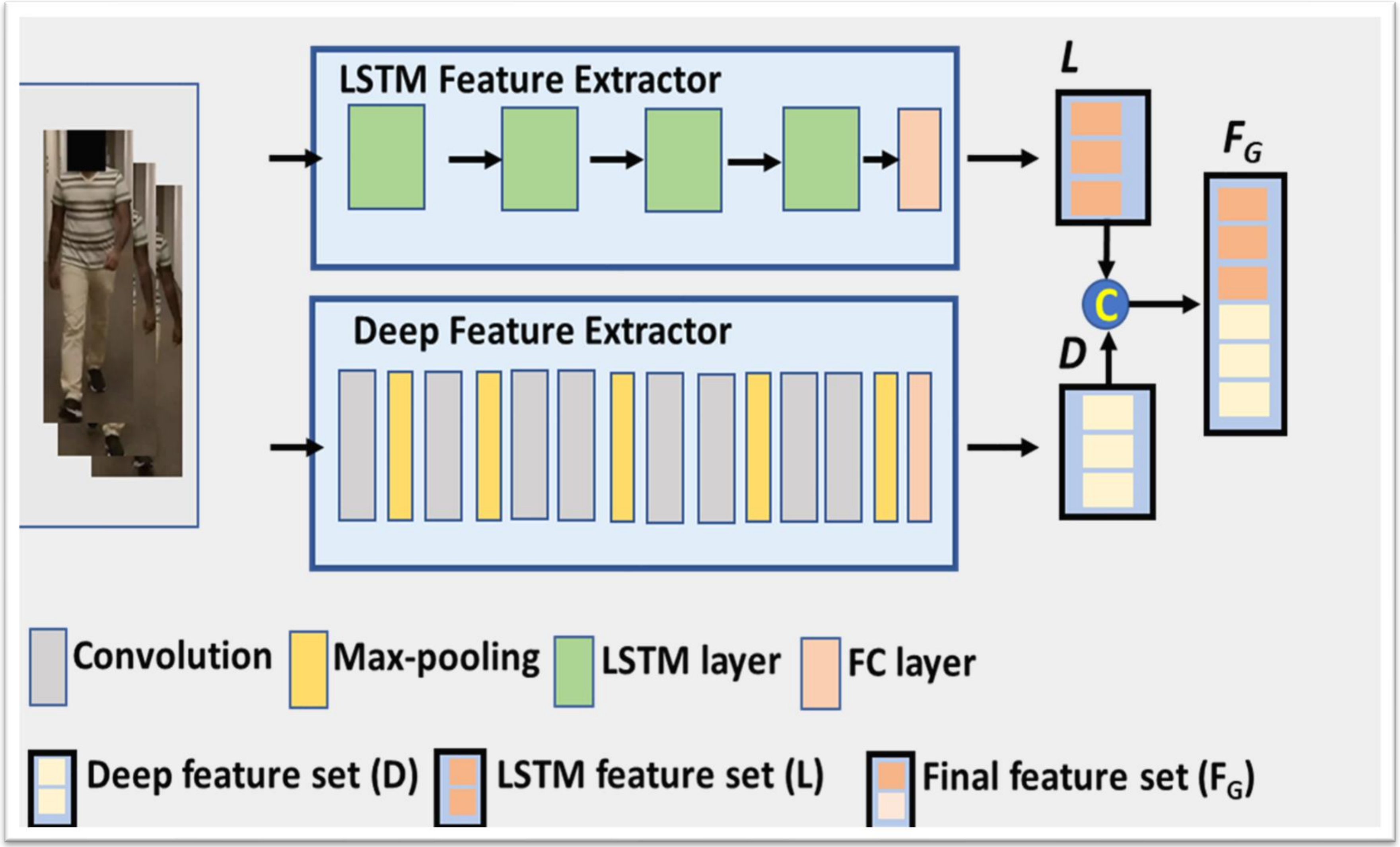


Visible feature Type	Description
Angle	At neck by both shoulders
Distance	Between right hand and hips joint
Area	Triangle between both hands and neck

Gait Module

Two-stream architecture

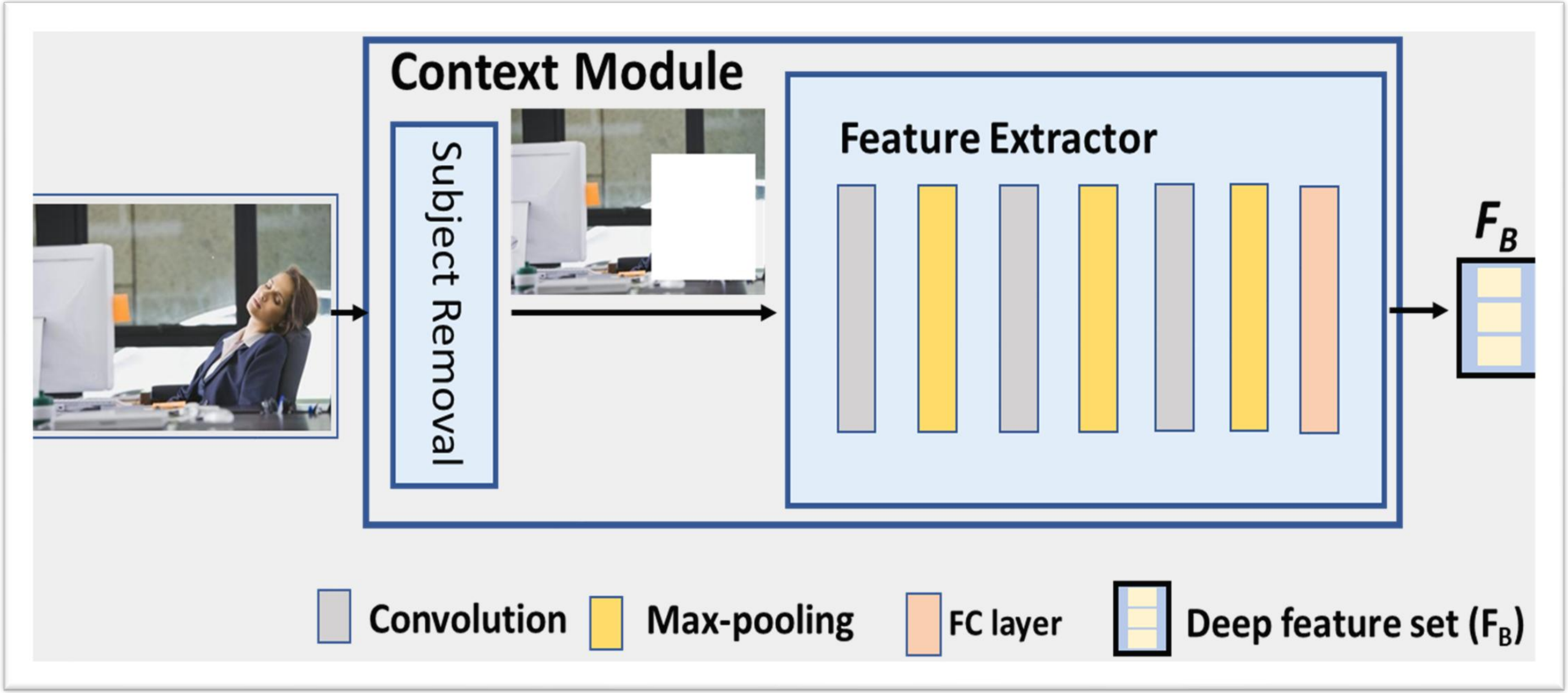
Upper stream: LSTM	Lower stream: 3D CNN



Background Module

Subject removal

Deep feature extraction



Place Module

Place dataset

Pre-trained AlexNet

Place category: “Bedroom”

Adjective-Noun Pair (ANP) Module

SentiBank 2.0: CNN based ANP classifier

Trained on one million images from Flickr

Emotion	Top ANPs
joy	happy smile, innocent smile, happy christmas
trust	christian faith, rich history, nutritious food
fear	dangerous road, scary spider, scary ghost
surprise	pleasant surprise, nice surprise, precious gift
sadness	sad goodbye, sad scene, sad eyes
disgust	nasty bugs, dirty feet, ugly bug
anger	angry bull, angry chicken, angry eyes
anticipation	magical garden, tame bird, curious bird

Colorful butterfly

Crying baby

Evaluation Setup

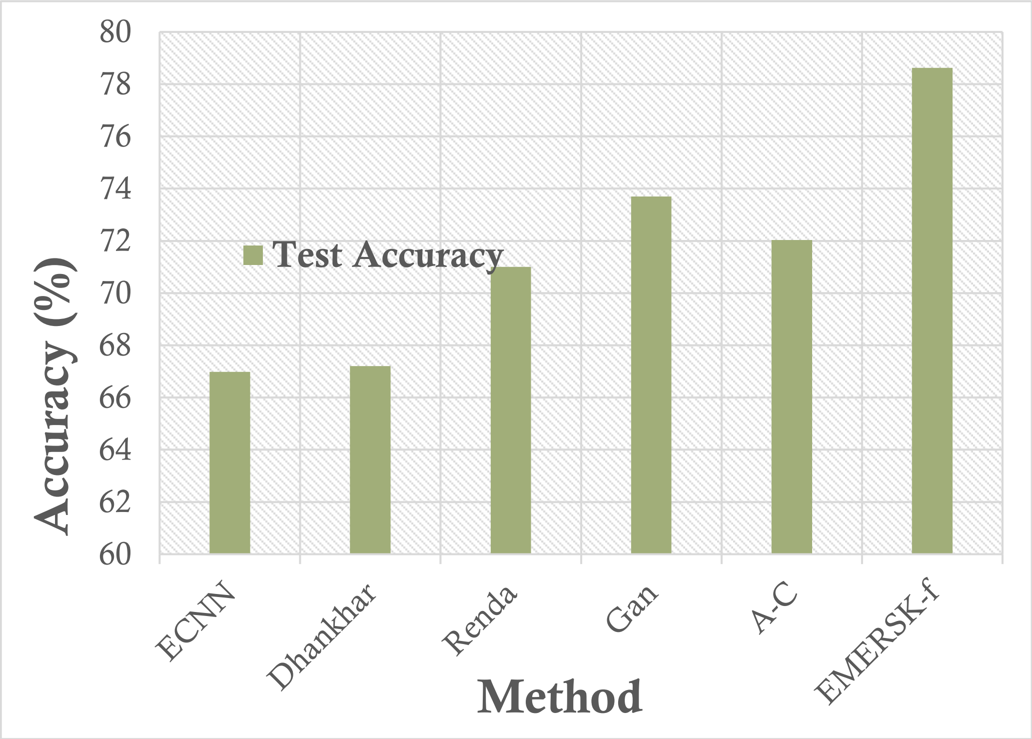
- **PC:**
  - 2.6 GHz 20 Cores Intel Xeon CPU
  - 96 GB of RAM
  - 3 NVIDIA TESLA GPUs with 24 GB of memory each
- **Dataset preparation**
  - Split into training, validation and test sets in an 80:10:10 ratio
  - Images resized to 224 × 224
  - Augmentation: cropping, rotation, brightness, and contrast adjustments
- **Evaluation Metrics**
  - Accuracy (%): $\frac{\#\text{samples correctly predicted}}{\#\text{total samples}} \text{x} 100$

Evaluation Setup: Similar Works

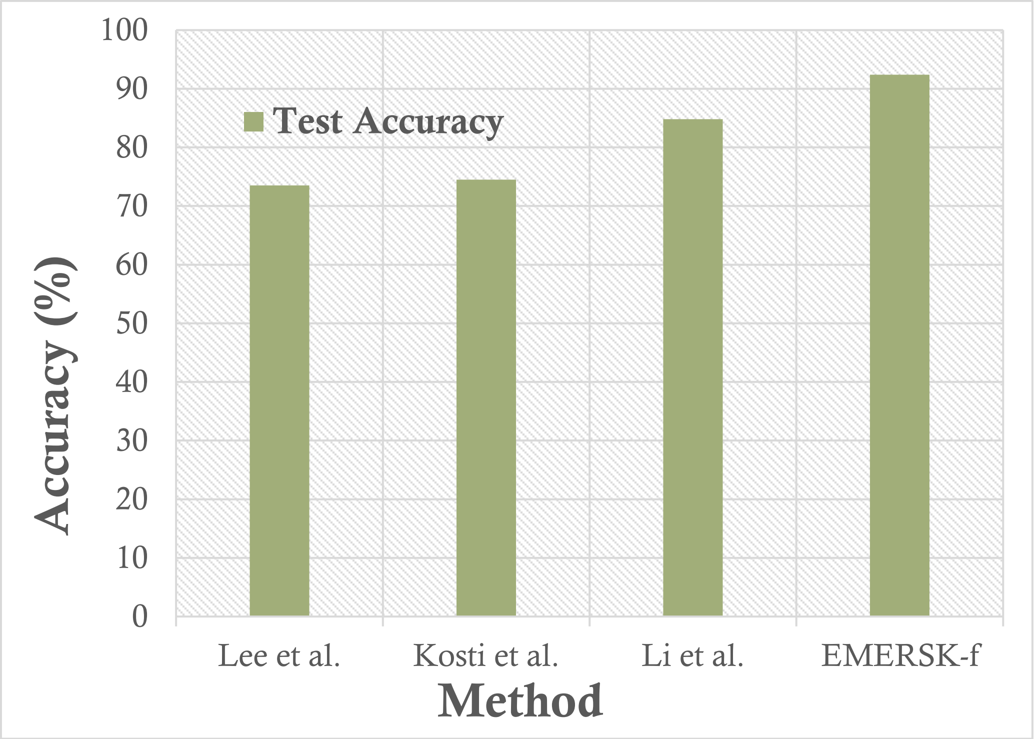
Name	Method	Limitations	Explainable?
CMEFA [52]	Broad deep learning fusion network (BDFN) on Face and Posture	Limited evaluation	No
Bhatia et al. [57]	Layered LSTM on gait	Gait mode only	No
Kosti et al. [5]	Dual stream CNN on body and background	Considers whole body as a single mode	No
Lee et al. [6]	Two stream CNN with adaptive fusion on face and background	Posture and gait not considered	No
Santosh et al. [63]	ConvLSTM	Not modular, treats the video as a single mode	No
Tahghighi et al. [47]	HOG-KLT+ SVM	Considers whole body as a single mode	No

Experimental Results and Findings: Face Module

Research question: Can face module perform standalone?



Results on FER-2013

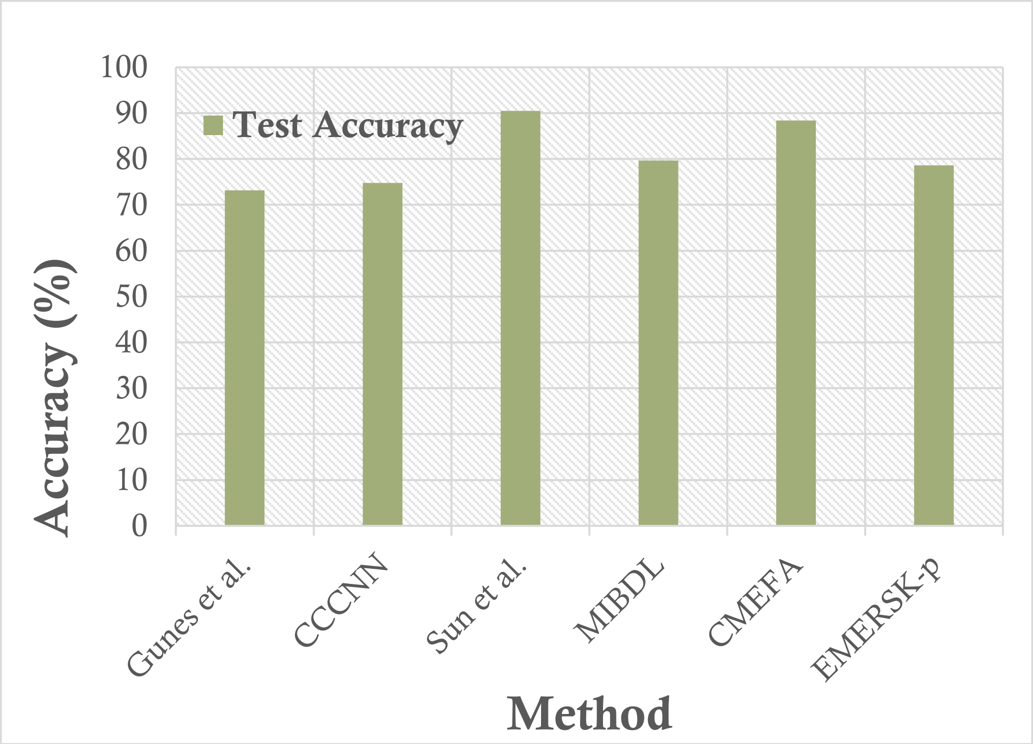


Results on CAER-S

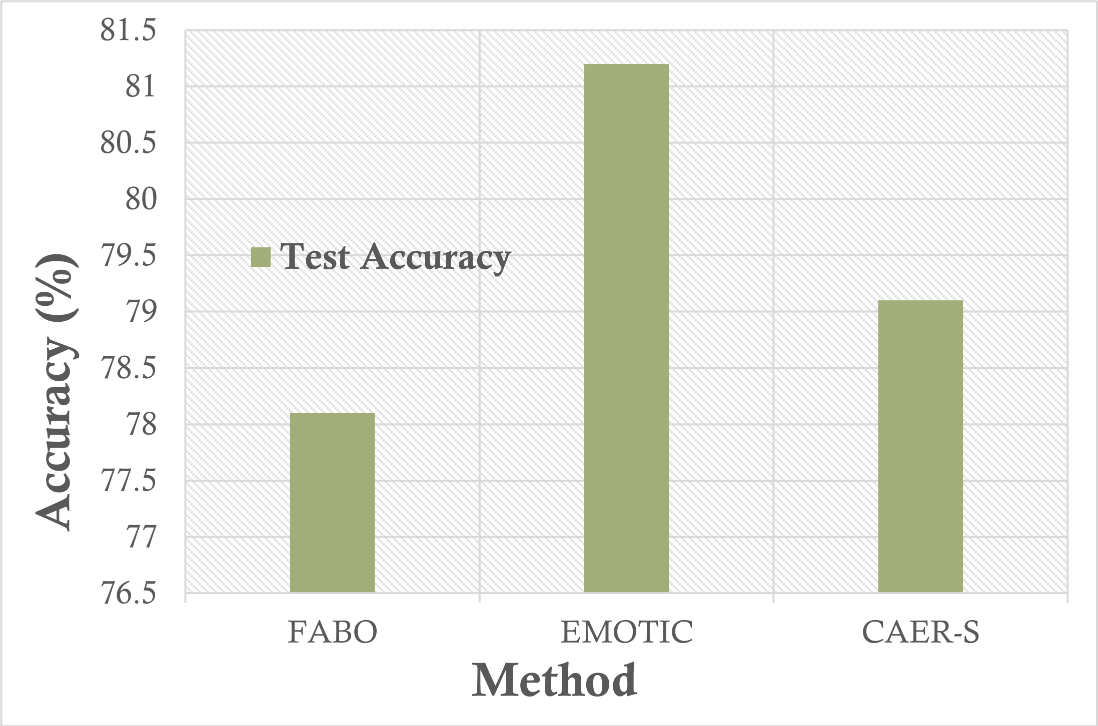
Findings: Face module provide superior standalone performance.

Experimental Results and Findings: Posture Module

Research question: Can posture module perform standalone?



Results on FABO dataset

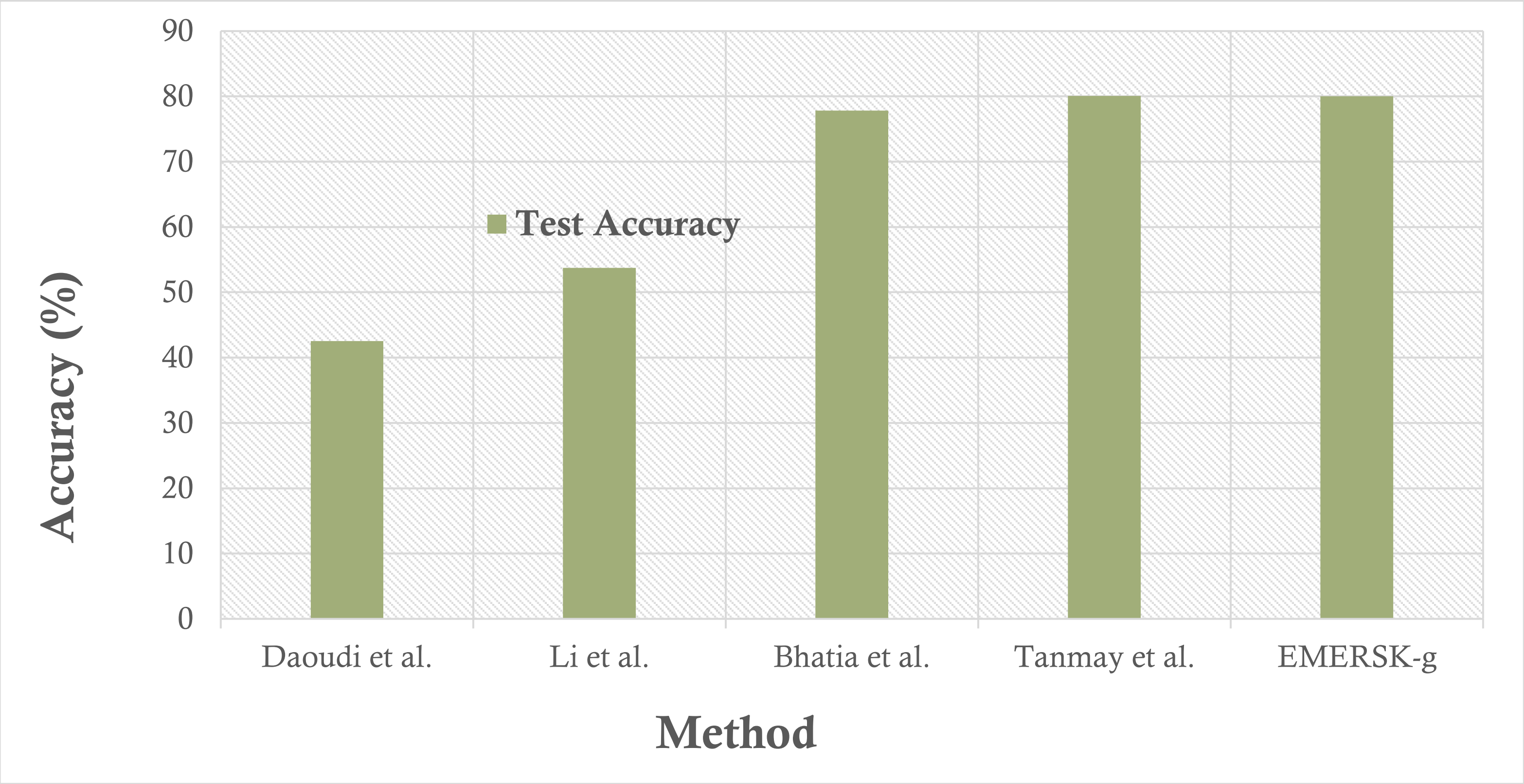


Results on various datasets

Findings: Posture module provide comparable and generalized standalone performance.

Experimental Results and Findings: Gait Module

Research question: Can gait module perform standalone?

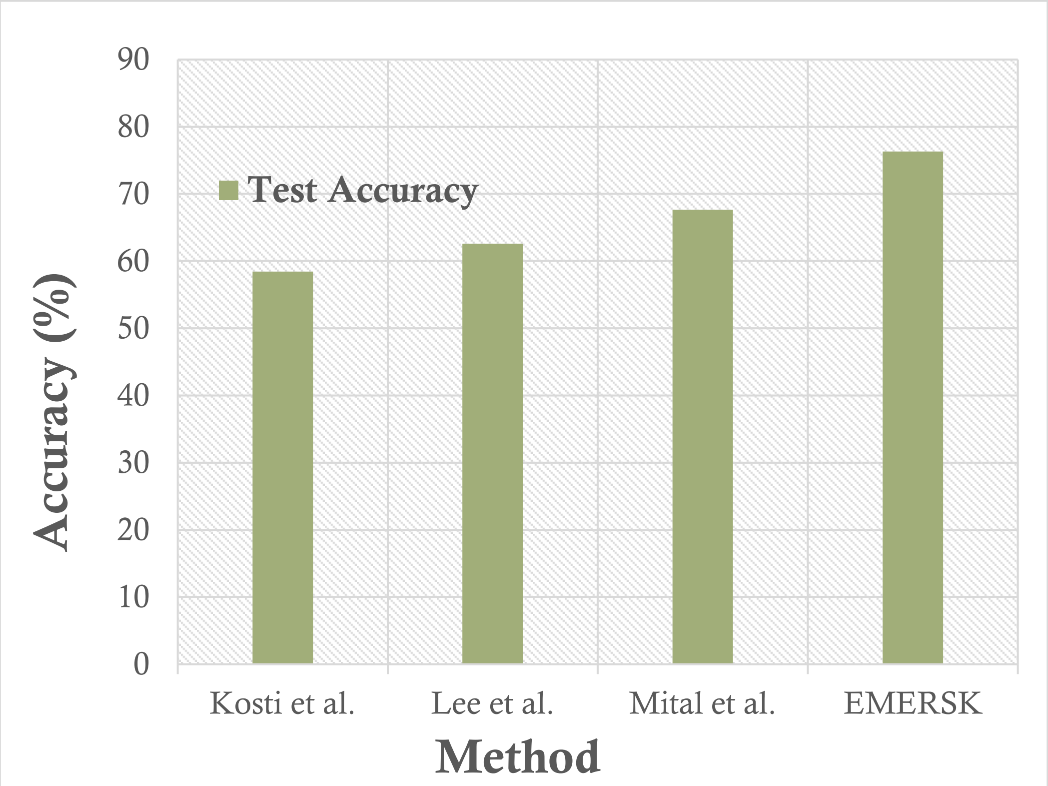


Results on FABO dataset

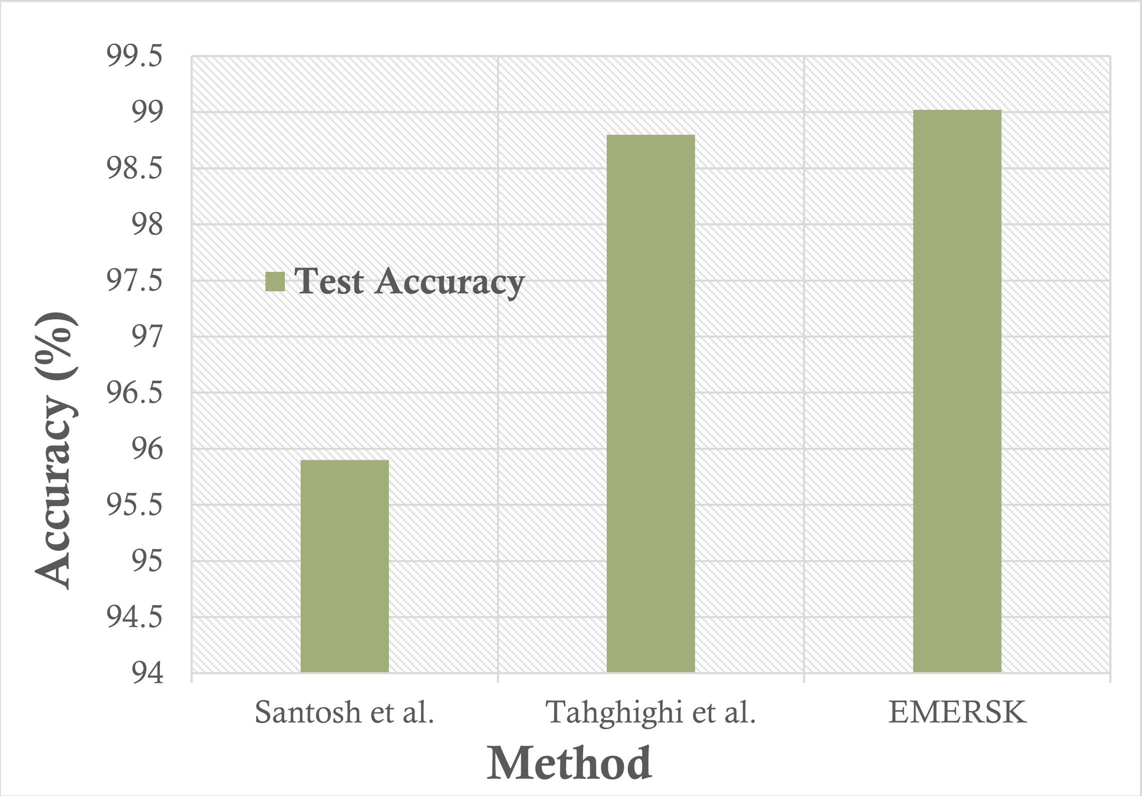
Findings: Gait module provide superior standalone performance.

Experimental Results and Findings: Multimodal Operation

Research question: Does multimodal improve performance?



Experiments on GroupWalk dataset

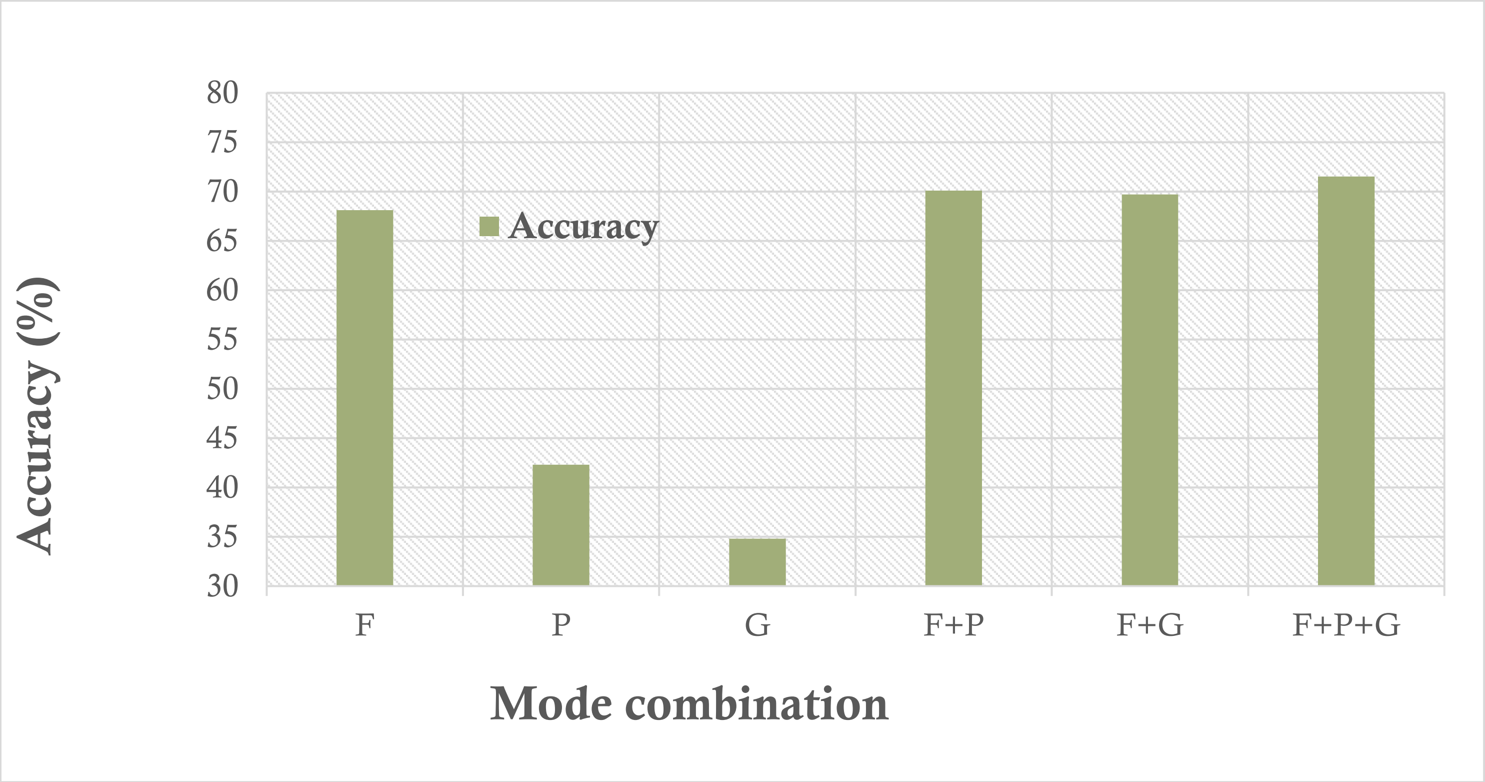


Experiments on GEMEP dataset

Findings: Multimodal provides superior performance than state-of-the-arts.

Ablation Study

Research question: What is the best combination of the modes?

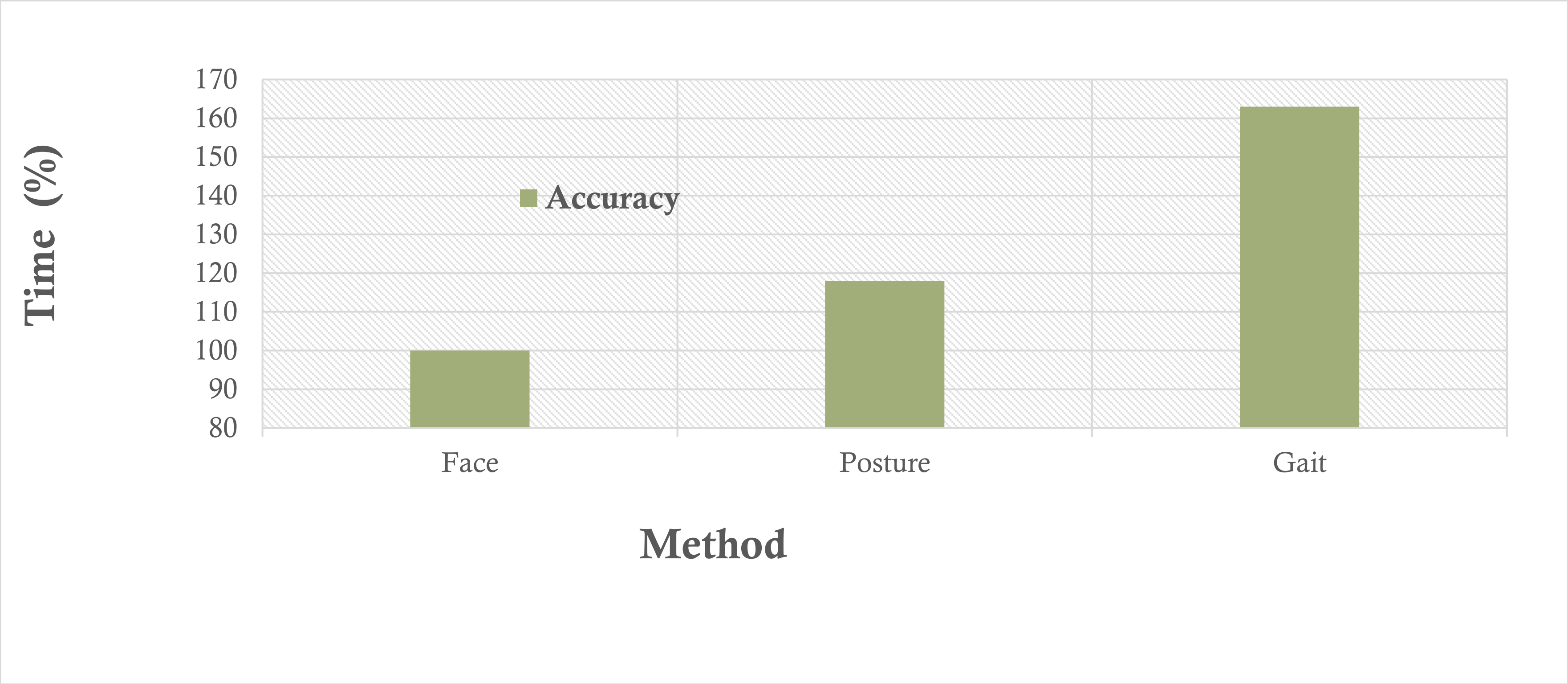


Experiments on GroupWalk dataset

Findings: Face is the most expressive mode and multimodal beats standalone methods.

Computational Cost

Research question: What is the computational cost of going multimodal?



Results on EWALK dataset

Findings: Face is the fastest mode and gait is the slowest.

Explanation Generation

Research question: How do we explain the output?

- Individual mode result
- Place type
- Adjective-Noun pair
- Average emotion

Explanation: "Emotion output is "happiness". The place is "nursery_classroom", it is "positive" environment, with "creative_work" and "smiling_kid". Subject face is: "happy",and posture is: "happy"".

Research Contribution of EMERSK

A modular architecture for emotion recognition from multiple modes

A novel approach for situational knowledge generation

A novel approach for explanation generation

Contributions

SAFER: Improved facial emotion recognition

EMERSK: Explainable multimodal emotion recognition

CoNERS: Novelty detection and handling

What About Novelty?

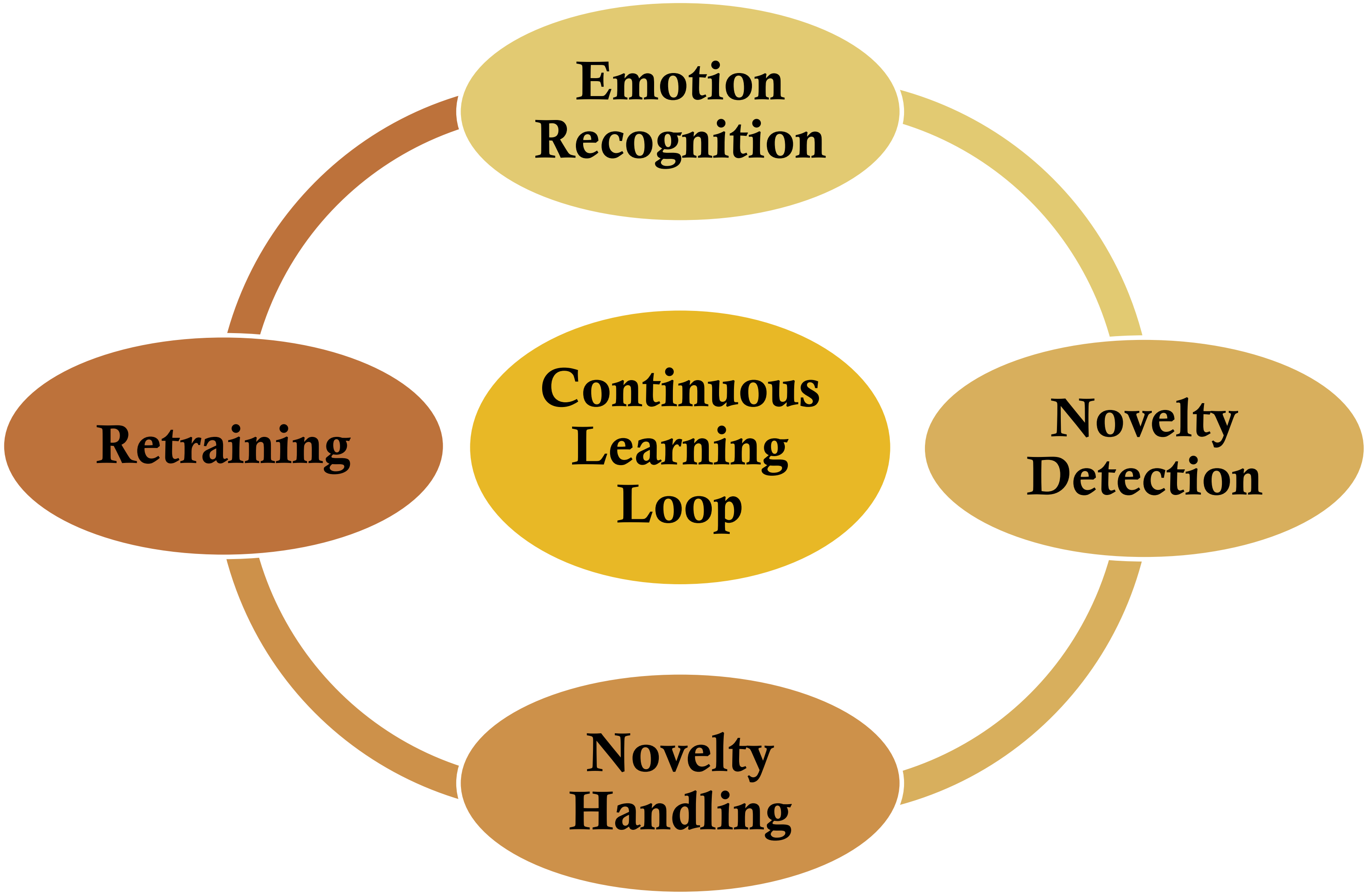
Research question: What happens with frequent novel or unexpected samples?

Novelty is defined as a new or unusual instance that deviate from the expected norm!

Example of Novelty: A rhino freely roaming the streets of west Lafayette

Need to detect and handle novelty

CoNERS: Continuous Learning Based Novelty Aware Emotion Recognition System



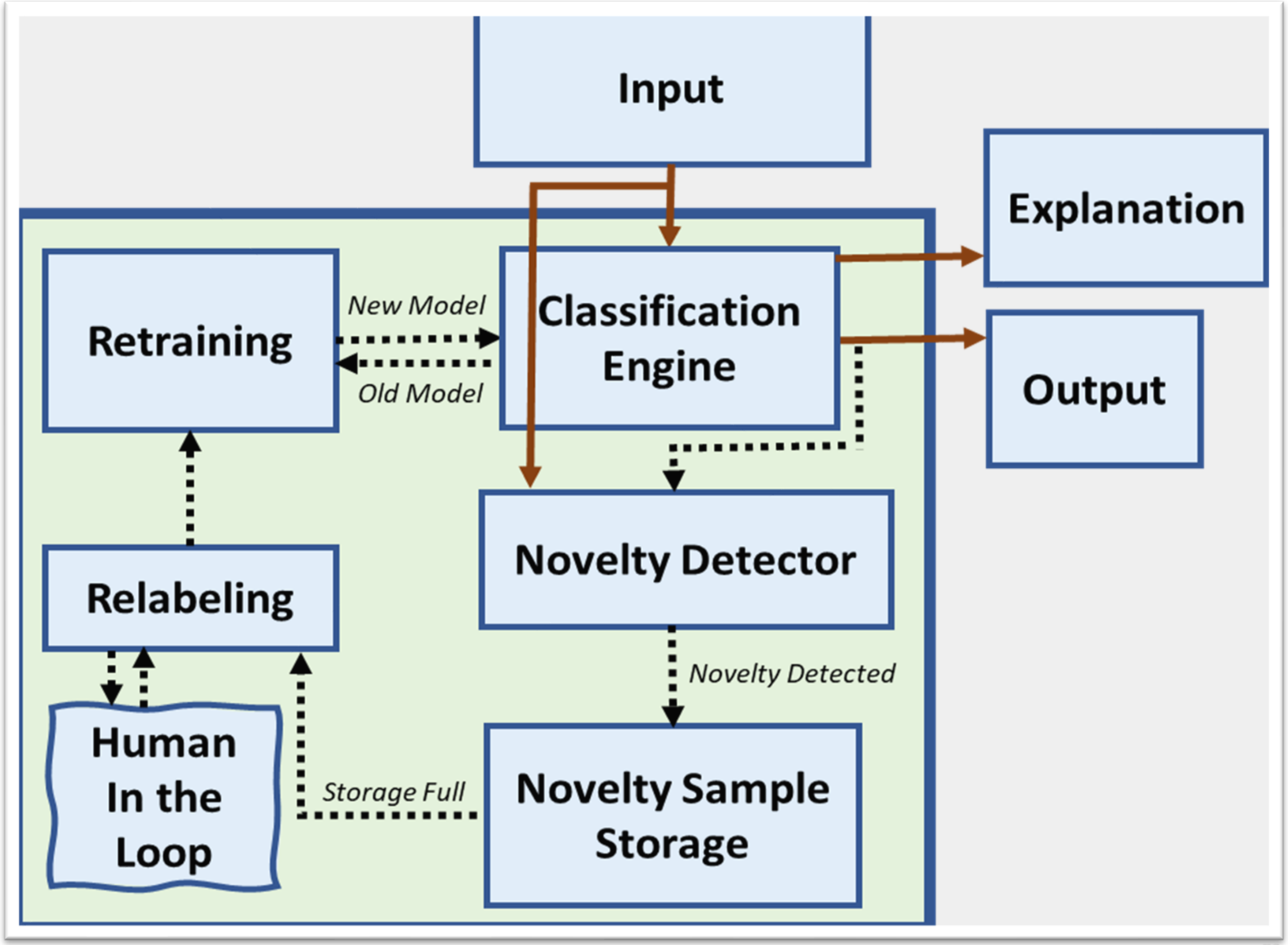
CoNERS: Architecture

Classification engine

Novelty detector

Re-labeler

Re-trainer



Classification Engine

Recognizes emotion and generates explanation

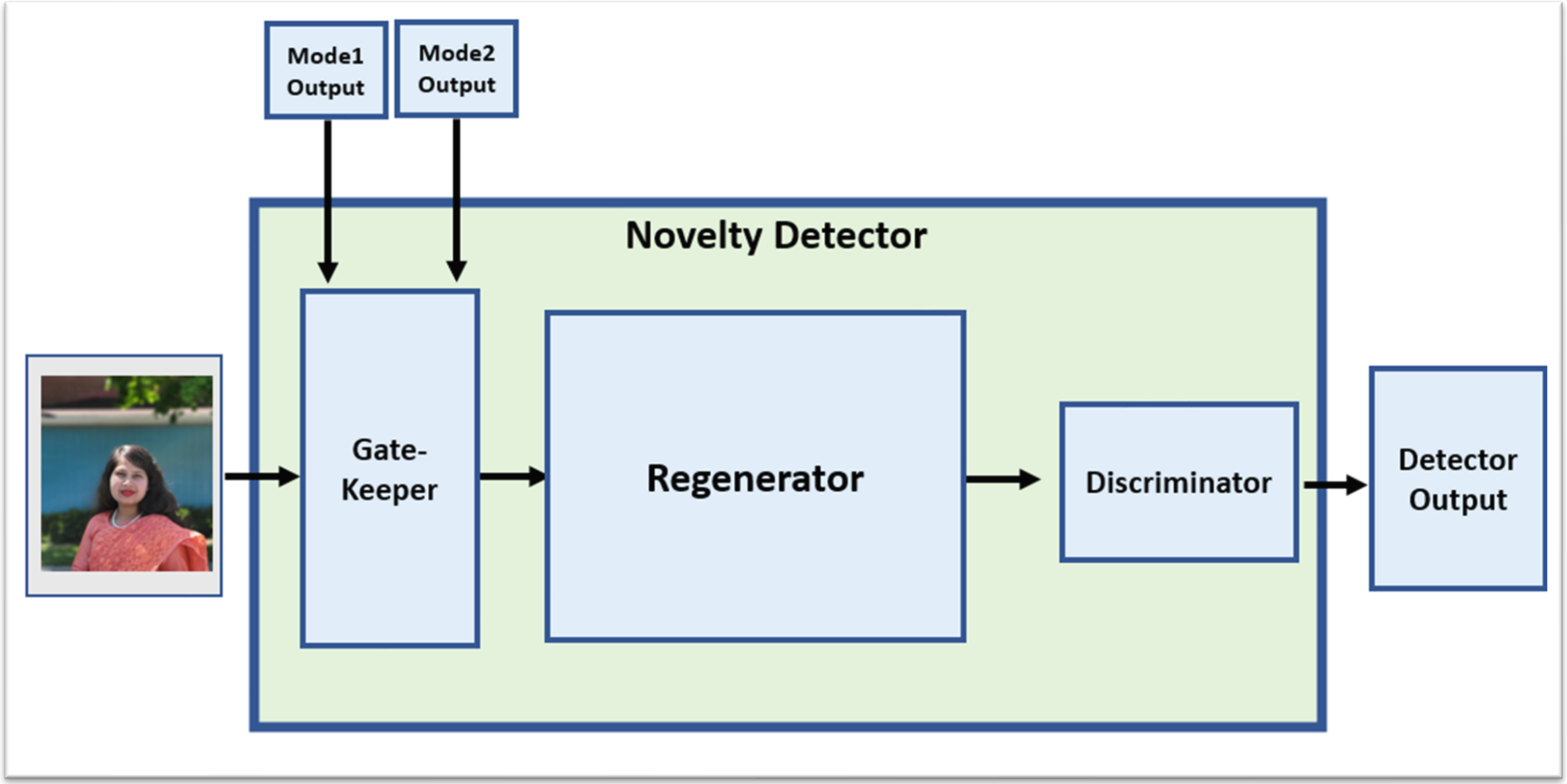
Any multimodal classification model such as EMERSK can be used

Novelty Detector

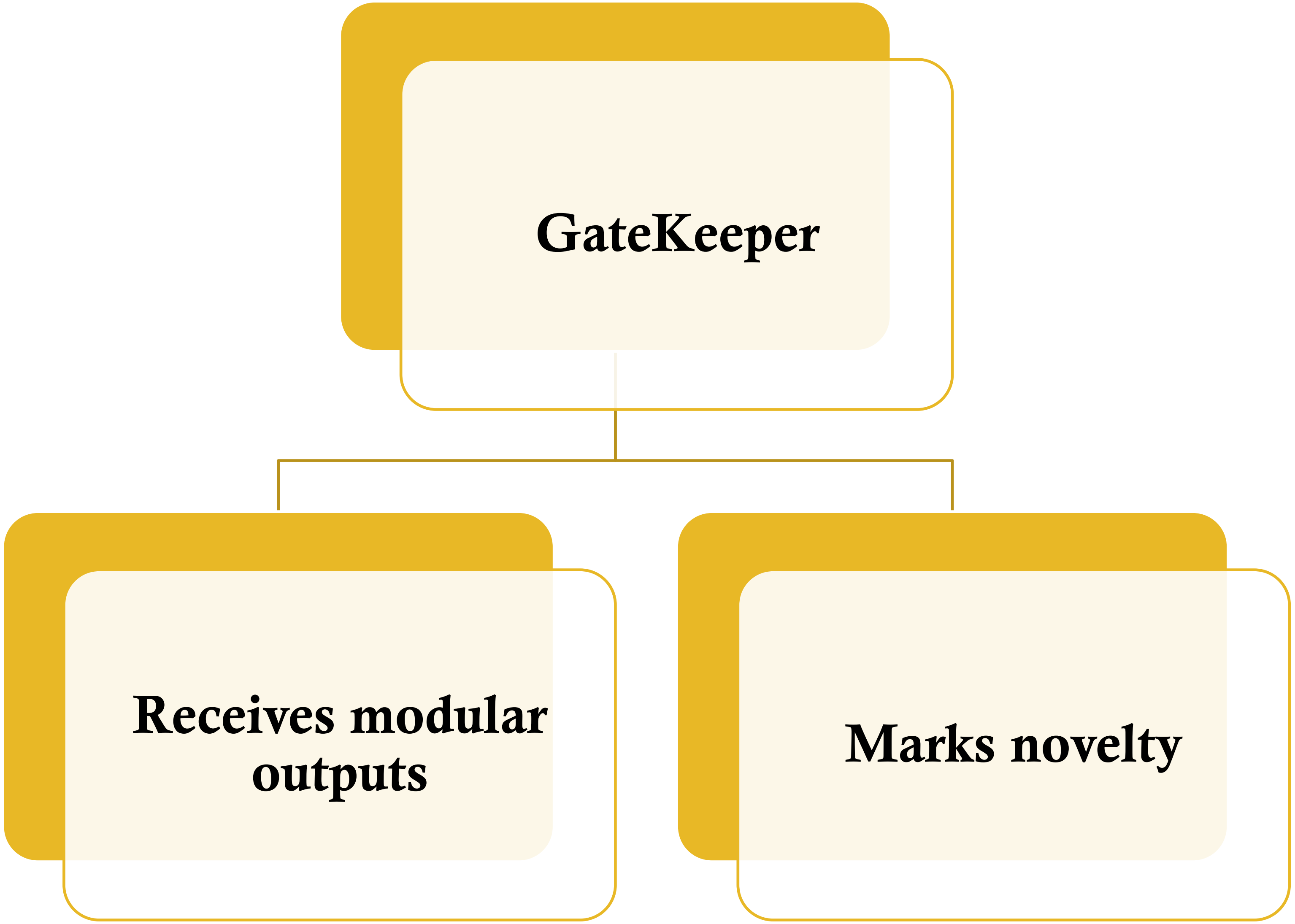
GateKeeper

Regenerator

Discriminator



GateKeeper



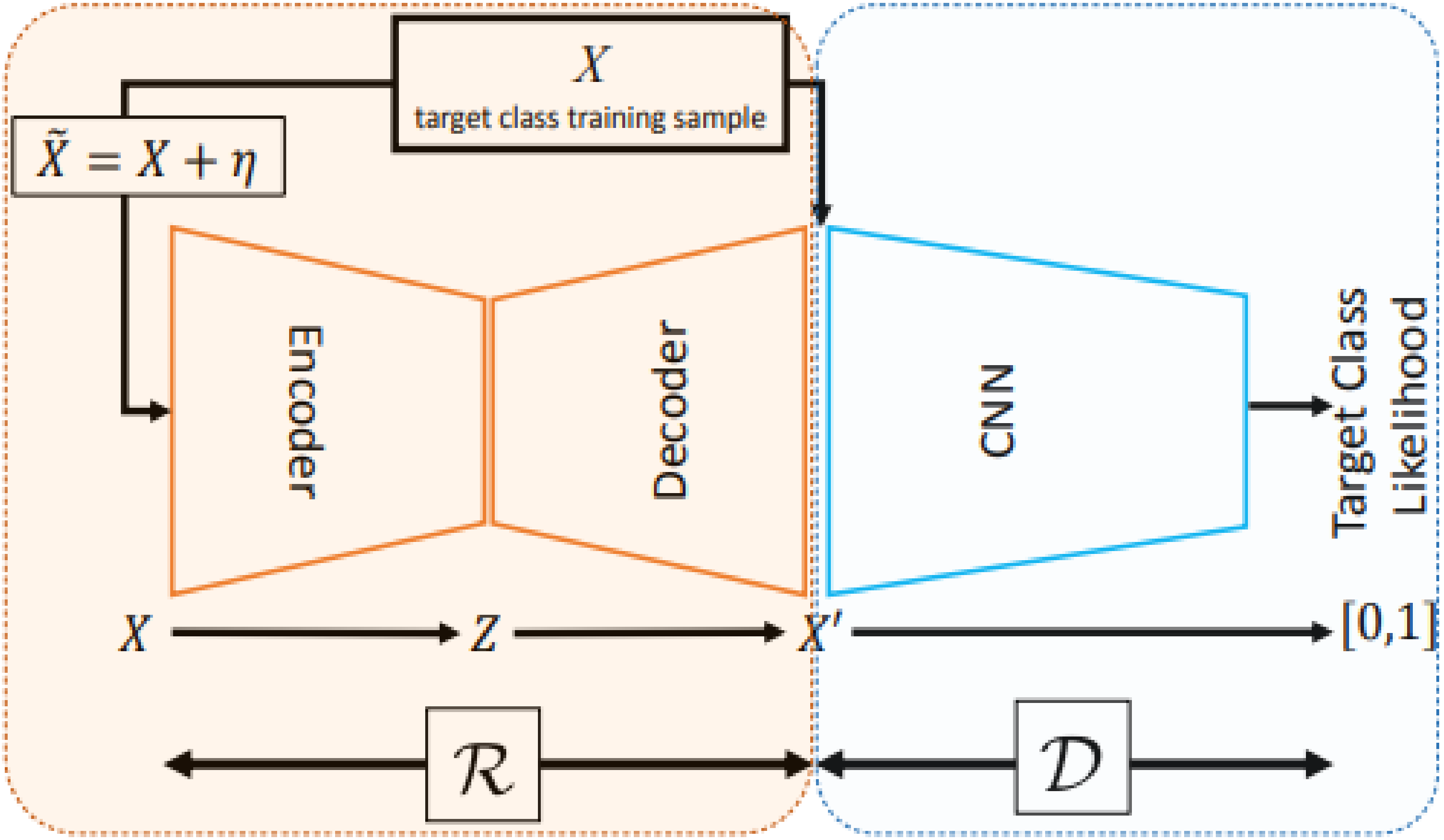
Regenerator and Discriminator

Autoencoder based regenerator and CNN based discriminator

Encoder compresses the sample and Decoder reconstructs the sample

Reconstruction training with reconstruction loss

Adversarial training of R+D with minimax loss function



Reconstruction Loss: $||X - R(X)||^2$

Minimax Loss: $E[\log(D(X))] + E[1 - \log(D(R(Z)))]$

Novelty Handling

Relabeling

- ❑ Periodic
- ❑ Label correction
- ❑ New class creation

Retraining

- ❑ Periodic
- ❑ Human in the loop

Evaluation Setup: Similar works

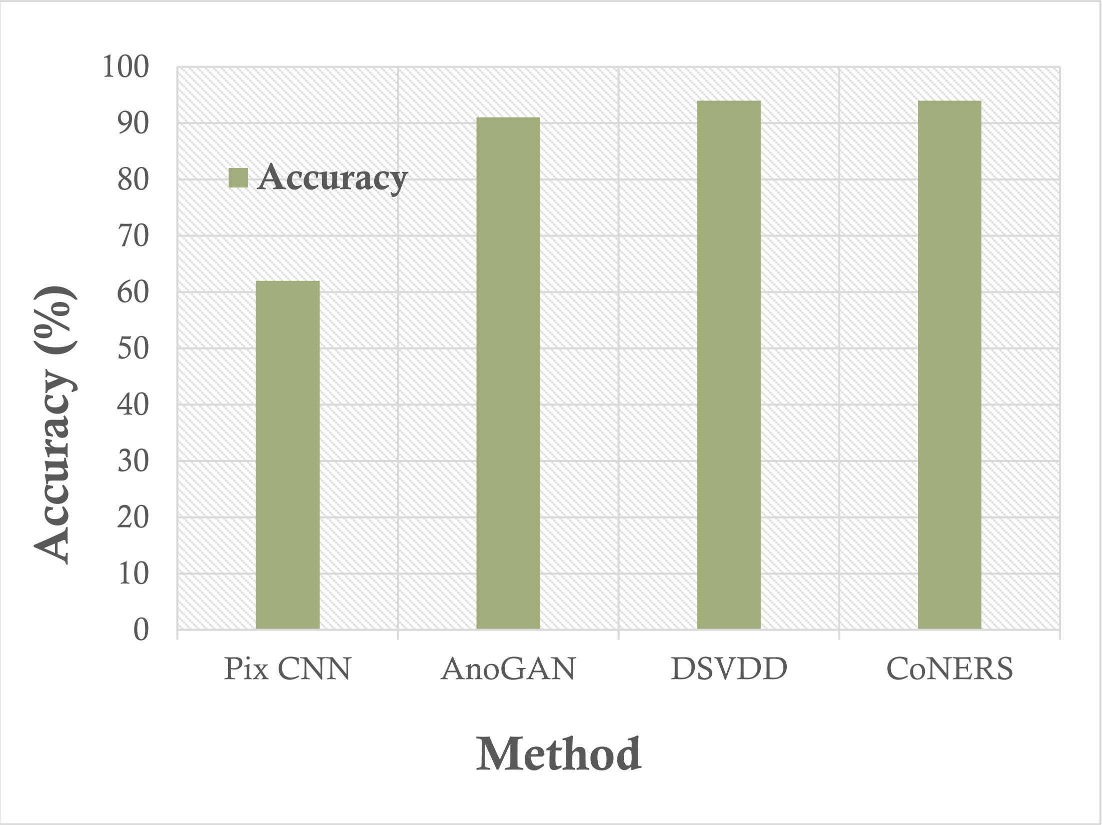
Name	Method	Limitations	Continuous learning loop?
Pix CNN [64]	Gated PixelCNN	Low accuracy	No
AnoGAN [65]	GAN+ coupled mapping	Inefficient	No
DSVDD [66]	Kernel-based one-class classification + minimum volume estimation	Limited generalization	No

Experimental Results and Findings: Novelty Detection

Research question: Is our detector reliable?

- ❑ **Trained in MNIST dataset**
- ❑ **50% samples of the test set are novelty**
- ❑ **Example: Digit “1” to “8” inliers and “9” novelty**

MNIST samples



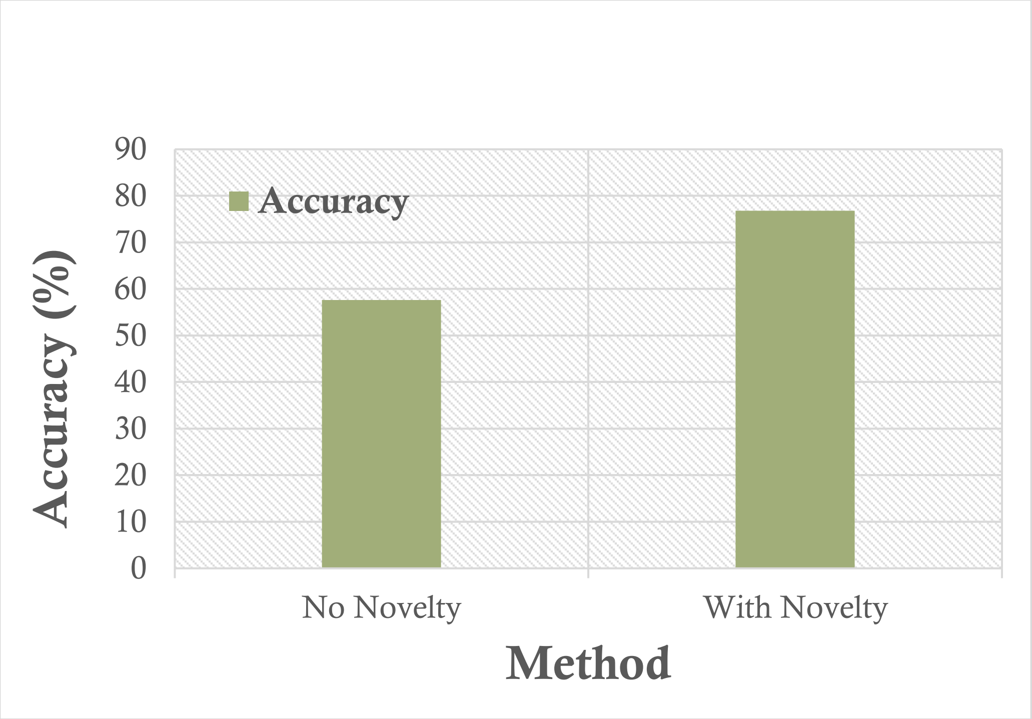
Results on MNIST dataset

Findings: Our novelty detector offers superior detection capability.

Experimental Results and Findings: Retraining

Research question: Does our system improves performance?

- **<u>First cycle (no novelty):</u>** Regular model with 20% novelty samples in the test set---→ **Degraded accuracy!!**
- Samples detected using the detector and model retrained
- **<u>Second cycle (with novelty):</u>** Updated model--→ **Improvement!!**



Results on FER-2013 dataset with 20% Novelty samples

Findings: In situations with frequent novel samples, our method can adapt and offer improved performance!

Research Contribution of CoNERS

Formalization of the novelty in
the automatic emotion recognition task

An adversarially trained auto-encoder based detector for novelty detection

A system that addresses novelty in a continuous learning manner for emotion recognition

Conclusions and Future Works

Conclusion

Proposed SAFER, a novel system for emotion recognition from facial expressions

Proposed EMERSK, a multimodal emotion recognition for additional reliability and explainable output

Proposed CoNERS, a novelty-aware emotion recognition system for real world situation

Future Works

Enhanced multimodal fusion

Anxiety and depression detection

Fast and light-weight system building for real time operation

Addressing the ethical, security and privacy issues

Acknowledgement

DARPA

Ford

NORTHROP GRUMMAN

Special Thanks

PURDUE UNIVERSITY®

Department of Computer Science

Thanks Everyone!

Extra Slides

Publications

Mijanur Rahaman Palash, Bharat Bhargava, **SAFER: Situation Aware Facial Emotion Recognition**, under review, ***Elsevier Artificial Intelligence 2023***

Mijanur Rahaman Palash, Bharat Bhargava, EMERSK -Explainable Multimodal Emotion Recognition with Situational Knowledge, accepted for publication with minor revisions in ***IEEE Transaction on Multimedia 2023***

Mijanur Rahaman Palash, Bharat Bhargava, Continuous Learning Based Novelty Aware Emotion Recognition System, ***AAAI Spring Symposium 2022***

Mijanur Rahaman Palash, Voicu Popescu, Amit Sheoran and Sonia Fahmy, CoRE- Non-Linear 3D Sampling for Robust 360 Degree Video Streaming ***IEEE INFOCOM 2021***

Mijanur Rahaman Palash, Bharat Bhargava, CoNERS: Continuous Learning Based Novelty Aware Emotion Recognition, under review, ***IEEE Transaction on Neural Networks and Learning Systems 2023***

Reference:

[1] "The Seven Universal Emotions We Wear on Our Face," CBC, accessed on [date], available at: [https://www.cbc.ca/natureofthings/features/theseven-universal-emotions-we-wear-on-our-face].
[2] K. Patel, D. Mehta, C. Mistry, et al., "Facial Sentiment Analysis Using AI Techniques: State-of-the-Art, Taxonomies, and Challenges," IEEE Access, vol. 8, pp. 90,495-90,519, 2020.
[3] "Reading Facial Expressions of Emotion," American Psychological Association, accessed on [date], available at: [https://www.apa.org/science/about/psa/2011/05/facialexpressions.].
[4] T. Mittal, P. Guhan, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emoticon: Context-Aware Multimodal Emotion Recognition Using Frege's Principle," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14,234-14,243.
[5] R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza, "Context-Based Emotion Recognition Using Emotic Dataset," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2,755-2,766, 2019.
[6] J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn, "Context-Aware Emotion Recognition Networks," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10,143-10,152.
[7] S. Knobloch-Westerwick, J. Abdallah, and A. C. Billings, "The Football Boost? Testing Three Models on Impacts on Sports Spectators' Self-Esteem," Communication & Sport, vol. 8, no. 2, pp. 236-261, 2020.
[8] J. Jayalekshmi and T. Mathew, "Facial Expression Recognition and Emotion Classification System for Sentiment Analysis," in 2017 International Conference on Networks & Advances in Computational Technologies (NetACT), IEEE, 2017, pp. 1-8.
[9] N. B. Kar, K. S. Babu, A. K. Sangaiah, and S. Bakshi, "Face Expression Recognition System Based on Ripplet Transform Type II and Least Square SVM," Multimedia Tools and Applications, vol. 78, no. 4, pp. 4,789-4,812, 2019.
[10] H. M. Shah, A. Dinesh, and T. S. Sharmila, "Analysis of Facial Landmark Features to Determine the Best Subset for Finding Face Orientation," in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), IEEE, 2019, pp. 1-4.
[11] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "BlazeFace: Sub-Millisecond Neural Face Detection on Mobile GPUs," arXiv preprint arXiv:1907.05047, 2019.
[12] R. S. Jadhav and P. Ghadekar, "Content-Based Facial Emotion Recognition Model Using Machine Learning Algorithm," in 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), IEEE, 2018, pp. 1-5.
[13] "FER-2013 Learn Facial Expressions from an Image," Kaggle, accessed on [date], available at: [https://www.kaggle.com/msambare/fer2013].
[14] S. Datta, D. Sen, and R. Balasubramanian, "Integrating Geometric and Textural Features for Facial Emotion Classification Using SVM Frameworks," in Proceedings of the International Conference on Computer Vision and Image Processing, Springer, 2017, pp. 619-628.
[15] A. R. Kurup, M. Ajith, and M. M. Ramón, "Semi-Supervised Facial Expression Recognition Using Reduced Spatial Features and Deep Belief Networks," Neurocomputing, vol. 367, pp. 188-197, 2019.
[16] Y. Gan, J. Chen, and L. Xu, "Facial Expression Recognition Boosted by Soft Label with a Diverse Ensemble," Pattern Recognition Letters, vol. 125, pp. 105-112, 2019. "[date]" and "[URL]" should be replaced with the specific date and URL of access.

Reference:

[17] P. Dhankhar, "ResNet-50 and VGG-16 for Recognizing Facial Emotions," International Journal of Innovations in Engineering and Technology (IJIET), vol. 13, no. 4, pp. 126-130, 2019.
[18] A. P. Fard and M. H. Mahoor, "AD-Corre: Adaptive Correlation-Based Loss for Facial Expression Recognition in the Wild," IEEE Access, vol. 10, pp. 26,756-26,768, 2022.
[19] A. H. Farzaneh and X. Qi, "Facial Expression Recognition in the Wild via Deep Attentive Center Loss," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 2402-2411.
[20] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion-Aware Facial Expression Recognition Using CNN with Attention Mechanism," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439-2450, 2018.
[21] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition," IEEE Transactions on Image Processing, vol. 29, pp. 4057-4069, 2020.
[22] J. She, Y. Hu, H. Shi, J. Wang, Q. Shen, and T. Mei, "Dive into Ambiguity: Latent Distribution Mining and Pairwise Uncertainty Estimation for Facial Expression Recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6248-6257.
[23] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 Million Image Database for Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1452-1464, 2017.
[24] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-Scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs," in Proceedings of the 21st ACM International Conference on Multimedia, 2013, pp. 223-232.
[25] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, 2017.
[26] S. Li, W. Deng, and J. Du, "Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2852-2861.
[27] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94-101. doi: [DOI].
[28] J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn, "Context-Aware Emotion Recognition Networks," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 10,143-10,152.
[29] H. Gunes and M. Piccardi, "Bi-Modal Emotion Recognition from Expressive Face and Body Gestures," Journal of Network and Computer Applications, vol. 30, no. 4, pp. 1334-1345, 2007.
[30] P. Ekman, "An Argument for Basic Emotions," Cognition & Emotion, vol. 6, no. 3-4, pp. 169-200, 1992.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015. arXiv: [arXiv link].
[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248-255.
[33] W. Li, X. Dong, and Y. Wang, "Human Emotion Recognition with Relational Region-Level Analysis," IEEE Transactions on Affective Computing, 2021.
[34] G. Wen, Z. Hou, H. Li, D. Li, L. Jiang, and E. Xun, "Ensemble of Deep Neural Networks with Probability-Based Fusion for Facial Expression Recognition," Cognitive Computation, vol. 9, no. 5, pp. 597-610, 2017.

Reference:

[35] A. Renda, M. Barsacchi, A. Bechini, and F. Marcelloni, "Comparing Ensemble Strategies for Deep Learning: An Application to Facial Expression Recognition," Expert Systems with Applications, vol. 136, pp. 1-11, 2019.
[36] D. Zeng, Z. Lin, X. Yan, Y. Liu, F. Wang, and B. Tang, "Face2Exp: Combating Data Biases for Facial Expression Recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 20,291-20,300.
[37] K. Wang, X. Peng, J. Yang, S. Lu, and Y. Qiao, "Suppressing Uncertainties for Large-Scale Facial Expression Recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6897-6906.
[38] CDC, accessed on [date], available at: [URL].
[39] J. Chakraborty, S. Majumder, and T. Menzies, "Bias in Machine Learning Software: Why? How? What to Do?" in Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 2021, pp. 429-440.
[40] Defi Dataset, accessed on [date], available at: [URL].
[41] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going Deeper in Facial Expression Recognition Using Deep Neural Networks," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016, pp. 1-10.
[42] T. Randhavane, U. Bhattacharya, P. Kabra, et al., "Learning Gait Emotions Using Affective and Deep Features," in Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games, 2022, pp. 1-10.
[43] S. K. D'mello and A. Graesser, "Multimodal Semi-Automated Affect Detection from Conversational Cues, Gross Body Language, and Facial Features," User Modeling and User-Adapted Interaction, vol. 20, no. 2, pp. 147-187, 2010.
[44] U. Bhattacharya, T. Mittal, R. Chandra, T. Randhavane, A. Bera, and D. Manocha, "STEP: Spatial Temporal Graph Convolutional Networks for Emotion Perception from Gaits," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 1342-1350.
[45] H. Liu, H. Cai, Q. Lin, X. Li, and H. Xiao, "Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6253-6266.
[46] J. L. Joseph and S. P. Mathew, "Facial Expression Recognition for the Blind Using Deep Learning," in 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-5. doi: [DOI].
[47] P. Tahghighi, A. Koochari, and M. Jalali, "Deformable Convolutional LSTM for Human Body Emotion Recognition," in Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part III, Springer, 2021, pp. 741-747.
[48] P. D. Marrero Fernandez, F. A. Guerrero Pena, T. Ren, and A. Cunha, "FerAtt: Facial Expression Recognition with Attention Net," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0-0.
[49] K. Sikka, K. Dykstra, S. Sathyanarayana, G. Littlewort, and M. Bartlett, "Multiple Kernel Learning for Emotion Recognition in the Wild," in Proceedings of the 15th ACM on International Conference on Multimodal Interaction, 2013, pp. 517-524.
[50] K. R. Scherer and H. Ellgring, "Multimodal Expression of Emotion: Affect Programs or Componential Appraisal Patterns?" Emotion, vol. 7, no. 1, p. 158, 2007.
[51] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated Analysis of Body Movement in Emotionally Expressive Piano Performances," Music Perception, vol. 26, no. 2, pp. 103-119, 2008.
[52] L. Chen, M. Li, M. Wu, W. Pedrycz, and K. Hirota, "Coupled Multimodal Emotional Feature Analysis Based on Broad-Deep Fusion Networks in Human-Robot Interaction," IEEE Transactions on Neural Networks and Learning Systems, 2023.

Reference:

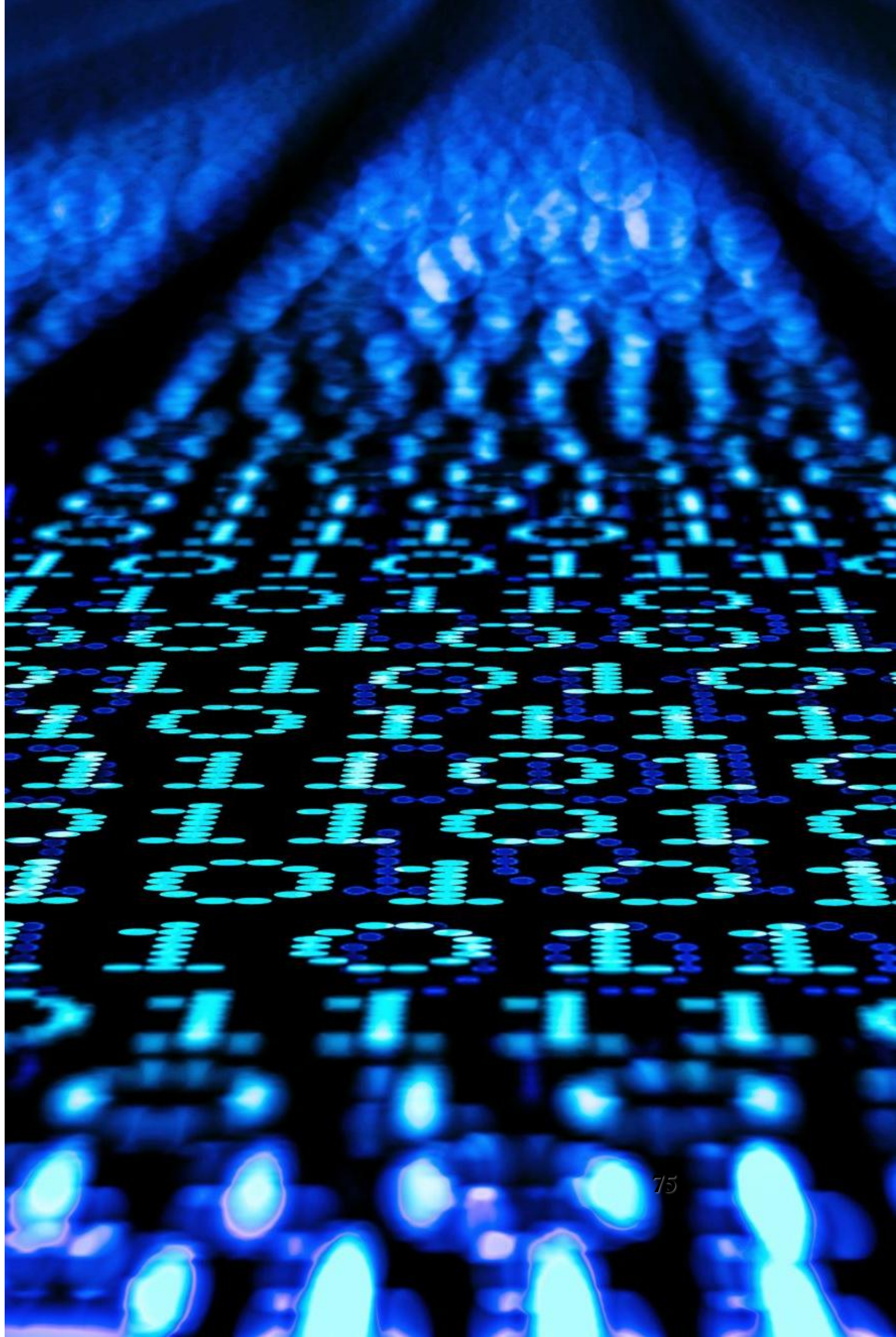
[53] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis," in 2016 IEEE 16th International Conference on Data Mining (ICDM), IEEE, 2016, pp. 439-448.
[54] B. Sun, S. Cao, J. He, and L. Yu, "Affect Recognition from Facial Movements and Body Gestures by Hierarchical Deep Spatio-Temporal Features and Fusion Strategy," Neural Networks, vol. 105, pp. 36-51, 2018.
[55] M. Li, L. Chen, M. Wu, W. Pedrycz, and K. Hirota, "Multimodal Information-Based Broad and Deep Learning Model for Emotion Understanding," in 2021 40th Chinese Control Conference (CCC), IEEE, 2021, pp. 7410-7414.
[56] T. Mittal, A. Bera, and D. Manocha, "Multimodal and Context-Aware Emotion Perception Model with Multiplicative Fusion," IEEE MultiMedia, vol. 28, no. 2, pp. 67-75, 2021.
[57] Y. Bhatia, A. H. Bari, and M. Gavrilova, "A LSTM-Based Approach for Gait Emotion Recognition," in 2021 IEEE 20th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), IEEE, 2021, pp. 214-221.
[58] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "M3ER: Multiplicative Multimodal Emotion Recognition Using Facial, Textual, and Speech Cues," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 1359-1367.
[59] K. Wang, X. Zeng, J. Yang, et al., "Cascade Attention Networks for Group Emotion Recognition with Face, Body and Image Cues," in Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018, pp. 640-645.
[60] T. Gedeon, A. Dhall, J. Joshi, J. Hoey, R. Goecke, and S. Ghosh, "From Individual to Group-Level Emotion Recognition: EmotiW 5.0," 2021.
[61] E. A. Veltmeijer, C. Gerritsen, and K. Hindriks, "Automatic Emotion Recognition for Groups: A Review," IEEE Transactions on Affective Computing, 2021.
[62] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-Device Real-Time Body Pose Tracking," arXiv preprint arXiv:2006.10204, 2020. [64] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 2017.
[63] R. Santhoshkumar and M. Kalaiselvi Geetha, "Vision-Based Human Emotion Recognition Using HOG-KLT Feature," in Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019), Springer, 2020, pp. 261-272.
[64] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al., “Conditional image generation with pixelcnn decoders,” Advances in neural information processing systems, vol. 29, 2016.
[65] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” in Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings, Springer, 2017, pp. 146–157.
[66] L. Ruff, R. Vandermeulen, N. Goernitz, et al., “Deep one-class classification,” in International conference on machine learning, PMLR, 2018, pp. 4393–4402.

Data Bias

- How do we deal with bias
  - Machine learning algorithms can discriminate based on classes like race and gender
  - A good model is dependent on a good dataset and without proper care a dataset can lack diversity
  - Biased dataset will perform poorly with minority:
    - If most of the samples are white males, the model will fail for women and people of color
  - Researchers (Buolamwini et. al.) showed:
    - Three commercially released facial-analysis programs from major technology companies demonstrate both skin-type and gender biases
    - Error rates in determining the gender of light-skinned men were never worse than 0.8 percent
    - For darker-skinned women, more than 20 percent in one case and more than 34 percent in the other two

Data Bias

- How bias is introduced in ER
  - Keyword searching in google is a popular method of collecting visual (image and video) data
  - In our search with keyword “angry face”- 85% of the acceptable images appeared are male
  - This pattern holds for other generic keywords like ”sad people”, ”happy human” etc.
  - Therefore, a dataset prepared by collecting results from these types of keyword search results in bias
  - Same applies to the volunteer choice for creating an acted dataset
  - Without careful selection of people from multiple genders and ethnic backgrounds, dataset bias can be easily incorporated into the model



Data Bias

- Bias reduction plan
  - Better representation of minority groups by using specific keywords:
    - Using both “happy man face” and “happy woman face” instead of “happy face” keyword
  - Choosing volunteers from diverse background
  - To produce new ML models which provide higher importance on less represented data samples
  - Data augmentation
  - Data cleaning algorithms
  - Transfer learning

Data Bias

- Bias in the ER datasets
  - Widely used ER dataset FER-2013 is an example of keyword search bias

TABLE VII: Gender bias- number of images with male subjects per 100 images returned from gender neutral-keyword searches on Google and number of images with male subjects per 100 images on FER-2013 dataset.

Keyword	# Male in Google(%)	# Male in FER-2013
"Angry people"	84.7	70
"Fear face:	60.1	52
"Happy human face"	55.8	58
"Sad human face"	40.0	45

CNN

- Input Image: 226x226x3

- Convolutional Layer 1:
  - Filter Size: 2x2
  - Stride: 1
  - Padding: 0
  - Output Dimension: 225x225x3
  - Activation: ReLU
- Max Pooling Layer 1:
  - Pooling Size: 2x2
  - Output Dimension: 112x112x3

- Convolutional Layer 2:
  - Filter Size: 2x2
  - Stride: 1
  - Padding: 0
  - Output Dimension: 111x111x3
  - Activation: ReLU
- Max Pooling Layer 2:
  - Pooling Size: 2x2
  - Output Dimension: 55x55x3

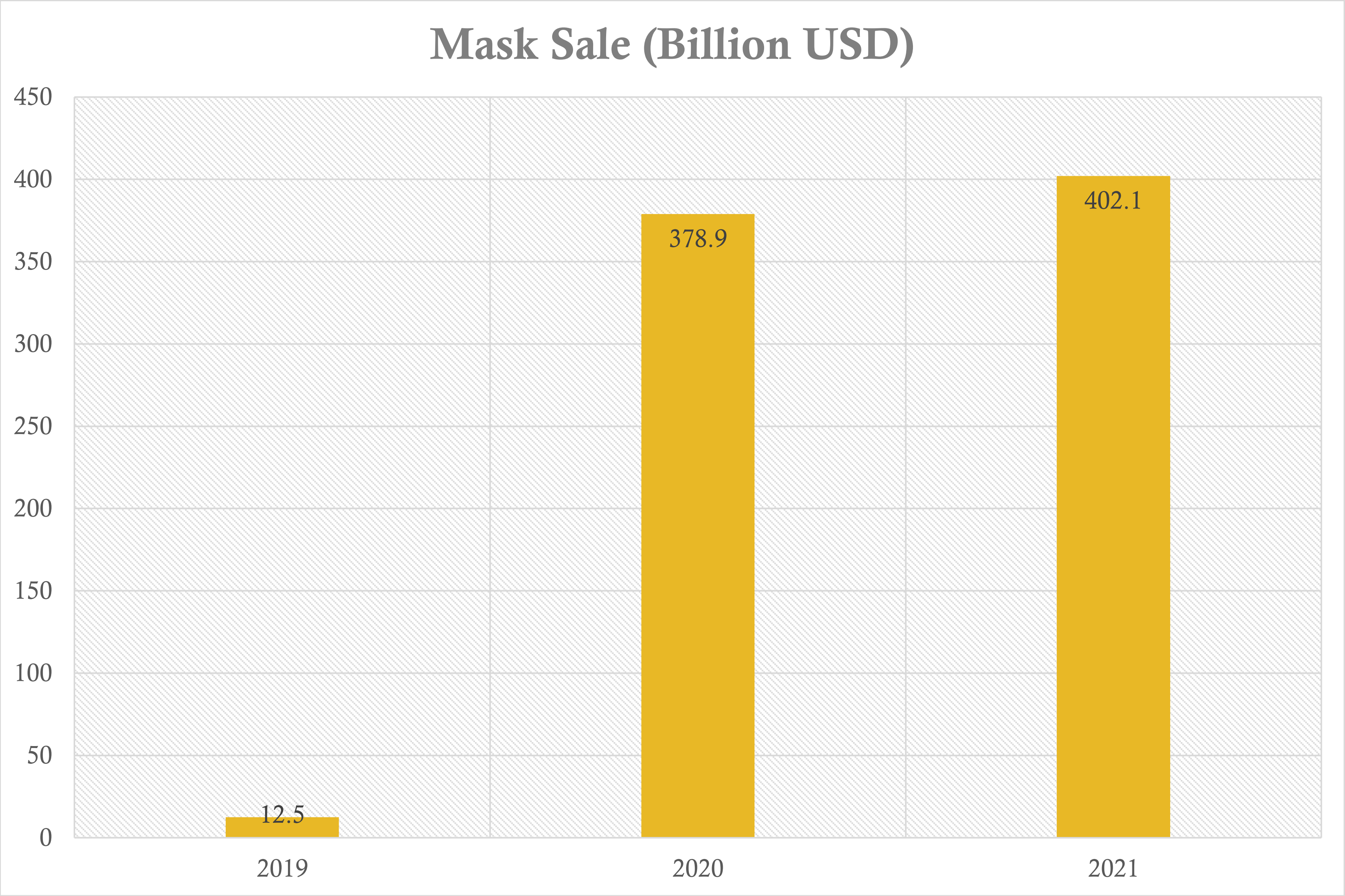
CNN

- Convolutional Layer 3:
  - Filter Size: 2x2
  - Stride: 1
  - Padding: 0
  - Output Dimension: 54x54x3
  - Activation: ReLU
- Max Pooling Layer 3:
  - Pooling Size: 2x2
  - Output Dimension: 27x27x3

- Fully Connected Layer:
  - Input Dimension: 27x27x3
  - Output Dimension: 256

Emotion Recognition Use Cases

Public Safety	• Identify mental health issue to prevent school shooting
Law Enforcement	• Identify suspicious behavior and criminal intent
Healthcare	• Detect medical conditions such as depression
Autonomous Car	• Trigger alarm for extreme emotional state (anger, fear etc.) of the driver
Interactive Gaming	• Adjust the gameplay based on the comfort level of the player


Mask Sale (Billion USD)
450
400
350
300
250
200
150
100
50
0
12.5
378.9
402.1
2019
2020
2021