Quantifying Viewer Engagement through Facial Mood Detection

Charumitra Ravi (chravi), Manideep Keerthi (mkeerthi), Shricharan Baskaran (shrib)
Luddy School of Informatics and Computing, Indiana University

1. Introduction

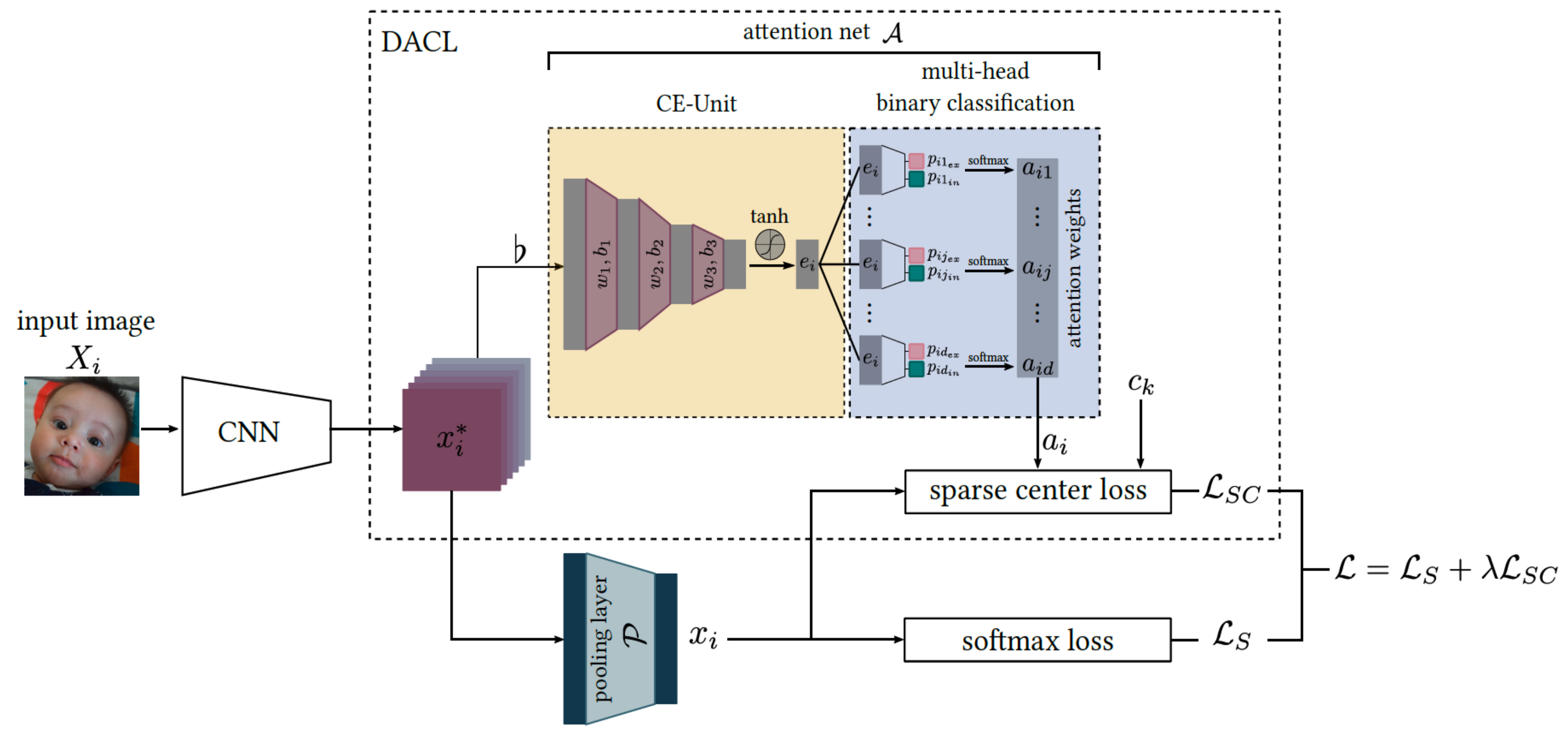
- The project aims to detect audience engagement using facial emotion recognition and provide Real-time feedback on audience mood.
- The project relies heavily on the paper "Facial Expression Recognition in the Wild via Deep Attentive Center Loss," which uses the Deep Attentive Center Loss (DACL) method for better facial expression recognition.
- Deep Metric Learning (DML) approaches have been used in Facial Expression Recognition (FER) methods to improve feature discrimination, but equally supervising all features with metric learning may include irrelevant features and hurt generalization ability.
- Deep Attentive Center Loss (DACL) overcomes this as it selects a subset of significant feature elements to improve discrimination.
- The DACL method has been proven to work well on varied emotion datasets like RAF-DB and AffectNet, and works particularly good for images taken in the wild (Non ideal or laboratory scenarios).

2. Motivation

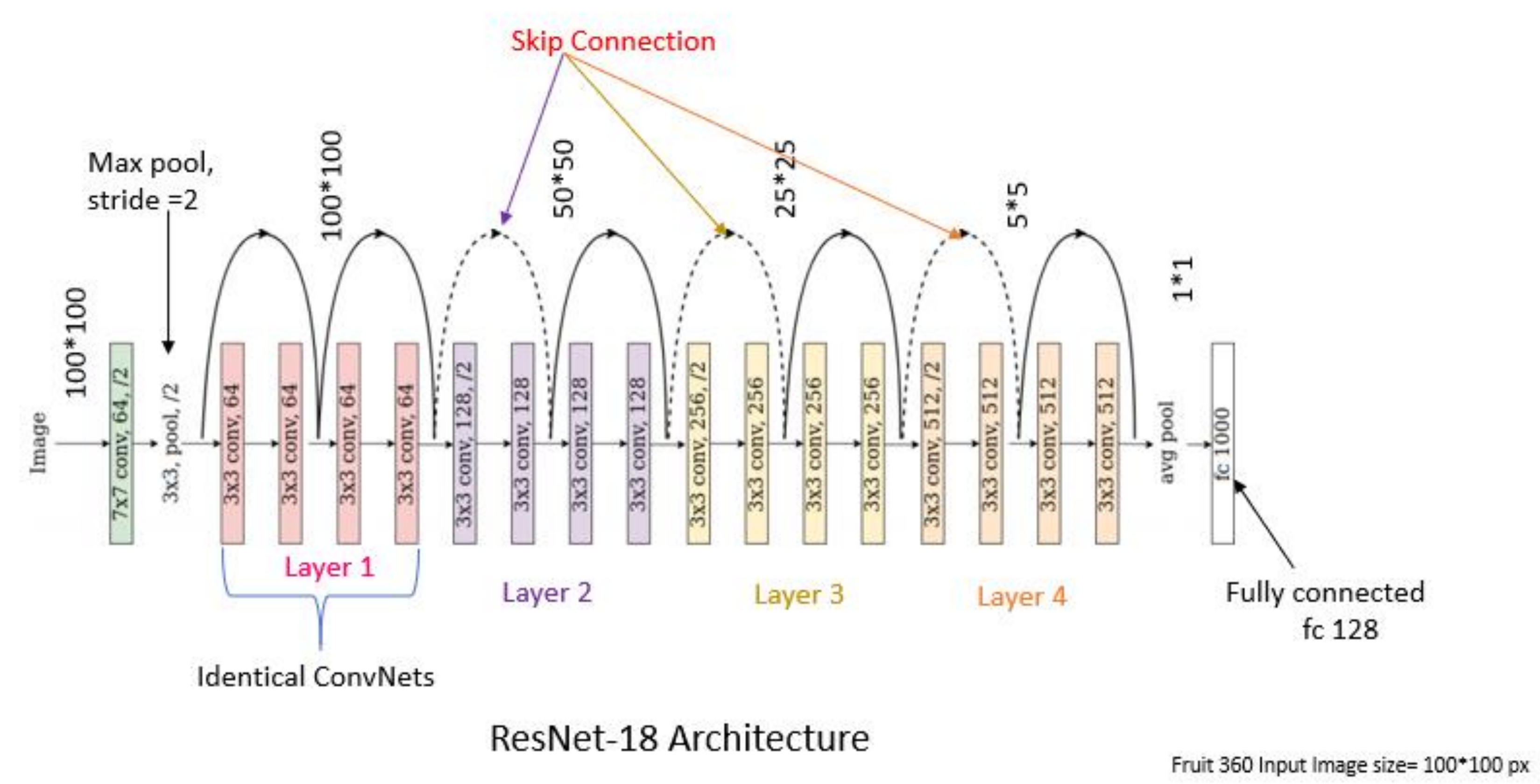
Initial motivation for the project was to enable speakers, especially in a virtual setting to get some feedback on the emotional response of the listeners, which is crucial for the speaker to adapt dynamically.

But the benefits extend to areas like,
Human-computer interaction: Enabling devices to respond to human emotions or expressions.
Marketing and advertising: Analyze customer reactions to marketing materials, to improve marketing strategies.

3. Methods



***Above:** The attention network calculates how much attention should be given to different aspects of the input features to effectively reduce the distance between similar instances while increasing the distance between dissimilar instances within the same class. This is done by selectively focusing on the most relevant information in the embedding space.*

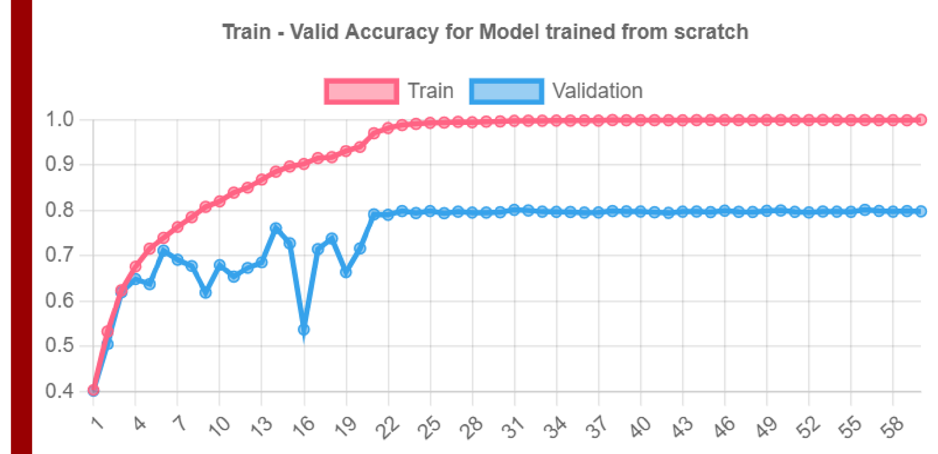


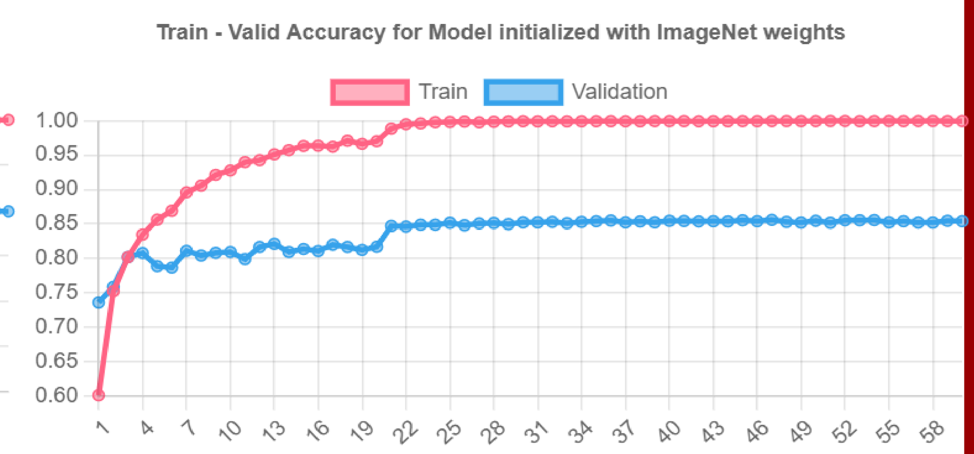
***Above:** Architecture of the ResNet-18 model used in this study.*

4. Experiments

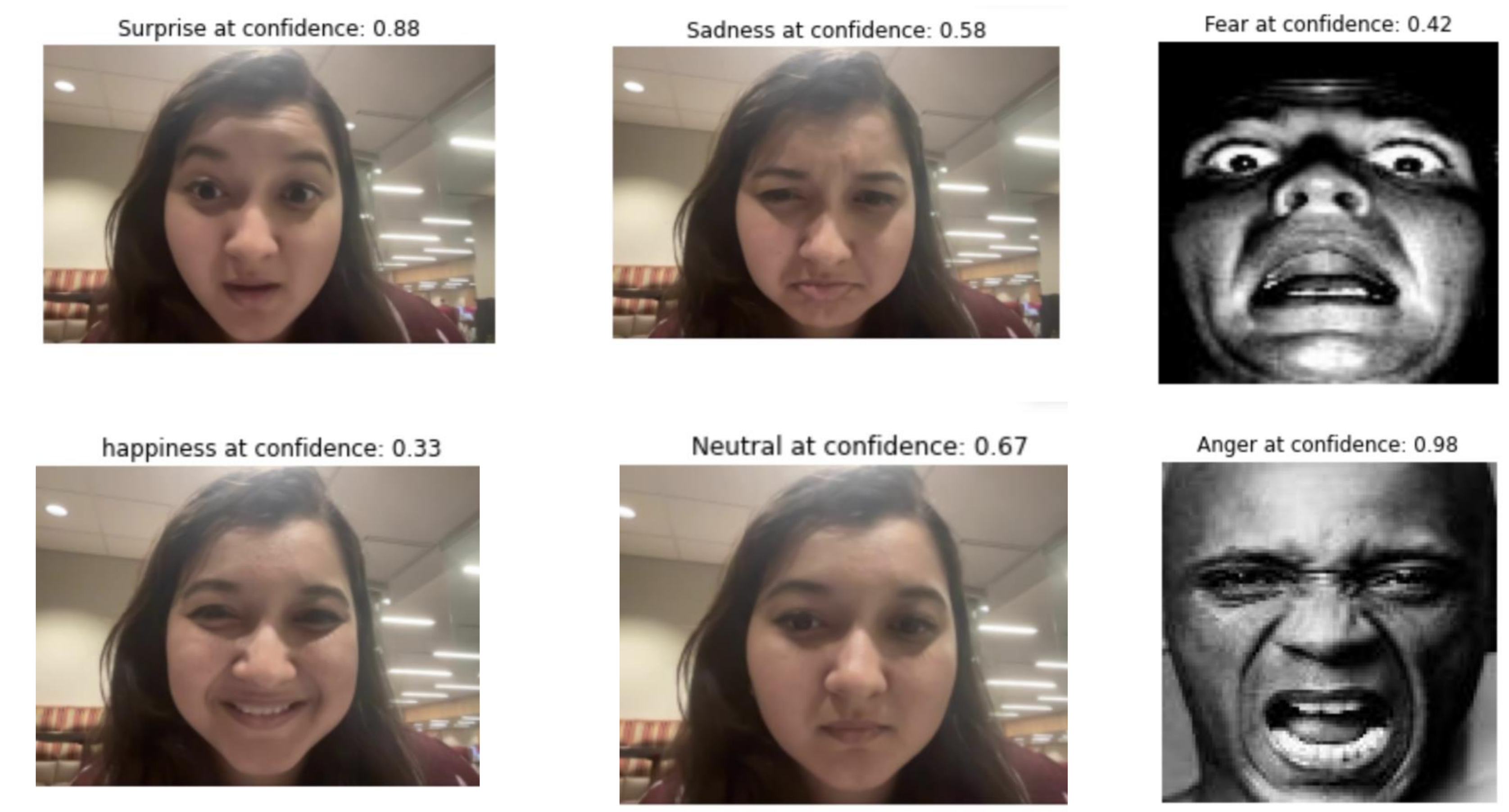
- We have trained and validated our model for 60 epochs both from scratch and using a pretrained model on ImageNet.
- Each run took approx. 12hrs.
- We have distributed the images in the RAF-DB into 7 respective classes for both training and validation data based on their annotations.

5. Results





Above: Comparison between Training and Validation Accuracy for Resnet18 trained from scratch on RAF-DB for 60 epochs and Resnet18 model initialized with weights from Pretrained Resnet18 on ImageNet DB.



6. Future Work

- Current model has an accuracy of approximately 85% in detecting emotions from facial images.
- Future work involves improving the accuracy further by incorporating more diverse and accurately annotated data.
- Obtaining such data is challenging due to the sensitive nature of the data and the manual effort required for annotation.
- Using AI-generated images may help address this challenge.
- The task requires making the algorithm as efficient as possible and high-quality hardware to provide real-time emotion detection for high-quality videos.