CHILD FACIAL EXPRESSION DETECTION

STUDENTS:

- EDEN BENHAMOU
- DEBORAH WOLHANDLER

SUPERVISORS:

- ALON ZVIRIN
- MICHAL ZIVAN

PROJECT MOTIVATION:

- Analyzing children **reaction** and **behavior** in different **study environments**:
  - Yoga and storytelling
- Three aspects in reaction interpretation:
  - Posture
  - Physical Movements
  - Facial Expression
- Facial Expression gives information about **emotional state** of a child during a lesson.
- Study importance of emotion in **learning capacity.**

PROJECT GOALS

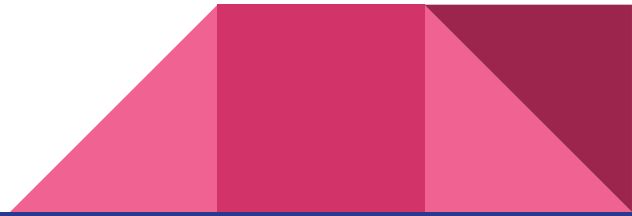
1. Detect **emotions** of children in videos

2. **Track** after children along the video

3. **Display results** for each child as function of time

WHAT ALREADY EXISTS:

- Face **detection** algorithms
  - Viola-Jones
- Emotion **classification** algorithms
  - CNN
- Face **recognition** algorithms
  - KNN, Eigenfaces based on PCA, Fisherfaces based on FLD



CHALLENGES

- **Humans** recognize emotions in **~65%** accuracy and use gesture, position, context…
- Some emotions are very similar, and it is **hard to differentiate** them
  - Anger and disgust for example
- **Differences** between **adults** and **children** facial expressions:
  - “open mouth”, “tongue out”, hands on face

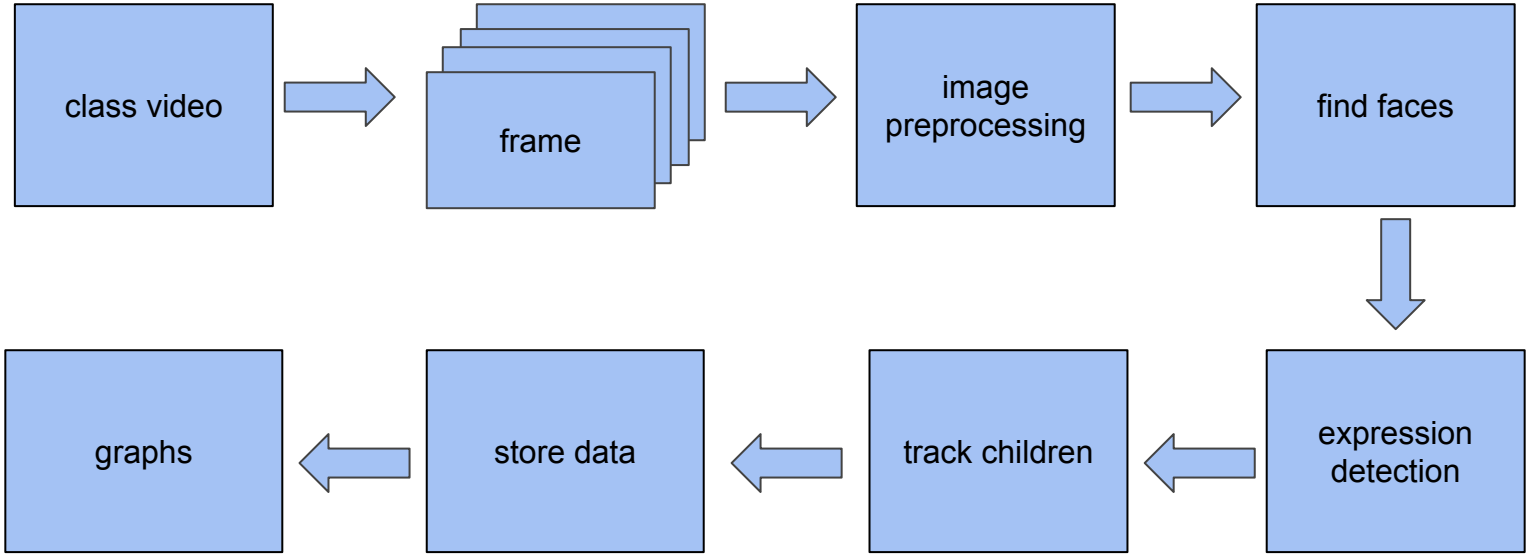
Child images removed due to privacy policy

CHALLENGES

- **Low** video **quality**:
  - Video compression
  - Unstable, not invasive cameras
- Small bounding boxes of faces
- Children **move** a lot
- **Rare** databases of children emotion

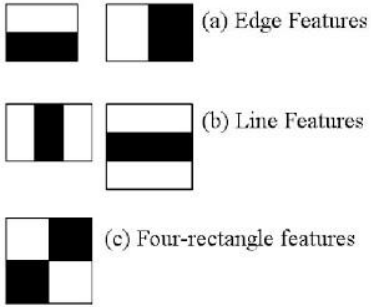
OUR SOLUTION

BLOCK DIAGRAM



FACE DETECTION - CASCADES

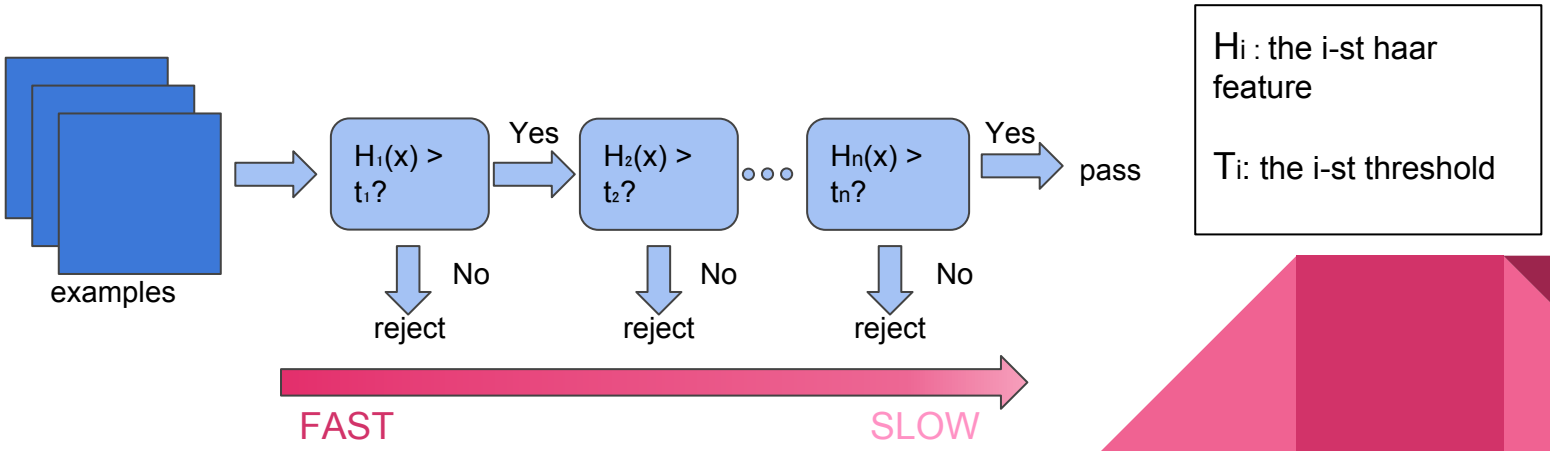
- Machine learning based approach
- **Haar-features** :
  - Each feature: a single value obtained by subtracting sum of pixels



- Use **Adaboost** to find the best features

FACE DETECTION - CASCADES

- Algorithm steps:
  - Choose threshold for low false negative rate
  - Fast classifiers early in cascade
  - Slow classifiers later, but most examples don't get there



FACE DETECTION

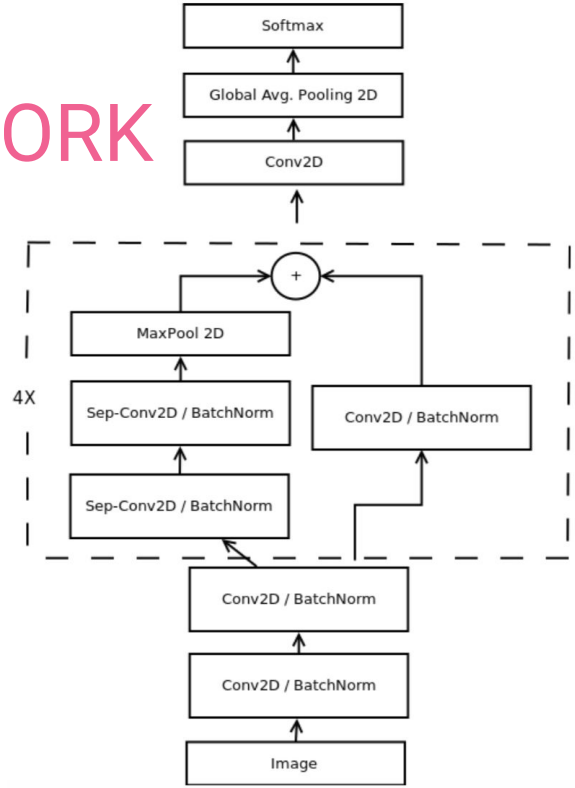
- Cascade for face detection
  - Adaboost selected first two features:
    1. 
    2. 
  - Easily interpretable
    1. The region of the eyes is often darker than the region of the nose and the cheeks
    2. The eyes are darker than the bridge of the nose

EMOTION RECOGNITION

- CNN solution : Mini-Xception architecture
- Input: 64*64*1 face image
- Output: 7 probabilities for each emotion

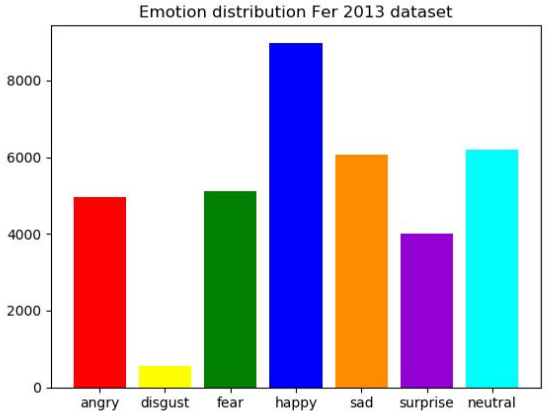
EMOTION RECOGNITION - NETWORK

- Based on **Xception** architecture:
  - Deletion of fully connected layer
  - Use of residual **modules**
  - **Depth-wise** separable convolutions
- Reduces number of parameters:
  - **Speeds** the algorithm
  - Provides better **generalization**
- Trained with **ADAM optimizer**



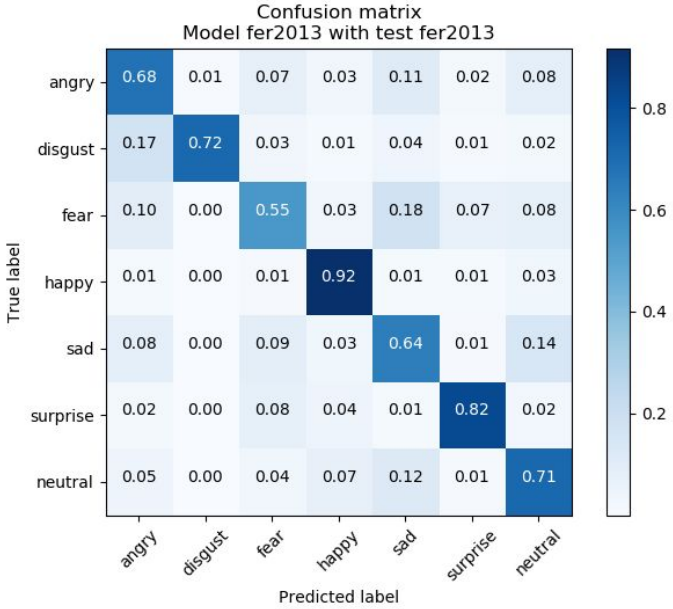
NETWORK - Initial training set

- Model trained with **FER2013** - faces labeled for seven basic expressions - it contains 35,887 images of size 48*48
- Fer2013:
  - Mostly adults
  - Have ~70% accuracy for the state-of-the-art network.
  - Imbalance of emotions in dataset



NETWORK - Initial training set

- Results on training set:

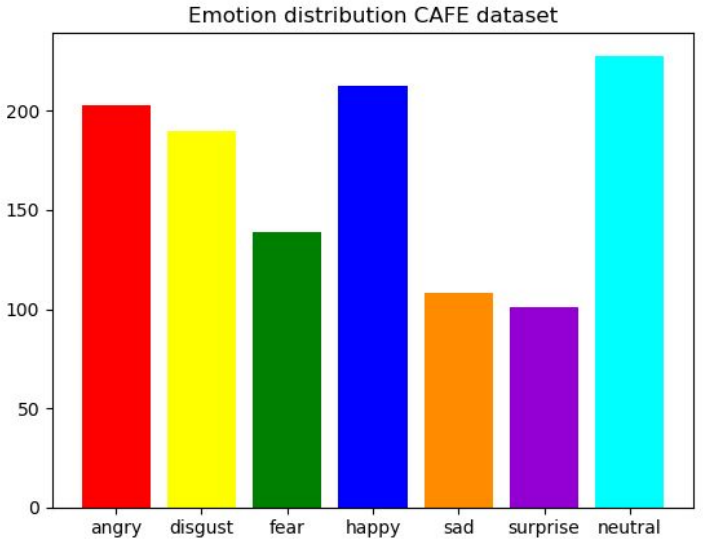


NETWORK - CAFE database

- Children database: CAFE
  - Children pictures posing for 7 emotions: sadness, happiness, surprise, anger, disgust, fear and neutral. Same emotions as fer2013.
  - 1192 pictures
- Highlights the differences between children and adults

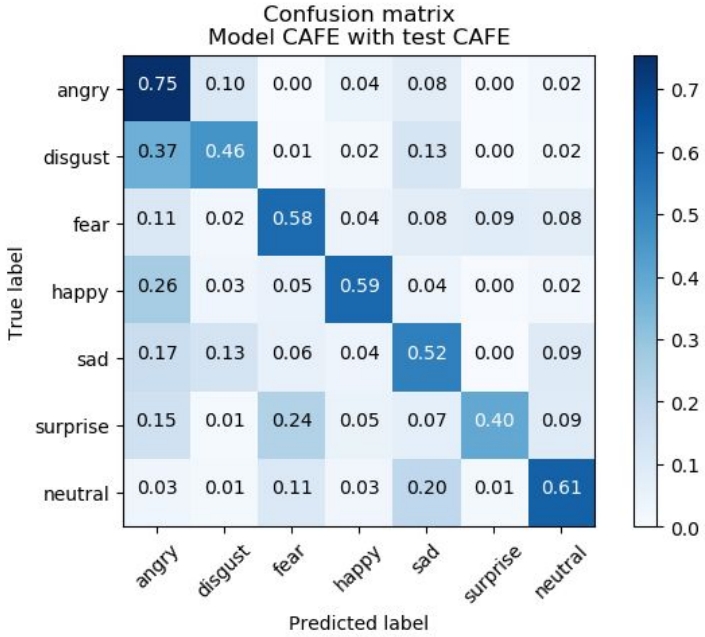
- Trained our model with this database

NETWORK - CAFE database



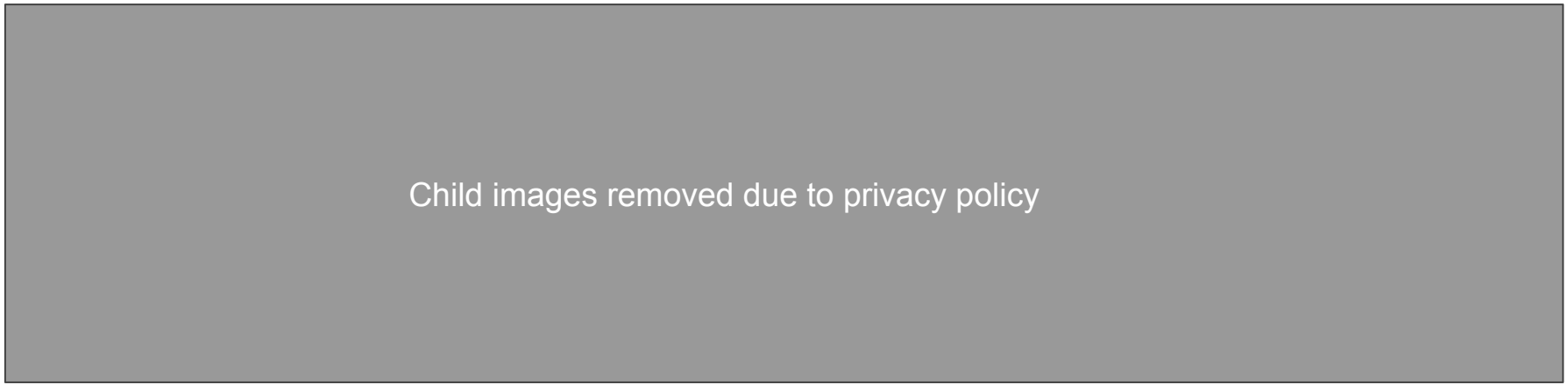
NETWORK - CAFE database

- Results on training set: model trained with fer2013 then with CAFE



NETWORK - CAFE database

- Examples of prediction mistakes:



True label: Disgust
Predicted: Angry

True label: Happy
Predicted: Angry

True label: Surprise
Predicted: Fear

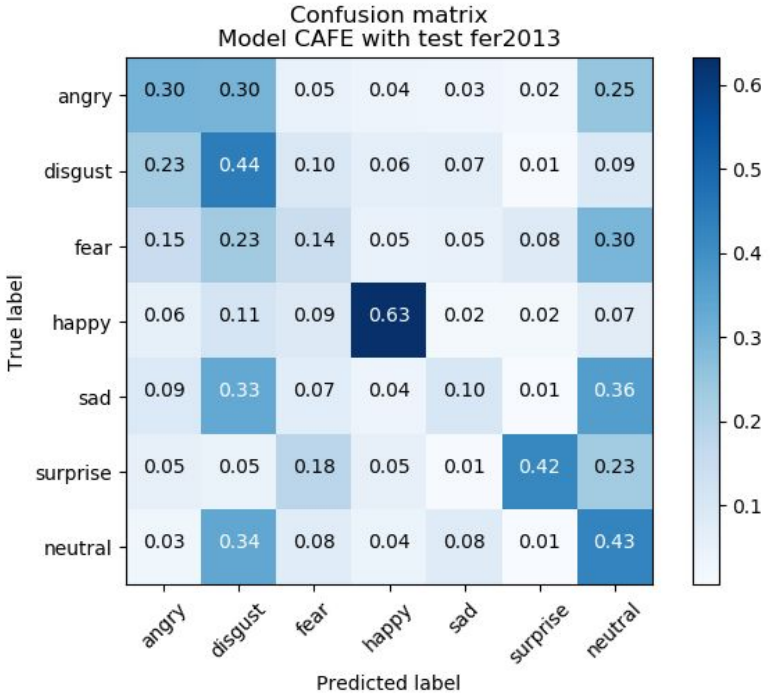
True label: Neutral
Predicted: Sad

NETWORK - CAFE database

- Results: confusion matrix of model trained with CAFE and tested with fer2013

- Errors due to:
  - very different resolution between fer2013 and CAFE
  - Frontal images only in CAFE
  - small database (overfitting)

- Neutral prediction bias:
  - pictures in CAFE shows forced emotions, then, unclear emotions are predicted as neutral



NETWORK - CAFE database

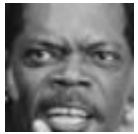
- Neutral prediction bias:

True label: Sad
Predicted: Neutral

True label: Angry
Predicted: Neutral

True label: Fear
Predicted: Neutral

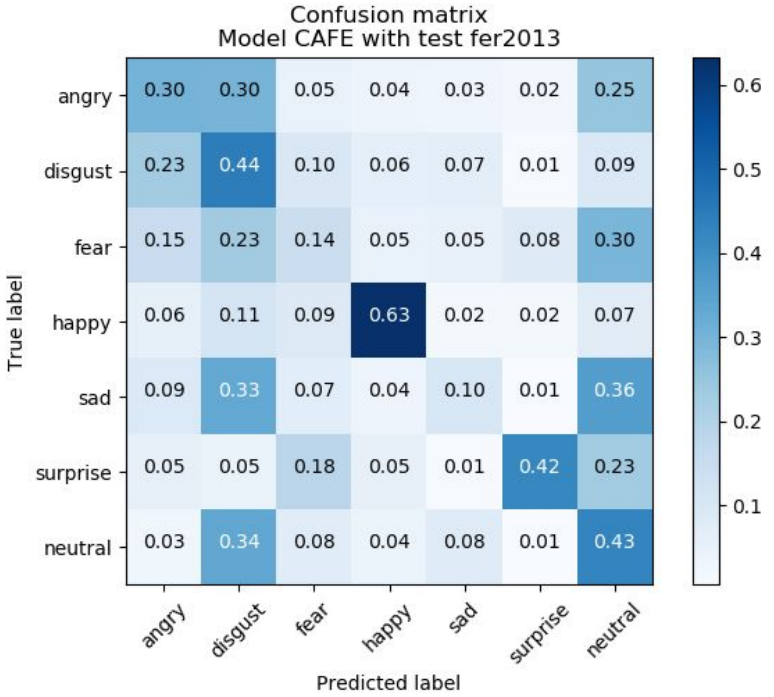
- Disgust prediction bias:
  - Seems like frowned eyebrows are interpreted as disgust



True label: Angry
Predicted: Disgust

True label: Fear
Predicted: Disgust

True label: Neutral
Predicted: Disgust

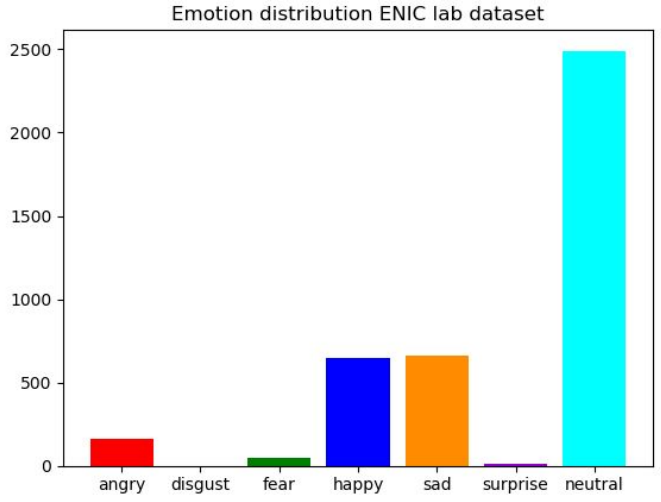


NETWORK - ENIC lab data

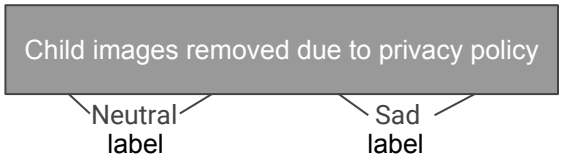
- Total number of videos : 255
- Length of video: 1 - 15 minutes
- Number of children: ~10 children per video
- Resolution of Bounding Box of detected faces: 48*48 up to 100*100

NETWORK - ENIC lab data

- Prepared bounding boxes of children pictures from the videos
- Data labelled by the ENIC lab
- About ~4000 pictures
- Unbalanced dataset:



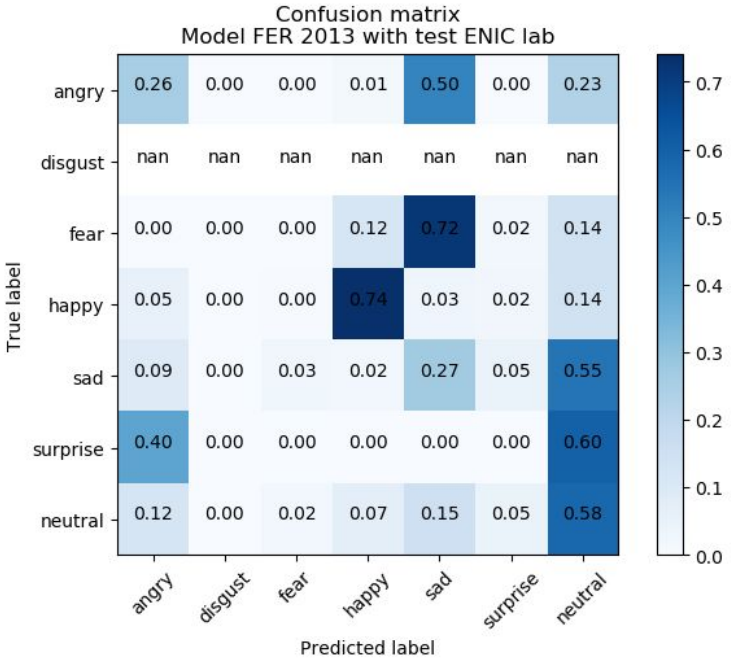
NETWORK - ENIC lab data

- Used for **testing** models accuracy in emotion detection
- Also used for **training** emotion recognition model
- Example of sequence
  - Predicted: all sad
  - 

NETWORK - ENIC lab data

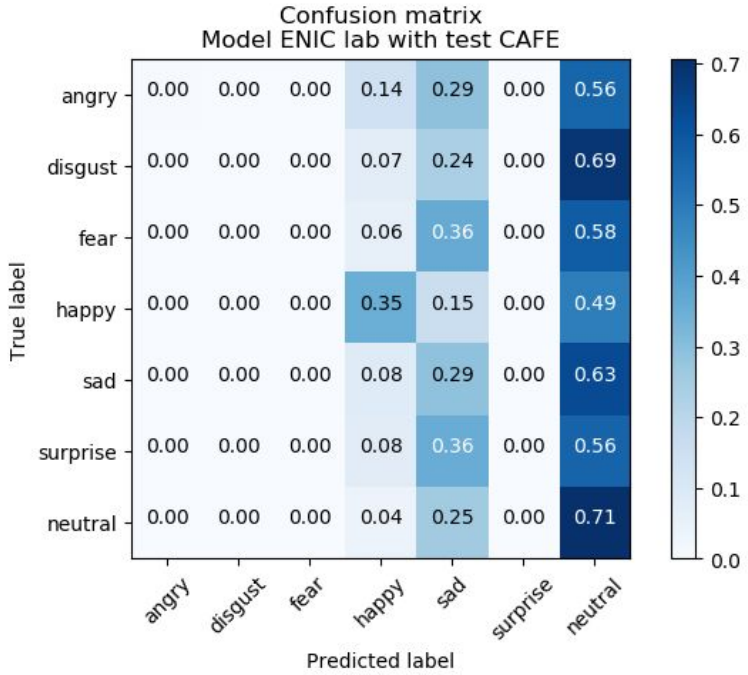
- Results: confusion matrix of the model with fer2013 and tested with ENIC lab dataset
- Explanations:
  - Neutral, happy and sad are the principal classes
  - Examples of sad predicted as neutral:

Child images removed due to privacy policy



NETWORK - ENIC lab data

- Results: confusion matrix of the model trained with ENIC and tested with CAFE
- Explanations:
  - All the pictures in the test set : classified as happy, sad or neutral.
  - Indeed : training database has been classified in these three categories mostly.

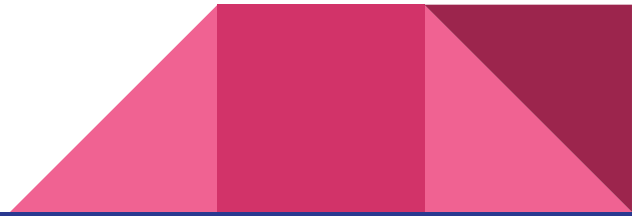


TRACKING - original algorithm

- For each child, looking for the corresponding BB in the previous frames
- Evaluate the **distance**: find center coordinates of BB and calculate euclidean distance between them.
- Find the **nearest** child in the previous frames
- **Maximal authorized distance**
  - if exceeded, new child discovered
- **Frames threshold**
  - If exceeded, consider we lose the child

TRACKING - Face recognition

- Works with KNN algorithm
- Reference directory: pictures for each person we want to recognize
- Trained with the reference directory
- Returns the child id or “unknown”
- Used in our tracking algorithm in two ways:
  - Semi automatic algorithm
  - Automatic algorithm



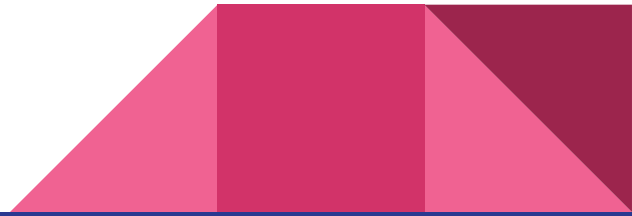
TRACKING - semi automatic algorithm

1. Reference directory with high resolution children pictures from ENIC lab

   Problem: not all the children were in the pictures, pretty bad results of face recognition.

   Possible solution: get pictures of **all** the children of the video.

2. Choose manually bounding boxes pictures and use them as reference.
   a. Fast training and better results
   b. For “unknown” cases, use the previous method with the distance
   c. Inconvenient: not user friendly, need to manually create the directories for each video.



TRACKING - automatic algorithm

- Automatically add new BB to the reference directory.
- First, looking for the child with the face recognition algorithm.
- If “*unknown*”, use the distance algorithm.
- If we found it, add the image to the ref dir of this child.
- Else, create new child in the ref directory.
- Results: for 1200 frames: we recognized 96% of the children:
  - more than 87% with the face recognition algorithm
  - 70% of the unknown BB left with the distance algorithm
- But need to train after each insertion of image in the ref dir.
- 30 times longer than the original !
-

SAVING SEQUENCES

- Saving sequences of bounding boxes for a specific child

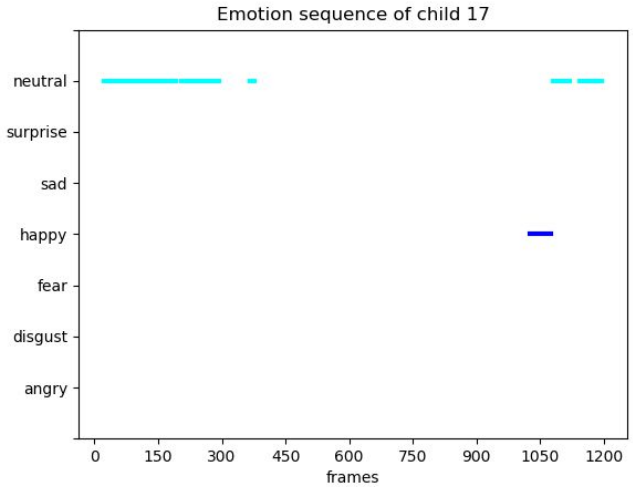
- One Sequence for one emotion: we want to output only stable emotions
  - Length of minimum sequence can be changed. By default equals 6

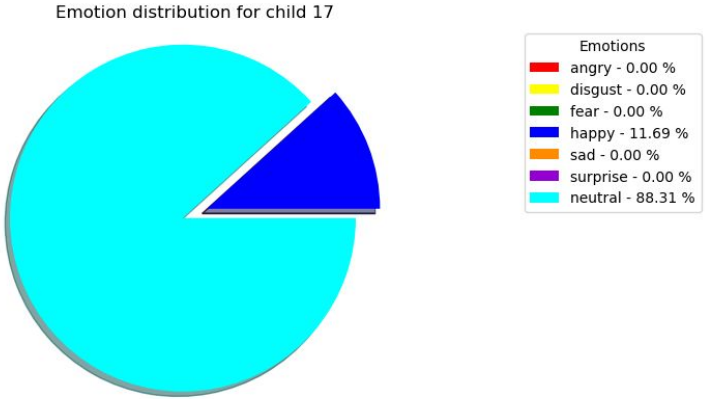
- Create a dataset with those images. Labelled by ENIC Lab

RESULTS

Child images removed due to privacy policy

happy neutral



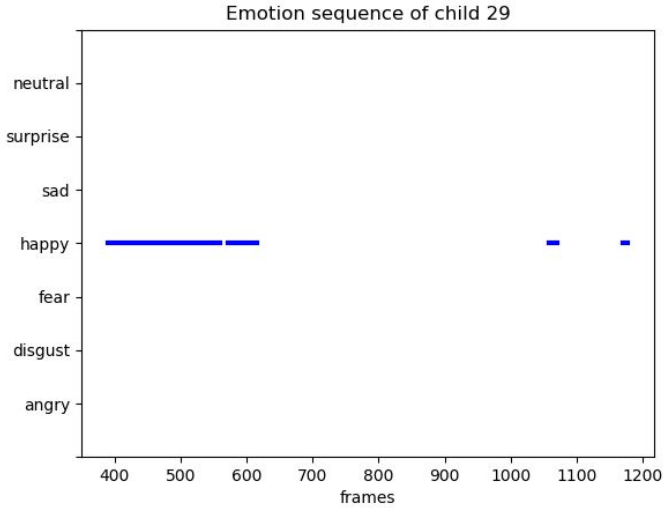


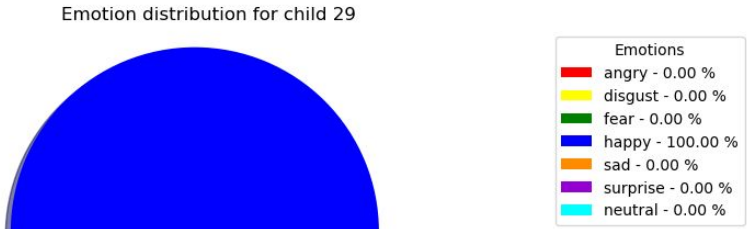
Hazav camera 1 220318 - minimum sequence = 6 for two minutes process

RESULTS

Child images removed due to privacy policy

happy happy



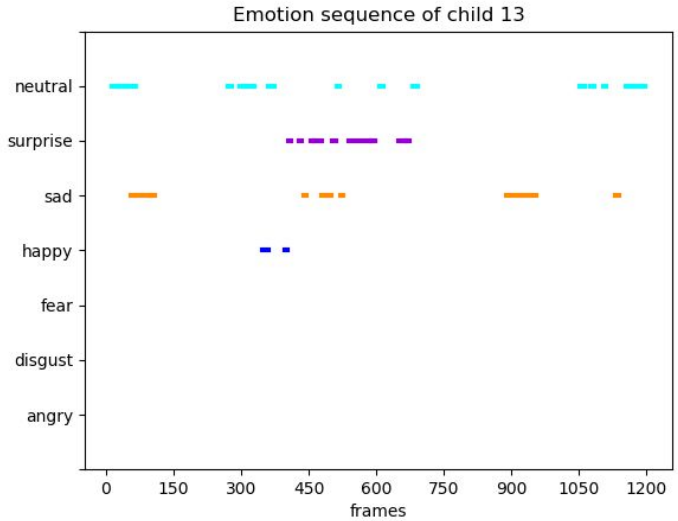


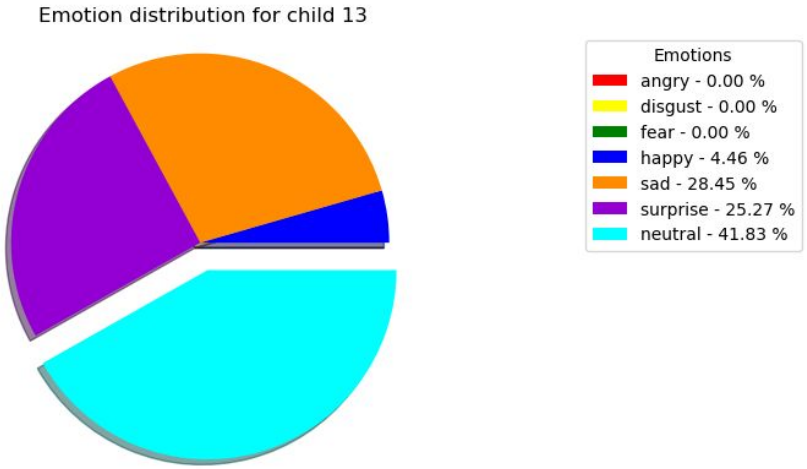
Hazav camera 1 220318 - minimum sequence = 6 for two minutes process

RESULTS

Child images removed due to privacy policy

sad neutral surprise happy





Hazav camera 1 220318 - minimum sequence = 6 for two minutes process

CONCLUSION

- Our goals were achieved
- Tracking children with 90% accuracy
- Emotion recognition performance is reasonable given the project challenges .
- Output statistical results that will boost the research of the ENIC lab.

NEXT STEPS

- Training:
  - Use more databases and compare results with the current model
  - Prepare more balanced dataset to label
- Filter emotion detection results:
  - Reduce number of emotions before training. For exampleת merge angry and disgust emotions.
- Tracking
  - Optimize parameters of current algorithm
  - Pass over the children directories in order to merge corresponding children
  - Try other face recognition algorithm deep learning based
- Super Resolution
  - Improve image resolution

LITERARY SURVEY

1. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, 2001.
2. Adam Geitgey, “Machine learning is Fun!”, https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471
3. Arriaga, O., Valdenegro-Toro, M., & Plöger, P. (2017). Real-time Convolutional Neural Networks for Emotion and Gender Classification. *arXiv preprint arXiv:1710.07557*.
4. Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*.
5. LoBue, V., Baker, L., & Thrasher, C. (in press). Through the eyes of a child: Preschoolers’ identification of emotional expressions from the Child Affective Facial Expression (CAFE) Set. Cognition & Emotion.
6. LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. Emotion Science, 5, 1532. http://doi.org/10.3389/fpsyg.2014.01532
7. Widen, S. C., Russell, J. A., & Brooks, A. (2004, May). Anger and disgust: Discrete or overlapping categories. In *2004 APS Annual Convention, Boston College, Chicago, IL*.