**Paper summaries**

**Real-Time Facial Expression Recognition App Development on Smart Phones (2017)**

* Automated facial expression recognition (AFER)
* Lists the possible challenges
* Most of the previous research has focused on basic feature descriptors and classification methods
* Lists early approaches to to FER such as graphic spectral methods
* **USES MANUAL FEATURE EXTRACTION**
* Interesting pre-processing of video to improve lightining conditions
* Has similar structure to what I have to do: train on something, deploy on something: see system evaluation
* Filtering the image to normalize it using Gausian smoothing and low-pass filter

**Facial Expression Recognition with Convolutional Neural Networks (2016)**

* *(FER) systems apply standard machine learning to ex- tracted image features, and these methods generalize poorly to previously unseen*

Some Stanfrod master’s students trying to do what I will do

* References the paper that lists sources of cross-cultural expressions Ekman et al. 1971
* Gives some nice baseline results
* Does transfer learning and references the code to fractional max-pooling
* Shows misclassifications and gives possible explanations
* Lists possible avenues for parameter experimentation
* Does fractional max-pooling work for my dataset?
* Low dropout?
* More filters in the deeper layers
* VGG16 not performing that wekk

**AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild (2017)**

* **Describes my dataset**
* Dimensional model (continuous, valance and arousal) and categorical model (discrete set of emotions)
* Gives some nice baseline results for **my dataset**
* reviews existing databases
* lists top approaches of Deep Neural Network approaches (Table 2)
* Section 3:
  + The database is created by querying emotion related keywords from search engines and annotated by expert human labelers
  + **The OpenCV face recognition** (use?) was used to obtain bounding boxes around each face.
  + Eleven discrete categories were defined in the categorical model of *AffectNet* as: Neutral, Happy, Sad, Surprise, Fear, Anger, Disgust, Contempt, None, Uncertain, and Non-face
* Section 4 Baseline entire thing is important:
  + Heavily imbalanced test set
  + The validation set is used for hyper-parameter tuning
  + The skew normalization is performed by random under-sampling of the classes in the test set. This process is repeated 200 times
  + Some mistakes of the algorithm could easily be made by a human
  + AlexNet architecture used as a baseline
  + Weighted loss works best bro
  + The described architecture works better than some shit from Microsoft

Deep convolution network based emotion analysis towards mental health care (2019)

* Shows performance of different network architectures on different datasets, including **AffectNet** and
* AlexNet seems to be best with some adjustments

**Deep Facial Expression Recognition: A Survey (2018)**

* Use cases
* Only talks about the categorical model
* Mentions why a lot of data is important
* Goes through the needed phases: pre-processing, deep feature learning and deep feature classification
* Pre-processing:
  + Face alignment (already done for affect-net, for new images do: https://www.pyimagesearch.com/2017/05/22/face-alignment-with-opencv-and-python/)
  + Data augmentation
  + Face normalization (not under angle) (**illumination** or pose normalization)
* Deep networks for feature learning methods
* Facial experession classification
* **State of the art:**
* incon- sistent annotations among different FER databases are inevitable which would damage the performance when the training set is enlarged by merging multiple datasets.
* 166 filters out the irrelevant features inside AlexNet
* Mentions special softmax layer to address the intra-class variation: 140 for Island loss and locality preserving loss by 44
* Points out important factors **about assembling ensemble networks**
* Describes the procedure for **dynamic image sequences** **(COME BACK LATER)**
* Aggregating frames in order to use a model trained on a static

**Facial Expression Recognition with Inconsistently Annotated Datasets (2018)**

* Introduces a framework for address inter-dataset varaince
* it tags multiple labels for each image with the human annotations or predicted pseudo labels, and then learns a FER model to fit the latent truth from the inconsistent pseudo labels.
* Fa
* Lists previous solutions to inconsistent labels: require multiplt annotators
* Lists previous solutions to inconsistent labels: requires a small set of clean data
* Provides github link with (light) implementation (<https://github.com/dualplus/LTNet>)
* Why does it perform better than step by step methods?
  + end-to-end methods can intrinsically capture the relations between the input image and inconsistent labels
  + the step-by- step methods separately capture the relations between the input images and the estimated labels, and the relations between the latent truths and inconsistent labels

**Facial Expression Recognition using Convolutional Neural Networks: State of the Art (2016)**

* ensemble methods yielding state of the art
* methods can vary in three big ways:
  + pre-processing
    - face detection, face registration and illimintaiton normalization (mean 0, norm of 100, histogram equalization and lnear plane fitting)
  + CNN architectures
    - Not necesearliy very deep architectures: shows that a CNN with depth 5 is already able to learn discriminative high-level features.
  + CNN Training and inference
    - Data augementation during training and sometimes test(crop10 times and take avg prediction)
    - Use ensembles!!!
* Rather shallow CNN’s, 5-6 layers
* Adding batch normalization layers after every convolutional and fully connected layer improves robustness. Add drouput after the first fully connected layer
* Gives nice parametrs for training (p.4)
* Do grid search to find out optimal droput rate for every architecture
* Outlines bottlenecks:
  + Proposed architectures are shallow compared to state-of-the-art architectures in related fields
  + Random crops and mirroring not optimized for the task at hand; proposes face-aware data augmentation via face registration improves performance
* Shows that deeper CNN’s perform better
* Backend consisting of a hidden layer of 1024 units
* Shows that all mentioned deep architectures perform better and ensemble is just a cherry on top

**Local Learning with Deep and Handcrafted Features for Facial Expression Recognitio (2019)**

* has a nice table with “state of the art results”
* maybe try transfer learning from VGG-13?