Suppressing Uncertainties for large-scale Facial Expression Recognition

- **Aim:** To design a network that automatically corrects uncertainties in the dataset arising from ambiguous facial expressions, low-quality images, incorrect data labeling etc. and then performs FER on the corrected dataset.
- **Methodology:**

- The paper proposes a Self-Cure Network (SCN); comprising of Self-Attention Importance Weighting, Rank Regularization and Relabelling Modules.
- The idea is to prepare a network which can automatically clean the noisy data and subsequently perform better FER with improved quality dataset, hence the name Self-Cure.
- Post feature extraction from each image by a backbone CNN network (Resnet18), each feature is passed through the self-attention importance weighting module which applies FC layer and sigmoid function to it and outputs the importance wts. (α_i for the i^{th} image).
- The logit weighted cross entropy loss is formulated as follows: $$\mathcal{L}_{WCE} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{\alpha_i \mathbf{W}_{y_i}^{\top}\mathbf{x}_i}}{\sum_{j=1}^{C} e^{\alpha_i \mathbf{W}_j^{\top}\mathbf{x}_i}},$$

Where W_j is the j^{th} classifier and L_{WCE} has a positive co-relation with α.

- The importance wts. are assigned in such a manner that certain images have higher wts. then uncertain images. The learned importance wts. are then ranked in a decreasing order and split into 2 groups (higher and lower importance) with a ratio β by the rank regularization module.
- For the purpose of ensuring that the mean attention wt. of the higher importance group is higher than that of the lower importance group, a rank regularization loss (RR- loss) is defined as follows:

δ_1 – Hyper-parameter (fixed or learnable)

α_h, α_l – mean wts. of High and Low Imp. Groups

the total loss func. is = $\gamma L_{RR} + (1-\gamma) L_{WCE}$

$$\mathcal{L}_{RR} = max\{0, \delta_1 - (\alpha_H - \alpha_L)\},$$

with

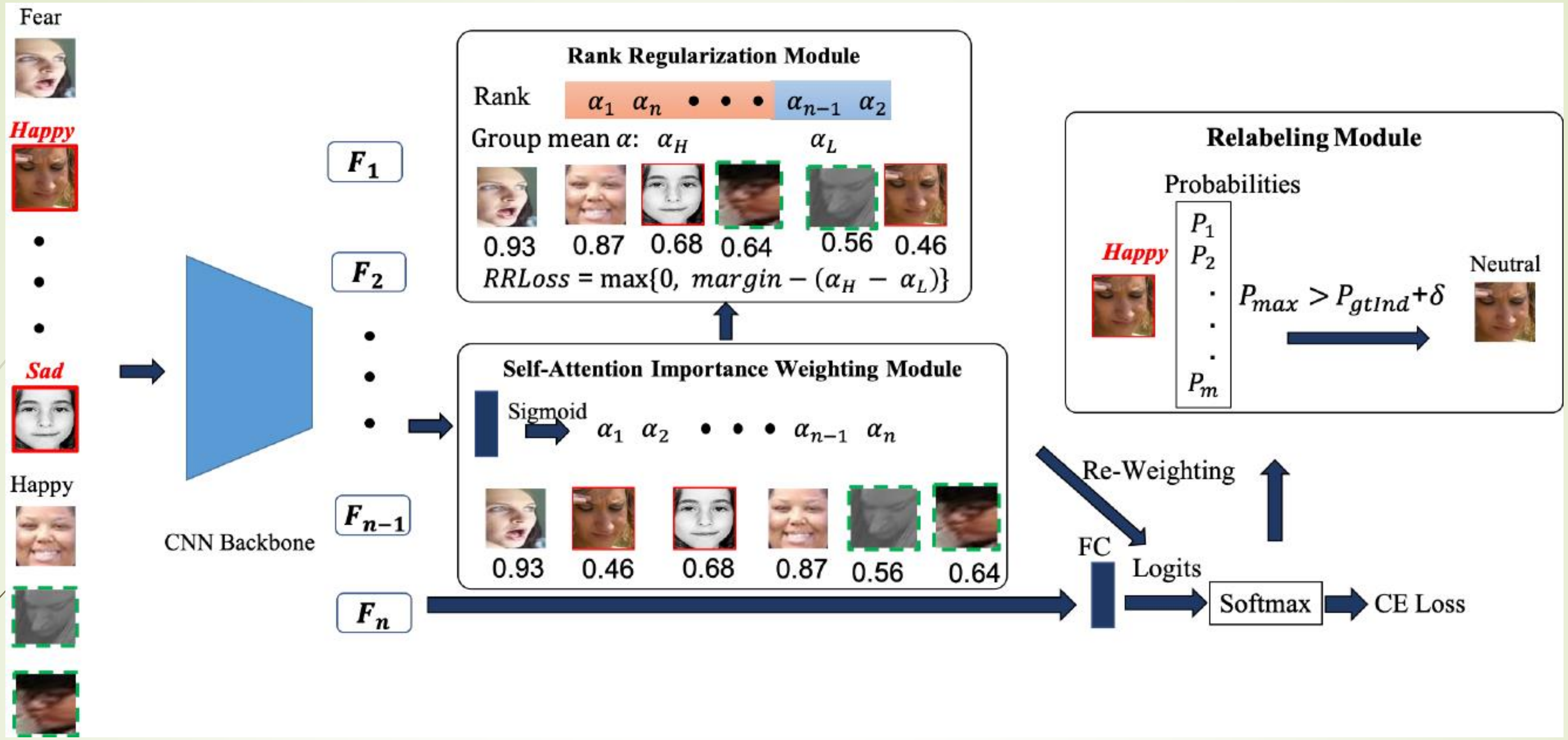
$$\alpha_H = \frac{1}{M}\sum_{i=0}^{M}\alpha_i, \alpha_L = \frac{1}{N-M}\sum_{i=M}^{N}\alpha_i,$$

- Finally, the Relabeling module compares the max. predicted probability to the probability of the given label for each low-imp. group sample. A sample is given the new index of maximum prediction only if the max. prediction probability is higher than the one of the given label by a threshold.

$$y' = \begin{cases} l_{max} & \text{if } P_{max} - P_{gtInd} > \delta_2, \\ l_{org} & \text{otherwise,} \end{cases}$$

Results:

- When compared with previous models on several different datasets with both synthetic (deliberately introduced) and real-world uncertainties, the SCN model outperforms all of them. This is because the SCN can correct the dataset and simultaneously train itself for FER, thereby learning better facial features and leading to better generalization; whereas the other models also learn incorrect information from the misleading data points and perform worse.



(a) Comparison on RAF-DB.

Method	Acc.
DLP-CNN [22]	84.22
IPA2LT [43]	86.77
gaCNN [24]	85.07
RAN [42]	86.90
Our SCN (ResNet18)	**87.03**
Our SCN (ResNet18) ‡	**88.14**

(b) Comparison on AffectNet.

Method	mean Acc.
Upsample [32]	47.00
Weighted loss [32]	58.00
IPA2LT‡ [43] (7 cls)	55.71
RAN [42]	52.97
RAN^+ [42]	59.5
Our SCN^+(ResNet18)	**60.23**

(c) Comparison on FERPlus

Method	Acc.
PLD* [5]	85.1
ResNet+VGG [18]	87.4
SeNet50* [1]	88.8
RAN [42]	88.55
RAN-VGG16* [42]	89.16
Our SCN (ResNet18/IR50)	88.01/**89.35**