Tom Kawchak

Progress (Literature Review):

- Tencent AI Research Paper: Range Loss for Deep Face Recognition with Long-Tailed Training Data
  - Tencent is the current leader for MegaFace and LFW competitions.
  - Presents a way to overcome the long-tailed distribution of data (few people with many photos, many people with few photos).
  - We empirically find that popular training methods in deep face recognition, i.e. contrastive loss, triplet loss, and center loss, all suffer from long tail distributions, while removing long tailed data can improve the recognition performance.
  - Extensive experiments and analysis on two famous face benchmarks demonstrate the proposed range loss can largely relieve the long-tail effect and achieve superior performance than previous methods.
  - This is applied to the VGG Deep 16 model, which we analyzed.

  - Some equations of loss functions:

$$\mathcal{L}_R = \alpha \mathcal{L}_{R_{intra}} + \beta \mathcal{L}_{R_{inter}} \tag{1}$$

    α, β - parameter constants.

$$\mathcal{L}_{R_{intra}} = \sum_{i \subset I} \mathcal{L}^i_{R_{intra}} = \sum_{i \subset I} \frac{k}{\sum_{j=1}^{k} \frac{1}{D_j}} \tag{2}$$

    I - The complete set of Identities in a batch.
    D_j - The jth largest distance.

$$\mathcal{L}_{R_{inter}} = max(m - D_{Center}, 0) = max(m - ||\bar{x}_{\mathcal{Q}} - \bar{x}_{\mathcal{R}}||_2^2, 0) \tag{3}$$

    D_{Center} - The shortest distance between class centers.
    m - a parameter that specifies a max margin for which the loss will not be computed.
    k - The index of the k^{th} largest within each class.

    The final loss function with softmax:

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_R = -\sum_{i=1}^{M} log(\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}) + \lambda \mathcal{L}_R \tag{4}$$

Tencent's Residual Net architecture:

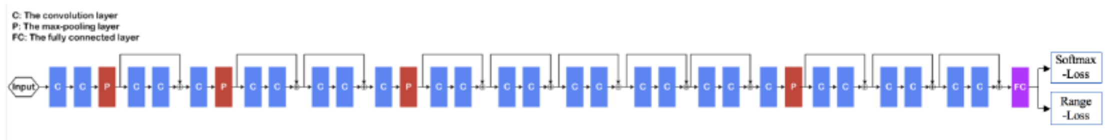

Figure 7. Residual Network's structure adopted in our experiment. All the convolutional filters' size are 3×3 with stride 1. Activation units ReLu layers are added after each convolutional layers. The number of the feature maps are 32 from the front layers to 512 in the last layers. We set the max-pooling's kernel size as 2×2 with stride 2. Features in the last convolutional layer and the penultimate convolutional layer are extracted and concatenated as the input of the last fully connected layers. The whole CNN is trained under the joint supervisory signals of soft-max and our range loss.