Margin-Mix: Semi–Supervised Learning for Face Expression Recognition

Corneliu Florea[1][0000−0001−9754−6795], Mihai Badea[1], Laura Florea[1], Andrei Racoviteanu[1], and Constantin Vertan[1]

Image Processing and Analysis Laboratory, Unversity Politehnica of Bucharest
{corneliu.florea; mihai_sorin.badea; laura.florea; andrei.racoviteanu; constantin.vertan }@upb.ro

1 Domain comparison: object recognition vs face expression recognition

In the main paper, we emphasized that face expression recognition (FER) domain is more compact than the object recognition domain. More precisely, FER has smaller inter class variance.

Given the original dimensionality of the two domains ($32 \times 32 \times 3 = 3072$ for CIFAR images and respectively $48 \times 48 = 2304$ for FER+ images), to our best knowledge there is no universally accepted technique to visualize the distribution of the points in the original space.

In this material, we propose some approaches widely used to visualize the original spaces and to provide an intuition about our proposal.

1.1 Low dimensionality in the CNN embedding space

In this case, for the convolutional neural network (CNN), we use the smaller AlexNet [2]. There, on the last layer before the prediction, we impose a dimensionality of 2 instead of the standard 4096. Instead of the WideResNet, the AlexNet type of network is able to accept such a modification with a dramatic change in performance (which nevertheless is smaller). In this scenario, the network was trained in purely supervised manner with the large margin loss. The distribution of the training points from the two domains, illustrated by the FER+ and CIFAR examples can be seen in figure 1.

One may easily note that data in the FER domain is more tightly clustered and in the first epochs, the clusters are overlapping no matter the classes. As epoch progress and the networks are adjusted for better separation, one may see that CIFAR corresponding become more distinct, while the FER are less so. Thus we argue that the difference in the distribution of points requires different strategy.

1.2 Visualization using t-SNE

Another popular technique for visualization high dimensional data is t-SNE [4]. In this case, we have considered the original architecture with full 4096 dimensionality and employed t-SNE for visualization of data structure with respect to the number of epochs. The plots may be seen in figure 2. This visualization technique also suggests more overlapped clusters in the case of FER problem than for object recognition (CIFAR).

2 Separated vs overlapping clusters

In the main paper, we compare our proposal with the MixMatch solution [1]. The main difference lies in the self-labeling step: our proposal, MarginMix, uses distances from each unlabeled data to the class centroids and those distances are turned into **soft** labels (i.e. probabilities / membership values). In contrast MixMatch uses PseudoLabels [3]. The latter bases the self-labels on the borders defined by the annotated data. Un-annotated data is thus **hard** labeled. We show an intuitive graphical comparison between these two strategies in figure 3 for a situation where class clusters are well separated and respectively in figure 4 for overlapping clusters.

The main idea is that PseudoLabels is very conservative on the original chosen border as it imposes that unlabeled data to have sharp labels (i.e only $(1, 0)$ or $(0, 1)$). Its advantage lies in better definition of the class borders in areas where there are too few points. However, in the case where the clusters corresponding to points from different classes are overlapping, the initial border is not well defined, as it depends on chance, and PseudoLabels does not encourage strong modifications from that initial case.

In contrast, our proposal uses self labeling based on large margins (distance to class centroids). It produces soft values for the labels (i.e $(0.34, 0.66)$ or $(0.72, 0.28)$, etc). It also behaves correctly in the case of separated clusters, but is being far more malleable in the case of overlapping class clusters and thus its performance is more robust.

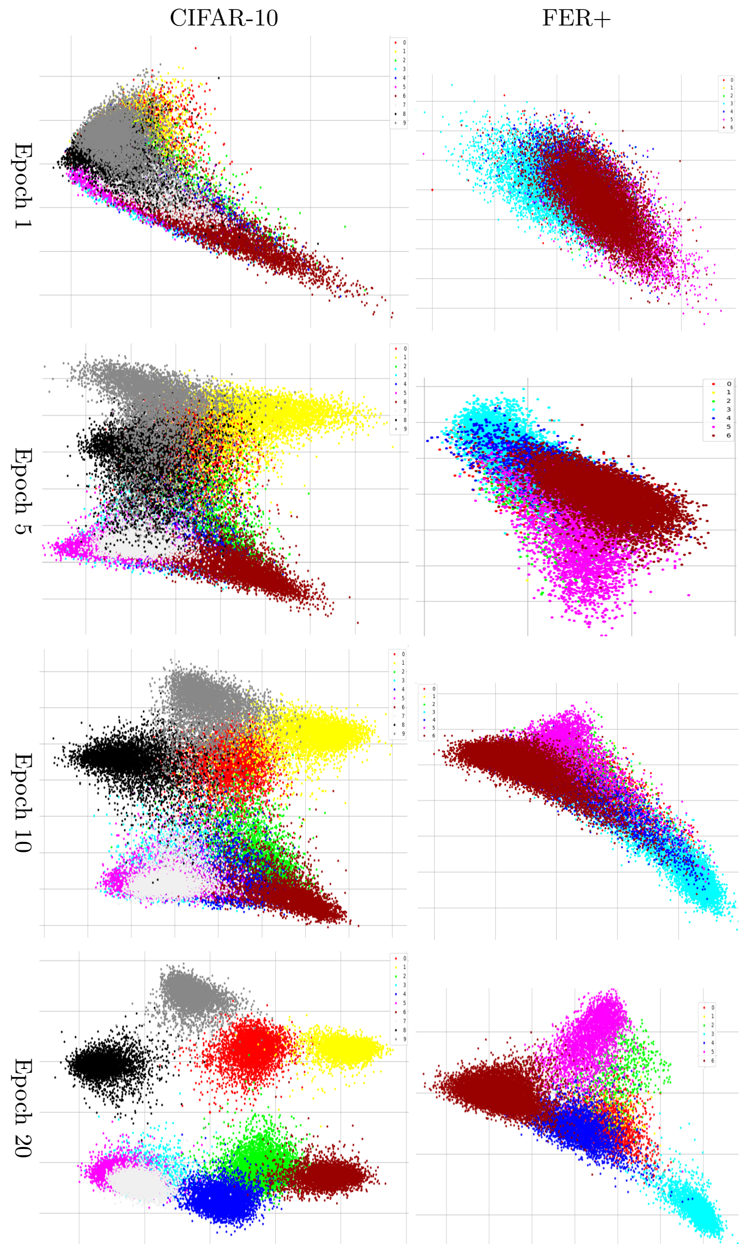

Fig. 1. Examples of the two problems approached. The training set is represented on the 2 neurons from the last layer fully connected before the decision one in an AlexNet. On the left column is represented the distribution of points in CIFAR-10 database while on the right column is FER+. One notes the degree to which the clusters corresponding to different classes are overlapped/separable in the two situations

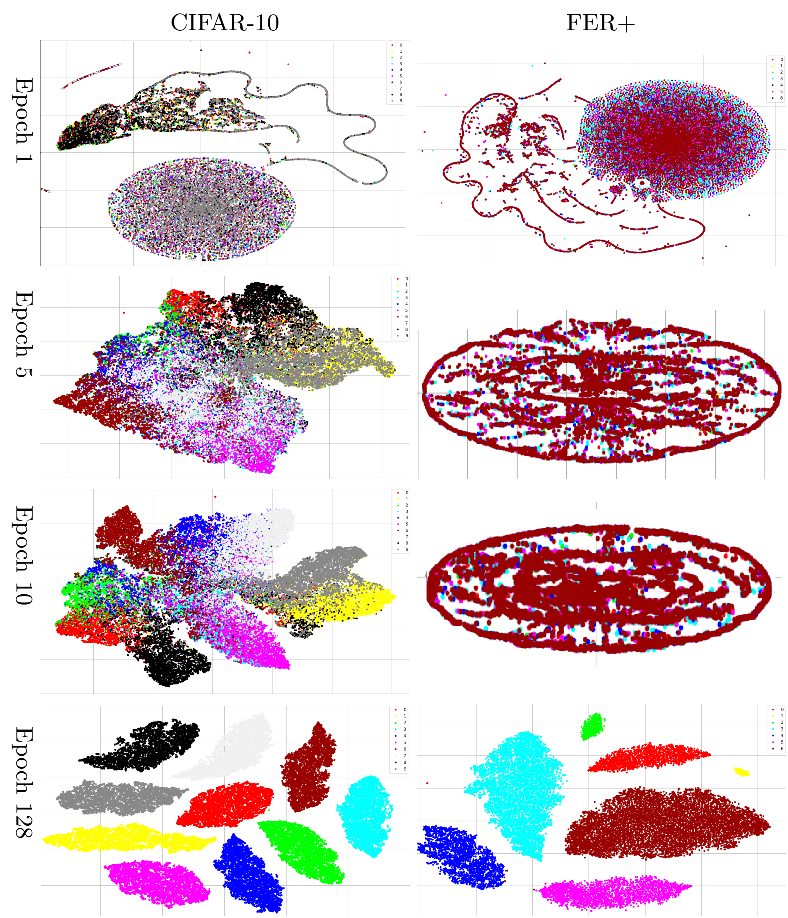

Fig. 2. Examples of the two problems approached. The training set is represented on the 4096 dimensional layer and visualized with t-SNE. On the left column is represented the distribution of points in CIFAR-10 database while on the right column is FER+. One notes the degree to which the clusters corresponding to different classes are overlapped/separable in the two situations

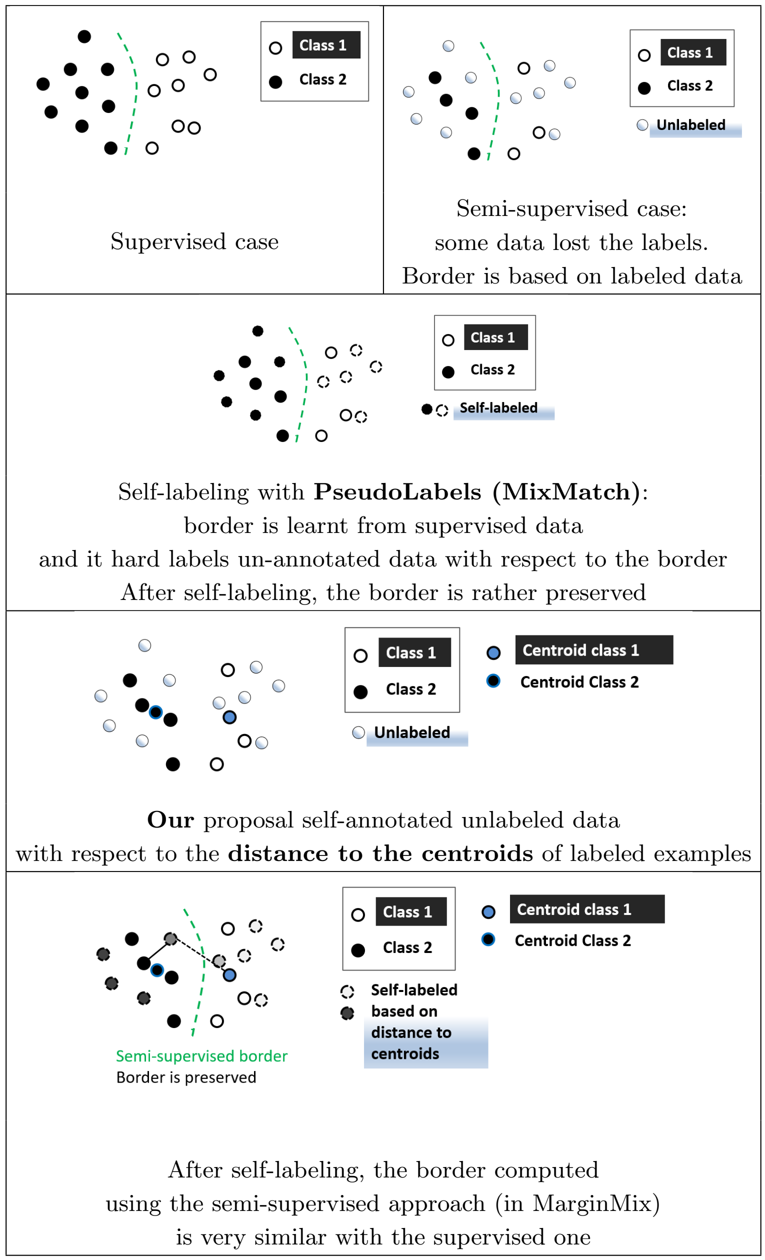

Fig. 3. Comparison between the proposed algorithm based on self labeling using distances to centroids and MixMatch that uses entropy regularization (PseudoLabels) on clusters that are well separated. Self-labeled data points color is proportional to the label value. Both algorithms are successful.

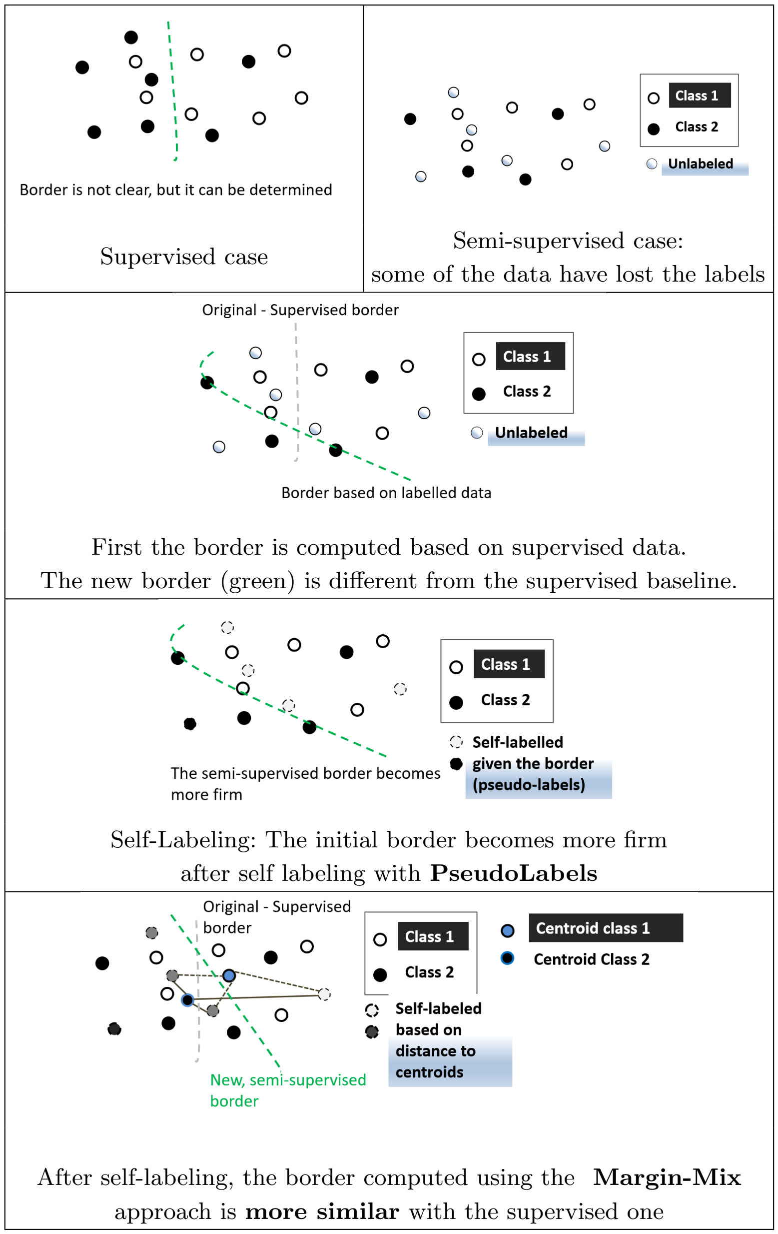

Fig. 4. Comparison between the proposed algorithm based on self labeling using distances to centroids and using entropy regularization (PseudoLabels) on overlapping clusters. Note the strength of the labels in the case PseudoLabels and in case of distances to centroids. Only our algorithm is successful in this case.

References

1. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. In: NIPS. pp. 5050–5060 (2019) 2
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012) 1
3. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICML Workshops (2013) 2
4. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008) 1