

A FLEXIBLE FRAMEWORK FOR DISCOVERING NOVEL CATEGORIES WITH CONTRASTIVE LEARNING

–SUPPLEMENTARY MATERIAL–

Anonymous authors

Paper under double-blind review

1 COMPARING WTA WITH OTHER ALTERNATIVES

We compare WTA and other alternatives to transfer pseudo labels on-the-fly in our framework. We compare WTA with cosine similarity, ranking statistics (Han et al., 2020), and nearest-neighbour. Table 1 shows the results. We can see that WTA performs significantly better than all other alternatives.

Table 1: **WTA vs other alternatives.**

| Method | Kinetics-400 | VGG-Sound |
|---------------------------------------|--------------|-----------|
| cosine | 35.4% | 29.8% |
| nearest-neighbour | 19.6% | 11.5% |
| ranking statistics (Han et al., 2020) | 37.4% | 38.6% |
| WTA | 56.5% | 50.0% |

2 QUALITATIVE RESULTS

We show the qualitative results on images and videos. In fig. 1 and fig. 2 we show the t-SNE projection for the features of data from the 5 unlabelled classes in CIFAR-10. The features are extracted using the model pretrained on the labelled data (fig. 1) and using our model (fig. 2) respectively. We can see that the embedding is rather cluttered using the model pretrained on the labelled data, while our model can successfully partition the unlabelled data into tight semantic groups. Similarly, we compare the features on videos in fig. 3 and fig. 4. For visualisation purpose, we randomly choose unlabelled instances from 10 classes from Kinetics-400, we can see that our model can successfully separate novel classes into compact groups, while the novel categories are projected very close to each other for the baseline model trained with full supervision on the labelled data.

REFERENCES

Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. In *ICLR*, 2020.

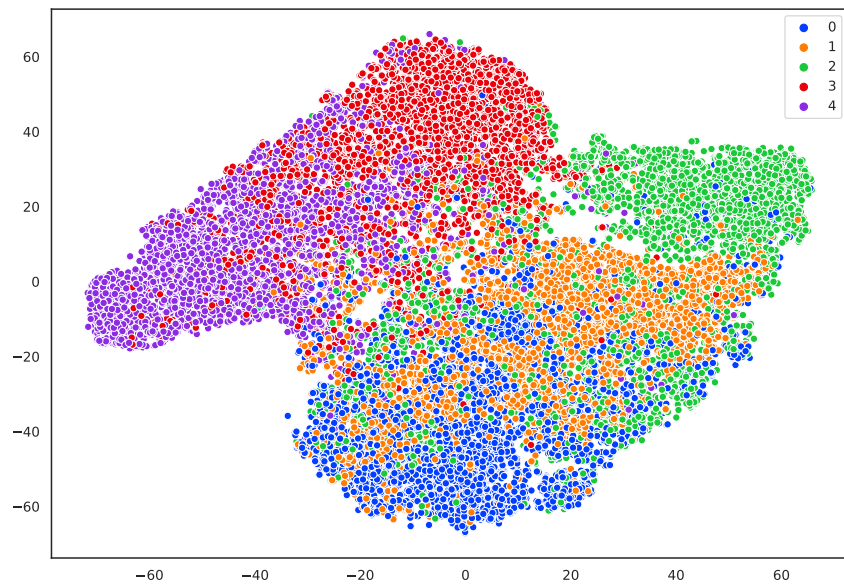


Figure 1: t-SNE visualization of features from unlabelled data on CIFAR-10, using the model pretrained on the labelled data with fully supervised training.

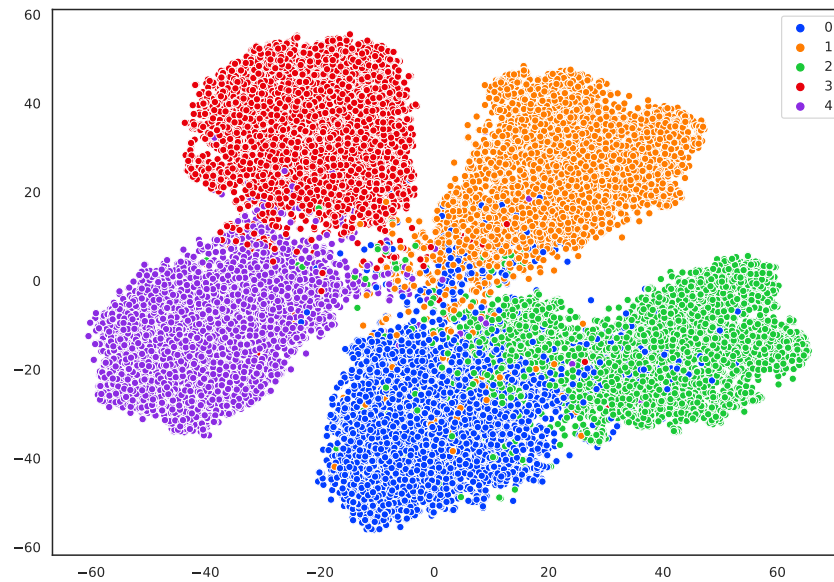


Figure 2: t-SNE visualization of features from unlabelled data on CIFAR-10, using our model after end-to-end training on both labelled and unlabeled data.

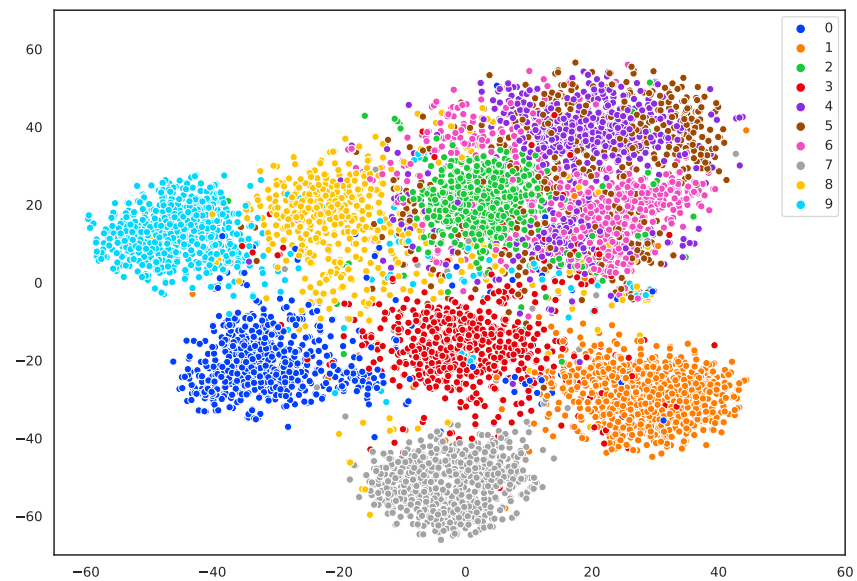


Figure 3: t-SNE visualization of features from unlabelled data on Kinetics-400, using the model pretrained on the labelled data with fully supervised training.

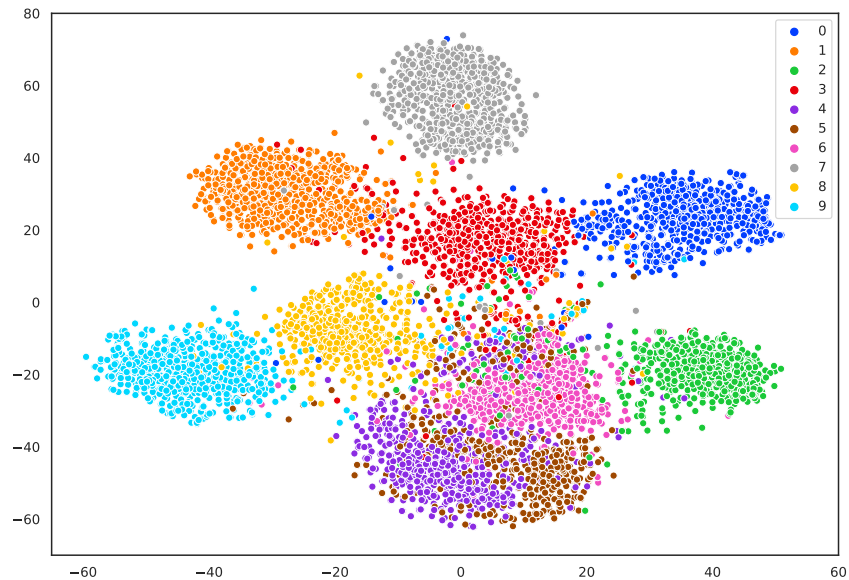


Figure 4: t-SNE visualization of features from unlabelled data on Kinetics-400, using our model after end-to-end training on both labelled and unlabelled data.