Temperature as Uncertainty in Contrastive Learning

Oliver Zhang[1], Mike Wu[1], Jasmine Bayrooti[1], Noah Goodman[1,2]
Departments of Computer Science[1] and Psychology[2], Stanford University

Introduction

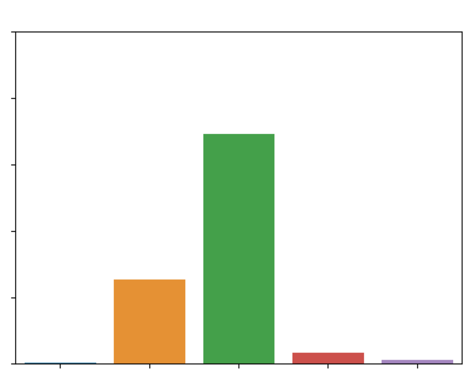
- Contrastive learning learns strong embeddings without labels.
- We induce uncertainties over the learned contrastive embeddings by repurposing the temperature term in the contrastive learning loss.
- We have the model predict a new temperature per training example as a measure of uncertainty. We call this method "Temperature as Uncertainty (TaU).
- We demonstrate that TaU can do out of distribution (OOD) detection, outperforming other embedding uncertainties. Moreover, TaU can be learned post-hoc on frozen models—even ones which were not trained with contrastive learning.

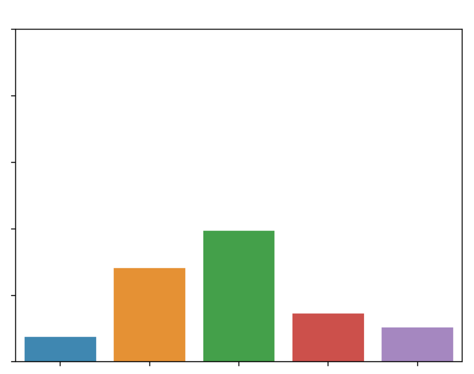
Methods

Let $f(x)$ be a function, parameterized by a neural networks, which maps an image x to an L_2 normalized representation. Let $p(t)$ be a distribution of augmentations and τ the softmax temperature. TaU has τ depend on x. The TaU objective function is then:

$$\mathscr{L}(x_i) = \mathbf{E}_{t,t',t_{1:k}\sim p(t)}\mathbf{E}_{x_{1:k}\sim p(x)}\left[\log \frac{e^{f(t(x_i))\cdot f(t'(x_i))/\tau(x_i)}}{\frac{1}{k}\sum_{j\in\{1:k\}} e^{f(t(x_i))\cdot f(t_j(x_j))/\tau(x_i)}}\right]$$

In short, this can be viewed as the SimCLR objective [2] with learned temperature. Temperature modifies how "sharp" the distribution is, which makes it good for representing uncertainty.





Finally, we employ many of the stability tricks from [1].

Out of Distribution Detection

Out of distribution (OOD) detection involves comparing the distributions of uncertainty scores for two distributions. Specifically, we can quantify how different the distributions are. In all the experiments, we use the AuROC when thresholding the uncertainty scores for classification.

Experiments

We test TaU, trained from scratch, on out-of-distribution (OOD) detection. TaU outperforms a variety of baselines in quantifying embedding uncertainty. For clarity, SimCLR + kNN and MoCo-v2 + kNN represent applying a kNN algorithm over SimCLR and MoCo-v2 embeddings and using the average distance of the k nearest neighbors as uncertainty.

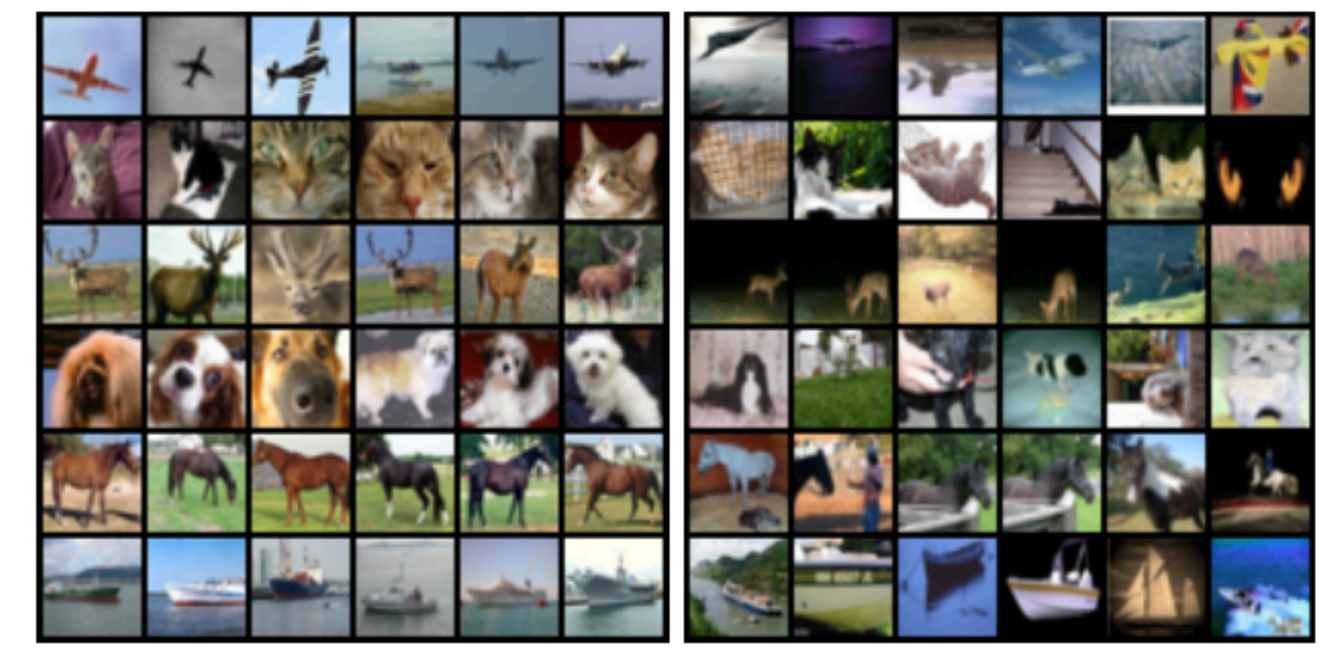
Method (CIFAR10)	**CIFAR100**	**SVHN**	**TinyImageNet**
TaU + SimCLR	**0.746**	0.964	**0.760**
TaU + MoCo-v2	0.728	**0.968**	0.746
SimCLR + kNN	**0.746**	0.829	0.756
MoCo-v2 + kNN	0.712	0.800	0.726
SimCLR + MC Dropout [9]	0.504	0.684	0.512
Supervised + MC Dropout [9]	0.659	0.745	0.722
Hedged Instance Embedding [23]	0.509	0.834	0.508
Ensemble of 5 SimCLRs	0.532	0.525	0.513

Finally, we test TaU's OOD detection using pretrained embeddings. We reshape images to 224x224 and compare against ImageNet images. We find that TaU develops good OOD detection capabilities, even when the pretrained embeddings are frozen. Surprisingly, TaU works even with embeddings from non-contrastive models such as a ResNet trained on a supervised learning objective.

Method (ImageNet)	**CIFAR10**	**CIFAR100**	**SVHN**	**TinyImgNet**	**LSUN**	**COCO**	**CelebA**
TaU + Supervised	**0.913**	**0.874**	**0.978**	**0.771**	0.657	0.458	0.657
TaU + SimCLR [5]	0.823	0.870	0.968	0.747	0.552	0.554	0.717
TaU + BYOL [12]	0.763	0.808	0.955	0.686	0.471	0.497	0.840
TaU + CLIP [27]	0.056	0.044	0.071	0.154	**0.779**	**0.579**	**0.883**

Visualization

We visualize Cifar10 images using TaU. The most certain images are on the left and the most uncertain images are on the right.



Limitations & Future Work

We notice two important limitations with TaU. To start, training TaU from scratch does not work with stop-gradient based contrastive learning methods, such as SimSiam or BYOL. For these methods, the embeddings must be pretrained. Moreover after employing the stability tricks from [1], the actual temperature values quickly converge to the lower bound of one. In the end, the differences between the temperatures become very small quickly.

Future work includes trying TaU on a variety of other cases in which uncertainty may be useful (e.g., corruptions, calibration, adversarial examples). Other future work might consist of applying TaU to other modalities, such as language.

References

[1] Lukas Neumann, Andrew Zisserman, and Andrea Vedaldi. Relaxed softmax: Efficient confidence auto-calibration for safe pedestrian detection. 2018.
[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.