Contrastive Learning with Latent Tension Regularization for Tight Orbits

Ritwik Ghosal
ritwikghosal1999@gmail.com

NEURAL INFORMATION PROCESSING SYSTEMS

What is the work about ?

This work introduces **Orbit Regularization Loss (ORL)**, a contrastive learning objective designed to enforce geometric consistency within the latent orbits formed by augmented views of the same sample. ORL reweights negative pairs using a **tension score**, which quantifies how a negative sample's direction aligns with the positive-pair displacement. This *strengthens* the geometric structure of the learned representation without requiring architectural modifications or additional supervision. Experiments on MNIST and CIFAR-10 show that ORL yields tighter orbits and more consistent latent trajectories.

In a less formal way :

ORL strengthens negative repulsion based on how similarly a negative is oriented relative to the anchor. For example, representation of digit '7' lies in a direction more similar to digit '1' than '8' does, then ORL treats '7' as a more *dangerous* negative and pushes it harder. In other words, negatives that intrude into the anchor's preferred latent direction receive ***stronger penalties***, helping keep the orbit of augmented views compact and directionally stable.

Where does it exactly hit :

$$\mathcal{L}_{\mathrm{NTX}}(z_i, z_j) = -\log \frac{\exp\left(\frac{\mathrm{sim}(z_i,z_j)}{\tau}\right)}{\sum_{k=1}^{2N} \mathbf{1}_{k\neq i} \cdot \exp\left(\frac{\mathrm{sim}(z_i,z_k)}{\tau}\right)}$$

The Nt-Xent loss

$$\mathcal{L}_{\mathrm{ORL}}(z_i, z_j) = -\log \frac{\exp\left(\frac{\mathrm{sim}(z_i,z_j)}{\tau}\right)}{\sum_{k=1}^{2N} \mathbf{1}_{k\neq i} \cdot \exp\left(\frac{\mathrm{sim}(z_i,z_k)\cdot T_{ik}}{\tau}\right)}$$

The OR Loss

Where,

$$T_{ik} = \cos(\theta_{ik}) = \frac{\delta_{pos}^{\top}\delta_{ik}}{\|\delta_{ik}\|}$$

Given a positive pair with embeddings $z_1, z_2 \in \mathbb{R}^d$

$$\delta_{pos} = \frac{z_2 - z_1}{\|z_2 - z_1\|} \quad \text{and} \quad \delta_{ik} = z_k - z_i$$

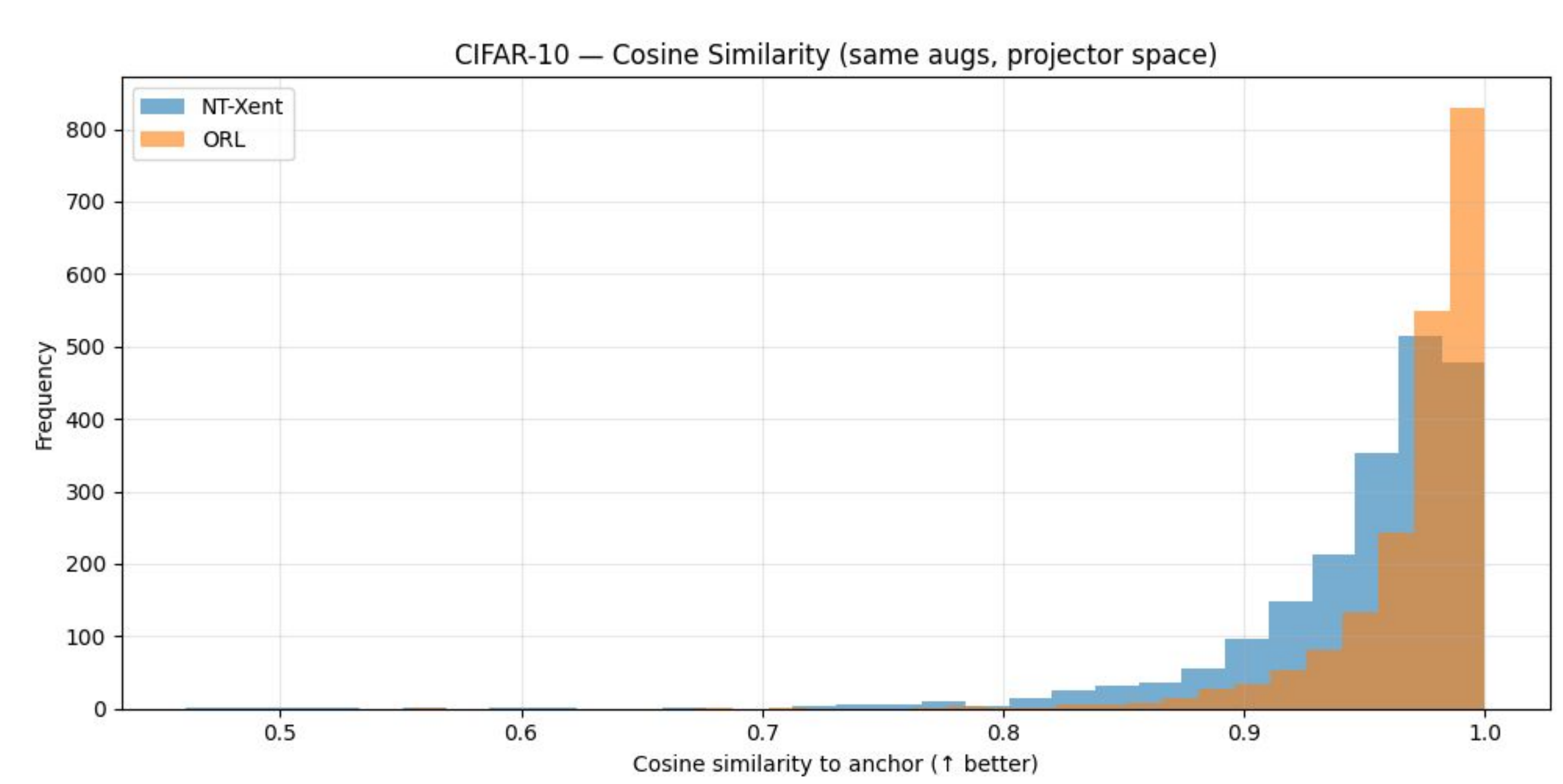
$$T_{ik} \rightarrow Tension\ score$$

Intuitively, the 'tension' captures how *on-axis* or *off-axis* a negative is, relative to the motion between two augmented views.

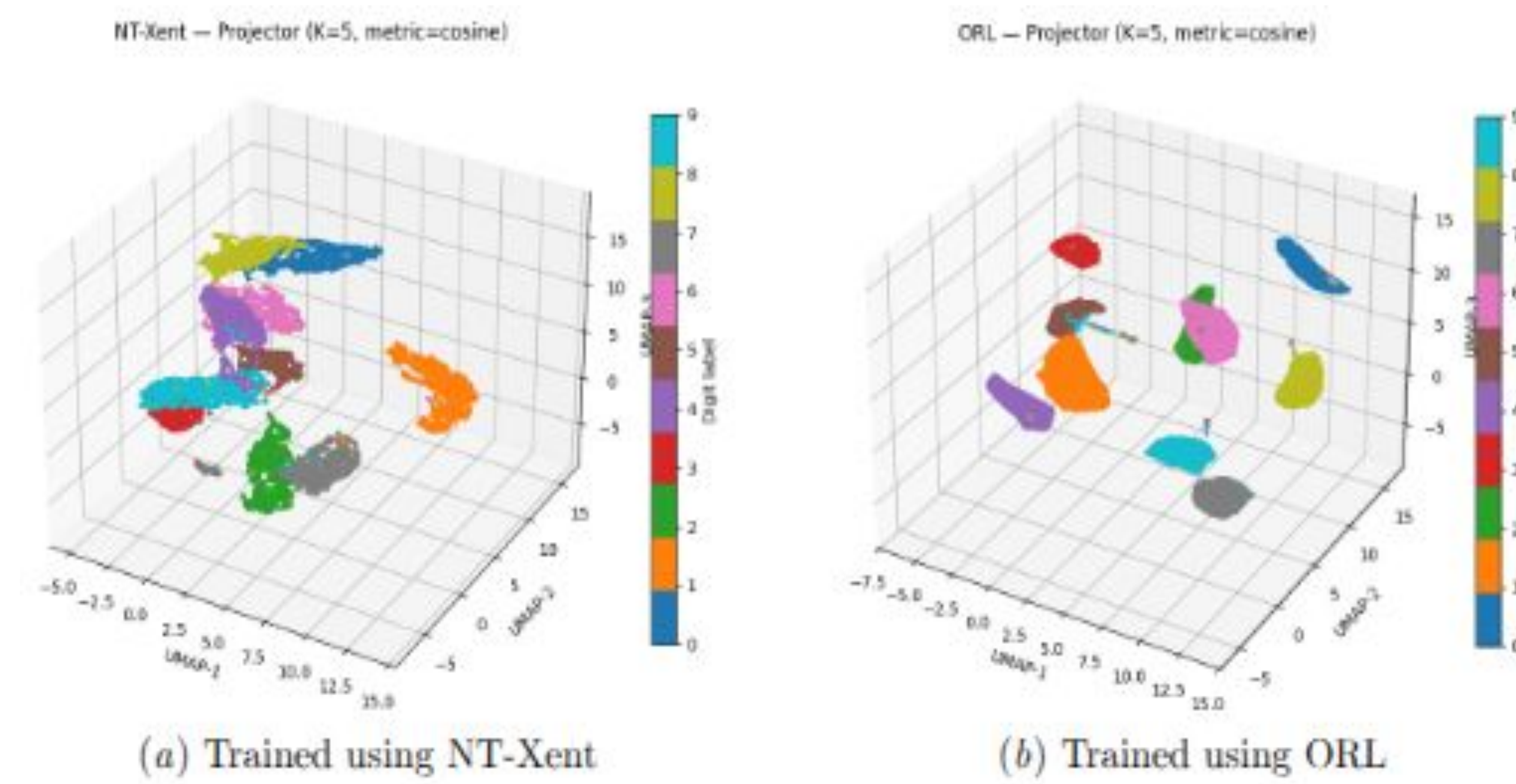
For the potential sign pitfall, the tension score is limited as:

$$T_{ik} \in [10^{-6}, 1]$$

ORL preserves the fundamental contrastive principle: pull positives together, push negatives apart. Its novelty lies in recognizing that not all negatives are equally informative. By modulating their influence with tension, ORL filters noisy gradients that can distort orbit geometry, yielding a principled, direction-aware generalization of NT-Xent.



Cosine similarity histogram between anchors and augmented views for NT-Xent and ORL (projector space). Evaluated on 200 anchors (images from the CIFAR-10 test set) with K = 10 views each (200 × 10 = 2000 anchor–view pairs per model).



UMAP of MNIST projector-space embeddings with cosine metric (K=5 views/image)

As we allow gradients to flow through the tension score itself, ORL introduces an additional alignment-shaping force during training. This extra gradient term adjusts the positions of negatives according to how well they align with the orbit direction, actively sculpting the latent geometry rather than relying solely on standard similarity-based repulsion.

To sum up:

- The work, at present, does not aim to improve downstream classification performance, instead, try to extend the contrastive learning work[1].
- Introduces a lightweight refinement of the NT-Xent contrastive loss focused on geometric consistency.
- Modifies contrastive learning only through the loss - no additional parameters, networks, or architectural changes.
- Achieves higher cosine similarity between positive pairs and consistently lower orbit variance.
- **Overall Positioning**: Aimed as a lightweight and effective directional enhancement to standard contrastive learning of SimCLR.

References

[1] 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, 2020a.