

Attempting To Boost TCR/Epitope Binding Affinity With Supervised Contrastive Loss

Nicholas Moran, Computer Science (BS)

Mentor: Dr. Heewook Lee, Assistant Professor

Ira A. Fulton Schools of Engineering

Introduction

- TCR/Epitope binding is crucial to the immune response
- Several deep learning models have been developed with the goal of accurately predicting this binding affinity
- Supervised Contrastive Loss (SCL) separates label classes in latent space
- In this work, SCL is used as a pre-training step, before Cross Entropy is used for classification

Supervised Contrastive Loss

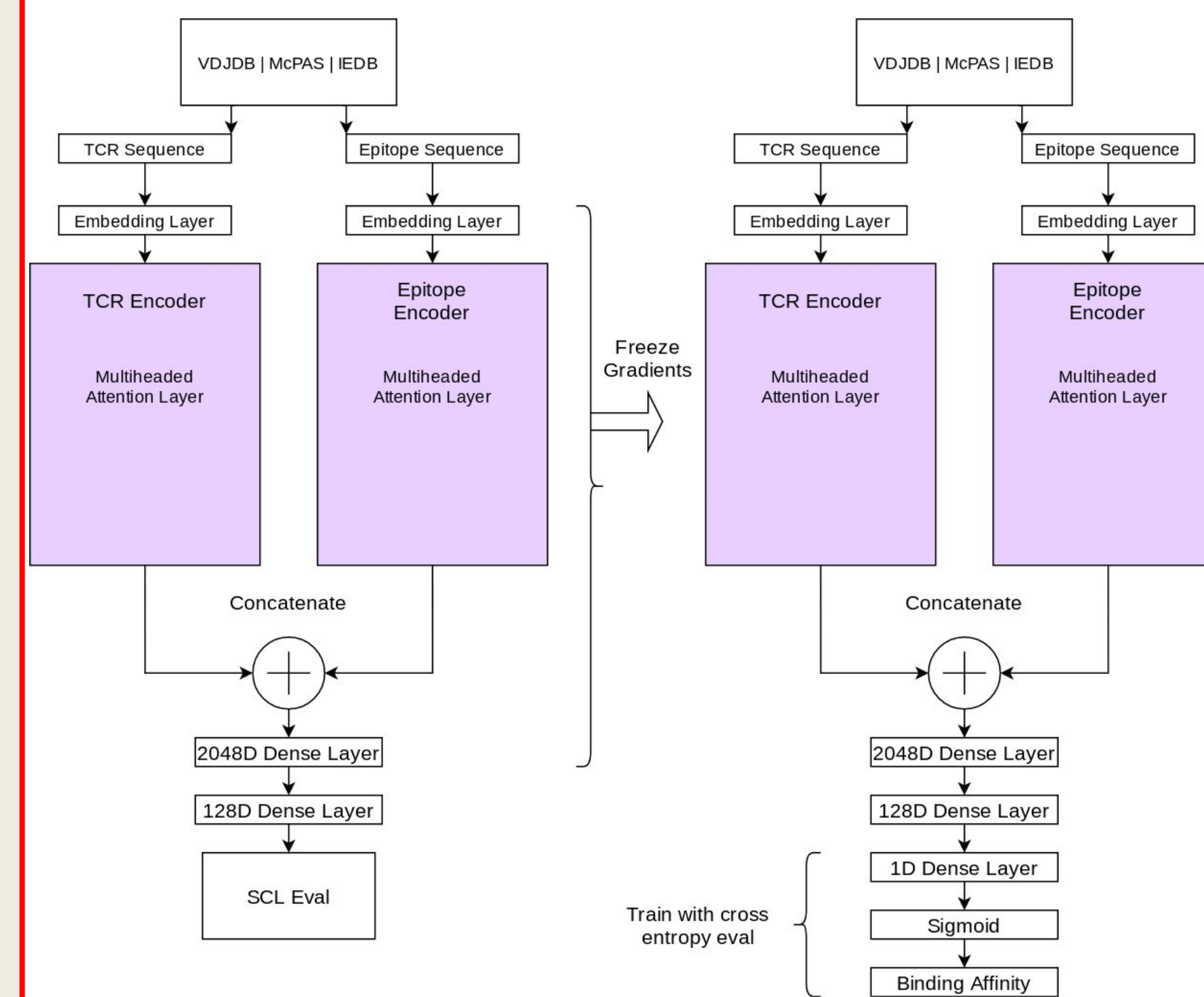
- Supervised contrastive loss attempts to pull classes apart in latent space, so later classification is an easy task
- In order to evaluate the model, a high dimensional projection network is trained against SCL. The projection network is thrown out at inference time.
- The function below is the canonical SCL function.

$$\mathcal{L}_{out}^{sup} = \sum_{i \in I} \mathcal{L}_{out,i}^{sup} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)}$$

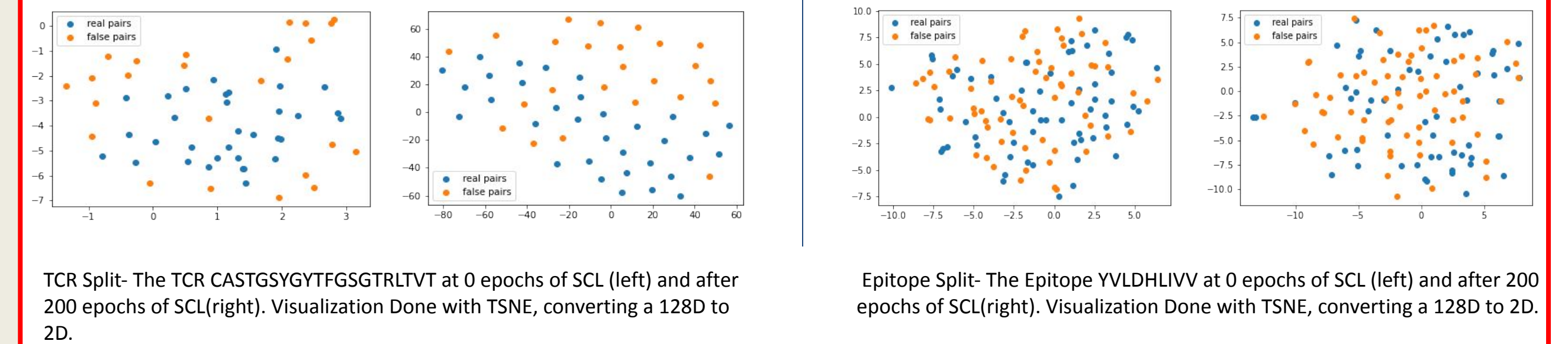
Data Preprocessing

- Data is built from VDJDDB, McPAS, and IEDB, databases
- We have two split types: tcr and epitope
- In each type, 5-fold validation is used, where the test set in each is a set of unseen types. This is done because our goal is to accurately predict out-of-class TCR/Epitope bindings
- The TCR sequences are 20 characters long, and the Epitope sequences are 22

Model Architecture



Visualizations



Results

Model Type	TCR Split AUC	Epitope Split AUC
Base Model- No SCL	77.3%	47.0%
Model with SCL	67.6%	53.2%

Conclusions and Further Work

- The model with SCL does not beat the base model on the TCR split, but does outperform on the Epitope Split.
- This is likely due to the fact that the embedding is being learned in the SCL phase, not the binding affinity phase
- In future work, pretrained embeddings should be used in the SCL phase. This would allow the model to focus on optimizing the distance between classes far better

References

- Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2020). Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning.
- Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised Contrastive Learning.
- Weber, A., Born, J., & Rodriguez Martínez, M. (2021). TITAN: T-cell receptor specificity prediction with bimodal attention networks. *Bioinformatics*, 37(Supplement_1), i237-i244.
- Michael Cai, Seojin Bang and Heewook Lee. TCR-epitope binding affinity prediction using multi-head self attention model