

## Appendix I. Training Environment, Hyperparameters, and RTT Settings

### Training Environment

All experiments for this work are done on a single NVIDIA A800 GPU on Ubuntu 20.04, python 3.6, torch 1.7, and transformers 4.2.1.

### Hyperparameters and RTT Settings

We report the important hyperparameters and datasets of the pre-trained encoder, TNCSE-BERT/RoBERTa(D, UC D), in Table I-1.

TNCSE is divided into two stages of training: the first stage is the pre-training encoder, and the second stage is the training of the dual-tower model, which adds tensor norm constraints. For the first stage, we set the learning rate search space from 1e-5 to 2e-5, and the step size is set to 2.5e-6. The second stage must reduce the learning rate, which is uniformly set to 1e-6. TNCSE-D and TNCSE-UC D are cumulative training, and the learning rate search space and step size are the same as the first stage of TNCSE. The batch size is set to 64, 128, 256, and we do not tune for random seeds, which is uniformly set to 42. We evaluate the performance of the model on the STS-B validation set once every one hundred steps of training, as a checkpoint saved metric.

For the training set, we uniformly employ Wiki1M and unlabelled SICKR. For the selection of RTT languages, four randomly selected languages, en-fr-en, en-ru-en, en-nl-en, and en-zh-en, are used for the first stage of the training of the encoder for TNCSE-BERT/RoBERTa. For the other models in the TNCSE series, we uniformly do not use RTT Data Augmentation.

### Appendix II. Impact of Pooling methods

We follow the approach of the SimCSE to take the last hidden state corresponding to the CLS token as the model’s output. We further explore the impact of the average pooling of all tokens and pooler output, the experimental results are reported in Table II-1. Average pooling may weaken the features of high-weight outliers. In our approach, the pooler output is only used to assist in optimizing the gradient update of the encoder, which is less rich in semantic distribution information than the last hidden state and results less effectively. CLS pooling aggregates the features of the whole sample and performs best in the STS task.

| Pooling Method | CLS          | Avg.  | Pooler Output |
|----------------|--------------|-------|---------------|
| 7 STS Avg.     | <b>79.58</b> | 78.19 | 69.28         |

Table II-1: This table reports the impact of pooling methods.

### Appendix III. Alignment and Uniformity

Wang and Isola (2020) proposes to use two key metrics, Alignment and Uniformity, to evaluate the representation. Given that a pair of positive samples  $p_{pos}$ ,  $l_{align}$  is defined as the expectation of the second-paradigm of the difference

between the vector representations of this pair of positive samples, defined as Eq. III-1:

$$l_{align} = \mathbf{E}_{(x, x^+ \sim p_{pos})} \|f(x) - f(x^+)\|^2, \quad (\text{III-1})$$

where  $\|\cdot\|$  denotes the Euclidean paradigm.  $l_{uniform}$  is defined as the uniform case of the distribution between sample pair that are mutually soft-negative, defined as Eq. III-2:

$$l_{uniform} = \log \mathbf{E}_{i.i.d} e^{-t\|f(x)-f(y)\|^2}, t > 0, \quad (\text{III-2})$$

where  $x, y \sim p_{data}$  and  $t$  is set to 2. The significance of both of the above in unsupervised comparative learning is that the representations of both mutually positive samples should be close to each other, and the mutually soft-negative samples should be evenly distributed. The smaller both metrics are, the better the model works.

## Appendix IV.. Results of MTEB Downstream Experiments

In this section, we report the results of the zero-shot tests of the TNCSE on the MTEB(Muennighoff et al. 2023). For a more comprehensive evaluation, we evaluate the multilingual and cross-lingual STS tasks of TNCSE and baselines on MTEB, reporting STS17 in Table IV-1 and STS22 in Table IV-2, respectively, and the results show that TNCSE is the optimal one. Since we have limited computational resources to evaluate hundreds of other downstream tasks, we randomly select 30 tasks from each of the text classification and retrieval tasks and the nine re-ranking tasks available in the test set to evaluate TNCSE and baselines. The results of these tests are reported in Tables IV-3, IV-4, and IV-5, respectively. The results show that, overall, TNCSE remains the best. All of the above experiments use **Avg** as the primary metric.

### References

- Muennighoff, N.; Tazi, N.; Magne, L.; and Reimers, N. 2023. MTEB: Massive Text Embedding Benchmark. In Vlachos, A.; and Augenstein, I., eds., *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023, 2006–2029*. Association for Computational Linguistics.
- Wang, T.; and Isola, P. 2020. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, 9929–9939. PMLR.

| Model                | Learning Rate | Batch Size | Datasets                       |
|----------------------|---------------|------------|--------------------------------|
| <b>TNCSE-BERT</b>    |               |            |                                |
| <b>Encoder I</b>     | 1.5e-5        | 64         | Wiki1M+Unlabelled en-fr-en RTT |
| <b>Encoder II</b>    | 2e-5          | 64         | Wiki1M+Unlabelled en-ru-en RTT |
| <b>TNCSE-BERT</b>    | 1e-6          | 64         | Wiki1M+Unlabelled              |
| <b>+D</b>            | 1e-5          | 64         | Wiki1M+Unlabelled              |
| <b>+UC D</b>         | 1.5e-5        | 64         | Wiki1M+Unlabelled              |
| <b>TNCSE-RoBERTa</b> |               |            |                                |
| <b>Encoder I</b>     | 1.5e-5        | 64         | Wiki1M+Unlabelled en-nl-en RTT |
| <b>Encoder II</b>    | 1.25e-5       | 64         | Wiki1M+Unlabelled en-zh-en RTT |
| <b>TNCSE-RoBERTa</b> | 1e-6          | 64         | Wiki1M+Unlabelled              |
| <b>+D</b>            | 1e-5          | 256        | Wiki1M+Unlabelled              |
| <b>+UC D</b>         | 1.5e-5        | 128        | Wiki1M+Unlabelled              |

Table I-1: We report important hyperparameters for model training.

| STS17        | SimCSE       | ESimCSE | DiffCSE | InfoCSE      | SNCSE | WhitenedCSE | RankCSE      | TNCSE | +D    | +UC D        |
|--------------|--------------|---------|---------|--------------|-------|-------------|--------------|-------|-------|--------------|
| <b>ar-ar</b> | 48.98        | 52.72   | 51.00   | <b>54.99</b> | 35.37 | 47.79       | 48.79        | 53.63 | 53.00 | 54.63        |
| <b>en-ar</b> | -2.43        | 4.60    | -4.86   | 0.13         | 7.66  | -4.09       | <b>11.73</b> | -1.94 | -1.61 | 0.21         |
| <b>en-de</b> | 27.59        | 32.13   | 32.68   | 39.82        | 11.64 | 29.20       | 28.90        | 37.95 | 36.12 | <b>39.68</b> |
| <b>en-en</b> | 83.90        | 85.63   | 86.26   | 85.05        | 53.11 | 85.15       | 85.88        | 85.41 | 85.36 | <b>86.38</b> |
| <b>en-tr</b> | 10.58        | 10.67   | -6.25   | 6.83         | 4.11  | -9.39       | <b>15.58</b> | 0.58  | 0.41  | 4.01         |
| <b>es-en</b> | 12.08        | 20.38   | 11.90   | 20.04        | 12.46 | 6.78        | 13.22        | 20.16 | 18.96 | <b>22.46</b> |
| <b>es-es</b> | 69.49        | 71.30   | 71.87   | <b>74.94</b> | 56.06 | 73.27       | 69.64        | 72.54 | 71.95 | 72.39        |
| <b>fr-en</b> | 36.18        | 28.08   | 26.71   | 27.47        | 12.28 | 26.50       | <b>43.60</b> | 40.03 | 39.49 | 39.29        |
| <b>it-en</b> | 20.08        | 15.72   | 11.75   | 12.63        | 6.92  | 13.36       | <b>23.39</b> | 16.73 | 14.61 | 18.51        |
| <b>ko-ko</b> | <b>52.62</b> | 49.67   | 52.00   | 51.06        | 39.64 | 51.00       | 52.58        | 52.05 | 50.86 | 51.96        |
| <b>nl-en</b> | 17.40        | 21.10   | 14.65   | 22.50        | 9.77  | 16.34       | 20.20        | 22.39 | 20.38 | <b>25.52</b> |
| <b>Avg.</b>  | 34.22        | 35.64   | 31.61   | 35.95        | 22.64 | 30.54       | 37.59        | 36.32 | 35.41 | <b>37.73</b> |

Table IV-1: The performance of the TNCSE and baseline models on 11 multilingual/cross-language semantic similarity tasks on the STS17 test set.

| STS22        | SimCSE       | ESimCSE      | DiffCSE      | InfoCSE      | SNCSE | WhitenedCSE  | RankCSE      | TNCSE        | +D           | +UC D        |
|--------------|--------------|--------------|--------------|--------------|-------|--------------|--------------|--------------|--------------|--------------|
| <b>ar</b>    | <b>38.33</b> | 32.48        | 34.94        | 21.08        | 33.58 | 36.08        | 35.33        | 31.26        | 32.88        | 34.03        |
| <b>de</b>    | 24.70        | 28.50        | 24.47        | 18.02        | 2.58  | 24.99        | 24.70        | 28.04        | 27.75        | <b>29.18</b> |
| <b>de-en</b> | 13.13        | 29.80        | 33.63        | <b>37.03</b> | 20.78 | 30.33        | 35.51        | 33.11        | 31.83        | 34.89        |
| <b>de-fr</b> | 35.92        | 32.68        | 38.29        | 2.44         | 25.42 | 31.45        | <b>39.27</b> | 35.12        | 39.16        | 33.72        |
| <b>de-pl</b> | 18.82        | 12.78        | 11.30        | -26.67       | 7.08  | 9.58         | 5.67         | <b>28.52</b> | 26.64        | 23.06        |
| <b>en</b>    | 59.11        | 60.66        | 61.15        | 54.96        | 54.23 | 60.16        | 62.46        | <b>63.23</b> | <b>63.79</b> | <b>63.28</b> |
| <b>es</b>    | 49.23        | 52.14        | 55.03        | 49.06        | 39.98 | 55.16        | 54.96        | <b>58.14</b> | <b>58.14</b> | <b>58.43</b> |
| <b>es-en</b> | 30.44        | 37.84        | 36.83        | 38.53        | 21.28 | 34.14        | 38.50        | <b>39.70</b> | 37.45        | <b>39.57</b> |
| <b>es-it</b> | 31.48        | 42.50        | 40.91        | 44.44        | 22.54 | 31.27        | 42.16        | 43.34        | 44.28        | <b>44.58</b> |
| <b>fr</b>    | 61.55        | 61.31        | 60.06        | 52.95        | 31.47 | 52.96        | 65.35        | <b>66.84</b> | <b>66.36</b> | <b>66.51</b> |
| <b>fr-pl</b> | 39.44        | <b>50.71</b> | -5.63        | 16.90        | 16.90 | 16.90        | 39.44        | 39.44        | 39.44        | 39.44        |
| <b>it</b>    | 54.67        | 59.89        | 57.61        | 52.94        | 27.64 | 53.46        | 60.60        | <b>63.98</b> | <b>63.31</b> | <b>62.98</b> |
| <b>pl</b>    | 22.79        | 26.72        | 23.77        | 8.23         | 6.78  | 23.42        | 26.09        | 25.72        | 25.79        | <b>27.13</b> |
| <b>pl-en</b> | 15.44        | <b>36.41</b> | 30.43        | 29.48        | 28.67 | 22.82        | 33.62        | 35.19        | 30.77        | 31.79        |
| <b>ru</b>    | 15.71        | 17.87        | 24.03        | 6.77         | 14.03 | <b>24.59</b> | 18.89        | 20.37        | 20.09        | 20.22        |
| <b>tr</b>    | 28.09        | 31.56        | 29.18        | 24.27        | 16.92 | 28.33        | 28.61        | 34.01        | <b>34.42</b> | <b>32.61</b> |
| <b>zh</b>    | 46.42        | 37.76        | <b>48.78</b> | 47.06        | 40.12 | 40.45        | 46.38        | 43.93        | 41.83        | 41.65        |
| <b>zh-en</b> | 4.82         | 9.87         | 13.14        | <b>27.61</b> | 15.06 | 11.94        | 8.60         | 18.22        | 17.94        | 23.58        |
| <b>Avg.</b>  | 32.78        | 36.75        | 34.33        | 26.48        | 23.61 | 32.67        | 37.01        | <b>39.34</b> | <b>38.95</b> | <b>39.26</b> |

Table IV-2: The performance of the TNCSE and baseline models on 18 multilingual/cross-language semantic similarity tasks on the STS22 test set.

| Tasks   | SimCSE       | ESimCSE      | DiffCSE      | InfoCSE      | SNCSE        | WhitenedCSE | RankCSE      | TNCSE        | +D           | +UCD         |
|---|--------------|--------------|--------------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|
| Banking77Classification                                 | 74.43        | 73.87        | 76.09        | <b>78.17</b> | 65.88        | 75.59       | 75.69        | 76.10        | 75.98        | 77.22        |
| BulgarianStoreReviewSentimentClassification             | 29.29        | 30.44        | 31.92        | 32.58        | <b>34.51</b> | 31.54       | 31.54        | 30.00        | 29.56        | 30.22        |
| CSFDZMovieReviewSentimentClassification                 | 20.32        | 20.28        | 20.91        | 20.92        | 20.11        | 21.21       | 20.71        | 21.03        | 21.07        | <b>21.30</b> |
| CUADAffiliateLicenseLicensorLegalBenchClassification    | 78.41        | 75.00        | <b>85.23</b> | 76.14        | 70.45        | 81.82       | 72.73        | 70.45        | 68.18        | 72.73        |
| CUADChangeOfControlLegalBenchClassification             | <b>71.15</b> | 63.70        | 69.71        | 67.79        | 71.15        | 67.07       | 66.11        | 69.95        | 68.99        | 68.03        |
| CUADInsuranceLegalBenchClassification                   | <b>94.17</b> | 88.64        | 91.36        | 91.84        | 89.51        | 86.02       | 90.78        | 89.42        | 90.00        | 92.62        |
| CUADMotFavoredNationLegalBenchClassification            | 78.13        | <b>84.38</b> | 75.00        | 75.00        | 64.06        | 76.56       | 71.88        | 79.69        | 78.13        | 76.56        |
| CUADThirdPartyBeneficiaryLegalBenchClassification       | 77.94        | 83.82        | 77.94        | <b>85.29</b> | 73.53        | 82.35       | 80.88        | 77.94        | 80.88        | 79.41        |
| Diversity2LegalBenchClassification                      | 74.00        | 73.67        | 74.33        | 74.67        | <b>76.00</b> | 74.67       | 74.67        | 74.33        | 74.33        | 74.67        |
| Diversity5LegalBenchClassification                      | 60.00        | 61.00        | 60.67        | 61.67        | <b>88.33</b> | 59.67       | 63.00        | 62.00        | 61.33        | 61.00        |
| FunctionOfDecisionsSectionLegalBenchClassification      | 16.35        | 15.26        | 15.80        | 15.53        | <b>23.16</b> | 16.35       | 14.71        | 17.44        | 17.44        | 14.44        |
| GermanPoliticiansTwitterSentimentClassification         | 38.21        | 36.67        | 37.98        | 39.41        | <b>40.14</b> | 38.68       | 38.74        | 38.01        | 37.45        | 37.65        |
| IndonesianIdClickbaitClassification                     | 54.26        | 53.89        | 54.09        | 54.56        | <b>57.57</b> | 54.15       | 53.44        | 54.68        | 55.02        | 55.36        |
| Itacola   | 48.56        | 49.02        | 47.28        | <b>50.62</b> | 50.07        | 47.82       | 48.13        | 48.87        | 48.94        | 49.64        |
| JCrewBlockerLegalBenchClassification                    | 77.78        | 74.07        | 72.22        | 57.41        | 87.04        | 72.22       | 75.93        | 66.67        | 72.22        | 68.52        |
| LearnedHandsDivorceLegalBenchClassification             | 76.00        | 79.33        | 75.33        | 69.33        | 64.67        | 80.67       | 83.33        | <b>84.00</b> | <b>84.00</b> | 83.33        |
| LearnedHandsDomesticViolenceLegalBenchClassification    | 78.16        | 77.59        | <b>78.74</b> | 70.69        | 72.41        | 75.86       | 75.29        | 78.16        | 77.01        | 76.44        |
| LearnedHandsFamilyLegalBenchClassification              | 70.75        | 77.10        | 68.65        | 71.48        | 64.99        | 68.85       | 71.53        | <b>79.20</b> | <b>80.66</b> | <b>78.52</b> |
| LearnedHandsHealthLegalBenchClassification              | 57.52        | <b>66.81</b> | 60.18        | 62.83        | 59.29        | 62.83       | 66.81        | 64.16        | 63.72        | 65.04        |
| LegalReasoningCausalityLegalBenchClassification         | 65.45        | 61.82        | 58.18        | 60.00        | <b>74.55</b> | 63.64       | 67.27        | 63.64        | 69.09        | 69.09        |
| MacedonianTweetSentimentClassification                  | 35.72        | 36.09        | 37.44        | 36.74        | <b>37.94</b> | 37.16       | 36.50        | 36.89        | 36.72        | 37.38        |
| OPPI15UserAccessEditAndDeletionLegalBenchClassification | 62.99        | 64.72        | 66.23        | 72.51        | 60.61        | 67.53       | 67.53        | 64.72        | 64.72        | 67.97        |
| OralArgumentQuestionPurposeLegalBenchClassification     | 22.44        | 15.71        | 21.47        | 19.87        | 24.04        | 23.08       | 21.79        | <b>25.64</b> | <b>25.00</b> | <b>25.00</b> |
| PersianFoodSentimentClassification                      | 55.98        | 55.76        | 56.77        | 58.01        | 53.55        | 57.43       | 55.95        | 57.25        | 57.18        | <b>57.60</b> |
| RestaurantReviewSentimentClassification                 | 50.59        | 51.39        | 50.29        | 50.72        | 53.03        | 51.88       | 50.10        | <b>52.04</b> | 51.60        | 51.77        |
| SCDDCertificationLegalBenchClassification               | 67.46        | 70.63        | 59.26        | <b>73.81</b> | 66.14        | 57.94       | 58.20        | 63.49        | 64.29        | 65.34        |
| SCDDTrainingLegalBenchClassification                    | 58.31        | 59.10        | 50.66        | <b>63.32</b> | 46.44        | 48.28       | 60.42        | 53.03        | 53.56        | 54.09        |
| ToxicChatClassification                                 | 69.47        | 68.50        | 66.31        | 71.01        | 69.78        | 64.60       | 68.21        | 71.32        | 68.57        | <b>71.41</b> |
| TweetEmotionClassification                              | 26.48        | 27.20        | 28.53        | 28.30        | <b>29.37</b> | 28.59       | 26.07        | 27.10        | 27.12        | 27.54        |
| TweetSentimentExtractionClassification                  | 54.27        | 52.36        | 54.54        | 54.23        | 53.25        | 53.91       | <b>56.53</b> | 54.30        | 54.95        | 55.31        |
| Avg.  | 58.15        | 58.26        | 57.44        | 58.15        | 58.05        | 57.60       | 58.15        | <b>58.38</b> | <b>58.59</b> | <b>58.84</b> |

Table IV-3: We report TNCSE and baselines performance on 30 classification tasks.

| Tasks                           | SimCSE       | ESimCSE      | DiffCSE      | InfoCSE      | SNCSE        | WhitenedCSE | RankCSE     | TNCSE        | +D           | +UCD         |
|---------------------------------|--------------|--------------|--------------|--------------|--------------|-------------|-------------|--------------|--------------|--------------|
| AILAStatistics                  | 8.17         | <b>9.78</b>  | 7.03         | <b>8.83</b>  | 7.43         | 6.25        | 6.58        | 9.36         | 8.84         | 8.55         |
| AlloprofRetrieval               | 3.68         | 6.27         | 2.71         | 2.74         | <b>2.67</b>  | 1.56        | 3.57        | 6.13         | <b>6.46</b>  | 5.26         |
| ArguAna-PL                      | 5.47         | 6.65         | 6.76         | 8.05         | 6.58         | 6.25        | 4.76        | 7.39         | <b>7.30</b>  | <b>8.15</b>  |
| CQADupstackEnglishRetrieval     | 16.78        | 16.16        | <b>16.61</b> | <b>22.06</b> | 13.27        | 16.68       | 15.49       | 15.25        | 16.31        | 19.64        |
| CQADupstackGamingRetrieval      | <b>23.83</b> | 23.51        | 25.41        | <b>31.13</b> | 20.73        | 25.18       | 22.34       | 24.92        | 25.25        | 29.20        |
| CQADupstackMathematicaRetrieval | <b>4.39</b>  | 5.33         | 5.26         | 6.89         | 4.41         | 4.88        | 5.14        | 6.08         | 5.98         | <b>7.72</b>  |
| CQADupstackTexRetrieval         | 4.42         | <b>6.40</b>  | 6.21         | 8.44         | 5.77         | 6.38        | 5.48        | 6.57         | 6.72         | <b>8.55</b>  |
| CQADupstackWebmastersRetrieval  | 13.32        | 13.41        | 14.16        | <b>16.00</b> | 13.49        | 14.89       | 12.20       | 14.31        | 15.11        | <b>17.33</b> |
| EstQA                           | 21.14        | 31.41        | 25.73        | 28.58        | <b>17.63</b> | 22.02       | 19.55       | 31.17        | <b>32.43</b> | 31.70        |
| FQuADRetrieval                  | 27.77        | 25.71        | 24.55        | <b>31.10</b> | <b>19.41</b> | 17.30       | 28.04       | 27.91        | 28.58        | 30.05        |
| GeorgianFAQRetrieval            | 2.88         | 2.59         | 3.45         | 2.85         | <b>3.03</b>  | 3.16        | 2.35        | 3.08         | 3.40         | <b>3.66</b>  |
| GerDaLIRSmall                   | 1.65         | 1.93         | 1.60         | 1.72         | <b>1.60</b>  | 1.38        | 1.32        | 1.94         | 1.95         | <b>2.12</b>  |
| GermanGovServiceRetrieval       | 19.73        | 28.04        | 19.45        | <b>36.45</b> | <b>12.49</b> | 15.37       | 18.23       | 28.50        | 27.15        | 29.20        |
| GermanQuAD-Retrieval            | 26.60        | 37.19        | 22.51        | <b>22.42</b> | 24.00        | 11.63       | 29.97       | <b>38.45</b> | 38.21        | 36.68        |
| JaQuADRetrieval                 | 3.00         | 3.16         | 2.72         | 4.01         | 3.21         | 3.27        | 3.89        | <b>4.99</b>  | 4.53         | 4.40         |
| LegalBenchCorporateLobbying     | 77.23        | <b>78.34</b> | 77.50        | 78.02        | 74.84        | 74.41       | 69.61       | <b>77.18</b> | 77.10        | <b>78.18</b> |
| LegalSummarization              | 45.57        | 45.80        | <b>44.45</b> | 46.20        | 43.43        | 43.29       | 44.25       | 48.62        | 48.30        | <b>50.96</b> |
| LEMBNarrativeQARetrieval        | 8.63         | 8.96         | 8.36         | 8.98         | 7.29         | 8.20        | 9.93        | <b>11.90</b> | <b>11.13</b> | <b>11.35</b> |
| LEMBQMSumRetrieval              | 9.64         | <b>11.54</b> | 10.10        | 10.82        | 9.47         | 9.09        | 10.84       | 12.82        | 12.47        | <b>13.12</b> |
| LEMBSummScreenFDRetrieval       | 39.88        | 45.34        | 41.75        | 38.93        | <b>28.58</b> | 41.46       | 40.40       | <b>52.69</b> | 51.54        | 52.36        |
| LEMBWikimQARetrieval            | 27.48        | 26.72        | 25.67        | 31.47        | <b>27.86</b> | 25.77       | 25.63       | 29.38        | 28.50        | <b>31.61</b> |
| MedicalQARetrieval              | 20.25        | 18.06        | 17.35        | <b>27.73</b> | 17.66        | 22.09       | 19.20       | 21.47        | 21.81        | 21.30        |
| NFCorpus                        | 3.66         | 3.55         | 3.43         | <b>4.44</b>  | 2.51         | 3.81        | 2.10        | <b>3.97</b>  | <b>3.70</b>  | <b>3.67</b>  |
| NQ                              | 9.79         | 8.53         | 9.50         | <b>15.84</b> | 9.04         | 9.41        | 8.51        | 8.92         | <b>10.58</b> | 10.93        |
| QuoraRetrieval                  | 75.58        | 74.23        | 75.66        | 77.57        | 70.06        | 74.54       | 76.41       | <b>77.08</b> | 76.76        | <b>78.21</b> |
| SCIDOCS                         | 4.22         | 4.26         | 4.45         | <b>5.06</b>  | 3.97         | 4.17        | 3.01        | 4.37         | 4.55         | 4.68         |
| SciFact                         | 32.07        | 34.37        | 33.05        | <b>34.83</b> | 27.34        | 32.22       | 23.13       | 30.65        | 29.47        | 33.51        |
| SlovakSumRetrieval              | 5.51         | 24.52        | 10.18        | 8.71         | 7.00         | 5.68        | 12.13       | 29.96        | <b>30.40</b> | 27.86        |
| SyntecRetrieval                 | 26.56        | 31.72        | 26.42        | 25.72        | <b>21.15</b> | 18.91       | 25.49       | 27.52        | 29.19        | <b>34.37</b> |
| TRECCOVID                       | 0.64         | 0.56         | <b>0.65</b>  | 0.63         | 0.56         | 0.61        | <b>0.48</b> | 0.53         | 0.51         | 0.54         |
| Avg.                            | 18.98        | 21.13        | 19.09        | 21.54        | 16.88        | 17.66       | 18.33       | <b>22.10</b> | <b>22.14</b> | <b>23.16</b> |

Table IV-4: We report TNCSE and baselines performance on 30 retrieval tasks.

| Model                     | SimCSE | ESimCSE      | DiffCSE | InfoCSE      | SNCSE | WhitenedCSE | RankCSE      | TNCSE        | TNCSE D      | TNCSE UC D   |
|---------------------------|--------|--------------|---------|--------------|-------|-------------|--------------|--------------|--------------|--------------|
| AlloprofReranking         | 30.46  | 32.74        | 28.33   | 27.97        | 25.31 | 27.34       | 29.25        | <b>33.63</b> | 31.76        | <b>33.94</b> |
| AskUbuntuDupQuestions     | 51.88  | 52.28        | 52.08   | 52.83        | 45.53 | 51.60       | <b>53.76</b> | 52.44        | 53.65        | 52.56        |
| CMedQAv1                  | 13.07  | 13.63        | 14.05   | <b>17.23</b> | 11.01 | 14.61       | 14.04        | 14.42        | 15.61        | 14.25        |
| CMedQAv2-reranking        | 13.97  | 14.78        | 15.26   | <b>17.21</b> | 11.69 | 15.06       | 14.47        | 14.42        | 16.01        | 14.68        |
| MMarcoReranking           | 2.48   | 3.77         | 3.64    | <b>4.96</b>  | 2.70  | 4.02        | 3.34         | 3.54         | 4.32         | 3.51         |
| SciDocsRR                 | 67.87  | 70.48        | 70.37   | 71.29        | 58.90 | 67.63       | 69.89        | 69.87        | <b>71.47</b> | 69.74        |
| StackOverflowDupQuestions | 39.57  | 40.64        | 42.77   | <b>44.21</b> | 31.06 | 42.64       | 41.18        | 42.05        | 43.25        | 41.94        |
| SyntecReranking           | 56.53  | <b>58.92</b> | 51.37   | 52.87        | 49.25 | 53.98       | 52.43        | 56.55        | 57.93        | 57.40        |
| T2Reranking               | 55.20  | 55.87        | 56.27   | <b>56.71</b> | 52.10 | 56.16       | 55.59        | 56.44        | 56.41        | 56.61        |
| Avg.                      | 36.78  | 38.12        | 37.13   | 38.36        | 31.95 | 37.01       | 37.11        | 38.15        | <b>38.93</b> | 38.29        |

Table IV-5: We report TNCSE and baselines performance on nine reranking tasks.