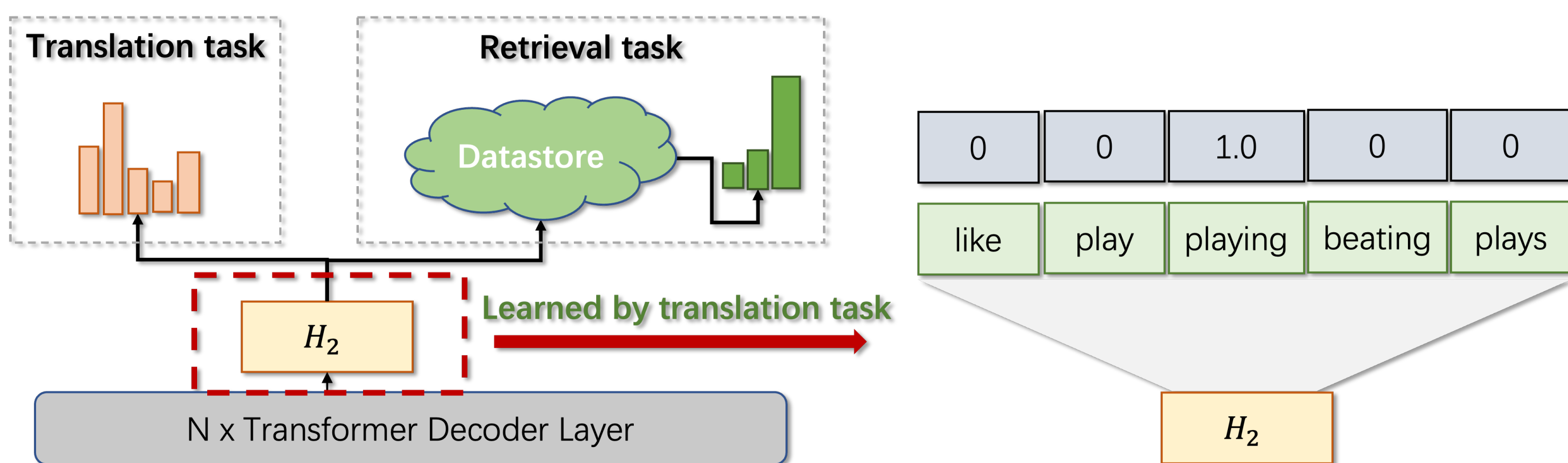


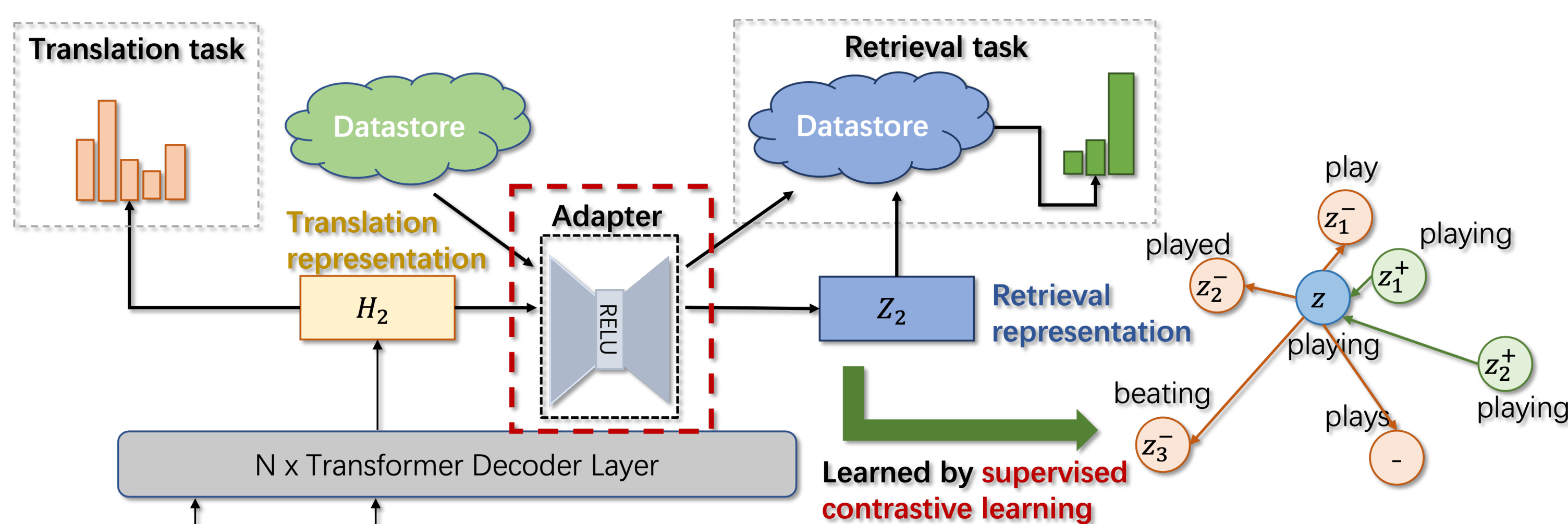
1. Motivation



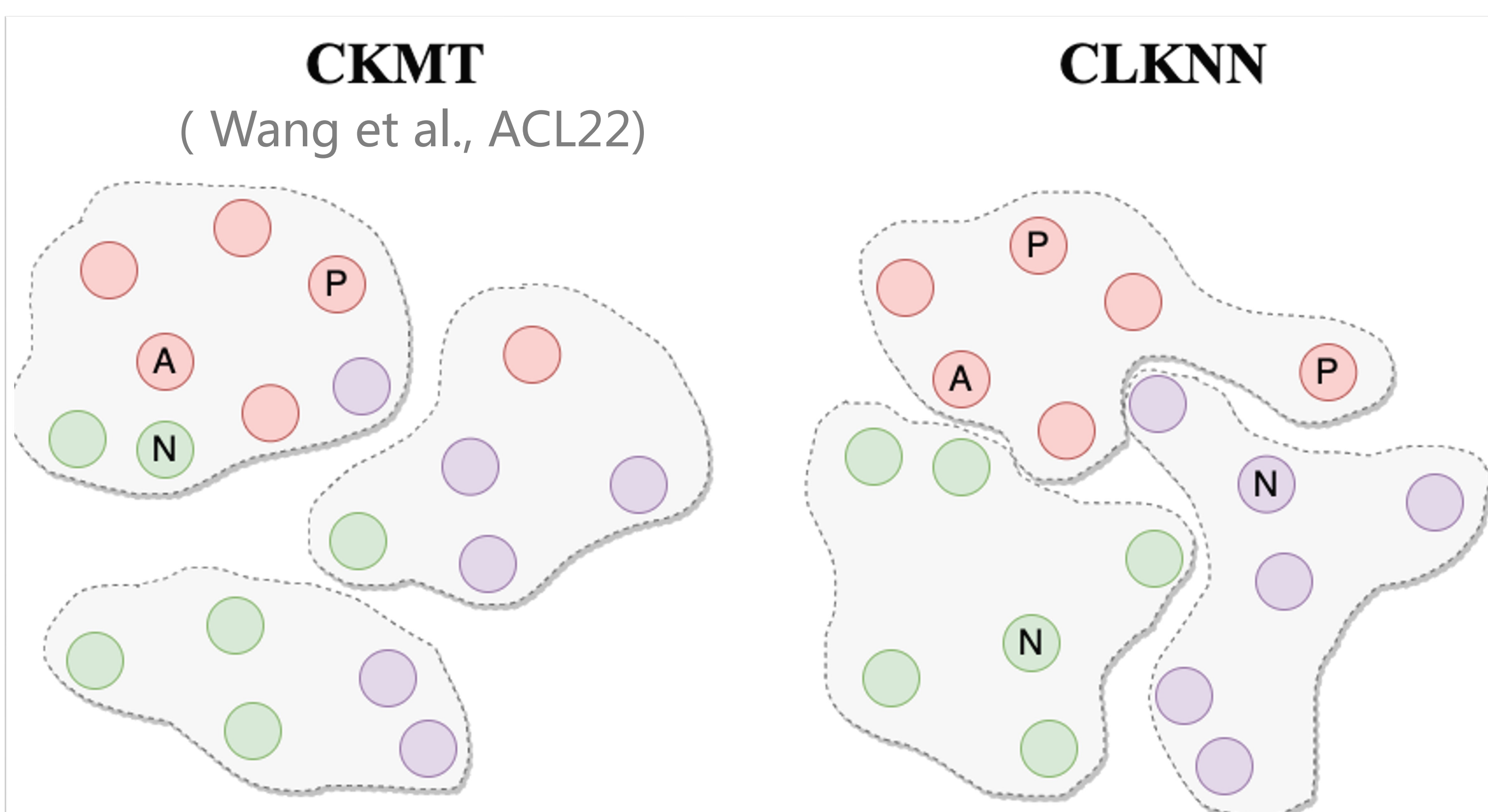
- kNN-MT couples the context representation in both translation and retrieval tasks
- Sub-optimal for retrieval

2. Approach

- Add a simple feedforward network as the adapter to transform translation context to retrieve context
- Learn adapter via supervised contrastive learning with multiple positive and negative samples



- Fast construction of hard negative samples



- | | |
|----------------------------|-----------------------------|
| - Explicitly do clustering | - Natural token label |
| - Spend 30 minutes | - Spend 3 minutes |
| - Single negative sample | - Multiple negative samples |

3. Experiment

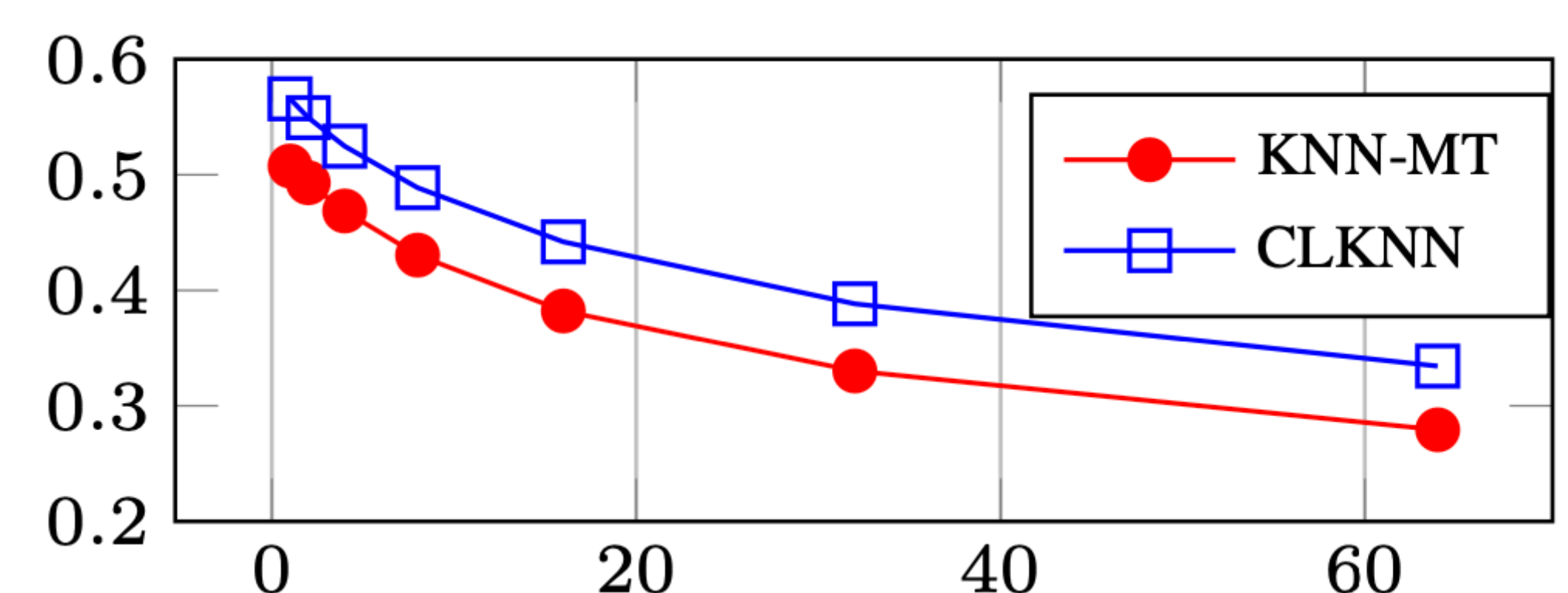
- SacreBLEU scores [%] on five in-domain German-English tasks

Method	Medical	Law	IT	Koran	Subtitle	Avg.
Baseline (WMT19 winner, Ng et al. (2019))	39.91	45.71	37.98	16.3	29.21	33.82
kNN-MT (Khandelwal et al., 2021)	54.35	61.78	45.82	19.45	31.73[†]	42.63
kNN-MT (our implementation)	54.41	61.01	45.20	21.07	29.67	42.27
<i>train by out-domain data</i>						
CLKNN	56.37	61.54	46.50	21.52	30.81	43.35
CLKNN + λ^*	56.52	61.63	46.68	21.60	30.86	43.46
<i>train by in-domain data</i>						
CLKNN	55.86	61.92	47.77	21.46	31.02	43.61
CLKNN + λ^*	55.87	62.01	47.84	21.81	31.05	43.72

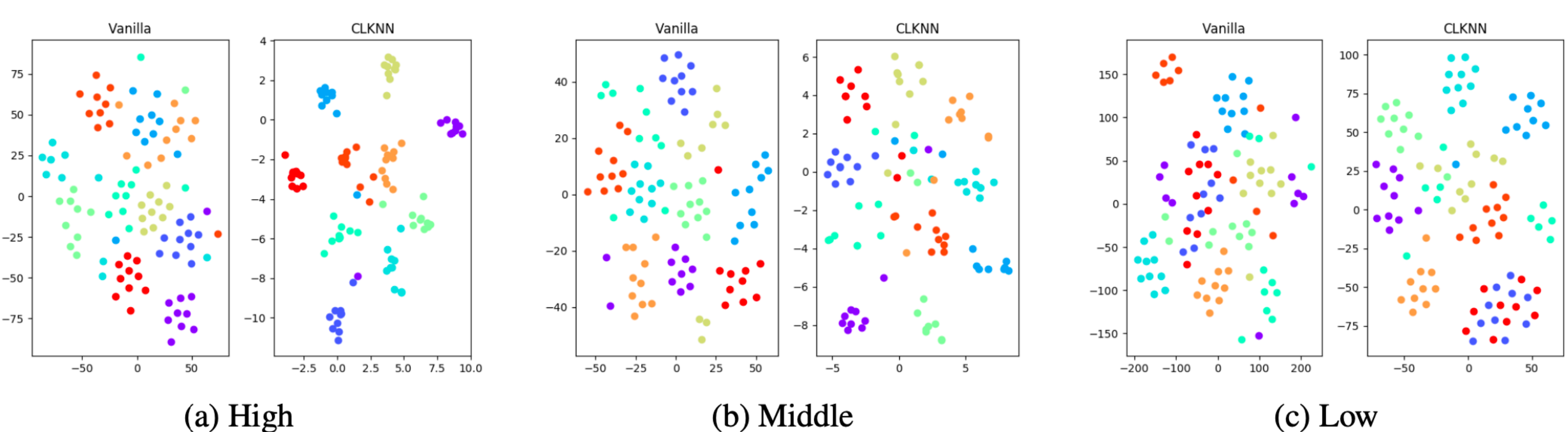
- Effects of multiple positive/negative samples

M	N	BLEU	M	N	BLEU
1	1	45.54	2	16	46.37
1	16	45.91	2	32	46.68
1	32	46.13	2	64	46.55
1	64	45.88	4	32	46.29

- Quantitative analysis on retrieval representation



- Qualitative analysis on retrieval representation



4. Conclusion

- We propose CLKNN to decouple retrieve representation by supervised contrastive learning with multiple positive and negative samples
- CLKNN uses a simple and effective method to construct hard negative samples
- Experimental results show that CLKNN improves more than 1 BLEU point than vanilla kNN-MT due to better retrieve representation