

# WENYE LIN

✉ lwy20@mails.tsinghua.edu.cn · ☎ (+86) 17708209207 · 🌐 <https://github.com/Linwenye>

## 🎓 EDUCATION

---

**Tsinghua University**, Beijing, China Aug. 2020 – Jul. 2023

*Third Year Graduate student* in Computer Science

**Nanjing University**, Nanjing, China Sept. 2015 – Jun. 2019

*B.Eng.* in Software Engineering

## 🧑‍🔬 RESEARCH

---

Alleviating Over-clustering in Lightweight Contrastive Models Apr. 2022 – Aug. 2022

**Wenye Lin, Yifeng Ding, Zhixiong Cao, Hai-tao Zheng ICLR 2023 submitted**

- Small contrastive models tend to suffer from the over-clustering problem. We propose Soft-InfoNCE, a novel loss to alleviate this problem.
- We dramatically boost the performance of lightweight contrastive models on downstream tasks. Compared with baselines, Soft-InfoNCE improves the top-1 accuracy from 42.2% to **59.3%** on EfficientNet-B0 and from 36.3% to **56.3%** on MobileNet-V3-Large on ImageNet.

Tree-Supervised Auxiliary Online Knowledge Distillation Feb. 2021 – May 2021

**Wenye Lin, Yangning Li, Yifeng Ding, Hai-tao Zheng IJCNN 2022 accepted**

- We design tree-structured auxiliary (TSA) online knowledge distillation to perform one-stage distillation when the teacher is unavailable.
- With TSA, we gain an average of **3%** to **4%** improvement in accuracy on CIFAR-100. On ImageNet, ResNet-34 obtains 74.97% accuracy, which is 1.8% above the vanilla one. On IWSLT translation tasks, we gain an average of 0.9 BLEU improvement over vanilla Transformer for three datasets.

## 🧑‍💼 EXPERIENCE

---

Research Intern: **Tencent Inc.** AI Lab, NLP Center Sept. 2021 – Aug. 2022

Manager: Lemao Liu

Graph Word Embedding May 2022 – Aug. 2022

**Wenye Lin, Lemao Liu, Haiyun Jiang, Hai-tao Zheng**

- Transformer-based models utilize word embeddings to represent the semantic inputs. Every word correspond to one embedding, and is updated independently. In our work, we combine GCN and Transformer to build embeddings where similar words are related.
- With fasttext, we significantly decrease the perplexity of Transformer-xl on wikitext-2 from 95.6 to **69.5**.

Efficient Structured Knowledge Distillation Oct. 2021 – Feb. 2022

**Wenye Lin, Yangming Li, Lemao Liu, Shuming Shi, Hai-tao Zheng ARR submitted**

- Performing knowledge distillation for structured prediction models is not trivial due to their exponentially large output space. We transfer the knowledge from a teacher model to its student model by locally matching their computations on all internal structures.
- It avoid adopting some time-consuming techniques like dynamic programming (DP) for decoding output structures, which permits parallel computation. Besides, it encourages the student model to better mimic the internal behavior of the teacher model.

## ⚙️ SKILLS

---

- Research Interest: *Representation Learning, Improving Generalization of Networks*
- Programming: Python, C++, Java; Pytorch