LINFENG ZHANG

(+86) 138-4030-1165 zhanglinfeng1997@outlook.com Google Scholar Homepage

EDUCATION & AWARDS

Tsinghua University (Ph.D. Student) 09/2019 - 06.2024 (Expectation)

Institute for Interdisciplinary Information Sciences, Supervised by Kaisheng Ma

Research Direction: Knowledge Distillation, Computer Vision

MSRA Fellowship, 2022 (Top-12 in Asia)

National Scholarship, 2020 (1%), Tang Lixin Scholarship (First Prize, 1%), 2021

Doctoral Forum Best Poster Award, 2020

Jiangnanxiang Scholarship, 2023 (Top-10 in Tsinghua University)

Northeastern University (Bachelor) 09/2015 - 07/2019

School of Computer Science And Engineering

Outstanding Graduates (<2%), Outstanding Graduation Dissertation (<2%), The First Prize Scholarship(<3%)

Outstanding Winner in Microsoft Student Club Practice Space, 2018

The First Prize Winner of ACM and MCM in NEU

ENACTUS China Regional Competition First Prize, 2018

ENACTUS China Regional Innovation Competition Gold Award, 2018

INTERNSHIP

DIDI Global, Cargo Autonomous Driving 04/2022 - 04.2023

Acceleration of Point Cloud Detection and Multi-view Detection: We have proposed two knowledge distillation methods for acceleration of **point cloud-based and multi-view images-based 3D detection**. Detectors trained with these methods have been **deployed on autonomous driving trucks** and tested in the streets of Beijing. The corresponding paper has been published in **CVPR2023**.

Kuaishou Technology, Y-tech Lab 04/2022 - 10.2022

Acceleration of GAN for Cartoon Style Transfer: Cartoon style transfer is an essential AI function in the **live stream and video editing** in Kuaishou APP. I have proposed a novel **knowledge distillation** method for GAN compression and successfully compressed GANs from **5G FLOPs to 200M FLOPs**. This model has been used in business and the corresponding paper has been published in **CVPR2022**.

PUBLICATION

Top Conference (Journal) & First Author × 10, Others × 8, Citation 800+

1. **Linfeng Zhang**, Jiebo Song, Anni Gao, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. Be Your Own Teacher: Improve the Performance of Convolutional Neural Networks via Self Distillation. IEEE International Conference on Computer Vision (ICCV2019).

2. **Linfeng Zhang**, Zhanhong Tan, Jiebo Song, Jingwei Chen, Chenglong Bao, and Kaisheng Ma. SCAN: A Scalable Neural Networks Framework Towards Compact and Efficient Models. Neural Information Processing Systems (NeurIPS2019).

3. **Linfeng Zhang**, Muzhou Yu, Tong Chen, Zuoqiang Shi, Chenglong Bao, and Kaisheng Ma. Auxiliary Training: Towards Accurate and Robust Models. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2020).

4. **Linfeng Zhang**, Yukang Shi, Zuoqiang Shi, Kaisheng Ma, and Chenglong Bao. Task-Oriented Feature Distillation. Neural Information Processing Systems (NeurIPS2020).

5. **Linfeng Zhang**, Chenglong Bao, and Kaisheng Ma. Self-Distillation: Towards Efficient and Compact Neural Networks. IEEE Transactions of Pattern Analysis and Machine Intelligence (IEEE TPAMI).

6. **Linfeng Zhang**, and Kaisheng Ma. Improve Object Detection with Feature-based Knowledge Distillation: Towards Accurate and Efficient Detectors. The International Conference on Learning Representations (ICLR2021).

7. **Linfeng Zhang**, Xin Chen, Xiaobing Tu, Pengfei Wan, Ning Xu, Kaisheng Ma. Wavelet Knowledge Distillation: Towards Efficient Image-to-Image Translation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2022).

8. **Linfeng Zhang**, Xin Chen, Junbo Zhang, Runpei Dong, Kaisheng Ma. Contrastive Deep Supervision. European Conference on Computer Vision (ECCV2022 **Oral Presentation**).

9. **Linfeng Zhang**, Runpei Dong, Huang-Shuo Tai Kaisheng Ma. Pointdistiller: structured knowledge distillation towards efficient and compact 3d detection. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2023).

10. **Linfeng Zhang**, Kaisheng Ma. A Good Data Augmentation Policy Is Not All You Need: A Multi-Task Learning Perspective" in IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT),

11. Yuni Lai*, **Linfeng Zhang*** **(co-first author)**, Donghong Han, Rui Zhou, Guoren Wang, Fine-grained emotion classification of Chinese microblogs based on graph convolution networks. World Wide Web Journal (WWW).

12. Runpei Dong, Zhanhong Tan, Mengdi Wu, **Linfeng Zhang**, Kaisheng Ma. Finding the Task-Optimal Low-Bit Sub-Distribution in Deep Neural Networks. International Conference on Machine Learning (ICML2022 **Spotlight**).

13. Runpei Dong, Zekun Qi, **Linfeng Zhang**, Junbo Zhang, Jianjian Sun, Zheng Ge, Li Yi, Kaisheng Ma. Autoencoders as Cross-Modal Teachers: Can Pretrained 2D Image Transformers Help 3D Representation Learning? International Conference on Learning Representation (ICLR2023)

14. Xiaolong Ma, Sheng Lin, Shaokai Ye, Zhezhi He, **Linfeng Zhang**, Geng Yuan, Sia Huat, Tan, Zhengang Li, Deliang Fan, Xuehai Qian, Xue Lin, Kaisheng Ma, Yanzhi Wang. Non-Structured DNN Weight Pruning – Is It Beneficial in Any Platform? IEEE Transactions on Neural Networks and Learning Systems (IEEE TNNLS).

15. Zimo Liao, Zhicheng Luo, Qianyi Huang, **Linfeng Zhang**, Fan Wu, Qian Zhang, Yi Wang. SMART: screen-based gesture recognition on commodity mobile devices. Annual International Conference on Mobile Computing and Networking (MobiCom21 **Oral Presentation**).

16. **Linfeng Zhang**, Xiaoman Zhang, Chenglong Bao, Kaisheng Ma. Wavelet J-Net: A Frequency Perceptive on Convolutional Neural Networks. International Joint Conference on Neural Networks (IJCNN2021).

17. Muzhou Yu, Sia Huat Tan, Kailu Wu, Runpei Dong, **Linfeng Zhang**, Kaisheng Ma, CORSD: Class-Oriented Relational Self Distillation. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2023).

PATENTS

1. 基于图卷积网络的社交网络文本情感细粒度分类方法（第一申请人，已授权，ZL 2019 1 0728633.7）
2. 一种通过辅助分类器在训练阶段提高卷积神经网络鲁棒性的方法 (第一申请人，已递交申请)
3. 一种基于频域小波基处理的神经网络图像分类处理方法 (第一申请人，已递交申请)
4. 一种通过关系对其提高卷积神经网络精确性和鲁棒性的方法 (第一申请人，已递交申请)
5. 一种卷积神经网络的自蒸馏训练方法 (第一申请人，已递交申请)
6. 基于知识蒸馏的图像目标检测方法和检测器及其训练方法 (第一申请人，已递交申请)
7. 任务牵引的特征蒸馏深度神经网络训练方法 (第一申请人，已递交申请)
8. 一种针对对抗生成网络的基于离散小波变换的知识蒸馏算法 (第一申请人，已递交申请)
9. 基于环境感知深度卷积神经网络的图像识别方法及系统（第一申请人，已递交申请）

PROJECTS

(1) Self-Distillation: Towards Efficient and Compact Models 2019 - 2020

Project Leader, First Author, **ICCV2019×1, NeurIPS2019×1, TPAMI×1**

· **Motivation:** Existing knowledge distillation methods need a pre-trained teacher model, which leads to unstable student performance and costs long training time.

· **Solution:** We propose the first teacher-free knowledge distillation method named self-distillation, which regards the deep layers of a model as the teacher and the shallow layers of a model as the student. Besides, self-distillation enables the student to work in a dynamic manner depending on the current input data. On CIFAR100, it achieves >3.25% accuracy improvements, 7.7 × compression and 3.1 × acceleration.

Self-distillation has already been used in Qiming AI chips, Huawei phones and Taobao recommendation system. It is considered as the pioneer and the most representative method in the domain of teacher-free knowledge distillation by many papers and surveys. Abundant researchers have proposed incremental work based on self-distillation in its theory and application on the other tasks, such as semantic segmentation and pre-trained language models.

(2) Auxiliary Training: Towards Accurate and Compact Models October 2020

Project Leader, First Author, **CVPR2020**

· **Motivation:** Deep neural networks are sensitive to the natural image corruption (e.g. noise, blurring). Naive data augmentation can alleviate this problem but leads to accuracy drop.

· **Solution:** We propose auxiliary training, which formulates data augmentation learning as a multi-task learning task, and solves it by adding several auxiliary classifiers for different types of augmentation. Besides, we introduce input-aware self distillation and selective batch normalization for further performance improvements. Our method first improves model accuracy and robustness at the same time.

(3) Task-Oriented Feature Distillation June 2021

Project Leader, Co-first Author, **NeurIPS2021**

- **Motivation:** We study the question: which kind of information of teachers are crucial for knowledge distillation and which kind of transformation function can extract these information.
- **Solution:** We assume that the task-oriented information are the most crucial information for knowledge distillation. Then, we apply several auxiliary classifiers to extract these task-oriented information and distill them from teachers to students. Besides, we also propose an orthogonal regularization method for better distillation performance. Experiments demonstrate the performance of our method on both RGB images and point cloud data.

(4) Improve Object Detection with Feature-based Knowledge Distillation October 2021
Project Leader, First Author, **ICLR2021**

- **Motivation:** Most of knowledge distillation methods are designed for classification and usually fail in the more challenging task such as object detection
- **Solution:** we design two novel knowledge distillation methods (attention-guided distillation and non-local distillation) for object detection, which show excellent improvements on both one-stage and two-stage detectors, both anchor-based and anchor-free method, both object detection and instance segmentation. For example, it leads to 4.1 AP improvements on Faster RCNN on COCO2017.

(5) Wavelet Knowledge Distillation Towards Efficient Image-to-Image Translation June 2022
Project Leader, First Author, **CVPR2022**

- **Motivation:** GAN-based image-to-image translation models usually have a tremendous amount of parameters and thus suffer from low efficiency and bulky memory usage.
- **Solution:** In this project, we have analyzed the performance of GANs from a frequency perspective, which quantitatively shows that GAN, especially small GAN lacks the ability to generate high-quality high frequency information in images. Based on the above observation, we propose wavelet knowledge distillation to address this issue by only distilling the high frequency information, instead of all the information from images generated by the teacher.
- Besides, we have studied the relation between discriminators and generators during model compression. It shows that compression on discriminators is necessary for maintaining its competition with compressed generators in adversarial learning, which further benefits the performance of generators.

(6) Fine-Grained Emotion Classification with Graph Convolution Networks October 2018
Project Leader, Co-first Author, **WWW Journal**

- We propose a novel syntax-based graph convolution network (GCN) model for emotion detection of Chinese microblogs. In addition, we collected and labeled 15,664 microblogs to construct a highquality Chinese microblogs dataset which is open to other researchers.

(7) Non-Structured DNN Weight Pruning–Is It Beneficial in Any Platform? October 2018
Project Member, **IEEE TNNLS**

- **Motivation:** Although unstructured weight pruning leads to significant sparsity in deep neural networks, however, they incurs index accesses due to irregular weights. In this project, we study the question that is non-structured DNN weight pruning really beneficial?
- **Solution:** We develop a methodology for fair and fundamental comparison of non-structured and structured pruning in terms of both storage and computation efficiency. Our experimental results show that non-

structured pruning is not competitive in terms of both storage and computation efficiency. Thus, we conclude that non-structured pruning is considered harmful.

(8) Wavelet J-Net: A Frequency Perspective on Convolutional Neural Networks May 2019
Project Leader, First Author, **IJCNN2021**

- **Motivation:** It is generally acknowledged in image processing domain that different frequencies of images convey different information. However, existing neural networks always ignore it and process all the frequency of images together.
- **Solution:** We propose a discrete wavelet transformation based deep neural network named J-Net, which first process the low frequency information, and then the high frequency information. We show that the proposed J-Net is not only more accurate but also more explainable and controllable.

(9) SMART: Screen-based Gesture Recognition on Commodity Mobile Devices Ocotber 2021
Project Member, **MobiCom21**

- We propose SMART, an in-air gesture recognition system leveraging the screen and ambient light sensor (ALS), which are ordinary modalities on mobile devices. For the transmitter side, we design a screen display mechanism to embed spatial information and preserve the viewing experience; for the receiver side, we develop a framework to recognize gestures from low-quality ALS readings. We implement and evaluate SMART on both a tablet and several smartphones. Results show that SMART can recognize 9 types of frequently used in-air gestures with an average accuracy of 96.1%.

(10) Finding Task-Optimal Low-Bit Sub-Distribution in Deep Neural Networks Ocotber 2021
Project Member, **ICML2022**

- **Motivation:** Quantized neural networks typically require smaller memory footprints and lower computation complexity, which is crucial for efficient deployment. However, quantization inevitably leads to a distribution divergence from the original network, which generally degrades the performance.
- **Solution:** we present an adaptive mapping quantization method to learn an optimal latent sub-distribution that is inherent within models and smoothly approximated with a concrete Gaussian Mixture (GM). In particular, the network weights are projected in compliance with the GM-approximated sub-distribution. This sub-distribution evolves along with the weight update in a co-tuning schema guided by the direct task-objective optimization.

(11) Tiny Updater: Towards Efficient Neural Network Software Updating October 2021
Project Leader, First Author, NeurIPS2022 Under Review

- **Motivation:** When neural networks based softwares are updated, software users have to re-download all the parameters of neural networks, which severely harms the user experience.
- **Solution:** By adopting pruning and knowledge distillation methods, Tiny Updater can update the neural network-based software by only downloading a few parameters (10%$\sim$20%) instead of all the parameters in the neural network. Besides, we also propose a new knowledge distillation method designed for video recognition, which aims to distill the frame-wise relation from teachers to students.

(12) Region-aware Knowledge Distillation for Efficient Image Translation October 2021
Project Leader, First Author, ECCV2022 Under Review

- **Motivation:** GANs usually contain a huge number of parameters, which lead to intolerant memory and computation consumption and limit their deployment on edge devices.
- **Solution:** we propose Region-aware Knowledge Distillation (ReKo) to compress image-to-image translation models. Firstly, ReKo adaptively finds the crucial regions in the images with an attention module. Then, patch-wise contrastive learning is adopted to maximize the mutual information between students and teachers in these crucial regions. Our 7.08× compressed and 6.80× accelerated CycleGAN student outperforms its teacher by 1.33 and 1.04 FID scores on Horse→Zebra and Zebra→Horse, respectively

(13) Contrast Deep Supervision January 2021

Project Leader, First Author, ECCV2022 Oral Presentation

- **Motivation:** The traditional training method only supervises the neural network at its last layer and propagates the supervision layer-by-layer, which leads to hardship in optimizing the intermediate layers. Deep supervision is proposed to add auxiliary classifiers to the intermediate layers of deep neural networks. By optimizing these auxiliary classifiers with the supervised task loss, the supervision can be applied to the shallow layers directly. However, deep supervision conflicts with the well-known observation that the shallow layers learn low-level features instead of task-biased high-level semantic features.
- **Solution:** We propose a novel training framework named Contrastive Deep Supervision, which supervises the intermediate layers with augmentation-based contrastive learning. Experimental results on nine popular datasets with eleven models demonstrate its effects on general image classification, fine-grained image classification and object detection in supervised learning, semi-supervised learning and knowledge distillation.

(14) A Good Data Augmentation Policy Is Not All You Need March 2022

Project Leader, First Author, IEEE TCSVT

- **Motivation:** Usually, the design of augmentation policies faces a diversity-difficulty trade-off. Automatic augmentation methods have been proposed to address this issue by searching the optimal data augmentation at the cost of expensive computation overhead.
- **Solution:** We propose to break the diversity-difficulty trade-off from a multi-task learning perspective. By formulating model learning on the augmented images and the original images as the auxiliary task and the primary task in multi-task learning, respectively, the hard augmentation does not directly influence the training of the primary branch, and thus its negative influence can be alleviated. Hence, neural networks can learn valuable semantic information even with a totally random augmentation policy.

(15) Point Distiller: Knowledge Distillation for Efficient 3D Objectors May 2022

Project Leader, First Author, NeurIPS2022 Under Review

- **Motivation:** Compared with image based 2D detectors, point clouds based 3D detectors usually require more computation, which has limited their deployment on edge devices.
- **Solution:** Based on the properties of point clouds, we propose local distillation to extract the local geometric structure information of point clouds, and then distill it from the teacher to the student. Besides, we introduce a reweighted learning strategy, which aims to only distill the points or voxels which make a more crucial influence for the prediction. Experimental results show that our method achieves 4.00× compression of PointPillars, SECOND, and PointRCNN detectors on KITTI without mAP drop.

(16) Refinements based Knowledge Distillation for Image-to-Image Translation May 2022

Project Leader, First Author, NeurIPS2022 Under Review

- **Motivation:** Due to the ill-posed property of image-to-image translation, directly applying knowledge distillation to image-to-image translation may make the student to learn the average between multiple correct answers, which harms student performance.
- **Solution:** To address this problem, we propose to replace the teacher in knowledge distillation with a student-conditioned refined network, which can provides not only better but also consistent learning targets for the student. During the training period, the refined network is trained to refine the outputs of the student to improve their quality. And the refined results are further utilized as the learning targets for the students.

(17) Revisiting Data Augmentation in Model Compression: A Comprehensive Study May 2022
Project Leader, First Author, NeurIPS2022 Under Review

- **Motivation:** Previous research usually studies in model compression from the persepctives of architecture (*e.g.* pruning and AutoML) and training method (*e.g.* knowledge distillation). In this paper, we propose to study the influence of data augmentation in model compression.
- **Solution:** We give an empirical and comprehensive analysis to the data augmentation in data augmentation. Our results show that: (1) Models in different sizes prefer data augmentation with different magnitudes. Hence, data augmentation with decayed magnitudes should be utilized in pruning. (2) Small models can not directly learn with strong data augmentation, but then can still benefit from strong data augmentation by learning them with additional parameters. (3) Big pre-trained models can be utilized to filter the hard data augmentations for the small models.

(18) Autoencoders as Cross-Modal Teachers ICLR2023

We revisit masked modeling in a unified fashion of knowledge distillation, and we show that foundational Transformers pretrained with 2D images or natural languages can help self-supervised 3D representation learning through training Autoencoders as Cross-Modal Teachers (ACT). The pretrained Transformers are transferred as cross-modal 3D teachers using discrete variational autoencoding self-supervision, during which the Transformers are frozen with prompt tuning for better knowledge inheritance.

(19) CORSD: Class-Oriented Relational Self Distillation ICASSP2023

We propose a novel training framework named Class-Oriented Relational Self Distillation which captures inter-class and intra-class relations of structured input, and the transferring of the relational knowledge from the the deepest layer to shallow layers.

PROFESSIONAL SERVICE

- Review papers for ICCV, CVPR, NeurIPS, ICLR, TIP, TMM, ECCV, ICML, AAAI, IJCAI, etc.
- Establish the Machine Learning Club in Northeastern University and work as its president in 2018.
- Work as the publicity minister in the postgraduate students Union at IIIS Tsinghua University.