Ming Lin
Ph.D.
Senior Applied Scientist
linming04@gmail.com

Amazon.com LLC
Bellevue
WA 98004
https://minglin-home.github.io/

Research Interest

- Mathematical foundation of deep learning
- Statistical machine learning
- Deep neural network design, optimization and acceleration
- Applications in Computer Vision / Neural Architecture Search / Mobile AI / Recommendation Systems

Skills & Capabilities

- 14 Years of research and industrial experiences in machine learning, deep learning and statistics.
- Programming
    - Python, C/C++, SQL, Java, Bash
- Deep Learning Frameworks
    - Pytorch, Tensorflow, MXNet

Research Highlight

Mathematical Neural Network Design **2021 — 2022**

- A mathematical framework for designing efficient deep learning models.
    - Design high performance deep neural networks with nearly zero-cost.
    - Theoretical understanding of deep learning
- Publication
    - **DeepMAD: Mathematical Architecture Design for Deep Convolutional Neural Network.** CVPR 2023.
    - **Entropy-Driven Mixed-Precision Quantization for Deep Network Design on IoT Devices.** NIPS 2022.
    - **MAE-DET: Revisiting Maximal Entropy Principle in Zero-Shot NAS for Efficient Object Detection**. ICML 2022.
    - **Zen-NAS: A Zero-Shot NAS for High-Performance Deep Image Recognition**. ICCV 2021.
- Invited talk
    - Radu Marculescu, Ming Lin, Atlas Wang, Kartikeya Bhardwaj. **Training-Free Approaches for Edge AI: Challenges, Opportunities and Progress**. In MLSys 2022 Tutorial. [website]
    - VALSE Talk [video] (Chinese)
- Hugging-face Demo [link]

Efficient Attention Modules **2021 — 2022**

- Innovate new attention-based building blocks for deep learning.
- Publication
  - **Making Vision Transformers Efficient from A Token Sparsification View.** CVPR 2023.
  - **Robust Graph Structure Learning over Images via Multiple Statistical Tests**. NIPS 2022.
  - **KVT: k-NN Attention for Boosting Vision Transformers**. ECCV 2022.

Projects

Tiny-NAS **2020 — 2023**

- Homepage [https://github.com/alibaba/lightweight-neural-architecture-search]
- Tiny-NAS is a **full-stack framework for designing deep networks**. It is able to design efficient image recognition models and object detection models in 1~2 hours, with state-of-the-art performance on ImageNet, MS COCO, etc.
  - Able to optimize multiple inference metrics including **hardware latency.**
  - Training with adaptive quantization and knowledge distillation.
- **Application in Product**
  - Alibaba Pailitao (拍立淘), 10 millions active users per day. (Similar to Google Lens)
    - Shopping With Your Camera: Visual Image Search Meets E-Commerce at Alibaba [https://medium.com/coinmonks/shopping-with-your-camera-visual-image-search-meets-e-commerce-at-alibaba-8551925746d0].
  - DAMO-YOLO: a fast and accurate object detection method with some new techs, including NAS backbones, efficient RepGFPN, ZeroHead, AlignedOTA, and distillation enhancement. [GitHub]
  - Dingding M1X FaceID Check-in System, 5+ million devices.
    - Product Homepage [https://tms.dingtalk.com/markets/dingtalk/dingding_M1X_pc] .
- **Application in NAS Competition**
  - Used in the 3rd-ranked solution PGONAS in CVPR 2022 Lightweight NAS Challenge. [arXiv]
  - Used in the winning solution of WebFace260M Track of ICCV21-MFR, 1.2% improvement over manually designed networks.
  - Used in the winning solution of LPIRC-2019 Competition.

GiraffeDet: High-Performance Object Detection Backbone **2020 — 2021**

- Github [https://github.com/damo-cv/GiraffeDet].
- The GiraffeDet uses an extremely lightweight backbone and a very deep and large neck module which encourages dense information exchange among different spatial scales as well as different levels of latent semantics simultaneously. This design paradigm allows GiraffeDet achieves SOTA performances in object detection tasks with low resource budget.
- **Application in Product**

- Satellite Remote Sensing Segmentation, 50+ TB high resolution dataset.

Deep Image Compression **2019 — 2022**

- Develop deep learning models for better image compression
- Save 21% bit-rate over BPG
- Publication
  - Entroformer: A Transformer-based Entropy Model for Learned Image Compression. ICLR 2022.
  - Learning Accurate Entropy Model with Global Reference for Image Compression. ICLR 2021.
- **Application in Product**
  - Compress images for online storage of DingDing.
  - DingDing [https://www.dingtalk.com/en] is the largest enterprise instance message app in China.

Real-time High-Precision GPS Grid System **2018 — 2019**

- Improve GPS precision to 1 centimeter in real-time (1 query per second).
- Positioning successful rate 99.5 % over-all in good weather, 80%-90% for single query.

Early Clinical Diagnosis of Alzheimer's Disease and Major Depressive Disorder **2015 — 2018**

- Use large-scale data and machine learning techniques to assist clinical diagnosis of Alzheimer's Disease.
- **Able to predict Alzheimer's Disease development in four years.**
- Cooperation project with Janssen Research & Development, LCC Michigan Alzheimer's Disease Center.

Automated Low-Level Analysis and Description of Diverse Intelligence Video (ALADDIN) **2014 — 2015**

- Develop a large-scale content-based video retrieval system.
- **Ranked top-1 in 6 out of 8 tasks in MED 2014.**
- Sponsored by Informedia@CMU.

Work Experience

2022 July — Senior Applied Scientist — Last Mile of Amazon.com, Inc.

- Tech lead of an internal start-up team to innovate an AI system to assist last 100 yards delivery
- Cross-team collaboration with Hardware Computing, Amazon DSP and Geospatial

		Science.
2018 — 2022	Staff Algorithm Engineer	DAMO Academy of Alibaba Group (U.S.) • Leading a research group of 5 research engineers and several research interns • Efficient deep architecture design • Neural Architecture Search • Drive 15 top conference papers
2015 — 2018	Research Investigator	Department of Computational Medicine and Bioinformatics University of Michigan Medical School Ann Arbor, Michigan 48109, United States. • Early diagnosis of Major Depressive Disorder and Alzheimer's disease • Theoretical guarantees for the second order linear models • 2 top-tier conference papers in statistical machine learning • Cooperation with Janssen Research & Development, LCC and Michigan Alzheimer's Disease Center • Mentor: Jieping Ye
2014 — 2015	Postdoctoral Research Fellow	School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213, United States. • Adviser: Alexander G. Hauptmann. • Major developer of the ALADDIN project • Video retrieval and feature fusion • 3 top conference papers with 600+ citation in total
2013 — 2014	Visiting Student	School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213, United States. • Adviser: Alexander G. Hauptmann.
2012 — 2013	Visiting Student	Department of Computer Science and Engineering Michigan State University East Lansing, MI 48824, United States. • Adviser: Rong Jin.

Education

2008 — 2014	Ph.D.	Department of Automation Tsinghua University Beijing 100084, China. • Adviser: Changshui Zhang
2004 — 2008	Bachelor	Department of Automation Tsinghua University Beijing 100084, China.

Peer-Review Service

- Reviewer of AAAI, IJCAI, ICML, NIPS, ICLR, CVPR, ICCV etc.
- Reviewer of Journals:
  - IEEE Transactions on Pattern Analysis and Machine Intelligence
  - ACM Transactions on Knowledge Discovery from Data
  - Pattern Recognition
  - Neurocomputing
  - Computer Vision and Image Understanding.
  - Data Mining and Knowledge Discovery

Impact

Google citation 1256, h-index 17, i10-index 27

Bibliography

Peer-Reviewed Conferences

1. Xuan Shen, Yaohua Wang, Ming Lin, Yilun Huang, Hao Tang, Xiuyu Sun, Yanzhi Wang. DeepMAD: Mathematical Architecture Design for Deep Convolutional Neural Network. In Proceedings of the IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2023.
2. Shuning Chang, Pichao Wang, Ming Lin, Fan Wang, David Junhao Zhang, Rong Jin, Mike Zheng Shou. Making Vision Transformers Efficient from A Token Sparsification View. In Proceedings of the IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2023.
3. Junyan Wang, Zhenhong Sun, Yichen Qian, Dong Gong, Xiuyu Sun, Ming Lin, Maurice Pagnucco, Yang Song. Maximizing Spatio-Temporal Entropy of Deep 3D CNNs for Efficient Video Recognition. In Proceedings of the Eleventh International Conference on Learning Representations (ICLR), 2023.

4. Yaohua Wang, Fangyi Zhang, Ming Lin, Senzhang Wang, Xiuyu Sun, Rong Jin. Robust Graph Structure Learning over Images via Multiple Statistical Tests. In Proceedings of the Conference on Neural Information Processing Systems (NIPS), 2022.
5. Zhenhong Sun, Ce Ge, Junyan Wang, Ming Lin, Hesen Chen, Hao Li, Xiuyu Sun. Entropy-Driven Mixed-Precision Quantization for Deep Network Design on IoT Devices. In Proceedings of the Conference on Neural Information Processing Systems (NIPS), 2022.
6. Pichao Wang, Xue Wang, Fan Wang, Ming Lin, Shuning Chang, Hao Li, Rong Jin. KVT: k-NN Attention for Boosting Vision Transformers. In Proceedings of the European Conference on Computer Vision (ECCV), 2022
7. Zhenhong Sun, Ming Lin, Zhiyu Tan, Xiuyu Sun, Hao Li, Rong Jin. MAE-DET: Revisiting Maximal Entropy Principle in Zero-Shot NAS for Efficient Object Detection. In Proceedings of the International Conference on Machine Learning (ICML), 2022.
8. Yichen Qian, Xiuyu Sun, Ming Lin, Zhiyu Tan, Rong Jin. Entroformer: A Transformer-based Entropy Model for Learned Image Compression. In Proceedings of the International Conference on Learning Representations (ICLR), 2022.
9. Yaohua Wang, Yaobin Zhang, Fangyi Zhang, Senzhang Wang, Ming Lin, YuQi Zhang, Xiuyu Sun. Ada-NETS: Face Clustering via Adaptive Neighbour Discovery in the Structure Space. In Proceedings of the International Conference on Learning Representations (ICLR), 2022.
10. Yiqi Jiang, Zhiyu Tan, Junyan Wang, Xiuyu Sun, Ming Lin, Hao Li. GiraffeDet: A Heavy-Neck Paradigm for Object Detection. In Proceedings of the International Conference on Learning Representations (ICLR), 2022.
11. Ming Lin, Pichao Wang, Zhenhong Sun, Hesen Chen, Xiuyu Sun, Qi Qian, Hao Li, Rong Jin. Zen-NAS: A Zero-Shot NAS for High-Performance Deep Image Recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
12. Yichen Qian, Zhiyu Tan, Xiuyu Sun, Ming Lin, Dongyang Li, Zhenhong Sun, Li Hao, Rong Jin. Learning Accurate Entropy Model with Global Reference for Image Compression. In Proceedings of the International Conference on Learning Representations (ICLR), 2021.
13. Hesen Chen, Ming Lin, Xiuyu Sun, Qian Qi, Hao Li, Rong Jin. MuffNet: Multi-Layer Feature Federation for Mobile Deep Learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV workshop), 2019.
14. Ming Lin, Xiaomin Song, Qi Qian, Hao Li, Liang Sun, Shenghuo Zhu, Rong Jin. Robust Gaussian Process Regression for Real-Time High Precision GPS Signal Enhancement. In Proceedings of the 25TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (SIGKDD), 2019.
15. Ming Lin, Shuang Qiu, Jieping Ye, Xiaomin Song, Qi Qian, Liang Sun, Shenghuo Zhu, Rong Jin. Which Factorization Machine Modeling is Better: A Theoretical Answer with Optimal Guarantee. The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI), 2019.
16. Tieliang Gong, Guangtao Wang, Jieping Ye, Zongben Xu, Ming Lin. Margin Based PU Learning. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI), 2018.
17. Xiang Li, Aoxiao Zhong, Ming Lin, Ning Guo, Mu Sun, Arkadiusz Sitek, Jieping Ye, James Thrall, Quanzheng Li. Self-paced Convolutional Neural Network for Computer Aided Detection in Medical Imaging Analysis. In Proceedings of the 8th International Workshop on Machine Learning in Medical Imaging (MLMI), 2017.
18. Zhenzhong Lan, Shoou-I Yu, Dezhong Yao, Ming Lin, Bhiksha Raj ; Alexander Hauptmann. The Best of BothWorlds: Combining Data-Independent and Data-Driven Approaches for Action

Recognition. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Pages 1196-1205, 2016.

19. Ming Lin, Jieping Ye. A Non-convex One-Pass Framework for Generalized Factorization Machine and Rank-One Matrix Sensing. In Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS), Pages 1633-1641, 2016.
20. Chuang Gang, Ming Lin, Yi Yang, Gerard de Melo, Alexander G. Hauptmann. Concepts Not Alone: Exploring Pairwise Relationships for Zero-Shot Video Activity Recognition. In Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI), Pages 3487-3493, 2016.
21. Chuang Gang, Ming Lin, Yi Yang, Alexander G. Hauptmann. Exploring Semantic Inter-Class Relationships (SIR) for Zero-Shot Action Recognition. In Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI), Pages 3769-3775, 2015
22. Zhenzhong Lan, Ming Lin, Xuanchong Li, Alexander G. Hauptmann, Bhiksha Raj. Beyond Gaussian Pyramid: Multi-skip Feature Stacking for Action Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pages 204-212, 2015.
23. Ming Lin, Zhenzhong Lan, Alexander G. Hauptmann. Density Corrected Sparse Recovery when R.I.P. Condition is Broken. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Pages 3664-3670, 2015.
24. Shoou-I Yu, Lu Jiang, Zexi Mao, Xiaojun Chang, Xingzhong Du, Chuang Gan, Zhenzhong Lan, Zhongwen Xu, Xuanchong Li, Yang Cai, Anurag Kumar, Yajie Miao, Lara Martin, Nikolas Wolfe, Shicheng Xu, Huan Li, Ming Lin, Zhigang Ma, Yi Yang, Deyu Meng, Shiguang Shan, Pinar Duygulu Sahin, Susanne Burger, Florian Metze, Rita Singh, Bhiksha Raj, Teruko Mitamura, Richard Stern, Alexander Hauptmann. Informedia@ TRECVID 2014 MED and MER. NIST TRECVID Video Retrieval Evaluation Workshop, 2014.
25. Ming Lin, Rong Jin, Changshui Zhang. Efficient Sparse Recovery via Adaptive Non-Convex Regularizers with Oracle Property. In Proceedings of the 30th Conference on Uncertainty in Artificial Intelligenre (UAI), Pages 505-514, 2014.
26. Lijun Zhang, Jinfeng Yi, Ming Lin, Xiaofei He. Online Kernel Learning with a Near Optimal Sparsity Bound. In Proceedings of the 30th International Conference on Machine Learning (ICML), Pages 621 – 629, 2013.

Peer-Reviewed Journals

1. Jian Liang, Kun Chen, Ming Lin, Changshui Zhang, Fei Wang. Robust Finite Mixture Regression for Heterogeneous Targets. Data Mining and Knowledge Discovery, Volume 32, Issue 6, pp 1509–1560, November 2018.
2. Ming Lin, Pinghua Gong, Tao Yang, Jieping Ye, Roger L. Albin, Hiroko H. Dodge. Big Data Analytical Approaches to the NACC Dataset: Aiding Preclinical Trial Enrichment, Volume 32, Issue1, Pages 18-27, January-March 2018.
3. Daqing Chang, Ming Lin, Changshui Zhang. On the generalization ability of online gradient descent algorithm under the quadratic growth condition. To appear in TNNLS 2018.
4. Ming Lin, Vaibhav Narayan, Wayne C. Drevets, Jieping Ye, Qingqin Li. Application of Growth Mixture Modeling in Antidepressant Treatment Response Studies. Biological Psychiatry, Volume 81, Issue 10, Supplement, Page S224, May 2017.

5. Xiaojun Chang, Zhigang Ma, Ming Lin, Yi Yang, Alexander G. Hauptmann. Feature Interaction Augmented Sparse Learning for Fast Kinect Motion Detection. IEEE Transactions on Image Processing, Volume 26, Issue 8, Pages 3911-3920. 2017.
6. Shoou-I Yu, Yi Yang, Zhongwen Xu, Shicheng Xu, Deyu Meng, Zexi Mao, Zhigang Ma, Ming Lin, Xuanchong Li, Huan Li, Zhenzhong Lan, Lu Jiang, Alexander G. Hauptmann, Chuang Gan, Xingzhong Du, Xiaojun Chang. Strategies for Searching Video Content with Text Queries or Video Examples (Invited Paper). ITE Transactions on Media Technology and Applications 4.3, Pages 227-238, 2016
7. Ming Lin, Lijun Zhang, Rong Jin, Shifeng Weng, Changshui Zhang. Online Kernel Learning with Nearly Constant Support Vectors. Neurocomputing. Volume 179, Pages 26-36, 2016.
8. Zhen Hu, Ming Lin, Changshui Zhang. Dependent Online Kernel Learning with Constant Number of Random Fourier Features. IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Volume 26, Issue 10, Pages 2464-2476, 2015.
9. Zheng Pan, Ming Lin, Guangdong Hou, Changshui Zhang. Damping proximal coordinate descent algorithm for non-convex regularization. Neurocomputing, vol 152, Pages 151-163, 2015.
10. Ming Lin, Fei Wang, Changshui Zhang. Large-Scale Eigenvector Approximation via Hilbert Space Embedding Nystrom. Pattern Recognition (PR), 48(5), Pages 1904-1912, 2015.
11. Ming Lin, Shifeng Weng, Changshui Zhang. On the Sample Complexity of Random Fourier Features for Online Learning. ACM Transactions on Knowledge Discovery from Data (TKDD), Volume 8 Issue 3, Pages 13:1--13:19, June 2014.
12. Shizhun Yang, Ming Lin, Chenping Hou, Changshui Zhang, Yi Wu. A General Framework for Transfer Sparse Subspace Learning. Neural Computing and Applications. Volume 21, Number 7, Pages 1801-1817, August 2012.