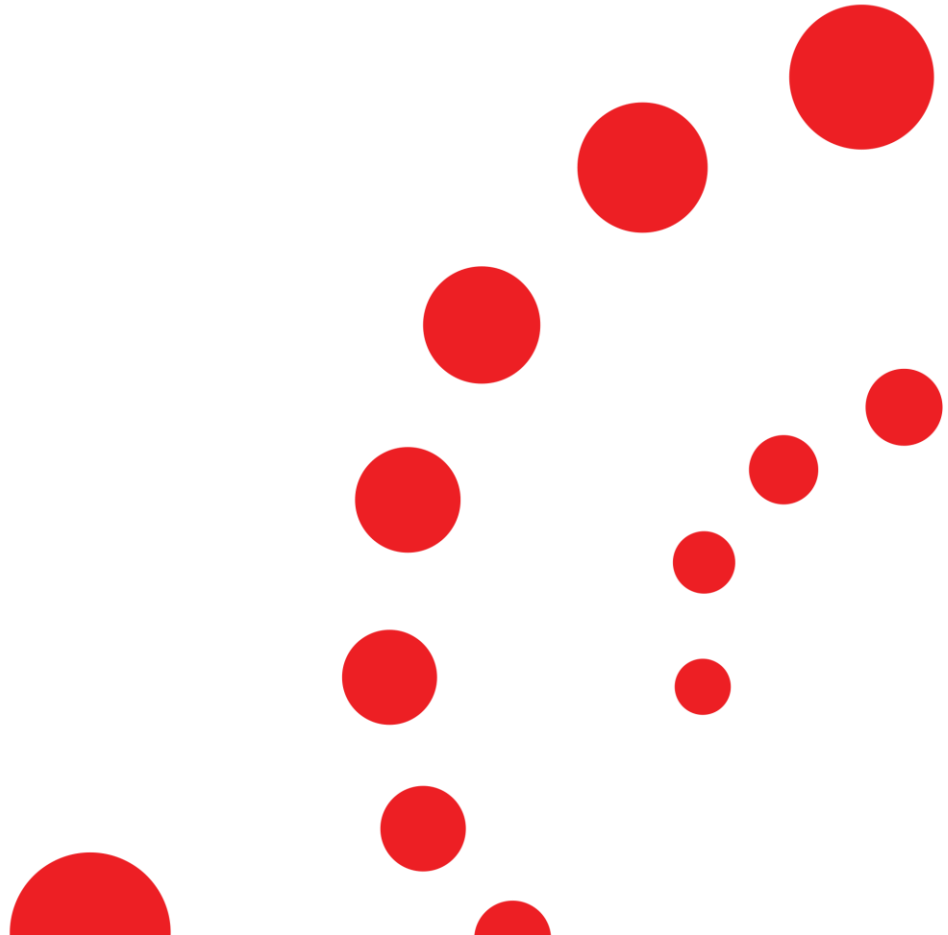
ICML 2022 Report

11/11/2022

OpenMindSpore Project
Silicon Valley System Software Lab



Contents

- Overview
- Awards
- Selected papers of interests
- Comments

38th International Conference on Machine Learning (ICML)

- Date: July 17-23, 2022
- Location: Baltimore, MD, USA

ICML 2022 Sponsors

- Diamond sponsors
  - Google Research, G Research, AstraZeneca, Meta
- Platinum sponsors
  - Apple, DeepMind, Amazon Science
- Gold sponsors
  - Microsoft, XTX Markets, Citadel
- Silver sponsors
  - Yokogawa, QuantumBlack, Absci, Two Sigma
  - Micron, Capital One, Morgan Stanley, Alegion
  - Point72, Invenia Labs, Hudson River Trading, Xyla
  - Squarepoint, Criteo, Netflix, Bosch
  - Huawei, Baidu, Bloomberg
- Bronze sponsors
  - Edgestream, Sea AI Lab, Intel, Boltzbit
  - Mohamed Bin Zayed Univ. of Artificial Intelligence, D.E. Shaw & Co, Roblox, Qualcomm

ICML 2022 Awards

- Test of time award (1)
  - Poisoning Attacks Against Support Vector Machines
- Test of time honorable mention (2)
  - Building high-level features using large scale unsupervised learning
  - On causal and anticausal learning
- Outstanding paper (10)
  - Understanding Dataset Difficulty with *V*-Usable Information
  - Learning Mixtures of Linear Dynamical Systems
  - Privacy for Free: How does Dataset Condensation Help Privacy?
  - Solving Stackelberg Prediction Game with Least Squares Loss via Spherically Constrained Least Squares Refomulation
  - Bayesian Model Selection, the Marginal Likelihood, and Generalization
  - G-Mixup: Graph Data Augmentation for Graph Classification
  - Stable Conformal Prediction Sets
  - Do Differentiable Simulators Give Better Policy Gradients?
  - Causal Conceptions of Fairness and their Consequences
  - The Importance of Non-Markovianity in Maximum State Entropy Exploration
- Outstanding paper runner up (5)
  - Monarch: Expressive Structured Matrices for Efficient and Accurate Training
  - Adversarially Trained Actor Critic for Offline Reinforcement Learning
  - Minimum Cost Intervention Design for Causal Effect Identification
  - Active fairness auditing
  - Learning inverse folding from millions of predicted structures

Invited Talks

- Towards a Mathematical Theory of Machine Learning
- Solving the Right Problems: Making ML Models Relevant to Healthcare and the Life Sciences
- Synthetic Control Methods and Difference-In-Differences
- Design for Inference in Drug Discovery and Development

Tutorials

- Causality and Deep Learning: Synergies, Challenges and the Future
- Quantitative Reasoning About Data Privacy in Machine Learning
- Validity, Reliability, and Significance: A Tutorial on Statistical Methods for Reproducible Machine Learning
- Bridging Learning and Decision Making
- Learning for Interactive Agents
- Climate Change and Machine Learning: Opportunities, Challenges, and Considerations
- Sampling as First-Order Optimization over a space of probability measures
- Welcome to the “Big Model” Era: Techniques and Systems to Train and Serve Bigger Models (UC Berkeley)
  - Hao Zhang, Lianmin Zheng, Zhuohan Li, Ion Stoica
  - https://alpa.ai/icml22-tutorial.html

Sessions of Main Conference

- July 19
  - Probabilistic Methods/Applications
  - Deep Learning: Robustness
  - Optimization: Convex
  - Theory: Online Learning/Bandits
  - Theory
  - Reinforcement Learning: Deep/Batch/Offline
  - Optimization
  - Theory: Bandits/RL/Everything Else
  - SA: Trustworthy Machine Learning
  - T: Learning/Deep Learning Theory
  - PM: Monte Carlo and Sampling Methods
  - Theory

Sessions of Main Conference

- July 20
  - PM: Variational Inference/Bayesian Models and Methods
  - OPT: First Order
  - T: Game Theory/RL/Planning
  - PM: Bayesian Models and Methods
  - SA: Trustworthy Machine Learning
  - T: Online Learning and Bandits/Learning Theory

Sessions of Main Conference

- July 21
  - T: Bandits/Online Learning/Reinforcement Learning
  - Reinforcement Learning
  - Theory/Social Aspects
  - Theory: Game Theory and Optimization
  - Reinforcement Learning
  - Optimization/Reinforcement Learning
  - Reinforcement Learning
  - Probabilistic Methods/MISC
  - Reinforcement Learning/Optimization
  - Social Aspects/Optimization

Selected Papers of Interests

Flashlight: Enabling Innovation in Tools for Machine Learning (Meta AI)

- Open-source minimalist ML library designed to support research in machine learning frameworks
  - Modular component-based customizable architecture
  - Compact and high-performant reference implementation
  - Comprehensive set of benchmarks representative of SOTA in machine learning
  - Available at https://github.com/flashlight/flashlight

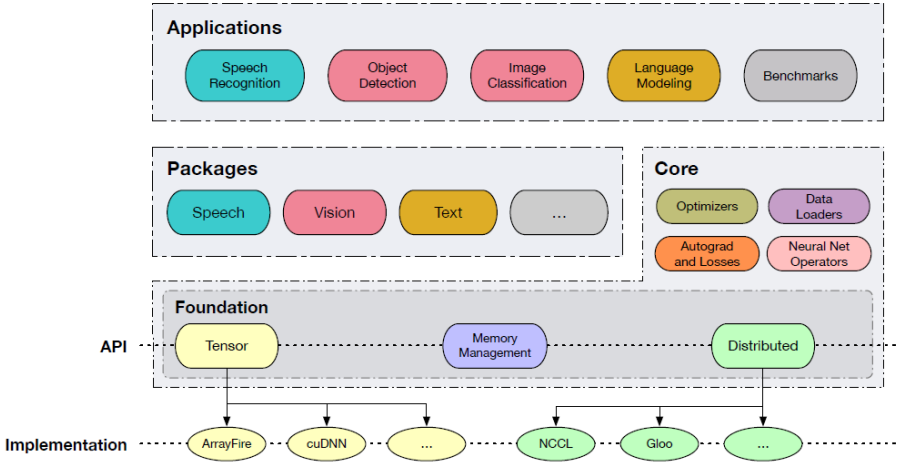


Figure 1. Components of the Flashlight library.

Metric	PyTorch	TensorFlow	(Ours) Flashlight
Binary Size (MB)	527	768	10
Lines of Code	1,798,292	1,306,159	27,173
Number of Operators	2,166	1,423	60
Approx num. ops. that perform:			
add	55	20	1
conv	85	30	2
sum	25	10	1

FUTUREWEI Technologies

Monarch: Expressive Structured Matrices for Efficient and Accurate Training (Stanford, SUNY, UM)

Unlock new ways to train and fine-tune sparse and dense models:

- Hardware efficiency
- End-to-end training a sparse (Monarch) model can be 2x faster than dense training
- Sparse-to-dense “reverse sparsification” can speed up training of large models such as GPT-2;
- Dense-to-sparse Monarch projection algorithm can transfer knowledge from pretrained dense model to Monarch model and speed up BERT fine-tuning
- This technique can be used to ViT and other models
- Available at [2204.00595.pdf (arxiv.org)]

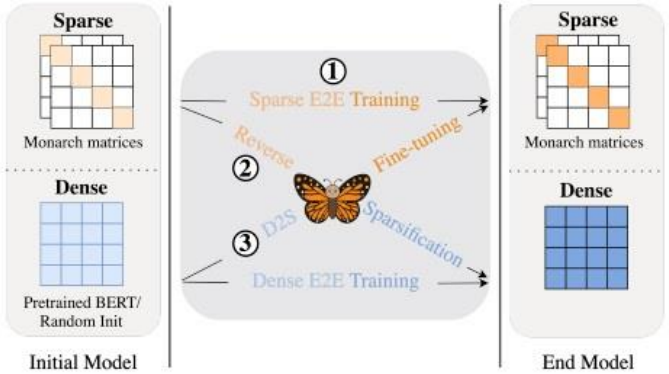


Table 8: The performance of Monarch matrices in finetuning BERT on GLUE.

Model	GLUE (avg)	Speedup	Params	FLOPs
BERT-base	78.6	-	109M	11.2G
Monarch-BERT-base	78.3	1.5×	55M	6.2G
BERT-large	80.4	-	335M	39.5G
Monarch-BERT-large	79.6	1.7×	144M	14.6G

A Study of Face Obfuscation in ImageNet (Princeton U and Stanford U)

- Even public dataset have privacy concern.
  - ImageNet dataset has 3 people categories (scuba diver, bridegroom, and baseball player) in 1000 categories.
  - However, the dataset exposes many people, e.g., 562,626 faces from 243,198 images (17% of all images have at least one faces).
  - More than 90% of categories have faces, even though they are not people categories.
- Contributions
  - Privacy-enhanced version of ILSVRC via blurring and overlaying of faces
  - https://github.com/princetonvisualai/imagenet-face-obfuscation
  - Validation accuracy drops only slightly (0.1%-0.7% for blurring, 0.3%-1.0% for overlaying)
  - Face-obfuscated pretraining data are equally transferable.

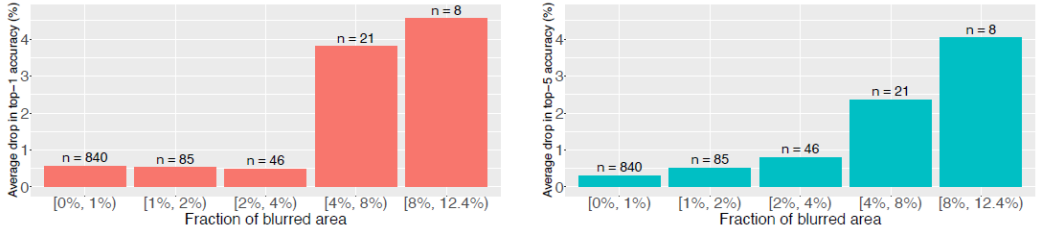


Figure 3. The average drop in category-wise accuracies vs. the fraction of blurred area in images. *Left*: Top-1 accuracies. *Right*: Top-5 accuracies. The accuracies are averaged across all different model architectures and random seeds.

FUTUREWEI Technologies

Privacy for Free: How does Dataset Condensation Help Privacy? (SJTU, U of Edinburgh, Sony AI, Outstanding paper)

- Background
  - Machine learning models suffers from privacy attacks
    - E.g., model inversion attack, membership inference attack (MIA), property inference attack
  - Previously solution using GAN (Generative Adversarial Networks) tried to overcome the issue by training with synthetic data, but privacy risks still exist.
  - Differential privacy (DP) has been used as a de facto privacy standard, but the generated low-quality data negatively impact accuracy of models trained on those data.
  - Instead, data condensation (DC) condenses a large training set into a small synthetic set which is comparable to the original one for DNNs.
- Contributions
  - First to (1) introduce DC techniques to privacy protection and (2) to build the connection between DC and differential privacy, and (3) to contribute theoretical analysis of this approach.
  - Experiments validate DC methods reduce adversary advantage of membership privacy, and DC-synthesized data are irreversible to original data (Figure 5 for example).

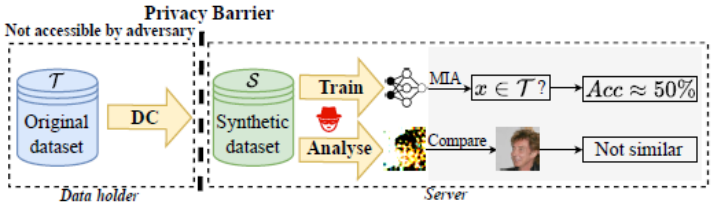


Figure 1. DC-synthesized data can be used for privacy-preserving model training and cannot be recovered through MIA and visual comparison analysis.

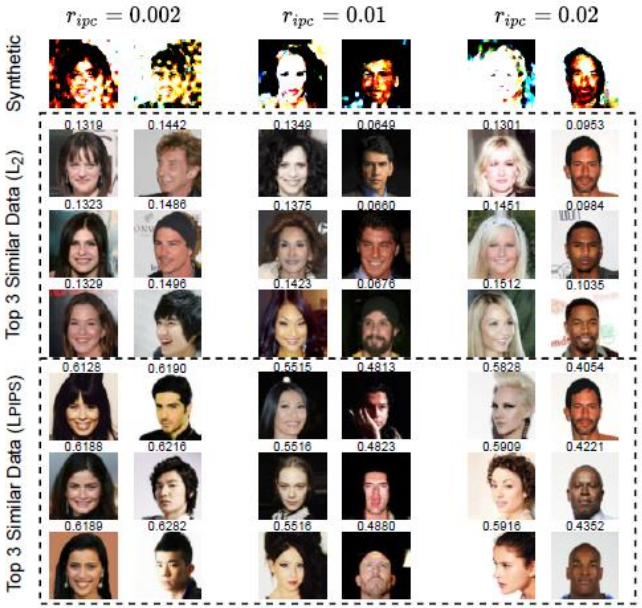
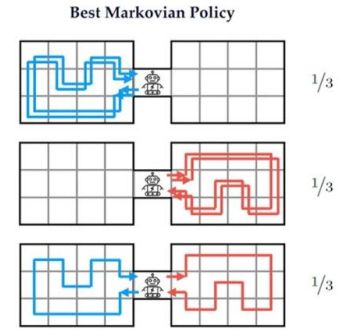


Figure 5. Examples of facial images that are most similar to synthetic data generated by DM with *random* initialization. The value

The Importance of Non-Markovianity in Maximum State Entropy Exploration (U of Bologna and ETH Zurich, Outstanding Paper)

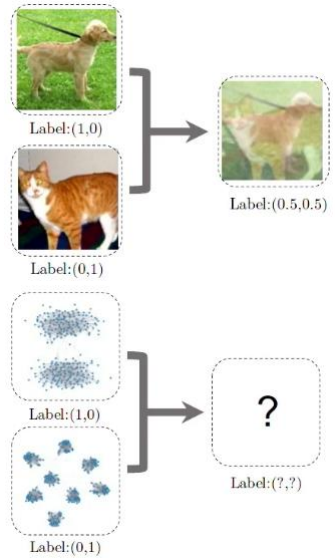
- Background
  - An agent interacts with a reward-free environment to learn a policy maximizing the entropy of the expected state visitations.
  - Previously it is known that the class of Markovian stochastic policies is sufficient for the maximum state entropy objective.
  - Exploiting non-Markovianity is considered pointless.
- Contributions
  - Non-Markovian policies are better for finite-sample convex objectives.
    - E.g., Agent exploring two rooms domain (in right figure).
    - Human takes intentional decisions in the middle of position to visit a room before the other.
    - Hidden infinite-samples assumption contrasts with optimization over a finite batch of interactions.
  - Optimizing non-Markovian policies exactly is often intractable.
    - Optimizing the finite-sample MSE within the space of non-Markovian policies is NP-hard.
  - Approximate methods to optimize non-Markovian policies for convex objectives.

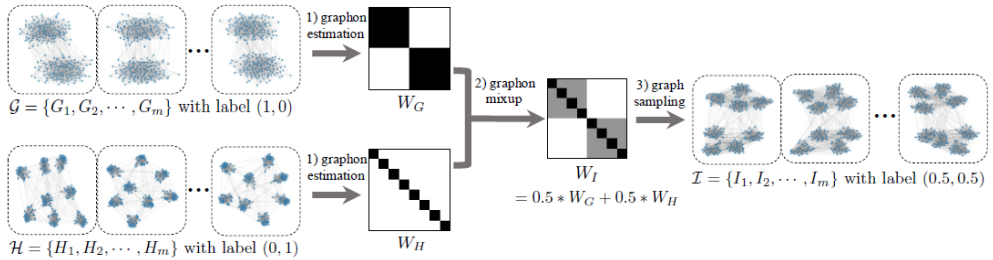


FUTUREWEI Technologies

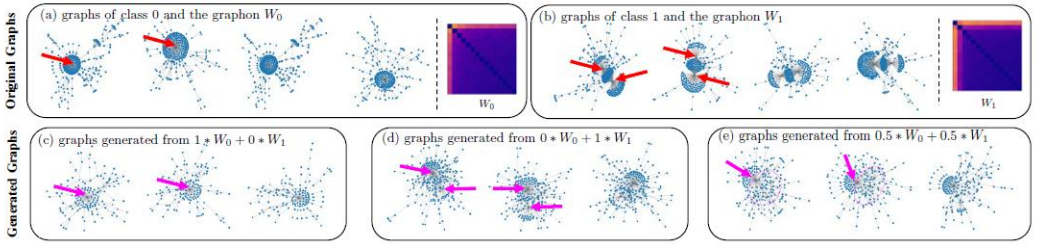
G-Mixup: Graph Data Augmentation for Graph Classification (Texas A&M, U of Georgia, Rice U, Outstanding Paper)

- Background
  - Mixup is a cross-instance data augmentation method, which linearly interpolates random sample pair to generate more synthetic training data.
  - Mixup have been empirically and theoretically shown to improve the generalization and robustness of DNNs.
  - Graph mixup is different and difficult.
    - Image data is regular, well-aligned, and grid-like data.
    - Graph data is irregular, not well-aligned, and has divergent topology information.
- G-Mixup
  - Mixing up graph generator (graphon) to achieve the input graph mixup.
    - Graph estimation, graph mixup, graph sampling
  - Real-world graphs of different classes have different graphons.
  - G-Mixup can improve the performance of GNNs on various datasets.
  - The loss curve of G-Mixup are lower than the vanilla model
  - G-Mixup can improve the generalization of graph neural networks.
  - Code is available at https://github.com/ahxt/g-mixup





Overview of G-Mixup. The task is binary graph classification.

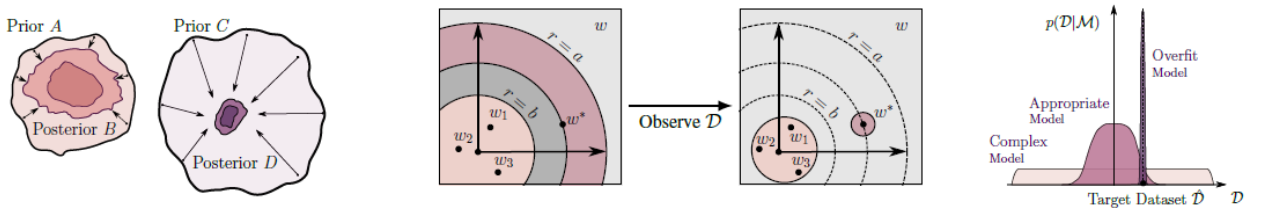


Visualization of generated synthetic graphs on REDD-BINARY dataset.

FUTUREWEI Technologies

Bayesian Model Selection, the Marginal Likelihood, and Generalization (NYU, Outstanding Paper)

- Background
  - The marginal likelihood or Bayesian evidence is the probability of generating our observations from a prior, and widely applied to hypothesis testing and model selection – which trained model is most likely to provide the best generalization.
  - Also, marginal likelihood optimization has been applied for hyperparameter learning with great success, which is known as empirical Bayes.
- Contributions
  - The paper shows that marginal likelihood can be negatively correlated with generalization in neural architecture search and can lead underfitting and overfitting in hyperparameter learning.
  - A conditional marginal likelihood is more aligned with generalization, and practically valuable for large-scale hyperparameter learning, such as deep kernel learning.
    - Dropping first (m-1) terms of LML (Log Marginal Likelihood) decomposition



(a) Posterior contraction with a diffuse prior (b) Marginal likelihood underfitting (c) Marginal likelihood overfitting

FUTUREWEI Technologies

Tutorials

- Validity, Reliability, and Significance: A Tutorial on Statistical Methods for Reproducible Machine Learning

  https://www.cl.uni-heidelberg.de/statnlpgroup/empirical_methods/

- Learning for Interactive Agents
- Welcome to the "Big Model" Era: Techniques and Systems to Train and Serve Bigger Models

- Quantitative Reasoning About Data Privacy in Machine Learning
- Bridging Learning and Decision Making
- Causality and Deep Learning: Synergies, Challenges and the Future
- Climate Change and Machine Learning: Opportunities, Challenges, and Considerations

Welcome to the “Big Model” Era: Techniques and Systems to Train and Serve Bigger Models

- Trends driving big model
- Preliminaries
  - History and problem overview
  - A new view of DL parallelism: inter- and intra-op parallelism
- Inter-op parallelism
- Intra-op parallelism
- New frontiers: auto-parallelism
- Tools, pretrained weights and services