ICML2021

July 18th–July 25th, 2021

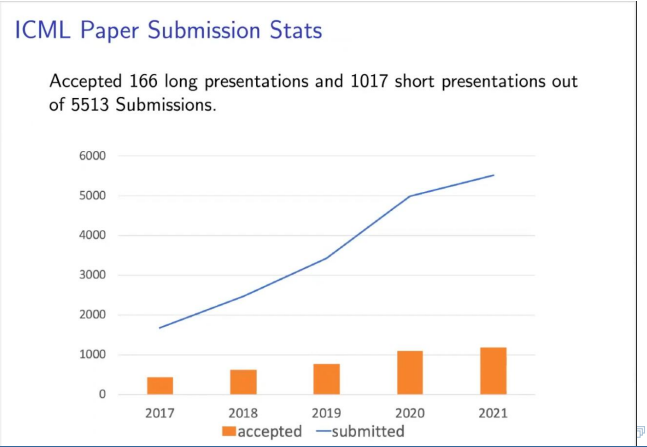
What's ICML

- International Conference on Machine Learning.
  - the leading international academic conference in machine learning, attracting annually about 500 participants from all over the world;
  - began in 1980, in Pittsburgh;
  - recently, co-hosted with the Uncertainty in Artificial Intelligence conference (UAI) and with the Conference on Learning Theory (COLT).
  - Proceedings online: see website for each year
  - https://slideslive.com/icml-2021

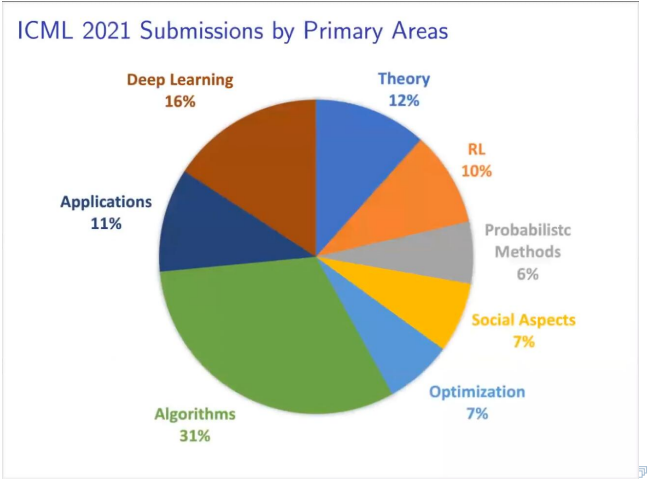
Related Conference

- ML: ICML, NeurIPS, COLT
- Representation Learning: ICLR
- Probability Learing: UAI
- DM: KDD, ICDM
- AI: IJCAI, AAAI
- Applications:
  - CV: CVPR, ICCV, ECCV
  - NLP: ACL
  - MultiMedia: MM
- Other Related:
  - Database: ICDE, SIGMOD, VLDB
  - Software Engineering: ICSE

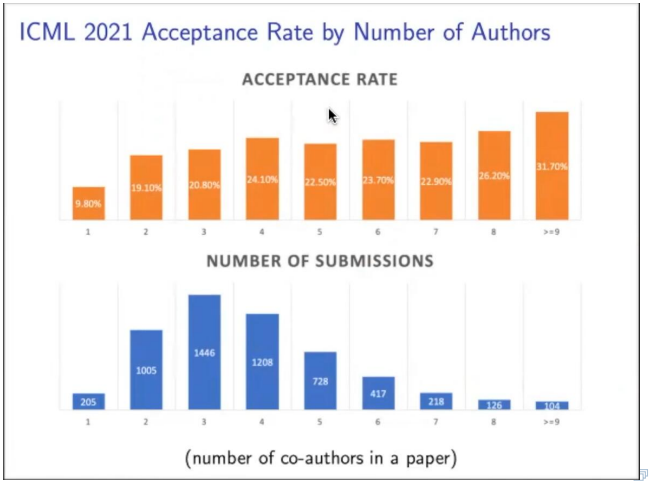
- ICML 2021 共有 5513 篇论文投稿，投稿数量再创新高，共有 1184 篇接受（包括 1018 篇短论文和 166 篇长论文），接受率 21.48%。与往年相比，接受率逐年走低。



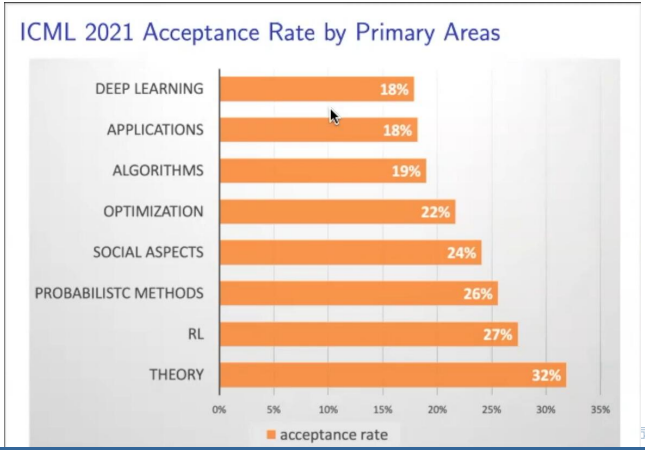
- ICML 2021 提交论文所属领域主要包括算法、深度学习、理论、应用、强化学习（RL)、社会层面、优化以及概率方法。



- ICML 2021 提交论文的论文作者数量，大多数论文都有不止一位共同作者。



- ICML 2021 按论文所属领域的接收率分布如下，其中理论论文接收率最高，随后依次为强化学习（RL）、概率方法、社会层面、优化、算法、应用和深度学习。



- ICML 2021 高频词汇：
  如下图所示，RL 出现频率最高，随后依次是 Noise、Planning、Bandits、Monte Carlo、Redution、Flow、Private 和 Provable。

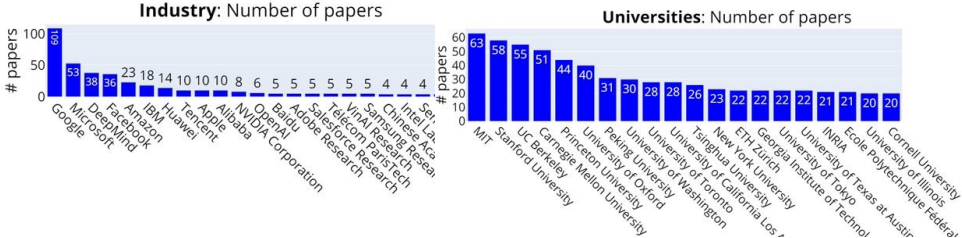
- 按照国家划分，前两位是美国和中国，分别是 729 篇和 166 篇，接着是英国（124 篇）。其中，中国大陆 159 篇相较于去年（122 篇，排名第三）进步明显。

papers	country
729	USA
166+	China
124	UK
79	Canada
48	Germany
45	Switzerland
44	Israel
41	France
38	Japan
35	Singapore
33	South Korea
27	Australia
22	Netherlands
13	Italy
12	India
10	Russia

- 按照机构划分，排在第一位的毫无疑问仍然是 Google，以 109 篇论文霸榜。
- 北京大学以 31 篇排在第 11 位，清华大学以 26 篇排在第 15 位。

papers	organization
109	Google
63	MIT
58	Stanford University
55	UC Berkeley
53	Microsoft
51	Carnegie Mellon University
44	Princeton University
40	University of Oxford
38	DeepMind
36	Facebook
31	Peking University
30	University of Washington
28	University of Toronto
28	University of California Los Angeles
26	Tsinghua University
23	New York University
23	Amazon
22	University of Tokyo
22	University of Texas at Austin
22	Georgia Institute of Technology

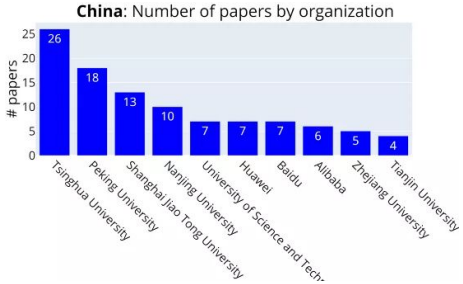
- 学术界的论文有 935 篇，工业界的有 352 篇，意味着大多数论文至少与一所大学有联系。
- 工业界排名前四分别是 Goole，Microsoft，DeepMind，Facebook
- 排名最高的大学 MIT、Stanford、UC Berkeley(整体排名也仅次于谷歌)



- 按照作者来排名，华人的表现非常亮眼

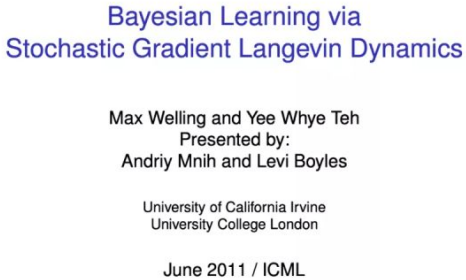
	papers	author
0	14	Masashi Sugiyama (RIKEN / The University of Tokyo)
1	13	Sergey Levine (UC Berkeley)
2	11	Zhuoran Yang (Princeton)
3	11	Zhaoran Wang (Northwestern U)
4	11	Gang Niu (RIKEN)
5	8	Quanquan Gu (University of California, Los Angeles)
6	8	Bo Li (JD)
7	7	Pieter Abbeel (UC Berkeley & Covariant)
8	7	Percy Liang (Stanford University)
9	7	Csaba Szepesvari (DeepMind/University of Alberta)
10	7	Chelsea Finn (Google)
11	7	Bo Han (HKBU / RIKEN)
12	7	Andreas Krause (ETH Zurich)
13	6	Wei Chen (State Key Lab of CAD&CG, Zhejiang University)
14	6	Tongliang Liu (The University of Sydney)
15	6	Shimon Whiteson (University of Oxford)
16	6	Ruslan Salakhutdinov (Carnegie Mellen University)
17	6	Michal Valko (DeepMind / Inria / ENS Paris-Saclay)
18	6	David Woodruff (Carnegie Mellon University)
19	6	Bernhard Schölkopf (MPI for Intelligent Systems Tübingen, Germany)
20	6	Alexandros Dimakis (UT Austin)

- 中国正在迎头赶上。


China: Number of papers by organization
papers
0
5
10
15
20
25
26
18
13
10
7
7
7
6
5
4
Tsinghua University
Peking University
Shanghai Jiao Tong University
Nanjing University
University of Science and Tech
Huawei
Baidu
Alibaba
Zhejiang University
Tianjin University
China: Number of papers by organization
papers
0
10
20
30
31
26
14
14
13
10
10
9
6
Peking University
Tsinghua University
University of Science and
Huawei
Shanghai Jiao Tong Unive
Tencent
Alibaba
Zhejiang University
University of Hong Kong

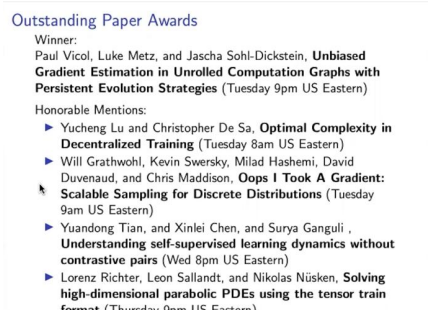
Test of Time Award

- 时间检验奖。



Outstanding Paper

- 来自多伦多大学和谷歌大脑的论文《在展开图中基于持续进化策略的无偏梯度估计》获得了此次会议的杰出论文奖，此外共四篇论文获得了杰出论文荣誉提名奖，其中包括康奈尔大学博士生陆昱成、Facebook 人工智能研究院研究员田渊栋等人参与的研究。



Outstanding Paper Awards

Winner:

Paul Vicol, Luke Metz, and Jascha Sohl-Dickstein, **Unbiased Gradient Estimation in Unrolled Computation Graphs with Persistent Evolution Strategies** (Tuesday 9pm US Eastern)

Honorable Mentions:

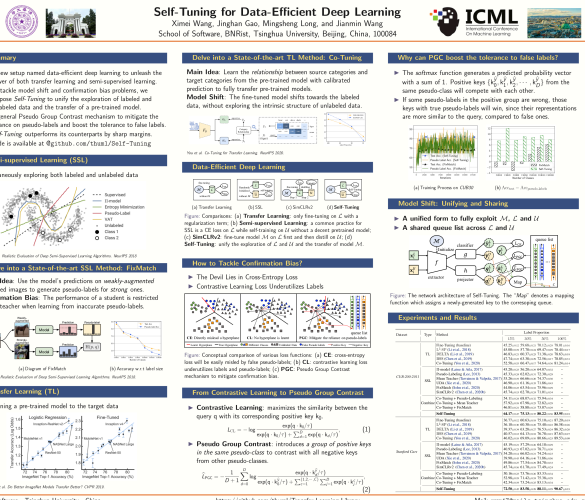
- Yucheng Lu and Christopher De Sa, **Optimal Complexity in Decentralized Training** (Tuesday 8am US Eastern)
- Will Grathwohl, Kevin Swersky, Milad Hashemi, David Duvenaud, and Chris Maddison, **Oops I Took A Gradient: Scalable Sampling for Discrete Distributions** (Tuesday 9am US Eastern)
- Yuandong Tian, and Xinlei Chen, and Surya Ganguli , **Understanding self-supervised learning dynamics without contrastive pairs** (Wed 8pm US Eastern)
- Lorenz Richter, Leon Sallandt, and Nikolas Nüsken, **Solving high-dimensional parabolic PDEs using the tensor train format** (Thursday 9pm US Eastern)

Invited Talk

Invited Talks — ML in Science

- Tue Jul 20 11:00 AM – 12:00 PM (EDT) **Rethinking Drug Discovery in the Era of Digital Biology**, Daphne Koller
- Tue Jul 20 11:00 PM – 12:00 AM (EDT) **Cryospheric Science and Emergence of Machine Learning**, Cunde Xiao (on behalf of Dahe Qin)
- Wed Jul 21 11:00 AM – 12:00 PM (EDT) Title TBA, Ester Duflo
- Wed Jul 21 11:00 PM – 12:00 AM (EDT) **Encoding and Decoding Speech From the Human Brain**, Edward Chang
- Thu Jul 22 11:00 AM – 12:00 PM (EDT) **Machine Learning for Molecular Science**, Cecilia Clementi
- Thu Jul 22 11:00 PM – 11:30 PM (EDT) **Test of the Time Award Talk**

Example of Poster

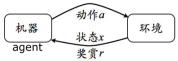


Application on Reinforcement Learning

- 强化学习近几年吸引了大量的关注，e.g.,Deep Mind 的 Alpha Go，OpenAI 的智能机器人。但主要集中在游戏行业，其本身距离具体工业落地应用还存在很大距离，主要有两大原因：
  - 效率问题：包括数据使用效率和算法开发效率。
  - 试错代价高。

强化学习

关键要素：<A, X, R, P>



action space: A
state space: X
reward: $R : X \times A \times X \to \mathbb{R}$
transition: $P : X \times A \times X \to \mathbb{R}$

策略： $a = \pi(x)$

$$P(a|x) = \pi(x, a) \quad \sum_{a \in A} \pi(x, a) = 1 \quad \forall a \in A, \pi(x, a) \geq 0$$

策略评价：累积回报

T-step: $\frac{1}{T}\sum_{t=1}^{T} r_t$ discounted: $\sum_{t=1}^{\infty} \gamma^t r_t$

学习目标：学习最大回报策略

Transformer[1]

- Transformer 最初是作为机器翻译的 Seq2Seq 模型提出的。后来的工作表明，基于 Transformer 的预训练模型 (BERT,GPT3) 可以在各种任务上实现 SOTA。因此，Transformer 已成为 NLP 中的首选架构。除了语言相关的应用，Transformer 还被 CV，音频处理甚至其他学科采用.
- 复旦大学的邱锡鹏教授整理了一份关于 Transformer 的各种变体的 Survey https://arxiv.org/abs/2106.04554.

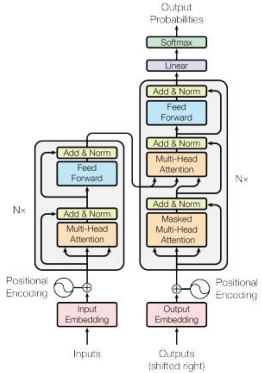


Figure 1: The Transformer - model architecture.

[1] Ashish Vaswani et al. "Attention is All you Need". In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*. 2017, pp. 5998–6008.

Adversarial learning.[2]

- 对抗样本：对抗样本是指将真实的样本添加扰动而合成的新样本，是由深度神经网络的输入的数据和人工精心设计好的噪声合成得到的，但它不会被人类视觉系统识别错误。
  - 如何攻击：白盒攻击、黑盒攻击、真实世界攻击等。
  - 如何防御：主动式防御、反应式防御等。

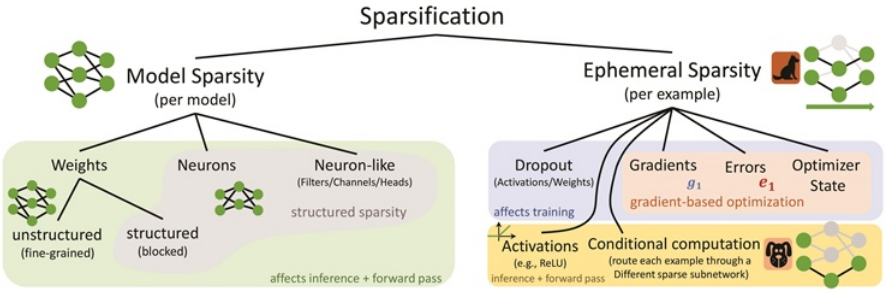
[2] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. "Explaining and Harnessing Adversarial Examples". In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Yoshua Bengio and

ML 中的“多快好省”

- 数据规模问题
  - 学习方法要能够处理大规模数据
- 学习效率问题
  - 学习方法要能够保障学习速率
- 性能保障问题
  - 学习方法要能有性能保障
- 代价抑制问题
  - 学习方法要能有对不同问题的代价保障

1.Larger but Efficent Model[3]

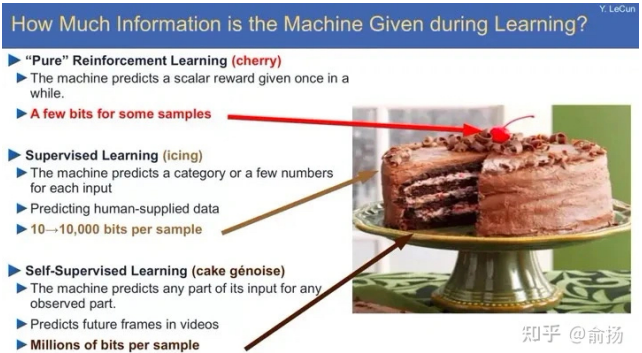
- Sparsity in Deep Learning:Pruning and growth for efficient inference and training.
  - Model Sparsity
  - Ephemeral Sparsity



3 Torsten Hoefler et al. "Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks". In: *CoRR* abs/2102.00554 (2021).

2.Contrastive Learning.

- 对比学习是自监督学习的一种，不依赖标注数据，要从无标注图像中自己学习知识。图像领域里的自监督可以分为两种类型：生成式自监督学习，判别式自监督学习。



2.Contrastive Learning.

- VAE[4]和 GAN[5]是生成式的两类典型方法，即它要求模型重建图像或者图像的一部分。
- 而对比学习则是典型的判别式自监督学习，相对生成式，对比学习的任务难度要低一些。
  - 比较典型的模型包括 SimCLR[6], MoCo[7], BYOL[8], SimSiam[9]等

4 Diederik P. Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. 2014.

5 Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems,NeurIPS 2014*.

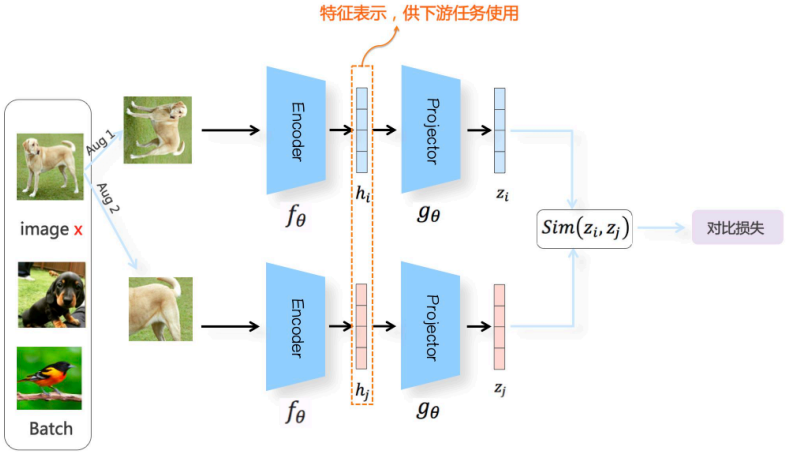
6 Ting Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations". In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. 2020, pp. 1597–1607.

7 Kaiming He et al. "Momentum Contrast for Unsupervised Visual Representation Learning". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 9726–9735.

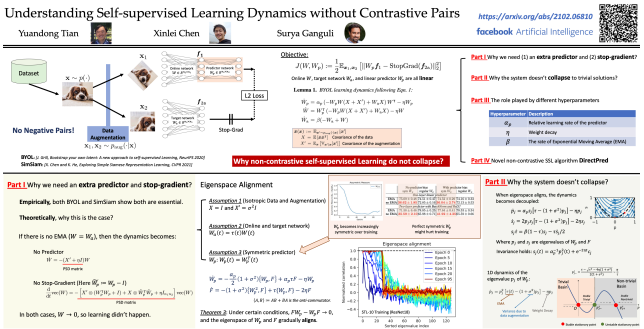
8 Jean-Bastien Grill et al. "Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning". In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. 2020.

9 Xinlei Chen and Kaiming He. "Exploring Simple Siamese Representation Learning". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. Computer Vision Foundation / IEEE, 2021, pp. 15750–15758.

2.Contrastive Learning.

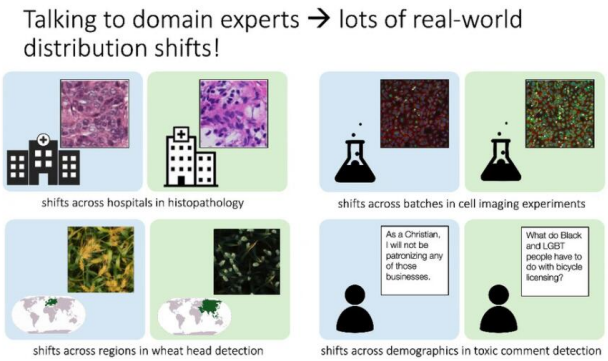


2.Contrastive Learning.



3.Uncertainty and Robustness in Deep Learning

- 现实生活中的数据样本与实验测试用的样本往往存在分布偏差。
- 一个好的模型不仅在 ID(in distribution) 内有着高性能，也要在 OOD(out of distribution) 有着好的泛化性能。



Thanks!