

Course Project

Ju Sun

Computer Science & Engineering

University of Minnesota, Twin Cities

October 18, 2022

General logistics

Project logistics

Project ideas

- HW2 is out. Due: Oct 30th

General logistics

Project logistics

Project ideas

Timeline & L^AT_EX template

- Proposal (5%, 1–2 page): Oct 28
- Progress lightning talks (5%, 5 mins): Nov 29
- Progress report (10%, 3–4 pages): Dec 02
- Final report (20%, 7–8 pages): Dec 20

All page counts exclude references

Template for all writeups: ICLR 2023 L^AT_EX style

<https://github.com/ICLR/Master-Template/raw/master/iclr2023.zip>

Add `\iclrfinalcopy` to the L^AT_EX preamble to make your names visible

Groups

Team ID	A	B	C	D	E	F
Team ID	Student 1 (Name, Email ID)	Student 2 (Name, Email ID)	Student 3 (Name, Email ID)	Student 4 (Name, Email ID)	Student 5 (Name, Email ID)	
Instruction Group	Ju Sun, jusun	Hengkang Wang, wang9881	Yash Travad, trava029			
	Asal Shavandi, shava006	Ali Aalipour, aalip002	Hamidreza Alai, alai0003	Ramin Zandvakili, zandv003		
	Mario Serrafero, serra082	Laura Arias Fernandez, arias046	Harrison Carter, cart0477	Anna Martin, mart5877	Josh Spitzer-Resnick, jresnick001	
	Vishnu Ravichandran, ravico38	Alper Basar, basar027	Zuhayr Nibras Islam, islam075	Cody Wang, wang8974		
	Reese Kneeland, kneel027	Shonal Gangopadhyay, gango010	Joan Petrovic, petro204	Shivam Bhandari, bhand092		
	Zee Myung Kim, kim01756	Min Namgung, namgu007	Minoh Jeong, jeong316			
	Michelle Galarneau, gatar025	Claire Chen, chen6242	Keith Willard, willa013	Brandon Herrera, herre350		
	Bharath Sivaram, sivar019	Tristen Butol, buto002	Pranav Julakanti, julak004	Arun Cherkkil, cherko67		
	Faizel Khan, khanx370	Ahmad Hassan, hassa654	Bharat Jayaprakash, jayap015			
	The Deep Learners	Ryan Devera, dever120	Matthew Tran tran0923	Harrison Russell-Pribnow russe82	Henry Sanford sanfo200	
Kah Meng Soh, soh00009		Jane Huynh, huynh369	Honghui Du, du000211	Nan Wang, wang7254	RuoLei Zeng, Zeng020	
Yoshitaka Inoue, inoue019		Hummin Lee, lee03915	Yunong Xia, xia00045			
Brandon Ung, ung00001		Carl Anderson, and08571	Matthew Senjem, senje001	Vismay Mehta, mehta266		
Anneke von Seeger, vorse006		Irene Noharinaivo, nohar003	Francis Zhang, zhan6806	Harry Hong, hong0506		
Tianhao Zhang, zhan7594		Wenjie Lu, lu000005				
Mouhmad-Ali Elamine, elami016		Tony Liu, liux4408	Nicholas Eickhoff, eickh032			
Abhilash Ojha, ojha0013		Vishal Subedi, subed029	Praveen Ravirathinam, pravr1at			
Frank Bender, bende263		Jashwin Acharya, achar061				
Shan Chen		Daowuan Li, li002504	Yuchi Zhano	Galni Zheno, zhen0256		

- All submissions as a group (in Canvas as group assignment); the group gets the same score

- Prototyping
 - * Google Colab <https://colab.research.google.com/>
 - * Local installation of Jupyter Notebook
<https://jupyter.org/>
 - * MSI notebook notebooks.msi.umn.edu
(<https://www.msi.umn.edu/support/faq/how-do-i-get-started-jupyter-notebooks>)
- Large-scale jobs: **submit them to MSI GPU queues**
 - * MSI quick start
<https://www.msi.umn.edu/quick-start-guides>
 - * Slurm scheduler tutorial
<https://www.msi.umn.edu/slurm>

Five necessary components

- What problem?
- Why interesting?
- Previous work
- Your goal
- Plan and milestones

General logistics

Project logistics

Project ideas

Roughly by ascending level of difficulty

- Literature survey/review
- Novel applications
- Novel methods
- Novel theories

Excerpt from a research project is fine, but you should describe your own contributions

A coherent account of recent papers in a focused topic

- Description and comparison of main ideas, or
- Implementation and comparison of performance, or
- Both of the above

should **complement** the topics we cover in the course



<https://paperswithcode.com/rc2022>

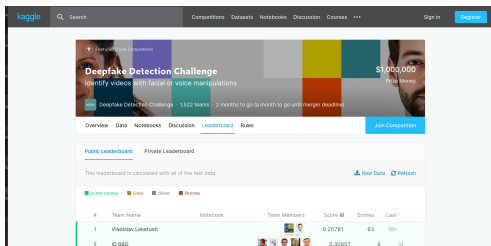
Random topics

- DL for noneuclidean data (e.g., graph NN, manifold NN)
- transformer models for sequential data
- generative models (e.g., GAN, VAE, normalization flow, diffusion models)
- 2nd order methods for deep learning
- constrained optimization for deep learning
- differential programming
- universal approximation theorems
- DL for 3D reconstruction
- DL for video understanding and analysis
- DL for solving PDEs
- DL for inverse problems
- RL for games
- RL for robotics
- DL for medical imaging
- DL for (astro)physics
- DL for chemistry
- adversarial attacks; robustness of DL
- privacy, fairness in DL
- visualization for DNN
- network quantization and compression
- hardware/software platforms for DL
- automated ML; architecture search
- optimization/generalization theory of DL

Novel applications

Apply DL to **new** application problems

- A good place to start: Kaggle <https://www.kaggle.com/>



The screenshot shows the Kaggle website interface for the 'Deepfake Detection Challenge'. The challenge title is 'Deepfake Detection Challenge' with a subtitle 'Identify videos with facial or voice manipulations'. A prize of '\$1,000,000' is displayed. The challenge has '1,522 teams' and '2 months to go (a month to go until merger deadline)'. The 'Leaderboard' tab is selected, showing a table of team rankings. The top team is 'Vladislav Lebedev' with a score of 0.25781. The second team is 'ID:RAD' with a score of 0.20857.

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Vladislav Lebedev			0.25781	83	10h
2	ID:RAD			0.20857	8	1d

- Think about data availability

Google dataset search

<https://datasetsearch.research.google.com/>

- Think about GPUs

Where to find inspirations

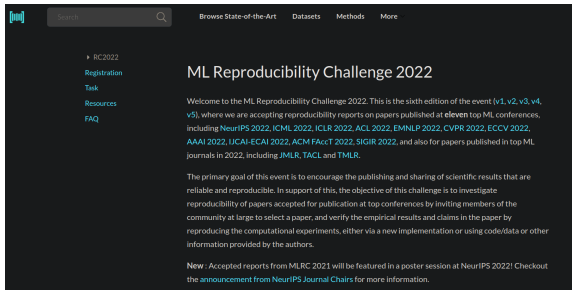
- arXiv machine learning
<https://arxiv.org/list/cs.LG/recent>
- Recent conference papers
 - ML: NeurIPS, ICML, ICLR, etc
 - CV: ICCV, ECCV, CVPR, etc
 - NLP: ACL, EMNLP, etc
 - Robotics: ICRA, etc
 - Graphics: SIGGRAPH, etc
- Talk to researchers!

Novel methods

Create new **NN models or training algorithms** to improve the state-of-the-art

Where to start:

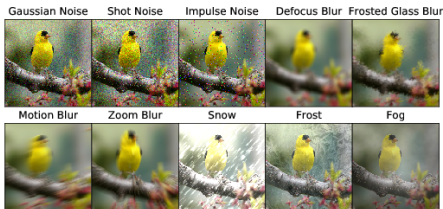
- Kaggle (again)!
- arXiv machine learning and recent conference papers
- MLRC



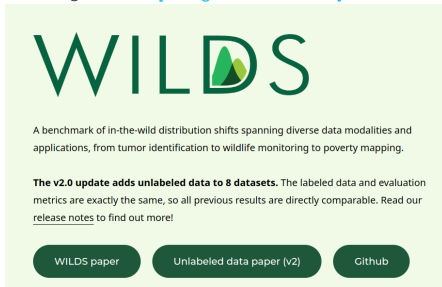
The screenshot shows the homepage of the ML Reproducibility Challenge 2022. The page has a dark theme with a white sidebar on the left containing navigation links: RC2022, Registration, Task, Resources, and FAQ. The main content area features the title 'ML Reproducibility Challenge 2022' and a welcome message. The welcome message states that this is the sixth edition of the event, accepting reports on papers published at eleven top ML conferences: NeurIPS 2022, ICML 2022, ICLR 2022, ACL 2022, EMNLP 2022, CVPR 2022, ECCV 2022, AAAI 2022, UCAI-ECAI 2022, ACM FAccT 2022, SIGIR 2022, and also for papers published in top ML journals in 2022, including JMLR, TACL, and TMLR. Below this, the primary goal of the event is described: to encourage the publishing and sharing of scientific results that are reliable and reproducible. The objective is to investigate the reproducibility of papers accepted for publication at top conferences by inviting members of the community at large to select a paper, verify the empirical results and claims in the paper by reproducing the computational experiments, either via a new implementation or using code/data or other information provided by the authors. At the bottom, a 'New' section announces that accepted reports from MLRC 2021 will be featured in a poster session at NeurIPS 2022 and directs users to check out the announcement from NeurIPS Journal Chairs for more information.

<https://paperswithcode.com/rc2020>

Novel methods



Credit: ImageNet-C <https://github.com/hendrycks/robustness>



WILDS

A benchmark of in-the-wild distribution shifts spanning diverse data modalities and applications, from tumor identification to wildlife monitoring to poverty mapping.

The v2.0 update adds unlabeled data to 8 datasets. The labeled data and evaluation metrics are exactly the same, so all previous results are directly comparable. Read our [release notes](#) to find out more!

WILDS paper Unlabeled data paper (v2) Github

Credit: WILDS <https://wilds.stanford.edu/>

Equally interesting to fool/fail the state-of-the-art, e.g., exploring robustness of DL, finding common limitations of state-of-the-art

Novel theories

Nothing is more practical than a good theory. – V. Vapnik

- universal approximation theorems
- nonconvex optimization
- generalization

Where to start:

- Analyses of Deep Learning (Stanford, fall 2019)
<https://stats385.github.io/>
- Theories of Deep Learning (Stanford, fall 2017)
https://stats385.github.io/stats385_2017.github.io/
- Toward theoretical understanding of deep learning (ICML 2018 Tutorial)
<https://unsupervised.cs.princeton.edu/deeplearningtutorial.html>

Questions?