Beyond ERM: What Optimization can help Large Foundation Models

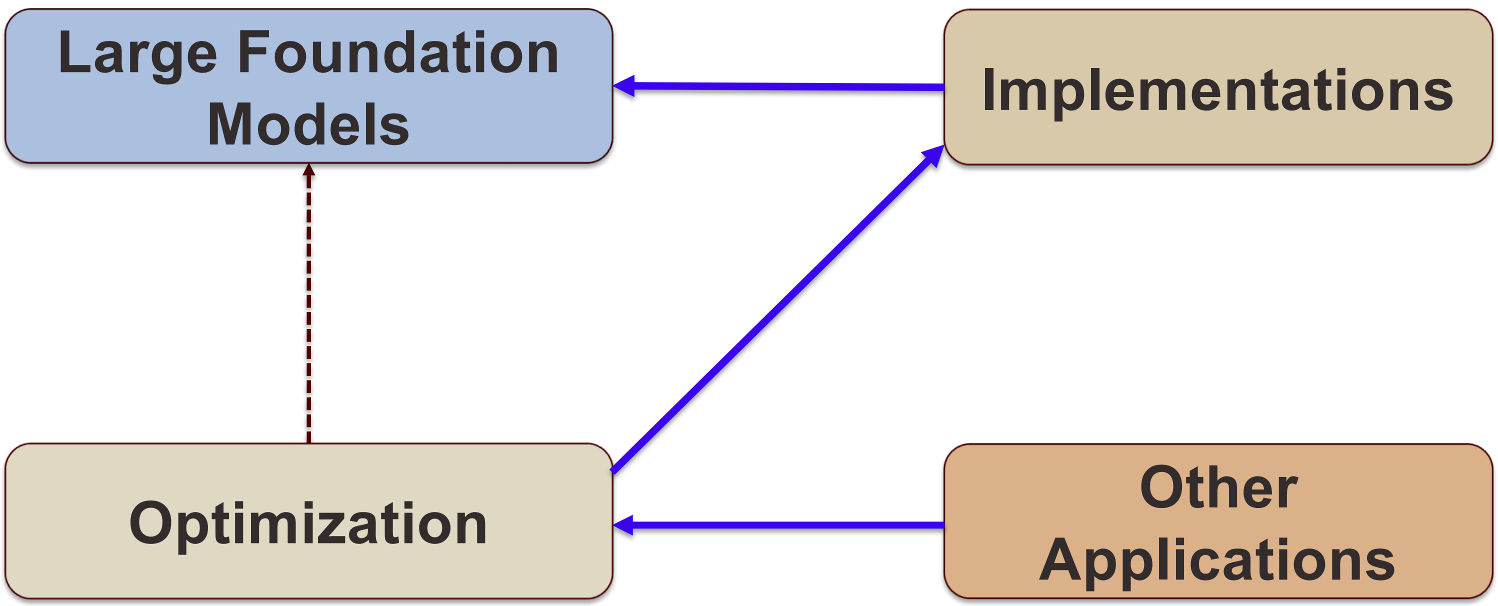
Tianbao Yang
CSE@TAMU

Theory and Practice

“The best theory is inspired by practice. The best practice is inspired by theory” - Donald Knuth

Overview

$$\min_{\mathbf{w}} F(\mathbf{w})$$



New Learning Paradigm

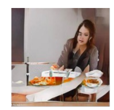
Labels

Supervised Learning

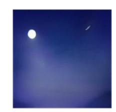
Expensive to Label

Free-texts Description

Self-Supervised Learning (SSL)



A woman is sitting on the chair at the table on which a full plate pizza and a tissue and fork are there. Behind her there is a wall.



In this image I can see a aeroplane which is white in color flying in the air, and in the background I can see the sky and the moon..

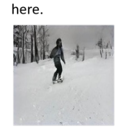
In this image we can see five sheep, in the front we can see some grass and some rocks, in the background we can see some trees, in the middle there is water, we can also see sky and clouds here.

In the image we can see there is a dog who is sitting on the ground and there is a kennel, there is a dog bed in it and there is a dog bowl.

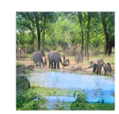
In this picture we can see a person skating on a skateboard, in the background we can see a shed, a pole, some of the trees here, on the top we can see a cloudy sky, this person wore a black color t-shirt.

This is the picture of the river. There is a boat on the water. There are two persons and a dog in the boat and there is a flag at the end of the boat. At the back there are trees. At the top there is a sky. At the bottom there is a water.

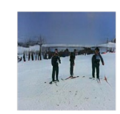
In the image there is a spring roll cheese pizza on a wooden plate and back of it there is another pizza on a wooden plate, Both are on floor.



In the picture there is a snow, in a which a person is diving in the snow with the skateboard, there are many trees covered with the



This is a picture taken in the outdoors. It is sunny. There are group of elephants drinking water in the river. Behind



Here we can see three persons are skating on the snow with ski boards. They wear a helmet and he has goggles. In the background there is a sky.

Easy to obtain

Multimodal Generative AI

Q: what is the flower in this image?

Gemini

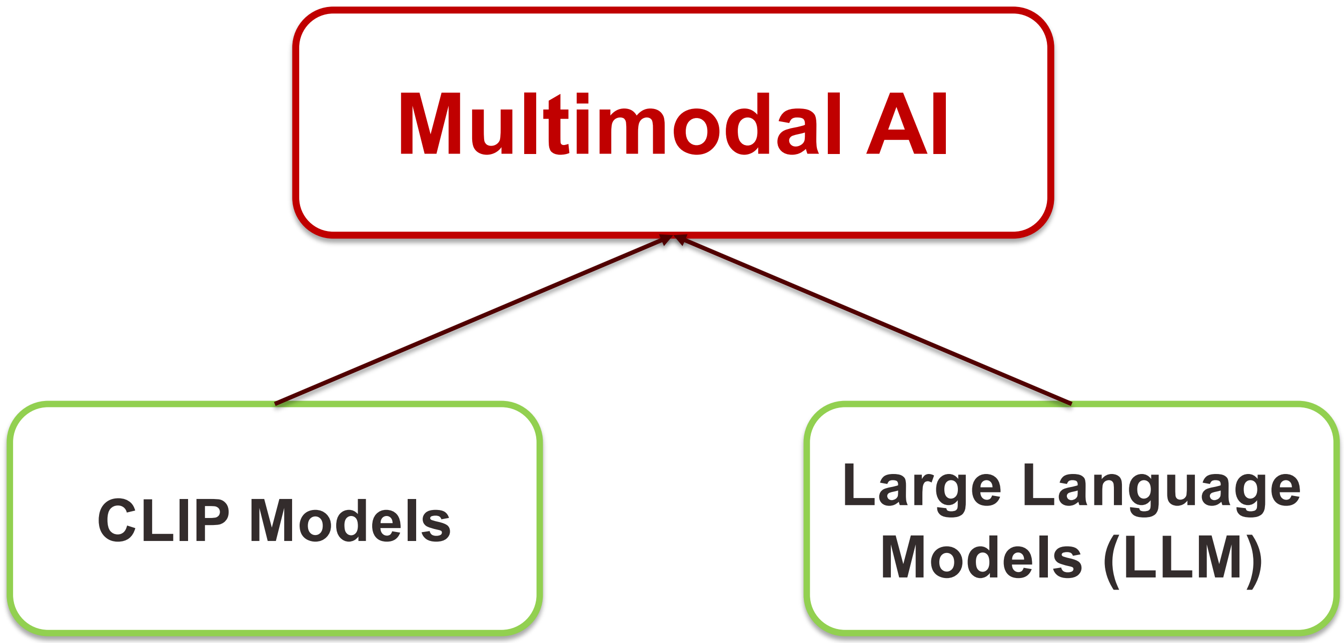
The flower in the image is most likely a showy lady's slipper (Cypripedium reginae). It is a perennial flowering plant native to temperate regions of North America, Europe, and Asia.

Showy lady's slippers are known for their large, pink or white pouch-like flowers with a spotted or striped interior. The flowers bloom in late spring or early summer. The plant grows in moist, shady woodlands and prefers acidic soil.

Here are some other features of showy lady's slippers:

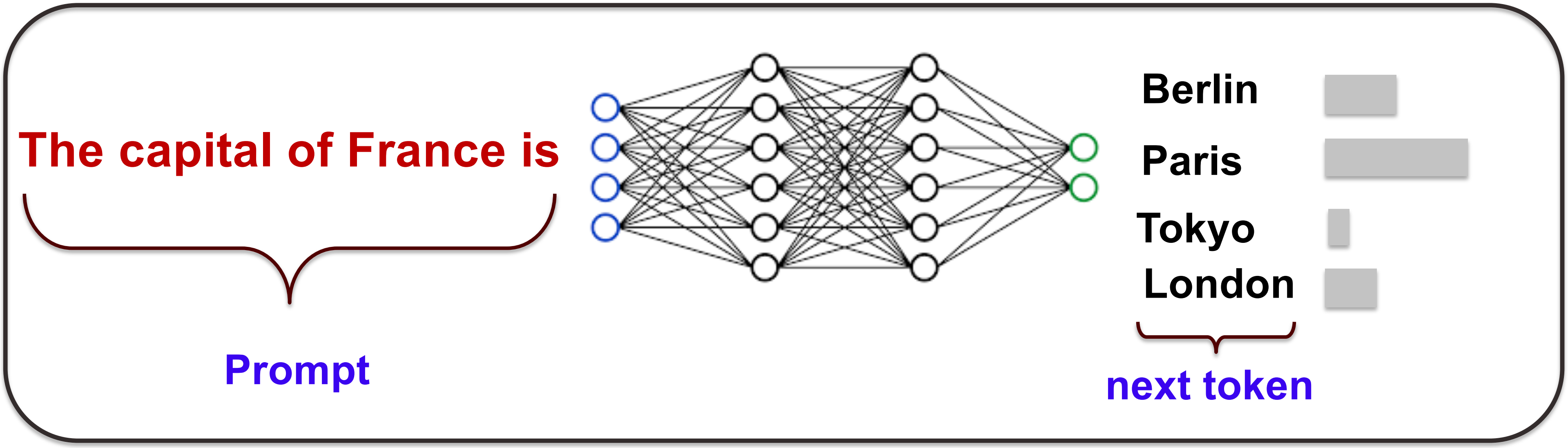
- They have a single, hairy stem that grows up to 3 feet tall.
- The leaves are broad and elliptical, and they are arranged in a whorl around the stem.
- The flowers are solitary and terminal, meaning they grow at the end of the stem.
- The pouch-like lip of the flower is called the labellum. It is designed to trap insects, which pollinate the flower.

Large Foundation Models



Large Language Models

Predicting the Next Token



Example

Instruction: Provide a name for the dish given the ingredients and instructions.
INGREDIENTS: 2 (5 oz) cans Bumble Bee® Solid White Albacore Tuna, drained 1 avocado
2 Tbsp Sriracha
1 Tbsp Dijon mustard
2 to 3 Tbsp celery, chopped
2 Tbsp red onion, chopped
2 green onions, chopped
1 Tbsp fresh cilantro, chopped
Salt and pepper, to taste
2 heaping cups leafy green lettuce
1 cup matchstick carrots
4 (10 inch) whole wheat tortillas
INSTRUCTIONS: In a medium bowl, mash together tuna and avocado until combined. Add in the rest of the ingredients through the salt and pepper, mixing well. To assemble, top each tortilla with a 1/2 cup leafy greens, 1/4 cup matchstick carrots and divide the tuna mixture evenly among the wraps. Tightly roll up the tortilla, slice and enjoy!

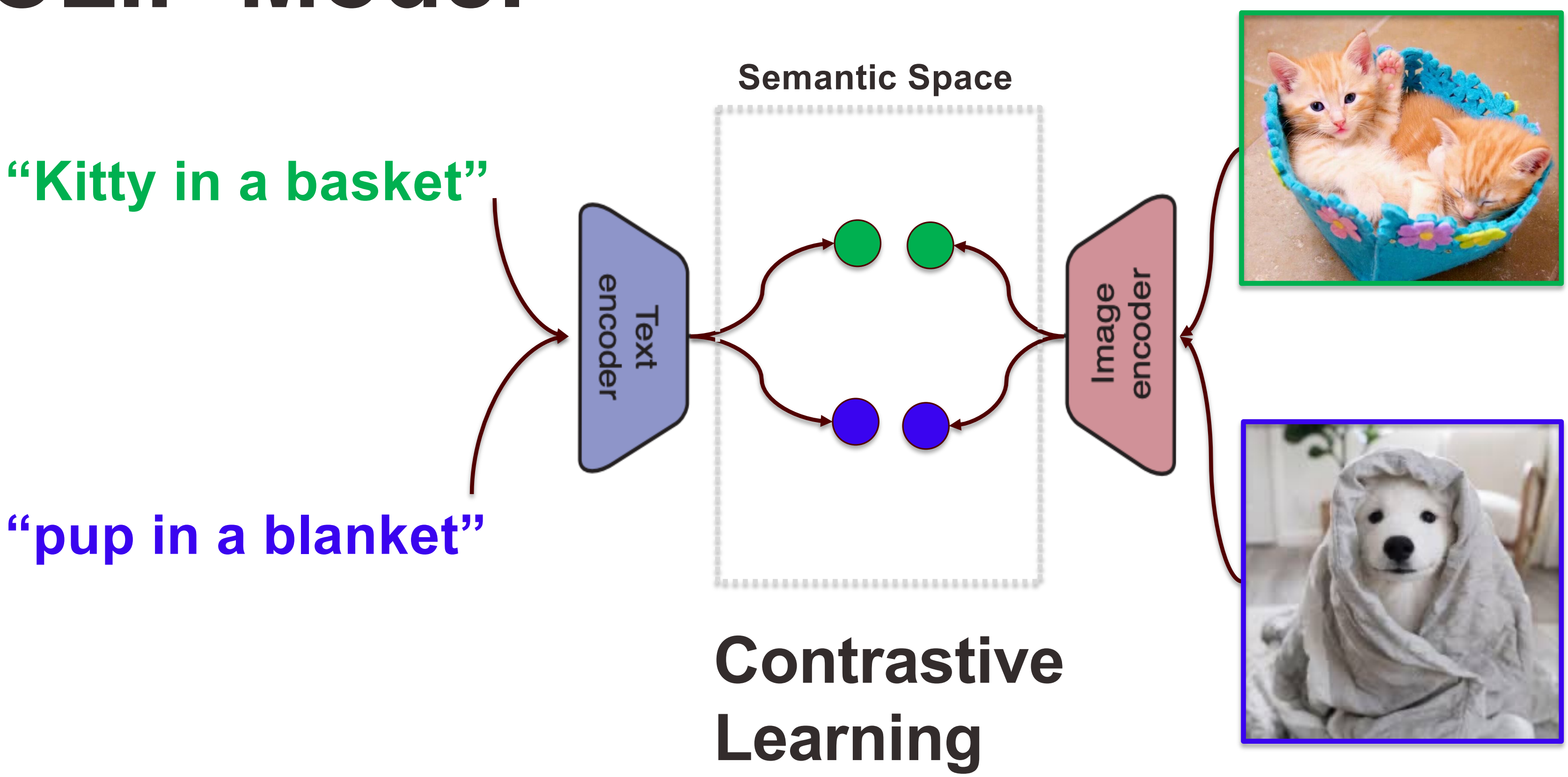
Gemini. "**Spicy Tuna Avocado Wraps with Crunchy Vegs**"

CHATGPT … "**Spicy Tuna Avocado Roll-Ups**" …

TempNet … "**Tunanadoes**" …

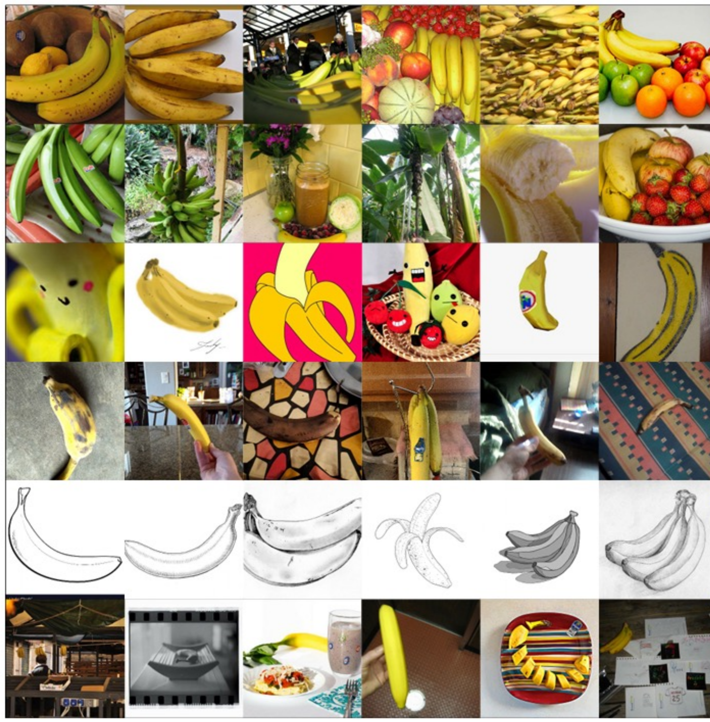
Qiu et al. 2024

CLIP Model



"Kitty in a basket"
"pup in a blanket"
Text encoder
Semantic Space
Image encoder
Contrastive Learning

Example

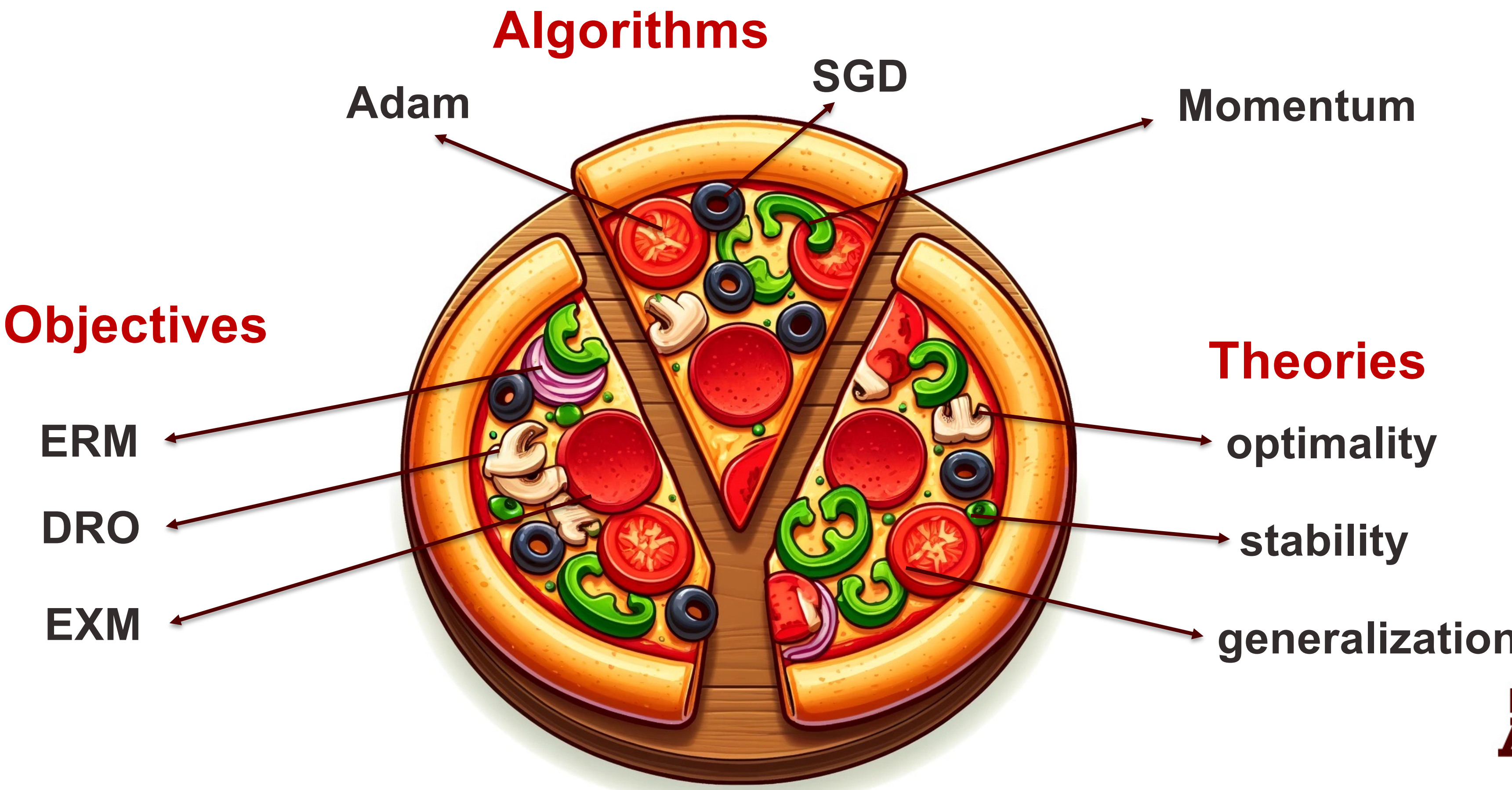


Dataset Examples	ImageNet ResNet101	Zero-Shot CLIP	Δ Score
ImageNet	**76.2**	**76.2**	0%
ImageNetV2	64.3	**70.1**	+5.8%
ImageNet-R	37.7	**88.9**	+51.2%
ObjectNet	32.6	**72.3**	+39.7%
ImageNet Sketch	25.2	**60.2**	+35.0%
ImageNet-A	2.7	**77.1**	+74.4%

Radford et al. 2021

Optimization for ML/AI

Optimization for ML/AI



Empirical Risk Minimization (ERM)

$$\min_{\mathbf{w}} F(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} \ell(\mathbf{w}, \mathbf{x}_i, y_i)$$

SGD

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \frac{1}{B} \sum_{i \in \mathcal{B}_t} \nabla \ell(\mathbf{w}_t, \mathbf{x}_i, y_i)$$

Standard Deep Learning Pipeline

Sample Mini-batch Samples

↓

Define Mini-batch (MB) Losses

↓

Back-propagation on MB Losses

↓

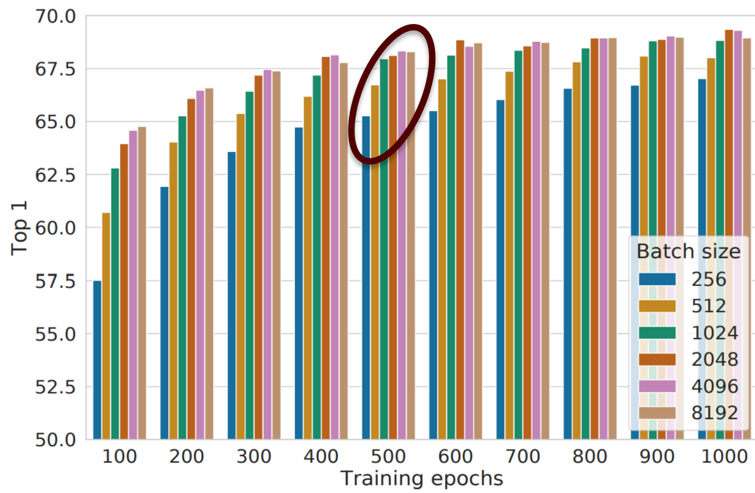
Update Model Parameters

PyTorch → Back-propagation on MB Losses

TensorFlow → Update Model Parameters

Training Challenge for Contrastive Learning

Self-supervised Learning



SimCLR: Chen et al. 2020.

Ordinary batch size has a large error

Google OpenAI

Scaling up #of GPUs

Not Sustainable

Global Contrastive Objective

Contrastive Learning

"Kitty in a basket" ... "pup in a blanket"

	T_1	T_2	T_3	...	T_N
I_1	$I_1 \cdot T_1$	$I_1 \cdot T_2$	$I_1 \cdot T_3$	...	$I_1 \cdot T_N$
I_2	$I_2 \cdot T_1$	$I_2 \cdot T_2$	$I_2 \cdot T_3$	...	$I_2 \cdot T_N$
I_3	$I_3 \cdot T_1$	$I_3 \cdot T_2$	$I_3 \cdot T_3$	...	$I_3 \cdot T_N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
I_N	$I_N \cdot T_1$	$I_N \cdot T_2$	$I_N \cdot T_3$	...	$I_N \cdot T_N$

Motivation

 "pup in a blanket"

 "Kitty in a basket"

$$\ell_{ij} = I_i^\top T_j - I_i^\top T_i$$

good idea?

$$L_i = \frac{1}{N} \sum_{j=1}^{N} \ell_{ij}$$

Global Contrastive Loss

Temperature

$$L_i = \tau \log \left(\frac{1}{N} \sum_{j=1}^{N} \exp(\ell_{ij}/\tau) \right)$$

Previous studies just use a mini-batch inside the *log*

(Yuan et al. '22)

Distributionally Robust Optimization (DRO)

$$L_i = \max_{p \in \Delta} \sum_j p_j \ell_{ij} - \tau \mathrm{KL}(\mathbf{p}, 1/N)$$

$$\Updownarrow$$

$$L_i = \tau \log \left(\frac{1}{N} \sum_{j=1}^{N} \exp(\ell_{ij}/\tau) \right)$$

(Qiu et al. ’23)

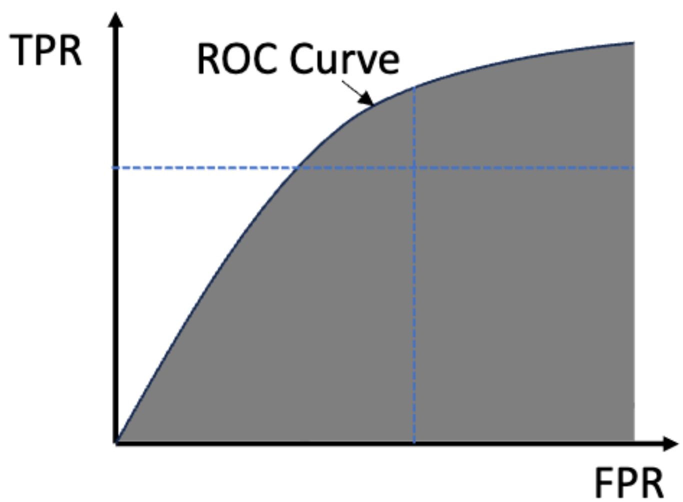
Global Contrastive Objective

$$\frac{1}{N}\sum_{i=1}^{N}\tau\log\left(\frac{1}{N}\sum_{j=1}^{N}\exp(\ell_{ij}/\tau)\right)$$

A Story about Application

1st Place at Stanford CheXpert Competition

$$\min_{u} \max_{v} f(u, v)$$



(Yuan et al. '20)

Leaderboard

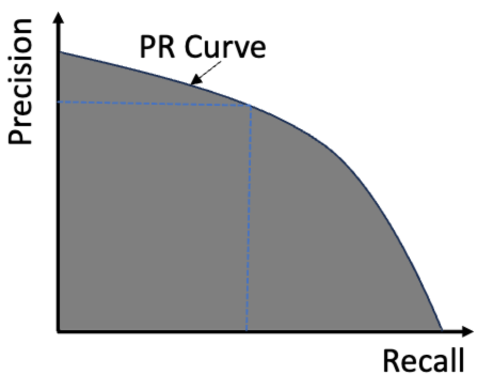
Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	DeepAUC-v1 *ensemble* https://arxiv.org/abs/2012.03173	0.930	2.8
2	Sep 01, 2019	Hierarchical-Learning-V1 (ensemble) *Vingroup Big Data Institute* https://arxiv.org/abs/1911.06475	0.930	2.6
3	Oct 16, 2019	Conditional-Training-LSR *ensemble*	0.929	2.6
4	Dec 04, 2019	Hierarchical-Learning-V4 (ensemble) *Vingroup Big Data Institute* https://arxiv.org/abs/1911.06475	0.929	2.6

MIT AICures Challenge

Fighting Secondary Effects of Covid

Evaluation Metric: **AUPRC**



(a) Test PRC-AUC

Rank	Model	Author	Submissions	Test PRC-AUC
1	MolecularG	AIDrug@PA	7	0.725
2	-	AGL Team	20	0.702
3	MoleculeKit	DIVE@TAMU	7	0.677
4	GB	BI	6	0.67
5	Chemprop ++	AICures@MIT	4	0.662
6	-	Mingjun Liu	3	0.657
7	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	2	0.651
8	RF + fingerprint	Cyrus Maher@Vir Bio	1	0.649
9	Graph Self-supervised Learning	SJTU_NRC_Mila	3	0.622
10	-	Congjie He	10	0.611

(b) Test ROC-AUC

Rank	Model	Author	Submissions	Test ROC-AUC
1	MoleculeKit	DIVE@TAMU	7	0.928
2	Chemprop ++	AICures@MIT	4	0.877
3	-	Gianluca Bontempi	7	0.848
4	-	Apoorv Umang	1	0.84
5	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	2	0.837
6	-	Kexin Huang	1	0.824
7	Chemprop	Rajat Gupta	7	0.818
8	MLP	IITM	7	0.807
9	Graph Self-supervised Learning	SJTU_NRC_Mila	3	0.8
10	-	Congjie He	10	0.8

Estimator: Average Precision

$$\mathrm{AP}(h) = \boxed{\frac{1}{n_+} \sum_{\mathbf{x}_i \in \mathcal{S}_+}} \boxed{\mathrm{Precision}(\boxed{h(\mathbf{x}_i)})}$$

Positive Examples

$$\mathrm{Precision}(h(\mathbf{x}_i)) = \frac{\sum_{\mathbf{x}_j \in \mathcal{S}_+} \mathbb{I}(h(\mathbf{x}_j) \geq h(\mathbf{x}_i))}{\sum_{\mathbf{x}_j \in \mathcal{S}} \mathbb{I}(h(\mathbf{x}_j) \geq h(\mathbf{x}_i))}$$

All Examples

1st Place at MIT AICures Challenge

Fighting Secondary Effects of Covid

Rank	Model	Author	Submissions	10-fold CV ROC-AUC	10-fold CV PRC-AUC	Test ROC-AUC	Test PRC-AUC
1		DIVE@TAMU	11			0.957	0.729
2	MolecularG	AIDrug@PA	9			0.7	0.725
3		AGL Team	20			0.675	0.702
4		phucdoitoan@Fujitsu	14	0.898 +/- 0.113	0.508 +/- 0.253	0.867	0.694
5	GB	BI	6			0.698	0.67
6	Chemprop ++	AICures@MIT	4			0.877	0.662
7		Mingjun Liu	3			0.72	0.657
8	Pre-trained OGB-GIN (ensemble)	Weihua Hu@Stanford	2	0.905 +/- 0.133	0.494 +/- 0.333	0.837	0.651
9	RF + fingerprint	Cyrus Maher@Vir Bio	1	0.896 +/- 0.074	0.481 +/- 0.338	0.799	0.649
10	Graph Self-supervised Learning	SJTU_NRC_Mila	3	0.825 +/- 0.210	0.530 +/- 0.342	0.800	0.622

1st Place at MIT AICures Challenge

Rank	Model	Author	Submissions	10-fold CV ROC-AUC	10-fold CV PRC-AUC	Test ROC-AUC	Test PRC-AUC
1		DIVE@TAMU	11			0.957	0.729

AUROC

1	MoleculeKit	DIVE@TAMU	7	0.928

AUPRC

w/o LibAUC

3	MoleculeKit	DIVE@TAMU	7	0.677

5% Improvement in **AUPRC**, **3%** Improvement in **AUROC**

Wang et al. (Bioinformatics'22)

Empirical X-risk Minimization: Algorithms & Theories

Empirical X-risk Minimization (EXM)

$$\min_{\mathbf{w}} F(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} f_i(g(\mathbf{w}, \mathbf{x}_i, \mathcal{S}_i))$$

X-risk

A family of **Compositional Risks** in which the loss function of each data point is defined in a way that **Contrasts** the data point with **Many** others.

$$g(\mathbf{w}, \mathbf{x}_i, \mathcal{S}_i) = \frac{1}{|\mathcal{S}_i|}\sum_{\mathbf{x}_j \in \mathcal{S}_i} \ell(\mathbf{w}; \mathbf{x}_i, \mathbf{x}_j)$$

Yang (arXiv'22), Yuan et al. KDD'23

Global Contrastive Objective

$$\frac{1}{N}\sum_{i=1}^{N}\tau\log\left(\boxed{\frac{1}{N}\sum_{j=1}^{N}\exp(\ell_{ij}/\tau)}\right)$$

$$\min_{\mathbf{w}} F(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} f_i(g(\mathbf{w}, \mathbf{x}_i, \mathcal{S}_i))$$

Optimization Challenge

EXM

$$\nabla f_i(g(\mathbf{w}, \mathbf{x}_i, \mathcal{S}_i))\nabla g(\mathbf{w}, \mathbf{x}_i, \mathcal{S}_i)$$

$$\mathbb{E} \neq$$

$$\nabla f_i(g(\mathbf{w}, \mathbf{x}_i, \mathcal{B}_i))\nabla g(\mathbf{w}, \mathbf{x}_i, \mathcal{B}_i)$$

Mini-batch

BSGD (Hu et al' 20)

ERM

$$\frac{1}{n}\sum_{i=1}^{n}\nabla \ell(\mathbf{w}, \mathbf{x}_i)$$

$$\mathbb{E} =$$

$$\frac{1}{B}\sum_{\mathbf{x}\in\mathcal{B}}\nabla \ell(\mathbf{w}, \mathbf{x})$$

Principles of Algorithm Design

1. **Simple** as SGD for ERM
2. **Constant** Batch Size Ensures Convergence
3. **Same-order** Complexity as SGD for ERM
4. **Parallel** Speed-up using Mini-Batch

SOX: update

Gradient **E**stimator

$$G_t = \frac{1}{B} \sum_{\mathbf{x}_i \in \mathcal{B}_1^t} \nabla f_i(\boxed{u_i^t}) \nabla g(\mathbf{w}_t, \mathbf{x}_i, \mathcal{B}_i)$$

Estimator of g_i

$$u_i^{t+1} = \begin{cases} (1-\gamma_t)u_i^t + \gamma_t g(\mathbf{w}_t, \mathbf{x}_i, \mathcal{B}_i), & \mathbf{x}_i \in \mathcal{B}_1^t \\ u_i^t & \text{o.w.} \end{cases}$$

Model **U**pdate

$$\mathbf{v}_{t+1} = \beta_1 \mathbf{v}_t + (1-\beta_1)G_t$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \mathbf{v}_{t+1}$$

Momentum or **A**dam

Wang & Yang. ICML'22

SOX: complexity

Assumption: smoothness & Lipschitz continuity

Iteration Complexity

$$\mathbb{E}\|\nabla F(\mathbf{w})\| \leq \epsilon \Leftarrow O\left(\frac{\boxed{n}}{\boxed{B_1 B_2}\boxed{\epsilon^4}}\right)$$

SCD of u

Match SGD

Parallel Speed-up

Better Total Complexity than BSGD when $n \leq \frac{1}{\epsilon^2}$

Wang & Yang. ICML'22

Improvements

Assumption: *f* is cvx, monotone, *g* convex

Matching our lower bounds

Strongly-Convex

$$O\left(\frac{n}{B_1 B_2 \epsilon} + \frac{1}{\mu \min(B_1, B_2)\epsilon}\right)$$

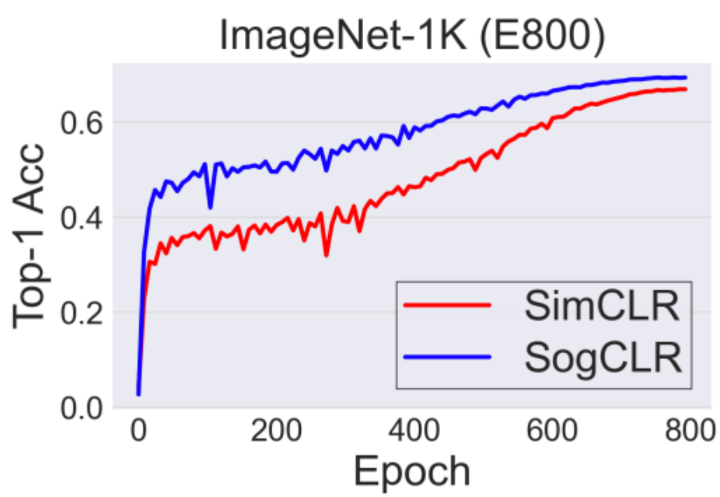
Convex

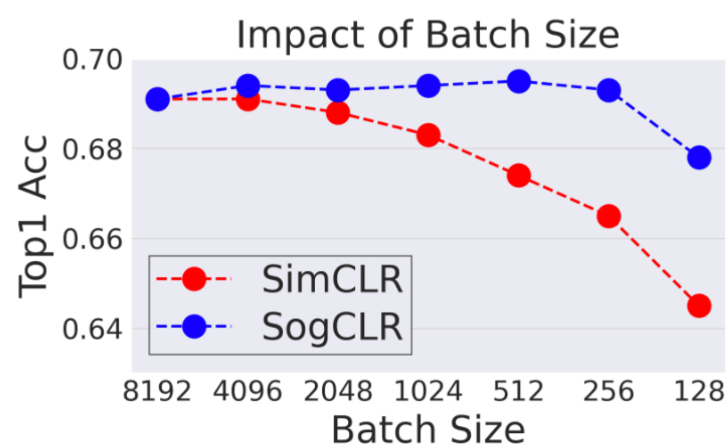
$$O\left(\frac{n}{B_1 B_2 \epsilon^2}\right)$$

Wang & Yang arxiv'24

SogCLR: for image modality

1.2M Images

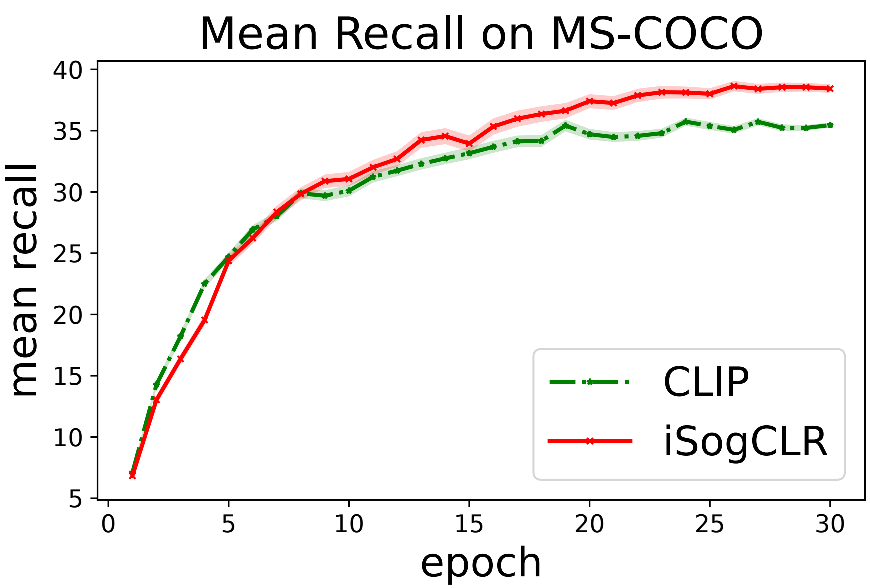


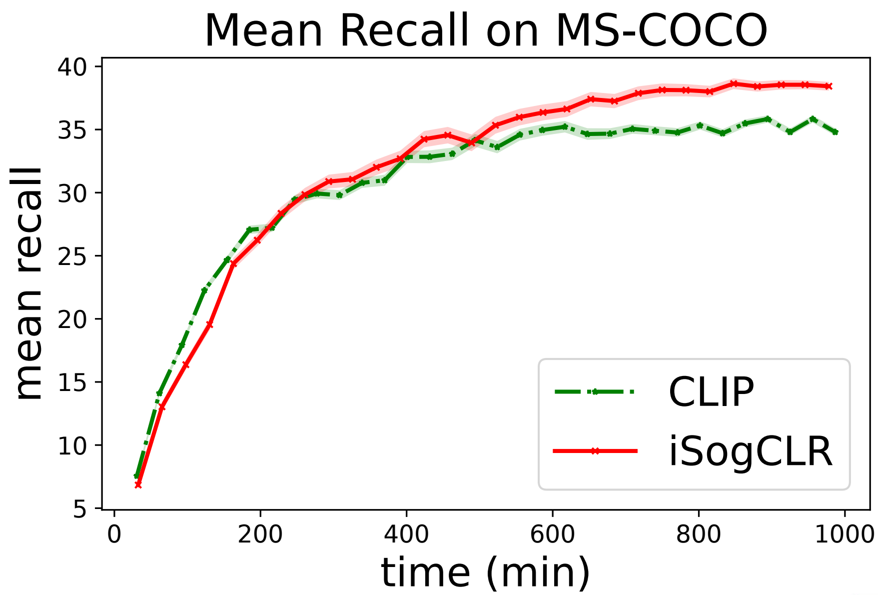


Yuan et al. (ICML’22)

iSogCLR: for CLIP training

3M Image-text pairs

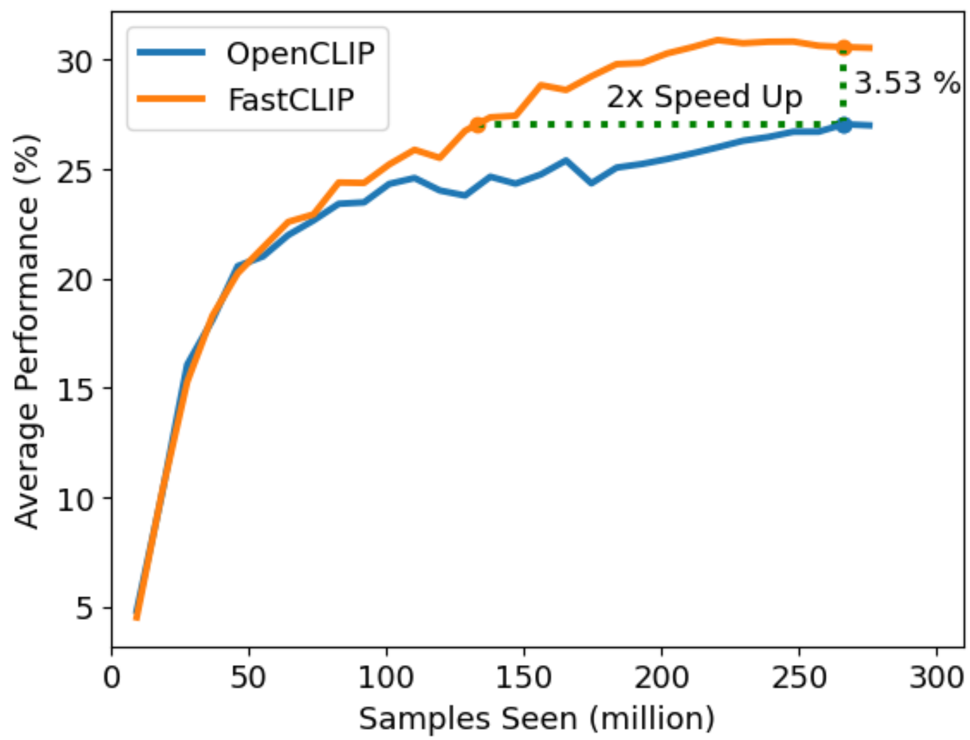




Qiu et al. (ICML'23)

FastCLIP: distributed training

12M Image-text pairs



Orange: Ours **Blue: OpenCLIP**

Personalized Temperature and Constrained DRO

How to set the temperature

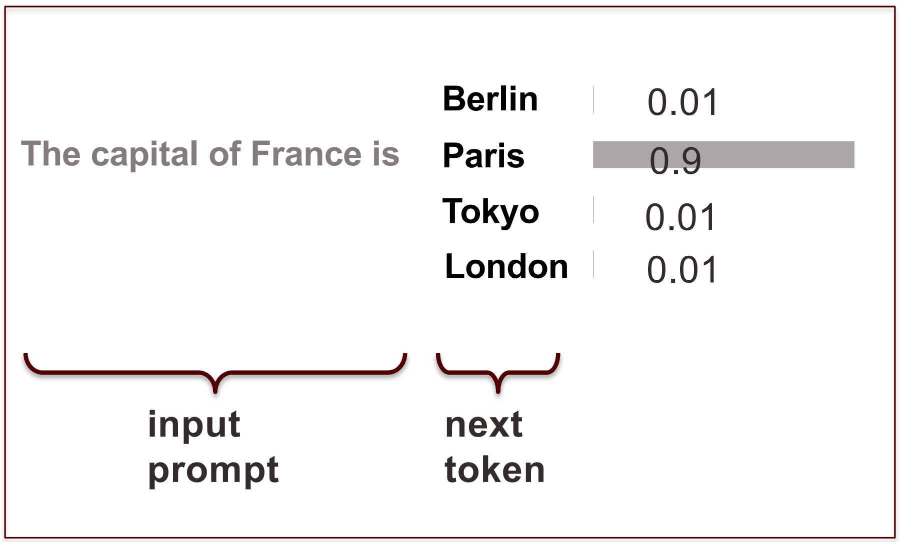
$$L_i = \max_{p \in \Delta} \sum_j p_j \ell_{ij} - \tau \mathrm{KL}(\mathbf{p}, 1/N)$$

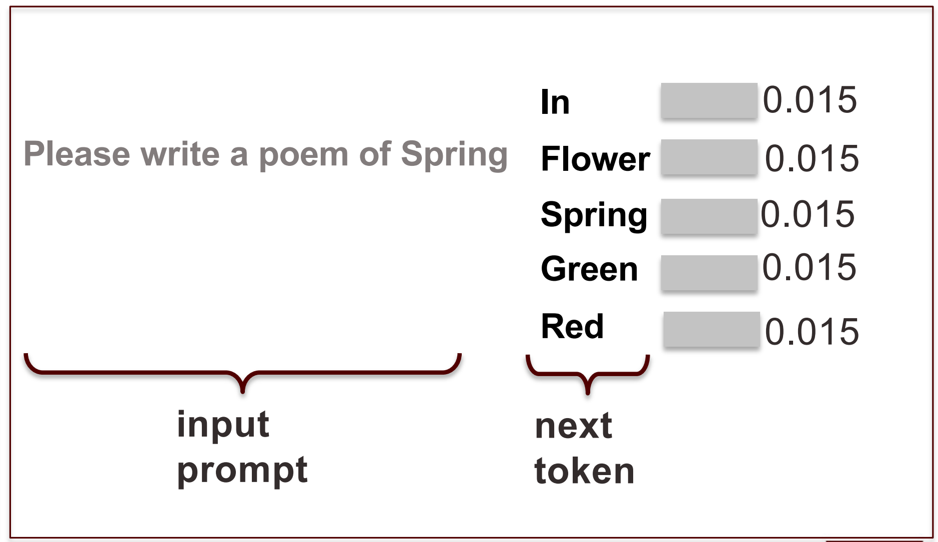
$$L_i = \underbrace{\tau}_{\text{Temperature}} \log \left(\frac{1}{N} \sum_{j=1}^{N} \exp(\ell_{ij}/\tau) \right)$$

(Qiu et al. '23)

Temperature in LLM

Compute probabilities of next token $\Pr(t_j|\mathbf{x}_i) = \dfrac{\exp(s_{ij}/\tau)}{\sum_{k=1}^{N} \exp(s_{ik}/\tau)}$





Personalization by Constrained DRO

$$L_i = \max_{p \in \Delta} \sum_j p_j \ell_{ij} \quad \text{s.t.} \quad \mathrm{KL}(p, 1/N) \leq \rho$$

$$\Updownarrow$$

$$L_i = \min_{\tau_i \geq \tau_0} \tau_i \log \left(\frac{1}{N} \sum_{j=1}^{N} \exp(\ell_{ij}/\tau_i) \right) + \rho \tau_i$$

(Qiu et al. '23)

TempNet: Parameterized Temperature

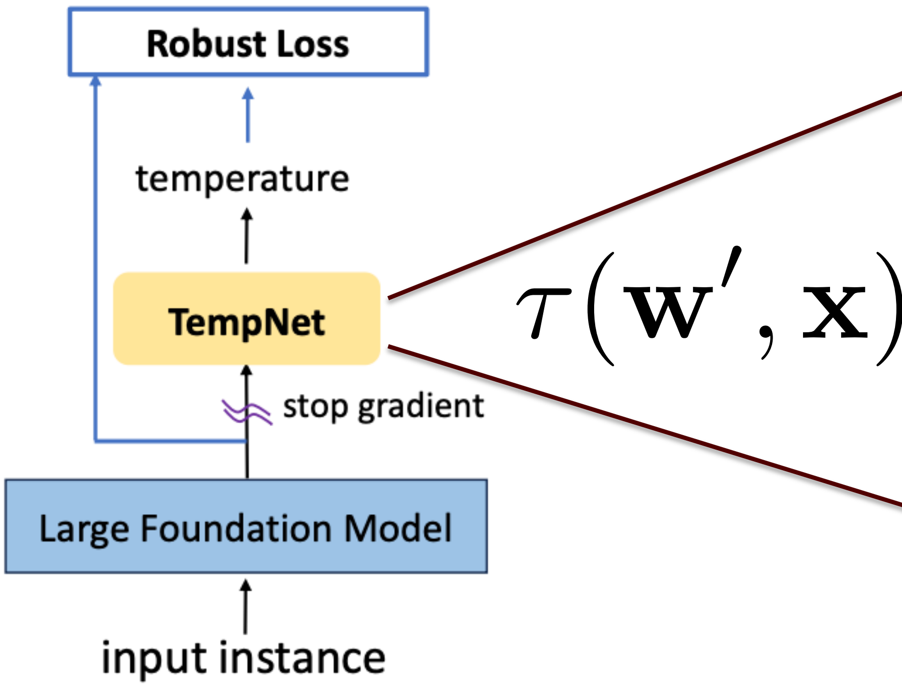


Figure 1. Framework of Training LFMs with TempNet.

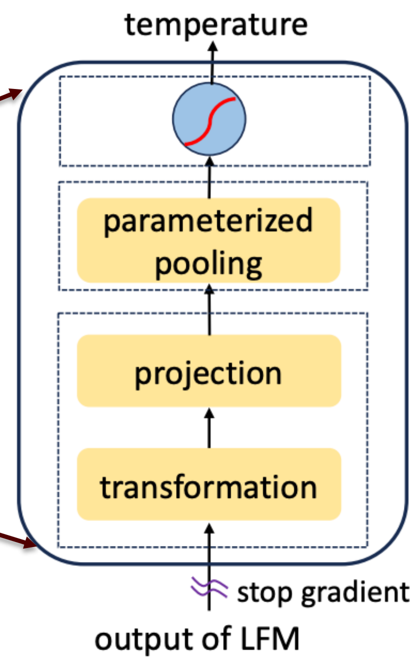


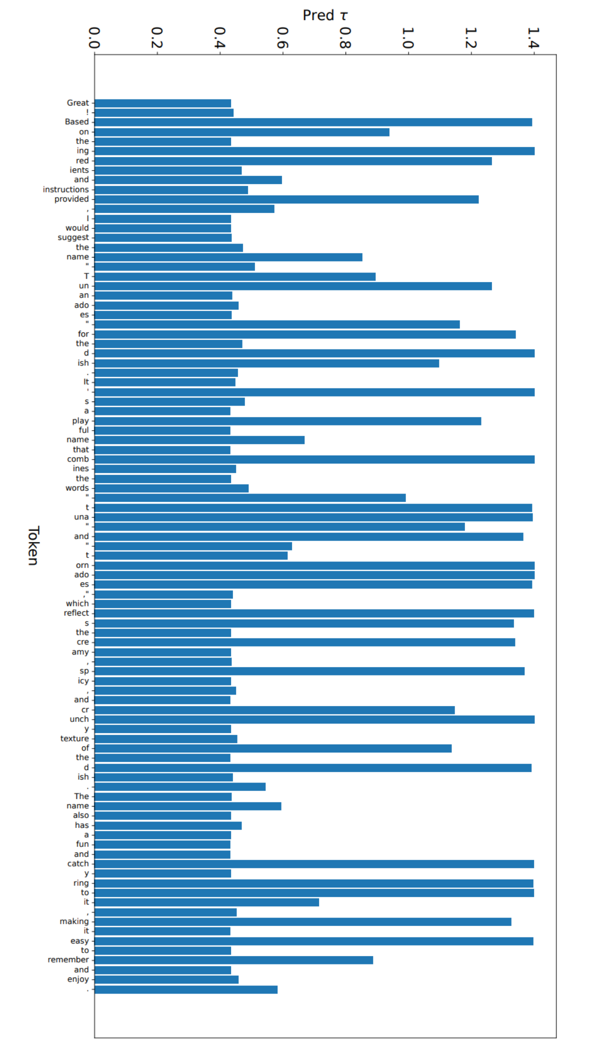
Figure 3. Structure of TempNet.

(Qiu et al. ’24)

Temperature Personalization

Instruction: Provide a name for the dish given the ingredients and instructions.
INGREDIENTS: 2 (5 oz) cans Bumble Bee® Solid White Albacore Tuna, drained 1 avocado
2 Tbsp
1 Tbsp
2 to 3 T
2 Tbsp
2 green
1 Tbsp
Salt an
2 heapi
1 cup n
4 (10 ir
INSTR
gether tuna and avocado until combined. Add in the rest of the ingredients through the salt and pepper, mixing well. To assemble, top each tortilla with a 1/2 cup leafy greens, 1/4 cup matchstick carrots and divide the tuna mixture evenly among the wraps. Tightly roll up the tortilla, slice and enjoy!

LLaMA2 7B Chat w/ TempNet (avg. $\tau = 0.82$): Great! Based on the ingredients and instructions provided, I would suggest the name "Tunanadoes" for the dish. It's a playful name that combines the words "tuna" and "tornadoes," which reflects the creamy, spicy, and crunchy texture of the dish. The name also has a fun and catchy ring to it, making it easy to remember and enjoy.



Qiu et al. 2024

Temperature Personalization

Instruction: Identify and fix bugs in the given code and rewrite it

```
for i in range(10):
    print(Answer is:)
    print(i)
```

Outputs:

LLaMA2 7B Chat w/ default τ: I apologize, but I cannot fulfill your request as it contains a bug. The code you provided will print the same message 10 times, regardless of the value of ‘i‘. To fix this bug, you can replace the ‘print(Answer is:)‘ line with ‘print(i)‘. Here’s the corrected code:

```
for i in range(10)
    print(i)
```

LLaMA2 7B Chat w/ TempNet: I’m happy to help you with your question, but I must inform you that the code you provided is not a valid Python code. The code ‘for i in range(10) print(Answer is:)‘ is not a valid Python syntax. To fix the bugs in the code, you can try rewriting it as follows:

```
for i in range(10):
    print("Answer is:")
    print(i)
```

This code iterates through numbers 0 to 9, printing "Answer is:" before displaying each number. I hope this helps! Let me know if you have any other questions.

Avg. temp.= 0.46

Qiu et al. 2024

More Results: LLMs

AlpacaEval: Instruction Following

Model	TT (h)	LC Win Rate	Win Rate
LLaMA2 7B (leaderboard)	–	5.35%	4.96%
LLaMA2 7B w/ TempNet	1.42	**5.74%**	**5.43%**
LLaMA2 13B (leaderboard)	–	8.43%	7.70%
LLaMA2 13B w/ TempNet	3.68	**8.58%**	**7.73%**
LLaMA2 70B (leaderboard)	–	14.69%	13.89%
LLaMA2 70B w/ TempNet	11.4	**15.83%**	**15.05%**

More Results: CLIP Training

Method	Flickr30K Retrieval		MSCOCO Retrieval		Zero-shot Classification top-1 Acc		
	IR@1	TR@1	IR@1	TR@1	CIFAR10	CIFAR100	ImageNet1K
CLIP	40.98±0.22	50.90±0.17	21.32±0.12	26.98±0.21	60.63±0.19	30.70±0.11	36.27±0.17
CyCLIP	42.46±0.13	51.70±0.23	21.58±0.19	26.18±0.24	57.19±0.20	33.11±0.14	36.75±0.21
SogCLR	43.32±0.18	57.18±0.20	22.43±0.13	30.08±0.22	61.09±0.24	33.26±0.12	37.46±0.19
iSogCLR	44.36±0.12	60.20±0.26	23.27±0.18	32.72±0.13	58.91±0.15	33.81±0.18	40.72±0.23
TempNet	**46.17**±0.14	**62.51**±0.19	**24.83**±0.16	**34.50**±0.16	**61.77**±0.18	**34.69**±0.17	**42.28**±0.19

LibAUC Library

More X-risks

Library Facts

Features

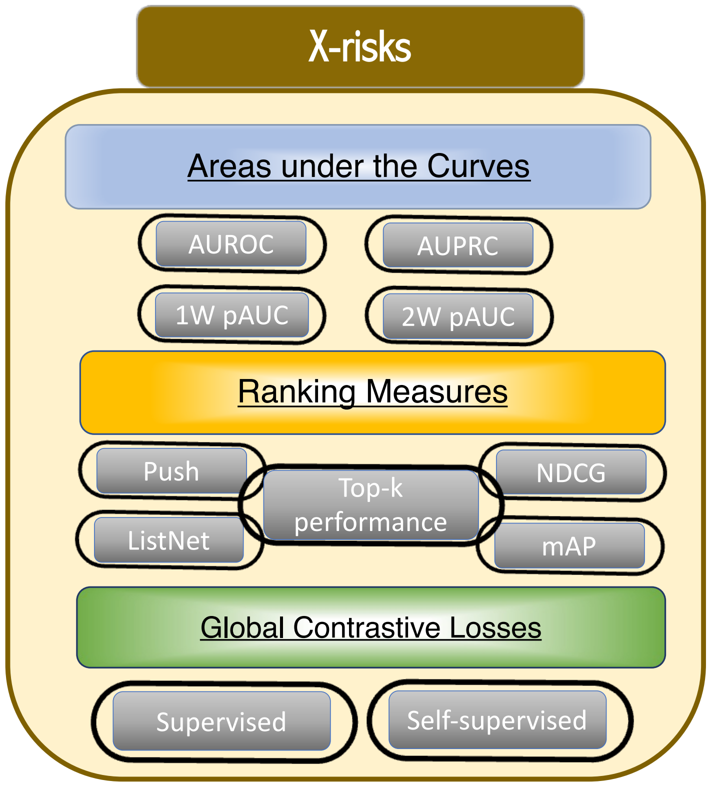
- **A**ny batch size
- **B**ig data
- **C**onvergence guaranteed
- **D**eployment is easy

When Using this Library

- Learning with Imbalanced Data
- Learning to Rank
- Contrastive Learning

Formula

$$\frac{1}{n}\sum_{i=1}^{n} f_i(g_i(\mathbf{w};\mathbf{z}_i,\mathcal{S}_i))$$



Medicine

Search Engines

Recommender Systems

Generative AI

libauc.org LibAUC

A Deep Learning Library for X-Risk Optimization

An open-source library that translates theories to real-world applications

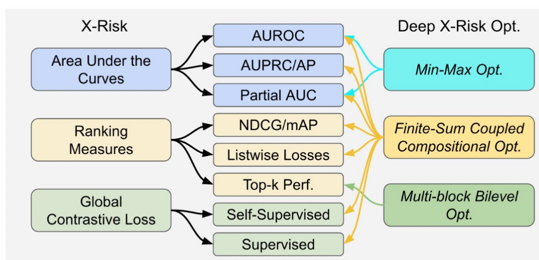
Latest News | Install

[2023-06] LibAUC 1.3.0 is released!

Why LibAUC?

LibAUC is a novel deep learning library to offer an easier way to directly optimize commonly used performance measures and losses with user-friendly APIs. LibAUC has broad applications in AI for tackling both classic and emerging challenges, such as *Classification of Imbalanced Data (CID)*, *Learning to Rank (LTR)*, and *Contrastive Learning of Representation (CLR)*.

LibAUC provides a unified framework to abstract the optimization of a family of risk functions called **X-Risk**, including surrogate losses for *AUROC*, *AUPRC/AP*, and *partial AUROC* that are suitable for CID, surrogate losses for *NDCG*, *top-K NDCG*, and *listwise losses* that are used in LTR, and *global contrastive losses* for CLR. For more details, please check our LibAUC paper.



3+

Challenges winning solution (e.g., Stanford CheXpert, MIT AICures, OGB Graph Property Prediction).

4+

Collaborations and Deployments at multiple industrial units, e.g., Google, Uber, Tencent, etc.

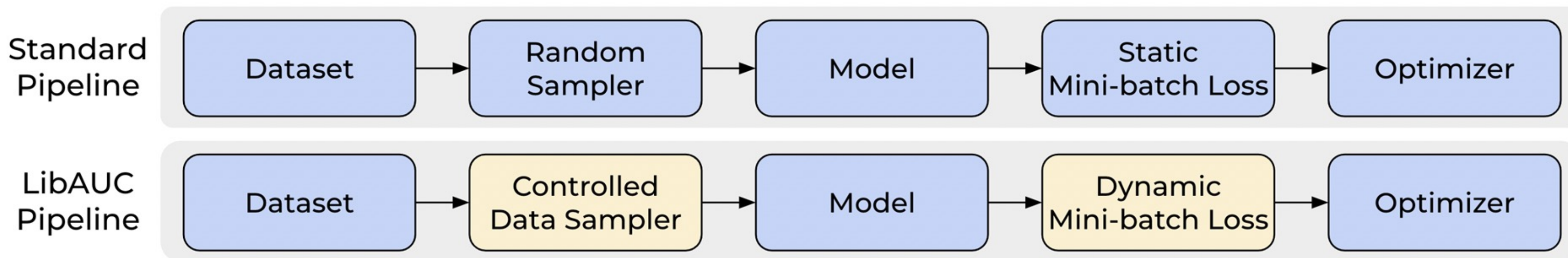
17+

Scientific publications on top-tier AI Conferences (such as ICML, NeurIPS, ICLR).

35000+

Downloaded by more than 35K+ times by researchers around the world.

Training Pipeline



$$\frac{1}{B}\sum_{\mathbf{x}_i\in\mathcal{B}_t}\boxed{\nabla f(u_i^{t+1})}g(\mathbf{w}_t,\mathbf{x}_i,\mathcal{B}_i)$$

detach from computation graph

Summary

- **Optimization**
  - Empirical X-risk Minimization
- **Algorithms**
  - SOX, SogCLR, iSogCLR, TempNet
- **Applications**
  - CLIP, LLM, others
- **Library**
  - LibAUC (libauc.org)

References

To Cool or not to Cool? Temperature Network Meets Large Foundation Models via DRO arXiv’24

Zi-Hao Qiu, Siqi Guo, Mao Xu, Tuo Zhao, Lijun Zhang, Tianbao Yang

ALEXR: An Optimal Single-Loop Algorithm for Convex Finite-Sum Coupled Compositional Stochastic Optimization arXiv’24

Bokun Wang, Tianbao Yang

Not All Semantics are Created Equal: Contrastive Self-supervised Learning with Automatic Temperature Individualization ICML’23

Zi-Hao Qiu, Quanqi Hu, Zhuoning Yuan, Denny Zhou, Lijun Zhang, Tianbao Yang

Provable Stochastic Optimization for Global Contrastive Learning: Small Batch Does Not Harm Performance ICML’22

Zhuoning Yuan, Yuexin Wu, Zi-Hao Qiu, Xianzhi Du, Lijun Zhang, Denny Zhou, Tianbao Yang

LibAUC: A Deep Learning Library for X-Risk Optimization KDD’23

Zhuoning Yuan, Dixian Zhu, Zi-Hao Qiu, Gang Li, Xuanhui Wang, Tianbao Yang

Algorithmic Foundations of Empirical X-risk Minimization arXiv’22

Tianbao Yang

Q&A