



Carnegie
Mellon
University



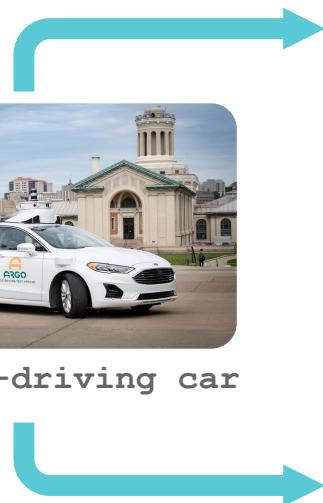
CLEAR

Challenge of Continual LEArning on Real-World Imagery

CVPR 2022 VPLOW Workshop Challenge Track

Organizers: Zhiqiu Lin, Siqi Zeng, Jia Shi, Shihao Shen

Visual perception systems need to cope with **changing environments..**



Pittsburgh

New cities?



Miami



Domino's car (2013)

New car models?



Domino's car (2023?)

But vision benchmarks **stay the same over time..**

ImageNet (2010)



same as in 2010



COCO (2015)



same as in 2015



2010

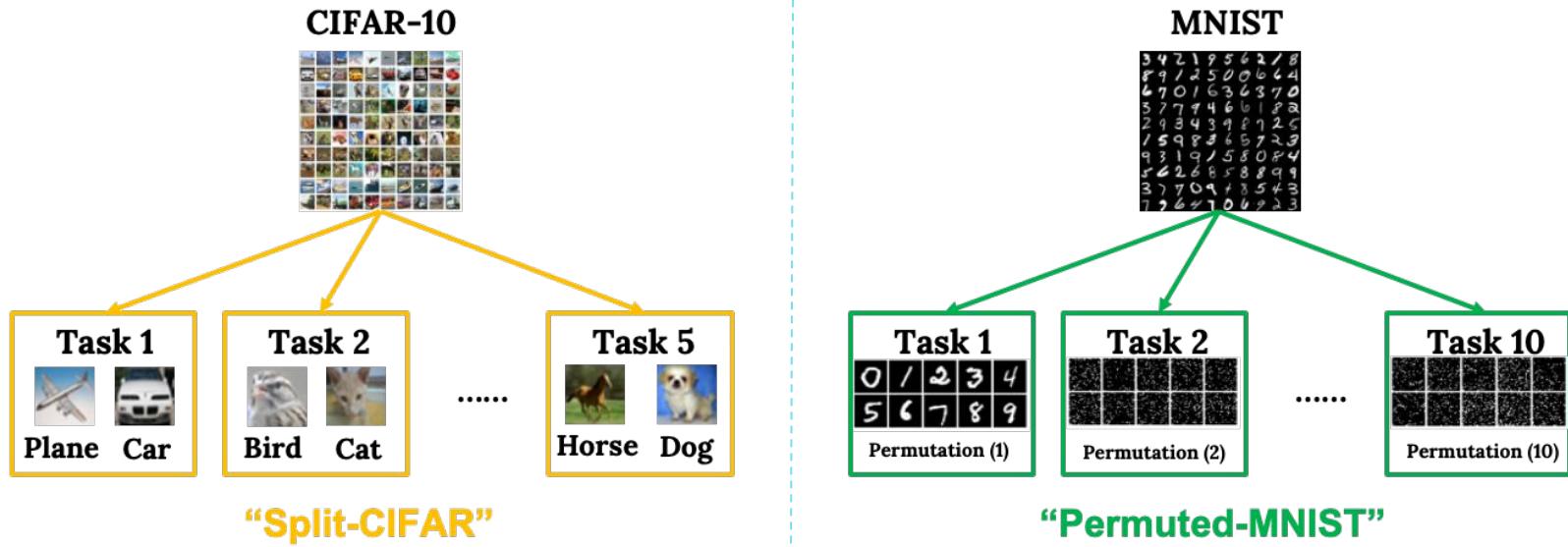
2014

2018

2022



Prior works simulates changing environments via **continual/lifelong learning** benchmarks



Issue: Extreme distributions shifts between tasks..

Real-world distributions shifts are **smooth**, such as computer make and models.

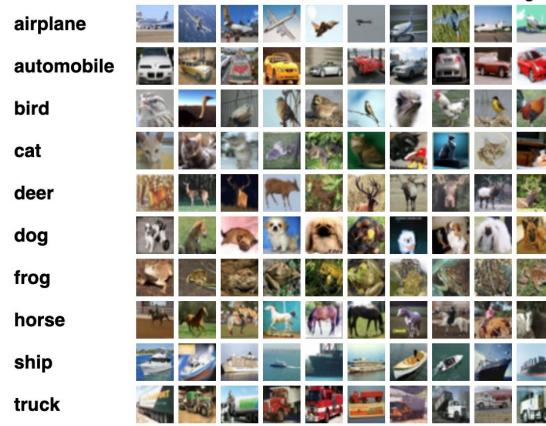


Idea: To collect a benchmark with natural distribution shifts!



CLEAR: Continual LEArning with Real-world Imagery

→ First CL benchmark for open-world vision



[1]

Superclass

- aquatic mammals
- fish
- flowers
- food containers
- fruit and vegetables
- household electrical devices
- household furniture
- insects
- large carnivores
- large man-made outdoor things
- large natural outdoor scenes
- large omnivores and herbivores
- medium-sized mammals
- non-insect invertebrates
- people
- reptiles
- small mammals
- trees
- vehicles 1
- vehicles 2

Classes

- beaver, dolphin, otter, seal, whale
- aquarium fish, flatfish, ray, shark, trout
- orchids, poppies, roses, sunflowers, tulips
- bottles, bowls, cans, cups, plates
- apples, mushrooms, oranges, pears, sweet peppers
- clock, computer keyboard, lamp, telephone, television
- bed, chair, couch, table, wardrobe
- bee, beetle, butterfly, caterpillar, cockroach
- bear, leopard, lion, tiger, wolf
- bridge, castle, house, road, skyscraper
- cloud, forest, mountain, plain, sea
- camel, cattle, chimpanzee, elephant, kangaroo
- fox, porcupine, possum, raccoon, skunk
- crab, lobster, snail, spider, worm
- baby, boy, girl, man, woman
- crocodile, dinosaur, lizard, snake, turtle
- hamster, mouse, rabbit, shrew, squirrel
- maple, oak, palm, pine, willow
- bicycle, bus, motorcycle, pickup truck, train
- lawn-mower, rocket, streetcar, tank, tractor

[1]

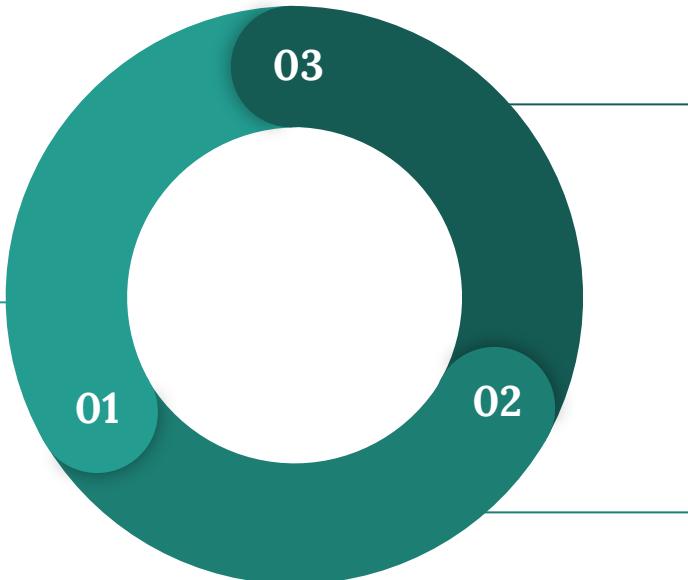
CIFAR10 (2009)**CIFAR100 (2009)**

How about CLEAR10 / CLEAR100
for Real-World Continual Learning?

Highlights

Natural Distribution Shift Over A Decade

CLEAR captures real distribution shifts of Internet images from 2004 to 2014 in YFCC100M.



Assets For Future CL Research

Unlabeled data

→ continual unsupervised learning

Metadata

→ continual multimodal learning

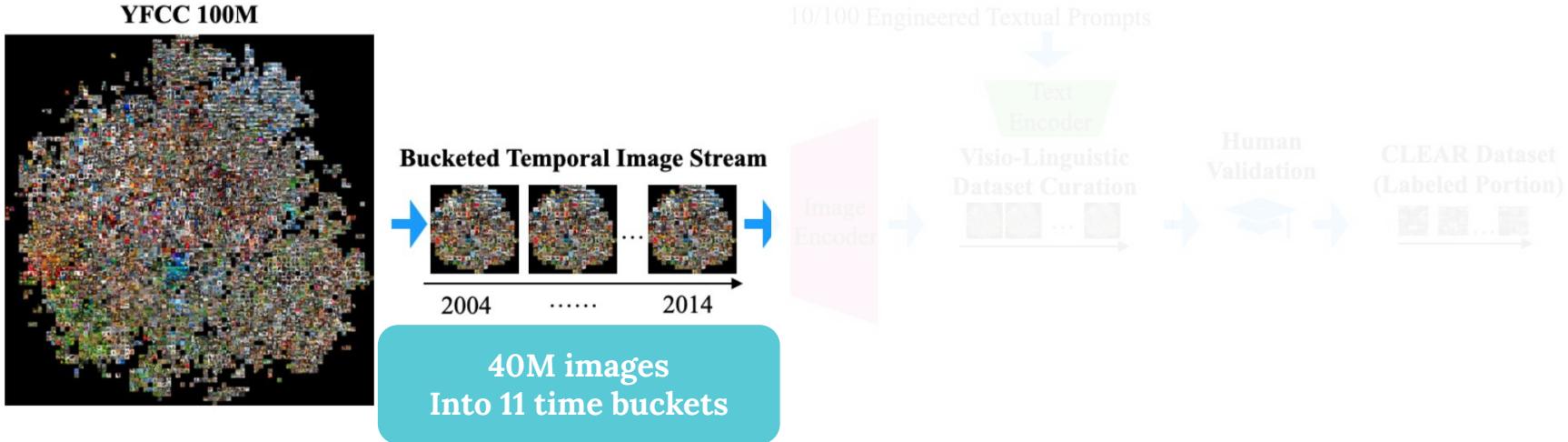
Instruction set

→ dataset curation/transparency

Efficient & Faithful Dataset Curation

To avoid working with massive data in YFCC, we create an efficient semi-automated visio-linguistic dataset curation pipeline followed by human verification.

We start from **Flickr YFCC100M** with **timestamped images from 2004 to 2014**.



We split the temporal image stream into 11 buckets:

- 0th bucket reserved for unsupervised pretraining
- 1st - 10th buckets with annotation for continual classification

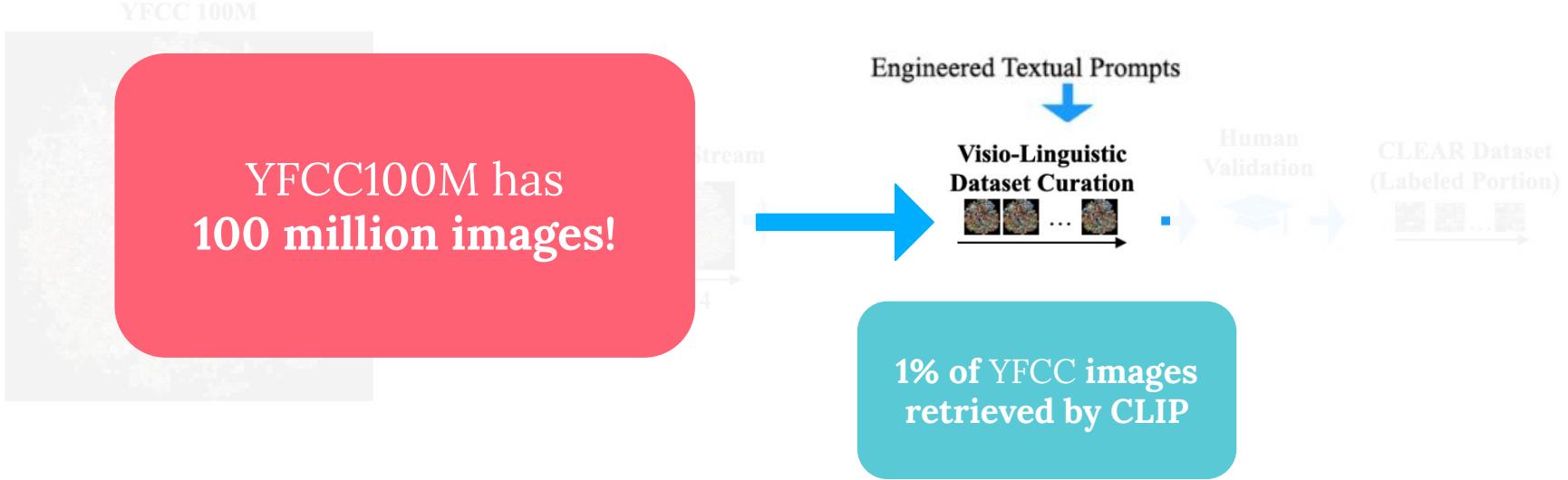
Visual Concepts in CLEAR10 and CLEAR100

bus camera
computer
CLEAR10
dress racing
pullover soccer cosplay
baseball hockey

watch gloves glasses violin piano
ring necklace backpack graffiti statue fountain bookstore observatory
scarf tie hat anime guitar billboard stadium temple
laptop camera beer ice_cream bridge lab bathroom castle opera_house
microphone chocolate lamppost road_sign gym
golf tennis canned_food
skateboarding horse_riding
ice_skating roller_skating swimming firefighter shopping_mall
field_hockey basketball volleyball policeman casino
surfing ice_hockey baseball chef laundry
billiard bowling diving bus coser soldier
football soccer subway helicopter
table_tennis skiing train airplane ferry
racing_car tractor bicycle
food_truck blackboard
umbrella plush_toys power_plant
lego mug vase vending_machine
pet_store garage robot

CLEAR100

We propose a **visio-linguistic approach** utilizing OpenAI's pretrained CLIP model to automatically retrieve images of particular visual concepts.



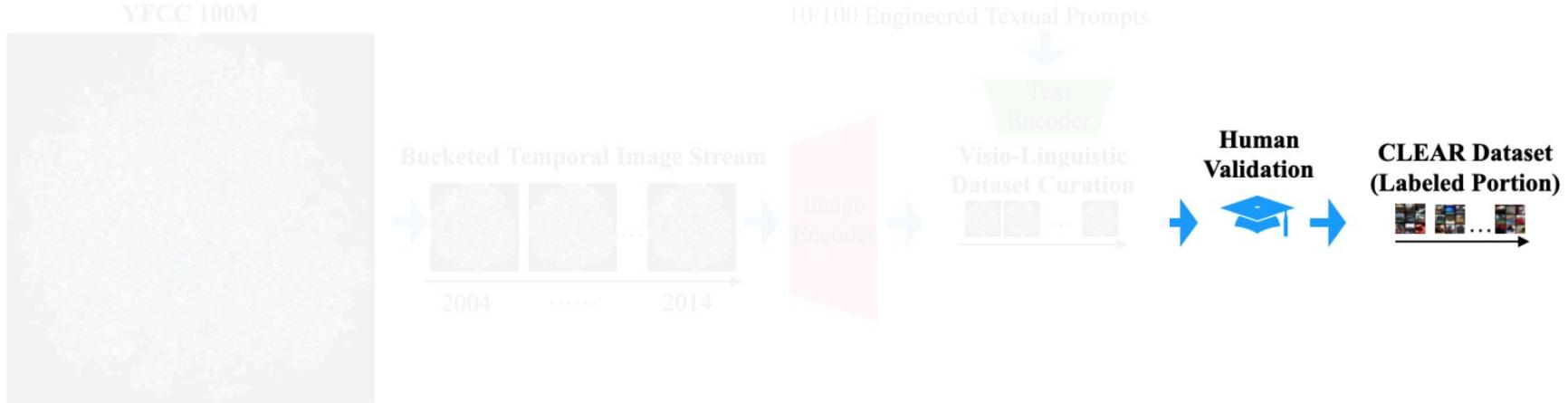
A Snapshot of CLEAR

Natural Temporal Evolution

The figure displays a horizontal timeline from 2004 to 2014, with a green arrow pointing from left to right at the bottom. Above the timeline, there are three rows of images corresponding to different CLIP prompts. The first row shows images related to 'laptop, computer'. The second row shows images related to 'a photo of watch'. The third row shows images related to 'a screenshot of video game'. Each row has 11 images, indexed from 0 to 10. The images illustrate how the visual concept for each prompt has changed over time, with the most recent images showing more complex and diverse examples.

CLIP Prompt	Visual Concept	0	1	2	3	4	5	6	7	8	9	10	Bucket
laptop, computer	computer												2004
a photo of watch	watch												2009
a screenshot of video game	video game												2012
													2014

CLIP generated labels are **verified by human** to ensure the label quality.



- crowd-sourced & professional labeling service for human validation
- high-quality labels!

Data Statistics

CLEAR

CLEAR10 (Labeled set)
~3.5K images x 10 buckets (1st - 10th)

Train
~3K images x 10 buckets

Test
~0.5K images x 10 buckets

CLEAR10 (Unlabeled set)
~0.8M images x 11 buckets (0th - 10th)

CLEAR100 (Labeled set)
~15K images x 10 buckets (1st - 10th)

Train
~10K images x 10 buckets

Test
~5K images x 10 buckets

CLEAR100 (Unlabeled set)
~3.6M images x 11 buckets (0th - 10th)

Assets for Future CL Research



Abundant Unlabeled Images

→ unsupervised continual learning

Metadata

Time/Location/Social Media Hashtag/Text Description/...

→ multimodal learning

4. Problems found during labeling

- (1) The definition for Places is not c observatory, temple, garage, power classes. All high buildings are define confusing during verification.
- ...
- (3) In fundamental rules, statue ima waterbodies. In this case, any statu bronze statue, or the Statue of Libe

2. Extra Label Policy:

- (1) If words exist in the picture, in general choose Y. If there is a sign saying "NO/Stop ..(class related keywords)" then select N.
- (2) If a non-lego class image is a toy or a model, choose Y, but it can't be a lego.
- (3) For classes except video game and anime, cartoon style object is N.
- (4) Drawings of an object is N in general, except for some extremely realistic images.

Labeling Policy

contains computer screen, and/or mouse, and/or keyboard
Lens and body of camera, or people using camera
skinny cylinder, might have foam around top, people usi
if not in its original package, yellow liquid with foam and
dark brown, brown, white chocolate bar. Packaged choc

Toy Piano Y
Milk N

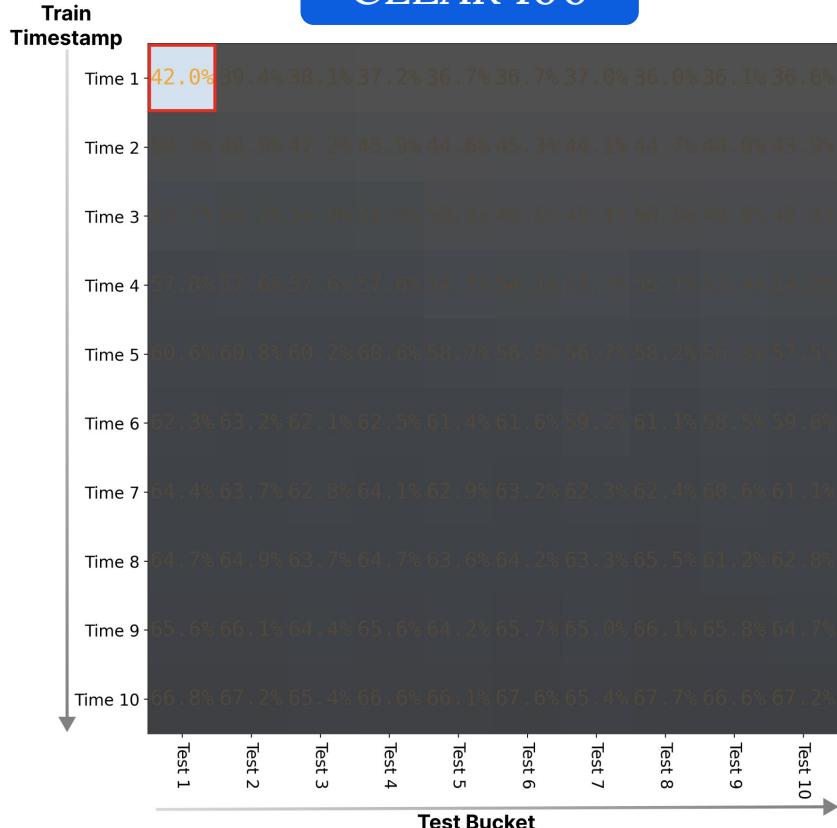
200+ Pages of Instruction Set & Corner Cases

→ dataset curation/transparency



→ Simulating Real-World Continual Learning

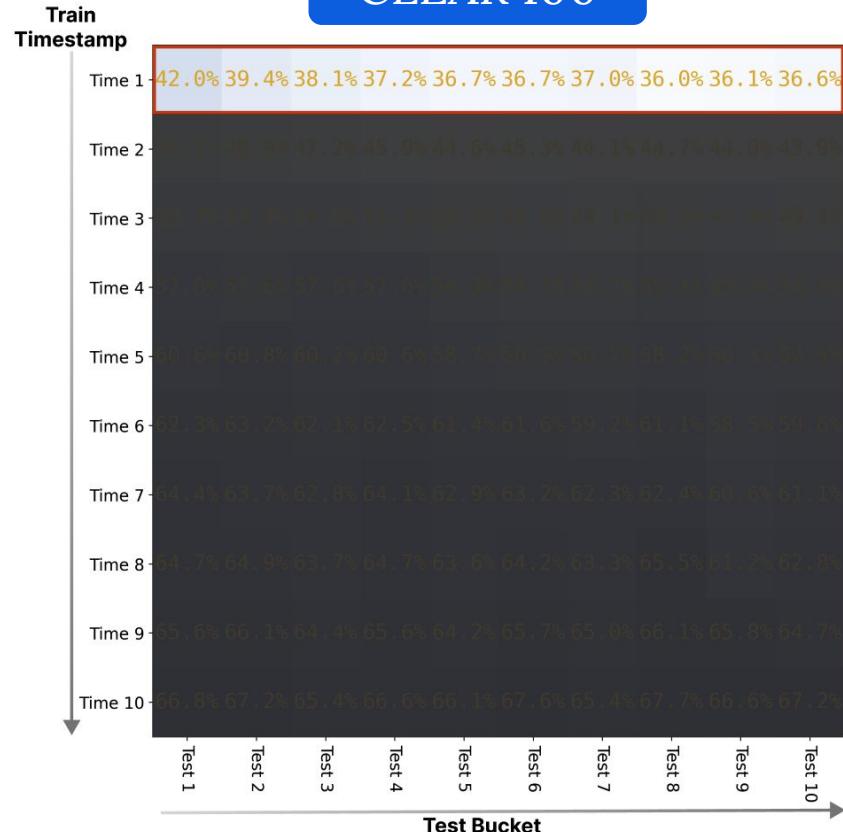
CLEAR 100



Train on 1st,
test on 1st
Acc = 42.0%

Standard classification model (ResNet18) can achieve reasonable test accuracy on 1st bucket..

CLEAR 100



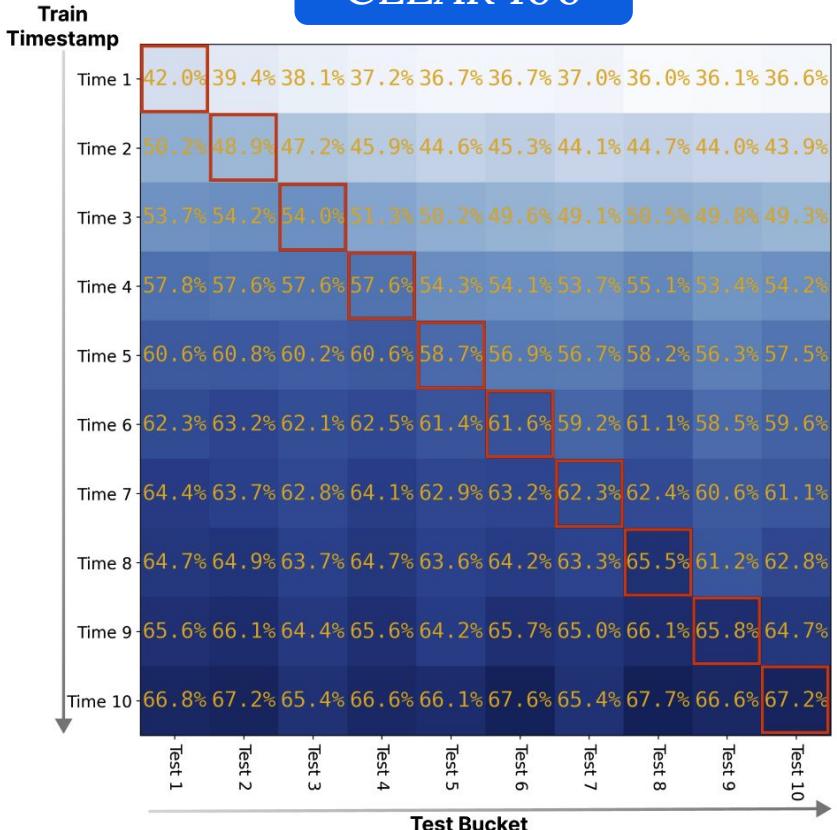
Train on 1st,
test on 1st
Acc = 42.0%

Train on 1st,
test on 2nd
Acc = 39.4%

.....
Train on 1st,
test on 10th
Acc = 36.6%

Without continual learning, performance suffers by **5.4%** (from 42.0% to 36.6%) over time..

CLEAR 100



Train on 1st,
test on 1st
Acc = 42.0%

Train on 1st,
test on 2nd
Acc = 39.4%

Train on 1st,
test on 10th
Acc = 36.6%

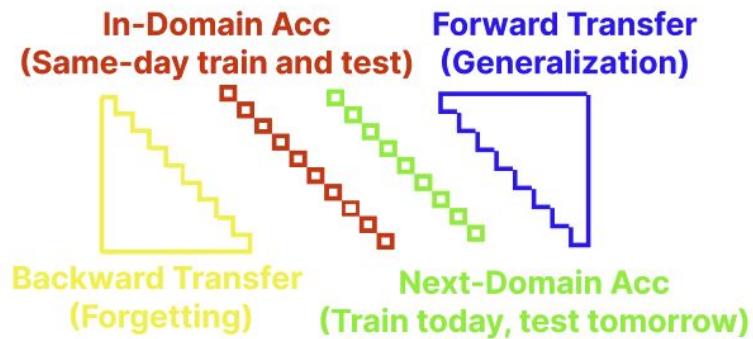
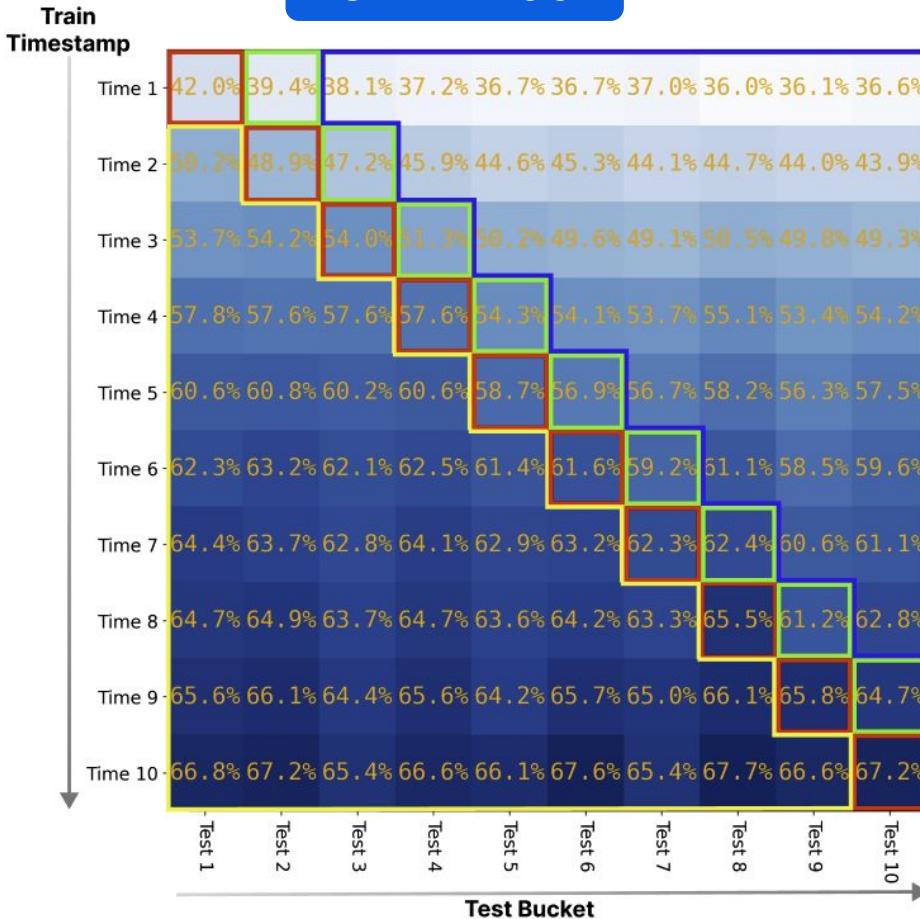
Train on [1+2],
test on 2nd
Acc = 48.9%

Train on
[1:10],
test on 10th
Acc = 67.2%

Continual learning helps – simply “finetuning” on accumulated data boosts on average **20%** accuracy!

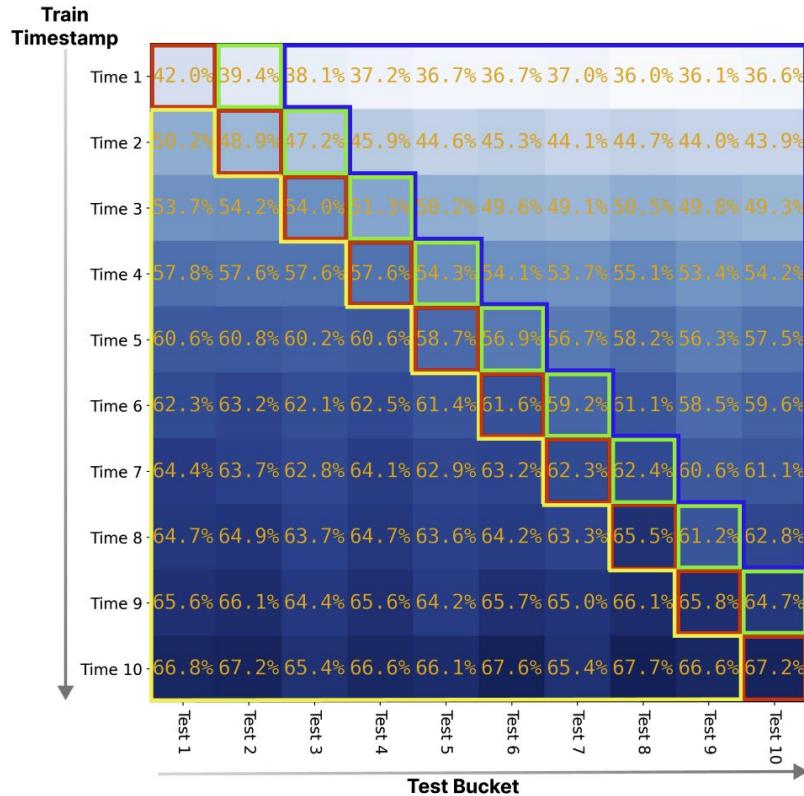
Metrics to quantify CL performances..

CLEAR 100



Next-Domain Acc is more realistic than **In-Domain Acc** due to time delay between **data arrival** and **model deployment**

CLEAR 100



Backward Transfer = 63.1%
(Forgetting)



In-Domain Acc = 58.4%
(Same-day train and test)



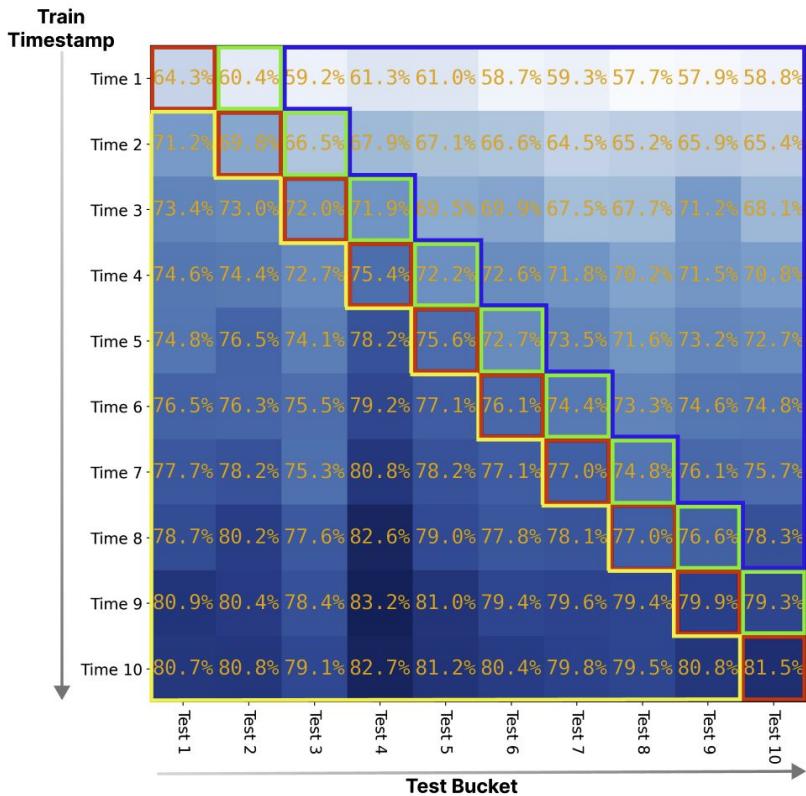
Next-Domain Acc = 55.2%
(Train today, test tomorrow)



Forward Transfer = 50.3%
(Generalization)

Next-Domain Acc / **Forward Transfer** are more challenging than **In-Domain Acc** / **Backward Transfer**, leaving large room for improvement.

CLEAR 10



Backward Transfer = 78.9%
(Forgetting)



In-Domain Acc = 74.9%
(Same-day train and test)



Next-Domain Acc = 72.1%
(Train today, test tomorrow)



Forward Transfer = 68.9%
(Generalization)

Same trends hold for
CLEAR10!

Though **CLEAR10**
performance is on
average **15%** higher than
CLEAR100 as it is a
simpler task.

CLEAR 10



Method	Evaluation Metrics			
	In-domain Acc	Next-domain Acc	Backward Transfer	Forward Transfer
Continual Finetuning	74.9% \pm .3%	72.1% \pm .2%	78.1% \pm .2%	68.9% \pm .1%
EWC (Elastic Weight Consolidation)	76.6% \pm .2%	74.3% \pm .6%	76.5% \pm .4%	71.1% \pm .6%
SI (Synaptic Intelligence)	76.0% \pm .2%	73.6% \pm .2%	76.0% \pm .5%	71.0% \pm .4%
LwF (Learning w/o Forgetting)	77.8% \pm .3%	75.7% \pm .3%	79.6% \pm .3%	72.5% \pm .3%
CWR	69.5% \pm .2%	67.8% \pm .3%	68.8% \pm .3%	66.6% \pm .3%
GDumb	66.0% \pm .4%	64.3% \pm .5%	68.9% \pm .4%	61.4% \pm .5%
ER (Experience Replay)	77.3% \pm .1%	75.6% \pm .3%	79.3% \pm .1%	72.4% \pm .2%
A-GEM (Gradient Episodic Memory)	76.2% \pm .3%	73.6% \pm .2%	75.8% \pm .2%	70.2% \pm .2%

Similar Performances!

Classic CL algorithms (Avalanche-based implementation), originally designed to combat forgetting, perform only marginally better or about the same as simple continual finetuning on CLEAR Benchmark.

CLEAR is now publicly available on Avalanche (a snapshot of the API)



Benchmarks based on the [CLEAR](#) dataset.

`CLEAR (*[, data_name, evaluation_protocol, ...])`

Creates a Domain-Incremental benchmark for CLEAR 10 & 100 with 10 & 100 illustrative classes and an n+1

Benchmarks for learning from pretrained models or multi-agent continual learning scenarios. Based on the [Ex-Model paper](#). Pretrained models are downloaded automatically.

`ExMLMNIST ([scenario, run_id])`

ExML scenario on MNIST data.

`ExMLCoRE50 ([scenario, run_id])`

ExML scenario on CoRE50.

`ExMLCIFAR10 ([scenario, run_id])`

ExML scenario on CIFAR10.

Datasets

The `datasets` sub-module provides PyTorch dataset implementations for datasets missing from the `torchvision/audio/*` libraries. These datasets can also be used in a standalone way!

`CORE50Dataset (root, ~pathlib.Path) = None, *`

CORe50 Pytorch Dataset

`CUB200 (root, ~pathlib.Path) = None, *[, ...]`

Basic CUB200 PathsDataset to be used as a standard PyTorch Dataset.

Try it out!



→ Summary of CVPR 2022 Challenge

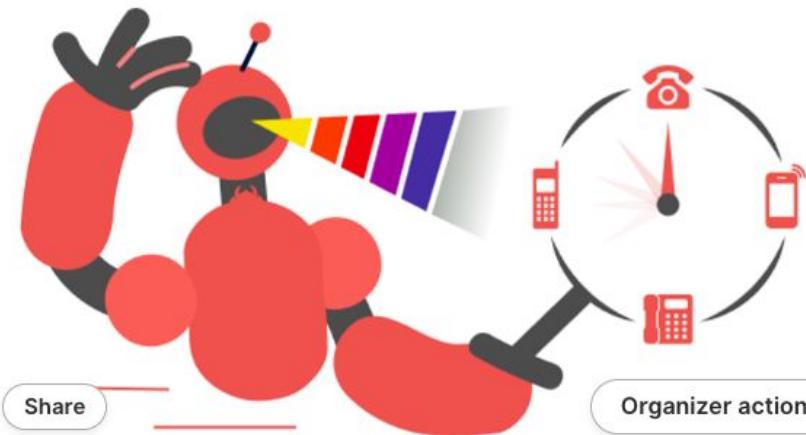
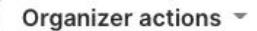


CVPR 2022 CLEAR Challenge

CVPR 2022 Workshop Challenge on CLEAR:
Continual LEArning on Real-world Imagery



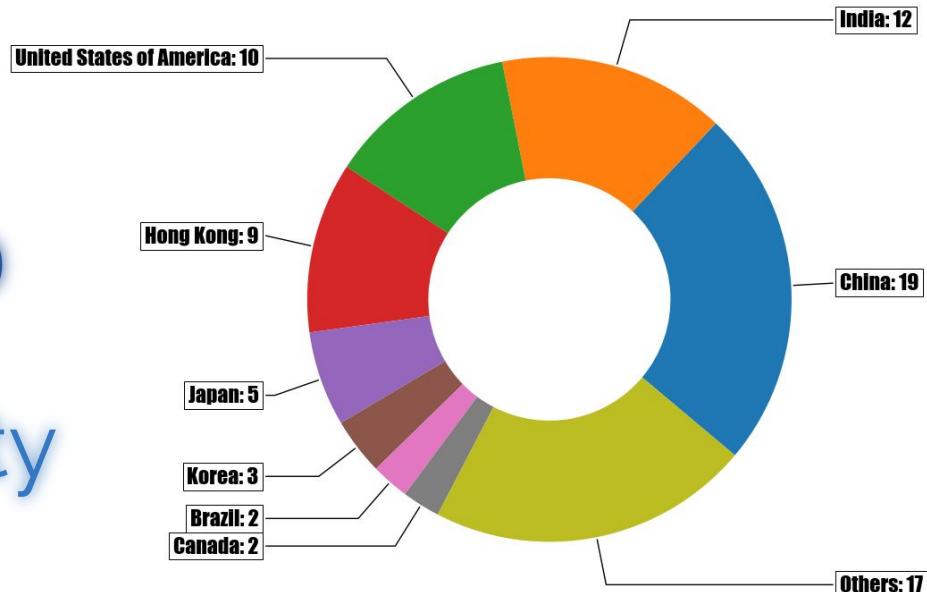
By Carnegie Mellon University



39 Days, 79 Participants, 15 Teams, 547 Submissions

Composition

IIT Kharagpur IIIT Hyderabad
Freelancer
Federal University of Pernambuco James Cook University
AI Prime UNIST
Tencent YouTu Lab
Carnegie Mellon University
Shanghai Jiao Tong University
Tsinghua University
Southern University of Science and Technology
BOE Information Technology University of the Punjab
Beihang University Zhejiang University



CLEAR10 Leaderboard

△	#	Participants	Weighted Average Score	Next-Domain	In-Domain	BwT	FwT
●	01	shennong3  	0.905	0.901	0.912	0.923	0.885
▲	02	BOE_AIoT_CTO   	0.895	0.891	0.904	0.915	0.871
▼	03	AI_PRIME   	0.889	0.885	0.896	0.911	0.864
●	04	Lge 	0.867	0.859	0.879	0.895	0.833
●	05	unist-milil   	0.728	0.711	0.748	0.781	0.670
●	06	try 	0.677	0.654	0.705	0.728	0.619
‡		Baseline clear10_naive_streaming_resnet18, script at https://github.com/ContinualAI/avalanche/blob/master/examples/clear.py	0.663	0.649	0.698	0.685	0.618
●	07	chen_sun 	0.644	0.630	0.680	0.671	0.593

>20% Jump from baseline

CLEAR100 Leaderboard

Δ	#	Participants	Weighted Average Score	Next-Domain Accuracy	In-Domain Accuracy	Backward Transfer	Forward Transfer
●	01	shennong3 	0.9146	0.9125	0.9199	0.9340	0.8920
●	02	AI_PRIME 	0.9124	0.9077	0.9178	0.9379	0.8863
▲	03	BOE_AIoT_CTO 	0.8873	0.8829	0.8960	0.9074	0.8630
▼	04	Lge 	0.8606	0.8536	0.8696	0.8965	0.8229
●	05	unist-mill 	0.6216	0.6078	0.6329	0.6890	0.5568
●	06	chen_sun 	0.5455	0.5363	0.5704	0.5689	0.5065
✖		Baseline Baseline clear100_naive_streaming_resnet18, script at https://github.com/ContinualAI/avalanche/blob/master/examples/clear.py	0.4935	0.4810	0.5220	0.5342	0.4367



>40% Jump

The Most Promising Strategies on CLEAR

- Experience Replay to utilize both current and previous buckets' data
- Strong Data Augmentation (e.g. AutoAug, CutMix, Mixup)
- Enhancing Generalization via
 - Sharpness Aware Minimization
 - Supervised Contrastive Loss
 - Unsupervised Domain Generalization
 - Meta Learning
 - Larger Backbone for Over-Parameterization

Winners



1st Place -- \$1000

Xinkai Guo, Bo Ke, Sunan He, Ruizhi Qiao
Tencent, YouTu Lab

"Bucket-Aware Sampling Strategy for Efficient Replay"



2nd Place -- \$300

Jiawei Dong, Mengwen Du, Shuo Wang
AI Prime

"Comprehensive Studies on Sampling, Architecture and Augmentation Strategies"



3rd Place -- \$100

Xiaojun Tang, Pan Zhong, Tingting Wang, Yuzhou Peng
BOE Technology Group

"Adaptive Loss for Better Model Generalization in Real World"



4th Place -- \$100

Ge Liu
Shanghai Jiao Tong University

"Improving Model Generalization by Contrasting Features across Domains"



Innovation Prize

Solang Kim, Jin Hyuk Lim, Sung Whan Yoon
Ulsan National Institute of Science and Technology

"Domain Generalization & Meta Learning for Robustness against Distribution Shifts"



→ Invited Team Presentation: 1st Place



Bucket-Aware Sampling Strategy for Efficient Replay

In Workshop Visual Perception and learning in an Open World at CVPR2022

Team: shennong3

Members: Xinkai Gao, Bo Ke, Sunan He, Ruizhi Qiao

Affiliation: Tencent Youtu Lab



→ Lessons Learned & Future Directions

Lessons we learned in this competition



Lesson 1: Sampling matters for efficient learning

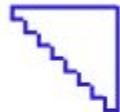


Lesson 2: Augmentation improves generalization in CL



Lesson 3: Generalization is the bottleneck for real-world CL

Future Step: Generalization Bottleneck for Real-World CL



**Forward Transfer = 89%
(Generalization)**



**Next-Domain Acc = 91%
(Train today, test tomorrow)**



**Backward Transfer = 93%
(Forgetting)**

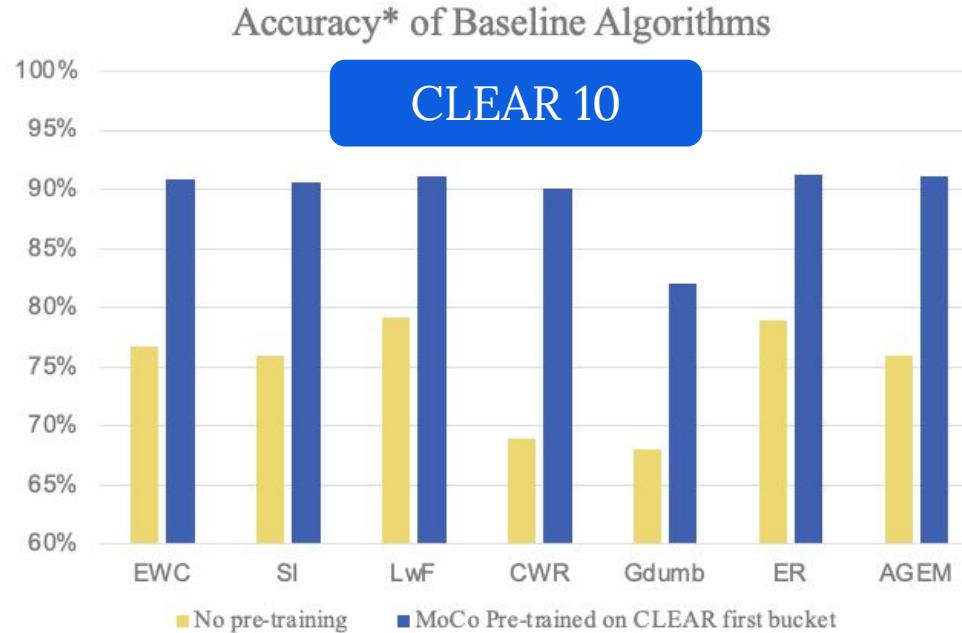


**In-Domain Acc = 92%
(Same-day train and test)**

Next-Domain Acc / Forward Transfer are more challenging than **In-Domain Acc / Backward Transfer**, suggesting the **generalization bottleneck** for real-world CL.

Domain generalization/domain adaptation/meta learning could be promising research directions.

Future Direction: Continual Unsupervised Learning



We use an unsupervised MoCo V2 model pretrained on **CLEAR's 0th bucket** of unlabeled data, and this simple pre-training steps boosts on average **15%** for all baseline methods.

It could be promising to perform **continual unsupervised learning**, using the unlabeled data of 1st-10th buckets.

Future Direction: ImageNet-scale Real-world CL Benchmark



for real-world CL?

We are trying to expand CLEAR to an ImageNet-scale benchmark!

Stay tuned!



Thank You!

Carnegie Mellon University
School of Computer Science

Alcrowd A red cartoon-style devil icon with horns and a mischievous expression.

