



CVPR 2024 Workshop on

Representation Learning with Very Limited Images

Zero-shot, Unsupervised, and Synthetic Learning in the Era of Big Models

LIMIT Organization Team

Organizers in CVPR 2024 LIMIT Workshop



Hirokatsu Kataoka
AIST



Yuki M. Asano
University of
Amsterdam



Christian Rupprecht
University of Oxford



Rio Yokota
Tokyo Tech/AIST



Nakamasa Inoue
Tokyo Tech/AIST



Dan Hendrycks
Center for AI Safety



Xavier Boix
Fujitsu Research



Manel Baradad
MIT



Connor Anderson
BYU



Ryo Nakamura
Fukuoka University/
AIST



Ryosuke Yamada
University of Tsukuba/
AIST



Risa Shinoda
Kyoto University/AIST



Ryu Tadokoro
Tohoku University/
AIST



Erika Mori
Keio University/AIST



Program committee

Akira	Sakai	Fujitsu Laboratories Ltd.
Alina E	Marcu	University Politehnica of Bucharest
Anestis	Kastellos	ITI/CERTH
Athanasios	Psaltis	ITI-CERTH, University of West Attica
Atsushi	Hashimoto	OMRON SINIC X Corp.
Atsushi	Kanehira	Microsoft
Balu	Adsumilli	YouTube/Google
Chia-Hsien	Shih	University of Illinois at Urbana-Champaign
Daiki	Ikami	Tokyo University of Agriculture and Technology
Debaditya	Roy	Agency for Science, Technology and Research, A*STAR, Singapore
Dragos	Costea	University "Politehnica" of Bucharest
Emil	Slusanschi	University "Politehnica" of Bucharest
Fabio	Ferreira	University of Freiburg
Hankook	Lee	Sungkyunkwan University (SKKU)
Hannah	Pinson	Eindhoven University of Technology
Hoàng-Ân	Lê	IRISA, University of South Brittany
Hojin	Jang	Korea University
Hyung-Il	Kim	ETRI
Ikuro	Sato	Tokyo Institute of Technology / Denso IT Laboratory
Jiadong	Chen	sensetime
Jingjing	Chen	Fudan University
Jose A	Sosa Martinez	University of Luxembourg

Many thanks!



cvpaper.challenge

Program committee

Kensho	Hara	National Institute of Advanced Industrial Science and Technology (AIST)
Kota	Matsui	Nagoya University
Kuniaki	Saito	OMRON SINIC X
Kyungjune	Baek	TwelveLabs
Lile	Cai	Institute for Infocomm Research
Marek	Wodzinski	AGH UST
Matthieu	Armando	Naver Labs
Mayu	Otani	CyberAgent
Mihai Cristian	Pîrvu	University "Politehnica" of Bucharest
Minho	Park	ETRI
Moyuru	Yamada	Fujitsu Limited
Naoshi	Kaneko	Tokyo Denki University
Naoya	Chiba	Tohoku University
Nikolas	Ebert	HS Mannheim
Nishant	Rai	Stanford University
Pavan C	Madhusudana	Samsung Research America
Pengwan	Yang	University of Amsterdam
Qianru	Sun	National University of Singapore
Renaud	Vandeghen	University of Liège
Ryosuke	Furuta	The University of Tokyo

Many thanks!



Program committee

Said	Ladjal	Télécom Paris
Sangbeom	Lim	Korea University
Sanjana	Srivastava	Stanford University
Takuma	Yagi	National Institute of Advanced Industrial Science and Technology (AIST)
Tommi	Kerola	Preferred Networks, Inc.
Tomoyuki	Suzuki	CyberAgent, Inc.
Vikash	Sehwag	Sony AI
Xiang	Wang	SenseTime Research
Youssef	Dawoud	Friedrich-Alexander-Universität Erlangen-Nürnberg
Youwei	Liang	UC San Diego
Yusuke	Mukuta	The University of Tokyo
Ziad	Al-Halah	University of Utah

Many thanks!



cvpaper.challenge

Today's schedule

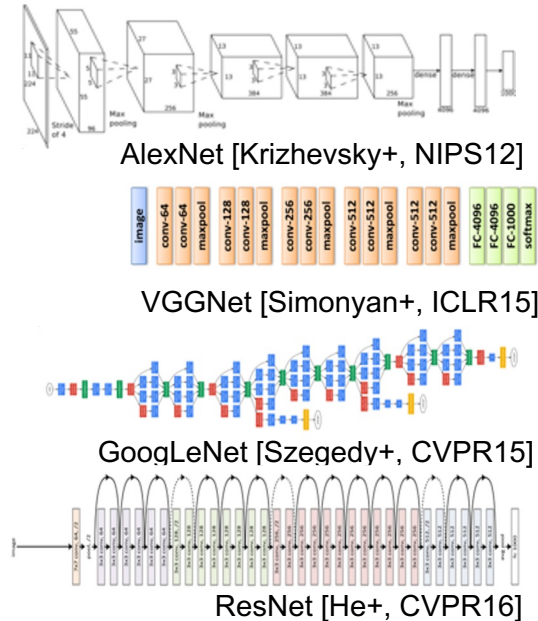
- 13:30 - 13:40: Opening remarks (Talk: Hirokatsu Kataoka)
 - 13:40 - 14:20: Invited Talk 1 (Talk: Phillip Isola, Chair: Yuki M. Asano)
 - 14:20 - 15:00: Oral Session (Chair: Ryosuke Yamada)
 - 15:00 - 15:40: Coffee Break
 - 15:40 - 16:20: Invited Talk 2 (Talk: Zeynep Akata, Chair: Christian Rupprecht)
 - 16:20 - 16:30: Short Break
 - 16:30 - 17:30: Poster Session (Room: Arch Building 4E, Chair: Erika Mori)
- * In-person only

Paper statistics

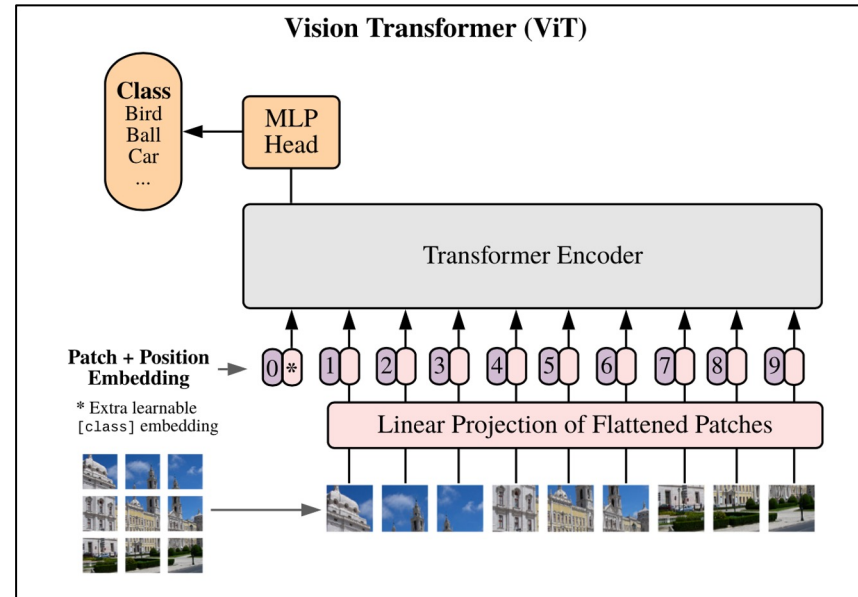
- ❑ 15 valid submissions
- ❑ 10 accepts (66.6% acceptance rate)
- ❑ 4 orals
- ❑ 15 posters (including +5 invited posters)

Scaling for models and datasets

Models



Vision Transformer (ViT)



[Dosovitskiy+, ICLR21]

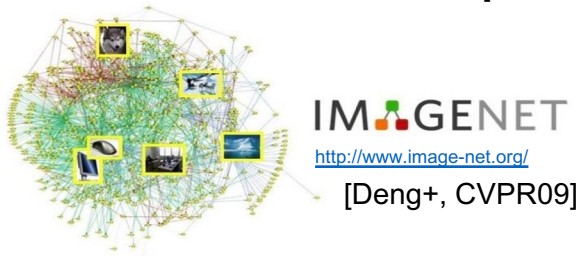
[Zhai+, CVPR22] [Dehghani+, ICML23]

ViT Models	Params
22B	22B
Giant	2B
Huge	632M
Large	307M
Base	86M
Small	22M
Tiny	5M

Datasets

JFT-300M/3B/4B

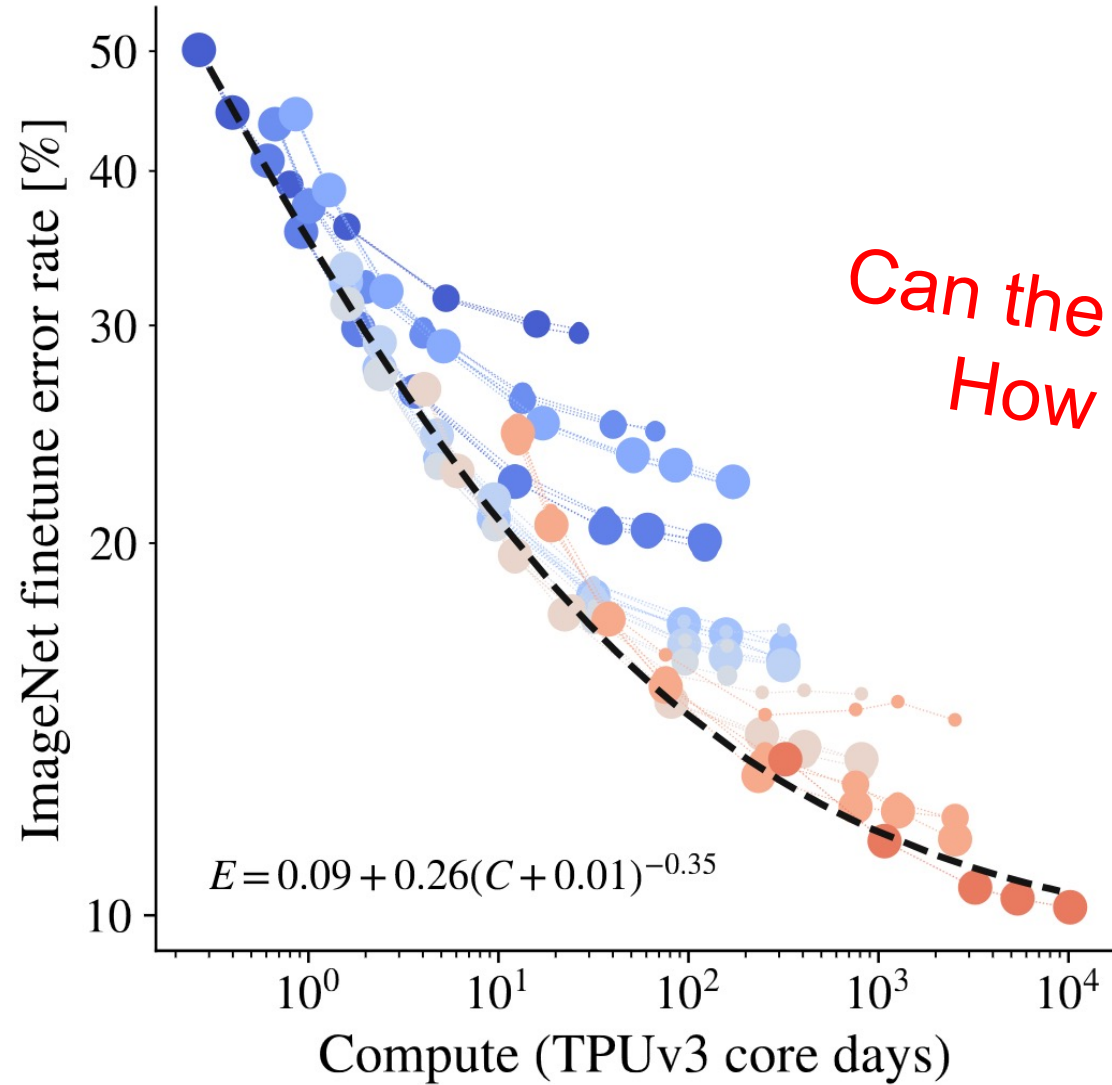
[Sun+, ICCV17][Zhai+, CVPR22][Dehghani+, ICML23]



Large size is
'JUSTICE'

Scaling laws in 'vision' transformers

Getting larger, getting better!



*Can the game only be played at specific labs?
How do academic conduct a research?*

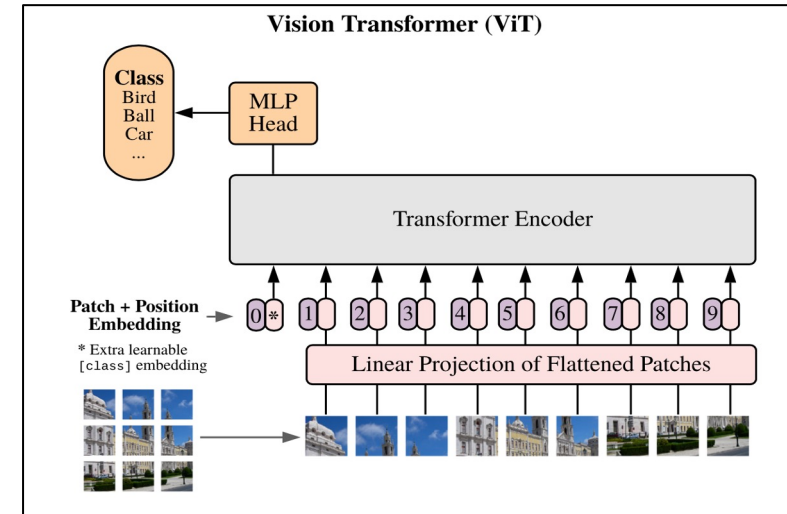
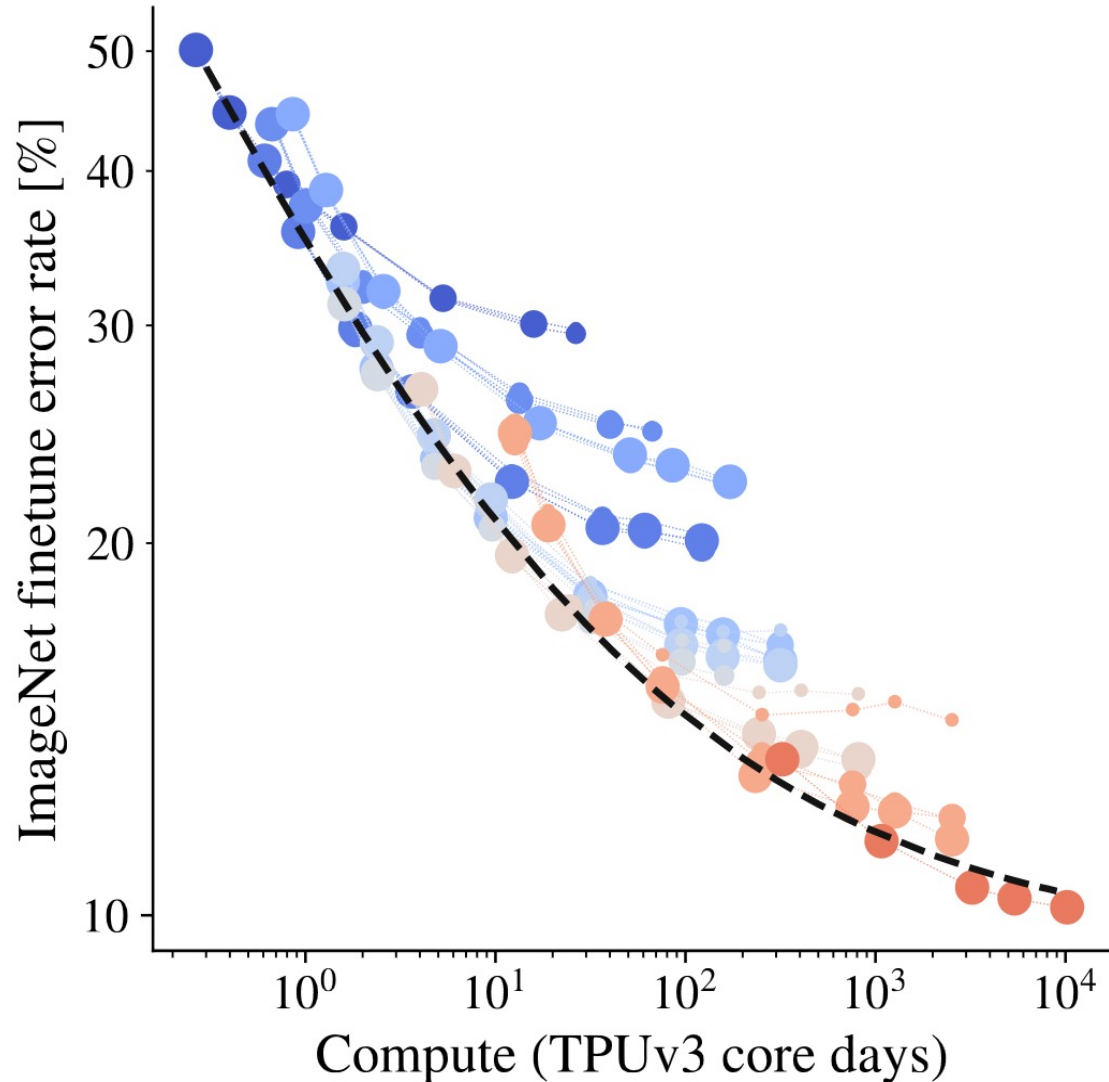
Scaling laws in 'vision' transformers

Getting larger, getting better!



A larger-curated dataset tends to occur labor/ethical concerns
less data, less computational resources, higher performance is preferable!

Scaling laws in 'vision' transformers



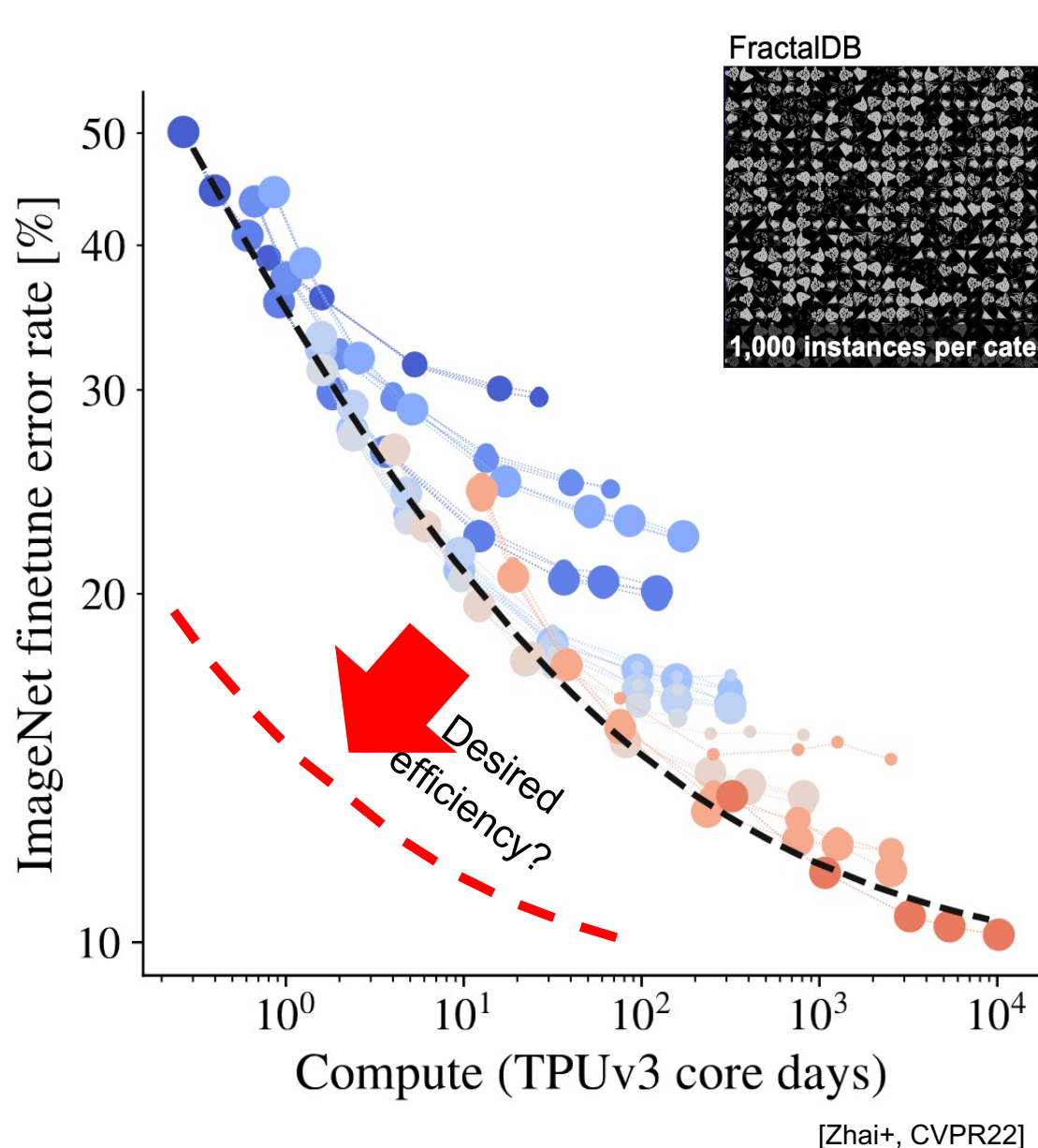
+

JFT-3B

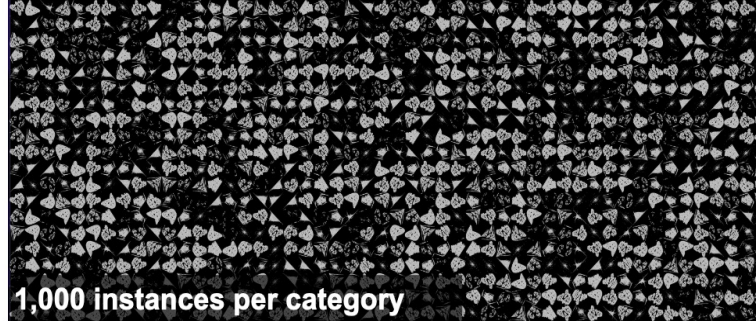
3,000,000,000 real images
are required for ViT pre-training



Breaking scaling laws in 'vision' transformers?



FractalDB



OFDB

One-instance FractalDB



[Nakamura+, ICCV23]

OFDB

1,000 synthetic images
are enough for ViT pre-training

1-video SSL

1 real video
is enough for ViT pre-training



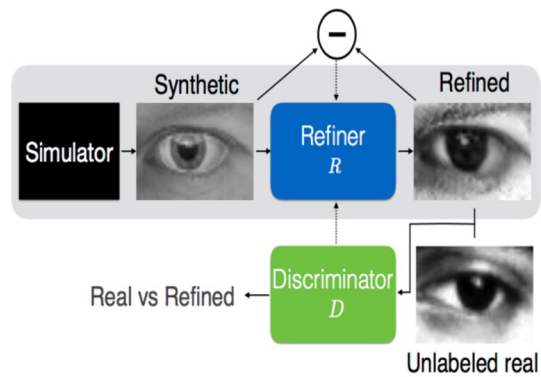
Best Paper Honorable Mention



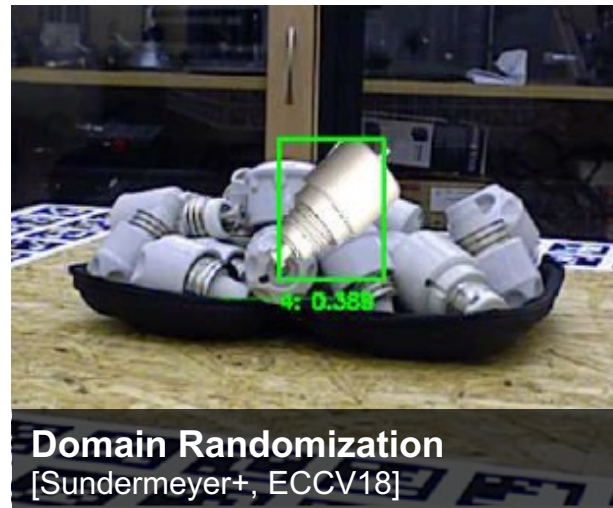
cvpaper.challenge

More wonderful solutions for 'LIMIT' community!

Our community has awesome solutions



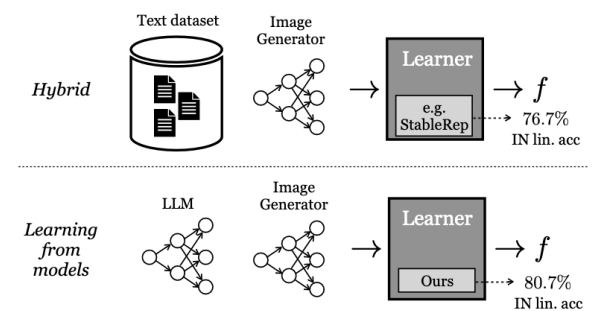
Sim-to-Real
[Shrivastava+, CVPR17]



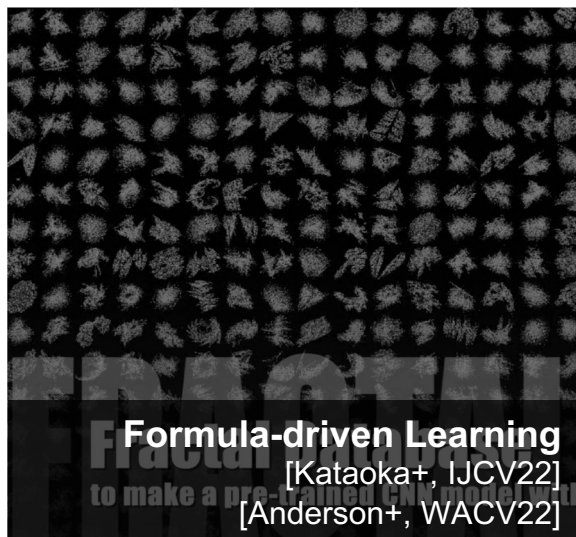
Domain Randomization
[Sundermeyer+, ECCV18]



Learning to See by Looking at Noise
[Baradad+, NeurIPS21]



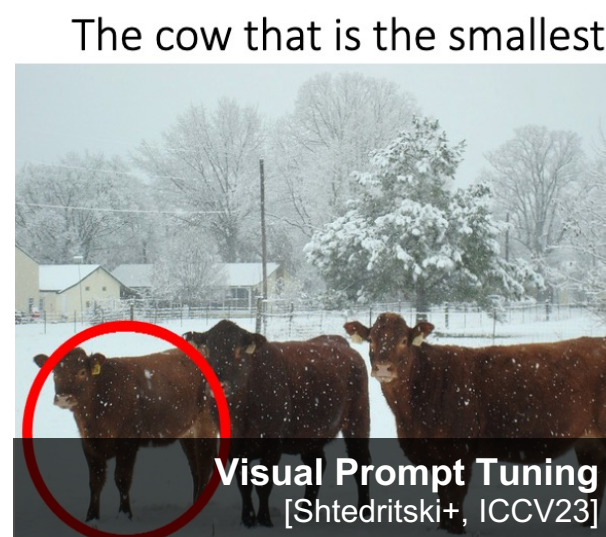
Representation Learning from synthetic images
[Tian+, CVPR24]



Formula-driven Learning
[Kataoka+, IJCV22]
[Anderson+, WACV22]



1-image SSL
[Asano+, ICLR20, ICLR23]



Visual Prompt Tuning
[Shtedritski+, ICCV23]

The cow that is the smallest

Keyword: Synthetic Data,
Adversarial Learning, Data
Augmentation, Domain
Randomization, Formula-driven
Supervised Learning, Visual
Prompt Tuning

Invited talk 1: Phillip Isola (MIT)

Title: N=0: Learning Vision with Zero Visual Data



Phillip Isola is the Class of 1948 Career Development associate professor in EECS at MIT. He studies computer vision, machine learning, robotics, and AI. He completed his Ph.D. in Brain & Cognitive Sciences at MIT, and has since spent time at UC Berkeley, OpenAI, and Google Research. His work has particularly impacted generative AI and self-supervised representation learning. Dr. Isola's research has been recognized by a Google Faculty Research Award, a PAMI Young Researcher Award, a Samsung AI Researcher of the Year Award, a Packard Fellowship, and a Sloan Fellowship. His teaching has been recognized by the Ruth and Joel Spira Award for Distinguished Teaching. His current research focuses on trying to scientifically understand human-like intelligence. (Refer from: <http://web.mit.edu/phillipi/www/bio.html>)

Invited talk 2: Zeynep Akata (TUM / Helmholtz Munich)

Title: Learning with Small Number of Images in Multimodal Large Language Models



Zeynep Akata is a Liesel Beckmann Distinguished professor of Computer Science at Technical University of Munich and the director of the Institute for Explainable Machine Learning at Helmholtz Munich. After completing her PhD at the INRIA Rhone Alpes with Prof Cordelia Schmid (2014), she worked as a post-doctoral researcher at the Max Planck Institute for Informatics with Prof Bernt Schiele (2014-17) and at University of California Berkeley with Prof Trevor Darrell (2016-17) and as an assistant professor at the University of Amsterdam with Prof Max Welling (2017-19). Before moving to Munich in 2024, she was a professor of computer science (W3) within the Cluster of Excellence Machine Learning at the University of Tübingen. She received a Lise-Meitner Award for Excellent Women in Computer Science from Max Planck Society in 2014, a young scientist honour from the Werner-von-Siemens-Ring foundation in 2019, an ERC-2019 Starting Grant from the European Commission, The DAGM German Pattern Recognition Award in 2021, The ECVA Young Researcher Award in 2022 and the Alfried Krupp Award in 2023. Her research interests include multimodal learning and explainable AI. (Refer from: <https://www.eml-unitue.de/people/zeynep-akata>)

