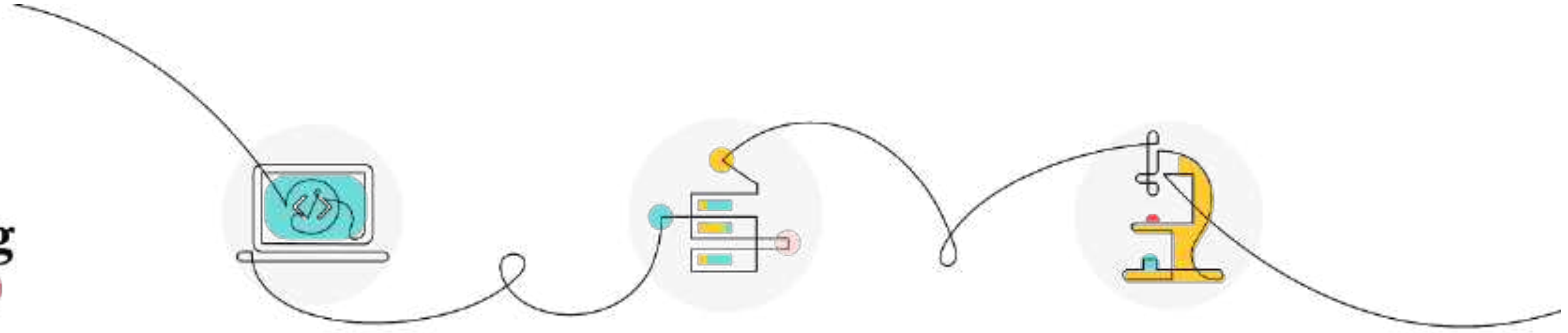
https://sebastianraschka.com

@rasbt

Chan Zuckerberg Initiative



Seed Networks
Computational Biology Meeting

Modern machine learning

An introduction to the latest techniques

Sebastian Raschka

About Myself

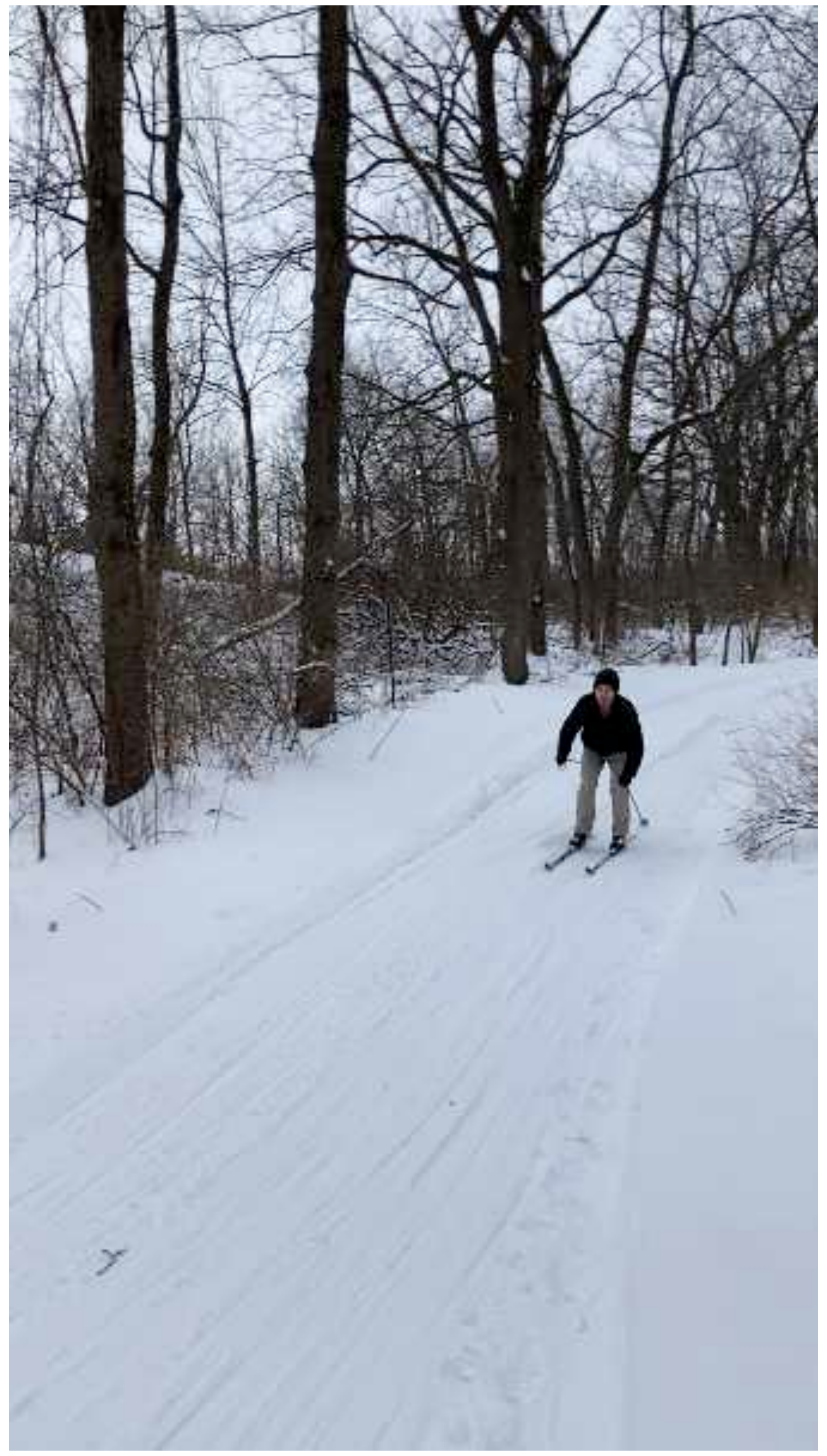
Contact:

https://sebastianraschka.com

@rasbt

Background & Specialties:

- Computational Biology
- Machine learning
- Deep learning
- Wisconsin State Parks



Affiliation:

Assistant Professor

Department of Statistics

https://stat.wisc.edu

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Slides: http://sebastianraschka.com/pdf/slides/2021-04_czi.pdf

Topics

(1) Intro to Machine Learning
- What is Machine Learning
- Deep Learning Frameworks

(2) Methods that Work
- Tabular Data
- Images
- Sequences & Text
- Improving Performance

(3) Challenges
- Small Data
- Ordinal Data
- Adversarial Attacks
- Bias

(4) Recent Trends
- Graphs
- Self-supervised Learning
- Transformers

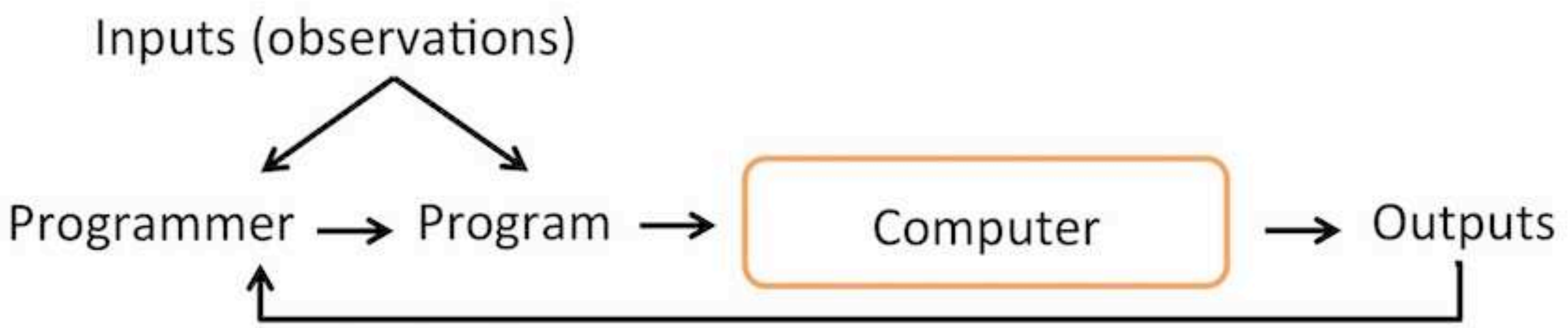
Part 1

(1) Intro to Machine Learning

What is Machine Learning

Deep Learning Frameworks

The Traditional Programming Paradigm



Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed
– Arthur Samuel (1959)

Machine Learning

Image source: https://sebastianraschka.com/blog/2020/intro-to-dl-ch01.html

The 3 Broad Categories of ML (and DL)

Focus of today's talk

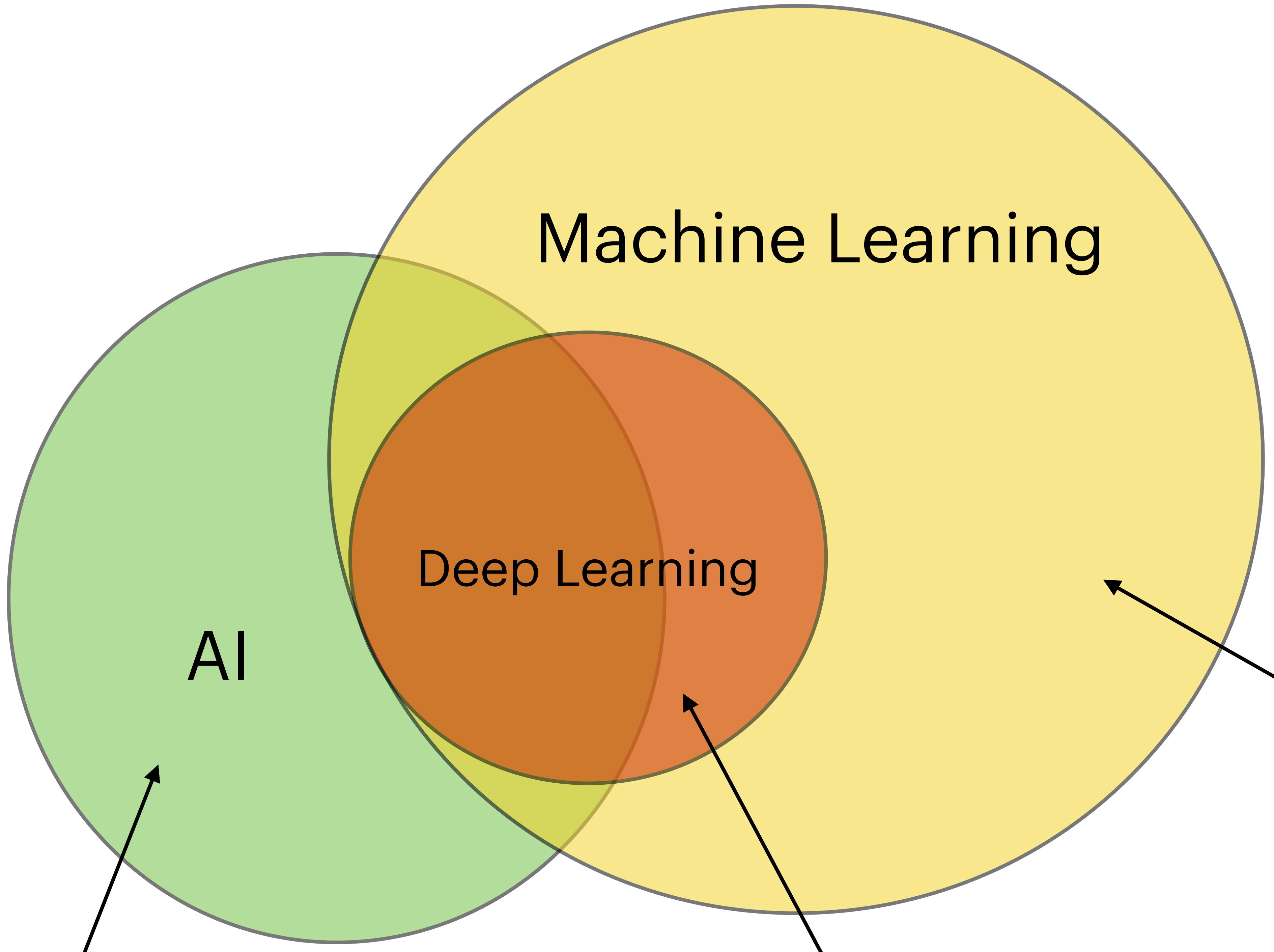
Supervised Learning	Labeled data Direct feedback Predict outcome/future
Unsupervised Learning	No labels/targets No feedback Find hidden structure in data
Reinforcement Learning	Decision process Reward system Learn series of actions

"Label learning"

- Regression
- Classification

Image source: Raschka and Mirjalili (2019). *Python Machine Learning, 3rd Edition.*
https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750

The Connection Between Fields



E.g.,
generalized linear models,
tree-based methods,
"shallow" networks,
support vector machines,
nearest neighbors, ...

E.g., symbolic expressions,
logic rules / "handcrafted"
nested if-else programming
statements ...

E.g., ML deep neural networks
capable of automatic feature
extraction

Image source: https://sebastianraschka.com/blog/2020/intro-to-dl-ch01.html

Deep Learning Frameworks: An Abbreviated History

2000s:
- OpenNN, Torch, Matlab

2010s:
- (Multi)-GPU support: Caffe, config files; Chainer imperative; Theano declarative

2015s:
- TensorFlow (Google), declarative
- Caffe2 (FAIR, by TensorFlow dev)
- CNTK (Microsoft)
- DyNet (Carnegie Mellon University)
- Paddle Paddle (Baidu)
- MXNet (Amazon support), declarative & imperative "mix"
- Keras API
- PyTorch (FAIR), imperative (Torch and Chainer)

Things Looks Much Simpler in 2021

2000s:

- OpenNN, ~~Torch~~, Matlab

2010s:

(PyMC3)

- ~~Caffe, config files~~; ~~Chainer imperative~~; ~~Theano declarative~~

2015s:

- ~~TensorFlow (Google), declarative~~
- ~~Caffe2 (FAIR, by TensorFlow dev)~~
- ~~CNTK (Microsoft)~~
- ~~MXNet (Amazon support), declarative & imperative "mix"~~

...

- ~~Keras~~
- PyTorch (FAIR), imperative (Torch and Chainer)

2021:

- TensorFlow v2
- PyTorch
- JAX

Part 2

(2) Methods that Work

- Tabular Data
- Images
- Sequences & Text
- Improving Performance

Structured vs Unstructured Data

A

Feature vector of the 1st training example

Class label

Index	Sepal length	Sepal width	Petal length	Petal width	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
...	...	...	...	...	...
150	5.9	3	5.1	1.8	Iris-virginica

B

Image source: https://sebastianraschka.com/blog/2020/intro-to-dl-ch01.html

Supervised Learning Methods for Tabular Data

Linear classifier/regressor as a good baseline:

Linear / (Multinomial) logistic regression

Robust non-linear classifier without tuning:

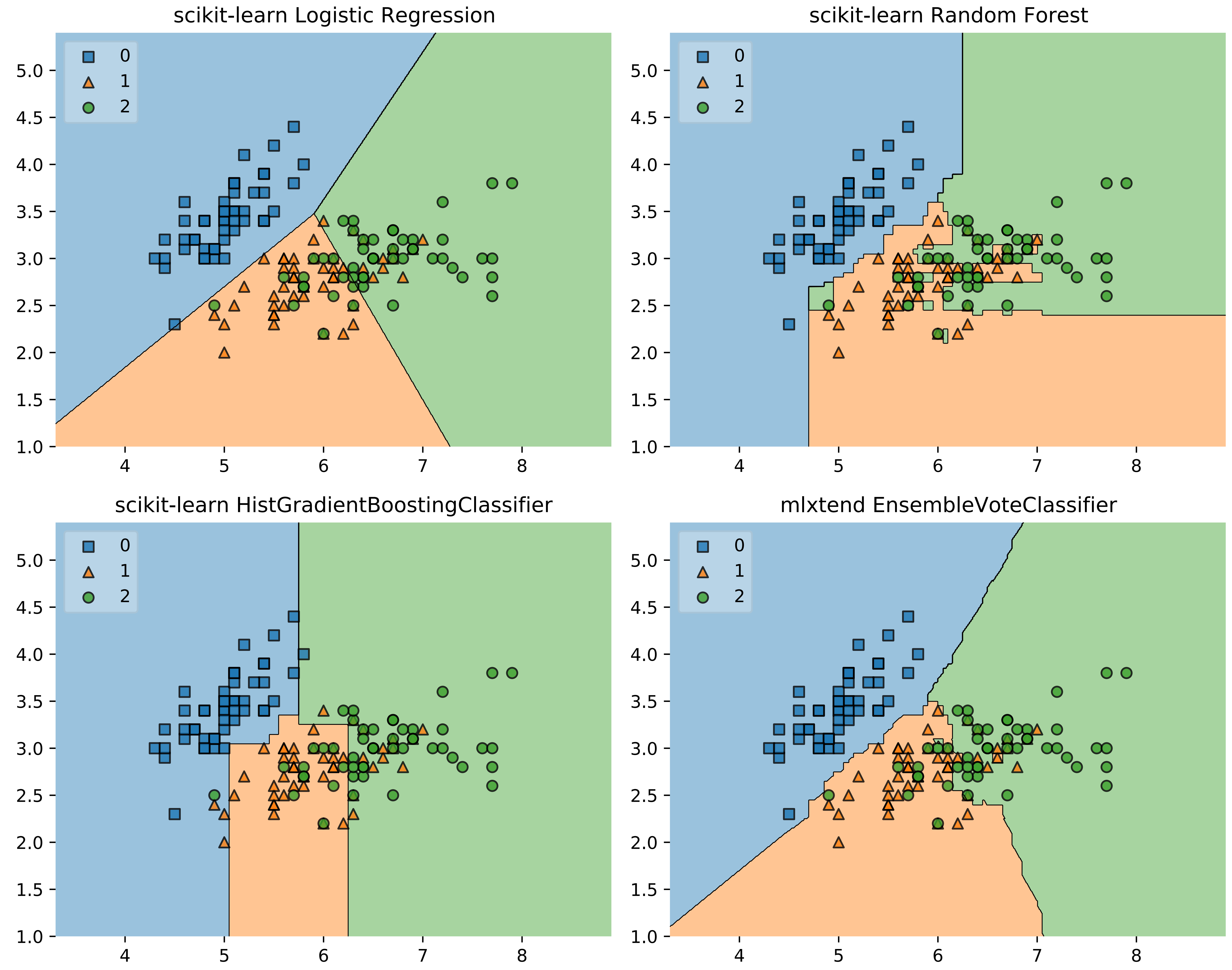
Random forests

State-of-the-art model for tabular data:

Gradient boosting (XGBoost, LightGBM, HistGradientBoostingClassifier...)

Supervised Learning Methods for Tabular Data

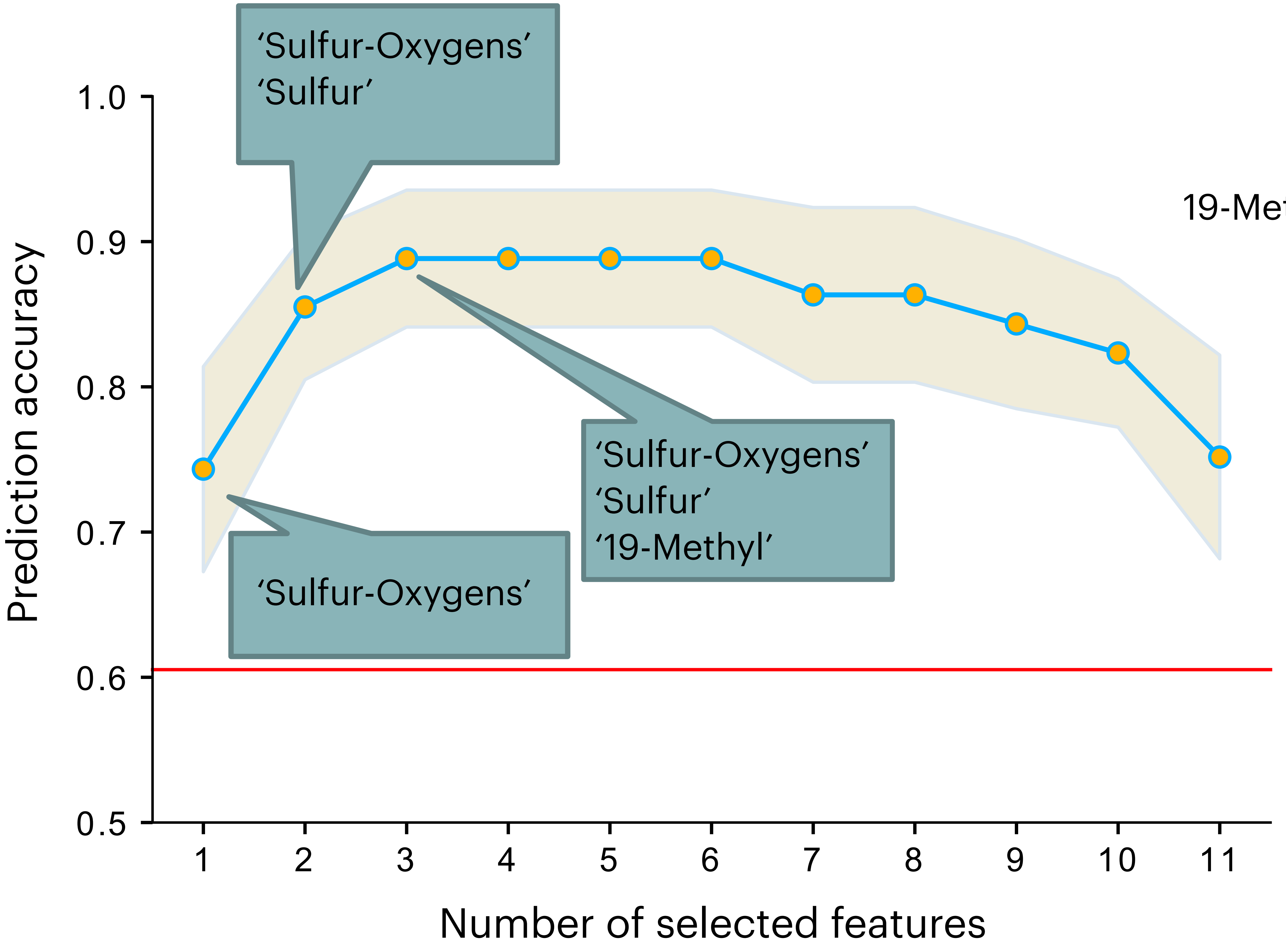
Iris classification toy example: sepal lengths & widths

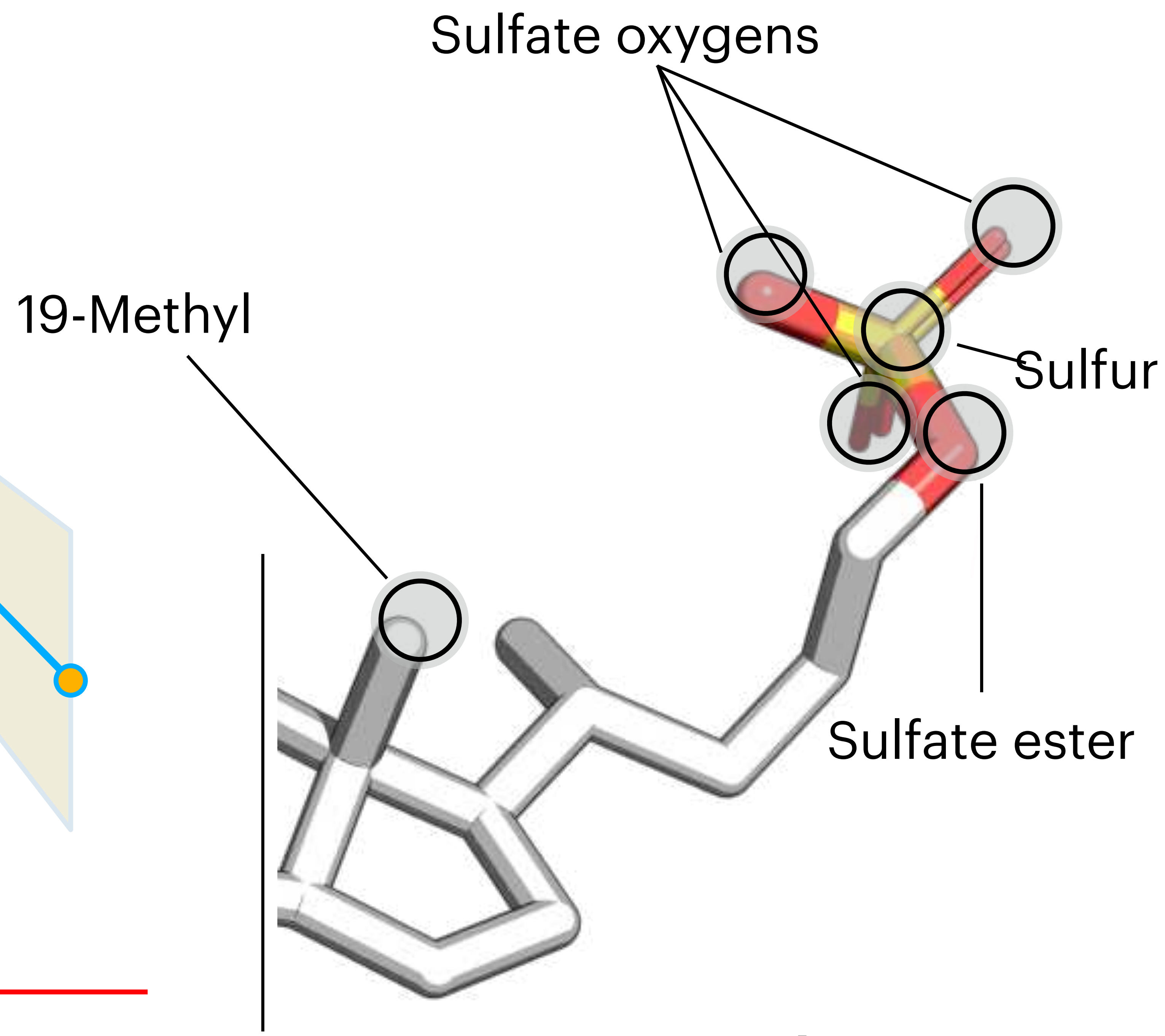


Feature Selection

`SequentialFeatureSelector`

Sebastian Raschka (2018) *MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack.* The Journal of Open Source Software 3.24.





DKPES pheromone

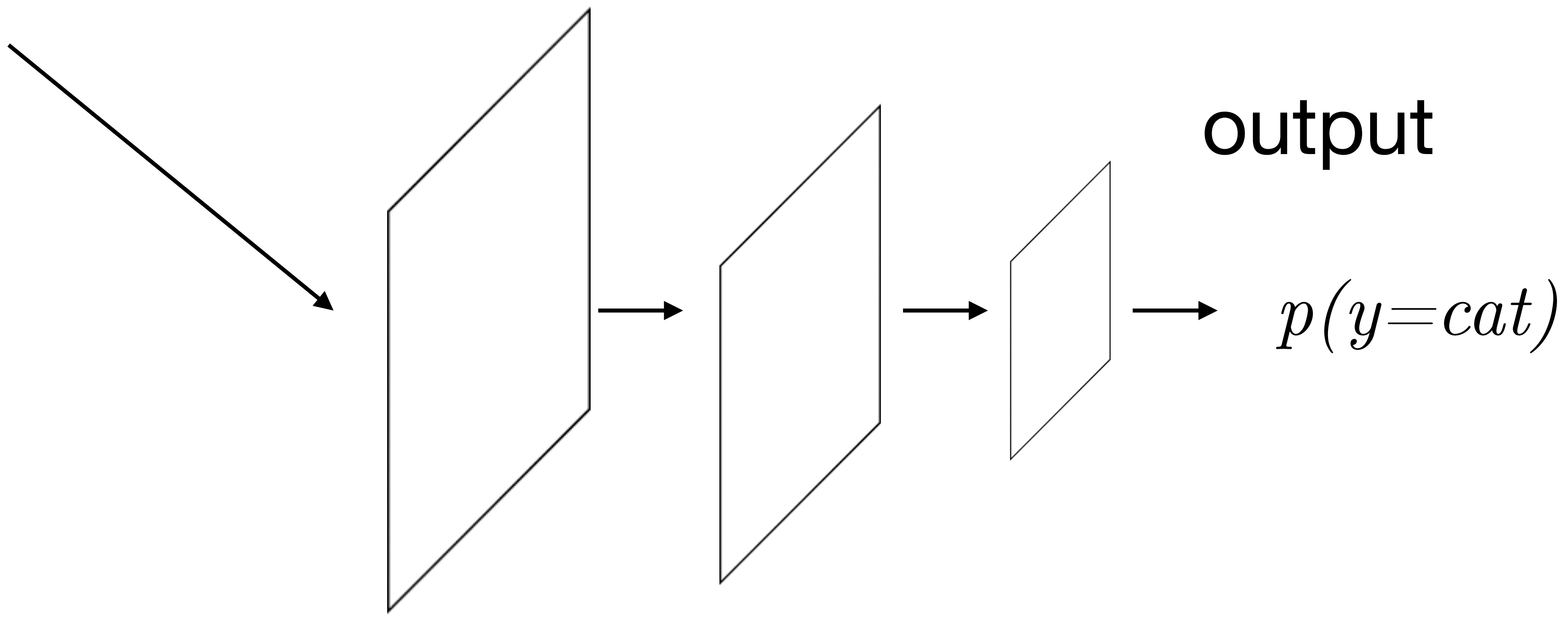
Raschka, Kuhn, Scott, Li (2018) Computational Drug Discovery and Design: *Automated Inference of Chemical Group Discriminants of Biological Activity from Virtual Screening Data*. Springer. ISBN: 978-1-4939-7755-0

Raschka, Liu, Gunturu, Scott, Huertas, Li, and Kuhn (2018) *Facilitating the Hypothesis-driven Prioritization of Small Molecules in Large Databases: Screenlamp and its Application to GPCR Inhibitor Discovery*. Journal of Computer-Aided Molecular Design, 32(3), 415-433.

Convolutional Neural Networks for Image Data

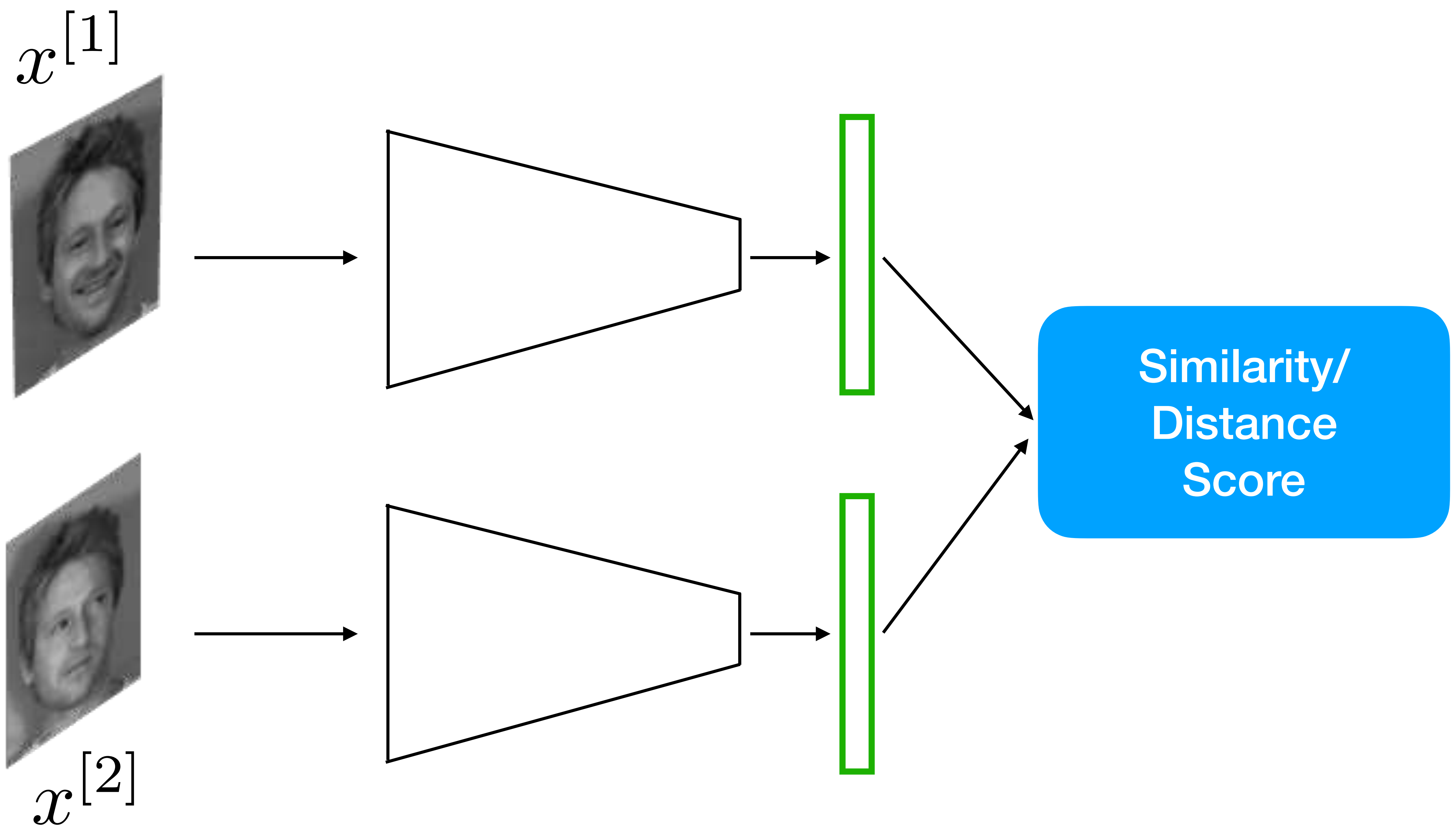
Image Source:
twitter.com%2Fcats&psig=AOvVaw30_o-PCM-K21DiMAJQimQ4&ust=1553887775741551

Image Source: https://www.pinterest.com/pin/244742560974520446



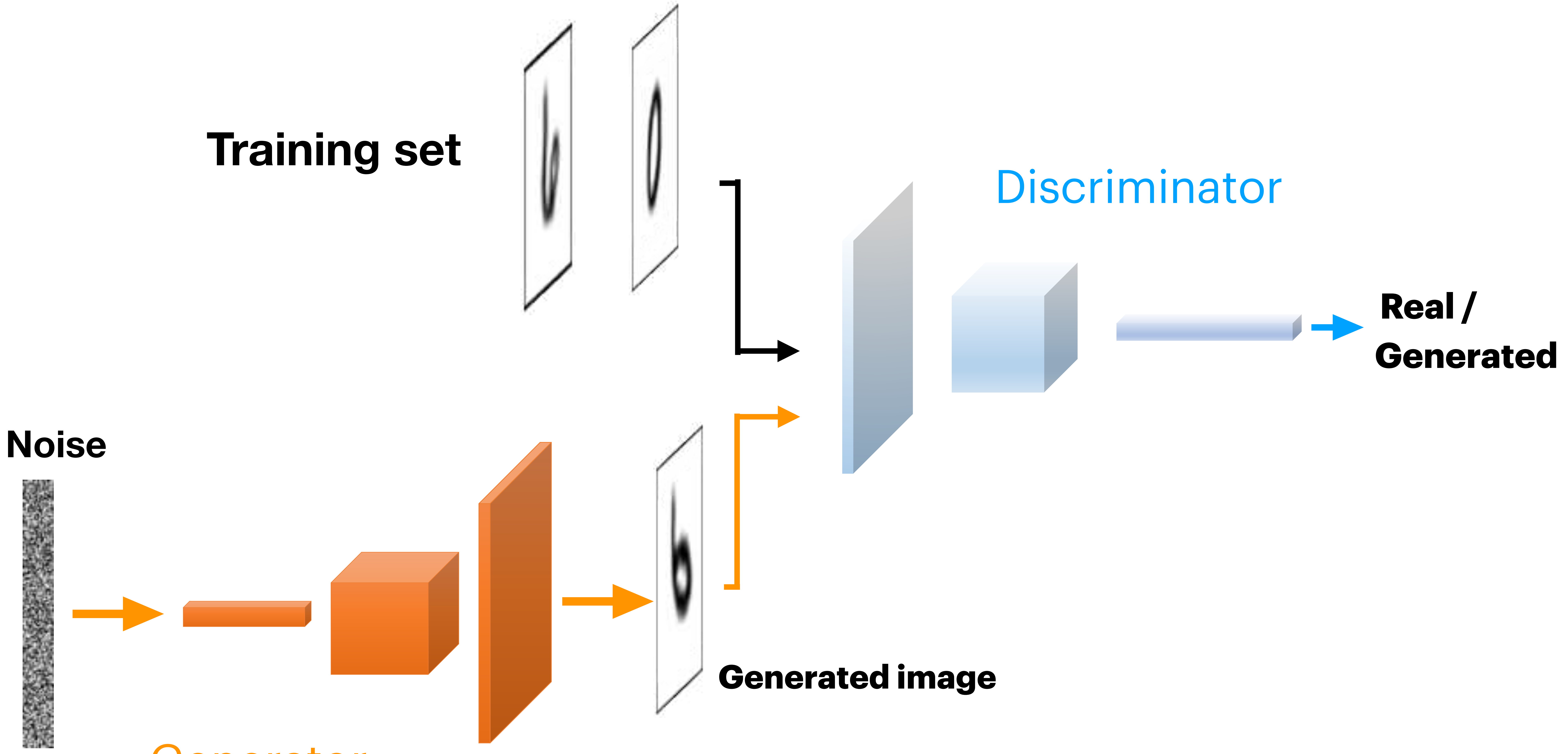
Convolutional Neural Networks (CNNs) for Image Classification

Image Comparison (e.g., Face Recognition)



Source: MUCT dataset

Image Synthesis (e.g., Generative Adversarial Network)



Source: MNIST dataset

Convolutional Neural Network Architectures (~2019)

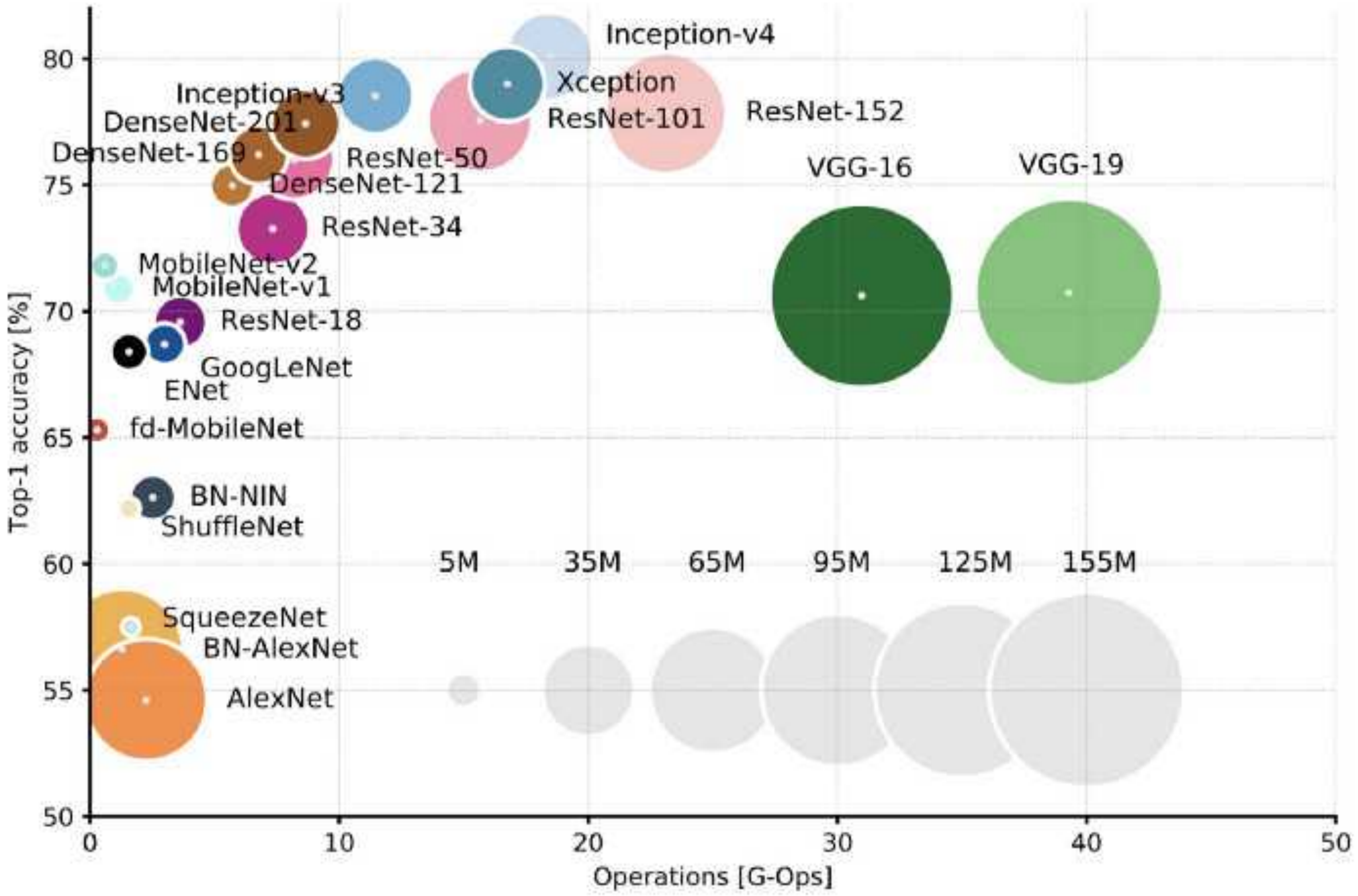


Image source:
Analysis of deep neural networks
By Alfredo Canziani, Thomas Molnar, Lukasz Burzawa, Dawood Sheik, Abhishek Chaurasia, Eugenio Culurciello
https://culurciello.medium.com/analysis-of-deep-neural-networks-dcf398e71aae

CNNs Also Work for 1D and (*here*) 3D Data

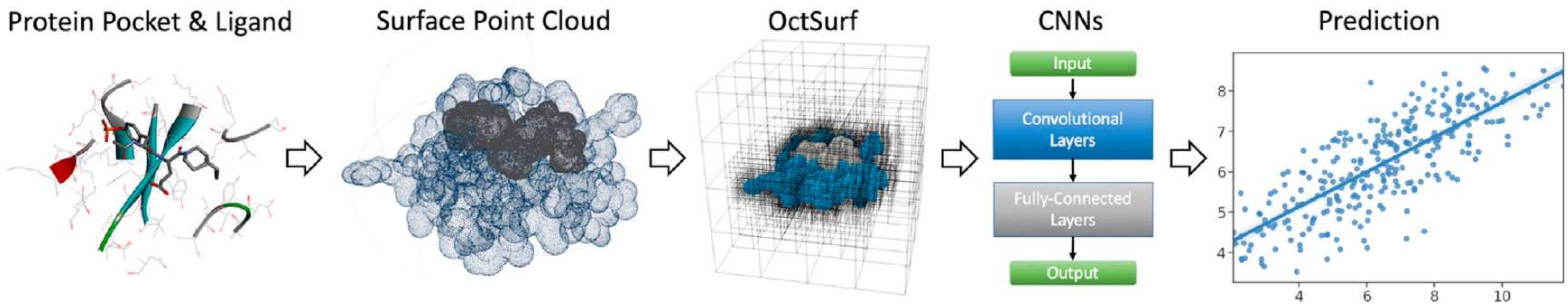


Fig. 2. The pipeline of our 3D-CNN implementation for the protein-ligand affinity prediction based on the OctSurf representation. Surface point clouds of binding pockets and bound ligands are rasterized into the octree-based volumetric representation, OctSurf, which are fed into the 3D-CNNs for binding affinity prediction.

Liu Q, Wang PS, Zhu C, Gaines BB, Zhu T, Bi J, Song M. OctSurf: Efficient hierarchical voxel-based molecular surface representation for protein-ligand affinity prediction. Journal of Molecular Graphics and Modelling. 2021 Jun 1;105:107865.
https://www.sciencedirect.com/science/article/pii/S1093326321000346

Recurrent Neural Networks for Text (and Sequence Data in General)

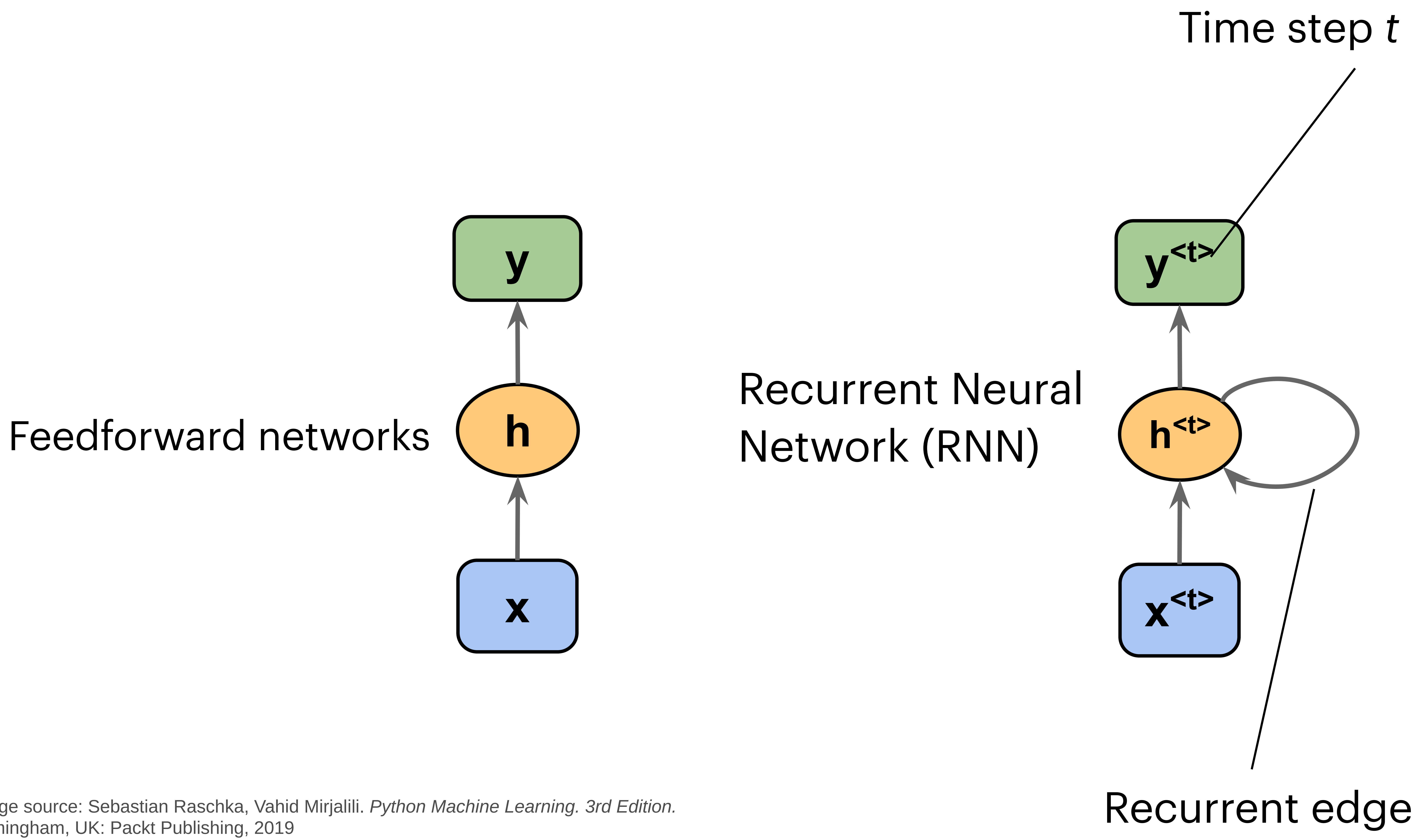
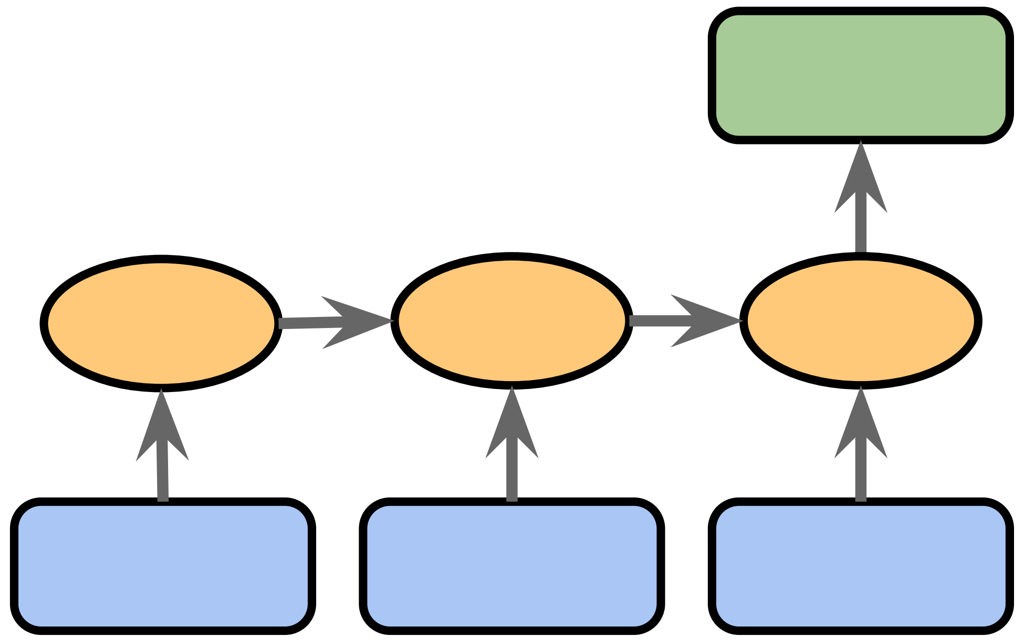


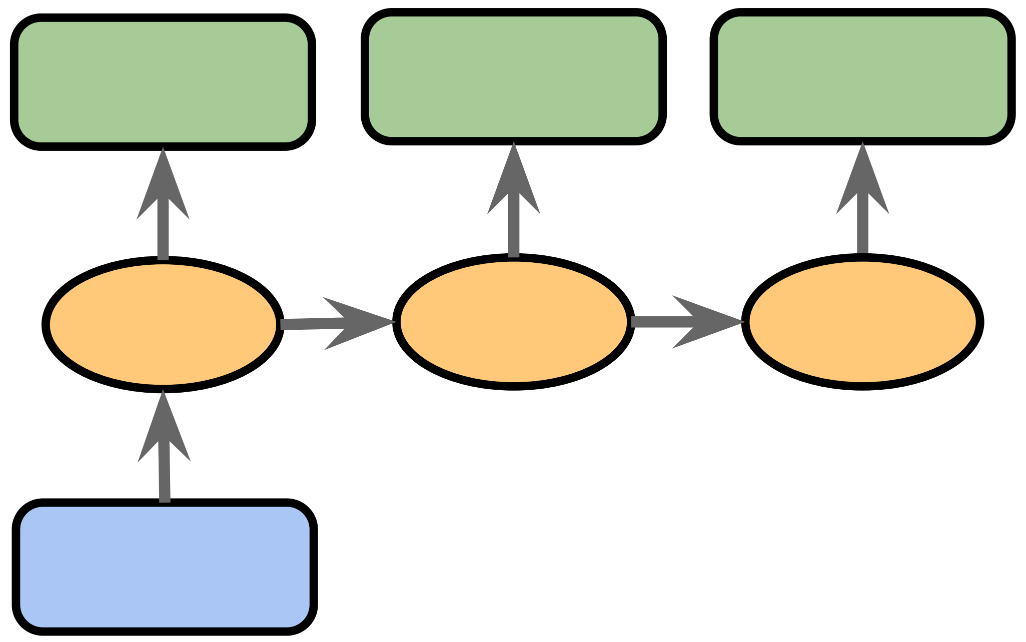
Image source: Sebastian Raschka, Vahid Mirjalili. *Python Machine Learning. 3rd Edition.* Birmingham, UK: Packt Publishing, 2019
https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750

RNNs Are Versatile With Respect to Prediction & Generation Tasks

E.g., sentiment analysis



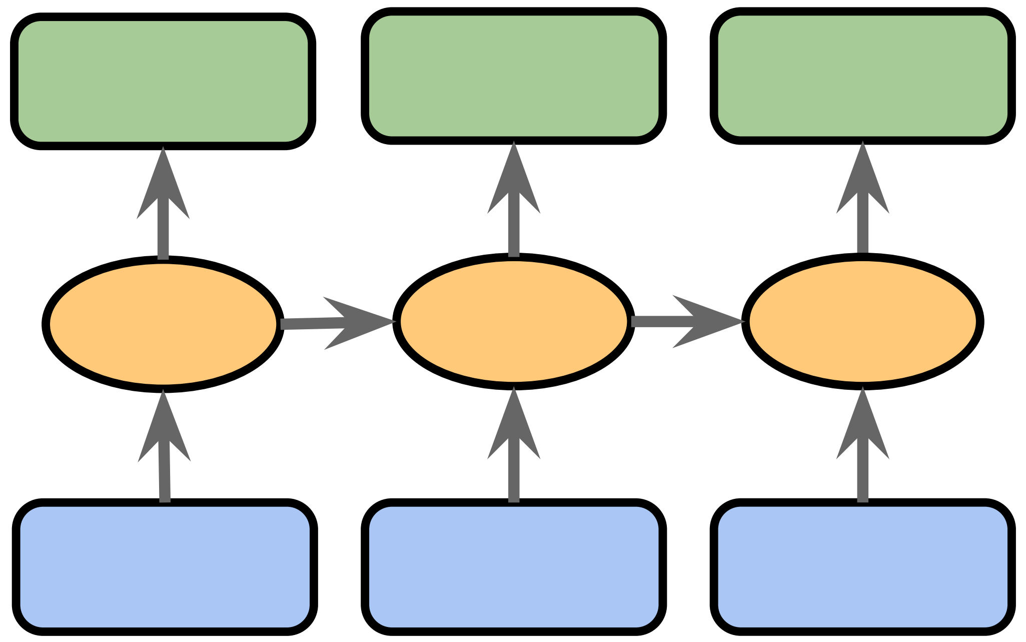
many-to-one



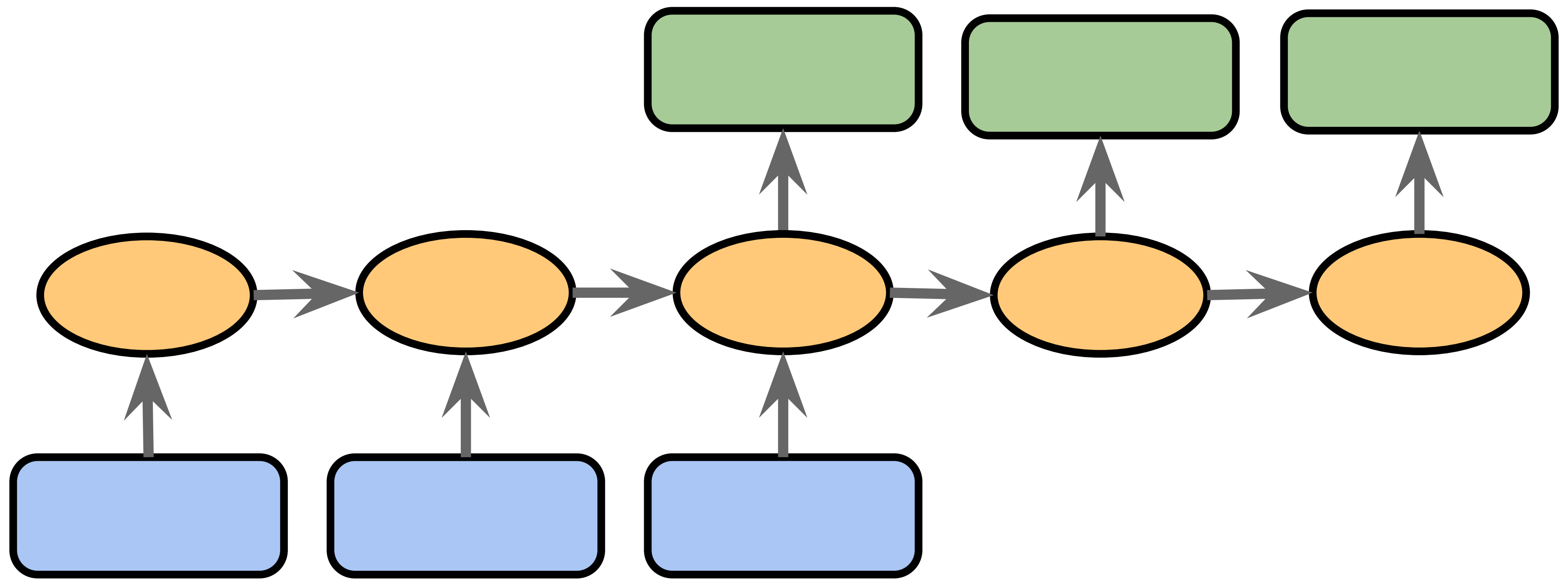
one-to-many

E.g., image captioning

E.g., video captioning



many-to-many



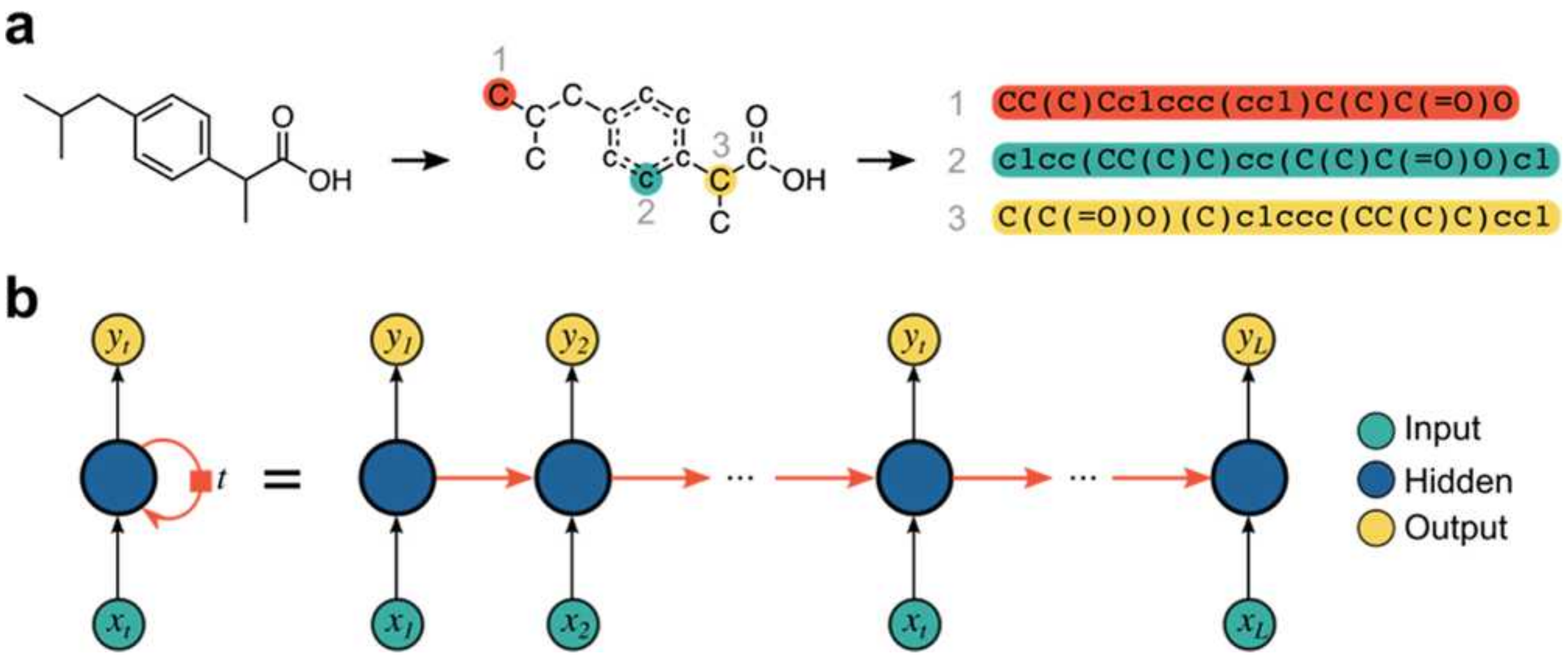
many-to-many

E.g., language translation

Image source: Sebastian Raschka, Vahid Mirjalili. *Python Machine Learning. 3rd Edition.* Birmingham, UK: Packt Publishing, 2019 https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750

Figure based on:
The Unreasonable Effectiveness of Recurrent Neural Networks by Andrej Karpathy (*http://karpathy.github.io/2015/05/21/rnn-effectiveness/*)

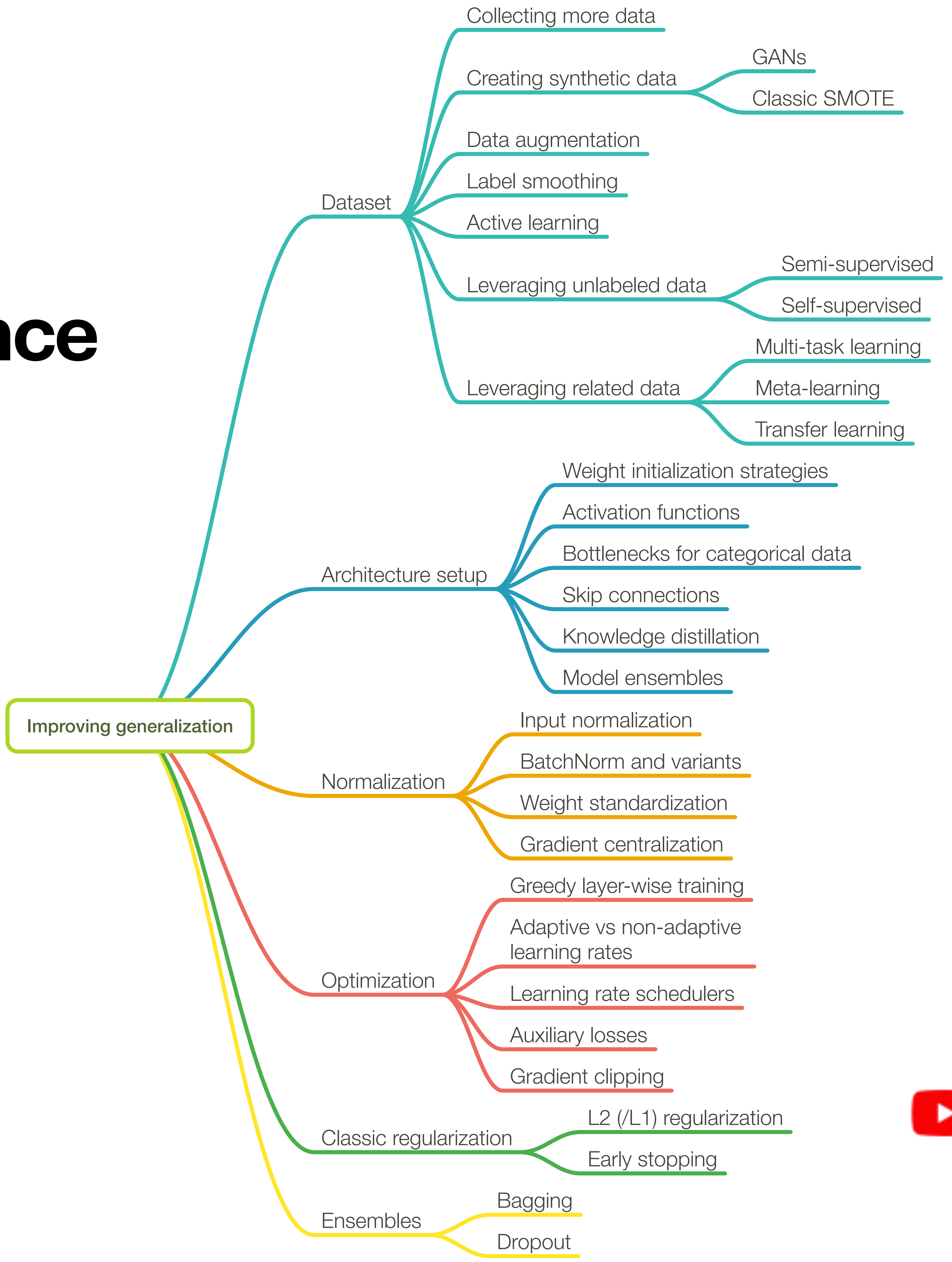
RNNs Can Be Used for Predictive and Generative Modeling



SMILES strings of Ibuprofen

Grisoni F, Moret M, Lingwood R, Schneider G. *Bidirectional molecule generation with recurrent neural networks*. Journal of Chemical Information and Modeling. 2020 Jan 6;60(3):1175-83.
https://pubs.acs.org/doi/abs/10.1021/acs.jcim.9b00943

Tuning Models to Improve Performance



L10.1 Techniques for Reducing Overfitting
https://youtu.be/KOBmBjlMVAE

Academia Vs Industry

Model-Centric Approach

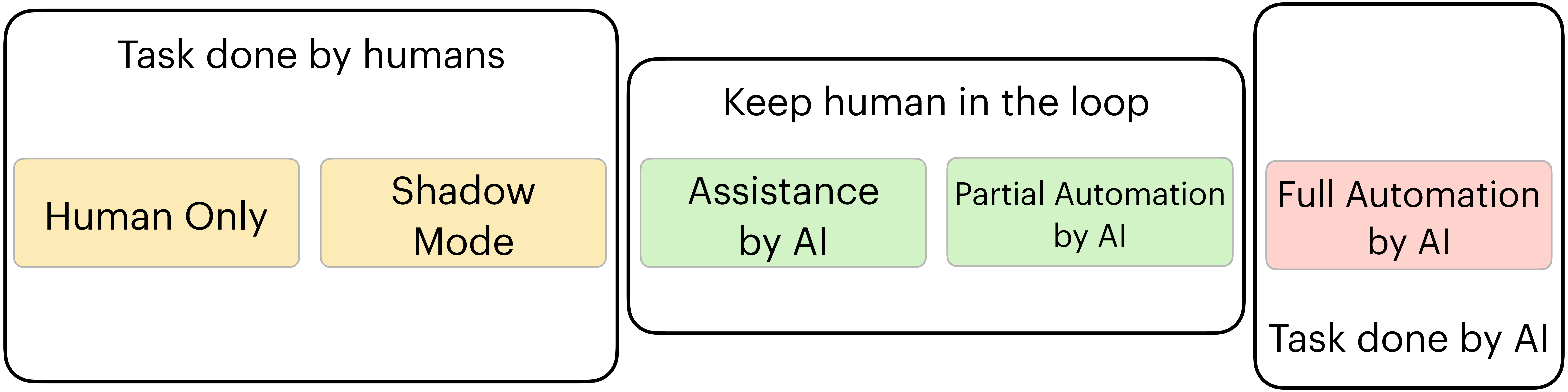
Primary focus is on tuning and developing models to improve performance on a fixed benchmark set

Data-Centric Approach

Primary focus is on how one can improve the dataset (collect more, select, relabel) to improve model performance

Source: Andrej Karpathy, Andrew Ng

What Problem Do You Want To Solve?



Source: Andrew Ng

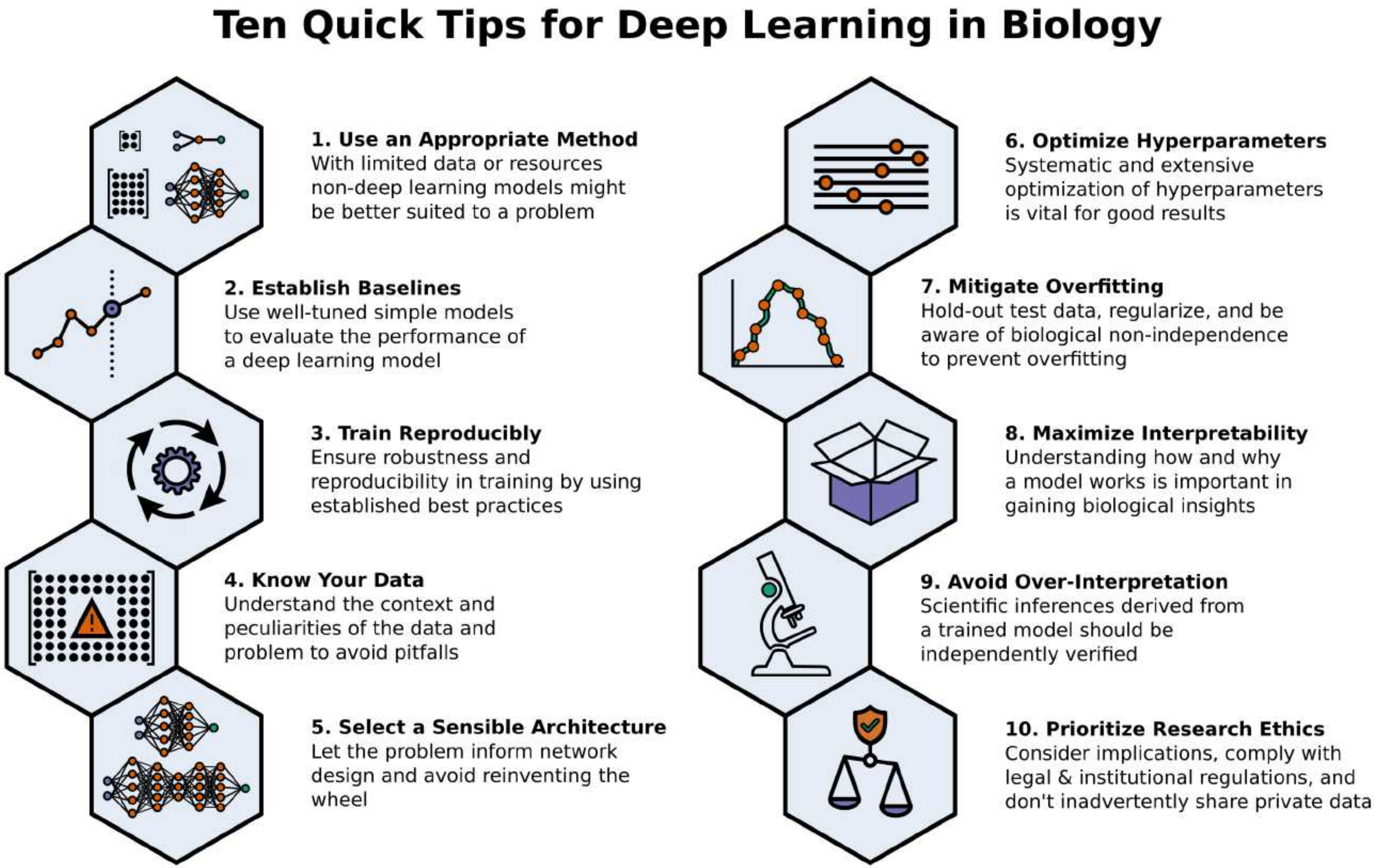
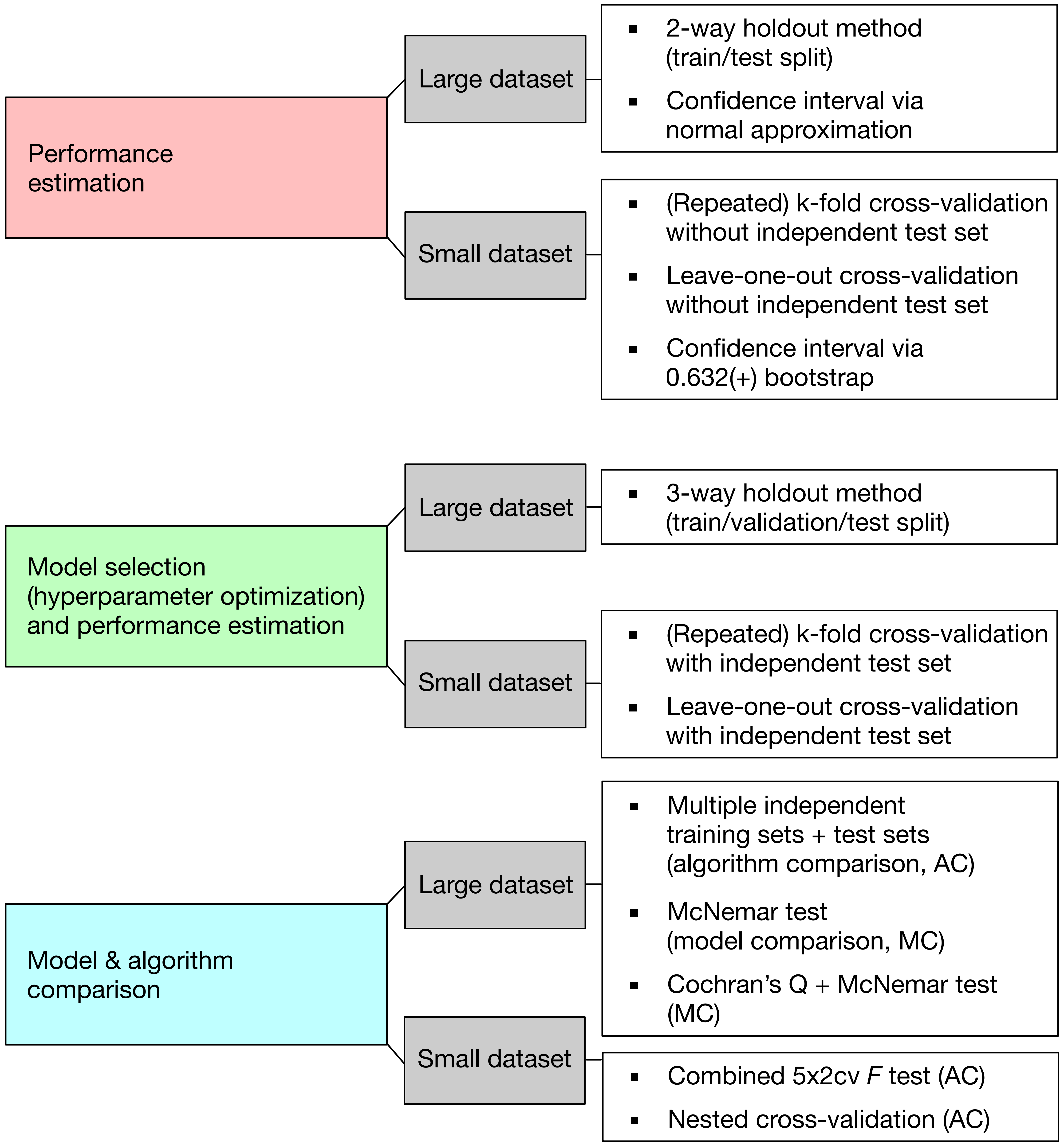


Image source:
Lee BD, Gitter, A, Greene CS, Raschka S, Maguire F, Titus A, Kessler M, Lee AJ et al. ***Ten Quick Tips for Deep Learning in Biology*** (under review)
https://benjamin-lee.github.io/deep-rules/manuscript.pdf

What is the Best/ Recommended Model Evaluation Strategy? It Depends!



AC = Algorithm comparison
MC = Model comparison

Image Source:
Sebastian Raschka (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. https://arxiv.org/abs/1811.12808

Part 3

(3) Challenges

- Small Data
- Ordinal Data
- Adversarial Attacks
- Bias

Tackling Small Data Problems

Active learning
Optimize data order and labeling

Transfer learning
Pre-train on larger related dataset with labels

Semi-supervised learning
Incorporate unlabeled data into the training

Few-shot learning
Special cases with very few examples
per class (incl. transfer learning, metric learning,
semi-supervised, meta-learning)

Self-supervised learning
Pre-train on unlabeled dataset by creating
leveraging data structure to create labels

Ordinal Data: Integrating Label Order Info

- **Ranking:** Predict Correct order
  (0 loss if order is correct, e.g., rank a collection of movies by "goodness")

$\prec$

$\prec$

- **Ordinal regression:** Predict correct (ordered) label
  (E.g., age of a person in years; here, regard aging as a non-stationary process)

Excerpt from the UTKFace dataset
https://susanqq.github.io/UTKFace/


18

$\prec$

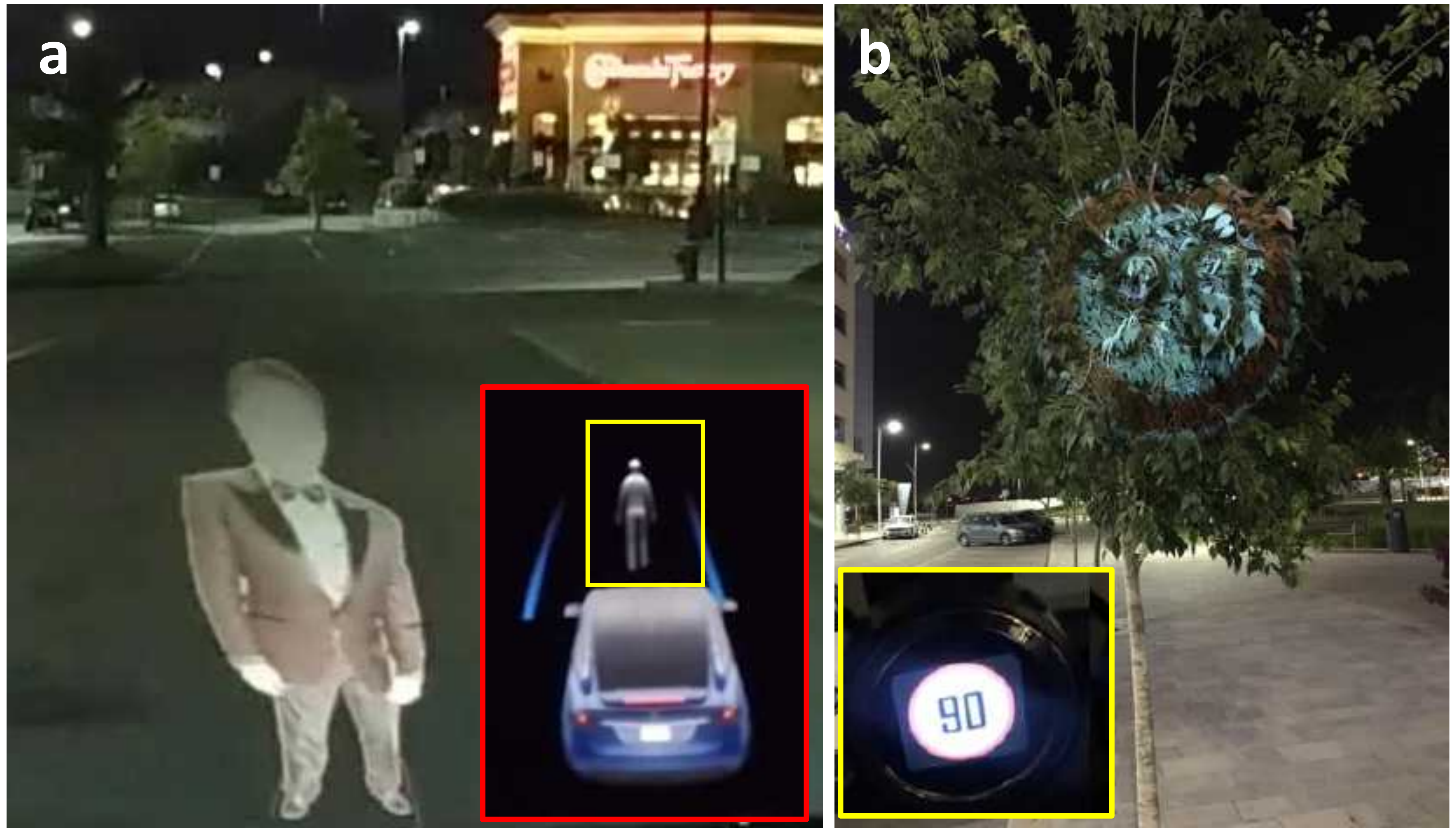

29

$\prec$


41

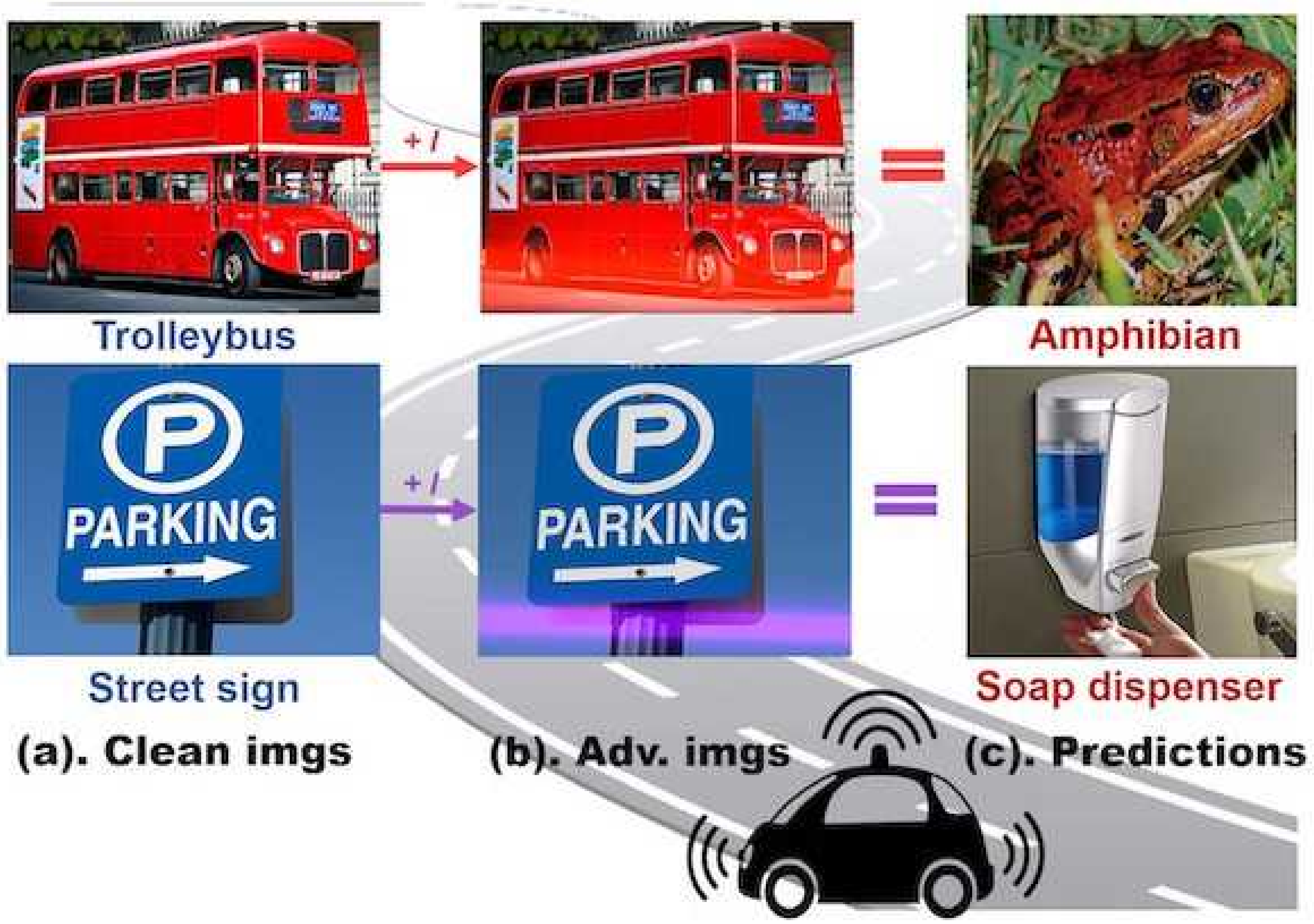
Cao, Mirjalili, Raschka (2020)
Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation
Pattern Recognition Letters. 140, 325-331
https://www.sciencedirect.com/science/article/pii/S016786552030413X

Beyond Pandas & Gibbons: Real-World Adversarial Attacks



Tesla Autopilot considers (a) as a real person and (b) as a real road sign

Nassi, Mirsky, Nassi, Ben-Netanel, Drokin, Elovici. *Phantom of the ADAS: Securing Advanced Driver-Assistance Systems from Split-Second Phantom Attacks*. ACM SIGSAC Conference on Computer and Communications Security, 2020
https://eprint.iacr.org/2020/085.pdf



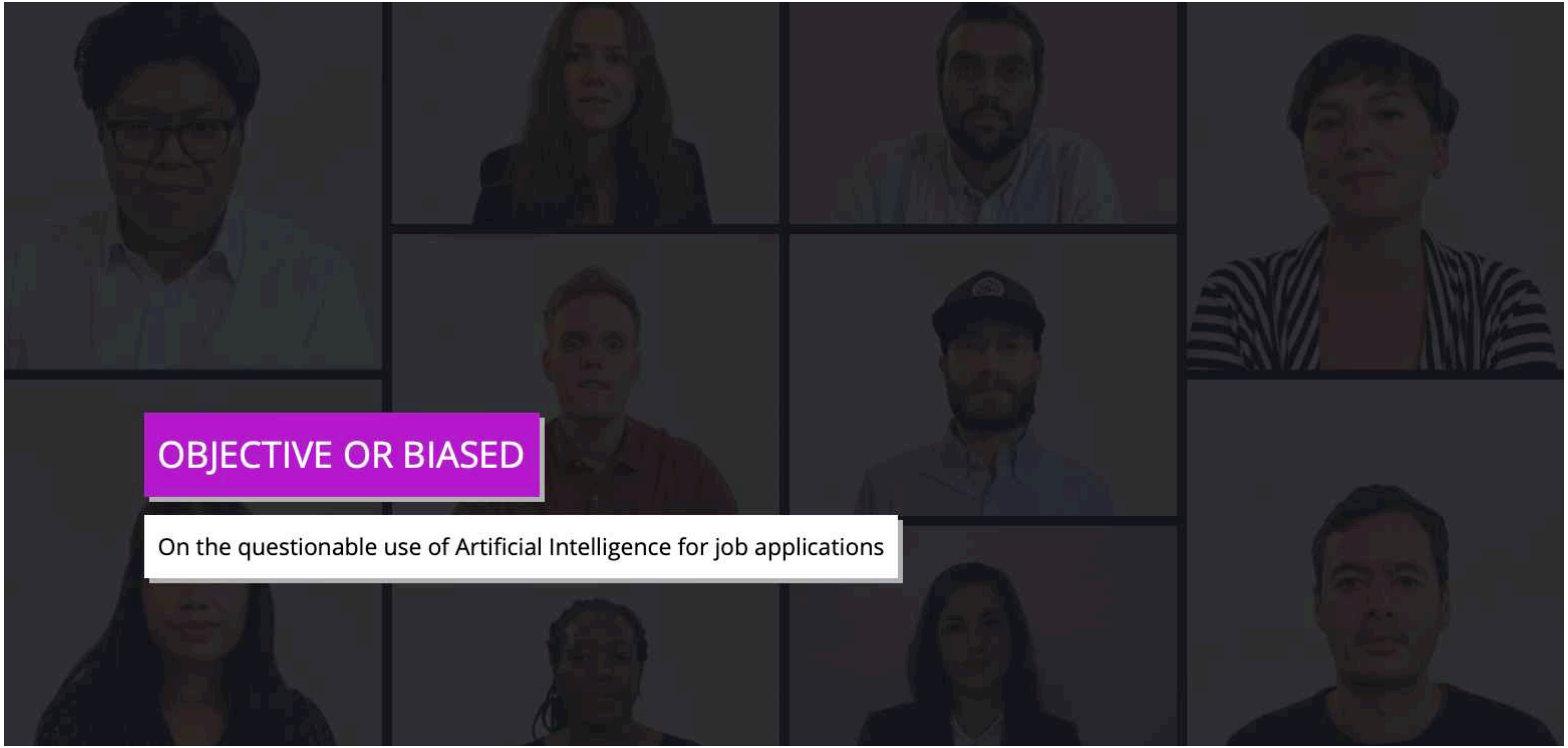
Laser beams turn buses into amphibians and street signs into soap dispensers

Duan, Mao, Qin, Yang, Chen, Ye, He. *Adversarial Laser Beam: Effective Physical-World Attack to DNNs in a Blink*. arXiv:2103.06504. 2021 Mar 11.
https://arxiv.org/abs/2103.06504

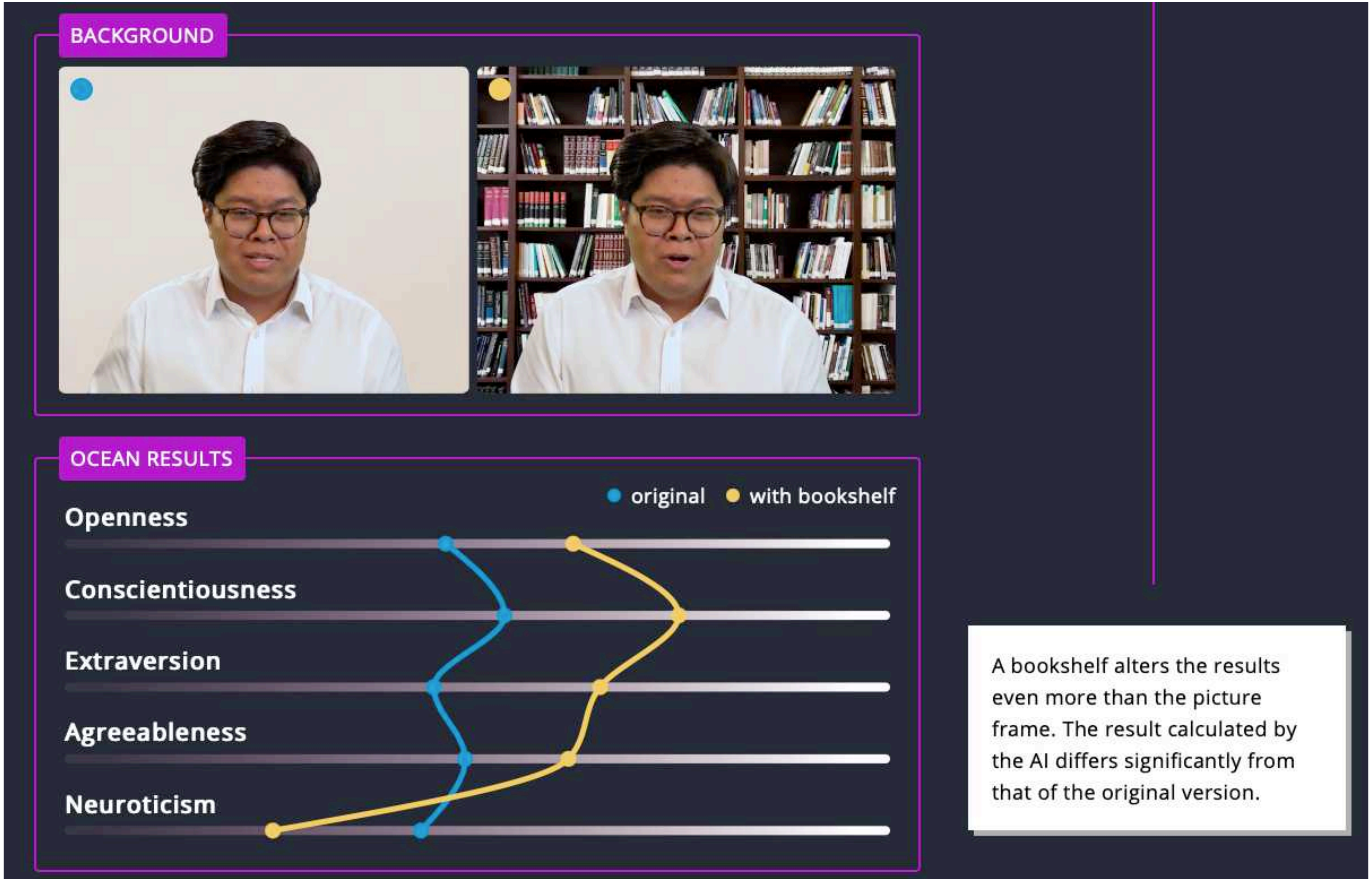
Some Common Adversarial Attacks & Defenses

	Cleverhans v3.0.1	FoolBox v2.3.0	ART v1.1.0	DEEPSEC (2019)	AdvBox v0.4.1
Supported frameworks					
TensorFlow	yes	yes	yes	no	yes
MXNet	yes	yes	yes	no	yes
PyTorch	no	yes	yes	yes	yes
PaddlePaddle	no	no	no	no	yes
(Evasion) attack mechanisms					
BLB [163]	yes	no	no	yes	no
AMD [170]	yes	no	no	no	no
ZOO [171]	no	no	yes	no	no
VA [172]	yes	yes	yes	no	no
AP [173]	no	no	yes	no	no
STA [174]	no	yes	yes	no	no
DTA [175]	no	no	yes	no	no
FGSM [176]	yes	yes	yes	yes	yes
R+FGSM [177]	no	no	no	yes	no
R+LLC [177]	no	no	no	yes	no
U-MI-FGSM [178]	yes	yes	no	yes	no
T-MI-FGSM [178]	yes	yes	no	yes	no
BIM [179]	no	yes	yes	yes	yes
LLC / ILLC [179]	no	yes	no	yes	no
UAP [180]	no	no	yes	yes	no
DeepFool [181]	yes	yes	yes	yes	yes
NewtonFool [182]	no	yes	yes	no	no
JSMA [183]	yes	yes	yes	yes	yes
CW/CW2 [184]	yes	yes	yes	yes	yes
PGD [185]	yes	no	yes	yes	yes
OM [186]	no	no	no	yes	no
EAD [187]	yes	yes	yes	yes	no
Boundary Attack [188]	no	yes	yes	no	no
HopSkipJumpAttack [189]	yes	yes	yes	no	no
MaxConf [190]	yes	no	no	no	no
Inversion attack [191]	yes	yes	no	no	no
SparseL1 [192]	yes	yes	no	no	no
SPSA [193]	yes	no	no	no	no
HCLU [194]	no	no	yes	no	no
ADef [195]	no	yes	no	no	no
DDNL2 [196]	no	yes	no	no	no
Local Search [197]	no	yes	no	no	no
Pointwise attack [198]	no	yes	no	no	no
GenAttack [199]	no	yes	no	no	no
Defense mechanisms					
Feature Squeezing [200]	no	no	yes	no	yes
Spatial Smoothing [200]	no	no	yes	no	yes
Label Smoothing [200]	no	no	yes	no	yes
Gaussian Augmentation [201]	no	no	yes	no	yes
Adversarial Training [185]	no	no	yes	yes	yes
Thermometer Encoding [202]	no	no	yes	yes	yes
NAT [203]	no	no	no	yes	no
EAT [177]	no	no	no	yes	no
DD [204]	no	no	no	yes	no
IGR [205]	no	no	no	yes	no
EIT [206]	no	no	yes	yes	no
RT [207]	no	no	no	yes	no
PixelDefend [208]	no	no	yes	yes	no
Regr.-based classfication [209]	no	no	no	yes	no
JPEG compression [210]	no	no	yes	no	no

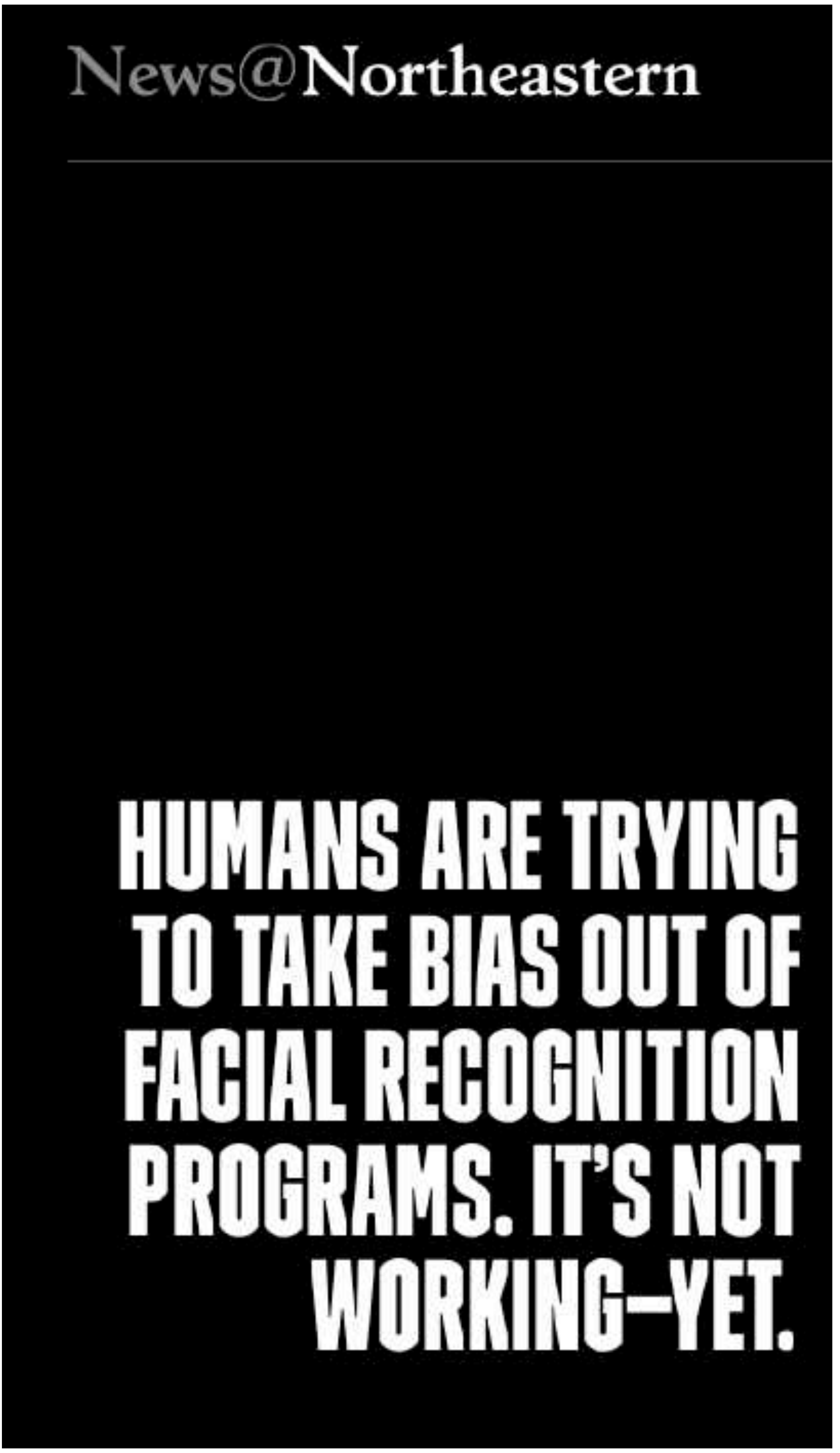
Raschka S, Patterson J, Nolet C. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. Information. 2020 Apr;11(4):193. https://www.mdpi.com/2078-2489/11/4/193



https://web.br.de/interaktiv/ki-bewerbung/en/



https://web.br.de/interaktiv/ki-bewerbung/en/



Common approach: Address lack of diversity in datasets.
--> provide algorithms with datasets that represent all groups equally and fairly

Does it work? Only for a stereotypical sense of fairness according to Zaid Khan:

> "The people in the images appeared to fit racial stereotypes. For example, an algorithm was more likely to label an individual in an image as 'white' if that person had blond hair."

https://news.northeastern.edu/2021/02/22/humans-are-trying-to-take-bias-out-of-facial-recognition-programs-its-not-working-yet/

Paper:
Khan Z, Fu Y.
One Label, One Billion Faces: Usage and Consistency of Racial Categories in Computer Vision.
ACM Conference on Fairness, Accountability, and Transparency 2021 Mar 3
https://dl.acm.org/doi/abs/10.1145/3442188.3445920

arXiv.org > cs > arXiv:2104.00170

Search...

Help | Advanced

Computer Science > Machine Learning

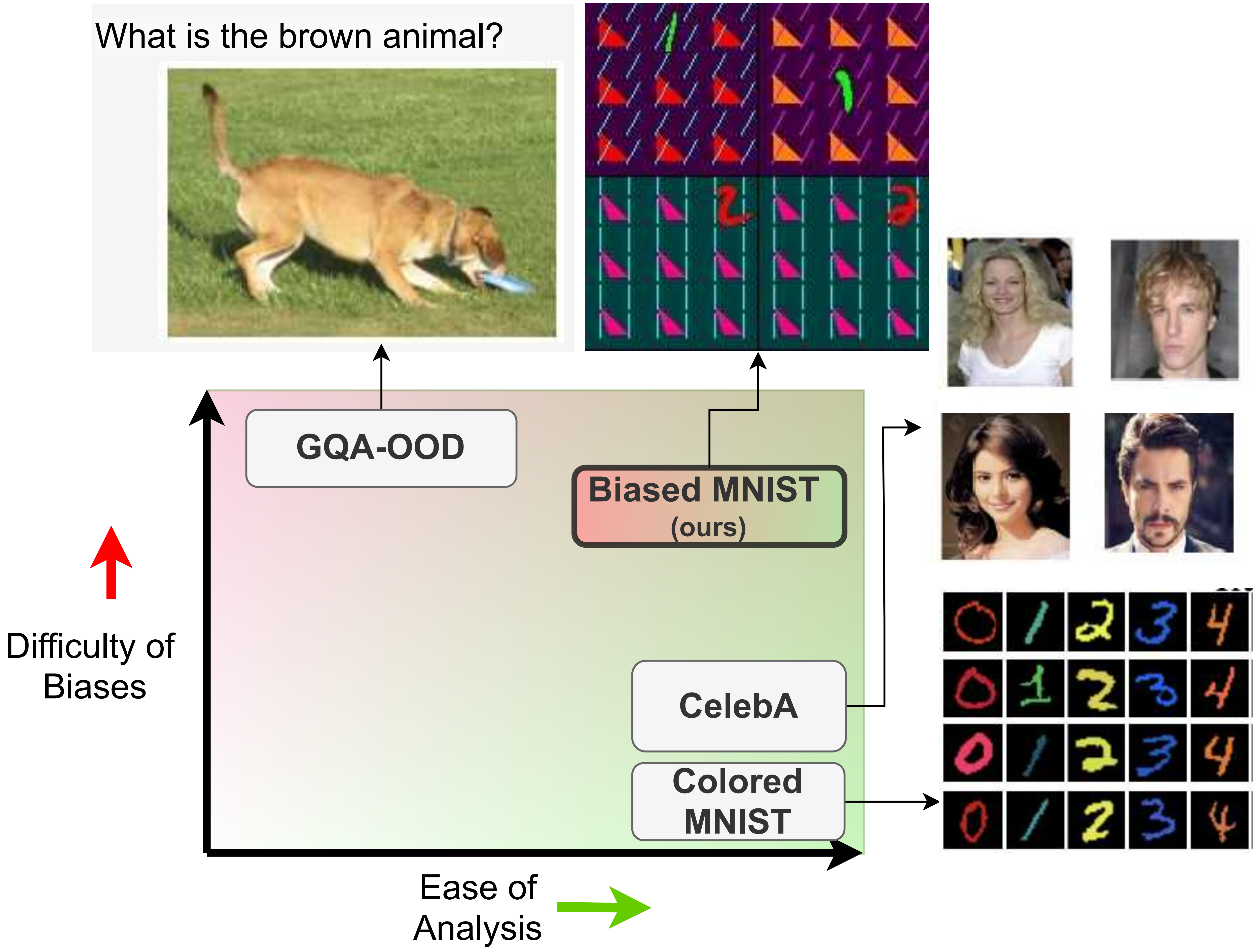
[Submitted on 1 Apr 2021]

An Investigation of Critical Issues in Bias Mitigation Techniques

Robik Shrestha, Kushal Kafle, Christopher Kanan

https://arxiv.org/abs/2104.00170

- Learning inappropriate biases can cause DL models to perform badly on minority groups
- Several methods were developed to address this, but do they work?
- Here:
  - Improved evaluation protocol & dataset
  - Evaluation of 7 methods
  - Biased MNIST dataset
- Code and data: https://github.com/erobic/bias-mitigators



arXiv.org > cs > arXiv:2104.00170

Computer Science > Machine Learning

[Submitted on 1 Apr 2021]

An Investigation of Critical Issues in Bias Mitigation Techniques

Robik Shrestha, Kushal Kafle, Christopher Kanan

https://arxiv.org/abs/2104.00170

We define two more metrics to help measure bias resistance. **Majority/Minority Difference (MMD)** simply measures the difference between majority and minority groups:

$$MMD = [Acc_{majority} - Acc_{minority}].$$

High MMD indicates that methods rely on factors that work for majority groups, but not for minority groups. The sec-

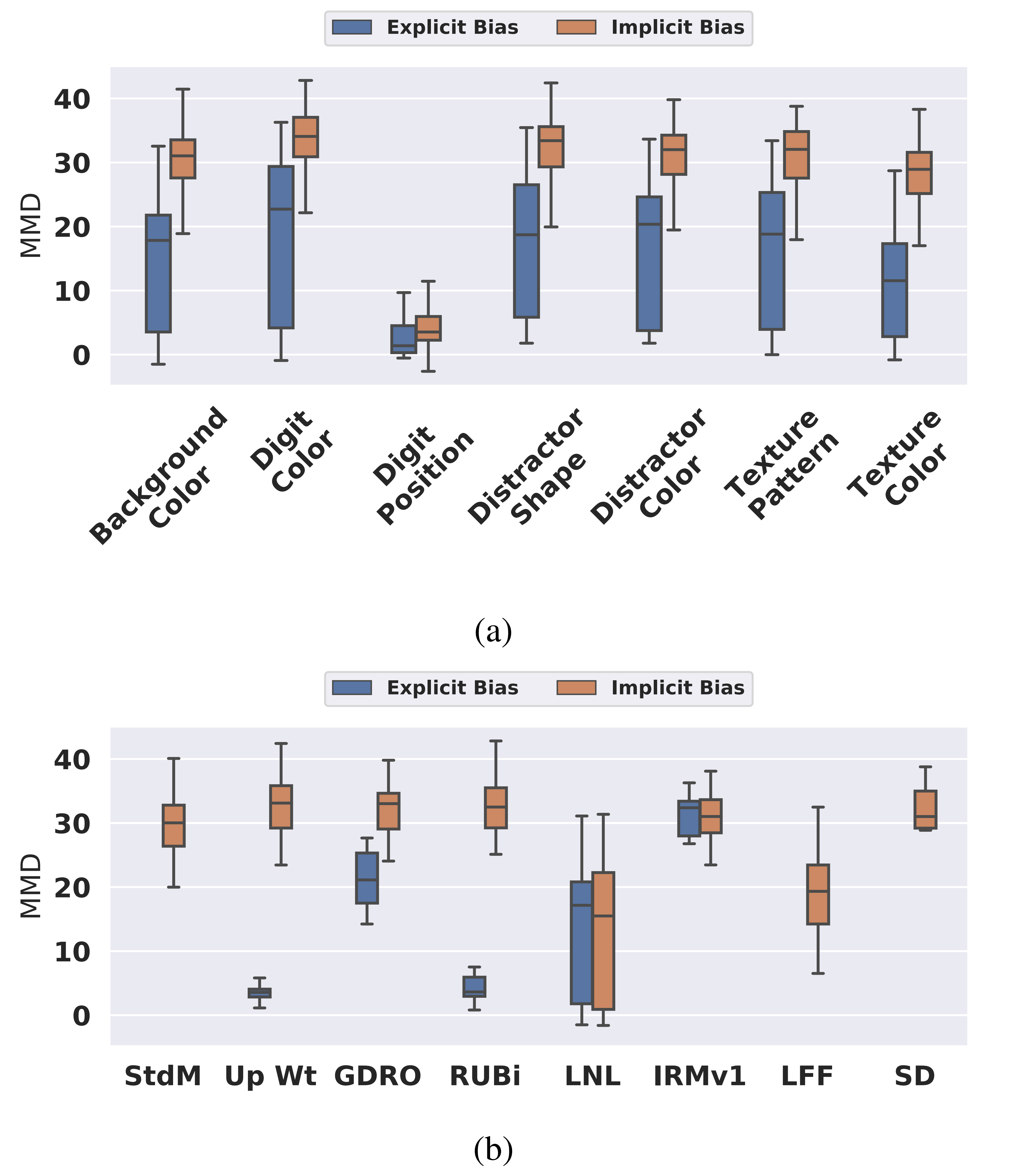


Figure 3: Boxplots of differences between majority and minority groups (MMD) on Biased MNIST over: a) bias variables and b) different methods.

Part 4

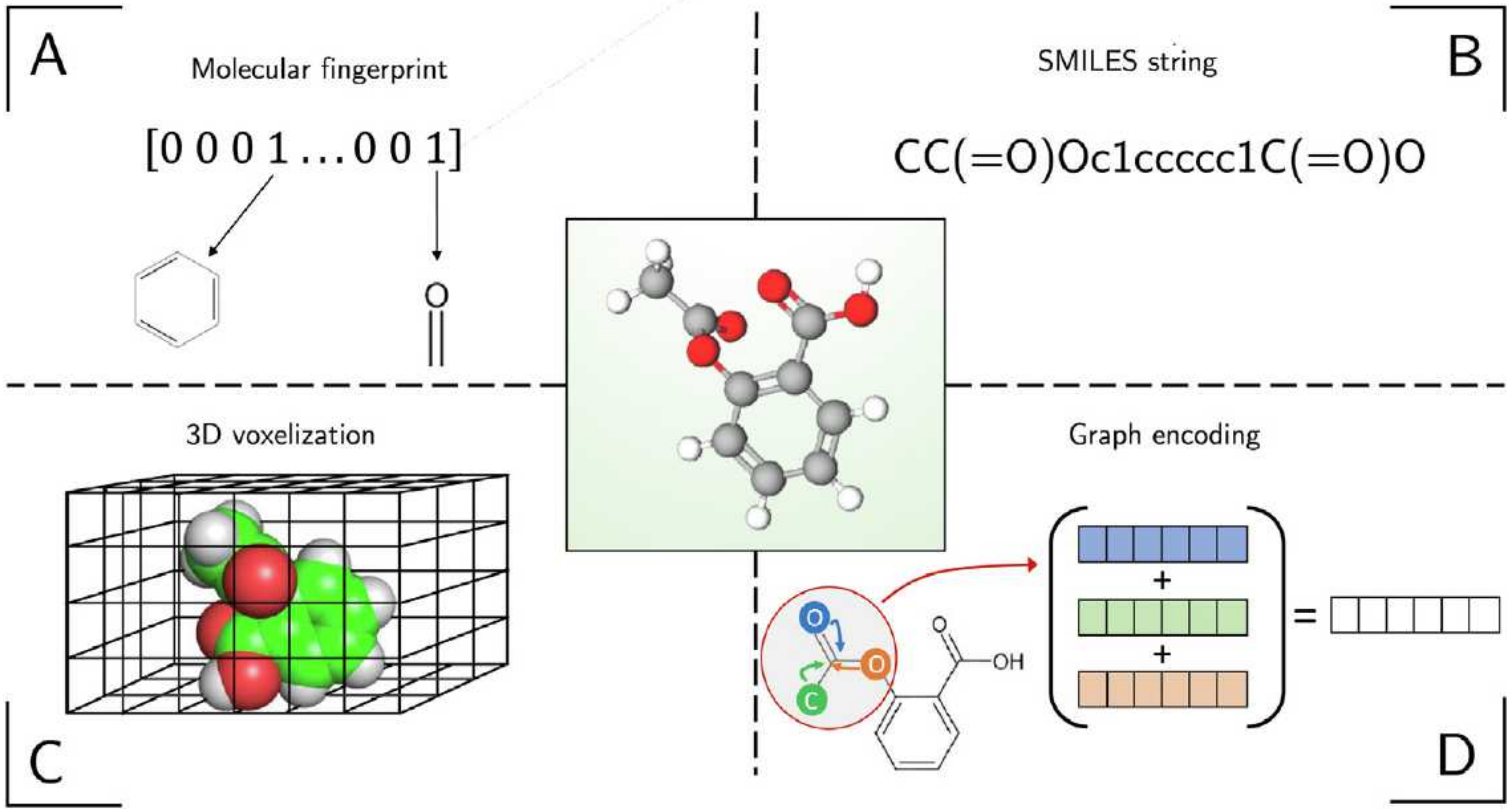
(4) Recent Trends

Graphs

Self-supervised Learning

Transformers

Why Are Graph Neural Nets Interesting?



Sebastian Raschka and Benjamin Kaufman (2020)
Machine Learning and AI-based Approaches for Bioactive Ligand Discovery and GPCR-ligand Recognition
Elsevier Methods, 180, 89–110
https://www.sciencedirect.com/science/article/pii/S1046202319302762

PyTorch geometric

https://github.com/rusty1s/pytorch_geometric

As of this writing: 82 graph neural net methods already implemented

Self-Supervised Learning

"Assisted Label Learning"

Leverage structure of data to create labels for supervised learning, to utilize large amounts of unlabeled data

1. Create labels (pre-text task) by leveraging structure of the data
2. Pre-train in self-supervised fashion to learn embeddings
3. Fine-tune in transfer learning fashion

Classic Self-Supervised Learning Example

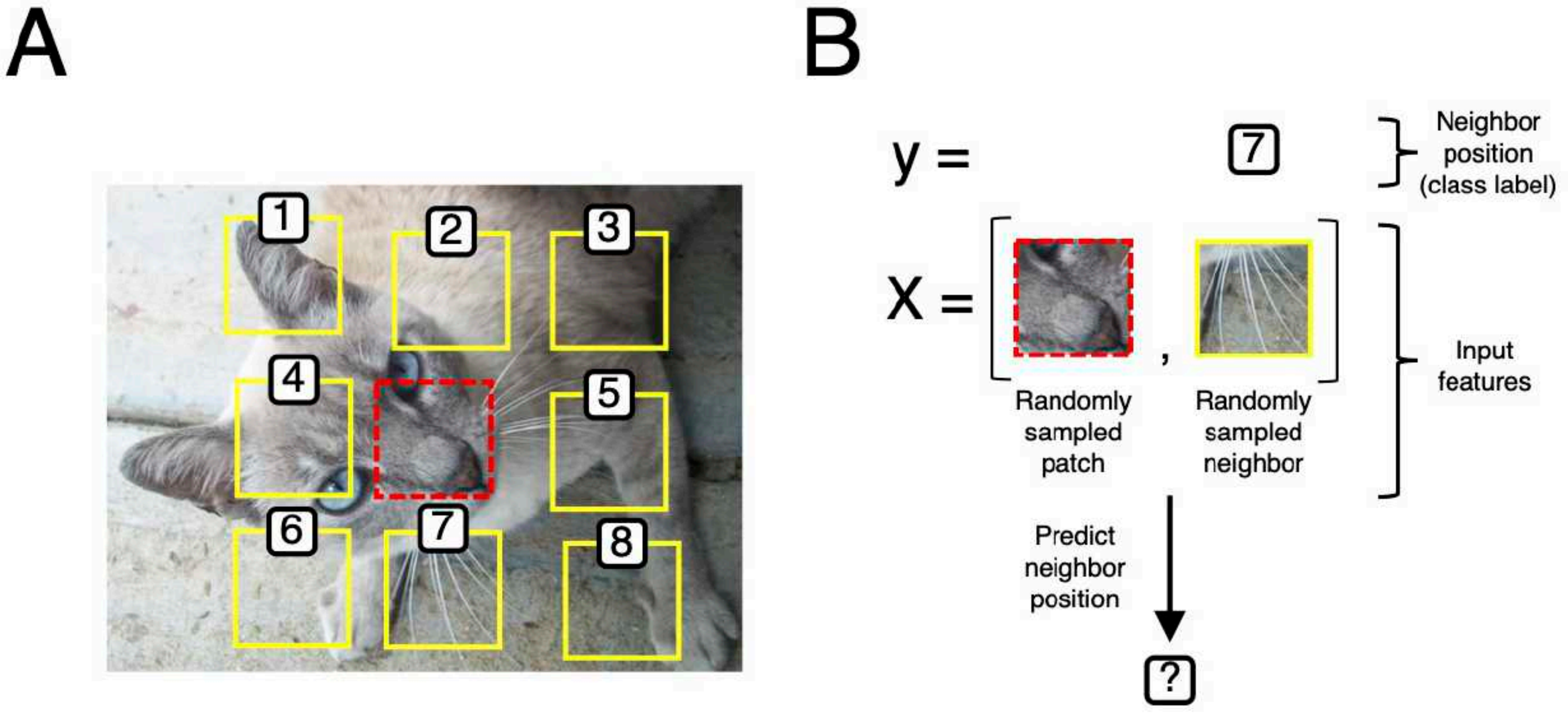


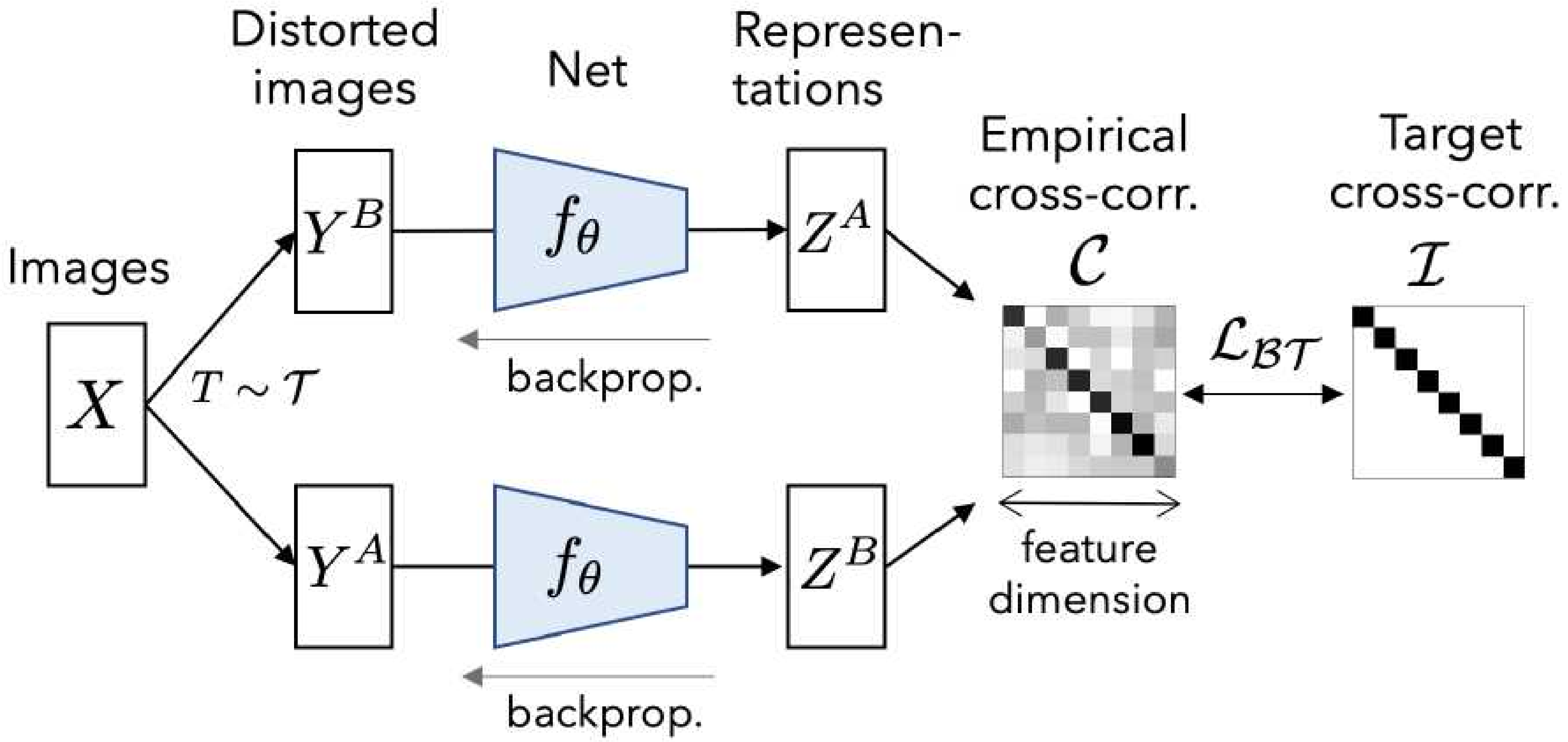
Image source: https://sebastianraschka.com/blog/2020/intro-to-dl-ch01.html

Based on: Doersch, C., Gupta, A., & Efros, A. A.. *Unsupervised visual representation learning by context prediction*. CVPR 2015
https://arxiv.org/abs/1505.05192

Zbontar, Jing, Misra, LeCun, Deny.

Barlow Twins: Self-Supervised Learning via Redundancy Reduction

arXiv:2103.03230, 2021 Mar 4.



1. Run original and distorted image through same network
2. Compute correlation matrix
3. Add objective to make correlation matrix close to identity matrix

↓

Forces representation vectors of similar samples to be similar

https://arxiv.org/abs/2103.03230

Goyal, Caron, Lefaudeux, Xu, Wang, Pai, Singh, Liptchinsky, Misra, Joulin, Bojanowski. **Self-supervised Pretraining of Visual Features in the Wild**. arXiv:2103.01988, 2021 Mar 2.
https://arxiv.org/abs/2103.01988

- SEER = SElf-supERvised
- new billion-parameter self-supervised computer vision model
- pretraining on a **billion** random, **unlabeled** and uncurated public Instagram images
- self-supervised SOTA: reaching 84.2 percent top-1 accuracy on ImageNet
- SwAV (https://arxiv.org/abs/2006.09882) uses online clustering to rapidly group images with similar visual concepts and leverage their similarities (doesn't need pair-wise comparisons; fast)

Self-Supervised Learning (Text Example)

Input sentence: A quick brown fox jumps over the lazy dog

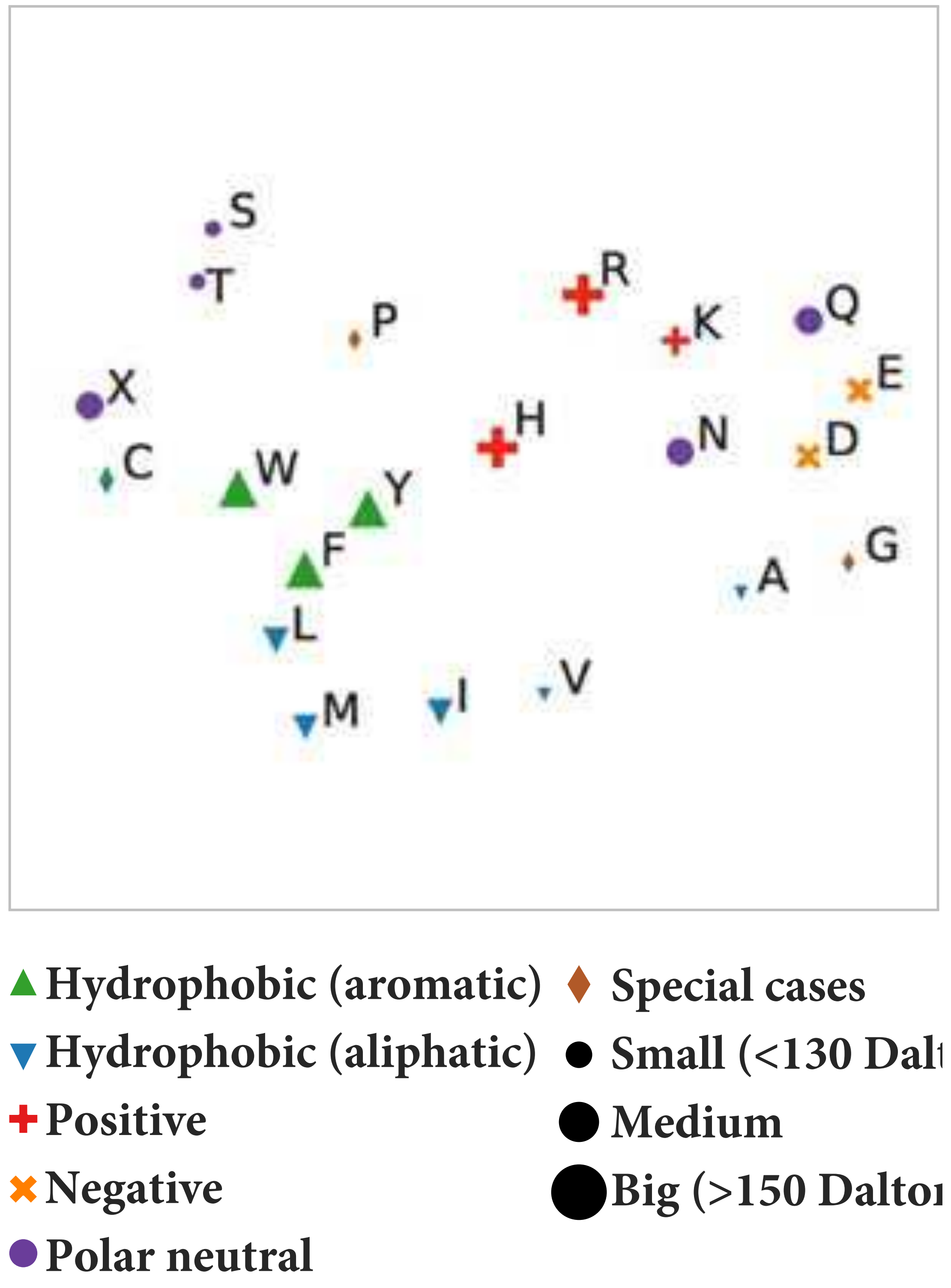
15% randomly masked: A quick brown [MASK] jumps over the lazy dog

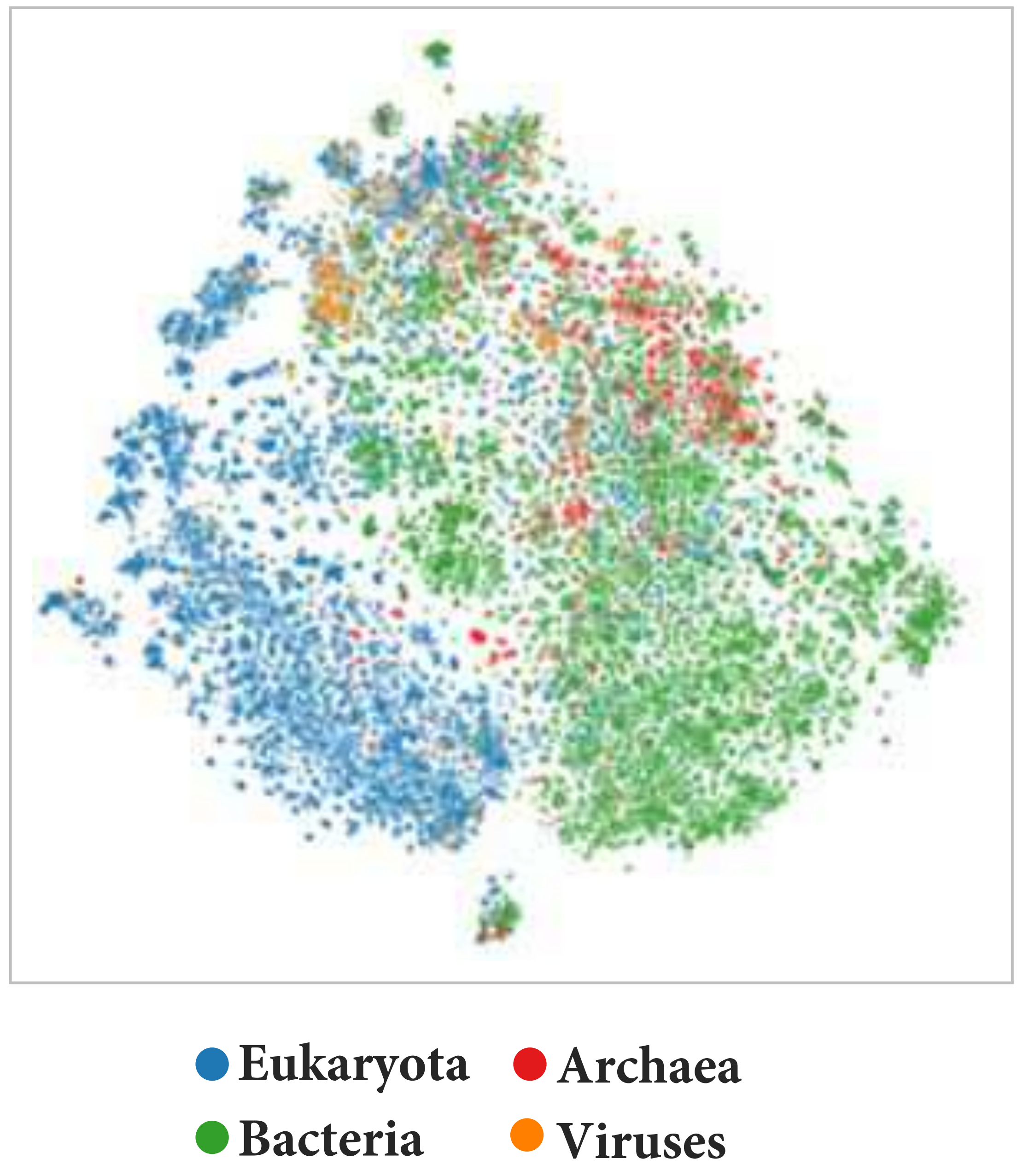
BERT

Possible classes (all words)

0.2%	ant
...	...
11%	fox
...	...
0.01%	zoo

Patterns Emerge When Training on Large Amounts of Unlabeled Amino Acid Sequence Data in Self-Supervised Fashion





Elnaggar A, Heinzinger M, Dallago C, Rihawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Bhowmik D, Rost B. ProtTrans: Towards Cracking the Language of Life's Code Through Self-Supervised Deep Learning and High Performance Computing. arXiv preprint 2020 Jul 13. https://arxiv.org/abs/2007.06225

"Old" Language Transformer Models

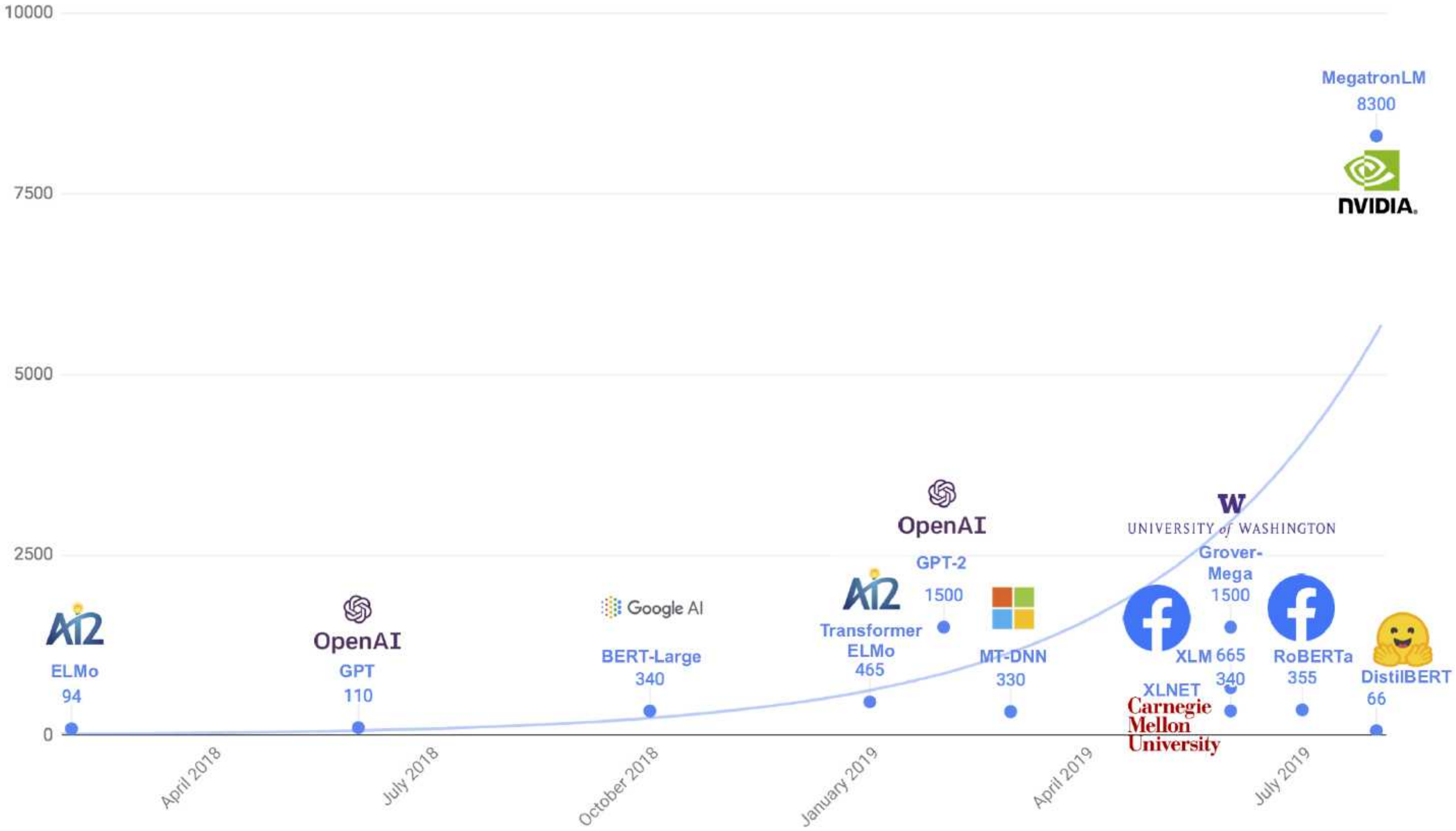


Image Source: https://medium.com/huggingface/distilbert-8cf3380435b5

THE COST OF TRAINING NLP MODELS
A CONCISE OVERVIEW

Or Sharir
AI21 Labs
ors@ai21.com

Barak Peleg
AI21 Labs
barakp@ai21.com

Yoav Shoham
AI21 Labs
yoavs@ai21.com

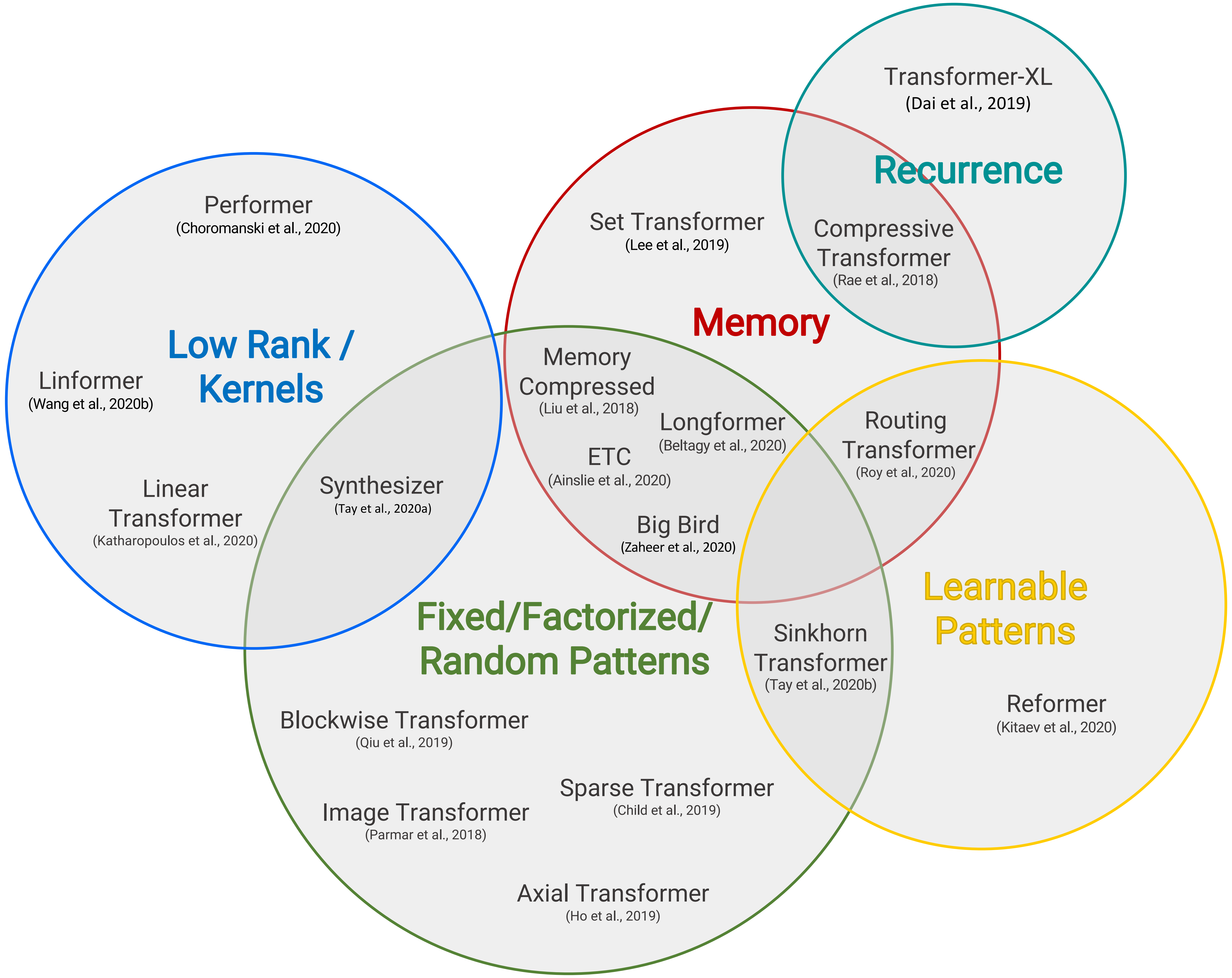
April 2020

http://arxiv.org/abs/2004.08900

Costs: Not for the faint hearted

- $2.5k - $50k (110 million parameter model)
- $10k - $200k (340 million parameter model)
- $80k - $1.6m (1.5 billion parameter model)

In Parallel: Increased Focus on Making Transformers Accessible



Tay, Dehghani, Bahri, Metzler. **Efficient Transformers**: A Survey. arXiv:2009.06732, 2020
https://arxiv.org/abs/2009.06732

"Transformers for Computer Vision" is a Fast Growing Field

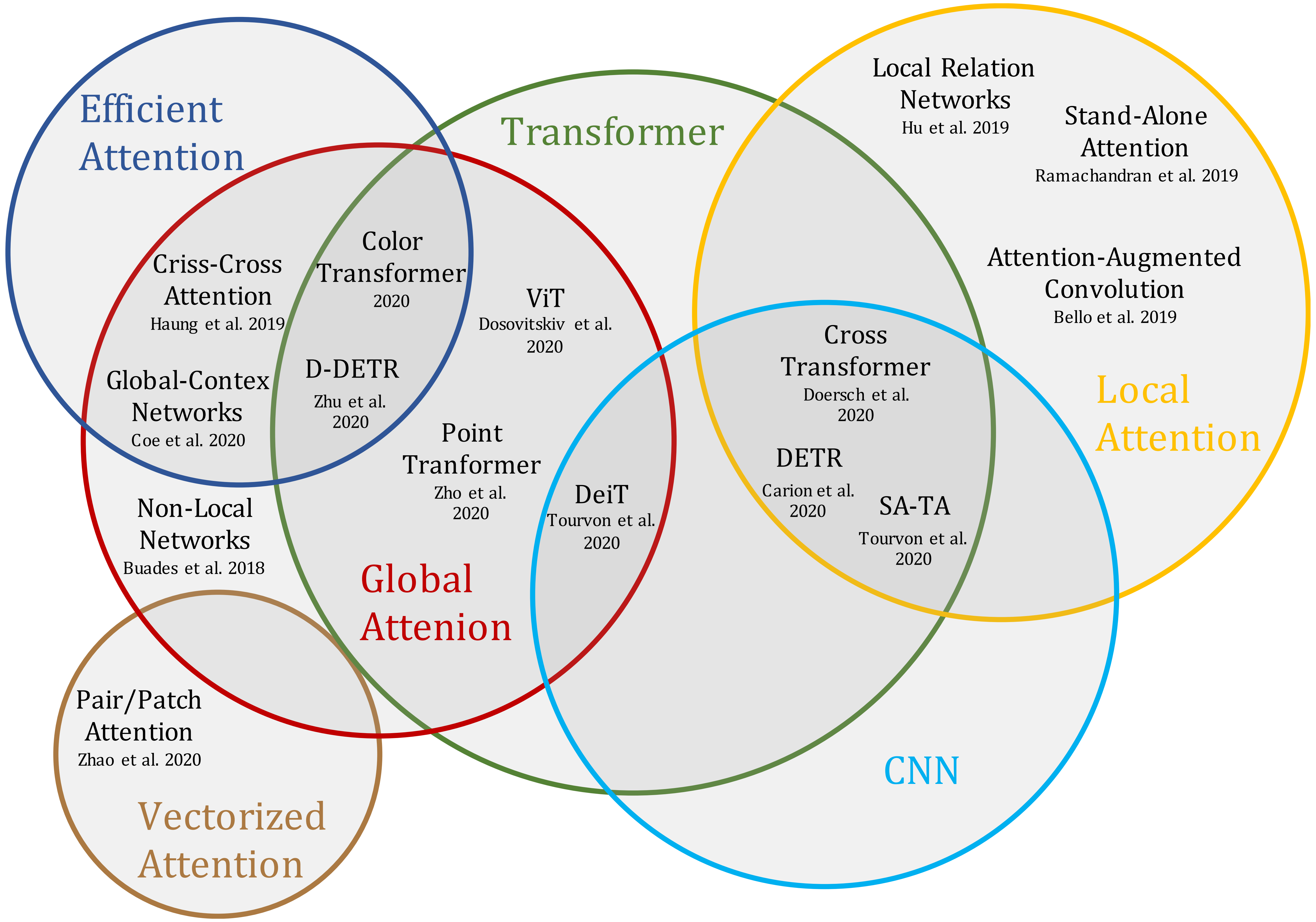


Fig. 3. A taxonomy of self-attention design space.

Khan, Naseer, Hayat, Zamir, Khan, Shah. **Transformers in Vision: A Survey**. arXiv preprint arXiv:2101.01169. 2021 Jan. https://arxiv.org/abs/2009.06732

arXiv.org > cs > arXiv:2104.00298

Search...

Help | Advanced S

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 1 Apr 2021]

EfficientNetV2: Smaller Models and Faster Training

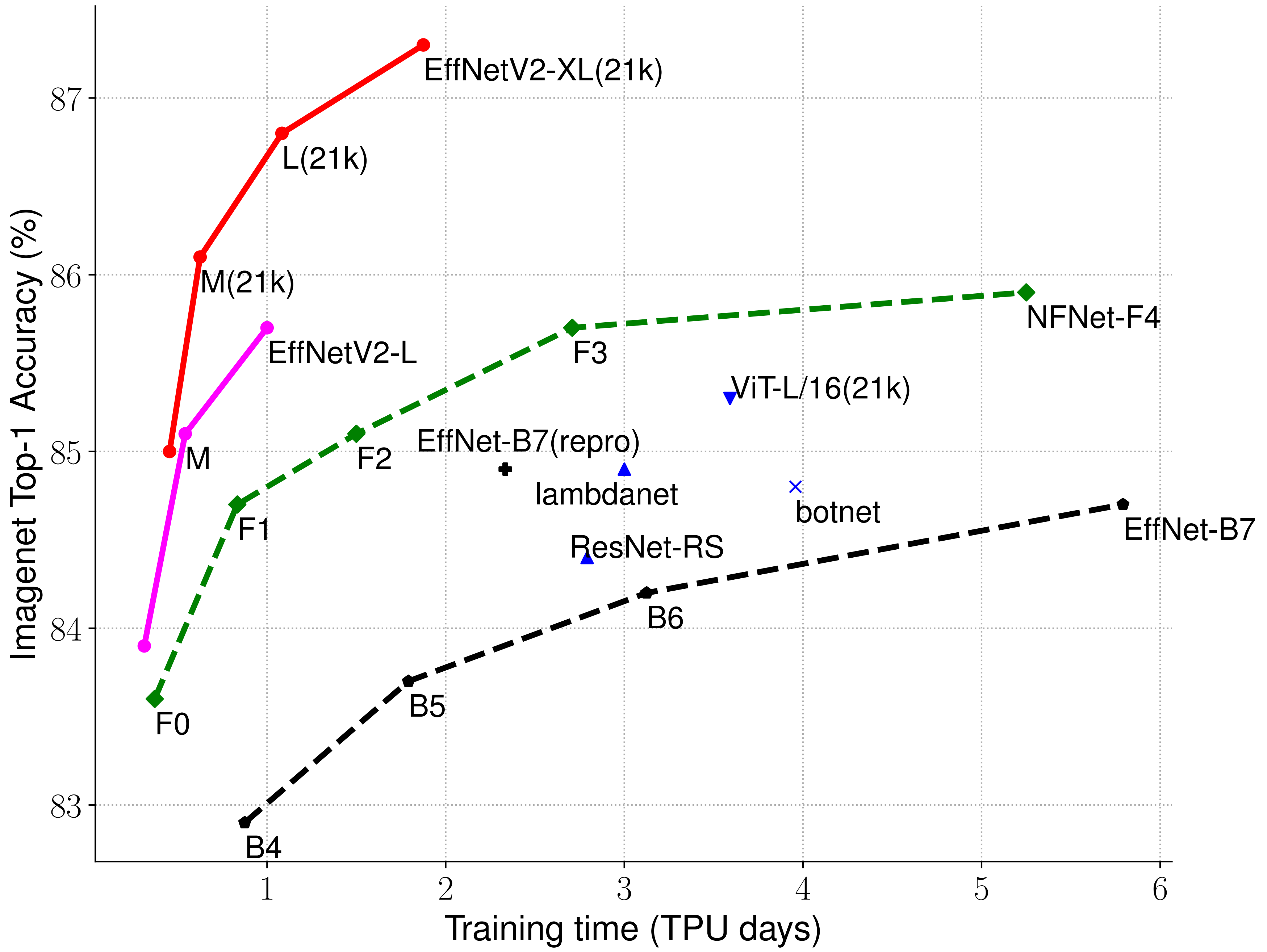
Mingxing Tan, Quoc V. Le

https://arxiv.org/abs/2104.00298

CNNs remain relevant for image data

- **EfficientNetV2:**
  Large improvement over EfficientNets V1
  Also beats Visual Transformers ;)

- Introduces
  new ops such as Fused-MBConv
  progressive increasing of image size during training
  -> adaptively adjusting regularization via dropout and data augmentation



(a) Training efficiency.

	EfficientNet (2019)	ResNet-RS (2021)	DeiT/ViT (2021)	EfficientNetV2 (ours)
Top-1 Acc.	84.3%	84.0%	83.1%	83.9%
Parameters	43M	164M	86M	24M

(b) Parameter efficiency.

Contact:

https://sebastianraschka.com

@rasbt

Sebastian Raschka

Slides: http://sebastianraschka.com/pdf/slides/2021-04_czi.pdf