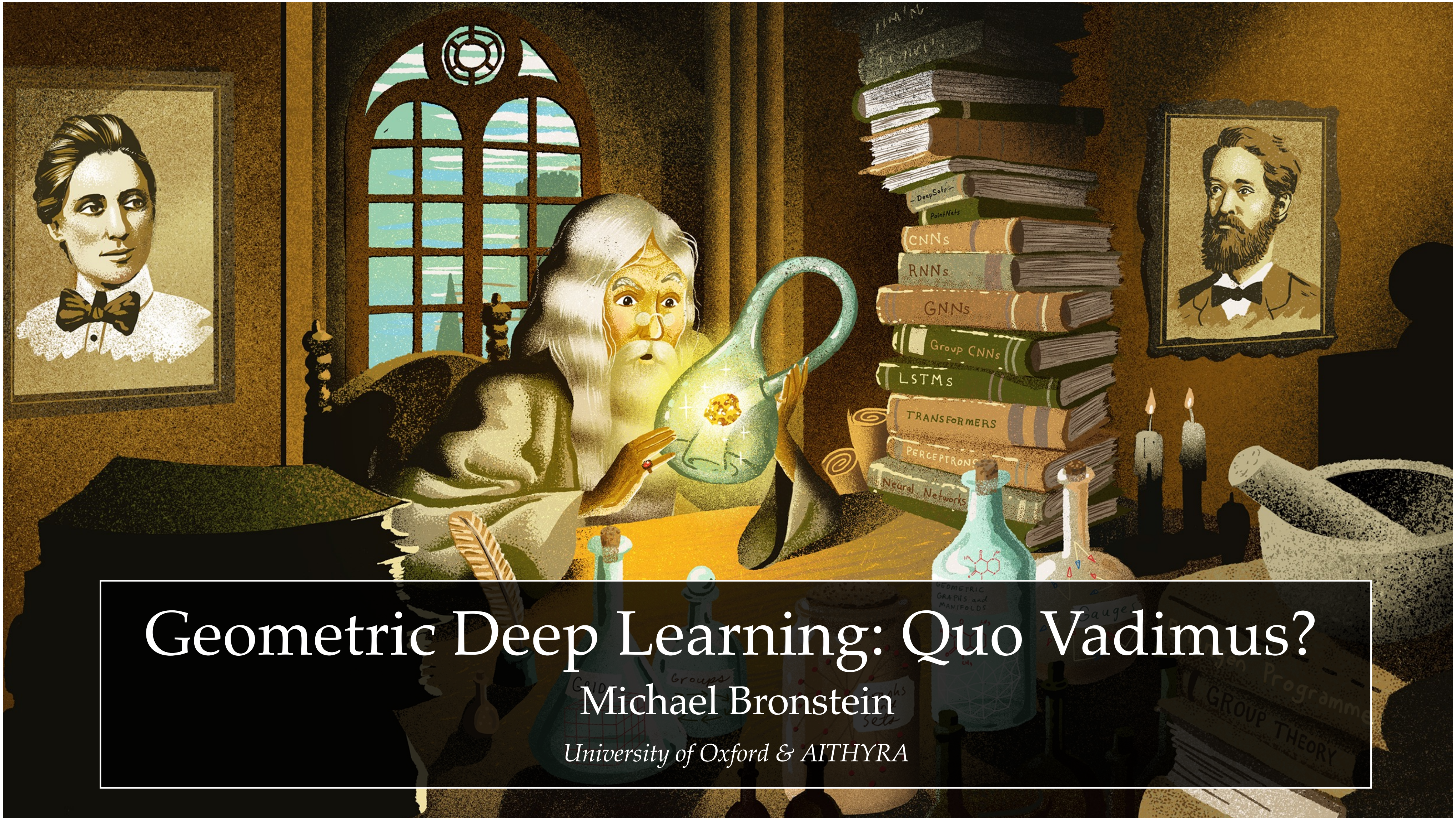


Geometric Deep Learning: Quo Vadimus?

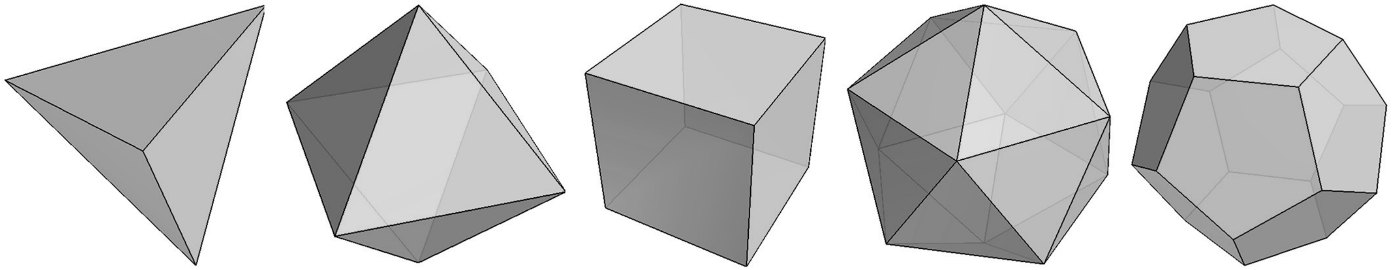
Michael Bronstein

University of Oxford & AITHYRA

“Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection”

H. Weyl

Weyl 1952



Plato

~370 BC

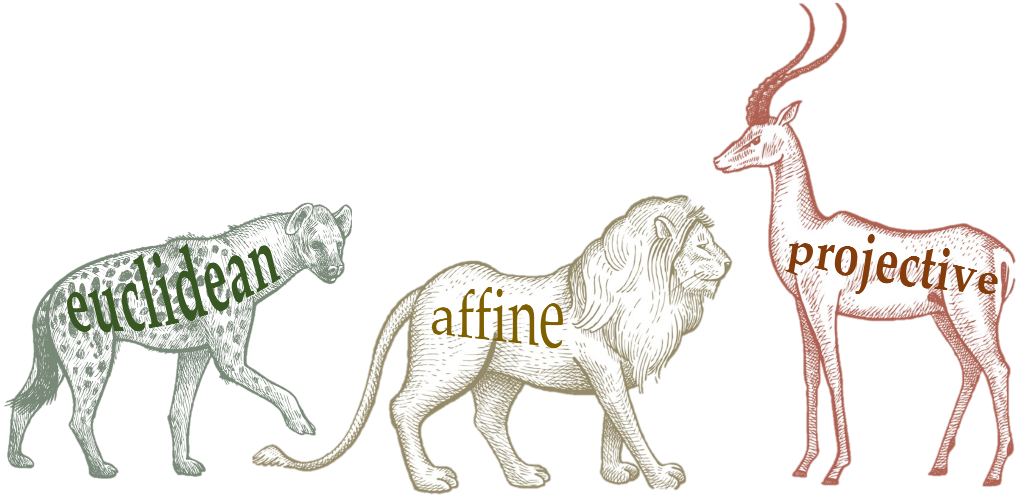
Portrait: Ihor Gorskyi

Oxford Math Institute

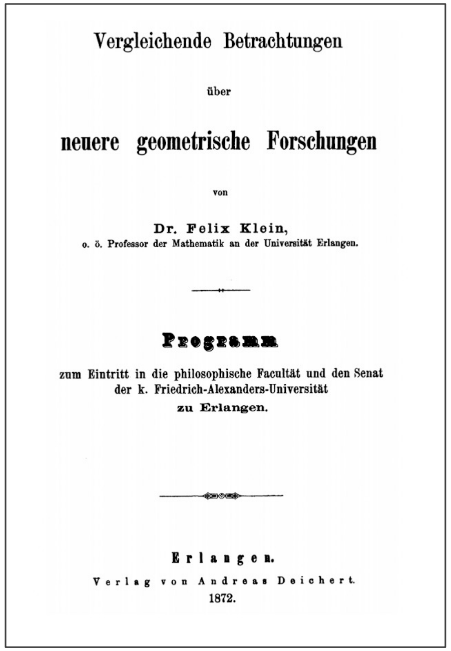
XIX century

euclidean
affine
projective
ΕΥΚΛΕΙΔΗΣ

The Erlangen Programme



Geometry = space + transformation group



Vergleichende Betrachtungen

über

neuere geometrische Forschungen

von

Dr. Felix Klein,
o. ö. Professor der Mathematik an der Universität Erlangen.

Programm

zum Eintritt in die philosophische Facultät und den Senat
der k. Friedrich-Alexanders-Universität
zu Erlangen.

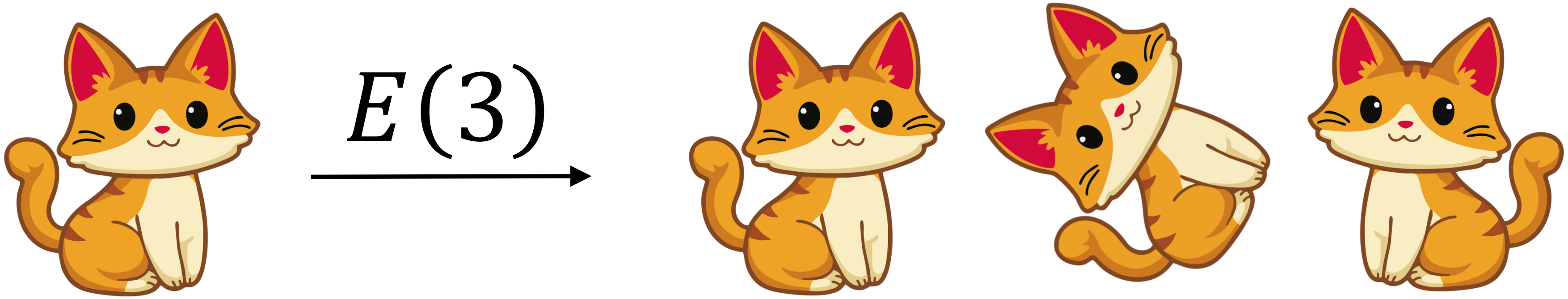
Erlangen.
Verlag von Andreas Deichert.
1872.

F. Klein

1872

Klein 1872

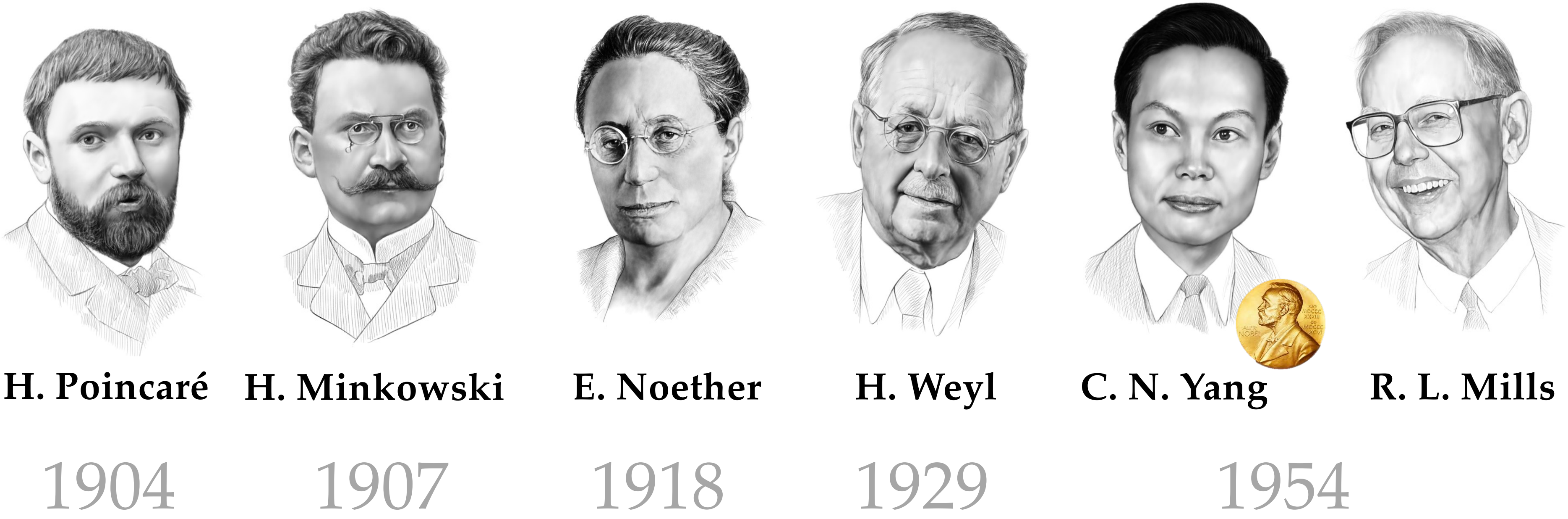
Euclidean geometry



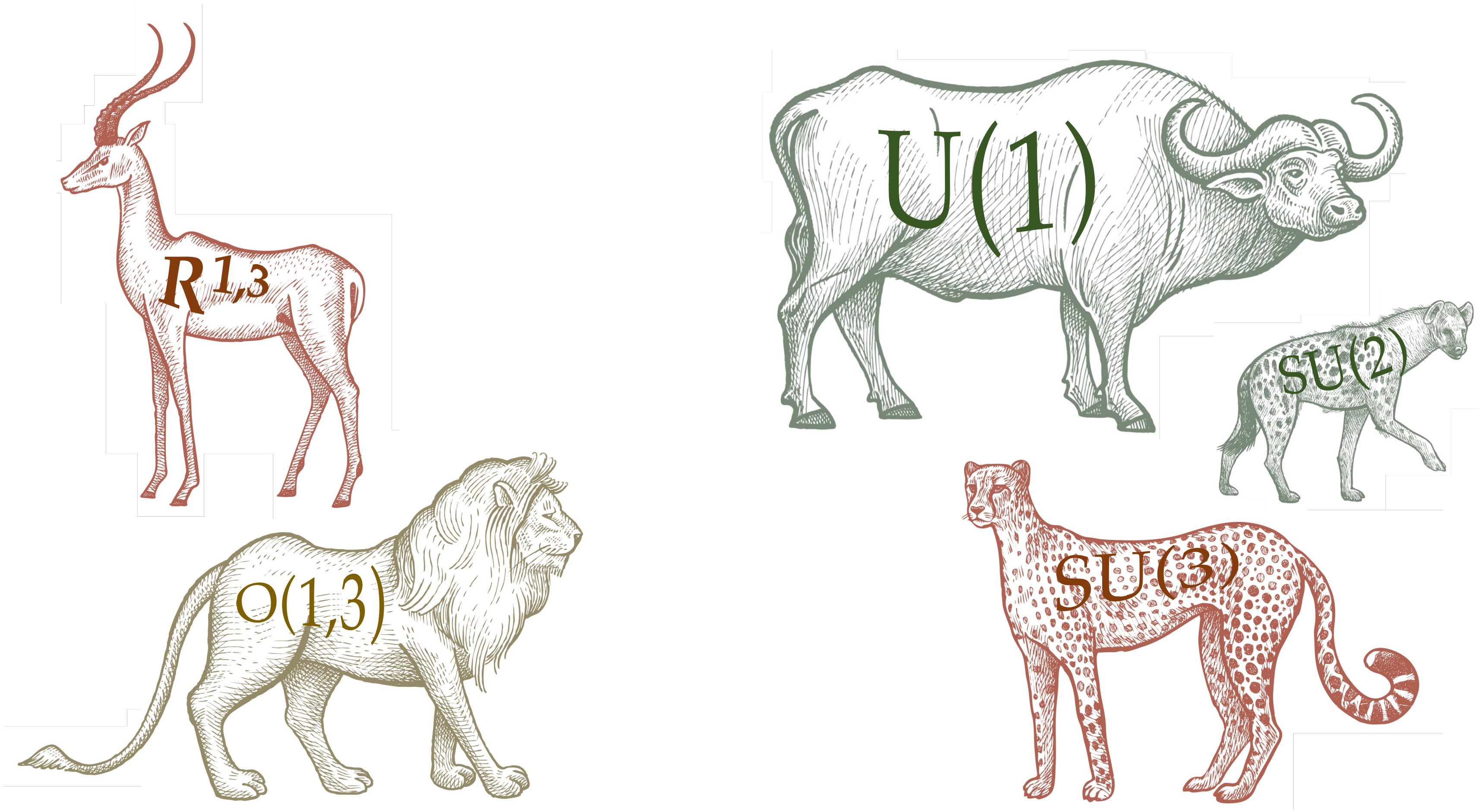
$E(3)$

Translation Rotation Reflection

Klein 1872



Poincaré 1904; Noether 1918; Weyl 1929; Yang & Mills 1954; Portraits: Ihor Gorskyi



External symmetry

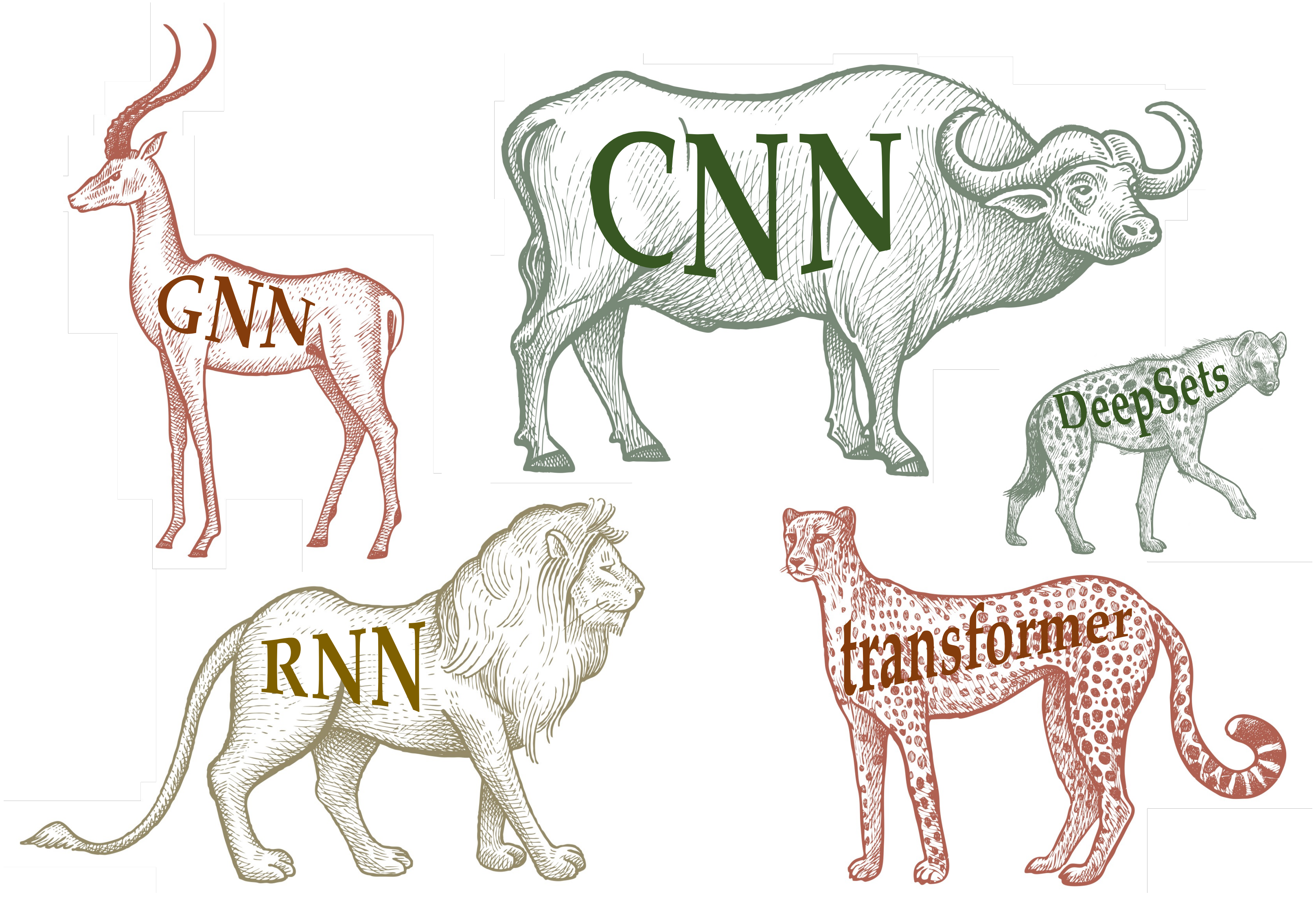
Internal symmetry

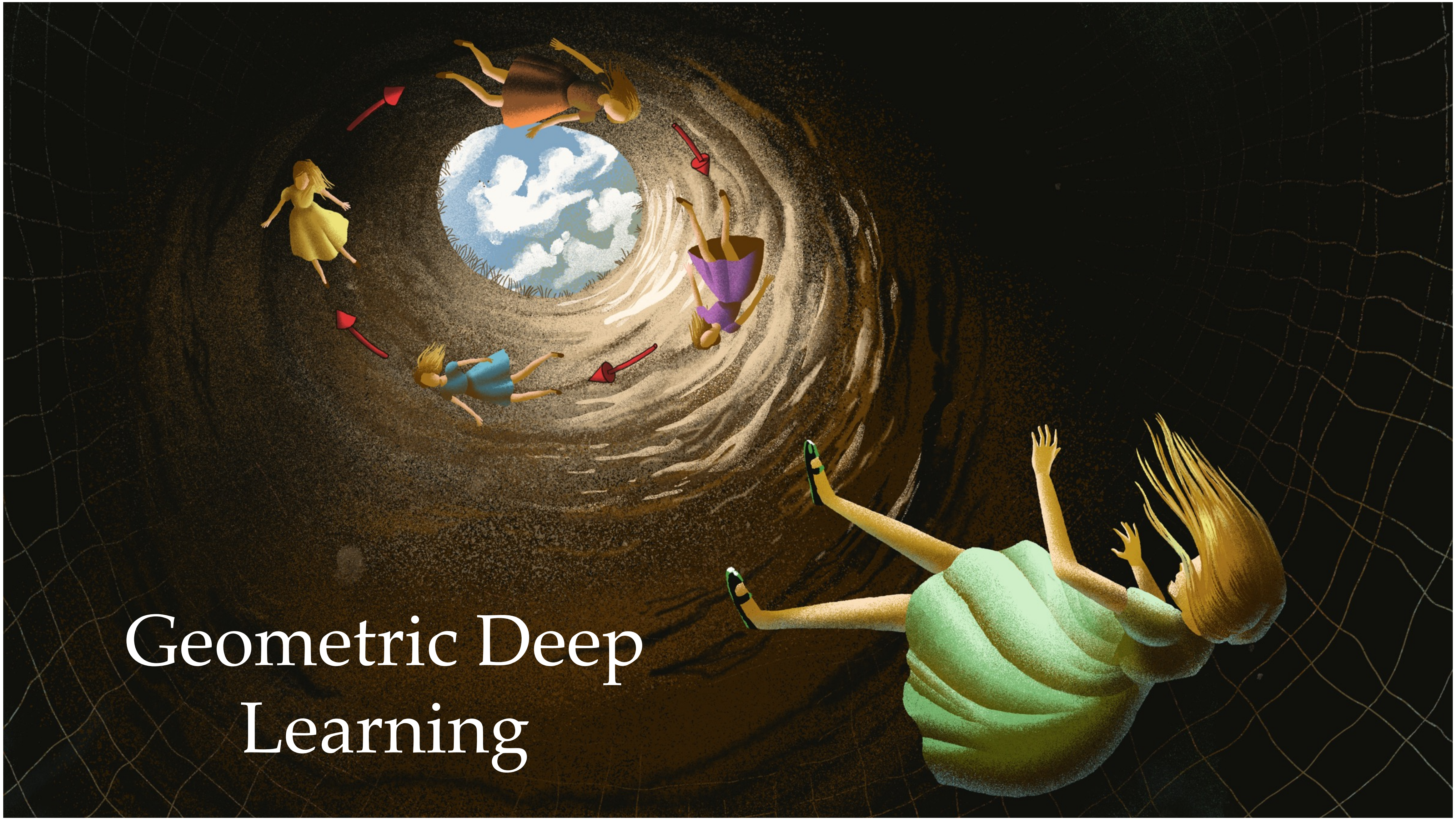
“It is only slightly overstating the case to say that Physics is the study of symmetry”

— *More is different*

P. Anderson

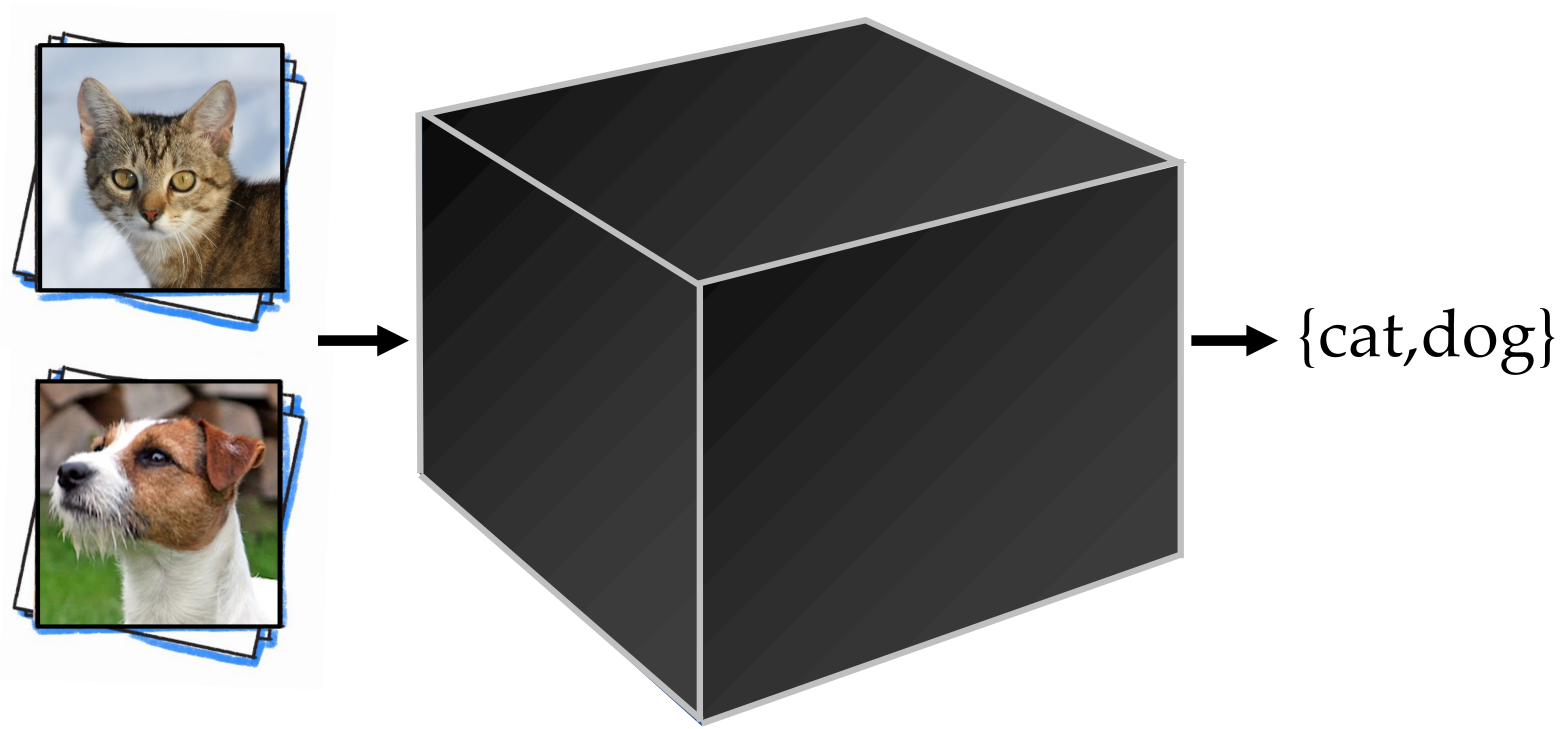
Anderson 1972


GNN
CNN
DeepSets
RNN
transformer



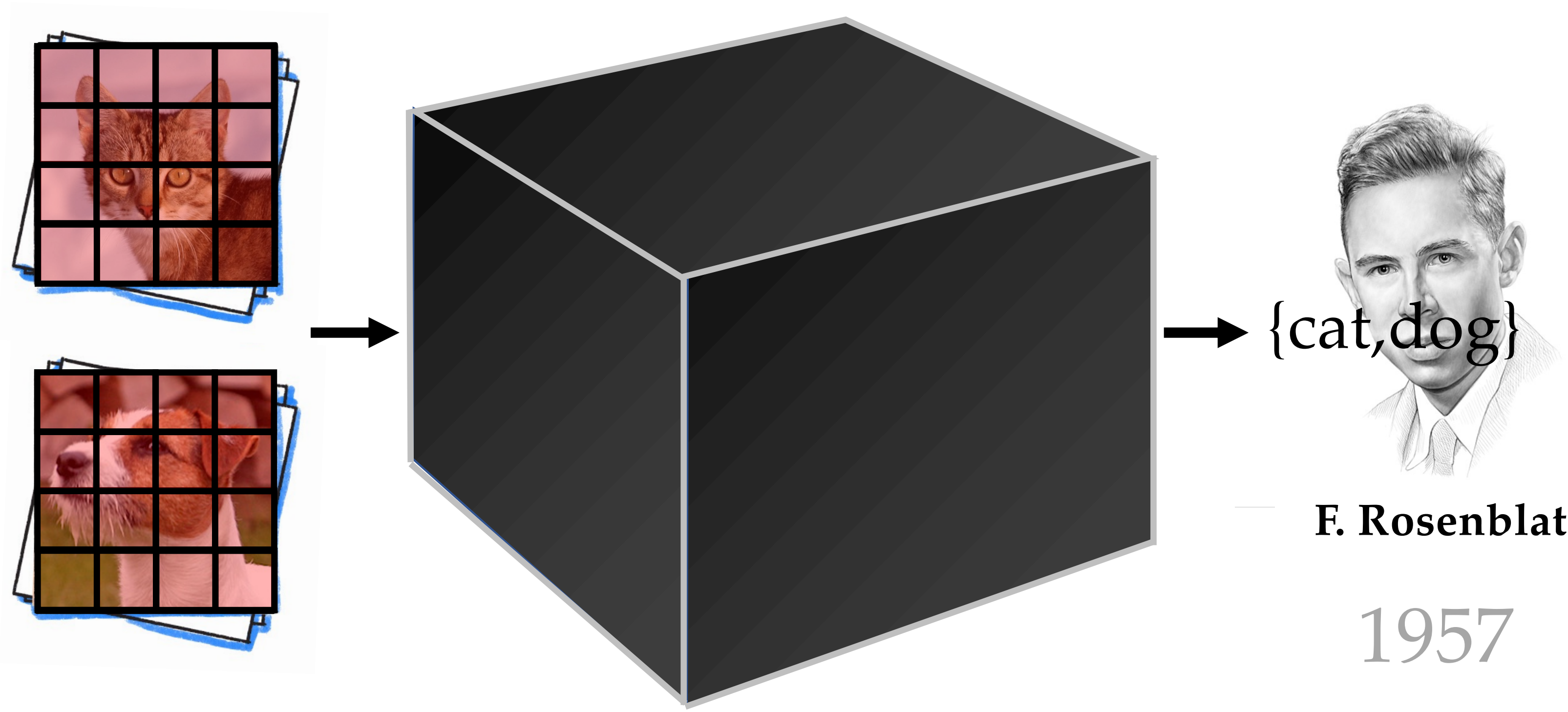
Geometric Deep Learning

Supervised ML = Function Approximation



{cat,dog}

Supervised ML = Function Approximation

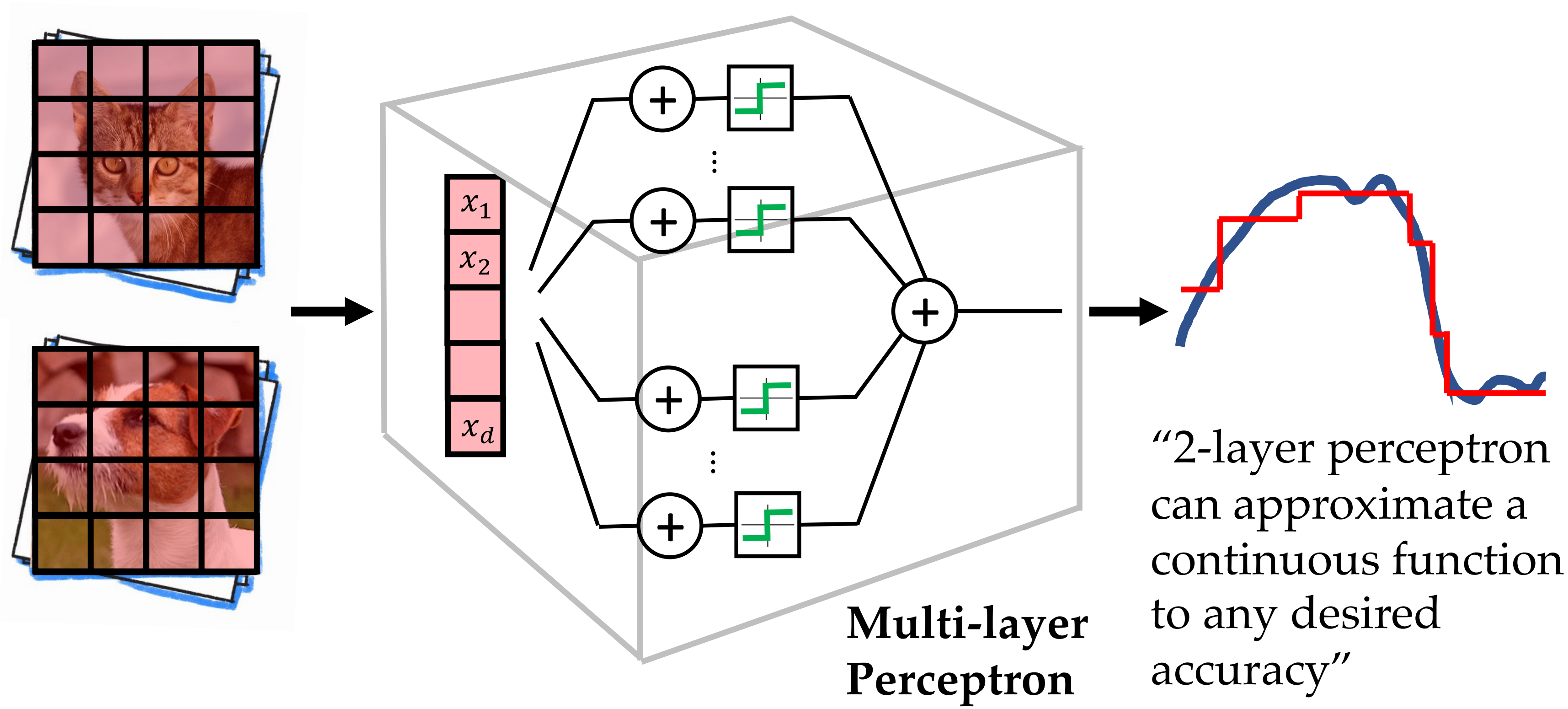


F. Rosenblatt

1957

Rosenblatt 1957; Portrait: Ihor Gorskyi

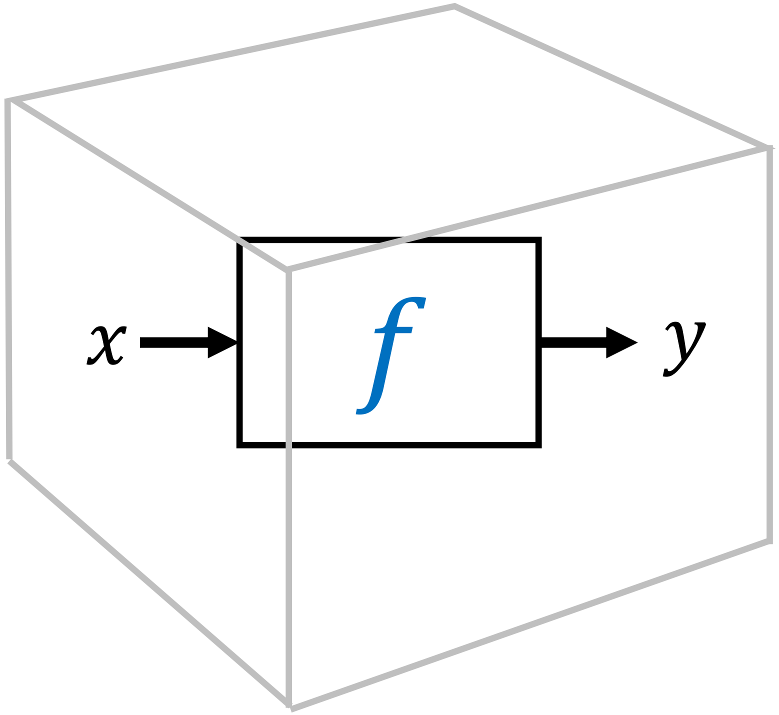
Universal Approximation

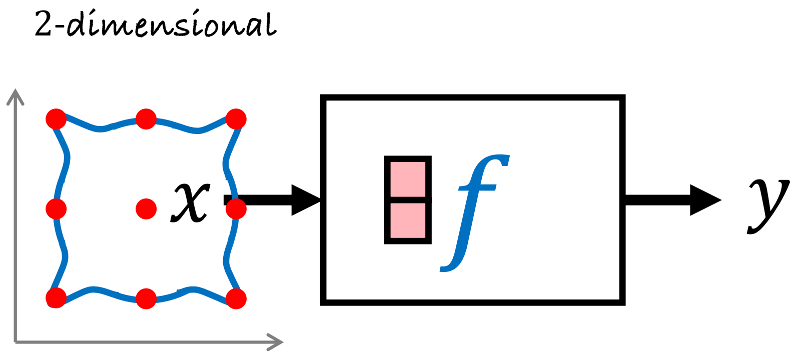


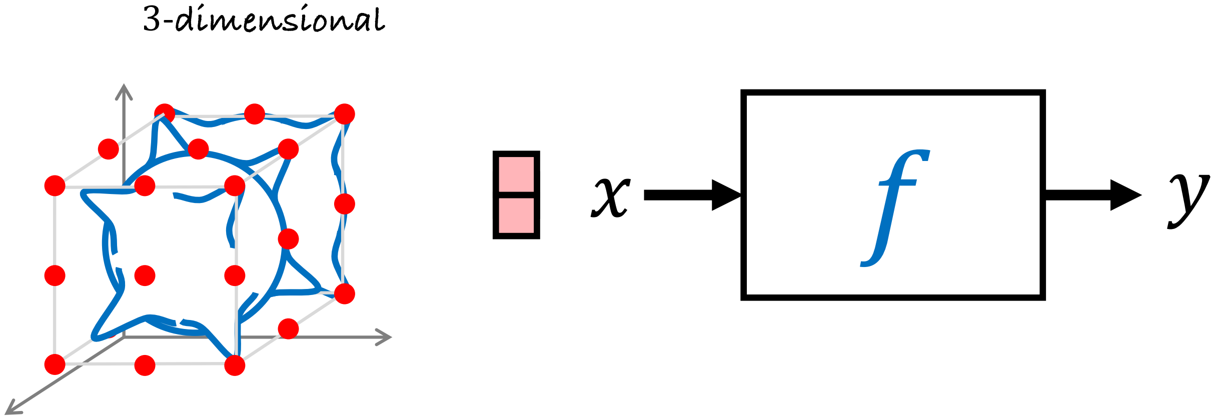
Multi-layer Perceptron

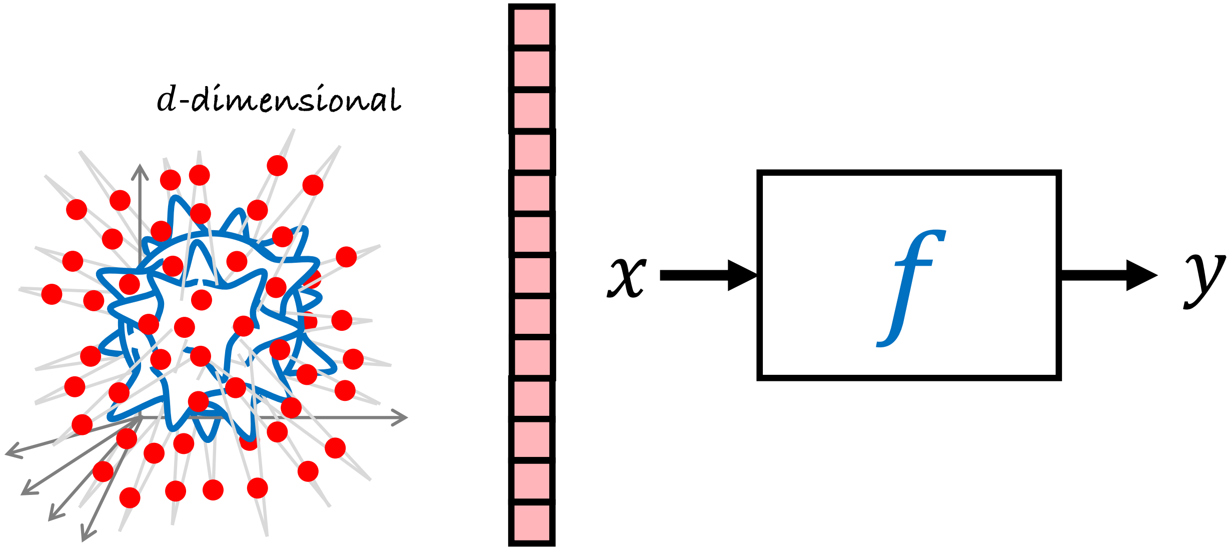
“2-layer perceptron can approximate a continuous function to any desired accuracy”

Universal Approximation: Hilbert’s 13th problem 1900; Kolmogorov 1956; Arnold 1957; Cybenko 1989; Hornik 1991; Barron 1993; Leshno et al 1993; Maiorov 1999; Pinkus 1999

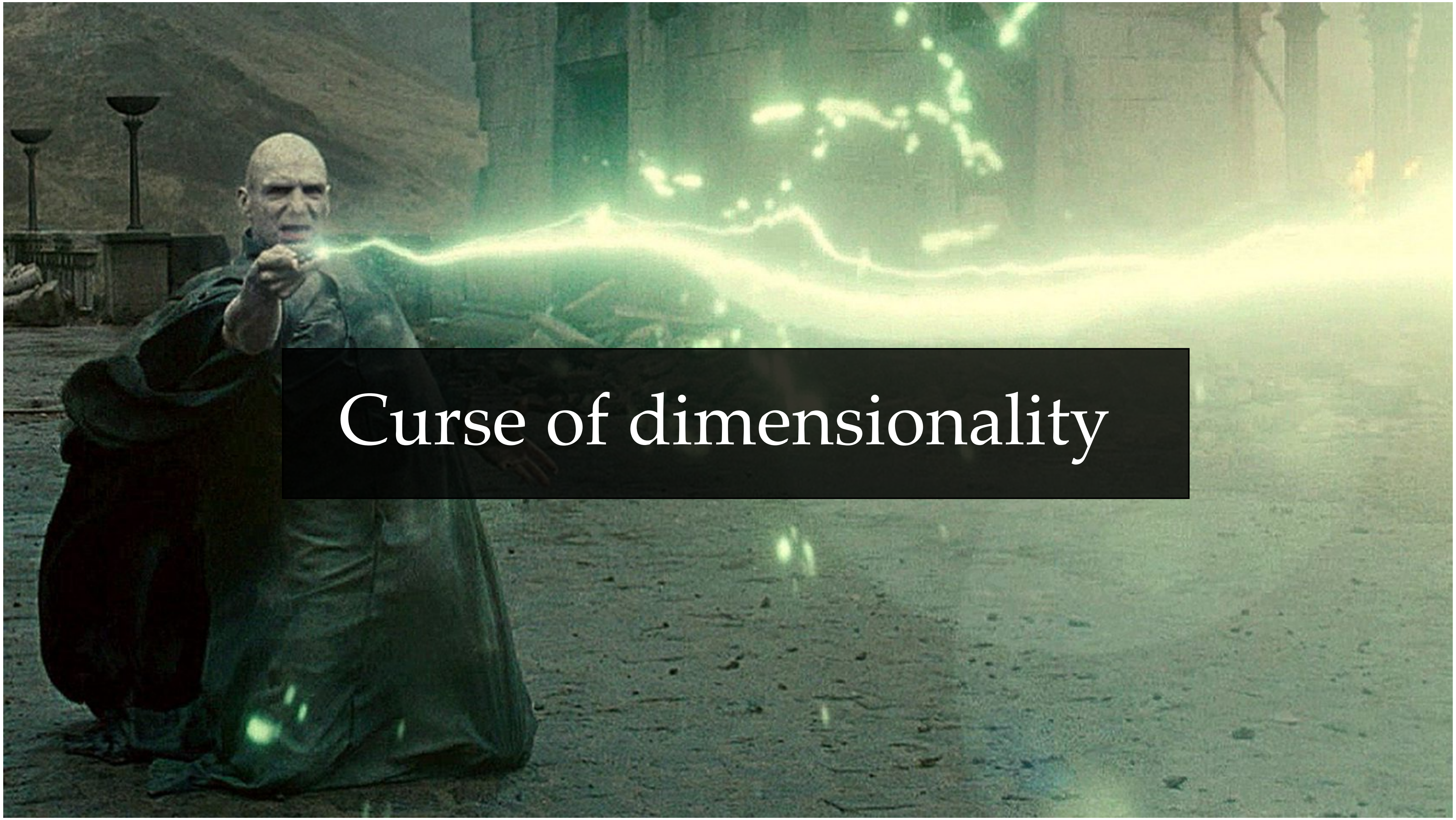

x
f
y


2-dimensional
x
f
y


3-dimensional
x
f
y

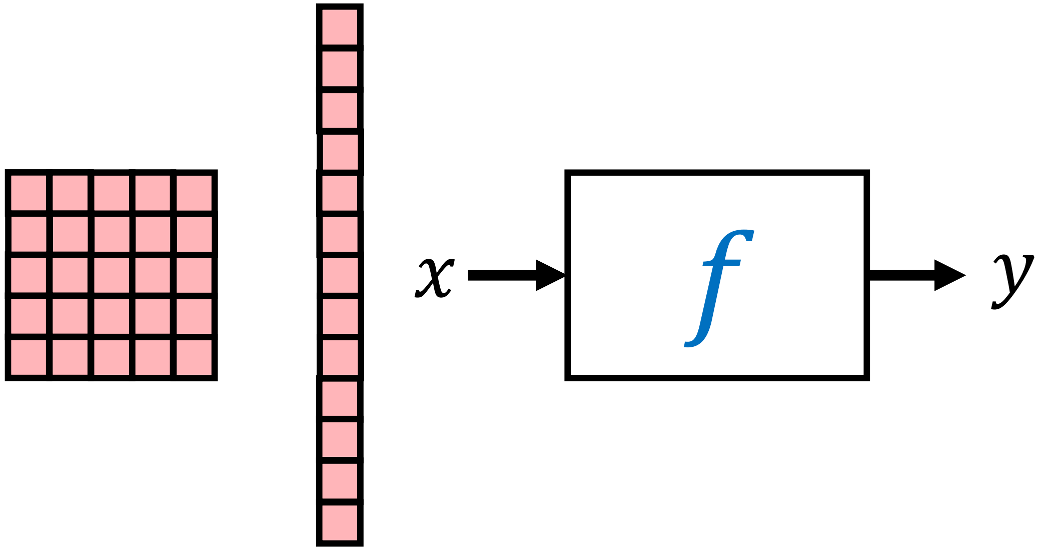


$O(\varepsilon^{-d})$ samples

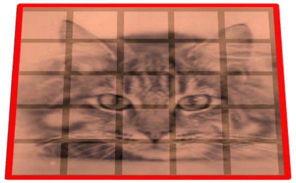


Curse of dimensionality

Geometric priors

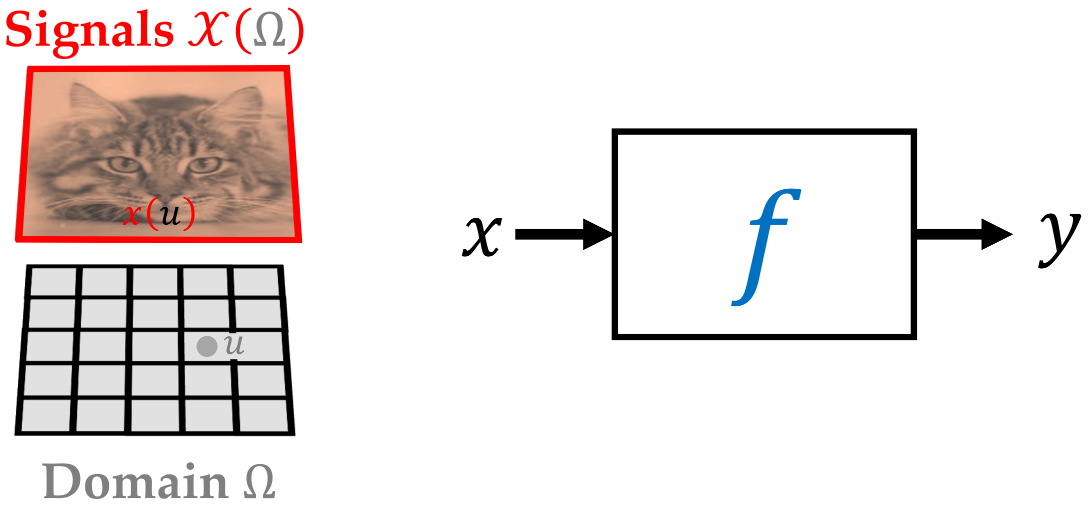


Geometric priors

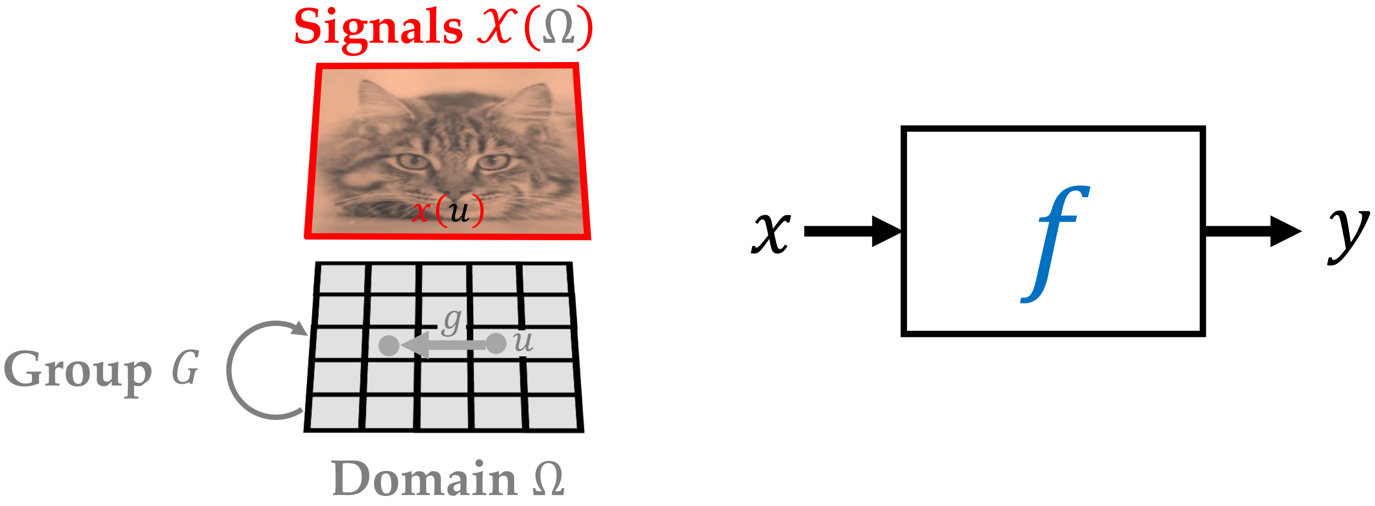


$$x \longrightarrow \boxed{f} \longrightarrow y$$

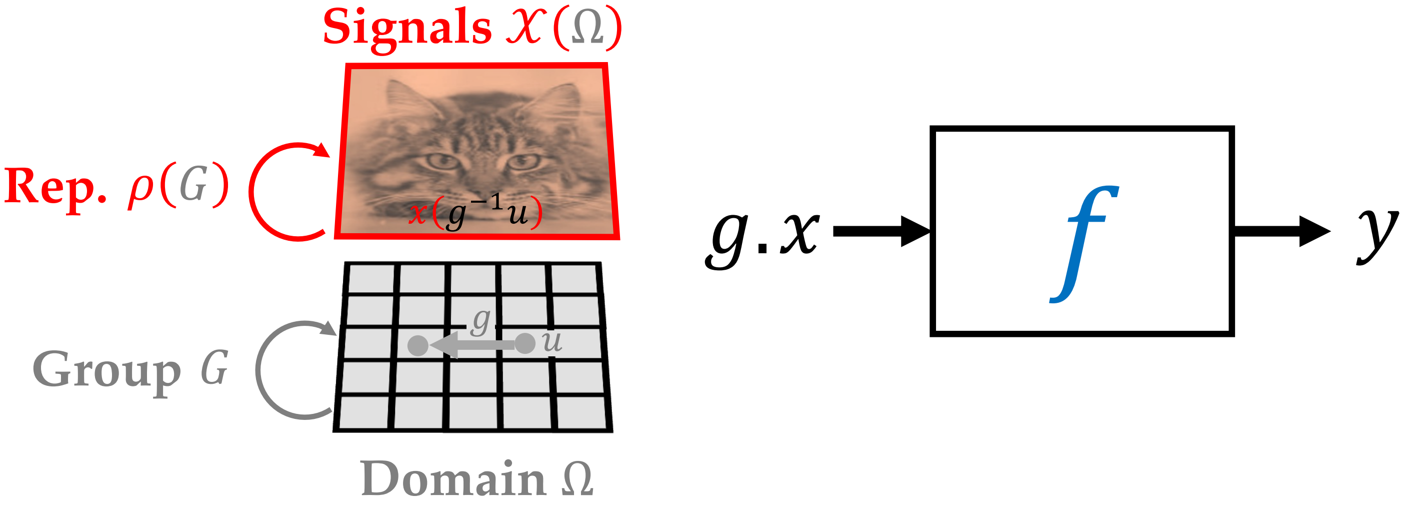
Geometric priors



Geometric priors

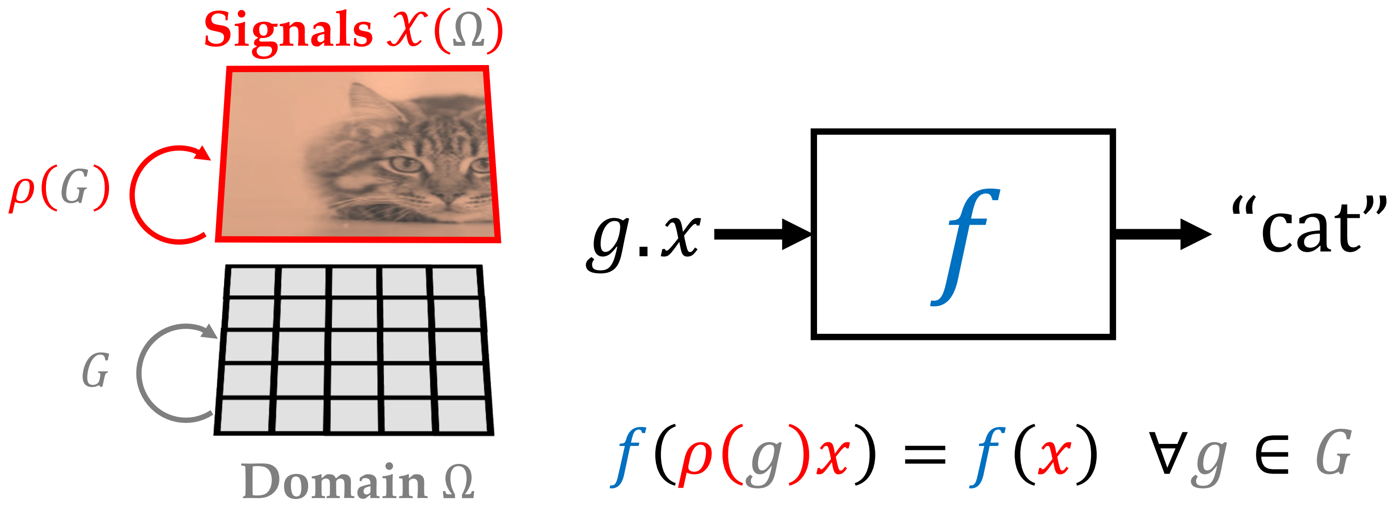


Geometric priors



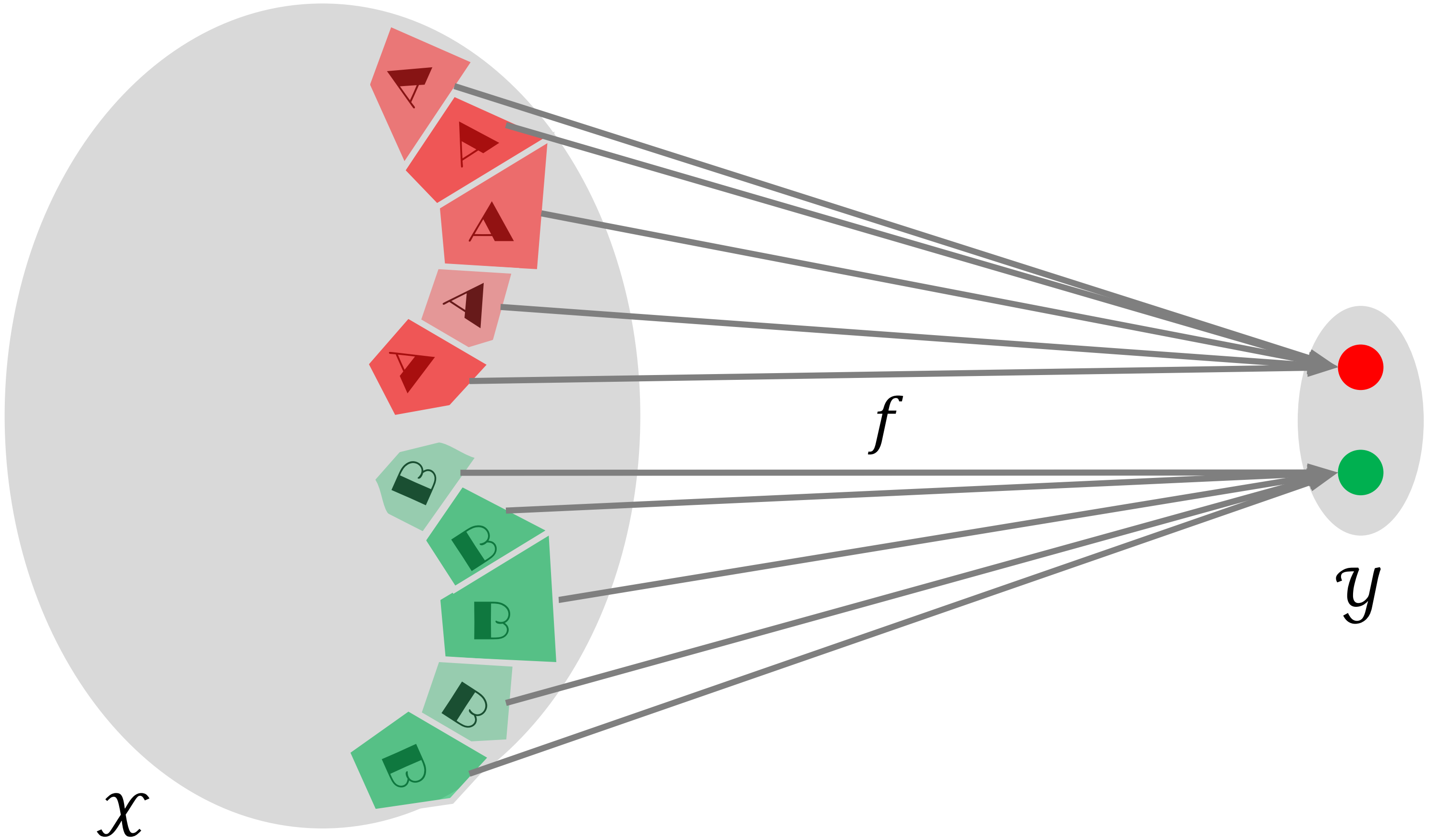
"How f interacts with the group G?"

Geometric priors: Invariance

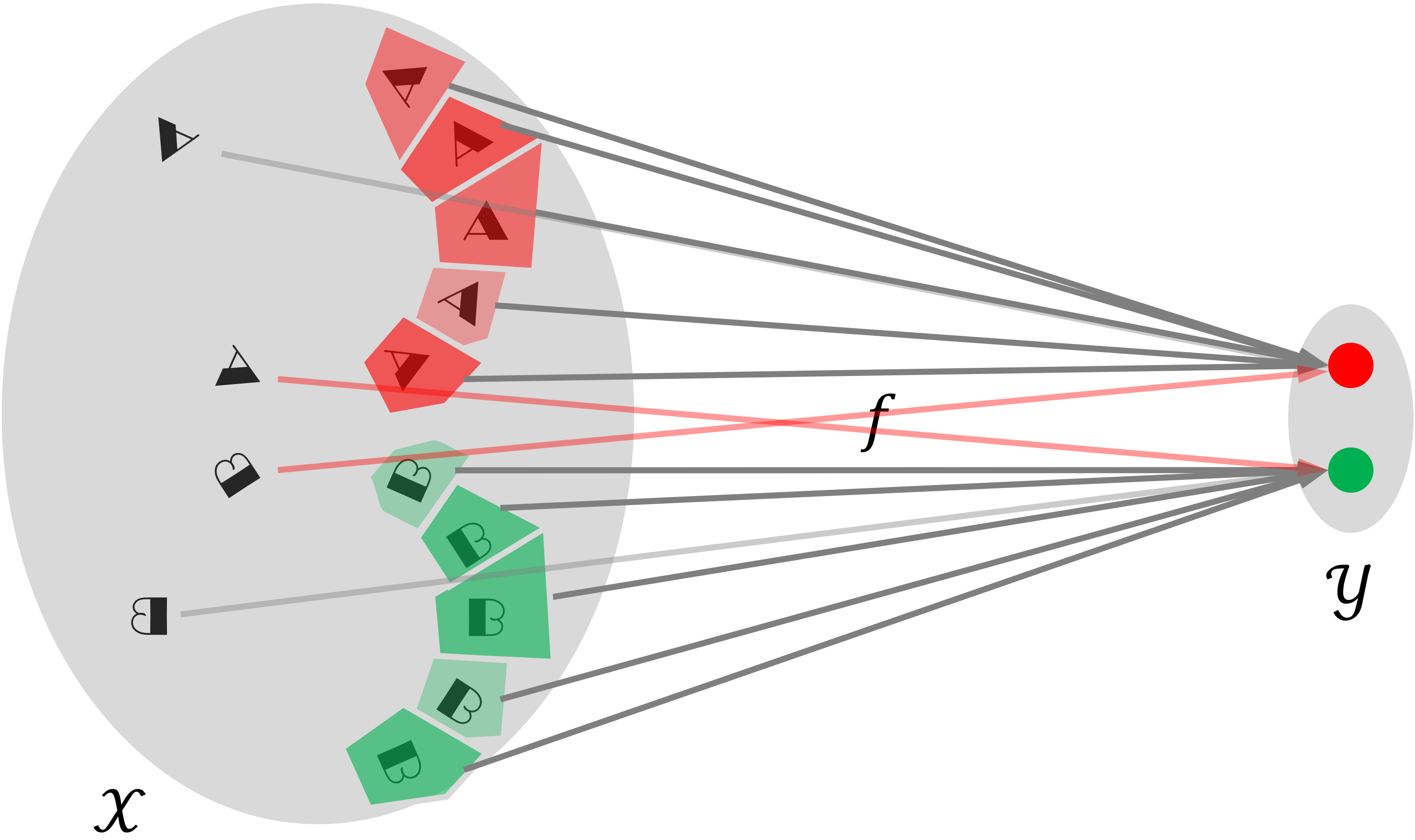


$$f(\rho(g)x) = f(x) \quad \forall g \in G$$

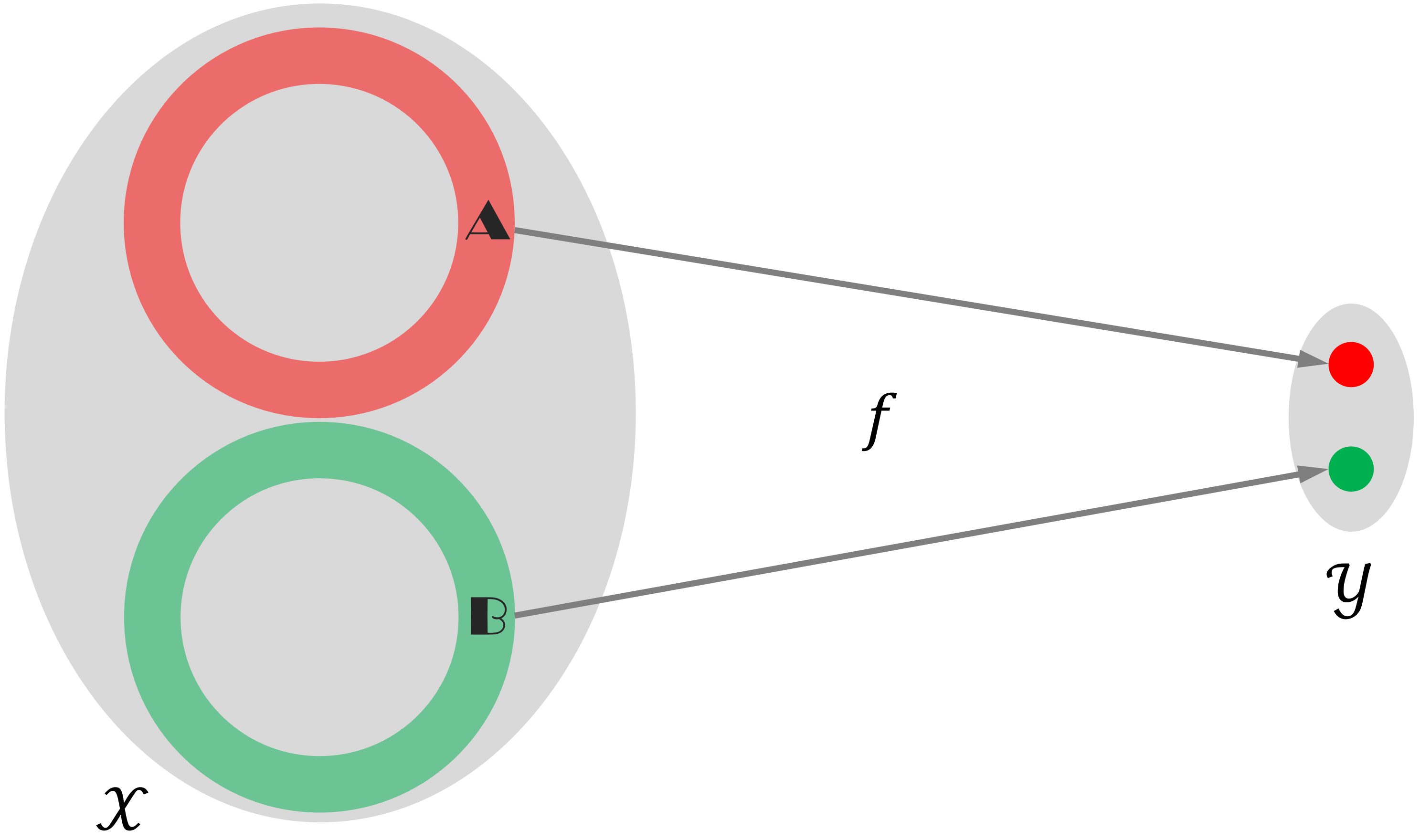
Symmetries of the Label Function



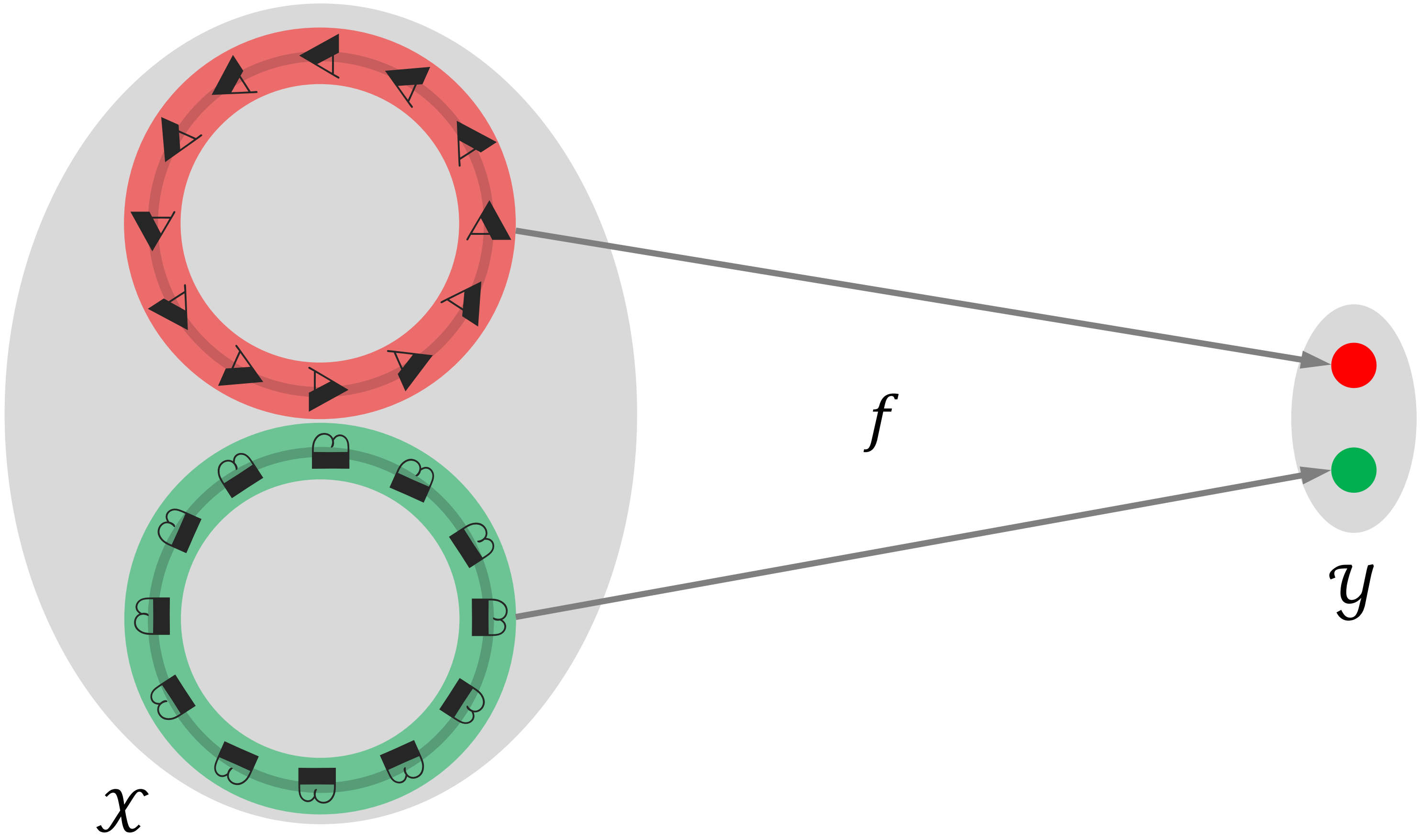
Symmetries of the Label Function



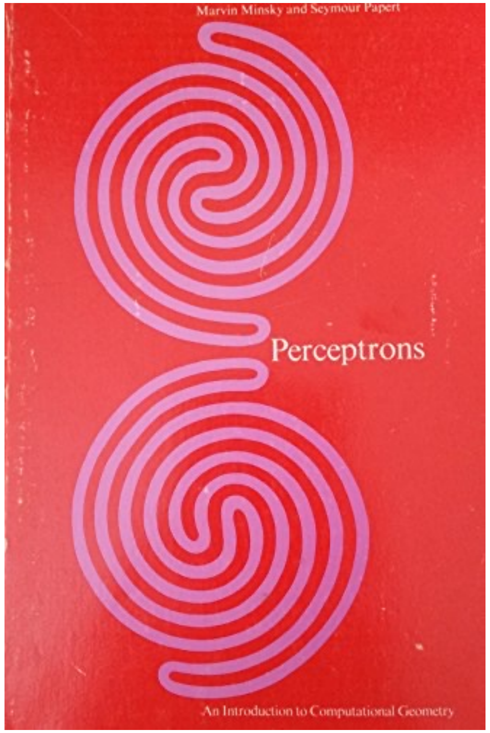
Symmetries of the Label Function

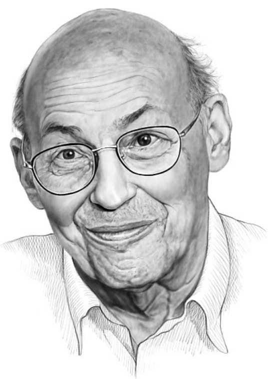


Symmetries of the Label Function

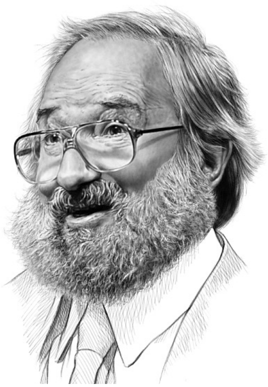


First "geometric" machine learning


Marvin Minsky and Seymour Papert
Perceptrons
An Introduction to Computational Geometry



M. Minsky



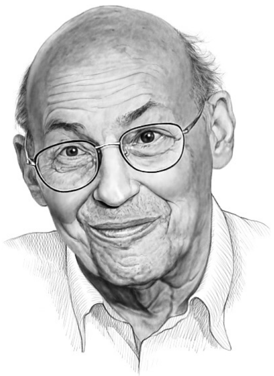
S. Papert

1969

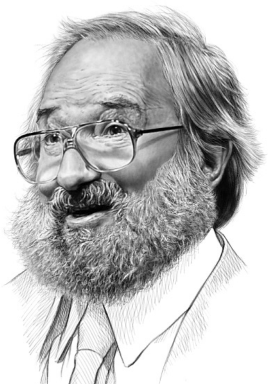
Minsky, Papert 1969

First "geometric" machine learning

Group Invariance Theorem: ″if a neural network is invariant to a group, then its output can be expressed as functions of the orbits of the group″



M. Minsky

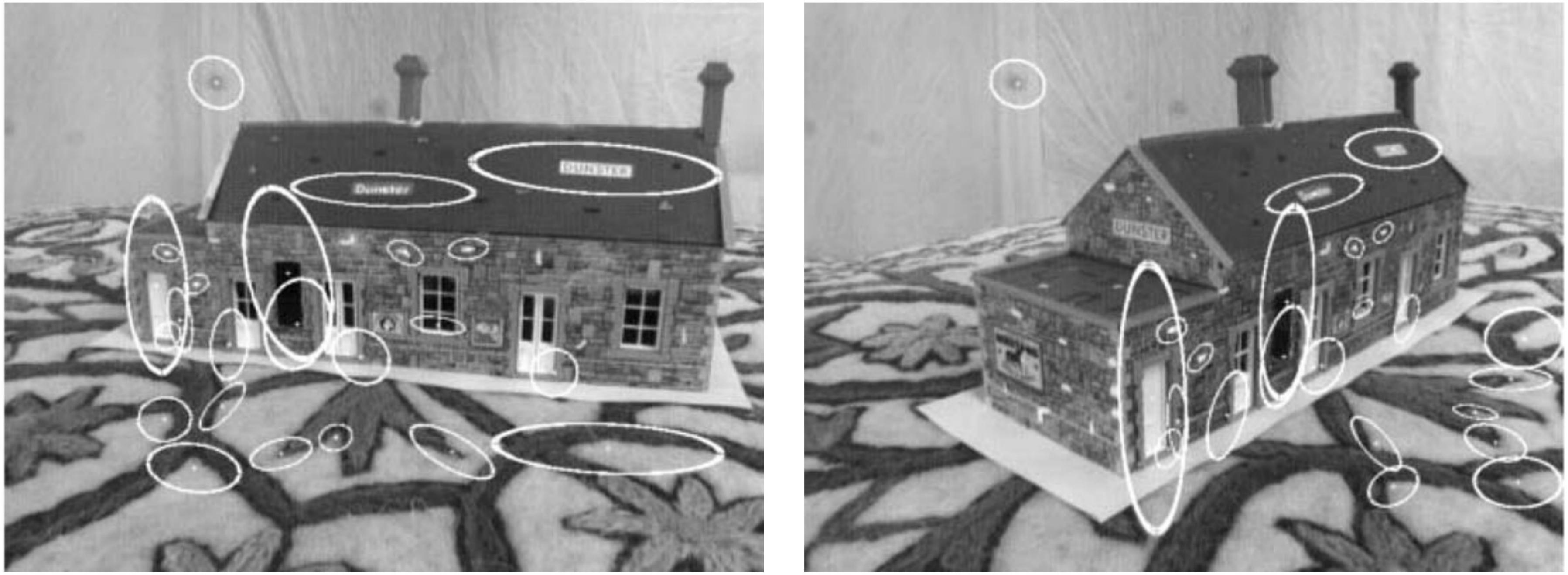


S. Papert

1969

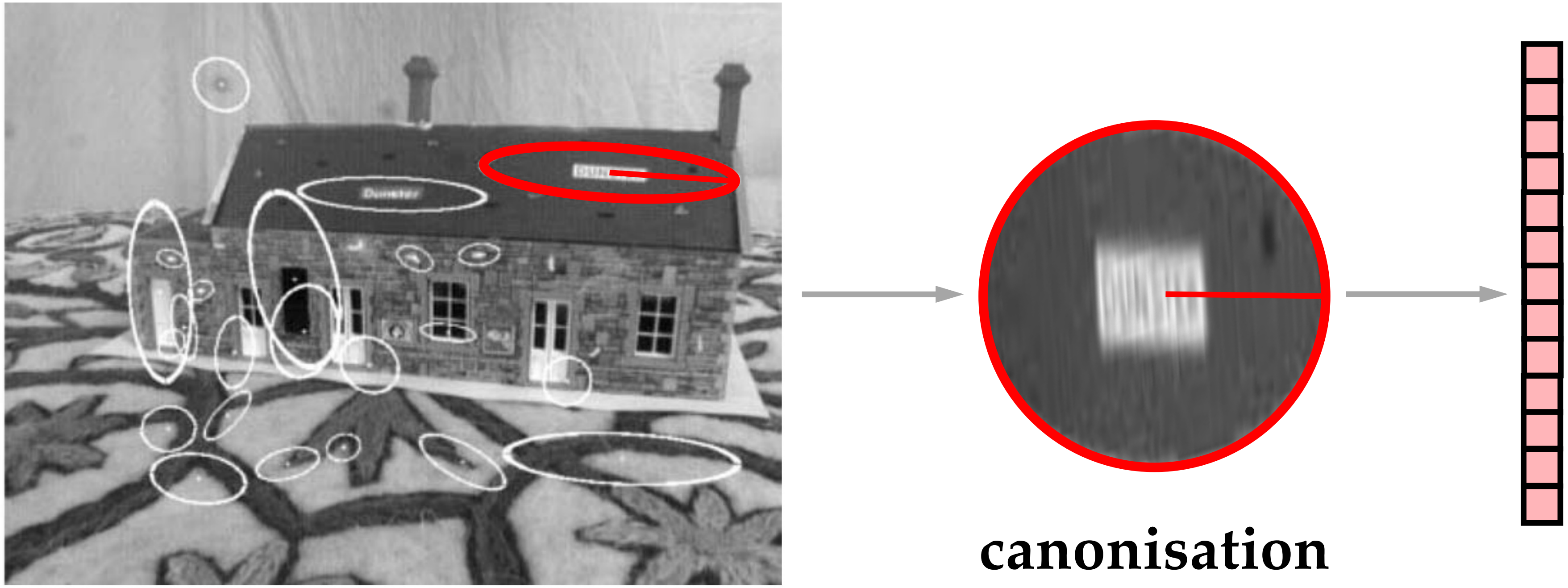
Minsky, Papert 1969

Canonisation



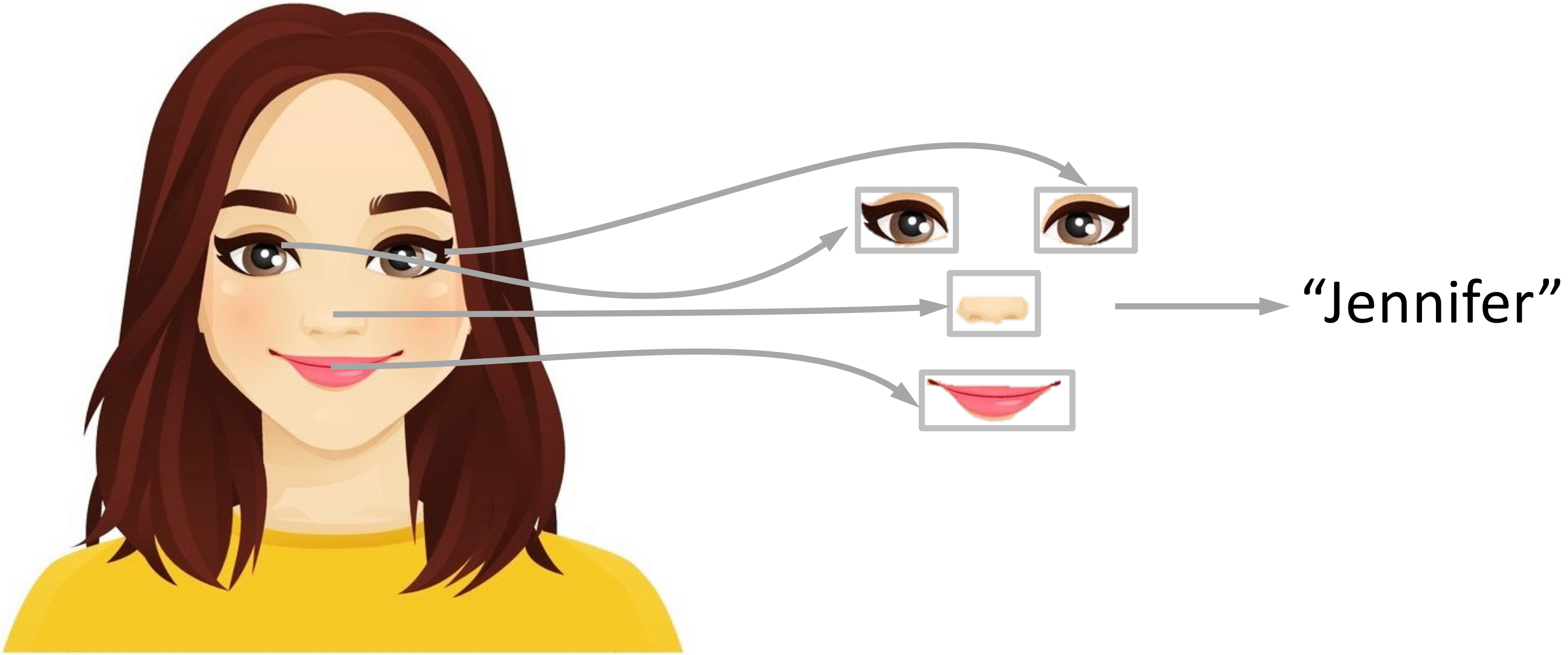
Mikolajczyk, Schmid 2004

Canonisation

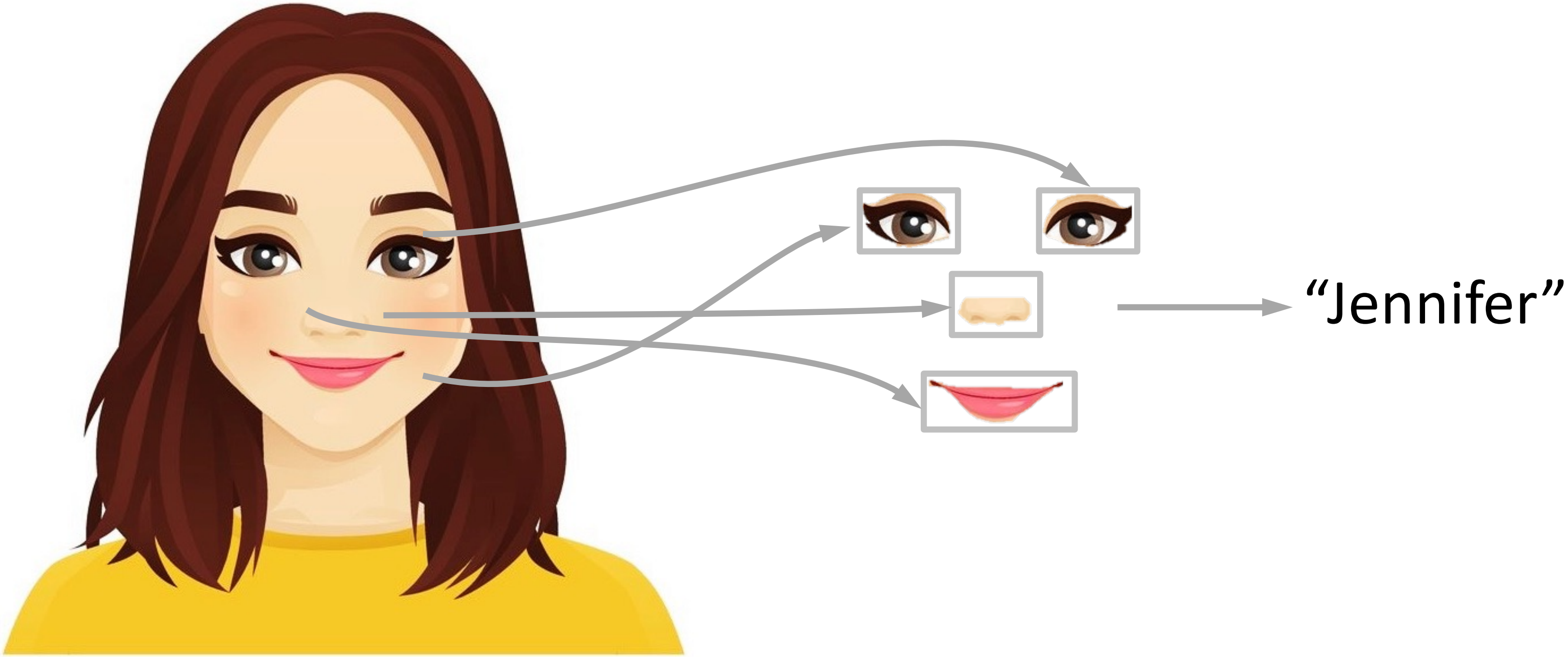


canonisation

Canonisation

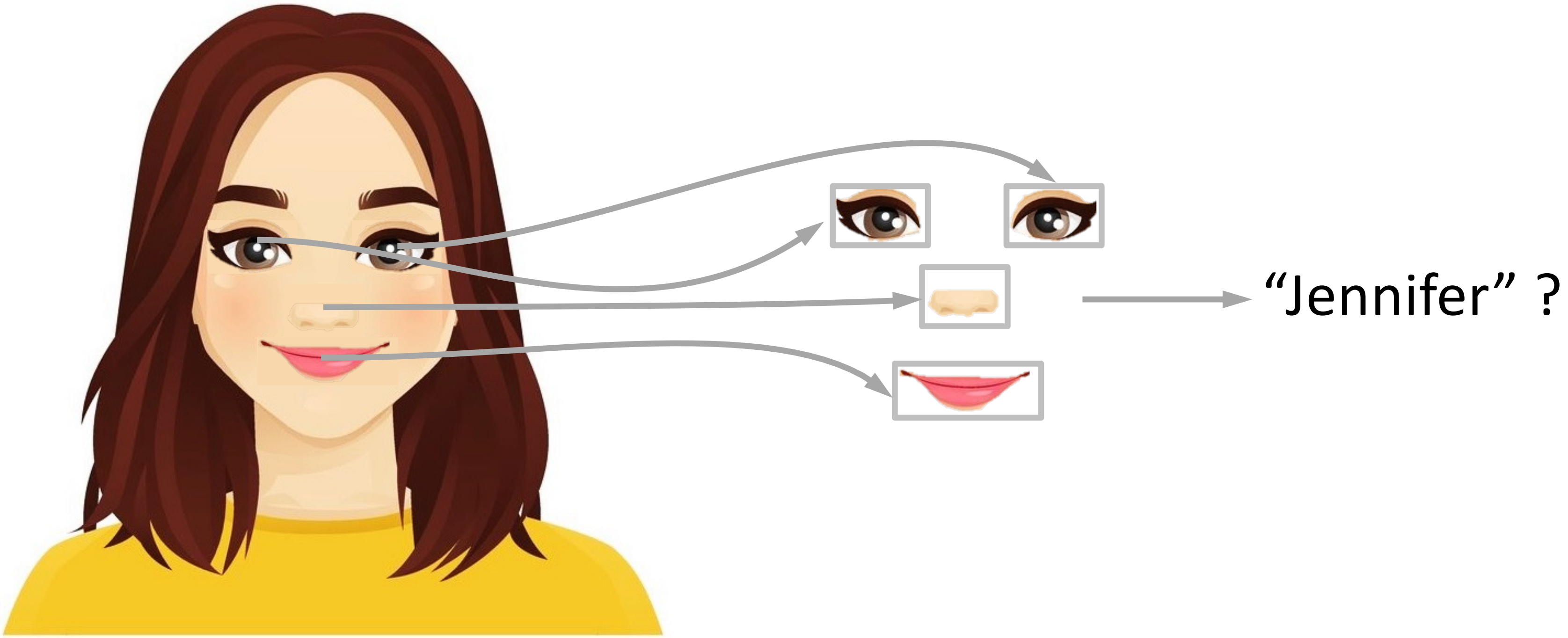


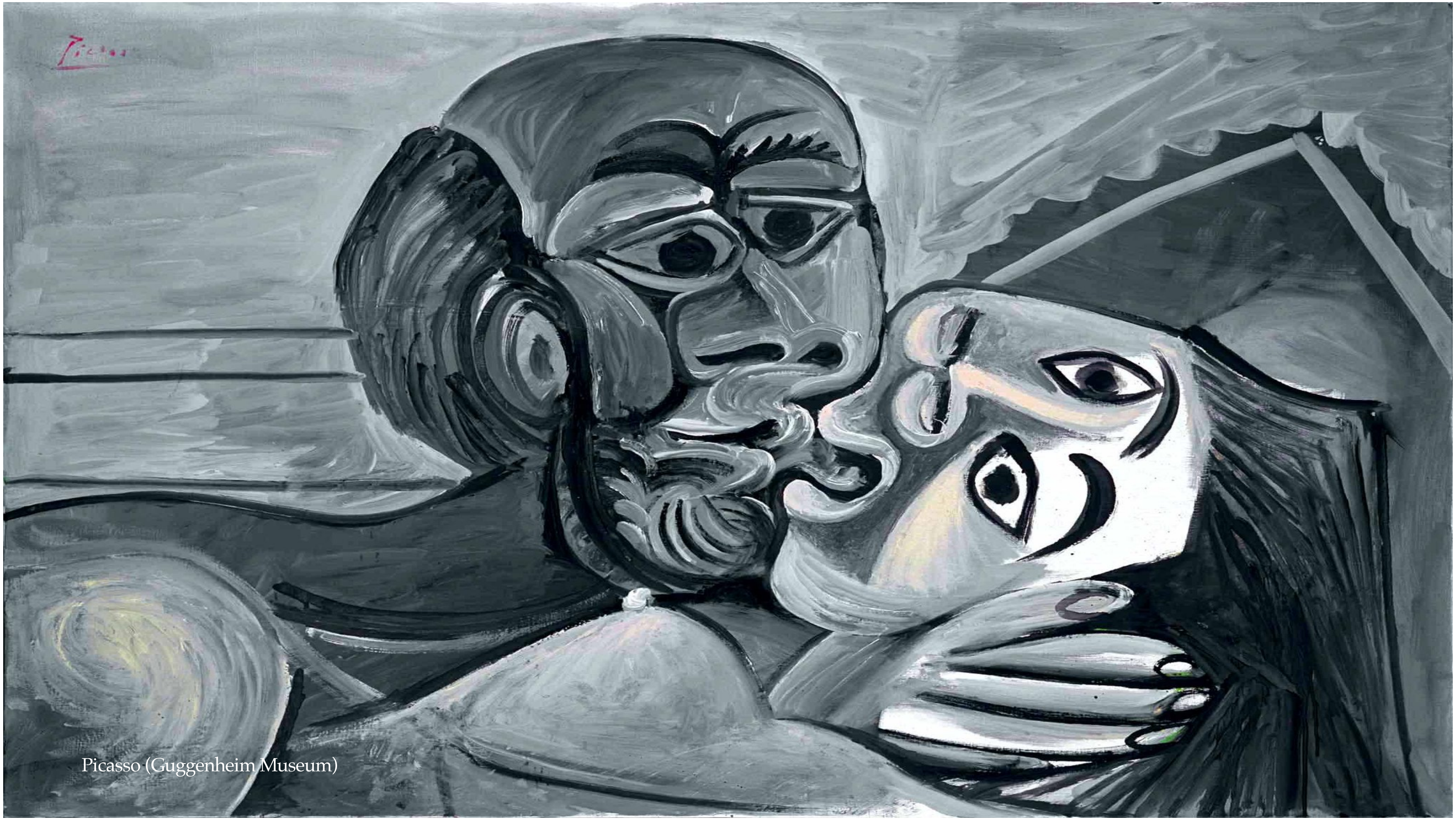
Canonisation



“Jennifer”

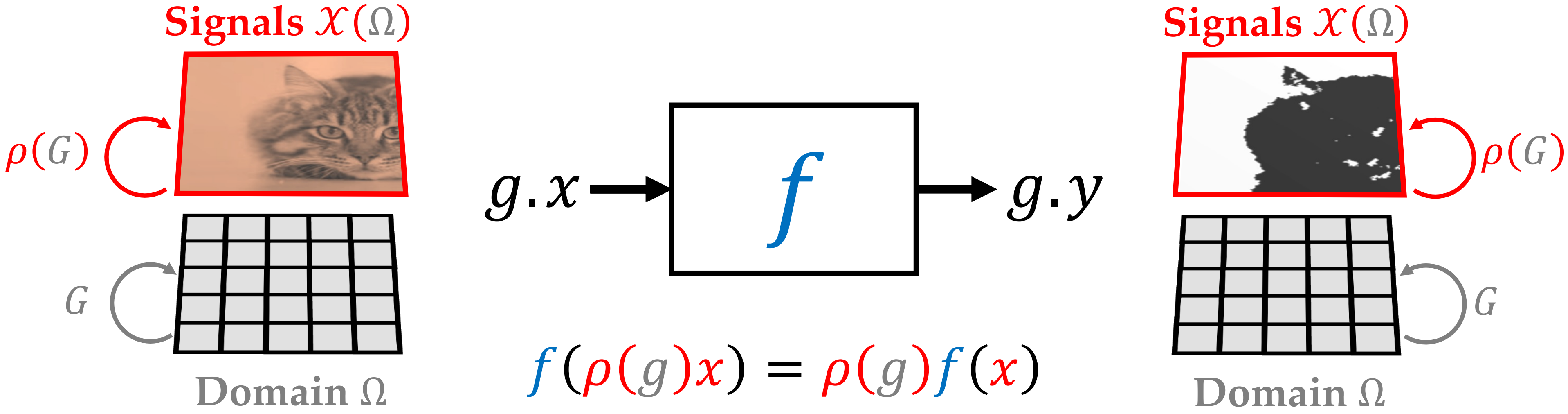
Canonisation





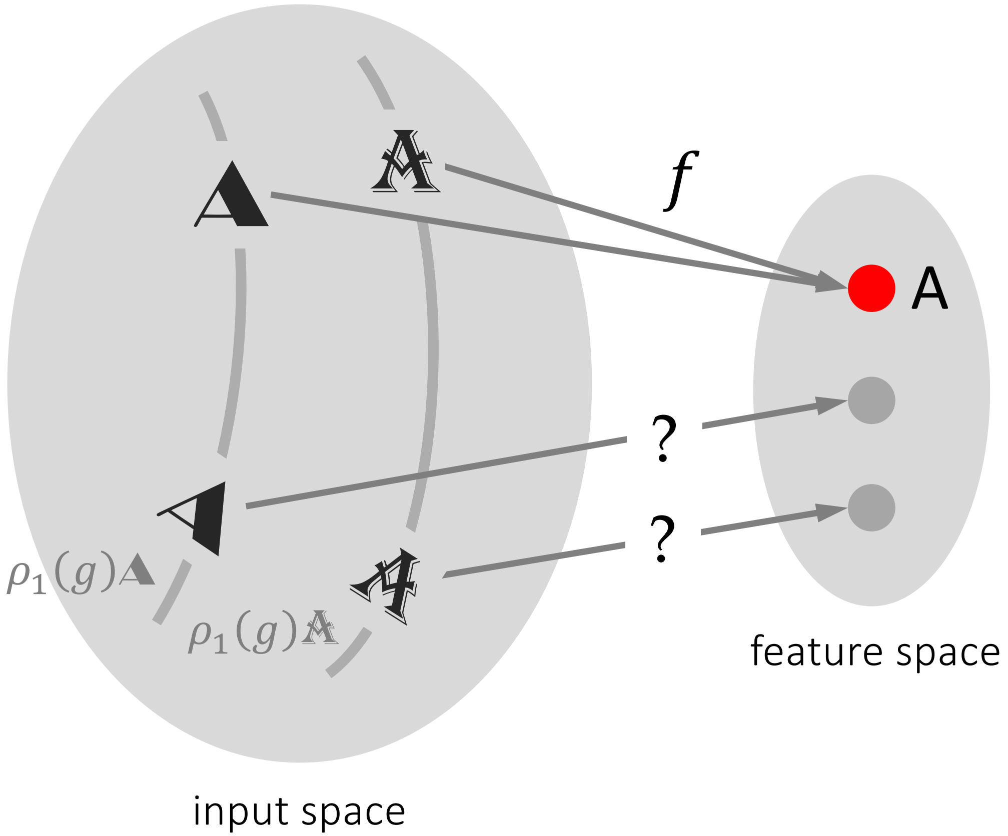
Picasso (Guggenheim Museum)

Geometric priors: Equivariance



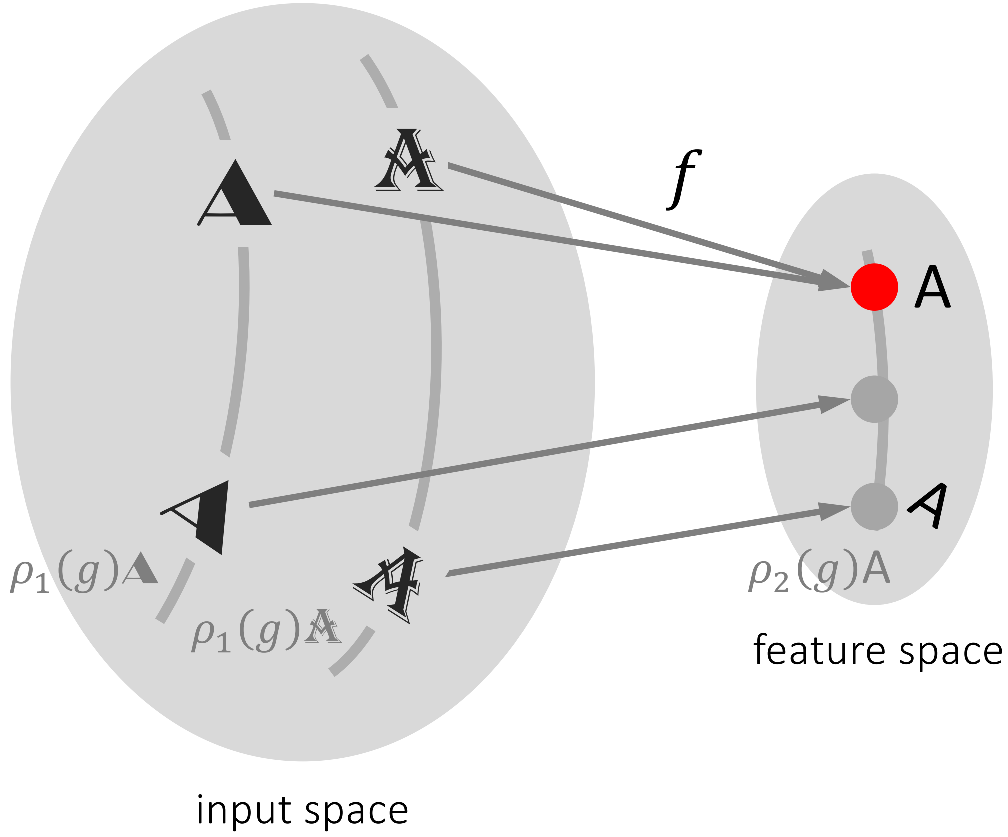
$$f(\rho(g)x) = \rho(g)f(x)$$
$$\forall g \in G$$

Equivariance = Symmetry-consistent generalisation



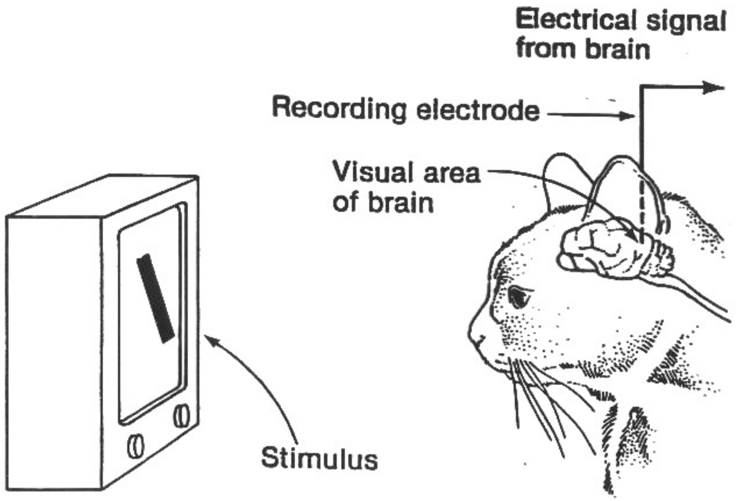
$$\begin{array}{rl} f(\rho_1(g)\mathbf{A}) = \rho_2(g) & \boxed{f(\mathbf{A})} \\ & \quad \| \\ f(\rho_1(g)\mathcal{A}) = \rho_2(g) & \boxed{f(\mathcal{A})} \end{array}$$

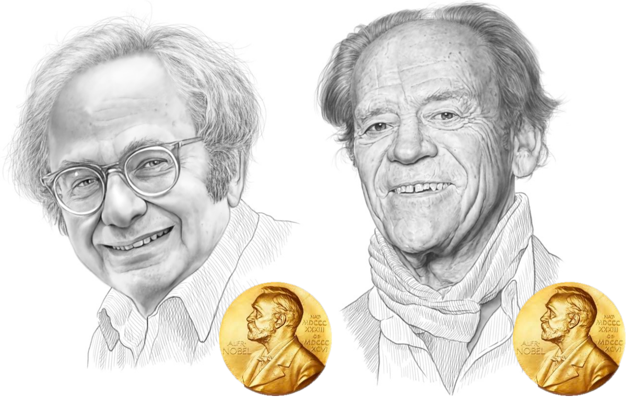
Equivariance = Symmetry-consistent generalisation



$$
\begin{aligned}
f(\rho_1(g)\mathbf{A}) &= \rho_2(g)\, f(\mathbf{A}) \\
& \qquad\qquad \parallel \\
f(\rho_1(g)\mathfrak{A}) &= \rho_2(g)\, f(\mathfrak{A})
\end{aligned}
$$

Early Geometric Architectures



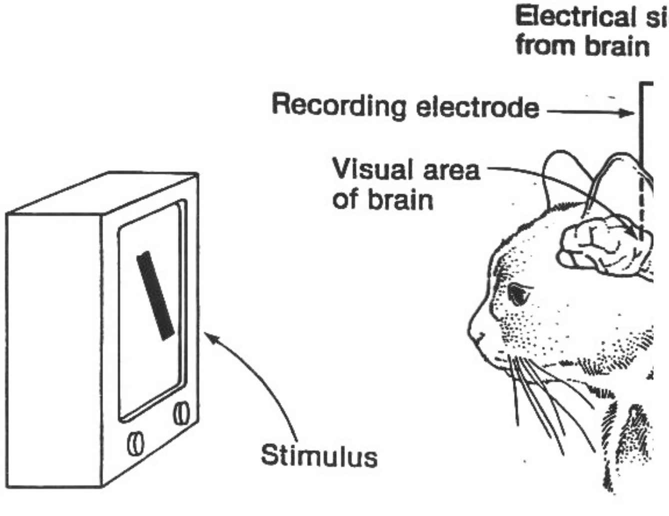


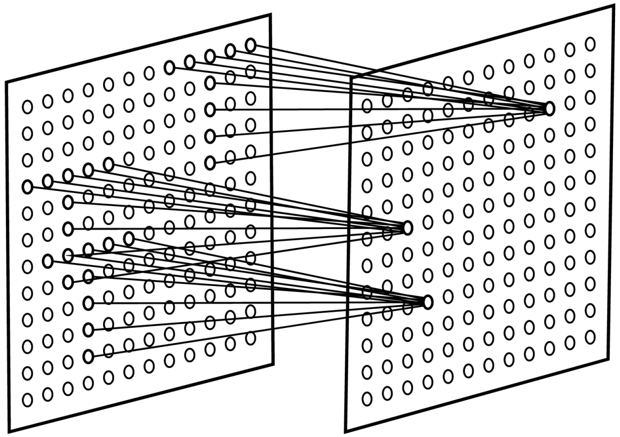
D. Hubel **T. Wiesel**

1959

Hubel, Wiesel 1959, 1962; Portraits: Ihor Gorskyi

Early Geometric Architectures





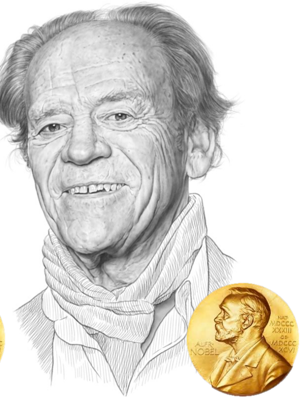
K. Fukushima

1959 1980

Hubel, Wiesel 1959, 1962; Fukushima 1980; Portraits: Ihor Gorskyi

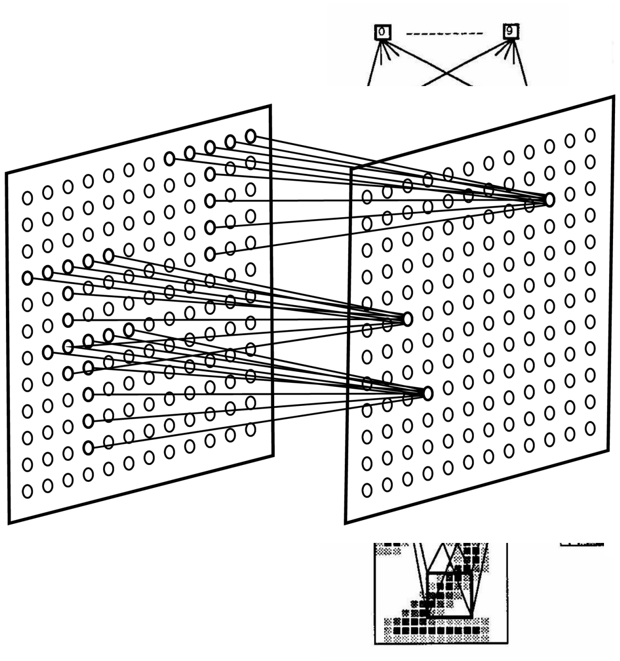
Early Geometric Architectures

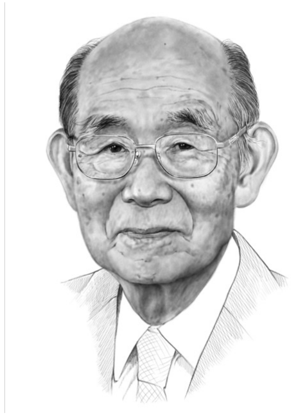
D. Hubel



T. Wiesel

1959



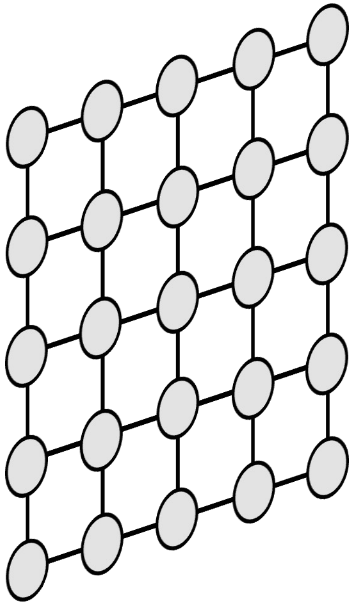


K. Fukushima

1989

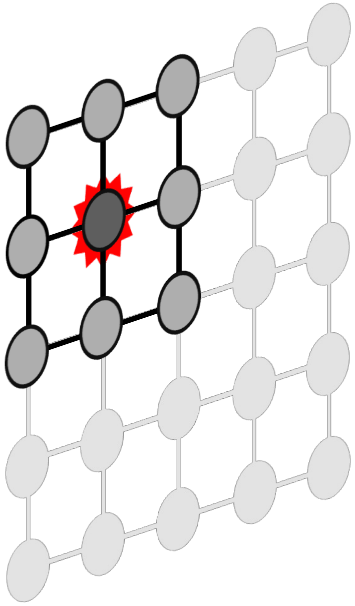
Hubel, Wiesel 1959, 1962; Fukushima 1980; LeCun et al. 1989; Portraits: Ihor Gorskyi

Convolutional Neural Networks



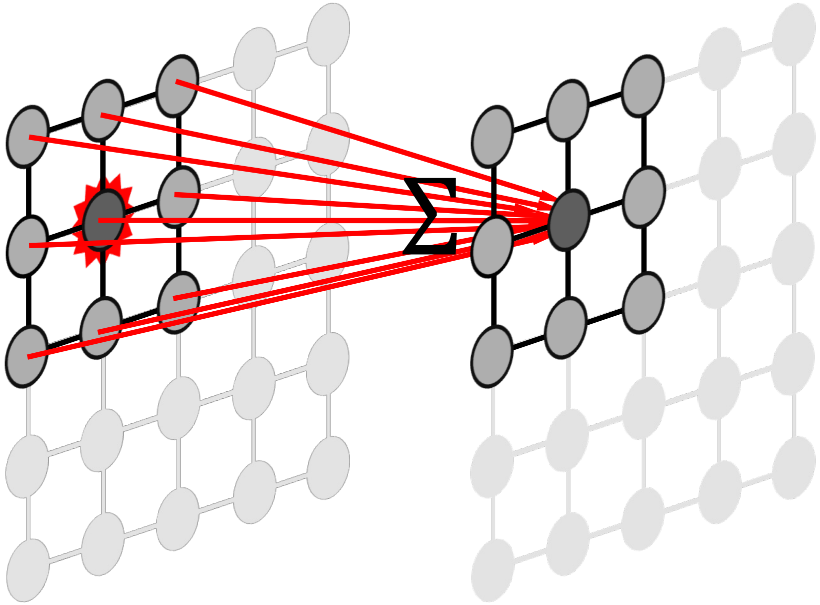
LeCun et al. 1989

Convolutional Neural Networks



LeCun et al. 1989

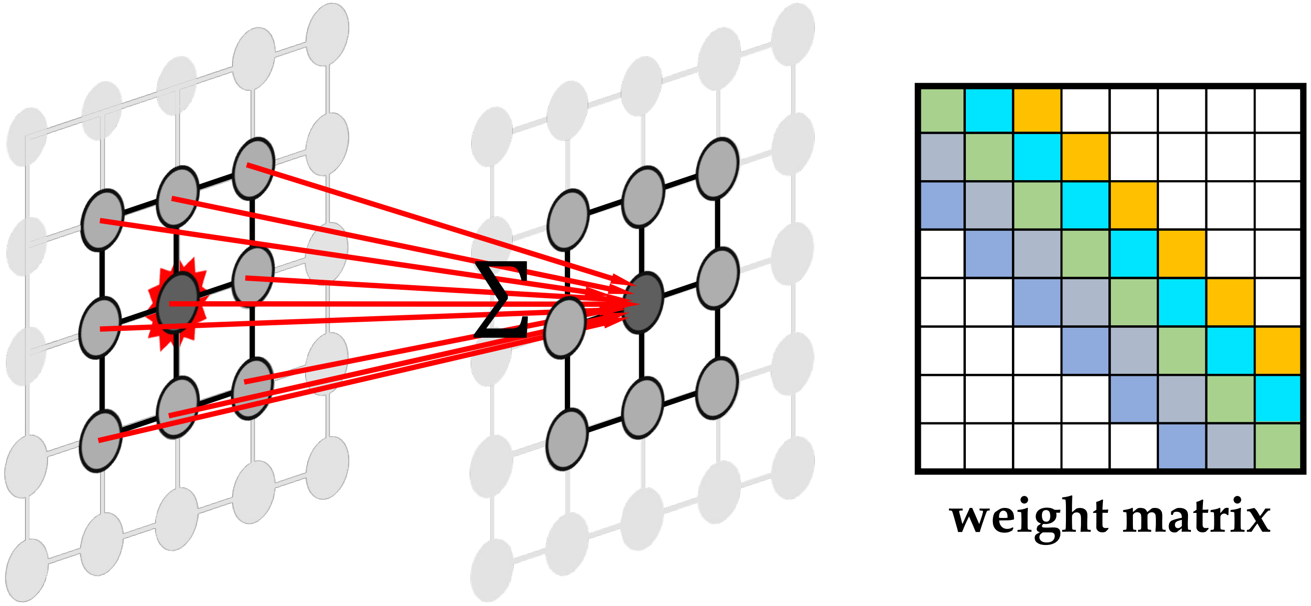
Convolutional Neural Networks



Locality + Shared parameters

LeCun et al. 1989

Convolutional Neural Networks

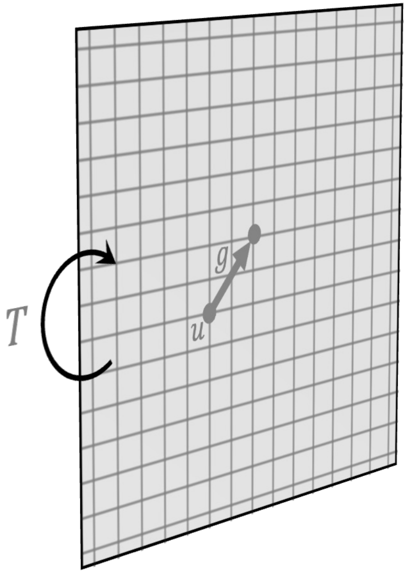


Locality + Shared parameters

LeCun et al. 1989

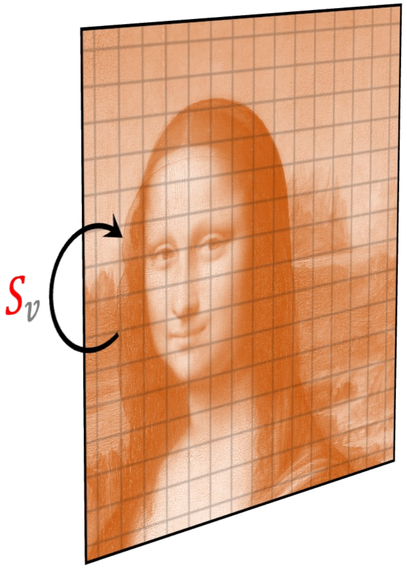
Convolutional Neural Networks

Plane $\mathbb{R}^2$



Translation group $T(2)$

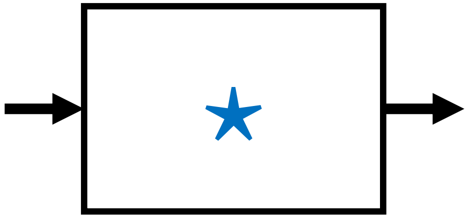
Images $\mathcal{X}(\mathbb{R}^2)$



Shift operator S

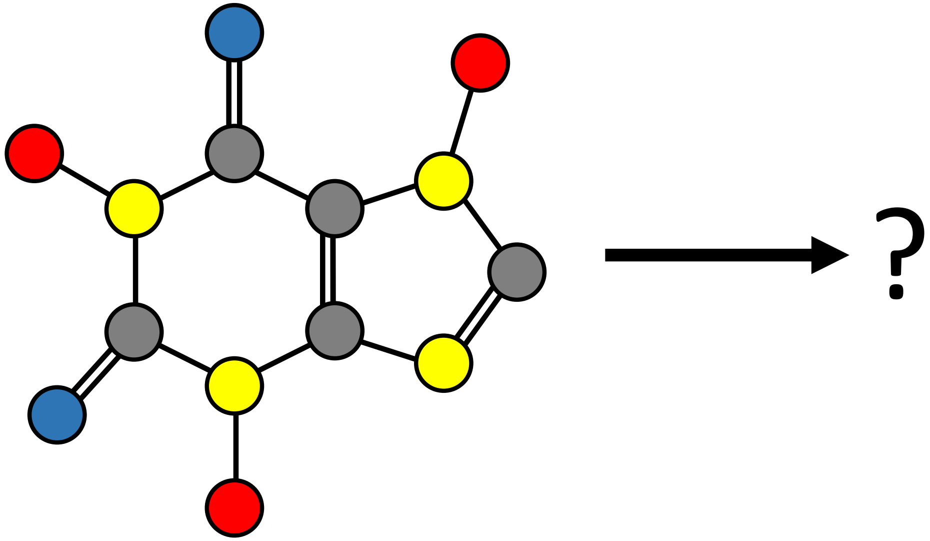
$$S_v x(u) = x(u - v)$$

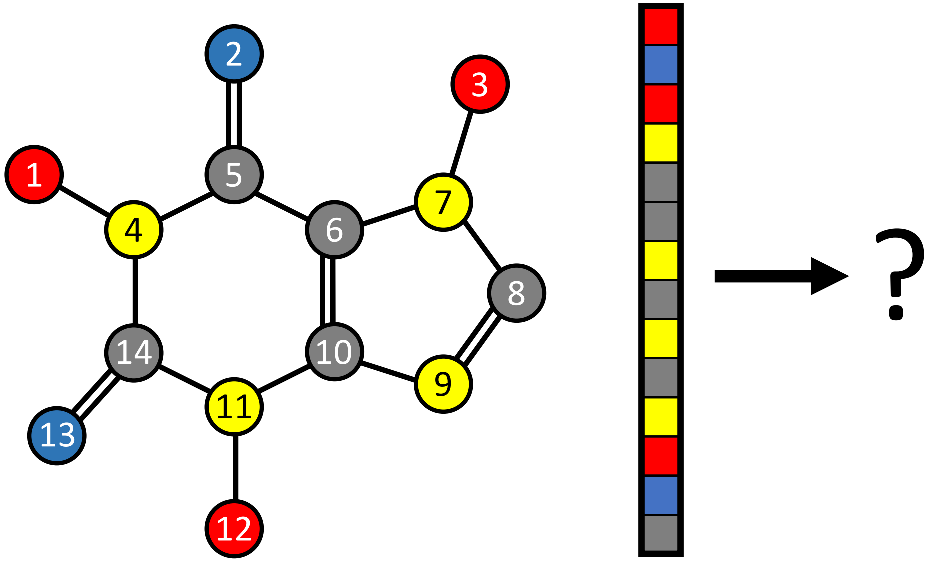
Functions $\mathcal{F}(\mathcal{X}(\mathbb{R}^2))$

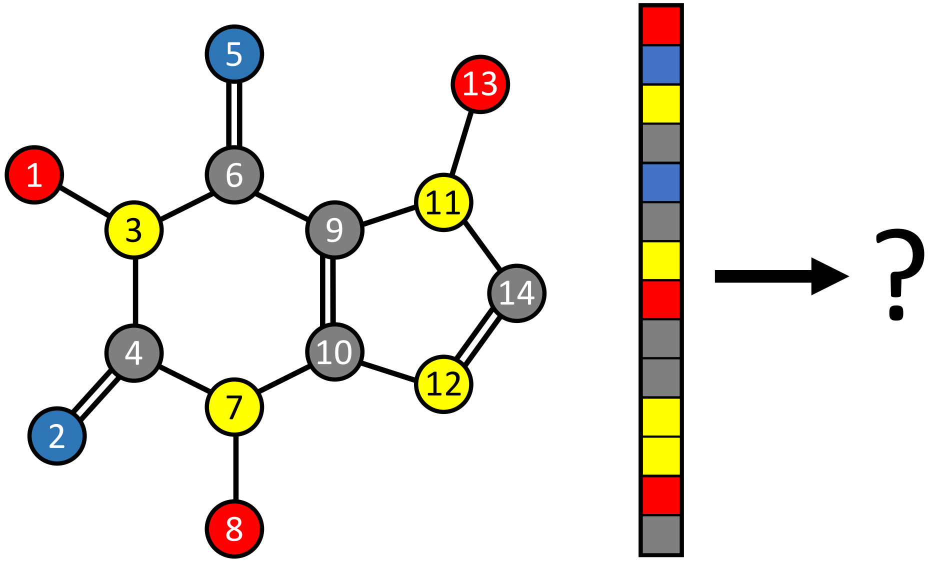


Convolution

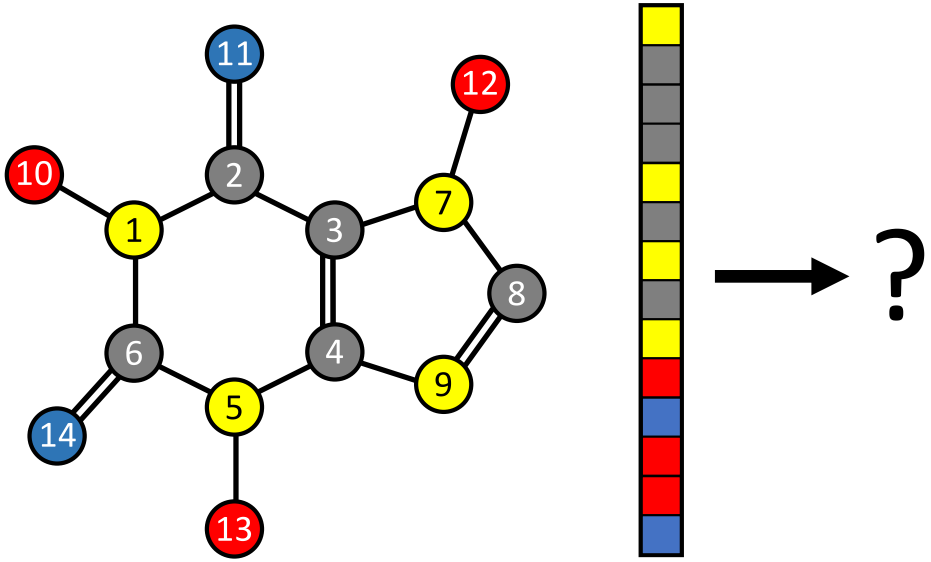
$$(Sx \star y) = S(x \star y)$$




1
2
3
4
5
6
7
8
9
10
11
12
13
14
?

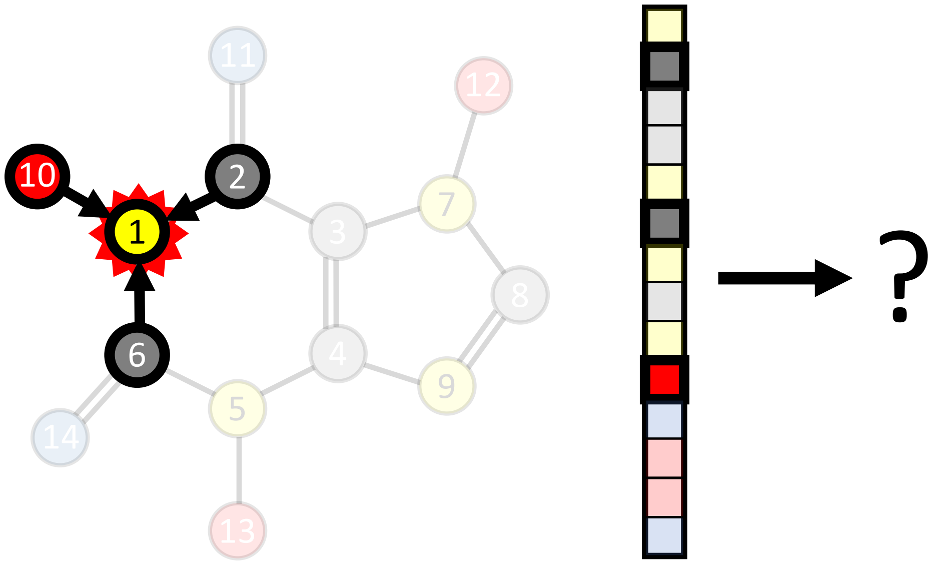

1
2
3
4
5
6
7
8
9
10
11
12
13
14
?

Permutation Invariance



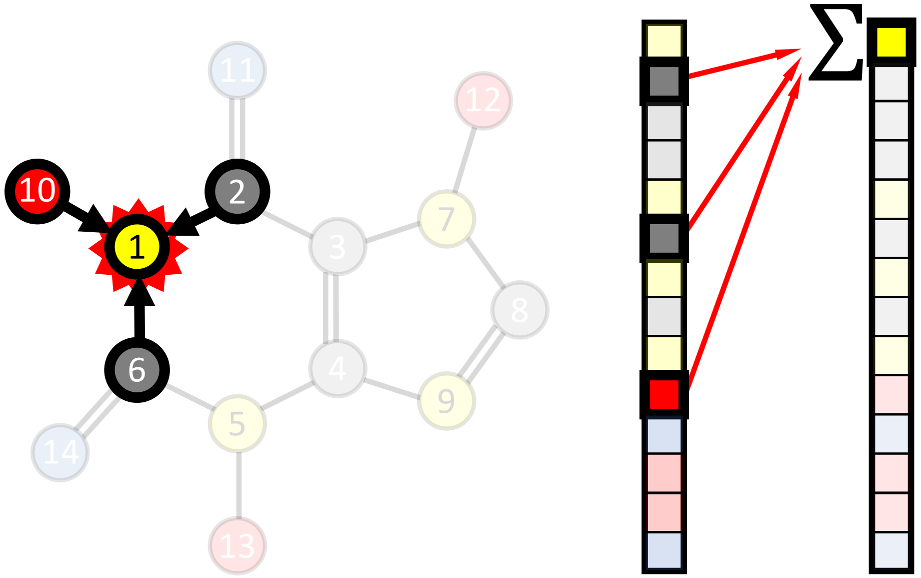
"properties of a molecule do not change if we reorder the atoms"

Graph Neural Networks



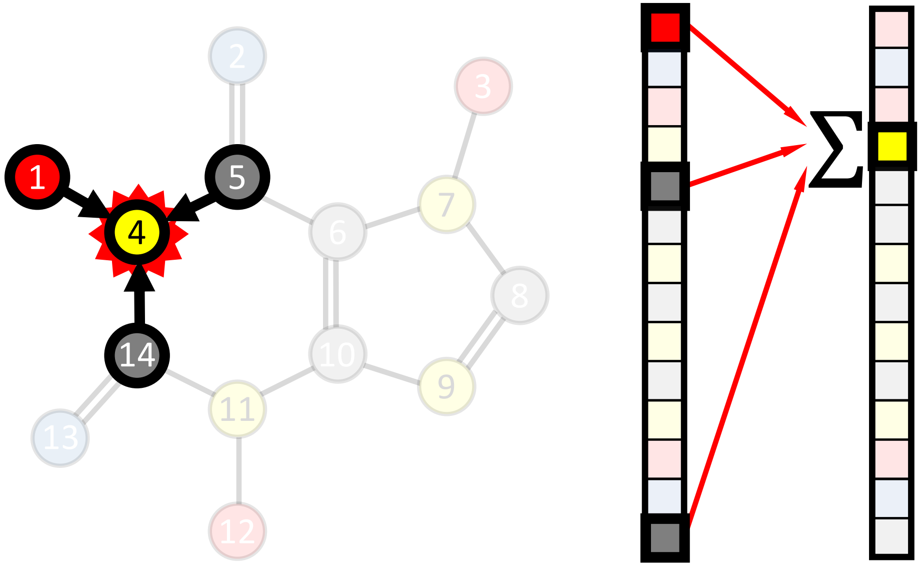
Locality + Shared parameters

Graph Neural Networks



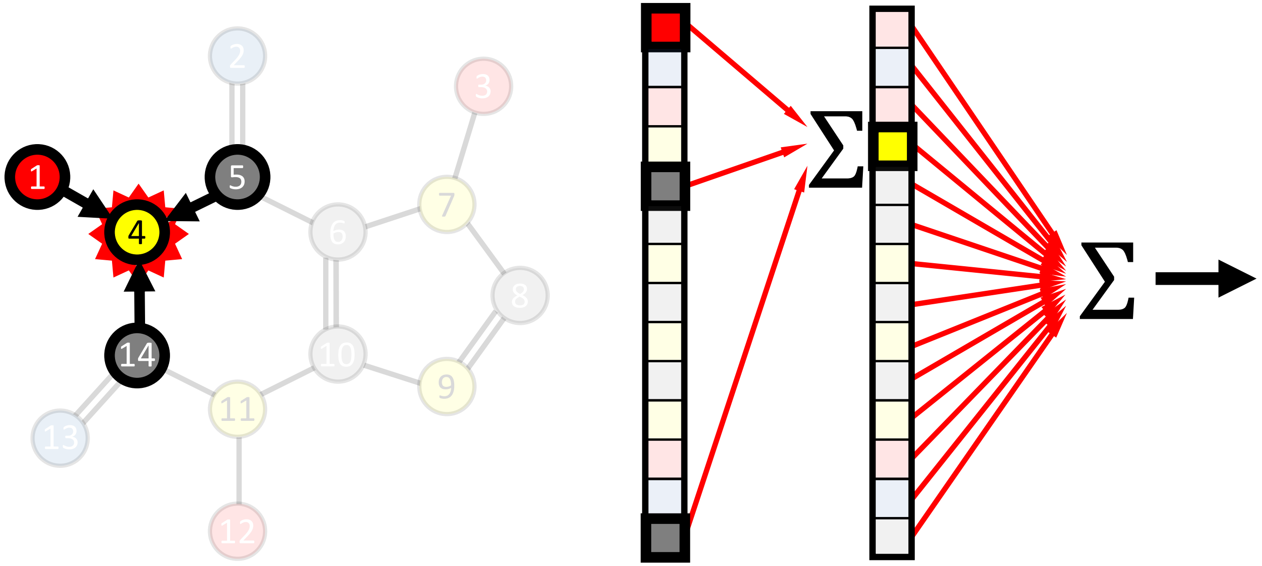
Permutation-equivariant layer

Graph Neural Networks



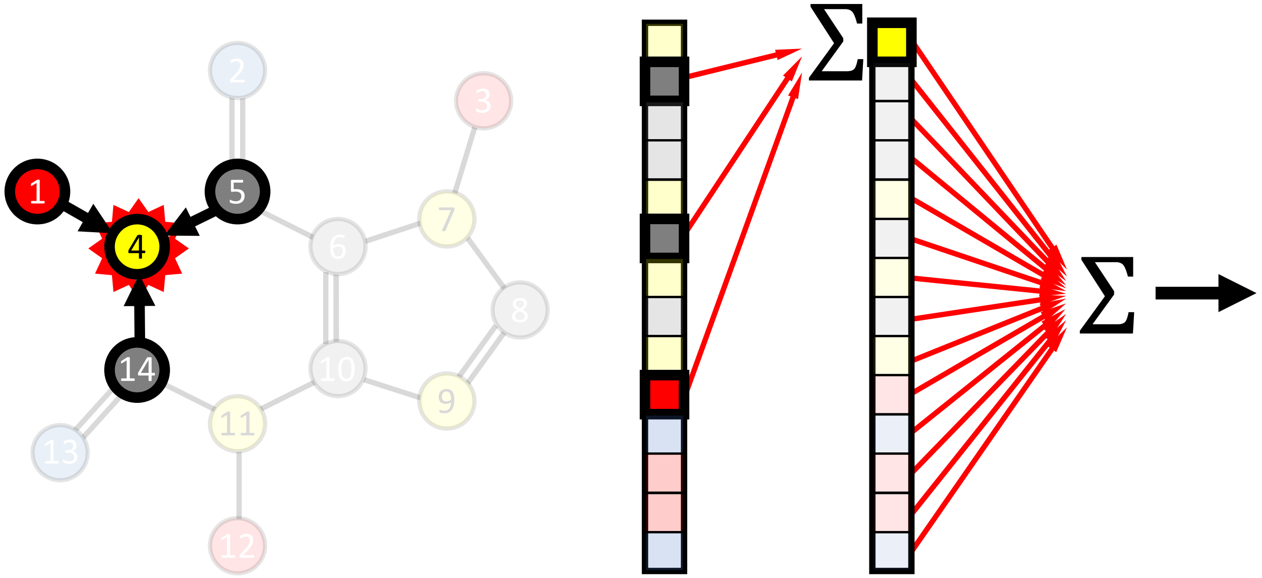
Permutation-equivariant layer

Graph Neural Networks



Permutation-invariant readout

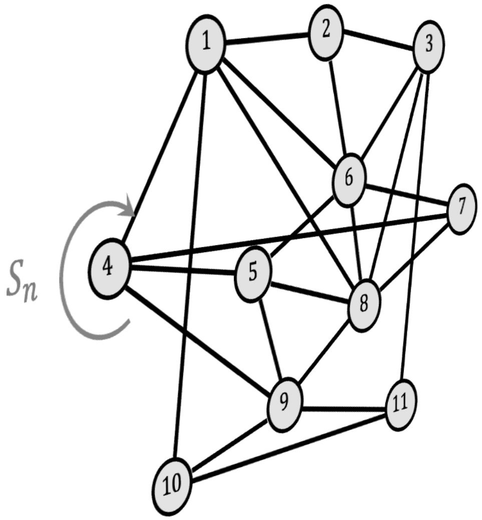
Graph Neural Networks



Permutation-invariant readout

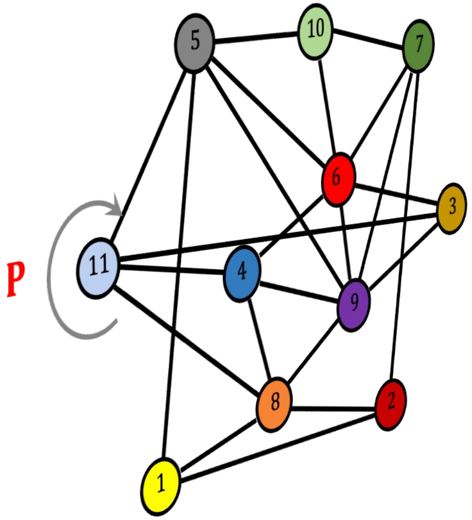
Graph Neural Networks

Graph $G = (V, E)$



Permutation group S_n

Node features $\mathcal{X}(G)$



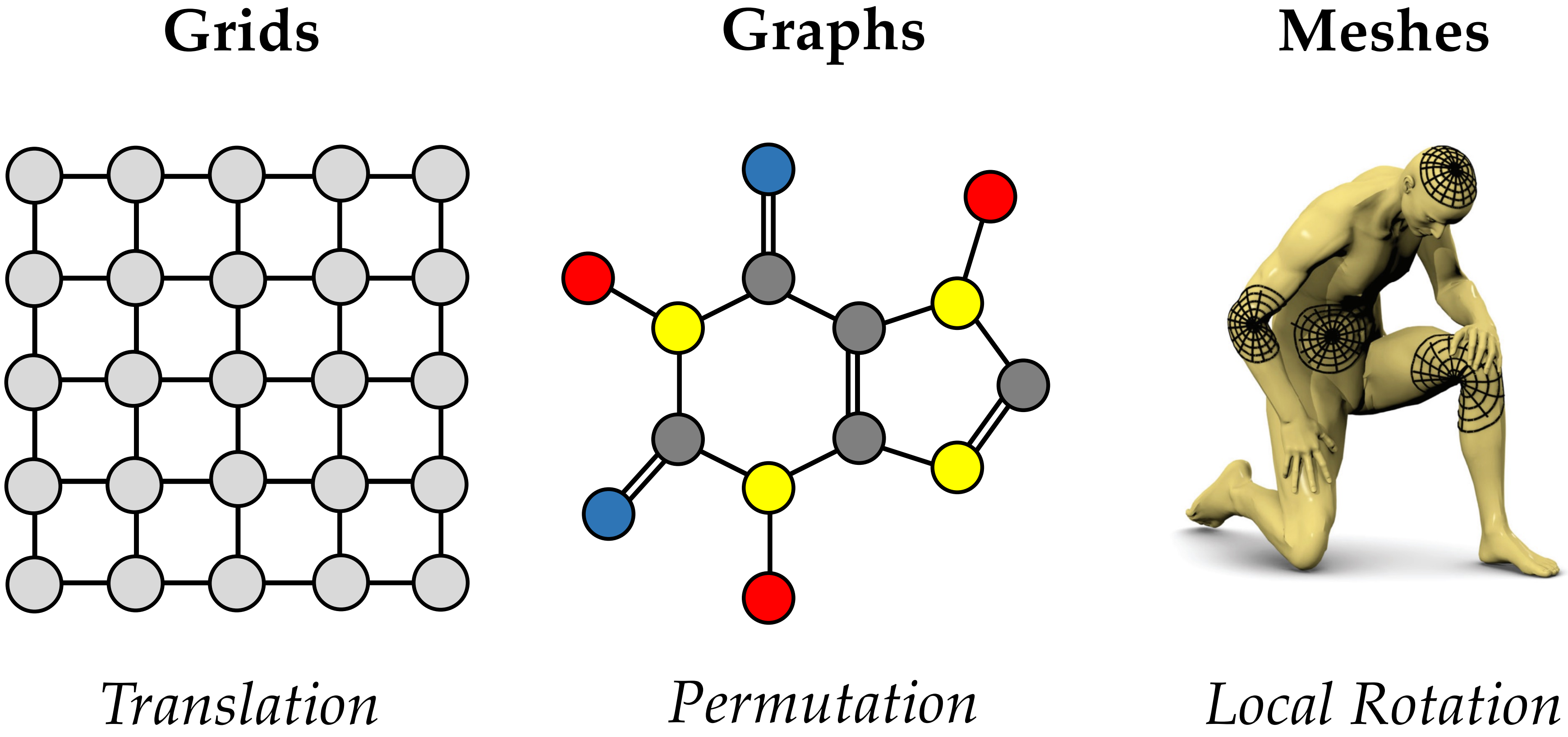
Permutation matrix P

$$\mathbf{PX} = \left(x_{\pi^{-1}(i),j}\right)$$

Functions $\mathcal{F}(\mathcal{X}(G))$

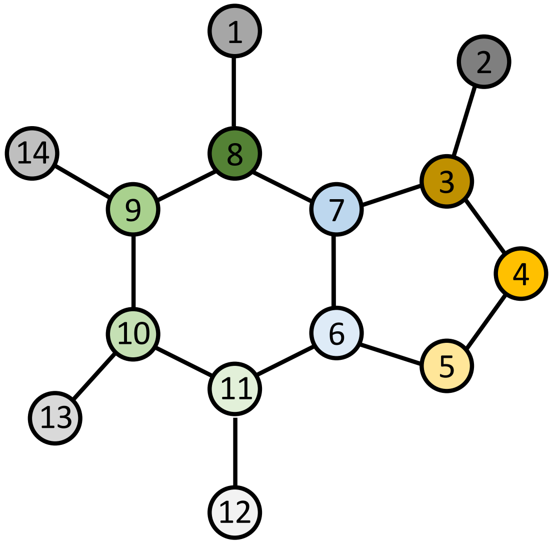
Message passing

$$\mathbf{F}\left(\mathbf{PX}, \mathbf{PAP}^\top\right) = \mathbf{PF}(\mathbf{X}, \mathbf{A})$$


Grids
Graphs
Meshes
Translation
Permutation
Local Rotation

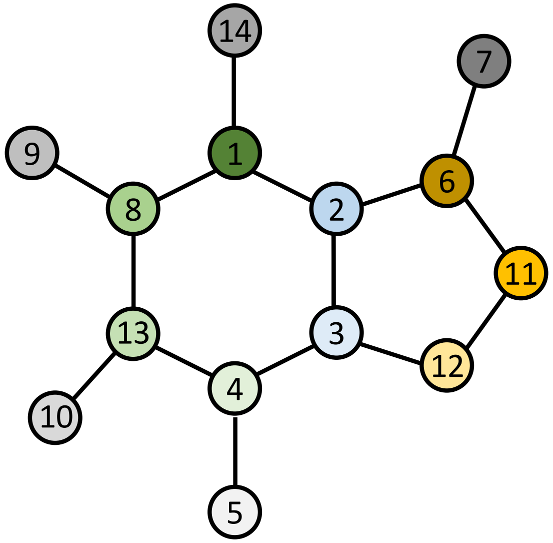
$$+\ T(2) = \mathrm{CNN}$$

B et al 2017; 2021



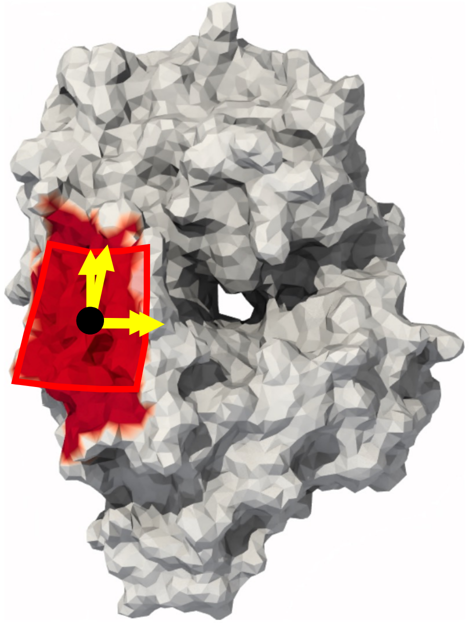
$$+ \quad S_n \quad = \text{GNN}$$

B et al 2017; 2021



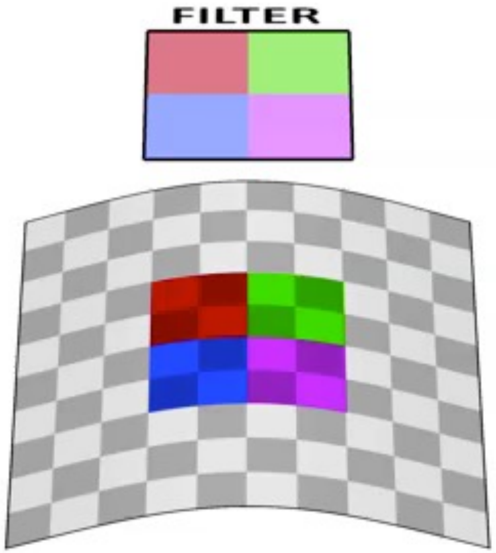
$$+ \quad S_n = \mathrm{GNN}$$

B et al 2017; 2021

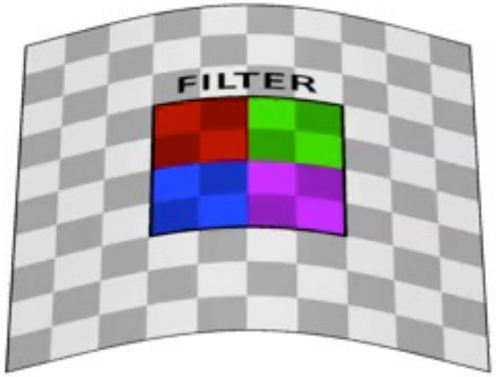


+ SO(2) = MeshCNN

B et al 2017; 2021

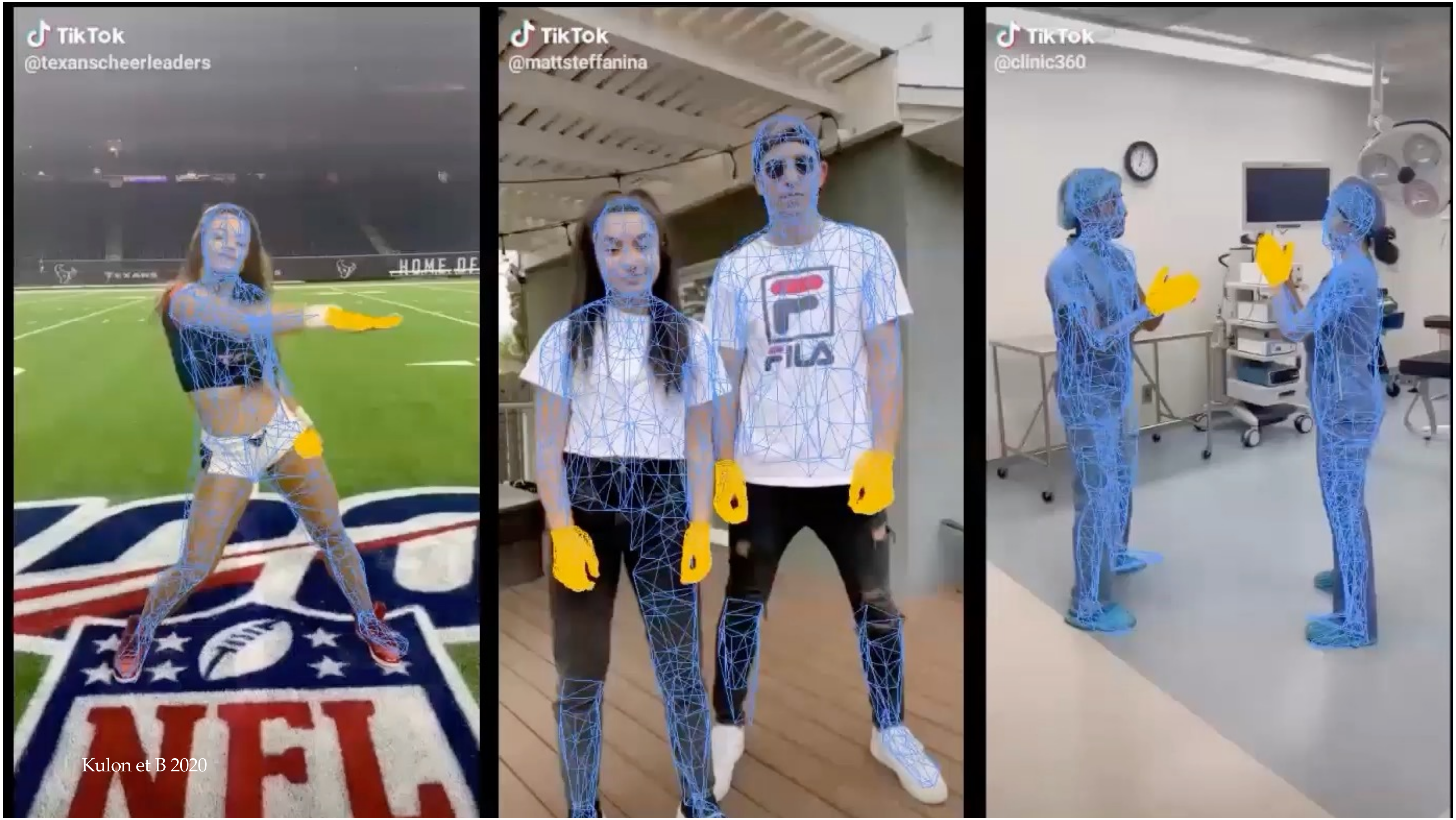


Euclidean (extrinsic) convolution



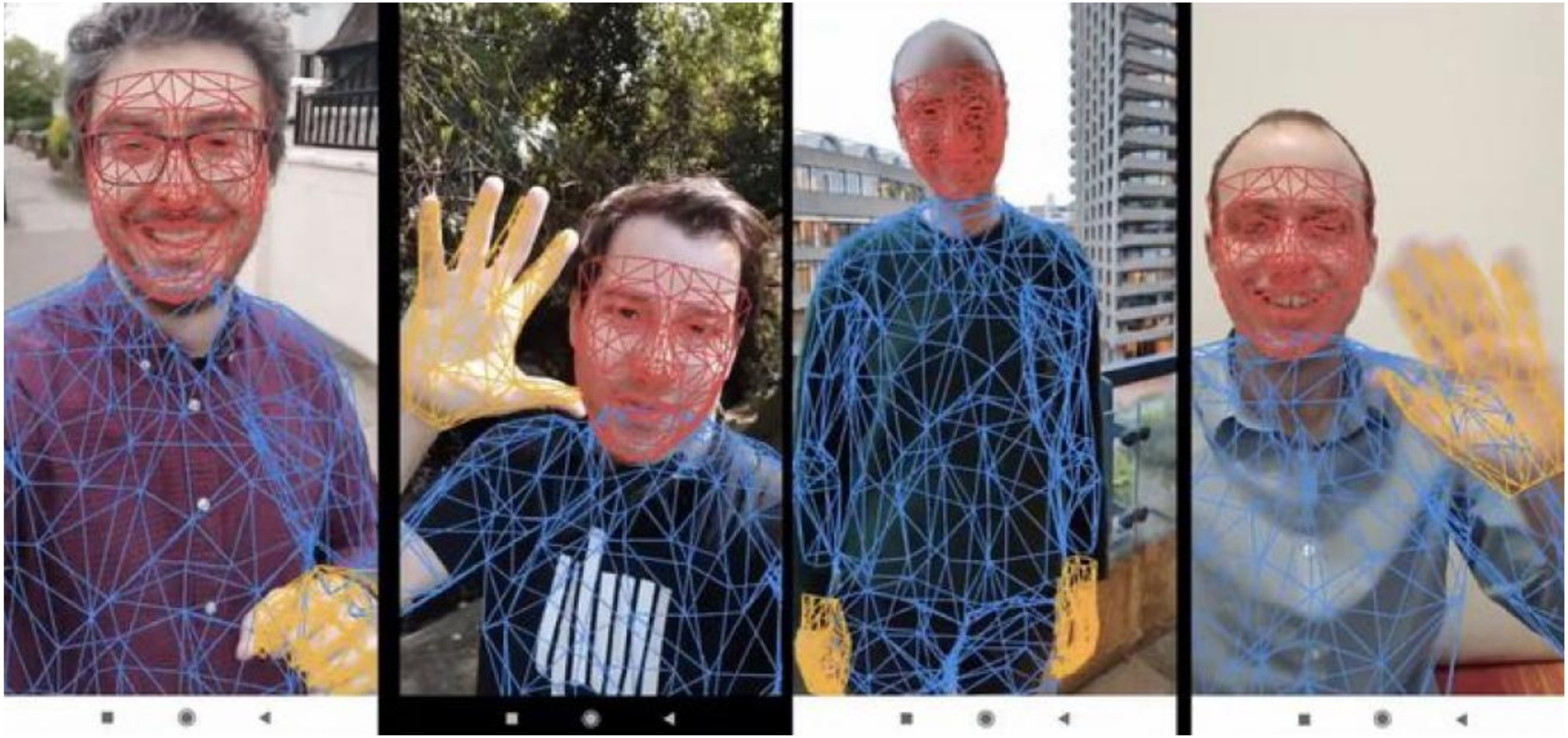
Geometric (intrinsic) convolution

Masci et B 2015; Monti et B 2017

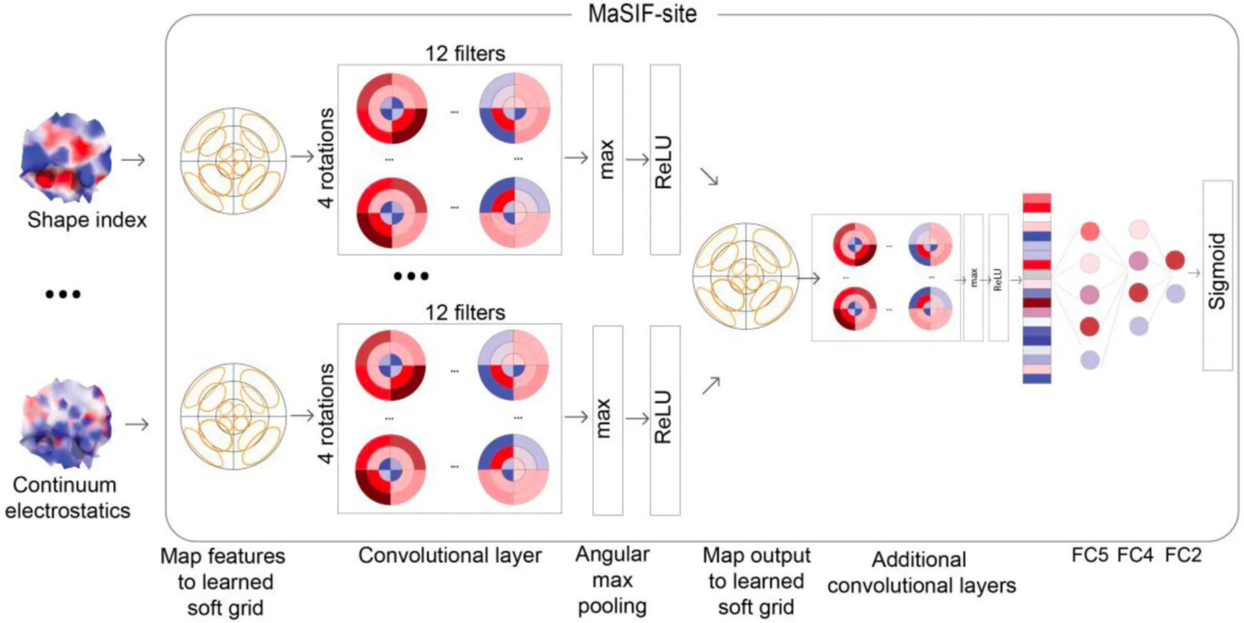


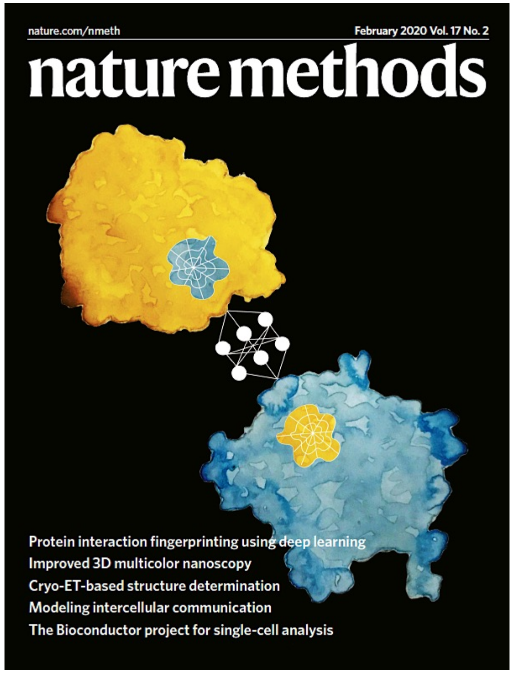
Kulon et B 2020

Snap Acquires Ariel AI To Enhance AR Features



MaSIF: Geometric ML for Protein Function Prediction & Design



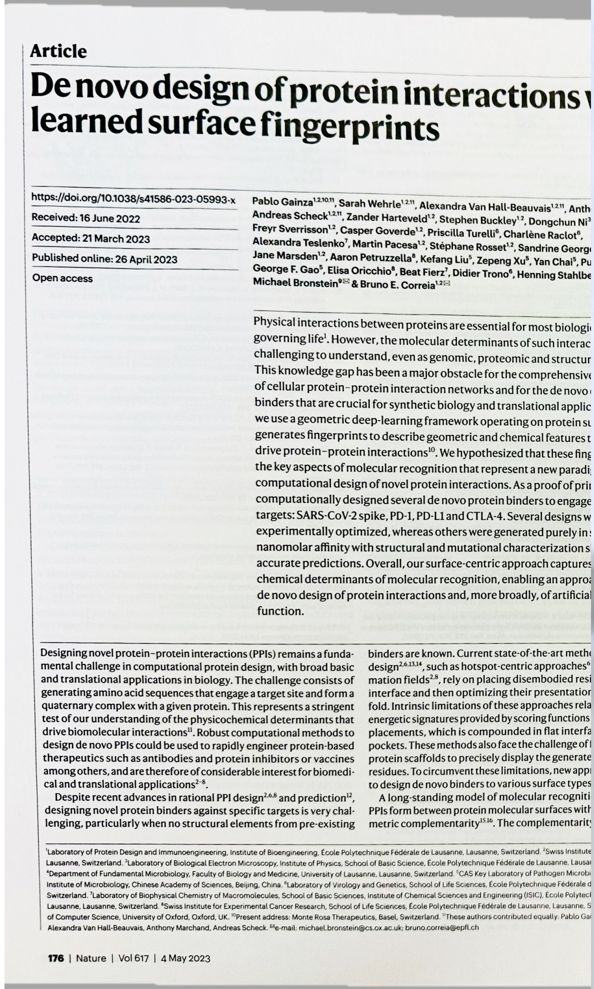

nature.com/nmeth

February 2020 Vol. 17 No. 2

nature methods

Protein interaction fingerprinting using deep learning
Improved 3D multicolor nanoscopy
Cryo-ET-based structure determination
Modeling intercellular communication
The Bioconductor project for single-cell analysis

Gainza et B, Correia 2020



Article

De novo design of protein interactions w
learned surface fingerprints

https://doi.org/10.1038/s41586-023-05993-x

Received: 16 June 2022

Accepted: 21 March 2023

Published online: 26 April 2023

Open access

Pablo Gainza[1,2,10,11], Sarah Wehrle[1,2,11], Alexandra Van Hall-Beauvais[1,2,11], Anth
Andreas Scheck[1,2,11], Zander Harteveld[1,2], Stephen Buckley[1,2], Dongchun Ni[3]
Freyr Sverrisson[1,2], Casper Goverde[1,2], Priscilla Turelli[6], Charlène Raclot[6],
Alexandra Teslenko[7], Martin Pacesa[1,2], Stéphane Rosset[1,2], Sandrine Georg
Jane Marsden[1,2], Aaron Petruzzella[8], Kefang Liu[5], Zepeng Xu[5], Yan Chai[5], Pu
George F. Gao[5], Elisa Oricchio[8], Beat Fierz[7], Didier Trono[6], Henning Stahlbe
Michael Bronstein[9]✉ & Bruno E. Correia[1,2]✉

Physical interactions between proteins are essential for most biologi
governing life[1]. However, the molecular determinants of such interac
challenging to understand, even as genomic, proteomic and structur
This knowledge gap has been a major obstacle for the comprehensive
of cellular protein–protein interaction networks and for the de novo
binders that are crucial for synthetic biology and translational applic
we use a geometric deep-learning framework operating on protein su
generates fingerprints to describe geometric and chemical features t
drive protein–protein interactions[10]. We hypothesized that these fing
the key aspects of molecular recognition that represent a new paradi
computational design of novel protein interactions. As a proof of prin
computationally designed several de novo protein binders to engage
targets: SARS-CoV-2 spike, PD-1, PD-L1 and CTLA-4. Several designs w
experimentally optimized, whereas others were generated purely in s
nanomolar affinity with structural and mutational characterization s
accurate predictions. Overall, our surface-centric approach captures
chemical determinants of molecular recognition, enabling an approa
de novo design of protein interactions and, more broadly, of artificial
function.

Designing novel protein–protein interactions (PPIs) remains a fundamental challenge in computational protein design, with broad basic and translational applications in biology. The challenge consists of generating amino acid sequences that engage a target site and form a quaternary complex with a given protein. This represents a stringent test of our understanding of the physicochemical determinants that drive biomolecular interactions[11]. Robust computational methods to design de novo PPIs could be used to rapidly engineer protein-based therapeutics such as antibodies and protein inhibitors or vaccines among others, and are therefore of considerable interest for biomedical and translational applications[2–8].

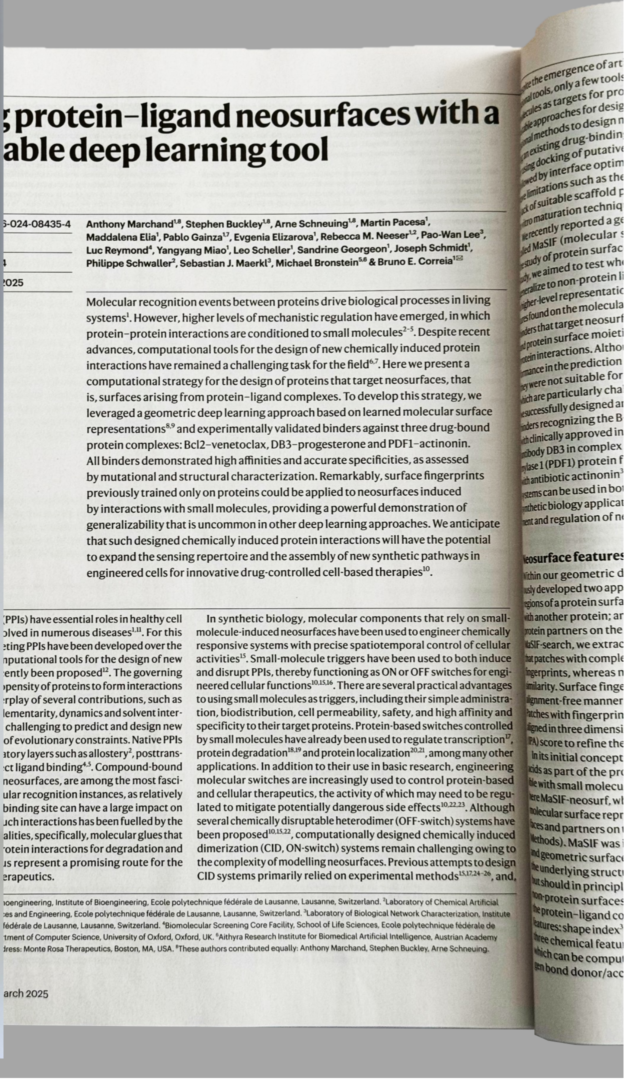
Despite recent advances in rational PPI design[2,6,8] and prediction[12], designing novel protein binders against specific targets is very challenging, particularly when no structural elements from pre-existing

binders are known. Current state-of-the-art metho
design[2,6,13,14], such as hotspot-centric approaches[6]
mation fields[2,8], rely on placing disembodied resi
interface and then optimizing their presentatior
fold. Intrinsic limitations of these approaches rela
energetic signatures provided by scoring functions
placements, which is compounded in flat interfa
pockets. These methods also face the challenge of
protein scaffolds to precisely display the generate
residues. To circumvent these limitations, new appr
to design de novo binders to various surface types

A long-standing model of molecular recogniti
PPIs form between protein molecular surfaces with
metric complementarity[15,16]. The complementarity

[1]Laboratory of Protein Design and Immunoengineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [2]Swiss Institute
Lausanne, Switzerland. [3]Laboratory of Biological Electron Microscopy, Institute of Physics, School of Basic Science, École Polytechnique Fédérale de Lausanne, Lausa
[4]Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland. [5]CAS Key Laboratory of Pathogen Microbi
Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. [6]Laboratory of Virology and Genetics, School of Life Sciences, École Polytechnique Fédérale d
Switzerland. [7]Laboratory of Biophysical Chemistry of Macromolecules, School of Basic Sciences, Institute of Chemical Sciences and Engineering (ISIC), École Polytec
Lausanne, Lausanne, Switzerland. [8]Swiss Institute for Experimental Cancer Research, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, S
of Computer Science, University of Oxford, Oxford, UK. [10]Present address: Monte Rosa Therapeutics, Basel, Switzerland. [11]These authors contributed equally: Pablo Ga
Alexandra Van Hall-Beauvais, Anthony Marchand, Andreas Scheck. ✉e-mail: michael.bronstein@cs.ox.ac.uk; bruno.correia@epfl.ch

176 | Nature | Vol 617 | 4 May 2023



g protein–ligand neosurfaces with a
able deep learning tool

6-024-08435-4

4

2025

Anthony Marchand[1,8], Stephen Buckley[1,8], Arne Schneuing[1,8], Martin Pacesa[1],
Maddalena Elia[1], Pablo Gainza[1,7], Evgenia Elizarova[1], Rebecca M. Neeser[1,2], Pao-Wan Lee[3],
Luc Reymond[4], Yangyang Miao[1], Leo Scheller[1], Sandrine Georgeon[1], Joseph Schmidt[1],
Philippe Schwaller[2], Sebastian J. Maerkl[3], Michael Bronstein[5,6] & Bruno E. Correia[1]✉

Molecular recognition events between proteins drive biological processes in living systems[1]. However, higher levels of mechanistic regulation have emerged, in which protein–protein interactions are conditioned to small molecules[2–5]. Despite recent advances, computational tools for the design of new chemically induced protein interactions have remained a challenging task for the field[6,7]. Here we present a computational strategy for the design of proteins that target neosurfaces, that is, surfaces arising from protein–ligand complexes. To develop this strategy, we leveraged a geometric deep learning approach based on learned molecular surface representations[8,9] and experimentally validated binders against three drug-bound protein complexes: Bcl2–venetoclax, DB3–progesterone and PDF1–actinonin. All binders demonstrated high affinities and accurate specificities, as assessed by mutational and structural characterization. Remarkably, surface fingerprints previously trained only on proteins could be applied to neosurfaces induced by interactions with small molecules, providing a powerful demonstration of generalizability that is uncommon in other deep learning approaches. We anticipate that such designed chemically induced protein interactions will have the potential to expand the sensing repertoire and the assembly of new synthetic pathways in engineered cells for innovative drug-controlled cell-based therapies[10].

(PPIs) have essential roles in healthy cell
olved in numerous diseases[1,11]. For this
eting PPIs have been developed over the
nputational tools for the design of new
ently been proposed[12]. The governing
pensity of proteins to form interactions
rplay of several contributions, such as
lementarity, dynamics and solvent inter-
challenging to predict and design new
of evolutionary constraints. Native PPIs
atory layers such as allostery[2], posttrans-
ct ligand binding[4,5]. Compound-bound
neosurfaces, are among the most fasci-
ular recognition instances, as relatively
binding site can have a large impact on
uch interactions has been fuelled by the
alities, specifically, molecular glues that
otein interactions for degradation and
is represent a promising route for the
erapeutics.

In synthetic biology, molecular components that rely on small-molecule-induced neosurfaces have been used to engineer chemically responsive systems with precise spatiotemporal control of cellular activities[15]. Small-molecule triggers have been used to both induce and disrupt PPIs, thereby functioning as ON or OFF switches for engineered cellular functions[10,15,16]. There are several practical advantages to using small molecules as triggers, including their simple administration, biodistribution, cell permeability, safety, and high affinity and specificity to their target proteins. Protein-based switches controlled by small molecules have already been used to regulate transcription[17], protein degradation[18,19] and protein localization[20,21], among many other applications. In addition to their use in basic research, engineering molecular switches are increasingly used to control protein-based and cellular therapeutics, the activity of which may need to be regulated to mitigate potentially dangerous side effects[10,22,23]. Although several chemically disruptable heterodimer (OFF-switch) systems have been proposed[10,15,22], computationally designed chemically induced dimerization (CID, ON-switch) systems remain challenging owing to the complexity of modelling neosurfaces. Previous attempts to design CID systems primarily relied on experimental methods[15,17,24–26], and,

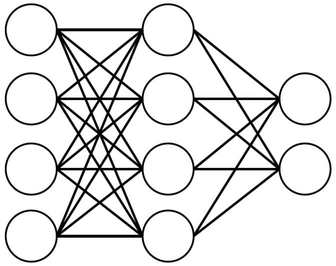
ite the emergence of art
nal tools, only a few tools
ecules as targets for pro
ble approaches for desig
nal methods to design n
n existing drug-bindin
sing docking of putative
wed by interface optim
e limitations such as the
k of suitable scaffold p
tro maturation techniq
We recently reported a ge
led MaSIF (molecular s
e study of protein surfac
dy, we aimed to test wh
eralize to non-protein li
gher-level representatio
es found on the molecula
nders that target neosurf
d protein surface moieti
otein interactions. Altho
mance in the prediction
ey were not suitable for
ich are particularly cha
e successfully designed a
nders recognizing the B
th clinically approved in
tibody DB3 in complex
ylase 1 (PDF1) protein f
th antibiotic actinonin[3]
stems can be used in bo
nthetic biology applica
ent and regulation of n

eosurface features

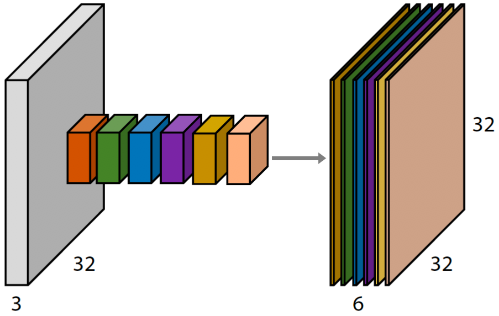
ithin our geometric d
usly developed two app
egions of a protein surfa
ith another protein; ar
rotein partners on the
aSIF-search, we extrac
at patches with comple
ngerprints, whereas n
imilarity. Surface finge
lignment-free manner
Patches with fingerprin
igned in three dimensi
PA) score to refine the
In its initial concept
cids as part of the pro
ble with small molecu
ere MaSIF-neosurf, wh
olecular surface repr
ces and partners on t
Methods). MaSIF was i
nd geometric surface
he underlying struct
ut should in principl
on-protein surfaces
he protein–ligand co
eatures: shape index[3]
hree chemical featu
which can be comput
gen bond donor/acc

oengineering, Institute of Bioengineering, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland. [2]Laboratory of Chemical Artificial
ces and Engineering, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland. [3]Laboratory of Biological Network Characterization, Institute
fédérale de Lausanne, Lausanne, Switzerland. [4]Biomolecular Screening Core Facility, School of Life Sciences, Ecole polytechnique fédérale de
tment of Computer Science, University of Oxford, Oxford, UK. [6]Aithyra Research Institute for Biomedical Artificial Intelligence, Austrian Academy
dress: Monte Rosa Therapeutics, Boston, MA, USA. [8]These authors contributed equally: Anthony Marchand, Stephen Buckley, Arne Schneuing.

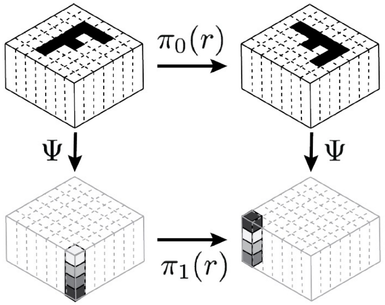
arch 2025



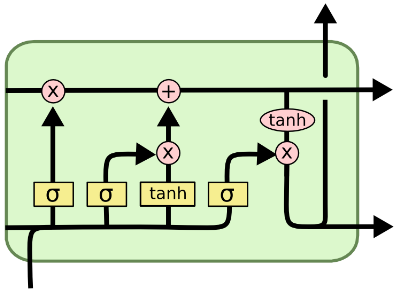
Perceptrons
Function regularity



CNNs
Translation

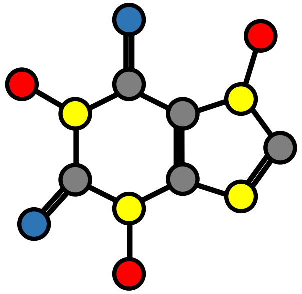


Group-CNNs
Translation+Rotation,
Global groups



LSTMs
Time warping

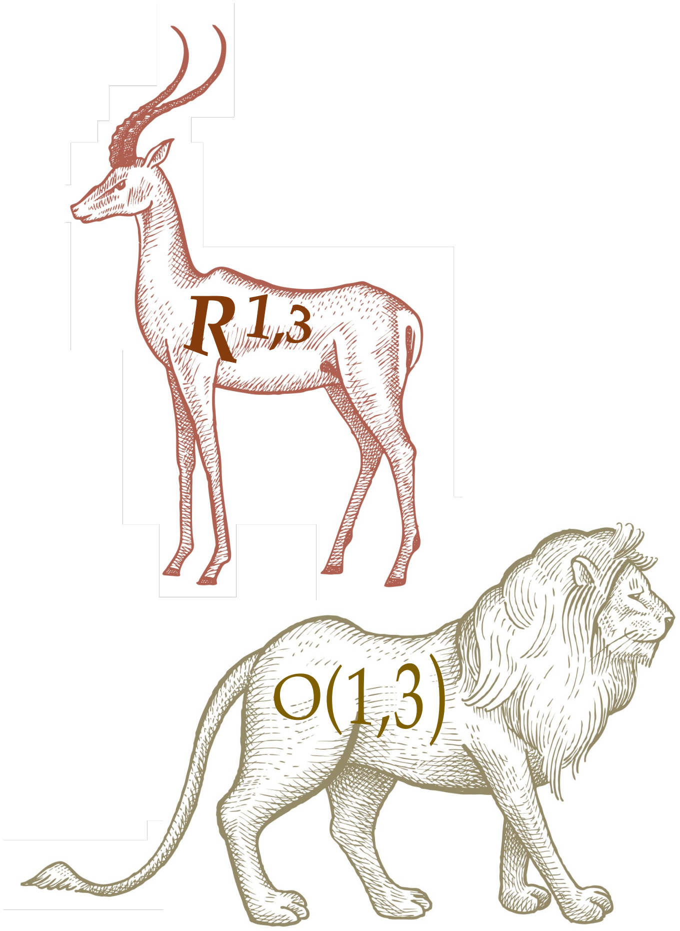
DeepSets / Transformers
Permutation



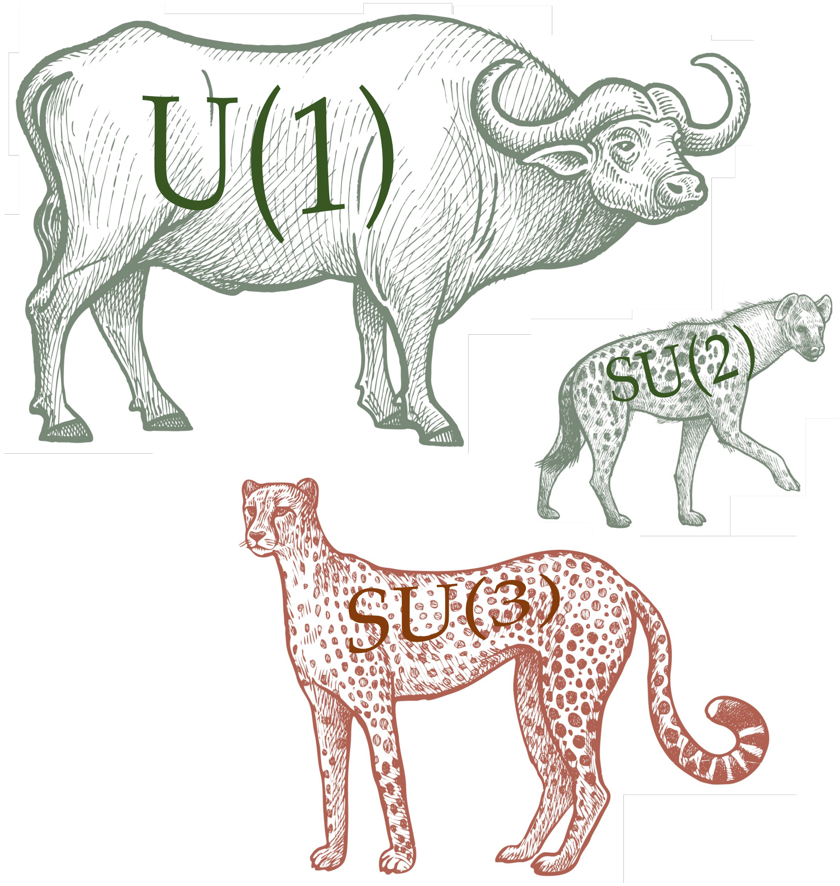
GNNs
Permutation

Intrinsic CNNs
Isometry / Gauge choice

ANY NN ARCHITECTURE
GEOMETRIC DEEP LEARNING
TikTok
@tired_actor

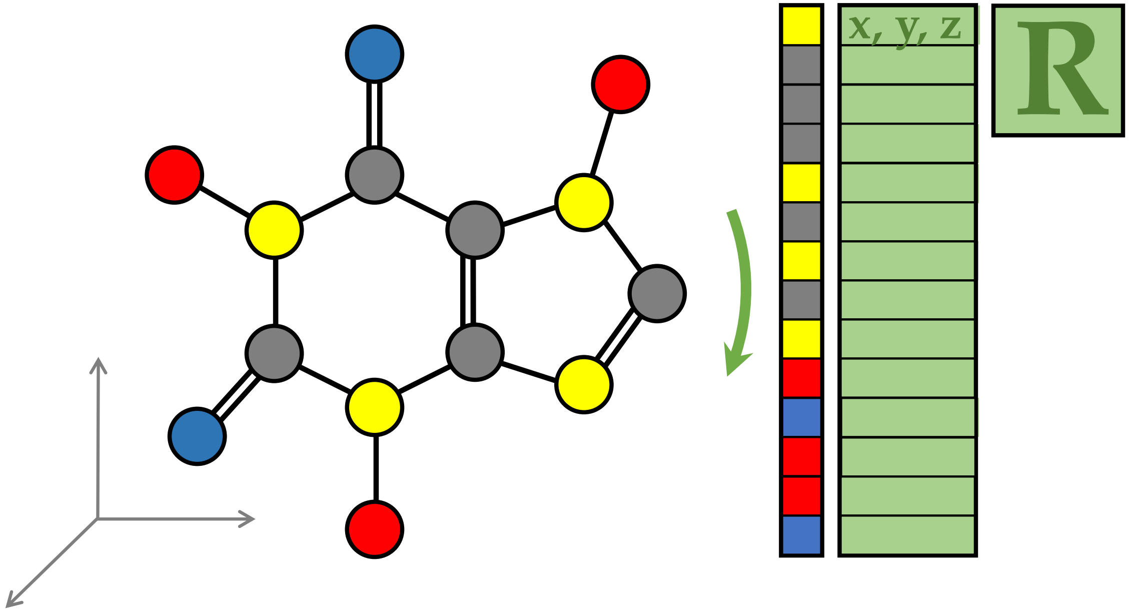


External symmetry



Internal symmetry

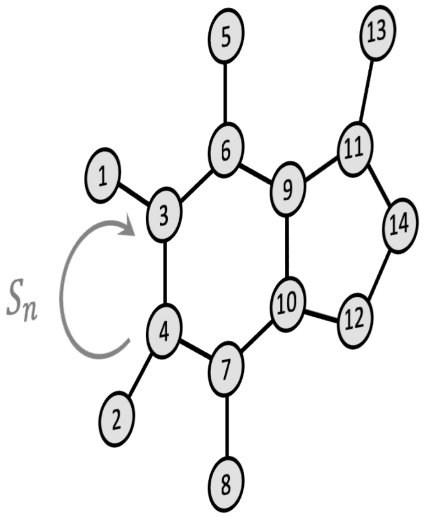
SO(3)-*invariance*



“properties of a molecule do not change if we rotate it”

Geometric ("Equivariant") Graph Neural Networks

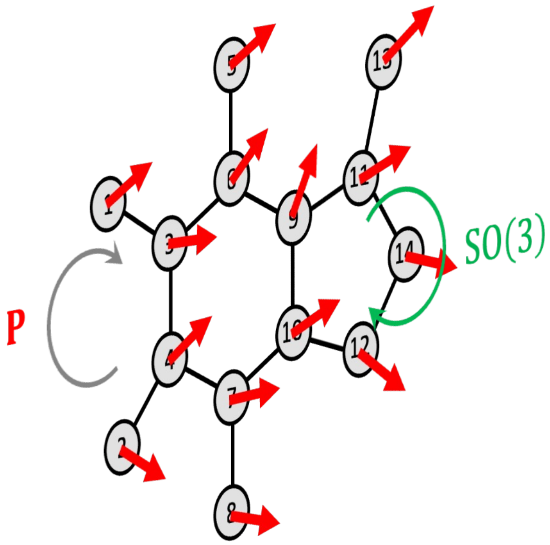
Geometric Graph G



Permutation group S_n

"domain symmetry"

Node features $\mathcal{X}(G)$



Permutation matrix P

Rotation R

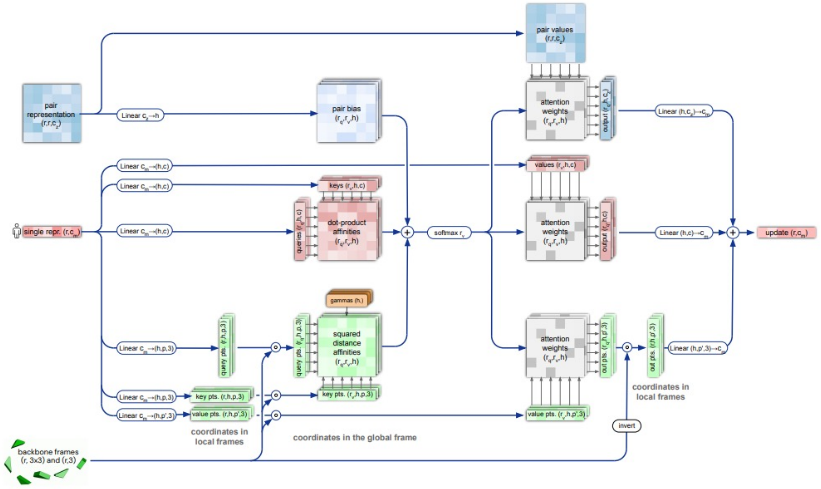
"data symmetry"

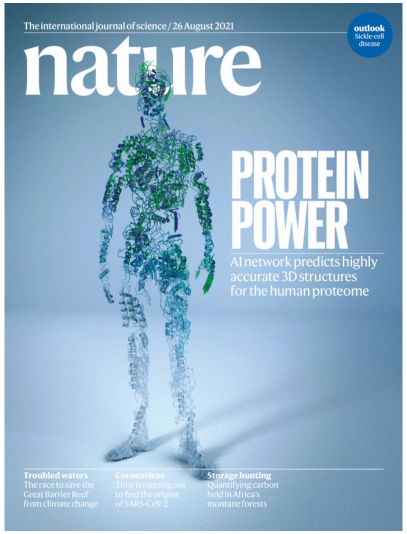
Functions $\mathcal{F}(\mathcal{X}(G))$

Geometric message passing

$$\mathbf{F}\left(\mathbf{PXR}, \mathbf{PAP}^\top\right) = \mathbf{PF}(\mathbf{X}, \mathbf{A})\mathbf{R}$$

Revolution in Structural Biology





The international journal of science / 26 August 2021
outlook Sickle-cell disease
nature
PROTEIN POWER
AI network predicts highly accurate 3D structures for the human proteome
Troubled waters The race to save the Great Barrier Reef from climate change
Coronavirus Time is running out to find the origins of SARS-CoV-2
Storage hunting Quantifying carbon held in Africa's montane forests

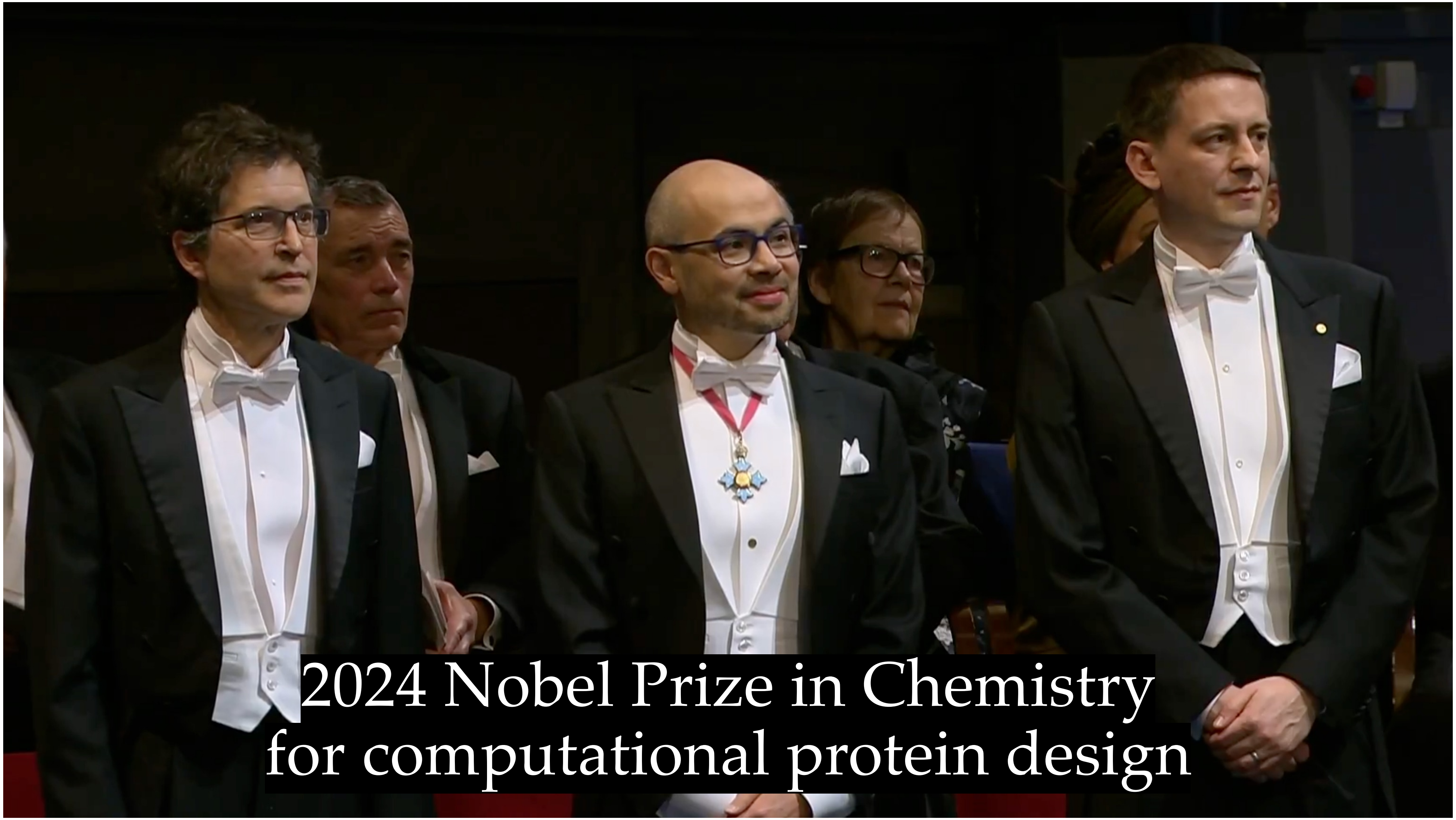
Jumper et al. 2021

SARS-CoV-2 is evolving fast. What will it do next? p. 844
Effects on surface water from hydraulic fracturing pp. 853 & 896
Baby bats babble like human babies p. 923
Science
$15
20 AUGUST 2021
sciencemag.org
AAAS
PREDICTING STRUCTURES
Deep learning accurately folds proteins p. 871

Baek et al. 2021

AlphaFold 2
“Invariant point attention”

RosettaFold
SE(3)-equivariant Transformer



2024 Nobel Prize in Chemistry
for computational protein design

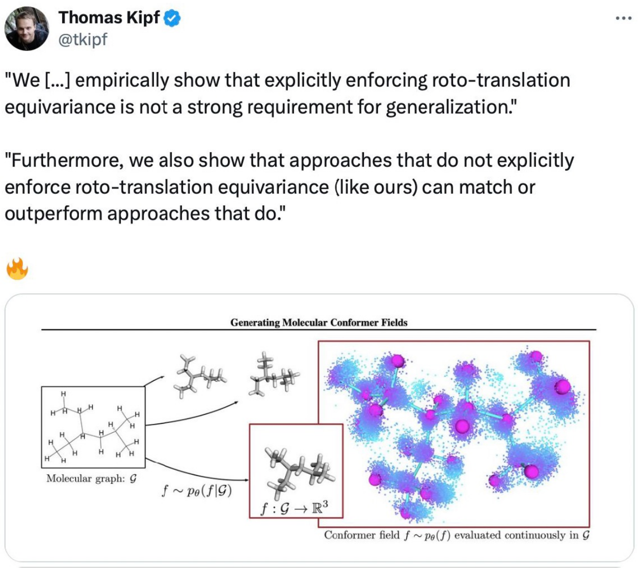
AND THEY LIVED

HAPPILY EVER AFTER

The Bitter Lesson: Equivariance is dead…long live equivariance!

Equivariance is the idea of giving a model the inductive biases to natively handle rotations, translations and (sometimes) reflections. It has been at the core of Geometric Deep Learning and biomolecular modelling research since AlphaFold 2. However, recent works by top labs have questioned the existing mantra.

- The first shots were fired by Apple, with a paper that obtained SOTA results on predicting the 3D structures of small molecules using a non-equivariant diffusion model with a transformer encoder.
- Remarkably, the authors showed that using the domain-agnostic model did not deleteriously impact generalization and was consistently able to outperform specialist models (assuming sufficient scale was used).
- Next was AlphaFold 3, which infamously dropped all the equivariance and frames constraints from the previous model in favour of another diffusion process coupled with augmentations and, of course, scale.
- Regardless, the greatly improved training efficiency of equivariant models means the practice is likely to stay for a while (at least for academic groups working on large systems such as proteins).



Thomas Kipf
@tkipf

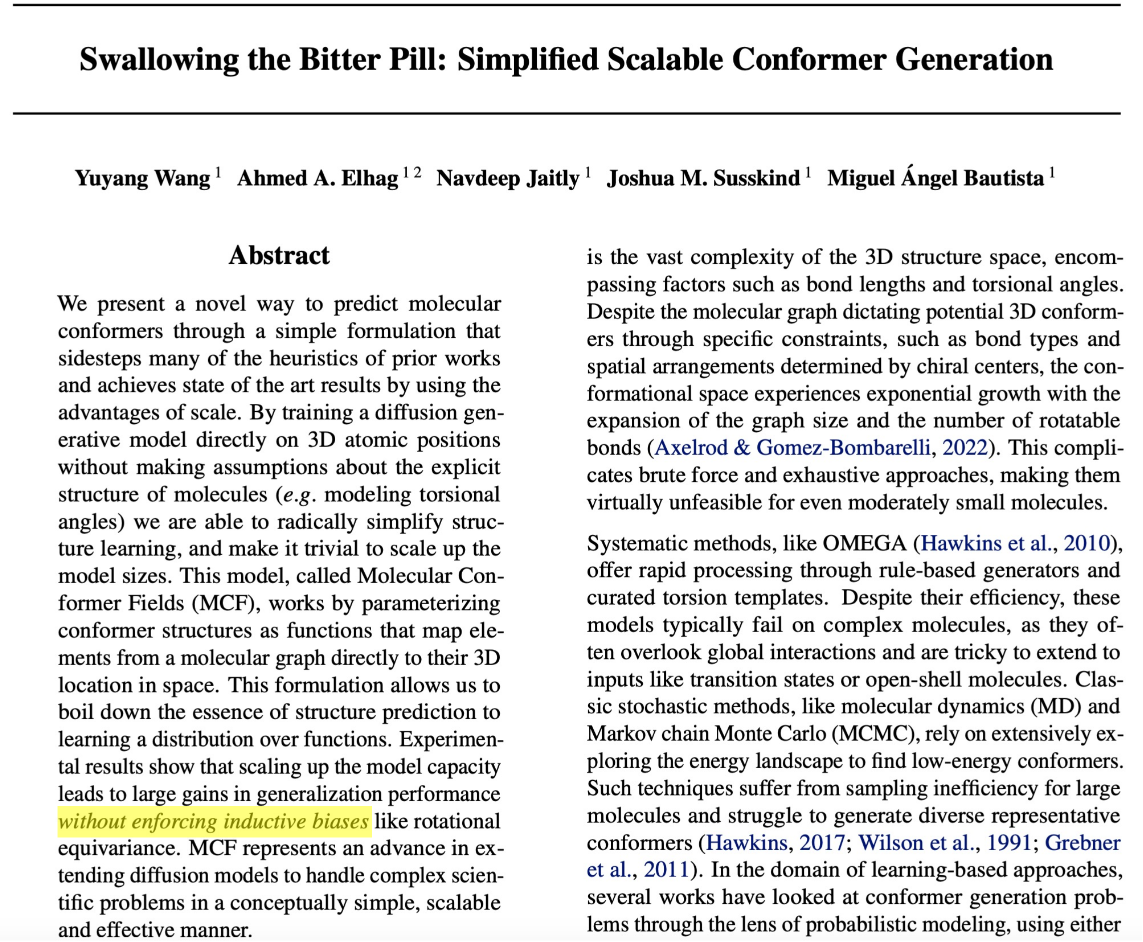
"We [...] empirically show that explicitly enforcing roto-translation equivariance is not a strong requirement for generalization."

"Furthermore, we also show that approaches that do not explicitly enforce roto-translation equivariance (like ours) can match or outperform approaches that do."

🔥

Generating Molecular Conformer Fields
Molecular graph: $\mathcal{G}$
$f \sim p_\theta(f|\mathcal{G})$
$f : \mathcal{G} \to \mathbb{R}^3$
Conformer field $f \sim p_\theta(f)$ evaluated continuously in $\mathcal{G}$

Benaich et al., State of AI report 2024



Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yuyang Wang [1] Ahmed A. Elhag [1 2] Navdeep Jaitly [1] Joshua M. Susskind [1] Miguel Ángel Bautista [1]

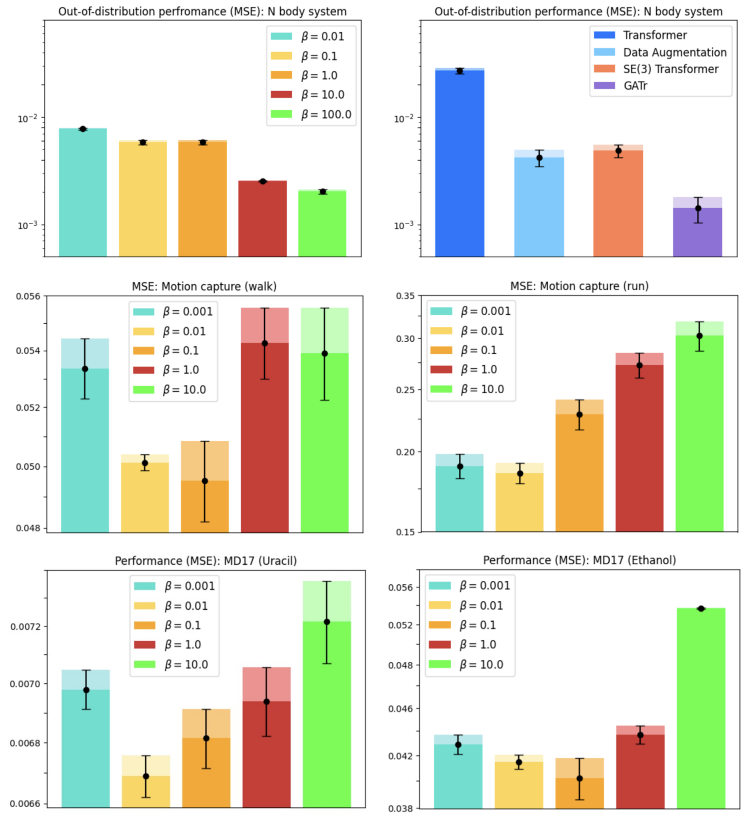
Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state of the art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (*e.g.* modeling torsional angles) we are able to radically simplify structure learning, and make it trivial to scale up the model sizes. This model, called Molecular Conformer Fields (MCF), works by parameterizing conformer structures as functions that map elements from a molecular graph directly to their 3D location in space. This formulation allows us to boil down the essence of structure prediction to learning a distribution over functions. Experimental results show that scaling up the model capacity leads to large gains in generalization performance *without enforcing inductive biases* like rotational equivariance. MCF represents an advance in extending diffusion models to handle complex scientific problems in a conceptually simple, scalable and effective manner.

is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Axelrod & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules.

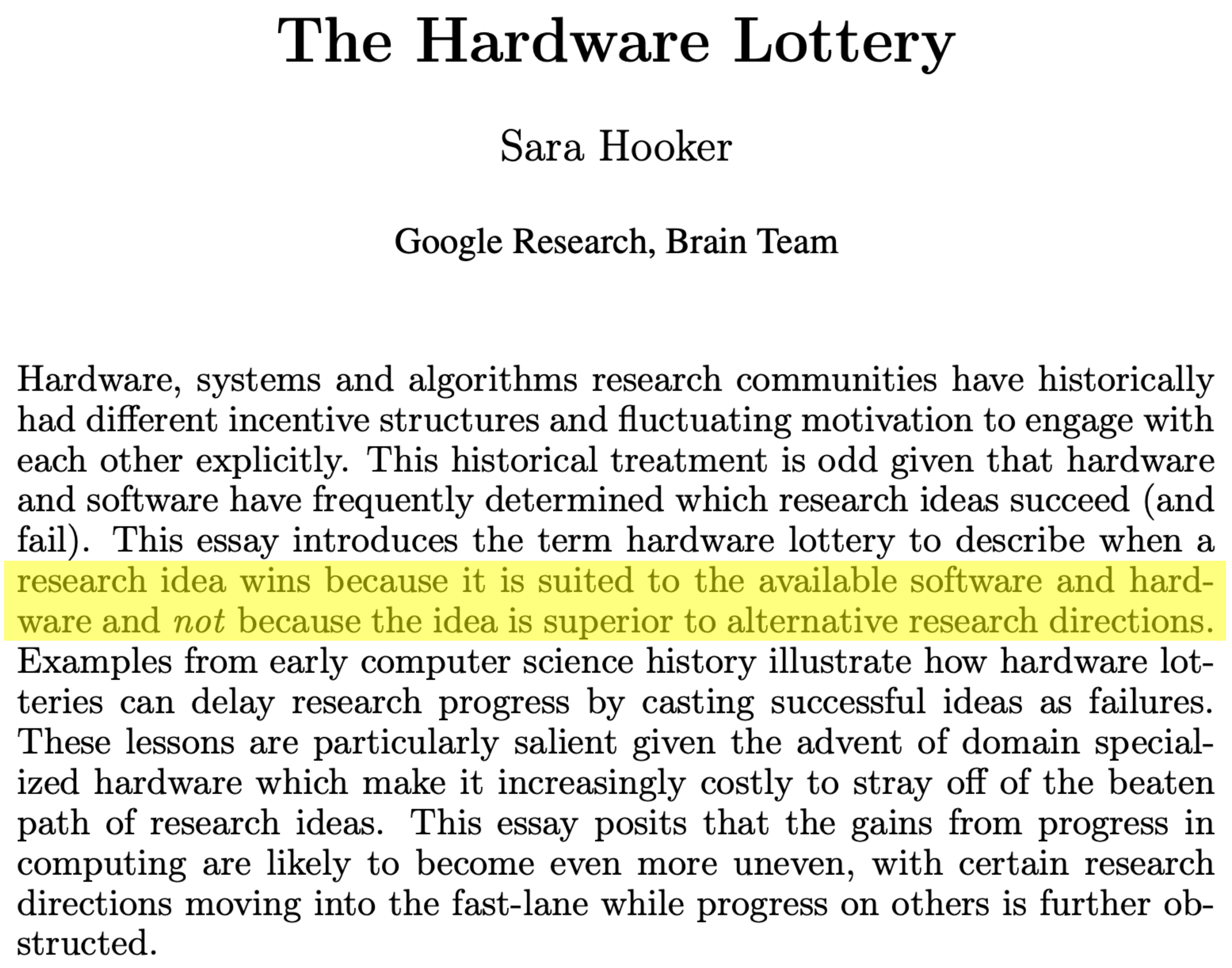
Systematic methods, like OMEGA (Hawkins et al., 2010), offer rapid processing through rule-based generators and curated torsion templates. Despite their efficiency, these models typically fail on complex molecules, as they often overlook global interactions and are tricky to extend to inputs like transition states or open-shell molecules. Classic stochastic methods, like molecular dynamics (MD) and Markov chain Monte Carlo (MCMC), rely on extensively exploring the energy landscape to find low-energy conformers. Such techniques suffer from sampling inefficiency for large molecules and struggle to generate diverse representative conformers (Hawkins, 2017; Wilson et al., 1991; Grebner et al., 2011). In the domain of learning-based approaches, several works have looked at conformer generation problems through the lens of probabilistic modeling, using either

Wang, Elhag et al. 2023



Elhag et B, 2024

EQUIVARIANCE
BRUTE-FORCE
DATA AUG-MENTATION
Quo vadimus?

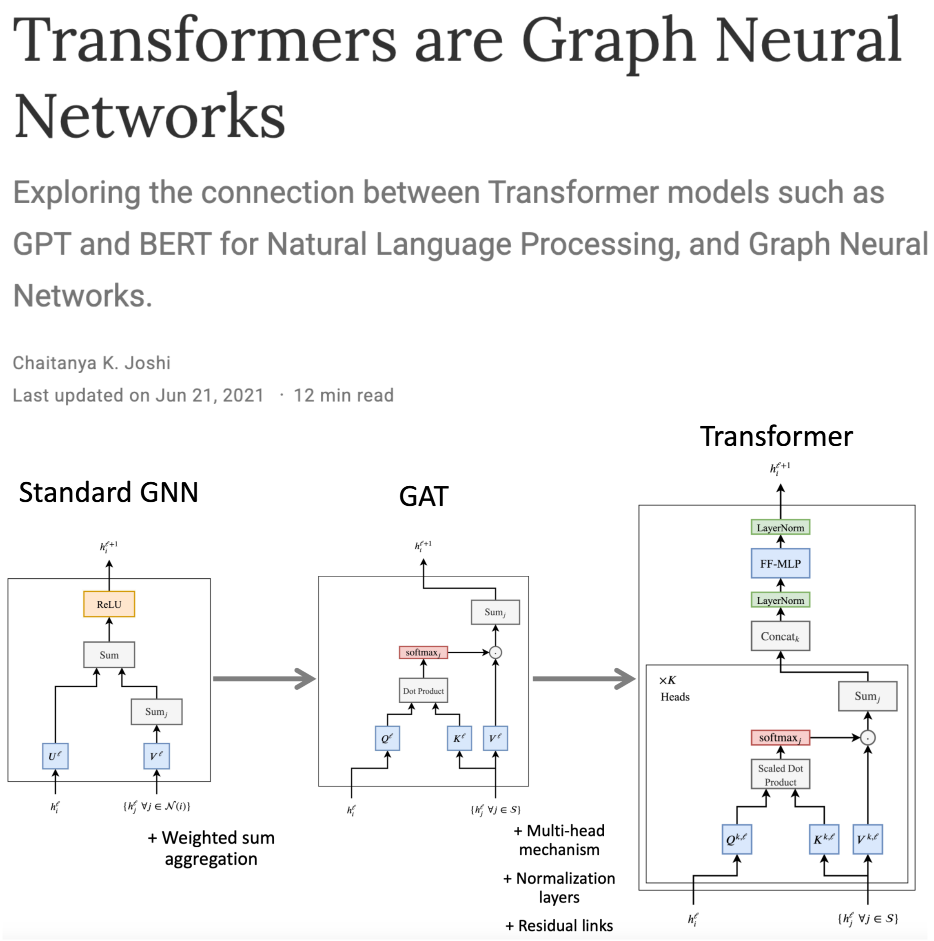


The Hardware Lottery

Sara Hooker

Google Research, Brain Team

Hardware, systems and algorithms research communities have historically had different incentive structures and fluctuating motivation to engage with each other explicitly. This historical treatment is odd given that hardware and software have frequently determined which research ideas succeed (and fail). This essay introduces the term hardware lottery to describe when a research idea wins because it is suited to the available software and hardware and *not* because the idea is superior to alternative research directions. Examples from early computer science history illustrate how hardware lotteries can delay research progress by casting successful ideas as failures. These lessons are particularly salient given the advent of domain specialized hardware which make it increasingly costly to stray off of the beaten path of research ideas. This essay posits that the gains from progress in computing are likely to become even more uneven, with certain research directions moving into the fast-lane while progress on others is further obstructed.



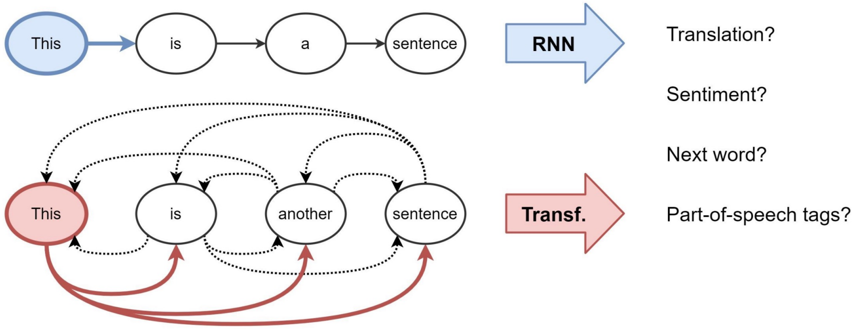
Transformers are Graph Neural Networks

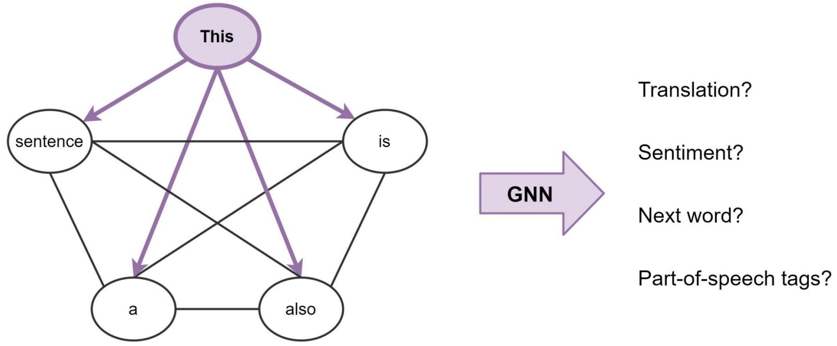
Exploring the connection between Transformer models such as GPT and BERT for Natural Language Processing, and Graph Neural Networks.

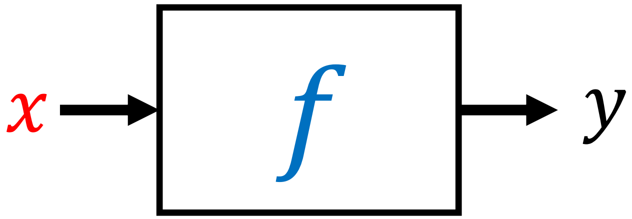
Chaitanya K. Joshi

Last updated on Jun 21, 2021 · 12 min read

Joshi 2021

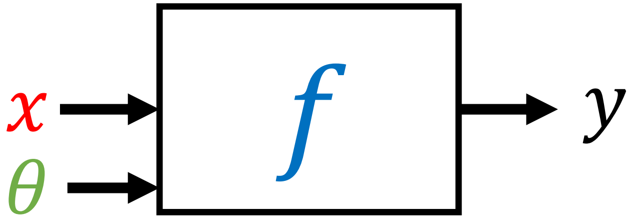






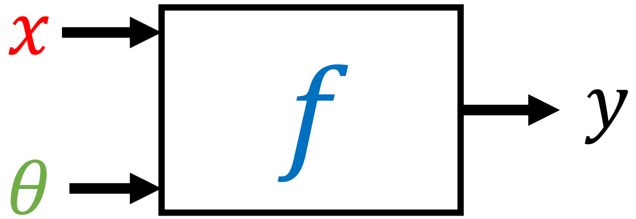
“How f interacts with the group G acting on x?”

$$f(g.x) = f(x)$$



"How f interacts with the group G acting on x?"

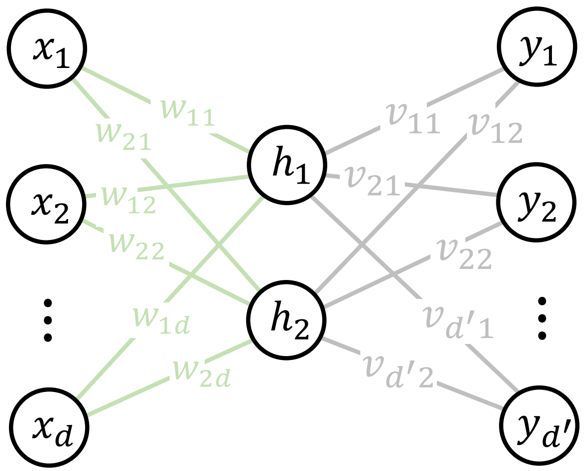
$$f(g.x) = f(x)$$

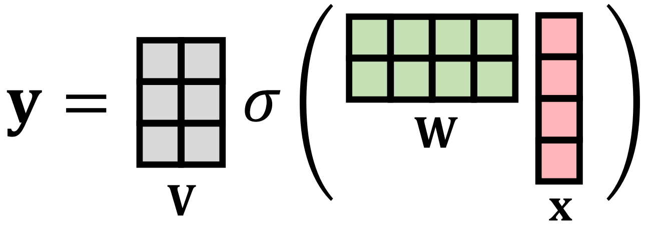


"How f interacts with the group G acting on x and H acting on θ?"

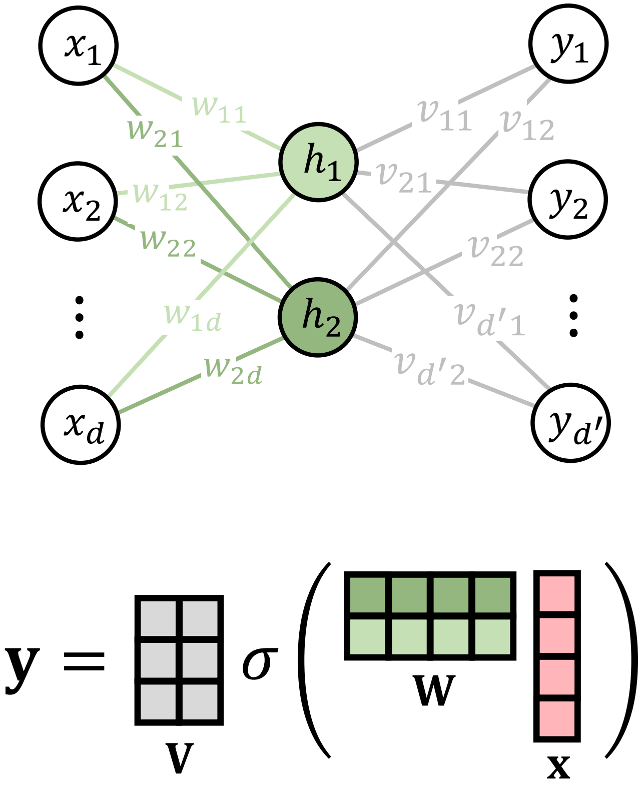
$$f(g.x, h.\theta) = f(x, \theta)$$

Symmetries of the Weights

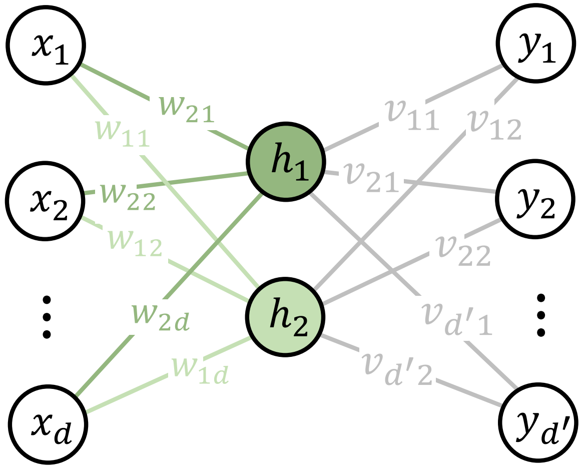


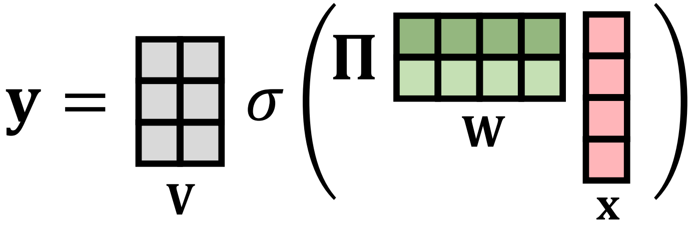


Symmetries of the Weights

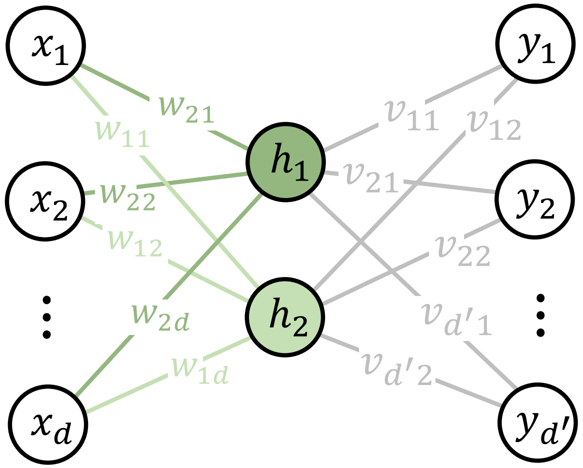


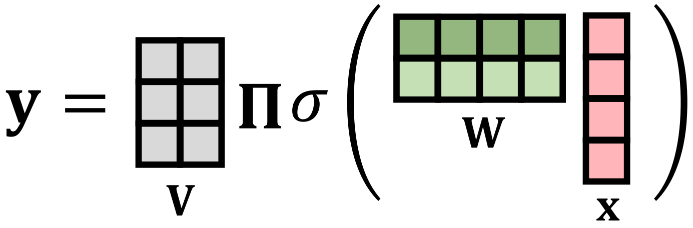
Symmetries of the Weights



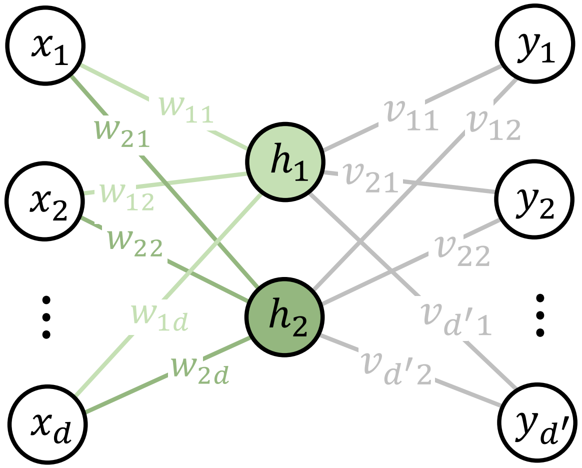


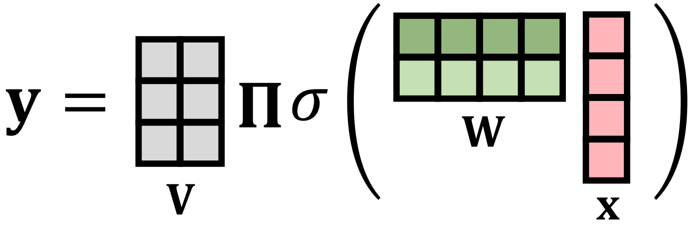
Symmetries of the Weights



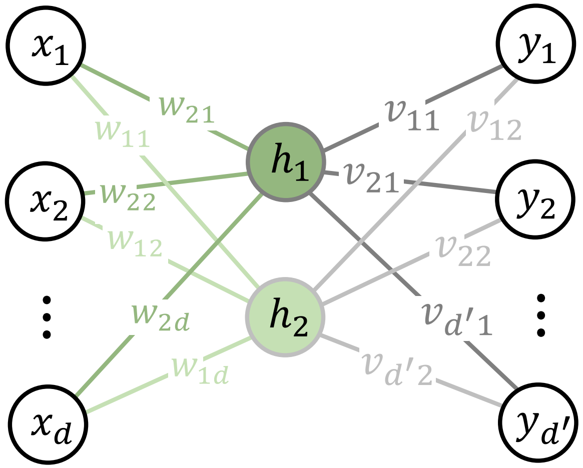


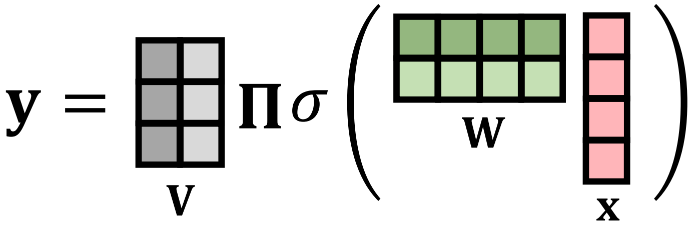
Symmetries of the Weights



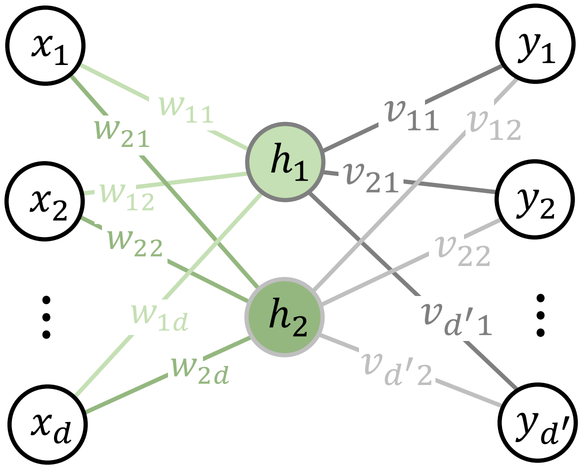


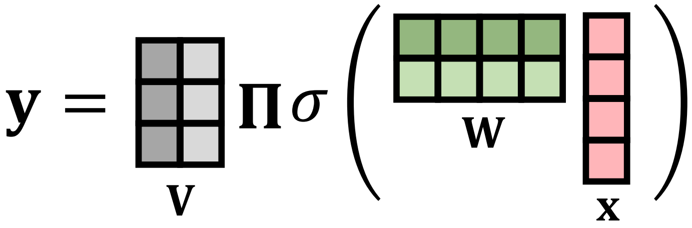
Symmetries of the Weights



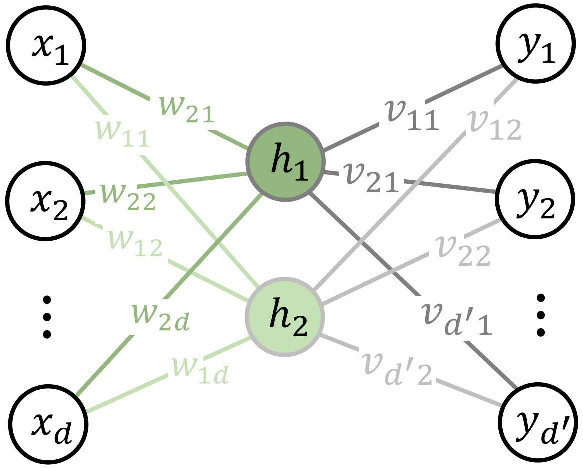


Symmetries of the Weights



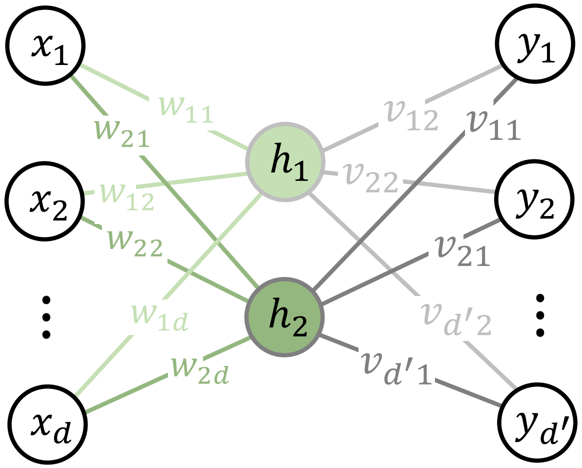


Symmetries of the Weights



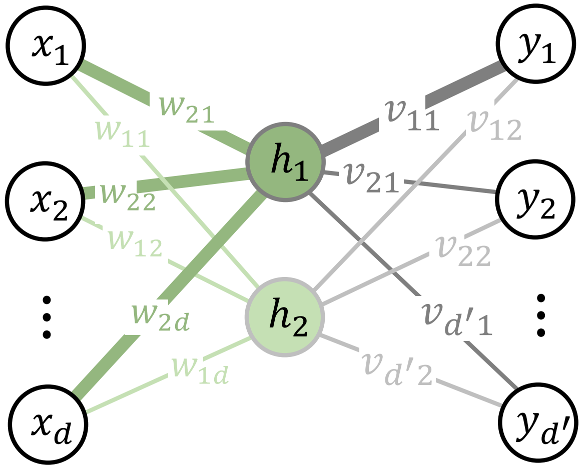
$$\mathbf{y} = \underbrace{\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}}_{\mathbf{V}} \mathbf{\Pi}^{\mathrm{T}}\mathbf{\Pi}\,\sigma\left(\underbrace{\begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}}_{\mathbf{W}} \underbrace{\begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}}_{\mathbf{x}} \right)$$

Symmetries of the Weights



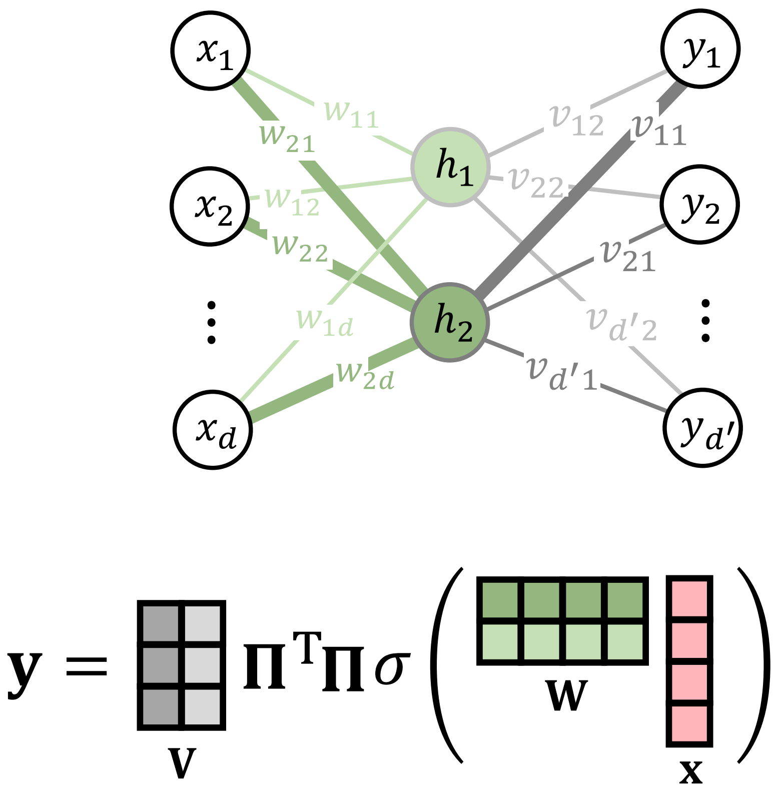
$$\mathbf{y} = \underset{\mathbf{V}}{\begin{bmatrix} \; & \; \\ \; & \; \\ \; & \; \end{bmatrix}}\,\mathbf{\Pi}^{\mathrm{T}}\mathbf{\Pi}\,\sigma\left(\underset{\mathbf{W}}{\begin{bmatrix} \; & \; & \; & \; \\ \; & \; & \; & \; \end{bmatrix}}\;\underset{\mathbf{x}}{\begin{bmatrix} \; \\ \; \\ \; \\ \; \end{bmatrix}}\right)$$

Symmetries of the Weights



$$\mathbf{y} = \underset{\mathbf{V}}{\boxed{}}\,\mathbf{\Pi}^{\mathrm{T}}\mathbf{\Pi}\,\sigma\left(\underset{\mathbf{W}}{\boxed{}}\;\underset{\mathbf{x}}{\boxed{}}\right)$$

Symmetries of the Weights



Symmetries of the Weights

- L-layer neural network with weights $\boldsymbol{\theta} = (\mathbf{W}_1, \mathbf{b}_1, \dots, \mathbf{W}_L, \mathbf{b}_L)$
- Parameter space symmetry $G = \mathrm{S}_{d_1} \times \cdots \times \mathrm{S}_{d_L}$

$$\begin{aligned}
\mathbf{W}_1' &= \boldsymbol{\Pi}_1^{\mathrm{T}} \mathbf{W}_1 & \mathbf{b}_1' &= \boldsymbol{\Pi}_1^{\mathrm{T}} \mathbf{b}_1 \\
\mathbf{W}_l' &= \boldsymbol{\Pi}_l^{\mathrm{T}} \mathbf{W}_l \boldsymbol{\Pi}_{l-1} & \mathbf{b}_l' &= \boldsymbol{\Pi}_l^{\mathrm{T}} \mathbf{b}_l \\
&\vdots & &\vdots \\
\mathbf{W}_L' &= \mathbf{W}_L \boldsymbol{\Pi}_{L-1} & \mathbf{b}_L' &= \mathbf{b}_L
\end{aligned}$$

such that $f(\cdot, g\boldsymbol{\theta}) = f(\cdot, \boldsymbol{\theta})$

Symmetries in the Gradient Space

$$\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{B}} \underset{\mathcal{L}_{(\mathbf{x},\mathbf{y})}}{\ell(f(\mathbf{x}, \boldsymbol{\theta}), \mathbf{y})}$$

$$\nabla_{\mathbf{b}_l} \mathcal{L}_{(\mathbf{x},\mathbf{y})} = \mathbf{g}_l \qquad \text{where } \mathbf{g}_l = \frac{\partial \ell(f(\mathbf{x},\boldsymbol{\theta}),\mathbf{y})}{\partial \mathbf{u}_l}$$

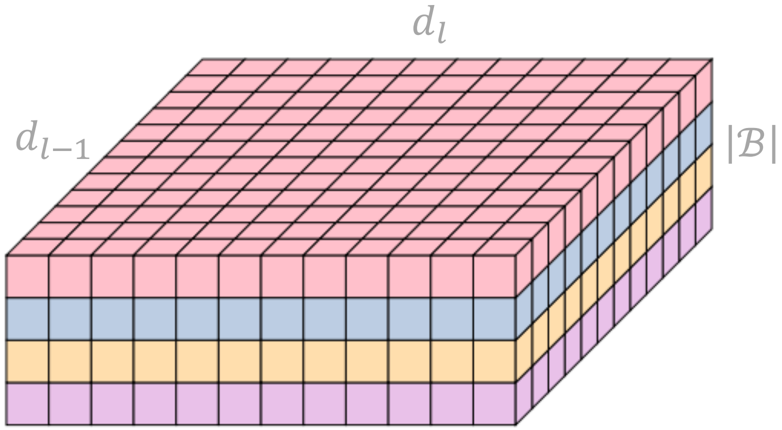
$$\nabla_{\mathbf{W}_l} \mathcal{L}_{(\mathbf{x},\mathbf{y})} = \mathbf{g}_l \mathbf{a}_{l-1}^{\mathrm{T}}$$

Gelberg et B, Maron 2025

Symmetries in the Gradient Space

$$\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{B}} \ell(f(\mathbf{x}, \boldsymbol{\theta}), \mathbf{y})$$

$$\nabla_{\mathbf{W}_l} =$$

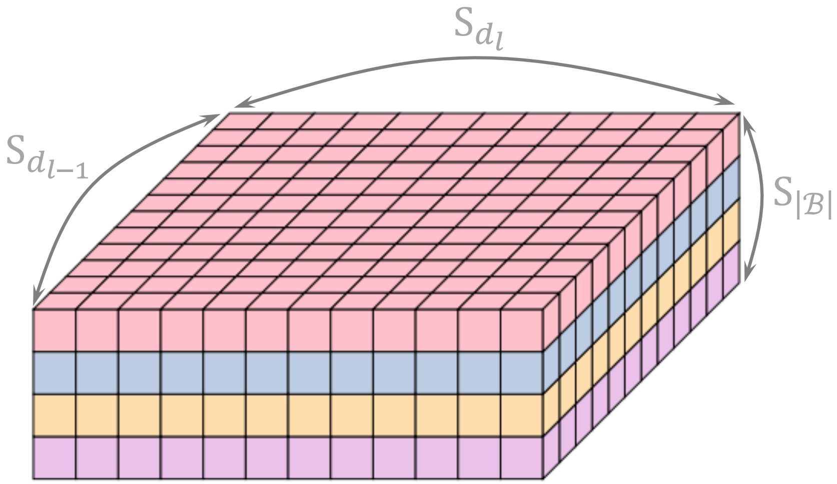


Gelberg et B, Maron 2025

Symmetries in the Gradient Space

$$\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{B}} \ell(f(\mathbf{x}, \boldsymbol{\theta}), \mathbf{y})$$

$\nabla_{\mathbf{W}_l} =$

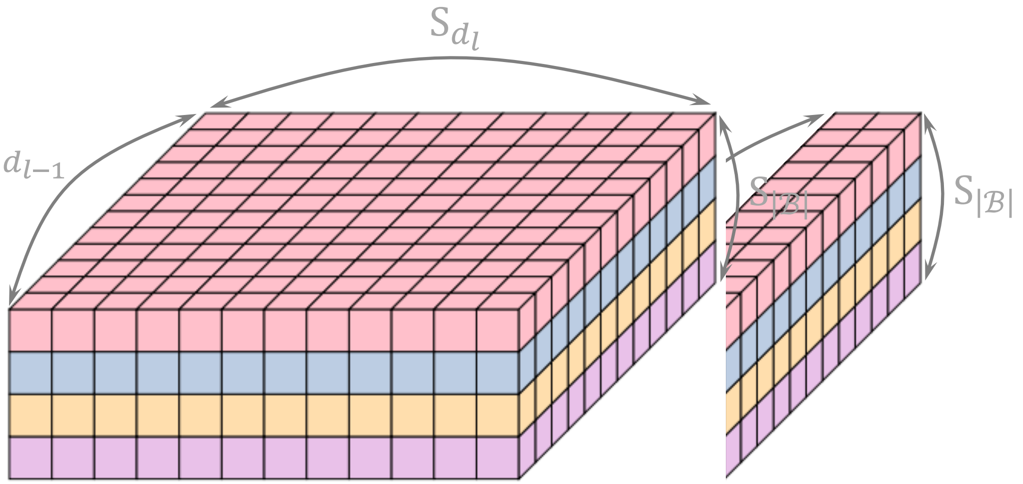


Gelberg et B, Maron 2025

Symmetries in the Gradient Space

$$\mathcal{L}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|} \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{B}} \ell(f(\mathbf{x},\boldsymbol{\theta}),\mathbf{y})$$

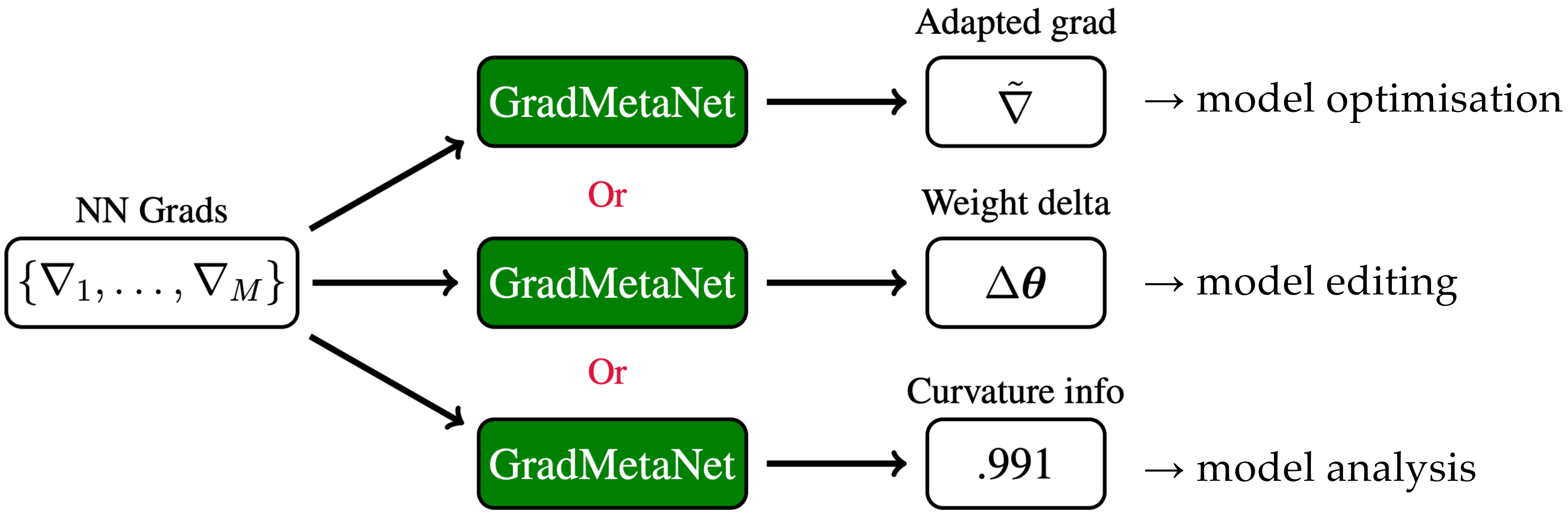
$\nabla_{\mathbf{W}_l} =$



Gelberg et B, Maron 2025

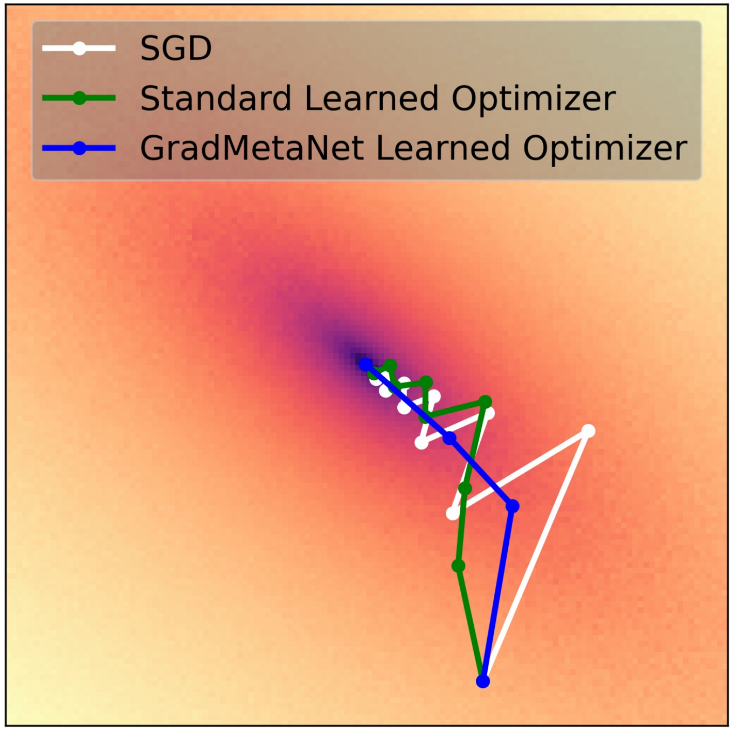
$$G_b = S_{|\mathcal{B}|} \times S_{d_1} \times \cdots \times S_{d_L}$$

GradMetaNet



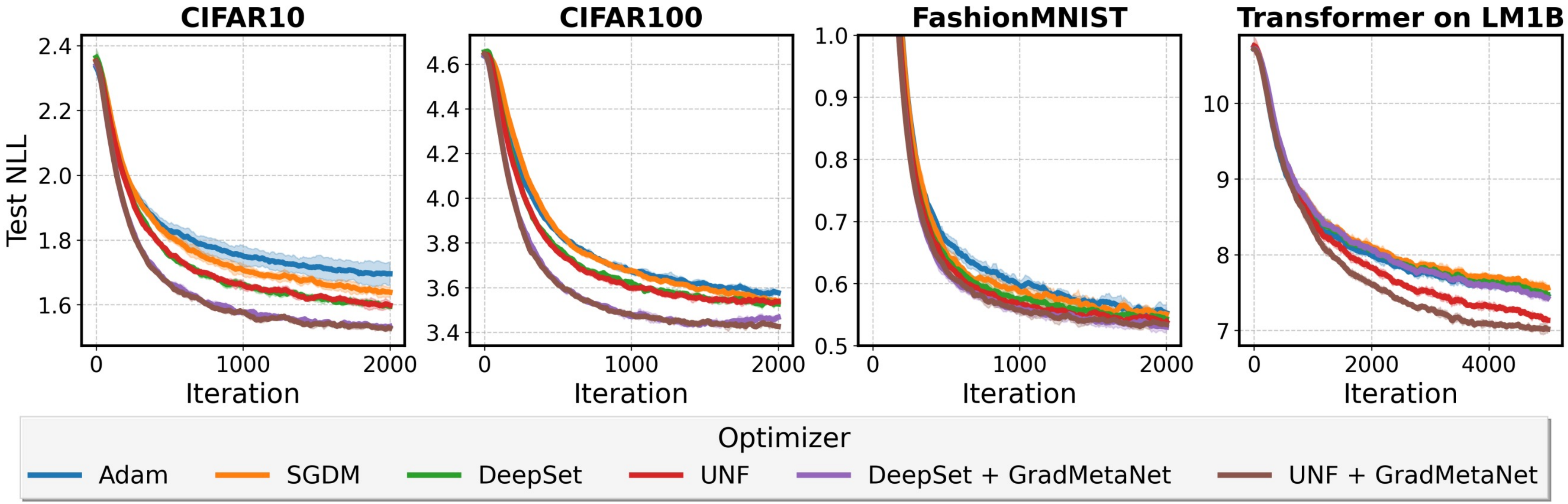
Gelberg et B, Maron 2025

GradMetaNet: Learning Optimisation



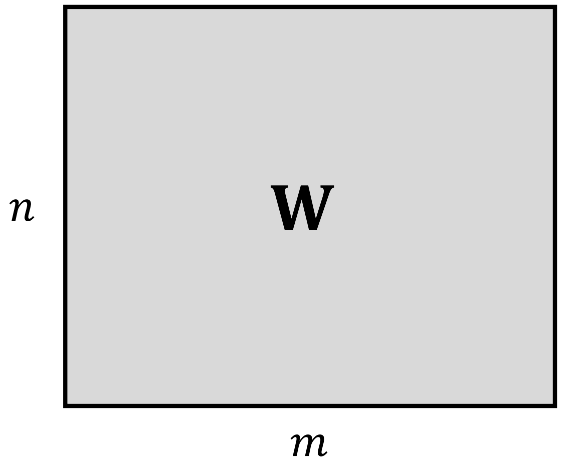
Gelberg et B, Maron 2025

GradMetaNet: Learning Optimisation

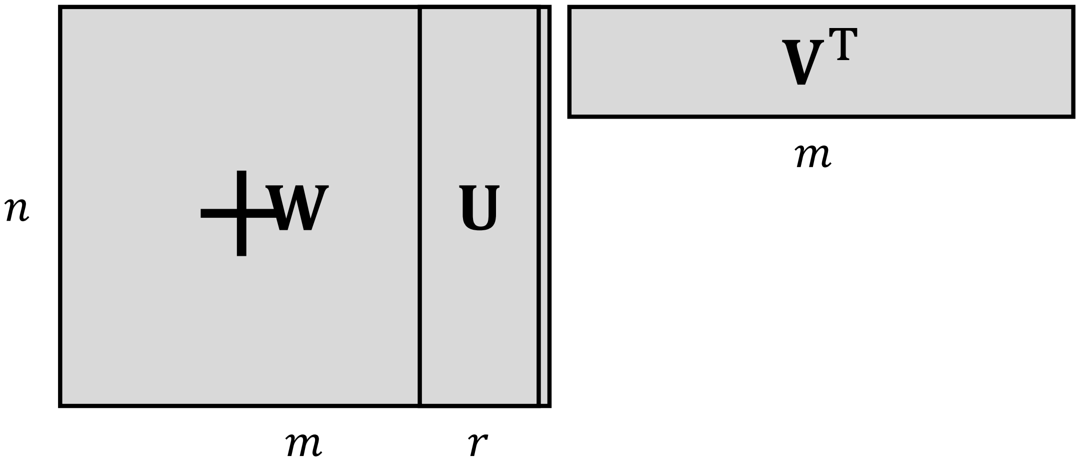


Gelberg et B, Maron 2025

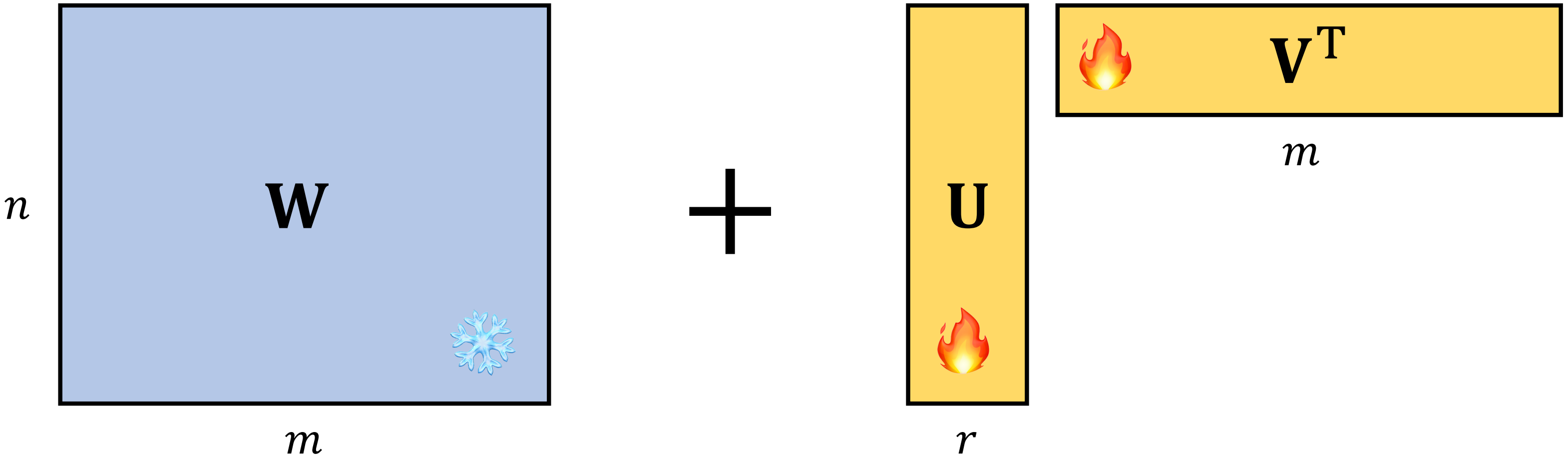
Low-Rank Adaptors (LoRA)



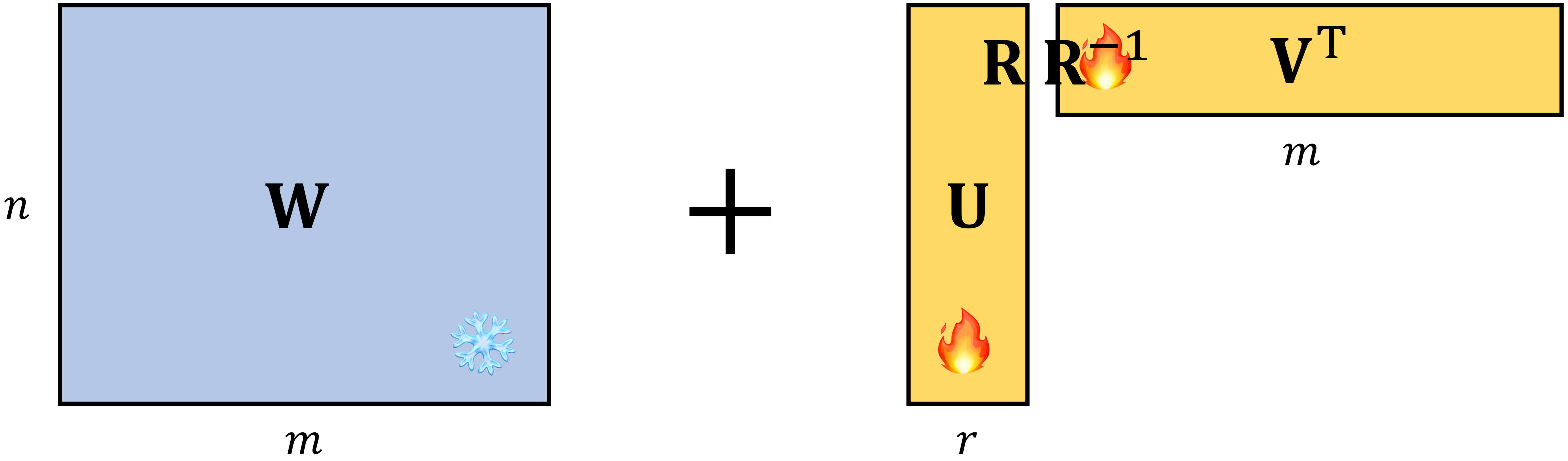
Low-Rank Adaptors (LoRA)



Low-Rank Adaptors (LoRA)



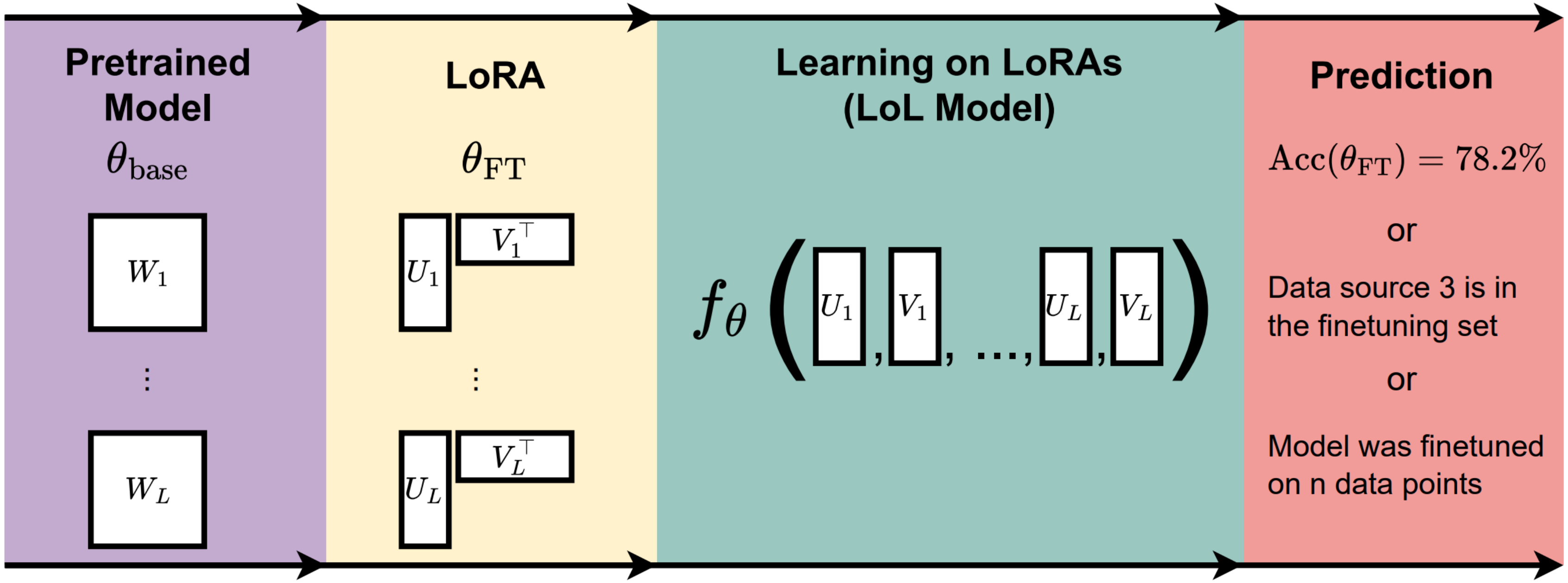
Low-Rank Adaptors (LoRA)



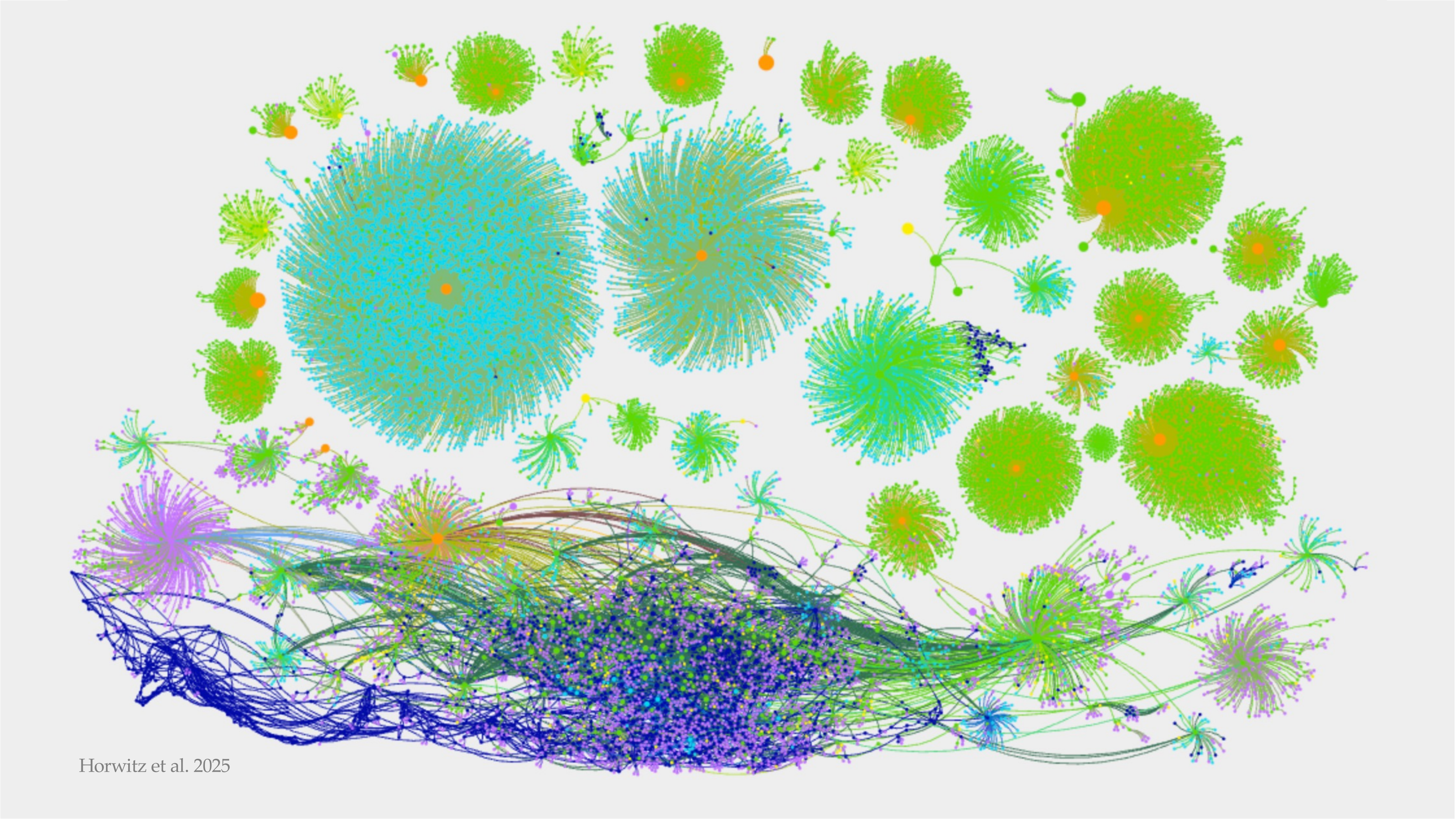
LoRA $(\mathbf{U}, \mathbf{V})$ *defined up to* $\mathrm{GL}(r)$

Putterman, Lim, Gelberg et B 2025

Learning on LoRAs (LOL)

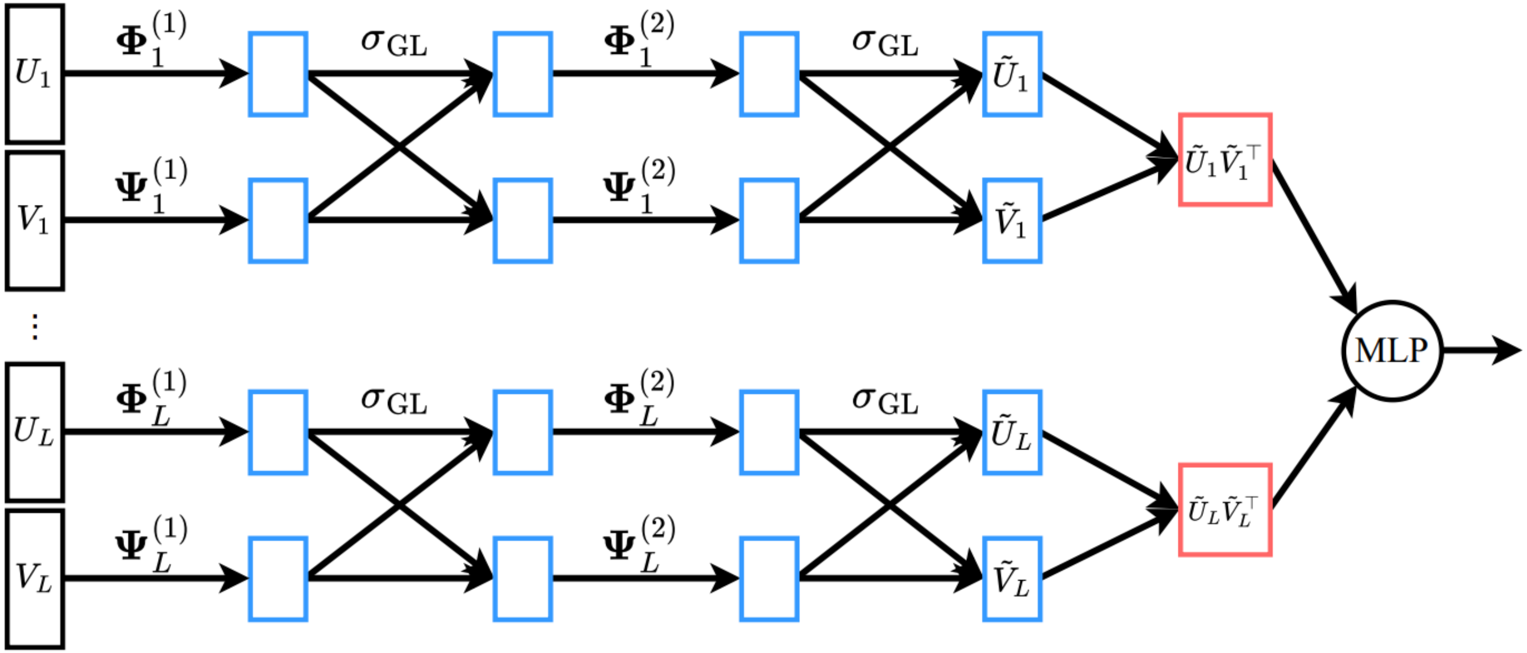

Pretrained Model
θ_{base}
W_1
⋮
W_L
LoRA
θ_{FT}
U_1
$V_1^\top$
⋮
U_L
$V_L^\top$
Learning on LoRAs (LoL Model)
$f_\theta\left(U_1, V_1, \ldots, U_L, V_L\right)$
Prediction
$\text{Acc}(\theta_{\text{FT}}) = 78.2\%$
or
Data source 3 is in the finetuning set
or
Model was finetuned on n data points

Putterman, Lim, Gelberg et B 2025



Horwitz et al. 2025

Learning on LoRAs (LOL)



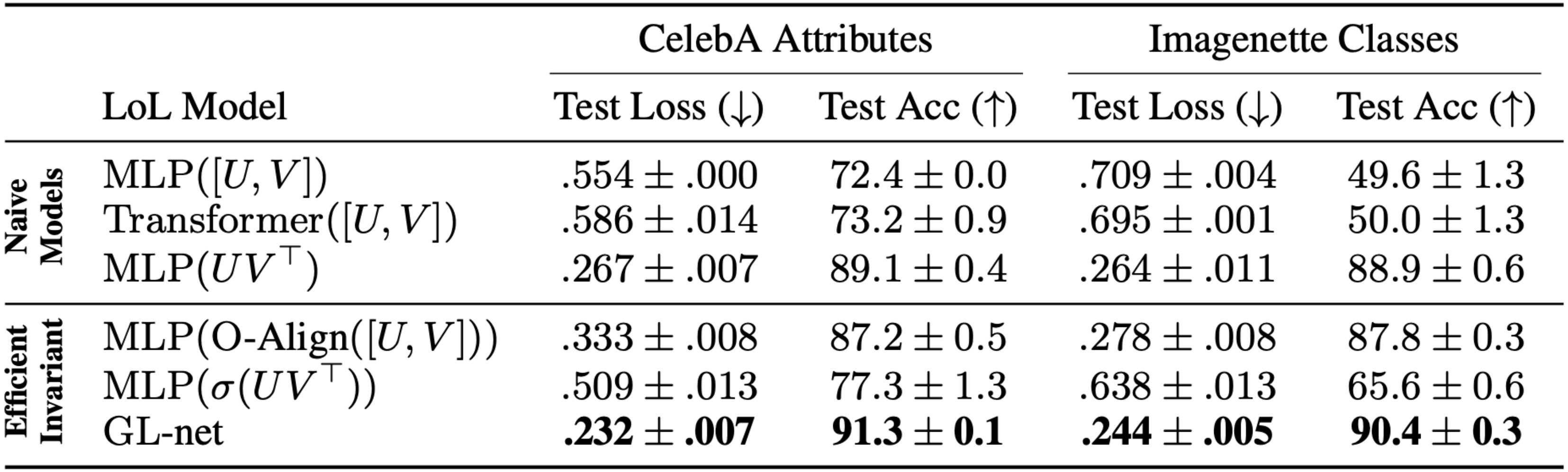
Putterman, Lim, Gelberg et B 2025

Learning on LoRAs (LOL)

LoL Model	GL-Inv.	O-Inv.	Expressive	Preprocess Time	Forward Time
Naive architectures					
MLP($[U, V]$)	✗	✗	✓	$O((m+n)r)$	$O((m+n)r)$
Transformer($[U, V]$)	✗	✗	✓	$O((m+n)r)$	$O((m+n)r)$
MLP($UV^\top$)	✓	✓	✓	$\boldsymbol{O(mnr)}$	$\boldsymbol{O(mn)}^2$
Efficient symmetry-aware architectures					
MLP(O-Align($[U, V]$))	✗	✓	✓	$O((m+n)r^2)$	$O((m+n)r)$
MLP($\sigma(UV^\top)$)	✓	✓	✗	$O((m+n)r^2)$	$O((m+n)r)$
GL-net	✓	✓	✓	$O((m+n)r)$	$O((m+n)r)$

Putterman, Lim, Gelberg et B 2025

Learning on LoRAs (LOL)

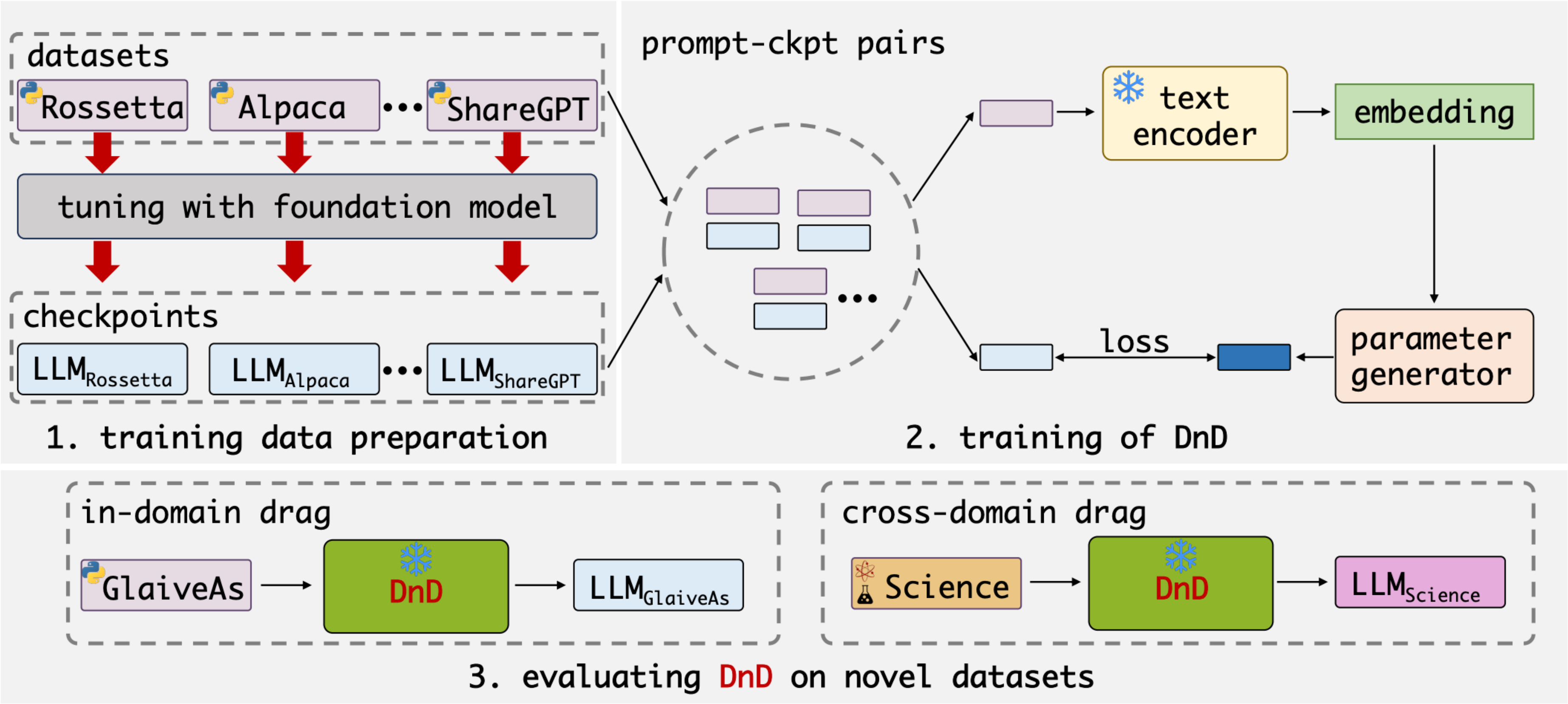


	LoL Model	CelebA Attributes		Imagenette Classes	
		Test Loss (↓)	Test Acc (↑)	Test Loss (↓)	Test Acc (↑)
Naive Models	$\mathrm{MLP}([U, V])$	$.554 \pm .000$	72.4 ± 0.0	$.709 \pm .004$	49.6 ± 1.3
	$\mathrm{Transformer}([U, V])$	$.586 \pm .014$	73.2 ± 0.9	$.695 \pm .001$	50.0 ± 1.3
	$\mathrm{MLP}(UV^\top)$	$.267 \pm .007$	89.1 ± 0.4	$.264 \pm .011$	88.9 ± 0.6
Efficient Invariant	$\mathrm{MLP}(\text{O-Align}([U, V]))$	$.333 \pm .008$	87.2 ± 0.5	$.278 \pm .008$	87.8 ± 0.3
	$\mathrm{MLP}(\sigma(UV^\top))$	$.509 \pm .013$	77.3 ± 1.3	$.638 \pm .013$	65.6 ± 0.6
	GL-net	$\mathbf{.232} \pm \mathbf{.007}$	$\mathbf{91.3} \pm \mathbf{0.1}$	$\mathbf{.244} \pm \mathbf{.005}$	$\mathbf{90.4} \pm \mathbf{0.3}$

Using LoL models to predict *CelebA* attributes (left) and *Imagenette* classes (right) of the finetuning data of diffusion models, given only the LoRA weights.

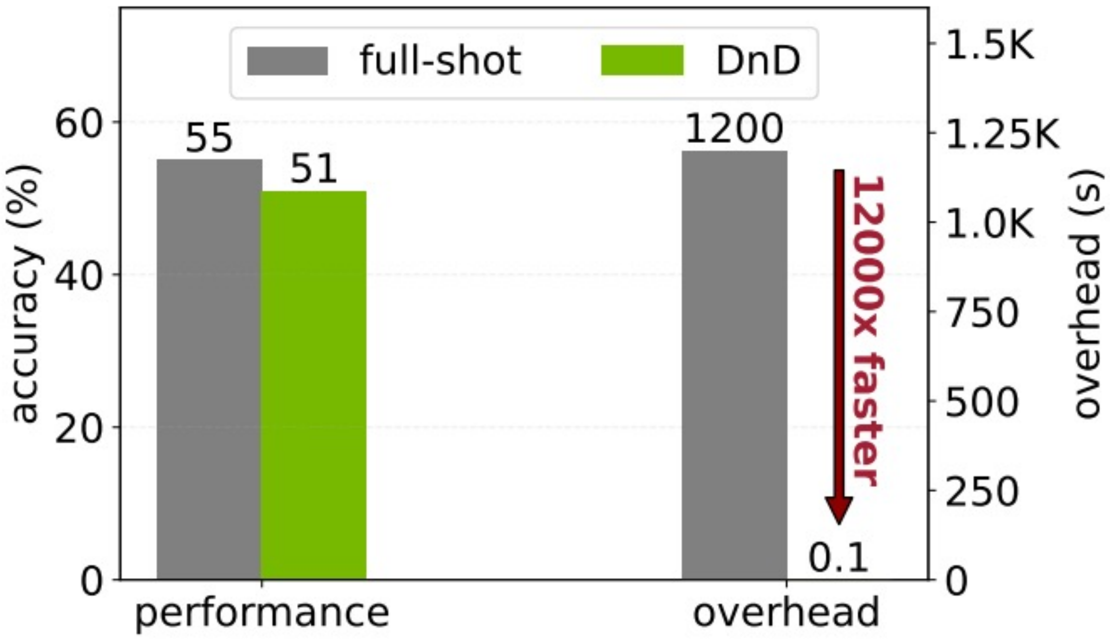
Putterman, Lim, Gelberg et B 2025

Drag-and-Drop LLMs


datasets
Rossetta
Alpaca
ShareGPT
tuning with foundation model
checkpoints
$\mathrm{LLM}_{\mathrm{Rossetta}}$
$\mathrm{LLM}_{\mathrm{Alpaca}}$
$\mathrm{LLM}_{\mathrm{ShareGPT}}$
1. training data preparation
prompt-ckpt pairs
text encoder
embedding
parameter generator
loss
2. training of DnD
in-domain drag
GlaiveAs
DnD
$\mathrm{LLM}_{\mathrm{GlaiveAs}}$
cross-domain drag
Science
DnD
$\mathrm{LLM}_{\mathrm{Science}}$
3. evaluating DnD on novel datasets

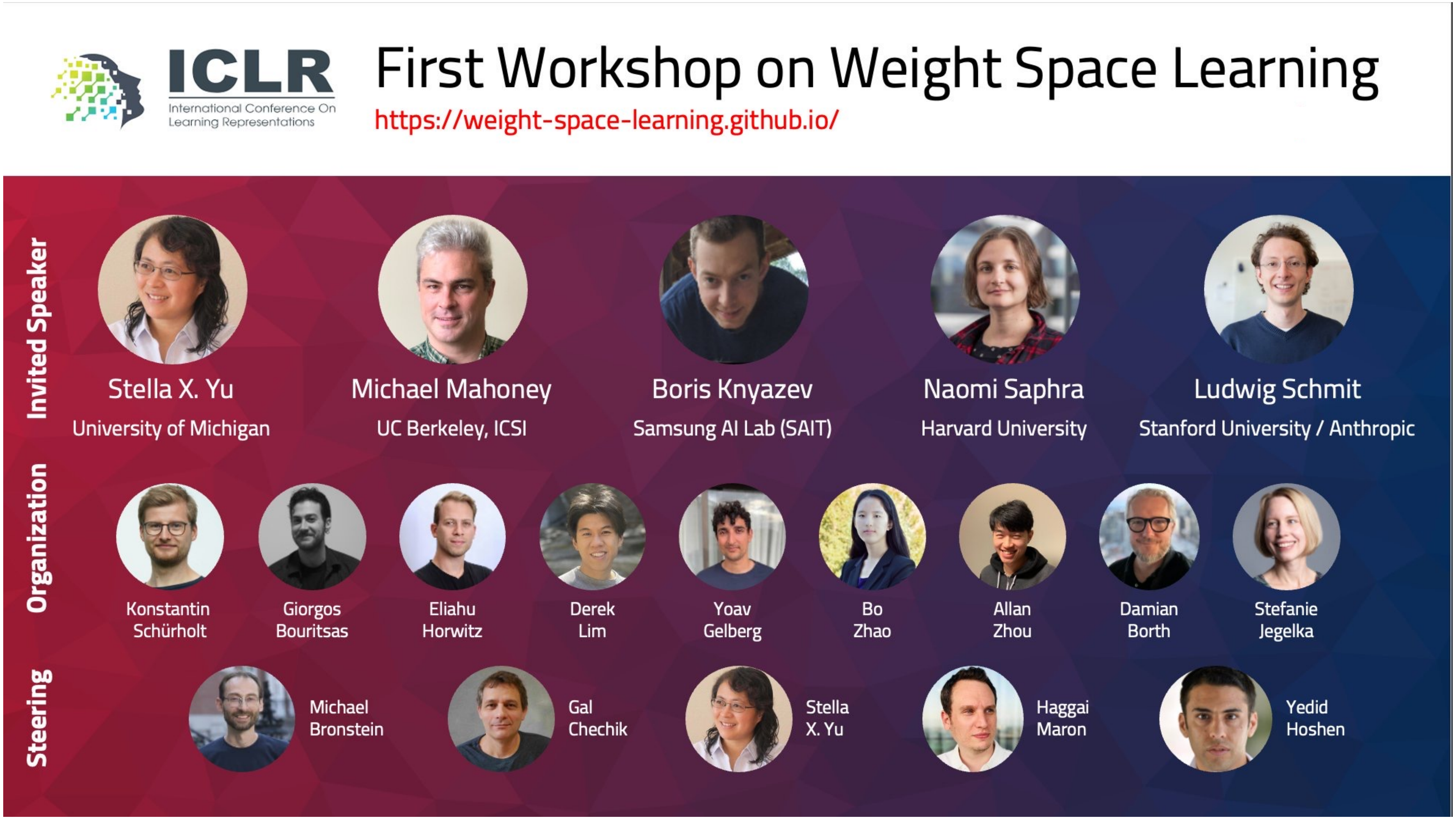
Liang, Tang, Zhou et B 2025

Drag-and-Drop LLMs



Liang, Tang, Zhou et B 2025

THE STORY CONTINUES IN THE WEIGHT SPACE



ICLR International Conference On Learning Representations

First Workshop on Weight Space Learning

https://weight-space-learning.github.io/

Invited Speaker

Stella X. Yu	Michael Mahoney	Boris Knyazev	Naomi Saphra	Ludwig Schmit
University of Michigan	UC Berkeley, ICSI	Samsung AI Lab (SAIT)	Harvard University	Stanford University / Anthropic

Organization

Konstantin Schürholt

Giorgos Bouritsas

Eliahu Horwitz

Derek Lim

Yoav Gelberg

Bo Zhao

Allan Zhou

Damian Borth

Stefanie Jegelka

Steering

Michael Bronstein

Gal Chechik

Stella X. Yu

Haggai Maron

Yedid Hoshen

Thank you!