The Machine Learning of Time & Dynamics

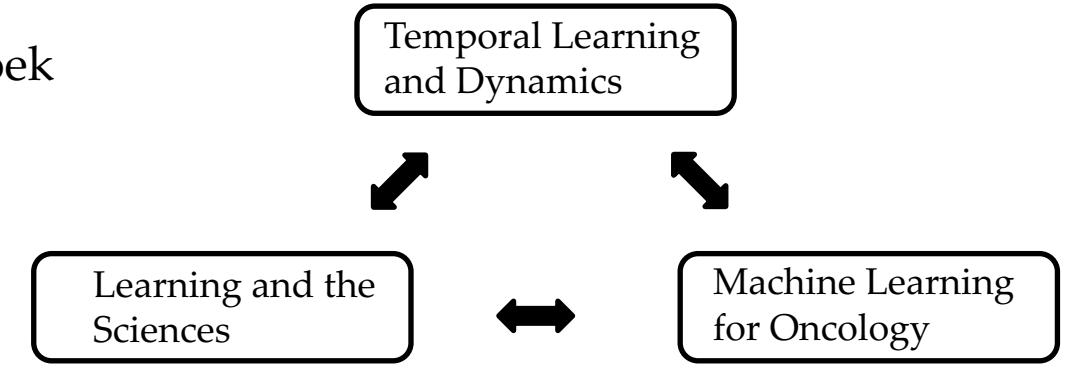
…with an outlook to the Sciences

Efstratios Gavves
Associate Professor

University of Amsterdam

Who am I?

- Associate Professor at the University of Amsterdam
  - ERC Starting Grant EVA (hiring!)
  - NWO VIDI TIMING
  - QUVA Lab w. Qualcomm + M. Welling +C.G.M. Snoek
  - POP-AART Lab w. Elekta & NKI +J.J. Sonke
- Teaching Deep Learning (http://uvadlc.github.io/)
- Scholar at ELLIS Network of Excellence in AI



- Co-founder of Ellogon.AI to personalize immunotherapy in oncology

OUTLINE

- Role of Time in machine learning and vision, key challenges, and ways forward

- First steps towards Temporal Learning & Dynamics
  - Learning dynamics in deep stochastic models
  - Equivariances in dynamical systems
  - Causal structure discovery

- Redefining spatiotemporal processes with an eye to the sciences (?)

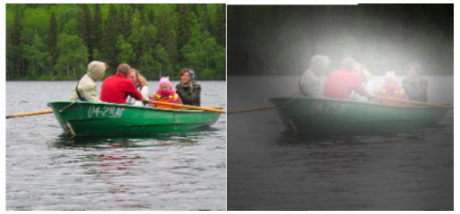
Time – Learning - Vision

The Golden Age of Learning Algorithms

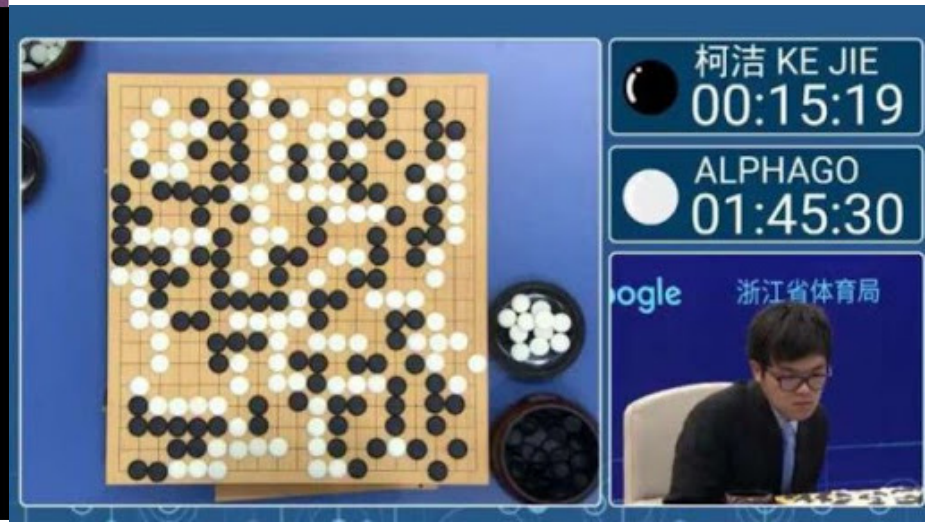
A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



[1] Ghodrati, Gavves, Snoek, Video Time: Encoders, Properties, Evaluation, BMVC, 2018
[2] Zhou, Andonian, Oliva, Torralba, Temporal Relational Reasoning in Videos, ECCV, 2018

An Urgent Paradox

Recognize apple falling

- Whether videos reversed, shuffled or normal, no difference[1,2]

Normal video: 83.1%

Reversed frames: 82.9%

Urgent for forecasting

State-of-the-art spatiotemporal models ignore time

[1] Ghodrati, Gavves, Snoek, Video Time: Encoders, Properties, Evaluation, BMVC, 2018
[2] Zhou, Andonian, Oliva, Torralba, Temporal Relational Reasoning in Videos, ECCV, 2018

An Urgent Paradox

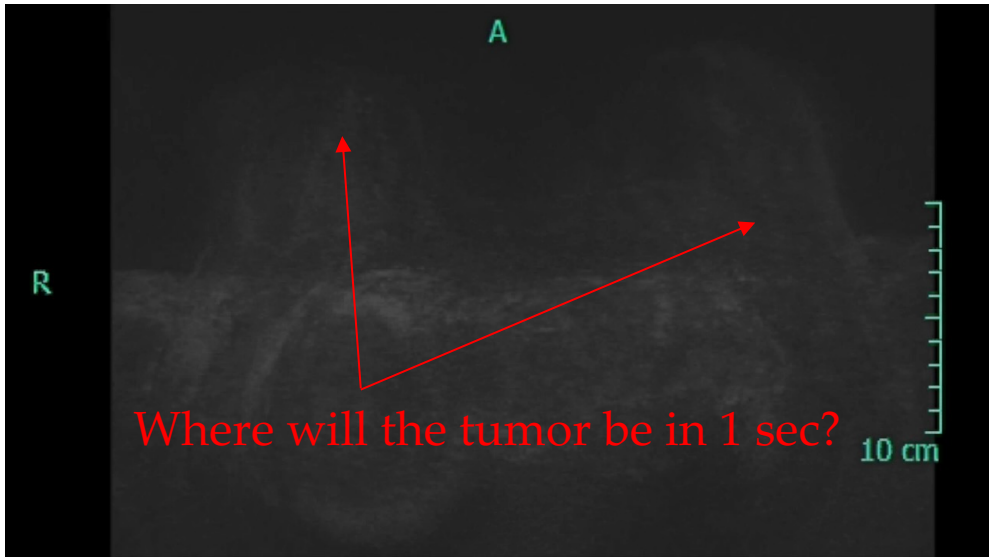
Recognize apple falling

- Whether videos reversed, shuffled or normal, no difference[1,2]

Normal video: 83.1%

Reversed frames: 82.9%

Urgent for future planning



State-of-the-art spatiotemporal models ignore time

[1] Ghodrati, Gavves, Snoek, Video Time: Encoders, Properties, Evaluation, BMVC, 2018
[2] Zhou, Andonian, Oliva, Torralba, Temporal Relational Reasoning in Videos, ECCV, 2018

An Urgent Paradox

Recognize apple falling

- Whether videos reversed, shuffled or normal, no difference[1,2]

Normal video: 83.1%

Reversed frames: 82.9%

Urgent for autonomous driving

State-of-the-art spatiotemporal models ignore time

Central Question

What is the role of time in recognition and learning?

TIME IN THE LITERATURE

HAMILTONIAN GENERATIVE NETWORKS

Peter Toth* DeepMind petertoth@google.com
Danilo J. Rezende* DeepMind danilor@google.com
Andrew Jaegle DeepMind drewjaegle@google.com
Sébastien Racanière DeepMind sracaniere@google.com
Aleksandar Botev DeepMind botev@google.com
Irina Higgins DeepMind irinah@google.com

ABSTRACT

The Hamiltonian formalism plays a central role in classical and quantum physics. Hamiltonians are the main tool for modelling the continuous time evolution of systems with conserved quantities, and they come equipped with many useful properties, like time reversibility and smooth interpolation in time. These properties are important for many machine learning problems —from sequence prediction to reinprovided out of the paper, we introduce h capable of consisobservations (such d, we can use HGN ackward in time and istrate how a simple ful normalising flow ltonian dynamics to rst practical demono deep learning.

Hamiltonian Neural Networks

Sam Greydanus Google Brain sgrey@google.com
Misko Dzamba PetCube mouse9911@gmail.com
Jason Yosinski Uber AI Labs yosinski@uber.com

FOURIER NEURAL OPERATOR FOR PARAMETRIC PARTIAL DIFFERENTIAL EQUATIONS

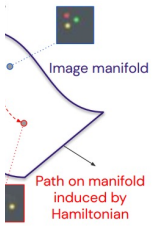
Zongyi Li zongyili@caltech.edu
Nikola Kovachki nkovachki@caltech.edu
Kamyar Azizzadenesheli kamyar@purdue.edu
Burigede Liu bgl@caltech.edu
Kaushik Bhattacharya bhatta@caltech.edu
Andrew Stuart astuart@caltech.edu
Anima Anandkumar anima@caltech.edu

ABSTRACT

The classical development of neural networks has primarily focused on learning mappings between finite-dimensional Euclidean spaces. Recently, this has been generalized to neural operators that learn mappings between function spaces. For partial differential equations (PDEs), neural operators directly learn the mapping from any functional parametric dependence to the solution. Thus, they learn an entire family of PDEs, in contrast to classical methods which solve one instance of the equation. In this work, we formulate a new neural operator by parameterizing the integral kernel directly in Fourier space, allowing for an expressive and efficient architecture. We perform experiments on Burgers' equation, Darcy flow, and Navier-Stokes equation. The Fourier neural operator is the first ML-based method to successfully model turbulent flows with zero-shot super-resolution. It is up to three orders of magnitude faster compared to traditional PDE solvers. Additionally, it achieves superior accuracy compared to previous learning-based solvers under fixed resolution.

1 INTRODUCTION

Many problems in science and engineering involve solving complex partial differential equation (PDE) systems repeatedly for different values of some parameters. Examples arise in molecular dynamics, micro-mechanics, and turbulent flows. Often such systems require fine discretization in order to capture the phenomenon being modeled. As a consequence, traditional numerical solvers



Video Transformer Network

Daniel Neimark Omri Bar Maya Zohar Dotan Asselmann
Theator
{danieln, omri, maya, dotan}@theator.io

Abstract

This paper presents VTN, a transformer-based framework for video recognition. Inspired by recent developments in vision transformers, we ditch the standard approach in video action recognition that relies on 3D ConvNets and introduce a method that classifies actions by attending to the entire video sequence information. Our approach is generic and builds on top of any given 2D spatial network. In terms of wall runtime, it trains 16.1× faster and runs 5.1× faster during inference while maintaining competitive accuracy compared to other state-of-the-art methods. It enables whole video analysis, via a single end-to-end pass, while requiring 1.5× fewer GFLOPs. We report competitive results on Kinetics-400 and Moments in Time benchmarks and present an ablation study of VTN properties and the trade-off between accuracy and inference speed. We hope our approach will serve as a new baseline and start a fresh line of research in the video recognition domain. Code and models are available at: https://github.com/bomri/SlowFast/blob/master/projects/vtn/README.md

1. Introduction

Attention matters. For almost a decade, ConvNets have ruled the computer vision field [21, 7]. Applying deep

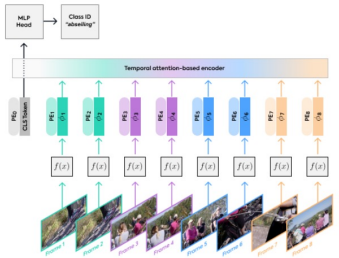
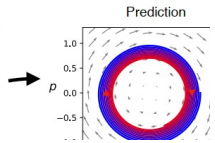


Figure 1. Video Transformer Network architecture. Connecting three modules: A 2D spatial backbone ($f(x)$), used for feature extraction. Followed by a temporal attention-based encoder (Longformer in this work), that uses the feature vectors (ϕ_i) combined with a position encoding. The $[CLS]$ token is processed by a classification MLP head to get the final class prediction.

mation later in the data flow by using attention mechanisms on top of the resulting features. Our approach input only RGB video frames and without any *bells and whistles (e.g.,* optical flow, streams lateral connections, multi-scale infer-

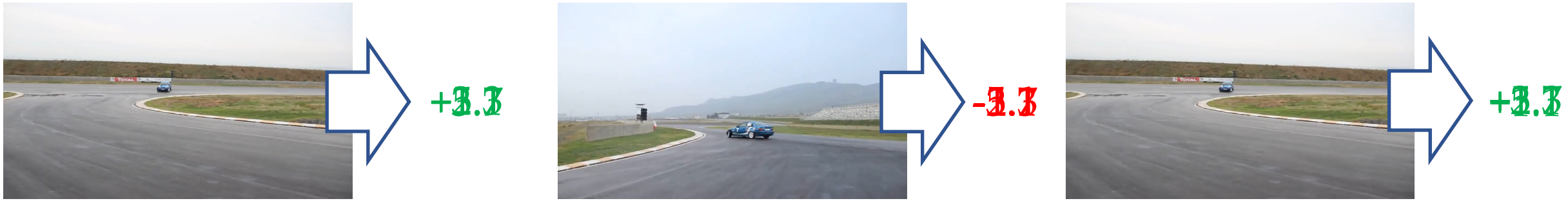


My Vision

Model that learn temporality in entangled space-time sequences

My Vision

Model that learn temporality in entangled space-time sequences



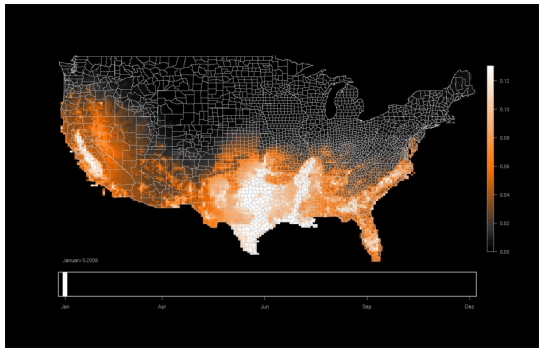
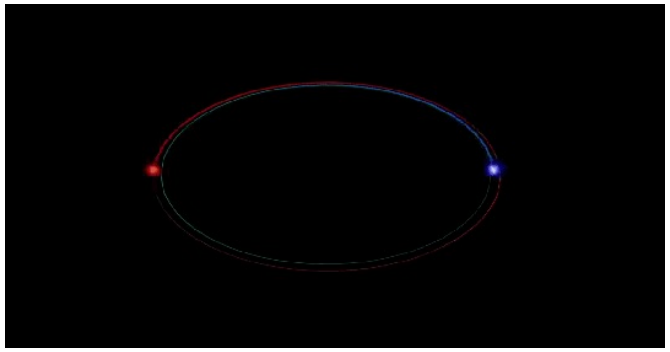
Entangled Spatiotemporal Data

- Any data with a spatial and temporal nature

(1) Thousands of spatio + temporal dimensions
(2) Confounding multi-scales in space & time

- Long & complex commercial videos, ecological sequences, geological sequences, astronomical sequences, particles, biomedical sequences, biochemical sequences…



Current Paradigm

Recognize bicycles in videos

Current Paradigm

Recognize bicycles in videos




Current Paradigm

Recognize bicycles in videos

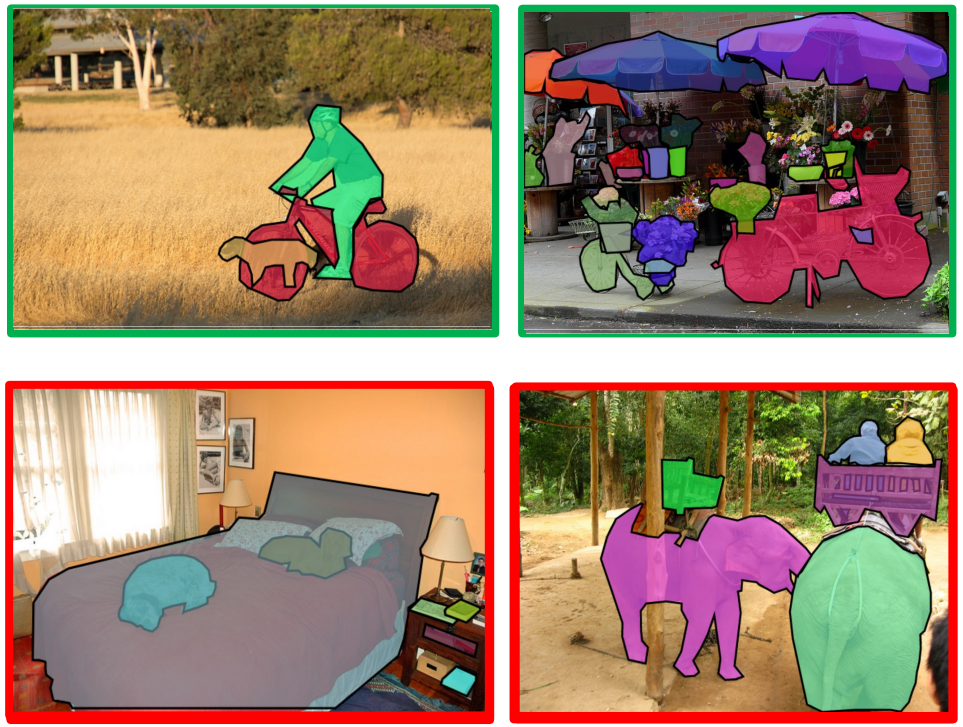
- Annotate bicycles in frames




Current Paradigm

Recognize bicycles in videos

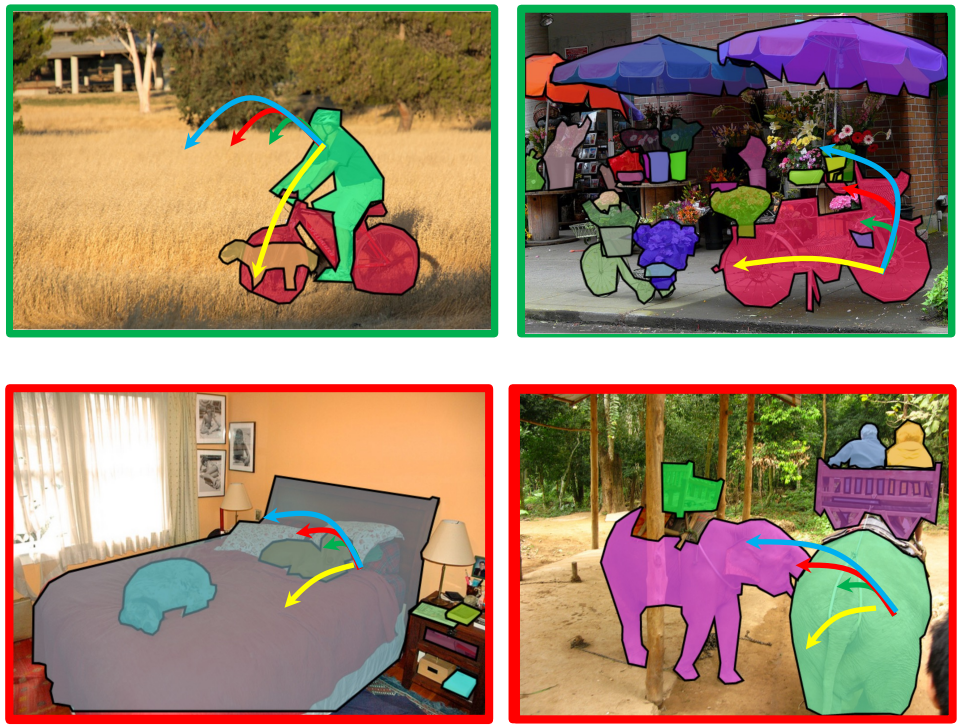
- Annotate bicycles in frames or pixels

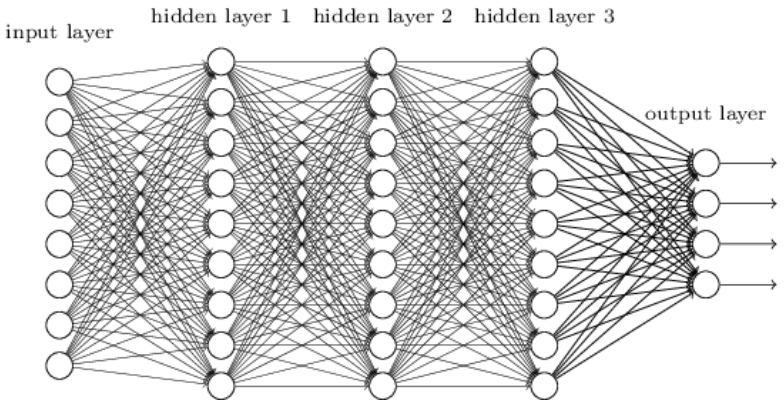


Current Paradigm

Recognize bicycles in videos

- Annotate bicycles in frames or pixels
- Neural net aggregates correlations locally

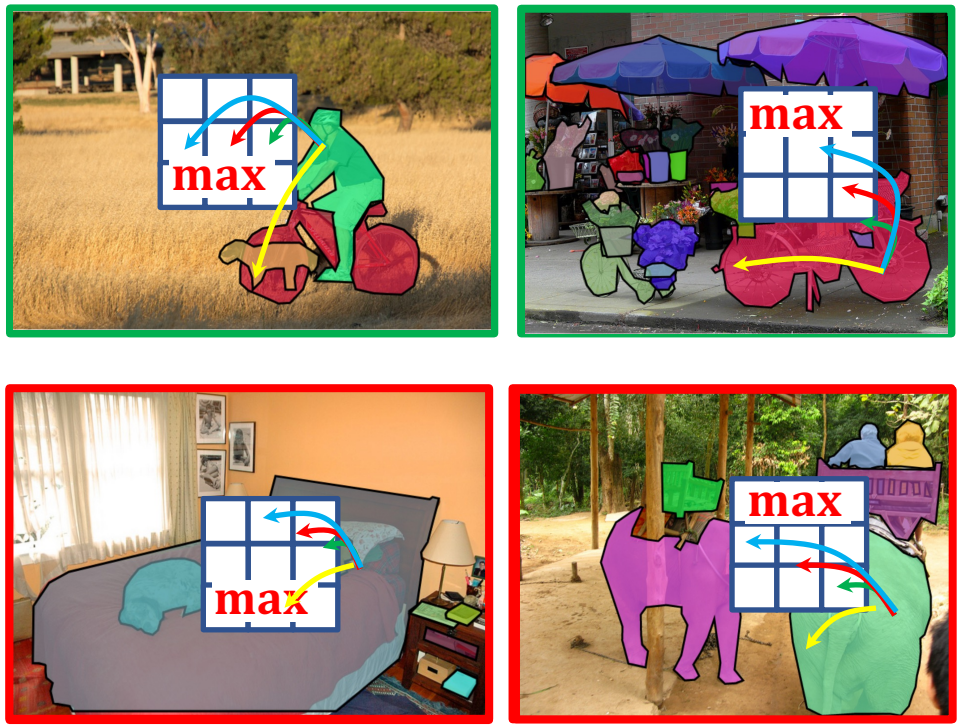


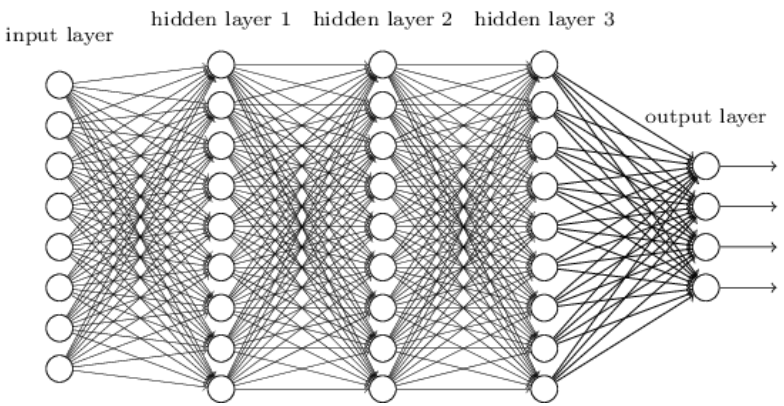


Current Paradigm

Recognize bicycles in videos

- Annotate bicycles in frames or pixels
- Neural net aggregates correlations locally with set operations

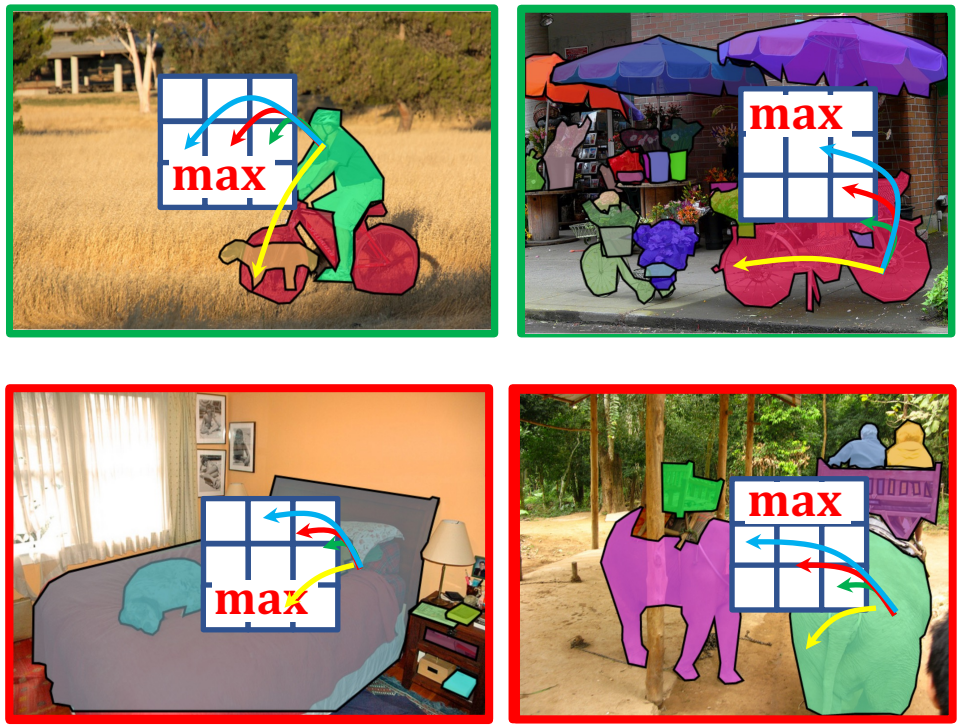


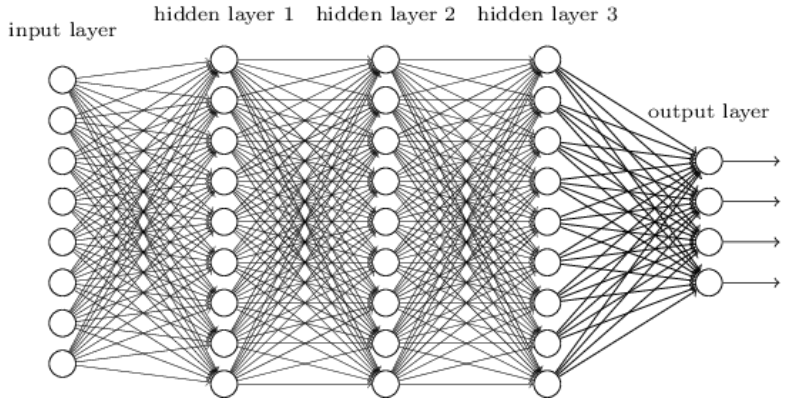


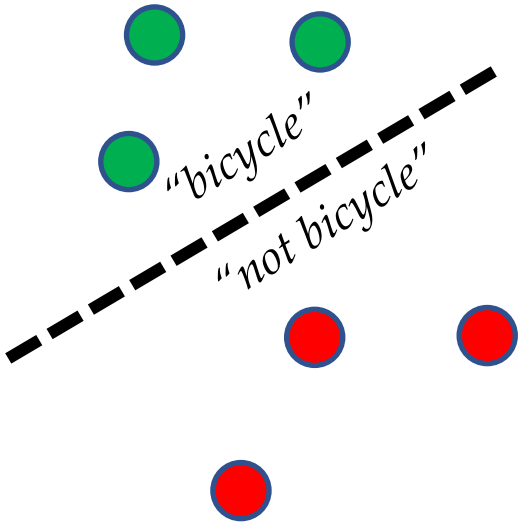
Current Paradigm

Recognize bicycles in videos

- Annotate bicycles in frames or pixels
- Neural net aggregates correlations locally with set operations
- To maximize separation

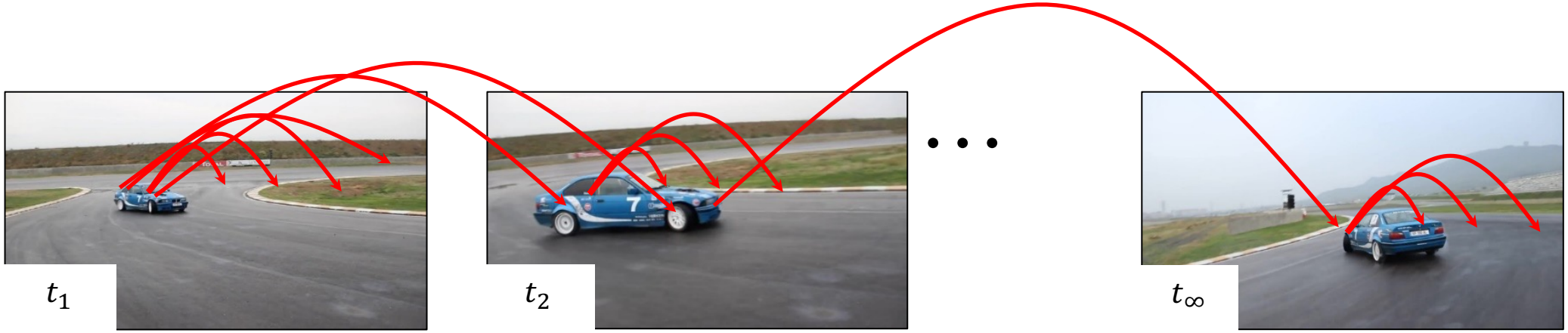






KEY CHALLENGE

- Thousands of frames even for moderate sequences
- → Innumerable correlations and dynamics

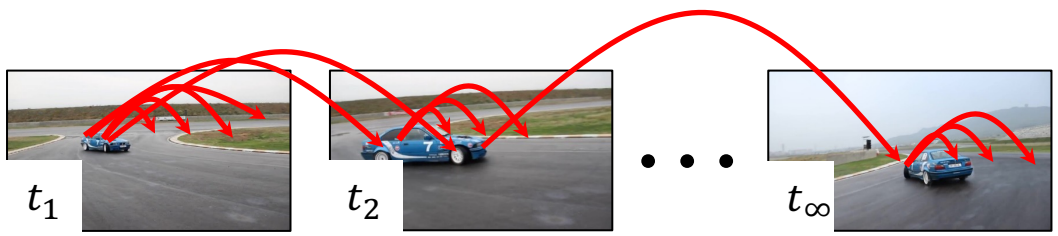


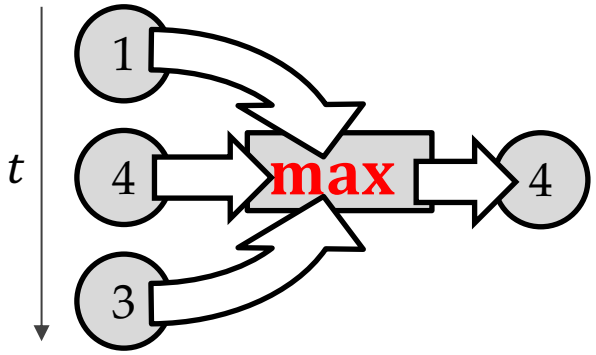
BREAKING DOWN THE KEY CHALLENGES

Challenge #0: Innumerable correlations and dynamics

Challenge #1: State-of-the-art discards time by aggregating with set operations

Challenge #2: Hard to annotate manually → supervised learning debatable → shorter videos





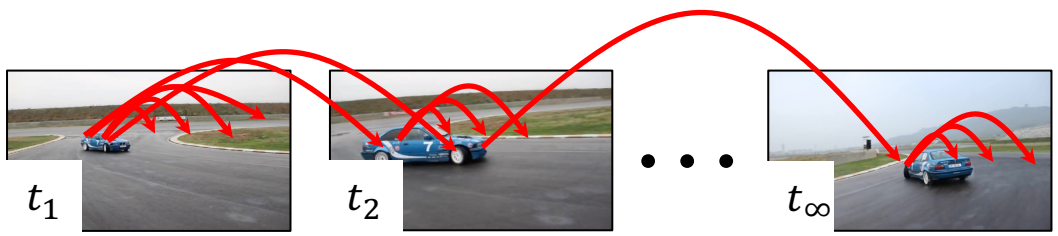
BREAKING DOWN THE KEY CHALLENGES

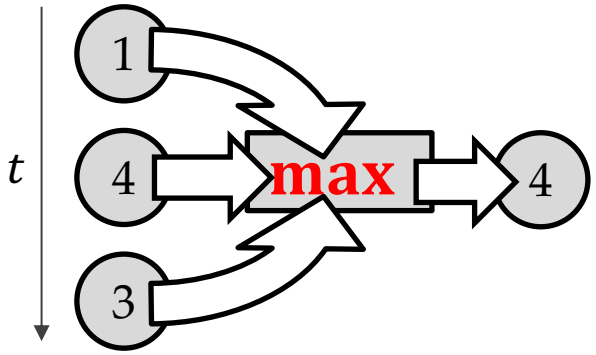
Challenge #0: Innumerable correlations and dynamics

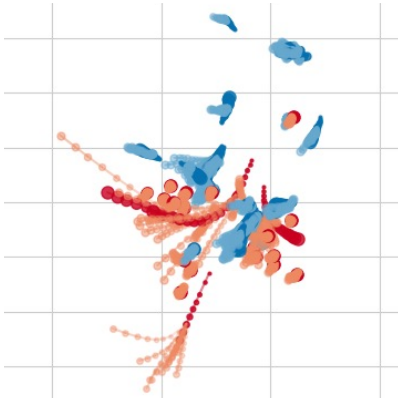
Challenge #1: State-of-the-art discards time by aggregating with set operations

Challenge #2: Hard to annotate manually → supervised learning debatable → shorter videos

Challenge #3: A sequence is one of myriad possibilities → generative modelling critical







BREAKING DOWN THE KEY CHALLENGES

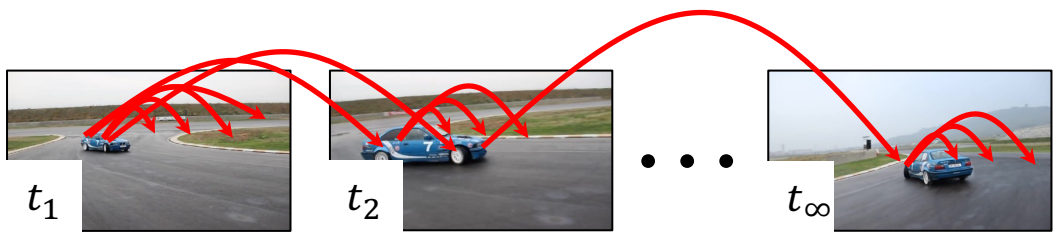
Challenge #0: Innumerable correlations and dynamics

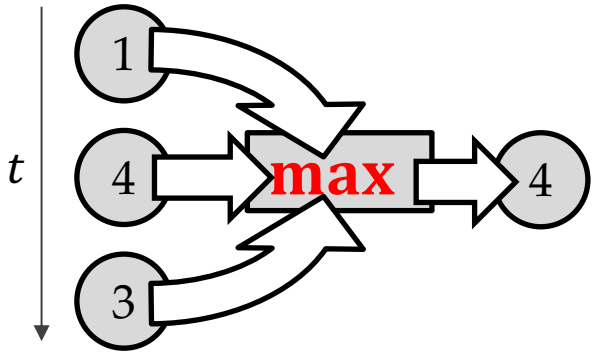
Challenge #1: State-of-the-art discards time by aggregating with set operations

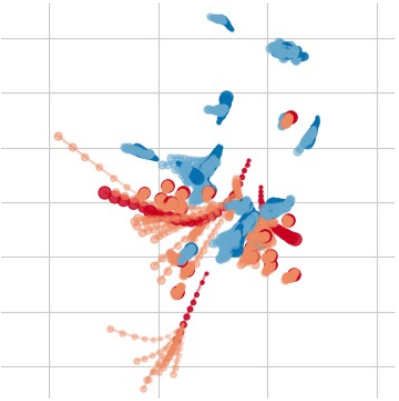
Challenge #2: Hard to annotate manually → supervised learning debatable → shorter videos

Challenge #3: A sequence is one of myriad possibilities → generative modelling critical

Challenge #4: Lack of standardization ← data is too huge & algorithms too complex



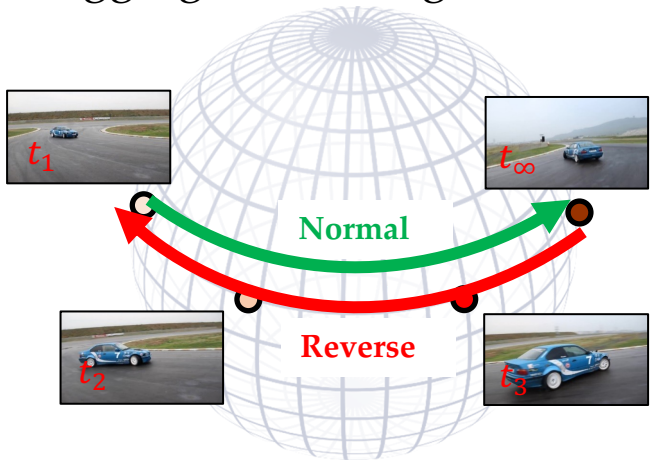




WAY FORWARD

Space-Time Geometry

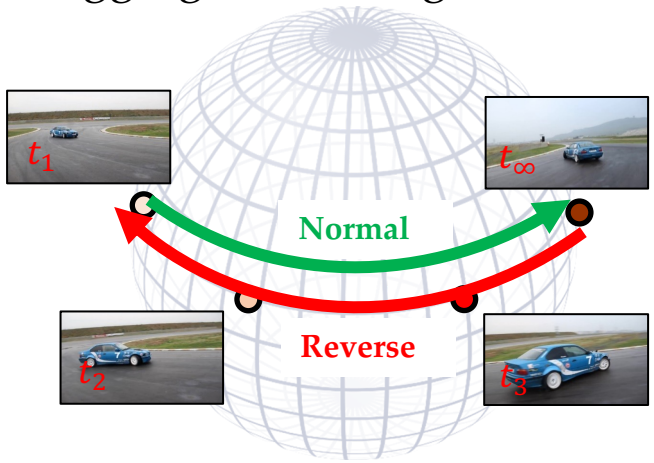
- Learn spatiotemporal geometric manifolds
- Aggregate over a geodesic time path on manifold



WAY FORWARD

Space-Time Geometry

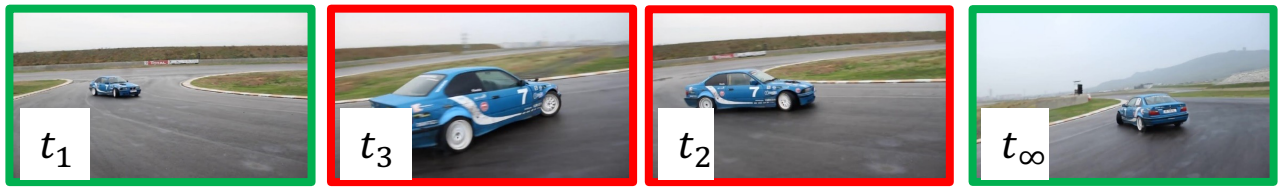
- Learn spatiotemporal geometric manifolds
- Aggregate over a geodesic time path on manifold



Space-Time Supervision

- *E.g.*, by causality, feature slowness, continuity, …
- Caveat: must combine with time-sensitive models

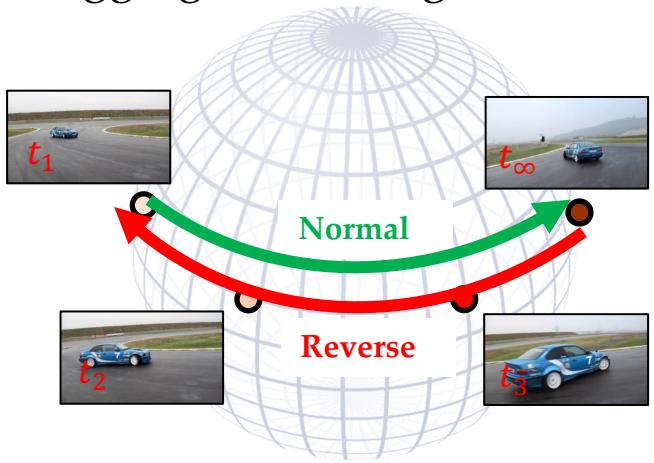
Are frames correctly ordered or causally logical?



WAY FORWARD

Space-Time Geometry

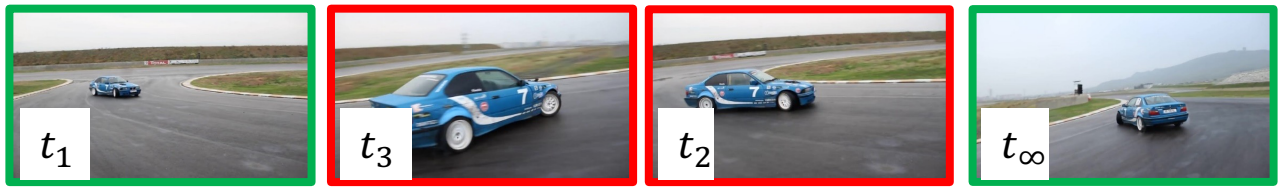
- Learn spatiotemporal geometric manifolds
- Aggregate over a geodesic time path on manifold



Space-Time Supervision

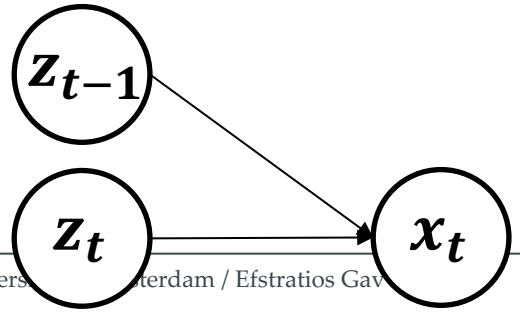
- *E.g.*, by causality, feature slowness, continuity, …
- Caveat: must combine with time-sensitive models

Are frames correctly ordered or causally logical?



Space-Time Stochastic

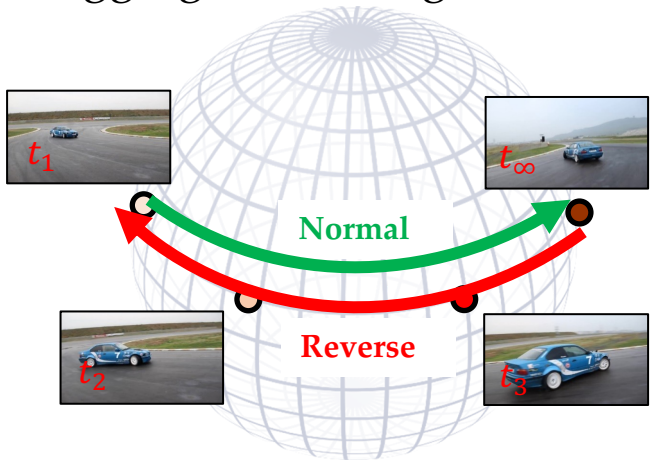
- Models that imagine all possible futures
- Spatiotemporal generative/bayesian models



WAY FORWARD

Space-Time Geometry

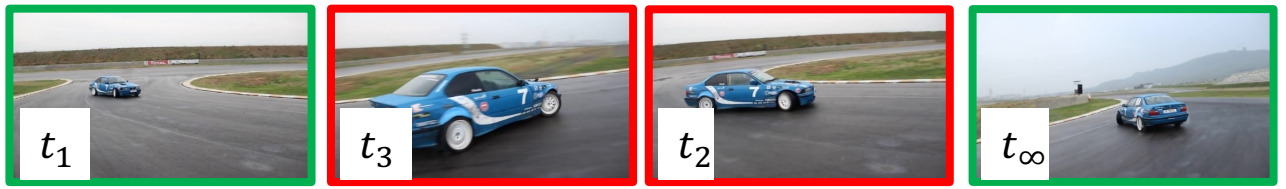
- Learn spatiotemporal geometric manifolds
- Aggregate over a geodesic time path on manifold



Space-Time Supervision

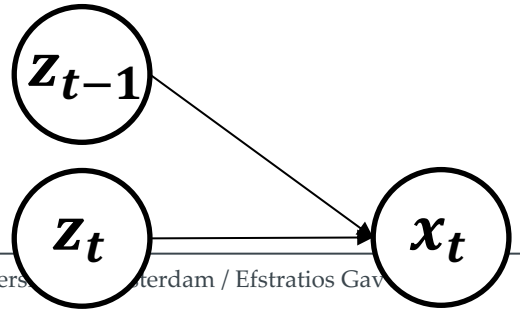
- *E.g.*, by causality, feature slowness, continuity, …
- Caveat: must combine with time-sensitive models

Are frames correctly ordered or causally logical?



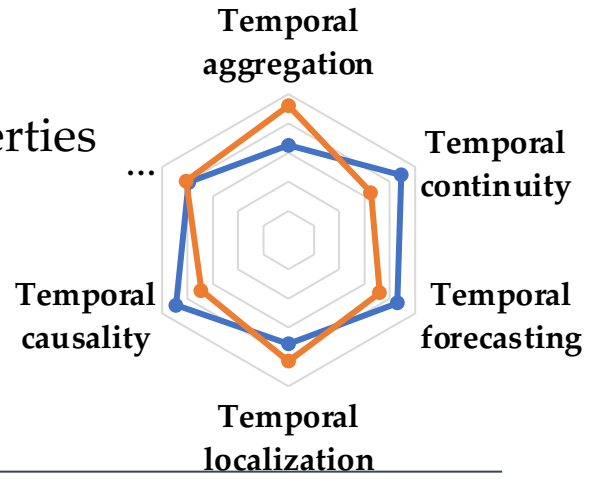
Space-Time Stochastic

- Models that imagine all possible futures
- Spatiotemporal generative/bayesian models



Space-Time Evaluation

- Standardize data
- Evaluate on temporal properties



Temporal Learning & Dynamics

First steps

Dynamics of Learning

Spectral smoothing unveils phase transitions in hierarchical variational autoencoders

ICML 2021

A. Pervez E. Gavves

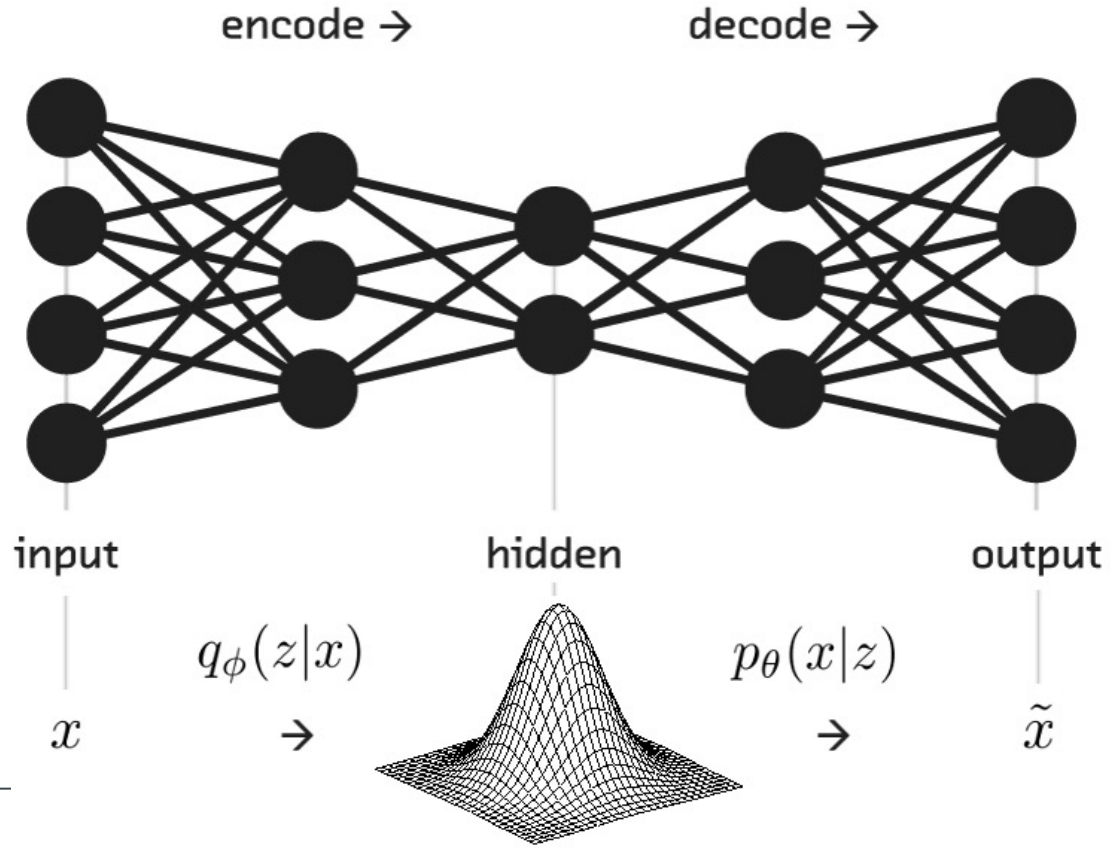
VARIATIONAL AUTOENCODERS

- Optimize ELBO

$$\mathbb{E}[\log p_\theta(x|z)] - \mathrm{KL}(q_\varphi(z|x) \parallel p(z))$$

Good' reconstructions

aggregate posterior
$\frac{1}{N}\sum_i q_\varphi(z|x_i)$ close to prior $p(z)$



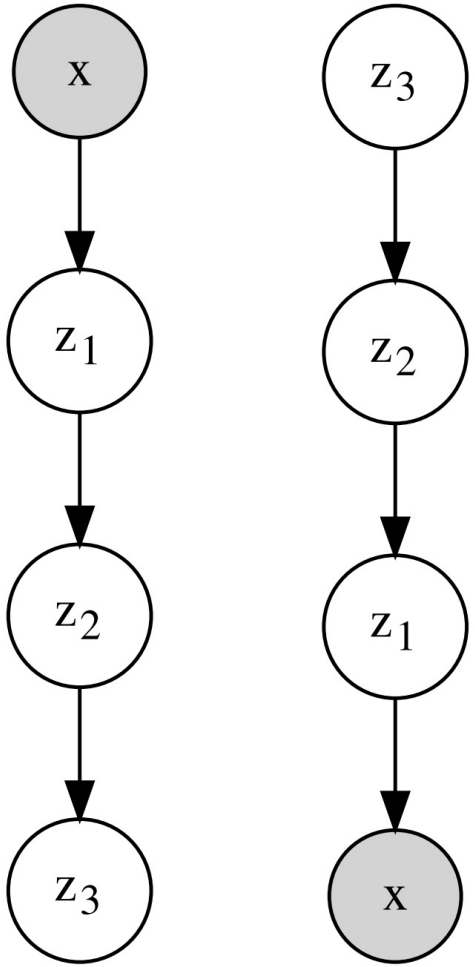
HIERARCHICAL VARIATIONAL AUTOENCODERS

- How to 'enhance' VAE representation capacity?
- Make them deep by stacking stochastic layers

$$z_1, z_2, \ldots$$

- Translates to adding multiple KL terms per layer

$$\mathrm{KL}(q(z_{i+1}|z_i) \parallel p(z_i|z_{i+1}))$$



Posterior collapse

- When stacking stochastic layers, often the individual posteriors fall back to the prior
  - Not just the aggregate posterior
- This phenomenon is called 'posterior collapse' (Bowman et al., 2016)
  - The respective dimensions are 'inactive'
  - Pathological behavior
- Posterior collapse is especially pronounced with hierarchical VAEs

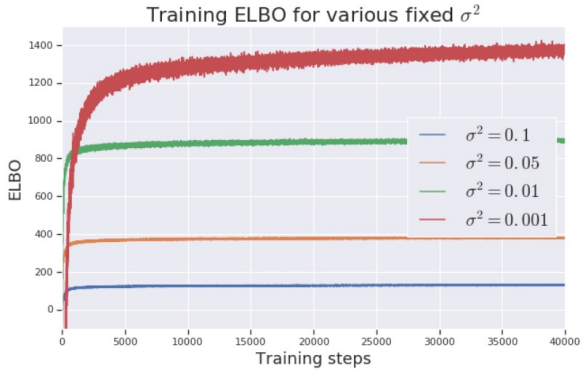


Figure 3: ELBO during training of MNIST VAEs with Gaussian observation model. A better ELBO is achieved with a smaller choice of σ^2.

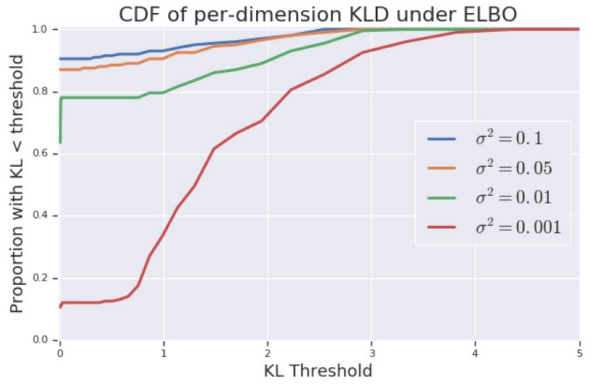


Figure 4: The proportion of inactive units in trained MNIST VAEs which, on average, are less than the specified threshold.

[1]

A. Lucas et al, Understanding posterior collapse in generative latent variable models, 2019

INTUITIVE HYPOTHESIS: VARIANCE CAUSES COLLAPSE

- Stacking stochastic variables increases variance exponentially
- In its inability to 'control' the stochasticity, the optimizer 'prefers' to switch off dimensions/layers
  - Especially, the farther these dimensions are from the output

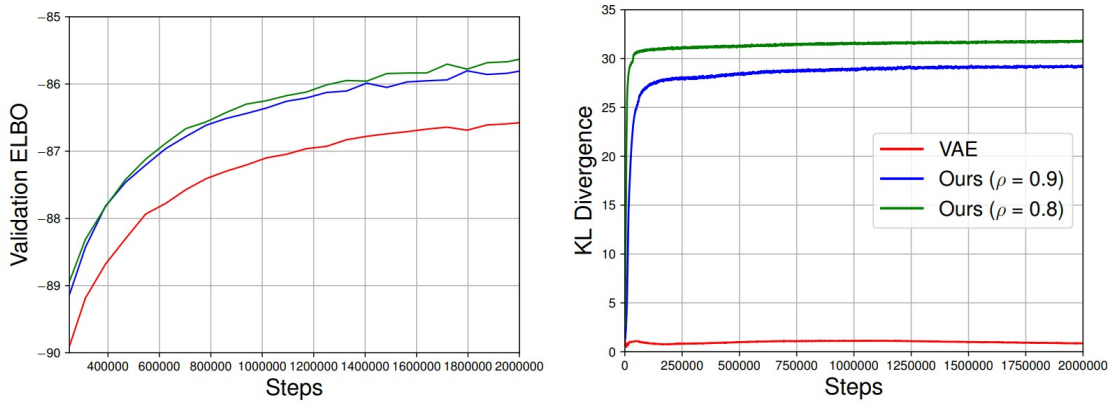


Figure 2. Training 4 layer VAE models on static MNIST with validation ELBO on the left and $KL(q(z_4|z_3)||\mathcal{N}(0,1))$ on the right. The VAE shows posterior collapse, while OU-VAE avoids it alongside improved validation ELBO.

ORNSTEIN-UHLENBECK SMOOTHING

- Ornstein-Uhlenbeck semigroup from Gaussian Analysis

$$U_\rho f(z) = \mathbb{E}_{z'|z}[f(z')] = \mathbb{E}_{z_1}\left[f\left(\rho z + \sqrt{1-\rho^2} z_1\right)\right]$$

Where z_1 is a Gaussian and $\rho \in (0,1)$ is a smoothing parameter

- Ornstein-Uhlenbeck semigroups

Ornstein-Uhlenbeck smoothing

- Ornstein-Uhlenbeck semigroup from Gaussian Analysis

$$U_\rho f(z) = \mathbb{E}_{z'|z}[f(z')] = \mathbb{E}_{z_1}\left[f\left(\rho z + \sqrt{1-\rho^2} z_1\right)\right]$$

Where z_1 is a Gaussian and $\rho \in (0,1)$ is a smoothing parameter

ORNSTEIN-UHLENBECK SMOOTHING: PROPERTIES

(1) Ornstein-Uhlenbeck semigroups preserve expectations

(2) Behaves nicely with Hermite expansion of Gaussian functions $f = \sum_{\alpha \in \mathbb{N}^n} \hat{f}(\alpha) h_\alpha$ in that it reduces higher order terms by some power of ρ

$$\mathrm{U}_\rho f = \sum_{\alpha \in \mathbb{N}^n} \rho^{|a|} \hat{f}(\alpha) h_\alpha$$

- U_ρ can be seen as operator that interpolates between $f(z)$ and $\mathbb{E}[f]$

ORNSTEIN-UHLENBECK VAE

- Replace the intermediate parameterized Gaussians

$$\mathcal{N}\left(z_i \middle| f_{\mu_{i+1}}(z_{i+1}), f_{\sigma_{i+1}}(z_{i+1})\right)$$

With smoothed ones

$$\mathcal{N}(z_i | U_\rho f_{\mu_{i+1}}(z_{i+1}), U_\rho f_{\sigma_{i+1}}(z_{i+1}))$$

- We can train the model with the smoothed version and evaluate without smoothing

- It can be shown there is a bias-variance trade-off
- We reduce the variance with $O(\rho^2)$ while causing bias with $O(1-\rho^2)$

Comparing active units

- VAEs with 4 stochastic layer
- Comparing with the latest methods against posterior collapse (+KL: KL-annealing)
- Measuring the KL divergence in the 'top' latent (further from output)

Model	V. ELBO	KLD	Top KLD	Active Units
IWAE (64-32-16-8)	-84.46	23.98	4.88	64-30-15-3
IWAE (40-40-40-40)	-84.63	23.98	1.18	40-37-4-0
VAE+KL (64-32-16-8)	-84.6	28.8	6.2	49-25-11-6
VAE+KL (40-40-40-40)	-84.7	28.07	1.13	40-15-6-1
VAE+Freebits (64-32-16-8)	-85.5	25.1	3.8	21-9-4-2
VAE+Freebits (40-40-40-40)	-86.0	23.6	2.46	18-8-2-1
OU-VAE ($\rho = 0.95$) (64-32-16-8)	-81.6	26.1	8.99	54-32-16-8
OU-VAE ($\rho = 0.9$) (64-32-16-8)	-81.7	25.6	9.56	43-32-16-8
OU-VAE ($\rho = 0.95$) (40-40-40-40)	-84.4	23.7	9.34	40-40-40-40

RESULTS

- Competitive results with very simple architectures
  - Only modification: adding OU smoothing
- Compared to SoTA models, at least an order of magnitude fewer parameters

Test ELBO on dynamic MNIST with MLP

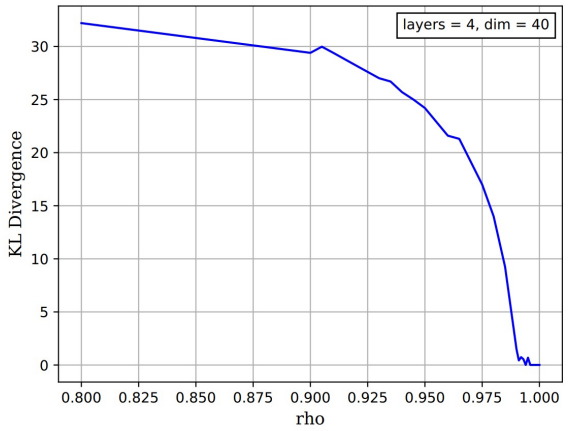
Model	ELBO
Ladder VAE (L=5)[4]	-81.7
VampPrior (L=2)[5]	-81.24
OU-VAE (L=4), $\mathcal{L}_{5000}$	-81.2
OU-VAE (L=5), $\mathcal{L}_{5000}$	-81.1

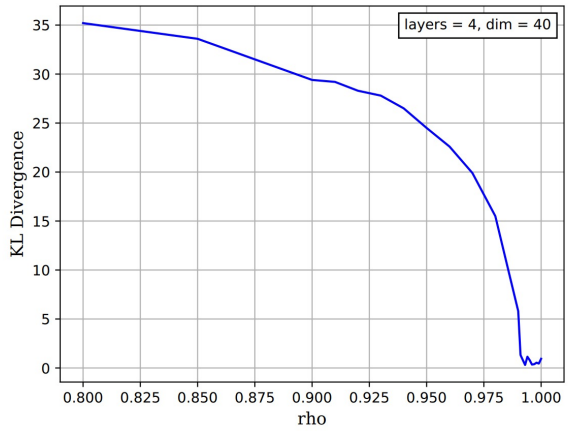
Table 4. Comparing bits per dimension, parameters and depths on CIFAR-10. 'OU-VAE+' include stochastic skip connections.

Model	BPD	Layers	Parameters
Vanilla feedforward networks, residual connections			
OU-VAE, $\mathcal{L}_1$	3.5	3	9.95M
OU-VAE, $\mathcal{L}_1$	3.46	4	12.4M
OU-VAE, $\mathcal{L}_1$	3.43	6	16.8M
OU-VAE+, $\mathcal{L}_1$	3.42	3	9.95M
OU-VAE+, $\mathcal{L}_{100}$	3.39	3	9.95M
Feedforward networks, residual connections, shared weights between encoder and decoder			
LVAE (Maaløe et al., 2019)	3.60	15	72.36M
LVAE+ (Maaløe et al., 2019)	3.41	15	73.35M
LVAE+ (Maaløe et al., 2019)	3.45	29	119.71M
BIVA (Maaløe et al., 2019)	3.12	15	102.95M
VAE+IAF (Kingma et al., 2016)	3.11	–	–
NVAE (Vahdat & Kautz)	2.91	–	–
Feedforward networks, residual connections, normalizing flow prior/autoregressive			
Disc. VAE++ (Vahdat et al., 2018)	3.38	–	–
NICE (Dinh et al., 2014)	4.48	–	–
RealNVP (Dinh et al., 2017)	3.49	–	–

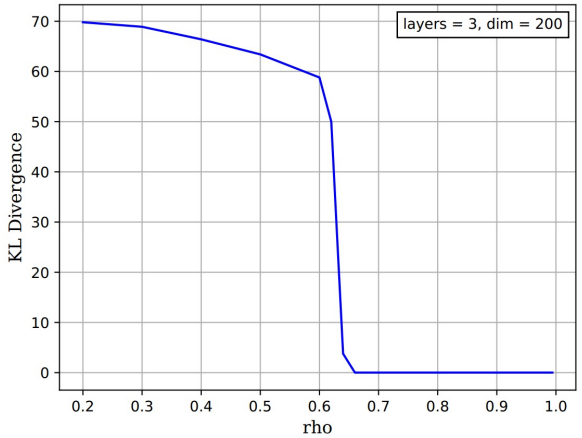
PHASE TRANSITIONS ARISE

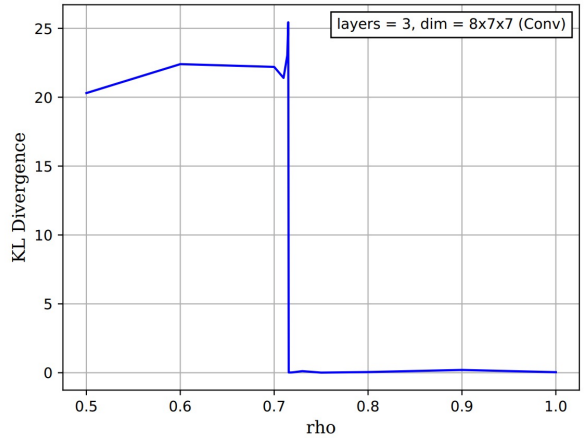
- When playing with ρ, we discovered very sudden transitions
- Even more attenuated when increasing 'architectural complexity'





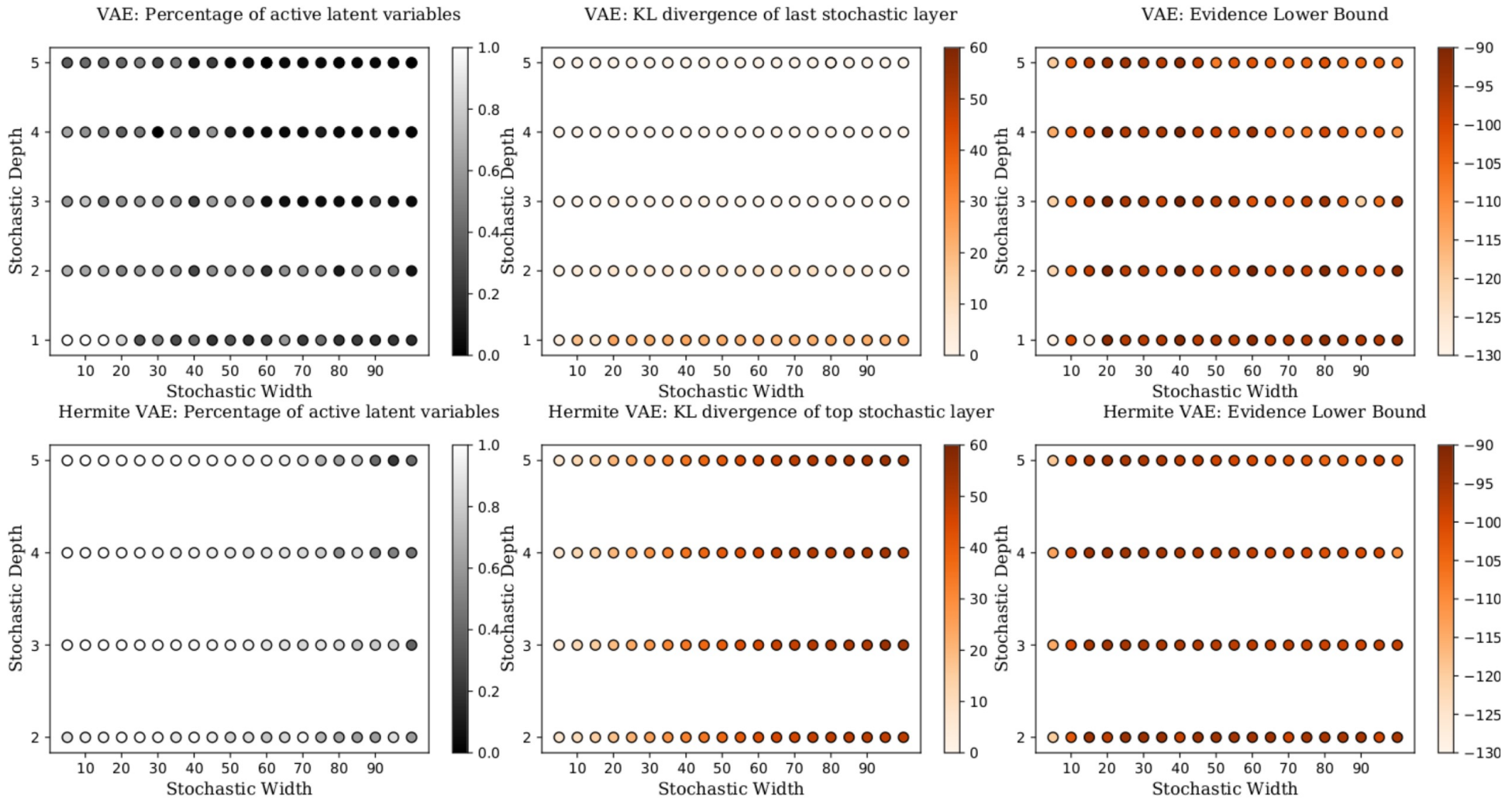
Top layer KLD v. ρ for a 4 layer VAE on MNIST (left) and OMNIGLOT (right





Increasing the latent dimensions: a 3-layer 200 unit MLP (left) and a 3-layer conv VAE (right)

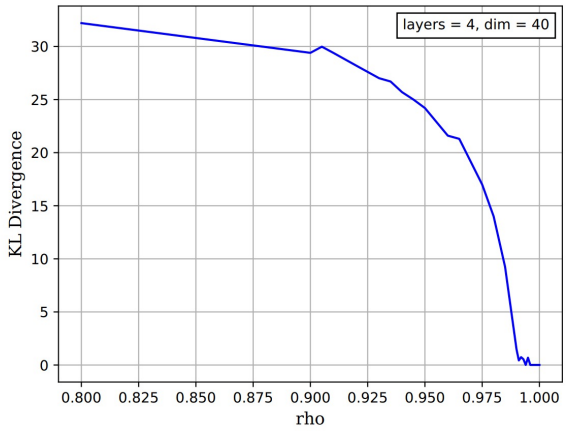
Phase transitions and posterior collapse

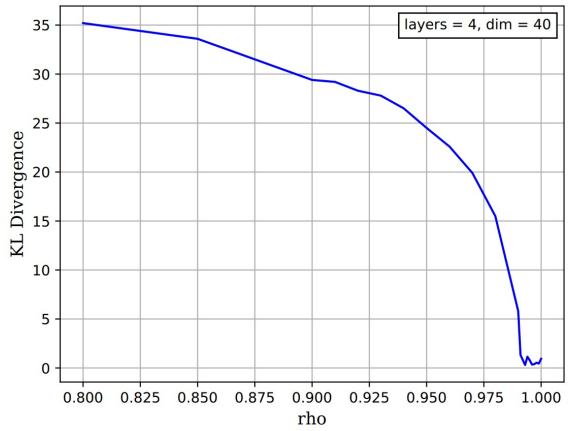


PHASE TRANSITIONS ARISE

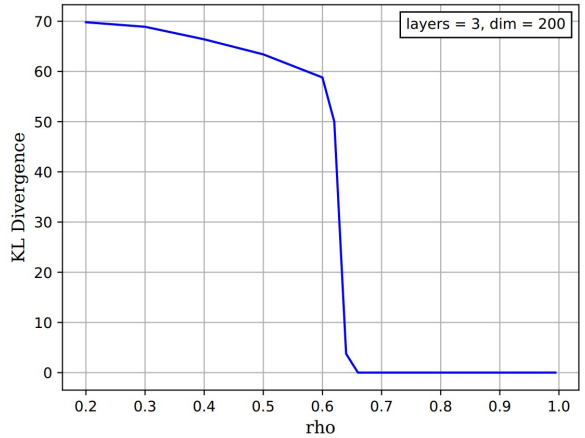
- When playing with ρ, we discovered very sudden transitions
- Even more attenuated when increasing 'architectural complexity'

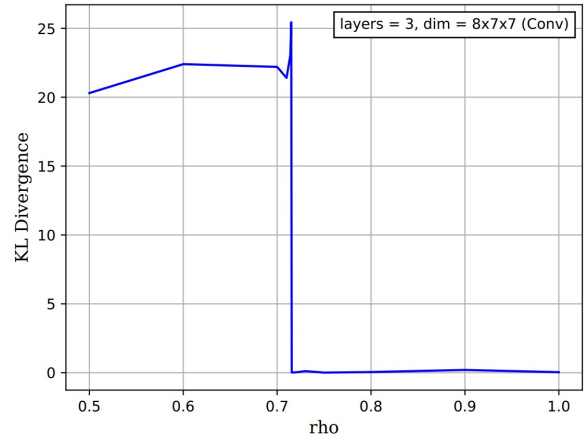
- Why is this happening? Not answered yet
- Phase transitions link to a statistical mechanics perspective?
- Is there a connection with what makes a system 'learnable'





Top layer KLD v. ρ for a 4 layer VAE on MNIST (left) and OMNIGLOT (right





Increasing the latent dimensions: a 3-layer 200 unit MLP (left) and a 3-layer conv VAE (right)

Code (soon) available

Causal Structure Discovery

Efficient Neural Causal Discovery without Acyclicity Constraint

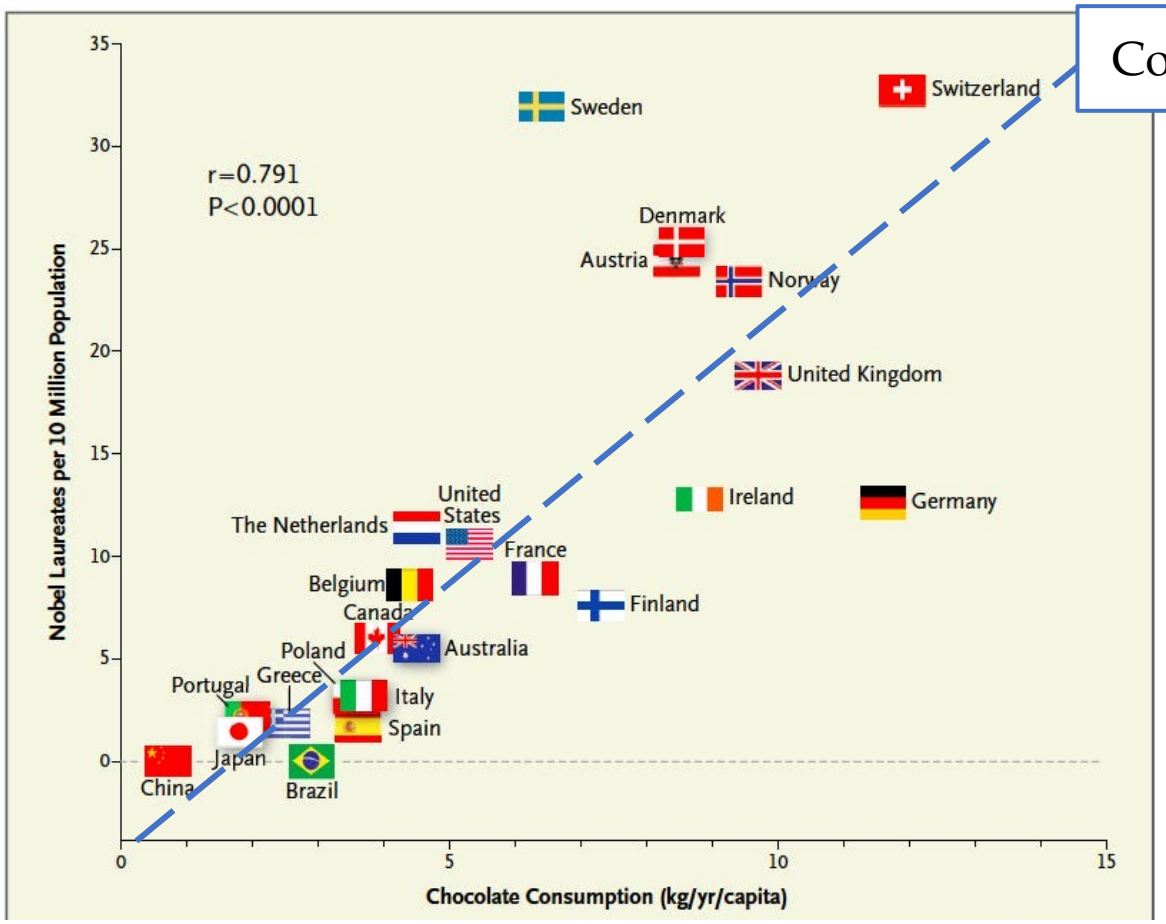
In submission

P. Lippe

T. Cohen

E. Gavves

Correlation vs Causation



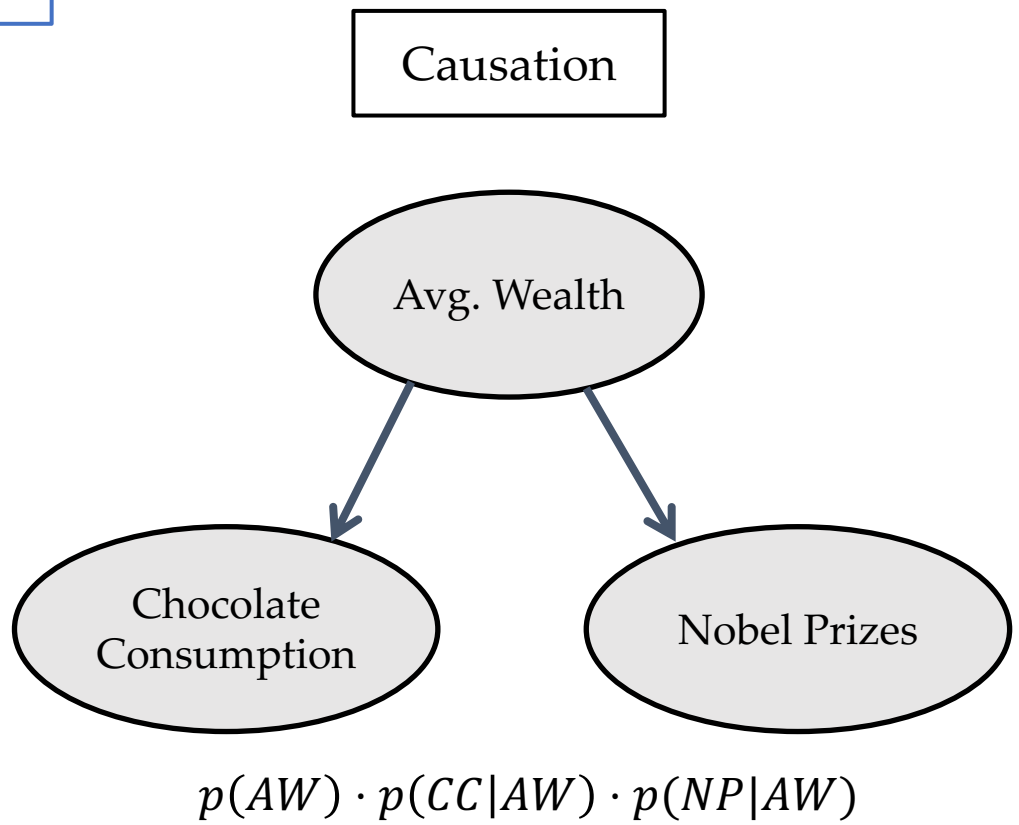


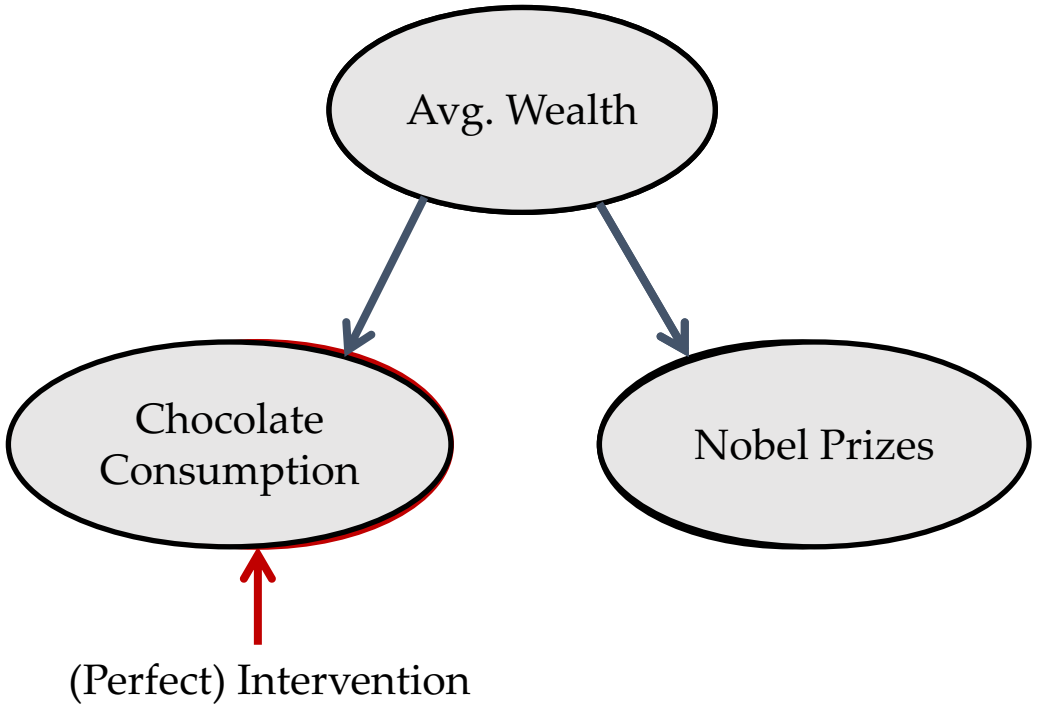
Figure credit: Franz H. Messerli, 2012. Chocolate Consumption, Cognitive Function, and Nobel Laureates

Background: Causal Discovery

- Causal Discovery: Find causal relations from data
- Intervention: changing variable distribution
- Causal Discovery from interventional data on all variables

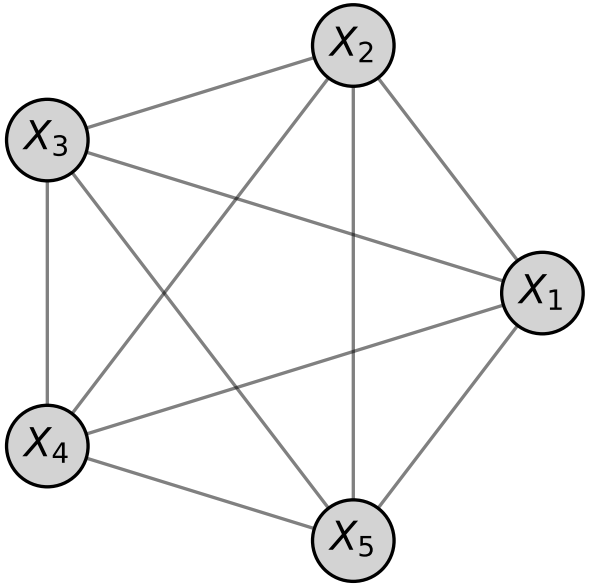
  ➜ In theory: possible

  ➜ In practice: difficult for large graphs/datasets



Background: Causal Discovery

- Recent related work: continuous-optimization score-based causal discovery
- Search space of possible graphs with gradient based methods
- Main problems:
- How can we limit the search space to directed acyclic graphs?
  → Currently: constraint-based optimization or regularization
- How can we efficiently search a discrete space?
- How scalable is the method?
- What guarantees do we have to find the correct graph?

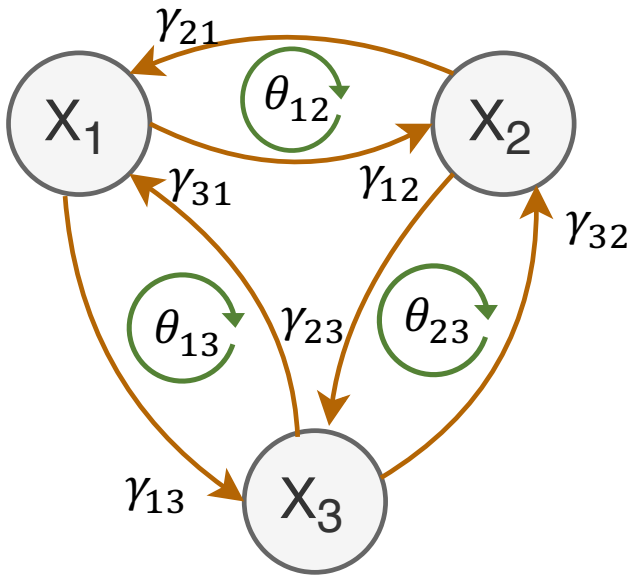


ENCO: Efficient Neural Causal Discovery

- Central idea: learn distributions $p(X_1|\ldots)$ from observational data,
  test generalization to interventional data
- Parameterize graph with edge existence and orientation parameters
- Orientation essential for breaking loops

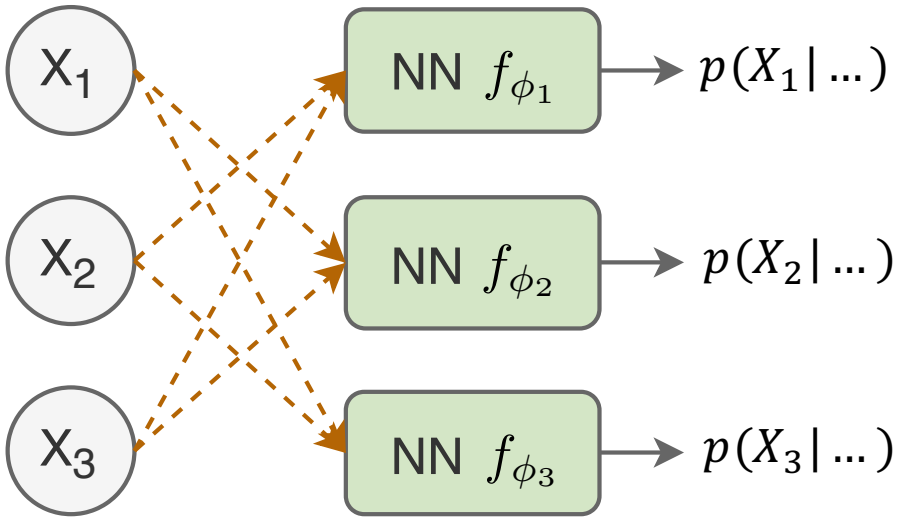
$$p(X_1)p(X_2|X_1) = p(X_2)p(X_1|X_2)$$

- Probability of an edge: $\sigma(\gamma_{ij}) \cdot \sigma(\theta_{ij})$
- Goal: the probability for correct edges goes to 1, all others to 0
- $\Rightarrow$ No constraint or regularization for acyclicity needed!



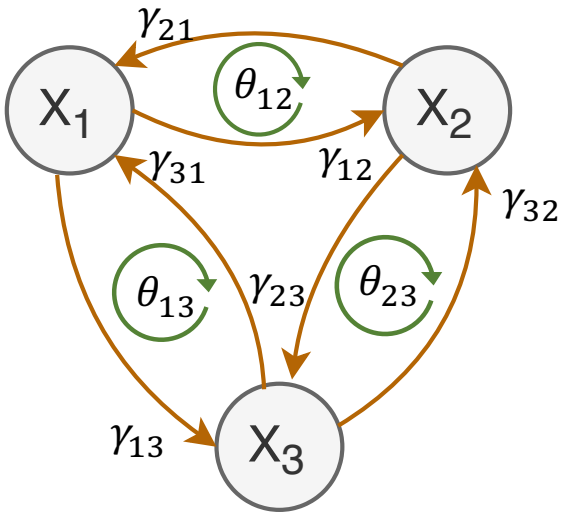
ENCO: Efficient Neural Causal Discovery

Distribution fitting



Alternate between both steps

Graph fitting



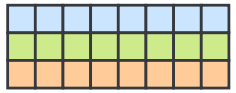
➜ Learn neural networks fitting conditional distributions on observational data

➜ Learn edge and orientation parameters based on fitted distributions

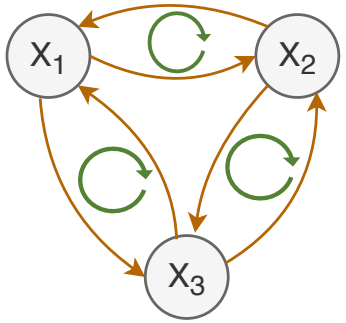
Graph fitting

Intervention

Data batch



Graph parameters



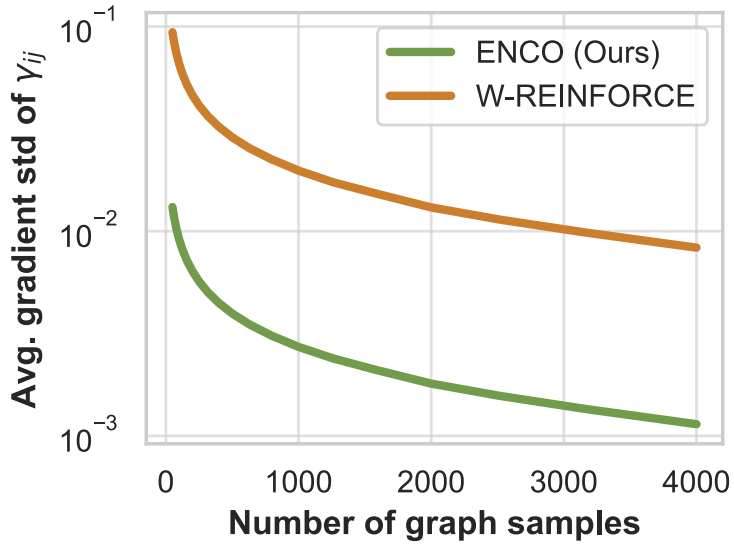
ENCO: Efficient Neural Causal Discovery

- Efficient low-variance gradient estimators for edge and orientation parameters
- Edge gradients:

$$\frac{\partial}{\partial \gamma_{ij}} \mathcal{L} = \alpha \cdot \mathbb{E}_{\boldsymbol{X}, C_{-ij}} \left[\mathcal{L}_{X_i \to X_j}(X_j) - \mathcal{L}_{X_i \not\to X_j}(X_j) + \lambda_{\text{sparse}} \right]$$

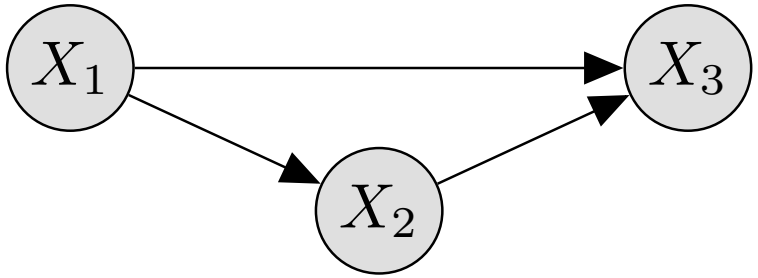
Graph/Data samples | Log likelihood w/o edge | Sparsity regularizer

- Sample and evaluate K graphs to estimate whether an edge is "beneficial" or not
- Similar idea for orientation parameters, but only with direct interventional data



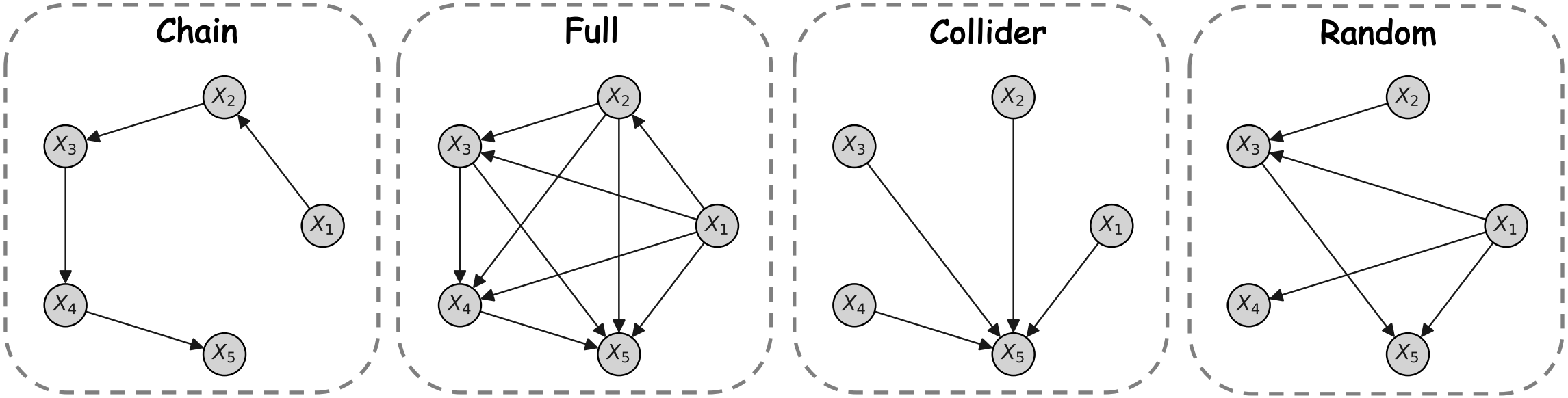
ENCO: Efficient Neural Causal Discovery

- ENCO provides guarantees for finding the true graph
- **Main conditions**: for every edge $X_i \to X_j$ in the causal graph,
  - edge $X_i \to X_j$ must not be disadvantageous for the log likelihood estimate of X_j under interventions on X_i
  - edge $X_i \to X_j$ must have a greater impact on the log-likelihood estimate than the sparsity regularizer λ_{sparse} in all situations
- If the assumption is not fulfilled, we still often find the correct graph



EXPERIMENTS

- Recover syntheticly generated graphs
- Testing various common graph forms to find weaknesses
- Graph size: 25 nodes
- Metric: Structural Hamming Distance (SHD) = FP + FN + wrongly orientated edges



Results

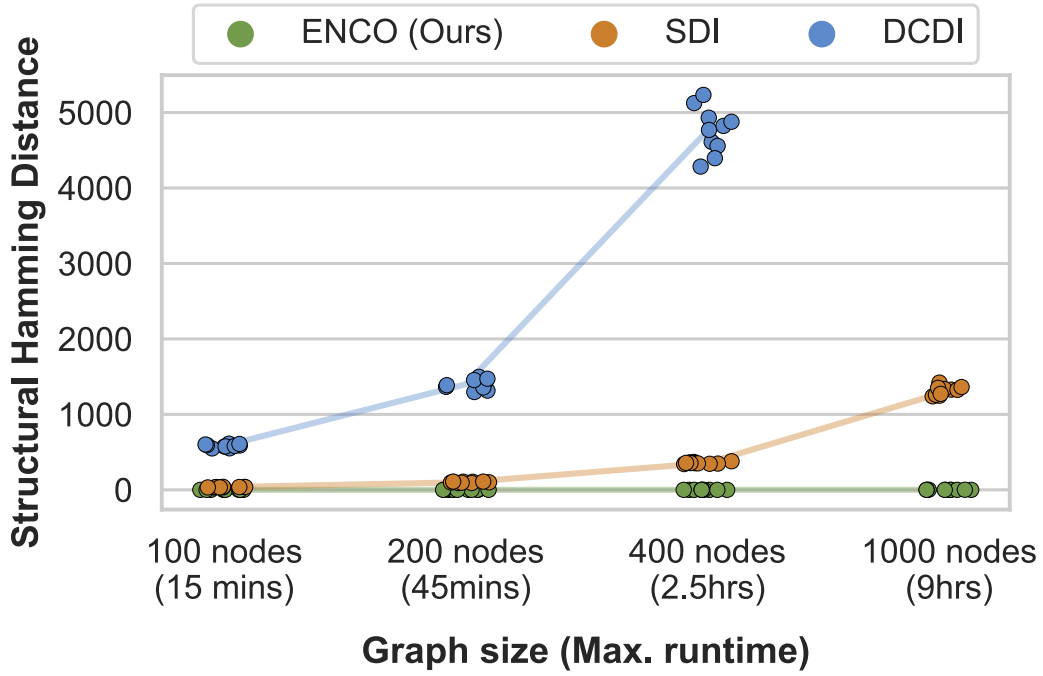
- Causality is even philosophically hard to grasp, quantify, evaluate
- Causal inference is the next frontier in Deep Learning, Machine Learning, and Computer Vision
- This method works very well but expect examples of interactions on all variables

Graph type	bidiag	chain	collider	full	jungle	random
GIES [15]	47.4 (±5.2)	22.3 (±3.5)	13.3 (±3.0)	152.7 (±12.0)	53.9 (±8.9)	86.1 (±12.0)
IGSP [47]	33.0 (±4.2)	12.0 (±1.9)	23.4 (±2.2)	264.6 (±7.4)	38.6 (±5.7)	76.3 (±7.7)
SDI [20]	2.1 (±1.5)	0.8 (±0.9)	14.7 (±4.0)	121.6 (±18.4)	1.8 (±1.6)	1.8 (±1.9)
DCDI [4]	3.7 (±1.5)	4.0 (±1.3)	**0.0** (±0.0)	2.8 (±2.1)	1.2 (±1.5)	2.2 (±1.5)
ENCO	**0.0** (±0.0)	**0.0** (±0.0)	**0.0** (±0.0)	**0.3** (±0.9)	**0.0** (±0.0)	**0.0** (±0.0)

Dataset	cancer [21] (5 nodes)	asia [22] (8 nodes)	sachs [37] (11 nodes)	child [41] (20 nodes)	alarm [2] (37 nodes)	diabetes [1] (413 nodes)	pigs [40] (441 nodes)
SDI [20]	3.0	4.0	7.0	11.8	24.6	422.4	18.0
ENCO	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**2.0**	**0.0**

Scaling up

- Testing scalability of the approach with synthetic graphs of up to 1000 nodes
- All baselines got the same computational resources
- On average, less than 1 mistake among 1 million edges for largest graph



Code (soon) available

SPACE-TIME - DYNAMICAL SYSTEMS - SYMMETRIES

ROTO-TRANSLATED LOCAL COORDINATE FRAMES FOR INTERACTING DYNAMICAL SYSTEMS

NEURIPS 2021

M. Kofinas

N. Shankar Nagaraja

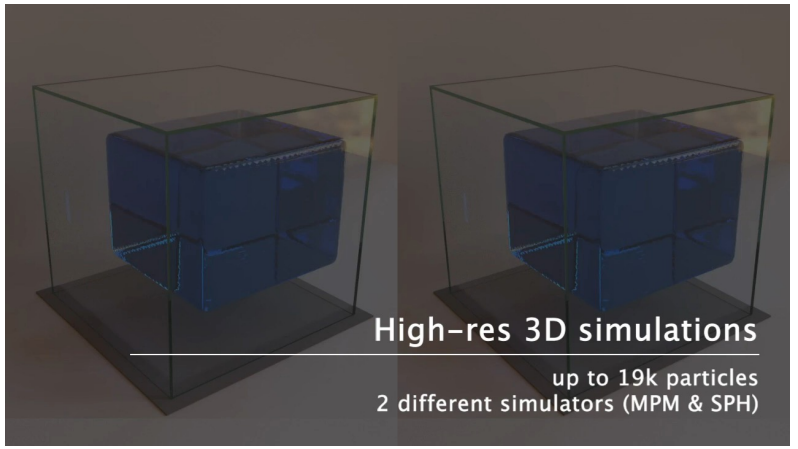
E. Gavves

University of Amsterdam / Efstratios Gavves

Dynamical Systems Interacting in Space-Time

- Systems of interacting objects with highly non-linear and time-dependent behavior
- Typically, hard-coded given 'forward' knowledge, *e.g., the trajectory of the single pendulum*
- Complexity increases very quickly, *e.g., going from single to double pendulum*
- Can we learn dynamical systems from observations?
  - Vehicle trajectories, particles colliding, material molecules, learning physics

Ma et al., AAAI, 2019



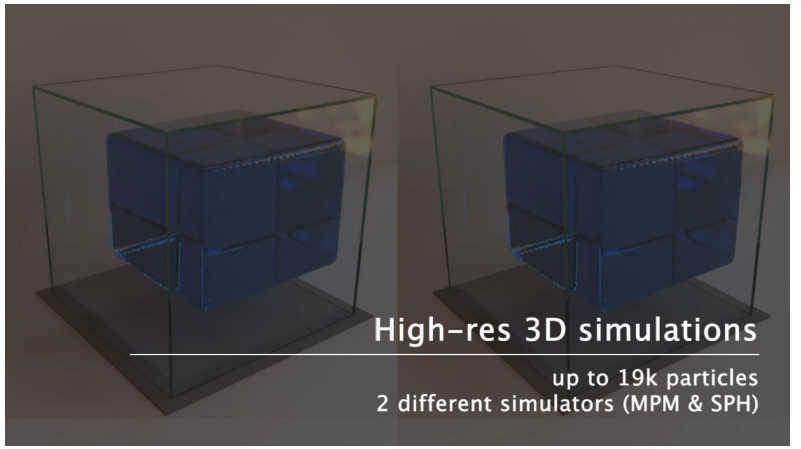
Sanchez-Gonzalez, ICML 20

GALILEAN INVARIANCE

- Dynamics are the same no matter the reference frame
  - Galilean invariance
- Input representations should be invariant to rotation and translation arbitrariness of reference frame
- And, output representations should be equivariant

- The way bicyclists move and avoid <u>does not depend on where I sit and look at them</u>

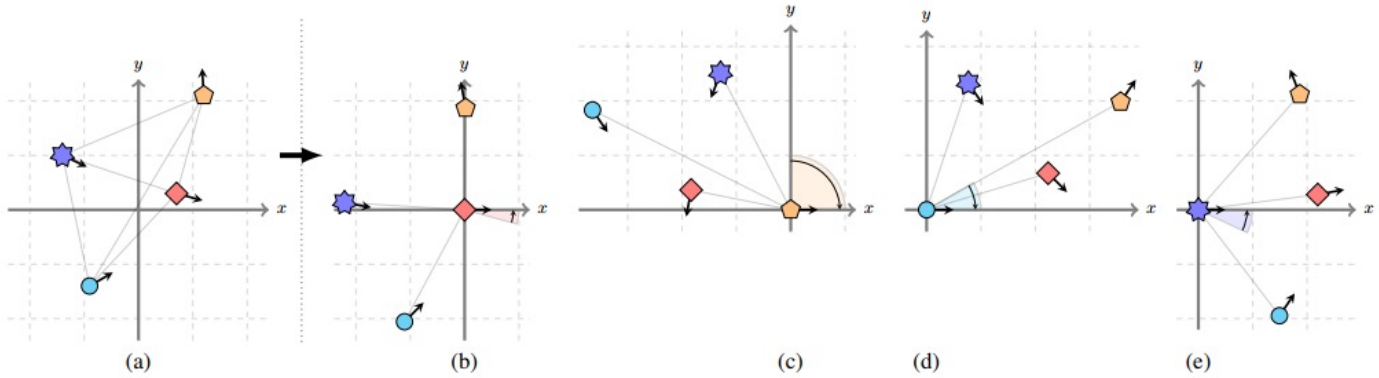
Ma et al., AAAI, 2019



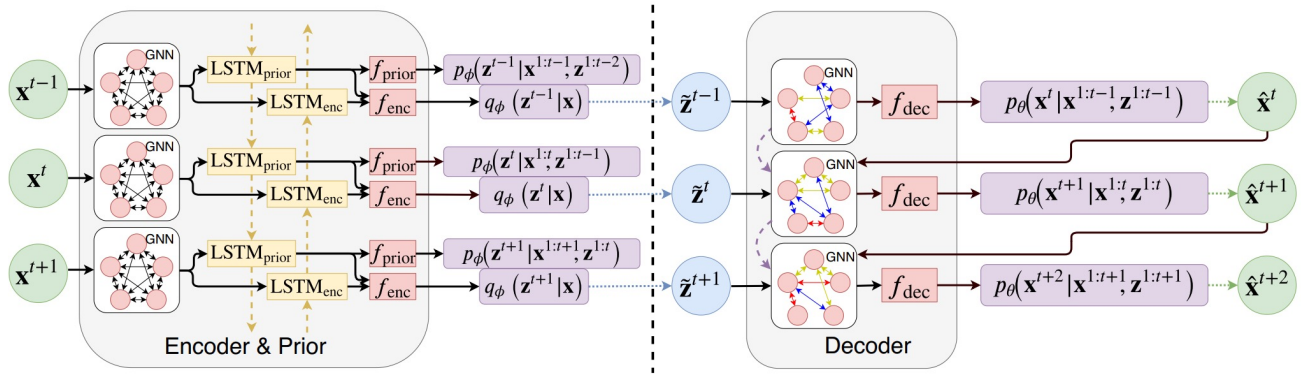
Sanchez-Gonzalez, ICML 20

Local Coordinate Frames

- Simply, change input representation

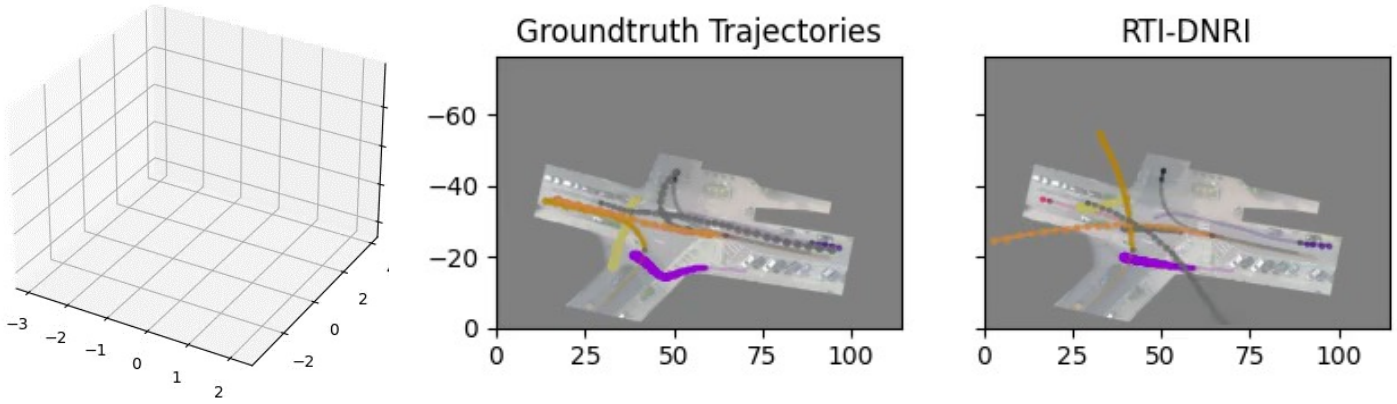


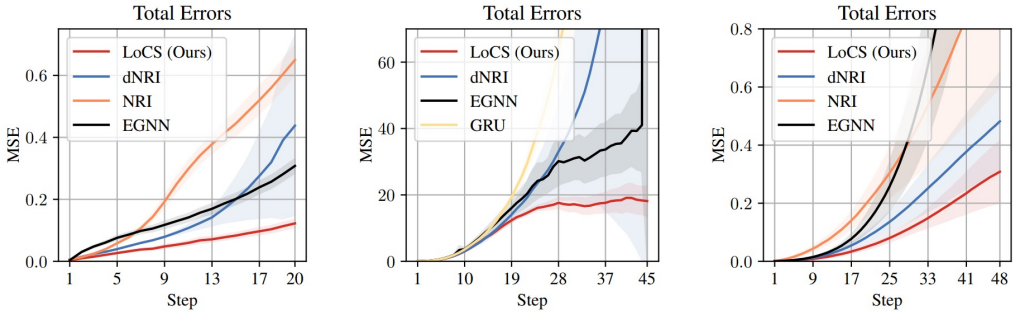
- Then, deploy existing neural dynamical inference models (Kipf 2019, Graber 2020)



Results & Take-away

- Considering local invariances/equivariances leads to great decrease in reducing forecasting errors
- One must take account others 'point-of-view' to make good future predictions
- Generalizes well to different dynamical systems: traffic scenes, colliding particles, 3D motion capture
- Exciting direction for research in machine perception and learning





(a) (b) (c)

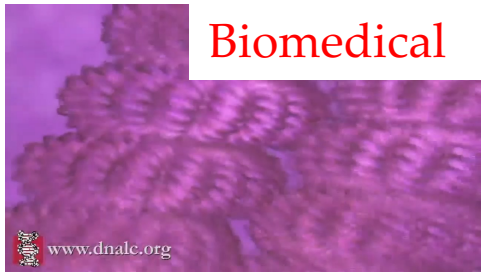
Figure 4: Total error curves in: (a) charged particles, (b) inD, (c) motion #35

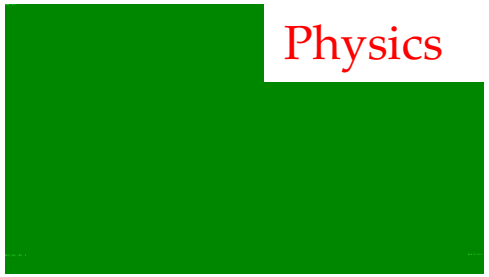
Spatiotemporal Learning in Sciences

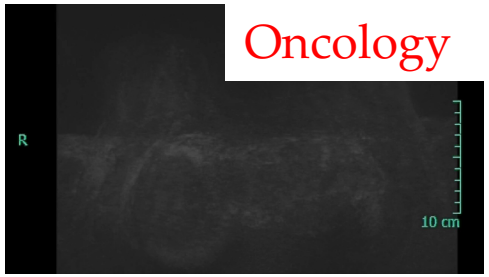
SCIENCES ⇔ SPATIOTEMPORAL RECORDINGS

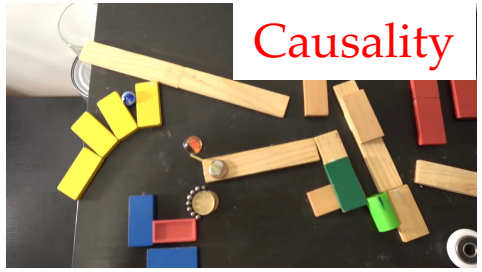
- Majority of **scientific recordings are videos** → Endless opportunities & exciting problems
- Our I3D's and trackers and deep networks should works with any **Space** and **Time** signal

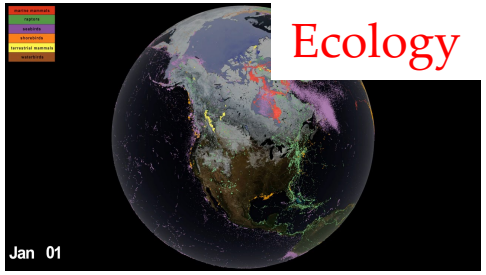
Space+Time = Video











Particle physics

SCIENTIFIC KNOWLEDGE AS GROUND TRUTH

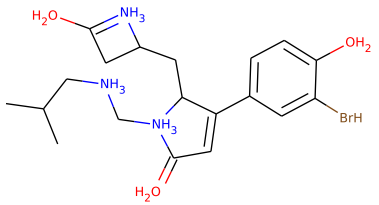
- What if we expand our definition of a ‘video’ to any sequence in space and time?
- What if we use scientific recordings as space-time data to learn and evaluate?
- Bonus: existing scientific knowledge as ‘perfect ground truth’

- Already some examples of particles colliding. What about glaciers melting? Astronomical objects rotating or pulsating? Biomedical sequences? Chemical reactions?
- Shouldn’t our I3Ds, ResNets, VideoLSTMs, Siamese Trackers work there?

- Can we prepare a decathlon on space-time data from various sciences for systematic evaluation

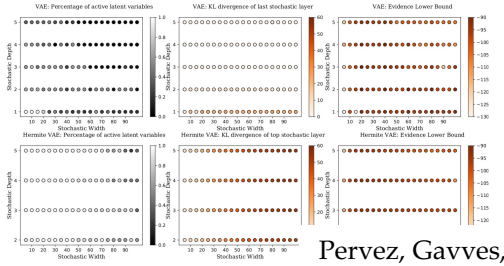
More research

Categorical Normalizing Flows



Lippe, Gavves, ICLR 2021

Equivariant Siamese Trackers

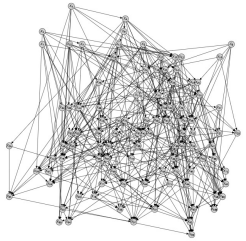


Pervez, Gavves, ICML 2021

Matching Implicit Functions

Chen, Fernando, Bilen, Mensink, Gavves, ICML 2021

Spiking Neural Networks



Lippe, Cohen, Gavves, Under review

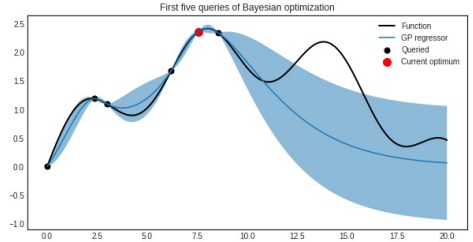
Boolean Neural Networks

$$\mathbf{f} : \{0,1\}^{\mathbf{n}} \rightarrow \{0,1\}$$

The Boolean function

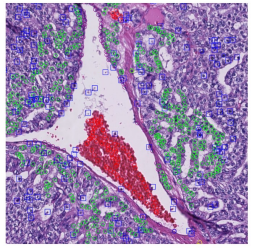
Pervez, Gavves, ICML 2020

Bayesian Optimization for Permutations



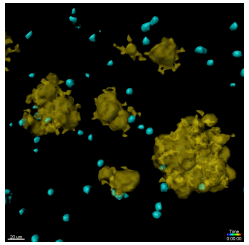
Oh, Bondesan, Gavves, Welling, Ongoing

Learning symmetries from data



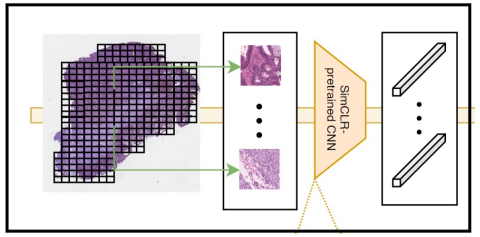
Panteli, Teuwen, Horlings, Gavves, ICCV 2021

Cell tracking for Cancer Drug Develop.



Lehman, Panteli, Teuwen, Gavves, Ongoing

Predicting genomic mutational burden



Schirris, Gavves, Horlings, Teuwen, Ongoing

JOIN THE TEAM

- Looking for a PostDoc to help guide the team and grow
- Motivated PhD candidates are always welcome
- Always more than interested in collaborations


Harmonic Analysis in Learning Systems


Spatiotemporal neural forecasting


Causality


Time-inspired generative models


Dynamical systems and memory nets


Neural networks as dynamical systems


Spate-time geometric deep learning


Physics-inspired machine learning


Self-supervised video representation learning


Long-term video representations


Efficient spatiotemporal representations


3D deep learning


Spate-time spatiotemporal CT


Transfer learning CT


Generative models for adaptive radiotherapy


Object detection in comp. pathology


Interactive medical segmentation


Self-supervised learning in computational pathology

TAKE-HOME MESSAGE

- Shifting paradigm: Static → Temporal
- Geometry & Dymamical Systems → Way forward
- Exciting time for CV/ML, and AI, especially w.r.t. the Sciences!
- Ping me in egavves@uva.nl if interested

Code (soon) available

SPACE-TIME - TRACKING - GEOMETRIES

ROTATION EQUIVARIANT SIAMESE INSTANCE SEARCH FOR TRACKING

CVPR 2021

D. Gupta

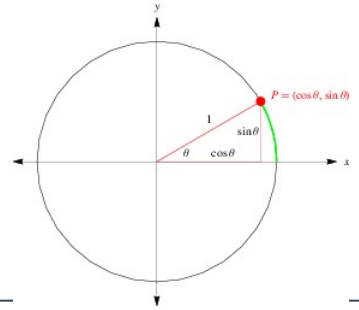
D. Arya

E. Gavves

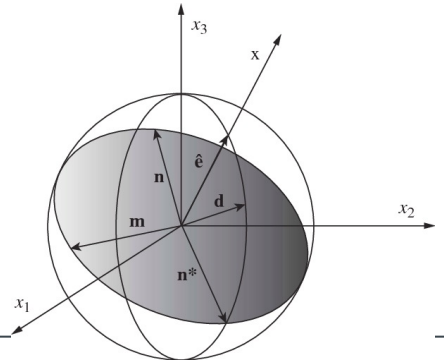
ROTATIONAL GEOMETRIES IN SPACE-TIME VIDEO SEQUENCES

- Objects constantly rotate
- Without changing their identity or semantics
- Ideally, our inferences should be either invariant or equivariant (proportional) to such rotations

In-plane rotation



Out-of-plane rotation



WHAT IS EQUIVARIANCE?

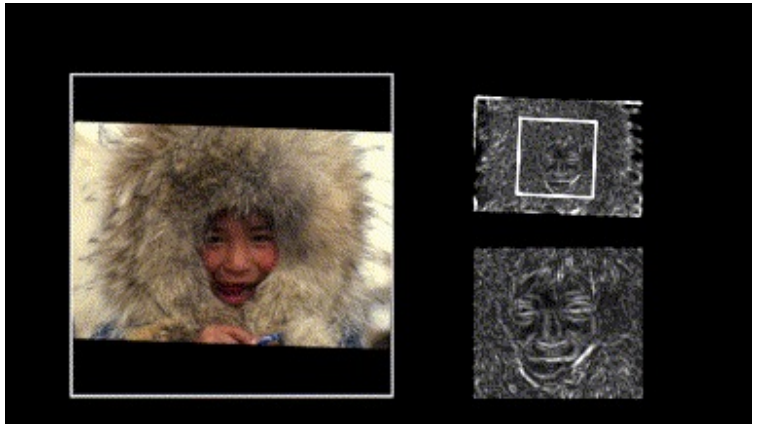
- Roughly, change in the input shall be proportionally transferred in the output of a function
- Mathematically, for a *Group* G a function $f: X \to Y$ if

$$f(g \cdot x) = g \cdot f(x) \text{ for } g \in G, x \in X$$

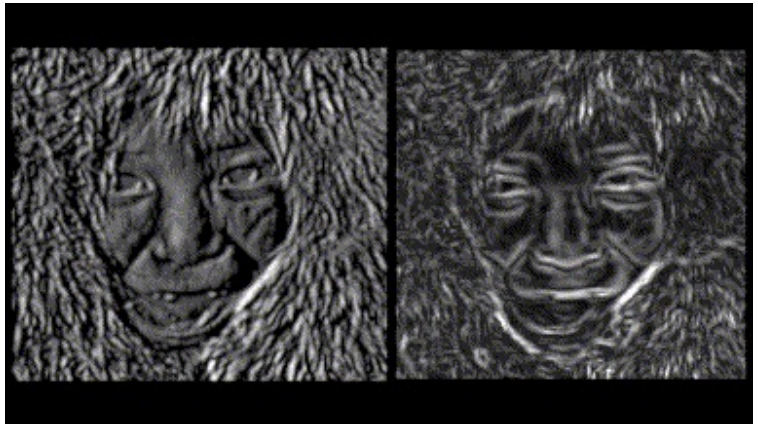
- Groups are mathematical objects far broader than rotations or translations
- Very much relate to the study of symmetries, check also M. Bronstein fantastic talks & books

Regular CNN

Rotation Equivariant NN

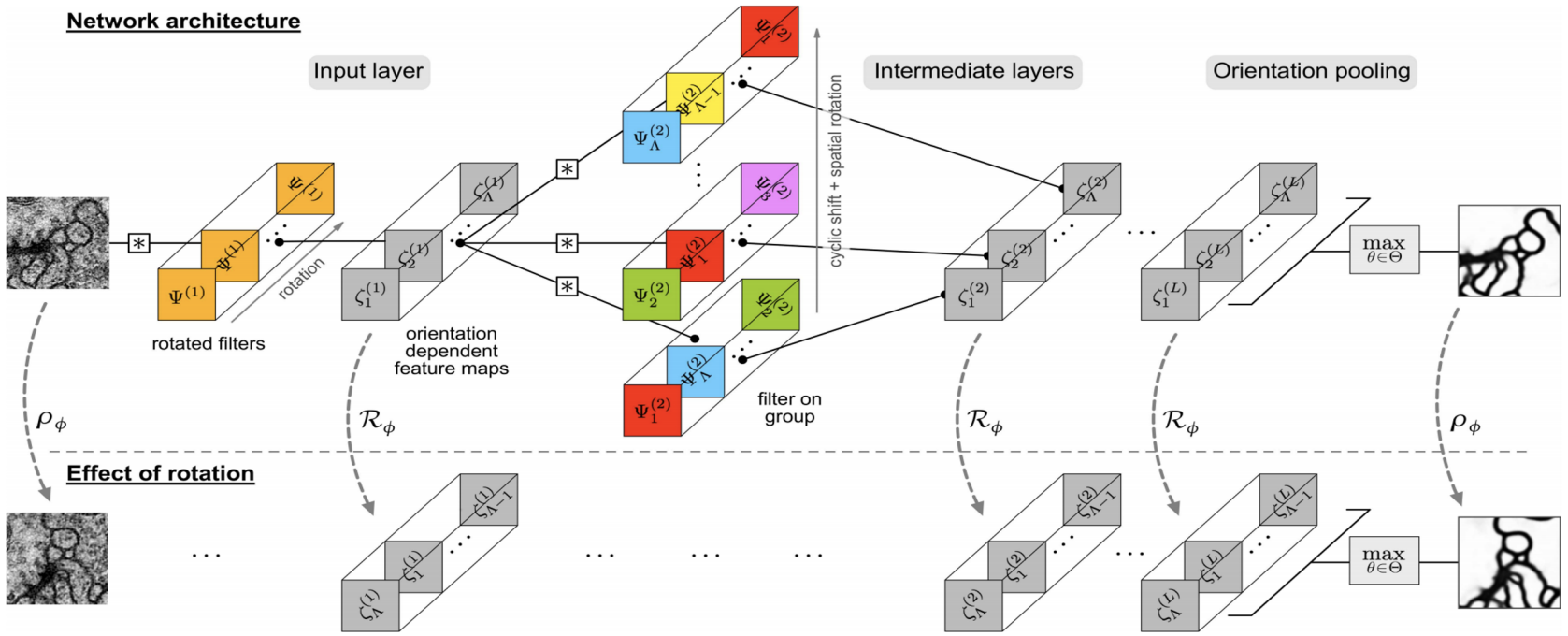


Difference



(H-Net; Worrall et al, CVPR 2017)

Rotation Equivariance with Circular Harmonics (Weiler et al., 2018)



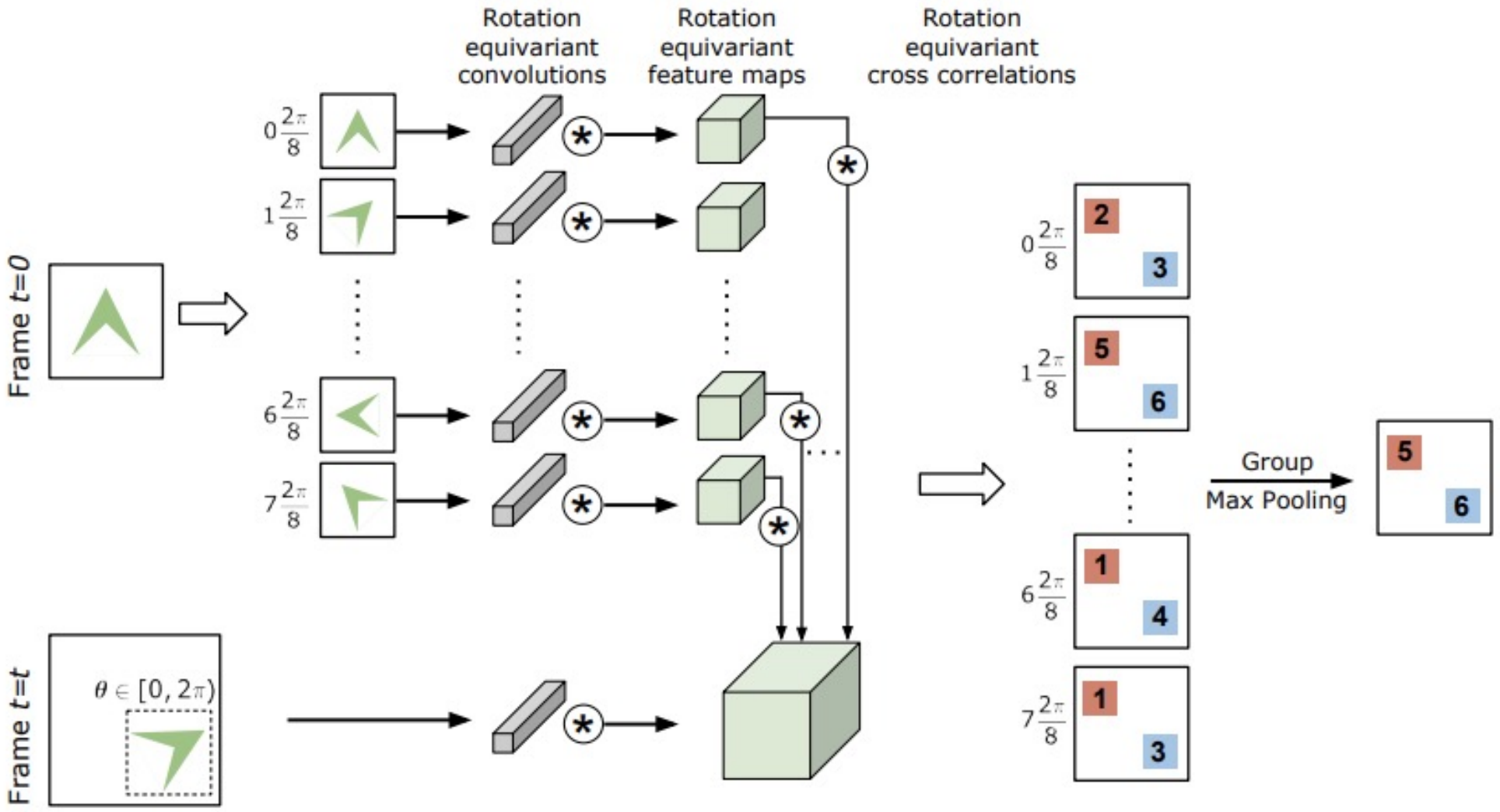
- Filters defined with circular harmonics as basis
- Rotating the filters means reverse rotating harmonics
- No need to optimize more filters to implicitly account for rotations

$$\psi_{jk}(r,\phi) = \tau_j(r)e^{ik\phi}$$

$$\rho_\theta \psi_{jk}(x) = e^{-ik\theta}\psi_{jk}(x)$$

$$\rho_\theta \Psi(x) = \sum_{j=1}^{J}\sum_{k=0}^{K} w_{jk} e^{-ik\theta}\psi_{jk}(x)$$

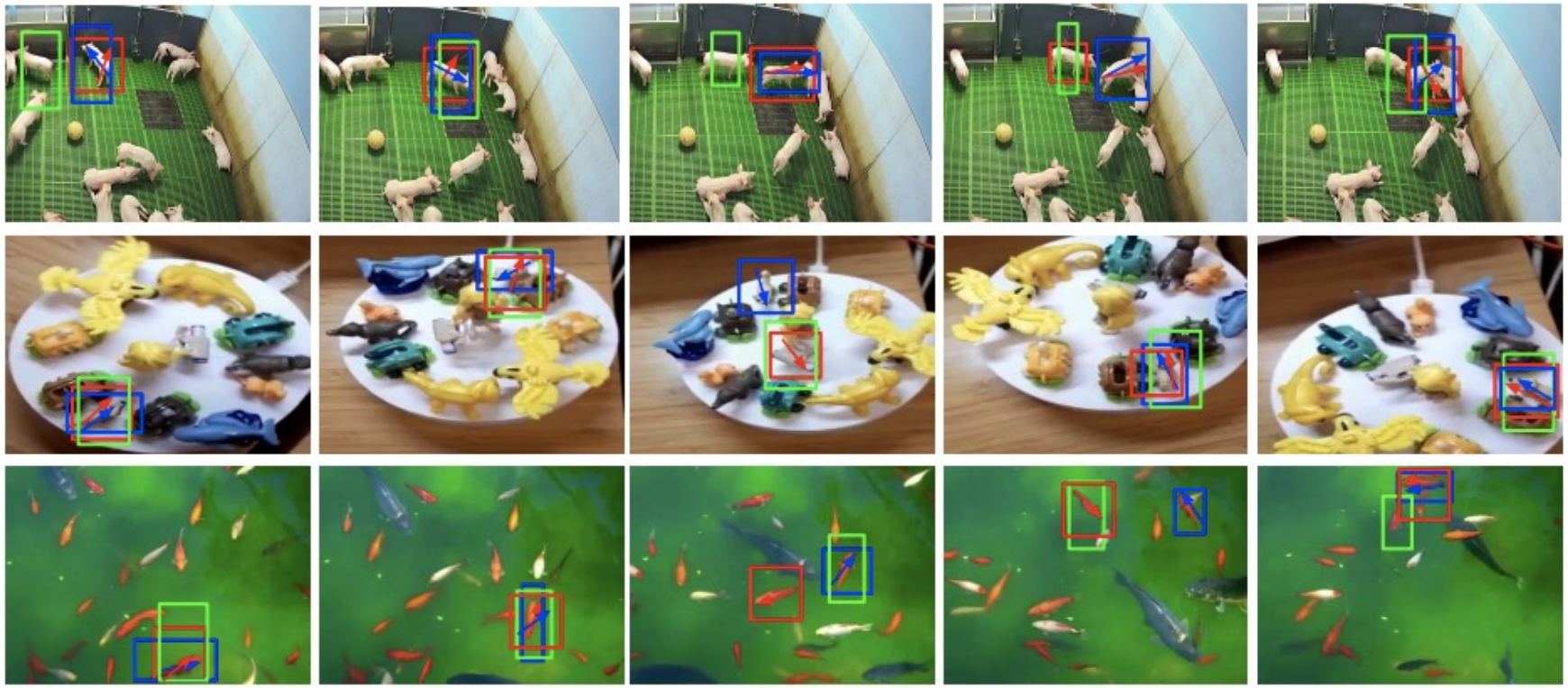
ROTATION EQUIVARIANT SIAMESE TRACKERS



[1] Gupta, Arya, Gavves, Rotation Equivariant Siamese Networks, CVPR 2021
[2] Tao, Gavves, Smeulders, Siamese Instance Networks for Tracking, CVPR 2016

RESULTS & TAKE-AWAY

- Improves greatly (10% or more) when object rotate and stable when not
- Works only for in-plane rotations but cannot do out-of-plane
- Considering rotational geometries in videos makes a difference in precise inferences



[1] Gupta, Arya, Gavves, Rotation Equivariant Siamese Networks, CVPR 2021
[2] Tao, Gavves, Smeulders, Siamese Instance Networks for Tracking, CVPR 2016

Code (soon) available

SPARSE-SHOT LEARNING WITH EXCLUSIVE CROSS-ENTROPY FOR EXTREMELY MANY LOCALIZATIONS

ICCV 2021




A. Panteli J. Teuwen H. Horlings E. Gavves

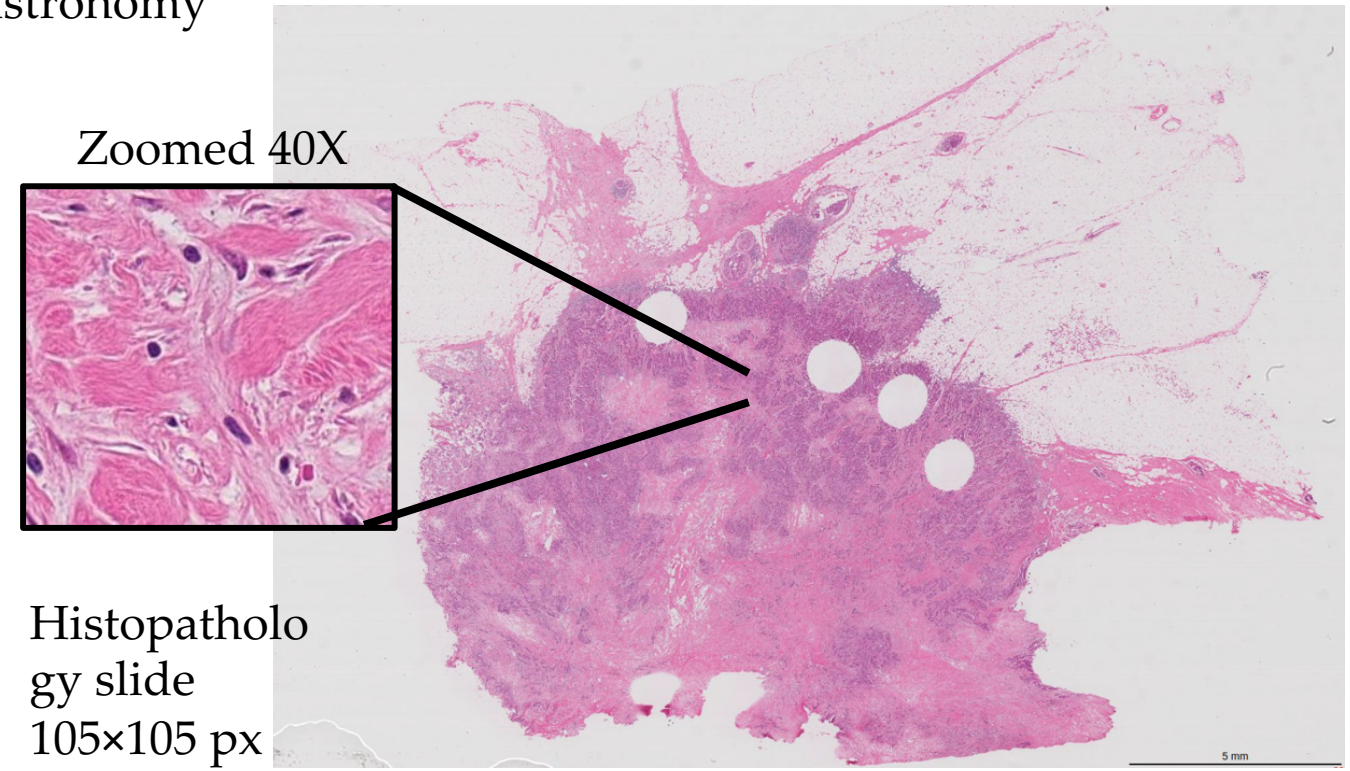
University of Amsterdam / Efstratios Gavves

Sparse Annotations

- Images of extreme dimensionalities → millions of objects to annotate
- Impossible to use ordinary machine learning models out of the box
- Difficult domains: pathology, microscopy, astronomy

MNIST
28×28 px
<10 sec

MS COCO
500×500 px
<5 min



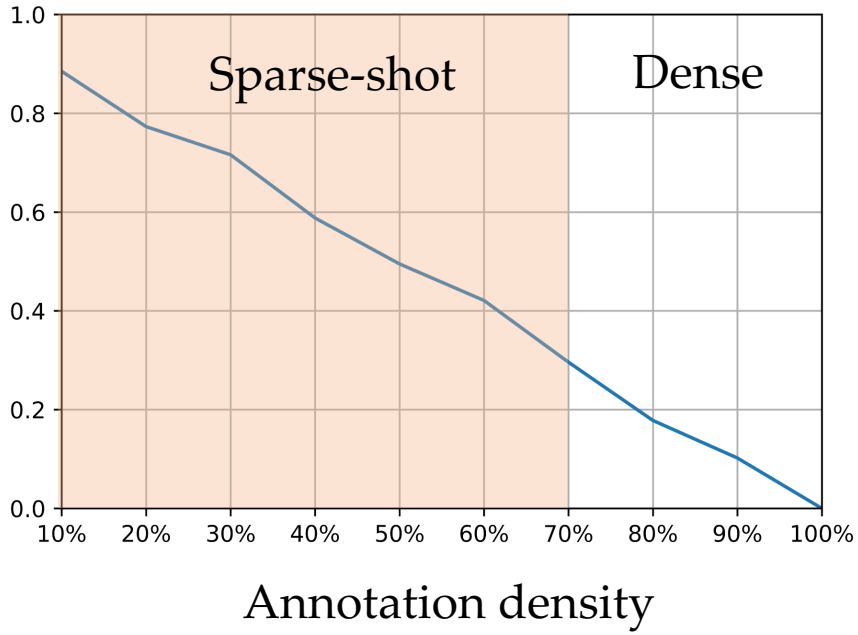
Histopatholo
gy slide
105×105 px
>100 days

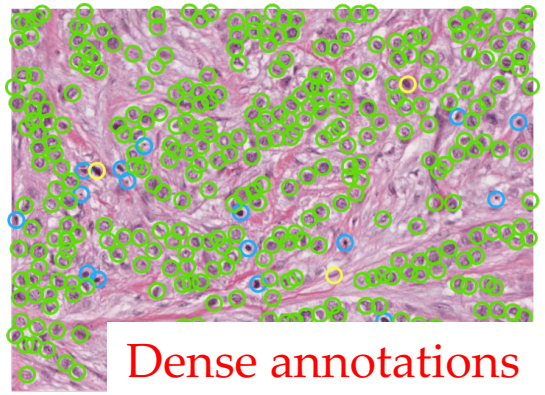
Zoomed 40x

Catastrophic biasing due to sparsity

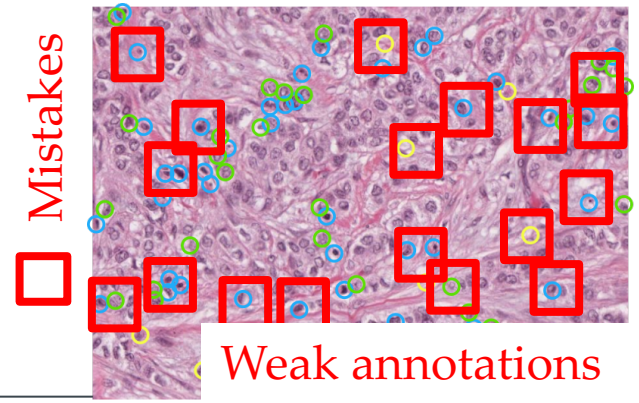
- Often less than 10% of available annotations → extreme information loss
- We cannot assume that background is all ‘true negatives’
- Weak (pseudo) annotations extremely wrong dominating gradients
- More than 30% missing annotations → catastrophic learning

$$bias = -\sum_{\text{Unannotated}} \log\left(1 - p_i\right)$$




Dense annotations

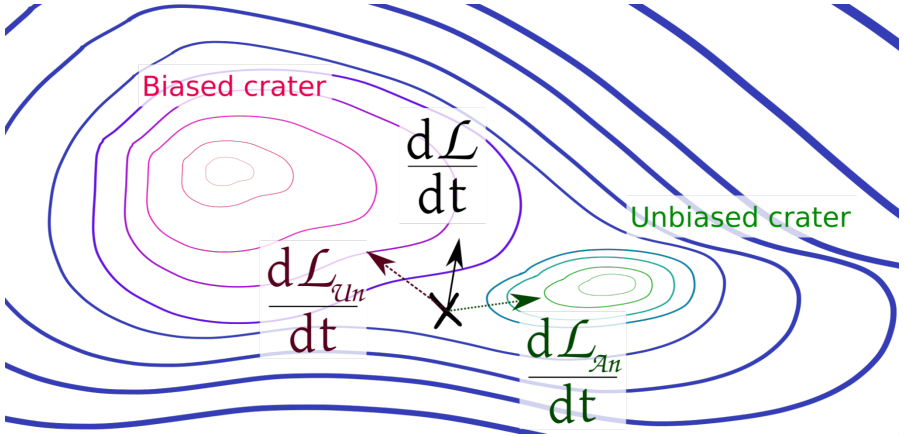

Sparse annotations



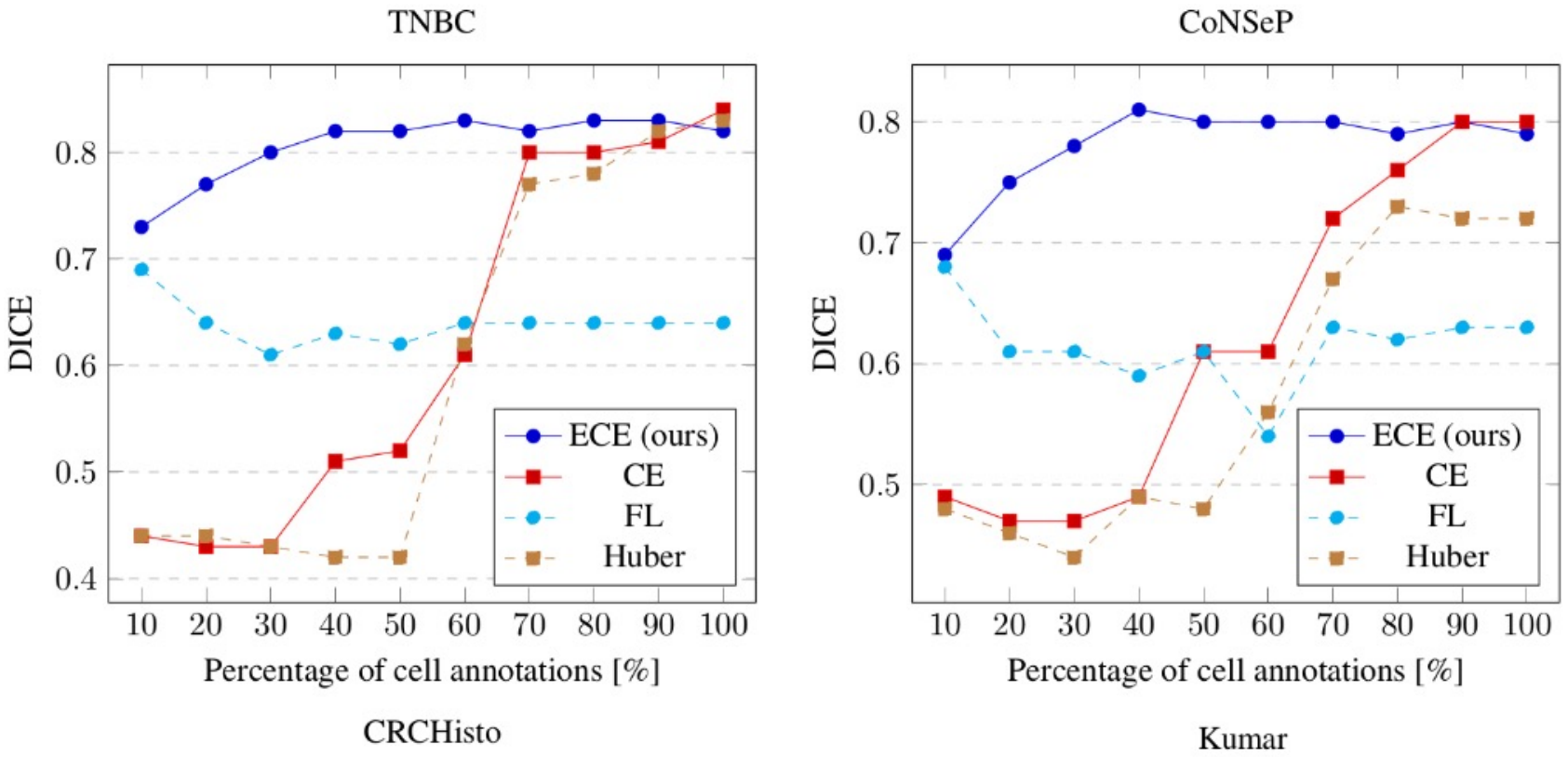
Weak annotations

KEY IDEA

- Biasing gradients push the model towards suboptimal craters
  - Especially, in early training where pseudo-annotations are untrustworthy
  - Even in dense learning early confidence is unreliable
- We could ignore pseudo-annotations, but then discard most of our dataset
- Instead, we slow down speed of learning from pseudo-annotations
  - And learn unbiased features first, before incorporate more background information



Results



- Impressive performance even with 90% of annotations missing
- Qualitatively, the results were even more impressive, but hard to show due to missing labels
- With only ~40% of annotations same performance as with full annotations
- Likely a strong connection with noisy labels and a great fit to video

QUALITATIVE RESULTS

- Almost all boxes detected with only 30% annotations
- High-quality localization avoiding background as false positives