Synthesis Lectures on Computer Vision

SYNTHESIS
COLLECTION OF TECHNOLOGY

Yue Song · Thomas Anderson Keller ·
Nicu Sebe · Max Welling

Structured Representation Learning

From Homomorphisms and Disentanglement to Equivariance and Topography

Springer

Synthesis Lectures on Computer Vision

This series publishes on topics pertaining to computer vision and pattern recognition. The scope follows the purview of premier computer science conferences, and includes the science of scene reconstruction, event detection, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, and image restoration. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems, such as those in self-driving cars/navigation systems, medical image analysis, and industrial robots.

Yue Song · Thomas Anderson Keller ·
Nicu Sebe · Max Welling

Structured Representation Learning

From Homomorphisms and Disentanglement to Equivariance and Topography

Springer

Yue Song
Department of Computing and Mathematical Sciences
California Institute of Technology
Milan, Italy

Thomas Anderson Keller
Kempner Institute
Harvard University
Cambridge, MA, USA

Nicu Sebe
Department of Computer Science and Information Engineering
University of Trento
Trento, Italy

Max Welling
Department of Machine Learning
University of Amsterdam
Amsterdam, The Netherlands

ISSN 2153-1056 ISSN 2153-1064 (electronic)
Synthesis Lectures on Computer Vision
ISBN 978-3-031-88110-7 ISBN 978-3-031-88111-4 (eBook)
https://doi.org/10.1007/978-3-031-88111-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface I

The art of machine learning (ML) is to optimally combine inductive bias (a.k.a. priors) with data. With more data, we need to include less prior information and we can let the data speak. This is the modern “scaling” paradigm of LLMs and Foundation models with trillions of parameters, trained on hundreds of thousands of GPUs on the entirety of the internet. For these models, the architecture of choice are usually Transformers, precisely because they scale well.

Are Transformers the final answer? That seems unlikely. In this book, we ask ourselves if there are architectures based on better priors about the world but that also scale to internet-level models. These models will not only be able to learn from fewer data but also exhibit improved scaling laws, ideally with a steeper slope.

What are some interesting priors to build into deep architectures? In this book, we are inspired by both neuroscience and physics. Neuroscience has been ML’s companion right from the start. Early architectures such as Rosenblatt’s Perceptron were already inspired by biological neurons. It’s only more recently that the two fields have gone their own separate ways. But given the huge gap in energy efficiency between artificial and biological neural networks, it may make sense to look at neuroscience again for inspiration.

In this book, we explore the possibility to use oscillators and traveling waves as a new computing paradigm, rather than static representations. Waves have the potential to collect and combine information from long distances, both on space and in time.

Another interesting prior is that in the world around us at the scale that we understand it (objects), things usually don’t change very fast. This is of course different than the nanosecond fluctuation at the level of individual atoms. As deep models build coarse-grained representations in their deeper levels, it seems reasonable to enforce this slowness at the level of abstract (deep) representations.

Since our models often model things in our physical world, we can also contemplate whether the symmetries of that physical world should be represented in our representations. There is now a rich literature on symmetries, such as translational and rotational symmetries that are exactly enforced through (irreducible or regular) equivariant group representations. However, often we do not know the symmetries present in the data, or some regularities might not be described by groups, or we may not simply know the representations of certain groups. In all these cases we need to generalize the concept of equivariance, on which these hard-coded symmetries are based.

In this book, we consider homomorphisms between latent representations and the world as an appropriate signal to learn such approximate symmetries. That is to say, the dynamics of the latent representations should mirror the corresponding dynamics of the world. What dynamics, or set of transformation of our latent code, do we entertain when we try to learn homomorphisms between the world and our deep representations? Here we are inspired again by neuroscience and physics: we have modeled these representations as collections of interacting oscillators, or in the continuum limit, PDEs, that support wave-like solutions. In some sense, we can think of these representations as a fluid in which waves can develop to perform computations.

And this brings me to my final point. Due to availability of multi-electrode sensors, waves are now commonly detected in the brain, and neuroscience researchers are starting to ask what its computational benefits might be. Can waves transport and combine information in new ways that we have not yet discovered? This is an intriguing possibility about which I have no doubt we will hear a lot more over the course of the next decade.

I wish you an interesting journey as you travel through the chapters of this book and become inspired to think of new ways to build inductive biases into the next generation of ML models.

Amsterdam, The Netherlands
February 2025

Max Welling

Preface II

The field of machine learning stands at a critical juncture. While recent advances in deep learning have delivered remarkable breakthroughs across domains such as vision, language, and robotics, these successes often come at the cost of massive data requirements, computational inefficiency, a lack of interpretability, and poor generalization to novel scenarios. As machine learning systems are increasingly deployed in real-world applications, these challenges highlight the need for models that go beyond brute force learning and instead can learn more like humans—adaptively, efficiently, and intuitively.

This book, *Structured Representation Learning: From Homomorphisms and Disentanglement to Equivariance and Topography*, offers a timely exploration of how structured approaches can reshape the design and performance of machine learning systems. By embedding principles such as symmetry, topography, and compositionality directly into model architectures, structured representation learning provides a pathway to models that are more robust, efficient, and capable of generalization.

At its core, structured representation learning seeks to address fundamental questions: How can machine learning systems capture the inherent relationships within data, such as symmetries and invariances? How can models decompose complex phenomena into simpler, interpretable components? And how can we align computational representations with the physical and biological principles that govern the real world? This book explores these questions through key concepts such as:

- **Equivariance and Symmetry**: Learning approximately equivariant representations that respect the transformations of the data beyond group theory.
- **Disentanglement**: Designing latent representations that isolate meaningful factors of variation, serving as approximate learned equivariance.
- **Topographic Representations**: Drawing inspiration from biological systems to organize information spatially and temporally in ways that mimic biological neural networks.

- **Physical Priors**: Baking physical principles into machine learning systems to represent the physical relations in the real world.

Structured representation learning represents a paradigm shift. By incorporating the above beneficial inductive biases directly into learning systems, this approach opens the door to machine learning systems that are not only more efficient but also more interpretable and aligned with the complexities of the real world.

This book is written for researchers, practitioners, and students eager to explore the intersection of machine learning, computational neuroscience, and natural sciences. It provides a high-level perspective on the field's foundational ideas while delving into specific techniques and applications that demonstrate the power of structured representation learning. As the demands on machine learning systems continue to grow, structured representations offer a promising direction toward building models that can reason, adapt, and learn with greater data efficiency and generalization abilities. We invite readers to engage with the ideas in this book, explore the rich potential of structured representations, and join in shaping the future of machine intelligence.

Pasadena, CA, USA — Yue Song
Cambridge, MA, USA — Thomas Anderson Keller
November 2024

Contents

Acronyms

AI	Artificial Intelligence
AR	AutoRegressive Model
ELBO	Evidence Lower BOund
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HJE	Hamilton-Jacobi Equation
ICA	Independent Component Analysis
LLM	Large Language Models
LN	Layer Normalization
LSTM	Long Short-Term Memory
MLP	Multi-layer Perceptron
ODE	Ordinary Differential Equations
OoD	Out-of-Distribution
OT	Optimal Transport
PDE	Partial Differential Equation
PINN	Physics-Informed Neural Network
RNN	Recurrent Neural Network
SE(N)	Special Euclidean Group
SFA	Slow Feature Analysis
SO(N)	Special Orthogonal Group
SOM	Self-Organizing Map
VAE	Variational Auto-Encoder

List of Figures