On Contrastive Representations of Stochastic Processes

NeurIPS 2021

UNIVERSITY OF OXFORD

Emile Mathieu, Adam Foster* and Yee Whye Teh*

Stochastic processes

$$f \sim p(\cdot)$$

$$f : \mathcal{X} \to \mathcal{Y}$$

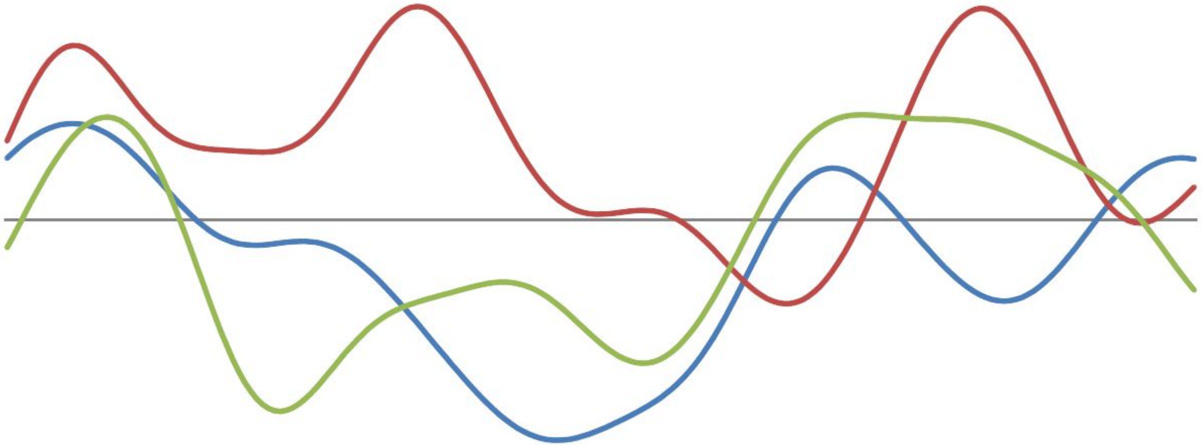
Stochastic processes

$$f \sim p(\cdot)$$

$$f : \mathcal{X} \to \mathcal{Y}$$

- 1D function

$$f : \mathbb{R} \to \mathbb{R}$$

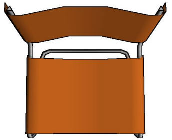


Stochastic processes (cont'd)

- 3D object view synthesis

$$f : \mathrm{SO}\left(3\right) \to 2D \text{ images}$$

$$\begin{bmatrix} \theta = 0 \\ \varphi = 0 \end{bmatrix} \longmapsto$$



$$\begin{bmatrix} \theta = \frac{\pi}{4} \\ \varphi = \frac{\pi}{4} \end{bmatrix} \longmapsto$$

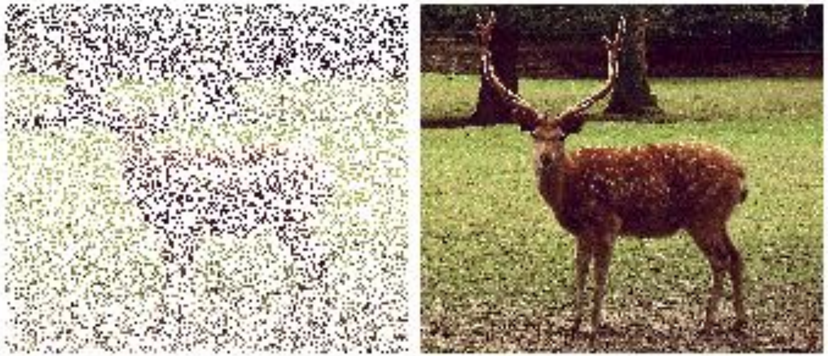
$$\begin{bmatrix} \theta = 0 \\ \varphi = 0 \end{bmatrix} \longmapsto$$

$$\begin{bmatrix} \theta = \frac{\pi}{4} \\ \varphi = \frac{\pi}{4} \end{bmatrix} \longmapsto$$

Stochastic processes (cont'd)

- Image in-filll

$$f : \mathbb{Z}^2 \to \mathbb{R}^3$$

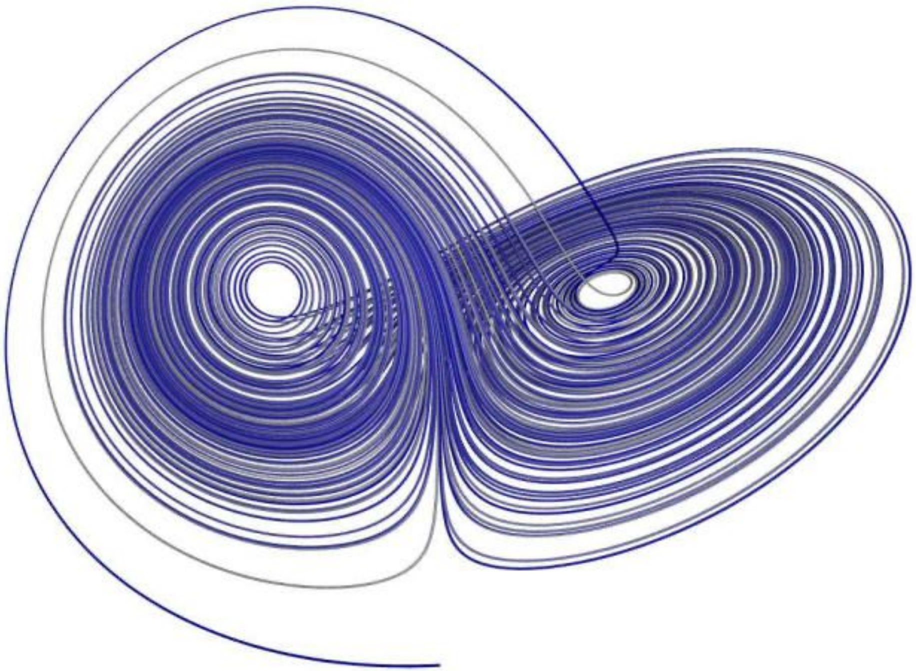


Stochastic processes (cont'd)

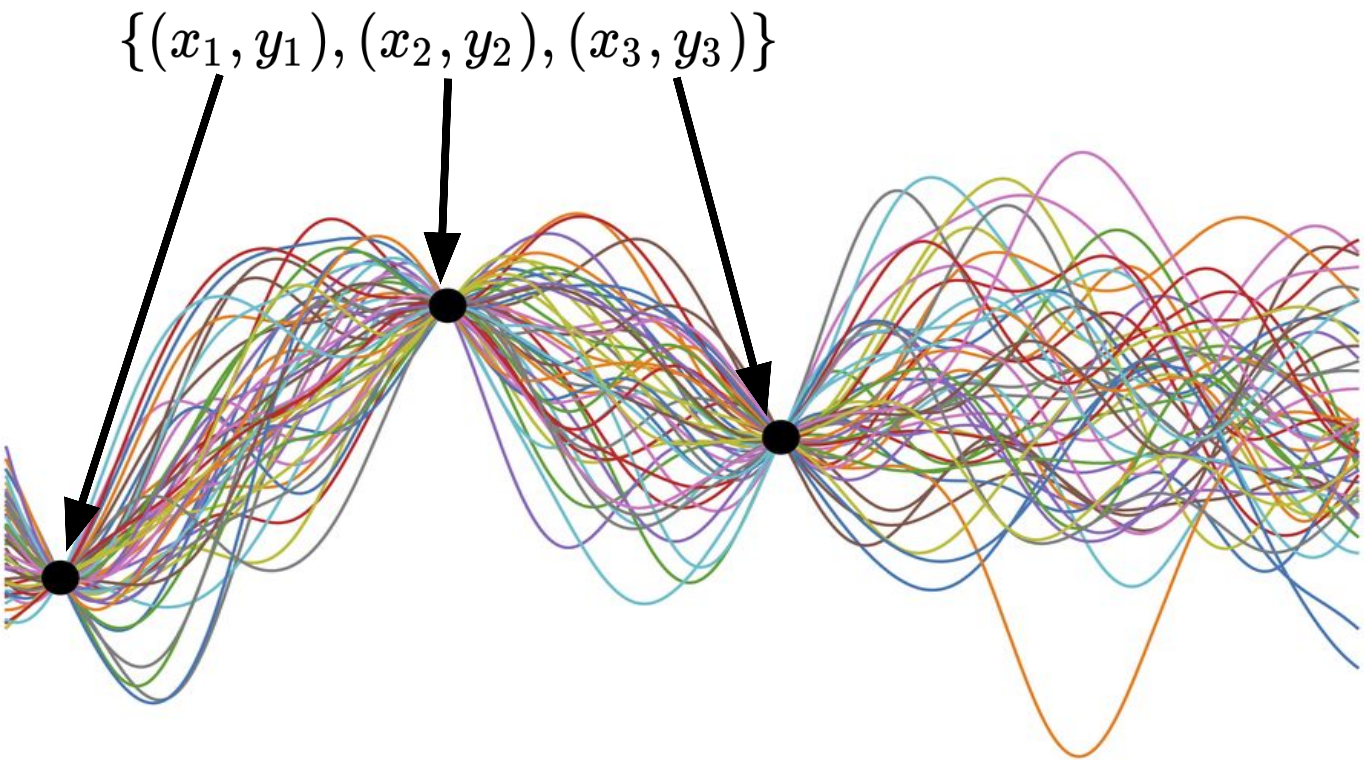
- Dynamical system

  time prediction

$$f : \mathbb{R} \to \mathbb{R}^3$$

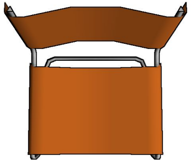


Context



Context (cont'd)

$$\Bigg\{ \left(\begin{bmatrix} \theta = 0 \\ \varphi = 0 \end{bmatrix}, \right.$$

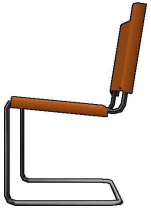


$$\left. \right),$$

$$\left(\begin{bmatrix} \theta = \frac{\pi}{4} \\ \varphi = \frac{\pi}{4} \end{bmatrix}, \right.$$

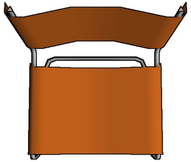
$$\left. \right),$$

$$\left(\begin{bmatrix} \theta = \frac{\pi}{2} \\ \varphi = \frac{\pi}{2} \end{bmatrix}, \right.$$

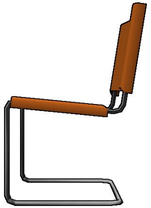


$$\left. \right) \Bigg\}$$

Representation learning

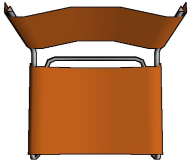
$$\left(\begin{bmatrix}\theta = 0 \\ \varphi = 0\end{bmatrix},\right.$$


$$\left(\begin{bmatrix}\theta = \frac{\pi}{4} \\ \varphi = \frac{\pi}{4}\end{bmatrix},\right.$$

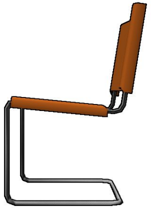

$$\left(\begin{bmatrix}\theta = \frac{\pi}{2} \\ \varphi = \frac{\pi}{2}\end{bmatrix},\right.$$


$\rightarrow$ *Representation*

Representation learning

$\begin{bmatrix} \theta = 0 \\ \varphi = 0 \end{bmatrix}$, 

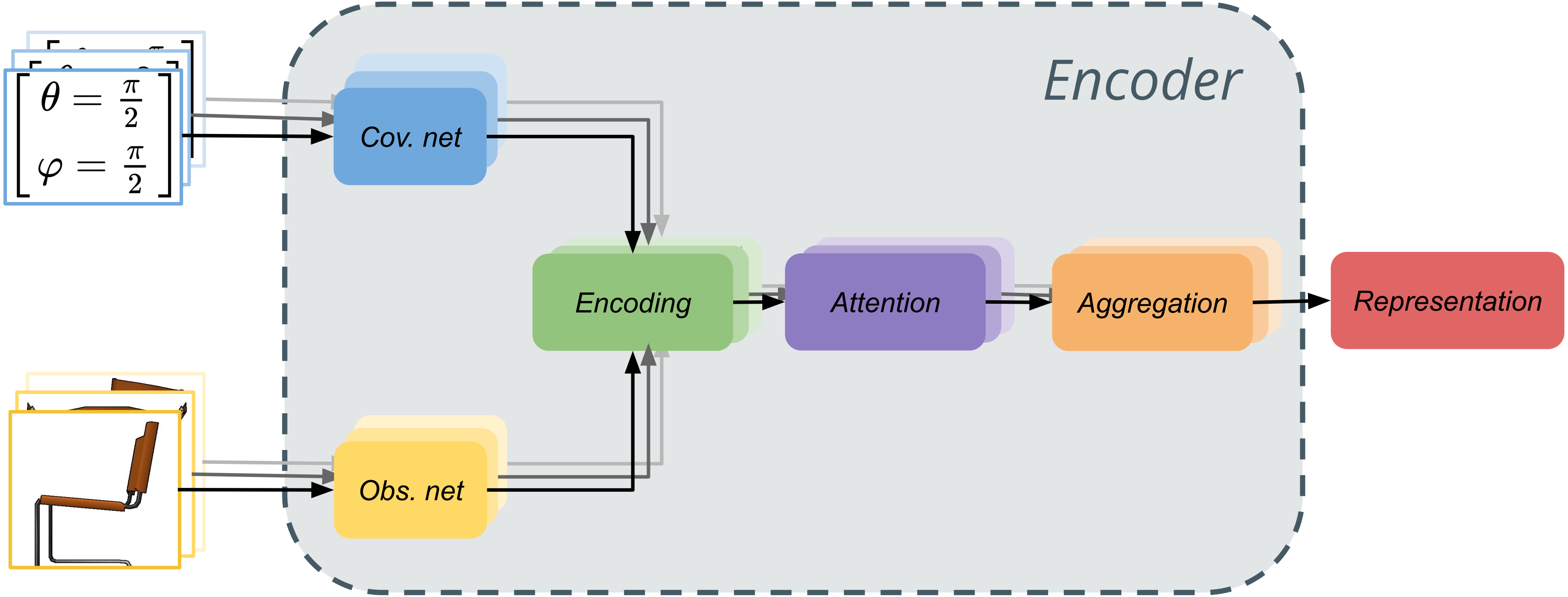
$\begin{bmatrix} \theta = \frac{\pi}{4} \\ \varphi = \frac{\pi}{4} \end{bmatrix}$, 

$\begin{bmatrix} \theta = \frac{\pi}{2} \\ \varphi = \frac{\pi}{2} \end{bmatrix}$, 

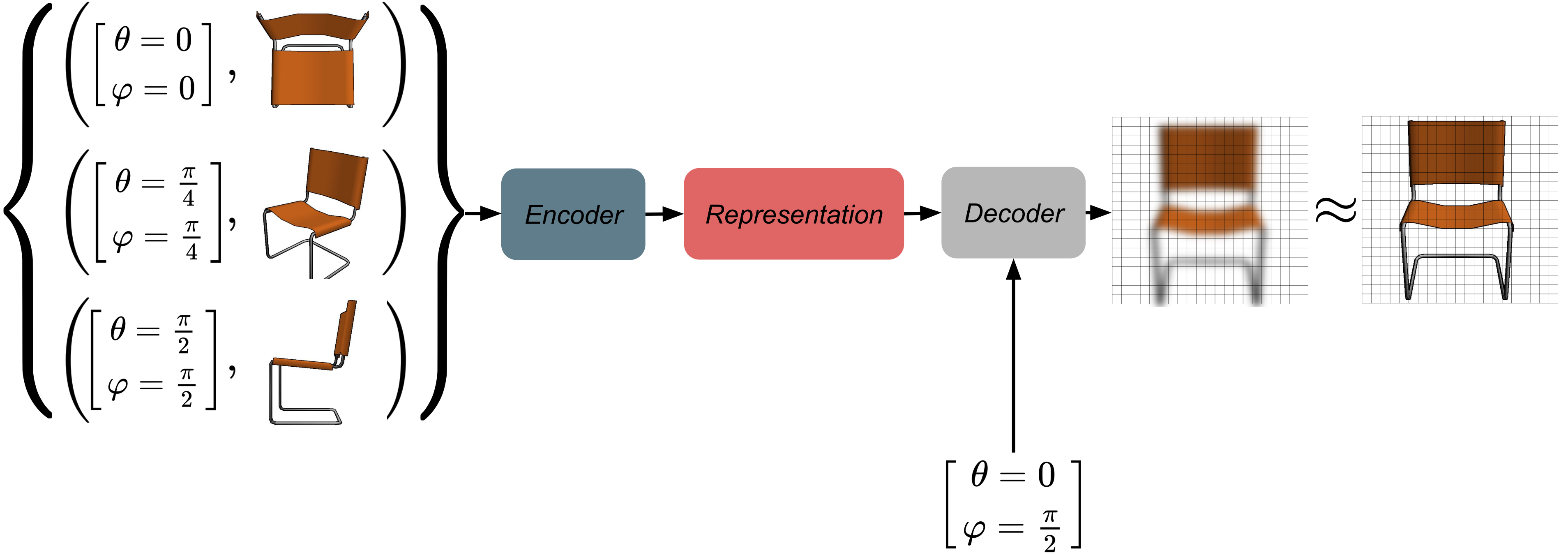
Representation

Downstream task

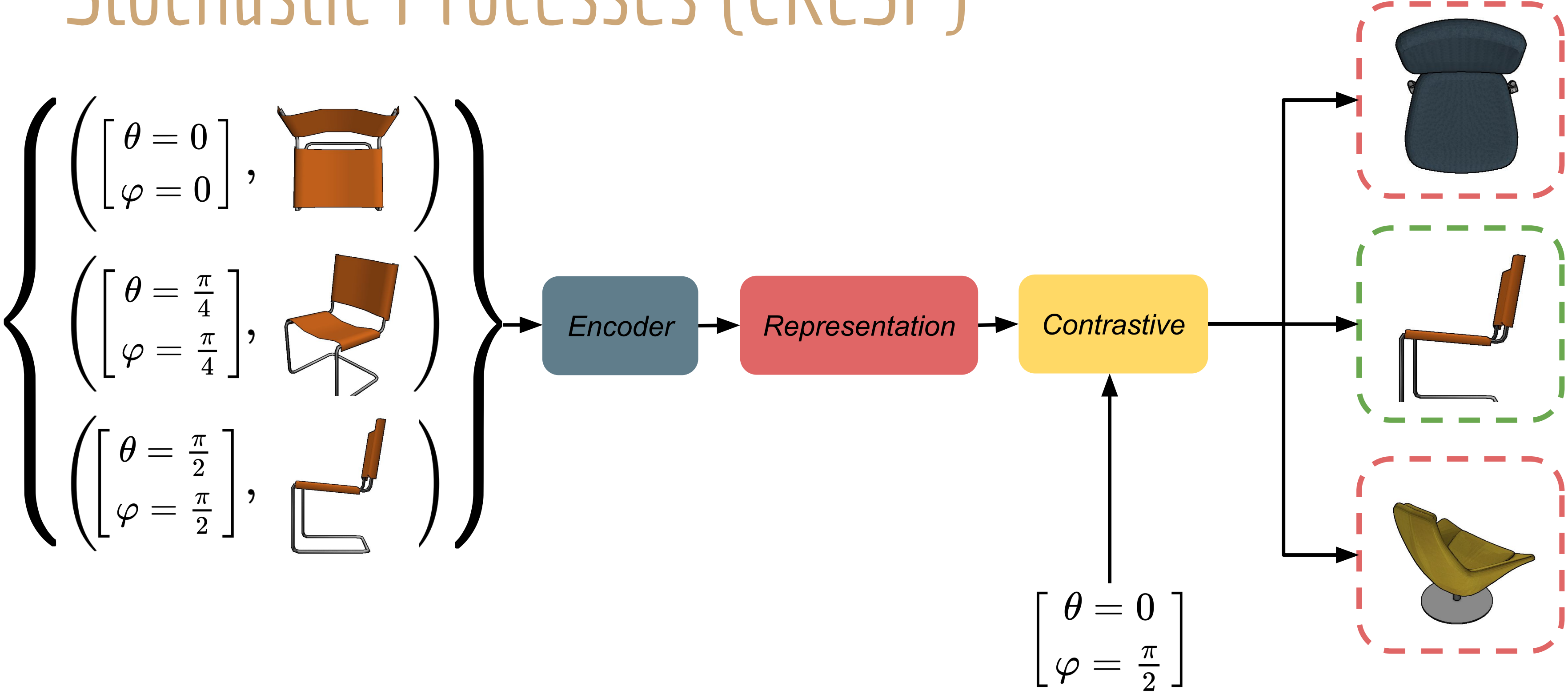
Representation learning



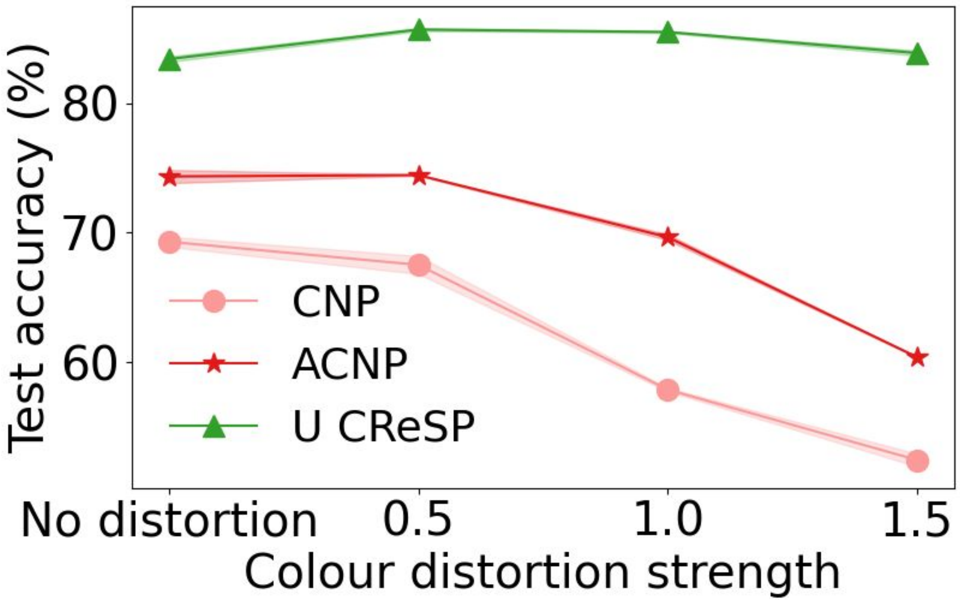
Reconstruction approach

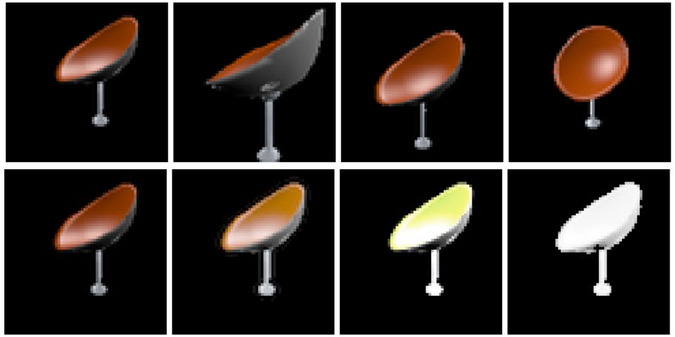


Contrastive Representations of Stochastic Processes (CReSP)



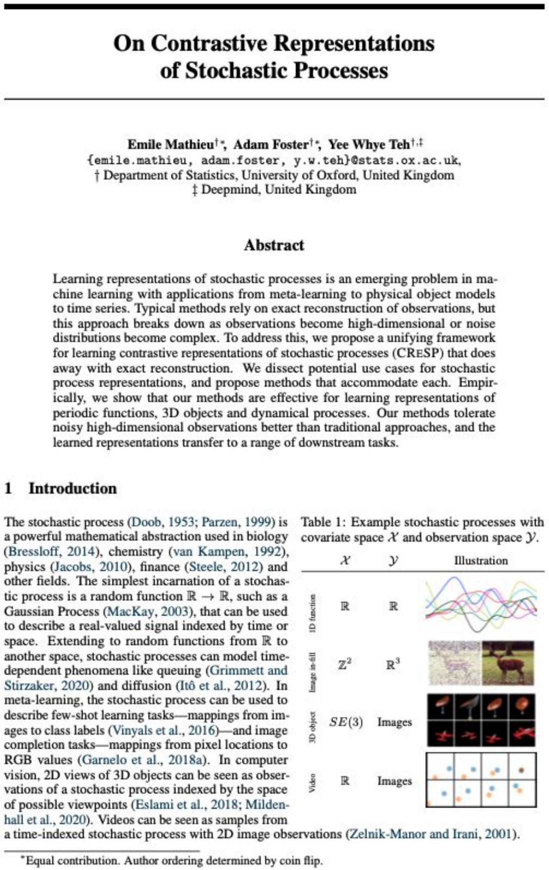
ShapeNet classification - effect of noise





Thank you for your attention!

Don't be shy and attend our poster session for a chat :)



On Contrastive Representations of Stochastic Processes

Emile Mathieu[†,*], **Adam Foster**[†,*], **Yee Whye Teh**[†,‡]
{emile.mathieu, adam.foster, y.w.teh}@stats.ox.ac.uk,
† Department of Statistics, University of Oxford, United Kingdom
‡ Deepmind, United Kingdom

Abstract

Learning representations of stochastic processes is an emerging problem in machine learning with applications from meta-learning to physical object models to time series. Typical methods rely on exact reconstruction of observations, but this approach breaks down as observations become high-dimensional or noise distributions become complex. To address this, we propose a unifying framework for learning contrastive representations of stochastic processes (CRESP) that does away with exact reconstruction. We dissect potential use cases for stochastic process representations, and propose methods that accommodate each. Empirically, we show that our methods are effective for learning representations of periodic functions, 3D objects and dynamical processes. Our methods tolerate noisy high-dimensional observations better than traditional approaches, and the learned representations transfer to a range of downstream tasks.

1 Introduction

The stochastic process (Doob, 1953; Parzen, 1999) is a powerful mathematical abstraction used in biology (Bressloff, 2014), chemistry (van Kampen, 1992), physics (Jacobs, 2010), finance (Steele, 2012) and other fields. The simplest incarnation of a stochastic process is a random function $\mathbb{R} \to \mathbb{R}$, such as a Gaussian Process (MacKay, 2003), that can be used to describe a real-valued signal indexed by time or space. Extending to random functions from $\mathbb{R}$ to another space, stochastic processes can model time-dependent phenomena like queuing (Grimmett and Stirzaker, 2020) and diffusion (Itô et al., 2012). In meta-learning, the stochastic process can be used to describe few-shot learning tasks—mappings from images to class labels (Vinyals et al., 2016)—and image completion tasks—mappings from pixel locations to RGB values (Garnelo et al., 2018a). In computer vision, 2D views of 3D objects can be seen as observations of a stochastic process indexed by the space of possible viewpoints (Eslami et al., 2018; Mildenhall et al., 2020). Videos can be seen as samples from a time-indexed stochastic process with 2D image observations (Zelnik-Manor and Irani, 2001).

Table 1: Example stochastic processes with covariate space $\mathcal{X}$ and observation space $\mathcal{Y}$.

	$\mathcal{X}$	$\mathcal{Y}$	Illustration
1D function	$\mathbb{R}$	$\mathbb{R}$	
Image in-fill	$\mathbb{Z}^2$	$\mathbb{R}^3$	
3D object	$SE(3)$	Images	
Video	$\mathbb{R}$	Images	

[*]Equal contribution. Author ordering determined by coin flip.