# Transfer of Status Report

Adam Kosiorek[1]

***Abstract*— Abstract goes here.**

## I. Introduction

We spend our lives roaming through the space-time continuum. Our senses have evolved to make use of the temporal dependencies omnipresent in real-world data. And yet the majority of machine learning (ML) algorithms either do not use temporal dependencies at all or rely on features extracted by models which do not take them into account. I am interested in and will focus on using neural networks for probabilistic time-series modelling, with the emphasis on unsupervised and self-supervised learning and the connection between learning and interacting with the environment. I am going to argue that time dependencies in data and the interaction of an agent with its environment are enough to create a powerful signal for self-supervised learning. My work as a PhD student at Oxford started with the problem of single object tracking in videos, which resulted in a successful NIPS 2017 submission (Kosiorek, Bewley, and Posner, 2017). This project gave me an opportunity to explore learning in the presence of temporal dependencies and to explore the concept of self-supervision: how to make the system learn better without using any additional external ( e. g., ground-truth) information? The rest of this paper is structured as follows: Section II covers prior work related to the areas in question. Specifically, I summarise the tasks of sequence prediction, predictive coding, variants of unsupervised learning and present a number of relevant approaches. In the section III, I describe the work on object tracking and how it ties with my interests and the planned future work on structured unsupervised learning for videos, predictive coding and model-based reinforcement learning. Section IV describes my future research plans, related risks and expected outcomes. Section V concludes this work.

## II. Related Work

### A. Unsupervised Learning via Generative Modelling

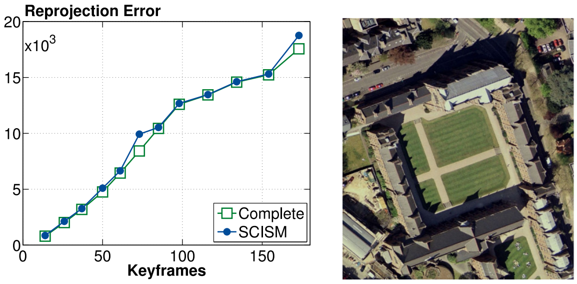
While data in general is abundant and cheap, data for supervised learning is often expensive and time-consuming to gather. The majority of ML algorithms require relatively large amounts of labelled training data. One of the explanation states that they start learning without any prior knowledge of the world . This is in stark contrast to humans, who not only have a vast amount of knowledge about the world, but also expand it continuously and without any supervision (Friston, 2009). One alternative is to perform generative modelling of the probability distribution $p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{z})\,\mathrm{d}\mathbf{z}$ of observations $\mathbf{x}$ in terms of some latent variables $\mathbf{z}$. The latent variables *explain* the observations and can make the joint distribution $p(\mathbf{x}, \mathbf{z})$ tractable even in the case of an intractable marginal distribution. The latent encoding can be used in upstream tasks e. g., for transfer or semi-supervised learning (Pan and Yang, 2010). Hinton, Osindero, and Teh, 2006 introduced Deep Belief Networks (DBN) which explain the observations in terms of Bernoulli latent variables. Alternatively, we can approximate the true data distribution by deriving the evidence lower bound (ELBO) on the log probability of the data, which results in variational autoencoders (VAE) (Kingma and Welling, 2013; Rezende, Mohamed, and Wierstra, 2014). VAEs are much more flexible than DBNs as they allow latent variables from arbitrary probability distribution functions (pdf) and can be trained end-to-end with off-the-shelf gradient-based methods. These approaches are primarily suited to modelling datasets of independent and identically distributed (*i.i.d.*) points.

Add reference(s).

### B. Sequence Modelling

Traditional approaches to sequence modelling often consider inference of latent variables, e. g., linear dynamical systems or hidden markov models, that explain the data (Bishop, 2006). They often require dynamics of the system to be known and often have too little capacity to model complex and high-dimensional real-world data. Neural networks, on the other hand, can learn both features and state dynamics from data and they can approximate functions of arbitrary complexity with arbitrary precision. Even early works on the topic demonstrated how useful neural networks are for prediction of chaotic time-series (Lapedes and Farber, 1988). Since then, neural networks have been successfully applied for sequence classification and prediction in different domains: written natural language, speech and audio, motion capture data or brain waves (Längkvist, Karlsson, and Loutfi, 2014). Unsupervised learning can be also done as sequence prediction, where the task is to predict the observation at time $t+1$ given a sequence of observations $\mathbf{x}_{1:t}$ up to time $t$. This task is flexible in that it admits many different model types, including Gaussian processes, support vector machines or feed-forward neural networks, although models which can explicitly use temporal structure of data such as Gaussian process dynamic models (GPDM; Wang, Fleet, and Hertzmann, 2008) or recurrent neural networks (RNN) tend to perform better. Recently, sequential counterparts of VAEs have been proposed, which allow efficient generative modelling of sequences (Fabius and Amersfoort, 2015; Bayer and Osendorfer, 2015; Karl et al., 2017).

[1] Oxford Robotics Institute, Dept of Engineering Science, University of Oxford.



(a) (b)

Fig. 1: Two figures wrapped in a table: (a) a graph and (b) a college.

### C. *Predictive Coding*

### D. *Learning of Abstract Ideas.*

## III. What I've done

We developed a biologically inspired single object tracker, which tracks objects in videos used a hierarchy of attention mechanisms that mimic the human visual cortex.

Even though it is fully supervised, we did it as a test bench for exploring the potential of self-supervision with temporal data. Attention mechanisms used in this work allowed us to increase computational efficiency of the approach but also to introduce learning cues that would be impossible to introduce otherwise. Specifically, we used self-supervision in terms of spatial masks for learning the foreground/background segmentation of the attention glimpse. It is self-supervision in the sense that we are not using any new information to introduce that loss; because the location of the object in the attention glimpse depends on the model and its parameters, it is different in every pass of the training algorithm, which is similar to augmenting the data. We don't use any new information since the bounding box information that we used was used previously to apply the bounding box loss.

## IV. Research Proposal

Things to write about:i

- Predictive Coding: next-frame prediction with VAEs, it's normalisation behaviour, inner feedback loop for corrections
- Model-based RL: sample efficiency, using a policy to improve learning speed of the perception module, learn a policy in the absence of a goal
- Unsupervised object tracking & detection

## V. Conclusions

Conclusions go here.

## Appendix

General theme: Representation learning for sequential data.

I'm interested in:

- timeseries
- unsupervised learning
- predictive coding
- learning abstract concept and ideas

What I'd like to do is:

I'd like to leverage unsupervised learning for time-series to learn abstract concepts that describe the world, and more importantly, its evolution. I would like to be able to:

- predict how the world evolves
- find out whether doing so in a probabilistic way has any benefits over deterministic approaches, e.g. is multi-modality of a probabilistic solution helpful
- see how imposing structure on the predictions affects representation learning, e.g. air-style unsupervised object tracking; using a policy to either minimise or maximise surprise to improve learning speed, predictive accuracy or both

## References

Bayer, Justin and Christian Osendorfer (2015). "Learning Stochastic Recurrent Networks". In: *Iclr*, pp. 1–9. arXiv: 1411.7610.

Bishop, Christopher M. (2006). *Pattern recognition and machine learning*. Springer, p. 738.

Fabius, Otto and Joost R. van Amersfoort (2015). "Variational Recurrent Auto-Encoders". In: *Iclr* 2013, pp. 1–5. arXiv: 1412.6581.

Friston, Karl (2009). "The free-energy principle: a rough guide to the brain?" In: *Trends in Cognitive Sciences* 13.7, pp. 293–301.

Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh (2006). "A Fast Learning Algorithm for Deep Belief Nets". In: *Neural Computation* 18.7, pp. 1527–1554.

Karl, Maximilian et al. (2017). *Deep Variational Bayes Filters*. arXiv: 1703.03129.

Kingma, Diederik P and Max Welling (2013). "Auto-Encoding Variational Bayes". In: arXiv: 1312.6114.

Kosiorek, Adam R., Alex Bewley, and Ingmar Posner (2017). "Hierarchical Attentive Recurrent Tracking". In: *NIPS*. arXiv: 1706.09262.

Längkvist, Martin, Lars Karlsson, and Amy Loutfi (2014). "A review of unsupervised feature learning and deep learning for time-series modeling". In: *Pattern Recognition Letters* 42, pp. 11–24.

Lapedes, AS and RM Farber (1988). "How neural nets work". In: *Neural information processing systems*.

Pan, Sinno Jialin and Qiang Yang (2010). *A survey on transfer learning*. arXiv: PAI.

Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra (2014). "Stochastic Backpropagation and Approximate Inference in Deep Generative Models". In: arXiv: 1401.4082.

Wang, Jack M., David J. Fleet, and Aaron Hertzmann (2008). "Gaussian process dynamical models for human motion". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2, pp. 283–298.