Linear Transformations in Autoencoder Latent Space Predict Time Translations in Active Matter System

Enrique Amaya*
Caltech
Pasadena, CA 91125
eamaya@caltech.edu

Shahriar Shadkhoo
Caltech
Pasadena, CA 91125
shahriar@caltech.edu

Dominik Schildknecht
Caltech
Pasadena, CA 91125
dominik.schildknecht@gmail.com

Matt Thomson
Caltech
Pasadena, CA 91125
mthomson@caltech.edu

Abstract

Machine Learning (ML) approaches are promising for deriving dynamical predictions of physical systems from data. ML approaches are relevant in active matter, a field that spans scales and studies dynamics of far-from-equilibrium systems where there are significant challenges in predicting macroscopic behavior from microscopic interactions of active particles. A major challenge in applying ML to active systems is encoding a continuous representation of time within a neural network. In this work, we develop a framework for predicting the dynamics of active networks of protein filaments and motors by combining a low-dimensional latent representation inferred through an autoencoder with a linear shift neural network that encodes time translation as a linear transformation within the latent space. Our method enables predicting the contraction and boundary deformations of active networks with various geometries. Although our method is trained to predict 20 time steps into the future, it can generalize to periods of 60 time steps and recapitulate the past 30 frames of a single given observation with less than 10% error. Finally, we derive an approximate analytic expression for the linear transformation in the latent space that captures the dynamics. Broadly, our study reveals that neural networks are powerful for forecasting the behavior of active matter systems in the complete absence of knowledge of the microscopic dynamics.

1 Introduction

Active matter systems are made of many autonomous agents that consume energy to generate motion. Examples of living and non-living active systems across scales include the cellular cytoskeleton, bacterial colonies, schools of fish, and drone swarms, to name a few [1]. To engineer active systems at all scales and in different settings, we need to predict how active systems self-organize into collective spatiotemporal patterns [2]. A fundamental limitation for mechanistic models of active matter arises from its out-of-equilibrium nature, which breaks detailed balance, and time-reversal symmetry [3]. Furthermore, the vast diversity of active agents makes it difficult to find a comprehensive theoretical framework of active matter.

*The code to reproduce our main findings can be found on the following Github repository: https://github.com/e-amaya/TCAE

NeurIPS 2021 AI for Science Workshop.

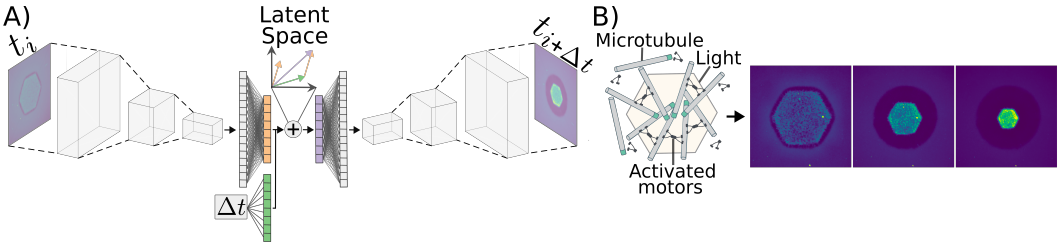

Figure 1: A) Schematic of the proposed DNN: The latent representation of the input image is linearly combined with an encoding of time translation. This combined representation is decoded to arrive at the output prediction. B) Active matter system used in this study: Microtubule filaments and motors cross-link under illumination, and three snapshots of the contracting network at t=50, 100, and 150.

An alternative approach to forecast the dynamics of active systems that bears great potential is Machine Learning. Specifically, Deep Neural Networks (DNNs) offer data-driven model-free predictions of complex systems [4, 5]. The multilayered hierarchical structure of DNNs provides an automated way to identify the most relevant degrees of freedom of a system by iteratively coarse-graining structured data without any prior knowledge of the underlying complexity [6]. However, the designing principles of DNNs' architectures that model active systems, as well as the physical interpretations of the representations learned by DNNs, are still largely unexplored.

In this work, we use DNNs to investigate the dynamics of an optically-controlled active matter system composed of microtubule filaments and light-switchable kinesin motor proteins [7]. In the active system under study, light patterns can be used to determine when and where motors reorganize microtubules to create non-equilibrium structures. We predict the dynamics of macroscopic patterns that emerge from light-induced microtubule reorganization using what we refer to as the Translational Convolutional Autoencoder (TCAE). TCAE learns a low-dimensional representation of microscopy data, constrained by a learned translation function. The translation function makes it possible to map input time differences to translation vectors in latent space. By feeding an initial image observation and varying scalar time differences to the TCAE, we can evolve the single observations through time and achieve low-error predictions. Although our method is trained to predict a fixed future time interval, it generalizes to prolongated periods. Surprisingly it can also infer the past dynamics if negative scalar time differences are provided to the TCAE. We hypothesize that the physical properties of the active system are embedded in the latent space our DNN constructs. This work shows the potential DNNs have for extracting physical properties and predicting the behavior of complex non-linear out-of-equilibrium systems.

2 Methodology

2.1 Data Description

The data consists of 16 active matter experiments as described in [8]. The light patterns include 4 different shapes (circle, hexagon, rectangle, and triangle), with four initial sizes (900, 750, 600, 450 μm). The raw videos have a resolution of 2048×2048, and a total video length of 170 frames. We down-sampled the videos to a resolution of 112×112 pixels, and normalized to have a range of values between 0 and 1. We chose 15 videos for training and left 1 for test evaluation.

2.2 Approach

We extend the input reconstruction capabilities of the Convolutional Autoencoder [9] for time propagation prediction. Our proposed method takes as input the image at time t and a change in time Δt to output a prediction for the image at time $t + \Delta t$. First, an encoder function $\mathcal{E} : \mathbb{R}^{m\times m} \mapsto \mathbb{R}^n$ maps an image $\mathbf{X}_t^s$ to a latent representation $\mathbf{z}_t^s$, where s indexes the list of movies, and t is time. Then, a linear time translation operator translates the image embedding $\mathbf{z}_t^s$ by a learned time-translation $\mathbf{\Phi}$, scaled by time shift Δt; namely $\mathbf{z}_t^s \mapsto \mathbf{z}_t^s + \mathbf{\Phi}\Delta t$. Finally, the decoder $\mathcal{D} : \mathbb{R}^n \mapsto \mathbb{R}^{m\times m}$ takes the translated image embedding and predicts the frame at time $t + \Delta t$; i.e. $\mathcal{D}(\mathbf{z}_t^s + \mathbf{\Phi}\Delta t) = \widetilde{\mathbf{X}}_{t+\Delta t}^s$. We

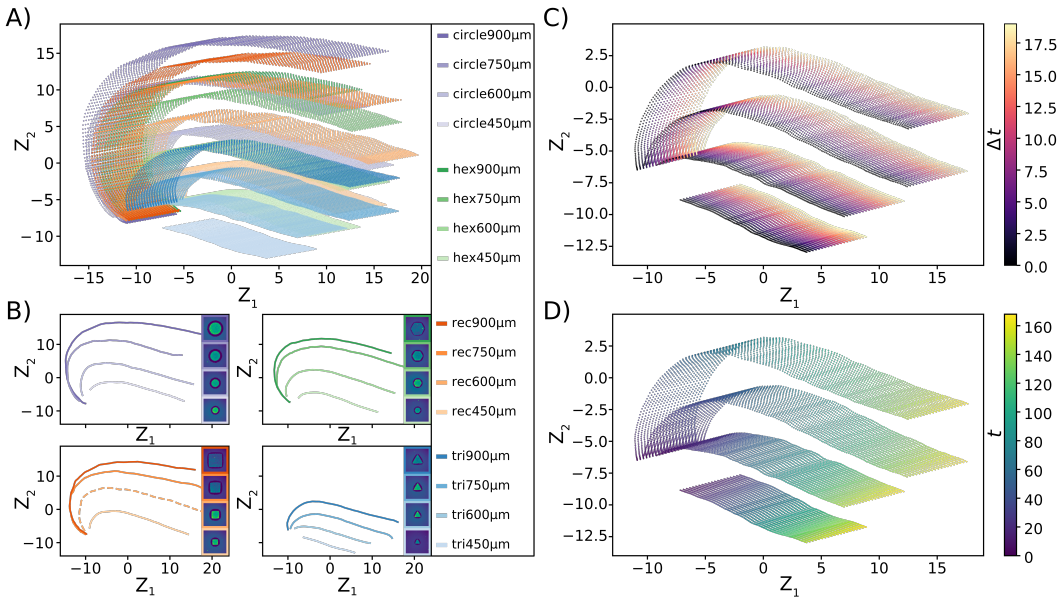

Figure 2: 2D visualization of TCAE latent space. A) Embeddings of all the images in the dataset and their corresponding translations using a Δt from 0 to 20, colored by light pattern. B) Embeddings without translations grouped by geometry. The dashed line is the test video. C,D) Embeddings of triangular geometries and their translations colored by Δt and time, respectively.

restrict our training task to predict for $\Delta t = \{1, \ldots, 20\}$ frames into the future. With training dataset $\mathcal{X} = \{\mathbf{X}_t^s\}\big|_{t\in[1,170]}^{s\in[1,15]}$; and a dissimilarity (i.e. loss) function $\mathscr{L}$, we find the network operators:

$$\underset{\mathcal{E},\mathcal{D},\mathbf{\Phi}}{\arg\min} \sum_{s} \sum_{t=1}^{150} \sum_{\Delta t=0}^{20} \mathscr{L}\left(\widetilde{\mathbf{X}}_{t+\Delta t}^{s}, \mathbf{X}_{t+\Delta t}^{s}\right)$$

We choose $\mathscr{L}$ to be a summation over all pixels of the pixel-wise binary cross-entropy function: $\mathscr{L}(\widetilde{\mathbf{X}}_t^s, \mathbf{X}_t^s) = \sum_{i,j} H(p_{ij}, \tilde{p}_{ij})$ where (i, j) labels the pixels, and $p_{ij}, \tilde{p}_{ij}$ are the values of pixels in $\mathbf{X}_t^s$ and $\widetilde{\mathbf{X}}_t^s$, respectively. The cross entropy reads $H(p, \tilde{p}) = -\left(p \log(\tilde{p}) + (1-p)\log(1-\tilde{p})\right)$.

3 Results

3.1 Interpretation of Time Translation in the Latent Space

Calculating the mismatch between the ground truth and predicted images exhibits up to $\sim 95\%$ accuracy of prediction; see Fig. (3). Such a high accuracy is suggestive of a dynamical mechanism that can be recapitulated by uniform translation of the latent space. In the 2D latent space spanned by coordinates (z_1, z_2). We expand the argument of the function $\mathcal{D}$, i.e. the latent space coordinates $\mathbf{z} + \mathbf{\Phi}\Delta t$; and demand that $\mathscr{L}\left(\mathcal{D}(\mathbf{z}(t) + \mathbf{\Phi}\Delta t) - \mathcal{D}(\mathbf{z}(t + \Delta t))\right)$ is minimized. Note that cross entropy is always non-negative; hence zero lower bound. For the first term we have: $\mathcal{D}(\mathbf{z}(t) + \mathbf{\Phi}\Delta t) = \mathcal{D}(\mathbf{z}(t)) + (\mathbf{\Phi}\Delta t)\nabla_{\mathbf{z}(t)}\mathcal{D}(\mathbf{z}(t)) + \mathcal{O}(\mathbf{\Phi}^2\Delta t^2)$, and for the second term: $\mathcal{D}(\mathbf{z}(t + \Delta t)) = \mathcal{D}(\mathbf{z}(t)) + (\partial_t \mathbf{z}(t)\Delta t)\nabla_{\mathbf{z}(t)}\mathcal{D}(\mathbf{z}(t)) + \mathcal{O}(\partial_t \mathbf{z}(t)^2\Delta t^2)$. Since the loss function is minimized over the entire data set, we claim that $\langle \mathbf{\Phi} - \partial_t \mathbf{z}(t)\rangle_{s,t,\Delta t} = 0$; thus $\mathbf{\Phi} = \langle \partial_t \mathbf{z}(t)\rangle_{s,t,\Delta t}$. Here $\langle \bullet \rangle_{s,t,\Delta t}$ denotes averaging over all the data sets, time points, and time translations. Therefore, to the first order approximation, $\mathbf{\Phi}$ can be thought of as the average over $\partial_t \mathbf{z}(t)$, namely the slopes of the lines connecting the points $\mathbf{z}(t)$ and $\mathbf{z}(t + \Delta t)$ for all the points along the $\Delta t = 0$ curves. The average over t, is equivalent to connecting the two ends of the curves: $\overline{\Delta \mathbf{z}} = \mathbf{z}(t_{\max}) - \mathbf{z}(t_{\min})$. Averaging these vectors over all datasets we obtain the translation $\mathbf{\Phi}$ that minimizes the loss function.

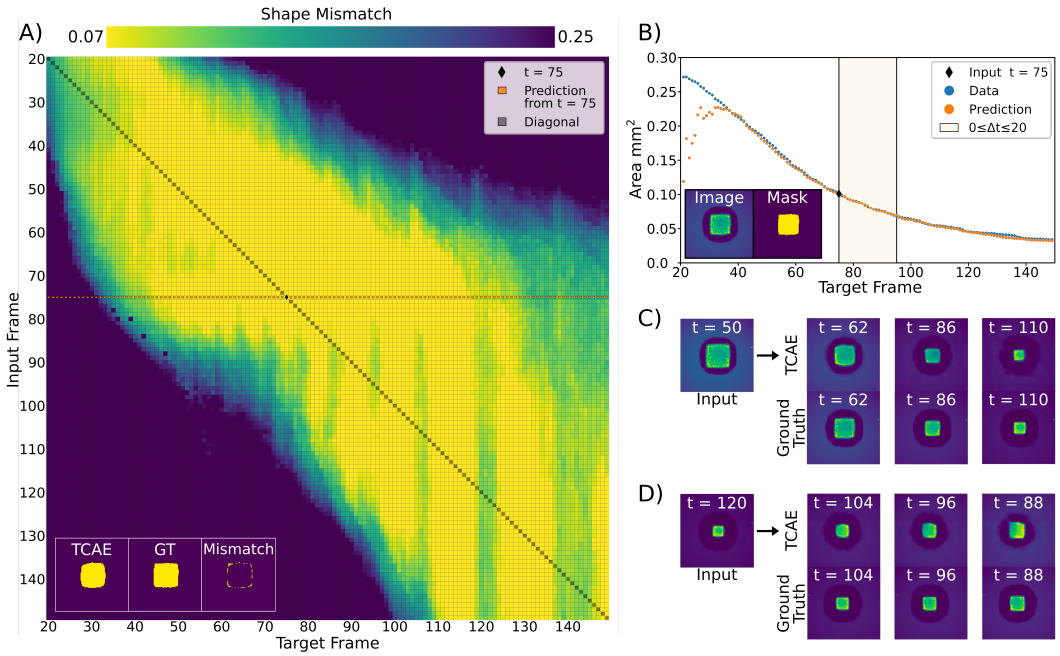

Figure 3: TCAE performance on the test set. A) Heatmap of shape mismatch in whole video prediction. Starting with one frame as input, we predicted the full video by varying Δt. Then we computed the shape mismatch between the predicted and ground truth videos. The error value for the frame used as input lies on the main diagonal (reconstruction error), and the rest of the error values for the predictions of the past and future frames lie on the row containing the input frame. B) Microtubule network size evolution: We predicted the past and future dynamics of the image at $t = 75$, then we used a segmentation algorithm to quantify the area of the microtubule network on both prediction and ground truth data. C,D) Perceptual comparison between the prediction of future and past time points with ground truth data.

Finally we shall mention that the qualitative similarity between the curves corresponding to different geometries suggests that they would possibly collapse onto a single curve upon performing rotation and scaling transformations in the latent space. Therefore, while the latent space carries information about the geometry and size, its 2D embedding predominantly captures the features that are independent of specific shapes, namely the contraction of the networks.

3.2 Active Matter Dynamics Prediction Analysis

An example of predictions on the test dataset made by the TCAE is shown in Fig. (3C). Perceptually, the predictions made by taking the test frame at $t = 50$ as starting point agree with unseen experimental data up to $\Delta t = 60$. We can observe the contraction and shape preservation of the microtubule network. Although in the training phase, we constrained Δt to take positive values between 0 and 20, the time embedding $\mathbf{\Phi}$ can be scaled by negative values, namely time translation in the opposite direction. Given that we can move in either time direction in the latent space, TCAE can perform whole video predictions from a single given frame. An example is shown in Fig. (3B). We see that TCAE predictions follow closely the microtubule network area dynamics observed in experimental data at least for 90 percent of the predicted frames from $t = 33$ to $t = 150$.

Finally, we define a simple metric to evaluate the shape mismatch between the ground truth and the predicted segmented (binary) images, $\mathbf{X}_{\text{TCAE}}$ and $\mathbf{X}_{\text{GT}}$. The normalized mismatch index ε is defined as deviation of the overlapping area from unity. The overlapping area is in turn defined as the area of the bit-wise product of the two binary images, normalized by the area of the ground truth image. Using the double-sum notation $A : B = \sum_{i,j} A_{ij} B_{ij}$, we get: $\varepsilon = 1 - (\mathbf{X}_{\text{TCAE}} : \mathbf{X}_{\text{GT}})/(\mathbf{X}_{\text{GT}} : \mathbf{X}_{\text{GT}})$. This metric is used to quantify the error of whole video prediction across all frames. Results are shown in Fig. (3A). For 90% of frames, it is possible to predict on average 38.5 frames into the future with less

than 10% error. A maximum of 57 future frames (with less than 10% error) can be predicted if we fix $t = 66\,,\ 67$. If we consider predicting past dynamics, we can take 90% of frames to predict on average 25.6 frames with less than 10% error. If we set $t = 79\,,\ 80$, we can reach a maximum of 36 frames in the past (with less than 10% error).

4 Summary

We propose TCAE, an autoencoder-based framework that incorporates a linear shift neural network that encodes time translation as a linear transformation within the latent space. Using our method, we construct a latent representation of an active matter system that predominantly captures the contraction of its microtubule networks. By leveraging the constructed latent representation, we can make whole video predictions from single observations. We obtained high accuracy (less than 10% error) in our predictions when compared to unseen ground truth data.

Acknowledgments and Disclosure of Funding

The authors are grateful to Emanuel Flores Bautista and Guruprasad Raghavan for scientific discussions. We acknowledge funding through the Packard Foundation, the Moore Foundation, and the Heritage medical research institute. The authors declare no competing financial interests.

References

[1] Gerhard Gompper, Roland G Winkler, Thomas Speck, Alexandre Solon, Cesare Nardini, Fernando Peruani, Hartmut Löwen, Ramin Golestanian, U Benjamin Kaupp, Luis Alvarez, et al. The 2020 motile active matter roadmap. *Journal of Physics: Condensed Matter*, 32(19):193001, 2020.

[2] Daniel Needleman and Zvonimir Dogic. Active matter at the interface between materials science and cell biology. *Nature reviews materials*, 2(9):1–14, 2017.

[3] M Reza Shaebani, Adam Wysocki, Roland G Winkler, Gerhard Gompper, and Heiko Rieger. Computational models for active matter. *Nature Reviews Physics*, 2(4):181–199, 2020.

[4] Jaideep Pathak, Brian Hunt, Michelle Girvan, Zhixin Lu, and Edward Ott. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Physical review letters*, 120(2):024102, 2018.

[5] Julia Ling, Andrew Kurzawski, and Jeremy Templeton. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807:155–166, 2016.

[6] Pankaj Mehta and David J Schwab. An exact mapping between the variational renormalization group and deep learning. *arXiv preprint arXiv:1410.3831*, 2014.

[7] Tyler D Ross, Heun Jin Lee, Zijie Qu, Rachel A Banks, Rob Phillips, and Matt Thomson. Controlling organization and forces in active matter through optically defined boundaries. *Nature*, 572(7768):224–229, 2019.

[8] Zijie Qu, Jialong Jiang, Heun Jin Lee, Rob Phillips, Shahriar Shadkhoo, and Matt Thomson. Programming boundary deformation patterns in active networks. *arXiv preprint arXiv:2101.08464*, 2021.

[9] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *International conference on artificial neural networks*, pages 52–59. Springer, 2011.