Deep anomaly detection using self-supervised learning: application to time series of cellular data

Romain Bailly, Marielle Malfante, Cédric Allier, Lamya Ghenim, Jérôme I. Mars

3rd International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAI' 2020), 17-19 November 2021, Porto, Portugal

Special session Topic: Biomedical signal processing

Deep anomaly detection using self-supervised learning: application to time series of cellular data

Romain Bailly [1, 4], Marielle Malfante [1], Cédric Allier [2], Lamya Ghenim [3], and Jérôme Mars [4]

[1] Univ. Grenoble Alpes, CEA, List, F-38000 Grenoble, France

[2] Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France

[3] Univ. Grenoble Alpes, CNRS, CRA, INSERM, IRIG, F-38000 Grenoble, France

[4] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France

Email: [1, 2, 3] firstname.name@cea.fr , [4] fistname.name@gipsa-lab.grenoble-inp.fr

Summary: We present a deep **self-supervised** method for **anomaly detection** on **time series**. We apply this methodology to detect anomalies from **cellular time series**. In particular, this study focuses on cell dry mass**,** obtained in the context of lens-free microscopy.

The method we propose is an innovative two-step pipeline using self-supervised learning. As a first step, a representation of the time series is learned thanks to a 1D-convolutional neural network **without any labels**. Then, the learned representation is used to feed a threshold anomaly detector. This new self-supervised learning method is tested on an unlabelled dataset of 9100 time series of dry mass and succeeded in detecting abnormal time series with a **precision** of **96.6%**.

Keywords: Self-supervised learning, 1D-CNN, Anomaly detection, Cellular anomaly, Time series, Lens-free microscopy

1. Introduction

Lens-free microscopy is a recently developed imaging technique [1] overcoming some limitations of classical microscopy. Typically, it allows the rendering of thousands of cells in a single frame with a much less cumbersome device. [2] proposes to analyse sequences of images, from which a dataset of time series of cells' dry mass is built.

The dry mass of a cell, measured in picograms (pg), is related to its metabolic and structural functions. Amongst the thousands of cells in a Petri dish, it may happen that some cells deviate from their typical behaviour, thus influencing their dry mass. It has been shown that cells deviating from healthy trajectories can further drive tissues toward diseases [3]. Detecting abnormal cells automatically is thus crucial.

We propose an innovative method for automatically detecting abnormal cells using their dry mass. Using methods that do not need any manually labelled data is of particular interest especially in the case of time series processing. Indeed, while expert have a good understanding of what a normal cell behaviour is, there is no *a priori* knowledge of what an abnormal cell behaviour is. Working without labels is therefore interesting especially when the datasets are complex or not yet fully understood.

The proposed approach is in two steps: first, a representation of the time series is trained using self-supervised learning. In a second step, an anomaly detection block is used over the learned representation to determine if a cell is abnormal. This self-supervised method benefits from the **representation power** of **deep learning without** the usual **labelling constraint**.

2. Related works

2.1. Anomaly detection

Anomaly detection is a broad field of research focusing on the detection of abnormal patterns within a given set of data. We focus on anomaly detection on time series as presented in [4]. In particular, prediction-based anomaly detection techniques, which tries to predict the future of, time series. An outlier score [4] is computed between the prediction and the true value of the time series to determine if it is abnormal.

Multiple predictors can be used such as support vector regression [5], multilayer perceptrons (MLP) [6] or mixture transition distribution [7]. While [8] proposes a vector ARIMA to identify outlier points, other methods focus on discovering multiple outliers such as Gibbs sampling and block interpolation [9] or re-weighted maximum likelihood [10].

2.2. 1D-convolutional neural networks

Convolutional neural networks (CNNs) are mainly known for their success in computer vision with AlexNet [11], VGG16 [12] or ResNet [13], since the emergence of huge labelled datasets such as CIFAR100 [14] or ImageNet [15].

Because of the state of the art performances for computer vision achieved by 2D-CNNs, the signal processing community started to renew interest in 1D-CNNs, in the past few years for a wide variety of applications. They range from healthcare with ECG classification [16], [17] to fault detection [18]–[23] including audio and speech recognition [24] and other

fields such as time series forecasting [25], or anomaly detection [26]. 1D-CNNs were introduced in the literature for the first time as Time Delay Neural Network (TDNN) in [27], [28].

2.3. Self-supervised learning

Self-supervised learning [29], [30] is a new training paradigm where **supervised methods** are used on an **unlabelled dataset**. The core idea is to **automatically** obtain a labelled dataset from the initially unlabelled dataset. A **pretext task** is associated to the **self-labelled** dataset and allows a supervised training of the neural network. [31] illustrates pretext tasks in computer vision: an input image is rotated $[0, 90, 180, 270]°$ and the neural network has to predict the rotation applied to the image. The network can only succeed if it has learned relevant visual features from within the images.

While a great deal of research exploiting pretext tasks can be found in the field of computer vision [32]–[36] very little of this work is related to time series processing, with the exception of some papers deeply linked to the temporality of the data. In [37], [38], a set of video images are given in a random order to the network that must order the frames. Finally, another time-related pretext task is presented in [39] where videos with modified playback speeds in range $[-5, +5]$ are given as inputs and the network must predict the playback speed.

3. Dataset

The acquisitions used in this study contain dry mass time series extracted from lens-free images of HeLa cells thanks to an upstream algorithm presented in [40]. A cell dry mass is a measure of how much the cell would weight if it had been deprived of its water. It is directly linked to the proteins content of the cell and is an indicator of its health.

Fig. 1. shows a normal cell behaviour on both the original images (**Fig. 1a-1d**) and the extracted dry mass time series in blue **Fig. 1e**. A typical track of dry mass contains a growing phase (**Fig. 1a** to **1c**) where the dry mass increases regularly and a division phase (**Fig. 1c** to **1d**) during which the mother cell is divided in two daughter cells of approximately equal sizes. The division appears as an abrupt decrease on the dry mass time series (between points *c* and *d* **Fig. 1e**).

The dataset is split into train, validation and test sets in a 80/10/10% distribution [41]. Each of those sub-dataset is augmented with window slicing [42]. Every full-length acquisitions is sliced into smaller ones. Every possible smaller time series are extracted from the full-length one *i.e.* there is a one-sample shift between two consecutive time series in the sub datasets.

4. Methods

4.1. Representation learning neural network

The neural network used to learn a representation of the time series is trained in a self-supervised framework. Self-supervision allows the model to learn a deep representation of the signal without any labels. It uses a pretext task, to learn this representation. In our application and in agreement with the experts, we chose the pretext task to be **time series prediction** as presented **Fig. 2**. In this study, the input vector length is set to 120 time steps and the label vector to 60 time steps.

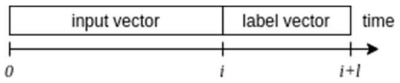


Fig. 2. Time series are split in an input vector of size i and label vector of size l.

A 1D-convolutional neural network architecture is used to capture the representation of the signal. The hyperparameter optimisation for the 1D-CNN representation learning neural network is presented section 4.2.

The neural network is trained using a Root Mean Squared Error (RMSE) loss eq. (1) between the true future of the time series and the predicted one with y_n the ground truth value at time step n and $\widehat{y_n}$ the prediction value at time step n. **Fig. 3.** describes the full anomaly detection pipeline, including the representation learning neural network.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \widehat{y_n})^2} \quad (1)$$

Neural networks in this study are trained on a single `NVIDIA Titan X` with a `batch size` of `32`, a `learning rate` of `0.001` and with `ADAM` optimizer.

4.2. Anomaly detection

The proposed method relies on a second anomaly detection block. Experimental results have shown that the use of a **threshold detector** over the prediction RMSE allow the model to detect abnormal cells. The threshold τ is computed following eq. (2) such that the metric values outside the 95% confidence interval of the metrics are flagged abnormal.

$$\tau = \mu_{test} \pm 2 \cdot \sigma_{test} \quad (2)$$

where μ_{test} and σ_{test} are respectively the mean and standard deviation of RMSEs over the test set. We assume the metric distribution over a dataset to be Gaussian.

4.3. Evaluation

The proposed method is designed to analyse **unlabelled datasets**. Therefore, it is not possible to fully annotate the dataset nor to compute classical precision/recall curves. We propose an evaluation method based on the annotation on solely positives detections, *i.e.* time series raised as anomalies. The precision is computed following equation (3). While the whole dataset cannot be annotated to compute the recall, we propose an evaluation of the recall $\hat{R}$ equation (4) by labelling a random 5% sample of the detected-normal cells (Negatives) to estimate the False Negative count.

$$P = \frac{TP}{TP + FP} \quad (3) \qquad \hat{R} = \frac{TP}{TP + \hat{FN}} \quad (4)$$

5. Results

5.1. Representation learning neural Network

The definition of the best architecture hyperparameter is achieved through an empirical study. Multiples neural networks are trained on the pretext prediction task. All the convolutional layers contain 64 filters and a pooling layer is always added every 3 convolutional layers. The features extracted from the convolutional layers are then fed in a dense layer of 128 neurons. **Table 1** shows the validation RMSE obtained for multiple architectures trained for this study.

Table 1. Architecture hyperparameters and their best MSE on the validation set. All the convolutional layers contain 64 filters and a pooling layer is always added every 3 convolutional layers.

# conv layer	kernel size	nb param	RMSE (pg)
3	3	155 708	87.726
3	8	196 988	94.053
3	16	263 036	92.035
3	32	395 132	85.677
3	64	659 324	84.608
3	120	1 121 660	83.089
5	3	82 108	97.824
5	16	295 932	93.178
5	64,32,16,8,4	282 620	93.706
9	3	303 548	80.941
9	**5**	**295 484**	**77.724**
9	8	393 980	85.643
12	3	266 876	81.877
12	5	357 116	86.096
12	6	402 236	88.993

The best 1D-convolutional neural network for the pretext task of prediction in the context of a cellular dry mass dataset is composed of 9 convolutional layers that contains 64 kernels of size 5. The RMSE on the test subset is computed to 76.62 pg.

Fig. 1e shows in blue the input given to the neural network, in dashed blue the ground truth to be predicted and in orange the network prediction. It shows on a specific example that the network is able to predict both a cell growing phase and a cell division.

5.2. Anomaly detection

The anomaly threshold on RMSE on the test set is computed to $\tau = 230.87$ pg thus raising **208** abnormal tracks. From a fully applicative point of view, the anomalies raised allowed domains experts to identify four possible causes of anomalies:

True positives **TP**:

1. Cellular Anomaly (CA): The cell grows in an unexpected way and should be analysed.
2. Measurement Anomaly (MA): the upstream dataset generation software was not able to track the cell properly.
3. Measurement Anomaly because of a cellular anomaly (CMA): because of a CA, an MA occurred.

False Positives **FP**:

4. Prediction Anomaly (PA): the neural network was not able to predict the cell future correctly whereas the cell is normal

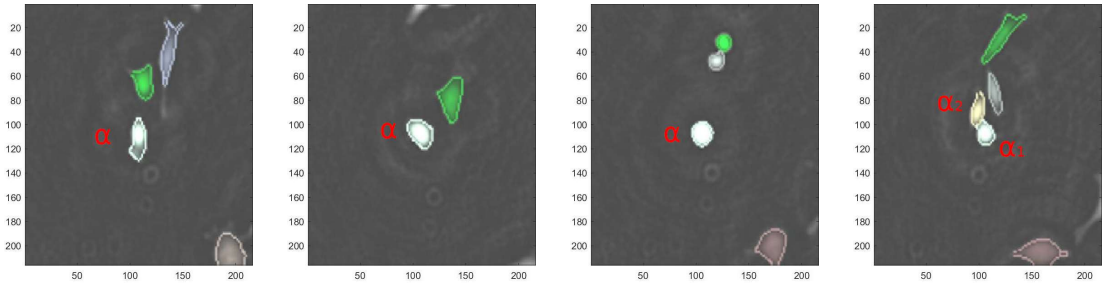
The category distribution of those abnormal cells is detailed in **Table 2**. Then, 31 false negatives were counted during the annotation of 447 samples (5%) of the cells predicted as normal. Anomaly detection has been achieved with a precision $P = 96.6\%$ and an estimated recall $\hat{R} = 24.5\%$.

Table 2. Expert classification of the anomalies raised.

Anomaly	CA	CMA	MA	PA
Ratio	40%	31%	26%	3%
	97%			3%

6. Conclusions

We propose an innovative two-step method for automatically detecting abnormal cells using their dry mass time series. This method focuses on unlabelled datasets thanks to the use of self-supervised learning. First, a representation of the time series is learned using a self-supervised 1D-convolutional neural network trained on a pretext prediction task. In a second step, the predicted dry mass value is compared to the ground truth. An anomaly is raised if the RMSE is above a given threshold. A precision of 96.6% and an estimated recall of 24.4% are achieved.

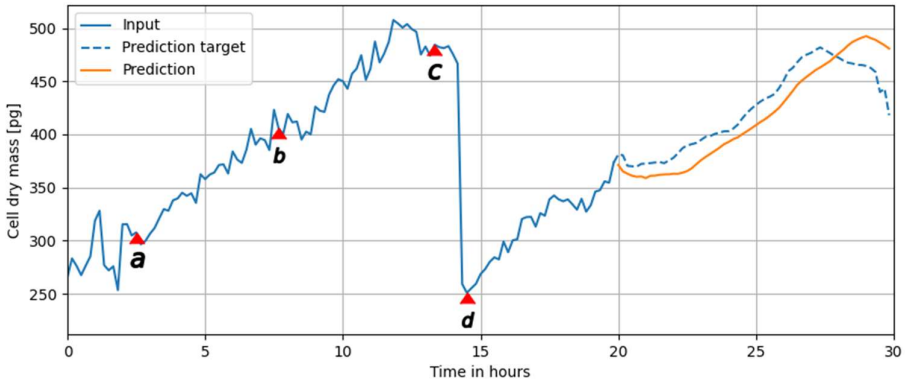


(a) Considered cell α is in pale white in the centre of the image

(b) Growing phase of the cell, its mass increases regularly

(c) Cell become spherical before division

(d) Division of the mother cell in two daughter cells in **pale white** α_1 and **pale yellow** α_2



(e) Dry mass of cell number 143. The plain blue line is the input feed into the network, the dashed blue line is the ground truth to be predicted and the plain orange line is the network prediction. Red triangles *a*, *b*, *c* and *d* are respectively the timestamps of figures **1a**, **1b**, **1c** and **1d**

Fig. 1. Tracking of cell number 143 in pale white tagged α which has a normal behaviour. Cell grows (**1a**-**1b**) and becomes spherical (**1c**) before division into two daughter cells (**1d**). One of them is given the same id (143) while the other is given the next available id

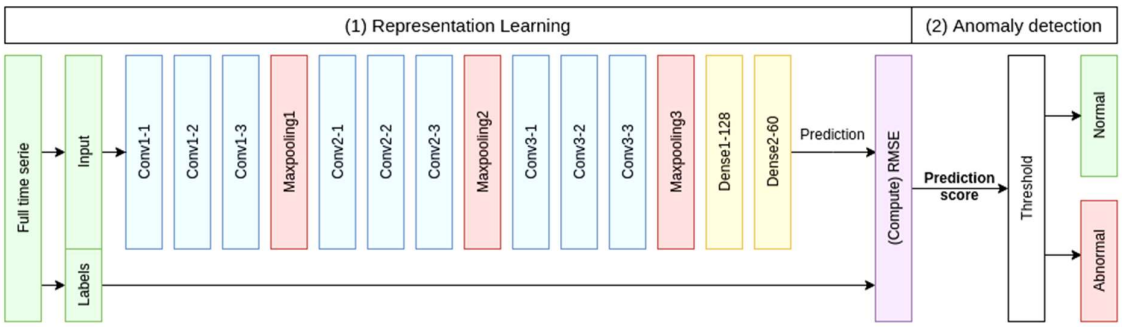


Fig. 3. Full anomaly detection pipeline. A 1D-CNN neural network is trained to predict the future of the time series. The RMSE between ground truth and prediction is compared to a threshold to define is a cell is abnormal

.

References

[1] T.-W. Su, S. Seo, A. Erlinger, and A. Ozcan, 'High-throughput lensfree imaging and characterization of a heterogeneous cell solution on a chip', *Biotechnology and Bioengineering*, vol. 102, no. 3, Art. no. 3, 2009

[2] C. Allier *et al.*, 'Imaging of dense cell cultures by multiwavelength lens-free video microscopy', *Cytometry Part A*, vol. 91, no. 5, Art. no. 5, 2017

[3] N. Rajewsky *et al.*, 'LifeTime and improving European healthcare through cell-based interceptive medicine', *Nature*, vol. 587, no. 7834, Art. no. 7834, Nov. 2020

[4] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, 'Outlier Detection for Temporal Data: A Survey', *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, Art. no. 9, Sep. 2014

[5] J. Ma and S. Perkins, 'Online novelty detection on temporal sequences', in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, Aug. 2003, pp. 613–618. doi: 10.1145/956750.956828.

[6] D. J. Hill and B. S. Minsker, 'Anomaly detection in streaming environmental sensor data: A data-driven modeling approach', *Environmental Modelling & Software*, vol. 25, no. 9, Art. no. 9, Sep. 2010

[7] N. D. Le, R. D. Martin, and A. E. Raftery, 'Modeling Flat Stretches, Bursts Outliers in Time Series Using Mixture Transition Distribution Models', *null*, vol. 91, no. 436, Art. no. 436, Dec. 1996

[8] R. S. Tsay, D. Peña, and A. E. Pankratz, 'Outliers in multivariate time series', *Biometrika*, vol. 87, no. 4, Art. no. 4, Dec. 2000

[9] A. Justel, D. Peña, and R. S. Tsay, 'Detection of outlier patches in autoregressive time series', *Statistica Sinica*, vol. 11, no. 3, Art. no. 3, 2001

[10] A. Luceño, 'Detecting possibly non-consecutive outliers in industrial time series', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 60, no. 2, Art. no. 2, 1998

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, 'ImageNet Classification with Deep Convolutional Neural Networks', in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. Accessed: Feb. 24, 2020. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[12] K. Simonyan and A. Zisserman, 'Very deep convolutional networks for large-scale image recognition', *CoRR*, vol. abs/1409.1556, 2015

[13] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', Jun. 2016.

[14] A. Krizhevsky, 'Learning Multiple Layers of Features from Tiny Images', p. 60, 2009

[15] O. Russakovsky *et al.*, 'ImageNet large scale visual recognition challenge', *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, Art. no. 3, 2015

[16] S. Kiranyaz, T. Ince, and M. Gabbouj, 'Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias', *Scientific Reports*, vol. 7, no. 1, Art. no. 1, Aug. 2017

[17] D. Li, J. Zhang, Q. Zhang, and X. Wei, 'Classification of ECG signals based on 1D convolution neural network', in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Oct. 2017, pp. 1–6. doi: 10.1109/HealthCom.2017.8210784.

[18] O. Abdeljaber, O. Avci, M. S. Kiranyaz, B. Boashash, H. Sodano, and D. J. Inman, '1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data', *Neurocomputing*, vol. 275, pp. 1308–1317, Jan. 2018

[19] O. Avci, O. Abdeljaber, S. Kiranyaz, and D. Inman, 'Structural Damage Detection in Real Time: Implementation of 1D Convolutional Neural Networks for SHM Applications', in *Structural Health Monitoring & Damage Detection, Volume 7*, Cham, 2017, pp. 49–54. doi: 10.1007/978-3-319-54109-9_6.

[20] L. Eren, T. Ince, and S. Kiranyaz, 'A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier', *Journal of Signal Processing Systems*, vol. 91, no. 2, Art. no. 2, Feb. 2019

[21] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, 'Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks', *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, Art. no. 11, Nov. 2016

[22] A. Khan, D.-K. Ko, S. C. Lim, and H. S. Kim, 'Structural vibration-based classification and prediction of delamination in smart composite laminates using deep learning neural network', *Composites Part B: Engineering*, vol. 161, pp. 586–594, Mar. 2019

[23] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, 'A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load', *Mechanical Systems and Signal Processing*, vol. 100, pp. 439–453, Feb. 2018

[24] A. van den Oord *et al.*, 'WaveNet: A generative model for raw audio', *CoRR*, vol. abs/1609.03499, 2016

[25] S. Du, T. Li, Y. Yang, and S. Horng, 'Deep Air Quality Forecasting Using Hybrid Deep Learning Framework', *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019

[26] S. Yi *et al.*, 'Interference Source Identification for IEEE 802.15.4 wireless Sensor Networks Using Deep Learning', in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2018, pp. 1–7. doi: 10.1109/PIMRC.2018.8580857.

[27] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, *Phoneme recognition: neural networks vs. hidden Markov models vs. hidden Markov models*. 1988. Accessed: Mar. 16, 2020. [Online]. Available: https://www.computer.org/csdl/proceedings-article/icassp/1988/00196523/12OmNCdk2By

[28] A. Waibel, 'Modular Construction of Time-Delay Neural Networks for Speech Recognition', *Neural Computation*, vol. 1, no. 1, Art. no. 1, Mar. 1989

[29] A. Kolesnikov, X. Zhai, and L. Beyer, 'Revisiting Self-Supervised Visual Representation Learning', in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 1920–1929. doi: 10.1109/CVPR.2019.00202.

[30] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, 'A simple framework for contrastive learning of visual representations', in *Proceedings of the 37th*

international conference on machine learning, Jul. 2020, vol. 119, pp. 1597–1607. [Online]. Available: http://proceedings.mlr.press/v119/chen20j.html

[31] S. Gidaris, P. Singh, and N. Komodakis, 'Unsupervised representation learning by predicting image rotations', Vancouver, Canada, Apr. 2018. [Online]. Available: https://hal-enpc.archives-ouvertes.fr/hal-01864755

[32] C. Doersch, A. Gupta, and A. A. Efros, 'Unsupervised Visual Representation Learning by Context Prediction', *arXiv:1505.05192 [cs]*, Jan. 2016

[33] M. Noroozi and P. Favaro, 'Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles', *arXiv:1603.09246 [cs]*, Aug. 2017

[34] G. Larsson, M. Maire, and G. Shakhnarovich, 'Learning Representations for Automatic Colorization', *arXiv:1603.06668 [cs]*, Aug. 2017

[35] S. Jenni and P. Favaro, 'Self-Supervised Feature Learning by Learning to Spot Artifacts', *arXiv:1806.05024 [cs]*, Jun. 2018

[36] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, 'Context Encoders: Feature Learning by Inpainting', *arXiv:1604.07379 [cs]*, Nov. 2016

[37] H.-Y. Lee, J.-B. Huang, M. Singh, and M.-H. Yang, 'Unsupervised representation learning by sorting sequences', Oct. 2017.

[38] I. Misra, C. L. Zitnick, and M. Hebert, 'Shuffle and Learn: Unsupervised Learning Using Temporal Order Verification', in *Computer Vision – ECCV 2016*, Cham, 2016, pp. 527–544. doi: 10.1007/978-3-319-46448-0_32.

[39] H. Cho, T. Kim, H. J. Chang, and W. Hwang, 'Self-Supervised Spatio-Temporal Representation Learning Using Variable Playback Speed Prediction', *arXiv:2003.02692 [cs]*, Mar. 2020

[40] C. Allier *et al.*, 'Quantitative phase imaging of adherent mammalian cells: a comparative study', *Biomed. Opt. Express*, vol. 10, no. 6, Art. no. 6, Jun. 2019

[41] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. 2009. doi: 10.1007/978-0-387-84858-7.

[42] Z. Cui, W. Chen, and Y. Chen, 'Multi-Scale Convolutional Neural Networks for Time Series Classification', *arXiv:1603.06995 [cs]*, May 2016