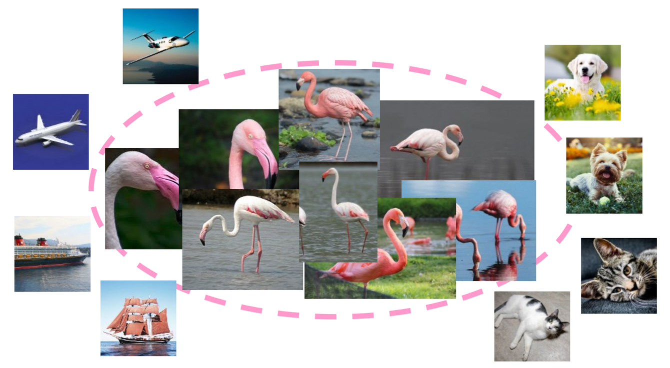
# Latent Space Autoregression for Novelty Detection


Davide Abati, Angelo Porrello, Simone Calderara, Rita Cucchiara

Department of Engineering, University of Modena and Reggio Emilia, Modena, Italy, {name.surname}@unimore.it


## 1. Problem Statement

*Novelty detection* is defined as the discrimination of new samples that significantly differ from training data. In this work, We focus on the **semi-supervised setting**, where the novelties are the ones exhibiting significantly different traits w.r.t. an underlying model of regularity, built from a collection of normal samples.



## 2. Model

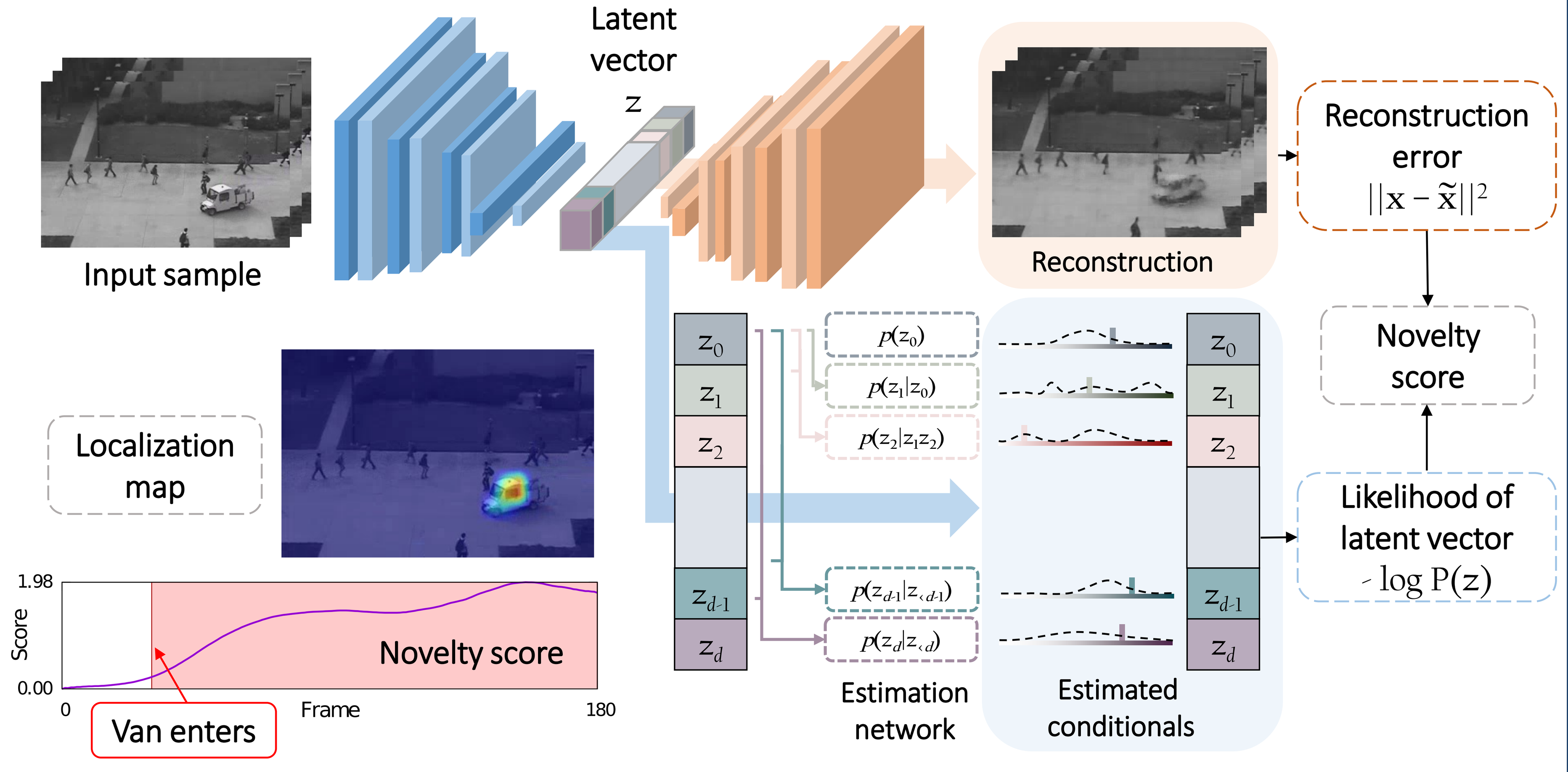
Our model leverages an auto-encoding architecture, composed by:

- an encoder $f(\mathbf{x};\theta_f) : \mathbb{R}^m \rightarrow \mathbb{R}^d$;
- a decoder $g(\mathbf{z};\theta_g) : \mathbb{R}^d \rightarrow \mathbb{R}^m$;
- an estimator $h(\mathbf{z};\theta_h) : \mathbb{R}^d \rightarrow [0,1]$.

Objective:

$$\mathcal{L} = \mathcal{L}_{\text{REC}}(\theta_f,\theta_g) + \lambda\mathcal{L}_{\text{LLK}}(\theta_f,\theta_h)$$

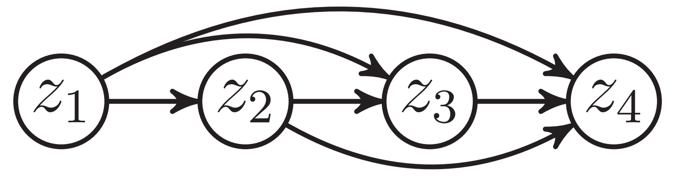
$$= \mathbb{E}_{\mathbf{x}}\Big[\underbrace{||\mathbf{x}-\tilde{\mathbf{x}}||^2}_{\text{reconstruction term}} - \lambda \underbrace{\log(h(\mathbf{z};\theta_h))}_{\text{log-likelihood term}}\Big]$$




## 3. Autoregression

Autoregressive models factorize the joint density function on $d$ variables through the chain rule of probability:

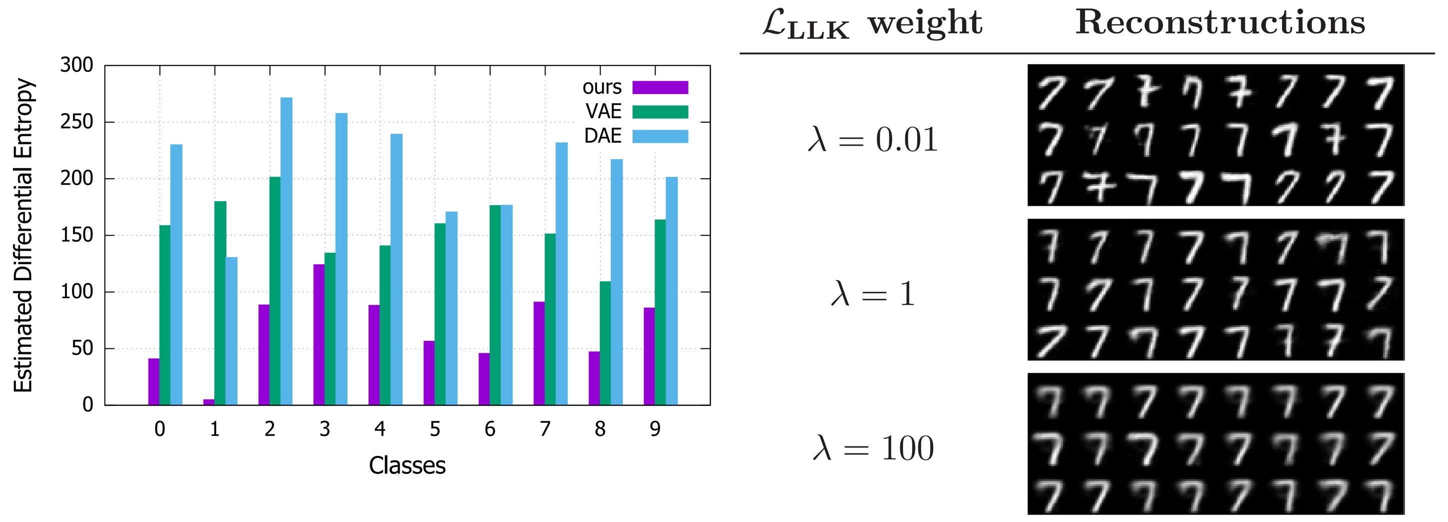
$$p(\mathbf{z}) = \prod_{i=1}^{d} p(z_i|\mathbf{z}_{<i})$$



We introduce the Masked Fully Connection (MFC) and Masked Stacked Convolution (MSC) layers enforcing an autoregressive procedure within the estimator $h(\mathbf{z};\theta_h)$.

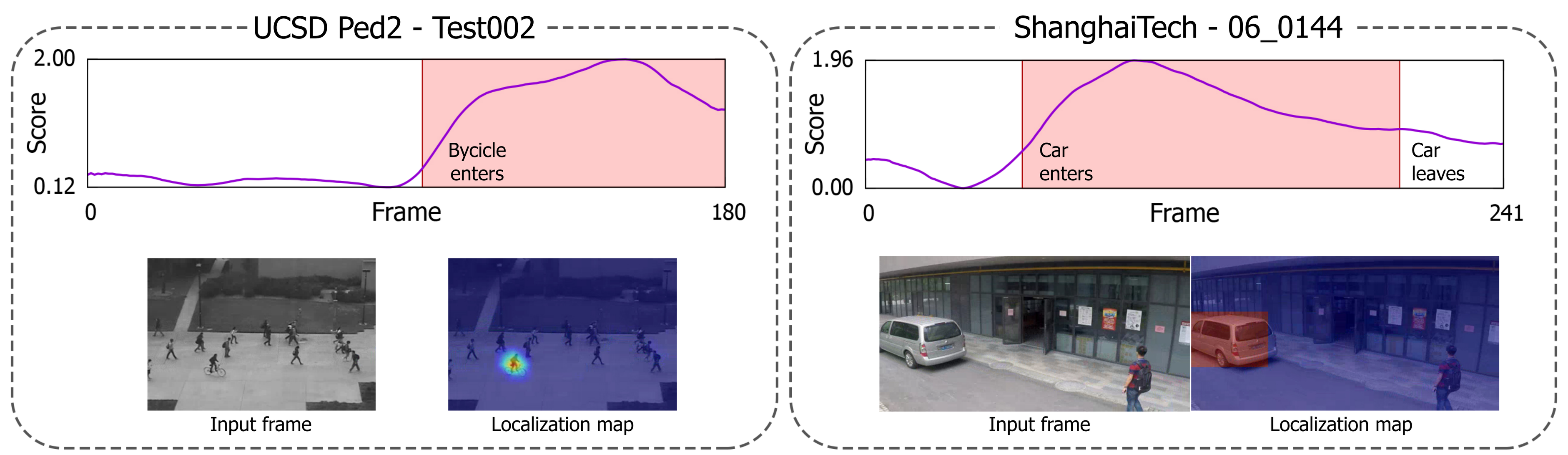
## 4. Entropy Regularization

The $\mathcal{L}_{\text{LLK}}$ objective leads to the minimization of the differential entropy underlyin the encoding distribution.




## 5. Video Anomaly Detection

Results on the UCSD Ped2 and ShanghaiTech datasets are reported as Area Under ROC Curve (AUROC).

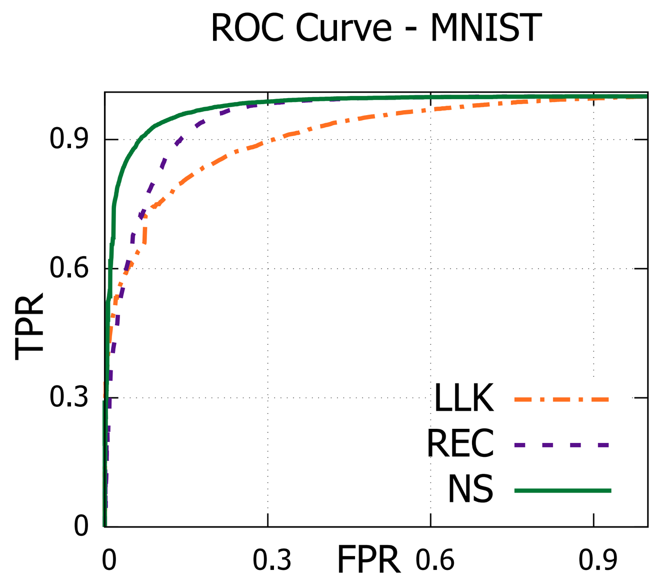



| | UCSD Ped2 | ShanghaiTech |
|---|---|---|
| ConvAE [1] | 0.850 | 0.609 |
| Hinami et. al [2] | 0.922 | - |
| TSC [4] | 0.910 | 0.679 |
| Stacked RNN [4] | 0.922 | 0.680 |
| FFP [3] | 0.935 | - |
| FFP+MC [3] | **0.954** | **0.728** |
| Ours | **0.954** | **0.725** |

## 6. One-Class Novelty Detection

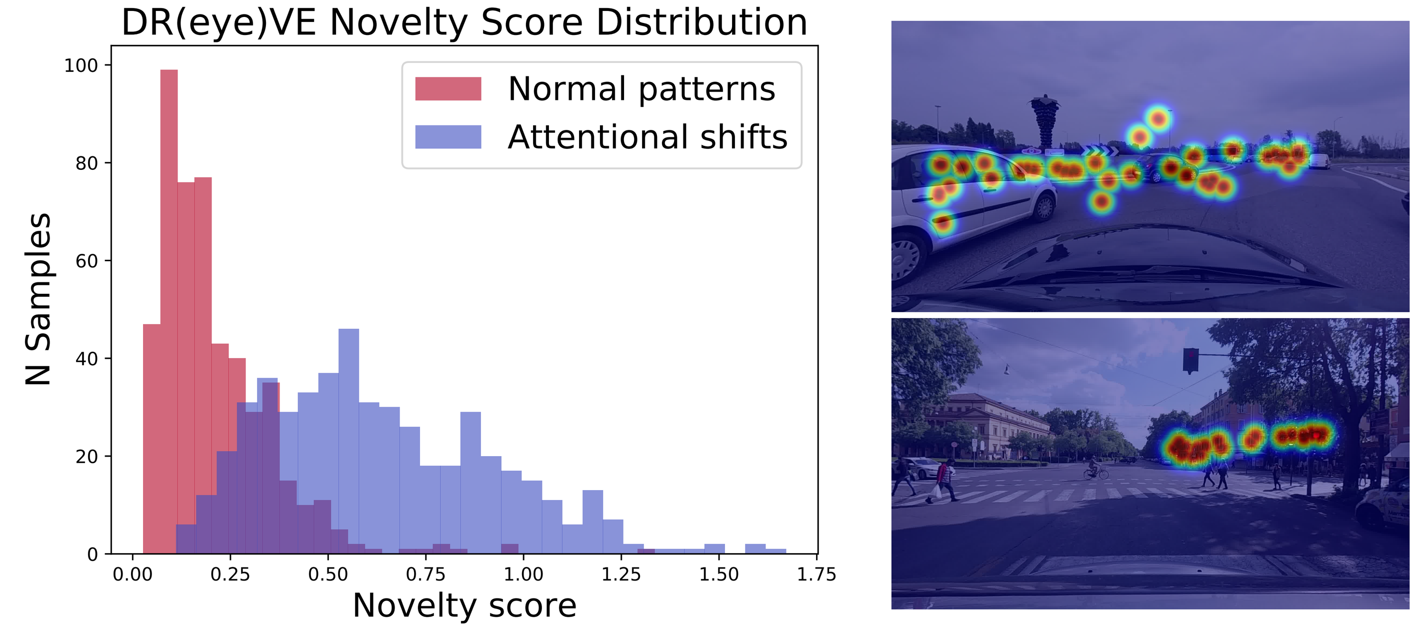
We test the model in one class settings, training it on each class of either MNIST or CIFAR-10 separately. We report the comparison in terms of average AUROC.

| | MNIST | CIFAR10 |
|---|---|---|
| OC SVM | 0.951 | 0.586 |
| KDE | 0.814 | 0.610 |
| DAE | 0.942 | 0.590 |
| VAE | 0.969 | 0.586 |
| Pix CNN | 0.618 | 0.551 |
| GAN | 0.866 | 0.592 |
| Ours | **0.975** | **0.641** |




## 7. DR(eye)VE Outlier Detection

We measure the correlation about the novelty score and the attentional shifts labeled in the DR(eye)VE dataset.




## 8. Model Analysis

We compare MFC and MSC against recurrent layers.

| CIFAR-10 | | UCSD Ped2 | |
|---|---|---|---|
| $\text{LSTM}_{[100]}$ | 0.623 | $\text{LSTM}_{[100]}$ | 0.849 |
| $\text{LSTM}_{[32,32,32,32,100]}$ | 0.622 | $\text{LSTM}_{[4,4,4,4,100]}$ | 0.845 |
| $\text{MFC}_{[100]}$ | 0.625 | $\text{MSC}_{[100]}$ | 0.849 |
| $\text{MFC}_{[32,32,32,32,100]}$ | **0.641** | $\text{MSC}_{[4,4,4,4,100]}$ | **0.954** |

## References


[1] M. Hasan et al. Learning temporal regularity in video sequences. In *CVPR*, 2016.

[2] R. Hinami et al. Joint detection and recounting of abnormal events by learning deep generic knowledge. In *ICCV*, 2017.

[3] W. Liu et al. Future frame prediction for anomaly detection – a new baseline. *CVPR*, 2018.

[4] W. Luo et al. A revisit of sparse coding based anomaly detection in stacked rnn framework. *ICCV*, 2017.