Spatio-Temporal Convolutional Autoencoders for Perimeter Intrusion Detection

Devashish Lohani, Carlos Crispim-Junior, Quentin Barthélemy, Sarah Bertrand, Lionel Robinault and Laure Tougne

LIRIS UMR 5205 CNRS

Reproducible Research in Pattern Recognition

FOXSTREAM GROUP | INTELLIGENT VIDEO

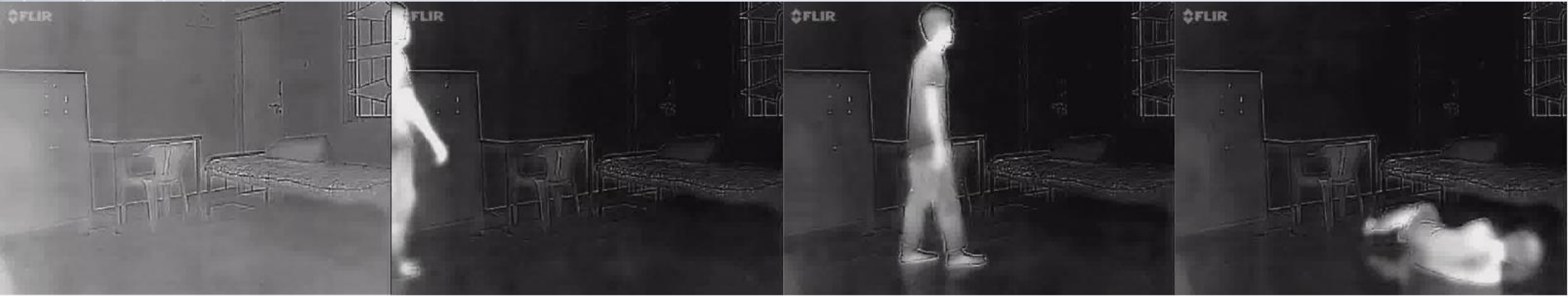
cnrs INSA UCB Lyon 1 Université Lumière Lyon 2 École Centrale Lyon

Outline

- Objectives
- Existing Works
- Approach
- Datasets
- Results
- Conclusion and Future Work

Objectives

- **Reproducibility on Fall detection task: Deepfall* framework**



a) Empty frame b) Person entering c) Person in frame d) Fall frame

- **Applicability on Perimeter intrusion detection task**

* Nogas, Jacob, Shehroz S. Khan, and Alex Mihailidis. "Deepfall: Non-invasive fall detection with deep spatio-temporal convolutional autoencoders." Journal of Healthcare Informatics Research (2020).

Perimeter Intrusion Detection System (PIDS)

- **Intrusion**
  - **Moving** object ∈ **unauthorized** category (like motorcycle, car, person, etc.)
    - unauthorized for **particular area** (space) and **time**
- **PIDS detects intrusion**
  - Detecting all frames containing an intrusion in the video

Existing PID Approaches

- **Moving object detection**
  - GMM, Background subtraction [Vijverberg14-ICIP]
- **Moving object detection & image classification**
  - Background subtraction + **HOG** & handcrafted features + **SVM**/ANN [Zhang15-ICCAS]
  - Background subtraction + **CNN** based classifier [Seung18-ANE]

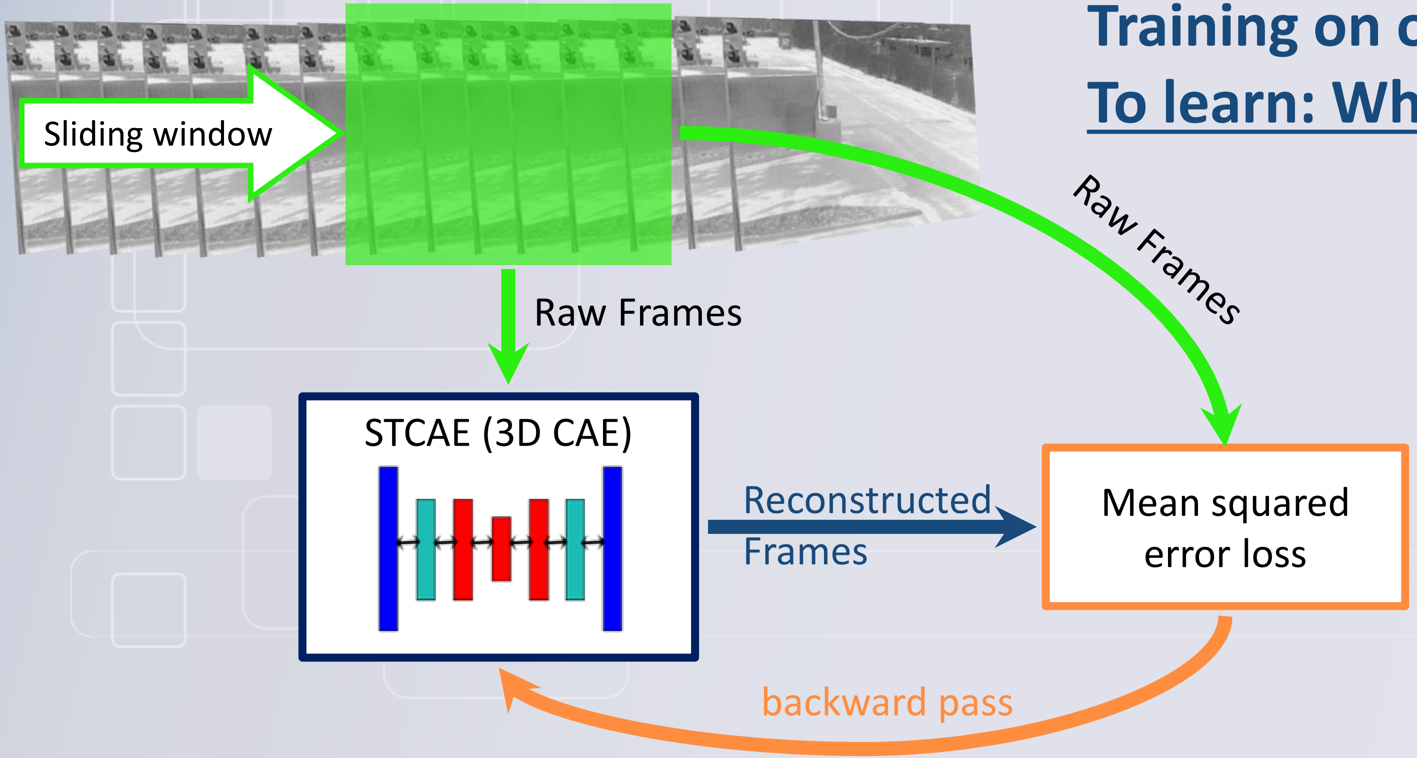
Limitations:

- Learns spatial & temporal features independently
- Supervised learning based PIDS (humans as intrusions)

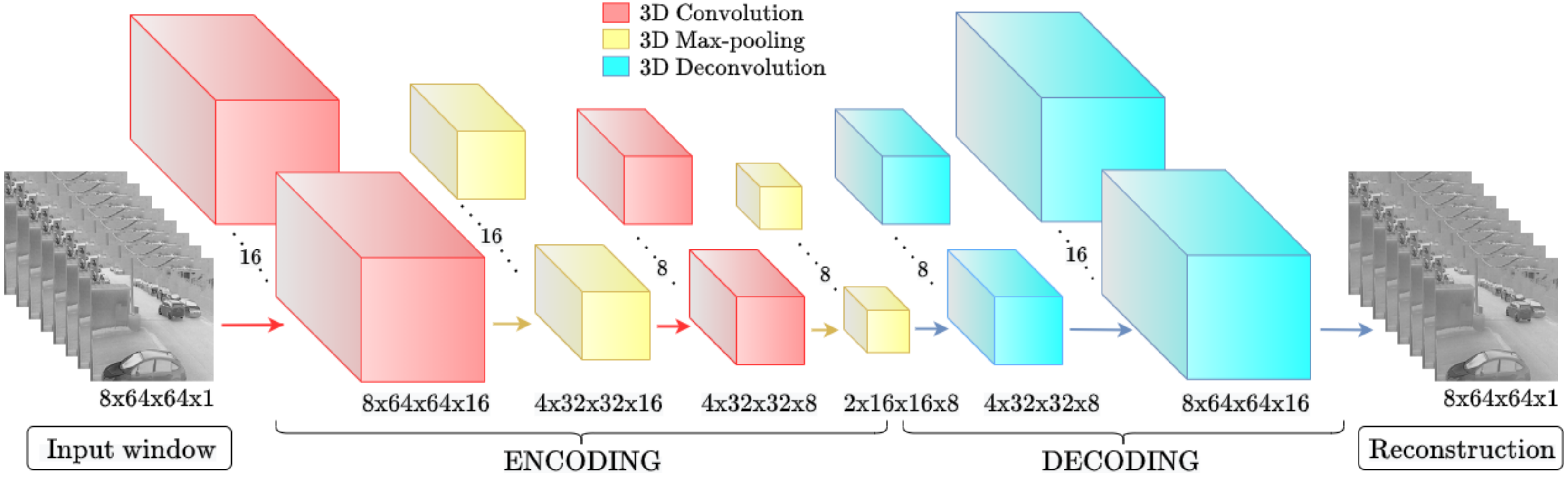
Approach

Training on only normal videos

<u>To learn: What is normal</u>

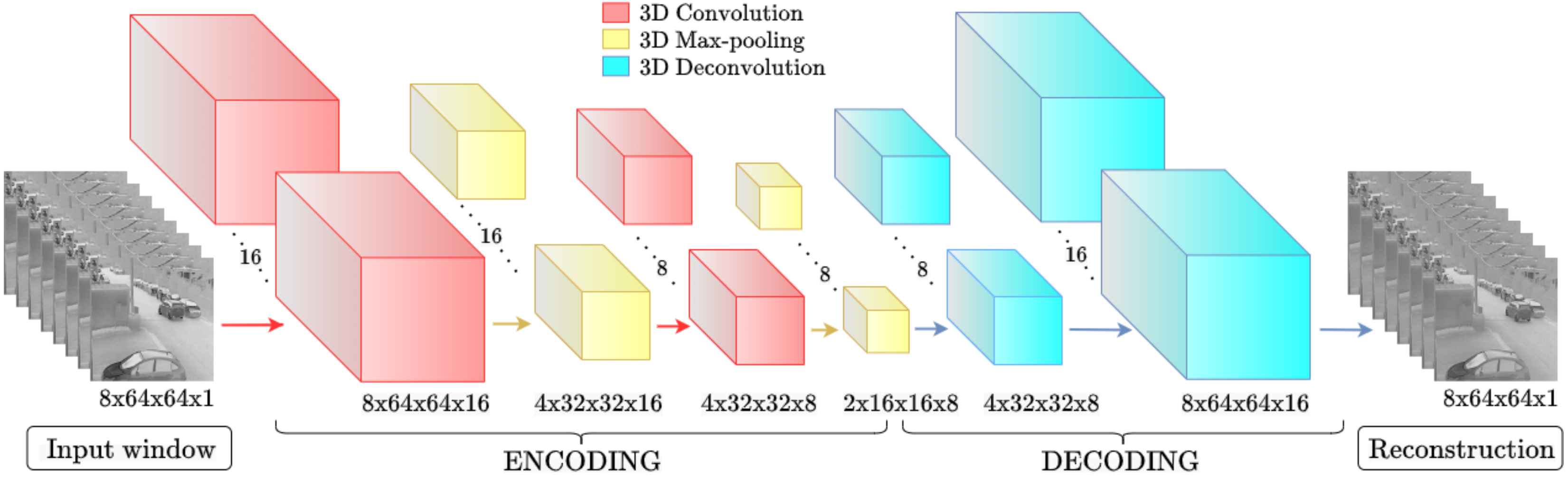


Spatio Temporal Convolutional AutoEncoder (STCAE)



1. **UpSampling**	3D Convolution + 3D Max-pooling	3D Upsampling + 3D Convolution
2. **Deconvolution**	3D Convolution + 3D Max-pooling	3D Deconvolution
3. **C3D**	3D Convolution + 2D/3D Max-pooling	2D/3D Upsampling + 3D Convolution

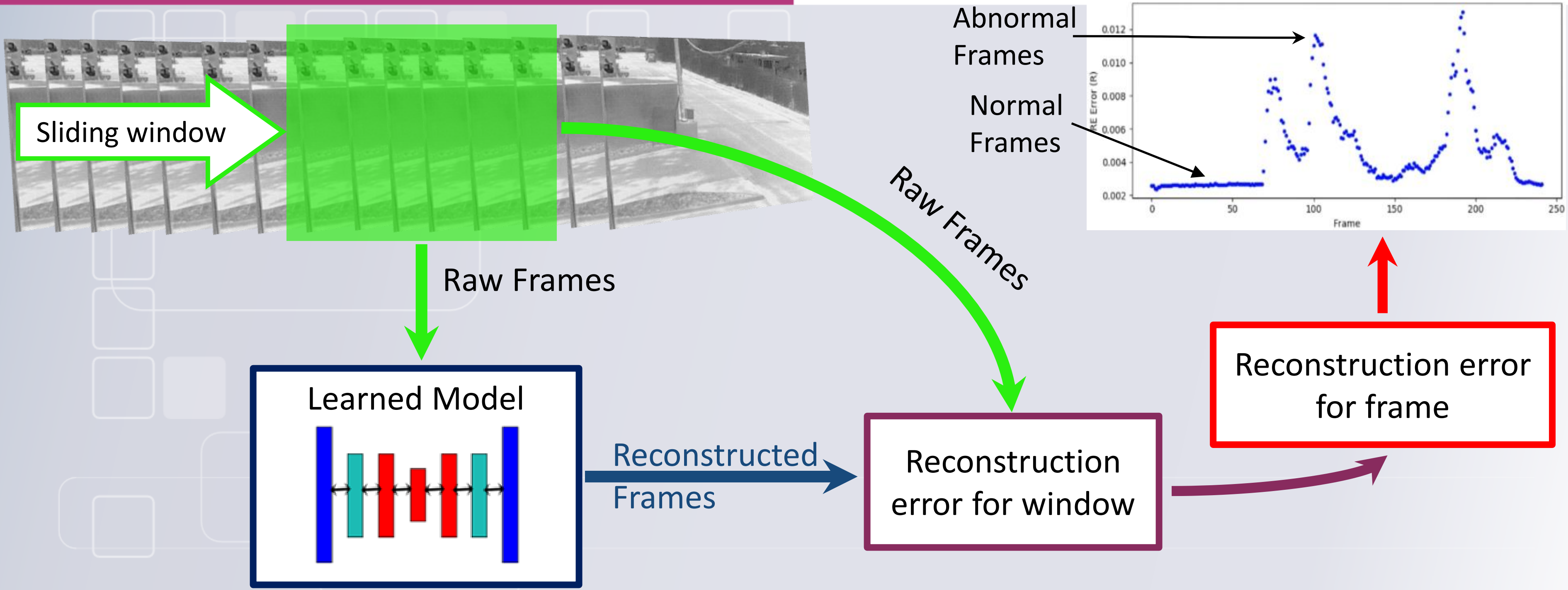
Spatio Temporal Convolutional AutoEncoder (STCAE)



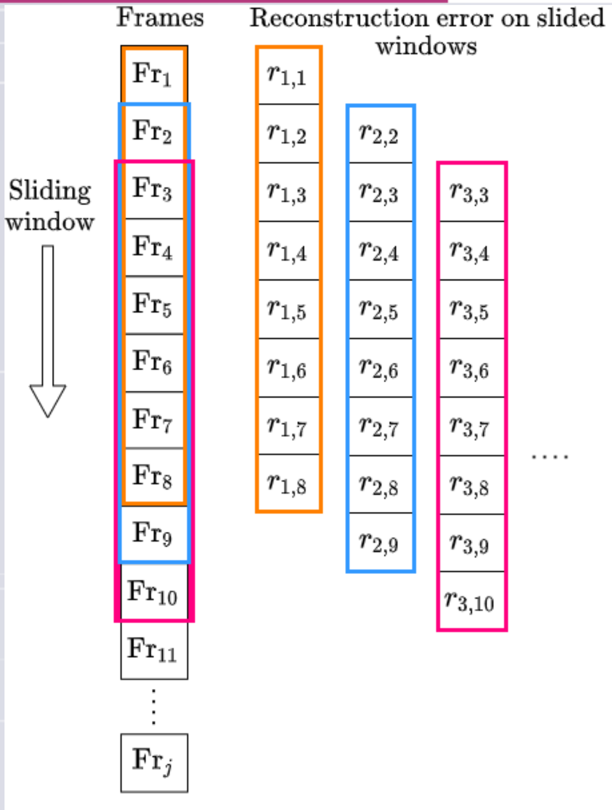
Training parameters:

- **Activation function : ReLU** - hidden layers, **tanh** - output layer
- **Optimizer :** Adadelta
- **Epochs :** 500

Approach: Detecting Abnormal events



Reconstruction Error for windows



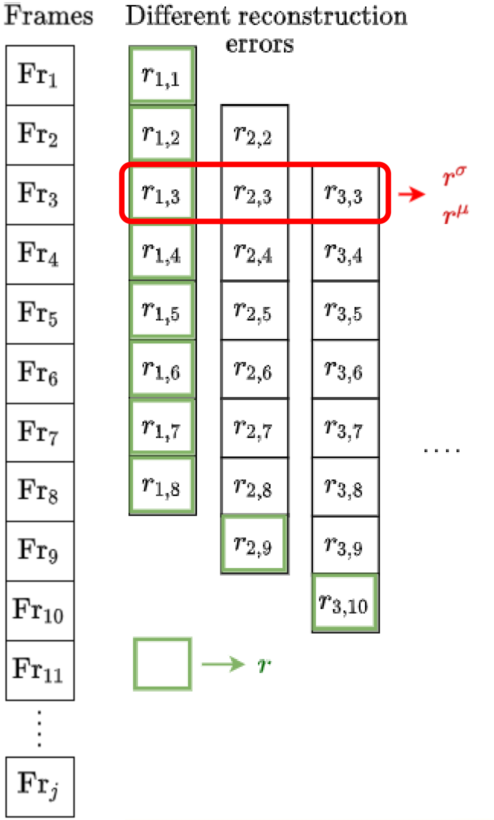
Reconstruction Error for frames

1. Reconstruction error r

- ❑ *RE of a frame from 1st window it appears in*

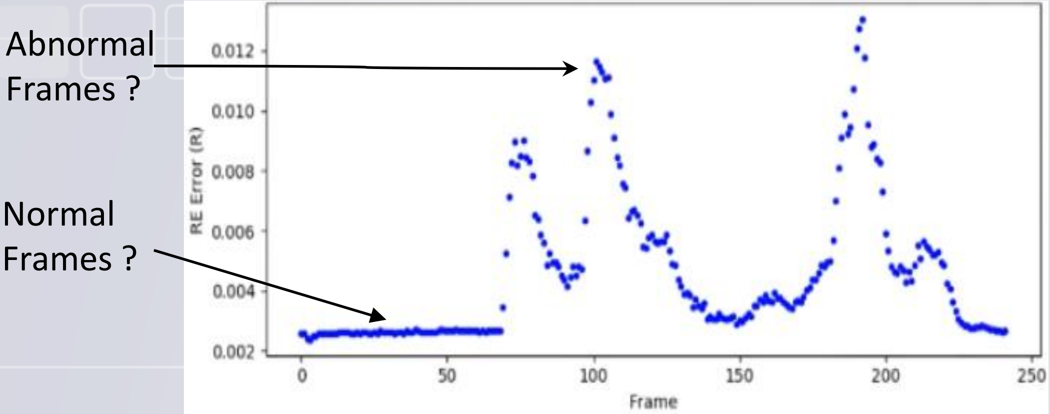
2. Cross-Window Reconstruction Errors

- ❑ r^{μ} : **mean** of REs **across temporal context**
- ❑ r^{σ} : **standard deviation** of REs **across temporal context**



Detection of Abnormal events

Depends on Threshold



- **ROC** and precision-recall (**PR**) curves
- **Overall score**: Area under the curve (**AUC**)
  - AUROC and AUPR scores

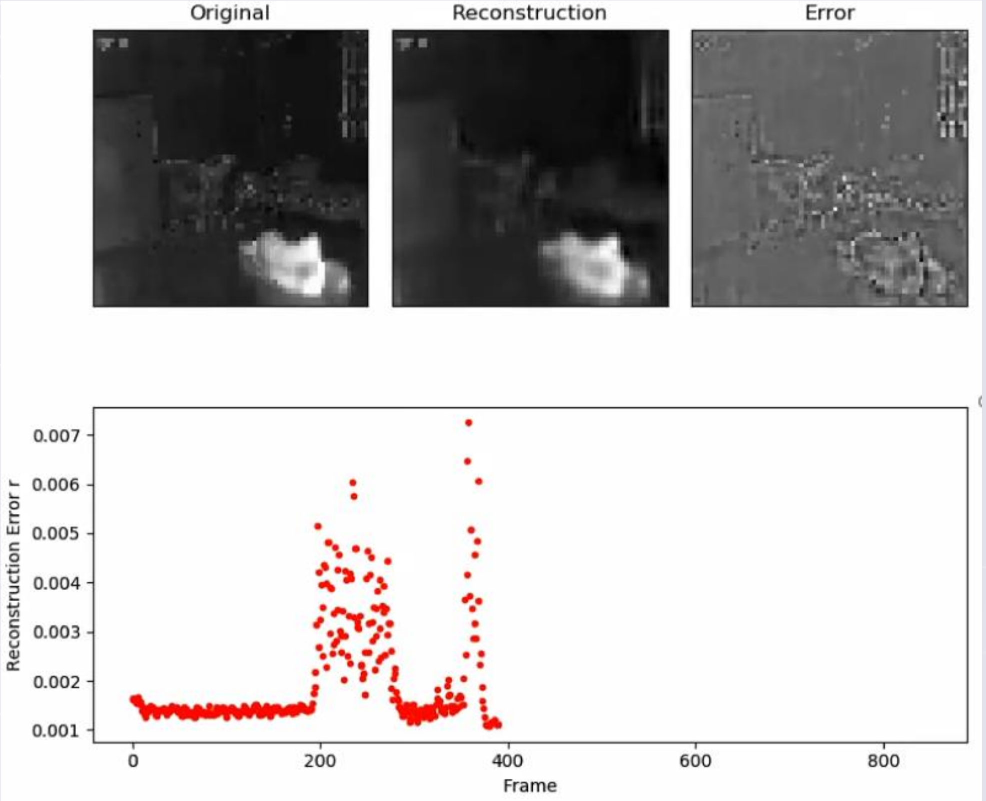
Datasets : Thermal Fall Dataset [Vadivelu16-ACCV]

- **Training:**
  - 9 non-fall videos
  - 22096 frames

- **Testing:**
  - 35 videos
  - 36166 frames
  - 2.19% Fall frames

➢ **Spatial Resolution:** 640 × 480, **Frame Rate:** 25 fps or 15 fps, **Single view**

Result demonstration on Fall Detection Task



Reproduced Results on Fall Detection Task

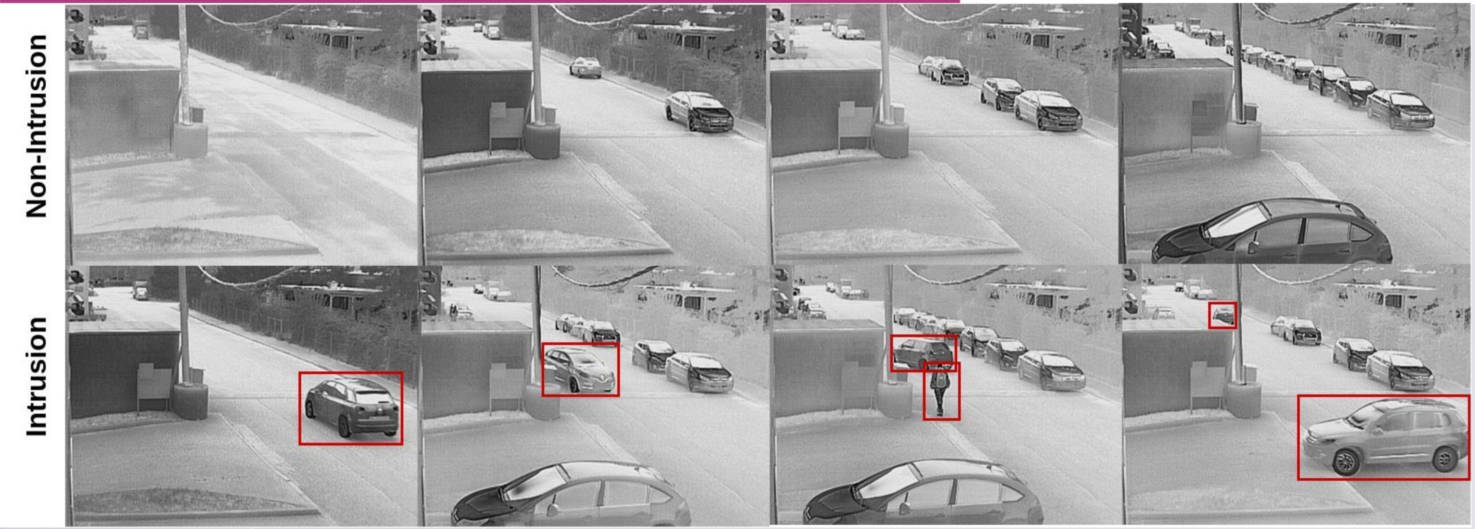
Models	RE	Testing Time	DeepFall	Ours		
			AUROC per video	AUROC per video	AUROC all videos	AUPR all videos
STCAE UpSampling	r^σ	49.88s	0.96(0.03)	0.96(0.02)	0.96	**0.29**
	r^μ	48.61s	0.95(0.04)	0.94(0.04)	0.88	0.23
	r	47.11s	–	0.94(0.04)	0.89	0.24
STCAE Deconvolution	r^σ	56.31s	0.96(0.02)	0.96(0.02)	0.96	**0.27**
	r^μ	55.94s	0.94(0.04)	0.94(0.04)	0.88	0.23
	r	54.92s	–	0.94(0.04)	0.89	0.21
STCAE C3D	r^σ	55.98s	0.97(0.02)	0.96(0.03)	0.95	**0.25**
	r^μ	54.52s	0.93(0.07)	0.90(0.07)	0.85	0.19
	r	54.23s	–	0.91(0.06)	0.87	0.21

- **UpSampling** models: fastest speed & highest performance
- Models with r^σ perform superior to others
- **Overall :**
  - Up to **6%** degradation of AUROC scores from “per video” to “all videos”
  - Poor AUPR all videos score

LIRIS

FOXSTREAM GROUP | INTELLIGENT VIDEO

Datasets : Perimeter Intrusion Dataset



- **Single View**
- **25** fps
- **400 × 296** frame resolution
- **Unauthorized classes:** human, car, bike, truck & other vehicles

- **Training - 80** videos
  - Only non-intrusion frames : **41941** frames
- **Testing - 100** videos
  - **38403** frames, **37.04%** intrusion frames

Result demonstration on Perimeter Intrusion Detection Task

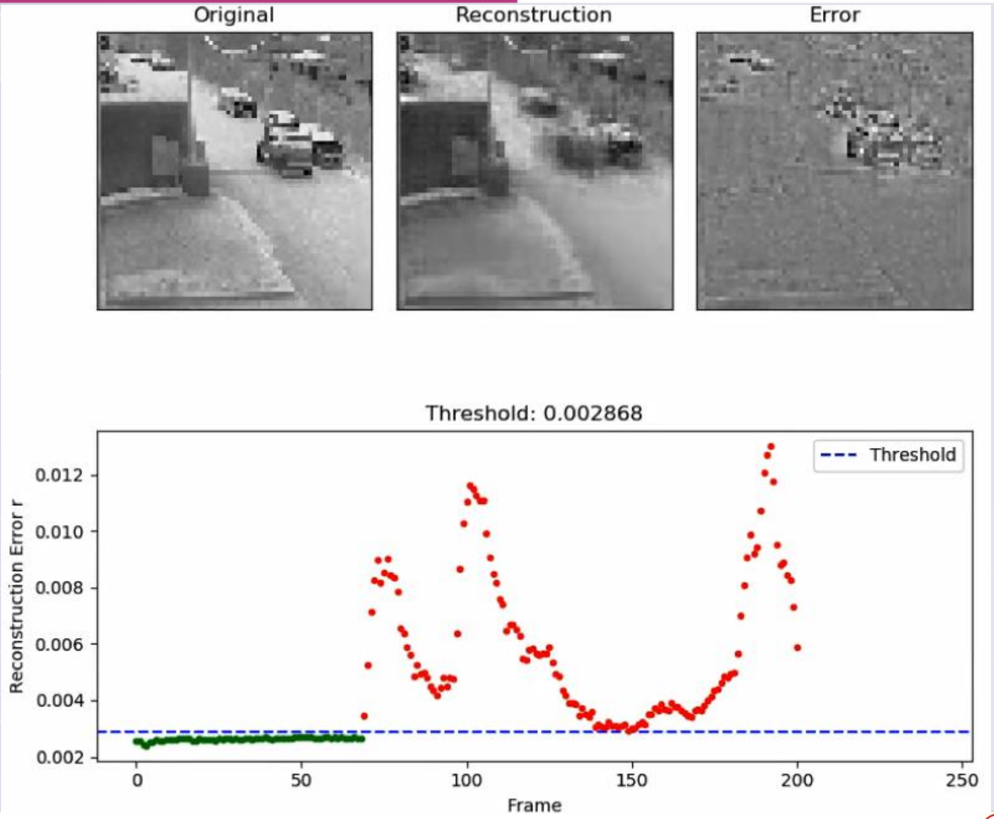
Optimal Threshold :

Where TPR is high,

FPR is low

Minimizing

tpr - (1-fpr)



Results on Perimeter Intrusion Detection Task

Models	RE	Time		AUC all videos	
		Training	Testing	ROC	PR
DSTCAE UpSampling	r^σ	590.25 min	55.19s	0.93	**0.88**
	r^μ		52.05s	0.91	0.81
	r		51.24s	0.92	0.83
DSTCAE Deconvolution	r^σ	594.95 min	61.15s	0.93	**0.86**
	r^μ		59.57s	0.91	0.80
	r		58.55s	0.91	0.82
DSTCAE C3D	r^σ	591.10 min	60.38s	0.90	0.81
	r^μ		59.46s	0.91	0.80
	r		57.98s	0.91	0.82

- **UpSampling** models: fastest speed & highest performance
- Models with r^σ perform superior to others
- **Overall :**
  - **Smaller gap** between AUROC and AUPR scores
  - **Good AUPR** score

Conclusion

- **STCAE application to 2 unsupervised tasks**
- **Fall detection task**
  - Correct results reproduction
    - Highest performance : UpSampling models and models with r^{σ}
  - Extended evaluation : AUC all videos
    - Large gap between AUROC and AUPR scores
    - Poor AUPR scores
- **Application on the Perimeter intrusion detection task**
  - UpSampling models and models with r^{σ} perform superior to others
  - Small gap between AUROC and AUPR scores
  - Good AUPR scores : STCAE suitable for a PIDS

Future Works

- **Testing model robustness on**
  - Lighting conditions
  - Intruder displacement speed
  - Intruder distance from camera
- **Cross video testing**
- **Adaptative threshold selection**

References

1. [Seung18-ANE] Seung Hyun Kim, Su Chang Lim, et al. Intelligent intrusion detection system featuring a virtual fence, active intruder detection, classification, tracking, and action recognition. Annals of Nuclear Energy, 2018.
2. [Zhang15-ICCAS] Zhang, Yong-Liang, et al. "Perimeter intrusion detection based on intelligent video analysis." 2015 15th International Conference on Control, Automation and Systems (ICCAS).
3. [Tran15-ICCV] Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." ICCV,2015.
4. [Kiran18-J. Imaging] Kiran, Ravi et al. "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos." Journal of Imaging, 2018.
5. [Nogas20-JHIR] Nogas, Jacob, Shehroz S. Khan, and Alex Mihailidis. "Deepfall: Non-invasive fall detection with deep spatio-temporal convolutional autoencoders." Journal of Healthcare Informatics Research (2020).
6. [Vadivelu16-ACCV] Vadivelu, Somasundaram, et al. "Thermal imaging based elderly fall detection." Asian Conference on Computer Vision. Springer, Cham, 2016.
7. [Vijverberg14-ICIP] Vijverberg, J.A., Janssen, R.T., de Zwart, R., de With, P.H.: Perimeter-intrusion event classification for on-line detection using multiple instance learning solving temporal ambiguities. In: ICIP. pp. 2408–2412 (2014)

UMR 5205 CNRS

Spatio-Temporal Convolutional Autoencoders for Perimeter Intrusion Detection

Devashish Lohani, Carlos Crispim-Junior, Quentin Barthélemy, Sarah Bertrand, Lionel Robinault and Laure Tougne

Contact : devashish.lohani@univ-lyon2.fr , d.lohani@foxstream.fr