Video Surveillance anomaly detection using Autoencoders

Ernesto Cruz-Esquivel | Advisor: Zobeida Guzman-Zavaleta
Universidad de las Americas Puebla, Mexico

Introduction

- Video surveillance monitors scenes to detect **critical events**, like accidents or crimes
- Such anomalies are deviations from the usual or the considered normal and are **context dependent**
- Current video anomaly detection methods use a **high amount of computational resources**

The **main purpose** of this research is to propose an architecture based on Autoencoders **balancing the time-accuracy tradeoff** in comparison with expensive related methods. The improvement of **spatiotemporal feature extraction for video** data is a priority.

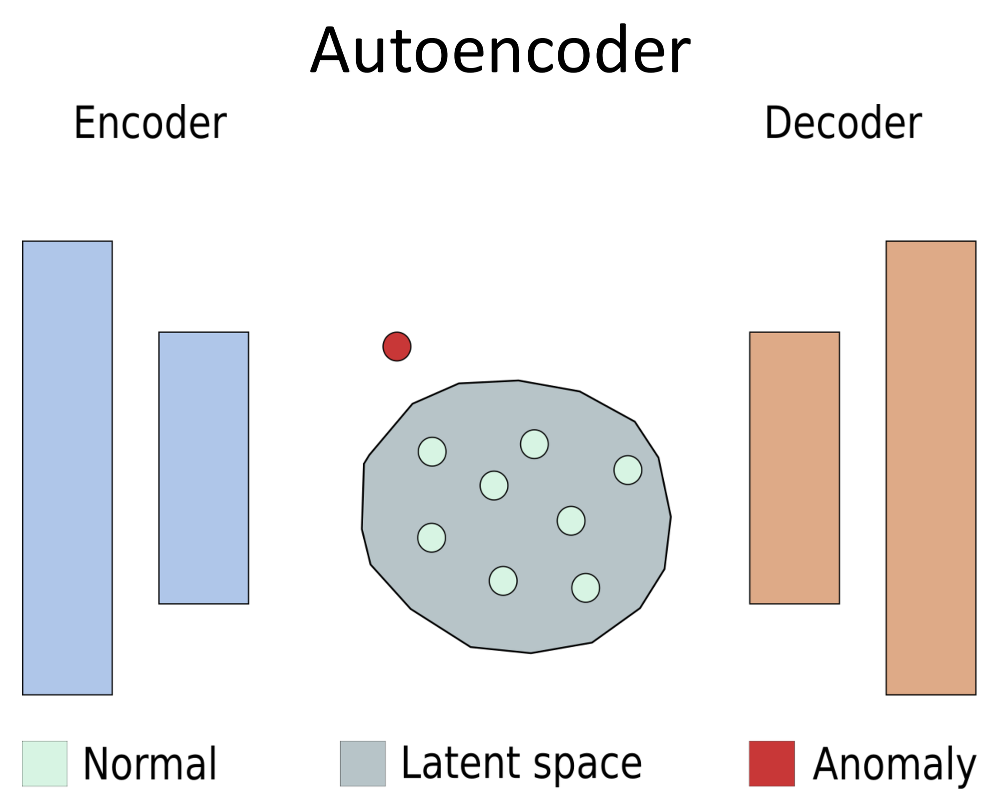

Fig. 1 Anomaly detection using Autoencoders

Autoencoder **trained** using only **normal data**. Trained autoencoder cannot **encode or decode anomalies**.

Acknowledgments

This work was supported in part by Consejo Nacional de Ciencia y Tecnología (CONACyT) student grant 975079, and in part by Fundación Universidad de las Américas Puebla (UDLAP) BAUI grant, and with infrastructure from Laboratorio Nacional de Supercómputo del Sureste de México LNS project 201902064C.

Contact Info:
Ernesto Cruz Esquivel – ernesto.cruzel@udlap.mx

Initial proposal

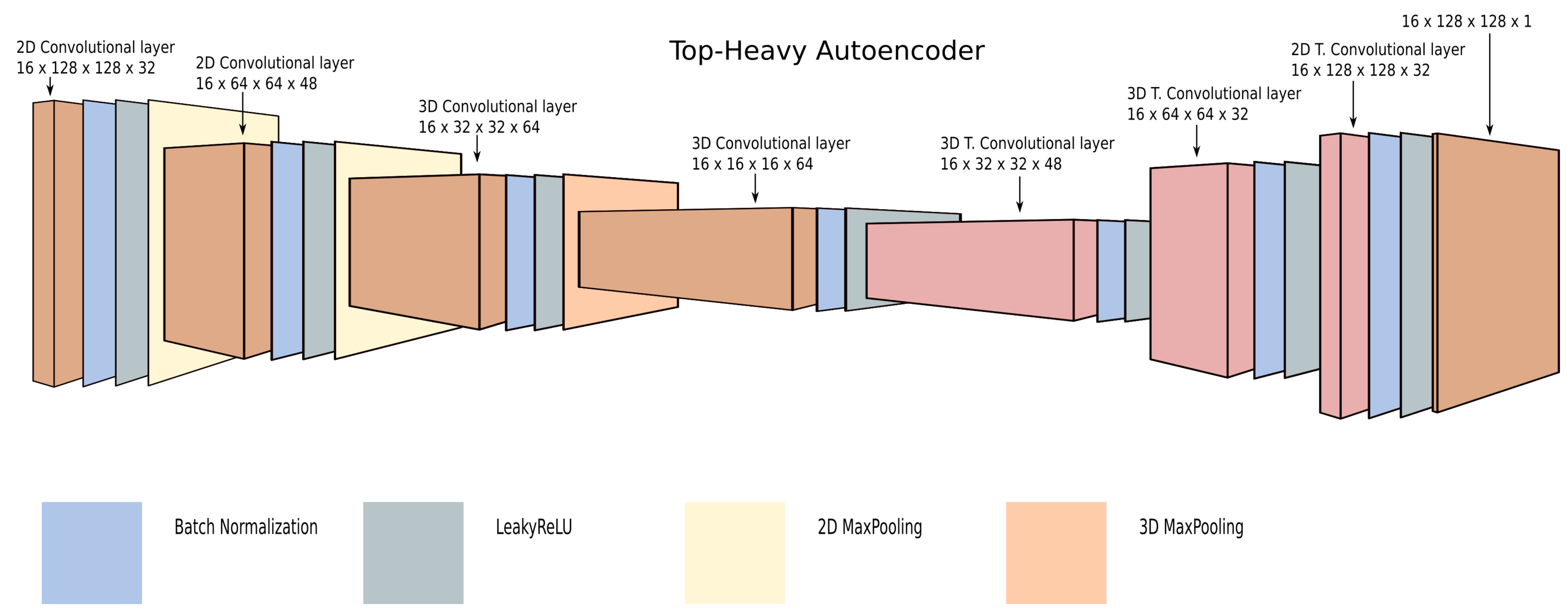

Fig. 2 Top-Heavy Autoencoder combines 2D and 3D convolutional layers and it was used for the tested models.

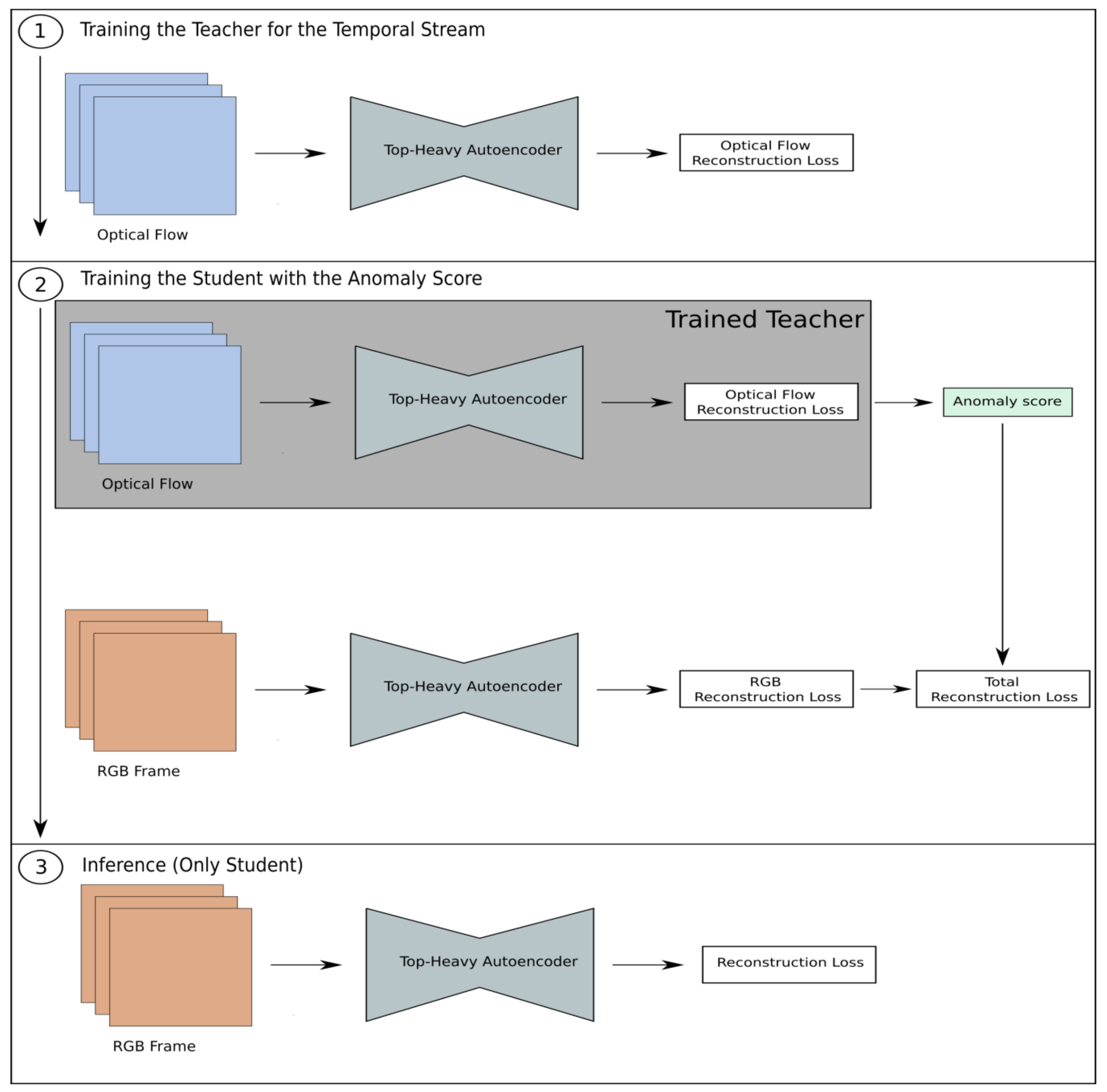

Fig. 3 Anomaly score distillation pipeline

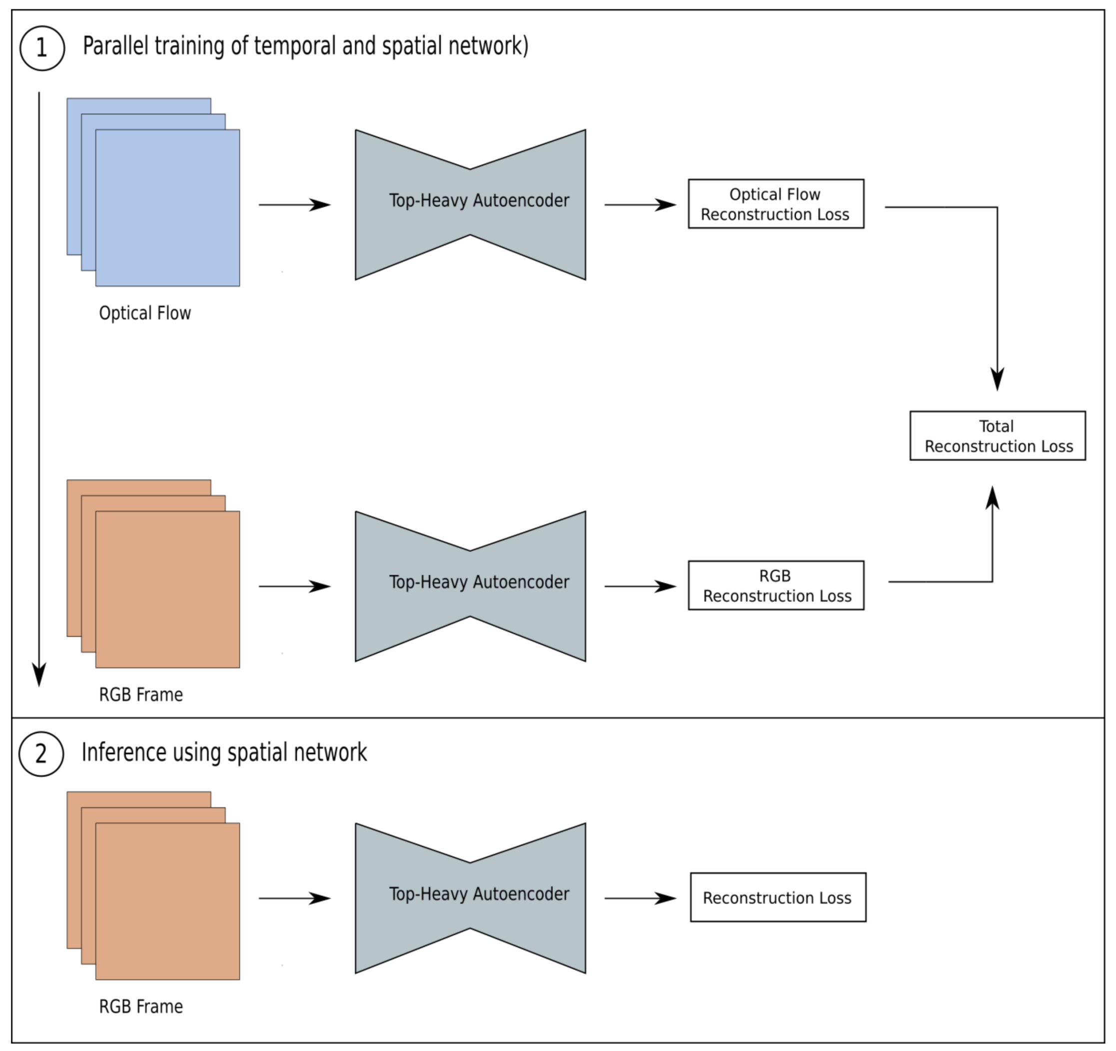

Fig. 4 Joint spatiotemporal training pipeline

Anomaly score distillation

- Extract **spatiotemporal features** using a single network
- $s(\theta, x^{(i)}) = [L_t(x^{(i)}) - L_s(x^{(i)};\ \theta)]^2$ Temporal reconstruction loss L_t, Spatial reconstruction loss L_s, Training parameters θ, Dataset $x^{(i)}$
- Anomaly score increases when there exists a **difference between the spatial and temporal** reconstruction loss

Joint spatiotemporal training

- Extract **spatiotemporal features** using a single network
- $L(\theta_1, \theta_2) = L_s(\theta_1) + [L_s(\theta_1) * L_t(\theta_2)]$ Temporal reconstruction loss L_t, Spatial reconstruction loss L_s, Spatial training parameters θ_1, Temporal training parameters θ_2
- Parallel training of the **temporal and spatial stream**

Results

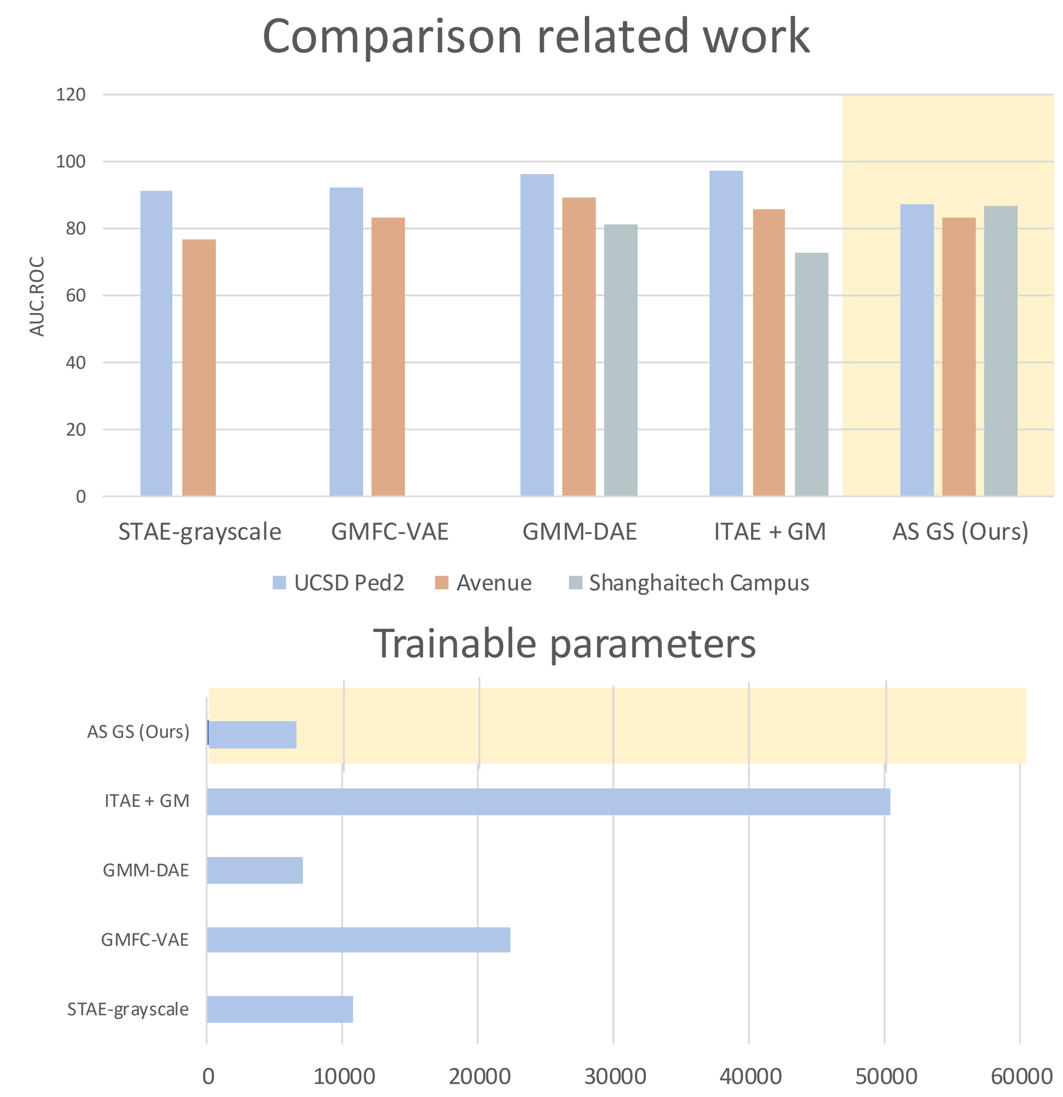

Fig. 5 Comparison against related work of AUC-ROC and trainable parameters. Our work highlighted in yellow background.

- Results of **Shanghaitech Campus** are the best of related work
- The least quantity of **trainable parameters**

Conclusions

- **Top-Heavy Autoencoder** uses only the **57%** of the trainable parameters compared to the 3D autoencoder with an average absolute difference of the **AUC-ROC** between both models of **1.2**
- The **AUC-ROC** difference between **Anomaly Score and Joint spatiotemporal** training methods has an average of **1.12**
- The results show a slight improvement in **temporal features extraction** and the extraction of more defining temporal features must be addressed for future work.

References

- Cruz-esquivel, E., & Guzman-zavaleta, Z. J. (2022). An Examination on Autoencoder Designs for Anomaly Detection in Video Surveillance. *IEEE Access*, *10*, 6208–6217. https://doi.org/10.1109/ACCESS.2022.3142247
- Cho, M., Kim, T., & Lee, S. (2020). *Unsupervised Video Anomaly Detection via Flow-based Generative Modeling on Appearance and Motion Latent Features*. http://arxiv.org/abs/2010.07524
- Ouyang, Y., & Sanchez, V. (2020). *Video Anomaly Detection by Estimating Likelihood of Representations*. http://arxiv.org/abs/2012.01468
- Fan, Y., Wen, G., Li, D., Qiu, S., Levine, M. D., & Xiao, F. (2020). Video anomaly detection and localization via Gaussian Mixture Fully Convolutional Variational Autoencoder. *Computer Vision and Image Understanding*, *195*(September 2018), 102920. https://doi.org/10.1016/j.cviu.2020.102920
- Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., & Hua, X. S. (2017). Spatio-temporal AutoEncoder for video anomaly detection. *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*, 1933–1941. https://doi.org/10.1145/3123266.3123451