Adversarial Regularized Reconstruction for Anomaly Detection and Generation

Angelica Liguori[1], Giuseppe Manco[2], Francesco Sergio Pisani[2] and Ettore Ritacco[2]

[1] Department of Computer Engineering, Modeling, Electronics and, Systems – University of Calabria, Italy
[2] Institute for High Performance Computing and Networking, National Research Council, Italy

ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Scenario

- **Goal**. Generate realistic outliers for enabling the learning of an outlier detector

- **Idea**. Develop an anomaly detection and generation system based on unsupervised deep learning model

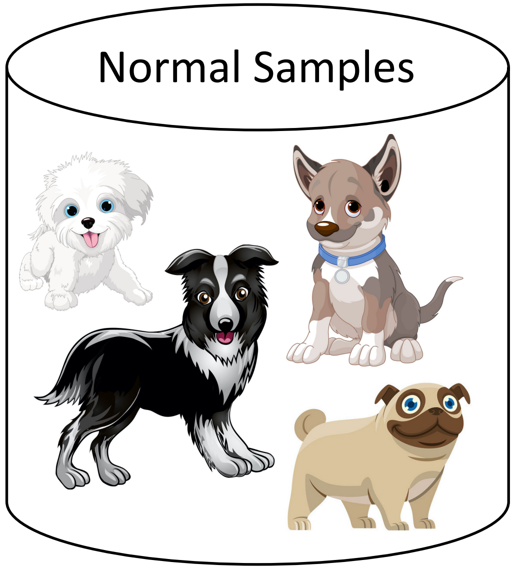
- **Solution**. Combine Variational Autoencoders and Generative Adversarial Networks

Methodology Overview

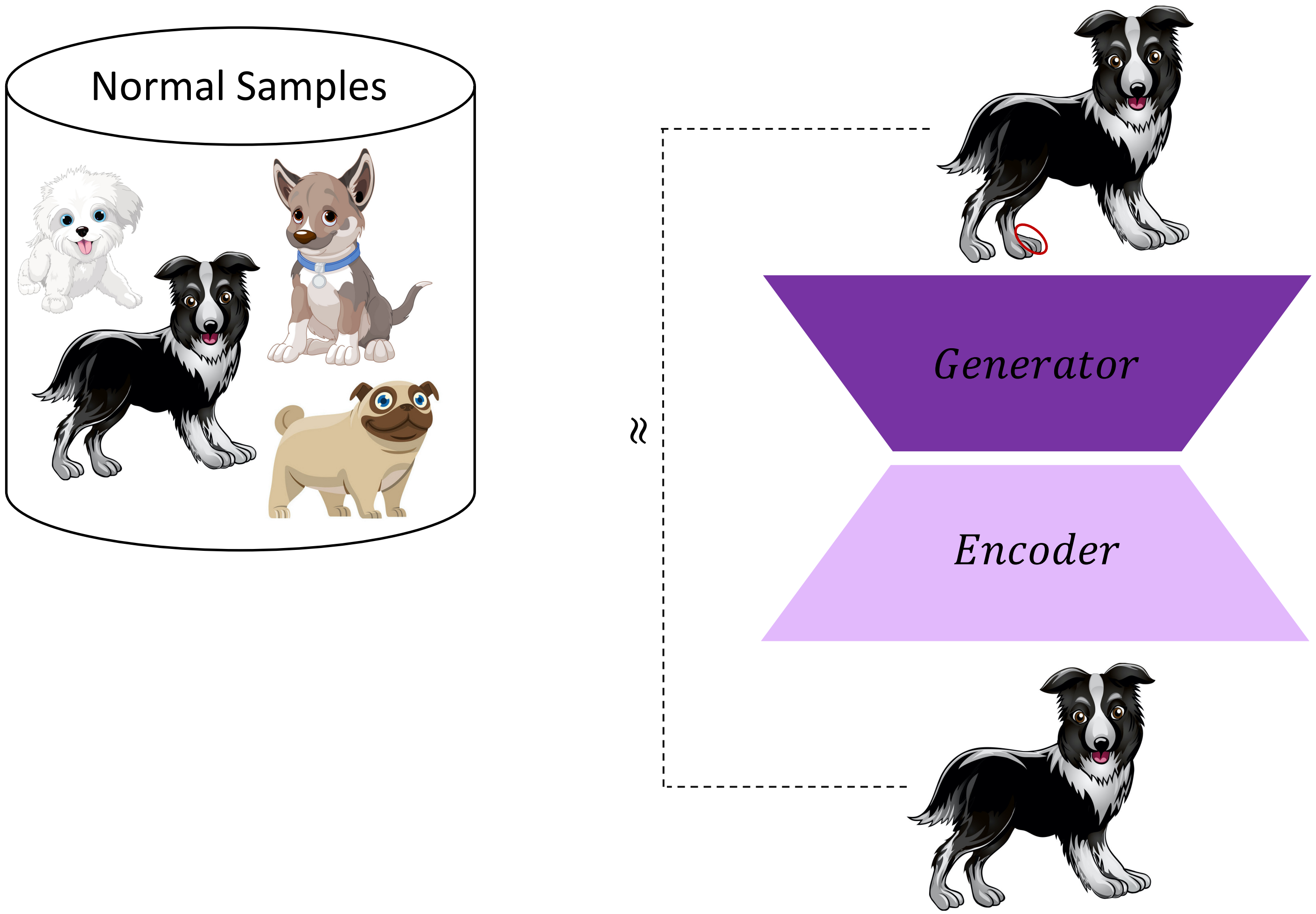
ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Methodology Overview

Adversarial Reconstruction Network

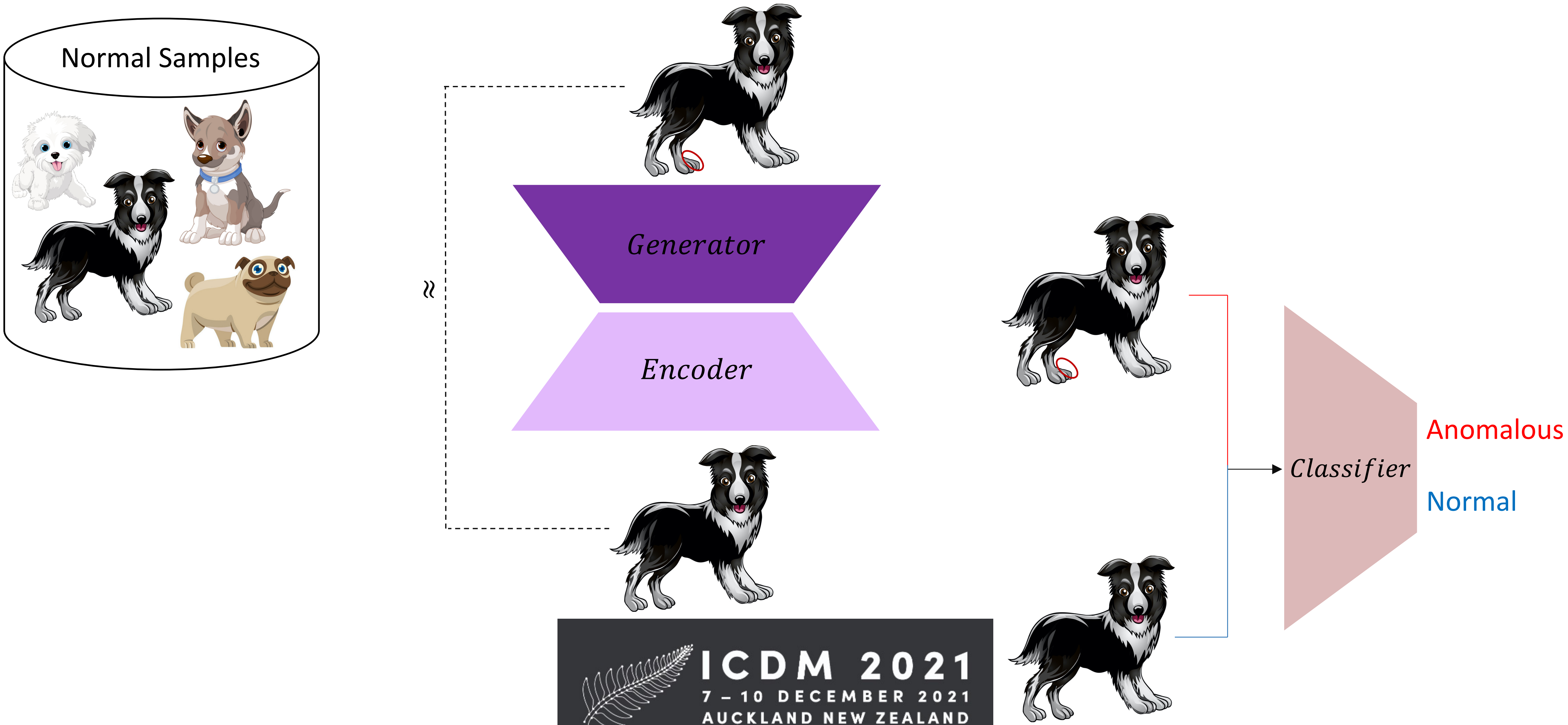


Adversarial Reconstruction Network



ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Adversarial Reconstruction Network



ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Adversarial Reconstruction Network

- The adversarial game has associated the discriminator loss

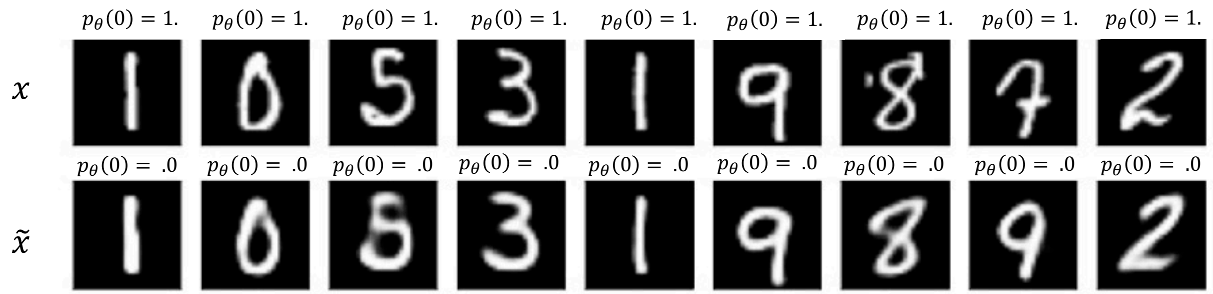
$$\mathcal{L}_{\mathcal{D}}(\theta|\phi,\psi) = \mathbb{E}_{x\sim\mathbb{P}_{\mathcal{D}}}[\log p_\theta(0|x)] + \mathbb{E}_{\substack{x\sim\mathbb{P}_{\mathcal{D}} \\ z\sim q_\psi(\cdot|x) \\ \tilde{x}\sim g_{\phi(z)}}}[\log p_\theta(1|\tilde{x})]$$

- and the generator loss

$$\mathcal{L}_G(\phi,\psi|\theta) = \mathbb{E}_{\substack{x\sim\mathbb{P}_{\mathcal{D}} \\ z\sim q_\psi(\cdot|x) \\ \tilde{x}\sim g_{\phi(z)}}}[\log p_\theta(0|\tilde{x})] + \mathbb{E}_{\substack{x\sim\mathbb{P}_{\mathcal{D}} \\ z\sim q_\psi(\cdot|x) \\ \tilde{x}\sim g_{\phi(z)}}}[\log p(x|\tilde{x})] - \mathbb{KL}\left[q_\psi(z|x)||p(z)\right]$$

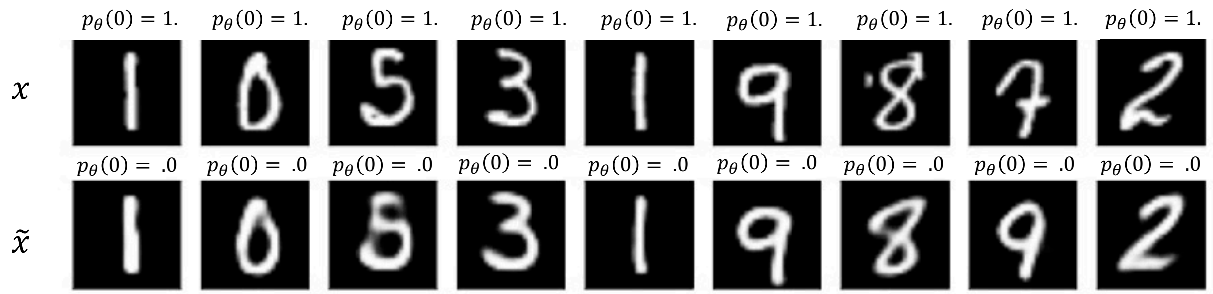
Experiments

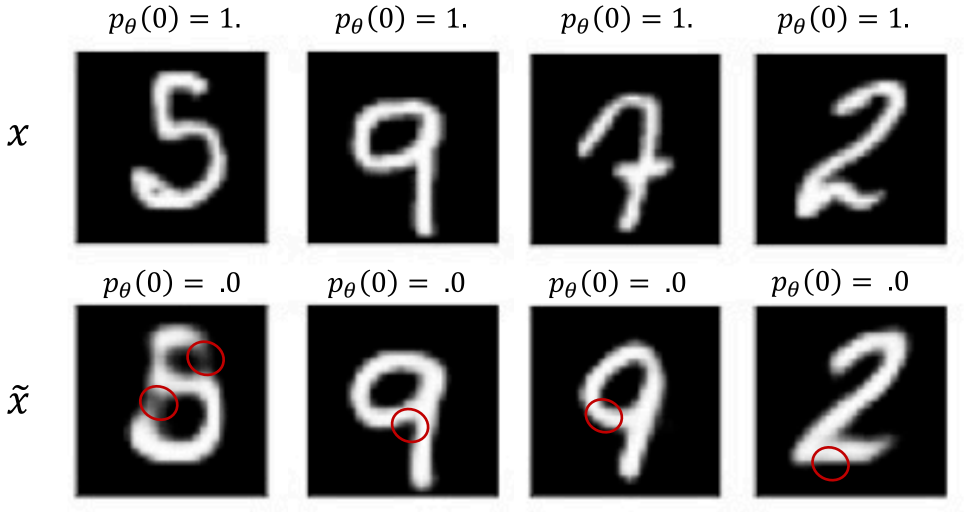
RQ1. Does the outlier generator produce realistic outliers? How does it affect the predictive power?



Experiments

RQ1. Does the outlier generator produce realistic outliers? How does it affect the predictive power?





ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Experiments

RQ2. In real-world scenarios, can the classifier component be used to predict unobserved anomalies? How does its predictive power compare to other state-of-the-art approaches?

Dataset	ARN^G		ARN^N		FenceGAN		GANomaly		OC-SVM		Baseline	
	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC	AUC	AUPRC
KDDCUP99	.99 ± .00	.99 ± .00	1.00 ± .00	.99 ± .00	.99 ± .00	.99 ± .00	1.00 ± .00	1.00 ± .00	.96 ± .00	.97 ± .00	1.00 ± .00	1.00 ± .00
KDDCUP99$_{Rev}$	.97 ± .01	.95 ± .02	**.99 ± .00**	.95 ± .02	.84 ± .01	.77 ± .01	.92 ± .01	.86 ± .01	.81 ± .00	.71 ± .00	.91 ± .01	.87 ± .01
KDDCUP99$_{Inv}$	1.00 ± .00	1.00 ± .00	1.00 ± .00	1.00 ± .00	.92 ± .03	.72 ± .08	.91 ± .04	.90 ± .03	.95 ± .00	.82 ± .00	1.00 ± .00	1.00 ± .00
NSL-KDD	.99 ± .00	.99 ± .01	.99 ± .00	.98 ± .01	.96 ± .00	.97 ± .00	.97 ± .01	.97 ± .01	.96 ± .00	.97 ± .00	.99 ± .00	.98 ± .00
DoH	.99 ± .01	1.00 ± .00	.99 ± .01	1.00 ± .00	.88 ± .02	.97 ± .00	.99 ± .00	1.00 ± .00	.88 ± .00	.97 ± .00	.96 ± .00	.99 ± .00
DoH$_{Inv}$	.98 ± .01	.97 ± .02	1.00 ± .00	**1.00 ± .00**	.89 ± .02	.44 ± .05	1.00 ± .00	.98 ± .01	.90 ± .00	.49 ± .01	.99 ± .00	.91 ± .04
CoverType	.94 ± .01	.95 ± .01	.92 ± .04	.93 ± .03	.70 ± .03	.41 ± .02	.56 ± .05	.30 ± .04	.73 ± .02	.43 ± .02	.53 ± .02	.28 ± .02
CreditCard	-	-	.99 ± .01	.59 ± .06	.90 ± .01	.51 ± .03	.84 ± .02	.36 ± .05	.92 ± .01	.57 ± .01	.99 ± .00	**.76 ± .01**
Bank	.77 ± .06	.63 ± .09	.69 ± .07	.50 ± .11	.56 ± .01	.23 ± .01	.53 ± .02	.22 ± .02	.60 ± .00	.28 ± .00	.65 ± .00	.32 ± .01

Experiments

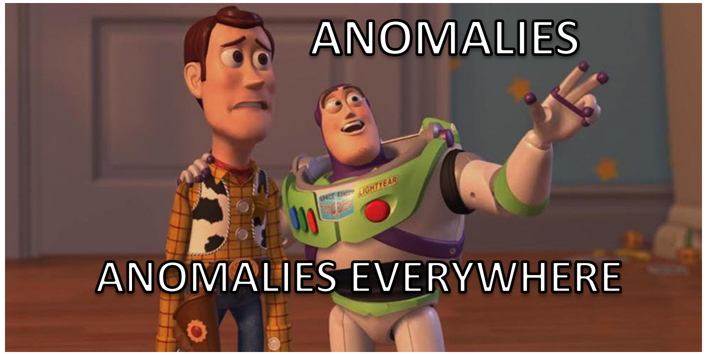
RQ3. Which components of the model contribute to the overall quality? How do the architectural choices affect the accuracy of the resulting predictions?

Dataset	$\mathbf{ARN}^{G}$	$\mathbf{ARN}^{N}$	$\mathbf{ARN}^{G-\mathbb{KLD}}$	$\mathbf{ARN}^{N-\mathbb{KLD}}$	$\mathbf{ARN}^{GE}$
KDDCUP99	.99 ± .00	1.00 ± .00	.98 ± .02	.98 ± .02	.74 ± 09
KDDCUP99$_{Rev}$	.97 ± .01	.99 ± .00	.98 ± .01	.96 ± .00	.83 ± .05
KDDCUP99$_{Inv}$	1.00 ± .00	1.00 ± .00	1.00 ± .00	1.00 ± .00	.88± .04
NSL-KDD	.99 ± .00	.99 ± .00	.99 ± .01	.98 ± .00	.74 ± .07
DoH	.99 ± .01	.99 ± .01	.73 ± .01	.79 ± .02	.99 ± .01
DoH$_{Inv}$	.98 ± .01	100. ± .00	.99 ± .00	.99 ± .00	.83 ± .04
CoverType	.94 ± .01	.92 ± .04	.94 ± .01	.94 ± .01	.77 ± .08
CreditCard	-	.99 ± .01	-	.96 ± .05	.75 ± .06
Bank	.77 ± .06	.69 ± .07	.74 ± .07	.63 ± .07	.62 ± .04

ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Conclusions and Future Work

- ARN: A twofold neural architecture aimed at generating and identifying anomalies
- Experiments prove the capability of the model in generating realistic outliers for enabling the learning of an outlier detector
- As future work, we plan to study a generalization of ARN towards a fully unsupervised setting
  - The proposed approach requires samples labeled as normal



ICDM 2021
7 – 10 DECEMBER 2021
AUCKLAND NEW ZEALAND

Thank you for your attention!