WP3.1 Robust Generative Neural Networks
UDRC
November 2021

- University of Edinburgh, UK

School of Engineering, Institute for Digital Communications (IDCOM)

- University Defence Research Collaboration (UDRC) in Signal Processing

RA: Dr Nikolaos Dionelis

Academics:

Dr Mehrdad Yaghoobi

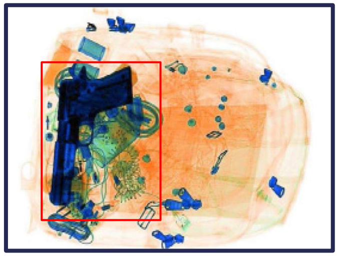
Prof. Sotirios Tsaftaris

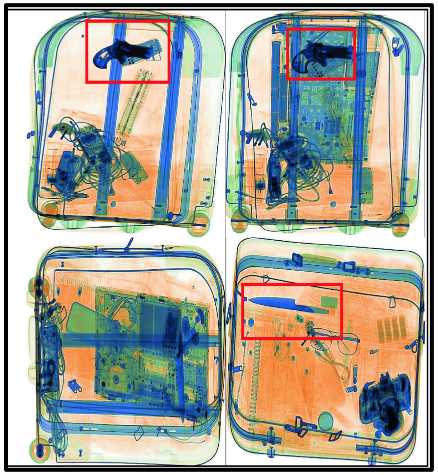
Dr Joao Mota

Dr Sen Wang

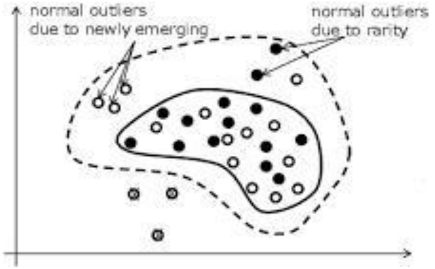
Current Work: Research in ML at UoE and UDRC

- Open-Set Recognition (OSR)
- **Few-shot** classification
  - Class-incremental learning
  - Cross-domain classification
- **Both recognition and OoD detection**
- Discriminative **and** generative models



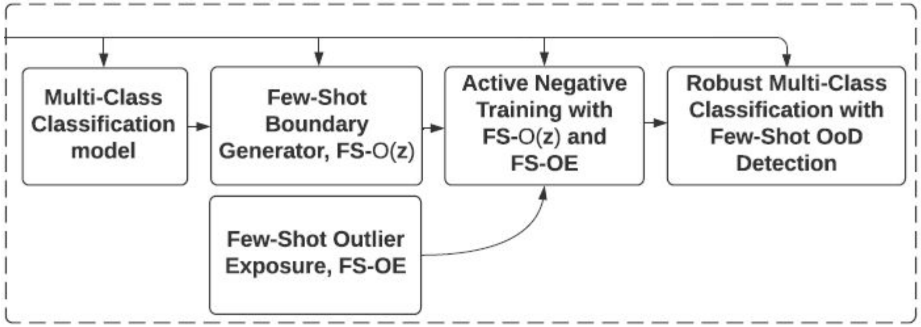


- **Main thrust of our research:**
  - 1) Classify objects in images
  - 2) Learn new objects fast with **few-shots**
  - 3) Identify **novel classes** as anomalies and learn them
  - 4) Maintain the capability of alerting the user for threats for **seen and unseen abnormal data**


normal outliers due to newly emerging
normal outliers due to rarity

Few-Shot Robust Classification and OoD Detection

- **Overconfidence:** Set high confidence to OoD samples away from training data
- Sample generation on the boundary
- Impose **low confidence** on boundary
- **Few-shot** OoD detection
  - **Robust to the number of few-shots**



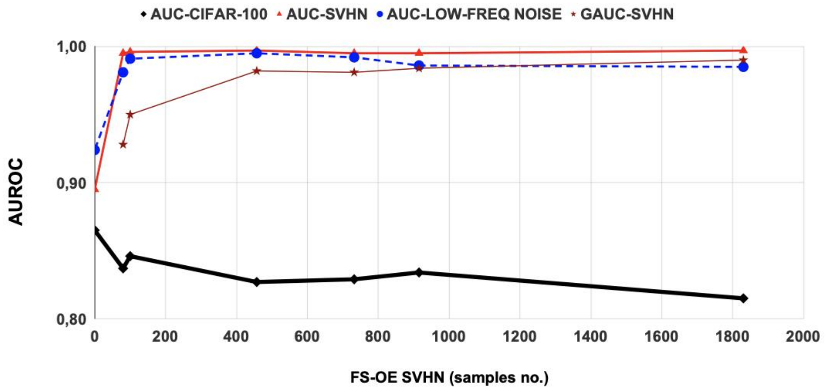
Multi-Class Classification model

Few-Shot Boundary Generator, FS-O(z)

Active Negative Training with FS-O(z) and FS-OE

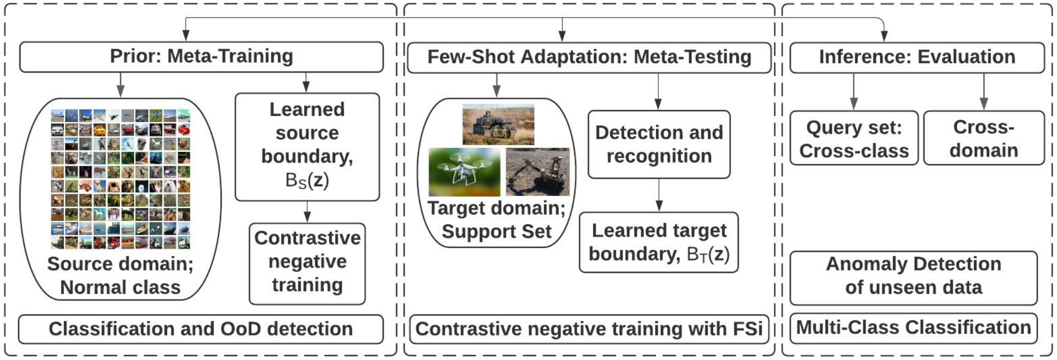
Robust Multi-Class Classification with Few-Shot OoD Detection

Few-Shot Outlier Exposure, FS-OE



Robust Few-Shot Class-Incremental OSR

- Learn a prior
- **Few-shot adaptation**
- Discern between **base classes, new FS classes, unknown OoD**



Prior: Meta-Training

Source domain; Normal class

Learned source boundary, $B_S(z)$

Contrastive negative training

Classification and OoD detection

Few-Shot Adaptation: Meta-Testing

Target domain; Support Set

Detection and recognition

Learned target boundary, $B_T(z)$

Contrastive negative training with FSi

Inference: Evaluation

Query set: Cross-class

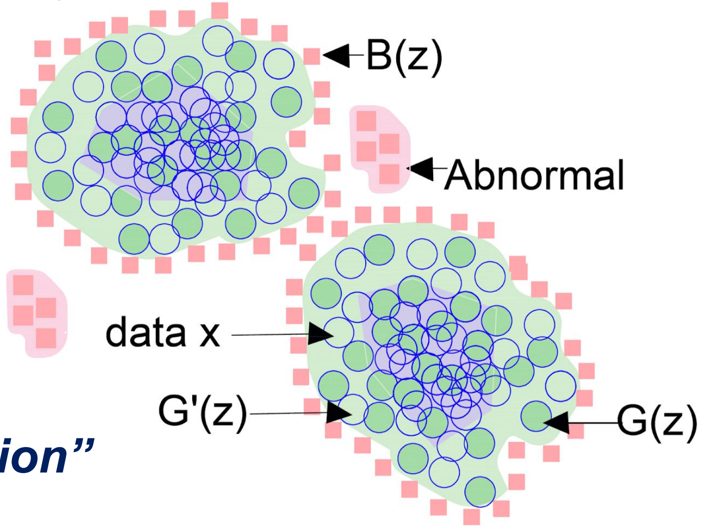
Cross-domain

Anomaly Detection of unseen data

Multi-Class Classification

Our Publication Outcomes

- ***"OMASGAN: Out-of-Distribution Minimum Anomaly Score GAN for Sample Generation on the Boundary"***
    - Contrastive negative training avoiding invertibility, 2021
- ***"REFGAN: Few-Shot Detection of OoC using GANs with Negative Retraining,"*** in Proc. ICTAI 2021
- GANs for detecting Objects of Concern with few-shots



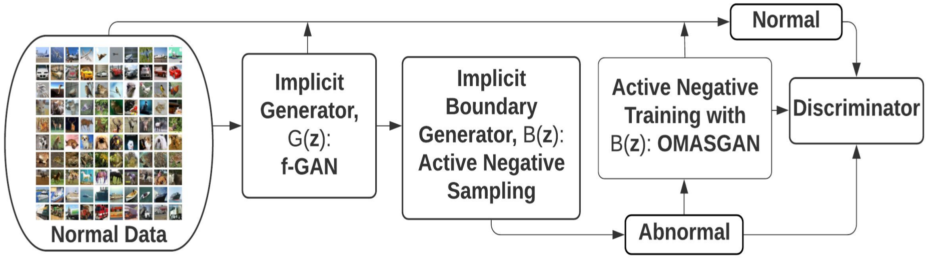
- ***"Few-Shot Robust Model for Classification and OoD Detection,"*** Submitted, 2021
- ***"Negative-Data Discriminative Classifier for Few-Shot Class-Incremental Open-Set Recognition"***

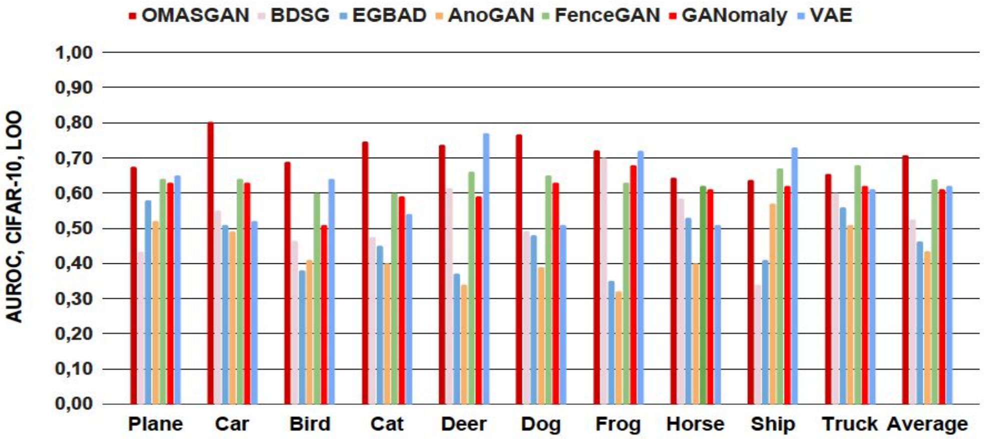
- Large scale MetaAudio paper
    - Benchmark and survey: Few-shot acoustic classification
- Multi-task learning
    - Cross-domain meta-learning

OoD Minimum Anomaly Score GAN

- **Rarity** of relevant OoDs: Learn directly from data only from the normal class
    - **Reduced** human intervention for supervision, e.g. feature extraction
    - Generate **minimum-anomaly-score OoDs**
        - Invertibility is **not necessary**

- **Retraining** by including OoD samples on the distribution boundary
    - Perform **self-supervised** negative data augmentation
- **Self-supervised learning:** Improve both unsupervised learning and AD
- **Evaluation:** Leave-one-out methodology
    - **Improvement** over benchmarks for AD


Normal Data
Implicit Generator, G(z): f-GAN
Implicit Boundary Generator, B(z): Active Negative Sampling
Active Negative Training with B(z): OMASGAN
Normal
Discriminator
Abnormal



“OMASGAN: OoD Minimum Anomaly Score GAN for Sample Generation on the Boundary,” 2021

Few-Shot Adaptive Detection of OoC: REFGAN

- **Robust** OoC detection
  - OoC: **Rare** & different from normality
    - Might be **unknown** during training
- Our proposed methodology:
  - Negative REtraining with Few-shots GAN (REFGAN)
- Learn a prior
- **Few-shot adaptation** of prior
  - Negative-data-based few-shot adaptation
- **Robust to few-shot samples**

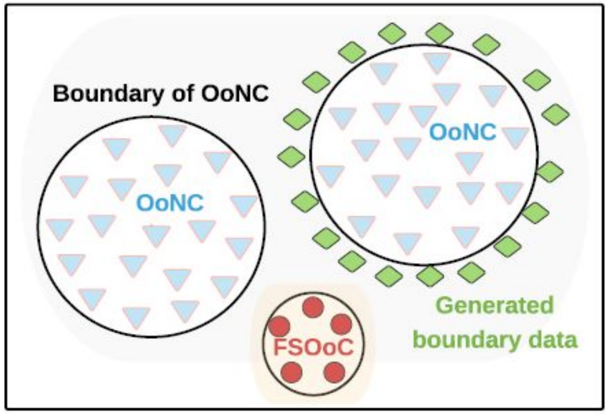


Fig. 1: REFGAN where the blue points are OoNC, the red points are FSOoC, and the green points are $B(\mathbf{z})$ samples.

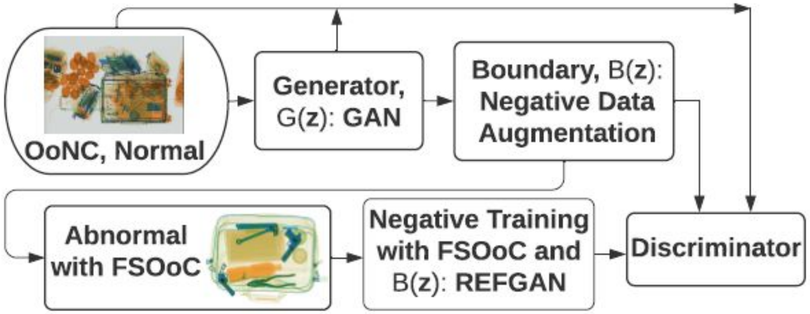
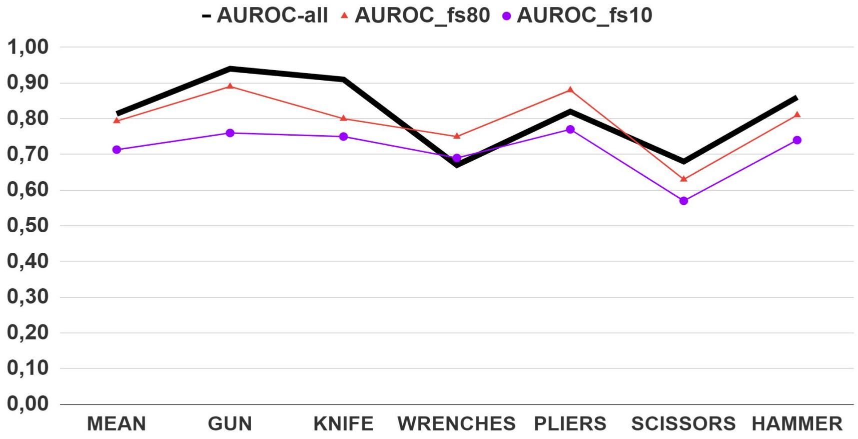

OoNC, Normal
Generator, G(z): GAN
Boundary, B(z): Negative Data Augmentation
Abnormal with FSOoC
Negative Training with FSOoC and B(z): REFGAN
Discriminator

Fig. 2: Training of the proposed REFGAN using the FSOoC samples, together with active negative sampling and training.



“REFGAN: Few-Shot Adaptive Detection of Objects of Concern using GANs with Negative Retraining,” ICTAI 2021