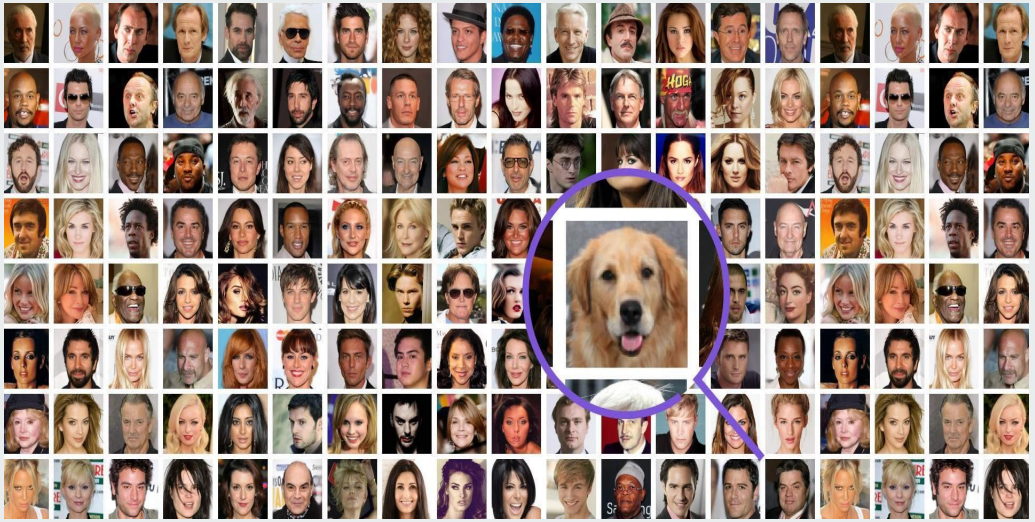
# Image/Video Deep Anomaly Detection: A Survey

By Tamanna Mehta



# Introduction:

- **Anomaly Detection** is a process of identifying/detecting unusual samples which seldom appear or do not even exist in the training dataset. These samples do not confirm the expected behavior and are thus called **outliers**. Anomalies occur very rarely in the data.
- Normal Data Instances follow target class distribution
- Anomalous Data Samples belong to out of class distribution are not present or rarely present at the expense of high computational cost.
- Deriving abnormal data leads to a very complicated learning process. So, researchers have tried to train models that are capable of classifying anomalous data from normal data.

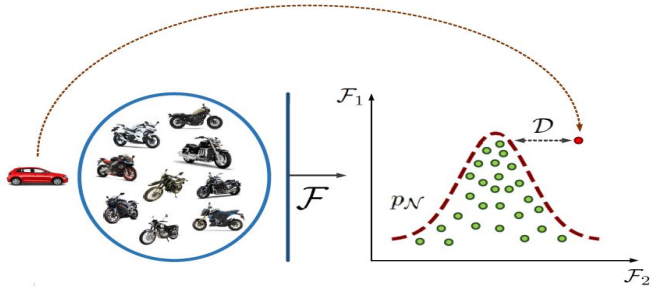



Figure 1: The general concept of AD is depicted in this figure. Here, motorcycles are considered as normal instances while the car is anomaly. $\mathcal{F}$ demonstrates a representation of the given data for analysis. For simplicity, samples are shown in two dimensions using $\mathcal{F}_1$ and $\mathcal{F}_2$ feature vectors. As is clear, motorcycles, which are denoted by green dots, follow the distribution of target class (normality), *i.e.*, $p_{\mathcal{N}}$. Therefore, an out-of-distribution instance (car in our case), which is represented by a red dot, has a deviation from the normal data calculated by a specific detection measures, *i.e.*, $\mathcal{D}$.

# Weakness of AD Algorithms:

1. High False Positive Rates
2. High Computational Cost
3. Unrealistic Dataset

Many deep learning-based solutions have been proposed in this area. In-depth investigated and most widely used approaches have considered shared concepts of normal data as a distribution or a reference model

# Problem Formulation:

$$\mathrm{AD}(\mathcal{F}(y)) \begin{cases} \text{Normal} & \mathcal{D}(\mathcal{F}(y), p_{\mathcal{N}}) \leq \tau \\ \text{Anomaly} & Otherwise \end{cases} \qquad (1)$$

$U(X_n)$-Unlabeled Images or video frames follows normal distribution

$p_N$-Normal Data distribution

F-Feature extractor that maps raw data to a set of discriminative features

D-metric used to compute the distance between a given instance and normal data distribution

# Anomaly Detection techniques:

1. Supervised (N+A):
2. Semi Supervised (N+A+U)
3. Unsupervised (U)

Where N-Normal data samples

A-Anomalous data samples

U in Unlabelled samples

# Supervised  Technique (N+A):

1. A CNN is made to learn on (N+A) samples in a supervised way to accurately make a distinction.
2. Supervised modeling produces high accuracy, it doesn't lead to generalized outcomes.
3. The performance of the supervised model is not optimal due to an imbalance in the dataset.
4. Due to diversity in anomalies, the training procedure is disturbed making it practically infeasible.

**Example-Accident Detection Applications**

# Semi Supervised Technique(N+A+U):-

1. Learn a model on numerous unlabelled samples along with a few normal and anomalous instances ( N+A<U).
2. Having access to both normal and abnormal events in most of the AD applications is practically impossible
3. Collecting such data is tedious and computationally expensive because of diversification in anomalies and their rare occurrences.

# UnSupervised Technique(U):-

1. Model is trained on only unlabelled data samples.
2. Outliers are detected based on the intrinsic properties of data samples.
3. It is assumed that just like realistic situations, abnormal events occur very occasionally in unlabelled samples.
4. Unsupervised methods are called as One-Class Classification or OCC (in short)
5. OCC involves fitting a model on the “normal” data and predicting whether new data is normal or an outlier/anomaly.

# Deep Image/Video Representation:

## Traditional Features

- Handcrafted
- trajectory-based
- low-level features such as Motion Boundary Histogram (MBH), Histogram of Gradient.

### Weaknesses :

high false-positive rates, low performance, computationally expensive, and inadequate in discriminating normal and abnormal data samples

### Video-Spatial & Temporal Features

**Eg-RNN, LSTM,3D CNN**

## Deep Features:

***Feature Learning:-***

Auto Encoders using Learned Features

***Pretrained Networks:***

- Transfer learning
- fine-tuning the hyperparameters

# Deep Networks for Anomaly Detection: End to End DNs

***Self Supervised Learning:-***

- Model access normal data or data with minimal abnormalities
- Researchers have tried to learn D and pN implicitly in order to a train model end to end on such data. DNN is trained under specific constraints to learn pN.
- A test sample say X, does not follow pN is if it does not satisfy the desired constraint and thus can be considered as an anomaly.
- Minimize Reconstruction Error, forcing latent representation to be sparse are well-known self-supervised tasks to learn distribution (pN).

# Deep Networks for Anomaly Detection: End to End DNs

**Encoder-Decoder Networks:**

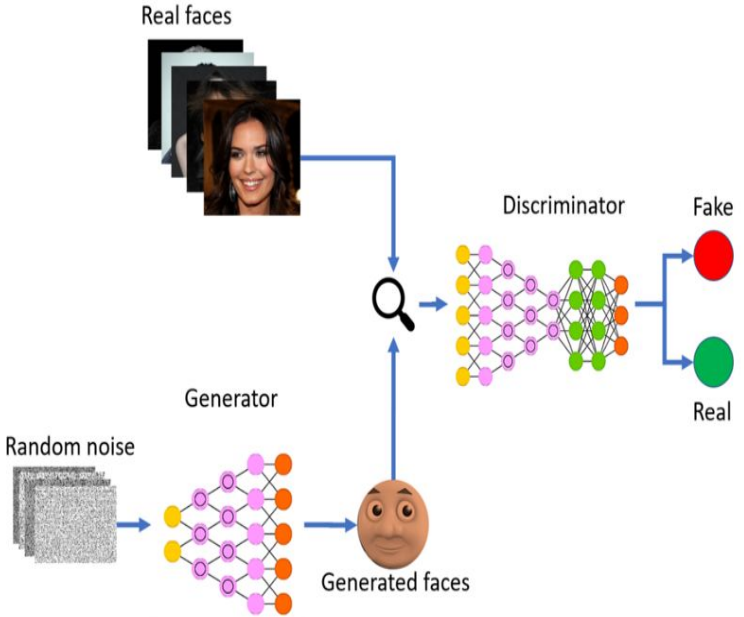
An encoder-decoder model generally consists of three parts: t**he encoder, the latent space representation, and the decoder.** The purpose of the encoder network Enc is to transform the input data into a latent space representation z that is often a vector; the decoder network Dec then produces the output by decoding z. During training, the encoder and the decoder are trained together to minimize the empirical risk. Proper design of the latent space representation z is crucial to the successful application of the encoder-decoder approach.

$$\mathbf{L} = \frac{1}{m} \sum ||X^i_{\mathcal{N}} - D(E(X^i_{\mathcal{N}}))||^2 \qquad (2)$$

D(E(X)) is an encoder-decoder network that learns normal distribution pN and not the anomalous data. In order to reconstruct normal instances, the parameters of D(E(X)) are optimized. If a reconstructed sample has a high Reconstruction Error(RE)it shows an anomaly.

# Generative Networks:

- Initially, G and D are naive to learning. So they take random decisions.
- Generator G is supplied with latent distribution(random noise) as input.
- Discriminator D is provided input with both real images and input from Generator.



# Generative Networks:

- G aims to generate data samples with the same normal distribution in order to fool the Discriminator so as to manipulate it to detect G(X) which is real data.
- Whereas, D tries to make a distinction between real image and data generated by G.
- Trained Adversarially

- During training, G tends to generate fake (anomalous) data for D
- D is trained to classify normal and abnormal image/video frames. D is a binary classifier.
- Eventually, D becomes capable of acting as a one-class classifier. G acts as an encoder-decoder as it recreates normal data instances.

Objective Function:-

$$\min_G \max_D \Big( \mathbb{E}_{X \sim P_{\mathcal{N}}} [\log(D(X))] + \mathbb{E}_{\tilde{X} \sim P_{\mathcal{N}} + \mathcal{N}_\sigma} [\log(1 - D(G(\tilde{X})))] \Big) \quad (3)$$

## Problem using GANs:

- In recent researches, G not only recreates normal data frames but also used them for pre-processing in order to improve the performance of D as end to end anomaly detector.
- Expensive training
- Instability

# Anomaly Generation:

- Binary classification problem - GANs can be used to generate abnormal data instead of directly using it.
- Idea was presented by Masoud Pourreza,
- The aim of this work is to train Wasserstein GAN on normal instances and exploit G before complete convergence.
- G generate Irregular data along with normal data from a training set that can be used for AD tasks.

## Datasets Used For Image/Video Anomaly Detection:

### *Image Datasets:*

- MNIST dataset (28 X 28) grayscale images of handwritten digits from 0–9 (10 classes)
- CIFAR-10 and CIFAR-100 consist of images (32 X 32) with 10 and 100 classes
- Caltech-256 included 30,607 images with 256 object categories

### *Video Datasets:*

- UMN dataset-normal (people wandering around) and abnormal events (running)
- UCSD Pedestrian 1 and Pedestrian 2 datasets
- 158 X234 and 240X360. Normal objects are pedestrians and cars, bicycles and skateboarders are anomalies.
- CUHK Avenue containing 47 abnormalities and UCF-Crime dataset

## Challenges & Future Directions:

1. **False-positive rate:** There are solutions to AD tasks with great accuracy in detecting outliers but they come with very high false-positive rates. Ideally, the solution must have accuracy with low false-positive rates.
2. **Fairness:** Skewed datasets, limited features are responsible for unfairness in AD. This is primarily due to the insufficient availability of anomalous data.
3. **Safety**: A minor manipulation to input sample confuses DNN in a way that misclassifies the input data. Thus, DNNs are prone to adversarial attacks.
4. **Realistic Datasets:** Datasets used for AD tasks are far from realistic situations.
5. **Early Detection:** Proposed Solutions correctly detect anomaly which is either over or near to end. Late Detection of an anomaly in the case of videos is unacceptable. So, early detection of such events is highly critical. A well-timed alarm can be beneficial in minimizing or preventing loss caused by anomalous events occurrences.

# References:

[1] https://medium.com/@rajatgupta310198/generative-adversarial-networks-a-simple-introduction-4fd576ab14a

[2]https://openaccess.thecvf.com/content/WACV2021/papers/Pourreza_G2D_Generate_to_Detect_Anomaly_WACV_2021_paper.pdf

[3]https://www.sciencedirect.com/science/article/abs/pii/S1361841518302640

**Reference paper**-https://arxiv.org/pdf/2103.01739.pdf