Distributed by:

SRS GLOBAL SERVICES

www.srsglobalservices.com

Report on the performance of various modules of Scylla AI Physical Threat Detection Solution

Whitepaper

Table of Contents

Abstract ›

Introduction ›

Baseline Installation & Configuration ›

Object Detection ›

Intrusion Detection & Perimeter Protection ›

Car Detection & Parking Area Monitoring ›

Anomaly Detection & Behavior Recognition ›

Face Recognition ›

Abstract

Scylla is a modern modular AI-powered threat detection system that enhances operational activities of security teams in five main areas:

- Object Detection (SKU: PTD)
- Anomaly Detection and Behavior Recognition (SKU: ADS)
- Thermal Screening (SKU: BTD)
- Intrusion Detection and Perimeter Protection (SKU: IDS)
- Drone Security (SKU: ASU)

Exclusive **benefits** that the solution offers:

- ➔ Scylla provides a unique security mechanism that has no counterparts in the world. It eliminates the negative effect of human-factor errors, enhances and augments the effectiveness of security personnel.
- ➔ Scylla reduces the response time of security units in an emergency. Every second can play a decisive role in saving lives.
- ➔ Scylla reduces security costs.
- ➔ Scylla seamlessly integrates into the existing security system, without requiring technical changes and additional costs.
- ➔ In the core of each module of Scylla, there are AI and machine-learning algorithms that work autonomously 24/7 and comprise the engine of the solution suite. Unveiling the higher-order logic of each of these modules as well as characterizing and reporting the accuracy is the scope of this paper.

Introduction

Scylla is a leading physical threat detection solution that aims to prevent crime before it happens. It is a revolutionary security system developed based on artificial intelligence and machine learning. Its goal is to detect and prevent crime and violence in public areas using existing CCTV cameras.

Scylla uses Artificial Intelligence to analyze data coming from stationary and portable cameras. It then processes the acquired data to generate meaningful information about the activity contained in the video stream. Matching the results of data through our Smart Decision Making algorithm Charon, Scylla identifies the suspect through facial recognition, as well potential threats posed by a suspect's actions or vehicle in the video feed. Finally, Scylla analyzes the collected data and distributes alerts to law enforcement response units through our web and mobile channels with information about the threat, location, and the identity of the suspect.

Scylla is powered by state-of-the-art algorithms that enable first responders and security teams to obtain instant identification and detection of security threats through a centralized dashboard. The system is capable of biometric identification of suspects. Furthermore, Scylla is capable of performing continuous scanning and search operations to identify and track suspects, most wanted individuals, registered sex offenders as well as missing persons nationwide. Scanning can be done both in real-time as well as in a forensic operation mode post-factum.

Based on the algorithm core Scylla modules can be divided into three groups: “object detection” (including classification and tracking), “action detection” and “hybrid”.

Object detection includes solutions like **Intrusion Detection**, **Occupancy Counting**, **Smart Parking**, **Intelligent Traffic Management**, etc. **Scylla Thermal Screening** also uses this engine for smart targeting.

Action detection is used for solutions like **Fight/Violence/Assault detection**, **Smoke** and **Fire detection**, **Shoplifting detection**, etc.

The hybrid of the two approaches is used in Object Detection where not only the object (weapon) detection matters but also the handling of it (action).

The solutions provided by Scylla are outstanding in a number of ways. To name a few:

- Solutions use proprietary analytics that works with moving backgrounds
- Solutions are flexible, easy to install, and intuitive to use
- Solutions use several times less hardware
- Can be installed locally or on cloud

and more.

Due to indifference towards moving backgrounds, different Scylla modules are ready to analyze video feeds from moving cameras, such as UAV-mounted or body-worn. In such implementations Scylla helps to monitor peripheral activity and help apprehend criminals. Scylla powered drones can augment security teams in mobile suspect chase operations. Finally, to help prevent acts of terrorism, Scylla tracks movement and identifies rogue activities and individuals. It can detect use of suspicious items including but not limited to different weapon types, track and alert on cargo movements, unattended bags, etc. It also supports tracking and localization of humans based on their appearance/outfit over time. The same tracking algorithm is initiated after a threat is detected when search is performed for armed humans on all other indoor cameras connected to Scylla, where any sighting of the person is highlighted on an indoor map across the entire video feed timeline and an alert is sent to operational units.

Scylla provides a unique security mechanism that is unmatched in its accuracy. It significantly reduces the response time in an emergency where every second can play a decisive role in saving lives. Additionally, using Scylla significantly cuts down security costs and adds a better understanding of blind spots and activity previously not detected by the naked eye. The system seamlessly integrates into the existing security infrastructure, without requiring major technical changes and additional costs. Scylla AI is additionally trained to identify weapons on IR images from night-vision enabled cameras. Lastly, Scylla is powered by Charon, a proprietary AI algorithm which plays the “trial by jury” role and is capable of constant evolution and improvement.

Baseline Installation & Configuration

In the most common installation architecture, Scylla suite initiates activation of the violence/threat detection module responsible for monitoring and analyzing video footage 24/7 on all connected cameras. The moment a threat is detected, Scylla sends alerts to all assigned endpoints. These include the main dashboard and the mobile devices in possession of Security Officers in the area under surveillance by Scylla. After that an optional Asset Tracking module can be activated manually. When this sequence is triggered, all cameras in the network switch to the “person identification and tracking” mode where they look for the subject identified as an attacker based on his/her appearance. The moment the attacker is spotted, Scylla creates an alert with relevant location information and sends it to Security Officers. The indoor navigation map in Scylla Mobile App is updated accordingly with detection frames, locations, and timings to facilitate steps of threat neutralization actions.

These complex actions are developed and polished in collaboration with the best state-of-the-art specialists and Special Operations advisors in the field. The architecture of Scylla protective suite is called to minimize the impact of potential threats, as well as to locate and assist in neutralizing the source of violence by employing AI-augmented counteraction scenarios. However, there are crucial and naturally occurring questions during every acquaintance with the solution, e.g. how robust Scylla modules are at doing their job, how good Scylla is at detecting and recognizing the objects of interest, etc.

To address these questions and back them up with exact figures, we have carried out extensive tests on each module. We present the results of each test in the sections below.

Note that the AI-based solution is inherently probabilistic. In case of object and action detection one usually faces a trade-off between counteracting goals expressed in two critical model evaluation metrics: **precision** and **recall**. While **recall** characterizes the sensitivity (ability to capture the events of interest), **precision** defines the rate of mistakes (the so-called false positives). With the development of Scylla solution we managed to achieve exceptional values and an optimal balance between both metrics.

Object Detection

Different Scylla solutions use proprietary Object Detection engine at some point of their analysis. It can be detection of a person in Intrusion Detection solution, detection of Face in Scylla Temperature Scanning or Face Recognition solution or detection of weapons in Scylla Preventive Threat detection solution.

We will assess the performance of this engine by probing first the flagman Scylla PTD.

Object detection in security monitoring environments

The detection of dangerous objects in CCTV is a difficult problem to solve. Nowadays, the security based on CCTV requires at least one person to be constantly monitoring everything that happens on one or more monitors simultaneously. This sometimes leads to overlooked threats and a delayed response to dangers. According to Velastin et al. [1] typically after 20 minutes of CCTV monitoring, operators often fail to detect objects in a video scene. Ainsworth [2] goes deeper into the analysis: after 12 minutes of continuous video monitoring, an operator is likely to miss up to 45% of screen activity and after 22 minutes of viewing, up to 95% of activity is overlooked. Therefore, combining human attention with a real-time detection system for dangerous objects can be highly valuable.

[1] S. A. Velastin, B. A. Boghossian, M. A. Vicencio-Silva, A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, Transportation Research Part C: Emerging Technologies 14 (2) (2006) 96–113.
[2] T. Ainsworth, Buyer beware, Security Oz 19 (2002) 18–26.

What Scylla Preventive Threat Detection offers is 24/7 accurate unbiased overwatch on the whole network of CCTV cameras. Given the weapon is visible and distinguishable for a human, Scylla detects it typically within the first 1-2 seconds. If the lighting conditions, camera quality, angles, and viewpoints are not optimal, detection happens within the subsequent 2-6 seconds. Moreover, for sophisticated high-resolution cameras placed, for instance, on drones, Scylla detects the threat even when it is hardly distinguishable, if ever, by the human eye on a typical monitor. The patented "zooming and tracking" technology it utilizes enhances the full capability of the surveillance unit thus augmenting protection capabilities.

Another equally important question is how versatile the detection module is at detecting, analyzing, and identifying other similar objects and distinguishing them from the designated sought-after objects it is trained to detect. Let's assume one is looking into CCTV frames trying to spot a weapon. A typical CCTV camera operates at 25 frames per second, so during 10 hours of the active daytime period, the system scans through **900,000** (!) frames. If one expects to have a maximum of one False Positive a day, then the False Positive Rate (FPR) (the ratio of False Positives and all "no-threat frames" to total frames) should be as low as 1.1 x 10-6. In other words, the system specificity should be 99,99988%! Such level of specificity is very ambitious to achieve with AI-based solutions. It is complicated even for an intelligent human being who is required to browse through all samples "manually." And, at Scylla, we proudly claim to have achieved it! Our system running in a very crowded office environment with 15-20 people regularly in the view typically makes no more than one mistake in 3 days (that's one False Positive per approximately every 2.7 million frames). Thus, Scylla's specificity is estimated to be more than **99,999963%**. Most experts agree that it's many nines after comma for an AI-based Computer Vision Detection solution.

Model testing

Now let's look at one of the base components of PTD - the preliminary Object Detection model. We assessed the model using several databases available online and compared it with results obtained by competing technologies. Below is a summary per model and databases.

YOLO

We compared the performance of it with the infamous YOLO - an open source object detection algorithm that is being used in 95% of cases. The table below summarizes our comparison results on the COCO Validation set:

Object detection model	Input size	FPS	mAP
YOLOv3-SPP	608	73	42.9%
YOLOv4-P5 (input size 896)	896	41	51.8%
YOLOv4-P6 (input size 1280)	1280	30	54.5%
YOLOv5x (input size 640)	640	167*	50.1%
ScyllaNet (input size 416)	416	340	50.3%

*YOLOv5x was tested at Float Precision 16 (vs. all others at 32 Float Precision)

Note that transferring to FP32 usually results in a 2-fold drop of FPS. Thus ScyllaNet performs more than 4-fold faster over the latest YOLOv5x at the same accuracy rates.

In addition to the above state-of-the-art metrics for each of the models Scylla uses, the algorithms the models are engaged in, the preprocessing and postprocessing as well as the final decision making stages are thoroughly designed and tailored for real-time speeds and production grade accuracy. The algorithm of Scylla incorporates the “zoom-in and tracking” logic where the system focuses on a particular object and double-checks it before making the final decision. In some cases, this results in a sub-second delay in decision making*.

Salamea test

Next we compared the results with those obtained Salamea at al. [3] Authors trained their own 2-step gun detection model to refine the metrics and reported the accuracies. We run our model on the dataset kindly provided by authors. In the table below we present numbers for both:

	Author’s	Scylla Test on “Shooting”
TP	1175	6340
FN	186	529
FP	192	51
Precision	**0.86**	**0.992**
Recall	**0.86**	**0.923**
F1Score	**0.86**	**0.96**

Statistics and Examples

To estimate and report the accuracy of the weapon detection module, we have tested the algorithm in different conditions. Among the variables of the module there are

1. **The weapon**
   1. 1. Gun (Desert Eagle 44 Magnum replica)
   1. 2. Rifle (HK G36, AK 47)
   1. 3. Shotgun (Mossberg 500 replica)
2. **The environment**
   2. 1. Well-lit with a light background
   2. 2. Average-lit with a dark background
   2. 3. Night vision mode*
3. **The distance**
   3. 1. 4 meters
   3. 2. 8 meters
   3. 3. 12 meters
   3. 4. 35 meters - for drones
4. **The position of the weapon**
   For the position of the weapon we have defined 5 vertical angles (at -90°, -45°, 0°, 45°, 90°) between the horizon and the pointing direction, and 7 horizontal positions (at -135°, -90°, -45°, 0°, 45°, 90°, 135°) between the direction facing the camera and the weapon-pointing direction, except when pointing straight up and down (-90°, 90°) in which case we neglect the horizontal angle. We derived the sample of results from 23 directions tested in total. The binary tests of each pointing position return a 1 or a 0, with the result of 1 indicating detection within the specified time and 0 indicating a miss. The average of all tests represents the overall sensitivity coefficient, presented in tables below. Note that detection sensitivity at “common, natural directions,” such as horizontal pointing of a gun/rifle is much higher, resulting in almost 100% of detection rate. However, to assess and demonstrate the **overall** versatility and sensitivity of the system, we have chosen to include less common angles as well.

We gated the detection time from the moment of drawing a weapon (based on motion detection) to the detection report time. In some cases, the test target would move with the weapon pointed in the specific direction keeping distance from the camera and other conditions constant.

Notes:
** when detection happened later than within the first 2 seconds for some pointing directions (see above), the sensitivity coefficient was calculated taking into account faster detections for other pointing directions.*
*** for night-vision mode distances of only up to 8 meters were tested.*

Results of tests in well-lit, light background environments:

	Detection time (distances: 4m / 8m / 12m)		
Weapon type	0-2 seconds	2-4 seconds	4-8 seconds
Gun	0.98 / 0.95 / 0.92	1.00 / 0.98 / 0.98	1.00 / 0.99 / 0.98
Assault rifle	0.98 / 0.96 / 0.93	1.00 / 0.99 / 0.96	1.00 / 0.99 / 0.97
Shotgun	0.96 / 0.93 / 0.90	0.98 / 0.93 / 0.94	0.98 / 0.97 / 0.96

Results of tests in fairly-lit, dark background environments:

	Detection time (distances: 4m / 8m / 12m)		
Weapon type	0-2 seconds	2-4 seconds	4-8 seconds
Gun	0.93 / 0.92 / 0.89	0.96 / 0.96 / 0.93	0.99 / 0.97 / 0.93
Assault rifle	0.94 / 0.93 / 0.90	0.96 / 0.96 / 0.94	0.98 / 0.97 / 0.94
Shotgun	0.93 / 0.92 / 0.90	0.96 / 0.95 / 0.90	0.98 / 0.95 / 0.90

Results of tests in night vision (IR) surveillance mode:

	Detection time (distances: 4m / 8m)		
Weapon type	0-2 seconds	2-4 seconds	4-8 seconds
Gun	0.88 / 0.75	0.94 / 0.89	0.96 / 0.93
Assault rifle	0.91 / 0.86	0.92 / 0.90	0.98 / 0.96
Shotgun	0.90 / 0.86	0.93 / 0.87	0.98 / 0.95

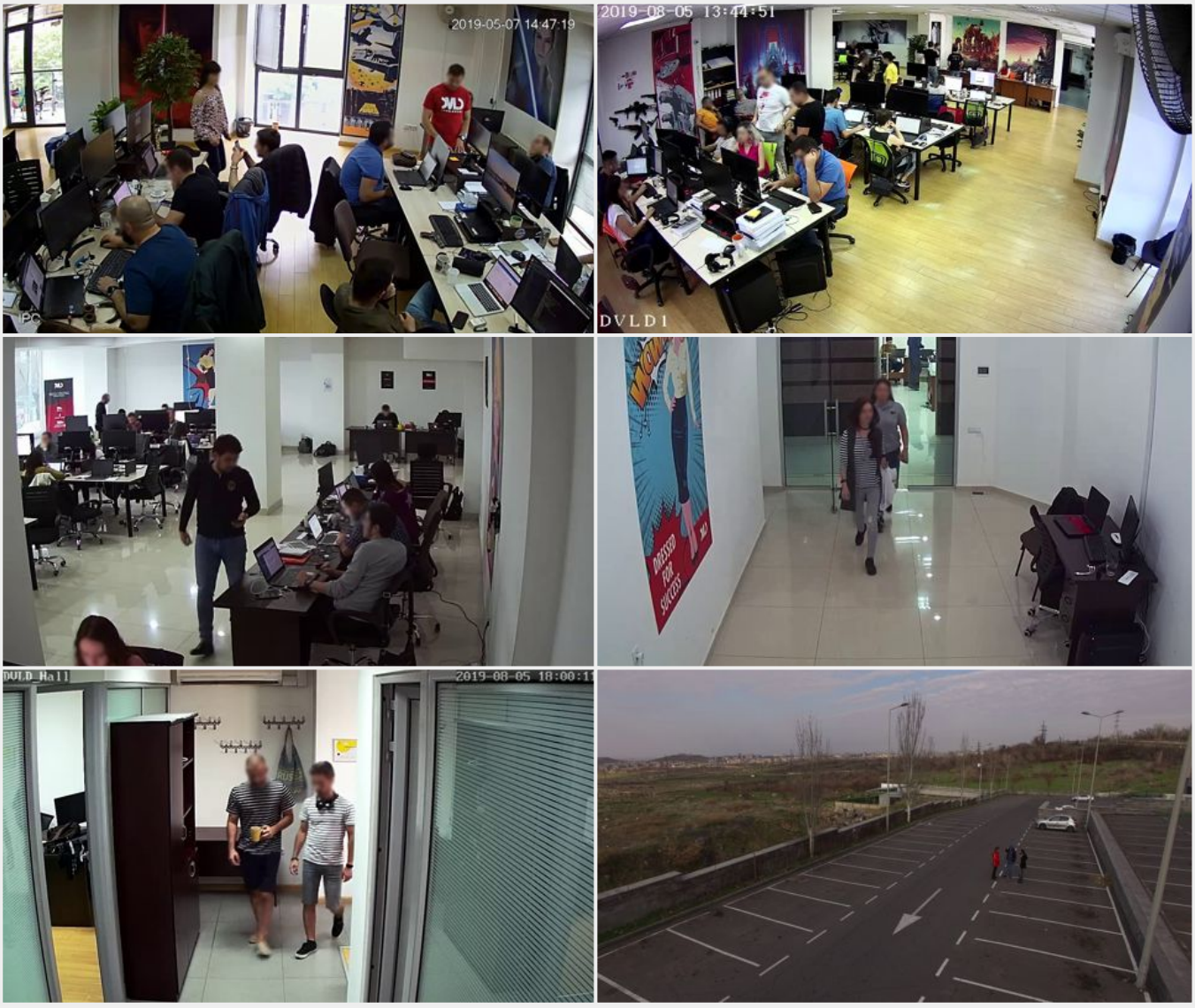
Samples of detections are available here.

Examples of detection from an IR-mode camera

Examples of detections from a Drone Camera (the video can be found here)

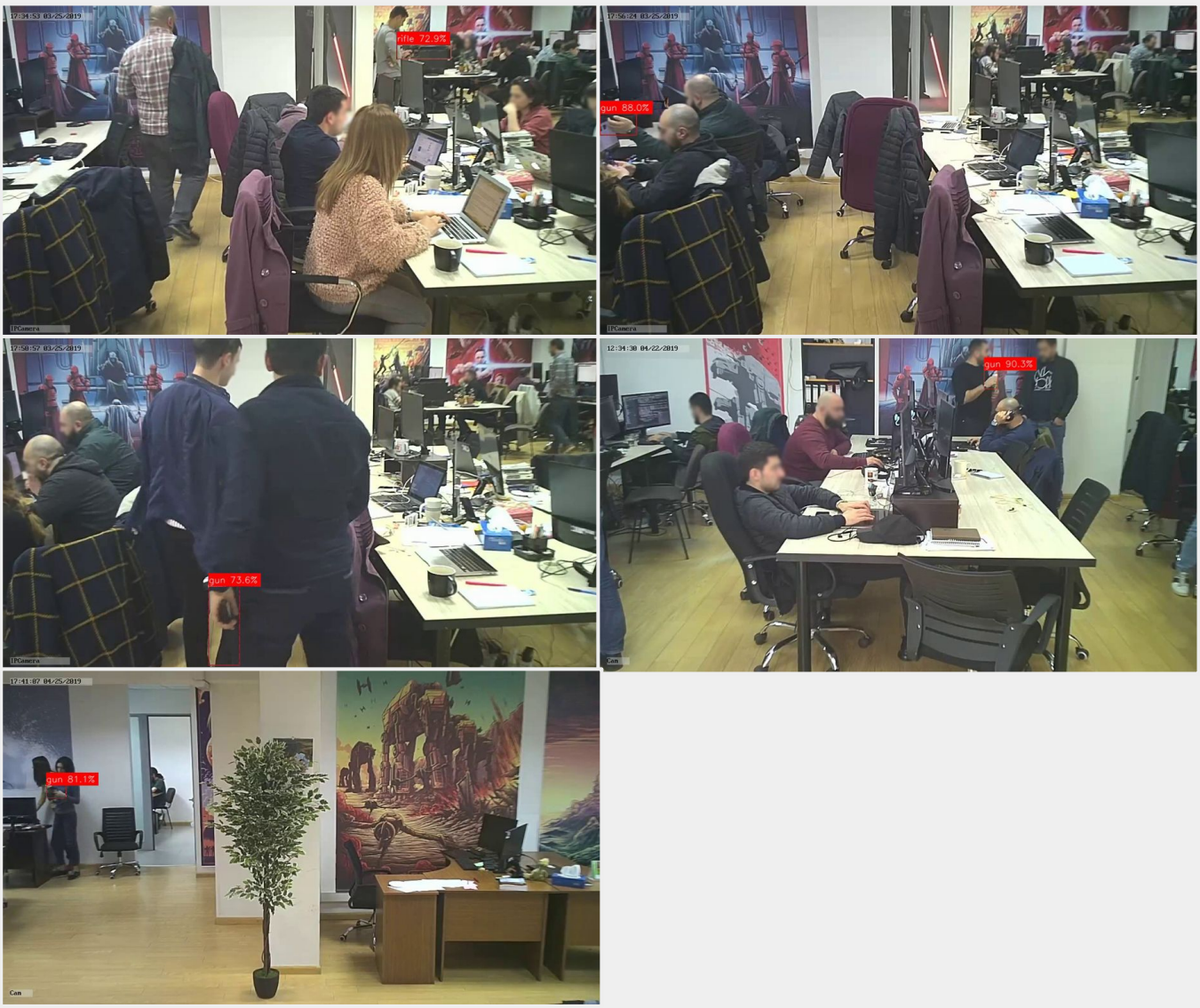
False Positive Rate

To address the False Positive Rate issue, we have tested the system in a few offices and outdoor installations (see sample camera viewpoints below):



We have considered office working hours only with the presence and activity of office personnel within the test space and environment encompassing diverse activities, settings, and movements. We chose these constraints of timing and environment to closely resemble a setting similar to a real-world business office environment. The system was live and operational in 8 locations for a total duration of 3-4 days (**287 hours**) of testing. During the testing, we registered 5 False Positives, all presented below:

False Positives registered during the long-term tests of the Weapon Detection Module



Taking into account the duration of tests and FPS, the calculated specificity of Scylla is estimated at **99.9999903%**.

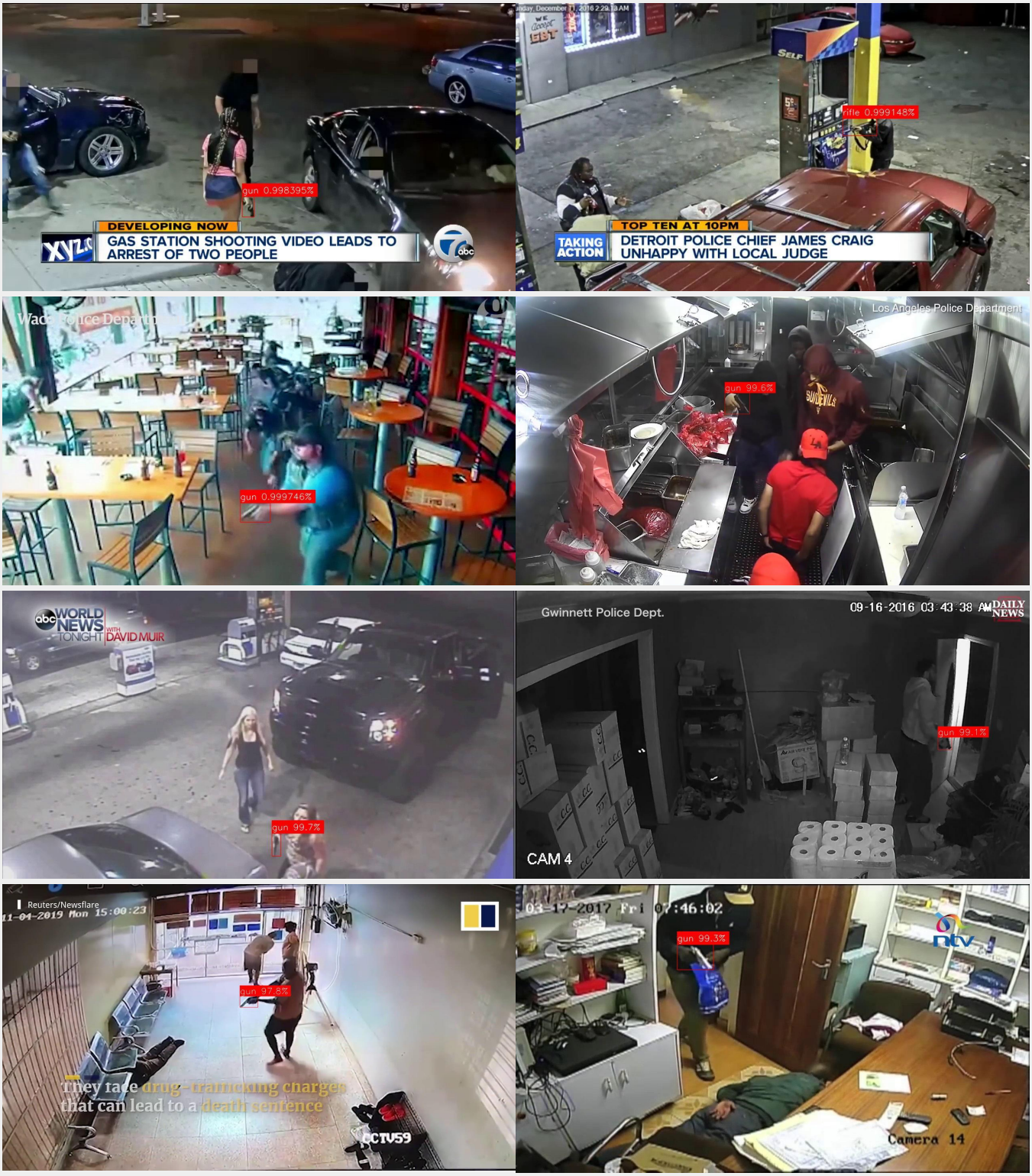
Tests from Youtube

Next we tested Scylla PTD on a number of publicly available real-life cases of assaults with weapons.

The table below summarizes the results of the testing:

Name	N
WATCH: Woman fires shots at home intruders https://youtu.be/thVhVjn59mg	57
Project Greenlight helps catch woman who shot at man at Detroit gas station www.youtube.com/watch?v=j8JtyM28WMc	29
Chief angry over robbery suspect release www.youtube.com/watch?v=cD1Dq7iU8gY	14
Waco biker gang shootout captured on restaurant's CCTV www.youtube.com/watch?v=uEhtPcMksK8	1
CCTV captures suspects robbing LA taco truck at gun point www.youtube.com/watch?v=xAzDrs5rKU8	5
Woman Pulls Out Gun During Road Rage Incident www.youtube.com/watch?v=SrB1yp2mVkI	1
VIDEO: Georgia woman shoots robbers, kills one www.youtube.com/watch?v=A804A3WAbB0	7
Accused drug dealer facing death penalty escapes www.youtube.com/watch?v=CjpMCqymmJk	1
Manhunt launched for Uthiru robbers caught on CCTV www.youtube.com/watch?v=GnfKCoiY5kY	5
Lonehill Johannesburg | Self-Defence Shooting www.youtube.com/watch?v=EfsZBaTvgzI	1*
CCTV of Moscow shooting www.youtube.com/watch?v=LcvTj2WqGdA	1
Caught on Camera 40-year-old Delhi man www.youtube.com/watch?v=B4O_-jj-Wqo	5

Examples of detections are presented below.



All detections can be accessed here.

Intrusion Detection & Perimeter Protection

Scylla IDS - is AI-based Intruder Detection and Perimeter Protection System that allows filtering out up to 99% of false alarms. It comes not only with conventional functionality (defining the zone of intrusion and the schedule of alerts), but also a possibility to specify the object of interest or run the suspected intruder through Scylla proprietary facial recognition system.

Object-detection based approach of Scylla IDS can work on cameras with moving backgrounds such as PTZ cameras (cameras that turn around to maximise their view coverage), drone- or body- mounted cameras. Meanwhile, Scylla IDS is invariant towards moving backgrounds and challenging conditions such as fisheye lens distortions, impaired illumination or noisy pictures. It can be easily integrated with the camera network that any retail area is equipped with.

There are two architectures for IDS solution: a) 24/7 real-time stream analysis and b) motion-detection based frame filtering.

a) As the description states - the first one analyses the video streams coming directly from the camera. It searches for acts of intrusion in every frame (can be reduced to 5 FPS) and reports directly to the Scylla Dashboard. In this scenario the server with Scylla is located either locally or on cloud.

b) second scenario is called “False Alarm Filtering” and is designed to filter false alerts coming from Motion detection cameras which usually caused by irrelevant events such as insects and dust particles flying in the view of the camera, environment-caused changes, luminescence light flickering, draft moved curtains and so on. To setup False Alarm Filtering usually “FTP upload of alarm triggered frames” is used. Usually the suspected frames are uploaded to cloud-hosted Scylla IDS instance, which analyses each frame, checks if there is a person or vehicle in the active area and forwards the alert to the end user in the positive case.

Occupancy Counting System (OCS)

Scylla OCS is a derivative of Scylla IDS. It is based on the same AI models. The major difference however that comprises the backbone of exceptional accuracy is the proprietary person tracking algorithm.

The spread of COVID-19 pandemic forced governments to apply restrictions and rules to minimize the impact it caused. It is mandatory to regulate and limit the maximum number of people allowed in closed spaces in several countries. Currently, managers hire people that stand at entrances for a whole day, manually and carefully counting people as they enter and exit. The task gets harder when there are multiple entrances/exits - the information needs to be synchronized all the time, a mistake made by one affects all.

This manual counting is mundane and expensive: tasking a single staff member to watch the door at all stores for one hour a day would cost three million dollars annually (source). Having someone outside for ten hours a day would cost even more.

What Scylla OCS offers is **automated people counting** - a system that utilizes IDS person detection and tracking engine and counts entrances and exits of individuals from any number of cameras.

The state-of-the-art tracker we developed monitors each customer who enters the building and can understand if the same person gets shortly out of and in the camera's view. This approach enables Scylla to be top-notch accurate at counting. Scylla OCS distributes alerts when total count of people in the setup area gets close to the defined threshold. In addition to the current total number of occupants Scylla OCS also provides statistical analysis, daytime distribution of occupancy and visitors, and more. Scylla OCS can be adjusted to measure other metrics, such as the time people spend in front of the cashier to assess customer service, how long employees spend in specific areas doing their job, etc. For instance, it can also be used to detect cases of loitering. One has to define the area, the maximum reasonable time of lingering in that area, and as soon as someone is detected to spend more time than the threshold, an alert will be created and sent to dashboards and configured endpoints. Similarly, a minimum "time of passage" can be defined to detect cases of "running through" with corresponding alarms. Lastly, similar to Scylla IDS here too the system outputs crowd heatmaps, occupancy per zone, "center of mass" maps and statistics on crowd flow pathways which can be provided as hourly, daily, weekly reports.

Person Search System (PSS)

Another derivative of IDS is the Person Search System. This solution provides an automatic search of a particular person(s) (Person Of Interest, POI) based on his/her appearance - where clothing plays the most important role. Similar to OCS it utilizes several modules and incorporates similar algorithms to improve accuracy.

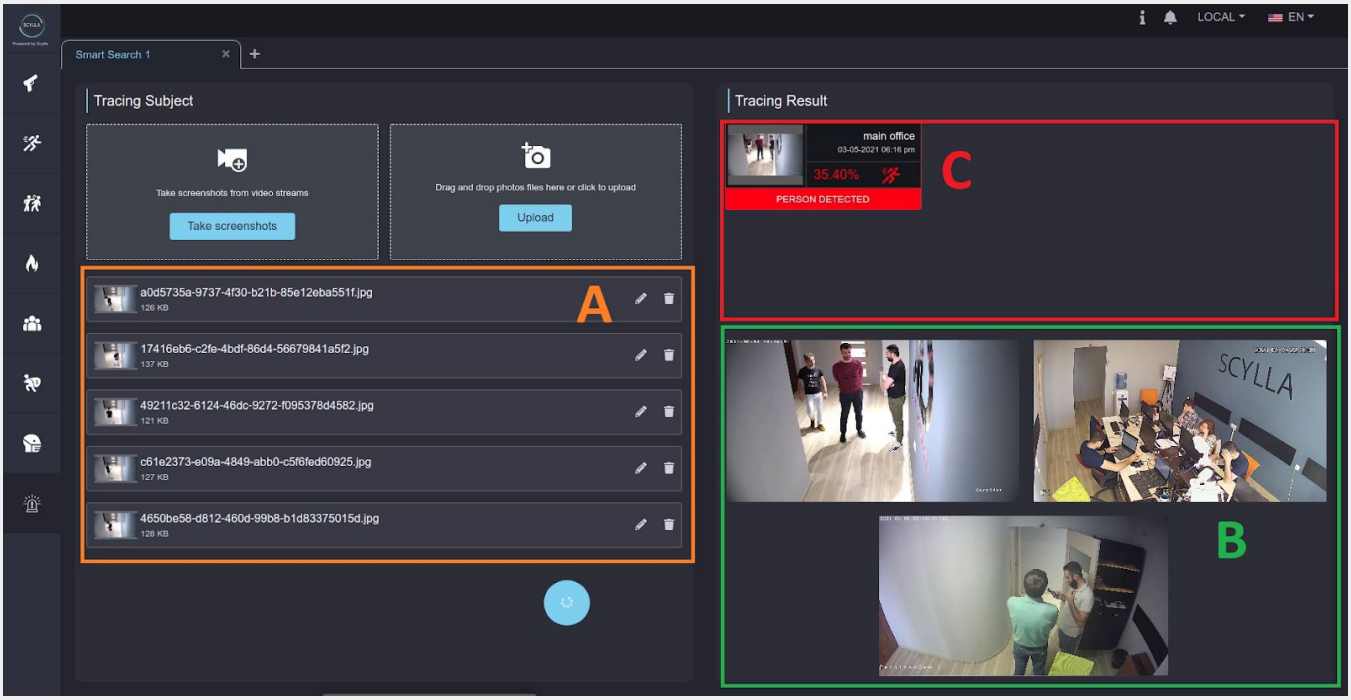
To initiate the search, photo(s) containing POI are uploaded to the system. Alternatively these frames can be collected from any of the cameras the system is connected to in real life. Once the frames are in the system POI needs to be specified. In case there are several people on the frame, the user is choosing which one is POI by clicking on the corresponding box. If the person is not outlined, the user needs to draw the contours of POI themselves.

The system can operate with only one image provided. However, the performance drastically improves if multiple images are provided, especially if the images are from different angles and perspectives. This requirement is more understandable if one takes into account the fact that clothing from the front and back might look different. The optimal number of photos of POI is 4 (this is the default number of images that the system takes from the camera once the real-time capturing is initiated). These appearances are transformed into feature vectors and stored to be used in search.

Once the frames are uploaded and POI is specified in each frame the search can be initiated. The search is running on all cameras simultaneously, assessing all people that are in the view of cameras. The PSS engine is engaging a number of Scylla modules. First it engages Person Detection from IDS, then each person is tracked using Scylla Object Tracking algorithm. Feature vectors from each occasion of each person are collected and compared with the ones from POI. The final conclusion is made based on statistical analysis from multiple assessments per each person in the view (note the tracking mentioned above). This approach eliminates the erroneous conclusions that otherwise would have been initiated based on resemblances from single sightings.

Once the suspicion of re-identification of POI is confirmed, the corresponding alert is shown on the dashboard. The user can either confirm or reject the suspicion. If the user is rejecting the proposed suspicion, the feedback is used by the engine as a “negative” feature vector set to reject proceeding errors, thus further narrowing the search and minimizing false positives. Similarly if the suspicion is confirmed, corresponding feature vectors are added to the search set and further aids the search accuracy.

Note that to experience best performance of PSS the camera placement and specifications should comply with the following requirements: the people on the screen should have 80 pixel height; the illumination should be enough for a human eye to confirm the similarity of the person of interest and the individuals in the view of the camera.



The dashboard of PSS. A) section with images containing the person of interest, B) the camera live feeds, C) the suspected sightings of the person of interest.

Simultaneous searching of multiple POI can be engaged on a single system.

The algorithm of re-identification can be used to detect re-appearance of individuals in use cases such as customer tracking from camera to camera. In this case “unrecognised” individuals are assigned a unique ID and added to a dynamic database of appearances. Once re-sighting of any individual is detected, a corresponding record is made in the tracking map.

The base model of re-identification that is used in PSS is different from the one that is used for tracking in OCS. It was trained on a different dataset where the accent is put on impressions of a single individual from different cameras and different view angles. Such an approach has drastically improved the accuracy of PSS which faces harder challenges when probing the similarity of individuals from cameras that have different qualities and positions, have different viewing angles, illumination, proximities etc.

The PSS re-identification model was tested using the test set of the most known re-identification dataset Market-1501.

The results are summarized in the table below:

Title	Title
Rank-1	83.2%
Rank-5	92.7%
Rank-10	95.6%
Rank-20	97.2%
mAP	65.6%

The numbers in the table are from cross-domain testing. Same domain testing mAP is 96.5% which is the second in the ranking.

Note that this model works alongside an ensemble of other models responsible for other steps of the solution (namely the detection, tracking and decision making), thus the final accuracy metrics depend on superposition of a set of conditions and parameters. And as usual the trade-off between sensitivity and specificity of the solution depends on the preset threshold and can be adjusted according to the demands of the use case. For instance, in case of the “lost child search” the search should be performed with a threshold set in favor of sensitivity due to the nature of the use case. Moreover, as such an emergency case search is supposedly performed under supervision of an overviewing user, the latter can overview the suggestions, guide and improve the search engine in real-time (see above) much similar to the so-called “reinforced machine learning” approach. In contrast, in cases where appearance-based re-identification is used in the unattended manner (i.e. customer tracking) the threshold is set in the favor of selectivity to minimize false guesses.

Expected results

IDS

The base model mAP (Mean Average-Precision) on Coco dataset is 0.93. Implemented algorithms result in filtering of up to 99.5% of false positives - frames that motion detection cameras attribute to suspicious acts of intrusion and that do not contain any humans. Running analytics on adjacent frames from the video sequence, implementing frame cross-checking and cascaded classifier, Scylla IDS reaches a rate of no more than 3 FPS per camera per day in real-world outdoor active installations and no more than 0.5 FPS a day for indoor installations.

OCS

The following estimations were derived based on the analysis of real-world installations and experimental setups.
In case of 100 people / hour traffic:
Error rate is ~ 1 person per hour.

Note: in many cases errors are related to camera positioning and can be improved once it is optimized. Observe the requirements and guidelines below to refine the view and improve the performance of Scylla OCS.

PSS

In case of 140 cameras:
False positive rate is ~ 0.3 cases per minute.

Note: False positive rate depends on the number and quality of the pictures of the POI provided (the number above is for 4 pictures, ~ 100 pixel height). It also depends on the average number of all people in all cameras that are connected to the system (~40 people in the case above). Lastly and naturally it also depends on the "uniqueness" of the appearance of POI.

Requirements

For IDS/OCS to operate in the most optimal conditions the following guidelines are to be observed.

Camera

The requirements on camera specifications are the following (the minimum is mentioned).

IDS

Resolution: minimum HD 1280x720 (optimal resolution Full HD 1920x1080)
FPS: 5
Positioning: camera can be positioned almost anywhere as the model is trained to detect and distinguish not only the “whole” human/vehicle in various positions, environments and illumination conditions but also parts of those. If only a hand of a human or a windshield of a car is within the camera view, Scylla IDS will still be capable of detecting the object.
Person/car size: 5% height of the vertical frame size.

OCS

Resolution: minimum HD 1280x720 (optimal resolution Full HD 1920x1080)
FPS: 15
Person size: 15% height of the vertical frame size.
Positioning: OCS is very picky towards the optimal positioning criteria. Partially the reason for that is the demanding nature of the tracking algorithm and it’s partially due to peculiarities of the task itself.

A few key points to take into account when setting up or adjusting existing cameras are the following:

1. Best case scenario is the “above head” placement of cameras where the camera is fixed above the entrance/exit tilted slightly downwards and looking towards a corridor of some sort. This type of positioning helps avoid cases when people obscure each other.
2. There should be some path that people pass before and after the crossline. Side exit/entries should be avoided (when a person can enter and exit avoiding being in both A and B regions).
3. The distance from people passing under the camera should be such that the height of people takes up no less than 1/6 of the frame vertical size.

An example of optimal placement is presented on Fig 1. Please take note of the area before the red line (4-5-6-3) and after (1-2-6-5) the red line. There should be substantial area on both sides of the line for the system to be able to detect the person and to understand the direction of the movement. Usually 1-2 steps should be enough.

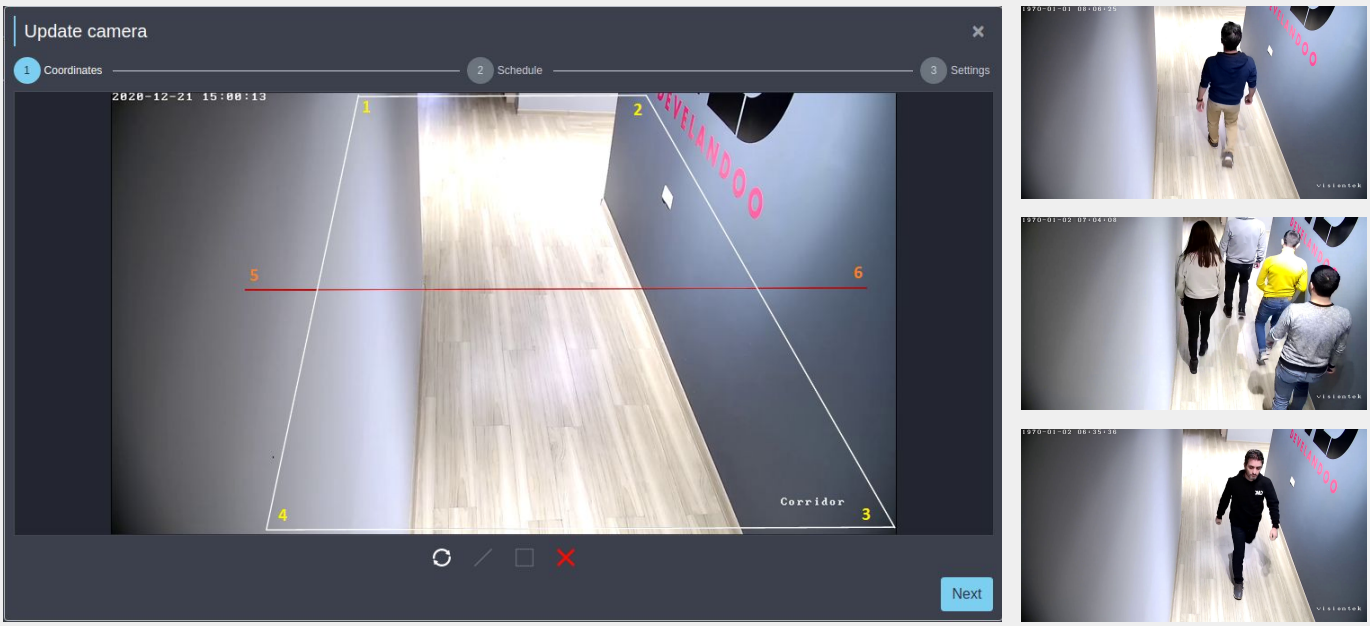


Fig. 1 An example of optimal placement of the camera for OCS.

PSS

Resolution: minimum HD 1280x720 (optimal resolution Full HD 1920x1080)
FPS: 15
Person size: 50 pixels in height.
Positioning: as PSS is using the overall appearance to search for the person, ideally it works best where the whole body is visible - both the top and bottom. An example of a non-optimal view is when the camera is facing an area where partial obscuring of all individuals takes place (i.e. at counters or straight downward facing turret cameras).

Car Detection & Parking Area Monitoring

In Scylla architecture, the Vehicle Identification and Tracking (VIT) module branched out of IDS solution - it uses the same engine for detection and tracking but the objects of interest here are vehicles. In many cases users can choose from their settings, if the detection/counting should include humans and/or vehicles.

VIT solutions provided by Scylla are used to solve smart traffic relevant problems, such as smart parking management, traffic counting, traffic rule violation etc.

In parking solutions the predominant algorithm is detection, however, the lite version of tracking is also implemented to verify parking bay entry/exit events and eliminate error cases related to temporary obscuration. For instance, for a parking solution to operate correctly FPS as low as 0.2 is enough.

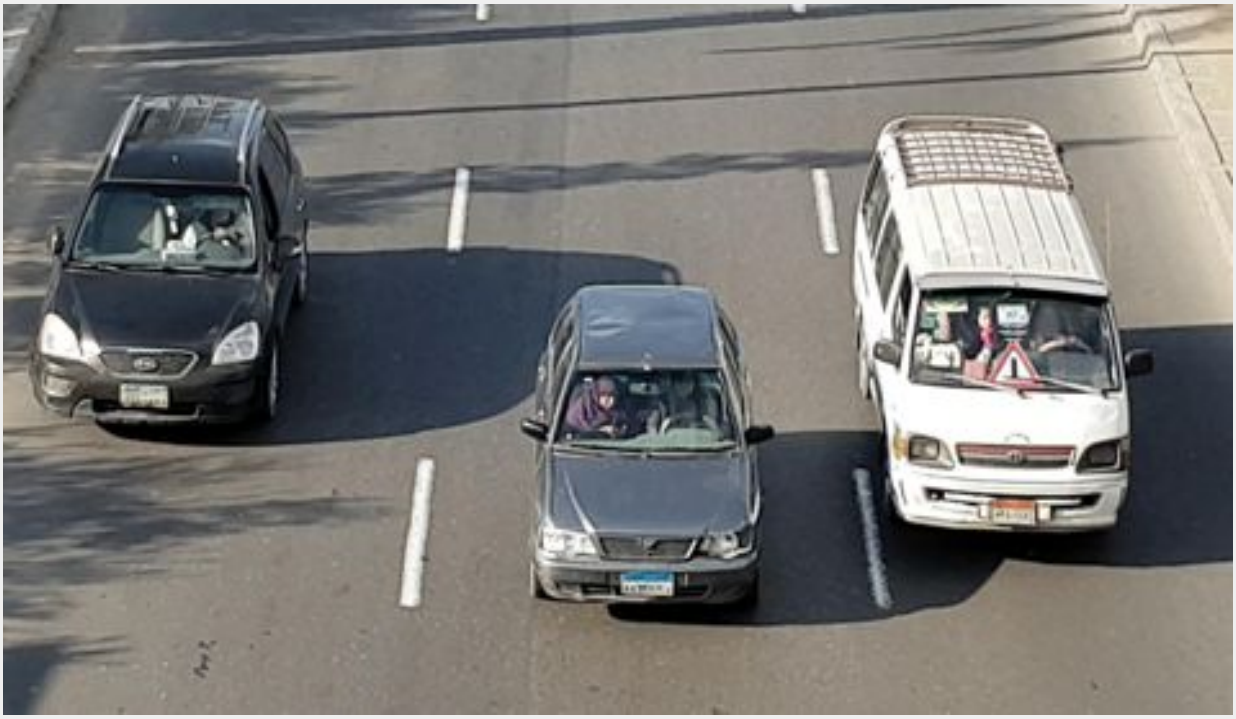
In contrast, car counting uses proprietary state-of-the-art tracking algorithms which are tailored for each object group (human vs vehicle vs face, etc). To count cars (as well as people) passing a certain pathway, each vehicle is tracked and the pathways are analysed. It allows Scylla to analyse and provide reports on:

- Car counting (per lane, per vehicle type)
- Flow density information
- Traffic rule violation
- Traffic jam detection
- Car flow heatmaps, etc.

Below there are examples of such operating installations:

Arlovskaya st., Minsk, Belarus (Car Counting VIT)

Khanjyan st., Yerevan, Armenia (Car Counting VIT)



Cairo Airport Bridge, Cairo, Egypt (Car Counting VIT)

Bangalore, India (Parking Solution VIT)

The overall accuracy of car detection/counting depends on the camera installation location, lighting conditions, distance from cars, camera quality as well as data transfer speed (video stream consistency). The documented individual results for several locations on car counting are listed in the table below:

Location	Precision	Recall	F1 score
Minsk, Belarus	99%	98%	98,5%
Yerevan, Armenia	100%	99%	99,5%
Cairo, Egypt	99%	98%	98,5%

The measured accuracy of the parking solution for four- and two-wheelers is more than 99,98% across 150 cameras (2267 parking bays). The scarce errors here predominantly resulted from poor lighting conditions, weather conditions induced obscurations, camera hardware-related frame distortions or similar hindrances.

Conclusion

All Scylla modules demonstrate state-of-the-art performance and accuracy hitting all the benchmark metrics to make it a commercially viable AI-powered solution. Additionally, extensive efforts were made to reduce hardware requirements of the system to support the scalability and cost-effectiveness of the technology.

All Scylla modules are capable of constant self-learning, which in turn will lead to further improvement of the reported results.

Anomaly Detection & Behavior Recognition

Abstract

In recent years, surveillance cameras have been widely used in public places. They can capture a wide variety of realistic anomalies. Unfortunately, these cameras provide evidence after the event has taken place and they are rarely used to prevent or stop criminal/abnormal activities in time. It is both time and labor-consuming to manually monitor a large amount of video data from surveillance video streams. Only 15% of an operator's time involved actively monitoring video feeds and searching for incidents (Wells, Allard, and Wilson, 2006). When a video surveillance operator has hundreds of individual cameras to monitor, it is unlikely that a potentially damaging event will be noticed in real-time (Sulman, Sanocki, and Goldgof, 2008). Scylla Anomaly Detection module has been developed specifically to close this gap and use video streaming information to accurately detect anomalies in real-time. Currently, the module has achieved state-of-the-art results on anomaly detection benchmark datasets.

Introduction

Vision-based action recognition is one of the most challenging topics in computer vision. The concept of video-based anomaly detection is defined as detecting anomalous behaviors in video data. It can be viewed as a subset of human action recognition which aims at recognizing general human actions (Cheng, Cai, and Li, 2019). Most deep learning based solutions for video recognition treat space and time symmetrically which might not lead to optimal results, since spatio temporal orientations are not equally likely.

Scylla Anomaly Detection module is based on deep learning architecture which takes into account spatiotemporal asymmetries and tries to treat them separately. Our method does not compute optical flow, hence the models are trained based on raw video data. This approach is demonstrated to be empirically more effective for more general action recognition tasks. There is no need for stream based retraining as is the case for the most anomaly detection systems that define violent or anomaly events as deviations from normal patterns. The solution built by Scylla team is outperforming the above-mentioned approach, since it is very difficult or even impossible to define a normal event which takes all possible normal patterns/behaviors into account (Sultani, Chen, and Shah, 2018). The definition of anomaly as a deviation from normal events also results in a large number of false positives, which is especially problematic for real-time crime detection systems.

How does the System Works?

The Anomaly Detection module is optimized to work on multiple video streams using a single GPU and providing real-time event tracking. The detection module was tested on various environments (universities, casinos, retail stores, metro stations, etc.) and has achieved superior test accuracy compared to the corresponding metric of other models with published results (for surveillance videos). The anomaly detector works offline as well, being able to analyze 24-hour videos in several minutes. Once an anomaly event has been recognized based on the chunk of frames given to the model, Scylla sends alerts to all assigned endpoints. The following examples demonstrate typical dashboard alerts for each of the submodules (figures 1, 2 and 3).

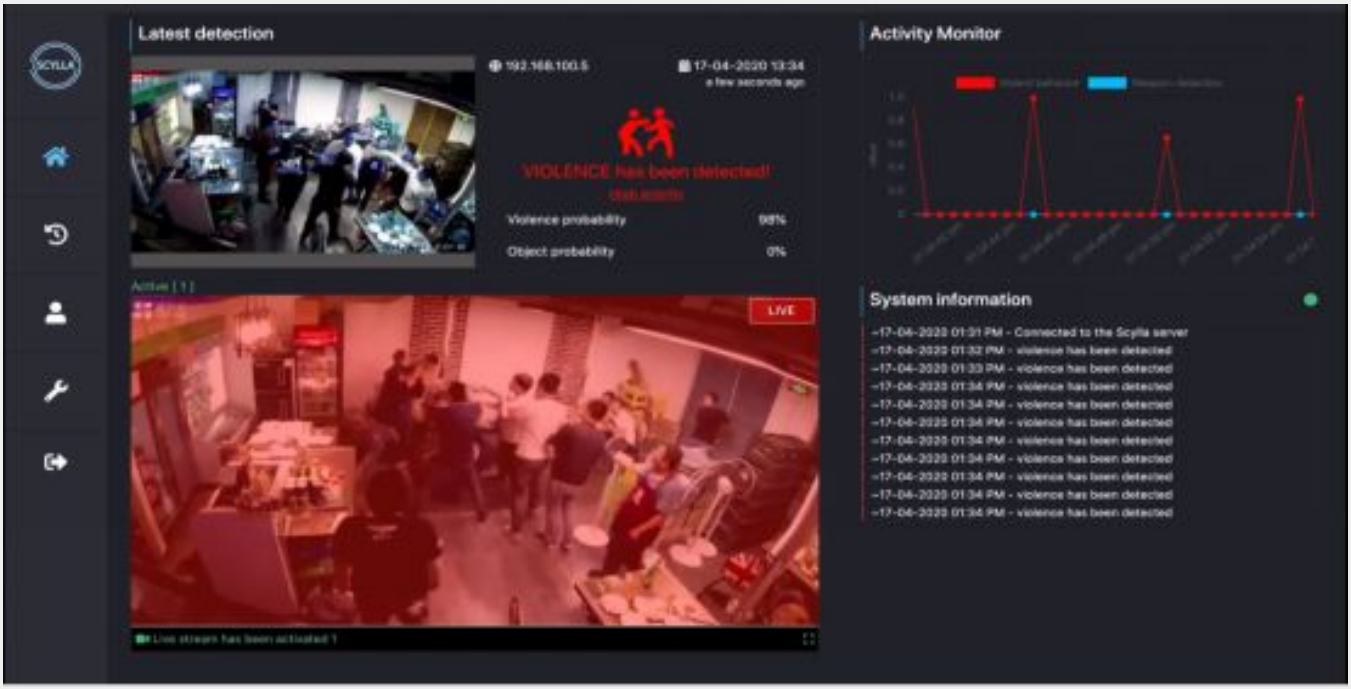


Figure 1. Security alert for the fight detection module

Figure 2. Security alert for the shoplifting detection module

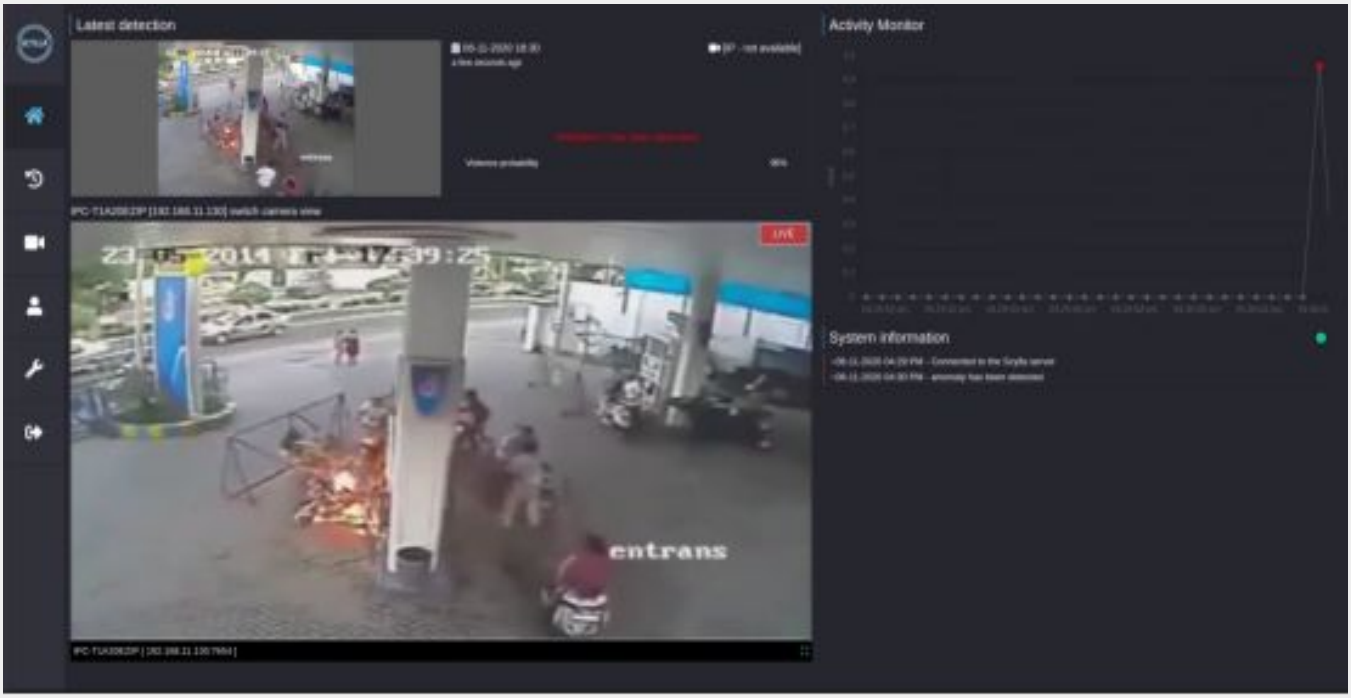


Figure 3. Security alert for the smoke/fire detection module

For offline video analytics, the video file could be uploaded for the model to run offline analysis in a couple of minutes (depending on the video duration and resolution). The Anomaly Detection submodules send back the starting frames of the abnormal events as well as information on the exact seconds where the system started to output the signals.

The system is self-learning, hence it can be adjusted for specific environments, in case the broad dataset of anomalous and normal events does not fully capture specific environmental details of the current installment.

The system outputs two types of signals. The red one is triggered immediately when the anomalous event is detected with a high probability specified by the threshold. The yellow signal is triggered when a signal is not strong enough to surpass the red threshold but is consistent enough to be worthy of attention.

One important limitation of the system that is worth mentioning is the fact that ADS is analyzing the big picture of what is happening in the stream. Hence if the event is taking place far enough (less than 10% of the height of the stream is covered) then the system might miss the event. In order to minimize these limitations it is recommended to mark an active area in the stream where the event is expected to occur (as an example the ceiling of the store is the place which could be skipped when analyzing the shoplifting case).

The video processing speed from one stream is 640 FPS on NVIDIA 1080 Ti graphic card and drops linearly with additional streams connected. The input resizing for the model is done in 2 ways: one simple square resizing (256 x 256), and the other one preserving the aspect ratio and scaling the shorter side to 256. Two methods work well, although the second option distorts the input image less. There are several versions of the model with modified architecture to offer different stream support/accuracy ratios. The biggest and most accurate one supports 14 streams on the NVIDIA 1060 GPU with 6GB video RAM (FP16 60 GFLOPs).

The databases for anomalous events are constantly being updated, hence the performance of the model is increasing with every new version released. The following sections cover all types of anomalies supported by the Anomaly Detection system as of the latest release.

Fight Detection Submodule

The fight detection model is trained on a large dataset of violent events recorded on surveillance cameras. Scylla AI-based system can detect a wide range of anomalous events, including:

- Fighting
- Assault
- Vandalism

Main Results

The current model has been tested on several benchmark datasets for video surveillance and outperformed existing models by a significant margin (10-25% more accurate on the test sets of the corresponding datasets).

The Fight detection module accuracy results are provided in the following table (Table 1):

Metrics	Accuracy	ROC	F1 score (binary)	Precision	Recall
Results	95.24	98.69	93.85	92.79	94.93

Table 1. Accuracy results of the Fight detection module

The test set contains a large number of surveillance videos with different backgrounds and day time information. It is designed to be representative of the true behavior of the model in real-life situations.

The corresponding ROC curve is presented in the figure below (Figure 4):

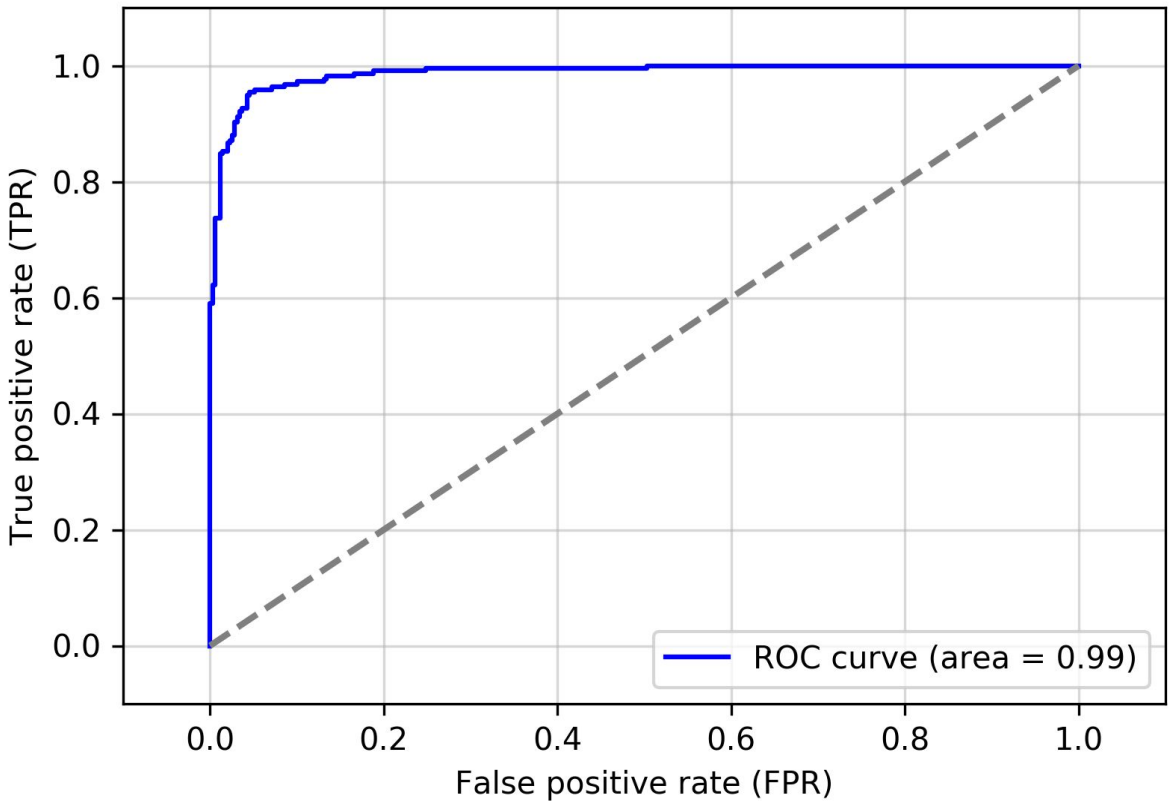


Figure 4. ROC curve of the predictive model (Fight)

One of the main concerns regarding the Anomaly Detection system is the false positive rate. As can be seen from the accuracy table, the precision of the model is 92.79%, which is significantly higher compared to other fight detection modules.

Below are presented several typical false positive examples from testing in real-life environments from different streams.

Figure 5. False positives of the fight detection system - Example 1

Figure 5. False positives of the fight detection system - Example 2

There are specific reasons for the model to signal a false positive alert in either of the cases. For example 1, the reason is the connection issues that have distorted the images, hence the system sees some kind of an anomaly in the chunk of frames presented as an input. The case also indicates the importance of a reliable network for the fight detection module to work efficiently without false alerts.

The second example is related to the fact that the camera is installed a little bit lower, hence some of the action happening in the frames presented is only partially visible. The latter is the indication of the system dependency on the correct positioning of cameras.

Shoplifting Detection Submodule

The shoplifting detection model is trained on a large dataset of videos from surveillance cameras, where shoplifting events occur. The submodule is designed to detect an action of taking an item and trying to conceal it. The system is supposed to analyze the streams coming from the retail environment excluding the cashier area, since after paying for an item a customer could take actions that are identical to concealing goods.

Main Results

The Shoplifting detection module accuracy results are provided in the following table (Table 2):

Metrics	Accuracy	ROC	F1 score (binary)	Precision	Recall
Results	84.94	86.58	68.94	60.00	81.00

Table 2. Accuracy results of the Shoplifting detection module

The test set is designed to contain challenging examples to reproduce an expected behavior in real-life environments, hence it contains normal videos which contain activities which might look similar to events of shoplifting. As it can be seen from the precision figure, the false positive rate of the submodule is much higher than in the case of the fight or smoke/fire detection, which is due to the fact that shoplifting behavior is far more complex in nature and might resemble normal behavior and vice versa quite often. The expected hourly rate for alerts is 4-5 per video stream, hence the submodule is advised to be treated as an early warning tool for a security officer to monitor.

The corresponding ROC curve is presented in the figure below (Figure 6):

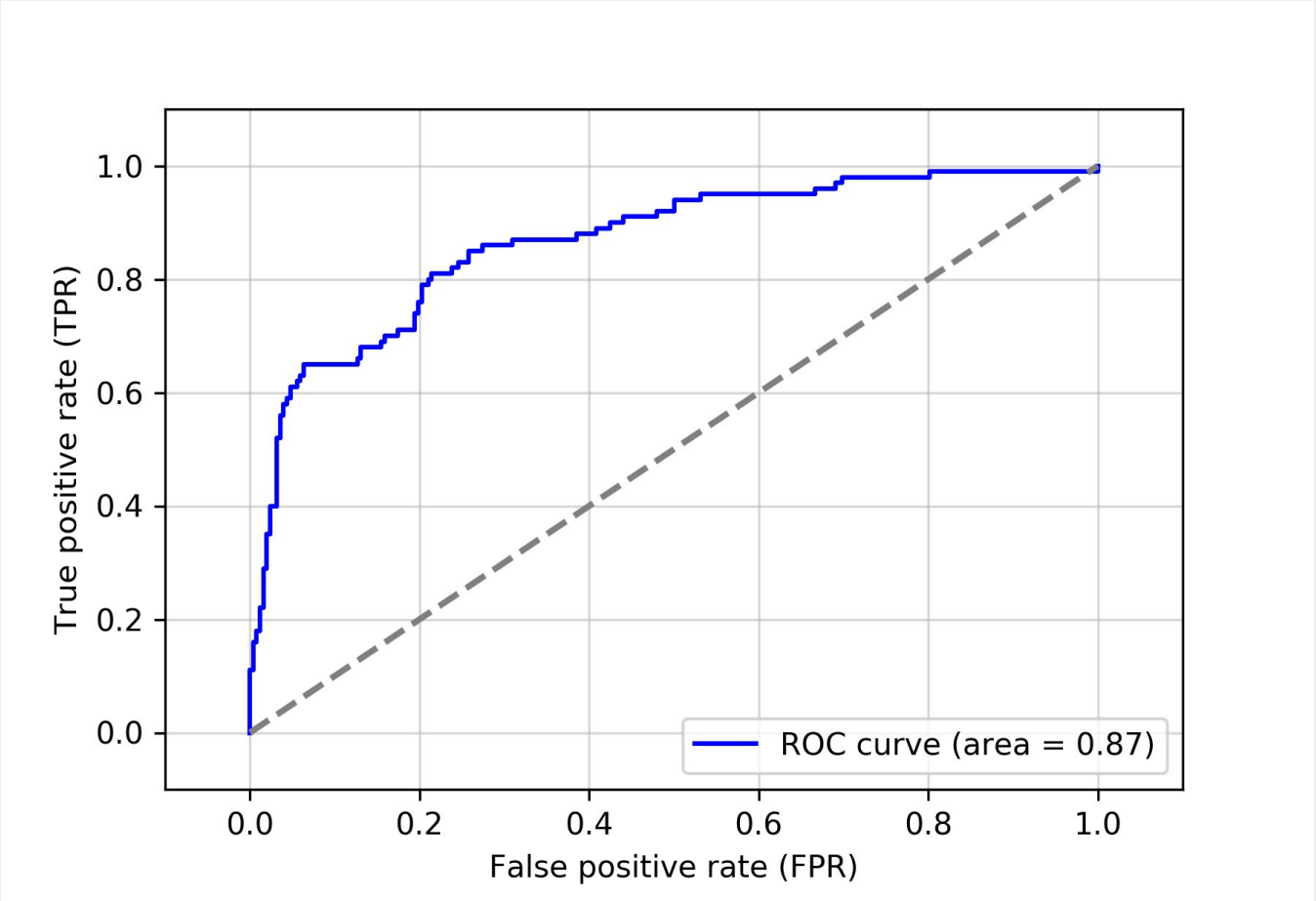


Figure 6. ROC curve of the predictive model (Shoplifting)

Below there are several typical false positive examples from testing in real-life environments from different streams:

Figure 7. False positives of the shoplifting detection system

As it can be seen from the cases above, the model could be considered as an early warning system, since it signals about the potential theft when the behavior somewhat similar to shoplifting occurs. Because of the complexity of the pattern of shoplifting events, there are several scenarios where visually the pattern could be indistinguishable from theft from the AI perspective and even for a human eye as well.

Smoke/Fire Detection Submodule

Smoke/Fire detection submodule is built on the largest dataset for the ADS models so far. As the title suggests, the system alerts in the event of either smoke or fire. The environment where the system could be installed is unconstrained and can be both indoor and outdoor.

Main Results

The Smoke/Fire detection module accuracy results are provided in the following table (Table 3):

Metrics	Accuracy	ROC	F1 score (binary)	Precision	Recall
Results	96.04	99.29	94.35	94.35	94.39

Table 3. Accuracy results of the Smoke/Fire detection module

As it can be seen from the table above, the false positive rate is the lowest among all three submodules.

The corresponding ROC curve is presented in the figure below (Figure 8):

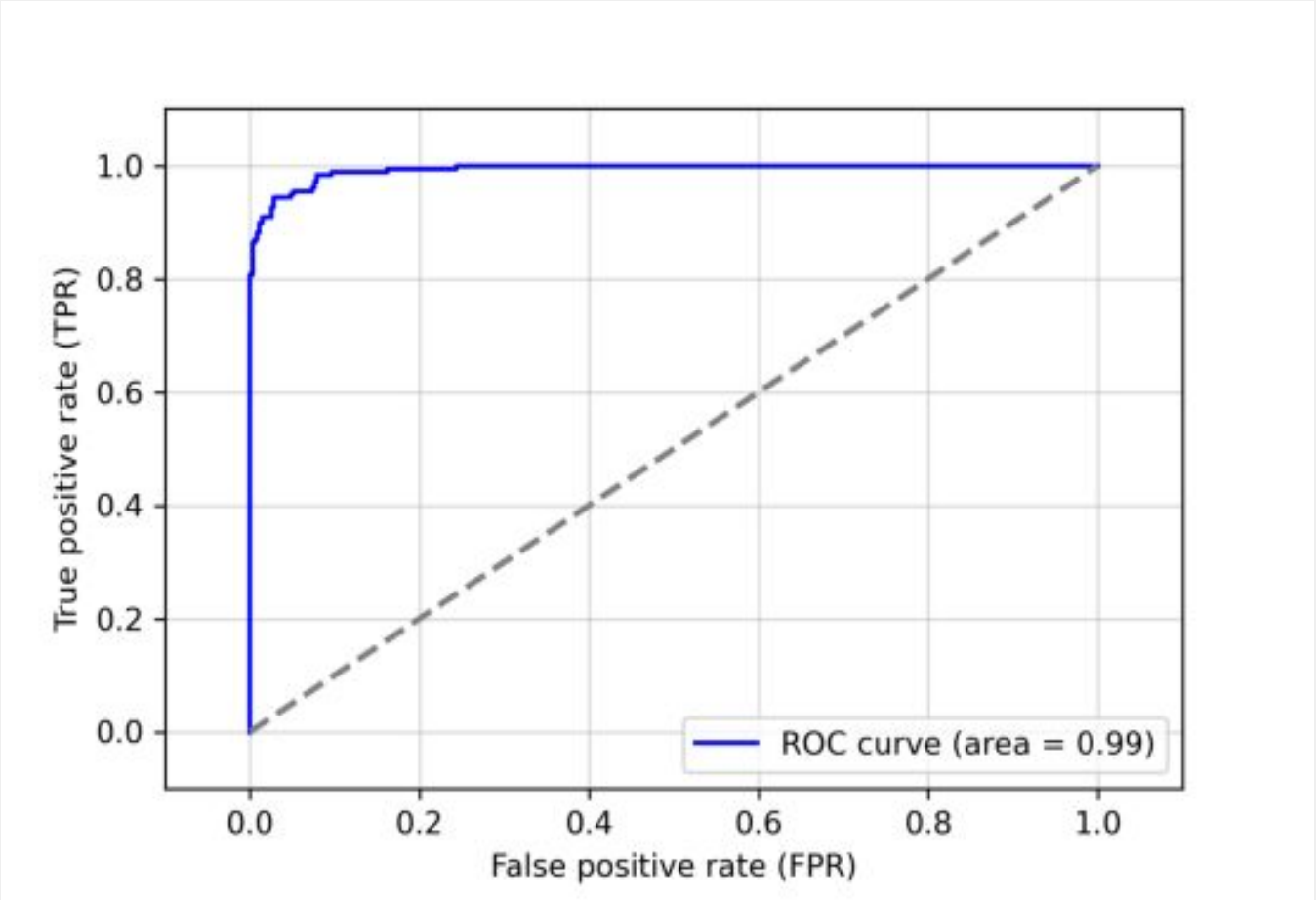


Figure 8. ROC curve of the predictive model (Smoke/Fire)

The system has been heavily adjusted not to be sensitive to several sources of false positives like strong lights at night, the sun, clouds, etc. However, it is still recommended to mark the active area in case of smoke/fire as well, since it would reduce unnecessary noise coming from the stream.

Conclusion

Scylla Anomaly Detection System is designed to solve one of the hardest problems in computer vision and pattern recognition. Most of the approaches used in operations nowadays rely on the system to be able to detect deviations from normal patterns. This technique is known to have low accuracy and a large number of false positives, besides it sometimes requires additional on-site training. The module designed by Scylla team beats all the currently available systems by a large margin in terms of accuracy. The module also supports real-time multiple stream processing as well as offline analysis of video recordings.

Face Recognition

Scylla Face Recognition module has a cascaded architecture, the overall performance of which depends on the performance of each unit. The process starts with the face detection algorithm. Then we employ our state-of-the-art tracking to capture and attribute a sequence of detected faces to a particular person. Next, a subset of representative faces is filtered by position and size and forwarded to the identification module. Finally, the system compares this subset with the dataset in the connected database and identifies the person.

Naturally, the performance of detection and identification steps depends on the quality of the video, face illumination, its size, and position to the camera. Illumination should be within the nominal range - illumination that is too dark or bright impairs identification accuracy.

Face Recognition accuracy depends on a number of factors three of which stand out: face **illumination**, the **angle** of the face towards the camera, and the face **size**.

Illumination is an external condition that should be considered when mounting the camera. For outdoor installations illumination change throughout the day may cause disruption of Face Recognition operation, if not regulated properly.

The **angle** of the face is amongst the most important limitations that can drastically affect the accuracy - both the face horizontal angle (between the facing direction and the camera) and face vertical tilt have to be within certain limits (usually less than ±30°) for face recognition to work properly. That's why tracking the person and regularly probing for a "properly positioned face" is the only reliable algorithm in cases where the identification process is not controlled (see below the explanation of whitelist/watchlist scenarios).

The **face** size expressed in pixels is yet another essential metric. It may depend on the resolution of the video frame, the camera view angle, and the distance between the person and the camera. Similar to the face angle, here too the best approach is to track each person in the view of the camera and select the cases when the individual approaches the camera and their face size exceeds the minimum threshold.

To benchmark and compare the accuracy of our algorithms we performed a few most common benchmarking tests. The results are presented in the table below.

Benchmark/dataset	Accuracy
Labeled faces in the Wild (1:1)	99.85%
IJBC@e4	97.5%
Megaface_Ver (1:1)	99.1%
Megaface_ID (1:N)	99.1%

Note the outstanding results on Megaface metrics which is considered the most challenging among the benchmarking tests.

The table below lists the dependencies of accuracy metrics on the listed factors. In Scylla SSIS algorithms we tend to set the lowest face size limit that is sent for identification at 40 pixels in height.

Conditions	Detection Accuracy, %		Identification Accuracy, %
	Precision	Recall	
1. Face horizontal angle < 42° 2. Face tilt < 22° 3. Face height >112px 4. Min interpupillary distance > 40px 5. Max face occlusion < 1/8 6. Min illumination > 80 CRI, > 1000 lx	97	90	96

Conditions	Detection Accuracy, %		Identification Accuracy, %
	Precision	Recall	
1. Face horizontal angle < 48° 2. Face tilt < 30° 3. Face height > 84px 4. Min interpupillary distance > 30px 5. Max face occlusion < 1/4 6. Min illumination > 50 CRI, > 200 lx	~80	~60	~70

The model used for identification has the following performance characteristics:

- Precision - 93.41
- Recall - 95.13
- F1 score - 94.26
- Accuracy - 94.53

In addition to the discussed parameters, factors such as obstruction of the facial area by other blocking objects and general camera quality can also affect Face Recognition accuracy. Scylla SSIS addresses many of these limitations by smart algorithms it incorporates. Depending on the use case the biometric identification is implemented either by **Whitelist** or **Watchlist** scenarios.

The **Whitelist** mode is used to identify personnel members and enable verification of Access Control protocols, biometric time-clocking, etc. It is a more “controlled” case where the person knows how to go through the verification process (i.e. stand in a certain spot close to the camera, look at the camera, wait for verification, etc.). The system resources here are used at a minimum. Most often in the **whitelisting** mode, Scylla SSIS is triggered occasionally through the day. For instance, it can be initiated as soon as a person stands still in front of the camera at a certain spot and distance. Identification takes a second, the flow is “one person at a time”, and after successful recognition, SSIS is hibernated. In many cases such sporadic use-per-demand nature of Whitelisting allows it to be installed on the same servers where other Scylla solutions are deployed.

In contrast, the **Watchlist** mode is used to identify strangers in the crowd based on the predefined “suspect” list. Here in most cases, individuals are not aware of the processes behind SSIS, so they do not necessarily “comply” with Face Recognition requirements (i.e. they do not always look at the camera, come close to it, stand still, or wait for the processes to finish, etc.). Thus, Scylla SSIS engages more resources for successful identification. More specifically, in the **Watchlisting** scenario, each individual is tracked through the view of the camera, identification is done continuously, and the analysis of this ensemble gives the most accurate output. These processes are continuous, and they need much more resources. Moreover, they depend on the maximum number of people found simultaneously in the view of the cameras. Constant engagement of non-stop people detection, tracking and identification results in elevated hardware requirements.

TIP

Whenever possible, the following camera placement guidelines should be observed to increase the accuracy of the **Watchlisting** scenario and chances of spotting the person in the list:

- place the cameras at face height or slightly higher,
- place cameras in corridors and passages, at entrances and doorways where the crowd is bound to walk towards the camera,
- limit the possibility of the “background crowd”,
- use high-resolution cameras so that a person’s face has a minimum of 50 pixels in height at least in some areas of the passageway.

For more information, please contact

SRS GLOBAL SERVICES

SRS Global Services LLC.
1-844-244-4211
info@srsglobalservices.com
www.srsglobalservices.com

Locations:
2553 Kincaid Ave
Baton Rouge, LA 70805

342 22 Ave
Nashville TN 37203