

AI-driven IVA Pro with FW 8.47 (CPP13) & FW 8.70 (CPP14)

Whitepaper



Table of contents

1 Introduction	3
1.1 Limitations	3
1.2 Supported cameras	3
2 Differences between IVA Pro Buildings, IVA Pro Traffic and IVA Pro Intelligent Tracking	4
2.1 Object classes	4
2.2 Tracking modes	4
3 Differences to IVA, IVA Pro Perimeter and Camera Trainer	5
4 Technical Background	7
4.1 Machine learning: Finding threshold between target object and world	7
5 Performance of IVA Pro Traffic	8
5.1 Dataset	8
5.2 Lane counting accuracy: >95%	8
5.3 Stop bar presence detection: >99%	8

1 Introduction

Intelligent Video Analytics (IVA) Pro is an AI-driven expert suite of video analytics that delivers valuable insights for improved efficiency, simplicity, security, and safety. Based on deep learning, it provides application specific video analytics with IVA Pro Buildings, IVA Pro Traffic, and specific PTZ camera functionalities with IVA Pro Intelligent Tracking.

IVA Pro Buildings is ideal for intrusion detection and operational efficiency in and around buildings. Without the need for any calibration, it can reliably detect, count, and classify persons and vehicles in crowded scenes. It can alert operators immediately as queues begin to form.

IVA Pro Traffic is designed for ITS applications such as counting and classification, as well as Automatic Incident Detection. It supports strategies that enhance mobility, safety, and the efficient use of roadways and solutions for intersection monitoring. It achieves accuracy levels beyond 95% for real-time event detection and aggregation of comprehensive data necessary for highway and urban infrastructure planning.

IVA Pro Intelligent Tracking adds AI to the Intelligent Tracking. It also brings AI, with the ability to separate close objects and classify them as person and vehicle, to a globally available video analytics outside of prepositions, which can be used even when the PTZ camera is moving.

In this whitepaper, the technology behind IVA Pro Buildings, IVA Pro Traffic and IVA Pro Intelligent Tracking will be explained, as well as the differences between each other, and towards the more traditional IVA Pro Perimeter and classical IVA.

1.1 Limitations

- ▶ Setup is only possible via the Configuration Manager.
- ▶ Detection of persons, bicycles, motorbikes, cars, trucks, busses.
 - Persons, bicycles and motorbikes may be confused, especially when seen from the front.
 - Busses and trucks may be confused.
- ▶ Minimum object size: 256 square pixel, e.g. 16x16 pixel, in the corresponding video analytics resolution
- ▶ Minimum object visibility: 50%. Objects occluded more than 50% may not be detected.
- ▶ Speed, geolocation and color are only available in 3D traffic mode.
- ▶ Top-down views (birds eye views) are not supported.

1.2 Supported cameras

IVA Pro is available on all CPP13 and CPP14 cameras with the exception of FLEXIDOME panoramic and FLEXIDOME multi.

2 Differences between IVA Pro Buildings, IVA Pro Traffic and IVA Pro Intelligent Tracking

There are two main differences between IVA Pro Buildings and IVA Pro Traffic:

- Available object classes (see Section 2.1)
- Available tracking modes (see Section 2.2)

IVA Pro Intelligent Tracking, on the other hand, extends IVA Pro Buildings and IVA Pro Traffic to PTZ cameras specific functionality by providing video analytics when the PTZ camera is moving and AI-based Intelligent Tracking.

2.1 Object classes

The following object classes and object filter are available:

IVA Pro Buildings:

- Person
- Vehicle

IVA Pro Traffic:

- Person
- Vehicle
 - o Bike
 - Bicycle
 - Motorcycle
 - o Car
 - o Truck
 - Bus

Filter by Object Class ?

Select preferred object classes to limit alarms. Some classes are only available for calibrated cameras with active 3D tracking.

- Person
- Vehicle
- Bike
- Bicycle
- Motorcycle
- Car
- Truck
- Bus

IVA Pro Intelligent Tracking inherits the object classes from IVA Pro Buildings and IVA Pro Traffic. So if a camera has only IVA Pro Buildings on it, then IVA Pro Intelligent Tracking has the two object classes person and vehicle. If the camera also has IVA Pro Traffic, then IVA Pro Intelligent Tracking inherits all of the IVA Pro Traffic classes.

Object classes are hierarchical. That means e.g. a bicycle is also a bike is also a vehicle, and a bus is also a truck is also a vehicle. Object class filters fully support this hierarchy, while visual class labels will only show the deepest level of classification, that is they will show person, bicycle, motorbike, car, truck and bus labels.

2.2 Tracking modes

The following tracking modes are available:

IVA Pro Buildings:

- Base tracking (2D)

IVA Pro Traffic:

- Base tracking (2D) (inherited from IVA Pro Buildings but with all object classes)
- Traffic tracking (3D) (needed for 3D object sizes, speed & geolocation)

IVA Pro Intelligent Tracking is separated from this, as Intelligent Tracking steers the PTZ camera to follow a selected object, and the global video analytics outside of the prepositions on the PTZ camera does not allow any tracking mode selection either.

3 Differences to IVA, IVA Pro Perimeter and Camera Trainer

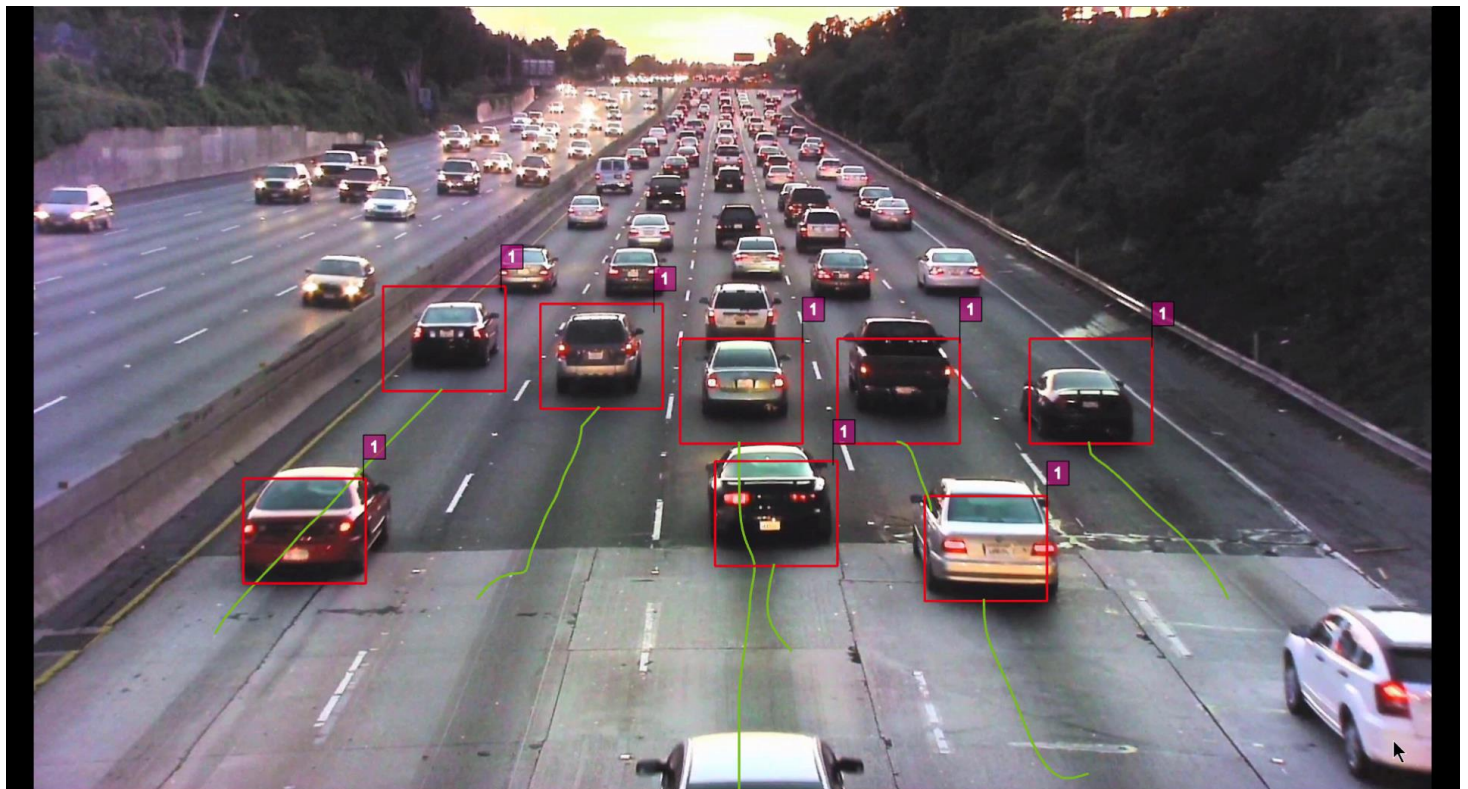
IVA and IVA Pro Perimeter were designed for intrusion detection and work well in scenes where objects are visually well separated. In scenes with lots of persons or vehicles, these algorithms soon reach their limits. In high-density traffic, where objects merge visually, IVA and IVA Pro Perimeter are not able to separate the objects anymore. Furthermore, they detect only moving objects.



IVA/ IVA Pro Perimeter works well in low / medium traffic

In high traffic, objects merge together

To separate vehicles in high traffic, or to detect parked vehicles, machine learning is needed. Camera Trainer was Bosch's initial solution for these applications. Camera Trainer's low processing power requirements made it ideal for use on Bosch IP cameras based on the CPP6, 7 and 7.3 common product platforms. However, Camera Trainer was limited to a short distance and needed to be trained by the user for every scene, resulting in high training effort. (The advantage of Camera Trainer is that any kind of rigid object can be trained. For more information please refer to the Camera Trainer TechNote).



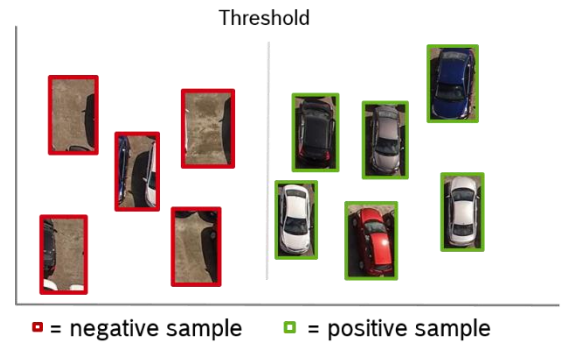
IVA Pro Buildings, Traffic and Intelligent Tracking provide a pre-trained vehicle and person detector which also supports greater detection distances than Camera Trainer, though less than IVA and IVA Pro Perimeter. IVA Pro Traffic separates persons, bikes, cars, trucks and busses even in dense congestion or traffic queues. Another benefit of IVA Pro Buildings, Traffic and Intelligent Tracking is that they are robust with regards to shadows or headlight beams.



4 Technical Background

4.1 Machine learning: Finding threshold between target object and world

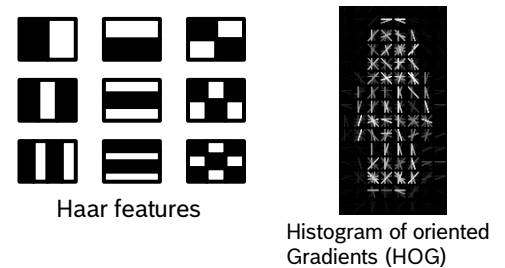
Machine learning for object detection is the process of a computer determining a good separation between positive (target) samples and negative (background) samples. In order to do that, a machine learning algorithm builds a model of the target object based on a variety of possible features, and of thresholds where these features do/don't describe the target object. This model building is also called the training phase or training process. Once the model is available, it's used to search for the target object in the images later on. This search in the image together with the model is called a detector.



Hand-crafted features typically describe edges and include:

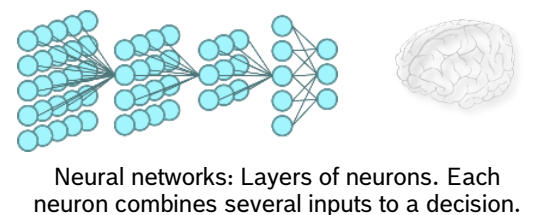
- Haar features: Binary edge descriptors via black/white tiles
- Histogram of oriented Gradients (HoG): Histogram of quantized edge direction and contrast

The resulting model will typically have around 2000 parameters. There are different methods of machine learning with hand-crafted features including support vector machines (SVNs), AdaBoost, and decision trees. Each of these methods has certain advantages, however, all of them result in similar performance levels. A detector based on these features can typically run in real-time in current network camera hardware.



Camera Trainer is based on an SVN using Histogramms of oriented Gradients. For more details on Camera Trainer, see its own Tech Note.

Neural networks are based on the visual cortex and are able to learn descriptive features on their own. They utilize a neural network structure of optimization parameters instead of using the handcrafted features. Typically, neural networks for image processing will also learn edge features and combine them first to parts of the object, and then the full target object itself. Deep neural networks for image processing use roughly ~20 million parameters, and can deliver a performance boost of up to 30%. On the other hand, deep neural networks are a brute force approach that requires hundreds of times more computational power.



Besides the model and training method, the samples of target objects and the background are very important. For a task like face or person detection, the positive samples need to show all possible variations including perspective and pose, while the background samples have to represent the full world. Therefore, machine learning needs tens of thousands of examples of the target object and billions of examples of what the rest of the world looks like. This is a huge effort both in collecting and preparing the sample images. For automated machine learning, either the target objects have to be marked in the image or the images restricted to show only the target object. Furthermore, modelling the complexity of the full world is one of the reasons machine learning is computationally expensive.

IVA Pro Buildings, Traffic and Intelligent Tracking are trained using a deep neural network.

5 Performance of IVA Pro Traffic

5.1 Dataset

The dataset consists of 253 video sequences from 19 different Bosch cameras deployed in urban environments and on highways. Each sequence is 30s long, the whole dataset consists of over 2 hours of video material. The video sequences were taken during day- and night-time, different weather conditions, and different traffic densities. Ground truth was generated by manual labelling.

5.2 Lane counting accuracy: >95%

Lane volume counting accuracy is defined as:

$$VA = 100 - \frac{|VT_{VD} - VT_{GT}|}{VT_{GT}} \cdot 100$$

with VT_{VD} = total volume of detected vehicles and VT_{GT} = total volume in the ground truth.

107 counting lines in total were defined and applied to the 253 sequences, where sequences from the same camera share the same counting lines.

	Amount of vehicles	Lane volume counting accuracy
Overall	3851	95,8%
Normal scenes	3215	97,0%
Challenging scenes	636	82,2%

Performance drops in challenging scenes due to reduced visibility due to weather, including rain drops on the camera lens, as well as headlight glares at night. Some extra counts can occur due to vehicle reflections, especially on wet surfaces.

5.3 Stop bar presence detection: >99%

Stop bar presence detection accuracy is defined as:

$$PA = \frac{|TT - CET|}{TT} \cdot 100$$

with TT = total time and CET = cumulative error time.

Error state is defined as any false positive or false negative detection in the detection zone.

15 presence detection areas were defined and applied to 98 sequences, where sequences from the same camera share the same presence detection areas. Stop bar presence detection accuracy stayed stable at 99% during day, night, normal and more challenging scenes.

Performance degrades if

- the detection zone is too close to the image border
- the lanes are hard to separate
- detection areas are placed in low-resolution image areas
- detection areas are often in deep shadows
- glare occurs often

Bosch Sicherheitssysteme GmbH

Robert-Bosch-Ring 5

85630 Grasbrunn

Germany

www.boschsecurity.com

© Bosch Sicherheitssysteme GmbH, 2023