Geo-spatial Traffic Behaviour Analysis and Anomaly Detection

TU/e. Surveillance and Safety lab

ePicture This, 2024

Dr. ir. Egor Bondarev, associate professor, SPS-VCA

VCA Video Coding & Architectures research group

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Intro: Traffic Anomalies

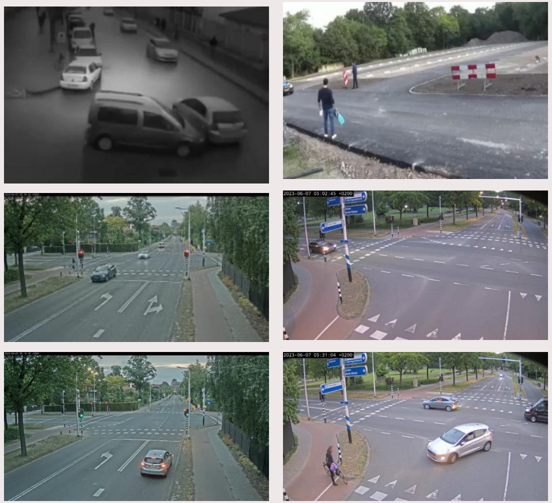
- Accidents
- Throwing and littering
- Illegal turning
- Opposite lane driving
- Zig zag driving
- Side lane parking
- Illegal crossing (biker, pedestrian)
- Violence
- Robbery
- Infrastructure collapse
- etc






Intro: Challenges in anomalous behaviour detection

1. Anomalies are rare events, training data cannot be collected in vast amounts
2. Anomaly types are limitless: you never know them in advance
3. Visual analysis of the actor's pixels is not enough: context is needed:
   - Robbery or helping an old lady to carry a bag?
   - Drug selling to a driver or helping the driver with directions?
   - Infrastructure attack or a repairman working?
   - Opposite lane driving – need to know the lane direction
   - Zig zag driving – need to know the lane shape
   - Side lane parking – need to know the traffic signs around
   - Illegal crossing (biker, pedestrian) – need to know road markings



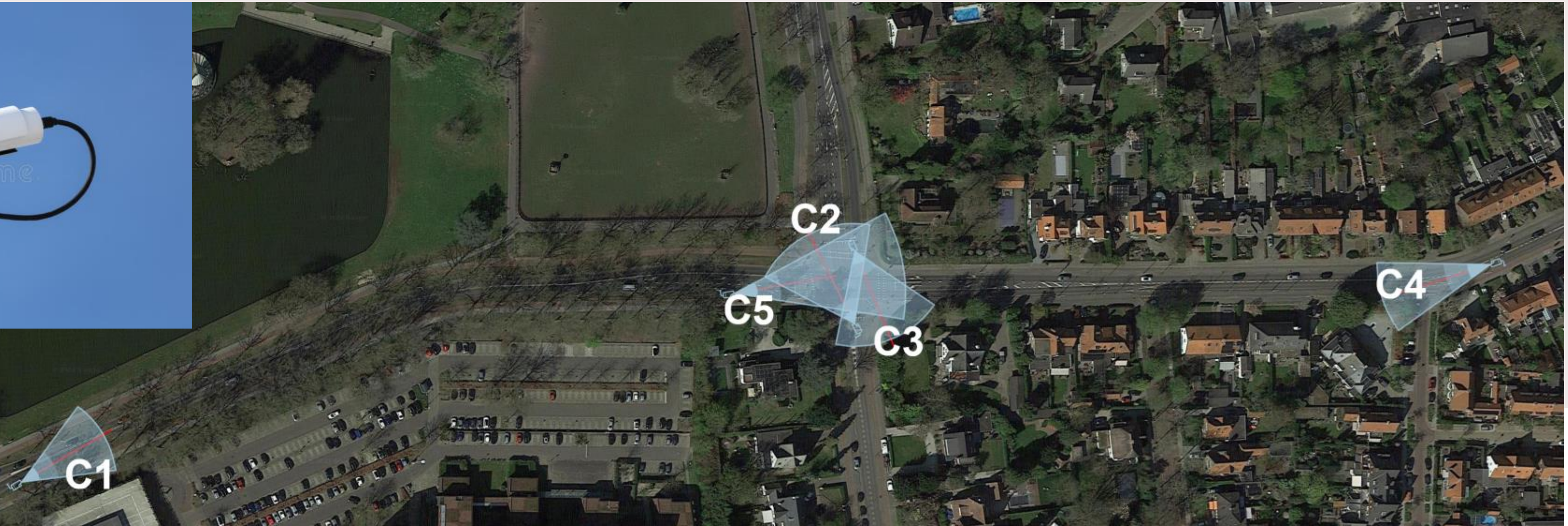
Intro: How to take the context into account?

Three principles

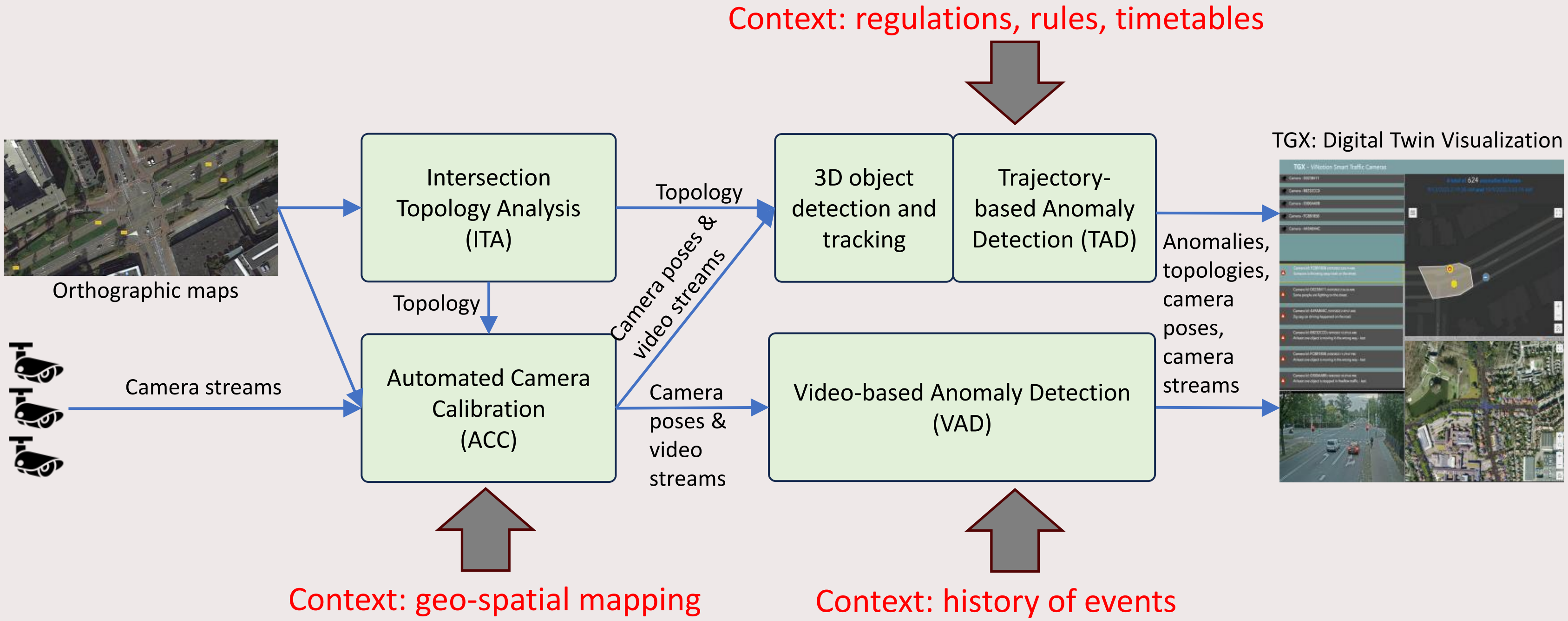
1. Observe, understand and remember the history of events preceding the anomaly
2. Bring the city/traffic regulations, rules, timetables into the analysis
3. Map the video (radar, acoustic) data onto the geo-spatial topological ground

SMART project: Fieldlab Helmond

- 2021 – 2023
- Partners: Royal Haskoning, Vinotion, CycloMedia, ESRI, TUE
- 5 cameras, ~100 TB traffic data
- Played anomalies
- Different traffic participants – vehicles, bikers, pedestrians

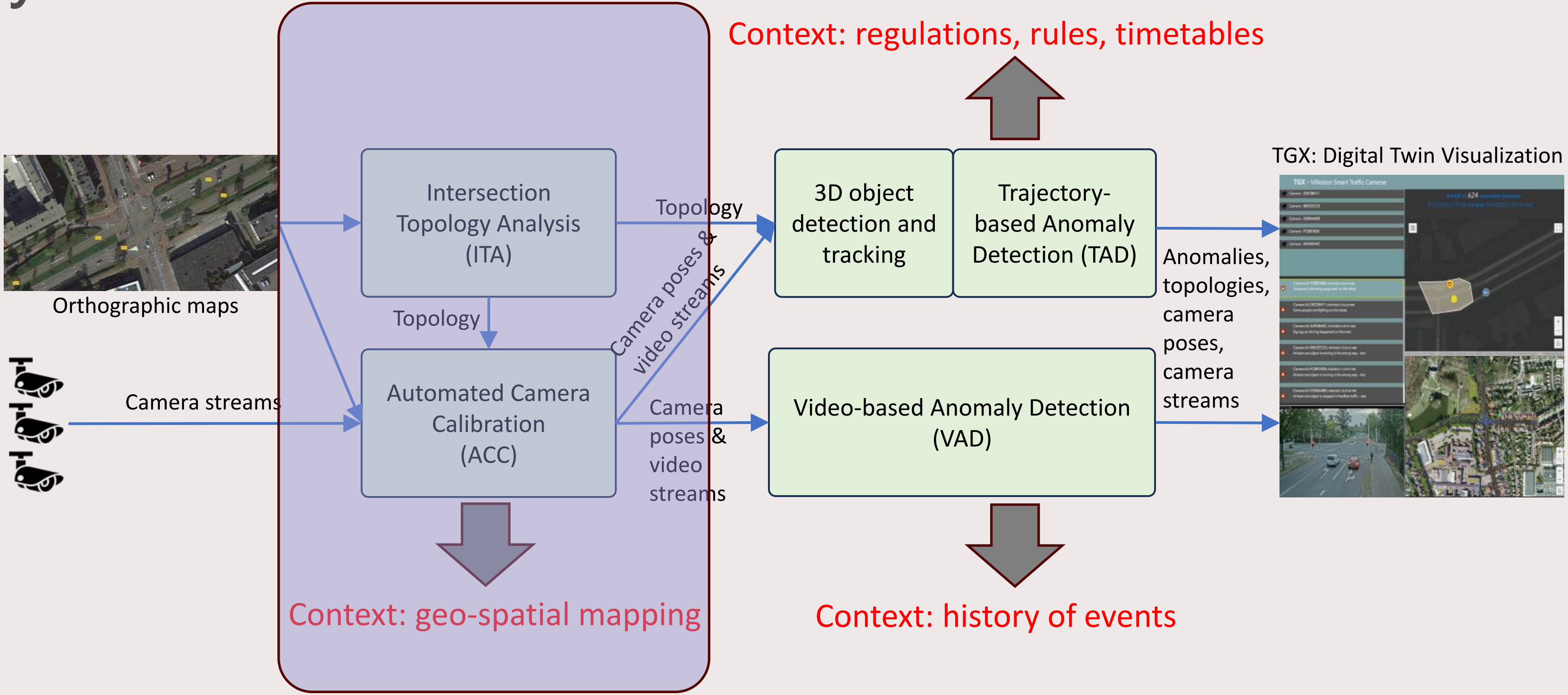


Architecture of SMART: geo-spatial anomaly detection system

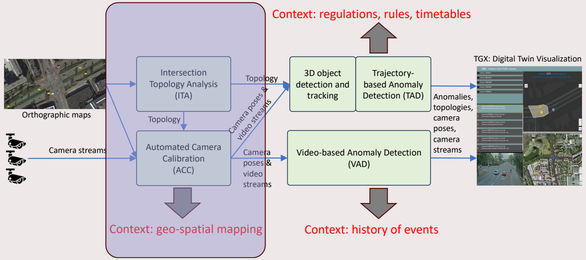


3D Vision Lab VCA TU/e

Architecture of SMART: geo-spatial anomaly detection system



Automated camera calibration



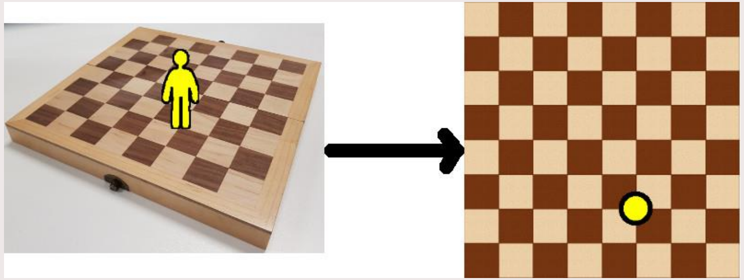
Detecting an anomaly in traffic requires an accurate and automated camera calibration:

- Some anomalies are correct actions performed in a wrong place → localizing the anomaly in real-world coordinates is very important

- Traffic cameras are mounted on lightposts and their pose can be altered by wind, vehicles passing by etc…

- The cameras need to be calibrated often and automatically with any weather or traffic condition

Calibration enables homography transformation



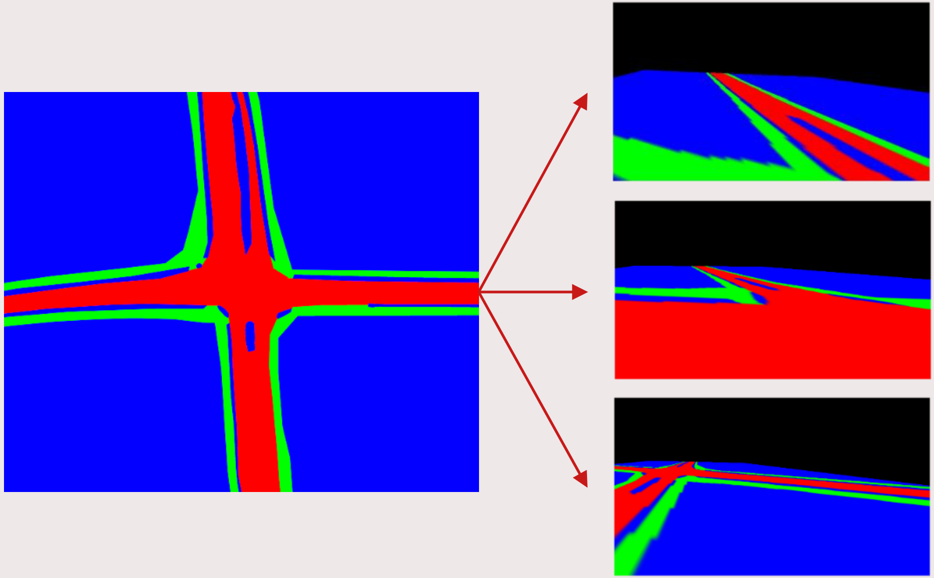
Idea: synthesize dataset of images from virtual camera poses

1. Take an intersection image from Google satellite imagery: birds-eye-view (BEV)

2. Semantically segment the image

3. Sample virtual camera with different focal length, locations and rotation angles to create homography matrices

4. Create training dataset: synthetic images (thousands) by warping the semantic bird's-eye-view with the sampled homographies

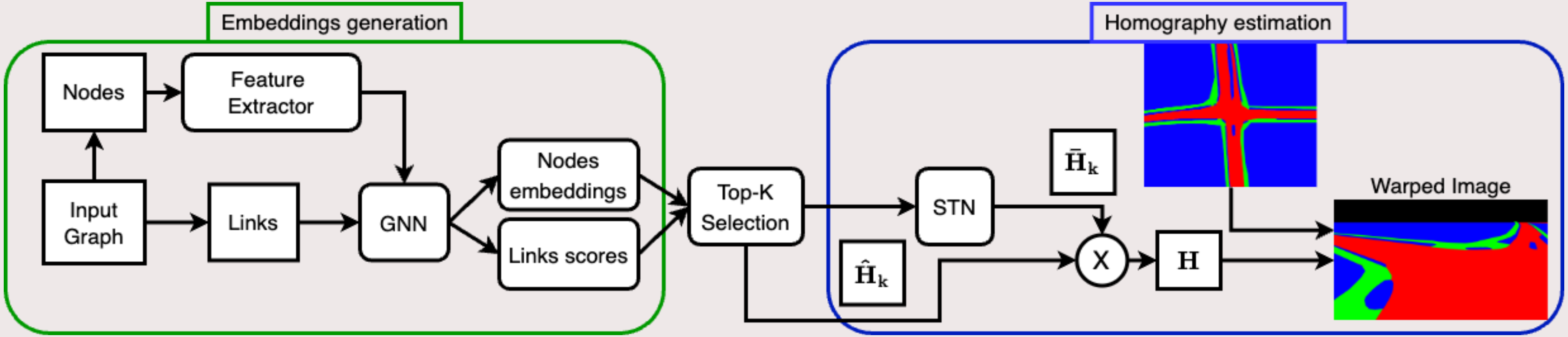


Homography estimation: automated camera calibration

The proposed framework consists in two main components:

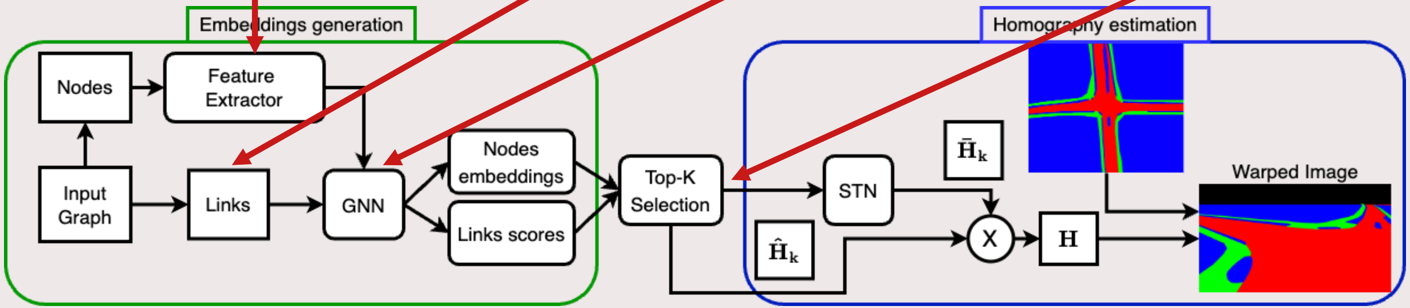
- **Embeddings generation – finding nearest virtual viewpoint for our camera view**
- **Homography estimation**

The embeddings generation component is trained individually and then end-to-end along with the homography estimation component



Methodology - Embedding Generation

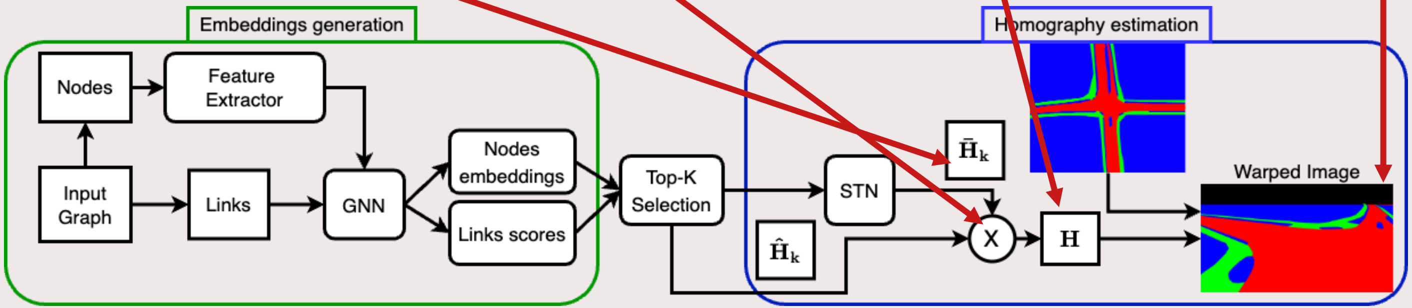
1) A CNN feature extractor is used to **obtain a feature vector for each node** in the mini-batch
2) We construct the matrix with **all possible links** between all the nodes in the mini-batch
3) The GNN is trained as a **link predictor** to score each possible link
4) The **nodes of the top-K links** for each *training/testing* node **are batched together**



Overview of the homography estimation framework

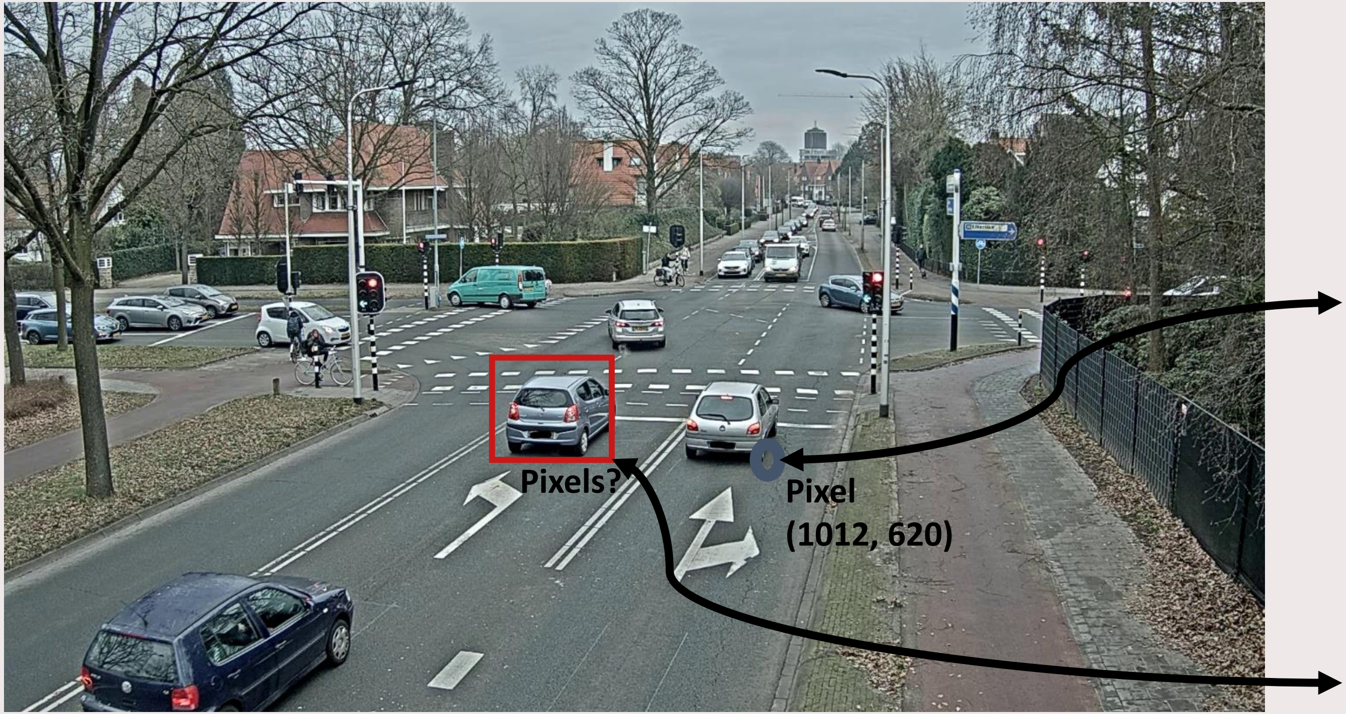
Methodology – Homography Estimation

1) The STN **estimates a homography $\overline{H}_k$** for each input node
2) $\overline{H}_k$ is multiplied with $\widehat{H}_k$, the homogrpahy **of the highest-scoring linked node**
3) H is used to warp the BEV of the intersection, producing a **warped image**
4) The complete framework is trained by comparing this image with the original input image



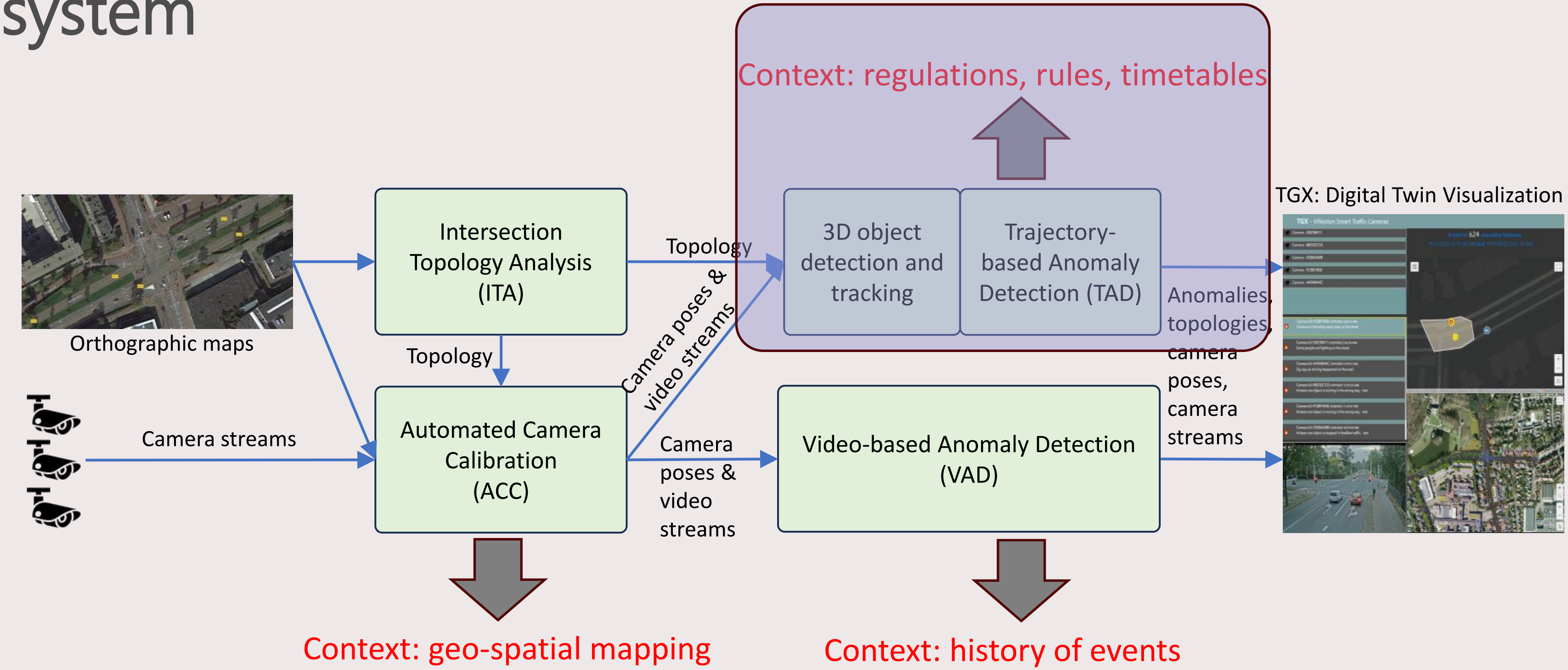
Overview of the homography estimation framework

Now we are able to map a pixel to real-world coordinates!



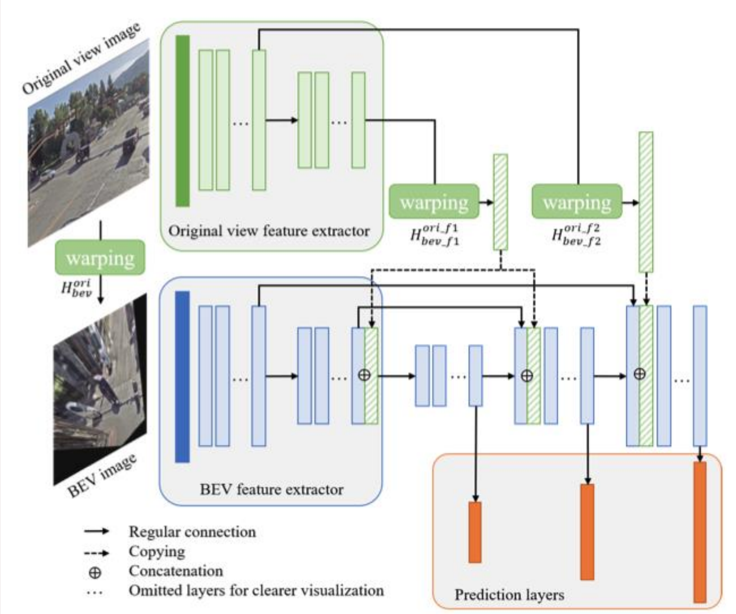
How can we map the agents to real-world coordinates?

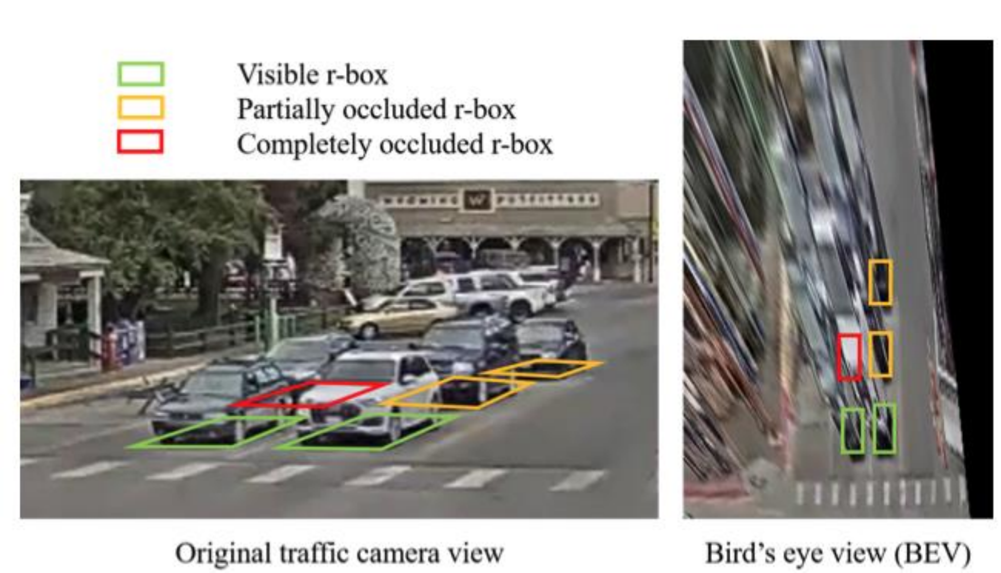
Architecture of SMART: geo-spatial anomaly detection system



Detect Agents and their 3D Bounding Boxes

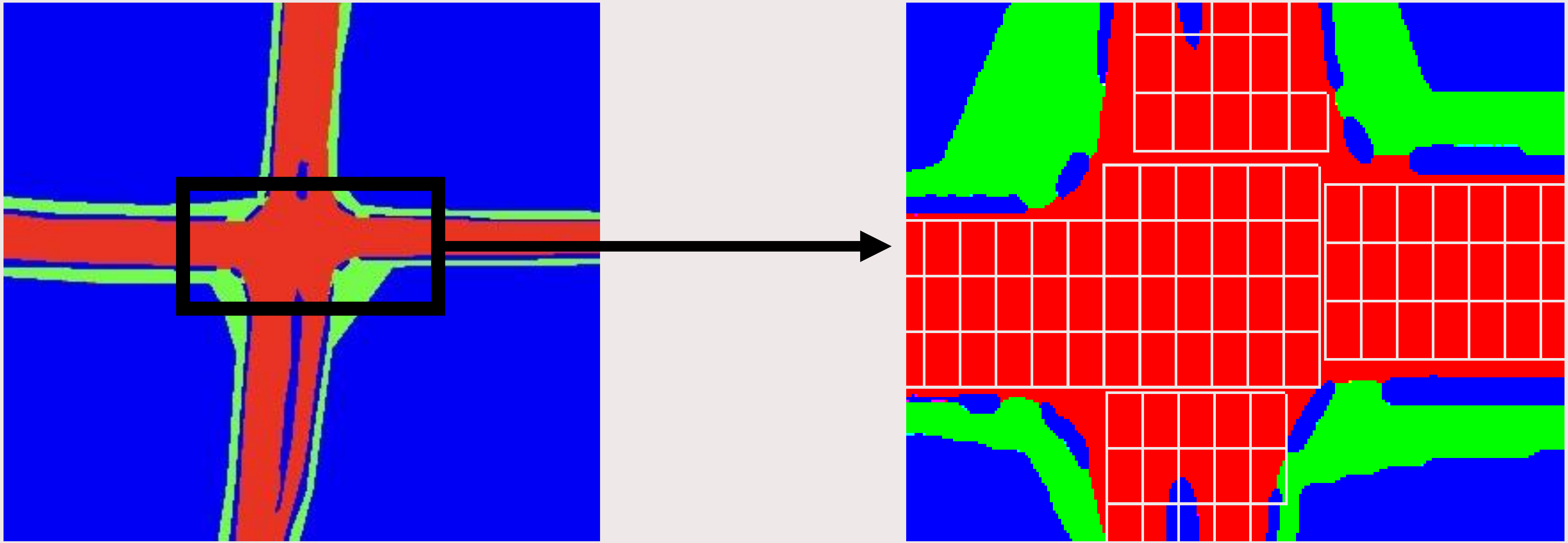
We leveraged the homography estimation component to detect the agents and obtain their 3D bounding boxes.





Create a grid for an intersection

The semantic regions of the intersection are split in grid with a fine granularity



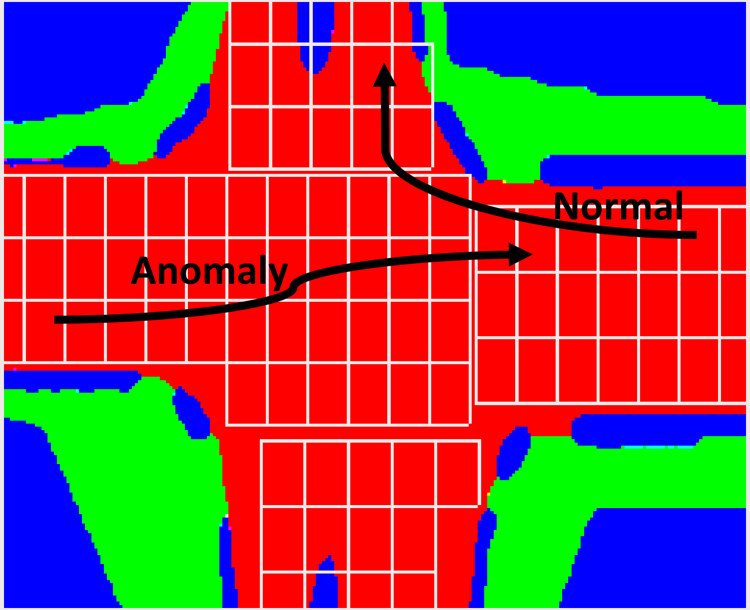
Train a normality model from standard moving trajectories

From 5TB of videos

- The agents are detected (3D bounding boxes) + at each timestamp their position is assigned to a grid cell.
- In this setting, we can learn the normal time spent in a cell and normality graph between cells.

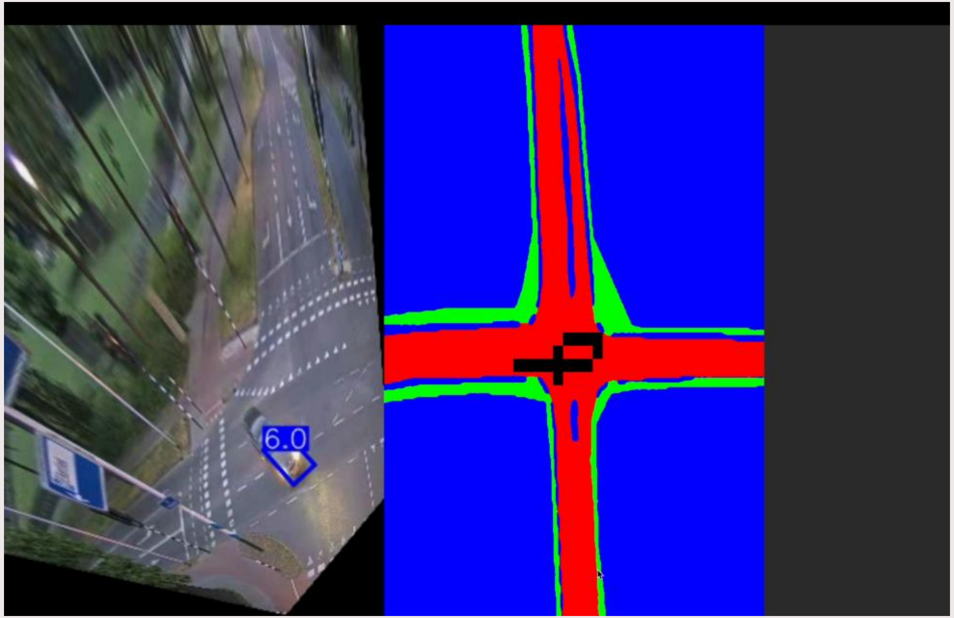
Most importantly, this allows us to:

- Define anomaly that we are looking for with simple if-else statements
  - E.g. if agent.Type == “Car” and agent.PatchType == “Terrain” then “Anomaly: Car on Curb”
- Simulate synthetic trajectories
- Detect & Classify anomalies that are not learned
- Further explore causal relations (Direct Acyclic Graphs)

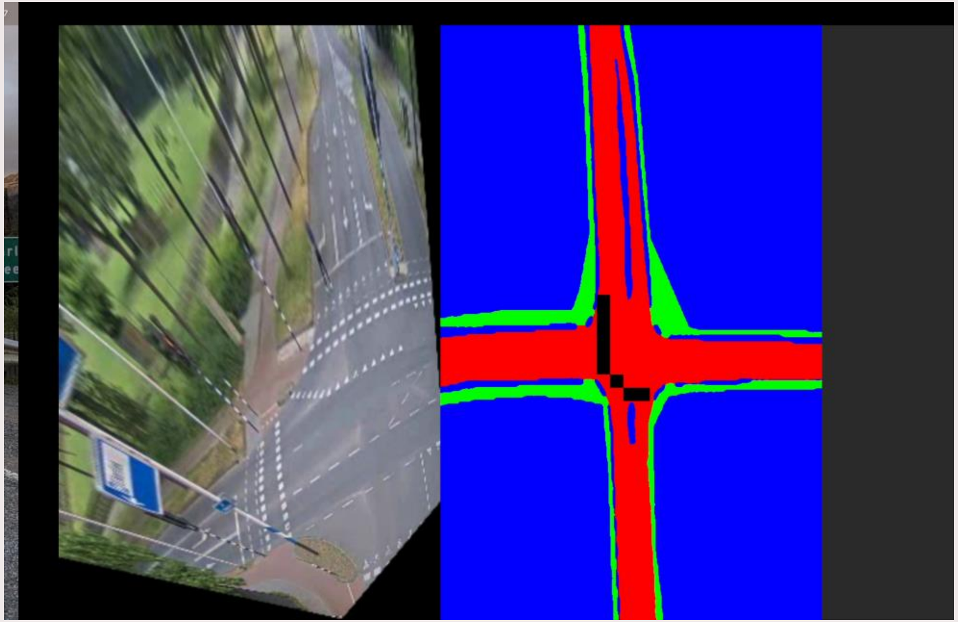


Anomalies

Donuts

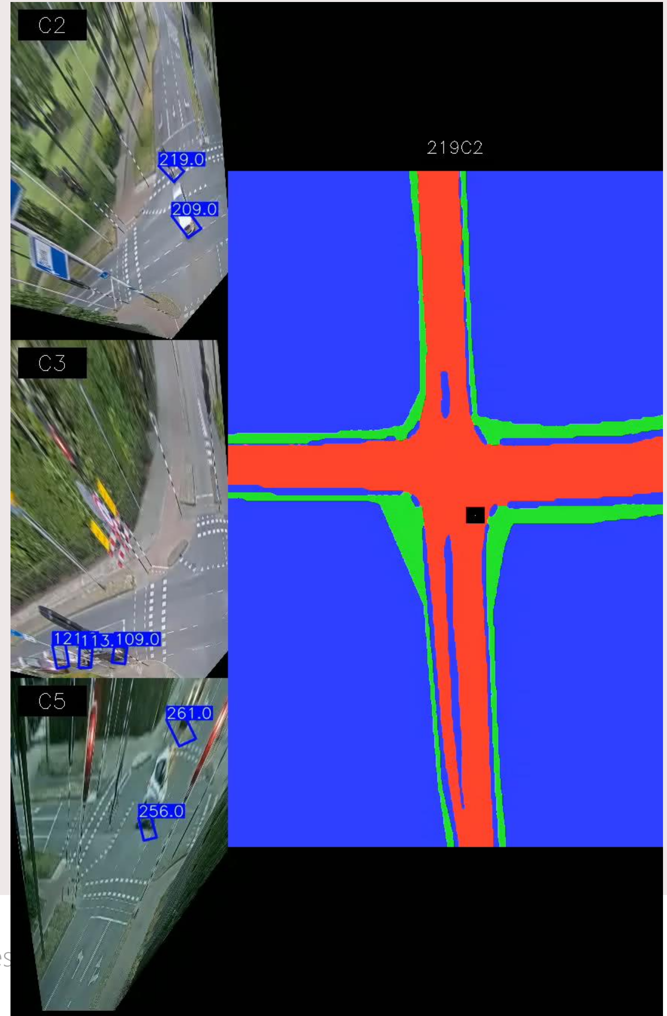


Jaywalking

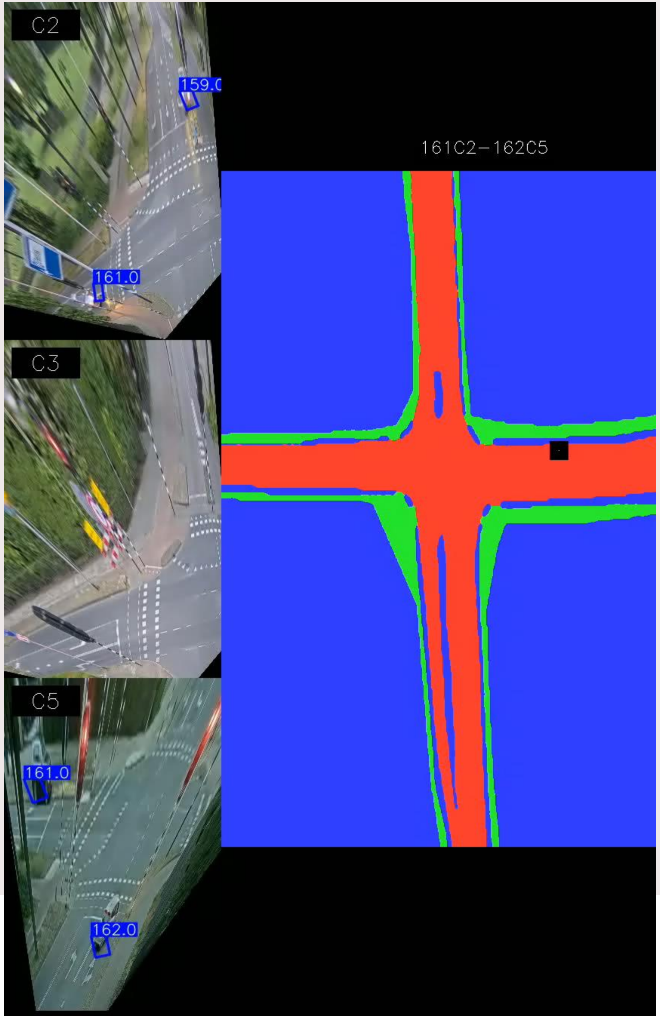


Anomalies

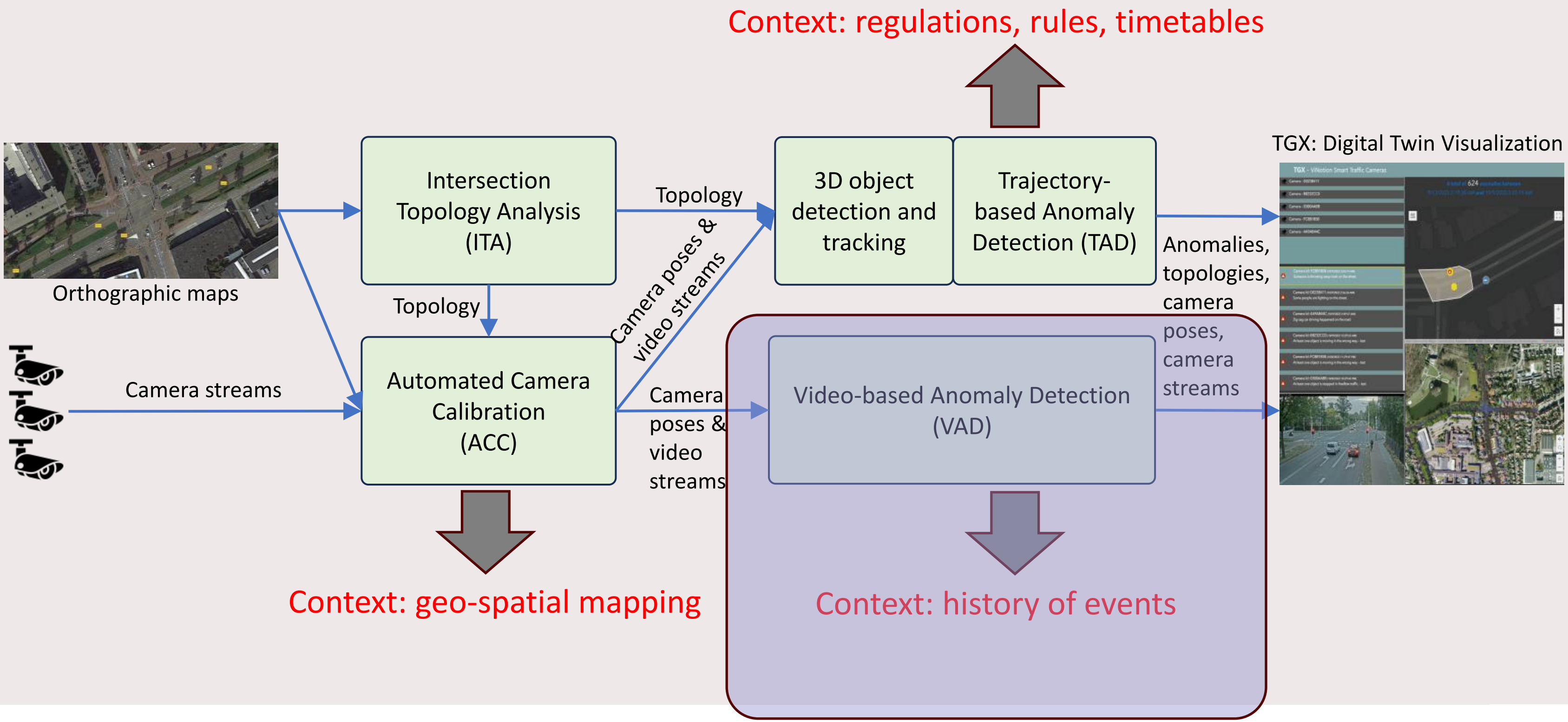
Biking on car lane



Wrong car turn



Architecture of SMART: geo-spatial anomaly detection system



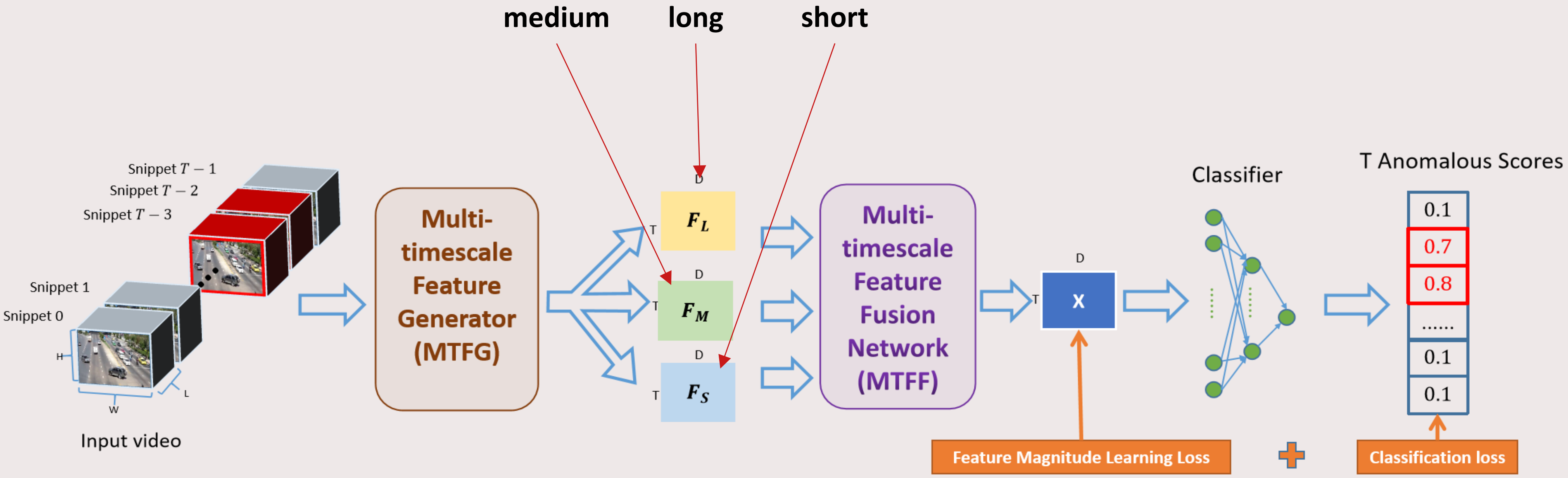
The Video Anomaly Dataset

- 2,202 train and 389 test= 2,591 videos
- 30 fps and 320×240 pixels
- EMV=Enclosed Motor Vehicles (cars and trucks)
- VRU=Vulnerable Road Users (motorbikes, bikes, and pedestrians)

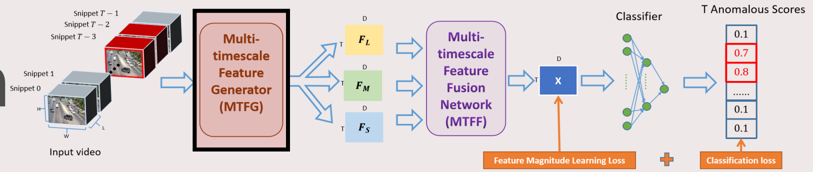
Class name	# Training	# Testing	Video sources
Normal	**900**	**170**	**1, 2, 3**
Dangerous Throwing	**154**	**25**	**2**
Littering	**81**	**12**	**2**
VRU vs VRU	**85**	**14**	**1, 3**
EMV vs EMV	**149**	**23**	**1, 3**
EMV vs VRU	**150**	**28**	**1, 3**
Abuse	48	2	1
Arrest	45	5	1
Arson	41	9	1
Assault	47	3	1
Burglary	87	13	1
Explosion	29	21	1
Fighting	45	5	1
Robbery	145	5	1
Shooting	27	23	1
Shoplifting	29	21	1
Stealing	95	5	1
Vandalism	45	5	1

Vision-based Anomaly Detection

1. Anomalies can be long, medium and very short in time
2. Learn various-duration anomalies by three different sampling strategies

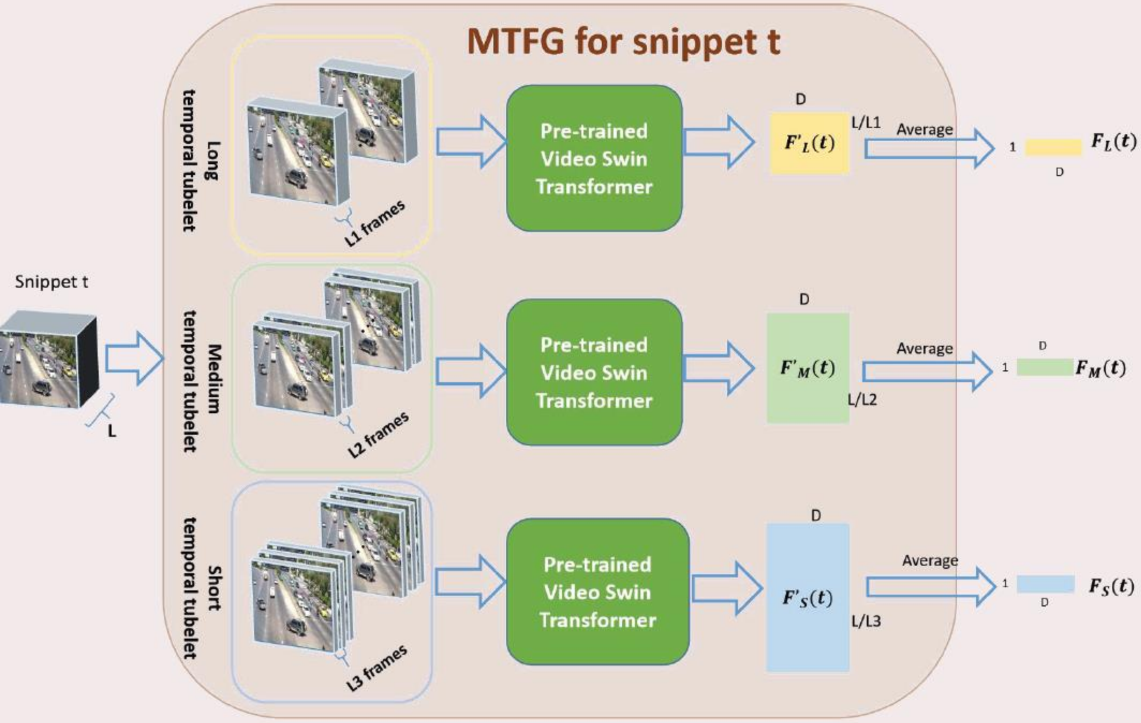


Vision-based Anomaly Detection

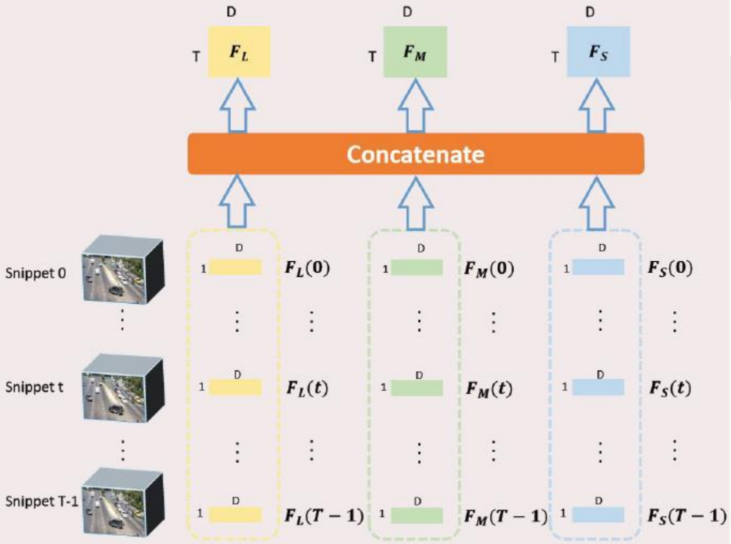


MTFG: Multi-timescale Spatio-temporal features

- Taking snippet t as an example



- Concatenating features



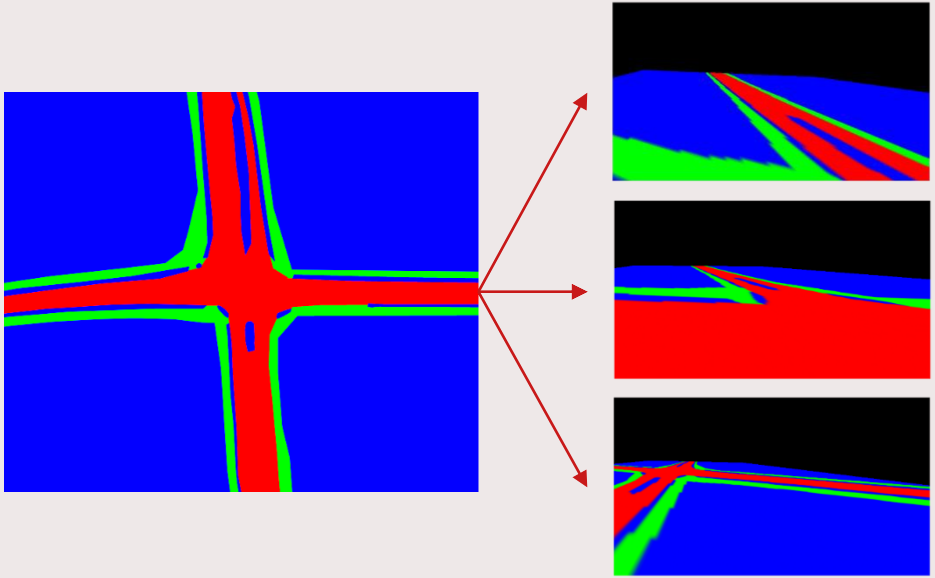
Examples of Video-based Anomaly Detection

Examples of Video-based Anomaly Detection

Discussion

Data Generation Pipeline

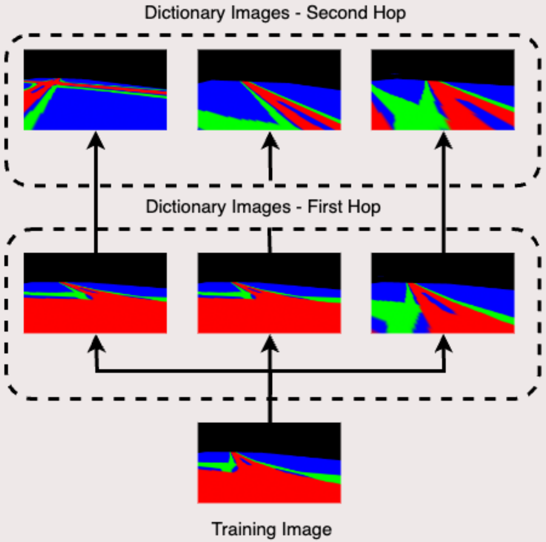
Warp the semantically segmented BEV with virtual cameras by sampling *focal length*, *rotation angles* and *location*



Randomly select *training*, *testing* and *dictionary* images

Structure the synthetic images in a graph

- Nodes = *training/testing* images
- Links = Top-20 most similar *dictionary* images

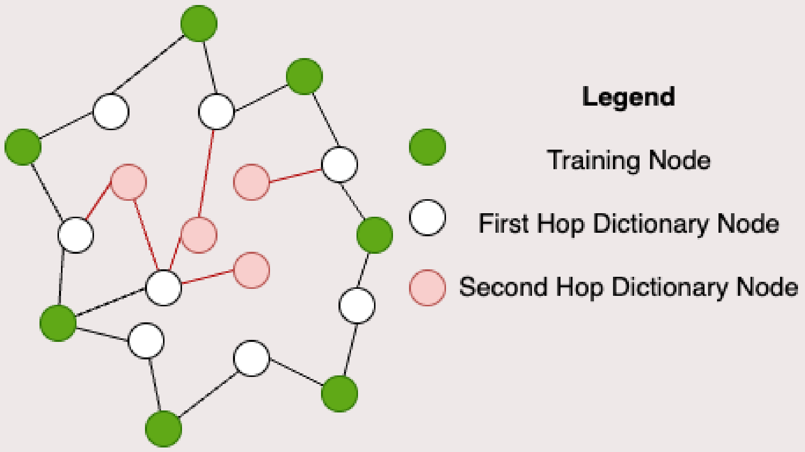


Data Generation Pipeline

A mini-batch is composed by sampling *training/testing* nodes and 10 of its dictionary nodes in the first or second hop

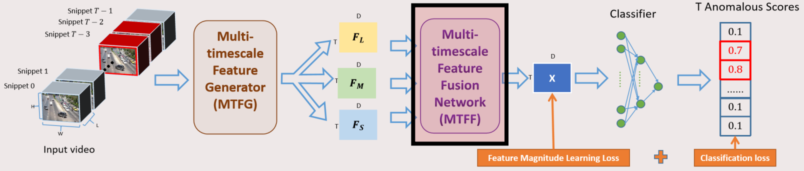
The dictionary nodes are shared between the *training/testing* graphs

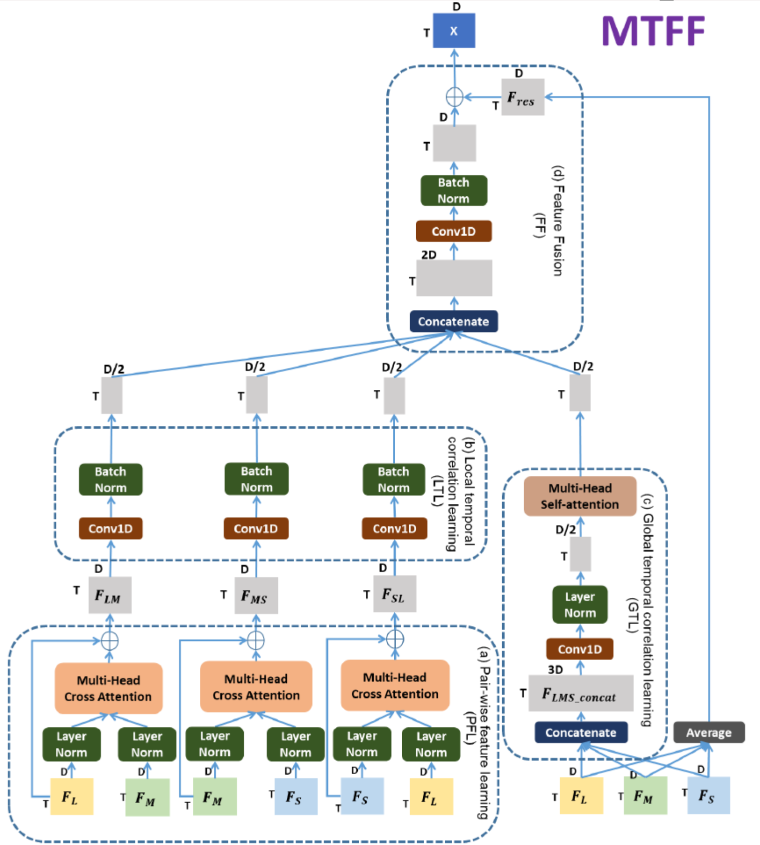
Some nodes may share dictionary nodes in the first or second hop



Reduced example of a mini-batch with 7 training nodes

Vision-based Anomaly Detection





- Pair-wise Feature Learning (PFL): Cross attention of short-medium-long temporal features.
- Local Temporal Correlation Learning (LTL): Scaling pairwise fusions by their local temporal correlation.
- Global Temporal Correlation Learning (GTL): Scaling the features based on the global temporal correlations.
- Feature Fusion (FF): Fusing features with a residual connection.

AI Detection of Traffic Anomalies

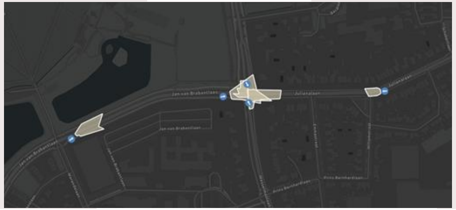
Real-time detection
All actors: bicyclists, cars, pedestrians
Detected anomaly types:

- Accidents
- Throwing and littering
- Illegal turning
- Opposite lane driving
- Zig zag driving
- Side lane parking
- Illegal crossing (biker, pedestrian)
- etc


Intersection topology
Camera streams

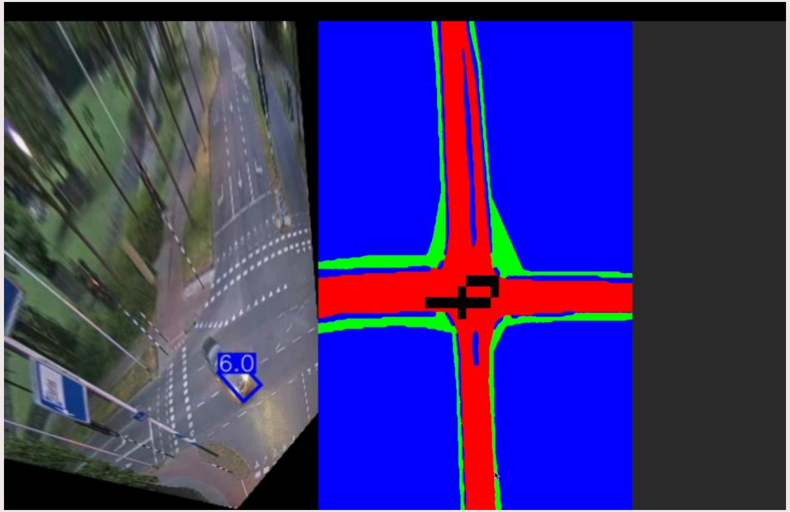
Automated camera calibration
3D localization of cars, bikers, pedestrians
Critical vehicle detection
Re-identification of cars, bikers, pedestrians
Image-based traffic anomaly detection
Trajectory-based traffic anomaly detection

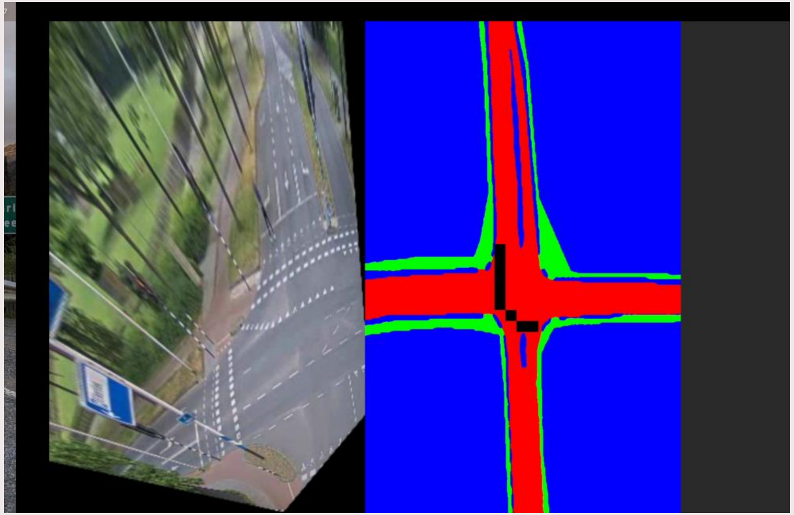
Visualization via TGX



VCA Video Coding & Architectures research group
TU/e

AI Detection of Traffic Anomalies

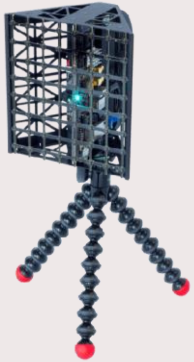




SINTRA project: Multi-modal sensor analysis against

- people trafficking,
- drugs smuggling,
- theft, intrusions

Vision-based people tracking in area: 2 pax



Acoustic-based people detection in area: 15 pax

Simple fusion: Early detection of possible people trafficking in containers

3D Vision Lab VCA TU/e