Utilizing deep object detector for video surveillance indexing and retrieval

Tom Durand, Xiyan He, Ionel Pop, Lionel Robinault

Context

- Foxstream – software for vidéo content analysis
  - Intrusion detection
  - Flow management

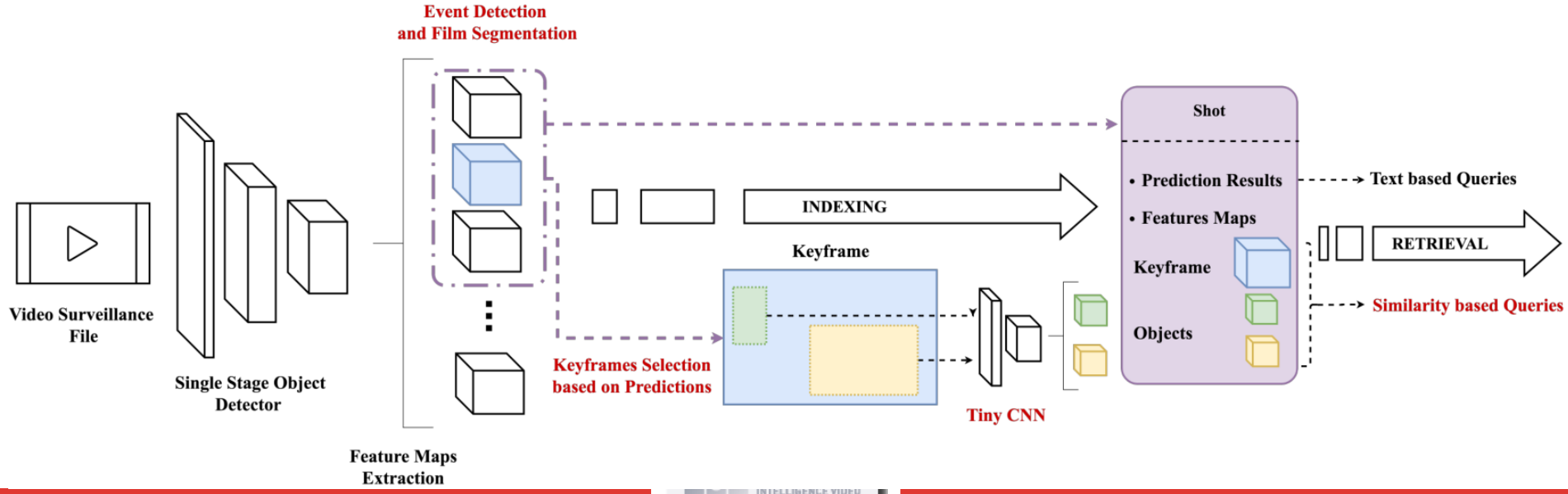
- Explore
  - new technologies
  - new applications

Purpose

- Technology : deep learning
  - Specialized (expensive) hardware
  - Hard to build (right), easy to use
- Application : video retrieval
  - Visual/text query
  - Fast
- Constraints
  - Low resolution video stream
  - Real time treatement (5 im/s)
  - Multi-stream

Solution idea

- Object detector/classifier
  - Video keyframes
  - Object extractor
- Tiny CNN
  - Object features

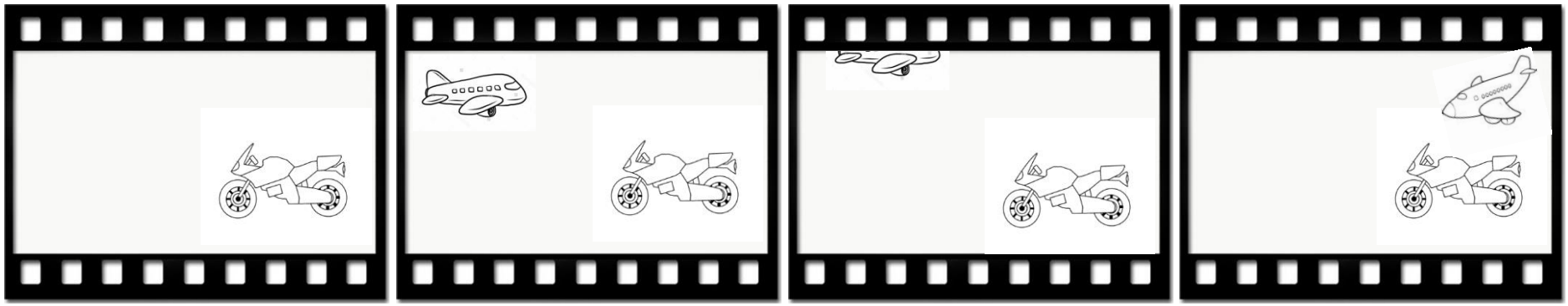


Keyframe extraction

↗ Target:

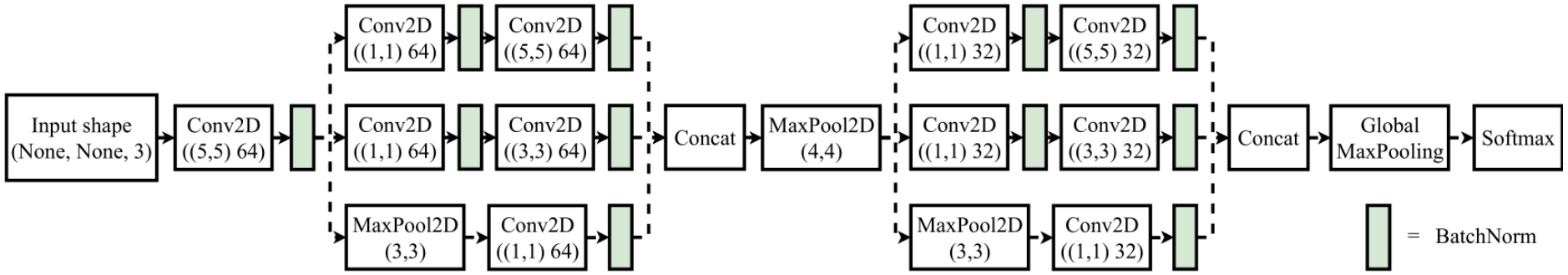
- Detect object apparition
- Each object is present at least once
- Limit the number of keyframes

↗ Use of class & objects position from YoloV2



Object features

- Target:
  - Fast
  - Low resolution input
  - Scale invariant



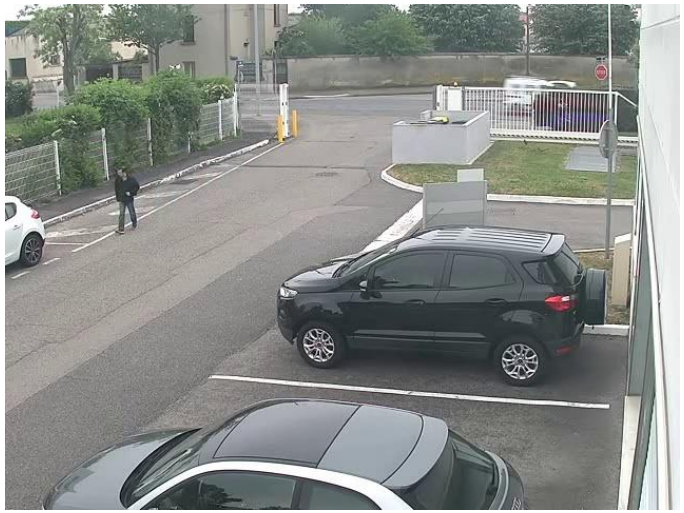
Indexation & retrieval

- Store
  - Semantic information (class + position)
  - Last layer from the classifier (scene descriptor)
  - Object features (object descriptor)

- Query
  - Scene
  - Object
  - Text (using Wordnet)

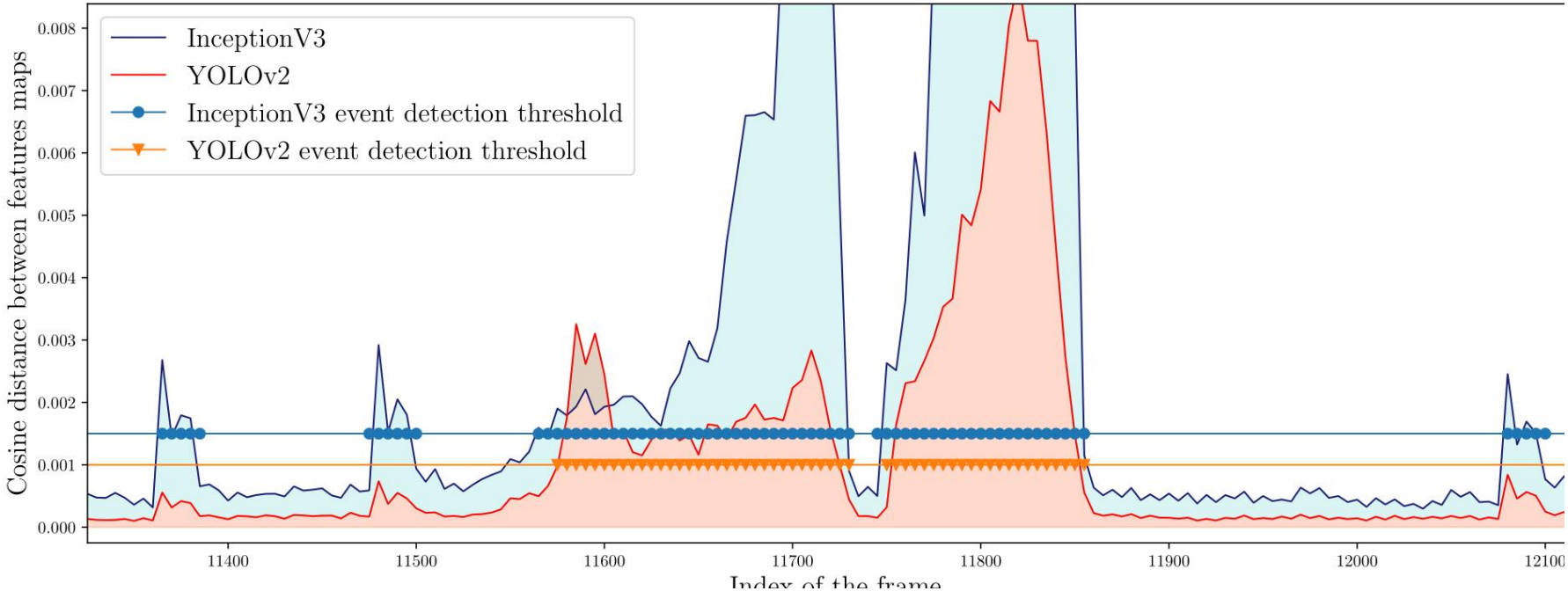
Results

- Data sets
  - PUT – Poznan University of Technology
  - CPM – Car Park Database



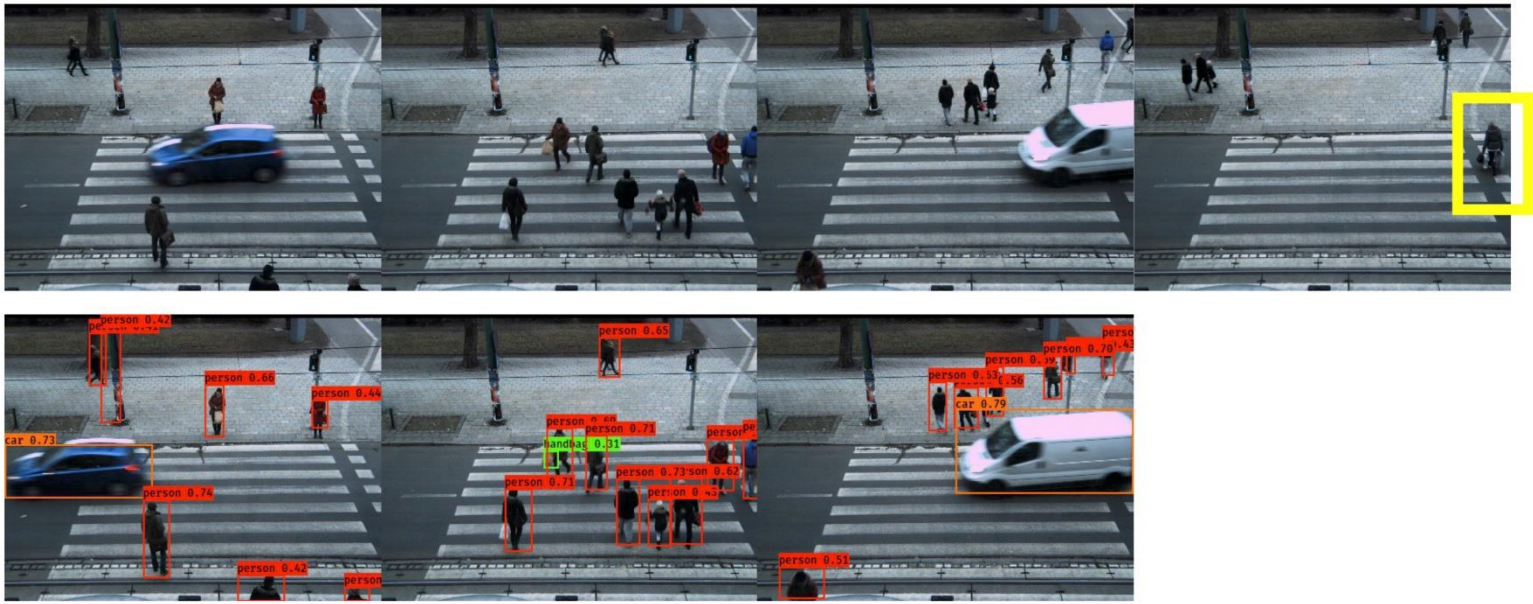
Results

- Feature pertinence
  - Similarity between consecutive frames
  - Compared against InceptionV3 trained on ImageNet



Results

- Key frame selection
  - Ground truth of a video of 1.5h with 100 events
  - 89.7% keyframes (of 117)
  - 1.5% missing objects (of 194)



Results

- TinyCNN
  - Compare against FisherVector and ResNet50
  - 6 classes (person, truck, bicycle, motorbike, other)
  - 7 hours of video

Methods	Fisher Vector	ResNet50	Tiny-CNN
Precision	21.3	**57.5**	46.3
Relevance	10	32.5	**36.3**
Inference Time (ms)	38	28	**6**
Number of Parameters	/	25M	**48K**

Methods	Fisher Vector	ResNet50	Tiny-CNN
Car	10	40	**60**
Truck	13.3	26.7	**40**
Bicycle	6.7	6.7	6.7
Motorcycle	0	**13.3**	6.7
People	20	**73.3**	60

Results

Conclusions & future works

- A prototype to index video data
  - At scene and object level
  - A long way to a product
- Tests on different scenario/scene/classes
- Evaluate different architectures of Tiny CNN
- Optimize the storage of indexed data