NATURAL VIDEO PROCESSING (NVP©) FOR BEHAVIOR RECOGNITION

Simon Polak - Chief Scientist, viisights
Menashe Rothschild - Chief Product Officer, viisights

viisights
Intelligence by vision

VIISIGHTS AT A GLANCE

TECHNOLOGY

BUSINESS

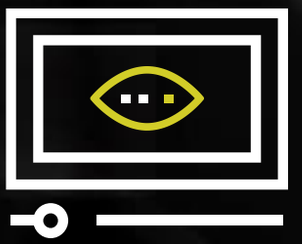
Cognitive Computing

Patented* Video Understanding Technology

Data Intelligence for Actionable Insights

Security & Safety Surveillance Intelligence

viisights
Intelligence by vision



DISRUPTING THE VIDEO ANALYTICS MARKET BY UNDERSTANDING VIDEO AND NOT ONLY CLASSIFYING IT'S OBJECTS

viisights
Intelligence by vision

PROBLEM STATEMENT

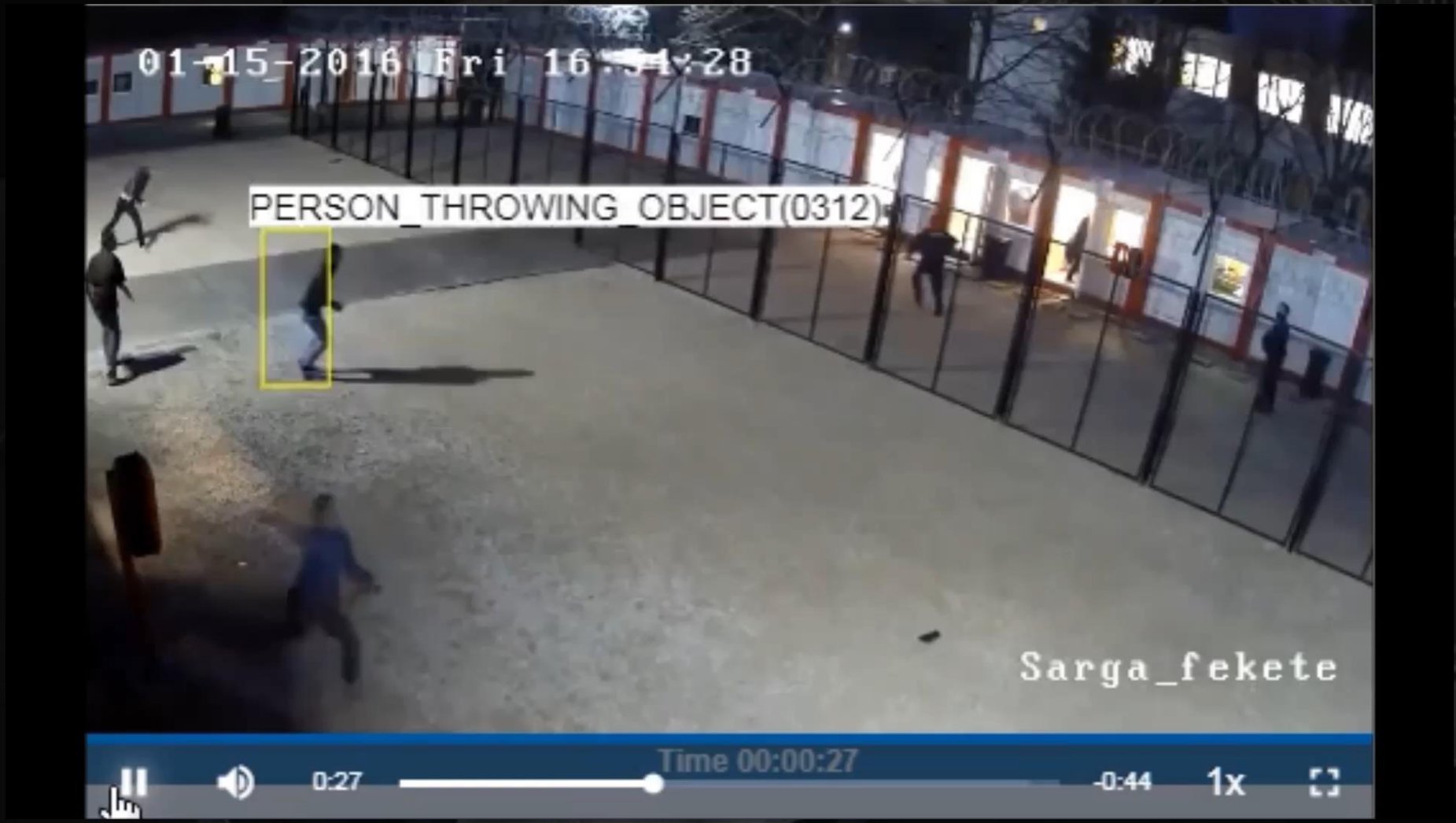
Monitoring of objects, **people, vehicles, appearance, behavior** and **activities** from a video stream for the purpose of detecting and predicting events of interest, such as dangerous behavior or criminal activity.

viisights
Intelligence by vision

BEHAVIOR RECOGNITION - EVENTS

viisights
Intelligence by vision

BEHAVIOR RECOGNITION – ACTIONS



viisights
Intelligence by vision

PROBLEM STATEMENT

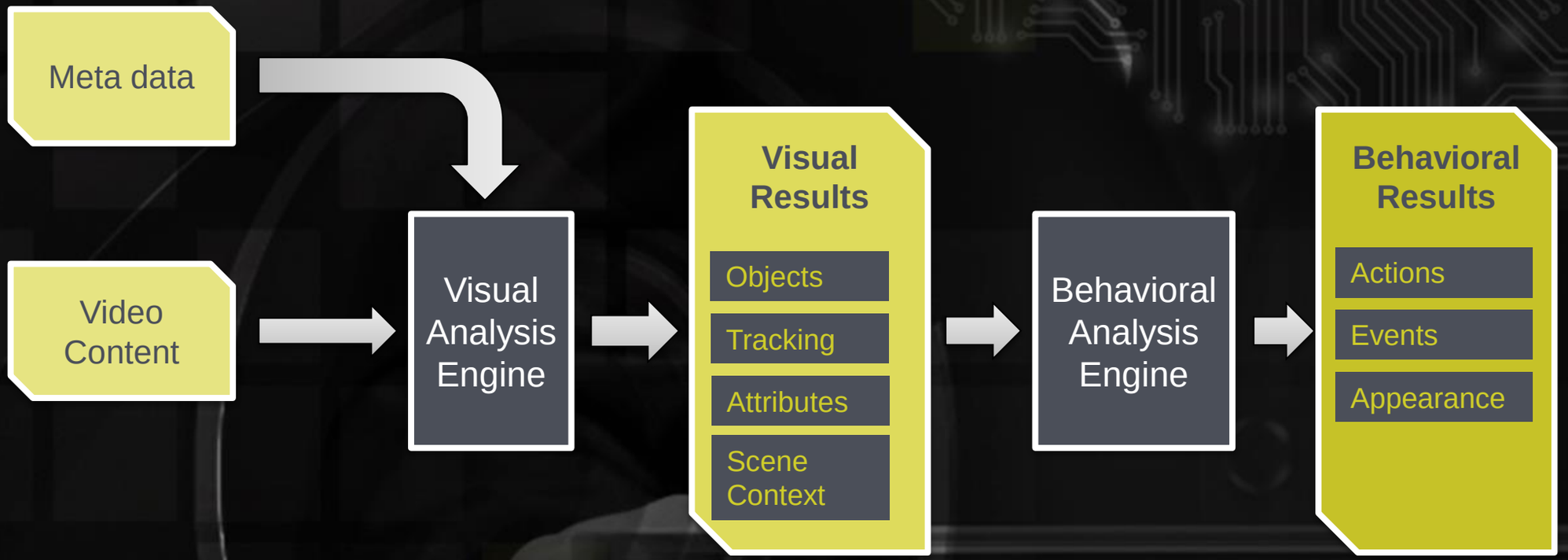
Monitoring of objects, **people, vehicles, appearance, behavior** and **activities** from a video stream for the purpose of detecting and predicting events of interest, such as dangerous behavior or criminal activity.

CHALLENGES

- Variable view points (street view, aerial, wearable camera)
- Variable video type (RGB, IR)
- Camera motion
- Small objects
- Real time, multiple streams per GPU processing
- High recall and low false positive rate requirement
- Rich object behavior and appearance description requirement

viisights
Intelligence by vision

HIGH LEVEL ARCHITECTURE

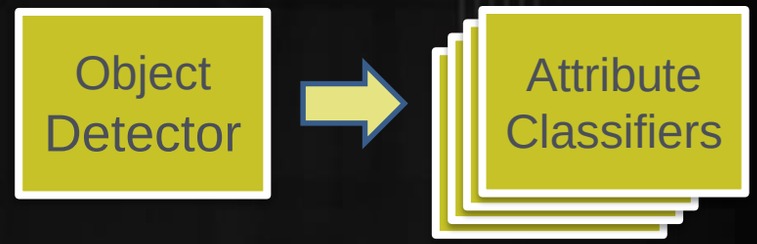


HIERARCHICAL OBJECTS AND OBJECTS' ATTRIBUTES DETECTION

HIERARCHICAL OBJECT DETECTION

- **The goal of the system is not only detect objects, but also provide object's description in terms of color, clothing, behavior etc.**

- **The standard approach for providing attributes of a detected object is :**
  - Extract windows of detected objects
  - Run additional classifiers in order to detect object's attributes

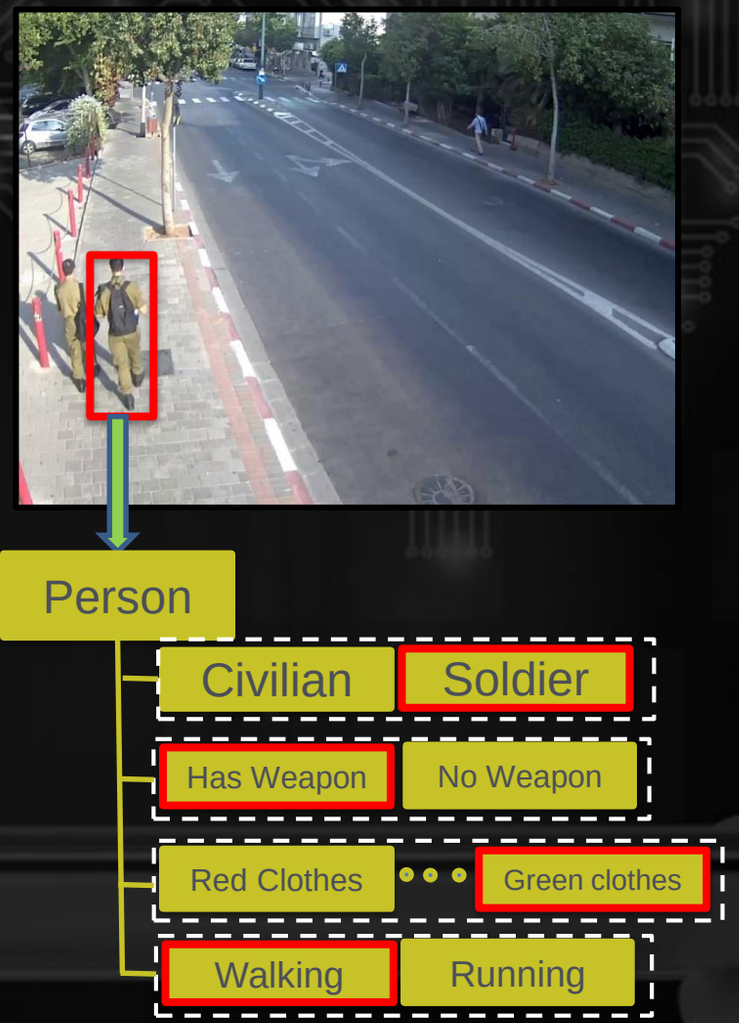


- **Drawbacks of this approach are:**
  - Higher time requirements – the same object is processed many times
  - Lower precision – attribute classifiers usually do not see the context and there is high dependency on the quality of the detected bounding boxes

HIERARCHICAL OBJECT DETECTION

- **viisights' object detector performs hierarchical classification of object type and its attributes for each detected object during the detection process.**
  - Objects are represented as a tree
  - The root representing "physical object", the second level has "person", "car", motorcycle etc.
  - Under each object there are branches representing the object's attributes.
  - For each tree edge the detector predicts a conditional probability, such as p("car" | "physical object") or p("soldier" | "person")
  - Probability of a specific attribute is calculated by following the tree branches. For example, p("soldier") = p("soldier" | "person") * p("person" | "physical object")

- **Advantages of this approach are:**
  - Lower time requirements – the additional time required for attribute classification is neglectable.
  - Higher precision – context is naturally represented, since the attributes classification is done directly inside the image
  - Enabling complex action and behavior recognition



viisights
Intelligence by vision

BEHAVIOR RECOGNITION

viisights
Intelligence by vision

BEHAVIOR RECOGNITION - REQUIREMENTS

- **Fast, in order to keep the system real time.**
  - Therefore, behavior recognition have to rely on the object & attributes detector output and not the raw video data
- **Robust and easily expandable.**
  - Thus, behavior analysis can not rely on heuristics
- **Support multiple concurrent actions and events, happening constantly and occupying only a small area of the video frame.**
  - Unlike the data in the standard academic datasets – AVA, UCF101
- **Utilize on a small training set, in order to be manageable.**
  - Therefore, a reliable augmentation and synthetic data creation method is needed

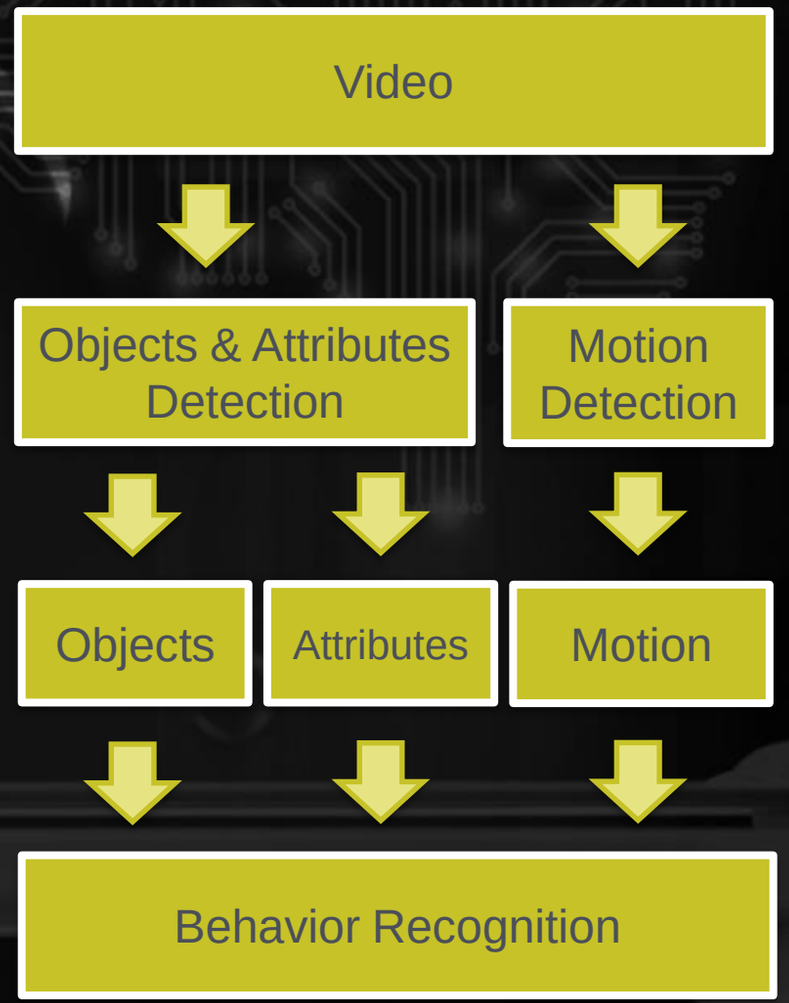
BEHAVIOR RECOGNITION - APPROACH

- **Recognizable behaviors types :**
  - Actions : behavior of a single object, such as person running.
  - Events : interactions between two or more objects, such as person getting into a car

- **Solution approach is similar to NLP**
  - *Objects*, *attributes* and *motion* are treated as words
    - Vocabulary of >200 words
    - Each word has an associated location
    - Each word has an associated (detection) probability
  - Behavior recognition => text paragraph classification

BEHAVIOR RECOGNITION

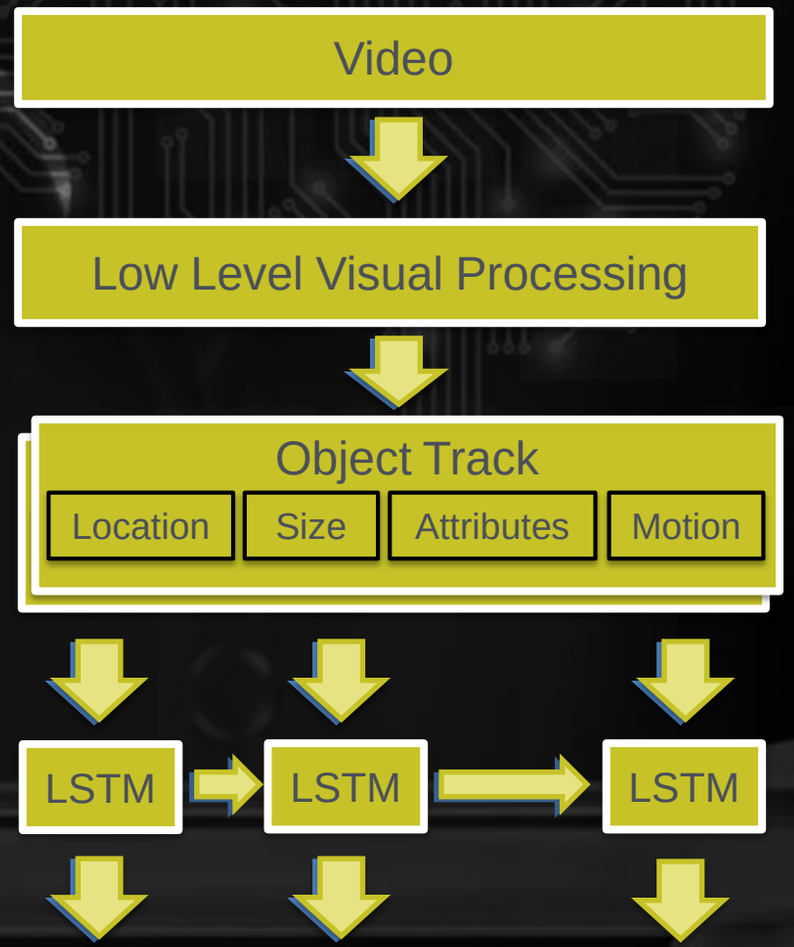
- The input to the behavior recognition module is the probabilistic semantic information produced by the hierarchical object detection module:
  - Objects – their location, size, detection confidence (nouns)
  - Attributes – their classification confidence (adjectives)
  - Motion – coarse motion values signifying which part of the object is moving (verbs)
- Actions recognition is based on following a track of an object in time
- Events recognition is based on following objects in an area in a video in time



BEHAVIOR RECOGNITION - ACTION

based on Long Short Term Memory (LSTM) neural network architecture

- Each instance in an object track is transformed into a vector containing object's locations, sizes, detections probabilities, attributes probabilities (coming from objects and attributes detection [OAD]) and motion data (coming from motion detection)
- Successive object's track instances are fed into LSTM network, which for each instance produces the probabilities vector for each defined action

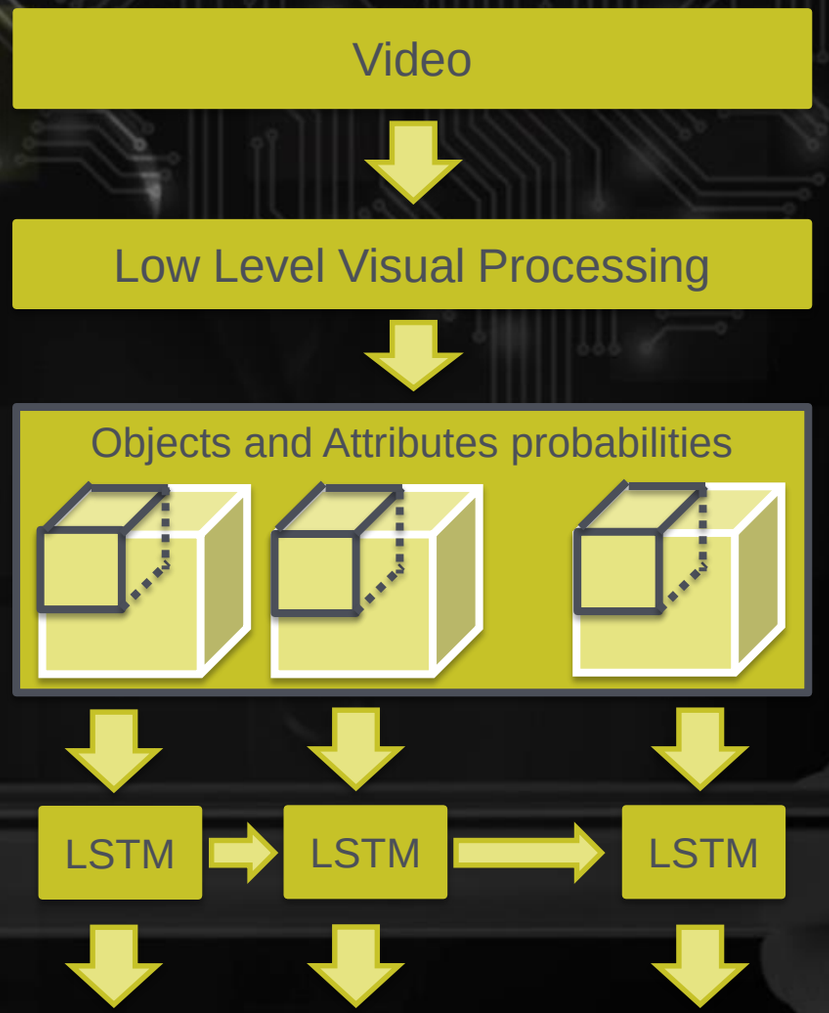


viisights
Intelligence by vision

BEHAVIOR RECOGNITION - EVENT

based on Long Short Term Memory (LSTM) neural network architecture

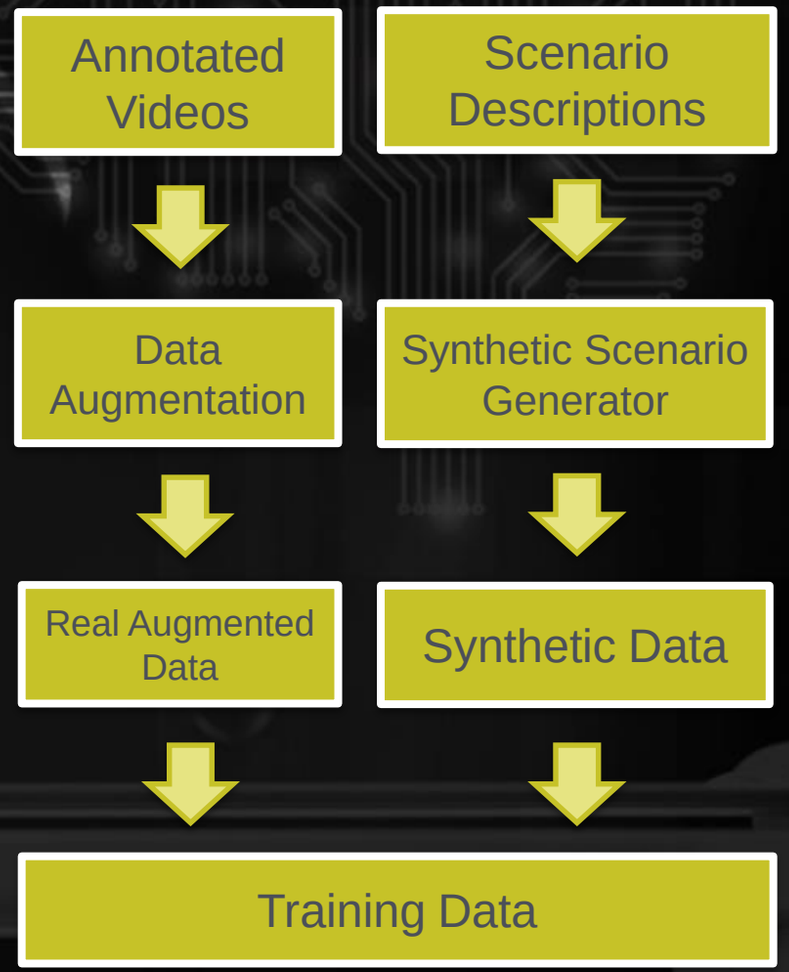
- Each frame in a video is processed by OAD and the result is represented as a tensor of objects' and attributes' probabilities. The frame is divided into NxN cell grid and a vector of length M in each cell represents the objects' data detected in this cell
- Windows in this tensor extracted from successive frames are fed into LSTM network which for each window in each frame produces the probabilities vector for each defined event.



viisights
Intelligence by vision

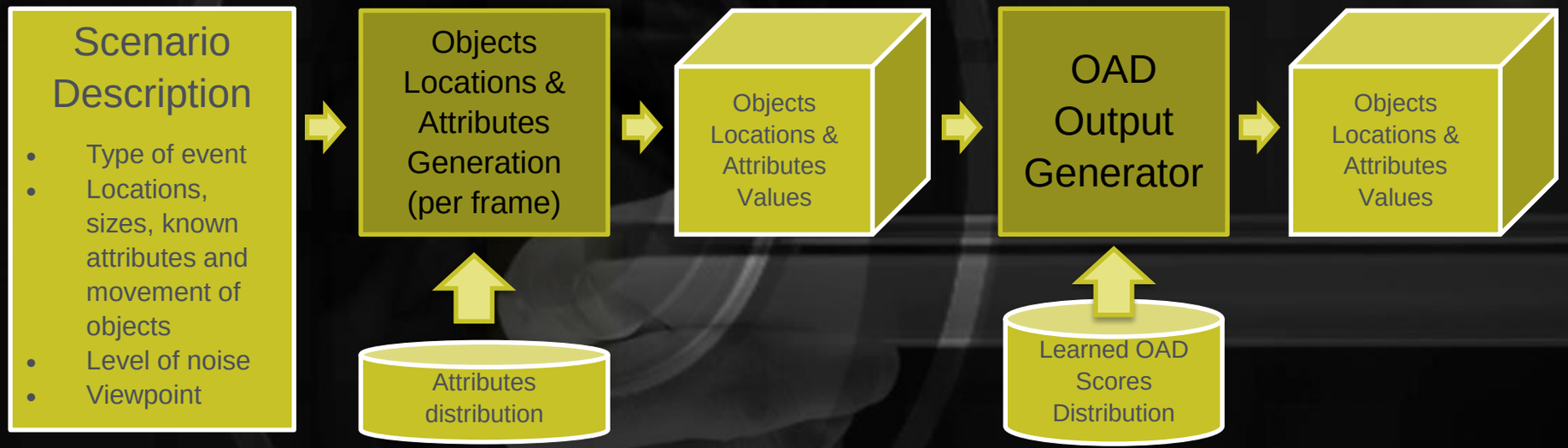
BEHAVIOR RECOGNITION - TRAINING

- Training a behavior recognition system can be prohibitively expensive – in order to reach sufficient precision you may need hundreds if not thousands of videos per each behavior.
- The design of the system, however, helps to make the training process much more manageable by allowing for fairly easy synthetic data creation.
- Since the input to the behavior recognition module are the objects and attributes probabilities and not the video frames, we model these input by measuring recognition probabilities of the object and attributes detector on a validation dataset and drawing from the estimated probability distributions.



SYNTHETIC SCENARIO GENERATOR

- Synthetic scenario generator takes as input a simple description of a scenario, containing the locations, types, motion paths and known attributes of objects.
- This description is used to create the objects' locations and their attributes values in each frame of the generated video.
- The output of the objects and attributes detector on this data is generated by drawing from estimated object detection probability distributions and attributes classification probability distributions



THANK YOU

for more information contact info@viisights.com

simon@viisights.com

menashe@viisights.com

viisights
Intelligence by vision