IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Computer Vision
Visual tracking using machine learning

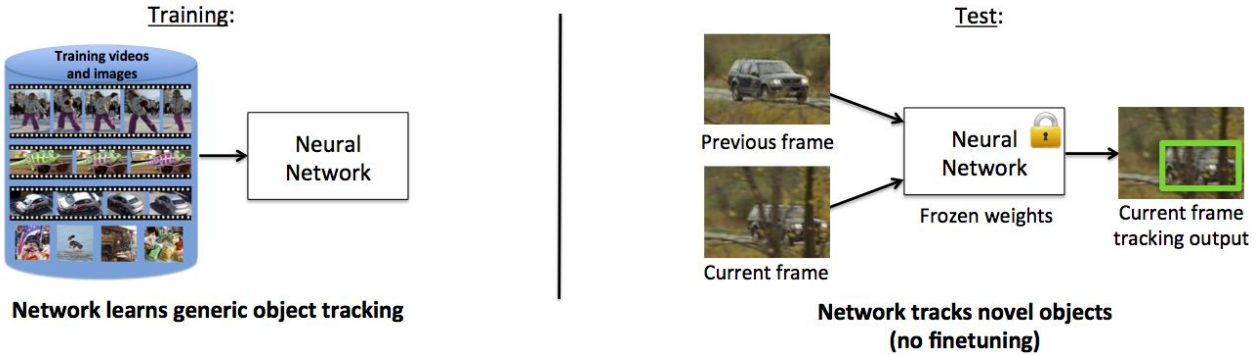
PASTORINO Martina
QUINTANA Gonzalo Iñaki
RIERA i MARÍN Meritxell
RODRIGUES DOS REIS Gustavo

1- GOTURN algorithm
2- Modification of the GOTURN algorithm
3- Datasets
4- Results
5- Perspectives

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

1- GOTURN algorithm

- Simple feed forward network with no online training required
- The tracker learns a generic relationship between object motion and appearance
- Can be used to track objects that do not appear in the training set
- Regression based approach



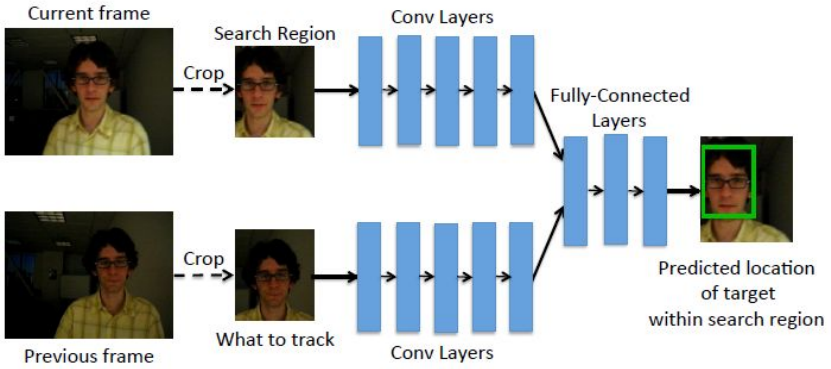
1- GOTURN algorithm

Inputs (two images):

- a *search region* from the current frame (time t)
- a *target* from the previous frame (time $t-1$)

Output:

- coordinates of the object in the current frame



The goal of the network is to predict the location of the target object.

2- Modifications of the GOTURN algorithm

Two **instances** of the GOTURN algorithm, which produce one bounding box each, for each frame. These bounding boxes should be combined in some way, in order to produce a final bounding box.

Forward-Backward method:

- One instance goes forward from the first frame.
- The other instance goes backwards from the last frame.
- Necessary the ground truth of the first and last frames.

Delta method:

- Two instances that go in the forward direction, separated by “delta” frames.
- Only the first ground truth for initialization is needed.

Two ways of combining the bounding boxes from the two instances:

- **IoU**: choose the bounding box that has the highest IoU with the ground truth. Unrealistic approach, as it is necessary to know all the ground truths.
- **Mean**: mean of the two bounding boxes. More realistic approach.

3- Datasets

Bag

Bear

Book

Camel

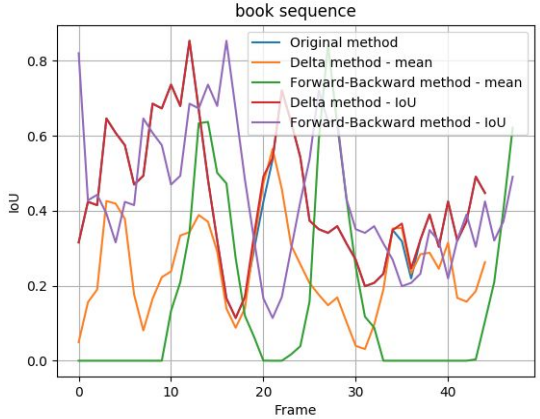
Rhino

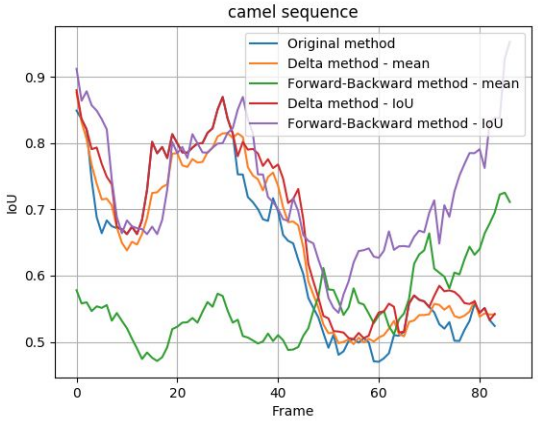
Swan

4- Results

To better quantify the results, the IoU between the ground truth and the predicted bounding boxes was calculated for the four most challenging datasets :

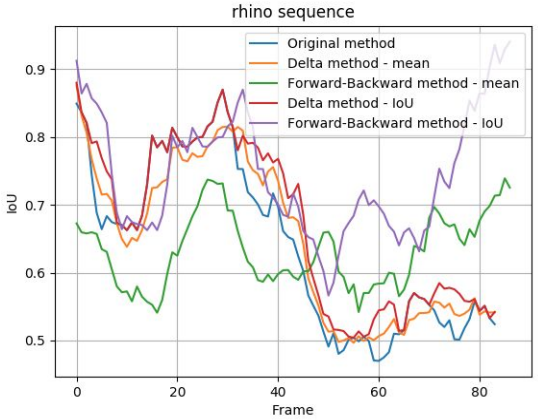
- **Book**: rotations of the object and sudden changes of size.
- **Camel**: similarities between the color of the background and another camel.
- **Rhino**: occlusions.
- **Swan**: shape (of the neck) and speed.

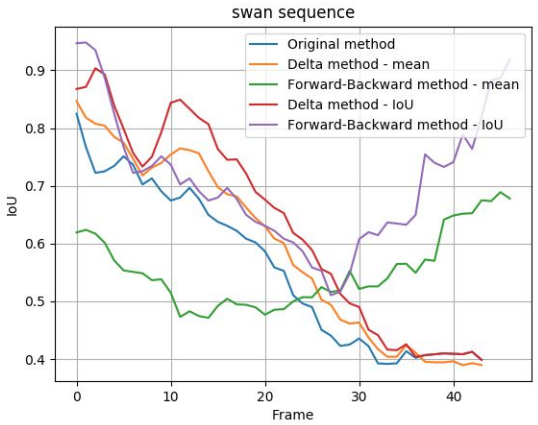




4- Results

- The Forward-Backward method clearly improves the IoU in the last frames, as the Backward instance begins in the last frame.
- For the rhino dataset, the Forward-Backward method helps to overcome the occlusion problem.
- For the book dataset, the best method is the Delta method, since considering two frames shifted by = 5 can help with the rotations and the sudden change of size. However, these good results were achieved using the IoU as the combination criterion, which is really unrealistic.
- The Delta method with the mean combination criterion didn’t produce really good results.





5.Perspectives

Some of the possible future improvements of the network:

- Enlarge the search region;
- Consider two instances that use different search regions;
- Use a prior for the combination method (way of combining the proposed bounding boxes of several instances):
    - For example, if it is known that in the video sequence there are no occlusions, the smooth variations in the bounding boxes should be prioritised;
    - On the contrary, if there are lots of occlusions, sudden changes of shape and rotations, big changes in the final bounding boxes should be prioritised.