DisNet: A novel method for distance estimation from monocular camera – 2018

Muhammad AbdulHaseeb, Jianyu Guan, Danijela Ristić-Durrant, Axel Gräser, Member, IEEE

**Introduction :**

Novel method for object distance estimation from a single image with no prior knowledge of the camera parameters or the scene.

**Hypothesis :**

* Reliable and accurate object detection is still a problem nowadays.

**Method :**

Utilisation de YOLO (base de données COCO)

YOLO 🡪 calcul de la boite d’ancrage (bounding box) des objets

* En fonction de la géométrie on classifie l’objet et on détermine sa distance

DisNet 🡪 Calcul distance par rapport à la caméra (relation inverse entre dimension et distance)

**Références :**

[11]: Stereo vision method for collision warning

[3]: Advanced driver assistance system using computer vision and laser scan

[1]: Obstacle detection technique based on stereo vision (using monocular cues)

[13]: 3-D Depth Reconstruction from a single still image

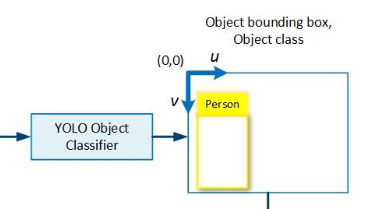
[16]: Stochastic optimization method Adam (appropriate for non-stationary objectives and problems with very noisy and sparse gradient)

**Technique:**

We first train the YOLO Object Classifier with COCO dataset.

The camera image is sent as an input to the YOLO object detector.

This same Object classifier gives the bounding box of each detected object in the images as an output and label each one of them.



For each extracted bounding-box, they calculate a six-dimensional feature vector like this:



Where:

Bh = height of the object bounding box in pixels/image height in pixels)

Bw = width of the object bounding box in pixels/image width in pixels

Bd diagonal of the object bounding box in pixels/image diagonal in pixels

Ch = Average height

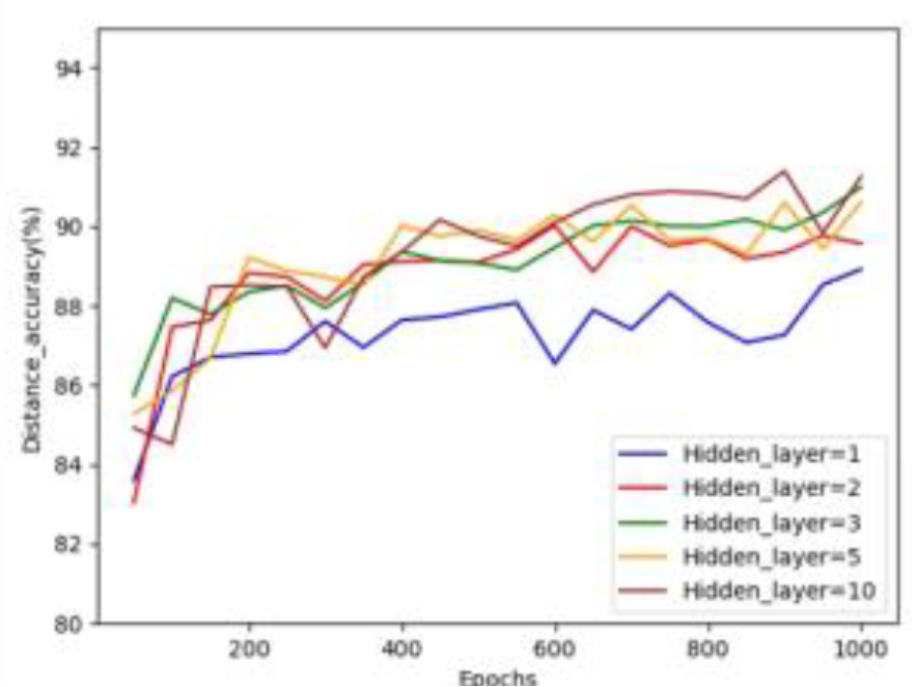
Cw = Average width

Cd = Average diagonal

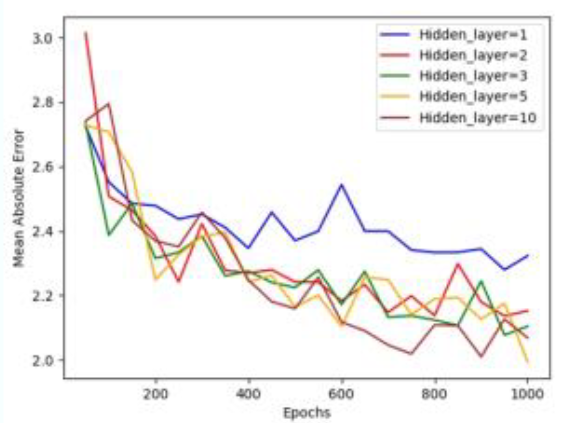
These bounding are then processed by the trained DisNet Multi Hidden-layer Neural Network.

**Disnet Structure :**

The DisNet algorithm has been tested with 1,2,3 ,5 and 10 Hidden Layer at first, after thorough test it has been shown that the 3 hidden layer structure was the best one because it has approximatively the same Mean Absolute Error value as the 5 and 10 hidden layers.



**Figure : Distance estimation accuracy achieved for different number of hidden layer**



**Figure : Mean Absolute error achieved for different number of hidden layer**

We can see that only one hidden layer is not nearly enough because it has a bad mean absolute error and a bad accuracy rate.

We can also see that 10 Hidden layer has the best results in both test but a trade off has been made between computational time and the final result.

And then to decide the number of hidden neurons in each layer they also tested the DisNet over 1000 epoch for 10, 30, 100 and 200 neurons.

Input (6 entrées)

Output (1 sortie)

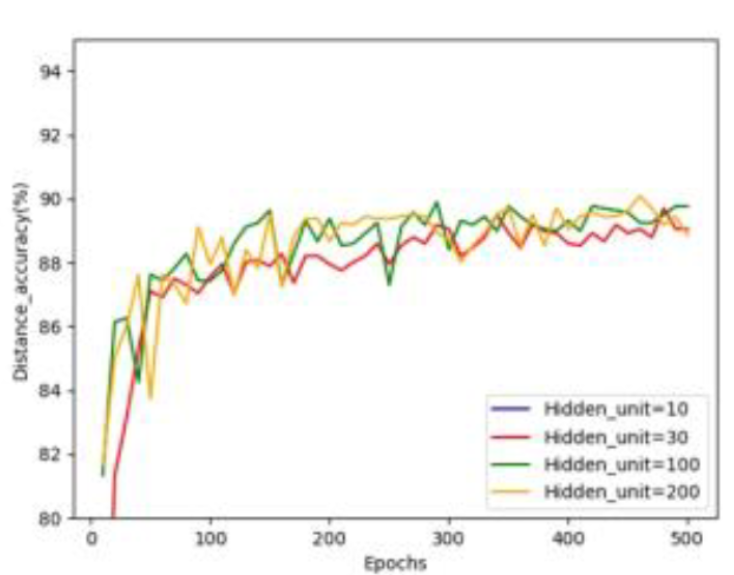
Hidden Layer (100 neurones)

Hidden Layer (100 neurones)

Hidden Layer (100 neurons)

**Input**: 6-dimensional feature vector

**Output**: Distance



**Figure : Distance estimation accuracy achieved for different number of hidden neurons**

In this graph we can see that there is a clear difference between 30 hidden neurons and 100/200 neurons, but not much of a difference can be seen between 100 and 200 neurons, so 100 hidden neurons have been chosen.

**Training :**

There is a relationship between the calculated dimension of the bounding boxes and the distance of the obstacle. Geometrically, by the projective transformations, the object bounding box size is expected to get smaller the further away the object is, so the inverse of the bounding box size is expected to increase as the distance increase.

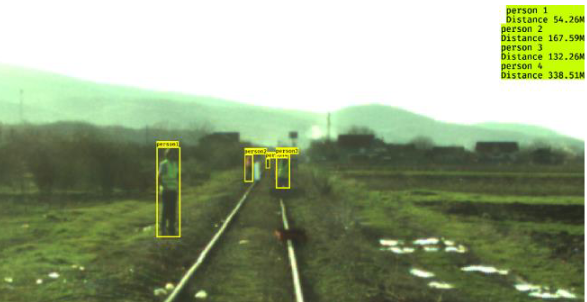
For training the network the input dataset was firstly randomly split into a training (80% of the data), validation (10% of the data) and test set (10% of the data). The DisNet was trained using the backpropagation method with the **Adam optimizer** backpropagation method on the dataset collected

But the algorithm was never trained with real railway track scene.

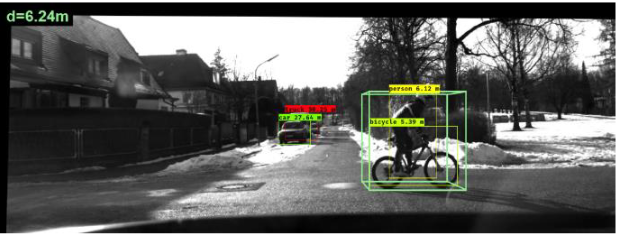
**Evaluation de la technique:**

The DisNet method is made for Railway situation so to evaluate the network they used straight rail tracks images of day and night location. They used a laser for the ground truth.

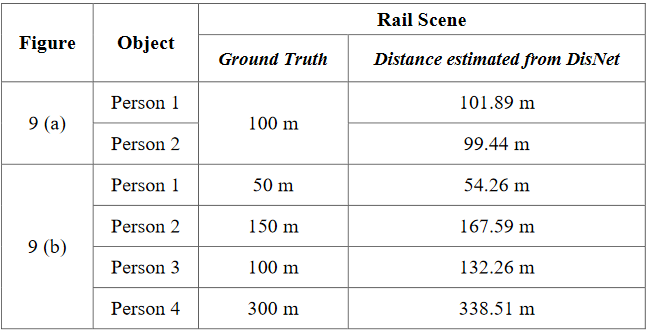
(9.a)

 (9.B)

The algorithm was then tested with road scene against a HiSpe3D-Vision which is a collision warning system for automobile.



**Résultats de la technique :**

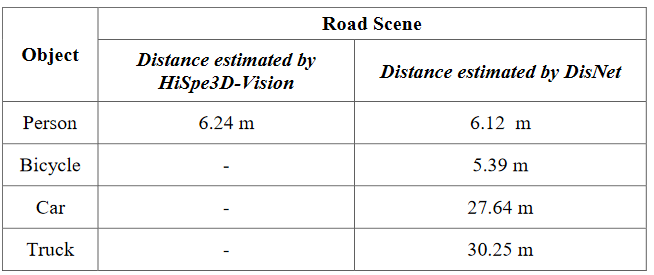


**Figure: Test on rail scene (*9a and 9b*)**

We can see with those first results that the Disnet is very accurate when the distance is less than 100m and there are just a few obstacles. Whereas in a situation with a lot of people we can see that the accuracy of the Disnet is really bad.

It is also due to the fact that the DisNet algorithm was not trained with real rail track scene which can lead to those inaccuracy.

It was then tested on road scene:



**Figure: Test on road scene**

The Disnet Algorithm gives actual result and similar result to HiSpe3D for the person but since we don’t have ground truth we can’t say for sure if DisNet works accurately.

**Conclusion :**

The DisNet algorithm recognized a lot more of obstacle but there is still a problem with the bounding boxes extraction

DisNet is really interesting because firstly it talks about the YOLO algorithm which is easy to implement and shows that it has a really good success rate, secondly the idea of using the bounding box dimension for distance estimation and having usable result can be useful for our project.

The training set used in this algorithm is really important because the results will vary a lot, for a usage in amateur video this can be a problem.

PRESENTATION: