



HAL
open science

Efficient binary segmentation through dense neural networks in a truncated frequency domain

Nils Defauw, Marielle Malfante, Olivier Antoni, Tiana Rakotovao, Suzanne Leseq

► **To cite this version:**

Nils Defauw, Marielle Malfante, Olivier Antoni, Tiana Rakotovao, Suzanne Leseq. Efficient binary segmentation through dense neural networks in a truncated frequency domain. EUSIPCO 2024 - 32nd European Signal Processing Conference, Aug 2024, Lyon, France. pp.566-570. cea-04688248

HAL Id: cea-04688248

<https://cea.hal.science/cea-04688248v1>

Submitted on 4 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Context

Binary segmentation applied to robotics applications

- Used in robotics perception : classification of pixels of an image (into vehicles, road, pedestrians, etc.) or points of a point cloud
- **Real-time** execution on **resources-constrained** embedded chips

State Of The Art

Typical Artificial Neural Networks (ANNs) for binary segmentation use **CNNs** directly inferred on images (U-Net [1]) or on a 2D projection of a LiDAR point cloud (VoxNet [2], RangeNet++ [3])

- **Computationally expensive** due to the size of the input tensor

Frequency transformation is used to enrich the input of ANNs, but also for image translation [4] and manipulation [5] in multimodal neural networks

- It is **not used to predict frequency domain output, nor to reduce the costs**

Objective

Binary segmentation under the following constraints

- **Reduction of computational costs** in order to envision an integration into embedded resource constrained hardware in the future
- **Fully configurable** method regarding the available **computational resources**

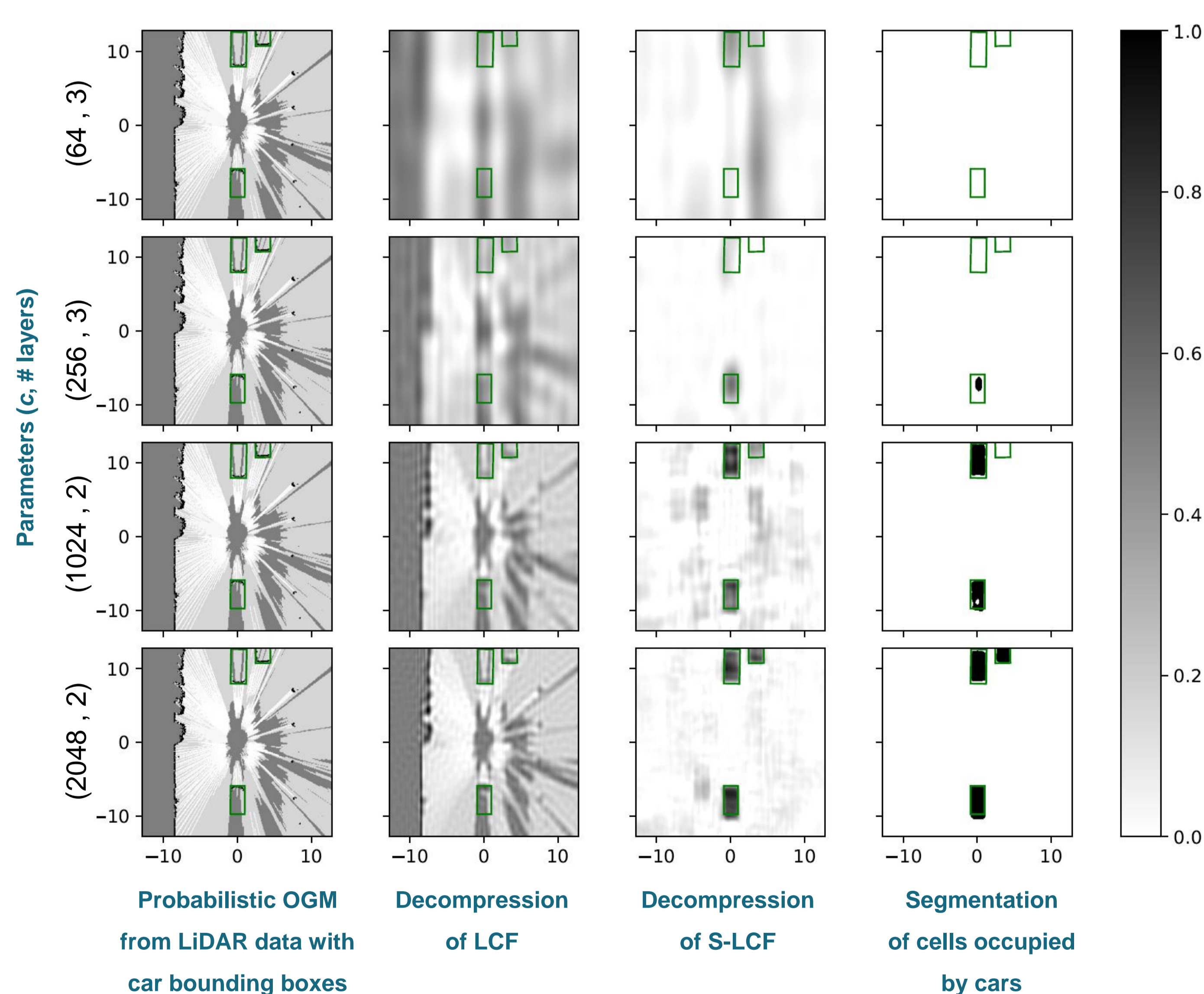
Use-case : Binary Segmentation of OGM

Probabilistic Occupancy Grid Map (OGM) [6]

- Representation of the environment around a car into a bird's eye view image composed of 65,536 squared cells of size 10 cm x 10 cm
- A probability value between 0 and 1 estimates the confidence about the occupancy state of each cell

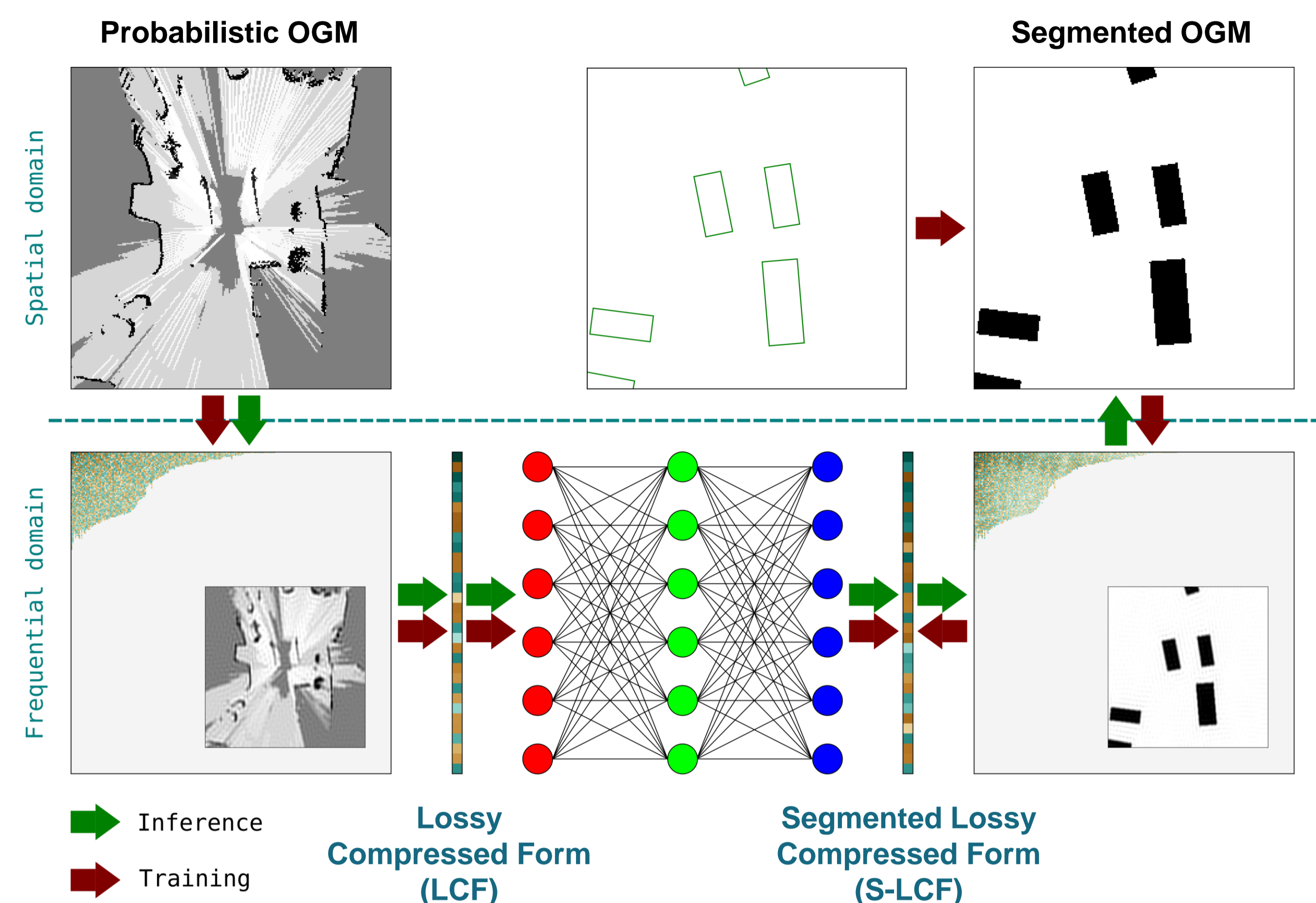
Objective : Segmentation of cells occupied by cars (binary segmentation)

Dataset : Original and segmented OGMs built from the bounding boxes of the labelled cars within the Waymo Open [7] dataset containing 198,068 traffic scenes



[1] Ronneberger et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation"
 [2] Maturana et al. "VoxNet: A 3D Convolutional Neural Network for real-time object recognition"
 [3] Milioto et al. "RangeNet++: Fast and Accurate LiDAR Semantic Segmentation"
 [4] Cai et al. "Frequency Domain Image Translation: More Photo-realistic, Better Identity-preserving"
 [5] Li et al. "Image Manipulation Localization Using Attentional Cross-Domain CNN Features"
 [6] Elfes "Using occupancy grids for mobile robot perception and navigation"
 [7] Sun et al. "Scalability in Perception for Autonomous Driving: Waymo Open Dataset"

Methodology



Step #1 - Compression into Frequency Domain

Input : Probabilistic Occupancy Grid Map (OGM)

1. Transformation of OGM into its spectrum obtained by the **Discrete Cosine Transform (DCT)**
2. **Truncation** of the obtained spectrum via an optimal mask (learnt based on the training set) in order to select and **keep only c number of frequencies**, where $c < \text{number of OGM cells}$
3. Flattening of the truncated spectrum into a 1D array

Output : **Lossy Compressed Form (LCF)** in the form of 1D array of size c containing the spectrum coefficients relative to the selected frequencies

Step #2 - Segmentation of Compressed Input

Input: Lossy Compressed Form (LCF)

1. Use **ANN to generate a Segmented Lossy Compressed Form (S-LCF)**
 - Architecture : 2 or 3 dense layers of c neurons, with ReLU activation function for all layers except for the last one
 - Training : Adam optimizer with early stopping
2. **Decompression** of S-LCF into spatial domain
3. Binary segmentation of the decompressed form by thresholding

Output: Binary OGM where cells occupied by cars are segmented

Experimental Results

Parameters		Quality	Computational costs	
c	# layers	AP	Time	# weights
64	2	0.272	0.76ms	8,320
64	3	0.280	0.77ms	12,480
128	2	0.337	0.84ms	33,024
128	3	0.405	0.91ms	49,536
256	2	0.456	0.83ms	131,584
256	3	0.484	0.88ms	197,376
512	2	0.531	0.85ms	525,312
512	3	0.530	0.92ms	787,968
1024	2	0.561	0.90ms	2,099,200
1024	3	0.536	1.04ms	3,148,800
2048	2	0.574	1.62ms	8,392,704
2048	3	0.527	2.16ms	12,589,056

- Quantitative results : **Inference times** < 2 ms for **Average Precision (AP)** > 0.5 , starting with compression parameter $c \geq 512$ (compression ratios ranging from 1/128 to 1/32)
- Qualitative results show an ability of the neural networks to **segment properly** OGMs even under very degraded compressed forms