Moving Object Segmentation for Underwater Videos: A Graph Neural Network Approach

Segmentation d'Objets Mobiles pour les Images Sous-Marines: Une Approche des Réseaux de Neurones en Graphes

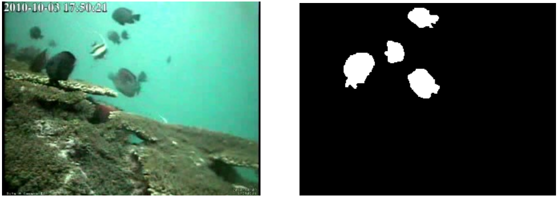
Wieke Prummel (PhD, 2nd year)
Co-supervisor: Anastasia Zakharova (MCF)
Supervisor: Thierry Bouwmans (MCF, HDR)

GAM - Gestion et Analyse des Données Maritimes

11th of May 2023 - Paris

La Rochelle UNIVERSITÉ
MATHEMATIQUES IMAGE APPLICATIONS
UNIVERSITÉ DE LA ROCHELLE
भारतीय प्रौद्योगिकी संस्थान जम्मू
INDIAN INSTITUTE OF TECHNOLOGY JAMMU
विद्याधनं सर्वधन प्रधानम्

- MOS is an important problem in computer vision, particularly in underwater surveillance system applications.
- The goal is to identify and separate the pixels or regions in a video that correspond to moving objects from the static background or other static objects.



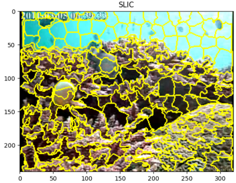
(a) Crowded Scene (b) Crowded Scene - Ground Truth

Figure – Fish4Knowledge Dataset

- Environmental issues (preservation of biodiversity, pollution at sea).
- Detecting motion in noisy and dynamic backgrounds is still an open research problem (Spampinato et al., 2014b ; Alshdaifat et al., 2020) due to various challenges (Spampinato et al., 2008).
- An underwater scene is affected by several challenges : light intensity (attenuated due to water salinity, and varies with changes in water viscosity), presence of a multimodal background (moving algae).
- In underwater videos MOS is generally more difficult than for other applications such as video surveillance.

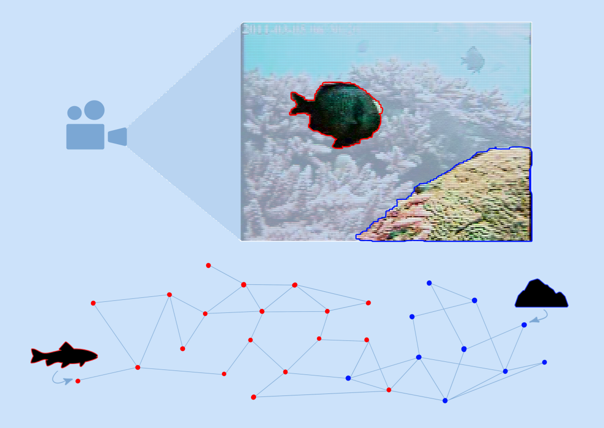
- Any system consisting of entities and relationships between them can be represented as a graph.
- For example, images are typically encoded as fixed-size 2-dimensional grids of pixels, and text as a 1-dimensional sequence of words. On the other hand, representing data in a graph-structured way reveals valuable information that emerges from a higher-dimensional representation of these entities and their relationships, and would otherwise be lost.

(a) Crowded Scene



(b) Crowded Scene - Super Pixel Segmentation

Figure – Fish4Knowledge image as fixed-sized 2-dimensional grids of pixels



- GraphMOS (*J. Giraldo et al., IEEE TPAMI, 2022*) and GraphMOD-Net(*J. Giraldo, et al., Frontiers of Computer Vision, 2021.*) use semi-supervised learning posed as a graph signal reconstruction problem.
- These methods are inspired by the theory of graph signal processing (*A. Ortega et al. Proceedings of the IEEE, 2018*) and have shown promising results.

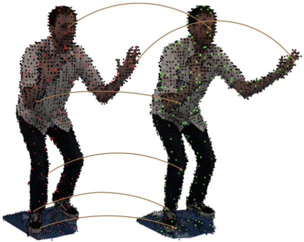


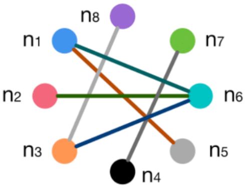
Figure – Example of motion estimation in a 3D point cloud sequence. Each frame is represented as a graph signal that captures the color and the geometry information of each voxel. (*A. Ortega et al. Proceedings of the IEEE, 2018*)

- Graph Neural Networks becomes the new trend in other domains too, and shows good performance for real-world deployment solutions.
- Uber, Google, Alibaba, Pinterest, Twitter and many others have shifted to GNN-based approaches.
- What are the actual advantages of Graph Machine Learning ?

Graph

- A Graph, G(V,E), is defined through a set of nodes V $= \{1, ..., \text{N}\}$ and a set of edges E $= \{(\text{i, j})\}$.
- $W \in \mathbb{R}^{NxN}$ is the adjacency matrix of G such that $W(i,j) = w_{i,j} \in \mathbb{R}^{+}$ is the weight connecting vertices i and j.
- W is symmetric for undirected graphs.
- Graph signal : y : V $\rightarrow \mathbb{R}$, $y \in \mathbb{R}^{N}$

Graph:



W:

	n1	n2	n3	n4	n5	n6	n7	n8
n1	0	0	0	0	1	1	0	0
n2	0	0	0	0	0	1	0	0
n3	0	0	0	0	0	1	0	1
n4	0	0	0	0	0	0	1	0
n5	1	0	0	0	0	0	0	0
n6	1	1	1	0	0	0	0	0
n7	0	0	0	1	0	0	0	0
n8	0	0	1	0	0	0	0	0

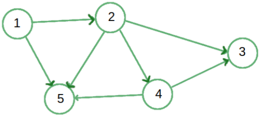
Figure – 'Hypergraph Neural Networks', Feng et al., AAAI [2019]

Definitions

Directed and Undirected Graph

Directed Graph

A directed graph has an ordered set of nodes.



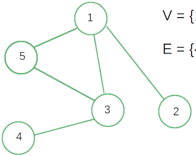
V = {1, 2, 3, 4, 5}

E = {(1, 2), (1, 5), (2, 3), (2 4), (2, 5), (4, 5), (4, 3)}

Figure – Directed graph

Undirected Graph

An undirected graph has an unordered set of nodes.



V = {1, 2, 3, 4, 5}

E = {{1, 2}, {1, 3}, {1, 5}, {3, 4}, {5, 3}}

Figure – Undirected graph

Unsupervised Learning

Uses no labeled training data.

Supervised Learning

Uses only labeled training data.

Semi-supervised Learning

Approach to Machine Learning combining a small set of labeled data and a large amount of unlabeled data during training.

GNN

- In GNNs, commonly a vector of features $\mathbf{x}_i \in \mathbb{R}^F$ is associated to each node i.
- The whole set of input features can be represented as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\mathsf{T} \in \mathbb{R}^{N \times F}$.
- The propagation rule of GCN is given by :

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}), \tag{1}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\tilde{\mathbf{D}}$ is the degree matrix of $\tilde{\mathbf{A}}$, $\mathbf{H}^{(l)}$ is the output matrix of layer l (with $\mathbf{H}^{(0)} = \mathbf{X}$), $\mathbf{W}^{(l)}$ is the matrix of trainable weights in layer l, and $\sigma(\cdot)$ is an activation function such as ReLU or softmax.

The most common deep models are mostly supervised and can be divided into four groups (*B. Hou et al., "A survey of efficient deep learning models for moving object segmentation," APSIPA, 2023*) :

1. 2D Convolutional Neural Networks (CNNs) (*M. Braham and M. Van Droogenbroeck, IEEE IWSSIP, 2016*),
2. 3D CNNs (*T. Akilan et al., IEEE TITS, 2019*),
3. Transformer neural networks (*I. Osmanet al., IEEE ICCV-W, 2021*),
4. Generative adversarial networks (GANs) (*M. Bakkay et al., IEEE ICIP, 2018*), (*M. Sultana et al., Neurocomputing, 2022*)

Some state-of-the-art (SOTA) techniques have been combined with deep methods to create novel approaches :

- MotionRec (*M. Mandal et al., IEEE WACV, 2020*),
- RT-SBS (*A. Cioppa et al., IEEE ICIP, 2020*),
- GraphMOS (*J. Giraldo et al., IEEE TPAMI, 2022*),
- GraphMOD-Net (*Giraldo, et al., Frontiers of Computer Vision, 2021.*).

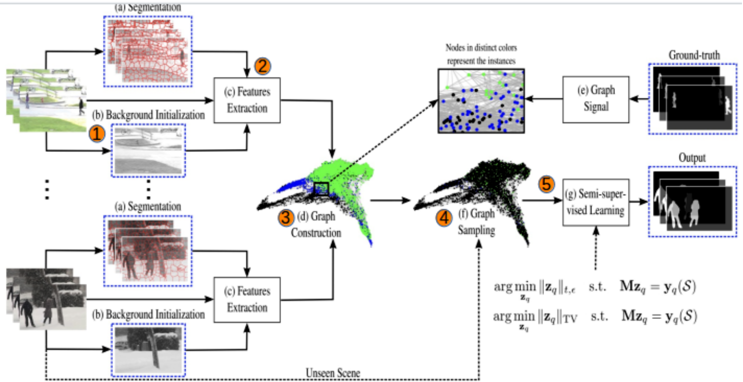


Figure – GraphMOS Architecture (*J. Giraldo et al., IEEE TPAMI, 2022*)

- While graph-based methods (ex. GraphMOS, *J. Giraldo et al., IEEE TPAMI, 2022*) have shown promising results in MOS, they have mainly relied on transductive learning which assumes access to the entire training and testing data for evaluation.
- This assumption is not realistic in real-world applications where the system needs to handle new data during deployment.

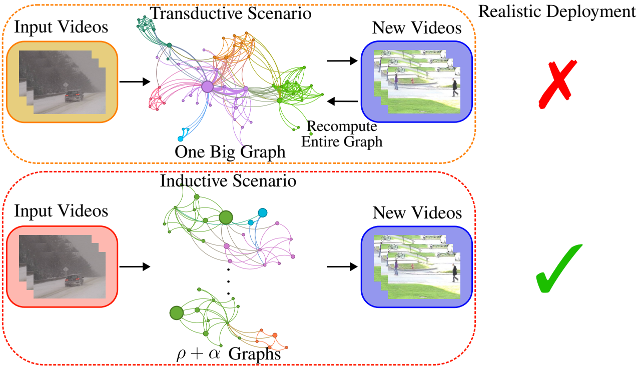


Figure – Transductive vs. Inductive Learning

W. Prummel, J. H. Giraldo, A. Zakharova and T. Bouwmans "Inductive Graph Neural Networks for Moving Object Segmentation".

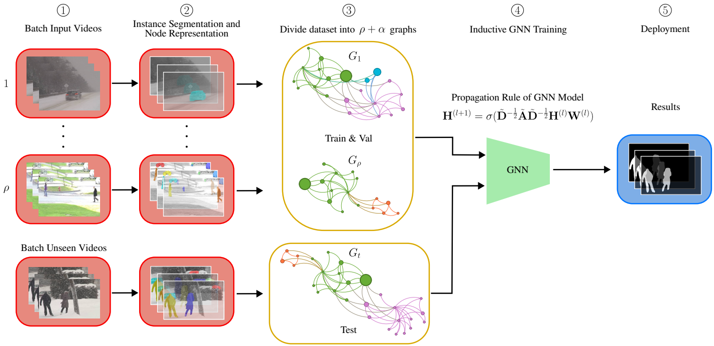


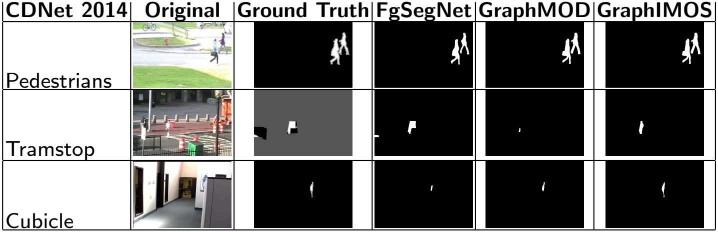
Figure – Pipeline : After instance segmentation and node feature extraction, the dataset is divided into graphs for training-validation and into testing graphs. The algorithm then classifies nodes in the graphs as either moving or static objects.

W. Prummel, J. H. Giraldo, A. Zakharova and T. Bouwmans "Inductive Graph Neural Networks for Moving Object Segmentation", submitted to IEEE ICIP 2023.s

Method	BSL	BWT	IOM	LFR	PTZ	THL	CJI	SHW	DBA	Overall
Transductive Learning Methods										
GraphMOS	0.9398	0.8294	0.3607	**0.5538**	0.7599	**0.7292**	0.7005	**0.9653**	0.7334	0.7302
GraphMOD-Net (Original)	**0.9550**	**0.8390**	**0.5540**	0.5210	**0.7700**	0.6820	**0.7200**	0.9420	**0.8510**	**0.7593**
Inductive Learning Methods										
FgSegNet	0.5641	0.2789	0.3325	0.2115	0.1400	0.3584	0.2815	0.3809	0.2067	0.3061
GraphMOD-Net (Modified)	0.6474	0.6268	0.5243	0.5337	0.5899	0.5484	0.4926	0.6587	**0.6254**	0.5831
GraphIMOS (Ours)	**0.7003**	**0.6377**	**0.5284**	**0.5478**	**0.5932**	**0.6453**	**0.6700**	**0.6807**	0.5868	**0.6211**

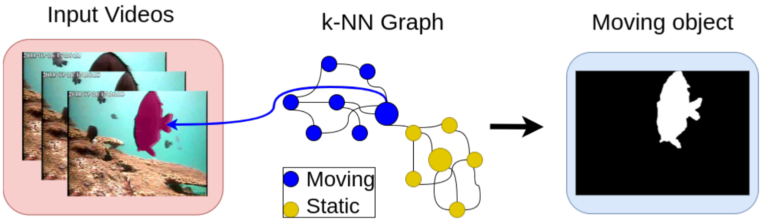
Average F-Measure for transductive and inductive methods. The best score of all methods appears in bold, and the best score of the inductive methods is bold underlined.

W. Prummel, J. H. Giraldo, A. Zakharova and T. Bouwmans "Inductive Graph Neural Networks for Moving Object Segmentation", submitted to IEEE ICIP 2023.



W. Prummel, J. H. Giraldo, A. Zakharova and T. Bouwmans "Inductive Graph Neural Networks for Moving Object Segmentation", submitted to IEEE ICIP 2023.

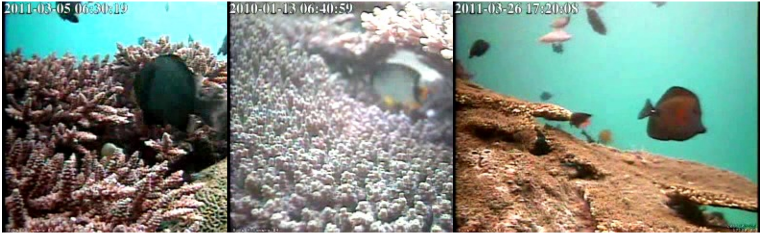
- To the best of our knowledge, GraphIMOS is the first approach that uses graph-based inductive learning for MOS, demonstrating its novelty and potential.
- Compared to previous works such as GraphMOD-Net, GraphIMOS offers improved performance and a better trade-off between performance and practical deployment.



- A new method for instance segmentation for underwater videos within the framework of the segmentation of mobile objects by graphs.
- A combination of feature learning and graphs to identify the main challenges of underwater MOS.

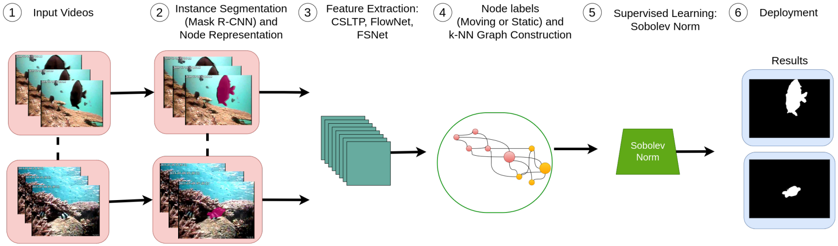
UNICT Underwater Background Modeling :

- 14 videos and 7 different classes representing complex challenges in the background modeling
- For each video : about 30 images annotated and provided as binary masks
- In total : more than 3500 labeled objects



GraphMOS-U

Architecture



Other Graph learning marine applications

- Side-scan sonars (SSS) provide high-resolution images and are used in underwater object detection and maritime search and rescue (SAR).
- Detect underwater maritime targets, such as shipwrecks,submerged containers, and their pollution (oil, sewage, chemicals), etc.
- Sector scanning sonars : moving object detection or segmentation.
- Acoustic cameras : moving object (ex. diver) detection or segmentation.

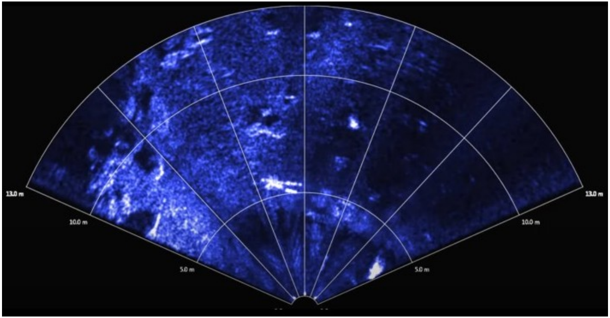


Figure – Acoustic 2D camera BlueView900

- **GraphMOS :** transductive graph learning algorithm. Performs well, but not adapted to real-world applications.
- **GraphIMOS :** novel inductive learning approach that uses GNNs for MOS. The proposed algorithm consists of four key components : in- stance segmentation using Mask R-CNN, feature extraction for node representation, k-NN for graph construction, and a GNN-based inductive learning algorithm.
- GraphIMOS is, to the best of our knowledge, the first approach that uses graph-based inductive learning for MOS. GraphIMOS offers improved performance and a better trade-off between performance and practical deployment.
- **GraphMOS-U :** First, we would like to validate GraphMOS on underwater videos and then if we have enough datasets, try inductive learning on underwater scenarios.

- **Book Chapter :** W. Prummel, A. Zakharova, T. Bouwmans, "Moving Objects Detection for Video Surveillance Applications in Society 5.0", Handbook on "Technological Prospects and Social Applications of Society 5.0" , Edited by L. Sharma and P. Garg, Taylor and Francis Group, 2023.
- **Conference Paper :** W.Prummel, J. Giraldo, A. Zakharova, T. Bouwmans, "Inductive Graph Neural Networks for Moving Object Segmentation", IEEE ICIP 2023 (**Submitted**).

B. Hou et al., “A survey of efficient deep learning models for moving object segmentation,” APSIPA, Transactions on Signal and Information Processing, vol. 12, no. 1, 2023.

M. Braham and M. Van Droogenbroeck, “Deep background subtraction with scene-specific convolutional neural networks,” in IEEE IWSSIP, 2016, pp. 1–4.

T. Akilan, Q. Wu, A. Safaei, J. Huo, and Y. Yang, “A 3D CNN-LSTM-based image-to-image foreground segmentation,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 3, pp. 959–971, 2019.

I. Osman, M. Abdelpakey, and M. Shehata, “Transblast : Self-supervised learning using augmented subspace with transformer for background/foreground separation,” in IEEE ICCV-W, 2021, pp. 215–224.

M. Bakkay, H. Rashwan, H. Salmane, L. Khoudour, D. Puig, and Y. Ruichek, “BSCGAN : Deep background subtraction with conditional generative adversarial networks,” in IEEE ICIP, 2018, pp. 4018–4022.

M. Sultana, A. Mahmood, T. Bouwmans, M. Khan, and S. Jung, "Moving objects segmentation using generative adversarial modeling," Neurocomputing, vol. 506, pp.240–251, 2022.

J. Giraldo, S. Javed, and T. Bouwmans, "Graph moving object segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 5, pp.2485–2503, 2022.

A. Ortega, P. Frossard, J. Kovaˇcevi c, J. MF Moura, and P. Vandergheynst, "Graph signal processing : Overview, challenges, and applications," Proceedings of the IEEE, vol. 106, no. 5, pp. 808–828, 2018.

M. Mandal, L. Kumar, and M. Saran, "MotionRec : A unified deep framework for moving object recognition," in IEEE WACV, 2020, pp. 2734–2743.

A. Cioppa, M. Van Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in IEEE ICIP, 2020, pp. 3214–3218.

L. A. Lim and H. Y. Keles, "Learning multi-scale features for foreground segmentation," Pattern Analysis and Applications, vol. 23, no. 3, pp. 1369–1380, 2020.

W. Prummel, J. H. Giraldo, A. Zakharova and T. Bouwmans "Inductive Graph Neural Networks for Moving Object Segmentation", submitted to IEEE ICIP 2023.

Thank You

Wieke Prummel, 2nd year of PhD
https ://sites.google.com/view/wiekeprummel/
wieke.prummel@univ-lr.fr