On Splitting Lightweight Semantic Image Segmentation for Wireless Communications

Ebrahim Abu-Helalah[*†], Jordi Serra[*], Jordi Perez-Romero[†],

[*]Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Barcelona, Spain

[†]Universitat Politecnica de Catalunya (UPC), Barcelona, Spain

{aebrahim, jserra}@cttc.es, {ebrahim.abu.helalah,jordi.perez-romero}@upc.edu

Abstract—Semantic communication represents a promising technique towards reducing communication costs, especially when dealing with image segmentation, but it still lacks a balance between computational efficiency and bandwidth requirements while maintaining high image segmentation accuracy, particularly in resource-limited environments and changing channel conditions. On the other hand, the more complex and larger semantic image segmentation models become, the more stressed the devices are when processing data. This paper proposes a novel approach to implementing semantic communication based on splitting the semantic image segmentation process between a resource constrained transmitter and the receiver. This allows saving bandwidth by reducing the transmitted data while maintaining the accuracy of the semantic image segmentation. Additionally, it reduces the computational requirements at the resource constrained transmitter compared to doing all the semantic image segmentation in the transmitter. The proposed approach is evaluated by means of simulation-based experiments in terms of different metrics such as computational resource usage, required bit rate and segmentation accuracy. The results when comparing the proposal with the full semantic image segmentation in the transmitter show that up to 72% of the bit rate was reduced in the transmission process. In addition, the computational load of the transmitter is reduced by more than 19%. This reflects the interest of this technique for its application in communication systems, particularly in the upcoming 6G systems.

Index Terms—semantic communication; semantic image segmentation; artificial intelligence; machine learning; 6G.

I. INTRODUCTION

The Ericsson Mobility Report from November 2024 [1] predicts that by 2030, there will be around 6.3 billion 5G mobile users worldwide. Moreover, there are substantial efforts in starting to shape the future 6G networks that will target challenging usage scenarios such as hyper-reliable and low-latency communication, immersive communication or ubiquitous connectivity [2]. It will also focus on artificial intelligence assistance in communications, encompassing standard applications such as automated driving, autonomous device collaboration for medical assistance, and distributing intensive computing tasks across devices and networks [2]. In light of this expected huge increase in the number of users and traffic demand and preparation for 6G network, it is necessary to develop communication models capable of transferring data faster and more efficiently.

Even though traditional communications have come a long way, they still have a hard time keeping up with the exponen-

tial growth in data traffic and the growing need to send large amounts of data, i.e., the bandwidth bottleneck. These systems transmit large amounts of raw data, which strains the channel capacity and leads to problems such as network congestion and increased latency [3]. This is a particular problem in real-time applications, such as live video streaming or autonomous systems, where even small delays can obstruct performance [4]. For instance, streaming a 4K video may require 30-50 Mbps [5], and virtual reality applications may require 25-100 Mbps for each user [6]. In this context, providing service to a large number of users leads to unprecedented demands in network infrastructure. Hence the need for a smarter approach that prioritizes the meaning of data over raw data. In this direction, semantic communication emerges as an effective model in meeting this objective [7].

In contrast to traditional communication, which focuses on transferring raw data without knowing their meaning or significance, i.e., pre-understanding transfer, semantic communication emphasizes the accurate transfer of meaning behind data, known as post-understanding transfer. The principles of semantic communication are widely applicable to systems that require intelligent and meaningful exchange of data [8].

Semantic communication addresses the issue of bandwidth limitations by concentrating on the transmission of only the essential semantic information, thereby lowering the necessary bit rate. However, this approach presents a new challenge, which is the limitation in computational resources. Transforming raw data into valuable insights, such as semantic segmentation, demands significant computing power [9]. Regardless of using lightweight models and optimization methods, the computational demands are still significant, particularly in environments with limited resources like when the transmitter is a mobile terminal [9]. The challenge of maintaining a balance between limited bandwidth and growing computational needs presents a significant obstacle to the broader implementation of semantic communication.

Among the applicability areas of semantic communication, semantic image segmentation involves creating a detailed map of an image, labeling each part with its corresponding meaning. Instead of sending the entire image, we can just send the labels for those meaningful parts, for example, "sky," "tree," "car," "person" [8]. This approach allows us to focus on the key information within the image, making communication

more efficient and easier to understand. Imagine describing a scene to someone instead of showing them a blurry image—semantic segmentation works similarly by providing a concise, meaningful representation of the visual world [10].

Semantic communication and semantic image segmentation, including the lightweight models, are important aspects in the development of intelligent systems based on post-understanding transmission. Despite the supposed interconnection between these two areas, previous research has often addressed them separately, leading to a research gap that prevents the desired integration. For example, papers such as Proportional-Integral-Derivative Network (PIDNet) [11] and Lightweight Context-Aware Network (LCNet) [12] have focused on developing lightweight semantic segmentation without paying attention to their integration with semantic transmission frameworks. Other papers, like [13], which looks at semantic transfer using polar codes for semantic-based communication systems, or [14], which looks at wireless semantic communication for image transmission, have focused on the framework of semantic transmission without considering lightweight semantic encoder methods. This research gap, in addition to the problems mentioned above, highlights the need for a study that combines both aspects to achieve a better integration between lightweight semantic image segmentation and semantic transfer frameworks.

Based on the above, the aim of this paper is to investigate the application of lightweight image segmentation models within semantic communication pipelines. Specifically, we will explore the trade-offs between computational resources at the transmitter, communication resources, and application accuracy. To achieve this, the main contribution of this paper is the proposal of splitting a lightweight semantic encoder (distributing its components between the sender and the receiver) and analyzing its impact on both communication and computational resource consumption.

Specifically, the novelty of the paper is the application of PIDNet in semantic communication by splitting the network after the fifth stage. This approach allows the sender to transmit low-resolution, semantically rich features, while the receiver completes the semantic inference process using the final layers of the model. Unlike traditional partitioning or edge inference methods, our design explicitly considers the balance between feature size, which reduces the communication resources used, and semantic accuracy across the communication channel.

The rest of the paper is organized as follows: Section II presents a comprehensive description of the considered problem. Section III presents the proposed solution that is experimentally evaluated in Section IV. Finlay, Section V concludes this work.

II. PROBLEM STATEMENT

The problem addressed by this paper is to achieve the balance between accuracy, bit rate, and computational requirements of semantic image segmentation within the framework of semantic communication. Semantic image segmentation is a fundamental task in the field of computer vision that involves classifying each pixel in an image into a predefined class. Unlike image classification, which assigns a single label to the entire image, semantic segmentation provides a detailed understanding of a scene by distinguishing objects and their boundaries at the pixel level. To ensure high accuracy in image semantic segmentation, convolutional neural networks (CNNs) have been widely used. Where CNNs extract hierarchical spatial features from input images through layers of convolutions, nonlinear activations, and pooling operations [10]. However, these techniques also introduce different challenges, the most notable of which is dealing with objects of different sizes within the image, because CNN networks have a fixed receptive field, which limits their ability to capture multi-scale contextual information. The Pyramid Scene Analysis Network (PSPNet) introduces the Pyramid Pooling Module (PPM) to address this challenge [15]. The purpose of PPM is to combine contextual data at different levels by using global pooling. This allows to make several feature maps with varying levels of accuracy. This technique improves segmentation performance, especially in complex scenes containing objects of different sizes. Even though PPM has benefits, it comes with extra computing costs. This is because multi-scale pooling operations need more memory and processing power, which can be expensive for real-time applications, especially when working with high-resolution images.

Moreover, implementing this technology in a resource-limited environment poses significant challenges due to the computational requirements needed for extracting this meaning, e.g. through semantic segmentation. Devices like edge sensors, IoT nodes, and mobile platforms face strict limitations, such as low-performance processors, limited memory, and restricted power supply [16]. For instance, a smart home sensor designed for real-time anomaly detection needs to process data locally while conserving battery life, which makes it difficult to use resource-heavy semantic algorithms. Similarly, drones that rely on image semantic segmentation for navigation must find a balance between lightweight processing models and the accuracy required to navigate. These examples underscore the difficulties of applying semantic communication systems in environments with limited computational resources.

The requirement to find an optimal compromise between the computational load and the complexity of CNNs has led to the design of new methods of semantic image segmentation. So, researchers have been developing lightweight neural networks (e.g., MobileNet [17], Tiny YOLO [18], and PIDNet) to perform the AI tasks within resource-constrained devices with minimum processing cost, which could be implemented on edge devices. But these strategies may still suffer in the resource-limited environment.

To achieve the balance mentioned earlier, we have proposed an innovative solution based on semantic encoder splitting between the transmitter and the receiver which will be discussed

in the next section.

III. A Splitting-Based Approach To Semantic Communication

The proposed solution provides a scalable framework for diverse communication environments that consists in splitting lightweight semantic image segmentation processing models between the transmitter, which is assumed to be a constrained resource device such as mobile, and the receiver, which is assumed to be a more powerful equipment such as an edge server. Fig. 1 shows the proposed method for semantic image segmentation. It has two parts, namely the "semantic encoder P1," at the transmitter side that executes part of the semantic image segmentation process, and the "semantic encoder P2," at the receiver that executes the rest of the process. Moreover, at the transmitter the channel coding together with the rest of radio protocol stack are applied to the output of the "semantic encoder P1" prior to delivering the resulting information through the channel. Similarly, at the receiver side, the received information is processed at the radio protocol stack including the channel decoding before delivering it to the "semantic encoder P2".
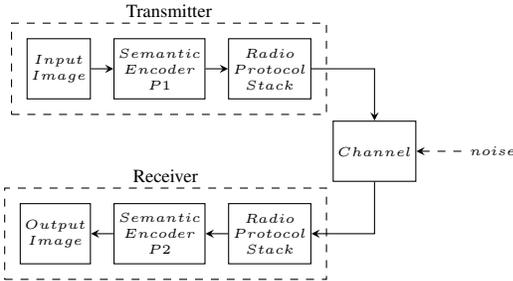

Fig. 1: Splitting approach for semantic encoder.

Among the existing lightweight semantic image segmentation models this paper focuses on the proportional-integral-derivative network PIDNet [11]. This model was chosen because it is an efficient model designed for real-time semantic image segmentation, making it well suited for this study as its balanced architecture ensures high accuracy while maintaining a low computational cost. PIDNet semantic encoder consists of six stages of semantic extraction, which allows us to split it easily.

Table I shows the stages of the PIDNet architecture. PIDNet processes the image in several steps, each one improving the representation of features while decreasing the image size. Assuming an input image with 1024 x 1024-pixel size, stage 0 applies convolutional layers to extract initial features, downscaling the image to 256 x 256 pixels and preparing the input for deeper feature extraction in subsequent stages. Stage 1 applies residual blocks (RB), maintaining the resolution at 256×256 while enhancing feature extraction. It uses skip connections to keep important features and stop gradients vanishing. Additional RBs are used in stage 2, reducing the resolution to 128×128 pixels and enabling more abstract feature learning while maintaining essential spatial information.

Stage 3 introduces a three-branch structure to process features (patterns, textures, shapes, and object details extracted from the image) in parallel. The first branch extracts local features, the second downsamples and upsamples to capture contextual information, and the third refines the representations. This stage reduces the resolution to 64×64 and merges multi-scale features for richer information. The three-branch structure continues in stage 4 processing features more deeply while reducing the resolution to 32×32. The combination of extracted local and contextual information strengthens the representation of complex patterns in the image. Stage 5 employs residual bottleneck blocks (RBB), smaller version of residual blocks that make computations faster by both shrinking and expanding the sizes of features, reducing computational costs while preserving important details with the continuation of the three-branched structure. The output of this stage is a 16×16-pixel feature map, maintaining segmentation-relevant information at a coarse scale. Last but not least, stage 6 adds a PPM to capture multi-scale contextual information and creates the final 128×128-pixel feature map, which is then run through convolutional layers to produce the final segmentation output.

TABLE I: Stages of the PIDNet for semantic segmentation. [11]

Stage	Operation			Output
0	$2 \times Conv$ layers			256×256
1	RB			256×256
2	RB			128×128
3	RB	RB	RB	64×64
4	RB	RB	RB	32×32
5	RBB	RBB	RBB	16×16
6	PMM $2 \times Conv$ layers			128×128

According to the above, we split the PIDNet after stage 5 to make sure the encoder creates small, low-resolution semantic features with rich contextual information. This splitting reduces transmission overhead while preserving task-relevant semantics for accurate reconstruction at the receiver. The splitting process after stage 5 is also easier compared to the splitting at previous stages, in which the presence of three branches complicates the separation process. Unlike typical model-splitting methods that focus on saving computing power or memory, our split is based on a communication-specific balance: reducing the bit rate while keeping important information intact, even when the channel conditions change.

Consequently, we can summarize the advantages and expected results of the splitting as follows:

- Reducing the bit rate: By transmitting images with a smaller size as shown in Table I. The size of the transmitted image will be 16x16 instead of 128x128. This way, we can significantly reduce the amount of data transmitted over the network, which leads to saving bandwidth.
- Reducing GPU Computing & Memory Usage: By moving the highly computational demanding PPM process of the

6th stage and the additional two convolutional layers to the receiver, the computing demands for the resource constrained transmitter can be substantially reduced. Hence, this approach can be used on low-performance devices to perform the transmission process.

IV. Experiments and Results

To evaluate the performance of our proposed solution, we conducted a series of experiments using a specially designed simulator that we programmed for this purpose. This section shows the experimental setup, describes the methodology, and presents the results, highlighting the advantages of our approach.

The experiments were done with a Python-based simulator. The images that are segmented have been obtained from the Cityscape dataset, which contains a variety of stereoscopic video sequences recorded in street scenes from 50 different cities [19]. A total of 500 images are sent in each experiment. The radio transmission is modelled considering two possible modulation schemes, namely Quadrature phase shift keying (QPSK) and 16 quadrature amplitude modulation (16QAM). As such, the bandwidth of the transmitted signal depends on the amount of information to send based on the segmentation process and on the used modulation. An Additive White Gaussian noise (AWGN) channel is assumed in the link between transmitter and receiver. The effect of the AWGN channel is modelled by means of the signal-to noise ratio (SNR).

The proposed solution is benchmarked against two reference schemes. Thus, three experiments have been executed. The first one consists in the traditional communications scheme, where we send an image without any processing through the channel and then apply the complete lightweight semantic image segmentation process at the receiver. The second experiment consists in executing the full semantic segmentation at the transmitter and the result is sent through the channel. The third experiment represents the proposed solution that splits the lightweight semantic segmentation model between the sender and the receiver as we explained in Section III. In all experiments, we send the data twice, once using QPSK and the second time using 16QAM.

The metrics to assess the performance of the experiments are carefully selected to capture critical aspects of semantic image segmentation, communication resources, and computational resources. They are the following ones:

- The mean of Intersection over Union (mIoU): which is the average of the IoU across all categories, i.e. different object classes or labels in the dataset (e.g., road, car, pedestrian, building, sky). The IoU provides a measure of image segmentation accuracy by comparing the predicted and ground truth regions (i.e. the actual, manually labeled segmentation categories) [20] and calculated as and is given:

$$IoU = \frac{TP}{TP + FP + FN} \qquad (1)$$

Where:

- TP is the true positive results. Refers to pixels that are correctly predicted as belonging to a specific category.
- FP is the false positive results. Refers to pixels that are incorrectly predicted as belonging to a category when they do not.
- FN is the false negative results. Refers to pixels that belong to a category but are either unclassified or misclassified by the model.

- Required bit rate in (Mbps): This measures the average bit rate required to be sent through the channel depending on the amount of information sent as a result of the semantic segmentation processes executed at the transmitter.
- GPU Processing Usage: This is the percentage of GPU processing power used by the operations executed at the transmitter.
- GPU Memory Usage: Calculates the percentage of GPU memory (VRAM) used by the operations in the transmitter for tasks such as storing models, results, or input data.

Fig. 2 shows the bit rate required to transmit data in the three experiments. It is observed that the bit rate with the proposed approach is reduced by 91% compared to traditional communication and by 72.6% for the full semantic segmentation before transmission.
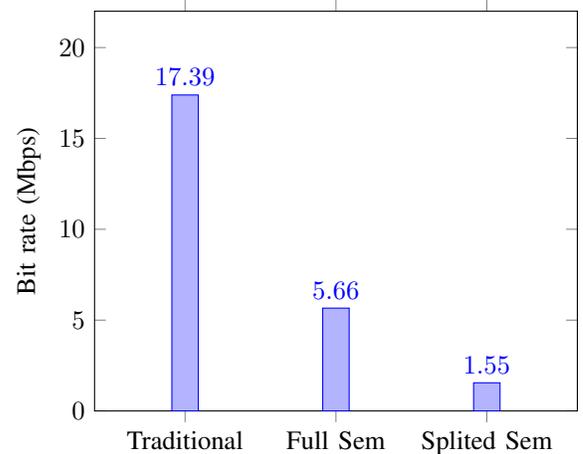

Fig. 2: Bit rate in Mbps for different experiments.

Fig. 3 shows the GPU processing utilization and GPU memory utilization in percentage. As shown, we were also able to cut down on the use of computing resources. For the full semantic segmentation before transmission, in the transmitter, the GPU processing usage dropped by 19.8% and the GPU memory utilization dropped by 3.4%.
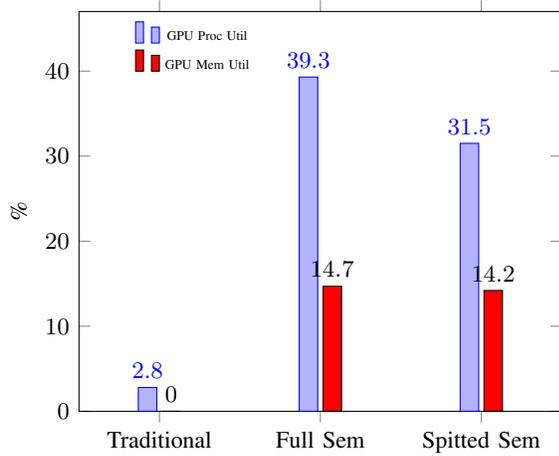
Fig. 3: GPU processing utilization and GPU memory utilization usage in percentage.

Fig. 4 shows the mIoU for different executions with the two modulation schemes and for SNR values varying in the range 5-30 dB. This lets us see how the splitting process affected the accuracy of semantic segmentation. The proposed approach is able to achieve the same mIoU than the full semantic segmentation at the transmitter with 1-1.5 dB less SNR. We anticipated this outcome because of the decrease in the data rate. We also observe that under good SNR conditions the accuracy is similar to the rest of the experiments.
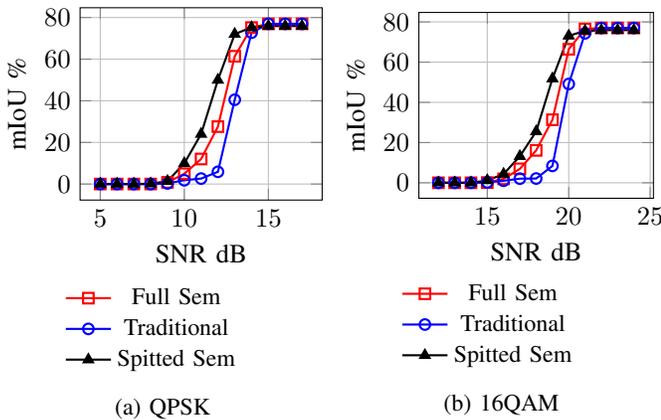

(a) QPSK

(b) 16QAM

Fig. 4: SNR vs mIoU for (a) QPSK and (b) 16QAM.

V. CONCLUSION

This paper has proposed the use of splitting in lightweight semantic image segmentation. This enables the model to operate on resource-constrained devices like mobiles, drones, etc. The approach intends to find a balance between communication resources and computation resources while maintaining the accuracy of image segmentation. The proposed approach has been experimentally evaluated to highlight its effectiveness. The results indicate that the required bit rate decreases by over 90% compared to executing image segmentation entirely at the receiver and by 70% compared to executing it entirely at the transmitter. Furthermore, the proposed approach managed to reduce the computation required for semantic image segmentation without compromising accuracy. Specifically, it reduced the GPU processing usage by more than 19%. The promising results obtained reflect the potential of our solution to maintain segmentation accuracy while distributing the computing load across multiple devices. Although the results are promising, the simulator does not fully take into account all channel impairments, including dispersion and fast fading. Future work will focus on expanding the experiments to include the possibility of making the model lighter and more compact and applying it to 6G communication scenarios.

REFERENCES

[1] Ericsson, "Ericsson mobility report 2024," Ericsson AB, Tech. Rep., November 2024. [Online]. Available: https://www.ericsson.com/4adb7e/assets/local/reports-papers/mobility-report/documents/2024/ericsson-mobility-report-november-2024.pdf

[2] ITU-R, "Recommendation itu-r m.2160-0: Framework and overall objectives of the future development of imt for 2030 and beyond," International Telecommunication Union, Tech. Rep., 2023. [Online]. Available: https://tinyurl.com/yc2bx9py

[3] A. Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005. [Online]. Available: https://doi.org/10.1017/CBO9780511841224

[4] F. Liu, G. Tang, Y. Li, Z. Cai, X. Zhang, and T. Zhou, "A survey on edge computing systems and tools," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1537–1562, 2019.

[5] A. Lustica, M. Jurcevic, and J. Bozek, "Live uhd/4k video transmission over 5g network," in *2021 International Symposium ELMAR*, 2021, pp. 17–20.

[6] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "Vr is on the edge: How to deliver 360° videos in mobile networks." Association for Computing Machinery (ACM), 2017. [Online]. Available: https://doi.org/10.1145/3097895.3097901

[7] Y. Shen, Z. Bao, L. Li, C. Li, H. Zhang, B. Liu, and C. Dong, "Digital wireless semantic communication enabled by the discrete semantic knowledge base and updating," in *2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2024, pp. 651–656.

[8] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–219, 2022.

[9] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," 2022. [Online]. Available: https://arxiv.org/abs/2201.01389

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440.

[11] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," 2023. [Online]. Available: https://arxiv.org/abs/2206.02066

[12] M. Shi, S. Lin, Q. Yi, J. Weng, A. Luo, and Y. Zhou, "Lightweight context-aware network using partial-channel transformation for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 7401–7416, 2024.

[13] H. Rezaei, T. Sivalingam, and N. Rajatheva, "Automatic and flexible transmission of semantic map images using polar codes for end-to-end semantic-based communication systems," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023, pp. 1–6.

[14] M. U. Lokumarambage, V. S. S. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, vol. 11, pp. 37 149–37 163, 2023.

[15] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," 2017. [Online]. Available: https://arxiv.org/abs/1612.01105

[16] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Will 6g be semantic communications? opportunities and challenges from task oriented and secure communications to integrated sensing," *IEEE Network*, vol. 38, no. 6, pp. 72–80, 2024.

[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[18] D. Padilla Carrasco, H. A. Rashwan, M. A. Garcisa, and D. Puig, "T-yolo: Tiny vehicle detection based on yolo and multi-scale convolutional neural networks," *IEEE Access*, vol. 11, pp. 22 430–22 440, 2023.

[19] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," 2016. [Online]. Available: https://arxiv.org/abs/1604.01685

[20] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," 2019. [Online]. Available: https://arxiv.org/abs/1902.09630