https://doi.org/10.31891/csit-2024-4-11
UDC 004.4

Markijan DURKOT, Nataliia MELNYK
Lviv Polytechnic National University

COMPARATIVE ANALYSIS OF REAL-TIME SEMANTIC SEGMENTATION ALGORITHMS

Semantic segmentation is a fundamental task in computer vision that enables machines to interpret and understand images at the pixel level, providing a deeper understanding of scene composition. By assigning a class to each pixel, this technique is critical for applications requiring detailed visual comprehension, such as autonomous driving, robotics, medical imaging, and augmented reality. This article presents a comprehensive comparative analysis of deep learning models specifically designed for real-time semantic segmentation, focusing on their performance metrics, architectures, and various application contexts. This study compares advanced deep learning models, including PIDNet, PP-LiteSeg, BiSeNet, SFNet, and others, using key metrics such as Mean Intersection over Union (mIoU) and Frames Per Second (FPS), alongside the hardware specifications on which they were tested. Models like PIDNet, known for its multi-branch architecture, emphasize detailed, context, and boundary information to improve segmentation precision without sacrificing speed. On the other hand models like PP-LiteSeg, with its Short-Term Dense Concatenate Network (STDCNet) backbone, excels in reducing computational complexity while maintaining competitive accuracy and inference speed, making it well-suited for resource-constrained environments. The analysis evaluates the trade-offs between accuracy and computational efficiency using benchmark datasets such as Cityscapes and DeepScene. Additionally, we examine the adaptability of these models to diverse operational scenarios, particularly on edge devices like NVIDIA Jetson Nano, where computational resources are limited. This discussion extends to the challenges faced in real-time implementations, including maintaining robustness across varying environments and achieving high performance with minimal latency. Highlighting the strengths, limitations, and practical implications of these models, this analysis can serve as a valuable resource for researchers and practitioners aiming to advance the field of real-time semantic segmentation.

Keywords: semantic segmentation, real-time image processing, neural networks, machine learning, deep learning

Маркіян ДУРКОТ, Наталія МЕЛЬНИК
Національний університет «Львівська політехніка»

ПОРІВНЯЛЬНИЙ АНАЛІЗ АЛГОРИТМІВ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ В РЕАЛЬНОМУ ЧАСІ

Семантична сегментація є фундаментальним завданням комп'ютерного зору, яке дозволяє машинам інтерпретувати та розуміти зображення на рівні пікселів, забезпечуючи глибше розуміння складу сцени. Призначаючи кожному пікселю клас, ця технологія є критично важливою для застосувань, що потребують детального візуального сприйняття, таких як автономне водіння, робототехніка, медична візуалізація та доповнена реальність. Ця стаття пропонує всебічний порівняльний аналіз моделей глибокого навчання, спеціально розроблених для семантичної сегментації в реальному часі, з акцентом на їх показниках продуктивності, архітектурі та різних контекстах застосування. Дослідження порівнює сучасні моделі глибокого навчання, включаючи PIDNet, PP-LiteSeg, BiSeNet, SFNet та інші, використовуючи ключові метрики, такі як середнє перетинання над об'єднанням (mIoU) та кількість кадрів за секунду (FPS), а також апаратні характеристики, на яких їх тестували. Моделі, такі як PIDNet, відомі своєю багатогілковою архітектурою, акцентують увагу на деталях, контексті та межах для підвищення точності сегментації без шкоди для швидкості. З іншого боку, моделі на кшталт PP-LiteSeg з основою Short-Term Dense Concatenate Network (STDCNet) відзначаються зниженням обчислювальної складності при збереженні конкурентоспроможної точності та швидкості роботи, що робить їх ідеальними для середовищ із обмеженими ресурсами. Проведений аналіз оцінює компроміси між точністю та обчислювальною ефективністю, використовуючи еталонні набори даних, такі як Cityscapes і DeepScene. Додатково ми досліджуємо адаптивність цих моделей до різних операційних сценаріїв, зокрема на пристроях із низьким енергоспоживанням, таких як NVIDIA Jetson Nano, де обчислювальні ресурси обмежені. Ця дискусія також охоплює виклики, з якими стикаються в реальних умовах, включаючи підтримання надійності в різних середовищах і досягнення високої продуктивності з мінімальною затримкою. У цій роботі підкреслено сильні та слабкі сторони, а також практичні аспекти розглянутих моделей. Проведений аналіз може бути корисним для дослідників і практиків, у сфері семантичної сегментації в реальному часі.

Ключові слова: семантична сегментація, обробка зображень в реальному часі, нейронні мережі, машинне навчання, глибоке навчання.

Introduction

Semantic segmentation is a fundamental task in image processing that classifies each pixel in an image into predefined categories, allowing machines to gain a detailed understanding of the visual content. Unlike traditional image classification, which assigns a single label to an entire image, semantic segmentation provides a detailed view and is therefore essential for applications such as autonomous driving, medical imaging and robotics [1-3].

The evolution of semantic segmentation techniques has seen a significant shift from traditional methods to deep learning approaches, particularly with the advent of convolutional neural networks (CNNs). Early algorithms relied on handcrafted features and simple classifiers, which limited their performance and adaptability to complex visual environments [4]. The introduction of Fully Convolutional Networks (FCNs) marked a turning point, enabling end-to-end training and achieving state-of-the-art results [5]. Following this, architectures such as U-Net [6] and DeepLab [7] further improved segmentation accuracy by incorporating advanced techniques like skip connections and atrous convolution.

Despite these advancements, many existing models are computationally intensive, posing challenges for real-time applications. As the demand for faster and more efficient algorithms grows, researchers have focused on developing lightweight architectures that maintain high accuracy while operating within the constraints of limited hardware resources. Models such as BiSeNet [8] and PP-LiteSeg [9] exemplify this trend, optimizing performance for real-time inference on edge devices.

Real-time semantic segmentation plays a vital role in various real-world applications by efficiently analyzing visual data to identify and classify different elements within a scene. Here are some prominent applications:

- **Autonomous Driving.** Real-time semantic segmentation helps self-driving cars understand their surroundings by identifying objects like vehicles, pedestrians, road signs, and lane markings. By processing visual data quickly, the system can make real-time decisions, enhancing both safety and navigation precision in dynamic environments [10].
- **Robotics.** Robots, especially those used for navigation and object manipulation, benefit from semantic segmentation to interact safely and accurately with their environment. In warehouses, for example, robots use segmentation to differentiate between items, shelves, and paths, enabling efficient movement and task execution [11].
- **AR (Augmented Reality).** Real-time segmentation enhances AR applications by recognizing and isolating objects and surfaces, allowing virtual objects to blend seamlessly into the real world. This technology is essential for apps that overlay graphics on specific objects, such as virtual furniture placements in home design tools [12].
- **Medical Imaging.** In healthcare, semantic segmentation aids in identifying and highlighting specific structures in medical scans (e.g., tumors in MRI scans). The ability to analyze images quickly is especially beneficial in procedures requiring immediate insights, such as during surgeries or emergency diagnostics [13].
- **Remote Sensing and Environmental Monitoring.** Real-time semantic segmentation is useful in satellite and drone imagery analysis for tasks like crop monitoring, forest fire tracking, and land use changes. It enables quicker response and management decisions in environmental conservation and disaster management efforts [14].

Real-time semantic segmentation is advancing through models that balance high accuracy with processing efficiency, ensuring these applications meet the demands of their specific operational environments.

This article provides a comparative analysis of algorithms for real-time semantic segmentation, examining their architectures, performance metrics, and suitability for various applications. By comparing traditional and modern approaches, we aim to highlight the strengths and weaknesses of each method, offering insights into their practical implications and future research directions.

Problem statement

Real-time semantic segmentation is a critical challenge in the emerging field of autonomous vehicles, requiring precise pixel-level classification of scenes under stringent time constraints. The task demands high accuracy to identify and differentiate objects while ensuring low latency for real-time decision making. This problem is compounded by diverse environmental conditions, such as varying lighting, weather, and occlusions, as well as the image resolution and computational limitations of embedded hardware. Achieving an optimal balance between speed, accuracy, and resource efficiency is essential for ensuring safe and reliable decision making on computational limited embedded hardware.

Through out the years one of the most popular choices in lightweight semantic segmentation models suitable for real-time semantic segmentation was ResNet [15]. It achieves state-of-art results in semantic segmentation across various datasets and has many variations of it's architecture making it suitable for multiple application contexts. However, its performance on computational limited embedded devices suffers on low inference speed. The data displayed in Table 1, provided by NVIDIA Jetson [16], describes the performance of modern state-of-art semantic segmentation model ResNet-18 on NVIDIA's Jetson Nano and Jetson Xavier GPUs.

Table 1

Comparison of the performance of ResNet-18 on NVIDIA Jetson Nano and Jetson Xavier GPUs

№	Resolution	Dataset	mIoU	Jetson Nano	Jetson Xavier
1	512x256	Cityscapes	83.3%	**48 FPS**	**480 FPS**
2	1024x512	Cityscapes	87.3%	**12 FPS**	**175 FPS**
3	2048x1024	Cityscapes	89.6%	**3 FPS**	**47 FPS**
4	576x320	DeepScene	96.4%	**26 FPS**	**360 FPS**
5	864x480	DeepScene	96.9%	**14 FPS**	**190 FPS**

Analyzing the results, significant jumps in performance are observed, depending on number of classes in the dataset, data complexity, and image resolution. The comparison results indicate that inference speed is highly correlated with image resolution. The higher the image resolution – the lower the inferences speed. For a task like autonomous navigation for aerial vehicles a high resolution of input images is required. The intensity of details in the landscape can variate between flight areas and depends on the altitude. On high altitudes low resolution images can cause a lack of context for decision making models.

Literature Overview

In [17] the authors propose an innovative approach to solve semantic segmentation task. The biggest novelty seen in PIDNet family of models is the integration of the PID controller principals into the architecture of the model. PIDNet architecture makes a connection between Convolutional Neural Networks (CNN) and Proportional-Integral-Derivative (PID) controllers and reveal that a two-branch network is equivalent to a Proportional-Integral (PI) controller, which inherently suffers from overshoot issues. To alleviate this problem, the authors propose a novel three-branch network architecture, which contains three branches to parse detailed, context and boundary information, respectively, and employs boundary attention to guide the fusion of detailed and context branches.

The researches addressed to solve the task of scene flow estimation in computer vision provided SFNet (Scene Flow Network) [18] – a deep learning model initially proposed for scene flow estimation — the task of estimating 3D motion of points in a scene, which requires understanding both the motion and the geometry of the scene. Scene flow is an extension of optical flow to 3D, combining both motion and depth information. In the context of SFNet-R18, SFNet is tailored to work on segmentation tasks as well, such as semantic segmentation or instance segmentation, while still utilizing principles of scene flow and depth estimation. At the backbone of SFNet-R18 lies the ResNet-18 model.

The authors in [19] propose a highly innovative architecture that targets memory traffic reduction in deep learning networks – HarDNet, making it an excellent choice for low-resource environments such as mobile devices, embedded systems, and real-time applications. By focusing on memory-efficient designs like depthwise convolutions, HarD activation transformations, and pruned filters, HarDNet strikes a good balance between accuracy and memory efficiency, making it highly suitable for practical deployment in edge computing and autonomous systems. While the model may not be the absolute best choice for all high-performance use cases, its ability to reduce memory traffic without significant accuracy loss is a notable contribution to the field of resource-constrained deep learning systems.

Authors in [8] present a bilateral network design – BiSeNet, combining a context path for global information and a detail path for fine-grained features, allows it to deliver competitive segmentation performance without sacrificing speed. BiSeNet uses relatively simple backbones (ResNet-18 or VGG-16 [20]) compared to more complex networks (e.g., ResNet-50). While this helps with speed, it may limit the performance in extremely challenging segmentation tasks, where deeper architectures might perform better. Although BiSeNet performs well on typical benchmarks, its performance on very large-scale scenes or high-resolution input images could be a concern. The use of dilated convolutions in the context path helps to capture larger areas, but larger image sizes could still pose challenges in terms of memory consumption.

In [21] authors propose a module called STDC, or Short-Term Dense Concatenate, designed for semantic segmentation that efficiently extracts deep features with a scalable receptive field and multi-scale information. Its primary goal is to eliminate structural redundancy in the BiSeNet architecture. In particular, BiSeNet introduces an additional path to capture spatial information, which can be computationally expensive. In contrast, STDC progressively reduces the dimensionality of the feature maps and aggregates them for image representation. This method involves concatenating response maps from several consecutive layers, each of which encodes the input image or features at different scales and receptive fields, resulting in a rich multi-scale feature representation. To improve speed, the filter size in each layer is gradually reduced, with minimal impact on segmentation performance.

BiSeNet V2 [22] is a significant improvement over the original BiSeNet, introducing Guided Aggregation to effectively fuse multi-scale features and enhance both speed and accuracy for real-time semantic segmentation. BiSeNet Guided Aggregation is a technique to better integrate multi-scale features and improve context information. Enhancing the spatial and context streams to improve both accuracy and speed. Leveraging a lightweight design that maintains real-time performance while significantly improving segmentation quality.

PP-LiteSeg [9] is a highly efficient and lightweight real-time segmentation model that offers an optimal balance between accuracy and inference speed. While real-time segmentation models like BiSeNet, have demonstrated success in balancing performance and efficiency, PP-LiteSeg aims to improve upon these existing models by reducing the computational complexity while enhancing segmentation accuracy. By incorporating Flexible and Lightweight Decoder (FLD), Unified Attention Fusion Module (UAFM), Simple Pyramid Pooling Module (SPPM), the model achieves real-time performance while maintaining competitive segmentation results. Its ability to work efficiently on resource-constrained devices makes it a strong candidate for applications that require fast, pixel-wise segmentation in dynamic, real-world scenarios. One of the primary design goals of PP-LiteSeg is to reduce the computational complexity of the model. To achieve this, PP-LiteSeg uses STDCNet (Short-Term Dense

Concatenate Network) as its backbone that is faster and more efficient than traditional deep convolutional networks like ResNet.

Analysis of modern real-time semantic segmentation models performance

In this chapter we compare the real-time segmentation models and approaches reviewed in this article by their accuracy (mean intersection over union metric) and inference speed (frames per second). Still it is hard to objectively compare the real inference performance between models that were tested on different hardware and with different input image resolutions. In Table 2 we provide the list of models from recent researches that achieved a high mean Intersection over Union (mIoU) metric for real-time semantic segmentation task.

Table 2

Comparison of real-time semantic segmentation models by mIoU metric

№	Model	Resolution	Dataset	mIoU	FPS	GPU	Hardware specs	Year
1	PIDNet-L	2048×1024	Cityscapes	**80.6%**	31.1	RTX 3090	24 GB RAM 3584 CUDA cores 3.1Hz	2022
2	PIDNet-M	2048×1024	Cityscapes	**79.8%**	42.2	RTX 3090	24 GB RAM 3584 CUDA cores 3.1Hz	2022
3	PIDNet-S	2048×1024	Cityscapes	**78.6%**	93.2	RTX 3090	24 GB RAM 3584 CUDA cores 3.1Hz	2022
4	SFNet-R18	1024 × 2048	Cityscapes	**80.4%**	25.5	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2020
5	HarDNet	-	Cityscapes	**75.9%**	53	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2019
6	BiSeNet	1024 × 2048	Cityscapes	**74.7%**	65.5	Titan XP	12 GB RAM 3840 CUDA cores 1.5Hz	2018

The leader in high accuracy oriented real-time semantic segmentation models is the PIDNet. The results, provided by the authors, were achieved by testing the model on NVIDIA RTX 3090 GPU. PIDNet-S modification provides the best tradeoff between accuracy and inference speed with 78.6% mIoU score and 93.2 FPS. PIDNet-L has an increased accuracy of 80.6% but an inference speed only of 31.1 FPS. Although SFNet-R18 provides an accuracy at the level of PIDNet (80.4%), the inference speed is significantly lower than it's competitors making only 25.5 FPS on NVIDIA GTX 1080Ti. While maintaining an efficient memory usage, HarDNet has a drop in accuracy compared to PIDNet and SFNet hitting mIoU of 75.9% and 53 FPS on GTX 1080 Ti. Unfortunately, the authors did not provide the exact image resolution the model was tested on. The experimental results show how the 3 branch architecture of PIDNet over performs the 2 branch architecture of BiSeNet both in accuracy and inference speed, where BiSeNet achieves an mIoU only of 74.4% and 65.5 FPS on NVIDIA Titan XP.

While Table 2 highlights the comparative mIoU metrics of various real-time semantic segmentation models, showcasing their accuracy levels, Table 3 transitions to a comparison of high inference speed real-time semantic segmentation models.

Experimental results show that STDCNet alone can reach an accuracy of 76.8% on 768×1536 pixel resolution images with inference speed of 97 FPS achieved by STDC2-Seg75 model variation. With reduction of image resolution the accuracy of STDC while PP-LiteSeg makes further improvements on top of STDCNet. PP-LiteSeg-B2 version achieves an accuracy of 77.5% with inference speed of 102.6 FPS on NVIDIA GTX 1080 Ti, loosing only 1.1% of accuracy and gaining almost 10 FPS compared to PIDNet-S. PP-LiteSeg-T and PP-LiteSeg-B versions use STDC1 and STDC2 backbones respectively. While BiSeNet V2 achieves a good balance between speed and accuracy hitting 72.6% mIoU and 156 FPS on Cityscapes dataset, it may still not match the absolute accuracy of more complex models and it's real-time semantic segmentation competitors.

Table 3

Comparison of real-time semantic segmentation models by inference speed

№	Model	Resolution	Dataset	mIoU	FPS	GPU	Hardware specs	Year
1	PP-LiteSeg-T1	512 × 1024	Cityscapes	72.0%	**273.6**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2022
2	PP-LiteSeg-B1	512 × 1024	Cityscapes	73.9%	**195.3**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2022
3	PP-LiteSeg-T2	768 × 1536	Cityscapes	74.9%	**143.6**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2022
4	PP-LiteSeg-B2	768 × 1536	Cityscapes	77.5%	**102.6**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2022
5	STDC1-Seg50	512 × 1024	Cityscapes	71.9%	**250.4**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2021
6	STDC2-Seg50	512 × 1024	Cityscapes	73.4%	**188.6**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2021
7	STDC1-Seg75	768 × 1536	Cityscapes	75.4%	**126.7**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2021
8	STDC2-Seg75	768 × 1536	Cityscapes	76.8%	**97**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2021
7	BiSeNet V2	2048 × 1024	Cityscapes	72.6%	**156**	GTX 1080 Ti	11GB RAM 3584 CUDA cores 1.5Hz	2020

Conclusions

This paper has provided a comprehensive analysis of methods and tools for real-time semantic segmentation, focusing on their performance metrics, architectures, and application contexts. The findings from this study highlight the significant advancements in the field, particularly the shift from traditional handcrafted feature-based methods to sophisticated deep learning models, such as Fully Convolutional Networks (FCNs), U-Net to lightweight architectures like PP-LiteSeg and PIDNet.

Through a detailed evaluation of trade-offs between accuracy and computational efficiency, we identified that models like PIDNet and PP-LiteSeg exhibit exemplary performance in balancing these aspects, making them particularly suited for real-time applications in resource-constrained environments. PIDNet's novel three-branch network inspired by Proportional-Integral-Derivative controllers achieves a remarkable synergy between high accuracy and edge detail preservation, while PP-LiteSeg excels in lightweight design and processing speed, thanks to innovations like the STDCNet backbone and Unified Attention Fusion Module.

Performance assessments on benchmark datasets such as Cityscapes revealed the importance of optimizing inference times and mIoU scores for diverse applications. For instance, while autonomous driving demands high-resolution inputs for precise decision-making, edge computing scenarios require efficient models capable of operating on hardware like NVIDIA Jetson Nano and Raspberry PI5. Our comparative results illustrate the practical implications of selecting the right algorithm for specific operational contexts.

References

1. Khan, B. A., & Jung, J. W. (2024). Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions. Applied Sciences, 14(9), 3712.

2. Chepys, O. (2024, July). Algorithms for effective control of Unmanned aerial vehicles using machine vision. In VII International Scientific and Practical Conference «SCIENCE OF POST-INDUSTRIAL SOCIETY: GLOBALIZATION AND TRANSFORMATION PROCESSES». LLC UKRLOGOS Group. 260-264.

3. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259-265.
4. Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. Sustainability, 13(3), 1224.
5. Wang, J., Song, L., Li, Z., Sun, H., Sun, J., & Zheng, N. (2021). End-to-end object detection with fully convolutional network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 15849-15858).
6. Butko, I., Holubenko, O., Makoveichuk, O. (2024). Semantic segmentation in multispectral imagery. ITSynergy, (1), 16-29.
7. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.
8. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., & Sang, N. (2018). Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 325-341).
9. Peng, J., Liu, Y., Tang, S., Hao, Y., Chu, L., Chen, G., Wu, Z., Chen, Z., Yu, Z., Du, Y. and Dang, Q. (2022). Pp-liteseg: A superior real-time semantic segmentation model. arXiv preprint arXiv:2204.02681.
10. Yurtsever, E., Lambert, J., Carballo, A., & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. IEEE access, 8, 58443-58469.
11. Kinnari, J., Renzulli, R., Verdoja, F., & Kyrki, V. (2023). LSVL: Large-scale season-invariant visual localization for UAVs. Robotics and Autonomous Systems, 168, 104497.
12. Khorolska, K. (2022). THE POTENTIAL OF VARIOUS ARTIFICIAL INTELLIGENCE METHODS APPLICATION IN THE PROBLEM OF DRAWING RECOGNITION AND 2D - 3D TRANSFORMATION. Electronic scientific publication "Cybersecurity: Education, Science, Technique", 1(17), 21–30. DOI: 10.28925/2663-4023.2022.17.2130
13. Lukashchuk B. (2024). Human skin wound image segmentation using convolutional neural networks. Herald of Khmelnytskyi National University. Technical sciences, 339(4), 368-374.
14. Zhang, W., & Ling, M. (2024). An improved point cloud denoising method in adverse weather conditions based on PP-LiteSeg network. PeerJ Computer Science, 10, e1832.
15. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
16. Hello AI World guide to deploying deep-learning inference networks and deep vision primitives with TensorRT and NVIDIA Jetson. URL: https://github.com/dusty-nv/jetson-inference
17. Xu, J., Xiong, Z., & Bhattacharyya, S. P. (2023). PIDNet: A real-time semantic segmentation network inspired by PID controllers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 19529-19539).
18. Li, X., You, A., Zhu, Z., Zhao, H., Yang, M., Yang, K., & Tong, Y. (2020). Semantic flow for fast and accurate scene parsing. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16 (pp. 775-793). Springer International Publishing.
19. Chao, P., Kao, C. Y., Ruan, Y. S., Huang, C. H., & Lin, Y. L. (2019). Hardnet: A low memory traffic network. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3552-3561).
20. Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), 9(10), 143-150.
21. Fan, M., Lai, S., Huang, J., Wei, X., Chai, Z., Luo, J., & Wei, X. (2021). Rethinking bisenet for real-time semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9716-9725).
22. Yu, C., Gao, C., Wang, J., Yu, G., Shen, C., & Sang, N. (2021). Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. International journal of computer vision, 129, 3051-3068.

Markijan Durkot **Маркіян Дуркот**	PhD Student of Software Department, Lviv National Polytechnic University e-mail: markiian.durkot@lpnu.ua https://orcid.org/0000-0003-4147-3077	Аспірант кафедри програмного забезпечення, Національний університет "Львівська політехніка", Львів, Україна
Nataliia Melnyk **Наталія Мельник**	PhD, Associate Professor of Software Department, Lviv National Polytechnic University, Lviv, Ukraine e-mail: nataliia.b.melnyk@lpnu.ua https://orcid.org/0000-0003-2337-2395 Scopus: 26537815100	Кандидат фізико-математичних наук, доцент, доцент кафедри програмного забезпечення, Національний університет "Львівська політехніка", Львів, Україна