

**ПРОБЛЕМИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ВБУДОВАНИХ СИСТЕМ
КЕРУВАННЯ З МОДУЛЕМ КОМП'ЮТЕРНОГО ЗОРУ**

Гусев А.Ю.¹ [ORCID], Клим В.Ю.² [ORCID]

¹Дніпровський національний університет імені Олеся Гончара, аспірант, Україна

²Дніпровський національний університет імені Олеся Гончара

канд. наук, доц, Україна

Анотація. У роботі досліджується актуальна проблема підвищення ефективності вбудованих систем автоматичного керування, де алгоритми комп'ютерного зору (на базі глибоких нейромереж) виконують роль віртуальних датчиків. Проаналізовано методи компресії моделей (квантування, проріджування) у межах підходів Edge AI та TinyML. Висунуто припущення, що ізольована програмна оптимізація нейромереж без урахування динаміки фізичного об'єкта може знижувати запаси стійкості системи через виникнення стохастичного шуму та змінної затримки. Запропоновано комплексну методологію дослідження, яка поєднує використання цифрових двійників, симуляції (Vision-in-the-loop), емуляції та апаратної реалізації (Hardware-in-the-loop). Цей підхід спрямований на пошук оптимального компромісу між затримкою інференсу та точністю розпізнавання для підвищення надійності та ефективності автономних кіберфізичних систем.

Ключові слова: комп'ютерний зір, вбудовані системи, автоматичне керування, TinyML, компресія нейромереж.

Розвиток автономної робототехніки та безпілотних літальних апаратів (БПЛА) вимагає розширення сенсорних можливостей кіберфізичних систем. У сучасних системах керування все частіше відбувається заміна коштовних фізичних датчиків на модулі комп'ютерного зору (КЗ). Проте інтеграція сучасних алгоритмів КЗ у периферійні пристрої (Edge devices) ускладнюється жорсткими обмеженнями апаратного забезпечення: лімітованою пам'яттю, обчислювальною потужністю та енергоспоживанням. Це вимагає використання методів стиснення нейромереж, що, своєю чергою, впливає на загальну стійкість систем автоматичного керування (САК).

Для перенесення алгоритмів КЗ на бортові пристрої застосовуються підходи Edge Computing та TinyML, зокрема квантування (перетворення

значень ваг у форматі меншої розрядності, наприклад INT8) та проріджування (видалення найменш важливих синаптичних зв'язків) [1].

Проте виникає певна проблема: фахівці з Data Science оптимізують нейромережі переважно ізольовано від об'єкта керування, оцінюючи якість статичними метриками. Але інтеграція такого модуля в контур зворотного зв'язку докорінно змінює характер його роботи. Відеосенсор виступає як динамічна ланка із запізненням, що описується передавальною функцією:

$$W(s) = e^{(-sT)}, \quad (1)$$

де T — час затримки.

Перевищення граничної затримки робить систему нестійкою.

Щоб зменшити затримку, моделі компресують. Однак зниження розрядності ваг генерує шум квантування (хаотичне тремтіння координат обмежувальної рамки). Сигнал y_t на виході такого віртуального датчика описується рівнянням:

$$y(t) = f_{cnn}(x(t - \tau(t))) + w(t), \quad (2)$$

Де f_{cnn} (function of Convolutional Neural Network) - це функція перетворення згорткової нейромережі (модуля комп'ютерного зору), яка відображає реальний просторовий стан об'єкта у вихідні координати обмежувальної рамки; $x(t)$ - реальний стан керованого об'єкта, $\tau(t)$ - змінна затримка інференсу, а $w(t)$ - стохастичний шум вимірювань, спричинений компресією; t - час.

У контурі керування цей шум може діяти деструктивно на диференціальну складову ПІД (пропорційно-інтегрально-диференціальний) - регулятора, змушуючи його хаотично "смикати" виконавчі механізми [2].

Для вирішення цієї проблеми пропонується багаторівнева методологія комплексного дослідження, що складається з 6 етапів:

1. Створення цифрового двійника (3D-моделі) об'єкта в середовищах Unity або NVIDIA Isaac Sim.
2. Синтез ідеального керування (Ground Truth) за допомогою віртуальних датчиків симулятора.
3. Vision-in-the-loop: впровадження нестисненої нейромережі у контур керування.
4. Оптимізація та емуляція (TinyML) з різними ступенями квантування.

5. Апаратна реалізація (Hardware-in-the-loop) на реальних контролерах (STM32, Raspberry Pi, NVIDIA Jetson).
6. Порівняльний аналіз для оцінки взаємовпливу компресії та запасів стійкості САК.

Висновки

Підвищення ефективності вбудованих систем керування з модулем комп'ютерного зору потребує міждисциплінарного (Co-design) підходу. Ізольована оптимізація швидкості інференсу без урахування динаміки об'єкта може призводити до втрати стійкості через шум квантування. Запропонована методологія дозволить обґрунтувати гранично допустимий ступінь стиснення мережі, за якого кіберфізичний об'єкт зберігає керованість та задані показники якості перехідного процесу.

ЛІТЕРАТУРА

1. El Zeinaty C., Hamidouche W., Herrou G., Menard D. Designing Object Detection Models for TinyML: Foundations, Comparative Analysis, Challenges, and Emerging Solutions. ACM Computing Surveys. 2024. Vol. 56. P. 1–46. DOI: <https://doi.org/10.1145/3744339>.
2. Park J., Kim P., Ko D. Real-time open-vocabulary perception for mobile robots on edge devices: a systematic analysis of the accuracy-latency trade-off. Frontiers in Robotics and AI. 2025. Vol. 12. 1693988. DOI: <https://doi.org/10.3389/frobt.2025.1693988>.
3. Худяков І. В., Грицук І. В., Черненко В. В. та ін. Особливості моделювання та побудови інформаційної системи дистанційного моніторингу технічного стану транспортних засобів. Вісник машинобудування та транспорту. 2021. № 2 (14). С. 140–148. DOI: <https://doi.org/10.31649/2413-4503-2021-14-2-140-148>.

ISSUES OF IMPROVING THE EFFICIENCY OF EMBEDDED CONTROL SYSTEMS WITH A COMPUTER VISION MODULE

Artem Husiev, Viktoriia Klym

Abstract. *This paper investigates the relevant problem of improving the efficiency of embedded automatic control systems, where deep learning-based computer vision algorithms act as virtual sensors. Methods of model compression (quantization, pruning) within Edge AI and TinyML approaches are analyzed. It is hypothesized that isolated software optimization of neural networks, without considering the dynamics of the physical object, may reduce the system's stability margins due to the emergence of stochastic noise and variable latency. A comprehensive research methodology is proposed, combining the use of digital twins, simulation (Vision-in-the-loop), emulation, and*

hardware implementation (Hardware-in-the-loop). This approach aims to find the optimal trade-off between inference latency and recognition accuracy to enhance the reliability and efficiency of autonomous cyber-physical systems.

Keywords: *computer vision, embedded systems, automatic control, TinyML, neural network compression.*

REFERENCE

1. El Zeinaty, C., Hamidouche, W., Herrou, G., & Menard, D. (2024). Designing object detection models for TinyML: Foundations, comparative analysis, challenges, and emerging solutions. *ACM Computing Surveys*, 56, 1–46. <https://doi.org/10.1145/3744339>
2. Park, J., Kim, P., & Ko, D. (2025). Real-time open-vocabulary perception for mobile robots on edge devices: A systematic analysis of the accuracy-latency trade-off. *Frontiers in Robotics and AI*, 12, 1693988. <https://doi.org/10.3389/frobt.2025.1693988>
3. Khudiakov, I. V., Gritsuk, I. V., Chernenko, V. V., et al. (2021). Osoblyvosti modeliuвання ta pobudovy informatsiinoi systemy dystantsiinoho monitorynhu tekhnichnoho stanu transportnykh zasobiv. *Visnyk mashynobuduvannia ta transportu*, 14(2), 140–148. <https://doi.org/10.31649/2413-4503-2021-14-2-140-148>.