УДК 004.93’11

ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ U-NET, DPT ТА MIDAS ДЛЯ ОЦІНКИ ГЛИБИНИ У SLAM-СИСТЕМАХ

Мороз М. В.[1]; Вишнівський В. В., д.т.н., проф.[2]
[12]Державний університет інформаційно-комунікаційних технологій
[12] Україна, Київ
[1] m.moroz@stud.duikt.edu.ua; [2] v.vyshnivskyi@duikt.edu

Анотація. *Проведено порівняльний аналіз точності та швидкодії моделей U-Net, DPT та MiDaS для задачі оцінки глибини сегментованого зображення у системах SLAM за допомогою модульного поєднання з існуючим алгоритмом ORB-SLAM3. Результати показали, що DPT забезпечує найвищу точність оцінювання, тоді як MiDaS є більш придатною для задач реального часу з обмеженими ресурсами. Було запропоновано критерії вибору моделі на основі вимог до точності обчислення та ресурсних можливостей системи.*
Ключові слова: *SLAM; U-Net; DPT; MiDaS; оцінка глибини; машинне навчання.*

Вступ

Точна оцінка глибини отриманого зображення є важливою підзадачею у системах, що реалізують одночасну оцінку положення та побудову мапи середовища (SLAM). Конвенційні методи використовують стереозображення, або апаратне забезпечення як LiDAR, однак їх ефективність суттєво знижується у складних динамічних умовах або за наявності лише монокулярної камери.

Нейромережеві моделі, такі як U-Net, DPT та MiDaS, показують здатність до відновлення геометрії та глибини з монокулярного RGB-зображення [1–3]. Проте задача інтеграції таких моделей у систему SLAM потребує компромісу між точністю, швидкодією та стабільністю роботи.

Мета роботи – провести порівняння трьох вибраних моделей оцінювання глибини для визначення їх придатності до використання у візуальних монокулярних системах SLAM**.**

Основна частина

Для дослідження було використано існуючу модульну систему ORB-SLAM3, у якій було інтегровано модуль оцінки глибини. Кожна модель отримувала RGB зображення та генерувала оцінку глибини для кожного конкретного зображення. Експерименти проводилися на наборі даних Low Viewpoint Forest Depth [4]. Метрики оцінювання включали середню абсолютну похибку (MAE), а також швидкість обробки кадрових зображень (FPS).

Таблиця 1. Порівняння моделей на основі середньої абсолютної похибки та швидкодії

Модель	MAE	FPS
DPT	0.093	12
U-Net	0.128	30
MiDaS	0.106	22

Експериментальні результати показали (таблиця 1) (рис. 1), що модель DPT досягла найменшої середньої абсолютної похибки, але при цьому використовує значний обчислювальний ресурс. Модель U-Net забезпечує задовільну точність із високим ступенем швидкодії. Модель MiDaS продемонструвала компроміс між точністю і швидкодією.

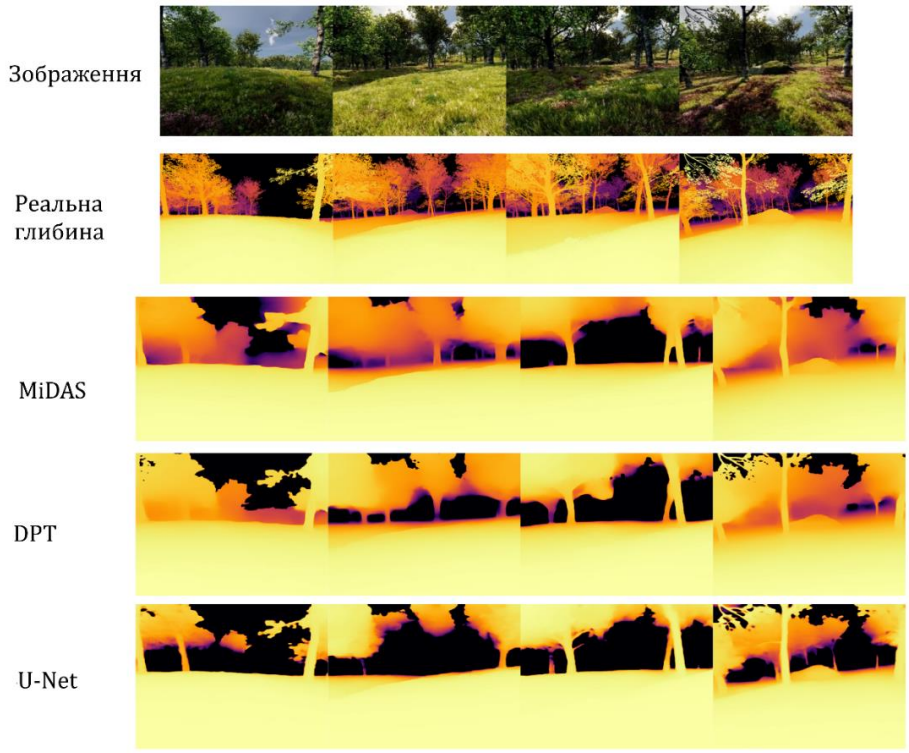

Рисунок 1. Візуальне порівняння згенерованих мап глибини та реального зображення.

Висновки

Було проведено порівняльний аналіз моделей U-Net. DPT та MiDaS, для задачі оцінки глибини зображення у системах SLAM. Вибір конкретної моделі повинен залежати від цільового застосування, як от модель DPT для задач високоточної оцінки глибини зображень із значними обчислювальними ресурсами, MiDaS - для динамічних середовищ та систем реального часу, U-Net – для автономних роботизованих платформ із обмеженими ресурсами. Подальші дослідження можуть включати об’єднання переваг конкретних моделей у гібридному рішенні та інтеграцію з різними типами систем SLAM.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI*, 234–241.

[2] Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision Transformers for Dense Prediction. *ICCV*, 12159–12168.

[3] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2020). Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. doi:10.48550/arXiv:1907.01341.

[4] Niu, C., Tarapore, D., & Zauner, K.-P. (2020). Low-viewpoint forest depth dataset for sparse rover swarms. doi:10.48550/arXiv:2003.04359.

Moroz M.V., Vyshnivskyi V.V.

Comparative analysis of U-Net, DPT and MiDaS models for depth estimation in SLAM systems

Abstract. *A comparative analysis of the accuracy and performance of the U-Net, DPT, and MiDaS models for the task of segmented image depth estimation in SLAM systems using a modular combination with the existing ORB-SLAM3 algorithm was conducted. The results showed that DPT provides the highest estimation accuracy, while MiDaS is more suitable for real-time tasks with limited resources. The model selection criteria were proposed based on the requirements for computational accuracy and system resource capabilities.*

Keywords: *SLAM; U-Net; DPT; MiDaS; depth estimation; machine learning.*