

УДК 004.93

Р. Н. КВЕТНИЙ, Р. В. МАСЛІЙ, О. М. КИРИЛЕНКО

ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЯ ОБ'ЄКТІВ ДОРОЖНЬОГО РУХУ ПРИ ВИКОРИСТАННІ СЕРЕДОВИЩА DIGITS

*Вінницький національний технічний університет
Хмельницьке шосе, 95, м. Вінниця, 21021, Україна
Тел. (0432) 59-80-06, e-mail: sasha.kyrylenko@gmail.com*

Анотація. Здійснено огляд архітектури нейронної мережі DetectNet з метою навчання моделі виявлення та класифікації об'єктів дорожнього руху. При цьому, розглянуто структуру нейронної мережі та формат вхідних даних. Здійснено навчання моделі за допомогою середовища DIGITS. Якість моделі було перевірено на валідаційному наборі зображень KITTI. Наведено результати навчання моделі нейронної мережі. Отримані результати порівняно з існуючими аналогами.

Ключові слова: глибоке навчання, виявлення об'єктів, класифікація об'єктів, DetectNet, DIGITS, KITTI.

Анотация. Осуществлен обзор архитектуры нейронной сети DetectNet с целью обучения модели обнаружения и классификации объектов дорожного движения. При этом, рассмотрена структура нейронной сети и формат входных данных. Осуществлено обучение модели с помощью среды DIGITS. Качество модели было проверено на валидационном наборе изображений KITTI. Приведены результаты обучения модели нейронной сети. Полученные результаты по сравнению с существующими аналогами.

Ключевые слова: глубокое обучение, обнаружение объектов, классификация объектов, DetectNet, DIGITS, KITTI.

Abstract. An overview of the architecture of the DetectNet neural network was conducted to study the model of detection and classification of traffic objects. In this case, the structure of the neural network and the format of the input data are considered. The modeling is done using the DIGITS environment. The quality of the model was tested on the image validation dataset KITTI. The results of studying the model of the neural network are presented. The results obtained compared with existing ones.

Keywords: deep learning, object detection, object classification, DetectNet, DIGITS, KITTI.

DOI: 10.31649/1681-7893-2020-39-1-14-20

ВСТУП

Машинне навчання дає можливість вирішувати деякі задачі, що зазвичай вирішуються людиною, без участі людини, наприклад, такі як діагностика захворювань, розпізнавання голосу, розпізнавання об'єктів у зображеннях тощо. Сьогодні одним із основних напрямків машинного навчання є глибоке навчання (deep learning). Одним з перспективних застосувань глибокого навчання є його використання для задач розпізнавання об'єктів дорожнього руху. Виявлення та класифікація об'єктів дорожнього руху у відеозображеннях отриманих з камер спостереження є важливими задачами, що вирішуються у системах керування дорожнім трафіком "розумних" міст. Такі системи не тільки роблять міське життя зручнішим, але й безпечнішим за рахунок зменшення кількості дорожньо-транспортних пригод [1].

Однак коректне виявлення та класифікація об'єктів дорожнього руху у відеозображеннях є не простою задачею. Це зумовлено багатьма факторами, серед яких і неконтрольовані умови освітлення, і значна кількість типів об'єктів, і перекриття одних об'єктів іншими, і різноманітність форм об'єктів, що зустрічаються у відео зображеннях [2]. Наприклад, у відеозображеннях можуть бути присутні пішоходи, автомобілі, велосипеди, дорожні знаки, світлофори тощо. Кожен з цих об'єктів може мати варіацію форм, наприклад, автомобілі мають різні форми залежно від їх типів: седан, вантажівка, фургон і позашляховик. Класичні методи машинного навчання, такі як бустінг методи, чи звичайні нейронні мережі не можуть у повній мірі розпізнати ці об'єкти [1].

Сьогодні за допомогою NVIDIA Deep Learning GPU Training System (DIGITS) дослідники-аналітики мають у своєму розпорядженні всю міць глибокого навчання, для вирішення загальних завдань, таких як: підготовка даних, визначення згорткових мереж, паралельне навчання декількох моделей,

спостереження за процесом навчання в реальному часі, а також вибір кращої моделі. У версії DIGITS 6 представлено новий підхід до задачі виявлення об'єктів, який дозволяє навчати глибокі нейронні мережі здійснювати виявлення об'єктів (таких як пішоходи та транспортні засоби) у зображеннях [3].

Метою даної статті є підвищення вірогідності виявлення та класифікації об'єктів дорожнього руху у зображеннях отриманих з камер відеоспостереження при використанні середовища DIGITS.

Задачі дослідження:

1. Дослідження нейромережевої архітектури DetectNet.
2. Навчання моделі виявлення та класифікації об'єктів при використанні середовища DIGITS.
3. Тестування навчених моделей на наборі даних KITTI.
4. Порівняння отриманих результатів з існуючими аналогами.

Архітектура мережі DetectNet

DetectNet – це нейромережева архітектура, що одночасно здійснює як виявлення так і класифікацію об'єктів у зображеннях. DetectNet є розширенням популярної нейромережі GoogleNet.

В цілому процес навчання мережі DetectNet складається з двох етапів: етап навчання мережі на тренувальних зображеннях та етап перевірки на валідаційних зображеннях. На рисунку 1 наведений етап навчання мережі DetectNet. В ньому можна виділити 3 важливі частини:

- a) зображення (images) і відповідні мітки (labels) навчальної вибірки надходять на вхід шару даних. Далі, шар перетворення (online data augmentation) «на льоту» здійснює доповнення даних, тобто генерує нові зображення на основі існуючих шляхом зсуву, зміни масштабу існуючих;
- b) повністю згортовка мережа (fully-convolutional network або FCN) здійснює витяг ознак і передбачення класів об'єктів і обмежуючих прямокутників (bounding boxes) по квадратах сітки;
- c) функції втрат (L1 Loss та L2 Loss), одночасно обчислюють помилку по двох задачах: передбачення покриття об'єкта (predicted coverage map) і розташування обмежуючих прямокутників по квадратах сітки.

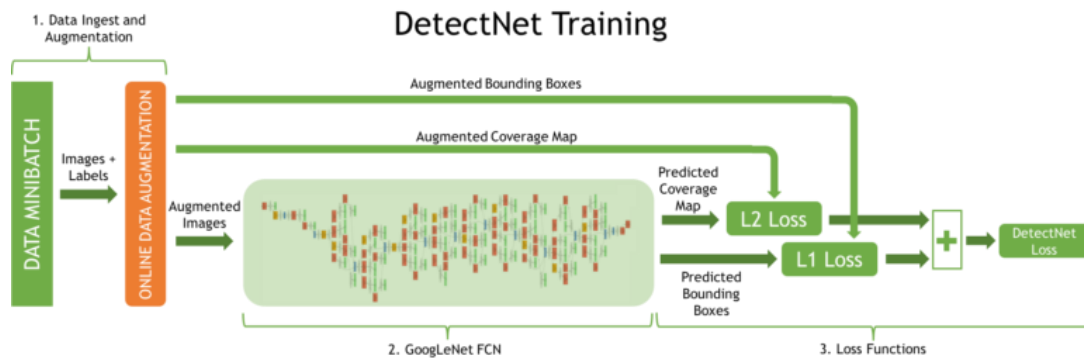


Рис. 1. Етап навчання мережі DetectNet

На рисунку 2 наведений етап перевірки мережі DetectNet на валідаційних зображеннях. На цьому етапі крім згорткової мережі задіяні такі частини як:

- a) кластеризація передбачених обмежуючих прямокутників для отримання остаточного набору обмежуючих прямокутників.
- b) підрахунок спрощеної метрики mAP (mean Average Precision) для оцінювання ефективності моделі на всьому валідаційному наборі (validation dataset) [4].

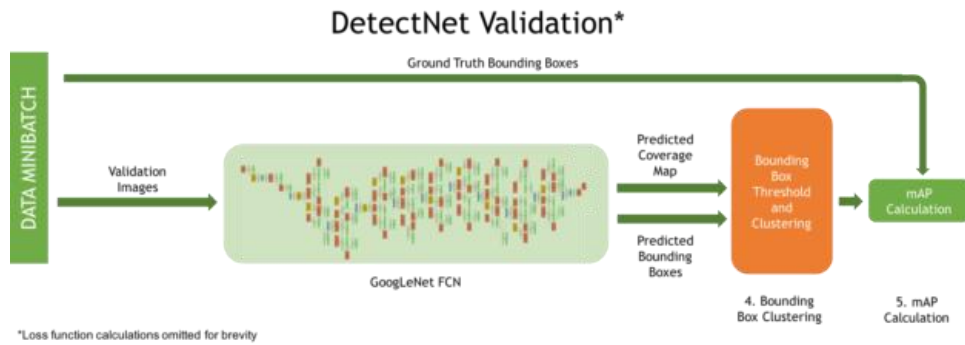


Рис. 2. Етап перевірки мережі DetectNet

Формат даних DetectNet

В якості вхідних даних навчальної вибірки для задачі класифікації зображень використовуються зображення (зазвичай невеликого розміру і містять один об'єкт) і класові мітки (зазвичай цілочисельний ідентифікатор класу або рядковий – назва класу). З іншого боку, для задачі виявлення об'єктів необхідно більше інформації для навчання. Зображення навчальної вибірки для DetectNet мають більший розмір і містять кілька об'єктів, і, для кожного об'єкта на зображенні мітка повинна містити не тільки інформацію про клас, до якого належить об'єкт, але і розташування кутів його обмежувального прямокутника. В даному випадку наївний вибір формату мітки зі змінною довжиною і розмірністю призводить до того, що визначення функції втрат (loss function) може бути ускладнене, оскільки кількість об'єктів у цьому зображенні може змінюватись. DetectNet вирішує цю проблему, використовуючи фіксований тривимірний формат підпису, що дозволяє працювати з зображеннями будь-якого розміру та різною кількістю присутніх у ньому об'єктів. Схема обробки зображень з навчальної вибірки з розміткою для навчання мережі DetectNet представлена на рисунку 3.

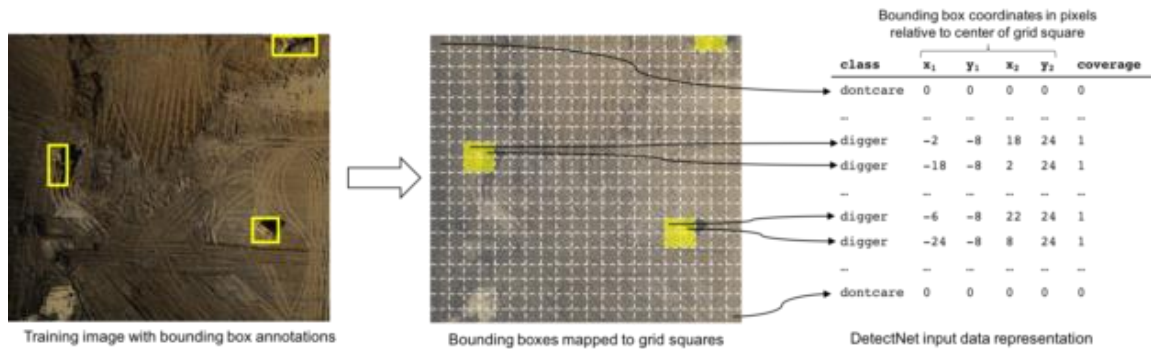


Рис. 3. Подання вхідних даних для мережі DetectNet

Спочатку, на вхідне зображення накладається фіксована сітка, яка розміром трохи менша, за найменший об'єкт, який потрібно виявити. Далі, кожен квадрат решітки розмічається з наступною інформацією: клас об'єкта, що знаходиться в квадраті решітки, і координати пікселя кутів прямокутника, що обмежує відносно центру квадрата решітки. У разі, якщо жоден об'єкт не потрапив в квадрат решітки, то використовується спеціальний клас "dontcare" для збереження фіксованого формату даних. Також у формат вхідних даних додається додаткове значення "coverage", яка приймає значення 0 або 1, щоб вказати, чи є об'єкт у квадраті чи ні. У разі, коли кілька об'єктів потрапляють в один квадрат решітки, DetectNet вибирає той об'єкт, який займає найбільшу кількість пікселів. Якщо кількість пікселів однакова, то вибирається об'єкт з найменшою ординатою (OY) обмежувального прямокутника. Такий вибір об'єктів не є принциповим для зображень аерофотозйомки, але має сенс для зображень з горизонтом, наприклад, для зображень з відеореєстратора, де об'єкт з найменшою ординатою обмежувального прямокутника знаходиться на ближчій відстані до камери [5].

Нехай I_n – зображення, $1 \leq n \leq N_3$, де N_3 – загальна кількість зображень, $N_3 = N_{tr} + N_{val}$, де N_{tr} – це кількість тренувальних зображень, N_{val} – це кількість валідаційних зображень. Нехай кожне зображення покривається сіткою, рівномірними прямокутниками, де кожна сторона прямокутника відповідно $N_{gr} \times M_{gr}$ пікселів.

Для зручності визначимо: N_{tbb} – істинна кількість обмежувальних прямокутників у зображенні n ; N_{pbb} – прогнозована кількість обмежувальних прямокутників у зображенні n ; N_g – кількість прямокутників

в сітці у зображенні n . Тому область покриття справжнього обмежувального прямокутника, P_{i_n} , $1 \leq i_n \leq N_{tbb}$, $1 \leq n \leq N_{\Sigma}$ з сіткою G_{j_n} є:

$$coverage(P_{i_n}) = \sum_{J_n=1}^{N_g} \frac{Area(P_{i_n} \cap G_{j_n})}{Area(G_{j_n})} \in [0,1], \quad (1)$$

і охоплення прогнозованого обмежувального прямокутника, K_{j_n} , $1 \leq j_n \leq N_{pbb}$, $1 \leq n \leq N_{\Sigma}$ з сіткою G_{j_n} є:

$$coverage(K_{j_n}) = \sum_{J_n=1}^{N_g} \frac{Area(P_{i_n} \cap G_{j_n})}{Area(G_{j_n})} \in [0,1], \quad (2)$$

Якщо

$$IOU_{P_{i_n}K_{j_n}} = \frac{Area(P_{i_n} \cap K_{j_n})}{Area(P_{i_n}) + Area(K_{j_n}) - Area(P_{i_n} \cap K_{j_n})} > \gamma, \quad (3)$$

де γ це деякий поріг, відмінний від K_{j_n} , що є істинно позитивним виявленням у зображенні, в іншому випадку це помилкове виявлення.

Показники точності (precision) та чутливості (recall) для валідаційних даних визначаються як:

$$Precision_{val} = \sum_{k=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FP_k}} \quad (4)$$

$$Recall_{val} = \sum_{k=1}^{N_{val}} \frac{N_{TP_k}}{N_{TP_k} + N_{FN_k}} \quad (5)$$

де N_{TP} визначається як число істинно позитивних виявлень у зображенні, N_{FP} визначається як кількість хибно позитивних виявлень у зображенні і N_{FN} – це кількість помилково негативних виявлень у зображенні.

$$CoverageLoss = coverage(P) - coverage(K) \quad (6)$$

$$mAP = Precision_{val} \times Recall_{val} \quad (7)$$

$BboxLoss$ визначається як різниця межі істинного обмежувального прямокутника в лівому нижньому кутку та правого верхнього кута та лівого нижнього кута передбачуваного обмежувального прямокутника. Якість моделей, що були навчені на основі архітектури DetectNet, визначається значеннями показників: $LossBox$ (training), $LossBox$ (validation), $LossCoverage$ (valuation), mAP (validation), $Precision$ (validation), $Recall$ (valuation) та $LossCoverage$ (training). Глибока нейронна мережа адаптує ваги, відповідно до методу градієнтного спуску [6].

Таким чином, метою навчання моделі за допомогою нейромережі DetectNet є визначення для кожного квадрата решітки присутності у ньому об'єкта та обчислення відносних координат кутів обмежувального прямокутника.

ОПИС ЕКСПЕРИМЕНТУ

Для створення та навчання моделей було використано середовище NVIDIA DIGITS [7].

Основні особливості середовища NVIDIA DIGITS:

- можливість створення, тренування та візуалізації глибоких нейронних мереж для класифікації, сегментації та виявлення об'єктів за допомогою Caffe, Torch та TensorFlow;
- можливість використання попередньо навчених моделей (AlexNet, GoogLeNet, LeNet та UNET);
- моніторинг та управління процесом навчання нейронних мереж;
- аналіз якості моделей у режимі реального часу;
- навчання при використанні декількох графічних процесорів.

Тренування та тестування моделі відбувалось на відкритому наборі даних КІТТІ [8]. Даний набір містить зображення, отримані зі стереокамери, що була розміщена на даху автомобіля. Набір містить 7481 навчальних зображень та 7518 валідаційних зображень, що в цілому мають близько 80 тисяч позначених об'єктів. Всі зображення кольорові та збережені у форматі png.

Навчання моделі здійснювалося при таких параметрах: кількість епох навчання – 100, solver type – Adam, policy – Exponential decay, gamma – 0.99, base learning rate – 0.000025. Візуалізація показників

якості моделі (mAP, precision, recall) в процесі навчання при використанні валідаційного набору зображень на кожній епоці навчання наведено на рисунку 4. На рисунку 5 наведено зміна параметра learning rate в залежності від епохи навчання.

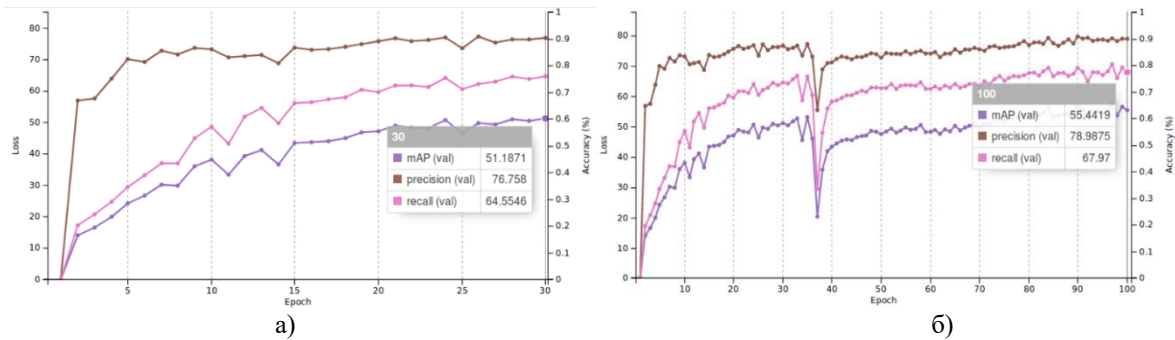


Рис. 4. Зміна показників якості моделі в процесі навчання
а) для 30 епох; б) для 100 епох

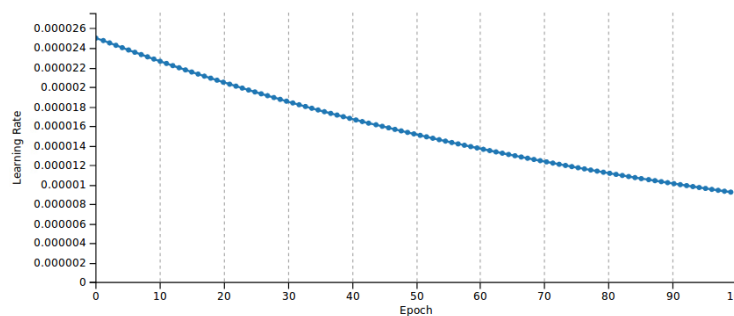


Рис. 5. Зміна параметру learning rate в залежності від епохи навчання

В результаті навчання нейронної мережі на 100 епохах було досягнуто таких показників якості моделі на валідаційному наборі зображень KITTI: precision (val) = 79%, recall (val) = 73%, mAP (val) = 55%. Порівняємо отримані результати з вже існуючою моделлю [9], що була навчена при використанні сучасної нейромережевої архітектури SSD MobileNet. На валідаційному наборі зображень KITTI ця модель показала такі показники якості: precision (val) = 69,7%, mAP (val) = 51,7%, що значно менші за показники вірогідності отриманої моделі.

Приклад виявлення автомобілів у зображенні з тестового набору KITTI при використанні навченої моделі зображено на рисунку 6.

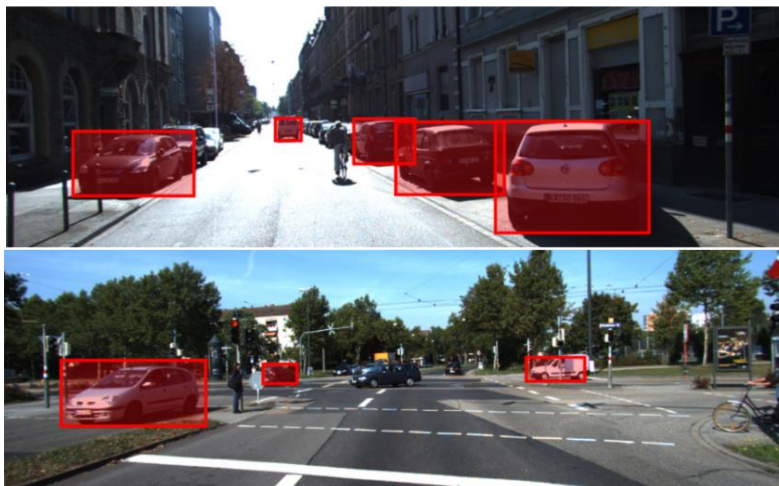


Рис. 6. Приклад виявлення автомобілів

ВИСНОВКИ

1. Проведено дослідження нейромережевої архітектури DetectNet. Розглянуто основні етапи навчання моделі, проаналізовано формат даних, що використовується для навчання мережі DetectNet.
2. За допомогою середовища DIGITS було здійснено навчання моделі виявлення та класифікації автомобілів у зображеннях отриманих з камер відеоспостереження.
3. Якість отриманої в результаті навчання моделі було перевірено на валідаційному наборі даних КІТТІ (mAP = 55%, precision = 79%, recall = 73%)
4. Доцільність продовження даних досліджень, для підвищення вірогідності виявлення та класифікації об'єктів дорожнього руху та здійснення перевірки ефективності отриманих результатів на реальних даних отриманих в реальному часі, підтверджено порівнянням отриманих результатів з існуючими аналогами.

ЛІТЕРАТУРА

1. Nagaraj, S. Edge-based street object detection. / Muthiyan, B., Ravi, S., Menezes, V., Kapoor, K., & Jeon, H. // IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI). – 2017. – С.1-4.
2. Roman Kyvetnyy, Roman Maslii, Volodymyr Harmash, Ilona Bogach, Andrzej Kotyra, Żaklin Grądz, Aizhan Zhanpeisova, Nursanat Askarova. Object detection in images with low light condition, Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017, 104450W (7 August 2017); doi: 10.1117/12.2281001;
3. Маслій Р. В. Використання локальних бінарних шаблонів для розпізнавання облич на напівтонових зображеннях [Електронний ресурс] / Р. В. Маслій // Наукові праці Вінницького національного технічного університету. - 2008. - № 4. - Режим доступу : <http://praci.vntu.edu.ua/index.php/praci/article/view/95>.
4. DetectNet: Deep Neural Network for Object Detection in DIGITS. [Електронний ресурс] – Режим доступу: <https://devblogs.nvidia.com/detectnet-deep-neural-network-object-detection-digits/>
5. Синеглазов В. Глибокі нейронні мережі для вирішення завдань розпізнавання і класифікації зображення [Електронний ресурс] / В. Синеглазов, О. Чумаченко. // – Інформаційні технології та комп'ютерне моделювання 2017. – Режим доступу до ресурсу: <http://itcm.comp-sc.if.ua/2017/Sineglazov.pdf>
6. Holupka, E. J. The Detection of Implanted Radioactive Seeds On Ultrasound Images Using Convolution Neural Networks. / Rossman, J., Morancy, T., Aronovitz, J., & Kaplan, M. D. // 1st Conference on Medical Imaging with Deep Learning (MIDL 2018). – 2018.
7. NVIDIA DIGITS [Електронний ресурс] – Режим доступу до ресурсу: <https://developer.nvidia.com/digits>
8. Zhang, S. Fcn-rlstm: Deep spatio-temporal neural networks for vehicle counting in city cameras. / Wu, G., Costeira, J. P., & Moura, J. M. // In Computer Vision (ICCV), IEEE International Conference. – 2017. pp. 3687-3696.
9. Zheng Lou. Dataset bias analysis on autonomous driving [Електронний ресурс] – Режим доступу до ресурсу: https://cs230.stanford.edu/projects_spring_2018/reports/8289902.pdf
10. Fritsch Jannik. A new performance measure and evaluation benchmark for road detection algorithms. / Tobias Kuehnl, Andreas Geiger. // 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). – 2013.

REFERENCES

1. Nagaraj, S. Edge-based street object detection. / Muthiyan, B., Ravi, S., Menezes, V., Kapoor, K., & Jeon, H. // IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI). – 2017. – С.1-4.
2. Roman Kyvetnyy, Roman Maslii, Volodymyr Harmash, Ilona Bogach, Andrzej Kotyra, Żaklin Grądz, Aizhan Zhanpeisova, Nursanat Askarova. Object detection in images with low light condition, Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017, 104450W (7 August 2017); doi: 10.1117/12.2281001;

3. Roman Maslii Using Local Binary Patterns for Face Detection in Half-Tone Imagery [Electronic resource] / Maslii Roman // Scientific papers of Vinnytsia National Technical University. - 2008. - No. 4. - [Electronic resource] Access mode: <http://praci.vntu.edu.ua/index.php/praci/article/view/95>
4. DetectNet: Deep Neural Network for Object Detection in DIGITS. [Electronic resource] Access mode: <https://devblogs.nvidia.com/detectnet-deep-neural-network-object-detection-digits/>
5. Victor Sineglazov Deep neural networks for solving problems of recognition and classification of images [Electronic resource] / Sineglazov V., Chumachenko O. // - Information Technologies and Computer Modeling 2017.– [Electronic resource] Access mode: <http://itcm.comp-sc.if.ua/2017/Sineglazov.pdf>
6. Holupka, E. J. The Detection of Implanted Radioactive Seeds On Ultrasound Images Using Convolution Neural Networks. / Rossman, J., Morancy, T., Aronovitz, J., & Kaplan, M. D. // 1st Conference on Medical Imaging with Deep Learning (MIDL 2018). – 2018.
7. NVIDIA DIGITS [Електронний ресурс] – Режим доступу до ресурсу: <https://developer.nvidia.com/digits>
8. Zhang, S. Fcn-rlstm: Deep spatio-temporal neural networks for vehicle counting in city cameras. / Wu, G., Costeira, J. P., & Moura, J. M. // In Computer Vision (ICCV), IEEE International Conference. – 2017. pp. 3687-3696.
9. Zheng Lou. Dataset bias analysis on autonomous driving [Електронний ресурс] – Режим доступу до ресурсу: https://cs230.stanford.edu/projects_spring_2018/reports/8289902.pdf
10. Fritsch Jannik. A new performance measure and evaluation benchmark for road detection algorithms. / Tobias Kuehn, Andreas Geiger. // 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). – 2013.

КВЕТНИЙ РОМАН НАУМОВИЧ – д.т.н., проф., завідувач кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, Вінниця, Україна.

МАСЛІЙ РОМАН ВАСИЛЬОВИЧ – к.т.н., доцент кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, Вінниця, Україна.

КИРИЛЕНКО ОЛЕКСАНДР МИХАЙЛОВИЧ – аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, Вінниця, Україна.