

УДК 004.93

О. М. КИРИЛЕНКО

РОЗРОБКА МЕТОДУ ПОВТОРНОЇ ІДЕНТИФІКАЦІЇ ЛЮДИНИ

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

Анотація. Зроблено огляд нейромережевої архітектури OSNet з метою навчання власних моделей повторної ідентифікації людини. Також було розглянуто структуру нейронної мережі. Досліджено існуючі набори даних для навчання моделей. Здійснено навчання моделей за допомогою PyTorch. Отримані власні моделі було перевірено на валідаційних наборах даних Market-1501 та DukeMTMC-reID. Представлено результати навчання нейромережевих моделей. Отримані результати порівняно з існуючими аналогами.

Ключові слова: глибоке навчання, повторна ідентифікація людини, OSNet, PyTorch, Market-1501, DukeMTMC-reID.

Анотация. Проведен обзор нейросетевой архитектуры OSNet с целью обучения собственным моделям повторной идентификации человека. Также была рассмотрена структура нейронной сети. Исследованы существующие наборы данных для обучения моделям. Осуществлено обучение моделей с помощью PyTorch. Полученные собственные модели были проверены на валидационных наборах Market-1501 и DukeMTMC-reID данных. Представлены результаты обучения нейросетевых моделей. Получены результаты были сравнены с существующими аналогами.

Ключевые слова: глубокое обучение, повторная идентификация человека, OSNet, PyTorch, Market-1501, DukeMTMC-reID

Abstract. The review of OSNet neural network architecture is made for the purpose of training of own models of re-identification of the person. The structure of the neural network was also considered. Existing data sets for model training are investigated. Models were trained using PyTorch. The obtained own models were tested on the validation databases Market-1501 and DukeMTMC-reID. The results of learning neural network models are presented. The results are obtained in comparison with existing analogues.

Keywords: deep learning, human re-identification, OSNet, PyTorch, Market-1501, DukeMTMC-reID.

DOI: 10.31649/1681-7893-2021-41-1-25-32

ВСТУП

Повторна ідентифікація об'єктів є одним із важливих завдань у відеоаналітиці, і їй все більше приділяється уваги. Повторна ідентифікація об'єктів, як основна технологія у відеоаналізі має на меті визначити, чи належать об'єкти, що з'являються на зображеннях, які не перекривається, одному об'єкту. Незважаючи на те, що дослідники доклали великих зусиль для вирішення цієї проблеми, вона все ще має певні прошарки через великі розбіжності у точках зору, фонах, ілюмінаціях та позах [1,2].

Повторна ідентифікація особи, як фундаментальне завдання розподіленого спостереження з кількома камерами, спрямована на те, щоб зіставити людей, які з'являються в різних видах камер, що не перекриваються. При розпізнаванні рівня екземпляра, повторна ідентифікація особи стикається з двома основними проблемами, як показано на рисунку 1.



Рис. 1 – Приклад зміни умов перегляду камери

По-перше, варіації внутрішнього класу (екземпляр/ідентичність) зазвичай великі через зміни умов перегляду камери. Наприклад, обидві людини на рис. 1 (а) та (б) несуть рюкзак; вид зміна між камерами (від фронтальної до задньої) призводить до значних змін зовнішнього вигляду в районі

рюкзака, що ускладнює узгодження однієї людини. По-друге, є також невеликі міжкласові варіації - люди у публічному просторі часто носять подібний одяг, тому здалеку, як зазвичай у відеоспостереженнях, вони можуть виглядати неймовірно схожими [3].

Метою даної статті є дослідження OSNet та навчання нових нейромережових моделей для підвищення ефективності повторної ідентифікації людини.

Задачі дослідження:

1. Дослідження OSNet для задачі повторної ідентифікації людини.
2. Аналіз існуючих наборів даних для задачі повторної ідентифікації людини.
3. Дослідження метрик оцінювання методів повторної ідентифікації людини.
4. Навчання власних моделей повторної ідентифікації людини та порівняння з аналогами

Архітектура мережі OSNet

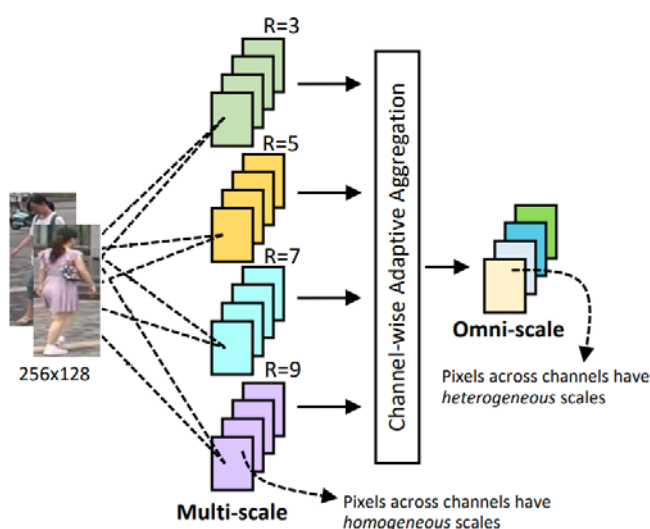


Рис. 2 – Архітектура мережі OSNet 1

об'єктів динамічно об'єднуються ваговими коефіцієнтами каналів, які генеруються уніфікованим шлюзом агрегації (Aggregation Gate).

Aggregation Gate (AG) — це міні-мережа, яка поділяє параметри для всіх потоків з низкою бажаних властивостей для ефективного навчання моделі. За допомогою AG, що піддається навчанню, генерується ваги каналу стають залежними від входу, отже, динамічне злиття масштабів. Цей новий дизайн AG дозволяє мережі вивчати всемасштабні представлення функцій: залежно від конкретного вхідного зображення, шлюз може сфокусуватися на одному масштабі, призначаючи домінуючу вагу певному потоку або масштабу; як альтернатива, він може підбирати та змішувати і таким чином створювати неоднорідні масштаби [5].

Таблиця 1 - Архітектура OSNet з розміром вхідного зображення 256×128

stage	output	OSNet
conv1	128×64, 64 64×32, 64	7×7 conv, stride 2 3×3 max pool, stride 2
conv2	64×32, 256	bottleneck × 2
transition	64×32, 256 32×16, 256	1×1 conv 2×2 average pool, stride 2
conv3	32×16, 384	bottleneck × 2
transition	32×16, 384 16×8, 384	1×1 conv 2×2 average pool, stride 2
conv4	16×8, 512	bottleneck × 2
conv5	16×8, 512	1×1 conv
gap	1×1, 512	global average pool
fc	1×1, 512	fc

Дослідження датасетів для навчання нейромережесвих моделей

Для дослідження було обрано відкриті набори даних для повторної ідентифікації людей.

Market-1501. Набір даних, що був створений з відео отриманого з шести камер перед супермаркетом у кампусі, у тому числі п'ять камер 1280 × 1080 HD та одна камера 720 × 576 SD. Серед цих камер є перекриття. Даний набір даних містить 32 668 зображень із 1501 особами. Зображення кожної особи фіксуються щонайбільше шістьма камерами. Кожна анотована особа фіксується принаймні двома камерами, щоб можна було здійснювати перехресний пошук [6]. Приклади набору даних показані на рисунку 3.

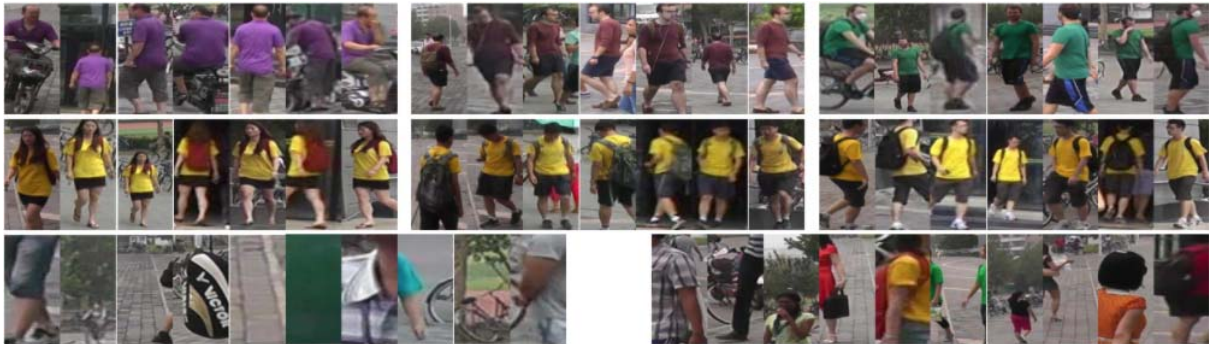


Рис. 3 - Приклади набору даних Market-1501

SUHK03. Набір даних складається з 14 097 зображень із 1 467 осіб, які було зібрано за допомогою камер в кампусі, і кожна особа фіксується двома камерами кампусу. Цей набір даних надає два типи анотацій, перший тип – це зображення розмічені вручну, а інший — розмічені за допомогою автоматичного детектора. Набір даних також містить 20 випадкових розподілів для навчання та тестування, у яких 100 осіб вибираються для тестування, а решта для навчання [7]. Приклади набору даних показані на рисунку 4.

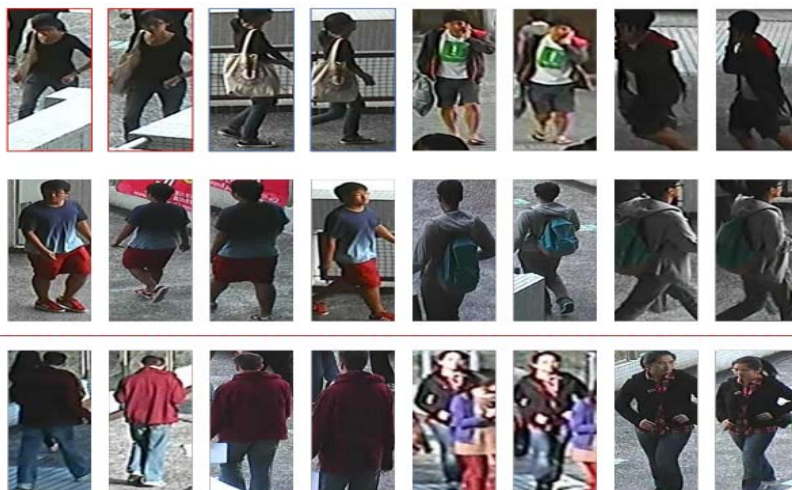


Рис. 4 - Приклади набору даних SUHK03

DukeMTMC-reID. Даний набір був зібраний в університеті Дюка і містить понад 14 годин відеопослідовностей, зібраних з 8 камер. Випадковим чином вибирається зображення з 702 особою як навчальний набір, а решту 702 як тестовий набір. У наборі для тестування вибирається одне зображення запиту для кожного особи в кожній камері, а інші зображення розміщуються в галереї. У результаті ми отримуємо 16 522 навчальні зображення з 702 особою, 2 228 зображень запиту інших 702 осіб та 17 661 зображення галереї [8]. Приклади набору даних показані на рисунку 5.



Рис. 5 - Приклади набору даних DukeMTMC-reID

Метрики оцінки

Для оцінки ефективності моделі в різних аспектах, які роблять оцінку комплексною використовують наступні метрики:

Cumulative Matching Characteristics (CMC) – це криві сукупних характеристик відповідності. На них побудований один з найпопулярніших підходів оцінки повторної ідентифікації особи. Це метод імовірності *top-k* пострілів для оцінки точності, в якому *top-k* відноситься до перших *k* вибірок з правилами ранжирування. Модель робить передбачення відповідно до зразка запиту та ранжирує їх за схожістю із запитом. Точність CMC *top-k* для запиту визначається як [9]:

$$Acc_k = \begin{cases} 1, & \text{якщо зразки галереї з найвищим рейтингом містять ідентифікатор запиту} \\ 0, & \text{інше} \end{cases}$$

яка є зсунутою ступінчастою функцією. Остаточна крива CMC обчислюється шляхом усереднення зміщених крокових функцій за всіма запитами.

Mean Average Precision (mAP) – широко використовується в задачах виявлення об'єктів і пошуку зображень, особливо там, де виникають завдання повторної ідентифікації особи [10].

Точність і відкликання є важливими компонентами середньої точності. Точність — це здатність моделі ідентифікувати лише відповідні об'єкти. Відкликання — це здатність моделі знаходити всі відповідні випадки. Точність і відкликання виражаються як:

$$Precision = \frac{TP}{TP+FP} \quad , \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad , \quad (2)$$

де *TP* - кількість істинно позитивних, *FP* - кількість хибнопозитивних, *FN* - кількість помилково негативних.

Середня точність (*AP*) означає середнє значення найвищої точності при різних відкликаннях, яке виражається таким чином:

$$AP = \frac{1}{R} \sum_{r \in R} Precision(r) \quad , \quad (3)$$

де *mAP* - це остаточний показник продуктивності моделі, оскільки він використовує середнє значення продуктивності моделі по всіх класах, в певному сенсі відображає середню продуктивність і визначається як [11]:

$$mAP = \frac{1}{M} \sum_{m \in M} AP(m) \quad . \quad (4)$$

Навчання та тестування нейромережесвих моделей

Для створення та навчання моделей було використано python бібліотеку глибокого навчання для повторної ідентифікації людей – Torchreid [12]. Основними особливостями цієї бібліотеки є:

- навчання при використанні декількох графічних процесорів;
- можливість використання попередньо навчених моделей (MuDeep, ResNet-mid, HACNN, PCB, MLFN, OSNet);
- аналіз якості моделей у режимі реального часу;
- підтримують як зображення, так і відео;
- проста підготовка наборів даних;
- навчання з кількома наборами даних;
- оцінка між наборами даних.

Навчання та тестування моделі відбувались на відкритих наборах даних Market-1501 та DukeMTMC-reID.

Для навчання моделей на основі моделі OSNet були обрані такі параметри: кількість епох навчання – 60, optim – adam [13], loss – softmax [14]. На рисунку 6 зображено, як змінювалось значення якості (acc) в процесі навчання.



Рис. 6. Зміна показників якості моделі в процесі навчання

Для тестування були обрані показники якості моделі – mAP та $СМС$ крива ($Rank-1$).

В результаті досліджень та навчання нейронних мереж було досягнуто таких показників якості моделі на валідаційному наборі зображень Market-1501: $mAP = 73,5\%$ та $Rank-1 = 90\%$. Отримані результати порівняємо з існуючою моделлю [15], яка була створена при використанні нейронної мережі OGNNet. Показники якості моделі на валідаційному наборі зображень Market-1501: $mAP = 69,52\%$ та $Rank-1 = 87,74\%$, що значно менші за показники навчаної власної моделі.

Результати досліджень та отриманих моделей на валідаційному наборі зображень Market-1501 наведено в таблиці 2.

В таблиці 3 представлено результати отримані на валідаційному наборі зображень DukeMTMC-reID.

Таблиця 2 - Результати навчання моделей на наборі зображень Market-1501.

model	dataset	optim	lr	epoch	mAP	Rank-1	Rank-5	Rank-10	Rank-20
OSNet	market-1501	adam	0.001	60	60.2%	82.6%	92.7%	95.0%	96.9%
OSNet	market-1501	adam	0.0015	60	67.9%	86.7%	95.0%	96.9%	98.0%
OSNet	market-1501	adam	0.002	60	68.1%	86.6%	94.5%	96.3%	97.8%
OSNet	market-1501	adam	0.003	60	70.5%	87.7%	95.0%	96.9%	97.8%
OSNet	market-1501	adam	0.0045	60	73.0%	89.4%	95.8%	97.3%	98.5%
OSNet	market-1501	adam	0.0055	60	73.5%	90.0%	96.1%	97.4%	98.2%
OSNet	market-1501	adam	0.007	60	72.2%	89.0%	96.0%	97.5%	98.4%
OSNet	market-1501	adam	0.0005	60	72.8%	89.0%	95.7%	97.2%	98.4%
OSNet	market-1501	adam	0.0085	60	71.7%	89.0%	95.3%	96.8%	98.0%

Таблиця 3 - Результати навчання моделей на наборі зображень DukeMTMC-reID.

model	dataset	optim	lr	epoch	mAP	Rank-1	Rank-5	Rank-10	Rank-20
OSNet	dukemtmc-reid	adam	0.001	60	51.7%	74.5%	85.6%	88.9%	91.4%
OSNet	dukemtmc-reid	adam	0.0015	60	56.0%	77.7%	86.8%	89.9%	91.9%
OSNet	dukemtmc-reid	adam	0.002	60	57.9%	78.0%	87.0%	90.4%	92.9%
OSNet	dukemtmc-reid	adam	0.005	60	62.1%	81.7%	89.5%	91.9%	93.7%
OSNet	dukemtmc-reid	adam	0.0055	60	33.6%	55.0%	72.6%	77.9%	82.7%

ВИСНОВКИ

1. Проведено аналіз нейронної мережі OSNet, визначено основні блоки та архітектуру мережі. Розглянуто основні метрики, що використовуються для оцінки кількісної ефективності моделей.
2. Проаналізовано існуючі відкриті набори даних для задачі повторної ідентифікації людини та обрано Market-1501 та DukeMTMC-reID для використання в навчання нейромережних моделей.
3. За допомогою мови програмування Python та бібліотеки глибокого навчання torchreid, було проведено дослідження та отримано власні нейромережні моделі. Отримані моделі було перевірено на валідаційних наборах даних Market-1501 ($mAP = 73,5\%$, $Rank-1 = 90\%$) та DukeMTMC-reID ($mAP = 62,81\%$, $Rank-1 = 81,7\%$).
4. Отриману модель було порівняно з існуючими аналогами та визначено шляхи та чинники, що впливають на підвищення вірогідності повторної ідентифікації людини. Проведено тестування та оцінку якості отриманих моделей на реальних даних.

ЛІТЕРАТУРА

1. O. Kyrylenko, R. Maslii, and Y. Marushchak «Analysis of methods of person reidentification in multi camera environment», Norwegian Journal of development of the International Science, №47, pp. 46-48, 2020.
2. О.М. Кириленко, Р.Н. Кветний, та Р.В. Маслій, «Дослідження атрибутів людини для задачі повторної ідентифікації», Інформаційні технології та комп'ютерна інженерія, вип. 49, вип. 3, с. 4–13, 2020.
3. H. Yu and W. Zheng "Weakly supervised discriminative feature learning with state information for person identification" Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 2020.
4. K. Zhou, Yo. Yang, A. Cavallaro, Tao Xiang "Omni-Scale Feature Learning for Person Re-Identification", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3702-3712.
5. K. Zhou, Yo. Yang, A. Cavallaro, Tao Xiang "Learning Generalisable Omni-Scale Representations for Person Re-Identification", Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
6. L. Zheng L. Shen L. Tian S. Wang J. Wang and Q. Tian "Scalable person re-identification: A benchmark" Proc. IEEE Int. Conf. Comput. Vis. pp. 1116-1124 Jun. 2015.
7. Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person reidentification. In CVPR, 2014.
8. Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In ECCVW, 2016.
9. Evaluation Metrics. Cumulative Matching Characteristics. [Електронний ресурс] – Режим доступу: https://cysu.github.io/open-reid/notes/evaluation_metrics.html
10. О.В. Бубенщиков, Є.В. Лепа. "Використання згортальних нейронних мереж для ідентифікації обличчя людини" Вестник Херсонського національного технічного університета, №1 (68), 2019, pp. 136-142.
11. Wu, S.; Gao, L. Multi-Level Joint Feature Learning for Person Re-Identification. Algorithms 2020, 13, 111. <https://doi.org/10.3390/a13050111>
12. K. Zhou and T. Xiang "Torchreid: A Library for Deep Learning Person Re-Identification in Pytorch".
13. Zijun Zhang et al. Normalized direction-preserving Adam. arXiv:1709.04546v2, 2017
14. The Softmax function and its derivative. [Електронний ресурс] – Режим доступу: <https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative/>
15. Z. Zhedong, N. Zheng and Yi Yang. "Parameter-efficient person re-identification in the 3d space." arXiv preprint arXiv:2006.04569, 2020.

REFERENCES

1. O. Kyrylenko, R. Maslii, and Y. Marushchak «Analysis of methods of person reidentification in multi camera environment», Norwegian Journal of development of the International Science, №47, pp. 46-48, 2020.
2. O. Kyrylenko, R. Kvyetnyy and R. Maslii, "Research of human attributes for the problem of re-identification", Information Technology and Computer Engineering, vol. 49, issue 3, p. 4–13, 2020.
3. H. Yu and W. Zheng "Weakly supervised discriminative feature learning with state information for person identification" Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 2020.
4. K. Zhou, Yo. Yang, A. Cavallaro, Tao Xiang "Omni-Scale Feature Learning for Person Re-Identification", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3702-3712.
5. K. Zhou, Yo. Yang, A. Cavallaro, Tao Xiang "Learning Generalisable Omni-Scale Representations for Person Re-Identification", Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
6. L. Zheng L. Shen L. Tian S. Wang J. Wang and Q. Tian "Scalable person re-identification: A benchmark" Proc. IEEE Int. Conf. Comput. Vis. pp. 1116-1124 Jun. 2015.
7. Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person reidentification. In CVPR, 2014.

8. Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In ECCVW, 2016.
9. Evaluation Metrics. Cumulative Matching Characteristics. [Electronic resource] – Access mode: https://cysu.github.io/open-reid/notes/evaluation_metrics.html
10. O. Bubenshchikov, E. Lepa. "The use of convolutional neural networks to identify a person" Bulletin of the Kherson National Technical University, №1 (68), 2019, p. 136-142.
11. Wu, S.; Gao, L. Multi-Level Joint Feature Learning for Person Re-Identification. Algorithms 2020, 13, 111. <https://doi.org/10.3390/a13050111>
12. K. Zhou and T. Xiang "Torchreid: A Library for Deep Learning Person Re-Identification in Pytorch".
13. Zijun Zhang et al. Normalized direction-preserving Adam. arXiv:1709.04546v2, 2017
14. The Softmax function and its derivative. [Electronic resource] – Access mode: <https://eli.thegreenplace.net/2016/the-softmax-function-and-its-derivative/>
15. Z. Zhedong, N. Zheng and Yi Yang. "Parameter-efficient person re-identification in the 3d space." arXiv preprint arXiv:2006.04569, 2020.

КИРИЛЕНКО ОЛЕКСАНДР МИХАЙЛОВИЧ – аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, Вінниця, Україна, email: sasha.kyrylenko@gmail.com.