

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І.І.МЕЧНИКОВА

(повне найменування вищого навчального закладу)

Факультет математики, фізики та інформаційних технологій

(повне найменування інституту, назва факультету (відділення))

Кафедра математичного забезпечення комп'ютерних систем

(повна назва кафедри (предметної, циклової комісії))

Кваліфікаційна робота

на здобуття ступеня вищої освіти «магістр»

(освітньо-кваліфікаційний рівень)

«Інформаційна технологія пошуку об'єктів відеопослідовності.

Технологія розпізнавання об'єктів у стиснутій відеопослідовності»

«Information technology of searching objects in video sequence.

Object recognition technology in a compressed video sequence»

Виконала: студентка денної форми навчання

напряму підготовки 126 – Інформаційні системи та технології

(шифр і назва напряму підготовки, спеціальності)

Освітня програма «Інформаційні системи та технології»

(назва освітньої програми)

Якушина Анастасія Олексіївна

(прізвище, ім'я, по-батькові)

Керівник к.ф.-м.н., доц. Шпінарева І.М.

(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рецензент к.ф.-м.н., доц. Петрушина Т.І.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рецензент к.техн.н, доц. Рудніченко М.Д.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:

Захищено на засіданні ЕК № ____

Протокол засідання кафедри

протокол № __ від «__» ____ 2022 р.

№ __ від «__» ____ 2022 р.

Оцінка ____ / ____ / ____

(за національною шкалою, шкалою ECTS, бали)

Завідувач кафедри

Голова ЕК

Євгеній МАЛАХОВ

(підпис)

(ім'я, прізвище)

Володимир ВИЧУЖАНІН

(підпис)

(ім'я, прізвище)

Одеса - 2022

АНОТАЦІЯ

У дипломній роботі розробляється тема «Технологія розпізнавання об'єктів у стиснутій відеопослідовності».

Технологія виявлення та відстеження облич є одним із важливих напрямків досліджень у задачах відстеження. Вона широко використовується в різних додатках у повсякденному житті, таких як відеоспостереження, взаємодія людини з комп'ютером, автоматичне супроводження цілей, безпечне водіння та медична діагностика. Більшість існуючих методів відстеження обличчя виявляють зображення обличчя в кожному кадрі відеопослідовності та реалізують динамічне відстеження обличчя.

Метою роботи є підвищення ефективності виявлення облич у стиснутій відеопослідовності шляхом розробки методу машинного навчання. Під ефективністю розуміється покращення точності і швидкості виконання виявлення облич.

В результаті проведених в роботі досліджень розглянуто методи виявлення об'єктів у відеопослідовностях, запропоновано метод виявлення облич з використанням переваг стиснутих відеопослідовностей. Використання векторів руху з неключових кадрів дозволило значно зменшити кількість викликів до модуля виявлення облич, який всередині використовує нейронну мережу SSD з вже попередньо навченими ваговими коефіцієнтами на наборі даних WIDER FACE. За допомогою інформації про вектори руху є можливість зміщати попередньо знайдені обмежувальні рамки на величину векторів руху.

У роботі сформувані вимоги до розробленої системи, спроектовано та реалізовано систему виявлення облич.

Шляхом модифікацій обраного методу виявлення облич в стиснутій відеопослідовності було досягнуто найбільшого значення AP – 82.5% з урахуванням зменшення частоти виклику модуля виявлення облич в 10 разів в порівнянні з традиційними методами виявлення об'єктів, які застосовують покадрове виявлення об'єктів.

ABSTRACT

In this thesis, the topic «Object recognition technology in a compressed video sequence» is designed.

Face detection and tracking technology is one of the important areas of research in tracking tasks. It is widely used in various applications in daily life, such as video surveillance, human-computer interaction, automatic target tracking, safe driving, and medical diagnosis. Most existing face tracking methods detect face images in each frame of a video sequence and implement dynamic face tracking.

The goal of the work is to improve the efficiency of face detection in a compressed video sequence by developing a machine learning method. Efficiency means improving the accuracy and speed of face detection.

As a result of the research carried out in the work, the methods of object detection in video sequences were considered, a face detection method using the advantages of compressed video sequences was proposed. The use of motion vectors from non-keyframes allowed to significantly reduce the number of calls to the face detection module, which internally uses an SSD neural network with already pre-trained weights on the WIDER FACE dataset. With the help of information about motion vectors, it is possible to shift previously found bounding frames by the amount of motion vectors.

In the work, the requirements for the developed system were formed, a face detection system was designed and implemented.

By modifying the selected face detection method in a compressed video sequence, the highest AP value was achieved – 82.5%, taking into account a 10-fold reduction in the call frequency of the face detection module compared to traditional object detection methods that use frame-by-frame object detection.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ І ТЕРМІНІВ.....	5
ВСТУП	6
1 ОГЛЯД МЕТОДІВ ВИЯВЛЕННЯ ОБ'ЄКТІВ У ВІДЕОПОСЛІДОВНОСТІ	8
1.1 Огляд методів виявлення об'єктів у відеопослідовності	8
1.1.1 Методи виявлення об'єктів на основі оптичного потоку	9
1.1.2 Методи виявлення об'єктів на основі LSTM	12
1.1.3 Методи виявлення об'єктів на основі механізму уваги	14
1.1.4 Методи виявлення об'єктів на основі відстеження	18
1.2 Порівняння методів виявлення об'єктів у відеопослідовності	20
1.3 Висновки та уточнення задач	22
2 ПРОЕКТУВАННЯ СИСТЕМИ ВИЯВЛЕННЯ ОБЛИЧ В СТИСНУТІЙ ВІДЕОПОСЛІДОВНОСТІ.....	23
2.1 Архітектура системи	23
2.2 Модуль захоплення та обробки відео	24
2.3 Модуль виявлення об'єктів	26
2.4 Модуль відстеження об'єктів	29
2.5 Набір даних	32
2.6 Засоби реалізації.....	33
3 РЕАЛІЗАЦІЯ СИСТЕМИ ВИЯВЛЕННЯ ОБЛИЧ В СТИСНУТІЙ ВІДЕОПОСЛІДОВНОСТІ.....	35
3.1 Модуль захоплення та обробки відео	35
3.2 Модуль виявлення об'єктів	36
3.3 Модуль відстеження об'єктів	40
3.4 Модуль збереження даних	46
3.5 Інструкція користувача.....	48
4 АНАЛІЗ РЕЗУЛЬТАТІВ.....	51
ВИСНОВКИ.....	56
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	58

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ І ТЕРМІНІВ

AP – Average Precision

CDT – Combining Detection And Tracking

DFF – Deep Feature Flow

FGFA – Flow-Guided Feature Aggregation

IoU – Intersection over Union

LSTM – Long Short Term Memory

mAP – Mean Average Precision

MEGA – Memory Enhanced Global-Local Aggregation

OGEMN – Object-Guided External Memory Network

PSLA – Progressive Parse Local Attention

RDN – Relation Distillation Networks

SELSA – Sequence Level Semantics Aggregation

SSD – Single-Shot Multibox Detection

ВСТУП

Виявлення об'єктів є фундаментальною проблемою комп'ютерного зору і має широкий спектр застосувань. Виявлення об'єктів у нерухомих зображеннях досягло значного прогресу на основі глибокого навчання. Однак у випадку виявлення відеооб'єктів, безпосереднє застосування детекторів об'єктів нерухомого зображення до кожного окремого відеокадру спричиняє надлишкові обчислювальні витрати та вимагає постобробки для отримання тимчасового контексту для більшої точності виявлення.

Виявлення обличчя є критично важливим кроком для всіх алгоритмів аналізу обличчя, включаючи вирівнювання обличчя, розпізнавання обличчя, верифікація обличчя і розбір обличчя. Хоча для людини це здається легким завданням, для комп'ютерів це дуже складне завдання. Проблеми, пов'язані з розпізнаванням обличчя, можна віднести до варіацій пози, масштабу, виразу обличчя, оклюзії та умов освітлення.

Обличчя людини є однією з природних структурних цілей із більш складними та детальними змінами, а виявлення обличчя є складною темою дослідження. Оскільки різні люди мають різні біологічні характеристики, такі як зовнішність, міміка, колір шкіри, то й саме людське обличчя має мінливість. Крім того, наявність допоміжних елементів, таких як людське волосся, очі та макіяж обличчя, ускладнює виявлення обличчя. Чіткість зображень обличчя, напрям і інтенсивність джерел світла також впливають на вираз обличчя. Тому дуже важливо забезпечити ефективні методи виявлення обличчя та відстеження у реальних відеопослідовностях із складними сценами.

Переваги стиснутого формату відеопослідовності забезпечують надзвичайно надійне рішення щодо виявлення обличчя шляхом поєднання методів із існуючих покадрових алгоритмів виявлення об'єктів на основі піксельної обробки та використання інформації з неключових кадрів для задач відстеження. Обчислювальна складність цього рішення повинна бути надзвичайно низька, так як для отримання векторів руху не потрібно задіяння

додаткових методів інтелектуального аналізу. Їх можна отримувати напряму з кодованого кадру, тому запропонованого рішення повинно бути достатньо для реалізації та використання в стаціонарних і навіть мобільних споживчих електронних пристроях з невеликими обчислювальними ресурсами.

Метою роботи є підвищення ефективності виявлення облич у стиснутій відеопослідовності шляхом розробки методу машинного навчання. Під ефективністю розуміється покращення точності і швидкості виконання виявлення облич.

Об'єктом дослідження є процес розпізнавання облич у стиснутій відеопослідовності.

Предметом дослідження є методи та алгоритми інтелектуального аналізу при розпізнаванні облич у стиснутій відеопослідовності.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- дослідження методів виявлення об'єктів у відеопослідовності;
- проектування системи для виявлення облич у стиснутій відеопослідовності;
- реалізація обраного методу виявлення облич та його модифікації;
- реалізація системи виявлення облич;
- аналіз модифікованого алгоритму.

ВИСНОВКИ

В ході виконання дипломної роботи проведено огляд та аналіз існуючих методів виявлення об'єктів у відеопослідовності, розглянуті переваги та недоліки цих методів та визначено найбільш перспективні методи для розпізнавання облич у відеопослідовності, запропоновано використання переваг стиснутої відеопослідовності задля зменшення кількості обчислювальних ресурсів. Зроблено висновок про доцільність виконання даного проекту з метою підвищення ефективності виявлення облич у стиснутій відеопослідовності. Під ефективністю розуміється покращення точності і швидкості виконання виявлення облич.

Система складається з модуля захоплення і обробки відеопослідовності, модуля виявлення об'єктів та модуля відстеження об'єктів, а також модуля збереження даних.

Також розглянуто загальну архітектуру згорткових нейронних мереж та запропоновано використання методу розпізнавання облич в відеопослідовності на основі каскадної структури нейронної мережі SSD з попередньо підготовленими ваговими коефіцієнтами на наборі даних WIDER FACE. Обрана нейронна мережа виявляє ознаки об'єкта на декодованому кадрі, як це відбувалося і в традиційних методах виявлення об'єктів з використанням згорткових нейронних мереж.

Далі в модулі відстеженні зіставляються відстежувані об'єкти з нововиявленими. Зіставлення виконується за допомогою угорського алгоритму, застосованого до матриці IoU усіх можливих пар рамок. Якщо ж ще не пройшов зазначений інтервал з останнього виявлення об'єктів, відбувається зміщення вже відстежуваних рамок за допомогою інформації о векторах руху з неключових кадрів. Таким чином отримується інформація о зміщенні об'єктів, але без використання згорткових мереж на кожному з кадрів, що істотно зменшує кількість обчислювальних витрат.

Запропонована модель виявлення облич має дві основні переваги:

1) значно прискорює процедуру отримання ознак для стиснутої відеопослідовності: повна мережу розпізнавання використовується з певним інтервалом між кадрами, а для всіх інших кадрів задіяний механізм відстеження на основі інформації про вектори руху;

2) на відміну від існуючих методів, що створюють додаткову мережу для моделювання руху кадрів, використано всі переваги векторів руху вільно доступних у стиснутих відеопослідовностях.

Шляхом модифікацій обраного методу виявлення облич в стиснутій відеопослідовності було досягнуто найбільшого значення AP – 82.5% з урахуванням зменшення частоти виклику модуля виявлення облич в 10 разів в порівнянні з традиційними методами виявлення об'єктів, які застосовують покадрове виявлення об'єктів.

Подальша робота в цьому напрямку має вестися в декількох напрямках. По-перше, це використання розробленої системи виявлення облич в режимі реального часу. Задля цього також можна додати можливість паралельної обробки декілька відеопотоків одночасно. По-друге, якісне доопрацювання – наприклад, можливість виявлення нових об'єктів за допомогою інформації про вектори руху з використанням легких нейронних мереж. З точки зору інтерфейсу, можливо розробити графічну обертку над побудованою системою.

За результатами даної роботи опубліковано тези на всеукраїнській конференції [29] та в спільній україно-китайській науковій студентській конференції [30].

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ghosh, A., Subudhi, B.N., Ghosh, S. Object Detection from Videos Captured by Moving Camera by Fuzzy Edge Incorporated Markov Random Field and Local Histogram Matching. *IEEE Trans. Circuits Syst. Video Technol.* 2012, 22, 1127–1135.
2. Zhu, X., Xiong, Y., Dai, J., Yuan, L., Wei, Y. Deep Feature Flow for Video Recognition. In *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, 22–29 October 2017*, pp. 4141–4150.
3. Zhu, X., Wang, Y., Dai, J., Yuan, L., Wei, Y. Flow-Guided Feature Aggregation for Video Object Detection. In *Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017*, pp. 408–417.
4. Hetang, C., Qin, H., Liu, S., Yan, J. Impression Network for Video Object Detection. *arXiv 2017*, arXiv:1712.05896
5. Zhu, X., Dai, J., Yuan, L., Wei, Y. Towards High Performance Video Object Detection. In *Proceedings of the 2018 IEEE/Cvf Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*, pp. 7210–7218.
6. Zhu, X., Dai, J., Zhu, X., Wie, Y., Yuan, L. Towards High Performance Video Object Detection for Mobiles. *arXiv 2018*, arXiv:1804.05830.
7. Liu, M., Zhu, M., White, M., Li, Y., Kalenichenko, D. Looking Fast and Slow: Memory-Guided Mobile Video Object Detection. *arXiv 2019*, arXiv:1903.10172.
8. Deng, J., Pan, Y., Yao, T., Zhou, W., Li, H., Mei, T. Relation Distillation Networks for Video Object Detection. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019*, pp. 7022–7031.
9. Wu, H., Chen, Y., Wang, N., Zhang, Z. Sequence Level Semantics Aggregation for Video Object Detection. In *Proceedings of the 2019 IEEE/CVF International*

- Conference on Computer Vision (ICCV), Gangnam-gu, Seoul, Korea, 27 October–2 November 2019.
10. Deng, H., Hua, Y., Song, T., Zhang, Z., Xue, Z., Ma, R., Guan, H. Object Guided External Memory Network for Video Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019, pp. 6677–6686.
 11. Chen, Y., Cao, Y., Hu, H., Wang, L. Memory Enhanced Global-Local Aggregation for Video Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–18 June 2020.
 12. Guo, C., Fan, B., Gu, J., Zhang, Q., Xiang, S., Prinet, V., Pan, C. Progressive Sparse Local Attention for Video object detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019.
 13. Kim, H.U., Kim, C.S. CDT: Cooperative Detection and Tracking for Tracing Multiple Objects in Video Sequences. In *Computer Vision—Eccv 2016, Part VI*, Leibe, B., Matas, J., Sebe, N., Welling, M., Eds., Springer: Cham, Switzerland, 2016, pp. 851–867.
 14. Mao, H., Kong, T., Dally, W.J. CaTDet: Cascaded Tracked Detector for Efficient Object Detection from Video. arXiv 2018, arXiv:1810.00434.
 15. Feichtenhofer, C., Pinz, A., Zisserman, A. Detect to Track and Track to Detect. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017, pp. 3057–3065.
 16. Battista, S., Casalino, F., Lande, C.. MPEG-4: a multimedia standard for the third millennium, Part 2. *Multimedia*, IEEE. 7. 2000. pp. 76 - 84. 10.1109/93.839314.
 17. Chen, J., Kao, C., Lin, Y. Introduction to H.264 advanced video coding. 2006. 10.1109/ASPDAC.2006.1594774.
 18. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Cheng-Yang F., Berg A. SSD: Single Shot MultiBox Detector. arXiv 2016, arXiv:1512.02325.

19. Shuo Y., Ping L., Chen C., Xiaoou T. WIDER FACE: A Face Detection Benchmark. arXiv 2015, arXiv: 1511.06523.
20. WIDER FACE: A Face Detection Benchmark [Електронний ресурс] - Режим доступу: <http://shuoyang1213.me/WIDERFACE/>
21. YouTube Faces Database Benchmark [Електронний ресурс] – Режим доступу: <https://www.cs.tau.ac.il/~wolf/ytfaces/>
22. TensorFlow. [Електронний ресурс] – Режим доступу: https://www.tensorflow.org/api_docs/python/tf
23. SciPy. [Електронний ресурс] – Режим доступу: <https://docs.scipy.org/doc/scipy/reference/index.html#scipy-api>
24. NumPy. [Електронний ресурс] – Режим доступу: <https://numpy.org/doc/stable/>
25. OpenCV. [Електронний ресурс] - Режим доступу: <https://opencv.org/>
26. MV-extractor. [Електронний ресурс] – Режим доступу: <https://libraries.io/pypi/mv-extractor>
27. Amazon S3. [Електронний ресурс] – Режим доступу: <https://aws.amazon.com/s3/>
28. Boto3. [Електронний ресурс] – Режим доступу: <https://boto3.amazonaws.com/v1/documentation/api/latest/index.html>
29. Якушина А.О., Шпінарева І.М. Перспективи використання методів глибокого навчання в задачах виявлення об'єктів в стиснутій відеопослідовності/ Інформатика, інформаційні системи та технології: тези доповідей дев'ятнадцятої всеукраїнської конференції студентів і молодих науковців. Одеса, 29 квітня 2022 р. – Одеса, 2022. – с. 81-82.
30. Shlemko O., Yakushyna A. Information technologies for object recognition and classification in video sequence / Conference proceedings of the 1st Student Scientific Conference of Joint Research Cooperation between Odesa I.I. Mechnikov National University and Huaiyin Institute of Technology (March 28-29, May 16, 2022) / MOES of Ukraine; Odesa I.I. Mechnikov National University; – Dnipro: Serednyak TK, 2022. – p.136-139.