DOI 10.31891/2307-5732-2024-341-5-55
УДК 044

ПІЦУН ОЛЕГ
Західноукраїнський національний університет
https://orcid.org/0000-0003-0280-8786
e-mail: o.pitsun@wunu.edu.ua
МЕЛЬНИК НАЗАР
Західноукраїнський національний університет
https://orcid.org/0009-0000-5917-1099
e-mail: 88nazar88@gmail.com
ЛІП'ЯНІНА-ГОНЧАРЕНКО ХРИСТИНА
Західноукраїнський національний університет
https://orcid.org/0000-0002-2441-6292
e-mail: kh.lipianina@wunu.edu.ua

ГІБРИДНИЙ ПІДХІД ДО ВИЗНАЧЕННЯ ДІПФЕЙКІВ НА ОСНОВІ ЛЮДСЬКОГО ОБЛИЧЧЯ

Генеративний штучний інтелект відіграє дедалі важливішу роль у сучасному житті, відкриваючи нові можливості в різних сферах, від медіа до технологій безпеки. Ця технологія активно використовується у графіці, дизайні та виробництві нових графічних об'єктів, що не лише розширює горизонти креативності, але й підвищує ефективність роботи професіоналів. Однак, зростання популярності генеративного інтелекту також породжує серйозні виклики, зокрема у вигляді діпфейків — відео та зображень, які маніпулюють реальністю, зображуючи людей у неправдивому контексті.

У даній роботі детально аналізуються сучасні підходи, інструменти та датасети, що використовуються для розробки діпфейків. Завдяки цьому аналізу, вдається виявити основні тенденції в їх розпізнаванні, які є критично важливими для розуміння способів захисту від дезінформації. Запропоновано узагальнений підхід до розпізнавання діпфейків на основі алгоритмів комп'ютерного зору та архітектур згорткових нейронних мереж. Цей комплексний підхід інтегрує різноманітні методи, що дозволяє ефективно детектувати елементи, які характеризують фейки, зокрема у контексті людських облич.

Крім того, робота обговорює виклики, що постають перед сучасними системами розпізнавання, такі як адаптація до нових технік створення діпфейків та забезпечення високої точності розпізнавання в умовах постійної еволюції технологій. Важливим аспектом є вивчення етичних та правових аспектів використання діпфейків, що підкреслює необхідність розвитку відповідних норм та стандартів у цій динамічно змінній сфері. Окрім технічних аспектів, увага приділяється й соціальним наслідкам діпфейків, що включають можливість маніпуляцій з громадською думкою та вплив на особисту безпеку.

Ключові слова: генеративний штучний інтелект, графічний об'єкт, діпфейк, датасет, відео, людське обличчя, машинне навчання, комп'ютерний зір.

PITSUN OLEH, MELNYK NAZAR
West Ukrainian National University
LIPIANINA-HONCHARENKO KHRYSTYNA
West Ukrainian National University

A HYBRID APPROACH TO THE DETECTION OF DEEPFAKES BASED ON THE HUMAN FACE

Generative artificial intelligence plays an increasingly significant role in contemporary life, unlocking new possibilities across various fields, from media to security technologies. This technology is actively utilized in graphics, design, and the creation of new graphical objects, not only expanding the horizons of creativity but also enhancing professionals' efficiency. However, the growing popularity of generative intelligence also poses serious challenges, notably in the form of deepfakes—videos and images that manipulate reality by depicting individuals in a false context.

This paper provides a detailed analysis of current approaches, tools, and datasets used in the development of deepfakes. Through this analysis, it identifies key trends in their recognition, which is critical for understanding ways to combat misinformation. A generalized approach to deepfake recognition is proposed, based on computer vision algorithms and convolutional neural network architectures. This comprehensive approach integrates various methods, enabling effective detection of elements characterizing fakes, particularly in the context of human faces.

Furthermore, the paper discusses the challenges faced by modern recognition systems, such as adapting to new deepfake creation techniques and ensuring high detection accuracy amid the continuous evolution of technologies. An important aspect is the examination of the ethical and legal dimensions of deepfake usage, emphasizing the necessity for the development of relevant norms and standards in this dynamically changing field. In addition to technical aspects, the social implications of deepfakes are addressed, including the potential for manipulation of public opinion and impacts on personal security.

Keywords: generative artificial intelligence, graphical object, deepfake, dataset, video, human face, machine learning, computer vision.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

З розвитком штучного генеративного інтелекту все частіше виникає необхідність у розподіленні корисного контенту від фейкового. Фейкові дані використовуються не лише у текстовій формі, але й у вигляді фото або відео. Діпфейки можуть бути використані для шахрайства, наприклад, для видавання себе за іншу особу в онлайн-спілкуванні. Діпфейки можуть бути використані для поширення неправдивої інформації або пропаганди. Діпфейки можуть використовуватися для створення фальшивих відео або зображень з метою дискредитації або компрометації людей.

Будь хто має можливість згенерувати фейкові зображень облич використовуючи інструменти

DeepFake та Face2Face Для реалізації задач створення фейкових облич зазвичай використовують технології глибокого машинного навчання та згорткові нейронні мереж адже саме ці технології використовуються для роботи із зображеннями та відео. До найросповсюдженіших архітектур згорткових нейронних мереж відносять AlexNet, LeNet, MobileNet, Resnet, VGG19.

При створенні фейкових облич зазвичай використовують різні способи маніпуляції над обєктом, зокрема зміна розміру, застосування фільтрів різного роду, контрастування, зміна рівнів яскравості тощо.

У більшості випадків використовують підхід пов'язаний із маніпулюванням виразу обличчя та перетворенням особи.

Незважаючи на розвиток генеративного інтелекту, люди можуть використовувати його не лише для позитивних намір, але й для недобрих. Тому актуальною є необхідність у пошуку механізмів визначення реальності того чи іншого фото чи відео.

Глибинне навчання часто використовують для створення фейків та фото та відео.

Один з основних підходів до аналізу зображення є визначення того, чи піддавалось зображення обробці, наприклад застосування фільтрів, контрастності, зміну розміру конвертування тощо.

Найпопулярніша функція методом вилучення є просторово насичена модель (SRM) [1] спочатку запропонував виділити стеганалітичні ознаки. Архітектура згорткової нейронної мережі MISLNet включає обмежену згортку, адаптивне вивчення функцій трасування маніпуляцій за допомогою зворотного поширення.

Серед найпоширеніших інструментів для маніпулювання людським обличчям є DeepFakes, Face2Face, FaceSwap. Зміна обличчя часто використовується для створення діпфейків завдяки розвитку технологій компютерного зору в галузі виділення елементів людського зображення.

Виділяють два підходи до створення фейків:

– маніпуляція виразом обличчя

– маніпуляція ідентифікацією обличчя

Виявлення та запобігання діпфейків є важливим аспектом у забезпеченні безпеки, правопорядку та захисту прав людини.

Аналіз досліджень та публікацій

Модель згорткової нейронної мережі, яка використовує зміни в виразах обличчя запропоновано у робіт [2]. Також автори пропонують фреймворк для виділення елементів на зображенні, що характрені для фейкових зображень. У роботі [3] наведено підходи до визначення фейків на відуо та зображеннях базуючих на розміщенні обличчя та інших показниках. Розробку назвали FakeCatcher, що базується на визначенні біологічних сигналів. У роботі [4] наведено порівняльний аналіз існуючих підходів до розпізнавання фейків та запропоновано мережу MesoNet, що використовує згорткові нейронні мережі. У роботі [6] проведено комплексний огляд технологій створення та виявлення дипфейків з використанням підходів глибокого навчання. Даний аналіз дозволяє отримати порівняльний аналіз та детальний опис останніх методів і наборів даних.

При вирішенні задачі визначенні діпфейків варто звернути увагу на сучасні підходи у обчисленнях великого набору даних та загальні підходи до обчислень. У роботі [8] наведено аналіз засобів обчислювального інтелекту. В процесі аналізу обличчя виникає необхідність у пошуку певних регіонів, наприклад очей, губ тощо. Сучасні підходи до сегментації зображень розглянуто у роботі [9].

Формулювання цілей статті

Метою роботи є: розпізнавання діпфейків на основі алгоритмів комп’ютерного зору та архітектур згорткових нейронних мереж

Для досягнення мети необхідно виконати такі задачі:

– провести аналіз існуючих підходів, інструментів, датасетів для визначення сучасних тенденцій у розробці систем для визначення діпфейків на основі аналізу людського обличчя.

– розробити узагальнений підхід до класифікації діпфейків з використанням як алгоритмів компютерного зору так і з використанням згорткових нейронних мереж, навчених на основі відомих датасетів.

– розробити структуру terraform – файлу для можливості розгортання проекту на хмарного середовищі з врахуванням необхідності зберігання датасету.

Виклад основного матеріалу

Для визначення діпфейків потрібно звернути увагу на такі фактори:

– занадто гладка шкіра на зображенні, деякі елементи обличчя , наприклад очі можуть бути старішими

– DeepFakes може не повністю відобразити природну фізику сцени. Наприклад діпфейк може некоретно відобразити тіні. В цьому випадку варто звернути увагу на очі та брови

– DeepFakes може не повністю відобразити природну фізику освітлення. В даному випадку варто звернути увагу на окуляри та чи змінюється кут відблиску при русі людини

– Варто звернути увагу на реалістичність волосся на обличчі.

– Варто звернути увагу на родимки

– Варто звернути увагу на рух губ, чи вони є реальними природніми чи надто синхронізованими [5]

Для створення діпфейків наразі найчастіше використовують генеративний штучний інтелект, зокрема GAN мережі. Однак для визначення діпфейків використовують засоби комп’ютерного зору та згорткові нейронні мережі, інколи і інші типи нейронних мереж.

Один з підходів, це визначення статистичних характеристик зображення на основі використання

малгоритмів комп'ютерного зору. Деякі з підході в розглянуто у [7].

Інший підхід полягає у розробці нових архітектур згорткових нейронних мереж для можливості якісної класифікації реальних та фейкових.

Одним із важливих підходів до визначення фейків є аналіз кліпання очима людини на відео.

Окрім того, варто звернути увагу на наявність розмитого простору навколо зображення голови.

До основних датасетів для визначення фейків можна віднести такі:

Flickr-Faces-HQ (FFHQ) – вміщує 70,000 зображень в хорошій якості та розмір 1024 на 1024 пікселі. Особливістю даної мережі є наявність фото з різними етнічними та віковими характеристиками, що дозволяє зробити нейронну мережі більш універсальнішою.

100K-Faces - вміщує 100,000 зображень. Особливість полягає у тому, що зображення згенеровані з допомогою нейронної мережі StyleGAN.

mEBAL – датасет вміщує матеріали, для можливості визначення кількості кліпань очима. Зазвичай цей датасет використовують у задачах attention level estimation, analysis of neuro-degenerative diseases, deception recognition, drive fatigue detection

VGGFace2 одна із найпотужніших баз, що вміщує більше 3 мільйонів зображень. ЇЇ особливість полягає у тому, що вона вміщує зображення відомих людей. Політиків, акторів тощо.

На рисунку 1 наведено узагальнений підхід до визначення діпфейків на основі алгоритмів комп'ютерного зору та архітектур згорткових нейронних мереж.

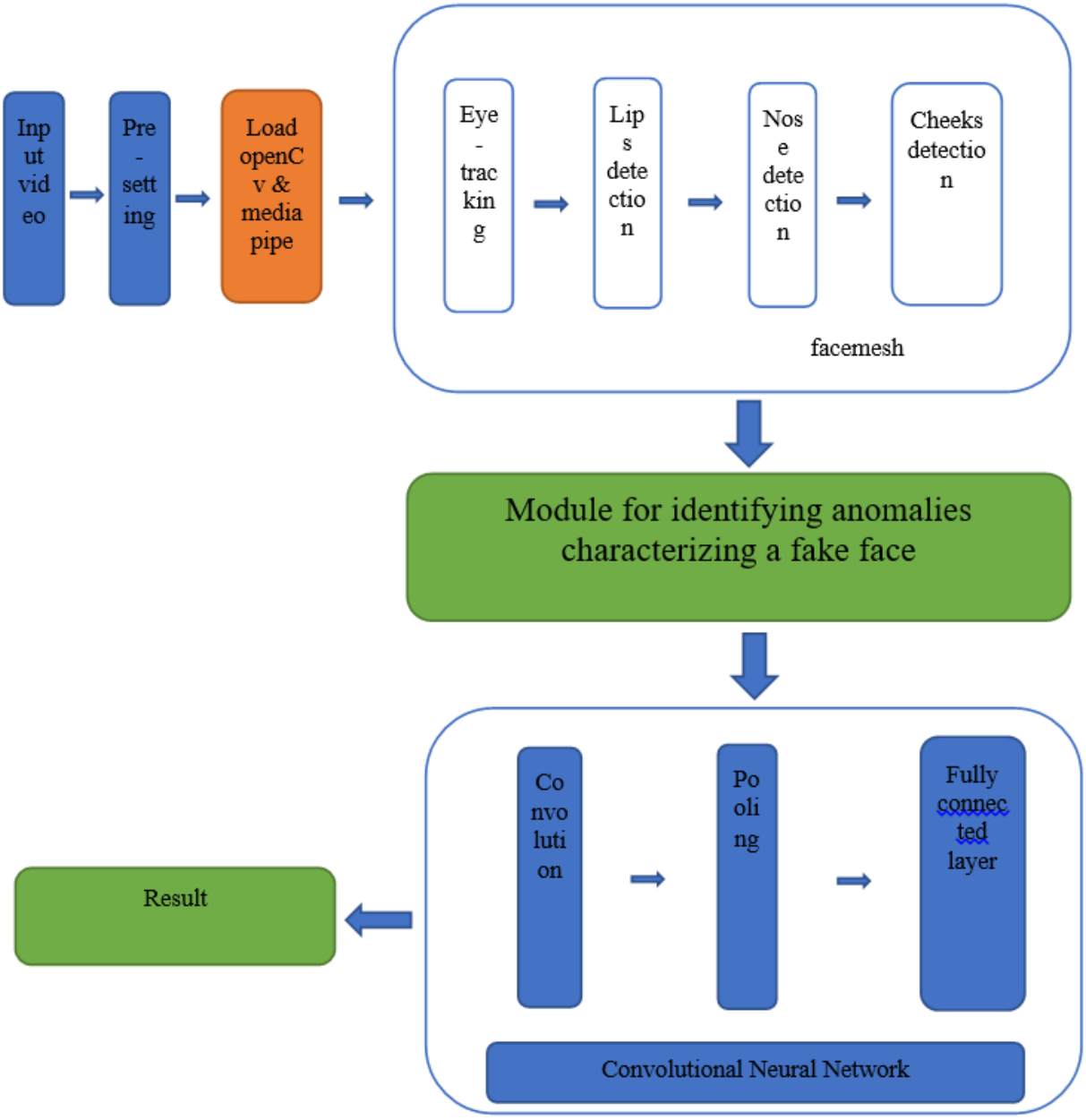

Рис. 1. Узагальнений підхід до визначення діпфейків

На першому етапі відбувається захоплення елемента відео з людськи обличчям. На наступному етапі відбувається етап попередньої підготовки відео. Для виділення необхідних елементів на обличчі потрібно

використати сучасні бібліотеки компютерного зору, такі як openCV та Mediapipe. Додатково необхідно налаштувати середовище розробки Python та бібліотеки для візуалізації та обчислення.

Принцип роботи полягає у тому, що задаються точки на обличчі людини, необхідні для ідентифікації певної частини обличчя, наприклад, очі, брови, губи, щоки тощо. Дані задаються у вигляді масиву точок. Після цього, в режимі реального часу відбувається відслідкування цих точок та обчислення відстані між ними. Наприклад для обчислення кількості кліпань використовується евклідова відстань між опрними токами на очах. Також з допомогою точок на обличчі можна визначити відстань між краєм губ, що дозволяє визначити чи людина посміхається чи ні, або інші аспекти, наприклад симетрія тощо. Детекцій щік або лоба необхідна для того щоби потім можна було визначити рівень колірного тону шкіри. Якщо цей тон надто ідеальний – це може свідчити про наявність діпфейків.

На фінальному етапі відбувається застосування згорткових нейронних мереж та різних архітектур нейронних мереж для визначення приналежності зображення до класу реальних або фейкових обєктів.

На основі такого комбінованого підходу відбувається формування кінцевого рішення щодо реальності обличчя на відео чи зображенні.

Використання бібліотеки mediapipe для задачі виділення ключових точок на обличчі наведено на рисунку 2.

Рис. 2. Виділення ключових точок на обличчі

В результаті використання такого підходу можна визначити необхідні точки, наприклад навколо очей чи губ для можливості подальшого обчислення їхніх характеристик.

Важливою складовою будь-якої системи, що використовує штучний інтелект та великі набори даних є можливість їхнього використання в хмарних середовищах. В таких випадках важливим аспектом є використання MLOps – підходів. Приклад конфігураційного файлу наведено на рисунку 3.

```
resource "digitalocean_droplet" "fake_detection" {
  ssh_keys = [
    digitalocean_ssh_key.default.fingerprint
  ]
  image     = "ubuntu-20-04-x64"
  name      = " fake_detection "
  region    = "nyc1"
  size      = "s-1vcpu-1gb"
  user_data = file("fake_detection_app.yaml")

  connection {
    host = self.ipv4_address
    user = "root"
    type = "ssh"
    private_key = file("tf-digitalocean")
    timeout = "2m"
  }
  provisioner "remote-exec" {
    inline = [
      "export PATH=$PATH:/usr/bin",
      # install nginx
      "sudo apt-get update",
      "sudo mkdir experiments",
      "sudo cd experiments",
      "sudo unzip link_to_dataset",
      "sudo git clone link_to_project"
    ]
  }
```

Рис. 3. Приклад конфігураційного файлу terraform

Висновки з даного дослідження
і перспективи подальших розвідок у даному напрямі

На основі аналітичного підходу проведено аналіз існуючих підходів, інструментів, датасетів для визначення сучасних тенденцій у розробці систем для визначення діпфейків на основі аналізу людського обличчя.

У роботі запропоновано узагальнений підхід до класифікації діпфейків з використанням як алгоритмів компютерного зору так і з використанням згорткових нейронних мереж, навчених на основі відомих датасетів.

Перевагою запропонованого підходу є те, що при оцінці фейк-реальне зображення використовуються не лише нейронні мережі, що вимагають багато обчислювальних ресурсів, GPU , хмарних технологій, але й використовуються алгоритми комп'ютерного зору, які дозволяють точково дослідити певні елементи на обличчі для подальшого аналізу та пошуку елементів, що характеризують діпфейки.

Додатково у роботі запропоновано структуру terraform – файлу для можливості розгортання проекту на хмарного середовищі з врахуванням необхідності зберігання датасету.

Література

1. Goljan M. CFA-aware features for steganalysis of color images / M. Goljan, J. Fridrich // Proc. SPIE. – 2015. – Vol. 9409. – Art. no. 94090V. https://doi.org/10.1117/12.2078399

2. Kim E. Exposing fake faces through deep neural networks combining content and trace feature extractors / E. Kim, S. Cho // IEEE Access. – 2021. – Vol. 9. – Pp. 123493-123503. https://doi.org/10.1109/ACCESS.2021.3110859

3. Ciftci U. A. Fakecatcher: Detection of synthetic portrait videos using biological signals / U. A. Ciftci, I. Demir, L. Yin // IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/TPAMI.2020.3009287.

4. Afchar D. MesoNet: a Compact Facial Video Forgery Detection Network. / D. Afchar, V. Nozick, J. Yamagishi, I. Echizen // Arxiv: 1809.00888v1 [cs.CV] 4 Sep 2018, https://arxiv.org/pdf/1809.00888v1

5. Detect DeepFakes: How to counteract misinformation created by AI. [Електронний ресурс]. – Режим доступу : https://www.media.mit.edu/projects/detect-fakes/overview/

6. Almars A. M. Deepfakes Detection Techniques Using Deep Learning: A Survey / A. M. Almars // Journal of Computer and Communications. – 2021. – Vol. 9. – P. 20-35. doi: 10.4236/jcc.2021.95003.

7. Xuan X. On the Generalization of GAN Image Forensics / X. Xuan, B. Peng, W. Wang, J. Dong // In: Sun, Z., He, R., Feng, J., Shan, S., Guo, Z. (eds) Biometric Recognition. CCBR 2019. Lecture Notes in Computer Science. Springer, Cham. – 2019. – Vol. 11818. – P. 134-141. https://doi.org/10.1007/978-3-030-31456-9_15

8. Berezsky O. Computational Intelligence in Medicine / O. Berezsky, O. Pitsun, P. Liashchynskyi, B. Derysh, N. Batryn // in: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies. Springer, Cham. – 2023. – Vol. 149. – P. 488–510. https://doi.org/10.1007/978-3-031-16203-9_28

9. Berezsky O. Regions matching algorithms analysis to quantify the image segmentation results / O. Berezsky, G. Melnyk, Y. Batko, O. Pitsun // in: Proceedings of the 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine. – 2016. – P. 33-36. doi: 10.1109/STC-CSIT.2016.7589862.

References

1. Goljan M. CFA-aware features for steganalysis of color images / M. Goljan, J. Fridrich // Proc. SPIE. – 2015. – Vol. 9409. – Art. no. 94090V. https://doi.org/10.1117/12.2078399

2. Kim E. Exposing fake faces through deep neural networks combining content and trace feature extractors / E. Kim, S. Cho // IEEE Access. – 2021. – Vol. 9. – Pp. 123493-123503. https://doi.org/10.1109/ACCESS.2021.3110859

3. Ciftci U. A. Fakecatcher: Detection of synthetic portrait videos using biological signals / U. A. Ciftci, I. Demir, L. Yin // IEEE Transactions on Pattern Analysis and Machine Intelligence. doi: 10.1109/TPAMI.2020.3009287.

4. Afchar D. MesoNet: a Compact Facial Video Forgery Detection Network. / D. Afchar, V. Nozick, J. Yamagishi, I. Echizen // Arxiv: 1809.00888v1 [cs.CV] 4 Sep 2018, https://arxiv.org/pdf/1809.00888v1

5. Detect DeepFakes: How to counteract misinformation created by AI. [Elektronnyi resurs]. – Rezhym dostupu : https://www.media.mit.edu/projects/detect-fakes/overview/

6. Almars A. M. Deepfakes Detection Techniques Using Deep Learning: A Survey / A. M. Almars // Journal of Computer and Communications. – 2021. – Vol. 9. – P. 20-35. doi: 10.4236/jcc.2021.95003.

7. Xuan X. On the Generalization of GAN Image Forensics / X. Xuan, B. Peng, W. Wang, J. Dong // In: Sun, Z., He, R., Feng, J., Shan, S., Guo, Z. (eds) Biometric Recognition. CCBR 2019. Lecture Notes in Computer Science. Springer, Cham. – 2019. – Vol. 11818. – P. 134-141. https://doi.org/10.1007/978-3-030-31456-9_15

8. Berezsky O. Computational Intelligence in Medicine / O. Berezsky, O. Pitsun, P. Liashchynskyi, B. Derysh, N. Batryn // in: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies. Springer, Cham. – 2023. – Vol. 149. – P. 488–510. https://doi.org/10.1007/978-3-031-16203-9_28

9. Berezsky O. Regions matching algorithms analysis to quantify the image segmentation results / O. Berezsky, G. Melnyk, Y. Batko, O. Pitsun // in: Proceedings of the 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine. – 2016. – P. 33-36. doi: 10.1109/STC-CSIT.2016.7589862.