

Пухальський Б. М., Кирієнко О. В.

## СИСТЕМИ ОБРОБЛЕННЯ НЕСТРУКТУРОВАНОЇ МУЛЬТИМЕДІЙНОЇ ІНФОРМАЦІЇ

*У роботі представлено огляд систем пошуку зображень і методів, які використовують для вирішення проблем комп'ютерного зору. У першій частині увагу зосереджено на потребах користувачів і вимогах до систем пошуку зображень, після чого розглянуто наявні системи, як-от Google Photos, Apple Photos, Amazon Photos, Microsoft Photos і Flickr; оцінено їхні переваги та недоліки та обговорено запропоновані шляхи вирішення виявлених недоліків. Детально проаналізовано запропоновані методи комп'ютерного зору, зокрема виявлення об'єктів (облич) і текстовий опис зображень.*

**Ключові слова:** системи пошуку зображень, комп'ютерний зір, виявлення об'єктів, текстовий опис зображень.

### Вступ

Швидке зростання цифрових зображень зумовило збільшення потреби в ефективних системах пошуку зображень, які могли б ефективно задовольнити потреби користувачів. У цій статті зроблено докладний огляд систем пошуку зображень і методів, які використовують для вирішення проблем комп'ютерного зору шляхом покращення можливостей пошуку та індексування.

Основною метою цього дослідження є аналіз потреб користувачів і вимог до систем пошуку зображень, визначення ключових аспектів, які сприяють ефективній та зручній роботі. Розуміючи ці вимоги, ми можемо оцінити системи, доступні на ринку, і вивчити їхні сильні та слабкі сторони. Зокрема, ми зосередимося на добре відомих системах пошуку зображень: Google Photos, Apple Photos, Amazon Photos, Microsoft Photos і Flickr. Ці платформи пропонують різноманітні підходи до пошуку зображень і цінну інформацію про поточний стан галузі. Окрім того, наше дослідження сприятиме знаходженню нових способів застосування комп'ютерного зору у виявленні об'єктів.

Результати цієї роботи можуть слугувати цінним ресурсом для дослідників, розробників і професіоналів галузі, які займаються розробленням і впровадженням систем пошуку зображень. Запропоновані рішення та висновки, отримані в результаті цього дослідження, можуть керувати майбутніми досягненнями в галузі.

### 1. Системи пошуку зображень

Мультимедіа — це форма комунікації, що поєднує такі види інформації, як текст, аудіо, зображення, анімація та відео [8]. На відміну від «традиційних» засобів подання інформації, мультимедіа надає користувачеві засоби взаємодії. Мультимедійна інформація репрезентується, обробляється та передається через численні сенсорні канали, як-от камери спостереження, датчики звуку тощо. Поєднання різних типів медіа підвищує насиченість і ефективність спілкування, створюючи захопливий та інтерактивний досвід.

У контексті систем пошуку зображень мультимедіа відіграє вирішальну роль у захопленні та представленні візуальної інформації. Мультимедіа охоплює широкий діапазон типів медіа: статичні (текст, зображення), динамічні (аудіо, відео) та багатовимірні (віртуальна реальність) [10] (рис. 1).

Текстова інформація, як-от підписи, описи та метадані, надає контекст і допомагає зрозуміти вміст мультимедійних об'єктів [13]. Зображення — це візуальні уявлення, які можуть передавати інформацію, викликати емоції і бути засобом спілкування [3]. Аудіокомпоненти містять звукові доріжки, голос за кадром, фонову музику або будь-яку форму аудіозаписів, які супроводжують мультимедійні презентації [6]. Відео поєднує в собі рухомі зображення та аудіо, що дає змогу знімати та відтворювати динамічні сцени, події або оповідання [3]. Анімація перед-

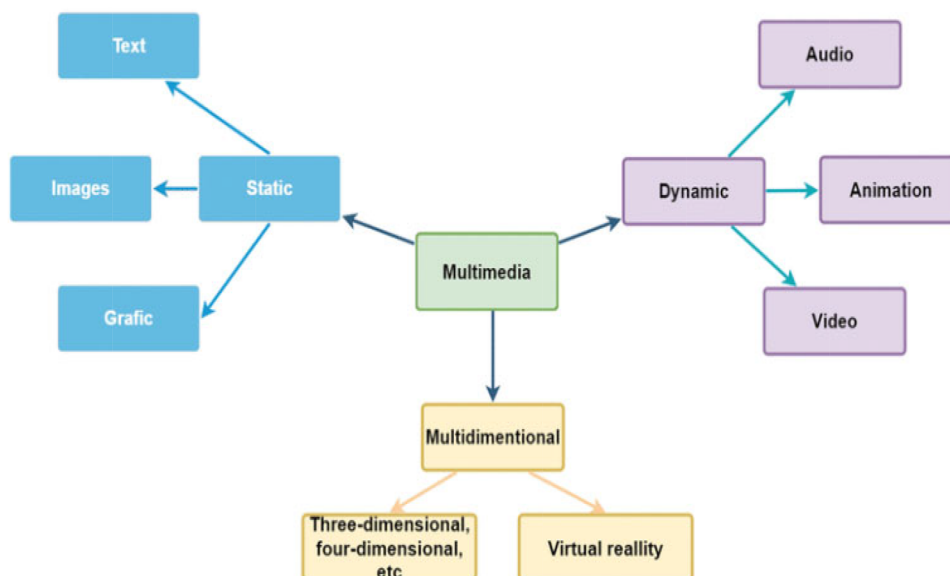


Рис. 1. Схема класифікації мультимедійної інформації [10]

бачає послідовне відображення зображень або візуальних елементів для створення ілюзії руху або зміни [13].

Кожен із цих типів потребує спеціальних методів і алгоритмів для ефективного зберігання, індексування та пошуку.

Фотографії посідають важливе місце в мультимедіа, і їх ефективний пошук є ключовою метою систем пошуку зображень. Для отримання відповідних фотографій із великих колекцій потрібні надійні механізми індексування [12].

Індексація фотографій передбачає вилучення значущих характеристик із зображень, таких як колір, текстура, форма та просторова інформація [3]. Потім ці функції використовуються для створення компактних представлень або дескрипторів зображень, які можна ефективно зберігати і здійснювати пошук [13].

Щоб увімкнути подальший пошук та індексацію, ці дескриптори зображень організуються та індексуються за допомогою відповідних структур даних, наприклад інвертованих файлів, хеш-таблиць або деревоподібних структур [12]. Цей процес індексування дає змогу здійснювати ефективний пошук на основі різних критеріїв, як-от схожість, релевантність або специфічні візуальні атрибути [3]. Пошук за дескрипторами не є новим, він широко використовується у різноманітних пошукових системах. Ці системи спрямовані на подолання семантичного розриву між низькорівневими візуальними функціями та високорівневими запитами користувача, забезпечуючи ефективний пошук і перегляд мультимедійного вмісту [6]. Вони відіграють життєво

важливу роль у різних сферах, зокрема у цифрових бібліотеках, базах даних зображень, платформах соціальних медіа та особистих колекціях фотографій.

Зазвичай користувачам систем пошуку мультимедійної інформації потрібно знайти зображення: за датою, за геолокацією, за об'єктами, зображеними на фото, за особами, зображеними на фото, за описом дій, що виконуються на фото.

Перші дві згадані функції реалізовано в більшості систем пошуку. Така інформація міститься у метаданих зображення, а перед тим наповнюється програмним забезпеченням, встановленим на пристрої, із якого роблять фотографію. Решта функцій здебільшого реалізовані в інформаційних системах із використанням нейромереж.

У таблиці 1 відображено реалізацію цих функцій у відомих системах.

Таблиця 1

#### Порівняльна характеристика функціоналу розглянутих систем

	Пошук за об'єктами	Пошук за обличчями	Пошук за діями
Google Photos	+	+	-
Apple Photos	+	+	-
Amazon Photos	+	+	-
Microsoft Photos	+	+	-
Flickr	+	+	-

Як бачимо, у кожному продукті реалізовано функціонал розпізнавання об'єктів, і облич також. Однак за допомогою жодного розглянутого застосунку не можна знайти фото за описом.

Продуктивність є важливим атрибутом якості інформаційних систем. Надмірне очікування

є неприпустимим для користувача системи. Крім того, продуктивність тісно пов'язана із масштабованістю, оскільки такі показники, як пропускна здатність і час очікування відповіді, визначають момент для розширення обчислювальних потужностей системи.

Реалізацію зазначеного функціоналу можна розглядати як розв'язання типових задач комп'ютерного зору. Є чотири основні класи задач комп'ютерного зору: класифікація зображень, виявлення об'єктів, сегментація зображень і текстовий опис зображень [9].

## 2. Методи розв'язання задач комп'ютерного зору

Найпоширенішими задачами комп'ютерного зору є класифікація зображень, визначення об'єктів (зокрема облич), сегментація зображень, текстовий опис зображень. Особливий інтерес становлять задачі з виявлення об'єктів (облич) і текстового опису зображень.

Виявлення об'єктів — це важливе завдання комп'ютерного зору, яке має справу з виявленням екземплярів візуальних об'єктів певного класу (як-от люди, тварини чи автомобілі) у цифрових зображеннях. Метою виявлення об'єктів є розроблення обчислювальних моделей і методів, які надають одну з основних частин інформації, необхідної для програм комп'ютерного зору, тобто дані про те, де які об'єкти розташовані [1; 11].

Застосування виявлення об'єктів у сфері охорони здоров'я зумовило революцію в медичній візуалізації та діагностиці, давши змогу автоматизовано виявляти та локалізувати анатомічні структури, ураження і аномалії. Ці програми використовують методи комп'ютерного зору, алгоритми глибокого навчання і передові технології візуалізації для підвищення точності та ефек-

тивності виявлення захворювань, діагностики і планування лікування. Завдяки автоматизації процесу виявлення об'єктів медичні працівники можуть заощадити час, покращити результати лікування пацієнтів і сприяти більш персоналізованому та цілеспрямованому втручанням [7].

Шлях розвитку рішень для задачі з виявлення об'єктів можна умовно поділити на два періоди. Перший період — це традиційні методи рішення задачі, розвиток яких закінчився у 2014 р., а саме алгоритми, що базуються на ознаках об'єкта, заданих вручну. Із розвитком нейронних мереж і глибокого навчання розпочався другий період, методи якого базуються на алгоритмах машинного навчання. Розвиток у цьому напрямі триває [16]. Хронологію розвитку рішень задачі з виявлення об'єктів зображено на рис. 2.

Традиційні методи розв'язання задачі з виявлення об'єктів поділяють на такі, що основані на особливостях, такі, що ґрунтуються на шаблонах, і такі, що основані на зовнішньому вигляді.

Відповідно до технік, які базуються на особливостях, визначаються риси, притаманні певному об'єктові. Для обличчя це очі, ніс, рот тощо. Наступним кроком є перевірка правдоподібного географічного розташування виявлених рис.

Техніки, що базуються на шаблонах, здатні виявляти об'єкт у різноманітних позиціях і поданнях. Однак для них потрібна належна ініціалізація, і їх швидкодія не є високою.

Згідно з технікою, що ґрунтується на зовнішньому вигляді, зображення поділяється на прямокутні зони, які можуть перекривати одна одну. Згодом у цих зонах буде знайдено кандидатів на об'єкт, який потрібно виявити. Після цього вибір можна уточнити, використовуючи каскад складніших, але селективних алгоритмів виявлення. Більшість підходів, основаних на зовнішньому вигляді, сильно покладаються на навчання кла-

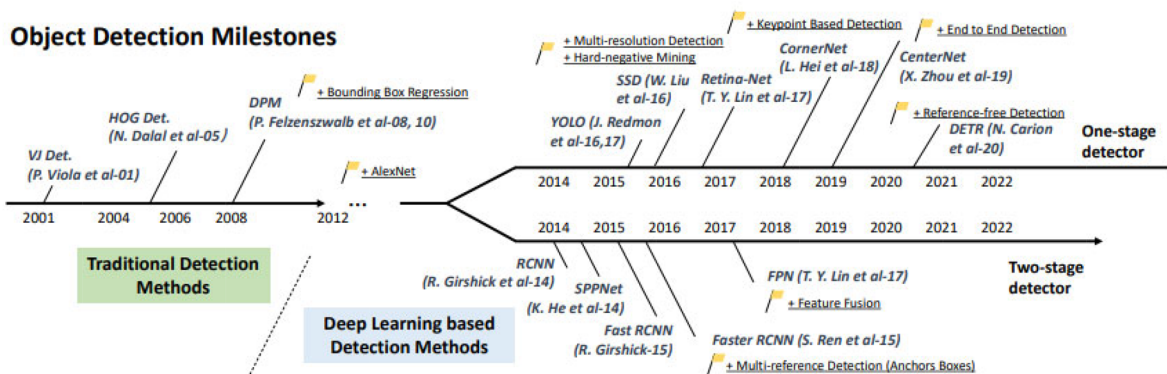


Рис. 2. Історія розвитку способів рішення задачі з виявлення об'єктів [16]

сифікаторів із використанням наборів мічених об'єктів і необ'єктних фрагментів. Серед «традиційних» детекторів найвідомішими є Viola Jones Detector, HOG Detector і Deformable Part-based Model (DPM).

Способи рішення періоду глибинного навчання поділяють на два типи, які розвивалися паралельно: виявлення у два етапи та виявлення в один етап. Перший спосіб визначає виявлення як процес «від грубого до тонкого», а другий є виявленням за один крок.

Найвідомішими двоетапними нейронними мережами є RCNN, SPPNet, FastRCNN, FasterRCNN. Короткий опис кожної із них подано нижче.

Першим етапом роботи нейромережі Regions with Convolutional Neural Networks (RCNN) є діставання пропозиції об'єктів. Потім кожна пропозиція масштабується до фіксованого розміру й передається до CNN моделі, натренованої на ImageNet, щоб дістати особливості. І наостанок, використовуються SVN класифікатори, щоб визначити наявність об'єкта у кожній області зображення та розпізнати категорії об'єкта.

Spatial Pyramid Pooling Networks (SPPNet) містить шар просторової піраміди, що дає змогу репрезентувати фіксовану довжину незалежно від розміру вхідного зображення або області без масштабування.

FastRCNN дає змогу одночасно тренувати детектор і регресор границь області за незмінної конфігурації мережі. Ця мережа є першим наскрізним і близьким до реального часу детектором. Головною особливістю є запровадження шару RPN, який забезпечує нересурсозатратні припущення областей. Крім того, окремі блоки системи виявлення об'єктів (виявлення припущення, вилучення особливостей, регресія границь області) поступово інтегрувалися в уніфікований наскрізний фреймворк навчання.

Серед одноетапних нейронних мереж для виявлення об'єктів варто згадати YOLO, SSD і Retina Net.

You Only Look Once (YOLO) — перший одноетапний детектор ери глибинного навчання. За принципом роботи ділить зображення на області, визначає границі та ймовірності для них. Основною відмінністю Single Shot Multibox Detector (SSD) є запровадження методів виявлення із декількома посиланнями та з різною роздільною здатністю, що значно підвищує точність виявлення одноетапного детектора, особливо для малих об'єктів. Оскільки одноетапні детектори не вирізняються точністю роботи, у Retina Net було введено нову функцію витрат

за назвою «фокальна витрата», яка змінює стандартні витрати перехресної ентропії, щоб під час навчання детектор більше зосереджувався на складних, неправильно класифікованих прикладах.

Визначено, що за принципом своєї роботи двоетапні нейронні мережі дають точніший результат, натомість одноетапні є швидшими. Вибираючи нейронну мережу, що буде застосовуватись у прикладному програмному забезпеченні, потрібно маневрувати між точністю та продуктивністю.

Окремо варто згадати про таку задачу комп'ютерного зору, як розпізнавання облич, що є окремим випадком задачі з виявлення об'єктів. Цю технологію застосовують у автоматизованому відстежуванні дотримання правил дорожнього руху, в розслідуванні кримінальних справ, у виявленні емоцій і накладанні масок у мобільних застосунках та в інших сферах. Хоч виявлення облич є підкласом більш загальної задачі, воно є дещо специфічним.

У процесі виявлення облич часто можна зіткнутися із такими складнощами, як варіативність, перекриття, виявлення у різних масштабах та виявлення у реальному часі [14]. Людське обличчя здатне виражати багато емоцій, причому кожна людина виражає їх по-своєму, до того ж існує велика варіативність кольору та відтінку шкіри, що ускладнює процес виявлення. Обличчя можуть бути частково закритими різноманітними об'єктами та аксесуарами, як-от волосся, головні убори, окуляри, макіяж, маска та інші. Масштаб фотографій облич також різниться. Лице може бути сфотографовано зблизька, здалеку або в натовпі. Для виявлення у реальному часі обличчя, яке містить досить багато характерних рис, потрібно залучати швидкі обчислювальні потужності, особливо коли йдеться про мобільні пристрої.

Згаданий детектор Viola Jones був однією з перших спроб рішення цієї проблеми і став основою для розвитку пізніших детекторів [2].

Текстовий опис зображень є порівняно новою проблемою у сфері комп'ютерного зору. Визначення проблеми полягає у генеруванні опису, який є найбільш доречним до заданого зображення. Опис містить об'єкти, які зображені на заданому мультимедійному файлі, та зв'язки між цими об'єктами. Серед зв'язків можуть бути дії, які виконують знайдені об'єкти, сполучники, прийменники та інші частини мови, що забезпечують цілісність опису. Не існує вимог щодо структури опису. Його можна надати в довільній формі [5].

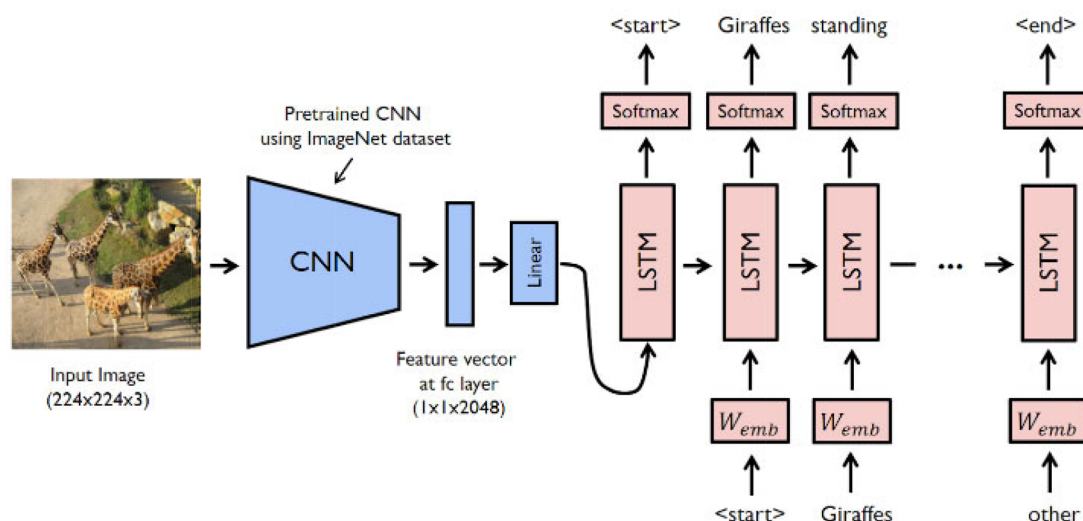


Рис. 3. Схема моделі Show and Tell [4]

Можна знайти багато застосувань вирішенням цієї проблеми. Найбільш очевидне — це генерація опису всього, що відбувається на вихідному зображенні, для підвищення ефективності пошуку самого зображення. Це дає нам змогу забезпечити посередництво між людиною з її високорівневими запитами та системою аналізу й пошуку зображень з її низькорівневими функціями, що суттєво підвищує зручність використання подібних систем.

Розглядувану функцію можна використовувати там, де є потреба забезпечити інклюзивність. Наприклад, люди із вадами зору, які не здатні чітко бачити мультимедійний вміст, отримують текстовий опис зображення зручним для них шрифтом.

По своїй суті текстовий опис зображень є комбінацією виявлення згаданих раніше об'єктів [9]. Протягом останніх майже десяти років виявлення виконують методами глибинного навчання, а опис — такими нейронними мережами, як RNN (Recurrent Neural Net) та LSTM (Long Short Term Memory), яка є розширеною версією RNN. Існує також модель Show and Tell, яка поєднує два згадані рішення з використанням LSTM [15]. На рис. 3 подано структуру цієї моделі.

Під час розроблення системи аналізу та пошуку зображень буде використано такий підхід із конфігураціями шарів для виявлення об'єктів.

## Висновки

Цю роботу було зосереджено на дослідженні систем пошуку зображень і використання методів комп'ютерного зору для ефективного пошуку та індексування. Проведено огляд систем пошуку, акцентовано на важливості пошуку зображень у ширшому мультимедійному контексті. Досліджено потреби та вимоги користувачів до систем пошуку зображень, що дало розуміння факторів, які сприяють ефективній взаємодії з користувачем.

Наведено порівняльну характеристику сучасних систем пошуку зображень, зокрема Google Photos, Apple Photos, Amazon Photos, Microsoft Photos і Flickr. Визначено переваги та недоліки цих систем.

У статті також описано основні методи вирішення проблем комп'ютерного зору. Акцент було зроблено на виявленні об'єктів, зокрема на розпізнаванні облич, і використанні текстового опису у пошуку зображень. В основному дано уявлення про методи, які використовують для розпізнавання обличчя та застосування текстового опису зображень, створення основи для подальшого розвитку системи пошуку та індексування зображень. Наведено порівняльну характеристику розглянутих способів.

## Список літератури

1. Як працює OSINT-розвідка? Від бізнес-аналізу до оборони України [Електронний ресурс] // ISSP Training Center. — 2022. — Режим доступу: [www.issp.training/post/yak-pratsyuye-osint-rozvidka-vid-biznes-analizu-do-oborony-ukrainy](http://www.issp.training/post/yak-pratsyuye-osint-rozvidka-vid-biznes-analizu-do-oborony-ukrainy).
2. Cheney J. Unconstrained face detection: State of the art baseline and challenges / Jordan Cheney, Ben Klein, Anil K. Jain, Brendan F. Klare // International Conference on Biometrics (ICB). — Phuket, Thailand, 2015. — Pp. 229–236. <https://doi.org/10.1109/ICB.2015.7139089>.



3. Datta R. Image retrieval: Ideas, influences, and trends of the new age / Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang // *ACM Computing Surveys*. — 2008. — Vol. 40, issue 2. — P. 60. <https://doi.org/10.1145/1348246.1348248>.
4. Image Captioning [Electronic resource]. — Mode of access: <http://shikib.com/captioning.html>.
5. Image Captioning [Electronic resource] // Papers with Code. — Mode of access: <https://paperswithcode.com/task/image-captioning>.
6. Lew M. S. Content-based multimedia information retrieval: State of the art and challenges / Michael S. Lew, Nicu Sebe, Chabane Djeraba, Ramesh Jain // *ACM Transactions on Multimedia Computing, Communications, and Applications*. — 2006. — Vol. 2, issue 1. — Pp. 1–19. <https://doi.org/10.1145/1126004.1126005>.
7. Litjens G. A survey on deep learning in medical image analysis / Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, Clara I. Sánchez // *Medical Image Analysis*. — 2017. — Vol. 42. — Pp. 60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
8. Multimedia [Electronic resource]. — Mode of access: <https://en.wikipedia.org/wiki/Multimedia>.
9. Murali S. An analysis on computer vision problems [Electronic resource] / Shravan Murali. — 2018. — Mode of access: <https://medium.com/deep-dimension/an-analysis-on-computer-vision-problems-6c68d56030c3>.
10. Nikulin D. What is Multimedia Data? [Electronic resource] / Dima Nikulin // DEV Community. — 2022. — Mode of access: <https://dev.to/dimanikulin/what-is-multimedia-data-111f>.
11. Raikote P. Object Detection — Part 1: Introduction [Electronic resource] / Pranav Raikote // *Applied Singularity*. — 2021. — Mode of access: <https://appliedsingularity.com/2021/05/18/object-detection/>.
12. Sivic J. Video Google: a text retrieval approach to object matching in videos / Josef Sivic, Andrew Zisserman // *Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France*. — 2003. — Vol. 2. — Pp. 1470–1477. <https://doi.org/10.1109/ICCV.2003.1238663>.
13. Smeulders A. Content-based image retrieval at the end of the early years // A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 2000. — Vol. 22, no. 12. — Pp. 1349–1380. <https://doi.org/10.1109/34.895972>.
14. Varun. What is Face Detection? Ultimate Guide 2023 + Model Comparison [Electronic resource] / Varun // *LearnOpenCV*. — 2022. — Mode of access: <https://learnopencv.com/what-is-face-detection-the-ultimate-guide/>.
15. Vinyals O. Show and Tell: A Neural Image Caption Generator / Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. — June, 2015.
16. Zou Z. Object Detection in 20 Years: A Survey / Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, Jieping // *Proceedings of the IEEE*. — 2023. — Vol. 111, no. 3. — Pp. 257–276. <https://doi.org/10.1109/JPROC.2023.3238524>.

### References

- Cheney, J., Klein, B., Jain, A. K., & Klare, B. (2015). Unconstrained face detection: State of the art baseline and challenges. *2015 International Conference on Biometrics (ICB)*. <https://doi.org/10.1109/icb.2015.7139089>.
- Datta, R., Joshi, D., Li, J., & Wang, J. (2008). Image retrieval. *ACM Computing Surveys*, 40 (2), 1–60. <https://doi.org/10.1145/1348246.1348248>.
- Image Captioning. (n.d.). Captioning. <http://shikib.com/captioning.html>.
- Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2 (1), 1–19. <https://doi.org/10.1145/1126004.1126005>.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>.
- Murali, S. (2018, June 17). An analysis on computer vision problems – Shravan’s Blog - Medium. *Medium*. <https://medium.com/deep-dimension/an-analysis-on-computer-vision-problems-6c68d56030c3>.
- Nikulin, D. (2022, July 7). *What is Multimedia Data?* DEV Community. <https://dev.to/dimanikulin/what-is-multimedia-data-111f>.
- Papers with Code - Image Captioning. (n.d.). <https://paperswithcode.com/task/image-captioning>.
- Raikote, P. (2021, May 18). *Object Detection – Part 1: Introduction*. *Applied Singularity*. <https://appliedsingularity.com/2021/05/18/object-detection/>.
- Sivic, & Zisserman. (2003). Video Google: a text retrieval approach to object matching in videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2. <https://doi.org/10.1109/iccv.2003.1238663>.
- Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380. <https://doi.org/10.1109/34.895972>.
- Varun. (2023, May 9). What is Face Detection? Ultimate Guide 2023 + Model Comparison. *LearnOpenCV – Learn OpenCV, PyTorch, Keras, Tensorflow with examples and tutorials*. <https://learnopencv.com/what-is-face-detection-the-ultimate-guide/>.
- Vinyals, O., Toshev, A., Bengio, S., Erhan, D., & Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015). Show and Tell: a neural image caption generator. *These CVPR 2015 Papers Are the Open Access Versions, Provided by the Computer Vision Foundation*.
- Wikipedia contributors. (2023, November 6). *Multimedia*. Wikipedia. <https://en.wikipedia.org/wiki/Multimedia>.
- Yak pratsiuiie OSINT-rozvidka? Vid biznes-analizu do oborony Ukrainy. (2022, September 20). ISSP Training. <https://www.issp.training/post/yak-pratsyuye-osint-rozvidka-vid-biznes-analizu-do-oborony-ukrainy>.
- Zou, Z., Chen, K., Shi, Z., Guo, Y., & Guo, Y. (2023). Object Detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3), 257–276. <https://doi.org/10.1109/jproc.2023.3238524>.

*B. Pukhalskyi, O. Kyriienko*

## SYSTEMS FOR PROCESSING UNSTRUCTURED MULTIMEDIA INFORMATION

*This paper presents an overview of image search systems and the methods used to solve computer vision problems. The first part focuses on the needs of users and the requirements for image search systems. Existing systems, such as Google Photos, Apple Photos, Amazon Photos, Microsoft Photos, and Flickr, are then considered. Their advantages and disadvantages are evaluated, and proposed solutions to the identified shortcomings are discussed. Proposed computer vision methods, including object detection (face detection) and image text description, are analyzed in detail.*

*The rapid growth of digital images has led to an increasing need for effective image search systems that can effectively meet the needs of users. The research presented in this paper aims to provide a detailed overview of image search systems and the methods used to solve computer vision problems by improving search and indexing capabilities.*

*The main purpose of this research is to analyze the needs of users and the requirements for image search systems, to identify key aspects that contribute to effective and convenient operation. By understanding these requirements, we can evaluate existing systems available on the market and study their strengths and weaknesses. Specifically, we will focus on well-known image search systems, including Google Photos, Apple Photos, Amazon Photos, Microsoft Photos, and Flickr. These platforms represent a variety of approaches to image search and offer valuable information about the current state of the industry. In addition, our research will contribute to the discovery of new ways to apply computer vision to object detection.*

*The results of this work can serve as a valuable resource for researchers, developers, and professionals in the field who are involved in the development and implementation of image search systems. The proposed solutions and conclusions obtained as a result of this research can guide future advancements in the field.*

**Keywords:** image search systems, computer vision, object detection, image text description.

*Матеріал надійшов 25.06.2023*



Creative Commons Attribution 4.0 International License (CC BY 4.0)