

КОНСОЛІДОВАНЕ РАНЖУВАННЯ СУЧАСНИХ ХМАРНИХ ОНЛАЙН-СЕРВІСІВ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННЯХ

¹Донецький національний університет імені Василя Стуса;

²Вінницький національний технічний університет

Автоматизація розпізнавання об'єктів на зображеннях є доволі поширеною задачею з очевидним практичним застосуванням у промисловому виробництві, медицині, транспорті, безпеці та інших галузях. Сьогодні існує кілька хмарних сервісів, які пропонують онлайн-інструменти для вирішення різних завдань, пов'язаних з розпізнаванням зображень. Вони мають низку переваг над традиційними інструментами. Наразі відсутні методи, які б дозволили користувачу виявити, якій саме хмарний сервіс найкраще підходить під його задачі. При цьому оцінювання має бути стислим і базуватися на використанні обмеженого набору профільних зображень. Відповідно типовий метод грубої сили, який потребує завантаження та аналізу великої кількості зображень є неприйнятним. Метод має ґрунтуватися на деталізованому аналізі результатів розпізнавання за малим набором тестових зображень з урахуванням особливостей хмарних сервісів. В роботі запропоновано метод ранжування хмарних сервісів за малих тестових датасетів. При цьому користувач формує тестові датасети з урахуванням профілю власних задач розпізнавання об'єктів. Запропонований метод базується на трьох частинних метриках, кожна з яких враховує ту чи іншу особливість хмарних сервісів. Перша метрика — це різниця між добутком рівнів впевненості правильно розпізнаних об'єктів та добутком рівнів впевненості хибно розпізнаних об'єктів. Друга метрика — це медіанне значення довжини списку виявлених об'єктів до першої помилки. Список об'єктів має бути відсортовано за спаданням впевненості. Перші дві частинні метрики є традиційними, а третя — новою. Фінальне рішення ухвалюється за консолідованим рейтингом, який агрегує три частинні метрики. Застосування методу ілюструється на задачі ранжування хмарних сервісів Microsoft Azure AI Vision Studio, Amazon Rekognition, Google Cloud Vision та Imagga.

Ключові слова: розпізнавання об'єктів, хмарні сервіси, метрика, ранжування, зображення.

Вступ

У сучасному цифровому світі велика кількість інформації представляється у формі зображень — різноманітних картинок та фотографій. Автоматизація розпізнавання об'єктів на зображеннях є доволі поширеною задачею з очевидним практичним застосуванням у промисловому виробництві, медицині, транспорті, безпеці та інших галузях [1]. В маркетингу важливо аналізувати зображення для вивчення поведінки споживачів, визначення популярності продуктів та вдосконалення збутових стратегій. В автономних системах з комп'ютерним зором — в автомобілях, дронах та роботах — розпізнавання об'єктів відіграє ключову роль в ухваленні рішень щодо забезпечення безпеки руху. Щодня з'являються нові різновиди задач, ефективність розв'язання яких може бути значно покращена за рахунок використання автоматичного розпізнавання об'єктів.

Під час автоматичного розпізнавання об'єктів на зображеннях виникає низка проблем, які зумовлені певними особливостями задач. Зображення можуть містити велику кількість об'єктів з різними розмірами, формою, кольорами та текстурою. Для ефективного навчання моделей розпізнавання об'єктів необхідно використовувати великі обсяги різноманітних зображень. Збір та розмітка великих датасетів є часомістким та вартісним процесом. Розпізнавання об'єктів також включає в себе розуміння контексту та семантики зображення. Деякі об'єкти можуть бути омонімічними — мати схожі форми, але різні семантичні значення.

Передові методи розпізнавання об'єктів потребують значних обчислювальних ресурсів, особливо для задач реального часу або за великого обсягу даних. Все це робить задачу розпізнавання об'єктів високозатратною та технічно складною [2]. Одним з популярних виходів з такої ситуації є використання хмарних сервісів. Таких сервісів є кілька, тому виникає задача вибору кращої альтернативи.

Метою статті є розроблення методу, який дозволяє оцінити та проранжувати доступні сучасні хмарні сервіси щодо якості розпізнавання об'єктів на зображеннях.

Аналіз останніх досліджень і публікацій

Найближчою за тематикою дослідження є стаття [3]. В ній автори порівнюють сервіси від Amazon, Google та Azure шляхом розрахунку схожості знайдених на зображеннях категорій за власним методом COSLAB. Цей метод базується на ідеях алгоритмів Word2vec та GloVe. Недоліком методу є те, що класифіковані об'єкти не аналізуються на безпомилковість розпізнавання, зокрема не аналізують випадки правильно розпізнаних (True Positive) та помилково розпізнаних (False Positive) об'єктів. Також, як недолік можна відмітити те, що алгоритм GloVe для деяких категорій видає напрочуд високу схожість, до прикладу, коефіцієнт схожості між кішкою та собакою дорівнює 0,922. Відповідно, якщо сервіс замість kota виявив собаку, то ця помилка буде незначною. Але з таким висновком можна погодитися лише для окремих картинок. Для багатьох інших сцен такий низький штраф за плутання kota з собакою є неприпустимим. Але коефіцієнти схожості є сталими і не залежать від змісту аналізованої сцени.

В іншій релевантній публікації [4] аналізуються сервіси від Clarifai, Google та Azure методом тренування моделі на підготовлених зображеннях з подальшою класифікацією. Але, в підготовлених датасетах використовуються лише зображення з трьома категоріями, що є суттєвим обмеженням — реальні зображення можуть містити значно більше об'єктів.

Є кілька публікацій, в яких автори аналізують здатність хмарних сервісів якісно виявляти деякі специфічні об'єкти. Серед новітніх таких публікацій виділимо роботу [5], в якій на основі тестування набору стандартизованих зображень Військової академії США встановлено, що хмарні сервіси доволі часто неправильно розпізнають об'єкти одягу. Причому, кількість протегованих об'єктів суттєво залежить від гендерних та расових груп, які присутні на зображенні.

В проаналізованих публікаціях та в інших роботах відсутні методи, які б дозволили користувачу виявити, якій хмарний сервіс найкраще підходить саме під його задачу. При цьому оцінювання має бути стислим і базуватися на використанні обмеженого набору профільних зображень. Метод має враховувати обмеження доступу до сервісів. А саме те, що демоверсії кожного з них мають лімітовану кількість запитів. Відповідно метод грубої сили, який потребує завантаження та аналізу великої кількості картинок є неприйнятним. Метод має ґрунтуватися на деталізованому аналізі результатів розпізнавання за малим набором тестових зображень з урахуванням особливостей хмарних сервісів.

Особливості хмарних сервісів розпізнавання об'єктів

На сьогодні існує кілька хмарних сервісів, які пропонують онлайн-інструменти для розв'язання різних задач, пов'язаних з розпізнаванням зображень. Вони мають багато переваг перед традиційними інструментами, а саме можливість використання хмарних ресурсів для складних обчислень, готові алгоритми та натреновані моделі, можливість їхньої інтеграції на різних популярних мовах програмування з використанням API, детальна статистика щодо використання, розподілені центри обчислення даних по різних регіонах світу. Відносним недоліком можна вважати вартість використання таких сервісів, на яку розробник не може вплинути.

Для дослідження вибрано такі чотири сервіси: Amazon Rekognition, Google Cloud Vision, Microsoft Azure AI Vision Studio та Imagga. Усі вони мають можливість інтеграції за допомогою API та надають відповідну детальну документацію. Додатковою перевагою є наявність певної кількості безкоштовних запитів на місяць. До того ж, всі вибрані сервіси мають онлайн демоверсію для зареєстрованих користувачів, яка дозволяє швидко і легко завантажити власне зображення та отримати результати розпізнавання. Це може бути корисним для попереднього відбору сервісу шляхом дослідження невеликої вибірки з відповідної предметної області прикладної задачі. Такий підхід не вимагає багато часу і дозволяє приблизно оцінити якість без необхідності розроблення програмної інтеграції під кожний сервіс.

Особливості сервісів, які потрібно врахувати під час розробки методу ранжування, є такими.

1. Довжина списку виявлених на тому самому зображенні об'єктів є різною для різних сервісів. При цьому, апріорі неможливо спрогнозувати, якою буде ця довжина.

2. Загальні списки можливих категорій для різних сервісів неузгоджені. Довжина їхнього списку для кожного сервісу невідома.

3. Для кожного виявленого об'єкта наводиться рівень впевненості в достовірності його присутності на зображенні.

Метрики для аналізу якості розпізнавання об'єктів

На сьогодні не існує універсальної метрики, за якою можна було б порівняти хмарні сервіси розпізнавання об'єктів з урахуванням перерахованих особливостей. Тому, пропонуємо використовувати кілька метрик, кожна з яких враховує ту чи іншу особливість. Фінальне рішення пропонується вибирати за консолідованою метрикою, яка є сумою рангів за кожною частинною метрикою. Всього пропонується застосувати три частинні метрики.

Перша метрика *MeanQuality* [6] базується на визначенні середнього значення впевненості сервісу у виявленому об'єкті на зображенні з подальшим усередненням за усім тестовим датасетом. У разі правильно виявленого об'єкта значення впевненості додається, а за хибно виявленого — віднімається. Отримана сума ділиться на кількість виявлених об'єктів з подальшим усередненням за усіма зображеннями

$$MeanQuality = \frac{1}{N} \sum_{i=1, 2, \dots, N} \frac{1}{L_i} \sum_{j=1, 2, \dots, L_i} p_{ij} \cdot \text{sign}(i, j), \quad (1)$$

де N — кількість тестових зображень; L_i — кількість виявлених об'єктів на i -му зображенні; p_{ij} — значення впевненості сервісу для j -го об'єкта на i -му зображенні; $\text{sign}(i, j)$ — функція, яка набуває значення 1 у випадку правильного розпізнавання j -го об'єкта на i -му зображенні та -1 у випадку хибного.

Що більше значення метрики, тим краща якість сервісу. Це проста та зрозуміла метрика, але невідомо який сенс закладено в числові оцінки достовірності в різних сервісах. Також невідомо, чи є вони адитивними. У формулі (1) вважається, що внесок у метрику двох об'єктів з рівнем впевненості 1 та 0,6 такий самий, як і для двох об'єктів з рівнем впевненості 0,8 та 0,8. Проте невідомо, чи так є насправді.

Друга метрика *MeanAccuracy* [7] розраховується як відношення кількості правильно визначених об'єктів до загальної кількості. Як і для метрики *MeanQuality*, зробимо агрегований розрахунок по всьому набору вибраних для тесту картинок. Тобто загальна оцінка буде сумою результатів роботи сервісу по всіх зображеннях, поділеною на кількість зображень

$$MeanAccuracy = \frac{1}{N} \sum_{i=1, 2, \dots, N} \frac{1}{L_i} \sum_{j=1, 2, \dots, L_i} tp_{ij}, \quad (2)$$

де tp_{ij} дорівнює 1 у випадку правильного розпізнавання j -го об'єкта на i -му зображенні та 0 — за хибного розпізнавання.

Що більше значення метрики (2), тим краща якість сервісу. Це також проста та зрозуміла метрика. Але, якщо сервіс видає короткий список об'єктів, то він буде менше помилятися, ніж сервіс, який намагається детально описати зображення і видати довгий список об'єктів. Відповідно, сервіс з короткою видачею буде, очевидно, кращим за метрикою (2) — він зрідка помилятиметься, але водночас повнота його опису замала.

Для доповнення двох вищеописаних метрик пропонуємо третю — *TrueBeforeFalse*. Ця метрика враховує ту особливість, що користувачу не завжди потрібен повний опис картинки. Для деяких задач йому важливо отримати такий список виявлених об'єктів, який дозволить зрозуміти зміст картинки, тобто дозволить знизити ентропію до рівня, достатнього для ухвалення правильного рішення. Для одних задач буде достатньо списку з двох об'єктів, для других — списку з трьох об'єктів, а для третіх — семи чи десяти. Які об'єкти користувач має вибрати з довгого списку виявлених об'єктів? Логічно, що вибирати він буде в порядку зниження рівня впевненості. І якщо в цей короткий список потрапить неправильно розпізнаний об'єкт, тоді шанси на помилкове рішення зростуть. На підставі цих міркувань авторами пропонується метрикою призначити довжину

списку до першого помилково виявленого об'єкта. Список об'єктів має бути відсортованим в порядку спадання впевненості. Усередненим значенням метрики за усім датасетом виберемо медіану довжин таких списків для кожного зображення

$$TrueBeforeFalse = \text{median}(tbp_1, tbp_2, \dots, tbp_N), \quad (3)$$

де tbp_i — кількість правильно визначених об'єктів до першої помилки для i -го зображення.

Набір зображень для тестування методу

Для перевірки методу відберемо 8 тестових зображень різної тематики (рис. 1), які містять велику кількість категорій з різних предметних областей. Розглянемо кожне зображення детальніше.



Рис. 1. Тестові зображення

Тематика «Вуличне фото» включає два тестові зображення. Кожне зображення має різний ракурс, різну кількість та набір категорій об'єктів на світлинці — люди, транспортні засоби, міські будівлі, елементи дорожньої інфраструктури тощо.

Тематика «Інтер'єр» також включає два тестові зображення. Кожне зображення — це фото кімнати, кухні та спальні відповідно. Присутня велика кількість та широкий набір категорій об'єктів на світлинці — меблі, елементи декору, посуд тощо.

Тематика «Екстер'єр» представлена фото зовнішнього фасаду будинку та галявиною перед ним. Набір категорій об'єктів на світлинці — деталі екстер'єру будинку, різні види рослин (трава, квіти, кущі, дерева).

Тематика «Спорт» представлена фото перегонів транспортних засобів під час спортивних змагань. Набір категорій об'єктів на світлинці — транспортні засоби, окремі їхні елементи, деталі та спортивні аксесуари.

Тематика «Тварини» представлена фото тварин у їхніх природних умовах проживання. Набір категорій об'єктів на світлинці — різні види тварин, рослини, елементи природного ландшафту тощо.

Тематика «Натюрморт». Набір категорій об'єктів на світлинці — різні види фруктів та ягід.

Розрахунок метрик та ранжування сервісів

Кожне зображення завантажимо у вибрані сервіси. Результати розпізнавання за кожним зображенням представляються у форматі, показаному на рис. 2. На основі цих даних розрахуємо значення трьох метрик для усіх тестових зображень та усереднимо їх (табл. 1—3).

За метрикою *MeanQuality Azure* та *Amazon* мають найкращі показники. Проте, розпізнавання зображення типу «Натюрморт» виявилось найпроблемнішою задачею для всіх, окрім *Imagga*. Разом з тим, *Imagga* дуже помиляється для таких зображень, як «Вуличне фото 1» та «Тварини», які легко розпізнаються іншими сервісами. Це може бути ознакою того, що для навчання *Imagga* використано інші методи та/або набори даних. Такі властивості *Imagga* є сигналом, що цей сервіс можна розглядати як кандидата для включення в ансамблеві моделі. Проте, ця гіпотеза потребує додаткової перевірки.

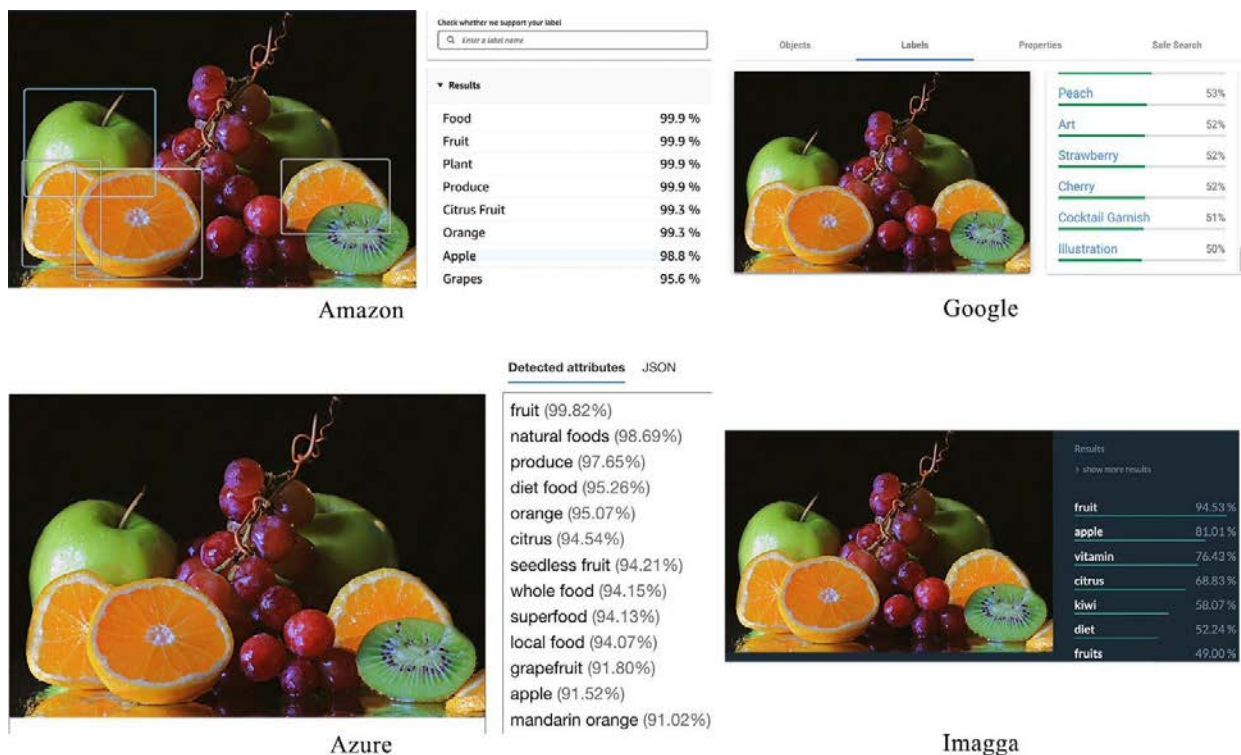


Рис. 2. Результати розпізнавання зображення «Натюрморт» в демо-версіях аналізованих сервісів

Таблиця 1

Оцінювання за метрикою *MeanQuality*

Зображення	Amazon	Google	Azure	Imagga
Вуличне фото 1	88,46	43,65	67,78	8,40
Вуличне фото 2	33,43	72,35	63,02	19,35
Інтер'єр 1	55,46	9,8	66,45	28,04
Інтер'єр 2	67,29	8,7	85,56	31,83
Екстер'єр	54,27	64,25	49,33	23,54
Спорт	34,61	52,05	76,46	22,94
Тварини	55,28	44,9	70,03	12,70
Натюрморт	18,86	-5,6	27,80	26,54
<i>MeanQuality</i>	50,96	36,26	63,30	21,67

Таблиця 2

Оцінювання за метрикою *MeanAccuracy*

Зображення	Amazon	Google	Azure	Imagga
Вуличне фото 1	0,95	0,75	0,8667	0,55
Вуличне фото 2	0,6842	0,95	0,85	0,75
Інтер'єр 1	0,7895	0,55	0,875	0,7
Інтер'єр 2	0,9412	0,55	0,95	0,7
Екстер'єр	0,7692	0,9	0,8	0,9
Спорт	0,7143	0,8	0,9412	0,8
Тварини	0,8125	0,75	0,9	0,6
Натюрморт	0,5714	0,45	0,65	0,75
<i>MeanPrecision</i>	0,779	0,7125	0,8541	0,7188

Оцінювання за метрикою *TrueBeforeFalse*

Зображення	Amazon	Google	Azure	Imagga
Вуличне фото 1	1	3	10	5
Вуличне фото 2	4	15	0	9
Інтер'єр 1	0	2	9	6
Інтер'єр 2	15	1	8	7
Екстер'єр	3	6	2	4
Спорт	1	3	13	0
Тварини	1	11	13	1
Натюрморт	2	1	2	2
<i>TrueBeforeFalse</i>	1,5	3	8,5	4,5

За метрикою *MeanAccuracy* Azure та Amazon мають найкращі показники. Натомість, на відміну від попередньої метрики, Imagga виявився кращим за Google. Зауважимо, що якість сервісів суттєво залежить від типу зображень.

За метрикою *TrueBeforeFalse* Azure знов виявився кращим. Зазначимо, що всі сервіси мають досить великий розкид значення метрики за окремими зображеннями. Можна припустити, що це зумовлено тим, що всі зображення мають зовсім різний набір об'єктів і жодний сервіс не може показати стабільну поведінку на усіх тестових зображеннях.

Тепер зробимо фінальний розрахунок за консолідованою метрикою (табл. 4). З таблиці видно, що Azure є лідером. Інші сервіси є приблизно однаковими. В межах цієї трійки позиція за консолідованою метрикою є чутливою до одиничної зміни рангової позиції за однією будь-якою частинною метрикою.

Таблиця 4

Ранжування за консолідованою метрикою

Ранг	Azure	Amazon	Imagga	Google
Ранг за <i>MeanQuality</i>	1	2	4	3
Ранг за <i>MeanAccuracy</i>	1	2	3	4
Ранг за <i>TrueBeforeFalseRanking</i>	1	4	2	3
Сума частинних рангів	3	8	9	10
Консолідований Ранг	1	2	3	4

Висновки

Задачу розпізнавання об'єктів на зображеннях можна ефективно розв'язати за допомогою хмарних онлайн-сервісів Microsoft Azure AI Vision Studio, Amazon Rekognition, Google Cloud Vision та Imagga. Запропоновано метод ранжування, який базується на трьох частинних метриках *MeanQuality*, *MeanAccuracy* та *TrueBeforeFalse*, кожна з яких враховує ту чи іншу особливість зазначених хмарних сервісів. Фінальне рішення ухвалюється за консолідованим рейтингом. Відмінність запропонованого методу полягає у деталізованому аналізі результатів розпізнавання за трьома метриками, що дозволяє компенсувати деякою мірою малий обсяг тестового набору зображень. Якість розпізнавання хмарними сервісами суттєво залежить від типу зображень. Тому, для вибору кращого сервісу користувачу слід сформулювати тестові датасети з урахуванням профілю власних задач розпізнавання об'єктів. Користувач також може врахувати додаткові знання щодо відповідності частинних метрик цілям його діяльності шляхом введення у консолідовану метрику вагових коефіцієнтів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proceedings of the IEEE*, no. 111 (3), pp. 257-276, 2023. <https://doi.org/10.1109/JPROC.2023.3238524>.
- [2] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" *PLoS Computational Biology*, no. 4 (1), pp. 0151-0156, 2008. <https://doi.org/10.1371/journal.pcbi.0040027>.
- [3] A. Berg, and M. Nelimarkka, "Do you see what I see? Measuring the semantic differences in image-recognition services' outputs," *Journal of the Association for Information Science and Technology*, no. 74(11), pp. 1307-1324, 2023. <https://doi.org/10.1002/asi.24827>.

[4] T. Araujo, I. Lock, and B. van de Velde, “Automated Visual Content Analysis (AVCA) in Communication Research: a Protocol for Large Scale Image Classification with Pre-Trained Computer Vision Models,” *Communication Methods and Measures*, no. 14(4), pp. 239-265, 2020. <https://doi.org/10.1080/19312458.2020.1810648>.

[5] Ethan Shafer, Jesse Wood, Sheyla Street, Enoch Crow, Calvin Lu, “Social Bias and Image Tagging: Evaluation of Progress in State-of-the-Art Models,” in book *Advances in Visual Computing* https://doi.org/10.1007/978-3-031-47966-3_26.

[6] *How To Calculate the Confidence Interval*. [Electronic resource]. Available: <https://www.indeed.com/career-advice/career-development/how-to-calculate-confidence-interval>.

[7] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. Da Silva, “A comparative analysis of object detection metrics with a companion open-source toolkit,” *Electronics (Switzerland)*, no. 10 (3), pp. 1-28, 2021. <https://doi.org/10.3390/electronics10030279>.

Рекомендована кафедрою комп'ютерних систем управління ВНТУ

Стаття надійшла до редакції 13.12.2023

Шульгін Олександр Якимович — студент факультету інформаційних і прикладних технологій, e-mail: a.shulhin@gmail.com.

Донецький національний університет імені Василя Стуса, м. Вінниця;

Штовба Сергій Дмитрович — д-р техн. наук., професор, професор кафедри інформаційних технологій Донецького національного університету імені Василя Стуса, м. Вінниця; професор кафедри комп'ютерних систем управління Вінницького національного технічного університету, e-mail: shtovba@vntu.edu.ua

O. Ya. Shulhin¹
S. D. Shtovba^{1,2}

Consolidated Ranking of Modern Cloud-Based Online Object Recognition Services on Images

¹Vasyl' Stus Donetsk National University, Vinnytsia;

²Vinnytsia National Technical University

Automating the object recognition in images is a widely encountered task with evident practical applications in industrial production, medicine, transportation, security, safety, and other fields. Today, there are several cloud services that offer online tools for solving various image recognition tasks. They have several advantages over traditional tools. Today, there are no methods that allow users to identify which cloud service is best suited for their tasks. Moreover, the identification should be concise and based on small set of profile images. Accordingly, a typical brute-force method that requires uploading and analyzing a large number of images is unacceptable. The method should be based on a detailed analysis of the recognition results using a small set of test images, taking into account the features of cloud services. This paper proposes a method for ranking cloud services using small test datasets. In this case, the user forms test datasets that take into account the profile of his or her object recognition tasks. The proposed method is based on three particular metrics, each of which takes into account some features of cloud services. The first metric is the difference between the sum of the confidence levels of correctly recognized objects and the sum of the confidence levels of incorrectly recognized objects. The second metric is the accuracy. The third metric is the median of the length of the service output until the first misrecognized object. The objects should be sorted in descending order of confidence. The first two metrics are traditional, the third metric is new. The final decision is made based on a consolidated score that aggregates the 3 metrics. The application of the method is illustrated by the task of ranking the following cloud services: Microsoft Azure AI Vision Studio, Amazon Rekognition, Google Cloud Vision, and Imagga.

Keywords: object recognition, cloud services, metric, ranking, images.

Shulhin Oleksandr Ya. — Graduate Student of the Department of Information and Applied Technologies, e-mail: a.shulhin@gmail.com ;

Shtovba Serhiy D. — Dr. Sc. (Eng.), Professor, Professor of the Chair of Information Technology of Vasyl' Stus Donetsk National University; Professor, of the Chair of Computer Control Systems of Vinnytsia National Technical University, e-mail: shtovba@vntu.edu.ua