

МОДЕЛІ ТА МЕТОДИ ТРАНСФЕРНОГО НАВЧАННЯ ДЛЯ ОБРОБКИ ЗОБРАЖЕНЬ СЦЕН НА МОБІЛЬНИХ ПРИСТРОЯХ

Вінницький національний технічний університет

Анотація

Розглянуто задачу класифікації сцен на мобільних пристроях з використанням трансферного навчання. Запропоновано метод дистиляції знань з глибоких нейронних мереж у нечіткі класифікатори, що дозволяє ефективно визначати візуальні зв'язки між об'єктами при низьких обчислювальних витратах.

Ключові слова: класифікація сцени, виявлення об'єктів, визначення візуальних зв'язків, дистиляція знань, гранульований нечіткий класифікатор на основі прототипу.

Abstract

The problem of scene classification on mobile devices using transfer learning is considered. A method of distilling knowledge from deep neural networks into fuzzy classifiers is proposed, which allows for effective determination of visible relationships between objects at low computational costs.

Keywords: scene classification, object detection, visual relationship detection, knowledge distillation, prototype-based granular fuzzy classifier.

Вступ

Люди є свідками подій, які фільмують за допомогою мобільного телефону. Тому виникає задача класифікації фото у галереї мобільного пристрою. Категоризація сцени насамперед потребує розуміння сцени на основі ознак контенту і контексту. Дана задача відноситься до задачі аналізу сцен, яка потребує виявлення множини об'єктів і встановлення візуальних відношень між ними, тобто складання опису сцени [1, 2]. При цьому є відомою інформація про структуру сцен класів, які аналізуються [1]. Успіхи глибокого навчання в області комп'ютерного зору зумовили високу точність розпізнавання об'єктів. Однак продуктивність розпізнавання сцен все ще недостатня через складні конфігурації, що формалізуються контекстними зв'язками між об'єктами [3].

За останні роки, алгоритми розпізнавання сцен зазнали важливих змін у результаті нових архітектурних рішень при проектуванні глибоких згорткових нейронних мереж (DCNN), спрямованих на виявлення визначальних областей, кодування ознак, навчання шаблонів просторової компоновки або прототипів [3, 4]. Проте розгортання DCNN на пристроях з обмеженими обчислювальними ресурсами є проблематичним [5]. В той же час, в теорії гранулярних обчислень [6, 7], об'єкти на зображеннях можуть розглядатись як гранули, а прототипи — як гранулярні структури, що об'єднують гранули за допомогою нечітких правил. В цьому випадку візуальні фрази на зображенні сцени описуються матрицями нечітких відношень «гранули — прототипи», а задача встановлення візуальних відношень потребує рішення системи нечітких реляційних рівнянь (SFRE) [8]. Дистиляція знань із DCNN у гранулярні нечіткі класифікатори, що формалізують контекстні зв'язки при описі сцен [9, 10], є можливим рішенням для обробки зображень на мобільних пристроях. Тому є доцільним розробка методу класифікації сцен шляхом дистиляції знань у прототип-орієнтовану систему гранулярних нечітких правил, що дозволить понизити обчислювальну складність із збереженням точності розпізнавання.

Основна частина

Відповідно до [9], у реальних складних сценаріях сцена розглядається як набір атракторів уваги, які містять семантично конгруентні, тобто релевантні та відповідні, та неконгруентні об'єкти. Зображення містять як очікувані в даному контексті об'єкти, так і об'єкти, що порушують семантичні правила. Порушення семантичних правил не заважає людині точно розпізнавати сцени. У моделях, натхнених людиною, просторова увага передбачає виявлення семантичних відносин через

послідовність включень і виключень об'єктів відповідно до встановлених шаблонів або прототипів сцен. Тому образи сцени класифікуються шляхом розв'язання зворотної задачі, яка полягає у реконструкції візуальних зв'язків із текстового опису структури сцени певного класу [11].

При використанні на пристроях з обмеженими обчислювальними ресурсами дистиляція знань забезпечує реконструкцію та класифікацію зображення конкретної сцени. Спочатку навчений детектор об'єктів використовується для розпізнавання семантично залежних компонентів візуальної фрази у вигляді суміжних об'єктів. Потім знання у вигляді мір значущості виявлених об'єктів переносяться в систему нечітких правил IF-THEN [12], де візуальні зв'язки моделюються ієрархічною системою реляційних матриць «гранули (об'єкти) – гранулярні структури – прототипи». Реконструкція сцени передбачає пошук матриць включення об'єктів для формування зернистих структур, які підлягають порівнянню з прототипами в текстовому описі структури сцени певного класу. Після аналізу сцени шляхом відновлення візуальних зв'язків зображення сцени класифікується за допомогою гранульованих нечітких правил на основі прототипу [1].

Проблема реконструкції сцени математично виражена в термінах нечітких реляційних рівнянь. Нечітке реляційне числення [8, 13] забезпечує потужну теоретичну базу для візуального виявлення зв'язків. Нечіткі реляційні рівняння, які пов'язують міри значущості гранул, гранульованих структур і прототипів, будуються на основі ієрархічного $\max - \min$ і двоїстого $\min - \max$ композиційного правила виведення [14, 15]. Проблема реконструкції сцени полягає в пошуку невідомих реляційних матриць, які можна перевести як набір гранульованих нечітких правил [16]. У нечіткому реляційному численні цей тип задач відноситься до класу обернених задач, які вимагають розв'язування SFRE за допомогою методів оптимізації [13, 15].

Задачу оптимізації реконструкції сцени можна сформулювати так [2, 17]. Для мір значущості виявлених об'єктів слід знайти кількість зернистих структур і вектори їх координат, а також матриці включення для об'єктів і прототипів, які забезпечують мінімальну різницю між теоретичним і експериментальним описом структури сцени окремого класу. Текстове представлення, створене для заданого зображення сцени, не є структурно унікальним. Відповідно до [8, 13] набір рішень SFRE визначається унікальним верхнім рішенням для правила максимальної довжини та нижнім рішенням для набору правил мінімальної довжини. Довжина візуальної фрази визначається набором рішень SFRE з $\max - \min$ композицією, а межі зернистих структур визначаються набором рішень SFRE з подвійним $\min - \max$ композицією [18]. У цьому випадку декомпозиція матриці включення об'єктів забезпечує передачу повідомлень між залежними компонентами візуальної фрази, які мають спільні об'єкти. Декомпозиція матриці включення прототипу формалізує перехід між семантично незалежними компонентами візуальної фрази.

Візуальні зв'язки відновлюються шляхом навчання заснованої на прототипі нечіткої реляційної нейронної мережі, ізоморфної SFRE [19]. Вхідні дані подвійних $\min - \max$ і $\max - \min$ нечітких нейронних мереж є мірами значущості гранул (об'єктів) і гранульованих структур відповідно. Виходи ієрархічної мережі $\min - \max/\max - \min$ є мірами значущості прототипів, отриманими на основі коефіцієнтів включення гранул і гранульованих структур у текстовому описі сцени. Експериментальні показники значущості прототипів порівнюються з теоретичними показниками значущості в описі структури сцени кожного класу. Відповідно до [19, 20], візуальна фраза формується шляхом повторних прогонів циклів навчання $\max - \min$ мережі для створення набору блоків-кандидатів, які представляють гранулярні структури. Кожен запуск передбачає пошук правила мінімальної довжини. Подвійна $\min$ - $\max$ мережа переналаштована, щоб розширити обмежувальну рамку для конкретної гранульованої структури. Пошук буде продовжено, якщо будуть знайдені нові рамки, що охоплюють гранулярні структури. Потім фрагменти візуальної фрази, пов'язані з одним прототипом у заданій області зображення, об'єднуються в засноване на прототипі правило максимальної довжини та укладаються в об'єднану обмежувальну рамку. Знайдені гранулярні структури додаються до текстового опису шляхом активації ваг прототипів для формування повного представлення сцени для висновку на основі відношень.

Запропонована рекурсивна техніка представляє інтерес для мобільних додатків реального часу. Виявлення візуальних зв'язків за допомогою вирішення SFRE дозволяє спростити процес реконструкції сцен укладання об'єктів - *stacking objects* - об'єкти, що накладаються (вони закривають один одного). Завдяки властивостям набору рішень сцена реконструюється шляхом формування правил мінімальної та максимальної довжини замість відстеження всіх можливих попарних зв'язків між щільно розташованими об'єктами. Точність реконструкції залежить від деталізації прототипів, які можна налаштувати відповідно до вподобань користувача.

Метою дослідження є розробка методу реконструкції сцени шляхом дистиляції знань у гранульовані нечіткі правила на основі прототипу. Метод повинен забезпечувати точне та швидке виявлення візуальних зв'язків під час реконструкції сцен укладання об'єктів. Для досягнення мети необхідно розробити модель представлення сцени на основі SFRE, яка пов'язує міри значущості гранул, гранульованих структур і прототипів за допомогою матриць включення об'єктів; розробити метод навчання реляційної нечіткої нейронної мережі на основі прототипу, ізоморфної SFRE.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Duda R.O., Hart P.E. (1973) Pattern Classification and Scene Analysis. John Wiley & Sons, Hoboken.
2. H. Li, G. Zhu, L. Zhang, Y. Jiang, Y. Dang, H. Hou, P. Shen, X. Zhao, S. A. Ali Shah, M. Bennamoun, Scene Graph Generation: A comprehensive survey, Neurocomputing, Vol. 566, 2024, 127052, <https://doi.org/10.1016/j.neucom.2023.127052>.
3. L. Xie, F. Lee, L. Liu, K. Kotani, Q. Chen. Scene recognition: A comprehensive survey, Pattern Recognition, Volume 102, 2020, 107205, <https://doi.org/10.1016/j.patcog.2020.107205>.
4. X. Wang, Z. Zhu, Context understanding in computer vision: A survey. Computer Vision and Image Understanding, Volume 229, 2023, 103646, <https://doi.org/10.1016/j.cviu.2023.103646>.
5. V. Kamath, A. Renuka, Deep learning-based object detection for resource constrained devices: Systematic review, future trends and challenges ahead, Neurocomputing, Vol. 531, 2023, P. 34-60, <https://doi.org/10.1016/j.neucom.2023.02.006>.
6. Pedrycz W. Granular Computing: Analysis and Design of Intelligent Systems. CRC Press, Boca Roca (2018) <https://doi.org/10.1201/9781315216737>
7. Yao Y. Three-way decision and granular computing, International Journal of Approximate Reasoning, Vol. 103, 2018, P. 107-123, <https://doi.org/10.1016/j.ijar.2018.09.005>.
8. Rotshtein A., Rakytyanska H. Fuzzy Evidence in Identification, Forecasting and Diagnosis, vol. 275. Studies in Fuzziness and Soft Computing. Springer, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-25786-5>
9. Angelov P., Gu X. Deep rule-based classifier with human-level performance and characteristics. Information Sciences. Vol. 463–464, P. 196–213 (2018). <https://doi.org/10.1016/j.ins.018.06.048>
10. Rakytyanska H. (2023). Knowledge Distillation in Granular Fuzzy Models by Solving Fuzzy Relation Equations. In: Pedrycz, W., Chen, SM. (eds) Advancements in Knowledge Distillation: Towards New Horizons of Intelligent Systems. Studies in Computational Intelligence, vol 1100, pp. 95-133. Springer, Cham. https://doi.org/10.1007/978-3-031-32095-8_4
11. Li, H., Zhu, G., Zhang, L., Jiang, Y., Dang, Y., Hou, H., Shen, P., Zhao, X., Ali Shah, S.A., Bennamoun, M.: Scene graph generation: A comprehensive survey. Neurocomputing 566, art. no. 127052 (2024). <https://doi.org/10.1016/j.neucom.2023.127052>
12. Rakytyanska, H.: Knowledge distillation in granular fuzzy models by solving fuzzy relation equations. In: Pedrycz, W., Chen, S. (eds.) Advancements in Knowledge Distillation: Towards New Horizons of Intelligent Systems. Studies in Computational Intelligence. vol. 1100, pp. 95–133. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-32095-8_4
13. Peeva, K., Kyosev, Y.: Fuzzy Relational Calculus. Theory, Applications and Software. World Scientific, New York (2004)
14. Rakytyanska, H.: Inverse inference based on interpretable constrained solutions of fuzzy relational equations with extended max–min composition. Soft Computing 28, 5461–5478 (2024). <https://doi.org/10.1007/s00500-023-09301-7>
15. Rotshtein, A., Rakytyanska, H.: Fuzzy Evidence in Identification, Forecasting and Diagnosis. Studies in Fuzziness and Soft Computing, vol. 275. Springer, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-25786-5>
16. Rotshtein, A., Rakytyanska, H.: Fuzzy genetic object identification: multiple inputs multiple outputs case. In: Hippe, Z., Kulikowski, J., Mroczek, T. (eds.) Human-Computer Systems Interaction. Part II. Advances in Intelligent and Soft Computing. vol. 99, pp. 375–394. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23172-8_25
17. Rotshtein, A., Rakytyanska, H.: Fuzzy logic and the least squares method in diagnosis problem solving. In: Sarma, R. (ed.) Genetic Diagnoses. pp. 53–97. Nova Science Publishers, New York (2011)
18. Rakytyanska, H.: Classification rule hierarchical tuning with linguistic modification based on solving fuzzy relational equations. Eastern-European Journal of Enterprise Technologies 1(4), 50–58 (2018). <https://doi.org/10.15587/1729-4061.2018.123567>
19. Rakytyanskaya, A., Rotshtein, A.: Fuzzy relation-based diagnosis. Automation and Remote Control 68(12), 2198–2213 (2007). <https://doi.org/10.1134/S0005117907120089>

Прус Богдан Вікторович — аспірант кафедри програмного забезпечення, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: bohdan.prus.vntu@gmail.com.

Ракитянська Ганна Борисівна — канд. техн. наук, доцент кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, rakit@vntu.edu.ua.

Prus Bohdan V. — Ph. D. student, Faculty for Information Technologies and Computer Engineering,, Vinnytsia National Technical University, Vinnytsia, bohdan.prus.vntu@gmail.com.

Rakytyanska Hanna B. — Cand. Sc. (Eng), Assistant Professor of the Software Department, Vinnytsia National Technical University, Vinnytsia, rakit@vntu.edu.ua.