НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
Факультет прикладної математики
Кафедра прикладної математики

Пояснювальна записка до курсового проекту
із дисципліни
«Алгоритми і системи комп'ютерної математики»
на тему
«Автоматична аннотація зображень за допомогою нейронних мереж»

Виконав:
студент групи КМ-01
Скорденко Д. О.

Керівник:
асистент кафедри ПМА
Ковальчик-Химюк Л. О.

Київ --- 2024

АНОТАЦІЯ

В даній роботі описано мультимодальну систему маркування зображень, в якій зроблено акцент на трьох апспектах: висока точність, використання тексту в якості додаткової інформації, явна підсистема для передбачення к-сті лейблів. Всі ці рішення значно підвищують точність в порівнянні із існуючими рішеннями.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

* DNN – Глибинна нейронна мережа (Deep Neural Network)
* CNN – Згорткова нейронна мережа (Convolutional Neural Network)
* RNN – Рекурсивна нейронна мережа (Recursive Neural Network)
* Анотація зображень, маркування зображень, мульти-лейбел класифікація – взаємозамінні поняття

ЗМІСТ

1 ВСТУП

Задача класифікації – це одна із основних задач в аналізі зображень, вона полягає у присвоєнні кожному зображенню один із класів. Таким чином дане формулювання накладає обмеження – зображення містить тільки один об'єкт. Поява DNN [3] та її подальший розвитком у CNN [12, 11] разом із створенням великих датасетів як-от ImageNet [4] дало змогу вирішувати задачу класифікації зображень значно швидше і якісніше ніж люди.

Зрозуміло, що зображення – це той тип даних, який у абсолютній більшості випадків містить більше одного об'єкта. Для поглиблення опису існує задача маркування зображень (image labeling). На відміну від класифікації, вона полягає у маркуванні зображення більше ніж одним класом. Таким чином якість опису зображення кратно зростає у порівнянні із звичайною класифікацією, однак привносить декілька складних завдань.

По-перше, наявність декількох класів у одного зображення створює можливість описувати значно ширший спектр візуальної інформації: різні об'єкти, стилі, дії, і тд. Поява великих хостингів зображень таких як Imgur, Flickr, та ін., де користувачі можуть як завантажувати різноманітні зображення, так і додавати до них описову інформацію у вигляді тегів / анотацій, дала змогу створити досить різноманітні датасети: ImageNet [4], MS-COCO [10], NUS-WIDE [2], та ін.

По-друге, анотація зображень передбачає не лише маркування більше ніж одним класом, а і передбачення к-сті класів. Для опису зображенням із широким спектром понятть необхідно N класів, для зображення із простим вмістом – 2-3 класи.

По-третє, анотація зображень потребує оцінки якості проведеного маркування. Оскільки будь який датасет буде містити в собі дизбаланс класів в тій чи іншій мірі, важливо оцінювати маркування із урахуванням цього.

Все це робить задачу маркування зображення досить складною.

2 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ

Базове рішення

Базовим рішенням для більшості робіт із маркування зображення є використання CNN. Більшість робіт використовує різні архітектури ResNet [6], AlexNet [1], GoogleNet [13]. Спільним між ним є те що вони вже натреновані на великому датасеті, здебільшого ImageNet [4]. Для адаптації моделі до обраного контексту така модель дотреновуєтсья (fine tune), замінюючи базовий класифікатор на такий же простий із адаптованою к-стю вихідних класів [5], або ж на більш складний класифікатор (який надає більш точні результати) [16]. Це працює завдяки тому, що всі архітектури сучасних CNN моделей є багатошаровими, і в них перші шари розпізнають базові особливості (features) зображення, які можна навіть візуалізувати, однак останні шари вивчають більш глибинні особливості зображення, таким чином роблячи модель більш універсальною при зміні класифікатора.

Додаткова інформація

Більш нові роботи також розглядають додавання сторонньої інформації для класифікації зображень. Існує два основних підходів:

а) Семантичний аналіз лейблів. Даний підхід аналізує зв'язок між різними класами. Схожі за контекстом лейбли знаходяться поруч (наприклад: риба, вода) [7, 9]

б) Аналіз додаткової інформації. Даний підхід аналізує додаткову до зображення інформацію. Це може бути як текстова інформація (теги / анотації) [18], так і метадані зображення [8, 14]

К-сть лейблів

Всі наведені вище роботи розглядаєють задачу вибору к-сті лейблів як найкращі k (top k) маркувань. k найбільше ймовірних класів, де k - наперед задана константа. Очевидно, що такий вибір к-сті класів не є оптимальним,

так як більш змістовні зображення будуть містити менше описової інформації і навпапки - менш змістовні будуть містити лишню інформацію, яка до того ж може не мати нічого спільного із цим зображенням (Рис.2.1)

Рис. 2.1 – Приклад адаптивної к-сті лейблів

Один із сучасних підходів як-от CNN-RNN [15], розглядає задачу маркування як задачу перекладу зображення в текст (image to text), де CNN - це кодувальник (encoder), а RNN (decoder) автоматично виконує як задачу маркування, так і задачу динамічного вибору кількості лейблів, однак є певні обмеження накладенні на порядок класів.

3 МОДЕЛЮВАННЯ

На основі проведного аналізу альтернатив, дана робота пропонує розглянути мультимодальну систему, яка складається із чотирьох компонентів (Рис.3.1)

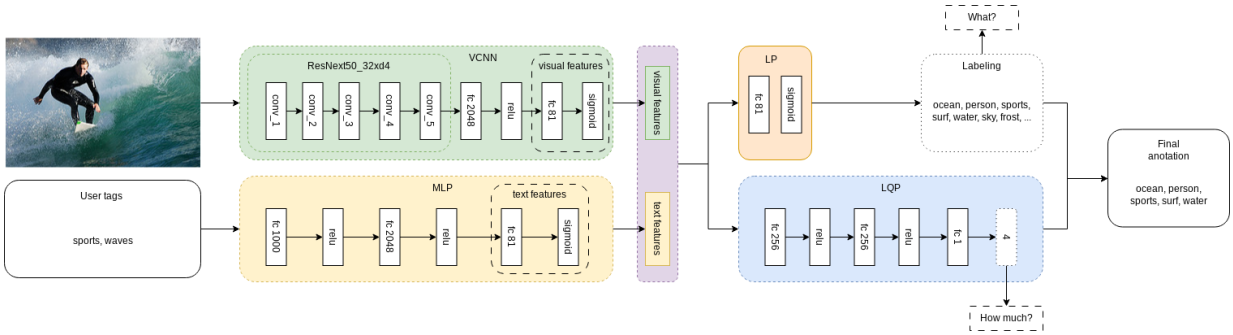

Рис. 3.1 – Архітектура композитної системи

3.1 VCNN

Дана модель призначена для вивчення особливостей (features) із зображення. Отримує на вхід піселі зображення I, у формі матриці розмірності (B,C,W,H), де B - к-сть зображень у групі для тегування, C - к-сть каналів у зображеннях зазвичай 1 або 3, Grey або RGB відповідно, W,H - розмірність зображень.

VCNN - це композитна модель, яка використовує за базове рішення ResNext50_32xd4 [17] (сучасна версія resnet).

3.2 MLP

Дана модель аналізує текстові особливості (text features) тегів до зображення. Теги до зображення i репрезентуються як $I =$

3.3 LP

3.4 LQP

ВИСНОВКИ

ПЕРЕЛІК ПОСИЛАНЬ

[1] Krizhevsky Alex, Sutskever Ilya та Hinton Geoffrey. „ImageNet Classification with Deep Convolutional Neural Networks“. В: *Advances in Neural Information Processing Systems*. 3а ред. F. Pereira та ін. Т. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[2] Tat-Seng Chua та ін. „NUS-WIDE: A Real-World Web Image Database from National University of Singapore“. В: *Proc. of ACM Conf. on Image and Video Retrieval (CIVR'09)*. Santorini, Greece., July 8-10, 2009.

[3] Dan Cireşan, Ueli Meier та Juergen Schmidhuber. „Multi-column Deep Neural Networks for Image Classification“. В: (2012). arXiv: 1202.2745.

[4] Li Deng та ін. „Imagenet: A large-scale hierarchical image database“. В: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, С. 248—255.

[5] Yunchao Gong та ін. *Deep Convolutional Ranking for Multilabel Image Annotation*. 2013. eprint: arXiv:1312.4894.

[6] Kaiming He та ін. *Deep Residual Learning for Image Recognition*. 2015. eprint: arXiv:1512.03385.

[7] Hexiang Hu та ін. *Learning Structured Inference Neural Networks with Label Relations*. CVPR 2016. 2016.

[8] Justin Johnson, Lamberto Ballan та Li Fei-Fei. *Love Thy Neighbors: Image Annotation by Exploiting Image Metadata*. ICCV 2015. 2015.

[9] Qing Li та ін. *Learning Category Correlations for Multi-label Image Recognition with Graph Networks*. 2019. eprint: arXiv:1909.13005.

[10] Tsung-Yi Lin та ін. „Microsoft COCO: Common Objects in Context“. В: *CoRR* abs/1405.0312 (2014). arXiv: 1405.0312. URL: http://arxiv.org/abs/1405.0312.

[11] Keiron O'Shea та Ryan Nash. *An Introduction to Convolutional Neural Networks*. 2015. eprint: arXiv:1511.08458.

[12] Karen Simonyan та Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014. eprint: arXiv:1409.1556.

[13] Christian Szegedy та ін. *Going Deeper with Convolutions*. 2014. eprint: arXiv:1409.4842.

[14] Kevin Tang та ін. *Improving Image Classification with Location Context*. 2015. eprint: arXiv:1505.03873.

[15] Wei Wang та ін. „CNN-RNN: A Unified Framework for Multi-Label Image Classification“. В: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Черв. 2016.

[16] Yunchao Wei та ін. „CNN: Single-label to Multi-label“. В: (2014). DOI: 10.1109/TPAMI.2015.2491929. eprint: arXiv:1406.5726.

[17] Saining Xie та ін. „Aggregated residual transformations for deep neural networks“. В: (листоп. 2016). arXiv: 1611.05431 [cs.CV].

[18] Fengtao Zhou, Sheng Huang та Yun Xing. *Deep Semantic Dictionary Learning for Multi-label Image Classification*. AAAI 2021. 2021.