

ПІДХІД ДО МАРКУВАННЯ ЗОБРАЖЕНЬ

Остапенко М.С., Штогріна О.С.

Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна

E-mail: mak7.ostapenko@gmail.com

Approach to marking images

In this paper propose the method to speed up labeling of the image for creating new datasets. We need datasets for training convolutional neural networks that solve the various task of computer vision for example: image classification, object detection, and localization.

Із розвитком технологій та зростанням потреб людства, постає необхідність збільшення автономності виконання виробничих процесів у всіх галузях народного господарства: від охорони здоров'я до транспорту і зв'язку. Для виконання більшості дій, що виконуються людиною, необхідне візуальне сприйняття світу. Тому не можливо повністю автоматизувати виробничі процеси без використання алгоритмів комп'ютерного зору. На даний момент вже існують алгоритми по розпізнанню об'єктів на зображеннях, що мають більшу точність від людини не залежно від критеріїв оцінювання. Їх використовують у системах автоматичного керування транспортними засобами, аутентифікації, аналізу емоційного стану людини, первинної діагностики хвороб.

Найбільш ефективними алгоритмами комп'ютерного зору для розпізнавання об'єктів є алгоритми, що базуються на штучних нейронних мережах (НМ). Проблема використання таких алгоритмів, полягає в необхідності мати велику кількість зображень для їх навчання. Людині достатньо один раз показати об'єкт і після цього вона зможе його з високою точністю розпізнавати незалежно від освітлення, кута зору, кольору та інших умов. Для навчання НМ потрібно подавати на вхід сотні варіантів одного й того ж або схожих зображень зроблених в різноманітних умовах. Після такого навчання НМ знаходить приховані залежності між зображеннями, та набуває можливості здійснювати бінарну класифікацію, тобто відрізнити об'єкт, на який вона була навчена, від усіх інших. Для покращення процесу навчання використовуються розмічені зображення, в яких окрім самого зображення, присутня інформація з описом об'єкта.

Існуючі програмні інструменти забезпечують зручний процес маркування вмісту зображень: об'єктів, їхнього розміщення та взаємодії між собою. Опис виконується в залежності від поставленої задачі, яку потрібно вирішити, це може бути розпізнавання, локалізація, класифікація, сегментація. Разом розмічені зображення будуть являти собою набір даних, що складається із самих зображень та файлу опису вмісту кожного зображення.

Серед розробників найбільш широко використовуються два засоби маркування зображень Sloth [1] та LabelImg [2]. Sloth надає гнучкі можливості

опису вмісту зображень, за рахунок налаштування процесу маркування кожним окремим користувачем. Він підтримує всі формати зображень і дозволяє представляти файл опису зображення у форматах XML, JSON, CSV. LabelImg є зручним у використанні, але у порівнянні із Sloth не надає можливості індивідуального налаштування процесу маркування і опис зображення формують тільки у форматі XML.

На даний момент не існує достатньої кількості розмічених зображень для навчання НМ, за винятком, найбільш популярних та комерціалізованих областей використання. А вище описані інструменти надають можливість тільки ручного маркування. Здійснюваний людиною процес маркування є дуже повільним, що призводить до неможливості якісно розмітити потрібну кількість зображень. Отже актуальним завданням є розробка підходу до маркування зображень, який би включав попереднє автоматизоване маркування, з можливістю подальшого ручного корегування отриманої розмітки.

Для автоматизації процесу маркування зображень пропонується використовувати алгоритми, що виконують класифікацію, локалізацію об'єктів та описують їх взаємодію між собою на зображенні.

Для класифікації об'єктів на зображенні використана переднавчена нейронна мережа ResNet [3]. На відміну від задачі класифікації об'єктів, задача локалізації вимагає точного встановлення положення об'єкта на зображенні, отримання координат його центру, ширини та довжини обмежувального регіону. Для локалізації об'єктів на зображенні пропонується використовувати один із алгоритмів YOLO [4] або Faster R-CNN [5].

YOLO (You only look once) розділяє вхідне зображення на комірки. Для кожної комірки передбачує вірогідність, того що вона належить до певного класу і далі об'єднує комірки в обмежувальні блоки. Використання тільки однієї згорткової нейронної мережі, що виявляє всі об'єкти на зображенні лише за одне поширення зображення через мережу, надає значну перевагу у швидкості. За рахунок цього YOLO одночасно виконує задачу локалізації та класифікації. Його можна використовувати для обробки відео потоків та відстеження об'єктів. Але його недоліком є точність віднесення шуканого об'єкту до певного класу у випадках коли об'єкт займає відносно малу площу по відношенню до всього зображення.

Faster R-CNN для локалізації використовує детектор на основі вікна, що рухається (sliding window detector). Це допомагає переформулювати задачу локалізації в задачу класифікації в регіоні, що надає високу точність, але вимагає більше обчислювальних ресурсів. Faster R-CNN являє собою суму детектора та класифікатора. Алгоритм Faster R-CNN є точнішим за YOLO, але більш повільнішим, що не дає змоги працювати ефективно з відео потоком, а при роботі із зображеннями може виникати короточасна затримка.

Описані алгоритми є доступними для загального використання. Їх можна завантажувати і для більшої точності дотреновувати нейронні мережі, на яких

вони базуються, на власних даних. Також алгоритми можна завантажити разом із навченими параметрами та зразу використовувати.

Запропонований підхід до маркування зображень складається із кроків:

1. Налаштування середовища маркування, для отримання необхідних інструментів та обрання алгоритму автоматизації, відповідно до заданих параметрів.

2. Автоматизоване маркування зображення, для отримання його первинного опису, що включає класи об'єктів, їх ключові точки та обмежувальні блоки.

3. Ручне корегування опису зображення, отриманого на попередньому кроці.

4. Збереження файлу опису зображення у зазначеному форматі.

Для налаштування середовища маркування, перед початком роботи, вказується тип задачі, що виконується, це може бути або класифікація, або локалізація. Тип задачі визначає вихідні параметрів. Для задачі класифікації вихідним параметром є лише назва класу до якого належить об'єкт, що знаходиться на зображенні. Для задачі локалізації вихідними параметрами є назва класу об'єкту, координати центру, ширина та довжина обмежувального блоку навколо об'єкта. Також вхідним параметром є множина назв класів можливих об'єктів, що можуть бути присутніми на зображеннях. Використовуючи назви класів алгоритми автоматизованого маркування будуть їх виділяти серед всіх інших об'єктів на зображенні. Для збереження вихідного файлу опису зображень потрібно вибрати його формат, який придатний для подальшого використання.

Запропонований підхід дозволяє підвищити швидкість маркування зображень, для подальшого їх використання при навчанні або тестуванні НМ.

Подальша робота буде полягати у підвищенні точності і швидкості класифікацій та локалізації об'єктів, за рахунок зміни кількості шарів НМ, використання технік зменшення розмірності або інших архітектур НМ. Також планується додати можливості маркування зображень, для задачі сегментації.

Література

1. Computer Vision for Human Computer Interaction Karlsruhe Institute of Technology. Sloth [Електронний ресурс] – режим доступу: <http://sloth.readthedocs.io/en/latest/>, вільний.
2. Tzu Ta Lin. LabelImg [Електронний ресурс] – режим доступу: <https://github.com/tzutalin/labelImg>, вільний.
3. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, arXiv:1512.03385, 2015.
4. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, arXiv:1506.02640, 2015.
5. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, arXiv:1506.01497, 2015.