Жмакін А.О.
студент,
Чернівецький національний університет
імені Юрія Федьковича

СЕМАНТИЧНА СЕГМЕНТАЦІЯ ЗОБРАЖЕНЬ НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Автоматичний аналіз зображень на основі методів машинного навчання в нинішній момент переживає нове народження і поступово входить в усі галузі людської діяльності. При цьому складні і важливі завдання аналізу зображень вирішуються методами машинного навчання, які застосовуються: на безпілотних автомобілях з використанням камер і радарів для правильної взаємодії суб'єкта з об'єктом (наприклад, у Google, Uber тощо); у військових системах по виявленню об'єктів з використанням високоточних камер; на вивісках, які автоматично виконують визначення статі людини, що проходить через зону дії, з наступною демонстрацією таргетованої реклами; на МРТ і рентгенівських знімках з автоматичним розпізнаванням аномалій.

На сьогодні, позитивних результатів при вирішенні задач автоматичної сегментації зображень досягли згорткові нейронні мережі, архітектуру яких запропонував Ян Лекун [1].

Завдання розбиття зображення (семантичної сегментації) полягає в знаходженні на зображенні областей (сегментів), а також їх класифікація по заздалегідь заданому набору класів. На даний момент не існує такого універсального методу для вирішення такого завдання, тому найчастіше вибір методу ґрунтується на предметнІй області, в якій ця задача ставиться. При цьому рішення часто шукають побудовою умовного Марківського випадкового поля і оптимізацією відповідної функції енергії [1–3]. Також використовують локальні властивості пікселів і регіонів зображень (таких як колір або текстура), а з метою поліпшення якості сегментації застосовується глобальна або контекстна інформація (наприклад, абсолютне [2] або відносне [4] розташування об'єктів різних класів, або локальні параметри сусідніх регіонів [5]).

До методів машинного навчання при семантичній сегментації зображень відноситься метод опорних векторів (SVM), який використовується для класифікації та регресійного аналізу. Даний метод передбачає машинне навчання з учителем. Також до методів машинного навчання з учителем відносяться випадковий ліс (Random Forest) і логістична регресія (Logistic Regression) [2; 3]. У цілому, принципи роботи класичних методів машинного навчання при семантичній сегментації зображень полягають: у фіксуванні кількості найближчих сусідів, ознаки яких будуть використовуватися для класифікації кожного пікселя зображення; у визначенні ознак пікселя, які будуть використовуватися, наприклад значення RGB; у складанні навчальної вибірки, в якій кожному пікселю (і його мітці класу), для якого потрібно передбачити клас буде відповідати вектор ознак, що складається з ознак сусідів

пікселя; у визначенні вектора ознак для пікселів, які знаходяться на границі зображення, з необхідністю в доповненні до необхідної розмірності за допомогою заповнення значення; в навчанні однієї або декількох моделей (в залежності від підходу). Але класичним методам машинного навчання характерна обчислювальна неефективність через необхідність проходження вікнами по зображенню і найгірша точність семантичної сегментації в порівнянні з методами, заснованими на нейромережевих моделях.

Методи для вирішення багатьох завдань семантичної сегментації при машинному навчанні, що використовують згорткові нейронні мережі є «state of the art». Досягненнями переваг за якістю згорткової нейронної мережі є автоматичне виділення ознак зображень (і мережа робить таку процедуру набагато краще, ніж фахівці люди), а також збільшені потужності і можливості комп'ютерних систем (зокрема поява потужних GPU), які дозволяють працювати згортковим нейромережевим архітектурам, виконувати їх різноманітне навчання, яке раніше було неможливим. На даний момент, на навчання нейромережевих моделей витрачається незначний період часу з можливим застосуванням в режимі реального часу навіть на мобільних пристроях.

Згорткові нейронні мережі дозволяють відійти від обчислювально-неефективної процедури ковзного вікна на користь архітектури енкодера-декодера, яка працює з усією картинкою цілком, а не з обмеженою її областю.

Прийнято вважати [6], що згорткові нейронні мережі, в основному, використовуються для обробки зображень. Відомий, досить ефективний спосіб вирішення цього завдання, який повертає на свою користь саму структуру зображення: передбачається, що пікселі, що знаходяться близько один до одного, тісніше «взаємодіють» при формуванні даної ознаки, ніж пікселі, розташовані в протилежних кутах. Крім того, процес класифікації зображення відзначається відсутністю значення, на якій ділянці зображення виявлені ці ознаки.

Існуюча модель SegNet [7] є типовим автокодувальником, який заснований на згортковій нейронній мережі. Мережа даної моделі складається з блоків, а кожен блок вміщує згортки і пулінги (субдискретизуючі шари) або шари, які підвищують дискретизацію (апсемплінг шари), а також активаційні шари ReLU [8] і шари нормалізації (BatchNorm) [8]. Архітектура SegNet є повністю симетричною, крім шару м'якого максимуму (Softmax) на кінці декодера, який виконує перетворення кожного пікселя вихідної матриці в ціле число, що показує клас даного пікселя. Головною відмінністю SegNet від звичайних згортальних автокодувальників є інформаційне з'єднання апсемплінг шарів декодера з відповідними пулінг шарами енкодера. Тобто, навчання апсемплінг шарів мережі не відбувається, а самі шари отримують інформацію про необхідне підвищення розмірності і відновлення стисненої (втраченої) топології, інформація при цьому надходить від відповідних пулінг шарів, які зберігають індекси активованих пікселів (пікселів з найбільшим значенням у вікні).

Також відомі й інші сегментаційні енкодер-декодер моделі, наприклад, Unet, Enet і тощо. Unet [5] добре зарекомендував себе як бейзлайн для вирішення практичних завдань і виконання сегментації, з архітектурою, що містить дві частини: звужуючу (енкодер) і розширюючу (декодер).

Enet [5] включає блоки з нетривіальною структурою, так званої bottleneck (пляшкове горлечко), архітектура якої складається з основної гілки з макс-пулінгом і паддінгом і допоміжної (зі згортками), а результати гілок зливаються і проходять через активацію ReLU.

Слід зазначити, що Enet на відміну від SegNet і Unet, має декодер набагато менше енкодера, оскільки вважається, що завдання виділення ознак побудови латентного простору набагато складніше, ніж відновити маску з гарних ознак. Enet володіє великою глибиною і містить набагато менше параметрів, ніж Unet і SegNet, що добре позначається на швидкості роботи і робить цю модель придатною для сегментації на мобільних пристроях в реальному часі.

Список використаних джерел:

1. Cun Y. Le, et al. Learning Hierarchical Features for Scene Labeling. URL: http://yann.lecun.com/exdb/publis/pdf/farabet-pami-13.pdf (Last accessed: 25.04.2019).

2. Agresti A. Logistic regression. Wiley Online Library, 2002.

3. Liaw A. Classification and regression by randomForest / A. Liaw, M. Wiener et al. // R news. – 2002. – Vol. 2, No. 3. – P. 18–22.

4. The Cityscapes Dataset for Semantic Urban Scene Understanding / M. Cordts, M. Omran, S. Ramos et al. // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

5. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation / A. Paszke, A. Chaurasia, S. Kim, E. Culurciello // CoRR. 2016. Vol. abs/1606.02147. – 1606.02147.

6. Krizhevsky A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, H. Geoffrey E. // Advances in neural information processing systems. 2012. – P. 1097–1105.

7. Badrinarayanan V. Segnet: A deep convolutional encoder-decoder architecture for image segmentation / V. Badrinarayanan, A. Kendall, R. Cipolla // IEEE transactions on pattern analysis and machine intelligence. 2017. – Vol. 39, No. 12. – P. 2481–2495.

8. Ioffe S. Batch normalization: Accelerating deep network training by reducing internal covariate shift / I. Sergey, S. Christian // arXiv preprint. – 2015. – arXiv:1502.03167.