

Технічні науки

УДК 004.9

Малярчук Роман Васильович

студент

Інституту прикладного системного аналізу

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

Малярчук Роман Васильевич

студент

Института прикладного системного анализа

Национального технического университета Украины

«Киевский политехнический институт имени Игоря Сикорского»

Maliarchuk Roman

Student of the

Institute of applied systems analysis of the

National technical university of Ukraine

"Igor Sikorskiy Kyiv Politechnical Institute"

Науковий керівник:

Кухарев Сергій Олександрович

асистент

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

UNET, ЯК РІШЕННЯ ДЛЯ ЗАДАЧІ СЕГМЕНТАЦІЇ

UNET, КАК РЕШЕНИЕ ДЛЯ ЗАДАЧИ СЕГМЕНТАЦИИ

UNET AS A SOLUTION FOR THE SEGMENTATION PROBLEM

Анотація. Огляд повністю згортової нейронної мережі та її архітектурних компонент для задачі сегментації.

Ключові слова: згорткова нейронна мережа, функція втрат, dice коефіцієнт.

Аннотація. Обзор полностью згорткових нейронной сети и ее архитектурных компонент для задачи сегментации.

Ключевые слова: сверточная нейронная сеть, функция потерь, dice коэффициент.

Summary. Overview of a fully convoluted neural network and its architectural components for the segmentation problem.

Key words: convolutional neural network, loss function, dice coefficient.

Нейронна мережа є звичайною комп'ютерною програмою, яка базується на ідеї роботи людського мозку. Власне в самій назві закладено багато сенсу, адже нейронні мережі називають мережами якраз через те, що вони спроектовані, як композиція багатьох функцій, які «навчаються» визначаючи взаємозв'язки та структурну схожість в даних, використовуючи попередні знання.

На сьогоднішній день існує багато типів нейронних мереж, наприклад повнозв'язні нейронні мережі, згорткові нейронні мережі, рекурентні нейронні мережі.

В задачах класифікації зображень найчастіше використовують згорткові нейронні мережі.

Чому не можна було взяти звичайну повнозв'язну нейронну мережу для реалізації задачі сегментації, адже сігмоїдними повнозв'язними нейронними мережами можна наблизити будь-яку обмежену функцію з скінченною кількістю розривів? Відповідь проста – такі мережі погано справляються з даними, які складаються з взаємозалежних компонентів, наприклад, послідовно корелюючих між собою пікселів для картинок. Звичайна нейронна мережа виходячи з того, що вона бачила побудує

маску для об'єкту. Якщо об'єкт буде знаходитися постійно в центрі картинки (наприклад), то мережа не розпізнає цей ж об'єкт в лівому верхньому куті. Звісно, можна взяти величезну вибірку, яка покриватиме всі можливі варіанти розташування об'єкту, всі розміри, кути нахилу, але тоді мережа буде мати велику кількість масок, відповідно потребуватиме великої кількості нейронів, що в свою чергу вимагає значної кількості пам'яті. Оскільки операція згортки інваріантна відносно положення, розміру, куту нахилу, то результат роботи на одному і тому ж об'єкті буде однаковий, просто в різних позиціях. Тож за допомогою згортки вирішується проблема обмеженості обчислювальних ресурсів. Також згортка враховує, які пікселі знаходяться близько один до одного.

Серед найвідоміших згорткових нейронних мереж, такі: LeNet(1998), AlexNet(2012), VGG(2014), GoogLeNet(2015), ResNet(2015), UNet(2015).

Для реалізації задачі сегментації найкраще підходить повністю згорткова нейронна мережа UNet, яка була модифікована таким чином, щоб давати кращу сегментацію медичних зображень[1]. Її архітектура зображена на рисунку 1.

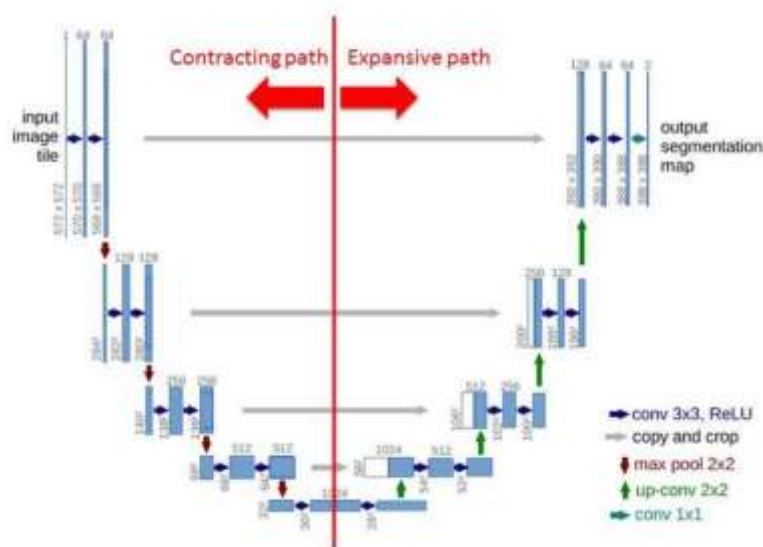


Рис. 1. Архітектура мережі UNet

Серед її основних переваг такі:

– UNet поєднує інформацію про місцеположення зі шляху зменшення розмірності з контекстною інформацією на шляху збільшення розмірності, щоб нарешті отримати загальну інформацію, що поєднує локалізацію та контекст, які необхідні для побудови хорошої сегментації

– Немає звичайного повнозв'язного шару, тому вхідні зображення можуть мати різні просторові розмірності (оскільки єдиними параметрами, які нейронна мережа вивчає, є ядро згортки, а розмір ядра не залежить від розміру вхідного зображення).

– Витискання з даних максимально можливої кількості інформації є важливим у таких сферах, як, наприклад, біомедична сегментація, оскільки кількість анотованих зразків зазвичай обмежена.

Для даної мережі доцільно використати функцію активації ReLU, адже вона має ряд переваг серед яких ефективність в обчисленні, уникнення затухання градієнту на $\pm\infty$, швидша збіжність ніж \tanh , сигмоїди.

Що стосується функції втрат, то вона використовується для оптимізації моделі під час її навчання на тренувальних даних. Від вибору функції втрат залежить як швидко та ґрунтовно вчитиметься нейронна мережа.

Для задач сегментації критерій оцінки якості моделі не сильно відрізняються по принципу побудови. Ми порівнюємо дані, які були розмічені вченими, так звана Ground Truth (Істина), з даними, які були отримані в результаті роботи моделі, так званий Prediction. Що в нашому випадку є Ground Truth? Для кожного з класів ми маємо маску з областями, які належать до відповідного класу. Маска, як вже було сказано складається з пікселів, кожен з яких, означає, належить відповідний піксель оригінального зображення до відповідного класу маски, чи ні. Що в нашому випадку буде Prediction`ом? Для кожного з

класів ми побудуємо маски, призначення яких повністю співпадає з Ground Truth. Тобто нам потрібно якось попарно порівняти всі маски та агрегувати їхні оцінки в одну оцінку. Порівняння повинно відбуватися попіксельно. Виберемо дві функції втрат Dice loss та BCE loss, отримуємо результуючу функцію втрат:

$$loss = BCELoss(y_{pr}, y_{gt}) + DiceLoss(y_{pr}, y_{gt}) \quad (1)$$

Таке рішення було прийнято на основі того, що BCELoss краще мінімізує функцію втрат, а DiceLoss добре відпрацьовує на даних, у яких деякі класи домінують над іншими [2].

BCELoss [3] - критерій, що вимірює бінарну перехресну ентропію між справжнім значенням та результатом роботи моделі.

$$bce(x, y) = -(y \log(x) + (1 - y) \log(1 - x)) \quad (2)$$

Також існує варіант BCE, який називається BCEWithLogits. Її перевага над BCE в тому, що вона на вхід приймає значення з діапазону $[-\infty; \infty]$, на відміну від $[0; 1]$, як в BCE (дані пропускаються через сигмоїду, перед передачею їх в BCE). Функція втрат BCEWithLogits поєднує застосування сигмоїди та BCELoss в одній дії, що в свою чергу сильно зменшує кількість операцій на однакових даних, аніж звичайна BCE.

Dice loss походить від коефіцієнта Соренсена, який є статистикою, розробленою в 1940-х роках для оцінки схожості двох зразків.

Вперше в комп'ютерному зорі цей коефіцієнт був використаний для трьохвимірної сегментації медичних зображень.

Dice коефіцієнт, де p_i та g_i пара пікселів, prediction та ground truth відповідно:

$$D = \frac{2 \sum p_i g_i}{\sum_i^n p_i^2 + \sum_i^n g_i^2} \quad (3)$$

Коли ми маємо справу з масками, то значення p_i і g_i є або 0, або 1, що представляє належність пікселя (значення 1) до класу, чи неналежність

(значення 0). Отже, в знаменнику сума пікселів, які ми визначили, як ті, які належать до класу, так і сума пікселів, які насправді належать до класу. В чисельнику сума правильно передбачених пікселів, оскільки сума зростає лише тоді, коли p_i і g_i однакові і рівні 1.

Схематично цей коефіцієнт зображений на рисунку 2

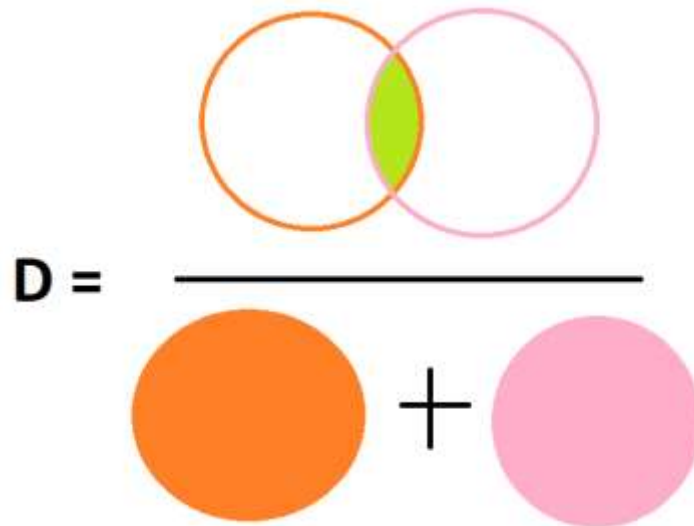


Рис. 2. Dice коефіцієнт з точки зору теорії множин

Тобто з точки зору теорії множин Dice коефіцієнт це міра перетину двох множин. Наприклад, якщо дві множини A і B ідеально перекриваються одна одною, то dice коефіцієнт набуває максимального значення 1. В іншому випадку dice коефіцієнт починає знижуватися, аж до мінімального значення, яке рівне 0, якщо ці множини взагалі не перетинаються. Тому діапазоном dice коефіцієнту є $[0; 1]$. Таким чином, ми можемо використовувати $1-D$ як функцію втрат для максимізації перекриття між двома множинами.

Отже, для задачі сегментації можна використати таку комбінацію архітектурних компонент для побудови нейронної мережі: за основу взяти мережу UNet, в якості функції активації взяти ReLU, за функцію втрат взяти суму критеріїв $\text{dice} + \text{bce}$.

Література

1. Роннебергер О., Фішер П., Брокс Т. U-Net: Світкові мережі для сегментації біомедичних зображень. Кафедра інформатики та Центр біологічних сигналів BIOSS, Університет Фрайбурга, Німеччина. 2015 Р. 1-4. URL: <https://arxiv.org/pdf/1505.04597.pdf> (дата зверення: 01.05.2020)
2. Чен К.Ф., Денг Дж. Ф., Статистичні атласи та обчислювальні моделі серця. Змагання ACDC та MMWHS: 8-й міжнародний семінар, STACOM 2017, що проводиться спільно з MICCAI 2017, Квебек, Канада, 10-14 вересня 2017 року. Springer. 2018 Р. 135, 145.
3. Арай К., Бхатія Р., Капур С. Матеріали конференції технологій майбутнього (FTC) 2019: Вип. 1. Springer Nature. 2019 Р. 279.

References

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany. 2015 P. 1-4. URL: <https://arxiv.org/pdf/1505.04597.pdf> (дата зверення: 01.05.2020)
2. Chien C.F., Deng, J.F. Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017. Springer. 2018 P. 135, 145.
3. Arai K., Bhatia R., Kapoor S. Proceedings of the Future Technologies Conference (FTC) 2019: Vol. 1. Springer Nature. 2019 P. 279