

Бучко О. А., Плахотна Д. О.

ПОРІВНЯННЯ АРХІТЕКТУР НЕЙРОННИХ МЕРЕЖ ДЛЯ СЕГМЕНТАЦІЇ ПУХЛИН МОЗКУ

У роботі досліджено ефективність сучасних архітектур глибокого навчання для сегментації медичних зображень у задачі виявлення пухлин мозку. Проведено порівняльний аналіз моделей FCN, DeepLabv3+, PSPNet та Attention U-Net. Особливу увагу приділено впливу попереднього самоконтрольованого навчання на якість сегментації в умовах обмеженої кількості розмічених даних. Для оцінювання результатів використано метрику подібності.

Ключові слова: сегментація зображень, глибоке навчання, нейронні мережі, Attention U-Net, BraTS2020, самоконтрольоване навчання.

Вступ

Сегментація зображень є однією з фундаментальних задач комп'ютерного зору, що полягає у поділі зображення на семантично однорідні області. Залежно від типу поставленої задачі, сегментація може мати різні форми: від знаходження фону та об'єкта до точного піксельного виділення багатьох класів. Особливу актуальність ця задача набуває в галузі медицини, де автоматизоване оброблення візуальних даних дає можливість покращити точність діагностики та прискорити процес ухвалення клінічних рішень.

Традиційні алгоритми сегментації, основані на простих евристичних, таких як порогове оброблення, виявлення контурів або кластеризація, мають низьку стійкість до змін у структурі даних, неоднорідності освітлення, шумів або варіативності об'єктів. Їхня ефективність різко падає в умовах, де спостерігається складна морфологія, як, наприклад, у випадках сегментації пухлин головного мозку за МРТ-знімками.

З огляду на це зростає потреба у використанні методів, здатних адаптуватися до різноманітних сценаріїв, урахувати як локальні, так і глобальні закономірності, а також навчатися з великих обсягів розмічених даних. Саме такі властивості реалізуються в сучасних підходах, що базуються на глибоких згорткових нейронних мережах (CNN), які сьогодні становлять основу комп'ютерного зору.

У цій статті представлено результати експериментального порівняння кількох сучасних сегментаційних архітектур глибокого навчання, а також досліджено вплив попереднього самоконтрольованого навчання енкодера на якість сегментації. За мету було поставлено визначення архітектурно-тренувальних факторів, які найбільше впливають на точність, узагальнюваність та обчислювальну ефективність моделі сегментації в умовах обмежених ресурсів та доступу до анотованих даних.

Різновиди сучасних підходів до сегментації зображень

Сучасні підходи до сегментації зображень базуються насамперед на глибоких згорткових нейронних мережах, що забезпечують автоматичне вилучення багаторівневих ознак і точне піксельне розмежування об'єктів [6]. Еволюцію методів доцільно розглядати крізь призму архітектурних парадигм, які розв'язують типові проблеми: низьку роздільну здатність глибоких представлень, втрату просторового контексту та брак анотованих даних у спеціалізованих доменах, зокрема в медицині [7].

Повністю згорткові мережі (англ. fully convolutional networks, FCN) стали першими моделями, що усунули повнозв'язні шари класичних CNN, здійснивши пряме згорткове перетворення зображення у карту класів [10]. Використання транспонованих згорток (англ. upsampling) дозволило відновлювати початкову роздільну здатність, однак через відсутність симетричного декодера FCN схильні до розмивання меж об'єктів, критичних у медичній діагностиці.

Найширше застосування отримала енкодер-декодерна архітектура, зокрема модель U-Net, яка поєднує контракційну гілку з розширювальною та реалізує прямі пропуски (англ. skip connections), що компенсують втрату низькорівневих деталей [1]. Завдяки здатності ефективно навчатися на малих вибірках U-Net стала «золотим стандартом» для медичних томографічних даних [4], зокрема для набору BraTS2020, використаного у нашому дослідженні [2]. Її модифікації, як-от Feature Pyramid Network (FPN) та SegNet, удосконалюють багаторівневе представлення ознак або підвищують ефективність декодування шляхом збереження індексів підсумплювання [5].

Проблема обмеженого сприймального поля без втрати роздільної здатності вирішується за допомогою розширених згорток (англ. dilated convolutions) та модуля Atrous Spatial Pyramid Pooling (ASPP), реалізованого в архітектурі DeepLabv3+ [3]. Остання поєднує ASPP із енкодер-декодером і використовує глибокі сепарабельні згортки, що забезпечує збалансованість між локальною точністю та глобальним контекстом.

Альтернативний спосіб урахування різнорівневого контексту реалізує Pyramid Scene Parsing Network (PSPNet), що агрегує глобальні статистики через пірамідальний пулінг [12]. У високоточних клінічних сценаріях PSPNet дає змогу врахувати складну топологію пухлин, хоча поступається DeepLabv3+ за деталізацією дрібних структур.

Високоточні задачі, які вимагають субпіксельної локалізації, мотивували розроблення High-Resolution Network (HRNet), яка підтримує паралельні гілки з різною дискретизацією та поступовим об'єднанням ознак [11]. Цей підхід дає змогу досягти чіткого збереження геометричних меж, що є критичним при визначенні, наприклад, інвазивних країв пухлин.

Для підсилення дискримінативності було запропоновано механізми уваги (англ. attention mechanisms). Attention U-Net, що використовується в нашій роботі, вводить модулі просторової уваги, які пригнічують нерелевантні регіони й підсилюють семантично значущі [9]. Подальший розвиток призвів до створення гібридних архітектур на основі трансформерів (як-от TransUNet), де self-attention-блоки використовуються для моделювання довготривалих залежностей.

Недостатня кількість розмічених медичних зображень сприяла поширенню самоконтрольованого навчання (англ. self-supervised learning, SSL). Одним із базових підходів є передбачення обертання зображення, тоді як сучасні методи (MoCo, DINO, BYOL) використовують контрастивні втрати для навчання інваріантних до перетворень представлень [7]. У межах нашого дослідження було застосовано самоконтрольоване попереднє навчання енкодера Attention U-Net на основі передбачення кута обертання зображення, що дозволило покращити якість сегментації при обмеженій кількості анотованих даних.

Експериментальна оцінка

Для проведення дослідження було використано відкритий медичний датасет BraTS2020 (Brain Tumor Segmentation Challenge), що містить МРТ-знімки мозку з багатоканальними 3D-об'ємами та відповідними анотованими масками. У дослідженні було задіяно 369 пацієнтів, розділених на тренувальну (70 %), валідаційну (15 %) і тестову (15 %) підвибірку на рівні пацієнтів. Для кожного зразка доступні чотири МРТ-последовності: T1, T2, T1ce та FLAIR, що дають змогу охопити різні морфологічні характеристики пухлинної ділянки.

На етапі підготовки даних було здійснено нормалізацію значень піксельної інтенсивності кожного каналу в діапазоні [0, 1]. Сегментаційні маски, представлені в датасеті BraTS2020 у формі трьох бінарних каналів, де кожен канал відповідає окремому типу пухлинної тканини, було перетворено в одноканальний формат. У цьому форматі кожному пікселю надавалося одне з чотирьох значень: 0 — фон, 1 — некротична ділянка, 2 — набряк, 3 — активна пухлина. Для формування навчального набору було використано двовимірні аксіальні зрізи з тривимірних об'ємів.

Для сегментації було обрано чотири сучасні архітектури згорткових нейронних мереж: FCN, DeepLabv3+, PSPNet та Attention U-Net, реалізовані за допомогою бібліотеки Segmentation Models PyTorch. Обрані архітектури представляють різні підходи до оброблення просторового контексту: ASPP у DeepLabv3+, пірамідальний пулінг у PSPNet та просторові модулі уваги в Attention U-Net.

Мережі було навчено на GPU з використанням оптимізатора Adam, функції втрат у вигляді суми Dice Loss та зваженої CrossEntropy Loss, а також з агресивним застосуванням аугментацій: випадкового повороту, горизонтального відображення, зміни яскравості та випадкового масштабування.

Приклад вхідного зрізу та відповідної маски наведено на рис. 1.

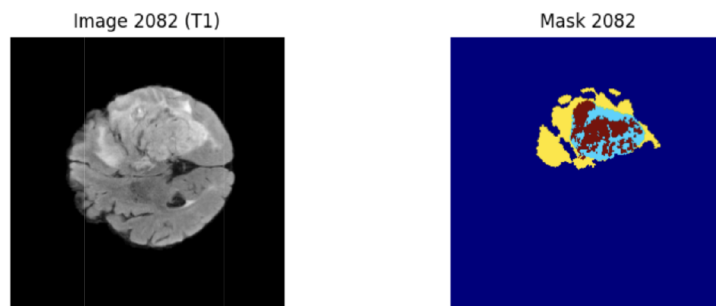


Рис. 1. Приклад МРТ-зрізу (T1) та відповідної сегментаційної маски з датасету BraTS2020

Кількісну оцінку якості сегментації було здійснено за допомогою коефіцієнта подібності кубиків (англ. Dice Similarity Coefficient) [8], визначеного за формулою:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (1)$$

де TP — кількість пікселів, правильно класифікованих як позитивні, FP — кількість хибнопозитивних пікселів, FN — кількість хибнонегативних пікселів. Ця метрика є чутливою до якості локалізації об'єкта та широко використовується у медичних задачах.

Для оцінювання якості сегментації було враховано чотири класи: фон, некроз, перитуморальний набряк і активна пухлина (табл. 1). Для кожної моделі було обчислено середні значення коефіцієнта Dice за кожним класом на тестовій вибірці.

Таблиця 1. Середні значення Dice-коефіцієнта для кожного класу (BraTS2020, тестова вибірка)

Модель	Фон (0)	Некроз (1)	Набряк (2)	Активна пухлина (3)	Середнє значення
FCN	0.975	0.744	0.802	0.763	0.821
Attention U-Net	0.982	0.784	0.843	0.813	0.856
DeepLabv3+	0.978	0.767	0.828	0.796	0.842
PSPNet	0.974	0.719	0.794	0.755	0.811

Як видно з таблиці, Attention U-Net досягла найвищих показників Dice-коефіцієнта як для кожного окремого класу, так і в середньому. Найбільший приріст спостерігався у класах активної пухлини та набряку, що є критичними для клінічної оцінки.

Особливої уваги заслуговує високий показник для класу фону (0) у всіх моделей (>0.97), що свідчить про низький рівень хибнопозитивних спрацювань поза межами пухлинної області. Водночас відносно нижчі значення для некротичної тканини (1) підтверджують, що саме ця ділянка є найскладнішою для сегментації через розмитість меж та внутрішню неоднорідність.

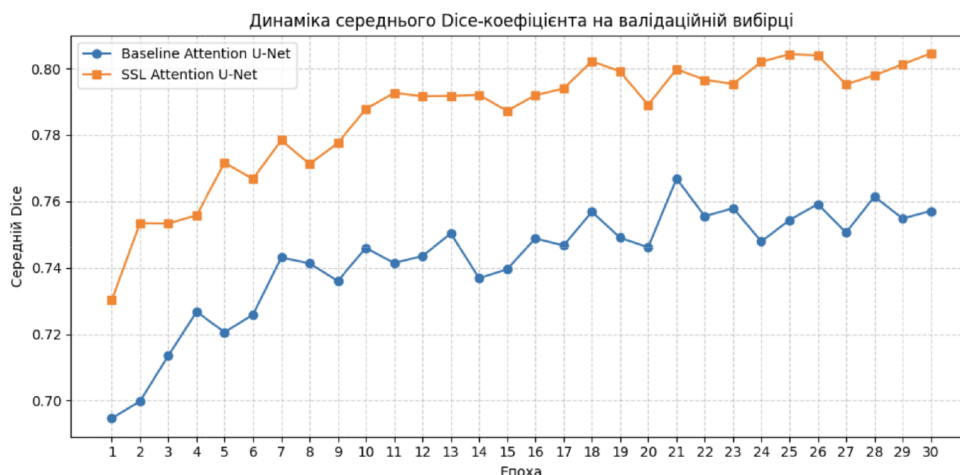


Рис. 2. Динаміка середнього Dice-коефіцієнта на валідаційній вибірці для Attention U-Net з і без самоконтрольованого попереднього навчання

Із метою підвищення узагальнюваності моделі Attention U-Net в умовах обмеженої кількості анотованих прикладів було застосовано попереднє самоконтрольоване навчання енкодера на основі завдання передбачення кута обертання (0° , 90° , 180° , 270°). Це дозволило моделі попередньо навчитися розпізнавати базові структурні особливості зображень без використання анотованих міток.

Після подальшого навчання на задачі сегментації середній Дісе-коефіцієнт зріс з 0.856 до 0.879. Покращення виявлено насамперед у складних класах. Динаміку зміни Дісе-коефіцієнта на валідаційній вибірці протягом 30 епох наведено на рис. 2, що демонструє стабільну перевагу SSL-ініціалізованої моделі над базовою.

Висновки

Із проведеного дослідження можна зробити висновок, що застосування сучасних архітектур глибокого навчання дає змогу досягти високої точності сегментації медичних зображень навіть в умовах обмежених анотованих даних. Найкращі результати було отримано для моделі Attention U-Net, зокрема у варіанті з попереднім самоконтрольованим навчанням, що підтверджує ефективність таких підходів у задачі сегментації зображень.

Переваги таких моделей відображено у кількісних метриках Дісе, які перевищують показники класичних архітектур, а також у візуальній якості сегментацій. Разом із тим слід враховувати обчислювальну складність та вимоги до апаратного забезпечення, що може обмежувати практичне використання деяких моделей у клінічних умовах.

Використання нейромережових методів сегментації, зокрема з урахуванням контексту та попереднього навчання, є перспективним напрямом для подальшого розвитку автоматизованих систем медичної діагностики, але потребує подальшої оптимізації та перевірки на більш різноманітних даних.

Список літератури

1. Badrinarayanan V. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [Electronic resource] / Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2017. — Vol. 39, no. 12. — Pp. 2481–2495. — <https://doi.org/10.1109/tpami.2016.2644615>.
2. BraTS2020 Dataset [Electronic resource]. — Mode of access: <https://www.med.upenn.edu/cbica/brats2020/>.
3. Chen L.-C. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [Electronic resource] / Liang-Chieh Chen [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2018. — Vol. 40, no. 4. — Pp. 834–848. — <https://doi.org/10.1109/tpami.2017.2699184>.
4. Çiçek Ö. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [Electronic resource] / Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. — Cham, 2016. — Pp. 424–432. — https://doi.org/10.1007/978-3-319-46723-8_49.
5. Drozdal M. The Importance of Skip Connections in Biomedical Image Segmentation [Electronic resource] / Michal Drozdal [et al.] // Deep Learning and Data Labeling for Medical Applications. — Cham, 2016. — Pp. 179–187. — https://doi.org/10.1007/978-3-319-46976-8_19.
6. Haralick R. M. Image Segmentation Techniques [Electronic resource] / Robert M. Haralick, Linda G. Shapiro // Computer Vision, Graphics, and Image Processing. — 1984. — Vol. 27, no. 3. — P. 389. — [https://doi.org/10.1016/0734-189x\(84\)90043-4](https://doi.org/10.1016/0734-189x(84)90043-4).
7. Image Segmentation Techniques: A Comprehensive Review [Electronic resource] // International Research Journal of Modernization in Engineering Technology and Science. — 2024. — <https://doi.org/10.56726/irjmets49604>.
8. McGurk R. TH-C-WAB-08: Modeling of the Dice Coefficient for PET Segmentation Studies [Electronic resource] / R. McGurk [et al.] // Medical Physics. — 2013. — Vol. 40, no. 6, part 32. — P. 538. — <https://doi.org/10.1118/1.4815765>.
9. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation [Electronic resource] / Olaf Ronneberger, Philipp Fischer, Thomas Brox // Lecture Notes in Computer Science. — Cham, 2015. — Pp. 234–241. — https://doi.org/10.1007/978-3-319-24574-4_28.
10. Shelhamer E. Fully Convolutional Networks for Semantic Segmentation [Electronic resource] / Evan Shelhamer, Jonathan Long, Trevor Darrell // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2017. — Vol. 39, no. 4. — Pp. 640–651. — <https://doi.org/10.1109/tpami.2016.2572683>.
11. Sun K. Deep High-Resolution Representation Learning for Human Pose Estimation [Electronic resource] / Ke Sun [et al.] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, 15–20 June 2019. — [S. l.], 2019. — <https://doi.org/10.1109/cvpr.2019.00584>.
12. Zhao H. Pyramid Scene Parsing Network [Electronic resource] / Hengshuang Zhao [et al.] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 21–26 July 2017. — [S. l.], 2017. — <https://doi.org/10.1109/cvpr.2017.660>.

References

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (12), 2481–2495. <https://doi.org/10.1109/tpami.2016.2644615>.
- BraTS2020 Dataset. (n.d.). *The Cancer Imaging Archive (TCIA)*. <https://www.med.upenn.edu/cbica/brats2020/>.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (4), 834–848. <https://doi.org/10.1109/tpami.2017.2699184>.

- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49.
- Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications* (pp. 179–187). Springer. https://doi.org/10.1007/978-3-319-46976-8_19.
- Haralick, R. M., & Shapiro, L. G. (1984). Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 27 (3), 389–405. [https://doi.org/10.1016/0734-189x\(84\)90043-4](https://doi.org/10.1016/0734-189x(84)90043-4).
- Image segmentation techniques: A comprehensive review. (2024). *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/irjmets49604>.
- McGurk, R., Klein, S., Yushkevich, P., & Rohlfing, T. (2013). TH-C-WAB-08: Modeling of the Dice coefficient for PET segmentation studies. *Medical Physics*, 40 (6, part 32), 538. <https://doi.org/10.1118/1.4815765>.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28.
- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 (4), 640–651. <https://doi.org/10.1109/tpami.2016.2572683>.
- Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA, June 15–20, 2019. IEEE. <https://doi.org/10.1109/cvpr.2019.00584>.
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 21–26, 2017. IEEE. <https://doi.org/10.1109/cvpr.2017.660>.

O. Buchko, D. Plakhotna

COMPARISON OF NEURAL NETWORK ARCHITECTURES FOR BRAIN TUMOR SEGMENTATION

Image segmentation plays a crucial role in medical diagnostics, where precise identification of tumor boundaries is essential for treatment planning and prognosis. However, accurate segmentation remains a challenge due to the complexity of anatomical structures and the limited availability of annotated data. Traditional methods are not robust to variability in medical images, which often results in inconsistent and inaccurate outcomes.

This paper investigates the effectiveness of modern neural network-based architectures for brain tumor segmentation using MRI data. The primary goal of the study is to evaluate and compare the accuracy of several convolutional segmentation models under identical training conditions, and to examine whether self-supervised pretraining can improve segmentation quality in cases with limited labeled samples.

The research is based on the BraTS2020 dataset, which contains multi-modal MRI scans with manual annotations of glioma subregions. Five segmentation models were trained: FCN, FPN, PSPNet, DeepLabv3+, and Attention U-Net. All models were evaluated using the Dice Similarity Coefficient. The best result was achieved by Attention U-Net with a mean Dice score of 0.842. A self-supervised learning (SSL) strategy was further applied to pretrain the encoder of this model using a rotation prediction task, which increased the Dice score to 0.869.

The findings confirm that neural network-based methods provide higher segmentation quality compared to classical approaches. More importantly, the integration of SSL enables performance improvements without requiring additional labeled data. This is particularly valuable in the medical field, where collecting expert annotations is expensive and time-consuming.

This article demonstrates that high-quality segmentation of brain tumors is possible even under limited supervision, provided that suitable architectures and training strategies are selected. The presented approach can be adapted for other medical image segmentation tasks and may support the development of practical clinical decision support systems.

Keywords: image segmentation, neural networks, Attention U-Net, self-supervised learning, BraTS2020.

Матеріал надійшов 14.06.2025



Creative Commons Attribution 4.0 International License (CC BY 4.0)