

В.М. СИНЕГЛАЗОВ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: svm@nau.edu.ua

К.Д. РЯЗАНОВСЬКИЙ

Національний технічний університет України «Київський політехнічний інститут
імені Ігоря Сікорського», Київ, Україна, e-mail: kir_ryaz@tk.kpi.ua

О.В. КЛАНОВЕЦЬ

Національний технічний університет України «Київський політехнічний інститут
імені Ігоря Сікорського», Київ, Україна, e-mail: alex.klanovets@gmail.com

ТРИЕТАПНИЙ 2D-3D АНСАМБЛЬ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ СЕГМЕНТАЦІЇ ЗЛОЯКІСНИХ ПУХЛИН ГОЛОВНОГО МОЗКУ НА МРТ-ЗОБРАЖЕННЯХ

Анотація. Розв'язано задачу бінарної семантичної сегментації пухлин головного мозку за МРТ-зображеннями. Виконано попиксельне визначення межі аномальної ділянки за наявності шуму в навчальній вибірці та входних даних. Показано, що у разі використання 2D-моделей для розв'язання задач 3D-сегментації просторову інформацію між сусідніми зрізами не враховують та не використовують. Запропоновано новий підхід до оптимізації оброблення 3D медичних зображень із застосуванням ансамблевих топологій у три етапи. На першому етапі здійснюють 2D-оброблення зображень ансамблем у трьох площинах для максимізації критерію різноманітності і точного захоплення області інтересу (region of interest, ROI). Другий етап передбачає ансамблеве оброблення 3D-областей ROI, виділених нейронними мережами, з різними 3D-розмірами входних блоків для забезпечення різноманітності. На третьому етапі об'єднують виділені аномальні ділянки (злоякісні пухлини) з першого та других етапів шляхом їхнього зваженого підсумовування та виконання операції порогового оброблення (thresholding) для отримання остаточної бінарної 3D-маски пухлини мозку. Проведено тестування запропонованого підходу на датасеті LGG Brain MRI Segmentation Dataset. Покращено точність сегментації за суттєвими метриками dice score та mIoU завдяки скороченню обсягу використовуваних обчислювально-витратних 3D-мереж.

Ключові слова: згорткова нейронна мережа, ансамблева топологія, пухлини головного мозку, МРТ, 3D нейронні мережі.

ВСТУП

Сегментація зображення є однією з найбільш фундаментальних проблем оброблення зображень і має широке застосування в різних галузях, зокрема під час виявлення об'єктів і медичної візуалізації. Пікселі (або вокселі) у кожному сегменті повинні мати певні схожі характеристики, наприклад, інтенсивність однорідності кольору, які свідчать про те, що вони належать одній і тій самій категорії або одному й тому самому об'єкту, що цікавить. Однак, у конкретних галузях застосування сегментація зображень часто має різні інтерпретації або цілі.

Через дедалі більшу доступність сучасних медичних систем сканування зображень, зокрема магнітно-резонансної томографії (МРТ) або комп'ютерної томографії (КТ), виникає потреба в автоматизованому обробленні сканованих даних. Оцінювання результатів, отриманих за допомогою цих методів сканування, лікарі зазвичай виконують вручну. Ця задача може бути повторюваною і трудомісткою навіть для досвідченого радіолога. Тому автоматизація цього процесу є вкрай необхідною і може допомогти лікарям швидше поставити правильний діагноз, коли їм будуть представлені сегментовані відскановані дані протягом декількох секунд.

Високоточна сегментація дає змогу безпомилково виділити аномальну ділянку, розв'язати задачу класифікації, правильно та швидко поставити діагноз

і тим самим сприяти успішному лікуванню. У цьому конкретному контексті мета сегментації полягає в тому, щоб окреслити анатомічні структури або вказати межі органів або інших областей інтересу (ROI). Після сегментації областей інтересу можна отримати геометричні властивості сегментів та іншу інформацію, пов'язану з формою, що полегшує клінічний аналіз.

Автоматична сегментація медичних зображень є дуже складною задачею через високу мінливість і складність медичних зображень, а також через те, що зображення часто спотворюються шумом. Протягом останніх двох десятиліть у цій галузі досягнуто значного прогресу та опубліковано низку досліджень щодо сегментації медичних зображень [1–4].

Однією із цілей медичної сегментації є сегментація людського мозку. Хоча домогтися точного розмежування важко, кіркові та підкіркові парцеляції є дуже важливими для виявлення пухлин, набряків та некротичних тканин. Точне виявлення цих тканин часто є основою надійного діагнозу.

Останнім часом методи глибокого навчання з використанням різних типів глибоких штучних нейронних мереж успішно та широко застосовують у багатьох типових задачах комп'ютерного зору, зокрема для розпізнавання зображень, класифікації зображень та сегментації зображень [5–8]. Розвиток та успіх застосування моделей глибокого навчання у семантичній сегментації природних зображень відкриває великі перспективи для розв'язання багатьох складних задач сегментації у медицині.

1. ПОСТАНОВКА ЗАДАЧІ

Розглянемо задачу у такій постановці. Для вибірки сканів $(X_1, Y_1), \dots, (X_l, Y_l)$, де $X_i \in \mathbb{R}^{n \times n \times m}$ — тензор одного МРТ-скану і $Y_i \in \{0, 1\}^{n \times n \times m}$ — бінарний тензор такого самого розміру, як оригінальний скан (n — ширина та висота зображення, m — кількість зрізів у скані), де елемент 0 означає відсутність, а елемент 1 — наявність пухлини у цьому вокселі скану, створити класифікатор $g(X)$, який коректно визначає бінарний тензор $Y \in \{0, 1\}^{n \times n \times m}$ для нового невідомого скану $X \in \mathbb{R}^{n \times n \times m}$.

Слід зазначити, що математична модель цього МРТ-скану має такий вигляд:

$$X_{ij} = L_{ij} + E_{ij},$$

де L_{ij} — матриця істинного (без шумів) МРТ-зображення j -го скану i -го пацієнта, E_{ij} — матриця шумів (артефактів) j -го скану i -го пацієнта, X_{ij} — матриця реального МРТ-зображення j -го скану i -го пацієнта, який аналізують.

Тому здійснюють попіксельне визначення межі аномальної ділянки за наявності шуму в навчальній вибірці та вхідних даних.

У цій статті запропоновано триетапну структуру ансамблю як такий класифікатор $g(X)$ для оптимізації 3D-оброблення нейронними мережами за допомогою попереднього 2D-оброблення та виділення обмежувальних рамок ROI.

2. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ

У більшості робіт із сегментації даних медичних зображень увагу зосереджено на сегментації аномальних ділянок тканини для визначення правильного діагнозу та особливостей прогресування ракових пухлин.

Останнім часом рішення для глибокого навчання, зокрема глибокі згортові нейронні мережі (CNN), широко використовують для розв'язання проблем розпізнавання зображень у різних сферах, включаючи комп'ютерний зір [9, 10].

Ці рішення забезпечують високий рівень їхнього розв'язання. У більшості робіт використовують двовимірні згорткові мережі, які приймають пакети зображень як вхідні дані, ігноруючи глобальну контекстну інформацію. У роботі [11] запропоновано застосовувати повністю згорткові мережі (FCN) для щільного маркування вхідних зображень через наскрізну глибоку мережеву архітектуру. Модель U-Net [12] покращує FCN, щоб вона працювала з дуже невеликою кількістю навчальних зображень і все ж давала більш точні результати сегментації медичних зображень.

Нещодавно опубліковано кілька методів сегментації, що ґрунтуються на глибокому 3D-навчанні [13–15] для безпосередньої роботи з 3D-медичними зображеннями органів на МРТ- та КТ-сканах. Загалом ці методи можна розділити на дві групи: застосування моделей двовимірної сегментації до 2D-зрізів і накладання/об'єднання результатів 2D-сегментації у певний спосіб; розроблення нової архітектури, яка безпосередньо приймає об'єм тривимірного зображення як вхідні дані. Обидві групи мають свої проблеми. Недоліком використання 2D-моделей для розв'язання проблем 3D-сегментації є те, що просторова інформація між сусідніми зрізами не враховується та не використовується. Отже, повна тривимірна контекстна інформація може бути не повністю відновлена під час наступного етапу оброблення (накладення або об'єднання фрагментів). Пряме застосування 3D-моделі CNN потребувало б включення набагато більшої кількості параметрів і відповідно набагато більшого обсягу даних про навчання, щоб уникнути проблем із перенавчанням. Однак, зазвичай немає можливості отримати багато анотованих 3D-зображень медичного призначення. Інша проблема застосування тривимірних моделей згорткової нейронної мережі до тривимірних медичних даних пов'язана з тим, що розмір зображення зменшується для економії пам'яті. Стиснення зображення для 3D глибокого навчання може погіршити результати виявлення уражень.

Протягом останніх років розроблено різні методи використання ансамблевого аналізу 2D- і 3D-зображень МРТ для сегментації пухлин головного мозку. Наприклад, автори [16] запропонували модель глибокого навчання з використанням 3D U-Net для сегментації пухлин головного мозку. При цьому вони не вибирають одну найкращу модель, а використовують сукупність кількох моделей з різними гіперпараметрами, що зменшує випадкові помилки кожної моделі та покращує продуктивність. Крім того, щоб підкреслити клінічну цінність розробленого методу сегментації, автори побудували лінійну модель, що ґрунтується на радіологічних ознаках, здобутих із сегментації, та інших клінічних ознаках і дає змогу прогнозувати загальне виживання пацієнтів.

Автори [17] запропонували повністю автоматизований ансамблевий метод сегментації пухлин головного мозку на основі глибокого навчання для чотирьох різних тривимірних сканів МРТ. Сегментацію виконують три найефективніші моделі глибокого навчання енкодера-декодера для сегментації, а їхні результати оцінюють за допомогою добре відомих показників сегментації. Далі проведено статистичний аналіз моделей і розроблено модель ансамблю для врахування найвищого коефіцієнта кореляції Метьюза з використанням певної модальності МРТ.

У статті [18] запропоновано нову структуру ансамблю для сегментації 3D біомедичних зображень, яка поєднує в собі переваги 2D- і 3D-моделей. Запропонований підхід складається з двох основних компонентів: 1) групи базових 2D-моделей, які досліджують навчальні 3D-дані з різних геометричних точок зору (x_y , x_z , y_z — довжина—ширина, довжина—висота, ширина—висота відповідно), та окремої 3D-мережі; 2) системи ансамблевого навчання, яка вико-

ристовує метамодель глибокого навчання для об'єднання результатів базових моделей. Масштабні експерименти на двох відкритих наборах даних показали, що їхній підхід може досягти кращої продуктивності, ніж поточні методи.

У роботі [19] запропоновано архітектуру Multi-Branch Ensemble Learning на основі 3D-CNN. Метод поєднує в собі три ключові ідеї: побудова 3D-CNN для максимально точного відображення просторової інформації легеневих вузликів у 3D-просторі; вбудовування архітектури мережі з паралельними ланками у 3D-CNN, яка добре адаптована до гетерогенності конкрецій; використання ансамблів для ефективного покращення продуктивності узагальнення моделі 3D-CNN. Експерименти на наборі даних LUNA16 показали кращі результати скринінгу. Недоліком методу в контексті цієї роботи є розв'язання задачі бінарної класифікації наявності патологій на частині зрізу, а не бінарної сегментації патології.

Залежно від роздільної здатності медичних зображень доступне обладнання графічного процесора відіграє важливу роль у 3D-аналізі сегментації. Наразі неможливо повністю вмістити розмір вибірки MPT високої роздільної здатності $400 \times 512 \times 512$ у поточні моделі згорткової нейронної мережі через потребу у величезному обсязі пам'яті графічного процесора.

Отже, тривимірні дані медичних зображень можна або розрізати на менші прямокутні секції, або аналізувати шар за шаром подібно до набору двовимірних зображень, щоб повністю використовувати інформацію за всіма трьома осями.

У цій статті розв'язано проблему оброблення MPT-зображень високої роздільної здатності шляхом попереднього вибору досліджуваних ділянок швидкими 2D-мережами і оброблення лише цих вибраних фрагментів даних за допомогою 3D-мереж. Цей підхід дає змогу значно скоротити обсяг обчислень, що є принциповим під час оброблення зображень.

3. ОПИС ДАНИХ

Оброблення форматів даних медичних 3D-зображень передбачає перетворення наданих анованих вхідних даних із форматів медичних даних на простий набір зображень. Це може дезорієнтувати тих, хто не має детальної інформації про різні медичні формати. Медичне зображення — це внутрішня структура анатомічної ділянки у вигляді масиву елементів, які називають пікселями у 2D або вокселями у 3D.

Формати даних медичних зображень містять загальну інформацію як-от: глибина пікселів, метадані, піксельні дані та фотометрична інтерпретація. Ці формати також зберігають кілька 3D-сканів пацієнта у часі, тому їх можна використовувати для оброблення 4D-зображень. Але усі дані, використані в цій статті, складаються лише з одного 3D-скану на пацієнта, тобто без часової складової.

Найпопулярнішим форматом для медичних зображень є Dicom (Digital Imaging and Communications in Medicine). Це поширений формат даних зображень, який найчастіше застосовується для MPT, КТ тощо і має суфікс .dcm.

У цьому дослідженні використано набір даних LGG Brain MRI Segmentation Dataset [20], попередньо оброблений і збережений у форматі TIFF. Набір даних складається з 112 3D-сканів; кожний скан містить приблизно 22 зрізи мозку. Для кожного зрізу є також бінарні маски з сегментованими ділянками патології. Приклади зображень наведено на рис. 1.

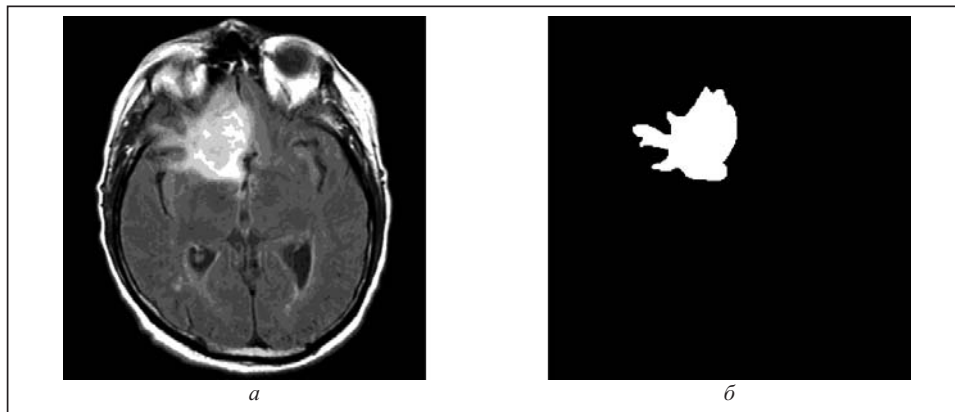


Рис. 1. Приклад зображення з набору даних (а) і маски (б)

4. МЕТОДИ ТА ПІДХОДИ

4.1. Опис запропонованого підходу. Підхід ґрунтується на модифікації, комбінуванні, архітектурному вдосконаленні підходів з [18] і [19] для розв'язання інших класів задач (сегментації), а також скороченні обсягу обчислень та часу навчання моделей.

У [18] 2D- і 3D-мережі нейронного оброблення об'єднано для аналізу 3D-сканів медичних зображень. Недоліки цього комплектування є такими.

— Повне 3D-оброблення всього скану, яке потребує великого обсягу обчислень та тривалого часу (удвічі більшого, ніж для 2D-оброблення [21–23]).

— Для економії пам'яті розмір зображення зменшено. Стиснення зображення для 3D глибокого навчання може погіршити результати виявлення уражень.

— Є тільки один розмір 3D-блоку, який може не охоплювати всіх патологій.

У [19] розглянуто ансамблевий підхід з кількома гілками оброблення 3D нейронних мереж вихідного 3D-скану з різними масштабами блоків ($18 \times 18 \times 18$, $30 \times 30 \times 30$, $40 \times 40 \times 40$) (рис. 2). Хоча ідея підходу є ефективною, він має такі недоліки:

— повне оброблення всього 3D-скану з усіма блоками, що потребує ще більшого обсягу обчислювальних ресурсів і пам'яті;

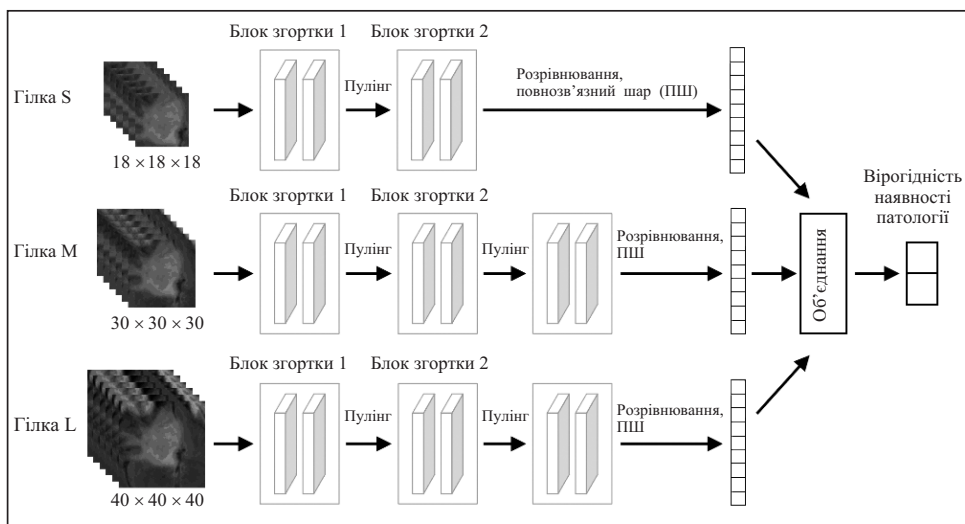


Рис. 2. Ансамблеве оброблення 3D-скану з різними масштабами вхідних блоків

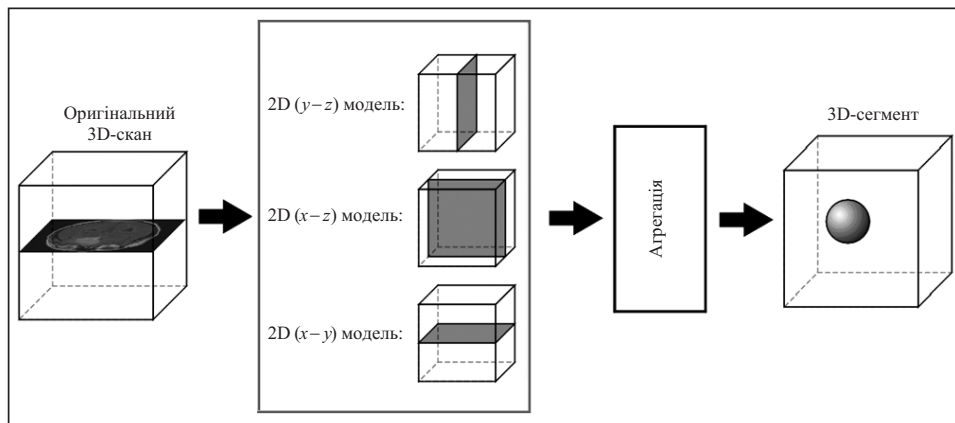


Рис. 3. Ансамблеве 2D-оброблення в різних ракурсах

— розв’язання задачі бінарної класифікації наявності патології у 3D-блоці, а не задачі сегментації патології;

— розмір мережі є досить великим: 42–44 млн параметрів для навчання.

Основна проблема полягає в тому, що оброблення всього 3D-скану з 3D-мережевими блоками може зайняти багато часу та пам’яті, хоча це важливо для просторового прояснення патологій на зображенні. Крім того, стиснення зображення для глибокого 3D-навчання може погіршити результати виявлення уражень.

Щоб поліпшити цей процес за критеріями розмірності моделей, часом навчання, характеристиками пам’яті та точності, запропоновано здійснювати триетапне ансамблеве оброблення 3D-медичних зображень. Воно містить такі етапи.

Етап 1. Двовимірне оброблення сканів для попереднього пошуку малих ROI. Використання ансамблю з трьох 2D-мереж для аналізу 3D-сканів у трьох ракурсах (з дотриманням підходу [18], але без 3D-мережі та метамоделі (рис. 3)). На виході отримують 3D бінарні маски патологій і ROI у вигляді обмежувальних рамок.

Етап 2. Застосування ансамблю нейронної мережі з вхідними 3D-блоками різного масштабу для сегментації вибраних малих 3D ROI з етапу 1 (з дотриманням підходу [19]). Розмір вхідних 3D-блоків вибирають автоматично залежно від розміру патологій на оброблюваному датасеті. Суть цього 3D-оброблення полягає в тому, щоб просторово уточнити аномалії, отримані в результаті 2D-оброблення.

У цьому випадку обробляють не повний скан (розміром, наприклад, $512 \times 512 \times 20$), а лише невеликі блоки ROI для їхнього уточнення (розміром, наприклад, $100 \times 100 \times 5$, що приблизно в п’ять разів менше оригінального 3D-скану). Це дає змогу не стискати дані для оброблення (збереження точності) і прискорити навчання.

Етап 3. Агрегування вкладених у 3D-маску 2D-масок з етапу 1 та 3D-масок з етапу 2 для отримання кінцевого результату для максимізації різноманітності оброблення даних.

4.2. Попереднє оброблення даних. Нормалізація інтенсивності пікселів.

Несумісність діапазонів інтенсивностей сигналу зображення може істотно вплинути на ефективність методів сегментації. Діапазони сигналів біомедичних зображень значно відрізняються між наборами даних через різні формати зображень, різне обладнання/інструменти (наприклад, різні сканери), технічні невідповідності та просто біологічні варіації. Крім того, алгоритми машинного

навчання, що лежать в основі сегментації зображення, зазвичай працюють краще з тими ознаками, які мають нормальний розподіл.

Щоб досягти узгодженості динамічного діапазону інтенсивності сигналу, рекомендовано масштабувати та стандартизувати дані зображення. Масштабування інтенсивності сигналу проєктує початковий діапазон значень у попередньо визначений діапазон, зазвичай між $[0, 1]$ або $[-1, 1]$, тоді як стандартизація центрує значення близько до нормального розподілу шляхом розрахунку Z -оцінки нормалізації. У цьому дослідженні було обрано масштабування вхідних 8-бітних зображень TIFF у діапазоні $[0, 1]$.

4.3. Детальний опис запропонованих етапів оброблення. Розглянемо кожен етап детальніше.

Етап 1. Використання ансамблю з трьох 2D-мереж для аналізу 3D-скану у трьох ракурсах (з дотриманням підходу [18], рис. 3, але без 3D-мережі та метамоделі) і генерування бінарних масок із ROI як обмежувальних рамок. Як такі 2D-мережі використано модифікацію моделі Fully Convolutional Network (FCN) [24], що забезпечила значні успіхи у сегментації біомедичних зображень та була інтегрована в останні досягнення в галузі дизайну мереж глибокого навчання, як-от батч-нормалізація, залишкові мережі та бутлнек-дизайн. Додатковою перевагою моделі є зменшена кількість параметрів, необхідних для навчання, яка становить 3.9 млн.

Кількість ракурсів (а отже і 2D-мереж) дорівнює трьом, бо здійснюється оброблення 3D-скану у трьох його площинах (xy , yz , xz).

У такий спосіб урізноманітнено оброблення даних для ефективного 2D-ансамблю. 2D-мережі будуть швидшими за 3D-мережі у разі повного сканування мозку. Це дає змогу оптимізувати виявлення початкової ділянки патології [21–23].

1.1. З кожної 2D-мережі отримуємо набір 2D-сегментованих масок

$$Y \in \{y \in \mathbb{R} \mid 0 \leq y \leq 1\}^{n \times n},$$

де n — ширина і висота зображення.

1.2. Далі вкладаємо всі 2D-матриці з однієї мережі в одну 3D маску-модель патології

$$Y \in \{y \in \mathbb{R} \mid 0 \leq y \leq 1\}^{n \times n \times m},$$

де n — ширина та висота зображення, m — кількість 2D-зрізів в одному повному 3D-скані мозку.

1.3. Оскільки кожна мережа працює у власному ракурсі, ці 3D-маски можуть відрізнятися, тому додатково об'єднуємо їх за допомогою операції максимуму ($\max$) в одну 3D-маску.

Операцію $\max$ використовують для захоплення всіх можливих виділених ділянок патології. Значення кожного пікселя агрегованої 3D-маски визначають так:

$$y_{ijk} = \max_{l \in N} y_{ijk}^{(l)},$$

де $y_{ijk}^{(l)}$ — i, j, k компонента вихідного 3D-тензора (3D-маски) l -го сегментатора, N — кількість сегментаторів.

За допомогою цього підходу можна захопити додаткові пікселі, які не належать досліджуваній ділянці. Тому етап 2 із 3D-обробленням вибраних досліджуваних ділянок потрібний для уточнення патологій.

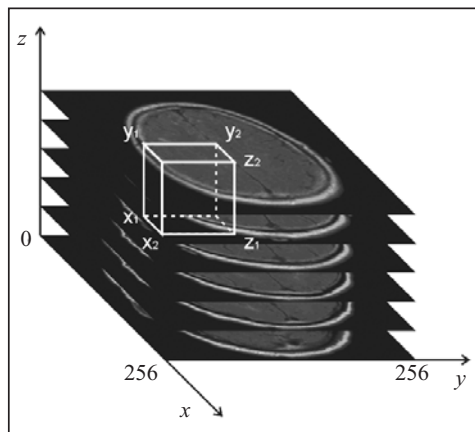


Рис. 4. Тривимірний приклад обмежувальних рамок патології

1.4. Після отримання агрегованої матриці для її бінаризації застосовують поріг зі значенням 0.5 (тобто ймовірність того, що піксель містить пухлину, перевищує 50 %). Отже, отримано результувну матрицю $Y \in \{0, 1\}^{n \times n \times m}$, де елемент 0 означає відсутність патології в цьому пікселі зображення, а елемент 1 — її наявність.

1.5. Далі будують мінімальні 3D-обмежувальні рамки (тензори), які містять 3D-маски патологій, отримані після оброблення агрегації 2D-мереж (п. 1.4) (рис. 4).

1.6. Отримані обмежувальні прямокутники застосовують до початкового 3D-скану мозку, щоб вирізати лише ті досліджувані ділянки з виявленими патологіями. Отже, із усього великого 3D-скану ($n \times n \times t$, де n — ширина та висота скану, t — кількість зрізів у скані) будуть вирізані маленькі 3D-об'єкти з можливими патологіями для аналізу ($k \times p \times l$, де k — ширина виділеної області, p — висота виділеної області, l — кількість фрагментів із площею, $k \ll n$, $p \ll n$, $t \ll l$). Інакше кажучи, простір для 3D-мереж значно звужено.

Загальну схему етапу 1 наведено на рис. 5.

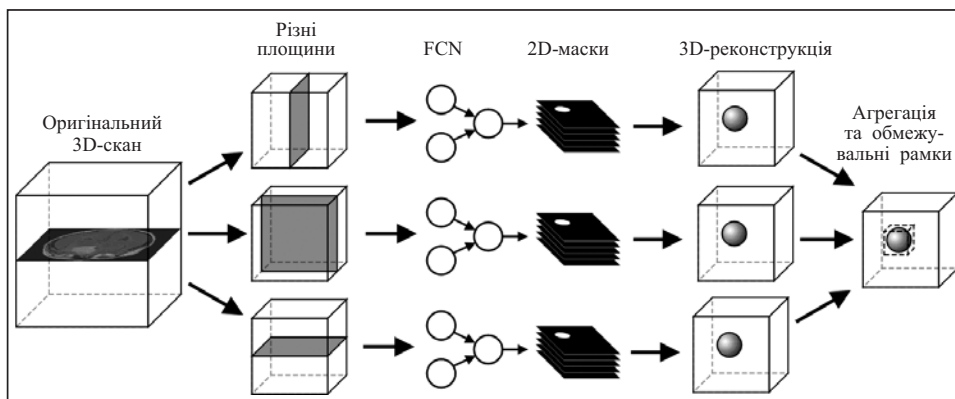


Рис. 5. Загальна схема етапу 1

Етап 2. Застосування ансамблю нейронної мережі з 3D-блоками різного масштабу для аналізу вибраних 3D ROI з етапу 1 (з дотриманням загального багатоланкового підходу [19]). Сенс цього 3D-оброблення полягає в тому, щоб просторово уточнити патології, отримані в результаті 2D-оброблення. Оскільки у підході [19] розмір мереж є досить великим (42–44 млн параметрів) та розв'язано задачу класифікації, у цій статті запропоновано модифікацію цієї архітектури під задачу сегментації з використанням сучасних Dense-зв'язків. При цьому зменшено розміри мереж та час навчання для більш ефективної роботи.

Автори архітектури Multi-Branch Ensemble Learning [19] використовують три паралельні гілки 3D-блоків з вхідними розмірами $18 \times 18 \times 18$, $30 \times 30 \times 30$ та

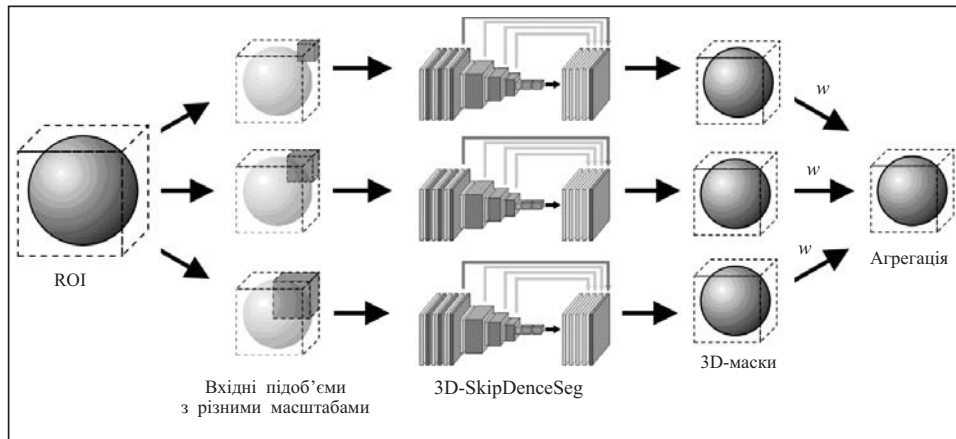


Рис. 6. Загальна схема етапу 2

$40 \times 40 \times 40$, які обробляють різними згортковими блоками: IResBlock, ConvBlock та DenseBlock. Кількість параметрів при цьому є надлишковою, а блоки орієнтовані на задачу класифікації. Нами запропоновано скористатися сучасною 3D-архітектурою 3D-SkipDenseSeg [25], яка ефективно розв'язує задачу сегментації, маючи при цьому невеликий розмір (1.55 млн параметрів) завдяки наявності Dense-зв'язків. Це дасть змогу модифікувати кожну гілку ансамблевої топології архітектурою 3D-SkipDenseSeg. Тоді кожна гілка буде розв'язувати задачу сегментації, а агрегація їхніх виходів забезпечить єдину 3D-маску патології (рис. 6).

Розглянемо детально кожен крок етапу 2.

Комбінування різних масштабів тривимірних блоків сприяє підвищенню різноманітності ансамблю та кінцевої точності [19]. Розмір вибраних ROI з першого етапу набагато менший за розмір повного скану (більше ніж у п'ять разів) і відповідає загальному розміру патології, яка сегментується. Тому розмір вхідних 3D-блоків вибирається автоматично залежно від розміру аномалії, яка знає сегментації. В контексті цієї задачі розмір 3D-блоків зменшено з $18 \times 18 \times 18$, $30 \times 30 \times 30$, $40 \times 40 \times 40$ до $10 \times 10 \times 10$, $20 \times 20 \times 20$ та $30 \times 30 \times 30$ відповідно.

На цьому кроці отримують набір тривимірних бінарних масок від кожної гілки. Застосовано зважене середнє для об'єднання отриманих масок ROI після оброблення 3D-блоків різних масштабів. Вага гілок блоку визначається пропорційно до точності. Як метрику точності використано попіксельний dice score:

$$DS = \frac{2TP}{2TP + FN + FP + \varepsilon},$$

де TP — кількість істинно позитивних результатів, FN — кількість хибно-негативних результатів, FP — кількість хибно-позитивних результатів, ε — дуже мала константа, щоб уникнути ділення на нуль.

Отже, ваги гілок визначають за такою формулою:

$$w^{(i)} = \frac{DS_i}{\sum_{j=1}^n DS_j}, \quad (1)$$

де $w^{(i)}$ — вага i -ї гілки, DS_j — dice score (точність на тестовій вибірці) j -ї гілки, n — кількість паралельних гілок.

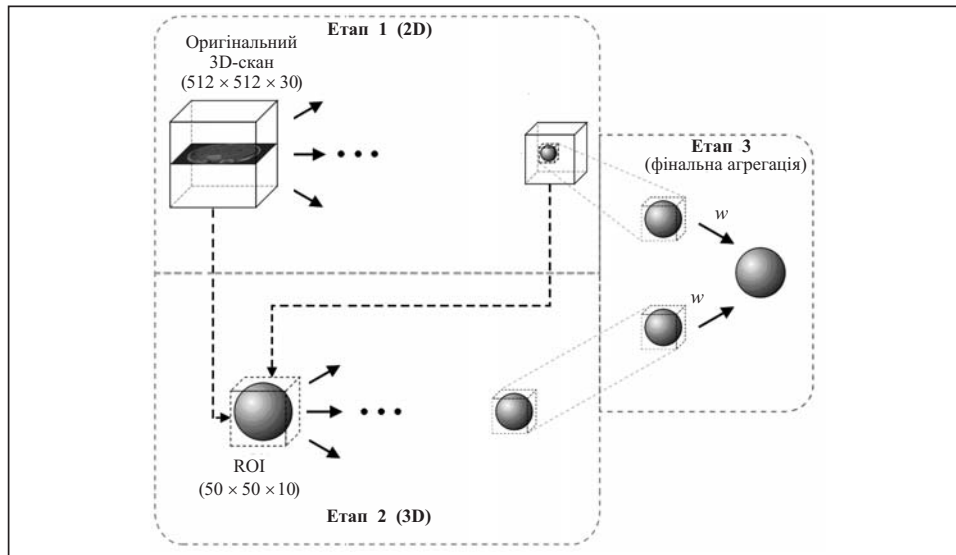


Рис. 7. Схема запропонованого триетапного підходу для оброблення 3D-сканів

Значення кожного пікселя агрегованої 3D-маски визначають за формулою

$$y_{ijk} = \sum_{l=1}^N w^{(l)} y_{ijk}^{(l)}, \quad (2)$$

$$\sum_{l=1}^N w^{(l)} = 1,$$

де i, j, k — координати елемента тензора, N — кількість гілок з різними масштабами 3D-блоків, $y_{ijk}^{(l)}$ — значення i, j, k елемента l -ї гілки 3D вихідного тензора, $w^{(l)}$ — вага l -ї гілки.

Етап 3. Поєднання 3D-масок з етапу 1 і 3D-масок з етапу 2. Отже, об'єднуємо 2D- і 3D-оброблення, але значно зменшуємо кількість 3D-обчислень, зберігаючи баланс різноманітності та точності. Результати об'єднано за допомогою формули (2) (середньозважено), ваги визначено за формулою (1) (відповідно до точності (dice score) 2D-мереж з етапу 1 і 3D-мереж з етапу 2).

Загальну схему підходу наведено на рис. 7.

5. РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ

5.1. Підготовка датасету та метрик. Запропоновані підходи перевірено на LGG Segmentation Dataset [20]. Для навчання мереж сегментаторів використано 112 сканів із набору даних. Їх нормалізовано до діапазону $[0, 1]$. Виконано 5-fold кросс-валідацію та розраховано довірчі інтервали з рівнем 95 % для більш статистично коректного відображення результатів точності моделей. Тестові набори містять скани лише нових пацієнтів, яких не було в навчальному наборі. Це забезпечує максимальний ступінь незалежності тестування. Окремі зрізи на кожному скані використано для навчання двовимірних мереж; повний 3D-скан використано для навчання 3D-мереж. Як функцію втрат застосовано попіксельний dice loss:

$$DL = 1 - \frac{2TP}{2TP + FN + FP + \epsilon},$$

де TP — кількість істинно позитивних результатів, FN — кількість хибно-негативних результатів, FP — кількість хибно-позитивних результатів, ε — дуже мала константа, щоб уникнути ділення на нуль.

Як оптимізатор використано алгоритм Адам зі швидкістю навчання $1e-4$.

Для перевірки точності моделей бінарної сегментації застосовано такі піпиксельні показники: точність (accuracy); mean intersection over union (mIoU); dice score.

5.2. Результати навчання. Результати запропонованих двовимірних моделей бінарного сегментатора FCN у різних ракурсах на тестовій вибірці представлено в табл. 1.

Результати тривимірних моделей бінарного сегментатора 3D-SkipDenseSeg з різними масштабами на тестовій вибірці представлено в табл. 2.

Підсумки агрегації результатів 3D і 2D (етап 3) на тестовій вибірці є такими: точність 99.12 ± 0.81 %, mIoU 0.968 ± 0.013 , dice score 0.981 ± 0.011 .

Результати порівняння запропонованого методу з іншими методами представлено в табл. 3.

5.3. Порівняння розмірів і продуктивності моделей. Запропоновані у підрозд. 4.3 модифікації архітектур та підходів мали на меті пришвидшення навчання

Таблиця 1. Результати застосування 2D-мереж на тестовій вибірці

2D FCN	Точність (%)	mIoU	Dice score
Ракурс 1 ($x-y$)	94.96 ± 1.71	0.912 ± 0.023	0.921 ± 0.024
Ракурс 2 ($y-z$)	95.15 ± 1.47	0.904 ± 0.014	0.932 ± 0.021
Ракурс 3 ($x-z$)	95.19 ± 2.11	0.918 ± 0.021	0.938 ± 0.017
Агрегація max	95.43 ± 1.53	0.922 ± 0.015	0.947 ± 0.016

Таблиця 2. Результати 3D мереж на тестовій вибірці

3D-SkipDenseSeg	Точність (%)	mIoU	Dice score
Масштаб 10×10	98.12 ± 1.22	0.932 ± 0.015	0.951 ± 0.012
Масштаб 20×20	96.51 ± 1.48	0.921 ± 0.024	0.956 ± 0.021
Масштаб 30×30	97.15 ± 2.12	0.936 ± 0.019	0.946 ± 0.017
Зважена агрегація	98.14 ± 1.42	0.946 ± 0.012	0.964 ± 0.018

Таблиця 3. Результати порівняння запропонованого методу з наявними методами

Метод	Точність (%)	mIoU	Dice score
CNN та GA (Anaraki et al. [26] (2019))	90.90	—	—
NS-CNN (Ozyurt et al. [27] (2019))	95.62	—	—
U-Net: ResNeXt50 (Ghosh et al. [28] (2021) (SOTA))	99.60	—	0.932
FC CRF (Li et al. [29] (2017))	—	—	0.8504
3-stage 2D-3D ensemble (Ця стаття)	99.12 ± 0.81	0.968 ± 0.013	0.981 ± 0.011

та роботи мереж-сегментаторів за рахунок зменшення кількості параметрів для навчання та поліпшення точності роботи. Проаналізуємо розміри архітектур запропонованого методу порівняно з іншими. Розмір запропонованого 2D-3D ансамблю становить 16 млн параметрів, що є значно меншим, ніж у підході з трьома 3D-гілками [19] (42–44 млн параметрів), який модифіковано під задачу сегментації. Порівняно з другим підходом, який також було взято за основу, з 2D- та 3D-обробленням та метамоделлю [18] (16 млн параметрів), розроблений підхід має дещо більше параметрів, але вони нівелюються значно збільшеним різноманіттям 3D-оброблення та зменшеним розміром вхідного 3D-блоку (не все зображення, а лише невеликий блок). Порівняно зі state-of-the-art (SOTA) методом [28] (32 млн параметрів) розроблений ансамблевий метод має у два рази менше параметрів.

5.4. Аналіз результатів. Як видно, архітектура ансамблю з максимальною агрегацією 2D-мереж забезпечила підвищення метрик dice score та mIoU (див. табл. 1), також додавши більше релевантних пікселів. Різноманітність мереж у трьох вимірах спрацювала позитивно.

Наступний етап 3D-сегментації з різними масштабами вхідних блоків дав кращу точність, ніж 2D-мережі, але швидкість роботи та навчання зайняли значно більше часу. Зважене ансамблювання дало змогу уточнити результати порівняно з двовимірною сегментацією на 2.7 % точності, 0.024 mIoU та 0.017 dice score (табл. 2). Отже, просторове уточнення результатів працює.

Остаточна точність усього підходу (етап 3) після об'єднання результатів гілки 2D-оброблення та 3D-оброблення є такою: точність 99.12 ± 0.81 %, dice score 0.981 ± 0.011 , mIoU 0.968 ± 0.013 .

Порівняно з іншими методами [26–29] на цьому наборі даних із цієї задачею (бінарна сегментація) (табл. 3) запропонований ансамблевий метод показав значну перевагу. На жаль, не всі автори зазначених робіт надали метрики, які використано авторами цієї роботи, але все ж вдалося провести порівняння, застосовуючи принаймні одну метрику.

Результати показали достовірність і перевагу застосування запропонованого методу для бінарної сегментації пухлин головного мозку за точністю та швидкістю навчання. Порівняно з SOTA методом [28] досягнуто приросту майже на 0.05 dice score (або 5 %), що є досить гарним показником. Аналіз розмірності сучасних методів та запропонованого ансамблевого підходу показав значне зменшення кількості параметрів та більш ефективну роботу.

Порівняння розробленого підходу з методами, взятими за основу [18, 19], свідчить про таке.

— Видалення повного 3D-оброблення з методу [18] (перший крок розробленого підходу) дало змогу зменшити кількість параметрів для навчання більше ніж на 5 млн у разі використання 3D UNet3+ або на 40 млн у разі використання 3D VNet. Також немає потреби навчати великі мережі, що скоротило час навчання на кілька днів.

— Зменшення розміру 3D-блоків з роботи [19] і розміру вхідних даних (більше ніж у п'ять разів) скоротило час навчання мереж також більше ніж у п'ять разів.

— Точність результатів розробленого методу статистично вища на 0.05 dice score порівняно із сучасним методом [28].

— Кількість параметрів для навчання стала меншою (порівняно з [19]) або майже не змінилась (порівняно з [18]). Забезпечено додаткові переваги у кількості обчислень та часі навчання).

Новизна цієї роботи полягає у побудові послідовного конвеєра для оброблення зображень на основі модифікації архітектур робіт [18, 19] та власних архітектурних доробків, де кожен крок залежить від попереднього, але зрештою всі вони впливають на результат. Новими рішеннями для оптимізації 3D-оброблення є етап 1, який попередньо вибирає невеликі 3D-ділянки з можливими патологіями, використовуючи швидші 2D-мережі, і етап 2 з точним 3D-просторовим обробленням цих невеликих областей для уточнення патологій.

ВИСНОВКИ

Запропоновано послідовний пайплайн ансамблевого оброблення 3D-сканів мозку для бінарної сегментації пухлин. Новизна полягає у використанні ансамблю з трьох 2D-мереж для аналізу 3D-сканів у трьох ракурсах, результатом якого є 3D бінарні маски патологій і ROI як обмежувальні рамки; застосуванні ансамблю нейронних мереж з 3D входними блоками різних масштабів для аналізу вибраних малих 3D ROI з етапу 1, що забезпечує просторове уточнення патологій, отриманих в результаті 2D-оброблення; поєднанні 3D-маски з аналізу 2D-даних з етапу 1 і 3D-маски етапу 2 для отримання кінцевого результату.

Ця конструкція ансамблю дає змогу оптимізувати складне, але важливе 3D-оброблення нейронними мережами шляхом попереднього 2D-вилучення невеликих областей ROI.

Запропонований підхід перевірено на LGG Segmentation Dataset бінарної сегментації пухлини мозку. Як критерії для оцінки ефективності цього підходу використано такі показники: точність, mIoU, dice score.

Порівняння з іншими сучасними підходами [18, 19, 26–29] показало значну перевагу запропонованого підходу щодо точності сегментації та розмірності моделей (кількості параметрів).

Напрямок майбутніх досліджень є аналіз різноманітних варіантів об'єднання моделей в ансамбль, подальша оптимізація 3D-оброблення для кращого захоплення просторових особливостей та поглиблений аналіз/складання нейронних мереж та їхніх модулів на прикладі моделі UNet.

СПИСОК ЛІТЕРАТУРИ

1. Malhotra P., Gupta S., Koundal D., Zaguia A., Enbeyle W. Deep neural networks for medical image segmentation. *Journal of Healthcare Engineering*. 2022. Vol. 2022. Article ID 9580991. 15 p. <https://doi.org/10.1155/2022/9580991>.
2. Wang R., Lei T., Cui R., Zhang B., Meng H., Nandi A. Medical image segmentation using deep learning: A survey. *IET Image Processing*. 2022. Vol. 16, Iss. 5. P. 1243–1267. <https://doi.org/10.1049/ipr2.12419>.
3. Rizwan-i-Haque I., Neubert J. Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*. 2020. Vol. 18. Article 100297. <https://doi.org/10.1016/j.imu.2020.100297>.
4. Müller D., Kramer F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Medical Imaging*. 2019. Vol. 21. Article 12. <https://doi.org/10.1186/s12880-020-00543-7>.
5. Xin M., Wang Y. Research on image classification model based on deep convolution neural network. *EURASIP Journal on Image and Video Processing*. 2019. Vol. 2019. Article 40. <https://doi.org/10.1186/s13640-019-0417-8>.
6. Krishna M., Neelima M., Mane H., Matcha V. Image classification using deep learning. *International Journal of Engineering & Technology*. 2018. Vol. 7, Iss. 7. P. 614–617. <https://doi.org/10.14419/ijet.v7i2.7.10892>.

7. Minaee S., Boykov Yu., Porikli F., Plaza A., Kehtarnavaz A., Terzopoulos D. Image segmentation using deep learning: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022. Vol. 44, Iss.7. P. 3523–3542. <https://doi.org/10.1109/TPAMI.2021.3059968>.
8. Islam M.T., Karim Siddique B.M.N., Rahman S., Jabid T. Image recognition with deep learning. *Proc. 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)* (21–24 October 2018, Bangkok, Thailand). Bangkok, 2018. P. 106–110. <https://doi.org/10.1109/ICIIBMS.2018.8550021>.
9. Chai J., Zeng H., Li A., Ngai E. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*. 2021. Vol. 6. Article 100134. <https://doi.org/10.1016/j.mlwa.2021.100134>.
10. Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E., Andina D. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*. 2018. Vol. 2018. Article ID 7068349. <https://doi.org/10.1155/2018/7068349>.
11. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. *Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (7–12 June 2015, Boston, MA, USA). Boston, 2015. P. 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>.
12. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proc. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. MICCAI 2015* (5–9 October 2015, Munich, Germany). Munich, 2015. Navab N., Hornegger J., Wells W., Frangi A. (Eds.). *Lecture Notes in Computer Science*. 2015. Vol. 9351. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
13. Milletari F., Navab N., Ahmadi S. V-Net: fully convolutional neural networks for volumetric medical image segmentation. *Proc. 2016 Fourth International Conference on 3D Vision (3DV)* (25–28 October 2016, Stanford, CA, USA). Stanford, 2016. P. 565–571.
14. Seol Y.J., Kim Y.J., Kim Y.S., Cheon Y.W., Kim K.G. A study on 3D deep learning-based automatic diagnosis of nasal fractures. *Sensors*. 2022. Vol. 22, Iss. 2. Article 506. <https://doi.org/10.3390/s22020506>.
15. Charles R.Q., Su H., Kaichun M., Guibas L.J. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (21–26 July 2017, Honolulu, HI, USA), 2017. P. 77–85. <https://doi.org/10.1109/CVPR.2017.16>.
16. Feng X., Tustison N.J., Patel S.H., Meyer C.H. Brain tumor segmentation using an ensemble of 3D U-Nets and overall survival prediction using radiomic features. *Frontiers in Computational Neuro Science*. 2020. Vol. 14. Article 25. <https://doi.org/10.3389/fncom.2020.00025>.
17. Das S., Bose S., Nayak G.K., Saxena S. Deep learning-based ensemble model for brain tumor segmentation using multi-parametric MR scans. *Open Computer Science*. 2022. Vol. 12, N 1. P. 211–226. <https://doi.org/10.1515/comp-2022-0242>.
18. Zheng H., Zhang Y., Yang L., Liang P., Zhao Z., Wang C., Chen D.Z. A new ensemble learning framework for 3D biomedical image segmentation. *Proc. Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'19/IAAI'19/EAAI'19)* (27 January – 1 February 2019, Honolulu, Hawaii USA). Honolulu, 2019. Article 725. P. 5909–5916. <https://doi.org/10.1609/aaai.v33i01.33015909>.
19. Cao H., Liu H., Song E., Ma G., Xu X., Jin R., Liu T., Hung Ch.-Ch. Multi-branch ensemble learning architecture based on 3D CNN for false positive reduction in lung nodule detection. *IEEE Access*. 2019. Vol. 7. P. 67380–67391. <https://doi.org/10.1109/ACCESS.2019.2906116>.
20. Brain MRI segmentation. LGG Segmentation Dataset. URL: <https://www.kaggle.com/datasets/mateuszbeda/lgg-mri-segmentation>.
21. Zhou X., Yamada K., Takayama R., Zhou X., Hara T., Fujita H., Wang S., Kojima T. Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images. *Proc. SPIE 10575, Medical Imaging 2018: Computer-Aided Diagnosis* (10–15 February 2018, Houston, Texas, USA). Houston, 2018. Vol. 10575. P. 105752C. <https://doi.org/10.1117/12.2295178>.

22. Srikrishna M., Heckemann R.A., Pereira J.B., Volpe G., Zettergren A., Kern S., Westman E., Skoog I., Schöll M. Comparison of two-dimensional- and three-dimensional-based U-Net architectures for brain tissue classification in one-dimensional brain CT. *Front. Comput. Neurosci.* 2022. Vol. 15. Article 785244. <https://doi.org/10.3389/fncom.2021.785244>.
23. Stamoulakatos A., Cardona J., Michie C., Andonovic I., Lazaridis P., Bellekens X., Atkinson R., Hossain Md.M., Tachtatzis C. A comparison of the performance of 2D and 3D convolutional neural networks for subsea survey video classification. *Proc. OCEANS 2021 San Diego — Porto* (20–23 September 2021, San Diego, Portugal). San Diego, 2021. P. 1–10. <https://doi.org/10.23919/OCEANS44145.2021.9706125>.
24. Yang L., Zhang Y., Chen J., Zhang S., Chen D.Z. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention — MICCAI 2017. MICCAI 2017* (11–13 September 2017, Quebec City, QC, Canada). Quebec City, 2017. *Lecture Notes in Computer Science*. Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (Eds.). 2017. Vol. 10435. P. 399–407. https://doi.org/10.1007/978-3-319-66179-7_46.
25. Bui T. D., Shin J., Moon T. Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. *Biomed. Signal Process. Control.* 2019. Vol. 54, Article 101613. <https://doi.org/10.1016/j.bspc.2019.101613>.
26. Anaraki A.K., Ayati M., Kazemi F. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering.* 2019. Vol. 39, N. 1. P. 63–74.
27. Ozyurt F., Sert E., Avci E., Dogantekin E. Brain tumor detection based on a convolutional neural network with neutrosophic expert maximum fuzzy sure entropy. *Measurement.* 2019. Vol. 147, Article 106830.
28. Ghosh S., Santosh Kc. Tumor segmentation in brain MRI: U-Nets versus feature pyramid network. *Proc. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (1–9 June 2021, Online event). P. 31–36. <https://doi.org/10.1109/CBMS52027.2021.00013>.
29. Li Z., Wang Y., Yu J., Shi Z., Guo Y., Chen L., Mao Y. Low-grade glioma segmentation based on CNN with fully connected CRF. *J. Healthc. Eng.* 2017. Vol. 2017. Article Gs9283480. <https://doi.org/10.1155/2017/9283480>.

V. Sineglazov, K. Riazanovskiy, O. Klanovets

A THREE-STAGE 2D-3D CONVOLUTIONAL NETWORK ENSEMBLE FOR SEGMENTATION OF MALIGNANT BRAIN TUMORS ON MRI IMAGES

Abstract. In this paper, the problem of brain tumor binary semantic segmentation from MRI images is solved. The pixel-by-pixel determination of the anomaly region boundary is performed given the presence of noise in the training sample and input data. It is shown that in the case of using 2D models for solving 3D segmentation problems, spatial information between neighboring slices is not considered and not utilized. A new approach for optimizing the processing of 3D medical images using ensemble topologies in three stages is proposed. The first stage involves 2D ensemble processing of images in three dimensions to maximize the diversity criterion and accurately capture the region of interest (ROI). The second stage involves ensemble processing of 3D ROI regions extracted by neural networks with different 3D input block sizes to ensure diversity. In the third stage, the extracted abnormal regions (malignant tumors) from the first and second stages are aggregated by weighted summation and thresholding to obtain the final binary 3D mask of the brain tumor. The proposed approach was tested on the LGG Brain MRI Segmentation Dataset. It is shown that the segmentation accuracy is significantly improved in terms of dice score and mIoU, reducing the use of computationally expensive 3D networks.

Keywords: convolutional neural network, ensemble topology, brain tumor, MRI, 3D neural network.

Надійшла до редакції 26.10.2022