

СУЧАСНІ ТОПОЛОГІЇ НЕЙРОМЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ ГРАФІЧНИХ ЗОБРАЖЕНЬ

Топології нейромоделей відносяться до структурної організації штучних нейронних мереж – того, як розташовані нейрони (вузли) і як визначаються зв'язки між ними. Це включає кількість шарів, тип шарів (наприклад, згортковий, повторюваний, повністю пов'язаний) та структуру з'єднань (наприклад, послідовні, пропускні з'єднання, посилення на основі уваги).

Топологія нейромоделі є критично важливою, оскільки вона безпосередньо впливає на здатність моделі навчатися та представляти складні закономірності в даних. Різні структури підходять для різних типів завдань: наприклад, згорткові топології дуже ефективні для обробки зображень завдяки своїй здатності фіксувати просторові ієрархії, тоді як повторювані або трансформаторні топології краще підходять для послідовних або контекстних даних. Добре розроблена топологія може значно підвищити ефективність навчання, зменшити параметри та покращити узагальнення, тоді як невідповідна структура може призвести до зниження продуктивності, перенавчання або надмірних обчислювальних витрат.

Сучасні топології нейронних мереж для розпізнавання графічних зображень значно еволюціонували за останнє десятиліття, перейшовши від класичних моделей, заснованих на згортці, до більш гнучких і масштабованих архітектур. Ці розробки зумовлені необхідністю підвищення точності, ефективності обчислень та здатності фіксувати складні просторові взаємозв'язки у візуальних даних.

Основною топологією розпізнавання зображень є згорткова нейронна мережа (CNN). CNN спеціально розроблені для обробки даних у вигляді сітки, таких як зображення, шляхом застосування згорткових фільтрів, які витягують ієрархічні ознаки. Ранні шари виявляють прості візерунки, такі як краї та текстури, тоді як глибші шари виявляють складні структури, такі як предмети та сцени. Таке ієрархічне виділення ознак робить CNN дуже ефективними для візуальних завдань, і вони залишаються домінуючим підходом у багатьох додатках. Сучасні архітектури CNN, такі як ResNet та ConvNeXt, містять такі вдосконалення, як залишкові з'єднання та оптимізовані згорткові блоки, що дозволяють ефективно навчати глибші мережі та досягати високої точності на великих наборах даних.

В останні роки в якості потужної альтернативи з'явилися моделі на основі трансформерів. Vision Transformer (ViT) представляє зображення у вигляді послідовностей фрагментів і обробляє їх, використовуючи механізми саморегуляції. На відміну від CNN, які зосереджуються на локальних полях сприйняття, ViT фіксують глобальні залежності по всьому зображенню на ранніх стадіях обробки. Ця можливість моделювати довгострокові взаємозв'язки дозволяє підвищити продуктивність при роботі зі складними наборами даних, особливо при наявності великомасштабної попередньої підготовки. Однак трансформаторам зазвичай потрібно більше даних та обчислювальних ресурсів порівняно з CNN, що може обмежити їх застосовність в умовах обмежених ресурсів.

Для усунення обмежень обох підходів були розроблені гібридні топології. Ці архітектури поєднують згорткові шари для ефективного вилучення локальних об'єктів з трансформаційними блоками для моделювання глобального контексту. Такі гібридні моделі використовують переваги обох парадигм і часто перевершують чисто згорткові або чисто трансформаторні системи. Недавні дослідження показують, що ці комбіновані архітектури забезпечують кращий баланс між точністю та обчислювальною складністю, що робить їх придатними для реальних застосувань.

Іншим напрямком, що розвивається, є використання альтернативних архітектур, таких як моделі на основі MLP (наприклад, MLP-мікшер), які покладаються виключно на багат шарові перцептрони для обробки просторової та каналної інформації. Ці моделі демонструють, що конкурентних результатів можна досягти навіть без використання звивин або механізмів уваги, підкреслюючи різноманітність можливих нейронних топологій для завдань розпізнавання зображень.

Сучасні топології нейромоделей для розпізнавання зображень включають CNN, Vision Transformers, гібридні архітектури CNN-transformer та новіші моделі на основі MLP. Кожен підхід має унікальні переваги: CNN забезпечують ефективність та сильне індуктивне зміщення, Трансформатори забезпечують розуміння глобального контексту, а гібридні моделі поєднують обидві переваги. Вибір топології залежить від конкретного завдання, розміру набору даних та обчислювальних обмежень, що робить цю область активною та швидко розвивається сферою досліджень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Optimized Modified K-Nearest Neighbor Classifier for Pattern Recognition / P. Dhabe et al. *Computational Intelligence in Image and Video Processing*. Boca Raton, 2023. P. 305–316.
URL: <https://doi.org/10.1201/9781003218111-16>

2. Shen Z., Kong B., Dong X. MAIM: a mixer MLP architecture for image matching. *The Visual Computer*. 2023. URL: <https://doi.org/10.1007/s00371-023-02851-9>
3. Transformers in Vision: A Survey / S.Khan et al. *ACM Computing Surveys*. 2022. URL: <https://doi.org/10.1145/3505244>