

ПЕРСПЕКТИВИ ПОЄДНАННЯ ВЕЙВЛЕТ- ПЕРЕТВОРЕННЯ З АРХІТЕКТУРОЮ ТРАНСФОРМЕР У ВИРІШЕННІ ЗАДАЧІ SUPER-RESOLUTION

Вінницький національний технічний університет¹

Анотація

Розглянуто особливості застосування нейронних мереж на основі архітектури трансформер до задачі super-resolution. Обґрунтовано перспективність застосування вейвлет-перетворення з метою підвищення якості відтворення зображень, та підвищення ефективності роботи нейронної мережі. Запропоновано декілька потенційних напрямків інтеграції вейвлет-перетворення у моделі super-resolution на основі архітектури трансформер.

Ключові слова: super-resolution, вейвлет-перетворення, трансформер, глибоке навчання, нейронні мережі, комп'ютерний зір

З появою Vision Transformer (ViT) [1] у 2021 році, архітектура трансформер отримала широке застосування у сфері комп'ютерного зору, зокрема і до задачі super-resolution (SR). Запропонована пізніше модель SwinIR [2], показала значний приріст у якості відтворення зображень в порівнянні з актуальними ЗНМ моделями, заклавши ґрунтовну основу для майбутнього розвитку. Ключовими перевагами архітектури трансформер відносно ЗНМ є механізм просторової самоуваги та здатність враховувати далекосяжні залежності у даних. Подальші дослідження дозволили досягти нових рівнів продуктивності, завдяки поєднанню архітектури трансформер з ЗНМ, поєднанню каналної і просторової уваги, дослідженню різних форм вікон просторової уваги [3, 4].

Проте, архітектура трансформер має і ряд властивостей, що ускладнюють її застосування для вирішення задачі SR. ViT в процесі виведення схильні фокусуватись на низькочастотних компонентах зображень, на протилежність ЗНМ, які природним чином виділяють високочастотні компоненти зображень [5]. Це, в поєднанні з функціями втрат на основі норм, призводить до втрат високочистої компоненти у відновлених зображеннях. Додатковим підтвердженням цього факту є ряд досліджень, у яких доповнення мереж на основі архітектури трансформер шарами ЗНМ демонструє позитивні зміни, як і у об'єктивних, так і у перцептивних оцінках відновлених зображень. Слабке індуктивне упередження, властиве для ViT, ускладнює процес навчання і вимагає значних обсягів навчальних даних. Посилення індуктивного упередження позитивно впливає, як на якість представлення, так і на стабільність та швидкість процесу навчання, що можна легко побачити на прикладі застосування локальних вікон уваги [5]. Ймовірно, що подальше обмеження ступенів свободи може додатково підвищити ефективність застосування архітектури трансформер у задачі SR.

Зменшити негативний вплив означених вище властивостей можливо за допомогою застосування вейвлет-перетворення [6]. Просторово-частотне представлення зображення дозволить як явно змістити фокус в сторону роботи з високочастотною складовою зображення, так і забезпечити модель апріорними знаннями щодо частотного розподілу даних, посиливши індуктивне упередження. Наскільки нам відомо, на сьогоднішній день існують лише дві роботи, що безпосередньо досліджують розширення нейронної мережі на основі архітектури трансформер за допомогою вейвлет-перетворення для вирішення задачі SR [7, 8]. Обидва дослідження продемонстрували вагоме збільшення ефективності (значне зменшення розмірів моделі) зі збереженням конкурентної якості відновлених зображень. Варто окремо відзначити [9], тут автори лише розширили функцію втрат за допомогою компоненту, який надає більшої ваги втратам у високочастотних піддіапазонах просторо-частотного представлення зображення, і експериментально показали, що така функція втрат забезпечує як приріст PSNR, так і кращу візуальну якість відтвореного зображення. У [10] досліджено застосування вейвлет-декомпозиції

для побудови вкладень ViT у задачі класифікації, для SR такий метод токенизації також виглядає привабливим, оскільки дозволить забезпечити модель вхідними даними у просторово-частотній формі.

Узагальнивши, можна виділити декілька напрямків щодо того, як вейвлет-перетворення може бути інтегровано до нейронної мережі SR на основі архітектури трансформер:

- модифікація функції втрат з метою збільшення ваги високочастотної компоненти, або обрахунок функції втрат виключно для зображень представлених у просторово-частотній формі;
- представлення вхідних даних у просторово-частотній формі або токенизація на основі вейвлет-декомпозиції;
- розробка механізмів самоуваги, які надаватимуть перевагу областям зображення з вищою концентрацією деталей;
- застосування зворотного вейвлет-перетворення для відтворення зображення високої роздільної здатності (підхід альтернативний до деконовлюції або pixelshuffle).

Висновки

Інтеграція вейвлет-перетворення з архітектурою трансформер має значний потенціал у вирішенні задачі SR. Очікується, що просторово-частотне представлення даних дозволить забезпечити якісніше відтворення високочастотної складової SR зображень. А також, посилить індуктивне упередження моделі, що згідно останніх досліджень, позитивно впливає як на процес навчання, так і на ефективність застосування архітектури трансформер. Наявні дослідження демонструють підвищення ефективності роботи моделей та якості відновлених зображень. Однак, їх кількість обмежена, тому напрямок має потенціал для розвитку.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. An image is worth 16x16 words: transformers for image recognition at scale [Електронний ресурс] / Alexey Dosovitskiy [та ін.] // The Ninth International Conference on Learning Representations, 3–7 трав. 2021 р. – [Б. м.]. – Режим доступу: <https://openreview.net/pdf?id=YicbFdNTTy> (дата звернення: 17.04.2025). – Назва з екрана.
2. SwinIR: Image Restoration Using Swin Transformer [Електронний ресурс] / Jingyun Liang [та ін.] // 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 жовт. 2021 р. – [Б. м.], 2021. – Режим доступу: <https://doi.org/10.1109/iccvw54120.2021.00210> (дата звернення: 17.04.2025). – Назва з екрана.
3. Activating More Pixels in Image Super-Resolution Transformer [Електронний ресурс] / Xiangyu Chen [та ін.] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 черв. 2023 р. – [Б. м.], 2023. – Режим доступу: <https://doi.org/10.1109/cvpr52729.2023.02142> (дата звернення: 17.04.2025). – Назва з екрана.
4. Dual Aggregation Transformer for Image Super-Resolution [Електронний ресурс] / Zheng Chen [та ін.] // 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 жовт. 2023 р. – [Б. м.], 2023. – Режим доступу: <https://doi.org/10.1109/iccv51070.2023.01131> (дата звернення: 17.04.2025). – Назва з екрана.
5. Park N. How do vision transformers work [Електронний ресурс] / Namuk Park, Songkuk Kim // The Tenth International Conference on Learning Representations, 22–29 квіт. 2022 р. – [Б. м.]. – Режим доступу: <https://openreview.net/pdf?id=D78Go4hVcxO> (дата звернення: 17.04.2025). – Назва з екрана.
6. Wavelet Tour of Signal Processing: The Sparse Way. – [Б. м.] : Elsevier Science & Technology Books, 2008.
7. Ran J. Lightweight Wavelet-Based Transformer for Image Super-Resolution [Електронний ресурс] / Jinye Ran, Zili Zhang // 19th Pacific Rim International Conference on Artificial Intelligence, Shanghai, 10–13 квіт. 2022 р. – [Б. м.]. – Режим доступу: https://doi.org/10.1007/978-3-031-20868-3_27 (дата звернення: 17.04.2025). – Назва з екрана.
8. SVTSR: image super-resolution using scattering vision transformer [Електронний ресурс] / Jiabao Liang [та ін.] // Scientific Reports. – 2024. – Т. 14, № 1. – Режим доступу: <https://doi.org/10.1038/s41598-024-82650-x> (дата звернення: 17.04.2025). – Назва з екрана.
9. Korkmaz C. Training Transformer Models by Wavelet Losses Improves Quantitative and Visual Performance in Single Image Super-Resolution [Електронний ресурс] / Cansu Korkmaz, A. Murat Tekalp // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 17–18 черв. 2024 р. – [Б. м.], 2024. – С. 6661–6670. – Режим доступу: <https://doi.org/10.1109/cvprw63382.2024.00660> (дата звернення: 17.04.2025). – Назва з екрана.
10. Zhu Z. Wavelet-Based Image Tokenizer for Vision Transformers [Електронний ресурс] / Zhenhai Zhu, Radu Soricut. – [Б. м. : б. в.], 2024. – 13 с. – (2405.18616). – Режим доступу: <https://arxiv.org/pdf/2405.18616> (дата звернення: 17.04.2025). – Назва з екрана.

Козлов Сергій Леонідович – аспірант кафедри комп’ютерних наук, Вінницький національний технічний університет, Вінниця, e-mail: serhii.kozlov@gmail.com

Науковий керівник: **Колесницький Олег Костянтинович** – кандидат технічних наук, професор кафедри комп’ютерних наук, Вінницький національний технічний університет, Вінниця

Abstract

This paper explores the application of transformer-based neural networks to the super-resolution task. The potential advantages of employing wavelet transforms to enhance image reconstruction quality and network efficiency are discussed. Several promising directions for integrating wavelet transforms into transformer-based super-resolution models are proposed.

Keywords: super-resolution, wavelet transform, transformer, deep learning, neural networks, computer vision

Serhii Leonidovich Kozlov – PhD student at the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia, email: serhii.kozlov@gmail.com

Supervisor: ***Oleh Kostiantynovych Kolesnytskyi*** – PhD, Professor at the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia.