

А.А. Демчишин

ОЦІНКА ЕФЕКТИВНОСТІ МОДЕЛЕЙ МЕРЕЖ ГЛИБИННОГО НАВЧАННЯ ПІДВИЩЕННЯ РОЗДІЛЬНОЇ ЗДАТНОСТІ ЗОБРАЖЕНЬ

Анотація. На сьогоднішній день мережі глибинного навчання демонструють значну перевагу перед іншими алгоритмами комп'ютерного бачення. Розглядається задача оцінки ефективності підвищення інформативності зображень, які отримано з використанням двох найновіших на даний час моделей нейронних мереж: Residual Dense Network for Image Super-Resolution та Enhanced Super-Resolution Generative Adversarial Network. В якості метрик інформативності зображення застосовано: функцію розподілу щільності вірогідності відтінків пікселів; значення ентропії, яке характеризує міру різноманітності кожного окремого пікселу вибірки; та значення ємності каналу для передачі повідомлення, яке стиснено на основі алгоритму LZ77.

Одержані результати свідчать про підвищення інформативності зображення у вигляді росту ентропії за Шенноном з 7.43 до 7.81 та 7.62 для мережі ESRGAN та RDN відповідно. Водночас, показано, що класична білінійна інтерполяція вибірки може з успіхом використовуватись для підвищення її інформативності (7.62). Продемонстровано, що метрика LZ77, яка вимірює ємність каналу для передачі повідомлення, є більш адекватною для оцінки візуальної цінності повідомлення, ніж метрика ентропії за Шенноном.

Ключові слова: мережа глибинного навчання, згорткова нейронна мережа, роздільна здатність зображення, гістограма, ентропія, LZ77.

Постановка проблеми. Алгоритми штучного інтелекту почали застосовуватись ще в 70-х роках минулого сторіччя, але через нерозвиненість теоретичної бази та слабкість обчислювальних потужностей машин того часу, результати проєктів на його основі були досить обмеженими [1]. Глибинне навчання – це набір алгоритмів машинного навчання, які націлено на тренування моделі, що складається з декількох шарів обробки даних. На сьогодні глибинне навчання показало значну перевагу перед іншими комп'ютерними алгоритмами в таких сферах, як: комп'ютерне бачення, розпізнання та обробка природної мови [2].

Фундаментальна проблема передачі інформації між двома агентами, яку сформулював Шеннон [3], полягає в якомога точнішому відтворенні повідомлення, яке передано іншою стороною. Типовий ланцюг комунікаційної системи представлено на рис.1:

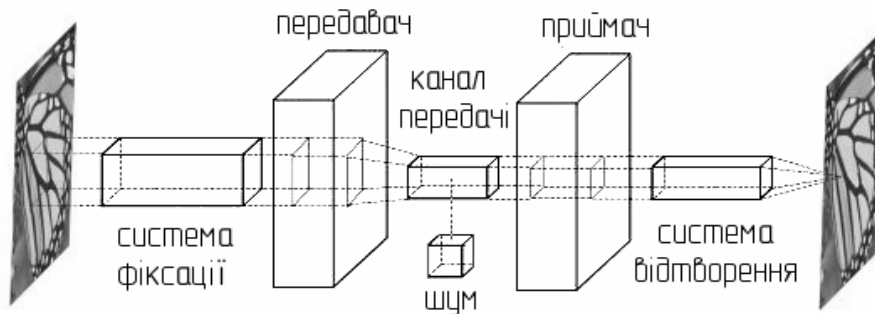


Рисунок 1 – Діаграма системи комунікації

Кожний з елементів системи характеризується власною пропускнуною здатністю, та може спричиняти деградацію повідомлення. Задача підвищення роздільної здатності зображення є актуальною не тільки в контексті комунікаційних систем, але й в системах синтезу зображень, які, наприклад, використовують алгоритм зворотного трасування променів [4].

Успіхи мереж глибокого навчання, які будуть описані нижче, надали змогу підвищувати роздільну здатність зображень, реконструюючи інформацію на основі попереднього досвіду.

Аналіз останніх досліджень. Задача реконструкції зображень та підвищення просторової розрізненості синтезованих зображень добре представлена в роботах науковців Дніпровської школи - розвиток цифрової обробки зображень відбувався за рахунок вдосконалення алгоритмів компресії [5], застосування фільтрів, що прибирають зайву інформацію [6], алгоритмів геометричної корекції зображення [7], та розширенню застосування методів вейвлет-аналізу [8].

Набутки глибокого навчання базуються на двох основних складових: прогрес обчислювального обладнання (загальна доступність процесорів *Single Instruction Multiple Data* архітектури [9]) та розвиток самих алгоритмів навчання. Одними з найбільш перспективних моделей мереж підвищення роздільної здатності на даний час є: мережа вивчення відхи-лу на основі густо зв'язаних шарів (*Residual Dense Network*) [10] та гене-

ративна змагальна мережа (Generative Adversarial Network) на основі RDN [11].

Зазвичай, оцінка якості (рівня спотворень) зображення проводиться на основі значення співвідношення пікового рівня сигналу до шуму (*Peak Signal Noise Ratio*), яке визначається на основі середньоквадратичного відхилення (*Mean Square Error*). Для двох монохромних зображень I та K розміром $m \times n$, одне з яких вважається спотвореним наближенням іншого, MSE розраховується наступним чином:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |I(i, j) - K(i, j)|^2 ; \quad (1)$$

в свою чергу PSNR визначається як:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (2)$$

де MAX_I – максимальне значення, яке може прийняти піксел зображення.

Не дивлячись на те, що адекватність метрики $PSNR$ викликає багато питань [12], вона є стандартом при порівнянні результатів роботи мереж.

Іншою метрикою сприйняття зображення, яка дає кількісну оцінку різниці двох зображень, є індекс структурної подібності (*Structure SIMilarity*). Розрахунок $SSIM$ ведеться на основі трьох порівняльних коефіцієнтів [13]:

$$\begin{aligned} luminance(x, y) &= \frac{2\mu_x\mu_y + const_1}{\mu_x^2 + \mu_y^2 + const_1}, \\ contrast(x, y) &= \frac{2\sigma_x\sigma_y + const_2}{\sigma_x^2 + \sigma_y^2 + const_2}, \\ structure(x, y) &= \frac{\sigma_{xy} + const_3}{\sigma_x\sigma_y + const_3}, \end{aligned} \quad (3)$$

де $x = \{x_i | i = 1, 2, \dots, N\}$ та $y = \{y_i | i = 1, 2, \dots, N\}$ – два дискретних невід’ємних сигнали, які було синхронізовано між собою; $const_{1,2,3}$ – константи, введені для попередження ділення на нуль в разі малого знаменника.

Автор роботи [14] аргументовано підкреслює, що індекс $SSIM$ є лише критерієм оцінки подібності між вибірками, і не може використо-

уватись для формування судження про те, яке зображень має кращу якість.

Як буде показано нижче, повідомлення, реконструйоване мережею глибокого навчання, може відрізнитись від початкової вибірки (автори роботи [11] вживають фразу «вбірка набуває деталей завдяки галюцинаціям мережі»), але, в той же самий час, його елементи виглядають цілісно, об'єкти мають чіткі краї; навіть якщо еталонне зображення містило артефакти JPEG компресії, вони можуть бути прибрані. Іншими словами, зображення, реконструйоване нейронною мережею, може виглядати суб'єктивно краще за оригінал.

Нееталонні метрики - це такі, що не потребують опорного або еталонного зображення в процесі підрахунку оцінки якості зображення.

Формування цілей статті. Таким чином актуальною є задача дослідження ефективності моделей мереж підвищення роздільної здатності зображень з використанням критеріїв, які не потребують еталонних зображень, що надасть можливість окреслити межі адекватності моделей.

Основна частина. В якості вихідної вибірки візьмемо контрастне зображення з великою кількістю деталей, яке виконано в низькому ключі (рис.2). Для підготовки тестового зображення низької роздільної здатності застосуємо алгоритм децимації. Суть алгоритму полягає в передачі лише кожного n -го пікселя вихідного зображення високої роздільної здатності. Вибір алгоритму децимації (без попереднього застосування фільтру низької частоти) визначено умовою імітації ефекту *аліасингу* - типового артефакту зображення при його аналого-цифровому перетворенні.

Найперше задамося питанням, як змінюється розподіл яскравості зображень, отриманих на основі нейронних мереж глибокого навчання моделі RDN [10] та ESRGAN [11], відносно вихідного. Оцінку ефективності роботи мереж будемо проводити, порівнюючи із зображенням, роздільну здатність якого піднято на основі класичного алгоритму бікубічної інтерполяції.

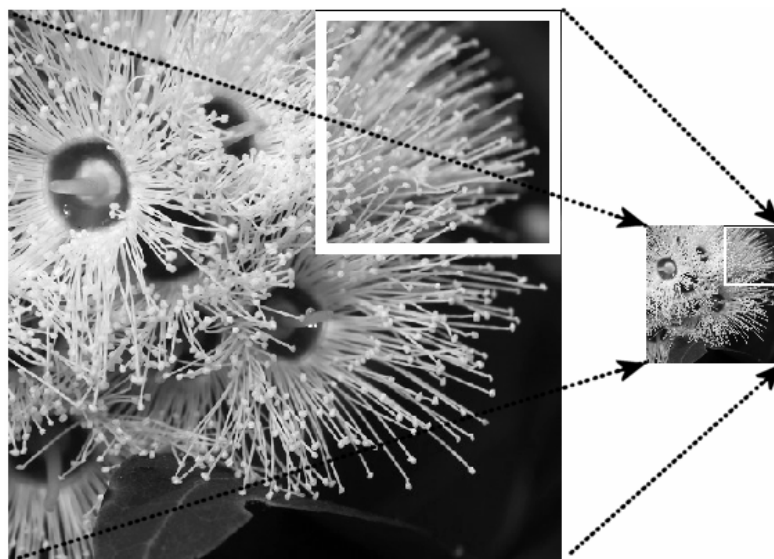


Рисунок 2 – Підготовка тестової фотографії низької роздільної здатності (100x100) на основі алгоритму децимації (квадратом виділено фрагмент детального аналізу)

Якщо X - зображення у відтінках сірого, то функцію розподілу густини вірогідності появи пікселя з відтінком r може бути розраховано, як:

$$p_r(r_j) = n_j/n, \quad 0 \leq j < L, \quad (4)$$

де n_j - кількість пікселів з відтінком r_j ; n - загальна кількість пікселів у зображенні; L - загальна кількість присутніх відтінків. Функція (4) є гістограмою зображення X .

Порівняємо гістограми зображень із підвищеною роздільною здатністю:

- гістограма зображення на рис.3а є зміщеною вліво, що є типовим для фото, яке містить світлі об'єкти на темному фоні. Шум графіку функції походить від застосування цифрової операції вирівнювання діапазону яскравості вихідного зображення. Недостатня роздільна здатність вибірки сигналу проявляється у вигляді спотворенням типу «східці»;

- гістограма зображення на рис.3б повторює гістограму оригінального зображення з різницею, що шум графіку згладжено;

- графік розподілу густини вірогідності розподілу яскравості зображення на основі моделі ESRGAN (рис.3в) демонструє тенденцію зміщення яскравих півтонів ближче до точки білого; в той же час з'являється відсічення темних та яскравих тонів, що є індикатором втра-

ти інформативності зображення завдяки виходу за межі допустимого діапазону;

- гистограма зображення на основі моделі RDN (рис.3г) показує перерозподіл кількості темних тонів фону в бік півтонів, що є маркером підвищення контрастності фотографії. Відношення кількості півтонів до світлих тонів - те саме, що й у випадку моделі ESRGAN.

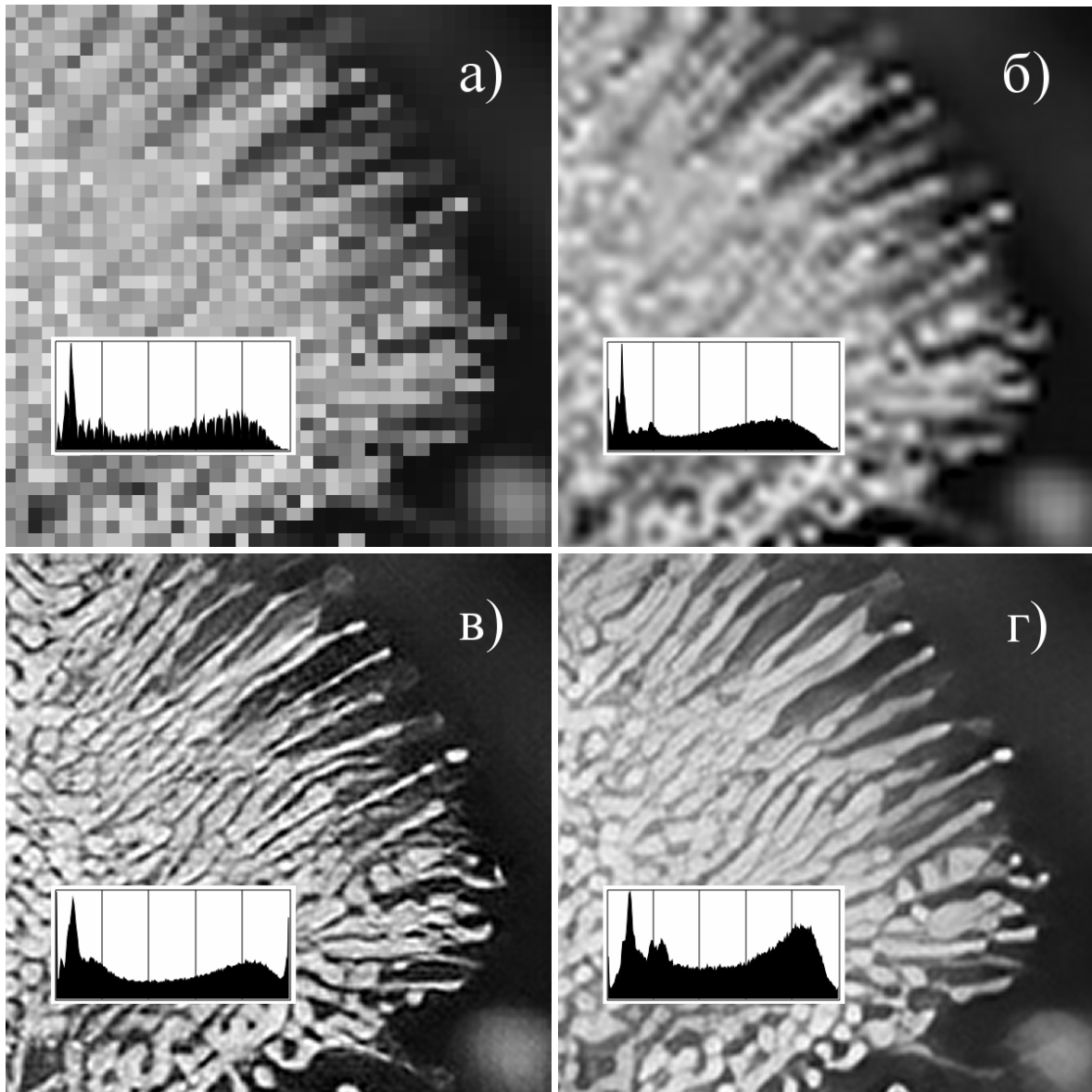


Рисунок 3 – Ілюстрація чотириразового підвищення роздільної здатності зображення (з 42x42 до 168x168 пікселів):

- а) початкова вибірка у масштабі 4:1;
- б) на основі бікубічної інтерполяції;
- в) на основі мережі SRGAN; г) на основі мережі RDN

Іншою об'єктивною метрикою, яка є тісно пов'язаною з функцією (2), є ентропія за Шенноном, що виступає мірою усередненого інформаційного вмісту повідомлення. Якщо розглянути послідовність пікселів зображення як динамічну систему, тоді $p_r(r_j)$ є вірогідністю знаходження системи у стані r_j власного фазового простору. Базовий принцип теорії інформації говорить [3], що інформативність повідомлення, яке передається, залежить від того, в якій мірі зміст повідомлення є неочікуваним. Якщо стан, в який переходить система, є дуже ймовірним, то передача такого повідомлення несе мало інформації. Маючи розраховані вірогідності появи піксела з відтінком r_j , функція ентропії за Шенноном записується у наступному вигляді:

$$H = -\sum_k p_r(r_j) \log_2(p_r(r_j)) \quad (5)$$

Порівнюючи стовпчики зі значеннями міри для вихідних зображень високої та низької роздільної здатності, з таблиці 1 видно, що останнє побіднішало внаслідок видалення пікселів. В свою чергу, інформаційний зміст вихідного зображення при застосування моделей підвищення роздільної здатності зріс. Міра H , розрахована на основі функції вірогідності (4), повністю проігнорувала просторову структуру повідомлення, - зображення, отримані на основі бікубічної інтерполяції та моделі мережі RDN, мають однакову розраховану міру інформативності.

З тієї ж теорії інформації [3] відомо, що ентропія визначає ємність каналу передачі інформації, стиснутої з використанням найбільш ефективного кодування. Таким чином, задачу оцінки інформативності зображення може бути переформульовано в задачу оцінки ємності, необхідної для передачі стиснених пікселів зображення. Таку оцінку може бути проведено із використанням, наприклад, алгоритму стиснення без втрат LZ77 [15]. Алгоритм Лемпеля-Зіва стикає дані за рахунок побудови словника фраз, присутніх в повідомленні раніше. В такий спосіб цінність повідомлення вираховується не тільки виходячи з окремих пікселів, а на основі горизонтальних елементів зображення.

Як видно з останнього рядка таблиці 1, оцінка LZ77, порівняно з оцінкою за функцією (5), краще відображає значну втрату інформативно-

сті вихідного зображення внаслідок децимації, та чіткішу структуру елементів зображень, отриманих на основі мереж глибинного навчання, відносно до вмісту зображення на основі бікубічної інтерполяції. Слід зазначити й певну кореляцію між двома оцінками - зображення з підвищеною роздільною здатністю, отримані на основі бікубічної інтерполяції та з використанням нейронних мереж, мають більшу інформативність, ніж знімок первинного джерела інформації.

Таблиця 1

Значення міри усередненого інформаційного вмісту зображення

	Вихідне зображення високої розд.здат.	Вихідне зображення низької розд.здат.	Бікубічна інтерполяція	ESRGAN	RDN
Ентропія за Шенноном	7.43	7.38	7.62	7.81	7.62
LZ77	24.67	4.49	25.42	28.47	27.16

Висновки. Одержані результати свідчать про підвищення інформативності зображень при використанні мереж глибинного навчання моделей ESRGAN та RDN підвищення роздільної здатності. Водночас, показано, що класична білінійна інтерполяція вибірки може з успіхом використовуватись для підвищення її інформативності. Продемонстровано, що метрика LZ77, яка вимірює ємність каналу для передачі повідомлення, є більш адекватною для оцінки візуальної цінності повідомлення, ніж метрика ентропії за Шенноном.

На даний час ведуться подальші дослідження, пов'язані з адаптацією моделей мереж глибинного навчання для роботи з груповими фотографіями людей.

ЛІТЕРАТУРА / LITERATURE

1. The History of Artificial Intelligence / C. Smith, B. McGuire, T. Huang, G. Yanhg // History of Computing, Univ. of Washington. – 2006. – 27 p.
2. 50 Years of Artificial Intelligence / M. Lungarella, F. Iida, J. Bongard, R. Pfeifer // Essays Dedicated to the 50th Anniversary of Artificial Intelligence. – Springer, 2007. – 399 p.
3. A Mathematical Theory of Communication / C. Shannon // Reprinted with corrections from The Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423, 623–656.

4. NVIDIA Reinvents Computer Graphics with Turing Architecture [Електронний ресурс] // Gail Laguna. – 2018. Режим доступу до ресурсу: <https://nvidianews.nvidia.com/news/nvidia-reinvents-computer-graphics-with-turing-architecture>
5. Гнатушенко В.В. Реконструкція растрових зображень на основі їх інваріантної геометричної моделі / В.В. Гнатушенко, О.В. Реута // Прикладна геометрія та інженерна графіка / Праці Таврійської державної агротехнічної академії. – Мелітополь: ТДАТА, 2004. – Вип. 4, т. 24. – С. 99–102.
6. Корчинський В.М. Аналіз методів ідентифікації геометричних форм на фотограмметричних цифрових зображеннях дистанційного зондування Землі / В.М. Корчинський, О.В. Спирінцева // Прикладна геометрія та інженерна графіка / Праці Таврійського державного агротехнічного університету. – Мелітополь: ТДАТУ, 2014. – Вип. 4, т.58. – С. 50–57.
7. Гнатушенко В.В. Оцінка ефективності геометричних і радіометричних перетворень багатоканальних цифрових зображень / В.В. Гнатушенко, О.О. Сафаров // Прикладна геометрія та інженерна графіка / Праці Таврійського державного агротехнічного університету. – Мелітополь: ТДАТУ, 2014. – Вип. 4, т. 58. – С. 29–33.
8. Гордієнко О.М. Синтез растрових проекційних зображень на основі суміщення просторових та спектральних складових / О.М. Гордієнко // Прикладна геометрія та інженерна графіка / Праці Таврійської державної агротехнічної академії. – Мелітополь: ТДАТА, 2004. – Вип. 4, т. 28. – С. 109-113.
9. Tearing apart Google's TPU 3.0 AI coprocessor [Електронний ресурс] // Paul Teich. – 2018. Режим доступу до ресурсу: <https://www.nextplatform.com/2018/05/10/tearing-apart-googles-tpu-3-0-ai-coprocessor/>
10. Residual Dense Network for Image Super-Resolution [Електронний ресурс] / Yulun Zhang, Yapeng Tian, Yu Kong та ін. // arXiv. – 2018. – Режим доступу до ресурсу: <https://arxiv.org/abs/1802.08797>.
11. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [Електронний ресурс] / Xintao Wang, Ke Yu, Shixiang Wu та ін. // arXiv. – 2018. – Режим доступу до ресурсу: <https://arxiv.org/abs/1809.00219>.
12. Image quality assessment: from error visibility to structural similarity / Zhou Wang, A. Bovik, H. Sheikh, E. Simoncelli // IEEE Transactions on Image Processing. – 2004. – vol. 13, no. 4. – P. 600–612.
13. Wang Z. Multiscale structural similarity for image quality assessment / Z. Wang, E.P. Simoncelli, A.C. Bovik // Conference Record of the Thirty-Seventh

Asilomar Conference on Signals, Systems and Computers. – 2004. – Vol. 2. – P. 1398–1402.

14. Старовойтов В. В. Индекс SSIM не является метрикой и плохо оценивает сходство изображений / В. В. Старовойтов // Системный ан. и прикл. информатика. – Минск: БНТУ, 2019. – № 2. – С. 12–17.

15. Avinery R. Universal and Accessible Entropy Estimation Using a Compression Algorithm / R. Avinery, M. Kornreich, R. Beck // Phys. Rev. Lett. 123, 178102, 2019. – 13p.

REFERENCES

1. The History of Artificial Intelligence / C. Smith, B. McGuire, T. Huang, G. Yanhg // History of Computing, Univ. of Washington. – 2006. – 27 p.

2. 50 Years of Artificial Intelligence / M. Lungarella, F. Iida, J. Bongard, R. Pfeifer // Essays Dedicated to the 50th Anniversary of Artificial Intelligence. – Springer, 2007. – 399 p.

3. A Mathematical Theory of Communication / C. Shannon // Reprinted with corrections from The Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423, 623–656.

4. NVIDIA Reinvents Computer Graphics with Turing Architecture [Internet resource] // Gail Laguna. – 2018. Rezhym dostupu do resursu: <https://nvidianews.nvidia.com/news/nvidia-reinvents-computer-graphics-with-turing-architecture>

5. Hnatushenko V.V. Rekonstruktsiia rastrovyykh zobrazhen na osnovi yikh invariantnoi heometrychnoi modeli / V.V. Hnatushenko, O.V. Reuta // Prykladna heometriia ta inzhenerna hrafika / Pratsi Tavriiskoi derzhavnoi ahrotekhnichnoi akademii. – Melitopol: TDATA, 2004. – Issue. 4, vol. 24. – P.99–102.

6. Korchynskiy V.M. Analiz metodiv identyfikatsii heometrychnykh form na fotogrammetrychnykh tsyfrovyykh zobrazhenniakh dystantsiinoho zonduvannya Zemli / V.M. Korchynskiy, O.V. Spirintseva // Prykladna heometriia ta inzhenerna hrafika / Pratsi Tavriiskoho derzhavnoho ahrotekhnichnoho universytetu. – Melitopol: TDATU, 2014. – Issue. 4, vol.58. – P. 50–57.

7. Hnatushenko V.V. Otsinka efektyvnosti heometrychnykh i radiometrychnykh peretvoren bahatokanalnykh tsyfrovyykh zobrazhen / V.V. Hnatushenko, O.O. Safarov // Prykladna heometriia ta inzhenerna hrafika / Pratsi Tavriiskoho derzhavnoho ahrotekhnichnoho universytetu. – Melitopol: TDATU, 2014. – Issue. 4, vol. 58. – P. 29–33.

8. Hordiienko O.M. Syntez rastrovyykh proektsiinykh zobrazhen na osnovi sumishchennia prostorovykh ta spektralnykh skladovykh / O.M. Hordiienko // Prykladna heometriia ta inzhenerna hrafika / Pratsi Tavriiskoi derzhavnoi

ahrotekhnichnoi akademii. – Melitopol: TDATA, 2004. – Issue. 4, vol. 28. – P.109-113.

9. Tearing apart Googles TPU 3.0 AI coprocessor [Internet resource] // Paul Teich. – 2018. Source: <https://www.nextplatform.com/2018/05/10/tearing-apart-googles-tpu-3-0-ai-coprocessor/>

10. Residual Dense Network for Image Super-Resolution [Internet resource] / Yulun Zhang, Yapeng Tian, Yu Kong and others // arXiv. – 2018. – Source: <https://arxiv.org/abs/1802.08797>.

11. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [Elektronnyi resurs] / Xintao Wang, Ke Yu, Shixiang Wu ta in. // arXiv. – 2018. Source: <https://arxiv.org/abs/1809.00219>.

12. Image quality assessment: from error visibility to structural similarity / Zhou Wang, A. Bovik, H. Sheikh, E. Simoncelli // IEEE Transactions on Image Processing. – 2004. – vol. 13, no. 4. – P. 600–612.

13. Wang Z. Multiscale structural similarity for image quality assessment / Z. Wang, E.P. Simoncelli, A.C. Bovik // Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers. – 2004. – Vol. 2. – P. 1398–1402.

14. Starovoitov V.V. Indeks SSIM ne iavliaetsia metrikoi i plokhо otsenivaet skhodstvo izobrazhenii / V. V. Starovoitov // Sistemnii an. i prikladnaia informatika. – Minsk: BNTU, 2019. – № 2. – P. 12–17.

15. Avinery R. Universal and Accessible Entropy Estimation Using a Compression Algorithm / R. Avinery, M. Kornreich, R. Beck // Phys. Rev. Lett. 123, 178102, 2019. – 13p.

Received 20.11.2019.

Accepted 24.11.2019.

Оценка эффективности моделей сетей глубокого обучения повышения разрешения изображений

В работе рассматривается задача оценки эффективности повышения информативности изображений, полученных с использованием двух новейших на теперешнее время моделей нейронных сетей: Residual Dense Network for Image Super-Resolution и Enhanced Super-Resolution Generative Adversarial Network.

Полученные результаты свидетельствуют о повышении информативности изображений в виде роста энтропии по Шеннону с 7.43 до 7.81 и 7.62 для сети ESRGAN и RDN соответственно. В то же время, показано, что классическая билинейная интерполяция выборки может с успехом использоваться для повышения ее информативности (7.62). Продемонстрировано, что метрика LZ77, которая измеряет емкость канала для передачи сообщения, является более адекватной для оценки визуальной ценности сообщения, чем метрика энтропии по Шеннону.

Evaluation of the efficiency of models of deep learning networks increasing the resolution of images

Since a decade, deep learning networks have shown a significant advantage over other algorithms of computer vision. The paper considers a task of evaluating the efficiency of increasing information content (self-information) of images generated using the two latest models of neural networks: the Residual Dense Network (RDN) for Image Super-Resolution and Enhanced Super-Resolution Generative Adversarial Network (ESRGAN). The increase in information content occurs in a context of creating the image with a fourfold resolution in comparison to the original image. The low resolution test sample is prepared using a decimation algorithm. The algorithm preserves only every n -th pixel of the high resolution original image. The way the low resolution image is prepared simulates the aliasing effect very well. Aliasing effect is a typical image artifact inherent to analog-to-digital conversion.

Usually the quality (distortion) of the image is evaluated basing on values of Peak Signal Noise Ratio and Structure SIMilarity (so called full-reference metrics that perform comparison against etalon image). At the same time the image reconstructed by the deep learning network may subjectively differ from the original sample. The paper uses three no-reference metrics to measure the information content of the image: a histogram displaying the probability density function of pixel shades; entropy value characterizing the degree of diversity of each individual pixel in the sample; and a channel capacity value needed for transmitting the message compressed using LZ77 algorithm. These metrics relate to the so-called "objective".

Acquired results indicate an increase of Shannon entropy when using deep learning networks of the ESRGAN and RDN models from 7.43 to 7.81 and 7.62 respectively. At the same time, it is shown that classical bilinear interpolation of the sample can be successfully applied to increase information content (7.62). It has been demonstrated that LZ77 metric measuring the capacity of a channel for transmitting a message is more adequate for assessing visual quality of the message than Shannon entropy metric.

Демчишин Анатолій Анатолійович – доцент кафедри автоматизації проектування енергетичних процесів та систем, Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського".

Демчишин Анатолий Анатольевич – доцент кафедры автоматизации проектирования энергетических процессов и систем, Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского".

Demchyshyn Anatoliy – assistant professor, department of automation of design of energy processes and systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".