

Ієрархічна згорткова нейронна мережа для підвищення роздільної здатності інфрачервоних зображень

Ярошенко^f М. О., ORCID [0000-0002-3092-3856](https://orcid.org/0000-0002-3092-3856)

Варфоломєєв^s А. Ю., к.т.н. доц., ORCID [0000-0002-6990-7140](https://orcid.org/0000-0002-6990-7140)

Яганов^s П. О., к.т.н. доц., ORCID [0000-0001-7358-9846](https://orcid.org/0000-0001-7358-9846)

Кафедра конструювання електронно-обчислювальної апаратури, Факультет електроніки
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського» ROR [00syn5v21](https://ror.org/00syn5v21)
Київ, Україна

Анотація—Через відносно високу ціну інфрачервоних сенсорів існує необхідність у методах обробки інфрачервоних зображень, отриманих із недорогих ІЧ-камер, які б підвищували їхню роздільну здатність. Одним з напрямків таких методів є надроздільна здатність (SR – Super Resolution) на основі згорткових нейронних мереж (CNN – Convolutional Neural Network), що активно розвивається останнє десятиріччя. Однак більшість існуючих нейронних SR-мереж створено для збільшення роздільної здатності зображень у видимому спектрі. В цьому ж дослідженні на інфрачервоних зображеннях перевірено працездатність однієї з кращих мереж для збільшення роздільної здатності зображень у видимому спектрі – BCLSR, та запропоновано власну мережу, яка є швидшою та точніше відновлює (збільшує) роздільну здатність вхідних зображень.

Ключові слова — підвищення роздільної здатності зображень; згорткові нейронні мережі; рекурентні нейронні мережі; інфрачервоні зображення

I. ВСТУП

В останні десятиліття застосування тепловізорів розширюється завдяки експлуатації наземних, бортових та космічних платформ. Наземні спостереження можуть використовуватися, наприклад, для виявлення втрат теплової енергії будівлями або для моніторингу робочих умов протікання промислових виробничих процесів, тоді як аерофотозйомка застосовується для підтримки міського планування та точного землеробства і т.д. Усі зазначені застосування потребують точної і докладної інформації про температуру поверхонь та їх просторову структуру.

Незважаючи на значний розвиток технологій виробництва теплових сенсорів, викликаний зростаючим попитом, вони досі мають низьку роздільну здатність, яка є помітно нижчою в порівнянні з зображеннями, отриманими у видимій смузі або в ближній інфрачервоній області. Виключаючи спеціалізоване супутникове обладнання, більшість наявних на ринку інфрачервоних камер мають роздільну здатність менше одного мегапікселя. З апаратної точки зору подальше збільшення роздільної здатності ускладнене тим, що або зменшення розміру пікселя призведе до зниження відношення сигнал/шум, або збільшення їхньої кількості призведе до збільшення розміру сенсора, що було б занадто дорого для пристроїв загального призначення.

У цьому контексті надроздільна здатність, що являє собою процес генерування зображення з більш високою роздільною здатністю (HR – High Resolution) на основі з одного або декількох кадрів нижчої роздільної здатності (LR – Low Resolution), є перспективним підходом для подолання поточних технологічних обмежень і забезпечення можливості аналізу дрібних деталей температурних характеристик поверхонь [1].

Провідні виробники тепловізорів почали впроваджувати алгоритми надроздільної здатності в програмне забезпечення своїх останніх моделей. Наприклад, компанія FLIR розробила технологію UltraMax, яка заснована на швидкому отриманні 16 теплових кадрів менш ніж за 1 с [2]. Через природні рухи оператора, кожне зображення буде дещо зміщене щодо інших. В результаті виходить нове зображення з подвоєною роздільною здатністю за умови, що користувач або ціль майже нерухомі [1]. Замість цього Infratech використовує апаратне рішення, що складається з швидкого обертового колеса MicroScan [3], яка дозволяє отримувати чотири експозиції, кожна з яких зміщена в поперечному напрямку на половину пікселя. Інші програмні рішення засновані на об'єднанні даних із зображеннями у видимій смузі. FLIR MSX [4], наприклад, підвищує чіткість зображення шляхом нанесення деталей зображення видимого спектру на теплові зображення. Однак такий алгоритм, імовірно, спотворить теплове вимірювання, тому що більш



дрібні деталі беруться з видимої смуги і не обов'язково відповідають змінам температури поверхонь, що спостерігаються [1].

Принципово іншим є підхід із застосуванням нейронних мереж, який ґрунтується на наявності попереднього «знання» алгоритмом типових форм об'єктів у зображенні низької роздільної здатності [5]. У 2015 році була вперше [6] застосована згорткова нейронна мережа (SRCNN – Super Resolution Convolutional Neural Network) для реконструкції зображення. SRCNN є згортковою нейронною мережею для відновлення високої роздільної здатності зображень, яка навчається на великій кількості пар зображень із різною роздільною здатністю. Хоча запропонований у [6] алгоритм і є кращим за звичайну бікубічну інтерполяцію, тришарова структура SRCNN все ж не дозволяє досягти більш високої якості зображень, оскільки мала кількість шарів дає лише «поверхневу» інформацію про вміст зображення. Щоб отримати більш детальні характеристики, а отже і кращий результат реконструкції, можна спробувати використовувати більш глибоку мережу, що власне і було запропоновано в методі VDSR [7], який має вже 20 згорткових шарів [5].

Зображення з низькою роздільною здатністю у вищезазначених підходах спочатку бікубічно інтерполюється, а потім оброблюється нейронною мережею. Для досягнення продуктивності в реальному часі було запропоновано мережу ESPCN [8], яка безпосередньо обробляє зображення з низькою роздільною здатністю і забезпечує її підвищення за рахунок збільшення числа каналів характеристичних ознак (feature channels) [5].

SRGAN [9] використовує GAN (змагальна породжуюча мережа) для поліпшення деталей зображення. SRGAN, хоч і має значення PSNR (Peak Signal-to-Noise Ratio – пікове співвідношення сигнал-шум) нижче за ESPCN, але згенеровані нею зображення мають вищий показник MOS (Mean Opinion Score), тобто, на думку незалежних оцінювачів, суб'єктивно виглядають більш якісними та близьким до оригіналу.

Робота [10] є однією з перших, де розглядається питання підвищення роздільної здатності теплових зображень за допомогою згорткових нейронних мереж. В цій роботі використовуються ідеї з праці [6], а її автори порівнюють точність мережі, навченої на зображеннях з різних спектральних діапазонів, щоб отримати краще уявлення про можливості застосування згорткових нейронних мереж для підвищення роздільної здатності теплових зображень. Вони, зокрема, прийшли до висновку, що мережа, навчена на відтінках сірого, забезпечує кращий результат масштабування, ніж навчена на безпосередньо теплових зображеннях MWIR-спектру (Mid-wavelength infrared) [11].

В роботі [12] запропоновано згорткову нейронну мережу для реконструкції теплових зображень, що навчається методом залишкового навчання (residual learning). При цьому автори роботи, з огляду на обмеженість навчальних вибірок з тепловими зображеннями, використовують для навчання звичайні

кольорові зображення, представлені у спеціальних просторах кольору та оцінюють ефективність використання для навчання однієї з кольорових компонент даних просторів, а саме: відтінки сірого, яскравість L в просторі HSL, інтенсивність I в просторі HSI та компонент V з кольорового простору HSV. Автори приходять до висновку, що навчання на V компоненті дає найкращий результат відновлення, а запропонований ними підхід також ще й покращує швидкість навчання [11].

Слід відмітити, що в роботі [11], на відміну від [10] та [12], запропоновано нейронну мережу, навчену виключно на (FIR – far infrared) теплових зображеннях, яка показала кращі результати, ніж дві попередні архітектури.

Згадані до цього моменту архітектури нейронних мереж складаються в основному зі згорткових шарів. Втім, глибока модель CNN (створена шляхом додавання більшій кількості згорткових шарів) призводить до суттєвого збільшення кількості параметрів та ускладнює процедуру навчання. Збільшення глибини нейронних мереж підвищує також і ризик перенавчання [13]. Для боротьби із цим існує 3 основні підходи (рис. 1):

- пропуск шарів (рис.1, а): як глобальне залишкове з'єднання (residual connection), яке пов'язує вхідні дані і вихідний шар, як, наприклад, в [7, 11, 12], так і локальне залишкове з'єднання з ResNet [14], зменшують труднощі в навчанні;
- структури з щільними зв'язками (dense connection structures) пов'язують всі шари в мережі і об'єднують карти об'єктів всіх попередніх рівнів (рис.1, б), щоб зменшити зникаючі градієнти і повторно використовувати елементи з неглибоких шарів (shallow layers);
- рекурсивні нейронні (рис.1, в) мережі багаторазово застосовують одні й ті ж згорткові шари для керування параметрами з досягненням великих рецептивних полів [13].

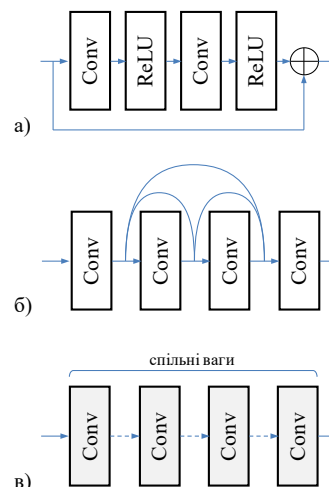


Рис. 1. Основні підходи до зменшення глибини нейронної мережі: а) пропуск шарів; б) структури з щільними зв'язками; в) рекурсивні нейронні мережі (сірим позначено рекурсивні згорткові шари)

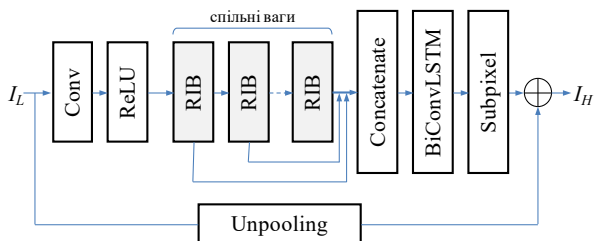
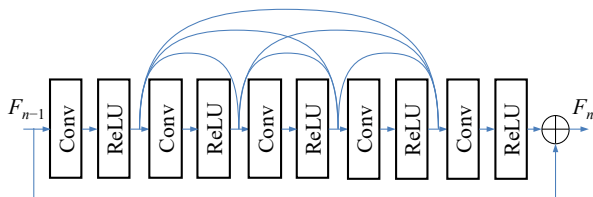


Рис. 2. Архітектура мережі BCLSR

Рис. 3. Блок рекурсивного виводу (RIB) мережі BCLSR. F_{n-1} та F_n позначають вихід попередньої ітерації та вхід наступної ітерації відповідно

У задачі надроздільної здатності рекурентні нейронні мережі зазвичай використовуються для захоплення довгострокової залежності сусідніх кадрів у відео [15]. Однак, в роботі [13] запропоновано нейронну мережу BCLSR (рис.2), що складається з двох структур – набору блоків рекурсивного виводу (RIB) та двонапрявленого згорткового LSTM шару (BiConvLSTM), що оброблюють послідовність, утворену з'єднаними виходами блоків рекурсивного виводу. Тобто результат кожної з ітерацій рекурсивного блоку є елементом послідовності для рекурентного LSTM шару.

Автори BCLSR стверджують, що запропонована ними мережа перевершує аналоги за рахунок таких аспектів:

- 1) Через збільшення поля сприйняття з кожним рекурсивним застосуванням RIB шару більше контекстної інформації може використовуватись для відновлення якісного зображення без додаткових параметрів.
- 2) Завдяки повторному використанню низькорівневих ознак потік інформації через мережу може бути підсилений без «вибуху» або «зникнення» градієнту.
- 3) Двонаправлений згортковий LSTM шар вибірково витягує доповнюючу інформацію з усіх рекурсій та запобігає надходженню зайвої інформації до шару відновлення.

Таким чином, нейромережі на базі LSTM шарів на даний момент є кращими за якістю відновлення високої роздільної здатності, складністю навчання та кількістю параметрів. В літературних джерелах не виявлено жодної згадки про застосування LSTM шарів до обробки інфрачервоних зображень, що і спонукало присвятити дане дослідження в тому числі використанню архітектури BCLSR для підвищення роздільної здатності інфрачервоних зображень. Окрім цього

в даній роботі було запропоновано власну архітектуру мережі без застосування рекурентних шарів та порівняно якість її роботи з мережею BCLSR.

Таким чином розробка власної архітектури нейронної мережі та порівняння її з однією з найкращих мереж для підвищення роздільної здатності зображень і складає основну мету даного дослідження.

II. BiConvLSTM МЕРЕЖА ДЛЯ ПІДВИЩЕННЯ РОЗДІЛЬНОЇ ЗДАТНОСТІ ІНФРАЧЕРВОНИХ ЗОБРАЖЕНЬ

Завдяки описаним в минулому розділі перевагам BCLSR мережі, її було обрано за основу запропонованого методу підвищення роздільної здатності інфрачервоних зображень з метою перевірки її архітектурних особливостей на FIR-зображеннях та можливості удосконалення або адаптації до цього класу зображень. Для повноти викладення розглянемо основні елементи архітектури BCLSR більш детально.

A. Блок рекурсивного виводу (RIB)

Як зазначалося раніше, одним з основних елементів мережі є блок рекурсивного виводу (RIB – Recursive Inference Block) – розроблений під впливом архітектури іншої мережі ResNet [14], що застосовується для класифікації зображень. Структура даного блоку показана на рис. 3.

В цілому, використання шарів та блоків із мереж, які застосовуються для класифікації зображень у мережах, що призначені для підвищення роздільної здатності, є виправданим, оскільки припускається, що в обох випадках використовуються одні й ті самі ознаки вхідних зображень. Однак, результат їх виділення є різним: відгуки згорткових шарів під час класифікації збуджують нейрони, що відповідають різним класам, тоді як в задачах підвищення роздільної здатності вони є основою для додавання деталей до зображення.

B. Двонаправлені LSTM шари (BiConvLSTM)

Окремо варто розглянути роботу двонаправлених LSTM-шарів. Оригінальні LSTM-шари запропоновано в 1997 в [16] з метою збереження інформації про довготривалі залежності між послідовними даними. В BCLSR використано згортковий LSTM комірки [17], що мають наступні рівняння виходів та станів:

$$\begin{aligned} i_t &= \sigma(x_t * w_{xi} + h_{t-1} * w_{hi} + b_i), \\ f_t &= \sigma(x_t * w_{xf} + h_{t-1} * w_{hf} + b_f), \\ o_t &= \sigma(x_t * w_{xo} + h_{t-1} * w_{ho} + b_o), \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tanh(x_t * w_{xc} + h_{t-1} * w_{hc} + b_c), \\ h_t &= o_t \cdot \tanh(c_t), \end{aligned} \quad (1)$$

де $\sigma(\cdot)$ позначає сигмоїдну функцію; “*” позначає згортку; “.” позначає по-елементне множення; x_t – поточний (t -й) елемент оброблюваної послідовності; w_{mn} – ваги фільтру для згортки з сигналу m для розрахунку вентилі n ($m = \{x, h\}$, $n = \{i, f, o, c\}$); b_n – зміщення для вентилі n ($n = \{i, f, o, c\}$); f_t – вентиль «забування», що визначає яку частину інформації з попередньої ітерації відкинути при обробці поточного елемента послідовності; i_t – вхідний вентиль;

o_t – вихідний вентиль; c_t – стан комірки, що відповідає за збереження довготривалих залежностей при обробці послідовностей; h_t – вихід комірки.

У BCLSR використовується два BiConvLSTM шари, кожен з яких містить по дві згорткові LSTM комірки, через які послідовності сигналів пропускаються в прямому та зворотному напрямках [13].

B. Depth-to-Space шар (Subpixel)

Для узгодження за розмірами перед додаванням до інтерпольованого вхідного зображення вихід попереднього шару спочатку оброблюється згорткою, щоб отримати тензор розміром $[h_{LR}, w_{LR}, c \times f_{upscale}^2]$, а потім об'єднується в один тензор, що має розмір $[h_{LR} \times f_{upscale}, w_{LR} \times f_{upscale}, 1]$ ($f_{upscale}$ – коефіцієнт масштабування) як показано на рис. 4 [18].

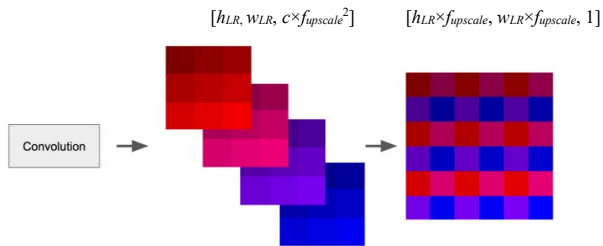


Рис. 4. Приклад операції Depth-to-Space

III. ЗАПРОПОНОВАНА МЕРЕЖА

BCLSR за даними, наведеними в оригінальній статті [13], дійсно перевершує попередні відомі нейронні мережі для підвищення роздільної здатності, однак вона має деякі недоліки. По-перше, застосування рекурсивного блоку хоч і зменшує кількість параметрів для навчання та дозволяє регулювати глибину нейронної мережі, втім, таке рішення може бути менш гнучким, ніж застосування нейронної мережі із такою самою глибиною, але різними параметрами шарів. По-друге, схожий недолік має й рекурсивний LSTM шар – не дивлячись на те, що при розрахунку кожної ітерації поточне вхідне значення та значення минулої ітерації «комбінуються» в різних пропорціях, протягом обробки вхідної послідовності значення ваг не змінюються зі збільшенням глибини, що також впливає на гнучкість нейронної мережі.

Для подолання вказаних недоліків було розроблено власну мережу, архітектура якої наведена на рис. 5.

Суть її роботи дуже схожа на BCLSR – вхідні дані послідовно оброблюються однаковими за структурою блоками шарів, після чого результати їх роботи комбінуються задля отримання кінцевого зображення з великою роздільною здатністю. Відмінності запропонованої архітектури:

- 1) відмова від LSTM-шарів із їхньою заміною на «гілки» з кількох однакових блоків;
- 2) застосування RIB з BCLSR як структурного блоку, але не рекурсивно.

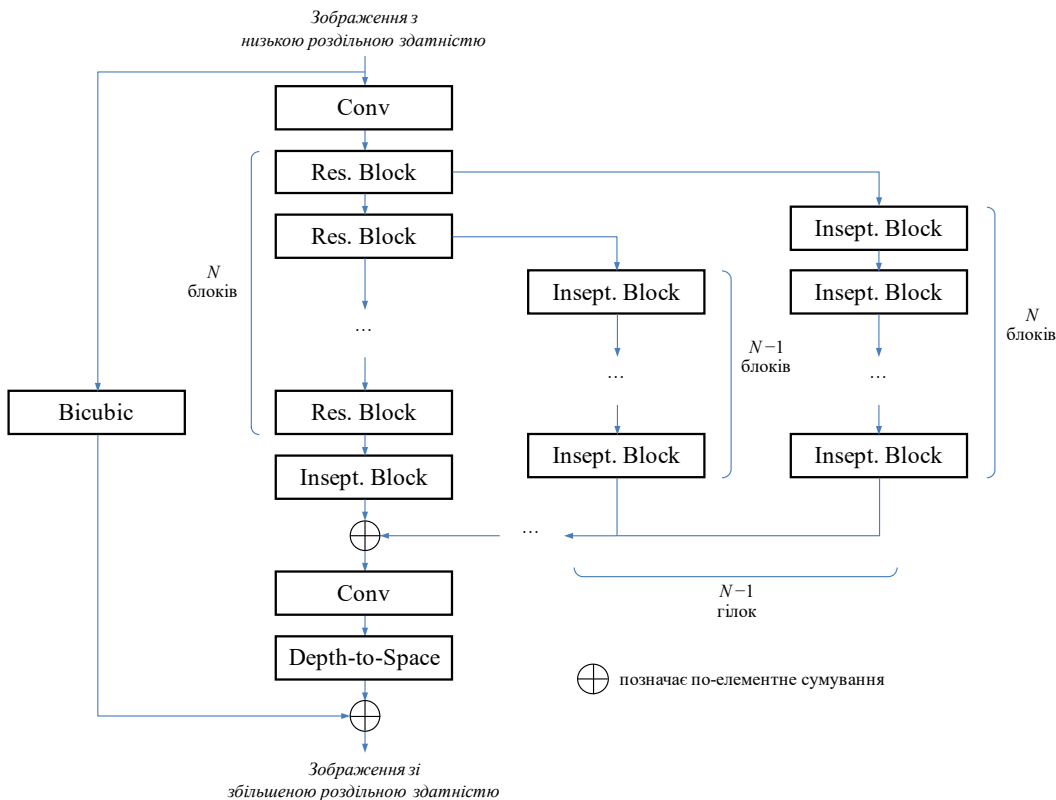


Рис. 5. Архітектура запропонованої мережі

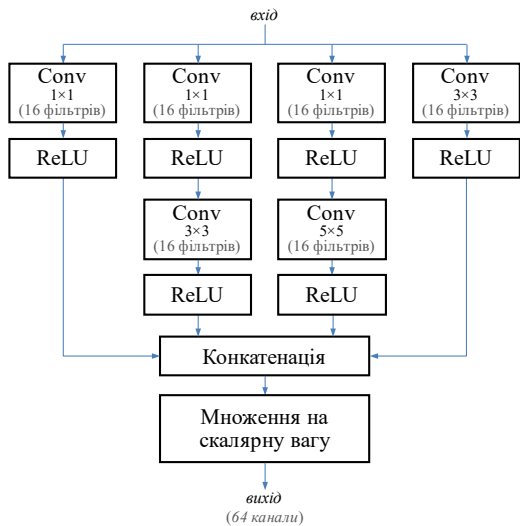


Рис. 6. Запропонований Inception-блок

Архітектура має деревоподібну (ієрархічну) структуру, а вихід кожного із послідовних Residual блоків (структура даного блоку аналогічна RIB блоку, наведеному на рис. 3, на рис. 5 позначений як Res.block) утворює гілку, що складається із послідовно з'єднаних Incept-блоків (на рис. 5 позначений як Incept.block), структура яких розглядається нижче. Виходи усіх гілок додаються і до отриманої суми застосовується перетворення Depth-to-Space. Наступним і останнім кроком є додавання результату останньої операції до вхідного зображення, розширеного за допомогою бікубічної інтерполяцією.

Для подальшої обробки ознак, вилучених після кожного Residual-блоку, як вже зазначалося раніше, створюється окрема «гілка». Кожна така «гілка» складається із послідовно з'єднаних однакових за структурою блоків (рис. 6), створених на основі однієї з найкращих нейронних мереж для розпізнавання зображень Inception [19], та, відповідно, позначених як Incept.block.

Після кожного зі згорткових шарів в Incept блоках додано dropout шари. За умови невеликої кількості прикладів у навчальних вибірках з FIR зображеннями, додавання dropout шару має запобігти перенавантаженню мережі.

IV. ПРАКТИЧНІ РЕЗУЛЬТАТИ

A. Особливості навчання

Моделювання запропонованої нейронної мережі виконано засобами програмного пакету Tensorflow в середовищі Google Colaboratory. Досліджувався варіант запропонованої мережі із 8 гілками ($N=8$ на рис.3), яка збільшує роздільну здатність в 2 рази. В усіх подальших порівняннях застосовується BCLSR із 8 застосуваннями рекурсивного блоку, яку також реалізовано та навчено засобами Tensorflow для цього дослідження. Для наближення умов порівняння із запропонованим рішенням, згорткові LSTM-шари також використовують dropout. В RIB та Residual блоках встановлено однакову кількість каналів в згорткових шарах – 64.

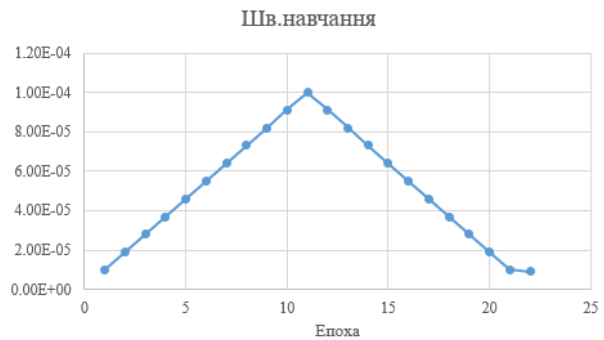


Рис. 7. Графік зміни швидкості навчання

Для навчання та тестування запропонованої мережі та BCLSR використано FIR-зображення, зняті вночі, з навчальної вибірки CVC-14 [20]. Дана вибірка не є великою, тому задля її розширення з кожного зображення випадковим чином було вирізано 20 фрагментів розміром 100×100 , до значень яскравості яких було додано випадкові числа та застосовано випадкову зміну контрастності (функції `random_brightness` та `random_contrast` пакету Tensorflow), до деяких зображень також застосовано віддзеркалення. Отримана навчальна вибірка поділяється на підвибірки (mini-batch) розміром по 8 зображень.

Для оптимізації моделі застосовано оптимізатор Adam із $\beta_1 = 0,9$, $\beta_2 = 0,999$, $\epsilon = 10^{-7}$. Функція втрат $L = \frac{1}{mn} \|f - f_{GT}\|^2$ – середньоквадратична похибка між зображенням, що генерує мережа f та еталонним зображенням f_{GT} з високою роздільною здатністю $m \times n$ пікселів.

У роботі [21] показано доцільність циклічної зміни швидкості навчання (одного з параметрів оптимізатора), подібний підхід застосовано й під час навчання запропонованої мережі – швидкість навчання зростає з 10^{-5} до 10^{-4} за 10 епох, потім так само спадає, додатково після цього одну епоху вона дорівнює $9,1 \cdot 10^{-6}$. Графік зміни швидкості навчання наведений на рис. 7.

Б. Результати тестування мереж

Після одного циклу навчання отримано нейронну мережу, що відновлює зображення із високою роздільною здатністю із середнім PSNR на тестовому наборі (FIR-зображення 640×480 пікселів) в 42,2749 дБ, в той час як для бікубічної інтерполяції цей показник дорівнює 40,5721 дБ. Таким чином, маємо приріст PSNR в 1,7028 дБ.

Навчена аналогічним чином мережа BCLSR на тому ж тестовому наборі нічних FIR-знімків CVC-14 забезпечує середній PSNR в 42,2119 дБ. Тобто дає приріст PSNR в 1,6398 дБ порівняно з бікубічною інтерполяцією.

Таким чином, якість зображень відновлених запропонованою мережею є дещо вищою в порівнянні з BCLSR. Візуальні результати відновлення для одного з зображень тестової вибірки CVC-14 показано на рис. 8.



Рис. 8. Візуальне порівняння оригінального зображення, його ути-снуготого варіанту (для покращення сприйняття збільшеного мето-дом найближчого сусіда), відновленого BCLSR, відновленого запропонованою нейронною мережею («Proposed») та відновленого бікубічною інтерполяцією

Запропонована мережа є помітно більшою за BCLSR. Так, зокрема, вона має 2697512 параметрів, тоді як BCLSR – 591940 параметрів. Не зважаючи на це, запропоноване рішення є більш швидкодіючим: час обробки одного зображення на відеокарті NVIDIA Tesla P100-PCIE-16GB для нього становить 0,126 с, хоча для BCLSR це значення дорівнює 0,24 с. Причиною цього є відносна складність операцій в LSTM шарах мережі BCLSR – як вже зазначалося, згортковий LSTM шар включає кілька згорток, по-елементне множення, підсумовування та, що важливо, обчислення сигмодних функцій та гіперболічних тангенсів (1).

ВИСНОВКИ

В статті вперше досліджено застосування нейронної мережі BCLSR для збільшення роздільної здатності FIR-зображень, а також запропоновано власну нейронну мережу, що є модифікацією BCLSR. Запропонована мережа відрізняється від BCLSR відсутністю рекурсивних і рекурентних елементів, а також використанням Insertion-блоків, що в результаті дозволило збільшити швидкість роботи майже в два рази та навіть дещо покращити якість відновлених зображень на 0,063 дБ. Процедура навчання при цьому виконувалась в один цикл із одноразовим зростанням та спаданням параметру швидкості навчання на нічних FIR-зображеннях навчальної вибірки SVC-14 із однаковими параметрами як для запропонованої, так і для BCLSR мережі.

Незважаючи на те, що запропоноване рішення має більшу кількість параметрів (2,7 М) порівняно з BCLSR (0,6 М), обидві мережі за розміром можуть

вважатися малими і використовуватись як у додатках для звичайних персональних комп'ютерів, так навіть і у вбудованих рішеннях.

Можливими напрямками подальших досліджень є вдосконалення архітектури запропонованої мережі шляхом додаванням нових типів шарів та (або) зміни гіперпараметрів поточних шарів. Покращення якості роботи також може бути досягнуте переглядом процедури навчання із застосуванням інших функцій втрат та стратегій зміни параметру швидкості навчання.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] E. Mandanici, L. Tavasci, F. Corsini and e. al., "A multi-image super-resolution algorithm applied to thermal imagery," *Appl. Geomat.*, vol. 11, p. 215–228, 2019. DOI: [10.1007/s12518-019-00253-y](https://doi.org/10.1007/s12518-019-00253-y).
- [2] "ULTRAMAX – The Ultimate Resolution," [Online]. Available: https://www.flirmedia.com/MMC/THG/Brochures/IND_023/IND_023_EN.pdf. [Accessed 05 02 2021].
- [3] "MicroScan significantly increases the geometrical resolution capability of the ImageIR® camera series," [Online]. Available: <https://www.ndt.net/search/docs.php3?id=21250&content=1>. [Accessed 03 01 2021].
- [4] "FLIR MSX Multi-Spectral Dynamic Imaging," [Online]. Available: <https://www.tequipment.net/FlirMSXMulti-SpectralDynamicImaging.html>. [Accessed 03 01 2021].
- [5] S. Liu, Y. Yang, Q. Li, H. Feng, Z. Xu and Y. Chen, "Infrared Image Super Resolution Using GAN With Infrared Image Prior," in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, Wuxi, China, 2019. DOI: [10.1109/SIPROCESS.2019.8868566](https://doi.org/10.1109/SIPROCESS.2019.8868566).
- [6] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence (TPAMI)*, vol. 38, no. 2, pp. 295-307, 2014. DOI: [10.1109/TPAMI.2015.2439281](https://doi.org/10.1109/TPAMI.2015.2439281).
- [7] J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016. DOI: [10.1109/CVPR.2016.182](https://doi.org/10.1109/CVPR.2016.182).
- [8] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken and R. Bishop, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016. DOI: [10.1109/CVPR.2016.207](https://doi.org/10.1109/CVPR.2016.207).
- [9] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. DOI: [10.1109/CVPR.2017.19](https://doi.org/10.1109/CVPR.2017.19).
- [10] Y. Choi, N. Kim, S. Hwang and I. S. Kweon, "Thermal image enhancement using convolutional neural network," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, Korea (South), 2016. DOI: [10.1109/IROS.2016.7759059](https://doi.org/10.1109/IROS.2016.7759059).
- [11] R. E. Rivadeneira, P. L. Suárez, A. D. Sappa and B. X. Vintimilla, "Thermal Image SuperResolution Through Deep Convolutional Neural Network," *Image Analysis and Recognition. ICIAR 2019. Lecture Notes in Computer Science*, vol. 11663, 2019. DOI: [10.1007/978-3-030-27272-2_37](https://doi.org/10.1007/978-3-030-27272-2_37).
- [12] K. Lee, J. Lee, J. Lee, S. Hwang and S. Lee, "Brightness-based convolutional neural network for thermal image enhancement,"

- IEEE Access*, vol. 5, pp. 26867-26879, 2017. DOI: [10.1109/ACCESS.2017.2769687](https://doi.org/10.1109/ACCESS.2017.2769687).
- [13] Y. Chang and B. Luo, "Bidirectional Convolutional LSTM Neural Network for Remote Sensing Image Super-Resolution," *Remote Sens.*, vol. 11, no. 2333, 2019. DOI: [10.3390/rs11202333](https://doi.org/10.3390/rs11202333).
- [14] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [15] M. S. M. Sajjadi, R. Vemulapalli and M. Brown, "Frame-Recurrent Video Super-Resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018. DOI: [10.1109/CVPR.2018.00693](https://doi.org/10.1109/CVPR.2018.00693).
- [16] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. (5-6), p. 602-610, 2005. DOI: [10.1016/j.neunet.2005.06.042](https://doi.org/10.1016/j.neunet.2005.06.042).
- [17] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong and W.-c. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting," in *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015.
- [18] Z. Wojna, V. Ferrari, S. Guadarrama, N. Silberman, L.-C. Chen, A. Fathi and J. Uijlings, "The Devil is in the Decoder: Classification, Regression and GANs," *International Journal of Computer Vision (IJCV)*, vol. 127, p. 1694-1706, 2019. DOI: [10.1007/s11263-019-01170-8](https://doi.org/10.1007/s11263-019-01170-8).
- [19] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," in *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [20] "CVC-14: Visible-FIR Day-Night Pedestrian Sequence Dataset," [Online]. Available: <http://adas.cvc.uab.es/elektra/enigma-portfolio/cvc-14-visible-fir-day-night-pedestrian-sequence-dataset/>. [Accessed 04 02 2021].
- [21] L. N. Smith, "A disciplined approach to neural network hyperparameters: Part 1 - learning rate, batch size, momentum, and weight decay," *CoRR*, arXiv: [1803.09820](https://arxiv.org/abs/1803.09820), 2018.

Надійшла до редакції 05 березня 2021 року

UDC 004.932.4

Hierarchical Convolutional Neural Network for Infrared Image Super-Resolution

M. O. Yaroshenko^f, ORCID [0000-0002-3092-3856](https://orcid.org/0000-0002-3092-3856)

A. Y. Varfolomeiev^s, PhD Assoc.Prof., ORCID [0000-0002-6990-7140](https://orcid.org/0000-0002-6990-7140)

P. O. Yaganov^s, PhD Assoc.Prof., ORCID [0000-0001-7358-9846](https://orcid.org/0000-0001-7358-9846)

Department of Design of Electronic Computing Equipment, Faculty of Electronics
National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute" ROR [00syn5v21](https://ror.org/00syn5v21)
Kyiv, Ukraine

Abstract—Due to the high price of thermal imaging sensors, methods for high quality upscaling of infrared images, acquired from low-resolution inexpensive IR-cameras become in high demand. One of the very promising branches of such kinds of methods is base on super-resolution (SR) techniques that exploit convolutional neural networks (CNN), which are developed rapidly for the last decade. During the review of existing solutions, we found that most of the super-resolution neural networks are intended for the upscaling of images in the visible spectrum band. Among them, the BCLSR network has proven to be one of the best solutions that ensure a very high quality of image upscaling. Thus, we selected this network for further investigation in the current paper. Namely, in this research, we trained and tested the BCLSR network for upscaling of far-infrared (FIR) images for the first time. Moreover, inspired by the BCLSR architecture, we proposed our own neural network, which differs from the BCLSR by the absence of recursive and recurrent layers that were replaced by series-connected Residual- and parallel-connected Inception-like blocks correspondingly. During the tests, we found that the suggested modifications permit to increase the network inference speed almost twice and even improve the quality of upscaling by 0,063 dB compared to the basic BCLSR implementation. Networks were trained and tested using the CVC-14 dataset that contains FIR images acquired at the night. We used data augmentation with random dividing dataset images onto 100×100 pixel patches and with subsequent application random brightness, contrast, and mirroring to the obtained patches. The training procedure was performed in a single cycle with single increase and decrease of the learning rate and used the same parameters for the proposed and the BCLSR networks. We employed the Adam optimizer for the training of both networks. Nevertheless, the proposed model has more parameters (2,7 M) compared to the BCLSR (0,6 M), both of the networks can be considered as the small ones, and thus can be used in applications for conventional personal computers, as well as in embedded solutions. The direction of the further research can be focused on the improvements of the proposed network architecture by introducing new types of layers as well as on the modifying of hyperparameters of the used layers. The quality of the upscaling can be increased also by using other loss functions and by the change of learning rate-varying strategies.

Keywords — *super-resolution; convolutional neural networks; recurrent networks; far infrared images*

