

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)Кафедра Інформатики
(повна назва)Рівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«____» _____ 2022 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУстудентові Кухарчуку Василю Андрійовичу
(прізвище, ім'я, по батькові)1. Тема роботи Дослідження нейромережного підходу для покращення якості зображення

затверджена наказом по університету від 9 листопада 2022 року №1469Ст

2. Термін подання студентом роботи до екзаменаційної комісії 25 листопада 2022 р.

3. Вихідні дані до роботи науково-методична та науково-технічна література, матеріали конференцій, дані інтернет-мережі.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Аналіз сучасного стану розвитку технологій покращення якості зображень в світі.

2. Дослідження принципів покращення якості зображень із використанням нейромереж.

3. Дослідження ефективності нейромережних методів покращення якості зображення.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п. 5 включається до завдання за рішенням випускової кафедри) Актуальність проблеми покращення зображень, постановка задачі, сучасні моделі суперроздільності, метрики оцінювання якості, порівняння ефективності моделей, напрямки майбутніх досліджень.

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Консультант з дотримання діючих стандартів та норм	доц. Творошенко І. С.		

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	09.11.2022	
2	Аналіз завдання, підбір літератури	09.11.22-11.11.22	
3	Аналіз літератури з досліджуваної проблеми	11.11.22-13.11.22	
4	Аналіз проблеми	13.11.22-14.11.22	
5	Дослідження методів	14.11.22-19.11.22	
6	Аналіз методів	19.11.22-23.11.22	
7	Оформлення пояснювальної записки	23.11.22-29.11.22	
8	Перевірка на плагіат	30.11.22	
9	Рецензування	02.12.21	
10	Підготовка презентації та доповіді	03.12.22-04.12.22	
11	Занесення роботи в електронний архів	04.12.22	
12	Попередній захист кваліфікаційної роботи	05.12.22	

Дата видачі завдання 9 листопада 2022 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Руденко Д.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 79 с., 2 табл., 30 рис., 79 джерел.

КОМП'ЮТЕРНИЙ ЗІР, ОБРОБКА ЗОБРАЖЕНЬ, НЕЙРОННІ МЕРЕЖІ, ГЛИБОКЕ НАВЧАННЯ, ПОКРАЩЕННЯ ЯКОСТИ ЗОБРАЖЕННЯ, СУПЕРРОЗДІЛЬНІСТЬ, ПІДВИЩЕННЯ ДИСКРЕТИЗАЦІЇ.

Об'єктом дослідження є нейромережні методи покращення якості зображень.

Метою дослідження є всебічний огляд та аналіз нейромережного підходу до покращення якості зображень.

Проведено дослідження методів покращення якості зображення із застосуванням глибоких нейронних мереж, їхні складові, підходи до проєктування та оцінювання результатів, напрямки застосування в реальному світі, проаналізовано показники продуктивності сучасних найдосконаліших моделей суперроздільності зображення. Визначено напрямки майбутніх досліджень в галузі покращення якості зображень.

Результат роботи апробовано у вигляді тез доповідей під час міжнародної наукової та практичної конференції «MODERN RESEARCH IN WORLD SCIENCE».

COMPUTER VISION, IMAGE PROCESSING, NEURAL NETWORKS, DEEP LEARNING, IMAGE QUALITY ENHANCEMENT, SUPER RESOLUTION, UPSAMPLING.

The object of the research is neural methods for image quality enhancement.

The aim of the research is the comprehensive review and analysis of the image quality enhancement based on the neural network approach.

Methods of image quality enhancement using deep neural networks, their components, approaches to designing and evaluating results, their application in the real world, and performance results of the most state-of-the-art super resolution image models were analyzed. The directions of future research in the field of image quality enhancement are determined.

The result of the work was tested in the form of the abstract of the report during the Scientific and Practical Conference «MODERN RESEARCH IN WORLD SCIENCE».

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ	8
1 Аналіз сучасного стану розвитку технологій покращення якості зображень в світі.....	10
1.1 Сучасний стан розвитку технологій покращення якості зображень у світі.....	10
1.1.1 Задача суперроздільності зображення та наявні методи її розв’язання.....	11
1.1.2 Основні галузі застосування технології СР.....	15
1.2 Аналіз літературних джерел щодо наявних нейромережних підходів для покращення якості зображення.....	22
1.3 Постановка задачі дослідження	26
2 Дослідження принципів покращення якості зображень із використанням нейромереж.....	27
2.1 Методи підвищення дискретизації зображення.....	27
2.1.1 Інтерполяція	27
2.1.2 Розгортка	29
2.1.3 Субпіксельний шар.....	30
2.1.4 Модуль метарозширення.....	31
2.2 Структури мереж СР	32
2.2.1 Лінійні мережі.....	33
2.2.2 Залишкові мережі	36
2.2.3 Рекурсивні мережі	39
2.2.4 Мережі зі щільними зв’язками.....	41
2.2.5 Мережі на основі уваги	43
2.2.6 Генеративні змагальні мережі.....	45
2.2.7 Розгалужені мережі	47
2.3 Функції втрат у моделях СР.....	50

	6
2.3.1 Піксельна втрата	51
2.3.2 Втрата вмісту	52
2.3.3 Текстурна втрата.....	52
2.3.4 Змагальна втрата	53
2.3.5 Априорна втрата	55
3 Дослідження ефективності нейромережних методів покращення якості зображення	57
3.1 Метрики оцінювання якості реконструкції.....	57
3.1.1 Пікове співвідношення сигнал/шум (PSNR).....	57
3.1.2 Структурна подібність (SSIM).....	58
3.1.3 Оцінка на основі усередненої думки (MOS).....	59
3.1.4 Перцептивна якість заснована на навчанні	60
3.1.5 Інші методи ОЯЗ.....	60
3.2 Набори даних для СР.....	61
3.3 Порівняння ефективності мереж СР.....	64
3.4 Сучасні виклики та напрямки дальших досліджень у сфері СР	65
3.4.1 Мережеві архітектури.....	66
3.4.2 Стратегії навчання	67
3.4.3 Метрики оцінювання якості.....	68
3.4.4 СР із довільним коефіцієнтом розширення.....	69
3.4.5 Взаємодія СР із завданнями КЗ високого рівня	69
3.4.6 СР в реальних сценаріях застосування	70
Висновки.....	71
Перелік джерел посилання	72

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ**

ШІ – штучний інтелект

МН – машинне навчання

ГН – глибоке навчання

ЗНМ – згорткова нейронна мережа

СКЗ – система комп'ютерного зору

РЗ – роздільна здатність

НР – низька роздільність

ВР – висока роздільність

СР – суперроздільність

ГЗМ – генеративні змагальні мережі

ОЯЗ – оцінювання якості зображення

ВСТУП

Технологія машинного навчання сприяє багатьом аспектам сучасного суспільства: від вебпошуку до фільтрації вмісту в соціальних мережах і рекомендацій на сайтах електронної комерції, і вона все більше присутня в споживчих товарах, таких як камери та смартфони. Системи машинного навчання використовуються для ідентифікації об'єктів на зображеннях, переведення мови в текст, аналізу новин, публікацій або продуктів з інтересами користувачів і вибору релевантних результатів пошуку. Ці застосунки все частіше використовують клас методів, які називають глибоким навчанням.

З моменту своєї появи у 2012 році глибоке навчання призвело до дуже хороших результатів у розв'язанні різноманітних задач, таких як візуальне розпізнавання, розпізнавання мови та обробка природньої мови. Серед різних типів глибоких нейронних мереж, згорткові нейронні мережі є одними з найбільш цікавих і широко вивчених. Завдяки стрімкому зростанню обсягу анотованих даних і значному вдосконаленню графічних процесорів, дослідження згорткових нейронних мереж розпочалося вельми швидко й за короткі терміни було досягнуто значних вислідів у різних завданнях.

ЗНМ – це тип моделі глибокого навчання для обробки даних, які мають структуру сітки (наприклад зображень), що натхненний організацією зорової кори тварин і призначений для автоматичного й адаптивного вивчення просторових ієрархій ознак із низько- до високорівневих. ЗНМ – це математична конструкція, яка зазвичай складається з трьох типів шарів (або будівельних блоків): шари згортки, шари об'єднання та повністю зв'язані шари. Перші два, шари згортки та об'єднання, виконують виділення ознак, тоді як третій, повністю пов'язаний рівень, приводить вилучені ознаки до кінцевого результату, наприклад класифікації. Згортковий рівень відіграє ключову роль у ЗНМ. Ядро згортки, що є оптимізованим екстрактором ознак, застосовується до кожної позиції зображення, що робить ЗНМ дуже

ефективними для обробки зображень, оскільки ознака може з'являтися в будь-якому місці зображення. Оскільки один шар передає свій вихід на наступний рівень, вилучені ознаки можуть ієрархічно та поступово ставати складнішими. Процес оптимізації таких параметрів, як ядра, називається навчанням, яке виконується таким чином, щоб мінімізувати різницю між виходами мережі та сподіваним результатом за допомогою певного алгоритму оптимізації.

Хоча ЗНМ досягли великого успіху в експериментальних оцінках, все ще є багато моментів, які заслуговують на дальше дослідження. По-перше, оскільки останні ЗНМ стають все глибшими, вони вимагають розлогіх наборів даних і великої обчислювальної потужності для навчання. Ручне збирання міченого набору даних вимагає величезних людських зусиль. Тим часом, щоб прискорити процедуру навчання, все ще варто розробити ефективні та масштабовані алгоритми паралельного навчання. Іще однією проблемою застосування ЗНМ є складність підбору параметрів мережі, таких як швидкість навчання, розмір ядра згортки, кількість шарів тощо.

Так чи так, як було зазначено раніше, дана технологія є надзвичайно актуальною й перспективною і тому потребує нових досліджень в усіх напрямках, в яких вона може застосовуватись. Одним із цих напрямків, що останнім часом активно розвивається, є підвищення якості зображень.

1 АНАЛІЗ СУЧАСНОГО СТАНУ РОЗВИТКУ ТЕХНОЛОГІЙ ПОКРАЩЕННЯ ЯКОСТИ ЗОБРАЖЕНЬ В СВІТІ

1.1 Сучасний стан розвитку технологій покращення якості зображень у світі

Роздільна здатність зображення – це кількість пікселів на квадратний дюйм цифрового зображення. Суперроздільність зображення – це процес відновлення зображення високої роздільної здатності (ВР) із зображення низької роздільної здатності (НР) [1]. Це також можна назвати масштабуванням, підвищенням дискретизації або збільшенням [2]. Метою суперроздільності зображення є одержання високої щільності пікселів і чітких деталей із зображень із ВР, які неможливо побачити голим оком.

Оскільки цифрові зображення з ВР бажані в більшості випадків, в останні десятиліття технології одержання зображень швидко розвивалися, і РЗ вийшла на новий рівень. Тому постає питання: чи все ще потрібні методи покращення якості зображення?

Річ у тім, що хоч на сьогодні дисплеї високої чіткості й записувальні пристрої досягли дуже високих показників, проблема недостатньої якості зображення все ще лишається в багатьох випадках практичного використання. Наприклад, щоб гарантувати тривалу стабільну роботу записуючих пристроїв, а також відповідну частоту кадрів, цифрові продукти відеоспостереження, як правило, певною мірою жертвують роздільною здатністю. Також, в деяких випадках доводиться знімати на далекій віддалі (через фізичну неможливість зняти ближче чи з міркувань безпеки), через що потрібні елементи на зображенні виходять надто малими й розмитими. Що стосується медичної візуалізації, то мають вплив певні фізичні закони, які визначають значення шуму та чутливість процесу візуалізації.

Тому підвищення РЗ все ще є необхідним, особливо в таких сферах, як відеоспостереження, медична діагностика та програми дистанційного

зондування. Враховуючи високу вартість і технічні обмеження підвищення РЗ за допомогою апаратних засобів, методи обробки сигналу, відомі як суперроздільність, стали потенційним способом одержання високої РЗ. На рисунку 1.1 продемонстровано приклад роботи методу СР.

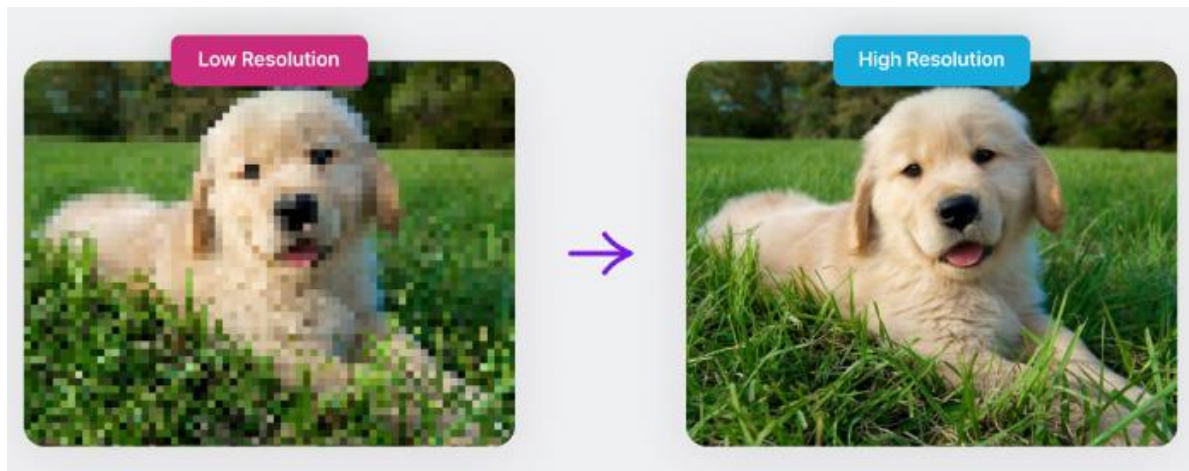


Рисунок 1.1 – Зображення низької роздільності й покращене зображення з високою роздільністю

1.1.1 Задача суперроздільності зображення та наявні методи її розв'язання

СР охоплює набір алгоритмів і методів, завдання яких полягає в тому, щоб збільшити ширину та висоту вхідного зображення з мінімальним (в ідеалі нульовим) погіршенням якості. Проте це набагато легше сказати, ніж зробити. Припустімо, що зображення має роздільну здатність (тобто розміри зображення) 64×64 пікселів і СР 256×256 пікселів. У цьому випадку процес називається 4-кратним підвищенням дискретизації, оскільки просторові розміри (висота та ширина зображення) збільшуються в чотири рази.

Зображення з низькою роздільною здатністю може бути змодельовано із зображення високої роздільної здатності, використовуючи функцію деградації δ і компонент шуму η таким чином:

$$I_{LR} = \delta(I_{HR}, \eta). \quad (1.1)$$

Відповідно до кількості вхідних зображень HR, SR можна поділити на SR на основі одного зображення (CPOЗ) і SR на основі кількох зображень (СРКЗ). Якщо можна одержати кілька зображень однієї сцени з субпіксельним зсувом, додаткову інформацію між ними можна використати для реконструкції зображення або послідовності зображень з вищою роздільною здатністю. Однак кілька зображень ВР часто недоступні, а технологія СРКЗ вимагає більше часу та обчислювальних ресурсів, через що СРОЗ наразі набагато популярніший [3].

Основні алгоритми СРОЗ поділяються на три категорії: методи на основі інтерполяції, методи на основі реконструкції та методи на основі навчання. Одразу варто зазначити, що неможливо зробити так, щоб відновлене зображення мало такий самий вигляд, як і оригінал. Жодна розумна функція чи алгоритм не зможе замінити відсутню інформацію. Найкраще, що можна зробити, це наблизити та заповнити прогалини на основі сусідніх пікселів [4].

Кожен, хто коли-небудь відкривав невелике зображення в Photoshop або GIMP, а потім намагався змінити його розмір, знає, що вихідне зображення виглядає піксельним та розмитим.

Це тому, що Photoshop, GIMP, Image Magick, OpenCV (через функцію `cv2.resize`) тощо використовують класичні методи та алгоритми інтерполяції для збільшення роздільної здатності зображення. Зазвичай це інтерполяція найближчого сусіда, що просто збільшує розмір вихідного зображення, а потім заповнює прогалини, копіюючи значення найближчих сусідніх пікселів (рис. 1.2).

Звичайно, результат можна поліпшити, взявши середнє зважене значення сусідніх пікселів, а не просто їхню копію. Це робиться за допомогою білінійної [5] або бікубічної [6] інтерполяції (рис. 1.3).

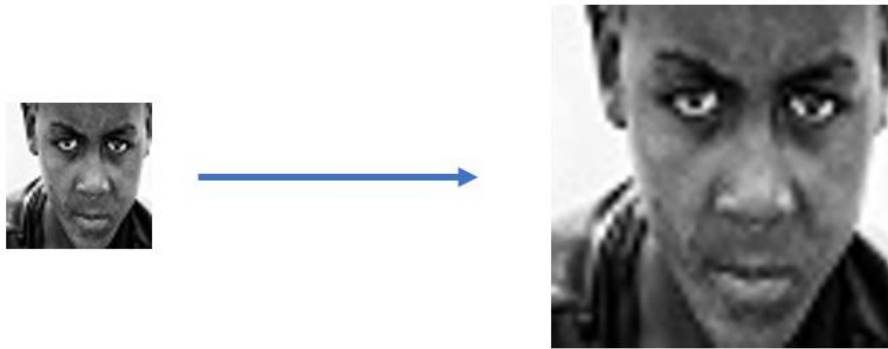


Рисунок 1.2 – Приклад розширення зображення методом інтерполяції найближчого сусіда



Рисунок 1.3 – Приклади розширених зображень методами білінійної та бікубічної інтерполяції

Ці методи дуже швидкі та прості, але високочастотні деталі зображення не відновлюються і якщо ви надто сильно збільшите просторові розміри, вихідне зображення буде піксельним, матиме артефакти та загалом матиме «естетично неприємний» для людського ока вигляд.

Методи СР на основі реконструкції [7] часто використовують складні попередні знання, регуляризацію та розгортку, щоб обмежити можливий простір рішень для створення чітких деталей. Однак продуктивність багатьох методів, заснованих на реконструкції, різко погіршується, коли масштабний коефіцієнт збільшується, і робота цих методів зазвичай вимагає багато часу.

Методи СРОЗ, засновані на навчанні привертають увагу через їх швидке обчислення та видатну продуктивність [8]. У цих методах зазвичай використовуються алгоритми машинного навчання для аналізу статистичних

взаємозв'язків між зображенням НР та його відповідним аналогом ВР на основі наданих навчальних прикладів.

Серед алгоритмів СР, заснованих на навчанні, можна виділити випадкове поле Маркова, метод розрідженого кодування, випадковий ліс та інші, а в багатьох роботах для одержання кращих результатів поєднуються переваги цих методів з підходами, заснованими на реконструкції.

З розвитком технологій машинного навчання основні дослідження зараз націлені на новий підхід заснований на навчанні, яким є метод глибокого навчання. ГН продемонструвало кращу продуктивність, ніж традиційні моделі машинного навчання, а через його швидкий розвиток результати безперервно покращуються (рис. 1.4).

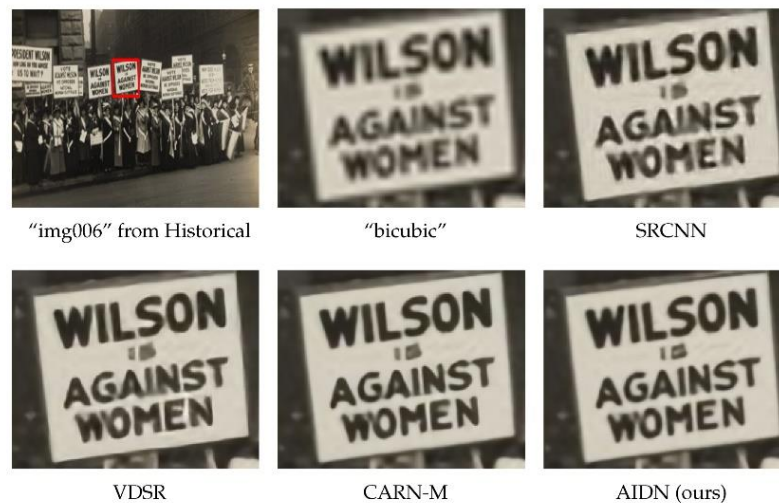


Рисунок 1.4 – Порівняння результатів СР зображення з навчального набору «Historical» методами бікубічної інтерполяції та кількох алгоритмів ГН

Алгоритми ГН переглядають тисячі й мільйони зразків і запам'ятовують закономірності, тож їм не доводиться приблизно визначати відсутню інформацію, вони її продукують на основі минулих даних. Також можна виділити ще дві переваги підходу на основі ЗНМ. По-перше, метод на основі ЗНМ виділяє значно більше ознак із вхідних даних порівняно з іншими згаданими методами, що забезпечує модель гнучкістю в оптимізації. По-друге, підхід на основі ЗНМ має механізм зворотного зв'язку, який дозволяє точніше

налаштовувати параметри. Кінцевою метою тонкого налаштування параметрів є одержання мінімальних втрат у прогнозі моделі [9].

У 2014 році було розроблено перший алгоритм СР зображення на основі згорткової нейронної мережі. Згорткова нейронна мережа для СР (SRCNN) [10], розроблена Донгом та іншими була піонерською розробкою в алгоритмах на основі ЗНМ. Використовуючи досить неглибоку нейронну мережу для СРОЗ, вони змогли досягти вельми високих показників у задачі 4-кратного збільшення для одного з навчальних наборів даних для СР зображень. Після цієї події галузь СР на основі ГНМ стала стрімко розвиватися, створено дуже багато моделей, технік та підходів і вже досягнуто досить непоганих результатів реконструкції.

1.1.2 Основні галузі застосування технології СР

Спостереження. Системи відеоспостереження широко використовуються в усьому світі для моніторингу безпеки та запису [11]. Відео з системи відеоспостереження часто використовується для допомоги в розкритті кримінальних справ. Однак на даний момент неможливо обладнати великогабаритні НР-пристрої, через що якість оптики та систем замкнутого телебачення (ССТV) є низькою і одержуване відео виходить нечітким. Окрім того, деякі зображення часто погіршуються за певних умов. Наприклад, знімання в умовах слабого освітлення чи туману (через недостатню кількість інформації, отриманої од об'єкта сцени) призводить до втрати контрасту та до розмитості деталей зображення. Тому для подолання низької якості зображення з відеокамер можна застосовувати техніки СР зображення.

Медична діагностика. Різні методи медичної візуалізації можуть надати як анатомічну, так і функціональну інформацію про структуру тіла людини [12]. Однак обмеження РЗ завжди знижують цінність медичних зображень у діагностиці. Технології СР використовуються з ключовими

методами медичної візуалізації, включаючи комп'ютерну томографію (КТ), магнітно-резонансну томографію (МРТ), позитронно-емісійну томографію (ПЕТ) та зображення судин очного дна сітківки, які почасти мають низьку РЗ, властивий шум і відсутність структурної інформації, через що викликають великі складнощі із тим, щоб поставити правильний діагноз. Причинами обмеження якості зображень у наведених методах є час сканування, рух тіла, міркування щодо комфорту пацієнтів, конфігурація апаратного забезпечення або обмеження дози випромінювання (наприклад, питання, як отримати 3D-моделі будови людини із зображеннями високої РЗ при зниженні рівня радіації все ще залишається проблемою), і методи ГН можуть бути дуже корисними для компенсації якості (рис. 1.5) [13].

Системи медичної візуалізації можуть працювати в пильно контрольованому середовищі, тому можна легко одержати безперервні та багаторакурсні зображення. Проте, порівняно з СР природніх зображень, медичним зображенням потрібна додаткова попередня інформація для конкретних випадків. Окрім того, є деякі складнощі з одержанням навчальних наборів даних, оскільки медична діагностика пов'язана з конфіденційністю і є проблема одержання відповідних клінічних зображень високої та низької РЗ.

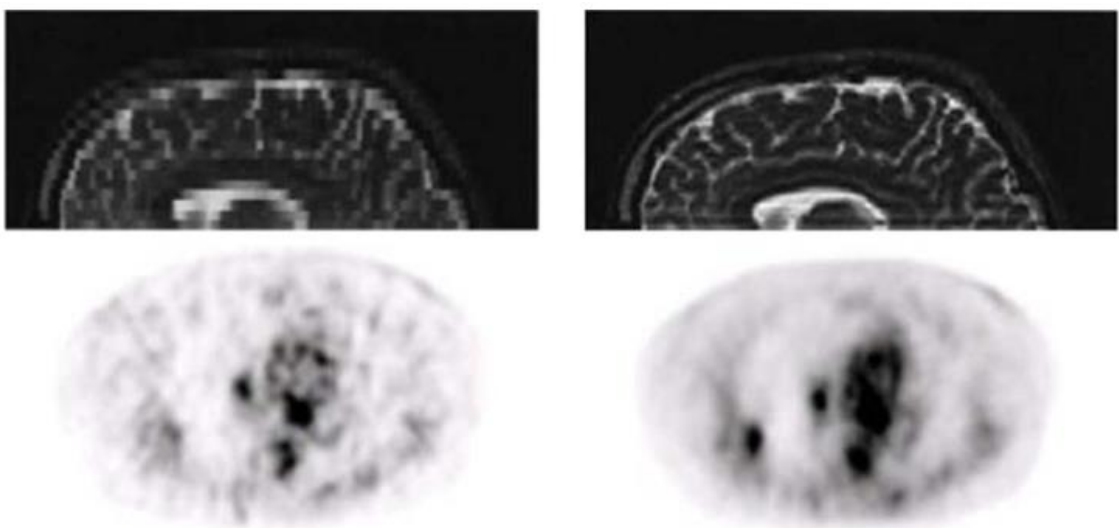


Рисунок 1.5 – Результати СР для даних МРТ головного мозку людини та респіраторного синхронізованого зображення ПЕТ

Дистанційне зондування Землі. Супутникові зображення та зображення аеророзвідки можна додатково збільшити без втрати якості та без додаткового обладнання, просто застосувавши для них СР (рис. 1.6) [14]. Наразі вже застосовуються супутники із вбудованою системою СР.

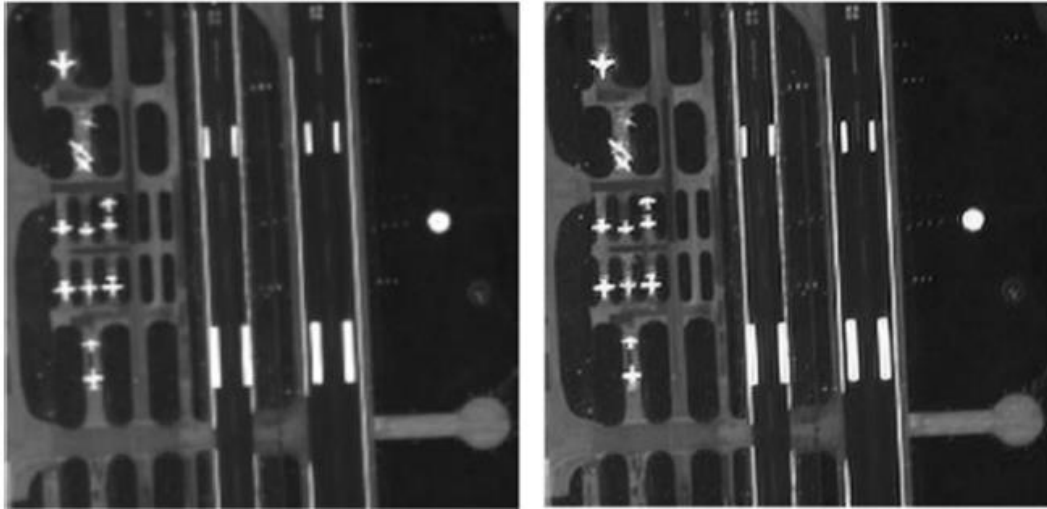


Рисунок 1.6 – Приклад застосування СР для світлини летища, зробленої із супутника

Астрономічні спостереження. Хоча ми можемо виявляти далекі зорі та планети, навіть за допомогою великих телескопів, таких як космічний телескоп Габбл, ми можемо робити лише розмиті зображення. Завдяки прогресу в техніці СР ми можемо одержати більш чіткі зображення небесних тіл [15]. Астрономічні системи зазвичай можуть збирати серії зображень для СР. Покращуючи роздільну здатність астрономічних зображень, СР може допомогти астрономам у дослідженні космічного простору. Конкретний приклад наведено на рисунку 1.7, на якому зображено приклад застосування СР до знімку зорь.

Перше в історії зображення чорної діри, яка розташована за мільйони світлових років од Землі, технічно не було зафіксовано. Зібрану інформація було оброблено й приведено до кінцевого зображення з високою РЗ, яке зображено на рисунку 1.8.

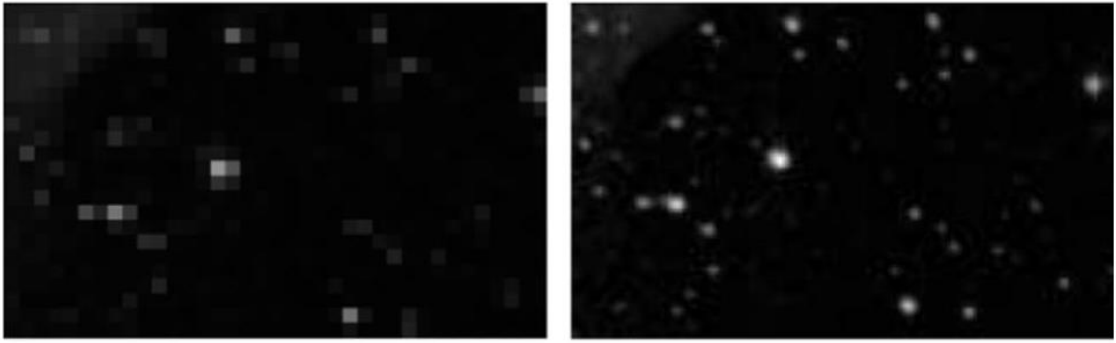


Рисунок 1.7 – Приклад СР для космічного знімку

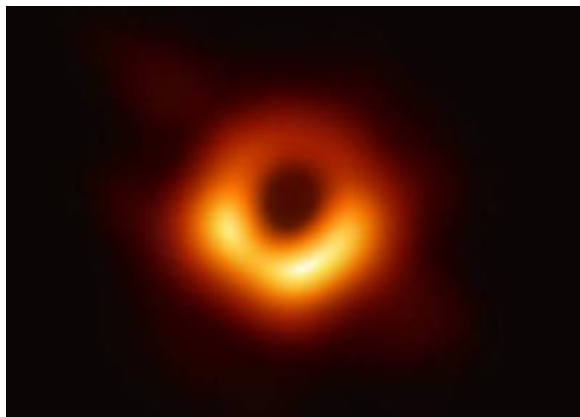


Рисунок 1.8 – Перше в історії фото чорної діри

Зараз також відправляються супутники для дослідження космосу, напр. програми дослідження Місяця та місія Mars Odyssey. На рисунку 1.9 показано приклад СР китайських знімків Місяця, де результат був реконструйований на основі трьох фото. СР може підвищити РЗ зображень і, таким чином, покращити чіткість малих об'єктів на поверхні небесних тіл. Окрім цього, використовуються теплові інфрачервоні та видимі зображення із різних спектральних областей системи тепловипромінювання, щоб створити покращене теплове інфрачервоне зображення поверхні Марса.

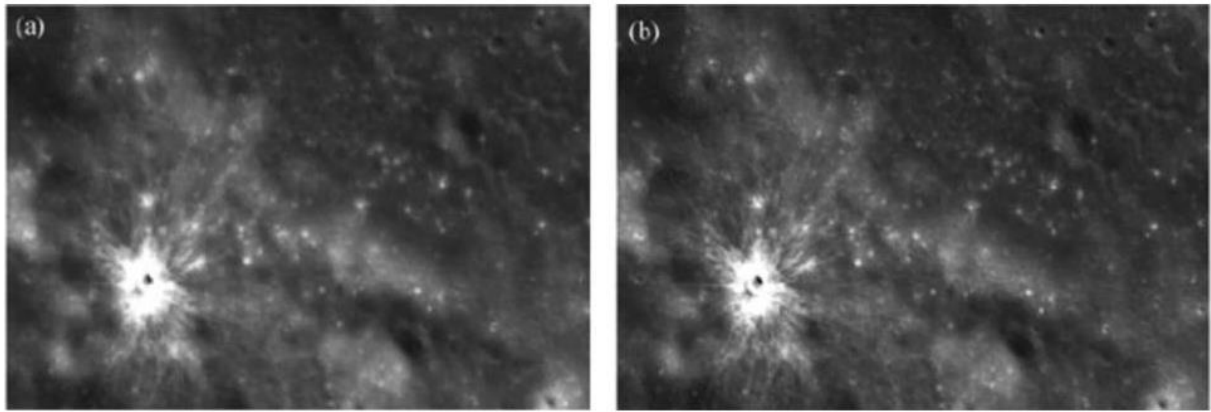


Рисунок 1.9 – Застосування SR для знімків Місяця

Біометрична ідентифікація. SR також важливий для біометричного розпізнавання, включно з покращенням РЗ для облич, відбитків пальців і зображень райдужної оболонки (рис. 1.10). РЗ біометричних зображень є ключовою в процесі виявлення й розпізнавання. Ґрунтуючись на надлишковості та подібності структурних особливостей біометричних зображень, СРОЗ на основі інформації із бази даних є ефективним способом підвищення РЗ. Використовуючи SR, деталі форм і текстури покращуються, тоді як глобальна структура ефективно зберігається, що може покращити здатність розпізнавання у відповідних програмах [16].

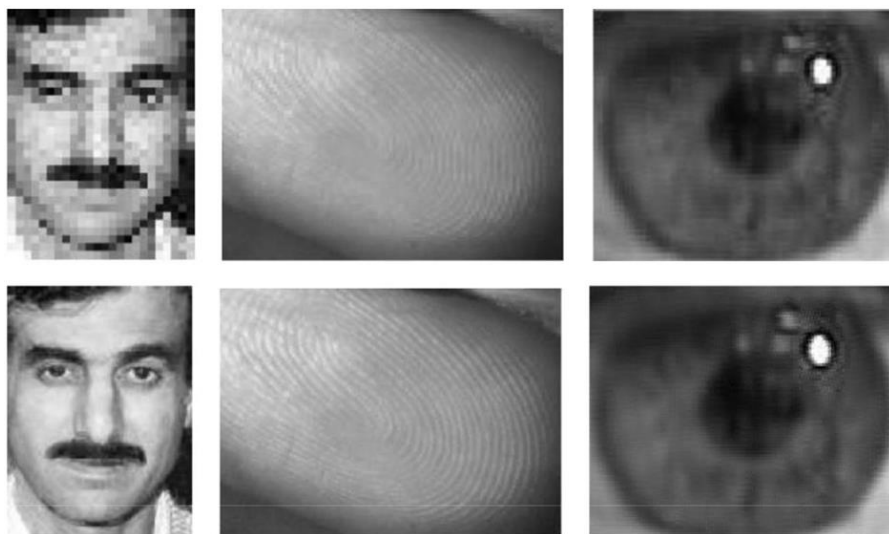


Рисунок 1.10 – Приклад застосування SR для зображень обличчя, відбитку пальців та райдужної оболонки ока

Стиснення даних, збереження пропускної здатності. Алгоритми стиснення зображень зазвичай призводять до втрати якості й за поміччю СР можна компенсувати втрачену інформацію. Що стосується збереження пропускної здатності, то можна передавати зображення з низькою РЗ, які займають небагато місця в пам'яті, і вже на боці клієнта переглядати їх з використанням системи СР.

Торгівля. Ще одним напрямом застосування технології СР є царина бізнесу. Наприклад, у роздрібній торгівлі, де автоматичне оброблення зображень для покращення їхньої якості призводить до збільшення привабливості для потенційних покупців. Те саме стосується сфери ріелторства, де якість світлин домівок грає важливу роль (рис. 1.11).

Для друку великих рекламних плакатів доконечні зображення високої чіткості й розміру, що подеколи повною мірою не можуть задовольнити навіть сучасні професійні камери [17].



Рисунок 1.11 – Приклад застосування СР для обробки рекламної світлини кімнати

Побутове використання. Застосування методів СР увійшло в наше повсякденне життя. Компанія Hitachi Ltd. винайшла метод перетворення телебачення стандартної чіткості (SDTV) на телебачення високої чіткості (HDTV) за допомогою технології СР для відео (рис. 1.12), що робить СР особливо цікавою та актуальною темою для досліджень.



Рисунок 1.12 – Приклад SR для покращення чіткості зображення в сучасних телевізорах

В усіх сучасних смартфонах використовується технологія SR для покращення якості світлин під час фотографування, що стало можливим завдяки збільшенню обчислювальної потужності, в тому числі графічних процесорів, цих пристроїв [18]. Окрім того, старі світлини, зроблені на старі пристрої зі значно меншими оптичними й обчислювальними можливостями, також можливо покращити за поміччю технології SR (рис. 1.13). Вже існує багато доступних сервісів SR. Навіть така передова американська компанія, як Adobe, вже створила власну реалізацію програмного забезпечення для збільшення роздільності зображень [19].

Таким робом, проаналізувавши технологію SR, царини її застосування та темпи розвитку, можна зробити висновок, що технологія є вельми актуальною та перспективною, на неї є значний попит, тому вона заслуговує на дальші дослідження та розвиток.



Рисунок 1.13 – Приклад застосування SR для звичайної світлин

1.2 Аналіз літературних джерел щодо наявних нейромережних підходів для покращення якості зображення

Покращення якості зображення за поміччю методів ГН – вкрай містка тема, яка є об'єктом досліджень вже протягом багатьох років. За цей час було написано безліч наукових робіт, що зачіпають різні особливості розроблення та проблеми, так чи так пов'язаних із цим напрямом обробки зображень. Вельми важко осягнути всі наявні праці, присвячені цьому питанню, проте в процесі роботи все ж було опрацьовано низку актуальних робіт, перелік яких наведено далі.

У праці [10] представлено найперший алгоритм для задачі SR одного зображення із використанням ЗНН. Наведений метод напряду вивчає наскрізне відображення між зображеннями з високою та низькою роздільністю. Це відображення представлено у вигляді ЗНН і приймає на вхід зображення з низькою роздільністю, а видає зображення з високою роздільністю. У статті досліджуються різні мережеві структури та настройки

параметрів для досягнення компромісу між продуктивністю та швидкістю. Окрім того, наводиться розширення мережі для роботи з трьома кольоровими каналами, що призводить до кращої загальної якості реконструкції.

Автори роботи [20] розвивають ідею використання ЗНН для СР зображення і представляють SRGAN – генеративну змагальну мережу для СР. Цей метод дозволяє вирішити проблему відновлення дрібних деталей при великих коефіцієнтах масштабування. Щоб досягти цього, вводиться функція втрати сприйняття, що складається зі змагальної та змістовної втрати. Функція змагальної втрати потрібна для мережі-дискримінатора, яка порівнює вихідне зображення із оригінальним зображенням високої роздільності і підштовхує до створення більш реалістичних результатів. Функція змістовної втрати, що ґрунтується на перцептивній подібності замість подібності в піксельному просторі. Наведений підхід призводить до найліпших показників, в порівнянні з тими, що були до нього.

У наступній роботі [21] пропонується дуже глибока, повністю згортована мережа автокодування для відновлення зображень. Мережа складається з багатьох рівнів операторів згортки та зворотної згортки, що вивчають відображення зіпсованих зображень на оригінальні. Щоб вирішити проблему, пов'язану з тим, що глибші мережі, як правило, складніше навчати, автори пропонують симетрично зв'язати згорткові та дезгорткові шари, завдяки чому навчання сходиться набагато швидше та досягає кращих результатів. Використовуючи ту саму структуру, модель навчають виконувати завдання зменшення шуму, СР, усунення артефактів стиснення JPEG й усунення розмиття зображення, та досягають високих показників для всіх чотирьох завдань.

У джерелі [22] досліджується проблема СР для зображень з дуже низькою РЗ (цебто СР з надвисоким коефіцієнтом масштабування), коли інформації замало й одному зображенню НР може відповідати багато зображень ВР. Автори наводять нову імовірнісну глибоку нейромережну

архітектуру, піксельну рекурсивну модель суперроздільності, що є розширенням PixelCNN для вирішення зазначеної проблеми.

У статті [23] відібрано 18 статей, із яких вибрано й оглянуто 19 алгоритмів СР зображень. Переглядаються та порівнюються певні аспекти цих алгоритмів, включаючи функції втрат, використані набори даних, методи оцінювання, модулі підвищення дискретизації та техніки проєктування. Для кожного модуля підвищення дискретизації та методу проєктування також було підсумовано відповідні переваги та вади.

Автори наступної праці [24] пропонують нову мережу концентрації уваги з дистиляцією інформації (AIDN) для ефективної та точної СР зображення, яка адаптивно модулює реакцію функцій шляхом моделювання взаємодії між виміром каналу та просторовими характеристиками. Зокрема, вводиться перетворення закритого каналу (GCT) для збору глобальної контекстної інформації між різними каналами для моделювання проміжних високорівневих функцій. Крім того, пропонується відкалібрований модуль уваги (RAM) для зміни розмірів виходів цих функцій, а RAM концентрує основний вміст навколо розріджених ділянок. Це підвищує ефективність обчислень і точність реконструкції. Комплексні кількісні та якісні оцінки демонструють, що AIDN перевершує сучасні моделі з погляду продуктивності реконструкції та якості зображення.

Розвідка [3] наводить огляд наявних методів СР зображення методами ГН і групують їх відповідно до ефективності реконструкції, точності реконструкції та точності сприйняття. Розглядаються набори даних, методи підвищення дискретизації, оптимізацію, методи оцінювання роботи мереж, а також царини застосування СР зображень. Також обговорюються деякі проблеми, що досі є в завданні СР одного зображення, та наводяться тенденції та напрямки майбутнього розвитку підходу.

Одним з основних обмежень поточних підходів СР зображення є те, що вони передбачають попередньо визначену модель деградації або ядро, напр. бікубічне, що контролює процес деградації зображення для створення

навчальних наборів даних. Проте в реальному середовищі зображення з низькою роздільністю можуть не підходити під певне ядро, що призводить до зменшення ефективності роботи моделей. У роботі [25] представлено модель SR зображення, що навчає мережу SR тільки на зображеннях з низькою якістю, не покладаючись на зовнішні навчальні набори. Автори наводять нову нейромережу DGDML-SR, що складається з двох окремих мереж: мережі SR зображення й мережі моделювання деградації (DAN), що орієнтується на конкретні зображення і створена для вирішення зазначеної вище проблеми.

Праця [26] вводить мережу уваги до спільних ознак залишкового каналу (RCSFANet), яка використовує ознаки подібності різних каналів для захоплення матриці уваги. Зокрема, матриця уваги подібності застосовується до всієї моделі, щоб упередити розподіл доступних ресурсів оброблення та вибірково підкреслити взаємозалежні карти каналів. Також автори розробили залишок в структурі залишків, що дозволяє одержувати велику кількість інформації в зображеннях НР прямо через кілька пропускових з'єднань, завдяки чому мережа більше зосереджується на вивченні високочастотної інформації. Все це призводить до ліпшої точності та візуальних покращень.

Й остання робота, про яку хотілося б згадати, це [27]. У цій розвідці розглядається застосування SR зображень для систем біометричної ідентифікації особистості, а саме: обличчя (2D та 3D), райдужки ока, відбитків пальців та ходи. Автори підходять до проблеми SR в біометрії з кількох поглядів, включно з розрідженою та частотною областями, одним і кількома вхідними зображеннями, підходами на основі навчання та реконструкції. Також обговорюються поточні проблеми та виклики для дослідників та надаються рекомендації на майбутнє щодо поліпшення використання SR у біометрії.

1.3 Постановка задачі дослідження

Актуальність цього дослідження полягає у важливості систем покращення якості зображень, їх широкому спектрі сфер застосування в сучасному світі, а також існуванні ряду пов'язаних із ними проблем, потребі в поліпшенні ефективності наявних та пошуку нових методів СР, заснованих на глибокому навчанні.

Об'єктом дослідження є нейромережні методи покращення якості зображень.

Метою дослідження є всебічний огляд та аналіз нейромережного підходу до покращення якості зображень.

Враховуючи мету роботи, необхідно вирішити такі завдання:

- дослідити модулі підвищення дискретизації, що застосовуються в моделях СР;
- дослідити нейромережні архітектури та підходи для СР зображень;
- проаналізувати функції втрат, що використовуються для навчання моделей СР;
- проаналізувати метрики оцінювання якості роботи моделей СР;
- дослідити набори даних, що застосовуються в навчанні та тестуванні систем СР;
- проаналізувати продуктивність сучасних моделей СР зображення;
- визначити майбутні виклики та напрямки дальших досліджень.

2 ДОСЛІДЖЕННЯ ПРИНЦИПІВ ПОКРАЩЕННЯ ЯКОСТІ ЗОБРАЖЕНЬ ІЗ ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖ

2.1 Методи підвищення дискретизації зображення

Як вже було зазначено раніше, мета СР полягає в тому, щоб збільшити зображення меншого розміру до більшого й при цьому зробити збільшене зображення якомога точнішим. Тому операція розширення, яка також називається підвищенням дискретизації, є важливим ступенем у СР. У цьому розділі розглянуто декілька основних методів підвищення дискретизації, що застосовуються у сучасних мережах СР.

2.1.1 Інтерполяція

Інтерполяція зображення, або ж масштабування зображення, стосується до зміни розміру цифрових зображень і широко використовується в програмах, пов'язаних із зображеннями. Традиційні методи інтерполяції включають інтерполяцію найближчого сусіда, білінійну [5] та бікубічну [6] інтерполяцію, фільтр Ланцоша [28] тощо. Оскільки ці методи просто інтерпретуються та є легкими для реалізації, деякі з них все ще широко використовуються в моделях СР на основі ЗНМ (рис. 2.1).

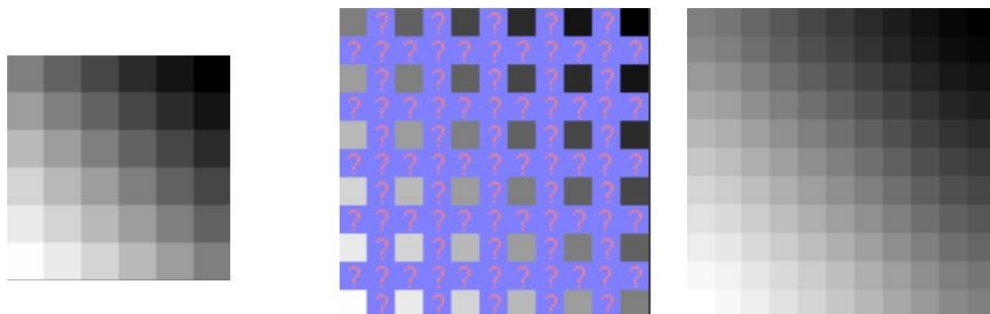


Рисунок 2.1 – Приклад застосування інтерполяції для збільшення зображення на 80%

Інтерполяція найближчого сусіда є простим та інтуїтивно зрозумілим алгоритмом. Він вибирає значення найближчого пікселя для кожної позиції, що підлягає інтерполяції, незалежно від будь-яких інших пікселів. Таким робом, цей метод є дуже швидким, але зазвичай дає результат у вигляді блоків низької якості.

Білінійна інтерполяція виконує лінійну інтерполяцію спочатку на одній, а потім на іншій осі зображення. Оскільки це призводить до квадратичної інтерполяції з рецептивним полем розміром 2×2 , вона показує набагато кращу продуктивність, ніж інтерполяція найближчого сусіда, зберігаючи при цьому відносно високу швидкість.

Подібним чином бікубічна інтерполяція виконує кубічну інтерполяцію на кожній із двох осей. Порівняно з білінійною, бікубічна інтерполяція враховує 4×4 пікселя, що забезпечує більш плавні результати з меншою кількістю артефактів, але набагато меншою швидкістю. Насправді бікубічна інтерполяція зі згладжуванням є основним методом створення наборів даних СР (тобто, погіршення зображень ВР до зображень НР), а також широко використовується у механізмі СР перед підвищенням дискретизації, що його буде розглянуто далі.

Методи підвищення дискретизації на основі інтерполяції покращують РЗ зображення лише на основі власних сигналів зображення, не надаючи додаткової інформації. Натомість вони мають деякі побічні ефекти, такі як обчислювальна складність, посилення шуму, розмиття результатів [29]. Окрім того, згорткові ядра інтерполяції неможна навчити, а значить із використанням цього методу втрачається певна корисна інформація. Отже, сучасна тенденція полягає в заміні методів на основі інтерполяції на шари підвищення дискретизації, здатні до навчання.

2.1.2 Розгортка

Шар транспонованої згортки, яку також називають розгорткою [30], намагається виконати перетворення, що є протилежним нормальній згортці, цебто прогнозує можливий вхід на основі карт ознак із розміром як вихід згортки. Зокрема, він збільшує РЗ зображення, розширюючи зображення шляхом вставки нулів і виконання згортки. Взявши як приклад $2 \times \text{CP}$ із ядром 3×3 (рис. 2.2), вхідні дані спочатку розгортаються вдвічі порівняно з початковим розміром, де додані значення пікселів встановлюються на 0. Потім застосовується згортка з ядром розміру 3×3 , кроком 1 і відступом од країв 1. Таким способом, вхідний сигнал збільшується в 2 рази, і в цьому випадку рецептивне поле становить щонайбільше 2×2 .

Оскільки транспонована згортка збільшує розмір зображення наскрізним способом, зберігаючи шаблон зв'язку, сумісний зі звичайною згорткою, він широко використовується як шар підвищення дискретизації в моделях СР. Однак цей шар може легко спричинити «нерівномірне перекриття» на кожній осі, а множення результатів на обох осях призводить до появи шахового візерунка різної величини та, таким чином, погіршує продуктивність СР.

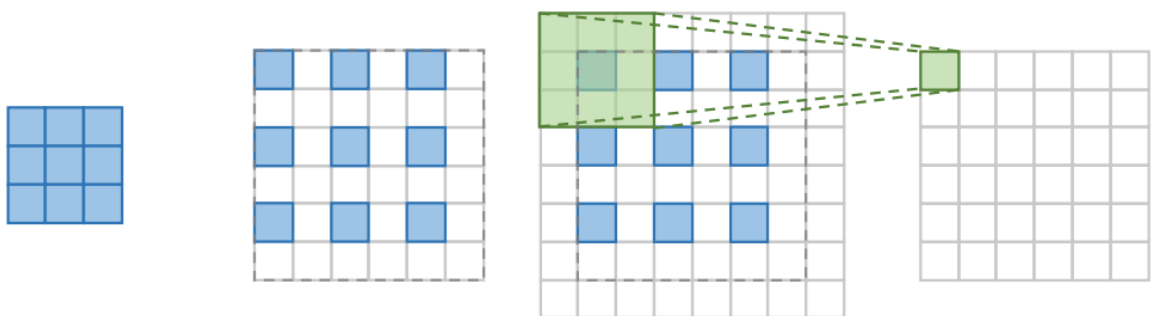


Рисунок 2.2 – Шар транспонованої згортки

2.1.3 Субпіксельний шар

Субпіксельний шар [31] – ще один наскрізний навчальний шар підвищення дискретизації. У межах цього шару спочатку застосовується згортка для отримання виходів із s^2 каналами, де s – коефіцієнт масштабування. Якщо припустити, що вхідний розмір $h \times w \times c$, то вихідний розмір буде $h \times w \times s^2 c$. Після цього застосовується оператор періодичного тасування, щоб одержати виходи розміром $sh \times sw \times c$ (рис. 2.3). У цьому випадку рецептивне поле може бути до 3×3 . Загалом субпіксельний згортковий шар можна визначити наступним чином:

$$I = f^L(I_x) = PS(W_L * f^{L-1}(I_x) + b_L), \quad (2.1)$$

де PS – оператор періодичного тасування, який переносить тензор $h \times w \times s^2 c$ до тензора розмірності $rh \times rw \times c$, де $rh \times rw$ – розмір зображення ВР, c – розмірність оброблюваних каналів. Крім того, згортковий фільтр W_L має форму $n_{L-1} \times s^2 c \times K_L \times K_L$, де n_L – кількість карт ознак у шарі $L-1$.

Порівняно з шаром транспонованої згортки субпіксельний шар має більше рецептивне поле, що надає більше контекстної інформації, щоб допомогти створити більш реалістичні деталі, тому він також широко використовується в моделях СР на основі ГН. Однак, оскільки розподіл рецептивних полів є нерівномірним і блокові області фактично мають одне й те саме рецептивне поле, це може призвести до деяких артефактів на кордоні сусідніх блоків.

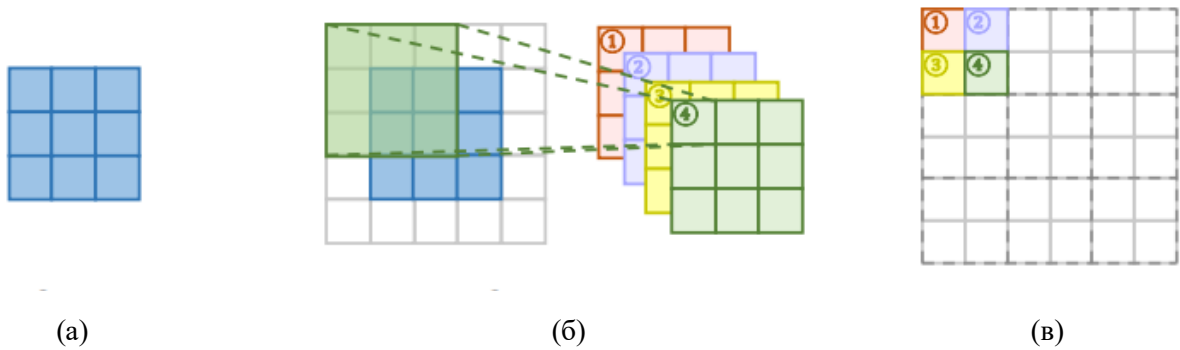


Рисунок 2.3 – Принцип роботи субпіксельного шару: (а) – початок; (б) – згортка; (в) – переформатування

2.1.4 Модуль метарозширення

Зазначені методи потребують попереднього визначення коефіцієнту масштабування, тобто навчання різних модулів підвищення дискретизації для різних коефіцієнтів, що є неефективним і не відповідає реальним потребам. У роботі [32] запропонували модуль метарозширення, що вирішує завдання SR із довільними коефіцієнтами масштабування і ґрунтується на метанавчанні.

Зокрема, для кожної цільової позиції на зображеннях VR цей модуль проєктує її на невеликий окіл на картах ознак HR (тобто $k \times k \times c_{in}$), прогнозує ваги згортки (тобто $k \times k \times c_{in} \times c_{out}$) відповідно до зміщення проєкцій та коефіцієнту масштабування та виконує згортку. Таким чином, модуль метарозширення може безперервно збільшувати зображення з довільними коефіцієнтами за допомогою однієї моделі. А завдяки великій кількості навчальних даних (одночасне навчання кількох факторів) модуль може демонструвати порівняну або навіть ліпшу продуктивність на фіксованих факторах. Незважаючи на те, що цей модуль потребує прогнозування вагових коефіцієнтів під час створення висновку, час виконання модуля підвищення дискретизації становить лише близько 1% часу вилучення ознак. Однак цей метод передбачає велику кількість вагових коефіцієнтів згортки для кожного цільового пікселя на основі кількох значень, незалежних од вмісту

зображення, тому результат передбачення може бути нестабільним і менш ефективним у разі більшого розширення.

На теперішній час ці здатні до навчання шари є найпоширенішими методами підвищення дискретизації. Особливо в механізмі SR після підвищення дискретизації, де ці шари використовуються на завершальному щаблі підвищення дискретизації для реконструкції зображень VR на основі представлень високого рівня, вилучених у просторі низького рівня, і в такий спосіб досягають наскрізної SR, уникаючи великозатратних операцій у просторі високого рівня.

2.2 Структури мереж SR

Проектування мережі є однією з найважливіших частин глибокого навчання. У галузі суперроздільності дослідники застосовують різноманітні стратегії проектування мереж поверх чотирьох основних каркасів SR, що поділяються відповідно до того, на якому місці в мережі розташований модуль підвищення дискретизації:

- SR до підвищення дискретизації;
- SR після підвищення дискретизації;
- прогресивне підвищення дискретизації;
- ітеративне підвищення та пониження дискретизації.

У цьому розділі розглядаються основні підходи до проектування глибоких мереж, що їх застосовують у завданні суперроздільності. Загальну структуру цих підходів зображено на рисунку 2.4.

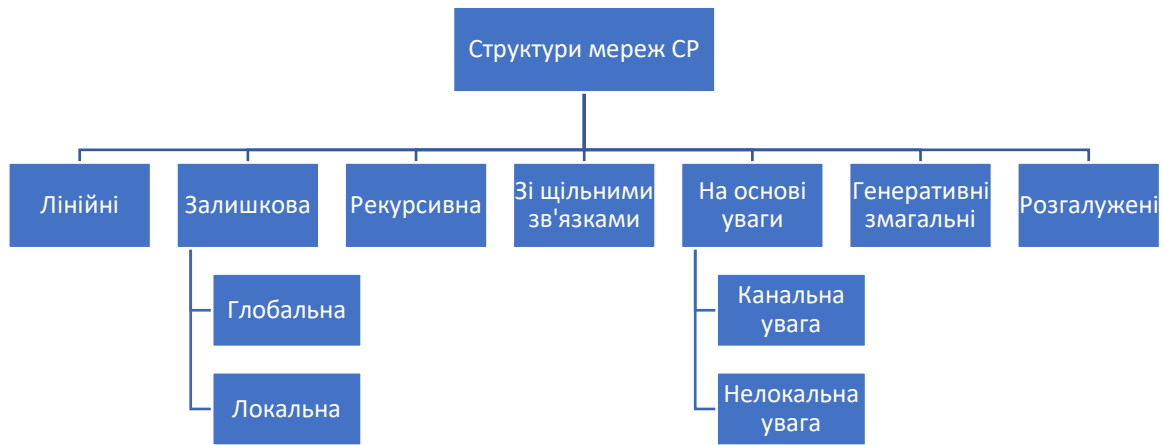


Рисунок 2.4 – Основні структури мереж CP

2.2.1 Лінійні мережі

Найпершою архітектурою CP була лінійна мережа, тобто найпростіший варіант глибокої нейронної мережі. Прикладами лінійних мереж є SRCNN [10], FSRCNN [33] і ESPCN [31].

SRCNN була першою роботою, що використовувала згорткової нейронної мережі (ЗНМ) для реконструкції зображень методами CP. Ідея SRCNN була надихнута методами CP на основі розрідженого кодування. SRCNN – це дуже проста архітектура, що складається з трьох основних рівнів (рис. 2.5), а саме вилучення патчів, нелінійного відображення та реконструкції зображення. Шар вилучення патчів витягає ознаки з бікубічної інтерполяції вхідного зображення HP. Потім виділені ознаки проходили через нелінійне відображення, де кожен з об'єктів великої розмірності відображався на іншій об'єкті великої вимірності. Шар нелінійного відображення складається зі згорткових фільтрів 1×1 , які використовуються для зміни кількості каналів і додавання нелінійності. Нарешті, вихідні ознаки з останнього шару нелінійного відображення реконструюються в зображення BP за допомогою процесу згортки. Функцією втрат для оптимізації SRCNN є

середньоквадратична помилка (MSE), яку буде розглянуто в наступному розділі.

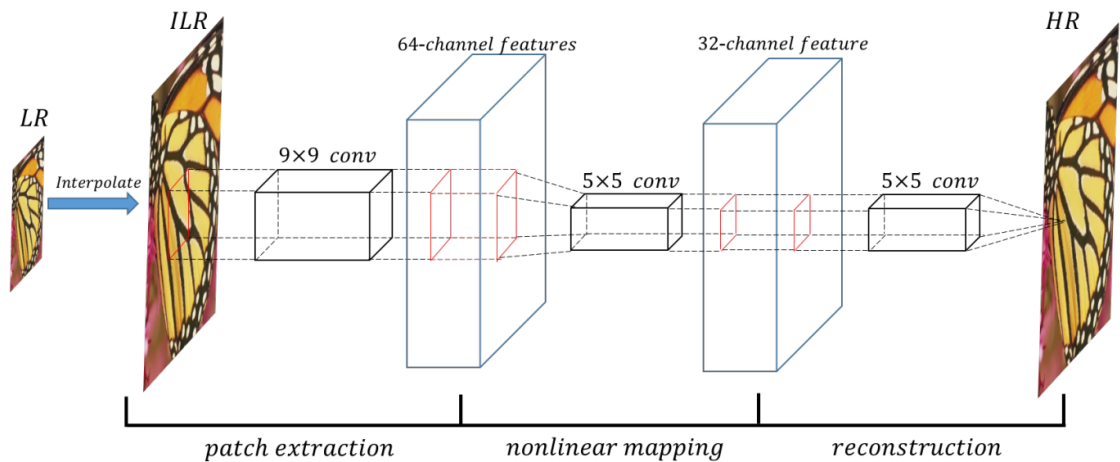


Рисунок 2.5 – Структура мережі SRCNN

SRCNN продемонстрував величезну перевагу над традиційними методами, проте деякі проблеми надихнули на створення більш ефективних архітектур:

- на вході в SRCNN стоїть модуль бікубічної інтерполяції, що призводить до розмивання деталей зображення, а дальша робота з вже інтерпольованим зображенням вимагає вельми багато часу;
- SRCNN – це лише трирівнева архітектура, а тому має обмежений простір для виділення ознак і обробки вхідного зображення;
- попередня інформація, що відбиває властивості зображення ВР, у функції втрат є тривіальною.

Пізніше було з'ясовано, що SRCNN вимагає більше згорткових шарів у нелінійному відображенні, щоб отримати кращий результат. Однак збільшення кількості шарів збільшує час роботи, і показнику PSNR важко сходиться під час навчання. Для подолання цих проблем було запропоновано FSRCNN [33] (Fast SRCNN).

Як видно на рисунку 2.6, FSRCNN містить п'ять основних частин: виділення ознак, звуження, нелінійне відображення, розширення та

реконструкція зображення. Отже, першою відмінністю FSRCNN од SRCNN є відсутність попередньої обробки або розширення на початку, тобто подальше вилучення ознак відбувається у просторі НР. Наступним нововведенням була операція звужування, тобто згортка 5×5 , щоб зменшити кількість каналів, а значить і обчислень. Крім того, використовується кілька згорток 3×3 замість великого згорткового фільтра. Для виконання операції підвищення дискретизації застосовується розгортка і FSRCNN є першою роботою, яка використовує цей шар для реконструкції зображень ВР з карт ознак НР.

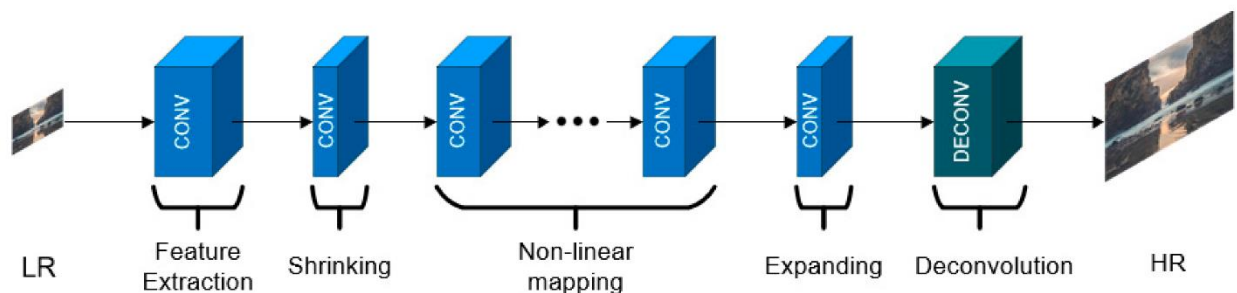


Рисунок 2.6 – Структура мережі FSRCNN

FSRCNN досяг покращення якості результату на 1,3% порівняно з SRCNN, а також скоротив час роботи в середньому на 78% порівняно з SRCNN.

ESPCN [31] було розроблено, щоб подолати проблему складності в SRCNN, оскільки вона зростала квадратично. ESPCN має таку саму структуру, як і SRCNN, за винятком модуля підвищення дискретизації: у ESPCN замість бікубічної інтерполяції використовувалася субпіксельна згортка, що є значно ефективнішим рішенням. Субпіксельна згортка має перевагу й над шаром розгортки у FSRCNN, а саме вирішує проблему появи візерунків у вигляді шахової дошки, що виникає через перекривання операції згортки.

2.2.2 Залишкові мережі

У дослідженні ГН існує теоретична робота, яка показує, що простір рішень ГНМ можна розширити шляхом збільшення її глибини або ширини. Однак пряме укладання шарів (розширення) спричиняє зникнення/вибухання градієнтів і проблему деградації. Тим часом додавання більшої кількості шарів призведе до вищої похибки навчання та дорожчих витрат на обчислення.

У мережі ResNet [34] було запропоновано метод залишкового навчання, в якому вивчаються тільки залишки замість підгонки ретельного відображення. Стратегії залишкового навчання можна розділити на глобальне та локальне остаточне навчання.

Оскільки зображення SR є завданням трансляції із зображення в зображення, де вхідне зображення сильно корелює з цільовим, досить легко змодельовати залишкове відображення між ними, що називається глобальним залишковим навчанням. В цьому випадку уникається вивчення складного перетворення від повного зображення до іншого, натомість треба лише вивчити залишкову карту для відновлення відсутніх високочастотних деталей. Оскільки залишки в більшості регіонів близькі до нуля, складність моделі та труднощі навчання значно зменшуються. Таким чином, він широко використовується моделями SR.

Локальне залишкове навчання подібне до залишкового навчання в ResNet і використовується для полегшення проблеми деградації, спричиненої глибиною мереж, що повсякчас зростає, зменшення складності навчання та покращення здатності до навчання.

На практиці вищезазначені методи обидва реалізуються за допомогою коротких зв'язків (часто масштабованих невеликою константою) і поелементним додаванням, тоді як різниця полягає в тому, що перший безпосередньо з'єднує вхідні та вихідні зображення, тоді як другий зазвичай додає кілька ярликів між шарами з різною глибиною всередині мережі.

VDSR [35] є першою дуже глибокою моделлю, яка використовується в СР. VDSR є 20-шаровою VGG-мережею (рис. 2.7). Архітектура VGG встановлює всі розміри ядра як 3×3 . Окрім інноваційної архітектури, VDSR зробив ще два внески. Перший полягає в тому, що одна модель використовується для кількох масштабів, оскільки процеси СР з різними коефіцієнтами масштабу мають міцний зв'язок один з одним. Другим внеском є залишкове навчання. VDSR запровадив залишкове навчання між входом і виходом останнього шару відображення ознак. Залишкове навчання додало вихідні ознаки з останнього рівня до інтерпольованих функцій через пропуск з'єднання. Оскільки низько- та високорівневі функції сильно корельовані, пропуск з'єднання допоміг використати функції низького рівня шляхом поєднання їх із високорівневими функціями та покращив продуктивність моделі. Таким способом, пропуск з'єднання змогло вирішити проблему зникнення градієнтів, спричинене збільшенням кількості шарів у моделі.

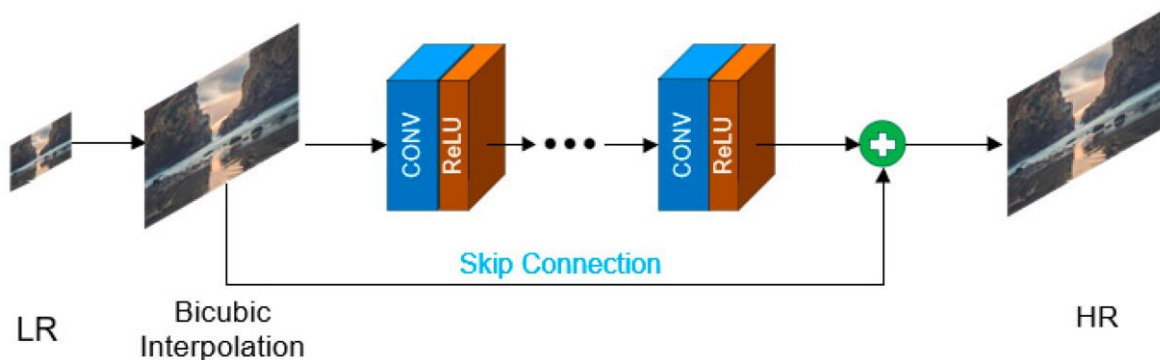


Рисунок 2.7 – Структура мережі VDSR

Розгортання залишкового навчання у VDSR має дві переваги перед SRCNN. По-перше, це допомогло мережі сильно скоротити час виконання. Загалом у моделі VDSR спостерігається скорочення часу роботи на 93,9%. По-друге, VDSR забезпечив краще значення точності, ніж SRCNN.

Надихаючись залишковою мережею VDSR та архітектурою SRResNet [36], було запропоновано мережу EDSR [37], що покликана вирішити проблему великого часу обчислення та споживання пам'яті через

застосування бікубічної інтерполяції. EDSR містить три основні модулі, якими є вилучення ознак, модуль залишкового блоку та модуль підвищення дискретизації (рис. 2.8). Ідея залишкового блоку походить од SRResNet, тоді як ідея пропуску з'єднання – од VDSR. В окремому залишковому блоці також було застосовано залишкове навчання між вхідними та вихідними функціями. Різниця між VDSR і EDSR полягала в тому, що деякі шари у VDSR були замінені залишковими блоками. Крім того, EDSR використовує субпіксельну згортку як модуль підвищення дискретизації.

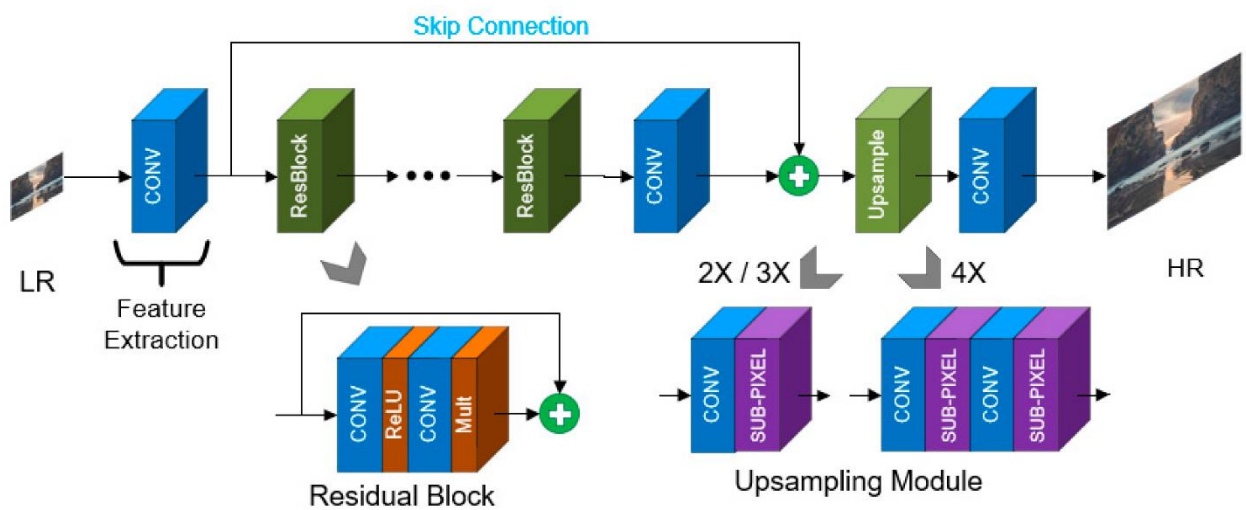


Рисунок 2.8 – Структура мережі VDSR

EDSR використовує інформацію з кожного залишкового блоку шляхом введення залишкового навчання всередині блоку, що значно покращило значення показника якості PSNR приблизно на 3,84% порівняно з VDSR. Крім того, застосування субпіксельної інтерполяції також допомогло скоротити час обчислення та споживання пам'яті.

2.2.3 Рекурсивні мережі

Задля вивчення функцій високого рівня без введення великої кількості параметрів, запропоновано використовувати рекурсивне навчання, що означає багаторазове застосування одних і тих же модулів рекурсивним способом.

DRCN [38] був першим алгоритмом, що застосував рекурсивний метод для завдання SR. Той самий згортковий шар у DRCN застосовується 16 разів, що призводить до одержання рецептивного поля розміром 41×41 , що набагато більше, ніж 13×13 у SRCNN, і при цьому без введення великої кількості параметрів. DRCN складається з трьох основних частин, а саме мережі вбудовування, мережі висновків і мережі реконструкції (рис. 2.9). Мережа вбудовування вилучала функції з інтерпольованого зображення. Потім виділені ознаки пропускалися через мережу висновків, в якій усі фільтри мали однакову вагу. Усі проміжні результати кожного згорткового шару в мережі висновків та інтерпольовані функції згортаються перед тим, як їх додають разом для формування зображення HR.

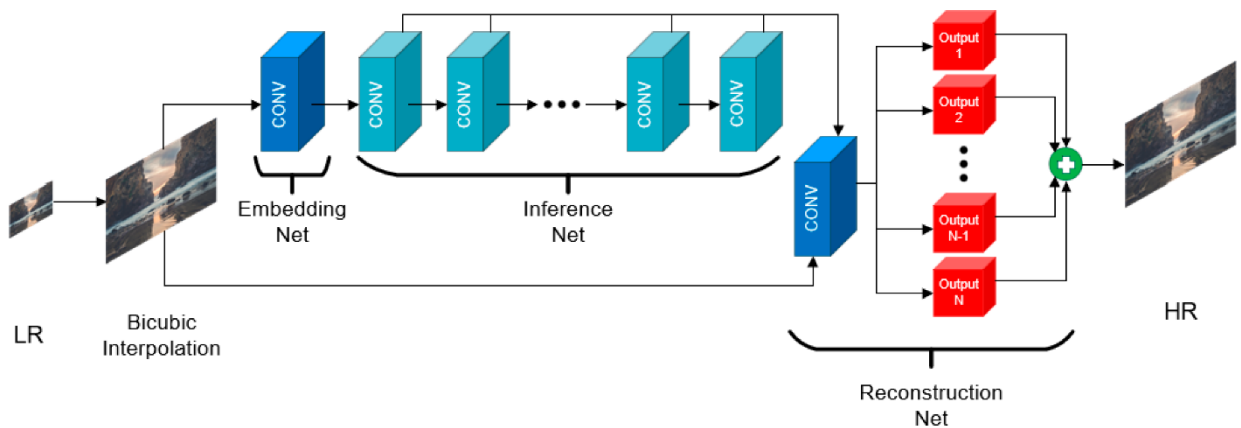


Рисунок 2.9 – Структура мережі DRCN

DRCN продемонструвала вагоме покращення якості реконструкції в порівнянні з SRCNN, проте в ній наявні значні недоліки. Одним із них є те, що DRCN вимагає нагляду за кожною рекурсією, що є обтяжливим процесом. По-друге, між усіма згортковими шарами в мережі висновків розподіляється лише

один тип ваги. Окрім того, накладене використання рекурсивних модулів також створює деякі проблеми: градієнтний вибух або його зникнення.

З огляду на ці проблеми, було розроблено мережу DRRN [39]. Використовуючи основну ідею з DRCN, мережа висновків була замінена на рекурсивний блок. Рекурсивний блок складався з кількох залишкових модулів, причому кожний залишковий модуль має два згорткових шари (рис. 2.10). Один із згорткових шарів у кожному залишковому блоку має одну однакову вагу, тоді як інший згортковий шар має іншу однакову вагу.

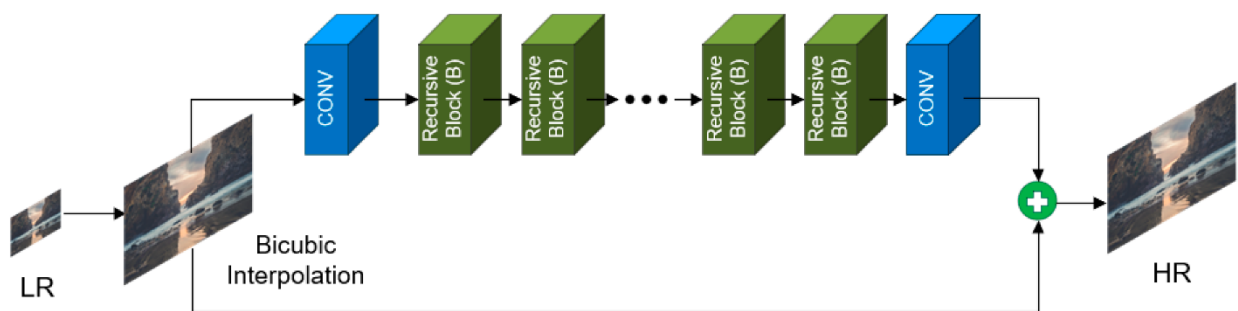


Рисунок 2.10 – Структура мережі DRRN

Значення PSNR в DRRN показало покращення на 0,7% порівняно з DRCN. Поліпшення спостерігалось з таких причин. По-перше, залучення залишкового навчання в залишковий блок допомогло вирішити проблему деградації, що впливає на продуктивність моделі. По-друге, заощаджується багато обчислювальних ресурсів, оскільки багато шарів мають однакові ваги. Крім того, DRRN звільнив од нагляду за кожною рекурсією, розробивши рекурсивний блок із розгалуженою структурою.

Пізніше було представлено MemNet [40] на основі блоку пам'яті, який складається з 6-рекурсивного залишкового блоку, де виходи кожної рекурсії об'єднані та проходять додаткову згортку 1×1 для запам'ятовування та забування. Каскадна залишкова мережа (CARN [41]) також використовує подібний рекурсивний блок, що включає кілька залишкових блоків.

Крім того, EBRN використовує рекурсивне навчання для досягнення диференціації інформації з різною частотою, тобто низькочастотна інформація

обробляється короткими модулями в мережі, а високочастотна інформація обробляється глибокими модулями.

2.2.4 Мережі зі щільними зв'язками

Механізм щільного зв'язку, що його представлено в мережі DenseNet [42], набув значного поширення в задачах комп'ютерного зору. Для кожного шару в щільному блоці карти ознак усіх попередніх шарів використовуються як вхідні дані, а власні карти ознак використовуються як вхідні дані для всіх наступних шарів (рис. 2.11). Це призводить до $l \cdot (l - 1)/2$ зв'язків у щільному блоці l -того шару ($l \geq 2$). Щільні з'єднання не тільки допомагають запобігти зникненню або вибухові градієнта й поліпшують поширення сигналу та заохочують повторне використання функцій, але також значно зменшують розмір моделі завдяки використанню невеликої швидкості зростання і стискання каналів після об'єднання всіх вхідних карт функцій.

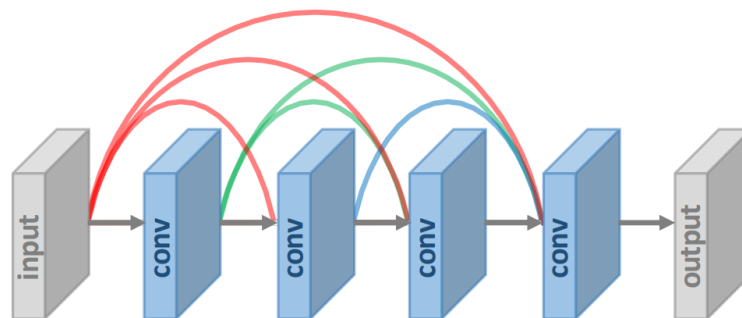


Рисунок 2.11 – Принцип роботи щільного блоку

Щоб поєднати функції низького та високого рівнів і одержати ціннішу інформацію для реконструкції високоякісних деталей в царину CP ввели механізм щільних зв'язків. Дизайн SRDenseNet [43] був натхненний ідеєю DenseNet, яка має можливість покращити потік інформації через мережу. SRDenseNet використовує щільні зв'язки не тільки на рівні шару, але й на рівні

блоку, де вихід кожного щільного блоку з'єднаний щільними зв'язками (рис. 2.12). Таким робом, функції низького рівня та функції високого рівня поєднуються та повною мірою використовуються для проведення реконструкції.

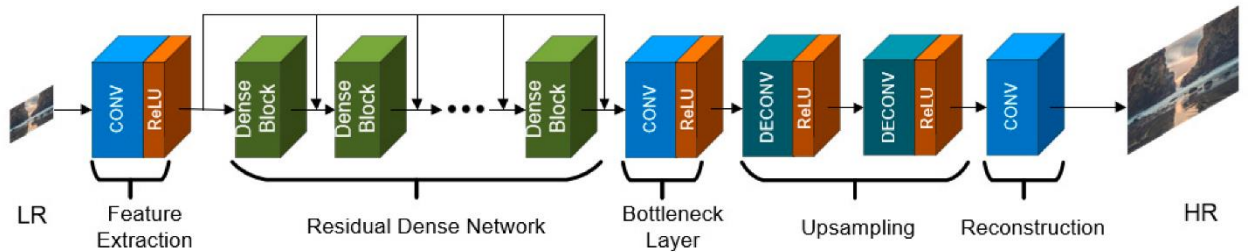


Рисунок 2.12 – Структура мережі SRDenseNet

У RDN [44] щільні з'єднання поєднуються із залишковим навчанням для формування залишкового щільного блоку (рис. 2.13), який дозволяє обходити низькочастотні функції через кілька пропускних з'єднань, роблячи основну гілку зосередженою на навчанні високочастотної інформації, тим самим сильно полегшуючи процес навчання мережі. Поєднання глобального залишкового навчання, локального залишкового навчання та щільного зв'язку призвело до того, що RDN працює краще, ніж SRDenseNet. Було досягнуто покращення точності на 1,3% порівняно з SRDenseNet.

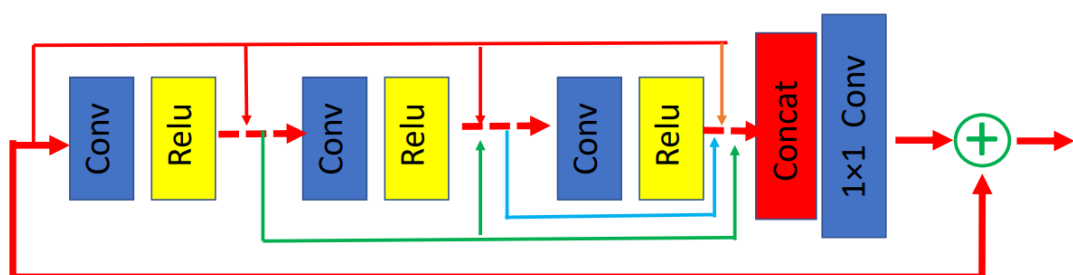


Рисунок 2.13 – Приклад залишкового щільного блоку

MemNet [40] рекурсивно використовує залишковий модуль для заміни нормальної згортки в блоці базової DenseNet і додає щільні зв'язки між

різними блоками. Автори пояснили, що локальні зв'язки в одному блоці нагадують короткочасну пам'ять, а зв'язки з попередніми блоками нагадують довготривалу пам'ять, через що мережа й одержала свою назву.

Окрім вищезгаданих моделей, щільне з'єднання також застосовується в мережах CARN, ESRGAN, RPMNet, MFNet та інші. DBPN [45] також широко використовує щільні зв'язки, проте вона заснована на принципі ітераційного підвищення та зменшення дискретизації і її щільні зв'язки існують між усіма модулями підвищення дискретизації та пониження дискретизації.

2.2.5 Мережі на основі уваги

Механізм уваги був запропонований у зв'язку з тим, що згорткові нейронні мережі більше зосереджені на локальній інформації та ігнорують глобальні особливості. Механізм уваги широко використовується в різних задачах комп'ютерного зору, часто вставляється в основну мережу як додатковий компонент, і його основною метою є розподіл обчислювальних ресурсів для більш важливих завдань. Коротше кажучи, механізм уваги допомагає мережі ігнорувати нерелевантну інформацію та зосереджуватися на важливих, інформативних деталях. Механізми уваги в основному поділяються на каналну увагу та нелокальну увагу.

Канальна увага. У SISR ми головним чином хочемо відновити якомога більше цінної високочастотної інформації. Однак звичайні методи, засновані на ЗНМ, однаково обробляють функції каналів, що не забезпечує гнучкості роботи з різними типами інформації. Розглядаючи взаємозалежність і взаємодію репрезентацій ознак між різними каналами, Ху та інші [46] пропонують блок «стискання та збудження» (СЗ) для покращення здатності до навчання шляхом явного моделювання взаємозалежності каналів (рис. 2.14). У цьому блоці кожен вхідний канал стискається в дескриптор каналу (тобто константу) за допомогою глобального середнього об'єднання

(GCO), потім ці дескриптори подаються на два щільні шари для створення коефіцієнтів масштабування каналів для вхідних каналів.

Пізніше Жанг та інші [47] запропонували новий модуль, заснований на механізмі СЗ, – RCAB. Цей модуль включений у мережу RCAN, яка помітно поліпшила здатність представлення моделі та продуктивність СР. Як показано на рисунку 2.14, шар глобального середнього об'єднання, за яким іде функція Sigmoid, змінює масштаб кожного каналу функції, що дозволяє мережі зосередитися на більш корисних каналах і покращує здатність до розрізнявального навчання.

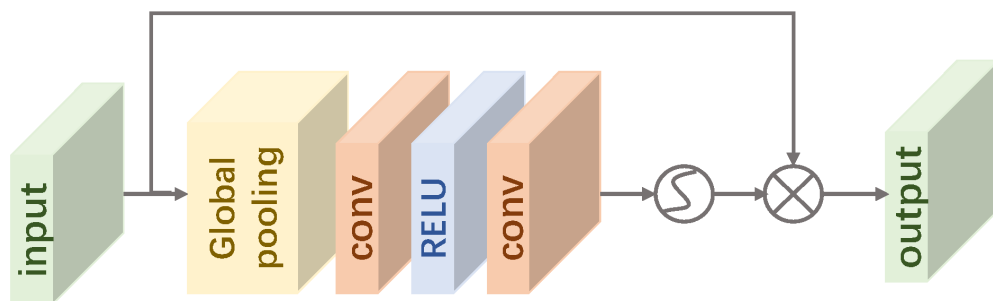


Рисунок 2.14 – Механізм каналної уваги

Оскільки увага каналу розглядає кожен згортковий шар як окремий процес і ігнорує кореляцію між різними шарами, використання цього алгоритму призводить до втрати деяких проміжних функцій під час реконструкції зображення. Тому створено цілісну мережу уваги (HAN [48]), що складається з модуля уваги шару (LAM) і модуля уваги простору каналу (CSAM). LAM може призначати різні ваги уваги об'єктам у різних шарах, одержуючи залежності між об'єктами різної глибини, а потім використовувати CSAM для вивчення кореляцій у різних місцях на кожній карті функцій, щоб більш ефективно охоплювати глобальні об'єкти.

Подібним чином модуль каналної уваги другого порядку (SOCA) у SAN [49] вивчає міжканальні кореляції ознак за допомогою статистики ознак другого порядку. Велика кількість експериментів показала, що канална увага

другого порядку може допомогти мережі одержати більш дискримінаційні уявлення, що призводить до вищої точності реконструкції.

Нелокальна увага. Через обмежений розмір рецептивного поля більшість мереж СР зображень здатні лише виділяти локальні особливості на зображеннях, ігноруючи кореляцію між дальніми ознаками. Однак вони можуть надати важливу інформацію для реконструкції зображень. Враховуючи цю проблему, було запропоновано механізм нелокальної уваги як алгоритм фільтрації для обчислення середнього зваженого значення всіх пікселів зображення.

Наприклад, призначення нелокального модуля RL-NL регіону в SAN полягає в тому, щоб розділити вхідне зображення на певні розміри та виконати нелокальні операції над кожним регіоном. Лю та інші [50] запропонували нелокальну рекурентну мережу (NLRN) для введення нелокальних операцій у рекурентні нейронні мережі для завдання відновлення зображень, щоб одержати кореляцію глибинних ознак у кожному регіоні з їхніми сусідніми ознак.

Блок нелокальної уваги пропонується в RNAN [51], де механізми уваги як каналної, так і просторової використовуються одночасно в його гілці-масці, щоб краще керувати виділенням ознак у основній гілці. Тим часом, у HAN пропонується цілісна мережа уваги, яка складається з модуля уваги шару та модуля просторово-каналної уваги, щоб моделювати цілісну взаємозалежність між шарами, каналами та позиціями.

2.2.6 Генеративні змагальні мережі

Мережі, що обговорювалися до цього, оптимізують різницю в пікселях між прогнозованим та вихідним зображенням ВР. Хоча цей показник робить нормально, він не ідеальний, адже люди розрізняють зображення не за різницею в пікселях, а за якістю сприйняття.

У 2014 році Гудфеллов та інші [52] запропонували генеративні змагальні мережі (ГЗМ, GAN), які широко використовувалися в задачах КЗ, таких як передача стилю та відновлення зображень. ГЗМ складається з генератора та дискримінатора. Генератор створює зображення з високою роздільною здатністю, яке дискримінатор повинен визначити як реальне зображення високої РЗ або штучно створене зображення з високою РЗ. Якщо дискримінатор не бачить різниці між вихідним зображенням моделі та справжнім ВР-зображенням, то цей одержане зображення вважається хорошим наближенням зображення високої РЗ. В такий спосіб, дискримінатор обмежує вихід відповідно до попередньо визначеного розподілу. У моделях ГЗМ значення оцінки якості PSNR знижуються, але якість сприйняття загалом покращується.

Натхненні ГЗМ, Ледиг та інші запропонували SRGAN [20]. У SRGAN генератор по суті є моделлю СР, яка навчена обманювати дискримінатор, а дискримінатор навчений відрізнити зображення СР од зображень ВР. Таким чином, генератор навчається створювати результати, які дуже схожі на зображення ВР, а потім реконструювати природніші СР зображення.

Генератором у SRGAN є SRResNet, а дискримінатор використовує архітектуру, запропоновану Радфордом та іншими [53] (рис. 2.15). SRGAN використовує багатокомпонентну функцію втрат із кількох частин:

- втрата MSE, яка сприяє подібності пікселів;
- втрата подібності сприйняття, що використовується для захоплення інформації високого рівня, використовуючи глибоку мережу;
- спеціальна змагальна втрата дискримінатора.

У ESRGAN [54], що базується на SRGAN, зробили дві модифікації SRResNet:

- замінили оригінальний залишковий блок на щільний блок залишку в залишку;
- видалили шари групової нормалізації для покращення здатності моделі до узагальнення.

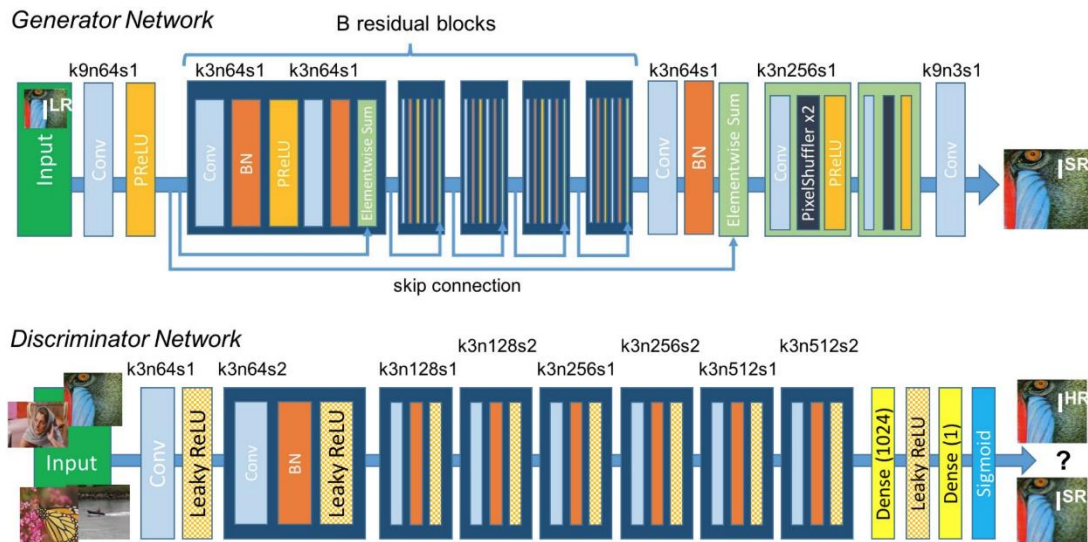


Рисунок 2.15 – Структура мережі-генератора та мережі-дискримінатора

У моделі також використовується розширений дискримінатор. Покращена втрата сприйняття вводиться за допомогою функцій VGG.

SRFeat [55] є ще однією CP ГЗМ моделлю із розрізненням ознак. Було досліджено, що методи CP на основі ГЗМ, як правило, створюють менш значущий високочастотний шум у реконструйованих зображеннях. Тому в SRFeat введено додатковий дискримінатор, щоб допомогти генераторові генерувати високочастотні структурні ознаки, а не шумові артефакти.

2.2.7 Розгалужені мережі

Як було зазначено раніше, глибші мережі дають кращі результати. Але навчання більш глибоких мереж є важким через проблему потоку інформації [56]. Залишкові мережі певною мірою вирішують це за допомогою коротких з'єднань. Розгалужені мережі покликані покращити потік інформації, маючи кілька гілок, через які може проходити інформація, що призводить до об'єднання інформації з багатьох рецептивних полів i , отже, до

кращого навчання. Розгалужене навчання можна розділити на глобальне, локальне та масштабо-специфічне розгалужене навчання.

Глобальне розгалужене навчання. Глобальне розгалужене навчання означає використання кількох шляхів для виділення функцій різних аспектів зображень. Ці шляхи можуть перетинатися під час розповсюдження і таким чином значно підвищити здатність до навчання. Однією з таких моделей є мережа LapSRN [57], що заснована на підході прогресивного підвищення дискретизації, завдяки якому ітеративно підвищує роздільність і дозволяє збільшувати зображення до 8 разів.

Її архітектура має дві гілки: одна для вилучення ознак, інша служить для реконструкції (рис. 2.16). На кожній ітерації LapSRN спочатку застосовує каскад згорткових шарів для одержання різноманітних карт функцій. Потім використовується транспонований згортковий шар для підвищення дискретизації карт функцій до вищого рівня. Шари розгортки у гілці реконструкції ініціалізуються за допомогою білінійного ядра, що є важливим для того, щоб змусити гілку вилучення ознак вивчати залишкові ознаки. Вихідні дані блоку розгортки також передаються на наступний етап, який використовується для вдосконалення виходу високої РЗ цього етапу та його масштабування до наступного рівня. Оскільки вихідні дані з нижчою роздільною здатністю використовуються для вдосконалення подальших етапів, застосовується спільне навчання, яке допомагає мережі працювати краще.

Оскільки кожен вхідний сигнал НР може мати кілька представлень ВР, функція втрат L_2 продукує згладжений вихід по всіх представленнях, через що зображення виходять нечіткими. Щоб впоратися з цим, використовується функція втрат Шарбоньє, яка може краще обробляти викиди.

Подібним чином DSRN використовує дві гілки для вилучення інформації в просторі низького та високого виміру відповідно, і постійно обмінюється інформацією для дальшого вдосконалення здатності до навчання.

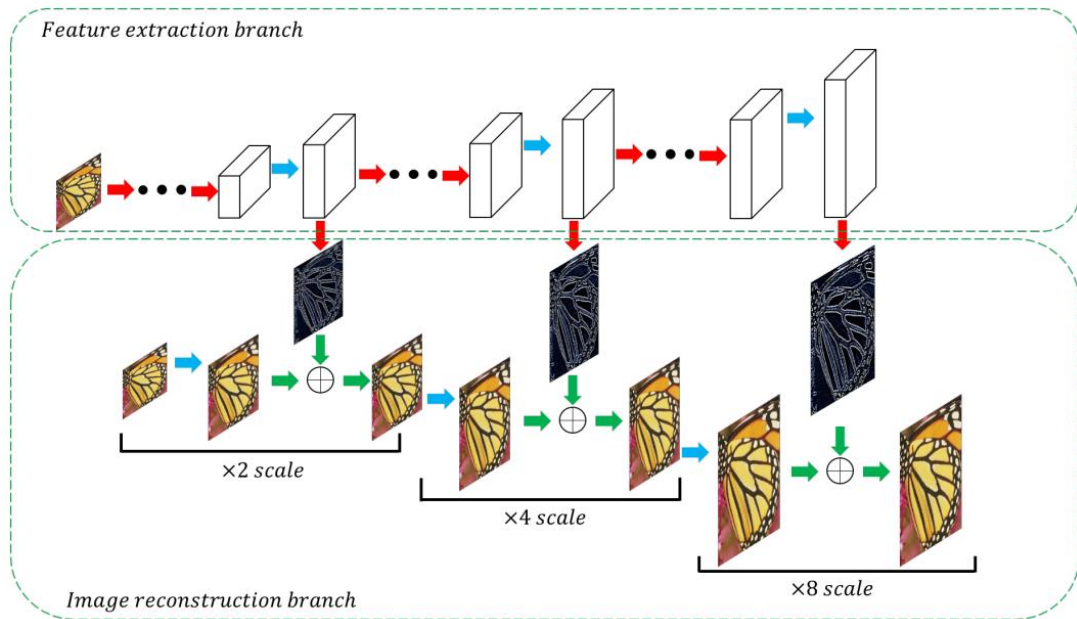


Рисунок 2.16 – Структура мережі LapSRN

Локальне розгалужене навчання. MSRN [58] вводить новий блок для багатомасштабного вилучення ознак (рис. 2.17). У цьому блоці два шари згортки з розміром ядра 3×3 і 5×5 застосовуються для одночасного вилучення ознак, потім виходи об'єднуються та знову проходять ті самі операції, і, нарешті, застосовується додаткова згортка 1×1 . Ярлик з'єднує вхід і вихід шляхом поелементного додавання. Завдяки такому локальному розгалуженому навчанню моделі SR можуть краще витягувати характеристики зображення з кількох масштабів і додатково покращувати продуктивність.

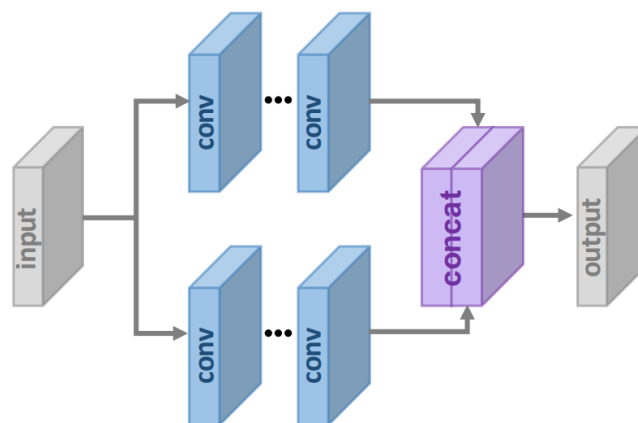


Рисунок 2.17 – Схема локальної розгалуженої мережі

Масштабо-специфічне розгалужене навчання. Враховуючи, що моделі SR для різних масштабів повинні пройти подібне виділення ознак, Лім та інші [37] пропонують масштабо-специфічне розгалужене навчання, яке дозволяє розв'язати задачу багатомасштабної SR за допомогою однієї мережі. Конкретніше кажучи, вони поширюють основні компоненти моделі (тобто проміжні рівні для виділення ознак) і приєднують масштабо-специфічні гілки попередньої обробки та шляхи підвищення дискретизації на початку та в кінці мережі відповідно. Під час навчання доступними для оновлення є лише гілки, що відповідають вибраному масштабові. Таким способом, запропонований MDSR значно зменшує розмір моделі завдяки спільному використанню більшості параметрів для різних масштабів і демонструє порівнянну з одномасштабними моделями продуктивність. Подібне розгалужене навчання, залежне од масштабу, також використовується в CARN [41] і ProSR [59].

2.3 Функції втрат у моделях SR

В галузі суперроздільності функції втрат використовуються для вимірювання похибки реконструкції та оптимізації моделі. Раніше дослідники використовували попіксельну втрату L2, але пізніше виявили, що вона не може досить точно виміряти якість реконструкції. Тому для кращого вимірювання похибки реконструкції та одержання більш реалістичних і якісних результатів застосовуються різноманітні функції втрат (наприклад, втрати вмісту, змагальні втрати). Сьогодні ці функції втрат відіграють важливу роль. У цьому розділі розглянуто функції втрат, які широко використовуються. В усіх рівняннях \hat{I} позначено цільове зображення високої роздільності, а I – реконструйоване зображення.

2.3.1 Піксельна втрата

Піксельна втрата вимірює попіксельну різницю між двома зображеннями та зазвичай включає втрату L1 (середня абсолютна похибка, MEA) і втрату L2 (середня квадратична похибка, MSE):

$$L_1(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} |\hat{I}_{i,j,k} - I_{i,j,k}|, \quad (2.2)$$

$$L_2(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} (\hat{I}_{i,j,k} - I_{i,j,k})^2, \quad (2.3)$$

де h , w та c – це висота, ширина та кількість каналів оцінюваних зображень.

Окрім того, є ще один варіант піксельної втрати L1, що зветься втратою Шарбоньє [57]:

$$L_{cha}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} \sqrt{(\hat{I}_{i,j,k} - I_{i,j,k})^2 + \epsilon^2}, \quad (2.4)$$

де ϵ – константа (напр. 10^{-3}) для числової стабільності.

Піксельна втрата обмежує згенероване зображення ВР \hat{I} , щоб значення його пікселів були досить близькими до істинного цільового зображення I . Порівняно з втратою L1, втрата L2 штрафуює більші помилки, але є більш терпимою до невеликих помилок, і тому часто призводить до надто розмитих результатів. На практиці втрата L1 показує ліпшу продуктивність і конвергенцію порівняно з втратою L2. Оскільки значення оцінки якості реконструкції PSNR сильно корелює з попіксельною різницею, а мінімізація втрати пікселів безпосередньо максимізує PSNR, піксельна втрата поступово стала найпоширенішою функцією втрат. Однак, оскільки втрата пікселів фактично не враховує якість зображення (наприклад, якість сприйняття та

текстури), результати часто не мають високочастотних деталей, через що вони є незадовільними через занадто гладкі текстури.

2.3.2 Втрата вмісту

Для того, щоб оцінити якість сприйняття зображень, в СР введено втрату вмісту [60]. Зокрема, вона вимірює семантичні відмінності між зображеннями за допомогою попередньо навченої мережі класифікації зображень. Позначаючи цю мережу як ϕ , а вилучені представлення високого рівня на l -му шарі як $\phi^{(l)}(I)$, втрата вмісту позначається як евклідова відстань між представленнями високого рівня двох зображень:

$$L_{Content}(\hat{I}, I; \phi, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} \left(\phi_{i,j,k}^{(l)}(\hat{I}) - \phi_{i,j,k}^{(l)}(I) \right)^2}, \quad (2.5)$$

де h_l , w_l та c_l – висота, ширина та кількість каналів представлення на шарі l .

По суті, втрата вмісту переносить отримані знання про ієрархічні функції зображення з класифікаційної мережі ϕ до мережі СР. На відміну од піксельної втрати, втрата вмісту спряє тому, щоб вихідне зображення \hat{I} було сприйнятливо схожим на цільове зображення I замість того, щоб змушувати їх точно збігатися за пікселями. Отже, він дає більш приємні візуально результати і також широко використовується в цій галузі, де VGG і ResNet є найбільш часто використовуваними попередньо навченими ЗНМ.

2.3.3 Текстурна втрата

З огляду на те, що реконструйоване зображення повинно мати той самий стиль (наприклад, кольори, текстури, контраст) із цільовим зображенням, в СР

введено текстурну втрату (так звану втрату реконструкції стилю) [61]. Текстура зображення розглядається як кореляція між різними каналами ознак і визначається як матриця Грема $G^{(l)} \in \mathbb{R}^{c_l \times c_l}$, де $G_{ij}^{(l)}$ – скалярний добуток між векторизованими картами функцій i та j шару l :

$$G_{ij}^{(l)}(I) = \text{vec}(\phi_i^{(l)}(I)) \cdot \text{vec}(\phi_j^{(l)}(I)), \quad (2.6)$$

де $\text{vec}(\cdot)$ – операція векторизації;

$\phi_i^{(l)}(I)$ – i -й канал карти функцій на шарі l зображення I .

Тоді текстурну втрату можна визначити як:

$$L_{\text{Texture}}(\hat{I}, I; \phi, l) = \frac{1}{c_l^2} \sqrt{\sum_{i,j,k} \left(G_{i,j}^{(l)}(\hat{I}) - G_{i,j}^{(l)}(I) \right)^2}. \quad (2.7)$$

Використовуючи текстурну втрату, EnhanceNet [62] створює набагато реалістичніші текстури та забезпечує візуально більш задовільні результати. Незважаючи на це, визначення розміру патчу для відповідності текстурам все ще є емпіричним. Занадто малі патчі призводять до артефактів у текстурованих областях, тоді як занадто великі патчі призводять до артефактів по всьому зображенню, оскільки статистика текстури усереднюється для областей із різними текстурями.

2.3.4 Змагальна втрата

В останні роки, завдяки потужній здатності до навчання, ГЗМ приділяють все більше уваги та впроваджують у різні завдання КЗ. Як було описано раніше, ГЗМ складається з генератора, що виконує генерування і дискримінатора, що приймає згенеровані результати та екземпляри з цільового

розподілу як вхідні дані та визначає, чи схожі вони між собою. Під час навчання по черзі виконуються два кроки: а) настроїти генератор і навчити дискриміратор краще розрізняти, б) настроїти дискриміратор і навчити генератор обманювати дискриміратор. Завдяки правильному ітеративному змагальному навчанню кінцевий генератор може виробляти результати, що відповідають розподілу реальних даних, тоді як дискриміратор не може відрізнити згенеровані дані од реальних.

Змагальне навчання вельми просто адаптувати до суперроздільності, у цьому випадку нам потрібно лише розглядати модель CP як генератор і визначити додатковий дискриміратор, який визначатиме, чи є вхідне зображення згенерованим, чи ні. Тому в SRGAN пропонується використовувати змагальну втрату на основі перехресної ентропії (CE) наступним чином:

$$L_{CE_Gen}(\hat{I}; D) = -\log D(\hat{I}), \quad (2.8)$$

$$L_{CE_Dis}(\hat{I}, I_s; D) = -\log D(I_s) - \log(1 - D(\hat{I})), \quad (2.9)$$

де L_{CE_Gen} та L_{CE_Dis} – змагальні втрати генератора (тобто CP моделі) та дискримінатора (тобто бінарного класифікатора);

I_s відбиває зображення, випадково відібране з цільового розподілу.

Крім того, Вонг та інші [59] і Юань та інші [63] використовують змагальну втрату на основі помилки найменшого квадрата для більш стабільного процесу навчання та вищої якості результатів:

$$L_{LS_Gen}(\hat{I}; D) = (D(\hat{I}))^2, \quad (2.10)$$

$$L_{LS_Dis}(\hat{I}, I_s; D) = (D(\hat{I}))^2 + (D(I_s) - 1)^2, \quad (2.11)$$

На відміну од наведених вище робіт, які зосереджуються на конкретних формах змагальності, в роботі [55] стверджують, що дискримінатор рівня пікселів викликає генерування високочастотного шуму, і пропонують інший дискримінатор рівня ознак, щоб працювати з представленнями високого рівня, одержаними попередньо навченою ЗНМ, що захоплює більш значущі атрибути реальних зображень ВР. Сю та інші [64] представили багатокласову ГЗМ, що складається з генератора та кількох специфічних для класу дискримінаторів. ESRGAN використовує відносний GAN, щоб передбачити ймовірність того, що справжні зображення є відносно більш реалістичними, ніж фальшиві, замість ймовірності того, що вхідні зображення є справжніми чи фальшивими, і, таким способом, відновити детальніші текстури.

Розширений тест MOS показує, що навіть якщо моделі СР, навчені зі змагальною втратою та втратою вмісту, досягають нижчого PSNR порівняно з моделями, навченими з піксельною втратою, вони приносять значний приріст у якості сприйняття. По суті, дискримінатор виділяє деякі приховані шаблони реальних зображень ВР і приводить згенеровані зображення ВР до відповідності, таким чином допомагаючи створювати реалістичніші зображення. Проте, на теперішній час, навчальний процес ГЗМ залишається складним і нестабільним. Хоча було проведено деякі дослідження щодо того, як стабілізувати навчання ГЗМ, проте все ще лишається проблемою те, як забезпечити коректне навчання і активну роль ГЗМ, інтегрованих у моделі СР [65].

2.3.5 Априорна втрата

На додаток до вищевказаних функцій втрат, для обмеження генерування також застосовуються зовнішні попередні знання. Зокрема, робота [66] зосереджується на СР зображень облич і запроваджує мережу FAN, щоб обмежити узгодженість ключових точок обличчя. FAN попередньо навчається

та інтегрується для надання попередньої інформації про вирівнювання обличчя, а потім навчається спільно з СР. Таким чином, запропонований Super-FAN покращує продуктивність як для НР вирівнювання обличчя, так і для СР зображень обличчя.

По суті, втрата вмісту та втрата текстури, які вводять мережу класифікації, забезпечують попередні знання про ієрархічні ознаки зображення для СР. Вводячи більше попередніх знань, продуктивність СР може бути додатково покращена.

3 ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТИ НЕЙРОМЕРЕЖНИХ МЕТОДІВ ПОКРАЩЕННЯ ЯКОСТИ ЗОБРАЖЕННЯ

3.1 Метрики оцінювання якості реконструкції

У зв'язку з широким використанням методів СР зображень, оцінка якості реконструйованих зображень стає все більш важливою. Якість зображення відноситься до візуальних атрибутів зображень і зосереджується на оцінці сприйняття глядачів [67]. Загалом методи оцінки якості зображення (ОЯЗ) включають суб'єктивні методи, засновані на сприйнятті людини (тобто наскільки реалістичним виглядає зображення), і об'єктивні обчислювальні методи. Перші більше відповідають нашим потребам, але часто дорогі й займають багато часу, тому другі наразі є основними. Однак ці два типи методів не обов'язково узгоджуються між собою, оскільки об'єктивні методи часто не в змозі дуже точно відбити зорове сприйняття людини, що може призвести до великої різниці в результатах ОЯЗ [20]. В цьому розділі представлено декілька найбільш часто використовуваних методів ОЯЗ, як об'єктивних, так і суб'єктивних.

3.1.1 Пікове співвідношення сигнал/шум (PSNR)

Пікове відношення сигнал/шум (PSNR) є одним із найпопулярніших способів вимірювання якості реконструкції перетворень з утратами (наприклад, стиснення або відновлення зображення). Для СР зображень PSNR визначається через максимальне значення пікселя (позначене як L) і середню квадратичну помилку (MSE) між зображеннями. Нехай дано цільове зображення BP I з N пікселями та реконструйоване зображення \hat{I} , тоді PSNR між I та \hat{I} можна визначити таким чином:

$$\text{PSNR} = 10 \log_{10} \left(\frac{L^2}{\text{MSE}} \right), \quad (3.1)$$

де L – максимальне значення яскравості пікселя зображення (для 8-бітних зображень $L = 255$).

Оскільки PSNR залежить лише од MSE, дбаючи тільки про відмінності між відповідними пікселями, а не про візуальне сприйняття, це часто призводить до низької продуктивності в представленні якості реконструкції в реальних прикладах, оскільки нам зазвичай важливіше людське сприйняття. Однак через необхідність порівнювати з літературними роботами та відсутність повністю точних метрик сприйняття, PSNR все ще є найбільш широко використовуваним критерієм оцінки для моделей СР.

3.1.2 Структурна подібність (SSIM)

Враховуючи, що зорова система людини (ЗСЛ) добре адаптована для виділення структур зображення, було запропоновано індекс структурної подібності (SSIM) [68], що вимірює структурну подібність між зображеннями на основі незалежних порівнянь яскравості, контрасту і структури. Для зображення I з N пікселями яскравість μ_I і контраст σ_I обчислюються як середнє значення і стандартне відхилення інтенсивності зображення відповідно. А порівняння яскравості та контрасту, позначених як $C_l(I, \hat{I})$ та $C_c(I, \hat{I})$ відповідно, визначають за формулами:

$$C_l(I, \hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1}, \quad (3.2)$$

$$C_c(I, \hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2}, \quad (3.3)$$

де $C_1 = (k_1L)^2$ і $C_2 = (k_2L)^2$ – константи для уникнення нестабільності, $k_1 \ll 1$ та $k_2 \ll 1$.

Крім того, структура зображення представлена нормалізованими значеннями пікселів (тобто $(I - \mu_I)/\sigma_I$), чії кореляції (тобто внутрішній добуток) вимірюють структурну подібність, еквівалентну коефіцієнту кореляції між I та \hat{I} . Таким робом, функція порівняння структури $C_s(I, \hat{I})$ визначається як:

$$\sigma_{I\hat{I}} = \frac{1}{N-1} \sum_{i=1}^N (I(i) - \mu_I)(\hat{I}(i) - \mu_{\hat{I}}), \quad (3.4)$$

$$C_s(I, \hat{I}) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I \sigma_{\hat{I}} C_3}, \quad (3.5)$$

де $\sigma_{I, \hat{I}}$ – коваріація між I та \hat{I} ;

C_3 – константа для стабільності.

Нарешті, SSIM визначається як:

$$\text{SSIM}(I, \hat{I}) = [C_l(I, \hat{I})]^\alpha [C_c(I, \hat{I})]^\beta [C_s(I, \hat{I})]^\gamma, \quad (3.6)$$

де α, β, γ – контрольні параметри для регулювання відносної важливості.

Оскільки SSIM оцінює якість реконструкції з погляду ЗСЛ, він краще відповідає вимогам оцінки сприйняття [69] і також широко використовується.

3.1.3 Оцінка на основі усередненої думки (MOS)

Оцінка на основі усередненої думки (MOS) – це широко використовуваний суб'єктивний метод ОЯЗ, коли живих людей просять дати перцептивну оцінку якості для зображень, які тестують. Як правило, оцінки

од 1 (погано) до 5 (добре). Остаточний MOS розраховується як середнє арифметичне всіх оцінок.

Хоча MOS-тестування виглядає слухним методом ОЯЗ, йому властиві вади, такі як нелінійне сприйняття масштабів, упередженість та розбіжність критеріїв оцінювання. Далєбі існують деякі моделі СР, які погано роблять із загальноприйнятими показниками ОЯЗ (як PSNR), але значно перевершують інших з погляду якості сприйняття, і в цьому випадку MOS-тестування є найнадійнішим методом ОЯЗ для вимірювання якості сприйняття.

3.1.4 Перцептивна якість заснована на навчанні

Щоб лїпше оцінити перцептивну якість зображення, зменшуючи ручне втручання, дослідники намагаються оцінити перцептивну якість, вивчаючи великі набори даних. Зокрема, Ма та інші [70] і Талєбі та інші [71] пропонують метрики Ма та NIMA, що навчаються на оцінках візуального сприйняття та прямо передбачають показники якості без дослідження цїльових зображень ВР.

Незважаючи на те, що ці методи демонструють кращу ефективність у наслідуванні людського зорового сприйняття, питання про те, яка якість сприйняття нам потрібна (наприклад, реалістичніші зображення чи відповідність оригінальному зображенню), лишається питанням для досліджень, тому об'єктивні методи ОЯЗ (PSNR, SSIM) все ще є популярними.

3.1.5 Інші методи ОЯЗ

Окрім вже розглянутих методів ОЯЗ, існують інші менш популярні метрики СР. Багатомасштабна структурна подібність (MS-SSIM) забезпечує більшу гнучкість, ніж одномасштабна SSIM, у врахуванні варіацій умов

перегляду. Подібність ознак (FSIM) [72] виділяє ключові точки людського інтересу, на основі фазової конгруентності і величини градієнта зображення для оцінки якості зображення. Оцінювач якості природнього зображення (NIQE) [73] використовує вимірні відхилення од статистичних закономірностей, що спостерігаються в природніх зображеннях, без впливу на спотворені зображення.

Нещодавно було математично доведено [74], що спотворення (наприклад, PSNR, SSIM) і якість сприйняття (наприклад, MOS) суперечать одне одному, і показано, що зі зменшенням спотворення якість сприйняття повинна бути гіршою. Отже, як точно виміряти якість СР, все ще залишається актуальною проблемою, яку потрібно вирішити.

3.2 Набори даних для СР

Глибоке навчання – це метод навчання на даних, і якість даних відіграє важливу роль у результатах навчання моделі, оскільки високоякісні дані можуть посприяти покращенню продуктивності реконструкції глибоких моделей СР [75]. За останні роки було запропоновано багато різноманітних наборів даних для навчання та тестування систем СР, які значно відрізняються за кількістю зображень, якістю, роздільною здатністю, типом зображень тощо. Деякі з них містять пари зображень НР-ВР, тоді як інші містять лише зображення НР, які доводиться попередньо оброблювати.

Набори даних, які зазвичай використовуються для навчання моделі СР, включають BSDS300, BSDS500, DIV2K і Flickr2K. Крім того, існує багато випробувальних наборів даних, які можна використовувати для ефективного тестування продуктивності моделей, таких як Set5, Set14, Urban100 і Manga109.

Серед цих наборів даних DIV2K є найпоширенішим високоякісним набором даних для навчання моделі, який містить 800 навчальних зображень,

100 затверджувальних зображень та 100 випробувальних зображень. Flickr2k – це великий набір даних, що містить 2650 2К-зображень із сайту Flickr. Set5 і Set14 – це класичні випробувальні набори даних для завдань СР, що можуть точно відбивати продуктивність моделі. RealSR – це перший справді зібраний набір даних SISR із парними зображеннями низької та високої РЗ. Докладний опис найпопулярніших наборів даних, що застосовуються для навчання та випробування моделей СР, показано в таблиці 3.1.

На додаток до перелічених наборів даних, деякі набори даних, які широко використовуються в інших галузях КЗ, також використовуються як додаткові навчальні набори даних для СР, серед них: ImageNet, MS-COCO, VOC2012 і CelebA. Крім того, також широко застосовується об'єднання кількох наборів даних (наприклад, поєднання T91 і BSDS300, поєднання DIV2K і Flickr2K) для навчання і дальшого поліпшення продуктивності моделі.

Таблиця 3.1 – Опис наборів даних для моделей СР

Назва	Кількість	Серед. РЗ	Формат	Опис
1	2	3	4	5
BSDS300	300	435×367	JPG	Загальні зображення
BSDS500	500	432×370	JPG	Загальні зображення
Set5	5	313×336	PNG	Загальні зображення
Set14	14	492×446	PNG	Загальні зображення
DIV2K	1000	1972×1437	PNG	Високоякісний датасет загальних зображень
Flickr2K	2650	2048×2048	PNG	2К-зображення з сайту Flickr
Manga109	109	826×1169	PNG	Зображення японської манги
Urban100	100	984×797	PNG	Зображення міської архітектури та конструкцій
BSDS100	100	432×370	JPG	Піднабір з набору BSDS500 для випробування

Продовження таблиці 3.1

1	2	3	4	5
BSDS200	200	435×367	JPG	Піднабір з набору BSDS500 для випробування
T91	91	264×204	PNG	Загальні зображення
PIRM	200	617×482	PNG	Загальні зображення
ImageNet	>3.2 млн	-	JPG	Один з найбільших наборів даних для завдань КЗ

Через специфіку завдання СР важко створити великомасштабний реальний парний набір даних. Тому дослідники часто застосовують шаблони деградації до вищезгаданих наборів даних, щоб одержати відповідні деградовані зображення для створення парних наборів даних. Найпоширеніший метод деградації можна виразити таким рівнянням:

$$\mathcal{D}(I; \theta_{\mathcal{D}}) = (I \otimes \kappa) \downarrow_s + n, \quad (3.7)$$

де $I \otimes \kappa$ – згортка між ядром розмиття κ та зображенням I ;

\downarrow_s – операція зменшення дискретизації з коефіцієнтом масштабу s ;

n – зазвичай додатковий білий шум Гауса зі стандартним відхиленням σ .

Однак, зображення в реальному світі легко спотворюються різними чинниками (наприклад, шумом датчика, розмиттям руху та артефактами стиснення), що призводить до того, що справжні зображення є складнішими, ніж змодельовані. Щоб полегшити ці проблеми та навчити ефективнішу загальну модель СР, деякі дослідники моделюють метод деградації, комбінуючи кілька операцій.

На основі (3.7) було запропоновано три найбільш широко використовуваних методи деградації: ВІ, ВD і DN. Серед них ВІ є найпоширенішим режимом деградації для імітації НР-зображень, який, по суті, є бікубічною операцією зменшення дискретизації. Для ВD зображення ВР розмиваються ядром Гауса розміром 7×7 зі стандартним відхиленням 1,6, а

потім зменшуються із коефіцієнтом масштабування 3. Щоб одержати НР-зображення методом DN, на зображенні ВР виконується бікубічне зниження дискретизації з коефіцієнтом масштабування $\times 3$, а потім до зображення додається шум Гауса з рівнем шуму = 30.

3.3 Порівняння ефективності мереж СР

В таблиці 3.2 наведено порівняння популярних моделей СР за метриками PSNR та SSIM на чотирьох найпоширеніших випробувальних наборах даних з коефіцієнтом масштабування $\times 4$, із зазначенням кількості параметрів мережі (цей показник вказує на пам'ять, яку споживає мережа під час роботи). Моделі відсортовані за показником якості для набору даних Set14 од найгіршого до найліпшого.

Таблиця 3.2 – Порівняння моделей СР

Модель	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Кількість параметрів
1	2	3	4	5	6
SRGAN	29,40/0,8472	26,02/0,7397	25,16/0,6688	-	1,6 М
SRCNN	30,48/0,8628	27,50/0,7513	26,90/0,7101	24,52/0,7221	57 К
FSRCNN	30,71/0,8657	27,59/0,7535	-	24,62/0,7280	13 К
ESPCN	30,90/0,8646	27,73/0,7562	-	24,60/0,7360	20 К
ZSSR	31,13/0,8796	28,01/0,7651	27,12/0,7211	-	-
VDSR	31,35/0,8838	28,01/0,7674	27,29/0,7260	25,18/0,7540	665 К
DRCN	31,53/0,8854	28,02/0,7670	27,23/0,7233	25,14/0,7510	1,77 М
LapSRN	31,54/0,8855	28,19/0,7720	27,32/0,7280	25,21/0,7560	812 К
DRRN	31,68/0,8888	28,21/0,7720	27,38/0,7284	25,44/0,7638	297 К
IDN	31,82/0,8903	28,25/0,7730	27,41/0,7297	25,41/0,7632	-
MemNet	31,74/0,8893	28,26/0,7723	27,40/0,7281	25,50/0,7630	677 К
DBCN	31,78/0,8885	28,27/0,7733	27,40/0,7291	25,55/0,7656	-
NLRN	31,92/0,8916	28,36/0,7745	27,78 0,7444	25,79/0,7729	330 К
EnhanceNet	31,74/0,8869	28,42/0,7774	27,50/0,7326	25,66/0,7703	-
SRResNet	32,05/0,8910	28,49/0,7800	27,58/0,7620	-	1,5 М
SRDenseNet	32,02/0,8934	28,50/0,7782	-	26,05/0,7819	2 М
MSRN	32,07/0,8903	28,60/0,7751	27,52/0,7273	26,04/0,7896	6,5 М
CARN	32,13/0,8937	28,60/0,7806	27,58/0,7349	26,07/0,7837	1,6 М
ESRT	32,19/0,8947	28,69/0,7833	27,69/0,7379	26,39/0,7962	751К
MDSR	32,50/0,8973	28,72/0,7857	27,72/0,7418	26,67/0,8041	8 М
ERN	32,39/0,8975	28,75/0,7853	27,70/0,7398	26,43/0,7966	-
EDSR	32,46/0,8968	28,80/0,7876	27,71/0,7420	26,64/0,8033	43 М

Продовження таблиці 3.2

1	2	3	4	5	6
DBPN	32,47/0,8980	28,82/0,7860	27,72/0,7400	26,38/0,7946	10 М
RCAN	32,63/0,9002	28,87/0,7889	27,77/0,7436	26,82/0,8087	16 М
HAN	32,64/0,9002	28,90/0,7890	27,80/0,7442	26,85/0,8094	16,1 М
SAN	32,64/0,9003	28,92/0,7888	27,78/0,7436	26,79/0,8068	15,7 М
RDN	32,61/0,9003	28,92/0,7893	27,80/0,7434	26,61/0,8028	22,6 М
ESRGAN	32,73/0,9011	28,99/0,7917	27,85/0,7455	27,03/0,8153	-
SwinIR	32,92/0,9044	29,09/0,7950	27,92/0,7489	27,45/0,8254	11,8 М

Аналізуючи таблицю 3.2 можна зробити такі висновки:

- жоден з підходів не забезпечує явно ліпших результатів, якщо оцінювати об'єктивними метриками;
- легкі моделі (з кількістю параметрів менше 1 мільйона) наразі перебувають в меншості й показують об'єктивно гірші результати, ніж важкі моделі з великою кількістю параметрів;
- SRGAN показує найгірший результат за об'єктивними метриками, проте його покращена версія, ESRGAN, перебуває майже на найвищій позиції, при тому, що моделі ГЗМ сконцентровані на перцептивній якості реконструкції, яку важко оцінити сучасними об'єктивними метриками;
- найвищі показники має мережа SwinIR [76] – новітня модель, що використовує популярну мережу Swin Transformer, яка поєднує в собі переваги ЗНМ і мереж-трансформерів. Інша мережа, що ґрунтується на використанні трансформерів, – ESRT [77], що також має вельми високі показники, маючи при цьому лише 751 тис. параметрів, що забезпечує значну обчислювальну продуктивність.

3.4 Сучасні виклики та напрямки дальших досліджень у сфері СР

Розглянуті раніше моделі досягли багатонадійних результатів і значною посприяли розвитку СР. Не дивлячись на це, в завданні СР все ще є багато

невирішених питань. У цьому розділі розглянуто деякі виклики й підсумовано деякі перспективні тенденції та майбутні напрямки розвитку.

3.4.1 Мережеві архітектури

Хороший дизайн мережі не тільки збільшує гіпотетичну межу якості роботи, але й допомагає ефективно вивчати представлення без надмірних просторових та обчислювальних витрат. Нижче наведено декілька перспективних напрямків удосконалення мереж.

Поєднання локальної та глобальної інформації. Велике перцептивне поле надає більше контекстної інформації та допомагає одержати більш реалістичні результати. Для СР зображень використання локальної інформації робить текстуру зображення більш реалістичною, а використання глобальної інформації робить вміст зображення більш відповідним контексту. Отже, перспективним є поєднання локальної та глобальної інформації для надання контекстної інформації різного масштабу для СР зображень.

Поєднання інформації низького та високого рівнів. Неглибокі шари в ЗНМ мають схильність витягувати низькорівневі ознаки, такі як кольори та границі, тоді як більш глибокі шари вивчають представлення вищого рівня, такі як тип об'єктів. Поєднання ознак низького рівня з семантикою високого рівня може бути дуже корисним для реконструкції зображень ВР.

Контекстно-специфічна увага. У різних контекстах люди схильні звертати увагу на різні аспекти зображень. Наприклад, дивлячись на траву людей може більше цікавити кольори та текстура, тоді як розглядаючи тіло тварини людина більше звертає увагу на вид і деталі волосся. Включення механізму уваги для посилення уваги до ключових ознак полегшує створення реалістичних деталей.

Легші та ефективніші архітектури. Наявні моделі СР вже показують досить хороші результати реконструкції, проте, як правило, ігнорують розмір

моделі та швидкість роботи. Деякі сучасні моделі СР навіть на найпотужніших графічних процесорах витрачають десятки секунд на оброблення зображення. Такий довгий час прогнозування є неприйнятним у практичних застосуваннях і повністю забирає можливість використання систем СР на периферійних пристроях (смартфонах, камерах тощо), тому ефективніші архітектури є доконечними. Як зменшити розмір моделі та прискорити прогнозування, зберігаючи продуктивність, залишається проблемою. Цього, зокрема, можливо добитися такими методами, як бінаризація та квантування мережі.

Методи підвищення дискретизації. Наявні методи підвищення дискретизації (розділ 2.1) мають більші чи менші вади: методи інтерполяції призводять до дорогих обчислень і не можуть бути наскрізно навчені, транспонована згортка створює артефакти шахової дошки, субпіксельний шар приносить нерівномірний розподіл рецептивних полів, а модуль метарозширення може спричинити нестабільність або неефективність і має додаткові можливості для вдосконалення. Як виконати ефективне підвищення дискретизації, особливо з високими коефіцієнтами масштабування, ще потрібно вирішити.

Останнім часом стає дедалі популярнішою техніка пошуку нейронної архітектури (ПНА), що значно покращує продуктивність та ефективність з невеликим штучним втручанням. Для галузі СР поєднання досліджень вищезазначених напрямків із ПНА має великий потенціал.

3.4.2 Стратегії навчання

Окрім грамотної архітектури, для досягнення задовільних результатів також потрібні надійні стратегії навчання [78]. Далі представлено кілька перспективних напрямків розвитку стратегій навчання.

Функції втрат. У минулому більшість моделей СР поклалися на втрати L1 або MSE. Хоча було запропоновано інші, новіші функції втрат, такі

як втрата вмісту, втрата текстури та змагальна втрата, вони все ще не можуть досягти хорошого балансу між точністю реконструкції та якістю сприйняття. На практиці ці функції втрат часто комбінуються, і найкраща функція втрат для СР все ще невідома. Тому дослідження нових функцій втрат, які можуть досягти компромісу між сприйняттям і попиксельною схожістю, залишається важливою темою дослідження.

Нормалізація. Хоча групова нормалізація широко використовується в завданнях КЗ, що значно прискорює навчання та покращує продуктивність, доведено, що він є неоптимальним для СР. Таким робом, конечно вивчити інші ефективні методи нормалізації СР.

3.4.3 Метрики оцінювання якості

Показники оцінки є одним із найважливіших компонентів машинного навчання. Якщо продуктивність неможливо точно виміряти, дослідникам буде важко покращити результат. На сьогодні, PSNR і SSIM є найпопулярнішими метриками для СР. Однак PSNR, як правило, призводить до надмірного згладжування, і результати можуть сильно відрізнятися між майже нерозрізненими зображеннями. SSIM виконує оцінку з погляду яскравості, контрастності та структури, та все ще не може точно виміряти перцептивну якість. Крім того, MOS є найбільш наближеним до зорової реакції людини, але потребує значних зусиль, є абсолютно суб'єктивним і невідтворюваним. Хоча дослідники запропонували різні метрики (підрозділ 3.1), наразі не існує уніфікованих і визнаних метрик оцінки якості СР. А значить негайно потрібні точніші метрики для оцінки якості реконструкції.

3.4.4 СР із довільним коефіцієнтом розширення

SISR має застосування в різних сценаріях реального життя та користувачів, тому необхідно розробити гнучку та універсальну багатомасштабну модель СР, яка може бути використана із будь-яким масштабом, включаючи асиметричні та нецілі масштабні коефіцієнти. Наразі більшість моделей СР на основі ГН можна застосовувати лише до одного чи обмеженої кількості коефіцієнтів підвищення дискретизації і незважаючи на те, що також було запропоновано кілька довільних методів СР із довільним масштабом (підпункт 2.1.4), їм, як правило, не вистачає гнучкості у використанні та простоти реалізації, що значно обмежує сценарії їх застосування. Отже, створення такої ж простої та гнучкої багатомасштабної моделі СР на основі ЗНМ, як бікубічна інтерполяція, має вирішальне значення для поширення технології СР.

3.4.5 Взаємодія СР із завданнями КЗ високого рівня

Завдання КЗ високого рівня (наприклад, класифікація зображень, сегментація зображень та аналіз зображень) сильно залежать од якості вхідного зображення, тому технологія СР зазвичай використовується для попередньої обробки. Водночас якість реконструйованих СР зображень значною мірою впливає на точність цих завдань. Тому є рація використовувати точність завдань КЗ високого рівня, як показник для оцінювання якості СР зображень. Отже, варто досліджувати моделі СР, які можуть взаємодіяти із завданнями КЗ високого рівня, таким чином СР та інші завдання можуть помагати й навчати одне одного.

3.4.6 SR в реальних сценаріях застосування

Як зазначено в підрозділі 3.2, вкрай важко зібрати зображення з різною РЗ однієї сцени, тому для побудови наборів даних SR широко використовується бікубічна інтерполяція. Однак моделі SR, навчені на цих наборах даних, можуть вивчати лише зворотний процес попередньо визначеної деградації, а зображення в реальному світі зазнають дуже складних спотворень, таких як розмивання, додатковий шум і артефакти стиснення. Через це моделі, навчені на штучних наборах даних, зазвичай погано роблять у сценах реального світу. Отже, перспективним напрямком для майбутнього розвитку є навчання моделей SR без нагляду (тобто навчання на наборах даних без парних зображень HR-VR), а також створення нових якісних реальних наборів даних.

ВИСНОВКИ

У рамках кваліфікаційної роботи було досліджено нейромережний підхід до покращення якості зображення.

Було виконано всі поставлені завдання, а саме:

- досліджено модулі підвищення дискретизації, що застосовуються в моделях СР, визначено їхні переваги та вади, розглянуто проблеми, пов'язані з підвищенням дискретизації зображень;

- досліджено основні сучасні підходи до проєктування глибоких нейромереж СР, їхні переваги та вади, розглянуто основні нейромережі, що ґрунтуються на кожному з підходів;

- проаналізовано функції втрат, що використовуються для навчання моделей СР, визначено п'ять основних типів втрат, доцільність застосування кожної з них;

- проаналізовано найпоширеніші метрики оцінювання якості реконструйованих зображень, визначено проблеми, пов'язані з ними;

- досліджено наявні набори даних, що використовуються в навчанні та випробуванні моделей СР, означено проблему створення якісних наборів для завдання СР;

- розглянуто продуктивність роботи сучасних моделей СР, а саме якість реконструкції зображень за двома метриками та кількість параметрів мережі, та проаналізовано ці результати;

- визначено майбутні виклики та напрямки дальших досліджень у сфері СР зображень, які виходять з досліджених підходів та методів в досліджуваній області;

Мету роботи досягнуто.

Результати роботи апробовано у вигляді тез доповідей під час Міжнародної наукової та практичної конференції «MODERN RESEARCH IN WORLD SCIENCE» [79].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Anwar, S., Khan, S., & Barnes, N. (2020). A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3), 1-34.
2. Anwar, S., Khan, S., & Barnes, N. (2020). A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, 53(3), 1-34.
3. Li, J., Pei, Z., & Zeng, T. (2021). From beginner to master: A survey for deep learning-based single-image super-resolution. *arXiv preprint arXiv:2109.14335*.
4. Dahl, R., Norouzi, M., & Shlens, J. (2017). Pixel recursive super resolution. In *Proceedings of the IEEE international conference on computer vision* (pp. 5439-5448).
5. Sun, N., & Li, H. (2019). Super resolution reconstruction of images based on interpolation and full convolutional neural network and application in medical fields. *IEEE Access*, 7, 186470-186479.
6. Kumar, G., & Singh, K. (2013). Image super resolution on the basis of DWT and bicubic interpolation. *International Journal of Computer Applications*, 65(15).
7. Sun, J., Xu, Z., & Shum, H. Y. (2008, June). Image super-resolution using gradient profile prior. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.
8. Bodyanskiy, Y. V., Shafronenko, A., & Rudenko, D. (2019). Online Neuro Fuzzy Clustering of Data with Omissions and Outliers based on Completion Strategy. In *CMIS* (pp. 18-27).
9. Lyashenko, V., & Rudenko, D. (2021). Modeling Deformation of Spur Gear.
10. Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In *European conference on computer vision* (pp. 184-199). Springer, Cham.

11. Постніков А.В. Руденко Д.О. (2021). Аналіз загальної класифікації та структури в event - індустрії. In TOPICAL ISSUES OF MODERN SCIENCE, SOCIETY AND EDUCATION (pp. 892-897).
12. Zeleniy, O., Rudenko, D., Lyubchenko, V., & Lyashenko, V. (2022). Image Processing as an Analysis Tool in Medical Research.
13. Li, Y., Sixou, B., & Peyrin, F. (2021). A review of the deep learning methods for medical images super resolution problems. *Irbm*, 42(2), 120-133.
14. Hung, S. C., Wu, H. C., & Tseng, M. H. (2021). Integrating Image Quality Enhancement Methods and Deep Learning Techniques for Remote Sensing Scene Classification. *Applied Sciences*, 11(24), 11659.
15. Willett, R., Jermyn, I. H., Nowak, R., & Zerubia, J. (2003). Wavelet-based superresolution in astronomy. *Astronomical Society of the Pacific*.
16. Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., & Zhang, L. (2016). Image super-resolution: The techniques, applications, and future. *Signal Processing*, 128, 389-408.
17. Колосок Е.В. Руденко Д.О. (2021). Огляд можливостей Google Analytics для роботи з даними. In Results of modern scientific research and development (pp. 162-165).
18. Филатов, В. А., & Руденко, Д. А. (2013). Формальный подход к описанию свойств данных в информационных системах.
19. Boost your image quality to the max with Super Resolution. URL: <https://www.adobe.com/products/photoshop-lightroom/super-resolution.html> (дата звернення 12.11.2022).
20. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).
21. Mao, X. J., Shen, C., & Yang, Y. B. (2016). Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint arXiv:1606.08921.

22. Dahl, R., Norouzi, M., & Shlens, J. (2017). Pixel recursive super resolution. In Proceedings of the IEEE international conference on computer vision (pp. 5439-5448).
23. Ooi, Y. K., & Ibrahim, H. (2021). Deep learning algorithms for single image super-resolution: a systematic review. *Electronics*, 10(7), 867.
24. Zang, H., Zhao, Y., Niu, C., Zhang, H., & Zhan, S. (2022). Attention Network with Information Distillation for Super-Resolution. *Entropy*, 24(9), 1226.
25. Cheng, X., Fu, Z., & Yang, J. (2020, August). Zero-shot image super-resolution with depth guided internal degradation learning. In European Conference on Computer Vision (pp. 265-280). Springer, Cham.
26. Sun, Y., Li, Y., & Zhang, Y. (2022). Super-Resolution Using Residual Channel Similar Feature Attention Network. Available at SSRN 4084707.
27. Nguyen, K., Fookes, C., Sridharan, S., Tistarelli, M., & Nixon, M. (2018). Super-resolution for biometrics: A comprehensive survey. *Pattern Recognition*, 78, 23-42.
28. Duchon, C. E. (1979). Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology and Climatology*, 18(8), 1016-1022.
29. Руденко, Д. О. (2017). Метод семантичної інтеграції локально незалежних даних (Doctoral dissertation, ДРУКАРНЯ МАДРИД).
30. Zeiler, M. D., Krishnan, D., Taylor, G. W., & Fergus, R. (2010, June). Deconvolutional networks. In 2010 IEEE Computer Society Conference on computer vision and pattern recognition (pp. 2528-2535). IEEE.
31. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).
32. Hu, X., Mu, H., Zhang, X., Wang, Z., Tan, T., & Sun, J. (2019). Meta-SR: A magnification-arbitrary network for super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1575-1584).

33. Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In European conference on computer vision (pp. 391-407). Springer, Cham.
34. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
35. Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).
36. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).
37. Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).
38. Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637-1645).
39. Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3147-3155).
40. Tai, Y., Yang, J., Liu, X., & Xu, C. (2017). Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE international conference on computer vision (pp. 4539-4547).
41. Ahn, N., Kang, B., & Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European conference on computer vision (ECCV) (pp. 252-268).

42. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
43. Tong, T., Li, G., Liu, X., & Gao, Q. (2017). Image super-resolution using dense skip connections. In Proceedings of the IEEE international conference on computer vision (pp. 4799-4807).
44. Xu, J., Chae, Y., Stenger, B., & Datta, A. (2018, October). Dense bynet: Residual dense network for image super resolution. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 71-75). IEEE.
45. Haris, M., Shakhnarovich, G., & Ukita, N. (2020). Deep back-project networks for single image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4323-4337.
46. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).
47. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 286-301).
48. Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., ... & Shen, H. (2020, August). Single image super-resolution via a holistic attention network. In European conference on computer vision (pp. 191-207). Springer, Cham.
49. Dai, T., Cai, J., Zhang, Y., Xia, S. T., & Zhang, L. (2019). Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11065-11074).
50. Liu, D., Wen, B., Fan, Y., Loy, C. C., & Huang, T. S. (2018). Non-local recurrent network for image restoration. *Advances in neural information processing systems*, 31.
51. Zhang, Y., Li, K., Li, K., Zhong, B., & Fu, Y. (2019). Residual non-local attention networks for image restoration. arXiv preprint arXiv:1903.10082.

52. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
53. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
54. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).
55. Park, S. J., Son, H., Cho, S., Hong, K. S., & Lee, S. (2018). Srfeat: Single image super-resolution with feature discrimination. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 439-455).
56. Rudenko, D., Serhienko, O., Zeleniy, O., & Lyashenko, V. (2022). Model for Predictive Analysis of International Trade Based on the Dynamics of Stock Indices (Example of Data from the USA, Canada and UK).
57. Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 624-632).
58. Li, J., Fang, F., Mei, K., & Zhang, G. (2018). Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 517-532).
59. Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., & Schroers, C. (2018). A fully progressive approach to single-image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 864-873).
60. Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (pp. 694-711). Springer, Cham.

61. Gatys, L., Ecker, A. S., & Bethge, M. (2015). Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28.

62. Sajjadi, M. S., Scholkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE international conference on computer vision* (pp. 4491-4500).

63. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., & Lin, L. (2018). Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 701-710).

64. Xu, X., Sun, D., Pan, J., Zhang, Y., Pfister, H., & Yang, M. H. (2017). Learning to super-resolve blurry face and text images. In *Proceedings of the IEEE international conference on computer vision* (pp. 251-260).

65. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In *International conference on machine learning* (pp. 214-223). PMLR.

66. Bulat, A., & Tzimiropoulos, G. (2018). Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 109-117).

67. Руденко, Д. А., & Филатов, В. А. (2013). Формальный подход к описанию свойств данных в информационных системах. *Вестник Херсонского национального технического университета*, (1 (46)), 146-149.

68. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612.

69. Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1), 98-117.

70. Ma, C., Yang, C. Y., Yang, X., & Yang, M. H. (2017). Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158, 1-16.

71. Talebi, H., & Milanfar, P. (2018). NIMA: Neural image assessment. *IEEE transactions on image processing*, 27(8), 3998-4011.

72. Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8), 2378-2386.

73. Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3), 209-212.

74. Blau, Y., & Michaeli, T. (2018). The perception-distortion tradeoff. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6228-6237).

75. Таняньський О.С., Руденко Д.О. (2021). Принципи передобробки даних для машинного навчання. In *TOPICAL ISSUES OF MODERN SCIENCE, SOCIETY AND EDUCATION* (pp. 381-385).

76. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1833-1844).

77. Lu, Z., Liu, H., Li, J., & Zhang, L. (2021). Efficient transformer for single image super-resolution. *arXiv preprint arXiv:2108.11084*.

78. Shafronenko, A., Bodyanskiy, Y., & Rudenko, D. (2020). Neuro-fuzzy clustering of Distorted Data Using Cat Swarm Optimization. LAP LAMBERT Academic Publishing.

79. Руденко Д.О., Кухарчук В.А. (2022). Сучасні напрямки застосування методів супер роздільності зображення. In *MODERN RESEARCH IN WORLD SCIENCE* (pp. 393-399).