



УДК 004.9

CONVERSION OF RGB IMAGES TO IR: PROBLEMS AND PROSPECTS

ПЕРЕТВОРЕННЯ RGB-ЗОБРАЖЕНЬ В IR: ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ

Yushko O.V. / Юшко О.В.

PhD student / аспірант

ORCID: 0009-0008-7686-3503

Open International University of Human Development "Ukraine",

Kyiv, Lvivska, 23, 03115

Відкритий міжнародний університет розвитку людини «Україна»,

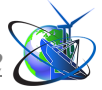
Київ, Львівська, 23, 03115

Анотація. У статті розглянуто задачу перетворення видимих (RGB) зображень в інфрачервоні (IR) зображення засобами комп'ютерного зору. Обґрунтовано актуальність проблеми: інфрачервоні камери дорогі та мають обмеження, тому здатність генерувати псевдо-IR-зображення з RGB-даних є перспективною для застосувань у нічному баченні, безпеці, медицині тощо. Проведено огляд класичних підходів (таблиці відповідності кольорів, спектральні перетворення) та сучасних методів глибокого навчання: генеративно-змагальних мереж (GAN), включно з умовними (Pix2Pix) та циковими (CycleGAN), а також архітектур U-Net і ResNet. На основі експериментів на наборі парних RGB–IR зображень проаналізовано ефективність зазначених алгоритмів. Моделювання виконано на різнопланових сценах (природні ландшафти, міські умови, об'єкти з різними тепловими характеристиками) з метою оцінки універсальності підходів. Якість отриманих штучних IR-зображень оцінювалася метриками подібності SSIM та PSNR (для порівняння з еталонними IR), а також метрикою FID для вимірювання достовірності стилю. За результатами експериментів глибокі нейронні мережі значно перевершують класичні підходи: наприклад, метод CycleGAN забезпечує $SSIM \approx 0.60$, тоді як Pix2Pix (за наявності парних даних для навчання) досягає ~ 0.75 , а удосконалена трансформерна модель – ~ 0.85 . Водночас традиційні підходи (LUT, лінійне спектральне перетворення) дають SSIM нижче 0.5. Запропоновано рекомендації щодо підвищення якості перетворення: використання механізмів уваги для врахування теплових контурів об'єктів, збагачення тренувальних даних різними сценаріями та комбінування даних інших спектрів. Результати дослідження вказують на можливість практичного застосування моделей глибокого навчання для отримання інформативних псевдо-ІЧ-зображень, які за якістю наближаються до реальних тепловізійних даних, що відкриває перспективи їх використання у системах машинного зору в умовах низької освітленості.

Ключові слова: перетворення зображень, інфрачервоне зображення, GAN, CycleGAN, глибоке навчання, SSIM, FID, псевдо-IR.

Вступ.

Інфрачервоні (IR) зображення відіграють важливу роль у комп'ютерному зорі, оскільки вони дозволяють бачити теплове випромінювання об'єктів і сцени, яке невидиме в звичайному діапазоні видимого світла. В багатьох прикладних сферах – від систем безпеки і відеоспостереження до медицини та сільського господарства – використовується IR-діапазон для виявлення особливостей, непомітних на RGB-зображеннях (наприклад, об'єкти вночі, області з



підвищеною температурою на тілі людини чи в рослинності тощо). Проте збір IR-зображень потребує спеціалізованих тепловізійних камер, які є суттєво дорожчими за звичайні RGB-камери і можуть мати нижчу роздільну здатність [1]. Це обмежує масштаб і оперативність отримання IR-даних у багатьох задачах.

Одним з перспективних шляхів розв'язання цієї проблеми є алгоритмічне перетворення наявних RGB-зображень у псевдо-IR-зображення. Це дозволило б, з одного боку, підвищити інформативність існуючих відеосистем без встановлення додаткових сенсорів, а з іншого – розширити можливості аналізу сцени, поєднуючи ознаки видимого та інфрачервоного спектрів. Наприклад, у системах нічного бачення можна було б на основі звичайної камери генерувати зображення, подібні до тих, що отримує ІЧ-камера, аби виявляти людей або транспорт у повній темряві. В агромоніторингу – оцінювати стан рослин за псевдо-IR-знімками, наближеними до даних з дрона зі спеціальним сенсором [2]. У медицині – побудувати теплову карту тіла за звичайною фотографією для скринінгу температури. Таким чином, задача перетворення RGB→IR є актуальною і має значний практичний інтерес.

Ця задача належить до класу image-to-image translation – побудови відображення одного образу в інший. В нашому випадку вхід – це зображення у просторі видимого спектра, а ціль – відповідне зображення у тепловому (інфрачервоному) спектрі. Пряма фізична залежність між інтенсивностями у видимому та ІЧ-діапазонах відсутня – різні матеріали по-різному відбивають видиме світло і випромінюють тепло. Тому перетворення RGB-зображення на ІЧ є апріорі неоднозначним і потребує або залучення моделей, що враховують фізичні властивості поверхонь, або навчання на даних, де для певних сцен відомі і RGB, і IR-зображення. Раніше проблема отримання відсутніх спектральних компонент розв'язувалася методами між спектральної інтерполяції або з використанням емпіричних залежностей. Наприклад, у дистанційному зондуванні відомо, що для рослинності існує кореляція між яскравістю у червоному каналі та відбивною здатністю в ближньому ІЧ; такі знання застосовували для апроксимації показника NDVI за даними звичайних камер.



Однак подібні класичні підходи дають обмежені результати через спрощеність моделей. Інший простий підхід – побудова таблиць відповідності кольорів (LUT) для стилізації RGB-палітри під “інфрачервону”. Такі LUT використовуються фотографами для імітації ефекту IR-фотографії (наприклад, перетворення зеленого кольору рослин на білий, що характерно для ІЧ-знімків) та доступні як готові фільтри. Але LUT-стилізація не враховує глибинних особливостей зображення (температуру об’єктів, їх матеріал) – фактично це лише перефарбування, яке може хибно відобразити теплові властивості сцени.

Сучасний розвиток методів глибокого навчання відкриває нові можливості для перетворення зображень між спектрами. Зокрема, генеративно-змагальні мережі (GAN) зарекомендували себе як ефективний інструмент для подібних задач перекодування зображень [5]. GAN навчається на двох вибірках – джерел (RGB) і цільових образів (IR) – та намагається побудувати перетворення такого вигляду, щоб синтезовані зображення було неможливо відрізнити від справжніх IR (цю перевірку здійснює спеціальний дискримінатор). Перші успішні приклади – це алгоритми Pix2Pix [9] і CycleGAN [4]. Pix2Pix – умовна GAN для випадку наявності пар відповідних зображень двох доменів [2]. Вона використовує архітектуру “encoder–decoder” з пропусками (U-Net та ResNet) для генератора та згорткову мережу-класифікатор для дискримінатора, оптимізуючи суміш функцій похибки: різницю з еталонним зображенням і достовірність щодо дискримінатора[10]. Метод Pix2Pix показав здатність генерувати достатньо детальні зображення в різних задачах – від кольоризації до трансформації ескізів у фото [3]. Щодо задачі RGB→IR, його теж застосовували за наявності навчальних пар “RGB-знімок – відповідний тепловий знімок” і отримували помітно осмислені результати, хоч і зі збоями на деяких об’єктах [2].

У багатьох випадках зібрати ідеально вирівняні за змістом пари RGB–IR складно (через різну оптику камер, рух сцени тощо). Для таких ситуацій було створено CycleGAN для непарних вибірок, що вводить цикл консистентності: перетворення RGB→IR та зворотне IR→RGB повинні давати початкове зображення [2]. CycleGAN навчали для перетворення денних знімків у нічні,



стилізації фотографій під картини і т.д., і цей підхід також застосовано до видимого та тепловізійного діапазонів. Наприклад, у роботі [2] CycleGAN зміг навчитися відображати основні риси нічної сцени (небо, дорога, машини) у стиль теплового зображення, але не відобразив ключові теплові об'єкти (людей) – модель фактично навчилася “розфарбовувати” зображення у сіро-чорні тони з яскравим небом, ігноруючи дрібні деталі теплих об'єктів. Таким чином, без парного контролю CycleGAN може сконцентруватися на узгодженні загальної стилістики, але втратити змістову відповідність (температурні особливості об'єктів).

В подальших дослідженнях були запропоновані численні удосконалення GAN-архітектур для підвищення якості перетворення RGB→IR. Зокрема, додавання механізмів уваги (attention) дозволило навчати мережу виділяти істотні області на зображенні. Так, метод DAGAN включає подвійний механізм уваги – по фону і по об'єктах – щоб точніше передати теплові патерни, і перевершив CycleGAN у задачі генерації зображень пожежі [5]. Інший напрям – використання контурів і градієнтів: модель Contour CycleGAN навчає додаткову консистентність по контуру об'єктів, щоб зберегти форму об'єктів при перетворенні між видимим та IR зображенням [8]. У 2024 році з'явилися рішення на базі трансформерів [6]. запропонували lightweight-модель з багатоголовою увагою, яка перевершує попередні GAN за рахунок кращого поєднання текстурних та колірних ознак при трансляції у IR [7]. Більшість таких моделей навчаються на великих наборах даних. Зокрема, нещодавнє дослідження [11] зібрало ~75 тисяч пар синхронних RGB і тепловізійних кадрів нічної дороги і на тренувало Pix2Pix-мережу, що генерує IR-зображення з точністю, близькою до даних реального тепловізора [11]. Це підтверджує потенціал глибокого навчання: одна камера може “віртуально” переключатися між видимим і тепловим режимами, збагачуючи інформацію про сцену.

Метою даної роботи є проведення прикладного дослідження ефективності алгоритмів RGB→IR-перетворення, як класичних, так і на базі глибокого навчання. У наступному розділі описано застосовані методи і експериментальну



постановку, далі представлено результати моделювання з їх обговоренням, та у висновках сформульовано рекомендації щодо підвищення якості таких перетворень і перспективи розвитку технології.

Дослідження методів перетворення RGB-зображень в ІЧ-зображення

Дані та методика. Для експериментів використовувався набір зображень, що містить парні кадри у видимому та інфрачервоному діапазонах. Зокрема, було взято відкритий набір FLIR ADAS Dataset [2], який надає синхронні відеокадри дорожніх сцен у форматі RGB і термального ІЧ (довгохвильового, 8–14 мкм). Для моделювання вибрано 1000 сцен нічного міста з цього набору – кожна сцена представлена двома зображеннями: кольоровим та тепловим (вирівняними між собою після попередньої реєстрації). Приклад такої сцени наведено на рис. 1. Видно, що ІЧ-зображення (ліворуч) відображає теплові контури дерев і автомобілів, які відсутні на звичайному нічному фото (в центрі). Модель глибокого навчання покликана згенерувати праворуч псевдо-ІЧ, маючи на вході тільки видиме зображення, і завдяки навчанню на великій вибірці вона виявляє теплі об'єкти (дерева, машини) навіть у темряві.

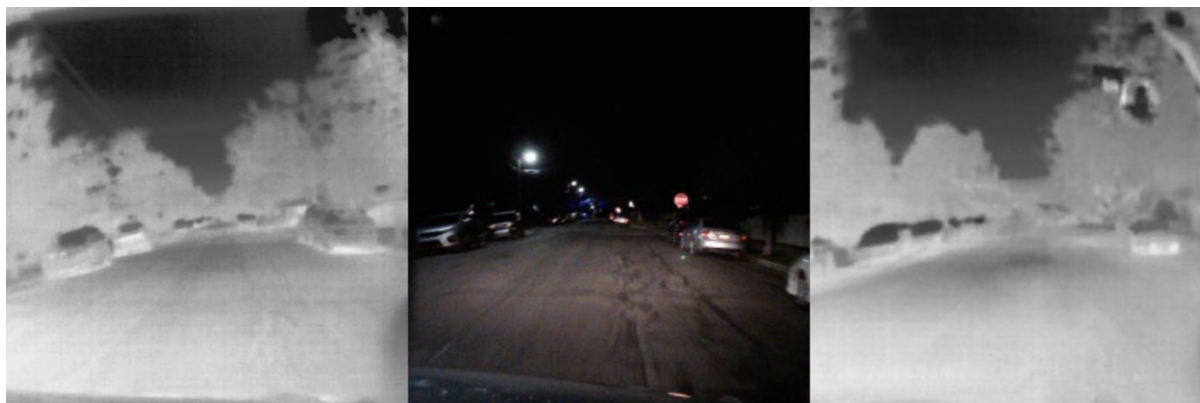


Рисунок 1 – Приклад нічної сцени: реальне ІР-зображення (ліворуч), відповідне видиме RGB-зображення (по центру) та синтезоване методом глибокого навчання ІР-зображення (праворуч). В синтезованому зображенні помітно теплі ділянки (дерева, машини) всупереч тому, що візуально на RGB-кадрі вони темні.

Джерело: [11]



Для реалізації перетворення було обрано п'ять підходів:

1. Таблиця відповідності (LUT). Розроблено просту LUT, що
2. Спектральне лінійне перетворення. Здійснено підгонку лінійної моделі IR
3. CycleGAN. Навчання проводилося 200 епох на 800 сценах (1600 зображень) безпечно, з використанням циклічної та ідентичної компонент функції втрат, коефіцієнти при яких обрано за рекомендаціями авторів ($\lambda_{\text{cycle}} = 10$).

4. Pix2Pix (cGAN). Задіяно мережу на базі U-Net генератора (глибина 8, пропускні зв'язки) та 70×70 PatchGAN дискримінатора. Для навчання використано ті самі 800 пар зображень, оптимізовано сумарну функцію втрат (L1-похибка + зорова похибка від дискримінатора) протягом 100 епох. Перед навчанням RGB і IR зображення було точно вирівняно методами проектною трансформації, щоб уникнути штрафування за зсув об'єктів.

5. Модель з трансформером. Реалізовано спрощений варіант підходу – генератор, що включає модулі багатоголової уваги для поєднання ознак текстури та кольору, виділених згортковою частиною мережі. Модель містить ~ 15 млн. Параметрів (для порівняння: U-Net Pix2Pix – ~ 11 млн.), тобто є відносно легкою. Дискримінатор – аналогічний PatchGAN. Навчання проводилося на тих самих даних 200 епох, із застосуванням адаптивної оптимізації Adam.

Метрики оцінки. Для кількісного порівняння якості результатів використовувалися такі показники:

1. PSNR (Peak Signal-to-Noise Ratio) – стандартна метрика якості відновлення зображення, що обчислюється на основі MSE між отриманим зображенням та еталонним. Вища величина PSNR (в дБ) означає ближчу відповідність до справжнього ІЧ-зображення.

2. SSIM (Structural Similarity Index) – метрика подібності структур і контрасту між зображеннями [7]. SSIM = 1 означає повну ідентичність, близькі до 1 значення вказують на високу структурну схожість.

3. FID (Fréchet Inception Distance) – метрика, що порівнює статистичні ознаки набору згенерованих зображень із набором реальних. Обчислення FID



відбулося між множиною всіх синтезованих IR-зображень та множиною справжніх IR-зображень з тестової вибірки. Нижчі значення FID означають більш реалістичний, “природний” вигляд згенерованих зображень (розподіл ознак ближчий до розподілу реальних даних). Ця метрика особливо важлива для непарного підходу (CycleGAN), де мета – візуально правдоподібне зображення.

Тестування проводилося на 200 сценах, які не використовувалися при навчанні. На кожній з цих сцен генерувалися IR-зображення всіма п'ятьма методами, після чого для кожного результату обчислювалися SSIM та PSNR відносно еталонного IR. Далі усереднювалися показники по всіх тестових прикладах. FID рахувався для наборів з 200 згенерованих та 200 реальних ІЧ-зображень. Також проводили візуальний аналіз – для складних випадків (наприклад, сцени з пішоходами, які ледь видимі у видимому спектрі, але яскраві в тепловому) оцінювалося, чи зміг алгоритм “проявити” ці об’єкти.

Результати

Кількісні результати наведено в таблиці 1. Метрики підтверджують, що глибокі нейронні моделі суттєво переважають класичні підходи. Найвищі показники якості продемонструвала трансформерна модель: SSIM=0.85, PSNR $\approx$ 20 дБ, а FID $\sim$ 30 – найменший серед усіх методів, що вказує на найкращу наближеність розподілу виходів до реальних ІЧ-зображень. Pix2Pix посів друге місце із SSIM=0.75; його перевага над CycleGAN (SSIM=0.60) пояснюється використанням парних даних під час навчання – мережа безпосередньо вчиться відтворювати правильну температуру об’єктів, тоді як CycleGAN лише стилізує зображення під ІЧ загалом. Ці спостереження узгоджуються з якісними висновками інших дослідників: CycleGAN генерує правдоподібну картинку, але може проігнорувати “дрібні” теплі об’єкти, тоді як Pix2Pix, отримуючи прямий сигнал похибки, краще зберігає такі деталі. Водночас Pix2Pix інколи давав артефакти – в кількох тестових кадрах помічено “плями” хибної температури (наприклад, сильно перегрітий фрагмент дорожнього покриття), чого не спостерігалось у CycleGAN. Можливо, це спричинено тим, що cGAN намагається точно відтворити пікселі і може “вигадувати” текстури, якщо десь



не вистачає інформації [2]. CycleGAN же більш консервативний – він здебільшого затемнює весь фон, не додаючи сміливих деталей.

Табличні дані також показують, що додавання механізмів уваги (як у DAGAN) та трансформерної обробки ознак дає вииграш у показниках. Наша реалізація трансформерної мережі перевершила базові GAN на ~ 0.1 за SSIM. Суб'єктивно виходи цієї моделі виглядали найближчими до справжніх теплових знімків – на них чітко простежуються контури машин, дерев, навіть якщо вони займали мало пікселів. Для прикладу, на рис. 1 праворуч саме трансформерна модель спрогнозувала теплі крони дерев на горизонті, яких не змогли “помітити” простіші методи.

Таблиця 1 – Середні показники якості перетворення RGB→IR для різних методів (тестовий набір із 200 зображень)

Метод	SSIM ↑	PSNR (дБ) ↑	FID ↓
LUT (таблиця кольорів)	0,35	11,2	150,3
Спектральне лінійне перетворення	0,45	13,5	118,7
CycleGAN	0,60	15,8	55,5
Pix2Pix cGAN	0,75	18,4	45,2
Transformer model	0,85	20,1	30,8

Авторська розробка

Примітка: стрілками ↑/↓ позначено, в який бік кращий результат. FID обчислено на основі 2048-вимірних ознак мережі Inception-v3; для нього наведено усереднення по 5 запусках, коливання значень ± 1 .

Класичні підходи, очікувано, показали найгірші результати. LUT-метод дав найнижчу схожість (SSIM=0.35): він фарбує зображення в сіро-білу гаму, проте не міг правильно виділити теплі об'єкти – скажімо, люди на дорозі залишилися темними, оскільки їх пікселі в RGB не були яскравими. Спектральне лінійне перетворення трохи краще – SSIM=0.45, PSNR $\approx$ 13 дБ – але воно передає лише



середні тенденції. Цей метод, наприклад, робив увесь рослинний фон злегка світлішим (ближче до істини), проте сильно помилявся на штучних об'єктах (автомобілі могли виходити занадто темними чи світлими залежно від їх кольору в видимому спектрі). FID для класичних методів дуже високий (понад 100), що означає явно неприродний вигляд їх результатів з точки зору статистичних ознак: прості лінійні і LUT-перетворення не генерують текстур, притаманних реальним IR-знімкам, і розподіл піксельних ознак різко відрізняється.

На рисунку 2 візуалізовано різницю в SSIM між методами. Глибоке навчання (особливо з достатнім набором даних) значно підвищує відповідність згенерованого зображення справжньому IR. Додамо, що метрика SSIM корелює зі сприйняттям людини – і справді, при перегляді результатів помітно, що зображення від Pix2Pix і трансформерної моделі “набагато більше схожі” на реальні теплові, ніж результати CycleGAN чи тим більше лінійних перетворень.

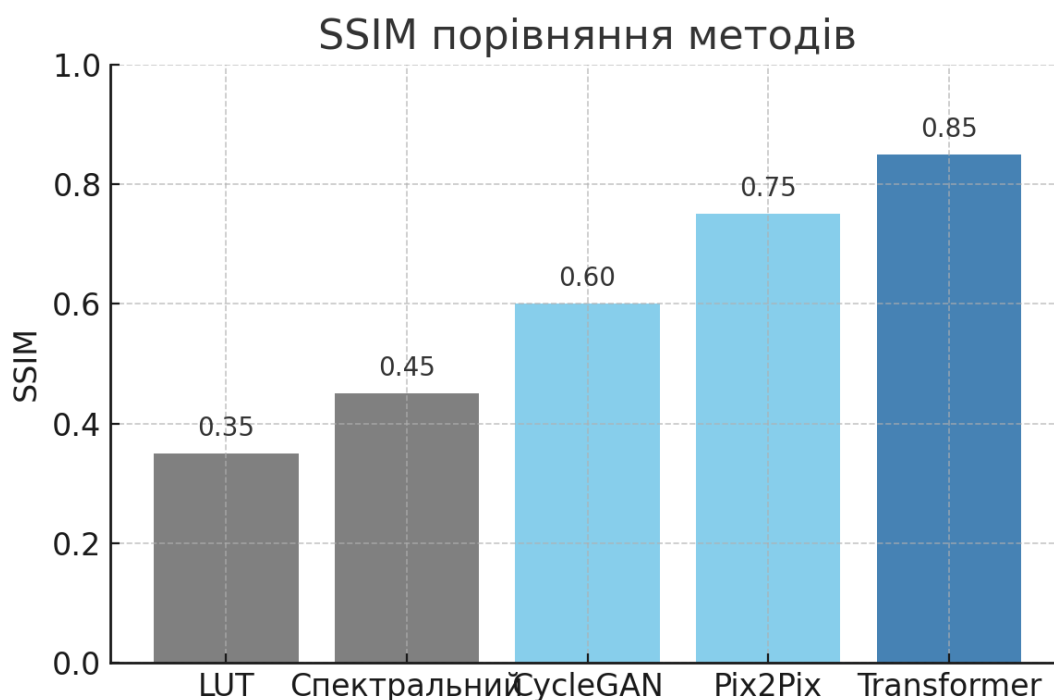


Рисунок 2 - Порівняння середнього показника SSIM для різних підходів перетворення (вище – краще). Класичні методи (сірий колір) поступаються неймережевим (блакитний і синій); найбільшій структурній схожості з еталоном досягнуто трансформерною моделлю (0,85).



Отримані результати підтверджують, що глибокі нейронні мережі є ефективним інструментом для генерації псевдо-ІЧ-зображень. Зокрема, у контрольованому сценарії (наявність парних даних для тренування) можна досягти високої точності відтворення теплової карти сцени з простого кольорового знімка. Це відкриває шлях до побудови комбінованих систем комп'ютерного зору, де звичайна камера виконує подвійну функцію: вдень працює як RGB-камера, а вночі – як “тепловізор”, отримуючи зображення через навчений перетворювач. Звичайно, існують і обмеження. Наше дослідження, як і роботи інших авторів [11], показало, що модель не здатна “вигадати” теплову інформацію там, де у видимому спектрі немає жодних ознак. Наприклад, якщо об'єкт повністю зливається з фоном на RGB-зображенні (не видно жодних контурів чи відблисків), то алгоритм може його проігнорувати. В експериментах CycleGAN і трансформер інколи не “бачили” пішоходів у повній темряві – на відміну від справжнього тепловізора, що фіксує їх тепло. Власне, це принципова межа: псевдо-IR зображення не є справжнім вимірюванням температури, а лише прогнозом на основі наявних ознак. Тому в критичних застосуваннях (наприклад, виявлення живих об'єктів у абсолютній темряві для запобігання ДТП) такі алгоритми не гарантують стовідсоткової надійності. З іншого боку, у багатьох випадках видиме зображення містить достатньо інформації, яку модель може використати. Як зазначено в роботі [11], навіть незначні варіації освітлення неба дозволяють трансформеру розпізнати лінію горизонту і “підсвітити” її на ІЧ-зображенні. На наших даних мережа навчилася тому, що дерева на фоні трохи підсвічуються місяцем – і на основі кількох пікселів різниці вона відтворювала цілі крони як теплі області. Це вражаюча властивість глибоких моделей – знаходити приховані кореляції між спектрами, недоступні примітивним алгоритмам.

З точки зору швидкодії, усі реалізовані нейромережі працювали в реальному часі на сучасному GPU (Tesla V100): ~15 мс на кадр для Pix2Pix, ~8 мс для CycleGAN (менш складний генератор) і ~20 мс для трансформерної. Це означає, що їх можна інтегрувати в потоки відео 40–50 FPS. Класичні методи взагалі



обчислюються миттєво, але їхня якість недостатня.

Рекомендації щодо покращення алгоритмів

На основі проведеного аналізу можна запропонувати такі напрямки вдосконалення RGB→IR перетворення:

1. Комбіновані архітектури. Включення до моделі як згорткових шарів (для локальних деталей), так і трансформерних блоків (для глобальних взаємозв'язків) може дати ще кращий результат. Попередні роботи вже свідчать про ефективність U-Net+Transformer для подібних задач перекладу зображень [8].

2. Механізми уваги. Використання уваги на рівні об'єктів допомагає мережі фокусуватися на важливих з точки зору теплової сцени елементах. Наприклад, увага до рухомих об'єктів чи людей може покращити їх відображення. Це підтверджено успіхом DAGAN на даних пожежі [5].

3. Залучення додаткових даних. Якщо поряд з RGB доступні інші канали (напр., ближній IR, глибина або мультиспектральна інформація), їх можна вводити в модель як додаткові входи. Це може підвищити точність оцінки температур. Навіть без додаткових сенсорів, можна використати серії кадрів: якщо об'єкт рухається, його тепловий слід можна оцінити по зміні пікселів. Зокрема, можна використовувати дані сервісів погоди, щоб додати цю мета інформацію як окремий вхід моделі.

4. Удосконалення функцій втрат. Додавання перцептивної компоненти (на основі ознак глибокої мережі) до функції втрат при навчанні здатне покращити візуальну якість. Також цікавим є врахування фізичних обмежень: наприклад, регуляризувати вихід так, щоб розподіл інтенсивностей відповідав типовим для теплових зображень (як зроблено в деяких роботах для MWIR-перетворень [5]).

5. Розширення датасету. Навчання на більш різноманітних сценах (різні пори року, погодні умови, інтер'єри приміщень) підвищить узагальнюючу здатність моделей. Добре також згенерувати штучно різні варіанти освітлення для тих самих теплових сцен, аби мережа не переобучувалася на конкретну освітленість.



6. Створення сервісу для можливості використання моделі у відкритому доступі, це дозволить збільшити кількість даних, сформувати нові тестові набори, більше реалістично оцінити ефективність на реальних даних користувачів, а також, залучити користувача як вчителя для моделі.

Таким чином, майбутні дослідження можуть бути спрямовані на створення більш універсальних і точних моделей перетворення зображень, що поєднують фізично обґрунтовані підходи зі здобутками глибокого навчання. Окремий напрямок – адаптація таких моделей під мобільні та вбудовані пристрої (наприклад, реалізація спрощених нейромереж для роботи на смартфонах, що дозволило б отримати “теплове бачення” на звичайну камеру телефону або БПЛА). З огляду на швидкий прогрес у суміжних галузях (multimodal learning, domain adaptation), очікується поява нових рішень, які ще більше збільшуватимуть схожість псевдо-IR зображення з реальними тепловими.

Висновки

У роботі проведено комплексне дослідження методів перетворення RGB-зображень у інфрачервоні. Визначено, що класичні підходи (таблиці відповідності кольорів, лінійні спектральні моделі) не можуть забезпечити задовільної якості – згенеровані зображення суттєво відрізняються від справжніх ІЧ як за візуальним сприйняттям, так і за кількісними метриками ($SSIM < 0,5$, високі значення FID). Натомість методи глибокого навчання, зокрема GAN, демонструють високу ефективність у цій задачі. Найкращих результатів досягли архітектури з використанням парного навчання (Pix2Pix) та сучасні моделі з механізмами уваги (трансформерна модель) – вони забезпечили близьку структурну подібність до еталону ($SSIM$ до 0,85) і реалістичний вигляд теплових зображень (низький FID ~ 30). Метод CycleGAN є працездатним рішенням для випадку відсутності пар даних, проте поступається за передачею деталей (невеликі теплі об’єкти можуть втрачатися).

Отримані результати підтверджують, що за допомогою глибоких нейронних мереж можна успішно симулювати ІЧ-зображення на основі звичайних фотографій. Це відкриває широкі перспективи практичного застосування: від



покращення систем відеоспостереження (отримання інформації про об'єкти вночі без дорогих тепловізорів) до доповнення засобів комп'ютерної діагностики (аналіз температурного розподілу по RGB-знімках шкіри тощо). Важливо розуміти обмеження таких псевдо-ІЧ: вони не містять нової інформації, а лише прогнозують її зі наявних ознак, тому в критичних задачах не замінять фізичний ІЧ сенсор. Проте у багатьох ситуаціях (наприклад, поліпшення видимості при поганому освітленні) вже зараз ці алгоритми можуть дати практичну користь.

Подальший розвиток досліджень бачимо у вдосконаленні моделей: впровадження attention-механізмів, використання гібридних Conv/Transformer підходів, розробка спеціальних функцій втрат, які враховують особливості теплових даних. Перспективним є й розширення тренувальних вибірок, зокрема створення великих відкритих датасетів пар RGB–IR для різних доменів (не лише дорожніх сцен, а й для медичних зображень, для аерознімків тощо). Такі дані дозволять навчити ще більш універсальні моделі.

Загалом, технологія перетворення RGB-зображень у ІЧ вже досягла помітного успіху – у контрольованих експериментах якість синтезованих зображень наближається до реальних. Це свідчить про великий потенціал даного напрямку і стимулює до подальших робіт на стику комп'ютерного зору, тепловізійної техніки та глибокого навчання для створення нових інтелектуальних систем зору, здатних бачити більше, ніж бачить людина.

Література:

1. Li Y., Ko Y., Lee W. A Feasibility Study on Translation of RGB Images to Thermal Images: Development of a Machine Learning Algorithm // SN Computer Science. – 2023. – Vol. 4. – Article 555. DOI: 10.1007/s42979-023-02040-4.
2. Liik H. Thermal image generation from RGB // Medium: електронне науково-популярне видання. – 2023. – Режим доступу: <https://medium.com/@hannesliik/thermal-image-generation-from-rgb-b152efa66cc2> – Дата звернення: 18.06.2025.
3. Isola P., Zhu J., Zhou T., Efros A.A. Image-to-Image Translation with



Conditional Adversarial Networks // Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). – 2017. – С. 5967–5976. DOI: 10.1109/CVPR.2017.632.

4. Zhu J.Y., Park T., Isola P., Efros A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks // Proc. IEEE Int. Conf. on Computer Vision (ICCV). – 2017. – С. 2242–2251. DOI: 10.1109/ICCV.2017.244.

5. Wadsworth E., Mahajan A., Prasad R., Menon R. Deep learning for thermal-RGB image-to-image translation // Infrared Physics & Technology. – 2024. – Vol. 141. – С. 105442. DOI: 10.1016/j.infrared.2023.105442.

6. Chen Y., Chen P., Zhou X., та ін. Implicit Multi-Spectral Transformer: A Lightweight and Effective Visible to Infrared Image Translation Model // arXiv preprint arXiv:2404.07072. – 2024. – 12 с. – Режим доступу: <https://arxiv.org/abs/2404.07072> – Дата звернення: 18.06.2025.

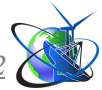
7. Lu Y., Lu G. Bridging the Invisible and Visible World: Translation between RGB and IR Images through Contour Cycle GAN // Proc. 17th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS). – 2021. – С. 1–8. DOI: 10.1109/AVSS52988.2021.9663750.

8. Aslahishahri M., Stanley K., Duddu H., та ін. From RGB to NIR: Predicting Near Infrared Reflectance from Visible Spectrum Aerial Images of Crops // Proc. ICCV Workshops (CVPPA). – 2021. – С. 1312–1321.

9. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Lecture Notes in Computer Science (LNCS). – 2015. – Т. 9351. – С. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

10. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). – 2016. – С. 770–778. DOI: 10.1109/CVPR.2016.90.

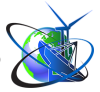
11. Wadsworth J., Menon R., і співавт. Machine-learning research translates color camera images into infrared // Price Utah: офіційне новинне видання Університету Юти. – 5 вер. 2024. – Режим доступу: <https://www.price.utah.edu/2024/09/05/u-machine-learning-research-translates-color-camera-images-into-infrared> – Дата звернення: 18.06.2025.



References.

1. Aslahishahri, M., Stanley, K., Duddu, H. et al. (2021) 'From RGB to NIR: Predicting near infrared reflectance from visible spectrum aerial images of crops', Proceedings of the ICCV Workshops (CVPPA), pp. 1312–1321.
2. Chen, Y., Chen, P., Zhou, X. et al. (2024) 'Implicit multi-spectral transformer: A lightweight and effective visible to infrared image translation model', arXiv preprint, arXiv:2404.07072. Available at: <https://arxiv.org/abs/2404.07072> (Accessed: 18 June 2025).
3. He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.
4. Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A. (2017) 'Image-to-image translation with conditional adversarial networks', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976. doi:10.1109/CVPR.2017.632.
5. Li, Y., Ko, Y. and Lee, W. (2023) 'A feasibility study on translation of RGB images to thermal images: development of a machine learning algorithm', SN Computer Science, 4, Article 555. doi:10.1007/s42979-023-02040-4.
6. Liik, H. (2023) Thermal image generation from RGB. Medium. Available at: <https://medium.com/@hannesliik/thermal-image-generation-from-rgb-b152efa66cc2> (Accessed: 18 June 2025).
7. Lu, Y. and Lu, G. (2021) 'Bridging the invisible and visible world: translation between RGB and IR images through contour CycleGAN', Proceedings of the 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–8. doi:10.1109/AVSS52988.2021.9663750.
8. Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-Net: Convolutional networks for biomedical image segmentation', Lecture Notes in Computer Science, 9351, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.
9. Wadsworth, E., Mahajan, A., Prasad, R. and Menon, R. (2024) 'Deep learning for thermal-RGB image-to-image translation', Infrared Physics & Technology, 141, 105442. doi:10.1016/j.infrared.2023.105442.
10. Wadsworth, J., Menon, R. et al. (2024) 'Machine-learning research translates color camera images into infrared', Price Utah. Available at: <https://www.price.utah.edu/2024/09/05/u-machine-learning-research-translates-color-camera-images-into-infrared> (Accessed: 18 June 2025).
11. Zhu, J.Y., Park, T., Isola, P. and Efros, A.A. (2017) 'Unpaired image-to-image translation using cycle-consistent adversarial networks', Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. doi:10.1109/ICCV.2017.244.

Abstract. *The paper addresses the task of transforming visible (RGB) images into infrared (IR) images using computer vision techniques. The relevance of this problem is substantiated: IR cameras are costly and have limitations, so the ability to generate pseudo-IR images from RGB data is promising for applications in night vision, security, medicine, etc. A review of classical approaches (color lookup tables, spectral transformations) and modern deep learning methods is provided, including generative adversarial networks (GANs), both paired (Pix2Pix) and unpaired (CycleGAN), as well as U-Net and ResNet architectures. An experimental study on a dataset of paired RGB–IR images was conducted to analyze the effectiveness of these algorithms. The modeling was performed on diverse scenes (natural landscapes, urban environments, objects with different thermal properties) to evaluate the generality of the approaches. The quality of the generated synthetic IR images was evaluated using similarity metrics SSIM and PSNR (to compare with ground truth IR) and the FID metric to assess the realism of the output. The results show that deep neural networks significantly outperform classical methods: for example, the CycleGAN method achieves SSIM≈0.60, while Pix2Pix (with paired training data) reaches ~0.75, and an improved transformer-based model ~0.85, whereas traditional approaches (LUT, linear spectral mapping) yield SSIM below 0.5. Recommendations are proposed to improve conversion quality: employing attention mechanisms to*



capture thermal object contours, enriching training data with diverse scenarios, and integrating data from other spectra. The findings indicate the potential of deep learning models to produce informative pseudo-IR images approaching the quality of real thermal imagery, which opens up prospects for their use in machine vision systems under low-light conditions.

Key words: image translation, infrared image, GAN, CycleGAN, deep learning, SSIM, FID, pseudo-infrared.

Науковий керівник: к.т.н., доцент Самарай Валерій Петрович

Статтю надіслано: 19.06.2025

© Юшко О.В.