

ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ ЗОБРАЖЕНЬ. МЕТОДИ РЕАЛІСТИЧНОГО ЗАМІЩЕННЯ ОБ'ЄКТІВ

Білий М. Д.

Науковий керівник – к.т.н, доц. Каук В. І.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: mykhailo.bilyi@nure.ua

This study explores advanced methods for realistic object substitution in digital images through machine learning techniques. Focusing on convolutional neural networks and generative models, it evaluates various algorithms including PatchMatch, GANs, DeepFill V2, and Stable Diffusion Inpainting, among others, for their effectiveness in seamless object integration. The analysis demonstrates the high efficiency of these methods in processing complex contextual relationships, ensuring that the integration of objects into original images appears natural and convincing.

Це дослідження розглядає передові методи реалістичної заміни об'єктів у цифрових зображеннях за допомогою технік машинного навчання. Зосереджуючись на згорткових нейронних мережах та генеративних моделях, воно оцінює різні алгоритми, включаючи PatchMatch, GANs, DeepFill V2 та Stable Diffusion Inpainting, серед інших, за їх ефективність у безшовній інтеграції об'єктів.

Метод PatchMatch та генеративні змагальні мережі (GANs) взаємодіють у сфері обробки зображень, доповнюючи один одного. PatchMatch ефективно вирішує задачі редагування зображень шляхом швидкого знаходження відповідностей між фрагментами, тоді як GANs забезпечують створення високореалістичних зображень, що важко відрізнити від справжніх. Вони навчаються відтворювати реалістичні текстури та деталі, заміщаючи пропущені сегменти так, що відновлене зображення здається цілісним і неперервним. Це досягається шляхом "навчання" однієї частини мережі (генератора) виробляти зображення, в той час як інша частина (дискримінатор) оцінює їх якість, покращуючи з часом результат (див. рис. 1).

Однією з проблемою є те, що згорткові нейронні мережі обробляють особливості зображення з локальним згортковим ядром шар за шаром, отже, не ефективні для запозичення ознак із віддалених просторових місць. Щоб подолати обмеження, ми розглядаємо механізм уваги та вводимо новий контекстний рівень уваги в глибокій генеративній мережі.

Рівень контекстної уваги дізнається, де запозичити або скопіювати інформацію про функції з відомих фонових латок, щоб створити відсутні латки. Він диференційований, тому його можна навчити в глибоких моделях, і повністю згортковий, що дозволяє тестувати на довільних роздільностях.

Зіставлення та спостереження. Ми розглядаємо проблему, у якій ми хочемо зіставити характеристики відсутніх пікселів (передній план) з оточенням (фон). Як показано на рисунку 2, ми спочатку витягуємо патчі (3×3) у фоновому режимі та змінюємо їх як згорткові фільтри. Щоб зіставити патчі переднього плану $\{f_{x,y}\}$ із фоновими $\{b_{x,y}\}$, ми вимірюємо нормалізований внутрішній добуток (косинус подібності).

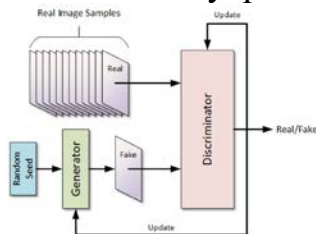


Рисунок 1 – GAN

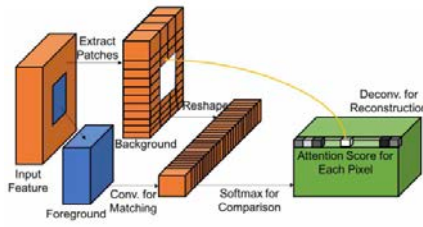


Рисунок 2 – Contextual Attention

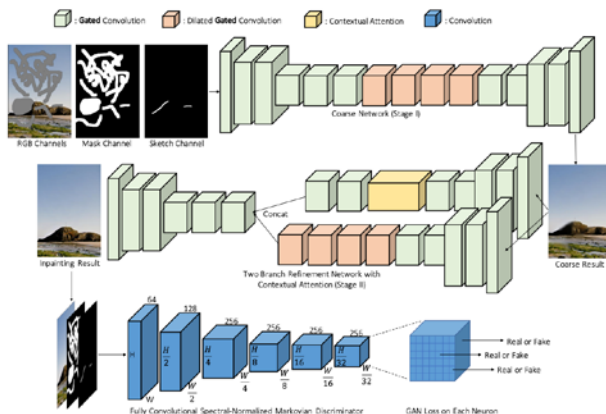


Рисунок 3 – Gated Convolution



Рисунок 4 – DeepFill V2

Gated Convolution вносить інновації в обробку зображень, вирішуючи проблему жорсткості звичайних згорткових мереж, які однаково обробляють всі частини зображення. Він додає "ворота" до згорткових операцій, дозволяючи мережі адаптивно змінювати вагу кожного пікселя на основі його контексту.

Таблиця 2 – Порівняння Gated Convolution з іншими методами.

Метод	Прямокутна маска		Вільної форми маск	
	l_1 помилка	l_2 помилка	l_1 помилка	l_2 помилка
PatchMatch	16.1%	3.9%	11.3%	2.4%
Global&Loca	9.3%	2.2%	21.6%	7.1%
ContextAttention	8.6%	2.1%	17.2%	4.7%
PartialConv	9.8%	2.3%	10.4%	1.9%
GatedConvolution	8.6%	2.0%	9.1%	1.6%

Gated Convolution ефективно вирішує проблеми з непередбачуваними або змінними областями зображення, значно підвищуючи точність та адаптивність моделі. Архітектуру можна побачити на рисунку 3.

DeepFill v2 розширює можливості інтегруючи Contextual Attention разом із Gated Convolution. Це поєднання дозволяє моделі більш точно відновлювати зображення, звертаючи увагу на контекстуальні відносини всередині пропущених або видалених ділянок (рис. 4).

Передовою моделлю генеративного машинного навчання є Stable Diffusion, розробленою для створення деталізованих зображень з текстових описів. Ця технологія використовує глибоке навчання для аналізу великих наборів даних із зображеннями та відповідними описами, навчаючись генерувати нові зображення, що відповідають заданим текстовим підказкам. Stable Diffusion здатна створювати високоякісні візуальні зображення з широкого спектру категорій, від реалістичних сцен до фантастичних ілюстрацій, відкриваючи нові можливості для творчості та дизайну.

В результаті дослідження, було встановлено, що кожен метод обробки зображень має свої переваги в певних умовах. Створення програмної системи, яке дозволяє завантажити зображення, нанести маску та вибрати метод обробки, стане важливим інструментом для фахівців. Це (рис. 5) дозволить не тільки вибрати найбільш підходящий метод для конкретної задачі inpainting, але й забезпечить можливість порівняння результатів, сприяючи вибору найкращого рішення для реалістичного заміщення об'єктів.

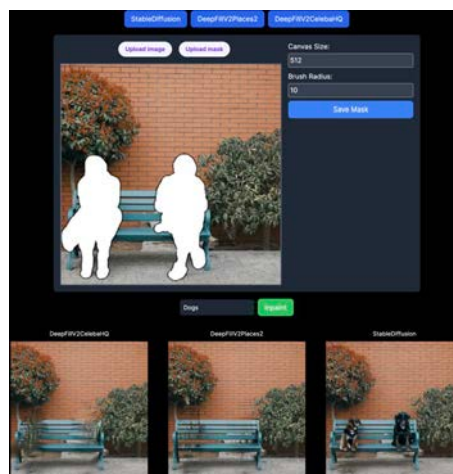


Рисунок 5 – Програмна система

Список використаних джерел:

1. Adaptive Image Enhancement Model for the Robot Vision System Vide. Tehnologija. Resursi – Environment, Technology, Resources, 2023, 3, с. 246–251.
2. Generative Image Inpainting with Contextual Attention [Електронний ресурс] / [J. Yu, Z. Lin, J. Yang та ін.]. – 2018. – Режим доступу до ресурсу: <https://arxiv.org/abs/1801.07892>.
3. Yu J. Free-Form Image Inpainting with Gated Convolution [Електронний ресурс] / J. Yu, Z. Lin, J. Yang. – 2019. – Режим доступу до ресурсу: <https://arxiv.org/abs/1806.03589>.