"ІНТЕЛЕКТУАЛЬНА СИСТЕМА СЕГМЕНТАЦІЇ ТА СТИЛІЗОВАНОГО ПЕРЕКЛАДУ ІЄРОГЛІФІЧНОГО ТЕКСТУ У МУЛЬТИМЕДІЙНИХ КОМІКСАХ"

Негруб Андрій, ФІКН БП-3
Керівник: ст. викладач Радзієвська О.В.
Кафедра інформатики

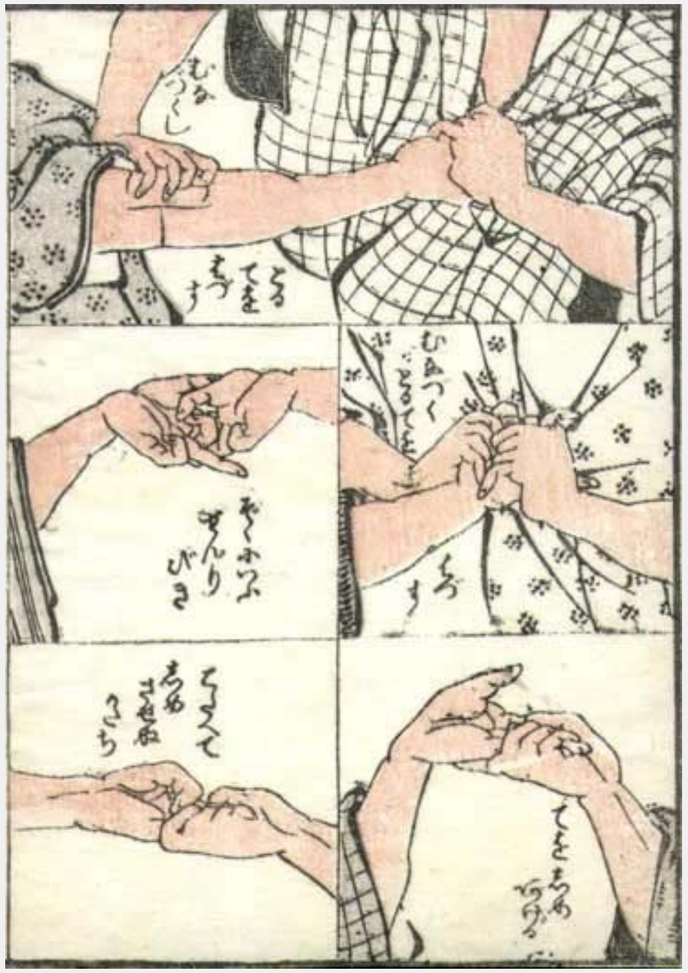
Світ коміксів

Релевантність дослідження

- Зростання попиту на локалізацію [1, 2], зокрема українською мовою;
- Недоліки існуючих систем перекладу для візуально-стилізованого тексту;

Мета роботи

- Використання візуальних характеристик текстових елементів для створення нового підходу до перекладу;
- Розробка модульної системи локалізації на основі цього підходу;



Манґа Хокусай із зображенням прийомів самозахисту (початок 19 століття)

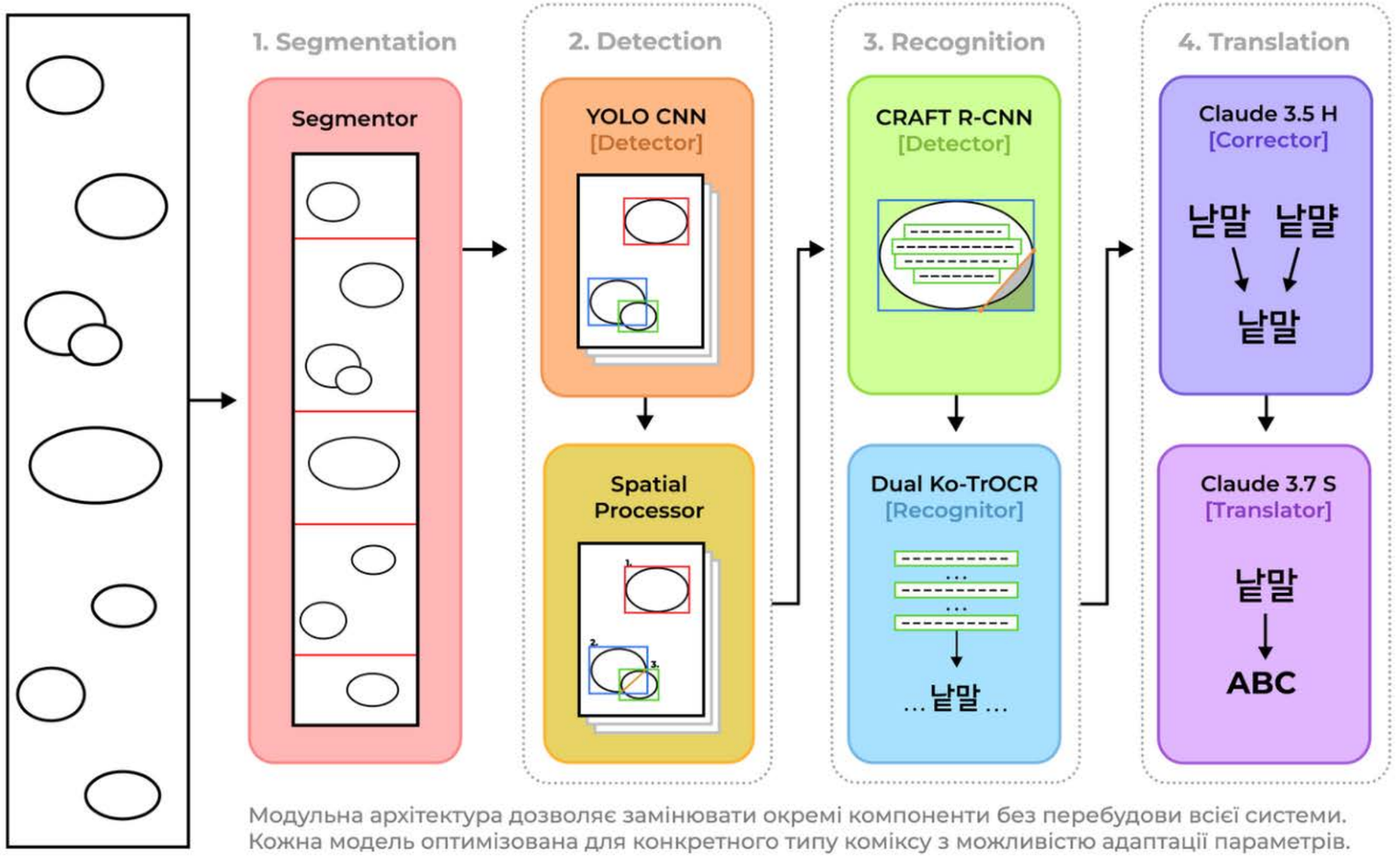
Візуальна стилізація

Кольорами заголовків зазначені відповідні класам розфарбування об'єктів на подальших візуалізаціях.

Характеристики класів текстових елементів:

- Кожен клас має унікальні візуальні ознаки, які використовуються для класифікації
- Агент детекції використовує ці візуальні ознаки для одночасної детекції та класифікації
- Класифікація текстових блоків враховується при подальшій обробці та перекладі

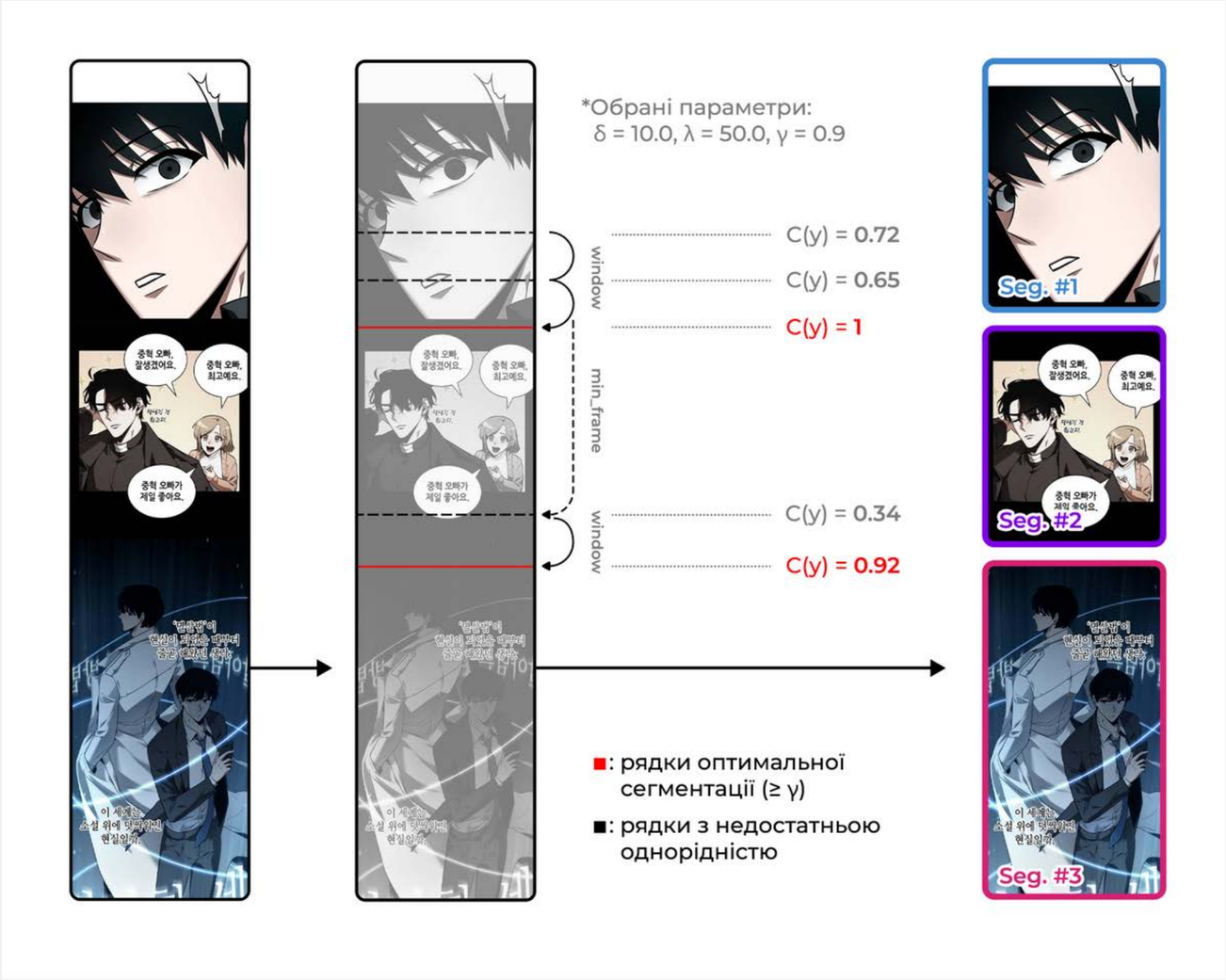
Запропонована архітектура



Модульна архітектура дозволяє замінювати окремі компоненти без перебудови всієї системи.
Кожна модель оптимізована для конкретного типу коміксу з можливістю адаптації параметрів.

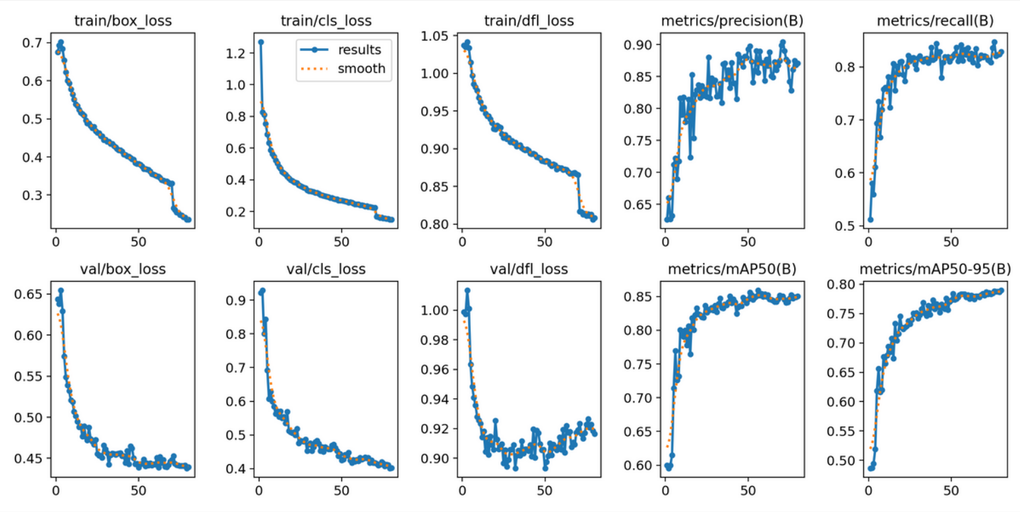
Сегментація сторінок

Розроблено спеціалізований алгоритм розбиття зображення на основі методу seam carving [3], що використовує метрику однорідності для визначення точок поділу зображення.

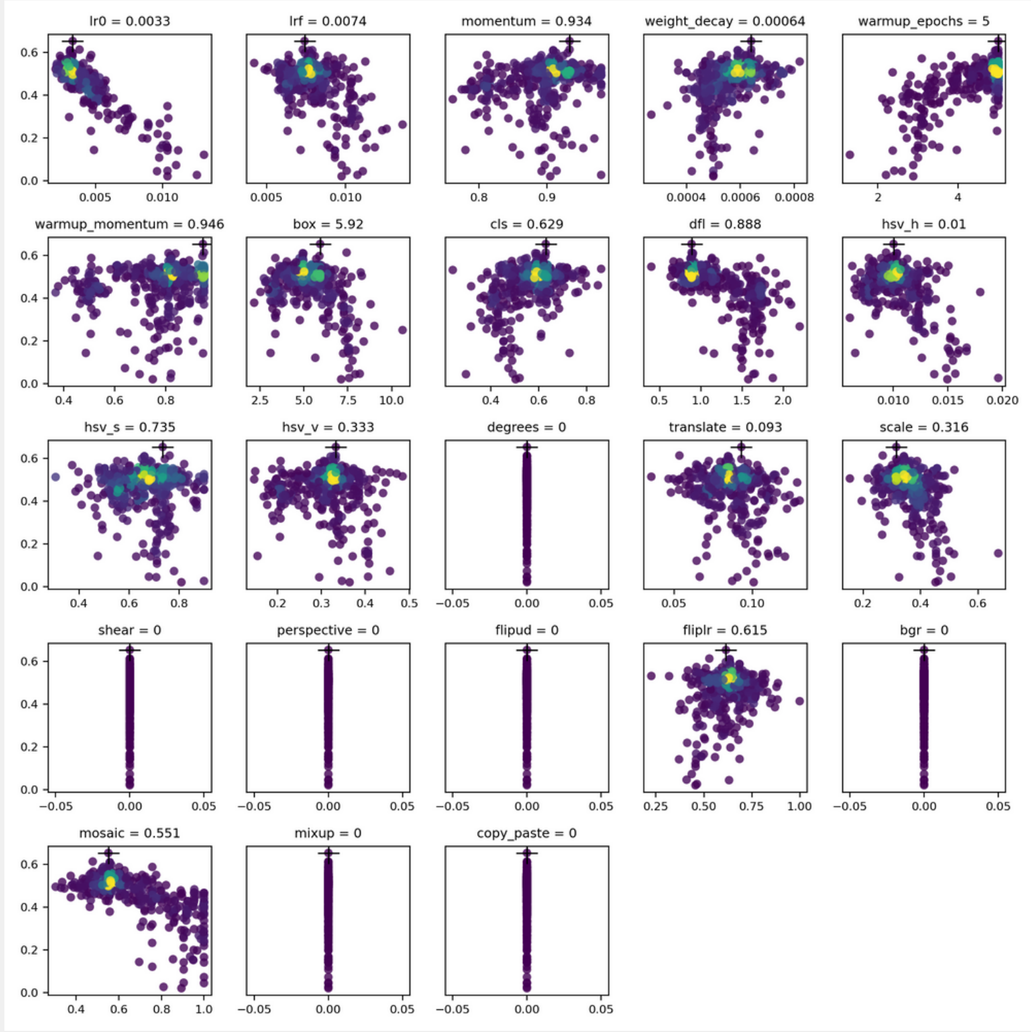


Детекція та класифікація

Експериментальне порівняння показало, що сімейство YOLO значно перевершує аналоги за співвідношенням точність/швидкість.



Динаміка навчання моделі YOLOv12: функції втрат та метрики якості

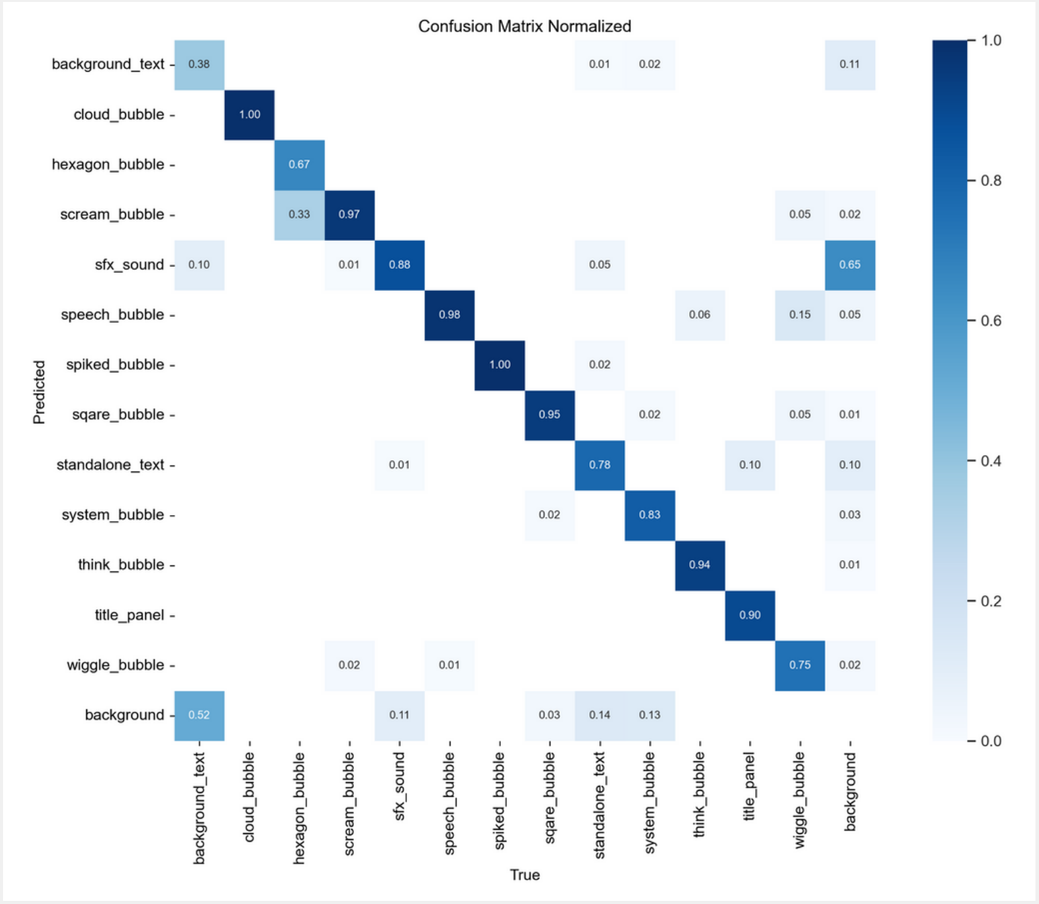


Матриця налаштування гіперпараметрів під час оптимізації

Результати тренування

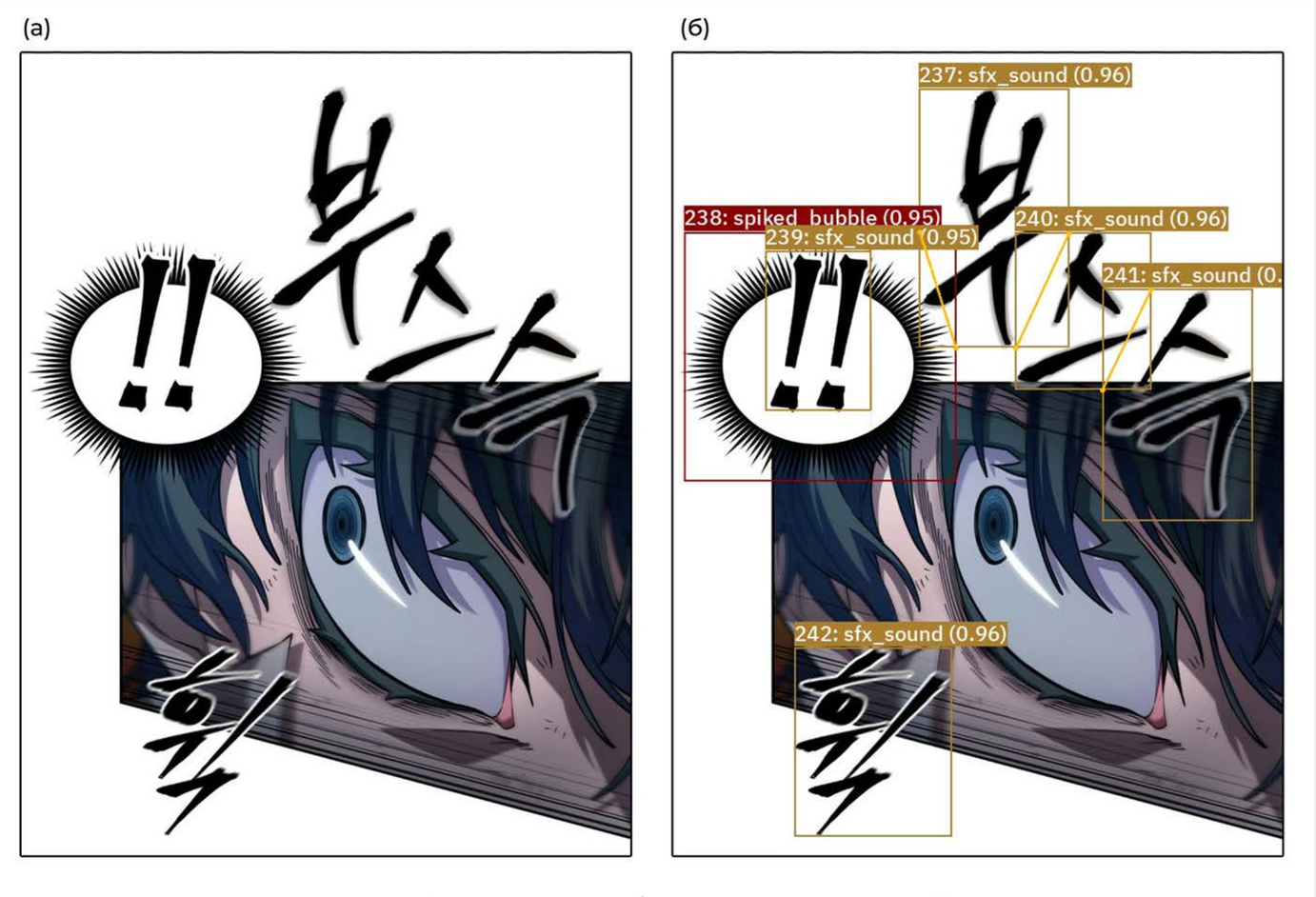
Результати експериментів з різними архітектурами моделей детекції

Модель	mAP50(B)	mAP50-95(B)	Precision(B)	Recall(B)	Inference (ms/img)
YOLOv11m	0.829	0.761	0.896	0.786	30.7
YOLOv11m (tuned)	0.85	0.79	0.871	0.828	30.7
YOLOv9m (tuned)	**0.863**	**0.8**	**0.872**	**0.831**	32.9
YOLOv12m (tuned)	0.841	0.781	0.853	0.806	29.7
YOLOv12s (tuned)	0.853	0.786	0.86	0.825	**27.2**
DETR ResNet-50	0.826	0.749	0.72	0.81	78



Нормалізована матриця плутанини для класів текстових блоків

Приклад роботи детектора



Результат роботи агента детекції
(а – стартове зображення; б - опрацьоване)

Сортування та маскування

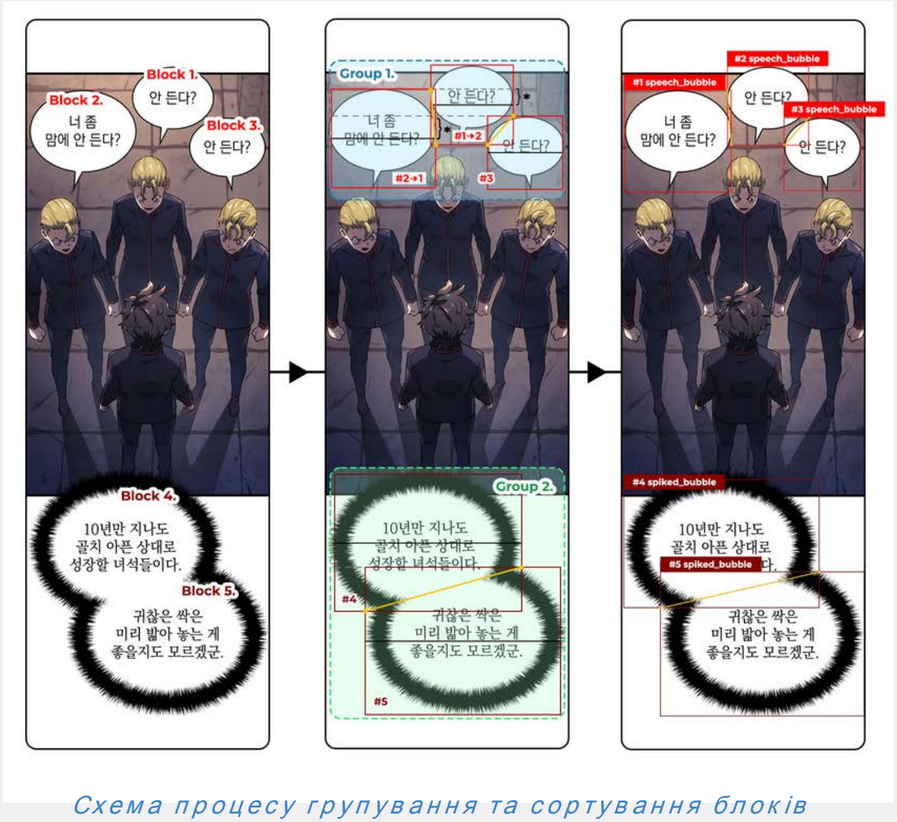
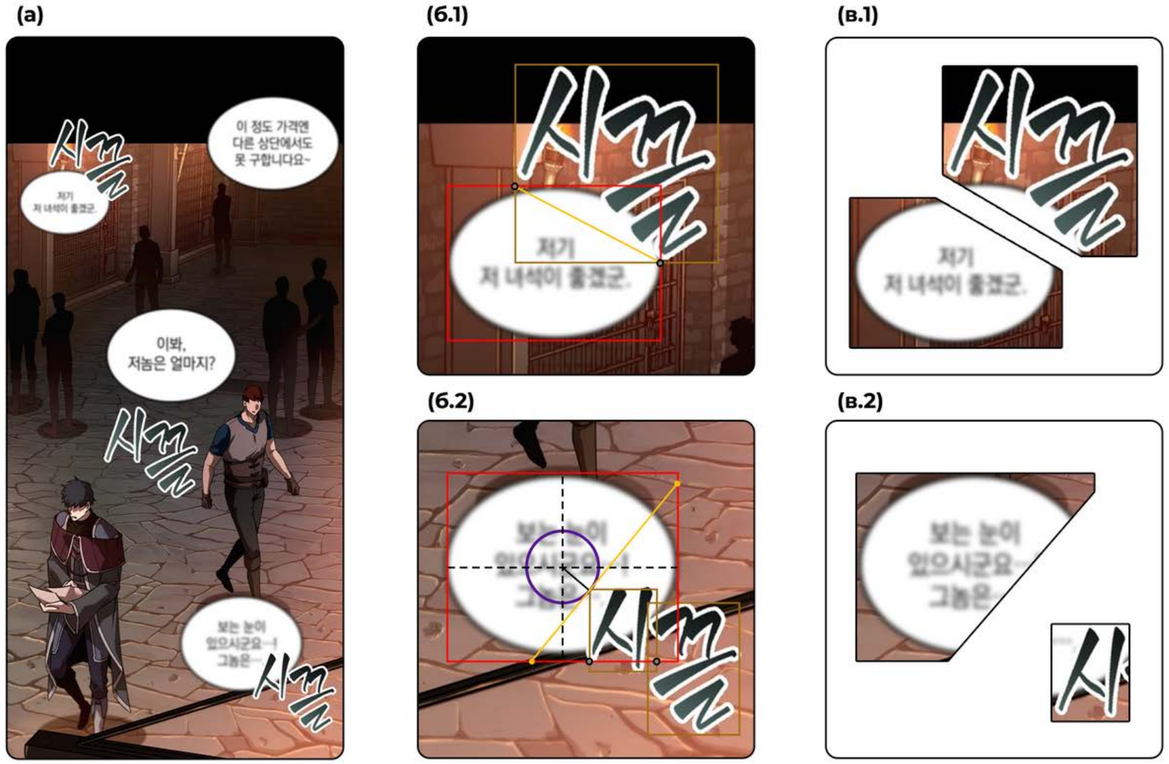


Схема процесу групування та сортування блоків



Приклад маскування перекриттів текстових блоків: (а) вихідне зображення, (б.1, б.2) визначення ліній відсікання, (в.1, в.2) результати маскування

Задача оптичного розпізнавання

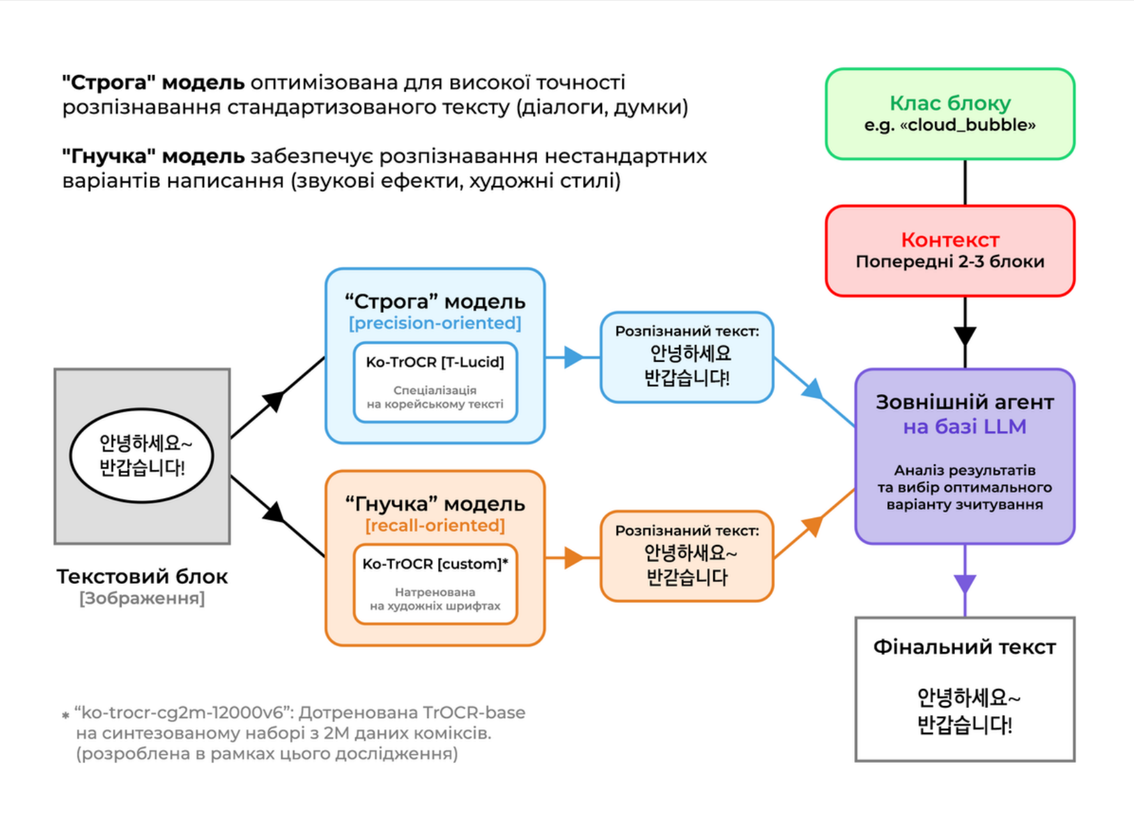
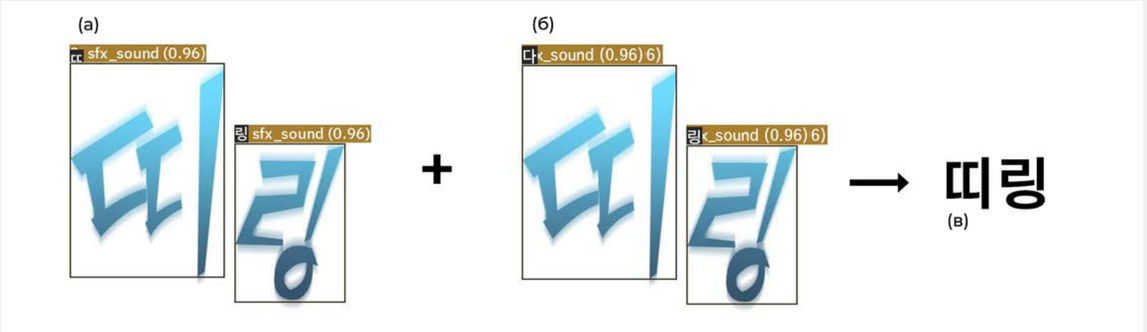


Схема дуальної архітектури розпізнавання тексту [4, 5]



Порівняння результатів розпізнавання:
(а – результат "строгої" моделі; б – результат "гнучкої" моделі; в – об'єднаний результат)

Синтетичний корпус і навчання OCR

Частина шрифтів, використаних для синтетичної генерації

그라나즈	그라나트	그라노베터	그라니
그라나치	그라네	그라노예르스	그라니에
그라나타	그라네로	그라노체	그라니코스

Приклад сформованого токенсету: Токени на «그라»

Приклад генерації синтетичних зображень (оригінальний текст, згенерована картинка, символьна маска)

Виявлення рядків CRAFT

Візуалізація роботи CRAFT на прикладі балону манхви [6]

Порівняння роботи CRAFT у детекції тексту запровонованими моделями (а – з використанням CRAFT; б – без використання)

Результат роботи агента розпізнавання (а – стартове зображення; б - опрацьоване)

Приклад роботи OCR розпізнавача

Порівняльна оцінка OCR-підходів

Метод	*CER (%)*	*WER (%)*	*BLEU Score*
Tesseract OCR	43.7	54.5	0.39
EasyOCR	18.3	26.8	0.68
Базова TrOCR	6.7	12.1	0.81
Адаптована TrOCR	8.2	10.7	0.84
Дуальна OCR-стратегія	**4.8**	**7.2**	**0.92**

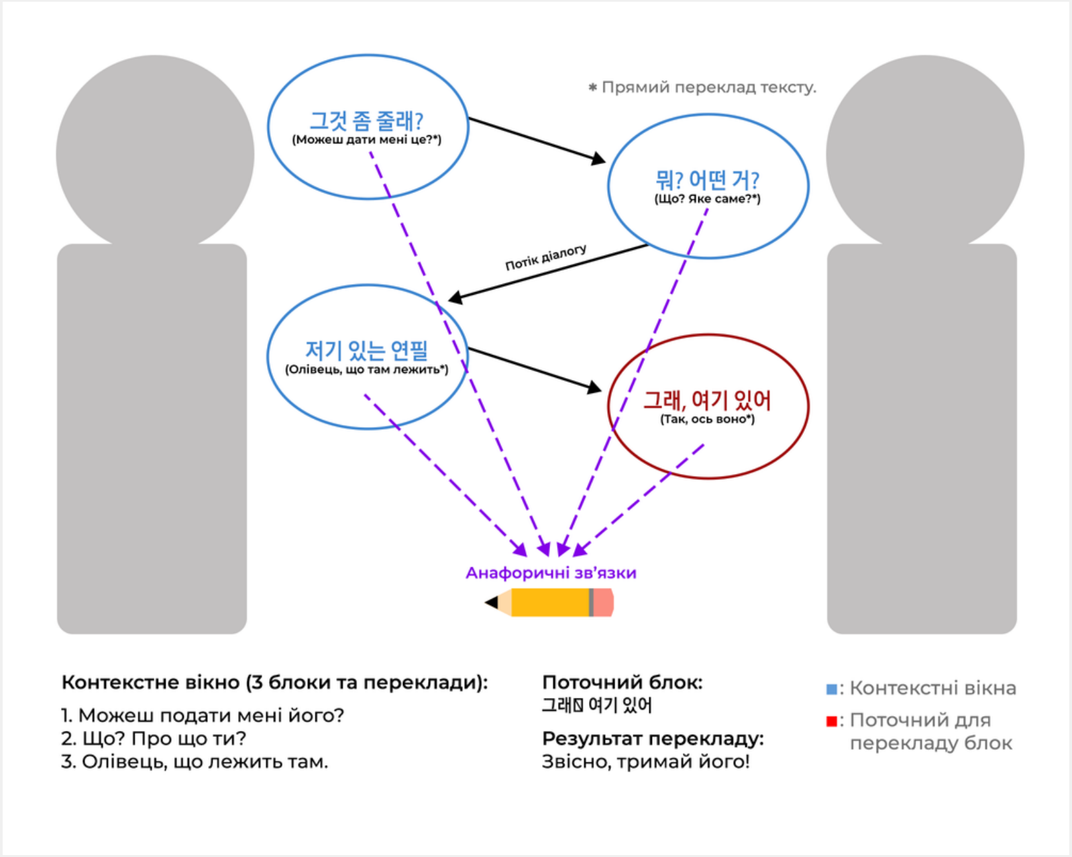
Стилізований переклад

Порівняння моделей для стилізованого перекладу

Модель	*Якість перекладу*	*Стилістична відповідність*	*Контекстуальна інтеграція*	*Швидкодія (с/блок)*
Helsinki-NLP	0.72	0.35	0.2	**0.12**
FlanT5	0.76	0.4	0.45	0.23
Claude 3.7	**0.93**	**0.89**	**0.87**	105
Qwen2.5	0.87	0.71	0.72	2.87
GPT-4	**0.91**	0.84	0.83	132

Приклад промпту для класу think_bubble:

```
This text represents inner thoughts of a character. Use
introspective, reflective style. Add thought fragments,
ellipses (...) where appropriate to convey thinking
process. Translate to create a distinct internal voice
that differs from spoken dialogue.
```



Діаграма принципу роботи контекстних вікон

Порівняльний аналіз систем

Система	*Точність виявлення (%)*	*CER (%)*	*Якість перекладу (0-1)*	*Збереження стилю (0-1)*	*Час обробки (с/фрейм)*
ComiTranslate [7]	72.7	16.2	0.62	0.41	5.3
ogkalu/Comic-Translate [8]	**88.9**	7.3	**0.89**	0.73	1.76
EasyOCR + DeepL	78.5	23.4	0.57	0.32	**0.63**
ComicGlot (розроблена)	86	**4.8**	**0.93**	**0.89**	1.44

Результати

0: Я дотрималась обіцянки, новачку.
나 약속 지켰어요,
초보자님.

1: Ви прибули раніше за нас.
우리보다
빨리 도착했군.
2: Я ж казала, що я ветеран!
말했잖아요,
나 베테랑이라고!

3: І я хочу познайомити тебе з однією людиною. Це твій фанат!
그리고
소개해 드릴
사람이 있어요.
당신의 팬이에요!
4: Фанат?
팬?

Напрямки вдосконалення

01.

Розробка спеціалізованої OCR-моделі для звукових ефектів;

02.

Впровадження локальної інференції для LLM-компонентів;

03.

Розширення функціональності з автоматичною вставкою тексту;

Використані джерела

[1] Comic book market size, share, value, trends | analysis, 2032. Fortune Business Insights | Global Market Research Reports & Consulting. URL: https://www.fortunebusinessinsights.com/comic-book-market-103903 (date of access: 05.05.2025).

[2] Webtoon: an annual report to security holders. Los Angeles, CA : WEBTOON Entertainment Inc., 2025. 152 p. URL: https://ir.webtoon.com/static-files/b98b7b60-c045-4f4b-8dce-3c424abaeade (date of access: 03.10.2023).

[3] Avidan S., Shamir A. Seam carving for content-aware image resizing. ACM transactions on graphics. 2007. Vol. 26, no. 3. P. 10. URL: https://doi.org/10.1145/1276377.1276390 (date of access: 05.05.2025).

[4] Li M., Lv T., Chen J., Cui L., Lu Y., Florencio D., Zhang C., Li Z., Wei F. TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models. arXiv preprint arXiv:2109.10282. 2021. URL: https://arxiv.org/abs/2109.10282 (date of access: 05.05.2025).

[5] Team LUCID. TrOCR-small-korean: Transformer-based OCR for Korean text recognition. Hugging Face. 2022. URL: https://huggingface.co/team-lucid/trocr-small-korean (date of access: 05.05.2025).

[6] Baek Y., Lee B., Han D., Yun S., Lee H. Character Region Awareness for Text Detection. arXiv preprint arXiv:1904.01941. 2019. URL: https://arxiv.org/abs/1904.01941 (date of access: 05.05.2025).

[7] Goh Yee Fong. ComiTranslate: empowering global readership through autotranslated comics and manga : Bachelor's Thesis. Faculty of Engineering and Science Universiti Tunku Abdul Rahman, Kajang, 2024. 85 p.

[8] Oghenekaro O. Comic-translate: A tool for translating text in comics. GitHub repository. 2022. URL: https://github.com/ogkalu2/comic-translate (date of access: 05.05.2025).

Дякую за увагу!

Q&A