

ВИЯВЛЕННЯ КОНТРАСТНИХ ОБ'ЄКТІВ З ВИКОРИСТАННЯМ КОМБІНАЦІЇ НАПІВКОНТРОЛЬОВАНОГО ТА НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

Степанчук Д.К.¹, Пишнограєв І.О.²

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна

¹ stepanchuk.dmytro@lil.kpi.ua, ² pyshnograiev@nas.gov.ua

Робота присвячена дослідженню методів напівконтрольованої детекції об'єктів (SSOD) для виявлення контрастних областей зображень в умовах обмеженої кількості розмічених даних. Проаналізовано еволюцію функцій втрат регресії обмежувальних рамок ($\text{IoU} \rightarrow \text{GIoU} \rightarrow \text{DIOU} \rightarrow \text{CIoU}$), механізми Teacher-Student фреймворків та архітектури сучасних детекторів (Faster R-CNN, YOLOv8, DETR). Обґрунтовано переваги індуктивного підходу SSOD для підвищення узагальнювальної здатності моделей. Мета роботи – аналіз теоретичного підґрунтя та вибір оптимальної комбінації архітектурних рішень для ефективної детекції контрастних об'єктів.

Ключові слова: напівконтрольоване навчання, детекція об'єктів, SSOD, Teacher-Student, CIoU, YOLOv8, DETR, кластеризація, навчання без учителя.

1. ВСТУП

Виявлення контрастних об'єктів – задача визначення областей зображення, що суттєво відрізняються від оточення за візуальними характеристиками – є фундаментальною проблемою комп'ютерного зору із застосуваннями у медичній діагностиці, промисловому контролі якості та автономній навігації. Класичні методи контрольованого навчання демонструють високу точність, але вимагають значних обсягів ретельно анотованих даних, що робить їх економічно неефективними.

Гіпотеза дослідження полягає в тому, що комбінування напівконтрольованого навчання (semi-supervised learning) з елементами неконтрольованого дозволить ефективно використовувати невелику кількість розмічених даних у поєднанні з великим обсягом нерозмічених зразків, підвищуючи узагальнювальну здатність моделі.

2. МЕТОДОЛОГІЯ ТА АРХІТЕКТУРА МОДЕЛЕЙ

2.1. Базові моделі

Розвиток функцій втрат для регресії обмежувальних рамок демонструє перехід від простого наближення до геометричної узгодженості (табл. 1):

Таблиця 1. Функції похибки для object detection

Функція	Геометричні фактори	Критичні недоліки
IoU	Перетин	Нульовий градієнт при відсутності перетину
GIoU	Перетин, охоплюючий контейнер	Повільна збіжність, деградація до IoU
DIoU	Перетин, відстань центрів	Ігнорує співвідношення сторін
CIoU	Перетин, відстань центрів, aspect ratio	Складність формули

CIoU Loss враховує три геометричні фактори та забезпечує найточнішу регресію. Додатково, Distributional Focal Loss (DFL) моделює невизначеність локалізації шляхом прогнозування розподілу координат.

2.2. Co-training: навчання з множинних поглядів

Co-training базується на припущенні, що вхідні дані можна розділити на два або більше наборів ознак (views), які є умовно незалежними за умови знання класу. Тренуються кілька моделей, кожна на своєму наборі ознак. Моделі співпрацюють, обмінюючись висококонфідентними прогнозами (псевдо-мітками) для нерозмічених даних, які додаються до розміченого набору. Цей ітеративний процес підвищує робастність системи.

У контексті детекції об'єктів природний поділ на незалежні види ознак є складним завданням. Сучасні адаптації, зокрема Multi-Head Co-Training, долають це обмеження через інтеграцію базових навчальних елементів у багатоголову структуру. Різноманітність досягається завдяки стратегії «Слабка та Сильна Аугментація»: одна модель отримує слабо аугментовані зображення (flip, crop), інша – сильно трансформовані (color jitter, cutout, mixup). Це запобігає колапсу моделей до ідентичних результатів та забезпечує взаємне збагачення через обмін псевдо-мітками.

2.3. Архітектури детекторів

Двоетапні детектори (Cascade R-CNN) – використовують каскад детекторів з прогресивно вищими порогами IoU для ітеративного уточнення рамок.

Одноступеневі детектори (YOLOv8) – прогнозують класи та рамки за один прохід. Архітектура включає блоки C2f (оптимізація градієнтного потоку) та SPPF (масштабна інваріантність). Використовує BCE + CIoU + DFL.

Трансформерні детектори (DETR) – End-to-End підхід без NMS та якорів. Угорський алгоритм забезпечує двобічне узгодження, але призводить до розрідженої супервізії та повільної збіжності (~500 epoch).

3. ПАЙПЛАЙН ІТЕРАТИВНОГО НАВЧАННЯ

3.1. Етап 1: Неконтрольована авторозмітка

Початковий етап експлуатує внутрішню властивість контрастних об'єктів – їх візуальну відмінність від оточення. Для автоматичного виявлення таких зон застосовується кластеризація у 5-вимірному просторі ознак:

$$\mathbf{F} = (\mathbf{x}, \mathbf{y}, \mathbf{R}, \mathbf{G}, \mathbf{B}),$$

де $(\mathbf{x}, \mathbf{y})$ – просторові координати пікселя, $(\mathbf{R}, \mathbf{G}, \mathbf{B})$ – значення кольорових каналів. Алгоритм кластеризації групує пікселі за евклідовою відстанню у цьому просторі, автоматично виділяючи компактні області з подібними просторово-колірними характеристиками.

Кластери, що суттєво відрізняються від домінуючого фону за кольором або просторовим розташуванням, ідентифікуються як потенційні контрастні зони.

3.2. Етап 2: Ініціалізація базового детектора

На основі мінімального розміченого набору (20 зображень з ручною анотацією) навчається базовий детектор. Цей детектор застосовується до зон, виявлених на етапі неконтрольованої кластеризації, генеруючи початкові псевдо-мітки з класифікацією та уточненими bounding boxes. Таким чином, неконтрольоване навчання забезпечує локалізацію потенційних об'єктів, а базовий детектор – їх класифікацію та геометричне уточнення.

3.3. Етап 3: Напівконтрольоване навчання

На отриманому наборі псевдо-міток паралельно навчаються кілька мереж із застосуванням **co-training** з різними аугментаціями для забезпечення різноманітності моделей. Моделі обмінюються високоякісними псевдо-мітками, взаємно покращуючи узагальнювальну здатність.

3.4. Етап 4: Відбір порогу впевненості та ручна дорозмітка

Після завершення раунду навчання емпірично визначається оптимальний поріг впевненості моделей. Зразки поділяються на три категорії:

- **Висока впевненість** ($p > \tau_{\text{high}}$) – псевдо-мітки приймаються автоматично;
- **Низька впевненість** ($p < \tau_{\text{low}}$) – зразки відкидаються як фонові;
- **Невизначені випадки** ($\tau_{\text{low}} \leq p \leq \tau_{\text{high}}$) – направляються на ручну дорозмітку експертом.

Така стратегія активного навчання концентрує людські зусилля виключно на складних, неоднозначних випадках, максимізуючи ефективність анотації.

3.5. Ітеративний цикл

Етапи 3–4 повторюються ітеративно: кожен новий раунд навчання використовує розширений набір даних (автоматичні псевдо-мітки + ручна дорозмітка складних випадків). З кожною ітерацією модель покращується, зменшуючи кількість невизначених випадків та потребу в ручному втручанні.

3.6. Валідація

Оцінка якості моделі проводиться двома комплементарними методами:

- **Крос-валідація між епохами** – порівняння прогнозів поточної моделі з моделями попередніх ітерацій для виявлення регресії або нестабільності;
- **Експертне оцінювання** – вибіркова перевірка результатів детекції людиною-експертом для контролю якості та виявлення систематичних помилок.

4. ВИСНОВОК

Дослідження обґрунтовує, що успіх SSOD критично залежить від якості локалізації Teacher-моделі. Еволюція функцій втрат від GloU до CIoU та DFL забезпечує геометричну точність та моделювання невизначеності, необхідні для генерації нешумних псевдо-міток.

Встановлено архітектурний компроміс: трансформерні детектори (DETR) пропонують теоретично елегантний End-to-End підхід, однак одноступеневі CNN-детектори (YOLOv8) залишаються кращим вибором для SSOD через ефективність тренування та високу швидкість інференсу.

Подальші дослідження спрямовані на практичну реалізацію гібридної системи, що поєднує co-training фреймворк з механізмами кластеризації для адаптивного виявлення контрастних об'єктів.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Liu Y.-C. et al. Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors. CVPR, 2022.
2. Xu M. et al. End-to-End Semi-Supervised Object Detection with Soft Teacher. ICCV, 2021.
3. Zheng Z. et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. AAAI, 2020.
4. Li X. et al. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. NeurIPS, 2020.
5. Carion N. et al. End-to-End Object Detection with Transformers (DETR). ECCV, 2020.
6. Ultralytics. YOLOv8 Documentation. 2023. URL: <https://docs.ultralytics.com/>
7. Ren S. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NeurIPS, 2015.