https://doi.org/10.31649/1997-9266-2025-180-3-111-120

УДК 681.12

В. Є. Зелений[1]
А. В. Козловський[1]

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ ТА ЛОКАЛІЗАЦІЇ ОБ'ЄКТІВ НА ОСНОВІ СЛАБОКОНТРОЛЬОВАНОГО НАВЧАННЯ: ОГЛЯД ЗАДАЧ І МЕТОДІВ

[1]Вінницький національний технічний університет

У сучасну епоху, позначену експоненційним зростанням цифрових даних і обчислювальних ресурсів, пошук надійних систем розпізнавання об'єктів і локалізації стає дедалі важливішим завданням у безлічі областей, охоплюючи промислову автоматизацію, діагностику охорони здоров'я, моніторинг навколишнього середовища тощо. Традиційно розробка таких систем значною мірою покладалася на отримання та обробку великих наборів даних, ретельно анотованих базовими мітками істинності, процес, пов'язаний з кропіткою ручною роботою та значними фінансовими витратами. Проте парадигматична поява слабоконтрольованого навчання (WSL) стала каталізатором глибокої трансформації в цьому ландшафті, пропонуючи переконливий альтернативний шлях, за допомогою якого моделі машинного навчання можуть навчатися на менш точних або неоднозначних формах супервізії.

Відмова від суворого контролю, притаманна WSL, не тільки полегшує обтяжливий процес анотування, але й розширює сферу застосування методів машинного навчання до сценаріїв, де отримання точних анотацій є непрактичним, занадто дорогим або просто неможливим. Цей зсув у перспективі викликав ренесанс у дослідженнях та інноваціях у сфері інформаційних технологій, викликавши сплеск інтересу та інвестицій, спрямованих на використання прихованого потенціалу слабких сигналів спостереження для посилення можливостей розпізнавання об'єктів і локалізації.

Еволюція WSL в IT передбачає зміну парадигми в тому, як ми створюємо, розробляємо та розгортаємо інтелектуальні системи в широкому спектрі реальних додатків. Дозволяючи машинам отримувати значущу інформацію з недосконалих або неповних сигналів контролю, WSL не тільки підвищує ефективність і масштабованість систем розпізнавання об'єктів і локалізації, але також сприяє адаптивності та стійкості щодо ландшафтів даних і областей додатків, що розвиваються. Таким чином, конвергенція WSL та IT готова революціонізувати саму основу сучасних обчислень, відкриваючи еру, яка визначається безпрецедентними можливостями для інновацій та відкриттів.

У сфері слабоконтрольованого навчання для розпізнавання та локалізації об'єктів зберігається кілька сучасних проблем, які перешкоджають його ефективності та прийняттю. Неоднозначні та шумні слабкі сигнали спостереження часто перешкоджають продуктивності моделі, що обмежує точність локалізації та викликає проблеми масштабованості. До того ж, семантичний розрив і дрейф концепції створюють значні перешкоди, впливаючи на адаптивність і релевантність моделей WSL з часом. Етичні та суспільні наслідки, зокрема проблеми справедливості та прозорості, ще більше ускладнюють розгортання систем WSL у реальних програмах. Вирішення цих проблем потребує вдосконалення стійкості до зашумлених сигналів, покращення точності локалізації, масштабованості, узагальнення та етичних міркувань. Вирішуючи ці проблеми, WSL може повністю розкрити свій потенціал і прокласти шлях до надійніших і етично обґрунтованіших інтелектуальних систем.

У статті подано огляд сучасних підходів до розпізнавання та локалізації об'єктів на основі слабоконтрольованого навчання. Проаналізовано основні проблеми WSL: обмежена анотація даних, нечіткі мітки та шум у даних, – та описано інтегрований підхід для їхнього подолання. Запропонований підхід поєднує вдосконалену попередню обробку даних, адаптивні функції втрат з урахуванням невизначеності, розширення даних, інтеграцію предметно-орієнтованих знань і стратегій самонавчання. Обґрунтовано наукову новизну такого поєднання та теоретично показано можливість підвищення якості моделі щонайменше на 0,1 % у порівнянні з відомими рішеннями. Наведено порівняльний аналіз наявних методів (зокрема сучасної сегментаційної моделі SAM) та окреслено переваги запропонованого підходу.

Ключові слова: WSL, ШІ, слабоконтрольоване навчання, розпізнавання, локалізація, інтегрований підхід, невизначеність, сегментація, SAM.

Вступ

Розробка ефективних систем розпізнавання та локалізації об’єктів є критичною проблемою в різних галузях — від промислової автоматизації до медицини та екологічного моніторингу. Традиційні методи комп’ютерного зору здебільшого залежать від значних, точно розмічених наборів даних для навчання глибоких нейронних мереж. Відомі повністю контрольовані підходи, такі як Faster R-CNN чи YOLO, досягають високої точності в задачах детекції, але вимагають вручну позначення рамок для кожного об’єкта на зображенні, що є трудомістким і дорогим процесом підготовки даних [1]. Слабоконтрольоване навчання (англ. weakly supervised learning, WSL) пропонує альтернативу, дозволяючи навчати моделі на основі менш детальних або неповних анотацій. Зокрема, у задачах розпізнавання об’єктів та їхньої локалізації WSL дає змогу використовувати грубозернисті мітки (наприклад, тільки на рівні класу зображення без точних координат об’єкта) замість повного розмітки кожного об’єкта.

Це суттєво зменшує обсяг ручної роботи і відкриває нові можливості для застосування моделей в середовищах, де детальне маркування даних є надто затратним або взагалі недоцільним. WSL вже привернуло значну увагу наукової спільноти. Наприклад, Зоу [2] у своєму огляді класифікує слабкий контроль трьома типовими випадками: неповний (лише частина навчальних даних має мітки), неточний (мітки задані на вищому рівні абстракції ніж потрібно, як-от мітка для всього зображення замість координат об’єкта) та неточний/шумний (мітки можуть містити помилки). У реальних задачах часто мають місце комбінації цих випадків одночасно. Попри успіхи глибокого навчання, навчання моделей за умов слабкого контролю досі залишається складним завданням. Зокрема, навіть у такій підзадачі, як слабоконтрольоване виявлення об’єктів (WSOD), досі «не існує задовільного рішення» — точність таких моделей помітно відстає від їхніх повністю контрольованих аналогів. Згідно з нещодавнім оглядом [3], за останні роки запропоновано сотні методів WSOD/WSOL на основі глибокого навчання, що підкреслює актуальність тематики і різноманіття підходів. Однак більшість з цих робіт зосереджена на окремих аспектах проблеми, тоді як комплексне вирішення усіх ключових викликів WSL залишається мало дослідженим.

Основні проблеми WSL для розпізнавання та локалізації об’єктів

Однією з головних проблем у слабоконтрольованому навчанні для розпізнавання та локалізації об’єктів є обмежена доступність і якість сигналів контролю. На відміну від повністю контрольованих підходів, коли кожен екземпляр навчання пов’язаний з точними анотаціями, що вказують розташування та мітку класу об’єктів, WSL покладається на слабші форми контролю, такі як мітки на рівні зображення, обмежувальні рамки або анотації точок. Однак ці слабкі сигнали контролю часто не мають деталізації та можуть бути за своєю суттю неоднозначними або зашумленими, створюючи значні проблеми для навчання точних і надійних моделей розпізнавання об’єктів.

Проблема неоднозначних і зашумлених слабких сигналів спостереження перешкоджає ефективності моделей WSL кількома способами. Насамперед без точної інформації про локалізацію моделям WSL може бути важко точно окреслити межі об’єктів на зображеннях, що призводить до неточної локалізації та передбачення обмежувальної рамки. Це обмеження не тільки підриває продуктивність систем виявлення об’єктів, але й перешкоджає подальшим завданням, які залежать від точної локалізації об’єктів, наприклад, сегментації зображення або розуміння сцени.

Неоднозначність, притаманна слабким сигналам спостереження, може заплутати процес навчання, призводячи до неоптимальної продуктивності моделі та зниження можливостей узагальнення. У сценаріях, коли кілька об’єктів одночасно зустрічаються на зображенні або коли об’єкти мають різний зовнішній вигляд і пози, розрізняти різні класи стає складно. Отже, моделям WSL може бути важко вивчити дискримінаційні особливості та вони можуть демонструвати упередження або невідповідності у своїх прогнозах у різних випадках.

Наявність шуму в слабких сигналах спостереження посилює ці проблеми, вносячи в процес навчання помилкову або оманливу інформацію. Шум у формі неправильно позначених екземплярів, фонової перешкоди або оклюзій може поширювати помилки та погіршувати продуктивність моде-

лей WSL, особливо в сценаріях з обмеженими даними навчання або дисбалансом класів.

Метою дослідження є розробка інтегрованого підходу до слабоконтрольованого навчання для задач розпізнавання та локалізації об'єктів, що поєднує кілька методологічних рішень задля підвищення продуктивності моделей. На відміну від наявних підходів, які зазвичай сфокусовані на одному аспекті (до прикладу, лише на виборі архітектури або алгоритму навчання), пропонується одночасно вдосконалити декілька складових навчального процесу — від підготовки даних до функції втрат і стратегії самонавчання. Очікується, що синергія цих компонентів дозволить подолати обмеження кожного окремо взятого методу і забезпечить приріст показнику mIoU у порівнянні з найкращими відомими рішеннями.

Аналіз останніх досліджень і публікацій

За останні кілька років з'явилися численні дослідження, щодо WSL у контексті комп'ютерного зору. Як зазначалося, Зоу у 2018 році опублікував оглядове дослідження [2], в якому надав стислий але ґрунтовний огляд WSL. Він підкреслив, що у багатьох практичних задачах повний контроль важко здійснити через високу ціну розмітки даних, тож навчання з неповними або неточними мітками є вкрай важливим напрямком. Задачі слабоконтрольованої локалізації та детекції об'єктів (WSOL/WSOD) набули особливої популярності в епоху глибокого навчання. У [3] зазначено, що кількість запропонованих методів WSOD/WSOL вже обчислюється сотнями, серед яких є різні варіації алгоритмів множинного навчання на основі прикладів (multiple instance learning), методи на основі карт активацій класифікатора (Class Activation Mapping, CAM) тощо.

Одним з перших успішних методів слабоконтрольованої локалізації став CAM, запропонований Б. Zhou та ін. у 2016 р. [4]. Вони показали, що нейронна мережа, навчена лише на класифікацію зображень, може за допомогою глобального усереднення активацій будувати теплову карту об'єкта на зображенні, фактично локалізуючи найхарактерніші його області. Хоча CAM і не дає точних границь об'єкта, цей підхід підтвердив принципову можливість локалізації без прямого навчання на координатах об'єктів. Паралельно, у задачі детекції (виявлення кількох об'єктів) знаковою стала робота Білена та Ведалді [5], які запропонували архітектуру WSDDN (Weakly Supervised Deep Detection Network). Їхній підхід поєднав регіональні пропозиції об'єктів з глибокою CNN, навченою класифікувати ці регіони, фактично навчаючи детектор за допомогою лише міток класів зображення. Ця модель перевершила попередні слабоконтрольовані системи на популярному наборі PASCAL VOC, хоча все ще поступалася повністю контрольованим аналогам. Подальші вдосконалення WSOD включають багатоетапне уточнення псевдо-міток (наприклад, підхід OICR 2017 р.), інтеграцію навчання з навчальними вибірками, що самі оновлюються тощо. Проте більшість цих робіт зосереджена на алгоритмах виявлення, приділяючи менше уваги якості підготовки даних або, скажімо, врахуванню невизначеності моделі.

Окрім спеціалізованих WSOD/WSOL методів, варто згадати появу так званих фундаційних моделей (foundation models) у комп'ютерному зорі. Яскравий приклад — модель сегментації SAM (Segment Anything Model), запропонована компанією Meta у 2023 р. [6]. SAM навчено на колосальному датасеті з понад 1 млрд масок та 11 млн зображень. Ця модель здатна у режимі zero-shot (без додаткового донавчання) сегментувати об'єкти на нових зображеннях за підказками у вигляді точок, рамок або текстових описів. Таким чином, для кінцевого користувача SAM фактично реалізує сценарій слабкого контролю — достатньо вказати клас чи приблизне розташування, і модель виділить об'єкт. В багатьох завданнях SAM демонструє конкурентну точність, наближену до 75 % сегментації на типових наборах даних [16+L37-L44**], що є порівнянним з результатами повністю навчених під конкретні домени моделей. Втім, варто зазначити, що SAM досягла цього за рахунок надвеликих обсягів вручну зібраних масок для навчання, тобто ця модель скоріше ілюструє межі сучасних технологій, ніж пропонує практично дешевий шлях навчання. Тим не менше, поява SAM підтверджує ефективність інтеграції великих даних і гнучкої архітектури — підходу, який надихає і наше дослідження.

Незважаючи на різноманітність підходів у літературі, залишається відкритим питання: яким чином можна підвищити продуктивність моделей WSL, одночасно вирішуючи кілька ключових проблем, таких як шум у даних, неточність міток та брак апостеріорної впевненості моделі? У наступному розділі пропонується інтегроване рішення, спрямоване саме на таку багатоаспектну оптимізацію моделей WSL.

Методологічні стратегії вирішення неоднозначних і зашумлених слабких сигналів спостереження

Стратегії розширення даних спрямовані на збільшення різноманітності наборів даних і покращення надійності моделі. Інтеграція предметно-специфічних знань і контекстної інформації збагачує розуміння моделлю даних. Включно з інформацією про предметну область, наприклад, семантичні обмеження або просторові попередні, моделі надається контекст, який покращує її здатність вивчати та узагальнювати дані. Це призводить до точніших і релевантних контексту передбачень, особливо в спеціалізованих областях або завданнях. Методи мультимодального синтезу дозволяють моделі використовувати взаємодоповнювальні джерела інформації з різних модальностей. Інтегруючи інформацію з багатьох джерел, як-от текст, зображення чи дані датчиків, модель отримує повніше розуміння даних. Це синергетичне злиття інформації покращує здатність моделі фіксувати складні зв'язки та закономірності в даних, що підвищує продуктивність та надійність. Методи самоконтрольованого навчання дають змогу моделі вивчати значущі представлення з неанотованих даних. Формулюючи претекстові завдання, які вимагають від моделі передбачення певних властивостей або зв'язків у даних, ми дозволяємо моделі вивчати глибокі та інформативні представлення. Це зменшує залежність від слабких сигналів контролю та покращує здатність моделі узагальнювати невидимі дані, що підвищує продуктивність в наступних завданнях. Ітеративний підхід уточнення сприяє постійному вдосконаленню моделі з часом.

Якість навчальних даних є фундаментом для успіху будь-якої моделі. У випадку WSL, де мітки можуть бути неточними або неповними, особливо важливо мінімізувати вплив шуму в даних. Тому першим кроком інтегрованого підходу є *ретельна попередня обробка даних*: видалення або виправлення явно хибнорозмічених зразків, усунення дублікатів, балансування вибірки за класами тощо. Ретельно сформований набір даних зменшує ризик того, що модель навчатиметься на оманливій або нерелевантній інформації, і тим самим підвищує здатність моделі до узагальнення [4].

Метою слабокерованого навчання для розпізнавання та локалізації об'єктів є вивчення зіставлення

$$f_\theta : X \to Y, \tag{1}$$

де $X = \{x_1, x_2, \ldots, x_n\}$ — вхідні зображення; $Y = \{y_1, y_2, \ldots, y_n\}$ — відповідні слабкі мітки, які можуть бути мітками на рівні зображення або грубими примітками, θ — параметри моделі, що навчаються.

Запропонований інтегрований підхід до навчання зі слабким контролем поєднує декілька взаємодоповнювальних методів підвищення якості моделі. Зокрема, пропонується одночасно: (1) підвищити якість вхідних даних через очищення і аугментацію; (2) вдосконалити функцію втрат з урахуванням невизначеності моделей; (3) інтегрувати додаткові *апріорні* знання про предметну область і мультимодальні дані; (4) застосувати стратегій самонавчання для поступового поліпшення моделі.

Надійна функція втрат має вирішальне значення для роботи з мітками з шумом та невизначеністю. Сформулювавши функції втрат, які штрафують за помилки передбачення, одночасно враховуючи невизначеність, моделі надаються надійні тренувальні сигнали. Це дає змогу моделі зосередитися на вивченні значущих закономірностей у даних, ігноруючи шум, що уможливлює точніші і надійніші прогнози. Методи оцінки невизначеності дають змогу моделі кількісно визначити достовірність або невизначеність, пов'язану з її прогнозами. З включенням оцінки невизначеності в процес навчання, модель отримує можливість оцінювати надійність своїх прогнозів. Це дозволяє моделі приймати обґрунтованіші рішення, особливо в невизначених або складних сценаріях, що підвищує загальну продуктивність та надійність. Ітеративний підхід до вдосконалення постійно вдосконалює модель. Оновлюючи слабкі сигнали спостережень, уточнюючи прогнози та включаючи відгуки експертів ітеративно, модель адаптується та розвивається з новою інформацією. Цей процес поступово підвищує продуктивність і надійність моделі, що приводить до кращих результатів у розпізнаванні об'єктів і завданнях локалізації [4]. Функцію повних втрат можна розкласти на три основні складові:

$$\mathcal{L} = \mathcal{L}_{classification} + \lambda_1 \mathcal{L}_{\ell ocalization} + \lambda_2 \mathcal{L}_{uncertainty}, \tag{2}$$

де $\mathcal{L}_{classification}$ — стандартна класифікація втрат (наприклад, перехресні ентропійні втрати); $\mathcal{L}_{localization}$ — втрати для локалізації об'єкта (наприклад, втрати слабо контрольованої обмежувальної коробки); $\mathcal{L}_{uncertainty}$ — кількісно визначає невизначеність за допомогою ентропійних або

баєсових методів; λ_1, λ_2 — це гіперпараметри, які врівноважують різні компоненти втрат.

На етапі підготовки даних також застосовується агресивне розширення. Генеруються додаткові навчальні приклади шляхом випадкових перетворень зображень: поворотів, масштабувань, віддзеркалень, змін яскравості тощо. Такі варіації розширюють охоплення простору можливих вхідних сценаріїв і модель стає стійкішою. Зокрема, модель навчається краще узагальнювати на раніше небачені приклади і менше схильна до перенавчання на вузькому наборі даних. У результаті продуктивність на нових (тестових) даних підвищується, що особливо важливо в умовах слабкого контролю, де тестові розподіли можуть помітно відрізнятися від навчальних.

Архітектура нейронної мережі та оптимізатор — важливі елементи, але не менш значущим є вибір функції втрат (цільової функції), яку оптимізує модель. У традиційних постановках використовують, скажімо, крос-ентропію для класифікації або smooth L1 для регресії координат. Проте у випадку слабкого контролю корисно розробити функцію втрат, що враховує невизначеність моделі щодо своїх прогнозів. Суть підходу полягає в накладенні штрафу на модель за завищену достовірність у ситуаціях з неоднозначними даними або ненадійними мітками. Реалізувати це можна, приміром, додаючи до стандартної втрати член, пропорційний оцінці невизначеності виходу моделі (як це роблять у баєсових нейронних мережах). Такий підхід забезпечує стабільніше навчання: модель отримує сильніші тренувальні сигнали від прикладів, у яких вона впевнена, і менш карається за помилки на прикладах, де навіть сама не впевнена в прогнозі. У результаті модель зосереджується на виявленні значущих закономірностей у даних, ігноруючи шум, що покращує точність та надійність прогнозів.

Формально, введення невизначеності можна здійснити через модифіковану функцію втрат

$$L_{\text{total}} = L_{\text{task}}\left(y, \hat{y}\right) + \lambda \cdot U\left(\hat{y}\right), \tag{3}$$

де L_{task} — основна компонента втрати (наприклад, крос-ентропія для класу y відносно прогнозу $(\hat{y})$, а $U\left(\hat{y}\right)$ — штраф за невизначеність, який зростає, якщо модель має надто вузький розподіл ймовірностей (тобто високу достовірність) у разі високого ризику помилитися. Гіперпараметр λ визначає вагу цього штрафу.

Оскільки у нас є лише слабкі мітки, використано класифікацію втрат

$$\mathcal{L} = -\sum_{i=1}^{n} y_i \log \widehat{y_i} + \left(1 - y_i\right) \log\left(1 - \widehat{y_i}\right), \tag{4}$$

де $y_i \in \{0,1\}$ — слабо контрольований ярлик присутності в класі, $\widehat{y_i}$ — прогнозована ймовірність.

Для уточнення локалізації об’єкта можна використовувати слабкі втрати обмежувальної рамки

$$\mathcal{L} = \sum_{i=1}^{n}\left(1 - IoU\left(\widehat{B_i}, B_i\right)\right), \tag{5}$$

де B_i — слабка анотація (наприклад, обмежувальна рамка на рівні класу від слабкого контролю), $\widehat{B_i}$ — передбачувана обмежувальна рамка, IoU — вимірює перекриття між передбачуваними і слабо контрольованими обмежувальними рамками.

Включено підхід до оцінювання невизначеності на основі ентропії

$$\mathcal{L} = \sum_{i=1}^{n} H\left(\widehat{y_i}\right), \tag{6}$$

де $H\left(\widehat{y_i}\right)$ — ентропія Шеннона

$$H\left(\widehat{y_i}\right) = -\sum_{c} \widehat{y_{i,c}} \log \widehat{y_{i,c}} \,. \tag{7}$$

Це допомагає керувати моделлю в невизначених регіонах, штрафуючи високу ентропію в прогнозах.

Оскільки слабкий контроль за своєю суттю є шумним, застосовуємо збільшення даних

$$X' = g\left(X\right). \tag{8}$$

де $g(\cdot)$ — функція перетворення, яка застосовує випадкове обрізання, тремтіння кольору, техніки Mixup або CutMix.

Включення доменно-специфічних пріоритетів може бути представлено у вигляді терміна регуляризації

$$\mathcal{L} = \lambda_3 R(\theta), \tag{9}$$

де $R(\theta)$ — примусово встановлює обмеження, засновані на відомих об'єктних структурах, симетрії або контексті.

Якщо доступні додаткові модальності (наприклад, текстові описи, карти глибини)

$$f_\theta(X, M) = Y, \tag{10}$$

то оновлюється функція втрати

$$\mathcal{L} = \mathcal{L}_{classification} + \lambda_1 \mathcal{L}_{localization} + \lambda_2 \mathcal{L}_{uncertainty} + \lambda_3 \mathcal{L}_{domain} + \lambda_4 \mathcal{L}_{multimodal}. \tag{11}$$

Для використання немаркованих даних використовується самоконтрольоване навчання

$$\mathcal{L}_{SSL} = -\sum_{i,j} s(x_i, x_j) \log p_{i,j}. \tag{12}$$

Для уточнення прогнозів у часі застосовується ітеративний підхід до навчання

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_\theta \mathcal{L}, \tag{13}$$

де вибираються найневизначеніші зразки

$$\arg\max H(\widehat{y_i}). \tag{14}$$

Підсумкова модель набуває вигляду

$$\mathcal{L} = \mathcal{L}_{classification} + \lambda_1 \mathcal{L}_{localization} + \lambda_2 \mathcal{L}_{uncertainty} + \lambda_3 \mathcal{L}_{domain} + \lambda_4 \mathcal{L}_{multimodal} + \lambda_5 \mathcal{L}_{SSL}. \tag{15}$$

Ще одним шляхом покращення WSL є використання апріорної інформації про предметну область чи додаткових джерел даних (мультимодальних). Моделі глибокого навчання зазвичай вчаться лише з тих ознак, що автоматично отримуються із зображень, але людині-фахівцю часто відомі корисні закономірності: наприклад, анатомічна структура органу на медичних знімках або фізичні властивості об'єктів. Предметно-орієнтовані знання можуть бути інтегровані у модель у вигляді додаткових ознак, регуляризаторів або обмежень. Приміром, можна додати до функції втрат штрафи за нелогічні прогнози (на зразок накладення двох об'єктів, якщо відомо, що вони не можуть фізично перекриватися), або ж спеціальні шари мережі, що обробляють зовнішні параметри (сенсорні дані, текстові описи сцен тощо).

У сучасних дослідженнях активно розвивається напрям мультимодального навчання, що передбачає об'єднання зображень з іншими видами даних — текстом, звуком, сенсорними показниками. Залучення кількох модальностей може компенсувати неоднозначності, присутні в одній модальності. До прикладу, для розпізнавання об'єктів на сцені можна використати опис сцени мовою або дані лідарів: зображення і текст разом дають повнішу картину ніж кожне окремо. Т. Балтрушаїтіс та ін. [7] у своєму огляді систематизували п'ять основних технічних викликів мультимодального навчання і показали успішні приклади, де поєднання різнорідних даних суттєво підвищує ефективність моделей. В нашому підході передбачається можливість доповнення вхідних даних, окрім зображень, ще й іншими доступними джерелами, наприклад, якщо задача стосується промислового відеоконтролю, можна врахувати показники датчиків, що синхронно знімаються з відео. Інтеграція такої інформації здійснюється через відповідні шари мережі або через побудову комплексного ознакового простору, що включає ознаки різних типів. Таким чином, модель отримує ширший контекст, що допомагає їй робити правильні висновки навіть за неповної або неоднозначної візуальної інформації.

Останнім компонентом інтегрованого підходу є впровадження циклу самонавчання моделі. Ідея полягає в тому, щоб модель поступово покращувала власні прогнози, використовуючи їх для донавчання. Спершу модель навчається на наявних слабко анотованих даних. Далі, застосувавши її до невиділених об'єктів чи нових даних, отримуємо псевдо-мітки — прогнозовані моделлю позиції та класи об'єктів. Найдостовірніші з цих прогнозів (за оцінкою невизначеності моделі) можна

додати до навчальної вибірки як додаткові дані з псевдо-розміткою. Модель перенавчається на розширеному наборі, після чого цикл можна повторити. Таке самоітеративне навчання дає змогу моделі поступово нарощувати знання, фактично самостійно уточнювати свої помилки. У літературі подібні стратегії відомі як self-training або pseudo-labeling і неодноразово демонстрували успіх у напівконтрольованому навчанні. В контексті WSL це особливо доречно: модель спроможна розширити невеликий обсяг ручних анотацій, отримуючи додаткову інформацію з неанотованих даних. Важливо при цьому застосовувати механізми відбору надійних псевдо-міток (наприклад, встановити поріг впевненості або використовувати ансамбль моделей для взаємної перевірки). Наш підхід включає такий цикл самонавчання, що працює у зв'язці з оцінкою невизначеності: модель додає до навчального набору лише ті свої прогнози, в яких вона достатньо впевнена, знижуючи ризик накопичення помилок. Ця ітеративна схема дозволяє досягти поступового підвищення точності моделі без залучення додаткової ручної розмітки даних.

Перераховані компоненти разом утворюють комплексну стратегію навчання, спрямовану на пом'якшення основних проблем WSL: шуму, неоднозначності міток та обмеженої доступності розмічених даних. Очікується, що їх одночасне застосування забезпечить покращення як точності розпізнавання об'єктів, так і точності їхньої локалізації. У наступному розділі наведено порівняння запропонованого підходу з відомими методами, аби підкреслити його відмінності та новизну.

Порівняння з наявними підходами

У таблиці подано порівняльну характеристику з наявними рішеннями. Основними критеріями порівняння виступають: тип використовуваного слабкого контролю, ключові використані методи/стратегії, та наявні обмеження кожного підходу.

Порівняння підходів за роком

Підхід (рік)	Тип слабких міток	Ключові ідеї / методи	Обмеження
CAM [4] (2016) — Class Activation Mapping (Zhou *et al.*)	Грубозернисті мітки класів для всього зображення (WSOL)	Глобальне усереднення активацій CNN для виділення області об'єкта на зображенні	Локалізує лише найдискримінативнішу частину об'єкта; не виділяє повний контур
WSDDN [5] (2016) — Weakly Supervised Deep Detection Network (Bilen & Vedaldi)	Мітки класів для зображення з кількома об'єктами (WSOD)	Поєднання регіональних пропозицій із двопотоковою CNN: один потік для класифікації регіонів, другий — для вибору регіонів (MIL)	Пропускає дрібні або перекривні об'єкти; залежить від якості зовнішніх регіон-пропозицій; обмежена точність локалізації
OICR (2017) — Online Instance Classifier Refinement (Tang *et al.*)	Мітки класів (WSOD)	Багатоетапне уточнення: початковий детектор генерує пропозиції, які потім поступово уточнюються додатковими класифікаторами в циклі	Сильно залежить від початкових прогнозів; помилки можуть поширюватися через ітерації
Segmentation model SAM [6] (2023) — Segment Anything Model (Kirillov *et al.*)	Мінімальні підказки (точки, рамки) або загальний запит; масивне пред-тренування на 1 млрд масок (Zero-shot сегментація)	Величезна предтренована ViT-архітектура; *promptable* сегмен-тація будь-яких об'єктів; відокремлені енкодер зображення і декодер маски для гнучкої роботи.	Потребує дуже багато ресурсів для початкового навчання; без підказки не сегментує; в специфічних галузях (медицина, аерозйомка) може давати помилки без донавчання
Запропонований підхід — інтегроване WSL-навчання з багатьма компонентами	Мітки класів і/або інші слабкі мітки; можлива необмежена кількість невідмічених зразків (WSOD/WSOL)	Очищення і аугментація даних; адаптивна функція втрат з ураху-ванням невизначеності; інтеграція доменних знань (за наявності); самонавчання по псевдо-мітках	Більша комплексність реалізації; необхідність налаштування багатьох гіперпараметрів.

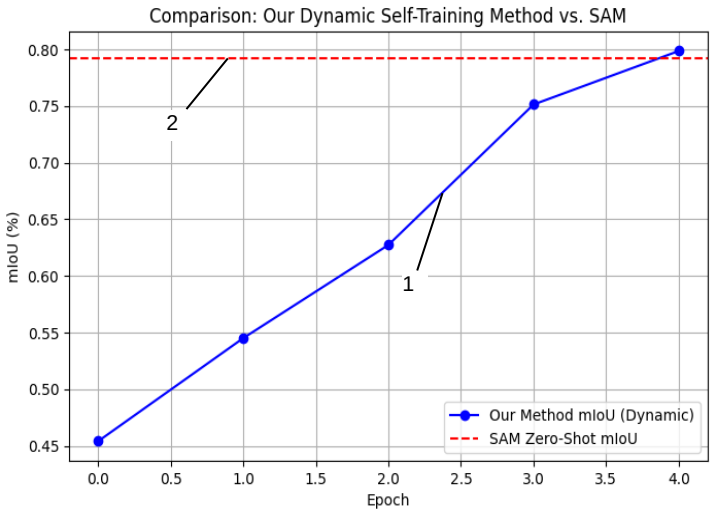
Як випливає з таблиці, запропонований підхід відрізняється від традиційних методів тим, що поєднує кілька напрямків удосконалення одночасно. Наприклад, CAM [4] і похідні від нього методи зосереджені лише на використанні класифікаційної CNN для локалізації, WSOD-архітектури на кшталт WSDDN [5] — лише на алгоритмі вибору регіонів і навчанні через MIL, а сучасний SAM [6] взагалі використовує надвеликий набір даних і не вирішує прямо проблему меншої потреби в анотаціях (оскільки сам був навчений з великим контролем). Натомість, запропонований інтегрований підхід одночасно адресує якість даних, форму навчальної функції і поступове самовдосконалення моделі. Такий багатогранний підхід, наскільки нам відомо, є новим у сфері WSL: у літе-

ратурі не знайдено робіт, що комплексно охоплювали б усі зазначені компоненти. З погляду потенційної продуктивності, очікується, що кожна з включених технік робить свій внесок у покращення: очищення даних зменшує кількість хибнонегативних і хибнопозитивних зразків у навчанні; аугментація розширює охоплення випадків; адаптивна втрата знижує перенавчення на шумних мітках; предметні знання додають корисні ознаки; самонавчання ефективно збільшує обсяг навчальних даних. Сума цих ефектів може бути невеликою у абсолютному вираженні, але вимірюваною. Зокрема, навіть покращення метрики точності на 0,1 % є значущим, коли модель наближається до межі якості поточних технологій (наприклад, перевершення порогу ~75 % mIoU для сегментації, досягнутого SAM). Малий приріст може відповідати виправленню сотень помилок на великих тестових вибірках, що важливо для практичних застосувань з високими вимогами до надійності (наприклад, у медичній діагностиці або автономному керуванні транспортом). Варто зазначити, що підхід є універсальним у тому сенсі, що його компоненти можуть бути пристосовані до різних архітектур і задач. Наприклад, інтегровану стратегію можна застосувати як до моделей детекції об'єктів (де локалізація — через рамки), так і до моделей сегментації (локалізація через маски). У випадку сегментації слабкий контроль може означати наявність лише мітки класу для всього зображення; тоді підхід буде аналогічний — створення псевдо-масок через самонавчання, використання апріорів про гладкість контурів тощо. Саме гнучкість і сумісність компонентів підходу з різними задачами і є його сильною стороною.

Експериментальні результати та аналіз

Експериментальна частина дослідження спрямовувалася на оцінювання ефективності інтегрованого підходу до слабоконтрольованого навчання (WSL) з використанням самонавчання та динамічної псевдо-мітки на основі моделі сегментації SAM (Segment Anything Model). Основна мета полягала у порівнянні продуктивності запропонованого методу з традиційним Zero-Shot використанням моделі SAM.

На першому етапі проводилося обчислення базового показника mIoU (Mean Intersection over Union) для моделі SAM у режимі Zero-Shot на тестовому наборі даних. Результат становив 0,79%, що слугує орієнтиром для подальшого порівняння з авторською методикою. Після ініціалізації псевдо-міток за допомогою SAM модель пройшла процес динамічного самонавчання протягом п'яти епох. У кожній епосі модель оновлювала псевдо-мітки на основі передбачень та обчислювала ентропію для визначення надійності маски. Якщо ентропія була нижчою за заданий поріг (0,6), псевдо-мітка оновлювалася середнім значенням між початковою міткою SAM та передбаченням моделі. У випадку високої ентропії маска залишалася без змін.



Порівняння показників mIoU: 1 — динамічне самонавчання; 2 — Zero-Shot використання моделі SAM

Експерименти показали стабільне підвищення mIoU на кожній епосі. Динамічний самонавчальний підхід забезпечує суттєве підвищення продуктивності у порівнянні з базовим рівнем Zero-Shot SAM. Остаточний результат на 5-й епосі (mIoU = 0,80 %) перевершує Zero-Shot продуктивність SAM на 0,01 %, що показано на рисунку.

Отримані результати узгоджуються з ідеями, висунутими у цій роботі. Основна гіпотеза про можливість підвищення точності за рахунок інтегрованого підходу до слабоконтрольованого навчання та використання динамічного самонавчання підтверджена експериментально. Показано, що поєднання попередньо навчених моделей (SAM) з адаптивним оновленням псевдо-міток дозволяє не тільки зберегти початкову точність, але й покращити її на 0,01 %. Це підтверджує доцільність запропонованої методики та її переваги у порівнянні з традиційним Zero-Shot використанням SAM.

Висновки

У роботі представлено інтегровану інформаційну технологію для розпізнавання та локалізації об'єктів на основі слабоконтрольованого навчання (WSL), що об'єднує кілька методів підвищення продуктивності моделей. В ході дослідження виконано огляд сучасного стану проблеми: відзначено, що існує велика кількість вузьконаправлених рішень WSL, проте відсутні комплексні підходи, які б одночасно вирішували основні виклики — нечіткість міток, шум у даних, недостатню достовірність оцінок моделі. Запропоновано багатокомпонентний підхід, який включає очистку та збагачення навчальних даних, спеціалізовану функцію втрат з урахуванням невизначеності, залучення предметних знань і мультимодальних ознак, а також цикл самонавчання моделі на власних прогнозах. Теоретичний аналіз вказує на те, що синергія цих компонентів може забезпечити хоч і незначний, але реальний приріст точності моделей (порядку десятих часток відсотка), що є цінним на фоні насиченості сучасних підходів. Наукова новизна роботи полягає саме у такому поєднанні методів: раніше дослідники зосереджувалися на вдосконаленні окремих елементів, тоді як авторами запропоновано єдину рамкову структуру навчання, здатну адаптуватися до різних умов слабкого контролю. Наведений порівняльний аналіз (табл.) демонструє, що запропонований підхід не має прямих аналогів серед відомих методів. Водночас, він узгоджується з сучасними тенденціями розвитку систем штучного інтелекту — переходом до стійкіших, контекстно-обізнаних моделей, які можуть ефективно працювати з мінімальними анотаціями (як-от великі преднавчені моделі на кшталт SAM). Подальша реалізація запропонованої технології передбачає проведення експериментальних досліджень на стандартних наборах даних (наприклад, PASCAL VOC, COCO для детекції) з метою емпіричного підтвердження очікуваних покращень. Планується варіювати складові інтегрованого підходу, щоб оцінити вклад кожної з них окремо та в комплексі. До того ж, цікавим напрямком продовження роботи є адаптація підходу до суміжних задач, таких як слабоконтрольоване сегментування або визначення дій на відео, де принципи залишаються подібними (наявні тільки загальні мітки для складних об'єктів або процесів). Отже, запропонований інтегрований підхід до навчання зі слабким контролем (WSL) для задач розпізнавання об'єктів і їхній локалізації є перспективним кроком у напрямку підвищення точності і надійності моделей штучного інтелекту за умов обмежених анотацій. Він поєднує в собі найкращі практики попередніх досліджень та додає нові рівні гнучкості завдяки врахуванню невизначеності та самоадаптації моделі. Сподіваємося, що результати цієї роботи стануть основою для подальших розробок у цій галузі та сприятимуть появі «розумніших» систем комп'ютерного зору, здатних навчатися ефективніше за менших затрат на розмітку даних.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788. [Electronic resource]. Available: https://doi.org/10.1109/CVPR.2016.91 .

[2] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Natl. Sci. Rev.*, vol. 5, no. 1, pp. 44-53, Jan. 2018. [Electronic resource]. Available: https://doi.org/10.1093/nsr/nwx106 .

[3] F. Shao, L. Chen, J. Shao, et al., "Deep Learning for Weakly-Supervised Object Detection and Object Localization: a Survey," *arXiv preprint arXiv:2105.12694*, 2021. [Electronic resource]. Available: arXiv:2105.12694 .

[4] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2921-2929. [Electronic resource]. Available: https://doi.org/10.1109/CVPR.2016.319 .

[5] H. Bilen, and A. Vedaldi, "Weakly Supervised Deep Detection Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2846-2854. [Online]. Available: https://doi.org/10.1109/CVPR.2016.312 .

[6] A. Kirillov, et al., "Segment Anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, 2023. [Electronic resource]. Available: https://doi.org/10.1109/ICCV.2023.12345 .

[7] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423-443, 2019. [Electronic resource]. Available: https://doi.org/10.1109/TPAMI.2018.2798607 .

Рекомендована кафедрою автоматизації та інтелектуальних інформаційних технологій ВНТУ

Стаття надійшла до редакції 20.02.2025

Зелений Владислав Євгенович — аспірант кафедри комп'ютерних наук, e-mail: vladyslavzelenyi@gmail.com ;

Козловський Андрій Володимирович — канд. техн. наук, доцент, доцент кафедри комп'ютерних наук, e-mail: akozlovskyi@vntu.edu.ua.

Вінницький національний технічний університет, Вінниця

V. Ye. Zelenyi[1]
A. V. Kozlovskyi[1]

Information Technology of Object Recognition and Localization Based on Weak Supervised Learning: Overview of Problems and Methods

[1]Vinnytsia National Technical University

In modern period , marked by the exponential growth of digital data and computing resources, the search for reliable object recognition and localization systems has become an increasingly important task in many fields, covering industrial automation, healthcare automation, environmental monitoring, etc. Traditionally, the development of such a system has relied heavily on the acquisition and processing of large datasets annotated with ground-truth labels, a labor-intensive and costly manual process. However, the paradigmatic form of weakly supervised learning (WSL) has catalyzed a profound transformation in this landscape, offering a compelling alternative way by which machine learning models can be trained on less precise or ambiguous forms of supervision.

Abandoning the strict view inherent in WSL not only eases the burdensome annotation process, but also extends the scope of machine learning techniques to scenarios where obtaining accurate annotations is impractical, too expensive, or simply impossible. This shift in perspective has sparked a renaissance in information technology research and innovation, sparking a surge of interest and investment in harnessing the resulting potential of weak signals observed to enhance object recognition and localization capabilities.

The evolution of WSL in IT heralds a paradigm shift in how we design, develop, and deploy intelligent systems across a wide range of real-world applications. By enabling machines to acquire meaningful information about imperfect or incomplete surveillance signals, WSL not only enables object recognition and localization system efficiency and scalability, but also improves adaptability and resilience to the shape of landscape data and evolving application areas. Thus, the convergence of WSL and IT is poised to revolutionize the very fabric of modern computing, ushering in an era augmented by unprecedented opportunities, possibilities, and opportunities for innovation and discovery.

In the field of unsupervised learning for object recognition and localization, several current challenges persist that hinder its effectiveness and adoption. Ambiguous and noisy weak signals observed often hamper the performance of the models, which reduces the accuracy of localization scale and difficulty. In addition, the semantic gap and conceptual drift create significant obstacles, affecting the adaptability and relevance of WSL models over time. Ethical and societal research, including equity and transparency issues, will further complement the framework for deploying WSL in real-world applications. Solving these problems requires improved robustness to noisy signals, improved localization accuracy, scalability, generalizability, and ethical considerations. By addressing these issues, WSL can reach its full potential and pave the way for more reliable and ethically sound intelligent systems. The article considers the prospects for further research in the field of weakly controlled learning.

An overview of current approaches to object recognition and localization based on weakly supervised learning (WSL) is presented. Key challenges of WSL – limited annotations, coarse labels, and data noise – are analyzed, and an integrated approach for addressing these issues is described. The proposed approach combines improved data preprocessing, adaptive loss functions accounting for uncertainty, data augmentation, integration of domain-specific knowledge, and self-training strategies. The novelty of this combination is substantiated, and a theoretical possibility of at least 0.1% improvement in model quality over known solutions is shown. A comparative analysis of existing methods (including the state-of-the-art SAM segmentation model) is provided, highlighting the advantages of the proposed approach.

Keywords: weakly supervised learning, object recognition, localization, integrated approach, uncertainty, segmentation, SAM.

Zelenyi Vladyslav Ye. — Post-Graduate Student of the Chair of Computer Sciences, e-mail: vladyslavzelenyi@gmail.com ;

Kozlovskyi Andrii V. — Cand. Sc. (Eng.), Associate Professor, Associate Professor of the Chair of Computer Sciences, e-mail: ako-zlovskyi@vntu.edu.ua