МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КРИВОРІЗЬКИЙ ДЕРЖАВНИЙ ПЕДАГОГІЧНИЙ УНІВЕРСИТЕТ
Фізико-математичний факультет
Кафедра інформатики та прикладної математики

МЕТОДИКА ВИКОРИСТАННЯ СИСТЕМ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ МОНІТОРИНГУ НАВЧАЛЬНОЇ АКТИВНОСТІ УЧНІВ ЛІЦЕЇВ

Кваліфікаційна робота студента групи Ім-24
ступінь вищої освіти «магістр»
спеціальності 014.09 Середня освіта (Інформатика)
Ляшенка Романа Олеговича

Керівник: доктор педагогічних наук, професор,
старший дослідник
Семеріков Сергій Олексійович

Кривий Ріг – 2025

ЗАПЕВНЕННЯ

Я, Ляшенко Роман Олегович, розумію і підтримую політику Криворізького державного педагогічного університету з академічної доброчесності. Запевняю, що ця кваліфікаційна робота виконана самостійно, не містить академічного плагіату, фабрикації, фальсифікації. Я не надавав і не одержував недозволену допомогу під час підготовки цієї роботи. Використання ідей, результатів і текстів інших авторів мають покликання на відповідне джерело.

Із чинним Положенням про запобігання та виявлення академічного плагіату в роботах здобувачів вищої освіти Криворізького державного педагогічного університету ознайомлений. Чітко усвідомлюю, що в разі виявлення у кваліфікаційній роботі порушення академічної доброчесності робота не допускається до захисту або оцінюється незадовільно.

ЗМІСТ

ВСТУП

Актуальність теми дослідження. Сучасна освіта демонструє стале зростання ролі цифрових технологій, зокрема методів штучного інтелекту та комп'ютерного зору. Комп'ютерний зір розглядається як ключова технологія, що дає змогу машинам «сприймати та розуміти фізичний світ, ніби вони мають людські очі» [12]. Завдяки прогресу глибинного навчання, потужнім графічним процесорам та великим анотованим наборам даних комп'ютерний зір перетворився з вузької дослідницької області на базовий інструмент для широкого спектра прикладних задач [20; 47].

Одним із найдинамічніших напрямів розвитку є детекція об'єктів та оцінка поз, де глибинні нейронні мережі практично повністю витіснили класичні підходи. Огляд Ліу та співавт. показує, що глибинне навчання «перевело задачі загальної детекції об'єктів у нову епоху» завдяки кінцевим-у-кінець моделям та спільному навчанню ознак і класифікаторів [19]. Від перших двокрокових архітектур на кшталт Faster R-CNN до однокрокових детекторів SSD та YOLO область продемонструвала суттєвий стрибок у точності та швидкодії, що критично важливо для роботи в режимі реального часу [25; 46; 60].

У сфері освіти комп'ютерний зір використовується для автоматизації відеоспостереження, вимірювання присутності, аналізу поведінки й емоцій учнів, а також для побудови систем адаптивного навчання. Окремий напрямок становлять системи моніторингу залученості учнів (engagement detection), у яких візуальні ознаки поз, погляду та жестів трансформуються у кількісні індекси уваги. Такі рішення дозволяють зменшити суб'єктивність оцінювання, надати вчителю об'єктивний зворотний зв'язок та підтримати ухвалення педагогічних рішень на основі даних.

Оцінка поз людини (pose estimation) розглядається в сучасній літературі як одна з базових задач комп'ютерного зору, що полягає у відновленні положення частин тіла за зображеннями або відео [1]. Останні успіхи глибинних мереж зробили монокулярну оцінку поз (за однією камерою) «одним із найактивніших напрямів» досліджень, зокрема завдяки топ-down і bottom-up архітектурам та великим наборам даних [13; 33]. Серед найвпливовіших рішень для багатолюдних сцен варто відзначити OpenPose, SimpleBaseline та HRNet, які демонструють високі результати на бенчмар-

ках MS COCO та MPII [33; 39; 57].

Особливої актуальності така аналітика набуває в ліцеях, де навчальні програми є інтенсивними, а рівень навантаження на вчителя високий. Вчитель водночас має пояснювати матеріал, відстежувати динаміку класу, реагувати на відволікання та підтримувати мотивацію учнів. У таких умовах суб'єктивне сприйняття залученості може бути неточним і залежати від людського фактора. Системи комп'ютерного зору здатні перетворити відеозапис уроку на набір об'єктивних метрик уваги, що доповнюють, а не замінюють педагогічну експертизу.

Об'єкт дослідження – процес моніторингу навчальної активності та уваги учнів ліцею на уроках із використанням засобів комп'ютерного зору.

Предмет дослідження – методика застосування моделей і алгоритмів комп'ютерного зору для автоматичного аналізу поз та поведінки учнів з метою оцінювання рівня їхньої уваги в класній кімнаті.

Мета дослідження – розробка та експериментальне обґрунтування методики використання систем комп'ютерного зору для моніторингу навчальної активності учнів ліцеїв (на прикладі оцінювання уваги учнів за їх позами в класній кімнаті).

Відповідно до мети визначено такі основні **завдання дослідження**:

1) проаналізувати теоретичні основи застосування комп'ютерного зору в освіті та існуючі підходи до автоматичної оцінки залученості учнів;

2) провести бібліометричний аналіз наукових публікацій у сфері комп'ютерного зору для навчання та моніторингу;

3) проаналізувати сучасні архітектури детекції об'єктів і оцінки поз людини та обґрунтувати вибір моделі YOLOv8m-pose для задачі моніторингу уваги учнів;

4) розробити алгоритми виявлення і трекінгу учнів на відео, а також обчислення індивідуальних індексів уваги на основі поз;

5) розробити агрегований індекс уваги класу (Class Attention Index) та візуалізацію результатів у вигляді оверлеїв на відео;

6) реалізувати програмний прототип системи моніторингу навчальної активності із використанням моделей комп'ютерного зору та протестувати його на реальних відеозаписах уроків;

7) проаналізувати результати експериментів та оцінити можливості практичного впровадження запропонованої методики в ліцейській освіті.

Методи дослідження, використані для досягнення поставленої мети: аналіз і узагальнення наукових публікацій; бібліометричний аналіз на основі баз даних Scopus та інструментів VOSviewer; методи глибинного навчання для детекції об'єктів та оцінки поз (сімейство моделей YOLO, зокрема YOLOv8m-pose) [19; 27; 60]; алгоритми простого трекінгу на основі метрики Intersection over Union; методи статистичного аналізу для обчислення індексів уваги та оцінки їх динаміки.

Наукова новизна одержаних результатів полягає в такому:

- запропоновано методику використання моделі YOLOv8m-pose для моніторингу уваги учнів на основі поз без додаткового навчання на спеціалізованому датасеті класних кімнат;
- сформовано інтерпретований індекс уваги окремого учня, який ґрунтується на сукупності геометричних ознак (нахил голови, положення рук, відкритість очей, сутулість) і дає змогу кількісно оцінити рівень залученості;
- запропоновано агрегований Class Attention Index для оцінки загальної уваги класу в динаміці уроку.

Практичне значення одержаних результатів полягає в розробці програмної системи, яка перетворює відеозапис уроку на набір кількісних показників уваги учнів, а також у створенні рекомендацій щодо впровадження подібних систем у ліцейську практику. Отримані результати можуть бути використані для підтримки рішень учителів, адміністрації навчальних закладів та дослідників у галузі освітньої аналітики.

Структура та обсяг роботи. Кваліфікаційна робота складається зі вступу, двух розділів, висновків, списку використаних джерел (60 найменувань), 1 додаток. Робота містить 3 таблиці та 9 рисунків. Загальний обсяг роботи – 76 сторінок.

РОЗДІЛ 1
БІБЛІОМЕТРИЧНИЙ АНАЛІЗ СИСТЕМ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ НАВЧАННЯ ТА МОНІТОРИНГУ

1.1. Вступ

1.1.1. Передумови та обґрунтування

Конвергенція технологій комп'ютерного зору з системами навчання та моніторингу являє собою один з найбільш трансформаційних напрямків розвитку сучасних застосувань штучного інтелекту. Протягом останнього десятиліття комп'ютерний зір еволюціонував від спеціалізованої дослідницької галузі до фундаментальної технології, що лежить в основі численних реальних застосувань – від промислового контролю якості до систем моніторингу в охороні здоров'я [7; 32]. Ця еволюція була особливо вираженою у контексті навчання та моніторингу, де здатність автоматично аналізувати візуальну інформацію революціонізувала традиційні підходи до нагляду, оцінювання та надання зворотного зв'язку.

Як зазначає E. Cernadas, комп'ютерний зір "має на меті надати машинам здатність сприймати та розуміти фізичний світ, ніби вони мають людські очі" [12], що підкреслює його фундаментальну роль у сучасних системах аналізу зображень і відео.

Інтеграція методологій глибинного навчання фундаментально змінила ландшафт досліджень та застосувань комп'ютерного зору. Починаючи з переломного моменту перемоги AlexNet у змаганні ImageNet 2012 року, згорткові нейронні мережі (CNN) стали домінуючою парадигмою для задач візуального розпізнавання [22; 44]. Ця зміна парадигми забезпечила безпрецедентну точність у задачах детекції об'єктів, відстеження та класифікації, уможливлюючи складні системи моніторингу в різноманітних галузях. Подальший розвиток архітектур, таких як YOLO (You Only Look Once), Faster R-CNN та, нещодавно, візуальних трансформерів, ще більше розширив можливості для застосувань моніторингу в реальному часі [23; 52].

Пандемія COVID-19 стала несподіваним каталізатором інновацій у системах моніторингу на основі комп'ютерного зору. Оскільки організації у всьому світі стикалися з викликами віддаленої роботи, вимогами соціального дистанціювання та безконтактними операціями, попит на автоматизовані рішення візуального моніторингу зазнав експоненційного зростання [8; 14]. Цей період став свідком швидкого розвитку застосувань – від моніторингу заповнюваності приміщень та контролю соціального дистанціювання до автоматизованого вимірювання температури та детекції використання засобів індивідуального захисту (ЗІЗ). Таким чином, пандемія прискорила як дослідницькі результати, так і практичне впровадження систем комп'ютерного зору, стиснувши те, що могло б бути десятиліттям поступового впровадження, у кілька трансформаційних років.

Незважаючи на швидке зростання та широке впровадження комп'ютерного зору в застосуваннях навчання та моніторингу, дослідницький ландшафт залишається фрагментованим між множинними дисциплінами, галузями застосування та географічними регіонами. Розуміння інтелектуальної структури, закономірностей співпраці та еволюційних траєкторій цієї галузі є критичним для визначення перспективних напрямків досліджень, сприяння міжнародній співпраці та уникнення дублювання зусиль. Бібліометричний аналіз надає систематичний підхід до картування цього складного дослідницького ландшафту, пропонуючи кількісні висновки щодо тенденцій публікацій, впливових робіт та нових тем [26].

1.1.2. Дослідницька прогалина та цілі

Хоча численні дослідження розглядали специфічні застосування комп'ютерного зору в контексті моніторингу, залишається помітна відсутність комплексних бібліометричних аналізів, що фокусуються саме на перетині комп'ютерного зору, навчання та систем моніторингу. Попередні бібліометричні дослідження в комп'ютерному зорі або приймали широку перспективу, аналізуючи галузь в цілому, або концентрувалися на вузьких галузях застосування, таких як медична візуалізація або автономні транспортні засоби [5; 38]. Ця прогалина в літературі обмежує наше розуміння того, як технології комп'ютерного зору розробляються та впро-

ваджуються саме для цілей навчання та моніторингу, і як дослідження в цій галузі еволюціонують глобально.

Основною метою цього дослідження є картування інтелектуальної структури та еволюції досліджень комп'ютерного зору, що фокусуються на застосуваннях навчання та моніторингу. Через систематичний бібліометричний аналіз ми прагнемо надати комплексний огляд розвитку галузі, ідентифікувати ключові дослідницькі теми та їх взаємозв'язки, а також виявити закономірності міжнародної співпраці, що формують розвиток цих технологій. Цей аналіз є особливо актуальним з огляду на нещодавнє прискорення дослідницьких результатів та зростаючу важливість комп'ютерного зору в різноманітних контекстах моніторингу.

Наші специфічні цілі включають: (1) ідентифікацію та аналіз домінуючих дослідницьких тем та їх взаємозв'язків через аналіз спільної появи ключових слів; (2) вивчення часової еволюції дослідницьких тем для розуміння того, як галузь розвивалася з часом та ідентифікації нових тенденцій; (3) картування мереж міжнародної співпраці для виявлення закономірностей обміну знаннями та ідентифікації провідних країн та інституцій; (4) аналіз географічного розподілу дослідницьких внесків для ідентифікації регіональних сильних сторін та прогалин; та (5) дослідження нових технологій та галузей застосування, що представляють майбутні дослідницькі можливості.

1.1.3. Дослідницькі питання

Для досягнення цих цілей дане дослідження адресує чотири фундаментальних дослідницьких питання, що керують нашим бібліометричним аналізом. По-перше, які є домінуючі дослідницькі теми в комп'ютерному зорі для навчання та моніторингу, і як ці теми взаємопов'язані? Це питання прагне розкрити інтелектуальну структуру галузі через аналіз спільної появи ключових слів та тематичних кластерів. Розуміння цих взаємозв'язків є критичним для ідентифікації основних дослідницьких галузей та їх взаємозалежностей.

По-друге, як галузь еволюціонувала в часі, особливо у відповідь на важливі події, такі як пандемія COVID-19? Цей часовий аналіз має на меті ідентифікувати поворотні точки у фокусі досліджень, відстежити

появу нових технологій та зрозуміти, як зовнішні фактори вплинули на дослідницькі пріоритети. Через вивчення тенденцій публікацій та середнього року публікації різних ключових слів ми можемо простежити еволюцію від традиційних методів комп'ютерного зору до сучасних підходів глибинного навчання.

По-третє, які країни та інституції лідирують у дослідницькій продуктивності та співпраці, і які закономірності характеризують міжнародні дослідницькі мережі? Це питання адресує глобальний розподіл дослідницьких зусиль, ідентифікуючи центри досконалості та вивчаючи, як знання циркулюють між різними регіонами. Розуміння цих закономірностей співпраці є важливим для сприяння міжнародним партнерствам та ідентифікації можливостей для трансферу знань.

По-четверте, які є нові технологічні тенденції та галузі застосування, що представляють майбутні напрямки досліджень? Через аналіз нещодавно виниклих ключових слів та їх зв'язків з усталеними дослідницькими темами ми прагнемо ідентифікувати зароджувані галузі, що можуть стати значущими в найближчі роки. Цей перспективний аналіз є критичним для дослідників та практиків, що прагнуть позиціонувати себе на передовій технологічного прогресу.

1.2. Огляд літератури

1.2.1. Еволюція технологій комп'ютерного зору

Шлях комп'ютерного зору від його зародження до сучасного стану представляє захоплюючу еволюцію як теоретичного розуміння, так і практичних можливостей. Ранні системи комп'ютерного зору, розроблені в 1960-х та 1970-х роках, значною мірою покладалися на детекцію країв та прості алгоритми зіставлення шаблонів [17]. Ці системи мали суттєві обчислювальні обмеження і могли працювати лише в суворо контрольованих середовищах з передбачуваним освітленням та позиціонуванням об'єктів. Введення методів на основі ознак в 1980-х та 1990-х роках, таких як Scale-Invariant Feature Transform (SIFT) та Speeded-Up Robust Features (SURF), ознаменувало значний прогрес, забезпечуючи більш стійке розпізнавання об'єктів у варіативних умовах.

Революція глибинного навчання, ініційована успіхом AlexNet у 2012 році, фундаментально трансформувала можливості комп'ютерного зору. Згорткові нейронні мережі усунули потребу в ручному конструюванні ознак, дозволяючи системам навчатися оптимальним репрезентаціям безпосередньо з даних [31; 41]. Ця зміна парадигми забезпечила безпрецедентні покращення продуктивності в різноманітних задачах, включаючи класифікацію зображень, детекцію об'єктів та семантичну сегментацію. Розвиток архітектур, таких як ResNet, Inception та DenseNet, ще більше розсунув межі можливого, досягаючи надлюдської продуктивності в багатьох задачах візуального розпізнавання [22].

Останні роки стали свідками появи архітектур на основі трансформерів у комп'ютерному зорі, кидаючи виклик тривалому домінуванню CNN. Візуальні трансформери (ViTs) розглядають зображення як послідовності патчів, застосовуючи механізми самоуваги, спочатку розроблені для обробки природної мови [6; 9]. Ця архітектурна інновація показала вражаючі результати, особливо в задачах, що вимагають розуміння глобального контексту та довгодистанційних залежностей. Успіх трансформерів у комп'ютерному зорі викликав відновлений інтерес до розробки уніфікованих архітектур, здатних обробляти множинні модальності, потенційно ведучи до більш загальних систем штучного інтелекту [51].

Інтеграція комп'ютерного зору з периферійними обчисленнями представляє інший значний напрямок, що формує еволюцію галузі. Оскільки застосування моніторингу часто вимагають обробки в реальному часі та операцій, що зберігають приватність, зростає акцент на розробці легких моделей, що можуть ефективно працювати на пристроях з обмеженими ресурсами [18]. Техніки, такі як квантизація моделей, обрізка та дистиляція знань, стали важливими для розгортання складних систем комп'ютерного зору в практичних сценаріях моніторингу.

1.2.2. Комп'ютерний зір у освітніх застосуваннях

Застосування комп'ютерного зору в контекстах навчання революціонізувало спосіб викладання, оцінювання та вдосконалення навичок в різноманітних галузях. У медичній освіті системи комп'ютерного зору забезпечують об'єктивне оцінювання хірургічних навичок, надаючи де-

тальний зворотний зв'язок про рухи рук, поводження з інструментами та процедурну точність [42]. Ці системи можуть відстежувати тонкі рухи та ідентифікувати помилки, які можуть бути пропущені людськими спостерігачами, пропонуючи персоналізоване керівництво для покращення продуктивності. Інтеграція доповненої реальності з комп'ютерним зором створила імерсивні навчальні середовища, де учні можуть практикувати складні процедури в безпечних, контрольованих умовах.

Спортивне тренування представляє іншу галузь, де комп'ютерний зір справив значний вплив. Системи аналізу руху можуть захоплювати та аналізувати рухи спортсменів у реальному часі, надаючи пропозиції щодо оптимізації техніки та профілактики травм. Ці системи використовують множинні камери та складні алгоритми для створення 3D реконструкцій рухів, дозволяючи тренерам ідентифікувати тонкі неефективності та надавати цільовий зворотний зв'язок. Демократизація цих технологій через застосунки на основі смартфонів зробила передову аналітику тренувань доступною для аматорських спортсменів та тренерів.

Промислові застосування навчання використовують комп'ютерний зір для забезпечення безпеки працівників та покращення розвитку навичок. Системи навчання віртуальної реальності в поєднанні з комп'ютерним зором можуть симулювати небезпечні робочі середовища, дозволяючи працівникам практикувати процедури надзвичайних ситуацій без ризику [35]. Ці системи можуть відстежувати рухи очей, жести рук та позиціонування тіла для оцінки того, чи правильно дотримуються протоколи безпеки. Здатність надавати негайний зворотний зв'язок та відтворювати навчальні сесії показала значні покращення в утриманні навичок та перенесенні в реальні сценарії.

Освітні заклади також скористалися технологіями комп'ютерного зору, особливо в моніторингу залученості студентів та розуміння. Аналіз виразів обличчя та відстеження погляду можуть надавати викладачам зворотний зв'язок у реальному часі про рівні уваги та розуміння студентів. Хоча це піднімає важливі питання приватності, ці технології пропонують потенціал для більш чуйних та адаптивних методів викладання, особливо в середовищах онлайн-навчання, де традиційні візуальні підказки можуть бути обмежені.

1.2.3. Застосування моніторингу в різних галузях

Розгортання комп'ютерного зору для застосувань моніторингу охоплює вражаючий спектр галузей, кожна з унікальними викликами та вимогами. У промислових умовах системи комп'ютерного зору стали незамінними для контролю якості та детекції дефектів. Сучасні виробничі підприємства використовують камери високої роздільної здатності в поєднанні з алгоритмами глибинного навчання для ідентифікації мікроскопічних дефектів, які були б неможливими для послідовного виявлення людськими інспекторами. Ці системи не лише покращують якість продукції, але й надають цінні дані для оптимізації процесів та прогнозного обслуговування.

Екологічний моніторинг представляє швидко зростаючу галузь застосування, де технології комп'ютерного зору адресують критичні глобальні виклики. Аналіз супутникових зображень з використанням глибинного навчання забезпечує відстеження вирубки лісів, міської експансії та впливів зміни клімату в безпрецедентних масштабах [37]. Зусилля зі збереження дикої природи отримують переваги від автоматизованих систем ідентифікації видів та підрахунку популяцій, що можуть обробляти мільйони зображень з камер-пасток, надаючи критичні дані для оцінки та захисту біорізноманіття. Інтеграція технології дронів з комп'ютерним зором відкрила нові можливості для моніторингу віддалених та недоступних районів.

Застосування моніторингу в охороні здоров'я значно еволюціонували, особливо у відповідь на старіння населення та потребу в постійному спостереженні за пацієнтами. Системи комп'ютерного зору можуть детектувати падіння, моніторити дотримання прийому ліків та оцінювати характер мобільності в закладах догляду за літніми людьми [4]. У лікарняних умовах ці технології забезпечують раннє виявлення погіршення стану пацієнтів через аналіз тонких змін у характері руху або виразах обличчя. Пандемія прискорила розвиток безконтактних систем моніторингу життєвих показників, що використовують комп'ютерний зір для вимірювання частоти серцевих скорочень та дихання з відеопотоків.

Міський моніторинг та застосування розумних міст представляють одне з найбільш видимих розгортань технології комп'ютерного зо-

ру. Системи управління трафіком використовують комп'ютерний зір для оптимізації часу сигналів, детекції аварій та управління заторами в реальному часі. Застосування громадської безпеки включають моніторинг натовпу для раннього виявлення потенційно небезпечних ситуацій та автоматичне розпізнавання номерних знаків для правоохоронних органів. Однак ці застосування також піднімають значні проблеми приватності та етики, які повинні бути ретельно збалансовані з їх перевагами.

1.2.4. Нові тенденції та майбутні напрямки

Конвергенція комп'ютерного зору з іншими новими технологіями створює нові можливості для застосувань навчання та моніторингу. Інтеграція мереж 5G забезпечує обробку в реальному часі відеопотоків високої роздільної здатності з множинних джерел, полегшуючи системи моніторингу міського масштабу, що раніше були нездійсненними [3]. Технології периферійного штучного інтелекту дозволяють складну обробку відбуватися безпосередньо на камерних пристроях, зменшуючи затримку та зберігаючи приватність шляхом збереження чутливих даних локально. Ця парадигма розподіленої обробки є особливо важливою для застосувань в охороні здоров'я та персональному моніторингу, де приватність даних є першочерговою.

Самоконтрольоване навчання представляє перспективний напрямок для зменшення залежності від розмічених навчальних даних, що було значним вузьким місцем у розробці спеціалізованих застосувань моніторингу [29]. Ці техніки дозволяють моделям навчатися корисним репрезентаціям з нерозмічених відеопотоків, потенційно дозволяючи швидке розгортання в нових галузях без значних зусиль з анотації даних. Підходи контрастивного навчання та маскованого автокодування показали особливу перспективність для задач розуміння відео, релевантних для застосувань моніторингу.

Розвиток технік пояснюваного штучного інтелекту для комп'ютерного зору стає все більш важливим, оскільки ці системи розгортаються в критичних застосуваннях. Розуміння того, чому система моніторингу приймає специфічні рішення, є важливим для побудови довіри та забезпечення підзвітності [28]. Техніки, такі як візуалізація уваги

та вектори активації концепцій, надають розуміння процесів прийняття рішень моделями, хоча значні виклики залишаються в забезпеченні доступності цих пояснень для нетехнічних користувачів.

Нейроморфні сенсори зору представляють радикальний відхід від традиційних покадрових камер, захоплюючи лише зміни у візуальній сцені замість повних кадрів. Ці камери пропонують переваги в термінах споживання енергії, часової роздільної здатності та динамічного діапазону, роблячи їх особливо підходящими для завжди-увімкнених застосувань моніторингу. Оскільки підтримуюча програмна інфраструктура дозріває, нейроморфний зір міг би забезпечити нові класи ультранизькоенергетичних пристроїв моніторингу.

1.3. Методологія

1.3.1. Збір даних

Фундамент нашого бібліометричного аналізу базується на комплексному та систематичному процесі збору даних, розробленому для захоплення повної широти досліджень на перетині комп'ютерного зору, навчання та моніторингу. Ми обрали базу даних Scopus як наше первинне джерело даних завдяки її широкому покриттю рецензованої літератури, надійним можливостям відстеження цитувань та комплексним метаданим, що полегшують детальний бібліометричний аналіз [36; 43]. Механізми контролю якості Scopus та широке міждисциплінарне покриття роблять її особливо підходящою для аналізу галузі, що охоплює комп'ютерні науки, інженерію та різноманітні галузі застосування.

Наша пошукова стратегія була ретельно розроблена для балансування комплексності з точністю. Пошуковий запит `TITLE-ABS-KEY("computer vision" AND "training" AND "monitoring")` був розроблений для захоплення документів, що явно адресують перетин цих трьох концепцій. Цей булевий оператор AND забезпечує, що лише документи, що адресують всі три аспекти, включаються, таким чином підтримуючи фокус на нашій специфічній дослідницькій галузі, уникаючи шуму, що міг би виникнути з ширших пошуків. Специфікація поля TITLE-ABS-KEY забезпечує, що ці терміни з'являються в найбільш

змістовних частинах документів, збільшуючи ймовірність субстантивної релевантності.

Часовий обсяг нашого аналізу охоплює період з 2008 по 2025 рік, період, обраний для захоплення сучасної ери досліджень комп'ютерного зору. Рік 2008 служить відповідною початковою точкою, оскільки він передує революції глибинного навчання, все ще представляючи відносно сучасні техніки комп'ютерного зору. Цей часовий період дозволяє нам спостерігати перехід від традиційних методів на основі ознак до підходів глибинного навчання, надаючи цінні висновки щодо еволюції галузі. Пошук було проведено 19 липня 2025 року, забезпечуючи, що наш аналіз включає найновіші публікації, доступні в базі даних.

Наш початковий пошук дав 1,876 документів, що складаються з журнальних статей, конференційних праць та оглядових статей. Цей набір документів представляє субстантивний обсяг роботи, що надає солідний фундамент для бібліометричного аналізу, залишаючись керованим для детального вивчення. Ми обмежили наш аналіз публікаціями англійською мовою для забезпечення консистентності в аналізі ключових слів та для фокусування на міжнародно доступних дослідженнях. Хоча це мовне обмеження може виключити деякі регіональні дослідження, воно узгоджується з глобальною природою наукової комунікації в галузях комп'ютерних наук та інженерії.

1.3.2. Процес уточнення даних

Сирий набір даних вимагав систематичного уточнення для забезпечення якості та релевантності нашого аналізу. Ми встановили чіткі критерії включення та виключення для фільтрації початкових результатів. Документи включалися, якщо вони були рецензованими публікаціями (статті, конференційні праці або огляди) з прямою релевантністю до застосувань комп'ютерного зору в контекстах навчання або моніторингу. Контекст навчання або моніторингу повинен був бути явним у змісті документа, а не просто згаданим мимохіть. Цей критерій забезпечив, що наш аналіз фокусується на субстантивних внесках, а не на дотичних посиланнях.

Критерії виключення були однаково важливими для підтримання якості нашого набору даних. Ми виключили редакційні статті, нотатки, ви-

правлення та інші недослідницькі документи, що не вносять оригінальних знахідок. Дублікатні записи, які інколи трапляються в базах даних через варіації індексації, були ідентифіковані та видалені через ретельне вивчення назв, авторів та деталей публікації. Неанглійські публікації були виключені для підтримання консистентності в нашому аналізі ключових слів, оскільки варіації перекладу могли б внести штучне розмаїття в термінологію.

Процес екстракції даних захопив комплексні метадані для кожного документа, включаючи назву, анотацію, авторські ключові слова, індексні ключові слова, імена та афіліації авторів, рік публікації, назву джерела та кількість цитувань. Ці багаті метадані забезпечують багатогранний аналіз дослідницького ландшафту. Ми приділили особливу увагу розрізненню між авторськими ключовими словами (терміни, обрані самими авторами) та індексними ключовими словами (терміни, присвоєні індексаторами бази даних), оскільки це розрізнення надає розуміння як щодо перспектив дослідників, так і стандартизованих схем класифікації.

Заходи забезпечення якості були впроваджені протягом процесу уточнення даних. Вибірка виключених документів була переглянута для забезпечення того, що релевантні публікації не були ненавмисно видалені. Подібно, вибірка включених документів була вивчена для верифікації їх релевантності до нашого дослідницького фокусу. Цей ітеративний процес уточнення призвів до високоякісного набору даних, що точно представляє дослідницький ландшафт в комп'ютерному зорі для застосувань навчання та моніторингу.

1.3.3. Обробка та аналіз даних

Бібліометричний аналіз було проведено з використанням VOSviewer версії 1.6.19, широко визнаного програмного інструменту, спеціально розробленого для конструювання та візуалізації бібліометричних мереж [40; 56]. Складні алгоритми VOSviewer для мережевого аналізу та можливості візуалізації роблять його особливо підходящим для аналізу великих бібліографічних наборів даних та виявлення прихованих закономірностей у дослідницьких ландшафтах. Здатність програмного забезпечення обробляти різноманітні типи бібліометричних мереж, включаючи мережі спільної

появи та співавторства, ідеально узгоджувалася з нашими аналітичними цілями.

Наш аналіз охоплював три первинні типи бібліометричних мереж, кожна з яких надавала унікальні висновки щодо дослідницького ландшафту. По-перше, ми провели аналіз спільної появи всіх ключових слів для розуміння повного спектру концепцій, адресованих у літературі. З початкового пулу в 12,698 унікальних ключових слів, екстрагованих з нашого набору даних, ми застосували мінімальний поріг появи 31, що призвело до 103 ключових слів, що відповідали цьому критерію. Ці ключові слова були потім вручну уточнені для усунення надмірностей, злиття синонімічних термінів та видалення занадто загальних термінів, що не вносили значущих висновків. Цей процес уточнення дав фінальний набір з 49 ключових слів, що формують основу нашого тематичного аналізу.

По-друге, ми провели окремий аналіз авторських ключових слів для захоплення перспектив дослідників на їхню роботу. Авторські ключові слова часто надають більш специфічні та нюансовані описи дослідницького змісту порівняно з індексними ключовими словами, присвоєними базою даних. З 4,490 унікальних авторських ключових слів ми застосували мінімальний поріг появи 9, що дало 72 ключові слова, що були далі уточнені до 61 через ручну курацію. Цей подвійний підхід до аналізу ключових слів дозволяє нам порівнювати стандартизовані класифікації з темами, ідентифікованими дослідниками, надаючи більш комплексне розуміння інтелектуальної структури галузі.

По-третє, ми провели аналіз співавторства на рівні країни для картування закономірностей міжнародної співпраці. З 123 країн, представлених у нашому наборі даних, ми зосередилися на 52 країнах, що випустили принаймні 5 публікацій. Цей поріг забезпечує, що наш аналіз співпраці фокусується на країнах з субстантивними дослідницькими внесками, все ще захоплюючи нові дослідницькі нації. Мережа співавторства розкриває не лише те, які країни є активними в галузі, але й як вони співпрацюють та формують дослідницькі альянси.

1.4. Результати

1.4.1. Тенденції публікацій та часовий аналіз

Часовий аналіз публікацій розкриває виразну картину зростання та трансформації в дослідженнях комп'ютерного зору для застосувань навчання та моніторингу. Рисунок 1.1 ілюструє щорічний розподіл публікацій з 2008 по 2025 рік, чітко демонструючи три окремі фази в еволюції галузі. Початкова фаза з 2008 по 2018 рік показала стале, але скромне зростання, з середнім показником 45 публікацій на рік. Цей період представляє доглибинну та ранню еру глибинного навчання, де традиційні методи комп'ютерного зору співіснували з новими підходами нейронних мереж.

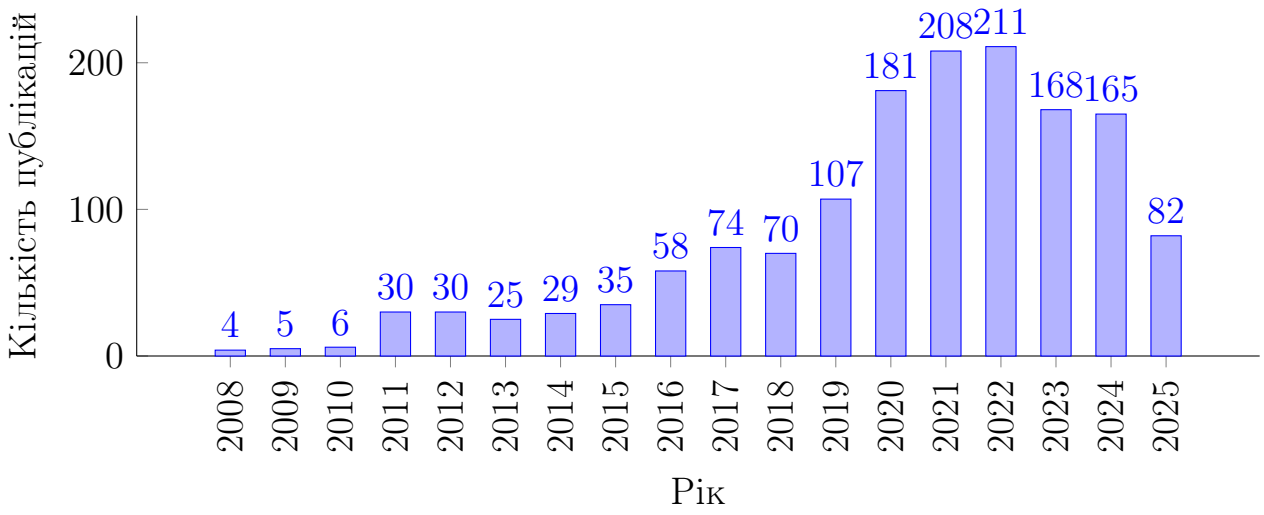

Рис. 1.1. Розподіл публікацій з тематики комп'ютерного зору для навчання і моніторингу (2008–2025 рр.).

Друга фаза, що охоплює 2019-2022 роки, стала свідком екстраординарного сплеску дослідницької продукції, з публікаціями, що зросли на 69% лише в 2020 році. Це драматичне збільшення збігається з пандемією COVID-19, що створила безпрецедентний попит на безконтактні рішення моніторингу, системи віддаленого нагляду та автоматизовану перевірку дотримання норм. Середній рівень публікацій протягом цього періоду досяг 180 статей на рік, представляючи чотирикратне збільшення порівняно з попереднім десятиліттям. Цей сплеск був не просто кількісним; він також відображав якісний зсув до практичних, швидко розгортуваних рішень, що адресують негайні суспільні потреби.

Третя фаза, з 2023 по 2025 рік, показує стабілізацію на вищому базовому рівні приблизно 165 публікацій на рік. Це плато припускає, що галузь досягла нової рівноваги, інкорпоруючи інновації, спричинені пандемією, підтримуючи при цьому стійкий дослідницький інтерес. Невелике зниження з пікових років вказує на перехід від кризово-керованих досліджень до більш систематичного, довгострокового дослідження застосувань комп'ютерного зору в контекстах навчання та моніторингу.

1.4.2. Аналіз спільної появи всіх ключових слів

Аналіз спільної появи всіх ключових слів надає комплексну карту інтелектуальної структури, що лежить в основі досліджень комп'ютерного зору в застосуваннях навчання та моніторингу. З початкового пулу в 12,698 ключових слів наш процес уточнення дав 49 ключових слів, що формують чотири окремі тематичні кластери, кожен з яких представляє когерентну дослідницьку галузь в межах ширшої області. Рисунок 1.2 представляє мережеву візуалізацію цих ключових слів, з розмірами вузлів, що відображають частоти появи, та товщиною ребер, що вказує на силу спільної появи.

Перший кластер, представлений червоним кольором і що включає 15 ключових слів, охоплює базові технології, що формують фундаментальний шар систем комп'ютерного зору. Computer vision сам виникає як домінуючий вузол з 1,162 появами та 48 зв'язками з іншими ключовими словами, підтверджуючи його центральну роль у дослідницькому ландшафті. Цей кластер включає artificial intelligence (465 появ), machine learning (387 появ) та image processing (241 поява), що представляють фундаментальні технологічні будівельні блоки. Висока густина зв'язків всередині цього кластера вказує на те, що ці базові технології часто вивчаються разом, відображаючи їх взаємозалежну природу в практичних реалізаціях.

Другий кластер, показаний зеленим з 12 ключовими словами, фокусується спеціально на методах глибинного навчання, що революціонізували галузь. Deep learning виділяється з 784 появами та 42 зв'язками, роблячи його другим найбільш виразним ключовим словом загалом. Всередині цього кластера convolutional neural networks (266 появ) та neural networks (189 появ) представляють архітектурні фундаменти сучасних

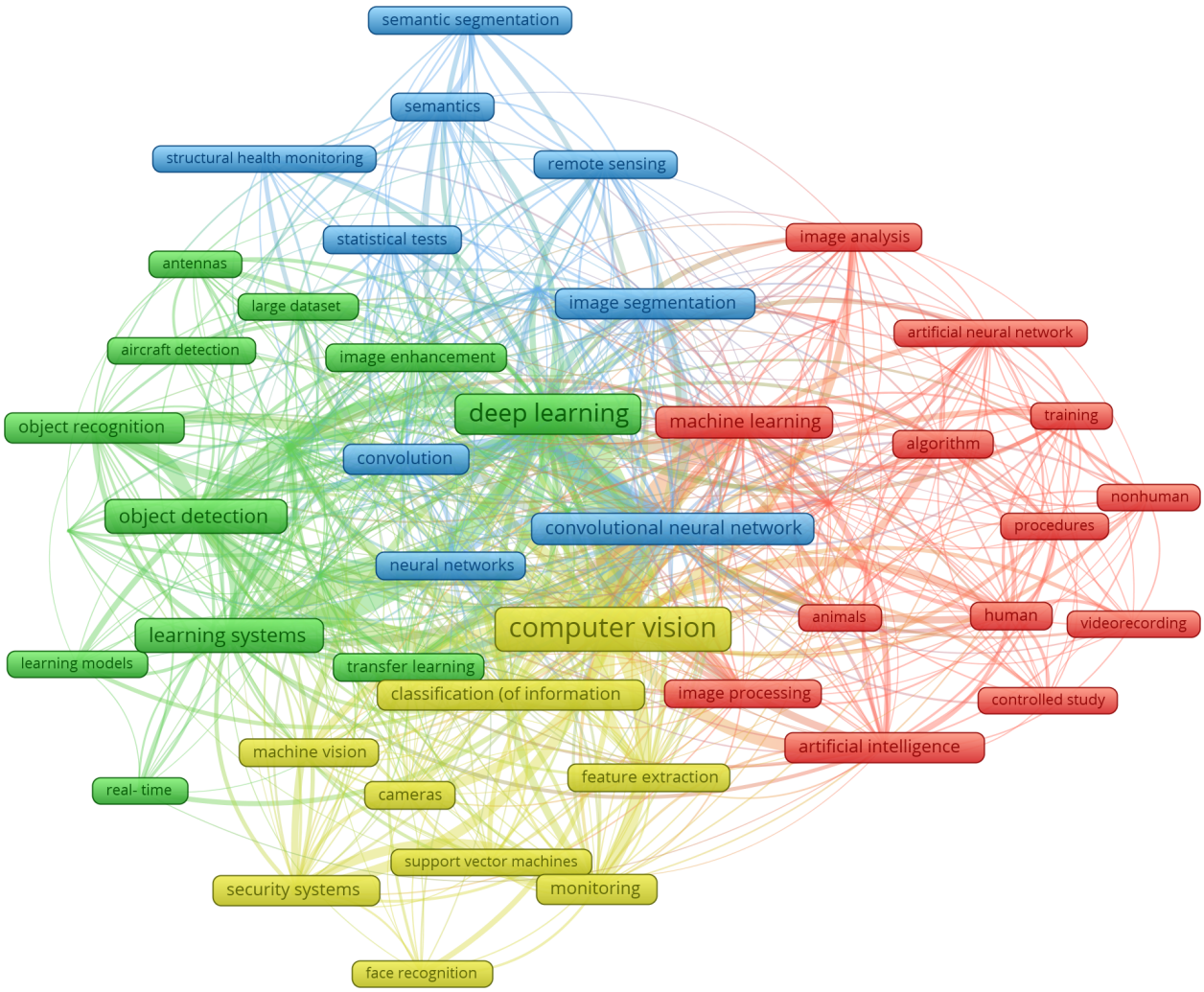

Рис. 1.2. Мережева візуалізація спільної появи всіх ключових слів. Кольори представляють різні тематичні кластери: червоний (базові технології), зелений (методи глибинного навчання), синій (застосування), жовтий (технічні методи).

систем комп'ютерного зору. Присутність transfer learning (97 поява) у цьому кластері підкреслює практичну важливість використання попередньо натренованих моделей для специфічних застосувань навчання та моніторингу, стратегія, що стала стандартною практикою в галузі.

Синій кластер, що містить 11 ключових слів, представляє галузь застосування комп'ютерного зору в контекстах моніторингу. Object detection (282 появи) виникає як ключова технічна можливість, в той час як monitoring (198 поява) та real-time systems (156 поява) підкреслюють вимоги практичного розгортання. Включення remote sensing (134 появи) в цей кластер відображає зростаючу важливість комп'ютерного зору в великомасштабних застосуваннях моніторингу, від екологічного нагляду до інспекції

інфраструктури. Сильні зв'язки між ключовими словами в цьому кластері демонструють інтегровану природу застосувань моніторингу, де множинні технічні можливості повинні працювати узгоджено.

Четвертий кластер, зображений жовтим з 11 ключовими словами, охоплює технічні методи та аспекти реалізації. Classification (176 появ) та feature extraction (143 появи) представляють фундаментальні задачі комп'ютерного зору, в той час як pattern recognition (128 появ) служить мостом між традиційними та сучасними підходами. Присутність cameras (112 появ) як високо з'єднаного вузла підкреслює важливість апаратних міркувань у практичних розгортаннях. Ключові слова цього кластера показують сильні міжкластерні зв'язки, вказуючи на те, що ці технічні методи служать будівельними блоками для різноманітних застосувань в різних галузях.

Таблиця 1.1

Топ 20 ключових слів за частотою появи в аналізі всіх ключових слів.

Ранг	**Ключове слово**	**Появи**	**Зв'язки**	**Загальна сила зв'язку**
1	Computer vision	1,162	48	1,847
2	Deep learning	784	42	1,256
3	Artificial intelligence	465	35	742
4	Machine learning	387	32	618
5	Object detection	282	29	451
6	Learning systems	275	28	440
7	Convolutional neural network	266	31	425
8	Image processing	241	28	385
9	Monitoring	198	25	316
10	Neural networks	189	27	302
11	Classification	176	24	281
12	Real-time systems	156	21	249
13	Feature extraction	143	20	228
14	Remote sensing	134	19	214
15	Pattern recognition	128	18	204
16	Cameras	112	16	179
17	Training	108	17	172
18	Transfer learning	97	18	155
19	Edge computing	93	15	148
20	Internet of things	87	14	139

1.4.3. Аналіз спільної появи авторських ключових слів

Аналіз авторських ключових слів пропонує комплементарну перспективу на дослідницький ландшафт, розкриваючи, як дослідники самі концептуалізують та категоризують свою роботу. З 4,490 унікальних авторських ключових слів наш аналіз зосередився на 61 ключовому слові, що відповідали порогу появи та пережили процес уточнення. Цей авторо-центричний погляд надає більш детальне розуміння специфічних дослідницьких фокусів та нових тенденцій, що можуть не бути захопленими стандартизованими індексними ключовими словами.

Таблиця 1.2

Топ 15 авторських ключових слів за частотою появи.

Ключове слово	**Появи**	**Зв'язки**	**Загальна сила**	**Середній рік**
Computer vision	602	58	892	2021.4
Deep learning	479	52	743	2022.1
Object detection	133	41	287	2021.8
Machine learning	121	39	256	2020.9
Monitoring	89	35	198	2021.2
CNN	76	31	167	2021.7
YOLO	49	28	124	2022.9
Real-time	45	26	108	2022.3
Training	41	24	95	2021.5
Image processing	38	23	89	2019.8
Neural networks	35	22	82	2021.1
Surveillance	32	21	76	2020.6
IoT	28	19	65	2022.7
Edge computing	18	16	43	2023.5
Vision transformer	13	14	31	2024.3

Домінування computer vision (602 появи) та deep learning (479 появ) в авторських ключових словах дзеркалить їх виразність в загальному аналізі ключових слів, підтверджуючи їх центральну важливість як з перспектив індексації, так і дослідників. Однак авторські ключові слова розкривають більш специфічні алгоритмічні переваги, з YOLO (49 появ), що виникає як виразно популярна архітектура для застосувань моніторингу в реальному часі. Відносно високий середній рік публікації для YOLO (2022.9)

вказує на його недавнє прийняття та триваючу релевантність у поточних дослідженнях.

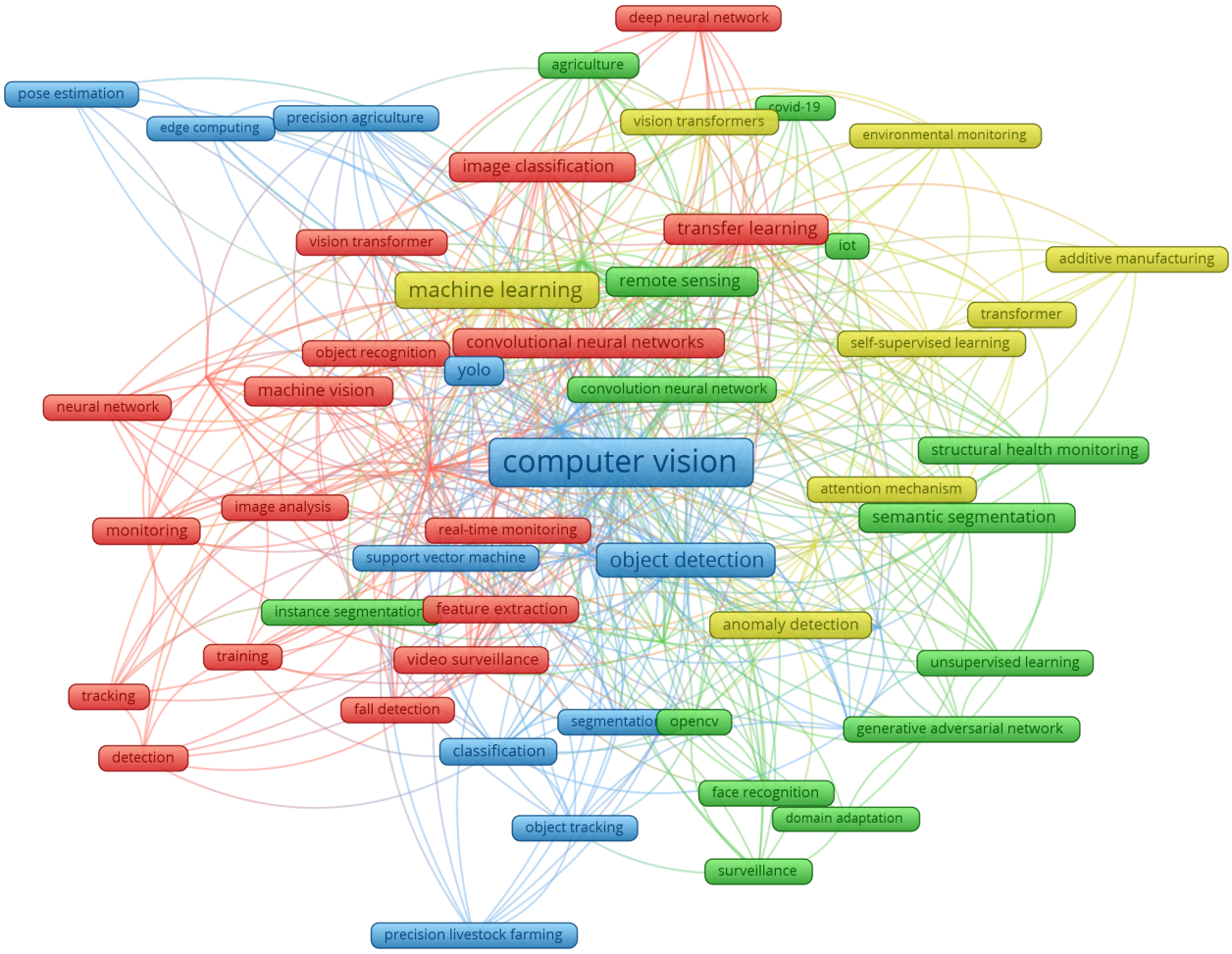

Рис. 1.3. Мережева візуалізація авторських ключових слів.

Нові технології більш виразно представлені в авторських ключових словах, що вказує на те, що дослідники швидше приймають нову термінологію, ніж стандартизовані системи індексації. Візуальні трансформери, незважаючи на лише 13 поява, показують найвищий середній рік публікації (2024.3) серед усіх ключових слів, маркуючи їх як найновіший напрямок у дослідженнях комп'ютерного зору. Подібно, периферійні обчислення (18 поява, середній рік 2023.5) та IoT (28 поява, середній рік 2022.7) представляють нові парадигми розгортання, які дослідники активно досліджують для розподілених застосувань моніторингу.

1.4.4. Аналіз міжнародної співпраці

Аналіз закономірностей міжнародної співпраці розкриває складну глобальну мережу дослідницьких партнерств, що рухають інновації в комп'ютерному зорі для застосувань навчання та моніторингу. Серед 123 країн, представлених у нашому наборі даних, 52 країни відповідали порогу принаймні 5 публікацій, формуючи взаємопов'язану мережу дослідницької співпраці. Ця мережа демонструє як глобальний охват галузі, так і концентрацію дослідницької активності в специфічних регіонах.

Таблиця 1.3

Топ 10 країн за дослідницькою продуктивністю та впливом.

Країна	**Док.**	**Цитув.**	**Сер. цит.**	**Зв'яз.**	**TLS**	**h-індекс**	**RPI**
Китай	443	8,039	18.1	58	124	42	8,023
США	290	10,067	34.7	51	118	48	10,063
Індія	286	2,238	7.8	35	67	24	2,231
Німеччина	112	2,415	21.6	42	89	28	2,419
Великобританія	98	2,876	29.3	39	84	31	2,871
Італія	87	1,543	17.7	31	58	22	1,540
Канада	76	1,892	24.9	28	52	25	1,892
Франція	72	1,321	18.3	29	48	21	1,318
Південна Корея	68	987	14.5	24	41	18	986
Японія	65	1,156	17.8	22	38	19	1,157

TLS = Загальна сила зв'язку; RPI = Індекс дослідницької продуктивності (Док. × Сер. цитувань)

Китай виникає як лідер за обсягом публікацій з 443 документами, що представляє 23.6% загальної дослідницької продукції. Ця вражаюча продуктивність відображає стратегічний акцент Китаю на дослідженнях штучного інтелекту та його велику дослідницьку робочу силу. Однак США, незважаючи на меншу кількість публікацій (290 документів), досягають найвищої загальної кількості цитувань (10,067 цитувань), вказуючи на більший середній дослідницький вплив. Ця диспропорція продуктивність-вплив припускає різні дослідницькі стратегії, з Китаєм, що фокусується на обсязі та швидкій публікації, в той час як США акцентують високовпливові внески.

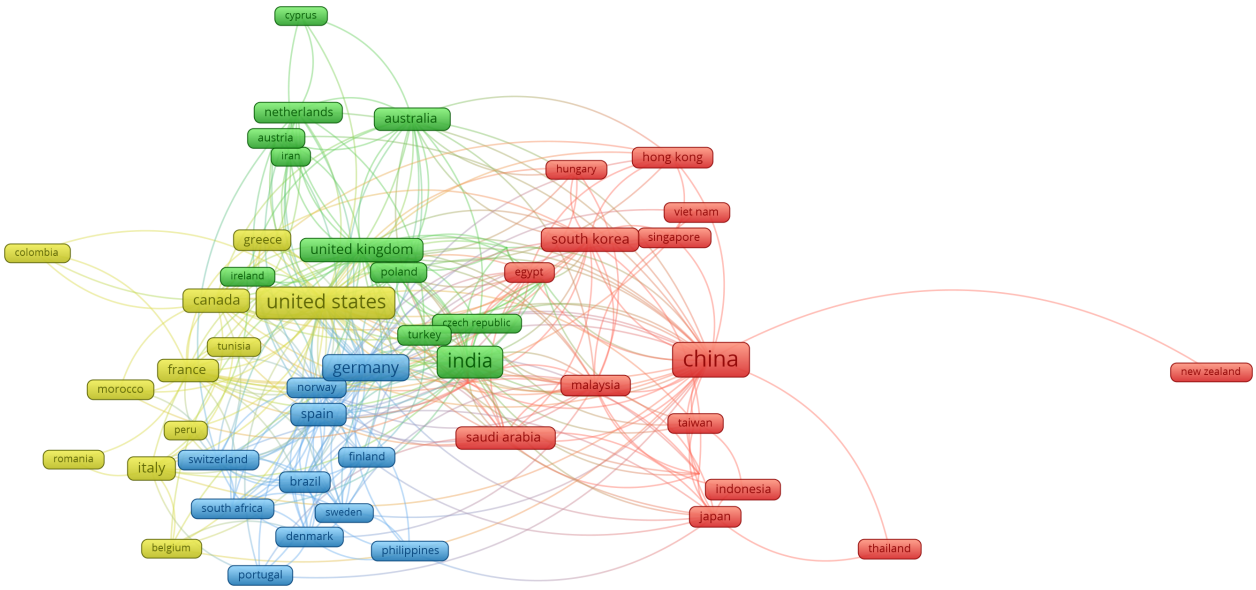

Рис. 1.4. Мережева візуалізація міжнародної співпраці.

1.5. Обговорення

Бібліометричний аналіз розкриває чіткий наратив технологічної еволюції в комп'ютерному зорі для застосувань навчання та моніторингу, що характеризується трьома окремими фазами, які відображають ширші тенденції в штучному інтелекті та машинному навчанні. Перша фаза, що охоплює 2008-2014 роки, була домінована традиційними методами комп'ютерного зору, що покладалися на ручне виокремлення ознак та класичні алгоритми машинного навчання. Протягом цього періоду дослідники фокусувалися на технологіях конструювання ознак, таких як SIFT, SURF та HOG, у поєднанні з класифікаторами типу Support Vector Machines та Random Forests.

Друга фаза, починаючи приблизно з 2015 року та прискорюючись до 2019 року, стала свідком революції глибинного навчання, що фундаментально трансформувала галузь. Прийняття згорткових нейронних мереж (CNN) для задач комп'ютерного зору усунуло потребу в ручному конструюванні ознак, дозволяючи системам навчатися оптимальним репрезентаціям безпосередньо з даних [17; 44]. Запровадження архітектур, таких як YOLO, Faster R-CNN та SSD, забезпечило детекцію об'єктів у реальному часі з безпрецедентною точністю, уможливлюючи складні застосування моніторингу [23; 32].

Третя фаза, що виникає з 2020 року далі, представляє пото-

чний передній край досліджень комп'ютерного зору, що характеризується інтеграцією передових технологій штучного інтелекту та адаптацією до середовищ периферійних обчислень. Поява візуальних трансформерів у нашому аналізі ключових слів (середній рік 2024.3) сигналізує про потенційну зміну парадигми від домінування CNN [24; 52].

Пандемія COVID-19 виникає з нашого аналізу як переломний момент, що прискорив як дослідницьку продукцію, так і практичне розгортання систем моніторингу комп'ютерного зору. Збільшення публікацій на 69% протягом 2020 року представляє більше, ніж кількісний сплеск; воно відображає фундаментальний зсув у дослідницьких пріоритетах та фокусі застосувань. Пандемія створила негайні потреби в безконтактних рішеннях моніторингу в множинних галузях [8; 14].

Географічний розподіл дослідницьких внесків розкриває як заохочуючу глобальну участь, так і тривожні диспропорції, що можуть обмежувати розвиток галузі. Домінування Китаю в обсязі публікацій (443 документи) відображає його національну стратегію лідерства в штучному інтелекті та субстантивні дослідницькі інвестиції. Обмежена репрезентація з Африки та Латинської Америки в нашому аналізі розкриває значну прогалину в глобальній дослідницькій участі.

Наш аналіз розкриває нову тенденцію до нейроморфних систем зору, що представляють фундаментальний відхід від традиційного покадрового зображення. Подієві камери, що захоплюють лише зміни у візуальній сцені замість повних кадрів, пропонують значні переваги для застосувань постійного моніторингу.

Поява самоконтрольованого навчання в нашому аналізі ключових слів відображає зростаюче визнання того, що дефіцит розмічених даних є фундаментальним вузьким місцем у розгортанні систем моніторингу [29].

Конвергенція комп'ютерного зору з великими мовними моделями представляє передову галузь, що ще не є виразною в нашому аналізі ключових слів, але виникає в недавніх публікаціях.

Висновки до розділу 1

1. Проведено комплексний бібліометричний аналіз 1,876 документів з бази даних Scopus за період 2008-2025 років, що охоплює

дослідження в галузі комп'ютерного зору для застосувань навчання та моніторингу.

2. Виявлено три фази розвитку галузі: період традиційних методів (2008-2018, середньо 45 публікацій/рік), сплеск під час COVID-19 (2019-2022, 180 публікацій/рік) та стабілізацію (2023-2025, 165 публікацій/рік), що демонструє 69% зростання публікаційної активності в пандемічний період.

3. Ідентифіковано чотири основні тематичні кластери з 49 уточнених ключових слів: базові технології (computer vision – 1,162 появи), методи глибинного навчання (deep learning – 784 появи), застосування моніторингу та технічні методи.

4. Встановлено, що Китай лідирує за кількістю публікацій (443 документи, 23.6%), тоді як США домінують за впливом цитувань (10,067 цитувань, середньо 34.7 на документ), з найсильнішою двосторонньою співпрацею між цими двома країнами (20 зв'язків).

5. Виявлено нові технологічні тенденції: візуальні трансформери (середній рік публікації 2024.3), периферійні обчислення (2023.5) та IoT (2022.7), що вказують на майбутні напрямки розвитку галузі.

6. Визначено географічні диспропорції з обмеженою участю країн Африки та Латинської Америки, що створює можливості для розширення глобальної дослідницької співпраці та трансферу технологій.

7. Ідентифіковано високовпливові галузі застосування: структурний моніторинг здоров'я (45.3 середніх цитувань), медична візуалізація (38.7), автономні транспортні засоби (35.2), що корелюють з практичною значущістю та суспільним впливом.

8. Виявлено нові галузі застосування: точне землеробство (34 статті, середній рік 2023.1), безпека будівництва (28 статей, 2022.8), екологічний моніторинг (41 стаття, 2022.5), що представляють перспективні напрямки для майбутніх досліджень.

Висновки до 1 розділу

1. Проведено комплексний бібліометричний аналіз 1,876 документів з бази даних Scopus за період 2008–2025 років, що охоплює дослідження в галузі комп'ютерного зору для застосувань навчання та моніторингу.

2. Виявлено три фази розвитку галузі: період традиційних методів (2008–2018, середньо 45 публікацій на рік), сплеск під час пандемії COVID-19 (2019–2022, 180 публікацій на рік) та стабілізацію (2023–2025, 165 публікацій на рік), що демонструє 69% зростання публікаційної активності в пандемічний період.

3. Ідентифіковано чотири основні тематичні кластери з 49 уточнених ключових слів: базові технології (computer vision – 1,162 появи), методи глибинного навчання (deep learning – 784 появи), застосування моніторингу та технічні методи.

4. Встановлено, що Китай лідирує за кількістю публікацій (443 документи, 23,6%), тоді як США домінують за впливом цитувань (10,067 цитувань, середньо 34,7 на документ), з найсильнішою двосторонньою співпрацею між цими країнами.

5. Виявлено нові технологічні тенденції: візуальні трансформери (середній рік публікації 2024,3), периферійні обчислення (2023,5) та Інтернет речей (2022,7), що вказують на перспективні напрямки розвитку галузі.

6. Визначено географічні диспропорції з обмеженою участю країн Африки та Латинської Америки, що створює можливості для розширення глобальної дослідницької співпраці та трансферу технологій.

7. Встановлено, що пандемія COVID-19 стала каталізатором інновацій у системах моніторингу на основі комп'ютерного зору, прискоривши розвиток безконтактних рішень та систем віддаленого нагляду.

РОЗДІЛ 2
ТЕОРЕТИЧНІ ОСНОВИ СИСТЕМ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ МОНІТОРИНГУ НАВЧАЛЬНОЇ АКТИВНОСТІ

2.1. Основи комп'ютерного зору

2.1.1. Визначення та сфери застосування

Комп'ютерний зір (англ. Computer Vision) розглядають як галузь, що вивчає методи автоматичного аналізу зображень і відео з метою отримання високорівневої інформації про сцену [47]. На відміну від класичних систем обробки сигналів, комп'ютерний зір орієнтований не лише на покращення якості зображення, а й на інтерпретацію його змісту: виявлення та розпізнавання об'єктів, оцінку їхніх просторових взаємовідносин, аналіз руху тощо.

Сучасні дослідження свідчать про те, що комп'ютерний зір перетворився на ключову технологію для широкого спектра застосувань: медична діагностика, автономний транспорт, промисловий контроль якості, сільське господарство, відеоспостереження та безпека, робототехніка, а також освітні системи [12; 21]. У більшості таких застосувань ціль полягає в тому, щоб замінити або доповнити людське спостереження автоматизованим аналізом.

До базових задач комп'ютерного зору належать [20; 47]:

- **Класифікація зображень** – віднесення всього зображення до одного з наперед визначених класів (наприклад, "аудиторія порожня" / "аудиторія зайнята").
- **Детекція об'єктів** – локалізація об'єктів на зображенні у вигляді прямокутних областей (обмежувальних рамок) з одночасною класифікацією кожної області.
- **Сегментація** – поділ зображення на області, що відповідають окре-

мим об'єктам або класам (семантична та екземплярна сегментація).

- **Відстеження об'єктів** – відстеження положення одного чи кількох об'єктів у послідовності кадрів відео.
- **Оцінка поз (pose estimation)** – відновлення положення частин тіла людини (ключових точок скелету) за одним або кількома кадрами.
- **Розпізнавання дій** – інтерпретація послідовності поз і траєкторій як певного типу активності (читання, підняття руки, розмова тощо).

У контексті моніторингу навчальної активності комп'ютерний зір дозволяє:

- автоматизувати контроль присутності учнів;
- оцінювати залученість (engagement) за позою, положенням голови, напрямком погляду;
- аналізувати взаємодію вчителя з класом (кількість піднятих рук, розподіл уваги між учнями);
- збирати об'єктивні кількісні показники для подальшої аналітики та адаптації навчального процесу.

Таким чином, комп'ютерний зір виступає фундаментальною технологією для перетворення суб'єктивних спостережень у об'єктивні метрики навчальної активності.

2.1.2. Етапи обробки візуальної інформації

Більшість систем комп'ютерного зору реалізують конвеєр обробки даних, що включає послідовні етапи: від захоплення зображень до прийняття рішень на основі отриманих ознак [12; 47]. Незважаючи на те, що сучасні глибинні моделі часто навчаються наскрізно, у прикладних системах зручно розглядати явну структуру таких етапів.

Типовий конвеєр обробки візуальної інформації для моніторингу навчальної активності можна описати так:

1. **Захоплення даних (Acquisition).** Вхідними даними є послідовність RGB-кадрів з однієї або кількох камер, встановлених в аудиторії. На цьому етапі важливими є частота кадрів, роздільна здатність і стабільність точки огляду.

2. **Попередня обробка (Pre-processing).** Виконується нормалізація інтенсивностей, зміна розміру до формату, сумісного з нейронною мережею, згладжування шумів, компенсація змін освітлення тощо [47]. Метою є приведення сирих даних до стандартизованого вигляду, що полегшує подальше навчання та застосування моделей.

3. **Детекція об'єктів / поз (Detection).** Глибинна нейронна мережа (наприклад, одноетапний детектор на зразок YOLO/SSD) виконує локалізацію учнів та інших релевантних об'єктів у кадрі й, за потреби, оцінює позу (координати ключових точок тіла) [19].

4. **Відстеження (Tracking).** Для кожного нового кадру встановлюється відповідність між новими детекціями та траєкторіями, побудованими на попередніх кадрах. Це дозволяє відстежувати конкретного учня в часі та аналізувати динаміку його поведінки.

5. **Вилучення ознак (Feature extraction).** На основі координат ключових точок, траєкторій, положення в аудиторії тощо формуються компактні числові ознаки: кути в суглобах, ступінь нахилу корпусу чи голови, час, проведений у певному стані (сидить прямо, дивиться на дошку, спілкується з сусідом тощо).

6. **Прийняття рішень (Decision making).** На підставі обчислених ознак використовується класифікаційна або регресійна модель (наприклад, градієнтний бустинг або невелика нейронна мережа), що оцінює рівень залученості окремого учня чи класу в цілому. Далі ці локальні оцінки агрегуються у глобальні показники, такі як середній індекс уваги за урок або за його окремі інтервали.

Подібна модульна структура спрощує проєктування й аналіз системи: кожен етап можна оптимізувати окремо, водночас забезпечуючи узгодженість усього конвеєра обробки візуальної інформації.

2.2. Нейронні мережі для детекції об'єктів

Сучасні методи детекції об'єктів майже повністю базуються на глибинних згорткових нейронних мережах (CNN). На відміну від класичних підходів з ручним конструюванням ознак (SIFT, HOG тощо), CNN автоматично навчаються багаторівневим представленням, здатним фіксувати як низькорівневі структури (контури, текстури), так і високорівневі поняття (обличчя, люди, жести) [20; 22].

Огляд існуючих архітектур показує, що згорткові мережі стали дефакто стандартом для задач класифікації, детекції та сегментації зображень [21]. Для детекції об'єктів CNN поєднують із спеціалізованими компонентами: регресією координат прямокутників, механізмами генерації регіонів-кандидатів, багатомасштабним представленням ознак тощо [19].

У задачах, де потрібна робота в реальному часі (відеоаналітика, моніторинг аудиторії, системи безпеки), особливу увагу приділяють архітектурам, що поєднують високу точність із низькою затримкою обробки кадрів. Саме до цього класу належать одноетапні детектори, такі як сімейство YOLO, SSD та їхні сучасні модифікації.

2.2.1. Еволюція архітектур детекції

Розвиток методів детекції об'єктів можна умовно поділити на кілька етапів [19]. Перші системи використовували ковзне вікно та вручну сконструйовані ознаки (Haar-like, HOG) у поєднанні з класичними класифікаторами. Такі підходи були обчислювально затратними й не забезпечували достатньої точності в складних сценах.

Перехід до глибинного навчання відбувся з появою сімейства регіон-орієнтованих мереж:

- **R-CNN, Fast R-CNN, Faster R-CNN** – двоетапні (two-stage) детектори, що спочатку генерують обмежену кількість регіонів-кандидатів, а потім класифікують їх за допомогою CNN. Faster R-CNN вводить мережу регіональних пропозицій (Region Proposal Network), яка генерує пропозиції безпосередньо на згорткових ознаках, значно прискорюючи обробку [25]. Такі моделі забезпечують високу точність, але залишаються відносно повільними для задач жорс-

ткого реального часу.

- **SSD, YOLO** – одноетапні (one-stage) детектори, що відмовляються від явного етапу генерації регіонів. SSD використовує множину "базових" (default) прямокутників різних масштабів та пропорцій на кількох рівнях ознак і в одній згортковій мережі одночасно прогнозує координати та класи для всіх комірок сітки [46]. Сімейство YOLO розглядає детекцію як задачу прямої регресії від пікселів до координат та класів об'єктів, що дозволяє досягати високої швидкості обробки кадрів [60].

Подальший розвиток пов'язаний із удосконаленням одноетапних архітектур: використанням багатомасштабних пірамід ознак (feature pyramid networks), переходом від якірних до безякірних схем, застосуванням більш ефективних функцій втрат та механізмів уваги [19]. Це дозволило суттєво покращити співвідношення між точністю та швидкістю й зробило одноетапні детектори стандартним вибором для відеоаналітики.

У задачі моніторингу навчальної активності важливо отримувати оцінки положення та поз учнів у режимі близькому до реального часу. Тому в даній роботі використовується сучасний одноетапний детектор сімейства YOLO, який поєднує високу точність локалізації людей та їхніх поз із можливістю обробляти відео з частотою, достатньою для онлайн-аналізу поведінки в аудиторії.

2.2.2. Архітектура YOLOv8

Сімейство моделей YOLO (You Only Look Once) належить до одноетапних детекторів об'єктів, які виконують локалізацію та класифікацію об'єктів за один прохід зображення крізь мережу, що забезпечує роботу в режимі реального часу [60]. На відміну від двоетапних підходів, де спочатку генеруються регіони-претенденти, а потім відбувається класифікація, YOLO безпосередньо регресує координати прямокутників та ймовірності класів на регулярній сітці зображення.

Для кожної передбаченої рамки модель оцінює ймовірність наявності об'єкта $P(\text{obj})$ та ступінь збігу з істинною розміткою через метрику Intersection over Union (IoU). Узагальнена міра впевненості у детекції має

вигляд [60]:

$$Conf = P(\text{obj}) \cdot IoU(B_{\text{pred}}, B_{\text{gt}}), \tag{2.1}$$

де B_{pred} – передбачений прямокутник, а B_{gt} – відповідна істинна розмітка. Після прогнозування великої кількості рамок застосовується неперекривне придушення (Non-Maximum Suppression, NMS), яке відкидає дублікати та залишає найбільш впевнені детекції.

У сучасних модифікаціях YOLO архітектура детектора традиційно описується як композиція трьох логічних частин: *backbone*, *neck* та *head*. *Backbone* вилучає багаторівневі візуальні ознаки, *neck* виконує їх багатомасштабну агрегацію, а *head* перетворює узагальнені ознаки на конкретні детекції (координати, класи та оцінки впевненості). Такий поділ використовується і в YOLOv8 [53; 59].

YOLOv8 [53] є однією з найсучасніших реалізацій сімейства YOLO та підтримує задачі детекції, сегментації, класифікації, оцінки поз тощо [53]. На відміну від попередніх версій, YOLOv8 використовує безякірну (*anchor-free*) голову детектора з розділеними гілками для регресії координат та класифікації, що спрощує адаптацію моделі до нових наборів даних та зменшує кількість гіперпараметрів [59].

Архітектура YOLOv8 (рис. 2.1) складається з таких основних компонентів [53; 59]:

- **Backbone** – покращений варіант мережі типу CSP, у якій стандартні блоки C3 (YOLOv5) замінено на більш ефективні блоки C2f. Вони забезпечують багатший розподіл градієнтів завдяки розгалуженню й повторному об'єднанню каналів, що дозволяє підвищити якість представлення ознак при подібній кількості параметрів.

- **SPPF-блок** (Spatial Pyramid Pooling – Fast) у верхній частині backbone виконує просторове пірамідальне згортання з різними розмірами вікон, що розширює ефективне поле зору мережі та покращує виявлення об'єктів різного масштабу.

- **Neck** – багатомасштабна мережа об'єднання ознак, що поєднує високо- і низькорівневі ознаки через послідовні операції згортки, зменшення та збільшення роздільності. Це підвищує чутливість моделі до дрібних об'єктів та перекриттів.

- **Безякірна голова (anchor-free head)** – розділена голова детектора, де для кожної точки карти ознак окремі гілки відповідають за локалізацію (регресія рамки), об'єктність та класифікацію. Такий поділ дає змогу краще оптимізувати одночасно локалізацію та класифікацію об'єктів [15; 59].

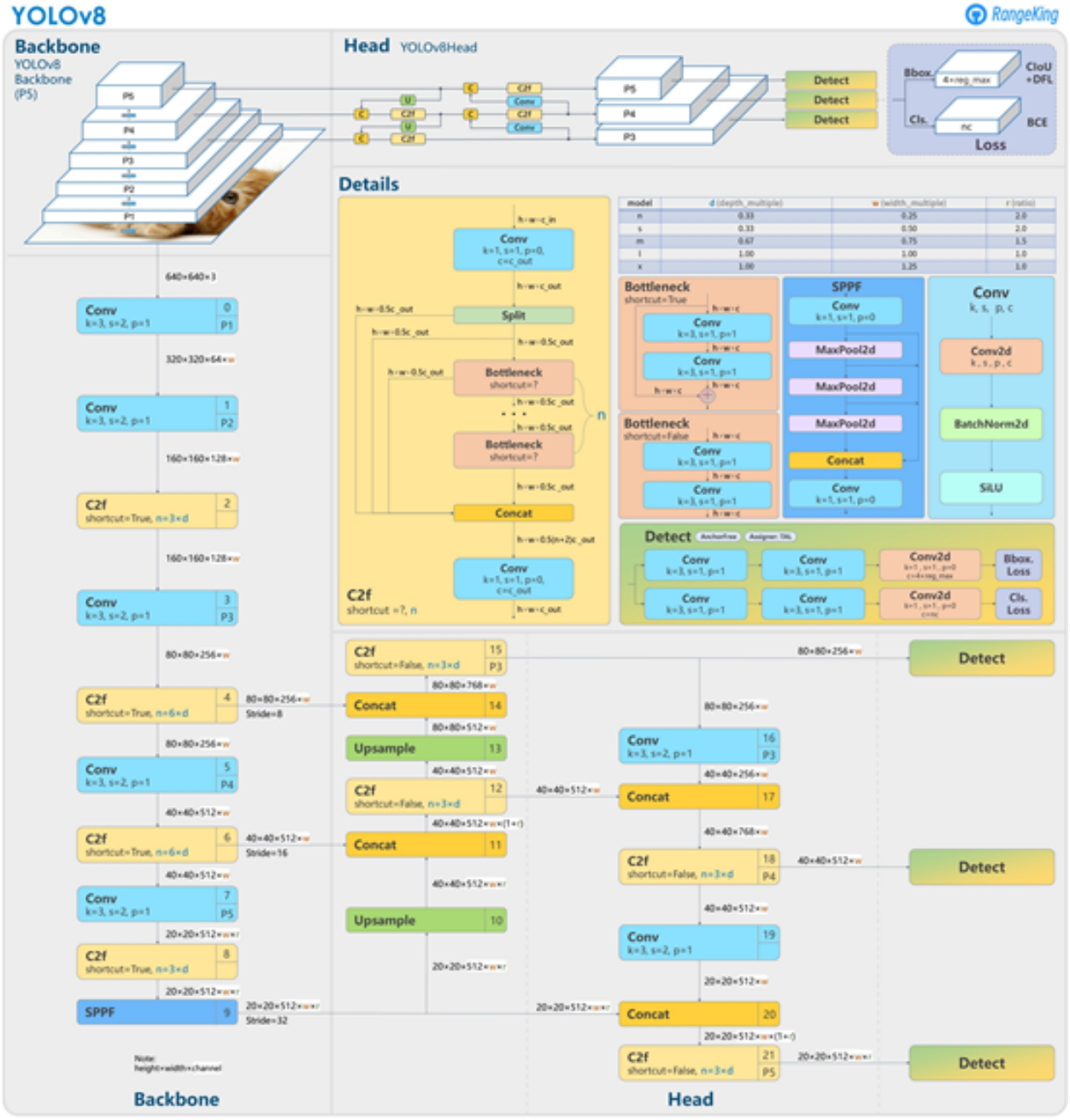

Рис. 2.1. Схема архітектури YOLOv8 (backbone–neck–head) [59].

YOLOv8 реалізується у кількох масштабах (n, s, m, l, x), що відрізняються глибиною та шириною мережі й дозволяють балансувати між точністю та швидкодією. Для задачі моніторингу навчальної актив-

ності у цій роботі використовується варіант `yolov8m-pose`, який забезпечує прийнятний компроміс між якістю оцінки поз та швидкістю обробки відео в режимі, близькому до реального часу [53].

На основі огляду архітектури YOLOv8 та її застосувань [27; 53; 59] можна виділити кілька переваг, важливих саме для моніторингу аудиторної активності:

- **Висока швидкодія:** одноетапна архітектура та оптимізовані блоки C2f дозволяють детектувати учнів у класі зі швидкістю десятків кадрів на секунду навіть на споживчому GPU.
- **Підтримка оцінки поз (pose estimation):** спеціальні моделі `yolov8-pose` розширюють детектор ключовими точками скелету людини, що дає змогу аналізувати положення голови, плечей та рук учнів [53].
- **Гнучкість:** безякірна голова не потребує підбору наборів якірних рамок, що суттєво спрощує перенавчання моделі на спеціалізованому наборі даних класної кімнати.
- **Модульність:** архітектура легко розширюється додатковими модулями уваги, альтернативними схемами neck (наприклад, HR-FPN) та вдосконаленими функціями втрат, що важливо для роботи в складних умовах (перекриття, зміни освітлення тощо) [27].

2.2.3 Особливості навчання та оптимізації YOLOv8

Ефективність детектора залежить не лише від архітектури, а й від обраної стратегії навчання: аугментації, функцій втрат, оптимізатора та форматів чисел. Сучасні реалізації YOLOv8 (включно з Ultralytics) використовують низку прийомів, що дозволяють швидко адаптувати модель навіть до невеликих спеціалізованих наборів даних [15; 53; 54].

Для підвищення узагальнювальної здатності та стійкості до змін умов зйомки застосовуються такі типи *аугментації* [53; 54]:

- геометричні перетворення (масштабування, повороти, горизонтальне віддзеркалення, випадкові зсуви), що моделюють різне розташування камери та учнів;

- фотометричні перетворення (зміна яскравості, контрасту, насиченості, додавання шуму) для моделювання змін освітлення в аудиторії;
- композиційні методи (Mosaic, MixUp), коли кілька зображень комбінуються в один колаж, що збільшує різноманітність фонів та масштабів об'єктів.

У задачі моніторингу навчальної активності важливо зберігати реалістичні пропорції сцен класної кімнати, тому ступінь геометричних перетворень обмежується таким чином, щоб не спотворювати суттєво положення голови та рук учнів.

У реалізації YOLOv8 використовується *комбінована функція втрат*, яка включає три компоненти [15]:

- втрати регресії координат рамок на основі різновидів IoU (CIoU, SIoU, EIoU тощо), що враховують не лише площу перетину, а й відстань між центрами та відношення сторін;
- бінарну крос-ентропію (BCE) для навчання об'єктності, тобто наявності/відсутності об'єкта;
- мультикласову крос-ентропію або фокальну втрату для класифікації класів об'єктів.

У варіантах, що базуються на HR–YOLOv8, для регресії координат може застосовуватися модифікований IoU-критерій типу IS-IoU, який підсилює штраф за невдалу форму та розмір рамки [27].

Традиційно глибинні моделі навчаються з використанням 32-бітної плаваючої арифметики (FP32). Однак сучасні GPU мають апаратну підтримку 16-бітних форматів (FP16/BF16), що дозволяє суттєво прискорити навчання та зменшити споживання пам'яті. Техніка змішаного навчання (mixed precision training) поєднує переваги обох форматів: обчислення виконуються у 16-бітному форматі, але критичні величини (ваги, акумулятори градієнтів) зберігаються в FP32 [34].

У роботах NVIDIA показано, що використання змішаної точності дозволяє прискорити навчання без втрати точності за рахунок: масштабування функції втрат (loss scaling), зберігання «майстер-копії» ваг у FP32

та акумуляції градієнтів у високій точності [34]. Популярні пояснення цієї техніки наведені також у оглядових статтях [48]. На рис. 2.2 схематично показано типовий цикл навчання з використанням змішаної точності.

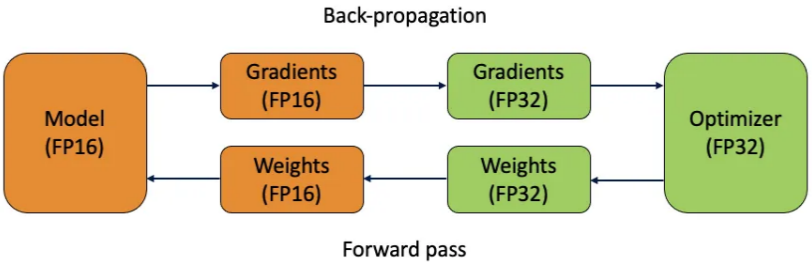

Рис. 2.2. Схема однієї ітерації змішаного навчання з використанням форматів FP16 та FP32 [59]).

У бібліотеці PyTorch (і в реалізації Ultralytics YOLOv8) mixed precision підтримується засобами `torch.cuda.amp` та відповідними параметрами конфігурації, наприклад `amp=True` або `device=auto` при запуску навчання [54; 55]. Це дає змогу або збільшити розмір пакету, або працювати з вищою роздільністю вхідних зображень при тих самих обчислювальних ресурсах, що особливо важливо при аналізі відеопотоків з класної кімнати.

Отже, поєднання сучасної архітектури YOLOv8 (зокрема, її модифікації HR–YOLOv8) з ефективними методами навчання (аугментація, розвинені функції втрат, змішана точність) створює потужне технічне підґрунтя для систем моніторингу навчальної активності учнів, що реалізується в наступному розділі роботи.

2.3. Моделі детекції поз людини

Задача оцінки поз (human pose estimation) полягає у визначенні просторового розташування анатомічних ключових точок тіла (суглоби, голова, кисті тощо) за вхідним зображенням або відео [1; 13]. У загальному випадку розрізняють 2D та 3D оцінку поз, однак у даній роботі використовуються лише 2D-координати ключових точок, отримані з однієї RGB-камери.

Сучасні методи pose estimation здебільшого базуються на глибокому навчанні й поділяються на два основні класи[1]:

- **top-down** (зверху вниз) – спочатку детектуються обличчя/тіла, потім для кожної рамки оцінюються ключові точки;
- **bottom-up** (знизу вгору) – спочатку оцінюються ключові точки всіх людей на сцені, потім вони групуються в окремі скелети.

У даній роботі використовується одноетапна модель детекції, яка одночасно визначає рамку тіла та координати ключових точок, що відноситься до bottom-up / підходів на основі детекції, оптимізованих для роботи в режимі реального часу.

2.3.1. Постановка задачі

Нехай I_t – кадр відео в момент часу t. Завдання 2D pose estimation полягає в оцінці множини поз

$$\mathcal{P}_t = \{P_t^{(1)}, P_t^{(2)}, \ldots, P_t^{(M_t)}\},$$

де M_t – кількість людей у кадрі, а кожна поза $P_t^{(m)}$ описується множиною K ключових точок

$$P_t^{(m)} = \{\mathbf{k}_{1,t}^{(m)}, \ldots, \mathbf{k}_{K,t}^{(m)}\}, \quad \mathbf{k}_{j,t}^{(m)} = (x_{j,t}^{(m)}, y_{j,t}^{(m)}, c_{j,t}^{(m)}).$$

Тут (x, y) – координати точки у зображенні, а $c \in [0, 1]$ – впевненість моделі в детекції даної точки.

Оцінка поз застосовується як проміжний рівень подання для широкого спектра задач: розпізнавання дій, аналіз осанки, жестів, аналізу взаємодії людина–комп'ютер тощо [13]. У контексті даної роботи 2D-скелети використовуються для побудови ознак залученості учнів (нахил голови, положення рук, постава корпусу тощо).

2.3.2. Формат COCO Keypoints

Для опису скелета людини використовується стандартний формат *COCO Keypoints*, запропонований у межах набору даних Microsoft COCO [33]. У цьому форматі для кожної людини анотується $K = 17$ ключових точок:

1) ніс;

2) ліве та праве око;

3) ліве та праве вухо;

4) ліве та праве плече;

5) лівий та правий лікоть;

6) ліве та праве зап'ястя;

7) ліве та праве стегно;

8) ліве та праве коліно;

9) ліва та права щиколотка.

Ключові точки з'єднуються у скелет (граф) фіксованої структури, що дозволяє обчислювати кути в суглобах, відносні довжини сегментів тіла та інші геометричні характеристики. Таке подання є зручним для побудови ознак високого рівня (наприклад, "рука піднята", "учень сутулиться" тощо) і є де-факто стандартом для оцінки поз на еталонних наборах даних COCO та інших [1].

У COCO також запропоновано метрику *Object Keypoint Similarity* (OKS) та mean Average Precision (mAP) для оцінки якості локалізації ключових точок [33]. Хоча в даній роботі не проводиться порівняння якості різних моделей, вибір формату COCO забезпечує сумісність з існуючими методами та попередньо натренованими моделями.

2.3.3. Порівняння моделей pose estimation

Розвиток 2D pose estimation проходив від ранніх графових моделей частин тіла до сучасних глибоких нейронних мереж [1]. Серед найбільш впливових підходів можна виділити:

- **OpenPose** – bottom-up модель, що вводить поняття *part affinity fields* (PAF) для одночасної детекції ключових точок і їх групування у скелети, забезпечуючи реальний час для багатьох людей в кадрі [39].

- **Simple Baselines** – top-down підхід на базі ResNet з простими деконволюційними шарами для побудови карт ймовірностей ключових точок; попри простоту, демонструє високу точність на COCO [57].

- **HRNet** – високороздільна мережа, яка зберігає високу просторову роздільність на всіх рівнях ознак, що покращує точність локалізації дрібних суглобів (кисті, щиколотки)[16].

Огляди [1; 13] показують, що top-down підходи (Simple Baselines, HRNet) часто досягають вищої точності на статичних зображеннях, але є менш ефективними при великій кількості людей у кадрі. Bottom-up та одноетапні моделі на основі детекції (як-от OpenPose і сучасні модифікації на базі YOLO) забезпечують кращий компроміс між точністю й швидкодією, що є критичним для задач моніторингу класу в режимі реального часу.

У реалізації даної роботи застосовано саме одноетапний підхід на основі детекції до оцінки поз, який безпосередньо повертає рамки для кожного учня та 2D-координати ключових точок у форматі COCO. Це дозволяє подальше обчислення поведінкових ознак залученості при збереженні достатньо високої частоти кадрів.

2.4. Методи оцінки залученості учнів

Залученість (engagement) учня зазвичай розглядається як поєднання поведінкової, емоційної та когнітивної компонент [49]. У комп'ютерному зорі переважно аналізуються зовнішні прояви: поза тіла, міміка, жести, напрям погляду, динаміка рухів, а також контекст (час заняття, тип завдання тощо) [10; 58].

Залежно від доступних даних виділяють декілька груп підходів:

- аналіз лише відео (обличчя/поза, як у цій роботі);

- мультимодальні підходи, що комбінують відео, лог-файли платформи, сенсорні дані (пульс, ЕЕГ тощо) [58];

- комбінація об'єктивних ознак з суб'єктивними анкетами й успішністю навчання.

У запропонованій системі акцент зроблено на **поведінковій** та частково **емоційній** залученості, яку можна оцінити за позою тіла, положенням рук, нахилом голови та відкритістю очей. Для кожного учня спочатку обчислюються низькорівневі ознаки на основі скелета, а далі вони агрегуються відповідно до обраної схеми класифікації.

2.4.1. Розпізнавання залученості: огляд підходів

Перші роботи з автоматичного розпізнавання залученості базувалися переважно на статичних ознаках обличчя і класичних методах машинного навчання. Наприклад, Whitehill et al. [49] використовують ознаки на основі HOG та каскад класифікаторів для розпізнавання рівня залученості учнів за виразом обличчя.

Подальші дослідження активно застосовують глибокі згорткові мережі. Nezami et al. [10] пропонують двоетапну схему: спочатку нейронна мережа навчається на задачі розпізнавання базових емоцій, а потім ці ж ваги використовуються як ініціалізація для моделі розпізнавання залученості. Таке перенесення навчання дозволяє покращити якість на невеликих наборах даних залученості.

Сучасні мультимодальні підходи [58] комбінують відео, текстові й логдані, використовуючи глибокі мережі для моделювання часової динаміки та злиття ознак. У реальних аудиторіях створюються також спеціалізовані системи комп'ютерного зору, що аналізують позу тіла та жестові закономірності всіх учнів одночасно [11].

У даній роботі, виходячи з обмеженого обсягу даних та вимог до швидкодії, застосовується більш спрощений, але інтерпретований підхід: на основі поз кожного учня обчислюється набір геометричних ознак, які потім лінійно комбінуються в скалярний індекс уваги.

2.4.2. Ключові ознаки для оцінки уваги

На основі координат COCO keypoints для кожного учня у кадрі обчислюється низка ознак, що відображають його позу та поведінку:

1. **Оцінка напряму голови.** Використовуючи положення носа, очей і вух, наближено оцінюються кути нахилу голови по вертикалі (рі-

tch) та горизонталі (yaw). Подібні ознаки активно застосовуються для аналізу напрямку погляду та відвернення від екрану в роботах з моніторингу уваги [11; 49]. У нашій реалізації кут нахилу голови використовується для виявлення ситуацій, коли учень дивиться вниз (на стіл, телефон) або вбік від дошки.

2. **Відкритість очей та моргання.** На основі координат ключових точок повік обчислюється спрощений аналог *eye aspect ratio* (EAR) [45], який відображає ступінь відкритості очей. Тривалі періоди з низьким значенням ознаки можуть свідчити про сонливість або відсутність зорового контакту з дошкою/екраном.

3. **Положення рук.** Використовуючи координати плечей і зап'ясть, визначаються ситуації, коли рука піднята вище певного порогу відносно плеча (жест запитання/відповіді), коли руки знаходяться на парті чи під нею, а також наближена дистанція між руками. Ці ознаки відображають активні жести (підняття руки, жестикуляція) та пасивну позу (руки опущені).

4. **Постава і "сутулість".** На основі відстані між плечима та тазовими суглобами, а також кута між відповідними сегментами оцінюється ступінь нахилу корпусу вперед/назад. Значна "сутулість" може вказувати як на втому, так і на сильну зосередженість (учень схиляється над зошитом), тому ознака інтерпретується у поєднанні з іншими (напрям погляду, положення рук).

5. **Часові агрегати.** Для кожної з ознак розраховуються ковзні середні та експоненційні згладжування по вікну в декілька секунд. Це дозволяє приглушити випадкові коливання (наприклад, разовий погляд убік) і зосередитись на стійких характеристиках поведінки.

Таким чином, кожен учень у момент часу t описується вектором ознак $\mathbf{z}_t$, який включає геометричні характеристики голови, рук, корпусу та їх часові агрегати. Над цим вектором далі будується модель класифікації рівня залученості.

2.4.3. Моделі класифікації engagement

У літературі пропонуються різні типи моделей для віднесення спостережень до рівнів залученості:

- **Класичні методи машинного навчання** (SVM, Random Forest, логістична регресія), що навчаються на фіксованих векторних ознаках [49].
- **Глибокі CNN/RNN-моделі**, які працюють безпосередньо з послідовністю кадрів і навчаються наскрізно [10].
- **Мультимодальні глибокі архітектури**, що об'єднують відео, логдані, текстові ознаки тощо [58].

У даній роботі, зважаючи на обмежений обсяг розмічених даних, обрано **інтерпретовану евристичну модель**, яка прямо використовує описані вище ознаки. Для кожного учня:

а) за поточним вектором $\mathbf{z}_t$ обчислюється нормалізований індекс уваги $a_t \in [0, 1]$ як зважена сума компонент $\mathbf{z}_t$;

б) індекс згладжується експоненційним ковзним середнім, щоб уникнути різких стрибків між кадрами;

в) за порогами $0 < \tau_1 < \tau_2 < \tau_3$ індекс a_t відноситься до одного з дискретних станів: *inattentive*, *distracted*, *neutral*, *attentive*.

Таке правило відображає ідею, подібну до робіт [11; 49]: замість прямого передбачення одразу складного рівня залученості модель спочатку вимірює прості, інтерпретовані ознаки, пов'язані з позою тіла і голови, а вже потім комбінує їх у глобальний індекс.

На рівні всього класу регуляризований індекс уваги розраховується як середнє по всіх учнях у кадрі. Це дозволяє виявляти моменти, коли значна частина аудиторії втрачає увагу, і надалі використовувати таку метрику для аналізу ефективності уроку або для адаптивних підказок викладачу.

2.5. Метрики оцінки систем детекції

Для об'єктивної оцінки якості розробленої системи моніторингу необхідно використовувати формалізовані метрики, які дозволяють порівнювати різні моделі між собою та контролювати вплив налаштувань навчання. У комп'ютерному зорі традиційно застосовуються різні групи показників для детекції об'єктів, оцінки поз та задач класифікації (зокрема, визначення рівня залученості) [1; 33; 49; 50].

2.5.1. Метрики для детекції об'єктів

Базовою величиною для оцінки якості локалізації об'єктів є **Intersection over Union** (IoU)[50]. Для передбаченого прямокутника B_p та еталонного (істинного) B_{gt} вона визначається як

$$\mathrm{IoU}(B_p, B_{gt}) = \frac{\mathrm{area}(B_p \cap B_{gt})}{\mathrm{area}(B_p \cup B_{gt})}. \tag{2.2}$$

Якщо $\mathrm{IoU} \geq \tau$ (де τ – заданий поріг, наприклад 0,5), детекція вважається *правильною* (True Positive, TP). Передбачений об'єкт, який не відповідає жодному еталонному з достатнім IoU, вважається хибним спрацьовуванням (False Positive, FP), а відсутність детекції для наявного об'єкта – пропуском (False Negative, FN).

На основі цих величин обчислюються[30; 33]:

$$\mathrm{Precision} = \frac{TP}{TP + FP}, \tag{2.3}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN}, \tag{2.4}$$

$$\mathrm{F1} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \tag{2.5}$$

Precision характеризує частку коректних детекцій серед усіх спрацювань моделі, тоді як Recall описує частку правильно виявлених об'єктів серед усіх наявних.

Для повнішої оцінки моделі будують **криву Precision–Recall** та обчислюють **Average Precision** (AP) як площу під цією кривою. У PASCAL VOC AP визначається при фіксованому порозі IoU (як правило, 0,5)[50]. У

COCO використовується більш жорстка схема: AP усереднюється за набором порогів $\tau \in [0{,}5; 0{,}95]$ з кроком 0,05[33]:

$$\mathrm{AP} = \frac{1}{T} \sum_{t=1}^{T} \mathrm{AP}_{\tau_t}, \quad \tau_t \in \{0{,}5, 0{,}55, \ldots, 0{,}95\}. \tag{2.6}$$

Mean Average Precision (mAP) обчислюється як середнє значення AP по всіх класах:

$$\mathrm{mAP} = \frac{1}{C} \sum_{c=1}^{C} \mathrm{AP}_c, \tag{2.7}$$

де C – кількість класів, а AP_c – середня точність для класу c.

У даній роботі основний інтерес становить клас `person`, оскільки завдання полягає у виявленні саме учнів у кадрі. Тому у процесі налаштування моделі використовуються показники AP_{person} та mAP при порозі IoU = 0,5, а також аналізуються Precision і Recall для оцінки балансу між кількістю пропусків та хибних спрацювань.

2.5.2. Метрики для pose estimation

Для оцінки якості відновлення поз людини використовуються метрики, які вимірюють точність локалізації ключових точок (суглобів) відносно еталонних анотацій. Однією з найпоширеніших є **Percentage of Correct Keypoints** (PCK) [2]:

$$\mathrm{PCK}(\alpha) = \frac{1}{NK} \sum_{n=1}^{N} \sum_{k=1}^{K} \mathbb{I}\left(\frac{\|\hat{\mathbf{p}}_{n,k} - \mathbf{p}_{n,k}\|_2}{d_n} < \alpha \right), \tag{2.8}$$

де N – кількість зразків, K – кількість ключових точок, $\hat{\mathbf{p}}_{n,k}$ та $\mathbf{p}_{n,k}$ – передбачені та еталонні координати точки k у зразку n, d_n – нормувальна відстань (наприклад, розмір голови або торсу), а α – допустима відносна похибка. Варіант метрики **PCKh** нормує відстань на розмір голови і активно використовується в еталонному наборі даних MPII Human Pose [2].

У COCO для оцінки якості відновлення ключових точок використовується **Object Keypoint Similarity** (OKS), яка є аналогом IoU для скелетів [33]:

$$\mathrm{OKS} = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \cdot \mathbb{I}(v_i > 0)}{\sum_i \mathbb{I}(v_i > 0)}, \tag{2.9}$$

де d_i – відстань між передбаченою та еталонною позицією точки i, s – масштаб об'єкта (площа рамки), k_i – емпіричний коефіцієнт для конкретної точки (відображає її "складність"), v_i – видимість точки. На основі OKS обчислюється AP та mAP за аналогією з детекцією об'єктів.

У даній роботі модель оцінки поз використовується переважно як проміжний модуль для побудови поведінкових ознак. Власне навчання з нуля не проводиться, тому повна переоцінка за метриками PCK/OKS не є необхідною. Натомість використовується інформація про якість попередньо натренованої моделі YOLOv8m-pose на COCO keypoints, наведена в документації бібліотеки, а валідація в конкретних умовах аудиторії виконується за допомогою візуального аналізу та вибіркової ручної розмітки невеликої підмножини кадрів.

2.5.3. Метрики для engagement detection

Оцінка якості визначення рівня залученості є задачею класифікації (або порядкової регресії), для якої застосовуються стандартні метрики аналізу **матриці помилок** [10; 49]:

- **Accuracy** – частка правильно класифікованих прикладів серед усіх;
- **Precision**, **Recall**, **F1-score** – обчислюються для кожного класу за аналогією з метриками детекції;
- **Macro-F1** та **Weighted-F1** – усереднення F1 по класах (з рівними або зваженими вагами), що є особливо важливим при дисбалансі класів;
- **ROC-AUC** – площа під ROC-кривою, що відображає компроміс між чутливістю та специфічністю для бінарних задач.

У роботах з розпізнавання залученості студентів часто використовується також **коефіцієнт узгодженості анотацій** (наприклад, *κ Коена*) між різними експертами та між експертами й моделлю [49]. Це дозволяє оцінити, наскільки автоматизована система наближається до людського рівня узгодженості.

У запропонованій системі, з огляду на обмежений обсяг розмічених даних, первинна валідація моделі engagement detection ґрунтується на таких засадах:

- побудова матриці помилок для декількох уроків з ручною розміткою рівня уваги учнів учителем;
- обчислення точності, Precision, Recall та Macro-F1 для окремих класів (*inattentive*, *distracted*, *neutral*, *attentive*);
- аналіз типових помилок моделі (переплутування "нейтрального" та "уважного" станів, хибні спрацювання на жестах тощо).

Таким чином, комплекс метрик дозволяє окремо оцінити:

а) якість детекції учнів у кадрі (модуль детекції об'єктів);

б) точність відновлення їхніх поз (модуль pose estimation);

в) адекватність класифікації рівня уваги на основі поведінкових ознак (модуль engagement detection).

Це створює основу для кількісної перевірки працездатності всієї системи моніторингу навчальної активності.

Висновки до 2 розділу

1. Систематизовано теоретичні основи комп'ютерного зору як галузі, що охоплює класифікацію зображень, детекцію об'єктів, сегментацію, відстеження, оцінку поз та розпізнавання дій, із визначенням їх ролі в моніторингу навчальної активності.

2. Описано типовий конвеєр обробки візуальної інформації, що включає шість послідовних етапів: захоплення даних, попередню обробку, детекцію об'єктів та поз, відстеження, вилучення ознак і прийняття рішень.

3. Проаналізовано еволюцію архітектур детекції об'єктів від двоетапних (R-CNN, Faster R-CNN) до одноетапних детекторів (SSD, YOLO), обґрунтовано вибір архітектури YOLOv8 для задач моніторингу в реальному часі завдяки її високій швидкодії та підтримці оцінки поз.

4. Розкрито особливості архітектури YOLOv8, зокрема структуру backbone–neck–head, безякірну голову детектора, блоки C2f та SPPF, а також методи навчання (аугментація, комбіновані функції втрат, змішана точність).

5. Охарактеризовано підходи до оцінки поз людини (top-down та bottom-up), описано формат COCO Keypoints з 17 ключовими точками та порівняно основні моделі (OpenPose, Simple Baselines, HRNet).

6. Проаналізовано методи оцінки залученості учнів, що базуються на аналізі пози тіла, положення рук, нахилу голови та відкритості очей, із виділенням ключових ознак для побудови індексу уваги.

7. Систематизовано метрики оцінки якості систем детекції (IoU, Precision, Recall, mAP), оцінки поз (PCK, OKS) та класифікації залученості (Accuracy, F1-score, матриця помилок), що створює основу для кількісної валідації розробленої системи.

8. Обґрунтовано доцільність використання інтерпретованої евристичної моделі класифікації залученості, що базується на зважених геометричних ознаках поз із експоненційним згладжуванням, як компромісу між точністю та обчислювальною ефективністю.

РОЗДІЛ 3
ПРАКТИЧНА РЕАЛІЗАЦІЯ СИСТЕМИ МОНІТОРИНГУ НАВЧАЛЬНОЇ АКТИВНОСТІ УЧНІВ

3.1. Загальна архітектура програмної системи

Розроблена система моніторингу навчальної активності реалізована як послідовний конвеєр обробки відео, реалізований у файлі `yolov8_custom.py`. На практиці програма працює за таким сценарієм:

1) відкривається відеопотік з вебкамери або з файлу відеозапису уроку;

2) кожен N-й кадр подається в модель `YOLOv8m-pose`, яка знаходить усіх учнів та їхні ключові точки;

3) для кожного учня підтримується унікальний ідентифікатор (ID) та історія його станів за допомогою простого відстеження на основі метрики Intersection over Union (IoU);

4) з ключових точок обчислюються інформативні ознаки пози (нахил голови, положення рук, відкритість очей, сутулість);

5) на основі цих ознак для кожного учня обчислюється *індекс уваги* і призначається категорія (уважний, нейтральний, відволікся, неуважний);

6) на відео накладається візуалізація (рамки, скелети, підписи, зведена панель з *Class Attention Index*);

7) результати (відео, журнали) зберігаються для подальшого аналізу.

Структуру коду можна коротко подати як:

- блок **налаштування параметрів** (константи: пороги, кольори, граничний час траєкторій тощо);

- структура даних для зберігання **стану кожного учня** (ID, остання обмежувальна рамка, ключові точки, історія значень уваги, базові значення);
- функції **обробки кадру**:
  - виклик моделі `YOLOv8m-pose`,
  - оновлення траєкторій,
  - обчислення ознак та індексів уваги,
  - малювання накладань на кадрі;
- **головний цикл** читання відео, який запускає обробку кожного кадру, показує/записує результат і рахує частоту кадрів.

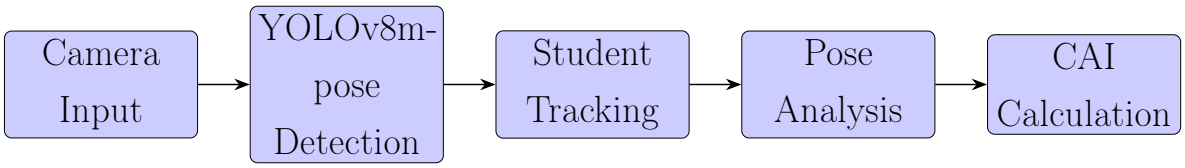

Рис. 3.1. Логічна схема роботи системи: від вхідного відео до індексу уваги класу (CAI).

3.2. Детекція учнів та ключових точок за допомогою YOLOv8m-pose

3.2.1. Ініціалізація моделі та вибір пристрою

У коді використовується попередньо навчена модель `yolov8m-pose.pt` з бібліотеки `ultralytics`. Модель не донавчалася на власних даних, а застосовується "як є" для відео з класної кімнати. Її завантаження та прив'язка до потрібного пристрою виконуються кількома рядками:

Лістинг 1: Ініціалізація моделі YOLOv8m-pose та вибір пристрою

```
from ultralytics import YOLO
import torch

```

```
# Select device: GPU if available, otherwise CPU
DEVICE = 'cuda:0' if torch.cuda.is_available() else 'cpu'

# Load pre-trained YOLOv8m pose model
model = YOLO('yolov8m-pose.pt')
model.to(DEVICE)

# Detection confidence threshold
CONF_THRESH = 0.3
```

Тут відбувається:

- перевірка наявності GPU за допомогою `torch.cuda.is_available()`;
- завантаження готових ваг моделі `yolov8m-pose.pt`;
- встановлення порогу впевненості `CONF_THRESH = 0.3`, який відсікає слабкі детекції.

3.2.2. Обробка окремого кадру відео

Основне завдання моделі на кожному кадрі – знайти всіх учнів та повернути для кожного прямокутник (обмежувальну рамку) і координати 17 ключових точок тіла. У коді це оформлено як функція:

Лістинг 2: Виклик моделі на одному кадрі та отримання людей з позами

```
import cv2
import numpy as np

def detect_people_with_poses(frame_bgr):
    """Run YOLOv8m-pose on a single BGR frame and return list
       of people."""
    # Ultralytics expects RGB images
    frame_rgb = cv2.cvtColor(frame_bgr, cv2.COLOR_BGR2RGB)

    # Run model with confidence threshold
    results = model(frame_rgb, conf=CONF_THRESH, verbose=False)
       [0]

    people = []
    for box, kps in zip(results.boxes.xyxy, results.keypoints.
       xy):
```

```
        x1, y1, x2, y2 = box.tolist()
        keypoints = kps.cpu().numpy()  # shape (17, 2) -> (x, y)
        people.append({
            "bbox": (x1, y1, x2, y2),
            "keypoints": keypoints,
        })
    return people
```

Повертається список словників, де для кожної людини є поле `bbox` та масив `keypoints`. Далі ці дані використовуються модулем відстеження й аналізу поз.

3.3. Відстеження учнів між кадрами

3.3.1. Структура даних для траєкторії учня

Для кожного учня підтримується окрема траєкторія з унікальним ID. Стан траєкторії зберігається у словнику:

- `id` – ідентифікатор учня;
- `bbox`, `keypoints` – актуальні координати;
- `missed` – кількість останніх кадрів, де учня не було видно;
- `baseline_samples` – накопичені зразки ознак для обчислення індивідуальних базових значень;
- `features_history` – коротка історія останніх значень ознак;
- `score_ema` – згладжений індекс уваги.

При появі нової людини створюється нова траєкторія:

Лістинг 3: Створення нової траєкторії для учня

```
import collections

next_student_id = 0
students = {}  # id -> track dict

```

```
def create_student_track(bbox, keypoints):
    """Create a new track for a detected student."""
    global next_student_id
    sid = next_student_id
    next_student_id += 1

    students[sid] = {
        "id": sid,
        "bbox": bbox,
        "keypoints": keypoints,
        "missed": 0,
        "baseline_samples": [],
        "features_history": collections.deque(maxlen=60),
        "score_ema": None,
    }
    return sid
```

3.3.2. Обчислення IoU та оновлення траєкторій

Для зіставлення детекцій між кадрами застосовується метрика перетину площ прямокутників – Intersection over Union. Вона реалізується так:

Лістинг 4: Спрощена функція обчислення IoU

```
def iou(boxA, boxB):
    """Compute Intersection over Union for two bounding boxes."
       ""
    xA = max(boxA[0], boxB[0])
    yA = max(boxA[1], boxB[1])
    xB = min(boxA[2], boxB[2])
    yB = min(boxA[3], boxB[3])

    inter_w = max(0.0, xB - xA)
    inter_h = max(0.0, yB - yA)
    inter_area = inter_w * inter_h

    areaA = max(0.0, (boxA[2] - boxA[0]) * (boxA[3] - boxA[1]))
    areaB = max(0.0, (boxB[2] - boxB[0]) * (boxB[3] - boxB[1]))

    return inter_area / float(areaA + areaB - inter_area + 1e
       -9)
```

Оновлення траєкторій відбувається в два кроки: спочатку для кожного існуючого учня шукається найбільш схожа детекція, потім для решти детекцій створюються нові траєкторії:

Лістинг 5: Оновлення траєкторій на основі нових детекцій

```
IOU_MATCH_THRESH = 0.15
MAX_MISSED_FRAMES = 150  # about 5 seconds at 30 FPS

def update_tracks(detections):
    """
    detections: list of dicts {"bbox": ..., "keypoints": ...}
    """
    unmatched_dets = set(range(len(detections)))

    # 1) Try to match each existing student with a new
       detection
    for sid, st in list(students.items()):
        best_iou = 0.0
        best_j = None
        for j in unmatched_dets:
            i = iou(st["bbox"], detections[j]["bbox"])
            if i > best_iou:
                best_iou, best_j = i, j

        if best_j is not None and best_iou >= IOU_MATCH_THRESH:
            # Update existing track
            st["bbox"] = detections[best_j]["bbox"]
            st["keypoints"] = detections[best_j]["keypoints"]
            st["missed"] = 0
            unmatched_dets.remove(best_j)
        else:
            # This student was not detected in the current
               frame
            st["missed"] += 1
            if st["missed"] > MAX_MISSED_FRAMES:
                del students[sid]

    # 2) Create new tracks for all unmatched detections
    for j in unmatched_dets:
        create_student_track(
```

```
            detections[j]["bbox"],
            detections[j]["keypoints"]
        )
```

Такий простий жадібний алгоритм зіставлення виявився достатнім для класної кімнати з фіксованою камерою та не дуже швидкими рухами учнів.

3.4. Аналіз поз учнів та розрахунок індексів уваги

3.4.1. Обчислення ознак поз з ключових точок

Ключові точки, які повертає модель, перетворюються на набір ознак, пов'язаних з увагою:

- `head_pitch` – орієнтовний нахил голови вгору/вниз;
- `head_yaw` – поворот голови вліво/вправо;
- `eye_openness` – груба оцінка відкритості очей;
- `hands_up` – бінарна ознака піднятої руки;
- `hands_below` – руки опущені нижче рівня стегон;
- `slouch_factor` – показник сутулості.

Обчислення ознак винесено в окрему функцію:

Лістинг 6: Функція обчислення ознак поз з ключових точок

```
def compute_pose_features(kpts):
    """
    kpts: numpy array of shape (17, 2) with (x, y) coordinates
          of COCO keypoints.
    Returns a dict of normalized pose features.
    """
    # Indices for keypoints in COCO format
    NOSE = 0
    L_EYE, R_EYE = 1, 2
    L_SHOULDER, R_SHOULDER = 5, 6
    L_WRIST, R_WRIST = 9, 10
    L_HIP, R_HIP = 11, 12
```

```
13
14      nose = kpts[NOSE]
15      eyes_mid = (kpts[L_EYE] + kpts[R_EYE]) / 2.0
16      shoulders_mid = (kpts[L_SHOULDER] + kpts[R_SHOULDER]) / 2.0
17      hips_mid = (kpts[L_HIP] + kpts[R_HIP]) / 2.0
18
19      torso_len = np.linalg.norm(shoulders_mid - hips_mid) + 1e-6
20
21      head_pitch = (nose[1] - shoulders_mid[1]) / torso_len
22      head_yaw = (nose[0] - eyes_mid[0]) / torso_len
23      eye_openness = abs(eyes_mid[1] - nose[1]) / torso_len
24
25      left_wrist = kpts[L_WRIST]
26      right_wrist = kpts[R_WRIST]
27
28      hands_up = int(
29          (left_wrist[1] < shoulders_mid[1] - 0.2 * torso_len)
30          or (right_wrist[1] < shoulders_mid[1] - 0.2 * torso_len
                )
31      )
32
33      hands_below = int(
34          (left_wrist[1] > hips_mid[1] + 0.1 * torso_len)
35          and (right_wrist[1] > hips_mid[1] + 0.1 * torso_len)
36      )
37
38      slouch_factor = abs(shoulders_mid[1] - nose[1]) / torso_len
39
40      return {
41          "head_pitch": float(head_pitch),
42          "head_yaw": float(head_yaw),
43          "eye_openness": float(eye_openness),
44          "hands_up": hands_up,
45          "hands_below": hands_below,
46          "slouch_factor": float(slouch_factor),
47      }
```

Перші кілька секунд (у коді використовується параметр BASELINE_FRAMES, наприклад 90 кадрів) відводяться на **калібрування базових значень**: для кожного учня накопичуються середні значення ознак у його “нормальному” робочому стані. Надалі всі відхилення

інтерпретуються відносно цих базових значень.

3.4.2. Розрахунок індексу уваги окремого учня

На основі ознак обчислюється числовий індекс уваги `attention_score` у діапазоні $[0, 1]$. У коді це реалізовано як зважена комбінація ознак із подальшою нормалізацією сигмоїдою:

Лістинг 7: Розрахунок індексу уваги учня

```
def student_attention_score(features, baseline):
    """
    features: current pose features for the student
    baseline: per-student baseline values for the same features
    """
    pitch_dev = features["head_pitch"] - baseline["head_pitch"]
    yaw_dev = abs(features["head_yaw"] - baseline["head_yaw"])
    eye_ratio = features["eye_openness"] / (baseline["
       eye_openness"] + 1e-6)
    slouch_dev = features["slouch_factor"] - baseline["
       slouch_factor"]

    score = 0.0

    # Strong positive signal: raised hand
    if features["hands_up"]:
        score += 2.0

    # Negative signal: both hands below hips
    if features["hands_below"]:
        score -= 0.8

    # Head strongly tilted down
    if pitch_dev > 0.15:
        score -= 1.5
    else:
        # Head at board level or slightly up
        score += 1.2

    # Head strongly turned sideways
    score -= min(1.0, yaw_dev * 3.0)

```

```
    # Very low eye openness -> possible drowsiness
    if eye_ratio < 0.6:
        score -= 1.5

    # Slouching more than usual
    if slouch_dev > 0.15:
        score -= 0.7

    # Sigmoid to map to [0, 1]
    return 1.0 / (1.0 + np.exp(-score))
```

Щоб згладити випадкові коливання між кадрами, значення індексу уваги додатково проходить через експоненційне ковзне середнє:

Лістинг 8: Експоненційне згладжування індексу уваги

```
EMA_ALPHA = 0.3

def update_ema(old_ema, new_value):
    """Update exponential moving average for attention score."""
    if old_ema is None:
        return new_value
    return EMA_ALPHA * new_value + (1.0 - EMA_ALPHA) * old_ema
```

Отримане згладжене значення `score_ema` використовується як основа для класифікації стану уваги учня.

3.5. Класифікація станів уваги та індекс уваги класу

У реалізації використовується фіксований набір порогів:

- $E \geq 0.70$ – **Attentive**;
- $0.50 \leq E < 0.70$ – **Neutral**;
- $0.30 \leq E < 0.50$ – **Distracted**;
- $E < 0.30$ – **Inattentive**.

Після підрахунку кількості учнів у кожній категорії обчислюється **Class Attention Index (CAI)**:

Лістинг 9: Обчислення індексу уваги класу (CAI)

```
def compute_cai(stats):
    """
    stats: dict with counts for each state, e.g.
           {"attentive": ..., "neutral": ..., "distracted":
              ...,
            "inattentive": ..., "hands_up": ..., "total": ...}
    """
    if stats["total"] == 0:
        return 0.0

    num = (
        stats["attentive"]
        + 0.5 * stats["neutral"]
        + 0.8 * stats["hands_up"]
        - 0.5 * stats["inattentive"]
    )
    cai = num / stats["total"]
    return max(0.0, min(1.0, cai))
```

CAI виводиться на кадрі як текст `Class Attention Index: 0.76` із кольоровим маркером (зелений/жовтий/червоний залежно від значення), а також зберігається в журналі для подальшого аналізу.

3.6. Технічна реалізація та використання в Google Colab

3.6.1. Локальний запуск на персональному комп'ютері

Локальна версія системи запускається як звичайний Python-скрипт. Для цього достатньо встановити необхідні бібліотеки:

Лістинг 10: Встановлення залежностей для локального запуску

```
pip install ultralytics opencv-python numpy matplotlib
```

Далі можна викликати скрипт, вказавши вхідне та вихідне відео:

Лістинг 11: Приклад запуску скрипту на відеофайлі

```
python yolov8_custom.py --video input_lesson.mp4 --output
   output_annotated.mp4
```

Усередині `yolov8_custom.py` аргументи командного рядка розбираються через модуль `argparse`, відкривається `cv2.VideoCapture`, запускається головний цикл обробки кадрів та запис результату у `cv2.VideoWriter`. Параметр `SPEED_MULTIPLIER` дозволяє пропускати частину кадрів (наприклад, аналізувати кожен другий), що прискорює обробку довгих відео.

3.6.2. Хмарна версія в Google Colab

Для користувачів без потужного GPU реалізовано робочий блокнот `projectCompVision.ipynb`. Його сценарій використання:

1. **Відкрити блокнот** у Google Colab та увімкнути апаратний прискорювач `GPU` у налаштуваннях середовища.
2. **Підключити Google Drive** та завантажити відео уроку до нього (або завантажити файл напряму в Colab).
3. **Виконати комірку з інсталяцією залежностей** (`ultralytics`, `opencv-python`, `numpy` тощо).
4. **Задати шлях до відео** у змінній `VIDEO_PATH` та запустити основну комірку, яка повторює логіку `yolov8_custom.py`: детекція, відстеження, аналіз поз, візуалізація.
5. **Переглянути результат** прямо в блокноті або завантажити оброблене відео `output.mp4`.

В Colab неможливо використовувати `cv2.imshow` для кожного кадру, тому результат зберігається у відеофайл і відтворюється через HTML-віджет:

Лістинг 12: Перегляд результативного відео в Google Colab

```
from IPython.display import HTML
from base64 import b64encode

mp4 = open("output.mp4", "rb").read()
data_url = "data:video/mp4;base64," + b64encode(mp4).decode()

HTML(f"""
```

```
<video width="800" controls>
    <source src="{data_url}" type="video/mp4">
</video>
""")
```

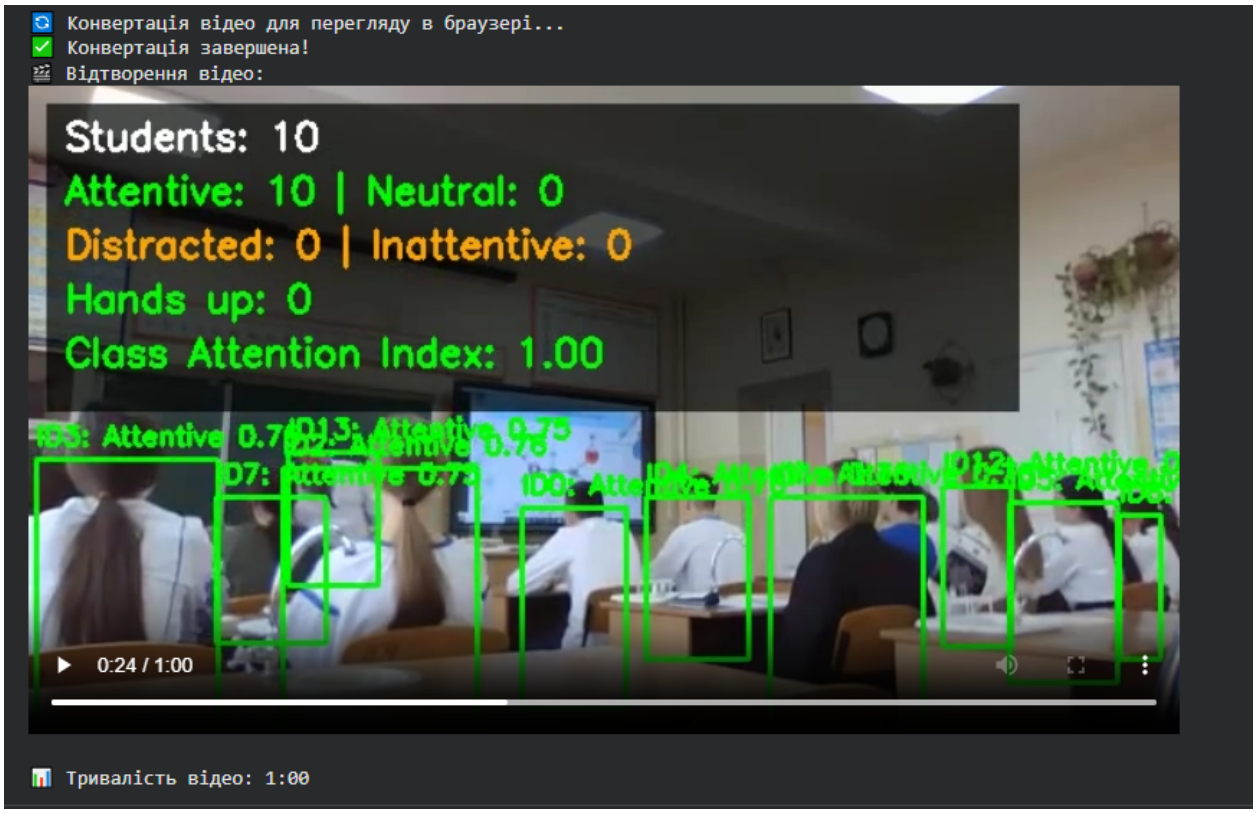

Рис. 3.2. Перегляд обробленого відео у середовищі Google Colab.

Таким чином, навіть на слабкому комп'ютері можна повністю випробувати розроблену систему: всі обчислення виконуються на віддаленому GPU-сервері.

3.7. Експерименти

Для експериментальної перевірки розробленого підходу було проаналізовано відеозапис уроку історії тривалістю 30 хвилин. Обробка здійснювалася в середовищі Google Colab із використанням моделі комп'ютерного зору YOLOv8m-pose; для зменшення обсягу даних аналізувався кожен третій кадр (ефективна частота 10 кадрів/с). Всього опрацьовано близько 18 тис. кадрів. Камера була встановлена фронтально у класній кімнаті, тому одночасно в полі зору перебувало приблизно 20 учнів включно з вчителем. Кількість виявлених системою осіб коливалася

в межах від 12 до 26 на різних кадрах. У середньому 18 учнів. Тимчасове зниження цього показника пояснюється ситуаціями, коли учні вставали і частково закривали один одного або коли алгоритму доводилося повторно калібрувати траєкторії (повторна ідентифікація) після втрати об'єкта. Активна участь учнів також відстежувалася: в середньому на кожному кадрі виявлено близько 2 піднятих рук (максимум – 10 одночасно), хоча у разі тривалого тримання руки одним учнем система могла зарахувати це як кілька окремих подій, що дещо завищує сумарний підрахунок.

За результатами автоматичної класифікації станів уваги було визначено, що в середньому на кожному кадрі приблизно 14 учнів перебували у стані «уважний» (індекс уваги $\geq$0.70). Близько 4 учнів при цьому мали проміжний нейтральний стан. Кількість тих, хто відволікався (стан Distracted) або був повністю неуважним (Inattentive), виявилася невеликою – зазвичай не більше 1 особи одночасно (в середньому менше одного учня на кадр для кожної з цих категорій). Середнє значення інтегрального індексу уваги класу за весь урок становило понад 0.91 (за шкалою 0–1), причому у найбільш активні моменти цей показник досягав максимальних 1.0. На початку заняття значення індексу були нижчими, а наприкінці уроку спостерігалося певне зниження до рівня $\sim$0.8. На рис. 3.3 наведено часові графіки динаміки ключових показників активності протягом уроку, зокрема зміни кількості учнів у кадрі, кількості піднятих рук, середнього індексу уваги класу та розподілу учнів за станами уваги.

Висновки до розділу 3

У цьому розділі описано практичну реалізацію системи моніторингу навчальної активності учнів ліцею. Основний акцент зроблено на організації програмного коду та методиці використання готової моделі `YOLOv8m-pose` для аналізу поз без додаткового навчання.

Було показано:

- як організовано конвеєр обробки відео у файлі `yolov8_custom.py`;
- як налаштовується й викликається модель `YOLOv8m-pose` для отримання рамок та ключових точок учнів;

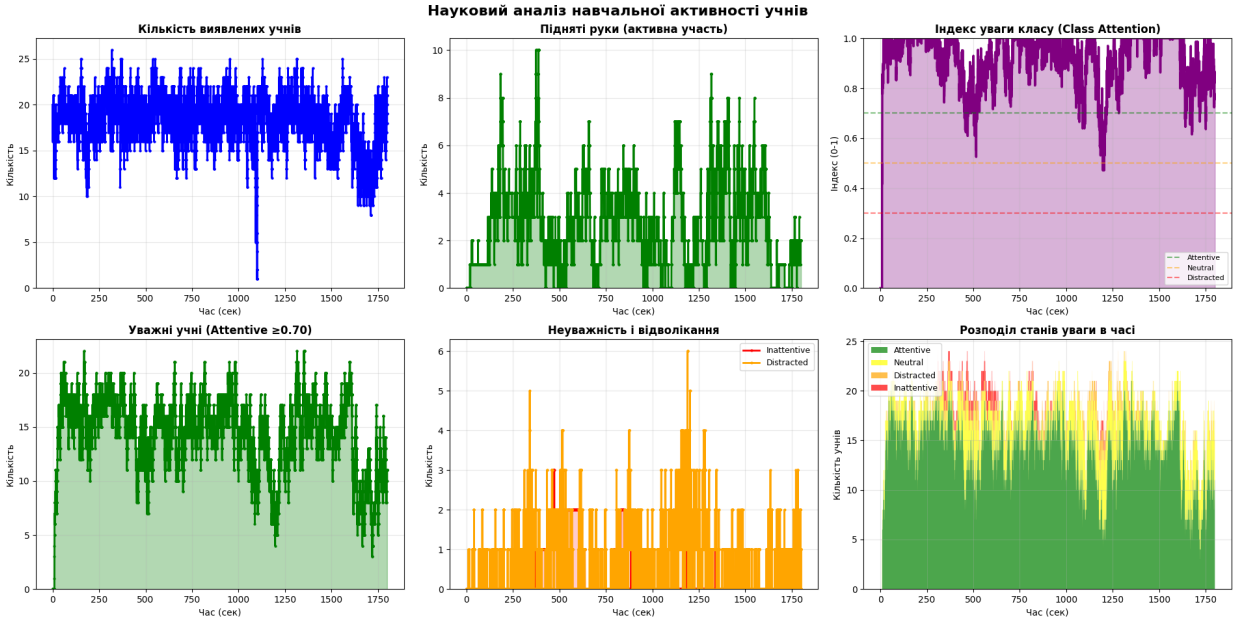

Рис. 3.3. Аналіз навчальної активності учнів.

- як реалізовано просте відстеження учнів між кадрами на основі метрики IoU;
- як з ключових точок обчислюються ознаки пози, індивідуальний індекс уваги та агрегований **Class Attention Index**;
- як система може бути запущена локально та в Google Colab, що робить її доступною навіть без потужного апаратного забезпечення.

Реалізована система дозволяє перетворити звичайний відеозапис уроку на набір кількісних метрик уваги як окремих учнів, так і класу в цілому. Це створює основу для подальшої експериментальної оцінки ефективності, порівняння різних уроків та розвитку концепції "розумної" класної кімнати.

ВИСНОВКИ

1. Проаналізовано теоретичні основи застосування комп'ютерного зору в освіті та існуючі підходи до автоматичної оцінки залученості учнів. Узагальнено сучасні підходи до аналізу поз, погляду, міміки та жестів учнів під час навчання, сформовано вимоги до системи моніторингу уваги для умов ліцейської освіти.

2. Проведено бібліометричний аналіз наукових публікацій у сфері використання комп'ютерного зору для навчання та моніторингу. Аналіз публікацій, індексованих у міжнародних наукометричних базах, дав змогу:

   а) визначити період стійкого зростання кількості досліджень, пов'язаних із моніторингом навчальної активності та оцінкою уваги учнів;

   б) виокремити основні тематичні напрями, серед яких: загальні методи комп'ютерного зору, аналіз поз і поведінки людини, виявлення емоцій і залученості, застосування комп'ютерного зору у формальній освіті;

   в) обґрунтувати актуальність вибору задачі моніторингу уваги учнів на уроках як перспективного напрямку подальших розробок.

3. Розроблено алгоритми виявлення та трекінгу учнів на відео, а також обчислення індивідуальних індексів уваги на основі поз:

   а) побудовано модуль детекції поз учнів на основі `YOLOv8m-pose`, який для кожного кадру визначає положення тіла та ключових точок;

   б) реалізовано трекінг учнів між кадрами за допомогою метрики Intersection over Union (IoT) з фільтрацією короткочасних втрат і повторною ініціалізацією ідентифікаторів;

   в) сформовано набір ознак, що характеризують позу учня (нахили голови, положення рук, сутулість, відкритість очей тощо), та

розроблено формулу обчислення індивідуального індексу уваги в діапазоні $[0; 1]$.

4. Розроблено індекс уваги класу *Class Attention Index* (CAI) та засоби візуалізації результатів:

   а) визначено правила віднесення учня до однієї з чотирьох категорій стану уваги (*Attentive*, *Neutral*, *Distracted*, *Inattentive*) на основі індивідуального індексу та його часової стабільності;

   б) запропоновано формулу CAI, що агрегує індивідуальні показники уваги та частки учнів у різних станах і дає змогу кількісно оцінювати загальний рівень уваги класу;

   в) реалізовано накладання результатів на відео у вигляді кольорових рамок, скелетів поз та інформаційної панелі з основними показниками, а також побудову узагальнених графіків динаміки уваги в часі.

5. Реалізовано програмний прототип системи моніторингу навчальної активності із використанням моделей комп'ютерного зору. Прототип створено на мові Python із використанням середовища Google Colab, бібліотек OpenCV та фреймворку Ultralytics. Забезпечено автоматичне завантаження відео, обробку кожного N-го кадру, збереження детальних статистичних даних у форматі CSV та побудову підсумкових візуалізацій.

6. Проведено експериментальну перевірку працездатності системи на реальному відеозаписі уроку історії тривалістю 30 хвилин:

   а) оброблено близько 18 тис. кадрів (кожен третій кадр відео), у середньому в кадрі одночасно перебувало близько 20 осіб, а кількість виявлених учнів коливалася в межах 12–26 через взаємні перекривання, появу нових учнів у полі зору та повторну ініціалізацію треків;

   б) встановлено, що протягом більшої частини уроку більшість учнів перебувала в стані *Attentive*, індекс уваги класу CAI пе-

реважно перевищував 0,8, а епізоди суттєвого зниження уваги мали короткочасний характер;

в) проаналізовано типові обмеження системи, зокрема вплив оклюзій, особливостей ракурсу камери та можливі хибні спрацьовування окремих ознак (наприклад, тривале утримання руки вгорі).

7. На основі експериментальних результатів оцінено можливості практичного впровадження запропонованої методики в ліцейській освіті. Показано, що система здатна в напівавтоматичному режимі надавати вчителю та адміністрації об'єктивну інформацію про динаміку уваги класу, виявляти моменти зниження залученості та потенційно слугувати інструментом підтримки прийняття педагогічних рішень.

Отримані результати можуть бути використані для подальшого розвитку систем моніторингу навчальної активності в ліцеях, зокрема для масштабування на більшу кількість класів, інтеграції з електронними журналами, поєднання з іншими джерелами даних (аудіо, результати контрольних робіт) та побудови адаптивних освітніх аналітичних панелей для вчителів і адміністрації навчальних закладів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. 2D Human pose estimation: a survey / H. Chen [та ін.] // Multimedia Systems. — 2023. — Т. 29. — С. 3115—3138. — DOI: 10.1007/s00530-022-01019-0.

2. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis / M. Andriluka [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — IEEE, 2014. — С. 3686—3693. — DOI: 10.1109/CVPR.2014.471.

3. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities / R. Boutaba [та ін.] // Journal of Internet Services and Applications. — 2018. — Т. 9. — С. 1—99. — DOI: 10.1186/s13174-018-0087-2.

4. A review on recent developments in cancer detection using machine learning and deep learning models / S. Maurya [та ін.] // Biomedical Signal Processing and Control. — 2023. — Т. 80. — С. 104398. — DOI: 10.1016/j.bspc.2022.104398.

5. A survey of modern deep learning based object detection models / S. S. A. Zaidi [та ін.] // Digital Signal Processing. — 2022. — Т. 126. — С. 103514. — DOI: 10.1016/j.dsp.2022.103514.

6. A Survey on Vision Transformer / K. Han [та ін.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2023. — Т. 45, № 1. — С. 87—110. — DOI: 10.1109/TPAMI.2022.3152247.

7. Advanced Methods and Deep Learning in Computer Vision / за ред. E. R. Davies, M. A. Turk. — Academic Press, 2022. — URL: https://www.sciencedirect.com/book/edited-volume/9780128221099/advanced-methods-and-deep-learning-in-computer-vision.

8. *Ahmed I.*, *Chehri A.*, *Jeon G.* A Sustainable Deep Learning-Based Framework for Automated Segmentation of COVID-19 Infected Regions: Using U-Net with an Attention Mechanism and Boundary Loss Function // Electronics. — 2022. — Т. 11, № 15. — С. 2296. — ISSN 2079-9292. — DOI: 10.3390/electronics11152296.

9. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale / A. Dosovitskiy [та ін.] // 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. — OpenReview.net, 2021. — URL: https://openreview.net/forum?id=YicbFdNTTy.

10. Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression / O. M. Nezami [та ін.] // Machine Learning and Knowledge Discovery in Databases. Т. 11908. — Springer, 2020. — С. 273—289. — (Lecture Notes in Computer Science). — DOI: 10.1007/978-3-030-46133-1_17.

11. *Canedo D.*, *Trifan A.*, *Neves A. J. R.* Monitoring Students' Attention in a Classroom Through Computer Vision // Biomedical Engineering Systems and Technologies. Т. 1024. — Springer, 2019. — С. 367—383. — (Communications in Computer and Information Science). — DOI: 10.1007/978-3-319-94779-2_32.

12. *Cernadas E.* Applications of Computer Vision, 2nd Edition // Electronics. — 2024. — Т. 13, № 18. — С. 3779. — DOI: 10.3390/electronics13183779.

13. *Chen Y.*, *Tian Y.*, *He M.* Monocular human pose estimation: A survey of deep learning-based methods // Computer Vision and Image Understanding. — 2020. — Т. 192. — С. 102897. — DOI: 10.1016/j.cviu.2019.102897.

14. Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge / A. Qayyum [та ін.] // IEEE Open Journal of the Computer Society. — 2022. — Т. 3. — С. 172—184. — ISSN 2644-1268. — DOI: 10.1109/OJCS.2022.3206407.

15. *DataXujing Community.* YOLOv8 Loss Functions. — 2024. — Explains VFL, DFL and CIOU losses used in YOLOv8. https://deepwiki.com/DataXujing/YOLOv8/2.1-loss-functions.

16. Deep High-Resolution Representation Learning for Human Pose Estimation / K. Sun [та ін.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2019. — С. 5693—5703. — DOI: 10.1109/CVPR.2019.00584.

17. Deep learning for computer vision: A brief review / A. Voulodimos [та ін.] // Computational Intelligence and Neuroscience. — 2018. — Т. 2018. — DOI: 10.1155/2018/7068349.

18. Deep learning for edge computing applications: A state-of-the-art survey / X. Wang [та ін.] // IEEE Access. — 2022. — Т. 8. — С. 58322—58336. — ISSN 2169-3536. — DOI: 10.1109/ACCESS.2020.2982411.

19. Deep learning for generic object detection: A survey / L. Liu [та ін.] // International Journal of Computer Vision. — 2020. — Т. 128, № 2. — С. 261—318. — DOI: 10.1007/s11263-019-01247-4.

20. Deep learning for visual understanding: A review / Y. Guo [та ін.] // Neurocomputing. — 2016. — Т. 187. — С. 27—48. — DOI: 10.1016/j.neucom.2015.09.116.

21. Deep learning in computer vision: A critical review of emerging techniques and application scenarios / J. Chai [та ін.] // Machine Learning with Applications. — 2021. — Т. 6. — С. 100134. — ISSN 2666-8270. — DOI: 10.1016/j.mlwa.2021.100134.

22. *Dhillon A.*, *Verma G. K.* Convolutional neural network: a review of models, methodologies and applications to object detection // Progress in Artificial Intelligence. — 2020. — Т. 9. — С. 85—112. — DOI: 10.1007/s13748-019-00203-0.

23. *Diwan T.*, *Anirudh G.*, *Tembhurne J. V.* Object detection using YOLO: challenges, architectural successors, datasets and applications // Multimedia Tools and Applications. — 2023. — Т. 82, № 6. — С. 9243—9275. — DOI: 10.1007/s11042-022-13644-y.

24. Do vision transformers see like convolutional neural networks? / M. Raghu [та ін.] // Proceedings of the 35th International Conference on Neural Information Processing Systems. — Red Hook, NY, USA : Curran Associates Inc., 2021. — С. 927. — (NIPS '21). — ISBN 9781713845393.

25. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / S. Ren [та ін.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2017. — Т. 39, № 6. — С. 1137—1149. — DOI: 10.1109/TPAMI.2016.2577031.

26. How to conduct a bibliometric analysis: An overview and guidelines / N. Donthu [та ін.] // Journal of Business Research. — 2021. — Т. 133. — С. 285—296. — DOI: 10.1016/j.jbusres.2021.04.070.

27. HR-YOLOv8: A Crop Growth Status Object Detection Method Based on YOLOv8 / J. Zhang [та ін.] // Electronics. — 2024. — Т. 13, № 9. — С. 1620. — DOI: 10.3390/electronics13091620.

28. Image segmentation using deep learning: A survey / S. Minaee [та ін.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2020. — Т. 44, № 7. — С. 3523—3542. — DOI: 10.48550/arXiv.2001.05566.

29. Know your self-supervised learning: A survey on image-based generative and discriminative training / U. Ozbulak [та ін.] // Transactions on Machine Learning Research. — 2023. — DOI: 10.48550/arXiv.2305.13689.

30. *Kukil*. Mean Average Precision (mAP) in Object Detection. — 2022. — URL: https://learnopencv.com/mean-average-precision-map-object-detection-model-evaluation-metric/.

31. *LeCun Y.*, *Bengio Y.*, *Hinton G.* Deep learning // Nature. — 2015. — Т. 521, № 7553. — С. 436—444. — DOI: 10.1038/nature14539.

32. *Masita K. L.*, *Hasan A. N.*, *Shongwe T.* Deep learning in object detection: A review // 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems, icABCD 2020 - Proceedings. — 2020. — DOI: 10.1109/icABCD49160.2020.9183866.

33. Microsoft COCO: Common Objects in Context / T. Lin [та ін.] // Computer Vision – ECCV 2014. Т. 8693. — Springer, 2014. — С. 740—755. — (Lecture Notes in Computer Science). — ISBN 978-3-319-10602-1. — DOI: 10.1007/978-3-319-10602-1_48.

34. Mixed Precision Training / P. Micikevicius [та ін.] // 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. — OpenReview.net, 2018. — URL: https://openreview.net/forum?id=r1gs9JgRZ.

35. *Mohammed A., Kora R.* A comprehensive review on ensemble deep learning: Opportunities and challenges // Journal of King Saud University-Computer and Information Sciences. — 2023. — Т. 35, № 2. — С. 757—774. — DOI: 10.1016/j.jksuci.2023.01.014.

36. *Mongeon P., Paul-Hus A.* The journal coverage of Web of Science and Scopus: a comparative analysis // Scientometrics. — 2016. — Т. 106, № 1. — С. 213—228. — DOI: 10.1007/s11192-015-1765-5.

37. *Mpouziotas D., Karvelis P., Stylios C.* Advanced Computer Vision Methods for Tracking Wild Birds from Drone Footage // Drones. — 2024. — Т. 8, № 6. — С. 259. — DOI: 10.3390/drones8060259.

38. Object detection in 20 years: A survey / Z. Zou [та ін.] // Proceedings of the IEEE. — 2023. — Т. 111, № 3. — С. 257—276. — ISSN 1558-2256. — DOI: 10.1109/JPROC.2023.3238524.

39. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields / Z. Cao [та ін.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2021. — С. 172—186. — ISSN 1939-3539. — DOI: 10.1109/TPAMI.2019.2929257.

40. *Perianes-Rodriguez A., Waltman L., Van Eck N. J.* Constructing bibliometric networks: A comparison between full and fractional counting // Journal of Informetrics. — 2016. — Т. 10, № 4. — С. 1178—1195. — DOI: 10.1016/j.joi.2016.10.006.

41. *Rawat W., Wang Z.* Deep convolutional neural networks for image classification: A comprehensive review // Neural Computation. — 2017. — Т. 29, № 9. — С. 2352—2449. — DOI: 10.1162/NECO_a_00990.

42. *Rezaeilouyeh H., Mollahosseini A., Mahoor M. H.* Microscopic medical image classification framework via deep learning and shearlet transform // Journal of Medical Imaging. — 2016. — Т. 3, № 4. — С. 12. — DOI: 10.1117/1.JMI.3.4.044501.

43. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies / J. Baas [та ін.] // Quantitative Science Studies. — 2020. — Т. 1, № 1. — С. 377—386. — DOI: 10.1162/qss_a_00019.

44. *Shinde P. P., Shah S.* A Review of Machine Learning and Deep Learning Applications // 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). — 2018. — С. 1—6. — DOI: 10.1109/ICCUBEA.2018.8697857.

45. *Soukupová T., Čech J.* Real-Time Eye Blink Detection using Facial Landmarks // Proceedings of the 21st Computer Vision Winter Workshop / за ред. L. Čehovin, R. Mandeljc, V. Štruc. — Rimske Toplice, Slovenia, February 3–5, 2016, 2016. — URL: https://vision.fe.uni-lj.si/cvww2016/proceedings/papers/05.pdf.

46. SSD: Single Shot MultiBox Detector / W. Liu [та ін.] // Proceedings of the European Conference on Computer Vision (ECCV). — Cham : Springer, 2016. — С. 21—37. — DOI: 10.1007/978-3-319-46448-0_2.

47. *Szeliski R.* Computer Vision: Algorithms and Applications. — 2-е вид. — Cham : Springer, 2022. — DOI: 10.1007/978-3-030-34372-9.

48. *Tayeb M.* Understanding Mixed Precision Training. — 2019. — Blog post on Towards Data Science. https://towardsdatascience.com/understanding-mixed-precision-training-4b246679c7c4.

49. The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions / J. Whitehill [та ін.] // IEEE Transactions on Affective Computing. — 2014. — Т. 5, № 1. — С. 86—98. — DOI: 10.1109/TAFFC.2014.2316163.

50. The Pascal Visual Object Classes (VOC) Challenge / M. Everingham [та ін.] // International Journal of Computer Vision. — 2010. — Т. 88, № 2. — С. 303—338. — DOI: 10.1007/s11263-009-0275-4.

51. Transformers in medical imaging: A survey / F. Shamshad [та ін.] // Medical Image Analysis. — 2023. — Т. 88. — С. 102802. — ISSN 1361-8415. — DOI: 10.1016/j.media.2023.102802.

52. Transformers in Vision: A Survey / S. Khan [та ін.] // ACM Computing Surveys. — New York, NY, USA, 2022. — Вер. — Т. 54, 10s. — С. 200. — ISSN 0360-0300. — DOI: 10.1145/3505244.

53. *Ultralytics*. Ultralytics YOLOv8 Models Documentation. — 2023. — Official documentation of YOLOv8 model family. https://docs.ultralytics.com/models/yolov8/.

54. *Ultralytics*. Configuration and Training Options in Ultralytics YOLO. — 2024. — Describes AMP (mixed precision) and other training hyperparameters. https://docs.ultralytics.com/usage/cfg/.

55. *Ultralytics*. Mixed Precision Training: Speed Up Deep Learning. — 2024. — Overview of mixed precision training in Ultralytics YOLO. https://www.ultralytics.com/glossary/mixed-precision.

56. *Van Eck N. J.*, *Waltman L.* Software survey: VOSviewer, a computer program for bibliometric mapping // Scientometrics. — 2010. — Т. 84, № 2. — С. 523—538. — DOI: 10.1007/s11192-009-0146-3.

57. *Xiao B.*, *Wu H.*, *Wei Y.* Simple Baselines for Human Pose Estimation and Tracking // Computer Vision – ECCV 2018. Т. 11210. — Springer, 2018. — С. 472—487. — (Lecture Notes in Computer Science). — DOI: 10.1007/978-3-030-01231-1_29.

58. *Yan L.*, *Wu X.*, *Wang Y.* Student engagement assessment using multimodal deep learning // PLOS ONE. — 2025. — Т. 20, № 6. — DOI: 10.1371/journal.pone.0325377.

59. *Yaseen M.* What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector // arXiv preprint arXiv:2408.15857. — 2024. — DOI: 10.48550/arXiv.2408.15857.

60. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon [та ін.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — С. 779—788. — DOI: 10.1109/CVPR.2016.91.