

DOI: 10.20998/2079-0023.2025.01.16
УДК 004.8+004.9

О. М. КОНДРАТОВ, аспірант, старший викладач кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; старший викладач кафедри цифрових технологій та проєктно-аналітичних рішень ТОВ «ТУ «МЕТІНВЕСТ ПОЛІТЕХНІКА»», Запоріжжя, Україна; ORCID: <https://orcid.org/0000-0001-6367-9944>; e-mail: kondratovolexy@gmail.com

О. М. НІКУЛІНА, д-р техн. наук, професор, завідувачка кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; професор кафедри цифрових технологій та проєктно-аналітичних рішень ТОВ «ТУ «МЕТІНВЕСТ ПОЛІТЕХНІКА»», Запоріжжя, Україна; ORCID: <https://orcid.org/0000-0003-2938-4215>; e-mail: elniknik02@gmail.com

ІДЕНТИФІКАЦІЯ ПАРАМЕТРІВ ДИНАМІЧНИХ ОБ'ЄКТІВ З ВИКОРИСТАННЯМ ТРАНСФОРМЕРА З ОПТИЧНИМ ПОТОКОМ ТА АНСАМБЛЕВИХ МЕТОДІВ

У статті розглянуто підхід до ідентифікації параметрів динамічних об'єктів у відеопотоці з використанням трансформерної архітектури, моделі GeoNet та ансамблевих методів машинного навчання, зокрема бегінгу та бустінгу. Ідентифікація параметрів таких об'єктів, як положення, швидкість, напрям руху та глибина, має важливе значення для широкого спектра застосувань, включаючи автономне водіння, робототехніку та системи відеоспостереження. У роботі описано комплексну систему, яка забезпечує інтеграцію просторово-часових характеристик відеопотоку через обчислення оптичного потоку та карти глибини за допомогою GeoNet, їх подальший аналіз із застосуванням трансформера, а також підвищення точності завдяки ансамблюванню результатів. GeoNet, як глибока згортоква нейронна мережа, об'єднує завдання оцінки глибини та оптичного потоку в єдину архітектуру, що дозволяє точно реконструювати тривимірну сцену. Використання трансформера дозволяє моделювати глобальні залежності в кадрах відео та покращити точність класифікації та виявлення об'єктів. Водночас, бегінг зменшує дисперсію шляхом усереднення результатів кількох моделей, навчених на різних підвбірках, а бустінг дозволяє фокусуватися на складних прикладах для підвищення точності прогнозу. Запропонована система забезпечує високу точність в умовах динамічного фону, зміни освітлення, оклюзії та шумів, завдяки чому може бути адаптована для використання в реальному часі в складних сценах. Наведено детальний опис кожного з компонентів системи: архітектури GeoNet, модулів трансформера, реалізації бегінгу та бустінгу, а також алгоритму об'єднання результатів. Очікувані результати мають демонструвати ефективність інтеграції методів глибокого навчання з класичними ансамблевими підходами для задач високоточної ідентифікації динамічних об'єктів. Запропонована методологія відкриває перспективи для створення інтелектуальних систем комп'ютерного зору нового покоління.

Ключові слова: Дистанційна ідентифікація динамічних об'єктів, виявлення об'єктів, комп'ютерний зір, ансамблеві методи, глибоке навчання, згорткові нейронні мережі, архітектура системи, програмне забезпечення, машинне навчання, штучний інтелект.

Вступ. Ідентифікація параметрів динамічних об'єктів у відеопотоці є важливим завданням у сфері комп'ютерного зору, що знаходить застосування в автономному водінні, відеоспостереженні, робототехніці. Одним із сучасних підходів є використання нейронних мереж для передбачення глибини сцени (depth estimation) та оцінки руху (optical flow). Однією з таких моделей є GeoNet, яка об'єднує ці два завдання в єдину архітектуру.

Завдяки одночасному вивченню геометрії сцени та руху камери, GeoNet демонструє підвищену точність у порівнянні з традиційними підходами, що розділяють ці завдання. У моделі використовується глибока згорткова нейронна мережа (CNN), яка дозволяє автоматично вивчати релевантні просторові ознаки для побудови карти глибини та оцінки векторів оптичного потоку.

Візуалізація роботи таких моделей, як GeoNet, може бути значно полегшена за допомогою спеціалізованих інструментів, таких як CNN Explainer інтерактивна система, що пояснює принцип роботи згорткових нейронних мереж через покрокову візуалізацію їх внутрішніх механізмів. Вона дозволяє краще зрозуміти, як мережа витягує ознаки з вхідних зображень та як ці ознаки сприяють прийняттю рішень. Крім того, в основі сучасних архітектур, включно з тими, що використовуються для обробки відео, все частіше лежить

механізм уваги (attention), представлений у Transformer-моделі, що дозволяє ефективно моделювати залежності між пікселями з урахуванням їх просторово-часових зв'язків. Це розширює можливості традиційних CNN, забезпечуючи кращу контекстну інтерпретацію динамічних сцен [1–2].

Однак такі моделі можуть страждати від локальних помилок через шум, зміну освітлення, оклюзії. Для покращення точності пропонується використовувати ансамблеві методи машинного навчання бегінг (bagging) та бустінг (boosting). Ці підходи дозволяють підвищити якість прогнозування, зменшуючи дисперсію та зміщення. Бегінг дозволяє зменшити дисперсію моделі за рахунок поєднання кількох незалежно навчених моделей, кожна з яких працює з випадковою вибіркою даних, тим самим згладжуючи вплив шумових артефактів. Натомість бустінг послідовно навчає серію слабких моделей, фокусуючись на тих прикладах, які були важкими для попередніх, що дозволяє знизити зміщення і покращити загальну точність передбачень.

Важливість точного виявлення та ідентифікації об'єктів у складних сценах, зокрема в умовах динаміки, підкреслюється в сучасних оглядах та роботах з обробки зображень і відео. Наприклад, у Transformer-архітектурі DETR (Detection Transformer) [3], яка поєднує переваги механізмів уваги та енд-ту-енд підходу до виявлення об'єктів, увага приділяється глобальному

© Кондратов О. М., Нікуліна О. М., 2025



Дослідницька стаття: Цю статтю опубліковано видавництвом **НТУ «ХПІ»** у збірнику «Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



контексту сцени, що підвищує стійкість до часткових спотворень і перешкод. Огляд методів обробки зображень за 20 років [4] підтверджує, що інтеграція класичних підходів з новітніми методами глибокого навчання та ансамблювання забезпечує найкращу продуктивність у складних задачах комп'ютерного зору, включно з ідентифікацією об'єктів у відеопотоці.

Ідентифікація параметрів динамічних об'єктів з використанням трансформера з оптичним потоком та ансамблевих методів є сучасним напрямом у галузі комп'ютерного зору, що дозволяє досягти високої точності в оцінці ключових характеристик об'єктів, таких як положення, швидкість, напрямок руху та глибина. У цій роботі пропонується комплексна система, яка поєднує можливості моделі GeoNet для оцінки глибини сцени та оптичного потоку, трансформерної архітектури для виявлення об'єктів та класифікації, а також методів ансамблювання для підвищення надійності та точності прогнозів. Такий підхід забезпечує більш повну картину динамічної сцени, дозволяючи одночасно виявляти об'єкти та оцінювати їхні кінематичні параметри у реальному часі.

Розроблена система включає попереднє завантаження відеопотоку, покадрову обробку із застосуванням CNN-базованих методів для обчислення оптичного потоку [5], подальше прогнозування глибини та руху за допомогою GeoNet [6], і виявлення об'єктів з використанням моделей на основі трансформерів або регіональних згорткових мереж [7]. Для покращення точності ідентифікації використовуються методи бегінгу та бустінгу, які дозволяють об'єднати результати декількох слабких моделей в єдине узгоджене рішення. Система розроблена для різних умов, зокрема на складних сценах із рухомим фоном та перешкодами, де вона має продемонструвати високу ефективність [8–9].

Мета та задачі дослідження. Мета статті є розробка та вдосконалення комплексної системи для ідентифікації параметрів динамічних об'єктів у відеопотоці, що поєднує можливості GeoNet для оцінки глибини та оптичного потоку, трансформерної архітектури для виявлення та класифікації об'єктів, а також ансамблевих методів (бегінг та бустінг) для підвищення надійності та точності прогнозів.

Для досягнення мети поставлені задачі дослідження.

1. Описати аналітичну модель GeoNet.
2. Описати модуль Бегінг.
3. Описати модуль Бустінг.
4. Описати використання трансформерної архітектури.
5. Описати алгоритм об'єднання результатів.
6. Аналіз очікуваних результатів.

GeoNet. Глибока нейронна архітектура, яка поєднує задачу оцінки глибини з задачею оцінки оптичного потоку, тобто векторного поля переміщення пікселів між двома послідовними зображеннями. Її основна мета інтегрувати просторову та часову інформацію в єдину структуру для більш точної реконструкції сцени та відстеження об'єктів. У контексті задачі ідентифікації параметрів динамічних об'єктів, GeoNet виступає як фундаментальний інструмент для попереднього збору

кінематичних характеристик об'єктів у відеопотоці. представлено загальну структуру GeoNet, де пара входних зображень послідовних кадрів обробляється мережею для одночасного визначення карти глибини та оптичного потоку. Такий підхід дозволяє об'єднати стереозору та часову динаміку, що критично важливо для ідентифікації просторового положення та напрямку руху об'єктів у сцені. Дослідження [9] демонструє, що використання GeoNet у поєднанні з трансформерною архітектурою дозволяє значно підвищити точність визначення параметрів руху завдяки можливості обробки глобального контексту у сцені. Трансформер аналізує отримані карти глибини та оптичного потоку, забезпечуючи високорівневе представлення руху, яке важливе для класифікації, прогнозування траєкторії та подальшої аналітики об'єктів.

Крім того, у реалізованій системі, запропонованій в попередній роботі [9], GeoNet може інтегруватись з ансамблевими методами машинного навчання, зокрема в цій статті багінгом та бустінгом, для підвищення стійкості до шумів, зміни освітлення та часткової втрати даних. У випадку багінгу, GeoNet запропоновано виконувати на різних модифікованих підвибірках відеокadrів, що дозволить отримати усереднені оцінки параметрів руху. У бустінгу результати першої моделі аналізуються, і друга модель фокусується на зонах з найбільшою похибкою, що забезпечує локальну точність.

Таким чином, використання GeoNet у зв'язці з трансформерами та ансамблевими підходами створює ефективну систему для високоточного аналізу динаміки візуальних сцен та ідентифікації параметрів руху об'єктів. Це особливо актуально в задачах дистанційного моніторингу, автономного транспорту та систем відеоспостереження, де важлива не лише фіксація наявності об'єкта, а й повноцінне розуміння його поведінки в просторі.

Показано на рис. 1 загальну модифіковану структуру GeoNet. Оригінальний GeoNet включає компоненти оцінки глибини (DepthNet), оцінки поз (PoseNet), оцінки оптичного потоку (ResFlowNet), і працює без вчителя (unsupervised). На рис. 1 присутні нові елементи, яких немає в оригінальній роботі трансформер, що моделює просторово-часові залежності (Transformer dynamic flow), згадки про бустінг/бегінг для підвищення точності (Ensemble methods), більш загальний підхід до трекінгу параметрів (швидкість, напрямок руху тощо), що виходить за межі задач (GeoNet Predicts the parameters), наявність блоків (Transformer parameters, Optical flow information) з підключенням до ансамблевих методів також свідчить про спроба об'єднати класичні компоненти (оптичний потік, геометрія), трансформери (для глобального контексту), та ансамблеві методи (для підвищення точності).

Оптичний потік. На рис. 2, представлено оптичний потік визначає напрям і швидкість переміщення пікселів між послідовними кадрами. Його застосування дозволяє отримати траєкторію руху об'єкта та визначити його кінематичні параметри. Оптичний потік є векторним полем, яке відображає переміщення пікселів

між двома послідовними кадрами відео. Цей метод дозволяє визначити напрямок і швидкість руху об'єктів у сцені. Завдяки цьому можна побудувати траєкторію об'єкта, визначити його миттєву швидкість і прискорення. У запропонованій системі оптичний потік виступає важливим джерелом інформації про кінематичні параметри динамічних об'єктів, особливо в ситуаціях, коли об'єкти частково або повністю перекриваються (оклюзія) [10, 11].

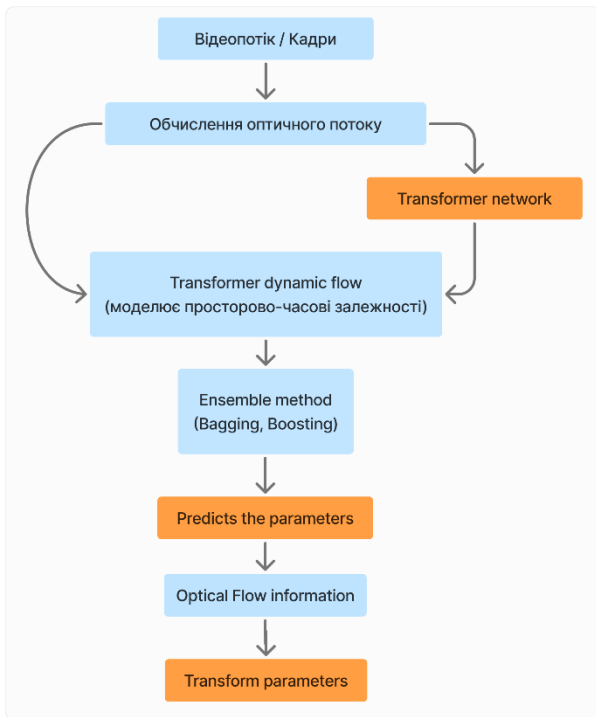


Рис. 1. Модифікована структура GeoNet

Оптичний потік. На рис. 2, представлено оптичний потік визначає напрям і швидкість переміщення пікселів між послідовними кадрами. Його застосування дозволяє отримати траєкторію руху об'єкта та визначити його кінематичні параметри. Оптичний потік є векторним полем, яке відображає переміщення пікселів між двома послідовними кадрами відео. Цей метод дозволяє визначити напрямок і швидкість руху об'єктів у сцені. Завдяки цьому можна побудувати траєкторію об'єкта, визначити його миттєву швидкість і прискорення. У запропонованій системі оптичний потік виступає важливим джерелом інформації про кінематичні параметри динамічних об'єктів, особливо в ситуаціях, коли об'єкти частково або повністю перекриваються (оклюзія) [10–11].

Ансамблевий метод Бегінг (bagging). На рис. 3, зображено бегінг (bagging), як створення кількох моделей на основі різних підвбірок вхідних даних, об'єднаних за допомогою середнього значення або голосування. Суть цього методу полягає у створенні кількох моделей, кожна з яких тренується на випадковій підвбірці вхідних даних (кадрів відео). Це дозволяє зменшити дисперсію моделі та знизити ймовірність пере-

навчання. У системі, що розглядається, на етапі попередньої обробки застосовуються варіації кадрів, зокрема випадкова зміна яскравості, масштабування та обрізка. Далі на кожній модифікації кадру запускається GeoNet, яка оцінює глибину та оптичний потік. Результати об'єднуються шляхом обчислення середнього значення карт глибини [9, 10].



Рис. 2. Візуалізація оптичного потоку в обраній моделі

Ансамблевий метод Бустінг (boosting). Послідовне навчання моделей, де кожна наступна фокусується на помилках попередньої, що дозволяє зменшити зміщення. На відміну від бегінгу, бустінг передбачає послідовне навчання моделей, де кожна наступна фокусується на помилках попередньої. У запропонованій системі це реалізовано шляхом поетапної оцінки глибини: початкова модель формує базову карту, а друга модель уточнює її у тих регіонах, де було виявлено найбільшу похибку. Це дозволяє суттєво зменшити зміщення та підвищити точність прогнозування, особливо на межах об'єктів або в умовах слабкої освітленості. У якості коректора використовується додаткова згортоква нейронна мережа, яка аналізує локальні помилки та вносить точкові покращення [9, 12].

Архітектура системи. Запропонована система складається з таких компонентів.

1. Модуль попередньої обробки відео (розбиття на кадри, нормалізація).
2. Модуль GeoNet (оцінка глибини та оптичного потоку).
3. Ансамблевий модуль Бегінг. Виконання GeoNet на різних підвбірках кадрів (випадкова зміна яскравості, кадрування, масштабування).
4. Ансамблевий модуль Бустінг. Поетапне навчання моделей, де друга модель навчається на зонах з великою похибкою від першої.

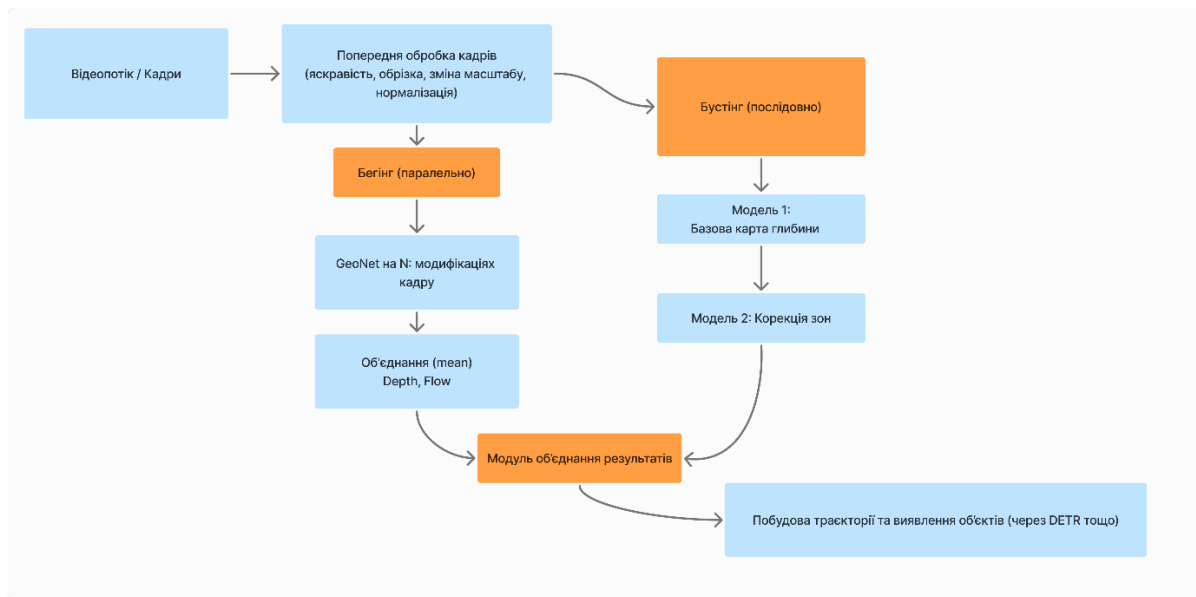


Рис. 3. Архітектура ансамблевих модулів у системі

5. DETR або інша трансформерна модель виявлення та класифікація об'єктів у кадрах.

6. Модуль об'єднання результатів. Інтеграція результатів з обох ансамблевих підходів для побудови єдиної траєкторії руху об'єктів.

Алгоритм об'єднання результатів. Для бегінгу об'єднання карт глибини методом середнього значення. Для бегінгу карти глибини, отримані з різних підвибірок кадрів, об'єднуються шляхом усереднення. Це дозволяє згладити локальні артефакти та підвищити стабільність результатів.

Для бустінгу зважене об'єднання результатів з урахуванням локальних похибок (приклад використання нейронної мережі-коректора). Для бустінгу застосовується зважене об'єднання результатів, де ваги призначаються на основі локальної похибки кожної моделі. У разі значного відхилення результатів у певній області рішення приймається на користь тієї моделі, яка показала нижчу похибку у сусідніх регіонах. Для цього використовується нейронна мережа-коректор.

На рис. 4, представлено результат візуалізації оптичного потоку та карти глибини після об'єднання. Результат отриманий за допомогою розробленого застосунку в якому Завантажується приклад відео з прикладів бібліотеки OpenCV. Вибираються два кадри. Обчислюється карта глибини (імітаційно через згладжений градієнт). Розраховується оптичний потік методом Фарнебака. Обидві карти візуалізуються окремо, а потім накладаються в одному зображенні.



Рис. 4. Результат візуалізації оптичного потоку та карти глибини після об'єднання

Очікувані результати. У майбутніх дослідженнях передбачається, що поєднання трансформерної архітектури DETR, оптичного потоку GeoNet та ансамблевих методів (зокрема бустінгу та багінгу) забезпечить підвищення точності ідентифікації параметрів руху динамічних об'єктів порівняно з традиційними підходами. Очікується, що така система зможе демонструвати стабільну роботу в складних умовах, зокрема при змінному освітленні, часткових оклюзіях та на фоні, що постійно змінюється, що є критично важливим для застосування в режимі реального часу.

Крім того, передбачається, що впровадження механізмів корекції локальних помилок за допомогою бустінгу дозволить суттєво покращити якість відновлення глибини в найбільш складних зонах зображення, таких як краї об'єктів, де традиційні методи часто демонструють знижену точність [13]. Такий підхід сприятиме більш точному просторово-часовому аналізу сцени та підвищить надійність системи для подальшого використання в автономних системах, системах відеоаналітики та моніторингу.

Висновки. Дослідження зосереджується на розробці комплексної системи для ідентифікації параметрів динамічних об'єктів у відеопотоці, що вирішує важливі завдання в комп'ютерному зорі, такі як автономне водіння, відеоспостереження та робототехніка. Основний підхід полягає в інтеграції кількох передових технологій для досягнення високої точності та надійності.

Ефективність інтеграції GeoNet. Використання GeoNet, яка одночасно оцінює глибину сцени (depth estimation) та оптичний потік (optical flow), є фундаментальним кроком. Цей підхід дозволяє отримати важливі кінематичні характеристики об'єктів, забезпечуючи більш точну реконструкцію сцени та відстеження об'єктів порівняно з традиційними роздільними методами.

Підвищення точності за допомогою трансформерів. Інтеграція трансформерної архітектури (наприклад, DETR) з даними GeoNet значно покращує точність визначення параметрів руху. Механізм уваги трансформерів дозволяє аналізувати глобальний контекст сцени, що є критично важливим для виявлення та класифікації об'єктів, а також для прогнозування їхніх траєкторій.

Переваги ансамблевих методів. Застосування ансамблевих методів, а саме бегінгу (bagging) та бустінгу (boosting), є ключовим для підвищення стійкості системи до локальних похибок, шуму, змін освітлення та оклюзій.

Бегінг зменшує дисперсію моделі, об'єднуючи результати GeoNet, отримані на різних модифікованих підвибірках кадрів, що згладжує вплив шумових артефактів.

Бустінг послідовно навчає моделі, фокусуючись на виправленні помилок попередніх етапів, що дозволяє суттєво зменшити зміщення та підвищити локальну точність, особливо на межах об'єктів.

Комплексна архітектура системи. Запропонована архітектура, що включає модулі попередньої обробки відео, GeoNet, ансамблеві модулі бегінгу та бустінгу, трансформерну модель для виявлення об'єктів та модуль об'єднання результатів, створює високоефективну систему для аналізу динаміки візуальних сцен.

Покращення якості візуалізації та аналізу. Об'єднання карт глибини методом середнього значення для бегінгу та зважене об'єднання з використанням нейронної мережі-коректора для бустінгу забезпечує більш стабільні та точні результати візуалізації оптичного потоку та глибини.

Список використаної літератури

1. Wang Z., Turko R., Shaikh O., Park H., Das N., Hohman F., Kahng M., Chau D. *CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization*. URL: <https://arxiv.org/abs/2004.15004> (дата звернення: 05.05.2025).
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. *Attention Is All You Need*. URL: <https://arxiv.org/abs/1706.03762> (дата звернення: 05.05.2025).
3. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. *End-to-End Object Detection with Transformers*. URL: <https://arxiv.org/abs/2005.12872v3> (дата звернення: 05.05.2025).
4. Zou Z., Chen K., Shi Z., Shi Z., Guo Y., Ye J. *Object Detection in 20 Years: A Survey*. URL: https://arxiv.org/pdf/1905.05055.pdf?fbclid=IwAR0ILGAWTWu-9-iH6lZyPFXyXA5JRWarM_XoSJ78QEhmn-txvr_iGEzCio (дата звернення: 05.05.2025).
5. Ammar A., Chebbah A., Fredj H., Souani C. *Comparative Study of latest CNN based Optical Flow Estimation*. URL: <https://ieeexplore.ieee.org/document/9806070/references#references>. (дата звернення: 05.05.2025).
6. Yin Z., Shi J. *GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose*. URL: <https://arxiv.org/abs/1803.02276v2> (дата звернення: 05.05.2025).
7. Girshick R., Donahue J., Darrell T., and Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2016. Vol. 38, no. 1. P. 142–158.
8. Нікуліна О. М., Северин В. П., Кондратов О. М., Ольховий О. М. Моделі дистанційної ідентифікації параметрів динамічних об'єктів з використанням трансформерів виявлення та оптичного потоку. *Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології*. Харків. НТУ «ХПІ», 2024. № 1 (11). С. 52–57.

9. Кондратов О. М., Нікуліна О. М. Програмна реалізація із використанням трансформера з оптичним потоком та GEONET для ідентифікації параметрів динамічних об'єктів. *Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології*. Харків. НТУ «ХПІ», 2024. № 2 (12). С. 86–91.
10. Нікуліна О. М., Северин В. П., Кондратов О. М., Рекова Н. Ю. Аналіз інформаційних технологій для дистанційної ідентифікації динамічних об'єктів. *Вісник НТУ «ХПІ». Серія: Системний аналіз, управління та інформаційні технології*. Харків. НТУ «ХПІ», 2023. № 1 (9). С. 110–115.
11. Gracyk A., Chen X. *GeONet: a neural operator for learning the Wasserstein geodesic*. URL: <https://arxiv.org/abs/2209.14440> (дата звернення: 05.05.2025).
12. Inomata T., Kimura K., Hagiwara M. *Object Tracking and Classification System Using Agent Search*. URL: https://www.jstage.jst.go.jp/article/ieejieiss/129/11/129_11_2065/_pdf/-char/ja (дата звернення: 05.05.2025).
13. Gavrylenko S., Chelak V., Hornostol O. Construction Method Of Fuzzy Decision Trees For Identification The Computer System State. *2022 XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA)*. 2022. P. 1–5.

References (transliterated)

1. Wang Z., Turko R., Shaikh O., Park H., Das N., Hohman F., Kahng M., Chau D. *CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization*. URL: <https://arxiv.org/abs/2004.15004> (accessed: 05.05.2025).
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. *Attention Is All You Need*. URL: <https://arxiv.org/abs/1706.03762> (accessed: 05.05.2025).
3. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. *End-to-End Object Detection with Transformers*. URL: <https://arxiv.org/abs/2005.12872v3> (accessed: 05.05.2025).
4. Zou Z., Chen K., Shi Z., Shi Z., Guo Y., Ye J. *Object Detection in 20 Years: A Survey*. URL: https://arxiv.org/pdf/1905.05055.pdf?fbclid=IwAR0ILGAWTWu-9-iH6lZyPFXyXA5JRWarM_XoSJ78QEhmn-txvr_iGEzCio (accessed: 05.05.2025).
5. Ammar A., Chebbah A., Fredj H., Souani C. *Comparative Study of latest CNN based Optical Flow Estimation*. URL: <https://ieeexplore.ieee.org/document/9806070/references#references>. (accessed: 05.05.2025).
6. Yin Z., Shi J. *GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose*. URL: <https://arxiv.org/abs/1803.02276v2> (accessed: 05.05.2025).
7. Girshick R., Donahue J., Darrell T., and Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2016, vol. 38, no. 1, pp. 142–158.
8. Nikulina O. M., Severyn V. P., Kondratov O. M., Olhovoy O. M. Models of remote identification of parameters of dynamic objects using detection transformers and optical flow. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., no. 1 (11), pp. 52–57.
9. Kondratov O. M., Nikulina O. M. Software implementation using transformer with optical flow and geonet for identifying parameters of dynamic objects. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., no. 2 (12), pp. 86–91.
10. Nikulina O. M., Severyn V. P., Kondratov O. M., Reкова N. Y. Analysis of information technologies for remote identification of dynamic objects. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., no. 1 (9), pp. 110–115.

11. Gracyk A., Chen X. *GeONet: a neural operator for learning the Wasserstein geodesic*. URL: <https://arxiv.org/abs/2209.14440> (accessed: 05.05.2025).
12. Inomata T., Kimura K., Hagiwara M. *Object Tracking and Classification System Using Agent Search*. URL: https://www.jstage.jst.go.jp/article/ieejieiss/129/11/129_11_2065/_pdf/-char/ja (accessed: 05.05.2025).
13. Gavrylenko S., Chelak V., Hornostal O. Construction Method Of Fuzzy Decision Trees For Identification The Computer System State. 2022 *XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA)*. 2022, pp. 1–5.

Надійшло (received) 15.05.2025

UDC 004.8+004.9

O. M. KONDRATOV, Postgraduate, senior lecturer of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; senior lecturer of Department IT, Analysis and Project Decisions, Technical University "METINVEST POLYTECHNICS", LLC, Zaporizhzhia, Ukraine; ORCID: <https://orcid.org/0000-0001-6367-9944>; e-mail: kondratovolexiy@gmail.com

O. M. NIKULINA, Doctor of Technical Sciences, Professor, Head of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; Professor of Department IT, Analysis and Project Decisions, Technical University "METINVEST POLYTECHNICS", LLC, Zaporizhzhia, Ukraine; ORCID: <https://orcid.org/0000-0003-2938-4215>; e-mail: elniknik02@gmail.com

IDENTIFICATION PARAMETERS OF DYNAMIC OBJECTS USING TRANSFORMER WITH OPTICAL FLOW AND ENSEMBLE METHODS

The article presents an approach to identifying the parameters of dynamic objects in a video stream using a transformer-based architecture, the GeoNet model, and ensemble machine learning methods, namely bagging and boosting. The identification of parameters such as position, velocity, direction of movement, and depth is of significant importance for a wide range of applications, including autonomous driving, robotics, and video surveillance systems. The paper describes a comprehensive system that integrates the spatiotemporal characteristics of a video stream by computing optical flow and depth maps using GeoNet, further analyzing them through a transformer, and enhancing accuracy via ensemble methods. GeoNet, as a deep convolutional neural network, combines the tasks of depth estimation and optical flow within a single architecture, enabling accurate 3D scene reconstruction. The use of a transformer allows modeling global dependencies across video frames and improves the accuracy of object classification and detection. At the same time, bagging reduces variance by averaging the results of several models trained on different subsets, while boosting focuses on difficult examples to improve prediction accuracy. The proposed system achieves high accuracy under conditions of dynamic background, lighting changes, occlusions, and noise, making it adaptable for real-time use in complex scenes. A detailed description of each system component is provided: the GeoNet architecture, transformer modules, implementation of bagging and boosting, and the result fusion algorithm. The expected results are intended to demonstrate the effectiveness of integrating deep learning methods with classical ensemble approaches for high-precision dynamic object identification tasks. The proposed methodology opens new prospects for the development of next-generation intelligent computer vision systems.

Keywords: Remote identification of dynamic objects, object detection, computer vision, ensemble methods, deep learning, convolutional neural networks, system architecture, software, machine learning, artificial intelligence.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Кондратов Олексій Михайлович, Kondratov Oleksii Mikhailovich

Автор 2 / Author 2: Нікуліна Олена Миколаївна, Nikulina Olena Mykolaivna