

*В.Д. Зівакін, О.О. Козачук, д.т.н. проф. П.О. Приставка
(Національний Авіаційний Університет, Україна)*

Навчальний датасет для нейромережевої обробки даних повітряного спостереження

Успіх широкого впровадження технологій штучного інтелекту зумовлює зростання цінності наборів даних, на основі яких відбувається створення та навчання інтелектуальних систем різноманітного призначення. Представлено набір даних для нейромережевого навчання, які отримані під час повітряного спостереження (аерозйомки). Надається результат роботи та розширення набору протягом чотирирічного періоду.

Важливо переоцінити важливість створення, поповнення та використання навчального набору для розпізнавання даних з камерного повітряного спостереження (даних аерозйомки). Саме тому, починаючи з 2018 року на кафедрі прикладної математики Національного авіаційного університету започатковано роботи за цим напрямком, зокрема було розпочато створення та поповнення навчального набору «Аерозйомка». У першу чергу зазначений набір даних розглядається як стартовий майданчик для розгортання системи машинного навчання, їх дослідження та постійного вдосконалення. За час роботи із набором «Аерозйомка» вчені кафедри студенти та аспіранти ставили та вирішували різноманітні задачі, вдосконалюючи, як сам навчальний набір, так і системи інформаційних технологій на його основі.

Динаміка проведених досліджень відображається у таблиці 1.

Таблиця 1.

Динаміка розвитку датасету та досліджень

Рік	Кількість класів	Кількість зображень	Види нейронних мереж та досліджень
2018	4	~2000	Згорткові нейронні мережі
2019	11-13	>16000	Згорткові мережі та мережі із ауто-енкодером. Дослідження якості визначення окремих (цільових) класів.
2020	7-13	~20000	Змагальні мережі та MASK R-CNN. Тренування мереж-класифікаторів на даних, що генеровані GAN.
2021	11-13	>22000	Робота із представленнями, отриманими із згорткового ауто-енкодера. Визначення класів, що не були формалізовані, тощо.

Дослідження останніх двох років представляють більшу цінність, адже проводилися із вдосконаленим набором даних та використовували складніші моделі нейронних мереж. Окрім того більшу цікавість викликають отримані результати.

Наприклад, цікавим досвідом стало дослідження моделі генеративних мереж (GAN). Близько двадцяти тисяч зображень були оброблені мережею, після чого була згенерована нова вибірка. Обидві вибірки впродовж проведення досліджень слугували як тренувальними, так і тестовими для згорткової мережі-класифікатора. Цінними є результати дослідження роботи цього класифікатора, які відображені у таблиці 2.

Таблиця 2.

Порівняння точності класифікації різних вибірок

		Датасет для перевірки (тестовий)	
		Реальні дані	Генеровані дані
Датасет для навчання (тренувальний)	Реальні дані	95%	80%
	Генеровані дані	98%	83%

Саме після цього експерименту почалася робота із різними видами представлень.

Загалом на сьогодні датасет складається із 16 класів, містить в собі понад 80000 зображень розміром 64x64 пікселя та має наступну структуру:

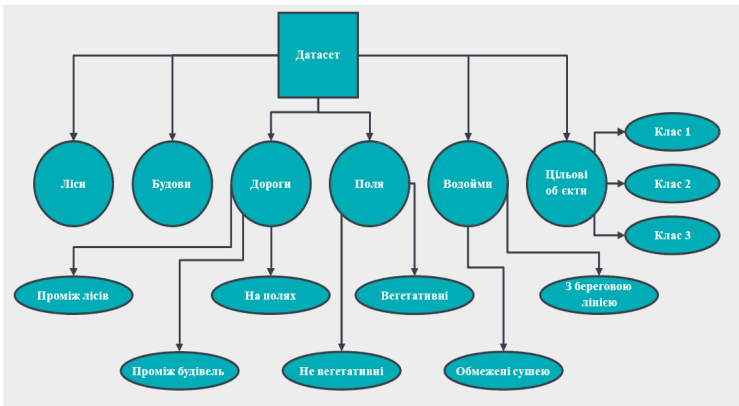


Рис. 1. Структура набору даних на сьогодні

Приклади зображень у наборі продемонстровану на рисунку 2.

Зображення	Клас зображення	Кількість
	Ліси	>40000
	Будови	>4000
	Дорога серед будівель	<1000
	Дорога серед дерев	> 2500
	Стояпи	>4000

Рис. 2. Приклади зображень датасету.

Як показує навіть стислий наведений огляд виконаних досліджень, наявність навчального набору «Аерозімка» дозволяє широко проводити науково-практичні дослідження в області машинного навчання та нейромережевого розпізнавання елементів місцевості. З прикладного боку, наявність такого набору вже зараз дозволяє вирішити цілий ряд завдань відеоаналітики, в тому числі і подвійного призначення: різноманітний повітряний моніторинг автошляхів, сільгосп-полів, лісових насаджень тощо, проведення аеророзвідки з пошуку цільових об'єктів у зоні бойових дій, використання в системах супроводу та цілевказання.

Список літератури

1. Тітенко Д.С. CNN-технологія розпізнавання об'єктів за даними аерофотозйомки:дип. ... бакалавр/ НАУ. Київ, 2018 58 с.
2. Карпенко Д.В. Автоматизована система розпізнавання об'єктів за даними аерофотозйомки:дип. ... бакалавр / НАУ. Київ, 2019. 60 с.
3. Сулейманова С.Р. Нейромережеве розпізнавання класів зображено за даними аерофотозйомки:дип. ... магістр / НАУ. Київ, 2020. 90 с.
4. Козачук О.О. Інформаційні технології розпізнавання класів доріг за даними аерофотозйомки: дип. ... бакалавр / НАУ. Київ, 2020. 61 с.
5. Карпенко Д.В. Технологія нейромережевої генерації навчального набору цифрових зображень:дип. ... магістр / НАУ. Київ, 2020. 81 с.
6. Тітенко Д.С. Автоматизація наповнення навчальної вибору за даними повітряного спостереження:дип. ... магістр/ НАУ. Київ, 2020 100 с.
7. Гончарук Ю.М. Ідентифікація класів даних аерофотозйомки, не формалізованих у навчальному наборі: дип. ... магістр / НАУ. Київ, 2021. 83 с.
8. Козачук О.О. Інформаційна технологія моделювання навчальних наборів даних аерозйомки на основі структур нейромереж з автоенкодером: дип. ... магістр / НАУ. Київ, 2021. 93 с.