# Вступ

Задача локалізації об’єктів на зображеннях є однією з найбільш інтенсивно досліджуваних проблем в області комп’ютерного зору і включає декілька пов’язаних задач: класифікація зображень, локалізація об’єктів шуканого класу, знаходження конкретного об’єкту, знаходження «схожих» зображень, у тому числі на з відео та ін.. Людина вирішує задачу локалізації і класифікації об’єктів напрочуд якісно та швидко, що не є дивним, оскільки вона навчається це робити усе своє життя.

# Постановка задачі

# Огляд проблеми

Задача локалізації об’єктів не є повністю вирішеною на сьогодні. Під час аналізу зображень комп’ютером виникають складнощі, пов’язані з тим, що об’єкти одного класу можуть сильно відрізнятися, а саме:

* мінливість світлотіньового малюнку, текстур (а)
* мінливість кута зору (б), масштабу (в)
* внутрішньокласова мінливість (рис. 1в)
* мінливість частин цілого, форм, поз (г)
* перекриття іншими об’єктами (д), тощо.

  

а б в

 

г д

1. Труднощі, що виникають під час локалізації класів об’єктів

Деякі проблеми вважаються вирішеними, наприклад пошук облич методом Віоли-Джонса[1], чи, наприклад, частково вирішеною є проблема локалізації пішоходів методом Далала-Триггса[2]. Значний прорив у вирішенні задачі локалізації об’єктів відбувся у 2012 р. з появою роботи Крижевського та ін.[3]. Алгоритм базувався на нейронній мережі, для його навчання використовувалися можливості графічних процесорів та велика за обсягом навчальна вибірка. Цей класифікатор був здатен розпізнавати об’єкти 1000 класів з такою точністю, що тільки у 26% випадків метод помилявся і не містив у п’яти найбільш ймовірних гіпотезах правильної відповіді. Згодом ці результати тільки покращувалися іншими авторами, але, не зважаючи на успіх нейронних мереж, вони мають свої недоліки. Зокрема якість роботи іноді може сильно залежати від незначних змін у параметрах. Саме тому наразі активно експериментують з нейронними мережами: майже кожна корпорація (наприклад, Google, Yandex, Baidu) має своїх дослідників. Для порівняння результатів досліджень є міжнародні конкурси з комп’ютерного зору, які є одночасно базою навчальних та тренувальних вибірок, наприклад Pascal VOC Challenge[[1]](#footnote-2) (2005 – 2012), або ImageNet[[2]](#footnote-3) (з 2010 року).

За роки існування дисципліна комп’ютерного зору здобула певний набір технік, що дозволяє покращувати вже існуючі алгоритми та «заточувати» їх під конкретну предметну область чи розпізнавання однотипних сцен (наприклад, машини, будівлі, пішоходів на вулиці). Деякі з них використані у дипломній роботі. Проблеми, наведені на початку розділу, частково вирішують великим об’ємом тренувальних даних. Поява та розвиток фотохостингів та графічних прискорювачів зробили можливим отримувати та обробляти величезні об’єми зображень за реальний час.

Деякі приклади програмного забезпечення з локалізації об’єктів можна знайти у відкритому доступі, наприклад у [4].

1. **P. Viola, M. Jones** Robust Real-time Object Detection / *IEEE ICCV Workshop Statistical and Computational Theories of Vision,* July 2001.
2. **N. Dalal, B. Triggs** Histograms of Oriented Gradients for Human Detection / *CVPR (1)*, pp. 886–893, 2005
3. **A. Krizhevsky, I. Sutskever, G. Hinton** ImageNet classification with deep convolutional neural networks / *Neural Information and Processing Systems,* 2012.
4. Visual Geometry Group. Department of Engineering Science, University of Oxford [Електронний ресурс] – Режим доступу <http://www.robots.ox.ac.uk/~vgg/research/>

1. Pascal2. Pattern Analysis, Statistical Modelling and Computational Learning [Електронний ресурс] – Режим доступу: <http://host.robots.ox.ac.uk/pascal/VOC/> [↑](#footnote-ref-2)
2. IMAGENET [Електронний ресурс] – Режим доступу: <http://image-net.org/> [↑](#footnote-ref-3)