# Тема 11. Машинний та комп'ютерний зір

Створення алгоритмів, що відтворюють функції людського зору, не просто складна проблема, а цілий набір взаємозалежних труднощів.

До прикладу, якщо хтось кидає м'ячик і людина його ловить. У реальності відбувається приблизно наступне: зображення м'ячика проходить крізь очі і потрапляє в сітківку, яка проводить елементарний аналіз і надсилає його в мозок, де зорова ділянка кори головного мозку робить більш глибокий аналіз зображення. Далі зображення надсилається в інші відділи кори, де порівнюється з вже відомими об'єктами і співвідноситься з якоюсь категорією. Мозок приймає рішення, як потрібно відреагувати на побачене: наприклад, підняти руку і зловити м'ячик (розрахувавши приблизну траєкторію його польоту). Все це відбувається за частки секунди, без жодних свідомих зусиль, і практично завжди спрацьовує без помилок.

Інтерес до створення технічного зору виник одним з перших в області штучного інтелекту. Навіть архітектура першої штучної нейронної мережі - перцептрону - була запропонована Френком Розенблатом, по аналогії з сітківкою ока, а її дослідження проводилося на прикладі задачі розпізнавання зображень символів.

Прогрес в області технічного зору визначається кількома факторами: розвиток теорії, алгоритмів, методів і апаратного забезпечення. Довгий час теорія і академічні дослідження випереджали можливості практичного використання систем технічного зору.

Масове застосування методи технічного зору отримали в 2010-х роках з досягненням відповідного рівня продуктивності процесорів комп'ютерів. В плані практичного застосування системи технічного зору пройшли ряд етапів:

* Індивідуальні рішення (як в частині апаратного забезпечення, так і алгоритмів) конкретних завдань.
* Застосування в професійних областях (особливо в промисловості і оборонній сфері) з використанням спеціальних процесорів.
* Спеціалізовані системи розпізнавання зображень і алгоритми, що призначені для роботи в умовах невизначеності.
* Масове застосування роботів.

В області досліджень і реалізації технічного зору існує два напрямки: машинний і комп'ютерний зір. Ці області тісно пов'язані між собою, термін  **комп'ютерний зір** позначає загальну назву набору технологій, а машинний зір – сферу застосування.

**Комп'ютерний зір** - це одночасно і теорія, і набір пов'язаних з нею технологій про те, як машини можуть візуально сприймати об'єктивну реальність. За технологіями комп’ютерного зору машини можуть знаходити, відстежувати, класифікувати та ідентифікувати об'єкти, витягуючи дані з зображень і аналізуючи отриману інформацію.

Комп'ютерний зір застосовується для розпізнавання об'єктів, відеоаналітики, опису вмісту зображень і відео, розпізнавання жестів і рукописних символів, а також для інтелектуальної обробки зображень.

**Машинний зір** - це технології, які допомагають пристроям побачити процес виробництва чого-небудь, проаналізувати дані і прийняти інформоване рішення. Машинний зір використовує аналіз зображень для того, щоб ефективно вирішувати промислові чи побутові завдання.

Машинний зір використовується в різних областях. В медицині - для точної діагностики, в промисловості - для зменшення собівартості товарів за рахунок автоматизації, в автомобільній індустрії - для навігації безпілотників, в торгівлі - для зчитування штрихкодів або підрахунку відвідувачів.

## 12.1. Машинний зір

Алгоритми машинного зору багато чого моделюють з функції природнього зору. Світлові частки (фотони) постійно відбиваються від різних об'єктів і потрапляють на сітківку очей. У кожному оці знаходиться приблизно 126 мільйонів чутливих до фотонів клітин, які розшифровують інформацію і надсилають її в мозок. Ці клітини є двох типів - колбочки і палички. Перші відповідають за розпізнавання кольору, другі дозволяють, зокрема, бачити вночі, працюючи з відтінками сірого. Колбочок є три типи - одні спеціалізуються на червоних кольорах, другі на зелених, треті - на синіх. Їх комбінації надають можливість людському оку сприймати мільйони відтінків.

Людська зорова система, втім, не сама просунута на планеті. Куди складніше влаштовані, наприклад, очі раків-богомолів. В них відразу 16 видів колбочок, їхні очі рухаються незалежно один від одного, і кожне око розділене ще на три частини. При цьому у раків-богомолів дуже маленький і примітивний мозок у порівнянні з людським. Він не може обробляти великі дані, але від очей отримує вже готову детальну інформацію. У людей навпаки - очі влаштовані дещо простіше, зате мозок - найпотужніший серед всіх видів.

В машинному зорі застосовуються обидва підходи отримання вхідних даних.

* Системи зі звичайними цифровими (іноді навіть аналоговими) камерами, які реагують на спеціальні давачі, отримують сире зображення, обробляють його, розпізнають елементи і їх закономірності, приймають рішення і надсилають сигнал до інших систем.
* Розумні камери, які самостійно проводять частину аналізу і розвантажують процесори системи.

Ще десяток років тому технології машинного зору були менш досконалими і успішно розпізнавали лише 65-70% об'єктів, що потрапляли в їх поле видимості. Це високий показник, але недостатній для того, щоб можна було довірити машинному зору відповідальні завдання. Зараз машини вже розпізнають до 98% об'єктів. Причому дійсно розпізнають - не лише фіксують наявність, а й визначають, що саме вони бачать і вирішують, що робити далі.

Системи сприйняття візуальної інформації в людини є більш гнучкими. Людина краще інтерпретує контекст. Машини старанно вивчають нові для них ситуації, а людина завжди може здогадатися, як діяти у невідомих для неї випадках. Тому, частка вдалих випадків розпізнавання машиною тримається на 98% і не досягає 100%.

Однак, у систем машинного зору є одна безперечна перевага перед людським зором. Людина не може сконцентруватися більш як на трьох-сімох об'єктах одночасно. Звісно, це залежить від особливостей конкретної людини, але мало хто може охопити увагою більшу кількість. Системи машинного зору фіксують абсолютно всі об'єкти і дії, які через зображення надходять до їх процесорів. Увагу комп'ютера неможливо відвернути - для нього все, що захоплено машинним зором має рівне значення.

Машинний зір зосереджується в основному на промисловому застосуванні, наприклад, автономні роботи, системи візуальної перевірки та вимірювань. Це означає, що технології давачів зображення і теорії управління пов'язані з обробкою відеоданих для управління роботом. Обробка отриманих даних в реальному часі здійснюється програмно або апаратно.

#### Завдання машинного зору

* Виявлення, розпізнавання та ідентифікація об’єктів.
* Аналіз оптичного потоку.
* Оцінка руху об’єктів.
* Виділення на зображеннях структур певного виду, сегментація об’єктів.

### 12.1.1. Популярні застосування машинного зору

#### Робототехніка

Традиційною сферою застосування машинного зору є робототехніка, як промислова так і побутова. Підвищення продуктивності процесорів дозволяє вирішувати широке коло завдань, зокрема:

* Набір завдань, пов'язаних з орієнтацією в навколишньому просторі, визначенням відстаней до об'єктів, прокладання маршрутів і подібне.
* Розпізнавання різних об'єктів і інтерпретація сцен в цілому.
* Виявлення людей, розпізнавання осіб і аналіз емоцій.

##### Системи допомоги водієві

Роботи з розпізнавання розмітки, перешкод на дорозі чи знаків активно велися і в 90-х роках. Однак достатнього рівня за точністю і надійністю методів, за продуктивністю процесорів, здатних в масштабі реального часу виконувати відповідні методи вони досягли переважно в останньому десятилітті.

* Стерео огляд для виявлення перешкод на дорозі.
* Розпізнавання пішоходів та дорожніх знаків.
* Контроль стану водія.
* Вдосконалення поведінки безпілотних автомобілів на незнайомих дорогах і за поганих погодних умов.

##### Мобільні додатки

Широкий клас додатків для мобільних пристроїв пов'язаний з завданнями доповненої реальності, які можуть бути дуже різними. Сюди відносяться ігрові програми (що вимагають узгодженого відображення віртуальних об'єктів над зображенням реальної сцени при переміщенні камери), а також різні розважальні програми в цілому, туристичні програми (розпізнавання пам'яток з виведенням інформації про них), а також багато інших додатків, що пов'язані з інформаційним пошуком і розпізнаванням об'єктів: розпізнавання написів на іноземних мовах з відображенням їх перекладу, розпізнавання візитних карток з автоматичним занесенням інформації до телефонної книги, а також розпізнавання осіб з витягуванням інформації з телефонної книги, розпізнавання постерів фільмів (з заміною зображення постера на трейлер фільму) тощо.

Системи доповненої реальності можуть створюватися у вигляді спеціалізованих пристроїв: окулярів чи шоломів, що ще більше збільшує інноваційний потенціал методів машинного зору.

##### Інформаційний пошук та навчання

Багато задач доповненої реальності (Augmented Reality, AR) тісно пов'язані з інформаційним пошуком.

* Розпізнавання об'єктів у реальному часі та визначення їхнього розташування в навколишньому середовищі. Це може бути корисним для інформаційного пошуку, коли користувач хоче отримати додаткову інформацію про конкретний об'єкт або місце, наприклад, пам'ятки, магазин або ресторан. AR-програми можуть надавати контекстуальну інформацію, таку як відгуки, рейтинги, розклад роботи та інші дані, які допоможуть користувачеві прийняти рішення.
* AR може використовуватися для створення інтерактивних візуальних навігаційних систем, які допомагають користувачам знайти місця, що їх цікавлять, або об'єкти в реальному світі. Наприклад, AR-програма може відображати на екрані напрямки та відстані до найближчих ресторанів, готелів, аптек та інших об'єктів, ґрунтуючись на інформації про місцезнаходження користувача та даних про точки інтересу.
* AR може використовуватись для розпізнавання тексту та об'єктів у реальному часі. Це може бути корисним для інформаційного пошуку, коли користувач хоче отримати додаткову інформацію про текст або об'єкт, з яким вони взаємодіють. Наприклад, AR-програма може розпізнати текст на вивісці та надати переклад або додаткову інформацію про магазин. Також AR може розпізнавати об'єкти, такі як книги, продукти або витвори мистецтва, та надавати пов'язані з ними дані чи контекст.
* AR може використовуватися для візуалізації додаткової інформації безпосередньо на екрані пристрою в реальному часі. Наприклад, при перегляді товару в магазині AR-програма може відображати ціни, описи, відгуки або супутні товари. Це допомагає користувачам прийняти більш поінформоване рішення та спрощує інформаційний пошук.

Ці завдання можуть зустрічатися в різних сучасних мобільних додатках. Вирішення цих задач зараз немислимо без залучення методів машинного навчання в області технічного зору.

## 12.2. Комп’ютерний зір

Завдання комп'ютерного зору в загальній постановці (розуміння, що зображено на картинці) вважається надто складним, вирішення якого еквівалентно штучному інтелекту. Для розуміння вмісту зображення людина використовує весь свій життєвий досвід, освіту, а іноді і інтуїцію. На сьогодні, «сильного» штучного інтелекту, що за своїми можливостями співмірний до людського поки не створено, тому, розробники систем комп'ютерного зору вирішують лише певні задачі.

Зорове сприйняття людини є досконалим надбанням природи. Людський мозок - це унікальна система розуміння і обробки тієї інформації, яку бачить людина. Вона схильна бачити те, що хоче бачити. Мозок часто шукає знайомі образи, «добудовує» картинку і додає семантику.

Людська зорова система самонавчається. Європейцеві складно розрізняти обличчя азіатів, оскільки, зазвичай, в житті він рідко їх зустрічає. Його зорова система навчилася вловлювати відмінності в європейських особах, а азіати, яких він бачив мало, здаються йому «на одне обличчя» і навпаки. Але достатньо провести довший час в Азії, європеєць помічає відмінності обличь і без ускладнень розпізнає осіб.

Ключовою проблемою у складності розпізнавання вмісту зображення є різниця у семантичному сприйнятті картинки людиною і комп’ютером. Людина легко виділяє на зображенні важливе і неважливе, наповнює її змістом, за неявними ознаками розуміє багато деталей (рис.12.1).

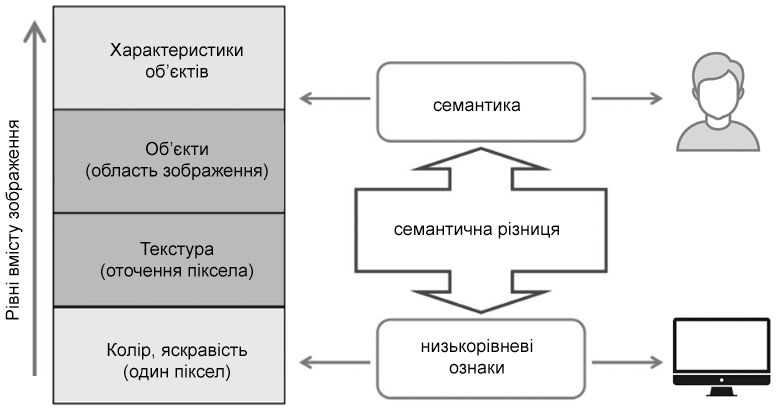


Рис.12.1. Особливості сприймання візуальної інформації

Комп'ютер на відміну від людини, оперує не образами, а числами. Для комп'ютера зображення - це набір пікселів, в кожного з яких є своє значення яскравості або кольору. Комп'ютер розуміє колір пікселів, вміє виділити контури і текстуру, що успішно спрацьовує для розпізнавання відомих об’єктів та композицій, але розпізнати приховані чи неявні ознаки зображеного для нього наразі складно. Щоб машина змогла отримати знання про вміст картинки, зображення обробляють за допомогою спеціальних алгоритмів.

Люди дуже добре вміють узагальнювати, за зображенням здатні здогадатися, що це за предмети, їх роль в загальній композиції, емоційний тон та неявні ознаки (вітер, спека, радість). Людині не потрібно знати всі модифікації об’єкта з одного класу, щоб віднести до нього зразок.

Щоб відповісти на питання, що зображено на фотографії, можна описати сцену в цілому (рис.12.2). Зрозуміло, що картинка зроблена на вулиці (поза приміщенням), десь в азіатській країні. Можливо хтось впізнає площу Тяньаньминь в Пекіні, якщо був там раніше. На зображенні легко розпізнати окремі об'єкти: автобус, портрет, дах, небо і фізичні властивості окремих об'єктів. Наприклад, дах - похилий, автобус їде, на стіні висить зображення Мао Цзедуна, вітер дує справа наліво (це можна визначити за напрямком руху прапорів).



Рис.12.2. Фотографія, що надана до розпізнавання

Отже, для відповіді на питання, що зображено на фотографії, людина використовує весь свій життєвий досвід. Наприклад, знання про те, що існує вітер (на зображенні його не можна побачити явно), що таке транспорт, базові знання історії Китаю. Відповідно, для розуміння вмісту фотографії людина дивиться не на пікселі, а використовує власні знання.

Інший приклад демонструє здібності людини здогадуватися та адаптувати свої знання до побаченого. На питання, що таке стілець, можна відповісти перше, що прийде в голову. Наприклад, стілець - це щось з чотирма ніжками і спинкою. А якщо стілець не підходить під стандартний опис, тоді навіть таку річ, як стілець, складно описати в термінах форм (рис.12.3).



Рис.12.3. Дизайнерська реалізація стільця

Стілець - це якесь концептуальне поняття: щось, на чому сидять. Можна уявити, як складно словами пояснити суть поняття, що таке сидіти. Перш ніж навчити систему знаходити на зображеннях стілець, варто, щоб вона розуміла концепцію «сидіти». В ідеалі, щоб система відповідала на питання про стільці так само добре, як людина, їй потрібно розуміти концепцію «сидіти».

Методи комп'ютерного зору використовуються для вирішення завдань, які умовно можна розділити на прості і складні.

* Складні завдання відповідають на питання, який об'єкт зображено, до якого класу він належить. Для вирішення цих завдань найчастіше використовуються методи машинного навчання, а комп’ютерний зір називається **високорівневим**.
* При вирішенні простих завдань здійснюються маніпуляції безпосередньо з пікселями, використовуються евристики, а методи машинного навчання, як правило, не застосовуються. Такі завдання називаються **низькорівневим** комп’ютерним зором

Завдання низькорівневого комп'ютерного зору часто використовуються як складова частина більш складних завдань розпізнавання. Наприклад, попередня обробка картинки дозволяє алгоритмам машинного навчання краще зрозуміти, що на ній зображено.

Головний прогрес відбувається в сфері високорівневого комп'ютерного зору, що пов'язана з розумінням зображень. Тут вирішуються завдання, які легко сформулювати. Наприклад, визначити об'єкти на фотографії, їх клас (людина, автомобіль, собака, дерево) та розташування. Або визначити, що роблять люди на відео, спокійна ситуація чи починається щось незвичайне, наприклад бійка. Тому, недостатньо просто встановити багато камер спостереження - потрібно, щоб потокове відео з них розумно оброблялося.

На початку 21 століття алгоритми високорівневого зору працювали погано. Традиційні класичні підходи не могли впоратися з такими завданнями, як відрізнити зображення котів від собак, або розпізнати їх породи. Сучасні алгоритми і глибинне навчання надають якість розпізнавання до 99%. Самими досконалими інструментами на сьогодні вважаються згорткові нейронні мережі, яким можна подати на вхід зображення, і отримати точну відповідь, що там зображено (кіт чи собака) і навіть якої породи.

Згорткові нейронні мережі мають величезну кількість параметрів, які самоналаштовуються в процесі машинного навчання. Навчання полягає на перегляді великої кількості навчальних фотографій, що спонукає мережу шукати однакові ознаки та властивості об’єктів.

Навчальна вибірка повинна бути досить великою та репрезентативною (необхідно, щоб вона відображала природу даних, що мають розпізнаватися). Важливою умовою для вибірки є наявність як вірних ( «на зображенні є кіт»), так і невірних ( «на зображенні не кіт») прикладів (рис.12.4).

Після складання вибірки, починається машинне навчання. В ході навчання комп'ютер аналізує зображення з вибірки, визначає, які ознаки та їх комбінації вказують на те, що на зображенні - котик, і обчислює їх значимість. Якщо навчання пройшло успішно (щоб упевнитися в цьому, проводяться перевірки), то машина може застосовувати отримані знання «на практиці» - знаходити котів на різних зображеннях.



Рис.12.4. Навчальна вибірка з зображеннями котівта інших тварин

Після навчання на одних об’єктах, нейронна мережа спроможна навчатися і для розпізнавання інших класів. Навчена нейромережа буде виявляти зображення, що містять схожі об'єкти або один і той же об'єкт, і розподіляти їх до відповідних класів. Це важлива властивість для всіх систем обробки інформації. Відбувається узагальнення і перенесення знань на нові класи, навіть якщо мережа під час навчання раніше не бачила цих об'єктів.

Значний прогрес високорівневого комп'ютерного зору відбувається з кількох причин.

* Комп'ютери стають потужнішими, що пришвидшує обчислення.
* З'являється більше якісних даних, і вони стають кращими за рахунок вдосконалення фото- і відеокамер, медичних сканерів, приладів у біологів і астрономів. Чим вище якість даних, тим простіше з них витягувати інформацію.
* Покращуються алгоритми, що пов'язано насамперед з глибинним навчанням і згортковими нейромережами.

Багато сфер людської діяльності потребує додатків з комп’ютерним зором, тому є величезний попит на різноманітні реалізації і талановитих фахівців та розробників.

### Популярні завдання комп'ютерного зору

Комп'ютерний зір знаходить широке застосування у багатьох областях.

* Розпізнавання об'єктів та класифікація. Системи навчаються розпізнавати об'єкти на зображеннях та надавати їм відповідні класи. Приклади включають розпізнавання осіб, автомобілів, тварин та інших об'єктів.
* Сегментація зображень. Розділення зображення на окремі сегменти, які відповідають різним об'єктам чи регіонам інтересу. Це важливо для розпізнавання об'єктів та аналізу сцен.
* Розпізнавання пози, рухів, жестів та жестових команд. Визначення пози людини або об'єкта на зображенні або відео, а також відстеження та аналіз рухів. Це може використовуватися в іграх, анімації, віртуальній реальності, робототехніці, медичних програмах та інших областях.
* Пошук та витягування інформації. Використовується для автоматичного витягування інформації з зображень, таких як текст, штрих-коди, номери та інші дані.
* Медична діагностика. Допомагає у розпізнаванні та аналізі медичних зображень, таких як рентгенівські знімки, МРТ для діагностики захворювань.
* Розпізнавання тексту (OCR). Перетворення тексту з зображень або документів в електронний текст. OCR використовується для сканування документів, розпізнавання номерів, автоматичного заповнення форм та інших додатків.
* Розпізнавання живопису та зображень. Аналіз художніх творів та розпізнавання стилів та характеристик картин.
* Виявлення об'єктів. Знаходження та локалізація конкретних об'єктів на зображенні. Це може бути виявлення осіб, автомобілів, пішоходів, тварин або будь-яких інших об'єктів, які цікавлять дослідника або розробника.
* Розпізнавання обличь. Ідентифікація та аутентифікація осіб на зображеннях або відео. Це може використовуватися в системах безпеки, системах керування доступом, системах відеоспостереження та інших програмах.
* Реконструкція 3D-моделей. Створення тривимірних моделей сцен чи об'єктів на основі двовимірних зображень чи відео. Це може бути корисним у віртуальній та доповненій реальності, архітектурному моделюванні та інших областях.

Це лише деякі з популярних завдань у галузі комп'ютерного зору. Існує багато інших завдань та додатків, які використовують можливості комп'ютерного зору для аналізу, обробки та розуміння зображень та відео. З розвитком технологій та методів машинного навчання, його застосування стають все більш різноманітними та різноманітними.

## 12.3. Пошук зображень в Інтернеті

### Пошук зображення за метаданими

На початковому етапі свого розвитку пошук зображень ґрунтувався виключно на отриманні та аналізі метаданих, пов'язаних безпосередньо з зображеннями: атрибутів alt, заголовків сторінок, текстів посилань на зображення. Поступово стали враховувати текст, що розташований на тій же веб-сторінці, що і зображення. Таким чином, завдання пошуку зображення деякий час обмежувалося знаходженням всієї можливої текстової інформації, що відноситься до нього і визначенням ймовірності, з якою ця інформація стосується саме до цього зображення.

### 12.3.1. Пошук схожих зображень за простими ознаками

Наступним кроком став пошук не за текстовим, а візуальним запитом. Перші алгоритми мали спроможності аналізувати лише прості ознаки наданого для пошуку зображення і відшукати у своїх базах ідентичну або подібну картинку. Це були перші спроби дивитися на зображення очима людини і розпізнавати його вміст: контури та форми об'єктів, їх кольори та текстуру, взаємне розташування та загальну композицію. Кожен метод ретельно відпрацьовував свій напрямок ознак, а у поєднанні вони надавали пристойні результати. Окрім ідентичних картинок знаходилися їх дублікати, що мали певні зміни чи спотворення: інший розмір, поворот, додавання або видалення певних фрагментів, зміна кольору, обрізування.

#### Межі об’єктів

Алгоритмів виділення меж є кілька, вони працюють досить добре, але подібна інформація слабо характеризує те, як картинка буде сприйматися візуально. Тому, цей клас ознак може служити тільки допоміжним (рис.12.5).



Рис.12.5. Виділення контурів зображення

#### Текстура

Текстура - це область зображення, де між сусідніми точками можуть бути суттєві перепади яскравості і кольору, але яка при цьому візуально сприймається як однорідна область (наприклад, трава, поверхня води, шкіра людини тощо). Ця інформація має певне семантичне навантаження щодо виявленого об’єкту, але ці ознаки теж можуть бути лише вторинними, хоча їх значення дещо більше (рис.12.6).



Рис.12.6. Текстура крапель води на рослині

#### Опорні точки

На зображенні вибирається певна множина точок (наприклад, самі контрастні кольори на зображенні). Далі обчислюються взаємне розташування опорних точок як характеристики картинки. В результаті для кожної картинки складається рівняння, що описує розташування цих точок. Алгоритм передбачає порівняння рівнянь всіх наявних картинок і знаходить схожу.

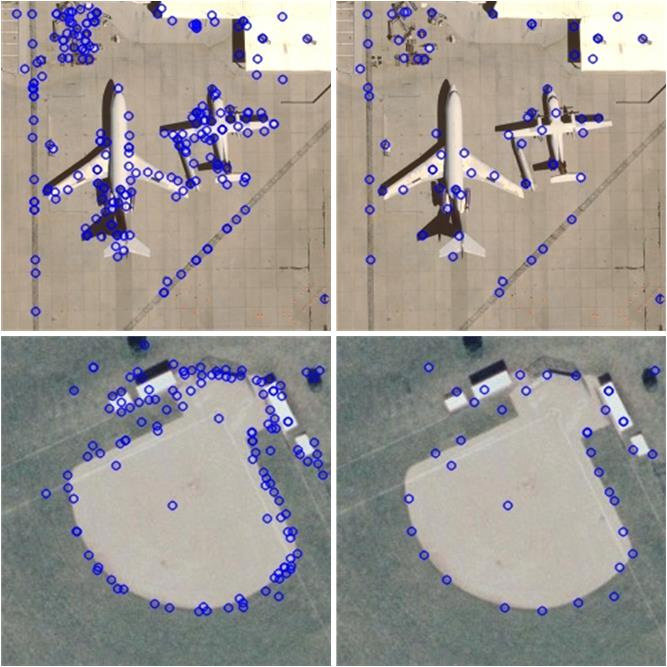


Рис.12.7. Обчислення опорних точок навколо об’єкта

Ця група методів досить хороша для виявлення деформованих копій зображення, однак для визначення подібності двох принципово різних, нехай і візуально схожих картинок, метод непридатний (наприклад, дві фотографії одного і того ж котика в різних позах матимуть мало спільного для таких методів). Тому, такий підхід може лише визначати, чи є в базі модифікації шуканого зображення, але не може знаходити подібні в будь-якому іншому сенсі.

#### Колір

Колір є, мабуть, найбільш значущою для аналізу характеристикою - це візуально дуже важлива ознака. Для більшості реальних зображень домінуючих кольорів, зазвичай, є не більше 6-7, а часто і зовсім 3-4. Це важливо, оскільки за такої невеликої кількості ознак пошук по базі картинок можна реалізувати значно швидше, ніж пошук по базі html-сторінок (рис.12.8).



Рис.12.8. Визначення домінуючих кольорів зображення

Перші успішні результати показала канадська компанія Tineye, в чому вона випередила Google. Визначаються домінуючі кольори (чим більше, тим краще), запам'ятовуються і знаходяться на інших картинках. Очевидно, що можна знайти багато картинок з подібним колірним наповненням, але іншого вмісту. Яскравим прикладом є сервіс [Multicolr](http://labs.tineye.com/multicolr/) від компанії Tineye, що здійснює пошук у багато мільйонній базі зображень Flickr.

#### Використання фільтрів

Значне покращення якості пошуку принесло додавання різноманітних фільтрів в розширеному пошуку: домінуючі кольори, кліпарт, графіка, фотографія, портрет. В основі роботи цих фільтрів покладено аналіз властивостей зображення: розмір, колір, реалістична фотографія чи ілюстрація та інші. Кожен раз, коли користувач використовує певний фільтр, відбувається порівняння всіх знайдених зображень з абстрактною моделлю-еталоном, яка відповідає даному типу зображення. Таким чином з результатів пошуку видаляються зображення, що не мають властивостей цієї ідеальної моделі, або мають не в тій мірі.

### 12.3.2. Пошук схожих зображень за складними обчисленнями

#### Хеш-значення

Зображення зменшується до невеликого розміру і для нього обчислюється хеш-значення: число, яке вимагає для зберігання невелику кількість пам'яті фіксованої довжини. Це значення порівнюється з базою хешів зображень. Якщо два зображення відрізняються хоч на один піксель, то вони мають різні хеші. Якщо хеші однакові - то зображення з певною ймовірністю можуть бути визнані схожими.

Плюс цього підходу в тому, що для зберігання і порівняння хешів зображень потрібно значно менше пам'яті і часу, ніж для зберігання і порівняння самих зображень. Але як недолік те, що для двох різних зображень міг порахуватися однаковий хеш. В результаті, пошук видавав зображення, що відрізнялися від шуканого. Такий підхід не надавав в результатах зображень в які внесено зміни: повороти, відзеркалення, зміна розміру, модифікації кольорів.

#### Візуальні слова

Спочатку на зображенні виявляють потенційно значимі місця - тобто ймовірні об'єкти чи їх межі. Це можна зробити декількома способами. Наприклад, дія алгоритму Difference of Gaussians (DoG, різниця за Гаусом) полягає в наступному: вихідну картинку кілька разів піддають розмиванню по Гаусу, кожен раз використовуючи різний радіус розмивання. Далі результати порівнюють між собою. Цей спосіб дозволяє виявити на зображенні найбільш контрастні фрагменти - наприклад, яскраві плями або злами ліній.

Після того як значущі місця знайдені, їх описують в числах. Запис фрагмента картинки (візуального слова) в числовому вигляді називається дескриптором. За допомогою дескрипторів можна швидко, повно і точно порівняти фрагменти зображення, не використовуючи самі фрагменти.

Оскільки дескриптор - це числовий опис даних, то порівняння зображень зводиться до порівняння чисел. Дескриптори виражені досить великими числами, тому їх порівняння може вимагати помітних обчислювальних ресурсів. Для пришвидшення обчислень, дескриптори розподіляють по групах або кластерах. В один кластер потрапляють схожі дескриптори з різних зображень.

Після кластеризації даний дескриптор зображення можна не розглядати; важливим стає лише номер кластера з дескрипторами, найбільш схожими на даний. Перехід від дескриптора до номера кластера називається квантуванням, а сам номер кластера - квантованим дескриптором. Квантування істотно скорочує обсяг даних, які потрібно обробити комп'ютеру.

Спираючись на квантовані дескриптори, комп'ютер виконує такі завдання, як розпізнавання об'єктів і порівняння зображень (рис.12.9).



Рис.12.9. Пошук зображення за візуальними словами

У випадку з розпізнаванням квантовані дескриптори використовуються для навчання класифікатора - алгоритму, який відокремлює зображення «з котиком» від зображень «без котика». У випадку з порівнянням картинок комп'ютер зіставляє набори квантових дескрипторів з різних зображень і робить висновок про те, наскільки схожі ці зображення або їх окремі фрагменти. Таке порівняння лежить в основі пошуку дублікатів і пошуку за завантаженим зображенням.

Такий підхід дозволяв шукати копії картинки або її варіанти: наприклад, ту ж картинку іншого розміру або без підписів. Однак вже тоді пошук показував перші здатності до узагальнення - і в деяких випадках знаходив не просто копії завантаженої картинки, а інші зображення, які містять такий же об'єкт.

#### Метрика подібності

Головною проблемою стало визначити метрику схожості зображень, оскільки вона може відрізнятися від завдання до завдання. Наприклад, потрібно знаходити вихідне зображення за кількістю об’єктів, а може за колірною гамою чи за схожими формами об'єктів? А чи це будуть схожі фотографії однієї і тієї ж собаки в різних ракурсах або схожі фотографії різних собак, але в одному ракурсі (рис.12.10).



Рис.12.10. Пошук зображення за метрикою близькості

Алгоритми пошуку за метрикою подібності є швидкими, знаходять дублікати з незначними спотвореннями. Але часто трапляються результати з зовсім іншими об’єктами, хоча зовні і доволі подібними.

#### Метаінформація

В мережі більшість зображень лежать не просто так - вони організовані або в фотоальбоми з назвами і коментарями, або є ілюстраціями до деяких текстів, або мають детальні підписи і навіть готові теги. Повноцінна система пошуку, зрозуміло, повинна використати максимум з цієї інформації, щоб тематично структурувати базу зображень.

#### Додаткові критерії

Цей підхід пропонує визначення категорії в наявних фільтрах, розширення візуального запиту текстовим уточненням та інші комбінації методів.

Сучасні пошукові системи навчилися добре розуміти сенс текстових запитів, сформульовані у вільній формі, наприклад «світлий породистий котик» або навіть «портрети Віталія Кличка в молодості». Запити такого типу несуть певний зміст і у поєднанні з пошуковою історією та персоналізацією є вдалим доповненням для кінцевого відбору зображень. Разом із завантаженою фотографією можна додати текст, що буде сприйматися як додаткове уточнення.

Можливе поступове уточнення запиту, коли користувач в процесі роботи системи пошуку позначає проміжні результати як «відповідні» або «незадовільні», і система продовжує працювати з уточненими запитом.

## 12.3. Застосування нейромереж в комп’ютерному зорі

Використання нейронних мереж для розпізнавання зображень має довгу історію розвитку. Однак прорив у галузі розпізнавання зображень з використанням нейронних мереж стався на початку 1990-х років з розробкою згорткової нейронної мережі (Convolutional Neural Network, CNN). Ця архітектура нейронної мережі була розроблена для імітації роботи зорової кори головного мозку і показала вражаючі результати задач розпізнавання зображень.

Одним з ключових моментів історії використання нейронних мереж для розпізнавання зображень був конкурс ImageNet Large Scale Visual Recognition Challenge (ILSVRC), який започатковано в 2010 році.

ILSVRC проводився щорічно з 2010 до 2017 року і став важливим мірилом прогресу в галузі комп'ютерного зору. Він включав кілька завдань, включаючи класифікацію зображень, виявлення об'єктів та локалізацію. У рамках змагання учасники навчали свої моделі на тренувальних даних ImageNet, а потім оцінювали їх на тестових даних, які не були доступні до моменту оцінки. ILSVRC суттєво сприяв розвитку згорткових нейронних мереж та привів до значного покращення результатів у завданнях розпізнавання зображень.

У ILSVRC учасники змагалися у задачі класифікації зображень на величезному наборі даних ImageNet, що містить мільйони зображень. У 2012 році команда дослідників з університету Стенфорда, на чолі з Джеффрі Хінтоном, представила нейронну мережу під назвою AlexNet, яка перевершила решту учасників з величезним відривом. AlexNet демонструвала високу точність у розпізнаванні та класифікації зображень та стала переломним моментом в історії комп'ютерного зору.

З того часу нейронні мережі, особливо згорткові нейронні мережі, стали домінуючим підходом у галузі розпізнавання зображень. Дослідники та інженери продовжують розробляти нові архітектури нейронних мереж, покращуючи точність та ефективність розпізнавання зображень. Прогрес у галузі нейронних мереж для розпізнавання зображень просувається також завдяки доступності великих обсягів даних та обчислювальних потужностей, а також вдосконалення алгоритмів навчання та методів оптимізації.

### Набори даних для навчання

**Набір даних ImageNet** (https://www.image-net.org/) є одним із найбільш популярних наборів даних для навчання та оцінки моделей комп'ютерного зору. Його створено у 2009 році і з того часу став одним із стандартних наборів даних в академічній та промисловій галузях комп'ютерного зору.

Основна частина набору даних ImageNet (https://www.kaggle.com/datasets) містить понад 14 мільйонів розмічених зображень, розділених на більш ніж 20 000 категорій. Спочатку набір даних ImageNet використовували для класифікації зображень, де моделі повинні визначити, якого класу належить дане зображення.

Важливою властивістю набору даних ImageNet є його різноманітність. Він містить зображення різних об'єктів, тварин, предметів та сцен, що дозволяє моделям навчатися на різноманітних ознаках та узагальнювати свої знання на нові зображення, які вони раніше не бачили.

У конкурсах ILSVRC використовується підвибірка з ImageNet розміром 1000 зображень в кожній з 1000 категорій. Таким чином, навчальна вибірка складається з приблизно 1.2 мільйонів зображень, перевірочна - 50000 зображень, тестова - 150000 зображень. Оскільки ImageNet складається з зображень різного розміру, то їх приведено до єдиного розміру 256х256. Якщо зображення є прямокутної форми, то воно масштабується і з нього вирізається центральна частина розміром 256х256.

У сфері комп'ютерного зору існує багато наборів даних, які широко використовуються для навчання та оцінки моделей. Велику колекцію датасетів наведено на ресурсі <https://paperswithcode.com/datasets?mod=images>

Кожен з датасетів має свої особливості, класи об'єктів і завдання, котрим вони найбільше підходять. Вибір конкретного набору даних залежить від завдання та доступних ресурсів для навчання моделі.

### 12.3.1. Класифікація зображень

На якісно іншому рівні завдання класифікації зображень почали вирішувати з 2013 року. Тоді на наборі даних ImageNet перетнули бар'єр в 15% помилок класифікації тисячі видів об'єктів. З того часу спроектовано і навчено дуже багато різних моделей нейромереж, і подолано бар'єр в 5% помилок. Більшість успішних моделей побудовано на основі згорткових нейромереж.

Одною з успішних моделей, що виграла змагання [ILSVRC-2014](https://neurohive.io/en/tag/segmentation/) є нейромережа VGG16. Нижче наведено загальний опис її архітектури (рис.12.11).

Детальний розгляд нейромережа VGG16 можна подивитися у відео уроці «Приклади архітектур згорткових мереж VGG-16 та VGG-19»  
<https://www.youtube.com/watch?v=IYxU985meTM>

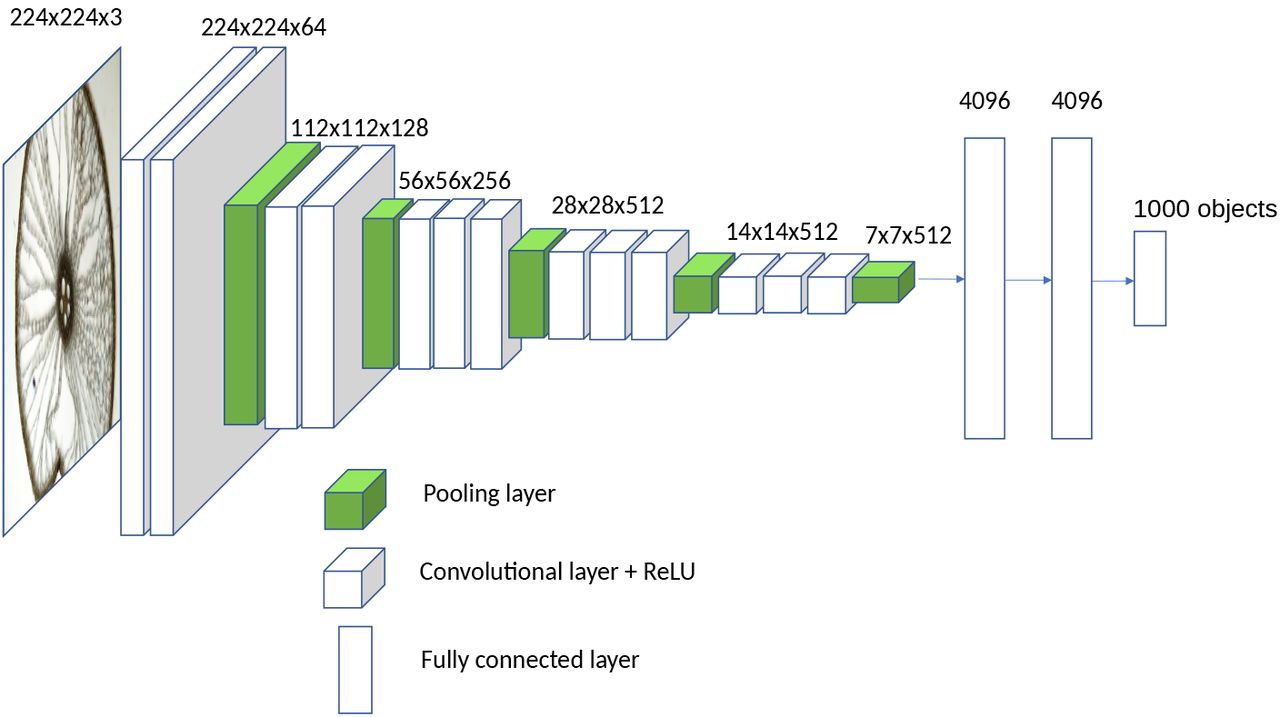


Рис.12.11. Нейронна згорткова мережа VGG16

Нейронна мережа VGG16 (VGG – Visual Geometry Group, 16 – кількість шарів) є моделлю глибокого навчання, яка використовується для вирішення завдань класифікації зображень. VGG16 використовує згорткові шари для отримання ознак зображень і пов'язані шари для класифікації об'єктів. Вона навчається на великій кількості зображень і може бути використана для класифікації нових зображень за заданими класами.

Порядок обробки зображення

* **Вхідні зображення.** На вхід подаються зображення, які можуть бути фотографіями, наприклад собаки або кішки.
* **Згорткові шари для витягування ознак.** VGG16 містить кілька згоркових шарів, які виконують операцію згортки з фільтрами, щоб витягти локальні просторові ознаки зображення. Згорткові шари мають різні розміри фільтрів та кількість фільтрів, щоб захоплювати різні характеристики зображення. На перших шарах фільтри згортки можуть виявляти прості ознаки, такі як межі, кути, форми, текстури або кольори на різних рівнях деталізації. На більш глибоких шарах фільтри можуть виявляти складні форми і об'єкти. Виявлені ознаки представлені у вигляді карт ознак, які містять інформацію про наявність цих ознак у різних частинах зображення. Кожен шар створює абстрактні представлення того, що знаходиться на зображенні.
* **Пулінг для зменшення розмірності.** Пулінгові шари виконують операцію пулінгу над картами ознак, щоб зменшити їх розмірність, зберігаючи важливі характеристики. Це дозволяє мережі знизити кількість параметрів та обчислень, а також досягти інваріантності до масштабування та невеликих трансляцій об'єктів на зображенні. Найпоширеніша операція пулінгу - це операція максимального пулінгу (max pooling), яка поділяє карту ознак на області, що не перекриваються, і вибирає максимальне значення в кожній області. Результатом операції пулінгу є зменшена у розмірі карта ознак з посиленими та найбільш значущими ознаками.
* **Повнозв'язані шари для класифікації та розпізнавання.** Після обробки згортковими та пулінговими шарами, інформація передається у пов'язані шари. Повноз'язані шари виконують класифікацію та розпізнавання об'єктів. У випадку VGG16 останній повнопов'язаний шар має 1000 вихідних вузлів, що відповідають 1000 класам, на які модель навчена., наприклад, чи є зображення собакою чи кішкою.
* **Вихід.** Ближче до виходу моделі утворюється невелике зображення - карта ознак розміру 7x7 пікселів з 512 фільтрами. З цієї тривимірної карті все ще неможливо зробити передбачення класів об'єктів на зображенні - котик або собака. Для того, щоб перейти вже до передбачення класів, цю карту укладають на площині і з'єднують з повнозв'язним прихованим шаром з 4096 нейронами. Повнозв'язані шари приймають абстрактні представлення (високорівневі ознаки) з карт ознак і перетворюють їх у ймовірності приналежності до різних класів об'єктів. А далі класична схема: ще один прихований шар з 4096 нейронами і вихідний шар від 1000 нейронами, кожен з яких видає ймовірність приналежності до одного з 1000 класів в наборі ImageNet. Зрештою, VGG16 видає ймовірності різних класів, такі як "собака" або "кішка". Найімовірніший клас вважається відповіддю моделі.
* **Навчання VGG16** відбувається шляхом подачі великої кількості розмічених зображень на вхід моделі та коригування ваг під час зворотного поширення похибки. Це дозволяє моделі налаштовувати свої параметри, щоб краще розпізнавати та класифікувати об'єкти на зображеннях.

Основними компонентами згорткових нейронних мереж є згорткові та пулінгові шари. Вони виконують важливу роль в аналізі та обробці зображень. Зазвичай, згорткові та пулінгові шари чергуються в згорткових нейронних мережах. Згорткові шари витягують ознаки зображення на різних рівнях деталізації, а пулінгові шари зменшують розмірність і підсумовують найбільш значущі ознаки. Кожен наступний шар "товстіше", оскільки в ньому більше фільтрів, але "менше", оскільки зображення спеціально зменшують. Це дозволяє мережі знайти більш абстрактні та узагальнені ознаки в міру глибини мережі та забезпечує її здатність класифікувати та розпізнавати об'єкти на зображеннях.

Таким чином, згорткові та пулінгові шари в CNN дозволяють мережі отримувати ієрархічні ознаки з зображень, починаючи з низькорівневих деталей і прогресуючи до абстрактніших представлень. Це дозволяє мережі виявляти та враховувати різні характеристики об'єктів та робити більш точні передбачення про їхній клас або ідентифікацію на зображенні.

Після навчання модель може використовуватись для класифікації нових зображень. Вона приймає вхідне зображення, пропускає його через згорткові та повнозв'язані шари і видає ймовірності класів для цього зображення.

#### Використання моделі

Як виявилося, мережі, що навчені на даних ImageNet можна використовувати для інших завдань комп'ютерного зору. Наприклад, для завдання відрізняти кішку від собаки не потрібно створювати окрему модель CatDog, шукати мільйони картинок і навчати з нуля мережу. Можна взяти навчену модель, забрати останні повнозв'язні шари нейронів, що відповідають за фінальну класифікацію, залишаючи тільки внутрішні 4096 нейронів, які з’єднані з 2 нейронами для виходу «котик-собака». Потрібно буде лише довчити модель цими 2\*4096 зв'язками, що робиться легко і швидко.

Який фізичний зміст в тому, що забирається останній шар мережі і підставляється новий? Виявляється, що всі шари згортки всередині себе набули здатність "розуміти" зображення. А оскільки навчання відбувалося на тисячах різних класів, то і узагальнююча здатність цих шарів досить сильна. В результаті зовнішні 4096 нейронів насправді видають вектор характеристик (ознак) будь-якого зображення в цілому.

Тому, для того, щоб здійснити класифікацію, що відмінна від початкової, потрібно лише довчити нейромережу і перевести цей 4096-мірний вектор в потрібний вектор передбачень приналежності до класів, не змінюючи існуючу глибоку згорткову мережу.

### 12.3.2. Пошук схожих зображень

Пошук схожих зображень з використанням згорткових нейронних мереж зазвичай виконується за допомогою витягування ознак з використанням попередньо навченої моделі.

Замість того, щоб навчати власну модель на великому наборі даних, можна скористатися заздалегідь навченою моделлю, яка вже навчена на величезному наборі зображень, такому як ImageNet. Попередньо навчена модель містить ваги і знання про різні ознаки, які можна використовувати для витягування ознак з нових зображень.

Зображення, для якого потрібно знайти схожі, пропускається через заздалегідь навчену модель. У цьому процесі отримують ознаки, що представляють високорівневі абстракції та характеристики зображення. Зазвичай використовуються виходи зі згорткових шарів або пов'язаних шарів, що передують фінальному класифікаційному шару.

Отримані ознаки перетворюються на векторне представлення. Це можна зробити, наприклад, шляхом застосування методу глобального пулінгу (global pooling) до карт ознак або шляхом використання методів стиснення розмірності. В результаті кожне зображення подається у вигляді вектору з високорівневими ознаками.

Після отримання векторного представлення застосовують різні метрики відстані, такі як евклідова відстань або косинусна відстань, для порівняння векторів наявних в базі зображень між собою. Чим менша відстань між двома векторами, тим схожими вважаються відповідні зображення.

На підставі відстані між векторами можна вибрати найбільш схожі зображення. Можна встановити граничне значення для відстані та вибрати всі зображення, відстань між векторами яких нижче за цей поріг. Також можна використовувати методи ранжирування, такі як пошук найближчих сусідів (nearest neighbors), щоб знайти найближчі зображення.

Це загальний підхід до пошуку схожих зображень з використанням згорткових мереж. Він ґрунтується на витягуванні ознак за допомогою попередньо навчених моделей та порівнянні векторних представлень зображень з використанням метрик відстані.

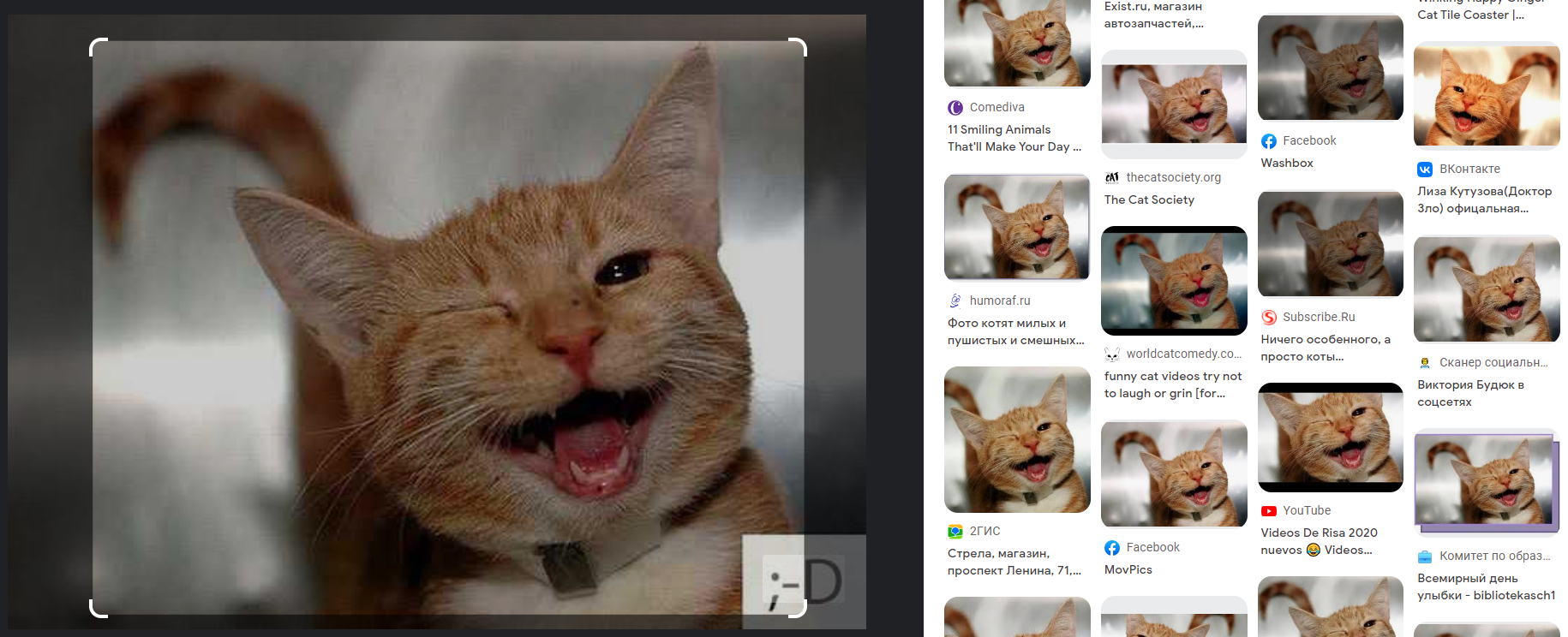


Рис.12.13. Результат пошуку схожих зображень

### 12.3.3. Виявлення (детекція) об’єктів на зображенні

Нейромережний детектор визначення типів об'єктів забезпечує однозначну ідентифікацію в кадрі різних типів об'єктів: людина, автомобіль, тварина, будівля. Результатом роботи детекції є, як правило, обведені прямокутниками об'єкти з описом (рис.12.14).

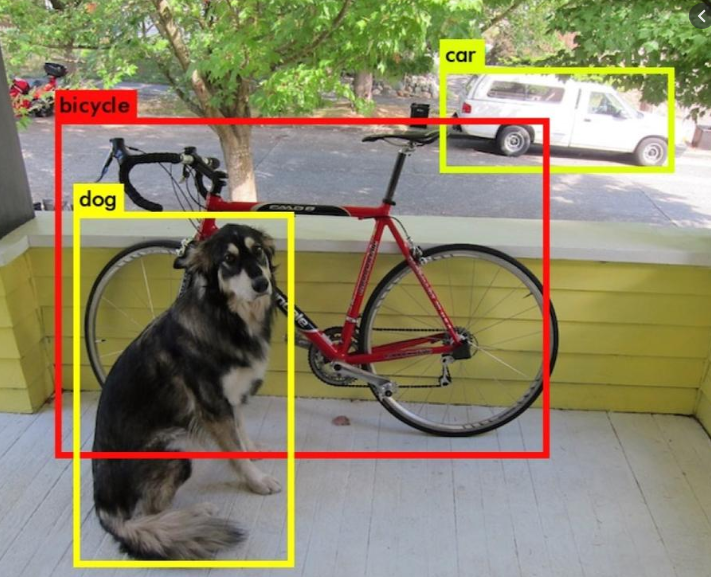


Рис.12.14. Виявлення нейромережею об’єктів на зображенні

Пошук та детекція об'єктів у зображенні за допомогою згорткових нейронних мереж (CNN) відбувається наступним чином.

* CNN аналізує зображення та знаходить у ньому різні особливості, такі як форми, текстури та контури об'єктів. CNN пропонує можливі розташування об'єктів, тобто регіони, де ці об'єкти можуть знаходитися.
* Кожен запропонований регіон перетворюється, щоб його розмір став фіксованим. Таким чином, всі регіони мають однаковий формат. Для кожного перетвореного регіону CNN визначає, до якого класу об'єктів він належить, та визначає межі об'єкта, створюючи обмежуючий прямокутник навколо об'єкта.
* Оскільки один і той же об'єкт може бути виявлений кілька разів у різних регіонах, застосовується метод, який видаляє повторювані детекції та залишає лише найбільш впевнені.

Для навчання детектора об'єктів використовуються розмічені дані, на яких модель навчається визначати об'єкти на зображеннях. Модель поступово покращується, щоб стати точнішою у виявленні об'єктів.

Таким чином, згорткові нейронні мережі в детекторах об'єктів аналізують зображення, знаходять об'єкти і створюють прямокутники навколо них. Це дозволяє автоматично виявляти та розпізнавати об'єкти на зображеннях з високою точністю. На точність розпізнавання типів об'єктів не впливають погодні умови, зміна часу доби, освітленість тощо.

Нейромережні детектори можна використовувати для запобігання порушень правил паркування, попередження про появу людини в небезпечній зоні та інших випадках, в яких достовірність визначення типу об'єкта грає істотну роль. Технології навчання нейромереж дозволяють створювати комплекси для ідентифікації будь-яких типів об'єктів.

### 12.3.4. Сегментація зображення

Сегментація - це об'єднання об'єктів в групи за загальними ознаками. Однокласова і багатокласова сегментація використовується, наприклад, для безпілотних автомобілів. Нейромережа розподіляє об'єкти по класах: машини, тротуар, будинки, люди. У всіх об'єктів є чіткі межі. Це робиться по одному знімку без жодної додаткової інформації. Алгоритми враховують не лише зовнішній вигляд об'єктів, а й їх взаємне розташування, що значно ближче до людського мислення, ніж формальний пошук зовні схожого об’єкта (рис.12.15).

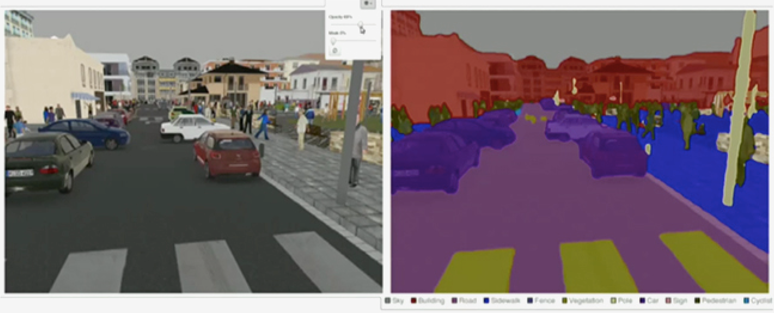


Рис.12.15. Сегментація зображення нейромережею

Як можна бачити на рисунку, сегментація має і недоліки. Наприклад, всі автомобілі відзначені однією великою плямою. Цього достатньо, щоб не врізатися, але може бути мало, щоб, наприклад, припаркуватися. Тому, існують спеціальні комбіновані підходи - на виході таких алгоритмів є і сегментація, і опис прямокутників (рис.12.16):

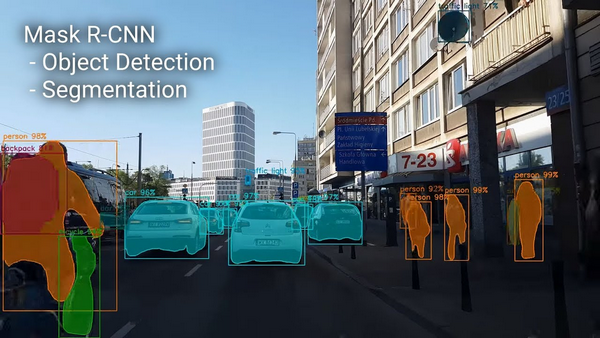


Рис.12.16. Сегментація та розпізнавання об’єктів зображення

## 12.4. Популярні сфери застосування комп’ютерного зору

* **Медицина.** Комп'ютерний зір використовується для аналізу медичних зображень, таких як рентгенівські знімки, МРТ, КТ та гістологічних зрізів. Це виявлення та класифікація пухлин, діагностика захворювань та автоматична сегментація органів.
* **Промисловість**. У промисловості комп'ютерний зір застосовується для контролю якості, інспекції та розпізнавання об'єктів. Це перевірка деталей, розмірів, форми та кольору виробів на виробничій лінії, визначення дефектів та відхилень, а також для сортування та класифікації продукції.
* **Транспорт та автомобільна промисловість.** Комп'ютерний зір використовується для автоматичного розпізнавання номерних знаків, системи допомоги при керуванні автомобілем, детектування об'єктів на дорозі та аналіз відео з супутників.
* **Безпілотні системи.** В області безпілотних систем, таких як безпілотні автомобілі та дрони, комп'ютерний зір застосовується для навігації, виявлення та уникнення перешкод, розпізнавання дорожніх знаків та сигналів, а також для аналізу навколишнього середовища.
* **Безпека та відеоспостереження.** Відеоспостереження, що засноване на комп'ютерному зорі, використовується для виявлення та розпізнавання осіб, аналізу поведінки, відеоаналітики та пошуку об'єктів на записах з камер відеоспостереження. Може використовуватись у громадських місцях, аеропортах, банках та інших об'єктах.
* **Робототехніка.** Комп'ютерний зір є невід'ємною частиною робототехніки. Роботи використовують відеокамери та системи комп'ютерного зору для сприйняття навколишнього середовища, навігації, виявлення об'єктів, розпізнавання жестів та взаємодії з людьми. Роботи-помічники здатні розпізнавати об'єкти, людей та довкілля, виконувати завдання, такі як прибирання, доставка та обслуговування, а також взаємодіяти з людьми.
* **Розпізнавання обличь та біометрія.** Комп'ютерний зір застосовується для розпізнавання обличь та інших біометричних даних. Це може використовуватися в системах аутентифікації, контролю доступу та інших додатках, де потрібна ідентифікація особи.
* **Роздрібна торгівля.** У роздрібній торгівлі комп'ютерний зір використовується для автоматизації процесів, таких як визначення та класифікація товарів на полицях, відстеження інвентарю, аналіз поведінки покупців та створення систем автоматичної оплати.
* **Агротехнології**. У сільському господарстві комп'ютерний зір застосовується для моніторингу рослин, виявлення хвороб та шкідників, визначення зрілості плодів, автоматизації збирання врожаю та оцінки якості сільськогосподарської продукції.
* **Реклама та маркетинг**. Комп'ютерний зір використовується в рекламі та маркетингу для аналізу зображень та відео, розпізнавання логотипів та брендів, визначення емоцій та реакцій споживачів, персоналізації рекламних повідомлень та створення віртуальних примірювальних.
* **Графічна обробка та доповнена реальність.** Комп'ютерний зір використовується для обробки та аналізу зображень та відео, створення спецефектів, трекінгу руху, створення віртуальних об'єктів та візуалізації доповненої реальності.

## Висновки

Комп’ютерний зір - швидко зростаюча область цифрових технологій, яка втілюється в багато сфер життя людей. Вже існує багато реалізацій технічного зору і його можливості розширюються щодня. Глибоке навчання, хмарні обчислення, швидкісні процесори і ефективні алгоритми сприяють появі нових сервісів, пристроїв та застосувань: розумні камери, складна робототехніка, поєднання реальності та віртуальності і багато іншого.

#### Можна виділити основні тенденції розвитку технічного зору:

* **Підвищення точності та продуктивності.** З розвитком алгоритмів технічного навчання та глибокого навчання, точність систем комп'ютерного зору значно зросла. Моделі стали більш точними та здатними розпізнавати об'єкти з високим ступенем точності. Також спостерігається покращення продуктивності систем технічного зору, що дозволяє обробляти великі обсяги даних у реальному часі.
* **Мобільний зір.** З розвитком мобільних пристроїв комп'ютерний зір стає більш доступним та інтегрованим у повсякденне життя. Мобільні програми з функціями технічного зору використовуються для розпізнавання обличь, сцен, штрих-кодів, а також доповненої реальності.
* **Об'єднання з іншими технологіями.** Технічний зір все частіше поєднується з іншими технологіями, такими як давачі, робототехніка, голосові інтерфейси та інтернет речей. Це дозволяє створювати інтегровані системи, які можуть сприймати довкілля та приймати рішення на основі виявлених даних.
* **Розширення сфери застосування.** Технічний зір знаходить все більше застосування в різних галузях, таких як охорона здоров'я, автомобільна промисловість, сільське господарство, роздрібна торгівля та промислове виробництво. Це з збільшенням обсягом даних, доступністю обчислювальних ресурсів і розробкою спеціалізованих моделей і алгоритмів для конкретних завдань.
* **Розпізнавання та аналіз відео.** Зростає інтерес до розробки систем комп'ютерного зору, здатних аналізувати відео. Це включає розпізнавання та класифікацію об'єктів у відеопотоці, відстеження руху, аналіз сцен та виявлення подій. Аналіз відео має широкий спектр застосувань у галузі відеоспостереження, робототехніки, автоматичного водіння та розважальної індустрії.
* **Обробка зображень у реальному часі.** Системи технічного зору дедалі більше орієнтовані обробку зображень у часі. Це вимагає оптимізації алгоритмів та архітектур моделей для досягнення високої швидкості обробки та відгуку. Обробка зображень у реальному часі має важливе значення для таких завдань, як автоматичне водіння, робототехніка та системи безпеки.

Це лише деякі з тенденцій, що спостерігаються, у розвитку технічного зору. З постійним розвитком алгоритмів і технологій очікується, що в майбутньому технічний зір продовжуватиме покращуватись і знаходити нові сфери застосування.

## Контрольні питання

1. Коротко перелічити особливості, що притаманні людському зору у сприйнятті реальності.
2. Пояснити різницю між поняттями комп’ютерний та машинний зір.
3. Перелічити популярні застосування машинного зору.
4. Назвіть різницю у сприйнятті зображення людиною і комп’ютером.
5. Які завдання покладаються на високорівневий та низькорівневий комп’ютерний зір?
6. Перелічити популярні завдання для комп'ютерного зору.
7. Яким чином здійснюється пошук схожих зображень за простими ознаками?
8. Назвіть підходи для пошук схожих зображень за складними обчисленнями.
9. Яким чином навчають нейронні мережі для використання в комп’ютерному зорі?
10. Перелічіть основні завдання, що спроможні вирішити навчені нейронні мережі для комп’ютерного зору.

## Використані джерела

1. Машинний зір <https://habr.com/ru/post/350918/>
2. Історія машинного зору <https://meduza.io/feature/2019/03/30/chto-takoe-mashinnoe-zrenie-i-chem-ono-otlichaetsya-ot-chelovecheskogo-seychas-ob-yasnim-ponyatno>
3. Комп’ютерний зір: технології, ринок, перспективи <http://www.tadviser.ru/index.php/Статья:Компьютерное_зрение:_технологии,_рынок,_перспективы>
4. Базові методи розпізнавання образів <https://habr.com/ru/post/208090/>
5. Стенфордський курс лекцій «Згорткові нейронні мережі для візуального розпізнавання (Convolutional Neural Networks for Visual Recognition» <https://www.reg.ru/blog/stenfordskij-kurs-lekciya-1-vvedenie/>
6. Як комп’ютери навчилися добре розпізнавати зображення <https://habr.com/ru/post/455331/>
7. Обробка відео нейромережею <https://dx.media/articles/how-it-works/remastering-s-ii-kak-neyroseti-vozrozhdayut-klassiku/>
8. Генеративно-змагальна нейромережа (GAN) <https://neurohive.io/ru/osnovy-data-science/gan-rukovodstvo-dlja-novichkov/>