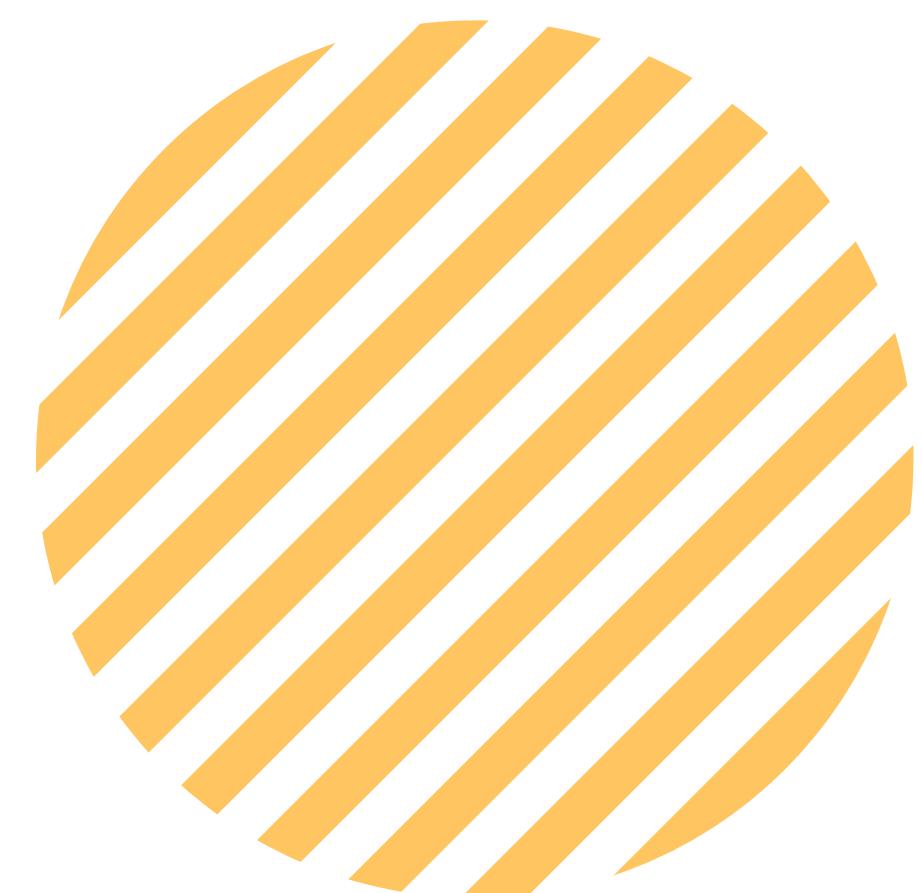


Аналіз сцен. Алгоритми обробки та генерації зображень

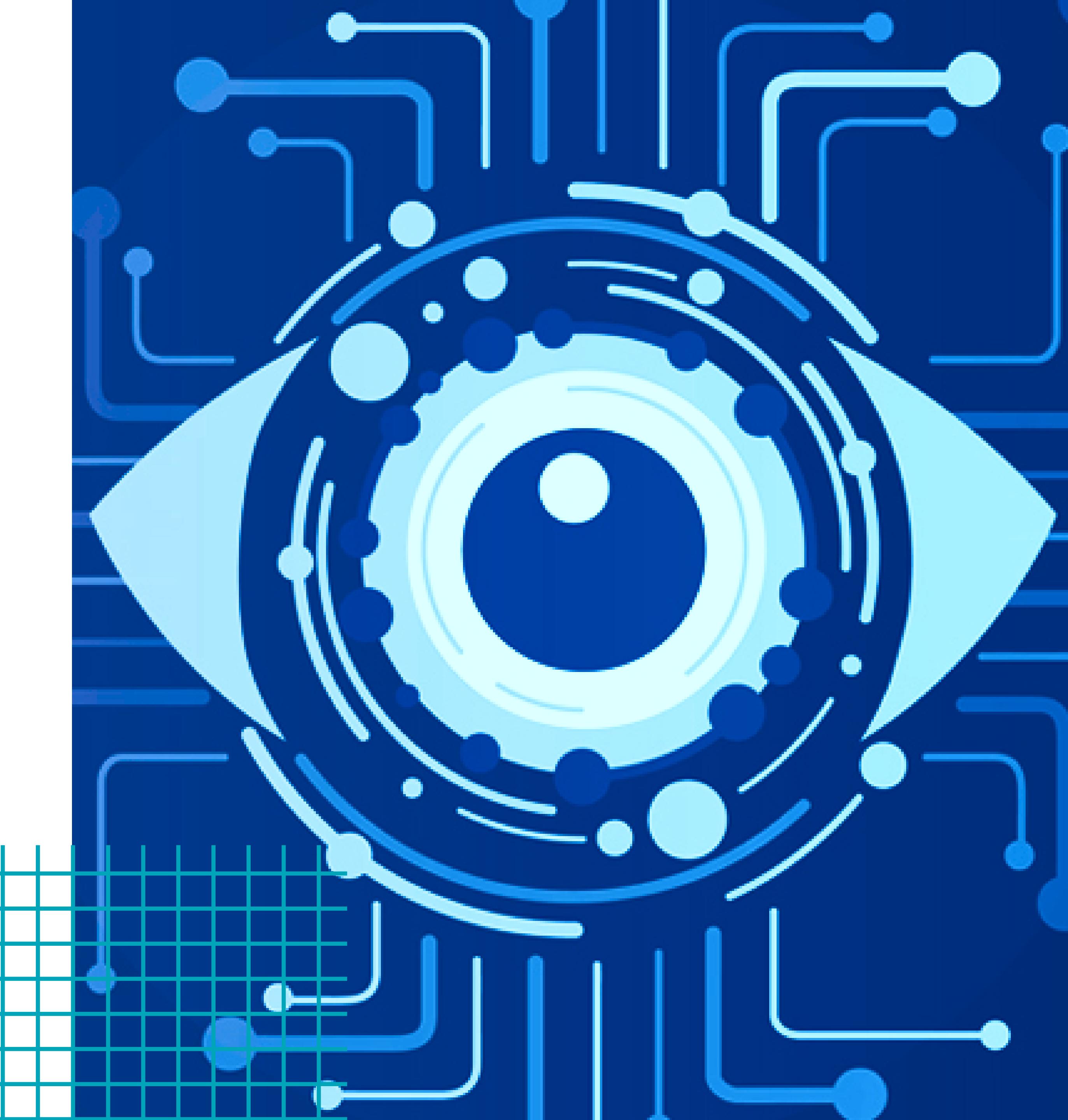
Студентки 4 курсу
Групи ДА-81 Желєзнової В.С.



Комп'ютерний зір (Computer Vision)

Комп'ютерний зір (Computer Vision, CV) - це область штучного інтелекту, пов'язана з аналізом зображень та відео. Вона включає набір методів, які наділяють комп'ютер здатністю «бачити» і витягувати інформацію з побаченого.

Системи складаються з фото- або відеокамери та спеціалізованого програмного забезпечення, яке ідентифікує та класифікує об'єкти. Вони здатні аналізувати образи (фотографії, картинки, відео, штрих-коди), а також обличчя та емоції.



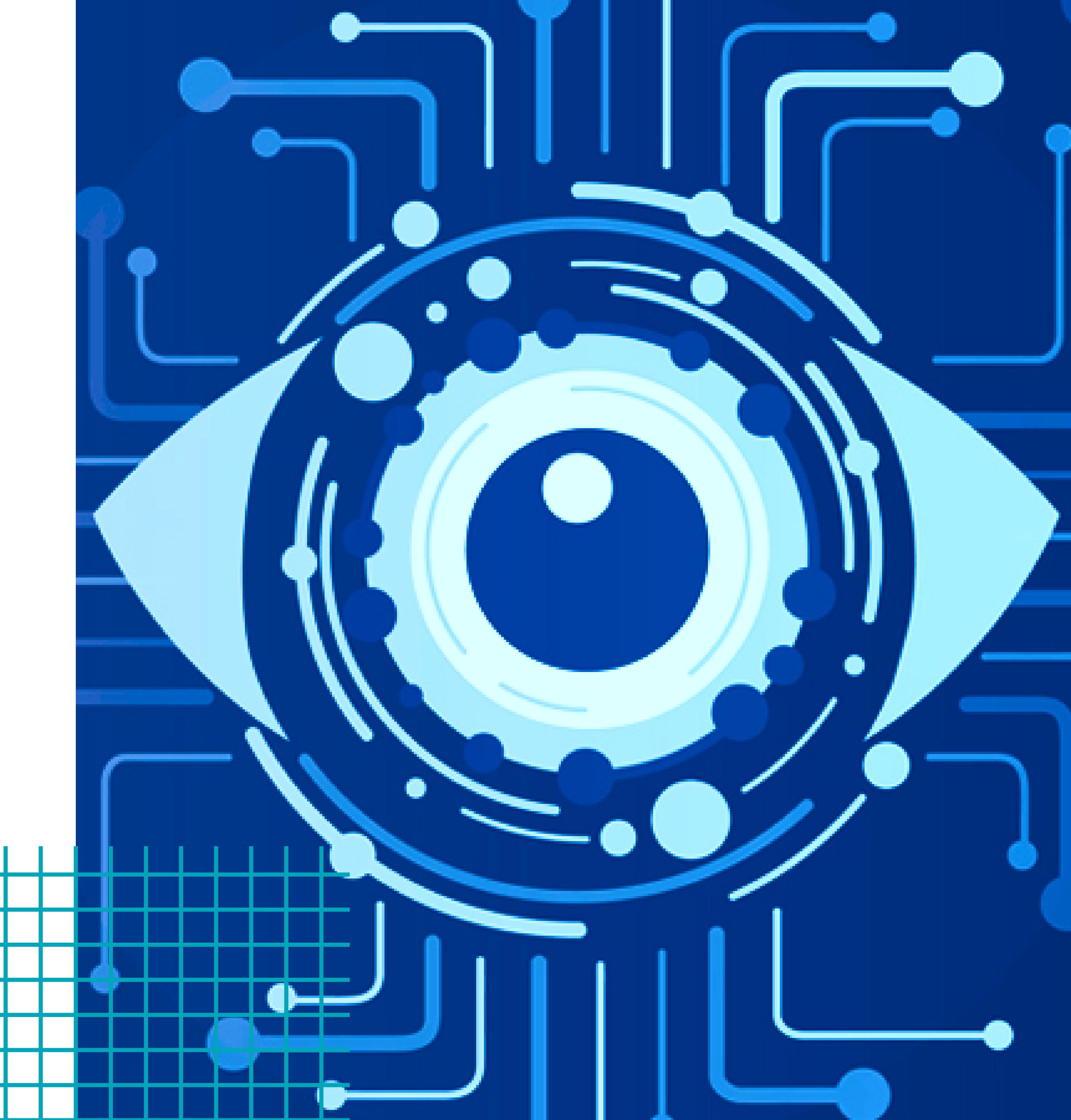
Приклади застосування комп'ютерного зору

Одним з найважливіших застосувань є обробка зображень у медицині. Ця область характеризується отриманням інформації з відеоданих для встановлення медичного діагнозу пацієнтам. У більшості випадків, відеодані отримують за допомогою мікроскопії, рентгенографії, ангіографії, ультразвукових досліджень та томографії. Прикладом інформації, яка може бути отримана з таких відео, є виявлення пухлин, атеросклерозу або інших злоякісних змін.

Військове застосування є, мабуть найбільшою областю комп'ютерного зору. Очевидними прикладами є виявлення ворожих солдатів і транспортних засобів та управління ракетами. Найбільш досконалі системи управління ракетами посилають ракету в задану область, замість конкретної мети, а селекція цілей проводиться, коли ракета досягає заданої області, виходячи з одержуваних відеоданих.

Типові завдання комп'ютерного зору

- Розпізнавання
- Рух
- Відновлення сцени
- Відновлення зображення



Розпізнавання

Класична задача в комп'ютерному зорі, обробці зображень і машинному зорі це визначення чи містять дані певний характерний об'єкт, особливість або активність. Це завдання може бути достовірно і легко вирішено людиною, але досі не вирішено задовільно у комп'ютерному зорі у випадку: випадкові об'єкти у випадкових ситуаціях.

Існуючі методи вирішення цього завдання ефективні лише для окремих об'єктів, таких як прості геометричні об'єкти (наприклад, багатогранники), людські особи, друковані або рукописні символи, автомобілі і лише в певних умовах, зазвичай це певне освітлення, фон та положення об'єкта щодо камери.

У літературі описано безліч проблем розпізнавання:

Розпізнавання: один або кілька попередньо заданих або вивчених об'єктів або класів об'єктів можуть бути розпізнані, зазвичай разом з двомірним положенням на зображені або тривимірним положенням в сцені.

Ідентифікація: розпізнається індивідуальний екземпляр об'єкта. Приклади: ідентифікація певної людини або відбитка пальців або автомобіля.

Виявлення: відеодані перевіряються на наявність певної умови.

Наприклад, виявлення можливих неправильних клітин або тканин у медичних зображеннях. Виявлення, засноване на відносно простих та швидких обчисленнях, іноді використовується для знаходження невеликих ділянок в аналізованому зображені, які потім аналізуються за допомогою прийомів, більш вимогливих до ресурсів, для отримання правильної інтерпретації.

Спеціалізовані задачі, засновані на розпізнаванні

Пошук зображень за змістом: знаходження всіх зображень у великому наборі зображень, які мають певний зміст. Зміст може бути визначений різними шляхами, наприклад, у термінах схожості з конкретним зображенням, або в термінах високорівневих критеріїв пошуку, що вводяться як текстові дані.

Оцінка: визначення положення або орієнтації певного об'єкта щодо камери. Прикладом застосування цієї техніки може бути сприяння руці робота у вийманні об'єктів зі стрічки конвеєра на лінії збирання.

Оптичне розпізнавання символів: розпізнавання символів на зображеннях друкованого або рукописного тексту, зазвичай для перекладу в текстовий формат, найбільш зручний для редагування або індексації (наприклад, ASCII).

Рух та відновлення сцени

Рух

Декілька завдань, пов'язаних з оцінкою руху, в яких послідовність зображень (відеодані) обробляються для знаходження оцінки швидкостіожної точки зображення або 3D сцени. Прикладами таких завдань є:

- Визначення тривимірного руху камери
- Спостереження, тобто слідування за переміщеннями об'єкта (наприклад, машин або людей)

Відновлення сцени

Дано два або більше зображень сцени, або відеодані. Відновлення сцени має завдання відтворити тривимірну модель сцени. У найпростішому випадку моделлю може бути набір точок тривимірного простору. Більш складні способи відтворюють повну тривимірну модель.

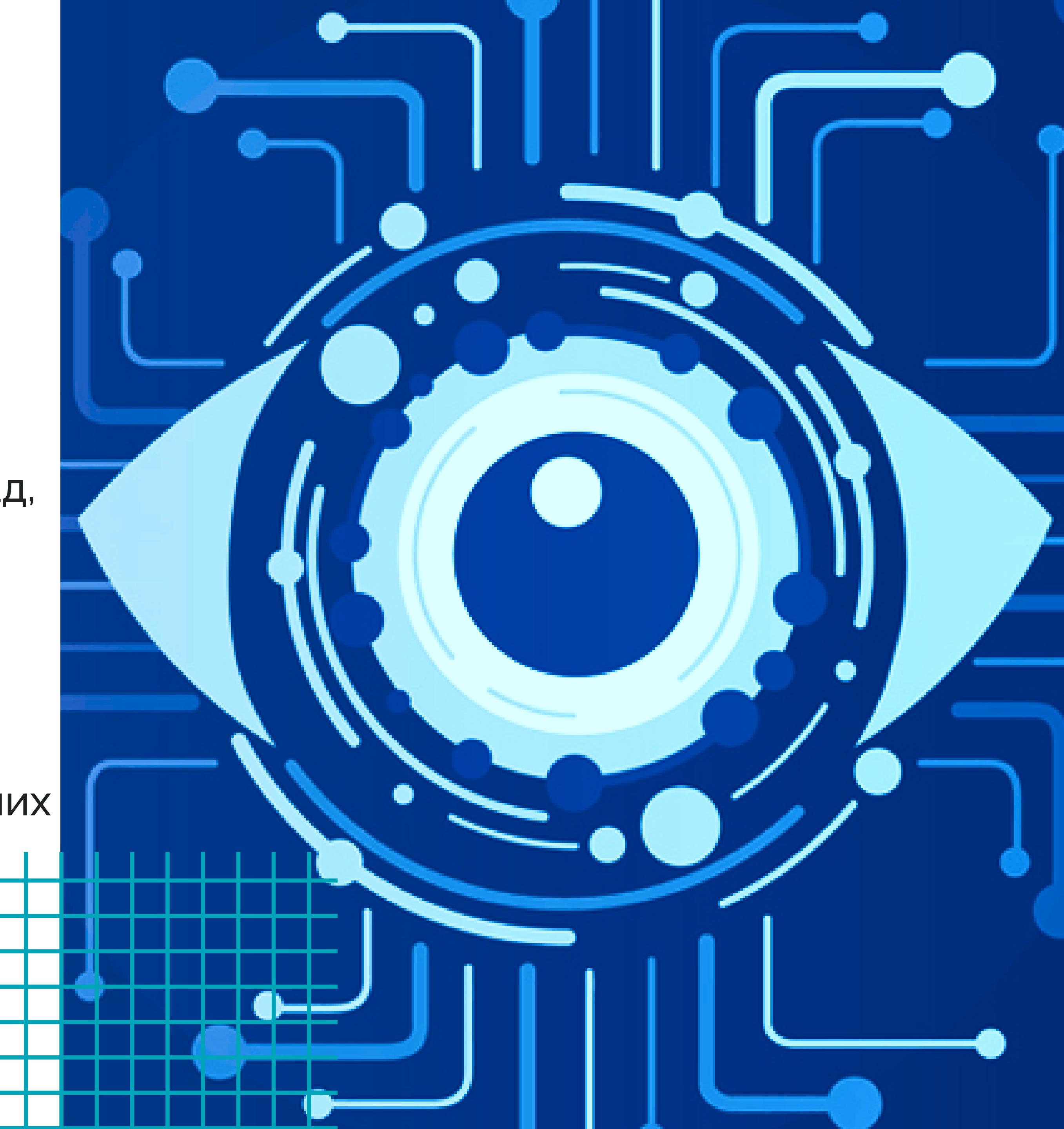
Відновлення зображення

Завдання відновлення зображень це видалення шуму (шум датчика, розмитість об'єкта, що рухається і т. д.). Найбільш простим підходом до вирішення цього завдання є різні типи фільтрів, таких як нижніх або середніх фільтри частот. Більш складні методи використовують уялення про те, як мають виглядати ті чи інші ділянки зображення, і основі цього їх зміна.

Вищий рівень видалення шумів досягається в ході початкового аналізу відеоданих на наявність різних структур, таких як лінії або межі, а потім управління процесом фільтрації на основі цих даних.

Обробка зображень

Обробка зображень – будь-яка форма обробки інформації, для якої вхідні дані представлені зображенням, наприклад фотографіями або відеокадрами. Обробка зображень може здійснюватися як для отримання зображення на виході (наприклад, підготовка до поліграфічного тиражування, телетрансляції і т. д.), так і для отримання іншої інформації (наприклад, розпізнання тексту, підрахунок числа і типу клітин в полі мікроскопа і т.д.). Крім статичних двомірних зображень, обробляти потрібно також зображення, що змінюються з часом, наприклад, відео.



Алгоритми обробки зображень

Алгоритм для видалення елементу із зображення та заміщення пропусків (англ. inpainting) - Belief Propagation.

Класичний алгоритм:

Нехай є ациклічний граф, в якому кожна вершина може приймати одно з k станів.

Відомі ймовірності $\Phi(X_k)$: для вершини x прийняти стан k та $\Psi(X_k; Y_q)$ - сумісний розподіл на x та y .

Максимізуємо функцію правдоподобності:

$$\max P = -\min [\log P] = \min_{x_1, \dots, x_n} \left[\sum_{i=1}^n \phi_i(x_i) + \sum_{\{k,j\} \in E} \psi_{j,k}(x_j, x_k) \right]$$

E - множина ребер графа
 P - правдоподібність
 X_i - стан вершини i

$\psi_{j,k}(x_j, x_k)$ - $(-\log)$ сумісного розподілу на дві вершини

$\phi_i(x_i)$ - $(-\log)$ ймовірностного розподілу на 2 вершини

Belief Propagation, формулювання для іnpainting

- Покриємо область мережею блоків, що перетинаються. Блоки - це вершини графа, перекриття задають ребра
- Норма перекриття двох блоків, або блоку і початкового наближення задає ймовірності
- Граф не ацикличний, але можна розраховувати на хороший локальний мінімум
- Алгоритм квадратично залежить від кількості станів, що унеможливлює використання алгоритму

Рішення:

- Priority message scheduling
- Dynamic pruning

Belief Propagation, методи оптимізації

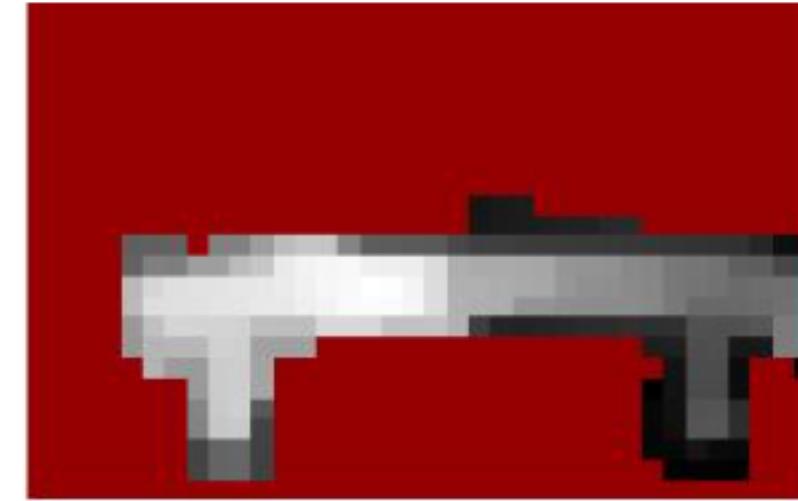
Ітеративно виконуватимемо алгоритм, як для ациклічного графа

- Dynamic pruning – зберігатимемо тільки найбільш ймовірні стани
- Priority message scheduling – виконуватимемо передачу повідомлень від вершин з найменшою кількістю станів.

При цьому такими вершинами виявляються блоки на границях об'єктів.

Їхня визначеність вища

Результаты



Originals

Restoration order
(Blacks are earlier)

Results



Originals

Restoration order
(Blacks are earlier)

Results

Веб-сервіси з функціоналом обробки зображень

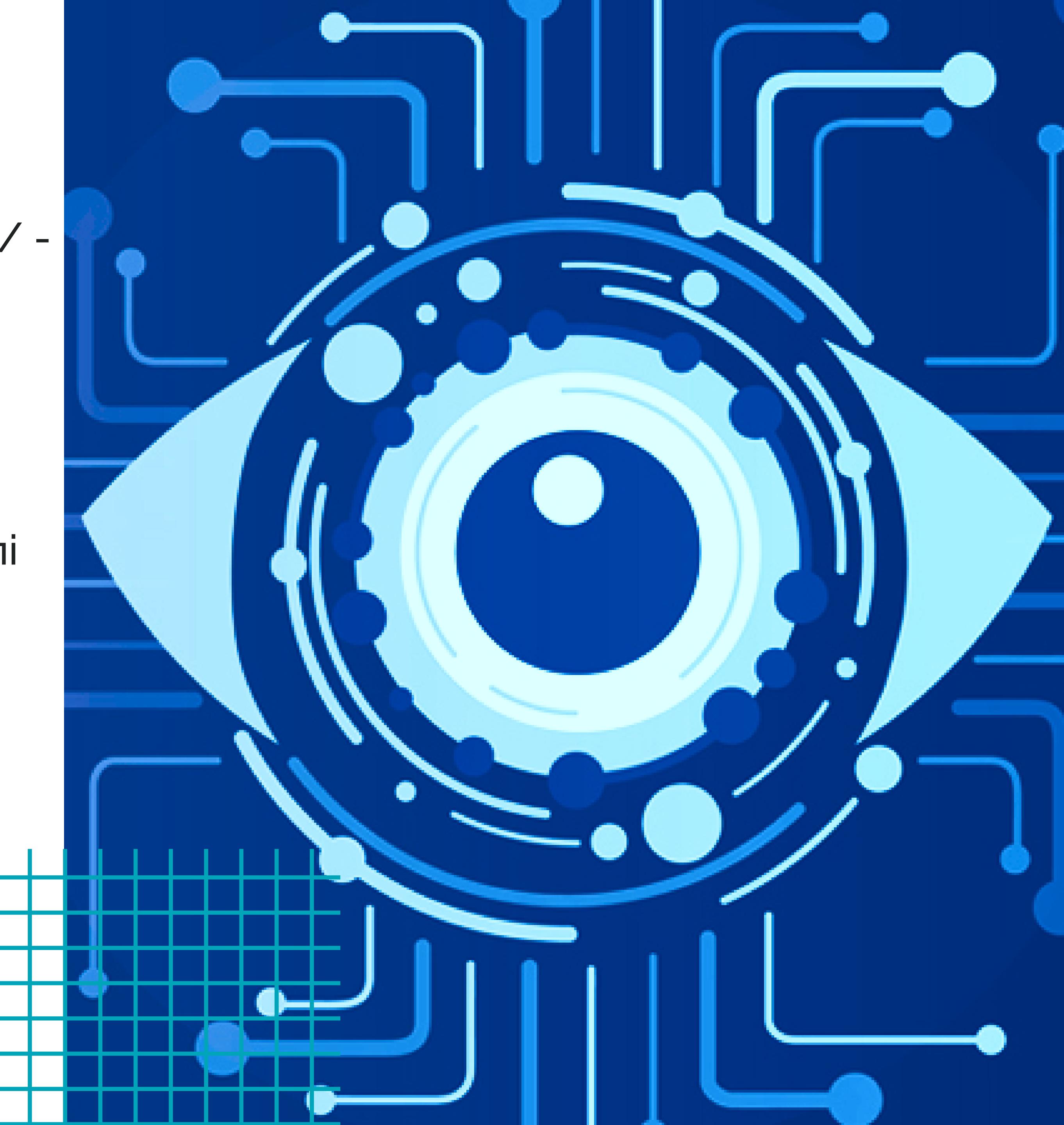
<https://www.nvidia.com/research/inpainting/> -

прибирає об'єкт із зображення, та заповняє отриману пустоту

<https://www.remove.bg/> - видаляє фон із зображення

<https://colorize.cc/> - розмальовує чорно-білі фотографії в реалістичні кольори

<https://cvl-demos.cs.nott.ac.uk/vrn/> - вміє робити 3D-моделі з одного зображення обличчя



Засоби ШІ генерації зображень

DALL-E – це програма штучного інтелекту, яка створює зображення з текстових описів, виявлених OpenAI 5 січня 2021 року. Вона використовує 12-мільярдну версію перетворювача GPT-3. Модель для інтерпретації вхідних даних природною мовою (наприклад, «зелений шкіряний гаманець у формі п'ятикутника» або «ізометричний вигляд сумної капібари») та створення відповідних зображень. Він може створювати зображення як реалістичних об'єктів («вітраж із зображенням блакитної полуниці»), так і об'єктів, що не існують насправді («куб з фактурою дикобраза»).

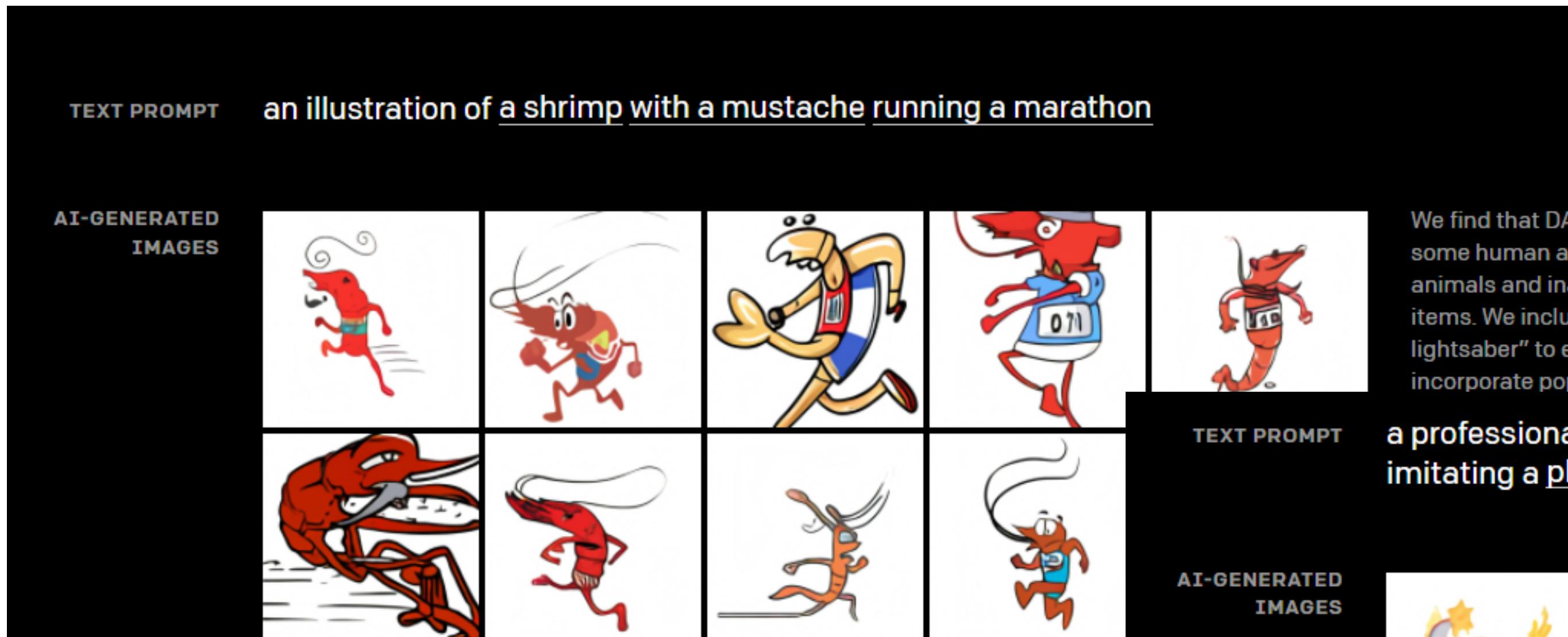
DALL-E був розроблений і оголошений публіці разом з CLIP (Contrastive Language-Image Pre-training) окремою моделлю, роль якої полягає в «розумінні та ранжируванні» її результатів. Зображення, які генерує DALL-E, куруються CLIP, який надає зображення найвищої якості для будь-якого запрошення. OpenAI відмовився випустити вихідний код для будь-якої моделі; керована демонстрація DALL-E доступна на веб-сайті OpenAI, де можна переглянути вихідні дані з обмеженого набору прикладів підказок.

Засоби ШІ генерації зображень

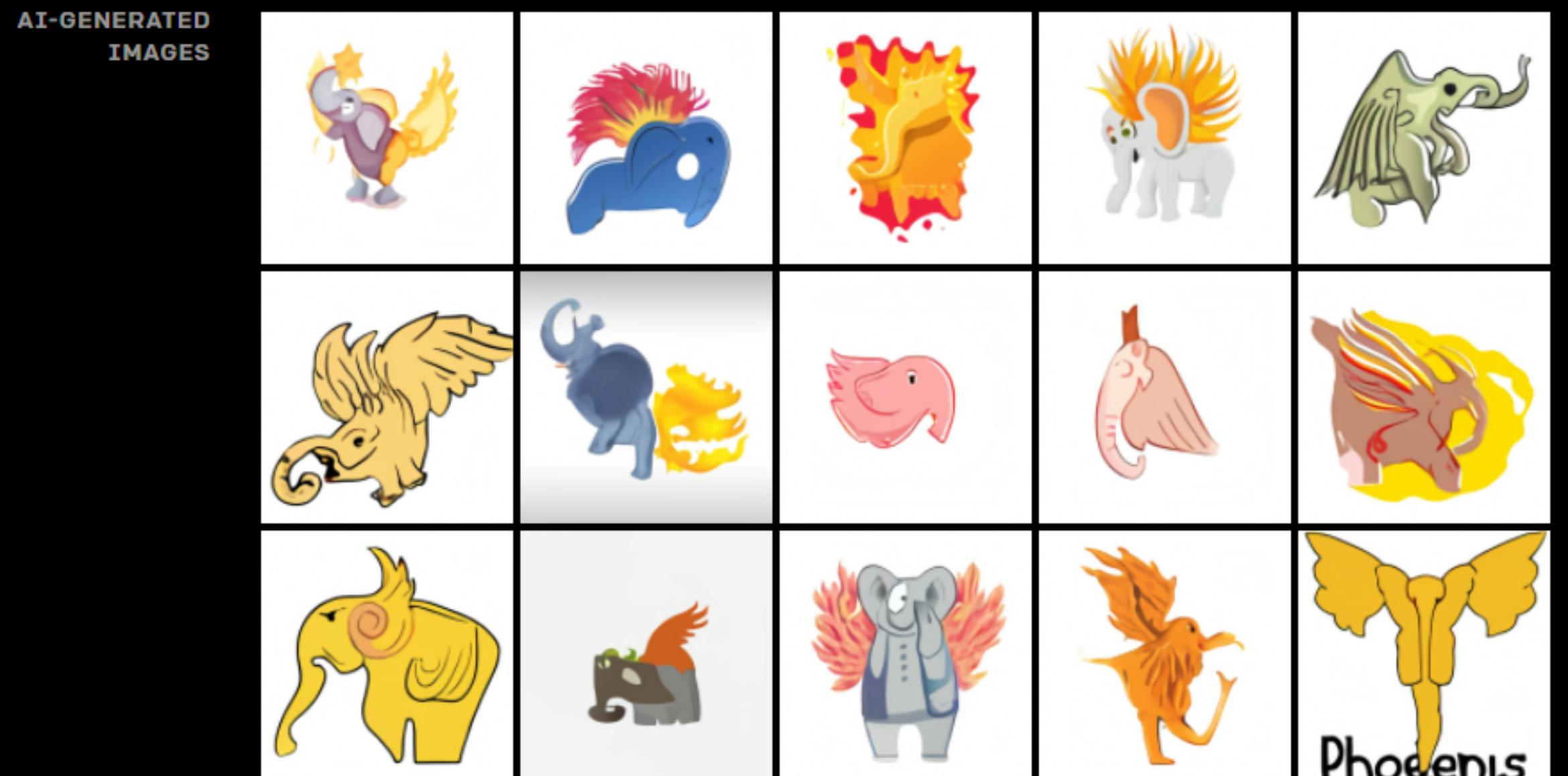
Модель DALL-E - це мультимодальна реалізація GPT-3 з 12 мільярдами параметрів (зменшена порівняно зі 175 мільярдами параметрів GPT-3), яка "змінює текст на пікселі", навчена на парах текст-зображення з Інтернету. Вона використовує нульове навчання до створення висновку з урахуванням описи і підказки без додаткового обучения.

DALL-E здатний генерувати зображення у різних стилях, від фотореалістичних зображень до картин та емодзі. Він також може "маніпулювати та переставляти" об'єкти у своїх зображеннях. Однією із здібностей, відмічених його творцями, було правильне розміщення елементів дизайну в нових композиціях без явних вказівок: "Наприклад, коли його просять намалювати редьку дайкон, що сморкається, потягує латте або катається на одноколісному велосипеді, DALL-E часто малює хустку, руки та ноги у правдоподібних місцях".

Приклад роботи моделі



TEXT PROMPT a professional high quality emoji of an elephant phoenix chimera. an elephant imitating a phoenix. an elephant made of phoenix. a professional emoji.



Веб-сервіси з функціоналом генерації зображень

<https://openai.com/blog/dall-e/> - вище описана модель

<https://thispersondoesnotexist.com/> - генерує обличчя людини

<https://thiscatdoesnotexist.com/> - генерує котика

<https://affinelayer.com/pixsrv/> - генерує зображення з контуру

edges2cats

