



Нейронні мережі

Лекція 6: Комп'ютерний зір I

Кочура Юрій Петрович
iuriy.kochura@gmail.com
[@y_kochura](https://twitter.com/y_kochura)

Сьогодні

Як змусити нейронні мережі «бачити»?

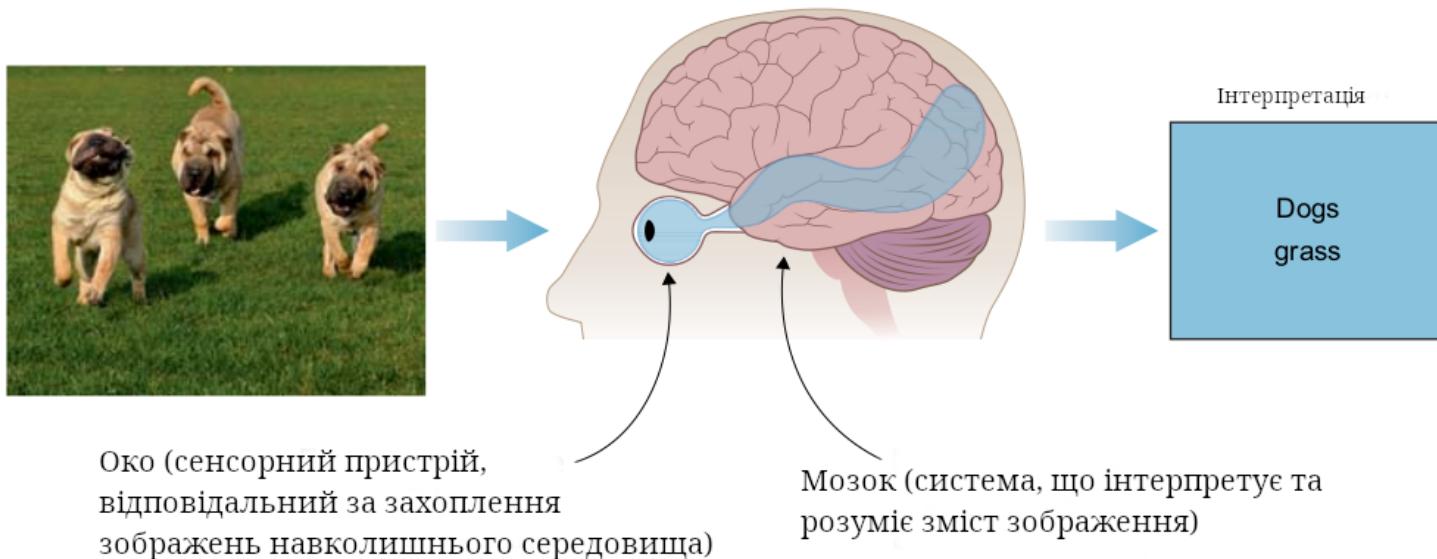
- Microphone icon Система людського зору
- Microphone icon Система комп'ютерного зору
- Microphone icon Застосування комп'ютерного зору

Як оцінити якість моделі та порівняти різні моделі на однакових завданнях?

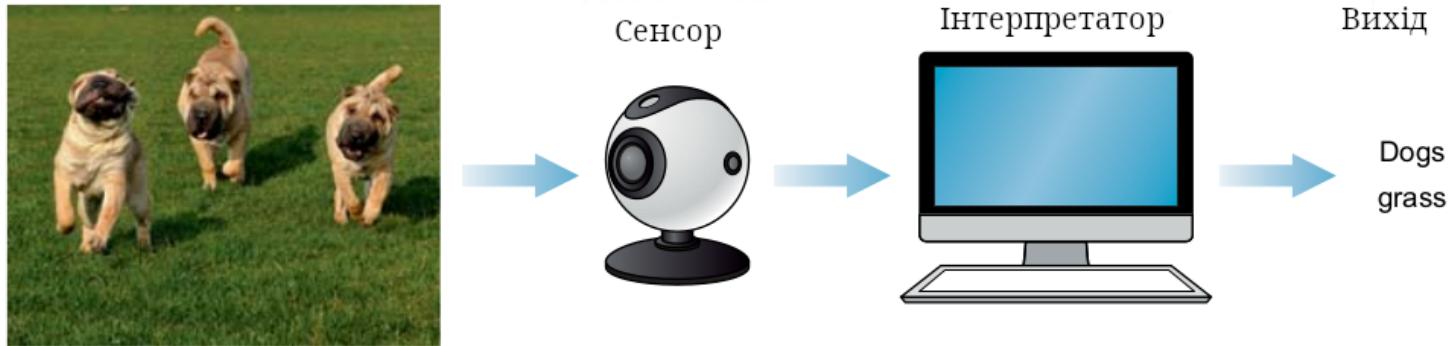
- Microphone icon Класифікація: точність, впучність, повнота, F1 міра
- Microphone icon Детекція: IoU, AP, AUC

Візуальне сприйняття

Система людського зору



Система комп'ютерного зору



Сенсорні пристрой

Розглянемо приклад автономного транспортного засобу (АТЗ).

Головна мета системи зору АТЗ – забезпечити здатність автомобіля розуміти навколошнє середовище та безпечно й вчасно переміщатися з точки А в точку В. Щоб досягти цієї мети, транспортні засоби оснащуються комбінацією сенсорів, здатних виявляти об'єкти в радіусі 360 градусів – пішоходів, велосипедистів, інші автомобілі, дорожні знаки та інші об'єкти – на відстані до трьох футбольних полів.

Деякі сенсорні пристрой, які зазвичай використовуються в АТЗ:

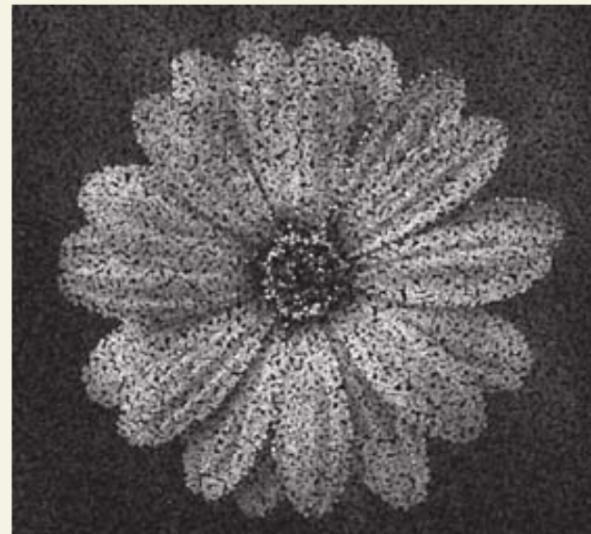
- Лідар – метод, подібний до радара, який використовує невидимі світлові імпульси для створення високоточної 3D-карти навколошньої місцевості.
- Камери допомагають розпізнавати дорожні знаки та розмітку, але не здатні точно вимірювати відстань.
- Радар може визначати відстань і швидкість, але не "бачить" з високою деталізацією.

Розпізнавання зображень

Тварини, люди та комахи мають очі як сенсорні пристрої, але структура, якість зображення та роздільна здатність очей різняться в залежності від потреб кожного організму. **Бджоли та інші комахи** мають складні очі, що складаються з багатьох лінз – до 30 000 в одному оці. Складні очі мають низьку роздільну здатність, що обмежує їх здатність розпізнавати об'єкти на великій відстані. Проте вони дуже чутливі до руху, що критично для виживання під час швидкого польоту. Бджоли не потребують високої роздільної здатності зображень. Їхні системи зору налаштовані на те, щоб помічати найменші рухи під час польоту.

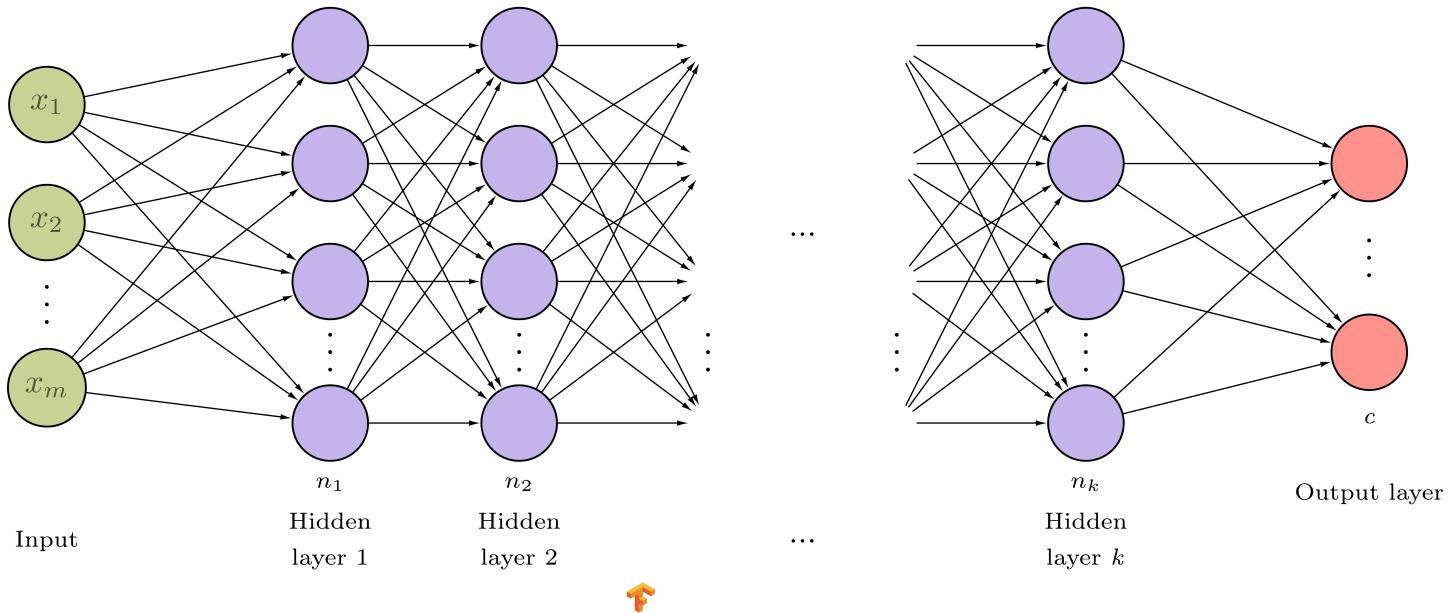


Складне око



Так бджоли бачать квітку

Інтерпретатор



```
import tensorflow as tf

model = tf.keras.Sequential([
    tf.keras.layers.Dense(n1),
    tf.keras.layers.Dense(n2),
    .
    .
    .
    tf.keras.layers.Dense(nk),
    tf.keras.layers.Dense(c)
])
```

Чи може машинне навчання досягти кращої продуктивності, ніж людський мозок?

Останні досягнення в ШІ та глибокому навчанні дозволили машинам перевершити людську зорову здатність у багатьох застосуваннях класифікації зображень та виявлення об'єктів. Можливості цих технологій швидко поширюються на нові сфери.

Застосування CV

Single-label multi-class classification



- Biking
- Running
- Swimming

Multi-label classification

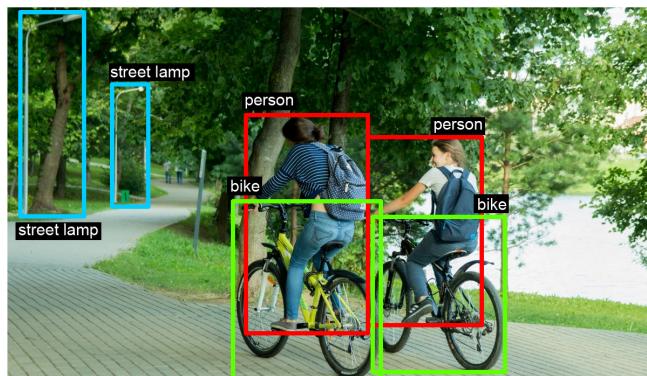


- Bike
- Person
- Tree
- Car
- Boat
- House

Image segmentation



Object detection



Класифікація зображень

Метою є присвоєння одного або кількох лейблів зображеню. Це може бути або класифікація з одною міткою (зображення може належати тільки до однієї категорії), або класифікація з кількома мітками (визначення всіх категорій, до яких належить зображення, як показано на попередньому слайді).

Наприклад, коли ви шукаєте ключове слово в додатку Google Photos, за лашунками ви запитуєте дуже велику модель для класифікації за кількома мітками — модель з більш ніж 20 000 різних класів, навчена на мільйонах зображень..

Детекція об'єктів

Метою є намалювати прямокутники (так звані [обмежувальні рамки](#)) навколо об'єктів інтересу на зображенні та асоціювати кожен прямокутник з певним класом.

Автономний автомобіль може використовувати модель детекції об'єктів для моніторингу автомобілів, пішоходів та знаків, що потрапляють в поле зору його камер.

Сегментація зображень

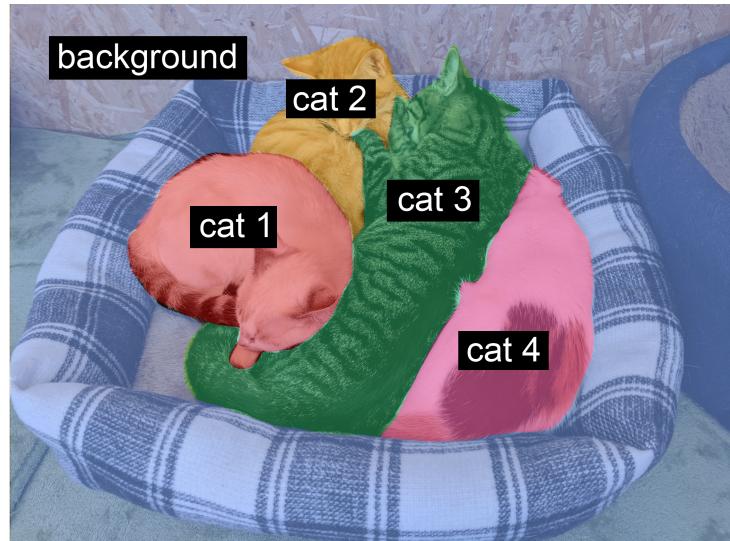
Метою є «сегментація» або «поділ» зображення на різні ділянки, при цьому кожна ділянка зазвичай представляє певну категорію.

Наприклад, коли Zoom або Google Meet показують користувацький фон під час відеодзвінка, вони використовують модель сегментації зображень, щоб точно відокремити ваше обличчя від того, що знаходиться позаду, з точністю до пікселів.

Приклади сегментації зображень



Семантична сегментація



Екземплярна сегментація

Демо

Computer Vision Explorer

Метрики оцінки

Класифікація

Матриця невідповідностей

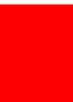
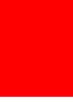
		Actual	
		+	-
Predicted	+	True Positive	False Positive
	-	False Negative	True Negative

- True Positive (TP)
- True Negative (TN)
- False Positive (FP)
- False Negative (FN)

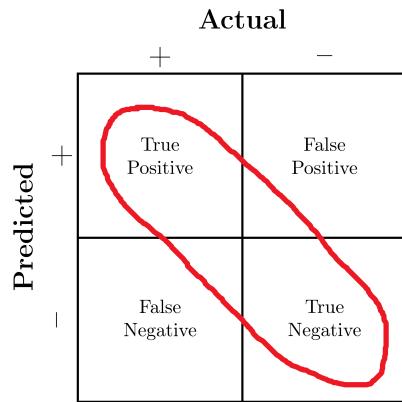
Приклад 1: Виявлення котів

Prediction	+	-	-	+	-	+
Image						
True Positive		True Negative		False Negative	False Positive	

Приклад 2: Виявлення пневмонії

	Normal		Pneumonia						
Test dataset									
Model1									
Model2									
Model3									

Точність



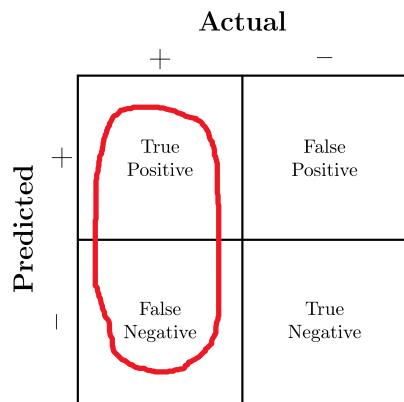
	Normal	Pneumonia
Test dataset	10 green	1 red
Model1	10 green	0 red
Model2	9 green	1 red
Model3	6 green	4 red

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Якщо використовувати точність (accuracy) як метрику оцінки, здається, що найкращою моделлю є Model1:

$$\text{Accuracy}(M_1) = \frac{9}{10} \quad \text{Accuracy}(M_2) = \frac{8}{10} \quad \text{Accuracy}(M_3) = \frac{5}{10}$$

Повнота



	Normal	Pneumonia
Test dataset	10	1
Model1	10	0
Model2	9	2
Model3	7	8

$$\text{Recall} = \frac{TP}{TP + FN}$$

Якою є чутливість кожної моделі?

$$\text{Recall}(M_1) = \frac{0}{1}$$

$$\text{Recall}(M_2) = \frac{1}{1}$$

$$\text{Recall}(M_3) = \frac{1}{1}$$

Влучність

		Actual	
		+	-
Predicted	+	True Positive	False Positive
	-	False Negative	True Negative

	Normal	Pneumonia
Test dataset	█	█
Model1	█	█
Model2	█	█
Model3	█	█

$$\text{Precision} = \frac{TP}{TP + FP}$$

Якою є влучність кожної моделі?

$$\text{Precision}(M_1) = \frac{0}{0} \quad \text{Precision}(M_2) = \frac{1}{3} \quad \text{Precision}(M_3) = \frac{1}{6}$$

F1 міра

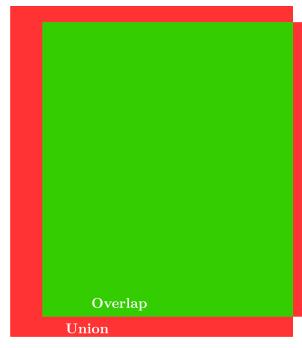
$$F_1 = \frac{2}{\frac{1}{\text{Повнота}} + \frac{1}{\text{Влучність}}} = 2 \cdot \frac{\text{Влучність} \cdot \text{Повнота}}{\text{Влучність} + \text{Повнота}}$$

Якою є міра F1 для кожної моделі?

$$F_1(M_1) = 0 \quad F_1(M_2) = \frac{1}{2} \quad F_1(M_3) = \frac{2}{7}$$

Детекція об'єктів

Intersection over Union (IoU)



$$\text{IoU} = \frac{\text{area of overlap}}{\text{area of union}}$$

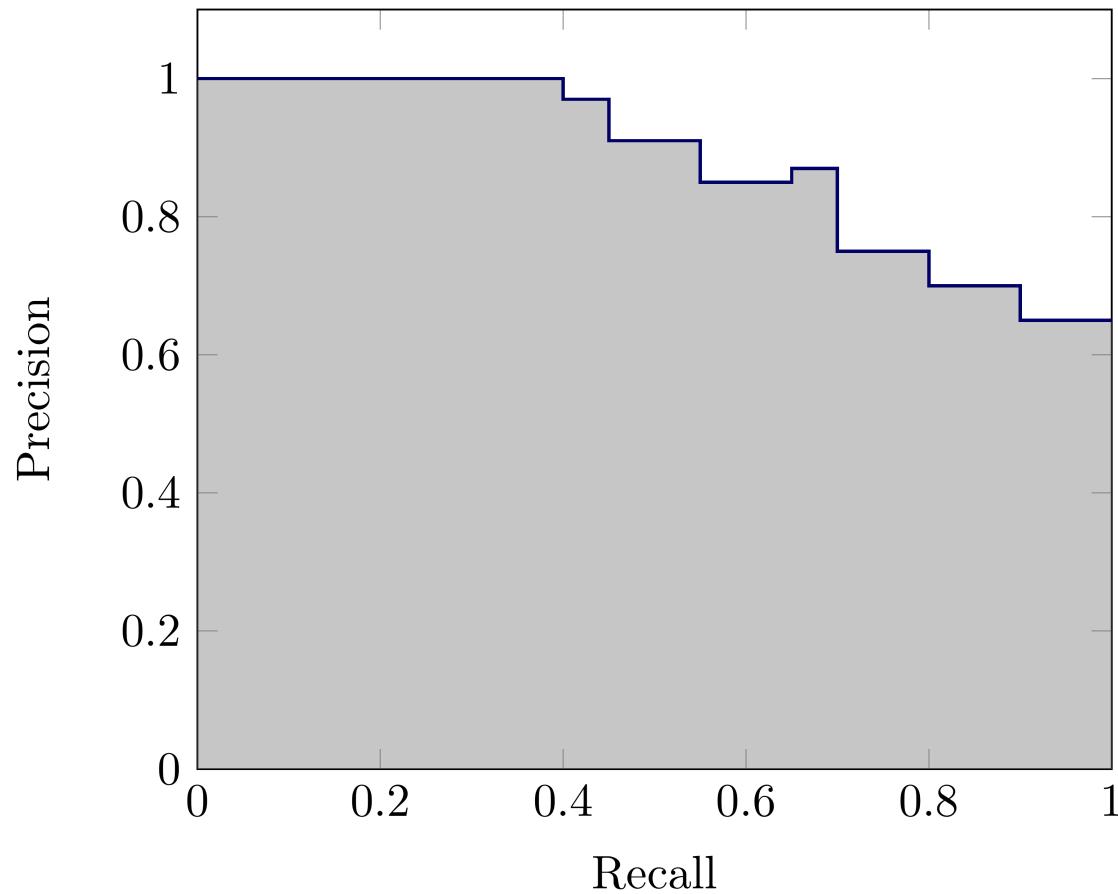


Ground Truth

Prediction

Average Precision (AP): the Area Under Curve (AUC)

2-class Precision-Recall curve: AP = 0.68



Демо

Evaluation Metrics

