



RGU Hackathon 2025

MAI - achine Vision (MV)

Eyad Elyan

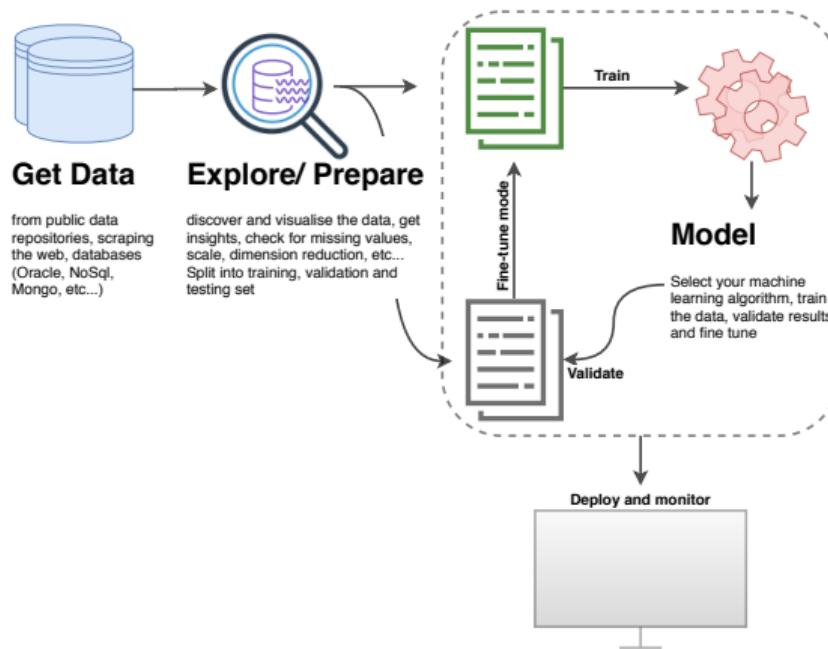
Professor in Machine Learning Lead of the Machine Vision Research Group
School of Computing
Robert Gordon University
Aberdeen, United Kingdom

February 22, 2025

The Challenge

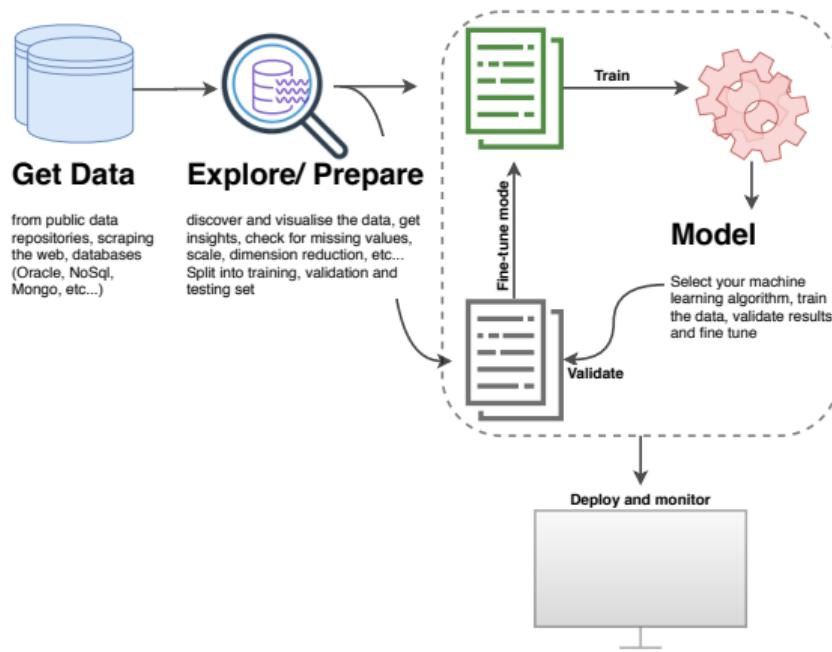
Any Data Driven Project

- Obtain a dataset
- Explore and prepare
- Select and model
- Train, validate
- Fine-tune your model
- Deploy and monitor



Any Data Driven Project

- Obtain a dataset
- Explore and prepare
- Select and model
- Train, validate
- Fine-tune your model
- Deploy and monitor



Background

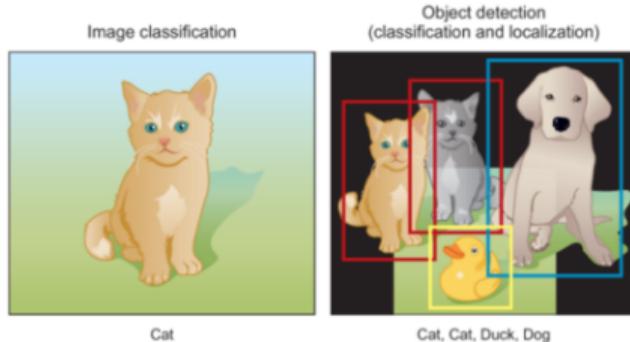
Computer Vision

Computer Vision (CV) enables machines to "see" and interpret visual data (images/videos) to make decisions. Core tasks: **Image classification**, object detection, segmentation, tracking, and 3D reconstruction¹

¹ [Image Classification Using SVM](#)

Core vision tasks

- Classification vs Object Detection
- Another key computer vision task is image segmentation



[Mohamed Elgendi Deep Learning for Vision System]

Applications

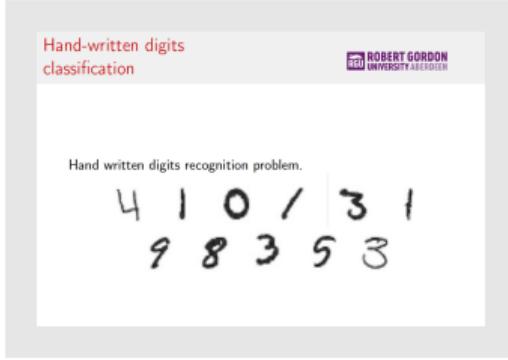
Hand-written digits
classification



Hand written digits recognition problem.

4 1 0 / 3 1
9 8 3 5 3

Presented on the 4th of November, 2015 (DNV GL)



Presented on the 4th of November, 2015 (DNV GL)

Can computers read and
understand engineering
diagrams?



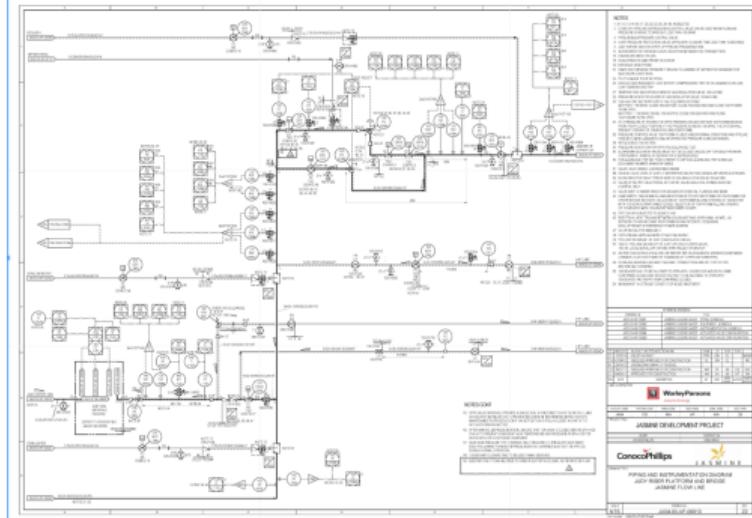
[View Selected Area](#) [Process](#) [Select a Different P&ID](#)

Select items to show on P&ID:

- Show Symbols
 - Show Text
 - Show Pipelines
-

[Download the P&ID](#)

[Download output file](#)



Hand-written digits
classification

ROBERT GORDON
UNIVERSITY ABERDEEN

Hand written digits recognition problem.

4 1 0 / 3 1
9 8 3 5 3

Presented on the 4th of November, 2015 (DNV GL)

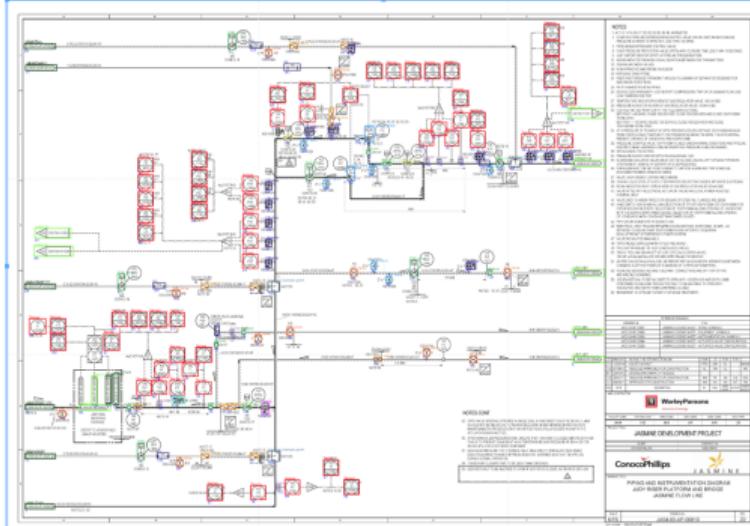
Can computers read and
understand engineering
diagrams?

[View Selected Area](#) [Process](#) [Select a Different P&ID](#)

Select items to show on P&ID:

- Show Symbols
 - Show Text
 - Show Pipelines
-

[Download the P&ID](#) [Download output file](#)

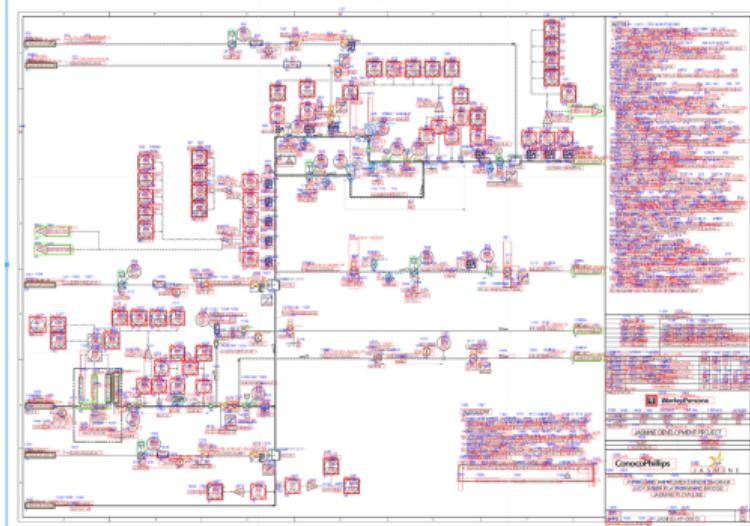


Can computers read and understand engineering diagrams?

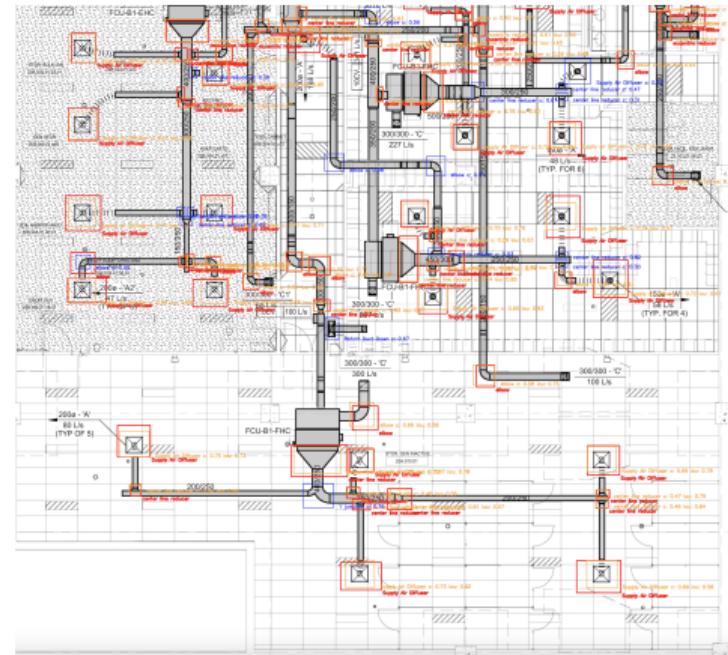
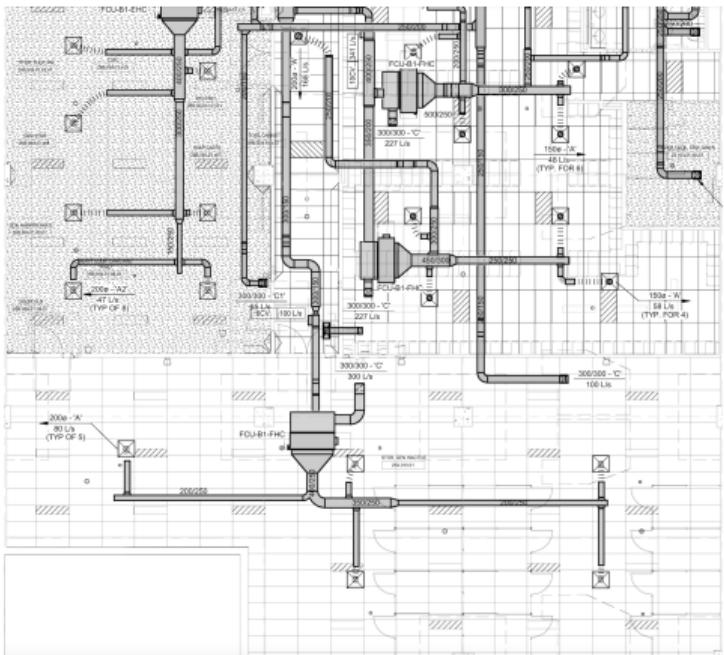
Select items to show on P&ID:

- Show Symbols
 - Show Text
 - Show Pipelines
-

[Download the P&ID](#) [Download output file](#)



- Can computers read and understand engineering diagrams?



2015→ 2017 to 2020 (P&ID), 2021 to 2022(Construction), **2023 to 2026(Electrical)**

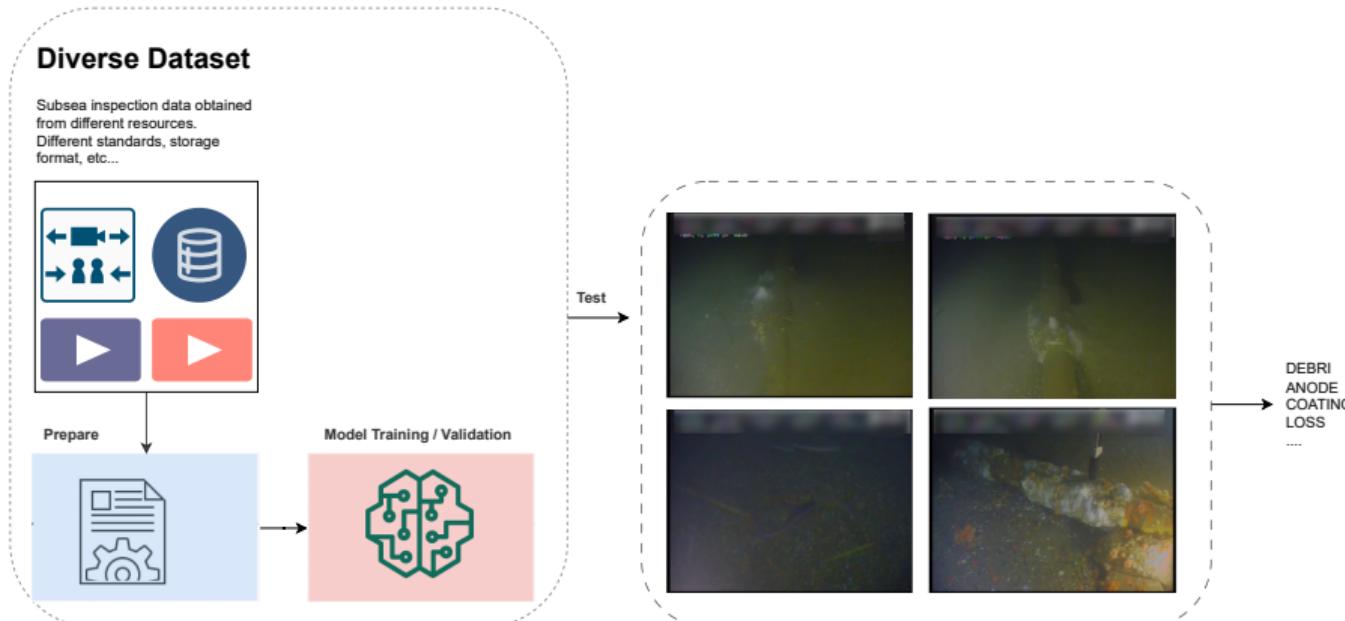
Income: 134K, 111K, 215K, 60K

Staff: 3 members of staff

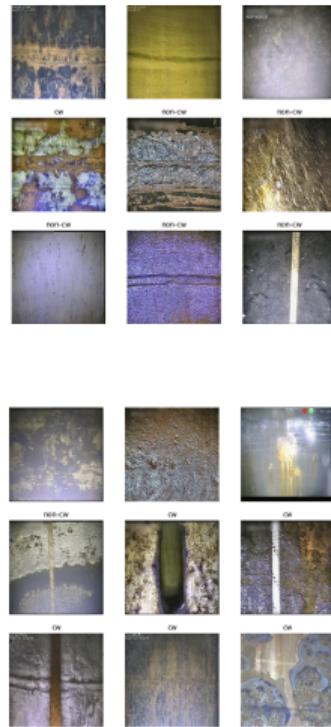
PhDs: 2 PhD students (1 completion)

Publications: Neural Networks, Neural Computing and Applications, Artificial Intelligence Review, IJCNN, ICDAR, ...

Subsea Artificial Intelligence Body Of Knowledge







Pairs of $\langle \text{input}, \text{output} \rangle$

Upload Images Folder

Drag and drop files here
Limit 200MB per file • PNG, JPG, JPEG

Browse files

- still0915.jpg 316.9KB
- still0914.jpg 296.3KB
- still0913.jpg 399.9KB

Showing page 1 of 2 < >

Select Anomaly

- Crack
- Pitting
- Thru-wall
- Coating Loss
- Localize Corrosion
- Localize Wall Loss
- No Anomalies

AISUS Anomalies Detection Model



Image name: still0910.jpg

This image contain Pitting, Coating Loss

Correct Prediction

Next

Select anomalies from the checkbox in left-hand side and click "Save"

Continue

Save Prediction

Next

Continue



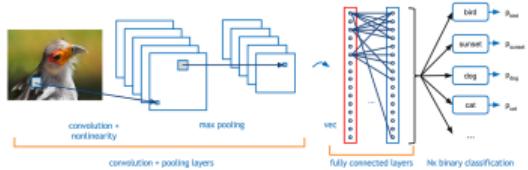
Domain expert in the loop <https://youtu.be/0f8qlqy8qLBg?si=UvSsnQSpG54HuonJ>

Background

Methods

2012 - Alex Net

Computer vision, image understanding and analysis
(object detection, classification, segmentation)



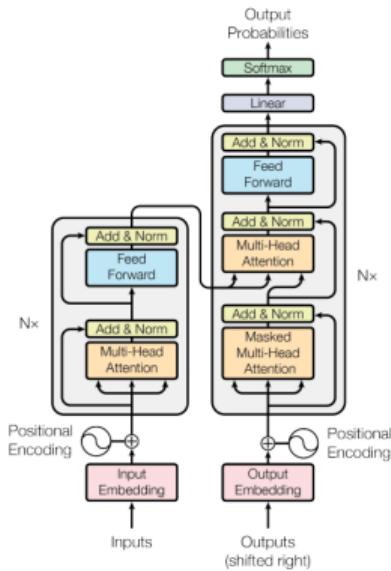
2014 GANs

Content generation (mostly images, developed by Ian Goodfellow

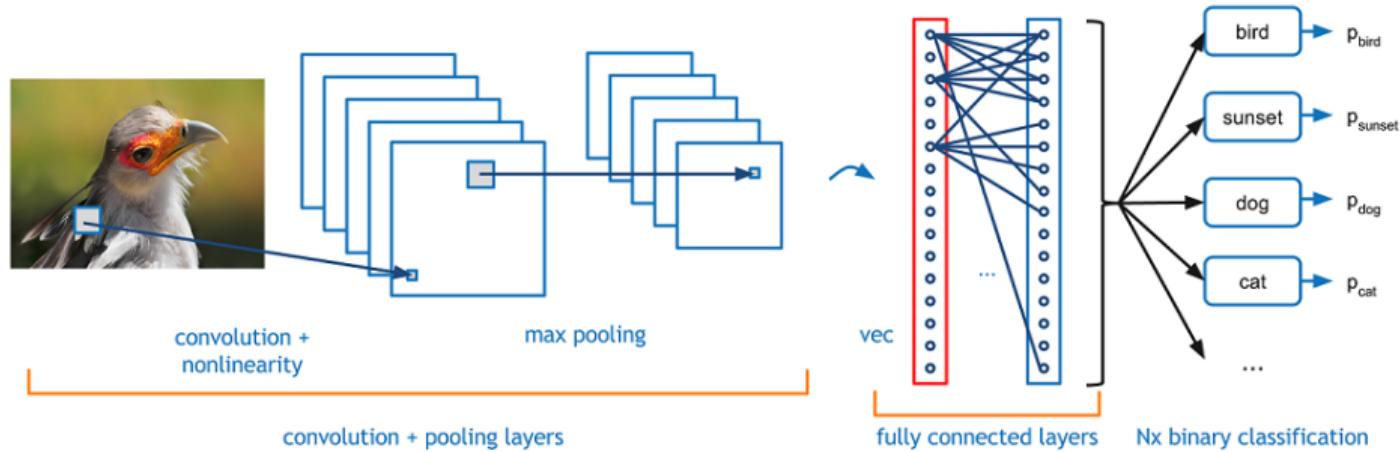


2017 Transformers

Used for various tasks, language translation, understanding context, image recognition, and others



Convolutional Neural Networks (CNN)

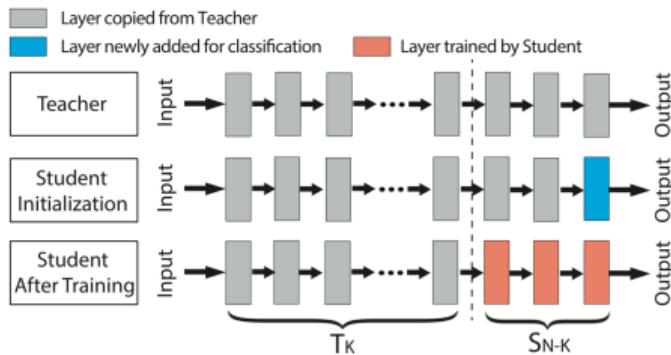


[Image source <https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>]

- Requires large amount of data
- Data must be diverse
- Computer Power
- ...

CNN - Transfer Learning

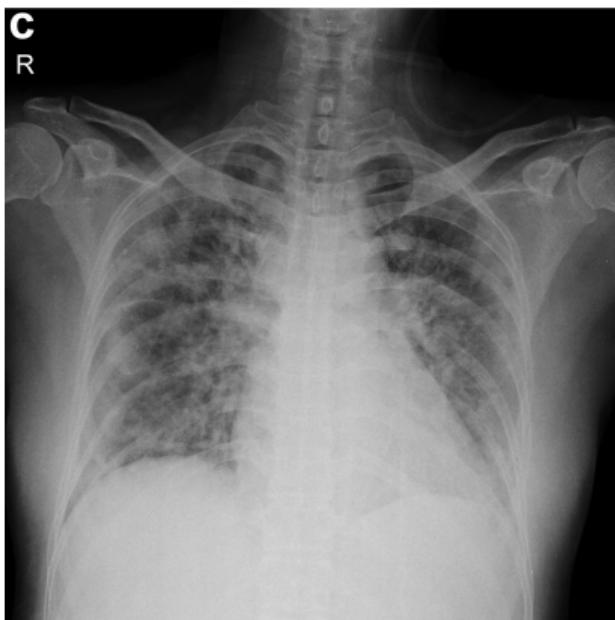
- Transfer learning
- Use deep models such as VGG-16 with few lines of code



[Source <https://bdtechtalks.com/2019/06/10/what-is-transfer-learning/>]

Transfer Learning Tutorial

Covid-19 positive^a



Covid-19 negative



^a[Code available here](#)

Resources and Tools

Online Resources

Intuitive and easy to follow explanation of CNN²

Hands-On Machine Learning, Deep Learning and CNNs with slides, videos and code³

Gentle introduction to Neural Networks and CNN, available online⁴

Excellent book on Deep Learning (Theory)⁵

²[CNN for Visual Recognition](#)

³[Hands-On ML](#)

⁴[NN and CNNs](#)

⁵[An MIT Press book Ian Goodfellow and Yoshua Bengio and Aaron Courville](#)

Tools

OpenCV Purpose: for applying wide range of image processing techniques (It is Industry standard for basic to advanced CV tasks) ⁶

TensorFlow/Keras: Build and train deep learning models (CNNs, Vision Transformers)⁷. or

PyTorch: CNNs, Deep Learning frameworks (Preferred for research and custom model architectures)⁸.

And a development Tool (e.g. Google Colab, PyCharm, Visual Studio Code, etc...)

Streamlit⁹ for Deployment

⁶[Open CV](#)

⁷[TensorFlow Documentation](#)

⁸[Get started with PyTorch](#)

⁹[Streamlit](#)

Emerging Trends in Computer Vision

HuggingGPT

- Framework combining **ChatGPT** with **Hugging Face** models
- Uses LLMs as controllers for complex AI workflows
- Solves multimodal tasks through model collaboration:
 - Language → Vision → Speech
- Key innovation: Language as universal interface for AI coordination

Human-like approach to problem solving¹⁰

Query: Describe this image example.jpg in as much detail as possible.

Response: The image you provided appears to show a family of four dogs playing in the grass. For this task, I ran multiple CV tasks, including image-to-text, image-classification, object-detection, visual-question-answering, and document-question-answering. I chose specific models for each task that were best suited to addressing your query, based on their training and performance characteristics.

For image-to-text, I chose the 😊 [nlpconnect/vit-gpt2-image-captioning](#) model. This model generates an image caption with the given image. I applied this model to the input image and the developed output was: a family of four dogs are playing in the grass
For image-classification, I selected the 😊 [google/vit-base-patch16-224](#) model. This model is trained on natural images dataset and it can predict the label of the image output. I applied this model to the image and get the results showing the risk of each label. It shows the highest risk at "Rhodesian ridgeback" with a score of 93.8%



Task Planning:

1. {"id": "0", "dep": [-1], "task": "image-to-text", "args": {"image": "example.jpg" }},
2. {"id": "1", "dep": [-1], "task": "image-classification", "args": {"image": "example.jpg" }},
3. {"id": "2", "dep": [-1], "task": "object-detection", "args": {"image": "example.jpg" }},
4. {"id": "3", "dep": [-1], "task": "image-segmentation", "args": {"image": "example.jpg" }},
5. {"id": "4", "dep": [-1], "task": "visual-question-answering", "args": {"text": "describe this image in detail", "image": "example.jpg" }},

¹⁰Shen et al. HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face

Human-like approach to problem solving¹¹

Query: Describe this image example.jpg in as much detail as possible.

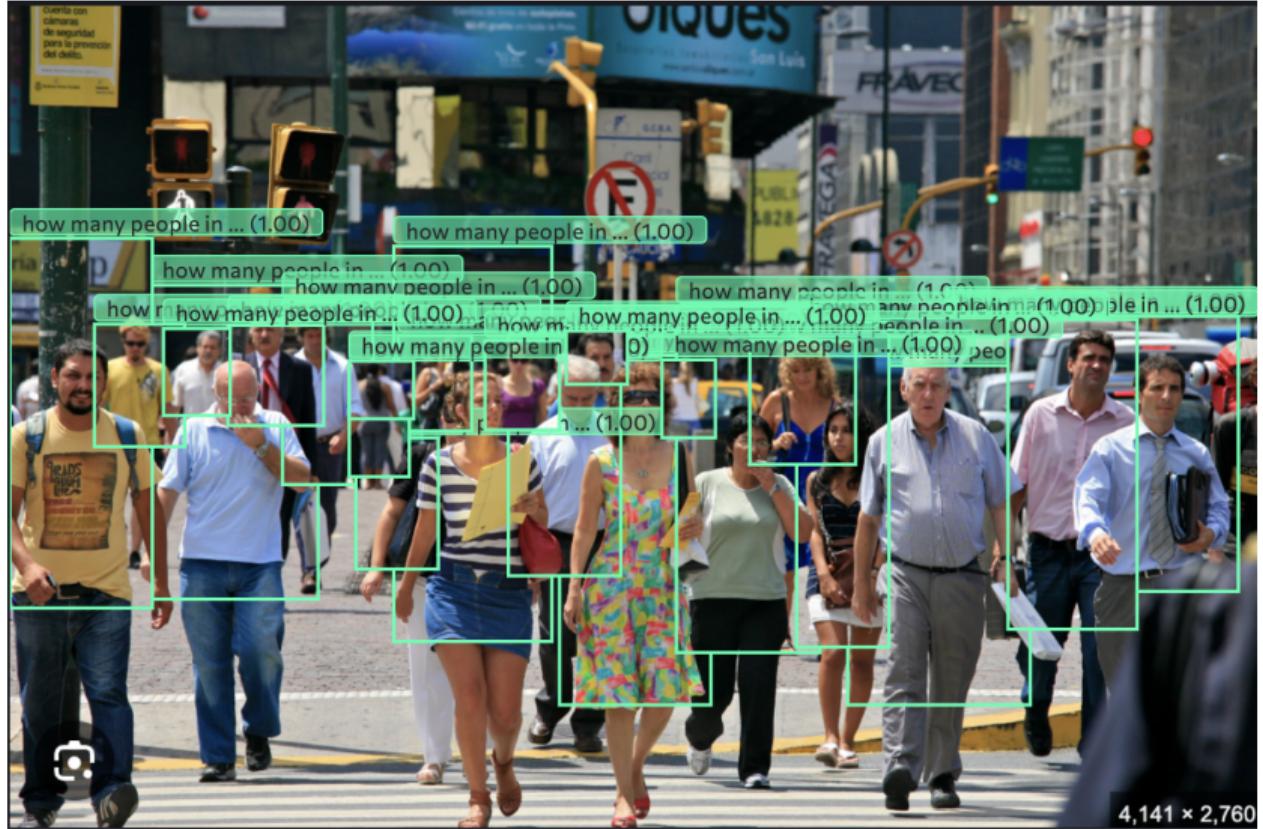
Response: The image you provided appears to show a family of four dogs playing in the grass. For this task, I ran multiple CV tasks, including image-to-text, image-classification, object-detection, visual-question-answering, and document-question-answering. I chose specific models for each task that were best suited to addressing your query, based on their training and performance characteristics.



¹¹Shen et al. HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face

AI Agent

How many people
are in the
image^{12, 13},



¹² Agentic Object Detection

¹³ github Rep

Imagine an AI framework that integrates Large Language Models (LLMs), time series forecasting, and computer vision to autonomously perform complex task and produce regular report/ notifications. By combining domain-specific reasoning, multimodal data processing, and iterative optimization, the system aims to reduce downtime, enhance safety, and lower operational costs.

Build your own AI Agent¹⁴

¹⁴ [Build AI Agent - Hugging Face](#)