Bottle, Can or Coffee Cup ?!

How Computer Vision and Machine Learning can be used to Recognise Different Materials to Make Recycling Easier

Glossary

Here are explanations of the highlighted technical terms used in the project documentation which will help with understanding the poster, booklet and webpages.

The terms are presented in alphabetical order.

For further information about the project this glossary relates to visit the Discover Materials website at

https://discovermaterials.co.uk/resource/bottle-can-or-coffee-cup/

or by scanning the QR code…

Discover Materials

visit **DISCOVER MATERIALS**
https://discovermaterials.co.uk
and learn more about what's happening in the world of materials science!

1.43 GHz

This is a measure of the number of operations the CPU performs every second, 1,430,000,000 operations per second.

A

Arm Cortex A57 CPU

The ARM Cortex-A57 is a high-performance central processing unit (CPU) designed by ARM Holdings, utilizing the ARMv8-A 64-bit instruction set. Known for its powerful processing capabilities it is often used in mobile devices and embedded systems

Artificial Intelligence

AI refers to the development of programs that behave intelligently and mimic human intelligence through a set of algorithms. The field focuses on three skills: learning, reasoning, and self-correction to obtain maximum efficiency. AI can refer to either machine learning-based programs or explicitly programmed computer programs.

Artificial Neural Network

An artificial neural network or ANN is a machine learning model inspired by the structure and function of biological neural networks in brains. An ANN consists of connected units or nodes called artificial neurons, which loosely model the neurons in the brain. Each artificial neuron receives signals from earlier connected neurons, then processes and sends a signal to other connected neurons. The "signal" is a real number, and the output of each neuron is computed by a non-linear function of the sum of its inputs, called the activation function. The strength of the signal at each connection is determined by a weights, which adjusts during the training. The signal is only sent on if it is greater than a bias value. Typically, neurons are aggregated into layers. Signals travel from the first layer (the input layer) to the final layer (the output layer), passing through multiple intermediate layers (hidden layers). Neural networks are discussed in detail in section 03 Machine Learning & Neural Networks at https://discovermaterials.co.uk/resource/03-machine-learning-neural-networks/

B

back-propagation

Short for "backward propagation of errors" this is an algorithm for the supervised learning of artificial neural network. Given an error between a prediction of a result and the true result, the method calculates the gradient (difference between the two), of the error with respect to the neural network's weights and biases. The "backwards" part of the name stems from the fact that calculation of the gradient proceeds backwards through the network, with the gradient of the final layer of weights being calculated first and the gradient of the first layer of weights being calculated last. Partial computations of the gradient from one layer are reused in the computation of the gradient for the previous layer. This backwards flow of the error information allows for efficient computation of the gradient at each layer versus the naive approach of calculating the gradient of each layer separately.

balanced dataset

to a dataset where all classes or categories are represented in roughly equal proportions, meaning there is a similar number of samples for each class, preventing the model from becoming biased towards any particular class and leading to more accurate predictions.

biases

the value a node in a neural network must reach before a signal is re-transmitted further into the network.

C

Computer Vision

Computer vision is a field of artificial intelligence that allows computers to interpret and understand visual information from digital images and videos, essentially enabling machines to "see" and make sense of the world around them by identifying and classifying objects, people, and scenes within the visual data, similar to how humans do with their eyes.

CPU

Central Processing Units are traditional micro-processors which are designed to perform one operation at a time, one after the other. Modern CPUs can perform many thousands of millions of operations every second and so have become powerful enough to do the billions of additions and multiplications needed to process the matrix that represents a large neural network.

CUDA compute power

CUDA stands for Compute Unified Device Architecture. It's a software platform that allows developers to use GPUs for general-purpose processing rather than just computer graphics. CUDA uses a parallel computing model, where many calculations happen at the same time and forms the bridge between the code you might and the GPU hardware which it runs on. NVIDIA created CUDA in 2006 and it has become a crucial part of high-performance computing and machine learning.

D

E

edge-computing

Edge computing is a system of distributed computing where data processing occurs closer to the source of the data rather than sending it to a centralized cloud, enabling faster real-time analysis, reduced latency, and improved responsiveness. The concept of edge-computing is discussed in detail in section 01 Computing on the Edge at https://discovermaterials.co.uk/resource/01-computing-on-the-edge/

F

features

In machine learning, features refer to individual, measurable characteristics or properties extracted from data that are used as input to train a model. They are the specific attributes of a data that the algorithm analyses to make predictions, In this project the features identified by the ResNet 18 network will tend to refer to the types of curves, edges and areas present in the images seen.

final layer

The final layer, also called the output layer, is the last layer of a neural network that produces the final prediction or output based on the information processed through the previous layers of the network, translating the numerical results of the network into a meaningful format for the task at hand. The number of neurons in this layer typically corresponds to the number of possible output categories in a classification problem.

G

GPU

A Graphical Processing Unit is a specialist microchip that was traditionally designed to perform the addition and multiplication of large matrix values in one parallel operation, rather than one value at a time like with a CPU. This was useful for providing the results of the calculations needed to display 3D computer graphics in real-time, hence the name Graphical Processing Unit. It was realised that the matrix maths needed for computer graphics, which these processors were designed to complete quickly, could be adapted for the matrix maths needed for training machine learning models quickly. These days the GPUs are designed more for performing this matrix maths as quickly and power-efficiently as possible, independent of its use in computer graphics and it is this improvement in the GPU hardware that has led to the current AI revolution.

H

High Performance Computing

High Performance Computing (HPC) refers to the practice of combining the processing power of many computers, often organized in clusters, to solve complex problems that are too large or time-consuming for one machine to handle, allowing for significantly faster computations compared to a standard desktop computer, Computing power is added together to tackle problems beyond the capabilities of a single workstation by utilizing parallel processing across multiple nodes.

hyperparameter

A configuration variable that is set before the training process begins and controls how the model learns, unlike parameters, such as the weights and biases which are values learned directly from the data during training. hyperparameters define the structure and behaviour of the learning algorithm, and their values can significantly impact model performance.

I

inference

the process of making a prediction, or *inferring* a result, from new images based on the patterns and structures the machine learning model has learnt from its training dataset.

inputs

In machine learning inputs refer to the data points or features that are fed into a model to make predictions. These are the characteristics or attributes of the data used by the algorithm to generate an output, the information the model is trying to learn from to make predictions based on new data.

iterations

an iteration is a single cycle of updating a model's parameters based on a small batch of training data, it is a single step in the learning process where the algorithm adjusts its weights and biases based on the information from that batch, it is a key part of the training process where the model learns from a portion of the data and improves its predictions based on that information.

J

K

L

latency

is the delay or lag between asking for a computer to do something and the computer appearing to have performed the task. Latency is caused by the time it takes to send large amounts of data through the machines and cables on which the internet operates.

layer structure

A layer refers to a fundamental building block in a neural network that receives input data, performs a specific transformation on it using a set of, and then passes the processed output to the next layer in the network. A layer is defined by the number of inputs and outputs it has, and the function applied within the layer to combine each of the inputs into outputs. The layer structure is the definition of the arrangement of layers stacked together between the first or input layer and the final or output layer.

M

Machine Learning

Machine learning is a subset of Artificial Intelligence, which uses algorithms that learn from data to make predictions. These predictions can be generated through supervised learning, where algorithms learn patterns from existing data, or unsupervised learning, where they discover general patterns in data. ML models can predict numerical values based on historical data, categorize events into different classes or cluster data points based on shared properties. Machine learning is discussed in detail in section 03 Machine Learning & Neural Networks at https://discovermaterials.co.uk/resource/03-machine-learning-neural-networks/

matrix

a matrix is a way of grouping together many connected values into an arrangement of rows and columns, to represent that an operation such as addition or subtraction is to be performed on all the numbers at the same time, rather than treating each value differently.

multi-core

A multi-core processor is a single computer chip that contains multiple independent processing units, cores, allowing it to execute multiple instructions at the same time, significantly improving performance compared to a single-core processor - by enabling parallel processing of tasks. Each core can work on a different task at the same time, boosting overall processing power and multitasking capabilities.

N

neurons

In a biological brain neurons, also called nerve cells, are the cells responsible for receiving sensory input from the external world, for sending motor commands to our muscles, and for transforming and relaying the electrical signals at every step in between.

In an artificial neural network neurons are the basic processing unit which receives input signals, performs a calculation based on assigned weights, and produces an output signal that can be passed on to other neurons. It is a node within the network that processes information and contributes to the overall computation.

nodes

In artificial neural networks nodes refers to a single computational unit, also called a neuron, which receives input signals, processes them using a weighted sum and activation function, and then outputs the result to the next layer in the network.

In High Performance Computing a node refers to a single computer or server within a cluster, a complete system with its own CPU, memory, and storage, that works together with other nodes to perform large-scale calculations in parallel, significantly boosting processing power.

NVIDIA Jetson Nano 2GB

The NVIDIA® Jetson Nano™ 2GB Developer Kit is ideal for teaching, learning, and developing AI and robotics. With an active developer community and ready-to-build open-source projects, you'll find all the resources you need to get started. It delivers incredible AI performance at a low price and makes the world of AI and robotics accessible to everyone with the exact same NVIDIA software and tools used to create breakthrough AI products across all industries. The Jetson Nano 2GB Developer Kit is supported by the comprehensive NVIDIA® JetPack™ SDK, and has the performance and capabilities needed to run modern AI workloads. The user guide can be found at https://developer.nvidia.com/embedded/learn/jetson-nano-2gb-devkit-user-guide

NVIDIA Maxwell GPU

The NVIDIA Maxwell GPU is a graphics processing unit (GPU) architecture developed by NVIDIA as the successor to the Kepler architecture and is used in the GeForce 700, 800M, 900, and Quadro series graphics cards being nearly twice as energy efficient as Kepler GPUs. The NVIDIA Maxwell GPU provides 128 cores of CUDA compute power for the machine learning to utilise.

O

optimisation

during training the machine learning model learns the patterns and structures by which to identify images by comparing its own predicted result with the true results, as labelled by a human. This gives an error value for that attempt. By systematically adjusting the millions of internal weights and biases within neural network by a small amount (through a process called back-propagation) the error is improved. Optimisation is the process of reducing the error value to a small as it can be over repeated attempts to improve it. Neural networks are discussed in detail in section 03 Machine Learning & Neural Networks at https://discovermaterials.co.uk/resource/03-machine-learning-neural-networks/

outputs

In machine learning outputs refer to the predictions or results generated by a trained model based on input data. The final outcome or decision that the model produces after analysing the given information can be a numerical value, a category label, a probability distribution, or any other format depending on the task and model type.

P

parallel

Parallel computing is a process where large problems are broken down into smaller tasks which can be solved at the same time by multiple processors. The processors communicate using shared memory and their solutions are combined using an algorithm once all the parts have been completed.

performance

In terms of machine learning performance refers to how well the model is able to accurately complete a given task, usually measured by how closely its predicted outputs match the expected true outputs on a set of data.

pre-train

it takes significant time to train a large neural network model on thousands of images so it is possible to store the resulting model so that that work does not have to be repeated, the model is pre-trained. With transfer learning it is possible to use that pre-trained model as the starting point for further training without having to start from scratch each time.

prediction

the process of taking a new set of inputs, passing them through a trained network and obtaining a result as output, which is the model's best prediction, or inference, of the result based on its training.

Q

R

re-train

taking a previous, or pre-trained model, and training all or part of the old model on new wither brand new data, or a combination of the old and new data. With transfer learning commonly only the final layer of the pre-trained model is re-trained on new data.

real-time

in computing real-time means the computer performs the task fast enough that a human does not recognise there is a delay.

ResNet 18

ResNet-18 is a convolutional neural network that is 18 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224 pixels.

S

Seeed Studio reComputer J1010

reComputer for Jetson series are compact edge computers built with NVIDIA advanced AI embedded systems. With rich extension modules, industrial peripherals, thermal management combined with decades of Seeed's hardware expertise, reComputer for Jetson is ready to help you accelerate and scale the next-gen AI product emerging in diverse AI scenarios. More information about the reComputer J1010 is available at:

https://web.archive.org/web/20230331031333/https://wiki.seeedstudio.com/reComputer_Jetson_Series_Introduction/

smart sensors

Smart sensors are devices that combine sensors with microprocessors to collect, analyse, and transmit environmental data. They are a key component of the internet of things (IoT).

training

the process of providing the machine learning model with many already identified historical images, from which to determine the patterns and structures by which to predict the identification of new images.

training epochs

In machine learning, a training epoch is a complete pass through a training dataset by a learning algorithm. Epochs are also known as training cycles.

training images

these are the set of images used to teach the computer model to recognize patterns or features. By providing human labelled examples of what objects look like, the labelled images are used to train the model during the learning process, enabling it to later make predictions on new, unseen images.

transfer learning

a machine learning technique where knowledge gained from training a model on one task is reused to improve the performance of a model on a related, but different task, applying previously learned information to a new problem, similar to how humans leverage past experiences to learn new skills more efficiently. For example, a model trained to recognize cars can be used as a starting point to learn to recognize trucks, as they share visual similarities. Instead of training a model from scratch, transfer learning utilizes a pre-trained model that has already learned features from a large dataset on a similar task as a starting point for the new task. Once the pre-trained model is loaded, it is often fine-tuned by adjusting the final layers of the network to adapt to the specific features of the new task and dataset. This requires less training data and time compared to training a model from scratch, especially when dealing with limited data for the new task, and leveraging knowledge gained from a large dataset can lead to better performance on a related smaller task.

transferred model

a transferred model in machine learning refers to a pre-trained model that is taken from one task and used as a starting point to train a new model for a related, but different task - leveraging the knowledge gained from the original model to improve performance on the new task. This allows for faster training with potentially less data needed for the new task.

U

V

W

weights

the amount of importance (represented numerically) given to the connection between two nodes in a neural network. Neural networks are discussed in detail in the section 03 Machine Learning and Neural Networks at https://discovermaterials.co.uk/resource/03-machine-learning-and-neural-networks/

X

Y

Z