

# Look, learn and inwardly digest!

Ordnance Survey is attempting to harness the power of Machine Learning by teaching networks to understand aerial imagery across Great Britain.

David Jones reports on progress

Aerial imagery has been an essential tool for keeping the national mapping database of Great Britain up-to-date for years.

Every day, production teams cast an expert eye over data captured in these pictures, interpreting it for contrasting reasons across the country.

The number of floors in a block of flats, the width of a river, the material of a factory roof which reveals its age - all these characteristics can be ascertained, thanks to the human eye and human knowledge.

This intelligence is then applied to help the likes of emergency services, local government, and businesses in real-time.

But what if machines could be trained to identify features from aerial images in the same way?

If it is possible for a computer to replicate human understanding of the different nuances from images it sees, how much bigger, quicker and more useful would geospatial data services become?

## Machine learning network

Ordnance Survey (OS) wants to know. Research scientists Izzy Sargent and Steve Coupland are working on a project to create a machine learning network run on OS's entire national data set.

Key to its success is being able to create metrics for machine learning that ensures machines can identify features in "generalised terms". A system that can train machines to operate over a multitude of different geographies rather than just cities or rural areas.

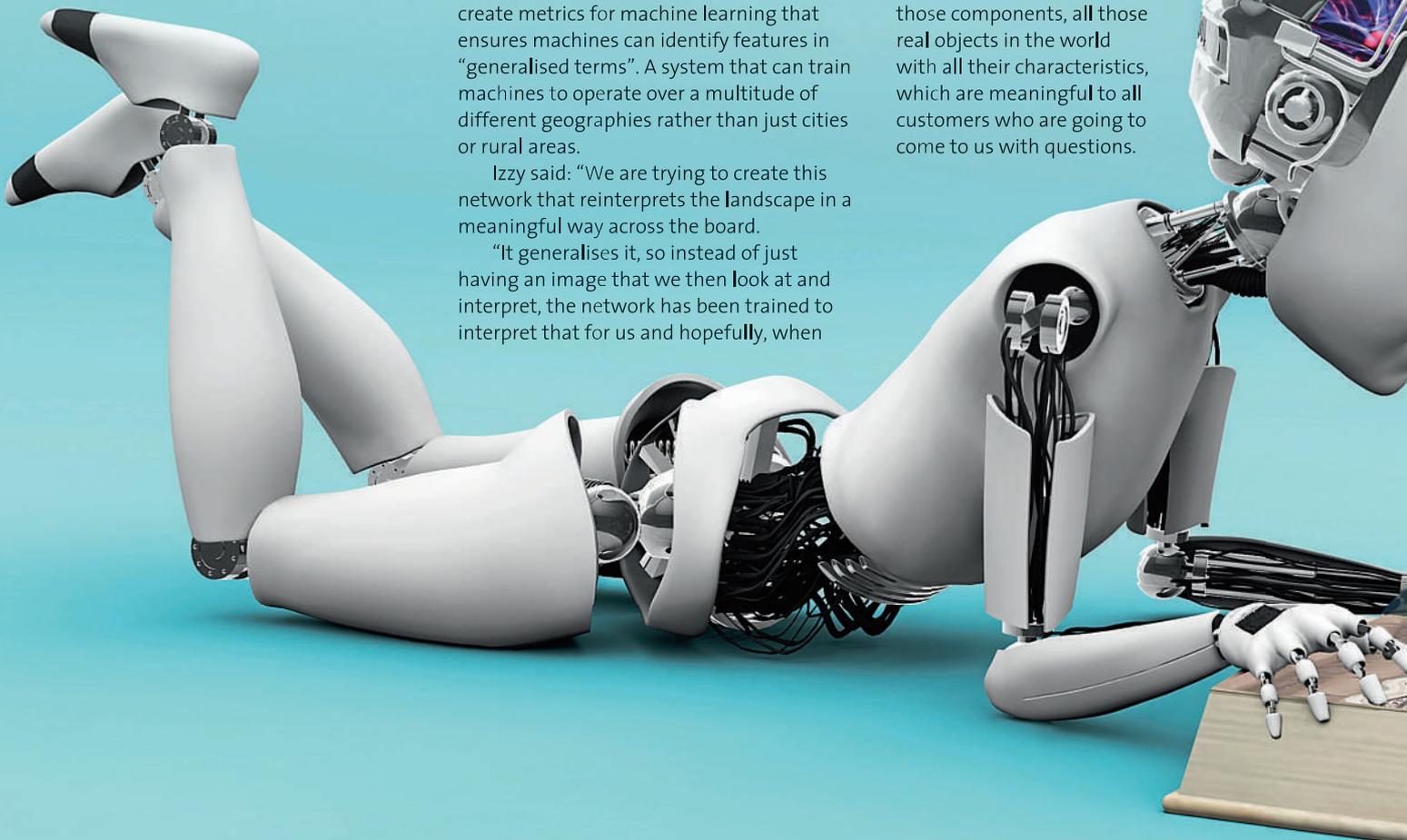
Izzy said: "We are trying to create this network that reinterprets the landscape in a meaningful way across the board.

"It generalises it, so instead of just having an image that we then look at and interpret, the network has been trained to interpret that for us and hopefully, when



OS research scientists Izzy Sargent and Steve Coupland

done well, it has identified all those components, all those real objects in the world with all their characteristics, which are meaningful to all customers who are going to come to us with questions.



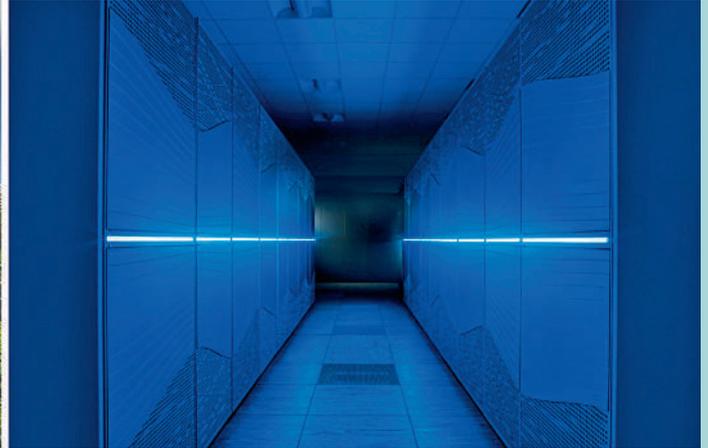
NPL



Ordnance Survey

Science and  
Technology  
Facilities Council

Three-way collaboration between the National Physical Laboratory, the Science Technology Facilities Council and Ordnance Survey is making the research possible



Ordnance Survey has harnessed the power of the powerful *Scafell Pike* supercomputer at STFC's Hartree Centre (pictured) to test and develop the machine learning techniques described in this article. Image: STFC

"By pre-processing all the data, it should be much faster for us to put together the right information for a customer answer.

"There might be one customer who works in insurance and wants to know the risks around the natural environment, and then contrast that with the Government wanting to understand the energy efficiency of the buildings within its housing stock and the different characteristics of the buildings involved.

"We are teaching it not by showing it, but by saying go and find and tell us what you find and give a name to it.

"Training neuro networks in this way directly doesn't answer customers' questions, but it tries to process the data in such a way that it makes it much faster to answer their questions later on."

### Innovate UK

The project was triggered by Innovate UK, the UK Government's innovation agency.

OS won a bid for its Analysts for Innovators (A4I) competition, which meant OS benefitted by teaming up with two significant partners.

The National Physical Laboratory (NPL) is

programming the metrics and standards the machine network needs to meet to become viable, while The Science and Technology Facilities Council (STFC) is providing its super computer to give the necessary power for the epic task to be computed.

Steve said: "Luckily NPL happened to be working towards creating a best practice for machine learning. They were developing their baseline for how you should perform machine learning, but they didn't have the data to check. Whereas we have all the data that needs checking, but don't have a best practice for machine learning yet. So together we are going to be a powerful combination.

"OS is supplying the data and the structure of the machine learning network.

"STFC is taking the data and applying it by using its supercomputer to create the new model, which is a huge undertaking.

"NPL is taking all our output data that we have created so far and that we will be creating with the new network, and it will be testing the quality and which metrics we need to concentrate on for our specific goals. There are various things they can test and they want to show us which ones would be the things to concentrate on to work for our purpose, which for us is generalisation."

OS has split the country into 56 x 56 metre grids to make it work. Instead of just overlaying squares on Britain's digital map, every small grid is centred upon a key feature. Features are divided into six categories and whatever is in the middle of an image determines what category it is.

Steve said: "If a building is in the middle, it would be the built environment, if a lake was in the middle or the sea then

that would be water, if it was a railway it would be rail, if it was a road, it would be a man-made surface, and so on.

"Every single building, every single railway, split into 56 by 56 metres. And then extracted from our data, which is a monumental task."

An advantage OS holds is the existing quality of its detailed OS data sets, such as aerial imagery, height data and OS MasterMap. Combining all these together makes it possible to create labels for machines to understand.

Steve said: "For example we've used a technique for height data, where by subtracting our (digital) terrain model away from our (digital) surface model data, we've been left with base heights for everything in the country. We can tell if something is a building, a tree, if its ground, and then we can combine that with OS Mastermap and all the attribution that OS Mastermap has. "It's become a very powerful labelling system."

### Complications

The scale of the challenge is massive. OS has 1.1 million aerial images from its original topo model and will use STFC's supercomputer to scale this up to around 80 million images. Then it will be a case of comparing the two to see if the new networks can identify them.

Steve explained it was important each image class was balanced to prevent bias skewing results.

He said: "We have to pick and choose which images make it through to being representative of that class. To avoid bias we need to make sure we have equal numbers of urban areas, rural areas, natural zones and coastal. We need to make sure we don't have just houses representing buildings.

It has to be houses, schools, industrial areas, all these sorts of things. It's so if you are showing a network a huge building it doesn't go 'I've not seen a building that big before I am going to assume it's a road.'

He added: "We also need to make sure we avoid bias in geography. If you only show the machine England it might not correctly identify the same features in Scotland or Wales due to the diverse architecture and landscapes."

Another setback to overcome has been the fallout from the Covid-19 pandemic. While there have been unexpected benefits (the fact all three organisations were forced to work from home and hold monthly meetings remotely helped establish effective behaviours and working patterns), the downside has been the inconvenience of transferring physical hard drives and devices from one place to another easily. Having to work with large datasets on home networks or requiring IT help to turn machines on and off again has also caused frustration.

But the biggest complexity of all is getting machines to recognise data in a general way.

Izzy explained it is not about simply feeding the networks different customer questions one after another and getting them to act as a gopher to find the answer.

She said: "Because we are not trying to directly answer a customer question, it is very hard to say how well our network has been trained in terms of customer questions.

"You can't just do a simple accuracy assessment.

"There are lots of different ways of training networks but how on earth do you compare the outcome of each when you don't know ultimately how the network will be used in future?

"This project is going to come up with metrics that will allow us to compare different networks, so we can use that to decide whether one network is likely to be better than another one when it comes to ultimately trying to answer many, many different customer questions using it."



Aerial imagery captured by Ordnance Survey is combined with its OS MasterMap and height data to create labels that machines can understand. Photo: Ordnance Survey

### Global ambitions

OS has global ambitions for the project. If the machine learning proves accurate and begins identifying the patterns it needs to from the data to the high standards required, the aim is for this to be replicated and adapted to help nations and governments elsewhere in the world.

Steve said: "There's quite a few things this will benefit. We have an international presence that is going to be using a lot more machine learning. We can use this network as a baseline for quality and, also, we can run a shallow train to repurpose the network that we create to work in other conditions and other countries.

"There is a lot of work around inference and discovery, relating to how the machine has learnt what features make a building and using that to find other things it uses to identify features. Through that you can derive more data from existing data.

"So we hope to identify solar panels or where certain housing estates are, things like that."

The project may create further opportunities for automation across OS processes. Steve believed OS had developed a way of labelling a mass amount of data through this process which was going to be very handy for all the

other international and domestic parts of the business.

### Further options

"Other options we can look into include using the network for automated quality control," he added.

"So a tool for checking OS Mastermap imagery and how it is up to date, how they blend together.

"There are a lot of possibilities we need to explore."

Ultimately, the aim is for machine learning to develop into a useful tool that makes life easier for employees and customers.

It is a change of technology to make things faster and better. New skills and retraining will help staff adapt.

Izzy said: "It is very unlikely you would get an AI that was as versatile as a human at interpreting data.

"But what an AI can do is a very large scale of interpretation, leaving the more interesting bits to the humans to do.

"That is why we need to think about how we are using humans and make sure that we are giving people the work that most suits them. The interesting stuff, the problems around the edges, the really hard decisions.

"Done properly, it will improve the quality of people's jobs."

**David Jones is Media Executive at Ordnance Survey in Southampton, Hampshire (<https://www.ordnance-survey.co.uk/>)**

