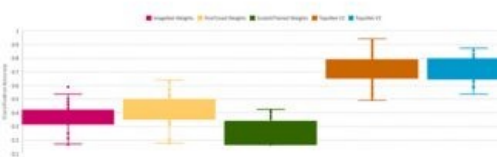


OS's Approach to Machine Learning for Inference and Discovery using Imagery



For 227 years, Ordnance Survey's role has been to create usable representations of the landscape to respond to the changing needs of our customers. We map physical aspects of the landscape such as the locations of buildings and field boundaries and the shape of the terrain. However, customers' needs are changing and becoming increasingly complex and bespoke. For a few years, we have been looking at novel approaches to responding to our customers' requirements.



The majority of OS's customers are within the UK Government. Through our engagement activities, we have identified a wish list of hundreds of real-world features and characteristics that our customers would like us to capture. It would be both expensive and time-consuming to create a new approach to

capture each item on the wish list and so we are working on an entirely different approach using deep learning. We believe our approach will enable the rapid development of bespoke products and services that respond to changing customer needs. It has not been a smooth ride from the early days of the idea and there is still a lot of work to do.

Exploiting Aerial Imagery

Ordnance Survey products are derived mostly from a combination of field survey and remote sensing. We can acquire aerial imagery during a long flying season which can last from March to November each year; being at least 25cm pixel resolution, these highly detailed images contain a great deal more information than we can possibly process. We know there is more that could be derived from our large archive of imagery and we are keen to find automatic means to extract new detail from this data.

For example, in figure 1, it is possible to identify the detail that we currently capture, such as roads, buildings and railway tracks, but different regions and types of development can be identified, such as differences in the type of housing or industrial land use. Shapes or materials of roofs can also be identified. The time of day and time of year can be inferred, based the vegetation, shadow and traffic. These things can be characteristic patterns that will manifest repeatedly in a large dataset. By creating methods to automatically find these repeating patterns in our data archive we hope to pull out considerably more detail than we are currently able. In finding these repeated patterns, we will be able to process the imagery not as an array of pixels, but as a set of descriptions of the landscape. We believe that landscape descriptors will enable us to directly respond to customer demands and discover new ways of representing the landscape – we call these two aspects of our work Inference and Discovery.

Background Research

Research in this field was started in 2015, when we set up a post-doctoral project at the University of Southampton, in conjunction with Lancaster University. At this stage, we only knew that representation learning had recently made some very large leaps in development, and that we wanted to know how we could apply it to our data.

Most of the advances in representation and machine learning used specialised datasets, such as the ImageNet dataset (Deng, et al. 2009), which comprises images of a few hundred pixels portraying objects (including animals), each image being labelled with the objects' name.

By contrast, our aerial images are big – several thousand rows and columns of pixels. Further, they are not 'composed' and so don't have a natural foreground and background composition, and there is no obvious label for an aerial image. We needed to determine how to apply

the technology that had been developed using datasets like ImageNet to our specific domain of data.

Over the course of the study, we tried different approaches to representation learning, settling on deep convolutional neural networks (DCNN), in which many layers of convolutional filters (sometimes called nodes) are learned using back-propagation. These networks are easier to create using one of the many frameworks that are now available that allow us to design networks more easily or re-use network architectures ‘off the shelf’. We tested several frameworks for deep learning before settling on Keras (Chollet 2018). We also developed approaches to cutting up the aerial imagery into patches (like those in figure 3) and labelling them using our topographic vector data. We used these labelled patches taken from imagery of Southampton to train our first deep network using the architecture known as AlexNet (Krizhevsky 2012), which has 13 layers, and then moved onto the 50-layer ResNet-50 (He, et al. 2015).

Imagery for Training DCNN

The most common approach to training convolutional neural networks is with thousands or millions of image-label pairs. The image is presented at the network input and ‘forward-propagated’ through to a final layer that predicts the image label. During training, the error of prediction is assessed and used to adjust the network weights so that during subsequent iterations the predictions become more accurate. One issue with deep networks is the requirement for so many labelled images. In many datasets, this labelling is done by hand. We realised that we were very lucky to have a vector dataset that parallels our imagery and so used this to label our patches of aerial imagery.

Training our first deep network required some heavy-lifting gear in terms of processing power (NVIDIA Titan X GPU), especially as we were trying to parallel the progress made in training deep networks with the ImageNet imagery by training with a similar number (1.2 million) of patches. The first ResNet-50 started with weights that had previously been learned using ImageNet, that were then adjusted, or ‘fine-tuned’, with our Southampton aerial imagery – firstly adjusting only the final layer or weights and then later allowing all the weights in the network to update. For comparison, we also trained a ResNet-50 ‘from scratch’ – that is starting with a randomised set of initial weights.

Responding to Customer Needs

Finally, we could start to challenge our belief that this approach would extract representations that would allow us to rapidly respond to customer needs. To test this, we identified problems that needed resolving to improve some of our capture processes. These included:

- the recognition of inland water,
- the differentiation of metalised and unmetalised roads; and
- the detection of playgrounds.

With a small dataset of patches giving positive and negative examples for these problems, we tested the two trained networks (‘FineTuned Weights’ and ‘ScratchTrained Weights’) against a network only trained with ImageNet data (‘ImageNet Weights’). We also obtained a 12 x 12 square of pixels directly from the centre of the image patch to produce a set of values of equivalent size to later layers in the deep networks so that we could compare the networks against the original image values (‘Image Values’). We obtained values for each of the three networks by forward-propagating the patches through the networks and extracting values at each filter. These values, and the values taken directly from the image patches, were then used as inputs to train shallow machine learning models (Support Vector Machines) to assess how good models learned using different sets of values performed against each of the problems.

We trained support vector machines with values taken from different layers of the networks and against each of the classification problems and repeated each of these with different random selections of the training and testing data. From this we obtained hundreds of classification results.

First Results

In our initial results, we didn’t expect to find that the ImageNet-trained network was, on the whole, producing better values for solving these problems than either of the networks that we had trained with aerial imagery. Despite this poor validation of our approach, we were sure that if good results were achievable with a network trained only with ImageNet, amazing results must be possible with a dataset that was in our domain – aerial imagery. Whilst the inference goal of our work didn’t seem to be going quite to plan, we also wanted to investigate how well our deep networks were discovering new ways of representing the landscape. Many commentators describe neural networks as ‘black boxes’ inside which the decision-making process is undiscoverable. However, alongside developments in deep learning approaches, there have been developments in interrogating what the networks are responding to. We have started looking at different approaches to interrogating the trained network, some of which look at what individual filters in the network respond to, and others which try to consider the activity of the entire network.

A reasonable assumption was that our training data was not representing the real-world well enough to train a deep network that detected the characteristic patterns in the landscape that we could interpret or would be useful for inference. We needed to improve this data.

Next, we looked at building a training dataset that represented the whole of Great Britain. We chose flying blocks from around the country and created more consistent data labels using principles based on the knowledge that deep convolutional networks’ filters respond most to the centre of the image patches (Zemel 2017). The resulting training patches, grouped into 12 equally-sized classes, certainly looked much more coherent, but had it been worth the effort?

We had improved our training time and refined our training data. We were ready to train a new deep network and run some tests, but first we needed one final tweak – we needed a name for our trained networks that could be incremented as we iterated through all our different options. We called the new network TopoNet and we now add a suffix that indicates the data and network version.

We carried out a test with a new set of problems based on building attributes – roof shape, roof material and whether there are solar panels on roofs. To ensure that TopoNet was benchmarked, we repeated these inference tests on a network trained only with ImageNet, as well as the FineTuned and ScratchTrained networks.

Better Results

It is hard to convey how keenly the results of this new set of inference tests were anticipated. They were to be the validation of our hunch that better training data was essential to creating a TopoNet that would help us achieve our goals. Our first sight of the inference results, summarised in figure 4, showed that our approach was consistently out-performing the ImageNet-trained network.

We are now working on an even better dataset for training TopoNet and looking at how we transfer our successes into a production operation. We're developing our disparate functions and scripts into a tracked and versioned repository, automatically recording data provenance, with principals that will ultimately support the reproducibility of our research. This repository is making it simpler for us to apply new versions of the training data and develop the network architecture so that, ultimately, we can produce a version of TopoNet that we are confident is capable of solving a range of inference problems and discovering new ways of portraying the landscape.

Future Work

We still have a long way to go in terms of building robust approaches to interrogating the trained network to understand if what it has discovered is meaningful, perhaps even useful. We also need to 'prove' what we are doing in an operational environment. But it is satisfying to look back and see how far we've come from a somewhat unformed idea in 2015.

Acknowledgements

We've been exceptionally fortunate to have been permitted to research the ideas before they were easily articulated and indebted to academics and PhD students to help us form the ideas into coherent research. In particular, Jonathon Hare and Iris Kramer, University of Southampton, Peter Atkinson and Ce Zhang, University of Lancaster, Dani Arribas-Bel, Melanie Green and Nikos Patias, University of Liverpool.

References

Chollet, François and others. 2018. Keras Documentation. 11 15. <https://keras.io>.

Deng, J, W. Dong, R, Socher, L.-J. Li, and K. and Fei-Fei, L Li. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." IEEE Computer Vision and Pattern Recognition (CVPR).

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. "Deep Residual Learning for Image Recognition." arXiv:1512.03385.

Krizhevsky, A. and Sutskever, I. and Hinton, G. E. 2012. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 1097-1105.

Zemel, Wenjie Luo and Yujia Li and Raquel Urtasun and Richard S. 2017. "Understanding the Effective Receptive Field in Deep Convolutional Neural Networks." arXiv:1701.04128.

This article was published in Geomatics World January/February 2019