Indianajones.ai: evaluating methods for mapping water management features across the South Indian Neolithic–Iron Age transition

Tanish Jain
Department of Computer Science
Stanford University
tanishj@stanford.edu

Haoran Shi
Department of Anthropology
Stanford University
hrs@stanford.edu

Abstract

We configured three convolutional neural networks (ConvNets) to perform semantic segmentation on high-resolution, multi-temporal satellite Imagery. Our best model correctly classifies seasonal reservoirs, which bear much archaeological significance in south Indian prehistory, with a balanced accuracy ($\bar{A}$) of 76% on the test set. The deep learning pipeline (including data pre-processing, augmentation, and model training) that we present here may provide insights for archaeologists and geospatial scientists alike who seek to map small features (*ca.* meters across) in arid conditions.

1 Introduction

A rock pool, or a *gnamma*, is a weathering pit formed during wet seasons (Figure 1); archaeologists have long speculated that these geomorphological features have afforded Neolithic–Iron Age communities the possibility to cope with deteriorating supply of, and perhaps increasing demand for, water across the Deccan region in South India [1]. Whereas their spatial distribution may contribute to our understanding of site growth, early state formation, and technopolitics in southern Indian prehistory, mapping these features is not always easy: they are often located on inaccessible inselbergs and usually too small (*ca.* meters across) to be discerned on satellite images — especially those available to public (and indeed to ill-funded archaeologists; *e.g.* Landsat 8 imagery).

We propose that deep neural networks, in particular Fully Convolutional Neural Networks (FCNNs), can be employed in conjunction with high-resolution, multi-temporal PlanetScope (PS) imagery to detect spectral signals of water that track a seasonal pattern in the Deccan region. We therefore implement semantic segmentation and produce per-pixel probability map that highlights these seasonal reservoirs.

2 Literature review

Multiple scholars have demonstrated the success of FCNNs — which typically generate pixel-wise output (or label map) following convolutional (conv), pooling, and unsampling layers — in detecting geomorphological features on remotely sensed data: notably, Wang et al. employed a pre-trained model to classify extra-terrestrial craters [2]. Similar to Palafox et al., Wang and his team proposed multi-scale receptive fields to tackle various sizes of landforms [3]; their binary classifiers are instrumental in the exploration outlined in this paper.

We recognized the unbalanced distribution of classes in our training data, and we referred to much work on the segmentation of medical images. In many cases, medical specialists emphasize recall (or sensitivity): Felzenszwalb et al., for example, adopted a 'hard negative mining' approach that reintroduces misclassified negatives in the training set [4]. This metric is useful in our analysis of rock pools, but we also want to consider the number of false positives: perennial reservoirs, often

CS 230 Deep Learning: project proposal 2020–2021 @tanishj @hrs, Stanford University, CA.

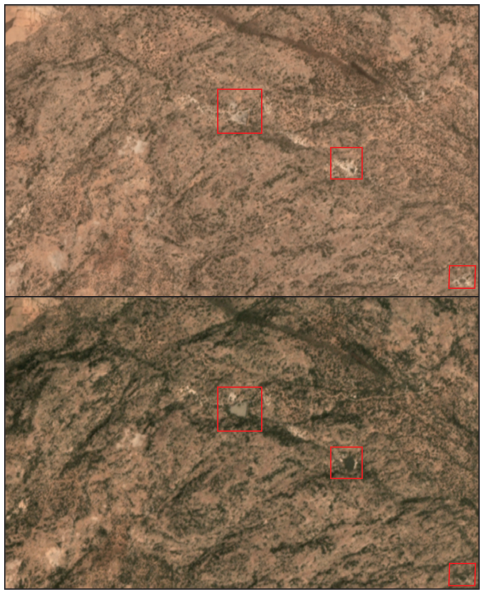

Figure 1: Multiple rock pools (red boxes) identified on 3-meter PlanetScope images, taken near (15.265N, 76.617E) in (*above*) dry season, April 2017 and (*below*) post-monsoonal season, October 2017. Scale: 1 : 3 000.

dated to the Medieval period, may be incorrectly classified as rock pools when they shrink in dry seasons. Câmara Neto et al., upon studying blood vessels in fundus image, proposed an alternative method to evaluate the performance of a model; balanced accuracy ($\bar{A}$), as defined below, provides a balanced weighting of sensitivity and specificity [5]. In this paper therefore we use $\bar{A}$ instead of the accuracy measure (which would be misleading for our dataset).

$$\bar{A} = \alpha \times sensitivity + \beta \times specificity \tag{1}$$

where α and β — two weighting factors — are set to 0.5, following the convention of Brodersen et al. [6].

3 Datasets

As shown in Figure 2, we manually labelled rock pools based on previous fieldwork in the Maski Archaeological Research Project (MARP) [7], Global Surface Water (GSW) data [8], and monthly composites generated from PS satellites [9]; 12 basemap quads — GeoTIFF files that contain 4096 $\times$ 4096 pixels — were fetched from a Web Map Tile Service (WMTS) server, covering six bands from two seasons (April 2017 and October 2017). In the 150 km^2 study area bounded by lat/long pairs: 76.4 E, 15.2 N and 76.6 E, 15.4 N, these basemaps were cropped into 512 $\times$ 512 images, where binary mask labels (including *ca*. 5,000 positive examples) were generated. 5% of the data are used for testing purposes.

To address our skewed dataset, we randomly selected crops that contained positive labels, rotating them by either 90, 180, or 270°(see Figure 2). As shown in Table 1, the data augmentation slightly improved the performance of our final model (ConvNet C — discussed below): despite improvement in F1 score, $\bar{A}$ only marginally increased. Nonetheless, we trained the model with the largest dataset.

4 Learning method

For the training purposes, we propose three ConvNet configurations, A–C (Figure 3): the ConvNet depth increases from left to right. ConvNet A, for example, consists of three blocks of encoders, which contains two or three convolutional layers with a receptive field of 3 $times$ 3 and a rectified

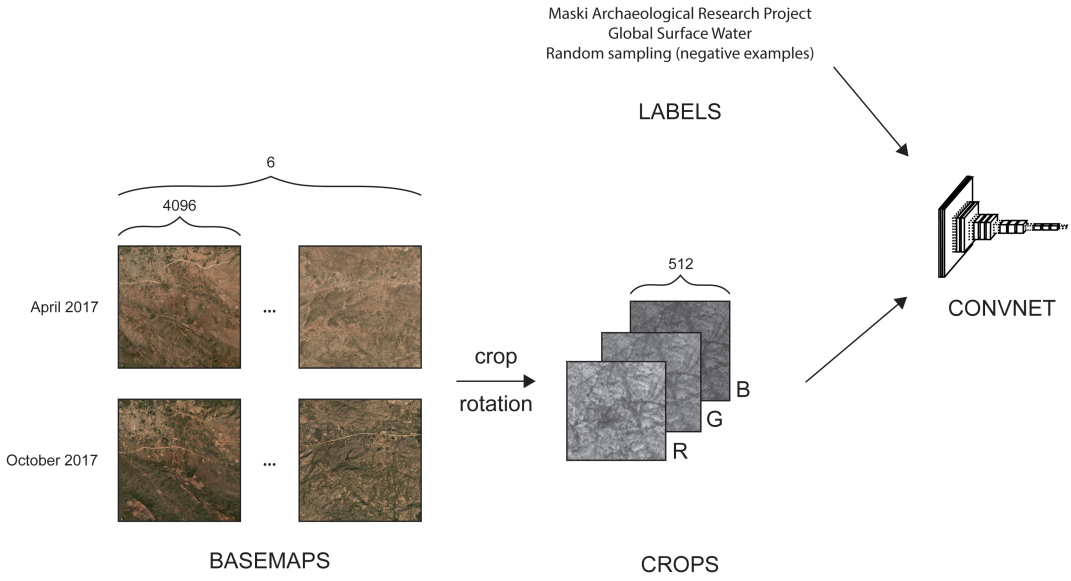

Figure 2: Data pre-processing and augmentation.

Table 1: Effects of data augmentation on model performance

$\times$ times of original size	Recall	F1 score	$\bar{A}$
2	0.586	0.551	0.793
5	0.604	0.664	0.802
10	0.602	0.704	0.801

linear unit (ReLU); a max pooling operation with stride 2 is used at the end of each block for downsampling.

A fully connected layer decodes the ConvNet; we then aimed to minimize the binary cross-entropy loss, defined as:

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} y_i \cdot log\hat{y}_i + (1 - y_i) \cdot (1 - \hat{y}_i) \tag{2}$$

where m is the number of examples, and y_i and $\hat{y}_i$ denote the i-th scalar value in the model output and label, respectively.

5 Implementation

The networks were trained using minibatch gradient-descent (MBGD) with a batch size of 2. Increasing the batch size did not seem to have an impact on the accuracy of the model, although smaller batch sizes used fewer computational resources and thereby resulted in faster training. A common problem with smaller batch sizes may be that the loss does not converge efficiently; however, we

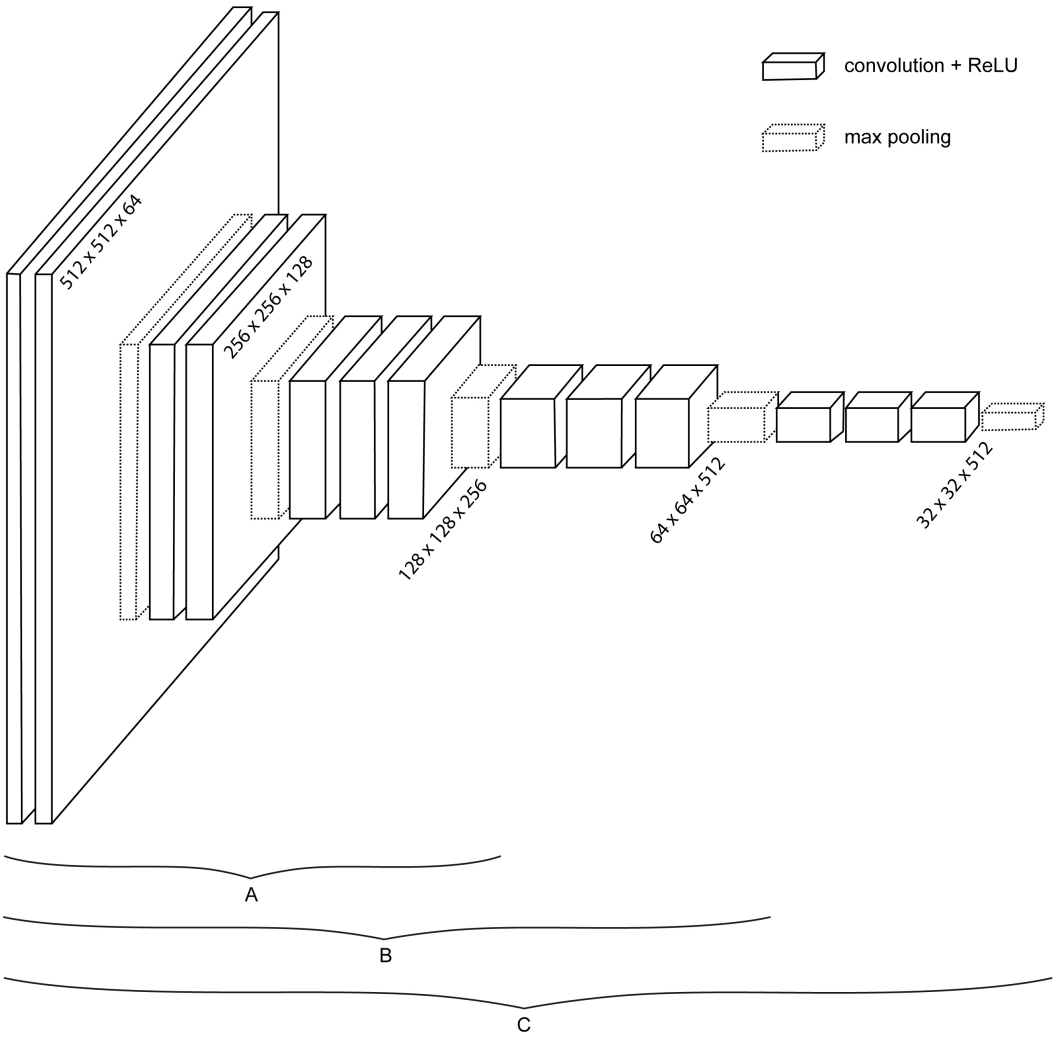

Figure 3: Three configurations of ConvNets; input and Dense layers are not included.

observed that, in this case, using a batch size of 2 resulted in a reasonably smooth loss curve while keeping the training time low. We used an ADAM optimizer with a learning rate of 0.001.

Table 2 was generated after we trained on each model for 200 epochs. The results are consistent with Simonyan and Zimerman's observation (2014) that, given sufficient data, a deeper ConvNet (in this case, C) outperforms its shallower counterparts. ConvNet C also shows high recall and precision compared to the other two configurations (53.3% vs 48.8% and 51.9%; 66.7% vs 51.7% and 59.8%). Furthermore, the results (*cf.* Table 2) indicate that our models are not subject to high variance despite the great number of learnable parameters.

Prediction maps were computed using ConvNet C: among the six correctly labelled rock pools shown in Figure 4, the classifier detected five. Qualitatively, false positives only occurred around the edges of these weathering pits; no perennial reservoirs were misclassified, and indeed this should give the archaeologist peace of mind.

Table 2: Effects of ConvNet depth on model performance

	Training set			Test set		
ConvNet configurations	Recall	F1 score	$\bar{A}$	Recall	F1 score	$\bar{A}$
A	0.488	0.598	0.744	0.378	0.437	0.689
B	0.519	0.624	0.759	0.463	0.522	0.732
C	0.602	0.704	0.801	0.524	0.587	0.762

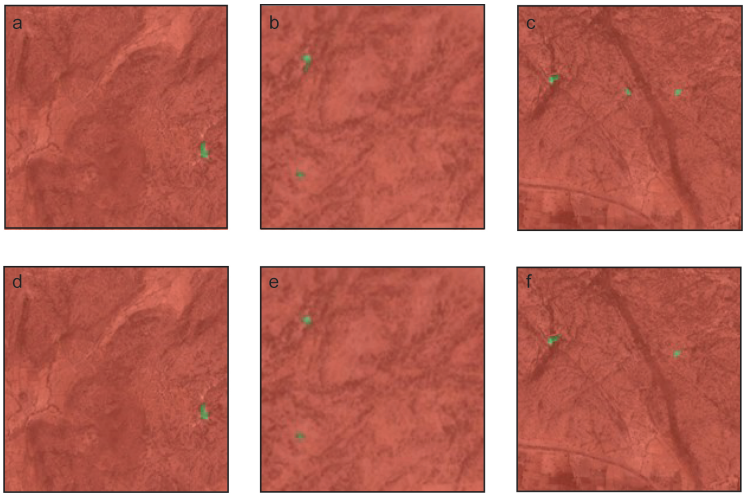

Figure 4: Prediction maps (d–f) vs groundtruth (a–c).

6 Next steps

Our deepest ConvNet achieved an $\bar{A}$ of 76% on the test set; the few false positives suggest that different receptive fields, as proposed by e.g. Wang et al., may be unnecessary for the application in question. The data augmentation however failed to improve our models significantly: it is possible to explore automated augmentation policies using, for example, Population Based Augmentation (PBA) [10].

In the future, we will work closely with archaeologists at Stanford: we will develop a model that is compatible with the Google Earth Engine (GEE) backend for image visualization and future prediction. The model will be able to estimate the location, density, and potentially volume of these rock pools, further contextualizing prehistoric settlement in south India.

7 Contributions

Jain implemented the convolutional neural network and analyzed the results of the model using statistical metrics for different network configurations and data augmentations.

Shi collected and manually labelled the data used for training. In addition, he used his domain expertise to visually analyze the results of the model. He also drafted the report and prepared the video.

We would like to thank Claudia A. Engel and Stace D. Maples for their support with the GSW and PS data. Andrew Bauer has generously offered his knowledge of South Asian archaeology and assisted our labelling process. We are also grateful for Shahab Mousavi's guidance and patience throughout the project: you made these stressful times less challenging for both of us.

Reference list

[1] P. Roberts, N. Boivin, M. Petraglia, P. Masser, S. Meece, A. Weisskopf, F. Silva, R. Korisettar, and D. Q. Fuller, "Local diversity in settlement, demography and subsistence across the southern Indian Neolithic-Iron Age transition: Site growth and abandonment at Sanganakallu-Kupgal," en, *Archaeological and Anthropological Sciences*, vol. 8, no. 3, pp. 575–599, Sep. 2016, ISSN: 1866-9557, 1866-9565. (visited on 03/17/2020).

[2] H. Wang, J. Jiang, and G. Zhang, "Crateridnet: An end-to-end fully convolutional neural network for crater detection and identification in remotely sensed planetary images," *Remote sensing*, vol. 10, no. 7, p. 1067, 2018.

[3] L. F. Palafox, C. W. Hamilton, S. P. Scheidt, and A. M. Alvarez, "Automated detection of geological landforms on Mars using Convolutional Neural Networks," en, *Computers & Geosciences*, vol. 101, pp. 48–56, 2017, ISSN: 00983004.

[4] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA: IEEE, 2008, pp. 1–8, ISBN: 978-1-4244-2242-5. DOI: 10.1109/CVPR.2008.4587597. [Online]. Available: http://ieeexplore.ieee.org/document/4587597/ (visited on 11/17/2020).

[5] L. Câmara Neto, G. L. B. Ramalho, J. F. S. Rocha Neto, R. M. S. Veras, and F. N. S. Medeiros, "An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images," *Expert Systems with Applications*, vol. 78, pp. 182–192, Jul. 15, 2017, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2017.02.015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417417300970 (visited on 11/17/2020).

[6] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *2010 20th International Conference on Pattern Recognition*, Istanbul, Turkey: IEEE, 2010, pp. 3121–3124, ISBN: 978-1-4244-7542-1. DOI: 10.1109/ICPR.2010.764. [Online]. Available: http://ieeexplore.ieee.org/document/5597285/ (visited on 11/17/2020).

[7] A. M. Bauer and P. G. Johansen, "The Maski Archaeological Research Project (2010–18): Initial results from a multi-period interdisciplinary project on the Raichur Doab, Karnataka, India," *Current Science*, vol. 117, no. 1, pp. 46–56, 2019, ISSN: 0011-3891. (visited on 02/24/2020).

[8] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, pp. 418–422, 2016, ISSN: 0028-0836, 1476-4687.

[9] Planet Labs Inc., *Planet basemaps product specifications*, 2018. [Online]. Available: https://assets.planet.com/marketing/PDF/Planet_Basemaps_Product_Specifications.pdf (visited on 02/11/2020).

[10] D. Ho, D. Liang, and R. Liaw. (2019). "1000x faster data augmentation," The Berkeley Artificial Intelligence Research Blog, [Online]. Available: http://bair.berkeley.edu/blog/2019/06/07/data_aug/ (visited on 11/17/2020).