The real tomb raider: A convolutional neural network application for remote sensing

Pablo Crespo

1 Introduction

Since 2014 the Swiss-Chinese Dzungaria Landscape Project has conducted surveys northern Xinjiang . A large diversity of monuments dating from the early Bronze Age to the ethnographic period was mapped during the first campaigns . Accumulations of gigantic early Iron Age burial mounds early on caught the attention of the researchers and it soon became clear that the southern Altai Mountains, especially the area around Heiliutan were a focus of intense funerary building activity especially during first millennium BCE. By far the most dominant anthropogenic features of the landscape were large burial mounds with circular ditches around them. They bore a striking resemblance with Saka burials from the Semirechye (eastern Kazakhstan) and the northern Tianshan.

The Xinjiang Uyghur Autonomous Region is known for its difficult access for foreign archaeologists. Successful fieldwork often depends on the ability to establish informal contacts to local decision makers and to navigate a complex social network. Sporadic eruptions of ethnic conflicts between the Uyghur minority and Han Chinese majority in southern Xinjiang carry with them the risk of last minute abort of long-planned projects. Militarized border zones further restrict access and even receiving a permit is not necessarily a guarantee that one is allowed to pursue the necessary work. Remote sensing mitigates these problems of access if high-resolution data can be acquired. But even open-source data like Google Earth is now of surprisingly good quality in many areas of northern Xinjiang. Hence, the ability of using classification algorithms to mitigate the problem of difficult access to these tombs via remote sensing is now more plausible than it has ever been.

By using the locations of tombs found in prior archeological digs, it is possible to use aerial pho-

Figure 1: *Tomb in the Heiliutan valley with clear morphological parallels to the northern Altai. North-south diameter 28.0 m, east-west diameter 29.5 m, well-preserved, large central depression (looted?), looking northeast.*

tographs to represent the problem of identification as a simply binary classification problem. We are only interested in whether a photograph contains a mound or not. Convolutional neural networks lend themselves well to the task since they are able to identify subtle features in photographs in each convolutional layer. In this work, we use the power of this methodology on data extracted using TerraIncognita and Google Earth to train a convolutional neural network. The results of the trained model are novel and encouraging regarding its use for eventual prediction and localization of tombs. The project for Udacity focuses in the training, testing and benchmarking of this network.

2 Data

The data used in the project was extracted using the Terra Incognita mapping tool with its access to Google Earth maps. Having prior knowledge from the locations of tombs from the Swiss-Chinese Dzungaria Landscape Project (special thanks to archeologist Dr. Gino Caspari) , the explored portion of the map was extracted and cut into 100×100 pixel images. Figure 1 shows one of these tombs exposed from the side. The figure makes clear why the monuments would be visible through aerial photography. The morphological similarities between tombs found through the project, as well as diameter, make them a good candidate for classification.

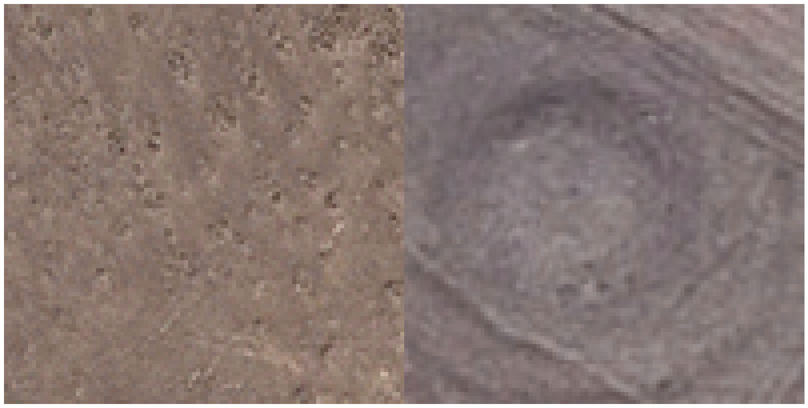

Figure 2: *Left: An example of a photograph labeled as "not mound." Right: An example of a photograph labeled as "mound."*

The extraction of the pictures ended with 1212 images, with 169 containing tombs and 1043 containing other patches of land. The dataset comprises all identifiable tombs from prior literature regarding physical archaeological expeditions. Figure 2 contains examples of howphtographs labeled as "mounds" and pieces of land labeled as "not mounds" appear in our dataset.

3 Classification methodology

Convolutional neural networks can detect specific patterns that are translation invariant in pictures, making it our methodology of choice. However, this does not come without challenges for our specific problem.Since each of the photographs have dimensions of 100×100 pixels, we have that the independent variables or features for a classification are $100 \times 100 \times 3$ arrays to account for the RGB channels. Considering that we have only 1043 of those arrays and only 169 of them are labeled as mounds, the dataset is relatively sparse for a multiparameter methodology such as convolutional neural networks. Hence, it is important to augment the data in order to perform a good fitting using a convolutional neural network. We use the Keras API for augmentation through generating data by transforming individual observations in our dataset.

The transformation step for generation included three steps:

- Shearing

- Zooming in slightly
- Flipping the image horizontally

Once the transformation process is complete, the second challenge faced by any convolutional neural network construction is the selection of the architecture. Since our problem is binary we code our dependent variable as:

$$Y = \begin{cases} 1 & if \text{ "mound"} \\ 0 & otherwise \end{cases}$$

We use two convolutional layers with RELU activations with a kernel size set to 32 and 3 strides on both directions, as a means to avoid the vanishing gradient problem. An extra RELU activated layer with a kernel size set to 64 is added to deal with larger possible trends in patterns. In addition, since the classification includes seemingly simple objects, three layers of RELU activations work well enough to assess even subtle patterns in the images. As the final layer we choose a sigmoid activation which generates the probability of a specific array being a mound. Furthermore, the prediction boundary allows for any probability above 0.5 to be labeled as 1 and 0 otherwise. In order to control for overfitting we use the dropout method and set the dropout rate to 0.5. We run the the fitting through 100 epochs with batches of size 16. The results of the fitted values for loss and accuracy (both training and validation) over 100 epochs are presented in Figure 3 and Figure 4.

From observing the graph in Figure 4 in particular we see that overtime the model oscillates for validation accuracy in between 0.93 and 1, and is concentrated around 0.98. This result becomes rather encouraging with the augmented data model. However, a more thorough evaluation needs to be performed in order to assess the goodness of fit of the model on testing data.

Figure 3: *Training and accuracy loss per epoch.*

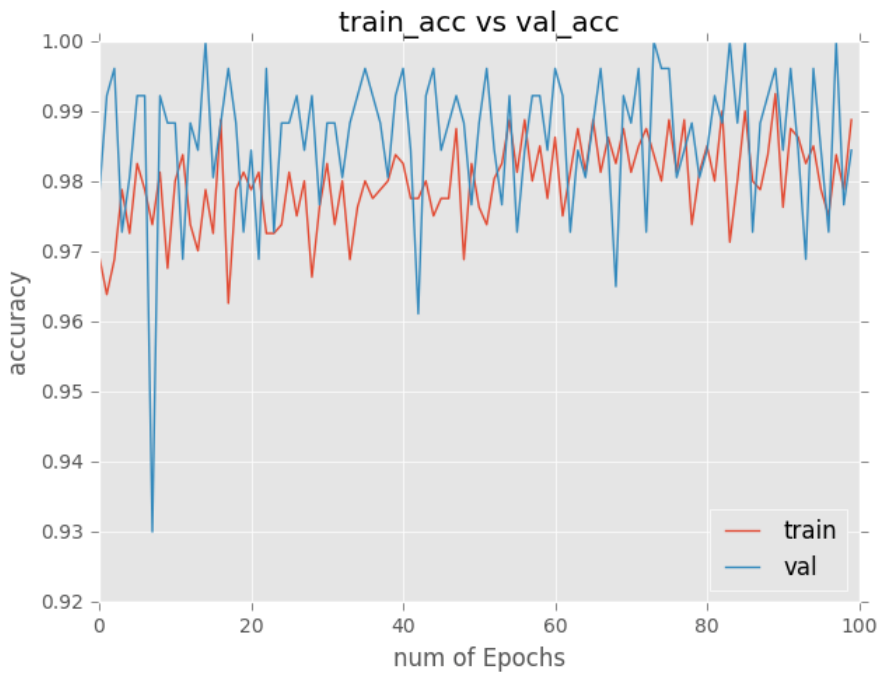

Figure 4: *Training and accuracy accuracy per epoch.*

	Model	Precision	Recall	$F1$ score
$Y = 0$	Random Guessing	0.89	1	0.94
	SVM with Linear Kernel	0.90	0.96	0.93
	SVM with RBF Kernel	0.96	0.97	0.97
	Convolutional Neural Network	0.98	1	0.99
$Y = 1$	Random Guessing	0	0	0
	SVM with Linear Kernel	0.29	0.15	0.20
	SVM with RBF Kernel	0.76	0.67	0.71
	Convolutional Neural Network	1	0.84	0.91
Average / Total	Random Guessing	0.79	0.89	0.84
	SVM with Linear Kernel	0.84	0.87	0.85
	SVM with RBF Kernel	0.94	0.94	0.94
	Convolutional Neural Network	0.98	0.98	0.98

Table 1: *Model evaluation metrics for random guessing, SVM classifiers with linear and radial basis function kernels, and convolutional neural network as specified in section 3 on a test dataset*

4 Detailed classification results and benchmarking

The preliminary results of fitting the convolutional neural network described in Section 3 seem encouraging. Yet, there is no reason to believe this is a comparably good model, and without further classification report metrics if it is a necessarily good model. Since the classification problem is binary and seemingly simple (mounds look very different to patches of land) then a "shallow" classification algorithm which can handle nonlinear decision boundaries is a good candidate as a competitor or benchmark model. As such we pick to use a support vector classifier with different kernels, with hyperparameters tuned with 5 fold cross-validation to control overfitting. The choices of kernel are the radial basis function and the linear kernel. Polynomial and sigmoid kernels were obtained, but the results were insufficiently good to consider them a good benchmark. Moreover, a random guessing classifier weighted by the relative frequencies of the labels as probabilities will serve as a secondary, yet common in practice and literature benchmark. The measures to use that reflect missclassified results are the usual metrics of precision, recall and $F1$ score. The reports in this work will only include the ones performed in the test dataset for expediency, simplicity and because these are the results that matter most when selecting a model for prediction using data that has not being seen in the fitting process. The results are compiled in Table 1.

Judiging the results from Table 1, we find that for all our purposes that the convolutional neural network outperforms all competing models as we have chosen them. Considering that we have unbalanced labels, having significantly more observations without mounds than those with, the panel in the table that needs to be brought to special attention is the one with $Y = 1$ on the title. It is of particular note that the convolutional network greatly outperforms all other models in the task of actually identifying mounds. The only close competitor is the SVM classifier with the radial basis function kernel. Yet the difference in scores is significant in each of the scores, between 0.17 to 0.24. Hence it seems that the architecture we have chosen works well enough for further implementation.

5 Conclusions

Given the difficulty to obtain training data due to the nature of the problem, and the possible problems caused by unbalanced labels, our proposed architecture is a plausible, tractable and appropriate methodology for mound identification through prediction. The main reason why it is possible for the model to do well is not only the data augmentation step for training, but the ability of CNNs to find subtle patterns in data arrays. The lack of need for flattening, and the shared weights can more easily deal with problems of identification that might have issues like translation variance. Moreover, the availability of Google Earth data of sufficient quality to yield high precision scores via prediction with the trained model allows for further work in the project to the point of allowing mound identification matched with location (latitude and longitude) as a possible tool for archeologists to organize future surveys or prepare for appropriate permissions for digs as well as finding tombs before looters do.