Application of CNN in classifying the structure of ancient ceramics

Pan Lei
12010919
MAE

Chen ZhaoRui
12010906
CS

Abstract

With the continuous development of AI technology, more and more industries are interlinked with AI, resulting in many fantastic reactions and bringing about huge industry changes. In archaeological technology, the use of convolutional neural network (CNN) algorithms for archaeological classification and analysis is necessary due to the gradual lack of talent and the high level of knowledge required. There are currently some attempts in different archaeological fields. Most of the attempts are relatively simple attempts at classification. In the case of ceramic structure classification, CNN is a highly accurate accessible and effective method of classifying ceramic structures that can be applied at a larger scale due to its high degree of texturization.

1. Introduction

In archaeological activities, the sorting and reorganization of large quantities of ceramic sherds, or the extraction of features or the analysis of their composition, requires a great deal of time and qualified expertise, which inevitably leads to errors in the sorting process due to various subjective factors. The introduction of convolutional neural networks as a tool for classifying ceramic structures is therefore necessary, not to abandon manual classification, but to combine the various expertise of different experts to achieve more objective and efficient classification tasks and more accurate results. As CNNs continue to evolve, more and more learning sets are being developed, but they have yet to be exploited in depth in the field of archaeology. Convolutional neural networks, with their highly discriminative nature for texture, are uniquely suited to classifying the composition of rock flakes.

In recent years CNN has been applied experimentally in various fields in the direction of archaeology, but all changes are the same, in a nutshell, surface features are extracted by scanning and then the extracted features are refined by machine learning methods for classification. The structural iterative specificity of different network models has different adaptations for processing archaeological data of smaller orders of magnitude.

2.1. An analysis of the context of ancient ceramic research in archaeology

The classification of ancient ceramics or ceramic sherds is of great importance in archaeology. Firstly, the carved or patterned surface features allow the local cultural characteristics of the ceramic's origin to be identified, which is of great importance for the study of cultural diffusion. Furthermore, the analysis of the composition of ancient ceramics, combined with the distribution of minerals in their place of origin, allows for an analysis of the geological composition of the period, which can also shed light on plate movements. In addition, the study of the classification of ceramics from different regions is also important for the study of the spread and distribution of ancient trade routes. Because the huge amount of classification work and the need for high accuracy make this a very time-consuming and labor-intensive task. CNNs are a very efficient and promising solution to this problem, which can be solved with the help of other autonomous analysis techniques.

2.2. The main methods of classifying ceramic structure

There are two main methods of classifying ceramics, one of which is to classify the macroscopic features on the surface of the ceramic, such as carved patterns, texts and the content of paintings. This approach requires ceramics and ceramic sherds to be complete, reflecting essentially the full range of features, and in the case of sherds a further layer of classification criteria is required to integrate sherds containing similar features. This is the most intuitive way of classifying ceramics and is less prone to error by chance, but for a time when painted ceramics or carving techniques were not widespread, this intuitive way of classification is not so easy, and it is difficult to analyse the colour of the fragments of clay alone to arrive at an efficient form of classification.

Another type of classification that is now more commonly used is the analysis of micro-composition components. This takes the form of thin slicing of ceramics, cutting them off and using a microscope and superimposing different polarised light to obtain a two-dimensional image, a three-colour map or a greyscale map, which allows a good analysis of the composition, with three main components: inclusions, clay matrix and gas interstices, in another way, the analysis of the glass phase, the mineral crystalline phase and the gas phase. This is essentially the same criterion as for the archaeological classification of rocks, for the reason that ceramics are also a man-made rock. Once the sections have been obtained, there are two types of experiments: direct textural and modal analyses, and indirect chemical analyses using different chemical properties. This classification method has the advantage of sufficient quantitative data and a high degree of objectivity, but is not very efficient, with a large number of slices and statistics. It also requires the extraction of ceramic structures, which can cause a certain amount of per se damage, especially when the data set is not quite sufficient with a small sample size. Rather, this classification, which is more dependent on the size of the dataset and has more objective and clear conditions for analysis, is more suitable for CNN use than the other convenience of 3D modelling that can be scanned for analysis.

3. Introduction to Convolutional Neural Network Models

3.1 Basic Concepts of Convolutional Neural Networks

3.1.1 Basic concepts

Convolutional neural networks (CNNs) are a class of feedforward neural networks that include convolutional computation and have a deep structure, and are one of the representative algorithms of deep learning. Compared to other deep learning structures, CNN can give better results in image and speech recognition. This model can also be trained using back propagation algorithms. Compared to other deep, feed-forward neural networks, CNN require fewer parameters to be considered, making them an attractive deep learning structure.

CNN consist of these main types of layers: an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer. The input layer is capable of processing multidimensional data and normalising the input data. The convolutional layer consists of three parts: the convolutional kernel, the convolutional layer parameters and the activation function, which works by regularly sweeping through the input features, multiplying them by matrix elements and superimposing deviations within the perceptual field. The convolutional layer parameters include the convolutional kernel size, step size and padding, which together determine the size of the convolutional layer's output feature map and are the hyperparameters of the convolutional neural network. The convolutional layer also contains an excitation function to assist in the representation of complex features. After feature extraction in the convolutional layer, the output feature map is passed to the pooling layer for feature selection and information filtering. The pooling layer contains a pre-defined pooling function that replaces the results of a single point in the feature map with the feature map statistics of its neighbouring regions. The pooling layer selects the pooled regions in the same way as the convolutional kernel scans the feature map, controlled by pooling size, step size and padding. The fully connected layer in a convolutional neural network is equivalent to the implicit layer in a traditional feedforward neural network. The fully-connected layer is located at the end of the implicit layer of the CNN and only passes signals to the other fully-connected layers. The feature map loses its spatial topology in the fully connected layer and is expanded into a vector and passed through the excitation function.

3.1.2 The history of CNNs

CNNs can be traced back to Hubel and Wiesel's 1968 paper on how the visual cortex of cats and monkeys contains neurons that respond individually to small areas of the visual field, and that if the eye does not move, the area of visual space where a visual stimulus affects a single neuron is called its Receptive Field. Neighbouring cells have similar and overlapping receptive fields. The size and location of the receptive fields vary systematically between cortices, forming a complete visuospatial map. This research laid the foundation for local perception in CNNs.

In 1980, the neocognitron was proposed, marking the first initial CNN and the first application of the receptive field concept in the field of artificial neural networks. The neocognitron decomposes a visual pattern into many sub-patterns (features), which then enter a hierarchically progressively connected feature plane for processing.

In 1988, the Shift-invariant neural network (SNN) was proposed, taking the CNN's capabilities one step further, enabling it to accomplish recognition even when objects are displaced or slightly deformed.

The feed-forward architecture of the CNN is extended by lateral and feedback connections in the Neural abstraction pyramid. The resulting recurrent convolutional network allows flexible incorporation of situational information to iteratively resolve local ambiguities. In contrast to previous models, the highest resolution image output is produced.

After the appearance of a GPU implementation of CNNs in 2005, which marked a more efficient way to implement CNNs, CNNs stood out for their high accuracy in the ImageNet competition in 2012, and deep learning was officially brought into the limelight.

Next we will describe several well-known CNN models.

3.2 AlexNet

AlexNet was designed by Hinton and his student Alex Krizhevsky (2012), and the model won the ImageNet competition in 2012. This model is the basis for many better and deeper neural networks.

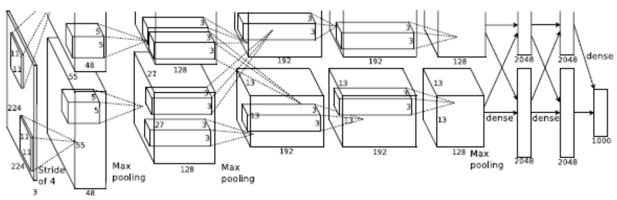

Figure 3.1 Structure of AlexNet

AlexNet contains several relatively new technical points:

--Successfully using ReLU as the activation function of CNN and verifying its effectiveness over Sigmoid in deeper networks, successfully solving the gradient dispersion problem of Sigmoid when the network is deeper.

--Dropout was used to randomly ignore some neurons during training to avoid overfitting the model.

--Maximum pooling using overlap in CNNs.

--A LRN layer is proposed to create a competition mechanism for the activity of local neurons, making the values in which the response is relatively large and suppressing other neurons with smaller feedback, enhancing the generalisation ability of the model.

--Using CUDA to accelerate the training of deep convolutional networks, the powerful parallel computing capability of the GPU is utilised to handle the large number of matrix operations during neural network training.

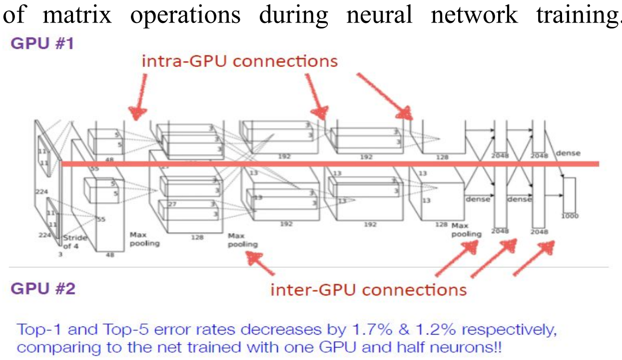

Figure 3.2 Multi-gpu parallel training of AlexNet

--The data enhancement, randomly intercepting a 224*224 sized region (and horizontally flipped mirror) from the original 256*256 image, corresponds to an increase of 2*(256-224)^2=2048 times the amount of data. Also, the AlexNet paper mentions that it will do PCA (Principal Component Analysis) process on the RGB data of the image and do a Gaussian perturbation of the principal component with a standard deviation of 0.1 to add some noise, and this Trick can bring the error rate down by another 1%.

3.3 VGGNet

VGGNet was developed in 2014 by researchers from the University of Oxford's Computer Vision Group and Google DeepMind, and achieved second place in the classification event and first place in the localization event of the ILSVRC 2014 competition. VGGNet can be seen as a deepened version of AlexNet, but in a simpler form.

3.3.1 Structure

VGG consists of 5 convolutional layers, 3 fully-connected layers, and 1 softmax output layer. The layers are separated from each other using a maxpool (maximisation pool), and the activation units of all hidden layers use the ReLU function. In the original paper, the authors designed six network structures, A, A-LRN, B, C, D and E, depending on the number of different sub-layers of the convolutional layers.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 3.3 Structure of VGGNet

The six networks are similar in structure, all consisting of five convolutional layers and three fully connected layers, the difference being that each convolutional layer has a different number of sub-layers, increasing from A to E. The total network depth ranges from 11 to 19 layers. The convolutional layer parameters in the table are expressed as "conv (receptive field size) - number of channels", e.g. con3-64 means a 3x3 convolutional kernel with 64 channels; maximum pooling is expressed as maxpool, with layers separated by maxpool; fully connected layers are expressed as "FC - number of neurons", e.g. FC-4096 indicates a fully connected layer containing 4096 neurons; and finally a softmax layer.

In the form, D denotes the well-known VGG16 and E denotes the well-known VGG19. For example, the structure of VGG16 is shown in the following figure:

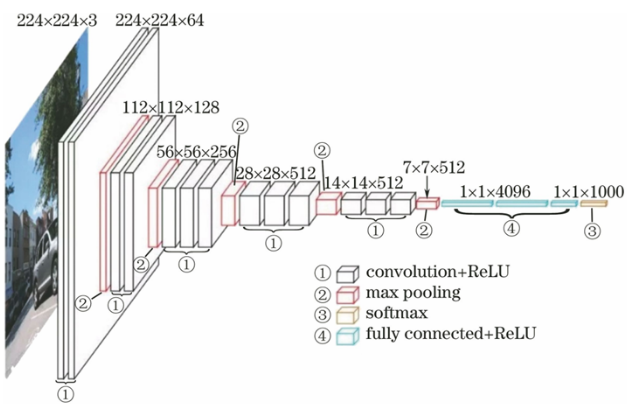

Figure 3.4 structure of VGG16

3.3.2 Model advantages

Compared to previous CNN models, VGGNet has the following advantages.

--Simple structure: Although VGGNet has more layers, its overall structure is not complex. As mentioned above, VGG consists of five convolutional layers (each with a different number of sub-layers), three fully-connected layers, a SoftMax output layer, and a maxpooling (maximization pool) separating the layers from each other, and a ReLU function for the activation units of all the hidden layers.

--Small convolutional kernels: Instead of using the larger convolutional kernel size in AlexNet (e.g. 7x7), VGG achieves the same performance by reducing the size of the convolutional kernels (3x3) and increasing the number of convolutional sub-layers (VGG: from 1 to 4 convolutional sub-layers, AlexNet: 1 sub-layer)

--Smaller pooling kernels: compared to AlexNet's 3x3 pooling kernels, VGG uses all 2x2 pooling kernels.

--More channels, wider features: each channel represents a Feature Map, more channels means richer image features. 64 channels in the first layer of the VGG network, doubled in each subsequent layer up to 512 channels, the increased number of channels allows more information to be extracted.

--Deeper layers: By using successive small 3x3 convolutional kernels instead of large ones, the network is deeper and fills in the edges, the convolution process does not reduce the image size. Also, only small 2x2 pooling units are used, reducing the size of the image.

--Fully connected layer to convolutional layer: this is a tested method proposed by the authors that enables the network model to accept image sizes of arbitrary size.

--Parameter stability: The depth of the six network structures, A, A-LRN, B, C, D and E, has increased from 11 to 19 layers, but the number of parameters has not changed much, due to the fact that basically small convolutional kernels are used and the parameters are mainly concentrated in the fully connected layer.

3.4 ResNet

ResNet was designed by Microsoft Research in 2015 and won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation in ILSVRC & COCO 2015 competitions.

Empirically, the depth of the network is critical to the performance of the model, and when the number of layers is increased, the network can perform more complex feature pattern extraction, so theoretically better results can be achieved when the model is deeper. However, the experiments revealed a Degradation problem in the deep network: when the depth of the network increases, the accuracy of the network saturates and even decreases. ResNet solves the Degradation problem by learning the residuals for a stacked layer structure (several layers are stacked together) when the input is x. The learned features are denoted as H(x), now We want it to learn the residual F(x) = H(x) - x, so that the original learned feature is actually F(x) + x. This is because residual learning is easier than direct learning of the original features. When the residuals are 0, the stacking layer is only doing constant mapping at this point, at least the network performance does not degrade, and in fact the residuals will not be 0. This will also allow the stacking layer to learn new features on top of the input features and thus have better performance. The structure of residual learning is shown in Figure 3.5. This is somewhat similar to a "short circuit" in a circuit, so it is a shortcut connection.

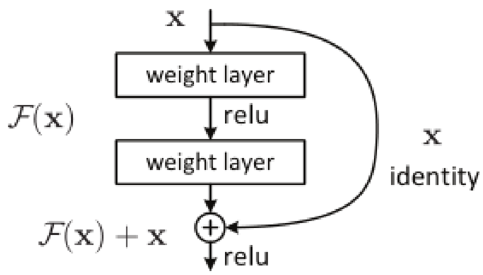

Figure 3.5 Residual learning: a building block

The ResNet network is a reference to the VGG19 network, modified from it and incorporating a residual unit through a short-circuiting mechanism, as shown in Figure 3.6.

An important design principle of ResNet is that the number of feature maps is doubled when the feature map size is halved, which maintains the complexity of the network layers. As can be seen in Figure 3.6, ResNet adds a short-circuiting mechanism between every two layers compared to a normal network, which creates residual learning, where the dashed line indicates that the number of FEATURE MAPs has changed. The 34-layer ResNet, shown in Figure 3.6, also allows for the construction of deeper networks as shown in Table 3.1. As can be seen from the table, for the 18-layer and 34-layer ResNet, it performs residual learning between two layers, and when the network is deeper, it performs residual learning between three layers, with three convolutional kernels of 1x1, 3x3 and 1x1, respectively. The number of feature maps in the implicit layer is relatively small and is 1/4 of the number of output feature maps.

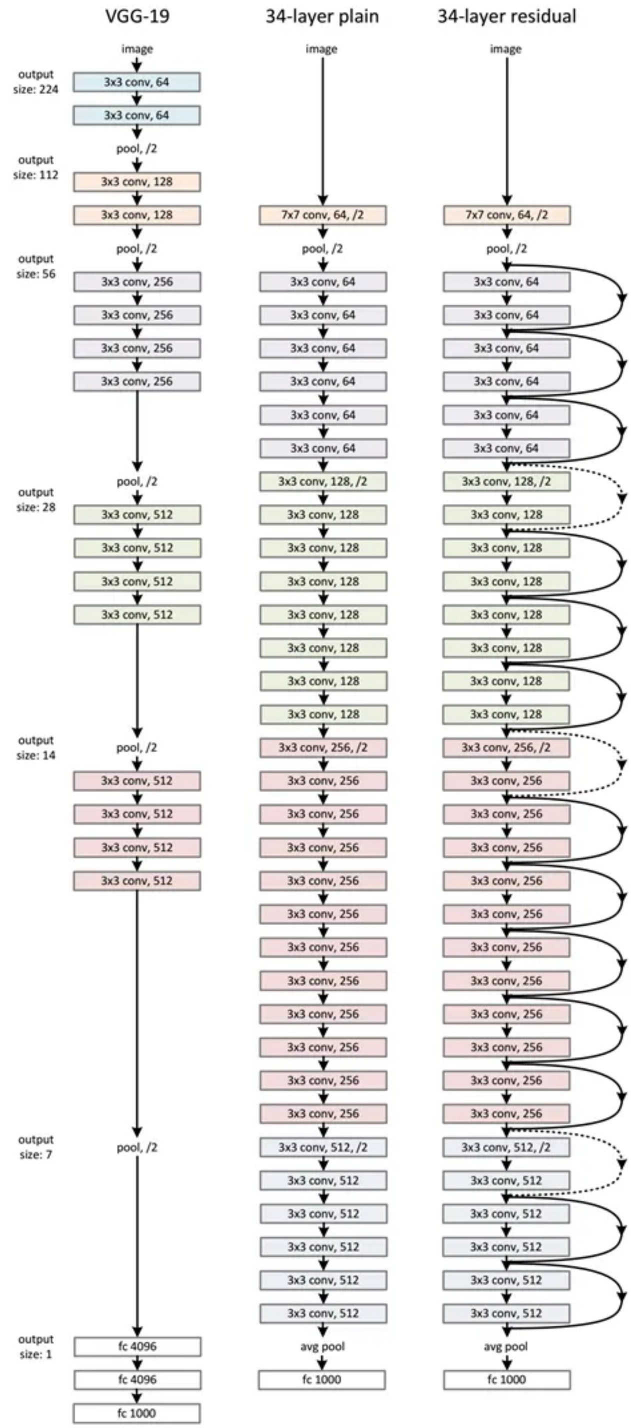

Figure 3.6 Example network architectures for ImageNet.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Table 3.1. Architectures for ImageNet.

ResNet uses two types of residual units, as shown in Figure 3.7. The left figure corresponds to a shallow network, while the right figure corresponds to a deep network. For short-circuit connections, when the input and output dimensions are the same, the input can be added directly to the output. But when the dimensions do not coincide (corresponding to a doubling of the dimensions), this cannot be added directly. There are two strategies: (1) using zero-padding to increase the dimensionality, which generally requires a downs amp first, and can be done by pooling with stride=2, which does not increase the parameters; (2) using a new mapping (projection shortcut), which generally uses a 1x1 convolution, which increases the parameters and also increases the computational volume. Short-circuit connections can, of course, use projection shortcut in addition to the direct use of constant mapping.

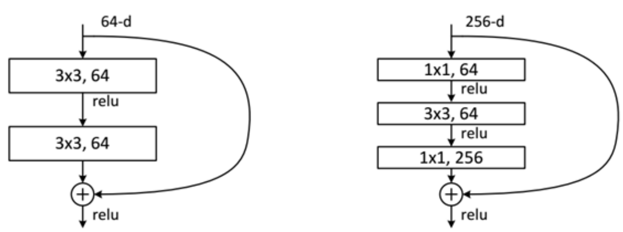

Figure 3.7 Example network architectures for ImageNet.

4. Feasibility analysis of convolutional neural networks applied to ceramic classification

As artificial intelligence continues to develop, advances in convolutional neural networks make it possible to use them on an increasingly broad level. In the archaeological direction, the large sample size and high level of knowledge required for feature recognition make such specialists in short supply, which makes the use of convolutional neural networks a strong necessity. Currently, there are already experts in different archaeological fields who have used the current popular network models for their applications and obtained good results.

4.1. Accuracy and confidence analysis of simple models that have already been applied

Bogacz and Mara (2020) used the ResNet50 network model in CNN when classifying cuneiform texts from different periods. A 3d model was first modelled for the different periods of cuneiform writing, a point cloud was formed and then machine learning was used and compared to accurate classification and an accuracy of 84% was obtained. Although the sample set for machine learning is not very large and the CNN is not very sensitive to macroscopic shapes, especially sharp corner turns, which can make the point cloud distribution not very accurate, a high accuracy rate was still achieved.This attempt also has significant implications for ceramic archaeology, particularly in terms of the macroscopic approach to classification, and the same applies to the carved patterns and painted surfaces of ceramics.

Figure 4.1 Example of cuneiform writing

In addition to this, there are many applications on a larger macro level, such as geographic surveys, for example, the location of graves and the determination of site zones.

Pawlowicz and Downum (2021) uses image recognition technology, developed using CNN, to identify the shape of various ceramic surface patterns. The researchers classified Tusayan white pottery from northeastern Arizona, which has a distinctive black and white patterned surface, and with the recognition system in operation, obtained an accuracy rate that was essentially the same as, and in some cases better than, that of an expert human discriminator.

Figure 4.2 Tusayan white ceramic

The purpose of the project is essentially the same as the one above, the ARCADIA Project (Chetouani et al. 2018; Chetouani et al. 2020). Designed to enhance the archaeological heritage of the ceramic sherds extracted in Saran (France). These sherds date back to the heyday of the Middle Ages. They propose to use a convolutional neural network (CNN) model to automatically classify these ceramic shards. The ultimate goal is to form clusters of shards that result in a map representing the movement of potters. In order to reduce the workload caused by modelling, the surface texture of the ceramics is printed directly on white paper by means of ink stamping, and then scanned to extract the corresponding texture. As this method is not directly applied to relic antique ceramics, damage to the ceramic surface does not need to be considered.

(a) (b) (c) (d)

Figure 4.3 Image processing steps

For the subsequent classification of simple ceramic texture shapes, a variety of current CNN network models were used for the experiments (ResNet18, VGG11, Alex Net) and the best performing network model was then selected and evaluated under different configurations. The results obtained show the relevance of the method and the best results are obtained with the ResNet18+SVM configuration.

The project above does not involve a specific project application. Pawlowicz and Downum used deep learning to quickly sort sherds according to texture in order to better classify Tusayan white porcelain.

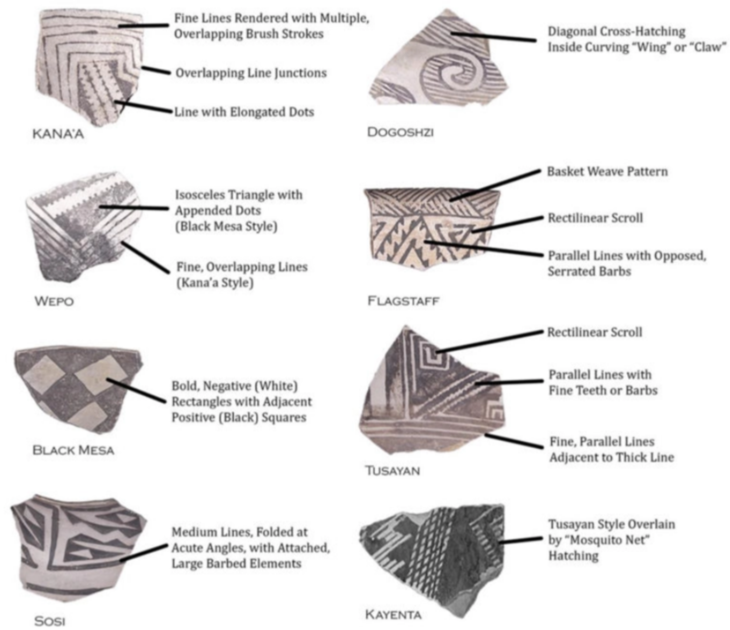

Figure 4.4 Texture classification of Tusayan white ceramics

In their study, Pawlowicz and Downum showed that, if trained properly, deep learning models can assign types to digital images of decorated cattle with an accuracy comparable to, and sometimes higher than, that of four expert contemporary archaeologists. Pawlowicz trained the AI model used for the study in just a few hours on his PC, which was equipped with an NVIDIA GPU running a pair of common convolutional neural network models, VGG16 and ResNet50. Far from competing with existing archaeologists, the model could even be a valuable tool for training new archaeologists.

Results from existing applications show that CNN holds great promise in terms of its ability to extract recognizable surface features, and that accuracy rates can be high.

4.2. Feasibility analysis of ceramic micro composition analysis and classification using CNN techniques

In the previous section of the technical background examination, most of the techniques used the simplest CNN models to classify the macroscopic structure of ceramics, and the application of CNN is not very mature for the direction of component analysis of microscopic slices. However, scientists have been using artificial neuronal network (ANN) techniques to classify ceramic flakes as early as ten years ago.

Aprile, Castellano, and Eramo (2014) constructed three network structures to separately identify the three different phases in the rock thin section. They approach combines image analysis techniques with artificial neural networks to automatically classify mineral inclusions and pores in archaeological pottery sherds using optical digital images. In particular, the automatic identification of quartz, calcareous aggregates and secondary porosity was considered. A collection of planar and cross-polarized light images obtained by means of a digital camera connected to an optical microscope in transmitted light was used.

It is shown that reliable features obtained by considering the optical properties of the class of interest and the specific neural network structure can provide good automatic classification.

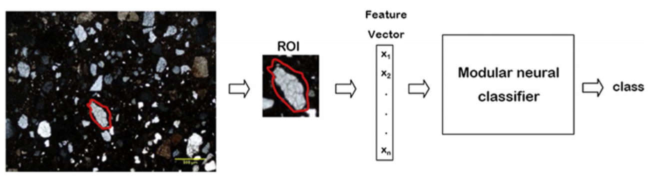

Figure 4.5 Flow chart for classifying ceramic slices using ANN

The insights from the above studies have led to the application of CNNs in the field of slice structure classification as well. Lyons, M. (2021) have built a deep learning model to automatically identify and classify ceramic fabrics, speeding up the classification process and reducing the need for experts. Thin-section images of rocks in cross-polarized light from the Coconut Tree period (1000-1525 AD) archaeological site in Guadalupe, on the northeast coast of Honduras, were classified using two divisions trained by VGG19 and ResNet50. Although the sample size and learning set are not large, the result of 93+% classification accuracy in four tests demonstrates an efficient classification method, and although this use case is narrowly defined, it serves as a proof of concept that suggests the method should be broadly applicable to any ceramic assemblage of definable ceramic fabrics.

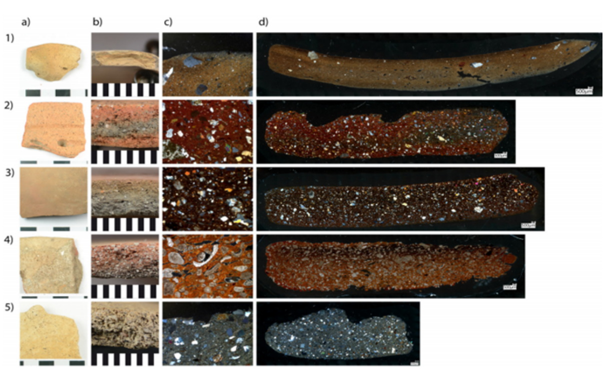

Figure 4.6 Examples of the five ceramic fabric types analyzed.

The unique sensitivity of convolutional neural networks to texture discrimination makes them an excellent predisposition for the analysis of ceramic flake compositions. Previous research has applied different network models not only for microscopic classification, but also for macroscopic characterization of ceramic fragments, all with high accuracy, in a way that reduces misclassification due to subjectivity, while allowing for simultaneous multi-threading to improve efficiency and solve the problem of insufficient scholarly reserves. In summary, CNNs have a strong need and a promising future in the direction of ceramic classification.

4.3. Limitations and defects

For many archaeological applications, where ML is an aid to more detailed work, such analyses may be sufficient. However, they may not be as reliable as traditional expert methods because they cannot yet handle the range of variability and inconsistency in archaeological data.

Often the biggest obstacle to constructing good ML models is that they work best when they are built on large databases of information, such as thousands of catalogued images or material from reliable sources, which can be difficult to achieve, particularly in terms of archaeological budgets and the diversity of data that may be available. Poor preservation of archaeological recovery samples makes the task more difficult as fragmentation and surface condition (including, for example, erosion, patina and vegetation cover) can affect the success of identification. Experts can often identify material that is difficult for ML models trained on ideal collections to handle.

Another aspect of ML is that models are very much a product of the data on which they are constructed. As such, these models tend to be classified according to the categories they know, which makes them (at least) susceptible to two main forms of bias.

The first involves focusing a combination into one of the previously identified categories. This means that rare and unusual objects can easily be categorized as more common types and ignored. For example, a ceramic vessel similar in shape to one of the models may look 'normal', but may have an unusual surface treatment that an archaeologist would immediately notice.

The second form of bias, and perhaps the most common, is that the model does not fully incorporate the variability of the features being categorized; ML analysis can easily lose the 'forest for the trees' because the data used to train the model is often stripped of contextual information (especially in the case of images) or manipulated on a limited number of pre-selected variables, which may not include sufficient variables to allow for the classification of the model. These variables may not include enough information to distinguish important (i.e. archaeologically relevant) categories.

ML techniques do have methods for checking 'performance', but these methods still rely on internal mathematical metrics and require the attention of archaeological research to ensure that they provide good results. Outside of archaeology, difficulties with ML have been repeatedly encountered, including exacerbating racial and gender bias in commercial settings (Gebru 2020).

5. Conclusion

From the above summarized study, we found that CNNs strongly tend to recognize texture rather than shape, and that ceramic fabrics are highly textural in nature. This is precisely the reason why CNN models can fit with the topic of classifying the structure of ancient ceramics. In addition, several studies have shown that CNNs are highly accurate in classifying ceramic structures, a conclusion that provides a piece of mind for researchers in the industry who have questioned the rigour of CNN technology.

As deep learning techniques continue to evolve, the accuracy and efficiency of neural network models are being optimized and improved. Computer-related practitioners are happy to see this happen, and there is a growing willingness in other industries like archaeology to use such techniques to aid their scientific work. We are seeing newer and more optimized convolutional neural network models being used in many of the latest archaeological research projects. The use of computers and big data is clearly more efficient than traditional manual excavation and identification. We expect archaeological research to make even greater breakthroughs with the help of technology.

References:

[1] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[2] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

[3]http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/

[4] Bogacz, B and Mara, H. 2020. Period Classification of 3D Cuneiform Tablets with Geometric Neural Networks. In: 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2020. IEEE, pp.246–251.

[5] Caron, M, Bojanowski, P, Joulin, A and Douze, M. 2018. Deep Clustering for Unsupervised Learning of Visual Features. In: Ferrari, V, Hebert, M, Sminchisescu, C and Weiss, Y (eds.). Computer Vision – ECCV 2018. Cham: Springer, pp. 139–156.

[6] Caspari, G and Crespo, P. 2019. Convolutional Neural Networks for Archaeological Site Detection – Finding "Princely" Tombs. Journal of Archaeological Science, 110: 104998.

[7] Cau, M-A, Day, PM, Baxter, MJ, Papageorgiou, I, Iliopoulos, I and Montana, G. 2004. Exploring automatic grouping procedures in ceramic petrology. Journal of Archaeological Science, 31(9): 1325–1338.

[8] Cheng, G and Guo, W. 2017. Rock images classification by using deep convolution neural network. Journal of Physics: Conference Series, 887: 012089.

[9] Chetouani, A, Debroutelle, T, Treuillet, S, Exbrayat, M andJesset, S. 2018. Classification of Ceramic Shards Based on Convolutional Neural Network. In: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, pp. 1038–1042.

[10] LeCun, Y, Boser, B, Denker, JS, Henderson, D, Howard, RE, Hubbard, W and Jackel, LD. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1(4): 541–551.

[11] LeCun, Y, Bengio, Y and Hinton, G. 2015. Deep learning. Nature, 521(7553): 436–444.

[12] Pawlowicz, LM and Downum, CE. 2021. Applications of Deep Learning to Decorated Ceramic Typology and Classification: A Case Study using Tusayan White Ware from Northeast Arizona. Journal of Archaeological Science, 130: 105375.

[13] Pires de Lima, R, Bonar, A, Coronado, DD, Marfurt, K andNicholson, C. 2019a. Deep convolutional neural networks as a geological image classification tool. The Sedimentary Record, 17(2): 4–9.

[14] Whitbread, I. 1989. A proposal for the systematic description of thin sections towards the study of ancient ceramic technology. In: Maniatis, Y (ed.). Archaeometry, Proceedings of the 25thInternational Symposium. Amsterdam: Elsevier, pp. 127–138.

[15] Wright, H and Gattiglia, G. 2018. ArchAIDE: Archaeological Automatic Interpretation and Documentation of cEramics. In: Workshop on Cultural Informatics co-located with the EUROMED International Conference on Digital Heritage 2018 (EUROMED 2018). Nicosia, Cyprus: Zenodo.

[16] Yosinski, J, Clune, J, Bengio, Y and Lipson, H. 2014. How Transferable Are Features in Deep Neural Networks? In: Proceedings of the 27th International Conference on Neural Information Processing Systems – Volume 2. NIPS'14. Montreal, Canada: MIT Press, pp. 3320–3328.