Semantic segmentation of ancient ceramics based on hyperspectral data

Gerard Graugés Bellver

Resum – Les ceràmiques han jugat un paper clau en el desenvolupament de l'arqueologia. Són els artefactes més habituals que es troben als jaciments arqueològics a causa de la seva composició que els fa molt duradors. Ser capaç de datar-los i classificar-los és una part essencial de la feina dels arqueòlegs. Aquest projecte se centra a classificar, a través de segmentació semàntica, imatges basades en dades hiperespectrals de ceràmiques de diferents regions i èpoques. Les dades hiperespectrals són útils per distingir entre diferents materials mitjançant la seva signatura espectral, que és el que s'utilitza per a la classificació. El sistema proposat s'ha dividit en tres mòduls. El primer mòdul consisteix en un algorisme de reducció de dimensionalitat basat en el "Principal Component Analysis" (PCA). En segon lloc, un mòdul de classificació i finalment un mòdul d'evaluació. El mòdul de classificació consisteix en dos algorismes d'aprenentatge computacional: el "Random Forest" i "Support Vector Machines", i un algorisme d'aprenentatge profund: la xarxa neuronal U-net. Els resultats obtinguts han demostrat que la U-net té els millors resultats i és capaç d'aprendre a classificar les ceràmiques agrupant-les segons trets comuns com categoria o origen. La segmentació semàntica ha permès implementar classificació per fragment, que ha incrementat lleugerament els resultats aconseguits per segmentació semàntica de la U-net.

Paraules clau – Imatge hiperespectral (HSI), signatura espectral, aprenentatge profund, reducció de dimensionalitat, Principal Component Analysis (PCA).

Abstract – Ceramics have played a key role in the development of archaeology. They are the most common artefact found on archaeological sites because their composition makes them very durable. Being able to date and classify them is an essential job for archaeologists. This project focuses on classifying through semantic segmentation images based on hyperspectral data from a number of ceramics from different types, regions and ages. Hyperspectral data is useful to distinguish among different materials through their spectral signature, which is used in the classification. The proposed system has been divided in three modules. The first module consists on a dimensionality reduction algorithm based on a PCA. Second, a classification module and finally the evaluation module. The classification module consists on two machine learning algorithms: the Random Forest and the Support Vector Machines, and a deep learning algorithm: the U-net neural network. The results obtained showed that the U-net has better results and is able to learn to classify the ceramics by grouping them based on common traits such as category or origin. Semantic segmentation has allowed to implement classifying by fragment, slightly improving the results obtained with semantic segmentation through the U-net neural network.
Keywords – Hyperspectral image (HSI), spectral signature, deep learning, dimensionality reduction, Principal Component Analysis (PCA).

1 INTRODUCTION

The craft of pottery started around 9000 years ago [1], back when humans first discovered that using fire to burn an abundant material, called clay, transformed it into a new hard and durable material: the ceramic. They discovered that giving shape to raw clay and burning it would allow them to create containers for storing food and water, as well as decorative objects, among other uses, [2,3].

• Contact e-mail: gerard.grauges@autonoma.cat
• Mención realitzada: Computació
• Project tutorized by: Daniel Ponsa Mussarra (Ciències de la computació)
• Academic year 2021/22

Ceramics have been essential in the development of archaeology, [4]. This artefacts are the most recovered class on the historical sites because they are able to survive on the ground due to their materials properties. Although they are commonly found only as fragments. They are easily datable, providing not only information about their period, but also about topics such as the social and economic status of their owners, technological changes and many others.

The techniques and materials used to build ceramics have improved and evolved ever since. Nowadays, ceramics are still used, but the process to make them is far more sophisticate. A mixture of different materials is used in order to get the best quality and ensure its durability. The evolution of the development of ceramics has not only varied through the ages, but also depending on the region where they were made.

An interesting characteristic from these materials is that they have different spectral signatures, which is the variance of their reflectance with respect to different wavelengths. The aim of this project is to develop a system capable of classifying ceramics from different parts of the world and different ages by analysing their reflectance in hyperspectral images.

A hyperspectral image is composed of hundreds of bands which are collected at different wavelengths for the same spatial area. It measures the spectrum of the light for each pixel of the scene at different wavelengths. These image bands conform a hyperspectral cube, where two dimensions represent the spatial area and the third one the spectral content, (Figure. 1). Hyperspectral images will be useful for our project because they provide how light behaves for each different material, meaning that we can identify each material by their spectral signature.

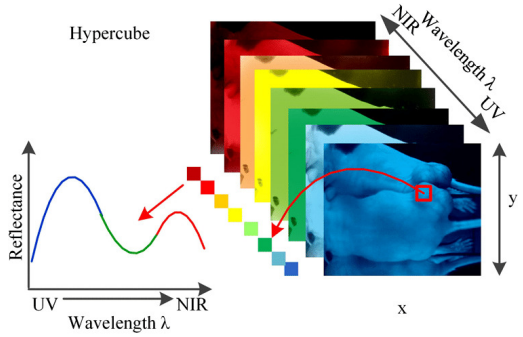

Fig. 1: Hyperspectral cube representation [5]

The issue that arises when using hyperspectral images is that the high dimensionality of the data brings the problem called the curse of dimensionality. This problem causes that with a small number of training samples, the accuracy of the classification decreases when the dimensionality increases. Acquiring a large number of training samples can be very costly and time-consuming. In order to solve this, we can use methods that aim to select the most suitable bands, called band selection, or combine them to compute a reduced number of features, called band compression, to reduce the spectral dimensionality.

This paper is organized as follows: section 2 reviews the state of the art algorithms that have been proposed so far of the project domain. Section 3 explains the methodology that has been followed to develop our system, presents the dataset used in our work and details the algorithms considered to build the system. The experimental work developed and the analysis of the results are presented in section 4. The last section describes the conclusions extracted from the experiments and results and the future work.

2 State of the art

For the sake of designing a proper approach to the introduced ceramics classification problem, a bibliographic research has been done. There are different parts involving this classification problem. On the one hand data reduction algorithms are necessary to deal with the curse of dimensionality problem and reduce the computational complexity required to process hyperspectral images. On the other hand, a robust classifier algorithm must be parameterised to assign each ceramic sample to its corresponding class.

This section briefly introduces different approaches on how to deal with hyperspectral imaging. First some papers aiming at classifying ceramics will be reviewed. Then some of the dimensionality reduction algorithms used with hyperspectral images will be described, followed by analysing different classification algorithms suitable for the project's purposes. Despite being in different subsections, classification algorithms can also include the dimensionality reduction on their own structure, being a part of them.

2.1 Ceramics classification papers

Some papers related to classifying ceramics have been found. The paper from George L. Miller, [6], tried to classify ceramics form the 17th and 18th centuries. This approach was based on classifying the ceramics using the decoration from them. The work in [7] implemented mathematical and computational tools for morphological description, classification and analysis of ceramics from the Iron Age. All image-based system found in the bibliography are based on standard color images, no paper has been found trying to classify ceramics using hyperspectral images.

2.2 Dimensionality reduction

In order to deal with the problem of the curse of dimensionality the most common approach is applying dimensionality reduction methods. Algorithms on this topic have also been studied for data compression, since they lead to removing redundant features, reducing the storage space required and faster transmission and processing time.

There are two different types of approaches which are reviewed in the following subsections.

2.2.1 Band compression

Band compression algorithms, also called feature extraction algorithms, aim to reduce the dimensionality of the data by transforming the original data into lower dimensional feature space with a defined criteria, [8].

Principal Component Analysis (PCA) has been widely used when dealing with unsupervised learning. This technique and a significant number of variants like PCA-

grounded correlation-based Segmented-PCA (SPCA) or Spectrally-Segmented-PCA (SSPCA), among others, have been developed to tackle the curse of dimensionality [9]. The PCA creates new uncorrelated variables, through linear combination, maximizing the variance so that the minimum information is lost.

Linear Discriminant Analysis (LDA) is used for supervised feature extraction. This algorithm has good results when data can be linearly divided. Otherwise, when data is non-linear it is used the kernel discriminant analysis (KDA), [10]. LDA differs from PCA because it maximizes the separability between classes, aiming for the best representation for each one instead of aiming at the variance between classes.

Compression algorithms can also be used as feature extractors for hyperspectral images. 3D-DCT (Discrete cosine transformation) is a transform-based algorithm that converts the raw pixels into a frequency domain using cosine and inverse cosine transformation functions in the three dimensions from the image, [11]. Another example of a compression algorithm used as feature extractor is the autoencoder [12]. It is a deep learning approach that is similar to PCA. However, the autoencoder is used because it is able to learn non linear transformations that are more adaptable to the data.

2.2.2 Band selection

The objective of band selection algorithms is selecting a subset of representative bands, from the set of bands of the image, to achieve a similar or better performance as using the original bands from the hyperspectral image in classification. This approach is feasible because, with hyperspectral data, there are always bands with redundant data. Moreover, there can be certain bands that have more relevant information than others depending on the objective of the classification problem.

A ranking based method called Minimum Misclassification Canonical Analysis (MMCA), derived from the Fisher's discriminant function, was proposed in [13] to rank bands according to their classification abilities. The same work also proposed the application of clustering based methods, such as k-means or the Affinity Propagation clustering algorithm (AP). The latest aims to select the most suitable bands to be the initial centroids of the clustering, taking into account the correlation and similarity between bands, instead of choosing the initial centroids randomly like the k-means.

A deep learning approach for band selection has also been proposed in [14]. The strategy designed is a Band Attention Module (BAM) included in a CNN architecture devoted to semantic segmentation. The proposed BAM consists of five 3x3 2D convolution layers and two 1D convolution layers. This module infuses global information learned from an image mask to increase the importance of the areas that are more likely to improve the accuracy. Although it has been referenced in this section, it is implemented within a classification neural network, in this paper being a VGGNet.

2.3 Classification approaches

Classifiers are those algorithms that aim to describe the category of new observations based on characteristics that have extracted from training data. Some widely used algorithms are Support Vector Machines (SVM) or Random Forests (RM). However, deep learning approaches are being used more and more because they have been able to further improve the results obtained compared to classical classifiers.

Classification can be implemented to classify at different levels. The first level is when algorithms classify the whole image with a specific label. The second level is to identify objects on an image and label them, whereas the last one consists on labeling each pixel from the image, task denoted as semantic segmentation.

To classify ceramics, there are two viable approaches, classification by fragment or by pixel. We don't consider classifying the whole image because it could contain fragments from different types of ceramics. On this project the chosen approach has been semantic segmentation. It has been selected because it is not always possible or easy to separate the fragments from the background in an image. Additionally, after the semantic segmentation fragments can be classified using the mean label of each pixel that belong to the same fragment. Several papers on this topic have been studied, which use hyperspectral images based on Convolutional Nerual Networks (CNN) models for the semantic segmentation task.

The work in [14] proposed a neural network following the structure of an eight-layer VGGnet. It was proposed a simple neural network architecture, but this could be also implemented with a CNN, ResNet or DenseNet.

CNN neural networks are usually used when considering the spatial information from the hyperspectral images. On the other hand, the CNN3D model was designed in order to take into account both the spatial and spectral information from the hyperspectral images, in the work of [15]. 3D convolutions are able to read the spatial information, as well as the band dimension of the image. They are usually used for medical images and action recognition in videos.

A hybrid neural network combining both CNN and CNN3D was developed in [16], called HybridSN. This consists of a CNN3D followed by a CNN, where the CNN3D joins the spatial-spectral information that is then learned by the CNN, which reduces the complexity of using a CNN3D alone.

A Multiscale Convolutional neural network (MCNN) was developed to learn the deep features of spatial relationships, [17]. It constructs a pyramid structure that present the spatial features at different scales, and then are concatenated with the spectral features to create a dataset for logistic regression.

A framework called self-taught semi-supervised autoencoder (SuSA) is also used [18]. This framework is made of two modules, one that extracts the spatial-spectral features and the second one that classifies them.

The work in [19] also studied the viability of the U-net neural network and some variations of it. The variations employed in this paper were the residual U-net, that benefits from the residual learning. Then the attention U-net,

which incorporates attention gates (AG) that highlight regions of interest and suppress irrelevant backgrounds. The last variant was the recurrent residual U-net, it had attention gates at each block from the architecture.

3 METHODS

This section introduces the methodology followed during the project. Then we can find a description of the dataset and the proposed system that has been developed for the classification task.

3.1 Methodology

A standard methodology has been followed in order to develop the project. First of all, numerous papers regarding similar topics have been studied to gain more knowledge of the subject, which have been analysed in the previous state of the art section. After reviewing the papers, the algorithms and techniques that could better adapt to the problem were chosen and implemented. The implementation step was followed by experiments to test the results from the algorithms and the analysis of the results.

The workflow mentioned was designed to achieve the objective of the project, which was to develop and train a model capable of classifying and distinguishing ceramics from different ages and zones from hyperspectral images. To do so, the project has been divided in different sub-objectives.

- Curate the dataset: Acquisition of a created dataset that was preprocessed to ensure the quality of the data.
- Select and implement dimensionality reduction methods.
- Select and implement classic algorithms for classification.
- Select and implement deep learning classifier.
- Select and implement performance evaluation methods.

3.2 Dataset

The dataset used for the ceramics classification was obtained through a collaboration between the ”Institut Català d’Arqueologia Clàssica” (ICAC) and the ”Centre de Visió per Computador” (CVC) from the ”Universitat Autònoma de Barcelona”.

The dataset consist on hyperspectral images from different kind of ceramics. The ceramics are between 17 different classes to classify. This classes can be grouped depending on different criteria such as region, kind of ceramic, age, etc. From the 17 classes, the first two, A and B, are the same kind: ”Campaniana”. The following 8 classes, from C to K, belong to the same kind ”Terra Sigillata”, but can also be separated according to their region. As well as L and M, which are more distinctive because of their region rather than the kind of ceramic they are. N, P and Q are from the same kind ”Àmfora” also differing on their region. Finally, R and S are the same kind of ceramic, but have different ages, being R from the ancient Rome and S contemporary.

There are classes which have more images than others, despite that, it was ensured that there would be the same number of fragments of ceramic per class: 30. However, having the same number of fragments doesn’t mean that there are the same number of pixels. In Figure 2 we can see the number of pixels per class, as well as the name for each class and the letter that represents it. The fragments per each ceramic have been divided in different sets because they couldn’t fit on the same image. For each set of fragment, three images were taken: one from the front, one from behind and the last one from the side.

The whole dataset has a total of 174 hyperspectral images from different types of ceramics of dimensions 1000*512 pixels and 224 bands. Out of all the images from the dataset, approximately 93% of the pixels are from the background while the 7% left correspond to the ceramics. For each image, a black and white mask was created indicating each fragment and the class they belong to.

The images were normalized using a white paper as a reference. Despite that, some images presented issues in certain pixels because they didn’t have the reference. In addition, Tipp-Ex had to be used to deal with imperfections on some fragments of ceramics, this meant that those pixels from the Tipp-Ex presented normalized values over the maximum of the reference. All this pixels were labelled as background on their respective masks.

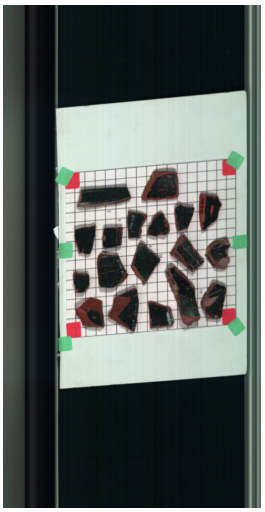

Fig. 3: Example of ceramic image, from the class ”Campaniana A”.

For the experiments, the dataset has been divided in three different sets, and we can see the division of classes that has been done on Table 1.

The camera used for the creation of the dataset was the Specim FX10, [20]. The camera can capture the spectral range of 400-1000 nm, with a high spatial resolution of 1024 pixels and 224 spectral bands of 5’5nm of spectral resolution.

3.3 Proposed system

The proposed system is conformed by three modules. First a dimensionality reduction algorithm, second a classification algorithm, which has been divided into two different versions to evaluate different algorithms of classification,

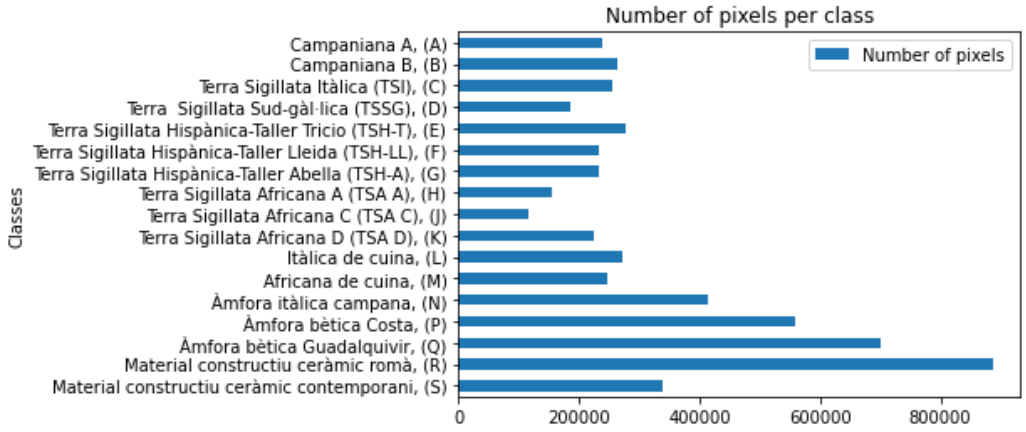

Fig. 2: Graph showing the classes of ceramics and the number of pixels per class

Class	Train	Validation	Test
A	3 (111117)	2 (77729)	1 (50471)
B	7 (187180)	1 (49608)	1 (27607)
C	4 (105219)	4 (118478)	1 (32724)
D	4 (129317)	1 (13362)	1 (42378)
E	5 (171966)	3 (89889)	1 (15734)
F	5 (122301)	3 (79044)	1 (31705)
G	5 (117159)	3 (81251)	1 (35744)
H	1 (28445)	4 (89439)	1 (36000)
J	4 (75512)	1 (11357)	1 (28717)
K	3 (91558)	5 (114097)	1 (17845)
L	6 (177391)	2 (60400)	1 (34367)
M	6 (164710)	2 (45716)	1 (34879)
N	10 (351014)	1 (44762)	1 (18786)
P	11 (429613)	3 (104625)	1 (23731)
Q	9 (337687)	8 (315144)	1 (47793)
R	16 (570077)	7 (285470)	1 (32823)
S	5 (233759)	3 (81658)	1 (22090)

Table 1: NUMBER OF IMAGES AND NUMBER OF PIXELS PER CLASS.

and third, the evaluation metrics.

Dimensionality reduction algorithm: PCA. To implement this module, the PCA has been chosen because it is a widely used dimensionality reduction algorithm. The PCA is a technique that evaluates the correlation between different variables, and then projects these variables into a new dimensional space where they are no longer correlated. The objective of the PCA is to trade a little accuracy for simpler data. The covariance matrix is used to find the correlation between variables. Once the covariance matrix is calculated, it is used to compute the eigenvectors and eigenvalues. They are then used to project the linear product of the variables that give the most information into the new reduced dimensional space.

We wanted to compress the spectral information from ceramics, to do so it was computed the mean and covariance matrix from the ceramic pixels of all the dataset. Because of the expensive computational cost, this calculation of the mean and covariance matrix was done incrementally. The combined mean was calculated using the following formula:

$$\bar{x}_c = \frac{n_1 * \bar{x}_1 + n_2 * \bar{x}_2}{n_1 + n_2} \tag{1}$$

where $\bar{x}_c$ is the combined mean, $\bar{x}_1$ is the mean of the spectral information of a set of ceramic pixels and $\bar{x}_2$ the mean of the spectral information of a second set of ceramic pixels, n_1 is the number of pixels in the first set and n_2 is the number of items in the second set.

To compute the covariance matrix first the variances of each variable are calculated following (Eq. (2)). Then the covariance matrix is calculated using (Eq. (3)).

$$S_c^2 = \frac{n1 * [S_1^2 + (\bar{x}_1 - \bar{x}_c)^2] + n2 * [S_2^2 + (\bar{x}_2 - \bar{x}_c)^2]}{n1 + n2} \tag{2}$$

where S_c^2 is the combined variance, S_1^2 is the variance of the first set of data and S_2^2 the variance of the second set, $\bar{x}_c$ is the combined mean of the two sets.

$$Cov = \frac{n1*[S_1^2+(\bar{x}_1-\bar{x}_c)*(\bar{y}_1-\bar{y}_c)]+n2*[S_2^2+(\bar{x}_2-\bar{x}_c)*(\bar{y}_1-\bar{y}_c)]}{n1+n2} \tag{3}$$

where Cov is the combined covariance, $\bar{x}_c$ is the combined mean of the two sets, $\bar{x}_1$ is the mean of the first variable of the first set and $\bar{y}_1$ is the mean of the second variable of the first set, the same way, $\bar{x}_2$ is the mean of the first variable of the second set and $\bar{y}_2$ is the mean of the second variable of the second set.

Classification algorithms. It has been divided into two approaches. First two classic algorithms have been trained in order to establish a baseline of results and define a reference. The second approach is a deep learning algorithm, which has been chosen because deep learning has proven an improvement of performance compared to classical approaches, achieved by extracting features that optimize the classification.

Both classification algorithms types will be implemented to do semantic segmentation. We have chosen semantic segmentation because this project wants to serve as a base to be able to use images from archaeological sites to find and classify ceramics. Different backgrounds, dirt and because ceramic fragment's can come in different shapes and sizes would make it too difficult to recognize whole ceramic fragments.

The classical algorithms selected have been Random forest and SVM. These two classic algorithms have been chosen because they are both widely used and well-known for their results with supervised classification.

The deep learning algorithm chosen has been the U-net. The U-net architecture has been the one presented in [21]. It was selected because it was designed for semantic segmentation and has been proven very efficient in fields such as bio-medicine or with satellite images for Geo sensing. It consists on two parts which are an encoder network followed by a decoder network. The encoder is used to project the input image into feature representations. Then, the learnt feature representations are projected in the pixels space in the decoder, restoring the full image.

The U-net classifier receives as an input the images that have the dimensionality reduced with the PCA. The ground truth correspond to the classes for each image and has been one-hot-encoded. Additionally, a model checkpoint callback has been implemented so that it stores the best model during training.

A focal loss function has been developed to more precisely assess the error from the neural network [22], because this loss function is designed for imbalanced classes. This loss function will be beneficial for the results because it focuses on the wrong prediction classes instead of the ones that have been predicted correctly. We are interested in this loss because it prioritizes the wrong predictions, which will help predicting those ceramics that are harder than the rest.

Evaluation of the model. To evaluate the results obtained from the classifiers, two metrics have been used. The metrics chosen were precision and recall. Precision computes the percentage of correct pixels that have been labeled from all the pixels classified to that class. The recall calculates from all the pixels from a certain class, how many were correctly classified. They have been chosen because they provide relevant information to assess the quality of the classification algorithms.

4 Experiments and results

In this section, we describe the experiments conducted, as well as the objectives intended and the results obtained from them.

There are two different sets of experiments. The first set consists on training the two classical classification algorithms that have been previously mentioned, the Random Forest classifier and the Support Vector Machine (SVM). These algorithms have been trained with a sampling from each class of ceramics. The second set consist on the training of the deep learning approach, the U-net neural network.

The first experimental results obtained have been the principal components from the PCA algorithm. Both sets of experiments have been trained using images after their dimensionality had been reduced through the PCA. We selected the first 5 principal components because we wanted to preserve 96.5 % of the whole variance, (Figure. 4). So, out of the 224 initial bands, they have been transformed into 5 selected principal components to reduce the dimensionality of the images while preserving most of the information.

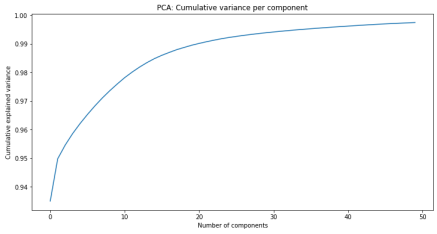

Fig. 4: Cumulative variance per principal component obtained with the PCA.

4.1 Baseline experiment: Classic approaches

The baseline experiment has been designed to train two classic approaches. By doing this, we are defining a baseline of results to try to improve with the deep learning approach.

Both algorithms, Random Forest and SVM, were trained with 10000 pixels for each class of ceramics and the pixels were chosen through random sampling. From the whole sample of pixels, 30% was used as validation and the rest for the training.

For the Random Forest, the number of estimators was 10. The results obtained showed that it was able to learn some differences between different categories of the ceramics. But the results were very poor.

For the SVM classifier, the regularization parameter was 1 and the kernel used was the radial basis function kernel (rbf). By analysing the results we could see that it performed a little better than the random forest classifier, but the results were still poor.

	Precision		Recall	
Class	RF	SVM	RF	SVM
A	0.31	0.42	0.32	0.14
B	0.43	0.55	0.40	0.61
C	0.12	0.14	0.15	0.39
D	0.27	0.43	0.32	0.36
E	0.10	0.21	0.10	0.02
F	0.12	0.26	0.12	0.12
G	0.13	0.25	0.13	0.08
H	0.09	0.11	0.09	0.40
J	0.22	0.50	0.21	0.19
K	0.09	0.09	0.09	0.05
L	0.08	0.40	0.08	0.00
M	0.18	0.42	0.17	0.20
N	0.17	0.28	0.16	0.09
P	0.14	0.14	0.18	0.38
Q	0.46	0.49	0.50	0.66
R	0.09	0.11	0.08	0.15
S	0.11	0.64	0.08	0.01

Table 2: Recall and precision for Random Forest and Support Vector Machine.

Both algorithms got good results for the first two classes of ceramics, that belong to the same category, (Table.2). Although they had trouble to tell them apart. A big difference that we can see between Random Forest and SVM is that the Random Forest results when classifying is able to learn a little the right class for each ceramic. However,

wrong predictions are distributed quite equally among all others classes. On the other hand, the SVM classifier also learns, better than the Random Forest, the right class for ceramics, but wrong predictions are distributed between two or three other classes that may belong or not to the same category of the ceramic. SVM predicted that a lot of pixels are from class C, H and P, being the three of them from different categories, while almost no pixels were predicted as class E, L and S.

4.2 Deep learning experiments

This set of experiments has been designed to see if the results obtained from the classic approaches can be improved through the deep learning classifier selected, the U-net.

The U-net was trained from scratch using Adam as the optimizer, with a learning rate of 0.001. It was defined to train for 350 epochs with an early stop on the validation precision metric with a patience of 50 epochs. The experiments involving the U-net architecture have been trained with a batch size of 2 due to limitations in storing the images during training.

The experiments done using the U-net model have used the dataset, which has been divided in three different subsets. The first consists on 104 images for the training of the model. A second set with 53 images is used as the validation of the training. Finally, a third set with 17 images, one per class, is used for testing the results from the training stage of the model.

All the pixels that belonged to the background, including the pixels that presented issues during the normalizing process, were masked in the loss function so that they were not taken into account. This was done because they introduced a huge imbalance in the dataset and we wanted the U-net to focus on learning to classify the ceramics.

4.2.1 Experiment 1: classification by class

The experiment by class consisted on training the model feeding it the images from the training subset with their dimensionality reduced. The U-net model had to classify between 17 classes.

The results were analysed using a confusion matrix, (Figure. 5), and can also be seen in Table 3. We could see that the model was learning to classify each different class. The mistakes that we can see are from ceramics that are very similar to each other, such as classes A and B, which are both ”Campaniana”. We can also see that the classes that are ”Terra Sigillata” are predicted between two sets, they are usually labeled as C or F, for classes C-F and G or J for classes G-K. This two sets coincide with the differences of origin between them, except for G. The classes from C to G are Hispanic, while classes from H-K are from Africa. We can also see confusion between P, Q and N, because they are all ”Àmfores”.

A visual analysis from the predictions of the ceramics showed how in those classes that are from the same category or region are confused. While ceramics from the classes ”Campaniana” have almost no errors and both are labeled as the same class, the rest of ceramics have quite a lot of discrepancies. Despite the errors of predicted pixels, we can

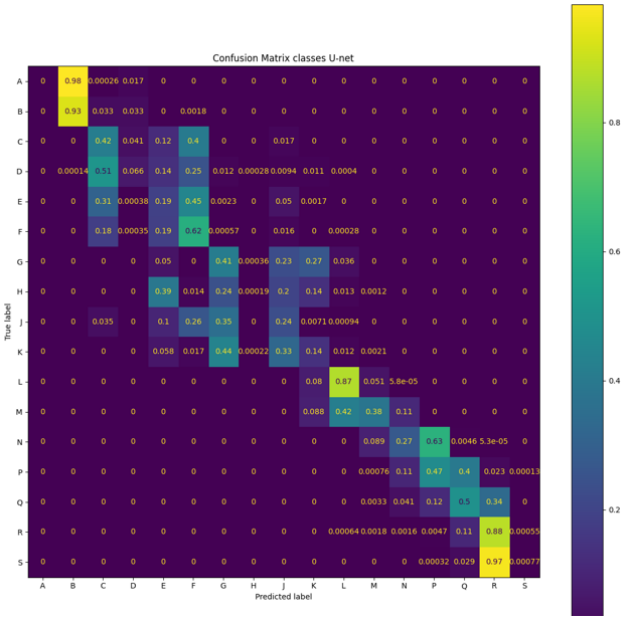

Fig. 5: Confusion matrix with U-net model. Experiment with background masked in the loss function.

see that they are from the classes that share common traits, such as the kind or region. So we can see that, although it makes mistakes, the model is able to learn relations between classes. This can be seen on figure 6, where the ceramics true class is represented as dark blue. The rest of colors belong also to ceramics that are ”Terra Sigillata”.

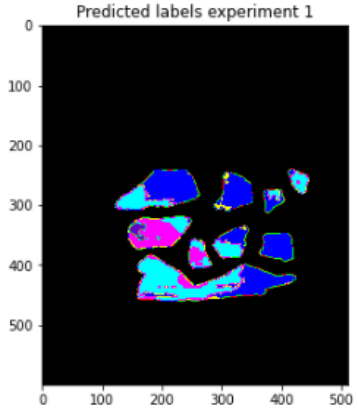

Fig. 6: Labels predicted for ceramic from class C in experiment 1.

4.2.2 Experiment 2: cropped images

An important hyperparameter from neural networks is the batch size. The batch size defines the number of samples that are passed at the network at one time. A big batch size can lead to poor generalization, whereas a small batch size is not guaranteed to find the global optima.

That is the reason why another experiment was developed in which we wanted to see if by increasing the batch size we could improve the results obtained. To be able to increase the batch size, the images were cropped into four parts. By doing that, we were able to increase the batch size to 12.

The results obtained were similar to the previous exper-

	Precision		Recall	
Classes	Exp. 1	Exp. 2	Exp. 1	Exp. 2
A	0.00	0.65	0.00	0.87
B	0.34	0.32	0.93	0.11
C	0.28	0.37	0.42	0.42
D	0.47	0.59	0.07	0.83
E	0.08	0.12	0.19	0.36
F	0.33	0.10	0.62	0.08
G	0.35	0.39	0.41	0.21
H	0.19	0.46	0.00	0.21
J	0.22	0.3	0.24	0.19
K	0.11	0.17	0.14	0.12
L	0.64	0.48	0.87	0.45
M	0.78	0.53	0.38	0.46
N	0.38	0.60	0.27	0.44
P	0.39	0.30	0.47	0.50
Q	0.63	0.41	0.50	0.43
R	0.43	0.45	0.88	0.63
S	0.45	0.72	0.00	0.07

Table 3: PRECISION AND RECALL PER CLASS FOR EXPERIMENT 1, CLASSIFICATION BY CLASS, AND EXPERIMENT 2, CROPPED IMAGES.

iment and can be seen on Table 3. It was able to learn to classify ceramics, distinguishing the same relations as the previous experiment. Although the distinction of region between the ceramics that are ”Terra Sigillata” wasn’t so clear.

4.2.3 Experiment 3: first classification by groups

This experiment was designed so that the classes from the ceramics were grouped by different categories. We wanted to see if by grouping them into different categories we are able to classify them with better results.

The categories for the groups were chosen based on the criteria specified by the archaeologists that provided the dataset. The first group was ”Campaniana” with classes A and B, because they were both the same kind of ceramic. The second group was ”Terra Sigillata Vermella” grouping C, D, E, F and G. This group presents ceramics that share a common age and some are imitations from others, which makes them very similar. The next group was ”Terra Sigillata Africana”, with classes H, J, K and M. This group was separated from the other one because they are all from Africa and have distinguishable traits from the rest of ceramics. The fourth group was ”Itàliques”, with classes L and N. This group criteria is the same as the previous one, being this classes of ceramics from Italy. Another group was ”Àmfores” which grouped P and Q for being the same kind of ceramic. Although class N is also from the kind ”Àmfora” it has been chosen to group with the classes originated from Italy. The last two classes R and S, were left as different groups.

This experiment used the full sized images, not the cropped images from the previous experiment. The results, (Table. 4), show that the classifier is able to learn with good results all groups of ceramics except one. We can see that the groups that were ”Terra Sigillata” but from different regions are able to be distinguished with good results, although the mistakes made are between this two groups.

Classes R and S haven’t been able to tell apart and they have also been confused with the group ”Àmfora”. The reason for them to be very difficult to tell apart is because when the PCA is applied their spectral signature becomes very similar, except for class Q. They have more notable differences when they haven’t had the transformation from the PCA, which indicates that a dimensionality reduction algorithm that preserves those differences, such as a band selection algorithm, may improve the results obtained. We can see their spectral curve in Figure 7 for each band and Figure 8 for each principal component obtained from the PCA. Images P and Q

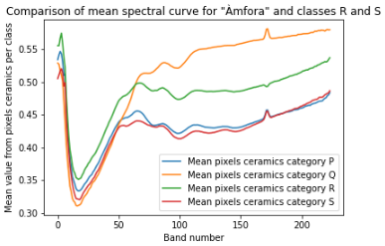

Fig. 7: Spectral curve per band for group ”Àmfora” and classs R and S.

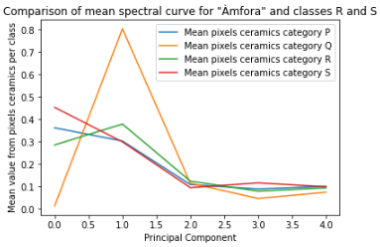

Fig. 8: Spectral curve per principal component for group ”Àmfora” and classs R and S.

Class	Precision	Recall
Campaniana (AB)	1.00	0.90
Terra Sigillata vermella (CDEFG)	0.8	0.79
Terra Sigillata Africana (HJKM)	0.65	0.74
Itàliques (LN)	0.70	0.62
Àmfores (PQ)	0.83	0.69
Mat. constructiu Romà”(R)	0.46	0.94
Mat. constructiu contemporàni”(S)	0.00	0.00

Table 4: PRECISION AND RECALL PER GROUPS FROM EXPERIMENT 3.

4.2.4 Experiment 4: second classification by groups

In this experiment we defined a different set of groups from the previous one. The difference between this experiment and the previous one is that classes L, M and N have been separated and classified as a group each one.

We want to see whether the classifier is able to tell this aforementioned classes apart or not. If they couldn’t be differentiated, we wanted to see if class N would be more similar to the category ”Àmfores” or to the other ceramic that is also from Italy, class L. The same way was for class M,

Class	Precision	Recall
Campaniana (AB)	1.00	0.95
Terra Sigillata vermella (CDEFG)	0.84	0.66
Terra Sigillata Africana (HJK)	0.53	0.72
Itàlica de cuina (L)	0.44	0.82
Africana de cuina (M)	0.78	0.32
Àmfora itàlica campana (N)	0.54	0.67
Àmfores (PQ)	0.73	0.64
Mat. constructiu Romà" (R)	0.39	0.69
Mat. constructiu contemporàni (S)	0.95	0.03

Table 5: PRECISION AND RECALL PER GROUPS FROM EXPERIMENT 4.

which we wanted to see if it would be similar with the ceramics from Africa or not because it's a different kind of ceramic.

After analysing the results, (Table. 5), we can see that class M has been mostly confused with class L. The class N has been proven that it has more similarity with the group "Àmfores", rather than with the other ceramic from Italy, class L. Class R has had good results, with a 0.69 recall, because it was usually able to classify ceramics of class R, but the precision is only 0.39 because ceramics from class S were also labeled as class R. When the classifier failed to predict a pixel from a ceramic of class R, it was confused with group "Àmfora".

4.3 Fragment classification

After analysing the results obtained from the classifiers through semantic segmentation, an interesting approach was to label whole fragments from their prediction. The predicted label for the whole fragment is chosen by computing the most predicted class out of all the pixels from that fragment.

First, the fragments were labeled using the results from the first deep learning experiment. We are able to see that the results by labeling whole fragments have a slight improvement compared to labeling by pixel. This is because by taking the maximum predicted label per fragment, the isolated errors from wrong predictions that are only in some pixels are corrected. We can see that on the confusion matrix in figure 9.

The results obtained when labeling per fragment with the others deep learning experiments showed that they also followed the same pattern. They also had a slight improvement of the results.

5 CONCLUSIONS AND FUTURE WORK

In this article, we have designed a computer vision system to classify ceramics based on hyperspectral data. Our system is composed of three modules. The first one consists on a dimensionality reduction algorithm based on a PCA. Second, a classification module and finally the evaluation module. The classification module consists on two classic machine learning algorithms, the Random Forest and the Support Vector Machines, and a deep learning approach: the U-net neural network.

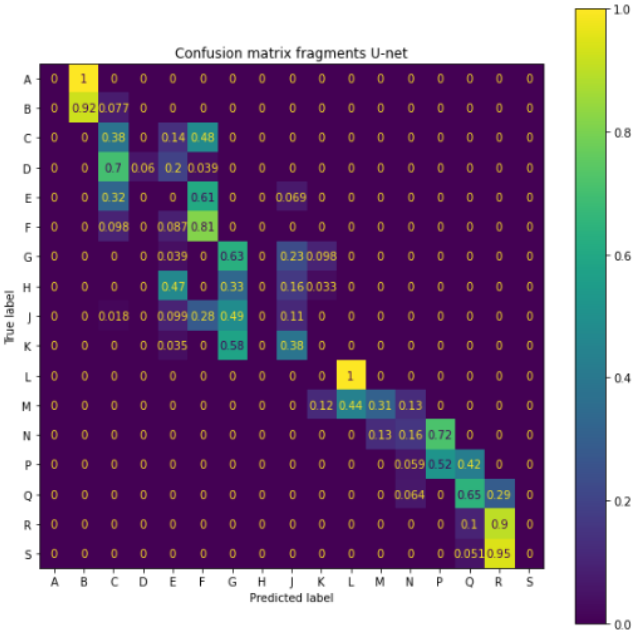

Fig. 9: Confusion matrix labeling fragment.

We have been able to see that the deep learning approach has obtained better results than the classical algorithms and that the ceramics can be classified by category using their spectral signature. This was expected, because the deep learning algorithm learns the information from the images taking into account the neighbours while the classic approaches learn each pixel independently.

The results obtained proved that there is a significant difference that can be used to classify them. Results have improved considerably, because the classic approaches had a precision and recall close or lower than 50% in their predictions. On the other hand, for classes such as the ones that were "Campaniana", the U-net achieved a 100% of precision and 90% of recall in experiment 3, and 100% precision and 95% recall in experiment 4. The groups "Terra Sigillata vermella" and "Terra Sigillata Africana" have also been able to be distinguished in both experiment 3 and 4 with notable results. Experiment 3 achieved a precision of 80% and a recall of 79% for the first group and a precision of 65% and a recall of 74% for the second group. Experiment 4 achieved a precision of 84% and a recall of 66% for the first group and a precision of 53% and a recall of 72% for the second group. Better results were expected for the ceramics that belong to the last classes. They were confused because key information about their spectral signature was lost due to the dimensionality reduction of the PCA.

Semantic segmentation has proven efficient in classifying the ceramics and also using the results to then label by fragments. The results obtained were slightly better than semantic segmentation. For example, using the predictions from experiment 3, the precision for "Terra Sigillata Vermella" increased 5% and the recall 6%. A similar increase happened for "Terra Sigillata Africana", where precision increased by 10% and the recall by 5%.

Due to the lack of state of the art surrounding the clas-

sification of ceramics based on hyperspectral images, this project has been able to establish and define the first experiments and a baseline of results to take into account and further improve.

After reviewing all the results, we believe that the U-net neural network has proven efficient in classifying the ceramics. We have left for future work implementing a band selection algorithm to see if it is able to preserve more important information to distinguish the ceramics.

REFERENCES

[1] Y. Maniatis, "The Emergence of Ceramic Technology and its Evolution as Revealed with the use of Scientific Techniques" ResearchGate, 2009. https://www.researchgate.net/publication/228776265_The_Emergence_of_Ceramic_Technology_and_its_Evolution_as_Revealed_with_the_use_of_Scientific_Techniques (accessed Feb. 24, 2022).

[2] *The evolution of ceramics processing in human history*. Available at: https://www.madeinitalyfor.me/en/info/evolution-ceramics-human-history (accessed: Feb. 27, 2022)

[3] *A brief history of ceramics and glass* Available at: https://ceramics.org/about/what-are-engineered-ceramics-and-glass/brief-history-of-ceramics-and-glass (accessed: Feb. 27, 2022)

[4] D. Barker and T. Majewski, "Ceramic studies in historical archaeology ** PROOF STAGE **," 2006. [Online]. Available: http://www.fiskecenter.umb.edu/Staff/Landon/Classes/Barker_Majewski_2006.pdf (accessed: Jun. 4, 2022)

[5] Vishvanathan, Sowmya & Kp, Soman Hassaballah, Mahmoud. (2019). Hyperspectral Image: Fundamentals and Advances. 10.1007/978-3-030-03000-1_16.

[6] Miller, G.L. Classification and economic scaling of 19th century ceramics. Hist Arch 14, 1–40 (1980). https://doi.org/10.1007/BF03373454

[7] A. Gilboa, A. Karasik, I. Sharon, and U. Smilansky, "Towards computerized typology and classification of ceramics," Journal of Archaeological Science, vol. 31, no. 6, pp. 681–694, Jun. 2004, doi: 10.1016/j.jas.2003.10.013.

[8] B. Rasti et al., "Feature Extraction for Hyperspectral Imagery: The Evolution From Shallow to Deep: Overview and Toolbox," IEEE Geoscience and Remote Sensing Magazine, vol. 8, no. 4, pp. 60–88, Dec. 2020, doi: 10.1109/mgrs.2020.2979764.

[9] Md. P. Uddin, Md. A. Mamun, and Md. A. Hossain, "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," International Journal of Remote Sensing, vol. 40, no. 18, pp. 7190–7220, Apr. 2019, doi: 10.1080/01431161.2019.1601284.

[10] "IEEE Xplore Full-Text PDF":, Ieee.org, 2022. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=arnumber=5765424 (accessed Feb. 27, 2022).

[11] "Comprehensive review of hyperspectral image compression algorithms," The Lens - Free Open Patent and Scholarly Search, 2013. https://www.lens.org/lens/scholar/article/048-152-860-932-428/main (accessed Feb. 20, 2022).

[12] J. Zabalza et al., "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," Neurocomputing, vol. 185, pp. 1–10, Apr. 2016, doi: 10.1016/j.neucom.2015.11.044.

[13] W. Sun and Q. Du, "Hyperspectral Band Selection: A Review," IEEE Geoscience and Remote Sensing Magazine, vol. 7, no. 2, pp. 118–139, Jun. 2019, doi: 10.1109/mgrs.2019.2911100.

[14] H. Dong, L. Zhang, and B. Zou, "Band Attention Convolutional Networks For Hyperspectral Image Classification," ResearchGate, Jun. 10, 2019. https://www.researchgate.net/publication/333717088_Band_Attention_Convolutional_Networks_For_Hyperspectral_Image_Classification (accessed Feb. 20, 2022).

[15] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 158, pp. 279–317, Dec. 2019, doi: 10.1016/j.isprsjprs.2019.09.006.

[16] S. Kumar, R. Krishna, R. Shiv, B. Dubey, and Chaudhuri, "HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification," doi: 10.1109/LGRS.2019.2918719).

[17] Xiaohui Yuan b, Jianfang Shi and Lichuan Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," ScienceDirect.

[18] R. Kemker, R. Luu, and C. Kanan, "Low-Shot Learning for the Semantic Segmentation of Remote Sensing Imagery," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–10, 2018, doi: 10.1109/tgrs.2018.2833808.

[19] M. S. Moustafa, S. A. Mohamed, S. Ahmed, and A. H. Nasr, "Hyperspectral change detection based on modification of UNet neural networks," Journal of Applied Remote Sensing, vol. 15, no. 02, Jun. 2021, doi: 10.1117/1.jrs.15.028505.

[20] admin, "Specim FX10 - Specim," Specim, Aug. 10, 2021. https://www.specim.fi/products/specim-fx10/ (accessed Jun. 21, 2022).

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015. [Online]. Available: https://arxiv.org/pdf/1505.04597.pdf

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018, doi: 10.1109/tpami.2018.2858826.