Classification of Historic Food Images

A pilot experiment on the example of the ChIA project

Amelie DORN, ACDH-CH OeAW, Austria
Renato ROCHA SOUZA, ACDH-CH OeAW, Austria
Gerda KOCH, Europeana Local – Österreich, Austria
Japesh METHUKU, Dublin City University, Ireland
Yalemisew ABGAZ, Dublin City University, Ireland

Abstract: This paper describes two image classification tasks, carried out in the context of the interdisciplinary Digital Humanities project ChIA (Accessing and Analysing Cultural Image with New Technologies). On a set of selected still life food images from the Europeana collection, two classification rounds were carried out by five human annotators. On the one hand, concrete food objects were annotated; on the other hand, also abstract cultural features. The degree of precision in the description of labels was varied. In the first annotation round, no description of labels was provided and annotators relied solely on their interpretation or intuition. In the second annotation round, annotators were given concrete definitions of the labels. Besides, the annotators varied in age, gender and cultural background. The aim of these tasks was to determine whether patterns in the classified data would emerge and if so, which parameters would be decisive. As part of the evaluation the inter-annotator agreement was calculated. Preliminary results suggest that the identification of cultural features in images is highly subjective. Agreements were higher for the classification of concrete objects than for abstract features. Some tentative patterns emerged with regard to gender, but a larger annotated data set is needed to draw further and more definite conclusions.

Keywords: *Historic Food Images—Image Classification—CNN—Digital Humanities—Image Annotation*

CHNT Reference: Dorn, A., Rocha Souza, R., Koch, G., Methuku, J., and Abgaz Y. (2022). 'Classification of Historic Food Images – a pilot experiment on the example of the ChIA project', in Börner, W., Rohland, H., Kral-Börner, C. and Karner, L. (eds.) *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, held online, November 2020*. Heidelberg: Propylaeum.
doi:10.11588/propylaeum.1045.c14477

Introduction

This paper is realised in the context of the interdisciplinary Digital Humanities (DH) project ChIA – accessing and analysing cultural images with new technologies (https://chia.acdh.oeaw.ac.at/) (Abgaz, Dorn, Koch, and Preza Diaz, 2020; Dorn, Abgaz, Koch, and Preza Diaz, 2020). The project brings together expertise from Digital Humanities, the Cultural Heritage Sector and Computer Science. It generally aims at testing different Semantic technologies and Artificial Intelligence (AI) tools on a selected set of Europeana (https://www.europeana.eu/de) food images for improving access and analysis possibilities of knowledge contained within images for research and education purposes. This paper reports a pilot experiment on the human classification of historic food images for the purpose of creating a training data set. In particular, the classification process of a selected set of Europeana historic food images is described, taking into account also cultural aspects. This specific combination of concrete (food) and abstract (cultural) features contained within the images poses a significant challenge for both humans and machines when it comes to image classification.

In this study, the aim was to determine whether certain patterns in the classified data emerge that could potentially correspond to specific image features (depicted objects, colour, etc.) or that are particular to human annotators (cultural background, gender, age, etc.). Here the authors report on the research process and provide insights into results as well as on challenges faced.

Background – Food, Culture and Art

The kind of food consumed by humans, the way it is produced, and the cultural features associated with it – these facets are all closely related to our political and economic history as food consumption evolves parallel to different major historic and cultural eras (Hirschfelder, 2001; Ott, 2017).

Images related to food, food production or consumption, typically carry strong cultural aspects, and can be found in archaeological representations, various art forms as well as in contemporary social media (Twiss, 2019). Throughout history, food related items have been captured in different social situations and art forms. A large number of paintings and pictures stored in museums, libraries and galleries depict food items carrying strong cultural symbolism, but the cultural meaning often remains hidden to the contemporary observer (Moon, 2015). Still life paintings and images (see Fig. 1) are a particularly prominent category, in which the depiction of food, its interpretation and symbolism has played a major role throughout the centuries (Bedaux, 1987; Piepmeier, 2018). Searching for such images with cultural connotations is frequently hampered by the fact that a detailed description of all items on a picture is missing, also from the title or metadata. That's why creating structured access to implicitly and explicitly contained knowledge can create significantly more opportunities for analysing possibilities across different scholarly fields.

The Experiment – Historic Food Image Classification

As a first step in the experiment, a specific dataset and vocabulary had to be selected, as a benchmark data containing images and their relevant semantic annotations were required in order to apply AI and CNN (Convolutional Neural Network) technologies for image recognition (cf. Ciocca et al., 2018). For ChIA, this data is provided by the European portal for cultural heritage, the Europeana Collections portal (http://www.europeana.eu) with more than 32 million images available under different licenses. For image collection and retrieval, the ChIA infrastructure hosts a service that supports the collection of test datasets according to predefined search criteria and provides the unique possibility to create selected test sets for further analysis with CV/CNN/AI tools out of the wealth of (open access) Europeana digital content.

Fig. 1. Example of a still life painting with fruit. Source: "Stilleven met artisjok, fruit op porseleinen schalen, een zoutvat en een pepervat", Osias Beert, ca. 1605 – ca. 1615, Rijksmuseum, Netherlands. Image: CC-BY-PD.

A concise set of cultural food images was carefully chosen, given the large and heterogeneous field of food images. For the purpose of this study, images on the topic of still life and fruits were selected through a targeted search on the ChIA/Europeana interface, yielding a set of a total of 392 images, including black and white, and colour paintings and drawings. The subject of still life images depicting fruit was chosen, since, on the one hand, still life images have been a popular subject of research across different disciplines and are rich in cultural symbolism. On the other hand, it also allowed us to create a more defined data set from among the vast amount of images depicting fruits in very different ways in the Europeana collection.

Next, the vocabulary for tagging was defined. Different approaches were considered based on the evaluation of frequently used concepts within the cultural images domain and iconographic descriptions, which also included the retrieval of a baseline of food concepts from the Getty Arts and Architecture Thesaurus (https://www.getty.edu/research/tools/vocabularies/aat/), the IconClass multilingual classification system for cultural content (http://www.iconclass.org/) and the FoodOn ontology (https://foodon.org/). As none of these common approaches yielded satisfactory solutions for our purpose, a different and simpler binary classification approach was chosen, taking a concrete object-related feature and one more abstract, cultural feature.

Images were tagged for food items (fruit/non-fruit) —a common task based on the presence or absence of some object; and on a cultural feature (formal/informal), which is based on personal views, cultural background, and tastes. The tagging was carried out using the MakeSense.AI application (https://www.makesense.ai/). Images were tagged by five independent annotators, in two rounds. Both tasks involved tagging the set of 392 images, for the two binary label choices (fruits/non-fruits and formal/informal). In the first round, the task was presented without detailed definitions on the classes and in the second one, more specific and concrete definitions were provided to the annotators. The five annotators came from culturally diverse backgrounds, pertained to different age groups and were of mixed gender (2 female, 3 male).

The evaluation was made using Cohen's kappa (κ), which is a common statistic to measure the inter-annotator agreement. The kappa function computes a score that expresses the level of agreement between two annotators on a classification problem. It is defined as:

$$K = (p_0 - p_e)/(1 - p_e)$$

where p_0 is the empirical probability of agreement on the label assigned to any sample (the observed agreement ratio), and p_e is the expected agreement when both annotators assign labels randomly. p_e is estimated using a per-annotator empirical prior over the class labels.

For each experiment and each task 25 pairwise comparisons were calculated (see Tables 1 and 2). The diagonal of the tables shows the self-comparison, which is by definition always equal to 1. Numbers in brackets indicate the number of images classified.

In what follows, the detailed procedures and results of the two classification experiments are outlined.

Experiment 1—Image Classification Results

As a first step, the images were distributed to annotators. Each person was asked to classify the images with the provided labels (set 1: fruit/non-fruit and set 2: formal/informal), where first the entire image set was classified using only label set 1, and in a second round the entire image set was classified again only using label set 2. For this experiment, no supplementary details, or definitions with regard to the labels were provided, relying solely on the annotators' own knowledge, intuition and interpretation.

The results from experiment 1 are presented in Table 1 below, showing the similarity scores (κ).

Table 1. Overview of similarity scores across the 5 annotators for experiment 1, applying label set 1 (upper panel) and label set 2 (lower panel).

Task_1	**User1**	**User2**	**User3**	**User4**	**User5**
User1	1.000 / (392)	0.928 / (392)	0.892 / (392)	0.907 / (392)	0.886 / (392)
User2	0.928 / (392)	1.000 / (392)	0.892 / (392)	0.938 / (392)	0.897 / (392)
User3	0.892 / (392)	0.892 / (392)	1.000 / (392)	0.923 / (392)	0.923 / (392)
User4	0.907 / (392)	0.938 / (392)	0.923 / (392)	1.000 / (392)	0.918 / (392)
User5	0.886 / (392)	0.897 / (392)	0.923 / (392)	0.918 / (392)	1.000 / (392)
Task_2	**User1**	**User2**	**User3**	**User4**	**User5**
User1	1.000 / (392)	0.330 / (392)	0.252 / (392)	0.316 / (392)	-0.091 / (392)
User2	0.330 / (392)	1.000 / (392)	0.210 / (392)	0.306 / (392)	0.153 / (392)
User3	0.252 / (392)	0.210 / (392)	1.000 / (392)	0.051 / (392)	-0.031 / (392)
User4	0.316 / (392)	0.306 / (392)	0.051 / (392)	1.000 / (392)	-0.028 / (392)
User5	-0.091 / (392)	0.153 / (392)	-0.031 / (392)	-0.028 / (392)	1.000 / (392)

Table 1 presents the pairwise comparisons of agreement scores across the five annotators applying the label sets for Task 1 (fruit/non-fruit) and for Task 2 (formal/informal). For Task 1, the agreement across the five annotators is relatively high with the score varying between 0.8 and 0.9. For

Task 2, a sharp drop in agreement scores compared to Task 1 is noted, with some scores falling even below zero (-0.2), with zero as an indication of chance agreement.

Experiment 2—Image Classification Results

In this experiment, the same setting as in the previous was given; however, at this time detailed definitions of the labels fruit/non-fruit and formal/informal were provided. This was done in order to determine whether the agreement among annotators would be any different than in the previous experiment. The definitions were derived from available monolingual English language dictionaries, e.g., Collins English Dictionary.

- Fruit: fruit or a fruit is something which grows on a tree or bush, and which contains seeds or a stone covered by a substance that you can eat. (e.g., strawberry, nut, tomato, peach, banana, green beans, melon, apple)
- Non-fruit: images that do not feature any type of fruit (for fruit definition see above)
- Formal: arranged in a very controlled way or according to certain rules; an official situation or context.
- Informal: a relaxed environment, an unofficial situation or context, disorderly arrangement

Results from experiment 2 are presented in the following Table, showing similarity scores across the five annotators.

Table 2. Overview of similarity scores across the 5 annotators for experiment 2 label set 1 (upper panel) and label set 2 (lower panel).

Task_1	**User1**	**User2**	**User3**	**User4**	**User5**
User1	1.000 / (392)	0.943 / (392)	0.913 / (392)	0.928 / (392)	0.913 / (392)
User2	0.943 / (392)	1.000 / (392)	0.886 / (392)	0.912 / (392)	0.866 / (392)
User3	0.913 / (392)	0.886 / (392)	1.000 / (392)	0.923 / (392)	0.928 / (392)
User4	0.928 / (392)	0.912 / (392)	0.923 / (392)	1.000 / (392)	0.913 / (392)
User5	0.913 / (392)	0.866 / (392)	0.928 / (392)	0.913 / (392)	1.000 / (392)
Task_2	**User1**	**User2**	**User3**	**User4**	**User5**
User1	1.000 / (392)	0.168 / (392)	0.216 / (392)	0.255 / (392)	0.167 / (392)
User2	0.168 / (392)	1.000 / (392)	0.188 / (392)	0.095 / (392)	0.419 / (392)
User3	0.216 / (392)	0.188 / (392)	1.000 / (392)	0.094 / (392)	0.358 / (392)
User4	0.255 / (392)	0.095 / (392)	0.094 / (392)	1.000 / (392)	0.089 / (392)
User5	0.167 / (392)	0.419 / (392)	0.358 / (392)	0.089 / (392)	1.000 / (392)

Table 2 presents the pairwise comparisons of agreement scores across the five annotators applying the label sets for Task 1 (fruit/non-fruit) and for Task 2 (formal/informal). Looking first at Task 1, the agreement across the five annotators is again relatively high, similar to those presented in Table 1, Task 1, with a score ranging between 0.8 and 0.9 as well. Looking next at Task 2, a considerable decrease in agreement scores compared to Task 1 can be noted, with scores ranging between 0.4 and ~0.09.

Discussion

For task 1 in both experiments, results showed that the average agreement was high, but increased slightly when definitions were provided in the second experiment (from 0.928 to 0.930). This is also a behaviour one would expect from non-human labelling systems (e.g. Neural Networks), given the clear correlation of the task and the graphical elements present in the figure.

For the second task, the average agreement was low in both experiments, and increased only slightly when definitions were provided (from 0.317 to 0.364). This was also expected, given the idiosyncrasies regarding classification of cultural aspects.

Additionally, the overall agreement among female and male classifiers was calculated. The overall agreement for the female classifiers, task 1 was 0.945 and for the male classifiers, task 1, was 0.934. For task 2, the agreement for the female classifiers was 0.318 and for the male classifiers was 0.242. Although there is no statistically significance with such a small sample, that may also indicate that gender related aspects can influence the levels of agreement in classification tasks.

Conclusions

In conclusion, our results from this experiment showed that classifying historic food images according to both physical and cultural aspects are fairly different tasks. While defined physical patterns are a domain in which the machines have acquired the state-of-the-art, cultural features pose a much more challenging task for both machines and humans, requiring at times different strategies to the commonly applied approaches. It is obvious that food interpretation is framed by an interdisciplinary science that includes primarily sociological, technological, economic, and cultural dimensions. And our findings support the proposition that food interpretation is time- and place-specific, and it is individual, not easily generalisable.

This paper aims to lay the grounds for further research in which it is intended to compare human and machine annotation performances. The analysis of inter-annotator-agreement showed so far that humans are able to capture subtle nuances and cultural features contained in images. In future experiments the aim is to compare human to machine annotations, as this can open ways to enhance image searches for digital cultural collections, a development that is to-date still in its infancy. A comprehensive discussion of the scientific work along with the final results of the study is available in Abgaz et al. (2021).

Funding

The ChIA project is funded by a go!digital Next Generation grant (GDNG 2018-051) of the Austrian Academy of Sciences.

Conflict of Interests Disclosure

The authors declared no conflict of interest.

Author Contributions

Conceptualization: Amelie Dorn, Gerda Koch, Japesh Methuku, Renato Rocha Souza, Yalemisew Abgaz

Data curation: Gerda Koch

Formal Analysis: Yalemisew Abgaz, Japesh Methuku, Renato Rocha Souza

Funding acquisition: Amelie Dorn, Yalemisew Abgaz

Investigation: Amelie Dorn, Gerda Koch, Japesh Methuku, Renato Rocha Souza, Yalemisew Abgaz

Methodology: Renato Rocha Souza, Yalemisew Abgaz, Japesh Methuku

Project Administration: Amelie Dorn

Resources: Amelie Dorn, Gerda Koch, Japesh Methuku, Renato Rocha Souza, Yalemisew Abgaz

Software: Renato Rocha Souza Yalemisew Abgaz

Supervision: Yalemisew Abgaz, Renato Rocha Souza

Validation: Yalemisew Abgaz, Japesh Methuku, Renato Rocha Souza

Visualization: Renato Rocha Souza, Yalemisew Abgaz, Japesh Methuku

Writing – original draft: Amelie Dorn, Gerda Koch, Japesh Methuku, Renato Rocha Souza, Yalemisew Abgaz

Writing – review & editing: Amelie Dorn, Gerda Koch, Japesh Methuku, Renato Rocha Souza, Yalemisew Abgaz

References

Abgaz, Y., Rocha Souza, R., Methuku, J., Koch, G., and Dorn, A.. (2021). 'A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies' *Journal of Imaging* 7(8), 121, doi:10.3390/jimaging7080121

Abgaz, Y., Dorn, A., Koch, G., and Preza Diaz, J. L. (2020). 'A picture is worth a thousand words? – From Project Inception to First Results: Describing Cross-disciplinary Collaboration in the Digital Humanities Project ChIA'. In: Krauwer, S. and Fiser, D. (eds.) *Proceedings of the Twin Talks 2 and 3 Workshop: Understanding and Facilitating Collaboration in Digital Humanities 2020, Ottawa Canada and Riga Latvia, July 23 and October 20 2020*. CEUR Workshop Proceedings Vol-2717. ISSN 1613-0073.

Bedaux, J. (1987). 'Fruit and Fertility: Fruit Symbolism in Netherlandish Portraiture of the Sixteenth and Seventeenth Centuries', Simiolus: Netherlands Quarterly for the History of Art17, No. 2/3 (1987). Pp. 150–68.

Ciocca, G., Napoletano, P., and Schettini, R. (2018). 'CNN-based features for retrieval and classification of food images', *Computer Vision and Image Understanding*, Volumes 176–177, 2018, pp. 70–77, ISSN 1077-3142, doi:10.1016/j.cviu.2018.09.001.

Dorn, A., Abgaz, Y., Koch, G., and Preza Diaz, J. L. (2020). 'Harvesting knowledge from cultural images with assorted technologies: the example of the ChIA project.' In: Lykke, M., Svarre, T., Skov, M., and Martínez-Ávila, D. (eds.) (2020). *Proceedings of the 16th International ISKO conference 2020. Aalborg, Denmark.* Advances in Knowledge Organization, Vol. 17. Pp. 470–473. Baden-Baden: Ergon Verlag.

Food in Art History. Website. https://www.widewalls.ch/magazine/food-in-at-history (Accessed: 20 August 2020)

Hirschfelder, G. (2001). *Europäische Esskultur: Eine Geschichte der Ernährung von der Steinzeit bis heute*. Campus Verlag.

Moon, M. (2015). *Interpreting Food at Museums and Historic Sites*. Lanham (Maryland): Rowman & Littlefield.

Ott, C. (2017). *Identität geht durch den Magen: Mythen der Esskultur*. Frankfurt am Main: S. Fischer Verlag.

Piepmeier, M. (2018). *The Appeal of Lemons: Appearance and Meaning in Mid Seventeenth-century Dutch Paintings*. Published MA Thesis. University of North Carolina at Chapel Hill, USA.

Twiss, K.C. (2019). *The Archaeology of Food: Identity, Politics, and Ideology in the Prehistoric and Historic Past*. Cambridge: University Press.