Improving object detection in paintings based on time contexts

Maria-Cristina Marinescu
Barcelona Supercomputing Center
Barcelona, Spain
mariacristina.marinescu@gmail.com

Artem Reshetnikov
Barcelona Supercomputing Center
Barcelona, Spain
artem.reshetnikov@bsc.es

Joaquim Moré López
Barcelona Supercomputing Center
Barcelona, Spain
joaquim.morelopez@bsc.es

***Abstract*—This paper proposes a novel approach to object detection for the Cultural Heritage domain, which relies on combining Deep Learning and semantic metadata about candidate objects extracted from existing sources such as Wikidata, dictionaries, or Google NGram. Working with cultural heritage presents challenges not present in every-day images. In computer vision, object detection models are usually trained with datasets whose classes are not imaginary concepts, and have neither symbolic nor time-specific dimensions. Apart from this conceptual problem, the paintings are limited in number and represent the same concept in potentially very different styles. Finally, the metadata associated with the images is often poor or inexistent, which makes it hard to properly train a model. Our approach can improve the precision of object detection by placing the classes detected by a neural network model in time, based on the dates of their first known use. By taking into account the time of inception of objects such as the TV, cell phone, or scissors, and the appearance of some objects in the geographical space that corresponds to a painting (e.g. bananas or broccoli in 15th century Europe), we can correct and refine the detected objects based on their chronologic probability.**

***Index Terms*—Object Detection, Computer Vision, Cultural Heritage, Deep Learning**

I. Introduction

Cultural heritage includes both historical and contemporary art, paintings, buildings, furniture, monuments, documents, archeological sites, as well as oral traditions and habits of the different populations worldwide. Cultural heritage is an essential part of everyday life, as it reflects our past, enriches the present, and informs the future. This provides an important motivation to explore and understand it at a more profound level. Reaching this type of insight automatically, using artificial intelligence techniques, is a challenging task due to several reasons. First, an important part of the cultural heritage artifacts - particularly those in visual arts - refer to imaginary concepts or have symbolic meaning. Every artifact is to a good extent a product of its time and thus reflects time-specific dimensions. Secondly, the artifacts are limited in number and represent the same concept in potentially very different styles. Finally, the metadata associated with the images is often poor, unstructured (i.e. in text form), or does not exist altogether. These facts add up to account for relatively small datasets of images with limited, poor, or unstructured associated metadata. This is a fundamental challenge for analysis tasks such as object detection or object classification, which traditionally rely on extensive datasets to perform well and provide quality metadata - and, as a result, facilitate a better understanding of cultural heritage.

Object detection, image classification, and caption generation have all been proposed to support understanding and metadata generation in the cultural heritage domain, with mixed success. In general, object detection in paintings - the type of artifacts we are focusing on in this work - is a problematic task in the absence of manually labeled images and a big enough dataset. This is basically due to specific aspects that have to do with the genre and style of the artists. In this context, transfer learning starting from models pre-trained over pictures (MS COCO, Flickr 7k, Flickr 30k, ImageNet) looks like the most appropriate solution. However, precision of detection can be quite low.

This paper presents a novel approach to improving the precision of object detection using a combination of deep learning and semantic metadata extraction. The basic idea is to start by using a model pre-trained on pictures (i.e. the MaskRCNN model based on the MS COCO dataset) to first generate a set of candidate object classes. Subsequent steps use information about the time the painting was executed and semantic information about the detected objects that will allow placing them in time and thus eliminate or refine them to concepts more appropriate to the time frame when they were executed. We extract the semantic information about the first-time-use of objects that appear in paintings from a dictionary and we transform it into a well-time-placed matrix, which we call the Time Matrix and use it for refining detected objects. To evaluate our implementation, we compare object detection results over images extracted from a Wikimedia Commons category with a significant number of paintings that relate to some of the most common painting class names, when using the Time Matrix vs. not using it. Results show improved precision and BLEU scores for our method.

II. Related work

The first approaches to object detection were based on template matching techniques and simple part-based models (Fischler and Elschlager et al., (1973)). Taking into account the performance of these approaches and the increasing amount of available data, these gave way to statistical classifiers such as, for instance, SVM or Bayesian Networks [Osuna et al. (1997),

DOI: 10.1109/ICDMW51313.2020.00133

Rowley et al. (1998), Sung and Poggio (1998), Schneiderman and Kanade (2000), Yang et al. (2000a,b), Fleuret and Geman (2001), Romdhani et al. (2001), and Viola and Jones (2001)]. These object detection mechanisms are still used nowadays. Since 2006, Deep Learning started to be used in object detection. due to several factors:

- Significant increase in the amount of annotated data
- Fast development of high performance parallel computing systems
- Advances in the design of network structure and training strategies

During this time, a lot of different models - such as ResNet, VGG-16, AlexNet, etc - were developed and started to be used in different application fields. Most of them were based on Convolutional Networks and were trained on large datasets. Meanwhile, to increase the performance of neural networks some research groups started to implement additional features by changing the structure of the networks, combining them, or including meta-information. There are two families of detectors based on Convolutional Networks: the first family detects more objects but with imprecise bounding boxes, while those of the second family do the opposite. Khaoula Drid et al.(2020) propose a solution by combining the two families, in a way similar to combining classifiers: some of these alternatives were successfully validated, such as it was the case for two famous detectors, Faster R-CNN - which detects more objects, and YOLO - which produces accurate bounding boxes.

Generic object detection focuses on locating, classifying, and labeling objects in images within bounding boxes by returning the probability of each possible label. The approach of generic object detection can be divided into two types. One follows the traditional object detection pipeline, first splitting images by regions of interests (RoI) and then classifying each of them as a class from the list of predefined classes. The other regards object detection as a regression or classification problem. The RoI based methods mainly include R-CNN, Fast R-CNN, Faster R-CNN and Mask R-CNN, some of which are related with each other. The regression/classification based methods mainly include MultiBox, AttentionNet, G-CNN, YOLO, YOLOv2 and DSOD.

III. Basic approach

Although transfer learning seems to be the most promising approach due to the size of the training sets, object detection for cultural heritage has specific challenges that can not be overcome by the straightforward use of this technique. Sure enough, some type of objects (person or horse) can be successfully detected in paintings by using transfer learning and architectures of neural networks such as Mask RCNN (Kaiming He et al.,2017) or Faster RCNN (Shaoqing Ren et al.,2015). However, objects in paintings often reflect symbolic and iconographic meaning, which cannot be simply learned from mechanical image processing, without additional meta-data. Likewise, they may illustrate objects that are now out of use - at least in their historical shape, objects whose shape is very similar to modern objects present in a much larger number of images, or it may be the case that a word has changed meaning over time. Without doubt, the names of the classes used in the MS COCO dataset for labeling photographs (Tsung-Yi Lin et al., 2014) can be used as classes for object detection in paintings. However, this approach doesn't reflect all the range of possible symbolic meanings of objects on paintings, nor those classes that represent imaginary beings not present in pictures. Furthemore, the time aspect can significantly decrease the precision of the object detection algorithm when this generates object labels which cannot appear in a certain time period, and erroneous labels may be generated for those objects that changed shape over time. The difficulty of transfer learning to cultural heritage domain lead us to an approach in which we combine deep learning with semantic metadata about classes, which we extract from external sources. The key is to place objects in the correct context; one of the most determinant context types is time, and this is the focus of the work we present in this paper. Our approach is based on filtering and refining the classes as generated by the Mask RCNN pretrained model based on the MS COCO dataset using transfer learning to best fit the time period of the painting.

Since this dataset only contains classes present in pictures, it will not contain imaginary beings such as angels, saints, or flying witches, nor will it allow refining a person to be a monk or Jesus. This would require introducing new classes and manually labeling many images to prepare a sufficiently large training dataset. We are not reporting on this line of work here, but instead we focus on the potential of using time constraints to refine classes (e.g. book vs cell phone, book vs laptop). To achieve this, we first extract semantic information about the candidate classes (e.g. person or cell phone) which reflects the time of first known use of the word. We then use this information to place concepts in time in the second step, in which our algorithm eliminates, refines, or replaces anachronic classes with the most probable candidates that are timewise viable.

The idea of using time to improve the results of image processing (object detection/caption generation) was first described in the presentation of Harald Sack (2019) and was based on comparing noun phrases of a generated caption with meta-information based on the year of the appearance of the concept. According to his example, if the caption contains the noun "skateboard", it cannot be accepted because the caption was generated for a painting of the 13th century. Considering the flexibility and ambiguous nature of natural language, it is a challenge to find the correct and trustworthy source of such meta-information. We are investigating other sources that may be able to give us better time information that a dictionary; this issue was triggered by possible limitations of using only English dictionaries, as well as observing that some concepts whose historical shape or functioning was drastically different may not appear (e.g. helicopter).

A. Extracting semantic metadata

Every object which can be detected in an image contains meta-information based on its shape, position, or color; it may also form part of contextual information about relationships between the objects. These are features that the neural networks can extract from the image itself. Other characteristics that are more semantic, such as iconographic and symbolic information, must be extracted from other sources. These features are fundamental to labeling objects of a painting coherently with each other and according to the time of the painting. A monk cannot hold a cell phone or eat bananas in a painting from the 13th century. Likewise, a person with a bat and a tall hat, painted in the 14th century, is more likely to be a warrior with some form of a weapon and helmet. In these examples, framing the painting in time will allow our algorithm, in the second step, to replace phones with books, bats with lances or swords, and tall hats with warrior helmets. It will basically filter out classes and recommend replacements for them - whether refinements of the replaced class or entirely different ones.

For each class name, we need to extract the time of the first use of the word. This is the minimum information that the algorithm requires to filter anachronic terms. We identified three sources that contain information for overlapping subsets of concepts. Some of them provide more precise information on the probability of use of a word during time, rather than just the year or period of first use.

1) Wikidata: Based on crowdsourcing information, Wikidata is a free and open knowledge base that contains data about most of the concepts that are represented as classes in our work. Concretely, it includes the inception time and all possible meanings of the concept. However, Wikidata does not necessarily cover concepts that appeared in the middle ages. As an example, pivoted scissors in our traditional understanding appeared only in the 15th century. Objects with similar shapes cannot be scissors in paintings earlier than the 15th century.

2) NGram Viewer: The Google Ngram Viewer is an online search engine that charts the frequencies of any set of comma-delimited strings using a yearly count of grams found in sources printed between 1500 and 2008 in Google's text corpora. The interesting property when using this tool is that it provides the probability of the appearance of a certain concept in a certain century. The results of the object detection model can be thus corrected based on probability, rather than just a boolean value. However, this approach has two main limitations. First, Google NGram Viewer only gathers concepts that appear between 1500 and 2008. Some of the objects that appeared earlier, such as a spoon or a wall clock, do not show up with full information. The second limitation refers to the evolving meaning of words throughout time, which can add some noise in the detection of objects. In Figure 1, we can see that the probability for the word "car" is significantly high for the 15th century. This is impossible for the meaning of "car" that we use nowadays; It would be necessary to disambiguate between the different meanings.

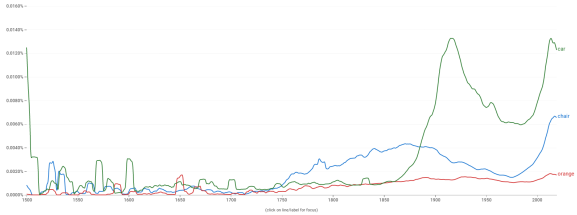

Fig. 1. Probabilities for words car(green), orange(orange), chair(blue)

3) Dictionary approach: Merriam-Webster has been America's leading and most-trusted provider of language information. Apart from the modern and archaic meaning of a word, it contains sections on the first known use, history, and etymology for the word. Based on this information we created a structure which stores, for each of the 80 classes, the period of inception of the concept it represents based on the correct understanding as related to its modern representation. Unfortunately, choosing the correct meaning can not be always automated; after the creation of this structure, this has to be checked manually for those classes that represent homonymic words or whose representation has changed significantly over time.

We have to understand that the correct creation of this time-holding structure has a significant impact on the accuracy of the method. There are a lot of variations in the representation of objects over time. Apart from that, we have to take into account the territorial context, given that the date of first use of an object in different countries can be different. For instance, fireworks were invented in the 10th century in China but were produced in Europe only in the 14th century, becoming popular during the 17th century. Taking this into account, we applied the following rules:

- Our area of interest is the European cultural heritage.
- The period of interest is limited between the 12th and the 19th centuries.
- The metadata we use for the time-holding structure and the object detection model should refer to the same shaped object. For instance, despite the fact that spring scissors were used in Europe before the 16th century, we use as the date of invention of scissors the 16th century because the object detection model can only detect the traditional form of scissors, which came in use in Europe during the 16th and 17th centuries. Spring scissors will only be correctly identified by object detection algorithms if enough images are included in the training set.

B. Combining Deep Learning and Semantic metadata

The second step of our algorithm relies on the quality of the semantic information extracted in the first step and summarized in the time-holding structure. Its correct creation is a key point for the successful implementation of our approach, and it isn´t straightforward because of aspects such as the different possible meanings of a word (homonyms), different

inception times for objects whose representation changes over time, etc(bananas, scissors, truck). Of the available sources described in the previous section, we choose the dictionary approach, in which we can choose the correct meaning of the word and take into account its historical context.

The implementation scrapes the data automatically using software such as Selenium. In some cases the time-holding structure needs manual correction, but most of the content it is generated automatically. Class "Tie" is an example that needs correction. In the modern understanding, a tie is a long piece of cloth worn around the neck or shoulders, which is used with official suits. However, in the 12th century a tie had the same meaning but different shapes. The pretrained object detection model based on MS COCO classes focuses on the detection of modern ties and doesn't know how ties looked like in the 12th century. This implies that all detections of ties (in their modern shape, the only one the pre-trained model recognizes) in paintings earlier than the 18th century (the time of inception of a tie in its modern form) will be falsely labeled as such when they can actually represent different objects, such as a hanging rope. The dictionary approach doesn't allow us to understand these aspects automatically; as a result, we have to manually change the inception date for the tie to the 18th century.

This step first detects the objects based on the Mask-RCNN model, which it then corrects using the information in the time-holding structure to generate a refined list of objects with associated classes. We describe these tasks in detail below.

1) Object detection model: We use the Mask-RCNN (Kaiming et al. (2017)) detection network to identify bounding boxes in paintings and generate candidate labels for each of them. We use the MS COCO (Tsung-Yi Lin et al. (2014)) pre-trained model, which can detect 80 classes. Some of them are general and can appear in paintings from any period, such as person, bird, dog, or cat. Some others appeared in middle ages, such as scissors or broccoli. Finally, others are modern concepts that appeared in the 19th and 20th centuries, such as motorbike, tv, remote, mouse, or keyboard. The present work focuses on paintings from the 12th century on; depending on the time the painting was executed, some of the concepts returned by the algorithm may not yet exist.

The model returns a list of instances (detected objects), one list per image. The instance contains the following information:

- rois: [N, (y1, x1, y2, x2)] a bounding box for each instance
- class_ids: [N] int class IDs for the objects
- scores: [N] float probability scores for the class IDs
- masks: [H, W, N] instance binary masks

2) Object corrector: We implement two approaches that use the time-holding structure to correct the results of object detection and generate a more accurate time matrix for an image.

Class correction of possible presence. Taking into account the date of creation of the painting, we can eliminate those classes whose first time use is later than this date. Figure 2 shows an example of this type of correction. As an illustration, we choose a painting by Alonso Cano, "The crucified Christ appears to Saint Teresa of Avila". Apart from person and book, the pretrained Mask-RCNN model detects the crucified Jesus as a bicycle, and the inkwell as a cell phone. By consulting the time-holding structure for the time of first use of bikes and cell phones, we are able to reject these two labels as anachronic.

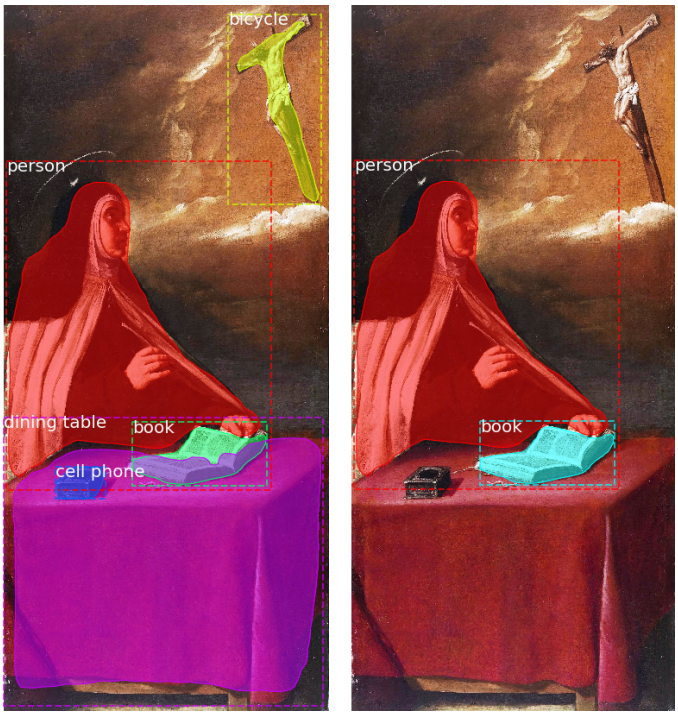

Fig. 2. Object detection with and without correction of possible presence

Class correction based on probabilities. Instead of simply removing an anachronic class, we can try to recommend more time-appropriate labels. The activation function for the typical object detection model architecture (normally a Softmax function) creates one vector with class probabilities for each detected object (i.e. bounding box). The class with the highest probability in the vector associated with a bounding box is the label recommended by the detection model.

It could happen that this class cannot appear in the painting if we consider the time context, in which case our algorithm checks the entire probability vector to find the next most probable class that fits the time of the painting.

To illustrate this, let us look at the "Saint George and the Dragon" painting by Raphael(Figure 3). The model detects a horse and a person, which is acceptable. However, instead of a person, the princess in the background was detected as a teddy bear. Here is the list of probabilities which corresponds to the princess bounding box:

The table I contains all classes with probability more than zero, in descending order. One by one we check if the class can appear in a painting according to the information in the time-holding structure and the date the painting was executed (the 16th century for this example). The algorithm stops with the first class which fits the time restrictions; in this case,

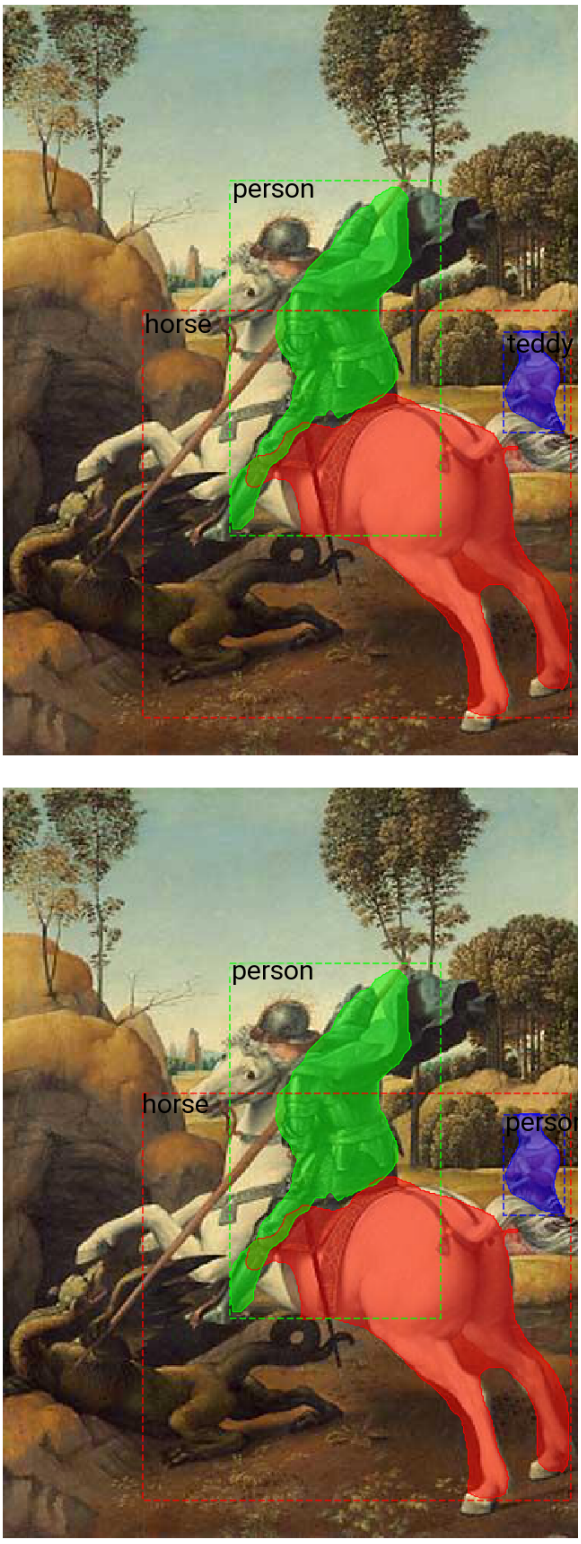

Fig. 3. Object detection with and without correction of classes

after teddy bear, the most probable class is person. If the list does not contain any suitable class, the detected object will be marked as background. Instead of deleting anachronic objects, this approach allows us to correct them based on external meta information, which increases the precision of the object detector.

TABLE I
LIST OF POSSIBLE CLASSES

Id	Name	Probability
3	person	0.222406
4	pot plant	0.138698
2	elephant	0.127589
0	bed	0.037131
1	dog	0.020135

IV. EVALUATION

The process of evaluating the effectiveness of our implementation does not follow the typical methodology, due to the absence of paintings in which objects have been labeled according to the classes in the COCO dataset. We basically had to find another way to evaluate based on external meta information, without requiring manual labeling. Our evaluation method is based on the BLEU score metric, computed between a reference string and the concatenated names of classes detected by the object detection model, with and without using the time matrix, for each painting. The reference string should reflect the set of main objects relevant to the painting, that can be potentially detected by the model. Hence, we decided to use only those sources of data that are clearly relevant for these objects; for instance, we don't use paintings of angels because MaskRCNN cannot detect this class. Concretely, we use Wikimedia Commons because of two reasons:

- The paintings are grouped by categories, whose names contain the semantic information we need for the reference string.
- All images are distributed under the CC license.

We choose two of the most popular classes from the MS COCO dataset, specifically person and book. To perform the evaluation, we choose the Wikimedia Commons category "Paintings of people holding books". The noun phrases present in the name of the category are people and books, which closely correspond to the COCO classes person and book and by concatenation give us the reference string: "person book". The "Paintings of people holding books" category contains 68 paintings. We downloaded them and applied the object detection algorithms with and without using the time matrix. Lastly, we concatenate the classes we detected as a result, and compute three different metrics between the resulting strings and the reference string: Precision, Brevity penalty, and the BLEU score. Table II contains the mean of each metric among all paintings in the category, with and without using the time-matrix.

TABLE II
EVALUATION

Approach	Precision	Brevity penalty	BLEU score
Without Time matrix	0.65	0.77	0.46
With Time matrix	0.76	0.70	0.51

The BLEU score is obtained by multiplying the precision by a measure that penalizes sentences that are shorter than reference strings. This measure is called the brevity penalty. If the output is as long or longer than the reference sentence, the penalty is 1. Since we're multiplying precision by it, that doesn't change the final output. On the other hand, if the output is shorter than the reference sentence, we divide the length of the reference by the length of the output, subtract one from that, and raise constant e to the power of the result of the subtraction to obtain the BLEU score. The longer the reference sentence and the shorter our output, the closer will the brevity penalty be to zero.

Although the mean of the brevity penalty values is higher in the approach that doesn´t use the time matrix, both precision and the BLEU score show an improvement in object detection when using the time matrix approach. The reason for the higher score of the brevity penalty without the time matrix is rooted in the implementation of class correction for possible presence. This approach deletes any object which is not related to the period of the painting, and, in the typical case, can return outputs with just one detected object (for instance "person"). This results in a brevity penalty of less than 1 and explains a higher score for object detection without the use of a time matrix.

V. Discussion

A. The problem of paintings of the 18th and 19th century

Using the time matrix to improve object detection doesn't always work well, especially for paintings from the 18th and 19th century, in which artists use blurring techniques, gradient color transitions, or other techniques specific to surrealism or impressionism. The main reason for this lack of precision when computing the list of class probabilities for bounding boxes is due to the fact that recognition is mostly based on the shapes of the objects. The more modern a painting, the less likely it is for it to be purely representational or follow the traditional school of painting.

B. The problem of relationships between objects

We clearly understand that relationships between objects may contain significant symbolic or iconographic meta-information that can be used in improving object detection. However, state-of-art object detection models don't allow extraction of this information. The lack of this information can evidently influence the accuracy of the time matrix method in the same way that it affects all the other object detection algorithms.

C. The problem of overlapping objects

Neural Networks are black-box models. They make great predictions, and you can easily check the computations they performed to make these predictions; nevertheless, it is usually hard to explain in intuitive terms why the predictions are as such. For example, if a neural network says that a particular person appears in a picture, it is hard to know what contributed to this prediction: did the model recognize that person's eyes? Her mouth? Her nose? Her shoes? Or even the couch that she was sitting on? This aspect of Neural Networks is a reason for the problem of "overlapped objects". Our challenge comes from the fact that applying class correction based on probabilities doesn't work for some cases where two objects significantly overlap. In such cases quite often it is the case that, when the first class label is incorrect for a bounding box, the second-highest probability label refers to an object with which the bounding box under analysis overlaps instead of another class contained in the bounding box - which may be more similar to the reality. Let's review this simple example. In Figure 4 there are two bicycles and one person. Let us use the time matrix method to identify these objects. For the sake of the argument, assume for a moment that this picture is from the 15th century. Below you see the list of class probabilities for each bounding box. (Tables III, IV).

Fig. 4. Object detection with correction of classes

TABLE III
List of possible classes for red bicycle. For class bicycle we chose class bench

Id	Name	Probability
3	motorcycle	0.004800
0	bench	0.002017
1	chair	0.001887
2	horse	0.000270
4	tennis racket	0.000136

TABLE IV
List of possible classes for purple bicycle. For class bicycle we chose class person

Id	Name	Probability
4	person	0.034217
3	motorcycle	0.004800
0	bench	0.002017
1	chair	0.001887
2	horse	0.000270
5	tennis racket	0.000136

For the red bicycle, where no person is present in the bounding box, our model suggests a bench as a better label of the (anachronic for the 15th century) bicycle. This seems like a reasonable possibility. The purple bicycle´s bounding box, on the other hand, includes parts of a person (hand, legs),

overlapping with the bicycle. As a result, the model suggests a person as the better label, which is evidently an error.

VI. Conclusion and future work

This paper proposes an approach for improving the precision of object detection in paintings based on combining Deep Learning and Semantic Metadata extraction. The metadata refers to the time of first use of the words representing the objects and form what we call a time matrix. The creation date of a painting is compared with the information in the time-holding structure to detect and replace anachronic objects with the most probable objects that fit the time period of the painting. The implementation is based on a "detector-corrector" structure, which we plan to implement at the level of the neural network. The architecture of the implementation will include a detector, a neural network which detects the possibility of the presence of specific objects based on semantic metadata, and a concatenation layer that will align the outputs of the two to correct the final list of detected objects.

Acknowledgment

This research has been supported by the Saint George on a Bike project 2018-EU-IA-0104, co-financed by the Connecting Europe Facility of the European Union.

References

[1] Harald Sack, "Artificial Intelligence for Cultural Heritage" Presentation at Europeana 2019,Lisbon, Portugal, November 2019.

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár, "Microsoft COCO: Common Objects in Context", arXiv: 1405.0312, November 2014.

[3] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, "Mask R-CNN", arXiv: 1703.06870, March 2017.

[4] Fischler, M. A., and Elschlager, R. The representation and matching of pictorial structures. IEEE Trans. Comput. C-22, 67–92. doi:10.1109/T-C.1973.223602, 1973

[5] Osuna, E., Freund, R., and Girosi, F. "Training support vector machines: an application to face detection," in Proc. of the IEEE Conference of Computer Vision and Pattern Recognition (San Juan: IEEE), 130–136. doi:10.1109/CVPR.1997.609310, 1997

[6] Rowley, H. A., Baluja, S., and Kanade, T. Neural network-based detection. IEEE Trans. Pattern Anal. Mach. Intell. 20, 23–28. doi:10.1109/34.655647, 1998

[7] Sung, K.-K., and Poggio, T. Example-based learning for viewed-based human face detection. IEEE Trans. Pattern Anal. Mach. Intell. 20, 39–51. doi:10.1109/34.655648, 1998

[8] Schneiderman, H., and Kanade, T. "A statistical model for 3D object detection applied to faces and cars," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (Hilton Head, SC: IEEE), 746–751., 2000

[9] Yang, M.-H., Ahuja, N., and Kriegman, D. "Mixtures of linear subspaces for face detection," in Proc. Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition (Grenoble: IEEE), 70–76., 2000

[10] Fleuret, F., and Geman, D., Coarse-to-fine face detection. Int. J. Comput. Vis. 41, 85–107. doi:10.1023/A:1011113216584, 2001

[11] Romdhani, S., Torr, P., Scholkopf, B., and Blake, A. "Computationally efficient face detection," in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, Vol. 2 (Vancouver, BC: IEEE), 695–700. doi:10.1109/ICCV.2001.937694, 2001

[12] Viola, P., and Jones, M., "Rapid object detection using a boosted cascade of simple features," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (Kauai: IEEE), 511–518. doi:10.1109/CVPR.2001.990517, 2001

[13] Khaoula Drid,Mebarka Allaoui, Mohammed Lamine KherfiObject Detector Combination for Increasing Accuracy and Detecting More Overlapping Objects, ICISP, 2020