# Automatic Matching of Paintings and Descriptions in Art-Historic Archives using Multimodal Analysis

**Christian Bartz*, Nitisha Jain*, Ralf Krestel**
Hasso Plattner Institute
University of Potsdam, 14482 Potsdam, Germany
{firstname.lastname}@hpi.de

**Abstract**

Cultural heritage data plays a pivotal role in the understanding of human history and culture. A wealth of information is buried in art-historic archives which can be extracted via digitization and analysis. This information can facilitate search and browsing, help art historians to track the provenance of artworks and enable wider semantic text exploration for digital cultural resources. However, this information is contained in images of artworks, as well as textual descriptions or annotations accompanied with the images. During the digitization of such resources, the valuable associations between the images and texts are frequently lost. In this project description, we propose an approach to retrieve the associations between images and texts for artworks from art-historic archives. To this end, we use machine learning to generate text descriptions for the extracted images on the one hand, and to detect descriptive phrases and titles of images from the text on the other hand. Finally, we use embeddings to align both, the descriptions and the images.

**Keywords:** cultural heritage, keyphrase identification, machine learning, natural language processing, computer vision

## 1. Introduction

In the age of big data, there is increasing attention on the digitization of cultural heritage collections and their availability as digital libraries to aid wider access and exploration of this previously opaque data. A number of museums, libraries, and other cultural institutions (e.g. Europeana, Getty Research Institute, Wildenstein Plattner Institute, and the Rijks Museum[1]) have invested significant efforts to digitize their collections consisting of old art books, catalogues for art exhibitions and auctions, etc. Initiatives, such as OpenGLAM[2], promote collaboration among these cultural institutions for research on shared resources.

The volume and heterogeneity of these digitized collections necessitates automated analysis of this data. Modern data science tools can assist in deriving insights from the images, as well as from the textual content of these collections. In addition to the actual content, cultural heritage datasets, such as art-historic corpora, are often enriched with meta-data that can provide useful information and context for automatic tools. One example of meta-data is the associations between the artwork images and the texts contained in catalogues and books. Art-historic corpora contain textual information in the form of captions of images (often depicting the titles of artworks), as well as the description of artworks including their creator, year, and, in case of auction catalogues, price information. During the digitization step, images from physical pages are typically scanned and the text is retrieved by means of Optical Character Recognition (OCR) technology. Although these techniques have been fairly improved to minimize the error rate, the information about the association between the images of artworks and their corresponding text excerpts is not retained. This is especially true when multiple images and text excerpts are present on a single page. The availability of such associations between images and texts can help with multimodal semantic analysis of artworks, wherein important descriptive features can be identified from the images, while the corresponding text might provide additional background information about the style and context of the artwork and the artist. In some cases, the text can also provide further evidence and confirmation for the features inferred from the images, and vice versa. For example, consider a case where image analysis correctly ascertains that a particular painting depicts a house with mountains in the background, and the associated text description not only contains terms such as mountains and house but also mentions that this painting is in landscape orientation, then the painting can be categorized and tagged as such. This meta-data derived from the associations between images and texts could be particularly useful in search and exploration of lost artworks, where only a few indicators about the sought-out artworks are known beforehand. An art historian would greatly benefit from image-text associations while retrieving images of artworks from a database by searching on the basis of a few keywords (style, motif, orientation and other features) that can be found in the corresponding description texts.

The matching of images with texts can be done at various levels of granularity based on the size of the data under consideration. Each level poses different challenges and demands unique techniques to achieve desired results. For instance, multiple images on a single catalogue page have a higher likelihood to belong to a common theme or topic. Matching at this level requires techniques to differentiate between similar images, as well as to extract the most distinctive keyphrases from the text descriptions. When the task is scaled to a large corpus of multiple types of catalogue pages, the matching will need to be performed between a large number of possible pairs. To narrow down the search space, the images could be classified on the basis of their art styles by identifying and leveraging common themes in the corpus. This would be followed by matching

*both authors contributed equally

[1] www.europeana.eu, www.getty.edu/research, wpi.art, www.rijksmuseum.nl

[2] openglam.org

on basis of differentiating characteristics as before.

In this work, we propose a generic framework to retrieve the associations between images of artworks and texts from art-historic archives by means of automated approaches. Due to the multimodal nature of this task, our solution is comprised of a combination of techniques from computer vision, as well as natural language processing. While image captioning techniques are employed to identify and tag the images of artworks, Named Entity Recognition (NER) and keyphrase identification techniques are used for the extraction of descriptive terms from the text excerpts. Lastly, to establish the associations between the images and texts, we perform the representation and alignment of the description texts obtained from above techniques via embeddings.

This paper describes an ongoing project on multimodal analysis of cultural heritage datasets. The project is a part of a larger collaboration[3] with the Wildenstein Plattner Institute[4] that was founded to promote scholarly research on cultural heritage collections. The contributions of this paper are : (1) Introduce the novel task of matching artwork images to their text descriptions in art-historic corpora. (2) Propose a framework to extract descriptive features from images and texts of artworks and perform their semantic alignment. (3) Identify evaluation methods for measuring the performance of the framework.

## 2. Related Work

The multimodal nature of our proposed framework is rooted in two different fields. The first field is text analytics for automatic understanding of the semantics of extracted texts. The second field is image analysis for the extraction of the semantics of images. In this section, we present and outline the relation of previous work that is related to the analysis of cultural heritage data for each of the two fields.

### 2.1. Text Analytics

Analysis of cultural heritage data has been of active research interest for the digital humanities where various works have performed use case driven text analysis of digitized art corpora. For example, there is existing work on performing event extraction for historical events (Segers et al., 2011) and finding parallel passages from cultural heritage archives (Harris et al., 2018). There have been several attempts to create knowledge repositories in the form of knowledge graphs and linked open data collections from art data (Hyvönen and Rantala, 2019; Van Hooland and Verborgh, 2014; Dijkshoorn et al., 2018; De Boer et al., 2012). While these works lay emphasis on extracting facts and useful information from the text, they do not necessarily identify the most representative terms and keyphrases from the text. NER is a related task which has been performed for the cultural heritage domain in several papers (Van Hooland et al., 2013; Ehrmann et al., 2016; Jain and Krestel, 2019). The challenges of this task in the context of noisy OCRed datasets have been discussed previously (Rodriquez et al., 2012) and (Kettunen and Ruokolainen, 2017). While we also require techniques to handle noise in datasets as proposed by these papers, this is not the primary focus of our work. For our text analytics approach, we need to broaden the scope beyond NER to identify the most important phrases from the digitized texts that contain descriptions of the artworks, which has not been addressed by any previous work.

### 2.2. Image Analysis

Automatic image analysis in the domain of art-historical research has been studied in several earlier research works (Huang et al., 2018; Elgammal et al., 2018; Yang et al., 2018; Thomas and Kovashka, 2019). One of the greatest problems of automatic image analysis in the art domain is the availability of suitable training data (Huang et al., 2018; Elgammal et al., 2018; Thomas and Kovashka, 2019). Methods in related work rely on fine-tuning image classification models, pre-trained on photographs, to overcome the problem of the non-availability of training data. Using such pre-trained models often leads to the problem of domain-adaptation, which arises because available models are pre-trained on photos and not on images of artworks. Thomas and Kovashka (Thomas and Kovashka, 2019) propose to use methods of neural style transfer (Gatys et al., 2015) to generate a sufficient amount of training data, based on photographs and a set of artworks that are used as baseline style images. All in all, related methods mainly concentrate on the problem of image classification (Thomas and Kovashka, 2019), style, genre, and artist classification (Huang et al., 2018; Elgammal et al., 2018; Lecoutre et al., 2017), or time period and type classification (Yang et al., 2018). So far, there has been no work on performing automatic image captioning for artworks, which is one of the focus points of our work.

### 2.3. Combination of Text Analysis and Image Analysis

A natural idea is to embed the features extracted from both modalities into a common semantic subspace (Kiros et al., 2014; Liu et al., 2019), where a model is learned, that embeds text and image features in a shared high dimensional embedding space. The goal of the embedding is to bring the concepts, obtained from text and image analysis that have the same meaning, as close to each other as possible. In our work, we want to follow this basic embedding approach and use the combined information from text analysis models and image analysis models for the matching of an image to its corresponding text in art-historic corpora.

## 3. Matching Paintings and Descriptions

In this section, we discuss our proposed framework for performing the matching of artwork images to associated texts and describe the different components in detail. We envision to create an automated pipeline that takes the raw scan of a page of any catalogue or book as input and performs several operations on it: (1) Text is localized and recognized using off-the-shelf OCR software. (2) The text analysis component extracts the most representative terms with help of NER and keyphrase identification. (3) In parallel, images on each page are localized and the image analysis

[3] https://hpi.de/naumann/projects/web-science/caart.html

[4] https://wpi.art/

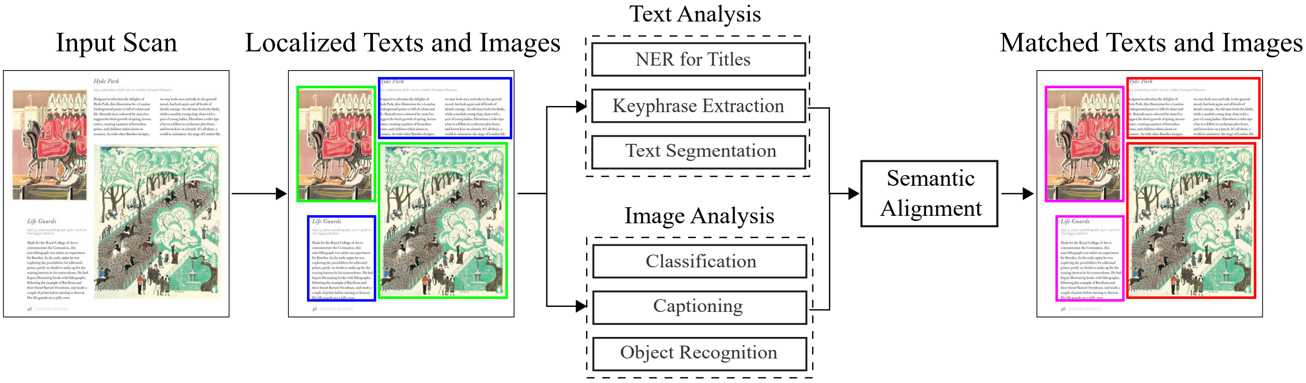

Figure 1: Overview of proposed framework

component extracts semantic meta-data from each image. (4) In the semantic alignment component, the results of step (1) and step (2) are embedded into a shared space and are used for matching and linking of the images to texts. Figure 1 provides a structural overview of the proposed framework. In the remainder of this section, we will explain the challenges and possible approaches towards a solution for each of the sub-tasks, namely text analysis, image analysis and semantic alignment of text and images.

## 3.1. Text Analysis

An intuitive way to match an image in an art catalogue with its description is via the *title* of the artwork. Assuming that the description of any given artwork will include the title, a human would be able to identify the relevant image on the page by matching the title with the caption of the image. Since any caption text associated with the images (including their titles) is usually not available after the digitization process, the matching for digitized datasets has to be performed solely on the basis of the features or tags that are extracted from the images. However, matching on the basis of titles alone is still not a viable approach due to several reasons. Firstly, as discussed in (Jain and Krestel, 2019), the identification of titles of artworks in textual descriptions is itself a non-trivial task and shows sub-optimal performance with existing NER tools. Secondly, even for a scenario where the titles are correctly identified from the text descriptions, they are not always sufficiently representative of the artworks. An example would be modern art paintings where the titles may not be descriptive of the motif in the painting and thus not helpful for matching. Titles are also not useful in the case of old portrait paintings where it is difficult to uniquely identify an image from the name of the depicted person (which is also the title in most cases). This illustrates that titles of artworks might not necessarily contain the required semantic information for the matching of texts with artwork images. As our approach relies on semantic alignment for the matching, it is important to focus on identifying the most salient parts of the description of paintings in the text.

To this end, there are two methods we would like to investigate. The first is to look at *keyphrase extraction*, which identifies and extracts the most representative phrases from a document. Supervised approaches for keyphrase identification are popular (Jiang et al., 2009), however they need annotated training data which is tricky to generate for art datasets. Owing to the subjective nature of the domain, a gold standard training dataset is difficult to obtain due to lack of agreement by non-expert annotators. Therefore, in this work, we would like to turn to unsupervised keyphrase extraction techniques (Hasan and Ng, 2010; Mihalcea and Tarau, 2004) where the task is performed with help of semantic relatedness. Further, to fine-tune this task for the art domain, we want to pursue domain-specific keyphrase extraction techniques (Wu et al., 2005; Hulth et al., 2001).

The second method is to directly *embed the text* in the semantic space. For this approach, we would need to perform the segmentation of the text excerpts, followed by identification of the relevant segments that contain descriptions of the artwork images. This is important particularly for art books where the texts include discussions not only about artworks, but also about the artists, art styles, etc.

## 3.2. Image Analysis

In order to analyze the semantic content of digitized images, we plan to use modern computer vision methods based on deep learning. Computer vision tasks which are very close to the tasks that we want to perform, are automatic image classification (Krizhevsky et al., 2012), image captioning (Xu et al., 2015), i.e. the generation of textual desciptions of depicted content, and object detection (Ren et al., 2015). All of these methods extract semantic information from images and have been shown to work very well on photographs. The most challenging problem in working with images of artworks is that photographs have a very different underlying data distribution than images of artworks, especially paintings. This makes it necessary to train machine learning models directly on images of artworks. However, large-scale annotated training data sets with artworks are not available.

There exist some datasets that contain artworks and annotations (e.g. art style), such as the WiKiArt database[5], or the OmniArt dataset (Strezoski and Worring, 2018). However, none of these datasets can be used for image classification or automatic image captioning, since they lack the annotations required for these tasks. We can, however, make use of photographies and their annotations, which are available in large-scale datasets.

[5] https://www.wikiart.org

To this end, we want to follow (Thomas and Kovashka, 2019) and use methods of neural style transfer (Gatys et al., 2015; Yao et al., 2019) to create new large-scale art centered datasets for image classification, image captioning, and object recognition on artworks. For image classification, we want to use the ImageNet dataset (Deng et al., 2009) and create a new ArtImageNet dataset that we will use as a base model in a subsequent step to train an image classification model. For image captioning and object detection, we want to use the COCO dataset (Lin et al., 2014) and fine-tune the image classification model that we created earlier for each of these tasks. For creating the artistic versions of the photographs from each dataset, we want use the WikiArt or the OmniArt dataset, as artistic style images.

### 3.3. Semantic Alignment

After performing the extraction of meaningful features from textual data and image data in parallel, the next step is to find ways of aligning the extracted information and match an image to its accompanying text. For this, we want to embed the output from the text analysis and image analysis component in a common semantic space (i.e via word embeddings), where we can represent similar concepts close to each other and thereby find text and image pairs that might be a good match. Another idea, is to use the feature vectors created by the image analysis methods and train a further model to embed them into the same semantic space as the word embeddings of the relevant texts and phrases. Such an alignment in a common semantic subspace will allow us to perform image retrieval for a given text query and also text retrieval for a given query image.

## 4. Evaluation Methods

In this section, we address the question of the evaluation of the proposed framework. This question can be divided into three parts: 1) How to evaluate the proposed text analysis methods regarding their adjustments to fit the challenges of extracting relevant information from art-historic archives. 2) How to evaluate the proposed image analysis methods in the context of art analysis, since state-of-the-art image analysis methods are mainly trained on photographs, which are quite different from artworks. 3) How to evaluate the framework that performs the alignment of the information from the text and image analysis components to enable matching of images with their textual descriptions.

**Evaluation of Text Analysis.** As discussed in Section 3.1., the availability of annotated datasets for training and evaluation is a major bottleneck for evaluating semantic representations, especially in the art domain. For this, we plan to enlist the help of domain experts for the creation of a smaller gold standard test dataset that will include annotations for the most important textual segments or keyphrases for identifying the corresponding images. The performance of our text analytics approaches can then be measured by comparing the results with the gold standard annotations in terms of precision and recall.

**Evaluation of Image Analysis.** The most important aspect in evaluating the image analysis methods is how well they can be adapted to work on images of art, despite having only a very small amount of annotated real training data available. Though there are datasets available, e.g. provided by Europeana[6], their annotations do not follow a common scheme which limits their utility for our purpose. As we propose in Section 3.2., we want to use methods of neural style transfer to create a sufficient amount of training data. On the one hand, we want to focus on the plain numerical evaluation of these models, using well known evaluation metrics, like classification accuracy for image classification, precision, recall and f-measure, as well as average precision for object detection, and metrics for image captioning evaluation, e.g. BLEU-score (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005), CIDEr (Vedantam et al., 2015), ROUGE (Lin, 2004), SPICE (Anderson et al., 2016), BERTScore (Zhang et al., 2019), or MoverScore (Zhao et al., 2019). On the other hand, we are interested in evaluating the influence of different base models that are used to create our image captioning for art, or object detection models. Here, we want to compare a standard ImageNet model with a model created with our ArtImageNet dataset. We want to use this to examine whether automatic methods can successfully be used to generate novel annotated data, based on already available data.

**Evaluation of Text and Image Alignment.** The task of matching a given text to an image in an art catalogue can be cast as a retrieval task. This retrieval task consists of two aspects. The first aspect is to retrieve an image, given a textual description and the second is to retrieve a textual description, given an input image. We can use standard image retrieval evaluation methods, also used in related work (Kiros et al., 2014; Liu et al., 2019), such as recall at k ($R@K$), for the evaluation. Here, we are interested in different values of $K$ based on the granularity of the current search. If we only consider a single page with text and several images, we are interested in the recall at $K = 1$, whereas if we want to retrieve an image to a given text over an entire catalogue, we are interested in the performance at higher values of $K$. Since the problem of extracting images and their textual descriptions from art-historic archives has not been studied before, there are no evaluation datasets available. For the evaluation of our method, it will be important to create an evaluation dataset with help from domain experts that includes different levels of granularity, for measuring the performance of this kind of retrieval task.

## 5. Conclusion

In this paper, we present the description of a project that deals with the novel task of matching artwork images to their corresponding text descriptions in digitized art-historic corpora. We provide an overview of the related work and challenges in this domain and describe a possible framework to tackle the problem of image and text alignment. Furthermore, we give an overview of the possible evaluation methods that we want to use for evaluating each component as well as the overall performance of our proposed framework.

**Acknowledgement.** We thank the Wildenstein Plattner Institute for providing access to their art-historic archives.

[6] https://www.europeana.eu/en/collections/topic/190-art

# 6. References

Anderson, P., Fernando, B., Johnson, M., and Gould, S. (2016). SPICE: Semantic Propositional Image Caption Evaluation. In *Proceedings of the European Conference on Computer Vision*, pages 382–398.

Banerjee, S. and Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72.

De Boer, V., Wielemaker, J., Van Gent, J., Hildebrand, M., Isaac, A., Van Ossenbruggen, J., and Schreiber, G. (2012). Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. In *Proceedings of the Extended Semantic Web Conference*, pages 733–747.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Dijkshoorn, C., Jongma, L., Aroyo, L., Van Ossenbruggen, J., Schreiber, G., ter Weele, W., and Wielemaker, J. (2018). The Rijksmuseum Collection as Linked Data. *Semantic Web*, 9(2):221–230.

Ehrmann, M., Colavizza, G., Rochat, Y., and Kaplan, F. (2016). Diachronic Evaluation of NER Systems on Old Newspapers. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)*, pages 97–107.

Elgammal, A., Liu, B., Kim, D., Elhoseiny, M., and Mazzone, M. (2018). The Shape of Art History in the Eyes of the Machine. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*.

Gatys, L. A., Ecker, A. S., and Bethge, M. (2015). A Neural Algorithm of Artistic Style. *arXiv:1508.06576 [cs, q-bio]*.

Harris, M., Levene, M., Zhang, D., and Levene, D. (2018). Finding Parallel Passages in Cultural Heritage Archives. *Journal on Computing and Cultural Heritage*, 11(3):1–24.

Hasan, K. S. and Ng, V. (2010). Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-art. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 365–373.

Huang, X., Zhong, S.-h., and Xiao, Z. (2018). Fine-Art Painting Classification via Two-Channel Deep Residual Network. In *Advances in Multimedia Information Processing – PCM 2017*, pages 79–88.

Hulth, A., Karlgren, J., Jonsson, A., Boström, H., and Asker, L. (2001). Automatic Keyword Extraction using Domain Knowledge. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, pages 472–482.

Hyvönen, E. and Rantala, H. (2019). Knowledge-based Relation Discovery in Cultural Heritage Knowledge Graphs. In *Digital Humanities in the Nordic Countries*, pages 230–239.

Jain, N. and Krestel, R. (2019). Who is Mona L.? Identifying Mentions of Artworks in Historical Archives. In *Proceedings of the International Conference on Theory and Practice of Digital Libraries*, pages 115–122.

Jiang, X., Hu, Y., and Li, H. (2009). A Ranking Approach to Keyphrase Extraction. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 756–757.

Kettunen, K. and Ruokolainen, T. (2017). Names, Right or Wrong: Named Entities in an OCRed Historical Finnish Newspaper Collection. In *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage*, pages 181–186.

Kiros, R., Salakhutdinov, R., and Zemel, R. S. (2014). Unifying Visual-Semantic Embeddings with Multimodal Neural Language Models. *arXiv:1411.2539 [cs]*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105.

Lecoutre, A., Negrevergne, B., and Yger, F. (2017). Recognizing Art Style Automatically in Painting with Deep Learning. In *Proceedings of the Asian Conference on Machine Learning*, pages 327–342.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Proceedings of the European Conference on Computer Vision*, pages 740–755.

Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Proceedings of Workshop on Text Summarization Branches out, Post2Conference Workshop of ACL*.

Liu, Y., Guo, Y., Liu, L., Bakker, E. M., and Lew, M. S. (2019). CycleMatch: A cycle-consistent embedding network for image-text matching. *Pattern Recognition*, 93:365–379.

Mihalcea, R. and Tarau, P. (2004). TextRank: Bringing Order into Text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 404–411.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28*, pages 91–99.

Rodriquez, K. J., Bryant, M., Blanke, T., and Luszczynska, M. (2012). Comparison of Named Entity Recognition Tools for Raw OCR Text. In *Konvens*, pages 410–414.

Segers, R., Van Erp, M., Van Der Meij, L., Aroyo, L., van Ossenbruggen, J., Schreiber, G., Wielinga, B., Oomen, J., and Jacobs, G. (2011). Hacking History via Event Extraction. In *Proceedings of the 6th International Conference on Knowledge Capture*, pages 161–162.

Strezoski, G. and Worring, M. (2018). OmniArt: A Large-scale Artistic Benchmark. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(4):88:1–88:21.

Thomas, C. and Kovashka, A. (2019). Artistic Object Recognition by Unsupervised Style Adaptation. In *Proceedings of the Asian Conference on Computer Vision ACCV 2018*, pages 460–476.

Van Hooland, S. and Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish your Metadata*.

Van Hooland, S., De Wilde, M., Verborgh, R., Steiner, T., and Van de Walle, R. (2013). Exploring Entity Recognition and Disambiguation for Cultural Heritage Collections. *Digital Scholarship in the Humanities*, 30(2):262–279.

Vedantam, R., Lawrence Zitnick, C., and Parikh, D. (2015). CIDEr: Consensus-Based Image Description Evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4566–4575.

Wu, Y.-f. B., Li, Q., Bot, R. S., and Chen, X. (2005). Domain-Specific Keyphrase Extraction. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 283–284.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A. C., Salakhutdinov, R., Zemel, R. S., and Bengio, Y. (2015). Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pages 2048–2057.

Yang, S., Oh, B. M., Merchant, D., Howe, B., and West, J. (2018). Classifying Digitized Art Type and Time Period. In *Proceedings of the 1st Workshop on Data Science for Digital Art History-Tacking Big Data*.

Yao, Y., Ren, J., Xie, X., Liu, W., Liu, Y.-J., and Wang, J. (2019). Attention-Aware Multi-Stroke Style Transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1467–1475.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2019). BERTScore: Evaluating Text Generation with BERT. *arXiv:1904.09675 [cs]*.

Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C. M., and Eger, S. (2019). MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP)*, pages 563–578.