Enhancing accessibility and discoverability of digital archive images through automated image recognition tool

Akara Thammastitkul[1], Jitsanga Petsuwan[2]
[1]Department of Information Studies, Faculty of Humanities and Social Sciences, Burapha University, Chonburi, Thailand
[2]Department of Computer Center, Chaiyaphum Hospital, Chaiyaphum, Thailand

Article Info

Article history:

Received Mar 3, 2023
Revised Nov 13, 2023
Accepted Dec 31, 2023

Keywords:

Archive image
Artificial intelligence
Image recognition
Metadata
Tagging

ABSTRACT

This research paper presents a comprehensive evaluation of the effectiveness of Imagga and Google cloud vision application programming interface (API) as image recognition tools for generating metadata in digital archive images. The assessment encompasses a diverse range of archive images, including those without text, images with text, and both color and black-and-white images. Through the use of evaluation metrics such as cosine similarity, word overlap similarity, recall, precision, and F1 score, the performance of these tools is quantitatively measured. The findings highlight the strong individual performance of both Imagga and Google cloud vision API, with the combined metadata outputs achieving significantly higher scores across all metrics. This emphasizes the potential benefits of employing a combined approach, leveraging the strengths of multiple tools to enhance the reliability and robustness of the metadata extraction process. The findings contribute to the advancement of metadata management in digital archives and underscore the importance of utilizing multiple tools for improved performance in image metadata generation.

Corresponding Author:

Akara Thammastitkul
Department of Information Studies, Faculty of Humanities and Social Sciences, Burapha University
169 Long-Hard Bangsaen Road, Saensook, Muang, Chonburi 20131, Thailand
Email: akara@buu.ac.th

1. INTRODUCTION

The digitization of archives and collections has significantly increased the availability of cultural heritage materials. By converting analog materials into digital format, these valuable resources can be easily accessed, searched, and analyzed through online platforms, overcoming limitations of physical access, preservation, and degradation [1]–[3]. The increased accessibility of digitized cultural heritage materials has democratized access to knowledge and enriched the scholarly community with diverse and previously inaccessible resources, fostering interdisciplinary research and advancing our understanding of human history, culture, and society [4]–[8]. The Internet Archive is a non-profit organization and online platform that arranges a comprehensive digital archive comprising web pages, books, audio recordings, movies, images, software, educational tools, and research materials [9], [10]. It has over 400 billion web pages from various historical periods, over 25 million digitized books, over 5 million audio recordings, over 2 million movies, and over 4 million photos. By digitizing physical archives and collections, the Internet Archive has significantly increased the accessibility of cultural heritage resources, empowering a global audience to explore and engage with these valuable materials. This accessibility benefits not only academic research but also promotes knowledge dissemination and facilitates the preservation of our collective cultural history. However, managing large volumes of digital images in a way that ensures their accessibility and discoverability can pose challenges. To address this, metadata, which refers to descriptive information about digital objects, plays a critical role in enabling access and discovery. Nevertheless, manualy creating metadata for large digital image collections has been time-consuming and costly. Automated metadata extraction using image recognition tools has been proposed as a promising solution to these challenges, facilitating process simplification and enhancing the accessibility and discoverability of digital archive images [11]–[14].

The significance of metadata for digital archive images has been widely recognized in the scholarly literature. Metadata serves as descriptive information that is crucial for the identification, retrieval, and management of digital objects. It enables the discovery and utilization of digital archive images, while also supporting their long-term preservation [15]–[17]. Automated image recognition using deep learning has revolutionized the field by leveraging artificial neural networks to automatically learn representations from large amounts of data [5]–[7], [18]. This has led to significant advancements in image classification [19]–[23], object detection [24]–[26], and image recognition [27]–[29]. These advancements have not only enhanced the accuracy and efficiency of image recognition systems, but have also enabled applications in a variety of fields such as healthcare, autonomous vehicles, and surveillance. Deep learning's significant contribution to image identification is well-documented and well acknowledged within the scientific community. With the ever-increasing availability of high-quality image data and the growing computational capabilities of modern hardware, deep learning has shown tremendous potential in improving the accuracy and efficiency of image recognition tasks. Several image recognitions tools [11], [19], [30] including Imagga, Google cloud vision application programming interface (API), Clarify, Amazon Rekognition, and Microsoft Azure computer vision, are available for metadata extraction. For this study, Imagga [31] and Google cloud vision API [32] were selected based on their outstanding performance and extensive capabilities in metadata generation, particularly in text extraction [30]–[33]. The ability to extract text from images is essential in evaluating archive images since it enables valuable information contained within the images to be retrieved. These tools, which use machine learning algorithms, automatically detect objects, faces, scenes, and other visual components in images, allowing specific metadata to be generated. Evidently, Google cloud vision outperformed Microsoft Cognitive Services when it came to text detection [19]. Zeng and Zhang [20] use google cloud vision API to identify invasive ductal carcinoma. Several studies have evaluated the effectiveness of image recognition tools for metadata extraction in digital archives. For instance, Samani *et al.* [21] utilized the Imagga tagging program to automatically generate image labels from Twitter and Flickr. Another study by the same authors compared the performance of Imagga and Google cloud vision servers in image analysis. The study found that Google cloud vision outperformed Imagga in terms of tag responses and response speed, leading to greater trust in the results. Alqahtani and Alsulaiman [24] used Imagga and Wordnet in their study investigating the security of image-based completely automated public turing test to tell computers and humans apart (CAPTCHA) against attacks based on machine learning. Furthermore, Google cloud vision API was also utilized for content-based image retrieval (CBIR) by Chen and Chen [25] and for Thai vehicle registration certificate by Thammarak *et al.* [26]. Fu and Rui [5] discussed the challenges of managing large personal photo collections due to the proliferation of mobile devices and media cloud services, where image tagging using a combination of models poses a challenge. In addition, Kubany *et al.* [30] conducted a study comparing various deep learning APIs for image multi-label classification using semantic metrics. The research evaluated and compared the performance of 13 commercial and open-source APIs on benchmark datasets. While conventional metrics revealed that Microsoft computer vision, Imagga, and international business machines (IBM) API succeeded well, semantic metrics revealed that InceptionResNet-v2, Inception-v3, and ResNet50 APIs, which are trained with a simple ImageNet dataset, were competition for top semantic performers.

Although image recognition tools are increasingly being used for automated metadata extraction, there is a lack of detailed literature that thoroughly evaluates their performance in diverse contexts. In this study, we aim to address the challenges of managing large volumes of digital images in digital archives by leveraging automated image recognition tools for metadata extraction. The novel aspect of this research involves the comprehensive assessment and comparative analysis of two prominent image recognition systems, Imagga and Google cloud vision API, for metadata extraction from digital archive images. We aim to provide valuable insights into the potential of these tools, individually and in combination, to enhance metadata quality and improve access and discoverability of digital archive images. This research contributes to the ongoing development of digital archives and metadata extraction by offering informative analysis and recommendations, ultimately assisting digital archive managers and practitioners working in the field of digital archives in selecting appropriate image recognition tools for their specific requirements.

This study aims to address this gap by assessing Imagga and Google cloud vision API as effective image recognition systems for extracting metadata from digital archives. The focus of our study is to conduct a comparative analysis of automated image recognition tools for metadata extraction in digital archive images. The study utilizes the collections of the Internet Archive as our dataset, and aims to provide valuable insights into the potential of Imagga, Google cloud vision API, and their combined models to enhance metadata quality, and improve access and discoverability for researchers and practitioners working in the field of digital archives and metadata extraction. The main objective of this study is to provide practical guidance for digital archive managers and researchers in selecting appropriate image recognition tools for metadata extraction. Additionally, this research aims to make a significant contribution to the ongoing development of the field of digital archives and metadata extraction by offering informative analysis and recommendations.

This paper is organized as: i) Section 2 provides the methodology employed in this study, including data collection, metadata extraction, metadata post-processing, and metadata evaluation. ii) Section 3 presents the results and discussion, offering a thorough analysis of the performance of the selected tools. iii) Section 4 concludes the important findings, discusses their implications, and providing recommendations for future research in the field of image metadata extraction.

2. METHOD

The methodology employed in this research study involves a multi-step approach to analyze and interpret the digitized book images obtained from the Internet Archive. The images were gathered exclusively from the Internet Archive's Flickr collection, which contains a wide range of book images submitted by various individuals and institutions. The study gains access to a wide range of book images by accessing this huge digital library, allowing for analysis and interpretation of the digitized content. The following sections provide a detailed description of the methodology employed for data collection, metadata extraction, metadata post-processing, and metadata evaluation. The proposed methodology is shown in Figure 1.

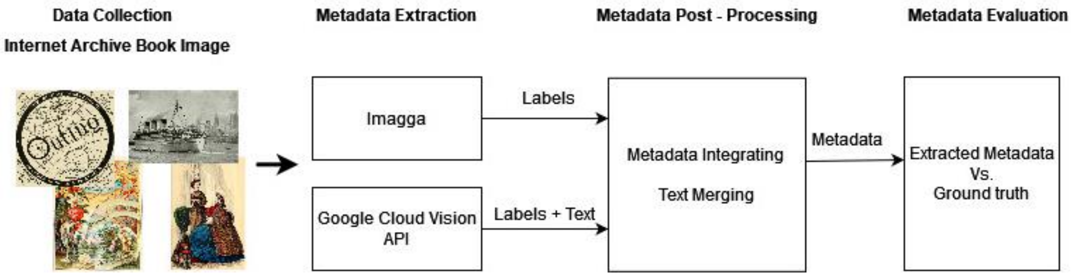


Figure 1. Proposed method

2.1. Data collection

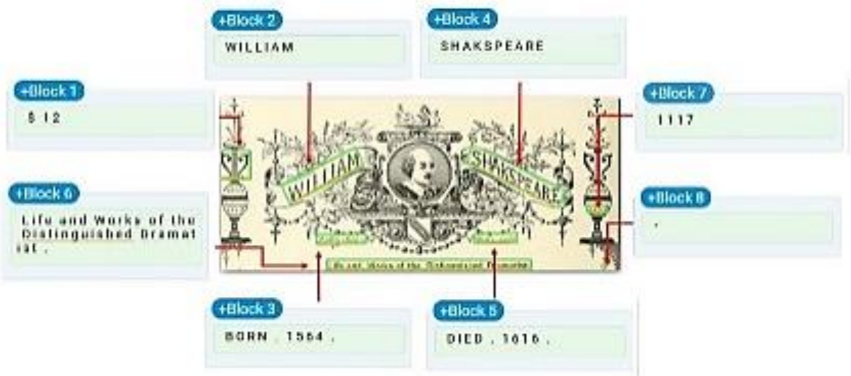
A total of 1,000 digital archive book images were carefully selected from the Internet Archive collection on Flickr. The selection process involved considering various inclusion criteria to ensure representation across book categories, publication periods, and visual attributes. The aim was to establish a well-rounded sample that encompassed diverse aspects of digital books. Conversely, exclusion criteria were applied to eliminate low-quality images, duplicates, and those subject to copyright restrictions. The image selection was categorized into two groups: 500 images without text and 500 images with text. Additionally, to increase the variety of materials evaluated, each group of 500 images consisted of an equal number of color and black-and-white images. In order to create an extensive data set for our study, the images were chosen at random from a collection of digitized texts that dates back to the 16th century and covers a wide range of subjects and times. The images were especially selected based on their license type, limited to those with a public domain or Creative Commons Zero (CC0) license, which grants unrestricted usage rights. We also ensured that the privacy and confidentiality of individuals depicted in the images were respected by not collecting any images that could be considered sensitive or inappropriate.

2.2. Metadata extraction

For the purpose of metadata extraction, we conducted an evaluation of two image recognition tools: Imagga and Google cloud vision API. These tools were selected based on their widespread adoption, availability, and renowned reputation for their accuracy and efficiency [19]–[22]. Imagga is known for its comprehensive image recognition capabilities, including object detection, color analysis, and image categorization, making it suitable for extracting a wide range of metadata from diverse images. Google cloud vision API is widely recognized for its cutting-edge machine learning algorithms and robust image analysis features, such as image labeling, facial recognition, and text recognition, making it a powerful tool for extracting detailed metadata from images. We used the API provided by each tool to extract metadata from the images in our dataset. Each tool was run separately and the metadata generated by each tool was recorded and analyzed. The images were uploaded systematically to the related API's of each tool. Both tools provided results that contained image labels and confidence scores. In addition, the Google cloud vision API results also included extracted text content, which enhanced the data for analysis. Figure 2 provides a representation of the extracted metadata with confidence values and text extracted. Figure 2(a) displays the metadatas using the Imagga tool, Figure 2(b) shows the metadatas using the Google cloud vision API and Figure 2(c) shows text extraction using Google cloud vision API.

(a)

(b)



(c)

Figure 2. Extracted metadata with confidence value using (a) Imagga,
(b) Google cloud vision API, and (c) text extraction using Google cloud vision API

2.3. Metadata post-processing

The metadata post-processing section consists of two main steps: metadata integration and text merging. Following the initial metadata extraction stage, the metadata is refined and processed to ensure accuracy, consistency, and completeness. The metadata obtained from each tool is then integrated and examined in order to eliminating duplicates and establishing a comprehensive and identical metadata representation for image. Another important aspect of metadata post-processing involves text merging. Since the text extraction results from the Google cloud vision API may come in multi-line text elements as shown in Figure 2(c), merging this text to a word can offer valuable context and information about the images. This process enriches the metadata and enhances its usability by providing comprehensive and cohesive textual information alongside other metadata elements.

2.4. Metadata evaluation

To assess the completeness and accuracy of the metadata output, we conducted a thorough evaluation by comparing the metadata extracted by each image recognition tool to the ground-truth metadata for each image. As the metadata obtained from the Internet Archive on Flickr for the digitized image book typically includes fields book identifiers (bookid), publication year (bookyear), decade (bookdecade), century (bookcentury), author (bookauthor), subject categories (booksubject), publisher (bookpublisher), contributing institution (bookcontributor), sponsor (booksponsor), leaf number (bookleafnumber), and collection information (bookcollection). Additional steps are taken to enhance its completeness and accuracy. The ground truth dataset of metadata for each image was careful collected from the text appearing before image and text appearing after image components which have been determined and attached together with images on Flickr. The extracted keywords or definition terms were then employed as reference metadata for our comprehensive evaluation. To assure the reliability and accuracy of the reference data, this selection process was carried out by a language expert. Descriptive statistics were employed to analyze the metadata generated by each image recognition tool, allowing a comparison of their accuracy and efficiency. For evaluation metrics, we utilized cosine similarity, word overlap similarity, precision, recall, and F1 score [34]–[37]. Cosine similarity measures the similarity between two sets of metadata by calculating the cosine of the angle between their corresponding vectors. Word overlap similarity is a metric used to compare the similarity of two sets of text by measuring the overlap of their individual words. Precision quantifies the accuracy of the tool's metadata by calculating the proportion of correctly identified metadata to the total number of identified metadata. Recall gauges the completeness of the tool's metadata by calculating the proportion of correctly identified metadata to the total number of metadata that should have been identified. The F1 score is a harmonic mean of precision and recall that provides a balanced evaluation of both. The cosine similarity, word overlap similarity, precision, recall, and F1 score are calculated using (1)-(5), respectively.

$$cos(\emptyset) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

A and B are vectors representing the metadata of two tools. A·B is the sum of the component-wise product of the two vectors. ‖A‖ and ‖B‖ are the magnitudes of the vectors representing the metadata of each tool.

$$sim_{overlap}(A, B) = \frac{|A \cap B|}{|A|} \tag{2}$$

$|A|$ is the number of metadata that come with the image. $|B|$ is to the number of metadata obtained from the tool. $|A \cap B|$ is to the number of metadata that are common to both the image and the tool.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

$$\text{F1-score} = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

3. RESULTS AND DISCUSSION

To evaluate the accuracy of the metadata generated, a comparison was conducted between the metadata extracted by the image recognition tools, which includes object labels and texts, and the ground-truth metadata. The ground-truth metadata refers to the manually metadata that is considered to be the accurate representation of the metadata for the given images. For example, in Table 1, ground-truth metadata set consists of ['hillsalbumofbiog00hill', '1887', '1880', '1800', 'Hill_Thomas_E_Thomas_Edie_1832_1915', 'Biography', 'Encyclopedias_and_dictionaries', 'Chicago_Hill_Standard_Book_Co', 'University_of_California_Libraries', 'Internet_Archive', '325', 'cdl', 'American', 'Painter', 'Art', 'Shakspeare']. This word set will be compared to the set obtained from the image recognition tools. This comparative analysis enabled us to evaluate the performance of the image recognition tools in accurately identifying and describing the content of the images. Table 1 and Table 2 present an example of comparison of the metadata produced by Imagga, Google cloud vision API, and the combined tool with the actual metadata.

The results from Table 1 show that Imagga mainly focuses on design and decorative elements such as floral patterns and vintage frames, whereas Google cloud vision API identifies more structural elements such as fonts and rectangles, as well as historical and artistic references such as line art and poster graphics in the given text. Table 2 reveals that Imagga predominantly provides labels related to religious and historical themes, while Google cloud vision API offers a wider range of labels, encompassing animals, plants, and fictional characters. The result shows that Imagga and Google cloud vision API produced different labels for the same image, but when their outputs were combined, the resulting metadata included a more comprehensive set of labels that encompassed a wider range of themes and concepts. In order to evaluate the accuracy and efficiency of the metadata generated by each image recognition tool, we conducted an analysis using descriptive statistics. This analysis involved assessing evaluation metrics such as cosine similarity, precision, recall, and F1 score. The results of this analysis were presented in Table 3 and Table 4, providing a comprehensive comparison of the performance of the image recognition tools in terms of these metrics.

Table 1. Example of metadata extracted from black-and-white images with text

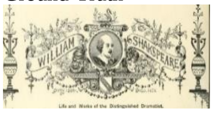
Tool	Metadata
Ground-Truth 	bookid:hillsalbumofbiog00hill bookyear:1887 bookdecade:1880 bookcentury:1800 bookcollection:cdl bookleafnumber:325 booksponsor:Internet_Archive bookcollection:American booksubject:Biography booksubject:Encyclopedias_and_dictionaries bookauthor:Hill_Thomas_E_Thomas_Edie_1832_1915 bookpublisher:Chicago_Hill_Standard_Book_Co_ bookcontributor:University_of_California_Libraries [Text Appearing Before Image] Painter, Art, Shakspeare [Text Appearing After Image] -
Imagga	[Label] Floral, Design, Pattern, Retro, Graphic, Vintage, Decoration, Art, Ornament, Frame
Google cloud vision API	[Label] Organ, Organism, Font, Poster, Art, Rectangle, Illustration, Line Art, Pattern Paper, Drawing, Paper Product, Graphics, History, Visual Arts, Circle [Text] William, Born. 1564. Shakspeare, Died. 1616. Life and Works of The Distinguished Dramatist.
Imagga + Google cloud vision API	Art, Circle, Decoration, Design, Distinguished, Drawing, Dramatist, Floral, Font, Frame, Graphic, Graphics, History, Illustration, Life, Line Art, Organ, Organism, Ornament, Paper, Product, Pattern, Poster, Rectangle, Retro, Shakspeare, Vintage, Visual Arts, William, Work

Table 2. Example of metadata extracted from colour images with text

Tool	Metadata
Ground-Truth 	bookid:hillsalbumofbiog00hill bookyear:1887 bookdecade:1880 bookcentury:1800 booksubject:Biography booksubject:Encyclopedias_and_dictionaries booksponsor:Internet_Archive bookleafnumber:176 bookcollection:cdl bookcollection:Americana bookauthor:Hill_Thomas_E_Thomas_Edie_1832_1915 bookpublisher:Chicago_Hill_Standard_Book_Co_ bookcontributor:University_of_California_Libraries [Text Appearing Before Image] Animal, Cattle, Dog, Fowls, Horse, Sheep, Swine [Text Appearing After Image] Domestic
Imagga	[Label] Comic Book, Altar, Structure, Church, Religion, Art, Print Media, Old, Antique
Google cloud vision API	[Label] Vertebrate, Organism, Art, Plant, Horse, Painting, Illustration, Visual Arts, Tree, Working Animal, Mythology, Fiction, Landscape, Pole, Font, History, Fictional Character [Text] Our Animal Friends What They Do For Us And What We May Do For Them
Imagga + Google cloud vision API	Altar, Animal, Antique, Art, Church, Comic Book, Fiction, Fictional Character, Friends, Font, History, Horse, Illustration, Landscape, Mythology, Old, Organism, Painting, Plant, Pole, Print Media, Religion, Structure, Tree, Vertebrate, Visual Arts, Working Animal

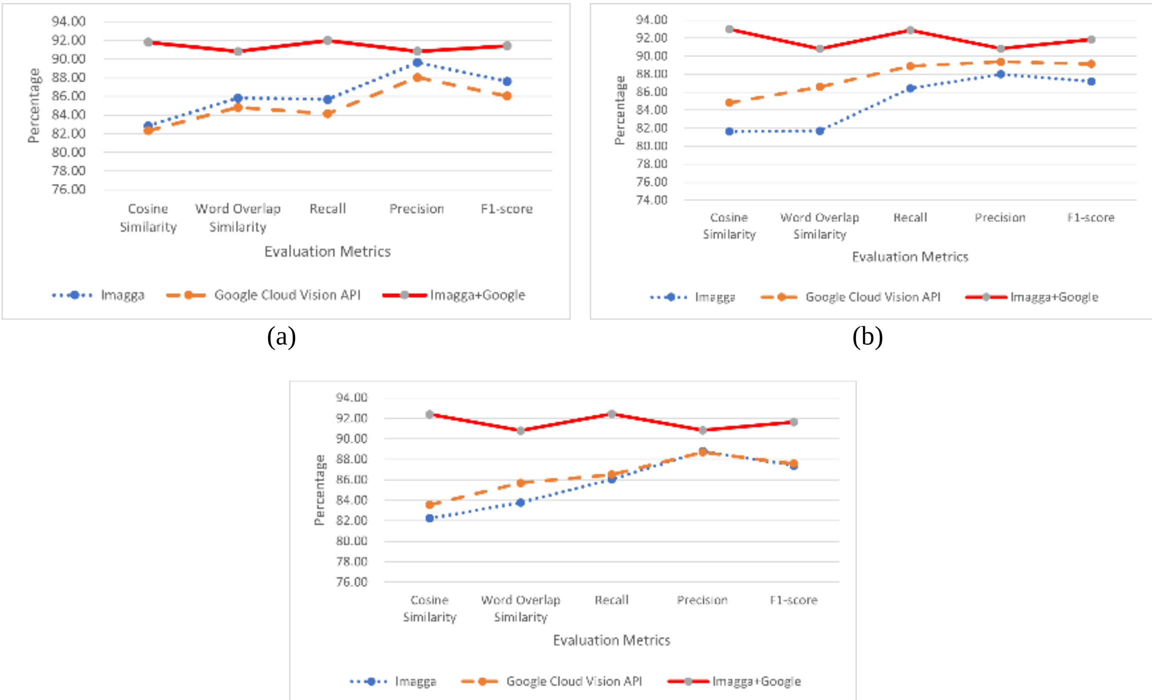
Table 3. Performance metrics of image recognition tools for images without text

Image type	Metric	Imagga	Google cloud vision API	Imagga + Google
Black-white image	Cosine similarity	82.15	81.77	91.12
	Word overlap similarity	85.56	84.21	90.41
	Recall	85.16	83.55	91.87
	Precision	89.14	87.88	90.14
	F1-score	87.10	85.66	91.00
Colour image	Cosine similarity	83.56	82.87	92.48
	Word overlap similarity	86.11	85.44	91.22
	Recall	86.18	84.74	92.12
	Precision	90.11	88.16	91.54
	F1-score	88.10	86.42	91.83

Table 4. Performance metrics of image recognition tools for images with text

Image type	Metric	Imagga	Google cloud vision API	Imagga + Google
Black-white image	Cosine similarity	81.15	83.77	92.21
	Word overlap similarity	80.51	85.52	90.41
	Recall	84.62	87.65	92.34
	Precision	87.44	89.59	90.14
	F1-score	86.01	88.61	91.23
Colour image	Cosine similarity	82.08	85.87	93.77
	Word overlap similarity	82.91	87.65	91.22
	Recall	88.23	90.12	93.41
	Precision	88.52	89.16	91.54
	F1-score	88.37	89.64	92.47

For images without text and image with text, both Imagga and Google cloud vision API exhibit strong individual performance, with comparable scores across various metrics for both black-white and color images. However, when the metadata outputs of both tools are combined, significantly higher scores are achieved in all metrics for both black-white and color images. The findings suggest that combining the strengths of both tools leads to a more robust and reliable metadata extraction process, indicating a synergistic effect. This can be advantageous for digital archive managers and researchers in enhancing the efficiency and accuracy of metadata management. By applying the combined tool, digital archive management can ensure that their metadata extraction approach is more comprehensive, reliable, and efficient. As a result, successful management of digital archives not only enables organizations to effectively handle their archives but also ensures the availability of reliable and consistent data, making tasks more convenient for academics, researchers, policymakers, industry professionals, educators, students, historians, cultural institutions, librarians, archivists, data scientists, journalists, and the general public. Figure 3 presents the performance evaluation of using Imagga, Google cloud vision API, and the combined tools on different types of images, including image without text in Figure 3(a), image with text in Figure 3(b), and all images in Figure 3(c).



(a) (b)

(c)

Figure 3. Efficiency of using Imagga, Google cloud vision API and combined tools on (a) image without text, (b) image with text, and (c) all images

According to the results shown in Figure 3, the study found variations in the accuracy and completeness of metadata fields between Imagga and Google vision API. Imagga performed slightly better than Google cloud vision API in terms of cosine similarity, precision, recall, and F1 score for images with text, while the opposite was observed for images without text. Nonetheless, the integration of both tools outperformed each tool individually in all image types based on the highest metric scores. For all 1,000 images, the combined tool achieved the highest cosine similarity, word overlap similarity, precision, recall, and F1 score of 92.40%, 90.82%, 92.44%, 90.84%, and 91.63%, respectively.

The findings of our study provide insight on the potential and limitations of employing image recognition techniques for metadata extraction in digital archive images. Firstly, our evaluation of the metadata output from Imagga and Google cloud vision API revealed their strengths in different aspects. Imagga exhibited high confidence in correcting metadata fields, indicating its reliability in generating accurate metadata outputs. This can be attributed to its robust image recognition algorithms and advanced machine learning models that can accurately identify and correct metadata fields based on context and semantic meaning. This makes Imagga a suitable choice for digital archive collections where metadata accuracy is crucial, such as historical archives or cultural heritage collections. On the other hand, Google cloud vision API demonstrated proficiency in accurately recognizing and extracting text from images. This can be attributed to its advanced optical character recognition (OCR) capabilities that can accurately detect and extract text from images with different fonts, sizes, and orientations. This makes Google cloud vision API a flexible option for digital archive collections where text-based metadata, such as keywords, subjects, or descriptions, are of primary importance. The strengths and weaknesses of each image recognition tool can inform decision-making in selecting an appropriate tool for metadata extraction, based on the specific requirements of the digital archive collection. For instance, if metadata accuracy is a top priority, Imagga may be a suitable choice. On the other hand, if text-based metadata extraction is of primary importance, Google cloud vision API may be a viable option. Furthermore, our results showed that the combined use of Imagga and Google cloud vision API resulted in higher accuracy and robustness in metadata extraction compared to either tool used alone, suggesting that utilizing the strengths of both tools can lead to improved metadata output in digital archive images. This information can aid in optimizing the metadata extraction process in digital archives, enhancing the overall efficiency and effectiveness of archival workflows. In accordance with the document developed by Wilkinson *et al.* [38] on community-driven governance of findable, accessible, interoperable, and reusable (FAIRness) assessment, it is essential to underscore the significance of FAIRness in optimizing the usability and value of digital archives. Our study findings align with this emphasis, highlighting the suitability of Imagga for collections that prioritize metadata accuracy, and the proficiency of Google cloud vision API in handling text-based metadata. By emphasizing FAIRness, organizations can effectively enhance the discoverability, accessibility, interoperability, and reusability of their digital archives, thereby addressing the core objectives of the FAIR metrics and data quality initiative advocated.

It is important to note that image recognition tools are constantly evolving, with advancements in machine learning and artificial intelligence [1], [5], [6], [8], [11], [30], [39]. As such, it is crucial to regularly evaluate and compare the performance of different tools in the context of specific digital archive collections. Further research and experimentation with different image recognition tools, as well as customization and fine-tuning of their parameters, can be explored to enhance the accuracy and efficiency of metadata extraction in digital archives. Despite the robustness of our methodology, there are some limitations to this study. Firstly, the evaluation of the image recognition tools was limited to the specific dataset of digital archive images from the Internet Archive collection on Flickr. The findings may not be generalizable to other types of image collections or domains. Secondly, the evaluation was based on the ground truth dataset obtained from text before image and text after image sections on Flickr, which may not be exhaustive or fully comprehensive. Lastly, the evaluation metrics used in this study may not capture all aspects of the performance of the image recognition tools, and other evaluation methods could be considered in future research. Overall, the study suggests that image recognition tools have the potential to streamline metadata creation and enhance the accessibility and discoverability of digital archive images. However, the effectiveness of these tools may depend on several factors, including the size and complexity of the image collection, the quality of the images, and the specific tool used. Further research is needed to evaluate the effectiveness of image recognition tools in different contexts and to develop best practices for their use in digital archives.

4. CONCLUSION

In this study, the application of image recognition tools, specifically Imagga and Google cloud vision API, for metadata extraction from digital archive images was explored. The utilization of these APIs allowed for the simplifying of the metadata extraction process by automating the identification and retrieval of relevant metadata from a large dataset of digital archive images. The findings of the study revealed that the combination of metadata generated by Imagga and Google cloud vision API resulted in an increased level of efficiency in metadata extraction, as the strengths of both tools were able to complement each other. Imagga exhibited high confidence in metadata correction, while Google cloud vision API accurately recognized and extracted text from images, ultimately leading to the improvement of the overall accuracy and completeness of the extracted metadata. The analysis carried out in this study can aid in informed decision-making by considering the specific requirements of the digital archive collection and the desired level of accuracy and efficiency in metadata extraction. The study provides valuable insights for researchers, practitioners, and developers working in the field of image recognition and metadata extraction, and further research can be conducted to explore other combinations of image recognition tools and their performance in different contexts. The results of this study contribute to the understanding of the capabilities of image recognition tools in metadata creation, and their potential applications in domains such as image retrieval, content management, and digital asset management. However, despite the high level of automation provided by these tools, a human final evaluation is still required.

ACKNOWLEDGEMENTS

We would like to thank Burapha University for all supports. This research was supported by Faculty of Humanities and Social Sciences, Burapha University (Grant No. HUSO001/2565).

REFERENCES

[1] G. Colavizza, T. Blanke, C. Jeurgens, and J. Noordegraaf, "Archives and AI: an overview of current debates and future perspectives," *J. Comput. Cult. Herit.*, vol. 15, no. 1, pp. 1–15, Feb. 2022, doi: 10.1145/3479010.
[2] M. Bell, "From tree to network: reordering an archival catalogue," *Rec. Manag. J.*, vol. 30, no. 3, pp. 379–394, Jun. 2020, doi: 10.1108/RMJ-09-2019-0051.
[3] G. M. Binmakhashen and S. A. Mahmoud, "Document layout analysis," *ACM Comput. Surv.*, vol. 52, no. 6, pp. 1–36, Nov. 2020, doi: 10.1145/3355610.
[4] Q. Cheng, Q. Zhang, P. Fu, C. Tu, and S. Li, "A survey and analysis on automatic image annotation," *Pattern Recognit.*, vol. 79, pp. 242–259, Jul. 2018, doi: 10.1016/j.patcog.2018.02.017.
[5] J. Fu and Y. Rui, "Advances in deep learning approaches for image tagging," *APSIPA Trans. Signal Inf. Process.*, vol. 6, p. E11, 2017, doi: 10.1017/ATSIP.2017.12.
[6] J. P. Landwehr, N. Kühl, J. Walk, and M. Gnädig, "Design knowledge for deep-learning-enabled image-based decision support systems," *Bus. Inf. Syst. Eng.*, vol. 64, pp. 707–728, Dec. 2022, doi: 10.1007/s12599-022-00745-z.
[7] N. Chaudhuri and I. Bose, "Exploring the role of deep neural networks for post-disaster decision support," *Decis. Support Syst.*, vol. 130, p. 113234, Mar. 2020, doi: 10.1016/j.dss.2019.113234.
[8] W. Su, L. Li, F. Liu, M. He, and X. Liang, "AI on the edge: a comprehensive review," *Artif. Intell. Rev.*, vol. 55, no. 8, pp. 6125–6183, Dec. 2022, doi: 10.1007/s10462-022-10141-4.
[9] R. Miller, "Millions of historic images posted to Flickr," *Internet Archive Blogs*, 2014. https://blog.archive.org/2014/08/29/millions-of-historic-images-posted-to-flickr/ (accessed Aug. 29, 1BC).
[10] Flickr, "Internet archive book images." https://www.flickr.com/photos/internetarchivebookimages
[11] A. Rees, *Image recognition as a tool in cataloguing born-digital photography*. 2020. [Online]. Available: https://eprints.whiterose.ac.uk/181398/
[12] M. Kuźma and A. Mościcka, "Evaluation of metadata describing topographic maps in a national library," *Herit. Sci.*, vol. 8, p. 113, Dec. 2020, doi: 10.1186/s40494-020-00455-3.
[13] R. A. McDougal, I. Dalal, T. M. Morse, and G. M. Shepherd, "Automated metadata suggestion during repository submission," *Neuroinformatics*, vol. 17, pp. 361–371, Jul. 2019, doi: 10.1007/s12021-018-9403-z.
[14] S. Pal, P. K. D. Pramanik, and P. Choudhury, "Enhanced metadata modelling and extraction methods to acquire contextual pedagogical information from e-learning contents for personalised learning systems," *Multimed. Tools Appl.*, vol. 80, pp. 25309–25366, Jul. 2021, doi: 10.1007/s11042-020-10380-z.
[15] W. Z. Alma'aitah, A. Z. Talib, and M. A. Osman, "Opportunities and challenges in enhancing access to metadata of cultural heritage collections: a survey," *Artif. Intell. Rev.*, vol. 53, pp. 3621–3646, Jun. 2020, doi: 10.1007/s10462-019-09773-w.
[16] V. Gautam, "Qualitative model to enhance quality of metadata for data warehouse," *Int. J. Inf. Technol.*, vol. 12, pp. 1025–1036, Dec. 2020, doi: 10.1007/s41870-018-0222-0.
[17] A. Rajesh, Y. Chang, M. S. Abedalthagafi, A. Wong-Beringer, M. I. Love, and S. Mangul, "Improving the completeness of public metadata accompanying omics studies," *Genome Biol.*, vol. 22, p. 106, Dec. 2021, doi: 10.1186/s13059-021-02332-z.
[18] M. S. Sayeed, I. Bin Yusof, M. F. A. bin Abdullah, M. A. Bari, and P. P. Min, "A comprehensive survey on deep-learning based gait recognition for humans in the COVID-19 pandemic," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 882–902, May 2023, doi: 10.11591/ijeecs.v30.i2.pp882-902.
[19] W. Torres, M. G. J. van den Brand, and A. Serebrenik, "Xamã : Optical character recognition for multi-domain model management," *Innov. Syst. Softw. Eng.*, Apr. 2022, doi: 10.1007/s11334-022-00453-7.
[20] Y. Zeng and J. Zhang, "A machine learning model for detecting invasive ductal carcinoma with Google Cloud AutoML Vision," *Comput. Biol. Med.*, vol. 122, p. 103861, Jul. 2020, doi: 10.1016/j.compbiomed.2020.103861.
[21] Z. R. Samani, S. C. Guntuku, M. E. Moghaddam, D. Preoţiuc-Pietro, and L. H. Ungar, "Cross-platform and cross-interaction study of user personality based on images on Twitter and Flickr," *PLoS One*, vol. 13, no. 7, p. e0198660, Jul. 2018, doi: 10.1371/journal.pone.0198660.
[22] M. H. Mutar, E. H. Ahmed, M. R. M. Alsemawi, H. O. Hanoosh, and A. H. Abbas, "Ear recognition system using random forest and histograms of oriented gradients techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 1, pp. 181–188, Jul. 2022, doi: 10.11591/ijeecs.v27.i1.pp181-188.
[23] R. B. Devareddi and A. Srikrishna, "Query-based image tagging model using ensemble learning with enhanced artificial bee colony optimization," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 870–881, May 2023, doi: 10.11591/ijeecs.v30.i2.pp870-881.
[24] F. H. Alqahtani and F. A. Alsulaiman, "Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study," *Comput. Secur.*, vol. 88, p. 101635, Jan. 2020, doi: 10.1016/j.cose.2019.101635.

[25] S.-H. Chen and Y.-H. Chen, "A content-based image retrieval method based on the Google cloud vision API and WordNet," in *Intelligent Information and Database Systems*, 2017, pp. 651–662. doi: 10.1007/978-3-319-54472-4_61.
[26] K. Thammarak, P. Kongkla, Y. Sirisathitkul, and S. Intakosum, "Comparative analysis of tesseract and google cloud vision for Thai vehicle registration certificate," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 2, pp. 1849–1858, Apr. 2022, doi: 10.11591/ijece.v12i2.pp1849-1858.
[27] S. A. A. Shah, A. A. Wahab, N. Ageelani, and N. Najeeb, "Street-crimes modelled arms recognition technique employing deep learning and quantum deep learning (SMARTED)," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 1, pp. 528–544, Apr. 2023, doi: 10.11591/ijeecs.v30.i1.pp528-544.
[28] S. Bekhet, A. M. Alghamdi, and I. F. Taj-Eddin, "Gender recognition from unconstrained selfie images: a convolutional neural network approach," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 2, pp. 2066–2078, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2066-2078.
[29] S. Adebayo, H. O. Aworinde, A. O. Akinwunmi, A. Ayandiji, and A. O. Monsir, "Convolutional neural network-based crop disease detection model using transfer learning approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, pp. 365–374, Jan. 2022, doi: 10.11591/ijeecs.v29.i1.pp365-374.
[30] A. Kubany, S. Ben Ishay, R. S. Ohayon, A. Shmilovici, L. Rokach, and T. Doitshman, "Comparison of state-of-the-art deep learning APIs for image multi-label classification using semantic metrics," *Expert Syst. Appl.*, vol. 161, p. 113656, Dec. 2020, doi: 10.1016/j.eswa.2020.113656.
[31] Imagga, "Image recognition applications with Imagga's API." https://imagga.com/
[32] Google Cloud, "Vision AI." https://cloud.google.com/vision
[33] J. Ma, Y. Wang, X. Wang, J. Wang, and J. Zhao, "A three-tiered semi supervised MTL mechanism and its application in dating apps," *Neural Comput. Appl.*, Dec. 2022, doi: 10.1007/s00521-022-08081-9.
[34] X. Ochoa and E. Duval, "Automatic evaluation of metadata quality in digital repositories," *Int. J. Digit. Libr.*, vol. 10, pp. 67–91, Aug. 2009, doi: 10.1007/s00799-009-0054-4.
[35] C. Tao, J. Gao, and T. Wang, "Testing and quality validation for ai software–perspectives, issues, and practices," *IEEE Access*, vol. 7, pp. 120164–120175, 2019, doi: 10.1109/ACCESS.2019.2937107.
[36] G. Mustafa, M. Usman, M. T. Afzal, A. Shahid, and A. Koubaa, "A comprehensive evaluation of metadata-based features to classify research paper's topics," *IEEE Access*, vol. 9, pp. 133500–133509, 2021, doi: 10.1109/ACCESS.2021.3115148.
[37] K. Kyriakou, P. Barlas, S. Kleanthous, and J. Otterbacher, "Fairness in proprietary image tagging algorithms: a cross-platform audit on people images," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 13, pp. 313–322, Jul. 2019, doi: 10.1609/icwsm.v13i01.3232.
[38] M. D. Wilkinson *et al.*, "Community-driven governance of FAIRness assessment: an open issue, an open discussion," *Open Res. Eur.*, vol. 2, p. 146, Sep. 2023, doi: 10.12688/openreseurope.15364.2.
[39] H. Elmannai and A. D. AlGarni, "Classification using semantic feature and machine learning: Land-use case application," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 19, no. 4, p. 1242, Aug. 2021, doi: 10.12928/telkomnika.v19i4.18359.

BIOGRAPHIES OF AUTHORS

Akara Thammastitkul received her B.Sc. and M.Sc. degrees in computer science from Burapha University and Mahidol University, Thailand, in 1998 and 2002 respectively. She holds her Ph.D. in Information Technology from Sirindhorn International Institute of Technology, Thammasat University, Thailand in 2009. She is now a lecturer at Department of Information Studies, Faculty of Humanities and Social Sciences, Burapha University, Thailand. Her research interests are image processing and image analysis. She can be contacted at email: akara@buu.ac.th.

Jitsanga Petsuwan earned his B.Sc. in Information Technology from Burapha University and subsequently gained valuable experience as a senior computer engineer in a startup company. Currently, he serves as an independent researcher and computer technical officer at Chaiyaphum Hospital. His expertise lies in the management and optimization of computer systems. He can be contacted at email: jitsanga@gmail.com.