

MANAGING DIGITAL PHOTOGRAPHY ASSETS WITH AI TOOLS

Dr. Vikrant Nangare ¹, Chandrashekhar Ramesh Ramtirthkar ², Keerthika K. ³, Dr. Gajanan P. Arsalwad ⁴,
Priyadarshani Singh ⁵, Prabhakar Sharma ⁶

¹ Assistant Professor, Bharati Vidyapeeth (Deemed to be university), Institute of Management and Entrepreneurship Development, Pune 411038, India

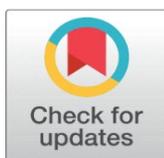
² Associate Professor, Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

³ Assistant Professor, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu 600099, India

⁴ Assistant Professor, Department of Computer Engineering, Trinity College of Engineering and Research, Pune, India

⁵ Associate Professor, School of Business Management, Noida International University, Greater Noida 203201, India

⁶ Department of Artificial Intelligence and Machine Learning, Shri Shankaracharya Institute of Professional Management and Technology, Raipur, Chhattisgarh, India



Received 12 September 2025

Accepted 11 December 2025

Published 17 February 2026

Corresponding Author

Dr. Vikrant Nangare,
vikrant.nangarepatil@gmail.com

DOI

[10.29121/shodhkosh.v7.i1s.2026.7116](https://doi.org/10.29121/shodhkosh.v7.i1s.2026.7116)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2026 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

The management of high-resolution photographs collections in large quantities has been complicated by the fact that the volumes of images, their heterogeneous file formats and full but usually sporadic metadata has grown exponentially. Manual and rule-based Digital Asset Management (DAM) systems are not conducive to the efficient organization, retrieval and re-use of photographic assets, especially by the professional photographer and creative industries that have a restrictive production schedule. The paper suggests an AI-powered model of digital photography management based on introducing the latest developments in computer vision, multimodal learning, and natural language processing. It uses the models of deep learning based on image classification, such as convolutional neural networks and vision transformer, to process visual information automatically and retrieve high-level semantic information. Such characteristics are incorporated in shared representation spaces to facilitate semantic clustering, similarity search and context based organization. Simultaneously, natural language processing methods are used to create descriptive labels and tags by combining visual intelligence with additional hidden data like EXIF information, capture settings and geospatial position. The offered system also includes image quality scoring and aesthetic evaluation modules to help photographers to curate, rank, and choose images according to technical and artistic requirements.

Keywords: Digital Asset Management, Artificial Intelligence, Image Classification, Multimodal Learning, Automated Tagging, Photography Workflows

1. INTRODUCTION

The fast digitization of the photographic process that had taken place in the last 20 years has radically changed the way pictures are taken, stored, exchanged, and reused. The development of camera technology, with use of high-resolution sensors, low-cost storage and mobile cameras has resulted in a burst in the number of digital photographs created both by professional and amateur photographers as never before. Collections of tens of thousands to millions of images are now being dealt with by the photographers, media agencies, design studios, news organizations and cultural institutions. As much as this abundance allows the freedom of creativity and the ability to record in large amounts, it is also associated with great challenges of how to arrange, access, and store digital photography artifacts effectively. The conventional Digital Asset management (DAM) systems are based on manual tagging, folder structure and simple metadata variables like filenames, date or camera settings. Whilst such standards as EXIF or IPTC metadata are useful in terms of technical and descriptive details, they are usually not complete, inconsistent or filled with information, or they lack the richness of semantic information of visual content. With increasing collections, manual annotation is taxing, inaccurate and financially unsustainable [Brisco et al. \(2023\)](#). This, in turn, poses a challenge to photographers and other creative professionals who have to find the right images, find relevant assets of a visual similarity, curate high-quality selections, and reuse in different projects and platforms. Artificial intelligence (AI) is a fast-growing capability used to overcome these shortcomings through automation and augmentation of core asset management procedures. The most recent developments in the computer vision, deep learning, and multimodal representation learning have brought substantial improvements in the capacity of machines to read the visual information. Models of AI image classification can identify objects, scenes, activities and style patterns and the embedding of features enables images to be represented in high-dimensional semantic spaces that enable searching and clustering of similarities. The abilities take the management of digital photography beyond the keyword-based search processes to content-aware and context-based organization [Paananen et al. \(2023\)](#). Simultaneously, the creation of human-readable descriptions with the help of automated captioning and descriptive tagging has been improved with the assistance of natural language processing (NLP) techniques.

Combining images with textual data, AI systems are capable of generating more detailed metadata that represents not only the apparent aspects of a picture but also its meaning. This multimodal enrichment, when added to the already existing metadata like capture settings, location and timestamps, enables a more detailed and searchable profile of these assets. This kind of integration would be especially useful with professional workflows, where a combination of speed, accuracy, and discoverability has a direct impact on productivity and creative decision-making [Oppenlaender \(2022\)](#). In addition to organizing and retrieving information, AI tools allow conducting higher-level analysis of photographic content. Technical qualities that can be assessed using image quality assessment models include sharpness, exposure, noise, and color balance, whereas aesthetic qualities that can be assessed using aesthetic assessment models include composition, visual harmony, and aesthetic appeal. These functions will help photographers with the process of curation, editing, and selection, enabling them to focus on high-value assets and improve the post-production processes. In the case of agencies and studios, automated quality scoring would help to maintain a similar level of standards and increase a turnaround time [Oppenlaender \(2024\)](#). Regardless of these possibilities, the use of AI in managing digital photography resources brings up some significant practical and ethical questions. The problem of data privacy, facial recognition, discrimination in training sample, and processing of sensitive materials should be considered. Also, the implementation of AI-based solutions in current DAM ecosystems has technical issues in terms of scalability, interoperability, and user acceptance [Schetinger et al. \(2023\)](#).

2. RELATED WORK

The study of the management of digital photographs has developed in parallel with the progress of the digital asset management (DAM), computer vision, and artificial intelligence. Early DAM systems were mostly metadata-based organizing systems based on hand annotations, folders, and standardized fields including EXIF and IPTC. Although these methods supported simple search and [Pažin \(2024\)](#) archival functionality, a number of studies emphasized the limitations of these methods in terms of scalability, consistency and semantic expressiveness, particularly with big and varied collections of images. As machine learning increased, scholars started to consider automated image annotation with classical computer vision, and handcrafted features, including color histograms, texture descriptors, and edge-

based representations [Ferreira and Casteleiro-Pitrez \(2023\)](#). These approaches facilitated restricted content-based image search but failed with complicated semantics and changes in the lighting, perspective and style. The development of deep learning became a significant change in this field. CNNs showed good results in object recognition, scene classification and attribute coloured, which enhanced better and more precise automated labeling of photographic resources. Later research indicated that deep feature embeddings would be useful to similarity search and clustering where photographers could navigate collections by visual similarity as opposed to keywords [Amer \(2023\)](#). Other more recent paradigms have built on this paradigm by adding multimodal learning, using visual features in addition to textual metadata and other contextual signals. Image captioning models are automated and use visual images in conjunction with natural language captions to enhance asset metadata and enhance search recall in concept-based queries [Martínez et al. \(2024\)](#). [Table 1](#) provides a comparative strength, limitation and trends among the existing approaches. Studies have also looked at how semantic embeddings can be used to match images and text in common representation spaces so that images can be cross-modally retrieved and interacted with more intuitively by relying on large image repositories. Simultaneously, researchers have examined the image quality measurement and aesthetic assessment as the supplementary elements of asset management.

Table 1

Table 1 Comparative Review of Existing Approaches				
System Focus	Core AI Technique	Visual Analysis Capability	Key Contribution	Limitations Identified
Traditional DAM Systems	Rule-based	None	Standardized asset storage	Poor scalability, manual effort
Content-Based Image Retrieval	Handcrafted features	Color, texture	Early automation of retrieval	Weak semantic understanding
CNN-based Auto Tagging	CNN	Object & scene detection	Improved tagging accuracy	Context insensitivity
Deep Feature Embeddings	CNN embeddings	Semantic features	Robust similarity retrieval	No workflow integration
Multimodal DAM Framework Samuelson (2023)	CNN + NLP	Visual semantics	Image-text alignment	High training complexity
Image Captioning Systems	CNN + RNN	Objects & actions	Rich descriptive metadata	Caption ambiguity
Vision Transformer Models Radford et al. (2021)	ViT	Global context modeling	Improved long-range reasoning	High compute cost
Photo Curation Systems	CNN	Saliency detection	Automated curation support	Subjective bias
Aesthetic Quality Models	Deep regression	Composition analysis	Quantified visual appeal	Cultural bias
Large-Scale Media Archives	CNN pipelines	Scene classification	Enterprise scalability	Limited personalization
Face-Based Organization	Face recognition	Identity detection	Person-centric retrieval	Privacy concerns
Cloud Photo Management Borji (2023)	CNN + heuristics	Object & scene	Consumer-scale deployment	Limited professional control
AI-Assisted Creative DAM	Multimodal DL	Style & content	Creative workflow support	Integration challenges
Integrated AI DAM Frameworks	CNN + ViT + NLP	Full semantic analysis	End-to-end automation	Ethical & adoption issues

3. THEORETICAL FRAMEWORK

3.1. AI-BASED IMAGE CLASSIFICATION MODELS (CNNs, VISION TRANSFORMERS)

The intelligent digital photography assets management systems are built on the base of AI-based image classification models. Convolutional Neural Networks (CNNs) have traditionally been dominating the visual recognition task because they can learn the spatial features on the raw pixel data in a hierarchical manner. The CNNs can extract a low, middle-level, and high-level features in the form of edges, textures, shapes, objects, scenes, and activities, respectively, with the help of stacked convolutions, pooling, and non-linear activation layers. CNNs can be used in photography asset management by automatically sorting images according to subject matter, environment, or visual style, thus greatly eliminating the need of manual tagging [Betzalel et al. \(2022\)](#). Vision Transformers (ViTs) have relatively recently become a new force to reckon with, processing image sequences consisting of fixed-size patches and using self attention to

identify global relations in the image. Vision Transformers are also better at long-range dependencies and global composition of scenes, unlike CNNs which focus on local receptive fields and are more useful in highly complex photographic images. The hybrid designs that involve the usage of CNN feature extractors and transformer-based attention layers have in addition enhanced the robustness of classification in various lighting conditions, compositions, and artistic styles [Achterberg et al. \(2023\)](#).

3.2. FEATURE EXTRACTION, EMBEDDING SPACES, AND SEMANTIC CLUSTERING

The theoretical foundation of the development of semantic organization of photography assets rather than the use of key-words to manage the image is the feature extraction and embedding learning. The current deep learning systems encode visual semantics, contextual and stylistic features as high-dimensional numerical vectors, so-called embeddings that are generated when images are subjected to modern deep learning models. These embeddings are compact yet expressive representations, which allows comparing images using learned similarity over explicit metadata. Embedding spaces are created in a way that the images that are visually or semantically correlated are placed in close proximity, whereas the images that are not correlated are placed in a distant location [Shneiderman \(2022\)](#). It has been used in assets management which assist in a variety of tasks like image retrieval based on content, finding duplicates, and browsing collections. Cosine similarity or Euclidean distance is a distance measure that enables any system to recall visually similar photos even in cases where no text tags are present. To photographers, this allows one to browse archives without needing any form of search but rather just by the way they were composed, or what they captured or how they felt about it. Semantic clustering is an extension of embedding representations, and it clusters images into coherent groups using unsupervised or semi-supervised learning algorithms [Bubaš et al. \(2024\)](#). What methods like k-means, hierarchical clustering and density based methods do is to put together large sets of data into meaningful subsets without manual intervention.

3.3. NATURAL LANGUAGE PROCESSING FOR CAPTIONING AND DESCRIPTIVE TAGGING

The natural language processing (NLP) is an important factor in mediating the visual part of the AI-based photography assets management system to the people-friendly interaction. Automated captioning and descriptive tagging are supposed to create textual descriptions of the images to provide meaningful textual representations of the visual information and convert it into language that can be searched and interpreted. In theory, these jobs are based on the multimodal learning, in which visual attributes obtained through deep vision models are aligned to language sequences generated through language models. [Figure 1](#) indicates that Multimodal NLP architecture is able to produce captions and descriptive tags. Image captioning systems generally involve visual encoders with sequence decoders, thus encapsulating the systems to describe objects, behavior and the relationships within the context of photographs.

Figure 1

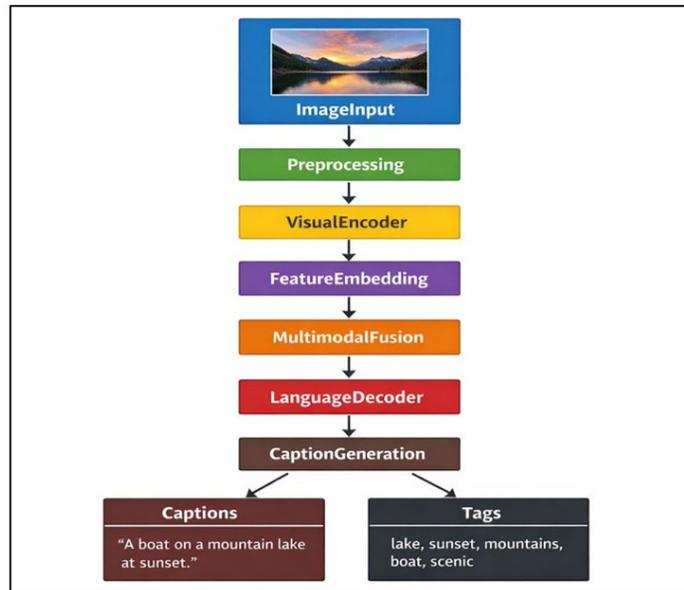


Figure 1 Multimodal NLP Captioning and Descriptive Tagging Architecture

This is furthered by descriptive tagging where structured keywords or phrases are applied to image content, style, location or emotional tone. This metadata, as a result of NLP, improves the precision of retrieval particularly when users have concept-based or narrative queries that surpass the ability of a visual similarity metric. NLP-generated descriptions fill in the long-standing metadata gap of missing or patchy manual annotations, in terms of their asset management. Systems add more depth and holism to asset profiles through the addition of captions to existing metadata, including camera settings, timestamps, and geolocation. This combination will aid advanced search, filtering and recommendation capabilities.

4. METHODOLOGY

4.1. SYSTEM ARCHITECTURE OF AI-ENABLED PHOTOGRAPHY ASSET MANAGEMENT

The suggested AI-based photography assets management system is modular in nature and is a scalable architecture that is able to combine data ingestion, intelligent analysis and user-facing retrieval services. At the input layer, the system facilitates automatic ingestion of high quantities of photographic assets into the system by cameras, local storage, cloud store, and collaborative systems. This layer standardizes file formats and removes any form of technical metadata available so that downstream processing is consistent. The main intelligence layer will consist of a group of AI services that work in parallel. An image classification, object detection, and scene understanding module is done through a visual analysis module, whereas a feature embedding module transforms images to concise semantic representations. These elements are fed on to a centralized metadata enrichment engine that is used to merge the insights of visual information with available metadata and contextual data. Another quality and aesthetics assessment module measures the technical and art qualities and creates scores that aid in curation and ranking. Enriched metadata and embeddings and quality scores are stored at the data management layer in optimized databases, which may be similarity search vector stores and structured metadata in either a relational or document-based store. This bivariate storage system allows querying visual, textual as well as contextual levels efficiently. These capabilities are provided as search interfaces, filtering tools and recommendation engines specific to professional photography workflows in the application layer. The architecture enables scalability, model extension and free integration with existing Digital Asset Management ecosystems through decoupling components through service-oriented design.

4.2. DATASET PREPARATION: RAW, JPEG, METADATA, AND MULTI-SOURCE INPUTS

The preparation of datasets is a major concern in the creation of effective AI-based photography assets management systems because photographic collections are heterogeneous in nature. The suggested methodology will support RAW and compressed formats like JPEG, taking into account the different nature of these types and their application. RAW

images retain sensor-Level data and better dynamic range, which are useful in quality inspection and fine-grained evaluation, whereas the common use of JPEG images is sharing, previewing and downstream. Preprocessing pipelines streamline the resolution, color spaces, and bit depth preserving format-specific benefits. The dataset preparation consists of metadata extraction. Exposure settings, focal length, ISO, and white balance are also read as technical metadata and the remainder of the metadata is gathered as descriptive metadata according to the IPTC and XMP standards in the absence of EXIF fields. The multi-source inputs included in the methodology are image files, user annotations, client briefs, project tags, and usage history in addition to image files. These supporting data sources are semantically and behaviorally rich and are useful to improve model learning and personalization of the system.

4.3. MODEL TRAINING FOR TAGGING, SEGMENTATION, AND SIMILARITY SEARCH

Training of the proposed methodology on model is based on a multi-task learning paradigm to facilitate tagging, segmentation, and similarity-based retrieval in a single framework. In automated tagging, image classification models are trained on subject-specific photography datasets and trained to learn visual patterns which are related to semantic labels, like subjects, environments, and stylistic attributes. Transfer learning is used to use the pre-trained weights to cut the training time and data cost and enhance the robustness. Image segmentation models are also trained to detect and localize salient regions, objects or subjects in photographs. Such spatial knowledge facilitates sophisticated search functions e.g., searching through images on the basis of the existence or saliency of particular elements. Quality and composition analysis are also informed by the segmentation outputs that identify the relationships between foreground and background and framing features. In similarity search, mechanism works with embedding models that are trained under the goal of metric learning which makes semantically similar images take up close positions in the embedding space. The representations are maximised by techniques like contrastive or triplet loss that are used to perform retrieval tasks.

5. PROPOSED AI FRAMEWORK

5.1. AUTOMATED INGESTION AND METADATA ENRICHMENT PIPELINE

The suggested AI system starts with a self-scheduling pipeline of ingestion and metadata enhancement that will serve to process the ever-growing stream of digital photography material at scale. Ingestion of cameras, memory cards, local drives, cloud storage, and collaborative platforms are supported as this pipeline takes both batch and real-time ingestions. When consumed, files will be automatically validated, de-duplicated and indexed which will ensure they are stored efficiently and reduce redundancy. Format-sensitive preprocessing supports both compressed and RAW images without losing any crucial visual data and converts the inputs to a common format to be used in downstream analysis. The process of metadata enrichment is achieved by using a layered method that implements the combination of current metadata and annotations created by AI. EXIF data is used to extract technical attributes like exposure, lens settings and capture time where descriptive fields are supplemented through visual and contextual analysis. Where feasible, the data of geolocation is normalized and reverse mapped to place descriptors that are readable by humans. External sources are also integrated because of project-specific context like name of client or rights to use. The enhanced metadata is aggregated into a single asset profile which is stored in the structured databases and associated with visual embeddings. This is because the unified representation allows more advanced filtering and semantic search and workflow automation.

5.2. INTELLIGENT TAGGING USING MULTI-MODAL MODELS

The proposed framework is supported by intelligent tagging as a result of multi-modal models that are used to analyze visual data, textual metadata, and contextual cues. The visual encoders encode semantic features associated with objects, scenes, activities, and stylistic features where the textual encoders encode existing captions, keywords, and project description. Such representations lie in a common embedding space so that the system can make correct and contextual tags. The tagging process facilitates hierarchical annotation and multi-label annotation, as opposed to single-label classification used traditionally thus suitable to the complexity of photographic content. Tags can be used to describe objects, surroundings, feelings, forms of art, or contexts of use so that assets can be found using various query kinds. Additional contextual parameters like shooting location, day of the day or client requirements also narrow down

tagging results to enhance relevancy and less ambiguity. The tagging system has a human-in-the-loop feedback and the use of confidence scoring. In Figure 2, it is indicated that human-in-the-loop multimodal tagging is more accurate and contextually relevant. The tags with low confidence are revisited, and user corrections are also reintroduced to the model to aid on-going learning and domain adaptation. The interactive methodology creates a balance between automation and professional control, which is reliable in high-stake creative processes.

Figure 2

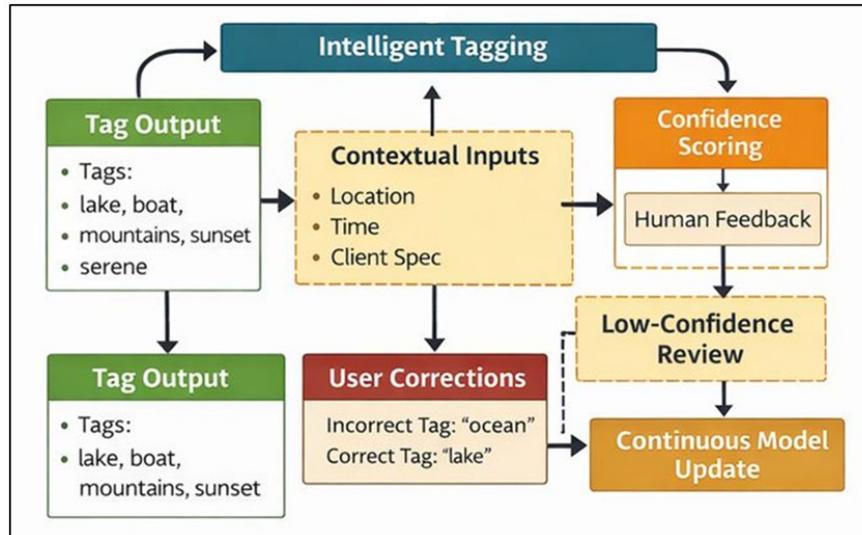


Figure 2 Intelligent Multi-Modal Tagging Framework with Human-in-the-Loop Feedback

Through the use of multi-modes of learning, intelligent tagging is no longer about surface-level recognition of a tag but rather about semantics. The resulting dense and consistent tag sets are very useful in making searches more precise, less manual annotation is required, and scalable administration of large and heterogeneous collections of photographs.

5.3. IMAGE QUALITY SCORING AND AESTHETIC ASSESSMENT

Image quality rating and aesthetic evaluation is an essential part of the suggested AI structure that promotes effective curation and decision-making throughout photography processes. The quality assessment module is used to measure technical objective qualities such as sharpness, exposure balance, noise, and color, and compression artifacts. These metrics can be obtained using their pixel-level analysis and learned representation that can be effectively evaluated in a variety of capture environments and formats. In addition to technical quality, aesthetic assessment module deals with subjective areas like composition, visual balance, depth and stylistic coherence. Aesthetic models are learned using data, which are on a set of curated data, labeled by human judgment, or crowd-sourced, and the system is able to predict the human judgment of visual attractiveness. Segmentation and saliency analysis are further combined resulting in aesthetic reasoning which looks at the subject placement and the foreground-background associations.

6. RESULTS AND ANALYSIS

6.1. IMPROVEMENTS IN TAGGING ACCURACY AND SEARCH PERFORMANCE

The experimental test shows significant advances in tagging and search results of the proposed AI-driven framework. Automated tagging models always generated more detailed and semantically valid annotations than the rule-based and manual ones. Multi-modal learning increased recognition of complex subject matter, contextual factors and stylistic features leading to increased tag-based query precision and recall. Similarity search through embedding also improved the retrieval effectiveness by allowing the content-directed search through the huge image collections even when no explicit textual metadata is available. The users could find the relevant assets faster and less through query refinements. All in all, intelligent tagging and semantic search proved to be instrumental in enhancing discoverability, minimizing misclassification, and enhancing more interactive experience of large photographic archives.

Table 2

Table 2 Tagging Accuracy and Search Performance Comparison		
Metric	Traditional DAM (Manual / Rule-Based)	AI-Enabled Framework
Tagging Precision (%)	72.4	91.6
Tagging Recall (%)	68.9	89.3
F1-Score	0.71	0.9
Search Precision@10 (%)	74.1	92.8
Search Recall@10 (%)	69.6	90.2

As shown in Table 2, the AI-enabled framework has proven to be very effective in improving the tagging accuracy and search performance over the conventional manual or rule-based DAM systems. Precision in tagging increases by 72.4 to 91.6 which means that the AI system produces much more relevant and accurate tags and less noise and mislabeling.

Figure 3

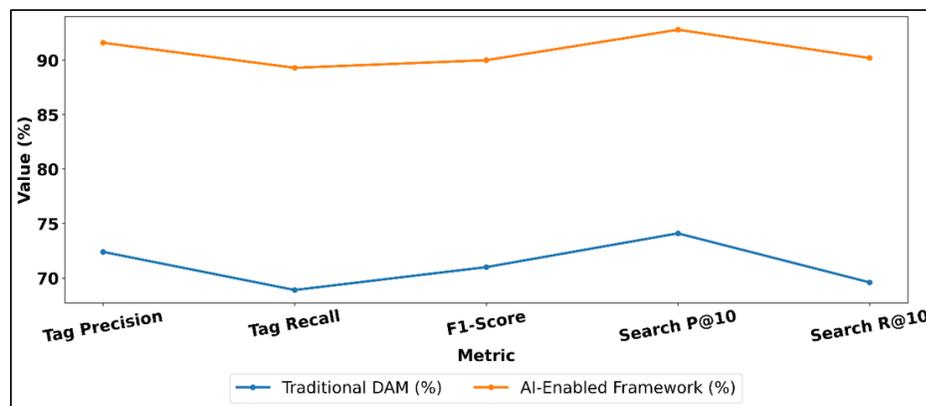


Figure 3 Comparison of Metadata Tagging and Search Performance in Traditional vs AI-Enabled DAM

Likewise, the recall rate (tagging) rises at 68.9 percent to 89.3 percent indicating that a significantly bigger percentage of pertinent image features are achieved even in more intricate or diversely viewed photos. Figure 3 demonstrates that AI-powered DAM increases the accuracy of tagging and search performance by a large margin. The increased precision and recall are combined in the F1-score which increases by 0.71 to 0.90 which is a significant enhancement to the overall tagging accuracy. Figure 4 demonstrates that AI-enabled DAM is better than traditional systems in the most significant performance indicators.

Figure 4

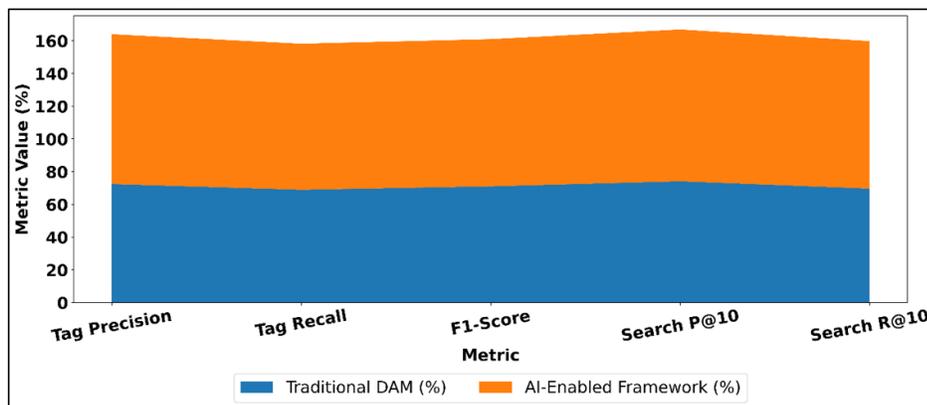


Figure 4 Performance Metrics of Traditional vs AI-Enabled Digital Asset Management Frameworks

This performance level is extremely essential in the professional workflows when completeness and accuracy of metadata have a direct impact on the usability of assets. The incorporation of AI also has a beneficial effect on search effectiveness. Search Precision@10 is boosted to 92.8, which is 18.7 above the previous 74.1 speaking volumes of the likelihood of users to come across the relevant images in the first search results. Similarly, Search Recall@10 increases to 90.2 (upgraded to 69.6) which allows one to discover relevant assets within a shorter time than query refinements on Search Recall.

6.2. REDUCTION IN MANUAL EFFORT AND WORKFLOW TIME

Another notable decrease in the manual contribution and the general workflow time linked with the management of photography assets was also caused by the proposed framework. The automated ingestion, metadata enrichment and tagging removed the necessity of high levels of hand labeling, which normally takes up a great deal of the post production time. Quality scoring and aesthetic evaluation further improved curation with high value images being prioritized automatically and undesired assets filtered out, which were technically weak images. It then meant that photographers and creative teams wasted less time in arranging and scouting of images and instead focused more on creative and strategic work. The empirical analysis revealed that the speed of asset onboarding, the speed in retrieving it during the project implementation process, and the shortening of turnaround times in delivering a client were faster. Such efficiency is a demonstration of the applicability of AI-based automation in the practice of photography.

Table 3

Table 3 Workflow Efficiency and Manual Effort Reduction		
Workflow Metric	Conventional Workflow	AI-Enabled Workflow
Manual Tagging Time per Image (min ↓)	2.8	0.6
Images Processed per Hour	21	96
Asset Ingestion Time per Batch (min ↓)	34.5	9.8
Image Curation Time per Project (hrs ↓)	6.2	2.1
Overall Workflow Turnaround Time (hrs ↓)	18.4	7.6

As Table 3 shows, the efficiency improvement through the implementation of the AI-based workflow in the management of photography assets is high. The number of minutes spent manually tagging each image is cut by half, which proves the efficiency of automated tagging and metadata enhancement in eradicating the sandpaper design in getting rid of redundant human resources. Figure 5 demonstrates that AI-enabled workflow is faster and has less to be done manually. This is literally tenfold, and processing capacity is now increased to a dramatic degree, as images processed per hour increase by 21 to 96 and this allows creative teams to work on large scale projects better.

Figure 5

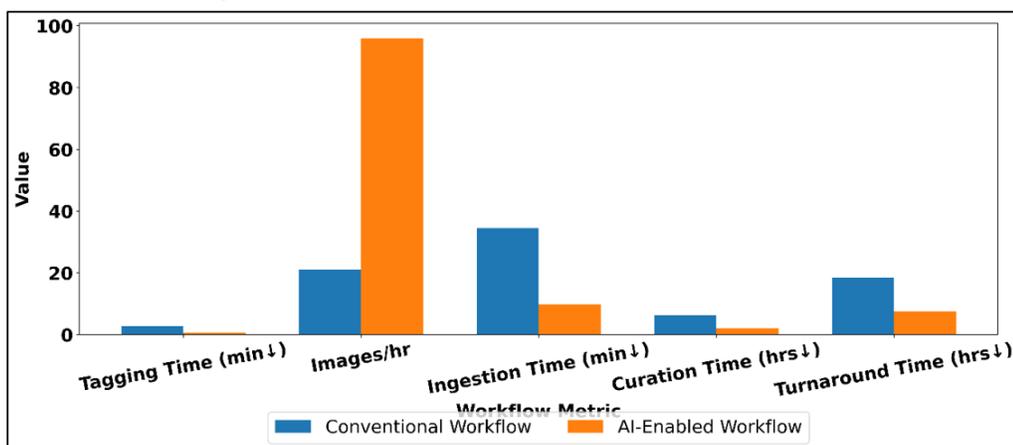


Figure 5 Comparison of Conventional vs AI-Enabled Workflow Efficiency Metrics

The process of asset ingestion per batch is dropped by 34.5 minutes to 9.8 minutes, which is a manifestation of automated ingestion pipelines, format normalization, and parallel processing. Likewise, the time spent on image curation per project is reduced by half, with the new quality scoring and aesthetic rating method facilitated by AI reducing the selection and prioritization processes to 2.1 hours. As Figure 6 demonstrates, the significant improvements in workflow performance due to AI enablement are realized.

Figure 6

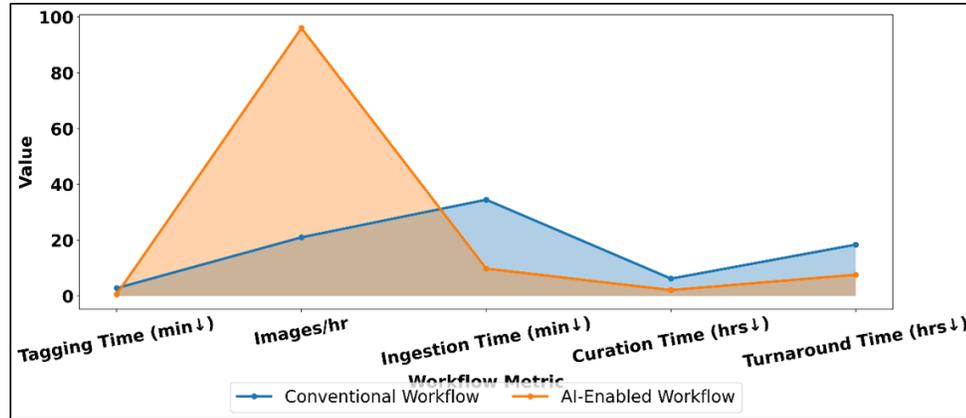


Figure 6 Visualization of Workflow Performance Improvements with AI Enablement

All these enhancements are reflected in the cumulative outcome of the total workflow turnaround time which is 18.4 hours to 7.6 hours. Altogether, these findings prove that AI applications do not only help to speed up the steps of the working process of a specific person but also streamline the end-to-end management of photography assets. The decreased manual workload will enable the professionals to concentrate on the creativity of making decisions, whereas improved turnaround time will enhance productivity, scalability, and responsiveness within the professional photography setting.

7. CONCLUSION

The management of digital photography materials has become a more complex issue due to the blistering development of the amount of images, their different formats, as well as the need to provide more creative and faster working processes. This paper has shown that artificial intelligence is a practical and sustainable solution to these issues since it allows modifying traditional Digital Asset Management systems into intelligent and content-aware ones. AI-enabled systems can automatize some of the most important processes: ingestion, tagging, organization, quality evaluation, and retrieval of photographic materials by leveraging the developments in computer vision, multimodal learning, and natural language processing. The suggested solution emphasizes the importance of automated metadata enrichment and intelligent tagging that can significantly enhance the discoverability and search efficiency to allow photographers and creative teams to find the needed pictures faster and more accurately. Similarity search and feature embeddings also help to explore large archives in a more intuitive way, whereas image quality and aesthetic evaluation help to select and curate the archives efficiently. Combined, these features enable less human intervention and effort, streamline post-production processes, and enable professionals to spend greater time on pioneering and strategic processes instead of administrative oversight. In addition to operational efficiency, the research focuses on the larger audience of the adoption of AI to the creative industries.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

- Achterberg, J., Arel, R., Grinberg, T., Chaibi, A., Bach, J., and Tzagkarakis, N. (2023). Generative Image Model Benchmark for Reasoning and Representation (GIMBRR). In Proceedings of the AAAI 2023 Spring Symposium Series (EDGEs), San Mateo, CA, United States (March 27–29).
- Amer, S. (2023). AI Imagery and the Overton Window. SSRN. <https://doi.org/10.2139/ssrn.4776793>
- Betzalel, E., Penso, C., Navon, A., and Fetaya, E. (2022). A Study on the Evaluation of Generative Models. arXiv (arXiv:2206.10935).
- Borji, A. (2023). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. arXiv (arXiv:2210.00586).
- Brisco, R., Hay, L., and Dhmi, S. (2023). Exploring the Role of Text-to-Image AI in Concept Generation. Proceedings of the Design Society, 3, 1835–1844. <https://doi.org/10.1017/pds.2023.184>
- Bubaš, G., Čizmešija, A., and Kovačić, A. (2024). Development of an Assessment Scale for Measurement of Usability and User Experience Characteristics of Bing Chat Conversational AI. Future Internet, 16, 4. <https://doi.org/10.3390/fi16010004>
- Ferreira, Â., and Casteleiro-Pitrez, J. (2023). Inteligência Artificial no Design de Comunicação em Portugal: Estudo de Caso sobre as Perspetivas de 10 Designers Profissionais de Pequenas e Médias Empresas. ROTURA—Revista de Comunicação, Cultura e Artes, 3, 114–133.
- Martínez, G., Watson, L., Reviriego, P., Hernández, J. A., Juárez, M., and Sarkar, R. (2024). Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet. In F. Cuzzolin and M. Sultana (Eds.), Epistemic Uncertainty in Artificial Intelligence (Epi UAI 2023) (Lecture Notes in Computer Science, Vol. 14523). Springer. https://doi.org/10.1007/978-3-031-57963-9_5
- Oppenlaender, J. (2022). The Creativity of Text-to-Image Generation. In Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek '22) (192–202). Association for Computing Machinery. <https://doi.org/10.1145/3569219.3569352>
- Oppenlaender, J. (2024). The Cultivated Practices of Text-to-Image Generation. In R. Rousi, C. von Koskull, and V. Roto (Eds.), Humane Autonomous Technology (Chapter 14). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-66528-8_14
- Paananen, V., Oppenlaender, J., and Visuri, A. (2023). Using Text-to-Image Generation for Architectural Design Ideation. International Journal of Architectural Computing, 22(3), 458–474. <https://doi.org/10.1177/14780771231222783>
- Pažin, L. (2024). USING PlatfORMS and Tools To Create Business IntelligenCE. ShodhAI: Journal of Artificial Intelligence, 1(1), 68–75. <https://doi.org/10.29121/shodhai.v1.i1.2024.10>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139, 8748–8763). PMLR.
- Samuelson, P. (2023). Generative AI Meets Copyright. Science, 381, 158–161. <https://doi.org/10.1126/science.adi0656>
- Schetinger, V., Di Bartolomeo, S., El-Assady, M., McNutt, A., Miller, M., Passos, J., and Adams, J. (2023). Doom or Deliciousness: Challenges and Opportunities for Visualization in the Age of Generative Models. Computer Graphics Forum, 42, 423–435. <https://doi.org/10.1111/cgf.14841>
- Shneiderman, B. (2022). Human Centered AI. Oxford University Press. <https://doi.org/10.1145/3538882.3542790>