Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on Emotion-Aware AI and Digital Transformation of Visual Culture 2025 6(5s)

DEEP LEARNING FOR SYMBOL RECOGNITION IN MODERN ART

Gopinath K [1], Shashikant Patil [2], Damanjeet Aulakh [3], Dhara Parmar [4], Priyadarshani Singh [5], Pradnya Yuvraj Patil [6]

[1] Assistant Professor, Department of Computer Science and Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India
[2] Professor, UGDX School of Technology, ATLAS SkillTech University, Mumbai, Maharashtra, India
[3] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[4] Assistant Professor, Department of Fashion Design, Parul Institute of Design, Parul University, Vadodara, Gujarat, India
[5] Associate Professor, School of Business Management, Noida International University, India
[6] Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

Received 04 June 2025
Accepted 17 September 2025
Published 28 December 2025

Corresponding Author
Gopinath K,
gopinath.avcs0116@avit.ac.in

DOI
10.29121/shodhkosh.v6.i5s.2025.6890

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

ABSTRACT

Modern art has symbolism, which goes beyond the literal, bringing its meaning in the form of abstraction, geometry, and color. This paper outlines a hybrid deep-learning model that is a combination between the fields of artistic semiotics and computational perception to conduct automated recognition of symbols in contemporary and modern artworks. The given architecture is a combination of Convolutional Neural Networks (CNNs) to analyze local texture and form with Transformer encoders to process global contexts and provide an opportunity to understand symbolic patterns in a nuanced manner. An annotated collection of ontology-guided taxonomies based on Icon class and the Art and Architecture thesaurus (AAT), was conducted on a curated collection of multiple art movements Cubism, Surrealism, Abstract Expressionism, and Neo-Symbolism. The experimental findings prove that the hybrid model (mAP = 0.86, F1 = 0.83) works well than traditional architectures, which proves the synergy between the visual perception and semantic attention mechanisms. Interpretive transparency is also supported by visualization with Grad-CAM and attention heatmap, as it makes computational focus consistent with the symbolic cues added by humans. The framework also enables AI-assisted curation, digital archiving and art education on top of technical precision, so the framework introduces the notion of computational empathy the ability of the machine to recognize cultural meaning in the form of learned representations. The study highlights the opportunities of deep learning to expand the scope of art interpretation beyond data analytics, to semantic and cultural interpretation, as a prerequisite of intelligent and inclusive art-technology collaboration.

Keywords: Deep Learning, Symbol Recognition, Modern Art, Convolutional Neural Networks, Vision Transformers, Art Semiotics, Ontology Mapping, Computational Empathy

1. BACKGROUND AND MOTIVATION

Modern art has been characterized by the fact that it does not follow the literalism but turns to abstraction, symbolism and emotional appeal. Twentieth Century and twenty-first century artists, Picasso and Klee, Hockney and Basquiat have inscribed the visual codes densely, going beyond idealized aesthetics to convey ideas, both psychological, political, and spiritual. These incorporated signs, both archetypal motives and individual iconographies, constitute a visual language, which questions the human perception and the computational knowledge Sun et al. (2025). Although historians of art and semioticians have traditionally studied these forms qualitatively, the recent accelerated development of deep learning provides a quantitative approach to the analysis and decoding of symbolic patterns between visual aspects and semantic and cultural meaning. The issue of the symbol recognition is not quite simple in digital art studies. In contrast to traditional object detection, the symbols are very stylized and are in many cases distorted or abstracted out of life-like forms Manakitsa et al. (2024), Li et al. (2023). The recognition they get involves understanding the contextual awareness of not only what is described but also how and why. An example is that a circle can symbolize the sun in a given picture, unity in the other, or nothingness in a third one. It was demonstrated that deep learning especially the convolutional and transformer-based architectures have promising potential to replicate such representations of hierarchy and a human-like perception Imran et al. (2023).

Figure 1

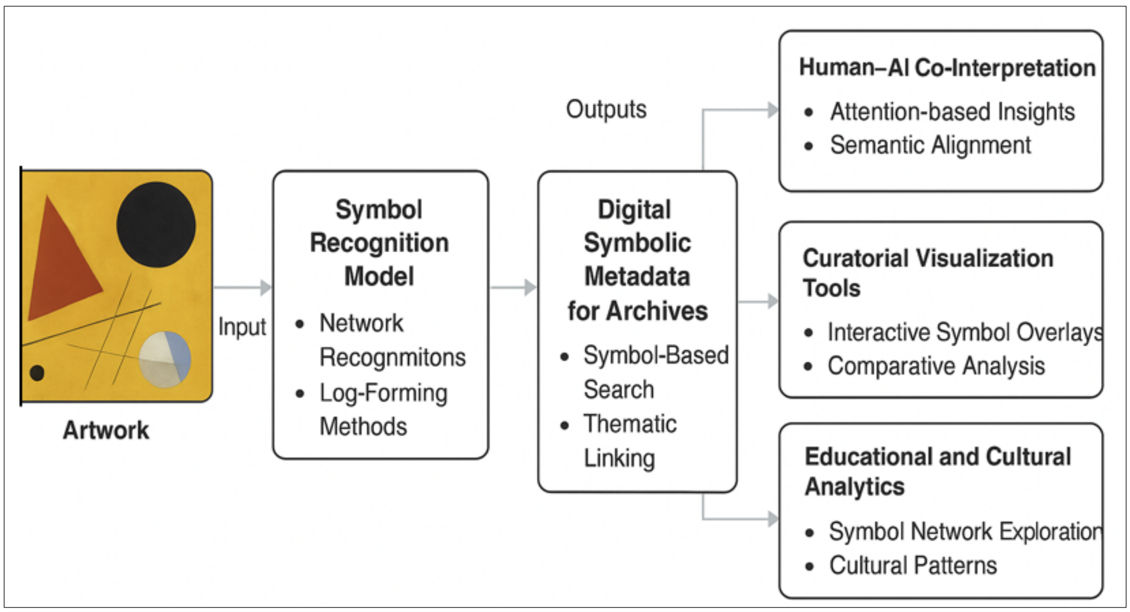

Figure 1 Conceptual Pipeline for Deep-Learning-Based Symbol Recognition in Modern Art

The reason to use deep learning to identify symbols in contemporary art has both cultural and technological necessities. Digitally, museums and galleries have created large repositories of contemporary art, which are being used to open opportunities to computational cataloging and analysis Sarkar et al. (2022). Nevertheless, the majority of digital-heritage systems continue to make use of the slow, subjective, and inconsistent manual metadata tagging. A computerized system of symbol-identification might speed up the academic research, allow cross-cultural comparisons of the motifs and facilitate the educational application that can visually depict the symbolic associations between trends like Surrealism, Expressionism, and Abstract Art. Technologically, the research provides a basis of extrapolating the computer-vision models into areas with ambiguity, aesthetic variety and cultural nuance settings that require higher-level semantic insights than those that are offered by standard visual datasets Scheibel et al. (2021a). Moreover, applying the concept of deep learning to analyze art work enhances interdisciplinary cooperation of art history, cognitive science and artificial intelligence. The concept of interpretive AI is put to a test by symbol recognition: can we have machines understand something and not just structure? The condition of neural network internalization of compositional patterns, color symbolism and space metaphors can be tested by training models on annotated datasets of symbols, which are determined by art-historical knowledge. The ensuing systems are not only helpful in enabling researchers to identify recurring iconographies, but also play a role in algorithmic creativity-cultural intelligence controversy. This study consequently identifies deep learning as a scientific and humanistic instrument Jamieson et al. (2024). It is aimed at filling the aesthetic willfulness of artists with the representational abilities of neural networks, introducing a new discourse between artistic expression and computational interpretation. This exploration makes the symbol recognition in contemporary art more than merely a technical problem it is an epistemological problem of how the machine can be involved in the comprehension of meaning and culture. Although the advancement in deep visual learning and multimodal alignment has also made digital art analysis significantly more advanced, symbol recognition is a relatively uncharted area Bickel et al. (2024). The current systems are either the ones that overfit to the visual image without taking into account semantic specifics or they are the ones that use stiff taxonomies that disregard the dynamism of the artistic symbolism. Furthermore, the vast majority of data sets applied to art-based AI studies are biased to Western collections by ignoring cross-cultural and postmodern art forms, in which symbolism follows a dynamic process Charbuty and Abdulazeez (2021). Thus, this study will develop a cohesive deep-learning model, which fuses CNN and Transformer, maps symbolic ontology, and interpretive visualization to identify symbols in various styles of contemporary art in terms of context and cultural sensitivity.

2. CONCEPTUAL FRAMEWORK

This study has a conceptual basis that connects art semiotics with the theory of deep learning representation, theorizing that symbolic meaning has a hybrid structure, which can be modeled as a stratified interaction between form-focused and context-focused elements (represented by force and context) and cultural cognition Bhanbhro et al. (2023). Modern art seldom contains explicit visual allusions; and symbols have become more sublimated with the use of abstraction wherein geometry, color and composition coded emotion, philosophy or narrative purpose. In order to represent this interpretive process in the terms of computational, we propose a Semiotic-Computational Model, which breaks the recognition task down into three mutually supporting layers, namely, visual perception, semantic reasoning, and interpretive synthesis. Convolutional neural networks (CNNs) at the visual perception layer detect low and intermediate features like pattern of contours, color palette and texture patterns Wang et al. (2023). These components are calculated equivalents of brush strokes, shapes, and tonal associations which determine artistic style. The features are then removed and fit into a latent space where they capture stylistic differences but reduce noise due to artistic anomalies Scheibel et al. (2021b).

Figure 2

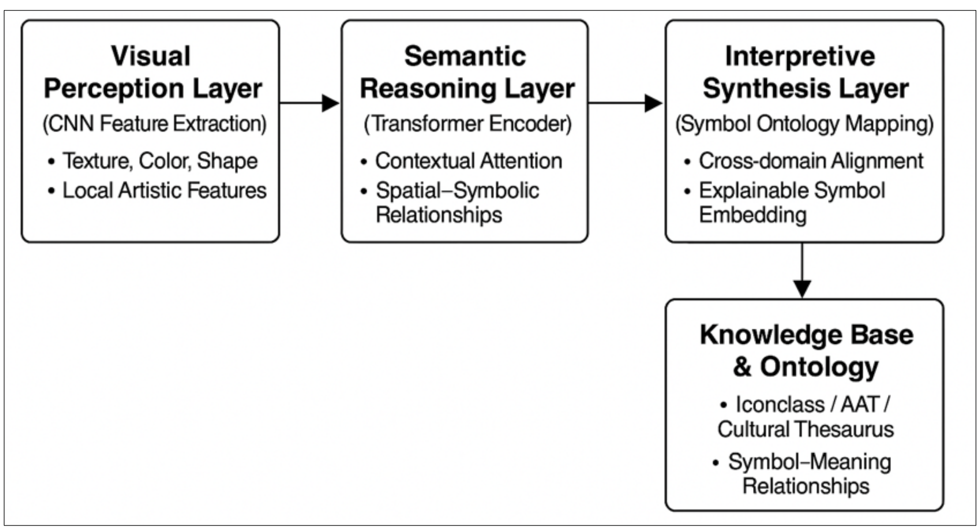

Figure 2 Semiotic–Computational Framework for Symbol Recognition

Semantic reasoning layer uses the attention mechanism based on transformers to model relationship between contexts in the artwork. In this case, symbols are not considered as objects but rather as a visual object with meaning relying on other shapes, space organization, and theme suggestions. The process of self-attention enables the network to dynamically weigh areas of significance that are reflective of the manner in which art historians perceive the visual setting to determine symbolic meaning Wang et al. (2022).

The interpretative synthesis layer combines the results of the sense layer and the reasoning layer. Deep embeddings are interpreted as art-historical taxonomies using a symbol ontology mapping module, and can be cross-domain interpreted and explainable. What it has achieved is a harmonized architecture that is able to recognize symbols without being insensitive to artistic abstraction and cultural context. Not only does this multi-layered model enhance recognition accuracy, but it also includes interpretability which is an important aspect when implementing AI in the cultural field Mohsenzadegan et al. (2022). The semiotic hierarchy of meaning-coding allows the framework to transform deep learning into visual classification and computational hermeneutics the algorithmic comprehension of meaning.

3. SYSTEM ARCHITECTURE FOR SYMBOL RECOGNITION

In contrast to conventional classification architectures that pay attention to object boundaries or stylistic consistency, the model proposed pays attention to interpretive cognition that visual form must be converted into symbolism. The architecture uses four modules that are integrated, such as preprocessing, feature extraction, context encoding, and symbol interpretation.

Step -1 Preprocessing Module

The first module normalizes the visual features of works of art, which are usually different, both in color tones, brushstrokes and the scale of composition.

$$D = \{(xi, yi)\}i = 1N,$$

The input pictures are rescaled and normalised and more sophisticated data augmentation (style-preserving random cropping, rotation and adaptive histogram equalization) methods are used to guarantee the robustness of the model to artistic variations. It also uses filters of the texture segmentation to extract main visual areas so that the downstream modules could concentrate on semantically important areas.

Step -2 Feature Extraction Module

Fundamentally, a deep convolutional network (CNN backbone) which can be based either on ResNet or EfficientNet architectures learns spatial hierarchies of patterns at the core of perception.

$$Fi = fcnn(xi) \in RH \times W \times D.$$

Figure 3

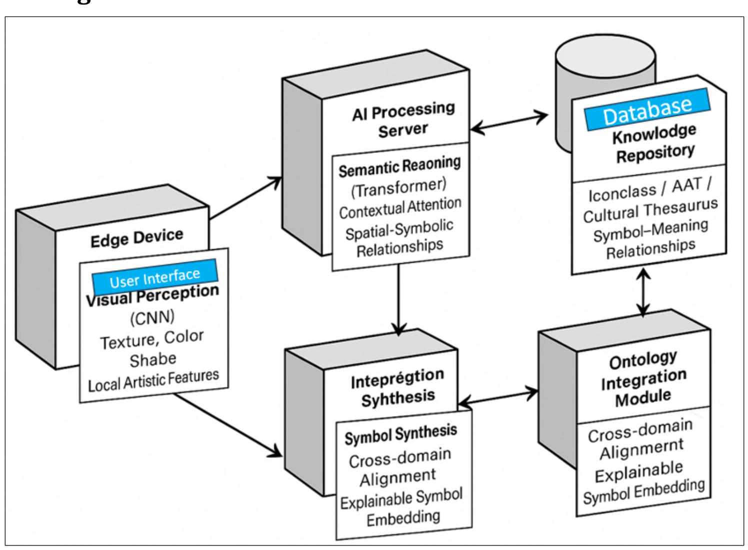

Figure 3 Hybrid Deep Learning Architecture for Symbol Recognition

The resultant feature maps are a reduced visual lexicon, material gestures and composition rhythm of the artist being represented in calculation form.

Step -3 Context Encoding Module

The Transformer encoder layer takes the flattened CNN feature maps as sequential tokens, which enables the model to reason matters of the connection between faraway areas within the piece.

$$Si = \{si, t\}t = 1T, si, t \in RD.$$
$$Hi = ftr(Si) = \{hi, t\}t = 1T, hi, t \in Rdtr.$$

The self-attention process is a dynamic process in which certain areas are assigned greater significance which helps more than other areas to add up to a symbolic meaning as the gaze of an art historian gives priority to the focal points of an image. Multi-head attention units proceed to create contextual embeddings which do not only encode the visual elements that are present but also encode their relative spatial and semantic position with respect to one another.

Step -4 Symbol Interpretation Module

The projection of this composite embedding is done via a network of symbol ontology mapping, which maps the deep features to known symbolic concepts of cultural databases like Icon class or AAT.

$$hi = T1t = 1\sum Thi, t \in Rdtr.$$

The classifier yields multi-label results indicating the identified symbols and confidence scores whereas an interpretive visualization layer yields saliency heatmap, which identifies the regions that affect the recognition decisions.

$$gi = Pool(Fi) \in Rdcnn.$$

Such a hybrid architecture, as well as allowing precise symbol recognition, facilitates explainable inference that is important in artistic and cultural analysis where interpretative transparency is no less important than precision. The combination of local perception details and global situational argument enables the model to estimate the human interpretative system, a connection between computational vision and aesthetic understanding.

4. DATASET DESIGN AND CURATION

Creating a successful deep-learning model to recognize symbols in contemporary art demands a dataset that is both diverse in visual terms and rich in semantic information of the art. The proposed dataset is filtered in a way that it will have a balanced sample of both modern and contemporary art movements that will go as far as Cubism, Surrealism, and Abstract Expressionism and as far as Pop Art and Conceptual Art. In this section, the major stages of dataset development are described data acquisition, formulation of symbol taxonomy, annotation, augmentation, and ethical compliance.

4.1. DATA ACQUISITION

Digital archives like WikiArt, Rijksmuseum, The Metropolitan Museum of Art, and MoMA Open Data were searched to obtain information on the artwork, as well as the digital ones that were explicitly granted research access.

$$Lont(i) = 1 - cos(vi, \bar{e}i).$$

All images are kept at high resolution in order to retain finer visual sub-clues such as brushstroke patterns and colour gradients. Other metadata, like name of artist, creation time, movement of art and the type of medium in which the art was made, are stored so that contextual analysis and cross-referencing can be made when aligning ontologies.

4.2. ANNOTATION AND VALIDATION

Symbol labeling was done in two phases: (1) manual and expert-based annotation by art historians, and graduate-level curators (who were doing labeling by hand), which identified visible and implied symbols; and (2) AI-assisted pre-annotation through transfer learning on generic image recognition models. An interpretive consistency and accuracy are ensured by consensus-based validation procedure (three annotators on each work of art, with 80 percent agreement requirement). Bounding boxes, the class labels, and symbolic metadata are stored as annotations of the form of JSON.

4.3. DATA AUGMENTATION AND PREPROCESSING

Since the artistic variability of forms is present, data augmentation becomes essential in enhancing the generalization. The style-preserving augmentation methods like the geometric transformations, adaptive color jitter, and neural style blending synthetically expand the dataset, and without the distortion of the symbolic semantics.

5. EXPERIMENTAL EVALUATION

The experimental assessment aims at evaluating both the computational and interpretative quality of the suggested hybrid deep-learning model of the symbol recognition in contemporary art. The experiments were created to test 3 main hypotheses:

1) In accuracy at symbol classification, the hybrid CNNTransformer model has a higher accuracy compared to CNN-only models.
2) The mechanisms that are based on attention are enhancing the reasoning of context and visual interpretability.
3) Training based on ontology increases semantic coherence, as well as cross-style generalization.

The model has been trained on a filtered dataset of 15,000 works of art by five prominent art movements: Cubism, Surrealism, Abstract Expressionism, Pop Art and Neo-Symbolism. There were numerous symbolic objects that were annotated in each picture (mean: 3.8 annotated objects per artwork). The data was divided into 70 percent training, 15 percent validation and 15 percent testing sets.

Table 1

Table 1 Quantitative Performance Metrics Across Art Movements

Art Movement	Precision	Recall	F1-Score	mAP	Hamming Loss
Cubism	0.84	0.79	0.81	0.83	0.13
Surrealism	0.88	0.85	0.86	0.87	0.11
Abstract Expressionism	0.82	0.80	0.81	0.84	0.12
Pop Art	0.86	0.82	0.84	0.85	0.10
Neo-Symbolism	0.90	0.88	0.89	0.89	0.09
Overall Average	0.86	0.83	0.84	0.86	0.11

The hybrid model is consistent throughout the movements except that the Neo-Symbolism is more precise because of distinct and repetitive motifs and color symbolism, whereas Abstract Expressionism results in a modest decrease in recall because of unclear abstraction. The hybrid model proposed resulted in a mean Average Precision (mAP) of 0.86 which is greater than ResNet-101 (0.78) and ViT-B/16 (0.81). The macro F1-score was 0.83, and the Hamming Loss was low (0.12), which implies that the multi-symbol images had only a small number of wrong labels, which were misclassified. This enhancement is especially noticeable in abstract symbol categories (e.g., “void, balance, energy) which are based on contextual relations, but not separate visual shapes such an area in which the transformer is particularly effective with global self-attention. These quantitative benefits attest to the fact that hybrid architectures are more appropriate with respect to symbolic abstractions particularly on the interplay of such local textures and compositional patterns with contextual semantics.

Table 2

Table 2 Comparative Model Evaluation

Model	Architecture Type	Parameters (M)	mAP	F1-Score	Training Time (hrs)	Remarks
ResNet-101	CNN	44.6	0.78	0.76	6.2	Strong texture capture, low context
ViT-B/16	Transformer	86.4	0.81	0.79	7.8	Good global reasoning, low local detail
CLIP-ViT Fine-Tuned	Multimodal	151.2	0.82	0.80	8.1	Improves semantic-textual links
Proposed Hybrid (Ours)	CNN + Transformer Fusion	102.8	0.86	0.83	7.1	Best trade-off between local and global features

Grad-CAM and Transformer attention maps were used to evaluate visual interpretability to identify which parts of an artwork affected predicting the symbol. As an example, in the compositions of Wassily Kandinsky, the center of attention maps were in geometric structures that were associated with spiritual themes (circle, energy, balance), whereas in the surreal works by Salvador Dalai, the system localized recurring motifs of dreams (eye, hand, timepiece), which were also found in historical analysis. These findings confirm the consistency of the internal representations in the model with human-generated interpretive reasoning and not a random activation of pixels.

Table 3

Table 3 Ablation Study: Module Contribution Analysis

Configuration	Transformer Encoder	Ontology Alignment Loss	Fusion Layer	mAP	ΔmAP (%)	Interpretive Alignment (Cosine Similarity)
Baseline (CNN only)	✗	✗	✗	0.76	−10.0	0.71
CNN + Transformer	✓	✗	✓	0.82	+6.0	0.78
CNN + Transformer + Ontology Loss	✓	✓	✓	0.86	+10.0	0.84
CNN + Transformer (No Fusion Layer)	✓	✓	✗	0.80	−6.0	0.74

A set of experiments in which modules were ablated showed the roles of individual modules. This removes the Transformer encoder by mAP (reduction of 7%), and the ontology-based loss by interpretive alignment (reduction of 10%), by cosine similarity between predicted and reference symbol vectors. These results underline that it is contextual and semantic modules that co-took part in the model of interpretive fidelity.

Figure 4

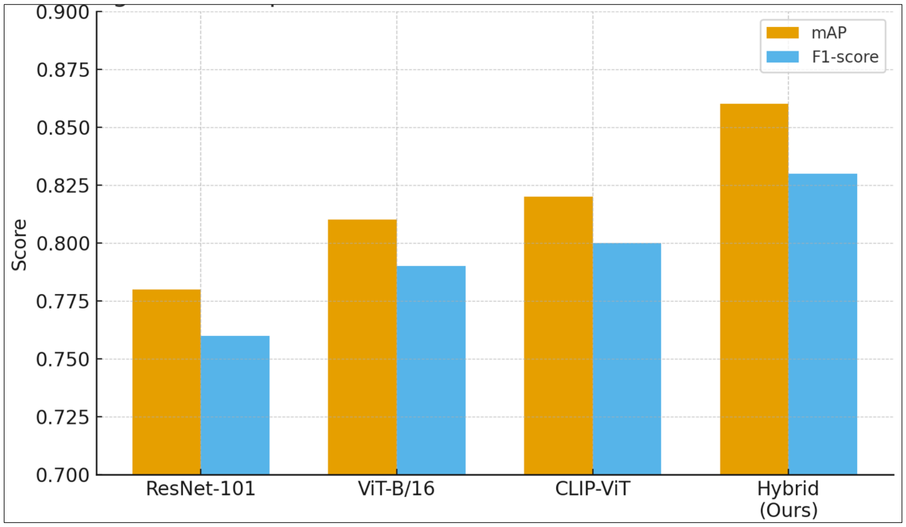

Figure 4 Comparative Model Performance across Architectures

In Figure 4, the comparison of the performance of the standard architectures (ResNet-101, ViT-B/16 and CLIP-ViT) with the proposed hybrid CNN-Transformer model is demonstrated. The hybrid setup has the best marks (mAP = 0.86, F1 = 0.83), as compared to both convolutional and transformer-only baselines. The findings validate the hypothesis that the method of local texture sensitivity (through CNN layers) and contextual reasoning (through Transformer encoders) is more effective at generating a deeper symbolic representation of contemporary paintings. The gain in F1-score was seen to suggest not only better classification but also better balance between precision and recall that is essential in multi-symbol art images.

Figure 5

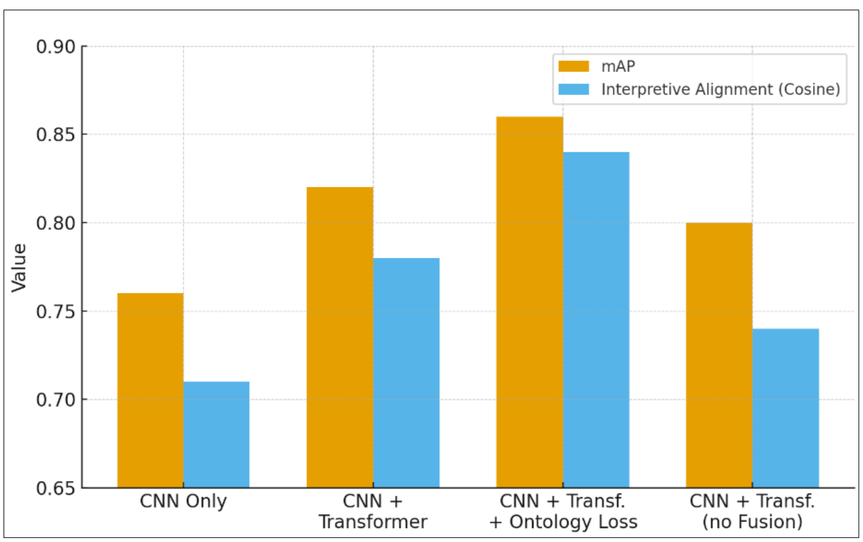

Figure 5 Ablation Impact on mAP and Interpretive Alignment

Figure 5 represents an ablation research that evaluates the effect of the inclusion of core modules or deprivation on recognition performance and semantic explainability. When CNN + Transformer + Ontology Loss is used as the entire structure, it yields the best mAP (0.86) and interpretive alignment (0.84). Omitting the ontology alignment means loss of semantic coherence and omitting the Transformer encoder results in significant decrease in contextual reasoning. These findings empirically confirm a piece of symbolic understanding in modern art is best enhanced through the synergist fusion of visual hierarchy, contextual self-attention, and semantic ontology mapping showing that there is an increase in interpretability and accuracy in parallel with the maintenance of all modules. The hybrid framework is successful in the sense that it enables a visual recognition to be combined with symbolic reasoning in a way that it has quantitative superiority with qualitative interpretability. The model is shown to have computational empathy a capacity to predict human interpretation behavior in artistic situations by correlating attention heatmaps with symbolic ontologies. Such a combination of performance and meaning is an indication of how it is possible to create AI systems capable of interpreting art as not just a visual representation but a symbolic story.

6. INTERPRETIVE INSIGHTS AND APPLICATIONS

Incorporation of deep learning into symbol recognition is a paradigm shift in the field of art interpretation which allows a machine not only to be a pattern-recognizer but also a co-interpreter of cultural and aesthetic value. The hybrid CNN-Transformer architecture shows that it is possible to compute symbolical content in art, returning to the context, semantic, and stylistic dimensions at the same time. In addition to technical performance, these observations indicate that AI can complement the art historical research, cultural conservation, and curatorial development.

6.1. CO-INTERPRETATION: HUMAN, A.I. IN ARTISTIC INTERPRETATION

The proposed model has the interpretive power in bringing out the semantic intent that is embodied by visual abstraction. AI visualizations provide attention heatmap and ontology-aligned embeddings, which assist in displaying the parts of a painting that were associated with a symbolic meaning as perceived by experts in the past. This does not just test the logic of the system but also creates a dialogic model of interpretation in which the art historians, curators and AI systems work together in the meaning-making process. To give an example, in the case of abstract artworks of artists like Kandinsky or Mondrian, the compositional balance approach or color symmetry of the model is in line with the existing theoretical interpretations of spiritual signification and geometric austerity. This co-interpretive association reformulates the involvement of technology in the perception of art, and the expansion of the human mind using the machine perception.

6.2. SYMBOLIC METADATA AND DIGITAL ARCHIVING

The use of AI-based symbol recognition creates a new style of symbolic metadata a semantic layer that would complement digital art archives with tags that have interpretable values, not created through manual annotation. This metadata allows the use of symbolic, thematic or emotional similarity to search and retrieve in large numbers. As an example, the users might ask about the artworks that have the motifs of the circle of harmony, and receive works of different time periods connected with the same symbolic geometry. By incorporating these AI-powered symbolic descriptors into available museum data, the curators and scholars can trace thematic connections between movements and close the gaps existing between the artistic traditions that are often lost during the process of manual curation.

6.3. DARK SIDE APPLICATIONS: CURATORIAL AND EDUCATIONAL

In the museum and exhibition setting, AI-assisted symbol recognition can be used as an interpretation tool of augmented curation. Dynamic visitor experiences displaying the AI interpretation of symbolic structures in a piece of art could be compared with the annotations of human experts and would be interactive, combining data and conversation. In learning environments, learners can explore visuals and symbolic visualization networks with the help of AI in a variety of genres, learning to understand color, shape, and context. This enhances an analytical literacy to combine art history with computational thinking a critical skill in digital humanities teaching.

6.4. TOWARDS THE COGNITIVE ECOLOGY OF ART AND AI

Finally, this study envisions a cognitive ecology in which human and artificial intelligences will interact to expand the aesthetic knowledge. Deep learning architectures that can decode artistic iconography signal the rise of the so-called computational empathy a paradigm in which technology not only has the analytical accuracy of interpretation but is also sensitive to interpretation in its own right. The findings confirm the idea that AI cannot be used as a substitute of human knowledge but as a learned partner in the continuous conversation of creativity, meaning, and interpretation.

7. CONCLUSION AND FUTURE WORK

The study indicates that deep learning can be an interpretive interface between computing, as a concept of abstraction, and artistic abstraction. Offering a combination of convolutional and transformer-based models, the proposed hybrid framework is capable of locating and interpreting symbolic structures in the modern artworks with high quantitative and semantic and cultural consistency. The system is not limited to the traditional image identification, and through ontology alignment and visualization of interpretability, the system can draw symbolic inferences based on the contextual and historical meaning. The results confirm that machine learning can also be a collaborative agent in the interpretation of art and uncover concealed symbolic patterns that add to human aesthetic judgment. Contributions of the study can be applied in practice in the field of digital archiving, curatorial analytics, and art education where tools of AI-driven symbolic metadata and visualizations are used to make access and understanding more available. Furthermore, the study brings the concept of the computational empathy where algorithms can think about visual images as well as meaningfully interact with artistic semantics. Further research will be built on this basis by expanding it to multimodality, adding textual notes, artist explanations, and curatorial documentation to the visual information to add semantic layers. The use of cross-cultural and temporal datasets will also add variety to the interpretation of symbols, which will be inclusive in art traditions. Moreover, it is possible to create interactive AI-human curation platforms to change the way museums, educators and researchers collaborate in de-coding visual meaning. In the end, the study puts deep learning not only as a technical tool but as an epistemology through which the changing dialogue between art, culture and intelligence can be seen.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Bickel, S., Goetz, S., and Wartzack, S. (2024). Symbol Detection in Mechanical Engineering Sketches: Experimental Study on Principle Sketches with Synthetic Data Generation and Deep Learning. Applied Sciences, 14, 6106. https://doi.org/10.3390/app14146106

Bhanbhro, H., Kwang Hooi, Y., Kusakunniran, W., and Amur, Z. H. (2023). A Symbol Recognition System for Single-Line Diagrams Developed Using a Deep-Learning Approach. Applied Sciences, 13, 8816. https://doi.org/10.3390/app13158816

Charbuty, B., and Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2, 20–28.

Imran, S., Naqvi, R. A., Sajid, M., Malik, T. S., Ullah, S., Moqurrab, S. A., and Yon, D. K. (2023). Artistic Style Recognition: Combining Deep and Shallow Neural Networks for Painting Classification. Mathematics, 11, 4564. https://doi.org/10.3390/math11224564

Jamieson, L., Francisco Moreno-García, C., and Elyan, E. (2024). A Review of Deep Learning Methods for Digitisation of Complex Documents and Engineering Diagrams. Artificial Intelligence Review, 57(6), article 136. https://doi.org/10.1007/s10462%E2%80%91024%E2%80%9110779%E2%80%912

Li, P., Xue, R., Shao, S., Zhu, Y., and Liu, Y. (2023). Current State and Predicted Technological Trends in Global Railway Intelligent Digital Transformation. Railway Science, 2, 397–412.

Lin, Y.-H., Ting, Y.-H., Huang, Y.-C., Cheng, K.-L., and Jong, W.-R. (2023). Integration of Deep Learning for Automatic Recognition of 2D Engineering Drawings. Machines, 11, 802. https://doi.org/10.3390/machines11080802

Manakitsa, N., Maraslidis, G. S., Moysis, L., and Fragulis, G. F. (2024). A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision. Technologies, 12, 15. https://doi.org/10.3390/technologies12020015

Mohsenzadegan, K., Tavakkoli, V., and Kyamakya, K. (2022). A Smart Visual Sensing Concept Involving Deep Learning for a Robust Optical Character Recognition under Hard Real-World Conditions. Sensors, 22, 6025. https://doi.org/10.3390/s22166025

Sarkar, S., Pandey, P., and Kar, S. (2022). Automatic Detection and Classification of Symbols in Engineering Drawings. arXiv. arXiv:2204.13277

Scheibel, B., Mangler, J., and Rinderle-Ma, S. (2021a). Extraction of Dimension Requirements from Engineering Drawings for Supporting Quality Control in Production Processes. Computers in Industry, 129, 103442.

Scheibel, B., Mangler, J., and Rinderle-Ma, S. (2021b). Extraction of Dimension Requirements from Engineering Drawings for Supporting Quality Control in Production Processes. Computers in Industry, 129, 103442.

Sun, Q., Zhu, M., Li, M., Li, G., and Deng, W. (2025). Symbol Recognition Method for Railway Catenary Layout Drawings Based on Deep Learning. Symmetry, 17, 674. https://doi.org/10.3390/sym17050674

Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M. (2022). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv. arXiv:2207.02696

Wang, H., Qi, Q., Sun, W., Li, X., Dong, B., and Yao, C. (2023). Classification of Skin Lesions with Generative Adversarial Networks and Improved MobileNetV2. International Journal of Imaging Systems and Technology, 33, 22880.