AI-Driven Recognition and Sustainable Preservation of Ancient Murals: The DKR-YOLO Framework

Zixuan Guo and Sameer Kumar *

Posted Date: 4 August 2025

doi: 10.20944/preprints202508.0180.v1

Keywords: cultural heritage preservation; ancient mural recognition; deep learning; YOLOv8; robustness; interpretability; inclusive AI; Sustainable digitalization; DySnake Conv; adaptive convolutional kernels

Article

AI-Driven Recognition and Sustainable Preservation of Ancient Murals: The DKR-YOLO Framework

Zixuan Guo and Sameer Kumar *

Universiti Malaya

* Correspondence: sameer@um.edu.my

Highlights

- Proposes DKR-YOLO for accurate and efficient recognition of ancient mural elements
- Integrates DySnake Conv layer to enhance fine-grained mural detail detection
- Employs adaptive convolutional kernels to optimize feature representation
- Incorporates RE-FPN and ECA mechanisms for prioritizing critical mural features
- Experimental results show a 43.6% reduction in FLOPs and a 5.1% increase in mAP

Abstract

This paper introduces DKR-YOLO, an advanced deep learning framework designed to empower the digital preservation and sustainable management of ancient mural heritage. Building upon YOLOv8, DKR-YOLO integrates innovative components—including the DySnake Conv layer for refined feature extraction and an Adaptive Convolutional Kernel Warehouse to optimize representation—addressing challenges posed by intricate details, diverse artistic styles, and mural degradation. The network's architecture further incorporates a Residual Feature Augmentation (RFA)-enhanced FPN (RE-FPN) and Efficient Channel Attention (ECA), prioritizing the most critical visual features and enhancing interpretability. Extensive experiments on mural datasets demonstrate that DKR-YOLO achieves a 43.6% reduction in FLOPs, a 3.7% increase in accuracy, and a 5.1% improvement in mAP compared to baseline models. This performance, combined with an emphasis on robustness and interpretability, supports more inclusive and accessible applications of AI for cultural institutions—including small museums and local heritage organizations—thereby fostering broader participation and equity in digital heritage preservation.

Keywords: cultural heritage preservation; ancient mural recognition; deep learning; YOLOv8; robustness; interpretability; inclusive AI; Sustainable digitalization; DySnake Conv; adaptive convolutional kernels

---

1. Introduction

Ancient murals are invaluable historical artifacts, providing a unique window into the cultural, religious, and social contexts of past civilizations. Found in religious sites, palaces, and tombs(Bolong et al., 2022), these artistic treasures offer tangible connections to the lives and beliefs of ancient societies and play a crucial role in reconstructing historical narratives and understanding artistic evolution (Veysi, 2022). However, the analysis and interpretation of these murals pose significant challenges due to deterioration, fading colors, and the complexity of the scenes depicted(Li et al., 2023). Traditional approaches, which often involve manual inspection and documentation, are time-consuming, error-prone, and difficult to scale. Furthermore, the intricate details and large scale of many murals make it challenging to capture and analyze all relevant information comprehensively(Zhong et al., 2020). The advent of digital image processing and computer vision

 doi:10.20944/preprints202508.0180.v1

techniques has transformed mural research, enabling automated extraction and analysis of mural elements with greater efficiency and accuracy(Amura et al., 2021).

The Dunhuang grottoes, first excavated in the second year of the Jian Yuan period during the Former Qin dynasty, encapsulate over a millennium of historical and cultural integration among various ethnic groups, ultimately forming a distinct Chinese Buddhist art system(Zhang, 2023). The Dunhuang murals are indispensable for comprehensively understanding the history of Chinese art and play a crucial role in fostering modern artistic innovation(Du, 2024). With the rapid development of information technology and the accelerated digitization of cultural heritage, restoration techniques and digital collection methods for Dunhuang murals have significantly advanced(Wang et al., 2024). Libraries, archives, and museums now house extensive collections of Dunhuang mural images. However, the abstract visual forms and obscure semantic content of these images often present challenges for users attempting to construct precise search queries, resulting in difficulties in retrieving relevant images and low resource utilization(Tekli, 2022). These challenges impede researchers' ability to study Dunhuang murals effectively and dampen public enthusiasm for exploring Dunhuang culture. The integration of computer image recognition technology into the search domain for Dunhuang murals presents a promising solution to enhance resource acquisition efficiency(Zeng et al., 2022). By improving the restoration and protection processes, as well as optimizing the digital collection and storage systems, this technology can contribute significantly to the safeguarding and transmission of Dunhuang's cultural heritage. Specifically, it supports the preservation of these murals by improving resource utilization and advancing information dissemination, ultimately ensuring the longevity and accessibility of this invaluable cultural legacy.

The YOLO series has been a pioneering and influential approach in object detection, fundamentally transforming how visual information is perceived and processed(Ali & Zhang, 2024). YOLOv8 integrates state-of-the-art advancements, pushing the boundaries of object detection and classification(Chappidi & Sundaram, 2024). Murals, as one of the most fundamental forms of painting, have become an integral part of human cultural history(Cetinic & She, 2022). The study of traditional artworks has increasingly benefited from the application of digital technologies. Machine learning techniques, including input vectors, have been used to categorize landscape paintings, while attention-based long short-term memory (LSTM) networks have been employed to classify Chinese paintings(Mei et al., 2021). Currently, most automated mural categorization techniques rely on computer vision. Common methods include contour-based similarity measurements, multi-instance grouping, and semantic retrieval models, which establish connections between the content of ancient murals and their underlying meanings(Jaworek-Korjakowska et al., 2023). Traditional methods for classifying wall paintings have shown some success; however, they are limited to extracting only the fundamental attributes of the images(Liu et al., 2021). The subjectivity and diversity inherent in wall paintings pose significant challenges in adequately capturing high-level features such as texture, color properties, and other intricate detail(Petracek et al., 2023). In recent years, numerous significant studies have focused on leveraging deep learning for mural classification(Menai, 2023). Zhou et al. selected 660 mural images with a shared thematic focus, but the limitation of using artworks from the same group restricts the generalizability of their approach(Huang et al., 2024).

Existing methods often fail to capture the hierarchical structure of ancient murals, which is crucial for understanding the spatial relationships between different elements(Liu et al., 2024). This limitation leads to poor localization and classification performance, particularly for small or partially occluded elements(Zhang et al., 2023). Furthermore, the absence of a robust feature pyramid network (FPN) in many current techniques hampers the extraction of multi-scale features, resulting in the loss of critical information across different levels of the image(Koo et al., 2021). Another major challenge is the inability of current algorithms to effectively address the unique issues posed by ancient murals, such as low-resolution images, uneven lighting, and the presence of artifacts. These factors significantly degrade the performance of object detection and classification models, leading to increased rates of false positives and false negatives. Moreover, the reliance on handcrafted features in traditional methods limits their adaptability and generalizability to diverse types of murals. The

 doi:10.20944/preprints202508.0180.v1

absence of dynamic feature extraction mechanisms further impedes the ability of these models to effectively learn and represent the complex patterns and textures inherent in ancient murals(Xiao et al., 2023).

In addition to YOLO, other object detection algorithms such as R-CNN, Faster R-CNN, and SSD have made considerable progress in deep learning applications(Aboyomi & Daniel, 2023). For instance, Tian et al. (2023) improved the YOLO model to achieve 97.5% accuracy and a 95.86% recall rate in belt surface defect detection. Zhang (2024) introduced the TPH-YOLOv5 model, incorporating the distortion predictor head (TPH) and integrating the volume block attention model (CBAM) to improve detection performance by approximately 7% compared to the baseline YOLOv5 model. Kim et al. (2022) proposed a trained free-package approach, which effectively combined adaptable and efficient training resources with a compound zoom technique. YOLOv7, with an accuracy of 56.8% average precision (AP), outperformed other real-time object detection systems, achieving over 30 frames per second in object recognition(Sakiba et al., 2023). Recent advancements in self-attention mechanisms and transformer-based models have significantly improved image anomaly detection and feature extraction. Shao et al. (2023)introduced COVAD, a self-attention-based deep learning model designed for content-oriented video anomaly detection. Their method leverages deep feature extraction and adaptive attention mechanisms to enhance anomaly localization. Similarly, Zhang and Tian (2023) proposed a transformer architecture based on mutual attention for image anomaly detection, demonstrating improved feature representation by focusing on interdependent image regions. The Efficient Attention Pyramid Transformer (EAPT) proposed by a study in IEEE Transactions on Multimedia introduced a multi-scale attention mechanism that effectively captures both local and global image features(Lin et al., 2021).

Dunhuang murals represent a key focus in the field of digital humanities, a discipline that merges computer science with the humanities and social sciences, facilitating the integration of digital technology into cultural heritage preservation and dissemination(Zhu et al., 2023). This interdisciplinary approach has reshaped traditional research paradigms and significantly advanced the study of cultural history. Wang et al. (2021) developed a semantic framework for Dunhuang murals, creating a domain-specific vocabulary to bridge the semantic gap in image retrieval. Zeng et al. (2024) employed the bag-of-visual-words method to extract features from mural images and used support vector machines (SVMs) for classification, exploring the thematic distribution and dynastic evolution of the murals. Ren et al. (2024) focused on mural restoration techniques based on generative adversarial networks (GANs), automating restoration by learning the relationship between degraded and restored mural textures. proposed a mural restoration algorithm based on sparse coding of line drawings, while Fei et al. (2023) enhanced the curvature diffusion algorithm with adaptive strategies for improved restoration. Mu et al. (2024) designed the “Restore VR” system, enabling users to experience the restoration of Dunhuang murals through virtual reality (VR) digital tours of the caves. Recent research in medical image analysis has also contributed valuable methodologies applicable to mural recognition. Ali et al. (2024) introduced an efficient glomerular detection network capable of identifying multiple pathological features in medical imaging, demonstrating robust anomaly detection through deep convolutional networks. Similarly, Nazir et al. (2021) employed an embedded clustering sliced U-Net with a fusing strategy for intervertebral disc segmentation and classification, achieving high precision in medical image segmentation. These models highlight the effectiveness of adaptive convolutional techniques in handling complex image structures.

The integration of deep learning into cultural heritage learning has also been explored in serious gaming and animation research. A recent study on using scaffolding theory in serious games for traditional Chinese murals culture learning demonstrated the effectiveness of interactive learning frameworks in preserving cultural knowledge(Li et al., 2024). Additionally, an animation line art colorization approach based on the optical flow method showcased a sophisticated method for handling line-based artworks, which could be adapted to enhance the automated recognition and restoration of ancient murals(Yu et al., 2024). These studies emphasize the role of deep learning in

 doi:10.20944/preprints202508.0180.v1

both cultural education and visual processing, which aligns with our objective of improving mural recognition.

The "Digital Dunhuang" project leverages scientific and technological methods to digitally collect, process, and preserve Dunhuang's cultural heritage, creating a multimodal and interconnected digital resource library accessible globally(Song, 2023). Despite these advancements, research on efficient and user-friendly search methods for Dunhuang murals remains insufficient, hindering the optimal utilization of cultural heritage resources(Lian & Xie, 2024). YOLOv8 is the latest iteration in the YOLO series, representing an end-to-end, compact neural network built on deep learning principles (Abo Jouhain, 2024). The DKR-YOLOv8, a refinement of YOLOv8 proposed in this study, enhances the feature extraction network of YOLOv8. This study introduces a mural image target recognition algorithm, DKR-YOLOv8, designed to address the unique challenges posed by ancient murals. The following is a summary of the key contributions of this work:

- Dataset Construction: A specialized dataset for Dunhuang grotto murals was compiled, encompassing 30 distinct classes of common mural art features. Rigorous evaluation procedures ensured the inclusion of only high-quality images, which were subsequently annotated manually to identify key elements within the mural artworks.
- Model Enhancement: Building upon the original YOLOv8 architecture, we incorporated DySnake Conv, a module known for its sensitivity to elongated topologies. This modification significantly improved the detection accuracy of mural art features, especially those with irregular shapes and elongated structures.
- Integrated Approach: Our method integrates an active target detection module, three-channel spatial attention mechanisms, and the dynamic convolution technique from Kernel Warehouse. These innovations reduce model parameters, mitigate overfitting, and alleviate computational and memory demands. Furthermore, this strategy substantially enhances the model's ability to accurately and reliably detect mural art elements.

The FPN structure has been substituted with the RE-FPN, a residual feature fusion pyramid structure that is more effective and lightweight. Furthermore, a layer of SOD is added to improve the model's ability to detect objects of various sizes, especially those that are small.

2. Materials and Procedures

2.1. The YOLOv8 Model's Architecture

YOLOv8 is an updated iteration of Ultralytics' publicly available single-stage object detection algorithm, offering variations in five editions: n, s, m, l, and x, each featuring an increased number of model parameters(Cederin & Bremberg, 2023). Similar to YOLOv5, YOLOv8 removes the feature extraction, feature fusion, and prediction modules. In addition, the cross-stage partial network (C3) from YOLOv5 is replaced by a more gradient-rich C2f module, which enhances information acquisition capabilities. The volume structure in the sampling phase is removed from the PAN-FPN, which increases computational speed. Moreover, the non-slip detector in YOLOv8 outperforms traditional frame-based methods in terms of accuracy, enabling faster object detection and identification. The loss function in YOLOv8 consists of classification losses and regression losses, where classification losses are computed using binary cross-entropy (BCE) loss, while regression losses are calculated using the distributed focal loss and union loss. Both categories of losses are weighted by predefined coefficients, and the total loss is computed through a weighted combination, as shown in Figure 1.

 doi:10.20944/preprints202508.0180.v1

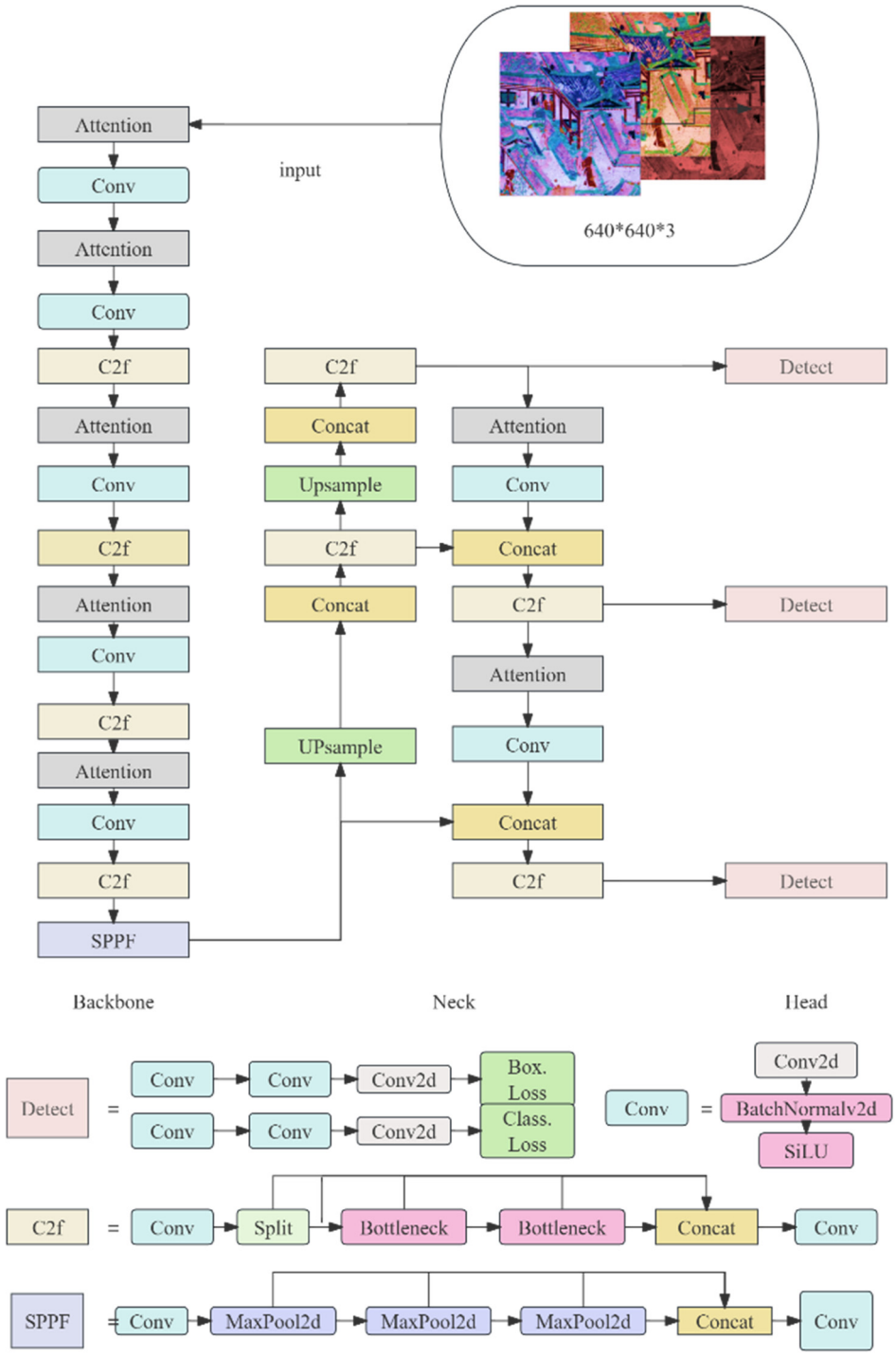


Figure 1. YOLOv8 model stench map.

Despite these advancements, there remains a pressing need for a comprehensive algorithm capable of effectively integrating these techniques to address the unique challenges posed by ancient mural images. The Kernel Warehouse, a repository of customizable kernel functions, presents a promising pathway for the development of such an algorithm. The proposed approach aims to enhance accuracy and robustness in recognizing and classifying elements within ancient mural images by combining the strengths of Kernel Warehouse, YOLOv8, feature pyramid networks, and Snake Convolution.

 doi:10.20944/preprints202508.0180.v1

2.2. Dynamic Snake Convolution

DySnake Conv, proposed in 2023, has predominantly been applied in the segmentation and recognition of medical images(Zheng et al., 2024). Its advantages over conventional convolutional kernels stem from its enhanced feature extraction capabilities and its flexibility in handling non-rigid object shapes. Building on this foundation, we have customized the DySnake Conv layer specifically for the complex task of identifying and predicting the intricate lines and patterns characteristic of ancient mural images. The flexible nature of the DySnake Conv kernel enables it to effectively learn the complex geometric features of the targets. By utilizing a restricted set of offsets as learning parameters, the model can dynamically identify the most optimal locations for feature extraction. This ensures that, even in the presence of substantial deformations, the detection area remains stable.

The topology, as depicted in Figure 2, has been fine-tuned to capture the subtle lines and artistic details inherent in mural images. This refinement leads to a more accurate and comprehensive analysis of ancient murals, allowing the deep learning network to extract local information with enhanced precision.

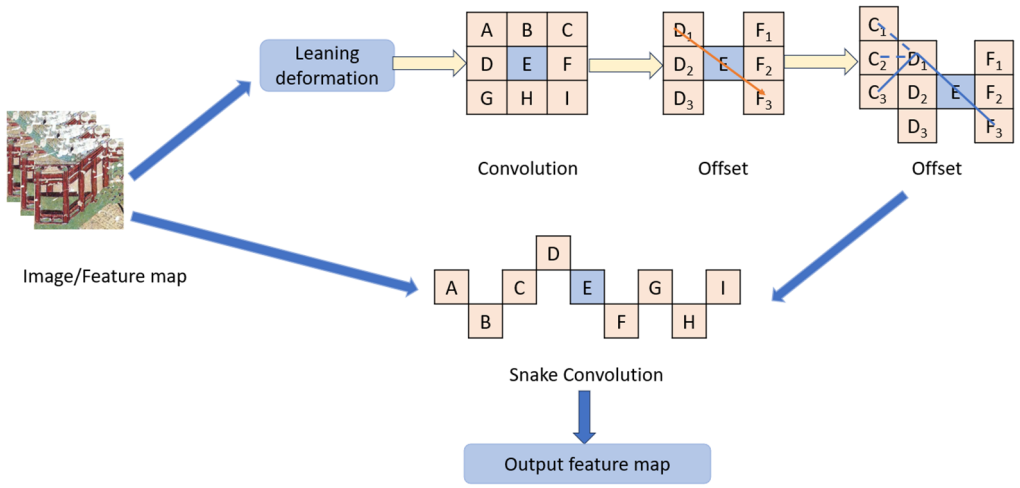


Figure 2. Structure of the DSConv.

The expression for a conventional 3x3 and 2D convolution kernel K is as follows:

$$K=\{(\alpha-1,\beta-1),(\alpha-1,\beta),\cdots,(\alpha+1,\beta+1)\} \tag{1}$$

Motivated by deformable convolution, the deformation offset Δ is introduced to enhance the flexibility of the convolution kernel, enabling it to focus on the intricate geometric features of the target. However, if the model is allowed to learn the deformation offset freely, the receptive field may drift away from the target, particularly when dealing with thin and elongated structures. To address this issue and maintain continuity in the attention mechanism, we employ an iterative approach to determine the subsequent position of each target to be processed. This iterative process ensures that the perception field remains focused on the target, preventing excessive expansion due to large deformation shifts.

$$K_{i+c}=(\alpha_{i+c},\beta_{i+c})=(\alpha_i+c,\beta_i+\sum\Delta\beta_{i+c}) \tag{2}$$

The deformation rule of DSConv along the *x*-axis is as follows:

 doi:10.20944/preprints202508.0180.v1

$$K_{i\pm c}=\begin{cases}\left(\alpha_{i+c},\beta_{i+c}\right)=\left(\alpha_i+c,\beta_i+\sum_{i}^{i+c}\Delta\beta\right)\\ \left(\alpha_{i-c},\beta_{i-c}\right)=\left(\alpha_i-c,\beta_i-\sum_{i-c}^{i}\Delta\beta\right)\end{cases} \tag{3}$$

The deformation rule along the *y*-axis is as follows:

$$K_{i\pm c}=\begin{cases}\left(\alpha_{j+c},\beta_{j+c}\right)=\left(\alpha_j+c,\beta_j+\sum_{j}^{j+c}\Delta\alpha,\beta_j+c\right)\\ \\ \left(\alpha_{j-c},\beta_{j-c}\right)=\left(\alpha_j-c,\beta_j-\sum_{j-c}^{j}\Delta\alpha,\beta_j-c\right)\end{cases} \tag{4}$$

When using bilinear interpolation, the offset Δ is usually expressed as a fraction. The process looks like this:

$$K=\sum_{\dot{K}}B\left(K',K\right)\cdot K' \tag{5}$$

The bilinear interpolation kernel is represented by B, and all of the listed integral spatial coordinates are denoted by K′. Two one-dimensional kernels comprise the bilinear interpolation kernel:

$$B\left(K,K'\right)=b\left(K_\alpha,K_\alpha^{'}\right)\cdot b\left(K_\beta,K_\beta^{'}\right) \tag{6}$$

DSConv is a deformation-based two-dimensional transformation process that spans a 9×9 area, offering enhanced perception of key features. This is especially advantageous for mural detection, which involves long-distance, small, and elongated targets. By covering a 9×9 range during the deformation process, DSConv expands the model's receptive field, improving the detection of crucial features and laying the foundation for accurate recognition. The introduction of dynamic serpentine convolution allows adaptive adjustments to the shape of the convolution kernel, facilitating the precise capture of local features and structural information in images. This approach humanizes the processing of complex images, resulting in better feature extraction capabilities and enhanced robustness.

Through this method, the next position of each target to be processed is selected sequentially, ensuring continuity of focus. Following this, conventional convolution is applied. Dynamic serpentine convolution ensures continuity in the convolution kernel's modifications by accumulating offset values. This design allows for flexible selection of the receptive field without excessive dispersion, maintaining more accurate and stable focus on elongated targets. Additionally, bilinear interpolation guarantees the smoothness and precision of the convolution process, further improving the effectiveness of convolution operations.

In the DSConv module, the combination of conventional convolution and dynamic serpentine convolution retains the stability and efficiency of traditional convolution while introducing the flexibility and adaptability of dynamic serpentine convolution. The three-layer convolution structure enhances and diversifies feature extraction. Experimental results validate that the improved model performance is better suited to handle the complex features of various targets.

2.3. Kernel Warehouse Dynamic Convolution

Dynamic rolling examines a mixed rolling nucleus made of n static rolling kernel linear mixture, weighting through their sample-related attention, demonstrating excellent performance compared to ordi-nary rolling(Dunkerley, 2023). However, previous designs have problems with parameter

 doi:10.20944/preprints202508.0180.v1

efficiency: they increase the number of rolling parameters by n times. This problem, together with the complexity of optimization, led to the lack of research advancement in the field of dynamic volume, which did not allow the use of bigger n values (e.g., n > 100, rather than conventional n < 10) to push the performance bounds.

Kernel Warehouse is broad version of dynamic envelope that can a beneficial balance be-tween parameter efficiency and representation. Its main idea was to reinterpret the fundamental terms "assembled convoluted nucleus" and "convolved nucleus" in relation to dynamic convolutions, with an emphasis on greatly expanding the number of nuclei while decreasing their dimensions. Kernel Warehouse increases the reliance of volume parameters within the same layer as well as between the con-tinuous layers through advanced kernel partition and warehouse sharing.

In particular, on any ConvNet volume layer, Kernel Warehouse splits the static nucleus into m noncompatible nuclei with the same dimensions. It then computes the linear mixture of each nucleus based on a predefined "warehouse" that contains n nuclei (e.g., n = 108) that is also shared by several ad-jacent volume layers. Finally, it assembles each m mixture sequentially, offering a higher degree of free-dom while adhering to the necessary parameter budget. A new attention function was also developed by the authors to aid in the process of learning to reconcile attention to the cluster nucleus.

Dynamic convolution is a method that uses input dependent attentions to learn a linear combination of n static kernels. This approach has been shown to outperform standard convolution in terms of perfor-mance. However, it results in an increase of the convolutional parameters by a factor of n, making it inef-ficient in terms of parameters. This lack of research advancement hinders researchers from exploring settings where n>100, which is far bigger than the common setting of n<10.Such exploration is crucial for advancing the performance boundary of dynamic convolution while maintaining parameter efficiency. In this paper, we introduce Kernel Warehouse as a solution to address this gap. Kernel Warehouse is a more comprehensive version of dynamic convolution that redefines the fundamental concepts of "kernels", "as-sembling kernels", and "attention function". It does so by leveraging the convolutional parameter depend-encies within the same layer and across adjacent layers of a ConvNet. We validate the efficacy of Kernel Warehouse on ImageNet and MS-COCO datasets by employing diverse ConvNet topologies. Kernel Warehouse can be applied to Vision Transformers, leading to a decrease in the size of the model's core while also enhancing the model's precision. For example, Kernel Warehouse (n=4) delivers 5.61%|3.90%|4.38% absolute top-1 accuracy gain on the ResNet18|MobileNetV2|DeiT-Tiny backbone, and Kernel Warehouse (n=1/4) achieves 2.29% gain on the ResNet18 backbone while reducing the model size by 65.10%.

Kernel Warehouse provides a use friendly interface that allows users to easily search, filter, and compare different kernels depending on their individual requirements and limits. This covers factors such as kernel size, stride, padding, and dilation, enabling users to rapidly identify the most suited kernel for their application. The schematic diagram describing the Kernel Warehouse is illustrated in Figure 3.

 doi:10.20944/preprints202508.0180.v1

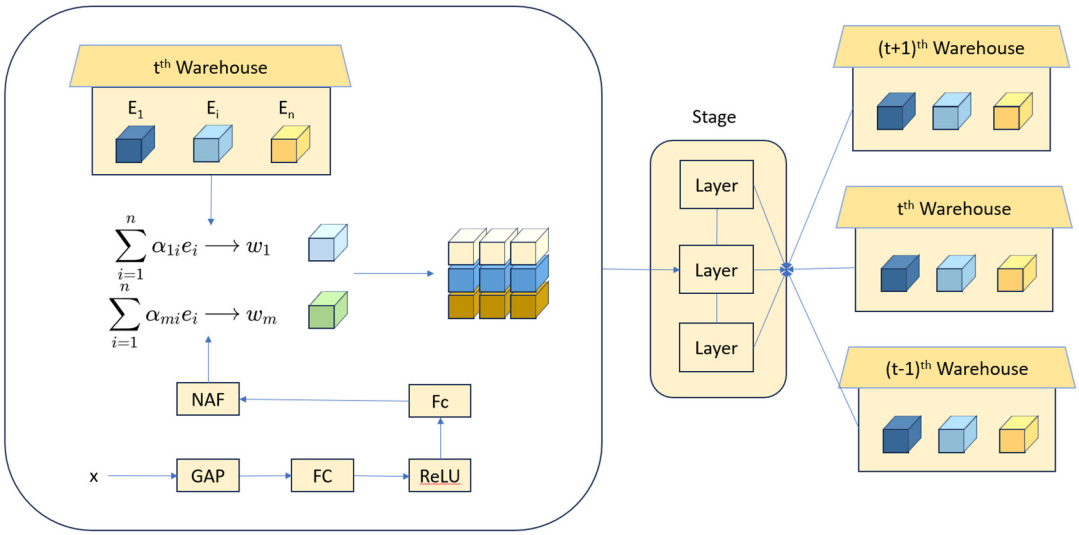


Figure 3. Schematic of the Kernel Warehouse dynamic convo-lution approach.

Another key feature of Kernel Warehouse is its efficiency in managing large-scale image datasets(Zhao et al., 2022). It incorporates an advanced data management system that ensures effective storage and retrieval of kernel functions, significantly reducing the computational cost associated with feature extraction(Ahsan et al., 2023). This capability allows our system to efficiently process vast volumes of ancient mural images without compromising the accuracy of recognition and classification. Kernel Warehouse seamlessly integrates with other components of our method, such as the Feature Pyramid Network and Snake Convolution. This synergistic interaction between the components enhances the overall performance of our system, enabling improved accuracy and efficiency in recognizing and classifying elements of ancient mural images. As the cornerstone of our proposed approach, Kernel Warehouse provides a versatile and efficient framework for feature extraction from ancient mural images. Its adaptability, speed, and seamless integration with other components make it an invaluable asset in our quest for precise and efficient recognition and classification of ancient mural elements.

2.4. Lightweight Residual Feature Pyramid Network

The YOLOv8-based recognition and classification method for ancient mural image elements relies significantly on the Feature Pyramid Network (FPN) architecture. FPN addresses the limitations of traditional object detection models, which often struggle to accurately detect objects of varying scales within an image (Bhalla et al., 2024). The FPN architecture utilizes a top-down approach with lateral connections, facilitating the extraction of multi-scale features from the input image. This structure is crucial in enhancing the YOLOv8 model's effectiveness in analyzing ancient mural paintings. FPN enables the model to capture both high-level semantic information and fine-grained details in mural images by constructing a pyramid of feature maps at varying resolutions. This multi-scale feature representation is essential for the accurate detection and classification of complex elements within the intricate patterns of ancient murals.

The FPN architecture consists of multiple stages, each responsible for extracting features at different scales. The initial stages focus on capturing low-level features, such as edges and textures, while later stages capture higher-level features, including object shapes and patterns. The lateral connections between these stages facilitate the fusion of information across different scales, resulting in a comprehensive feature representation well-suited for the recognition and classification of mural image elements. Additionally, the FPN architecture is highly customizable and can be seamlessly integrated with other components of the YOLOv8 model, such as the Snake Convolution module. This integration ensures efficient and effective feature extraction from mural images, significantly enhancing the overall performance of the recognition and classification algorithm. By leveraging the

 doi:10.20944/preprints202508.0180.v1

advantages of the FPN architecture, the YOLOv8-based method can reliably identify and classify a wide range of elements within ancient mural images, thereby contributing to a deeper understanding of these invaluable cultural artifacts.

$$L^{ut} = \sum_{u=1}^{U} \sum_{p} \| C^{*}_{u(p)} - C^{ut}_{u(p)} \|_2^2 \tag{7}$$

$$L^{bt} = \sum_{b=1}^{B} \sum_{p} \| C^{*}_{b(p)} - C^{bt}_{b(p)} \|_2^2 \tag{8}$$

Shallow features in deep neural networks contain rich spatial information and fine-grained details but are often accompanied by noise. As the network deepens, the semantic content in the features increases, while the spatial and minor detail information gradually diminishes, and the noise decreases. This study focuses on enhancing the efficiency of the neck architecture in the YOLOv8 model by leveraging the Feature Pyramid Network (FPN) to facilitate the integration of features. The goal is to enable the flow and interaction between features at different depths. Furthermore, we introduce a novel approach to address the core issue of spatial information loss caused by channel transformations in higher-level features within the FPN framework. This is achieved by incorporating the Residual Feature Augmentation (RFA) unit into the model design. The RFA module utilizes residual branches to inject contextual information from various spatial positions, thereby improving the feature representation of higher-level features. Additionally, an ECA (Efficient Channel Attention) module, a lightweight attention mechanism, is integrated into each branch of the FPN to further minimize the loss of spatial information. A lightweight residual feature fusion pyramid structure, called RE-FPN, is developed by applying a 3×3 Depthwise Convolution (DW Conv) operation to each feature map, as illustrated in Figure 4.

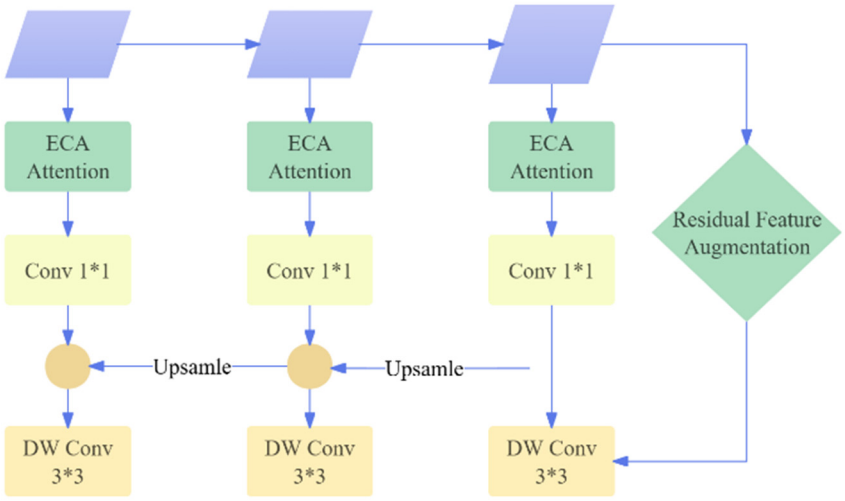


Figure 4. the pyramid structure for the lightweight residual feature fusion.

The primary goal of this approach is to enhance feature interaction, reduce spatial information loss, and preserve the lightweight nature of the YOLOv8 algorithm's FPN. By encouraging the movement and communication of features at different network levels, the RE-FPN structure strengthens high-level feature representation through the inclusion of contextual information via residual branches, tackling the issue of spatial information loss in higher-level feature channels. Furthermore, the integration of a simple attention mechanism in each FPN branch reduces spatial information loss. The application of 3×3 depthwise convolution on each feature map creates the RE-FPN structure, combining the advantages of feature interaction and spatial information retention

 doi:10.20944/preprints202508.0180.v1

while maintaining the network's lightweight characteristics. The RE-FPN thus enhances feature interaction and representation with minimal spatial information loss, improving the overall performance of the YOLOv8 algorithm in object detection tasks.

To address the issue of ineffective feature fusion across stages in the algorithm, this paper introduces the FPN structure at the neck position to enhance feature circulation and information interaction at various stages. The outputs from layers 2, 4, 6, and 7 of the MobileNet V2 network are selected as the input features for the FPN structure, denoted as C2, C3, C4, C5, with input feature map sizes set to 38×38×32, 19×19×96, 10×10×320, and 10×10×1280, respectively. To further enhance the model's lightweight characteristics, the number of channels in each stage of the FPN structure is reduced from 256 to 160 using 1×1 convolution, minimizing the model parameters and computational load. Additionally, the 3×3 standard convolution in the FPN structure is replaced by a 3×3 depthwise separable convolution, further reducing the model's volume. At the feature level, the FPN structure propagates strong semantic features from higher to lower levels, improving object detection performance through feature fusion. However, in the process of feature fusion, low-level features benefit from the strong semantic information of high-level features, while the features at the highest pyramid level lose information due to channel reduction. To address this issue, adaptive pooling is utilized to extract varying contextual information, reducing the loss of information in the highest-level features within the feature pyramid. By introducing the residual feature enhancement (RFA) module, the residual branch injects contextual information from different spatial positions, thus enhancing the feature expression of high-level features in a more concise manner, with reduced computational requirements.

3. Experiment and Results

3.1. Hardware and Software Configuration

The experimental platform for this study, as shown in Table 4, At the same time, the model uses unified super parameters in the training phase, where the optimizer is Random Gradient Decrease (SGD), the initial learning rate is 0.01, the weight decrease is 0.0005, the number of training rounds is 200 cycles.

Table 1. Experimental platform.

Name	Related Configuration
Operating system	Windows 11 (64 bit)
CPU	Intel (R) Core (TM) i7-14700HX
GPU	NVIDIA GeForce RTX 3050
Software and environment	PyCharm 2021.3, Python 3.8, Pytorch 1.10

The experimental setup for the Kernel Warehouse, Feature Pyramid Network (FPN), and Snake Convolution YOLO (DKR-YOLO) detection and classification method for ancient mural image elements based on YOLOv8 was carefully designed to ensure accurate and reliable results. Image preprocessing was conducted to remove noise, artifacts, and distortions that could hinder algorithm performance.

The Kernel Warehouse module was constructed using a combination of predefined and custom-designed kernel functions, optimized to capture the unique characteristics of ancient mural elements. The Feature Pyramid Network utilized a series of dilated convolutional layers, enabling the extraction of multiscale features. The Snake Convolution module was integrated into the YOLOv8 architecture to enhance the detection and classification of irregularly shaped objects, such as the intricate patterns and themes found in historical murals.

Data augmentation was applied based on the distribution of images in the dataset. This technique helps the network learn a broader range of features, improving its performance on unseen

 doi:10.20944/preprints202508.0180.v1

data and enhancing generalization while reducing the risk of overfitting. Data augmentation also acts as a regularization method, promoting more stable convergence during training, particularly for complex model structures or uneven data distributions. To enhance the diversity and robustness of the dataset, 20 different random combinations were generated to augment the image set, with bounding boxes adjusted accordingly. Figure 6 shows examples of the augmented images from each process.

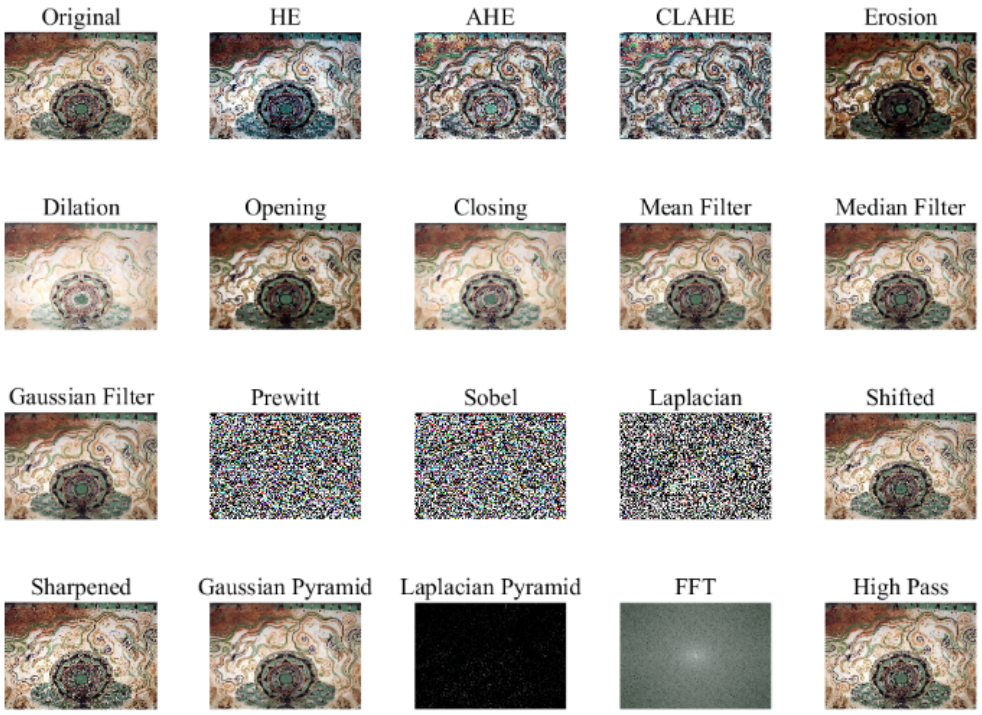


Figure 5. Examples of enhanced images.

Various image processing operations are applied using specific parameters and methods to enhance image quality. Histogram Equalization (HE) improves image contrast by equalizing the brightness histogram. Adaptive Histogram Equalization (AHE) further enhances local contrast, while Contrast-Limited Adaptive Histogram Equalization (CLAHE) improves contrast while limiting it to prevent noise amplification. Erosion and dilation operations utilize a disk-shaped structuring element with a radius of 5 to eliminate small noise or fill minor gaps. Similarly, opening and closing operations employ disk-shaped structuring elements to remove small objects and close small holes. Mean filtering is used for smoothing, applying a 5x5 filter to reduce noise. Median filtering, using a 5x5 median filter, effectively removes impulse noise. Gaussian smoothing, using a 5x5 Gaussian filter, reduces high-frequency noise. Gradient filtering is employed for edge detection and detail enhancement. Shift filtering operations translate the image by one pixel to the left to simulate image translation. Sharpening enhances image edges and fine details. The Gaussian pyramid method performs down-sampling, while the Laplacian pyramid achieves down-sampling by subtracting the up-sampled version of the down-sampled image. The Fourier Transform is applied to convert the image to the frequency domain, using a logarithmic function and normalized spectrum. Fourier low-pass and high-pass filtering, using Gaussian filters, are employed to retain the low-frequency and high-frequency components of the image, respectively.

3.2. Recognition Results

The Dunhuang murals, as representative examples of ancient Chinese grotto art, originated during the Sixteen Kingdoms period and have a history spanning over 1,500 years. These murals

 doi:10.20944/preprints202508.0180.v1

exhibit diverse styles across different historical periods, making them one of the world's most significant artistic treasures. The patterns and designs in Dunhuang murals are highly varied, commonly seen in architectural elements such as herringbone patterns, flat beams, and caissons, as well as in decorative features of Buddhist artifacts like niche lintels, canopies, lotus thrones, and halos. Additionally, strip borders are frequently used to divide and embellish architectural and mural spaces. The designs encompass a broad array of motifs, including flowers, plants, clouds, birds, beasts, flames, geometric shapes, and gold lamps. These elements are rich in artistic atmosphere and imbued with profound symbolic meanings. They not only reflect refined aesthetic tastes but also create unique spatial perceptions, contributing significant artistic, aesthetic, and scholarly value.

This study utilizes imagery from The Complete collection of Chinese Dunhuang murals(Duan, 2006). These murals feature a variety of images and styles. The DOI of the dataset for the images is 10.57967/hf/4516(URL is https://huggingface.co/datasets/jinmuxige/dunhuang). The dataset is divided evenly, with 50% allocated for the validation set and 50% for the training set. A selection of these motifs was processed for the experiment, and the results are presented in Table 2.

Table 2. Mural Recognition types.

category	name	introduce	sets
human landscap	Fans	These fans not only have practical uses but also carry rich cultural meanings, embodying the artistic achievements of ancient craftsmen.	85
	Honeysuckle	In the edge of Dunhuang grotts, such as caisings, flat tiles, wall layers, arches, niches, and canopies, honeysuckle patterns are used as edge decorations.	40
	Flame	Flames in Dunhuang murals often appear as decorative patterns such as back light and halo, symbolizing light, holiness, and power. Around religious figures like Buddhas and Bodhisattvas, the use of flame patterns enhances their holiness and grandeur.	35
	Bird	Birds are common natural elements in Dunhuang murals. They adding vivid life and natural beauty to the murals.	28
	Pipa	As an important ancient plucked string instrument, the pipa frequently appears in Dunhuang murals, especially in musical and dance scenes. These pipa images not only showcase the form of ancient musical instruments but also reflect the music culture and lifestyle of the time.	62
	Konghou	The konghou is also an ancient plucked string instrument and is a significant part of musical and dance scenes in Dunhuang murals.	34
	tree	Trees in Dunhuang murals often serve as backgrounds or decorative elements, such as pine and cypress trees. They not only add natural beauty to the mural but also symbolize longevity, resilience, and other virtuous qualities.	38
productive labor	Pavilion	Pavilions are common architectural images in Dunhuang murals. These architectural images not only display the artistic style and technical level of ancient architecture but also reflect the cultural life and aesthetic pursuits of the time.	76

 doi:10.20944/preprints202508.0180.v1

	Horses	Horses in Dunhuang murals often appear as transportation or symbolic objects, such as warhorses and horse-drawn carriages. These horse images are vigorous and powerful, reflecting the military strength and lifestyle of ancient society.	72
	Vehicle	Vehicles, including horse-drawn carriages and ox-drawn carriages, are also common transportation images in Dunhuang murals. These vehicles not only showcase the transportation conditions and technical level of ancient society but also reflect people's lifestyles and cultural habits.	49
	Boat	While boats are not as common as land transportation in Dunhuang murals, they do appear in scenes reflecting water-based life. These boat reflecting the water transportation conditions and water culture characteristics of ancient society.	22
	Cattle	Cattle in Dunhuang murals often appear as farming or transportation images, such as working cows and ox-drawn carriages. These cattle images are simple and honest, closely connected to the farming life of ancient society.	32
religious activities	Deer	Deer in Dunhuang murals often symbolize goodness and beauty. In some story paintings or decorative patterns, deer images add a sense of vivacity and harmony to the mural.	52
	Clouds	Clouds in Dunhuang murals often serve as background elements. They may be light and graceful or thick and steady, creating different atmospheres and emotional tones in the mural. The use of clouds also symbolizes good wishes such as good fortune and fulfillment.	72
	Alage wells	Algae Wells are important architectural decorations. Located at the center of the ceiling, they are adorned with exquisite patterns and colors. They not only serve a decorative purpose but also symbolize the suppression of evil spirits and the protection of the building.	126
	Baldachin	Canopies or halos in Dunhuang murals may appear as head lights or back lights, covering religious figures such as Buddhas and Bodhisattvas, symbolizing holiness and nobility.	43
	Lotus	The lotus is a common floral pattern in Dunhuang murals, symbolizing purity, elegance, and good fortune. Below or around religious figures such as Buddhas and Bodhisattvas.	24
	Niche Lintel	Niche lintels are the decorative parts above the niches in Dunhuang murals, often painted with exquisite patterns and colors. These niche lintel images not only serve a decorative purpose but also reflect the artistic achievements and aesthetic pursuits of ancient craftsmen.	10

 doi:10.20944/preprints202508.0180.v1

Pagoda	Pagodas are important religious architectural images in Dunhuang murals. These pagoda images not only showcase the artistic style and technical level of ancient architecture but also reflect the spread and influence of Buddhist culture.	66
Monk Staff	The monastic staff is a commonly used implement by Buddhist monks and may appear as an accessory to monk figures in Dunhuang murals. As an important symbol of Buddhist culture undoubtedly adds a strong religious atmosphere to the mural.	29

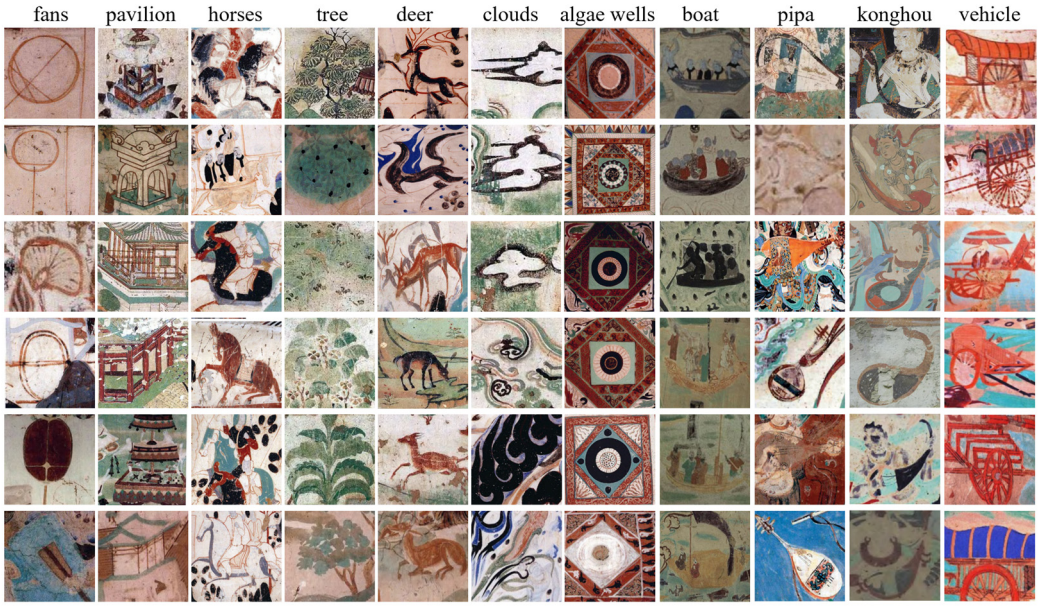


Figure 5. Examples of mural images.

The enhanced DKR-YOLO model was employed to identify 20 classes of symbols in photographs of the Mo Kao Grotto murals at Dunhuang. Figure 5 showcases some of the results from the mural recognition process. As depicted in Figure 6, the upgraded DKR-YOLO model effectively recognizes murals featuring multiple targets. It can also detect murals that have been partially hidden or modified using Gaussian blur.

 doi:10.20944/preprints202508.0180.v1

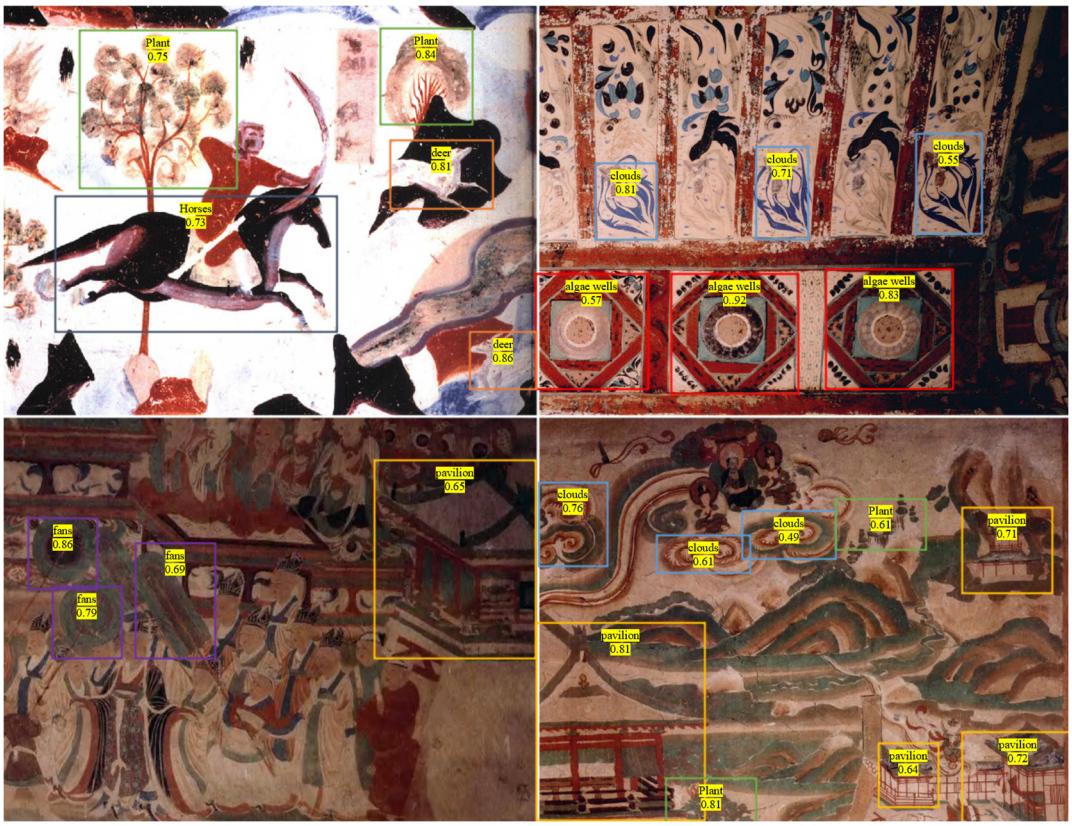
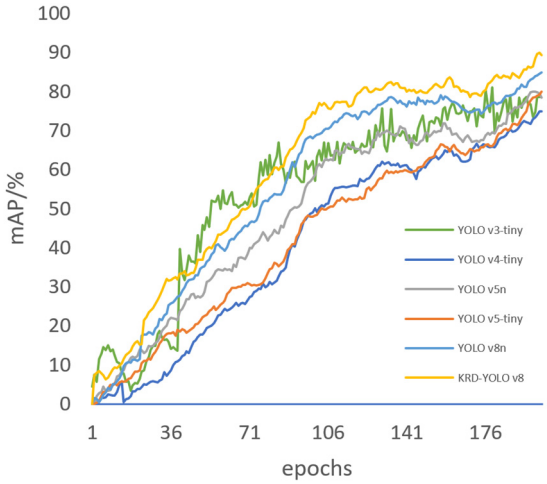


Figure 6. Mural image recognition findings.

3.3. Test Results on mural Dataset

We assessed effectiveness of revised DKR-YOLO model by contrasting it with six target detection models: YOLOv3-tiny, YOLOv4-tiny, YOLOv5n, YOLOv7-tiny, and YOLOv8.This assessment was completed using our bespoke ancient wall paintings dataset under identical experimental settings. In Figure 7, the mAP training curve shows the differences in average accuracy between the updated model and the models mentioned earlier.

The data in Figure 7 unequivocally shows that the enhanced DKR-YOLOv8 outperforms YOLOv3-TINY, YOLOv4-TINY, YOLOv5n, YOLOv7-TINY, and YOLOv8 in mean average precision (MAP). Although the convergence speed of DKR-YOLOv8 is significantly slower than that of YOLOv3-TINY, it achieves better convergence and higher MAP after 200 epochs of adequate training. To more effectively highlight the accuracy and performance of the improved DKR-YOLOv8 model, Table 1 presents a comparative experiment of recognition accuracy and performance among the DKR-YOLOv8 model and other models.



 doi:10.20944/preprints202508.0180.v1

Figure 7. MAP Variation Curves of Different Models.

Table 2. presents a comparison of network models in terms of recognition accuracy and performance.

Simulations	P/%	R/%	mAP/%	F1/%	FPS
YOLOv3-tiny	79.2	79.6	81.4	78.8	557
YOLOv4-tiny	81.4	74.8	82.6	78.1	229
YOLOv5n	80.1	75.2	82.3	77.2	326
YOLOv7-tiny	81.3	73.8	81.2	76.9	354
YOLOv8	78.3	75.4	80.6	77.2	526
DKR-YOLOv8	81.6	80.9	88.2	80.5	592
	(0.2%↑)	(1.6%↑)	(6.8%↑)	(2.1%↑)	(6.2%↑)

In Table 1, there is a detailed comparison of the accuracy in detecting murals between YOLOv7-tiny and DKR-YOLOv8.Despite YOLOv7-tiny having a precision that is 0.2 percentage points lower than DKR-YOLOv8, its recall, mAP, F1 score, and FPS do not match the efficiency of DKR-YOLOv8.Specifically, DKR-YOLOv8 obtains a recall of 80.9% and a mAP of 88.2%.The goal of this research is to build models with increased accuracy and faster detection speed, which necessitates taking into account factors like recognition accuracy and detection speed. Despite YOLOv3-tiny's impressive 557 frames/s FPS, it falls short of DKR-YOLOv8 in terms of accuracy, recall, mAP, and F1 score, with only a 35 frames/s difference between them. Primarily, the DKR-YOLOv8 model outperforms the YOLOv3-tiny, YOLOv4-tiny, YOLOv5n, YOLOv7-tiny, and YOLOv8 prototypes. The values enclosed in brackets show the difference in performance between DKR-YOLOv8 and the runnerup model, while the figures highlighted with underscores in Table 1 correspond to the second-highest scores in each category. Upward arrows imply increased outcomes.

3.4. Ablation Experiment

To evaluate the efficiency of the upgraded DKR-YOLOv8 model, a series of comparative ablation studies were conducted. Eight ablation experiments were performed to consistently assess model performance, utilizing the same dataset, training configuration, and methodology in each case. The validation analyses included precision, recall, integrated evaluation metrics, and mean Average Precision (mAP). The first row of Table 3 presents the training results for the original YOLOv8 model, without any augmentation techniques. The incorporation of Dynamic Snake Convolution (DSC), Kernel Warehouse dynamic convolution, or Lightweight Residual Feature Pyramid Network (LRFPN) into the baseline YOLOv8 model led to improvements in detection rates, accuracy, and computational efficiency across all updated models. Notably, the use of the Kernel Warehouse technique resulted in a reduction of approximately 50% in FLOPS. Furthermore, mAP improved by 5.1% with Kernel Warehouse and by 4.1% with DSC, while the LRFPN method enhanced the recall rate by 3.5%. Following these independent algorithm enhancements, the resulting model is designated as DKR-YOLOv8. This model demonstrates significant improvements, with a 4.7% increase in precision, a 3.8% increase in recall, a 7.0% increase in the F1 score, and a 5.6% increase in mAP compared to the original YOLOv8. It also achieves the lowest FLOPs and a frame rate of 590 FPS. In summary, the enhanced DKR-YOLOv8 model attains an 80.9% recall rate and an 81.2% mAP. The increase in FPS, along with reductions in model parameters and FLOPS, substantially boosts detection speed, enabling deployment on both mobile and stationary observation platforms.

 doi:10.20944/preprints202508.0180.v1

Table 3. Experiments on ablation of network models.

Models	Based Models	DSC	KW	RE-FPN	P/%	R/%	F1/%	mAP/%	FLOPs (G)	FPS
Model1	YOLOv8				77.9	77.9	75.2	83.5	28.4	529
Model2	YOLOv8			✓	81.6	73.9	78.8	80.8	27.7	868
Model3	YOLOv8		✓		77.3	74.8	78.7	81.3	26.9	640
Model4	YOLOv8	✓			84.0	83.2	84.5	82.1	14.2	474
Model5	YOLOv8		✓	✓	80.9	74.8	75.9	82.0	27.9	669
Model6	YOLOv8	✓	✓		80.6	79.8	76.5	86.2	13.4	539
Model7	YOLOv8	✓		✓	80.6	81.4	84.3	78.9	15.0	524
Model8	YOLOv8	✓	✓	✓	81.6	80.9	80.5	88.2	13.1	592

After 200 cycles, we retrieved the position loss values for the DKR-YOLOv8 and YOLOv8 models in order to do a more detailed analysis of the model's performance before to and after the upgrade. This comparative analysis is visually represented in Figure 8. By examining these position loss values, we can gain insights into the efficiency and accuracy improvements brought about by the upgraded DKR-YOLOv8 model. The data illustrate how the enhancements in the DKR-YOLOv8 model contribute to more precise object localization, ultimately leading to better overall detection performance.

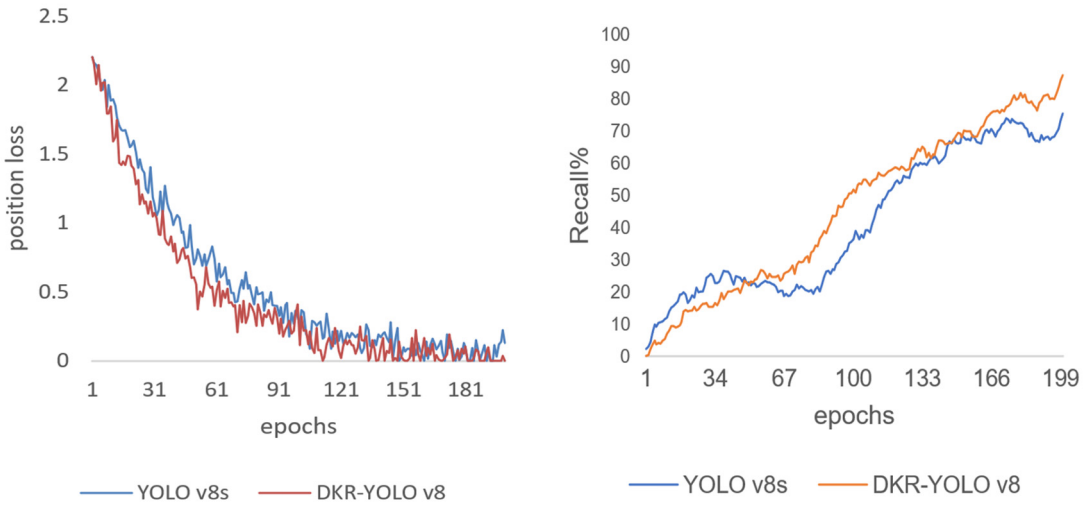


Figure 9. Curve of position loss rates and recall rates.

As shown in Figure 8, the improved DKR-YOLOv8 model exhibits faster convergence and a significantly higher convergence rate compared to the original YOLOv8 model. This improvement can be attributed to the integration of the Kernel Warehouse and Dynamic Snake Convolution techniques. Figure 9 presents the recall curves for the models, illustrating a steady increase in recall as the number of training iterations increases. Notably, after the 200th epoch, a significant difference is observed between the revised DKR-YOLOv8 model and the original YOLOv8 model.

In terms of clustering performance, the DKR-YOLOv8 model outperforms YOLOv8, demonstrating superior clustering capability. The clustering density is significantly higher, with 82.6% of the variables clustered within the range of -0.4 to 0.3. Figure 8 further illustrates a more robust clustering function, which enhances the model's ability to effectively group variables. This tighter clustering contributes to more accurate object detection, as the model is better able to distinguish between different objects within an image.

 doi:10.20944/preprints202508.0180.v1

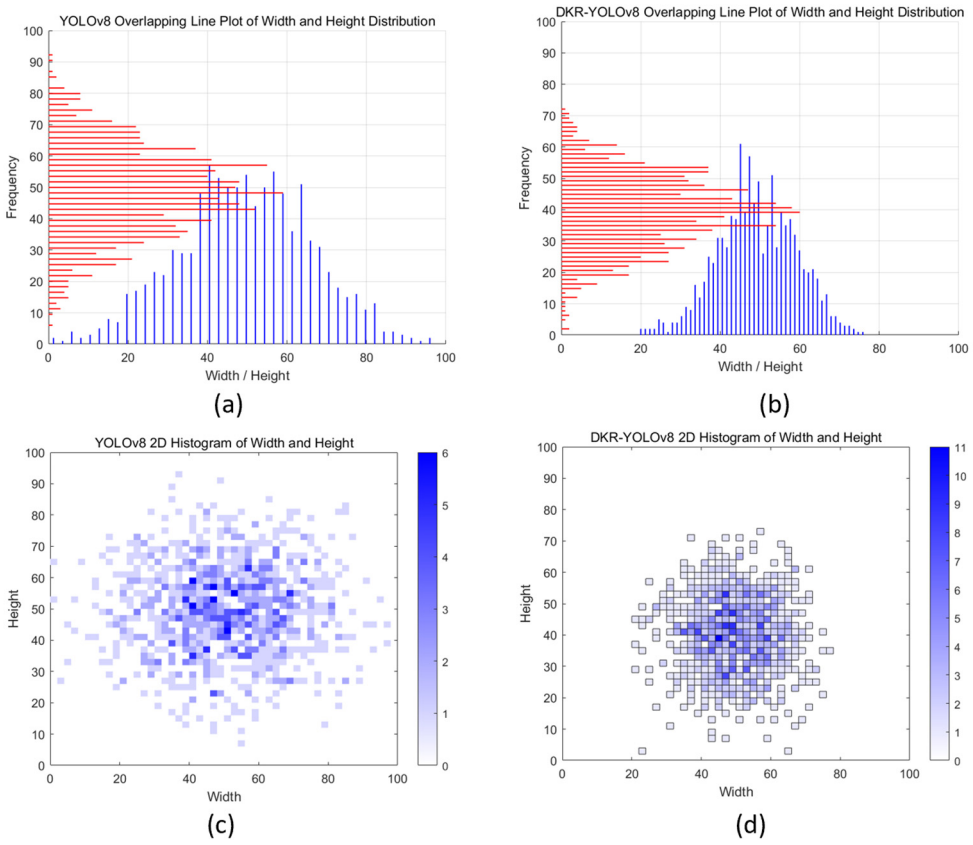


Figure 10. Academic Expansion and Analysis of DKR-YOLO v8 Compared to YOLO v8.

The 2D histogram of data extraction frequency for DKR-YOLO v8 exhibits a more cohesive and uniform distribution. The frequency counts predominantly fall within the range of 20-60, suggesting a balanced and consistent extraction process. This uniformity indicates that the model performs data extraction more evenly across different frames, reducing the likelihood of bias or overfitting to specific data segments. DKR-YOLO v8 shows a closer relationship across various angular directions, which is depicted by the near-circular distribution in the 2D histogram. This close-to-circular pattern signifies a more isotropic clustering of data points, meaning that the model's performance and accuracy are less dependent on the orientation of the objects within the frames. Such isotropy is advantageous as it ensures the model's robustness and consistency in detecting objects regardless of their orientation.

Confusion matrix is a two-dimensional matrix where rows represent the predicted classes by the model and columns represent the actual classes. Each row's data denotes the instances predicted as a particular class, with the total number in each row representing the number of instances predicted to belong to that class. Similarly, each column's data represents the actual classes, with the total number in each column indicating the number of instances belonging to that class. Figure 12 illustrates the confusion matrix of recognition results among different mural types in the test dataset. This matrix displays both proportional and quantitative information. Compared to YOLOv8, the clustering degree of the confusion matrix of DKR-YOLOv8 is significantly higher, the accuracy of DKR-YOLOv8 is about 94%, while YOLOv8 is only 85%. The improved model effectively reduces the probability of missed detections. Through the analysis of the confidence map, it is evident that the attention mechanism significantly enhances the model's detection accuracy, thereby reducing uncertainty and false detection rates. The analysis results of the confidence map show that the introduction of the attention mechanism significantly improves the model's ability to recognize different mural categories, reducing instances of false detections and missed detections, and thereby increasing overall detection accuracy. These improvements demonstrate that the attention

 doi:10.20944/preprints202508.0180.v1

mechanism plays a crucial role in optimizing the performance of the DKR-YOLOv8 model, making it exhibit higher reliability and accuracy in mural detection tasks within complex backgrounds.

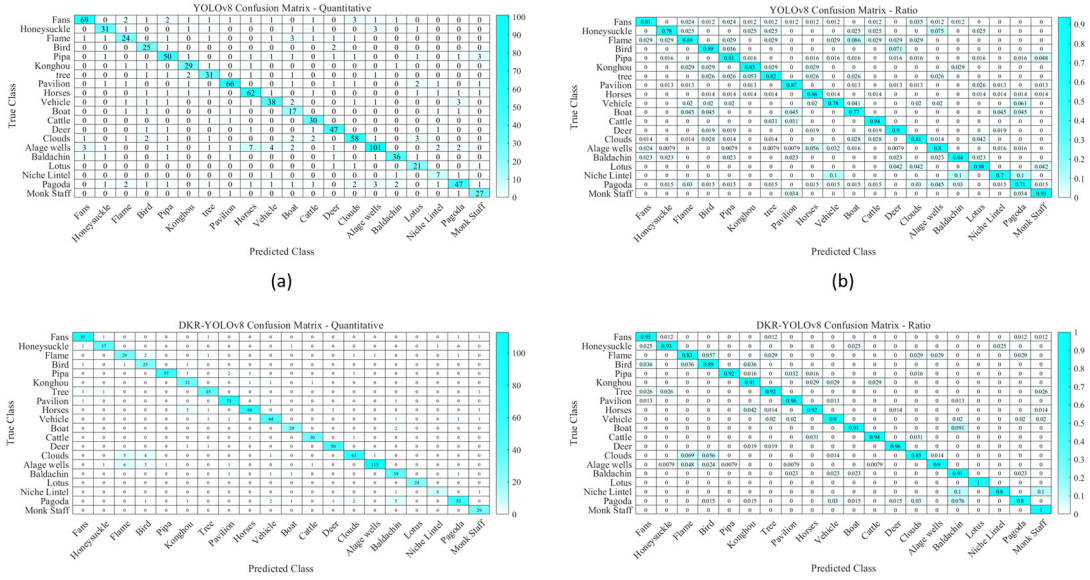


Figure 12. provides a comparison of the quantitative information and ratio information in the confusion matrix used for recognizing different types of mural elements across different models (a) Confusion matrix count information for YOLOv8. (b) Confusion matrix ratio information for YOLOv8. (c) Confusion matrix count information for DKR-YOLOv8. (d) Confusion matrix ratio information for DKR-YOLOv8.

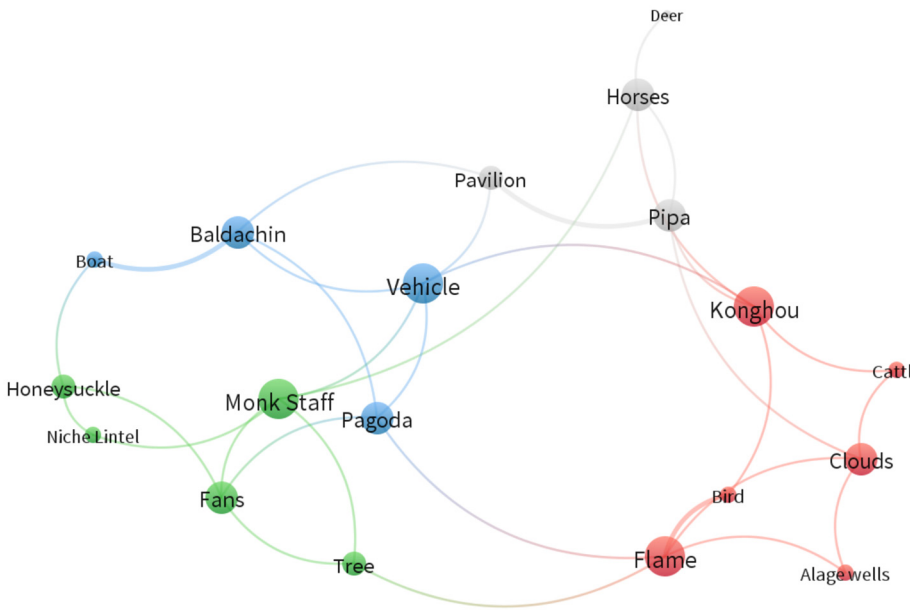


Figure 13. Map of misdetected co-occurrences in image recognition.

In image recognition, object categories are often prone to confusion, either between different categories or within the same category. Below is an analysis of the factors contributing to such misclassification for each category:

Category 1: Konghou, Cattle, Clouds, Bird, Flame, Algae Wells

Natural elements such as clouds, flames, and birds exhibit highly variable shapes, making it challenging for image recognition systems to capture clear boundaries and forms. This variability can

 doi:10.20944/preprints202508.0180.v1

lead to confusion with objects that share similar morphological features, such as flowing water or smoke. Decorative or artistic forms, including algae wells and the konghou, often feature intricate textures or curves that resemble certain natural forms (e.g., plants or ripples), increasing the likelihood of misclassification. Similarly, the textures of cattle skin or bird feathers can be mistaken for those of other furred or feathered animals due to insufficient differentiation in surface details.

Category 2: Deer, Horses, Pavilion, Pipa

Deer and horses share morphological similarities, particularly when viewed from specific angles or in low-quality images, often resulting in misidentification. Architectural structures like pavilions may be confused with similar structures (e.g., temples or covered bridges), especially if the algorithm lacks sensitivity to finer details. Likewise, the pipa may be misclassified as other string instruments, such as the guqin or yueqin, when their distinctive features are not adequately captured during recognition.

Category 3: Baldachin, Pagoda, Boat

Baldachins and pagodas frequently include multi-tiered structures or ornamented spires, which may cause them to be misclassified as other religious buildings (e.g., stupas or temple roofs). Their repetitive decorative elements often bear high visual similarity to other royal or religious architectural designs, complicating the recognition of their specific functions or symbolic meanings. Similarly, the diverse forms of boats can lead to confusion with other objects of comparable shapes, such as bridge components or lower sections of buildings.

Category 4: Monk Staff, Honeysuckle, Niche Lintel, Fans, Tree

Trees are often difficult to distinguish from other plants or natural objects, such as shrubs or vines, particularly in images with blurred details. The monk staff may resemble plant forms like bamboo, especially when the material is not prominently displayed. Decorative and functional overlap further contributes to confusion; for example, fans may be misclassified as other similarly shaped objects, such as folding fans or decorative screens, when shared patterns or structures are present. Likewise, niche lintels can be mistaken for other architectural elements, such as door lintels or window frames, due to their similar shapes and designs.

In image recognition, the primary reasons for object confusion include morphological similarity, surface texture complexity, the interplay of functional and decorative elements, and the visual similarity between natural and artificial objects. These factors can lead to misrecognition or classification errors if the algorithm fails to adequately capture the key features of the objects.

As shown in Figure 16, the loss functions and accuracy metrics during the model training process exhibit their respective convergence trends and oscillatory behaviors. The loss functions measure the discrepancy between the model's predictions and the actual labels. All three loss functions (Box Loss, Objectness Loss, and Classification Loss) demonstrate a clear convergence trend, eventually stabilizing. This indicates that the model progressively learns the features during training and effectively reduces the prediction error.

The Box Loss function is used to measure the difference between the predicted bounding boxes and the ground truth bounding boxes. The Objectness Loss function assesses the model's accuracy in predicting the presence of an object. The Classification Loss function evaluates the model's accuracy in predicting the class of the object. The model's accuracy exhibits noticeable oscillations during the training process. This oscillatory behavior could be due to the model continuously adjusting its parameters to achieve better performance on a complex dataset. However, as the training progresses, the accuracy gradually stabilizes, and around 160 epochs, it stabilizes at approximately 0.8. This indicates that the model has achieved high accuracy in the classification task. All three loss functions converging to a stable state indicates that the model has achieved the expected performance in various tasks. Despite the oscillations observed in the accuracy during training, it eventually stabilizes and remains at a high value. The introduction of the attention mechanism has significantly enhanced the model's detection accuracy, reduced uncertainty and false detection rates, and effectively improved the overall performance of the model.

 doi:10.20944/preprints202508.0180.v1

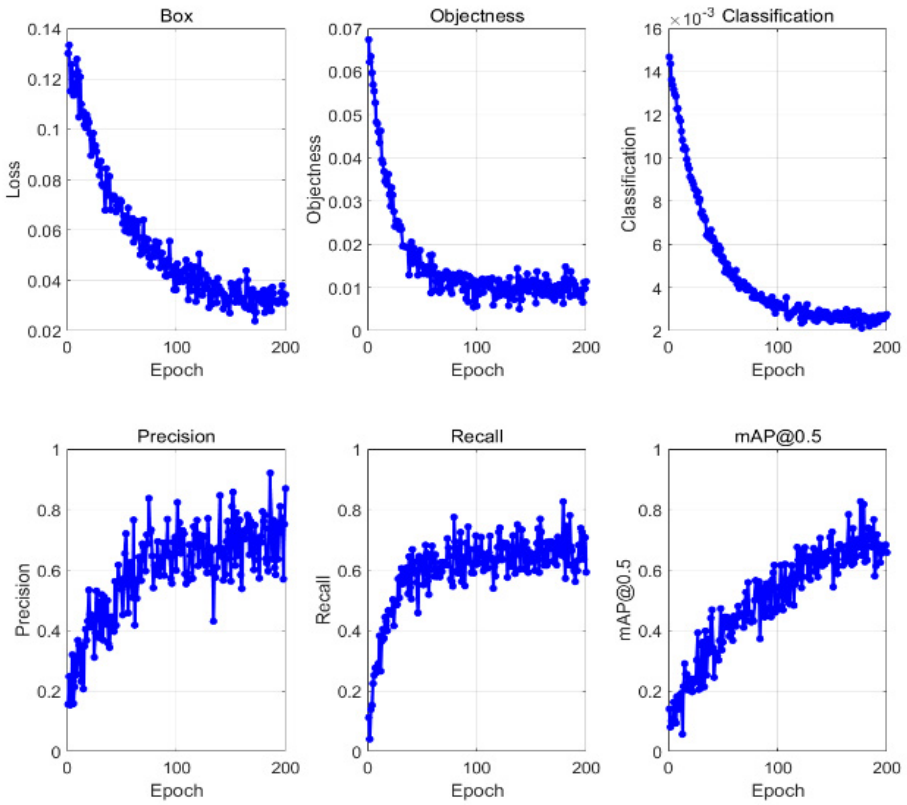


Figure 16. Result on dataset. The loss function, precision, recall, and map evaluation metrics are included in the result picture.

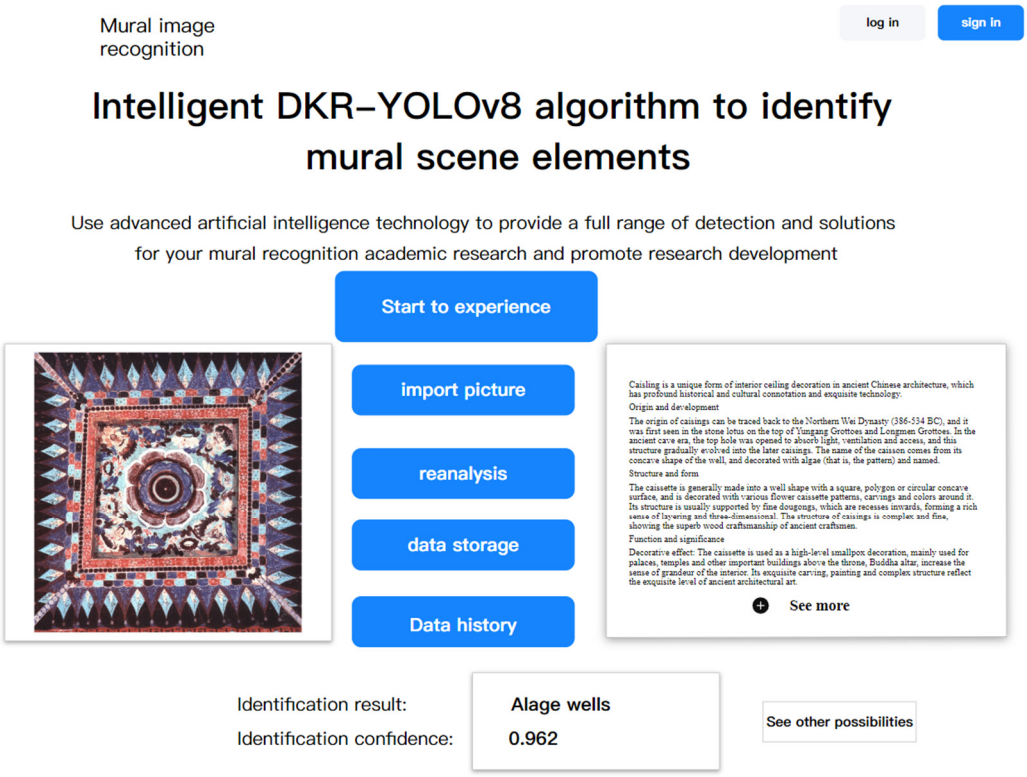


Figure 16. Mural recognition software page.

 doi:10.20944/preprints202508.0180.v1

The core function of the software is the automatic recognition of the elements in the mural. Through high-precision image recognition technology, the software can accurately capture and analyze the key information in the mural. To provide strong technical support for scholars in the field of mural research. Scholars can use the software to quickly obtain key data in murals, so as to more deeply study the history, culture, artistic style and other aspects of murals. The software has a user-friendly interface design, including functional modules such as starting the experience, importing pictures, analyzing data and viewing history. This makes it easy for users to get started and quickly complete the recognition and analysis tasks of mural elements. The development and application of the software involves many disciplines such as computer science, image processing, culture and art. Its successful application will promote the cross integration and deep cooperation between these disciplines, and promote the common development of related fields. The application of the software is helpful to improve the level of mural protection and research. By accurately identifying the elements and features in the murals, the preservation status of the murals can be more accurately evaluated, which provides strong support for the development of scientific conservation programs and research plans.

4. Conclusions

In conclusion, we supplied an antique mural picture model based on YOLOv8 architecture. Our proposed technique has shown to be successful in correctly and efficiently detecting mural pictures in a variety of environmental conditions, as proven by extensive testing and meticulous analysis.

4.1. Research Result

The key enhancements adopted to reach this purpose are the following:

- The creation of a large dataset comprising 20 diverse types of mural images has been a significant milestone in our study. This dataset serves as a robust foundation for validating the effectiveness of our proposed technique in environmental monitoring applications. By encompassing a wide variety of mural images, the dataset ensures that our algorithm can adapt to the diverse visual qualities and challenges present in real-world scenarios. This extensive collection not only enhances the reliability of our technique but also demonstrates its potential for application in specialized environmental surveillance systems, enabling more accurate and efficient mural monitoring across various contexts.
- The identification and analysis of mural images, particularly those from Dunhuang, present substantial challenges due to the diverse nature of the samples, intricate backgrounds, and the limited recognition accuracy of current YOLO detection techniques. To address these obstacles, we propose the DKR-YOLOv8 model, which integrates Kernel Warehouse dynamic convolution and Dynamic Snake Convolution. These enhancements collectively improve the feature extraction network by capturing finer and more nuanced image characteristics. Additionally, the model incorporates a Lightweight Residual Feature Pyramid Network, significantly boosting detection accuracy and operational efficiency, particularly in dry grassland environments. These innovations enable the DKR-YOLOv8 model to achieve superior performance in recognizing and classifying mural images, surpassing the limitations of previous YOLO models.
- The enhanced DKR-YOLOv8 model delivers outstanding performance in mural image recognition, achieving a precision rate of 81.6%, recall rate of 80.9%, F1 score of 80.5%, mean average precision (mAP) of 88.2%, 13.1 GFLOPs, and a processing speed of 592 frames per second (FPS). Compared to other YOLO models, the DKR-YOLOv8 model excels in both accuracy and speed, making it an ideal choice for mobile applications targeting mural analysis in environmental settings.

The integration of Kernel Warehouse, Feature Pyramid Network, and Snake Convolution within the YOLOv8 framework for ancient mural element recognition and classification offers significant practical implications for the field of mural analysis. This advanced computational approach enables more precise and efficient identification of diverse elements within historical murals, including

 doi:10.20944/preprints202508.0180.v1

figures, objects, and architectural details. By automating the recognition process, researchers can save considerable time and resources that would otherwise be spent on manual identification and documentation. Moreover, the enhanced precision provided by the YOLOv8-based algorithm ensures more accurate and consistent interpretations of mural elements. This capability is particularly valuable for comparative research, where identifying similar or identical features across murals can yield valuable insights into the cultural, religious, or historical contexts of the artworks. Additionally, the algorithm's ability to classify objects based on their attributes contributes to a deeper understanding of the artistic styles and techniques employed in ancient mural paintings. The practical implications of this research extend to the preservation and restoration of historical murals. By accurately recognizing and classifying mural elements, conservationists can assess the condition of the artworks more effectively and prioritize areas requiring restoration. This targeted approach to preservation ensures the longevity of these invaluable cultural heritages.

In conclusion, the implementation of Kernel Warehouse, Feature Pyramid Network, and Snake Convolution within the YOLOv8 framework provides a powerful tool for scholars, historians, and conservationists. By expediting the identification and analysis processes, this algorithmic technique has the potential to significantly enhance our understanding of historical murals and contribute to their preservation for future generations.

4.2. Research Prospects

This paper has conducted extensive research on the digital recognition of mural paintings. While some progress has been made in the study of ancient murals, it is evident that there are still numerous challenges to overcome. Due to existing issues such as the diversity and inconsistency of mural image standards, the algorithm presented in this paper has certain limitations. To truly unlock the potential of ancient mural research and address these complexities, further work is imperative in several key areas:

- Enhancing the Mural Database: The classification of images within the mural database faces challenges such as disputes over classification and a scarcity of comprehensive resources. These issues largely stem from the lack of large, publicly available mural image datasets. Furthermore, mural images are often subject to strict protection by local authorities, making their collection particularly difficult. Many mural images are concentrated in specific regions, sharing similar styles and content. However, their uneven distribution raises concerns about the reliability and representativeness of the collected data. To address these challenges, future research must include extensive fieldwork, close collaboration with local authorities, and the use of diverse references. The aim is to create a comprehensive, balanced, and extensive mural image dataset that accurately reflects the richness and diversity of ancient mural art.
- Addressing Data Imbalance in Neural Architecture Search: Data imbalance is a significant issue when employing neural architecture search algorithms for classification. This challenge often arises due to limitations in time and manpower, leading to datasets that are unevenly distributed across categories. To ensure effective and reliable classification, it is essential to construct a balanced dataset for ancient mural classification. This involves not only increasing the overall size of the dataset but also ensuring that each category of murals is adequately represented. A balanced dataset will enable more effective training of the algorithm, resulting in improved recognition accuracy and a deeper understanding of ancient mural art.
- Improving Recognition Accuracy and Handling Controversial Images: While significant improvements have been achieved in recognition accuracy for mural classification, challenges remain, particularly when addressing controversial mural images. These images often pose difficulties for the algorithm, leading to less satisfactory performance. To overcome this, future research should prioritize refining and enhancing the algorithm's performance. This includes developing advanced feature extraction techniques capable of accurately capturing and analyzing the intricate characteristics of mural images, such as their content, style, and historical

 doi:10.20944/preprints202508.0180.v1

context. By focusing on these advancements, the algorithm will be better equipped to handle controversial images, resulting in more reliable and accurate classifications.

In conclusion, while significant progress has been made in the digital recognition of ancient murals, there is still much work to be done. By enhancing the mural database, addressing data imbalance, and improving recognition accuracy, we can unlock the full potential of this research and gain a deeper understanding of the rich history and artistry of ancient mural paintings.

References

Abo Jouhain, A. (2024). *Optimization of Deep Learning Techniques for Real-Time Detection and Tracking of Marine Species utilizing YOLOv8 and Deep SORT Algorithms* UIS].

Aboyomi, D. D., & Daniel, C. (2023). A Comparative Analysis of Modern Object Detection Algorithms: YOLO vs. SSD vs. Faster R-CNN. *ITEJ (Information Technology Engineering Journals)*, *8*(2), 96-106.

Ahsan, F., Dana, N. H., Sarker, S. K., Li, L., Muyeen, S., Ali, M. F., Tasneem, Z., Hasan, M. M., Abhi, S. H., & Islam, M. R. (2023). Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review. *Protection and Control of Modern Power Systems*, *8*(3), 1-42.

Ali, M., & Zhang, Z. (2024). The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection. Computers 2024, 13, 336. In.

Ali, S. G., Wang, X., Li, P., Li, H., Yang, P., Jung, Y., Qin, J., Kim, J., & Sheng, B. (2024). Egdnet: an efficient glomerular detection network for multiple anomalous pathological feature in glomerulonephritis. *The Visual Computer*, 1-18.

Amura, A., Aldini, A., Pagnotta, S., Salerno, E., Tonazzini, A., & Triolo, P. (2021). Analysis of diagnostic images of artworks and feature extraction: design of a methodology. *Journal of Imaging*, *7*(3), 53.

Bhalla, S., Kumar, A., & Kushwaha, R. (2024). Feature-adaptive FPN with multiscale context integration for underwater object detection. *Earth Science Informatics*, *17*(6), 5923-5939.

Bolong, C., Zongren, Y., Manli, S., Zhongwei, S., Jinli, Z., Biwen, S., Zhuo, W., Yaopeng, Y., & Bomin, S. (2022). Virtual reconstruction of the painting process and original colors of a color-changed Northern Wei Dynasty mural in Cave 254 of the Mogao Grottoes. *Heritage Science*, *10*(1), 164.

Cederin, L., & Bremberg, U. (2023). Automatic object detection and tracking for eye-tracking analysis. In.

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *18*(2), 1-22.

Chappidi, J., & Sundaram, D. M. (2024). Novel Animal Detection System: Cascaded YOLOv8 with Adaptive Preprocessing and Feature Extraction. *Ieee Access*.

Deng, X., & Yu, Y. (2023). Ancient mural inpainting via structure information guided two-branch model. *Heritage Science*, *11*(1), 131.

Du, R. (2024). The cultural heritage of secular music and dance in the Mogao Caves of Dunhuang: AN ARTS EDUCATION PERSPECTIVE. *Arts Educa*, *38*.

Duan, W. (2006). *The Complete collection of Chinese Dunhuang murals*. The editorial board of the Complete Collection of Dunhuang Murals.

Dunkerley, D. (2023). Leaf water shedding: Moving away from assessments based on static contact angles, and a new device for observing dynamic droplet roll-off behaviour. *Methods in Ecology and Evolution*, *14*(12), 3047-3054.

Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., & Dai, B. (2023). Generative diffusion prior for unified image restoration and enhancement. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Huang, L., Wang, W., Wu, Z.-F., Dou, H., Shi, Y., Feng, Y., Liang, C., Liu, Y., & Zhou, J. (2024). Group diffusion transformers are unsupervised multitask learners.

Jaworek-Korjakowska, J., Yap, M. H., Bhattacharjee, D., Kleczek, P., Brodzicki, A., & Gorgon, M. (2023). Deep neural networks and advanced computer vision algorithms in the early diagnosis of skin diseases. In *State of the art in neural networks and their applications* (pp. 47-81). Elsevier.

 doi:10.20944/preprints202508.0180.v1

Kim, T., Zhou, X., & Pendyala, R. M. (2022). Computational graph-based framework for integrating econometric models and machine learning algorithms in emerging data-driven analytical environments. *Transportmetrica A: Transport Science*, *18*(3), 1346-1375.

Koo, B., Choi, H.-S., & Kang, M. (2021). Simple feature pyramid network for weakly supervised object localization using multi-scale information. *Multidimensional Systems and Signal Processing*, *32*(4), 1185-1197.

Li, Q., Wang, P., Liu, Z., Zhang, H., Song, Y., & Zhang, Y. (2024). Using scaffolding theory in serious games to enhance traditional Chinese murals culture learning. *Computer Animation and Virtual Worlds*, *35*(1), e2213.

Li, W., Lv, H., Liu, Y., Chen, S., & Shi, W. (2023). An investigating on the ritual elements influencing factor of decorative art: based on Guangdong's ancestral hall architectural murals text mining. *Heritage Science*, *11*(1).

Lian, Y., & Xie, J. (2024). The evolution of digital cultural heritage research: Identifying key trends, hotspots, and challenges through bibliometric analysis. *Sustainability*, *16*(16), 7125.

Lin, X., Sun, S., Huang, W., Sheng, B., Li, P., & Feng, D. D. (2021). EAPT: efficient attention pyramid transformer for image processing. *IEEE Transactions on Multimedia*, *25*, 50-61.

Liu, S., Yang, J., Agaian, S. S., & Yuan, C. (2021). Novel features for art movement classification of portrait paintings. *Image and Vision Computing*, *108*, 104121.

Liu, Y., Chen, W., Bai, Y., Liang, X., Li, G., Gao, W., & Lin, L. (2024). Aligning cyber space with physical world: A comprehensive survey on embodied ai. *arXiv preprint arXiv:2407.06886*.

Mei, S., Li, X., Liu, X., Cai, H., & Du, Q. (2021). Hyperspectral image classification using attention-based bidirectional long short-term memory network. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-12.

Menai, B. L. (2023). *Recognizing the artistic style of fine art paintings with deep learning for an augmented reality application* Université Mohamed Khider (Biskra-Algérie)].

Mu, R., Nie, Y., Cao, K., You, R., Wei, Y., & Tong, X. (2024). Pilgrimage to Pureland: Art, Perception and the Wutai Mural VR Reconstruction. *International Journal of Human–Computer Interaction*, *40*(8), 2002-2018.

Nazir, A., Cheema, M. N., Sheng, B., Li, P., Li, H., Xue, G., Qin, J., Kim, J., & Feng, D. D. (2021). Ecsu-net: an embedded clustering sliced u-net coupled with fusing strategy for efficient intervertebral disc segmentation and classification. *IEEE Transactions on Image Processing*, *31*, 880-893.

Petracek, P., Kratky, V., Baca, T., Petrlik, M., & Saska, M. (2023). New era in cultural heritage preservation: Cooperative aerial autonomy for fast digitalization of difficult-to-access interiors of historical monuments. *IEEE Robotics & Automation Magazine*.

Ren, H., Sun, K., Zhao, F., & Zhu, X. (2024). Dunhuang murals image restoration method based on generative adversarial network. *Heritage Science*, *12*(1), 39.

Sakiba, C., Tarannum, S. M., Nur, F., Arpan, F. F., & Anzum, A. A. (2023). *Real-time crime detection using convolutional LSTM and YOLOv7* Brac University].

Shao, W., Rajapaksha, P., Wei, Y., Li, D., Crespi, N., & Luo, Z. (2023). COVAD: Content-oriented video anomaly detection using a self attention-based deep learning model. *Virtual Reality & Intelligent Hardware*, *5*(1), 24-41.

Song, S. (2023). New Era for Dunhuang Culture Unleashed by Digital Technology. *International Core Journal of Engineering*, *9*(10), 1-14.

Tekli, J. (2022). An overview of cluster-based image search result organization: background, techniques, and ongoing challenges. *Knowledge and Information Systems*, *64*(3), 589-642.

Tian, Z., Huang, J., Yang, Y., & Nie, W. (2023). KCFS-YOLOv5: A high-precision detection method for object detection in aerial remote sensing images. *Applied Sciences*, *13*(1), 649.

Veysi, H. (2022). Megatsunamis and microbial life on early Mars. *International Journal of Astrobiology*, *21*(3), 188-196.

Wang, X., Tan, X., Gui, H., & Song, N. (2021). A semantic enrichment approach to linking and enhancing Dunhuang cultural heritage data. In *Information and Knowledge Organisation in Digital Humanities* (pp. 87-105). Routledge.

Wang, X., Zhao, K., Zhang, Q., & Liu, C. (2024). Digital deduction theatre: An experimental methodological framework for the digital intelligence revitalisation of cultural heritage. In *Intelligent Computing for Cultural Heritage* (pp. 203-220). Routledge.

 doi:10.20944/preprints202508.0180.v1

Xiao, H., Zheng, H., & Meng, Q. (2023). Research on Deep Learning-Driven High-Resolution Image Restoration for Murals From the Perspective of Vision Sensing. *Ieee Access*.

Yu, Y., Qian, J., Wang, C., Dong, Y., & Liu, B. (2024). Animation line art colorization based on the optical flow method. *Computer Animation and Virtual Worlds*, *35*(1), e2229.

Zeng, Z., Sun, S., Li, T., Yin, J., & Shen, Y. (2022). Mobile visual search model for Dunhuang murals in the smart library. *Library Hi Tech*, *40*(6), 1796-1818.

Zeng, Z., Sun, S., Li, T., Yin, J., Shen, Y., & Huang, Q. (2024). Exploring the topic evolution of Dunhuang murals through image classification. *Journal of Information Science*, *50*(1), 35-52.

Zhang, B. (2024). *Enhanced Safety of Autonomous Driving in Real-World Adverse Weather conditions* via *Deep Learning-Based Object Detection* Université d'Ottawa| University of Ottawa].

Zhang, J., Zhang, X., Huang, Z., Cheng, X., Feng, J., & Jiao, L. (2023). Bidirectional multiple object tracking based on trajectory criteria in satellite videos. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-14.

Zhang, M., & Tian, X. (2023). Transformer architecture based on mutual attention for image-anomaly detection. *Virtual Reality & Intelligent Hardware*, *5*(1), 57-67.

Zhang, X. (2023). The Dunhuang Caves: Showcasing the Artistic Development and Social Interactions of Chinese Buddhism between the 4th and the 14th Centuries. *Journal of Education, Humanities and Social Sciences*, *21*, 266-279.

Zhao, M., Agarwal, N., Basant, A., Gedik, B., Pan, S., Ozdal, M., Komuravelli, R., Pan, J., Bao, T., & Lu, H. (2022). Understanding data storage and ingestion for large-scale deep recommendation model training: Industrial product. Proceedings of the 49th annual international symposium on computer architecture,

Zheng, J., Fu, Y., Zhao, R., Lu, J., & Liu, S. (2024). Dead Fish Detection Model Based on DD-IYOLOv8. *Fishes*, *9*(9), 356.

Zhong, Q., Liu, Y., Ao, X., Hu, B., Feng, J., Tang, J., & He, Q. (2020). Financial defaulter detection on online credit payment via multi-view attributed heterogeneous information network. Proceedings of the web conference 2020,

Zhu, W., Du, X., Lyu, K., Liu, Z., Ren, S., Xiao, S., & Seong, D. (2023). Big Data in Art History: Exploring the Evolution of Dunhuang Artistic Style Through Archaeological Evidence. *Mediterranean Archaeology and Archaeometry*, *23*(3), 87-106.