PeerJ Computer Science

# Comparative evaluation of machine learning models for museum exhibit recognition from video-derived datasets

Madina Ipalakova, Zhiger Bolatov, Yevgeniya Daineko, Regina Sharshova, Kamila Abdugapparova and Dana Tsoy

International Information Technology University, Almaty, Kazakhstan

## ABSTRACT

This study evaluates the performance of multiple deep learning models for automatic recognition of museum artifacts using image frames extracted from real-world video footage. A comparative analysis is conducted across eight state-of-the-art architectures—MobileNetV3, ResNetV2, EfficientNetV2, You Only Look Once v8 (YOLOv8), Visual Geometry Group 16 (VGG16), ConvNeXtTiny, SwinV2-Base, and Dual Attention Vision Transformer (DaViT)—on a custom dataset collected in an actual museum environment. The dataset comprises labeled video frames categorized by artifact type and is used to train and test models for both classification and object detection tasks. Results indicate that YOLOv8, MobileNetV3, and DaViT achieve superior performance for real-time mobile and augmented reality (AR) applications, while ResNetV2 and SwinV2-Base provide high classification accuracy suitable for archival and cataloging systems. This work offers practical guidance on dataset design, model choice, and deployment strategies for artificial intelligence (AI)-powered cultural heritage technologies.

**Subjects** Artificial Intelligence, Computer Vision, Data Mining and Machine Learning, Neural Networks

**Keywords** Machine learning, Museum exhibit recognition, Video-derived datasets, Object detection, Cultural heritage, Real-time recognition, Augmented reality (AR)

Submitted 13 June 2025
Accepted 20 August 2025
Published 2 October 2025

Corresponding author
Madina Ipalakova,
m.ipalakova@iitu.edu.kz

Academic editor
Siddhartha Bhattacharyya

Additional Information and Declarations can be found on page 15

DOI **10.7717/peerj-cs.3207**



## INTRODUCTION

The increasing application of artificial intelligence in cultural heritage has significantly transformed how museums recognize and present exhibits. Machine learning-based object recognition systems provide promising solutions for real-time identification in digital museums, promoting more interactive and engaging visitor experiences (*Kiourexidou & Stamou, 2025*; *Meyer et al., 2024*). Unlike traditional static and labor-intensive approaches, modern computer vision techniques—especially those leveraging deep learning—enable dynamic, scalable, and precise artifact recognition. However, despite advancements in convolutional neural networks (CNNs) and single-stage detectors, comparative studies evaluating various architectures under realistic museum conditions, such as inconsistent lighting, object occlusion, and varying perspectives, remain limited (*Khan et al., 2022*; *Patil, Sharma & Jain, 2024*). Therefore, selecting appropriate architectures is critical for developing successful museum applications, given the continuous evolution of deep learning models across diverse fields such as healthcare, security, and education.

**How to cite this article** Ipalakova M, Bolatov Z, Daineko Y, Sharshova R, Abdugapparova K, Tsoy D. 2025. Comparative evaluation of machine learning models for museum exhibit recognition from video-derived datasets. **PeerJ Comput. Sci. 11:e3207 DOI 10.7717/peerj-cs.3207**

Integrating machine learning (ML) technologies into museum exhibitions represents a significant step toward the digital transformation of cultural heritage, offering multiple opportunities ranging from enhancing artifact classification accuracy to delivering personalized visitor interactions. *Luo & Li (2024)* illustrate the development of an intelligent real-time exhibit recognition system based on deep learning, significantly improving visitor engagement. Further, hybrid approaches, as highlighted by *Bobasheva, Gandon & Precioso (2022)*, demonstrate that combining symbolic artificial intelligence (AI) with ML substantially enhances cultural metadata quality, enabling more precise navigation through museum collections.

The effectiveness of ML models in museum contexts heavily depends on the quality and diversity of training data, particularly for visual recognition tasks. Data augmentation techniques have thus become essential for expanding datasets and improving model generalizability. Specifically, simulating variations in images captured by smartphones—such as adjustments in color profiles, sharpness, noise, dynamic range, and lens distortion—helps overcome the limitations associated with uniform data sources. *Lee (2025)* emphasizes structured data preparation and dataset diversity as critical components for successful ML implementation in cultural heritage. *Wang & Zhao (2023)* also confirm that improved data quality directly impacts museum collection management and exhibition optimization.

The broader potential of intelligent museum systems extends to spatial exhibition design, visitor behavior analysis, and ethical considerations related to inclusive representation (*Cai, Zhang & Pan, 2023*; *Tang et al., 2024*; *Acosta et al., 2021*; *Walsh et al., 2021*; *Huang & Liem, 2022*). *Li (2024)* outlines opportunities and challenges in implementing AI in museum development, highlighting the need for adaptive and high-performing models to enhance visitor experiences. Similarly, *Bazarbekov et al. (2024)* emphasize the successful application of AI-powered image analysis across medical and industrial domains, underscoring the importance of adaptive visual modeling. *Rahim et al. (2025)* further demonstrate the effectiveness of pretrained convolutional networks, reinforcing the value of robust datasets and augmentation strategies in successful ML applications.

These advances in deep learning are already evident across various domains, including medicine, cybersecurity, and autonomous systems. For example, the modified U-Net architecture (D-UNet), combining 2D and 3D features, significantly improved segmentation accuracy for stroke lesions and retinal blood vessels by leveraging dimension fusion and deformable convolutions (*Zhou et al., 2021*; *Jin et al., 2019*). Dense-UNet, another variation, effectively segmented cellular images under noisy conditions by enhancing information flow through dense connections, mitigating gradient issues (*Cai et al., 2020*). Such innovations highlight the extensive potential of deep learning architectures to transform museum exhibit recognition and interaction profoundly.

In cybersecurity, the X-NET architecture, grounded in explainable AI, achieved high accuracy in detecting network threats and offered interpretable results, although it can be computationally demanding (*Patel et al., 2024*; *Li et al., 2024*). X-NET's adaptations have also successfully segmented video scenes with dynamic backgrounds (*Zhang et al., 2019*).

CNN-based methods, like U-Net, similarly demonstrated effectiveness in seismic data reconstruction from sparse samples (*Huang & Nowack, 2020*).

Other deep-learning architectures such as SegNet, PSPNet, and V-Net showed effectiveness across various applications, including road infrastructure segmentation, medical diagnostics, and remote sensing. SegNet efficiently segmented road scenes and skin lesions but exhibited limitations in resolution and accuracy compared to specialized models (*Shabalina et al., 2021*; *Sokolov, 2024*; *Zaitseva & Kazankov, 2021*). PSPNet leveraged pyramid spatial pooling to capture contextual information but was computationally intensive, limiting real-time use (*Gorbachev et al., 2020*; *Taran et al., 2018*; *Yuan, Wang & Xu, 2022*). V-Net excelled in volumetric medical imaging and road network extraction, proving its effectiveness in complex 3D environments (*Milletarì, Navab & Ahmadi, 2016*; *Abdollahi, Pradhan & Alamri, 2020*; *Kato & Hotta, 2020*).

The DeepLab v3+ model, tested on TS1 (Plant Village) and TS2 (real vineyard) datasets, demonstrated improved mean intersection over union (mIOU), recall, and F1-score metrics for grapevine black rot spot segmentation, providing an effective method for assessing disease severity and potentially applicable to other plant diseases (*Yuan et al., 2022*).

Additional methods, such as meta pseudo labels and NoisyNN architectures, advanced the performance of deep-learning models by improving training efficiency and robustness to noise (*Pham et al., 2021*; *Go & Moon, 2024*; *Gesmundo & Dean, 2022*; *Gesmundo, 2022*; *Gao et al., 2024*; *Foret et al., 2020*; *Papers with Code, 2025*; *Yu et al., 2023*; *Liu et al., 2023*). Such innovations indicate significant potential for museum-related recognition tasks, yet practical challenges persist, including domain adaptation, limited data, and real-time performance constraints (*Zhang, Tas & Koniusz, 2018*; *Koniusz et al., 2018*; *Ypsilantis et al., 2021*; *Perera et al., 2020*; *Pasqualino et al., 2020*; *Wang & Li, 2022*).

Thus, deep neural network architectures are actively used today for the recognition of museum objects. However, challenges related to adaptation to real-world conditions, limited training data, and the need for real-time performance remain relevant. The research presented in this study complements existing approaches by offering a comparative evaluation of eight models—MobileNetV3, ResNetV2, EfficientNetV2, You Only Look Once v8 (YOLOv8), Visual Geometry Group 16 (VGG16), ConvNeXt-Tiny, SwinV2-Base, and DaViT—on video data collected in a museum setting, with a focus on their use in developing mobile AR applications. These architectures were selected based on their image classification and detection capabilities, computational efficiency, and potential for integration into various digital museum systems. MobileNetV3 was chosen for its compactness and speed, ResNetV2 for its deep architecture and high accuracy, EfficientNetV2 for its balance between computational load and accuracy, YOLOv8 for its fast and precise object detection, and VGG16 as a well-established solution with strong generalization capabilities. ConvNeXt-Tiny, a modern architecture that integrates CNN and transformer principles, was also evaluated, along with SwinV2-Base—a powerful visual transformer model known for its high accuracy and robustness, and DaViT, a hybrid

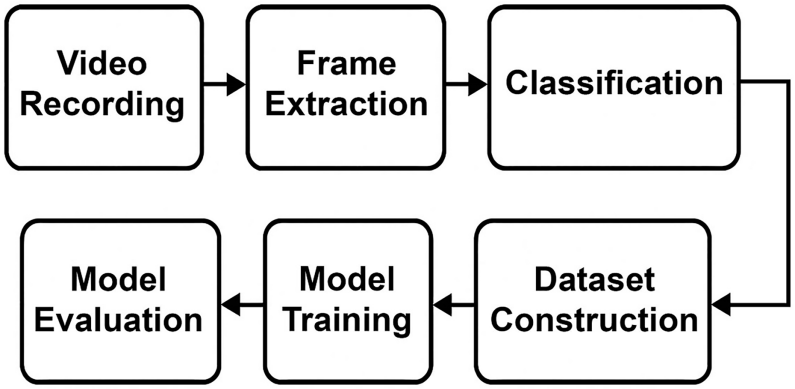


**Figure 1** **The visual depiction of the methodology.** Full-size DOI: 10.7717/peerj-cs.3207/fig-1

model that effectively combines the strengths of convolutional and transformer-based approaches for classifying museum exhibits. By systematically comparing eight neural network architectures on a custom video-derived dataset, this study not only fills a methodological gap in the literature but also provides practical recommendations for deploying AI in real-world museum settings.

## MATERIALS AND METHODS

To prepare the dataset used in this research, a comprehensive methodology involving multiple sequential stages was employed. As illustrated in Fig. 1, the primary steps included: video recording in a museum environment, frame extraction from collected videos, manual classification and labeling, and the subsequent dataset construction and expansion to ensure robust model training and evaluation. Each of these stages is described in detail in the following subsections.

### Video recording

This stage involved recording video footage in the museum using a smartphone camera. The recordings were made with iPhone 12 Pro and Samsung S22 Plus smartphones in .mov format, at Full HD resolution and 60 frames per second. As the authors collaborate with the A. Kasteyev State Museum of Arts in Almaty, the exhibits from this museum were used for the study.

The filming was conducted under natural conditions, without special preparation, simulating the behavior of regular museum visitors. Approximately 80 video clips were collected during this process, which were later categorized and used for analysis.

All collected video materials were categorized into the following categories:

- Sculpture—12 objects, including:
  - “Candlestick. Camel” (N. I. Pavlenko, 1970)
  - “Decorative Sculpture. Ram” (N. I. Pavlenko, 1971)
  - “Sculptural Composition. Caravan” (Yu. G. Popov, V. G. Popov, 1983)

- "Vase" (20th century, unknown author)
- "Kimono" (R. F. Kozhakhmetov, 2012)
- "Petroglyphs" (G. Z. Dosmagambetova, 1998)
- "Cry" (B. Abishev, 1991)
- "Four-Part Composition. Fragment" (R. Akhmetov, 1980)
- "Portrait of Olzhas Suleimenov" (T. Dosmagambetov, 1977)
- "Saule" (A. Yesenbayev, 1990)
- "Three Swords" (A. Zhumabay, 2006)
- "Family" (Kazaryan)

- Tapestry—six objects, including:
  - "Syrmak—Felt Floor Carpet" (unknown author)
  - "Legend of the Mountains and Steppes" (K. T. Tynybekov, 1979)
  - "Wall Panel. Jailau" (B. E. Zaurbekova)
  - "Tree of Life" (R. E. Bazarbayeva, 2016)
  - "Galaxy" (S. S. Bapanova, 2016)
  - "Tükki Kilem—Pile Carpe"t (unknown author, Southern Kazakhstan)

- Painting—18 objects, including:
  - "Herd in the Jailau" (A. Kasteyev, 1947)
  - "Blooming Apple Trees" (A. Kasteyev, 1958)
  - "Portrait of Kenesary" (A. Kasteyev)
  - "Portrait of Folk Akyn Zhambyl" (A. Kasteyev, 1937)
  - "Collective Farm Celebration" (A. Kasteyev)
  - "Turksib" (A. Kasteyev, 1969)
  - "Portrait of Ch. Valikhanov" (A. Kasteyev, 1951)
  - "Collective Farm Dairy Farm" (A. Kasteyev, 1936)
  - "Haymaking" (A. Kasteyev, 1934)
  - "Lake Issyk" (A. Kasteyev, 1953)
  - "Portrait of Amangeldy Imanov" (A. Kasteyev, 1950)
  - "On the High-Mountain Skating Rink" (A. Kasteyev)
  - "Milking Mares" (A. Kasteyev, 1936)
  - "Kapchagay Hydroelectric Station" (A. Kasteyev, 1972)
  - "Happiness" (S. Aitbayev, 1966)
  - "Sounds of the Kobyz" (K. Yesirkeyev, 1973)
  - "Earth and Time. Kazakhstan. Triptych. Harvest Time" (K. V. Mullashev, 1978)
  - "Milking the Red She-Camel" (A. Sydykhanov, 1986–1987)

- Decorative and Applied Arts—nine objects, including:
  - "Decorative Vase. Gemstone" (V. V. Tarasova, 1967)
  - "Saukele" (unknown author, 19th century)
  - "Torsyk—Leather Vessel for Kumis" (G. K. Zhuvaniyazova, 20th century)
  - "Kübi—Churn for Kumis" (K. Malaev, 1982)
  - "Zhaglan and Kebeje" (unknown author)
  - "Dombry" (unknown author)
  - "Er—Men's Saddle" (unknown author)
  - "Women's Saddle" (unknown author)
  - "Vase from the "Patches" Series" (R. F. Kozhakhmetov, 2006)
- Mixed Media—one object, including:
  - "Kobyz" (unknown author, 21st century)
- Jewelry Art—two objects:
  - Jewelry showcase
  - Jewelry items (exhibition display)

Filming was conducted in public areas of the museum without recording any personal data of visitors, ensuring complete anonymity and compliance with privacy regulations. All supplementary information about the exhibits (including descriptions, author attribution, and creation dates) was officially provided by the A. Kasteyev State Museum of Arts solely for the purpose of this research project and is not publicly available.

The use of this data is strictly limited to the scope of the research project and is carried out with the museum's permission. Currently, the dataset is closed. Its publication in open-access repositories (such as Zenodo or Kaggle) is only possible upon formal agreement with the museum and in full compliance with copyright regulations and data distribution requirements.

## Frame extraction

Video recordings were captured using smartphone cameras at 60 frames per second. From these recordings, 2–3 distinct frames per second were extracted. This frequency was chosen to ensure sufficient variation between frames in terms of object pose, lighting, and angle, thereby enriching the training data with diverse visual samples.

## Classification

Each extracted frame was manually reviewed and assigned to a specific exhibit class based on the depicted artifact. The labeling was performed with the help of metadata and descriptions provided by the museum.

## Dataset construction

Initially, a dataset of approximately 60 labeled images per class was created. However, this limited sample size led to insufficient recognition performance (~30% accuracy during in-museum testing). To address this, the dataset was augmented by repeating the same video processing methodology, resulting in an average of 120 images per class. This enhancement allowed the models to generalize better and significantly improved recognition accuracy (up to 96%).

These preprocessing steps ensured the dataset captured the variability of real-world museum conditions (*e.g.*, different lighting, partial occlusion), laying a solid foundation for robust model training and evaluation. The overall training and evaluation procedure is outlined in pseudocode (Algorithm 1, File S2), presenting the step-by-step workflow of the proposed approach.

As part of the study, eight deep learning models were tested: YOLOv8, MobileNetV3, ResNetV2, EfficientNetV2, VGG16, ConvNeXt-Tiny, SwinV2-Base, DaViT (*Redmon et al., 2016*; *Howard et al., 2017*; *He et al., 2015*; *Tan & Le, 2019*; *Simonyan & Zisserman, 2014*; *Liu et al., 2022*, *2021*; *Ding et al., 2022*). The selection of models was based on their effectiveness in image classification and detection tasks, as well as their computational efficiency.

- YOLOv8—a modern single-stage object detection model that delivers high accuracy and speed, making it well-suited for real-time processing applications.
- MobileNetV3—a compact and energy-efficient architecture optimized for mobile and embedded devices, offering high accuracy with minimal resource consumption.
- ResNetV2—a deep neural network featuring residual connections and pre-activation, enabling stable training and high classification performance, particularly for complex objects.
- EfficientNetV2—a well-balanced model focused on training speed and accuracy, suitable for cloud-based solutions and augmented reality applications.
- VGG16—a classic convolutional network known for its high accuracy but characterized by a large number of parameters and high resource consumption.
- ConvNeXt—Tiny—a modern architecture that combines convolutional network principles with Vision Transformer techniques, delivering strong recognition performance with moderate computational complexity.
- SwinV2-Base—an advanced transformer-based architecture with a hierarchical structure and window-based attention mechanism, offering high accuracy and stability when processing high-resolution images. This makes it particularly valuable for tasks requiring contextual understanding and detailed analysis.
- DaViT—a hybrid architecture combining convolutional and transformer blocks with both vertical and horizontal attention mechanisms. It provides an effective balance between accuracy and computational efficiency, making it suitable for server-side applications and, potentially, edge deployments.

**Table 1** A comparative analysis of the models.

| Model | Architecture | Advantages | Limitations | Suggested application |
|---|---|---|---|---|
| YOLOv8 | Single-stage object detector | Very high speed, real-time detection | Possible reduction in localization accuracy | Mobile and AR applications, and automatic navigation systems |
| VGG16 | Classic convolutional CNN | High accuracy, easy to interpret | Large model size, high resource consumption | Cataloging and offline analytics |
| ResNetV2 | Deep network with residual connections | Stable training, high accuracy, suitable for complex objects | Requires significant computational resources | Archival systems and complex classification tasks |
| MobileNetV3 | Lightweight CNN for mobile devices | Compact and efficient, high accuracy with low resource consumption | Challenging to configure and interpret (AutoML-generated) | Mobile applications and resource-constrained devices |
| EfficientNetV2 | Balanced accuracy and speed (NAS-based) | Fast training, good accuracy, parameter efficiency | Challenging to reproduce, resource-intensive in larger versions | Cloud systems and AR/VR applications |
| ConvNeXt-Tiny | Modern CNN with Vision Transformer elements | High accuracy, competitive with ViT, upgraded architecture | Accuracy fluctuations during early training stages, requires fine-tuning | Challenging conditions and multimodal systems |
| SwinV2-Base | Hierarchical Vision Transformer | High accuracy, scalability, robust to resolution changes | Higher computational load, complex configuration | Medical imaging and satellite analysis |
| DaViT | Hybrid CNN+Transformer with directional attention (ViT+CNN) | Balanced accuracy and efficiency, versatility, stable generalization | Novel architecture, limited framework support | Video surveillance, autopilot systems, and cross-domain classifiers |

**Table 2** Key hyperparameters.

| Hyperparameter | Value |
|---|---|
| Batch size | 8 (SwinV2-Base, DaViT)<br>16 (MobileNetV3, ResNetV2, EfficientNetV2, ConvNeXt-Tiny, YOLOv8)<br>32 (VGG16) |
| Image size | 224 × 224 pixels |
| Number of epochs | Up to 300 (with early stopping applied) |
| Optimizer | Adam |
| Initial learning rate | 0.0002 (2e−4) |
| Regularization | dropout (0.8) + L2-regularization (0.01) |

Each model was evaluated for its applicability in museum settings, including mobile applications, cloud services, and real-time systems. A comparative analysis of the models is presented in Table 1.

## EXPERIMENTAL SETUP

This chapter provides a detailed overview of the experimental configuration used for training and evaluating various deep neural network architectures in the tasks of classification and detection of museum exhibits. Table 2 describes the key hyperparameters, data preprocessing procedures, regularization methods, as well as the software and hardware environments employed. To ensure a fair comparison, all models were trained on a uniformly prepared dataset using consistent augmentation techniques,

**Table 3** Model training using both cloud-based and local computational resources.

| Model architectures | Hardware and software |
|---|---|
| VGG16, ConvNeXt-Tiny | Google Colab, GPU: NVIDIA Tesla T4, RAM: 16 ГБ |
| ResNetV2, EfficientNetV2, MobileNetV3 | Google Colab Pro+, GPU: enhanced versions of Tesla T4/A100 |
| YOLOv8 | Local server, GPU: GeForce RTX 2080 Super |
| SwinV2-Base, DaViT | Local server, GPU: GeForce RTX 3070 |

validation strategies, and early stopping mechanisms. Additionally, this chapter outlines the metrics used for the quantitative evaluation of model performance, including accuracy, precision, recall, F1-score, and mean average precision (mAP).

Data augmentation techniques included rotations, horizontal and vertical shifts, scaling (zooming), horizontal mirroring, and brightness adjustment—all applied randomly within a range of up to 20%.

Model training was conducted using both cloud-based and local computational resources, as detailed in Table 3. Specifically, VGG16 and ConvNeXt-Tiny were trained in Google Colab using NVIDIA Tesla T4 GPUs with 16 GB of RAM. For more demanding models such as ResNetV2, EfficientNetV2, and MobileNetV3, training was performed in Google Colab Pro+ with access to enhanced Tesla T4 or A100 GPUs. Meanwhile, YOLOv8 was trained on a local server equipped with a GeForce RTX 2080 Super, and SwinV2-Base and DaViT models were trained using a GeForce RTX 3070. The use of high-performance GPUs significantly accelerated training due to the parallel processing capabilities of the CUDA architecture.

For the MobileNetV3, ResNetV2, and EfficientNetV2 models, more powerful GPU accelerators were used, which enabled faster training and improved performance due to larger batch sizes and higher processing speeds.

The following training characteristics were also applied:

- Dataset split: 80% for training, 20% for validation (using ImageDataGenerator);
- Cross-validation: built-in validation with validation_split was used;
- Callbacks: saving the best weights based on the validation loss metric;
- Early stopping: prevention of overfitting in the absence of improvement.

Table 4 provides an overview of the evaluation metrics used in the study. The confusion matrix is presented as a heatmap, offering a clear visualization of classification errors across different classes. The F1-confidence curve was utilized to determine the optimal confidence threshold for classification. In addition, comparative charts were generated to assess model performance based on accuracy and inference time, supporting a comprehensive evaluation of each architecture.

Based on the testing results:

- MobileNetV3 and EfficientNetV2 demonstrated the best balance between speed and quality;
- VGG16 provided high accuracy but required more resources;

**Table 4** Models evaluation metrics.

| Metric | Description |
|---|---|
| Accuracy | Proportion of correct predictions |
| Precision, Recall, F1-score | Calculated separately for each class |
| Mean average precision (mAP) | mAP0.5 and mAP0.5: 0.95 |
| Confusion matrix | Heatmap for analyzing classification errors |
| F1-confidence curve | Curve showing the relationship between F1-score and model confidence |

- ConvNeXt-Tiny showed excellent performance without significantly increasing computational load;
- YOLOv8 delivered the best overall results;
- SwinV2-Base exhibited outstanding accuracy and strong robustness, confirming its status as a powerful ViT architecture for recognition tasks;
- DaViT combined transformer and convolutional approaches, providing stable results with acceptable training time and high versatility.

Training on powerful GPUs (A100/V100) significantly improved the performance of MobileNetV3, ResNetV2, and EfficientNetV2, nearly doubling the training speed compared to the T4. All models, regardless of their architecture, were trained under controlled conditions with enhanced augmentations and consistent regularization procedures. The use of accelerated hardware, particularly the GeForce RTX 2080 Super for YOLOv8, ensured efficient reduction of training time while maintaining high model accuracy. The SwinV2-Base and DaViT models were trained on the GeForce RTX 3070 GPU, where they demonstrated stable performance with high accuracy and moderate training times. Thus, the choice of the optimal model depends on the specific task, inference speed requirements, available computational resources, and the need for model adaptation to specific operating conditions.

## MobileNetV3 training results

MobileNetV3 demonstrated high classification accuracy (99.93%) already by the 99$^{th}$ epoch. During training, accuracy increased from 86.63% in the first epoch to nearly perfect accuracy, while the loss value decreased from 3.2130 to 0.7425. This model is especially effective for mobile and web applications due to its minimal computational resource requirements. However, MobileNetV3 may struggle with recognizing fine details of exhibits and under varying lighting conditions.

## ResNetV2 training results

ResNetV2 showed steady accuracy improvement, starting at 98.90% in the first epoch and reaching 99.78% by the 80$^{th}$ epoch. The use of pre-trained ImageNet weights accelerated the model's convergence. ResNetV2 performs particularly well in classifying complex museum objects such as textiles, ceramics, and relief items, but it requires substantial computational power, making it less suitable for mobile solutions.

## EfficientNetV2 training results

Starting with a low accuracy of 28.73%, EfficientNetV2 reached 87.8% by the 75$^{th}$ epoch. This is attributed to its optimized architecture, which enables high accuracy with a relatively small number of parameters. However, the model exhibited fluctuations in accuracy at different stages of training, indicating the need for more precise hyperparameter tuning. EfficientNetV2 can be effectively utilized in cloud computing environments and augmented reality (AR)/virtual reality (VR) applications.

## YOLOv8 training results

The YOLOv8 model was trained for museum exhibit detection and achieved a mean average precision (mAP@0.5) of 99.5%. The confusion matrix revealed that the model made virtually no errors, except when dealing with objects similar in shape and texture. The precision-recall (PR) curve confirmed high accuracy even under varying parameters. The F1-confidence curve determined an optimal confidence threshold of 0.691, at which the best balance between precision and recall was achieved. YOLOv8 is particularly effective for real-time applications, such as in automated museum navigation systems.

## Training results of VGG16

VGG16 initially showed low accuracy (5.18%) in the first epoch; however, by the 20th epoch, accuracy reached 99.47%. The model demonstrated a steady improvement in metrics and exhibited high validation accuracy. The use of pre-trained weights accelerated the training process, and final results indicated that VGG16 performs well in classification tasks but lags behind modern architectures in terms of performance. It is more resource-intensive, making it a suboptimal choice for mobile applications, but suitable for cataloging museum data.

## Training results of ConvNeXtTiny

ConvNeXtTiny achieved high accuracy (97%) at the final training stage, though significant fluctuations were observed in the early epochs. Unlike other models, ConvNeXtTiny combines the strengths of CNNs and transformers, making it promising for complex computer vision tasks. However, it proved less accurate than ResNetV2 and VGG16, indicating the need for further architectural refinement.

## Training results of SwinV2-Base

The SwinV2-Base model showed consistently high performance, achieving 97.72% accuracy in the final training epoch. By combining the benefits of convolutional networks and transformer architecture with a hierarchical structure, it effectively handles detailed images of exhibits. Thanks to architectural enhancements, including shifted window attention, the model exhibits excellent generalization ability and performs particularly well on images where both local and global contexts are important. However, compared to lighter models, it requires significant computational resources, limiting its use on low-end hardware.

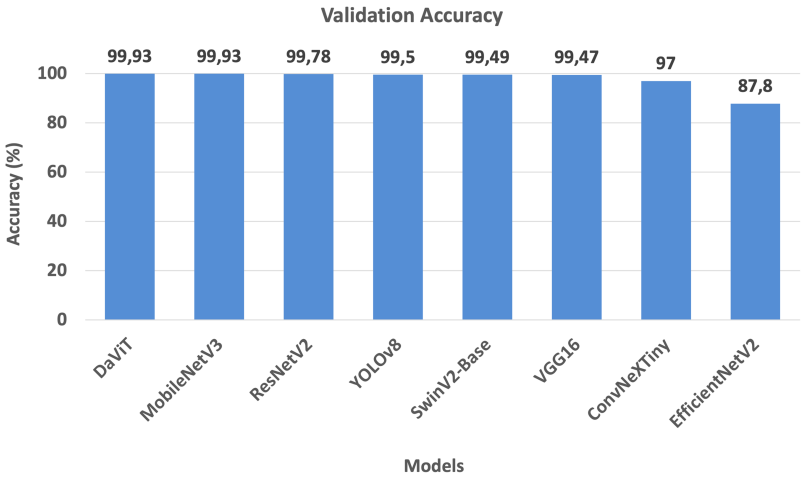


**Figure 2** **Training accuracy of the models.** Full-size DOI: 10.7717/peerj-cs.3207/fig-2

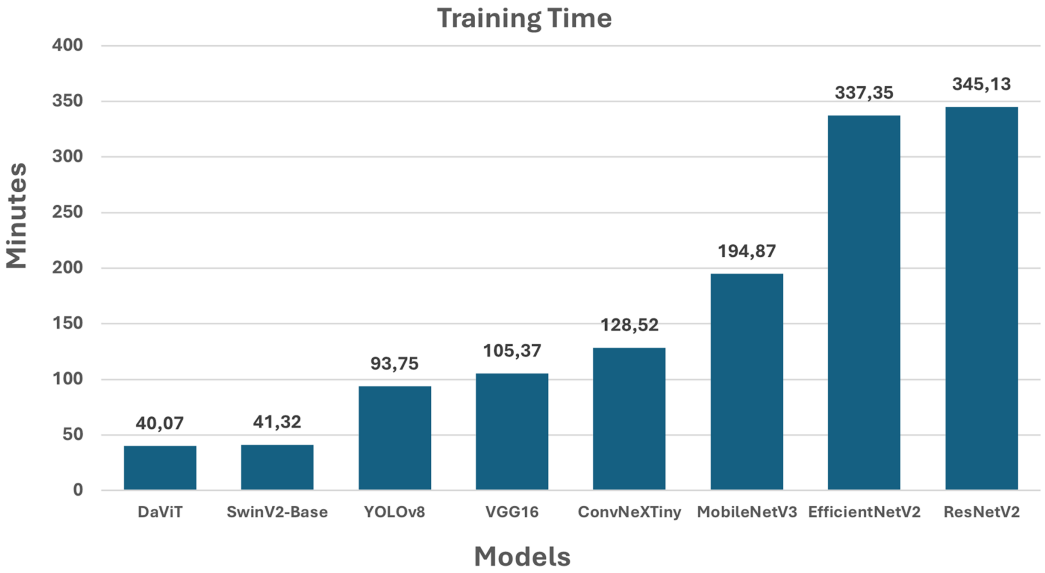


**Figure 3** **Training time of the models.** Full-size DOI: 10.7717/peerj-cs.3207/fig-3

## Training results of DaViT

The Dual Attention Vision Transformer (DaViT) model achieved 99.71% classification accuracy, demonstrating high stability and fast convergence throughout the epochs. Its hybrid architecture integrates axial and spatial attention, allowing more accurate recognition of complex elements in museum exhibits. DaViT showed excellent adaptability to class diversity and object variability. It occupies a middle ground between lightweight and heavy architectures in terms of computational load, making it suitable for desktop and server solutions in museum information systems.

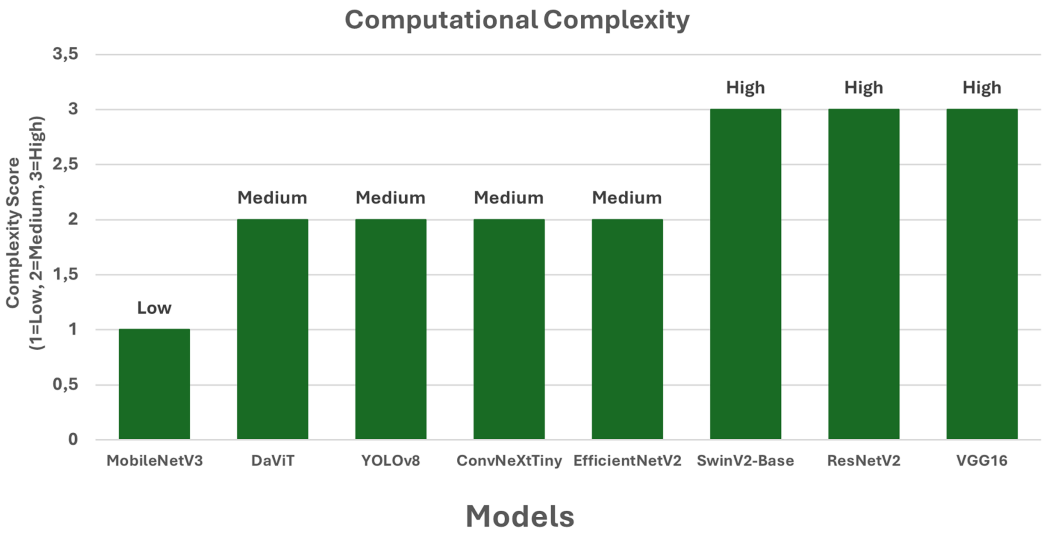


**Figure 4** **Computational complexity of the models.** Full-size DOI: 10.7717/peerj-cs.3207/fig-4

## RESULTS

To evaluate the effectiveness of various neural network architectures in the task of classifying museum exhibits, a comparative analysis of eight popular models was conducted. The study considered three key indicators: prediction accuracy, computational complexity, and training time. The results are visualized in the form of graphs (Figs. 2–4), which clearly illustrate the strengths and weaknesses of each architecture.

Figure 2 shows that MobileNetV3 and DaViT achieve the highest accuracy—99.93% —which is impressive for both a lightweight and a hybrid architecture. ResNetV2, YOLOv8, VGG16, and SwinV2-Base also demonstrate high accuracy values, ranging from 99.47% to 99.78%, confirming their reliability in recognizing museum exhibits. EfficientNetV2 lags significantly behind at 87.80%, which may be due to suboptimal model tuning or specific characteristics of the dataset. ConvNeXtTiny achieved 97.00% accuracy, demonstrating a good compromise between performance and computational cost. Overall, the chart highlights that both compact and transformer-based architectures can reach near-perfect accuracy with proper customization.

The training time of the models (Fig. 3) exhibit significant differences that should be considered when selecting an architecture for practical applications. DaViT and SwinV2-Base demonstrate the shortest training time—approximately 40 min—making them particularly appealing for rapid deployment and frequent fine-tuning. YOLOv8, VGG16, and ConvNeXtTiny occupy intermediate positions, with training times ranging from 93 to 128 min, striking a balance between speed and accuracy, especially for computer vision and classification tasks. MobileNetV3 requires more time (194.87 min), despite being typically classified as a lightweight model. This may be due to specific training configurations, or the characteristics of the dataset used. The models that are the most resource-intensive in terms of training time are EfficientNetV2 (337.35 min) and ResNetV2 (345.13 min). While

these models demonstrate high accuracy, their application in scenarios that necessitate rapid iteration may be constrained.

Consequently, the chart highlights the crucial importance of training time in applied tasks: when computational resources are scarce, priority should be given to models with shorter training times, even if it entails a compromise in accuracy.

According to the computational complexity chart (Fig. 4), MobileNetV3 remains the least resource-intensive model with a low level of complexity, making it ideal for mobile and embedded solutions. The most resource-demanding models are ResNetV2, VGG16, and SwinV2-Base—their architectural characteristics require significant computational power, which limits their use in systems with strict performance constraints. EfficientNetV2, YOLOv8, ConvNeXtTiny, and DaViT demonstrate a moderate level of complexity, offering a good balance between computational cost and accuracy. This makes them attractive for use in versatile museum systems where both classification quality and available resources must be taken into account. Thus, the chart highlights the importance of balancing accuracy and computational load when selecting an architecture.

The comparative analysis of eight deep learning models revealed that each has its own unique advantages and limitations that determine its applicability in museum digital systems.

Therefore, the choice of architecture should be guided by specific tasks: the need for real-time performance, resource availability, the characteristics of the museum environment, and the type of visual data. Combining models may further enhance recognition quality and system adaptability under real-world conditions.

## DISCUSSION

The training and testing results of eight machine learning models confirmed their high effectiveness in the task of automatic recognition of museum exhibits. However, each architecture demonstrated unique characteristics in terms of accuracy, training speed, and computational load.

MobileNetV3 remains the best choice for mobile applications due to its high accuracy and minimal computational cost. ResNetV2 delivers consistently high accuracy but requires significant resources, making it preferable for stationary archival solutions. EfficientNetV2 represents a compromise between accuracy and speed, proving especially effective in cloud and AR applications. YOLOv8 shows excellent performance in real-time tasks thanks to its high speed and detection accuracy. VGG16 still offers high accuracy, but due to its resource intensity, it falls behind more modern architectures and is better suited for offline cataloging tasks. ConvNeXtTiny is a balanced model that combines elements of CNNs and transformers but requires fine-tuning to achieve optimal performance. SwinV2-Base ranks among the top in terms of accuracy but demands substantial computational resources, which is justified in high-end systems. DaViT demonstrated outstanding accuracy with moderate complexity and training time, making it a versatile candidate for both cloud-based and local solutions.

Based on the obtained results, a mobile application utilizing augmented reality technology was developed for the A. Kasteyev State Museum of Arts (Almaty,

Kazakhstan). The system enables real-time object recognition using YOLOv8 and provides users with interactive information about the exhibits, thereby enhancing digital engagement within the museum environment. The developed application is currently undergoing testing.

Future research may focus on developing hybrid solutions that combine YOLOv8 for fast and accurate object detection with high-precision classification architectures such as DaViT or SwinV2-Base. Using ConvNeXtTiny as an intermediate feature extractor could improve the model's generalization capability. Another promising direction is the analysis of the effectiveness of various data augmentation strategies to improve model robustness under real-world conditions—including changing lighting, partial occlusion, and background variability.

## CONCLUSION

Thus, the study demonstrated the high effectiveness of modern deep learning models for automatic recognition of museum exhibits in conditions close to real-world scenarios. Eight models were comparatively analyzed: MobileNetV3, ResNetV2, EfficientNetV2, YOLOv8, VGG16, ConvNeXtTiny, SwinV2-Base, and DaViT. Each architecture was evaluated based on accuracy, training time, and computational complexity, allowing for the identification of the most suitable use cases within digital museum systems.

The results showed that MobileNetV3, YOLOv8, and DaViT are optimal for mobile solutions and augmented reality systems due to their combination of high accuracy and performance. ResNetV2, VGG16, and SwinV2-Base deliver high classification quality, making them effective for tasks such as cataloging, archiving, and analytics. EfficientNetV2 and ConvNeXtTiny demonstrate strong potential for use in cloud-based platforms and adaptive systems, though they require careful tuning and parameter control.

Based on the selected models, a mobile application with augmented reality functionality was developed for the A. Kasteyev State Museum of Arts, confirming the practical relevance of the proposed solution and its potential for widespread implementation in the museum sector.

Therefore, the integration of computer vision and deep learning technologies into the cultural heritage domain opens new horizons for the digital transformation of museums, enhancing visitor engagement and enabling the development of intelligent navigation and educational systems. Promising directions for future research include improving model robustness to varying imaging conditions, developing hybrid architectures, and integrating multimodal data (images, text, audio) into a unified interactive platform.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The study was carried out with the financial support of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19676803). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Grant Disclosures

The following grant information was disclosed by the authors:
Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan: AP19676803.

## Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

- Madina Ipalakova conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Zhiger Bolatov conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yevgeniya Daineko conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Regina Sharshova conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Kamila Abdugapparova conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Dana Tsoy conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplemental File.

The code is available at GitHub and Zenodo:

- https://github.com/ZhigerBolatov/KasteevMuseum.

- Ipalakova, M., Bolatov, Z., Daineko, Y., Sharshova, R., Abdugapparova, K., & Tsoy, D. (2025). Kasteev Museum Recognition Model Code. Zenodo. https://doi.org/10.5281/zenodo.16722603.

Data is also available at Zenodo:

Ipalakova, M., Bolatov, Z., Daineko, Y., Sharshova, R., Abdugapparova, K., & Tsoy, D. (2025). Kasteev Museum Models and Dataset Iteration 1 & Iteration 2 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.16941686.

## Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3207#supplemental-information.

## REFERENCES


**Abdollahi A, Pradhan B, Alamri A. 2020.** VNet: an end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *IEEE Access* **8**:179424–179436 DOI 10.1109/ACCESS.2020.3026658.

**Acosta H, Henderson N, Rowe J, Min W, Minogue J, Lester J. 2021.** What's fair is fair: detecting and mitigating encoded bias in multimodal models of museum visitor attention. In: *Proceedings of the 2021 International Conference on Multimodal Interaction. Presented at the ICMI '21: International Conference on Multimodal Interaction, Montréal QC Canada* DOI 10.1145/3462244.3479943.

**Bazarbekov I, Razaque A, Ipalakova M, Yoo J, Assipova Z, Almisreb A. 2024.** A review of artificial intelligence methods for Alzheimer's disease diagnosis: insights from neuroimaging to sensor data analysis. *Biomedical Signal Processing and Control* **92(106023)**:106023 DOI 10.1016/j.bspc.2024.106023.

**Bobasheva A, Gandon F, Precioso F. 2022.** Learning and reasoning for cultural metadata quality: coupling symbolic AI and machine learning over a semantic web knowledge graph to support museum curators in improving the quality of cultural metadata and information retrieval. *Journal on Computing and Cultural Heritage* **15(3)**:1–23 DOI 10.1145/3485844.

**Cai S, Tian Y, Lui H, Zeng H, Wu Y, Chen G. 2020.** Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative Imaging in Medicine and Surgery* **10(6)**:1275–1285 DOI 10.21037/qims-19-1090.

**Cai P, Zhang K, Pan Y. 2023.** Application of AI interactive device based on database management system in multidimensional design of museum exhibition content. DOI 10.21203/rs.3.rs-3074947/v1.

**Ding M, Xiao B, Codella N, Luo P, Wang J, Yuan L. 2022.** DaViT: dual attention vision transformers (Version 1). ArXiv DOI 10.48550/arXiv.2204.03645.

**Foret P, Kleiner A, Mobahi H, Neyshabur B. 2020.** Sharpness-aware minimization for efficiently improving generalization. ArXiv DOI 10.48550/arXiv.2010.01412.

**Gao L, Li H, Chen Q, Peng D. 2024.** A multimodal commodity hybrid recommender system incorporating M2net+ (Vit-L/16). SSRN DOI 10.2139/ssrn.4825404.

**Gesmundo A. 2022.** A continual development methodology for large-scale multitask dynamic ML systems. ArXiv DOI 10.48550/arXiv.2209.07326.

**Gesmundo A, Dean J. 2022.** An evolutionary approach to dynamic introduction of tasks in large-scale multitask learning systems. ArXiv DOI 10.48550/arXiv.2205.12755.

**Go J, Moon N. 2024.** Cleaned meta pseudo labels-based pet behavior recognition using time-series sensor data. *Sensors* **24(11)**:3391 DOI 10.3390/s24113391.

**Gorbachev VA, Krivorotov IA, Markelov AO, Kotliarova EV. 2020.** Semantic segmentation of airport satellite images using convolutional neural networks. *Computer Optics* **44(4)**:636–645 DOI 10.18287/2412-6179-CO-636.

**He K, Zhang X, Ren S, Sun J. 2015.** Deep residual learning for image recognition (Version 1). ArXiv DOI 10.48550/arXiv.1512.03385.

**Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. 2017.** MobileNets: efficient convolutional neural networks for mobile vision applications (Version 1). ArXiv DOI 10.48550/arXiv.1704.04861.

**Huang H-Y, Liem CCS. 2022.** Social inclusion in curated contexts: insights from museum practices. In: *2022 ACM Conference on Fairness, Accountability, and Transparency. Presented at*

*the FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea* DOI 10.1145/3531146.3533095.

**Huang J, Nowack R. 2020.** Machine learning using U-Net convolutional neural networks for the imaging of sparse seismic data. *Pure and Applied Geophysics* **177(6)**:2685–2700 DOI 10.1007/s00024-019-02412-z.

**Jin Q, Meng Z, Pham TD, Chen Q, Wei L, Su R. 2019.** DUNet: a deformable network for retinal vessel segmentation. *Knowledge-Based Systems* **178(5)**:149–162 DOI 10.1016/j.knosys.2019.04.025.

**Kato S, Hotta K. 2020.** Pitching classification and habit detection by V-Net. In: *Proceedings of the 17th International Joint Conference on e-Business and Telecommunications*, 615–621 DOI 10.5220/0009347106150621.

**Khan I, Chen T, Khan MAA, Aamir S, Menhaj W. 2022.** Object recognition in different lighting conditions at various angles by deep learning method. ArXiv DOI 10.48550/arXiv.2210.09618.

**Kiourexidou M, Stamou S. 2025.** Interactive heritage: the role of artificial intelligence in digital museums. *Electronics* **14(9)**:1884 DOI 10.3390/electronics14091884.

**Koniusz P, Tas Y, Zhang H, Harandi M, Porikli F, Zhang R. 2018.** Museums exhibit identification challenge for domain adaptation and beyond. ArXiv DOI 10.48550/arXiv.1802.01093.

**Lee BCG. 2025.** The "Collections as ML Data" checklist for machine learning and cultural heritage. *Journal of the Association for Information Science and Technology* **76(2)**:375–396 DOI 10.1002/asi.24765.

**Li Z. 2024.** Opportunities and challenges of artificial intelligence + enabling museum building. *Applied Mathematics and Nonlinear Sciences* **9(1)**:245–259 DOI 10.2478/amns-2024-2093.

**Li Y, Li W, Yu L, Wu M, Liu J, Li W, Hao M, Li S. 2024.** A novel paradigm for neural computation: X-Net with learnable neurons and adaptable structure. ArXiv DOI 10.48550/arXiv.2401.01772.

**Liu Z, Hu H, Lin Y, Yao Z, Xie Z, Wei Y, Ning J, Cao Y, Zhang Z, Dong L, Wei F, Guo B. 2021.** Swin Transformer V2: scaling up capacity and resolution. ArXiv DOI 10.48550/arXiv.2111.09883.

**Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. 2022.** A ConvNet for the 2020s (Version 2). ArXiv DOI 10.48550/arXiv.2201.03545.

**Liu Y, Yang H, Dong Z, Keutzer K, Du L, Zhang S. 2023.** NoisyQuant: noisy bias-enhanced post-training activation quantization for vision transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 20321–20330.

**Luo S, Li H. 2024.** The design of intelligent recognition and guide system for museum exhibits based on computer vision and deep learning. *Frontiers in Artificial Intelligence and Applications* **391**:115–122 DOI 10.3233/faia241093.

**Meyer L, Aaen JE, Tranberg AR, Kun P, Freiberger M, Risi S, Løvlie AS. 2024.** Algorithmic ways of seeing: using object detection to facilitate art exploration. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Article 29, New York: Association for Computing Machinery DOI 10.1145/3613904.3642157.

**Milletarì F, Navab N, Ahmadi S. 2016.** V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. Piscataway: IEEE, 565–571.

**Papers with Code. 2025.** Trending papers. *Available at https://paperswithcode.com*.

**Pasqualino G, Furnari A, Signorello G, Farinella GM. 2020.** Synthetic to real unsupervised domain adaptation for single-stage artwork recognition in cultural sites. ArXiv DOI 10.48550/arXiv.2008.01882.

**Patel N, Ramoliya F, Jadav NK, Gupta R, Tanwar S, Aujla GS. 2024.** X-NET: explainable AI-based network data security framework for Healthcare 4.0. In: *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*. Piscataway: IEEE, 481–486 DOI 10.1109/ICCWorkshops59551.2024.10615382.

**Patil P, Sharma A, Jain R. 2024.** A comprehensive study on object detection techniques in unfettered environments. In: *INCON XVII Conference Proceedings*.

**Perera WL, Messemer H, Heinz M, Kretzschmar M. 2020.** Detecting treasures in museums with artificial intelligence. In: *Workshop Gemeinschaften in Neuen Medien (GeNeMe)*, Dresden, Germany.

**Pham H, Dai Z, Xie Q, Le QV. 2021.** Meta pseudo labels. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 11557–11568.

**Rahim NM, Hairuddin MA, Megat Ali MSA, Tahir NM, Almisreb AA, Ashar NDK. 2025.** Pretrained convolutional neural network for fruit classification analysis of pineapple plantation images. *Engineering, Technology & Applied Science Research* **15(2)**:20819–20826 DOI 10.48084/etasr.9249.

**Redmon J, Divvala S, Girshick R, Farhadi A. 2016.** You only look once: unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 779–788 DOI 10.1109/CVPR.2016.91.

**Shabalina DE, Lanchukovskaya KS, Liakh TV, Chaika KV. 2021.** Semantic image segmentation in Duckietown. *Vestnik NSU. Series: Information Technologies* **19(3)**:26–39 DOI 10.25205/1818-7900-2021-19-3-26-39.

**Simonyan K, Zisserman A. 2014.** Very deep convolutional networks for large-scale image recognition (Version 6). ArXiv DOI 10.48550/arXiv.1409.1556.

**Sokolov IV. 2024.** The SegNet neural network for the segmentation of skin neoplasms in medical images. *Trends in the Development of Science and Education* **106(11)**:112–114 DOI 10.18411/trnio-02-2024-617 [In Russian].

**Tan M, Le QV. 2019.** EfficientNet: rethinking model scaling for convolutional neural networks. ArXiv DOI 10.48550/arXiv.1905.11946.

**Tang Q, Zheng L, Chen Y, Yan L, Chen J. 2024.** Artificial intelligence empowering museum space layout design: insights from China. *PLOS ONE* **19(11)**:e0310594 DOI 10.1371/journal.pone.0310594.

**Taran V, Gordienko N, Kochura Y, Gordienko Y, Rokovyi A, Alienin O, Stirenko S. 2018.** Performance evaluation of deep learning networks for semantic segmentation of traffic stereo-pair images. ArXiv DOI 10.48550/arXiv.1806.01896.

**Walsh D, Clough P, Hall MM, Hopfgartner F, Foster J. 2021.** Clustering and classifying users from the national museums Liverpool website. In: *Lecture Notes in Computer Science. Linking Theory and Practice of Digital Libraries*. Cham: Springer, 202–214.

**Wang Q, Li L. 2022.** Museum relic image detection and recognition based on deep learning. *Computational Intelligence and Neuroscience* **2022(2)**:9670191 DOI 10.1155/2022/9670191.

**Wang N, Zhao K. 2023.** An adaptive method based on fuzzy logic and AdaBoost for exhibition of museum collections. In: Batista P, Pachori RB, eds. *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2023)*. Cham: Springer, 159.

**Ypsilantis NA, Garcia N, Han G, Ibrahimi S, Van Noord N, Tolias G. 2021.** The met dataset: instance-level recognition for artworks. In: *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

**Yu X, Huang Z, Xue Y, Zhang L, Wang L, Liu T, Zhu D. 2023.** NoisyNN: exploring the influence of information entropy change in learning systems. ArXiv DOI 10.48550/arXiv.2309.10625.

**Yuan W, Wang J, Xu W. 2022.** Shift pooling PSPNet: rethinking PSPNet for building extraction in remote sensing images from entire local feature pooling. *Remote Sensing* **14(19)**:4889 DOI 10.3390/rs14194889.

**Yuan H, Zhu J, Wang Q, Cheng M, Cai Z. 2022.** An improved DeepLab v3+ deep learning network applied to the segmentation of grape leaf black rot spots. *Frontiers in Plant Science* **13**:795410 DOI 10.3389/fpls.2022.795410.

**Zaitseva EV, Kazankov VK. 2021.** Comparison of neural network architectures for image segmentation. In: *Current Issues in the Development of the Modern Digital Environment: Collection of Articles Based on the Proceedings of the Scientific and Technical Conference of Young Scientists*. Volgograd: Sirius Publishing, 468 [In Russian].

**Zhang J, Li Y, Chen F, Pan Z, Zhou X, Li Y, Jiao S. 2019.** X-Net: a binocular summation network for foreground segmentation. *IEEE Access* **7**:71412–71422 DOI 10.1109/ACCESS.2019.2919802.

**Zhang R, Tas Y, Koniusz P. 2018.** Artwork identification from wearable camera images for enhancing experience of museum audiences. ArXiv DOI 10.48550/arXiv.1806.09084.

**Zhou Y, Huang W, Dong P, Xia Y, Wang S. 2021.** D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **18(3)**:940–950 DOI 10.1109/TCBB.2019.2939522.