ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)*
ISSN 3083-7650 (0135-1710) (print) *Management Information System and Devices. 2026. No. 2 (189)*

UDC 004.89; 004.93 DOI: https://doi.org/10.30837/0135-1710.2026.189.182

Matvii Kuchapin, Kyrylo Smelyakov, Anastasiya Chupryna, Sergiy Luchenko

COMPREHENSIVE ANALYSIS OF METHODS AND TOOLS FOR HYBRID IMAGE ANNOTATION IN COMPUTER VISION SYSTEMS

The **subject matter** of the study is the methods, algorithms, and software tools for annotating visual data in computer vision systems within the Data-Centric AI paradigm. The study analyzes the processes of structuring unstructured information where labeling quality determines the accuracy of artificial intelligence models, considering the limitations of manual methods and risks of systematic errors in full automation. The **goal** of the work is to perform a comprehensive analysis of modern image annotation methods and tools in computer vision systems and a quantitative evaluation of the effectiveness of hybrid Human-in-the-Loop strategies to improve the efficiency of forming high-quality datasets within the Data-Centric AI paradigm. The following **tasks** were addressed in the article: a systematization of annotation types from image classification to panoptic segmentation and 3D scenes was conducted; a review of the toolset based on the Segment Anything Model and GroundingDINO was performed; comparative evaluations of manual, automatic, and hybrid labeling scenarios based on accuracy $(mIoU)$ and labor intensity were carried out; issues regarding operator trust and interaction ergonomics were identified. The following **methods** are used – systematic comparative analysis of hybrid Human-in-the-Loop strategies, cross-domain synthesis of active learning and interactive segmentation studies, and formalization of acceleration, quality, and manual labor indicators. The following **results** were obtained: the hybrid pipeline provides a 5.4x process acceleration. The formation time of a semantic mask for an object is reduced from 65.0 to 12.0 seconds while maintaining quality at $mIoU = 0.94$, with only a 0.02 loss relative to the reference standard. The hybrid scenario was found to be optimal within the threshold range of $0,82 < Q_{min} \le 0,94$, covering the broadest class of practical tasks from training production models to medical diagnostics. A direct dependence of annotator performance on the reliability of automatic prompts was identified. **Conclusions:** hybrid annotation is the optimal strategy for creating Ground Truth in critical domains such as autonomous driving and medicine, providing a balance between speed and accuracy. The proposed formalization of the optimization problem with Q_{min} threshold constraint enables informed selection of the annotation scenario for a specific domain. Future research involves improving synthetic data generation to overcome the domain gap and developing adaptive interfaces to reduce cognitive load.

Keywords: computer vision; image annotation; active learning; segmentation; synthetic data.

1. Introduction

The exponential growth in the volume of visual data is the foundation of modern computer vision (CV). However, in its unstructured form, this information has limited value without appropriate semantic interpretation [1]. Data structuring has become a key challenge, making annotation the most complex and costly stage of artificial intelligence (AI) development. Despite significant progress in the development of neural network architectures, within the Data-Centric AI paradigm, the quality of training datasets is recognized as the primary factor limiting model productivity [2].

Poor annotation quality has critical consequences for the effectiveness of systems in achieving accurate predictions. This illustrates the fundamental principle of Data Science – "Garbage In, Garbage Out" [3]. The situation is complicated by the fact that traditional manual methods scale poorly and are subjective, while full automation carries the risk of systematic errors [4]. Therefore, the most promising solution is hybrid Human-in-the-Loop approaches, which combine intelligent systems with expert verification [5].

In this context, the aim of this study is to conduct a comprehensive analysis of modern methods and tools for image annotation in computer vision systems and to quantitatively evaluate the effectiveness of hybrid Human-in-the-Loop strategies for improving the efficiency of creating high-quality datasets within the Data-Centric AI paradigm. The main focus is on researching hybrid strategies that allow for the effective combination of the computational capabilities of machine learning algorithms with expert verification, thereby mitigating the shortcomings of both purely manual and fully automated approaches. The analysis aims to systematize existing annotation types and software tools, as well as to quantitatively justify the effectiveness of implementing intelligent pipelines for creating high-quality training datasets in modern computer vision systems.

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

The scientific novelty of the study lies in the following:

– a systematic cross-domain comparative analysis of hybrid annotation systems from various subject areas, such as medical diagnostics, security systems, synthetic art classification, and object detection, was conducted, and a formalized criterion for selecting the optimal annotation scenario based on a quality threshold constraint was formulated, which allowed us to determine the limits of effective application of each approach depending on domain requirements;

– a formalized weighted effectiveness metric is proposed which, unlike known metrics, accounts for the criticality of accuracy for a specific domain through a parametric coefficient, allowing for a well-founded selection of the optimal annotation scenario for domains with varying quality requirements;

– based on a synthesis of independent research results, the influence of ergonomic factors in operator-system interaction on the effectiveness of hybrid annotation has been identified and systematized, in particular, the effect of negative framing, where the operator's bias regarding the unreliability of AI leads to the rejection of correct suggestions, which allows for the formulation of practical requirements to ensure the transparency of algorithmic recommendations and the communication of the model's confidence level when designing interfaces for hybrid annotation systems.

2. Analysis of the Current State of Research on Hybrid Annotation Systems

An analysis of scientific publications in recent years indicates a steady trend toward the implementation of Human-in-the-Loop (HITL) approaches, where the human factor is combined with machine learning algorithms to optimize annotation processes. Current research focuses on resolving the dilemma between annotation cost and training data quality. Hybrid systems demonstrate high adaptability across various domains, ranging from medical diagnostics and security systems to digital art analysis.

A key achievement of this period was the demonstration of the effectiveness of active learning and interactive segmentation methods. These approaches allow for a 50–87% reduction in the volume of manual annotation without sacrificing model accuracy, delegating routine tasks to algorithms and complex cases to experts.

To conduct a detailed study of the effectiveness of hybrid approaches, five recent papers were selected and analyzed based on the following criteria:

– the presence of quantitative results demonstrating the effectiveness of the HITL approach;

– coverage of various subject areas to ensure cross-domain representativeness;

– publication in peer-reviewed journals from 2023–2025.

The sample covers various subject areas, allowing for an assessment of the methods' versatility. Summary data on the models used, metrics, and key results of these studies are presented in Table 1.

A review of the literature shows that combining active learning with interactive annotation is the dominant strategy for conserving resources. The paper "HAL-IA: A Hybrid Active Learning framework using Interactive Annotation for medical image segmentation" [6] proposes a hybrid strategy where the model automatically selects the most ambiguous regions, and an expert refines them by clicking on superpixels. This approach allows for achieving accuracy metrics close to those of full annotation while maintaining high Dice/IoU scores, using significantly less annotated data, and minimizing manual intervention.

Table 1. *A comparative review of modern hybrid annotation systems*

Study type	Problem and data	Key results
Active Learning and Interactive segmentation [6]	Medical segmentation (ultrasound, CT, X-ray). 4 datasets.	High accuracy (Dice, IoU) with fewer clicks and annotated images; reduced annotation costs.
HITL Hybrid decision making [7]	Smuggling detection in X-ray images (security).	Improved recall (threat detection) compared to human performance; increased system throughput.
Human-in-the-Loop and Active Learning [8]	Classification: "Real vs. Artificial Art".	Accuracy of ~98.65% using 87.5% less labeled data; rapid adaptation to new generative models.
HITL with confidence estimation [9]	Emotional engagement assessment (DAiSEE video dataset).	A correlation was found between model reliability and annotator performance; AI errors cause frustration and reduce consistency.
Progressive active learning [10]	Object detection on floor plans.	mAP = 0.833 when trained on 500 labeled and 4,500 unlabeled images; significant reduction in manual labor.

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)
ISSN 3083-7650 (0135-1710) (print) Management Information System and Devices. 2026. No. 2 (189)

Similar effectiveness is confirmed in the paper "The power of progressive active learning in floorplan images for energy assessment" [10], where a progressive learning strategy was applied for object detection in floor plans. Starting training with just 500 labeled images, the model gradually learned on a dataset of 4,500 unlabeled examples, achieving an mAP accuracy of ~0.833, which is comparable to the results of full manual annotation.

A telling example is the classification of synthetic art in the paper "Hybrid intelligence approach for detecting synthetic art" [8], where the AL approach reduced the need for labels by 87.5%. Moreover, the system demonstrated high adaptability when images from new generative models appeared; accuracy initially dropped to 75%, but after several feedback iterations, it recovered to 98%. This indicates that selectively involving humans in the most informative and complex samples is significantly more effective than linear, continuous labeling.

A second critically important research area is the optimization of "human-machine" interaction and the ergonomics of decision-making. In the scientific study "Application of human-in-the-loop hybrid augmented intelligence approach in security inspection system" [7], in the context of security tasks such as X-ray screening, the superiority of hybrid strategies over fully automated or manual methods was demonstrated. In particular, the "deviation-priority" mode, where the machine focuses on detecting suspicious objects and a human verifies them, allowed for maximizing the completeness of threat detection, thereby expanding the system's "safety zone". At the same time, the alternative "clearance-priority" mode ensured a balance between reliability and speed of inspection.

However, the success of such collaboration directly depends on psychological factors, particularly the operator's trust in the algorithm. The study "Human-in-the-Loop Annotation for Image-Based Engagement Estimation: Assessing the Impact of Model Reliability on Annotation Accuracy" [9] revealed a direct correlation between the reliability of the model's prompts and the quality of the annotator's work. Experiments showed that when working with an unreliable model, users felt frustrated and demonstrated inconsistency in their decisions. Most interesting is the identified effect of negative framing. If users were biased regarding the system's unreliability, they tended to reject even correct AI prompts. Thus, algorithm transparency and clear communication of the model's confidence level are no less important than its mathematical accuracy, as they establish the necessary level of trust for productive collaboration.

Thus, the analysis confirms that the key to the effectiveness of modern annotation systems lies in the harmonious combination of algorithmic optimization through active learning and the deep integration of ergonomic principles. Technological accuracy must be complemented by transparency in interaction, as it is the operator's trust and cognitive comfort that become decisive factors in achieving high-quality annotation in hybrid environments.

3. Research Objectives and Tasks

The objective of this research is a comprehensive analysis of modern methods and tools for image annotation in computer vision systems to improve the efficiency of creating high-quality datasets within the Data-Centric AI paradigm. Unlike existing works that examine the effectiveness of hybrid approaches in individual domains, this study focuses on a unified cross-domain analysis with the formalization of a criterion for selecting the optimal annotation scenario, taking into account the criticality of the subject domain.

To achieve this goal, the following main tasks must be addressed:

– systematize annotation types and methods, covering various levels of detail from image classification to panoptic segmentation and 3D scene labeling;

– review modern software platforms and tools, in particular those based on the fundamental models Segment Anything Model and Grounding DINO;

– conduct a comparative analysis of the effectiveness of manual, automatic, and hybrid annotation approaches in terms of spatial accuracy (mIoU) and computational complexity;

– identify current scientific challenges in the field of data preparation, including issues related to the adaptation of synthetic domains and the reduction of cognitive load on annotators.

4. Methodological Foundations and Tools for Visual Data Annotation

4.1. Formalization of the Annotation Task and Selection of Annotation Strategies

In the context of building intelligent systems, annotation is defined as the process of enriching "raw"

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

data with metadata, which transforms unstructured visual signals into a format suitable for algorithmic processing and knowledge extraction [11]. For visual data, this process bridges the semantic gap between the low-level pixel representation of an image and the high-level semantics of a scene, with the level of detail ranging from image classification to pixel segmentation and 3D space labeling [12].

Within the data-driven paradigm, the quality, completeness, and representativeness of annotated data are recognized as the primary factor in model performance [13]. Annotations form a ground truth label that defines target values for calculating the loss function during training and serves as a reference standard during validation. Errors in labels, such as annotation noise, reduce the model's ability to generalize, forcing it to learn spurious correlations instead of true features [13], which makes annotation a key iterative component of the data preparation pipeline [14].

To establish a sound methodological framework for comparing annotation strategies, it is advisable to formalize the key components of this process. Formally, annotation can be defined as a mapping $A: X \to Y$, where X is the space of input images, and Y is the space of labels. For the task of semantic segmentation, each image $x \in X$ has dimension $H \times W$, and the corresponding label $y \in Y$ is a matrix of the same dimension, where each element $y(i, j) \in \{1,...,C\}$ determines the pixel's membership in one of the C semantic classes. An ideal mapping A^* generates a ground-truth annotation, but any real annotation process A introduces a certain level of noise η, formalized as $A(x) = A^*(x) + \eta$, where η depends on the chosen strategy. For manual annotation η is due to the operator's subjectivity; for automatic annotation, to systematic model errors; and for hybrid annotation, to a combination of both, taking into account corrections.

The relationship between the quality of the annotation and the model's productivity is formalized using the empirical risk function. For a model f_θ with parameters θ, the empirical risk is defined as:

$$R(\theta) = \frac{1}{N}\sum_{i=1}^{N} L\left(f_\theta(x_i), y_i\right) \quad (1)$$

where L – loss function.

In the presence of annotation noise η, the true learning risk takes the form of:

$$\tilde{R}(\theta) = \frac{1}{N}\sum_{i=1}^{N} L\left(f_\theta(x_i), y_i + \eta_i\right) \quad (2)$$

According to studies on the impact of noisy labels [13], systematic noise η causes a shift in the parameter optimum θ^*, resulting in the model learning false correlations instead of the true features of the objects. This relationship defines the fundamental requirement of minimization $\|\eta\|$ to ensure training quality, which justifies the need for expert verification of automatic labels in hybrid pipelines.

The problem of selecting the optimal annotation strategy can be formulated as a conditional optimization problem. Minimize the total cost $C_{\text{total}} = C_{\text{auto}} + C_{\text{human}}$ subject to the constraint $Q(\hat{A}) \geq Q_{\min}$, where Q is the annotation quality metric (e.g., mIoU), and $Q_{\min}$ is the minimum acceptable quality threshold for the target domain. In a fully manual scenario $C_{\text{auto}} = 0$, and $C_{\text{human}} = \max$. In a fully automatic scenario $C_{human} = 0$, but the constraint $Q \geq Q_{\min}$ may be violated due to systematic model errors. The Human-in-the-Loop hybrid scenario minimizes C_{total}, ensuring compliance with the quality constraint through targeted expert correction of only problematic areas. The parameter $Q_{\min}$ is domain-dependent: for critical applications, such as autonomous driving or medical diagnostics, it approaches the values of the ground-truth manual annotation, whereas for pre-filtering tasks, a significantly lower threshold is permitted.

4.2. Multidimensional Typology of Annotation Processes

The annotation of visual information is a multi-component and multi-faceted process, which in modern research is typically classified according to three interrelated dimensions: the type of source data, the level of detail of visual primitives, and the degree of automation in the annotation process.

From the perspective of information theory, the validity of this classification is determined by differences in the information capacity of various types of annotations. Image classification requires only $\log_2(C)$ bits of information per sample, where C is the number of classes. The bounding box adds spatial coordinates, increasing the amount of information to

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)
ISSN 3083-7650 (0135-1710) (print) Management Information System and Devices. 2026. No. 2 (189)

$\log_2(C)+4\log_2(R)$ bits, where R is the resolution of the coordinate grid. Pixel segmentation requires $H\times W\times\log_2(C)$ bits, which is orders of magnitude higher than the previous levels. This exponential growth in information capacity theoretically justifies the nonlinear increase in annotation complexity and explains why segmentation tasks are the prime candidates for automation.

The first vector defines the specific spatiotemporal characteristics of the information. The baseline here consists of 2D images. These are static arrays of pixels where annotation focuses exclusively on spatial features within the frame plane [15]. Video data imposes significantly higher processing requirements, as the addition of a temporal dimension necessitates ensuring temporal coherence and tracking unique object identifiers across a sequence of frames [16]. The highest level of complexity is characterized by 3D data, such as LiDAR point clouds, where annotation occurs in three-dimensional space, requiring the determination of physical dimensions and the precise orientation of the object, which is critical for autonomous navigation systems [17].

Along with the nature of the input data, an important classification parameter is the level of detail in the scene description, i.e., the type of annotation primitives. The most aggregated level is image-level annotation, where only a general category is established without specifying the localization of objects. For object detection tasks, the standard approach involves the use of bounding boxes – rectangular regions that can be either axially aligned or oriented at an arbitrary angle to more accurately capture the object's geometry [15]. The highest spatial accuracy is provided by polygonal or pixel segmentation methods, where each pixel is associated with a specific class or a specific instance of an object, allowing for the most detailed delineation of its shape [18]. A separate group consists of key points used to describe the human skeletal structure or mechanisms, as well as polylines used to mark road infrastructure elements, trajectories, and other linear objects.

The third defining vector of the typology is the annotation methodology, which ranges from fully manual approaches to complete process automation. Traditionally, manual annotation is considered the "gold standard" due to its high accuracy and expert oversight; however, it is expensive, labor-intensive, and has limited scalability. In response to these limitations, a semi-automated approach known as "Human-in-the-loop" has emerged, combining automatic preliminary tag generation by algorithms with subsequent human verification. This approach significantly reduces the amount of manual work and can accelerate the annotation process by a factor of 10–20 [19].

The most modern and rapidly growing field is synthetic annotation, in which labeled data is procedurally generated in artificial or simulated environments. This allows for the automatic generation of annotations with perfect accuracy and full control over scene parameters. At the same time, a key challenge of this approach remains the gap between the simulated and real domains, which requires the use of domain adaptation and style alignment methods to ensure the correct transfer of models to real-world conditions [20].

4.3. Main Tasks and Formats of Image Annotation

Image annotation covers a wide range of tasks that form the basis of computer vision models, from simple classifiers to autonomous navigation systems. These tasks differ significantly in terms of annotation effort, which directly influences the choice of annotation strategy in the context of the optimization problem formalized above C_{total}.

Image classification is a fundamental task in which models assign a global label (single-class or multi-class) to the entire image without localizing individual objects [21]. From an annotation perspective, this task is characterized by minimal labor intensity and the lowest sensitivity to noise η, since it does not require spatial detail.

Object detection requires determining not only the category but also the spatial localization in the form of bounding boxes: horizontal (COCO, PASCAL VOC formats [22]) or oriented at an arbitrary angle for aerial photography and satellite image analysis tasks [23]. The accuracy of the bounding boxes directly affects the quality of detector training, and the degree of object overlap in dense scenes significantly complicates both manual and automatic annotation.

Image segmentation provides the highest spatial accuracy by performing classification at the pixel level. We distinguish between semantic (class of each pixel), instance (identification of individual instances), and panoptic segmentation, which combines both approaches [18, 24, 25]. Segmentation tasks are the

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

most labor-intensive; creating a single pixel mask requires significantly more time than bounding box annotation, making them prime candidates for implementing hybrid Human-in-the-Loop strategies.

Pose and keypoint estimation involves annotating skeletal models of objects, ranging from 17 keypoints in the COCO format to 106 for facial analysis, which is used in augmented reality systems, sports analytics, and gesture recognition [26]. A separate area is the description of images in natural language, which is critical for training CLIP, BLIP-2, and GPT-4V models and for creating assistive systems [27].

The presented typology of tasks demonstrates that the labor intensity of annotation increases non-linearly with the level of detail, ranging from seconds for classification to minutes for pixel-level segmentation of a single image. This effect has a theoretical basis in the relationship between the complexity of the label and the minimum sample size required to train the model. According to statistical learning theory, more complex label spaces require a greater number of annotated examples to achieve an equivalent level of generalization, which places a double burden on resources. Each sample is both more expensive to annotate and requires a larger quantity. This increase directly determines the feasibility of using automated and hybrid strategies, where minimization C_{total} becomes critical precisely for tasks with a high label complexity component C_{human}.

4.4. The current state of annotation technologies and tools

The modern visual data annotation ecosystem is undergoing a clear shift from labor-intensive manual annotation processes to comprehensive, automated workflows driven by artificial intelligence algorithms. This transformation is driven by the rapid increase in demand for large-scale training datasets and the need to eliminate the bottleneck of slow and expensive manual annotation. While the industry previously relied heavily on crowdsourcing approaches, the "human-in-the-loop" paradigm is now becoming dominant, within which the annotator's role shifts from creating labels to that of an expert verifier and data curator [2]. The introduction of automatic pre-tagging reduces manual work by 90–95%, leaving only the most complex or ambiguous cases for human review. An important component of ensuring annotation quality is measuring the level of agreement among annotators, which is done using metrics such as Cohen's kappa or Fleiss's kappa [28]. Using these metrics helps reduce the subjectivity of annotation and improve its consistency in large-scale projects.

The application of deep active learning approaches has become a key tool for improving the efficiency of the annotation process. Unlike the traditional practice of annotating the entire dataset, this approach involves a phased, iterative selection of only those samples that are most valuable for further training of the model. The most common strategies include uncertainty-based sampling methods, where priority is given to examples with the lowest model confidence in its own prediction, as well as diversity-based sampling, which aims to cover a wide range of features and ensure the representativeness of the feature space. According to the results of a number of studies, Active Learning is capable of achieving an accuracy level comparable to that of a fully labeled dataset using only 15–20% of the initial volume of annotated data, which allows for a significant reduction in the financial and time costs of annotation [29].

The theoretical basis for the effectiveness of active learning is the principle of maximizing information gain at each iteration; for annotation, a sample is selected that minimizes the entropy of the posterior distribution of model parameters. Formally, the uncertainty-based sampling strategy implements approximate optimization of mutual information between the model parameters and the label of the selected sample. This theoretically guarantees faster training convergence compared to random sampling and explains the empirically observed 80–85% reduction in annotation volume.

A revolutionary shift in labeling practices was made possible by the emergence of large visual foundation models, which were trained on massive multimodal collections of billions of "image-text" pairs. One of the most significant developments is the Segment Anything Model (SAM) from Meta AI, which introduced the concept of "promptable segmentation" – segmentation activated via interactive or text prompts. SAM is capable of generating highly accurate object masks in real time, while demonstrating exceptional generalization ability to new domains without additional training, i.e., operating in zero-shot mode [30].

Modern automated annotation pipelines increasingly combine multiple models into a single pipeline. For example, Grounding DINO is used for open-vocabulary object detection based on text instructions. The resulting bounding boxes are then passed to SAM to generate accurate segmentation masks. This approach already enables the implementation of fully automated

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)
ISSN 3083-7650 (0135-1710) (print) Management Information System and Devices. 2026. No. 2 (189)

annotation systems that do not require pre-training on target domain classes [30]. In tasks requiring an immediate response, models from the YOLO family (YOLOv8, v9, v10) continue to play a key role; thanks to their high speed and accuracy, they are widely used for automatic pre-tagging, leaving humans with a minimal amount of work to verify and refine the results [31].

One of the key areas of development in modern annotation technologies is aimed at overcoming the problem of data scarcity by creating synthetic datasets. The use of highly realistic graphics engines, such as Unreal Engine or Unity, as well as specialized simulation platforms, including NVIDIA Omniverse and CARLA, enables the creation of complex scene configurations with automatically generated and highly accurate annotations. This approach is particularly important in the area of autonomous transportation, where collecting real-world data on rare or hazardous road situations is often impractical or risky [20].

5. Research Results

5.1 Formalization of the task and annotation scenarios

The research focuses on the task of semantic segmentation of objects in 2D images, as one of the most representative tasks for modern computer vision systems. Let there be a set of images D for which we need to construct a set of semantic masks Q .

$$D = \{I_i\}_{i=1}^{N}, \quad M = \{M_i\}_{i=1}^{N} \tag{3}$$

where I_i is a single input image from the sample; M_i is the corresponding semantic mask of the input image; N is the total number of images in the dataset.

The analysis considers three basic scenarios for generating annotations M :

– S_m fully manual annotation, which is considered the gold standard;

– S_{auto} fully automatic annotation based on YOLO/SAM models;

– S_{hyb} hybrid annotation with expert verification and correction.

The hybrid approach involves an iterative process in which the results of automatic generation are subject to expert review, refinement of object geometry, and correction of erroneous or missing segments. A generalized diagram of the hybrid pipeline is shown in Fig. 1.

The diagram shown illustrates a hybrid annotation pipeline in which the automatic generation of preliminary masks using YOLO and SAM neural networks is integrated with expert review in a human-in-the-loop mode. After the initial processing of complex 2D images of street scenes, a specialist performs a thorough correction of object geometry, removes false detections, and adds missing elements to achieve maximum accuracy. An important part of the process is an iterative feedback cycle, which allows for the rapid fine-tuning of system parameters and a gradual increase in its level of autonomy. The final output of the pipeline is high-quality Ground Truth-standard semantic masks, which provide an ideal balance between data preparation speed and reliability.

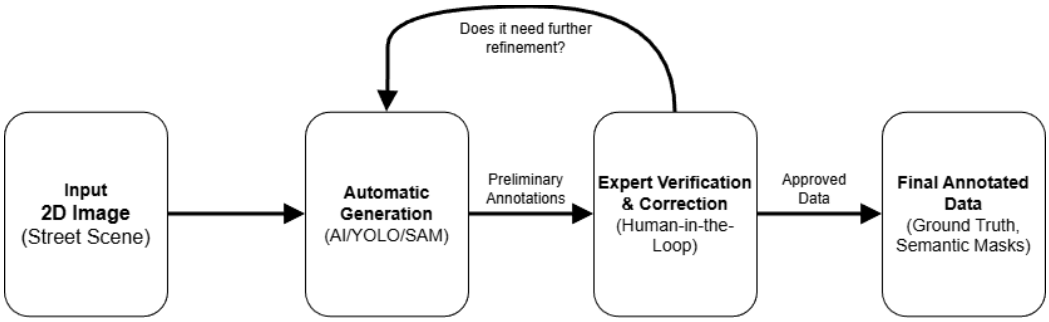


Fig 1. Block diagram of a hybrid image annotation pipeline

This approach mitigates the shortcomings of purely automated labeling, ensuring the creation of accurate datasets for autonomous driving systems while significantly optimizing time expenditure.

5.2. System of Quality and Labor-Intensity Metrics

A comprehensive assessment of the effectiveness of each scenario is based on two complementary groups of indicators: quality metrics and labor intensity metrics.

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

To quantitatively assess the quality of an annotation, the standard metric of mean Intersection over Union $(mIoU)$ is used:

$$mIoU = \frac{1}{C}\sum_{C=1}^{C}\frac{|P_C \cap G_C|}{|P_C \cup G_C|} \quad (4)$$

where P_C is the set of pixels assigned to class *c* in the generated or corrected mask; G_c is the corresponding set of pixels in the ground truth annotation; C is the number of classes.

The metric $mIoU$ characterizes the spatial consistency of segmentation masks and allows for a quantitative assessment of the accuracy of object boundary reproduction.

To formalize the labor intensity of the annotation process, the average time required to generate a mask for a single object is introduced:

$$T = T_{gen} + T_{corr} \quad (5)$$

where T_{gen} is the time required for automatic generation of the initial mask; T_{corr} is the time required for expert verification and correction.

It is worth noting that the component T_{corr} is not a constant and is directly dependent on the reliability of the automatic prediction. In the case of low model confidence, the cognitive load on the operator increases, which can lead to frustration and reduced consistency of actions. Thus, the effectiveness of the hybrid scenario is determined by the algorithm's ability to minimize the need for a complete mask reconstruction, limiting itself to its refinement.

To quantitatively assess the extent of manual intervention, the relative manual labor coefficient is used:

$$R = T_{corr}/T_{man} \quad (6)$$

where T_{man} is the average time required to fully annotate a single object manually.

Thus, the R metric reflects the proportion of human involvement in the hybrid scenario relative to fully manual annotation.

5.3. Quantitative analysis and comparison of scenarios

To derive numerical estimates, we used statistical aggregation of results from recent studies in the areas of interactive segmentation and progressive learning [6, 8, 10]. The validity of this approach is determined by the fact that the selected studies use a single quality metric $(mIoU)$, a common task of semantic segmentation of 2D images, and comparable architectural solutions based on the YOLO and SAM models, which ensures methodological consistency of the sample. Time values are normalized to a single object, allowing the results of different pipelines to be compared regardless of dataset size and hardware configuration. It should be noted that the presented metrics are indicative: specific values may vary depending on the complexity of the domain, hardware, and the skill level of the annotators. The unified metrics are presented in Table 2.

Table 2. Comparison of the effectiveness of image annotation scenarios

Annotation scenario	**Average time per object (sec)**	**Annotation quality (*mIoU*)**	**Manual labor**
Manual	~65.0	0.96 (Baseline)	100%
Automated (YOLO/SAM)	~0.5	0.82	0
Hybrid	~12.0	0.94	15–20%

To quantify the performance gain, an acceleration factor relative to fully manual annotation is introduced:

$$K_{speed} = T_{man}/T_{hyb} \quad (7)$$

For the values in Table 2, the acceleration factor is $K_{speed} = 65/12 \approx 5.4$. Thus, using the hybrid scenario allows the time required to generate a single semantic mask to be reduced by more than five times.

We define the loss of quality relative to manual annotation as:

$$\Delta IoU = mIoU_{man} - mIoU_{hyb} \quad (8)$$

For generalized data, the quality loss is $\Delta IoU = 0.96 - 0.94 = 0.02$.

This level of error is lower than typical values of inter-rater variability, which, according to [28], range from 0.0001 to $0.0003\,0.03 - 0.05$ s for semantic segmentation tasks when measured using Cohen's kappa coefficient. This allows us to conclude that the hybrid method provides annotation quality identical to or higher than that of a group of independent experts, but with better consistency of object boundaries due to the algorithmic stability of the YOLO and SAM models.

To analyze the trade-off between quality and time expenditure, a generalized performance metric is introduced:

$$E = mIoU/T \quad . \quad (9)$$

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)
ISSN 3083-7650 (0135-1710) (print) Management Information System and Devices. 2026. No. 2 (189)

Accordingly, for each scenario, we have the following metrics: $E_{man} = 0.96/65$, $E_{auto} = 0.82/0.5$, $E_{hyb} = 0.94/12$.

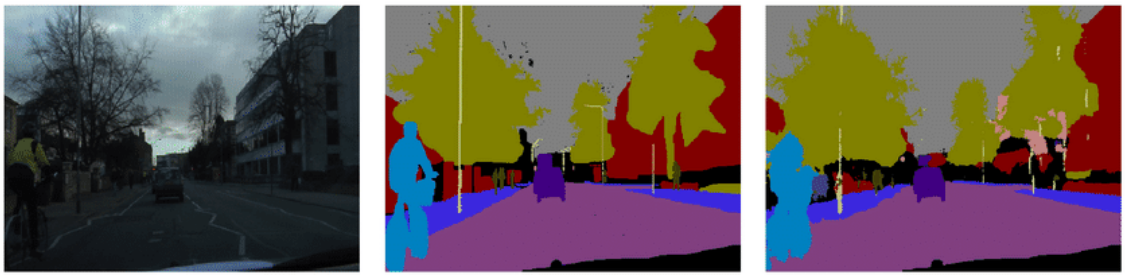
Although the "S_{auto}" scenario formally has the highest performance metric (E) due to its speed, it does not meet the requirements for creating a ground truth reference. For a more accurate assessment, it is advisable to introduce a weighting factor for accuracy criticality, n, which transforms the performance metric into the following form:

$$E_n = mIoU^n / T \ , \qquad (10)$$

where, for $n > 1$, the advantage of the hybrid scenario becomes evident, as it provides the necessary accuracy threshold for training models in critical domains such as autonomous driving or medical diagnostics.

Although the fully automatic scenario achieves the highest value of the indicator E due to its minimal processing time, it does not guarantee sufficient segmentation accuracy for tasks where correct object boundary geometry and semantic consistency of masks are critical. The hybrid scenario, in turn, provides an optimal balance between speed and quality, combining a significant reduction in annotation time with high performance metrics $mIoU$.

The results show that in hybrid mode, the operator performs only targeted corrections of problematic segmentation areas, specifically removing artifacts at object boundaries, refining the geometry of small elements, and correcting omissions in complex scene areas. A visual comparison of the results of automatic and hybrid segmentation using a complex urban scene as an example is shown in Fig. 2, where the reduction in noise and improvement in the spatial coherence of the masks following expert correction are clearly demonstrated.



a) input image b) result of human correction c) result of the model's output

Fig. 2. Comparison of automatic and hybrid segmentation:
a) input image; b) result of human correction; c) model output (contains noise) [32]

Quantifying the extent of manual labor is of particular practical importance. The value of the coefficient $R = 0.15 - 0.2$ indicates that in the hybrid scenario, manual labor is reduced by $80 - 85\%$ compared to fully manual annotation. This result is consistent with experimental observations from recent studies, which report the possibility of reducing the volume of manual annotation by $80 - 90\%$ when active learning and interactive segmentation are applied.

The results in Table 2 take on a non-trivial character when interpreted through the conditional optimization problem formalized in Section 4.1: $\min C_{\text{total}}$ subject to the constraint $Q(\hat{A}) \geq Q_{\min}$. The threshold $Q_{\min}$ is a domain-dependent parameter, and it is this parameter that determines the set of feasible scenarios for each specific application. Comparisons of admissibility based on aggregated data are presented in Table 3.

Table 3. *Admissibility of annotation scenarios depending on domain requirements*

Application type	$Q_{\min}$	S_{man} **(0.96)**	S_{auto} **(0.82)**	S_{hyb} **(0.94)**	**Optimal scenario**
Pre-filtration, prototyping	0.70	✓	✓	✓	Automatic (T ~ 0.5 s)
Training of general-purpose models	0.85	✓	✗	✓	Hybrid (T ~ 12 s)
Medical diagnostics	0.90	✓	✗	✓	Hybrid (T ~ 12 s)
Autonomous driving (Ground Truth)	0.95	✓	✗	✗	Manual (T ~ 65 s)

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

An analysis of Table 3 reveals a non-trivial result that refutes the naive assumption of a universal advantage of the hybrid approach. The advantage of the hybrid scenario is not universal but arises only within the range $0,82 < Q_{\min} \le 0,94$. At $Q_{\min} \le 0,82$, a fully automatic approach is sufficient, while at $Q_{\min} > 0,94$, only manual annotation ensures the required level of quality. It is precisely this range that covers the vast majority of real-world industrial and research tasks, as confirmed by the practice of creating datasets in the Waymo Open Dataset, nuScenes, and COCO projects, where the threshold quality requirements fall within the range $0.85-0.93$. Thus, the Human-in-the-Loop hybrid scenario is optimal for the broadest class of practical tasks, ranging from training production models to medical diagnostics, but not for the extreme cases of the spectrum.

To quantitatively assess the impact of the parameter n on the scenario rankings, the values of the weighted efficiency index $E_n = mIoU^n/T$ were calculated for each scenario at different levels of criticality, as shown in Table 4.

Table 4. *Sensitivity analysis of the weighted efficiency index* (E_n) *to the criticality parameter*

n	E_n **(manual)**	E_n **(automatic)**	E_n **(hybrid)**	**Rating**
1	0.015	1.640	0.078	auto ≫ hybrid > manual
2	0.014	1.345	0.074	auto ≫ hybrid > manual
5	0.013	0.741	0.061	auto > hybrid > manual
10	0.010	0.275	0.045	auto > hybrid > manual
20	0.007	0.038	0.024	auto > hybrid > manual
30	0.005	0.005	0.013	hyb > auto ≈ man

An analysis of Table 4 reveals two patterns. First, as n increases, the performance of the automatic scenario degrades significantly faster than that of the hybrid scenario: when $n=30$, the value of E_n (automatic) decreases by ~330 times compared to $n=1$, whereas E_n (hybrid) decreases by only ~6 times. This is due to the nonlinear nature of the penalty $mIoU^n$. The difference between $mIoU = 0,82$ and $mIoU = 0,94$ becomes decisive when raised to a high power. Second, even with moderate values of n, the weighted metric E_n does not shift the absolute ranking in favor of the hybrid scenario, indicating the need to supplement the quantitative analysis with a threshold constraint $Q_{\min}$, as shown in Table 3. It is the combination of the threshold constraint $Q_{\min}$ with the metric E_n that provides a comprehensive criterion for selecting an annotation strategy.

Thus, the analysis conducted allows us to formulate a well-founded criterion for selecting an annotation strategy: for a specific application with the threshold requirement $Q_{\min}$, the optimal scenario is the one with the minimum time T among those that satisfy the constraint $mIoU \ge Q_{\min}$. According to the aggregated data, the hybrid scenario is optimal for the widest range of practical tasks $(0.82 < Q_{\min} \le 0.94)$, providing a fivefold speedup relative to the manual approach while reducing the amount of manual labor to $15-20\%$. At the same time, the results of the sensitivity analysis (Table 4) demonstrate that the indicator E_n is not a self-sufficient criterion and requires supplementation with the threshold constraint $Q_{\min}$, which underscores the practical significance of the proposed formalization of the optimization problem.

At the same time, the study has several limitations. First, the quantitative analysis is based on the aggregation of results from independent studies with different experimental conditions, which makes it impossible to statistically test the significance of differences between scenarios. Second, the analysis is limited to the task of semantic segmentation of 2D images. For other types of annotation (3D point clouds, video tracking), quantitative ratios may differ significantly. Third, the values of time metrics depend on hardware configuration and the skill level of operators, which limits the direct transfer of results to specific production environments. However, these limitations do not diminish the value of the obtained results, since the proposed formalization of the optimization problem with a threshold constraint is invariant with respect to specific numerical values and remains methodologically valid for any domains and configurations. The quantitative estimates presented in Tables 2–4 should be considered representative

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)
ISSN 3083-7650 (0135-1710) (print) Management Information System and Devices. 2026. No. 2 (189)

indicative values demonstrating general patterns of the relationship between speed and quality for the three classes of scenarios.

6. Conclusions

Image annotation is a critically important step in the data preparation process for computer vision systems. In the context of the Data-Centric AI paradigm, it is the quality of the annotations, rather than the architectural complexity of the models, that determines the accuracy and reliability of predictions. The annotation process provides a semantic interpretation of "raw" pixel data, forming the ground truth annotations necessary for training and validating algorithms.

The formalization of the annotation task as a conditional optimization problem and the comparative analysis conducted indicate that the modern technological landscape is undergoing a gradual shift in approaches, moving from exclusively manual labor toward the implementation of hybrid Human-in-the-Loop systems. Such systems synergistically combine automated pre-labeling using fundamental models, notably SAM and YOLO, with subsequent expert verification. A quantitative analysis of scenarios demonstrated that the hybrid approach provides a fivefold acceleration of the process from 65 to 12 seconds per object while maintaining high quality. The recorded accuracy loss of 0.02 relative to the manual ground truth falls within typical inter-annotator variations, confirming the reliability of the hybrid method for creating professional datasets.

The economic feasibility of implementing such pipelines lies in their ability to reduce the amount of manual labor to. However, the effectiveness of such collaboration depends on the reliability of automatic prompts: low model confidence exponentially increases correction time and the cognitive load on the operator. For critical areas, such as autonomous driving or medical diagnostics, it is rational to use a weighted performance metric that prioritizes accuracy over speed.

The effectiveness of hybrid systems is determined not only by algorithms but also by the operator's trust in AI and the ergonomics of the interfaces. Future research aims to improve algorithms for generating synthetic samples to address the "domain gap" problem and to develop adaptive interfaces that minimize annotator fatigue.

Conflict of Interest

The authors declare that they have no conflicts of interest regarding this study, including financial, personal, authorship, or other conflicts that could influence the study and its results presented in this article.

Funding

The study was conducted without financial support.

Data Availability

Data will be provided upon reasonable request.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence technologies in the creation of this work.

References

1. Song, H. et al. (2020), "Weighted Topic Model Learned From Local Semantic Space for Automatic Image Annotation", *IEEE Access*, Vol. 8, pp. 76411–76422. DOI: https://doi.org/10.1109/ACCESS.2020.2989200
2. Monarch, R. M. (2021), "Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI", *New York*, 424 p.
3. Montezuma, D. et al. (2022), "Annotating for Artificial Intelligence Applications in Digital Pathology: A Practical Guide for Pathologists and Researchers", *United States & Canadian Academy of Pathology*, Vol. 36, pp. 100086. DOI: https://doi.org/10.1016/j.modpat.2022.100086
4. Demrozi, F. et al. (2023), "A Comprehensive Review of Automated Data Annotation Techniques in Human Activity Recognition", *Cornell University arXiv*. DOI: https://doi.org/10.48550/arXiv.2307.05988
5. Sun, Q. et al. (2025), "DocSpiral: A Platform for Integrated Assistive Document Annotation through Human-in-the-Spiral", *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, Vol. 3, pp. 267–274. DOI: https://doi.org/10.18653/v1/2025.acl-demo.26

6. Li, X. et al. (2023), "HAL-IA: A Hybrid Active Learning framework using Interactive Annotation for medical image segmentation", *Medical Image Analysis*, Vol. 88, pp. 102862. DOI: https://doi.org/10.1016/j.media.2023.102862
7. Huang, Y. et al. (2025), "Application of human-in-the-loop hybrid augmented intelligence approach in security inspection system", *Frontiers in Artificial Intelligence*, Vol. 8. DOI: https://doi.org/10.3389/frai.2025.1518850
8. Ramanathan, A. S., Oyelere, S. S., Baruah, N. (2025), "Hybrid intelligence approach for detecting synthetic art", Human-Intelligent *Systems Integration*, Vol. 7, pp. 325–340. DOI: https://doi.org/10.1007/s42454-025-00081-z
9. Yadnakudige Subramanya, S. et al. (2025), "Human-in-the-Loop Annotation for Image-Based Engagement Estimation: Assessing the Impact of Model Reliability on Annotation Accuracy", *Human-Computer Interaction. HCII 2025. Lecture Notes in Computer Science*, Vol. 15770, pp. 169–186. DOI: https://doi.org/10.1007/978-3-031-93864-1_12
10. Al-Turki, D. et al. (2023), "The power of progressive active learning in floorplan images for energy assessment", *Scientific reports*, Vol. 16238. DOI: https://doi.org/10.1038/s41598-023-42276-x
11. Sager, C., Janiesch, C., Zschech, P. (2021), "A survey of image labelling for computer vision applications", *Journal of Business Analytics*. DOI: https://doi.org/10.48550/arXiv.2104.08885
12. Bachani, V. et al. (2024), "Image Segmentation Survey: Classical and Deep Learning Methods", *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pp. 1–6. DOI: https://doi.org/10.1109/ICECET61485.2024.10698602
13. Song, H. et al. (2023), "Learning From Noisy Labels With Deep Neural Networks: A Survey", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 34, pp. 8135-8153. DOI: https://doi.org/10.1109/TNNLS.2022.3152527
14. Whang, S., Roh, Y., Song, H. (2023), "Data collection and quality challenges in deep learning: a data-centric AI perspective", *The VLDB Journal 32*, pp. 791–813. DOI: https://doi.org/10.1007/s00778-022-00775-9
15. Zou, Z. et al. (2023), "Object Detection in 20 Years: A Survey", *Proceedings of the IEEE*, Vol. 111, pp. 257–276. DOI: https://doi.org/10.1109/JPROC.2023.3238524
16. Ciaparrone, G. et al. (2020), "Deep learning in video multi-object tracking: A survey", *Science Direct, Neurocomputing*, Vol. 381, pp. 61–88. DOI: https://doi.org/10.1016/j.neucom.2019.11.023
17. Nagiu, A. S. et al. (2024), "3D Object Detection for Autonomous Driving: A Comprehensive Review", *2024 6th International Conference on Computing and Informatics (ICCI)*, pp. 01–11. DOI: https://doi.org/10.1109/ICCI61671.2024.10485120
18. Minaee, S. et al. (2022), "Image Segmentation Using Deep Learning: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, pp. 3523–3542. DOI: https://doi.org/10.1109/TPAMI.2021.3059968
19. Wu, X. et al. (2022), "A survey of human-in-the-loop for machine learning", *Science Direct, Future Generation Computer Systems*, Vol. 135, pp. 364–381. DOI: https://doi.org/10.1016/j.future.2022.05.014
20. Mumuni, A., Mumuni, F., Gerrar, N. K. (2024), "A Survey of Synthetic Data Augmentation Methods in Machine Vision", *Springer Nature Link, Machine Intelligence Research*, Vol. 21, pp. 831–869. DOI: https://doi.org/10.1007/s11633-022-1411-7
21. Chen, Z. M. et al. (2019), "Multi-Label Image Recognition With Graph Convolutional Networks", *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5172–5181. DOI: https://doi.org/10.1109/CVPR.2019.00532
22. Padilla, R. et al. (2021), "A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit", *Electronics*, Vol. 10, pp. 279. DOI: https://doi.org/10.3390/electronics10030279
23. Ajmera, F. et al. (2021), "Survey on Object Detection in Aerial Imagery", *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, pp. 1050–1055. DOI: https://doi.org/10.1109/ICICV50876.2021.9388517
24. Gu, W., Bai, S., Kong, L. (2022), "A review on 2D instance segmentation based on deep neural networks", *Science Direct, Image and Vision Computing*, Vol. 120, pp. 104401. DOI: https://doi.org/10.1016/j.imavis.2022.104401
25. Elharrouss, O. et al. (2021), "Panoptic Segmentation: A Review", *Cornell University arXiv*. DOI: https://doi.org/10.48550/arXiv.2111.10250
26. Chen, Y., Tian, Y., He, M. (2020), "Monocular human pose estimation: A survey of deep learning-based methods", *Science Direct, Computer Vision and Image Understanding*, Vol. 192, p. 102897. DOI: https://doi.org/10.1016/j.cviu.2019.102897
27. Stefanini, M. et al. (2021), "From Show to Tell: A Survey on Image Captioning", *Cornell University arXiv*. DOI: https://doi.org/10.48550/arXiv.2107.06912
28. Roh, Y., Heo, G., Whang, S. E. (2021), "A Survey on Data Collection for Machine Learning: A Big Data – AI Integration Perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33, pp. 1328–1347. DOI: https://doi.org/10.1109/TKDE.2019.2946162
29. Gu, F. et al. (2021), "A Survey on Deep Learning for Human Activity Recognition", *Association for Computing Machinery*, Vol. 54, p. 34. DOI: https://doi.org/10.1145/3472290

ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189)*
ISSN 3083-7650 (0135-1710) (print) *Management Information System and Devices. 2026. No. 2 (189)*

30. Kirillov, A. et al. (2023), "Segment Anything", *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3992–4003. DOI: https://doi.org/10.1109/ICCV51070.2023.00371
31. Terven, J. et al. (2023), "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS", *Machine Learning and Knowledge Extraction*, Vol. 5, p. 1680–1716. DOI: https://doi.org/10.3390/make5040083
32. Corbière, C. et al. (2019), "Addressing Failure Prediction by Learning Model Confidence", *Computer Vision and Pattern Recognition*, arXiv. DOI: https://doi.org/10.48550/arXiv.1910.04851

Received (Надійшла) 19.02.2026
Accepted for publication (Прийнята до друку) 15.04.2026
Publication date (Дата публікації) 29.05.2026

Відомості про авторів / About the Authors

Кучапін Матвій Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна;
Matvii Kuchapin – Kharkiv National University of Radio Electronics, Postgraduate Student at the Software Engineering Department, Kharkiv, Ukraine;
e-mail: matvii.kuchapin@nure.ua
ORCID ID: http://orcid.org/0009-0006-1953-2893

Смеляков Кирило Сергійович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри програмної інженерії, Харків, Україна;
Kyrylo Smelyakov – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Head at the Department of Software Engineering, Kharkiv, Ukraine;
e-mail: kyrylo.smelyakov@nure.ua
ORCID ID: http://orcid.org/0000-0001-9938-5489

Чуприна Анастасія Сергіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, Харків, Україна;
Anastasiya Chupryna – Candidate of Technical Sciences, Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of the Department of Software Engineering, Kharkiv, Ukraine;
e-mail: anastasiya.chupryna@nure.ua
ORCID ID: http://orcid.org/0000-0003-0394-9900

Лученко Сергій Васильович – кандидат технічних наук, Національний аерокосмічний університет "Харківський авіаційний інститут", старший викладач кафедри інженерії програмного забезпечення, Харків, Україна;
Sergiy Luchenko – Candidate of Technical Sciences, National Aerospace University "Kharkiv Aviation Institute", Senior Lecturer of the Department of Software Engineering, Kharkiv, Ukraine;
e-mail: s.luchenko@khai.edu
ORCID ID: http://orcid.org/0009-0006-9606-5774

КОМПЛЕКСНИЙ АНАЛІЗ МЕТОДІВ ТА ІНСТРУМЕНТІВ ГІБРИДНОГО АНОТУВАННЯ ЗОБРАЖЕНЬ У СИСТЕМАХ КОМП'ЮТЕРНОГО ЗОРУ

Предметом дослідження є методи, алгоритми та програмні інструменти анотування візуальних даних у системах комп'ютерного зору в межах парадигми Data-Centric AI. У статті проаналізовано процеси структурування неструктурованої інформації, де якість розмітки визначає точність моделей штучного інтелекту. Виявлено обмеження ручних методів і ризики систематичних помилок за умови повної автоматизації. **Мета дослідження** – комплексний

Автоматизовані системи управління та прилади автоматики. 2026. № 2 (189) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 2 (189) *ISSN 3083-7650 (0135-1710) (print)*

аналіз сучасних методів та інструментів анотування зображень у системах комп'ютерного зору та кількісне оцінювання доцільності гібридних стратегій Human-in-the-Loop для підвищення ефективності формування якісних наборів даних у межах парадигми Data-Centric AI. У статті необхідно виконати такі **завдання**: систематизувати типи анотувань від класифікації до паноптичної сегментації та розмітки 3D-сцен; розглянути інструментарій на основі моделей Segment Anything Model і GroundingDINO; здійснити порівняльне оцінювання ручного, автоматичного й гібридного сценаріїв за показниками точності (mIoU) та трудомісткості; визначити проблеми довіри оператора до алгоритмічних підказок та ергономіки взаємодії. **Методи**: систематичний порівняльний аналіз гібридних стратегій Human-in-the-Loop, крос-доменний синтез результатів досліджень активного навчання та інтерактивної сегментації, формалізація показників прискорення, якості й відносного обсягу ручної праці. **Досягнуті результати.** Доведено, що гібридний конвеєр (YOLO/SAM та експертна корекція) забезпечує прискорення процесу в 5,4 раза. Час формування семантичної маски об'єкта скорочується з 65 до 12 с за умови збереження якості $mIoU = 0{,}94$, де втрата щодо еталона становить лише 0,02. Установлено, що гібридний сценарій є оптимальним у діапазоні порогових вимог $0{,}82 < Q_{min} \le 0{,}94$, що охоплює найширший клас практичних завдань – від навчання виробничих моделей до медичної діагностики. Виявлено пряму залежність якості роботи анотатора від надійності автоматичних підказок, що підтверджує важливість прозорості алгоритмів. **Висновки.** Гібридне анотування є оптимальною стратегією для створення Ground Truth у критичних доменах (автономне водіння, медицина), що забезпечує баланс швидкості й точності. Запропонована формалізація задачі оптимізації з пороговим обмеженням Q_{min} дає змогу обгрунтовано обирати сценарій анотування для конкретного домену. Перспективи подальших досліджень полягають у вдосконаленні методів генерації синтетичних даних у симульованих середовищах і розробленні адаптивних інтерфейсів для зниження когнітивного навантаження на експертів.

Ключові слова: комп'ютерний зір; анотування зображень; активне навчання; сегментація; синтетичні дані.

Бібліографічні описи / Bibliographic descriptions

Кучапін М. Ю., Смеляков К. С., Чуприна А. С., Лученко С. В. Комплексний аналіз методів та інструментів гібридного анотування зображень у системах комп'ютерного зору. *Автоматизовані системи управління та прилади автоматики*. 2026. № 2 (189). С. 182–195. DOI: https://doi.org/10.30837/0135-1710.2026.189.182

Kuchapin, M., Smelyakov, K., Chupryna, A., Luchenko, S. (2026), "Comprehensive analysis of methods and tools for hybrid image annotation in computer vision systems", *Management Information System and Devices*, No. 2 (189), P. 182–195. DOI: https://doi.org/10.30837/0135-1710.2026.189.182