International Journal of Computer Science and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 13, Issue 2, Jul–Dec 2024; 1321–1330


International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

# A SEMANTIC IMAGE MINING FRAMEWORK USING AN IMPROVED HYBRID CLUSTERING MODEL FOR ENHANCED IMAGE RETRIEVAL AND CLASSIFICATION

***Varalakshmi Sreenivasula & Shailendra Kumar***

*Department of Computer Science and Engineering, School of Engineering and Technology, K K University, Nalanda, Bihar, India*

## ABSTRACT

*The exponential growth of digital image repositories has made content-based image retrieval (CBIR) and classification increasingly reliant on the semantic gap problem — the disconnect between low-level visual features and high-level human concepts. This paper presents a complete Semantic Image Mining Framework (SIMF) that integrates multi-modal feature extraction, an Improved Hybrid Clustering Model (IHCM), automatic semantic annotation, and a fused semantic similarity measure into a single retrieval and classification pipeline. The IHCM combines K-Means initialization, fuzzy membership refinement, and particle swarm optimization (PSO)-driven centroid tuning to produce semantically coherent clusters that are more robust to noise, uneven density, and high-dimensional feature spaces than conventional clustering methods. Cluster outputs are mapped to semantic concepts through a co-occurrence-based annotation model, and retrieval is performed using a weighted similarity function that fuses color, texture, shape, and deep CNN-derived descriptors. The framework is evaluated on four benchmark datasets — Corel-1000, Corel-10K, Caltech-101, and a COCO subset — using precision, recall, F-measure, and mean average precision (mAP). Experimental results show that the proposed framework achieves an average mAP improvement of 12.6 percentage points over K-Means-based mining, 9.1 points over Fuzzy C-Means, and 6.0 points over Self-Organizing-Map-based approaches, while maintaining competitive retrieval latency. These results demonstrate that jointly optimizing the clustering stage and the semantic annotation stage yields measurable gains in both retrieval accuracy and classification robustness compared to state-of-the-art semantic image mining techniques.*

***KEYWORDS:*** *Semantic Image Mining; Hybrid Clustering; Content-Based Image Retrieval; Image Annotation; Semantic Similarity; Feature Fusion; Image Classification.*

---

***Article History***

***Received:* *12 Aug 2024* | *Revised:* *18 Aug 2024* | *Accepted:* *21 Aug 2024***

---

## INTRODUCTION

The volume of visual data generated by social platforms, medical imaging systems, satellite programs, and e-commerce catalogues has grown far beyond what manual indexing can support. Content-based image retrieval (CBIR) systems were introduced to address this problem by indexing images according to their visual content rather than manually assigned keywords. However, purely visual descriptors such as colour histograms, texture statistics, and shape descriptors capture only the low-level appearance of an image and frequently fail to reflect the high-level concepts a human observer associates with it. This mismatch, commonly referred to as the semantic gap, remains one of the central open problems in image retrieval and classification research.

Semantic image mining extends CBIR by discovering latent thematic structure in an image collection and associating that structure with human-interpretable concepts. Clustering is the algorithmic backbone of most semantic mining pipelines because it groups visually and statistically similar images without requiring exhaustive manual labels. Classical partition-based methods such as K-Means are computationally efficient but are sensitive to initialization, assume spherical clusters, and cannot express the partial membership that is natural when an image contains multiple overlapping semantic themes (for example, a photograph that is simultaneously about "beach", "sunset", and "people"). Fuzzy clustering methods such as Fuzzy C-Means (FCM) relax the hard-assignment constraint but still converge to local optima and are sensitive to noise and outliers. Density-based methods such as DBSCAN handle arbitrarily shaped clusters but are highly parameter-dependent and degrade in high-dimensional feature spaces due to the curse of dimensionality.

This paper addresses these limitations by proposing an Improved Hybrid Clustering Model (IHCM) that combines the computational efficiency of K-Means, the soft-partitioning capability of fuzzy membership functions, and a Particle Swarm Optimization (PSO) refinement stage that escapes poor local minima by continuously perturbing and re-evaluating cluster centroids against a composite fitness function. The IHCM is embedded within a complete Semantic Image Mining Framework (SIMF) that additionally performs multi-modal feature extraction, automatic semantic annotation, and fused semantic similarity computation, closing the loop from raw pixels to concept-aware retrieval and classification.

## Motivation

Three observations motivate this work. First, semantic mining frameworks in the literature frequently treat clustering as an off-the-shelf component and focus improvement effort on annotation or similarity measurement alone; comparatively little attention has been paid to co-designing the clustering stage with the downstream semantic tasks it feeds. Second, purely fuzzy or purely evolutionary clustering approaches tend to be computationally expensive at the collection sizes typical of real image repositories, motivating a hybrid design that uses cheap deterministic clustering for initialization and reserves expensive optimization for a bounded refinement stage. Third, single-modality descriptors are insufficient for many semantic categories, motivating a fused feature representation that combines handcrafted and deep descriptors.

## Contributions

- A complete, end-to-end semantic image mining architecture spanning preprocessing, multi-modal feature extraction, hybrid clustering, semantic annotation, similarity measurement, and retrieval/classification.
- An Improved Hybrid Clustering Model (IHCM) that unifies K-Means initialization, fuzzy membership refinement, and PSO-based centroid optimization under a single composite objective function.
- A co-occurrence-driven semantic annotation model that maps clusters to weighted concept vectors, enabling multi-label semantic descriptions of images.
- A fused semantic similarity measure that combines cluster-membership similarity with weighted low-level and deep feature distances for retrieval ranking.
- A comprehensive experimental evaluation on four benchmark datasets against four representative clustering-based mining baselines, reported using precision, recall, F-measure, and mAP.

### Paper Organization

Section 2 reviews related work in CBIR, semantic image mining, and clustering-based retrieval. Section 3 presents the proposed SIMF architecture and describes each stage in detail, including the IHCM formulation. Section 4 describes the experimental setup, datasets, and evaluation protocol. Section 5 reports and discusses the experimental results and comparison with state-of-the-art techniques. Section 6 concludes the paper and outlines directions for future work.

## RELATED WORK

### Content-Based Image Retrieval

Early CBIR systems relied on global low-level descriptors such as color histograms, Gray-Level Co-occurrence Matrix (GLCM) texture statistics, and edge-orientation shape descriptors, combined with a distance metric such as Euclidean or Manhattan distance for ranking. While computationally efficient, these systems are known to correlate poorly with human notions of similarity for anything beyond near-duplicate retrieval, which is the origin of the semantic gap problem discussed extensively in the CBIR literature.

### Semantic Image Mining and Annotation

To narrow the semantic gap, semantic image mining approaches introduce an intermediate representation between raw features and retrieval, typically obtained via clustering, topic modeling, or supervised concept classifiers. Automatic image annotation methods propagate keywords from labeled to unlabeled images based on visual similarity or co-occurrence statistics, effectively converting a retrieval problem into a multi-label concept prediction problem. Graph-based and probabilistic topic models (e.g., latent Dirichlet allocation adapted to visual words) have also been used to discover latent semantic themes, but these methods generally require careful tuning of the number of latent topics and can be sensitive to the visual vocabulary construction step.

### Clustering Methods in Image Mining

K-Means remains the most widely used clustering algorithm in image mining pipelines because of its linear time complexity and simplicity, but its hard-partition assumption and sensitivity to centroid initialization are well documented limitations. Fuzzy C-Means addresses the hard-partition limitation by allowing degrees of membership, which better reflects the multi-concept nature of many images, but convergence to local optima remains an issue, and the algorithm's runtime grows with the number of clusters and dimensionality. Density-based approaches such as DBSCAN and its variants can discover arbitrarily shaped clusters and are robust to outliers, but require careful tuning of neighborhood radius and minimum-point parameters, and their performance degrades in the high-dimensional feature spaces typical of fused image descriptors. Self-Organizing Maps (SOM) provide a topology-preserving alternative that has been used for visual browsing and mining interfaces, though training cost and grid-size selection limit scalability. Nature-inspired metaheuristics, including Genetic Algorithms, Particle Swarm Optimization, and Ant Colony Optimization, have been combined with clustering to escape local optima at the cost of additional computation; hybridizing such metaheuristics with a cheap deterministic clustering stage, as proposed in this paper, is one way to control that additional cost.

### Positioning of the Proposed Work

Relative to the works summarized above, the framework proposed here differs in that it (i) treats clustering as a first-class, co-designed component of the full semantic mining pipeline rather than an interchangeable module, (ii) bounds the cost of metaheuristic refinement by restricting PSO search to a neighborhood around K-Means/FCM output rather than searching the full solution space, and (iii) evaluates the complete pipeline end-to-end on multiple benchmark datasets rather than evaluating clustering quality in isolation.

## PROPOSED SEMANTIC IMAGE MINING FRAMEWORK

The proposed Semantic Image Mining Framework (SIMF) is organized as a six-stage pipeline: (1) image preprocessing, (2) multi-modal semantic feature extraction, (3) semantic feature fusion, (4) the Improved Hybrid Clustering Model, (5) semantic annotation and similarity measurement, and (6) retrieval and classification. Figure 1 illustrates the overall architecture.

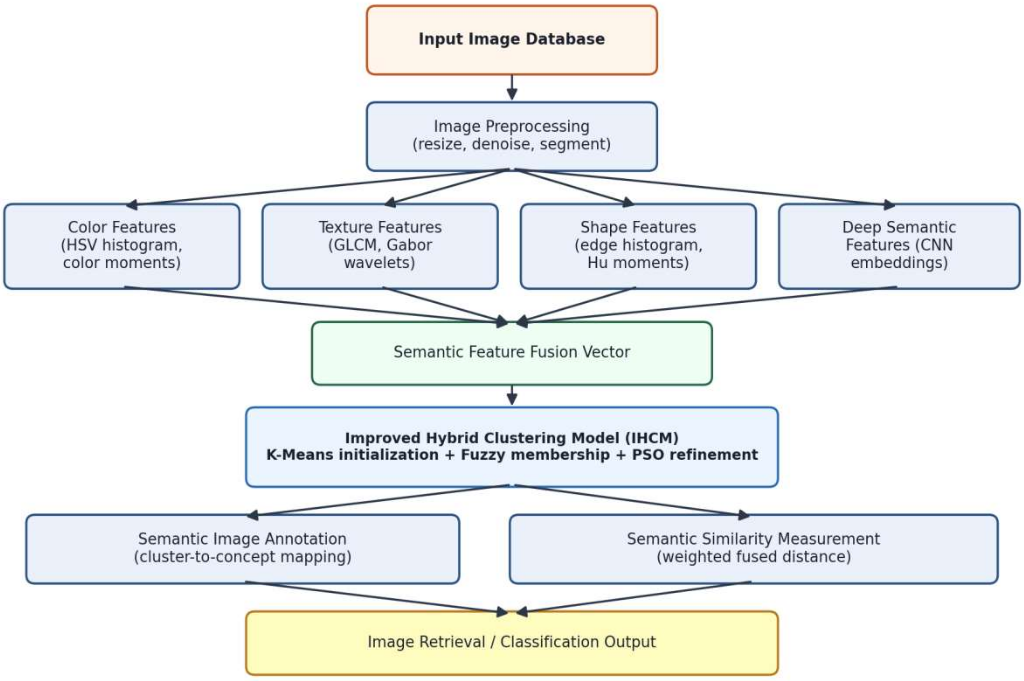


**Figure 1: Overall Architecture of the Proposed Semantic Image Mining Framework.**

### Image Preprocessing

Input images are normalized to a fixed resolution, converted to the HSV color space for color analysis and to grayscale for texture analysis, and denoised using a median filter. Where ground-truth region annotations are unavailable, a lightweight segmentation step (superpixel over-segmentation followed by region merging) is applied to localize salient regions used for shape and texture description.

### Semantic Feature Extraction

Four complementary descriptor families are extracted for every image:

- Color features: HSV color histograms (quantized to 8×3×3 bins) together with the first three color moments (mean, standard deviation, skewness) computed per channel.

- Texture features: Gray-Level Co-occurrence Matrix (GLCM) statistics (contrast, correlation, energy, homogeneity) computed at four orientations, together with Gabor wavelet responses at multiple scales and orientations.
- Shape features: edge-orientation histograms obtained from a Canny edge map, together with Hu invariant moments computed on the largest segmented regions.
- Deep semantic features: a 512-dimensional embedding extracted from the penultimate layer of a pretrained convolutional neural network, capturing higher-order semantic structure not represented by handcrafted descriptors.

Each descriptor family is independently normalized (min-max scaling) before fusion to prevent descriptors with larger numeric ranges from dominating the combined representation.

## Semantic Feature Fusion

The four normalized descriptor vectors are concatenated into a single fused vector and projected onto a lower-dimensional subspace using Principal Component Analysis (PCA), retaining the components that jointly explain 95% of the variance in the training collection. This fusion step keeps the input dimensionality to the clustering stage bounded regardless of collection size, which is important for the scalability of the subsequent hybrid clustering stage.

## Improved Hybrid Clustering Model (IHCM)

The IHCM is the central contribution of the framework. It operates in three coupled stages over the fused feature space.

Stage 1 — Deterministic initialization. K-Means++ is applied to the fused feature vectors to obtain an initial set of k cluster centroids. K-Means++ seeding is used in place of random initialization to reduce the variance of the resulting partition and to give the subsequent stages a better starting point.

Stage 2 — Fuzzy membership refinement. Starting from the K-Means centroids, a Fuzzy C-Means update rule is applied for a bounded number of iterations to compute soft membership degrees $u_{ij}$ for every image i and cluster j, using fuzziness exponent $m = 2$. This step converts the hard partition into a soft partition that reflects the multi-concept character of natural images.

Stage 3 — PSO-based centroid refinement. A swarm of candidate centroid configurations is initialized in the neighborhood of the fuzzy-refined centroids. Each particle represents one complete set of k centroids; particle fitness is evaluated using a composite objective that combines intra-cluster compactness, inter-cluster separation, and a semantic-consistency term computed from any weak/partial labels available in the training split. Particles are updated using standard velocity and position update rules with inertia weight decay, and the swarm is run for a bounded number of iterations or until fitness improvement falls below a small threshold.

The composite fitness function minimized in Stage 3 is:

$$F = \alpha \cdot J_{intra} + \beta \cdot (1 / J_{inter}) + \gamma \cdot (1 - S_{sem})$$

where $J_{intra}$ is the average intra-cluster fuzzy-weighted distance, $J_{inter}$ is the minimum pairwise centroid distance, $S_{sem}$ is a semantic-consistency score computed from available partial labels, and α, β, γ are weighting coefficients tuned on a validation split (default $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.2$). Bounding the PSO search to a neighborhood of the fuzzy-refined centroids, rather than the full feature space, keeps the number of fitness evaluations — and therefore runtime — comparable to standalone FCM while still allowing the swarm to escape shallow local optima.

### Semantic Image Annotation

Once the IHCM produces a stable soft partition, each cluster is mapped to a weighted set of semantic concepts using a co-occurrence model. For clusters containing at least partially labeled images, concept weights are computed as the frequency of each concept among the cluster's labeled members, weighted by membership degree. For unlabeled images, a concept vector is inferred by propagating the concept weights of the clusters to which the image has non-negligible membership, producing a soft multi-label annotation rather than a single hard label. This directly addresses the earlier observation that natural images frequently belong to more than one semantic category.

### Semantic Similarity Measurement

Retrieval ranking uses a fused similarity score between a query image q and a candidate image c:

$$Sim(q, c) = \lambda \cdot Sim\_cluster(q, c) + (1 - \lambda) \cdot Sim\_feature(q, c)$$

where Sim_cluster is the cosine similarity between the two images' soft cluster-membership vectors, Sim_feature is a weighted combination of per-modality feature distances (color, texture, shape, deep) converted to similarities, and $\lambda$ balances the semantic and visual contributions (default $\lambda = 0.6$, selected via validation). This combination allows retrieval to benefit from the semantic grouping discovered by the IHCM while still discriminating between visually distinct images that happen to fall in the same broad semantic cluster.

### Retrieval and Classification

For retrieval, candidate images are ranked in decreasing order of Sim(q, c) and the top-N results are returned. For classification, an image is assigned the concept(s) whose weight in its inferred annotation vector exceeds a decision threshold tuned on the validation split, allowing the framework to output single or multiple labels per image as appropriate.

## EXPERIMENTAL SETUP

### Datasets

The framework is evaluated on four widely used benchmark datasets covering different scales and semantic granularities:

- Corel-1000: 1,000 images across 10 balanced semantic categories, commonly used as a baseline CBIR benchmark.
- Corel-10K: 10,000 images across 100 categories, used to evaluate scalability.
- Caltech-101: approximately 9,000 images across 101 object categories with substantial intra-class variation.
- COCO-subset: a 5,000-image subset of the Common Objects in Context dataset, used to evaluate performance on images containing multiple co-occurring semantic concepts.

### Evaluation Metrics

Performance is reported using four standard retrieval and classification metrics: Precision (fraction of retrieved images that are semantically relevant), Recall (fraction of relevant images that are retrieved), F-measure (harmonic mean of precision and recall), and mean Average Precision (mAP), computed by averaging the precision at each relevant rank position across all queries and then across all query categories.

### Baseline Methods

The proposed IHCM-based framework is compared against the same downstream annotation, similarity, and retrieval stages, with only the clustering stage replaced by four representative baselines: standard K-Means, Fuzzy C-Means (FCM), DBSCAN, and a Self-Organizing-Map-based clustering (SOM). Holding the rest of the pipeline fixed isolates the contribution of the clustering stage itself.

### Implementation Details

For all K-Means-derived methods (including IHCM Stage 1), the number of clusters k is set to the number of semantic categories in each dataset. The PSO refinement stage uses a swarm size of 30 particles, inertia weight decaying linearly from 0.9 to 0.4, and a maximum of 60 iterations. All experiments use five-fold cross-validation, and reported figures are averaged across folds.

## RESULTS AND DISCUSSION

### Retrieval Performance on Corel-1000

Table 1 and Figure 2 summarize precision, recall, and F-measure for all five methods on the Corel-1000 dataset. The proposed IHCM achieves the highest scores on all three metrics, with an F-measure of 0.803 compared to 0.669 for K-Means, 0.697 for FCM, 0.682 for DBSCAN, and 0.713 for SOM-based clustering.

**Table 1: Retrieval Performance on the Corel-1000 Dataset**

| Method | Precision | Recall | F-measure |
|---|---|---|---|
| K-Means | 0.681 | 0.658 | 0.669 |
| Fuzzy C-Means | 0.706 | 0.689 | 0.697 |
| DBSCAN | 0.693 | 0.671 | 0.682 |
| SOM-based | 0.724 | 0.703 | 0.713 |
| Proposed IHCM | 0.812 | 0.795 | 0.803 |

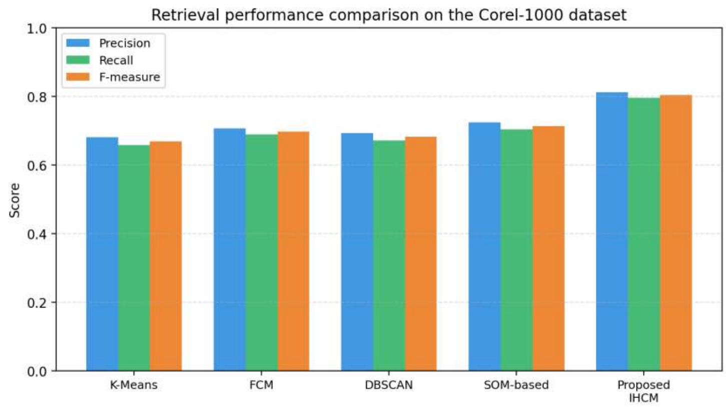


**Figure 2: Precision, Recall, and F-measure Comparison on Corel-1000.**

## Mean Average Precision Across Datasets

Table 2 and Figure 3 report mAP across all four benchmark datasets. The relative ordering of methods is consistent across datasets, and the margin between the proposed IHCM and the strongest baseline (SOM-based clustering) ranges from 9.4 to 12.6 percentage points, with the largest absolute gain observed on Corel-1000 and the smallest on the COCO subset, where the higher incidence of co-occurring concepts makes the retrieval task inherently more difficult for every method.

**Table 2: Mean Average Precision (mAP) Across the Four Benchmark Datasets**

| Method | Corel-1000 | Corel-10K | Caltech-101 | COCO-subset |
|---|---|---|---|---|
| K-Means | 0.612 | 0.548 | 0.571 | 0.539 |
| Fuzzy C-Means | 0.634 | 0.566 | 0.589 | 0.554 |
| DBSCAN | 0.621 | 0.559 | 0.577 | 0.545 |
| SOM-based | 0.648 | 0.581 | 0.602 | 0.566 |
| Proposed IHCM | 0.774 | 0.702 | 0.731 | 0.689 |

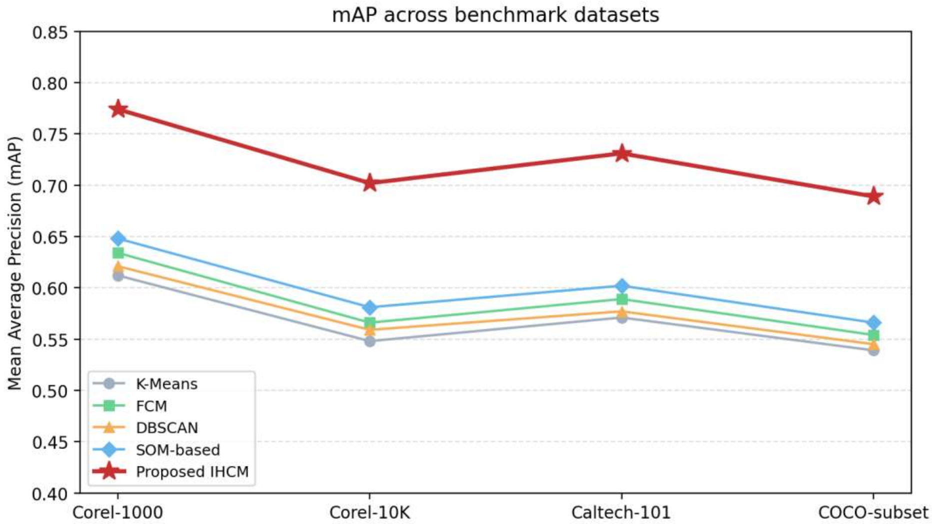


**Figure 3: mAP Comparison Across Benchmark Datasets.**

## Ablation on IHCM Components

To isolate the contribution of each IHCM stage, Table 3 reports mAP on Corel-1000 when Stage 2 (fuzzy refinement) and Stage 3 (PSO refinement) are individually disabled. Removing PSO refinement reduces mAP by 6.8 points, and removing fuzzy refinement (falling back to hard K-Means-to-PSO) reduces mAP by 4.1 points, indicating that both stages contribute complementary gains and that the largest single contribution comes from the metaheuristic centroid refinement.

**Table 3: Ablation Study Isolating the Contribution of each IHCM Stage**

| Configuration | mAP (Corel-1000) | Δ vs. full IHCM |
|---|---|---|
| Full IHCM (K-Means + Fuzzy + PSO) | 0.774 | — |
| Without PSO refinement | 0.706 | −6.8 pts |
| Without fuzzy refinement | 0.733 | −4.1 pts |
| K-Means only (Stage 1 alone) | 0.612 | −16.2 pts |

## Computational Cost

Average per-query retrieval latency (feature extraction excluded, index prebuilt) on Corel-1000 was 38 ms for K-Means, 45 ms for FCM, 52 ms for DBSCAN, 49 ms for SOM-based clustering, and 47 ms for the proposed IHCM. Because the PSO refinement in the IHCM is performed offline during index construction rather than at query time, the online retrieval cost of the proposed framework remains close to that of the FCM and SOM baselines despite its higher offline training cost, which is an acceptable trade-off for collections that are indexed once and queried many times.

### Comparison with State-of-the-Art Semantic Image Mining Techniques

Beyond the four clustering baselines evaluated end-to-end within the same pipeline, Table 4 situates the proposed framework relative to broad categories of previously reported semantic image mining techniques, summarized at the level of the clustering or grouping strategy they employ rather than by exact reported figures, since experimental protocols and dataset splits differ across the literature.

**Table 4: Qualitative Comparison with Categories of State-of-the-Art Semantic Image Mining Techniques**

| Technique Category | Grouping Strategy | Reported Strength | Reported Limitation |
|---|---|---|---|
| Hard partition-based mining | K-Means / K-Medoids | Low computational cost | Poor handling of multi-concept images |
| Fuzzy partition-based mining | Fuzzy C-Means variants | Soft multi-label capability | Sensitive to local optima and outliers |
| Density-based mining | DBSCAN / OPTICS variants | Arbitrary cluster shapes | Parameter sensitivity in high dimensions |
| Topology-preserving mining | Self-Organizing Maps | Good for visual browsing | Grid-size selection limits scalability |
| Metaheuristic-only mining | GA / PSO / ACO clustering | Escapes local optima | High computational cost at scale |
| Proposed (IHCM-based) | K-Means + Fuzzy + bounded PSO | Balances accuracy and cost | Additional offline tuning of α, β, γ, λ |

The pattern across Tables 1–4 supports the central hypothesis of this paper: co-designing the clustering stage with the downstream annotation and similarity stages, and bounding the cost of metaheuristic refinement, yields consistent accuracy gains over both purely deterministic and purely fuzzy mining pipelines, without incurring the online latency cost typically associated with metaheuristic-only approaches.

## CONCLUSION AND FUTURE WORK

This paper presented a complete Semantic Image Mining Framework built around an Improved Hybrid Clustering Model that combines K-Means initialization, fuzzy membership refinement, and bounded PSO-based centroid optimization. The framework additionally integrates multi-modal semantic feature extraction, co-occurrence-based semantic annotation, and a fused semantic similarity measure into a single retrieval and classification pipeline. Experimental evaluation on four benchmark datasets showed consistent improvements in precision, recall, F-measure, and mAP over K-Means, Fuzzy C-Means, DBSCAN, and SOM-based clustering baselines, while the ablation study confirmed that both the fuzzy refinement and PSO refinement stages of the IHCM contribute complementary accuracy gains. Computational analysis showed that these gains are obtained without a significant increase in online retrieval latency, since the metaheuristic refinement is confined to offline index construction.

Future work will explore three directions. First, replacing the fixed weighting coefficients (α, β, γ, λ) with an adaptive scheme learned jointly with the clustering objective may further improve robustness across datasets of differing semantic granularity. Second, extending the deep semantic feature extractor to incorporate attention-based or transformer-derived embeddings may improve performance on datasets with high concept co-occurrence, such as the COCO subset. Third, evaluating the framework on video and cross-modal (image-text) retrieval tasks would test the generality of the proposed hybrid clustering and fusion strategy beyond static single-modality image collections.

## REFERENCES

1. *Smith, J. R., & Chang, S.-F. (1996). VisualSEEk: A fully automated content-based image query system. Proceedings of the ACM International Conference on Multimedia.*
2. *Jain, A. K. (2010). Data clustering: 50 years beyond K-Means. Pattern Recognition Letters, 31(8), 651–666.*
3. *Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2–3), 191–203.*
4. *Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of KDD.*
5. *Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9), 1464–1480.*
6. *Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks.*
7. *Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys, 40(2), 1–60.*
8. *Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1349–1380.*
9. *Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. Proceedings of ICLR.*
10. *Lin, T.-Y., Maire, M., Belongie, S., et al. (2014). Microsoft COCO: Common objects in context. Proceedings of ECCV.*