УДК 004.8

STRUCTURAL CONCEPT FORMATION VIA GRAPH REDUCTION: AN EXPLAINABLE APPROACH TO FEW-SHOT IMAGE CLASSIFICATION

***Mykyta Lapin*[1]**, *Kostiantyn Bokhan*[2] **, *Kyrylo Perevoznyk* [1]**

[1] *Postgraduate student of the Department of System Analysis and Information and Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.*

[2] *PhD, Associate Professor of the Department of System Analysis and Information and Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine.*

Mykyta.Lapin@cit.khpi.edu.ua

Introduction. Modern perception systems achieve impressive accuracy yet operate as opaque pipelines. Post-hoc explanation techniques such as LIME and SHAP can be adversarially manipulated, undermining trust in domains requiring traceable decision paths [1, 2]. We propose making structure the carrier of explanations by encoding visual information using explicit graphs. For contour images, critical points (endpoints, corners, junctions) and line segments become graph nodes with attributes that preserve meaning directly, making explanations navigable: cycles justify closed contours, node degree reveals branching, attribute ranges encode variation.

Problem Statement. Post-hoc XAI techniques suffer from fundamental vulnerabilities: adversarial manipulation and sensitivity to hyperparameters [3]. Since explanations are generated after training, corroborating them against actual model reasoning is inherently difficult.

Graph-based representations offer an alternative paradigm where structure itself carries semantic information. Vision GNNs treat images as graphs with content-based connectivity, enabling interpretable visual reasoning [4]. Meanwhile, few-shot learning addresses concept formation from minimal data through meta-learning frameworks [5]. Graph-based contour representations and vectorization techniques provide foundations for encoding visual patterns using structural primitives [6, 7]. However, no existing approach combines structural explainability with few-shot capability through graph-based concept formation. This work addresses this gap by encoding visual patterns directly as attributed contour graphs and forming concepts through iterative structural reduction to stable attractors.

Main Material. *Graph Representation.* Image contours are encoded as graphs alternating between Point and Line nodes. Line segments are first-class nodes with attributes (length, direction, orientation) rather than edges. Point nodes classify into four types: EndPoint (terminals), CornerPoint (with angle attributes), IntersectionPoint (junctions), and StartPoint (traversal anchors). This separates topological structure from geometric properties. Nodes store (x, y) coordinates, angles, and directional attributes. Coordinates normalize to [−1, 1] via $normalized_x = (x - center_x)/center_x$ for scale and translation invariance.

Concept Formation. Concept formation follows principles of architectural information representation and energy-based neural network models [8, 9]. Given training samples $G_1, \dots, G_n$, the algorithm initializes $C_0 = G_1$ and iteratively refines via $C_{i+1} = CRO(C_i, G_{i+1})$ for $i = 1, \dots, n-1$, preserving only common structural elements. The process: (1) align start points; (2) preprocess critical points; (3) generate traversal paths; (4) identify common structure via similarity; (5) merge properties. The CRO employs three strategies in hierarchy IntersectionPoint → CornerPoint → Point: endpoint removal, intersection merging (relabeling nodes with degree ≤ 2), and path pruning via template matching. Parametric generalization:

numeric properties become ranges {min, max, center}; categorical properties preserve consistent values; list properties use set intersection.

Validation on MNIST. We validated on six MNIST classes (1, 2, 3, 6, 7, 9) with 2-4 samples per subclass (8 concepts total). Samples were augmented with 10 variants via rotation (±10 ◦) and translation (10%). Processing pipeline: binary thresholding, skeletonization, Growing Neural Gas fitting, simplification, graph construction, and normalization. Simple structures (digit 1) yielded 3-node graphs with mean degree 1.33; curved contours (2, 3, 7) produced 5-12 nodes with degrees 1.60-2.00; closed loops (6, 9) exhibited degree 2.00 with intersection points.

Classification via Graph Edit Distance (GED) with node substitution costs and range-based matching achieved 82.35% accuracy, 83.28% precision, 82.35% recall, 82.16% F1 across 5467 test images. Digits 6 and 9 achieved highest precision (94.23%, 91.55%) due to unique closed-loop topology. Digit 7 had lowest (74.38%) from ambiguity with digit 1. Primary confusion: 2/3 (152 errors) and 7/1 (118 errors)

Conclusions. This work presents a graph-based representation where structure carries explanations. Visual patterns are encoded as attributed graphs with critical points and line segments as nodes, semantic tags and geometric attributes attached directly. Multiple instances reduce to stable concept attractors through iterative composition, requiring only 2-6 training samples without gradient-based optimization, enabling few-shot learning with complete interpretability. MNIST validation shows concepts from 5-6 samples achieve 82.35% accuracy through explicit graph matching. Confusion patterns trace to structural similarity (2/3 share curved morphology, 7/1 share angular contours), confirming decisions derive from queryable topological and geometric properties rather than opaque features. By encoding semantics in graph topology instead of weight matrices, the approach advances explainable AI toward transparent, verifiable systems suitable for domains requiring trustworthy reasoning.

References:

1. *M. T. Ribeiro, S. Singh, and C. Guestrin*, ""why should i trust you?" explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

2. *D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju,* "Fooling lime and shap: Adversarial attacks on post hoc explanation methods," 2020. [Online]. Available: https://arxiv.org/abs/1911.02508

3. *Hooshyar and Y. Yang,* "Problems with shap and lime in interpretable ai for education: A comparative study of post-hoc explanations and neural-symbolic rule extraction," IEEE Access, vol. 12, pp. 137 472–137 490, 2024.

4. *K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu,* "Vision gnn: An image is worth graph of nodes," in Advances in Neural Information Processing Systems, vol. 35, 2022, pp. 8305–8319.

5. *J. Snell, K. Swersky, and R. Zemel,* "Prototypical networks for few-shot learning," in Proc. NeurIPS, 2017.

6. *Y. Parzhin*, "Principles of modal and vector theory of formal intelligence systems," 2013. [Online]. Available: https://arxiv.org/abs/1302.1334

7. *Y. Parzhin, S. Galkyn, and M. Sobol,* "Method for binary contour images vectorization of hand-written characters for recognition by detector neural networks," in 2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), Kharkiv, Ukraine, 2022, pp. 1–6.

8. *Y. Parzhyn,* "Architecture of information," 2025.

9. *Y. Parzhyn, M. Lapin, and K. Bokhan*, "A new approach to building energy models of neural networks," Advanced Information Systems, vol. 9, no. 4, pp. 100–119, 2025.