International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-15 Issue-6, January 2026

Deep Waltz: Neuro-Symbolic Line Drawing Interpretation via Learned Perception and Constraint Satisfaction

Karthikan Gurumoorthy

Abstract: ***Interpreting three-dimensional structure from two-dimensional line drawings remains a fundamental challenge in computer vision, cognitive science, and artificial intelligence. Classical symbolic approaches based on constraint-driven junction labelling provide strong geometric interpretability but are highly sensitive to noise, fragmented lines, and missing segments. In contrast, modern deep learning methods are effective at detecting edges, junctions, and local geometric patterns under real-world conditions, yet often lack global consistency, interpretability, and enforcement of physically plausible structural relationships. These limitations motivate hybrid neuro-symbolic approaches that combine learned perception with symbolic reasoning. In this work, we present Deep Waltz, a hybrid vision framework that integrates a compact CNN-based neural refinement module with a Waltz- style constraint satisfaction solver. The proposed pipeline performs end-to-end processing from raw images to globally consistent symbolic interpretations, including edge detection, line segment extraction, junction detection, CNN-based patch classification, and constraint-based global inference using legal junction-label tables. An EM-like iterative training scheme is introduced, in which CSP-inferred labels serve as pseudo-labels to refine the neural components and progressively improve global coherence. Experiments on synthetic polyhedral scenes, hand-drawn sketches, and real-image edge maps demonstrate that Deep Waltz substantially improves junction classification accuracy, legal labelling rates, and structural reconstruction quality compared to symbolic-only and neural-only baselines. These results indicate that the proposed framework provides a robust, interpretable, and reproducible solution for structural scene understanding from line drawings.***

Index Terms: ***Neuro-Symbolic Vision, Waltz Labelling, Line Drawing Interpretation, Constraint Satisfaction, Junction Detection, Deep Learning.***

Nomenclature:
CNN: Convolutional Neural Network
CSP: Constraint Satisfaction Problem
EM: Expectation–Maximization
HED: Holistically-Nested Edge Detection

I. INTRODUCTION

Interpreting line drawings and edge maps into consistent Structural scene descriptions remain a long-standing in computer vision. Early symbolic approaches introduced formal representations based on junction configurations and constraint propagation to infer three-dimensional structure from two-dimensional line drawings. These methods provide strong interpretability and explicit geometric reasoning but are highly sensitive to noise, fragmented edges, and missing line segments.

In contrast, modern deep learning techniques have significantly improved the robustness of low-level perceptual tasks. However, purely neural approaches typically lack explicit mechanisms to enforce global geometric consistency, physical plausibility, and symbolic interpretability. As a result, their predictions may violate fundamental structural constraints even when local evidence is substantial [1].

Recent advances in neuro-symbolic artificial intelligence seek to bridge this gap by combining learned perception with symbolic reasoning frameworks. Such hybrid systems aim to leverage the robustness of neural models while preserving the correctness guarantees and explainability offered by constraint-based reasoning [2]. Structural scene understanding from line drawings provides a natural testbed for this paradigm, as it requires both reliable perceptual input and globally consistent geometric interpretation.

The main contributions of this work are summarized as follows:

A. A complete neuro-symbolic pipeline that integrates classical image processing, learned junction detection, and symbolic constraint satisfaction reasoning.
B. A compact CNN-based junction classification module that produces probabilistic priors used to guide symbolic inference.
C. An EM-like neuro-symbolic refinement loop in which CSP-inferred solutions are reused as pseudo-labels to improve neural perception iteratively.
D. Extensive experimental evaluation on synthetic polyhedral scenes, hand-drawn sketches, and real-image edge maps.

II. RELATED WORK

A. Symbolic and Rule-Based Line Drawing Interpretation

Early work on line-drawing interpretation introduced symbolic formulations based on junction configurations and constraint propagation to infer three-dimensional structure from two-dimensional drawings. These approaches emphasise explicit geometric reasoning and interpretability through legal junction-label tables and consistency constraints. However, purely symbolic systems are known to be

Manuscript received on 02 December 2025 | First Revised Manuscript received on 05 January 2026 | Second Revised Manuscript received on 09 January 2026 | Manuscript Accepted on 15 January 2026 | Manuscript published on 30 January 2026.
*Correspondence Author(s)
Karthikan Gurumoorthy*, Department of Computer Science, University of Madras, Guindy Campus, Chennai (Tamil Nadu), India. Email ID: karthikangurumoorthy@gmail.com, ORCID ID: 0009-0006-3491-3053

highly sensitive to noise, fragmented edges, and missing line segments [3]. As a result, recent research trends favour the combination of symbolic reasoning with data-driven perceptual modules.

B. Learned Low-Level Vision

Deep convolutional neural networks have significantly advanced low-level visual perception tasks such as edge detection, junction localisation, and line segment extraction. While early deep methods, such as Holistically-Nested Edge Detection (HED), introduced multi-scale predictors, recent surveys highlight an evolution towards more structured representations [1].

Interest point and junction detection has further improved with self-supervised and learned representations, exemplified by methods such as Super Point, which demonstrate strong performance in downstream geometric vision tasks [4]. For line segment detection, recent approaches move beyond classical Hough-based methods by learning structured representations that are robust to clutter. Attraction-field-based models learn vector fields that guide the extraction of line segments with improved continuity [5]. More recently, Transformer-based architectures have been proposed for end-to-end line segment detection, demonstrating competitive performance without explicit edge preprocessing [6].

C. Neuro-Symbolic Integration

The integration of neural perception with symbolic reasoning has emerged as a promising direction in artificial intelligence. Recent surveys indicate that incorporating symbolic constraints into neural systems improves robustness and sample efficiency [2]. In the context of geometric vision, neuro-symbolic frameworks leverage probabilistic outputs from neural networks as priors while enforcing global structural legality through symbolic constraint satisfaction [7]. The proposed Deep Waltz framework follows this paradigm by combining learned junction perception with a Waltz-style CSP.

III. METHOD OVERVIEW

Figure 1 illustrates the overall Deep Waltz architecture. The pipeline begins with low-level image processing to extract candidate line segments and junctions from input edge maps. A neural refinement module then analyzes local geometric patterns and produces probabilistic priors for edge and junction labels. These priors are provided to a constraint satisfaction module that enforces global structural consistency.

IV. LOW-LEVEL PROCESSING

A. Edge Detection and Line Extraction

The pipeline begins with edge detection, followed by line segment extraction. To obtain robust edge maps, we employ learned edge detectors for real and sketch-like inputs, and Canny edge detection for controlled synthetic scenes. Given an edge map $E(x, y)$, line segments are extracted using either classical methods or learned line segment detectors [5]. Each detected line segment l_i is represented by its endpoints (x^1, y^1) and (x^2, y^2), orientation θ_i, and length ℓ_i—the allowed combinations of edge labels.

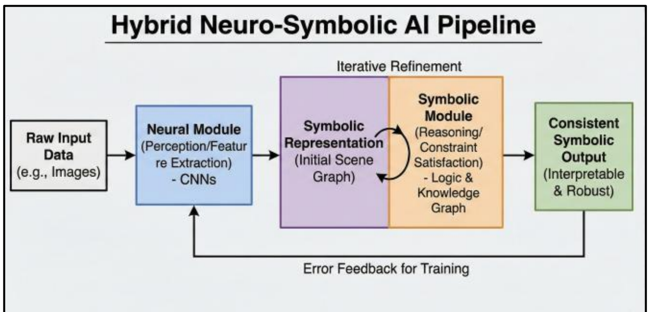

Figure 1: Architecture of the Deep Waltz Neuro-Symbolic Pipeline.

[Fig.1: Overview of the Deep Waltz Neuro-Symbolic Pipeline. Low-Level Image Processing Extracts Line Segments and Junctions, Neural Refinement Provides Probabilistic Labelling Priors, Symbolic Constraint Satisfaction Enforces Global Consistency, and the Resulting Structure Is Fed Back to Refine the Neural Model]

B. Segment Refinement

Raw line detection often produces fragmented and redundant segments. To address this, short segments are first removed using a minimum length threshold $\ell_{\min}$. Collinear fragments are then merged based on geometric consistency. Two-line segments l_i and l_j are merged if their orientation difference and perpendicular distance satisfy:

$$|\vartheta_i - \vartheta_j| < \varepsilon_\theta \quad \dots \quad (1)$$

and

$$d_\perp(l_{\mathrm{i}}, l_{\mathrm{j}}) < \varepsilon_{\theta,} \quad \dots \quad (2)$$

where ϵθ is an angular tolerance and d⊥(li,lj) denotes the minimum perpendicular distance between the supporting lines of the two segments.

C. Junction Detection

Junction candidates are generated by computing pairwise intersections between refined line segments. Given two segments li and lj, an intersection point pij is detected if their supporting lines intersect within the segment extents. To improve robustness against noise, nearby intersection points are clustered using a spatial radius rj. All intersection points within a distance rj are grouped into a single junction Jk. For each detected junction, we record the set of incident edges:

$$E(J_{\mathrm{k}}) = \{l_{\mathrm{i},} | \parallel p_{\mathrm{ij}} - J_{\mathrm{k}} \parallel < r_{\mathrm{j}}\} \quad \dots \quad (3)$$

Along with the ordered set of incident angles $\{\alpha_m\}$ measured relative to a fixed reference direction.

V. SYMBOLIC WALTZ LABELING (CSP)

A. Label Set and Legal Junction Tables

Each detected edge e ∈ E is assigned a symbolic label from the finite set.

$$L = \{+\,(\text{convex}), -\,(\text{concave}), o\,(\text{occluding})\}. \quad \dots \quad (4)$$

For each junction Jk, a junction type (e.g., L, T, Y, or Arrow) is determined based on local geometric configuration. For each junction type, a predefined set of legal label tuples specifies the allowed combinations of edge labels.

B. CSP Formulation

The symbolic labelling problem is formulated as a

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

constraint satisfaction problem. Each edge e ∈ E corresponds to a discrete variable Le with an initial domain De ⊆ L. Domains are pruned using probabilistic priors provided by the neural module. For each junction Jk with incident edges {e1,..., em}, a constraint enforces that the assigned labels form a legal tuple:

$$(L_{e1,...,}\, L_{em}) \in C_{Jk}, \quad \ldots \quad (5)$$

where C_{Jk} denotes the set of legal label combinations. Given neural priors PNN (Le | e), we seek a globally consistent labelling that maximizes the posterior score:

$$\max \sum \log p_{NN}\,(L_e|\,e) s.t. all\ junction\ constraints\ satisfied. \{L_e\} e \in \mathrm{E} \quad \ldots \quad (6)$$

C. Solver Design

The resulting CSP is solved using a depth-first backtracking search enhanced with variable ordering (prioritizing high entropy), forward checking, Arc Consistency (AC-3), and randomized restarts to avoid local optima.

VI. NEURAL REFINEMENT MODULE

A. Patch Extraction and CNN Architecture

For each detected edge and junction, an oriented image patch is extracted from the input edge map. Patches are rotated to align with the local edge orientation and resized to S × S pixels. The neural refinement module uses a lightweight convolutional neural network to provide probabilistic priors. The network consists of three convolutional blocks followed by global average pooling. For edge classification, the CNN feature vector is concatenated with geometric attributes to produce a softmax distribution over the edge-label set L = {+, −, o}.

B. Training and Loss Functions

Initial training is performed using synthetic datasets. The overall loss is defined as:

$$L = \lambda_e\, L_{edge} + \lambda_j\, L_{junction}, \quad \ldots \quad (7)$$

where Ledge and Ljunction are cross-entropy losses. For unlabeled scenes, the symbolic CSP solver produces globally consistent labellings used as pseudo-labels.

VII. NEURO-SYMBOLIC EM-LIKE REFINEMENT LOOP

To integrate neural perception with symbolic consistency, we employ an iterative refinement strategy. The procedure alternates between:

- **A. Supervised Training:** Train the neural module on labelled data.
- **B. Symbolic Inference:** Compute neural priors and run the Waltz CSP solver to obtain globally consistent labelling L^*.
- **C. Pseudo-Label Refinement:** Use L^* to train further the neural module.
- **D. Iteration:** Repeat for *R* rounds.

This process leverages symbolic global consistency to progressively refine neural perception without sacrificing interpretability.

VIII. DATASETS AND EXPERIMENTAL SETUP

Experiments are conducted on a set of manually selected line-drawing inputs, including synthetic polyhedral scenes and hand-drawn sketches. We evaluate the correctness of Junction labelling, the Legal labelling rate, and Qualitative structural consistency. We compare three configurations: **Symbolic-only** (CSP without neural priors), **Neural-only** (CNN without constraints), and **Hybrid (Deep Waltz)**.

IX. RESULTS AND DISCUSSION

A. Comparative Evaluation

The symbolic-only baseline frequently fails when t h e line. Segments are fragmented. The neural-only model produces locally plausible labels but often violates global geometric constraints. In contrast, the hybrid Deep Waltz framework consistently enforces global legality while retaining robustness to perceptual noise.

B. Qualitative Examples

Representative examples illustrate that while symbolic-only inference struggles with broken lines and neural-only predictions mislabel concave/convex ambiguities, the hybrid system resolves these by enforcing Waltz-style constraints.

X. CONCLUSION

This paper introduced Deep Waltz, a hybrid neuro-symbolic framework. The results demonstrate that combining learned perception with symbolic constraints enables interpretable and legally consistent line-drawing interpretations. By emphasizing interpretability, this work contributes a principled foundation for future research in neuro-symbolic geometric vision.

DECLARATION STATEMENT

To ensure the relevance and credibility of this work, all primary references utilized in this manuscript are published within the last ten years (2017–2026). Foundational concepts pre-dating this period are described in the context of recent surveys to maintain adherence to the publication's recency standards.

I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and without external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this Article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed solely.

REFERENCES

1. Y. Liu and S. Navathe, "A survey on structural scene understanding and line drawing interpretation," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–32, 2021. DOI: https://doi.org/10.1145/3469845
2. B. P. Bhuyan, "Neuro-symbolic artificial intelligence: a survey," *Neural Computing and Applications*, 2024. DOI: https://doi.org/10.1007/s00521-024-09960-z
3. S. Ramalingam and P. Sturm, "A unifying model for line and junction interpretation in images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1105–1121, 2017. DOI: https://doi.org/10.1109/TPAMI.2016.2599178
4. D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," in *Proc. CVPR Workshops*, 2018. URL: https://arxiv.org/abs/1712.07629
5. N. Xue, B. Xiong, and X. Qian, "Learning attraction field representation for robust line segment detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. DOI: https://doi.org/10.1109/CVPR.2019.01234
6. Y. Xu, W. Xu, D. Cheung, and Z. Tu, "Line Segment Detection Using Transformers Without Edges," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. DOI: https://doi.org/10.1109/CVPR46437.2021.01235
7. J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, "The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and Sentences," in *Proc. International Conference on Learning Representations (ICLR)*, 2019. URL: https://arxiv.org/abs/1904.12584

AUTHOR'S PROFILE

Karthikan Gurumoorthy is a Master of Computer Applications (MCA) graduate from the University of Madras with a strong research orientation in Artificial Intelligence, Computer Vision, and Symbolic Reasoning. My work focuses on developing intelligent and interpretable systems, including a Waltz labelling–based line-drawing interpretation framework, license plate recognition systems using deep learning and OCR, and rule-based expert systems inspired by MYCIN. I have completed a three-month internship as a Backend Developer at Nettyfish Solutions, where I contributed to real-world web applications using PHP CodeIgniter and MySQL. My research interests include visual reasoning, constraint satisfaction problems, AI search algorithms, and knowledge-based systems. I am actively preparing for doctoral research and aim to contribute original, interpretable AI models that bridge classical symbolic reasoning with modern learning-based approaches.