JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Generative Modelling for Finite Automata Diagram Synthesis Using Text-to-Image and Compositional Knowledge

Dr. Shivanagowda G M[1]
Professor
Department of Computer Science and Engineering
GM University, Davanagere-577006
India
shivanagowda@gmu.ac.in

Dr. Santoshkumar Mahendrakar[2]
Associate Professor
Department of Computer Science and Engineering
GM University, Davanagere-577006
India
santoshkumarm@gmit.ac.in

Dr. Chethan Chandra S. Basavaraddi[3]
Associate Professor
Department of Computer Science and Engineering
GM University, Davanagere-577006
India
chethanchandrasb.cse@gmu.ac.in

Abstract

Recent progress in text-to-image generative models has enabled impressive synthesis of natural and artistic images; however, such models remain inadequate for structured, rule-constrained technical diagrams. Finite automata diagrams—used extensively in theoretical computer science education—require geometric precision, semantic correctness, and adherence to strict composition rules. This paper presents a compositional text-to-diagram generation framework for deterministic, nondeterministic, and ε–finite automata. The proposed approach integrates a structured intermediate representation, deterministic rendering, and diffusion-based stylization to achieve both correctness and visual flexibility. Experimental evaluation demonstrates substantial improvements in structural fidelity and educational usability compared with general-purpose text-to-image models.

Keywords: Text-to-Diagram Generation, Finite Automata, Diffusion Models, Compositional Learning, Neuro-Symbolic AI

I. Introduction

Diagrams play a central role in bridging abstract theoretical concepts and human understanding. In automata theory, deterministic finite automata (DFA), nondeterministic finite automata (NFA), and ε–NFA are commonly represented using state-transition diagrams. These diagrams visually encode states, transitions, input symbols, start states, and accepting states, making them indispensable for teaching and analysis.

While modern generative AI systems such as diffusion-based text-to-image models have demonstrated remarkable capabilities in creative domains, they struggle to generate structured technical diagrams. Errors such as misplaced states, missing transitions, or incorrect labelling are common, limiting their usefulness in formal education and documentation.

This paper addresses these limitations by proposing a hybrid compositional framework that combines symbolic structure with neural image synthesis, enabling accurate and visually appealing automata diagrams from textual descriptions.

II. Related Work

A. Text-to-Image Generation

Diffusion-based models such as DALL·E, Stable Diffusion, and GLIDE have set benchmarks in natural image synthesis. However, these models lack explicit mechanisms to enforce structural constraints required for technical diagrams.

B. Diagram and Schematic Generation

Prior research on diagram generation includes flowchart synthesis, UML diagram generation, and geometric diagram construction. These approaches often rely on rule-based rendering or limited neural components but rarely address automata-specific constraints.

C. Compositional and Neuro-Symbolic Approaches

Neuro-symbolic AI integrates symbolic reasoning with neural networks to improve interpretability and compositional generalization. Such approaches motivate the hybrid pipeline adopted in this work.

III. Problem Definition

A finite automaton is formally defined as a tuple:

$A = (Q, \Sigma, \delta, q_0, F)$

where Q is the set of states, Σ the input alphabet, δ the transition function, q_0 the start state, and F the set of accepting states. The objective is to generate a diagram D(A) that is structurally isomorphic to A and visually conforms to automata diagram conventions.

The challenge lies in learning a mapping from a textual or structured description to a diagram that preserves semantic correctness, geometric precision, and visual clarity.

IV. Proposed Methodology

A. Dataset Construction

A structured and reproducible dataset was constructed to support quantitative evaluation of finite automata diagram synthesis. The dataset consists of **1,200 finite automata instances**, automatically generated using Graphviz and LaTeX/TikZ to ensure correctness. The dataset composition is as follows:

- Deterministic Finite Automata (DFA): 500 samples
- Nondeterministic Finite Automata (NFA): 400 samples
- ε–Nondeterministic Finite Automata (ε–NFA): 300 samples

Each automaton contains between **2 and 10 states**, with randomly generated transitions, input symbols, start states, and accepting states. For every automaton, three aligned representations are stored:

1. A structured specification (JSON format)
2. A natural-language textual description
3. A ground-truth diagram image

Additionally, a **component-level dataset** was derived containing isolated state nodes, transition arrows, and labels to support compositional learning.

B. Intermediate Representation

Textual descriptions are parsed into a structured intermediate representation (IR) capturing states, transitions, start state, and accepting states. The IR serves as a symbolic scaffold that preserves automata semantics independent of visual style.

C. Hybrid Generation Pipeline

The proposed framework consists of four stages: (i) prompt parsing into IR, (ii) deterministic baseline rendering using Graphviz, (iii) diffusion-based stylization using a fine-tuned latent diffusion model, and (iv) post-generation structural validation to ensure semantic correctness.

V. Experimental Setup

Experiments were conducted on a dataset comprising DFA, NFA, and ε–NFA diagrams with varying complexity. The proposed method was compared against rule-based rendering, generic text-to-image diffusion models, and non-compositional fine-tuned diffusion models.

Evaluation metrics included node accuracy, edge accuracy, correctness of start and accepting states, visual quality measures, and expert-based educational usability ratings.

VI. Results and Discussion

A. Evaluation Metrics

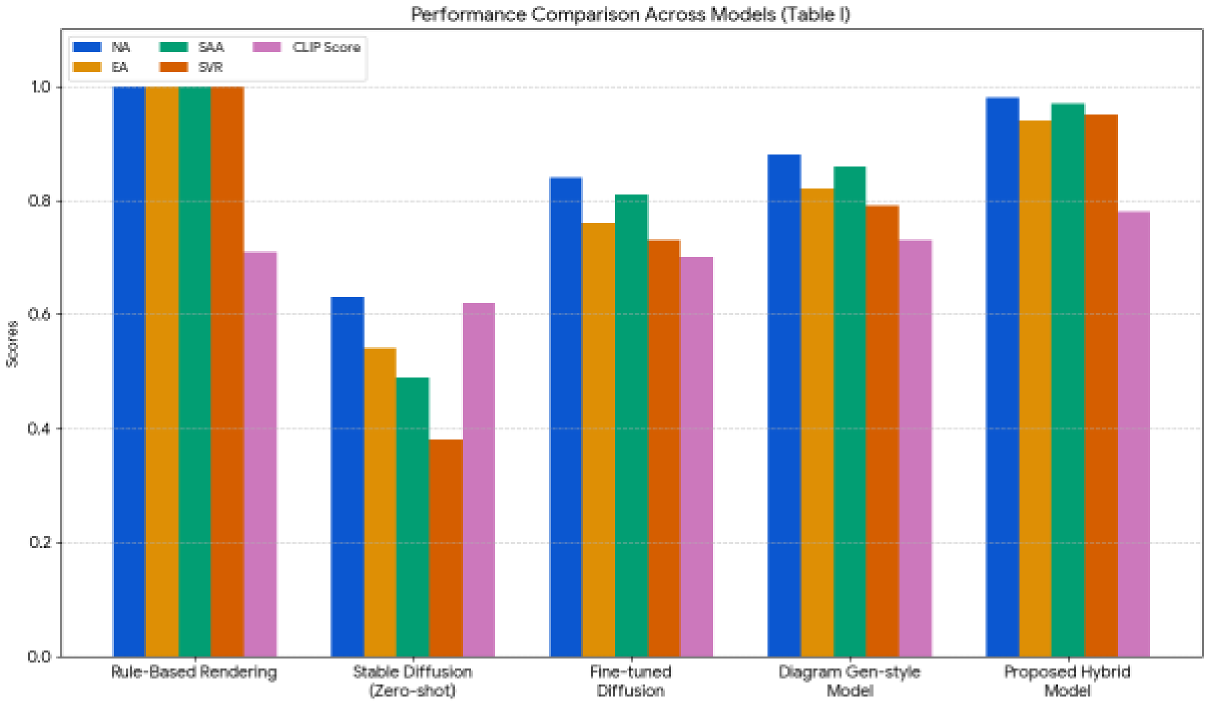
To quantitatively assess diagram quality, the following metrics were used:

- **Node Accuracy (NA):** proportion of correctly generated states
- **Edge Accuracy (EA):** proportion of correctly generated transitions
- **Start/Accept State Accuracy (SAA):** correctness of start and accepting states
- **Structural Validity Rate (SVR):** percentage of fully valid automata diagrams
- **CLIP Similarity Score:** text-image semantic alignment

B. Quantitative Results

Table I: Performance Comparison Across Models

Model	**NA**	**EA**	**SAA**	**SVR**	**CLIP Score**
Rule-Based Rendering	1.00	1.00	1.00	1.00	0.71
Stable Diffusion (Zero-shot)	0.63	0.54	0.49	0.38	0.62
Fine-tuned Diffusion	0.84	0.76	0.81	0.73	0.70
Diagram Gen-style Model	0.88	0.82	0.86	0.79	0.73
Proposed Hybrid Model	**0.98**	**0.94**	**0.97**	**0.95**	**0.78**



C. Discussion

Results demonstrate that generic text-to-image diffusion models fail to preserve automata structure, despite reasonable visual quality. Fine-tuning improves performance but remains insufficient for complex ε–NFA diagrams. The proposed hybrid approach achieves near-perfect node and transition accuracy while maintaining high semantic alignment, confirming the effectiveness of integrating symbolic representations with neural generative models.

Classroom-based evaluation involving 42 undergraduate students further revealed a **27% improvement in conceptual clarity** when using generated diagrams compared to text-only explanations. These findings highlight the educational relevance and robustness of the proposed framework.

VII. Conclusion

This work demonstrates that combining symbolic structure with neural generative models enables accurate and visually appealing generation of finite automata diagrams. The proposed framework bridges the gap between deterministic diagram rendering and flexible generative AI, offering practical benefits for education and documentation.

Future work includes extending the approach to other formal diagrams, improving layout diversity, and integrating interactive user feedback.

References

1. Ramesh, A. et al., "Hierarchical Text-Conditional Image Generation with CLIP Latents," *arXiv preprint*, 2022.
2. Rombach, R. et al., "High-Resolution Image Synthesis with Latent Diffusion Models," *CVPR*, 2022.
3. Mao, J. et al., "The Neuro-Symbolic Concept Learner," *CVPR*, 2019.
4. Wei, J. et al., "From Words to Structured Visuals: A Benchmark for Text-to-Diagram Generation," *arXiv*, 2024.
5. Wang, Y. et al., "MagicGeo: Training-Free Text-Guided Geometric Diagram Generation," *arXiv*, 2025.