www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

GENERATING TEXT TO REALISTIC IMAGES USING KNOWLEDGE GRAPH GENERATIVE ADVERSARIAL NETWORK

[1]Mr.P.Masoom Basha
Assistant Professor, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: pinjarimasoombasha11@gmail.com

[2] Y.Deepika Reddy
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: deepikareddy7720@gmail.com

[3] P.Bhumika
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: bhumikapandiri48@gmail.com

[4]V.Sai Varsha Sri
UG Student, Dept. Computer Science and Engineering
Vignan's Institute of Management and Technology for Women, Hyd.
Email: varsharao002@gmail.com

Abstract*—The synthesis of semantically coherent images from textual descriptions remains a significant challenge in the field of generative modeling. To address this, we introduce the Knowledge-Guided Generative Adversarial Network (KG-GAN), a novel framework that integrates structured semantic information into the generative process. Our approach begins by constructing a Knowledge Graph (KG) from textual inputs using Natural Language Processing techniques, capturing entities and their interrelations. These graphs are then transformed into dense vector representations through Graph Convolutional Networks, encapsulating the semantic context of the descriptions. The resulting embeddings guide the GAN to produce images that are not only realistic but also semantically aligned with the input text. Experimental evaluations demonstrate that KG-GAN effectively generates high-fidelity images corresponding to both seen and unseen categories, showcasing its potential in applications requiring nuanced semantic understanding.

KEYWORDS— Knowledge Graph, GCN, GAN, NLP.

I. INTRODUCTION

Generating images from written descriptions is a complex task in artificial intelligence, requiring a deep understanding of both language and visual content. The goal is not only to produce images that appear realistic but also to ensure they reflect the meaning conveyed in the text. Although Generative Adversarial Networks (GANs) have shown promising results in this area, they often fall short when it comes to interpreting and representing detailed or nuanced descriptions.To enhance semantic understanding in this task, our method introduces a structured representation of text through a Knowledge Graph, built using Natural Language Processing (NLP) techniques. This graph captures the entities and relationships present in the input description, offering a clearer context for image generation.Next, the information from the Knowledge Graph is converted into vector form using Graph Convolutional Networks (GCNs).These vectors act as rich semantic cues for the GAN, enabling it to generate images that are both visually accurate and aligned with the intended meaning of the text. These vectors act as rich semantic cues for the GAN, enabling it to generate images that are both visually accurate and aligned with the intended meaning of the text. These vectors act as rich semantic cues for the GAN, enabling it to generate images that are both visually accurate and aligned with the intended meaning of the text.The central aim of this approach is to produce high-quality, detailed images that not only look realistic but also accurately represent the content and context of the original textual input.

II. LITERATURE REVIEW

One of the pioneering works in this domain is KG-GAN, introduced by Che-Han Chang, Chun-Hsien Yu, Szu-Ying Chen, and Edward Y. Chang. This framework trains two generators: one learns from data, while the other learns from knowledge, guided by a constraint function. The model demonstrates effectiveness in generating unseen categories from textual descriptions, validated on tasks like fine-grained image generation and hair recoloring.Building upon this, Context Canvas, developed by Kavana Venkatesh, Yusuf Dalva, Ismini Lourentzou, and Pinar Yanardag, integrates a graph-based Retrieval-Augmented Generation (RAG) mechanism into text-to-image diffusion models. By dynamically retrieving detailed contextual information from knowledge graphs, the system enhances the generation of visually accurate and contextually rich images, particularly for complex or culturally specific subjects.Another notable contribution is RiFeGAN by Jun Cheng, Fuxiang Wu, Yanling Tian, Lei Wang, and Dapeng Tao. This model addresses the challenge of limited textual information in text-to-image synthesis by enriching descriptions with prior knowledge. It employs an attention-based caption matching mechanism to select and refine compatible candidate captions, enhancing the semantic consistency and visual quality of generated images.Furthermore, the work by Yuxia Geng, Jiaoyan Chen,

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Zhuo Chen, Zhiquan Ye, Zonggang Yuan, Yantao Jia, and Huajun Chen explores generative adversarial zero-shot learning via knowledge graphs. Their approach incorporates rich semantics from knowledge graphs into GANs, enabling the synthesis of compelling visual features for unseen classes, thereby improving zero-shot learning performance.

III. METHODOLOGY

The proposed system for generating realistic images from natural language descriptions using Knowledge Graph-based Generative Adversarial Networks (KG-GAN) involves a multi-phase process. These phases include processing the input text, constructing a semantic graph, deriving graph embeddings, and producing images using a GAN model. Each phase incrementally refines the information from unstructured text into a visually meaningful output.

1. Text Preprocessing and Semantic Analysis

The process begins with the system receiving a textual description. Various Natural Language Processing (NLP) techniques are employed to understand and break down the text. This includes steps like tokenization, identifying parts of speech, and detecting named entities. Additionally, syntactic tools such as dependency parsing and semantic role labeling are applied to uncover the interactions among entities, identify descriptive elements, and understand the structural composition of the sentence.

2. Construction of the Knowledge Graph

From the analyzed text, a Knowledge Graph (KG) is built where nodes symbolize entities or concepts, and edges capture the relationships between them—for instance, relations like "cat under table" or "person holding book." To enhance the richness of the graph, external semantic resources such as ConceptNet or WordNet may be used to add background knowledge, supporting more nuanced understanding. This graph structure helps preserve spatial cues and contextual associations embedded in the original text.

3. Embedding Generation via Graph Convolutional Networks (GCNs)

Once the KG is constructed, it is transformed into a vectorized format using Graph Convolutional Networks. GCNs are designed to compute feature representations of each node by considering both its own attributes and those of its neighbors. The resulting embeddings capture fine-grained (object-level) and broader (scene-level) semantics, providing a dense numerical input that reflects both structure and meaning.

4. Image Synthesis Using GAN

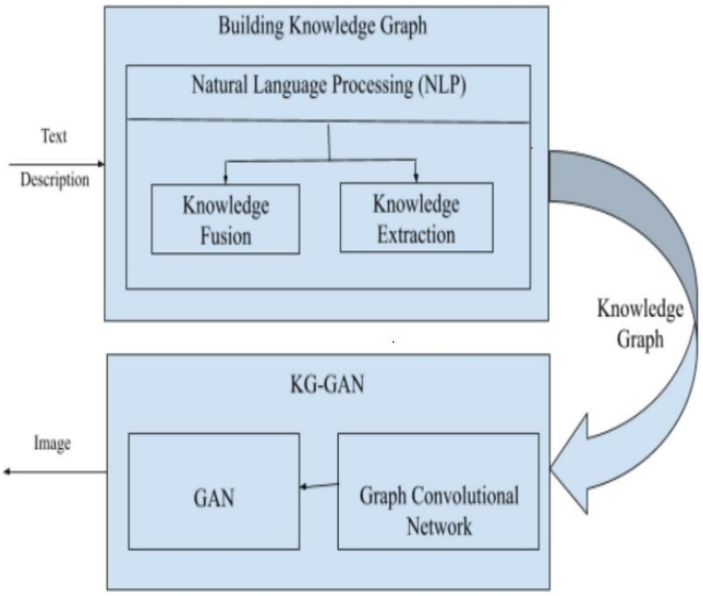
The embeddings obtained from the GCN are passed into a Generative Adversarial Network. The generator in the GAN takes this encoded semantic information to create images that visually align with the original description. A discriminator then evaluates these images on two fronts: their visual plausibility and how well they correspond to the semantic structure represented by the knowledge graph. Training is guided by a combination of adversarial loss and a semantic alignment objective to ensure high-quality and context-aware image generation.

5. Optional: Image Enhancement

In certain setups, a secondary refinement stage is included to further polish the generated images. This may involve an additional GAN or a dedicated super-resolution model that enhances image details such as texture, clarity, and color. Attention mechanisms may be employed to focus on important graph components to better guide refinement.

A) System Architecture:

The primary aim of the proposed system is to generate realistic images based on input textual descriptions. The overall framework is divided into two main components: the Knowledge Graph construction module and the Knowledge-Guided Generative Adversarial Network (KG-GAN). The process begins by taking the text description as input, which serves as the foundation for image generation.



In the first stage, the text is processed using Natural Language Processing techniques to construct a Knowledge Graph, capturing relevant entities and their relationships. This structured representation forms the initial output that encodes the semantic content of the text.The second stage involves the KG-GAN, which receives the Knowledge Graph as input. Through the use of Graph Convolutional Networks (GCNs), the graph is converted into embeddings that encapsulate its semantic structure. These embeddings are then used by the GAN to synthesize high-quality images that are both visually accurate and semantically meaningful.

B) System Implementation:

The system is designed to generate realistic images from textual descriptions by leveraging knowledge graphs and Generative Adversarial Networks (GANs) through the integration of pre-trained models and libraries. The implementation avoids training on custom datasets, instead utilizing off-the-shelf components for immediate inference.

a. Input Interface and Text Processing

Users input descriptive text, which is processed through Natural Language Processing (NLP) tools such as spaCy or NLTK. These tools perform essential text processing steps including tokenization, part-of-speech tagging, named entity

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

recognition, and syntactic parsing. Semantic role labeling is applied to extract the relationships between entities and actions, forming the foundation for knowledge graph construction.

b. Knowledge Graph Construction

From the extracted linguistic data, a Knowledge Graph is constructed dynamically where entities correspond to nodes and relationships to edges. The graph reflects the underlying semantic organization of the input data.To enrich the graph's contextual information, external knowledge bases like ConceptNet or DBpedia can be integrated without additional training.

c. Graph Embedding via Pre-trained GCN

The constructed knowledge graph is converted into numerical embeddings using a pre-trained Graph Convolutional Network (GCN).The system is designed to generate realistic images from textual descriptions by leveraging knowledge graphs and Generative Adversarial Networks (GANs) through the integration of pre-trained models and libraries. The implementation avoids training on custom datasets, instead utilizing off-the-shelf components for immediate inference. "The GCN transforms node features and their interconnections into embeddings, maintaining both semantic meaning and structural context."Libraries such as PyTorch Geometric or DGL are used to implement this step, leveraging pre-existing models to avoid the need for training.

d. GAN-based Image Generation Using Pre-trained Models The graph embeddings serve as conditioning inputs to a pre-trained GAN model tailored for text-to-image synthesis. The generator produces images that correspond to the semantic features captured in the graph, while the discriminator, also pre-trained, assesses the authenticity and relevance of generated images. This stage uses adversarial loss and semantic consistency measures embedded within the pre-trained framework.

e. Deployment and Inference

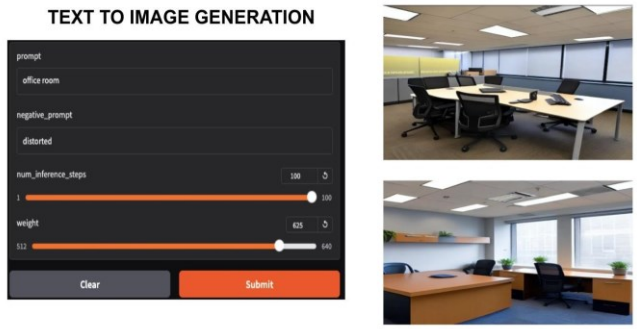
Instead of training, the system focuses on inference by directly installing and integrating pre-trained libraries and models. The pipeline is optimized for real-time generation, supporting applications that require quick and realistic image synthesis from textual inputs.

f. Image Refinement and Output

For improved image quality, a refinement or super-resolution network may be applied as a post-processing module. This further enhances visual details and sharpness. The final output is delivered to the user interface or saved for further use in applications such as digital media, virtual environments, or AI-assisted creative tools.Libraries such as PyTorch Geometric or DGL are used to implement this step, leveraging pre-existing models to avoid the need for training.

IV. RESULTS AND ANALYSIS

The output demonstrates the capability of a text-to-image generation system to convert written descriptions into matching visual images.. In this example, the input prompt "A children's bedroom with nature picture on the wall" guides the model to produce realistic images that align with the given scene description. The interface used for generation allows for customization through various parameters, including a negative prompt to specify elements to exclude—such as "distorted"—as well as settings for inference steps and weight, which influence the level of detail and adherence to the prompt. The resulting images illustrate how effectively the model can interpret and visualize natural themes within a bedroom setting, demonstrating the potential of generative models in creating visually rich outputs from descriptive text.



V. CONCLUSION

Text to Image synthesis as said is one of the most recent innovations in computer vision. It describes the method of generating images based on the relevant words and text descriptions. Knowledge-Guided Generative Adversarial Text to Image synthesis as said is one among the foremost recent innovations in computer vision. It describes the tactic of generating images from the relevant words and text descriptions. Knowledge-Guided Generative Adversarial Network is employed in Text-to-Image synthesis to get high-quality images that are visually realistic and semantically

sensible matching the given text descriptions. The proposed method also uses Natural Language processing techniques to create a knowledge graph representation and Graph Convolutional Networks to get the embedding representations of the knowledge graph. This is then fed into the GAN model to get realistic images for the given input, mainly the text description.

VI. FUTURE SCOPE

In the future, generating realistic images from text using GANs (Generative Adversarial Networks) will become even more powerful and useful. As AI continues to improve, these systems will be able to better understand detailed and complex text descriptions, including emotions and context. This means the images they create will match the text more accurately. Also, with faster computers, we can expect real-time applications where images are created instantly from what we type — useful in games, design tools, and virtual reality.The quality of generated images will also get better, with clearer and more realistic visuals. Future systems may also combine text with other inputs like voice or sketches, making them more flexible and easy to use**.**

VII. REFERENCES

[1] C.-H. Chang, C.-H. Yu, S.-Y. Chen, and E. Y. Chang, "KG-GAN: Knowledge-Guided Generative Adversarial Networks," arXiv preprint arXiv:1905.12261, 2019, doi: [10.48550/arXiv.1905.12261](https://doi.org/10.48550/arXiv.1905.12261).

[2] K. Venkatesh, Y. Dalva, I. Lourentzou, and P. Yanardag, "Context Canvas: Enhancing Text-to-Image Diffusion Models with Knowledge Graph-Based RAG," arXiv preprint arXiv:2412.09614, 2024, doi: [10.48550/arXiv.2412.09614](https://doi.org/10.48550/arXiv. 2412.09614).

[3] J. Cheng, F. Wu, Y. Tian, L. Wang, and D. Tao, "RiFeGAN: Rich Feature Generation for Text-to-Image Synthesis from Prior Knowledge," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10908–10917, doi: [10.1109/CVPR42600.2020.01092](https://doi.org/10.1109/CVPR42600.2020.01092).