International Journal of Science, Strategic Management and Technology
Volume **02** Issue **02** February-2026 | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

OPEN ACCESS

A Study on a Journey Through Stories in Sequential Frames

Mr.Md.Abdul Kalam*1, B Abhinaya Jyothi*2, Rokoti Sowmya*3, K Rishitha*4,

*1Assistant Professor, CSE (AI & ML), ACE Engineering College, Hyderabad, India. *2,3,4Department CSE (AI & ML), ACE Engineering College, Hyderabad, India.

https://doi.org/10.55041/ijsmt.v1i1.001

Cite this Article: Rishitha, B. A. J. ,. R. S. ,. K. (2026). A Study on a Journey Through Stories in Sequential Frames. International Journal of Science, Strategic Management and Technology, *Volume 10*(01). https://doi.org/10.55041/ijsmt.v2i2.138

ABSTRACT:

This project is aimed at educating machines to narrate visual stories through creating descriptive and coherent text from image sequences. Based on analyzing objects, actions, emotions, and scene changes, the system constructs narrative text reflecting the story in the images. It intends to simulate human storytelling by extracting temporal and contextual relationships among images. Applications are automatic photo album description, digital storytelling, and assistive technology for the visually impaired. By closing the gap between visual perception and language generation, this approach enhances the way machines interpret and describe sophisticated visual experiences.

Keywords: Visual Storytelling,Image Sequence Analysis, Narrative Generation,Scene Understanding, Object and Action Recognition,Emotion Detection,Temporal Context Modeling, Multimodal Learning,Assistive Technology, Image-to-Text Synthesis

INTRODUCTION:

Visual storytelling takes a series of images and turns them into a cohesive, coherent story. It's not simply adding captions—it's seeing the sequence of events between several pictures to produce a smooth, natural flow of narrative.This challenge combines two of the most important areas of AI: computer vision, which allows machines to understand images, and natural language processing, which allows them to produce text. With current deep learning methods, it is possible to instruct computers to learn about people, objects, and activities in images and then create a story in which everything sounds believable to humans.

It is difficult to build such a system. It's not merely about identifying objects, but also understanding the emotions, events, and narrative structure behind them. Due to recent breakthroughs in deep learning, particularly in computer vision and NLP, we are closer than ever to making this feasible.In this project, our goal is to create a system that can look at a set of images and write a clear, engaging story that feels like it was written by a human.

BACKGROUND OF THE PROJECT:

This work tackles the narrative visualization challenges posed by large-scale, complex, and time-evolving datasets. To enable both exploration and presentation of insights well, the authors introduce Temporal Summary Images (TSIs) — a

solution that integrates three components: temporal layout (e.g., a line chart), data snapshots (comic strip-style panels), and text annotations (identifying principal events). TSIs simplify the otherwise time-consuming and manually intensive process of narrating stories using interactive tools that suggest significant features automatically and create annotations, thereby simplifying data storytelling to become quicker and more intuitive.

LITERATURE REVIEW:

1.Title: Narrative Visualization

Authors : Segel and Heer [1]:

This paper classifies types of narrative visualization like annotated charts and comic-strip formats, highlighting the ways in which narrative frames (e.g., text, layout, and imagery) influence viewers' interpretations.

It is commonly applied to learning basic narrative structures and styles in data storytelling. It does not include implementation details, nor does it discuss interactive or automated mechanisms for storytelling**.**

2.Title: Sequential Vision Language as Story: Storytelling Dataset Benchmarking

Authors: Z. M. Malakan, S. Anwar[2]:

The authors present the SSID dataset, which contains hand-curated five-image sequences, and show its strength in producing more coherent, relevant, and informative stories compared to prior benchmarks suc h as VIST.It is mainly employed to benchmark and test AI storytelling models. Its weakness is fixed-length image sequence lengths and not supporting user-driven or dynamic storytelling.

3.Title: Vision Transformer Based Model for Describing a Set of Images as a Story

Authors: M. Hassan, A. Mian [3]:

This work introduces a storytelling model that combines a Vision Transformer for image feature extraction, a Mogrifier-LSTM for story generation, and a Bi-LSTM for sequence encoding, which performs better than previous models on a range of metrics. It is applied to automatic story generation from image sets with enhanced coherence. However, the model is computationally expensive and offers little transparency or user control over the generated stories.

4.Title: Temporal Summary Images

Authors: C. Bryan, K. L. Ma [4]:

The authors introduce a visualization technique called Temporal Summary Images (TSIs), combining a temporal layout, comic-strip-like data snapshots, and self-contained textual annotations to facilitate narrative construction from time-varying datasets. This is applied in both data analysis and storytelling, particularly in scientific and spatiotemporal settings. Nonetheless, its interface is difficult to use for laymen, and the numerous auto-generated annotations are overwhelming to non-experts who are not utilizing effective filtering mechanisms.

5.Title: Optimized Storytelling Comics-Based for Image Sequences Temporal

Authors: W.-T. Chu, C.-H. Yu, H.-H. Wang [5]:

This work positions visual storytelling as a labeling problem and uses a Genetic Algorithm to produce well-structured comic pages from sequential image data, which closely follows human layout preferences. It

is employed to automatically create visual storytelling from
media such as images. The method primarily optimizes the structure of layout but does not deal with semantic depth or narrative quality of the content

6.Title: Visual-Textual Modeling for Coherence

Authors: M. Alikhani, P. Sharma, S. Li, R. Soricut, M. Stone [6]:

This work integrates discourse coherence theory into image captioning models to enhance textual description-image content alignment for more coherent narrative flow. It is applied to improve the semantic quality and logical consistency of multimodal story generation. The limitation is that it adds model complexity and remains highly reliant on text models, with minimal multimodal integration beyond caption-level coherence

COMPARISON TABLE:

S. No	Author(s)	Title	Methodology Used	Findings from the Reference Paper
1	Zainy M. Malakan, Saeed Anwar, Ghulam Mubashar Hassan	Sequential Vision Language as Story: Storytelling Dataset Benchmarking	Constructed Sequential Storytelling Images Dataset (SSID) from open-source text; five sequential images per set	SSID is harder and more coherent than VIST; scored higher on coherence, relevance, and informativeness
2	Chris Bryan, Kwan-Liu Ma	Temporal Summary Images: An Approach to Narrative Visualization	Constructed Temporal Summary Images (TSIs) framework for visualizing multivariate, time-varying datasets	TSIs facilitate better data analysis and storytelling;automatic annotations emphasize most important data points

3	Wei-Ta Chu, Chia-Hsiang Yu, Hsin-Han Wang	Optimized Storytelling Comics-Based for Image Sequences Temporal	Labeling problem addressed by Genetic Algorithm (GA) to cluster images into coherent comics pages	Cosmicsclosely resemble human arrangements;generalized well to videos, photo albums, and films .
4	Zainy M. Malakan, Ghulam Mubashar	Vision TransformerBased Model for Describing a Set of Images as a Story	ViT for feature extraction, Mogrifier-LSTM for storytelling generation, Bi-LSTM for sequence encoding, Attention Mechanism	Model outperformed all previous methods in generating coherent and relevant image stories
5	Md Sultan Al Nahian, Tasmia Tasrin,Ryan Gaines, Brent Harrison	A Hierarchical Approach for Visual Storytelling Using Image Description	Introduced Hierarchical Context-Based Network (HCBNet) incorporating visual and textual information	HCBNet outperformed baselines (GLAC Net, AREL) on BLEU, CIDEr, and METEOR scores

6	Malihe Alikhani, Piyush Sharma,	Visual-Textual Modeling for Coherence	Integrated discourse theory coherence relations into image captioning models	Coherence-aware models generated more consistent andgoal-consistent captions
7	Segel and Heer	Narrative Visualization	The authors developed a taxonomy of genres,narrative organization, and patternsof user engagement	The pairing of visual narrative and interactive elements enhances comprehension and reader involvement. Effective story visualizations reconcile authorimposed organization withreader-driven inquiry.
8	Niklas Elmqvist	Storytelling in Information Visualizations	They employed a mixed-method strategy combining images and analysis with user studies to assess the impact of storytelling methods on user engagement in visualizations	Narrativeelements improve userengagement, understanding, andmemorabilityin information visualization. Nevertheless, too much structure limits exploration, emphasizing a balance between storytelling and user agency

Table 1: Review of Existing Research on A Journey Through Stories in Sequential Frames

RESEARCH GAPS IN EXISTING SYSTEMS:

Based on the literature review, several research gaps have been identified in the development of Flappy Bird AI using Reinforcement Learning (RL) and related intelligent learning approaches:

1.Limited Coherence Over Longer Sequences:

Most models are good at short image sequences but find it difficult to keep narrative coherence over longer, more intricate stories.

2.Low Diversity in Story Outputs

Stories tend to sound formulaic or repetitive because models are optimized for average performance (e.g., BLEU, METEOR) but not for diverse and creative storytelling.

3.Multimodal Alignment Issues

Multimodal alignment between visual (image) and textual (story) representations is still challenging, particularly whenscenes involve actions, intricate relationships, or uncertain environments**.**

PROPOSED SYSTEM:

The suggested system creates meaningful narratives from sequential images by combining cutting-edge deep learning methods in computer vision and natural language processing. Combining CNNs for image comprehension, RNNs or Transformers for sequence modeling, and attention-based decoders for storytelling allows the system to replicate human-like narrative abilities and push the boundaries of automated content production. An attentive decoder then produces a narrative that flows naturally. Language quality can be improved through post-processing using NLP models. The system is assessed using BLEU, ROUGE, and human feedback after being trained on datasets such as VIST.

CONCLUSION AND FUTURE SCOPE:

Conclusion

The proposed system successfully combines computer vision and natural language processing to generate coherent stories from sequential images. By leveraging CNNs for visual understanding and RNNs or Transformers for temporal and contextual modeling, the system can produce meaningful and fluent narratives that mimic human storytelling. This has vast potential in various fields including education, media, accessibility, and surveillance. The approach demonstrates that deep learning can bridge the gap between visual data and language generation effectively.

Future Scope

Future enhancements can focus on improving narrative creativity through reinforcement learning or generative models like GPT. Incorporating user inputs such as storytelling style or emotion could personalize outputs. Additionally, using multimodal data (e.g., audio, video) and expanding to multilingual storytelling would increase the system's versatility. Real-time story generation and integration into mobile or web platforms could also make the system more accessible and practical for everyday use.

REFERENCES:

1. Zainy M. Malakan, Saeed Anwar, Ghulam Mubashar Hassan, "Sequential Vision Language as Story: Storytelling Dataset Benchmarking," Conference Paper, 2022.
2. Chris Bryan, Kwan-Liu Ma, "Temporal Summary Images: An Approach to Narrative Visualization via Interactive Annotation Generation and Placement," IEEE Transactions on Visualization and Computer Graphics, vol. 23, no. 1, pp. 511-520, 2017.
3. Wei-Ta Chu, Chia-Hsiang Yu, Hsin-Han Wang, "Optimized Storytelling Comics-Based for Image Sequences Temporal," Multimedia Tools and Applications, Springer, 2015.
4. Zainy M. Malakan, Ghulam Mubashar Hassan, Ajmal Mian, "Vision Transformer Based Model for Describing a Set of Images as a Story," Research Paper, 2023.

5. Md Sultan Al Nahian, Tasmia Tasrin, Sagar Gandhi, Ryan Gaines, Brent Harrison, "A Hierarchical Approach for Visual Storytelling Using Image Description," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021.
6. Malihe Alikhani, Piyush Sharma, Shengjie Li, Radu Soricut, Matthew Stone, "Visual-Textual Modeling for Coherence," Findings of the Association for Computational Linguistics (ACL), 2020.
7. Linyi Jin, Yukun Zhu, Chen Sun, Kevin Murphy, Li Fei-Fei, Juan Carlos Niebles, "DeepStory: Video Story QA by Deep Embedded Memory Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
8. Peter J. Liu, Mohammad Saleh, Et al., "Generating Wikipedia by Summarizing Long Sequences," International Conference on Learning Representations (ICLR), 2018.

Harsh Agrawal, et al., "Don't Just Listen, Use Your Imagination: Leveraging Visual Common Sense for Non-Visual Tasks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017.