
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2023 SCPE. Volume 24, Issues 4, pp. 1137–1144, DOI 10.12694/scpe.v24i4.2434

MULTIMODAL ANALYSIS OF IMAGE TEXTS IN COLLEGE ENGLISH TEXTBOOKS USING VISUAL GRAMMAR THEORY FOR REAL-TIME EMBEDED SYSTEMS

DEXIA LI*

Abstract. The existence of images in the English textbook can bring several benefits to colleges. Research has been done on multimodality to analyze the impact of the image text within English textbooks. This study uses visual grammar to identify image and text elements within textbooks. The multimodal framework is involved in "Van Leeuwen's Grammar Visual Design, which has been used in this study to identify multimodal elements that can appear in the English textbook. Image or visual elements can provide pedagogical advantages to college students to understand the text.

Key words: Visual Grammar, Multimodal Analysis, Image Texts, real-time embedded system

1. Introduction. The multimodal analysis of the text image is one of the packages of software to analyze, learn, and teach multimodal texts. Multimodal analysis is one of the key software packages in the market that not only provide theoretical understanding but also the practicability of learning and teaching in different circumstances [1]. In the case of linguistics, multimodal analysis has become one of the key approaches to global language learning. The process of multimodal analysis is found using different forms of language, such as texts, images, and other forms. However, the process of image text is found most crucial as it helps to uncover the maximum number of semiotic resources.

The employment of appealing visual features in instructional materials has become increasingly common in today's quickly expanding digital world, notably in college English textbooks. It is commonly acknowledged that combining visual graphics and textual text may considerably improve the overall learning experience and successfully assist in a better grasp of complex subjects. As a result, this paper goes into the interesting worlds of multimodal analysis and visual grammar theory, focusing on their enormous influence on analyzing picture texts found in college English textbooks. Furthermore, this investigation focuses on the amazing applicability of these analytical frameworks to the dynamic world of real-time embedded systems.

The above figure presents the multimodal analysis framework involved in the context, disclosure, and genre, including the textual and visual layout of images. The research also discusses the use of English Grammar Theory in the process of multimodal text analysis. The above image shows the different structures revolving around multimodal text analysis and the role of visual grammar in this section [3, 2].

The presence of images within college English textbooks has become increasingly recognized for its potential to enhance the learning experience. The convergence of visual and textual elements in educational materials has prompted research into multimodality, focusing on how the combination of image and text impacts comprehension and engagement. This study aims to extend the understanding of this phenomenon by utilizing the principles of visual grammar to delve into the intricacies of multimodal analysis of image-text interactions within college English textbooks. Moreover, this exploration is conducted with an awareness of the implications of real-time embedded systems, which can potentially facilitate the integration and interpretation of these multimodal elements.

The steps of the Multimodal analysis framework of the image text have been given in the above table.

The aim of this study is a multimodal analysis of the image texts in English textbooks using the theory of visual grammar. While this study has shed light on these two domains' potential synergies and applications, a more detailed exploration of how real-time embedded systems can effectively facilitate and enhance multimodal

*School of International Education, Xianda College of Economics and Humanities Shanghai International Studies University, Shanghai 200083, China (dexialiresearc1@outlook.com)

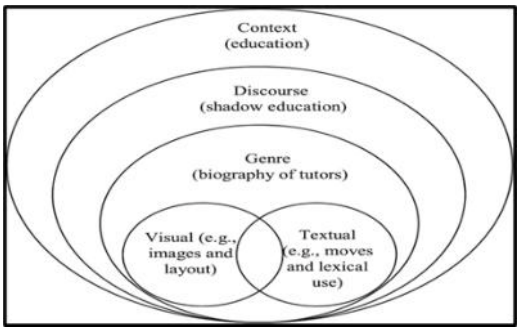

Fig. 1.1: Multimodal analysis framework of Image text

Table 1.1: Steps involved in Multimodal analysis framework of Image text

Steps	Consideration
Identification of Mode	This includes various visual modes like colors, images typography and layout
Semiotic analysis	Analysis of individual visual elements or semiotic resources like framing, composition, syntax as well as symbolism
Intermodal analysis	Exploring relationships and interactions between various modes of image text
Meaning-making	This step is related to the interpretation of meanings that are conveyed by image text and the audience has a crucial role here
Contextual factors	this Consider the cultural, social as well as historical context that can be present around the image text.
Audience analysis	This analysis reflects on how audiences can be different in terms of responding to and interpreting the image text, which depends on the personal, social, and cultural contexts.
Critical reflection	This step involves critical reflection as well as analysis of image text. Here, is an assessment of the potential implications, ethical considerations, and effects associated with the utilization of the visual modes of image text.

text analysis would provide a deeper understanding of their interplay. This could involve delving into the technical intricacies of integrating real-time processing capabilities within text analysis frameworks, ensuring seamless synchronization and efficient data flow. Additionally, investigating practical use cases where real-time embedded systems can mitigate challenges and optimize processes in multimodal text analysis, such as real-time sentiment analysis of social media streams or live transcription and translation of multimedia content, could further underscore the significance of this connection. By addressing these aspects, the study would contribute to a more comprehensive comprehension of the potential at the intersection of real-time embedded systems and multimodal text analysis.

2. Objectives.

1. To evaluate the concept of the multimodal analysis of image texts and the theory of visual grammar
2. To determine the impact of using the theory of visual grammar in the multiple analysis of image texts for English college textbooks
3. To identify the impacts of using visual learning on the “real-time embedded system” .

Motivation: The motivation behind this study lies in the growing recognition of images’ role in facilitating effective learning, particularly in the context of college English education. As students engage with complex textual content, visual elements can serve as valuable aids in comprehension. However, the nuanced ways images and texts interact require a deeper analysis considering linguistic and visual components. Furthermore, the emergence of real-time embedded systems presents opportunities to dynamically enhance the interaction

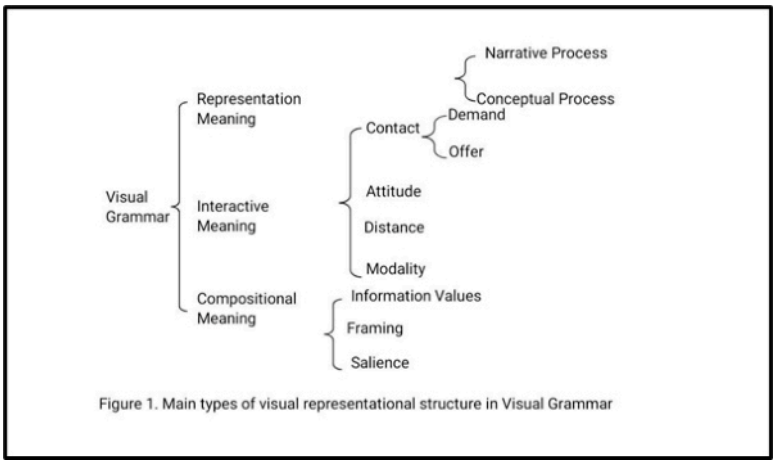

Fig. 4.1: Visual grammar framework

between image and text, potentially leading to more immersive and interactive learning experiences.

Contribution: This study contributes to the existing body of knowledge by applying the theory of visual grammar, specifically "Van Leeuwen's Grammar of Visual Design," to analyze image-text relationships within college English textbooks. By employing this framework, the study aims to systematically identify and categorize multimodal elements, shedding light on how images enhance textual understanding. Moreover, considering real-time embedded systems in this context adds a futuristic dimension, proposing ways in which technological advancements can further augment the multimodal learning experience.

3. Methodology. A descriptive research design has been employed to conduct the secondary qualitative research. This method is an effective way to code, screening as well, and interpret. This research design was in line as data have been collected, categorized as well and analyzed.

Secondary thematic analysis has been done as a data analysis process to conduct the study to determine the multimodal analysis of the image texts within English textbooks and the benefits of using the theory of visual grammar [4, 13]. Therefore, subjective data have been gathered and analyzed by making themes based on the study's objectives.

4. Concept of the Multimodal Analysis of Image Texts and Visual Grammar Theory. Based on the above figure, it can be seen that the main type of visual representation is visual grammar. The concept of multimodal analysis is embedded in the use of semiotic modes. These models are the primary factors of textbooks that work as an essential factor in analyzing image texts [7]. Due to this reason, the relationship between images and texts is one of the key things in the multimodal analysis of text images. The arrangement of their images, the position of images and texts in the visual fields, and the role of the images in the meaning-making process are very important in this case. The above picture discusses the different types of visual grammar used in standard multimodal analysis [6]. The representation of the images, the composition of the images, and the interaction between the images and the texts are the primary visual grammar used in the text analysis process. However, all these forms of images have different subsections and divisions based on the type of grammar used in the analysis process.

Understanding the intricacies of college English textbooks can be done through multimodal analysis, which is the research of various communication modalities within a given context. Along with written content, multimodal analysis also looks at images, sounds, and motions. By utilizing the concepts of visual grammar theory, researchers gain a deeper understanding of the visual components present in various communication channels. By giving researchers a framework for examining crucial elements like composition, color, perspective, symbolism, and visual metaphors, visual grammar theory enables them to understand the hidden visual cues contained in picture texts in college English textbooks.

Table 5.1: Elements of visual grammar

Elements	Description
Objects	The basic element of visual grammar can be concrete or abstract
Structures	Patterns of the visual grammar should be created based on the basic element as well as patterns
Activities	The process can be represented based on the basic element and the patterns
Relations	This includes the relationship among the process, patterns, and elements

5. Uses of Theory Visual Grammar for the Multimodal Analysis in College English Textbooks. The use of visual grammar has been found as one of the main tools for analyzing college English textbooks. The use of this technology in textbook representations is found to consist of conceptual and narrative representations [9]. The representation can be done correctly based on the structure of the display, texts, and the interrelation between texts and images. It has been found that cohesiveness is one of the key factors that can be achieved by the proper implementation of the visual grammar theory. The use of visual characteristics in college English textbooks can be better understood using the framework provided by visual grammar theory [10]. Using this approach, students can examine a variety of examples, charts, graphs, photographs, and other visual elements that are frequently included in textbooks in greater detail. By carefully examining the composition, color schemes, and visual metaphors used in these picture texts, researchers can decipher not just the intended emphasis but also the emotions and underlying concepts portrayed [11]. This thorough investigation is essential in enabling teachers to provide more useful teaching resources, which in turn aids students in deriving pertinent and significant information from the visual texts they encounter.

The assimilation between synonyms, repetition, and visual models creates a perfect cohesiveness of text and images, which helps students as well as teachers to convey the messages in the textbooks. In the case of English textbooks, it has been found that the higher the cohesiveness between the texts and images, the greater the utilization of knowledge [11]. Students can comprehend more knowledge from a textbook, which provides synchronization between the texts and the images present in the English textbooks. Multimodal analysis of picture texts using visual grammar theory has several advantages in the setting of real-time embedded systems. It enhances learners' comprehension and retention, to start. By examining the visual elements in real-time embedded systems featured in college English textbooks, students can learn difficult subjects and remember information better. Multiple entrance points for comprehension are made possible by the combination of written and visual data, suiting different learning styles.

6. Advantages of Analysis of the Image Text by using Visual Grammer Theory in case of the Real-time Embedded System. Platforms of interactive language learning can be the "real-time embedded system". This is the platform where learners can get the opportunity to engage in "real-time conversation". This system is able to analyze patterns of speech, give instant feedback as well as can offer vocabulary and grammar.

The above figure depicts the advantages of the multimodal analysis for the real-time embedded system. This system is capable of adopting the needs of an individual as well as fixing the progress of learners by monitoring the performance of learners. "Real-time embedded system", therefore can be integrated to analyze the image text for providing instant feedback and an interactive experience of learning [14]. College students can interact with the system by uploading images or capturing images of their textbooks or any specific image texts. These systems can employ the vision of the computer algorithms for processing texts and images as well as extract visual features for analysis.

The embedded system serves the pedagogical support to instructors and learners. For instance, tutorials of visual grammar can be system generated, and interactive exercises or supplementary materials can be developed based on the analysis of the image texts. Therefore, learners can access these resources anytime. On the other hand, it enables independent learning and provides various additional opportunities such as reinforcement and practice. This has advantages in the preventive algorithm in scheduling. However, it can be a disadvantage in the learning courses, as this system cannot prioritize any tasks [16]. One task can go through multiple rounds

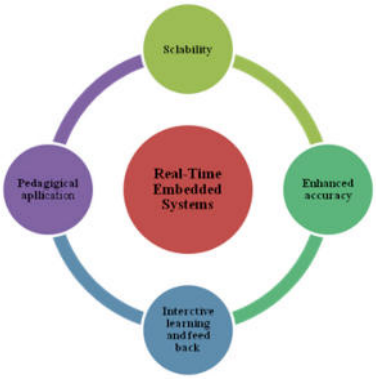

Fig. 6.1: Advantages of the real-time embedded system

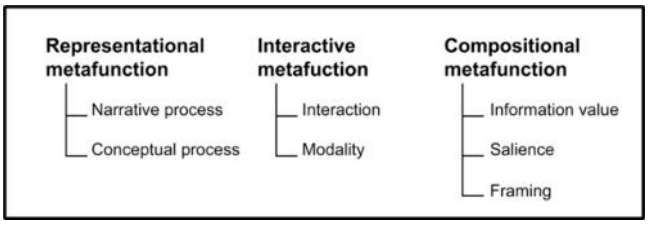

Fig. 7.1: Categories of the visual grammar theory

for completion; therefore, it can be said that this approach is relatively straightforward and easy. Improved Teaching Experience: Adding visually appealing image texts to college English textbooks can increase student engagement with the subject matter and lead to a better learning environment.

7. Theoretical Framework. Van Leeuwen's Grammar Visual Design

This framework focuses on the grammar as well as principles that are underlying visual design and communication. This theory provides a systematic analysis of the different types of visual images as well as their interaction for catering and managing within visual texts. In the case of visual grammar, this involves different categories as well as principles to analyze visual communication [18, 5]. This grammar involves the understanding of the process of the visual elements that can be combined and organized for developing and conveying a specific meaning. This includes several categories like compositional structures, representational structures as well and interactive Meta functions.

The above figure depicts that three categories of the visual grammar theory have different processes representational conceptual and narrative processes. While narrative representation shows every participant that is connected to one another by lines, which are called vectors. On the other hand, representational representation can show a stable concept.

8. Result. The visual grammar theory has been used in the image work to communicate messages just as words do. Like every language, the English language has its grammar as well as an understanding of the uses of images for communicating clearly by visual grammar [17].

The above figure depicts a relationship between teaching diversification in English and teaching effectiveness based on the above graph it can be said that diversification of the various modes of teaching as well as teaching effectiveness is more remarkable which has an impact on the rising intention of the student learning [8].

The real-time embedded system can be operated based on several principles like quick response, predictability, deadline of the task as well as operational failures. This system can work in fixed constraints of time. Apart from this system can be predictable or deterministic and does not allow any deviations. Meeting deadlines for

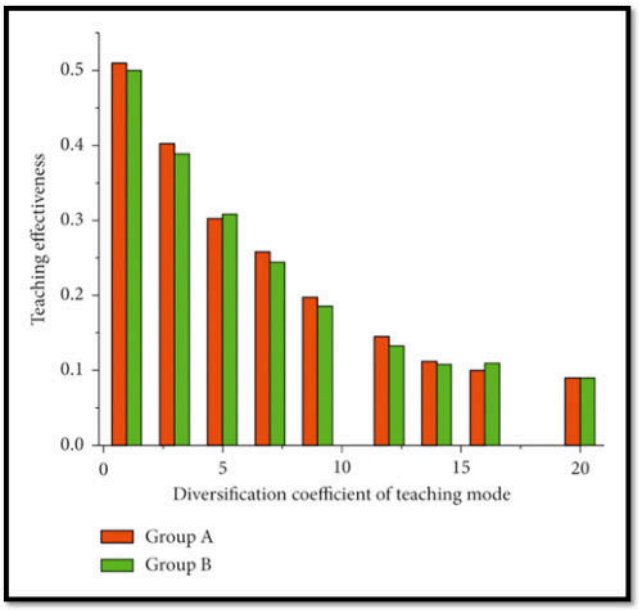

Fig. 8.1: Relationship between the teaching diversification and its effectiveness

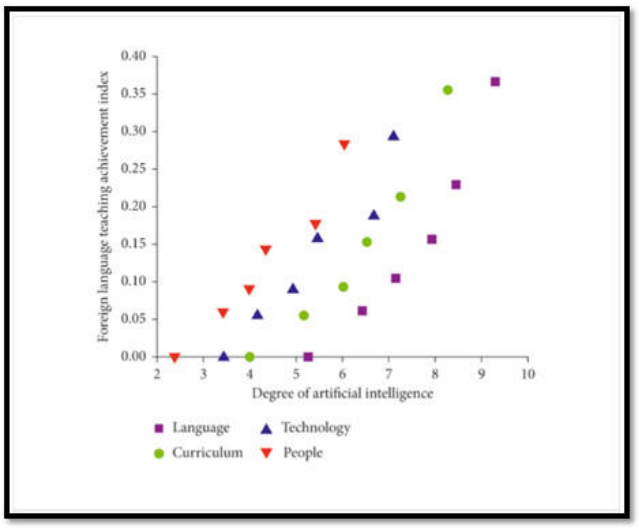

Fig. 8.2: Relationship between foreign language and uses of artificial intelligence

tasks is more crucial for this system than other characteristics. If any tasks fail to meet the deadline, then it can have a negative impact on the users as well as even lead the fatal results. Abstract objects of visual grammar are included in the lines, points, surfaces as well and volumes [18]. Volumes of the abstract objects in English learning textbooks involve three dimensions that are lines, points, and surfaces.

The above figure presents a scatter diagram to establish the relationship between the uses of AI in the case of real-time embedded systems. From the above figure, it can be seen that there are four-dimensional systems such as language, technology, curriculum as well as people or audiences [15]. This manual cooperation can increase the effectiveness of English teaching. In order to use multimodal analysis can produce multimodal

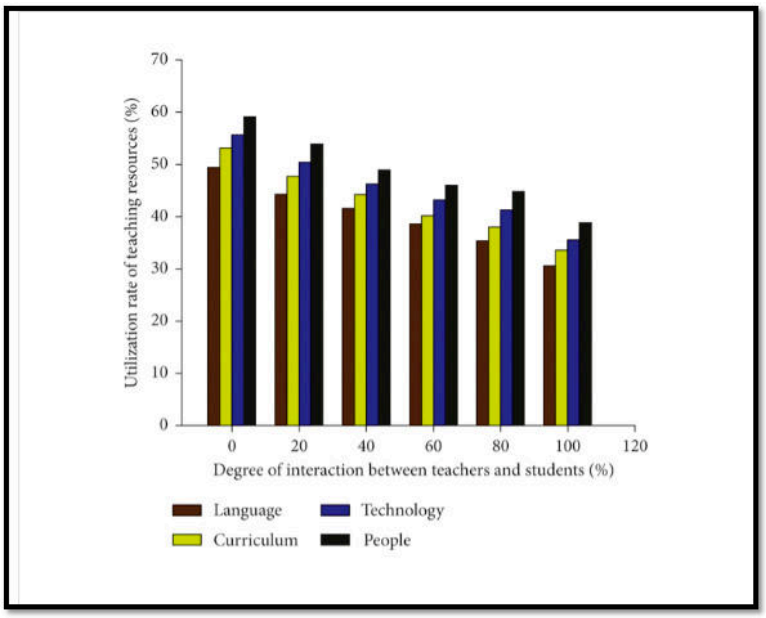

Fig. 9.1: Relationship between the various teaching resources and interaction among teachers and students

recognition within teaching space, time, resources, and audiences.

9. Discussion. A real-time embedded system can be defined as a particular type of embedded system that works with real-time computing through an RTOS embedded system. This system follows a quick response pattern to the external stimulus to work within a proper period. In the case of multimodal text analysis for English textbooks, the use of real-time embedded systems can be found very important [14].

The above figure presents the relationship between teaching resources and the highest teaching resource is its people than technology. The main basis of English learning is pure language however in the real-time embedded system the companion language is one of the crucial components. This has a role in the real-time embedded system as a supplementary, auxiliary, as well as reinforcing role within the transmission of the English language [12, 8]. This transmission can be done by using color, shape, spatial layout, sound, font size, and others.

Real-time embedded systems can be developed through the symbiosis of English textbooks based on the use of visual grammar. If the multimodal analysis is found fruitful, the cohesive communication between the textbook materials can accelerate the process of real-time embeddedness. Students' visual literacy skills are improved when visual grammar theory is used [8, 15]. Through multimodal analysis, students develop the ability to evaluate and grasp the visual elements in picture texts. Learners, who comprehend the foundations of visual composition, color, and symbolism may use and assimilate visual information more selectively. Since complex visuals are frequently seen in real-time embedded systems, this knowledge is extremely helpful.

10. Conclusion. Based on the above discussion it can be said that several empirical studies focus on the representation of digital English textbooks. Multimodal analysis can be divided into two groups, which depend on the process of making conceptualizing meaning. Multimodal elements encourage fostering multimodal relationships to foster a relationship between the readers and text producers. Visual grammar theory can be used examine how to textual and visual content can encode as well as communicate ideas about the real world in terms of representation. Real-time embedded systems can be involved in the soft and hard real-time as well as the non-preemptive and preemptive scheduling. "Real-time embedded system" is effective as it can simplify several processes as well as increase the quality of life.

While this study provides valuable insights, it is important to acknowledge its limitations. The scope of this

analysis primarily focuses on the theoretical aspects of visual grammar theory and its application to multimodal analysis. Further empirical studies could delve into the practical implementation of these theories in actual digital English textbooks, assessing their impact on student engagement and comprehension.

Moreover, the exploration of real-time embedded systems opens avenues for future research. As this study merely touches upon the potential of real-time embedded systems in enhancing various processes and improving quality of life, in-depth investigations can delve into the intricacies of their integration within educational settings. Future studies may also consider exploring the challenges and ethical considerations associated with the integration of technology, particularly in educational contexts. Additionally, further research can delve into the cognitive and pedagogical implications of different types of image-text interactions, contributing to a deeper understanding of effective instructional design. This study paves the way for future investigations into the dynamic fusion of visual grammar theory, multimodal analysis, and real-time embedded systems in educational contexts.

REFERENCES

[1] B. Asghar, M. Azher, and S. Fatima, *Gender portrayal on cover pages of pakistani english children's literature: A visual semiotic analysis*, Journal of Gender and Social Issues, (2022).

[2] V. B, M. S, P. N, J. L, N. V, and K. S, *Artificial conversational entity with regional language*, in 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1–6.

[3] N. Cohn, *Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity*, Narrative complexity and media: Experiential and cognitive interfaces, (2019), pp. 174–199.

[4] F. Huang, X. Zhang, Z. Zhao, J. Xu, and Z. Li, *Image–text sentiment analysis via deep multimodal attentive fusion*, Knowledge-Based Systems, 167 (2019), pp. 26–37.

[5] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, J. Lorandel, and O. Romain, *Radar signal processing for sensing in assisted living: The challenges associated with real-time implementation of emerging algorithms*, IEEE Signal Processing Magazine, 36 (2019), pp. 29–41.

[6] J. E. Liu and A. M. Lin, *(re) conceptualizing "language" in clil: Multimodality, translanguaging and trans-semiotizing in clil*, AILA Review, 34 (2021), pp. 240–261.

[7] Z. Liu, *Introducing a multimodal perspective to emotional variables in second language acquisition education: Systemic functional multimodal discourse analysis*, Frontiers in Psychology, 13 (2022), p. 1016441.

[8] L. Niu, *Analysis of multimodal teaching of college english under the background of artificial intelligence*, Security and Communication Networks, 2022 (2022).

[9] M. Petryk, M. Rivera, S. Bhattacharya, L. Qiu, and S. Kumar, *How network embeddedness affects real-time performance feedback: An empirical investigation*, Information Systems Research, 33 (2022), pp. 1467–1489.

[10] M. Purgina, M. Mozgovoy, and J. Blake, *Wordbricks: Mobile technology and visual grammar formalism for gamification of natural language grammar acquisition*, Journal of educational computing research, 58 (2020), pp. 126–159.

[11] R. P. Sari and T. Tawami, *Visual-text integration in digital public service announcement: A strategy to fight fake news spread*, Aksara, 35 (2023), pp. 125–137.

[12] S. Sun, Y.-C. Chen, L. Li, S. Wang, Y. Fang, and J. Liu, *Lightningdot: Pre-training visual-semantic embeddings for real-time image-text retrieval*, in Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 982–997.

[13] B. Vivek, A. Arulmurugan, S. Maheswaran, S. Dhamodharan, A. S. Dharunash, and N. Gowtham, *Design and implementation of physical unclonable function in field programmable gate array*, in 2023 8th International Conference on Communication and Electronics Systems (ICCES), 2023, pp. 152–158.

[14] X. Yang, S. Feng, D. Wang, and Y. Zhang, *Image-text multimodal emotion classification via multi-view attentional network*, IEEE Transactions on Multimedia, 23 (2020), pp. 4014–4026.

[15] F. Zhan, Y. Yu, R. Wu, J. Zhang, S. Lu, L. Liu, A. Kortylewski, C. Theobalt, and E. Xing, *Multimodal image synthesis and editing: A survey*, arXiv preprint arXiv:2112.13592, (2021).

[16] Z. Zhao, H. Zhu, Z. Xue, Z. Liu, J. Tian, M. C. H. Chua, and M. Liu, *An image-text consistency driven multimodal sentiment analysis approach for social media*, Information Processing & Management, 56 (2019), p. 102097.

[17] J. Zhu, Y. Zhou, J. Zhang, H. Li, C. Zong, and C. Li, *Multimodal summarization with guidance of multimodal reference*, in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 9749–9756.

[18] R. Zhussupova, *Teaching english using multimodal texts*, EDITOR: Dzintra ILIŠKO, Professor, Dr, (2023), p. 28.

Edited by: Sathishkumar V E
Special issue on: Scalability and Sustainability in Distributed Sensor Networks
Received: Jul 21, 2023
Accepted: Sep 8, 2023