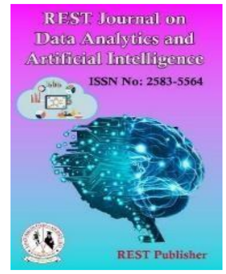
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/55



Poster Generation Using Stable Diffusion 3 and Auto Encoders

***G. Victor Daniel, Y Vinod, K Ruchitha**
Anurag University, Hyderabad, Telangana, India
Corresponding Author Email: 21eg107a53@anurag.edu.in

Abstract. *Creating posters can feel like a hassle, especially if design isn't your strong suit. It often requires a lot of manual adjustments and artistic decisions, which can be overwhelming. That's where PosterGen steps in—a framework designed to simplify and streamline the process. Using a large-scale visual-textual model, PosterGen can automatically find background images that complement your text. It also employs auto encoders to arrange the text in a visually appealing and easy-to-read way. On top of that, it selects fonts and colors that fit the theme, ensuring everything flows seamlessly. What's even more impressive is that PosterGen learns with very little labeled data by using weakly- and self-supervised methods. In our experiments, we've seen that it not only makes poster creation easier but also outperforms some of the current tools and techniques available.*

1. INTRODUCTION

Creating professional and engaging event posters is simplified by the automated poster generation system, which uses advanced machine learning technologies, including Stable Diffusion 3 (SD3). This robust generative model generates thematic background images from descriptive text prompts, enhanced by negative prompts to ensure visuals align with specific styles. Cascaded auto encoders optimize text placement to maintain visual harmony and readability, while text styling is achieved by matching fonts and colors to the event's theme using a predefined cluster of semantic styles. By employing weakly- and self-supervised learning, the system minimizes reliance on large datasets, allowing for scalability and continuous adaptation to new design trends and user preferences, thus making the poster creation process large-scale, efficient, and adaptive, improving output quality with minimal manual oversight [1-4]. This image generation process is further enhanced by the use of negative prompts, which filter out elements that may not fit the design. So, if the event's description calls for minimalism or certain visual styles, SD3 can ensure the output is clean and on-point, avoiding distractions or clutter [5-8]. Once the background is set, the next challenge is making sure the text fits well without disrupting the overall look. This is where cascaded auto encoders come into play [9]. Auto encoders are neural networks that are particularly good at compressing and reconstructing data—in this case, layout information. The cascaded approach means the system refines the layout over several stages to achieve the best possible balance between text and imagery [10]. The process starts by encoding the initial design into a compressed representation, which highlights key regions for text placement while avoiding cluttered or busy areas in the image. The auto encoder then decodes this compressed data back into a suggested layout. This iterative refinement ensures the text doesn't overwhelm the image but remains prominent and easy to read. The result is a balanced design that maintains the poster's visual appeal while effectively communicating the event's details [11].

2. LITERATURE SUMMARY

TABLE 1. 2. Literature Summary

S.No	Paper	Journal	Technique	Limitations
1	**Text2Poster:** Laying out Stylized Texts on Retrieved Images	IEEE(2024)	The models used in text-to-poster generation typically include CNNs ,RNNs, LSTMs ,and Transformers.	A limitation of the BriVL model's background image retrieval is that it relies on a fixed database, which may restrict the diversity and relevance of generated images for various themes
2	**PosterBot**: A System for Generating Posters of Scientific Papers with Neural Models	IEEE(2022)	The models used in Poster Bot are: Roberta, Transformer Encoder, Section Classifier, Extractive Summarization Model, and Self-Attention with Reference Relationships.	Requires manual customization and relies on human- labeled section importance

3. PROPOSED SYSTEM

The proposed system leverages advanced machine learning models to streamline the entire poster design process. It begins with the Stable Diffusion model generating images from text prompts, creating a visually appealing base. The Layouts Diatribe Model then determines optimal text placement using neural network techniques such as convolutional layers and batch normalization. For stylistic coherence, K-Means clustering organizes text attributes like font and color based on their semantic content. Finally, the system applies these optimized layouts to the image, ensuring a cohesive and attractive poster design that effectively communicates its intended message.

Image Generation with Stable Diffusion: The image generation begins when a user inputs a text prompt describing their envisioned image. Utilizing this prompt, the Stable Diffusion model, which harnesses a well-trained neural network, begins its work. This advanced model intricately synthesizes images by methodically applying learned transformations to a noise distribution. This process continues iteratively until the noise is transformed into a coherent and detailed image that corresponds with the initial text description [15-18]. Stable Diffusion's backbone is the demising diffusion probabilistic model, which is designed to gradually convert noise into a structured image through a sophisticated reverse process akin to a Markov chain. During the forward phase, where noise is systematically introduced into the image, the process is governed by the formula: xt+1= 1 − βtXt + $\beta t\epsilon t$ represent the predetermined noise levels, and ϵt is noise sampled from a normal (Gaussian) distribution [19-20]. This formula is pivotal as it dictates how noise is incrementally added to the image, setting the stage for the reverse process where this noise is methodically removed to recover the detailed final image from the initial textual prompt.

Layout Distribution with Layouts Distrib Model (Cascaded Auto-Encoder): The LayoutsDistribModel, a cascaded auto-encoder, is instrumental in determining optimal regions for text placement on images. At the heart of this model are convolutional layers (Conv2D), which extract spatial hierarchies of features from input masks. These layers operate by applying a convolution operation (∗) with filters defined by weights (W) and biases (b). The resulting feature maps are passed through a Rectified Linear Unit (ReLU) activation function, which introduces non-linearity by retaining only positive values and zeroing out negative values, thereby enhancing the model's ability to capture essential features without being overwhelmed by variations in lighting and color. Following convolution, batch normalization is applied to stabilize learning and accelerate convergence. This process adjusts and scales the activations by normalizing the output of the convolution layers using the formula:

$$y = ((x - \mu)/\sqrt{\sigma 2 + \epsilon})*\gamma+\beta$$

where x is the input to the layer, μ and σ^2 are the mean and variance of thatInput, and γ and β are parameters learned during training that scale and shift the normalized data. The model then employs average pooling to reduce the spatial dimensions of the feature maps. This step involves calculating the average of all values within a specific window on the feature map to down sample the data, which significantly reduces the computational complexity and helps to prevent over fitting. To enhance the resolution of the down sampled feature maps, convolutional transpose layers (ConvTranspose2d) are utilized. These layers effectively perform the reverse of convolutional layers by up sampling the input feature maps, thus increasing their resolution and enabling the model to construct a detailed output layout that specifies where text should be optimally placed on the image. Finally, the model employs a sigmoid activation function in the output layer, which is crucial for converting the output values to a range between 0 and 1. This is particularly useful for representing probabilities, such as the likelihood of a region being suitable for text placement. The sigmoid function is defined as $\sigma(x) = 1/(1 + e^{-x})$, which smoothly maps large negative values towards zero and large positive values towards one, making it ideal for binary classification tasks like determining suitable text areas. Overall, each component of the LayoutsDistribModel works in concert to ensure that the final text placement is both visually appealing and functionally effective, making the most of both the generated image and the textual content.

Text Stylizing Using K-Means Clustering: In the vibrant field of graphic design, text stylization plays a critical role, especially in the creation of compelling posters. To achieve a visually appealing and thematically cohesive design, the system employs K- Means clustering to effectively style the text. This method organizes text attributes—such as font, size, and color—into distinct groups that harmonize with the poster's overall message. The stylization process begins with Feature Extraction, where various attributes of the text, including font size and color, are analyzed and converted into feature vectors. These vectors capture the stylistic essence of the text, paving the way for the next step. Clustering Text Features is achieved through the K-Means algorithm, which partitions these feature vectors into predefined clusters. Each cluster is formed around shared characteristics such as similar font types or color schemes. The clustering is guided by the formula: $\min_S \sum_{i=1}^{k} \sum_{x \in S} \|x - \mu_i\|^2$ where x is a feature vector, S_i is a cluster, μ_i is the centroid of cluster SiS_iSi , and kkk is the number of clusters. This process ensures that text styles are grouped by visual similarity and potential functional use. Semantic Relevance is carefully considered during clustering. The algorithm adjusts clusters based on the semantic content of the text, ensuring that style choices are suitable for the text's purpose. For example, promotional materials may favor bold, attention-grabbing fonts, whereas formal communications might opt for more refined, elegant text styles. After clustering, Selection of Styles involves choosing the most appropriate text styles for the poster. The system selects from clusters that best match the semantic tags associated with the poster's content, ensuring that the chosen styles enhance both the visual appeal and the communicative effectiveness of the poster. The final step, Final Text Layout Application, involves meticulously applying the optimized text layout to the generated image. This crucial phase ensures that the text is integrated seamlessly with the visual elements, maintaining a cohesive and attractive design. This careful placement guarantees that each element of the poster works together to create a unified and compelling visual experience.

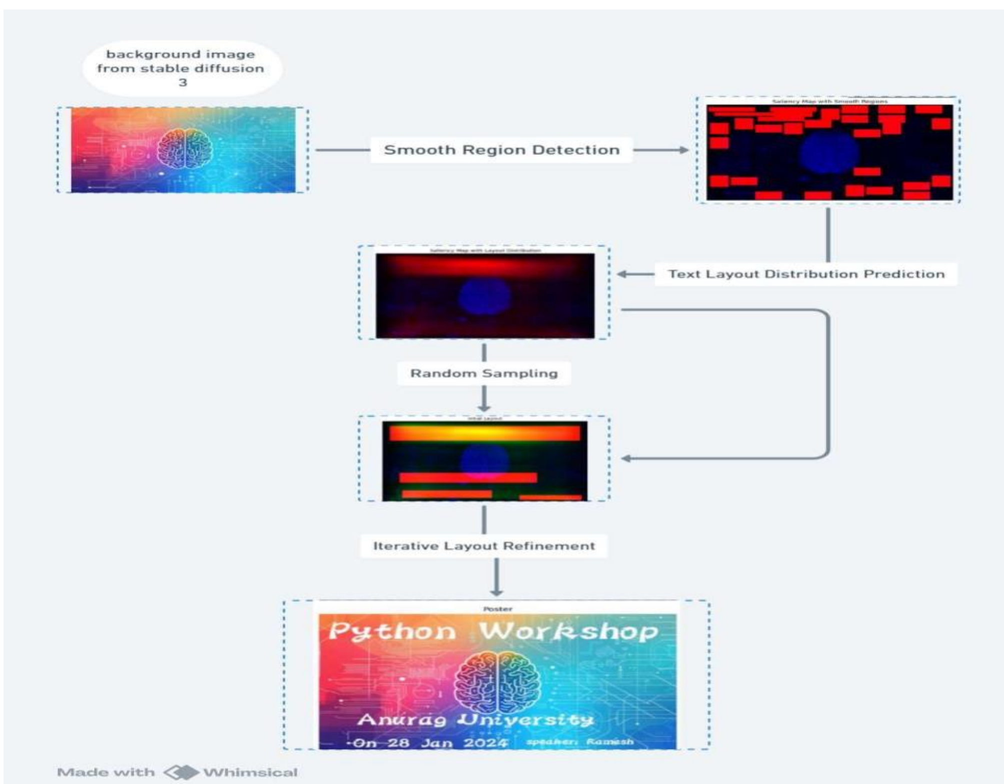


FIGURE 1. Layout detection using auto encoders

4. CONCLUSION

The Poster Gen project successfully addresses the challenges of automated poster design by leveraging advanced deep learning techniques for image generation, layout prediction, and text stylization. Through a systematic approach, the project harnesses the power of Stable Diffusion 3 (SD3) for generating contextually relevant background images, ensuring that each poster aligns with the intended message. The integration of spectral residual saliency detection allows for effective identification of smooth regions within the generated images, optimizing text placement for readability and aesthetic appeal. By utilizing cascaded auto-encoders for layout distribution prediction and iterative refinement through Convolutional Neural Networks (CNNs) and Bidirectional LSTMs, the system mimics the iterative nature of human designers, leading to optimal text positioning. Furthermore, the implementation of K-Means clustering for text stylization ensures that the font, size, and color of the text are semantically appropriate, enhancing the overall coherence of the poster. The final output is a visually appealing and contextually aligned poster, showcasing the effectiveness of combining various deep learning methodologies in creative design. Overall, Poster Gen not only streamlines the poster creation process but also sets the foundation for future advancements in automated graphic design, enabling users to produce high-quality promotional materials with minimal manual intervention. The project's success highlights the potential of machine learning to transform creative tasks, making them more accessible and efficient for a broader audience.

REFERENCES

[1] Papineni, S.L.V., Yarlagadda, S., Akkineni, H., Reddy, A.M. Big data analytics applying the fusion approach of multicriteria decision making with deep learning algorithms, International Journal of Engineering Trends and Technology, 2021, 69(1), pp. 24–28

[2] Reddy, A.M., Krishna, V.V., Sumalatha, L., Obulesh, A. Age classification using motif and statistical features derived on gradient facial images Recent Advances in Computer Science and Communications, 2020, 13(5), pp. 965–976

[3] Ravi, P., Obulesh, A., Reddy, A.M. Training-domain-process-evaluation framework for web-based industry- oriented mini-project Journal of Engineering Education Transformations, 2020, 33(Special Issue), pp. 329–333

[4] Prasad, M., Sreenivasu, M., Lakshmi, N., Reddy, A.M. Power consumption for routing improvement using AODV_EXT & AODV_EXT_BP, International Journal of Recent Technology and Engineering, 2019, 8(2 Special Issue 8), pp. 1639–1643

[5] Pruthvi Raj Goud, B., Anand Babu, G.L., Sekhar Reddy, G., Mallikarjuna Reddy, A. Multiple object detection interface using HSV, hough and haar-classifier International Journal of Innovative Technology and Exploring Engineering, 2019, 8(9), pp. 1512–1516

[6] S. K.Sarangi ,R.Panda & Manoranjan Dash," Design of 1-D and 2-D recursive filters using crossover bacterial foraging and cuckoo search techniques", Engineering Applications of Artificial Intelligence, Elsevier Science,vol.34, pp.109-121,May 2014.

[7] Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., "Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection", Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[8] Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[9] Manoranjan Dash, "Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment," International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[10] Manoranjan Dash et al.," Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique", Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[11] Manoranjan Dash et al.," Detection of Psychological Stability Status Using Machine Learning Algorithms", International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[12] Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[13] Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[14] Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[15] R. S. Kumar, K. K. Srinivas, A. Peddi and P. A. H. Vardhini, "Artificial Intelligence based Human Attention Detection through Brain Computer Interface for Health Care Monitoring," 2021 IEEE International Conference on Biomedical

Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 42-45, doi: 10.1109/BECITHCON54710.2021.9893646.

[16] Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.

[17] S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.

[18] Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055

[19] Rao, N.K., and G. S. Reddy. "Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm", International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.

[20] Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.

[21] Daniel, G. V., Chandrasekaran, K., Meenakshi, V., & Paneer, P. (2023). Robust Graph Neural-Network-Based Encoder for Node and Edge Deep Anomaly Detection on Attributed Networks. Electronics, 12(6), 1501. https://doi.org/10.3390/electronics12061501

[22] Victor Daniel, G., Trupthi, M., Sridhar Reddy, G., Mallikarjuna Reddy, A., & Hemanth Sai, K. (2025). AI Model Optimization Techniques. Model Optimization Methods for Efficient and Edge AI: Federated Learning Architectures, Frameworks and Applications, 87-108.

[23] Lakshmi, M.A., Victor Daniel, G., Srinivasa Rao, D. (2019). Initial Centroids for K-Means Using Nearest Neighbors and Feature Means. In: Wang, J., Reddy, G., Prasad, V., Reddy, V. (eds) Soft Computing and Signal Processing . Advances in Intelligent Systems and Computing, vol 900. Springer, Singapore. https://doi.org/10.1007/978-981-13-3600-3_3.