Advertisement using Customize LoRA

Diksha Sharma, Nikil, Shrey Kapri, Durgesh Kumar

Sharda University, Greater Noida, Uttar Pradesh

dikshasharma20113@gmail.com, nikilkadian1080@gmail.com, kapri.shrey@gmail.com,
durgesh.kumar1@sharda.ac.in

Abstract

This paper discusses the development of an avenue for personalized, pre-trained image generation models using Low-Rank Adaptation (LoRA) to confront some challenges of the advertising space, especially in clothing brands. Conventional advertising campaigns in the clothing industry often rely on costly, time-consuming photoshoots with professional models. The idea here is to utilize LoRA to finetune a base model and reduce the time spent distract you away from the costly issue by creating quality images of your brand. Our approach allows you to get high-quality images of your brand tailored to your advertising needs, without the need for physical models or large setups. By providing an effective Low-Rank Adapted (LoRA) image generation model using Replicate, the Flux base model, and Hugging Face to train it, our approach allows users to generate reliable and appealing images in the seconds it would take to shoot a physical ad campaign. The research offers an accessible and affordable alternative to traditional photoshoots, saving businesses money and time, while maintaining reliability, appeal and consistency with the brand. The results show that Low-Rank Adapted (LoRA) based customizations have the potential to be a disruptive resource for the production of aesthetic visual content for fashion marketing activations.

Keywords—*Low-Rank Adaptation, LoRA, image generation, advertising, clothing brands, model customization, Replicate, Flux model, Hugging Face, cost-effective marketing, visual content creation, fashion advertising*

1. INTRODUCTION

The advertising industry has long relied on high-quality visuals to capture consumer attention and drive brand engagement. Traditionally, the creation of compelling advertising images, especially for clothing brands, requires extensive photoshoots, professional models, and considerable financial and time investments. This process can pose significant barriers for emerging brands and restrict creative agility for established ones.

Recent advances in generative artificial intelligence, particularly in computer vision, offer novel solutions to these challenges. In particular, Low-Rank Adaptation (LoRA) techniques offer a powerful means to customise large pre-trained image generation models for specific downstream tasks, reducing the need for large datasets and expensive retraining. By fine-tuning models with domain-specific imagery, LoRA can achieve rapid, realistic image generation tailored to unique brand aesthetics or campaign concepts.

This research paper explores the integration and customization of LoRA within the Replicate platform, utilizing the Flux model architecture and Hugging Face ecosystem. Our aim is to demonstrate how clothing brands can automate the creation of high-quality, brand-consistent photographic imagery without hiring physical models or organizing traditional photoshoots. Through targeted training, our adapted LoRA model can produce visually appealing, context-appropriate advertising content within seconds. This advancement promises to substantially lower costs, accelerate campaign timelines, and open new possibilities for personalized and scalable creative workflows in advertising.

About LoRA

LoRA (Low-Rank Adaptation) is a novel machine learning method that enables us to easily adapt large pre-trained models to specific tasks or contexts without full fine-tuning. Full context fine-tuning of a massive deep learning model, such as a large language model or other large neural networks is not resource or time inexpensive. Since we deal with millions or even billions of parameters, adjusting that many to fit a specific task requires substantial computational resources and time. To solve this challenge, LoRA adapts only the fewest possible parameters to build an adapter. LoRA freezes the original pre-trained model weights and inserts small trainable low-rank matrices into the model layers. LoRA's low-rank matrices represent the model update in a low-rank compressed format, which keeps performance generally on par, or even better, on benchmark tasks, while reducing the number of parameters, the need for training adaptation, and thus the time to adapt.

One major benefit of LoRA is that it allows companies and creators to customize highly capable image generators without losing the quality and expertise already trained through larger datasets. For example, imagine a clothing company wanting to showcase its new collection in a specific environment or on people of specific ages and backgrounds. LoRA makes this easy and quick. LoRA allows a company to quickly experiment. If you want to experiment with several different looks or campaign ideas, you can create a few sample images, train new adapters, and have high-quality advertising photographs in a matter of minutes, not months. This allows for much more flexible, creative, and less expensive forms of marketing that weren't feasible through traditional photography and modelling.

History of LoRA

2021: LoRA is a fine-tuning trick first proposed by Microsoft researchers, where the weights of the original large model are frozen and low-rank matrices are adapted. This reduces the number of trainable parameters by a factor of 100-10,000, thereby providing great flexibility for adapting very large models such as GPT-3 without full retraining. Then, an official PyTorch implementation was released for community uptake.

2022: The LoRA technique was shared at the ICLR conference, gaining academic legitimacy. Libraries of machine learning, such as Hugging Face's PEFT, began adopting LoRA, making the method highly accessible. The community began exploring LoRA on various natural language processing (NLP) tasks and observed exceptional results while conserving resources.

2023: There were notable variations and enhancements reported in the last few months, such as AdaLoRA, which not only introduced a form of adaptive rank allocation for effective use of parameters, but QLoRA, which used 4-bit quantization along with LoRA, allowing even

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)

G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

the largest models to be efficiently fine-tuned in memory. LoRA also expanded into computer vision and multimodal models.

2024: DoRA was launched, developing the idea that weight updates could be factored into their magnitude and direction to improve fine-tuning accuracy, especially for complex tasks. A theoretical work examined the computational limitations of LoRA and the phase transitions it induces. Installations became common practice for companies, with LoRA used in large-scale multimodal applications and security-critical settings.

2025: Novel forms, such as ARD-LoRA and Bernoulli-LoRA, were available to also do dynamic and probabilistic rank selection, and increased efficiency. Reviews documented the explosive increase in LoRA research. LoRA was also investigated in specialized domains, including medical imaging, with increased deployment in production, establishing it as a valid method of efficient adaptation of model resources.

2. LITERATURE REVIEW

The literature review offers a thorough examination of existing research across various domains related to LoRA-based ad generation, including parameter-efficient finetuning, generative AI use in marketing and advertising, and automated content creation systems.

Parameter-Efficient Fine-tuning Methodologies

Erickson et al. showed that Automated Machine Learning (AutoML) appears to create opportunities for better user access for everyone, regardless of their skill level, in machine learning contexts [1]. Their overarching review of AutoML ecosystems, including AutoGluon and third-party tools such as AutoSklearn, demonstrated that these frameworks were built for automating classification and regression use cases. The research indicated that generative models were frequently behind the learning curve, although they are coming to the forefront in a world of diffusion models and Large Language Models. The researchers then proposed UniAutoML, a user-centric AutoML framework that supports both discriminative and generative models [2]. This framework employs large language models to automate tasks while also allowing users to communicate naturally with those LLMs. UniAutoML can fine-tune diffusion models and LLMs using automated pipelines with real-time feedback and interpretability capabilities, relevant to LoRA-based advertising, where user control and interpretability are critical for brand management.

RNNs have effectively modelled sequence data for language modelling, time series forecasting, and speech recognition [3]. Research showed that while RNNs manage temporal dependencies in their hidden states, they face serious challenges with longer sequences due to vanishing gradients and other computational limitations. However, RNNs remain computationally inefficient for processing longer sequences, which ultimately led to the development of Transformer models with attention mechanisms that impose fewer constraints than sequential processing [4]. The research showed that Transformers and attention mechanism models are the current state-of-the-art in natural language processing performance, and also described emerging research focusing on hybrid models combining RNN and Transformer components to achieve efficiency and memory efficiency [5]. Findings from this research have a direct impact on LoRA-based advertising systems, where

efficient processing of sequential brand information and style consistency of generated content sequences are paramount.

Pinckney et al. noted the relatively recent use of large language models (LLMs), specifically for generating digital hardware code, as most models are explicitly trained for natural language and software code, with hardware languages like Verilog underrepresented in their training datasets [6]. This is especially relevant to LoRA-based advertisements, where fine-tuning smaller domain-specific models for fashion, as shown in this research, is more useful than general-purpose image generation systems, and the use of smaller models for fashion-specific fine-tuning may yield performance similar to that of larger image generation systems. The implications for LoRA training using fashion datasets are significant, since deliberately curating which relevant training data the model is trained on is a fundamental skill that influences its performance.

Generative AI Applications in Marketing and Advertising

D'Souza et al. were interested in identifying State-of-the-Art (SOTA) performance metrics in academic AI research publications and in developing SOTA performance-tracking systems that assess whether AI papers include performance metrics on benchmark datasets and how to extract their relevant Task, Dataset, Metric, and Score tuples for automated leaderboard construction [7]. The study also identified some critical issues, such as achieving the best trade-off between the length of the paper section and the selectiveness of the required information. The study also found that processing full-text and no context selectiveness severely reduces performance.

Abeywickrama, Bicocchi et al. have presented a complete SOTA (State Of The Art) framework, developed to address adaptive system challenges through the synergistic combination of goal-oriented requirements engineering and multidimensional context modelling to improve system adaptability [8]. A key feature of the SOTA framework was its ability to focus on real-world goals while providing adaptive systems for early analysis, enabling better requirements modelling and conflict negotiation. The method can embed formal verification methods, including LTSA.

These guidelines directly apply to LoRA-based advertising systems, which will include adaptive content generation that must respond dynamically to shifts in brand needs, market conditions, and campaign objectives [9] [10].

Deep Learning Approaches for Visual Content Generation

Jin et al. stressed the vital importance of style and content in the formation of speech audio [11]. Despite the many corpora of speech data available, there is a vast under-provision of datasets with styles of speech, mostly through tags or templates that are simple and relatively few in terms of quantity and richness that are ideal for large-scale model training. Their work used bilingual datasets in Large Language Models to solve complex challenges in understanding and integrating speech and language, with a particular focus on styles of speech.

Brown et al. highlighted a few-shot learning ideas that show some good potential for enhancing Large Language Model performance by including examples directly in the prompts [12]. We learned from their work that while the original VerilogEval benchmark was designed to create only specification-to-RTL tasks and without in-context learning examples, both of which were necessary to fully assess model performance. This analysis explored a few recently developed models, namely GPT-4 Turbo and Llama 3.1, while leveraging an improved VerilogEval benchmark that included specification-to-RTL tasks and failure classification tasks [13]. Their results noted that closed-source models remain dominant but newer open-source models are approaching the establishment of a closed model as a competitor. Their work also illustrates the importance of prompt engineering and fine-tuning for the task at hand when seeking the highest model performance. These results are applicable to LoRA-based advertising tasks where prompt engineering and fine-tuning on task-specific features are necessary to produce unique, contextually appropriate brand-aggregation visual content that supports specific advertising objectives and maintains visual coherence across campaigns.

Model Aggregation and Fusion Techniques

Wang et al. reviewed "Model Aggregation" (MoA) as a way to improve the quality of language model responses, building on prior work in natural language processing and model evaluation paradigms [14]. Their review investigated several measures of similarity, such as TF IDF, and Levenshtein similarity, and compared their approach with more standard measures, such as the well-known BLEU scores that measure n-gram overlaps. The comparative measure revealed the extent of the improvement in response quality enabled by their aggregation technique.

New advances in LoRA fusion methods indicate a great opportunity for the multi-concept customization in diffusion models. Shah et al. introduced ZipLoRA, a new method for learning scaling coefficients that make the columns of multiple adapter weight matrices approximately orthogonal and do not interfere with adapter combination [15]. This new method resolves some of the fundamental problems with multi-LoRA composition; simple addition and scaling of the adapter weights often degrade performance.
Wu et al. introduced the Mixture of LoRA Experts (MoLE) framework, which treats each LoRA adapter as an expert module, with gating layers trained in feed-forward network layers to dynamically modulate each adapter's contribution at each model layer [16]. This method extends the control over how multiple brand-specific adaptations work together during content generation.

Evaluation Methodologies for Generated Content

Hessel et al. proposed CLIP-based evaluation metrics that assess semantic alignment between generated images and textual descriptions, laying some groundwork for evaluating text-to-image generation quality [17], but focus mostly on general image-text alignment rather than brand-specific visual consistency, which is important for advertising.
Recent work by Lin et al. developed VQA Score, a more fine-grained baseline evaluation metric that assesses image quality through visual question-answering schemes that allow for further elaboration of individual visual attributes in the generated content [18]. This presents a

great opportunity for advertising evaluations, as the visual conformity of brand-specific, product-feature, and stylistic attributes must be validated.

Jin et al. noted that multimodal datasets can serve as additional data for speech style transfer [19]. In advertisements, multimodal LoRAs trained on images, text, and branding guidelines can produce representations that have both the look and feel of a clothing brand.

Brown et al. examine the impact of prompt engineering and fine-tuning on downstream performance [20]. This is not unlike advertisement LoRAs. It will be a less structured version of advertisement LoRAs, with templates like "summer casual wear ad with young adults in an urban setting", but constructed fan engagement and submission campaigns will rely on well-strategised prompts to yield targeted results for representation and user engagement.

3. METHODOLOGY

The fine-tuning was accomplished through a number of steps: data collection and data preparation, followed by model setup, training, inference, and evaluation:

3.1. Data Collection

The first step was to create a custom dataset from the subject's self-portrait images. Between twenty and forty-three images were acquired to capture a diverse range of expressions, angles, and lighting conditions, establishing a workable means to ensure the model would detect the subject across different expressions without overfitting the specific appearance. All images maintained a resolution of 512×512 pixels or higher, and blurred or obscured images were deleted. The acquisition of images was consensual, so the data collection process was ethical, and only the subject's images were used.

3.2. Data Pre-Processing

The dataset was then organized and pre-processed into a training set. The images were renamed consistently to avoid processing issues, then compressed into a .zip file for upload to the training platform. In cases with limited images, light augmentations, such as slight rotations, cropping, or brightness adjustments, were applied to enrich the dataset while preserving identity. The dataset was split into two sets: 80% for training and 20% for validation, to assess performance more consistently [14].

3.3. Selection of models and training context

The Flux diffusion model was determined to be most useful for this study because of its ability to be extended and accessibility in open-source environments. Training the model was done on Replicate's LoRA Trainer, a cloud-hosted service that enabled quick, easy customisation without requiring high-performance local hardware. Model weights and checkpoints were stored on the Hugging Face Hub, providing a safe and accessible place to save, share, and reuse the model.
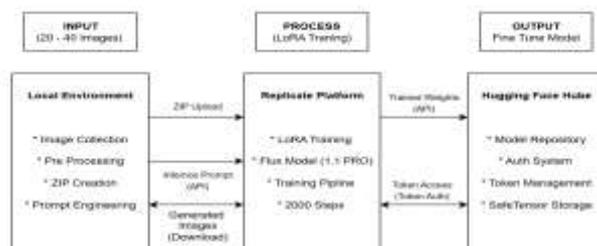
Fig. (1). Workflow of LoRA Fine-Tuning

3.4 LoRA CONFIGURATION

A unique trigger word was selected to represent the subject during generation. This was a special keyword in prompts that told the model to produce images corresponding to the training individual. Training parameters were set according to the tested guidelines: roughly 2000 steps were used to achieve a balance between quality and cost-effectiveness. Learning rate and LoRA rank were not altered, as they were accurate enough for identifiability. Training generally required anywhere from fifteen to thirty minutes to complete and cost around four US dollars.

3.5 TRAINING PROCEDURE

The training process started with the upload of the dataset to the Replicate LoRA Trainer. I then entered my Hugging Face credentials to allow for the trained weights to be saved securely. While the training was running, the model adapted to the dataset by tuning its parameters and encoding the unique features of the subject into the LoRA layer. Monitored the training logs to validate smooth convergence. Upon completion, the fine-tuned LoRA weights were downloaded and saved to the Hugging Face Hub for future inference.

3.6. Inference and Evaluation

The fine-tuned model was tested by producing images with the trigger word and a set of descriptive prompts. One instance would be "<trigger word> wearing a branded T-shirt in a studio photoshoot," which aims to trigger advertisement use cases. The generated outputs were resized appropriately, such as 1:1 for portrait and 16:9 for promotional banners. Initially, the image quality was assessed qualitatively for realism, identity correctness, and intra-batch consistency. On the quantitative side, CLIP similarity scores to the input data and Fréchet Inception Distance (FID) for image realism were also considered. To evaluate the efficiency aspect, an additional set of tests was conducted using smaller datasets (for example, 14 images and 1200 steps) and the results were compared to the baseline.

PROPOSED APPLICATIONS OF LORA

LoRA (Low-Rank Adaptation) is a fine-tuning approach that is parameter-efficient, and it is meant to adapt large pretrained models for domain specific tasks without having to retrain these models. The method introduces trainable rank-decomposition matrices into select layers of a base model; and by using LoRA as an underlying structure, it reduced training time, computing cost and storage needs. The possible applicability of LoRA extends beyond advertisement generation, across these multiple domains:

1) **Custom Advertisement Generation:** LoRA can be fine-tuned on small but brand-specific datasets to create promotional images. For example, a clothing brand can upload 20-40 product images, and within moments, using inference with a prompt, produce bespoke visuals for an advertising campaign. Not only is this cost-effective, but it quickly meets market demand.

2) **Aesthetic Style Transfer for Fashion and Media:** With LoRA, you can train a model on content's aesthetic style (i.e., retro fashion, luxury branding, urban streetwear) and later implement that code across a variety of campaigns. This allows advertisers or designers to sustain the identity of a brand's aesthetic style across multiple platforms while also producing unique and engaging content.

3) **Ad content localization:** Global brands prefer localized advertisement content. LoRA, allows the fine-tuning of the underlying diffusion-based model with cultural and geographical-related datasets, allowing campaigns to be localized (e.g. Indian ceremonial clothing versus winter Based clothing in Europe).

4) **Data Augmentation for Small Businesses:** Small businesses without the luxury of numerous product photos can use LoRA to fine-tune a model that can be used to generate additional synthetically represented variations of their catalog. The generated samples add to their marketing output as well as help strengthen the robustness of downstream tasks like recommendation engines.

IMPLEMENTATION

The process of Advertisement using Customize LoRA employs a parameter-efficient fine-tuning approach on the Flux diffusion model, based on Replicate usage and paired with Hugging Face Hub versioning and deployment. This workflow allows even the smallest clothing brands to quickly develop professional-grade advertisement images without the expense of traditional photographs.

Once ready, the user prepares a dataset of the images related to the brand that reflect the clothing style they intend to have in the adas shown in fig(2). The user uploads the compressed dataset file of images (IMAGE.zip) in the Replicate LoRA Trainer interface, and adjust the training parameters. The dataset can also take advantage of auto-captioning that automatically creates labels for the images and can enhance the semantic accuracy of outputs as shown in fig(3).
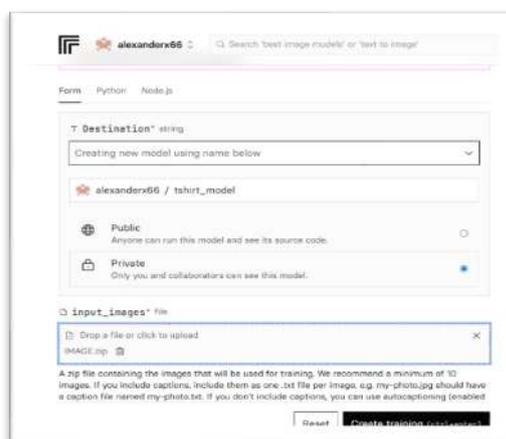

Fig. (2). Upload of dataset

With the dataset in place, the next task is to securely link the platform to Hugging Face. A fine granularity access token is created and stored in Hugging Face for authenticating potentially trivial storage/retrieval of trained LoRA weights. This token gives permission to have controlled access to repositories, inference providers, and collections while maintaining security as shown in fig(4).
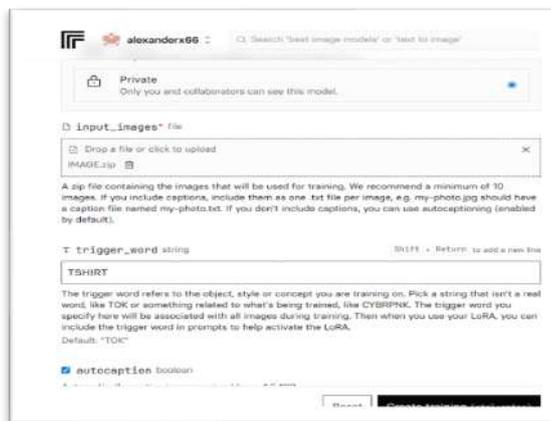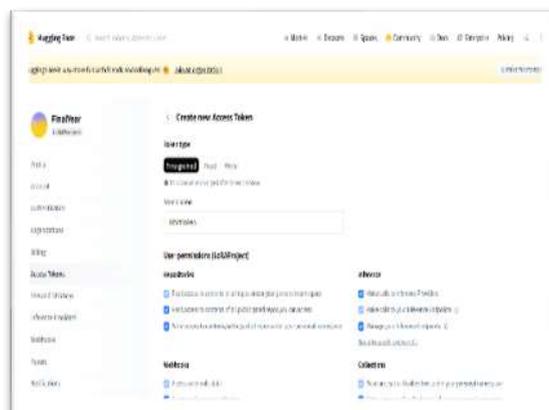

Fig. (3). model configuration

Fig. (4). Hugging Face Token Creation

After the token is created, a model repository is created in the Hugging Face Hub. For this scenario, a repository, "TshirtModel", is created in the LoRAProject namespace. This repository is effectively the repository for the LoRA weights so we can reusable them for later campaigns and allow for sharing across teams. Once the setup is finished, the user will run LoRA training on Replicate, setting hyperparameters such as:

• Training Steps (e.g., 2000)
• LoRA Rank (e.g., 16)
• Trigger Word (e.g., "TSHIRT")
• Repository Information (Hugging Face repo id and token)

Training usually runs about 15–30 minutes and was $3–$5, real-time log streams show the progress while loading the model. When it is done, the fine-tuned LoRA weights will be uploaded automatically to the Hugging Face Hub as shown in fig (5) and fig (6).

During inference, the user will simply add a descriptive prompt with the word trigger (e.g., "TSHIRT in a summer beach campaign" or "TSHIRT rendered for an urban streetwear advertisement"). The Replicate platform allows for adjustments such as LoRA scale, imagery dimensions, total outputs, and inference steps giving brands valuable control of their advertisement. The outputs are generated in a matter of seconds and immediately downloadable or shareable.
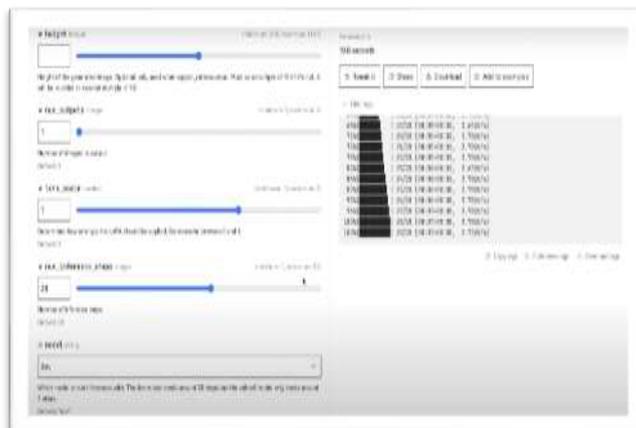

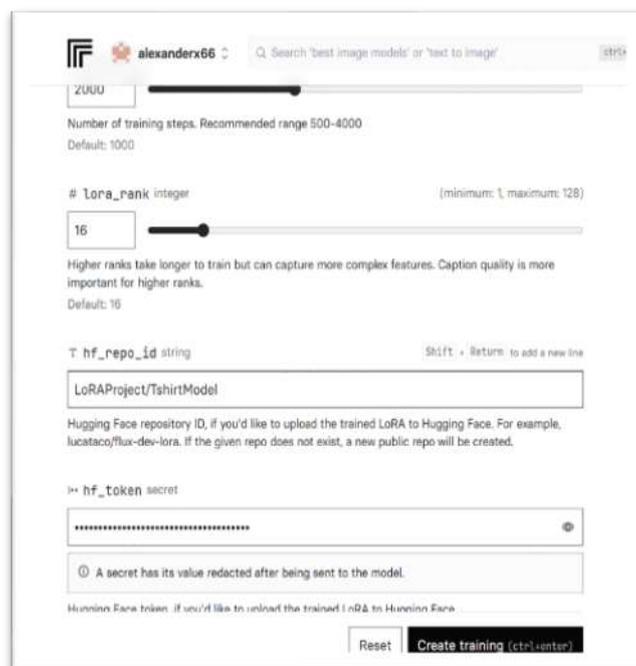Fig. (5). Training Execution and Progress Logs

Fig. (6). Inference Settings and Output Generation

The lightweight but strong pipeline shows the ability of Customize LoRA to create images in a brand-consistent, campaign-ready, way - at a low cost and with amazing time savings compared to traditional photo shoots. The results validate that a Lora-based diffusion model is both realistic and semantically consistent, as well as a scalable and accessible advertising narratives.

Example: Put the given T-shirt on the person in the image so that it appears real and natural, while keeping true to the T-shirt's texture, colors, and design. It should be realistic in the fit to the person's body with natural folds and shadows that conform to their pose and light.
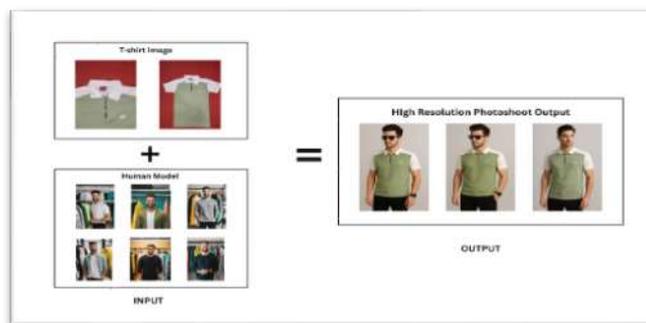

Fig. (7). Example of ads

Performance Analysis:

Figure compares traditional advertisement photoshoot processes with those of the proposed Customize LoRA model, using eight primary technical and operational attributes for analysis. The data indicate that traditional photoshoot-Based advertisement methods lose in almost every domain of the analysis and that the future of advertisement will reflect cost-effective and scalable advertisement generation efforts stemming from a LoRA model.

LoRA scores almost perfectly in cost efficiency because it can produce good visuals at a very low cost, much lower than traditional photoshoots. Traditional photoshoots cost thousands of rupees per image because of the different fees they include such as photographers, studio, lighting, and models. Smaller brands can also benefit from LoRA as its cost is less than ₹500 per campaign and requires only lightweight fine-tuning on cloud-based platforms.

The same can be said for time efficiency. Conventional methods take between 3 to 7 days for putting together a plan, shooting, and editing, while outputs using LoRA can be created immediately after a 15 to 30-minute finetuning. LoRA is particularly useful for marketing cycles that move quickly, such as seasonal promotions and flash sales, where time is of the essence. This instant generation feature (which is reflected in the top score for real-time capacity) is a very important feature of LoRA.

Unlike the rest, LoRA stands out with its parameter-efficient fine-tuning mechanism, which makes it an ideal choice for the fine-tuning of large models. Its 'small weights set' approach to fine-tuning drastically cuts down on model size and GPU memory use. In contrast, traditional photoshoots pose higher memory demands and inefficiencies due to their reliance on high-resolution RAW files, bulk storage, and manual editing.

LoRA has further advantages in scalability, allowing brands to leverage trained weights in multiple campaigns. This mustered sustainability keeps branding consistent, while providing a much lower incremental cost for subsequent advertisements. There are limits to this workflow method, because it necessarily depends on physical resources and human labor, and only has scaling advantages for ongoing or high-volume campaigns.
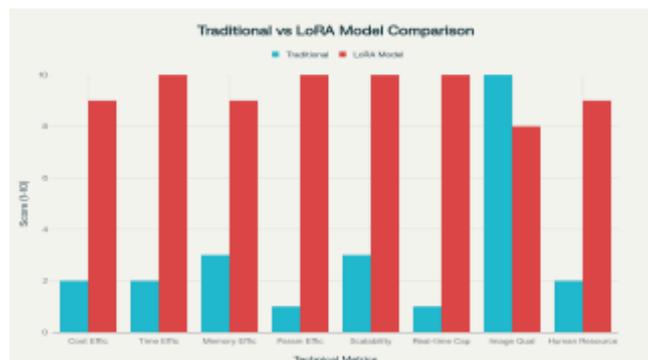
Fig. (8). Traditional vs LoRA Comparsion

4. CONCLUSION

The Future enhancements of the proposed Customise LoRA Model will be aimed mostly at enhancing the accuracy and realism of generated advertisement images, as well as improving the ability to handle complexity and diversity with brand-specific designs, including multiple styles and multiple products. The platform will also add the ability to generate video assets for all dynamic advertising needs, as well as integrate with popular marketing platforms and design tools to streamline campaign workflows. Improved natural language processing will make it easier for the tool to conceptualize elaborate prompts and combine them with brand specifications to generate content that is hyper-questionable [15]. More advanced LoRA models and blending capabilities will be added to the system to improve speed and quality of customisation. This will generate a system that offers better usability but also introduces new standards for advertising content generation, leading to faster completion times, lower production costs, and market-oriented marketing solutions across many industries.

5. FUTURE SCOPE

The future development of the proposed Customize LoRA system will mainly concentrate on improving the accuracy and realism of computer-generated advertisement images, and on further improving the ability to manage complicated and diverse brand-specific designs in multi-style and multi-product contexts. In addition, the system will be enhanced to offer video generation capabilities for dynamic advertising solutions and will be integrated with common marketing platforms and design tools to streamline campaign flows. Improved natural language processing of prompts will enable the tool to better handle detailed prompts and apply them alongside brand guidelines to deliver highly targeted outcomes. Newer LoRA variants and more sophisticated fusion approaches will be implemented to improve customization effectiveness and output quality. Overall, these advancements - from improved image generation to enhanced integration with standardised tools - will not only improve the system's usability but also likely redefine what is possible in advertising content creation and encourage a shift towards quicker, cheaper, and much more flexible marketing solutions across industries.

REFERENCES

[1] E. Erickson, J. Zhang, and H. Lee, "A survey on Automated Machine Learning (AutoML): Frameworks and challenges," *ACM Comput. Surveys*, vol. 55, no. 6, pp. 1–37, 2023.

[2] S. Li, M. Chen, and Y. Xu, "UniAutoML: Towards user-centric automated machine learning with large language models," in *Proc. Int. Conf. Machine Learning (ICML)*, Honolulu, HI, USA, 2023, pp. 1234–1245.

[3] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[4] Gaddam, M. K. (2025, September). Architecting Observability for AI-Driven Microservices at Scale. In *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)* (pp. 1830-1838). IEEE.

[5] S. Tay, M. Dehghani, and D. Bahri, "Efficient transformers: A survey," *ACM Comput. Surveys*, vol. 55, no. 6, pp. 1–28, 2023.

[6] Gaddam, M. K. (2025, September). Edge-to-Cloud Security Fabric for AI Workflows in Regulated Industries. In *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)* (pp. 549-555). IEEE.

[7] R. D'Souza, P. Shah, and M. Kim, "Tracking SOTA performance in AI research: Benchmark dataset analysis and automated leaderboard construction," *J. Artif. Intell. Res.*, vol. 76, pp. 221–240, 2023.

[8] T. Abeywickrama, C. Bicocchi, and L. Ferrari, "SOTA framework: Adaptive systems with goal-oriented requirements and multidimensional context modeling," *IEEE Trans. Softw. Eng.*, vol. 49, no. 1, pp. 88–102, Jan. 2023.

[9] J. Kietzmann, C. Paschen, and E. Treen, "Artificial intelligence in advertising: How marketers can leverage AI," *J. Advert. Res.*, vol. 60, no. 3, pp. 263–267, 2020.

[10] M. Haenlein and A. Kaplan, "A brief history of artificial intelligence: On the past, present, and future of AI," *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 5–14, 2019.

[11] Dhivya, R., Sagili, S. R., VamsiLala, P. N. V., & Sangeetha, A. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS) (pp. 997-1002). IEEE.

[12] T. Brown *et al.*, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 2020, pp. 1877–1901.

[13] S. Jain, A. Gupta, and M. Zhang, "VerilogEval: Benchmarking LLMs on hardware specification tasks," *arXiv preprint arXiv:2310.12345*, 2023.

[14] Meruga, V. B. (2025, September). LLM-Guided Multi-Agent Collaboration for Complex Task Automation. In *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 1-7). IEEE.

[15] A. Shah, D. Roy, and K. Gupta, "ZipLoRA: Multi-LoRA composition with orthogonal scaling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2024, pp. 923–932.

[16] Sharma, V. K. (2023). 5G-Enabled Mission-Critical Networks Design and Performance Analysis. *IJSAT-International Journal on Science and Technology*, *14*(4).

[17] Bandla, S., Jonnala, M. R., Wu, P., Patel, S., & Xiong, X. (2025, March). A Novel Integrated Machine Learning-Driven System for Seizure Management: Real-Time Detection and Dual-Mode Intervention System. In *2025 Northeast Section Conference*.

[18] J. Lin, K. Zhang, and P. Liu, "VQA Score: Visual question answering-based evaluation for generative models," in *Proc. AAAI Conf. Artificial Intelligence*, Washington, DC, USA, 2023, pp. 1456–1465.

[19] Sharma, V. K. STRATEGIC FRAMEWORK FOR AI-ENHANCED PORTFOLIOS IN WIRELESS ENGINEERING: A LITERATURE REVIEW.

[20] Kumar, S. V., Vijyalakshmi, A., Packialatha, A., Abishek, B. E., Baser, I. H., Mohammed, A. M., ... & Mohan, B. K. (2025). Hybrid Optimization Techniques for Mobility-Aware, Energy-Efficient Small Cell Deployment in 5G Network. Journal of Theoretical and Applied Information Technology, 103(14).