PixelForge: AI-Driven Digital Marketing System

FYP– II REPORT

BS(CS/AI) - Fall 2024

Muhammad Shaheer Luqman
21K-4655

Rabeet Tanveer
21K-3374

Ismail Qayyum
21K-3118

Supervisor: Sir Basit Jasani

Co-supervisor: Dr. R/\Fi

Department of Computer Science

FAST-National University of Computer & Emerging Sciences, Karachi

Project Code	F24-105
Supervisor	Dr. Basit Jasani
Co-Supervisor	Dr. Muhammad Rafi
Project Team	Shaheer Luqman(21K-4655) Khawaja Rabeet(21K-3374) Ismail Qayyum(21K-3118)
Submission Date	May 15, 2025

Supervisor

Mr. Basit Jasani

Head of Department

Dr. Ghufran Ahmed

ACKNOWLEDGEMENT:

All the thanks to the supervisor Dr. Basit Jasani and co-supervisor Dr. Mohammad Rafi, the members of the project Ismail Qayyum, Muhammad Shaheer Luqman, Khawaja Rabeet Tanveer, and people who cooperated in providing the data for the project. All the mentioned people have contributed to the project and helped us on this journey. Without their help we could not have done this.

Table of Contents

Abstract....5
Introduction....5
Related Work....6
Requirements....7
    Use Case Diagram....7
    Use Case Tables....8
    Non-Functional Requirements:....14
Design....15
Implementation....18
System Architecture Diagram....23
Conclusion....24
References....25
Similar Existing Apps....26

ABSTRACT

The Generative AI-based Advertisement System is an innovative solution designed to transform the process of digital advertisement creation. Leveraging cutting-edge natural language processing (NLP) and advanced machine learning (ML) algorithms, the system automates the generation of high-quality image posts, captions, and video advertisements tailored to specific products. Unlike existing video generation methods that rely on web scraping and video stitching, our system introduces a novel approach to true product-based video generation, creating advertisements frame-by-frame based on user-provided prompts and product images. The project delivers an all-in-one web application where users can seamlessly generate promotional image advertisements or brief video showcases, paving the way for efficient, scalable, and AI-driven content creation. While initially focused on generating short-duration advertisements, this system has the potential to revolutionize the advertising industry with fully AI-generated campaigns in the future.

INTRODUCTION

The evolution of Generative AI has marked a significant shift in various industries, including digital advertising, where automation and personalization are becoming indispensable. Current video generation systems, while advanced, rely heavily on techniques such as web scraping and video scripting. These methods involve generating a script based on a given prompt, and then searching for and collaging pre-existing video snippets to create an advertisement. While this approach is practical, it cannot create truly unique, product-specific advertisements that are custom-generated frame-by-frame.

This gap highlights the need for a generative system capable of true video synthesis, particularly for advertising, where personalization and creativity drive engagement and ROI. Existing research on video generation technologies, such as GANs, VAEs, and Diffusion Models, provides a solid foundation. However, these systems primarily focus on generating general-purpose videos rather than tailored advertisements.

The Scope of our project encompasses the creation of an all-in-one web application that allows users to submit their product image and prompt and choose between a Promotional Image Advertisement with slogans and captions provided or a generated video showcase of their product. Even though we are currently deploying our solution at a short scale creating only a few seconds worth of advertisements, future implications for this work could see purely AI-generated advertisements.

Related Work

Several web-based platforms currently support image and video generation, including commercial tools like RunwayML and Sora, which offer advanced video synthesis. However, these systems are generally built as standalone video generators, with limited integration of input images and no specific focus on advertisement content.

Other platforms primarily rely on text prompts or web scraping, lacking product-awareness and producing generic outputs. Most fail to provide an end-to-end solution that combines image-based inputs, dynamic captioning, and branded content generation.

In contrast, our system offers image and prompt-based video generation, image and prompt-based post generation, LLM-enhanced captioning capabilities all packed in a user-centric web app with little to no admin intervention, essentially providing an all in one package for users looking to create advertisements for their products.

Requirements

Functional requirements and the diagrams are given below:

Use Case Diagram

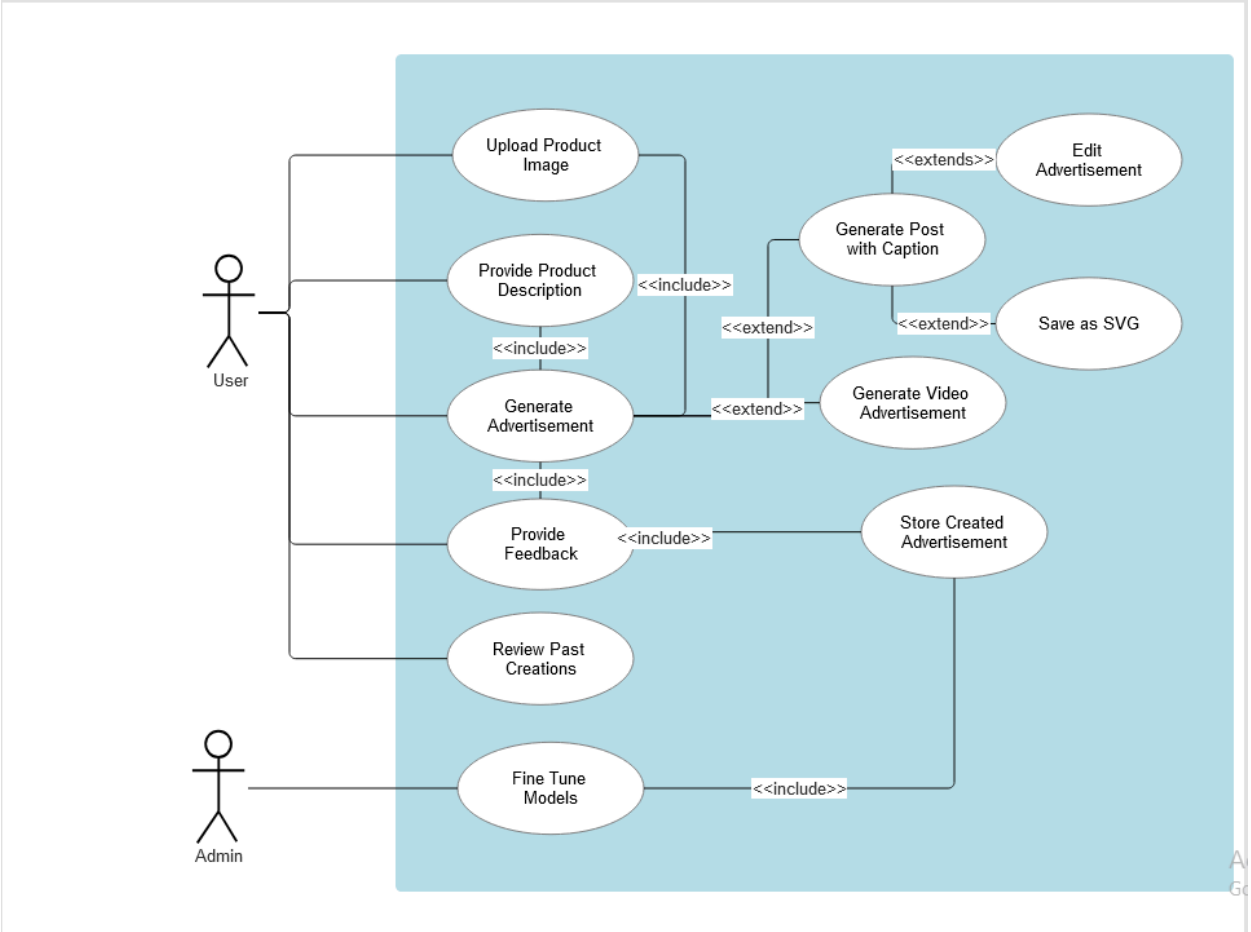


Figure 1: Use Case Diagram

Use Case Tables

<UC-USER-01>		
Use case Id:	Write a use case reference number.	
Actors: User.		
Feature: Upload Product Image		
Pre-condition:	Users must have a valid account in the system. Users must log in with the correct email and password. The system must have user info already stored.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User interacts with the upload button.	Displays upload image scenes.
2.	User selects Image to upload.	Loads Image for confirmation.
3.	User confirms.	Uploads Image to application.
Alternate Scenarios: None.		
N/A		
Post Conditions		
Step#	**Description**	
	User successfully uploads an image.	
Use Case Cross referenced	**UC-USER-03**: Relationship: Includes. Cannot generate Advertisement without product image.	

<UC-USER-02>		
Use case Id:	Write a use case reference number.	
Actors: User.		
Feature: Provide Product Description		
Pre-condition:	Users must have a valid account in the system. Users must log in with the correct email and password. The system must have user info already stored.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User interacts with the Description Textbox.	Prepares to write keyboard output to Textbox.
2.	User provides Description and submits	Uploads Product Description to application.
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
	User successfully submits description.	
Use Case Cross referenced	**UC-USER-03**: Relationship: Includes. Cannot generate Advertisement without product description.	

<UC-USER-03>		
Use case Id:	Write a use case reference number.	
Actors: USER.		
Feature: Generate Advertisement		
Pre-condition:	Users must have a valid account in the system. Users must log in with the correct email and password. The system must have user info already stored. User must have entered Product Description User must have uploaded Product Image	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User chooses between Post and Video Generation	Displays relevant models.
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
1.	Relevant Models Generation is started.	
Use Case Cross referenced	**UC-USER-01**: Relationship: Includes. Cannot generate Advertisement without product image. **UC-USER-02**: Relationship: Includes. Cannot generate Advertisement without product description. **UC-SYSTEM-01**: Relationship: Extends. Users can choose Post Generation. **UC-SYSTEM-02**: Relationship: Extends. Users can choose Video Generation.	

<UC-USER-04>		
Use case Id:	Write a use case reference number.	
Actors: **USER**.		
Feature: Provide Feedback		
Pre-condition:	Users must have a valid account in the system. Users must log in with the correct email and password. The system must have user info already stored. Users must have Generated Advertisement.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	Users enter feedback on a rating of 1 to 5.	Stores feedback rating locally.
Alternate Scenarios: User Chooses not to enter feedback.		
1a: If feedback is not entered, the created advertisement is stored given a rating of 2.5.		
Post Conditions		
Step#	**Description**	
	Users created advertisements and ratings as stored together to be used for fine tuning.	
Use Case Cross referenced	**UC-USER-03**: Relationship: Includes. Cannot give feedback unless Advertisement is generated. **UC-SYSTEM-03**: Relationship: Includes. Cannot Store Advertisement unless rating is given as feedback.	

<UC-USER-05>		
Use case Id:	Write a use case reference number.	
Actors: **USER**		
Feature:	Review Past Creations	
Pre-condition:	Users must have a valid account in the system. Users must log in with the correct email and password. The system must have user info already stored. Users must have Generated at least one Advertisement.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User Selects History Tab	System displays a History Page.
2.	Users can choose any past advertisement for feedback or download.	Open specific advertisements interacted with.
Alternate Scenarios: NONE		
N/A		
Post Conditions		
Step#	**Description**	
	NA	
Use Case Cross referenced	**UC-USER-03**: Relationship: Includes. Cannot view past Advertisement unless Advertisement is generated.	

<UC-Admin-01>		
Use case Id:	Write a use case reference number.	
Actors: **Admin**		
Feature:	Fine Tune Model	
Pre-condition:	Admin Access Device must be used.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	Admin uses Rating and new Advertisement Pairs to Finetune Models.	
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
	Model Accuracy and Results increase	
Use Case Cross referenced	**UC-SYSTEM-03**: Relationship: Includes. Cannot finetune model unless rating and advertisement are present in database.	

<UC-SYSTEM-01>		
Use case Id:	Write a use case reference number.	
Actors: **System**.		
Feature:	Generate Post Advertisement with caption	
Pre-condition:	User must have Generated Advertisement	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	.	Separates Product from Product Image.
2.		Extracts Color Scheme from Product.
3.		Generates Background to compliment Product.
4.		Generates concise captions for Post.
5.		Uses Gemini API to turn caption into Slogan
6.		Overlays Slogan and gives a separate descriptive Caption.
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
	Final Post is generated.	
	Can be saved or edited.	
Use Case Cross referenced	**UC-USER-03**: Relationship: Extends. Users can choose Post Generation. **UC-SYSTEM-04**: Relationship: Extends. Users can choose to Edit Post. **UC-SYSTEM-05**: Relationship: Extends. Users can choose to Save Post as SVG.	

<UC-System-02>		
Use case Id:	Write a use case reference number.	
Actors: **System**		
Feature:	Video Generation	
Pre-condition:	User must have Generated Advertisement	
Scenarios		
Step#	**Action**	**Software Reaction**
1.		Generate suitable script based on uploaded image
2.		Pass image and script(s) to RunwayML Gen1.
3.		Enhance Final product video
Alternate Scenarios: none		
N/A		
Post Conditions		
Step#	**Description**	
TBD	RBD	
Use Case Cross referenced	**UC-USER-03**: Relationship: Extends. Users can choose Video Generation.	

<UC-System-03>		
Use case Id:	Write a use case reference number.	
Actors: System		
Feature:	**Store Created Advertisement**	
Pre-condition:	User must have Provided Feedback Rating	
Scenarios		
Step#	**Action**	**Software Reaction**
1.		checks rating score.
2.		if > 5, generates pair variable
3.		creates pair of rating and advertisement
4.		Stores pair to database
Alternate Scenarios: if rating <5, created advertisement is discarded as to not worsen the model.		
Post Conditions		
Step#	**Description**	
	Entry is made into the database.	
	.	
Use Case Cross referenced	**UC-USER-04**: Relationship: Include. Cannot store advertisements unless rating is provided. **UC-ADMIN-01**: Relationship: Include. Cannot fine tune model unless advertisements are stored.	

<UC-SYSTEM-04>		
Use case Id:	Write a use case reference number.	
Actors:	**SYSTEM**	
Feature:	Edit Advertisement	
Pre-condition:	User must have generated a Post with Caption.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User chooses to edit the Advertisement.	Saves Advertisement as SVG to downloads folder. Redirects user to Photoshop editor
2.	User confirms redirection	Opens Online Photoshop editor and loads up SVG.
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
	Manual Editing can be done by User.	
Use Case Cross referenced	**UC-SYSTEM-01**: Relationship: Extends. Users can choose to Edit Advertisements.	

<UC-SYSTEM-05>		
Use case Id:	Write a use case reference number.	
Actors: **SYSTEM**		
Feature:	Save as SVG	
Pre-condition:	Users must have generated Post Advertisements.	
Scenarios		
Step#	**Action**	**Software Reaction**
1.	User selects save as SVG	System saves post to downloads folder on users device.
Alternate Scenarios: None		
N/A		
Post Conditions		
Step#	**Description**	
Use Case Cross referenced	**UC-SYSTEM-01**: Relationship: Extends. Users can choose to Save Advertisements as SVG.	

Non-Functional Requirements

Performance Requirements

- The system should handle up to more than 1 concurrent user without noticeable lag in processing or response time.
- CRUD operations (Create, Read, Update, Delete) on the database should execute within 1 second on average.
- The AI evaluation process should produce results for an employee within 3 minutes after data submission for image generation.
- The AI evaluation process should produce results for an employee within 2 hours after data submission for video generation.
- The system must maintain 99.9% uptime to ensure reliability for critical operations like ongoing generation.

Safety Requirements

- The system should prevent unauthorized deletion or modification of critical data (e.g., employee records, research papers).
- A backup mechanism must be implemented to automatically store data snapshots daily, ensuring recovery in case of system failure.
- Error handling must prevent the system from crashing under invalid inputs or unexpected actions by users.
- Ensure compliance with workplace safety standards for data handling, avoiding harm caused by incorrect evaluations or statistics.

Security Requirements

- The system must encrypt sensitive data (e.g., user credentials, employee performance scores) using **AES-256 encryption**.
- Admins and Users should have role-based access control (RBAC), ensuring they only access authorized features.
- Employee data privacy will be prioritized.

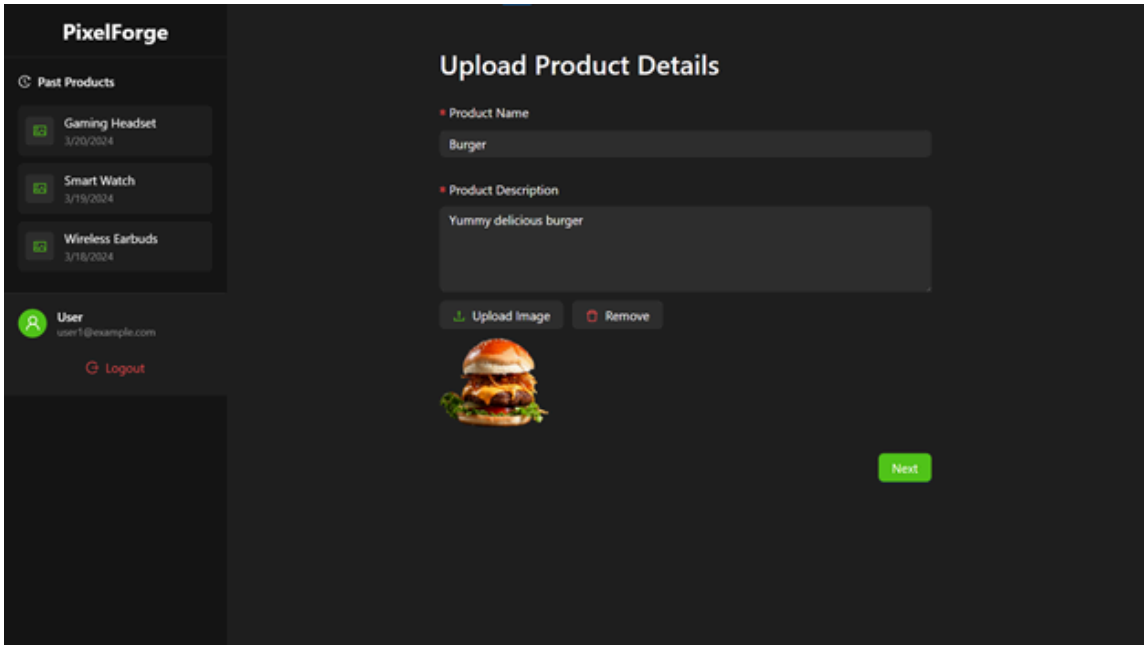
User Documentation

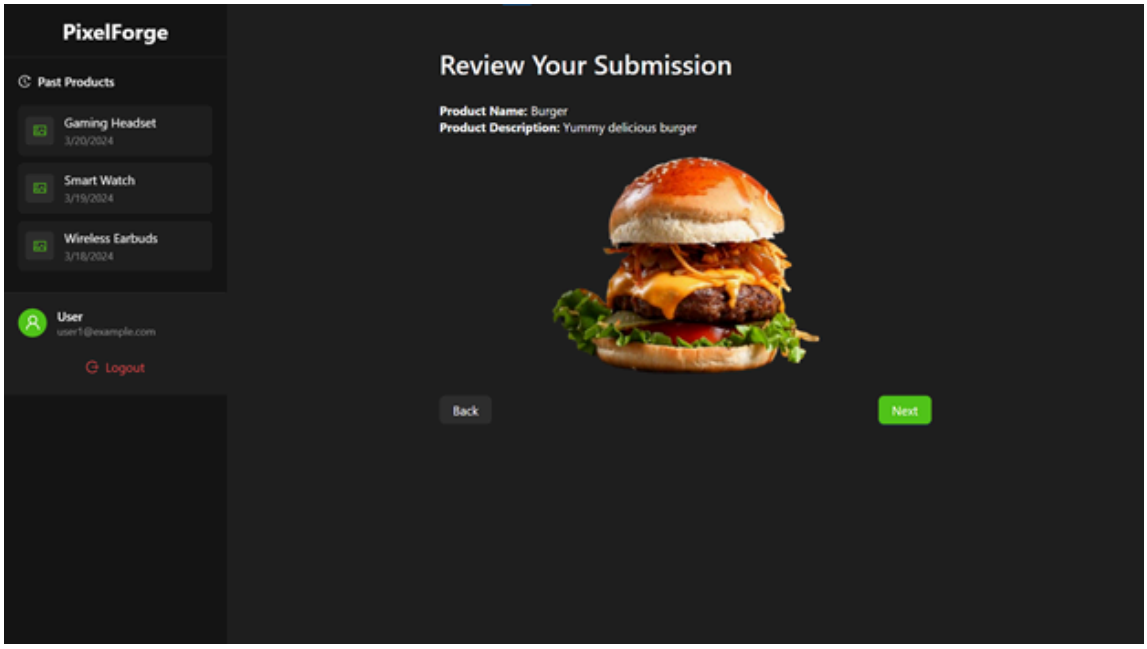
- **Online Help** integrated into the system will provide step-by-step instructions for using key features, accessible via home page.
- **Tutorial Videos** will be available for complex operations, such as configuring AI weights or writing promotion criteria on the blockchain.

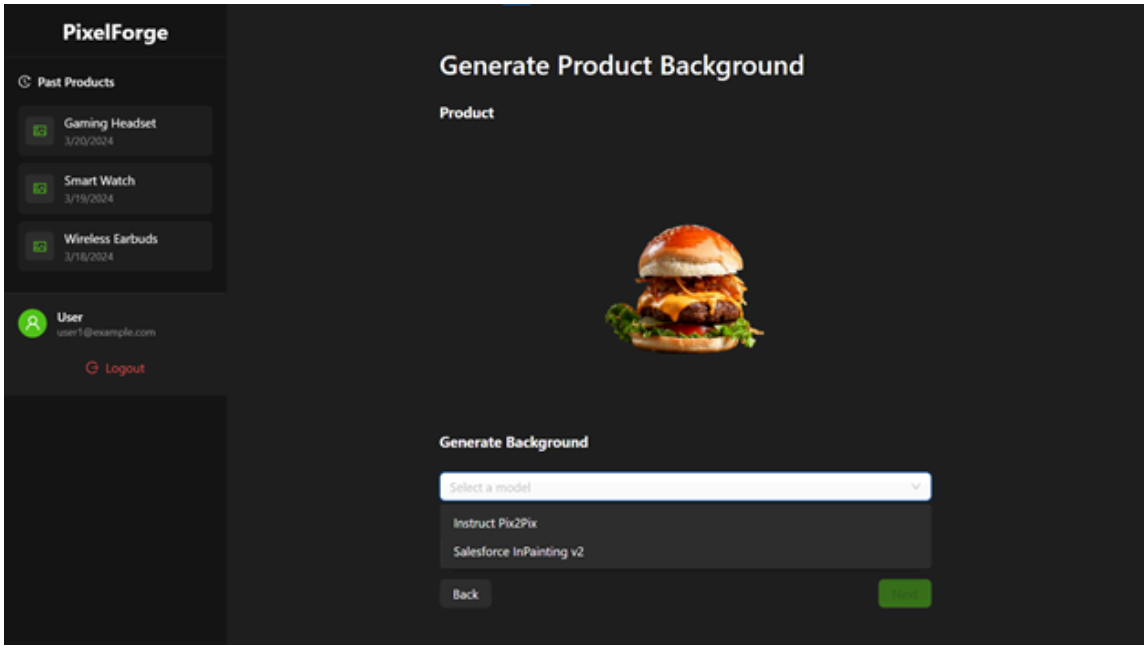
Design

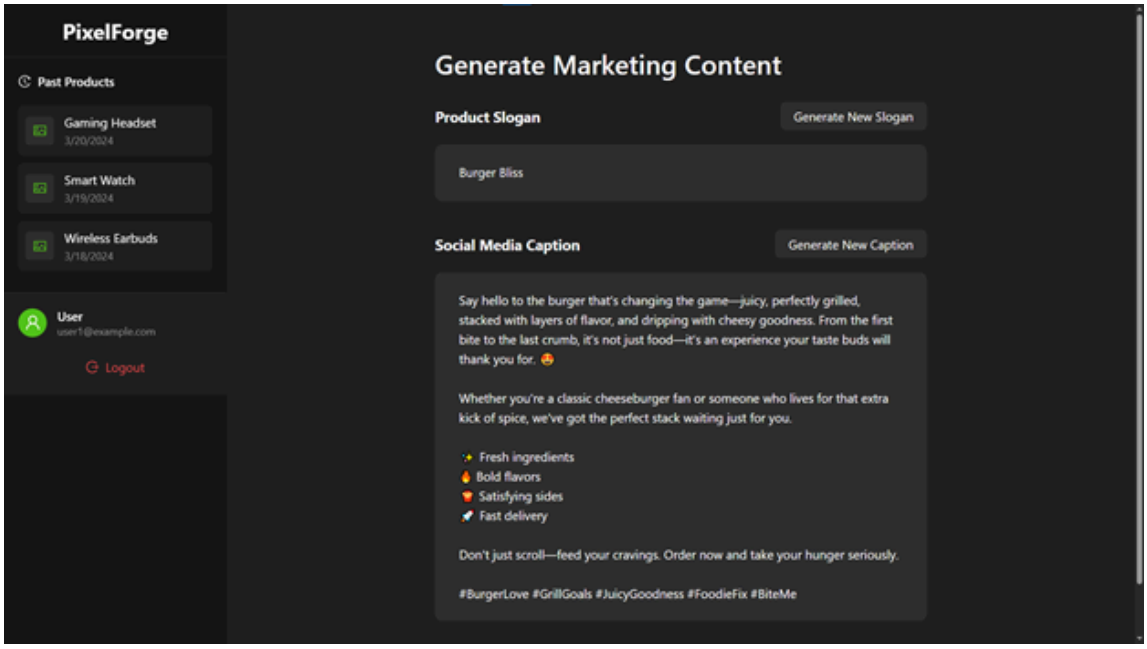
Our system is specifically designed to guide the user as he creates his advertisements with clearly labelled models and short descriptions regarding their use. Tutorial Videos regarding the main functionalities have also been provided. The goal is to create an intuitive, interactive, and user-friendly platform that simplifies the process for stakeholders. We have ensured seamless database connectivity to handle user data cleanly and securely.

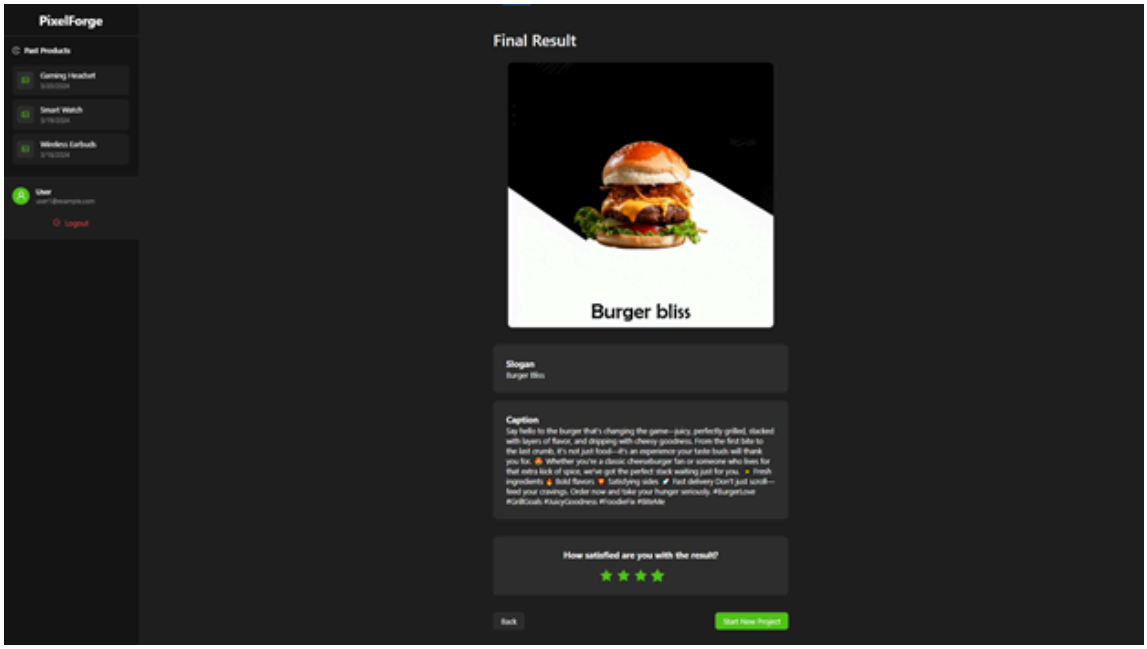
The user interface (UI) and user experience (UX) design are carefully crafted to deliver an aesthetically appealing and consistent layout, making the platform easy to navigate and engaging for all users. Input validation mechanisms and real-time feedback have been implemented to ensure data accuracy and enhance usability. Standard UX that is in line with most AI Model web apps is in place to keep a sense of familiarity with the clients. A step by step process of generation has been provided below.




PixelForge
Past Products
Gaming Headset
3/20/2024
Smart Watch
3/19/2024
Wireless Earbuds
3/18/2024
User
user1@example.com
Logout
Review Your Submission
Product Name: Burger
Product Description: Yummy delicious burger
Back
Next


PixelForge
Past Products
Gaming Headset
3/20/2024
Smart Watch
3/19/2024
Wireless Earbuds
3/18/2024
User
user1@example.com
Logout
Generate Product Background
Product
Generate Background
Select a model
Instruct Pix2Pix
Salesforce InPainting v2
Back


PixelForge
Past Products
Gaming Headset
3/20/2024
Smart Watch
3/19/2024
Wireless Earbuds
3/18/2024
User
user1@example.com
Logout
Generate Marketing Content
Product Slogan
Generate New Slogan
Burger Bliss
Social Media Caption
Generate New Caption
Say hello to the burger that's changing the game—juicy, perfectly grilled, stacked with layers of flavor, and dripping with cheesy goodness. From the first bite to the last crumb, it's not just food—it's an experience your taste buds will thank you for. 😍
Whether you're a classic cheeseburger fan or someone who lives for that extra kick of spice, we've got the perfect stack waiting just for you.
✨ Fresh ingredients
🔥 Bold flavors
🍟 Satisfying sides
🚀 Fast delivery
Don't just scroll—feed your cravings. Order now and take your hunger seriously.
#BurgerLove #GrillGoals #JuicyGoodness #FoodieFix #BiteMe


PixelForge
Final Result
Burger bliss
Slogan
Burger Bliss
Caption
How satisfied are you with the result?
Back
Start New Project

IMPLEMENTATION

PixelForge's web application is compatible with any operating system that supports internet connectivity. The technology stack for the application includes Flask, React, and PostGres SQL, enabling seamless interaction between the layers. Since our work mostly has to do with the Models being used, their separate workings have been presented below.

Captioning Model (SalesForce/BLIP)

Despite giving good captions in general, the BLIP model would often produce outliers that were completely wrong.

Fig 1. Wrong Captioning.

Even when it did work, the model produced captions that were far too small and non-descriptive.

Fig 2. Accurate but Incorrect Caption.

After fine tuning the model over a more precise dataset, the results improved quite a bit but were still not quite what was expected.

Fig 3. Correct Caption but no Uniqueness or Flair.

As a result of this dissatisfaction, we turned to other models where InternVL v2 was a promising choice however the resource requirements for this model at minimum were a GPU having 24GB of VRAM. Not having such equipment readily on hand, we used an alternative route wherein we used an API for Gemini. We would provide Gemini with the caption BLIP created and have it generate a slogan and a descriptive caption that would more likely be found under posts.

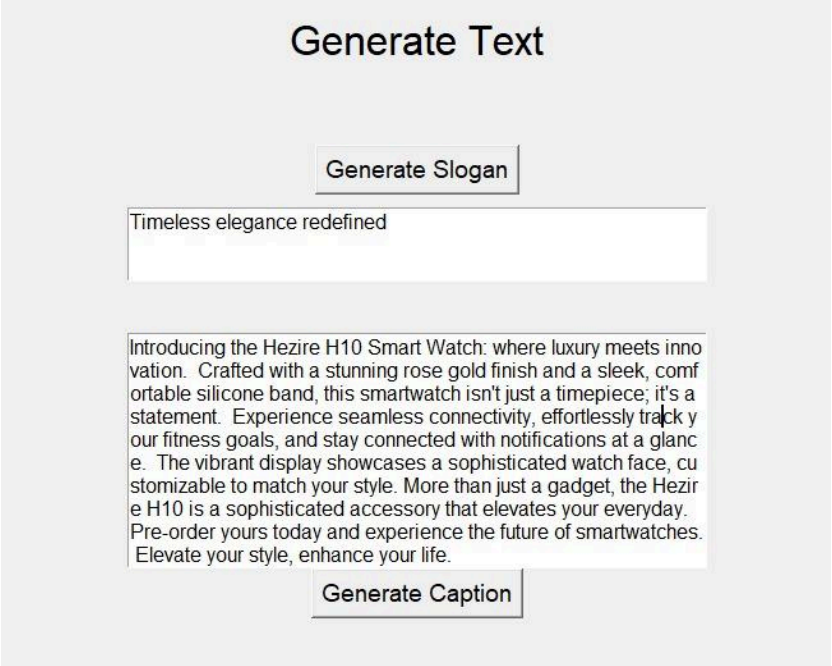


Fig 4. A completed Slogan and Descriptive Caption.

Post Generation Model (IP2P and SIPv2)

A similar line of results and finetuning was observed for the imaging models. The figures below will allow for an easier understanding.

Fig 1 and 2. Initial product Images that will be passed.

Fig 3. A faulty Image Generated by I2P2 model

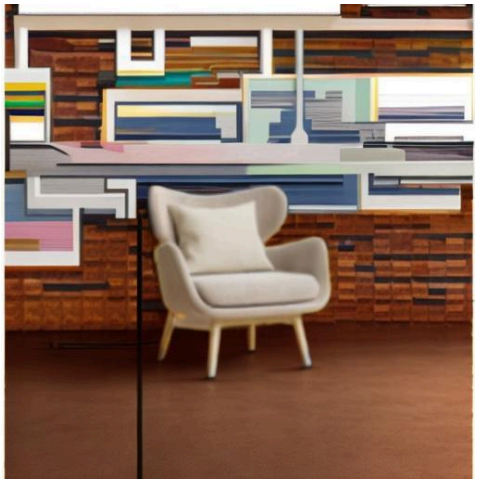


Fig 4. A faulty Image generated SIP v2.

As can be seen, IP2P ended up distorting the product beyond while SIP v2 while starting off with a good background ended up distorting the image towards the end.
After rounds of fine-tuning, and parameter altering, the following results were obtained.

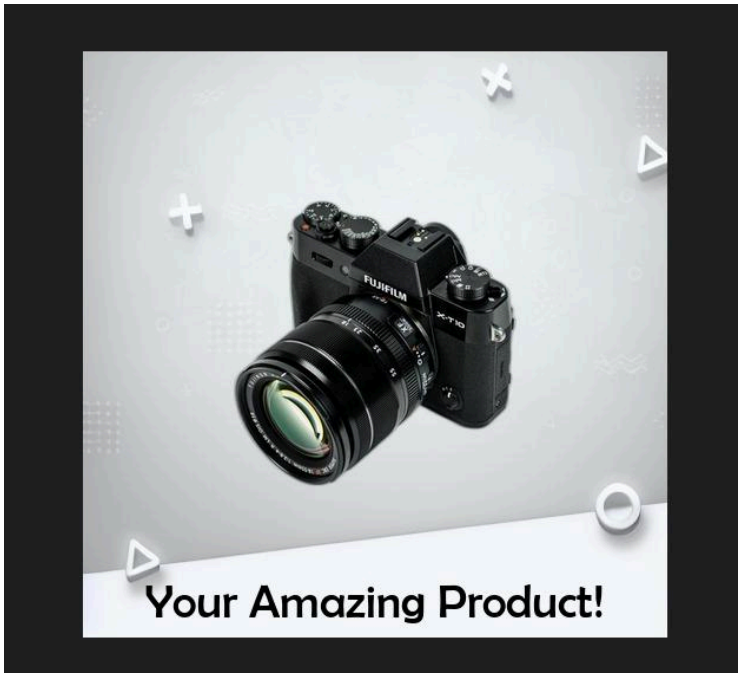


Fig 5. A Post generated by IP2P with a Slogan(generated by BLIP)

Fig 6. A Post Generated by SIP v2.

Video Generation Model (RunwayML Gen 1)

Regarding video generation, we had anticipated that normal GPUs, even the ones being offered by our university might fall short. These expectations were met almost immediately since mere inference of most models gave us an out of memory issue. As such for the sake of testing, we bought one of the commercially available GPUs on google collab for our purposes of inference and fine tuning. We soon came to the realization that while inference could be achieved, fine tuning was wholly out of the question. Even the commercial GPU we were using was a severely limited, and after browsing the forums and communities, we came to the conclusion that the memory required for fine tuning needed multiple costly commercial GPUs. As such we have decided to move forward with using RunwayMLs Gen 1 model as it is.

During our testing of the model, we came to the realization that the model was heavily dependent. A wrong or under-explained prompt leads to the generation of extremely distorted advertisements. As such we decided to let the prompt generation be done by a portion of our captioning model, which would fine tune to generate descriptive prompts regarding the image rather than captions. This approach takes the load off the user to provide hyper specific prompts, and allows them to upload only an image and allow our system to handle the rest. For obvious reasons, the example videos themselves cannot be attached here. but a few sample frames have been shown below.

Fig 7. A distorted frame generated due to an under explained prompt.

System Architecture Diagram

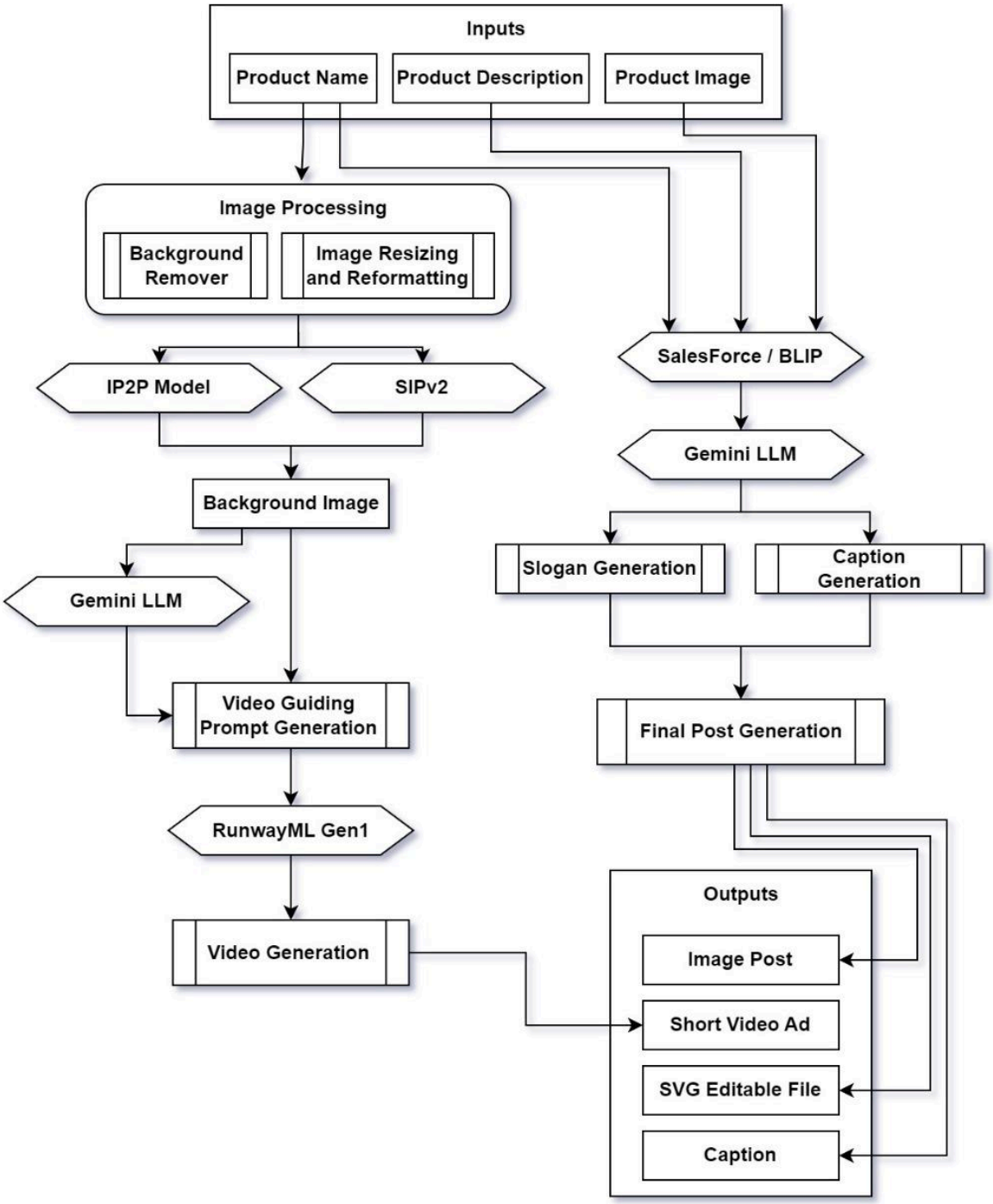


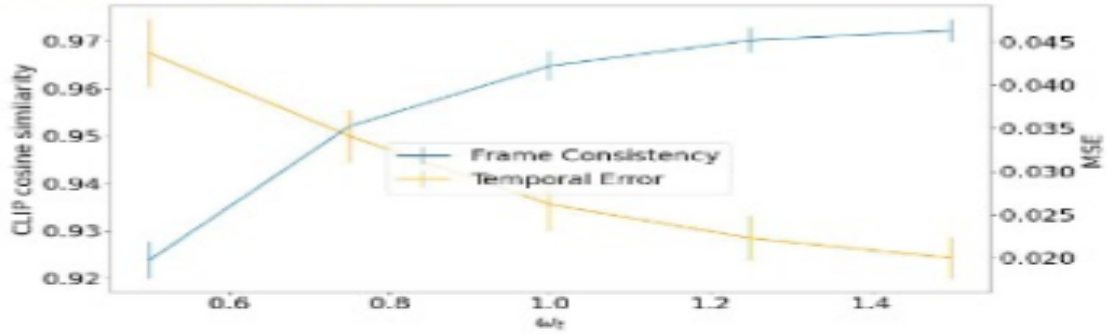
Figure 9: System architecture Diagram

RESULTS

The results of the Image and Caption Generation Models are displayed alongside the implementation section since they are easy to display and understand. Regarding the video model however, since whole videos cannot be provided within the document to properly explain the results, a variety of graphs and figures have been explained below to fully explain the evaluation metric of the RunwayML Gen 1 Model.

Temporal Consistency

In simple words, it refers to smoothness between the changing frames in the generated Video. The graph below will further explain this metric.



Frame Consistency refers to how similar 2 adjacent frames are in terms of context. For this purpose, CLIP Cosine Similarity is used to compute the Frame Consistency score at different values of the **temporal guidance scale (ω□)**. Temporal Error tracks how an individual pixel changes from one frame to the next. Lower Values mean smooth transitions. The Images below are frame by frame comparisons of a task assigned to the model with different **(ω□)** values. The initial row is the input, with the prompt “pencil sketch this man looking at the camera”, the second row has a low **ω□** value (0.5) while the last row has a higer **ω□** value (1.5). The difference is smoothness and consistency is obvious.

Due to hardware limitations on our end, inferencing with high values of **ω□** is too time consuming and computationally expensive. We have settled for a **ω□** value of 1.0. With this setup, most of our generated videos are consistent and smooth but some do break if a correct prompt isnt provided.

Future Work

Despite current hardware limitations making full fine-tuning of video models infeasible (e.g., >50GB VRAM requirements), future work can explore creative alternatives such as smarter prompt engineering, modular adapter-based models, or lightweight fine-tuning techniques as they become more accessible. Additionally, as more and more open source alternatives hit the public space, there are bound to be models with lower hardware requirements that may be easier to finetune and at some point in the future even train from scratch. Furthernore, building tools to refine LLM-prompted video generation and integrate user feedback loops can significantly enhance quality without retraining models. As open-source communities and cloud resources evolve, this system can scale toward more autonomous, domain-specific ad generation pipelines.

Conclusion

We are proud to conclude that PixelForge is a service designed to revolutionize the advertising market by providing its users an all in 1 platform that allows them to create all manner of advertisements. The platform offers a user-friendly interface that simplifies content generation while ensuring data privacy and security. Although we initially anticipated the need for a commercial GPU for fine-tuning the video model, the actual computational demands far surpassed our available resources. As a result, we adapted by leveraging the model solely for inference. Nonetheless, the project stands as a robust, scalable foundation for automated content creation

References

1 P. Esser, J. Chiu, P. Atighehchian, J. Granskog, and A. Germanidis, "Structure and Content-Guided Video Synthesis with Diffusion Models," *arXiv preprint arXiv:2302.03011*, Feb. 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.03011

2 *A. E. Eshratifar, J. V. B. Soares, K. Thadani, S. Mishra, M. Kuznetsov, Y.-N. Ku, and P. de Juan, "Salient Object-Aware Background Generation using Text-Guided Diffusion Models," arXiv preprint arXiv:2404.10157, 2024. [Online]. Available: https://arxiv.org/abs/2404.10157*

3 *J. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," arXiv preprint arXiv:2201.12086, 2022. [Online]. Available: https://arxiv.org/abs/2201.12086*

4 [2] T. Brooks, A. Holynski, and A. A. Efros, "InstructPix2Pix: Learning to Follow Image Editing Instructions," *arXiv preprint arXiv:2211.09800*, Jan. 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2211.09800

5 W. Yan, Y. Zhang, P. Abbeel, and A. Srinivas, “VideoGPT: Video Generation using VQ-VAE and Transformers,” *arXiv preprint arXiv:2104.10157*, 2021.

6 I. Skorokhodov, S. Tulyakov, and M. Elhoseiny, “StyleGAN-V: A Continuous Video Generator with the Price, Image Quality and Perks of StyleGAN2,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 362-371.

7 M. Chen, S. Mei, J. Fan, and M. Wang, "An Overview of Diffusion Models: Applications, Guided Generation, Statistical Rates and Optimization," *arXiv*, vol. abs/2404.07771, Jun. 2024. [Online]. Available: https://arxiv.org/abs/2404.07771.

8 *A. Dosovitskiy, P. Fischer, E. Ilg, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2211.09800, Nov. 2022. [Online]. Available: https://arxiv.org/abs/2211.09800*

Similar Existing Apps

- RunwayML
- Sora
- Web Scraping applications.