VIETNAM DATATHON 2023

AI VIDEO GENERATOR PROPOSAL

Prepared By

BÙI THỊ THANH NGÂN

ĐINH CÔNG HUY HOÀNG

TÔ GIA BẢO

ĐỖ NGUYÊN

I. Introduction

In the competitive landscape of contemporary commerce, the creation of compelling and appealing advertising materials plays a pivotal role in the marketing strategies of retail enterprises. However, the constraints of time and the requirement for specialized design skills often impede the timely production of high-quality advertising assets.

To address this challenge, we propose the deployment of an AI-powered video creation tool in an open world setting. This tool will provide customizable templates, diverse image options, and text variations, enabling users to easily generate professional videos without the necessity of deep design expertise.

The objective of this tool is to enhance the video creation process by simplifying design workflows and bolstering consistent and rapid production. This will aid the retail sector in producing consistently engaging advertising materials, thereby amplifying customer interaction and potentially increasing conversion rates.

With a focus on technological foundations and AI, the Video AI Generator MVP will serve as a critical tool, enabling retail businesses to swiftly and effectively produce high-quality advertising materials, while providing an opportunity to optimize their marketing strategies.

II. Problem Statement

According to research, 64% of consumers are more likely to purchase a product online after watching a video about it. A product video can be a powerful tool to increase your conversion rate, as it can help demonstrate the features, benefits, and advantages of a product in a more engaging way than text and images alone.

To be able to take advantage of valuable information sources from products, there exists a need to transform raw datasets containing product information from renowned brands into captivating advertising videos.

The problem at hand is to devise an efficient and innovative solution that automates the transformation of this raw data into visually engaging videos tailored to capture the attention of potential customers.

III. Solution Overview

Our AI-based solution for video promotion is a cutting-edge approach that leverages advanced techniques to revolutionize how products are showcased and marketed. In this solution, we've seamlessly integrated AI algorithms and models to create compelling video promotions, ensuring a harmonious blend of technical innovation and business functionality.

The AI-powered video promotion solution leverages various AI techniques to enhance the creation, and effectiveness of promotional content. Here's a detailed explanation of how these techniques are applied:

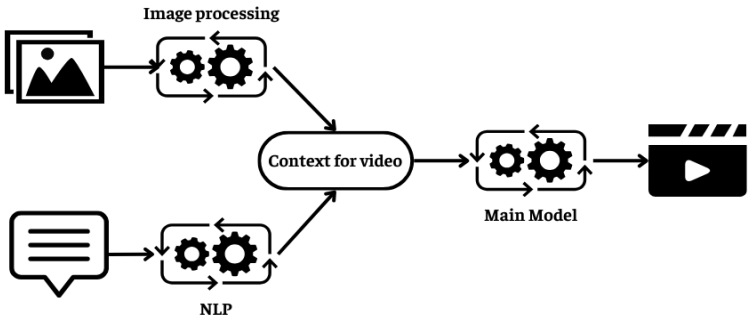
- AI Image Recognition: utilizing computer vision allows AI algorithms to analyze images of the product. This technique helps identify and understand visual elements, such as product features, colors, and shapes.

- Natural Language Processing: NLP techniques are employed to analyze textual descriptions of the product. By understanding the language used to describe the item, the solution extracts key information and sentiments, ensuring a nuanced and comprehensive representation. Analyzing the context of the text contributes to enhancing the relevance and personalization of the video.
- Dynamic Video Editing: AI-driven dynamic video editing ensures that the promotional content is customized based on the features and attributes identified through image recognition and NLP. The solution dynamically adjusts video elements to highlight specific product characteristics, creating a more engaging and tailored promotional experience.

Enhancement and advancement:

- The incorporation of AI into the proposed solution is anticipated to diminish workloads while concurrently augmenting operational efficiency in the pursuit of marketing endeavors.
- Develop a comprehensive marketing strategy, combining both traditional and digital approaches. Utilize creative advertising strategies to attract attention and create a strong impression in the minds of customers.
- The user-friendly technology not only significantly reduces both time and resources compared to manual labor but also ensures the security of data as it is entirely processed automatically.

IV. Methodologies



1. Key Components:

a. Image processing:

- Goal: Leveraging state-of-the-art image recognition algorithms for meticulous analysis empowers content generation for various inputs. For instance, we can create a captivating background for a video by using a flower image.
- Model: YOLOv8 pre-trained model to detect patterns and analyze images can be a powerful tool for generating creative content. This model needs to undergo retraining to align with specific requirements.

b. Text processing (NLP / ChatGPT):

- Goal: Integrate NLP techniques to analyze and extract meaningful insights from textual content. This can include sentiment analysis, keyword extraction, and comprehension of the context in which the text is presented.
- Model: The utilization of GPT-3.5 for the analysis and processing of meaning constitutes the generation of inputs for video content. The application of this process is facilitated through the utilization of its available API.

c. Video creator (Main model):

- Goal: Utilizing AI-driven dynamic video editing to adapt the video content in real-time. This involves adjusting pacing, transitions, and visual effects based on the features identified through image processing and NLP.
- Model: The main generation diffusion model consists of three sub-networks: text feature extraction model (extracted once again), text feature-to-video latent space diffusion model, and video latent space to video visual space model. The overall model parameters are about 1.7 billion. Currently, it only supports English input. The diffusion model adopts a UNet3D structure, and implements video generation through the iterative denoising process from the pure Gaussian noise video.

2. Processing flow:

a. Input:

- Start with input data, which includes textual descriptions and images related to the product or content you want to promote through the video.

b. Image Processing (YOLOv8):

- Feed images through YOLOv8 for object detection and pattern analysis.
- Output: Extracted visual features, patterns, and identified objects.

c. Text Processing :

- Process textual descriptions using GPT-3.5 for sentiment analysis, keyword extraction, and contextual comprehension.
- Output: Extracted sentiments, keywords, and contextual information.

d. Integration of Visual and Textual Insights

- Merge the outputs from Image Processing and Text Processing to create a unified set of visual and textual insights.

e. Main Generation Model:

- The processed video undergoes a series of iterations, incorporating visual patterns, textual insights, and dynamic adjustments.
- The final output is a customized, engaging, and visually appealing video promotion.

V. Core Functionality

1. Automated Banner Creation:

- **Functionality:** Automatically generate basic video banners based on the provided dataset. This includes integrating text, images, and basic design elements to create visually appealing banners, enabling users to tell a visual story of their properties using dynamic templates and visual elements.
- **User Benefit:** The automated video platform swiftly transforms product details into captivating introductory videos, providing a captivating viewing experience for potential buyers or renters, enhancing marketing strategies for businesses.

2. User-Friendly Interface:

- **Functionality:** The platform features an intuitive and user-friendly interface for easy navigation and operation.

- **User Benefit:** Enhance user experience, making the video creation process accessible to individuals with varying levels of technical expertise.

3. **Customization Options for Unique Properties:**

- **Functionality:** Basic customization options allow users to tailor visual elements to suit the unique features of each property. Also, users have the ability to select a region on the preview banner.
- **User Benefit:** Present properties in a way that highlights their individual strengths and characteristics, catering to specific target audiences.

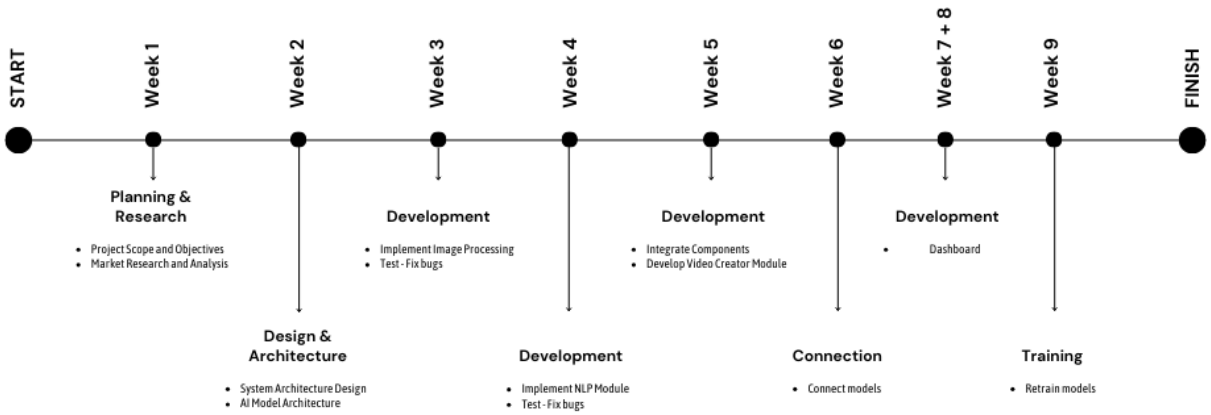
4. **Template-Based Video Generation:**

- **Functionality:** Implement pre-designed templates or frameworks that dynamically incorporate the provided product information (description, caption, and images) into video formats.
- **User Benefit:** Simplifies editing or updating videos by following a structured template format, enabling quick modifications to specific product details without redesigning the entire video.

5. **Streamlined Video Deployment:**

- **Functionality:** Users can quickly export and share finalized property videos in standard formats.
- **User Benefit:** Expedite the deployment of property videos across various digital platforms, reaching a wider audience and increasing property visibility.

VI. Timeline and Roadmap



VII. User Interface or Interaction

- Admin Dashboard: In this MVP, we will develop a simple admin dashboard where, alongside product management functions such as adding, editing, deleting, and updating product information (including descriptions, introductions, or additional images), we have introduced a new tab specifically for managing product introduction videos. In this tab, administrators

can select a specific product and click the 'Generate Video' button to create an introduction video for that particular product. At that point, the system will make an API call to the backend and return the corresponding introduction video.

VIII. Limitations and Future Enhancements

- *The limitations of the MVP*

- **Complexity of Creative Design:** While automation can handle certain aspects of creative design, achieving truly sophisticated and brand-specific visuals may require human intervention. The platform might struggle with intricate design nuances that are crucial for maintaining the unique brand identity of Nike and Adidas.
- **Potential for out of standardization:** Automation might lead to a degree of out of standardization in design, making it challenging to create truly unique and stand-out visuals for specific products or campaigns.
- **Content Understanding:** Automated systems might face challenges in fully grasping the nuanced meanings or cultural contexts embedded in product descriptions and captions. This could lead to misinterpretations and potential mismatches between visual elements and textual content.

- *Future Enhancements*

- **Natural Language Processing Improvements:** Enhance natural language processing capabilities to better understand the subtleties in product descriptions and captions. This could involve sentiment analysis and contextual understanding for more accurate content interpretation.
- **Collaborative Design Features:** Introduce collaborative features that allow human designers to work alongside the automated system, combining the efficiency of automation with the creativity and intuition of human designers.
- **Integrate User Feedback Mechanisms:** Develop features that allow users to provide feedback on generated banners, enabling the system to learn from real-world performance and user preferences.
- **Multi-Channel Integration:** Extend the platform's capabilities to seamlessly generate banners optimized for various digital channels, including social media, websites, and mobile apps, ensuring consistency across diverse platforms.

IX. Conclusion

In this proposal, we have presented a unique and innovative idea to build an automated system for creating product introduction videos. Utilizing data from the product database, we combine image processing, natural language processing (NLP) techniques, and AI models such as CycleGAN to generate dynamic and engaging videos.

Through the integration of NLP techniques, we can understand and analyze product descriptions, prominent features, and product names from textual data.

The system also comes with a user-friendly admin interface, enabling administrators to create videos easily and swiftly. This not only reduces time and effort investments but also enhances the advertising and marketing experience.

With significant potential in optimizing online marketing processes, this proposal marks not only a breakthrough in video creation but also boosts interaction and marketing performance for products. We hope that this idea will bring substantial value to the marketing strategy of businesses and create a unique experience for consumers.