Automous Image Descritption Genarator For ASU Lost And Found

Team 402

Ehtesham Sana, Ganesh Apparaju, Lin Liu, Ronald Chiang, Snigdha Pilli

TABLE OF CONTENTS

INTRODUCTION

OBJECTIVE

Automate ASU Lost and Found descriptions with computer vision for efficiency and accuracy.

Manual system's subjectivity causes errors, delays; computer vision ensures accuracy, efficiency.

CHALLENGES

BENEFITS

- Boosted accuracy via computer vision
- Accelerated processing for efficiency
- Improved searchability and satisfaction

TECHNOLOGICAL REQUIREMENTS

LIMITATIONS

- Manual methods are slow, yield incomplete or incorrect data.
- Inconsistent records hinder identification and retrieval for owners.

IMPROVEMENTS

- Transition to automated, image-driven system with computer vision.
- Simplified logging: Users upload photo, system generates accurate description.

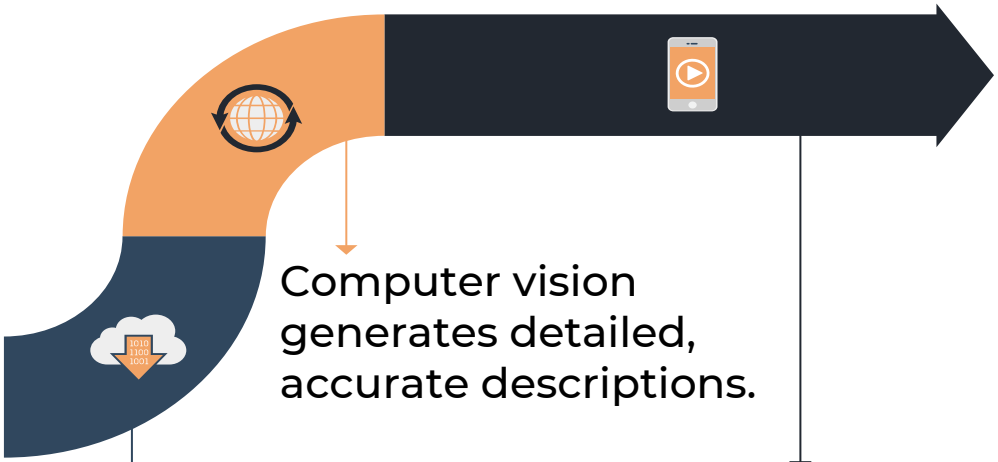
PROPOSED SOLUTION

PROBLEM DEFINITION

ASU's Lost and Found relies on inefficient, error-prone manual description process.

SOLUTION OBJECTIVE

Implement CV for automated, accurate description generation, enhancing efficiency.



Lost items uploaded by users to portal.

Computer vision generates detailed, accurate descriptions.

Descriptions and images auto-added, enhancing searchability and retrieval.

TECHNOLOGICAL REQUIREMENTS

ROLE OF CV

CV transforms manual, error-prone item description to swift, automated, accurate procedure, replacing subjective input with objective outputs.

CHOICE OF MODEL

- **BLIP model** for advanced image understanding and text generation.
- Renowned for efficiency and accuracy, ideal for immediate deployment without resource-intensive training.

MODEL OVERVIEW

01

BLIP – Berkeley Language and Image Processing

02

Integrates cutting-edge ML techniques for image analysis and text description generation.

03

Excels in comprehending intricate visual content and

04

Expresses it in natural language.

ARCHITECTURAL HIGHLIGHTS

Transformer-based architecture enhances speed and accuracy.

Utilizes attention mechanisms to focus on relevant image parts.

Excels in processing complex visual content into natural language.

BLIP MODEL

Aspect	Description
Model Type	Transformer-based architecture – combination of Vision Transformers and a language model for captioning
Architecture Features	Vision Encoder: Uses Vision Transformer to process image inputs Language Decoder: transformer-based language model to generate descriptions
Training Method	End-to-end training – joint optimization of both vision and language
Pre-training Tasks	Image-Text Matching (ITM) Masked Language Modeling (MLM) on image descriptions
Fine-tuning Tasks	Caption-based image retrieval Zero-shot and few-shot learning tasks to enhance model generalization
Datasets Used	COCO: Common Objects in Context – for general object detection and captioning Visual Genome: Rich annotations linking visual concepts & language descriptions Conceptual Captions: Automatic captions from web-crawled images
Output	Generates coherent and contextually relevant captions describing images

TECHNOLOGICAL REQUIREMENTS

INTEGRATION

- BLIP seamlessly integrates into ASU's Lost and Found via API.
- User uploads image, BLIP processes and returns descriptive caption.

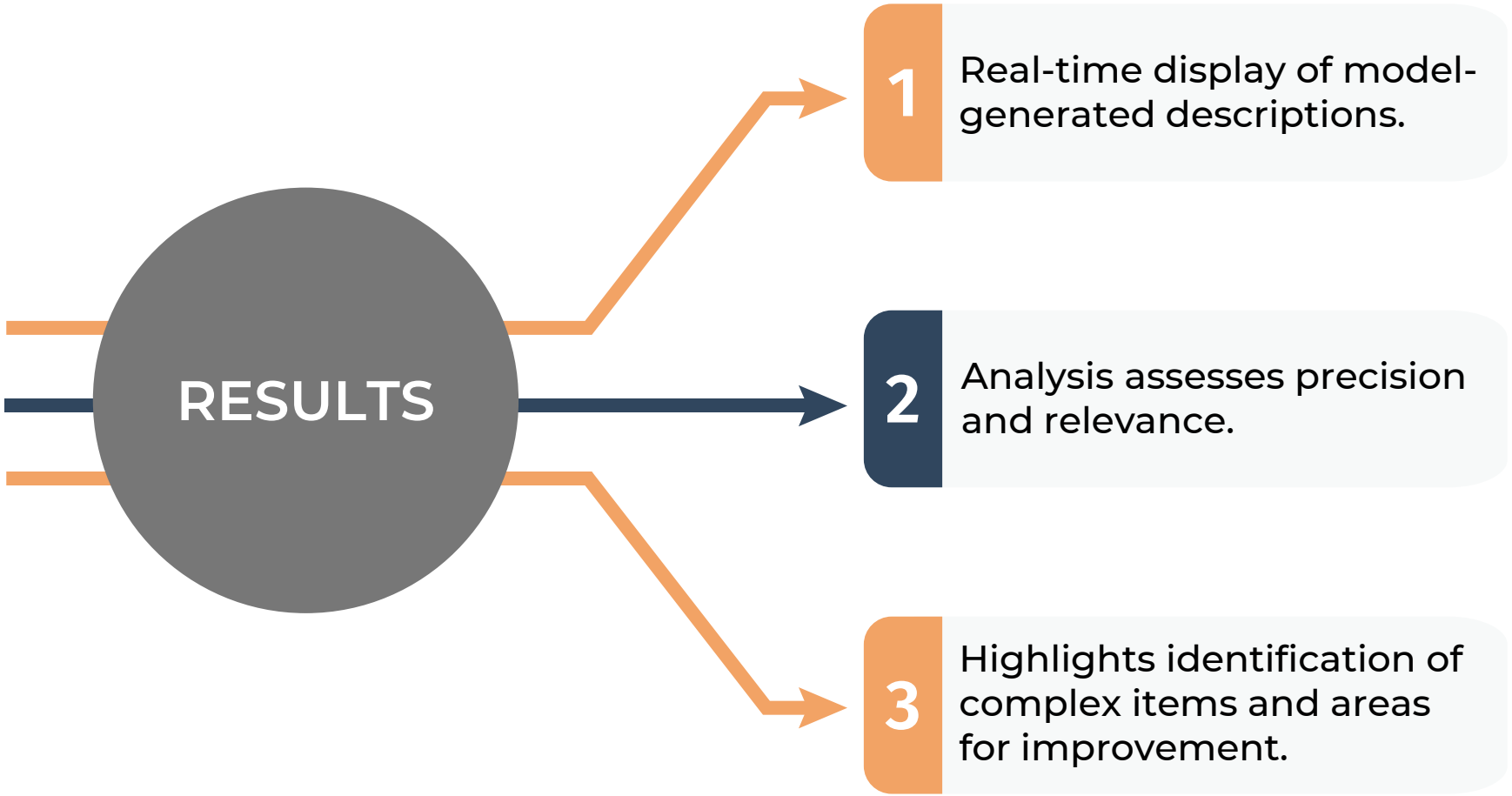
VALIDATION

Model outputs validated with curated Lost and Found images for accurate descriptions.

DEMONSTRATION SETUP

Live demonstration showcases BLIP model in action.

Utilizes typical ASU lost items like wallets, keys, and notebooks.


RESULTS
1
Real-time display of model-generated descriptions.
2
Analysis assesses precision and relevance.
3
Highlights identification of complex items and areas for improvement.

MONITORING & PROSPECTS

MONITORING

- Continuous assessment through metrics tracking.
- Regular updates based on new data, technological advances, and user feedback.

PROSPECTS

- Implement measures to identify and mitigate biases.
- Ensure fair and equitable descriptions across all item types.

TASK OWNERSHIP

Team Member	Role	Specific Tasks	Contribution to Final Product
Lin Liu	Project Management	Coordinated the project, oversaw documentation.	Ensured smooth project progress and documentation integrity.
Snigdha	Presentation Development	Developed and designed the PowerPoint slides.	Created a professional and coherent presentation to effectively communicate the project.
Ehtesham	Code Implementation	Implemented the BLIP model integration and functionality in Python.	Developed the core functionality for generating captions from images, crucial for the project's CV aspect.
Ganesh	Code Review, Testing and Debugging	Tested and debugged the code to ensure reliability and functionality.	Ensured that the code was efficient and bug-free, improving reliability.
Ronald	Proposal Writing	Wrote the project proposal and documented the process and findings.	Provided a detailed and structured proposal, clarifying project goals and methodologies for stakeholders.

THANK YOU!