Personalized Generative Storytelling with AI-Visual Illustrations for the Promotion of Knowledge in Cultural Heritage Tourism

Andrea Ferracani, Marco Bertini, Pietro Pala,
Gabriele Nannotti, Filippo Principi, Giuseppe Becchi

Presenter: **Niccolò Biondi**

SUMAC-24, 6th workshop on analy**S**is, **U**nderstanding and pro**M**otion of herit**A**ge **C**ontents
in conjunction with ACM Multimedia 2024, Melbourne, Australia, October 28th 2024

State of the Art (1/2)

Interactive Digital Storytelling

- Proven effective for engagement and education in cultural heritage contexts
- Previous approaches rely on **rule-based** or **manual storytelling**
- Recent advancements in **human-AI collaboration** leverage models like **GPT** for automated storytelling

State of the Art (2/2)

Generative Image Models:

- **Stable Diffusion** and **ControlNet** used for coherent image generation
- Existing models focus on enhancing immersion by creating illustrations that match the narrative flow
- **Challenges**: generating consistent character representations across scenes and managing **real-time generation**

Overview of the Application

Main Features

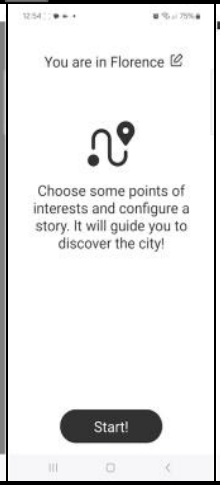
- **Personalized narratives** using GPT-4
- **User as the protagonist**: personalized storylines based on the user's selected **Points of Interest (POIs)** and **genre preferences**
- **Visual illustrations** dynamically generated to reflect user actions in the story
- Includes **multimedia content** such as panoramic images, videos and maps

WHAT THE TOURISM STORYTELLING APP DOES

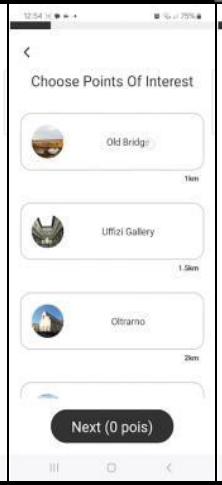
1. Select city
2. Select points of interest to be included in the tour, which will be the focus of the storytelling
3. Select the story genre
4. Select the type of narration: first or third person
5. Enter your name
6. Take a selfie for fine-tuning the image generation model
7. Follow the story
8. You'll find the generated image that reflects the representation of your actions described in the text

1

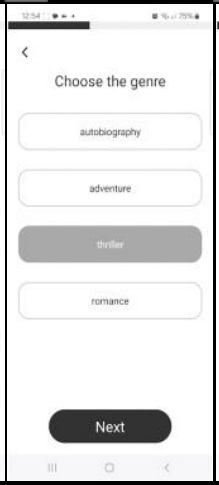
2



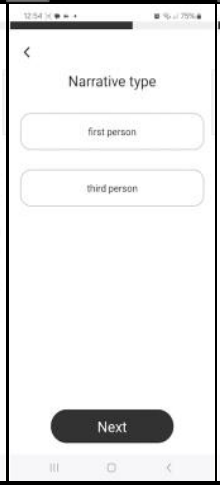
3



4

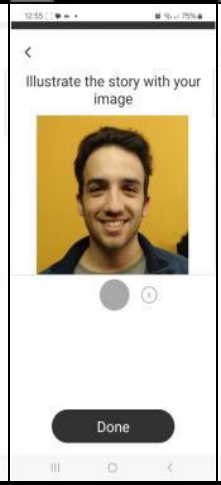


5



6

7



8

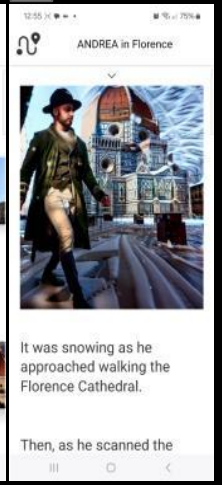


Image Generation Pipeline (1/2)

Overview of the Pipeline

The image generation pipeline consists of several stages:

- Step 1: modify the base image's **theme**, **weather**, and **time of day**
- Step 2: **inpainting** of the user's avatar in any pose
- Step 3: **photorealism enhancement**
- Step 4: smoothing of transitions between user inpainting and background
- Step 5: **user face inpainting** to integrate selfies into the images

Image Generation Pipeline (2/2)

Key Technologies

- **Stable Diffusion**: generates photorealistic base images.
- **ControlNet**: ensures the correct **pose** and **action** of the inpainted user
- **Low-Rank Adaptation (LoRA)**: maintains **identity consistency** when representing users across different scenes
- Optimized for **real-time performance** to ensure smooth storytelling

THE IMAGE GENERATION PIPELINE

THE IMAGE GENERATION PIPELINE

step 1: modifying the location theme, meteorological conditions and day-time/night-time

Process:

ControlNet:
- Analyzes structure and layout.

Image Generation Model:
- Creates final image.

Strength Adjustment:
- Emphasizes prompt for themes and weather.

Day to Night Conversion:
- Uses instruct-pix2pix if activated.

THE IMAGE GENERATION PIPELINE

step 2: inpainting a person in any pose using a control image

Process:

Inpainting Preparation:

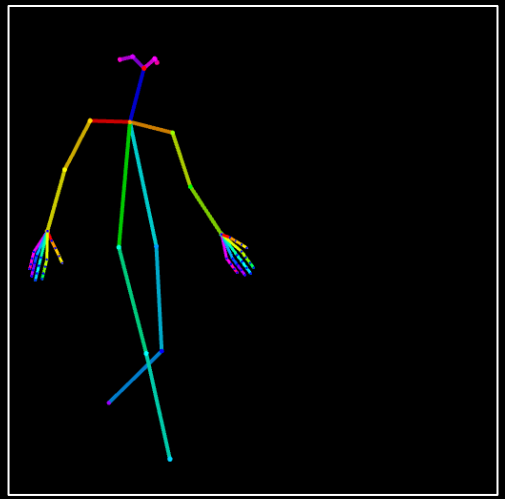
"runwayml/stable-diffusion-inpainting" prepares the model to focus on the inpainting area without impacting the background.

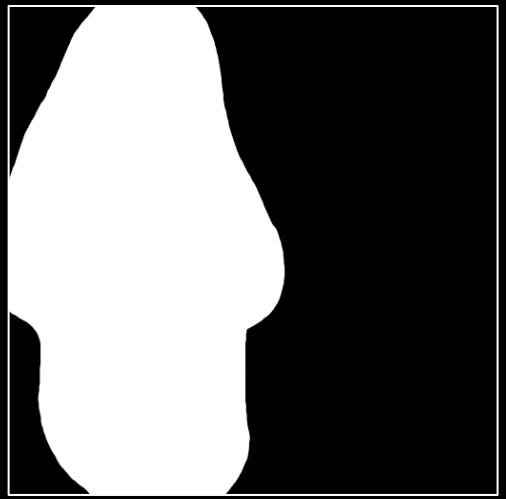
Pose Definition:

ControlNet defines the subject's pose or action. Uses a model to define and condition the subject's specific pose based on the skeleton image.

Subject Insertion

Inserts the subject into the mask defined area with the intended pose or action, ensuring proper integration.





THE IMAGE GENERATION PIPELINE

step 3: enhancing photorealism and image quality

Process:

Mask Application:

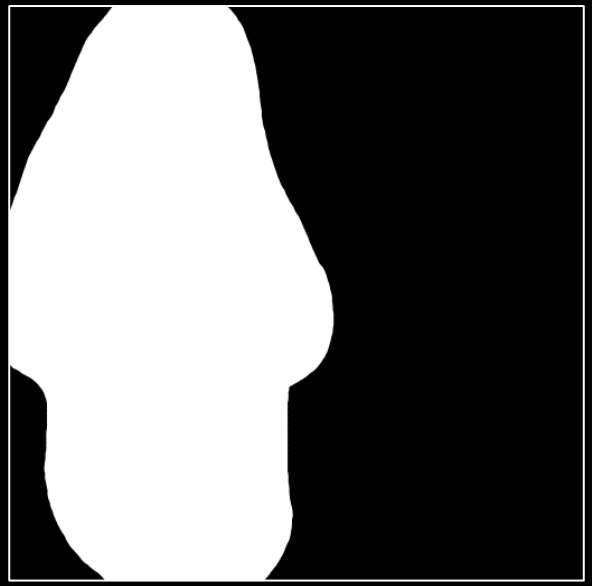
Apply a mask to define the specific area for enhancement, ensuring focus on the subject.

Photorealism Refinement:

Enhance the subject's photorealism using a specialized model:"dreamlike-art/dreamlike-photoreal-2.0".

Strength Control:

Carefully control the generation strength to maintain the original pose and minimize background impact.



THE IMAGE GENERATION PIPELINE

step 4: enhancing image homogeneity

Process:

Whole Image Processing:

Process the entire image to smooth transitions between the inpainted subject and the original background.

Boundary Smoothing:

Apply techniques to reduce the visibility of edges between the subject and background, creating a natural blend.

Strength Adjustment:

Use a low strength setting to ensure subtle adjustments, preserving overall content and achieving a cohesive look.

THE IMAGE GENERATION PIPELINE

step 5: inpainting the face of a specific person (the mobile app user)

Process:

Mask Application:

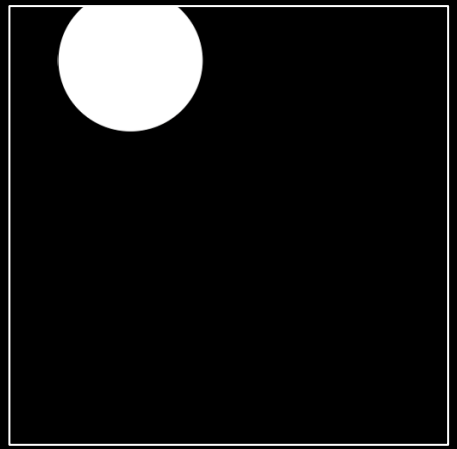
Apply a specific mask around the character's head to focus the inpainting process.

User Face Integration:

Use the LoRA model trained on the user's face to integrate facial features with the stable diffusion model.

Realistic Inpainting:

Ensure the realistic blending of the user's face onto the character to maintain photorealism and natural appearance.



Other Examples








Other Examples








User Evaluation (1/2)

Study Design

- Three user groups:
  1. **G1**: Printed materials (non-interactive)
  2. **G2**: App (non-interactive, no image personalization)
  3. **G3**: Full app experience (interactive with personalized images)
- Users experienced stories and were tested for **engagement**, **immersion** and **learning outcomes**

User Evaluation (2/2)

Measure	G1 (P)	G2 (ANI)	G3 (AI)
Satisfaction	6.4	5.5	5.8
Engagement	3.4	5.5	8.3
Immersion	2.9	6.5	7.8

Results – Satisfaction, Engagement, and Immersion

- **G3 (App Interactive)** showed the highest **engagement** and **immersion** scores
- Satisfaction was similar across all groups, but interactive elements contributed to a significant increase in **user immersion**

Conclusion and Future Work

Summary

- The app successfully integrates **personalized storytelling** and **AI-driven illustrations** to enhance cultural tourism experiences
- **User studies** confirmed improved **engagement**, **immersion**, and **learning outcomes**

Future Work

- Expand the application to include more **personalizations**
- Improve real-time performance of the image generation pipeline for higher-resolution outputs