

# Automatic Geo-alignment of Artwork in Children's Story Books

Aneesha Amodini  
Victor Suarez  
James Wald  
Jakub Dylag

Supervisor: Dr Shoaib Jameel  
GDP Group #11

## Introduction

By having access to familiar books, a child is more likely to read more, and strengthen their attitude towards reading. This project aims to:



## Methods

### 1 Prompt augmentation using keywords

Takes the raw text from a book, and automatically edit it (using Parts of Speech tagger) to produce a new prompt! New prompt: include key objects mentioned in the text + culturally sensitive.



"Riley and his dad took the kite to the park." → "Japanese childrens book style, Riley, his dad took kite to park"

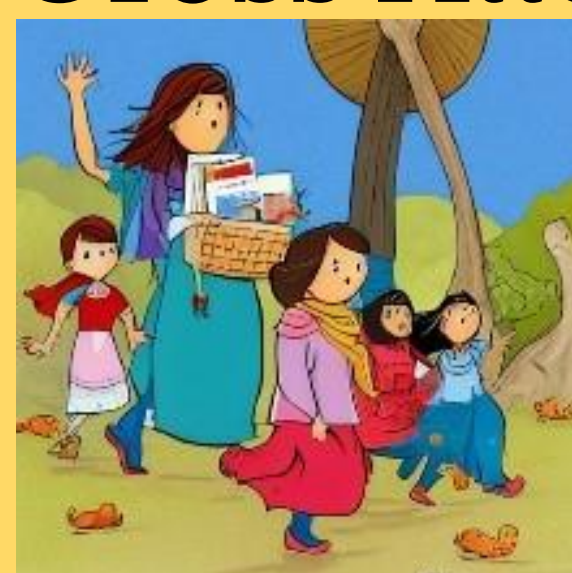
### 2 CLIP embedding mask

Use a CLIP Embedding Mask to encode the cultural information of the chosen geo-location. Identify which keywords will lead to the most success (based on relevance and editability) Removes the need to have domain specific knowledge.



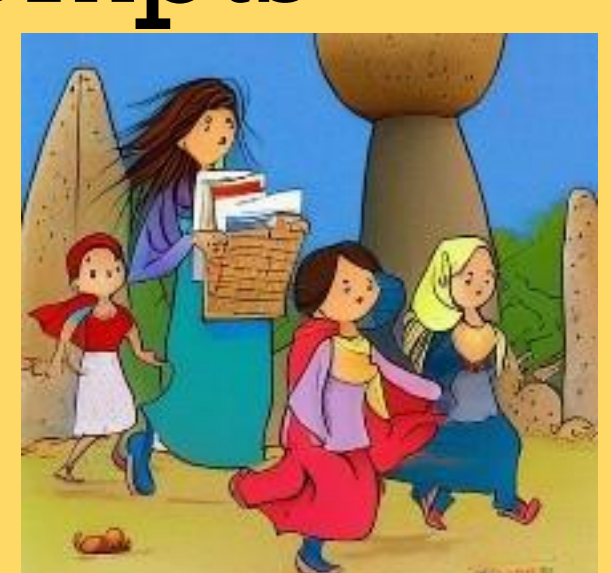
"Grandma has a big mango tree in her garden. Many birds come there." + CLIP embeddings

### 3 Cross Attention Control with Editorial Prompts



Method 1 prompt: "middle eastern childrens book style, The wind did not blow"

Editorial prompt: "middle eastern childrens book style, The **middle eastern** wind did not blow"



Characters are more culturally sensitive!

Fine tuning the prompt. Map the prompt to a desired culture, then introduce a **secondary editorial prompt** for further adjustment.

References:  
[1] "Stable Diffusion Online," [stablediffusionweb.com](https://stablediffusionweb.com/).  
<https://stablediffusionweb.com/>

## Stable Diffusion

An AI diffusion model that can generate photo-realistic images given a text input[1].



## Conclusion

All three methods created images in the **style of artwork originating from that culture**, changing the **characters' skin tone, hairstyle and attire**. Worked best with **simple, descriptive sentences**. Method 1 worked the best, fulfilling the project goal.

In the future, image quality can be increased by incorporating **Large Language Models** and personalised stylistic models, as well as adaptation to **Video and 3D sculpture generation for digital webbooks**.

## Evaluation

### Qualitative



An online questionnaire was submitted to the Japanese, Indian and Middle Eastern societies.

#### Japanese Questionnaire

section 1 Rank generated images from UK book to different countries art style:

Method 3! 👍

section 2 Similarity of generated images with original text:

Method 1! 👍

### Quantitative



#### Frechet Inception Distance (FID)

Distance between two feature vectors. Small FID = high similarity.

#### Method:

1. Real and generated images must be mapped to the same vector space
2. Average distance between all images in two sets of images is calculated

Method 2! FID=273.90 👍