



Relatório - Post Mortem

Equipe 1:

Alice Oliveira (aoqb)
Ana Letícia (alas3)
Guilherme Morone (gma2)
Williams Santiago (wssf)
Matheus Godinho (mrbg)
Enrique Laborão (elm2)
Marcello Menezes (meam)

Galeria

Vocais da Loirinha

Vocais da Loirinha.

 <https://criatividade-computacional.vercel.app/>

Geração de músicas

Foram testados alguns modelos encontrados na comunidade do discord [AI HUB](#).

1. O primeiro foi um modelo [RVC com 490 epochs](#).

Utilizamos o [colab](#) sugerido pela própria comunidade e seguimos os passos descritos nele.

Não obtivemos resultados satisfatórios.

A voz não era tão similar e tinha muitos ruídos. Por ter sido logo no início dos testes, pode ter havido também uma má escolha de áudios para serem testados

2. O segundo foi um modelo SVC 4.0 com 101k epochs.

Utilizamos o colab de modelos SVC também sugerido pela própria comunidade e seguimos os passos descritos nele.

Tivemos resultados muito bons, que foram utilizados no nosso projeto final.

No entanto tivemos algumas dificuldades já que o colab possui uma limitação de memória e alguns arquivos de áudio acabavam estourando esse limite no processo de conversão.

3. O terceiro foi um modelo RVC v2 com 300 epochs que já estava sendo utilizado no Kits.AI

O processo desse foi muito mais fácil, já que bastava apenas adicionar o arquivo de áudio no site, que ele mesmo se encarregava do resto.

Conseguimos fazer testes com os áudios que deram errado no modelo SVC e os resultados também foram muito bons.

Em todos os modelos tivemos dificuldades com ruídos, já que para a conversão ser feita era necessário que os áudios de entrada fossem apenas o vocal da música e era muito difícil encontrar esse conteúdo na internet com boa qualidade. Tentamos também utilizar ferramentas que separavam o vocal da instrumentação da música, como o Splitter, mas ele ainda mantinha alguns backing vocals e efeitos sonoros.

Processamento do áudio

Para utilizar os modelos mencionados anteriormente, era necessário o uso de músicas acapella. Inicialmente, buscamos áudios já existentes dos vocais extraídos das músicas. Posteriormente, tentamos utilizar o Splitter.ai para separar o vocal e o instrumental de algumas músicas cujo resultado não foi como esperado, mas os resultados obtidos através do uso dos vocais separados pelo Splitter não tiveram uma melhoria significativa, pois alguns ruídos de outros instrumentos e backing vocals ainda permaneceram presentes nos vocais separados.

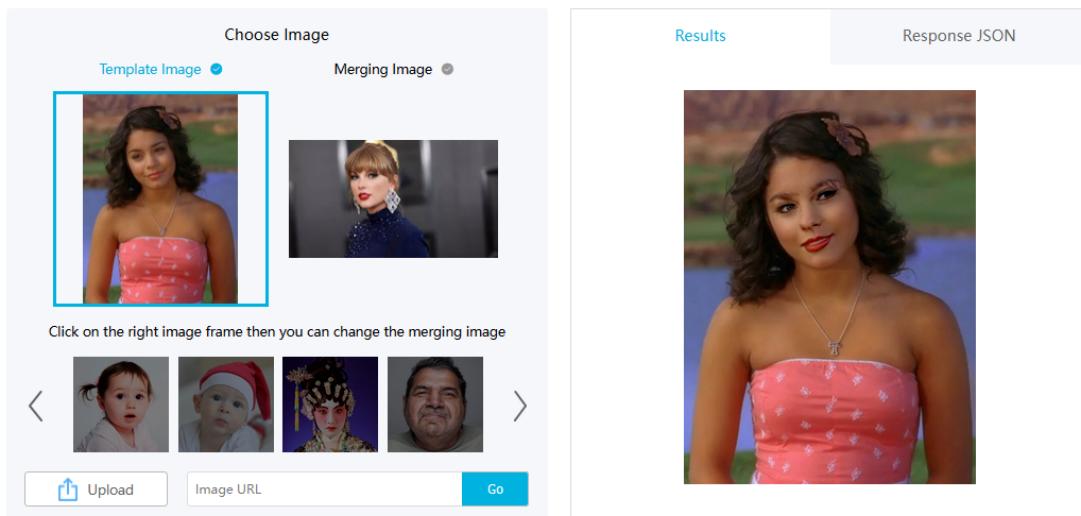
Geração de imagens

Para acompanhar cada música, foi proposta a geração de uma imagem que mesclava a foto do artista original e Taylor Swift. Para isso, foram testadas diferentes ferramentas:

1. Face++: Face Merging

Não obtivemos resultados satisfatórios.

Exemplo: resultado de mesclagem de Vanessa Hudgens (como a personagem Gabriella Montez de High School Musical) com Taylor Swift.



2. Lambda: Image Mixer

Também não obtivemos resultados bons.

Exemplo: resultado de mesclagem de Bruno Mars com Taylor Swift.



3. Playform: Face Mix

Com o propósito de só misturar as fotos, os resultados não foram tão bons e, além disso, havia limitação de créditos.

4. Playground AI

Foi a opção escolhida, apresentou resultados bons e mais opções de ajustes. Para gerar cada imagem foi escolhida uma foto do artista original para ser colocada como prompt juntamente com “taylor swift”. Abaixo estão detalhadas as configurações:

Filter: None

Prompt: “taylor swift”

Exclude from image: “nsfw, nude”

Image: (imagem do artista escolhido)

Model: Stable Diffusion 1.5

Guidance Scale: 7

Sampler: k_euler_ancestral

Seed: (aleatória)

Image Strength: (diferente para cada imagem)

Na galeria, é possível visualizar os parâmetros de cada imagem individualmente.

