Street View Images Generation

Alessandro Cesa

Università degli studi di Trieste

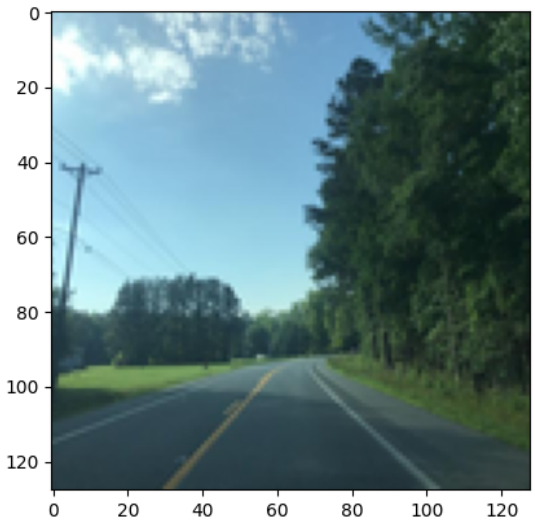
2024

Street View

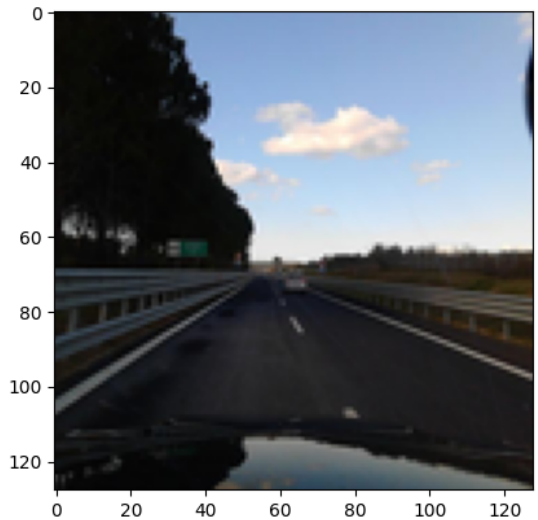
We want to generate, given a country, Google Street View style Images.

Open Street View Dataset

A free Dataset of 5 million Street View Images, available on Hugging face.



USA

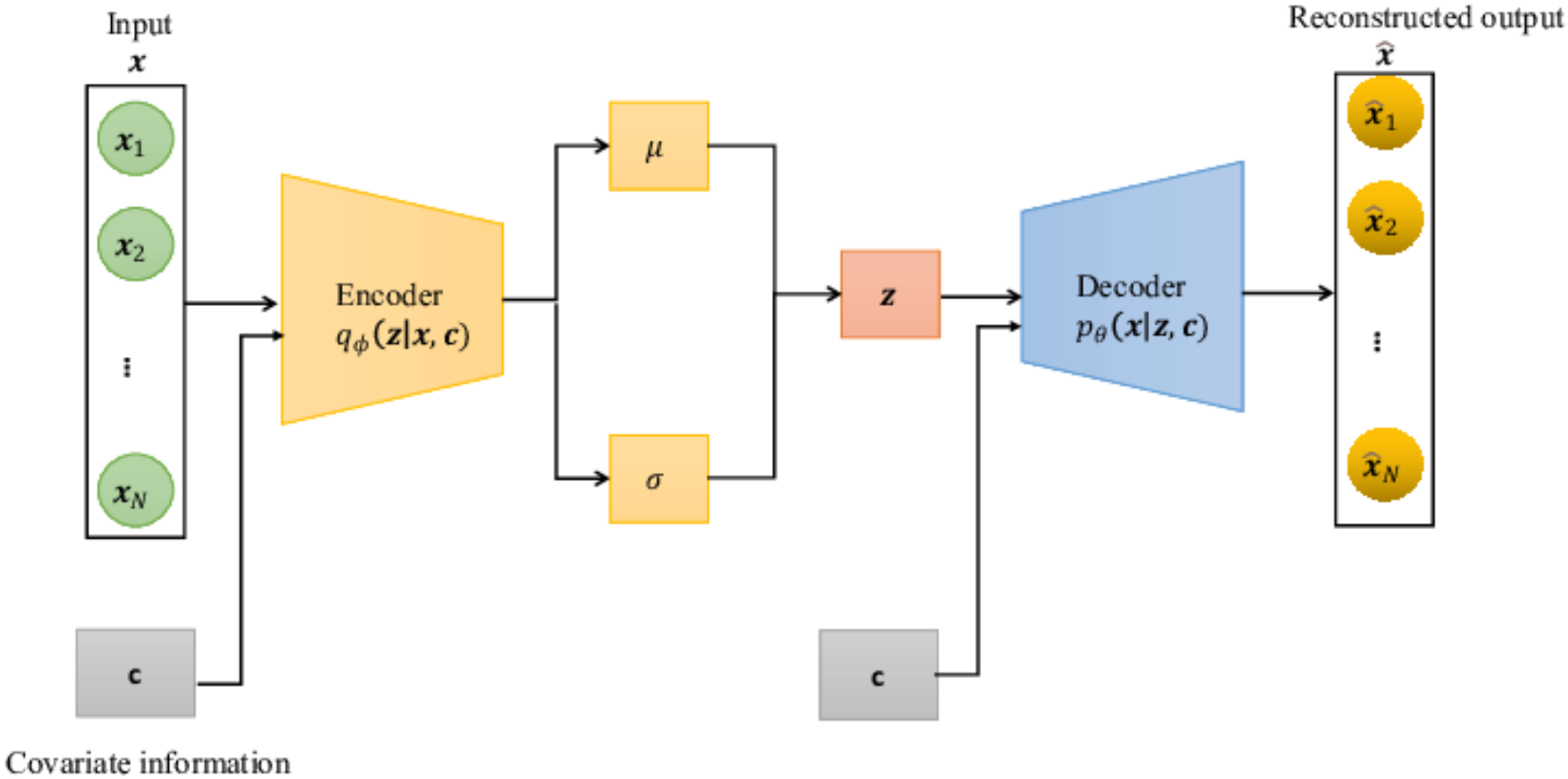


Italy

Pre processing

- Take only part of dataset: 160,050 images from 84 countries
- Reduce size of images from 1920x1080 to 128x128
- Extract Country
- One Hot Encode country names

Conditional Variational Auto Encoder



Conditional Variational Auto Encoder

Architecture:

- Image Encoding: 3 Convolutional Layers, with 3×3 kernel, stride 2, and padding 1, followed by BatchNormalization and with LeakyReLU activation
- Label Embedding: A fully connected layer that transforms the one-hot encoded label into an embedding of dimension
- Flattening of encoded image and concatenation with embedded label
- Fully connected layer with ReLU activation to transform the concatenation of encoded image and embedded label into the latent dimension
- Extraction of mean and variance of latent variable with fully connected layers
- Sampling of latent variable

Conditional Variational Auto Encoder

- The embedded label is re-concatenated with the latent variable
- Fully connected layer to transform concatenation of latent variable and embedded label
- Decoding:
  - Transposed Convolutional Layer, with 3×3 kernel, stride 1, and padding 1, followed by BatchNormalization and with LeakyReLU activation
  - Transposed Convolutional Layer, with 3×3 kernel, stride 2, and padding 1, followed by BatchNormalization and with LeakyReLU activation
  - Transposed Convolutional Layer, with 3×3 kernel, stride 2, and padding 1
  - Sigmoid to return output image

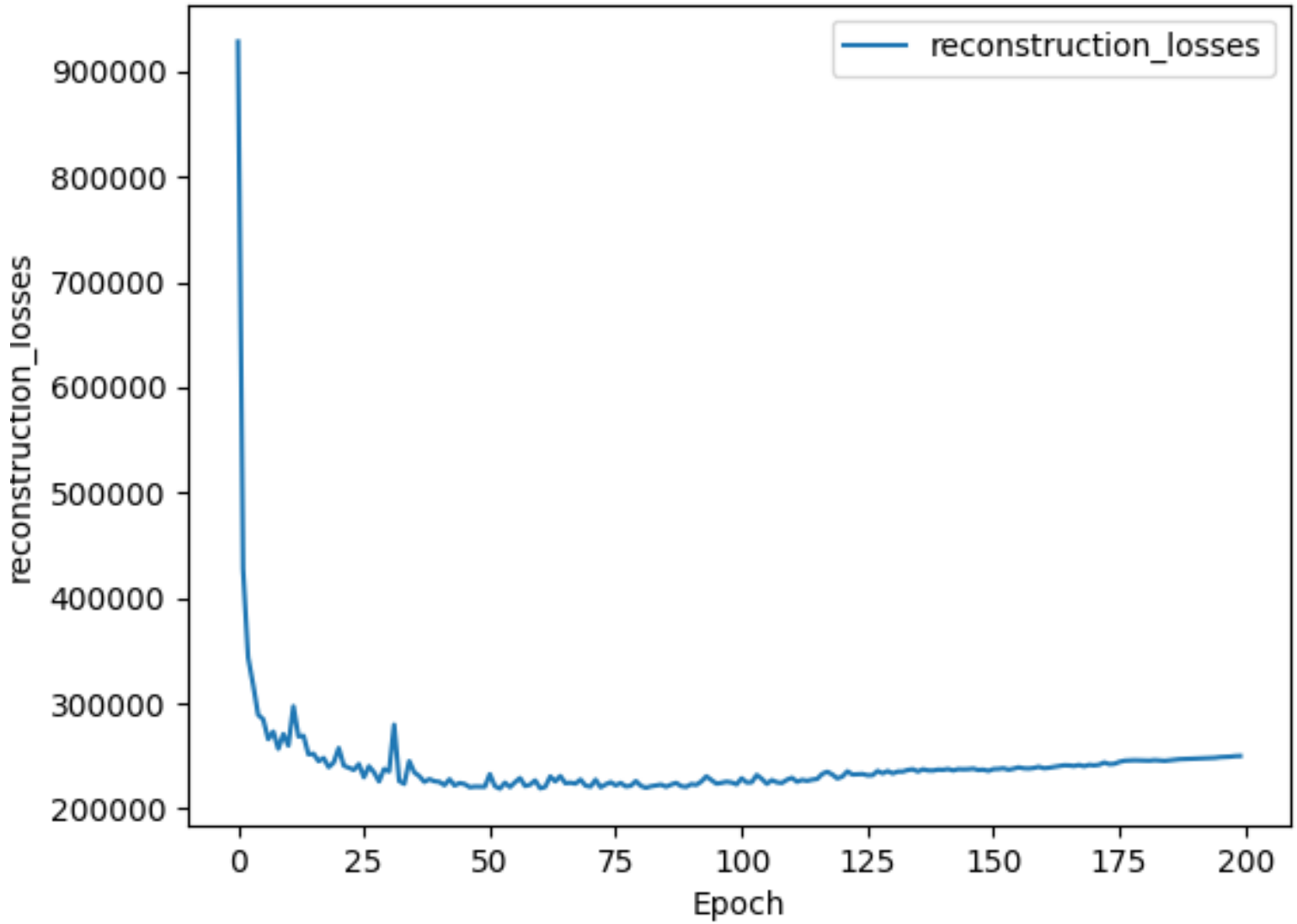
Optimizer and Losses

- Adam Optimizer
- Learning rate scheduler
- Mean Squared Error
- Kullback Leibler Divergence with growing weight

Training

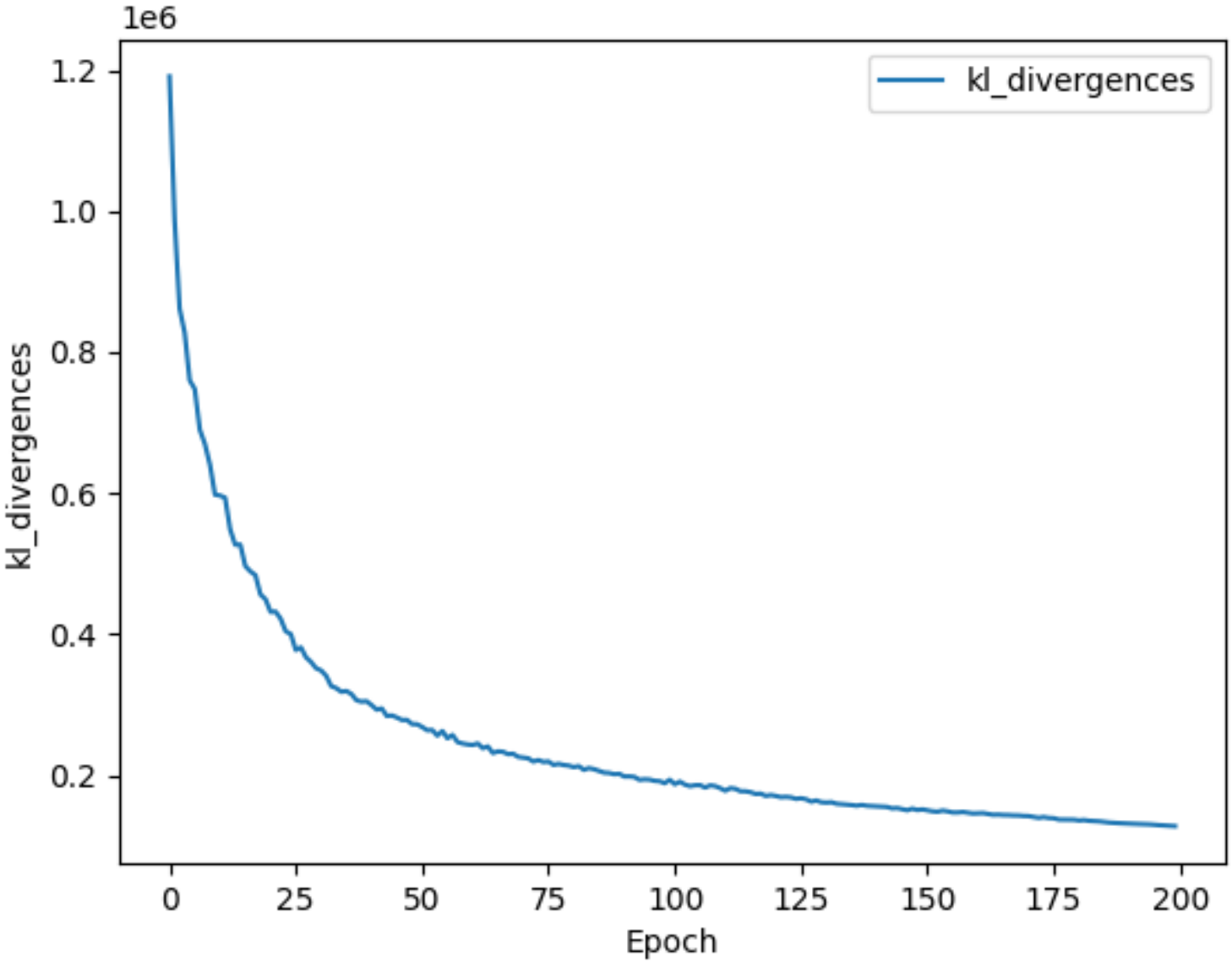
The model was trained on ORFEO GPUs for 200 epochs , for a total of 1 hour and 54 minutes.

Reconstruction Losses



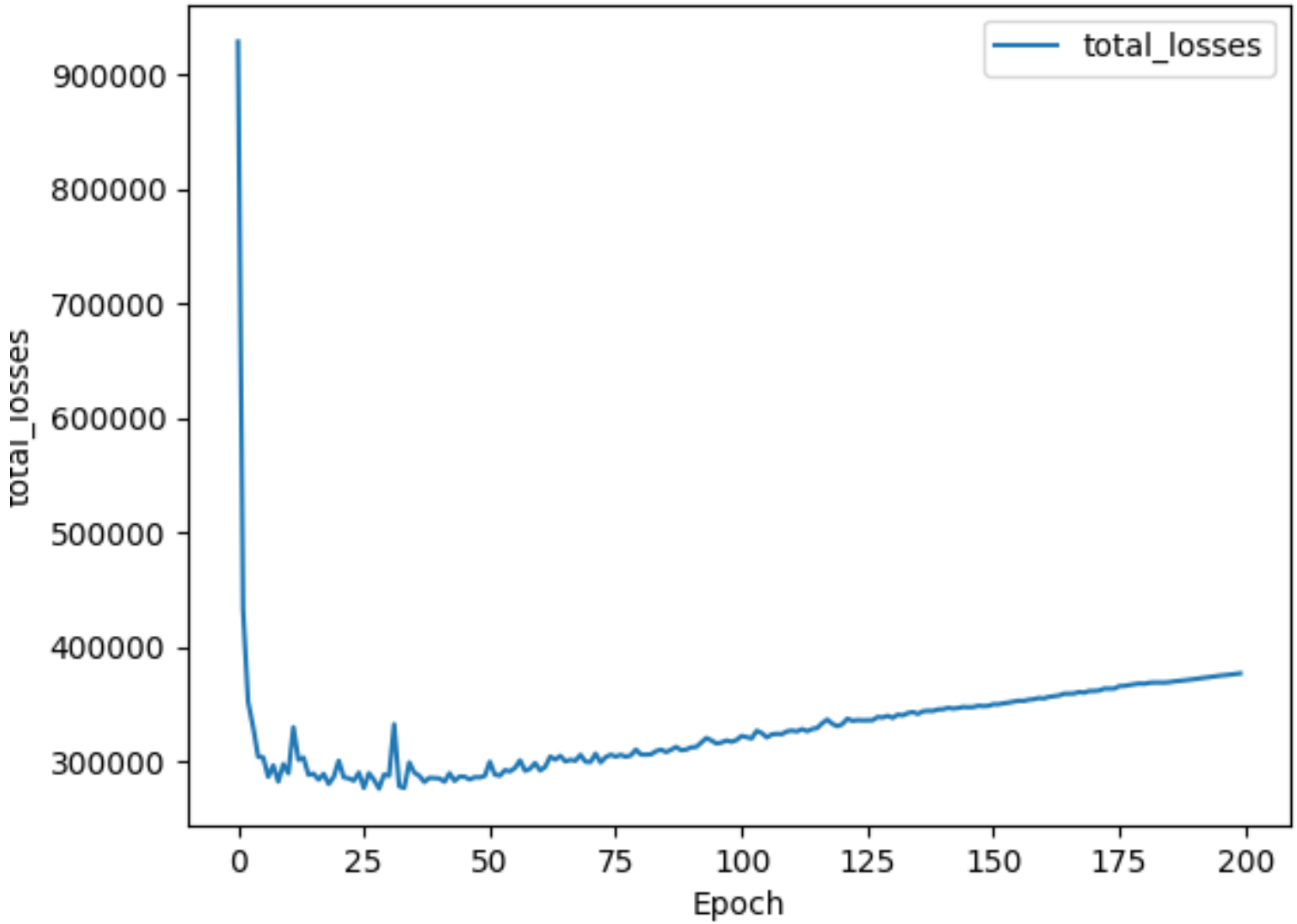
Reconstruction Loss

KL Divergences



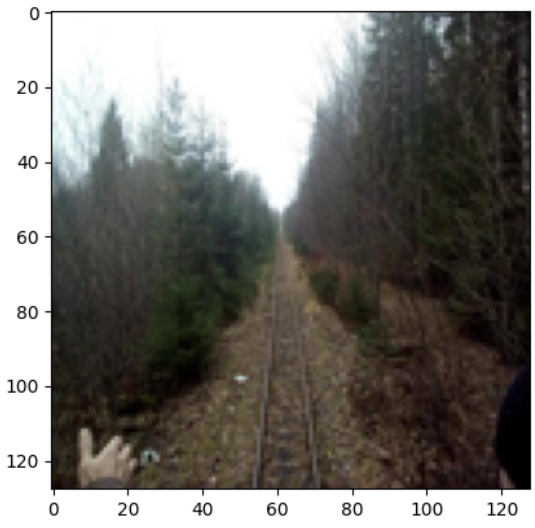
KL Divergence

Total Losses

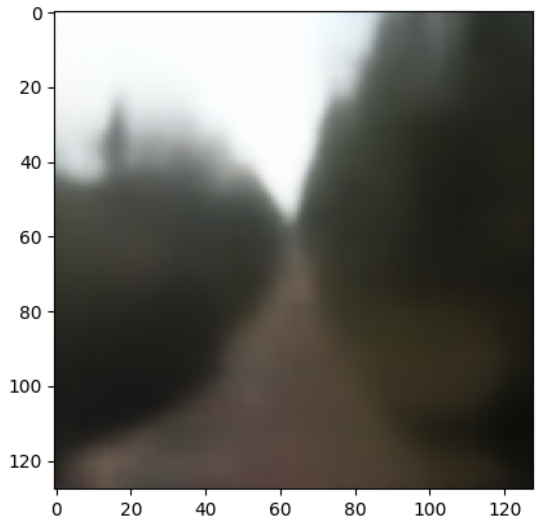


Total Loss

Results

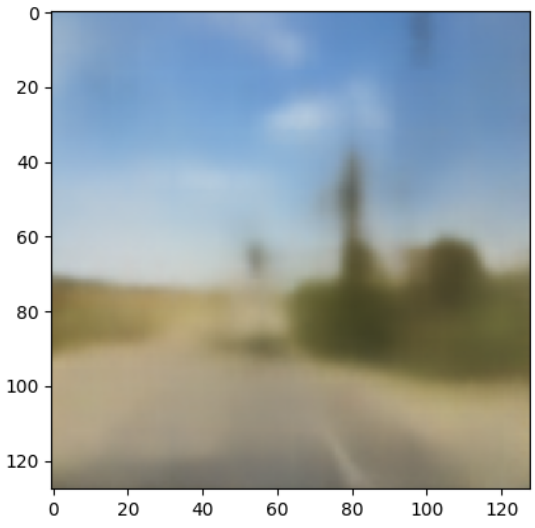


Original Image

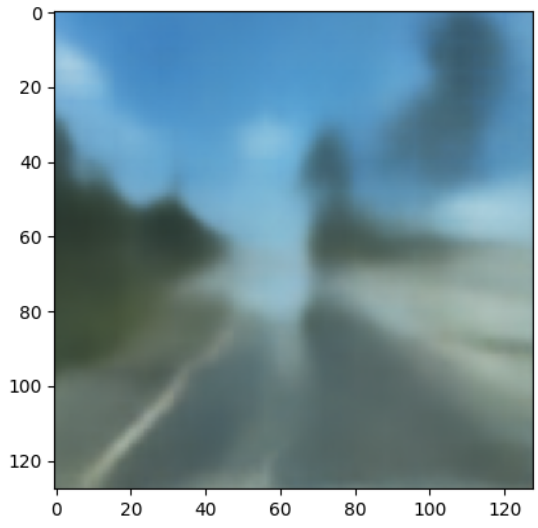


Reconstructed Image

Results

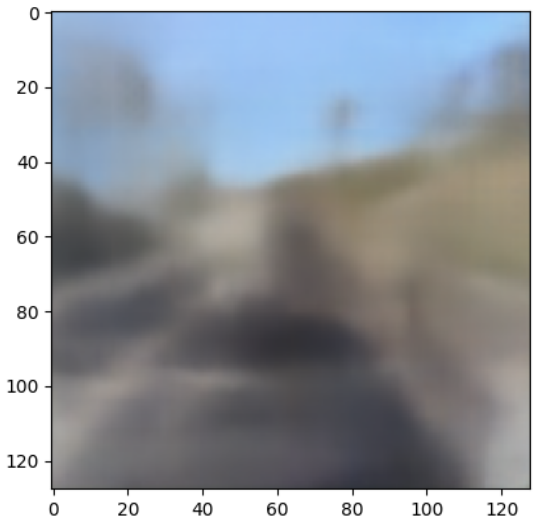


USA

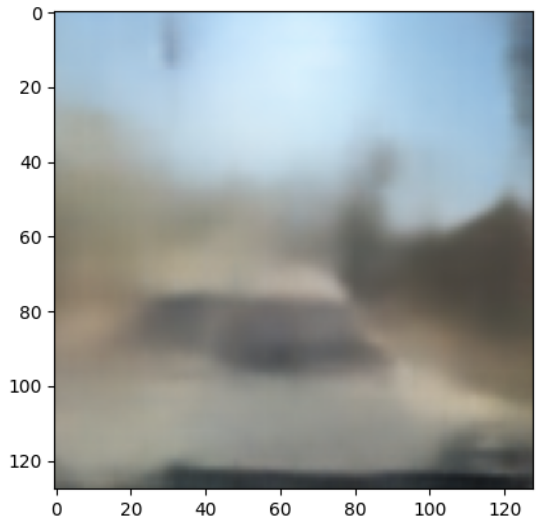


USA

Results



Italy



Morocco

Results

- Generates images that kind of look like Street View images
- Very blurred
- Not great difference between countries

Possible improvements

- More hyperparameter tuning
- Larger images
- Longer training
- Larger latent space and label embedding
- Perceptual loss