dSprites VSA

1   Introduction

The goal of this work is to extract the features of the image and represent them as multidimensional vectors, which, when summed up, will allow you to reconstruct the original image. Having started with a single object in the scene, we will try to move to a set of several objects.

2   Datasets

The task consists of three consecutive steps for each of which a different dataset was created. These datasets are based on the dSprites dataset Figure 1. dSprites is a dataset of 2D shapes procedurally generated from 6 ground truth independent latent factors Table 1
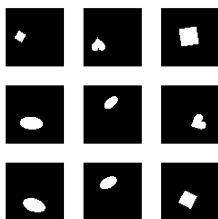

Figure 1: Example of images from dSprites dataset

Table 1: List of features

Feature	Distribution
Shape	square, ellipse, heart
Scale	6 values linearly spaced in [0.5, 1]
Orientation	40 values in [0, 2 pi]
Position X	32 values in [0, 1]
Position Y	32 values in [0, 1]

2.1 Paired dSprites

The purpose of creating this dataset is to make it possible to easily obtain paired images that differ in a single feature. This is possible due to the fact that in the original dataset the images are arranged in an orderly manner. An example of pairwise images in a dataset can be seen in the Figure 2
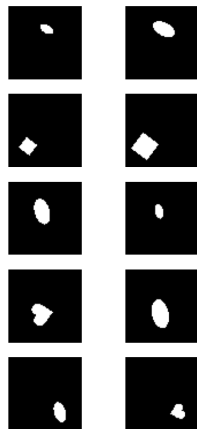

Figure 2: Visualizing Paired dSprites dataset elements

2.2 Scene-dsprites dataset

This dataset was created to test the model's ability to reconstruct a scene from the sum of object vectors. Dataset consists of 2 to 5 non-overlapping figures from dSprites dataset. An example of such images on Figure 3. This version differs from the original multi-dsprites dataset in that the figures do not overlap and have the same color.

Figure 3: Example of collected scenes in the scene-dsprites dataset

2.3   Paired-Scene-dSprites dataset

This dataset combines the capabilities of the first and second dataset. There are two objects in the scene image. One of these objects can change one feature. Figure 4
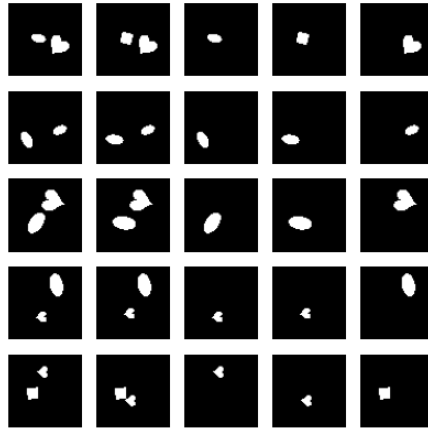

Figure 4: Example of images from paired-scene-dsprites dataset. From left to right, first scene, pair scene, the object to be changed, his pair, second object

3   Models

The main idea of the model is that it can reconstruct from the sum of features the object in the image Figure 5. There may be several objects with different features on the stage Figure 6
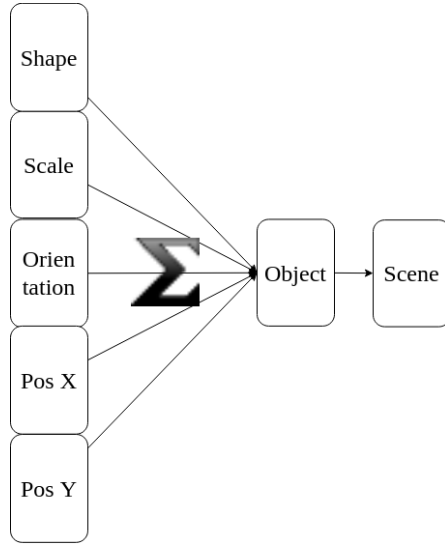
Figure 5: The summation of image features to obtain a latent representation of the object on the scene.
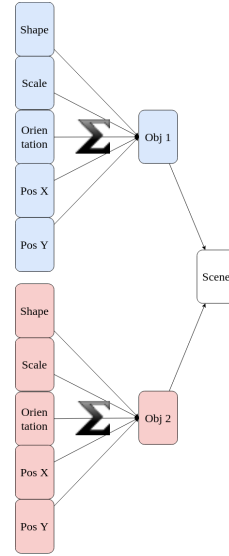

Figure 6: The summation of the latent representations of several objects on the scene.