**IBM Deep Learning Project – Car Image Generation**

This study looks into using deep learning to generate novel images of cars. To main approaches were considered for image generation: a generative adversarial network and a variational autoencoder.

**1 Introduction**

Data was taken from <http://ai.stanford.edu/~jkrause/cars/car_dataset.html> [1]. This dataset contained 16185 images of 196 classes of cars. The objective of this analysis was to determine whether GANs or VAEs were more effective in generating “new” images of cars – images that could be explored for future models of cars and new concept ideas.

**2 Data Exploration and Cleaning**

The dataset is split in a 50-50 proportion for training and testing. As the testing set was used for testing the VAE, only the training et was used to train both models to allow for fairer comparison. To improve results in future is definitely advised to make use of both sets of data.

The distribution of the counts of the classes can be seen in Figure 1. As all were between 48 and 136, none were removed. It was noted though that the train and test set had similar distributions of classes.

|  |
| --- |
|  |
| Figure 1: Counts of car models |

For pre-processing, several transforms were applied to the images. In order to make them all of consistent size, they were resized and centre-cropped to 64 by 64 pixels for both networks. Additionally, the colour values for each pixel were normalised. In each colour channel, the pixels were normalised with a mean on 0.5 and standard deviation of 0.5. This allowed for faster training of the networks.

|  |
| --- |
|  |
| Figure 2: Example batch |

**4 Models**

**4.1 GANs**

Generative Adversarial Networks (GANs) use a generator network and a discriminator network. This former has the aim of fooling the discriminator network into classifying the fake outputted image as real while the later aims to classify an image as real or fake correctly.

**4.2 VAEs**

Variational Autoencoders (VAEs) consist of an encoder network and decoder network. The encoder network maps the input image to the latent space and the decoder maps a distribution in latent space to a reconstructed image.

**5 Key Findings**

**5.1 The GAN**

|  |
| --- |
|  |
| Figure 3: Images produce by the GAN |

After 100 epochs of training the images produced by the GAN show a good improvement

**5.2 The VAE**

The variational autoencoder produced images as shown in Figure 4.

|  |
| --- |
|  |
| Figure 4: Images produced by the VAE |

As can be seen, the results are a lot worse than the GAN after 100 epochs of training – images were also converted to black and white to speed up training.

**6 Possible Flaws**

One of the larger flaws we found with our models was the mixing of geometries with respect to the orientation of the wheels of the cars.

**7 Next Steps**

Having trained these models on a wide range of cars, one could use transfer learning in order to make the model more specific to say supercars, producing more interesting results and leveraging the time spent training this model. However, in order to do this, more data is required, in the form of images of supercars. Another possible step is to make the variational autoencoder class specific.

**References**

[1] Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, **3D Object Representations for Fine-Grained Categorization**, *4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013* **(3dRR-13)**. Sydney, Australia. Dec. 8, 2013.  
[[pdf]](http://ai.stanford.edu/~jkrause/papers/3drr13.pdf)   [[BibTex]](http://ai.stanford.edu/~jkrause/papers/3drr13.bib)   [[slides]](http://ai.stanford.edu/~jkrause/papers/3drr_talk.pdf)