Unsupervised Learning to Explore Automotive Design

Mason Llewellyn[1],
mason747@stanford.edu,
[1]Department of Computer Science, Stanford University

Abstract

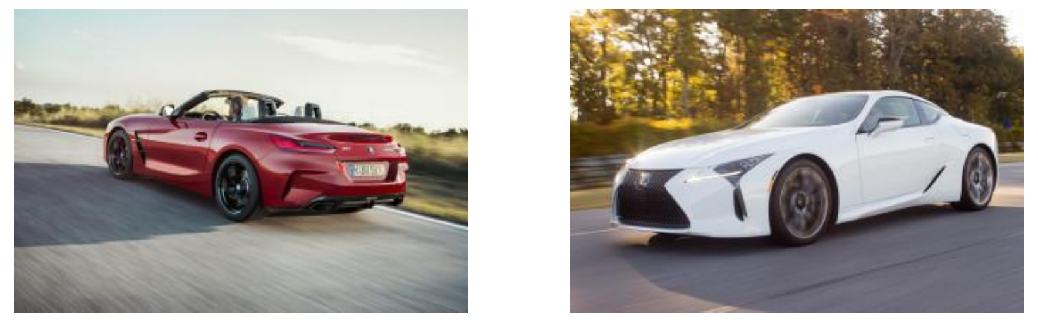
Within this paper, we investigate the application of unsupervised learning to image generation, with the intent to create new car designs as a mixture of k existing cars. To accomplish this, we encode images of existing cars into a lower-dimensional latent space, create a new latent vector that represents a convex combination of the existing cars, and finally decode the combined latent vector into a new image. We investigated various methods of compressing vehicles into a lower-dimensional space, including Principal Component Analysis (PCA), a Convolutional Neural Network-based Variational Autoencoder (VAE), and a Resnet18-based VAE. Our methods were evaluated both by their ability to reconstruct a single image of a car, and their ability to design new cars as a mixture of multiple existing cars. The generated images were assessed both on their apparent visual quality and their Frechet Inception Distance to the original dataset of real car images.

Introduction

In the century since the automobile's invention, its styling and design have been a core part of its appeal. To create the styling elements of new cars, automotive designers often borrow and combine styling elements of existing cars making many new cars a mixture of ones that came before. **Within our project, we apply unsupervised learning techniques to automate the design of new cars as a mixture of existing cars.**

- Using a dataset of automotive images, we learn to map vehicle images to a lower-dimensional latent space.
- We mix cars using by decoding a convex combination of multiple vehicles' latent vectors back into image space.

Dataset



(a) BMW Z4 (b) Lexus LC

Figure 1: Example Images from Car Connection Dataset

To train our model, we used the Car Connection Picture Dataset [1], which includes over 64,000 images of vehicles scraped from thecarconnection.com, a vehicle review website.

- To train our model, we used the Car Connection Picture Dataset [1], which includes over 64,000 images of vehicles scraped from thecarconnection.com, a vehicle review website.
- Before training we used a YOLOv5 object detection model to crop each image to only include the car and to remove all images not containing cars from our dataset. All Images are then resized to 64x64.
- After this process we were left with 45,376 images.

Methods

Baseline: Principal Component Analysis

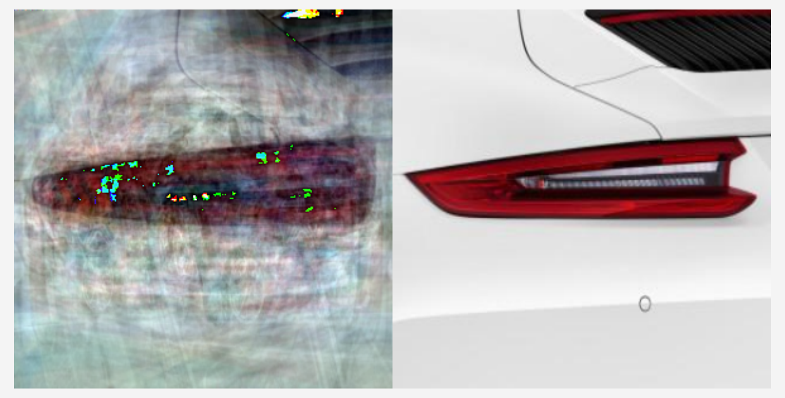


Figure 2: PCA-Based Reconstruction of unseen image

- As an initial baseline, we attempted to learn a compressed representation for the Porsche 911 sportscar using PCA.
- Though it works well for images within the training set, it fails to represent unseen images that differ slightly from training images.
- This sensitivity to pixel-level features of an image led us to move on from PCA without further testing in search of a method less sensitive to minor pixel-level variations in the

Variational Autoencoders

- To better encode high-level features in the image, we turned to a varitational autoencoder (VAE)[4] with convolutional encoders and decodeder networks.
- To further improve our encoder's representational power, we also experimented with a VAE with Resnet-18[2] based encoder and decoder networks.

Design Mixing

Mixture of k Designs

$$z_m = \frac{1}{k}\sum_{i=1}^{k}\mu(E(x_i))$$

$$x_m = D(z_m)$$

- We mix the designs of k vehicles by using our VAE's encoder to encode all images into the latent space and average all latent vectors.
- We then decode this average vector back into image space to get our new design.

Evaluation

- We evaluate the quality of our designs using both qualitative evaluation and Frechet Inception Distance (FID)[3].
- FID uses the activations from a pretrained inception network to quantify the realism of a set of generated images.
- μ_r, μ_g, Σ_r, and Σ_g are the means and covariance matrices of the inception network's activations for real and generated images respectively.

Frechet Inception Distance

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{trace}\left(\Sigma_r + \Sigma_g - 2\Sigma_r\Sigma_g\right)$$

Results

2-Car Mixing

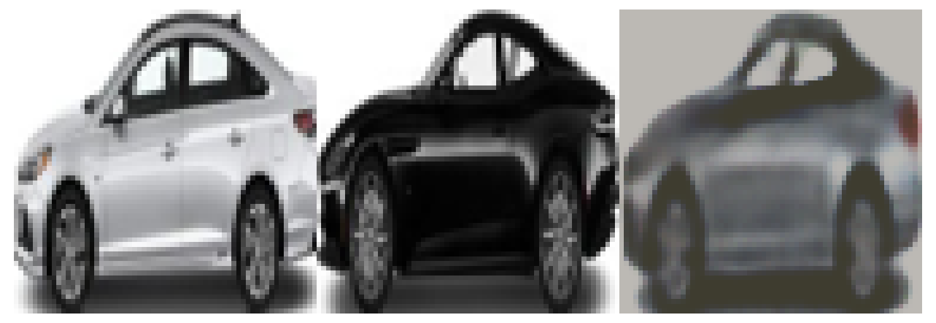


Figure 3: Convolutional VAE Combination of two cars

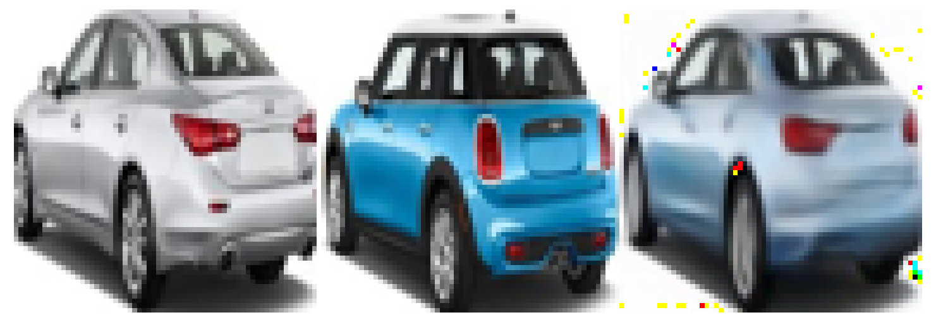


Figure 4: Resnet18-Successful Combination of Two cars

- The Resnet-18 VAE produces higher quality mixed designs than the convolutional VAE according to both visual inspection and the FID metric.
- As can be seen in 4, the Resnet-18 VAE's mixed design clearly draws its body shape from the Infiniti Q50 Sedan on the and it's color from the blue mini cooper.

3-Car Mixing

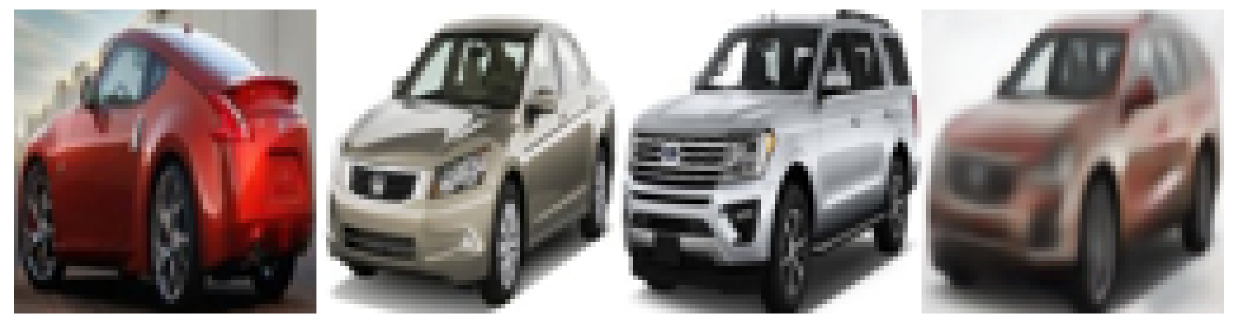


Figure 5: Resnet18 VAE combination of three cars

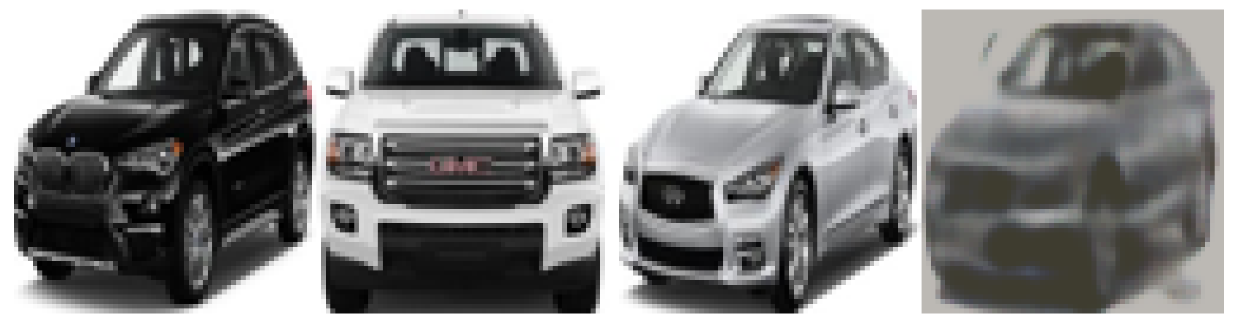


Figure 6: Convolutional VAE combination of three cars

- According to the FID metric and visual inspection 3-car mixed designs for both the convolutional and Resnet-18 VAEs were much less realistic than their two-car counterparts.
- In fact, aside from the cherry picked examples in figures 5 and 6, the majority of the results did not look like cars.

VAE	k	Frechet Inception Distance
Convolutional	2	77.96
Convolutional	3	87.31
Resnet-18	2	27.99
Resnet-18	3	32.33

Table 1: k-image mixing results

Conclusions

- Our mixing method works best when all original images show cars at the same angle. Mixing car images at different camera angles often leads to very unrealistic results. A potential fix for this issue would be to use a representation for vehicle design that does not rely on angle such as a 3D point cloud that captures the car's shape
- The improvement in image realism of the Resnet-18 VAE compared to the Convolutional VAE is likely due to the increased representational power of the architecture compared to our Convolutional architecture.
- Future work should investigate if further impruments in representational power (such as with Resnet-34 or Resnet-50) architecure leads to similar improvements in performance.

References

[1] N. Gervais. The car connection dataset. https://github.com/nicolas-gervais/predicting-car-price-from-scraped-data/tree/master/picture-scraper, 219.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[3] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[4] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2013.