

# RBE 474X

## Project 4

Colin Balfour  
Worcester Polytechnic Institute  
Worcester, Massachusetts 01609

Khang Luu  
Worcester Polytechnic Institute  
Worcester, Massachusetts 01609

Thinh Nguyen  
Worcester Polytechnic Institute  
Worcester, Massachusetts 01609

### I. PART 1 - DIFFUSION TRAINING

We trained the diffusion model to generate new images from a provided dataset. This training is done over linear betas. Below are the hyperparameters used for training:

- Batch size: 80
- T: 500 (number of steps)
- Channel: 128
- Channel-multiplier: [1, 2, 2, 2]
- Learning rate:  $1e-4$
- Beta 1:  $1e-4$
- Scheduling: linear
- Beta T: 0.028
- Loss Function: MSE loss

For training, we are using the CIFAR10 dataset. Figure 1 is the loss plot of the training:

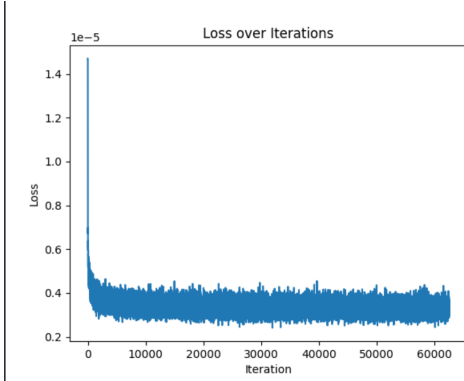


Fig. 1. Plot

Figure 2 to 11 show the images for every class generated by the diffusion model.

### II. PART 2 - DATA AUGMENTATION USING DIFFUSION

For this section, we are augmenting the cifar10 dataset with 5000 extra images per class (examples are shown in Figure 2 to 11).

The network used for this is a custom implementation of MLP. The hyperparameters used for training are:

- Learning rate:  $1e-4$
- Epoch count: 40
- Batch size: 32

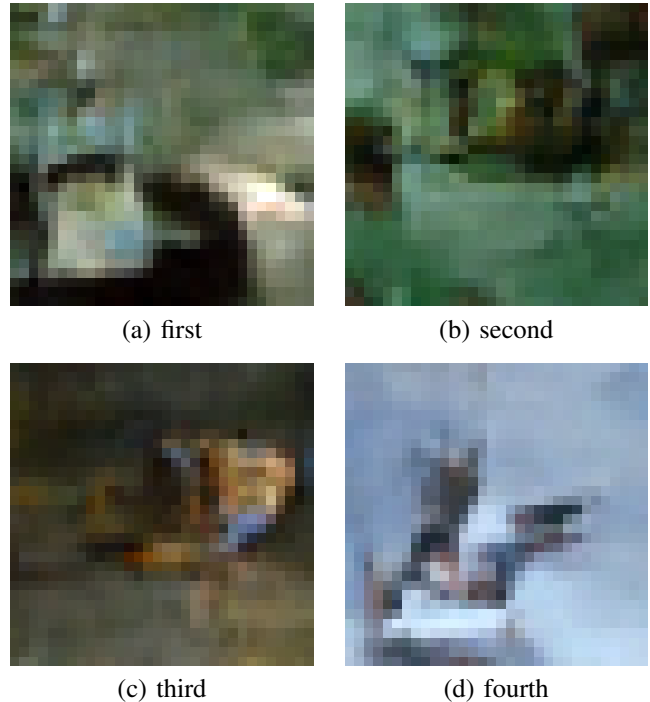


Fig. 2. Bird images, generated from the diffusion model

#### A. Results for non-augmented data

The accuracy for the non-augmented data is 44.05%. Figure 12 is the loss plot for the training.

And the confusion matrix for the training is shown in Figure 13.

#### B. Results for augmented data

The accuracy for the augmented data is 41.14%. Figure 14 is the loss plot for the training.

And the confusion matrix for the training is shown in Figure 15.

#### C. Comparison

The accuracy for the non-augmented data is 44.05%, while the accuracy for the augmented data is 41.14%. Although the augmented data includes additional images generated by the diffusion model, it does not improve the accuracy. This is likely because the generated images are somewhere between



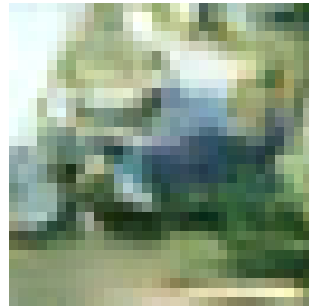
(a) first



(b) second



(c) third



(d) fourth

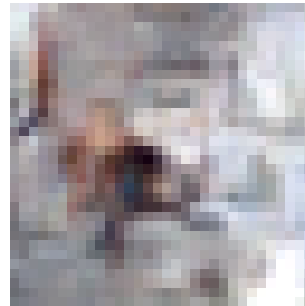
Fig. 3. Car images, generated from the diffusion model



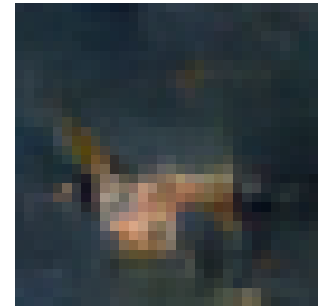
(a) first



(b) second



(c) third



(d) fourth

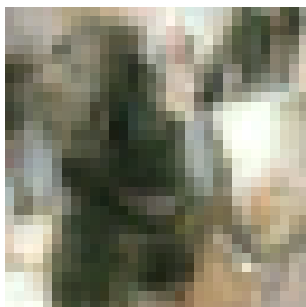
Fig. 5. Deer images, generated from the diffusion model



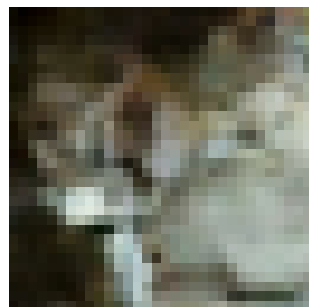
(a) first



(b) second



(c) third



(d) fourth

Fig. 4. Cat images, generated from the diffusion model



(a) first



(b) second

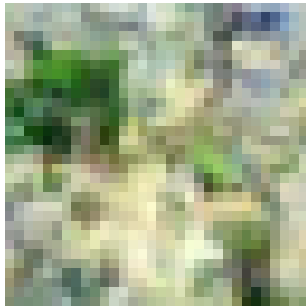


(c) third

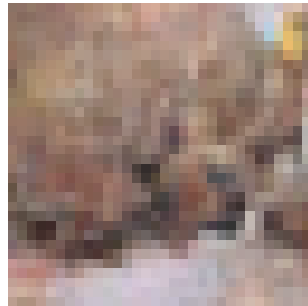


(d) fourth

Fig. 6. Dog images, generated from the diffusion model



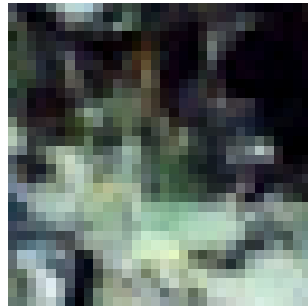
(a) first



(b) second

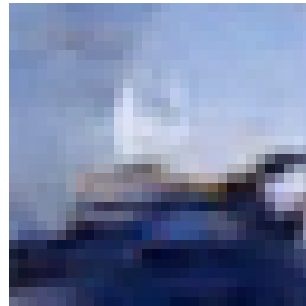


(c) third

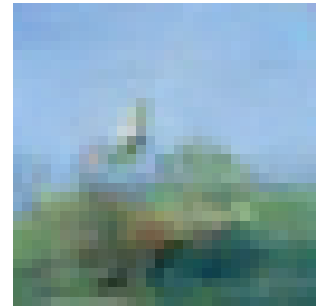


(d) fourth

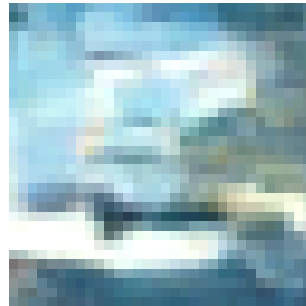
Fig. 7. Frog images, generated from the diffusion model



(a) first



(b) second

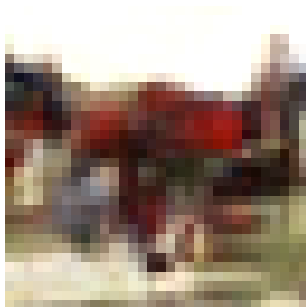


(c) third



(d) fourth

Fig. 9. Ship images, generated from the diffusion model



(a) first



(b) second



(c) third



(d) fourth

Fig. 8. Horse images, generated from the diffusion model



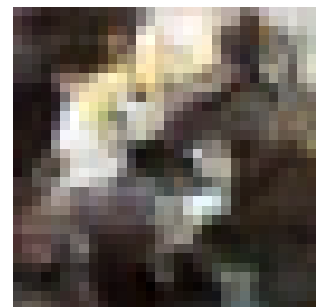
(a) first



(b) second



(c) third



(d) fourth

Fig. 10. Truck images, generated from the diffusion model

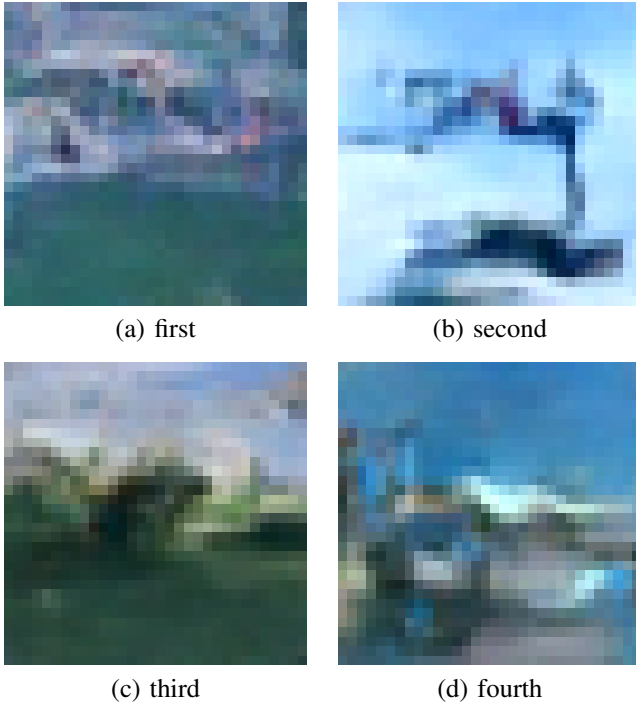


Fig. 11. Plane images, generated from the diffusion model

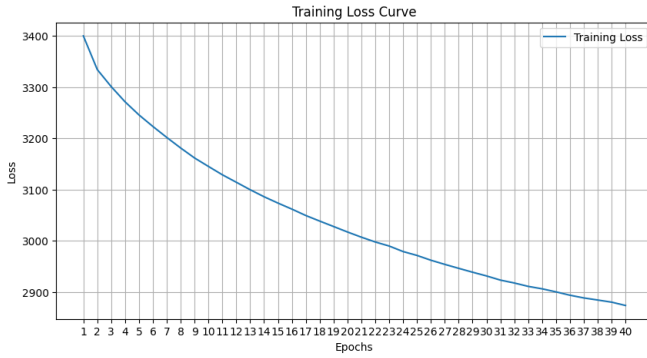


Fig. 12. Non-augmented data, loss plot

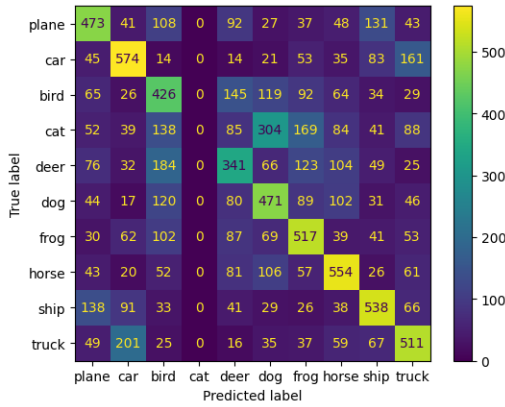


Fig. 13. Non-augmented data, confusion matrix

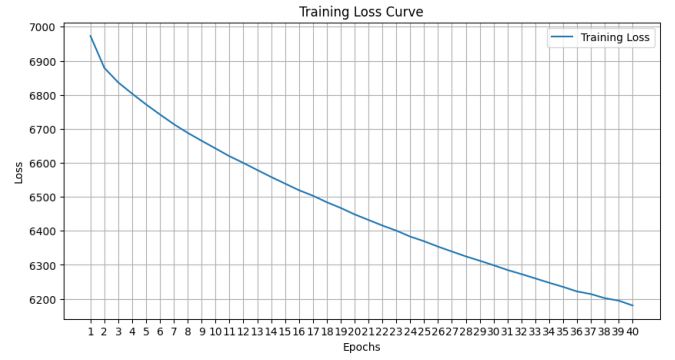


Fig. 14. Augmented data, loss plot

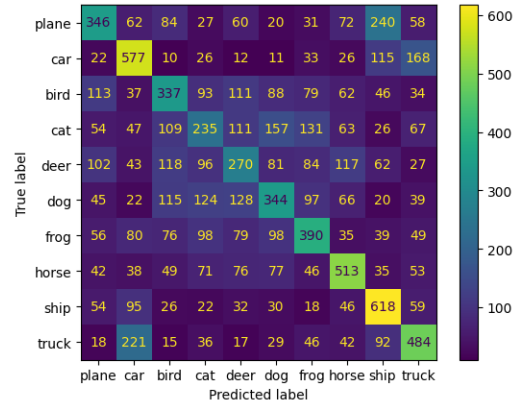


Fig. 15. Augmented data, confusion matrix

random noise and real data, resulting in a training process that is less effective. The network takes longer to train, and the loss is significantly higher for the augmented data, almost double that of the non-augmented data. This higher loss suggests that the model is not effectively learning from the augmented dataset and may even be forming incorrect patterns, which explains why the accuracy remains similar but doesn't improve with augmentation. In this case, the diffusion model's generated data does not contribute meaningful new information for the model to learn from.

### III. PART 3 - COSINE SCHEDULING TRAINING

The same training as in Part 1 is done, but with cosine scheduling. Below are the hyperparameters used for training:

- Batch size: 80
- T: 500 (number of steps)
- Channel: 128
- Channel-multiplier: [1, 2, 2, 2]
- Learning rate: 1e-4
- Beta 1: 1e-4
- Scheduling: cosine
- Beta T: 0.028
- Loss Function: MSE loss

Figure 1 is the loss plot of the training:

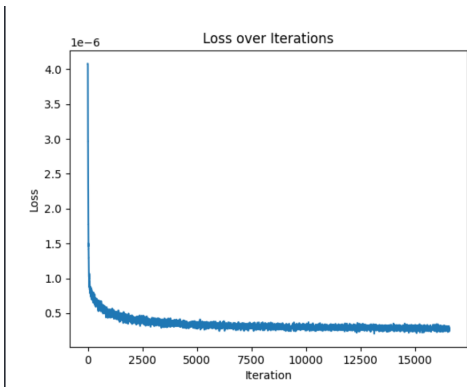
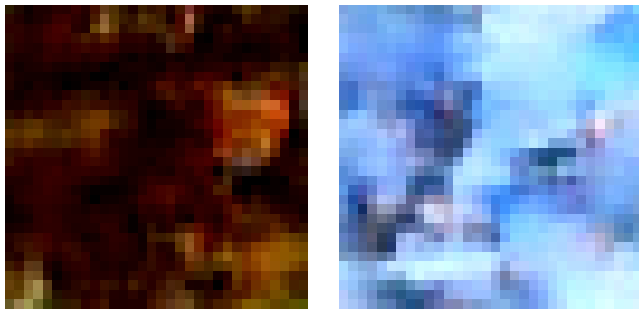


Fig. 16. Plot



(a) first

(b) second



(c) third

(d) fourth

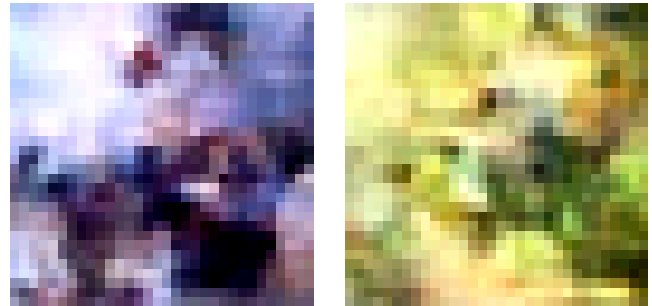
Fig. 17. Bird images, generated from the diffusion model

Figure 2 to 11 show the images for every class generated by the diffusion model.



(a) first

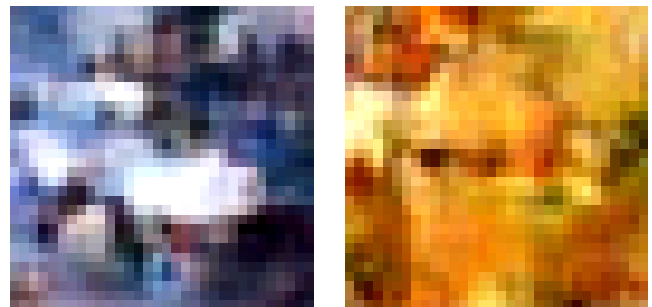
(b) second



(c) third

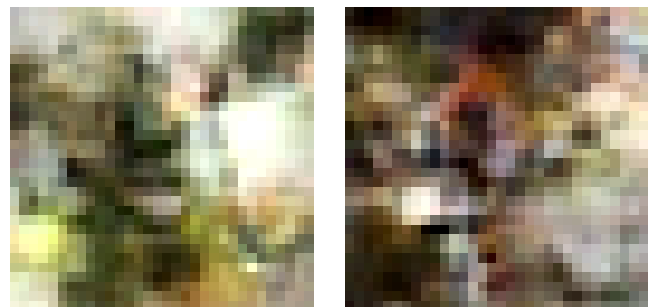
(d) fourth

Fig. 18. Car images, generated from the diffusion model



(a) first

(b) second



(c) third

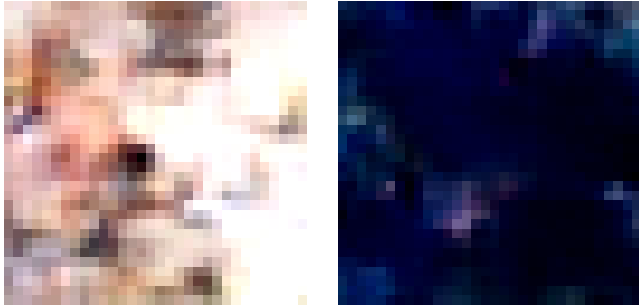
(d) fourth

Fig. 19. Cat images, generated from the diffusion model



(a) first

(b) second



(c) third

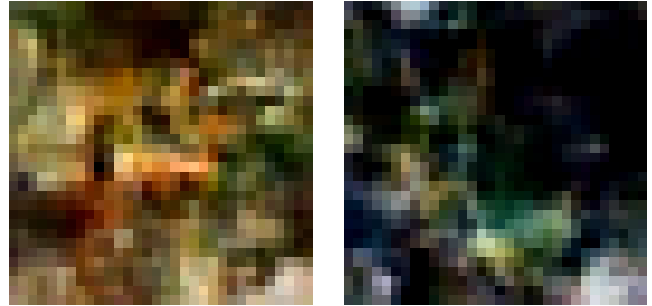
(d) fourth

Fig. 20. Deer images, generated from the diffusion model



(a) first

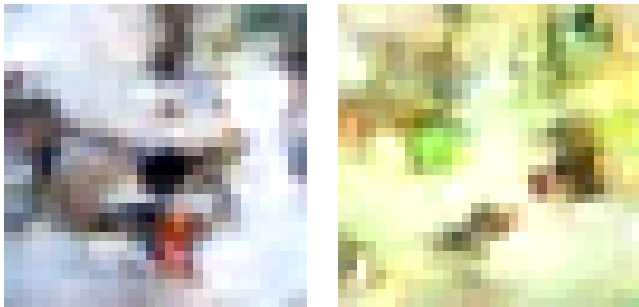
(b) second



(c) third

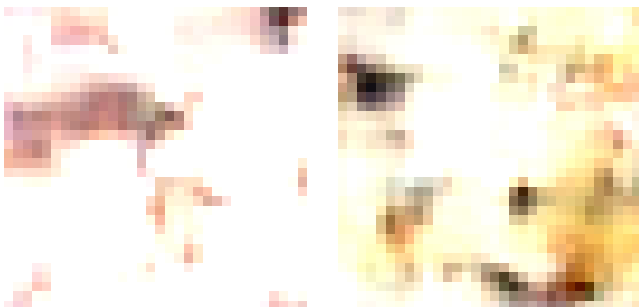
(d) fourth

Fig. 22. Frog images, generated from the diffusion model



(a) first

(b) second



(c) third

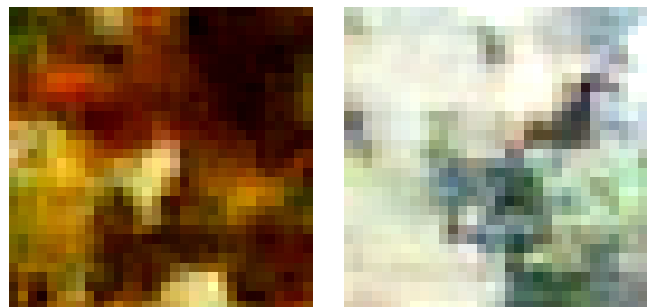
(d) fourth

Fig. 21. Dog images, generated from the diffusion model



(a) first

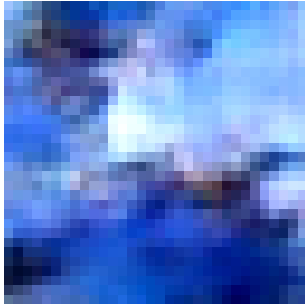
(b) second



(c) third

(d) fourth

Fig. 23. Horse images, generated from the diffusion model



(a) first



(b) second

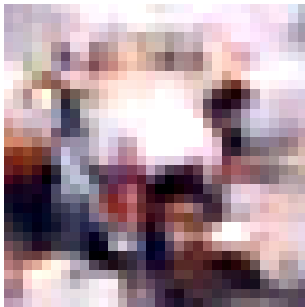


(c) third

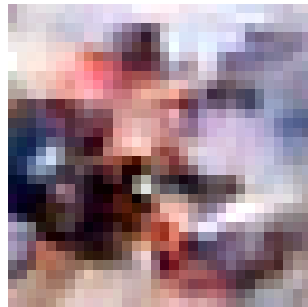


(d) fourth

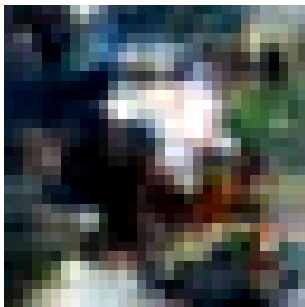
Fig. 24. Ship images, generated from the diffusion model



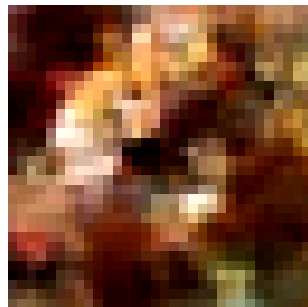
(a) first



(b) second

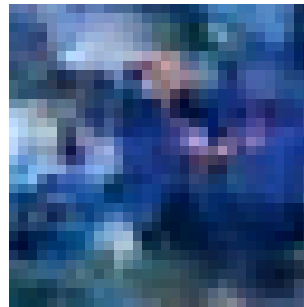


(c) third



(d) fourth

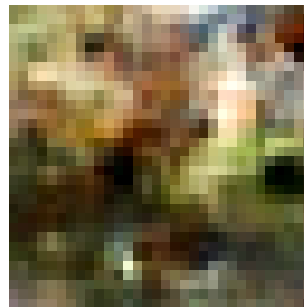
Fig. 25. Truck images, generated from the diffusion model



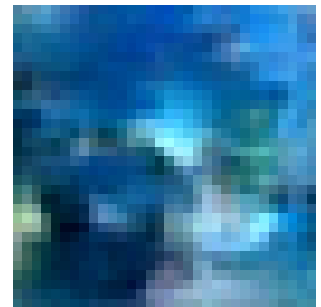
(a) first



(b) second



(c) third



(d) fourth

Fig. 26. Plane images, generated from the diffusion model