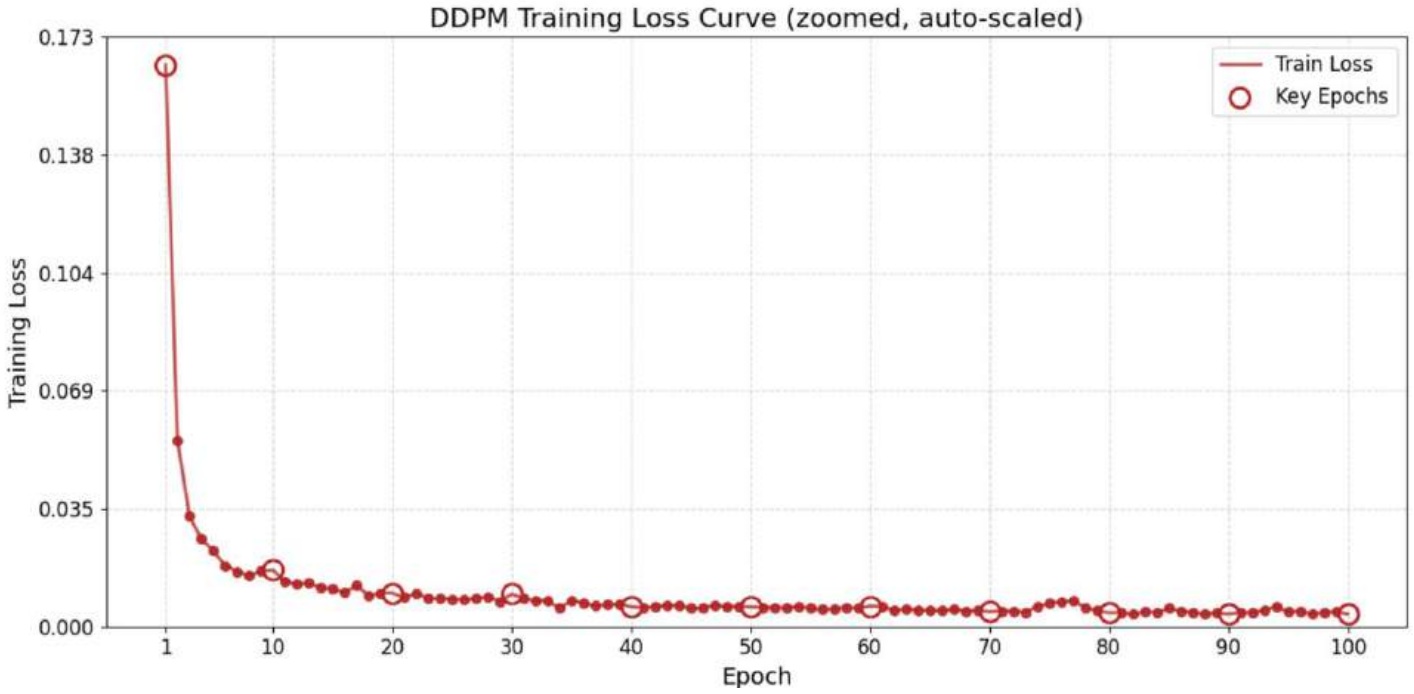
Deep Learning Lab

Semantic segmentation, image captioning and image-text retrieval

Gurmeher Singh Puri
Matriculation Number: 5776434

1.1 Training loss curve



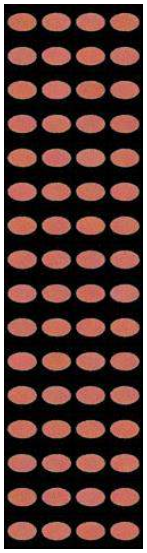
The model's training loss starts at 0.173 in epoch 1, then drops sharply to 0.0557 in epoch 2 and 0.0342 in epoch 3. By epoch 50, the loss continues declining to 0.0056, and finally reaches 0.0048 by the end of training at epoch 100, showing a consistent pattern of decreasing loss throughout the entire training process.

1.2 Generated images from seen conditions ["000","001","100","010"] with probe accuracy

All the images are saved in the main directory of the code as well

Accuracy for config 000: 1.0

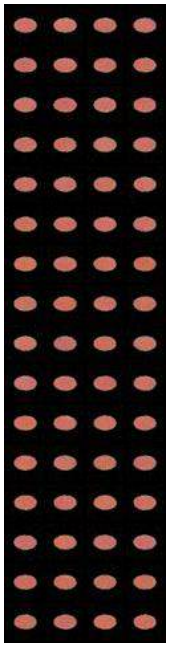
Generated_image_000



Mean_Image_000

Accuracy for config 001: 1.0

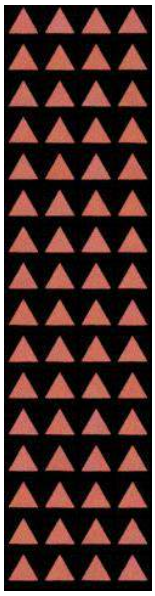
Generated_image_001



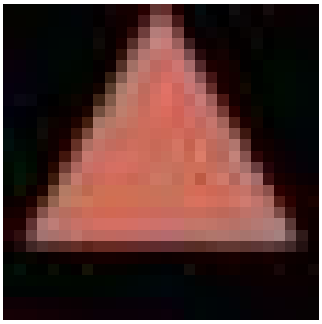
Mean_Image_001

Accuracy for config 100: 1.0

Generated_image_100

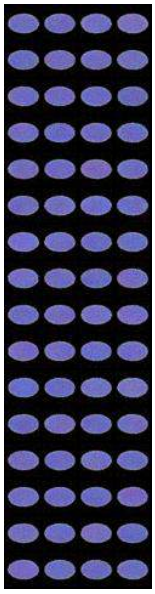


Mean_Image_100

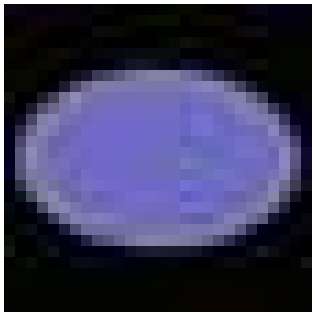


Accuracy for config 010: 1.0

Generated_image_010



Mean_Image_010

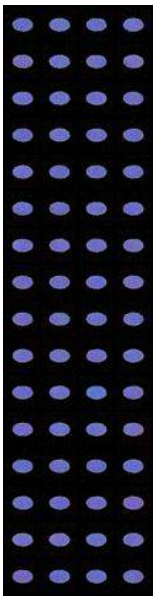


1.3 Generated images from unseen conditions ["011","110","101","111"] with probe accuracy

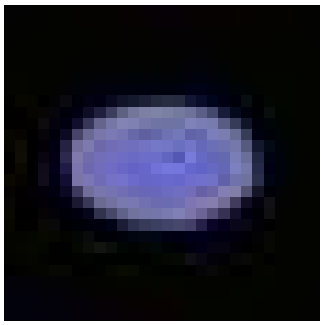
All the images are saved in the main directory of the code as well

Accuracy for config 011: 1.0

Unseen_Generated_image_011

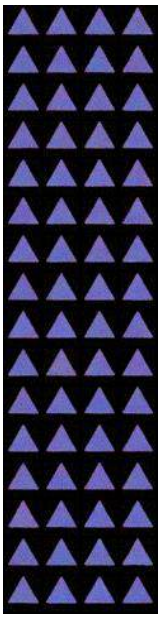


Unseen_Mean_Image_011

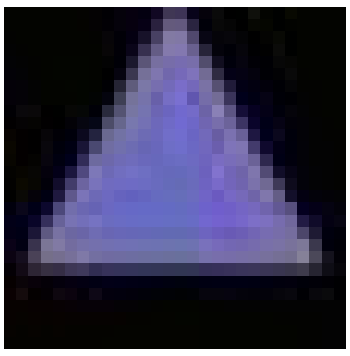


Accuracy for config 110: 1.0

Unseen_Generated_image_110

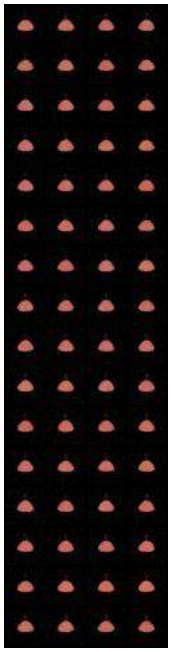


Unseen_Mean_Image_110

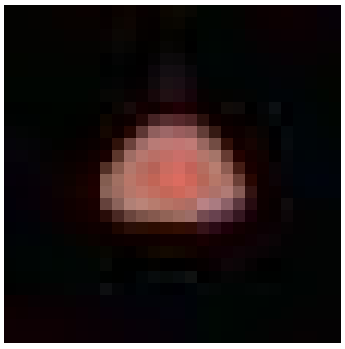


Accuracy for config 101: 1.0

Unseen_Generated_image_101

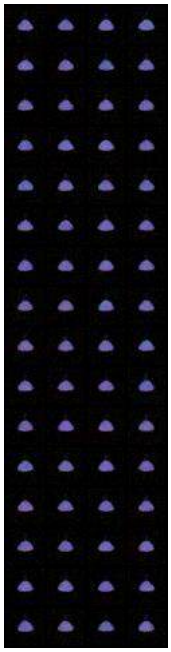


Unseen_Mean_Image_101

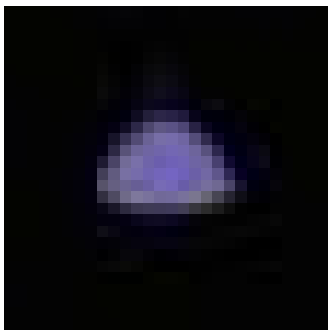


Accuracy for config 111: 1.0

Unseen_Generated_image_111



Unseen_Mean_Image_111



2.0 Image-Text

2.1 Image Captioning

2.1.1 Why do the results improve with lower temperature?

The temperature parameter is used to divide the raw predicted score; it controls the randomness of the prediction made by the model. Lower temperature scores indicate higher confidence in the prediction, hence improving the results. However, a lower temperature score might also lead to generic captions, as the model may stick to common phrases. Therefore, we should choose an ideal temperature score to balance the trade-off between diversity of captions and prediction accuracy.

2.1.2 Prompt Engineering

Prompt	BLEU score
a picture of	12.90%
This is a	9.09%
This image depicts	9.69%

Table 1: Experiment with different prompts

The prompt A picture of gives the highest BLEU score of 12.90%, other prompts are giving slightly lower BLEU score comparatively.

2.1.3 Student Hyperparameter search

Top K	Temperature	Prompt	BLEU score
50	0.7	a picture of	12.90%
50	1.0	a picture of	6.56%
10	0.7	a picture of	12.88%
70	0.7	a picture of	12.68%
100	0.7	a picture of	12.59%
100	1.0	a picture of	6.31%
False	-	a picture of	27.95%

Table 2: Experiment with different hyperparameters

When the TopK is false, it is greedy configuration; the score achieved is 27.95%
The highest score is achieved with TopK 50 which is 12.90%

2.2 Image-Text Retrieval

2.2.1 Finetuning

When finetuning with hyperparameters: learning rate = 1e-5, weight decay = 0 and training for 3 epochs gives 58.38% as the BLEU score, which is higher than the training-from-scratch (lr = 1e-3, weight decay = 1e-3, 5 epochs and temperature = 0.1) BLEU score of 45.13%.

2.2.2 Student Hyperparameter search

Learning Rate	Weight Decay	Epochs	Temperature	BLEU Score
1e-5	1e-3	3	0.1	59.48
1e-5	0	3	0.1	59.35
1e-2	1e-3	3	0.1	37.68
1e-3	1e-2	3	0.1	55.41
1e-2	1e-3	3	0.7	26.84

Table 3: Experiment with different hyperparameters