Assignment 4

CMSC 691 — Computer Vision

Faisal Rasheed Khan

VB02734

[vb02734@umbc.edu](mailto:vb02734@umbc.edu)

**Question 1-----------------------------------------------------**

Task 1

A screen shot of a computer program

Description automatically generated

A collage of a boat and an object

Description automatically generated

A screen shot of a computer program

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

Task 2

A screenshot of a computer

Description automatically generated

Task 3

1. The learning rate 0.001 is good learning rate, so tuned with various epochs and batch size increasing and the at epoch 30 and batch size 25 got the best accuracy of 65%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Epochs** | **Batch Size** | **Learning Rate** | **Accuracy** |
| **1.** | 2 | 4 | 0.001 | 54 |
| **2.** | 4 | 4 | 0.001 | 59 |
| **3.** | 2 | 6 | 0.001 | 53 |
| **4.** | 4 | 6 | 0.001 | 61 |
| **5.** | 4 | 10 | 0.001 | 54 |
| **6.** | 10 | 50 | 0.001 | 55 |
| **7.** | 10 | 50 | 0.01 | 63 |
| **8.** | 30 | 25 | 0.01 | 57 |
| **9.** | 30 | 25 | 0.001 | 65 |
| **10.** | 50 | 30 | 0.001 | 61 |
| **11.** | 15 | 20 | 0.001 | 63 |

Task 4

1. The better compared to the normal without augmentation is A computer code with text

   Description automatically generated

A screenshot of a computer

Description automatically generated

The accuracy I got is 55% which is just 1% better where the normal setting gave 54%. The rotation is 2 in magnitude, horizontal flip, random affine is 0 degree and 1% magnitude.

I have also tried for the rotation of 5,7 with the same affine of degree 0 and 10% magnitude, the 5 rotation gave 51% and 7 rotation gave 48%.

After some tries with better transformations I got better by 1% with 2 rotation and the accuracy would also be increased with increasing epochs

1. The performance improved by 1%
2. The data augmentation provides the diverse range of the training instances to the model to be trained.

The data augmentation helps in improving accuracy for the unseen data by those transformations.

The data augmentation help in avoiding overfitting by just having nice images, instead with data augmentation helps in testing the challenge images

1. Adding noise to the image

Resizing the image

Data Augmenting

Testing on the ensemble methods such that the training images learns various models

Task 5

1. The classifiers prediction will be of one of the classes of 10 CIFAR classes based on their trained weights.
2. The confidence for the unseen class might be low compared to the seen classes, as the model weights are trained on the seen weights. The model predicts for the unseen images using the trained weights of the seen classes.
3. The computer vision models to overcome these failures are to have transfer learning, ensemble learning, Data Augmentation.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

**Question 2-----------------------------------------------------**

Task 1

A screenshot of a computer program

Description automatically generated

Task 2

A screenshot of a computer code

Description automatically generated

A screenshot of a computer program

Description automatically generated

Task 3

A screen shot of a computer

Description automatically generated

A group of firefighters standing next to a car

Description automatically generated

A computer screen shot of a computer code

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A group of firefighters standing next to a car accident

Description automatically generated

Model 0:

A person standing on a sidewalk with a dog and a car

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A group of people standing on a street with a dog and a bus

Description automatically generated

A black and white image of a person and a dog

Description automatically generated

Task 4

A computer screen shot of a program code

Description automatically generated

A screen shot of a computer program

Description automatically generated

A close up of numbers

Description automatically generated

A black text with numbers

Description automatically generated with medium confidence

A screenshot of a computer code

Description automatically generated

A number and numbers on a white background

Description automatically generated

**Question 3-----------------------------------------------------**

**3.1-----------------------------------------------------**

1. Summarize the talk in a couple of paragraphs. You can write about the big-picture topic of the talk and specific projects (tasks, methods, results, etc.) (4 points)

The talk focuses on the evaluation metrics for Text-to-image models and its efficiency over the years. The main problem for these tasks are to integrate are multi-linguality, generative tasks and evaluation metrics. The author Micheal Sexon introduced the evaluation metrics T2IScoreScore which addresses the issue while evaluating the text-to-image for the improvements which lacked earlier. The evaluation performance are very important for the text-to-image tasks, the talk presented the technical comparisons between models and its metrics such as TS2. The author mentions the weakness of the TS2 metrics. With the proposed TSIScoreScore provides good evaluation compared to the previous methods which do not see those mistakes. TS2 aims for the development of better T2I prompt faithfulness metrics through improvements.

1. Questions/Discussions: If you asked a question, please summarize that question and the answer from the speaker. If you didnt ask a question, please summarize any question that was asked and the answer from the speaker. (3 points)

The question I asked was how important the evaluation metric is for the Text-to-image tasks, as the text give the output?

The answer I got was the improvement of the prompts to get the better image, the speaker gave comparison of the mistakes which was shown with the previous metrics and shown the improvements with their TS2 metric.

1. What was your favorite part of the talk / discussion?

My favorite part of the talk / discussion was the improvement of the prompts to generate the images based on the texts. And the details presented while showing the examples. The results shown ging the improved results was very nice to see. The critical examinations of benchmarks and metrics

1. Cite one of the speaker’s published paper that they discussed in detail during the talk. (2 points) (1 points)

Saxon, M., Jahara, F., Khoshnoodi, M., Lu, Y., Sharma, A., & Wang, W. Y. (2024). Who Evaluates the Evaluations? Objectively Scoring Text-to-Image Prompt Coherence Metrics with T2IScoreScore (TS2). arXiv preprint arXiv:2404.04251. <https://arxiv.org/pdf/2404.04251>

**3.2-----------------------------------------------------**

1. Summarize the talk in a couple of paragraphs. You can write about the big-picture topic of the talk and specific projects (tasks, methods, results, etc.) (4 points)

The talk focuses on the effective Scene Composing of the image by performing the image synthesis. The speaker Y. Zeng proposes the novel approach for generating the image. The text-to-image prompt is converted to the semantic-to-image to generate semantic syntesis of image called SceneComposer. Traditional methods focus on the low level details but the SceneComposer aims to focus on the semantic description to match the described objects to fine the details of the image. They map the text details into the 2D canvas to have the appropriate details in order to get the fine results. The results based on the evaluation metric T2I and S2I, both of them gave good results when they applied SceneComposee which is better than the Text-to-Image

1. Questions/Discussions: If you asked a question, please summarize that question and the answer from the speaker. If you didnt ask a question, please summarize any question that was asked and the answer from the speaker. (3 points)

The question I asked was the chance of handling the synthesis of hypothetical scenarios?

The answer I got was SceneComposer is designed to have flexibility in synthesizing semantic images at any level. The model is trained to handled on different types of data and can be handled based on the text provided and each has its own trained images and then they are combined based on the SceneComposer effectively.

1. What was your favorite part of the talk / discussion?

My favorite part of the talk / discussion was the approach of solving the image synthesis generating text to 2D Canvas and then generating the images and with this the images can be generated accurately which are compares with the metrics compared to the other techniques.

1. Cite one of the speaker’s published paper that they discussed in detail during the talk. (2 points) (1 points)

Y. Zeng, Z. Lin, J. Zhang, Q. Liu, J. Collomosse, J. Kuen, V. Patel, “Scenecomposer: Any-level semantic image synthesis,” CVPR, 2023 (Hightlight, top 2.5% ). https://arxiv.org/pdf/2403.09632

References:

[1] [COCO - Common Objects in Context (cocodataset.org)](https://cocodataset.org/#home)

[2] [Training a Classifier — PyTorch Tutorials 2.3.0+cu121 documentation](https://pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html#sphx-glr-beginner-blitz-cifar10-tutorial-py)

[3] [CMSC491/691 HW4 Starter Code and Instructions - Colab (google.com)](https://colab.research.google.com/drive/11OVrERQq2LDnlxffy9ODKnSKZKv3YKxi?usp=sharing)

[4] Saxon, M., Jahara, F., Khoshnoodi, M., Lu, Y., Sharma, A., & Wang, W. Y. (2024). Who Evaluates the Evaluations? Objectively Scoring Text-to-Image Prompt Coherence Metrics with T2IScoreScore (TS2). arXiv preprint arXiv:2404.04251. <https://arxiv.org/pdf/2404.04251>

[5] Y. Zeng, Z. Lin, J. Zhang, Q. Liu, J. Collomosse, J. Kuen, V. Patel, “Scenecomposer: Any-level semantic image synthesis,” CVPR, 2023 (Hightlight, top 2.5% ). <https://arxiv.org/pdf/2403.09632>

[6] [Matplotlib — Visualization with Python](https://matplotlib.org/)

[7] [Python Documentation contents — Python 3.12.2 documentation](https://docs.python.org/3/contents.html)

[8] [NumPy –](https://numpy.org/)

[9] [OpenCV - Open Computer Vision Library](https://opencv.org/)