North South University

Department of Electrical and Computer Engineering

School of Engineering and Physical Sciences (SEPS)

Directed Research

Text-to-Image Using Contrastive Learning on Paraphrasing Captions

Umnoon Binta Ali
1713013042
Department of Electrical and Computer Engineering
North South University

Faculty Advisor
Dr. Nabeel Mohammed
Associate Professor
North South University

DECLARATION

We, hereby, declare that the work presented in this report is the outcome of our four months' work performed under the supervision of Dr. Nabeel Mohammed, Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh. The work was spread over a span of one of the final year courses, CSE498R, Directed Research, in accordance with the course curriculum of the Department for the Bachelor of Science in Electrical and Electronics Engineering program.
Students' name & signature:

_____

Umnoon Binta Ali

APPROVAL

The Directed Research report on 'Text-to-Image Using Contrastive Learning on Paraphrasing Captions' has been submitted by Umnoon Binta Ali (ID #1713013042), student of the Department of Electrical and Computer Engineering, North South University, Bangladesh. This report partially fulfills the requirement for the degree of Bachelor of Science in Electrical and Electronics Engineering in December, 2021 and has been accepted as satisfactory.

Supervisor's Signature
Dr. Nabeel Mohammed
Associate Professor
Department of Electrical and Computer Engineering
North South University, Dhaka, Bangladesh

Department Chair's Signature
Dr. Mohammad Rezaul Bari
Associate Professor and Chair
Department of Electrical and Computer Engineering North
South University, Dhaka, Bangladesh

ACKNOWLEDGEMENT

First, we express our gratefulness to almighty ALLAH for His blessing which made it possible to complete the project. We are grateful and wish our profound indebtedness to **Dr. Nabeel Mohammed**, Associate Professor, Department of ECE, North South University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Pattern Recognition" to carry out this project. His guidance, constant supervision, enthusiastic encouragement, sagacious advice and an effective surveillance throughout the entire period of the project have made it possible to complete this project. We would like to thank our entire CSE 498R course mate in North South University, who took part in this discussion while completing the course work. At last, we must express our sincere heartfelt gratitude to all the staff members of the Computer Engineering Department who helped us directly or indirectly during this course of work.

Abstract

Text-to-image synthesis aims to generate a photorealistic image that corresponds to a given text description. In actuality, the captions added by people for the same image vary greatly in terms of word choice and content. The language disparity in the captions of similar photos causes synthetic images to deviate from reality. Modern state-of-the-art methods rely on a small number of captions. To address this problem, we present a contrastive learning strategy for enhancing the quality and semantic consistency of synthetic images by paraphrasing the captions. For paraphrasing the captions a large BART seq2seq model fine-tuned is used which is trained on 3 paraphrase datasets. In the pre-training phase, we employ the contrastive learning strategy to learn consistent textual representations for captions that match to the same image. In addition, in the subsequent stage of GAN training, we apply contrastive learning to improve the consistency between the images generated from captions associated with the same image. Four experiments are conducted to evaluate the best outcome.

Table of Content

List of Figures

List of Tables