UNIVERSITY OF SCIENCE
ADVANCED PROGRAM IN COMPUTER SCIENCE

HUỲNH LÂM HẢI ĐĂNG - HỒ THỊ NGỌC PHƯỢNG

ENHANCING VIDEO SUMMARIZATION WITH CONTEXT AWARENESS

BACHELOR OF SCIENCE IN COMPUTER SCIENCE

HO CHI MINH CITY, 2023

UNIVERSITY OF SCIENCE

ADVANCED PROGRAM IN COMPUTER SCIENCE

HUỲNH LÂM HẢI ĐĂNG 19125003

HỒ THỊ NGỌC PHƯỢNG 19125014

ENHANCING VIDEO SUMMARIZATION WITH CONTEXT AWARENESS

BACHELOR OF SCIENCE IN COMPUTER SCIENCE

THESIS ADVISOR

ASSOC. PROF. TRẦN MINH TRIẾT

DR. LÊ TRUNG NGHĨA

HO CHI MINH CITY, 2023

COMMENT OF THESIS'S ADVISOR

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

COMMENTS OF THESIS'S REVIEWER

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

. . . . . .

ACKNOWLEDGEMENTS

Authors

Huỳnh Lâm Hải Đăng & Hồ Thị Ngọc Phượng

THESIS PROPOSAL

Thesis title: ENHANCING VIDEO SUMMARIZATION WITH CONTEXT AWARENESS

Advisors: Assoc.Prof. Trần Minh Triết, Dr. Lê Trung Nghĩa

Duration: January 1st, 2023 to August 14th, 2023

Students: Huỳnh Lâm Hải Đăng (19125003) - Hồ Thị Ngọc Phượng (19125014)

Theme of Thesis: theoretical research, proposed improvements.

Content:

We aim to propose a novel approach for improving video summarization quality by integrating context awareness. We also aim to propose an evaluation metric that better suits the practical use of problem in real life.

The details include:

- Literature Review and Proposal Writing
  - Conduct a comprehensive literature review on video summarization, identifying the current state-of-the-art techniques and their limitations, as well as opportunities for improvement.
  - Analyze the importance of context in video summarization and compare existing methods and tools for context extraction in videos, in terms of performance and applicability for video summarization.
  - Develop a research proposal, including research questions, hypothesis, and methodology, based on the findings from the literature review.

- Dataset Collection
  - Collect datasets suitable for training and testing.
  - Analyze the current evaluation metrics for video summarization and identify their flaws.
  - Define relevant performance metrics for evaluating the effectiveness of the context awareness in improving the quality of video summarization.
- Model Development
  - Develop baseline model for the sake of benchmarking.
  - Develop different models to prove the proposed hypothesis.
  - Train and optimize the model using the collected datasets.
- Comparison with Existing Video Summarization Techniques
  - Conduct experiments on proposed enhancements with a thoroughly designed ablation study.
  - Analyze the strengths and weaknesses of the proposed approach.
  - Conduct surveys based on the proposed evaluation metric.
- Demo Application Development
  - Develop a demo application that can demonstrate the functionality and usability of the proposed framework for video summarization.

- Thesis Writing and Submission
  - Write up the thesis, including an introduction, literature review, methodology, results, discussion, and conclusion.
  - Submit the thesis for review and evaluation by the thesis committee.

Implementation plan:

- Literature Review and Proposal Writing: 01-01-2023 to 31-01-2023
- Dataset Collection: 01-02-2023 to 15-02-2023
- Saliency Detection Model Development: 16-02-2023 to 15-03-2023
- Video Summarization Model Development: 16-03-2023 to 15-04-2023
- Integration of Saliency Detection into Video Summarization: 16-04-2023 to 15-05-2023
- Comparison with Existing Video Summarization Techniques: 16-05-2023 to 31-05-2023
- Demo Application Development: 01-06-2023 to 30-06-2023
- Thesis Writing and Submission: 01-07-2023 to 31-07-2023

	December 26th 2022
Advisors	**Authors**
Assoc. Prof. Trần Minh Triết	Huỳnh Lâm Hải Đăng
Dr. Lê Trung Nghĩa	Hồ Thị Ngọc Phượng

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Video summarization is an emerging research field that addresses the need for efficient video browsing and retrieval in today's vast and ever-expanding video collections. With the exponential growth of multimedia data, the ability to effectively analyze and extract relevant information from video content has become crucial. Video summarization techniques aim to automatically generate a concise and meaningful representation of a video by selecting key frames, shots, or segments that capture the essence of the content. This process can significantly reduce the time and effort required to review and analyze video data, thereby improving the efficiency and accuracy of various applications, including video surveillance, education, entertainment, and social media.

Despite the wide-ranging usage of video summarization, there are only a few datasets available for this task, with the two most prominent are SumMe[1] and TVSum[2]. This limitation hinders the comprehensive evaluation and benchmarking of video summarization algorithms. The scarcity of diverse and representative datasets restricts the generalizability and effectiveness of developed techniques. Additionally, the evaluation metrics employed for video summarization are also flawed, as they fail to fully capture the inherent challenges and complexities involved in generating high-quality video summaries. This inadequacy hampers the accurate assessment of different algorithms and inhibits the advancement of the field.

However, the inherent nature of the video summarization task poses challenges in evaluating the quality of generated summaries without human involvement. It is difficult to determine objectively whether one video summary is superior to another without relying on subjective human judgment. Recognizing this limitation, we propose a self-supervised model that mitigates the issues associated with the data-intensive nature of video summarization. By moving away from fixed ground truth annotations and instead leveraging the inherent structure

and information within the video data itself, our self-supervised model learns to generate informative and representative summaries.

In addition to addressing the data scarcity challenge, we also introduce an innovative evaluation pipeline specifically tailored for the video summarization task. To ensure that our generated summaries effectively capture the essence of the original videos, we conduct a comprehensive survey involving human participants. The survey participants are provided with the original videos, ground truth summaries, and our generated summaries. They are then asked to evaluate and compare the informativeness of the generated summaries against the ground truth summaries. This human-centric evaluation approach enables us to obtain valuable insights into the performance and effectiveness of our proposed video summarization techniques.

By proposing a self-supervised model and an evaluation pipeline that incorporates human judgment, this thesis not only addresses the data scarcity and evaluation challenges but also provides a more realistic and meaningful assessment of the video summarization task. The experimental results and feedback obtained from the survey validate the efficacy and relevance of our proposed approaches, highlighting their potential for improving the accuracy and reliability of video summarization in practical applications.

CHAPTER 1
INTRODUCTION

This chapter describes the motivation, objectives, and content of this thesis. 1 5 1.4 1.2 1.3

1.1 Overview

This is overview.

1.2 Motivation

1 5 1.4 1.2 1.3

1.3 Objectives

This is section:intro-objectives.

1.4 Thesis Content

This is content.

CHAPTER 2

RELATED WORK

This chapter describes the related work.

2.1 Background

This is content.

2.2 Supervised approaches

This is motivation.

2.3 Unsupervised approaches

This is overview.

2.4 Other approaches

This is section:intro-objectives.

CHAPTER 3

PROPOSED METHOD

This chapter describes the proposed method.

3.1 Self-Supervised Pipeline for Summarization Learning

This is content.

3.2 Clustering-based Video Partitioning and Summarization

This is motivation.

3.3 Summarization Evaluation with Human Feedback

This is section:intro-objectives.

CHAPTER 4

EXPERIMENTS

This chapter describes the experiments conducted to evaluate the proposed method.

4.1 Datasets

This is content.

4.2 Evaluation methods

This is motivation.

4.3 Implementation details

This is section:intro-objectives.

4.4 Experimental results

This is overview.

4.5 Discussion

This is overview.

CHAPTER 5

CONCLUSIONS

This chapter concludes the thesis.

5.1 Future Directions

This is content.

5.2 Final Conclusions

This is motivation.

REFERENCES

[1] M. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool, "Creating summaries from user videos," in ECCV, 2014.

[2] Y. Song, J. Vallmitjana, A. Stent, and A. Jaimes, "Tvsum: Summarizing web videos using titles," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

APPENDICES