HACETTEPE UNIVERSITY
COMPUTER ENGINEERING DEPARTMENT

UNDERGRADUATE PROJECT PROPOSAL

Project Name	**Proposed Term**
WEB PAGE COMPLEXITY STUDY	☐ Fall ☐ Spring ■ Fall + Spring

Student Number(s)	**Student Name(s)**
21627543 21626901 21627802	Ece OMURTAY Deniz Ece AKTAŞ Ömer Bilal YAY
Supervisor(s)	**Expertise Area(s)**
Murat AYDOS	■ Software ☐ Hardware ■ Computer Science ☐ Other _____
Technical Complexity	**Research Dimension**
☐ Very Low ☐ Low ☐ Medium ■ High ☐ Very High	☐ Very Low ☐ Low ☐ Medium ■ High ☐ Very High

Owner(s) of intellectual property	**Does the Project require ethics approval?**
☐ Student(s) ■ Supervisor(s) ☐ Company	☐ Yes ■ No
Company Representative	**Is the Project supported by a formal body?**

Company Name: Contact Name: Contact Email:	☐ Yes ■ No If yes, name of body: _____

Project Coordinator	**Proposal Approval**
Date: _____	☐ Yes ☐ No If no, rational of rejection: _____

A. PROJECT VISION

I. PROJECT SUMMARY

Classification of the visual complexity level of web pages through CNN based learning methods.

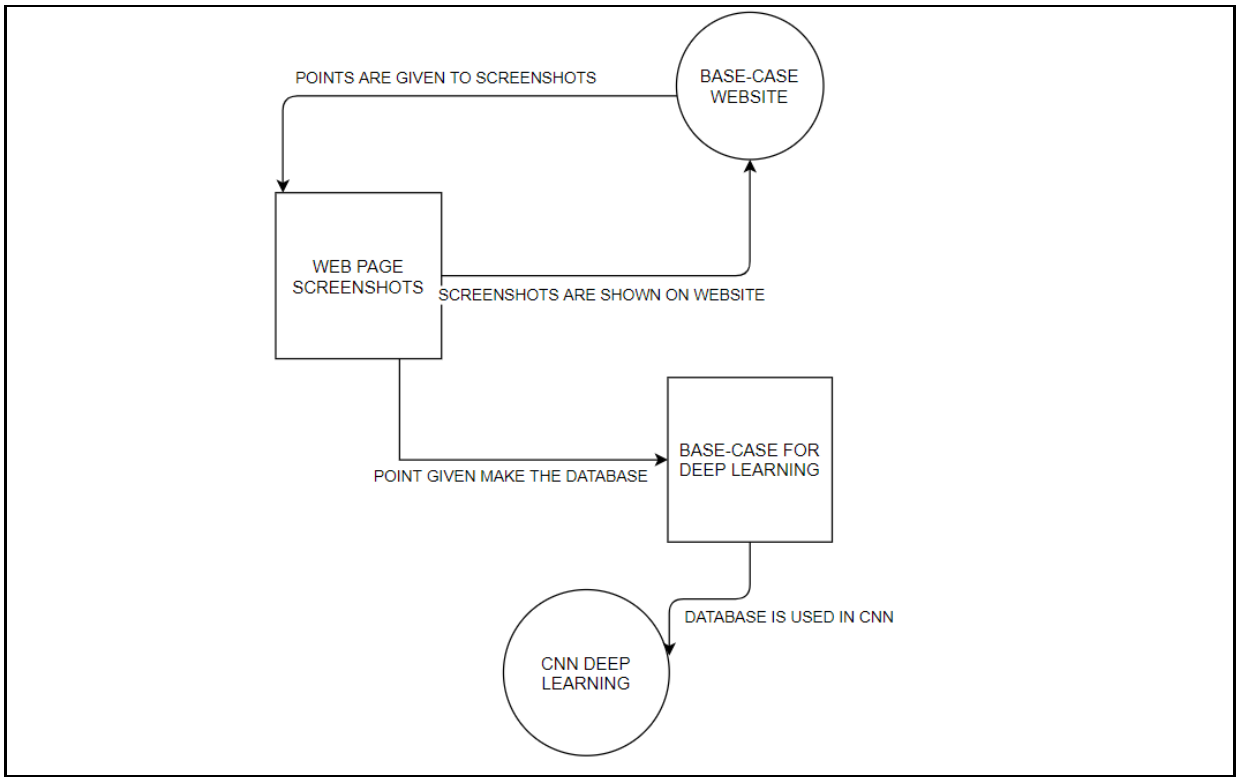
The visual appearance of web pages affects the way a user will interact with the web page contents. The layout which shapes the visual characteristics of overall appearance is composed of various visual components such as texts, images, form elements, and white spaces. Moreover, the placement and visual presentation of these components influence the perceived visual complexity. At this point, the literature suggests that the visual complexity of a web page impacts the cognitive load and effort of the users when they interact with web pages.

This project investigates the use of convolution neural network-based deep learning methods for understanding the visual complexity level of the web pages. To do this, we are going to first of gather a base case by making a website and gathering user scores that provide an insight of how visually complex a website is from the screenshots we will provide. Afterwards we will sector the base case and make a database and then we will find the feature vectors of websites and optimize the vectors and finally we are planning to work with different machine learning methods to classify the level of complexity of web pages based on dimensionally reduced vectors.

Keywords: CNN, Visual Complexity, Machine Learning

II. PRODUCT FEATURES

[For our project first of all we will make a website to collect users insight of the level of complexities of websites that we have screenshots of. We are planning to make this website using ASP.NET and it is shown as base-case website in the context diagram down below. In this website screenshots of the various websites we have will be shown and users will give points according to the complexity level afterwards these points will be used to create a database that will be used as a base-case for the machine learning methods and deep learning methods like CNN, we will implement.



III. SUMMARY OF STATE-OF-THE-ART

There are many studies on visual complexities and aesthetics of web sites and their impacts on user's reactions to these components and these research papers and results have been used to make better websites. There are even some websites like Wix that provide aesthetic and visual choices for websites designs.

Our supervisors provided us with some research papers to read and the papers are; visual complexity and aesthetic perception of web pages[1], layout-based computation of web page similarity ranks[2], visual aesthetics of e-commerce websites an eye tracking approach[3], brand recognition of Phishing web pages via Global Image Descriptors[4].

IV. INNOVATIVE ASPECTS

There are already works and researches done about visual complexities of websites but the difference of our study is that we will find the level of complexity of web pages specifically based on dimensionally reduced vectors using machine learning and deep learning.

V. POTENTIAL CONTRIBUTION(S) TO INDUSTRY AND ECONOMY

The potential contribution to the industry and the economy of our study is that if we could find the potential complexity levels of websites, in the future while designing a website the balance of complex and simple design could be found more easily and so that the user of the website would have a better experience with the website so that it would be used more frequently. This could affect the economy especially if the result are used on shopping and retail websites.

VI. REQUIRED DISCRETE MATHEMATICS, PROBABILITY AND STATISTIC BACKGROUND

After getting the dataset, to divide the dataset into label we are planning to use mean, median and mode of the dataset and take the dataset labels accordingly, also to collect the dataset we used math to keep track of how many results for data we had. We are also after doing the machine learning algorithms compare and decide which

algorithm gave the best results and for that we are planning to use mean, median, mode and probability as well.

VII. TECHNOLOGIES TO DEVELOP/USE AND UNIQUE ACHIEVEMENTS

We are planning to make a website to get our base case from users for this purpose we are going to use asp.net to develop our website and to gather the database we are planning to use SQL. Afterwards to implement methods like CNN we are planning to use python programming language.

VIII. METHOD TO FOLLOW

Our project investigates the use of convolution neural network-based deep learning methods for understanding the visual complexity level of the web pages. To do this, we are given a novel web page corpus involving 400 web pages screenshots throughout different sectors. During questionnaire part of the project, screenshots will be scored by 30 different people, then the scores will be divided into 5 ranges such as “very basic”, “basic”, “moderate”, “complex” and “very complex”. Datasets will be labeled according to these labels. We will work on VGG and Resnet types of supervised CNNs with obtained labels in order to extract feature vectors. Afterwards we are planning to apply a supervised technique named UMAP to reduce the dimensions in a supervised manner. We will do these steps to get better results from CNN. And, finally we are planning to work with various machine learning methods to classify the level of complexity of web pages based on dimensionally reduced vectors.

IX. REFERENCES

[1]. Michailidou, Eleni & Harper, Simon & Bechhofer, Sean. (2008). Visual complexity and aesthetic perception of web pages. Proceedings of the 26th Annual ACM International Conference on Design of Communication. 215-224. 10.1145/1456536.1456581.
[2]. Bozkir, Ahmet & Sezer, Ebru. (2017). Layout-based computation of web page similarity ranks. International Journal of Human-Computer Studies. 110. 10.1016/j.ijhcs.2017.10.008.
[3]. Pappas, Ilias & Sharma, Kshitij & Mikalef, Patrick & Giannakos, Michail. (2018). Visual Aesthetics of E-Commerce Websites: An Eye-Tracking Approach. 10.24251/HICSS.2018.035.
[4]. Eroglu, Esra & Bozkir, Ahmet & Aydos, Murat. (2019). Brand Recognition of Phishing Web Pages via Global Image Descriptors. European Journal of Science and Technology. 436-443. 10.31590/ejosat.638397.

B. PROJECT PLAN

I. PROJECT GOALS

The goals that we are trying to archive at the end of the project are understanding machine learning and deep learning algorithms and their implementations, get an understanding of computer vision and web design. Also to understand the correlation between the level of complexity of the website, how the complexity is received and computer's understanding of the complexity.

II. PROJECT ORGANIZATION

Our team consists of three members: Ece OMURTAY, Deniz Ece AKTAŞ and Ömer Bilal YAY. During the project we are working collectively but for the website we will develop at the start of the project Ece Omurtay and Deniz Ece Aktaş are working with ASP.NET while Ömer Bilal Yay is focusing on SQL and database management. After we get the base-case needed we are planning to work together on machine learning and deep learning parts of the project. Because we are working remotely the team uses Whatsapp and Discord for communication. To have meeting with the supervisors we are using Zoom and to share codes, datas, files and results we are going to be using Dropbox and GitHub.

III. PROJECT MILESTONES AND OBJECTIVES

Milestone #	Primary Objective	Due Date	Project Deliverable (if any)
1.	To finish the website for collecting data for base-case and to get to get the database we will be using from the user inputs on the website	November 2020	Database that will be used in machine learning and deep learning.
2.	Project process evaluation and project process report delivery	December 2020	Project process reports will be submitted.
3.	To work on VGG and Resnet types of supervised CNNs with obtained labels in order to extract feature vectors	December 2020	Feature vector that is extracted from websites.
4.	To apply a supervised technique named UMAP to reduce the dimensions in a supervised manner	December 2020	Optimized result is acquired.
5.	Final project delivery and presentations	January 2020	Final project reports and presentations will be delivered.

IV. PROJECT PRACTICES AND MEASURES

Task #	Task Description	Responsible Team Member	Start Date	Finish Date	Success Criteria
1.	To get the base-case webpage development	Ece Omurtay Deniz Ece Aktaş	October 2020	November 2020	Having a working website.
2.	Implementing the point acquired from website into database.	Ömer Bilal Yay	October 2020	November 2020	Having a database that is divided into labels.
3.	Using the data base and implementing CNN and testing.	Ece Omurtay Deniz Ece Aktaş Ömer Bilal Yay	December 2020	January 2020	Getting the optimal results from the machine learning algorithms.

V. PROJECT BUDGET

We are going to use our own computers that we have and as for software we are planning to use Visual Studio, PyCharm and SQL but we will use the student packets of these software so we will not be doing any expenses and also because of the COVID-19 pandemic we will be working remotely so we will not be having any commuting expenses. As for income we don't have any income that is related to design project.

VI. PROJECT RISKS

Risk Item #	Description	Probability	Effect	How to handle its occurrence? (Plan-B)
1.	If the user points are given in the start of the project is given with errors or given not seriously.	Possible	Results of study would be wrong	We are planning to implement a confirmation system for points in the background of the website.
2.	The machine learning methods we will try could give not the best results.	Possible	Results wouldn't be optimal	We are planning to use more than one machine learning methods to see which one gives the best results.