

Automatic Generation of Web Page from Hand Drawn Sketches

*A literature review report submitted in partial fulfilment of the requirements for
the award of the degree of*

Bachelor of Technology

in

Computer Science & Engineering

Submitted by

Yedhin Kizhakkethara

Ranjith R.K



Federal Institute of Science And Technology (FISAT)[®]
Angamaly, Ernakulam

Affiliated to

APJ Abdul Kalam Technological University
CET Campus, Thiruvananthapuram

FEDERAL INSTITUTE OF SCIENCE AND TECHNOLOGY
(FISAT)

Mookkannoor(P.O), Angamaly-683577



CERTIFICATE

This is to certify that literature review report for the project entitled “**Automatic Generation of Web Page from Hand Drawn Sketches**” is a bonafide report of the project presented during VIIth semester (CS451 - Seminar and Project Preliminary) by **Yedhin Kizhakkethara(FIT16CS125), Ranjith R.K(FIT16CS094)**, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology (B.Tech) in Computer Science & Engineering during the academic year 2019-20.

Staff in Charge

Project Guide

Dr. Prasad J C
Head of the Department

ABSTRACT

Transforming a hand drawn sketch created by a designer into computer code is a typical task conducted by a developer in order to build customized software, websites, and mobile applications. As webpages are essential for marketing and businesses, there's a need for faster iteration, easier accessibility to non-developers and less dependency on developers for initial prototyping. A deep learning model which can convert hand drawn wireframes of websites into corresponding digital webpages can resolve the above mentioned problems.

Contribution by Author

The contributions made as a team by the members were very much valuable for completing the Literature Survey. Both members performed comparative analysis and comparisons between different machine learning models which can be used for auto generation of the digital webpage from its sketch and in turn helped in developing a methodology to be followed in the project.

Key individual contributions and roles made by each of the member is as follows :

- Yedhin Kizhakkethara : Reviewed the base paper sketch2code, and a second supporting paper pix2code.
- Ranjith R.K : Reviewed the paper Canny Text Detector: Fast and Robust Scene Text Localization Algorithm, and another supporting paper Reverse Engineering Mobile Application User Interfaces with REMAUI.

Yedhin Kizhakkethara
Ranjith R.K

ACKNOWLEDGMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along for completing of the project.

I would like to express my deepest appreciation towards Dr. George Is-sac, Principal, Federal Institute of Science and Technology [FISAT], Dr. Prasad J C, Head of Department, Computer Science and Engineering, for helping us to avail the college facilities all the time.

I owe my deepest gratitude to our project coordinators Prof. Reshmi R, Prof. Jestin Joy, Prof. Merin Cherian and Prof. Sruthy Suresh, for their guidance and constant supervision as well as for providing necessary information regarding the project.

I am thankful to all other teaching and non-teaching staff for their support in completing the project.

Last but not the least, I would also like to thank my fellow teammate for his enthusiastic support through coordination and cooperation.

Yedhin Kizhakkethara
Ranjith R.K

List of Figures	vi
------------------------	-----------

List of Tables	vii
-----------------------	------------

1 Introduction	1
1.1 Overview	1
1.1.1 Software Prototyping Model	1
1.1.2 Deep Neural Network	1
1.2 Problem Statement	1
1.3 Objective	2
2 Literature Review	3
2.1 Sketch2Code	3
2.2 Pix2Code: Generating Code from a Graphical User Interface Screenshot	3
2.3 Canny Text Detector: Fast and Robust Scene Text Localization Algorithm	4
2.4 Reverse Engineering Mobile Application User Interfaces with REMAUI	4
2.5 Comparison	4
3 Design	6
3.1 Paper-1 : Sketch2Code - Method 1	6
3.1.1 Using Classical Computer Vision Techniques	6
3.2 Paper-2 : Sketch2Code - Method 2	6
3.2.1 Using Deep Learning Segmentation	6
3.3 Paper-3 : Pix2Code	7
3.3.1 Vision Model	7
3.3.2 Language Model	7
3.3.3 Decoder	8
3.3.4 Design Model	8
3.4 Paper-4 : Canny Text Detector: Fast and Robust Scene Text Localization Algorithm	8
3.4.1 Algorithm Details	9
3.4.2 Design Model	9
3.5 Paper-5 : Reverse Engineering Mobile Application User Interfaces with REMAUI	10
3.5.1 Optical Character Recognition (step 1)	10
3.5.2 Computer Vision (step 2)	10
3.5.3 Merging (step 3)	10
3.5.4 Identify Lists (step 4)	10
3.5.5 Export (step 5)	11
3.5.6 Design Model	11
3.6 Proposal	12

3.6.1	Preprocessing	13
3.6.2	Dataset Preprocessing	13
3.6.3	Segmentation	13
3.6.4	Post Processing	13
3.6.5	Expected Outcome	13
4	Work Plan	14
4.1	WorkPlan - Phase1	14
4.2	WorkPlan - Phase2	15
4.3	Budget	16
5	Conclusion	17

List of Figures

3.1	Training and Sampling in Pix2Code [4]	8
3.2	Input and MSERs [5]	9
3.3	ERs after NMS and Localization Results [5]	9
3.4	Overview of Processing Steps [6]	11
3.5	Workflow of the model with two modes of input	12

List of Tables

WorkPlan - August	14
WorkPlan - September	14
WorkPlan - October	14
WorkPlan - November	14
WorkPlan - January	15
WorkPlan - February	15
WorkPlan - March	15

Chapter 1

Introduction

1.1 Overview

As webpages are essential for marketing and businesses, there's a need for faster iteration, easier accessibility to non-developers and less dependency on developers for initial prototyping. The project aims to create a deep learning model (DNN) which can convert hand drawn sketches created by an user into their corresponding digital webpage, for faster iteration and reduction of dependency on developers for prototyping.

1.1.1 Software Prototyping Model

Prototype methodology is defined as a Software Development model in which a prototype is built, test, and then reworked when needed until an acceptable prototype is achieved. It also creates a base to produce the final system. The project uses Rapid Throwaway Prototyping strategy, where the prototyping is quickly developed to show how the requirement will look visually. The customer's feedback helps drives changes to the requirement, and the prototype is again created until the requirement is baselined.[1]

1.1.2 Deep Neural Network

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. The DNN finds the correct mathematical manipulation to turn the input into the output, whether it be a linear relationship or a non-linear relationship. The network moves through the layers calculating the probability of each output. For example, a DNN that is trained to recognize dog breeds will go over the given image and calculate the probability that the dog in the image is a certain breed. The user can review the results and select which probabilities the network should display (above a certain threshold, etc.) and return the proposed label. Each mathematical manipulation as such is considered a layer, and complex DNN have many layers, hence the name "deep" networks.[2]

1.2 Problem Statement

As webpages are essential for marketing and businesses, there's a need for faster iteration, easier accessibility to non-developers and less dependency on developers

for initial prototyping. A deep learning model which can convert hand drawn wireframes of websites into corresponding digital webpages can resolve the mentioned problems.

1.3 Objective

- Automatic transformation of hand drawn sketches created by a designer to a web page.
- Reduce the dependency on actual web developers for initial prototyping of a webpage.
- Enable fast and easier web page prototype development for non-developers and businesses.
- Use of deep learning techniques for realtime conversion of sketches to web page for prototyping.
- Provide an intuitive User Interface and software for the designer to customize the web page prototype by making each of the elements in the webpage draggable and droppable, for ultimate freedom of design.
- When machine learning is used to power front-end development, the designers and developers get more time to focus on creative tasks. Eventually, the optimization of time, resources and design budgets **can help companies to reduce costs and increase the return on investment (ROI).**

Chapter 2

Literature Review

2.1 Sketch2Code

Sketch2Code[3] discusses the usage of two methods, a classical computer vision technique and a deep segmentation method. The paper presents approaches which automates the process of generation of website from sketch. The classical computer vision technique follows several steps which include Element Detection, Structural Detection, Container Classification and Layout Normalization. This method's image classifier was very highly discriminatory. Another highlighted result is the input class having a precision of 0.190, i.e. many false positives. This is due to container elements being misclassified as input elements indicating the need for better post detection filtering. Also segmentation using DNN can be used for Element Detection where a segmentation network groups pixels related to an object together and these pixel boundaries can be extracted as element boundaries, Element Classification where a segmentation network will label related groups and these labels correspond to classes from the training set, as such, the network classifies elements and Element Normalisation where alterations such as rotation and scaling are done.

Generally the DNN based method outperform the results of the classical computer vision model with the exception of the paragraph class. The precision score of the paragraph element performed worse in the deep learning model compared with the classical computer vision model (0.562 vs. 0.747). This indicates that the classical computer vision models paragraph detection produced fewer false positives.

2.2 Pix2Code: Generating Code from a Graphical User Interface Screenshot

Pix2Code[4] discusses about a novel approach based on Convolutional and Recurrent Neural Networks allowing the generation of computer tokens from a single GUI screenshot as input. That is, no engineered feature extraction pipeline nor expert heuristics was designed to process the input data; the model learns from the pixel values of the input image alone. Deep Learning methods such as CNN and LSTM were leveraged to train a model end-to-end so as to automatically generate code from screenshots of code.

2.3 Canny Text Detector: Fast and Robust Scene Text Localization Algorithm

The paper, Canny Text Detector: Fast and Robust Scene Text Localization Algorithm [5], presents a text detection algorithm, Canny Text Detector, which takes advantage of the similarity between image edge and text for effective text localization with improved recall rate. The algorithm makes use of double threshold and hysteresis tracking to detect texts of low confidence. As closely related edge pixels construct the structural information of an object, the paper observes that cohesive characters compose a meaningful word/sentence sharing similar properties such as spatial location, size, color, and stroke width regardless of language. However, prevalent scene text detection approaches have not fully utilized such similarity, but mostly rely on the characters classified with high confidence, leading to low recall rate. The paper exploits this similarity to quickly and robustly localize a variety of texts. Experimental results on public datasets demonstrate that the algorithm outperforms some of the state of the art scene text detection methods in terms of detection rate.

2.4 Reverse Engineering Mobile Application User Interfaces with REMAUI

When developing the user interface code of a mobile application, in practice a big gap exists between the digital conceptual drawings of graphic artists and working user interface code. Currently, programmers bridge this gap manually, by which can be cumbersome and expensive. To bridge this gap, REMAUI[6] can be used. On a given input bitmap REMAUI identifies user interface elements such as images, texts, containers, and lists, via computer vision and optical character recognition (OCR) techniques. In the experiments conducted on 488 screenshots of over 100 popular third-party Android and iOS applications, REMAUI-generated user interfaces were similar to the originals, both pixel-by-pixel and in terms of their runtime user interface hierarchies. REMAUIs average overall runtime on a standard desktop computer was 9 seconds.

2.5 Comparison

The Method using classical computer vision technique shows very high discrimination and another highlighted result is that it shows many false positives based on the data. This is due to container elements being misclassified as input elements indicating the need for better post detection filtering.

The method using DNN generally outperforms the results of the classical computer vision method. This shows that the DNN method is better efficient, less discriminatory and have lesser chance for false positives. The significantly higher

F1 score on classification of all elements except paragraphs indicates that the implementation of deep learning segmentation outperforms classical computer vision techniques at the task of element detection and classification.

On the other hand the paper pix2code only works for screenshot images, where each element position and structure is well defined, with corresponding element description within the bounding boxes. This is the major disadvantage pix2code has when comparing to sketch2code.

Reverse Engineer Mobile Application User Interfaces (REMAUI) is a technique which can be used to bridge the gap between digital conceptual drawings of graphic artists and mobile user interface code. But the model's precision suffered if a subject contained a bitmap that contained both text and non-text imagery because OCR it is hard to distinguish if a given text is plain text or belongs to a bitmap of text and other elements. another disadvantage is that low image recall occurred when images overlapped that were of similar colour or where the top image is somewhat transparent. These scenarios are challenging for edge detection.

Canny text detector presents a novel scene text detection algorithm which takes advantage of the similarity between image edge and text for effective text localization with improved recall rate. This method can be used for text detection, which can be used in the proposed model for various purposes including the detection of writings from the user input.

Chapter 3

Design

3.1 Paper-1 : Sketch2Code - Method 1

3.1.1 Using Classical Computer Vision Techniques

The baseline method which the paper primarily uses is computer vision to solve the task of converting an image of a sketch into code. This approach involves four key phases:

- Element detection - use computer vision to detect and classify the position, sizes and types of all elements from the sketch. Required for reproducing equivalent elements in HTML.
- Structural detection - produce the hierarchical tree from the list of all elements. Required for correctly reproducing the element tree in HTML.
- Container classification - classify the types of container structures e.g. headers and footers. Required as aids producing the correct structure as well as creating semantically correct and human readable code.
- Layout normalisation - correct for human errors in the sketching process such as misaligned elements. Required to produce correct element tree in HTML.

3.2 Paper-2 : Sketch2Code - Method 2

3.2.1 Using Deep Learning Segmentation

In order to overcome the limitation of the first approach of classical computer vision techniques such as being highly discriminatory and generating false positives, the paper used a second methodology which uses Deep Learning Segmentation. There are three approaches using deep learning:

- Using convolutional neural networks to categorise elements and containers. This would involve using CNNs to classify both elements and containers. State of the art networks such as Yolo or Faster R-CNN can classify and detect accurate bounding boxes in real time. However, this approach would require additional steps for normalisation and hierarchical tree detection.

- Using an ANN to learn the relationship between an image of a wireframe and the code directly. This has the benefit of being a complete solution with no post processing steps. pix2code implemented this approach in a similar domain. pix2codes approach of using a CNN and long short-term memory (LSTM) could be modified for this problem.
- Using an ANN to learn the relationship between an image of a wireframe and an image of the result. ANN can be used to translate an image of a wireframe into an image which represents the structure. The structure image would be equivalent to a normalised website. The normalised form would classify elements and containers as well as mitigating human errors. Then apply a post-processing step to translate the resulting image into code.

3.3 Paper-3 : Pix2Code

The task of generating computer code written in a given programming language from a GUI screenshot can be compared to the task of generating English textual descriptions from a scene photography. In both scenarios, the aim is to produce variable-length strings of tokens from pixel values. The paper divides this problem into three sub-problems.

- First, a computer vision problem of understanding the given scene (i.e. in this case, the GUI image) and inferring the objects present, their identities, 2 positions, and poses (i.e. buttons, labels, element containers).
- Second, a language modeling problem of understanding text (i.e. in this case, computer code) and generating syntactically and semantically correct samples.
- Finally, the last challenge is to use the solutions to both previous sub-problems by exploiting the latent variables inferred from scene understanding to generate corresponding textual descriptions (i.e. computer code rather than English) of the objects represented by these variables.

3.3.1 Vision Model

The model used a CNN to perform unsupervised feature learning by mapping an input image to a learned fixed-length vector; thus acting as an encoder. The input images are initially re-sized to 256 256 pixels (the aspect ratio is not preserved) and the pixel values are normalized before to be fed in the CNN. No further pre-processing is performed.

3.3.2 Language Model

The model was only interested in the GUI layout, the different graphical components, and their relationships; thus the actual textual value of the labels is ignored. Additionally to reducing the size of the search space, the DSL simplicity also reduces the size of the vocabulary (i.e. the total number of tokens supported by the DSL). As a result, their language model can perform token-level language

modeling with a discrete input by using one-hot encoded vectors; eliminating the need for word embedding techniques such as word2vec that can result in costly computations.

3.3.3 Decoder

This first language model is implemented as a stack of two LSTM layers with 128 cells each. The vision-encoded vector p and the language-encoded vector q_t are concatenated into a single feature vector r_t which is then fed into a second LSTM-based model decoding the representations learned by both the vision model and the language model. The decoder thus learns to model the relationship between objects present in the input GUI image and the associated tokens present in the DSL code. The decoder is implemented as a stack of two LSTM layers with 512 cells each.

3.3.4 Design Model

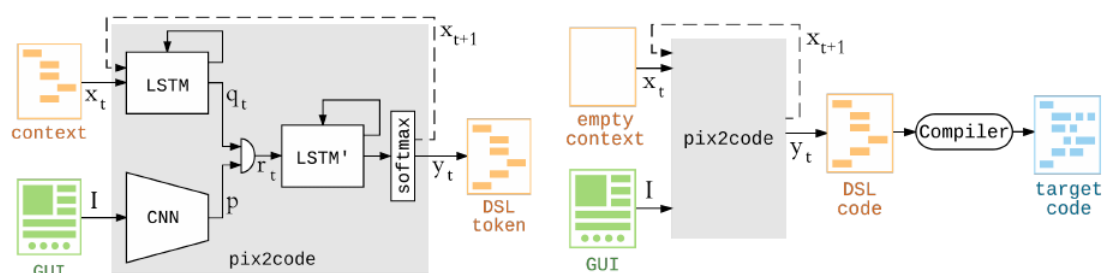


Figure 3.1: Training and Sampling in Pix2Code [4]

3.4 Paper-4 : Canny Text Detector: Fast and Robust Scene Text Localization Algorithm

As most text recognition applications require texts in images to be localized in advance, there is a significant demand for text detection algorithms that can robustly localize texts from a given scene image. To extract character candidates with better recall rate, the paper utilizes extremal regions (ERs) that are extracted with relatively weak constraints compared to those of the original MSER (Maximally Stable Extremal Regions). Overlapped candidates are reduced to a unique candidate by non-maximum suppression. Then classify the candidates with double threshold as one of strong text, weak text, and non-text. Strong text candidates are included in the final result, and weak text candidates that are connected to the strong texts are only selected by hysteresis. The surviving text candidates are grouped to compose sentence(s).

3.4.1 Algorithm Details

Character Candidate Extraction

Many of the previous approaches have adopted MSER to extract character candidates and achieved remarkable performance. However, the constraint for maximum stability is often too strong to embrace various kinds of scene text in practice. So mitigate the maximum stability constraint and employ only extremal regions (ERs) for better recall to satisfy the Recall criterion.

Non-Maximum Suppression

It is well known that MSERs have a large number of repeating components. Since ER is a superset of MSER, the initial ERs also suffer from the same problem. To guarantee the Uniqueness criterion, suppress the repeating ERs and allow only one ER that has the maximum stability.

Text Tracking by Hysteresis

Include the strong text in the final result, as they are classified with high confidence. However, the weak text can be either true text or non-text (e.g., window, leaf, and fence). So they are included if and only if they have similar properties to strong text candidates. To meet the Recall criterion with a high recall rate, start from each strong text Rs and track its neighborhood text candidates classified as weak text, Rw. Whenever Rw satisfies the similar text properties against Rs, then change the status of Rw to Rs and investigate its neighbors recursively

3.4.2 Design Model

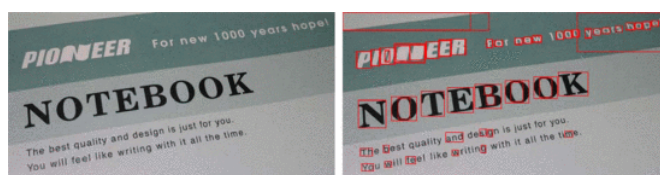


Figure 3.2: Input and MSERs [5]



Figure 3.3: ERs after NMS and Localization Results [5]

3.5 Paper-5 : Reverse Engineering Mobile Application User Interfaces with REMAUI

REMAUI automatically infers the user interface portion of the source code of a mobile application from screenshots or conceptual drawings of the user interface. On a given input bitmap REMAUI identifies user interface elements such as images, text, containers, and lists, via computer vision and optical character recognition (OCR) techniques. REMAUI further infers a suitable user interface hierarchy and exports the results as source code that is ready for compilation and execution. There are six main processing steps :-

3.5.1 Optical Character Recognition (step 1)

First, REMAUI applies on the given input bitmap off-the-shelf OCR word detection. Since optical character recognition suffers from false positives, REMAUI post-processes OCR results to remove candidate words that likely do not reflect true words in the input.

3.5.2 Computer Vision (step 2)

First detect the edges of each image element via Canny's widely used algorithm. But these edges themselves are not good candidates for atomic elements as, for example, each character or even minor noise would become its own element. To merge close-by elements with each other and with surrounding noise and to close almost-closed contours REMAUI dilates its detected edges.

3.5.3 Merging (step 3)

First, REMAUI removes OCR-detected words that conflict with vision-inferred element bounding boxes. This step addresses common OCR false positives such as classifying part of an image as a text fragment, classifying bullet points as o or a similar character, and merging lines of text that have too little spacing. REMAUI further removes OCR words that are not contained by an OCR line (using the OCR lines from step 1). REMAUI then merges OCR words and lines into text blocks.

3.5.4 Identify Lists (step 4)

In this step REMAUI identifies repeated items and summarizes them as collections, for two reasons. First, the final UI definition is more compact and efficient if each repeated resource is only represented once. Second, this step allows REMAUI to generalize from a few instances to a generic collection. REMAUI can then supply the observed instances as an example instantiation of the collection.

3.5.5 Export (step 5)

REMAUI exports all detected text content and format to Android strings.xml and styles.xml files. REMAUI exports layout files to the Android layout directory, for the layout shared between list entries and for the main screen.

- To extract character candidates with better recall rate, the paper utilizes extremal regions (ERs) that are extracted with relatively weak constraints compared to those of the original MSER (Maximally Stable Extremal Regions).
- Overlapped candidates are reduced to a unique candidate by non-maximum suppression. Then classify the candidates with double threshold as one of strong text, weak text, and non-text. Strong text candidates are included in the final result, and weak text candidates that are connected to the strong texts are only selected by hysteresis. The surviving text candidates are grouped to compose sentence(s).

3.5.6 Design Model

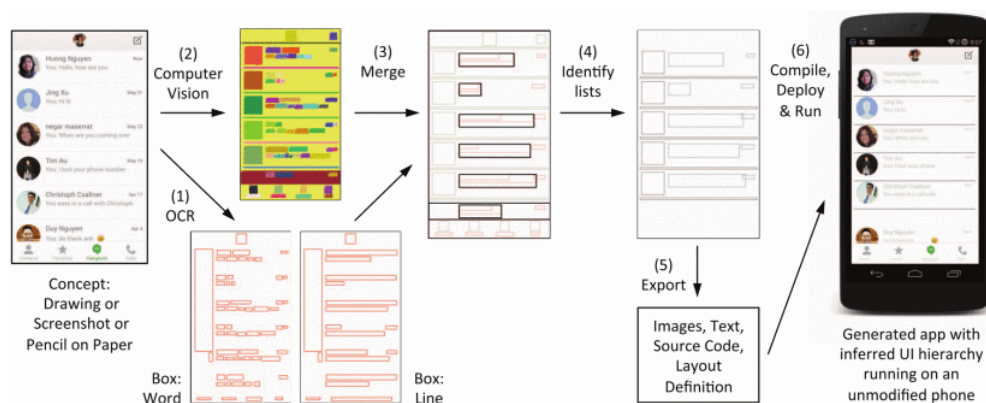


Figure 3.4: Overview of Processing Steps [6]

3.6 Proposal

A Deep Neural Network (DNN) will be used to learn the relationship between an image of a hand drawn wireframe and an image of the result. This can be utilized to translate an image of a wireframe into an image which would represent the structure of the webpage. The structure image would be equivalent to a normalised website. The normalised form would classify elements and containers as well as mitigating human errors. A post-processing step is used to translate the resulting image into code for the corresponding webpage. From the code, the model directly generates the webpage in html form. The following flowchart visualizes the above mentioned steps :-

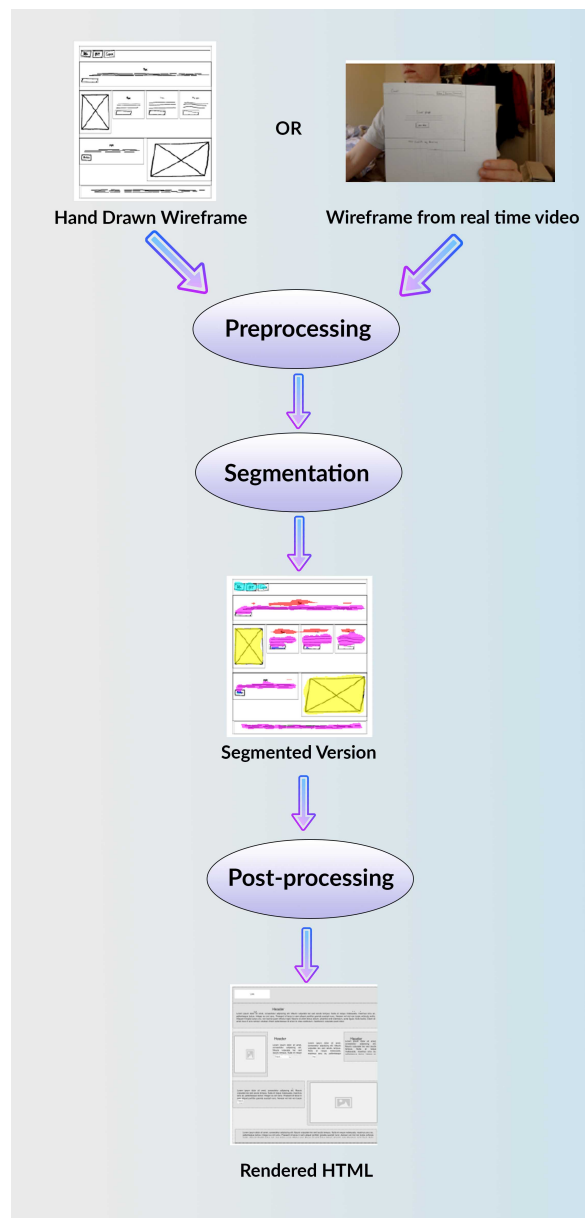


Figure 3.5: Workflow of the model with two modes of input

3.6.1 Preprocessing

The dataset contains sketches and their associated normalised version of the website. In order for the segmentation network to understand the normalised image, just convert it into a label map i.e. label each pixel with the element class it represents.

3.6.2 Dataset Preprocessing

The original website is normalised into a form where styles have been removed and maps HTML elements into wireframe element classes. The structure is then extracted by detecting the colours and using the structure inference algorithm. From the structure a sketched version of the website is created by replacing elements with a random sketched version.

3.6.3 Segmentation

An Xception model (a CNN based on Inception v3) trained on the custom dataset collected by scraping open webpages and converted to black and white 2D sketches and also the ImageNet - a large real world image database - to use as a backbone network for Deeplab v3+.

3.6.4 Post Processing

From the list of elements, the structure inference algorithm is applied to create a hierarchical tree structure. The hierarchical tree Domain Specific Language(DSL) is then fed into the post-processing phase of the framework in order to produce the HTML, which can in turn render the webpage.

3.6.5 Expected Outcome

Rapid generation of an accurate design of the webpage (in html form) from the rough hand drawn sketches.

Chapter 4

Work Plan

4.1 WorkPlan - Phase1

August	
Duration	Tasks
Week 1	Team Formation.
Week 2-4	Topic Selection and Approval..

September	
Duration	Tasks
Week 1	Review of different papers that support the topic.
Week 2	Comparison studies of the papers.
Week 3	First presentation preparation.
Week 4	First evaluation.

October	
Duration	Tasks
Week 1	Review and selection of 2 new papers that support the project topic.
Week 2	Comparison studies of the papers.
Week 3	Feasibility analysis.
Week 4	Methodology and system design.

November	
Duration	Tasks
Week 1	Report preparation.
Week 2	Final presentation preparation.
Week 3	Last evaluation of Phase-1.

4.2 WorkPlan - Phase2

January	
Duration	Tasks
Week 1	Collection of the openly available datasets of webpage and corresponding normalized sketches.
Week 2	Scraping webpages to create custom dataset and converting it to hand drawn sketches, inorder to increase dataset count.
Week 3-4	Preprocessing - Resizing sketch, Edge detection, Conversion to Label Map.

February	
Duration	Tasks
Week 1	Development of algorithms for structure inference.
Week 2	Perform Segmentation so as to produce a single channel image with pixels labelled from 0 to 10.
Week 3	Tuning a Xception model(an improved CNN based on Inception-v3) which can be used for the classification and prediction.
Week 4	Hyperparameter and Learning Rate tuning by trying out various parameter values and finding best accuracy for the model.

March	
Duration	Tasks
Week 1	Start post-processing steps : filtering of elements to avoid holes and non perfect edge problems, use container classification to produce bounding boxes for each element.
Week 2	Apply structure inference algorithm to create hierarchical tree structure, and then produce the HTML.
Week 3	Evaluation - Micro and Macro performance evaluation, Study results.
Week 4	Final report preparation.

4.3 Budget

- Consumables : 1000
- Equipments(Graphic Cards) : 6000
- Travel : 3000

Chapter 5

Conclusion

Literature Review on several papers which supports the idea of converting a hand drawn image into its corresponding digital webpage, gave the observation that a Deep Neural Segmentation based model can be the most effective way to achieve the goal. This method ensures less discrimination and lesser chance for false positives. Also a model generated using this method will have high f1-score and can be further extended to support the future scopes of this project. Thus the proposal is to use a Deep Neural Network (DNN) for generating a model which can be used to learn the relationship between an image of a hand drawn wireframe and its normalized image, and to convert the hand drawn sketch into its corresponding digital webpage.

Bibliography

- [1] "Software Prototyping - INGSOFTWARE". Retrieved 2018-06-27. www.ingsoftware.com
- [2] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2017, pp. 0588-0592. doi: 10.1109/ICCSP.2017.8286426
- [3] Alex Robinson. "Sketch2code: Generating a website from a paper mockup". In: CoRR abs/1905.13750 (2017). arXiv: 1905.13750. url: <https://arxiv.org/abs/1905.13750>
- [4] Tony Beltramelli. 2018. pix2code: Generating Code from a Graphical User Interface Screenshot. In Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '18). ACM, New York, NY, USA, Article 3, 6 pages. DOI: <https://doi.org/10.1145/3220134.3220135>
- [5] H. Cho, M. Sung, and B. Jun, Canny text detector: Fast and robust scene text localization algorithm, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Mar. 2016, pp. 3566-3573.
- [6] Tuan A. Nguyen and Christoph Csallner. 2015. Reverse engineering mobile application user interfaces with REMAUI. In Proc. 30th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 248-259.