International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 21st Jun 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-06

Title: CONVERTING UI SKETCH TO CODE USING CONVOLUTIONAL NEURAL NETWORKS

Volume 08, Issue 06, Pages: 245–247.

Paper Authors

PRAJWAL V BHARADWAJ, S.YASHAS, SRI HARI ANAND

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CONVERTING UI SKETCH TO CODE USING CONVOLUTIONAL NEURAL NETWORKS

[1]PRAJWAL V BHARADWAJ, [2]S.YASHAS, [3]SRI HARI ANAND
[1,2,3]UG Scholar, Department of ISE, SJBIT,Bangalore.

ABSTRACT

Transforming a Graphical user interface sketch into a computer code is a typical task conducted by a developer in order to build customized websites. Developing a GUI code for a given user interface is time-consuming for developers. By doing so, developers are prevented from dedicating their majority time of the work towards the logic, functionality, system verification and system validation of the software they are developing.In this paper, we demonstrate that machine learning models can be utilized to prepare a model end to end, to automatically produce code from a solitary sketch. Classification algorithms will classify the given image input into different elements and tags of HTML. Our system eliminates the tedious job of developing blueprint for every iteration of refinement in the User interface.

KEYWORDS

Machine Learning, , Convolutional neural networks, Deep Learning, Neural Networks, Long short term memory, Image Processing, Image Segmentation.

1. INTRODUCTION

In this advanced technological age, mundane job of visualizing a basic design can be done by the computers using Machine Learning, Image Processing and Program Synthesis. This will eliminate the requirement of a designer and hence save valuable resources like time and money for a company. The main objective of the proposed system is to bridge the gap between designer and developers a reduce the time between designing and development of blueprint. Writing the HTML code for blueprints is a relatively time consuming and boring task. We aim at completely automating it and ensure the developers focus on other crucial parts such as styling and program logic. The proposed system takes an image of a rough sketch of the design as an input and generates the equivalent HTML code to visualize it. When compared to the existing systems which use bulky Image Captioning models extensively, our system uses SSD (Single Shot Detector) which uses LSTM (Long Short Term Memory) to find the elements and their positions in the image. This information is then fed into a program which generates the equivalent HTML. Further in the paper we are going to explain the related work, the concepts used, proposed system and the system architecture.

2. RELATED WORK

In the current status of the automatic code generation of the graphical user interface, pix2code [1] has provided a novel method to automatically produce computer code for a given single GUI image as input. Generative Adversarial Networks (GANs), model is used for generating the computer code from the input image. The system mainly depends on components:

- Convolution Neural Network (CNN) maps the raw input image to a learned vector of fixed length.
- Recurrent Neural Network (RNN) performs language modeling on the image caption of the input image.

In [3] author introduced a hybrid system which detects the hand drawn images and classifies it between seven predefined classes. The initial detection of the image is done for separating between the hand written texts and the basic shapes. These shapes can be used for identifying different HTML tags. The different shapes are classified by passing it to K-NN classifiers. The classifiersuse all subsets of features for detecting between the geometrical shapes, lines, arrows and texts that are written. A fitting algorithm is used, for detecting the four different classes of the geometrical shapes (triangle, square, ellipse and rectangle), which executes a chain code histogram to make a format of the drawn object. The outcome of this fitting algorithm) is used for verification of mode detection and the final recognition (fit) of the object.

3. CONCEPTS USED

A. Machine Learning

Machine learning is a part of artificial. Intelligence that enables systems to learn and improve on their own, from experience without any hard-coded instructions. Machine learning aims to develop programs that can read data and learn from it. The first step of learning is observation (collecting data), like direct experience or instruction to find patterns in data for making better decisions.

There are two main categories of machine learning algorithms, namely, supervised and unsupervised. In supervised machine learning algorithms, we provide older data with the result to learn from, after it has learnt from the classified examples, it is used to predict the future examples. The algorithm analyzes from the known training dataset, then it produces an inferred function which can predict the output values. The algorithm will be able to provide prediction for any new input after it is trained sufficiently. The algorithm also compares its output with the correct prediction and automatically modifies the model accordingly.

B. Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). Image segmentation is typically used to locateobjects and boundaries (lines, curves, etc.) in images. Meaningful segmentation is the first step from low-level image processing transforming a greyscale or color image into one or more other images to high-level image description in terms of features, objects, and scenes.

C. Artificial Neural Network

Artificial NeuralNetworkisaloosemodel ofthehuman brainwhichis madeup ofmanylayersof nodescalled neurons. Thereare3typesoflayersnamely,aninputlayer, usuallymorethanoneintermediatelayerandanoutputlayer. Eachneuronreceivesweightedinputs fromnodesofthe previouslayer,processesitusinganactivationfunctionbefore passingitasaninputtothenextlayer.

Theactivationfunction introducesnon-linearity intheinputsbeforepassingthemto thenextlayer,withouttheactivation functionthepowerofa neural networkis reduced toalinearregressionmodelandis unabletorepresentreal-timeproblemswhichareusuallynon-linearinnature. Theweightsareadjustedasaprocess of learningaccording tothetrainingdata.Awelltrainedmodel canpredictwithgreataccuracy.

D. Convolution Neural Network

ConvolutionNeuralNetworkisaclassof NeuralNetwork whichcontainsatleastone convolutionlayer.Thislayer uses hierarchical patternsindataandassemblecomplexpatterns usingasmallerandsimplerpatterns. Thislayerreducesthe connectednessandcomplexityofthe neuralnetwork.

CNNis usedtoidentifytheelementthattheshaperepresents.Whena convolutionalneuralnetworksislearning,itcreatesasnapshot ofthepattertoberecognized.Thepatterithastostoreiscalled asafilter.During thetraining phasethemodel computesthe filtervaluesandrefinesthesevaluesineveryepoch.

E. Longshort-termmemory

SSDAlgorithm usesaLSTMwhichisanArtificial Recurrent NeuralNetwork tofindtheregionsoftheimage containingtheshapesandtheirposition among othershapes intheimage.LSTMunitconsists ofcellwhich remembers valuesoverarbitrary timeintervalsandthreegatesnamely, an inputgate, an outputgate andaforgetgate to regulatethe flowofinformationinandoutofthecell.Thecellremembers valuesoverarbitrary timeintervals andthethreegates regulatetheflowofinformationintoandoutofthecell.

LSTMsweredevelopedtodealwiththeexploding and vanishing gradientproblemsthatcanbeencounteredwhen trainingtraditionalRNNs.The cell keeps track of the dependencies of the input elements. The input gate is responsible for controlling the flow of new value into the cell. The forget gate tells for how long the value must be stored in a cell, while the output gate is used to tell when the stored value in the cell is to be used in computing the output activation of the unit.

F. Image Processing

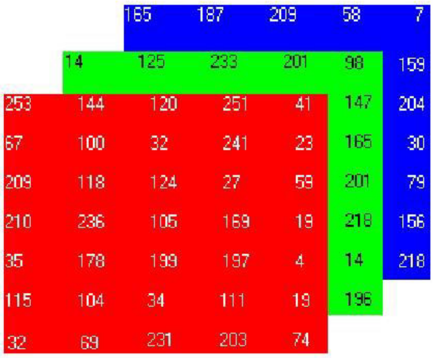
Image processing is a process of converting a picture or an image into digital format to be able perform any operation on it, like extracting useful and important features from it or for just getting an enhanced or better image. It takes any image whether taken from a video frame or a picture taken from a camera, as an input and processes it, to give a better quality image or the required features from the image. Image Processing treats images as two dimensional signals when applying a set signal to them for processing.

There are two methods used in Image Processing, namely, Analog and Digital Image Processing. The visual or analog techniques of image processing is generally used for the documents or pictures in the form of hard copies. In visual or analog image processing techniques, image analysts follow the fundamentals of interpretation. The image processing is not limited by the areas to be studied but it extends based on the amount of knowledge that the analyst has. Another important tool of visual image processing technique is the Association. Hence analysts can apply both their knowledge and the data gathered by studies in image processing.

4. PROPOSED SYSTEM

Theproposedsystem usesalight,computationally inexpensiveapproachforgenerating HTMLcodefromaUI sketch.

- Imagesareconvertedtoblack/whiteandthencropped totherequiredpixelwidth.
- Aboundedboxisdrawnaroundeachelementofthe sketch.
- Eachelementiscroppedfromthemainimageand thesecroppedimageisthenfedasaninput tothe convolutionalneuralnetwork.
- The convolutionalneuralnetworkclassifiesthe elementasdifferentHTMLtags.Thiscombinedwith theposition oftheelementisgiveasaninputtothe algorithmwhichthengeneratestheHTMLcode.



A. Preprocessing Images

Inthispaperwehaveusedexisting datasets from various sources. Wehaveuseddatasets of:pix2code,sketch2code, sketch-code,Screenshot-to-code.Wehavealsousedadataset of1600images of4differentshapes.Firstweconvert the givenimageintoablack/white imageusingthePILlibrary. Thismethodusestheluminosity leveltoconvertthe(r,g,b) values into (L)values.Themainaim ofthisstepistoreduce thecomputationalcost.Figure1.1showshowacolorimage is represented.

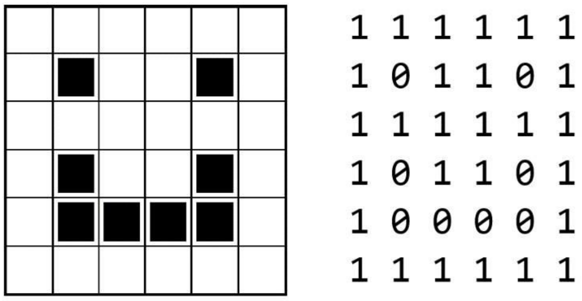


Figure 1.1 RGB Representation of a color Image

The(r,g,b)valueforeachpixelisconsideredan dthe Luminosityiscomputed.

$$L= R*299/1000+G*587/1000+B*114/1000$$

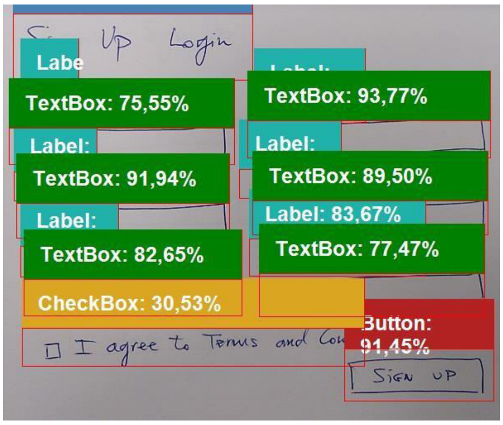
Theblack/white imagearereesizedinto255*255images forconsistency. ThisalsohelpsinoptimizingtheCNNand increases theaccuracy.Figure1.2representsanimageafter

Figure 1.2 Black/White representation of an Image

preprocessing.

B. Finding Region of Interest

Themostimportant partofthesystemistofindthe positionoftheelements.SSDalgorithm isusedtofindthe Region ofInterest. Region ofinterestistheregioninthe imagewhereanelementexists.SSDalgorithmis usedtofind theregionofinterest.SSD



usesfasterRCNNandLSTMfor multipleobjectdetection.

Therearealotofalgorithms forfindingtheregion of interest.WehavechosenSSDoverothersbecauseithasthe bestbalance between speedandaccuracy.A typical CNN networkgradually shrinksthefeaturemapsizeandincrease thedepthasitgoestothedeeperlayers.Thedeeplayerscover larger receptive fields andconstruct moreabstract representation, whiletheshallowlayerscoversmaller receptivefields.Figure1.3showstheboundingboxthatwill be visualizedbySSD model.

C. Recognizing the elements

TheSSDalgorithmgivestheco-ordinatesofthebounding boxofalltheelementsinthesketch. Theseco-ordinatescan

Figure 1.3 Bounded box

beusedtocroptheimage.Thesecroppedimagesarethensent toanotherNeuralnetwork.

ThismodelisaConvolutionalneuralnetwork. Thismodel isbuiltonthearchitecture ofVGG16.TheVGG16modelis alteredtothe requiredapplication. Few layersoftheVGG16 modelarefrozenandthenthemodelisagaintrained. The outputlayerisalteredtoclassify thedifferent HTML elements. Thismodel justoutputswhatelementispresentin thegivencroppedimage.

Theinputtocov1layeris offixedsize 224x224RGB image. Theimageispassedthrough astackofconvolutional (conv.) layers, wherethefilterswereusedwithavery small receptive field: 3×3(which isthesmallestsizetocapturethe notion ofleft/right,up/down,center). Inoneofthe configurations,italsoutilizes 1×1convolutionfilters,which canbeseenasalineartransformation oftheinputchannels (followedbynon-linearity).Theconvolutionstrideisfixedto1pixel;thespatialpaddingofconv.layerinputissuchthatthe spatialresolutionispreservedafterconvolution, i.e.the paddingis1-pixelfor3×3conv.layers.Spatialpoolingiscarriedoutbyfivemax-poolinglayers,whichfollowsomeof theconv.layers(notalltheconv.layersare followedbymax- pooling).Max-poolingisperformedovera2×2-pixel window, withstride2.

Three Fully-Connected(FC) layersfollow a stackof convolutionallayers(which hasadifferentdepthindifferent architectures).The final layer isthe soft-maxlayer.The configuration ofthefullyconnectedlayersisthesameinall networks. Allhiddenlayersareequippedwiththerectification (ReLU)non-linearity

.

D. Generating the HTML code

Thetwothingsthatarerequiredtogenerate theHTML codeistorecognizetheHTMLtagandtheposition ofthat elementintheUI sketch.TheSSDmodulegivesthe position oftheelementsintheUIsketch. Theconvolutional neural networkrecognizestheelement bytakingthecroppedimage generatedbytheSSDalgorithm.

Alltheelementsrecognizedbytheconvolutional neural networkarerepresented intermsofobjects.Theseobjects containinformation abouttheHTMLelementandthe horizontal andvertical pixelnumbers.Figure 1.5showsthe classdiagramoftheobjects.

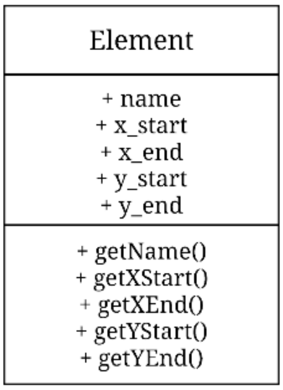


Figure 1.5 Class diagram

5. SYSTEM ARCHITECTURE

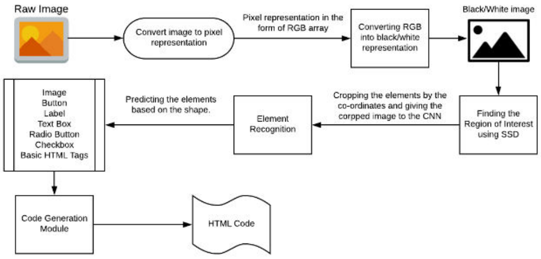

Raw Image
Convert image to pixel representation
Pixel representation in the form of RGB array
Converting RGB into black/white representation
Black/White image
Finding the Region of Interest using SSD
Cropping the elements by the co-ordinates and giving the corpped image to the CNN
Element Recognition
Predicting the elements based on the shape.
Image
Button
Label
Text Box
Radio Button
Checkbox
Basic HTML Tags
Code Generation Module
HTML Code

Figure 1.6 System Architecture

The proposedsystemhasfourmodules.Figure1.6 showthesystemarchitectureoftheproposesystem.

FirstarawimageofaUIsketch isgivenastheinputto thesystem.This image isthenconverted toBlack/White usingtheopenCV API.Thisconvertedimageisthenresized to255*255andrepresentedas a pixelarray.Sincetheimage isintheform ofBlack/White eachpixelcanberepresented byonlyoneBooleanvalueinsteadof(r,g,b)array .Thisarray representationispassedontothe nextmodule.

Thesecondmoduleisresponsibleforfinding theRegion ofInterest.SSDalgorithm isusedtodrawabounding box aroundalltheelement intheUIsketch.Thismodule gives theco-ordinatesoftheboundingbox.

Theco-ordinates ofthebounding boxareusedtocrop eachelementandthesecroppedimagesareagain passedtoa Convolutional NeuralNetwork.Thismodelthenclassifies theimagesasdifferentHTMLelements.

OncetheHTMLelementshavebeenidentified, theco-ordinatesandtheelementclassificationdataisrepresentedin theform ofanobject. Thisdataisfurthergivenasinputto anotheralgorithmwhichwill generatetheHTMLcode.

6. CONCLUSION

Theproposedsystem

isintendedtobridgethegap betweenUI/UXdesignersandfront-enddevelopers, not replaceanyofthem.Designers needtobeascreativeas possibletobetterserveendusers,anddevelopers todedicate theirtimeprogrammingthecorefunctionality andforget aboutrepetitive taskssuchasUIimplementation. With proper researchworkandfocusdeeplearningwillchange front-enddevelopment.Itwillincreaseprototypingspeedand lowerthebarrierforbuildingsoftware.

Theproposed systemcancurrentlygenerateHTMLcode forsketchesconsistingoflabels,textboxesandbuttons.This ismainlyduetotheunavailabilityof proper datasetstotrain themodelforotherHTMLelements. Withfurtherresearch andminorchangestotheexisting system, thismodelcanbe trainedto generatethecodeforanyHTMLpage.

7. REFERENCE

[1]Generating Code from a Graphical User InterfaceScreenshot. Tony Beltramelli,UIzardTechnologies,Copenhagen, Denmark 19th sept 2017.

[2]ReverseEngineering Mobile Application User Interfaces With REMAUI Tuan Anh Nguyen, Christoph Csallner ComputerScience and Engineering Department. The University of Texas at Arlington Arlington, TX 76019, USA.K. Elissa, “Title of paper if known,” unpublished.

[3] Mode detection in on-line pen drawing and handwriting recognition on Willems, Stephane Rossignol, and Louis Vuurpijl,Nijmegen Institute for Cognition and Information. Radboud University Nijmegen.

[4] Handwritten Objects Recognition Using Regularized Logistic Regression And Feedforward Neural Network , ShahamShabaniYaserNorouziMarjanFariborz , Amirkabir University of Technology Tehran, Iran.

[5]https://github.com/emilwallner/Screenshot-to-code

[6]https://github.com/harvardnlp/im2markup

[7]https://github.com/ashnkumar/sketch-code

[8]https://towardsdatascience.com/build-your-own-convolution-neural- network-in-5-mins-4217c2cf964f

[9] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. arXivpreprint arXiv:1409.0473,2014.

[10]https://github.com/emilwallner/Screenshot-to-code.