INTERPRETING PAINTINGS USING IMAGE SEMANTIC SEGMENTATION AND DECISION TREES

1

jasper haven s. briones, Ateneo de Davao University   
kennethe ann Q. mina, Ateneo de Davao University

Write you abstract here.

General Terms: Terms

Additional Key Words and Phrases:

# INTRODUCTION

## Background of the Study

Interpreting paintings is not a simple task. It requires to go beyond what our eyes have perceived. They do not look to what the artist is trying to express, but solely judging the surface of the painting. Moreover, it would be much more difficult if the painting expresses a deeper meaning which is harder to interpret for other or most people. Furthermore, each one of us has a different opinion in defining what expressions did the artist convey resulting to a possible off-track from what they want to express. Because of this, the purpose of the artist in creating their arts would be in vain and the message they are trying to pass would not reach to the people looking at their works.

It would be most feasible if there is a tool that would assist people looking to different works of art in interpreting the artist’s thoughts, feelings, and emotions. Such a tool would make the viewers of the painting have insights that they could never think about, thus having another perspective that would let them appreciate it more. Moreover, it would be easier for them to interpret these forms of art if the tool provides a list of connotations from the subject matter.

Image Semantic Segmentation is the process of understanding and recognizing an image by pixel level. It extracts features like shape, or color by dividing it into regions with boundaries in defining the objects present in an image.

Decision Tree is a diagram that branches out the possible outcomes of a certain input, which gives out a tree-like figure. This is commonly used when the factors affecting the outcomes are conditional statements. Each branch represents a statistical probability as to how the input should be interpreted.

In order to accomplish such a task, this study uses the Image Semantic Segmentation approach in getting the possible subject matters seen and depicted by the painting, and Decision Trees for weighing and choosing the best subject matter present in the artwork. Furthermore, in light to address the people such as viewers and critics interpreting artworks by artists whom have expressed their creativity, the proponents pursued this study for handing out assistance to them. Through this study, it would greatly ease their tasks in terms of time efficiency and work efficiency.

## Problem Statement

This study aims to give a list of potential subject matters from the artworks made by the artists through image semantic segmentation and weighing down the most probable subject mostly present in the art chosen through the use of decision trees. This study sought to address the following questions:

1. What are the characteristics that makes a theme unique and similar to the other?
2. What model is suitable for Semantic Segmentation in paintings?
3. How will the results of the Semantic Segmentation affect the Decision Trees?
4. How will the Decision Trees evaluate the results?

## Objectives

This study intends to extract and generate a list of possible topics from paintings that artists use to express their thoughts, emotions, and feelings through the process of image semantic segmentation and deciding the theme or subject that is mostly present from the art through the use of decision trees.

This study had the following general objective:

1. Successfully identify the subject matter of the paintings given.

This study intended to accomplish the following objectives:

1. Compare and contrast the characteristics that are unique and similar to the themes.
2. Find out which model is suitable for Semantic Segmentation in paintings.
3. Explain the relationship between Semantic Segmentation and Decision Trees.
4. Explain how will the Decision Trees evaluate the results.

## Significance of the Study

The results of this study can be used for those people who are critics in art, helping them in getting more thoughts and insights of the art they are currently evaluating. In addition, to the ones who are having trouble interpreting such art, this tool would give some assistance that would list possible interpretations besides their own, thus giving them an extended reach in knowing what the artist has to convey from their viewers of their art.

Furthermore, if there are art curators that would like to categorize paintings based on themes or its subject matter, this tool would be much of help for them. Moreover, this tool can also be used for categorizing their collection, especially if the collection they have is huge, it would be much of greater assistance.

## Scope and Limitations

The study will generate a list based on the given set of themes or subject matters. Semantic Segmentation and Decision Trees were used to interpret the paintings. This will not include abstract images in the data set to lessen the difficulty of the study. The output does not replicate the way a human would interpret the art. The generated list may also not be the actual theme the author had implied.

## Definition of Terms

   • Semantic Segmentation

   • Decision Trees

   • Characteristics

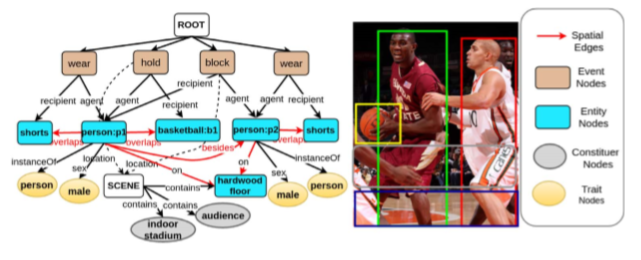
   • Connotation

# REVIEW OR RELATED LITERATURE

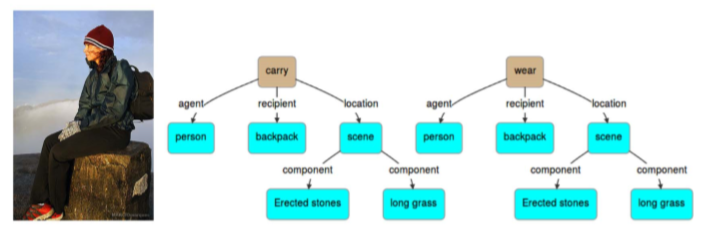
## Image Interpretation

Image interpretation is the analysis of a certain image that one must understand its context before drawing conclusions afterwards. There should be understanding of the main subject matter depicted in the image and having to derive the semantics of the interconnected relationships from its main context.

Based on a recent research, the proponents of that study faced a problem in the area of image understanding under computer vision that commonly most approaches would only describe the salient aspects of an image. It would not describe all the aspects with reasonings to back up the connections of its contents that are present. From this, they were motivated to model an architecture that is based on how humans would interpret an image. Moreover, this human perception of interpreting consists of an interaction of both visual input and language that would come afterwards, thus giving out the semantics and understanding.



**Fig. 1a** Corresponding ideal SDG encoding semantic, ontological, and spatial relations



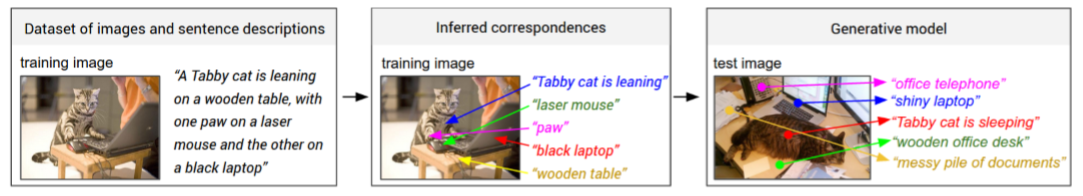
**Fig. 1b** One of the examples of an SDG

From this study, they have used the approach called Scene Description Graph (SDG). This approach is defined as a graph labeled directly representing objects, actions, regions, as well as their attributes together with the concepts inferred and semantic, ontological, and spatial relations. Furthermore, SDG depicts the semantics of a given scene, and having an integration of direct visual knowledge and background common sense knowledge. Moreover, SDG has a similar structure in comparison to semantic structure of sentences, thus having an interaction between Vision and Natural Language.

They had concluded their study to an extent where their evaluation from their generated output (sentences) is quite thorough and relevant. Their output was considered not as informative as the studies that have used existing neural approaches. Furthermore, they had ended it having said their proposed architecture of work can be used to properly elaborate the results shown and evaluate its error sources, be it from their visual detection, knowledge base or reasoning modules [1].

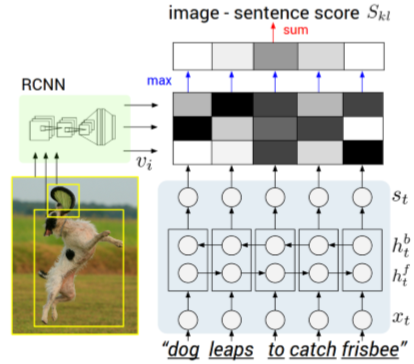
Our study would differentiate this from our method in terms of theirs uses a graph to formulate the semantics and relationships detected from their visual detection module while we would not use this kind of approach.

### Image Caption Generation

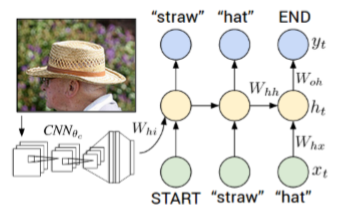


**Fig 2.** Overview of approach by the study of Karpathy et al.

Another study by Karpathy et al. introduces a problem where they recognized the remarkable ability of how humans can describe an image at first glance while existing and previous visual recognition models are having difficulty in accomplishing the same level as how humans would do. They had also mentioned that even though there are a lot of convenient models in labeling images with a fixed set of visual categories, it still has a great restriction compared to the vast and numerous descriptions that a human can think of.



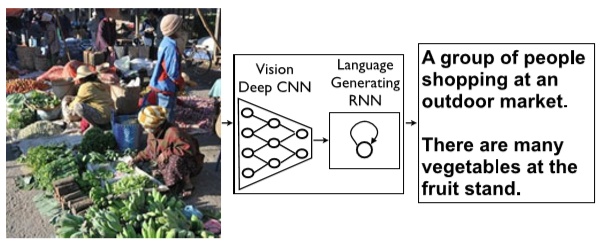
**Fig 3a.** CNN-BRNN architecture model



**Fig 3b.** Multimodal Recurrent Neural Network architecture model

From this, they were motivated to make a model in generating deep descriptions of images. They aimed to formulate a model design where it is rich enough in reasoning the contents of an image. Furthermore, in order to accomplish their goals, they have used Convolutional Neural Networks (CNN) over regions of an image, bidirectional Recurrent Neural Networks (RNN) over sentences, and a model that would align the previous two models through the use of multimodal embedding. Lastly, using those alignments from describing a Multimodal Recurrent Neural Network architecture, it learns to generate novel descriptions of image regions.

They concluded their study that it outperforms its retrieval baselines from its evaluation of performance on both full-frame and region-level experimentations [4].



**Fig 4a.** Neural Image Caption model; vision CNN followed by a language generating RNN



**Fig 4b.** Comparison of their initial model and their best model that was submitted to the competition

Moreover, there is another study also by Vinyals et al. that tackles about the fundamental problem in artificial intelligence that relates with computer vision. Having said that researches from computer visions aim to describe an image with a deep sense of semantic analysis. Their study uses deep CNN as an encoder for their image classification tasks and from having reached the last hidden layer will be then used as an input for their RNN as a decoder in generating sentences or captions from an image. They had called their approach or model as Neural Image Caption (NIC).

Furthermore, their study had quite good and accurate results from their initial model, and after having made their best model from their participation of a contest called 2015 MS COCO Challenge, they have reached first in rank from automatic and human evaluation. The researchers for this study said descriptions produced from an image through this automation process are one of many possible image interpretations, and that it is possible of having a direction where the system is able have more targeted descriptions rather than just only one stated description [5].

Of the two studies about Image Caption Generation being reviewed, they have used an integration of models between Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in generating descriptions of an image through natural language while our method would not have this kind of integration.

From this related literature, we proponents have decided to pursue an approach where we will use Image Semantic Segmentation as to extract the objects and features present in the painting image at the field of visual analytics. On the other hand, we will use Decision Trees to integrate the relationships of the features from a pre-trained dataset of known specific category of subject matters in giving out a possible list of ideas and semantics depicted from the chosen images, specifically digitized paintings.

## Approaches

### Semantic Segmentation

#### Convolutional Neural Networks

#### Fully Convolutional Network

Long *et al.* showed that fully convolutional networks (FCN) is capable of segmenting arbitrary-sized images whereas other studies that uses convolutional neural networks (CNN) failed to do so. Training the CNN end-to-end, pixels-to-pixels

#### Dense Image Labeling

Zeng *et al.* [2] used dense semantic segmentation to give a more accurate result in describing the concept of an image based on the objects and attributes present in the said image than the proposed approach by other studies such as the Joint attributes-objects Pixel-level fully-connected CRF. The said study used conditional random field (CRF) to model the relationships of the attributes and objects to label at the pixel level. A two-level hierarchical model was also used to label the objects and attributes at a region level. Results had 42% average label accuracy improvement on the object class segmentation. It was also stated that when both factors are not jointly used, such as removing pixel-level, region-level or even both of them, the accuracy reduces by 5%, 4.4%, and 10.1% respectively. This means that the attributes serve a significant role in object segmentation. The overall study had an average label-accuracy of 61.4% using the aNYU dataset.

[insert [2] image]

A similar study by Islam *et al.* [3] also had similar objectives involving dense image labeling. What differs [2] from [3] is that the latter used Deep Convolutional Neural Networks (DCNN) in segmenting the objects. The approach of the said study involves two dense labeling tasks, the semantic segmentation, which aims to label objects according to its category at a pixel level, and geometric labeling, which aims to label each pixel according to its geometrical class. The study had 5 different configurations in the experimentation. Results showed that the proposed approach outperforms all the baseline methods.

[insert [3] image]

#### Scene Recognition

### Decision Trees

#### Texton Forest

#### Random Forest based Classifiers

## Theoretical Framework

Body of Section 2 here.

# METHODOLOGY

## Section 1 (Replace the heading appropriately.)

Body of Section 1 here.

## Section 2 (Replace the heading appropriately.)

Body of Section 2 here.

# THEORETICAL BACKGROUND

## Section 1 (Replace the heading appropriately.)

Body of Section 1 here.

## Section 2 (Replace the heading appropriately.)

Body of Section 2 here.

REFERENCES

[1] S. Aditya, Y. Yanga, C. Barala, Y. Aloimonosb, C. Fermüllerb. *Image understanding using vision and reasoning through scene description graph*, 2017

[2] S. Zeng, M. Cheng, J. Warrel, P. Sturgess, V. Vineet, C. Rother, P. Torr. *Dense semantic image segmentation with objects and attributes.* In *CVPR*, 2014.

[3] A. Islam, N. Bruce, Y. Wang. *Dense image labeling using deep convolutional neural networks.*

[4] A. Karpathy, F. Li. *Deep visual-semantic alignments for generating image descriptions*, 2014.

[5] O. Vinyals, A. Toshev, S. Bengio, D. Erhan. *Show and tell: lessons learned from the 2015 MSCOCO image captioning challenge*, 2017

[6]