Dominant Color Analysis of the Prominent Object in the Images

1st Tarık Çarlı
Computer Engineering Department
Yildiz Technical University
Istanbul, Turkey
starikcarli@gmail.com

2nd Yusuf Anı
Computer Engineering Department
Yildiz Technical University
Istanbul, Turkey
yusufani8@gmail.com

3rd Doç. Dr. M. Elif Karslıgil
Computer Engineering Department
Yildiz Technical University
Istanbul, Turkey
elif@yildiz.edu.tr

***Abstract*—In the research, it was aimed to determine the dominant color of the most important object in an image. It can be used in many places in daily life in the study. It can be used to determine the colors preferred by the employees of a company in their clothes, to determine the dominant color of the clothes in a fashion show or to determine the dominant car color preference in a region.**

In the research, the intended result is obtained by applying a series of different processes to the image. First of all, the introduction image is converted into a suitable format and it is divided into sections with semantic integration method supported by deep learning. The linked component labeling method is used to detect the largest object in the segmented image. Then, only the pixel values of the largest object are sent to the color quantization function using k-mean. Finally, the pixels of the largest object whose colors are reduced are displayed to the user with the help of graphics.

The success of the transaction has been tested through a questionnaire. The output of our model for 20 images was shown to about 20 people and asked to score between 1 and 5. The success of the system is shown by taking the average of the points given by the users.

***Index Terms*—semantic segmantation , connected component labeling , color quantization , deep learning**

I. INTRODUCTION

Nowadays, the use of colors has increased much more compared to the old times. For this reason, the importance and popularity of color research projects have increased.

Determining the dominant color in a fashion show, the cars dominant color in a region in a visual content is a difficult task. In solutions where artificial neural networks are not used, the lack of background separation, which we can call noise in pictures, reduces the performance of the results.

In this project, in order to find the dominant color, the object and its location in the foreground were determined and the average colors were calculated. Basically, the project can be divided into 3 components to perform all these operations.

In the first component , A deep learning based semantic segmentation model is required for the detection of the object and background for high success rate in the results.

In the second component, in the results obtained from the segmentation model, the object or objects that are in the foreground should be identified. For this problem, the biggest object in the foreground was found by applying the connected component labeling to the results from the segmentation.

In the third component, the color of the found object is determined. To do this, using the K-means algorithm, it is provided to produce results in 4 or more colors.

II. REALTED RESEARCHS

According to latest research on the subject, some of these projects are only oriented to find the biggest object in a image. The other part of the projects aims to find the dominant color of the all image. All of the projects we encounter as a result of our research are not sufficient to achieve our goal. In our research, it was aimed to find the dominant color of the biggest foreground object in an image by combining these two projects.

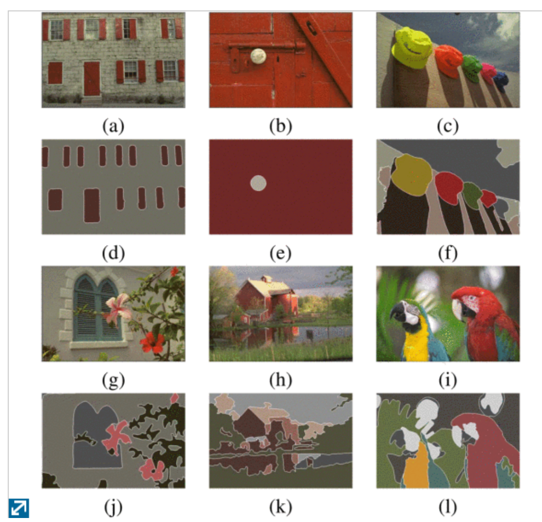

Fig. 1. The input and output of the research on image segmentation based on K-Means clustering

The success rates are low since the studies that find the biggest object in an image do not have deep learning in that times. So we decided to implement a deep learning based semantic segmentation to increase the success rate of our project.

The input and output of the research on image segmentation based on K-Means clustering [1] is given figure 1. As can be seen from the picture figure 1, this research can only detect objects. It doesn't do anything about the colors of objects.

III. Dataset

Data set selection is very important as it will affect the performance of segmentation and therefore the performance of the project. The dataset to be selected should be able to identify the basic objects and mark the parts of the background. In the Pascal VOC 2012 dataset, a total of 4 major classes and 20 of these subclasses are marked. The classes in the Pascal VOC 2012 dataset are listed below.

1) Human: Human.
2) Animals: Bird, Cat, Cow, Dog, Horse, Sheep.
3) Vehicle: Plane, Bike, Boat, Bus, Car, Motorcycles, Train.
4) Closed Area: Bottle, Chair, Dinner table, Potted plant, Sofa set, Television, Monitor.

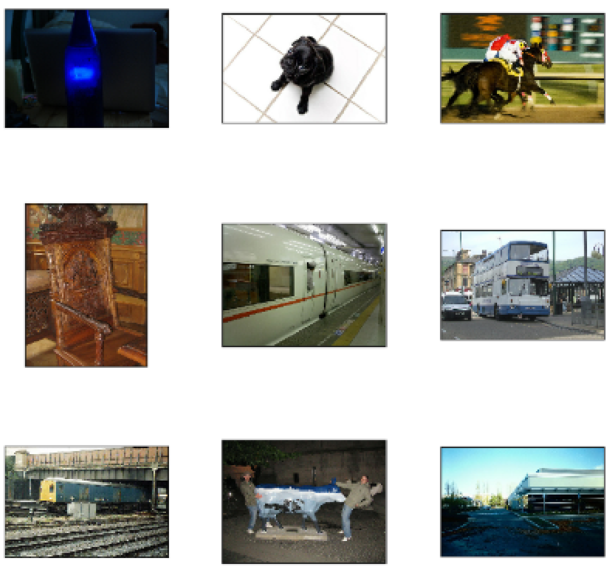

Fig. 2. Pascal VOC 2012 Image examples

Apart from these classes, the other advantage that Pascal VOC 2012 dataset offers us is class assigning as a background if any pixel in the photo cannot be associated with one of our classes. In this way, faster compound results can be applied by applying the linked labeling method to the objects of the classes that are really interested.

IV. Models Components

This section contains detailed explanation of components made under the model.

A. Segmentation

In the studies carried out, it is necessary to identify the object or objects that may be in the foreground. Thus, while finding the dominant color of the image, it is planned to increase the success by not including the background colors in the parts that the image is not interested in. To do this, it is necessary to identify the objects in the image and segment them. In recent image segmentation researchs, segmentation processes has been chosen based on deep learning because of deep learning based segmentation and success rates are higher than other methods.

There are multiple types of segmentation. Some of these are semantic segmentation, object detection, sample segmentation. The reason for choosing semantic segmentation in the study is to collect more information about the related object in the image by displaying objects of the same type in the same sections. For example, if there is more than one car in an image that contains street photography, the color of all cars has information.

		Top-1 accuracy	Top-5 accuracy
VGGNet – 1st Runner Up in ILSVRC 2014	VGG-16	0.715	0.901
ResNet – Winner in ILSVRC 2015	ResNet-152	0.770	0.933
Inception-v3 – 1st Runner Up in ILSVRC 2015	Inception V3	0.782	0.941
	Xception	**0.790**	**0.945**

ImageNet: Xception has the highest accuracy

Fig. 3. Performance of Different Architectures

There is more than one network architecture where the semantic part is applied. As can be seen from figure 3, Xception architecture has achieved higher success in the classification study on ImageNet dataset compared to others. For this, Xception was chosen as the newtork architecture.

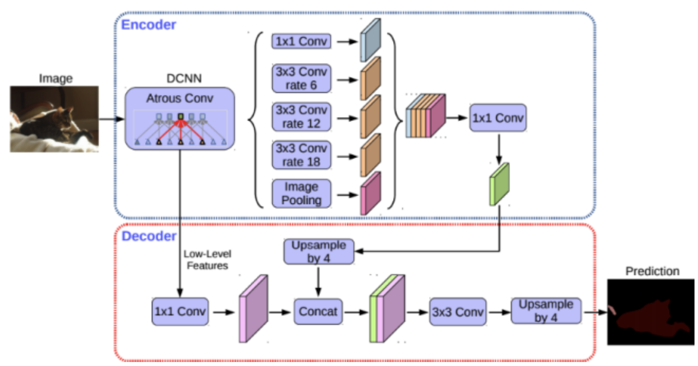

Fig. 4. DeepLabV3+ Model Architecture

Xception architecture uses Atrous convolution which is also known as dilated convolutions. It allows repurposing ImageNet pretrained Networks to extract denser feature maps by removing the downsampling operations from the last few layers and upsampling the corresponding filter kernels, equivalent to inserting holes between filter weights. With atrous convolution, we can able to control the resolution at which feature responses are computed within DCNNs without requiring to learn extra parameters. Atrous rate say r is defined as convolving the feature input with upsampled filters produced by inserting r - 1 zeros between two consecutive filter values. Standard convolution is a special case for rate r = 1. Employing larger value of atrous rate enlarges the model's field-of-view, enabling object encoding at multiple scales. Atrous convolution also allows to explicitly control how densely to compute feature responses in fully convolutional networks. Output stride is defined as the ratio of input image spatial resolution to final output resolution.

B. Connected Component labeling

Connected-component labeling is used in computer vision to detect connected regions in binary digital images, although color images and data with higher dimensionality can also be processed.

A graph, containing vertices and connecting edges, is constructed from relevant input data. The vertices contain information required by the comparison heuristic, while the edges indicate connected 'neighbors'. An algorithm traverses the graph, labeling the vertices based on the connectivity and relative values of their neighbors. Connectivity is determined by the medium; image graphs, for example, can be 4-connected neighborhood or 8-connected neighborhood.

It is assumed that the input image is a binary image, with pixels being either background or foreground and that the connected components in the foreground pixels are desired. The algorithm steps can be written as:

1) Start from the first pixel in the image. Set current label to 1.
2) If this pixel is a foreground pixel and it is not already labelled, give it the current label and add it as the first element in a queue, then go to (3). If it is a background pixel or it was already labelled, then repeat (2) for the next pixel in the image.
3) Pop out an element from the queue, and look at its neighbours (based on any type of connectivity). If a neighbour is a foreground pixel and is not already labelled, give it the current label and add it to the queue. Repeat (3) until there are no more elements in the queue.
4) Go to (2) for the next pixel in the image and increment current label by 1.

Note that the pixels are labelled before being put into the queue. The queue will only keep a pixel to check its neighbours and add them to the queue if necessary. This algorithm only needs to check the neighbours of each foreground pixel once and doesn't check the neighbours of background pixels.

C. Color Quantization

In computer graphics, color quantization is quantization applied to color spaces; it is a process that reduces the number of distinct colors used in an image, usually with the intention that the new image should be as visually similar as possible to the original image.

Fig. 5. Color quantization example

Most standard techniques treat color quantization as a problem of clustering points in three-dimensional space, where the points represent colors found in the original image and the three axes represent the three color channels. Almost any three-dimensional clustering algorithm can be applied to color quantization, and vice versa. After the clusters are located, typically the points in each cluster are averaged to obtain the representative color that all colors in that cluster are mapped to. The three color channels are usually red, green, and blue.

V. Experiment

In this section, the working version of the project will be tested using 3 pre-selected images. The first image is the dominant color, consisting of a single object and a background. In the second image, it is a multi-colored image consisting of a single object and a background. The third image is a multi-object and background image. The final output and intermediate outputs of the project are added separately for each sample image. The color histogram of the whole image has been added to the output figure for evaluation so that readers can compare results. All outputs of the image given as input are combined as a single output image so that the reviewers can easily compare. The output image consists of a combination of 8 images. Each image has a title and what it means relative to the title is explained in the list below.

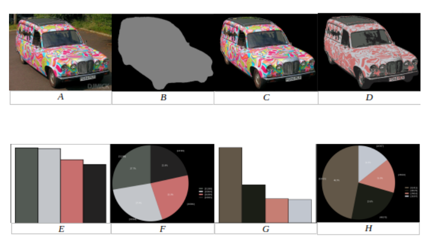

Fig. 6. Second image

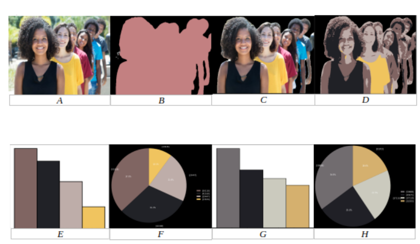

Fig. 7. Third image

- **A:** It is the input image.
- **B:** Semantic segmantation result
- **C:** Connected component labeling result.
- **D:** Largest object reduced to 4 colors using color quantization.
- **E:** Relevant colors of the largest object shown in colored bar graph.
- **F:** Relevant colors of the largest object shown in colored pie chart.
- **G:** Relevant colors of the all image shown in colored bar graph.
- **H:** Relevant colors of the all image shown in colored pie chart.

The accuracy of the outputs of the project has been tested by more than 20 people through questionnaires. The following output images of the project were shown to the survey respondents and the accuracy rate was asked to be scored between 1 and 5. (1 is very poor and 5 is excellent)

The average accuracy value of the first category in the survey is %86. Since the first category consist of one color and one object, the success of our project is high compared to the participants.

The average accuracy value of the second category in the survey is %87. Since the second images consisted of more than one color and one object, the success of our project according to the participants decreased compared to the first image. This is because our color quantization algorithm tries to reduce the colors in the largest object to 4 colors, and if the largest object has more colors, distortion occurs in the colors.

The average accuracy value of the third category in the survey is %87. If we use semantic segmentation technique and connected component labeling technique in our project, if two different objects of the same type remain connected in the image, our project perceives these two objects as one object and remains in the narrative of the desired performance. As a solution to this situation, the example segmentation technique can be used instead of semantic segmentation.

A. Evaluation

	Q -1	**Q -2**	**Q -3**	**Q -4**	**Q -5**	Q -All
Very Poor (1)	0	0	0	0	0	
Poor (2)	0	1	2	1	0	
Average (3)	4	3	3	3	2	
Average (3)	6	6	7	1	7	
Very Good (4)	10	11	8	15	11	
Average Rating	**4.3 ∼ %86.0**	**4.35 ∼ %87.0**	**4.05 ∼ %81.0**	**4.5 ∼ %90.0**	**4.45 ∼ %89.0**	**4.33 ∼ %86.6**

For performance analysis, 20 different people were given 20 different images and evaluate the dominant colors of this image. The users were asked how well the given colors reflect the true dominant colors of the picture. Very bad (1), Less bad, Average (3), Good (4), Very good (5) were given as the evaluation criteria. Being able to show in this article, first 5 images were selected and tabulated. For each question obtained sequential results are %86,%87,%81,%90,%89 and average success rate is %86.6.

TABLE I
DURATION OF COMPONENTS IN SECONDS

	Semantic Segmantation	Connected Component Labeling	Color Quantization	All
Time (second)	8.96	0.04	0.44	**9.45**

As seen in the above table, semantic segmentation takes about 9 seconds, while the other components take a total of half a second. These measurements, regardless of input / output processes, are the results when running with Intel (R) Xeon (R) CPU @ 2.30GHz processor with 12 GB of memory. Better results are expected for better hardware equipments.

VI. CONCLUSION

Our goal in this research is to detect the dominant color of the largest object in an image. The image is first divided into sections using semantic segmentation method with deep learning support. The connected component labeling method is used to find the largest object in the segmented image. Then, only the pixel values of the largest object are reduced to k colors by the color quantization with K-means method. Finally, the results we obtain are visualized and presented to the user.

Our model can be used in many areas today. For example, it can be used in cameras at a school to determine which colors the students prefer the most in their clothes, to identify the popular colors of the year in fashion shows, to identify the most preferred car colors in a country or city.

VII. FUTURE WORK

Since our model uses semantic segmentation, if the same type of bound objects are present in the image, it accepts them as one object. Experimental results As you can see in the images in category 3, in this case, the success score of our model decreases. To avoid this, instead of semantic segmentation, objects of the same type can be divided into different segments using the sample segmentation method. The second development is the K-means algorithm, if the image consists of many colors and is attempted to be quantified to less colors, it may impair the color quality of the image. Color quality can be tried to be preserved while maintaining color quality or when quantizing color with an algorithm that selects the number of k.

REFERENCES

[1] Tse-Wei Chen and Yi-Ling Chen and Shao-Yi Chien “Fast image segmentation based on K-Means clustering with histograms in HSV color space“ 2008 IEEE 10th Workshop on Multimedia Signal Processing.

[2] Everingham, M. and Eslami, S. M. A. and Van Gool, L. and Williams, C. K. I. and Winn, J. and Zisserman, A. “The Pascal Visual Object Classes Challenge: A Retrospective“ International Journal of Computer Vision.