An in-car camera system for traffic sign detection and recognition

Conference Paper · June 2017

DOI: 10.1109/IFSA-SCIS.2017.8023239

Shu-Chun Huang

Huei-Yung Lin
National Chung Cheng University

An In-Car Camera System for Traffic Sign Detection and Recognition

Shu-Chun Huang and Huei-Yung Lin
Department of Electrical Engineering
Advanced Institute of Manufacturing with High-Tech Innovation
National Chung Cheng University
168 University Road, Min-Hsiung, Chiayi 621, Taiwan

Chin-Chen Chang
Department of Computer Science and Information Engineering
National United University
2 Lienda, Miaoli 360, Taiwan

Abstract—**In the present society, driving safety becomes a very important issue. If there is an excellent driving assistance system, the possibility of a car accident can be significantly reduced. This paper presents a driving assistance system for traffic sign detection and recognition. The proposed technique consists of two subsystem for detection and recognition. First, the road sign detection subsystem adopts the color information to filter out most of irrelevant image regions. The image segmentation and hierarchical grouping are then used to select the candidate road sign region. For the road sign recognition subsystem, Convolution Neural Network (CNN) is adopted to classify the traffic signs for the candidate regions. In the experiments, the proposed technique is carried out using real scene images. The performance evaluation and analysis are provided.**

I. Introduction

In recent years, with the rapid development of science and technology, vehicles are used in various families. Compared with its applications and demand, many kinds of traffic safety questions have become more and more serious. At the same time, due to various types of sensing and positioning technologies such as laser, computer vision, GPS and other robot-related research boom, driving assistance is now a popular research topic. For unmanned vehicles and driving assistance system, the safety question is always the highest priority compared with the convenience or practicality for a project or system designer.

In the process of driving a vehicle, the driver can get a variety of messages based on the local road signs such as speed limit, double curves, slippery road, children crossing, etc. To prevent the driver ignores signs and quickly distinguishes them, traffic signs are often designed with eye-catching colors and easy-to-understand symbols. But if driving in a complex environment or the driver's mental state is not well, this might cause the driver overlooks the messages from the traffic sign. If there is an automatic detection and recognition system, it can promptly report the correct traffic signs to the driver and also reduce the burden of the driver. When the driver ignores a traffic sign, the system can give a timely warning. If this system is used in an unmanned vehicle, it can help the automatic driving system to judge the road condition so that the safety of the vehicle driving is greatly improved and the risk of accidents is reduced.

Fig. 1: Traffic sign detection and recognition system.

This paper focuses on the detection and identification of traffic signs. In addition to testing the system performance, we also experiment with a combination of different system architectures. We expect to achieve good results in the worse environment, in order to assist the driver and enhance the traffic safety. In this work, we first filter out most of the non-sign parts of the images using the color information. Then we extract the region where may have image blocks, and further extract the candidate region from the above image block. Finally, we use deep learning to verify the candidate areas of non-road signs and identify the type of the traffic sign is.

II. Related Work

In the detection of traffic signs, color and geometric information is the most basic characteristic, and commonly used for screening. In order to make the traffic signs easier to be noticed while driving, most of the warnings and restrictions on traffic signs choose to use the eye-catching red as a design feature.

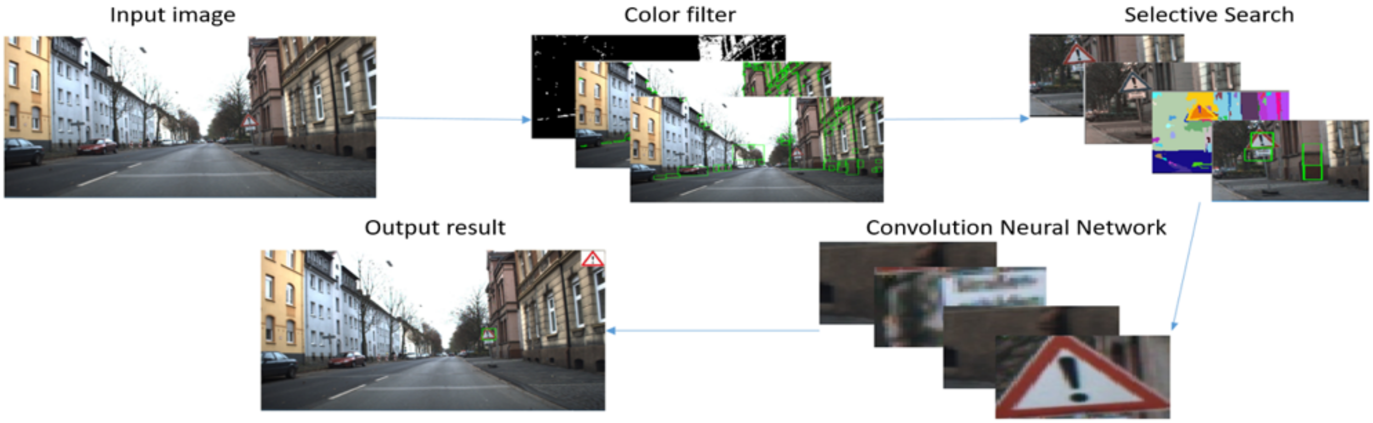


Fig. 2: The proposed system architecture. We first use of the color filter and selective search as a detection subsystem. A large number of candidate areas will be detected and then input to the convolutional neural network for the final screening and identification.

Therefore, this paper uses red as a preliminary screening feature. A commonly used color space is the HSI color space because the color is independent of a channel and less affected compared to the R, G, B on the brightness change. Escalera *et al.* proposed a method using the HSI color space [1]. They first turned the color space to the HSI color space, and then found the obvious red according to the range of hue and saturation. Similar approaches based on the HSV color space have also been applied in many works such as [2], [3], [4]. Shaposhnikov *et al.* [5] and Vitabile *et al.* [6] used the HSI color space to define the red area outside the mark and made further retrieval. For the existing work using other color spaces, Miura *et al.* [7] first converted the images to the Y, U, V color space, and then through the repeated experiments to select the red and blue range.

In terms of geometric and gradient characteristics, Broggi *et al.* [8] used pre-defined templates to align the normalized color modules. If its similarity is greater than a certain threshold, the shape can be determined. Escalera *et al.* [9] used a known angle mask and matched the pre-set search range to detect the triangular traffic signs of the three vertex or circular traffic signs on its circumference after capturing a specific color. The slope with the absolute value 1 of the 4 specific circle points is used then determine the shape and range. Belaroussi and Tarel [10], [11] abandoned the use of color information and focused on the shape features for the algorithm development.

Bahlmann *et al.* [12] used the Haar feature to deal with the color information. They use AdaBoost and Bayesian classification training to achieve the purpose of road sign detection and identification. Kuo and Lin [13] adopted Hough conversion and corner detection and other methods to detect traffic sign locations, and then used RBF neural network with K-d tree to identify the traffic signs. Greenhalgh and Mirmehdi [14] first converted the images to the HSI color space, and then used MSER to filter the circle traffic sign locations. It was then followed by capturing the characteristics of the HOG, and with the SVM classifier to perform the traffic sign identification.

Maldonado-Bascon *et al.* [15] first screened the images in the RGB color space, selected the red area with an aspect ratio to perform the restriction, and bound selected traffic signs of the candidate area. Then the DtBs feature is extracted from the region, and the SVM classifier was used to train and classify the traffic signs. Fang *et al.* [16] used the neural network as a basis and the image color and shape as features, and input to the two types of neural networks in order to achieve effect detection. The Kalman filter was then used to predict the possible position of the next frame under the traffic sign. There are also many people using deep learning to identify the road signs. [17], [18] used the convolution neural network relying on the network to iterate the appropriate weight, and designed the traffic sign identification system.

III. System Structure

Our system structure mainly focuses on the general area in the HSI color space, and then divides the candidate area into multiple small parts according to color, texture, size and region similarity. Finally, the convolution neural network is used to identify the candidate region. The system architecture shown in Fig. 2.

A. Color Information

The common color spaces include RGB, HSI, YCrCb and Lab. These color spaces all have three channels. Among them, the RGB color space consists of red, green and blue. The composition of the HSI channels include hue, saturation and brightness. The Lab is composed of brightness, interval of green to red and interval of blue to yellow. YCrCb is made of Lumen, red concentration offset (Cr) and blue concentration offset (Cb). In these color spaces, we choose HSI as our bases for color judgement. The reason we choose HSI is this color space only uses one channel to present the color interval. Its performance of three primary colors is 0 to 360 degrees. Because of the above properties, HSI is subject to minimal changes in light and shadow. The process shown in Fig. 3.

Fig. 3: Color selection process: Left: The binary image obtained according to the filter; Middle: Select the object outline box based on the binary image; Right: Save a rectangular image according to the object contour.

B. Selective Search

In general, the way to find objects from an image can be divided into two step. First, select the possible locations from the image. Second, extract features from the candidate area and recognize it with possibility calculation. Selective search is a kind of algorithm about the first step to identify the possible region. The calculation process for selective search is first to segment the image into a large number of super pixels as the initial split area. Hierarchical grouping is then used to combine the initial division areas. The merged large area is the candidate which will be used for identification in a later stage.

The existing algorithms usually perform the exhaustive search, but it can only identify the subjects in the region within its kernel. To prevent the loss of any targets, the exhausted search uses a simple but violent way to solve this problem. It uses a mask to scan the entire area of the image. Because of the uncertainty of the target object size, the exhausted search needs to repeat search for all images with different kernel size. The computation complexity of the exhausted search is very large, so its major drawback is the long processing time.

The selective search adopted in this work has three advantages. First, the selective search can identify different sizes of objects with the strategies of image segmentation and hierarchical grouping. Second, the basis to distinguish includes color, texture, size and region similarity. Rely on the parameters and weights, it can be implemented for most situations. Third, the processing speed is higher compared to the exhausted search. As a result, the selective search is able to produce a lot of candidate regions with high speed. The process shown in Fig. 5.

In addition, because the target objects have different characteristics. According to different situations, the selective search has some strategies to implement. Here, we list three different strategies as follows:

1) Because the selective search algorithm requires hierarchical merging for the initial regions, it is very important to use different initial algorithms according to the situations.
2) Use different color spaces to extract different color attributes.
3) In the case of region consolidation, we can change the estimation of region similarity with different situations.

This paper adopts graph-based image segmentation approach [19]. The algorithm uses image pixel as unit, according to the dissimilarity between the pixels, to determine whether two pixels are in the same region. The pixels are compared to the surrounding area in eight directions, and the similarity in eight directions is arranged from small to large, $e_1, e_2, \cdots, e_N$. The computation for its dissimilarity is as follows:

$$\sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2}$$

where r_i, g_i, b_i are the three channels of the color space. When the pixels make up a small area, repeat the calculation of dissimilarity so that we can spread to the entire image. In some different circumstances of the local area, the image-based segmentation does not use global thresholds, but adaptive thresholds. The adaptive thresholds are calculated by the interclass between different regions and intraclass in the same region.

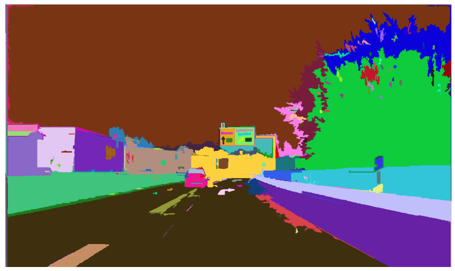
About the regional consolidation, the region similarity can be calculated by the image features. We divide the feature similarity into color similarity, texture similarity, size similarity and filling similarity. The color similarity is calculated using three color channels such as hue, saturation and brightness. The texture similarity uses Gaussian function to calculate the differentiation in each direction of all channels. In the end, we can get a texture histogram of 240-vector. The size similarity is based on the number of pixels in two regions. This is to increase the combined rate of small areas. The texture similarity identifies the border around the region. If the border size of two regions is similar and the overlap area is large, then we merge these two regions. Use the four different similarities above, we can adjust individually according to the real situations.

The selective search can produce a large number of candidate regions in a short time, but its disadvantage is also obvious. It has too many post-selected areas in the complex image, which causes the speed of deep learning decreases. In our system we have already filtered out most regions with HSI color space, so the influence of having too many candidate regions from the selective search can be reduced.

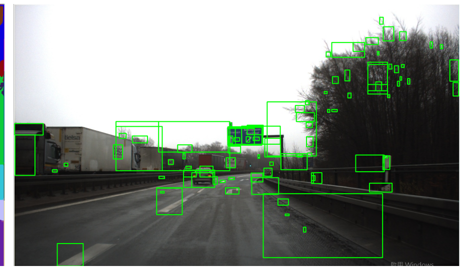
C. Convolutional Neural Network

Convolutional Neural Network (CNN) is an important technique for recognition which attracts much attention in

(a)



(b)



(c)

Fig. 4: The process diagram of the selective search. (a) The source image; (b) Image segmentation; (c) The result of the selective search.

recent years. Now, it has become one of the most important research topics in many scientific fields, especially in pattern classification. CNN is widely used because it does not require complicated preprocessing.

In general, the basic architecture of CNN contains two layers. The first layer is the feature extraction layer, and is also called a convolution layer. It consists of several convolution units. The input of each unit is connected to the local area of the previous layer and extracts the features of the local region. After extraction by multiple layers of convolution, the low-level features such as horns and edges can be enhanced to the more complex features of the structure. When the local feature is extracted, the positional relationship between features is thus determined. The second layer is the feature mapping layer, which consists of the feature map. Each feature map is a plane with the same neuron weight value. The feature map structure uses the sigmoid function as an activation function. Each convolution layer in a convolutional neural network is accompanied by a computational layer for local averaging and secondary extraction. This unique secondary feature extraction structure can reduce the feature resolution.

In this paper, we adopt two kinds of deep learning structure to train our system. One of the structure is Alexnet [20], in which the architecture mainly contains eight layers. The first five layers are convolutional layers and the latter three are all connected layers. In the last layer we set a 43-way softmax layer to connect with the full connected layer. AlexNet can be considered as an example architecture of deep convolution network. There are many network architectures such as ZF-net, SPP-net, VGG and other networks, taking AlexNet as a prototype. AlexNet's success consists of several parts: Rectified Linear Unit (ReLU), pooling, local response normalization and dropout. It not only improves the training speed and accuracy, but also reduces the over-fitting problem.

The second deep learning architecture used in the work is GoogLeNet. In the case with a large amount of data, the easiest way to improve the performance of the network is to increase the depth and width of the network. But doing this has two problems:

1) A large network generally needs more parameters, and using the fixed data, it is likely to cause the network over-fitting.
2) A large network needs more computing resources. For instance, increasing the number of convolution of the network will lead to increased the computing volume. In addition, if the increase of the network has not been effectively used such as the weight close to zero, it will cause a waste of computing resources.

To solve these problems, Arora [21] proposed the construction of inception. The main idea of inception is to find and describe the dense components of the local sparse structure in the convolutional network. It is assumed that translation invariance is established by convolutional blocks and the repetition is spatially extended. We need to do is to find the ideal local structure and expand the duplication in the space. GoogLeNet is also based on the above inception design concept.

IV. System Structure

The experimental results of this paper are divided into two parts, which are individually tested according to the detection and identification in the system.

A. Dataset

The datasets are used separately for detection and recognition in this work. We use German Traffic Sign Detection Benchmark (GTSDB) as the dataset for detection and German Traffic Sign Recognition Benchmark (GTSRB) as the dataset for recognition.

B. Traffic Sign Detection System

For the detection part, we use 300 discrete images in GTSDB dataset for testing. In addition, the system is designed for red traffic signs, so we only record red signs as our ground truth. The correctness of the test is based on whether the candidate area is selected by the selective search and can be correctly extracted with a bounding box. There are two points to check with the benchmark. First, the target area has to be selected more than 80%. Second, the area of the bounding box containing the target should not exceed 1.5 times of the target area. If one of the conditions fail, then we will consider the detection incorrect.

According to the previous part of the selective search, the algorithm can use different color spaces as its strategy in different situations. Therefore, this experiment in addition to

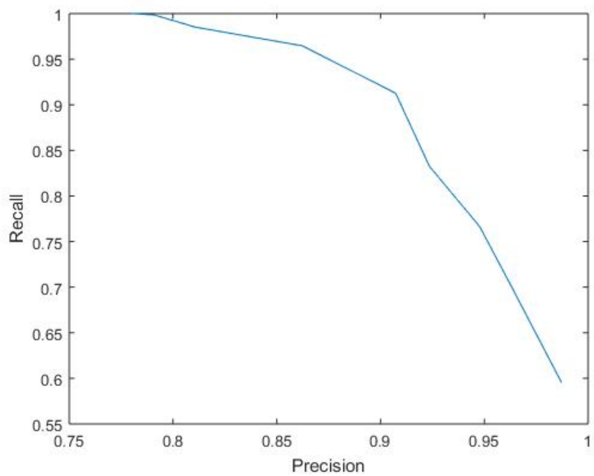


Fig. 5: Test result of GoogLeNet.

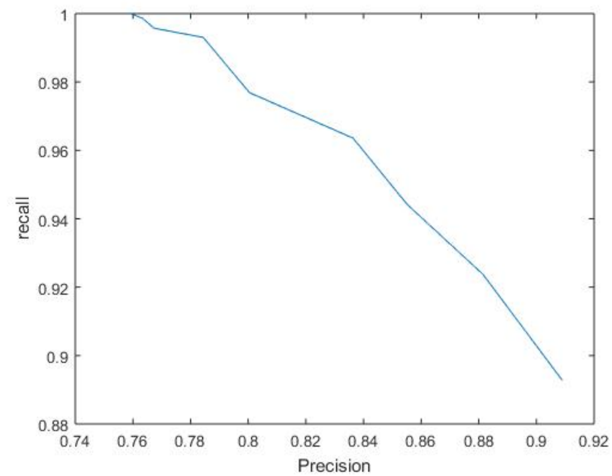


Fig. 6: Test result of AlexNet.

record the accuracy of the detection system, we also test the image segmentation for the selective search in different color spaces for the correction rate. Table I shows the evaluation results. The detection system only focuses on the correct capture of the traffic sign location. The selected road sign image region is integrated into the identification system. Thus, only the values of true positive, false negative and accuracy are reported in the table.

According to the experimental results in Table I, although the HSI color space is used to reduce the impact of light and shadow changes in the image, it still cannot fully reduce its influence. The reflective light and back-light will cause the detection failure. In different color spaces, although the CIE XYZ gives the highest correction rate, it still cannot distinguish the background from the traffic sign due to their similarity.

TABLE I: The selective search in the color space with the accuracy and recall rates, and the red filter screening results.

	True Positive	False Negative	Precision
RGB	238	19	92.6%
HSI	215	42	83.6%
YCrCb	230	27	89.4%
CIE XYZ	241	16	93.7%
CIE Lab	230	27	89.4%
CIE Luv	224	33	87.1%
Red filter	246	11	95.7%

C. Traffic Sign Recognition System

In this system, the part of deep learning is based on Caffe. We have 39209 training images from GTSRB as our training data, which can be divided into 43 road sign categories. The test data are randomly selected 1000 images from GTSRB. The results of training and testing via GoogLeNet are shown in Fig. 5. We use gradient descent method to carry out the training. The parameters include the basic learning rate set as 0.01, the number of iterations as 100000 times, and the weight attenuation as 0.0002.

Fig. 6 shows the results of using AlexNet. The gradient descent method is used for training, and the test parameters are as follows: The base learning rate is set as 0.01, the number of iterations is 600, and the weight attenuation is 0.0005. With the above parameters, when the iteration is more than 600 times, the model will incur the over-fitting issue.

V. Conclusion

This paper presents a set of driving assistance techniques designed to help general drivers or unmanned vehicles to detect and identify road signs and increase their safety. The system can be divided into road sign detection and identification with two subsystems. The road sign detection system uses the color, the image segmentation and the hierarchical grouping methods to select the candidate area of the road sign. We then use convolutional neural network for the road sign recognition system. The candidate area from detection is used for identification. The correct rates for the detection and recognition systems are 92.63% and 80.5%, respectively.

Acknowledgment

The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 104-2221-E-194-058-MY2 is gratefully acknowledged.

References

[1] A. de la Escalera, J. M. Armingol, J. M. Pastor, and F. J. Rodriguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *Trans. Intell. Transport. Sys.*, vol. 5, no. 2, pp. 57–68, Jun. 2004.

[2] W.-C. Huang and C.-H. Wu, "Adaptive color image processing and recognition for varying backgrounds and illumination conditions," *IEEE Transactions on Industrial Electronics*, vol. 45, no. 2, pp. 351–357, Apr 1998.

[3] D. L. Kellmeyer and H. T. Zwahlen, "Detection of highway warning signs in natural video images using color image processing and neural networks," in *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*, vol. 7, Jun 1994, pp. 4226–4231 vol.7.

[4] L. Pacheco, J. Batlle, and X. Cufi, “A new approach to real time traffic sign recognition based on colour information,” in *Intelligent Vehicles ’94 Symposium, Proceedings of the*, Oct 1994, pp. 339–344.

[5] W. D. Shaposhnikov, D. G. Shaposhnikov, L. N, E. V. Golovan, and A. Shevtsova, “Road sign recognition by single positioning of space-variant sensor,” in *Proc. 15th International Conference on Vision Interface*, 2002, p. 2002.

[6] S. Vitabile, A. Gentile, and F. Sorbello, “A neural network based automatic road signs recognizer,” in *Neural Networks, 2002. IJCNN ’02. Proceedings of the 2002 International Joint Conference on*, vol. 3, 2002, pp. 2315–2320.

[7] J. Miura, T. Kanda, and Y. Shirai, “An active vision system for real-time traffic sign recognition,” in *ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493)*, 2000, pp. 52–57.

[8] A. Broggi, P. Cerri, P. Medici, P. P. Porta, and G. Ghisio, “Real time road signs recognition,” in *2007 IEEE Intelligent Vehicles Symposium*, June 2007, pp. 981–986.

[9] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, “Road traffic sign detection and classification,” *IEEE Transactions on Industrial Electronics*, vol. 44, no. 6, pp. 848–859, Dec 1997.

[10] R. Belaroussi and J.-P. Tarel, “A real-time road sign detection using bilateral chinese transform,” in *Proceedings of the 5th International Symposium on Advances in Visual Computing: Part II*, ser. ISVC ’09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 1161–1170.

[11] R. Belaroussi and J. P. Tarel, “Angle vertex and bisector geometric model for triangular road sign detection,” in *2009 Workshop on Applications of Computer Vision (WACV)*, Dec 2009, pp. 1–7.

[12] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, “A system for traffic sign detection, tracking, and recognition using color, shape, and motion information,” in *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, June 2005, pp. 255–260.

[13] W. J. Kuo and C. C. Lin, “Two-stage road sign detection and recognition,” in *2007 IEEE International Conference on Multimedia and Expo*, July 2007, pp. 1427–1430.

[14] J. Greenhalgh and M. Mirmehdi, “Real-time detection and recognition of road traffic signs,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498–1506, Dec 2012.

[15] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, “Road-sign detection and recognition based on support vector machines,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264–278, June 2007.

[16] C.-Y. Fang, S.-W. Chen, and C.-S. Fuh, “Road-sign detection and tracking,” *IEEE Transactions on Vehicular Technology*, vol. 52, no. 5, pp. 1329–1341, Sept 2003.

[17] D. Cirean, U. Meier, J. Masci, and J. Schmidhuber, “A committee of neural networks for traffic sign classification,” in *The 2011 International Joint Conference on Neural Networks*, July 2011, pp. 1918–1921.

[18] P. Sermanet and Y. LeCun, “Traffic sign recognition with multi-scale convolutional networks,” in *The 2011 International Joint Conference on Neural Networks*, July 2011, pp. 2809–2813.

[19] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ser. NIPS’12. USA: Curran Associates Inc., 2012, pp. 1097–1105.

[21] S. Arora, A. Bhaskara, R. Ge, and T. Ma, “Provable bounds for learning some deep representations,” in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ser. ICML’14. JMLR.org, 2014, pp. I–584–I–592.