Image Segmentation using K-means and Unet Models

Rongjie Wang, Haiyun Zhou, Panyi Chen, Yunbing Bai

University of North Carolina at Chapel Hill

Abstract

Image segmentation aims to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Existing methods or models is already excellent on segmenting the images. We would use two existed models to try image segmentation by ourselves: first K-means model and second U-net model. The code can be found: https://github.com/cpysleeper/comp562_final.

1 Introduction

Machine learning is a powerful tool for thr modern computer science field and beyond. Many subjects teach related concepts to utilize Machine learning to conduct anlysis on data or natural languages. As searching for interesting topics, our group attempts to apply these concepts to image processing. Images compared with natural language tasks, are more visualized with the effects of color stimulus and concrete contours around the shape. Therefore, in this project, we want to try and apply some image segmentation models and compare their performances.

Before we dig deeper into image segmentation, let's first describe some main types of images: light intensity image, range image, nuclear magnetic resonance image, thermal image, et cetera(Pal and Pal, 1993). Human daily receptions of lights with eyes and cameras are light intensity images. They rely on the variations of light intensity to show the picture. A range image is a map of depth information laid out on a surface, while nuclear magnetic resonance images use the radio waves generated by biological bodies exposed to electromagnetic waves. Thermal image, as its name infers, uses temperatures as the counterpart.

Image segmentation has practical applications for processing images and provides solutions to target images or components from images. To be more specific, image segmentation at its basic level, is to assert a category for each pixel and group some common ones together. For example, the boundaries of an object in the image should have similar features such as color. This feature allows the computer to trace the boundaries of object and distinguish the contour and the filling. Image segmentation is a meaningful topic to study because it is widely used in the image recognition system that owns prevalent in modern life. With a simple search with "image segmentation" as the keyword, many results show up connecting distinctive fields. Some display strong connections with biochemistry and biomedical studies. Such capability of image processing motivates our group to learn more about this subject.

2 Related work

Intrigued by the broad application of segmentation, hundreds of techniques are proposed in the literature. But most of them are established and developed for one class of image so they may not maintain their superior performance when applied to other classes. Moreover, finding a unified method that is feasible for all types of images is still one of the challenging tasks in the literature on segmentation methods. Generally, Fu and Mui (1981) categorized these techniques by their objectives:

- Characteristic feature thresholding or clustering,
- Edge detection,
- Region extraction.

In a more up-to-date review by D. Kuar and Y. Kuar (2014) on image segmentation techniques methods, the authors confirmed the traditional ways of thresholding method, edge based segmentation method, region based segmentation method, and clustering based segmentation method. In addition, they included watershed based methods, partial differential equation based segmentation method, and artificial neural network based segmentation method.

Our project focuses on the clustering-based segmentation method and the neural network method, as the clustering method persists to be a useful approach while the neural network method emerges to be effective more recently. Specifically, the clustering method has two categories: hierarchical and partition based method. The hierarchical clustering uses a tree to with the root being the whole database and nodes being clusters. The partition based method more about calculating distances between data points and their cluster centers. Extensive work has been done to boost the performance. For instance, researchers use the Davies-Bouldin index (Davies and Bouldin, 1979), as a measure of the validity of clusters, to minimize the within-cluster scatter and maximize the between-cluster separation. Also, Ray and Turi (1999) shined lights on the optimal choice of k to balance the trade-off between variance and cluster validity.

On the other hand, Deep learning segmentation methods using convolution neural networks have been providing state-of-the-art performance in the past few years. Among which, U-Net has become one of the most attractive for applications like medical image classification, segmentation, and detection. The U-Net architecture follows an encoder-decoder hierarchical structure. The encoder gradually compresses information into a lower-dimensional representation. Then, the decoder decodes this information back to the original image dimension. Related extensions include modifications of U-net on fields such as the recognition of medical compartments . For example, Luo et al. (2019), added an attention mechanism after the encoder-decoder structure to add the dense connections to enhance the transmission efficiency of features to achieve their needs in segmenting target vessels from their images. The strong presence of both models in the field attracts our interests, so we adopt them for our project to explore their properties.

3 Method

3.1 K-means Clustering Model

As mentioned, the task of image segmentation is essentially grouping similar pixels of an image to extract part of the picture for use. K-means clustering algorithm is a commonly used method for that purpose. This model is a clustering algorithm which belongs to unsupervised algorithms. In other words, there is no labelled data available.

K in the name of the algorithm refers to the number of clusters during the grouping process. For example, we want to trace the contour of an animal image, the number of clusters k is set to 2. The idea of K-means model is to calculate the squared distances between points and the cluster center, randomly assigned at first, and rearrange some of the points to their nearest clusters if the distance to that center is closer than the distance to the current one (Siddheswar Ray, Rose H. Turi, 1999). The overall minimized clusters should contain points for distinct groups. Visually, each group could be presented with a different color to contrast each other.

3.2 U-Net Model

Besides the unsupervised learning approach, we also want to utilize convolution neural networks. U-Net is a well-known example for this task with the advantage of using available annotated sample efficiently in training (Olaf Ronneberger, 2015). In a U-Net model, there is an encoder and decoder, and in our project, we utilize tools provided by Tensorflow to avoid training the encoder by ourselves. The pretrained MobileNetV2 model helps downsample pixel values to decrease the bit rate during training. We use the upasmple block from pix2pix to decode. Setting the epoch of training results in a trainable model improving predicting accuracy each epoch. After training, the predicted mask of an image is shown.

4 Experiments and Results

In this section, we display a sample result and discuss the performance of the K-means Clustering model and the U-Net convolution networks model. Overall, the convolution networks with more training outperform the unsupervised learning method.

To demonstrate the difference, we use the dataset of pets in Parkhi et al. (2012), the Oxford-IIIT-Pet dataset, covering 37 different breeds of cats and dogs. This dataset is representative in the sense that these animals are appearing subtle differences between the breeds and making the visual task quite challenging. Figure 1 provides an dog photo in the dataset. From left to right, we plot one image of pet and a targeting segmentation related to it. The segmentation here is a trimap with regions corresponding to: foreground (the pet body), background, and ambiguous (the pet body boundary and any accessory such as collars).

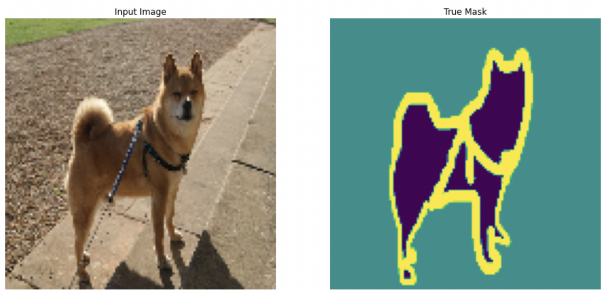

Figure 1: An example image in Oxford-IIIT-Pet and its segmentation.

4.1 Implementation details

4.1.1 K-means Clustering

After reshaping the image into a 2D array of pixels and 3 color values, we are capable to implement k-means clustering method. Since the segmentation we care is to built the trimap containing the pet's body, then we set the number of clusters k as 2, which separates the data points into two clusters, one consists of points belonging to the pet and one for the background. Besides, in our implementation, we set the criteria of convergence to be either the accuracy is 100% or the iteration reaching the maximum, 12 times. For the example image shown in Figure 1, we present the resulting segmentation in Figure 2.

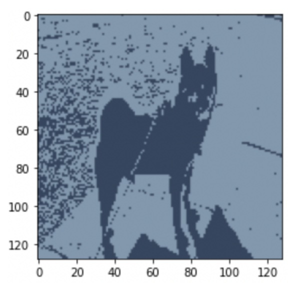

Figure 2: The resulting segmentation for the example image with k-means clustering method ($k = 2$).

In Figure 2, we can clearly see the contour of the dog as well as its body. However, some background points are also clustered into the region corresponding to the pet. That is caused by the limited number of clusters we used. As there is only two available clusters, some data points in the background which have similar color as the pet will possibly be grouped into the trimap of the pet.

4.1.2 U-Net Model

In our U-net model, we specify the input image with shape $128 \times 128 \times 3$, the number of filters used for the convolutional layers as 32. We also determine the number of output channels as 3, which implies 3 possible clusters. That is because there are three possible labels for each data points: pet, background, and outline.

In order to illustrate the superiority of U-net method, we compare the predicted segmentation of the deep learning method with and without U-net algorithm. The right figure in Figure 3 shows the predicted segmentation based on machine learning method only, which basically extracts zero information from the original image. But, after 12 epochs, the U-net method gives a segmentation close to the targeting one, as shown in Figure 4.

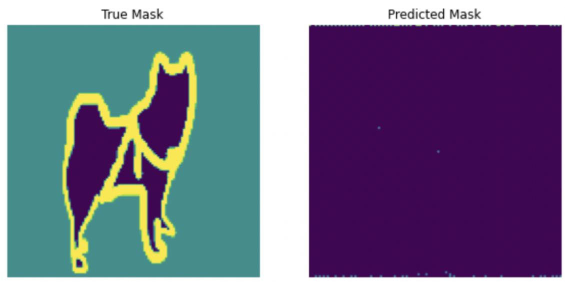

Figure 3: From left to right: the targeting segmentation, and the predicted segmentation based on machine learning method without U-net.

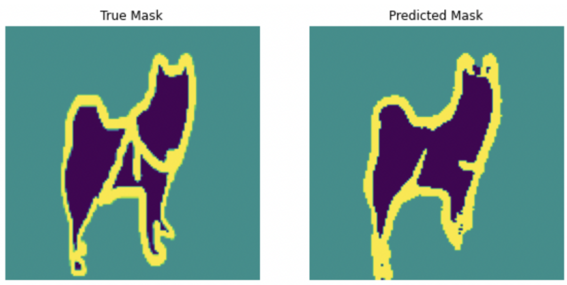

Figure 4: From left to right: the targeting segmentation, and the predicted segmentation based on machine learning method with U-net.

4.1.3 Result

For the k-means models, we found that it needs to train each time a new picture is targeted for segmentation, and the result is not clear enough. Borders between the target object and background are not obvious enough. However, our U-net model has Much better accuracy: 0.9098 for the train set and 0.8249 for the test set.

5 Conclusion and Discussion

The U-net model is easier to find the ideal segmentation which obviously divides the object and the background. The trained model can be easily applied for new pictures. If new pictures are very different from the training set, we need to feed the model with new data similar to the new category. Then we can get a positive result from the model.

Our parameters search is not enough due to the time limit, so the result is not optimized. Parameters like drop rate could be further trained to gain an optimized result. if the model is trained on a small training set, we could improve the performance by using a large training set or adding the technique of a few shots learning(Kang, 2022). A more direct method to optimize the image segmentation model is to use an enhanced U-Net-like model, such as Mask R-CNN(He, 2017).

6 Reference

N. Pal and S. Pal, A Review on Image Segmentation Techniques, Pattern Recognition, Vol. 26, No. 9, 1277-1294, 1993.

K.S. Fu and J.K. Mui, A survey on image segmentation, Pattern Recognition, Vol. 13, 3-16, 1981.

D.L. Davies and D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Machine Intell. Vol. 1, 224-227, 1979.

R. Siddheswar, and R.H. Turi. Determination of number of clusters in k-means clustering and application in colour image segmentation. In Proceedings of the 4th international conference on advances in pattern recognition and digital techniques, 137-143. 1999.

Z. Luo, Y. Zhang, L. Zhou, B. Zhang, J. Luo and H. Wu, "Micro-Vessel Image Segmentation Based on the AD-UNet Model," in IEEE Access, vol. 7, pp. 143402-143411, 2019.

R. Olaf, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, 234-241. Springer, Cham, 2015.

O.M. Parkhi et al., Cats and dogs, 2012 IEEE conference on computer vision and pattern recognition. IEEE, 3498-3505, 2012.

Kang, Dahyun, and Minsu Cho. "Integrative Few-Shot Learning for Classification and Segmentation." arXiv preprint arXiv:2203.15712 (2022).

He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.