Surgical Tool Classification Using Densely Connected Convolutional Networks

Duc My Vo

Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Korea

Sang-Woong Lee

Pattern Recognition and Machine Learning Lab, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam 13120, Korea

Abstract

Automatically surgical tool detection in videos has drawn increasing attention recently with the introduction of the challenging CATARACTS dataset. Detecting automatically 21 similar surgical tools in this dataset is not an easy task. This is because some surgical tools can appear in the same image and we do not have any information of their locations. For this reason, we developed a multi-label classification algorithm for surgical tool detection. Our proposed algorithm is mainly based on the Densely Connected Convolutional Networks (DenseNet). This deep learning network is currently attracting the attention of researchers because it obtains significant improvements over the state-of-the-art algorithms of Deep Convolutional Neural Networks (DCNN). We used the CATARACTS dataset to demonstrate the effectiveness of our proposed method on the problem of multi-label object classification.

Keywords: Densely Connected Convolutional Networks, surgical tool detection.

1. Introduction

Recently, deep convolutional neural networks have made noteworthy contributions to the development of object recognition systems. By training features

on very large datasets, several deep learning methods achieved very high recognition rates. The convolutional neural networks are usually designed to build a deep model that automatically transforms input images into a common set of distinct features. This model with a very deep architecture can be more successful for training than traditional models with the same number of parameters. Therefore, many excellent methods of training deep convolutional neural networks are successful in learning object representations from very large datasets.

Recently, a new convolution neural network architecture, Densely Connected Convolutional Network, has shown outstanding results on object detection tasks. In a Densely Connected Convolutional Network, each layer is directly connected to every other layer to ensure maximum information flow between layers. Thus, the network is more accurate and easier to train than other deep convolutional neural networks. Compared to Inception networks [2] and ResNet networks [4], Densely Connected Convolutional Networks are simpler and more efficient. For this reason, in this study, we focused on building a deep learning network based on a densely connected convolutional model for surgical tool detection in cataract surgery videos.

In our research, we also found that the problem of surgical tool annotation is challenging because, in many video frames, two or more surgical tools can appear at the same time. Since these surgical tools are not annotated by bounding boxes, it is very difficult to apply conventional algorithms for detecting them. In our analysis, the problem of surgical tool annotation can be categorized into problems of multi-label classification. For this reason, we extended the Densely Connected Convolutional Network to work as a multi-label classifier by adding a Euclidean loss layer in the end of the network. We aim to demonstrate that the Densely Connected Convolutional Network is not only the state-of-the-art method of object detection but also performs excellently in multi-label classification tasks.

2. Proposed Approach

We aim to build a deep learning network using the DenseNet-161 model [5] for surgical tool detection in cataract surgery videos. The DenseNet-161 model with 161 layers was pretrained on ImageNet [3] to accelerate the training process. This model was built from dense blocks and pooling operations, where each dense block is an iterative concatenation of previous feature maps. We replace the last layer of the DenseNet-161 network by a fully connected layer, consisting of twenty one units for twenty one individual surgical tools. To improve further computational efficiency of the DenseNet-161 network working as a multi-label classifier, we add a Euclidean loss layer in the end of the network to compute the sum of squares of differences of the predicted output and the ground truth input.

In the deployment process, a binary classification layer is added at the end of this network. The binary classification layer is used to threshold the outputs of the fully connected layer and classify them into binary labels 0, 1. The label 1 is represented for the corresponding tool found in the current video frame. The label 0 means the corresponding tool does not appear in this frame.

The input images are resized to 224×224 and they are randomly flipped for data argumentation. To improve the visibility and contrast of color images, we apply histogram equalization using cubic root distribution to obtain the uniform histogram. To normalize image inputs, we rescale all pixel values by dividing the data by 255 so that all the pixel values lie in the range [0,1].

Our network is trained using stochastic gradient descent (SGD). The initial learning rate is set to 0.001, and is divided by 10 after 50,000 iterations. To optimize the network, the momentum is set to 0.9 and the weight decay is set to 0.0005. Due to GPU memory constraints, our network is trained with a batch size of 16.

3. Experiments

In this section, we demonstrate the effectiveness of our proposed method on the problem of surgical tool detection in cataract surgery videos. We used the challenging cataract surgical dataset [1] performed in Brest University Hospital to evaluate the accuracy of our method. The original video frames are 1920×1080 and the frame rate is approximately 30 frames per second. This dataset includes 50 videos on cataract surgeries. Our task is to design a system that can recognize which surgery tools appear in images. In total, our system has to identify 21 surgery tools in 25 training videos and 25 testing videos. In each training image, each surgery tool is labeled by 0 or 1. The label of an arbitrary tool is set to 1 if this tool is being used in the current image and vice versa. To train the model and evaluate the performance of our model, we extracted training images from training videos with sampling rate of 15 frames per second.

We trained our system with the DenseNet-161 model [5], and with the ResNet-101 model [4] on this training dataset, respectively. In the final results, the system with the DenseNet-161 model outperforms its competitor. Thus, we selected the DenseNet-161 model for the multi-label classification task.

4. Conclusion

In this paper, we proposed a multi-label classifier for automatically surgical tool detection. Our classifier is built based on a Densely Connected Convolutional Network. Our network can recognize accurately any surgical tool that appears in video frames.

References

[1] CATARACTS. 2017. Challenge on Automatic Tool Annotation for CATARACT Surgery. https://cataracts.grand-challenge.org//. (2017).

[2] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385, 2015.

[3] A. Krizhevsky, I. Sutskever, and G. E Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[4] G. Huang, Z. Liu, K. Q. Weinberger, L. V. Maaten, : Densely connected convolutional networks. In: CVPR. pp. 2261–2269 (2017)

[5] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi. 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv:1602.07261. (2016).