DE-Net: Dilated Encoder Network for Automated Tongue Segmentation

Speaker: Hui Tang

University: Nanjing University of Finance & Economics

CONTENTS

PART ONE

Introduction

Introduction

- Tongue is one of the most important organs in the oral cavity, which plays a critical role in breathing, feeding and speech. Chinese medicine doctors believe that the tongue is closely related to the internal organs of human and can be used as an indicator of human health. Therefore, tongue analysis and recognition are of great significance in the diagnosis of diseases.
- Tongue image segmentation is a key step of automated tongue recognition system.
- With the popularity of mobile devices, it is easy for people to collect tongue images by mobile devices, making the task convenient and of low cost. However, images taken by mobile devices are easily affected by various environmental conditions, such as imaging background and illumination conditions, different imaging angle and huge variance of tongue positions, which makes fine segmentation a more challenging task compared with those taken by specialized acquisition devices.

Introduction

- Recently, some attempts have been made to apply deep learning to tongue image segmentation.
- However, the successive pooling layers of these methods lead to loss of information on image details, making them fail when segmenting low-quality images captured by mobile devices.
- To address this issue, we propose a novel network structure named Dilated Encoder Network (DE-Net) for automated segmentation of tongue images acquired from a mobile device in opening environment.
- In the proposed method, an efficient hybrid cascade dilated convolution (HCDC) block is designed to capture wider and deeper semantic features without losing resolution by using dilated convolutions.

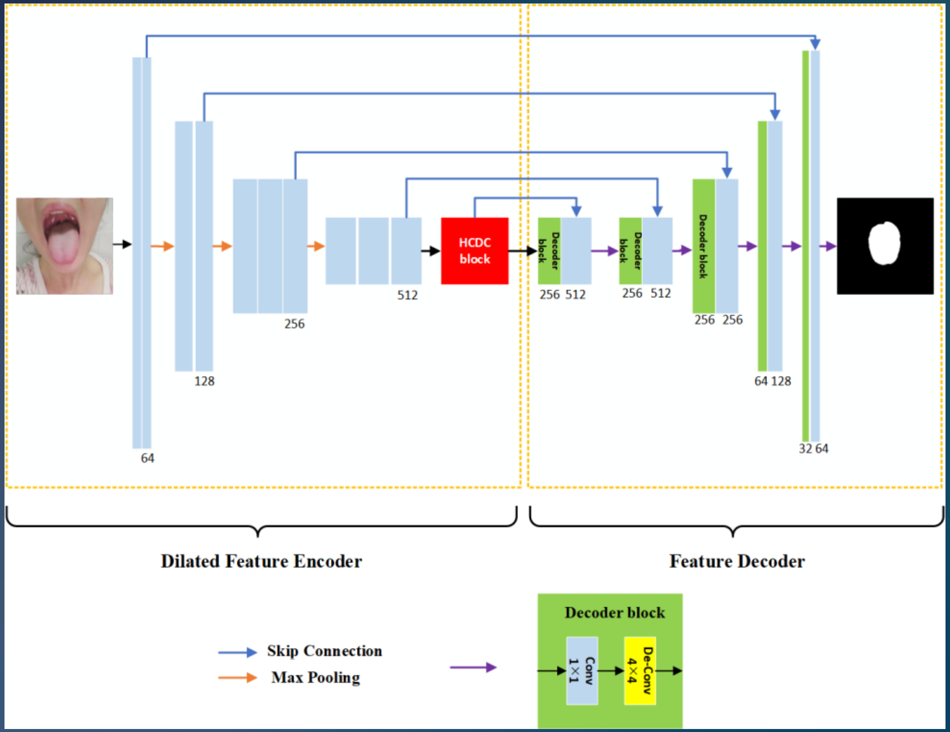
contents page

PART TWO

Methods

Methods

The proposed DE-Net for tongue image segmentation consists of two major modules: a dilated feature encoder module, and a feature decoder module.



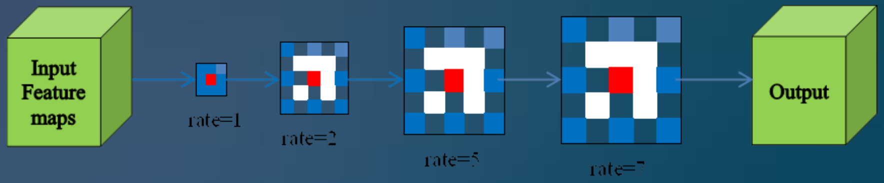
Methods

- Dilated Feature Encoder Module

  We adopt the VGG16 in the feature encoder module and removed the last two pooling layers, last three convolution layers and all the fully connected layers and replace each of them with an HCDC block. In addition, batch normalization mechanism is added after each output of convolutional layer before applying the activation function.

- HCDC

  The dilated convolution is stacked in cascade mode. And the dilation rates are 1, 2, 5 and 7, respectively. By combining the dilated convolution of different dilation rates, the HCDC block is able to extract features for objects with various sizes



Methods

- Feature Decoder Module

The feature decoder module is adopted to restore the high-level semantic features and spatial information extracted from the dilated feature encoder module. The skip connection takes some detailed information from the encoder to the decoder to remedy the information loss due to consecutive pooling and striding convolutional operations .

An efficient decoder block is used to enhance the decoding performance. The decoder block includes a 1×1 convolution and a 4×4 de-convolution that doubles the size of a feature map while reducing the number of channels by half.

PART THREE

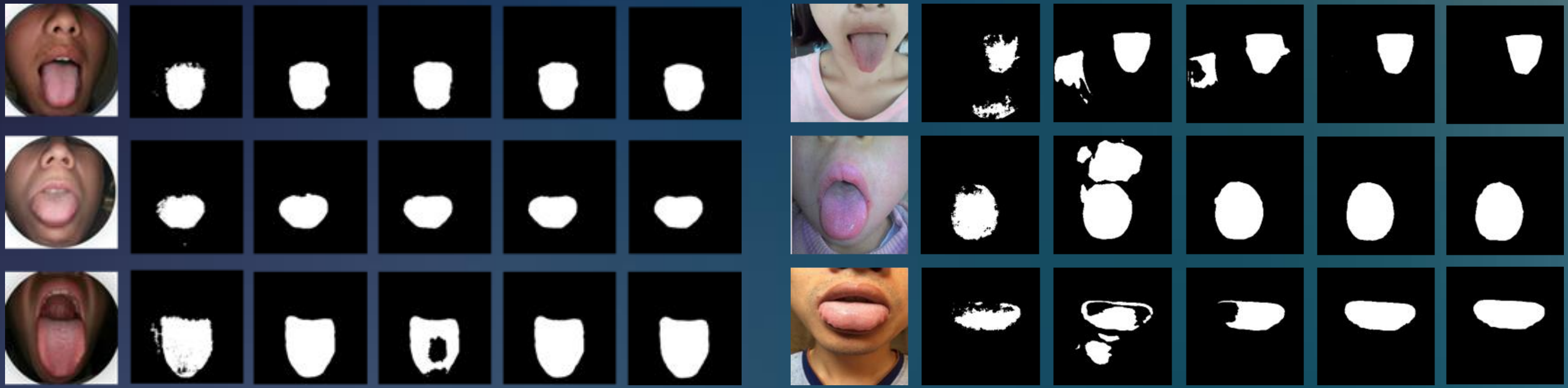
Results

Results

TABLE I. PERFORMANCE COMPARISON OF THE PROPOSED DE-NET WITH THE STATE-OF-THE-ARTS METHODS.

Method	Dataset	DSC (%)	Sen (%)	Spe (%)
SegNet [17]	TongueDataset1	93.65	94.04	98.17
	TongueDataset2	89.16	83.22	98.55
U-net [19]	TongueDataset1	97.01	98.65	98.47
	TongueDataset2	93.57	95.43	97.12
Yuan [26]	TongueDataset1	96.19	97.30	98.21
	TongueDataset2	91.22	92.60	95.41
DeepTongue [13]	TongueDataset1	97.13	98.66	98.49
	TongueDataset2	93.59	95.47	97.45
Proposed DE-Net	TongueDataset1	**98.28**	**99.74**	**99.52**
	TongueDataset2	**97.50**	**98.88**	**98.96**

Results



Results on TongueDataset1/TongueDataset2. From left to right: original tongue images, results obtained by SegNet, U-net, DeepTongue, DENet and labels.

PART FOUR

Conclusion

Conclusion

- In this paper, we introduce a dilated encoder network (DE-Net) to capture more high-level features and get high-resolution output for automated tongue image segmentation. In addition, the dice coefficient loss function is introduced to better balance the tongue and non-tongue regions. Moreover, we construct two tongue image datasets which images taken by specialized devices and mobile devices, respectively, to verify the effectiveness of the proposed method.
- Experimental results on both image datasets demonstrate that the proposed method outperforms the state-of-the-art methods for tongue image segmentation. It is believed that our approach will lay a good foundation for the subsequent extraction of color, coating, cracks, teeth marks and texture features in automated tongue recognition system.

THANKS!