AECNN: Adversarial and Enhanced Convolutional Neural Networks

Saeed Izadi[1], · **Ghassan Hamarneh**[1]

1 Introduction

Izadi and Hamarneh's method for segmenting gastrointestinal polyps from colonscopy images uses an adversarial and enhanced convolutional neural networks (AECNN). As the number of training images is small, the core of AECNN relies on fine-tuning an existing deep CNN model (ResNet152). AECNN's enhanced convolutions incorporate both dense upsampling, which learns to upsample the low-resolution feature maps into pixel-level segmentation masks, as well as hybrid dilation, which improves the dilated convolution by using different dilation rates for different layers. AECNN further boosts the performance of its segmenter by incorporating a discriminator competing with the segmenter, where both are trained through a generative adversarial network formulation.

2 Methodology

The architecture of our method is shown in Figure 1. Given the limited number of training images, we fine-tune a fully convolutional version of the ResNet152 [2] model, pre-trained on ImageNet [6], for segmenting the gastrointestinal polyps in colonoscopy images. To tackle the problem of low-resolution feature maps caused by max-pooling operations, we utilize the method of Wang et al. [7] to incorporate a dense upsampling convolution (DUC) module, as the final component of the network, which learns to upsample the low-resolution feature maps into pixel-level segmentation maps. Compared to non-learnable upsampling techniques, e.g., bilinear interpolation, the DUC technique leads to finer boundaries. We also exploit dilated convolutional operations [8], which

[1] School of Computing Science, Simon Fraser University
{saeedi, hamarneh}@sfu.ca

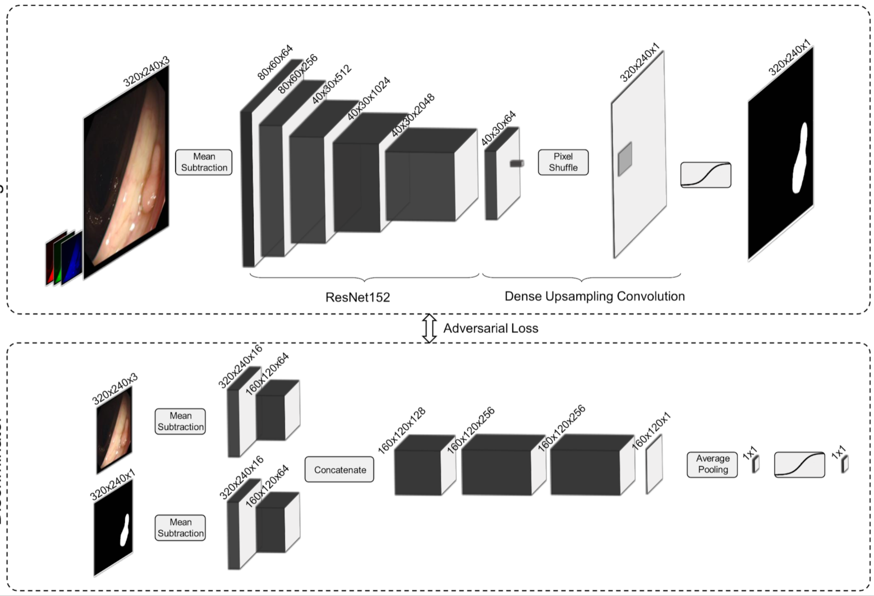

Fig. 1 The schematic of the proposed AECNN model for polyp segmentation. The error in the Discriminator is backpropagated through the Segmenter to make it produce more realistic segmentation masks.

enlarge the valid receptive field of our model, in order to improve the segmentation performance especially for large polyps. Wang et al. [7] also highlighted the "gridding effect" problem with dilated convolutions and proposed a simple yet effective solution to tackle it. Instead of using the same dilation rate after the downsampling stage, they suggested different dilation rates for each subsequent layer in a sawtooth wave-like fashion. Particularly, a number of layers are grouped together to form a "rising edge" of the wave that has an increasing dilation rate, and the next group repeats the same pattern. We also found, in our experiments, the approach of Wang et al. effective for segmenting objects with large.

Inspired by the works of Pan et al. [5] and Luc et al. [3], we further boost the performance of our model by adding a discriminator network to distinguish ground truth from generated prediction maps. Specifically, we feed the loss value of the discriminator network back to the segmenter. When the segmenter-discriminator are trained alternately, the adversarial scenario causes the two networks to compete against each other: The segmenter learns to produce prediction maps that are difficult for the discriminator to distinguish from the ground truth mask, while the discriminator attempts to correctly distinguish the true (i.e. ground truth) from the synthesized label masks. Our qualitative experiments show that this adversarial competition leads the segmenter model to uniformly highlight the polyp regions and ignore irrelevant features in the final prediction map. As data augmentation, we inflate the training set by applying rotation, horizontal and vertical flipping. All images are also resized to 240×320. We post-process the binarized prediction results with one iteration of morphological closing and opening operations with a 5×5 structuring element to remove any remaining isolated pixels or small holes.

3 Generative Adversarial Networks

Generative adversarial networks [1], GANs for short, have been recently introduced as way to train generative models in the scope of deep learning. Typically, GAN models consists of two sub-models, a generator and a discriminator which are trained jointly in an adversarial atmosphere. The generator network G receives a noise sample z from a random distribution p_z and produces a realistic sample x via capturing the data distribution p_{data} while the discriminator D takes the generated sample x as the input and determines whether it came from the true distribution p_{true} or the one learned by G. Once trained adversarially, the generator G attempts to produce realistic data samples that fools the discriminator. On the other hand, the discriminator's ultimate goal is to perfectly distinguish between the synthetic and real samples. Both G and D are trained simultaneously in a two-player training framework using the following objective function:

$$\min_G \max_D L(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z}(z)[\log(1 - D(G(z)))] \quad (1)$$

Conditional generative adversarial network [4], CGAN for short, is a variant of GANs where generator and/or discriminator are conditioned on some extra information. The conditioning is typically performed by exposing the information as inputs to the networks. Specific to binary image segmentation, extra information is the ground truth segmentation binary mask y and appears as the input to the discriminative model during the training. The objective function for CGAN is as follows:

$$\min_G \max_D L(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x|y)] + \mathbb{E}_{z \sim p_z}(z)[\log(1 - D(G(z)))] \quad (2)$$

It is noteworthy that for image segmentation, the input of the generator z is the image to be segmented.

4 References

The following are the most important references listed in descending order of importance: [7], [5], then [3].

5 Acknowledgements

SI and GH acknowledge funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada through the Collaborative Health Research Projects (CHRP), and NVIDIA Corporation for the donation of a Titan X GPU used in this research.

References

1. Goodfellow, I., et al.: Generative adversarial nets. In: NIPS, pp. 2672–2680 (2014)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
3. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408 (2016)
4. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
5. Pan, J., Canton, C., McGuinness, K., O'Connor, N.E., Torres, J., Sayrol, E., Giro-i Nieto, X.a.: Salgan: Visual saliency prediction with adversarial networks. In: CVPR Scene Understanding Workshop (SUNw) (2017)
6. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge. IJCV **115**(3), 211–252 (2015)
7. Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.: Understanding convolution for semantic segmentation. arXiv preprint arXiv:1702.08502 (2017)
8. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: ICLR (2016)