Convolutional Neural Network based Medical Imaging Segmentation: Recent Progress and Challenges

Jiaxing Tan
Road Map

- Introduction
- CNN based Models
- Encoder-Decoder based Models
- GAN Based Models
- Some Challenges
- Conclusion
Road Map

• Introduction
• CNN based Models
• Encoder-Decoder based Models
• GAN Based Models
• Some Challenges
• Conclusion
Introduction

- One key research topic in Medical Imaging is image segmentation.
- Image segmentation, or Semantic Segmentation, is a pixel level image understanding task which is to perform a pixel-by-pixel classification to decide the class of each pixel.

Figure 1: An example of semantic segmentation from the famous VOC dataset. Left: Input image. Right: semantic segmentation result.
Introduction

• In the context of Medical Imaging, such segmentation method could be utilized to solve the problems such as nodule detection, anomaly detection and organ segmentation.

• Medical image suffers the fact of high noisy and low quality, which makes it is much harder to perform segmentation on the medical images.
Introduction

Examples of lung CT slices with nodules in red mask, the nodules are quite small comparing to the whole image.

Finding the needle in the haystack
Introduction

- Traditionally, medical image segmentation is performed by hand-engineered feature based classification.
- One common attribute of such method is some empirical “magic numbers” are used for thresholding and preprocessing.
- Risk remains that those empirical values could be dataset specific which might impede the model to be a general solution.
Introduction

- As the final goal of medical imaging is to build CAD to serve the massive people, a robust and adaptive model is highly recommended.

- Convolutional Neural Network could perform automatically feature learning by itself. (Cast new light!)
Road Map

• Introduction
• CNN based Models
• Encoder-Decoder based Models
• GAN Based Models
• Some Challenges
• Conclusion
Basic Structure of CNN

Convolutional Neural Networks are similar to the deep neural networks as they are made up of neurons that have learn-able weights and biases.

In a tone of a more commonly seen introduction, CNN is a deep neural network, inspired by the biology study of human cortex, constructed by four types of layers:

- Input Layer
- Convolution Layer
- Sampling Layer
- Fully Connected Layer
An Example of CNN
CNN-Input Layer

• The Input layer is in charge of reading data with a predefined size without performing any changes to it.

• In figure below, the input layer reads in a CT scan image with size 256×256.
Convolution Layer

- 2D Convolution: An operation on two functions $f$ and $g$, which produces a third function that can be interpreted as a modified ("filtered") version of $f$.
  \[ f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) \cdot g(x - \tau) \, d\tau \]

- where * means convolution and $\cdot$ means ordinary multiplication.
CNN- Convolution Layer

- The output of the convolution layer is $k$ feature maps, each generated by a convolution operation with one kernel applied on the whole image.
- In figure below, there are 2 convolution layers. Conv1 has 32 kernels, each of size $7 \times 7$, conv2 has 64 kernels, each of size $5 \times 5$. 
Pooling Layer (2D)

Accepts a volume of size $W_1 \times H_1 \times D_1$

Requires two hyper-parameters:

- their spatial extent $F$,
- the stride $S$,

Produces a volume of size $W_2 \times H_2 \times D_2$ where:

- $W_2 = \frac{(W_1 - F)}{S} + 1$
- $H_2 = \frac{(H_1 - F)}{S} + 1$
- $D_2 = D_1$
Fully-connected layer

Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset.
The last layer of a CNN model is usually a fully connected layer which serves as an output layer. In the output layer, the number of neurons denotes the number of classes in the classification task.
Some Deep Learning Packages

• One advantage of CNN is that several public packages are available.

• Instead of building your CNN from scratch, you could take advantage of the publicly available GPU support packages.
Some Deep Learning Packages

- Caffe: http://caffe.berkeleyvision.org/
- TensorFlow: https://www.tensorflow.org/
- Theano: http://deeplearning.net/software/theano/
- Torch: http://torch.ch/
- Pytorch: http://pytorch.org/
- Mxnet: http://mxnet.io/
- Keras: https://keras.io/
- Caffe2: https://cae2.ai/

And this list keep increasing.....

Which one I like most? Depends on available code for the model.....
CNN Models

• The general idea is infer the class label of the center pixel(s) using its neighbors nearby.

• We classify them into three categories:
  • 2D CNN based
  • 3D CNN based
  • Holistic CNN based.
CNN Models

- Bounding Box model: The 2D and 3D methods are based on whether a 2D or 3D neighborhood is considered for classification of the centered pixel(s).
- As a complementary to bounding box method, we also introduce methods that does not rely on bounding box, named Holistic.
2D CNN based lung nodule detection

• Given a set of CT scans of a patient, which usually contains more than 300 slices depending on the body size of the patient, radiologists will check the scan slice by slice to detect the nodule.

• For each scan, radiologists will observe every sub-region in it. This procedure is performed in 2D slices.
2D CNN based lung nodule detection

- To simulate this procedure, 2D CNN is used.
- After applying some pre-processings such as lung segmentation and noise elimination, the slice would be cut into small sub-regions.
- Each region is an input into CNN and the output is a decision whether such region contains a nodule or not.
- The combined result shows if a nodule exists in this slice.
2D CNN based lung nodule detection

- Most recently, Nima, et. al compares the performance of several CNN structures on lung nodule detection with MTANN.

3D CNN based lung nodule detection

• When radiologists check each scan, for a better inspection, besides going through each region of the scan, they will also check the same region on the slices before or after the current slice to decide whether there is a nodule inside.

• Such detection procedure takes advantage of the 3D nature of the CT scan, which could also be an inspiration on the design of CNN based detection.
3D CNN based lung nodule detection

• Traditionally, a CNN takes a 2D matrix as an input.
• However, there are some recent publications in computer vision introduce 3D CNN for the task such as video scene recognition or 3D object recognition, which achieve promising result.
3D CNN based lung nodule detection

• In the area of lung nodule detection, due to the 3D nature of CT scan, it will be reasonable to apply 3D CNN. Some efforts have also been made.

• Rushil Anirudh et. al has applied 3D CNN on a weakly labelled lung nodule dataset. He uses a voxel $v$ as a input into the CNN to decide whether the center point $v$ located at $(x; y; z)$ to be a nodule or not.

• $v$ is defined as $(x-w : x+w; y-w : y+w; z-h : z+h)$, which means not only neighbors of $v$ in the same slice but also the neighbors on the previous and latter slice are considered to make the decision.
3D CNN based lung nodule detection

- This design is closer to how radiologist performs lung nodule detection.
- A sensitivity of 80% for 10 false positives per scan has been given on their weakly labelled dataset as a result.

Holistic CNN lung nodule detection

• All the methods we have mentioned in the previous two sections mainly obey the pipeline so that the detection result of a given slice is based on the results from a group of sub-tasks which perform nodule detection in each region of the slice with a sliding window.

• Such pipeline is not very efficient. The concern is whether we could perform nodule detection on the whole image to achieve the detection result without dividing it into sub-tasks.
YOLO

- It models detection as a regression problem.
- It divides the image into an even $(S \times S)$ grid and simultaneously predicts $(B)$ bounding boxes, confidence in those boxes, and $(C)$ class probabilities.
- Each bounding box consists of 5 predictions: center of the box (relative to the bounds of the grid cell), Width, height, and confidence.
- These predictions are encoded as an $S \times S \times (B \times 5 + C)$ tensor.
Holistic CNN lung nodule detection

Mingchen Gao et al. has applied a holistic classification on lung CT scans to detect 6 different kinds of diseases. Although the task is lung disease detection instead of lung cancer, this paper casts some light on using a different pipeline to perform nodule detection.

Fig. 1. Example images (segment of HRCT axial slices) for each of the six lung tissue types. (A) Normal (NM). (B) Emphysema (EM). (C) Ground Glass (GG). (D) Fibrosis (FB). (E) Micronodules (MN). (F) Consolidation (CD).
Road Map

• Introduction
• CNN based Models
• Encoder-Decoder based Models
• GAN Based Models
• Some Challenges
• Conclusion
Auto-encoder alike structure

- An **auto-encoder** neural network is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs.
- **My understanding**: Given an input $x$, the auto-encoder could pass only the information needed while filtering out all the noises.
The features are merged from different stages in the encoder which vary in **coarseness of semantic information**.

The upsampling of learned low resolution semantic feature maps is done using **deconvolutions which are initialized with bilinear interpolation filters**.

Excellent example for **knowledge transfer from modern classifier networks** like VGG16, Alexnet to perform semantic segmentation.
SegNet

- Auto-encoder Alike structure (symmetric).
- SegNet uses **unpooling** to upsample feature maps in decoder to use and keep high frequency details intact in the segmentation.
- This encoder doesn’t use the fully connected layers (by convolutionizing them as FCN) and hence is lightweight network lesser parameters.
• U-Net simply *concatenates* the **encoder** feature maps to upsampled feature maps from the **decoder** at every stage to form a ladder-like structure.

• The architecture by its skip concatenation connections allows the decoder at each stage to learn back relevant features that are lost when pooled in the encoder.
Link-Net

- Similar to U-Net
- Link each encoder with decoder
- Each encoder block is a ResNet Block
- Reduced parameters
PSP-Net

- Embed difficult scenery context features in an FCN based pixel prediction framework.
- First use a pretrained ResNet model with the dilated network strategy to extract the feature map.
- Then fuse different level features for further analysis.
• Given an input image (a)
• Use CNN to get the feature map of the last convolutional layer (b)
• A pyramid parsing module is applied to harvest different sub-region representations, followed by up-sampling and concatenation layers to form the final feature representation, which carries both local and global context information in (c).
• Finally, the representation is fed into a convolution layer to get the final per-pixel prediction (d).
Why Global Features?

An illustration to showcase the importance of global spatial context for semantic segmentation.
Road Map

- Introduction
- CNN based Models
- Encoder-Decoder based Models
- GAN Based Models
- Some Challenges
- Conclusion
Generative Adversarial Network (GAN)

The loss for the G network is:

\[-E_{x \sim P_z}[D(x)]\]

The loss for the D network is:

\[E_{x \sim P_z}[D(G(x))] - E_{x \sim P_{real}}[D(x)]\]
GAN Based Segmentation

Segmentor

Image
Convnet
Class predictions

Adversarial network

16
64
concat
128
256
512
0 or 1 prediction

Ground truth
GAN Based Segmentation

Given a data set of $N$ training images $x_n$ and a corresponding label maps $y_n$, the loss is defined as:

$$\ell(\theta_s, \theta_a) = \sum_{n=1}^{N} \ell_{mce}(s(x_n), y_n) - \lambda \left[ \ell_{bce}(a(x_n, y_n), 1) + \ell_{bce}(a(x_n, s(x_n)), 0) \right]$$

Training the adversarial model:

$$\sum_{n=1}^{N} \ell_{bce}(a(x_n, y_n), 1) + \ell_{bce}(a(x_n, s(x_n)), 0).$$

Training the segmentation model:

$$\sum_{n=1}^{N} \ell_{mce}(s(x_n), y_n) - \lambda \ell_{bce}(a(x_n, s(x_n)), 0)$$
Road Map

• Introduction
• CNN based Models
• Encoder-Decoder based Models
• GAN Based Models
• Some Challenges
• Conclusion
SOME CHALLENGES

• Data Source
• Data Preparation
• Some Other Challenge:
  – HPC support
  – High Cost
  – Multi-disciplinary Cooperation Required
Data Source

• CNN, as some other big data technologies, requires a large enough dataset to learn the classification rules.
• Different from computer vision area, where large and clean benchmark dataset is available, limited lung nodule dataset is available to the public.
• Most people have their own datasets containing different numbers of patients from various sources.
• Where to get data is a big challenge to perform a deep learning based detection.
Data Source

- **SPIE-AAPM-LUNGx** dataset: a dataset used for a lung challenge originally to decide whether a nodule is benign or malignant.

- **LIDC-IDRI**: contains 1018 cases, the largest public database founded by the Lung Image Database Consortium and Image Database Resource Initiative. On the website lung nodule CT scan is available for download.

- **ELCAP Public Lung Image Database**: contains 50 lowdose thin-slice chest CT images with annotations for small nodules.

- **NSCLC-Radiomics**: contains 422 non-small cell lung cancer (NSCLC) patients.
Data Preparation

• The major purposes of the data preparation are to make the training data less confusing, more fit to CNN and enrich data size.

• To make the data less confusing, some literatures perform lung segmentation to reduce noise.

• Then possible smooth methods could be applied to the segmented lung parts.

• Also, some other unnecessary parts, like some light dots or air, could be filtered out with threshold or other techniques.
Data Preparation

- For the purpose of modifying the data to be more fit for CNN, one challenge to be mentioned is the difference between CT scan and a RGB image.
- For a RGB image, it contains three channels, each channel has data ranging from 0 to 255. For a CT scan, it has only one channel with data ranging from \([-1000,3000]\], which is much larger.
- Based on our experiments, if one directly puts CT scan with such large range into CNN, the performance will be limited.
Data Preparation

• To make CT scan more similar to the image originally processed by CNN in computer vision, one solution is to rescale the data range of CT scan to $[0, 255]$.

• This could definitely cause information loss.
Data Preparation

• One idea has been raised in the literature that turn the one channel CT scan into three channels by separating attenuations into three levels: low, normal and high.

• Then the three channeled image would be rescaled into $[0, 255]$ . One benefit of this method is that the CT image now is in the same format with a RGB image.
Data Preparation

• To enlarge the size of dataset to meet the need of big data by CNN, some methods such as image translation could be applied to enlarge dataset.
• The generated ones are considered different from the original image.
• Also, adding random noise, such as white noise, to the original image, could also be a solution to enlarge dataset.
• One more thing is the issue of imbalanced dataset.

• As nodule detection is a binary classification problem (Nodule or Non-nodule), to train a classifier, the dataset should be a balanced one, which means both classes have equal number of samples.
Data Preparation

• However, obviously, in a set of CT scans, the number of slices containing nodule is much smaller than that of slices do not contain nodules. So when preparing the training dataset.

• We need to balance the dataset to make the number of two types of samples, containing nodule or not, to be equal.
Some Other Challenges

- **HPC support**: The training of CNN based model requires huge amount of calculations on huge amount of data. Even with the help of HPC can the CNN model be trained in a durable length of time. Nowadays, besides training CNN purely on CPU, CUDA accelerated GPU has also been used for training as well.

- **High Cost**: As with the need of HPC, another challenge is cost. The support of HPC consumes large amount of energy and requires facilities.
Some Other Challenges

• **Multi-disciplinary Cooperation Required**: The design of a CNN based lung nodule detection system requires the cooperation from multiple disciplinary such as medical, radiology and computer science.
Road Map

• Introduction
• CNN based Models
• Encoder-Decoder based Models
• GAN Based Models
• Some Challenges
• Conclusion
Conclusion

• We give a brief introduction on the recent progress of using CNN for lung nodule detection.
• A list of public packages as well as a list of public data are given.
• We can see that CNN has shown great potential in the area of Medical Imaging Segmentation.
• Meanwhile, challenges still remain and researchers are working on solving them.
Questions?