Convolutional Neural Networks in Medical Imaging

Mitchell Finzel

Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

15 April 2017,
Senior Seminar Conference

Introduction

- Convolutional neural networks or CNNs, have seen a rise in popularity in image related fields.
- CNNs have been having great success in biological segmentation tasks.
- These tasks include:
  - The automated detection of lymph nodes
  - Segmentation of knee cartilage
  - Detection of Alzheimer's

Introduction

- Convolutional neural networks or CNNs, have seen a rise in popularity in image related fields.
- CNNs have been having great success in biological segmentation tasks.
- These tasks include:
    - The automated detection of lymph nodes
    - Segmentation of knee cartilage
    - Detection of Alzheimer's

Introduction

- Convolutional neural networks or CNNs, have seen a rise in popularity in image related fields.
- CNNs have been having great success in biological segmentation tasks.
- These tasks include:
  - The automated detection of lymph nodes
  - Segmentation of knee cartilage
  - Detection of Alzheimer's

Introduction

- Convolutional neural networks or CNNs, have seen a rise in popularity in image related fields.
- CNNs have been having great success in biological segmentation tasks.
- These tasks include:
    - The automated detection of lymph nodes
    - Segmentation of knee cartilage
    - Detection of Alzheimer's

Introduction

- Convolutional neural networks or CNNs, have seen a rise in popularity in image related fields.
- CNNs have been having great success in biological segmentation tasks.
- These tasks include:
  - The automated detection of lymph nodes
  - Segmentation of knee cartilage
  - Detection of Alzheimer's

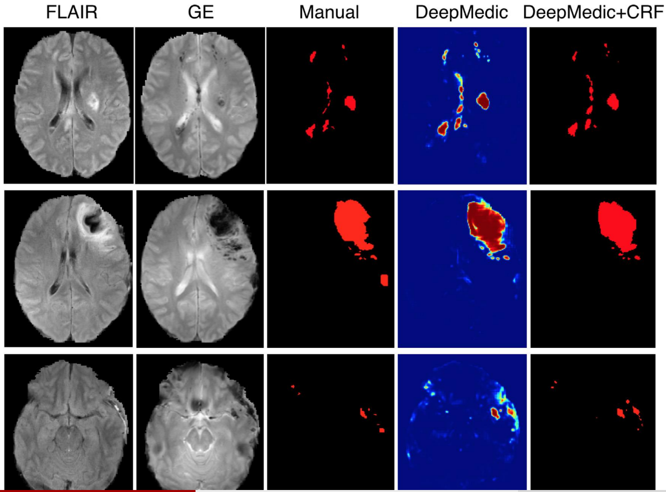
Introduction Continued

- We will be looking at two approaches to brain MRI segmentation.
- The goal of this work is to provide unsegmented MRIs to the network and receive properly segmented MRIs as output.
- Currently this requires time consuming labor from a skilled medical professional.

Introduction Continued

- We will be looking at two approaches to brain MRI segmentation.
- The goal of this work is to provide unsegmented MRIs to the network and receive properly segmented MRIs as output.
- Currently this requires time consuming labor from a skilled medical professional.

Input Output Example



Outline

- Background - Information about basic structural concepts for CNNs
- Methods used by Havaei, et al.
- Methods used by Kamnitsas, et al.
- Results
- Conclusions

Biological Segmentation

- Segmentation is the process of identifying the boundaries of different structures and classifying them
- Segmentation is loosely defined and can have a wide range of granularities.
  - Rough graularity such as identifying the different bones in a leg xray.
  - Smooth granularity such as determining the differing regions of a tumor.

Biological Segmentation

- Segmentation is the process of identifying the boundaries of different structures and classifying them
- Segmentation is loosely defined and can have a wide range of granularities.
  - Rough graularity such as identifying the different bones in a leg xray.
  - Smooth granularity such as determining the differing regions of a tumor.

Biological Segmentation

- Segmentation is the process of identifying the boundaries of different structures and classifying them
- Segmentation is loosely defined and can have a wide range of granularities.
  - Rough graularity such as identifying the different bones in a leg xray.
  - Smooth granularity such as determining the differing regions of a tumor.

Biological Segmentation

- Segmentation is the process of identifying the boundaries of different structures and classifying them
- Segmentation is loosely defined and can have a wide range of granularities.
    - Rough graularity such as identifying the different bones in a leg xray.
    - Smooth granularity such as determining the differing regions of a tumor.

Neural Networks

- Neural Networks are a form of machine learning
- Neural Networks can be thought of as pattern recognizers.
- They are loosely based on the neuronal structure of the cerebral cortex, the part of the brain that takes in sensory data.

Neural Networks

- Neural Networks are a form of machine learning
- Neural Networks can be thought of as pattern recognizers.
- They are loosely based on the neuronal structure of the cerebral cortex, the part of the brain that takes in sensory data.

Neural Networks

- Neural Networks are a form of machine learning
- Neural Networks can be thought of as pattern recognizers.
- They are loosely based on the neuronal structure of the cerebral cortex, the part of the brain that takes in sensory data.

Neural Network Structure

- Comprised of layers of nodes
- Each node has an activation function that triggers when it recognizes something in the input.
- These activations are then passed to neighboring nodes through weighted connections eventually leading to some type of output.
- The network can 'learn' by altering the weights of its connections based on the accuracy of the output to the goal result.

Neural Network Structure

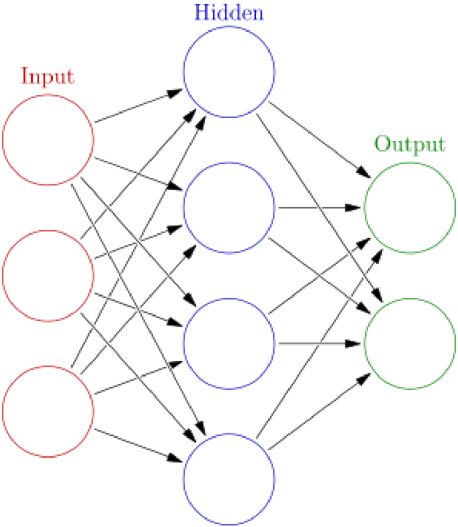
- Comprised of layers of nodes
- Each node has an activation function that triggers when it recognizes something in the input.
- These activations are then passed to neighboring nodes through weighted connections eventually leading to some type of output.
- The network can 'learn' by altering the weights of its connections based on the accuracy of the output to the goal result.

Neural Network Structure

- Comprised of layers of nodes
- Each node has an activation function that triggers when it recognizes something in the input.
- These activations are then passed to neighboring nodes through weighted connections eventually leading to some type of output.
- The network can 'learn' by altering the weights of its connections based on the accuracy of the output to the goal result.

Neural Network Structure

- Comprised of layers of nodes
- Each node has an activation function that triggers when it recognizes something in the input.
- These activations are then passed to neighboring nodes through weighted connections eventually leading to some type of output.
- The network can 'learn' by altering the weights of its connections based on the accuracy of the output to the goal result.


Hidden
Input
Output

Kernels

- Kernels, neurons and filters are interchangeable names
- Kernels are an array based representation of image features

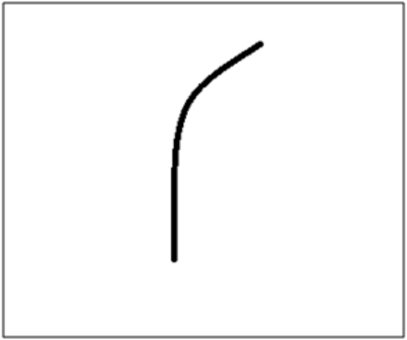
Kernels

- Kernels, neurons and filters are interchangeable names
- Kernels are an array based representation of image features

Kernels

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter



Visualization of a curve detector filter

Kernel
`https://adeshpande3.github.io/`

Convolutional Layers

- Convolutional Layers are where CNNs get their name
- Every CNN starts with a convolutional layer
- The Kernel slides or "convolves" around the input image
- The results of the convolutions are stored in the feature map

Convolutional Layers

- Convolutional Layers are where CNNs get their name
- Every CNN starts with a convolutional layer
- The Kernel slides or "convolves" around the input image
- The results of the convolutions are stored in the feature map

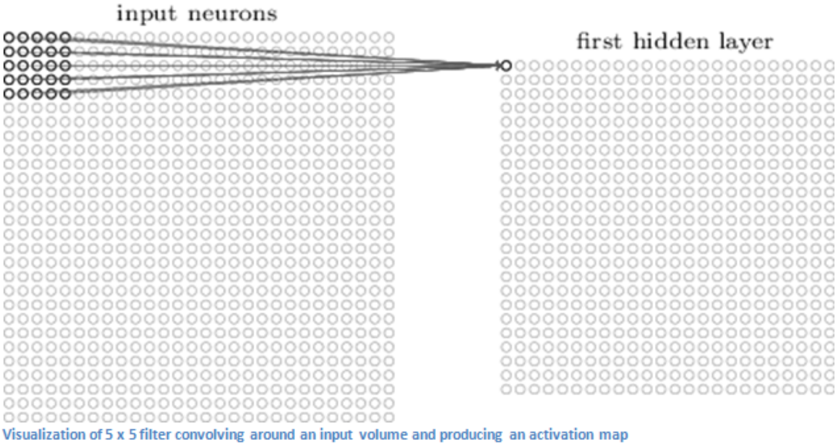
Convolutional Layers

- Convolutional Layers are where CNNs get their name
- Every CNN starts with a convolutional layer
- The Kernel slides or "convolves" around the input image
- The results of the convolutions are stored in the feature map

Convolutional Layers

- Convolutional Layers are where CNNs get their name
- Every CNN starts with a convolutional layer
- The Kernel slides or "convolves" around the input image
- The results of the convolutions are stored in the feature map

Convolutional Layers



Feature Map
https://adeshpande3.github.io/

Fully Connected Layers

- Can be thought of as the final layers in the network
- Their job is to convert the feature maps from previous layers into label probabilities.
- these probabilities come in the form of a vector that adds up to 1.

Fully Connected Layers

- Can be thought of as the final layers in the network
- Their job is to convert the feature maps from previous layers into label probabilities.
- these probabilities come in the form of a vector that adds up to 1.

Fully Connected Layers

- Can be thought of as the final layers in the network
- Their job is to convert the feature maps from previous layers into label probabilities.
- these probabilities come in the form of a vector that adds up to 1.

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
  - The forward pass
  - The loss calculation
  - The backward pass
  - Weight update

Training

- Training is the crux that makes everything work
- Training requires data that has already been properly segmented.
- Network is initialized with random kernel weights
- Training has four main steps:
    - The forward pass
    - The loss calculation
    - The backward pass
    - Weight update

Overview of Havaei, et al.

- Havaei, et al proposes a two pathway approach to the BRATS 2013 brain tumor segmentation challenge.
- Their approach has three main components
  - Two pathways
  - Two CNNs concatenated together
  - A two phase approach to training

Overview of Havaei, et al.

- Havaei, et al proposes a two pathway approach to the BRATS 2013 brain tumor segmentation challenge.
- Their approach has three main components
  - Two pathways
  - Two CNNs concatenated together
  - A two phase approach to training

Overview of Havaei, et al.

- Havaei, et al proposes a two pathway approach to the BRATS 2013 brain tumor segmentation challenge.
- Their approach has three main components
  - Two pathways
  - Two CNNs concatenated together
  - A two phase approach to training

Overview of Havaei, et al.

- Havaei, et al proposes a two pathway approach to the BRATS 2013 brain tumor segmentation challenge.
- Their approach has three main components
  - Two pathways
  - Two CNNs concatenated together
  - A two phase approach to training

Overview of Havaei, et al.

- Havaei, et al proposes a two pathway approach to the BRATS 2013 brain tumor segmentation challenge.
- Their approach has three main components
    - Two pathways
    - Two CNNs concatenated together
    - A two phase approach to training

Two Pathways

- Havaei, et al. setup their network with two pathways
  - The local pathway with a smaller 7x7 receptive field'
  - The global pathway with a larger 13x13 receptive field.
- These two pathways allow the combination of finite detail with greater locational context.

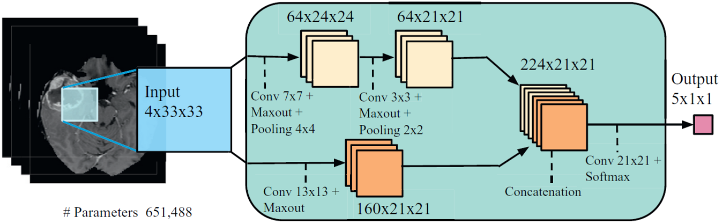
Two Pathways

- Havaei, et al. setup their network with two pathways
  - The local pathway with a smaller 7x7 receptive field'
  - The global pathway with a larger 13x13 receptive field.
- These two pathways allow the combination of finite detail with greater locational context.

Two Pathways

- Havaei, et al. setup their network with two pathways
  - The local pathway with a smaller 7x7 receptive field'
  - The global pathway with a larger 13x13 receptive field.
- These two pathways allow the combination of finite detail with greater locational context.

Two Pathways



Cascaded Architectures

- The second approach implemented by Havaei, et al. is their use of a 'cascaded architecture'.
- In this approach they take the output one of one CNN and concatenate it into a layer of another.
- The goal is to allow for joint segmentation models where different labels are allowed to correlate with one another.

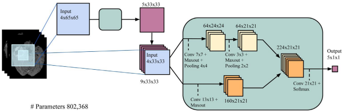
Cascaded Architectures

- The second approach implemented by Havaei, et al. is their use of a 'cascaded architecture'.
- In this approach they take the output one of one CNN and concatenate it into a layer of another.
- The goal is to allow for joint segmentation models where different labels are allowed to correlate with one another.

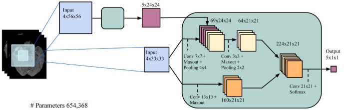
Cascaded Architectures

- The second approach implemented by Havaei, et al. is their use of a 'cascaded architecture'.
- In this approach they take the output one of one CNN and concatenate it into a layer of another.
- The goal is to allow for joint segmentation models where different labels are allowed to correlate with one another.

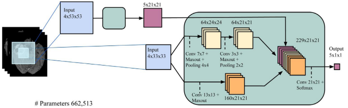
Cascaded Architectures



(a) Cascaded architecture, using input concatenation (InputCascadeCNN).



(b) Cascaded architecture, using local pathway concatenation (LocalCascadeCNN).



(c) Cascaded architecture, using pre-output concatenation, which is an architecture with properties similar to that of learning using a limited number of mean-field inference iterations in a CRF (MFCascadeCNN).

Two Phase Training

- The last approach implemented by Havaei, et al. is a two phase training system.
- This is done to alleviate the relative abundance of healthy tissue versus the small quantity of tumor tissue in each image.
- The two phases consist of:
  - First they train the network on image patches where the probability of each label being present is equal.
  - Then they retrain the final layer with the relative probabilities of each label.
- This allows for better label discrimination while maintaining proper output probabilities.

Two Phase Training

- The last approach implemented by Havaei, et al. is a two phase training system.
- This is done to alleviate the relative abundance of healthy tissue versus the small quantity of tumor tissue in each image.
- The two phases consist of:
  - First they train the network on image patches where the probability of each label being present is equal.
  - Then they retrain the final layer with the relative probabilities of each label.
- This allows for better label discrimination while maintaining proper output probabilities.

Two Phase Training

- The last approach implemented by Havaei, et al. is a two phase training system.
- This is done to alleviate the relative abundance of healthy tissue versus the small quantity of tumor tissue in each image.
- The two phases consist of:
  - First they train the network on image patches where the probability of each label being present is equal.
  - Then they retrain the final layer with the relative probabilities of each label.
- This allows for better label discrimination while maintaining proper output probabilities.

Two Phase Training

- The last approach implemented by Havaei, et al. is a two phase training system.
- This is done to alleviate the relative abundance of healthy tissue versus the small quantity of tumor tissue in each image.
- The two phases consist of:
  - First they train the network on image patches where the probability of each label being present is equal.
  - Then they retrain the final layer with the relative probabilities of each label.
- This allows for better label discrimination while maintaining proper output probabilities.

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output

Overview of Kamnitsas, et al.

- Kamnitsas, et al have five different architecture approaches
  - 3D kernels
  - Dense training
  - Two pathways
  - Deeper networks
  - 3D conditional random fields on the output