Outline

• Introduction

• Our encoder-decoder architecture
  • Input-aware residual convolutional units
  • Densely connected decoding units
  • Multi-resolution input

• Results
Introduction

• CNNs are the state-of-the-art architectures for segmenting natural and medical images.

• However, CNNs can’t be directly applied to breast biopsy WSI images due to their size.

<table>
<thead>
<tr>
<th>Diagnostic Category</th>
<th>#ROI (training)</th>
<th>#ROI (test)</th>
<th>#ROI (total)</th>
<th>Avg. size (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>$9K \times 9K$</td>
</tr>
<tr>
<td>Atypia</td>
<td>11</td>
<td>11</td>
<td>22</td>
<td>$6K \times 7K$</td>
</tr>
<tr>
<td>DCIS</td>
<td>12</td>
<td>10</td>
<td>22</td>
<td>$8K \times 10K$</td>
</tr>
<tr>
<td>Invasive</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>$38K \times 44K$</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>28</td>
<td>58</td>
<td>$10K \times 12K$</td>
</tr>
</tbody>
</table>
Introduction

• To segment these images, a simple strategy is to use a *sliding window-based* approach.

• Diving these large tissue structures limits the context available to CNNs and may affect the segmentation performance.

• We introduced a *new multi-resolution encoder-decoder architecture* that was specifically designed to handle the challenges of the breast biopsy semantic segmentation problem.

**Figure:** The set of tissue labels used in semantic segmentation. Note that the objects of interest (or tissues) are variable in size.
Encoder-decoder Network for Segmenting WSIs
Overview of encoder-decoder network

- Encoder-decoder network comprises of two networks:
  - Encoder
  - Decoder
- Encoder aggregate features at multiple spatial resolutions by performing different operations such as convolution and down-sampling operations.
- Decoder tries to invert the loss of spatial resolution due to down-sampling operations in the encoder.

Single vs Multi-resolution Network

• Patch-based approach divides large tissue structures into smaller structures and limits the context (surrounding tissue information) available to CNNs.
• Patch-based approach may affect the segmentation performance.
• To make the CNN model aware of the surrounding information, we introduce a multi-resolution network
Patch-wise Predictions: Single vs Multi-resolution Network

Figure: Patch-wise predictions of Plain Encoder-Decoder network with single and multiple resolution input. Multi-resolution input helps in improving the predictions, especially at the patch borders.
Overview of Convolutional Units

• Convolutional unit is a composite function comprising of convolutional layers, non-linearity operations (such as ReLU) and batch normalization.

• Two popular convolutional units are:

  ![Diagrams of VGG and ResNet]

  **ResNet** adds a bypass connection between input and output of the convolutional block to improve the information flow inside the network and avoid the vanishing gradient problem.

Source:


As we increase the depth of the network, we learn coarse features about the objects. These coarse features are useful for object classification, but not for segmentation.

Figure: FCN-32s architecture that upsamples the last CNN layer (VGG) output by 32x, so that input image and segmentation output are of the same resolution.

Input-aware Residual Convolution Units

• We introduce an input-aware residual convolutional unit that reinforces the input at different spatial levels of CNNs to learn input-specific features.
Activation Map Visualization

Residual Convolutional Unit (RCU) vs Input Aware Residual Convolutional Unit (IA-RCU)

Figure: Two examples visualizing the activation maps at different spatial resolutions. IA-RCU compensates the loss of spatial information due to down-sampling operations and helps in learning features that are relevant with respect to input.
Densely Connected Decoding Paths

- Similar to convolutional units, we can have skip connections between encoding and its corresponding decoding block.

- These skip-connections establishes a direct connection between encoder and decoder and improves the information flow.

- To further improve the information flow, we introduce direct connections between a decoding block and all encoding blocks that are at the same-level or lower-level.

- These connections establishes **long-range connections** and promote feature reuse.

**Figure:** Different encoder-decoder architectures
Visualization of Activation Maps of different networks

- Features learned by the plain network are noisy.
- Residual network helps in refining the feature maps by combining the low-level and high-level information.
- Dense connections promote the feature reuse and helps in efficiently combining the low-level and high-level information.
Our Encoder-Decoder Architecture for WSI Segmentation

Our encoder-decoder network for segmenting WSIs that incorporates:

• Multi-resolution input
• Input-aware residual convolutional units
• Densely connected decoding paths
• Sparse decoder

More details about the network architecture, see our paper: Learning to Segment Breast Biopsy Whole Slide Images, to appear in IEEE Winter Conference in Computer Vision (WACV-18)

Web Link: https://arxiv.org/abs/1709.02554

11/28/2017
Results
Training details

• Training Set: 30 ROIs
  • 25,992 patches of size 256x256 with augmentation
  • Split into training and validation set using 90:10 ratio

• Test Set: 28 ROIs

• Evaluation metric:
  • Pixel accuracy
  • Mean Region Intersection over Union
  • F1-score

• Stochastic Gradient Descent for optimization

• Implemented in Torch
  • [http://torch.ch/](http://torch.ch/)
Results

Key findings:

- **Singe-vs-multi-resolution**: For all models, multi-resolution improves the pixel accuracy by about 6%.
- **RCU-vs-IARCU**: IARCU improves the pixel accuracy by about 4% and 7% over RCUs (A1).
- **Residual vs Dense Connections**: The residual encoder-decoder has a 0.5% higher pixel accuracy (PA) than the plain encoder-decoder, and our model with dense connections (A3) has a 2% higher PA than plain encoder-decoder under both single and multiple resolution settings.
WSI Segmentation Results

RGB Image

Ground Truth

Predicted Semantic Mask

- background
- benign epithelium
- normal stroma
- secretion
- malignant epithelium
- desmoplastic stroma
- blood
- necrosis
WSI Segmentation Results

RGB Image  
Ground Truth  
Predicted Semantic Mask

- background
- benign epithelium
- normal stroma
- secretion
- malignant epithelium
- desmoplastic stroma
- blood
- necrosis

11/28/2017
WSI Segmentation Results

[Images of RGB images, ground truth, and predicted semantic masks with color-coded labels for various tissue components: background, benign epithelium, normal stroma, secretion, malignant epithelium, desmoplastic stroma, blood, necrosis.]

11/28/2017
Application to Cancer Diagnosis

• Segmentation labels have high descriptive power and therefore, leads to good classification accuracy even with simple classifiers such as SVM and Multi-layer perceptron (MLP)

• Multi-resolution network improves the pixel-wise classification accuracy of stroma tissue; which is an important tissue type for identifying invasive cancer.
Thank You!!

For more details about this work, please check DIGITAL PATHOLOGY project here:
https://homes.cs.washington.edu/~shapiro/digipath.html