Automatic Histopathology Image Analysis with CNNs

Le Hou, Kunal Singh, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, Roberta J. Seidman, Joel H. Saltz

Stony Brook University
Gagapixel Whole Slide Tissue Image

Biopsy

Gigapixel Tissue Image

[vector.me]
Analyzing Tissue Images

• Is crucial to study disease onset
• To develop targeted treatment
• Is a challenging problem
Gigapixel Resolution

50K × 50K

An ImageNet image

350 × 500
Nucleus Segmentation

Nuclear Attribute Recognition

- Shape
- Density
- Texture
- Mitosis?
- ...
Nucleus Segmentation
Region Attribute Recognition

• Necrosis?
• Pseudopalisading?
• Microvascular proliferation?
• ...

Nucleus Segmentation

• Shape
• Density
• Texture
• Mitosis?
• ...

Nuclear Attribute Recognition

• Shape
• Density
• Texture
• Mitosis?
Nucleus Segmentation

Region Attribute Recognition

• Cancer/non-cancer?
• Type
• Grade
• …

Cancer Attribute Recognition

Region Attribute Recognition

• Necrosis?
• Pseudopalisading?
• Microvascular proliferation?
• …

Nucleus Segmentation

Nuclear Attribute Recognition

• Shape
• Density
• Texture
• Mitosis?
• …
- Cancer/non-cancer?
- Type
- Grade
- ...

- Necrosis?
- Pseudopalisading?
- Microvascular proliferation?
- ...

- Shape
- Density
- Texture
- Mitosis?
- ...

Cancer Attribute Recognition
Region Attribute Recognition
Nucleus Segmentation
Nuclear Attribute Recognition

Feedback
Feedforward
Recognizing Attributes of Glioma Nuclei

- Perinuclear Halo
- Gemistocyte
- Nucleoli
- Hyperchromasia
- Mitosis

Automatic Recognition
Our Dataset

<table>
<thead>
<tr>
<th>Morphological Attributes</th>
<th>#. Present</th>
<th>#. Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinuclear halos</td>
<td>78</td>
<td>2000</td>
</tr>
<tr>
<td>Gemistocyte</td>
<td>51</td>
<td>2027</td>
</tr>
<tr>
<td>Nucleoli</td>
<td>77</td>
<td>2001</td>
</tr>
<tr>
<td>Grooved</td>
<td>14</td>
<td>2064</td>
</tr>
<tr>
<td>Hyperchromasia</td>
<td>505</td>
<td>1573</td>
</tr>
<tr>
<td>Overlapping nuclei</td>
<td>105</td>
<td>1973</td>
</tr>
<tr>
<td>Multinucleation</td>
<td>43</td>
<td>2035</td>
</tr>
<tr>
<td>Mitosis</td>
<td>53</td>
<td>2025</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>20</td>
<td>2058</td>
</tr>
<tr>
<td>No nucleus</td>
<td>545</td>
<td>1533</td>
</tr>
</tbody>
</table>
A Multi-label Problem

• There can be multiple classes (nuclear attributes) for each instance (glioma nuclear image).
• Existing approaches [Thibault, 2008] [Kong, 2011] ignored the multi-label nature.
A Multi-label Problem

• There can be multiple classes (nuclear attributes) for each instance (glioma nuclear image).

• Existing approaches [Thibault, 2008] [Kong, 2011] ignored the multi-label nature.

• Our contribution:
  first multi-label modeling on this problem.
A Multi-label Problem

• Existing approaches [Thibault, 2008] [Kong, 2011] only focus on a subset of classes at a time.

• Our contribution:
  Recognizing nine subtle and important attributes with good accuracy
Convolutional Neural Network (CNN)

• A popular image classification method

• Input training set:
  - Images with ground truth labels

• Output:
  - Predicted class(es)
CNN for Image Classification

Conv 1: Edge+Blob
Conv 3: Texture
Conv 5: Object Parts
Fc8: Object Classes

[James Hays]
Multi-label CNN

• Approach 1:
  • Predict each class independently
  • Drawback: do not capture inter-class dependency
    Example: mitosis are always hyperchromasia

• Approach 2:
  • A chain of CNNs [Read, 2009]
Multi-label CNN

Image → CNN → Perinuclear Halo? Gemistocyte? Nucleoli? ...

Semi-supervised CNN

- Getting ground truth labels is laborious
- Tissue Images have billions of unlabeled nuclei
- To utilize unlabeled nuclear images: Semi-supervised CNN
Training a CNN without Ground Truth Labels

Convolutional AutoEncoder (CAE)

[Mike Swarbrick Jones]
Semi-supervised CNN with CAE

1. CAE training

2. Parameter assignment

3. CNN training

Convolution + Pooling

Deconvolution + Unpooling

200 Encoding Neurons

Perinuclear halo?

Gemistocyte?

...?

No nucleus?
Reconstructed Images by CAE

Left: original nuclear images.
Right: CAE reconstructed images.
Training a CNN on Another Dataset

• Alternative:
  • Train a CNN on a different dataset that has ground truth labels.
  • Use this CNN as a feature extractor.
  • Use Support Vector Machine (SVM) as a classification model.

• We used the CNN trained by the Visual Geometry Group (VGG) as a feature extractor.
Results

<table>
<thead>
<tr>
<th>Morphological Attributes</th>
<th>Semi-supervised CNN</th>
<th>VGG16 + SVM</th>
<th>Best of two (per attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perinuclear halos</td>
<td>0.8789</td>
<td><strong>0.9257</strong></td>
<td>0.9257</td>
</tr>
<tr>
<td>Gemistocyte</td>
<td>0.8026</td>
<td><strong>0.9548</strong></td>
<td>0.9548</td>
</tr>
<tr>
<td>Nucleoli</td>
<td>0.8366</td>
<td><strong>0.9076</strong></td>
<td>0.9076</td>
</tr>
<tr>
<td>Grooved</td>
<td><strong>0.8956</strong></td>
<td>0.7296</td>
<td>0.8956</td>
</tr>
<tr>
<td>Hyperchromasia</td>
<td><strong>0.9450</strong></td>
<td>0.8854</td>
<td>0.9450</td>
</tr>
<tr>
<td>Overlapping nuclei</td>
<td><strong>0.8969</strong></td>
<td>0.8305</td>
<td>0.8969</td>
</tr>
<tr>
<td>Multinucleation</td>
<td>0.7329</td>
<td><strong>0.7507</strong></td>
<td>0.7507</td>
</tr>
<tr>
<td>Mitosis</td>
<td><strong>0.8731</strong></td>
<td>0.8559</td>
<td>0.8731</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>0.8676</td>
<td><strong>0.9767</strong></td>
<td>0.9767</td>
</tr>
<tr>
<td>No nucleus</td>
<td><strong>0.9828</strong></td>
<td>0.9639</td>
<td>0.9828</td>
</tr>
<tr>
<td>Averaged AUC</td>
<td>0.8712</td>
<td>0.8616</td>
<td><strong>0.9109</strong></td>
</tr>
</tbody>
</table>

Both methods perform well on some but not all morphological attributes and are complementary with each other.
Summary

• Automatically classify nuclei is important
• We model it as a multi-label problem
• We achieved promising results classifying nine subtle nuclear attributes
• Future work: combining two CNN-based methods. R-CNN