Automated Segmentation of Cervical Nuclei Pap Smear Images using Deformable Multi-path Ensemble Model

Jie Zhao¹, Quanzheng Li¹,²,³, Xiang Li²,³, Hongfeng Li¹, Li Zhang¹

1 Center for Data Science in Health and Medicine, Peking University, Beijing, China;
2 MGH/BWH Center for Clinical Data Science, Boston, MA 02115, USA;
3 Department of Radiology, Massachusetts General Hospital, Boston, MA 02115, USA;
Motivation

- **Pap smear testing** has been widely used for detecting cervical cancers based on the morphological properties of cell nuclei in microscopic image.

- Accurate nuclei segmentation is the first step for image-based screening.

- **Image features with detailed structural information and representation capability** are vital for accurate segmentation.
Illustration of preprocessing steps:
(a) raw Pap smear images.
(b) groundtruth with four regions.
(c) pre-processed groundtruth, bright area represents small nucleus and light gray area represents large nucleus.

7-class:
- normal
- columnar
- normal
- intermediate
- normal
- superficial
- light
- dysplastic
- moderate
dysplastic
- severe
dysplastic
carcinoma
in situ
Method: Deformable Multi-path Ensemble
Method: Deformable Multi-path Ensemble

• **Dense blocks** improves information flow between neural network layers and show better capability in feature extraction.

• **Deformable convolutions** captures detailed structures of nuclei since the abnormal cervical nuclei may display irregular shape rather than circular shape of normal nuclei.

• **A multi-path fashion** trains multiple networks simultaneously with different settings and integrates the results using a majority voting strategy to further improve segmentation result.
Dense Block

- Parameter efficiency
- Implicit deep supervision
- Feature reuse
Enhance transformation modeling capability for adjusting receptive fields of convolutional kernels.

Optimizing FOV offsets for subtle structure characterization by augmenting spatial sampling locations on feature maps.

(a) normal convolution; (b) deformable convolution with arbitrary offsets; (c-d) learned deformable convolution.
Multi-path Ensemble Model

- We found that: feature maps in contracting path are more related to **contextual information** and in expansive path are more related to **positional** and **morphological information**.
- Three networks are trained in a **multi-path fashion**:
  1. Dense-Unet;
  2. Dense-Unet with deformable convolutions in contracting path;
  3. Dense-Unet with deformable convolutions in expansive path;
Experiment and Results

- Pytorch 0.4.0 package, NVIDIA GTX 1080ti GPU with 11 GB of memory.
- Dice coefficient index as performance evaluation metric.
- All experiments are under the same set of hyper-parameters.
- Run on Herlev dataset consists of 917 cytological images from Pap smear tests.
- All images are normalized to have zero mean with unit variance intensity and are resized to size of 256×256.
Examples of segmentation results. (a) Pap smear images, (b) Manual annotations, (c) Segmentation results of Unet, (d) Segmentation results of dense U-Net, (e) Segmentation results of D-Con (deformable convolutions in contracting path), (f) Segmentation results of D-Exp (deformable convolutions in expansive path; (g) Segmentation results of D-MEM (multi-path ensemble).
Experiment and Results

<table>
<thead>
<tr>
<th>Method</th>
<th>ZSI</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised[15]</td>
<td>0.89±0.15</td>
<td>0.88±0.15</td>
<td>0.93±0.15</td>
</tr>
<tr>
<td>FCM[16]</td>
<td>0.80±0.24</td>
<td>0.85±0.21</td>
<td>0.83±0.25</td>
</tr>
<tr>
<td>P-MRF[17]</td>
<td>0.93±0.03</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SP-CNN[18]</td>
<td>0.90</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>Our Method</td>
<td><strong>0.933±0.14</strong></td>
<td><strong>0.946±0.06</strong></td>
<td><strong>0.984±0.00</strong></td>
</tr>
</tbody>
</table>

Quantitative comparison of state-of-the-art methods with our proposed method in terms of mean(± standard deviation) of ZSI, precision and recall rates.

Experiment and Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>U-Net</th>
<th>Dense-Unet</th>
<th>D-Con</th>
<th>D-Exp</th>
<th>D-MEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZSI</td>
<td>0.869</td>
<td>0.910</td>
<td>0.918</td>
<td>0.917</td>
<td><strong>0.933</strong></td>
</tr>
<tr>
<td>Precision</td>
<td>0.897</td>
<td>0.893</td>
<td>0.888</td>
<td>0.894</td>
<td><strong>0.946</strong></td>
</tr>
<tr>
<td>Recall</td>
<td>0.879</td>
<td>0.956</td>
<td>0.972</td>
<td>0.961</td>
<td><strong>0.984</strong></td>
</tr>
</tbody>
</table>

Comparison among different model settings.
**Effect of dense connection**: Experiment on Herlev dataset shows that performance of Dense-Unet (0.910) is better than U-Net (0.896).

(a) Input images. (b) Ground truth. (c) Segmentation results of U-Net. (d) Result of Dense-Unet.
Discussion

Effect of deformable convolution: deformable convolution layer is providing extra spatial prior and improves the results.

(a) Dense U-Net.
(b) D-Con.
(c) D-Exp.
Discussion

Effect of number of parameter (i.e. network size): the more parameters are used, the better segmentation performance is. If the number of parameters is increased further, we could extend the ensemble model to more varieties (currently ensemble 3 models) to further improve segmentation performance.
Thank You!