Deep Learning Approach for Tissue Segmentation in Whole Slide Images

By the students: Haneen Naaran, Amir Bishara
Supervised by: Tal Neoran, Roy Velich

234329 Winter 2022/23
Whole Slide Imaging (WSI)

- A technology to digitally scan and archive glass slides in high resolution.
- Glass-slide of size 20 mm × 15 mm results in image of 80,000 × 60,000 pixels.
- Histopathologists can provide an accurate diagnosis of biopsy specimens based on WSI data.
WSIs for cancer diagnosis

- WSIs are used in cancer diagnosis labs.

- Predicting hormonal receptor status
  - Patch level classifiers.

- Supervised learning models are used; patch level labels.

- But.. WSI has tons of patches!

- We need to discard background patches
  -> tissue segmentation.
Patch Extraction from WSI

* Tissue is stained with Hematoxylin & Eosin

Patches are sent for diagnosis pipeline
Thumbnails Segmentation

- WSIs are large – Gigapixel images
  - Pixelwise segmentation is infeasible.

- We segment (pixel level) **thumbnails of WSI**
  - Downsampling version of WSI – on average 1500x1500 pixels.

- Thumbnail segmentation is then upsampled back to the original WSI
  - Patches are extracted from WSI accordingly.
OTSU thresholding

• The algorithm iteratively searches for the threshold that minimizes the within-class variance

• Heuristic approach

• Pros
  • Fast and simple, no need for data
  • Usually works well

• But..
  • Poor generalization
  • Requires manual tuning
  • Misclassifies artifacts

Otsu segmentation of WSI thumbnail on image from Carmel
Deep learning Approach

- Train neural network on labeled images (thumbnails)
- Better generalization
- No need for extra modifications in inference time
- Robustness
OTSU as Ground Truth for DL models

• We will use OTSU masks as ground truth for DL models, why?
  • OTSU masks are easily available and “usually” accurate.
  • OTSU masks in the data are fine tuned.

• Potential Pitfalls:
  • DL models could “overfit” to the OTSU method.

• Alternatives:
  • Unsupervised methods
Model Architecture

• U-net architecture.
Training

• Breast tissue scans – **OTSU as a ground truth segmentation**
• U-net is trained by optimizing pixel-wise cross entropy loss
• Trained on random 512x512 crops from thumbnails.
• Each batch is 5 crops, 512x512 each.
• Augmentations:
  • Horizontal and vertical flips, Rotations, Color jitter

A batch of 5 crops 512x512 each, after augmentations
Inference

• Output of U-net is a mask of scores.
• Segmentation mask is obtained by thresholding the scores.
• We chose naive threshold of 0.5.
• How to solve class imbalance issue?
  • Using ROC curve.
  • Appropriate loss function.
Handling Class Imbalance

- Approach 1 –
  - Use ROC curve for threshold selection

- Approach 2 –
  - Use appropriate loss function – Focal Loss.

Focal Loss

\[
CE(p_i) = -\log(p_i)
\]

\[
FL(p_i) = -(1 - p_i)^\gamma \log(p_i)
\]
Sliding Window Inference

• Inference in a sliding window manner
  • Memory constraint

• Sliding windows of shape $512 \times 512$ (non over lapping).
  • Reflection padding.
Metrics we are using

• Visualizations
  • By looking at the visual comparisons (next slides).
• Dice score
• Pixel accuracy (averaged upon classes)
Experiment 1 – Train on TCGA

• **Aim of the experiment:**
  • Train on clean dataset, see generalization on real data.
  • Trained on **TCGA** ~3300 images, 0.15 validation ratio.
Experiment 1 Results – Numeric Metrics

- Does not overfit
- Trains for 28 epochs (early stopped), took ~3h
Experiment 1 Results
– Numeric Metrics

• Tested on Carmel, ABCTB and Haemek, 100 images each

<table>
<thead>
<tr>
<th>Data</th>
<th>Pixel Accuracy</th>
<th>Dice Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA (test set)</td>
<td>0.9794</td>
<td>0.9507</td>
</tr>
<tr>
<td>Carmel</td>
<td>0.9925</td>
<td>0.9301</td>
</tr>
<tr>
<td>HAEMEK</td>
<td>0.9902</td>
<td>0.9174</td>
</tr>
<tr>
<td>ABCTB</td>
<td>0.9391</td>
<td>0.8849</td>
</tr>
</tbody>
</table>
Experiment 1 – Train on TCGA

Visual Results - ABCTB

Original Image(ABCTB)

Visual Comparison

- Red: Pixel predicted as Tissue by U-net and Otsu
- Green: Pixel predicted as Tissue by U-net, Background by Otsu
- Blue: Pixel predicted as background by U-net, tissue by Otsu
Experiment 1 – Train on TCGA

Visual Results - Haemek

Original Image (Haemek)

Visual Comparison

- Red: Pixel predicted as Tissue by U-net and Otsu
- Green: Pixel predicted as Tissue by U-net, Background by Otsu
- Blue: Pixel predicted as background by U-net, tissue by Otsu
Experiment 1 – Train on TCGA

Visual Results - Carmel

*Model is trained on TCGA datset, image is from Carmel dataset*
But..

*Model is trained on TCGA*
Experiment 1 – Train on TCGA

Conclusions

- Good generalization on tissue regions
- Fails on artifacts
  - Because data is clean
- Should train on dataset with artifacts
Experiment 2 – Train on Carmel

• **Aim of the experiment:**
  • Does training on data with artifacts enhance the U-net robustness?
• Trained on **Carmel** – batches 1 – 6 ~6600 images in total
• 0.15 validation ratio.
Experiment 2 – Train on Carmel

Visual Results - Carmel

Original image (Carmel)  Visual comparison

- Pixel predicted as Tissue by U-net and Otsu
- Pixel predicted as Tissue by U-net, Background by Otsu
- Pixel predicted as background by U-net, tissue by Otsu
Experiment 2 Visual Results - ABCTB

Original image from ABCTB dataset

Predicted mask

visual comparison between U-net and Otsu segmentation

When trained on TCGA:
Conclusion from training on Carmel

• Training on Carmel does not solve the artifacts issue.
• Why?
  • Carmel train set has mistakes ground truth segmentations!
  • Artifacts don’t appear in random crops
• Should train on correctly segmented artifacts.
Custom transform – Add random artifacts

- We want to ensure training on artifacts, with correct segmentations.
- Idea: Add random artifacts on clean train images.
- Introduce 3 class segmentation – tissue, background, artifact.
Custom transform – Add random artifacts

Created a dataset of artifacts:
- Random crops of artifacts from Carmel, ABCTB (train sets)
- Some from external datasets - like MNIST
- Created segmentations for them (OTSU)
- 123 crops in total

Crops of artifacts with their OTSU segmentation
Custom transform

• Random resize the artifact image
• Randomly add the **segmented** artifact, with $p = 0.2$
• Mask now has 3 classes
Experiment 3 – train on TCGA with added artifacts

• **Aim:** Does training on clean dataset with added artifacts give better robustness?

• Trained on TCGA again, with added artifacts.

• Model now predicts 3 classes.
Experiment 3 – Visual result - ABCTB

* We ensured that artifacts in test images are not used in the custom transform.
Experiment 3 – Visual result - Carmel

Original image  Otsu comparison
Experiment 3 – Visual result - Haemek

- Predicted mask
- Original image (Haemek)
- Comparison with Otsu

Legend:
- Red: Pixel predicted as Tissue by U-net and Otsu
- Green: Pixel predicted as Tissue by U-net, Background by Otsu
- Blue: Pixel predicted as background by U-net, tissue by Otsu
- Yellow: Pixel predicted as Artifact by U-net
Explorations – Unet strengths

Pixel predicted as Tissue by U-net and Otsu
Pixel predicted as Tissue by U-net, Background by Otsu
Pixel predicted as background by U-net, tissue by Otsu
Pixel predicted as Artifact by U-net

Original image from ABCTB
Explorations – Unet strengths

U-Net Predicted mask  Original image from ABCTB  Otsu mask
Explorartions – Unet weaknesses
Explorations – Unet weaknesses

Images taken from Ipatimup:
Conclusions and Future Work

• Deep learning approach proved better generalization, compared to OTSU.
• Hence, could potentially replace OTSU method in patch extraction phase.
• Enables artifact detection (when trained on proper data).
• What’s next?
  • Enhance ground truth masks.
  • Exploring unsupervised segmentation methods.
  • Find better ways to evaluate model performance
  • Expand the study to other types of tissues and stains.