Extracting quantitative information from microscopy data: Benchmarking & controls

Ida-Maria Sintorn
Centre for Image Analysis
Swedish University of Agricultural Sciences
ida@cb.uu.se

How do you know that your measures and results are valid?
Some suggestions on how you can validate your methods and measures

Who am I?
• Researcher, Assistant Professor at CBA since 2008
• Research focus: segmentation, texture analysis, automated image acquisition for HCA and 2D TEM image analysis
• PhD in image analysis 2005
• 2 years as image analysis researcher at industrial research institute in Sydney (HCA methods)
• 2 years at Vironova AB, responsible for image analysis research & development

What is image analysis?
→ Extraction of meaningful information from (digital) images...
...by means of digital image processing/analysis techniques.

Image analysis fundamental steps

Benchmarking
To evaluate/compare the methods used to extract measures with the "ground truth" or standard methods

Common benchmarking methods:
• Counting
• Image wide segmentation
• Object outlining
• ranking
• Control/reference samples
Ground truth

- Comes from remote sensing
- The content of each pixel in the image is measured on the ground
- Widely used as a term describing the "truth or correct result"

Counting

- Compare the number of objects found in each image with manual counting. The manual counting is then the ground truth
- Compare the number of correctly detected objects as well as detected false objects and undetected objects as manually marked.

True and False positives and negatives

<table>
<thead>
<tr>
<th>Ground truth/expected</th>
<th>true objects</th>
<th>false objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>true positives - correct detections</td>
<td>tp</td>
<td>fp</td>
</tr>
<tr>
<td>false positives - incorrect detections</td>
<td>fn</td>
<td>tn</td>
</tr>
</tbody>
</table>

- tp = true positives = correct detections
- fp = false positives = incorrect detections
- fn = false negatives = undetected
- tn = true negatives = correctly undetected

Precision and Recall

- Precision is the fraction of retrieved (detected) objects that are relevant (true).
  \[
  \text{Precision} = \frac{tp}{tp + fp}
  \]
- Recall is the fraction of relevant (true) objects that are retrieved (detected).
  \[
  \text{Recall} = \frac{tp}{tp + fn}
  \]
- For the example image, precision = \(81/(81 + 12) = 0.87\)
- Probability that a (randomly selected) detected object document is true.
- For the example image, recall = \(81/(81 + 16) = 0.84\)
- Probability that a (randomly selected) true object is detected in a search.
F-measure

- Combines precision and recall
  \[ F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]
- For the example image
  \[ F = 2 \cdot \frac{81 \cdot 81}{81 + 81} = 0.85 \]

sensitivity, specificity, accuracy

- Sensitivity: recall
- Specificity: True neg. rate
  \[ \text{Recall} = \frac{tp}{tp + fn} \]
  \[ \text{Specificity} = \frac{tn}{tn + fp} \]
- Accuracy is the proportion of true results
  \[ \text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \]
- For the example image, specificity = 31/(31+16) = 0.70
  \[ \text{Accuracy} = \frac{31 + 16}{31 + 16 + 31 + 16} = 0.30 \]

How should one decide what a tn is for an example like this?

For the example image
\[ F = 2 \cdot \frac{81 \cdot 81}{81 + 81} = 0.85 \]

Object outlining

If the position of the boundaries are important.
- Can be evaluated as just described (segmented regions in image) but per object instead
  - Or by measuring the distance from the border of the automatically detected outline to the manually segmented outline.

Image wide segmentation benchmarking

- The region(s)/object(s) of interest are outlined or marked manually
- The manually segmented image is compared to the automatically segmented image (True positive area, false positive area, false negative area, true negative area)
Use a distance transform from the manually segmented border and analyze the values under the automatically segmented border.

For example, look at the percentage of pixels closer or further away than a certain distance from the manual border.

**Ranking**

- If a quantitative ground truth cannot be obtained to evaluate automatic measures, ranking of images can sometimes be useful.

**Manual/expert comparison**

- One or more experts
- One or more time points
- Random image presentation order

What is relevant depends on the question.

**Control or reference samples**

- Samples which you use to automatically compare/evaluate your method or effect on.
  - Negative control: No effect. No added substrate
  - Positive control: Clear effect. Added known substrate.
  - Reference sample: Sample with known characterized content

**Z-factor**

- Used in HCS as a measure of usefulness of measure
- Mean and std of positive and negative controls

\[ Z = \frac{1}{6} \left( \frac{\text{mean std of all doses}}{\text{std of all doses}} \right) \]

**V-factor**

- Alternative to Z-factor for dose response curves.
- It takes the variance of all doses into consideration

\[ V = 1 - 6 \left( \frac{\text{mean std}}{\text{abs}(m_p - m_n)} \right) \]
Example

• Known drug at 150nM conc. induces translocation of protein from cytoplasm to nucleus
• Want to search for alternative drugs with the same effect but less side effects
• Dose response images with DNA (nuclear) stain and GFP stained protein

What shall we measure?

• Intensity of GFP in cytoplasm
• Intensity of GFP in nucleus

We need to detect the nuclei and cytoplasms. How can we do that?

• Use DNA stain to segment nuclei
• Approximate the cytoplasm by ring around each nucleus

Mean cytoplasm intensity

Mean nucleus intensity

Ratio cytoplasm/nucleus

Taking the ratio produces a measure with good separation between control groups and reduced variance for each dose. Z=0.62
Repeated experiments

- Repetitions to get variance -> measure of how reliable your extracted measures are.
- The more steps that can/are repeated the better: sample prep, imaging, analysis etc.

Robustness

- Sensitivity/robustness to parameter settings: image quality, preparation, noise, etc.
- Test and evaluate result when varying the parameters
- Test performance when adding (artificial) noise
- Evaluate for repeated experiments

Conclusion/summary

- Plan your experiment
- Think/work through all analysis steps to understand the variation
- Look at your images
- Apply suitable benchmarking/quality control methods

⇒ Understand what you are doing and look at the images