Layout Evaluation Tests

Christian Clausner

August 2010

Open problems are marked **bold red**.

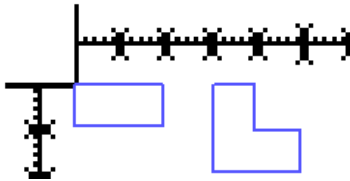## Contents

Region Area

- Ground-truth: test_1_region_area.xml
- Segmentation result: test_1_region_area.xml
- The ground-truth contains 2 regions. One with 200 pixels size and the other with 300 pixels.
- Checked the overall region area in the statistics: 500px ✓

Region Foreground Pixel Count

- Ground-truth: test_1_region_pixelcount.xml
- Segmentation result: test_1_region_pixelcount.xml
- The ground-truth contains a region with one element consisting of 250 black pixels.
- Checked the overall region pixel count in the statistics: 250px ✓
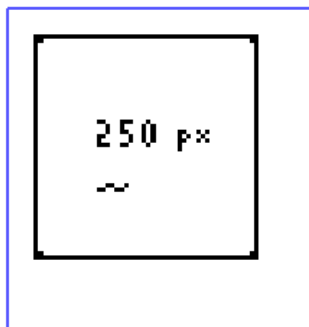
Merge Area / Pixel Count

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_merge_seg.xml
- The ground-truth has 6 regions. The segmentation result merges them to three regions.
- Checked the number of merges: 5 ✓
- Checked which region is merged with which other region. ✓
- Checked area and foreground pixel count of the overlaps. ✓
- Checked overall area and foreground pixel count of one merge:
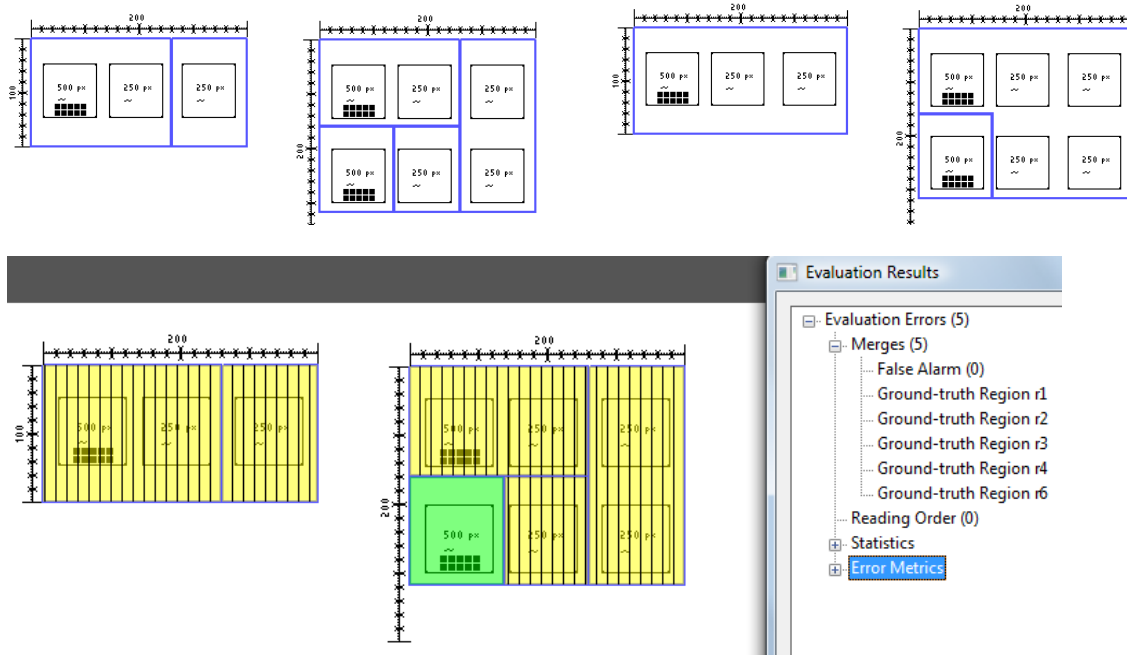
Merge False Alarm

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_merge_seg2.xml
- The ground-truth has two regions that are overlapped by a segmentation result region. But the segmentation result region doesn't contain black pixels.
- Checked if there are false alarm merges. ✓
- Checked the areas. ✓
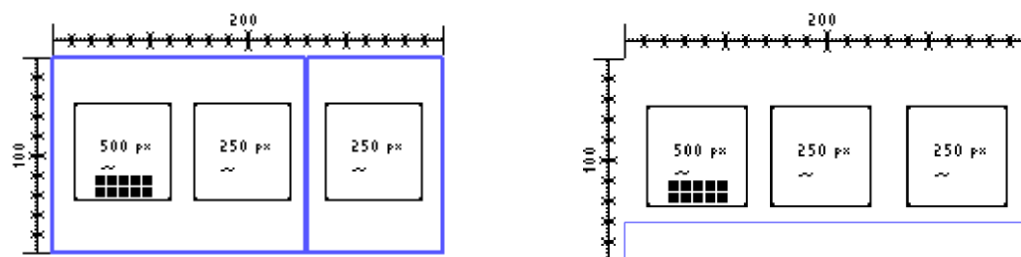
Allowable Merge

- Ground-truth: test_3_merge_gt.xml
- Segmentation result: test_3_merge_seg.xml
- The ground-truth contains several pairs of paragraphs. The segmentation result merges these pairs. The merges are partly allowable, partly not. Different reading orientations and reading directions have been used.
- Checked the allowable flag. ✓
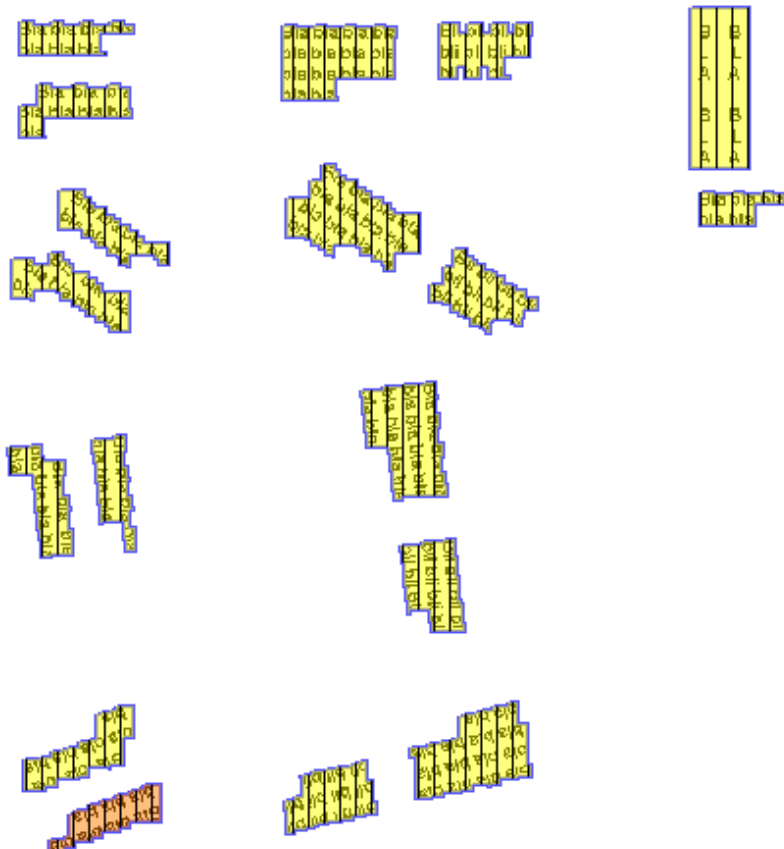
Allowable Merge 2

- Ground-truth: test_4_merge_gt.xml
- Segmentation result: test_4_merge_seg.xml
- Some additional reading orientations (-30, -80 and -160 degrees).

Checked the allowable flag. ✓

Split Area / Pixel Count

- Ground-truth: test_2_split_gt.xml
- Segmentation result: test_2_split_seg.xml
- The ground-truth has 3 regions. The segmentation result splits them into six regions.
- Checked the number of splits: 2 ✓
- Checked which region is split into which regions. ✓
- Checked area and foreground pixel count of the overlaps. ✓
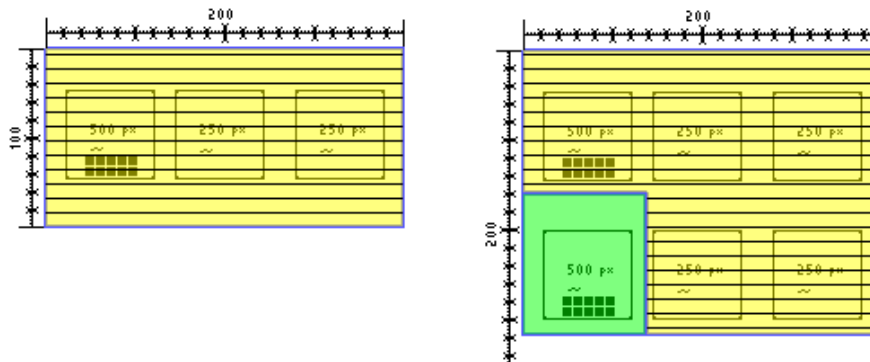
Split False Alarm

- Ground-truth: test_2_split_gt.xml
- Segmentation result: test_2_split_seg2.xml
- The ground-truth has one that is overlapped by two segmentation result regions. But the segmentation result regions don't contain black pixels.
- Checked if there are false alarm splits. ✓
- Checked the areas. ✓
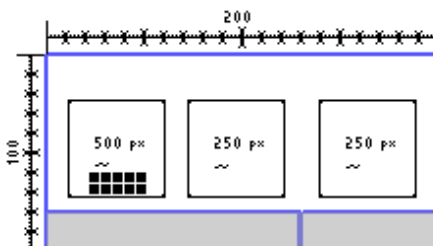
Allowable Split

- Ground-truth: test_3_split_gt.xml
- Segmentation result: test_3_split_seg.xml

- The ground-truth contains regions that are split into pairs in the segmentation result. The splits are partly allowable, partly not. Different reading orientations and reading directions have been used.
- Checked the allowable flag. ✓

Miss Area / Pixel Count

- Ground-truth: test_1_miss_gt.xml
- Segmentation result: test_1_miss_seg.xml
- The ground-truth contains a region that is not overlapped from any segmentation result region.
- Checked that the miss is there. ✓
- Checked the area. ✓
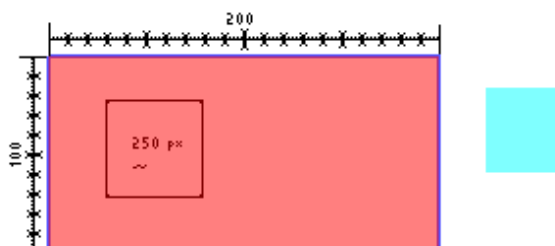
Miss False Alarm

- Ground-truth: test_1_miss_gt2.xml
- Segmentation result: test_1_miss_seg.xml
- The ground-truth contains a region that is not overlapped from any segmentation result region. But the ground-truth region doesn't contain black pixels.
- Checked that the miss is displayed as false alarm. ✓
- Checked the area. ✓
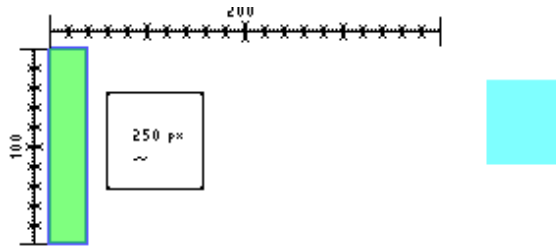
Partial Miss Area / Pixel Count

- Ground-truth: test_2_part_miss_gt.xml
- Segmentation result: test_2_part_miss_seg.xml
- The ground-truth contains a region including three glyphs. The segmentation result has a region including only two of the three glyphs.
- Checked that the partial miss is recognized.
- Checked area and pixel count: area: 7000px ✓, count: 250px ✓

Partial Miss False Alarm

- Ground-truth: test_2_part_miss_gt.xml
- Segmentation result: test_2_part_miss_seg2.xml
- The ground-truth contains a region including three glyphs. The segmentation result has a region including all three glyphs, but the region is smaller than the ground-truth region.
- Checked that the partial miss is recognized as false alarm. ✓
- Checked the area. ✓
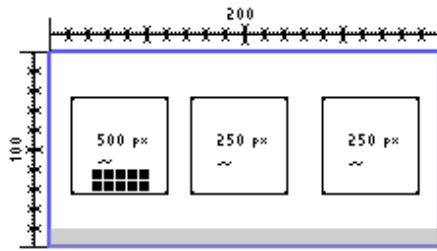
Misclassification Area / Pixel Count

- Ground-truth: test_2_ misclass_gt.xml
- Segmentation result: test_2_misclass _seg.xml
- The ground-truth contains a text paragraph region, a text header region and an image region. The segmentation result has (in the same order) a graphic region, a text footer region and a noise region.
- Checked that the misclassifications are listed and explained. ✓
- Checked area and pixel count. ✓
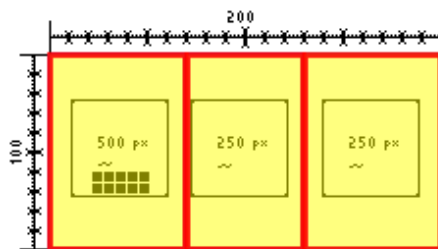
Multi-Misclassification

- Ground-truth: test_2_misclass _gt2.xml
- Segmentation result: test_2_misclass_seg.xml
- The ground-truth contains a text paragraph region. The segmentation result has split the ground-truth region into three regions: a graphic region, a text footer region and a noise region.
- Checked that the misclassifications are listed and explained. ✓
- Checked area and pixel count. ✓
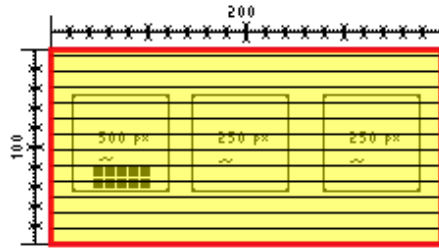
Misclassification False Alarm
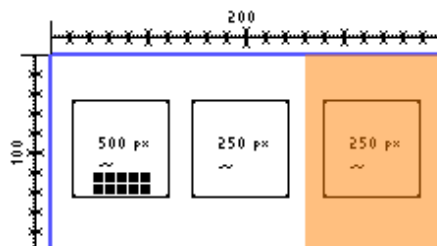
- Ground-truth: test_2_ misclass_gt.xml
- Segmentation result: test_2_misclass _seg2.xml
- The ground-truth contains a text paragraph region, a text header region and an image region. The segmentation result has (in the same order) a graphic region, a text footer region and a noise region. The segmentation result regions don't contain black pixels.
- Checked that the misclassifications are listed as false detection. ✓
- Checked area and pixel count. ✓
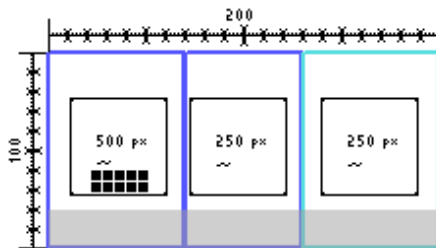
False Detection Area / Pixel Count

- Ground-truth: test_1_false_detection_gt.xml
- Segmentation result: test_1_false_detection_seg.xml
- The segmentation result contains a region that is not overlapped from any ground-truth region.
- Checked that the false detection is there. ✓
- Checked the area. ✓
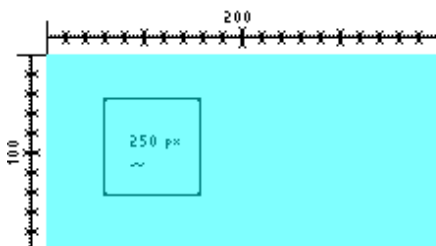
False Detection False Alarm

- There is no false alarm for false detection. Even if there are no black pixels in the false detected region, it is not flagged as false detection. Checked. ✓

Statistics Region Area / Pixel Count / Number of Regions

- Ground-truth: test_2_statistics_gt.xml
- Segmentation result: test_2_statistics_gt.xml
- The ground-truth contains regions of different types containing different glyphs.
- Checked overall area, pixel count and number of regions. ✓
- Checked area, pixel count and number of regions per region type. ✓

Statistics Image Area / Pixel Count

- Ground-truth: test_5_statistics.xml
- Segmentation result: test_5_statistics.xml
- The size of the document image is 500 * 500 pixels and it contains glyphs with a total of 5250 black pixels.
- Checked area and pixel count. ✓

Merge Overall Error / Success Rate

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_merge_seg.xml
- The ground-truth has 6 text regions. The segmentation result merges them to three regions.
- Count based error:
  - There are two merges, one including two regions and one including 3 regions.
  - A merge of two regions counts as 2; a merge of three regions counts as 3.
  - Checked the count based error: 5 ✓
  - The fifty-percent-x for the success rate lies at #GT regions / 2 = 3
  - The success rate function is then $1 / (0.333*x + 1)$
  - Checked the success rate: 37.5% ✓
  - Checked the overall error. ✓
  - Checked the relative error. ✓
  - The basic influence is $(5 * (1\text{-merge rate}) + 1) / 6 = 4.125 / 6 = 68.75\%$
  - The other influences are all 1/6
  - The overall arithmetic success rate: $(0.6875 * 37.5\% + 5 * 1/6 * 100\% + 100\%) / (0.6875 + 5*1/6 + 1) = (25.78 + 83.3 + 100) / (2.5208) = 82.9 \%$

- - - o The harmonic success rate: (0.6875 + 5*1/6 + 1) / (0.6875/37.5% + 5*(1/6)/100% + 1/100%) = 2.5208 / (0.01833 + 0.008333 + 0.01) = 2.5208 / 0.03667 = 68.7%
      - o Checked the overall success rate: arithmetic: 83.0% ✓   harmonic: 68.8% ✓
  - Area based error:
    - o The merge of the two regions (overlap weight is 1):
      - ▪ Left region merged with the right one. The weighted area is 250
      - ▪ Right region merged with the left one. The weighted area is 250
    - o The merge of the three regions (overlap weight is 0.5):
      - ▪ Top region merged with the right one. The 'merge area' is 500. The weighted area is 250.
      - ▪ Top region merged with the bottom one. The 'merge are' is 250. The weighted area is 125.
      - ▪ Right region merged with top one. Weighted area: 250
      - ▪ Right region merged with bottom one. Weighted area: 125
      - ▪ Bottom region merged with top one. Weighted area: 125
      - ▪ Bottom region merged with right one. Weighted area: 125
    - o Checked the area based error: 1500 ✓
    - o The fifty-percent-x lies at overall region area / 4 = 1500 / 2 = 750.
    - o The success rate function is then 1 / (0.001333*x+1)
    - o Checked the success rate: 33.3% ✓
    - o Checked the overall error. ✓
    - o Checked the relative error. ✓
    - o The basic influence is (5 * (1-merge rate) + 1) / 6 = 4.335 / 6 = 0.7225
    - o The other influences are all 1/6
    - o The overall arithmetic success rate: (0.7225 * 33.3% + 5 * 1/6 * 100% + 100%) / (0.7225 + 5*1/6 + 1) = (24.06 + 83.3 + 100) / (2.556) = 81.1 %
    - o The harmonic success rate: (0.7225 + 5*1/6 + 1) / (0.7225/33.3% + 5*(1/6)/100% + 1/100%) = 2.556 / (0.0217 + 0.008333 + 0.01) = 2.556 / 0.0400 = 63.9%
    - o Checked the overall success rate: arithmetic: 81.2% ✓   harmonic: 63.9% ✓
  - Checked the results per region type. ✓
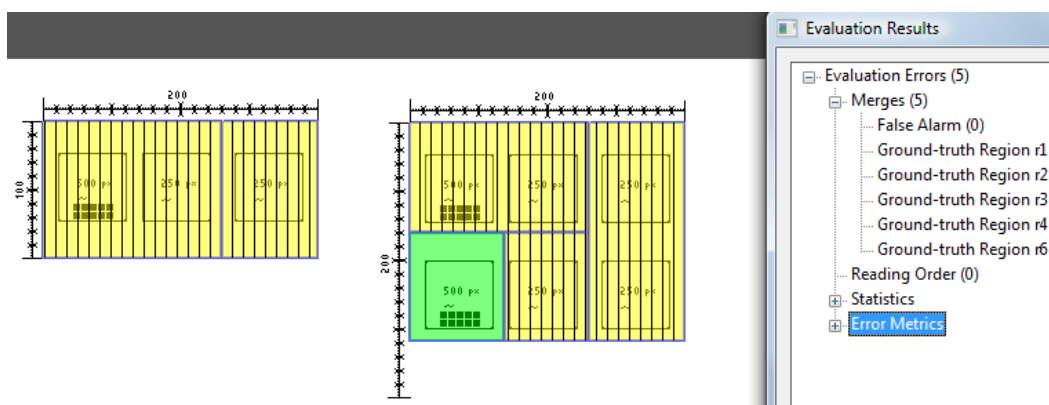
Split Overall Error / Success Rate

- Ground-truth: test_2_split_gt.xml
- Segmentation result: test_2_split_seg.xml
- The ground-truth has 3 regions. The segmentation result splits them into six regions.
- Count based error:
  - There are two splits. One ground-truth region is split into 2 parts and another is split into 3 parts.
  - A split count equals the number of regions it splits into (here 2 and 3).
  - Checked the count based error: 5 ✓
  - The fifty-percent-x for the success rate lies at #GT regions / 2 = 3
  - The success rate function is then $1 / (0.333*x + 1)$
  - Checked the success rate: 37.5% ✓
  - Checked the overall error. ✓
  - Checked the relative error. ✓
  - The basic influence is (5 * (1-split rate) + 1) / 6 = 4.125 / 6 = 68.75%
  - The other influences are all 1/6
  - The overall arithmetic success rate: (0.6875 * 37.5% + 5 * 1/6 * 100% + 100%) / (0.6875 + 5*1/6 + 1) = (25.78 + 83.3 + 100) / (2.5208) = 82.9 %
 - The harmonic success rate: (0.6875 + 5*1/6 + 1) / (0.6875/37.5% + 5*(1/6)/100% + 1/100%) = 2.5208 / (0.01833 + 0.008333 + 0.01) = 2.5208 / 0.03667 = 68.7%
  - Checked the overall success rate: arithmetic: 83.0% ✓ harmonic: 68.75% ✓
- Area based error:
  - The split into two regions:
    - Left split region: area is 250
    - Right split region: area is **750**
    - the relative split area is 750 / 1000 = 3/4 = 0.75
    - the weighted area is (1-0.75) * 2 * 1000 = 500
  - The split of the three regions:
    - Top split region: area is **750**
    - Right split region: area is 500
    - Bottom split region: area is 250
    - the relative split area is 750/1500 = 0.5
    - the weighted area is (1-0.5) * 3 * 1500 = 2250
  - Checked the area based error: 2750 ✓
  - The fifty-percent-x lies at overall region area / 2 = 3000 / 2 = 1500.
  - The success rate function is then $1 / (0.000667*x+1)$
  - Checked the success rate: 35,3% ✓
  - Checked the overall error. ✓
  - Checked the relative error. ✓
  - The basic influence is (5 * (1-split rate) + 1) / 6 = 4.235 / 6 = 0.7058
  - The other influences are all 1/6
  - The overall arithmetic success rate: (0.7058 * 35.3% + 5 * 1/6 * 100% + 100%) / (0.7058 + 5*1/6 + 1) = (24.91 + 83.3 + 100) / (2.539) = 82.0 %
  - The harmonic success rate: (0.7058 + 5*1/6 + 1) / (0.7058/35.3% + 5*(1/6)/100% + 1/100%) = 2.539 / (0.0200 + 0.008333 + 0.01) = 2.539 / 0.03832 = 66.3%
  - Checked the overall success rate: arithmetic: 82.0% ✓ harmonic: 66.2% ✓

- Checked the results per region type. ✓
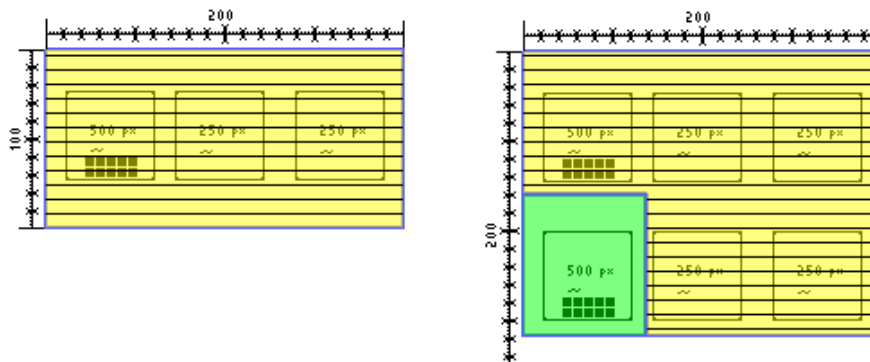
Miss Overall Error / Success Rate

- Ground-truth: test_2_miss_gt.xml
- Segmentation result: test_2_miss_seg.xml
- The ground-truth has 3 regions (2 text, 1 image). The segmentation result has only one text region.
- Count based error:
  - Two regions are missing
  - Checked the count based miss error: 2 ✓
  - The fifty-percent-x equals the number of ground-truth regions / 2 = 1.5
  - The success rate function is: $1 / (0.667*x + 1)$
  - Checked the success rate: 42.9% ✓
  - Checked the overall error. ✓
  - Checked the relative error. ✓
  - The basic influence is (5 * (1-miss rate) + 1) / 6 = 3.855 / 6 = 0.6425
  - The other influences are all 1/6
  - The overall arithmetic success rate: (0.6425 * 42.9% + 5 * 1/6 * 100% + 100%) / (0.6425 + 5*1/6 + 1) = (27.56 + 83.3 + 100) / (2.4758) = 88.2 %
  - The harmonic success rate: (0.6425 + 5*1/6 + 1) / (0.6425/42.9% + 5*(1/6)/100% + 1/100%) = 2. 4758 / (0.01498 + 0.008333 + 0.01) = 2. 4758 / 0.0333 = 74.3%
  - Checked the overall success rate: arithmetic: 85.2% ✓   harmonic: 74.3% ✓
- Area based error:
  - Area of missed regions: 500 + 250 pixels = 750px
  - Checked the area based miss error: 750 ✓
  - The fifty-percent-x lies at overall region area / 2 = 1750 / 2 = 875
  - The success rate function is then $1 / (0.00114*x+1)$
  - Checked the success rate: 53.8% ✓
  - Checked the overall error. ✓
  - Checked the relative error. ✓
  - The basic influence is (5 * (1-miss rate) + 1) / 6 = 3.31 / 6 = 0.5517
  - The other influences are all 1/6

- o The overall arithmetic success rate: (0.5517 * 53.8% + 5 * 1/6 * 100% + 100%) / (0.5517 + 5*1/6 + 1) = (29.68 + 83.3 + 100) / (2.385) = 89.3 %
- o The harmonic success rate: (0.5517 + 5*1/6 + 1) / (0.5517/53.8% + 5*(1/6)/100% + 1/100%) = 2.385 / (0.0103 + 0.008333 + 0.01) = 2.385 / 0.0286 = 83.4%
- o Checked the overall success rate: arithmetic: 89.3% ✔ harmonic: 83.5% ✔
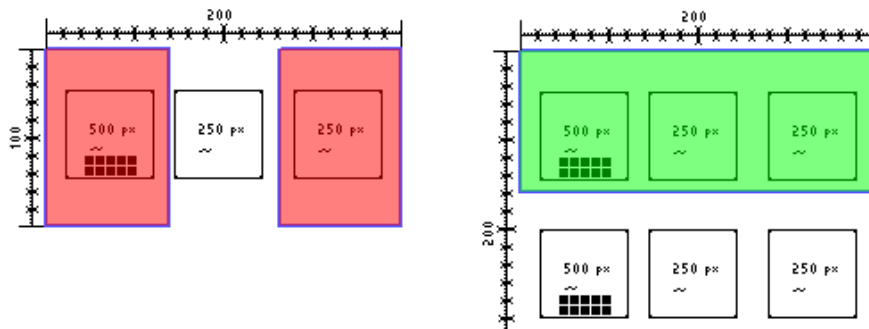- Checked the results per region type. ✔

Partial Miss Overall Error / Success Rate

- Ground-truth: test_2_part_miss_gt2.xml
- Segmentation result: test_2_part_miss_seg3.xml
- The ground-truth has two regions that are not fully recognized by the segmentation result.
- Count based error:
  - o Two regions have a partial miss
  - o Checked the count based partial miss error: 2 ✔
  - o The fifty-percent-x equals the number of ground-truth regions / 2 = 1
  - o The success rate function is: 1 / (x + 1)
  - o Checked the success rate: 33.3% ✔
  - o Checked the overall error. ✔
  - o Checked the relative error. ✔
  - o The basic influence is (5 * (1-partial miss rate) + 1) / 6 = 4.335 / 6 = 0.7225
  - o The other influences are all 1/6
  - o The overall arithmetic success rate: (0.7225 * 33.3% + 5 * 1/6 * 100% + 100%) / (0.7225 + 5*1/6 + 1) = (24.06 + 83.3 + 100) / (2.556) = 81.1 %
  - o The harmonic success rate: (0.7225 + 5*1/6 + 1) / (0.7225/33.3% + 5*(1/6)/100% + 1/100%) = 2.556 / (0.0217 + 0.008333 + 0.01) = 2.556 / 0.0400 = 63.9%
  - o Checked the overall success rate: arithmetic: 81.2% ✔ harmonic: 63.9% ✔
- Area based error:
  - o Missed area: 250 + 750 pixels = 1000px
  - o Checked the area based partial miss error. ✔
  - o The fifty-percent-x lies at overall region area / 2 = 3000 / 2 = 1500
  - o The success rate function is then 1 / (0.000667*x+1)
  - o Checked the success rate: 60% ✔

- o Checked the overall error. ✓
- o Checked the relative error. ✓
- o The basic influence is (5 * (1-partial miss rate) + 1) / 6 = 3.0 / 6 = 0.5
- o The other influences are all 1/6
- o The overall arithmetic success rate: (0.5 * 60.0% + 5 * 1/6 * 100% + 100%) / (0.5 + 5*1/6 + 1) = (30.0 + 83.3 + 100) / (2.333) = 91.4 %
- o The harmonic success rate: (0.5 + 5*1/6 + 1) / (0.5/60.0% + 5*(1/6)/100% + 1/100%) = 2.333 / (0.008333 + 0.008333 + 0.01) = 2.333 / 0.02667 = 87.5%
- o Checked the overall success rate: arithmetic: 91.4% ✓ harmonic: 87.5% ✓
- Checked the results per region type. ✓
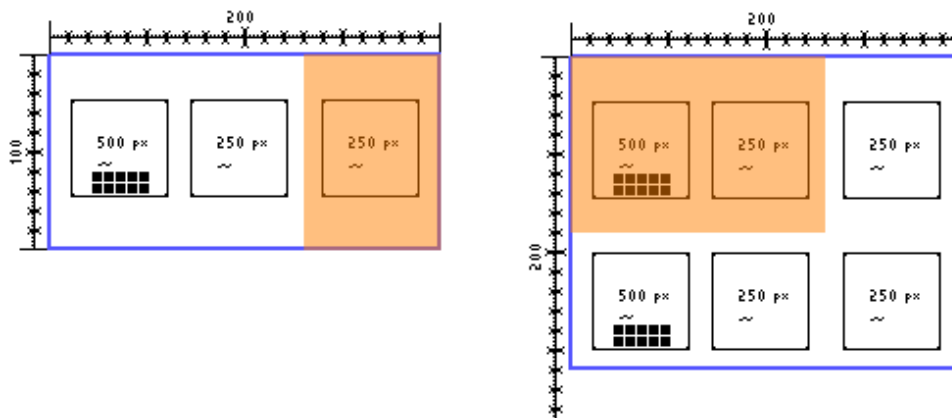
Misclassification Overall Error / Success Rate

- Ground-truth: test_2_misclass_gt.xml
- Segmentation result: test_2_misclass_seg.xml
- The ground-truth contains a text paragraph region, a text header region and an image region. The segmentation result has (in the same order) a graphic region, a text footer region and a noise region.
- Count based error:
  - o Three misclassified regions
  - o Checked the count based misclassification error: 3 ✓
  - o The fifty-percent-x equals the number of ground-truth regions / 2 = 1.5
  - o The success rate function is: 1 / (0.667*x + 1)
  - o Checked the success rate: 33.3% ✓
  - o Checked the overall error. ✓
  - o Checked the relative error. ✓
  - o The basic influence is (5 * (1-misclass rate) + 1) / 6 = 4.335 / 6 = 0.7225
 - o The other influences are all 1/6
  - o The overall arithmetic success rate: (0.7225 * 33.3% + 5 * 1/6 * 100% + 100%) / (0.7225 + 5*1/6 + 1) = (24.06 + 83.3 + 100) / (2.556) = 81.1 %

- - o The harmonic success rate: (0.7225 + 5*1/6 + 1) / (0.7225/33.3% + 5*(1/6)/100% + 1/100%) = 2.556 / (0.0217 + 0.008333 + 0.01) = 2.556 / 0.0400 = 63.9%
    - o Checked the overall success rate: arithmetic: 81.2% ✓   harmonic: 63.9% ✓
- Area based error:
  - o Misclassified area: 500 + 250 + 250 pixels = 1000px
  - o Checked the area based misclassification error. ✓
  - o The fifty-percent-x lies at overall region area / 2 = 1000 / 2 = 500
  - o The success rate function is then $1 / (0.002*x+1)$
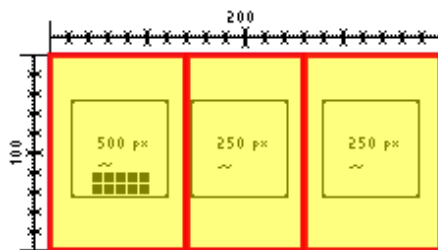 - o Checked the success rate: 33.3% ✓
  - o Checked the overall error. ✓
  - o Checked the relative error. ✓
  - o The basic influence is (5 * (1-misclass rate) + 1) / 6 = 4.335 / 6 = 0.7225
  - o The other influences are all 1/6
  - o The overall arithmetic success rate: (0.7225 * 33.3% + 5 * 1/6 * 100% + 100%) / (0.7225 + 5*1/6 + 1) = (24.06 + 83.3 + 100) / (2.556) = 81.1 %
  - o The harmonic success rate: (0.7225 + 5*1/6 + 1) / (0.7225/33.3% + 5*(1/6)/100% + 1/100%) = 2.556 / (0.0217 + 0.008333 + 0.01) = 2.556 / 0.0400 = 63.9%
  - o Checked the overall success rate: arithmetic: 81.2% ✓   harmonic: 63.9% ✓
- Checked the results per region type. ✓

False Detection Overall Error / Success Rate

- Ground-truth: test_2_false_detection_gt.xml
- Segmentation result: test_2_false_detection_seg.xml
- The ground-truth contains 4 regions (3 text, 1 graphic) and the segmentation result contains four additional regions (2 text, 1 graphic and 1 chart)
- Count based error:
  - o Checked the false detection count error: 4 ✓
  - o The fifty-percent-x equals the number of ground-truth regions = 4
  - o Checked the success rate: 20% ✓
  - o Checked the relative error. ✓
  - o The basic influence is (5 * (1-false detection rate) + 1) / 6 = 5 / 6 = 0.833
  - o The other influences are all 1/6
  - o The overall arithmetic success rate: (0.833 * 20% + 5 * 1/6 * 100% + 100%) / (0.833 + 5*1/6 + 1) = (16.66 + 83.3 + 100) / (2.667) = 75.0 %

- o The harmonic success rate: (0.833 + 5*1/6 + 1) / (0.833/20% + 5*(1/6)/100% + 1/100%) = 2.667 / (0.04165 + 0.008333 + 0.01) = 2.667 / 0.05998 = 44.5%
  - o Checked the overall success rate: arithmetic: 75.0% ✓   harmonic: 44.4% ✓
- Area based error:
  - o False detected area: 250 + 250 + 500 + 250 pixels = 1250px
  - o Checked the area based false detection error. ✓
  - o The fifty-percent-x lies at (image area / 10) = 6536 / 10 = 653.6
  - o The success rate function is then 1 / (0.00153*x+1)
  - o Checked the success rate: 34.3% ✓
  - o Checked the overall error. ✓
  - o Checked the relative error. ✓
  - o The basic influence is (5 * (1-false detection rate) + 1) / 6 = 4.285 / 6 = 0.714
  - o The other influences are all 1/6
  - o The overall arithmetic success rate: (0.714 * 34.3% + 5 * 1/6 * 100% + 100%) / (0.714 + 5*1/6 + 1) = (24.5 + 83.3 + 100) / (2.547) = 81.6 %
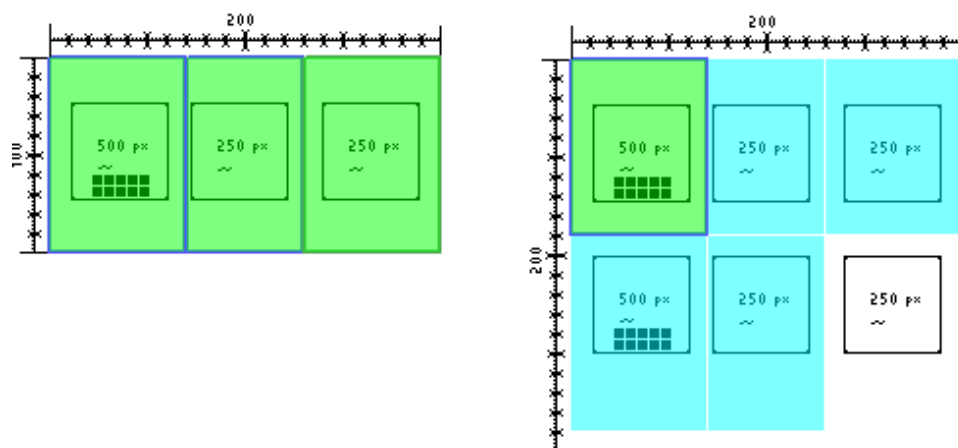 - o The harmonic success rate: (0.714 + 5*1/6 + 1) / (0.714/34.3% + 5*(1/6)/100% + 1/100%) = 2.547 / (0.0208 + 0.008333 + 0.01) = 2.547 / 0.0391 = 65.1%
  - o Checked the overall success rate: arithmetic: 81.6% ✓   harmonic: 65.1% ✓
- Checked the results per region type.

Region Type Weights

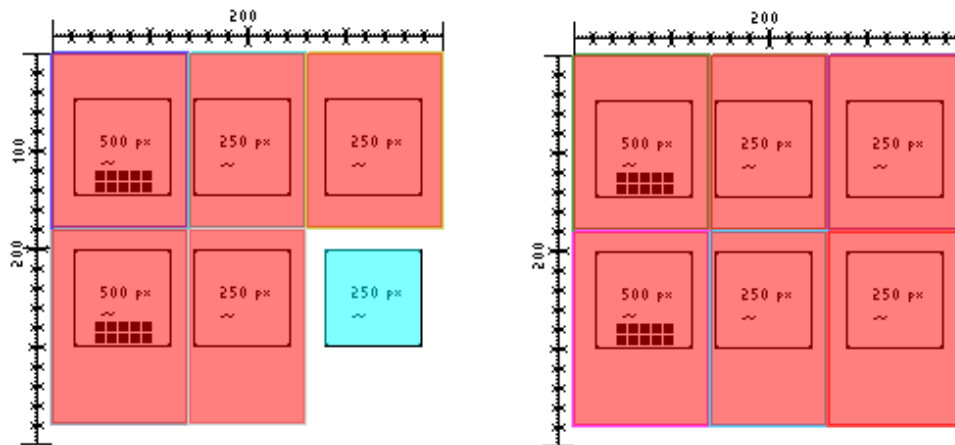- Ground-truth: test_6_miss_gt.xml
- Segmentation result: test_6_miss_seg.xml
- The ground-truth contains one region of each region type. The segmentation result has only one text region (false detection).
- Results without changing weights:
  - o Missed area: 3500
  - o False detected area: 250
  - o Area success rate: arithmetic: 89.1%   harmonic: 77.9%
  - o Misses: 10
  - o False detection: 1

- o Count success rate: arithmetic: 89.6% harmonic: 78.6%
- Text region weight = 0.49
  - o Missed area: 3245
  - o False detected area: 122
  - o Area success rate: arithmetic: 90.1% harmonic: 79.6%
  - o Misses: 9
  - o False Detection: 0
  - o Count success rate: arithmetic: 90.4% harmonic: 79.8%
- Image region weight = 2.03
  - o Missed area: 3757
  - o False detected area: 250
  - o Area success rate: arithmetic: 88.9% harmonic: 76.7%
  - o Misses: 11
  - o False detection: 1
  - o Count success rate: arithmetic: 89.2% harmonic: 77%
- Line drawing region weight = 3.01
  - o Missed area: 4002
  - o False detected area: 250
  - o Area success rate: arithmetic: 88.7% harmonic: 75.6%
  - o Misses: 12
  - o False detection: 1
  - o Count success rate: arithmetic: 89% harmonic: 75.5%
- Graphic region weight = 0.0
  - o Missed area: 3000
  - o False detected area: 250
  - o Area success rate: arithmetic: 89.7% harmonic: 80.3%
  - o Misses: 9
  - o False detection: 1
  - o Count success rate: arithmetic: 89.9% harmonic: 80.2%
- Table region weight = 3.99
  - o Missed area: 4247
  - o False detected area: 250
  - o Area success rate: arithmetic: 88.5% harmonic: 74.6%
  - o Misses: 12
  - o False detection: 1
  - o Count success rate: arithmetic: 88.7% harmonic: 74.1%
- Chart region weight = 0.28
  - o Missed area: 3320
  - o False detected area: 250
  - o Area success rate: arithmetic: 89.3% harmonic: 78.7%
  - o Misses: 9
  - o False detection: 1
  - o Count success rate: arithmetic: 89.8% harmonic: 79.7%
- Separator region weight = 5.04
  - o Missed area: 5520

- False detected area: 250
- Area success rate: arithmetic: 87.8%   harmonic: 69.5%
- Misses: 14
- False detection: 1
- Count success rate: arithmetic: 88.5%   harmonic: 72.6%
- Maths region weight = 0.07
  - Missed area: 3276
  - False detected area: 250
  - Area success rate: arithmetic: 89.4%   harmonic: 79%
  - Misses: 9
  - False detection: 1
  - Count success rate: arithmetic: 89.9%   harmonic: 80.1%
- Noise region weight = 6.01
  - Missed area: 4752
  - False detected area: 250
  - Area success rate: arithmetic: 88.2%   harmonic: 72.5%
  - Misses: 15
  - False detection: 1
  - Count success rate: arithmetic: 88.4%   harmonic: 71.3%
- Unknown region weight = 10.0
  - Missed area: 6000
  - False detected area: 250
  - Area success rate: arithmetic: 87.6%   harmonic: 67.8%
  - Misses: 20
  - False detection: 1
  - Count success rate: arithmetic: 87.6%   harmonic: 65.3%
  - **Note: Fixed a bug here. For all tests above the 'Unknown' region type was missing!**
- Each weight = 0.0
  - Missed area: 0
  - False detected area: 0
  - Area success rate: arithmetic: 100%   harmonic: 100%
  - Misses: 0
  - False detection: 0
  - Count success rate: arithmetic: 100%   harmonic: 100%
- Each weight = 10.0
  - Missed area: 37500
  - False detected area: 2500
  - Area success rate: arithmetic: 78.5%   harmonic: 24.8%
  - Misses: 110
  - False detection: 10
  - Count success rate: arithmetic: 79.6%   harmonic: 25.1%
- Paragraph weight = 0.0
  - Same as text region weight = 0.0 ✓
- Header weight = 0.0
  - No changes ✓

Error Type Weights

- Ground-truth: test_6_error_types_gt.xml
- Segmentation result: test_6_error_types_seg.xml
- The segmentation result contains all error types (merge, split, miss, partial miss, misclassification, false detection).
- Results without changing weights:
  - Area: merge: 500   split: 500   miss: 250   part. miss: 250   miscl.: 750   false d.: 250
  - Area success rate: arithmetic: 83.0%   harmonic: 81.5%
  - Count: merge: 2   split: 2   miss: 1   part. miss: 1   miscl.: 3   false d.: 1
  - Count success rate: arithmetic: 80.7%   harmonic: 78.8%
- Merge weight = 0.49
  - Area: merge: 245   split: 500   miss: 250   part. miss: 250   miscl.: 750   false d.: 250
  - Area success rate: arithmetic: 85.0%   harmonic: 84.2%
  - Count: merge: 0   split: 2   miss: 1   part. miss: 1   miscl.: 3   false d.: 1
  - Count success rate: arithmetic: 82.5%   harmonic: 80.8%
- Split weight = 2.03
  - Area: merge: 500   split: 1014   miss: 250   part. miss: 250   miscl.: 750   false d.: 250
  - Area success rate: arithmetic: 81.0%   harmonic: 79.0%
  - Count: merge: 2   split: 4   miss: 1   part. miss: 1   miscl.: 3   false d.: 1
  - Count success rate: arithmetic: 78.8%   harmonic: 76.8%
- **Note: the following results were calculated with the old merge and split formulas**
- Miss weight = 0.0
  - Area: merge: 750   split: 750   miss: 0   part. miss: 250   miscl.: 750   false d.: 250
  - Area success rate: arithmetic: 84.5%   harmonic: 82.8%
  - Count: merge: 2   split: 2   miss: 0   part. miss: 1   miscl.: 3   false d.: 1
  - Count success rate: arithmetic: 83.3%   harmonic: 80.8%
- Partial miss weight = 3.08
  - Area: merge: 750   split: 750   miss: 250   part. miss: 770   miscl.: 750   false d.: 250
  - Area success rate: arithmetic: 80.3%   harmonic: 79%
  - Count: merge: 2   split: 2   miss: 1   part. miss: 3   miscl.: 3   false d.: 1

- o Count success rate: arithmetic: 77.5%   harmonic: 74.9%
- Misclassification weight = 0.21
  - o Area: merge: 750   split: 750   miss: 250   part. miss: 250   miscl.: 157   false d.: 250
  - o Area success rate: arithmetic: 85.8%   harmonic: 84.6%
  - o Count: merge: 2   split: 2   miss: 1   part. miss: 1   miscl.: 0   false d.: 1
  - o Count success rate: arithmetic: 84.6%   harmonic: 83.7%
- False detection weight = 10.0
  - o Area: merge: 750   split: 750   miss: 250   part. miss: 250   miscl.: 750   false d.: 2500
  - o Area success rate: arithmetic: 76.5%   harmonic: 72.3%
  - o Count: merge: 2   split: 2   miss: 1   part. miss: 1   miscl.: 3   false d.: 10
  - o Count success rate: arithmetic: 74.6%   harmonic: 70.8%
- Each weight = 0.0
  - o Area: all 0
  - o Area success rate: both 100%
  - o Count: all 0
  - o Count success rate: both 100%
- Each weight = 10.0
  - o Area: merge: 7500 split: 7500 miss: 2500 part. miss: 2500 miscl.: 7500 false d.: 2500
  - o Area success rate: arithmetic: 41.5%   harmonic: 30.6%
  - o Count: merge: 20   split: 20   miss: 10   part. miss: 10   miscl.: 30   false d.: 10
  - o Count success rate: arithmetic: 38.8%   harmonic: 27%
- Merge text (paragraph) – image weight = 0.0
  - o Area: merge: 500   split: 750   miss: 250   part. miss: 250   miscl.: 750   false d.: 250
  - o Area success rate: arithmetic: 83.9%   harmonic: 82.8%
  - o Count: merge: 1   split: 2   miss: 1   part. miss: 1   miscl.: 3   false d.: 1
  - o Count success rate: arithmetic: 82.5%   harmonic: 80.8%
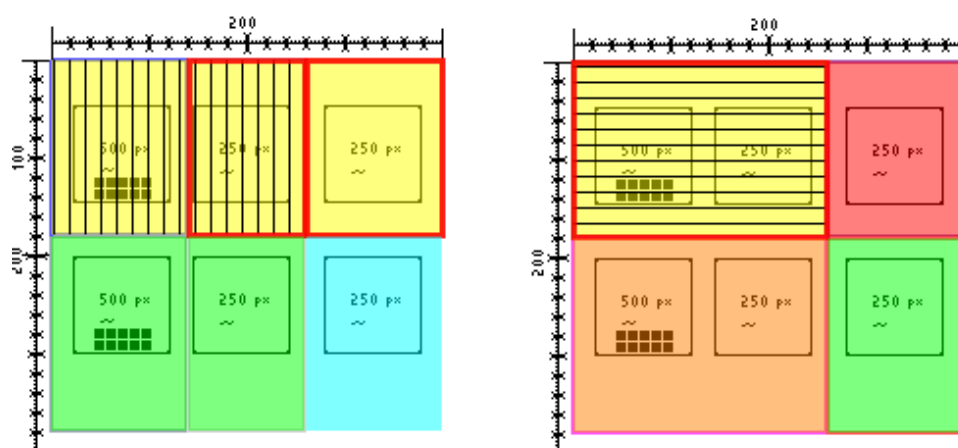
Reading Order Weight

- Ground-truth: test_6_reading_order_gt.xml

- Segmentation result: test_6_reading_order_seg.xml
- The document layout consists of 9 regions. The reading order of ground-truth and segmentation result are significantly different.
- Result without changing weights:
  - Reading order error: 320
  - Reading order success rate: 52.9%
  - Overall success rate: arithmetic: 93.3%   harmonic: 88.7%
- Reading order weight = 0.49
  - Overall success rate: arithmetic: 96.5%   harmonic: 93.7%
- Reading order weight = 2.03
  - Overall success rate: arithmetic: 88.1%   harmonic: 81.7%
- Reading order weight = 0.0
  - Overall success rate: arithmetic: 100%   harmonic: 100%
- Reading order weight = 10.0
  - Overall success rate: arithmetic: 70.6%   harmonic: 64.3%
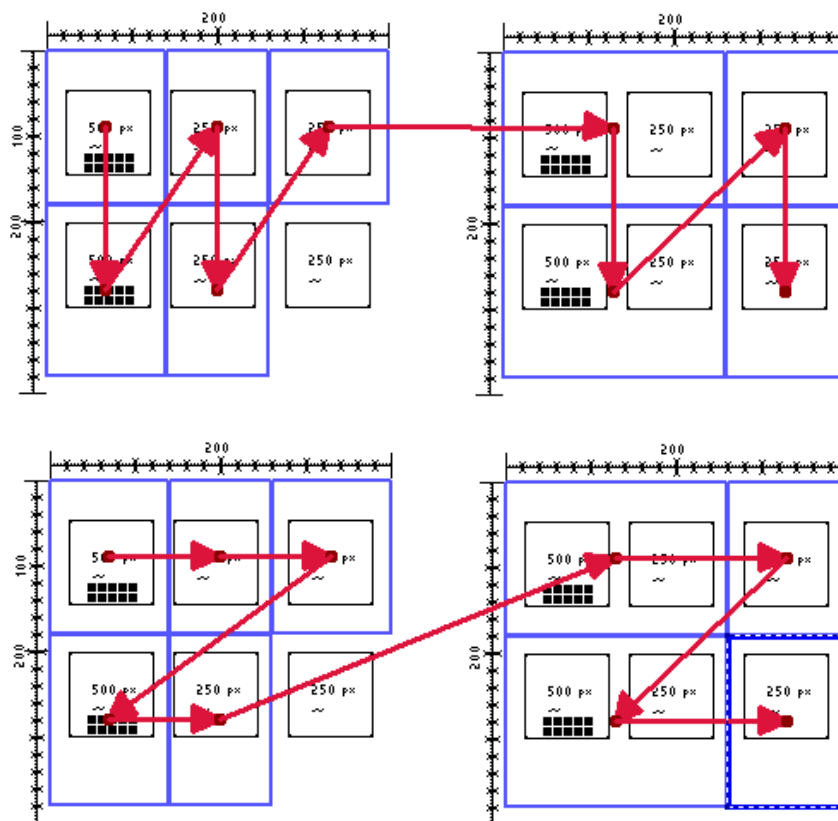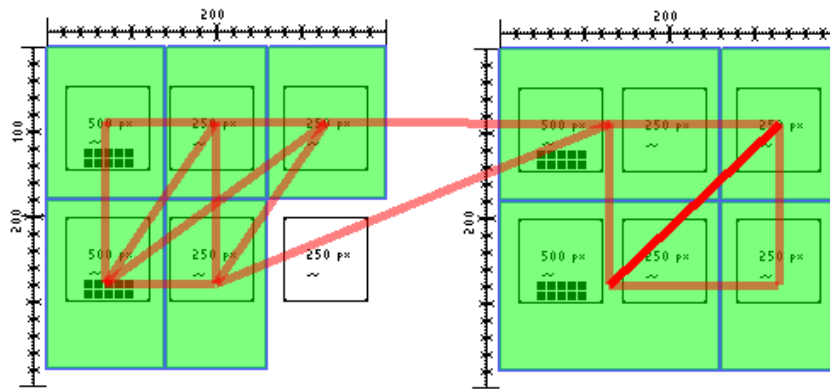
Allowable Merge Weight

- Ground-truth: test_3_merge_gt.xml
- Segmentation result: test_3_merge_seg.xml
- The ground-truth contains several pairs of paragraphs. The segmentation result merges these pairs. The merges are partly allowable, partly not. Different reading orientations and reading directions have been used.
- Result without changing weights:
  - Merge area: 20771
  - Area success rate: arithmetic: 83.3%   harmonic: 70%
  - Merge count: 18
  - Count success rate: arithmetic: 81.9%   harmonic: 69.2%
- Allowable merge weight = 0.49
  - Merge area: 16802
  - Area success rate: arithmetic: 84%   harmonic: 72.8%
  - Merge count: 13
  - Count success rate: arithmetic: 82.8%   harmonic: 72.5%
- Non-allowable weight = 2.03
  - Merge area: 34149
  - Area success rate: arithmetic: 81.9%   harmonic: 62%
  - Merge count: 28
  - Count success rate: arithmetic: 80.6%   harmonic: 62.2%
- Allowable weight = 0.0
  - Merge area: 12989
  - Area success rate: arithmetic: 84.9%   harmonic: 75.7%
  - Merge count: 10
  - Count success rate: arithmetic: 83.9%   harmonic: 75.9%
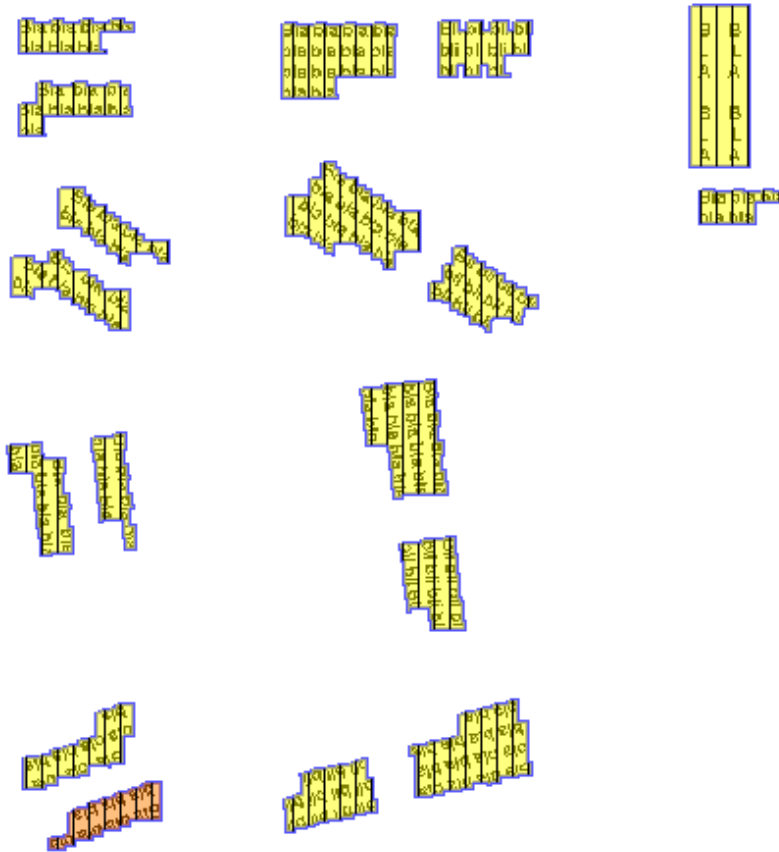
Allowable Split Weight

- Ground-truth: test_3_split_gt.xml
- Segmentation result: test_3_split_seg.xml
- The ground-truth contains regions that are split into pairs in the segmentation result. The splits are partly allowable, partly not. Different reading orientations and reading directions have been used.
- Result without changing weights:
  - Split area: 20771
  - Area success rate: arithmetic: 81.9%   harmonic: 66.7%
  - Split count: 18
  - Count success rate: arithmetic: 81.9%   harmonic: 66.7%
- Allowable split weight = 0.49
  - Split area: 16802
  - Area success rate: arithmetic: 82.6%   harmonic: 69.2%
  - Split count: 13
  - Count success rate: arithmetic: 82.8%   harmonic: 69.7%

Recall

- Ground-truth: test_6_recall_gt.xml
- Segmentation result: test_6_recall_seg.xml
- The ground-truth contains 4 regions of different types. The segmentation result contains also 4 regions but with different types and sizes.
- Type independent recall (non-strict):
    - The overall ground-truth region area is 2750 pixels
    - The recalled area is 2250
    - Checked the recall: 81.8% ✔
    - Checked the region type specific recall (strict). ✔
- Type dependent recall (strict):
    - The recalled area is 1750
    - Checked the recall: 63.6% ✔
- F-measure:
    - strict: 70% ✔   non-strict: 90% ✔
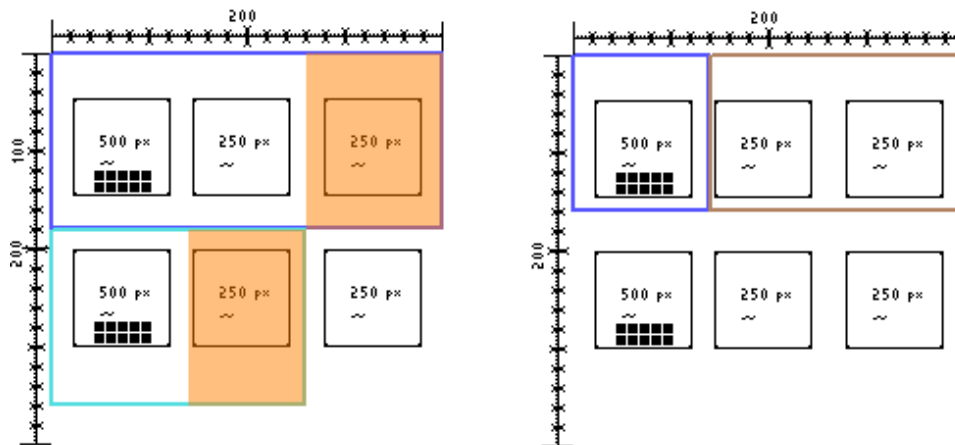
Precision

- Ground-truth: test_6_precision_gt.xml
- Segmentation result: test_6_precision_seg.xml
- Ground-truth and segmentation result simply have been swapped compared to the recall test.
- Type independent precision (non-strict):
    - The overall segmentation result region area is 2750 pixels
    - The recalled area is 2250
    - Checked the precision: 81.8% ✓
    - Checked the region type specific recall (strict). ✓
- Type dependent recall (strict):
    - The recalled area is 1750
    - Checked the recall: 63.6% ✓

For the following tests: 'Use Pixel Area' = FALSE (for area based error)

Merge False Alarm ('Use Pixel Area' = FALSE)

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_merge_seg2.xml
- The ground-truth has two regions that are overlapped by a segmentation result region. But the segmentation result region doesn't contain black pixels.
- Checked if there are NO false alarm merges. ✓
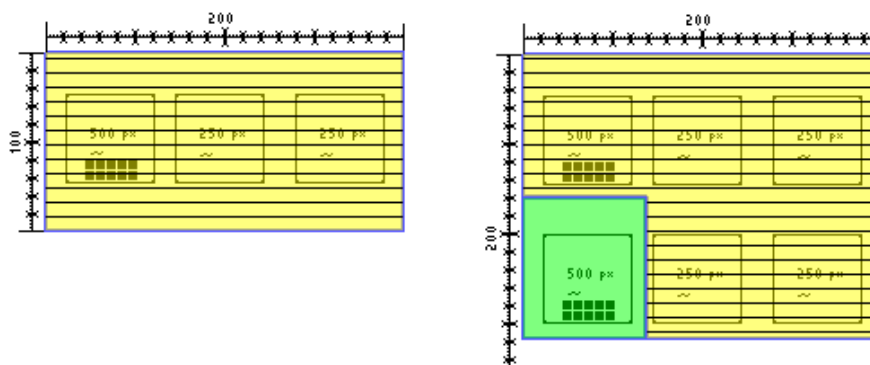- Checked the areas. ✓
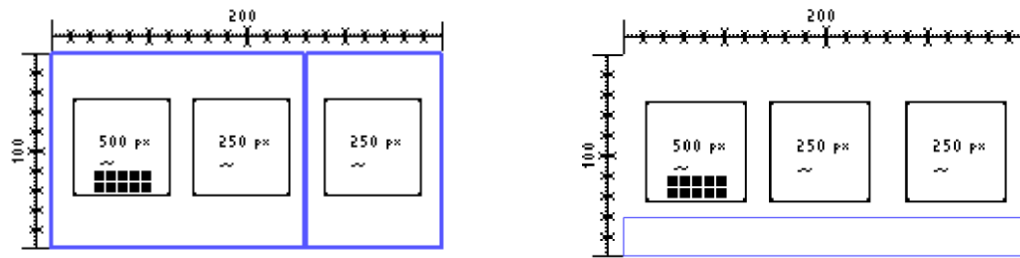
Merge Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)

- Ground-truth:　　　　　　test_2_merge_gt.xml
- Segmentation result:　　test_2_merge_seg.xml
- The ground-truth has 6 text regions. The segmentation result merges them to three regions.
- Area based error:
    - The merge of the two regions (overlap weight is 1):
        - Left region merged with the right one. The weighted area is 7000
        - Right region merged with the left one. The weighted area is 7000
    - The merge of the three regions (overlap weight is 0.5):
        - Top region merged with the right one. The 'merge area' is 10400. The weighted area is 5200.
        - Top region merged with the bottom one. The 'merge are' is 4819. The weighted area is 2409.5.
        - Right region merged with top one. Weighted area: 5200
        - Right region merged with bottom one. Weighted area: 2409.5
        - Bottom region merged with top one. Weighted area: 2409.5
        - Bottom region merged with right one. Weighted area: 2409.5
    - Checked the area based error: 34038 ✓
    - The fifty-percent-x lies at overall region area / 4 = 51800 / 4 = 12950.
    - The success rate function is then $1 / (0.0000772 * x + 1)$
    - Checked the success rate: 27,6% ✓
    - Checked the overall error. ✓
    - Checked the relative error. ✓
- Checked the results per region type. ✓
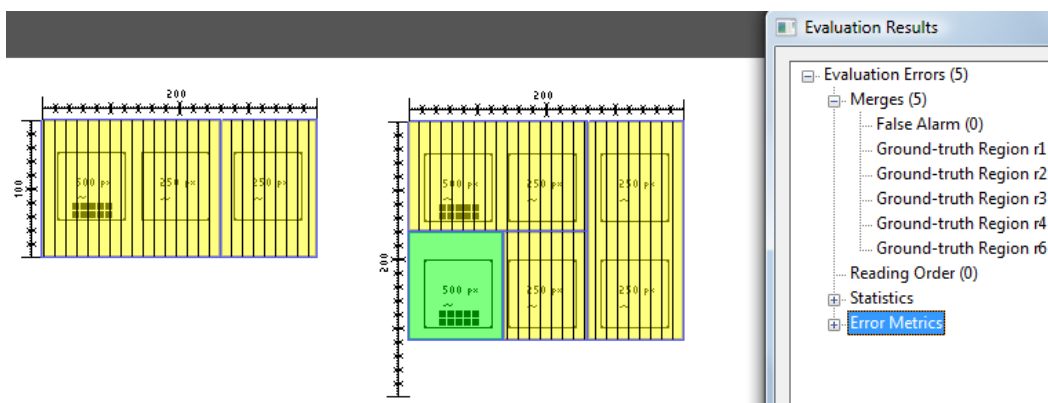
Split Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_2_split_gt.xml
- Segmentation result: test_2_split_seg.xml
- The ground-truth has 3 regions. The segmentation result splits them into six regions.
- Area based error:
    - The split into two regions:3
        - Left split region: area is 7000
        - Right split region: area is **13000**
        - The relative split area is 13000/20000 = 0.65
        - The weighted split area is (1-0.65) * 2 * 20000 = 14000
    - The split of the three regions:
        - Top split region: area is 10400
        - Right split region: area is **11130**
        - Bottom split region: area is 4819
        - The relative split area is 11130 / 26349 = 0.422
        - The weighted split area is (1-0.422) * 3 * 26349 = 45689
    - Checked the area based error: 59656 ✓
    - The fifty-percent-x lies at overall region area / 2 = 25900.
    - The success rate function is then 1 / (0. 0000386*x+1)
    - Checked the success rate: 30,3% ✓
    - Checked the overall error. ✓
    - Checked the relative error. ✓
- Checked the results per region type. ✓

Miss Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)
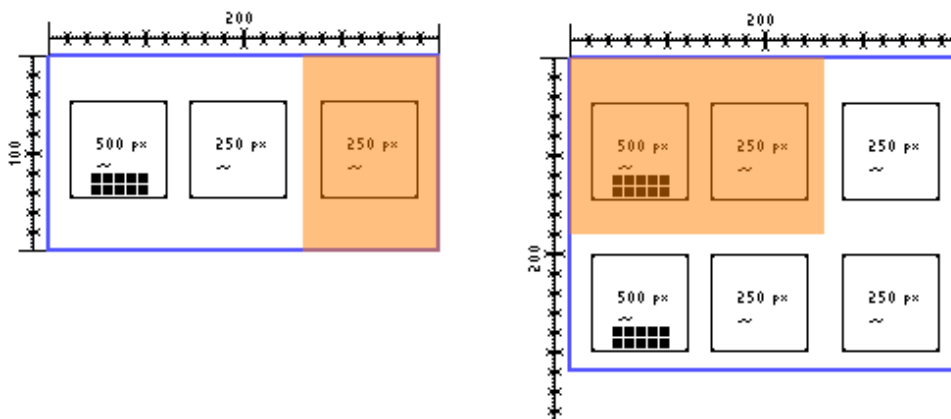
- Ground-truth: test_2_miss_gt.xml
- Segmentation result: test_2_miss_seg.xml
- The ground-truth has 3 regions (2 text, 1 image). The segmentation result has only one text region.
- Area based error:
    - Area of missed regions: 7000 + 6900 pixels = 13900px
    - Checked the area based miss error. ✓

- o The fifty-percent-x lies at overall region area / 2 = 29900 / 2 = 14950
- o The success rate function is then 1 / (0.0000669*x+1)
- o Checked the success rate: 51.8% ✓
- o Checked the overall error. ✓
- o Checked the relative error. ✓
- Checked the results per region type. ✓
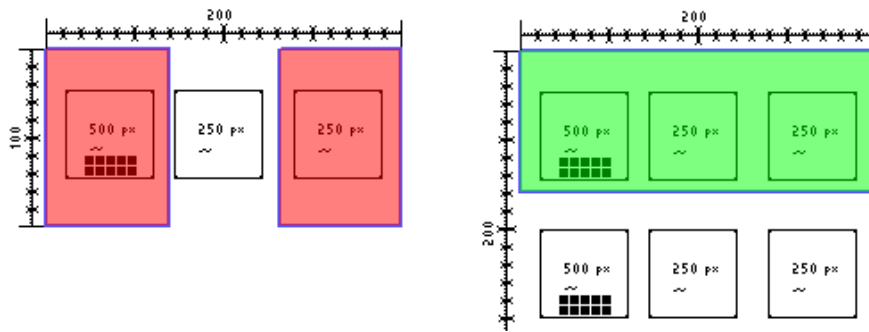
Partial Miss Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_2_part_miss_gt2.xml
- Segmentation result: test_2_part_miss_seg3.xml
- The ground-truth has two regions that are not fully recognized by the segmentation result.
- Area based error:
  - o Missed area: 7000 + 11700 pixels = 18700px
  - o Checked the area based partial miss error. ✓
  - o The fifty-percent-x lies at overall region area / 2 = 52000 / 2 = 26000
  - o The success rate function is then 1 / (0.0000385*x+1)
  - o Checked the success rate: 58.2% ✓
  - o Checked the relative error. ✓
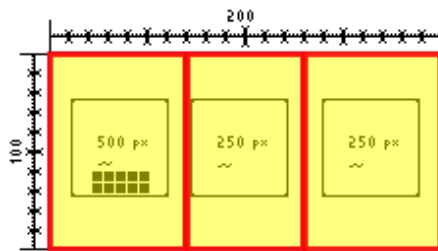- Checked the results per region type. ✓

Misclassification Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_2_ miss_gt.xml
- Segmentation result: test_2_ miss_seg.xml
- The ground-truth contains a text paragraph region, a text header region and an image region. The segmentation result has (in the same order) a graphic region, a text footer region and a noise region.
- Area based error:
  - Misclassified area: 7000 + 7000 + 6000 pixels = 20000px
  - Checked the area based misclassification error. ✓
  - The fifty-percent-x lies at overall region area / 2 = 20000 / 2 = 10000
  - The success rate function is then $1 / (0.0001 \ast x + 1)$
  - Checked the success rate: 33.3% ✓
  - Checked the relative error. ✓
- Checked the results per region type. ✓

False Detection Overall Error / Success Rate (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_2_false_detection_gt.xml
- Segmentation result: test_2_false_detection_seg.xml
- The ground-truth contains 4 regions (3 text, 1 graphic) and the segmentation result contains four additional regions (2 text, 1 graphic and 1 chart)
- Area based error:
  - False detected area: 6300 + 7000 + 6000 + 5400 pixels = 24700px
  - Checked the area based false detection error. ✓
  - The fifty-percent-x lies at (image area – overall region area) / 2 = (529000 – 26300) / 2 = 5036 / 2 = 251350
  - The success rate function is then $1 / (0.00000398 \ast x + 1)$
  - Checked the success rate: 91.1% ✓
  - Checked the relative error. ✓
- Checked the results per region type.
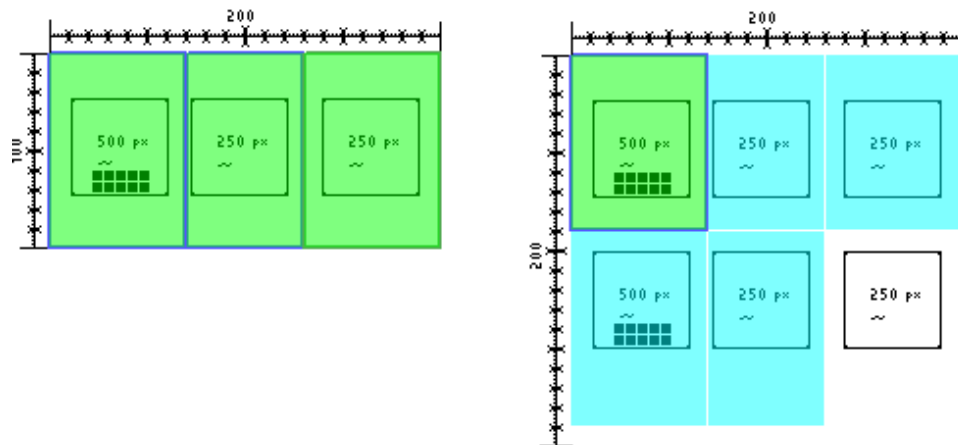
Region Type Weights (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_6_miss_gt.xml
- Segmentation result: test_6_miss_seg.xml
- The ground-truth contains one region of each region type. The segmentation result has only one text region (false detection).
- Results without changing weights:
    - Missed area: 69000
    - False detected area: 2500
    - Area success rate: arithmetic: 90.3%   harmonic: 77.7%
- Text region weight = 0.49
    - Missed area: 65787
    - False detected area: 1225
    - Area success rate: arithmetic: 90.6%   harmonic: 78.5%
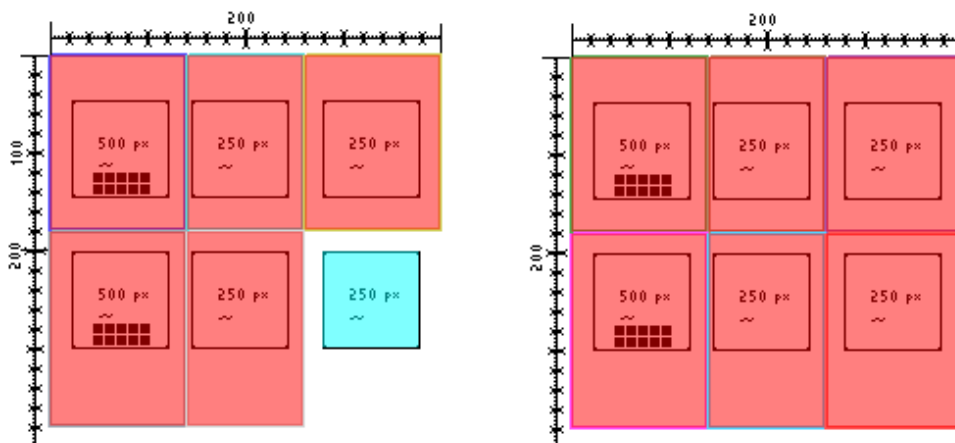
Error Type Weights (*'Use Pixel Area' = FALSE*)

- Ground-truth: test_6_error_types_gt.xml
- Segmentation result: test_6_error_types_seg.xml
- The segmentation result contains all error types (merge, split, miss, partial miss, misclassification, false detection).
- Results without changing weights:
  - Area: merge: 10800   split: 10800   miss: 6300   part. miss: 6000   miscl.: 17100 false d.: 7000
  - Area success rate: arithmetic: 81.6%   harmonic: 79.4%
- Merge weight = 0.49
  - Area: merge: 5292   split: 10800   miss: 6300   part. miss: 6000   miscl.: 17100 false d.: 7000
  - Area success rate: arithmetic: 83.8%   harmonic: 82.3%
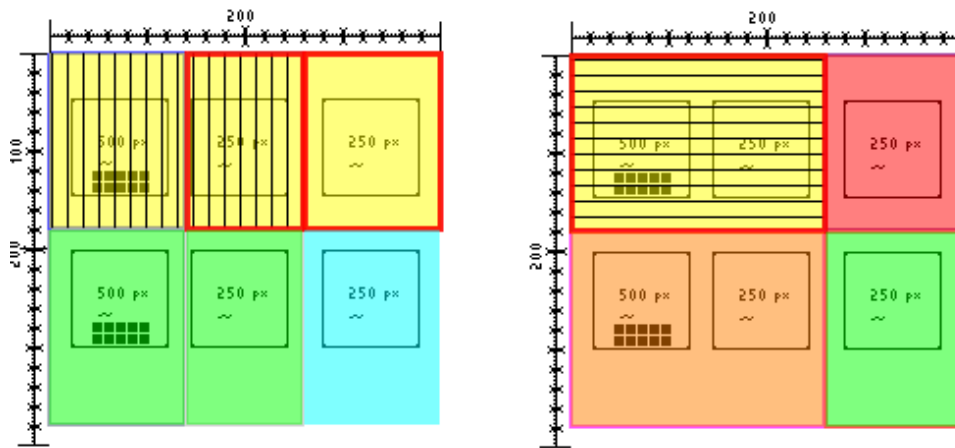
Save-Load-Save Test

- Ground-truth: mp00032.xml
- Segmentation result: mp00032.seg-finereader-converted.xml (competition 2009)
- Test:
  - run evaluation for the files above
  - save the result as test1.evx
  - close the document
  - open test1.evx
  - save as test2.evx
  - compare the two XML files

Ground-Truth - Ground-Truth Test

- Ground-truth: mp00032.xml
- Segmentation result: mp00032.xml
- Checked that the statistics are the same for ground-truth and segmentation result. ✓
- Checked that all errors are 0 and all success rates are 100%. ✓

Default Reading Direction Test

- Ground-truth: test_3_merge_gt.xml
- Segmentation result: test_3_merge_seg.xml
- Use reading direction as specified in ground-truth and allowable merge weight = 0.21:
    - Merge success rate: 41.54%
- Use default reading direction 'bottom-to-top' and allowable merge weight = 0.21:
    - Merge success rate: 47.35%
    - Checked which merges are allowable. ✓

Default Reading Orientation Test

- Ground-truth: test_3_merge_gt.xml
- Segmentation result: test_3_merge_seg.xml
- Use reading orientation as specified in ground-truth and allowable merge weight = 0.21:
    - Merge success rate: 41.54%
- Use default reading orientation 45 degrees and allowable merge weight = 0.21:
    - Merge success rate: 36%
    - Checked which merges are allowable. ✓

Empty Segmentation Result

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_empty.xml
- Checked the results. ✓

Empty Ground-Truth

- Ground-truth: test_2_empty.xml
- Segmentation result: test_2_merge_seg.xml
- Checked the results. ✓

Empty Ground-Truth and Segmentation Result

- Ground-truth: test_2_empty.xml
- Segmentation result: test_2_empty.xml
- Checked the results. ✓

Region Count Deviation Test

- Ground-truth: test_2_merge_gt.xml
- Segmentation result: test_2_merge_seg.xml
- The ground-truth has 6 regions and the segmentation result has 3 regions
- Checked the absolute region count deviation: 3 ✓
- Checked the relative region count deviation: 0.5 ✓

Overall Success Rates for Text Lines, Words and Glyphs Test

- Ground-truth: test_6_error_types_gt2.xml
- Segmentation result: test_2_error_types_seg2.xml
- Text regions only. The result has merges, splits, misses, partial misses and false detection. For each region there is a text line, word and glyph with the same size.
- Area Success Rates on Region Level:
  - Merge: 65.2% success, 45.7% influence
  - Split: 79.0% success, 34.2% influence
  - Miss: 88.2% success, 26.5% influence
  - Partial miss: 88.2% success, 26.5% influence
  - Misclassification: 100% success, 16.7% influence
  - False detection: 75.8% success, 36.8% influence
  - Reading order: 100% success, 100% influence
  - Overall:
    - Arithmetic: 248.15 / 2.86 = 86.8% ✓
    - Harmonic: 2.86 / 0.0339 = 84.4% ✓
- Area Success Rates on Text Line Level:
  - Merge: 65.2% success, 45.7% influence
  - Split: 79.0% success, 34.2% influence
  - Miss: 88.2% success, 26.5% influence
  - Partial miss: 88.2% success, 26.5% influence
  - False detection: 75.8% success, 36.8%influence
  - Overall:
    - Arithmetic: 131.45 / 1.7 = 77.3% ✓
    - Harmonic: 1.7 / 0.0222 = 76.6% ✓
- Checked that the other region levels have the same results.

Empty Segmentation Result Test for Text Lines, Words and Glyphs

- Ground-truth: test_6_error_types_gt2.xml
- Segmentation result: test_6_empty.xml
- Checked the results. ✓

Empty Ground-Truth for Text Lines, Words and Glyphs

- Ground-truth: test_6_empty.xml
- Segmentation result: test_6_error_types_seg2.xml
- Checked the results. ✓

Empty Ground-Truth and Segmentation Result for Text Lines, Words and Glyphs

- Ground-truth: test_2_empty.xml
- Segmentation result: test_2_empty.xml
- Checked the results. ✓