**Problem statement**

Compound image

![Automatic separation](image)

Segmented into subfigures

**Proposed algorithm**

1. **Figure image**
2. **Illustration classifier**
   - Illustration
   - Non-illustration
3. **Border band removal**
4. **Separator line detection**
   - **edge-based**
   - **band-based**
5. **Decide about horizontal or vertical separation**
   - Yes: **Have separator?**
   - No: **Return image**
6. **Subfigure separation**
7. **Subfigure images**

**Evaluation**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>images</th>
<th>illustrations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustration classifier training</td>
<td>1071</td>
<td>57%</td>
</tr>
<tr>
<td>Compound figure separation</td>
<td>3381</td>
<td>54%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound figure separation</th>
<th>Accuracy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>submitted</td>
<td>optimized*</td>
</tr>
<tr>
<td>band-based</td>
<td>30%</td>
<td>69%</td>
</tr>
<tr>
<td>edge-based</td>
<td>35%</td>
<td>55%</td>
</tr>
<tr>
<td><strong>proposed</strong></td>
<td>49%</td>
<td>73%</td>
</tr>
<tr>
<td>best competitor (NLM)</td>
<td>85%</td>
<td></td>
</tr>
</tbody>
</table>

* Parameter optimization using the training set was done only after paper submission.

**Further work**

Improve training set quality for classifier

Optimize decisions of illustration classifier

Eliminate small bounding boxes

Remove image markup

Extract image labels