Overview of the medical task of ImageCLEF 2016

Alba G. Seco de Herrera
Stefano Bromuri
Roger Schaer
Henning Müller
Tasks in ImageCLEF 2016

- Automatic image annotation
- **Medical image classification**
  - Sub-tasks in compound figure identification, separation, multi-label classification, modality classification and caption prediction
- Handwritten scanned document retrieval
  - New task in 2016
Motivation

- The medical literature contains large amounts of images
  - The majority are compound figures that require to have subfigures treated separately
- Image content and caption text can be used for the analysis
- Making the compound figure content accessible is the main goal
Changes in medical task in 2016

- More figures for all subtasks
- New subtask
  - **Caption prediction**: given a medical image, produce a caption associated with the image that is then compared to the real captions
ImageCLEFmed 2016 subtasks

- Compound figure detection
- Compound figure separation
- Multi-label classification
- Subfigure classification
- Caption prediction
Compound figure detection

• To **identify** if a figure is a compound figure or has a single content
To separate the compound figures into subfigures.
Multi-label classification

- To **label** compound figures with each of the modalities of the subfigures
- Without previously separating them
Image type hierarchy

Modality Classification for subfigures

Diagnostic images

- Radiology
  - Ultrasound
  - Magnetic Resonance
  - Computerized Tomography
  - X-Ray, 2D radiography
  - Angiography
  - PET
  - Combined modalities in one image
- Visible light photography
  - Dermatology, skin
  - Endoscopy
  - Other organs
- Printed signals, waves
  - Electroencephalography
  - Electrocardiography
  - Electromyography
- Microscopy
  - Light microscopy
  - Electron microscopy
  - Transmission microscopy
  - Fluorescence microscopy
- 3D reconstructions

Generic biomedical illustrations

- Tables and forms
- Program listing
- Statistical figures, graphs, charts
- Screenshots
- Flowcharts
- System overviews
- Gene sequence
- Chromatography, gel
- Chemical structure
- Mathematics, formula
- Non-clinical photos
- Hand-drawn sketches
Subfigure classification

- To classify subfigures into the 30 classes
Caption detection

- To **generate a caption** that is as close as possible to the real one.
Caption prediction task

• Predicting the caption text based on the image content (and training data)
  • Using a word similarity metric to evaluate success
  • Can also help to make the image content accessible
• Training data are available
Datasets

- ImageCLEFmed 2016
  - 26,456 figures
  - Distributed in training and test sets
- Subset of PubMed Central
  - Over 4 million images of over 1,000,000 articles (2016)
Dataset by subtasks

- Compound figure detection:
  - Full dataset: 26,456 compound and single figures

- Compound figure separation
  - Subset containing 8,397 compound figures

- Multi-label classification
  - Subset containing 2,651 compound figures

- Subfigure classification
  - 10,942 subfigures of compound figures

- Caption prediction
  - 20,000 diagnostic figures (non-compound)
Compound figures and subfigures

- 2,651 figures are:
  - Labeled with all subfigure types
  - Separated into subfigures
  - Using Crowdsourcing

- Figure ID:
  - “1297-9686-42-10-3”

- Subfigure IDs:
  - “1297-9686-42-10-3-1”, “1297-9686-42-10-3-2”, …, “1297-9686-42-10-3-4”
Participation

- Over 72 groups registered
- 8 groups from 4 continents submitted results
- 69 runs submitted
Results: compound figure detection

- **Multimodal approaches achieve best results**
- **Best results using deep convolutional neural networks (CNN)**

<table>
<thead>
<tr>
<th>Group</th>
<th>Run type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUTIR</td>
<td>mixed</td>
<td>92.70</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>90.74</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>90.39</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>90.39</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>85.47</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>69.06</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>mixed</td>
<td>52.25</td>
</tr>
<tr>
<td>MLKD</td>
<td>textual</td>
<td>88.13</td>
</tr>
<tr>
<td>DUTIR</td>
<td>textual</td>
<td>87.03</td>
</tr>
<tr>
<td>DUTIR</td>
<td>textual</td>
<td>86.05</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>textual</td>
<td>85.47</td>
</tr>
<tr>
<td>DUTIR</td>
<td>visual</td>
<td>92.01</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>89.64</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>89.29</td>
</tr>
<tr>
<td>DIS UDEL</td>
<td>visual</td>
<td>69.82</td>
</tr>
</tbody>
</table>
Results: compound figure separation

- CIS UDEL applied:
  - Connected component analysis
  - Post-processing to avoid over-fragmentation

<table>
<thead>
<tr>
<th>Group</th>
<th>Run type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>84.43</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>84.08</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>84.03</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>83.04</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>81.23</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>75.27</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>74.83</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>74.30</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>73.57</td>
</tr>
</tbody>
</table>
Results: multi-label classification

- BMET uses **CNN** and **deep learning**
- MLKD uses a **textual** approach with a **random forest classifier**

<table>
<thead>
<tr>
<th>Group</th>
<th>Hamming Loss</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMET</td>
<td>0.0131</td>
<td>0.295</td>
</tr>
<tr>
<td>BMET</td>
<td>0.0135</td>
<td>0.320</td>
</tr>
<tr>
<td>MLKD</td>
<td>0.0294</td>
<td>0.320</td>
</tr>
</tbody>
</table>
Results: compound figure separation

- **Multimodal** approaches achieve best result
- CNN, feature selection, multiple visual descriptors are used

<table>
<thead>
<tr>
<th>Group</th>
<th>Run type</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCSG</td>
<td>mixed</td>
<td>88.43</td>
</tr>
<tr>
<td>...</td>
<td>mixed</td>
<td>...</td>
</tr>
<tr>
<td>MLKD</td>
<td>textual</td>
<td>72.22</td>
</tr>
<tr>
<td>BCSG</td>
<td>textual</td>
<td>58.37</td>
</tr>
<tr>
<td>BCSG</td>
<td>visual</td>
<td>85.38</td>
</tr>
<tr>
<td>IPL</td>
<td>visual</td>
<td>84.01</td>
</tr>
<tr>
<td>BMET</td>
<td>visual</td>
<td>77.55</td>
</tr>
<tr>
<td>NWPU</td>
<td>visual</td>
<td>76.38</td>
</tr>
<tr>
<td>CIS UDEL</td>
<td>visual</td>
<td>72.46</td>
</tr>
<tr>
<td>NovaSearch</td>
<td>visual</td>
<td>65.31</td>
</tr>
<tr>
<td>...</td>
<td>visual</td>
<td>...</td>
</tr>
</tbody>
</table>
Main tendencies

- **CNNs** (convolutional neural networks) are prominent in 2016
- **Multimodal** approaches achieve generally best results
- **Multiple features** used for **visual classification**
- **Connected component** analysis for figure separation
Conclusions

• Participants present a large variety of techniques
• The trend is towards the use of neural networks
• There were no submissions in the caption prediction subtask
• The subfigure classification subtask was the most popular task
Thank you for your attention!!!

Questions?

http://imageclef.org/2016/medical

albagarcia@nih.gov