<table>
<tr><th rowspan="2">cases</th><th>doc_1</th><th>doc_2</th><th>decision</th><th>id</th></tr>
<tr>
<td>
<table>
<tr><th>authors</th><td>Yue He
Chen Chen
Jing Zhang
Juhua Liu
Fengxiang He
Chaoyue Wang
Bo Du</td></tr>
<tr><th>title</th><td>Visual Semantics Allow for Textual Reasoning Better in Scene Text Recognition</td></tr>
<tr><th>publication_date</th><td>2021-12-24 02:43:42+00:00</td></tr>
<tr><th>source</th><td>SupportedSources.ARXIV</td></tr>
<tr><th>journal</th><td>None</td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>http://arxiv.org/pdf/2112.12916v1
http://arxiv.org/abs/2112.12916v1
http://arxiv.org/pdf/2112.12916v1</td></tr>
<tr><th>id</th><td>id8715457885182187645</td></tr>
<tr><th>abstract</th><td>Existing Scene Text Recognition (STR) methods typically use a language model to optimize the joint probability of the 1D character sequence predicted by a visual recognition (VR) model, which ignore the 2D spatial context of visual semantics within and between character instances, making them not generalize well to arbitrary shape scene text. To address this issue, we make the first attempt to perform textual reasoning based on visual semantics in this paper. Technically, given the character segmentation maps predicted by a VR model, we construct a subgraph for each instance, where nodes represent the pixels in it and edges are added between nodes based on their spatial similarity. Then, these subgraphs are sequentially connected by their root nodes and merged into a complete graph. Based on this graph, we devise a graph convolutional network for textual reasoning (GTR) by supervising it with a cross-entropy loss. GTR can be easily plugged in representative STR models to improve their performance owing to better textual reasoning. Specifically, we construct our model, namely S-GTR, by paralleling GTR to the language model in a segmentation-based STR baseline, which can effectively exploit the visual-linguistic complementarity via mutual learning. S-GTR sets new state-of-the-art on six challenging STR benchmarks and generalizes well to multi-linguistic datasets. Code is available at https://github.com/adeline-cs/GTR.</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td>
<td>
<table>
<tr><th>authors</th><td>Yue He
Chen Chen
Jing Zhang
Juhua Liu
Fengxiang He
Chaoyue Wang
Bo Du</td></tr>
<tr><th>title</th><td>Visual Semantics Allow for Textual Reasoning Better in Scene Text Recognition</td></tr>
<tr><th>publication_date</th><td>2021-12-24 00:00:00</td></tr>
<tr><th>source</th><td>SupportedSources.INTERNET_ARCHIVE</td></tr>
<tr><th>journal</th><td></td></tr>
<tr><th>volume</th><td></td></tr>
<tr><th>doi</th><td></td></tr>
<tr><th>urls</th><td>https://web.archive.org/web/20220105103329/https://arxiv.org/pdf/2112.12916v1.pdf</td></tr>
<tr><th>id</th><td>id433245745519123933</td></tr>
<tr><th>abstract</th><td>Existing Scene Text Recognition (STR) methods typically use a language model to optimize the joint probability of the 1D character sequence predicted by a visual recognition (VR) model, which ignore the 2D spatial context of visual semantics within and between character instances, making them not generalize well to arbitrary shape scene text. To address this issue, we make the first attempt to perform textual reasoning based on visual semantics in this paper. Technically, given the character segmentation maps predicted by a VR model, we construct a subgraph for each instance, where nodes represent the pixels in it and edges are added between nodes based on their spatial similarity. Then, these subgraphs are sequentially connected by their root nodes and merged into a complete graph. Based on this graph, we devise a graph convolutional network for textual reasoning (GTR) by supervising it with a cross-entropy loss. GTR can be easily plugged in representative STR models to improve their performance owing to better textual reasoning. Specifically, we construct our model, namely S-GTR, by paralleling GTR to the language model in a segmentation-based STR baseline, which can effectively exploit the visual-linguistic complementarity via mutual learning. S-GTR sets new state-of-the-art on six challenging STR benchmarks and generalizes well to multi-linguistic datasets. Code is available at https://github.com/adeline-cs/GTR.</td></tr>
<tr><th>versions</th><td></td></tr>
</table>
</td>
<td>DUPLICATES</td>
<td>29</td>
</tr>
</table>