Possible Approaches to Decoding Strings in Scene Text Recognition

Jindřich Libovický

ÚFAL Annual Seminar, Sedec–Prčice, September 16, 2014
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
Why am I telling you this?
Why am I telling you this?
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
What is STR?

Source: ICDAR Robust Reading Dataset
What is STR?

Source: ICDAR Robust Reading Dataset

(200, 532) (358, 575)

FREEDOM
Hypotheses Decoding
Hypotheses Decoding
Hypotheses Decoding
Hypotheses Decoding

Scene Text Recognition

Jindřich Libovický, Charles University in Prague, September 16, 2014
Possible Approaches to Decoding Strings in Scene Text Recognition, 8/18
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
many features for the edges
many features for the edges
learned by Structured Perceptorn and Structured SVM
many features for the edges
learned by Structured Perceptron and Structured SVM

TAKPOROS → TANFORDS

character retrieval F1: .696 → .775
Learning the Decoding

Standard Structured Prediction Methods

- many features for the edges
- learned by Structured Perceptron and Structured SVM

TAKPOROS → TANFORDS

<table>
<thead>
<tr>
<th>character retrieval F1:</th>
<th>word retrieval F1:</th>
</tr>
</thead>
<tbody>
<tr>
<td>.696 → .775</td>
<td>.392 → .393</td>
</tr>
</tbody>
</table>
Moral of the story

- in fact stacking of various local classifiers + the features they were trained from
Moral of the story

- in fact stacking of various local classifiers + the features they were trained from
- in general hard to include non-linearity to the decisions

Source: ...
Moral of the story

- in fact stacking of various local classifiers + the features they were trained from
- in general hard to include non-linearity to the decisions
- missing global features

Source: ...
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
CRFs minimize energy of a variable assignment
CRFs minimize energy of a variable assignment

- binary variables – is / is not in word
CRFs minimize energy of a variable assignment
- binary variables – is / is not in word
- minimum energy + “true” assignments must form a path
CRFs minimize energy of a variable assignment
binary variables – is / is not in word
minimum energy + “true” assignments must form a path
can be encoded as linear programming task
Unsupervised feature learning

- tons of unlabeled data
Unsupervised feature learning

- tons of unlabeled data
- RBM, autoencoders ...
Back-prob in online learning

Local features optimization

Σ
Σ

1-1
1

edges from the decoded path
edges from the ground truth path
Outline

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization
Global features optimization

Greedy Agent

- choses edges greedily
Greedy Agent

▶ chooses edges greedily
▶ can use the previous decoding to make prediction
Greedy Agent

- choses edges greedily
- can use the previous decoding to make prediction
- generalized SEARN for learning
Thank you for your attention.