Reciprocal Feature Learning via Explicit and Implicit Tasks in Scene Text Recognition

Hui Jiang, Yunlu Xu, Zhanzhan Cheng, Shiliang Pu, Yi Niu, Wenqi Ren, Fei Wu, Wenming Tan
Hikvision Research Institute, Hangzhou, China
Zhejiang University, Hangzhou, China
Background

Method

Experiment
Background

**Scene Text Recognition Feat Single-Task Learning**

- CTC-Based
- Attention-Based

**Scene Text Recognition Feat Multi-Task learning**

- Additional information from another task or detailed supervision
- Exploiting original tasks and supervision
**Background**

*Drawback of current solution*

- Single Task solution
  - Limited Performance
  - Add extra annotations

*Multi-Task solution*

- Immature Technology Application
- Ignore the relation between tasks
- Task competition

*Motivation*

- Excavate implicit information from existing annotations to training a auxiliary task
- Excavate and utilize the relation between tasks to improve the performance
Method

**Overall Architecture**

- Backbone
- Character Counting Branch (CNT)
- Text Recognition Branch (RCG)
- Reciprocal Feature Adaptor (RF-Adaptor)
Method

**Character Counting Branch**

Text is a **hierarchically** information carrier

\[
T: 1 \text{ h : 1 e : 1 n : 1}
\]

Then

\[
\begin{align*}
T & \quad \text{Th} \quad \text{The} \quad \text{Then} \\
4 & 
\end{align*}
\]

**Text length** is a facilitated information in text information and correlate to the text recognition task

\[
L_{\text{cnt}} = \begin{cases} 
MSE(\hat{y}_{\text{cnt}}, y_{\text{cnt}}) & \text{if Regression} \\
\text{CrossEntropy}(\hat{y}_{\text{cnt}}, y_{\text{cnt}}) & \text{if Classification} 
\end{cases}
\]

**Metric**

\[
\begin{align*}
RMSE & = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{c}_i - c_i)^2} \\
\text{relRMSE} & = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{(\hat{c}_i - c_i)^2}{c_i + 1}}
\end{align*}
\]
Method

🌟 Reciprocal Feature Adaptor

Transfer the bi-directional complementary data from one to the other, assembling features and adapting to task

- Feature Fusion

  - RCG contains more information than CNT, replenish information via ⊕
  - CNT is feature selector like a learnable gate to suppress the noise via ⊗

- Feature Strengthen

  - Apply different self-enhancement module to strengthen the feature
Performance Summary

Compared with SOTA solution

| Methods         | Year | Training data | Benchmark | Avg. Acc |
|-----------------|------|---------------|-----------|----------|
|                 |      |               | HII, SVT, IC03, IC13, IC15, SVTP, CT | Regular | Irregular |
| CRNN [27]       | 2016 | MJ            | 78.2, 80.8, 89.4 | -        | -        |
| AON [5]         | 2018 | MJ+ST         | 87.0, 82.8, 91.5 | -        | 73.0, 76.8 |
| NRTR [26]       | 2018 | MJ+ST         | 90.1, 91.5, 94.7 | 79.4, 86.6 | 80.9, 82.3 |
| ASTER [28]      | 2019 | MJ+ST         | 93.4, 89.5, 94.5 | 91.8, 78.5 | 79.5, 92.3 |
| TPS-Bilstm-Attn [1] | 2019 | MJ+ST         | 87.9, 87.5, 94.9 | 93.6, 77.6 | 79.2, 91.0 | 76.9 |
| AutoSTR [40]*   | 2020 | MJ+ST         | 94.7, 90.9, 93.3, 94.2 | 81.8, 81.7 | -        | 93.2, - |
| RobustScanner [39]+ | 2020 | MJ+ST         | **95.3**, 88.1 | -        | -        | 79.5, **90.3** | - |
| Bilstm-Attn [1]$^3$ | 2019 | MJ+ST         | 93.7, 89.0, 92.3, 93.2 | 79.3, 81.2 | 80.6, 92.1 | 80.4 |
| Bilstm-Attn w. RF-L | -    | MJ+ST         | 94.1, 88.6, 94.9 | **94.5**, 82.4 | 82.0, 82.6 | 93.0 (+0.9) | **82.4** (+2.0) |
| DAN [35]$^4$    | 2020 | MJ+ST         | 93.4, 87.5, 94.2, 93.2 | 75.6, 80.9 | 78.0, 92.1 | 78.2 |
| DAN w. RF-L     | -    | MJ+ST         | 94.0, 87.7, 93.6, 93.5 | 76.7, 84.7 | 77.8, 92.2 (+0.1) | 79.7 (+1.5) |

Samples

| w.o RF-L | evil | gujarat | evil | gujarat | laugh | squiris |
| w. RF-L  | pipinang | alibaba | change | before |

| Samples |
|---------|
| w.o RF-L | Pipinang | Alibaba | Change | Before |
| w. RF-L  | Pipinang | Alibaba | Change | Before |
Experiment

🌟 Ablation Summary

- CNT Implementation Ablation

| Methods   | w.o. Class Balance | w. Class Balance |
|-----------|--------------------|------------------|
|           | Regular² Irregular | Regular Irregular|
| CE        | 89.5 78.5          | 93.2 83.5        |
| Regression| 93.3 82.3          | 94.6 84.5        |

- CNT Implementation compared with ACE

| Methods | Auxiliary | CNT RCG | RCG Accuracy (%) | CNT RMSE |
|---------|-----------|---------|------------------|----------|
| ACE     | IIT       | SVT     | IC03             | IC15     |
| w. RCG (RF-L) | √ | 87.5 81.8 89.9 67.5 | 0.477 0.963 0.555 0.889 |
| w. CNT (RF-L) | √ | 88.4 83.8 90.2 70.0 | 0.323 0.890 0.518 0.896 |

- Generalization Ablation

| Methods | Encoder      | Decoder      | w. CNT (RF-L) | IIT | SVT | IC03 | IC13 | IC15 | SVTP | CT | Avg.Gain |
|---------|--------------|--------------|---------------|-----|-----|------|------|------|------|-----|----------|
| IIT     | VGG          | Bilstm-Attn  | √             | 91.2 85.5 92.6 92.1 77.5 77.7 73.6 |
|         | VGG          | Bilstm-Attn  |               | 91.8 86.9 92.9 92.9 78.0 78.9 74.7 | +0.9 |
| SVT     | ResNet       | Bilstm-Attn  | √             | 93.7 89.0 92.3 93.2 79.3 81.2 80.6 |
| IC03    | ResNet       | Bilstm-Attn  |               | 94.1 88.4 94.5 94.5 81.9 82.0 82.6 | +1.2 |
| IC13    | ResNet       | CTC          | √             | 91.7 85.8 91.5 91.7 74.1 73.2 76.7 |
| IC15    | ResNet       | CTC          |               | 92.1 86.9 92.1 92.4 76.5 75.8 78.9 | +1.5 |
| SVTP    | ResNet       | Paral-Attn   | √             | 90.0 82.8 87.6 89.0 72.4 71.0 73.3 |
| CT      | ResNet       | Paral-Attn   |               | 90.3 85.8 92.2 93.0 73.8 75.8 77.8 | +3.8 |
## Ablation Summary

- Optimization Ablation

| Methods                        | Branch | Direction | Benchmark                   | Avg. Acc |
|-------------------------------|--------|-----------|-----------------------------|----------|
|                               | RCG    | RCG CNT   | R2C R2C H1T SVT IC03 IC13 IC15 SVTP CT |          |
| RCG                           | ✓      | ✓         | 90.0 82.8 87.6 89.0 72.4 71.0 73.3 | 81.3     |
| RCG w. CNT (JT-L)             | ✓      | ✓         | 89.6 83.9 92.6 91.7 72.6 74.0 78.1 | 82.4(+1.1) |
| RCG w. Fixed CNT (RF-L)       | ✓      | ✓         | ✓ 90.2 86.7 92.2 91.6 73.2 76.0 79.5 | 82.8(+1.5) |
| RCG w. CNT (Unidirectional RF-L) | ✓     | ✓         | ✓ 90.7 86.6 92.6 91.2 73.2 76.0 80.2 | 82.9(+1.7) |
| RCG w. CNT (Bidirectional RF-L) | ✓     | ✓         | ✓ 90.3 85.8 92.2 93.0 73.8 75.8 77.8 | 83.3(+2.0) |
| CNT                           | ✓      | ✓         | 92.5 93.0 96.3 95.6 84.2 85.0 85.8 | 89.4     |
| CNT w. RCG (JT-L)             | ✓      | ✓         | ✓ 93.0 94.3 96.2 96.1 84.9 86.4 83.7 | 89.8(+0.4) |
| CNT w. Fixed RCG (RF-L)       | ✓      | ✓         | ✓ 91.6 92.9 96.5 96.0 86.0 87.3 87.2 | 89.9(+0.5) |
| CNT w. RCG (Unidirectional RF-L) | ✓     | ✓         | ✓ 92.6 93.5 96.6 95.2 86.0 86.7 89.6 | 90.0(+0.6) |
| CNT w. RCG (Bidirectional RF-L) | ✓     | ✓         | ✓ 93.5 94.0 96.7 95.7 85.5 86.7 88.9 | 90.3(+0.9) |
Thank you