Motivation

- We address a challenging problem: recognizing multiple text sequences from an image by pure end-to-end learning.
- Multiple text sequences recognition (MSR). Each image can contain multiple text sequences of different content, location, and orientation.
- Pure end-to-end (PEE) learning. Each training image is annotated with only text transcripts.
- Most existing works cannot handle this problem. Some of them use both text transcripts and text locations in a non-end-to-end (NEE) or quasi-end-to-end (QEE) way. Some of them are PEE method but for single text sequence recognition problem.
- We develop a novel PEE method MSRA to solve the MSR problem, in which the model is trained with only sequence-level text transcripts.

Multiple Sequence Recognition Approach (MSRA)

- MSRA aims to transform a three-dimensional tensor \( X \) to a conditional probability distribution over multiple character sequences \( P(Z|X) \).

\[
X = \begin{bmatrix}
X^{(1)} & X^{(2)} & \cdots & X^{(w)}
\end{bmatrix}
\]

\[
P(Z|X) \triangleq \frac{1}{N} \sum_{i=1}^{N} p(l_i|X)
\]

\( Z \) is denoted as a set of text sequences \( l_i \), which is obtained by using the many-to-one \( B \)-mapping strategy for path \( l \) on the two-dimensional probability distribution \( X \).

The evaluation of \( P(l_i|X) \) turns to solve the two-dimensional probability path \( l \) search problem over \( X \).

\[
p(l_i|X) = \sum_{l \in B^{(1)}} p(l_i|l) = \sum_{l \in B^{(1)}} \prod_{t=0}^{l-1} z_{l,t}^{(i,t)}
\]

Objective Function

\[
O = -\sum_{(X,Z) \in S} \ln p(Z|X) = -\frac{1}{S} \sum_{l \in \mathcal{L}} \sum_{(i,t)} \alpha_i(j) \beta_j(l)
\]

Similar to \( \alpha_i(j) \) and \( \beta_j(l) \), the gradient of the objective function can be obtained based on them when \( \text{lab}(l,k) = \{ i : v_i = k \} \).

Experiments

- Evaluation metrics
  - NED(%): the normalized edit distance.
  - SA(%): the sequence recognition accuracy.
  - AQA(%): the image recognition accuracy.

Recognition results on MS-MNIST datasets

- Recognition results on real application scenarios datasets

Conclusion

Our contribution can be summarized as below:

- A new taxonomy of text recognition methods: NEE, QEE, PEE.
- A novel PEE method MSRA to solve MSR.
- Build up several datasets and conduct extensive experiments on them.