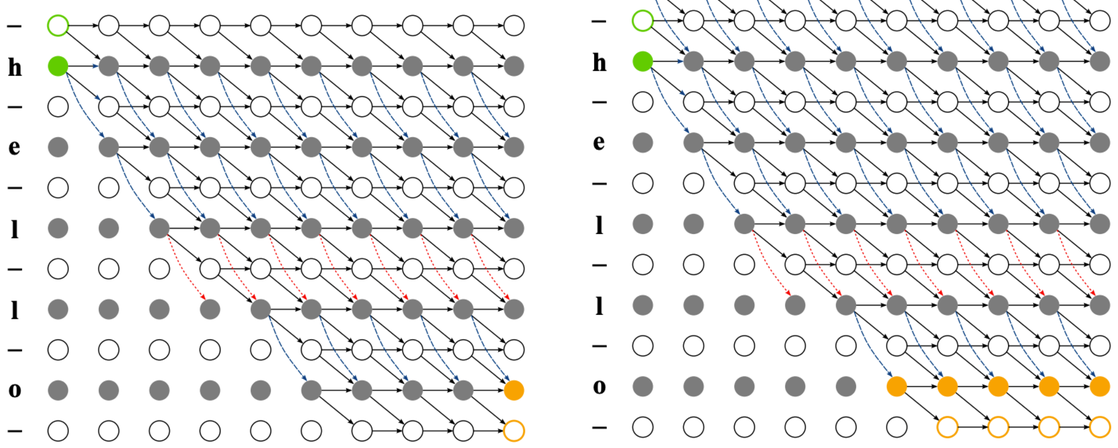
W-CTC: a Connectionist Temporal Classification Loss with Wild Cards

Xingyu Cai, Jiahong Yuan, Yuchen Bian,
Guangxu Xun, Jiaji Huang, Kenneth Church

Paper Summary

- This paper proposes a **wild-card enhanced CTC (W-CTC)**, that solves the sequence learning problem when the labels lost starting or ending parts.
- Comprehensive experiments on ASR (speech), OCR (vision) and CSLR (video) tasks, with different backbone networks, show that the proposed W-CTC consistently outperforms the standard CTC by a large margin when label is incomplete (up to 70% missing).

The Proposed W-CTC



(a) Standard CTC: The two green nodes at top-left corner are initial states for DP recursion. The two orange nodes at bottom-right corner are ending nodes for final loss calculation. The black and blue arrows are allowed transitions based on Equation 3. The red arrows are forbidden.

(b) W-CTC: The first row corresponds to the prepended wild-card "$*$" symbol (the blue nodes). There are three initial states. The ending nodes are the entire last two rows, rather than only two right-bottom nodes. These changes enable $\hat{Y}$ to match only a fraction of X.

Figure 1: Illustration of Dynamic Programming based CTC and proposed W-CTC loss calculation.

The 3 key improvements are:

1. Prepend a "$*$" symbol (wild-card, matches any characters in the label) to Y (label), resulting in an additional row on top of the trellis, denoted as the blue nodes in Figure 1(b).
2. There are 3 green nodes to start Dynamic Programming recursion.
3. The orange ending nodes (final states) are the entire last two rows in the trellis.

Notice that:

1. The wild-card symbol "$*$" is **NOT** equivalent to the "blank" symbol in standard CTC.
2. W-CTC only handles missing Y (label) cases, but not missing X (sequence) problem.
3. The missing part in Y needs to be at the sides, but not in the middle. Taking the example of "hello", it can handle "?ell?", but not "h???o", nor "?e?llo", where "?" represents the missing part.

OCR Experiments

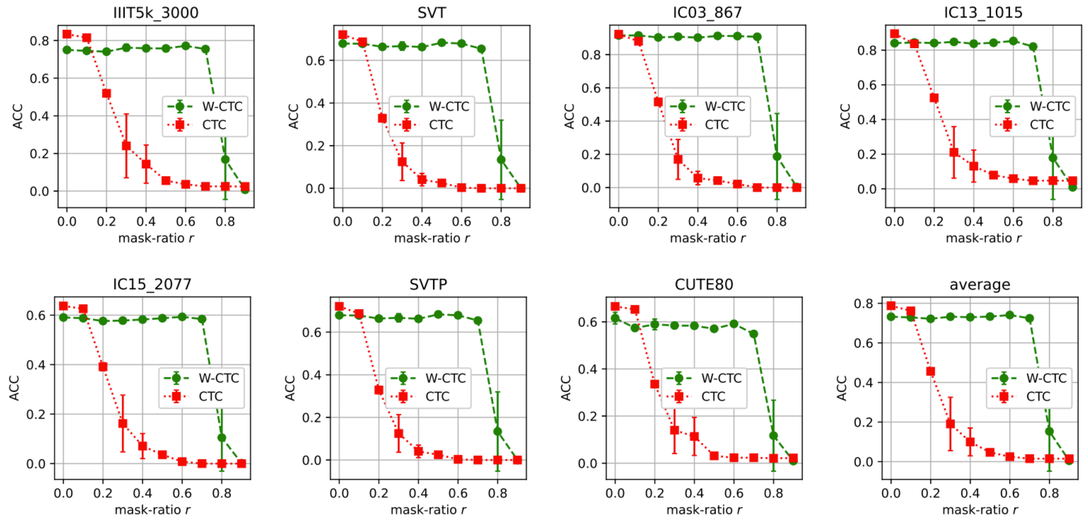
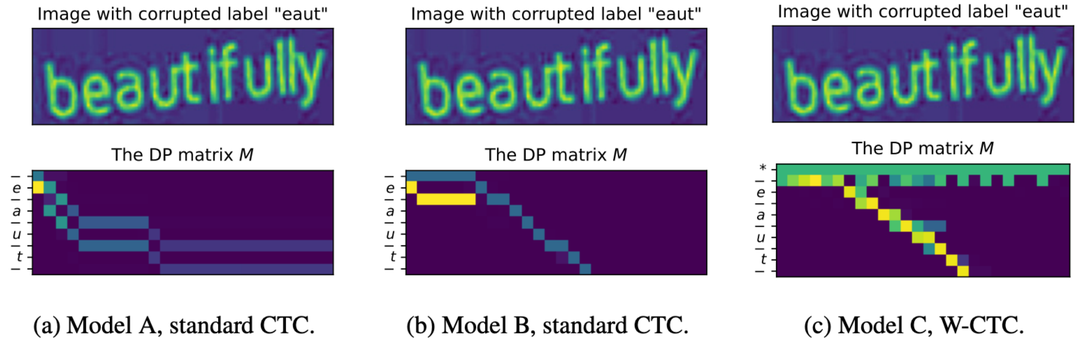


Figure 3: Test accuracy on 7 standard test sets, as a function of r (mask ratio). The last plot is the average. The proposed W-CTC has generally better accuracy than the standard CTC.



(a) Model A, standard CTC. (b) Model B, standard CTC. (c) Model C, W-CTC.

Figure 4: The alignment paths (in the trellis M) from CTC computation. Model A and C are trained on corrupted labels $r = 0.7$; model B is trained on clean data $r = 0$. Model A and B use standard CTC; model C uses W-CTC. Model A fails to produce correct alignment. Model B has correct alignment but with confusions. Model C provides clear path though trained on corrupted labels.

ASR and PR Experiments

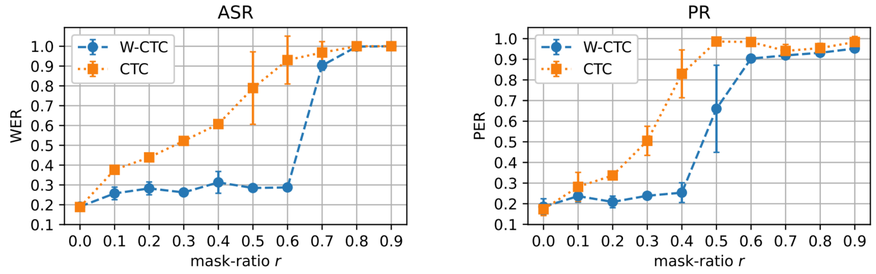


Figure 2: WER / PER vs mask-ratio in ASR and PR tasks, on TIMIT test set.