

Connectionist Temporal Classification with Maximum Entropy Regularization

Hu Liu Sheng Jin Changshui Zhang

Department of Automation
Tsinghua University

NeurIPS, 2018



Introduction to Connectionist Temporal Classification (CTC)



$$P(\text{'dog'} | \text{dog}) = P \left(\begin{array}{c|c} \text{dd_oo_g} & \\ \text{doo_gg} & \\ \text{— dogg} & \text{dog} \\ \vdots & \\ \text{dddoogg} & \end{array} \right)$$

↑
The most suitable path

$$\frac{\partial L_{ctc}}{\partial y_k^t} = -\frac{1}{p(l|X)y_k^t} \sum_{\{\pi | \pi \in \mathcal{B}^{-1}(l), \pi_t = k\}} p(\pi|X)$$

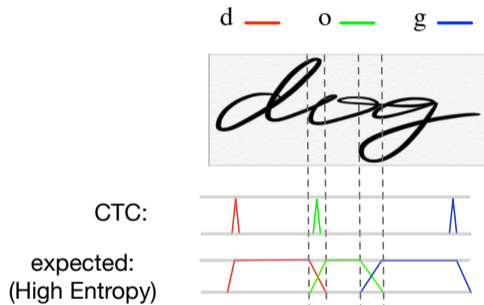
error signal
positive feedback

CTC Drawbacks:

- CTC lacks exploration and is prone to fall into worse local minima.
- Output overconfident paths (overfitting).
- Output paths with peaky distribution.



Maximum Conditional Entropy Regularization for CTC (EnCTC)



$$L_{enctc} = L_{ctc} - \beta H(p(\pi|l, X)) \leftarrow \text{Entropy-based regularization}$$

$$H(p(\pi|l, X)) = - \sum_{\pi \in \mathcal{B}^{-1}(l)} p(\pi|X, l) \log p(\pi|X, l).$$

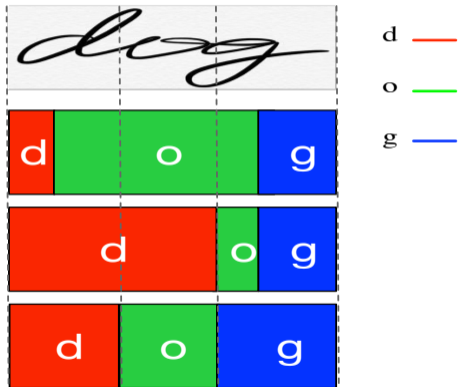
EnCTC:

- Better generalization and exploration.
- Solve peaky distribution problem.
- Depict ambiguous segmentation boundaries.



Equal Spacing CTC (EsCTC)

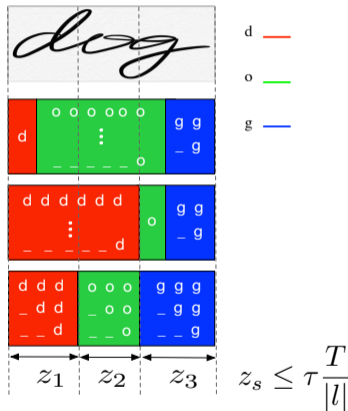
- The spacing of two consecutive elements is nearly the same in many sequential tasks.



- We adopt equal spacing as a pruning method to rule out unreasonable CTC paths.



Equal Spacing CTC (EsCTC)



Theorem 3.1. Among all segmentation sequences, the equal spacing one has the maximum entropy.

$$\operatorname{argmax}_z \max_p H(p(\pi|z, l, X)) = z_{es}$$

Equal spacing is the best prior without any subjective assumptions.

$$L_{esctc} = -\log \sum_{z \in C_{\tau, T}} \sum_{\pi \in \mathcal{B}_z^{-1}(l)} p(\pi|X)$$



Algorithm and Complexity Analysis

We propose dynamic programming algorithms for EnCTC, EsCTC, and EnEsCTC.

EnCTC

$$\bar{\gamma}(t, s) = \begin{cases} \gamma(t-1, s) + \gamma(t-1, s-1) & \text{if } l'_s = b \text{ or } l'_{s-2} = l'_s \\ \gamma(t-1, s) + \gamma(t-1, s-1) + \gamma(t-1, s-2) & \text{otherwise} \end{cases} \quad Q(l) = \gamma(T, |l'|) + \gamma(T, |l'| - 1)$$

EsCTC

$$\alpha_\tau(t, s) = \begin{cases} \sum_{t'=1}^{\tau} \prod_{t''=1}^{\tau} \alpha_\tau(t-t', s-1) \sigma(t-t'+1, t, s) & \text{if } l_{s-1} \neq l_s \\ \sum_{t'=2}^{\tau} \prod_{t''=2}^{\tau} \alpha_\tau(t-t', s-1) y_b^{t-t'+1} \sigma(t-t'+2, t, s) & \text{otherwise} \end{cases} \quad p_\tau(l|X_{1:T}) = \alpha_\tau(T, |l|) + \sum_{t'=1}^{\tau} \prod_{t''=1}^{\tau} \alpha_\tau(T-t', |l|) \sigma(T-t'+1, T, 0)$$

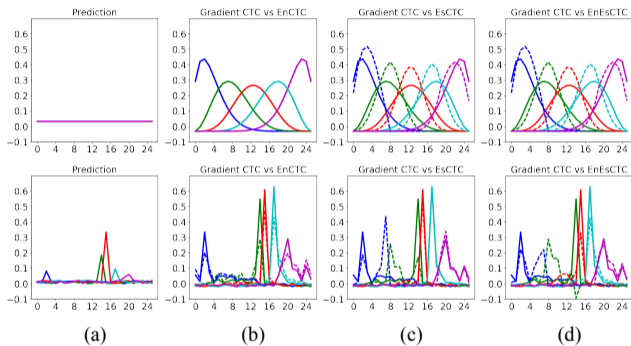
EnEsCTC

$$\gamma_\tau(t, s) = \begin{cases} \sum_{t'=1}^{\tau} \prod_{t''=1}^{\tau} \gamma_\tau(t-t', s-1) \sigma(t-t'+1, t, s) + \alpha_\tau(t-t', s-1) \eta(t-t'+1, t, s) & \text{if } l_{s-1} \neq l_s \\ \sum_{t'=2}^{\tau} \prod_{t''=2}^{\tau} \gamma_\tau(t-t', s-1) y_b^{t-t'+1} \sigma(t-t'+2, t, s) + \alpha_\tau(t-t', s-1) y_b^{t-t'+1} \eta(t-t'+2, t, s) + \alpha_\tau(t-t', s-1) y_b^{t-t'+1} \log y_b^{t-t'+1} \sigma(t-t'+2, t, s) & \text{otherwise} \end{cases} \quad Q_\tau(l) = \gamma_\tau(T, |l|) + \sum_{t'=1}^{\tau} \prod_{t''=1}^{\tau} \gamma_\tau(T-t', |l|) \sigma(T-t'+1, T, 0) + \alpha_\tau(T-t', |l|) \sigma(T-t'+1, T, 0) \log \sigma(T-t'+1, T, 0)$$

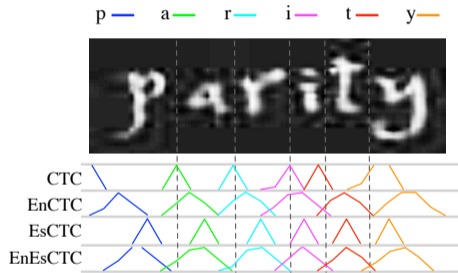
Complexity	CTC	EnCTC	EsCTC	EnEsCTC
Time	$O(T)$	$O(T)$	$O(\frac{T^2}{ l }) \sim O(T^2)$	$O(\frac{T^2}{ l }) \sim O(T^2)$
Space	$O(T l)$	$O(T l)$	$O(T^2 l)$	$O(T^2 l)$



Qualitative Analysis



Error Signal in Training



Alignment Evaluation



Results on Scene Text Recognition Benchmarks

Evaluation of model generalization.

Method	Synth5K
CTC	38.1
CTC + LS	42.9
CTC + CP	44.4
EnCTC	45.5
EsCTC	46.3
EnEsCTC	47.2

Comparisons with the state-of-the-art methods.

Method	IC03	IC13	IIIT5K	SVT
CRNN	89.4	86.7	78.2	80.8
STAR-Net	89.9	89.1	83.3	83.6
R2AM	88.7	90.0	78.4	80.7
RARE	90.1	88.6	81.9	81.9
EnCTC	90.8	90.0	82.6	81.5
EsCTC	92.6	87.4	81.7	81.5
EnEsCTC	92.0	90.6	82.0	80.6



For more results and analyses, please come

Poster: Room 210 & 230 AB #106

<https://github.com/liuhu-bigeye/enctc.crnn>

