Connectionist Temporal Classification with Maximum Entropy Regularization

Hu Liu Sheng Jin Changshui Zhang
Institute for Artificial Intelligence, Tsinghua University (THUAI)
Department of Automation, Tsinghua University, Beijing, P.R.China

Introduction for CTC

- Connectionist Temporal Classification (CTC) is a popular objective function for end-to-end sequence learning tasks, including speech recognition and scene text recognition.
- CTC learns by maximum likelihood estimation over the summation of all feasible path probabilities,

$$L_{ctc} = -\log p(l|X) = -\log \sum_{\pi \in B^{-1}(l)} p(\pi|X)$$

CTC peaky distribution problem

- Output paths with peaky distribution.
- Output overconfident paths.
- Lack exploration among feasible paths.

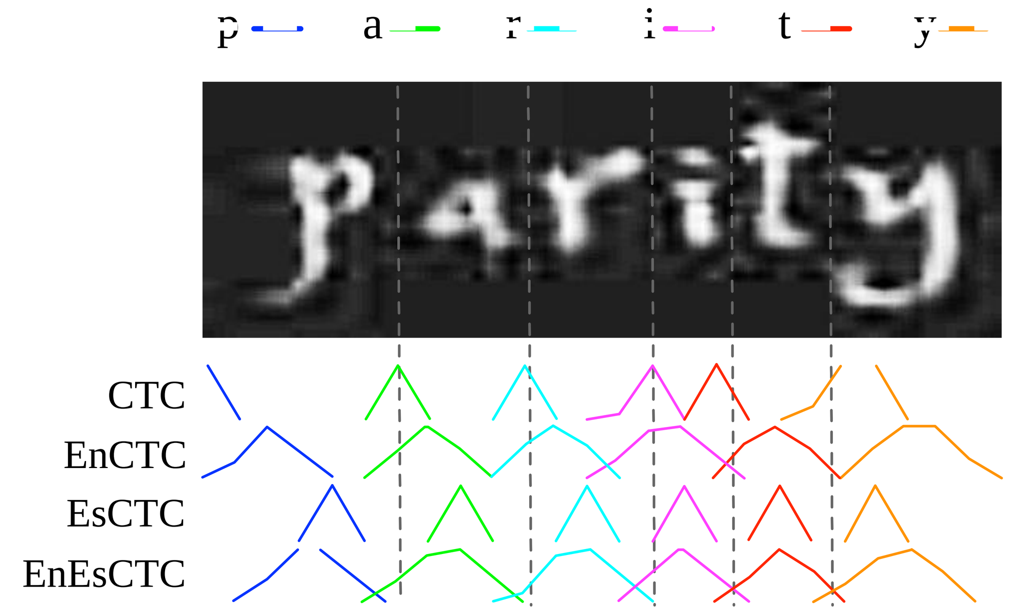

Figure1: scene text recognition predictions of an example image

Maximum Conditional Entropy Regularization for CTC (EnCTC)

- We propose an entropy based regularization term that prevents the entropy of the feasible paths from decreasing too fast, leading to better generalization and exploration.

$$L_{enctc} = L_{ctc} - \beta H(p(\pi|l,X)),$$

where β controls the strength of the regularization.

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

Equal Spacing CTC (EsCTC)

- We propose to enforce equal spacing constraints in order to explicitly rule out the unreasonable alignments.

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

where $C_{\tau,T} = \left\{ z \,|\, T - \tau \frac{T}{|l|} \le \sum_{s=1}^{|l|} z_s \le T, z_s \le \tau \frac{T}{|l|} \right\}.$

- We further prove that among all segmentation sequences, the equal spacing one can reach the maximum entropy.

$$\arg\max_{z} \max_{p} H\big(p(\pi|z,l,X)\big) = z_{es}$$

Results

- En/EsCTC achieve superior performance to the CTC baseline and show better generalization ability.

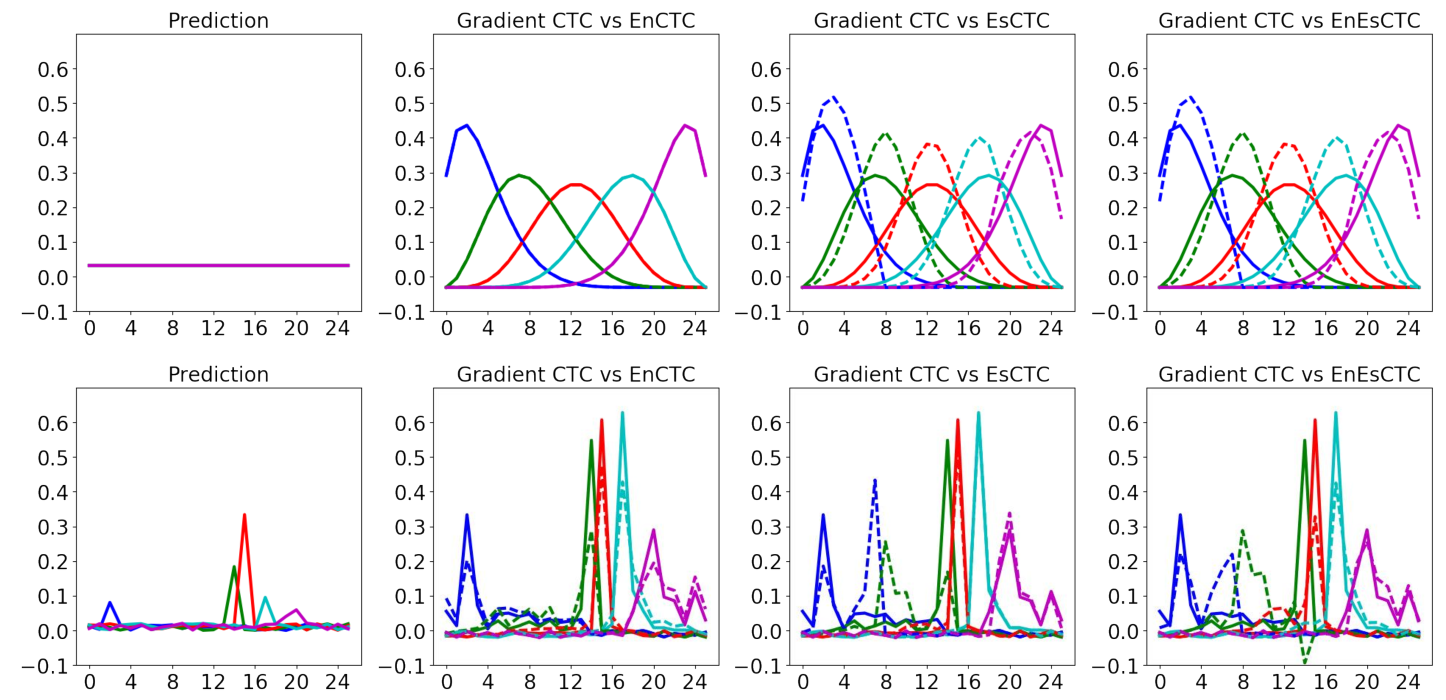

Figure 1: The evolution of prediction and error signal

Method	Synth5K
CTC	38.1
CTC + LS [31]	42.9
CTC + CP [27]	44.4
EnCTC	45.5
EsCTC	46.3
EnEsCTC	**47.2**

Table1: Evaluation of generalization

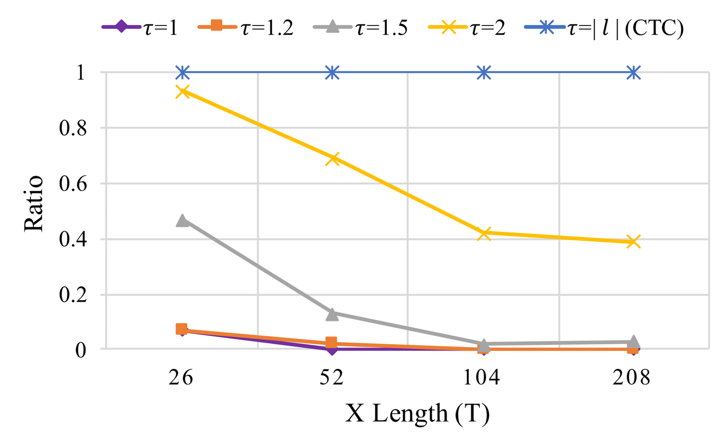

Figure2: Path pruning rate by EsCTC

Method	IC03	IC13	IIIT5K	SVT
CRNN [29]	89.4	86.7	78.2	80.8
STAR-Net [19]	89.9	89.1	**83.3**	**83.6**
R2AM [17]	88.7	90.0	78.4	80.7
RARE [30]	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

Table2: Comparisons with the state-of-the-art methods