When Counting Meets HMER: CAN for HMER

汇报人：刘翰文

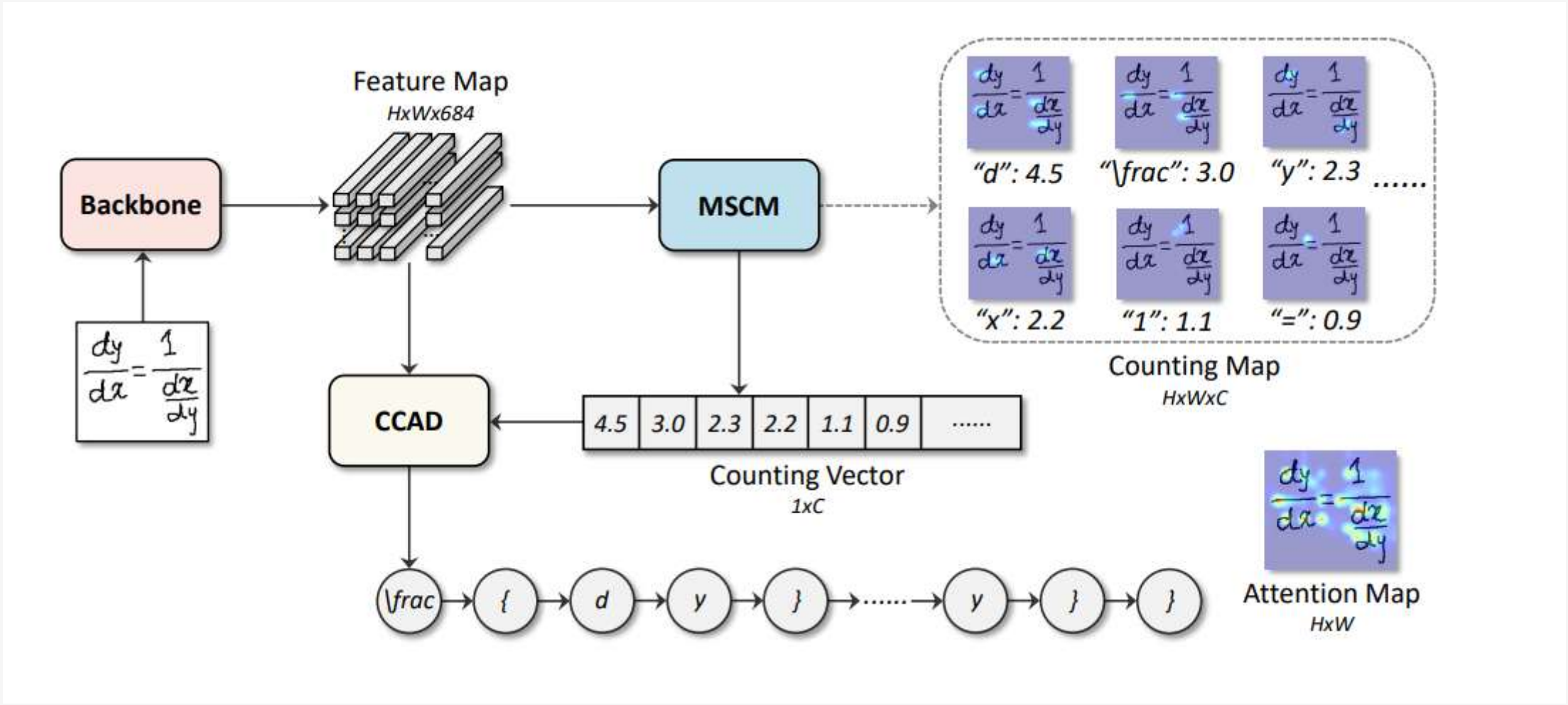
2024年6月4日

饮 水 思 源 • 爱 国 荣 校

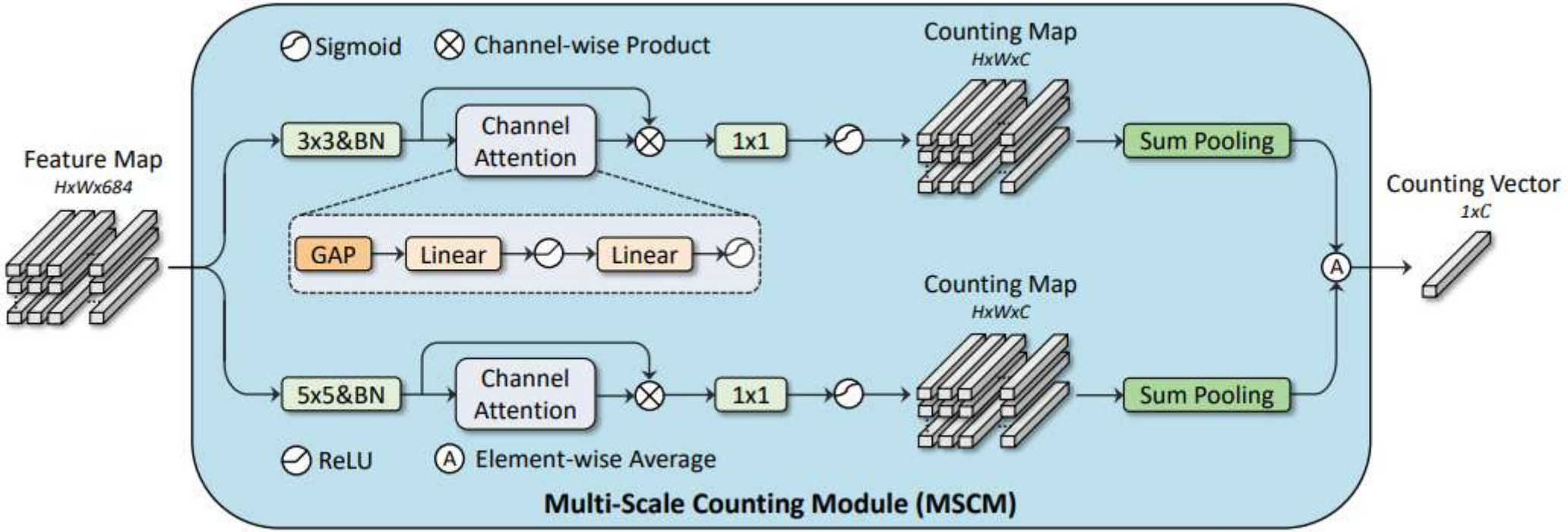
Traditional HMER

- Encoder-decoder architecture
- Encoder: Input Sequence → Output Context Vector in fixed length
- Decoder: Context vector → Output Sequence
- Weakness: Accuracy not guaranteed for complicated formula or long markup sequence

Structure of CAN



MSCM



Feature Map →

$$\mathcal{Q} = \sigma(W_1(G(\mathcal{H})) + b_1),$$

$$\mathcal{S} = \mathcal{Q} \otimes g(W_2\mathcal{Q} + b_2),$$

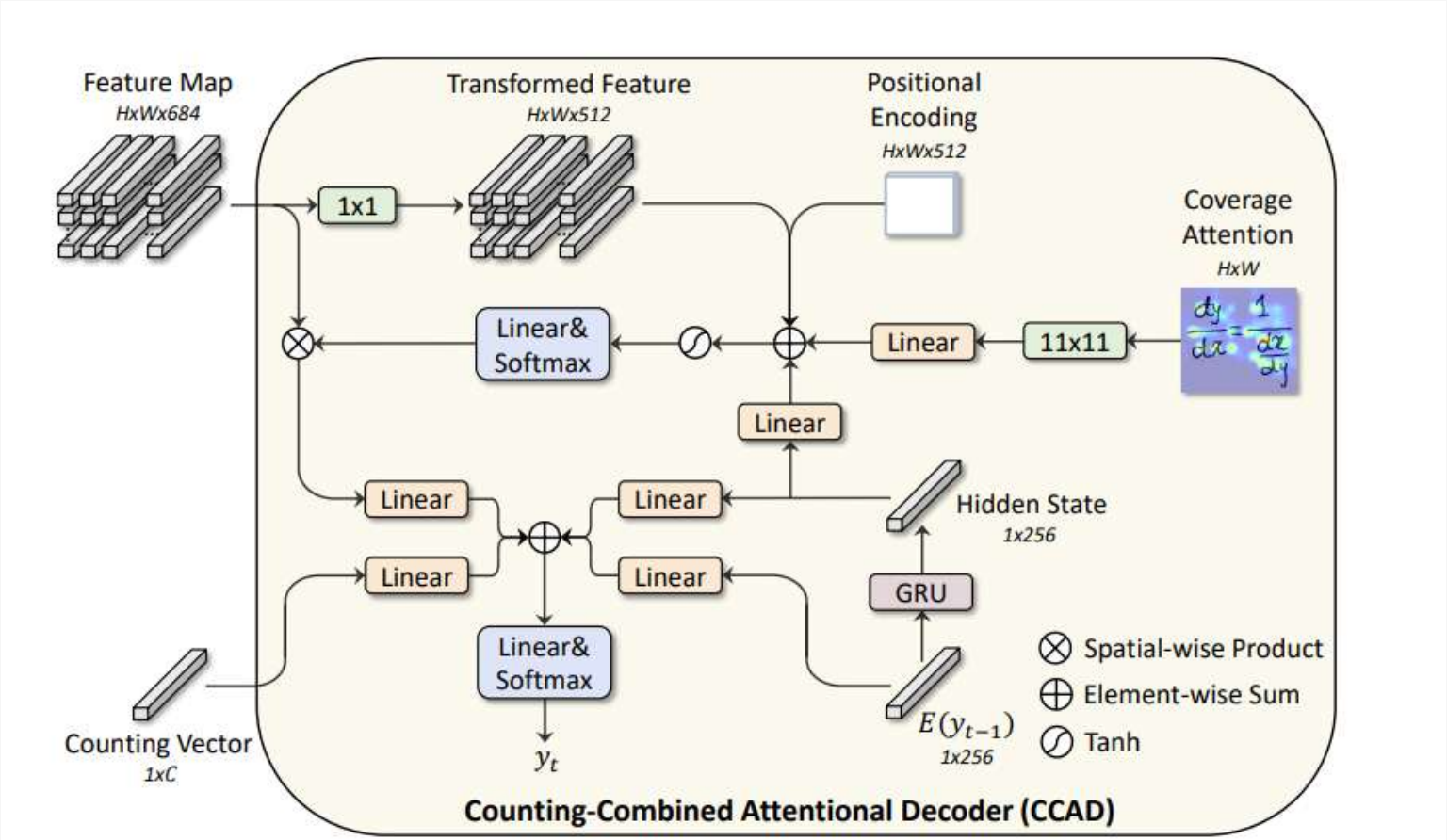
→ 1×1 Conv +Sigmoid →

$$\mathcal{V}_i = \sum_{p=1}^{H}\sum_{q=1}^{W} M_{i,pq}$$

Counting Vector 1xC →

Enhanced Feature　　Pseudo Density Map　　Sum Pooling

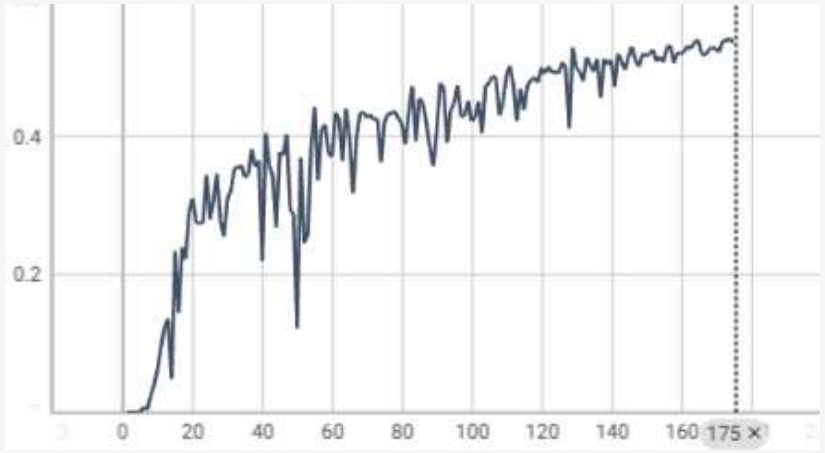
CCAD



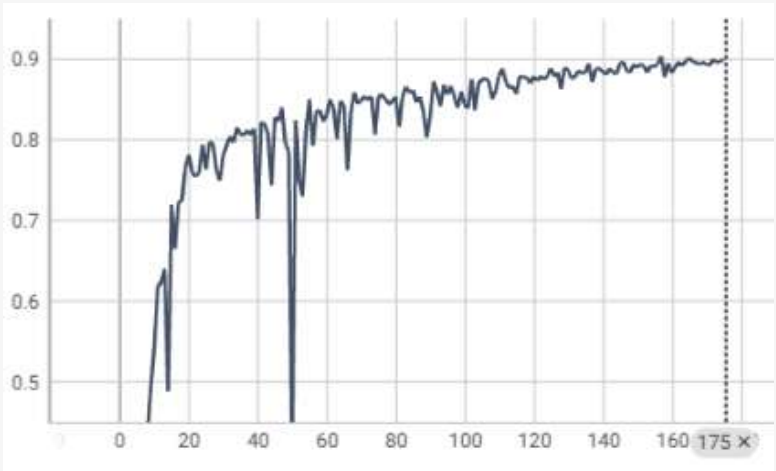
$$p(y_t) = \text{softmax}(w_o^T(W_c\mathcal{C} + W_v\mathcal{V} + W_t h_t + W_e E) + b_o,$$

$$y_t \sim p(y_t),$$

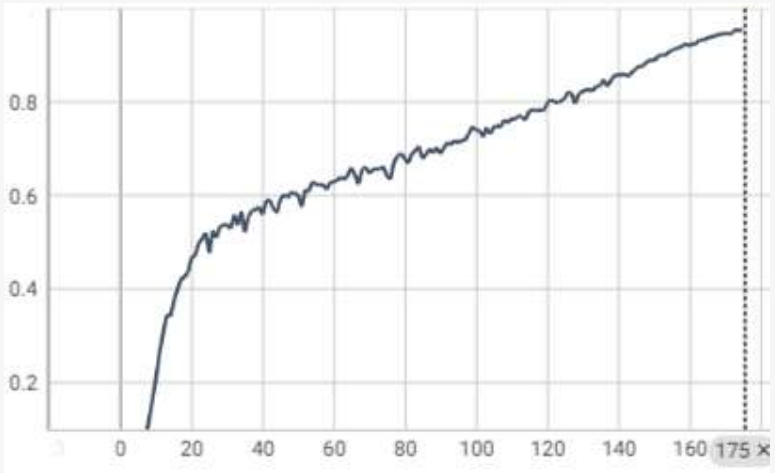
CAN Replicatioin



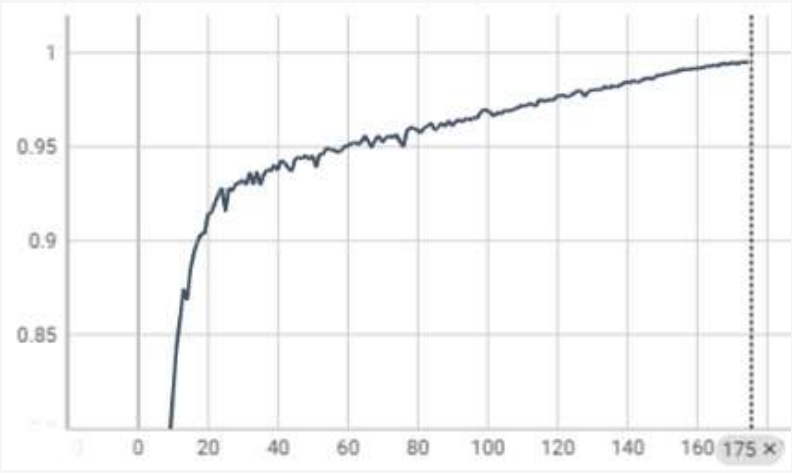
Evaluate ExpRate



Evaluate WordRate



Train ExpRate



Train WordRate

Improvement

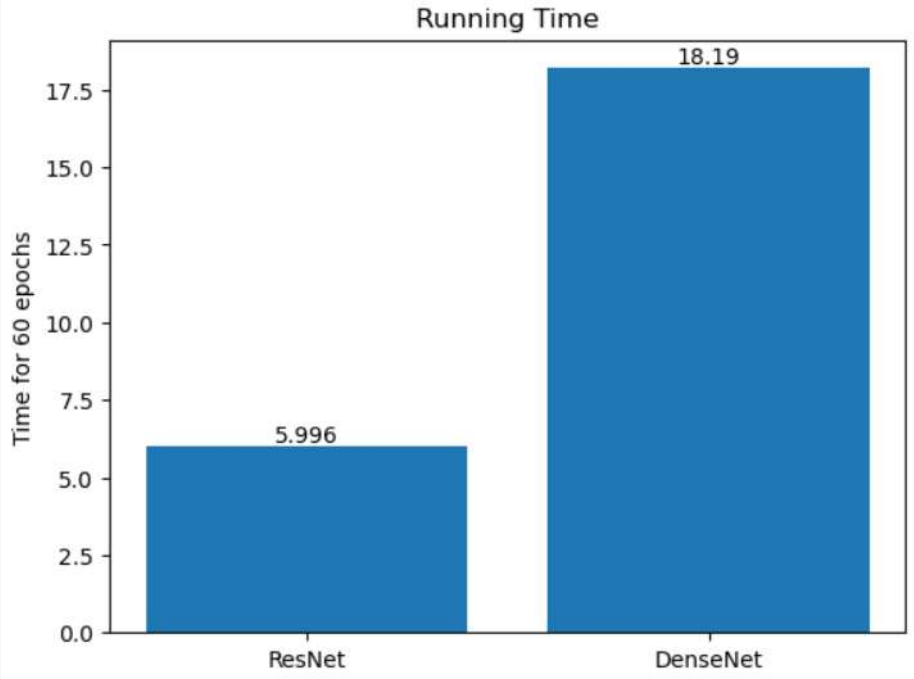
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

- Backbone: DenseNet→ResNet50

ResNet

Pros: Faster Training and Evaluating Speed
Faster Convergence Speed

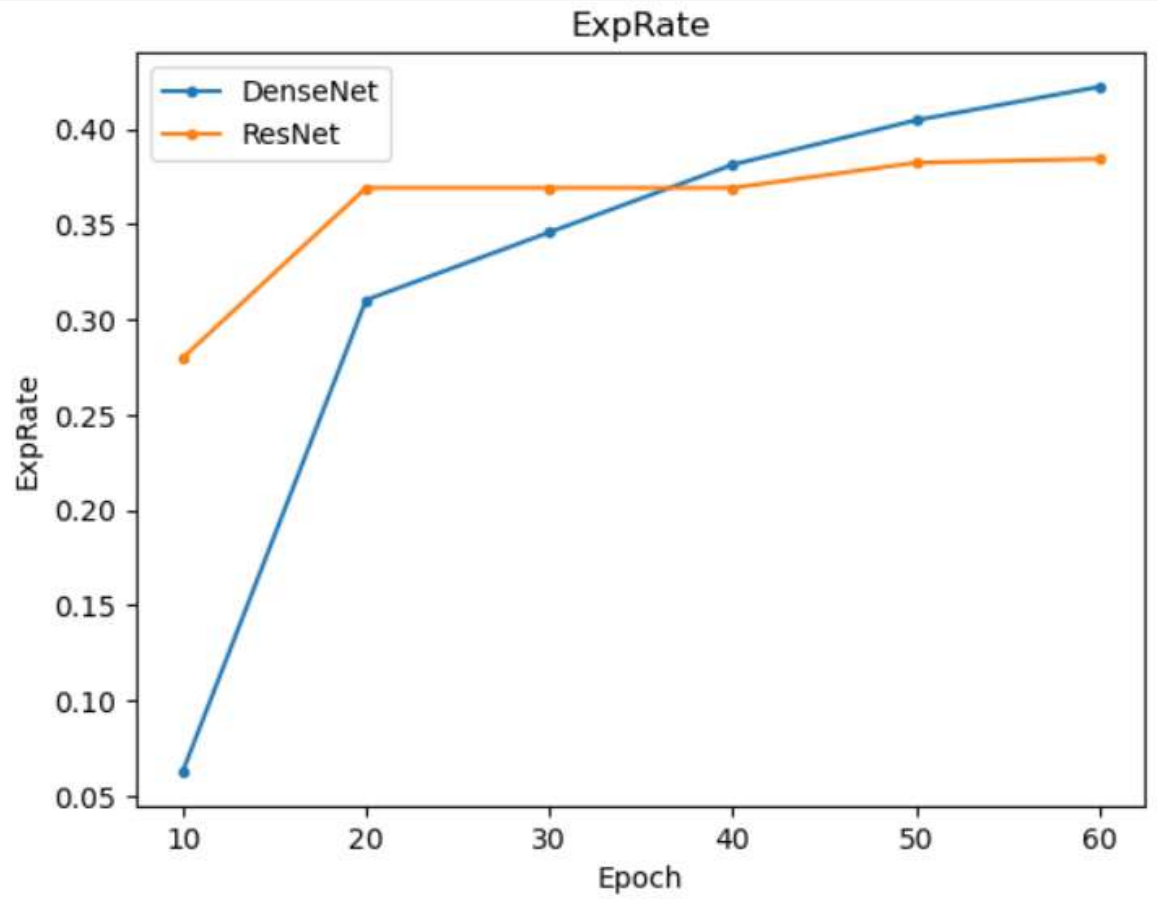
Cons: Lower Evaluating ExpRate
Higher Model Complexity(Overfitting)
3M params in DenseNet vs **15M** params in ResNet50

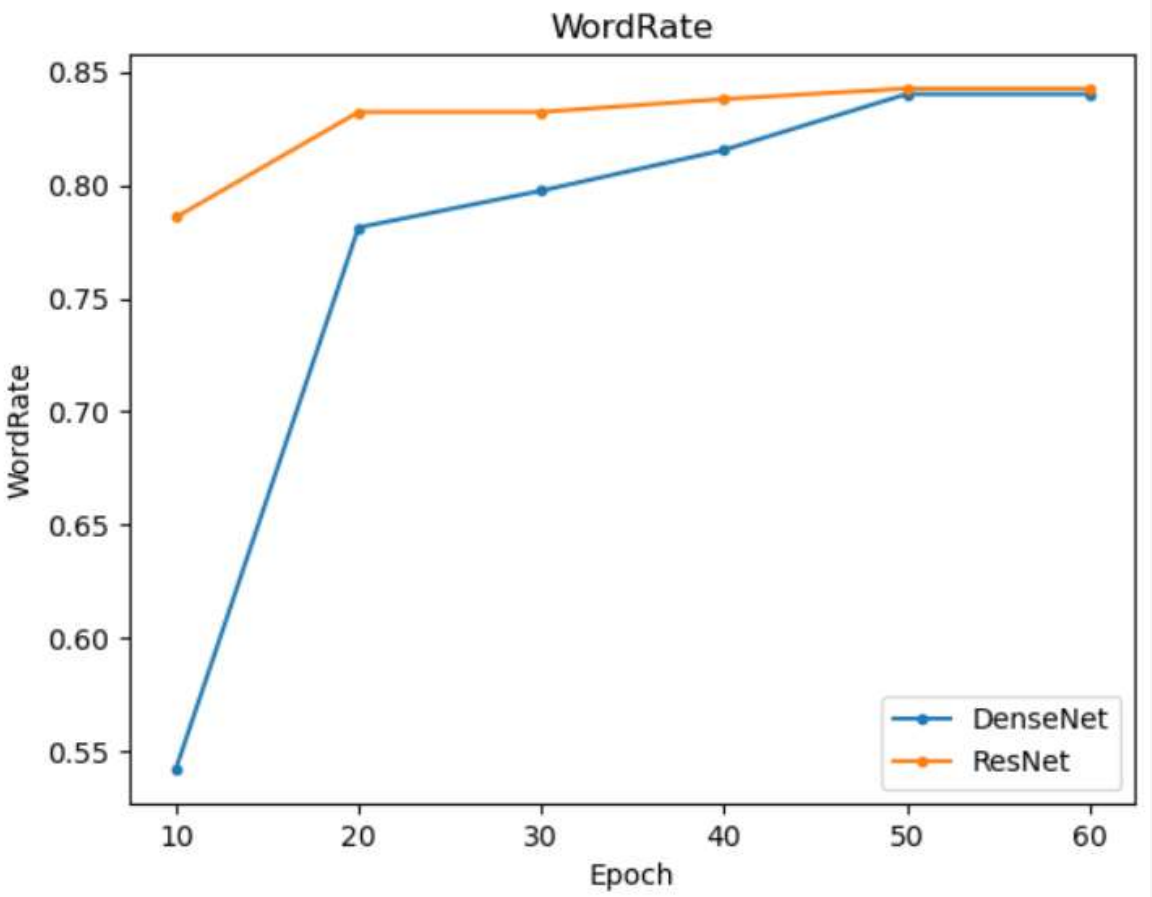


Improvement

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Evaluating

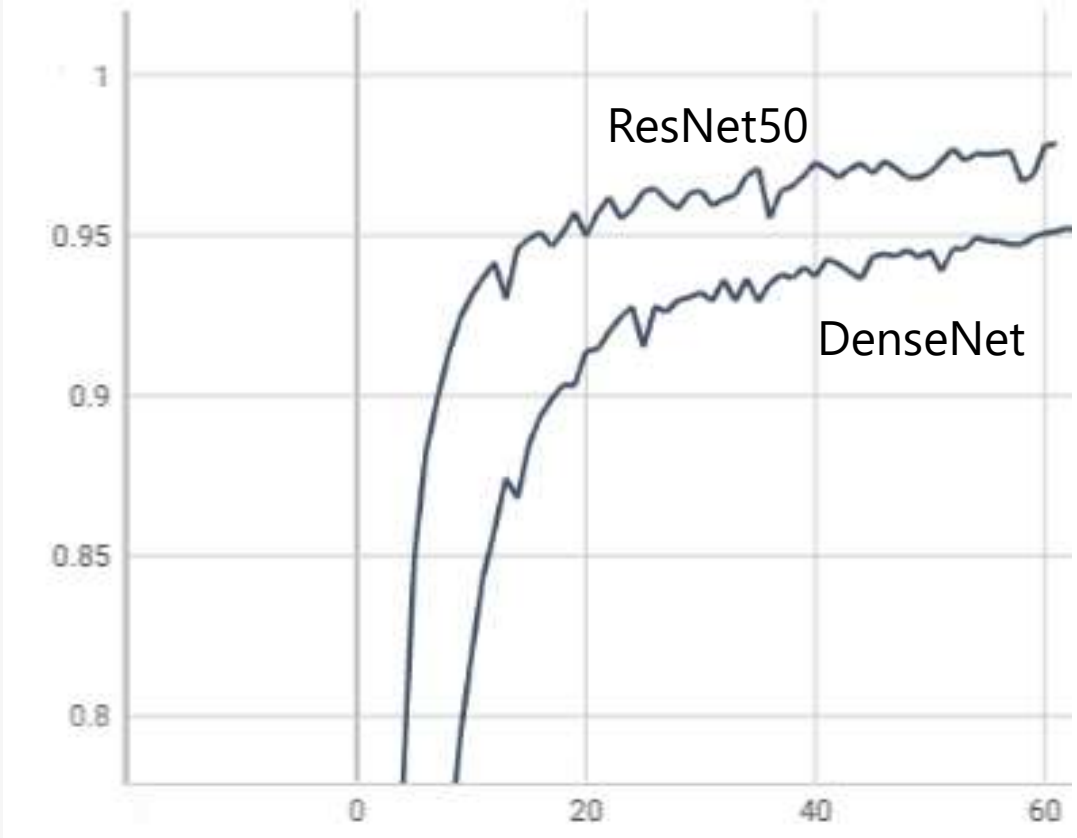




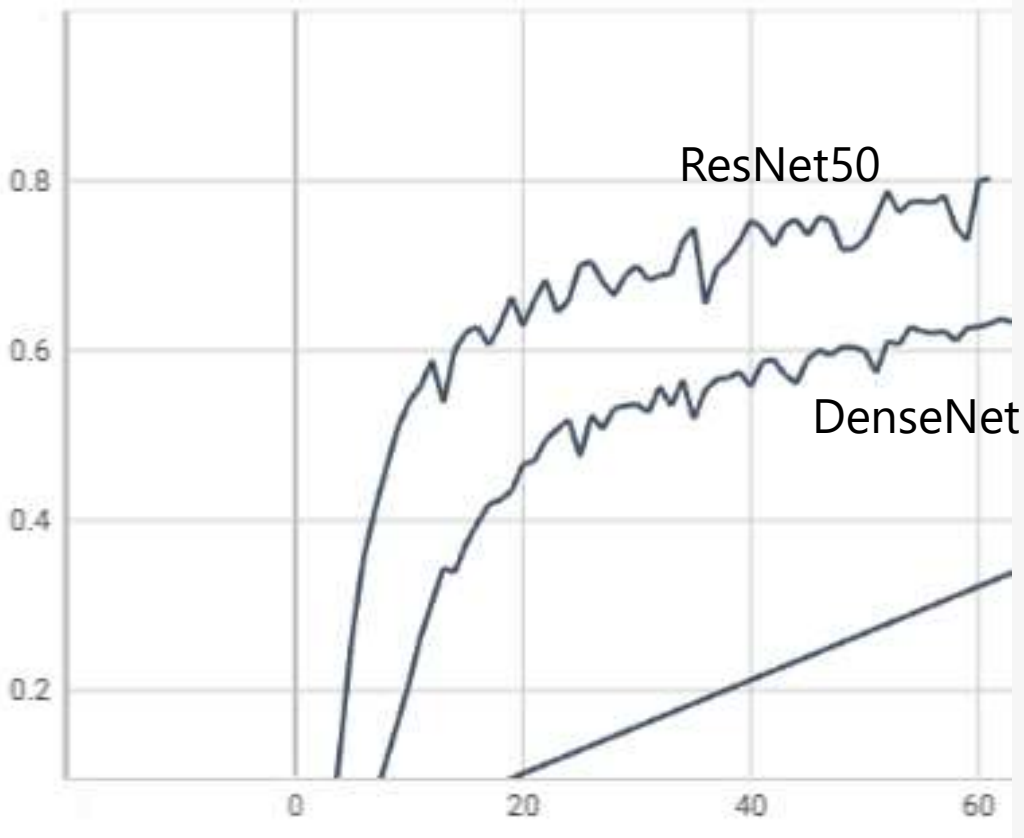
Improvement

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Training



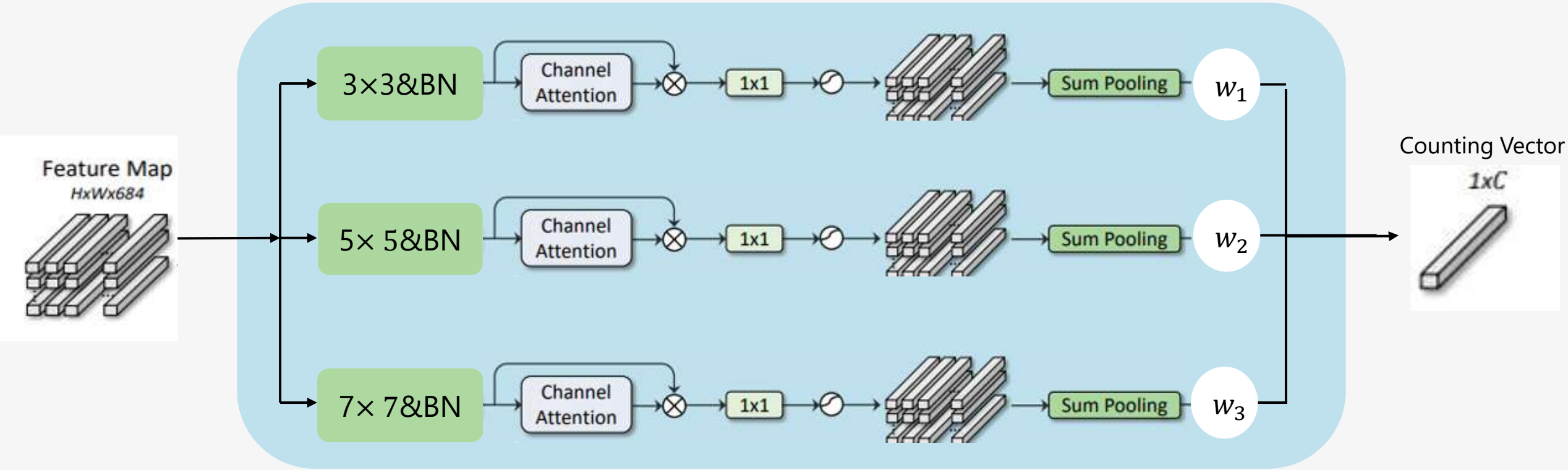
Training WordRate



Training ExpRate

Improvement

- MSCM: 3×3+5×5→ 3×3+5×5+7×7
- Average→Trainable weight w_1, w_2, w_3

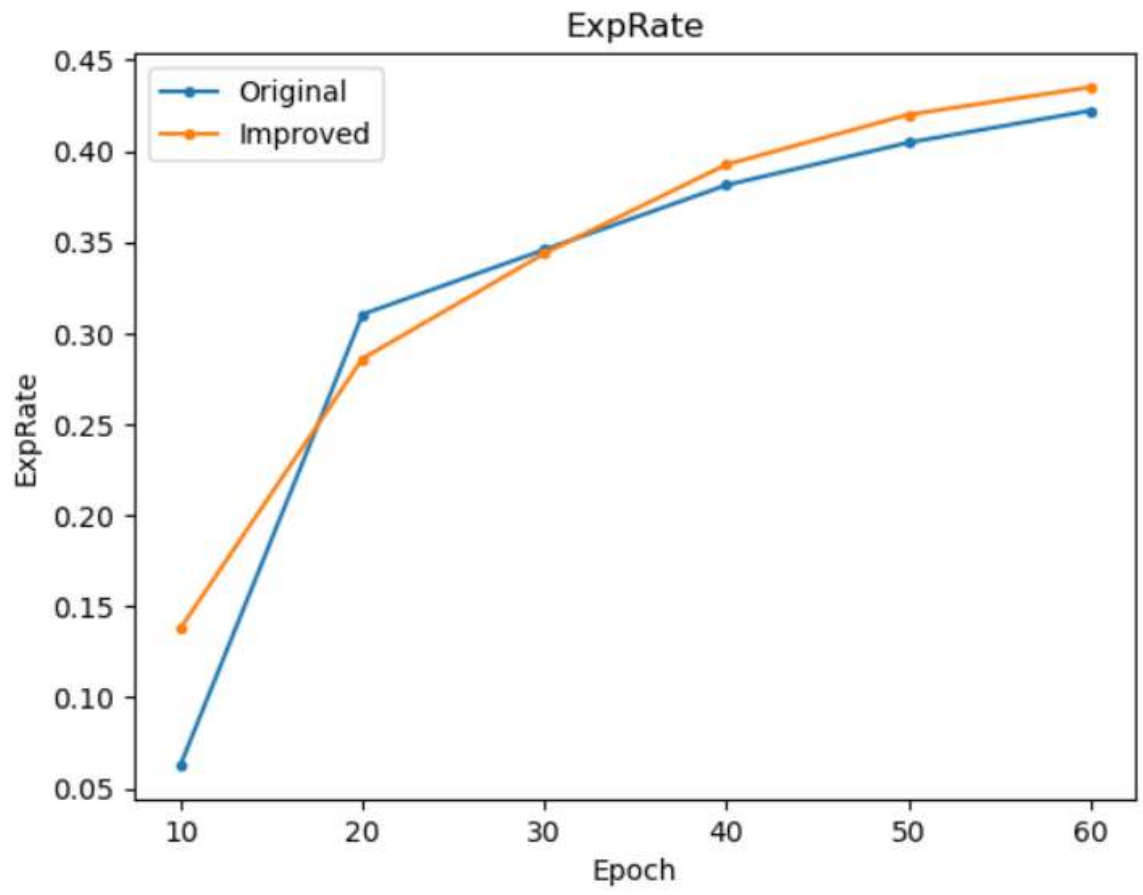


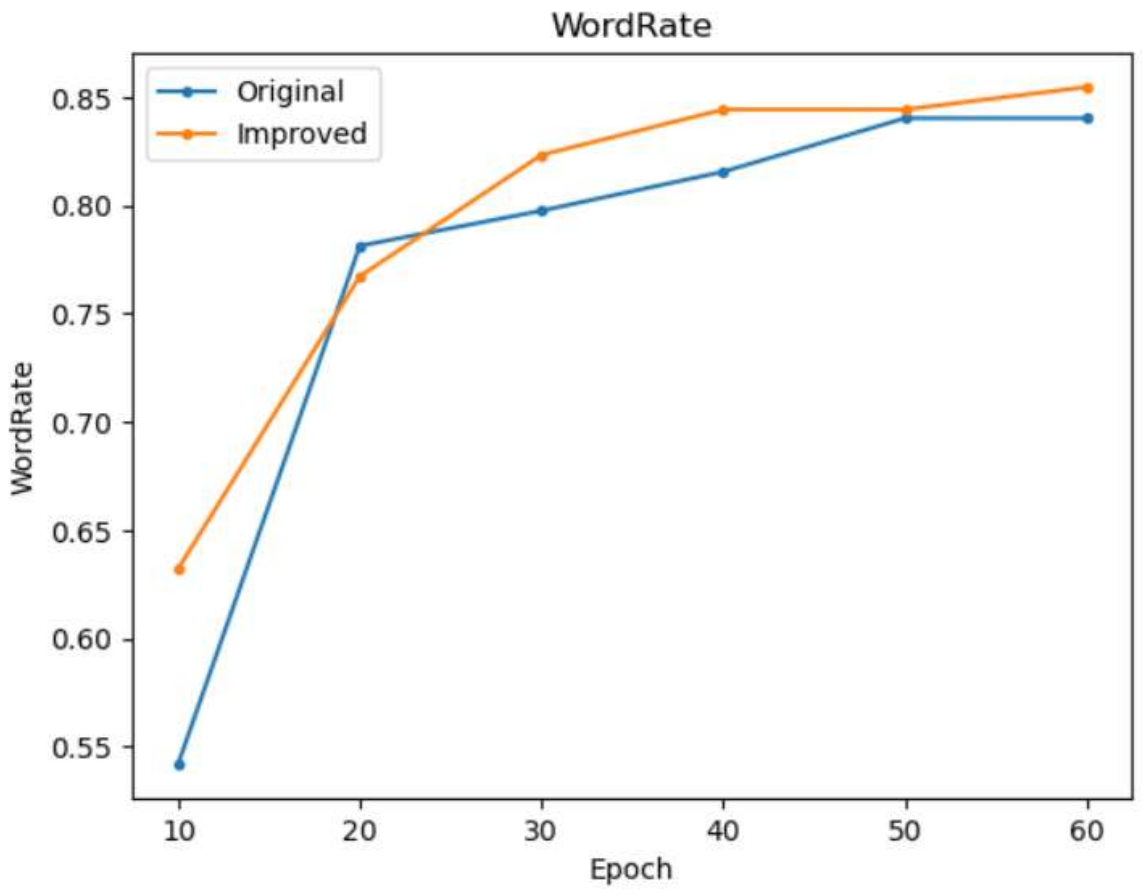
MSCM

Improvement

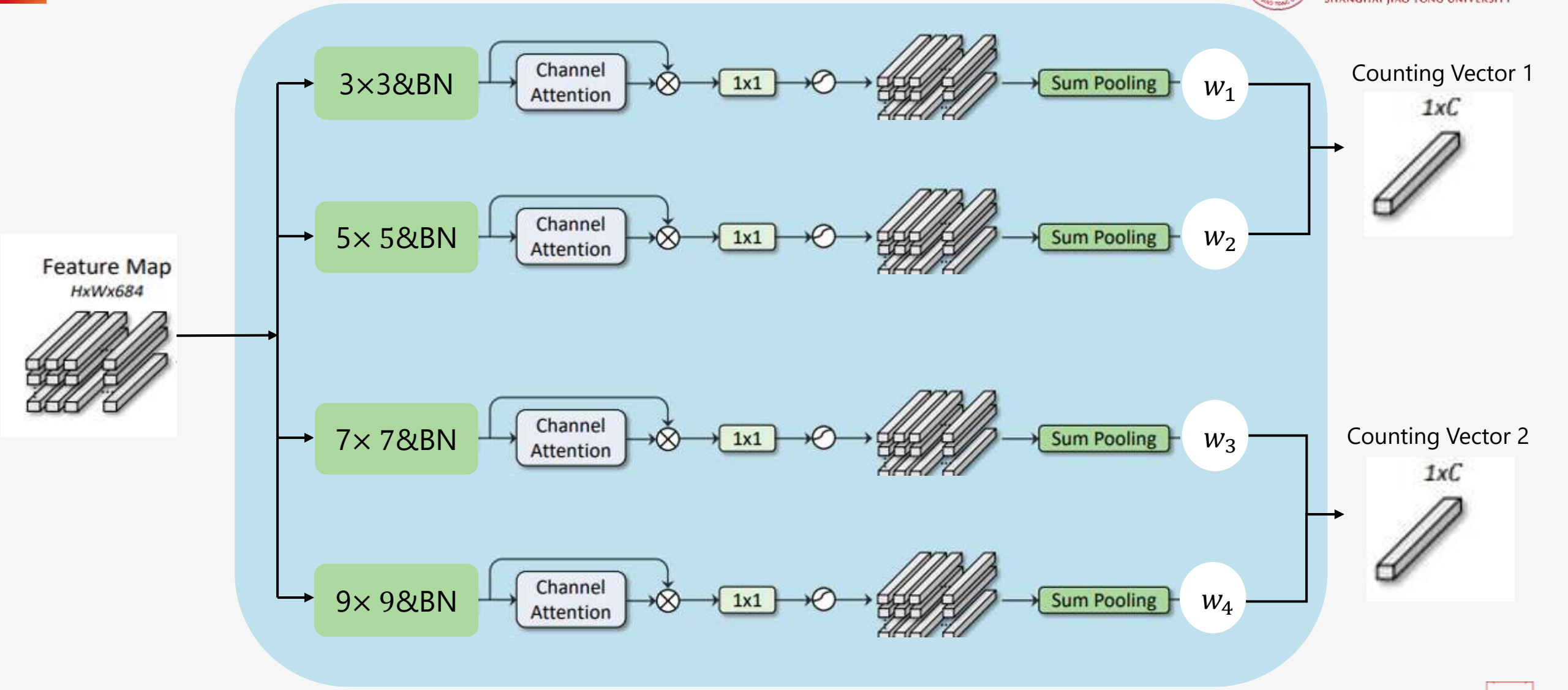
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Evaluating

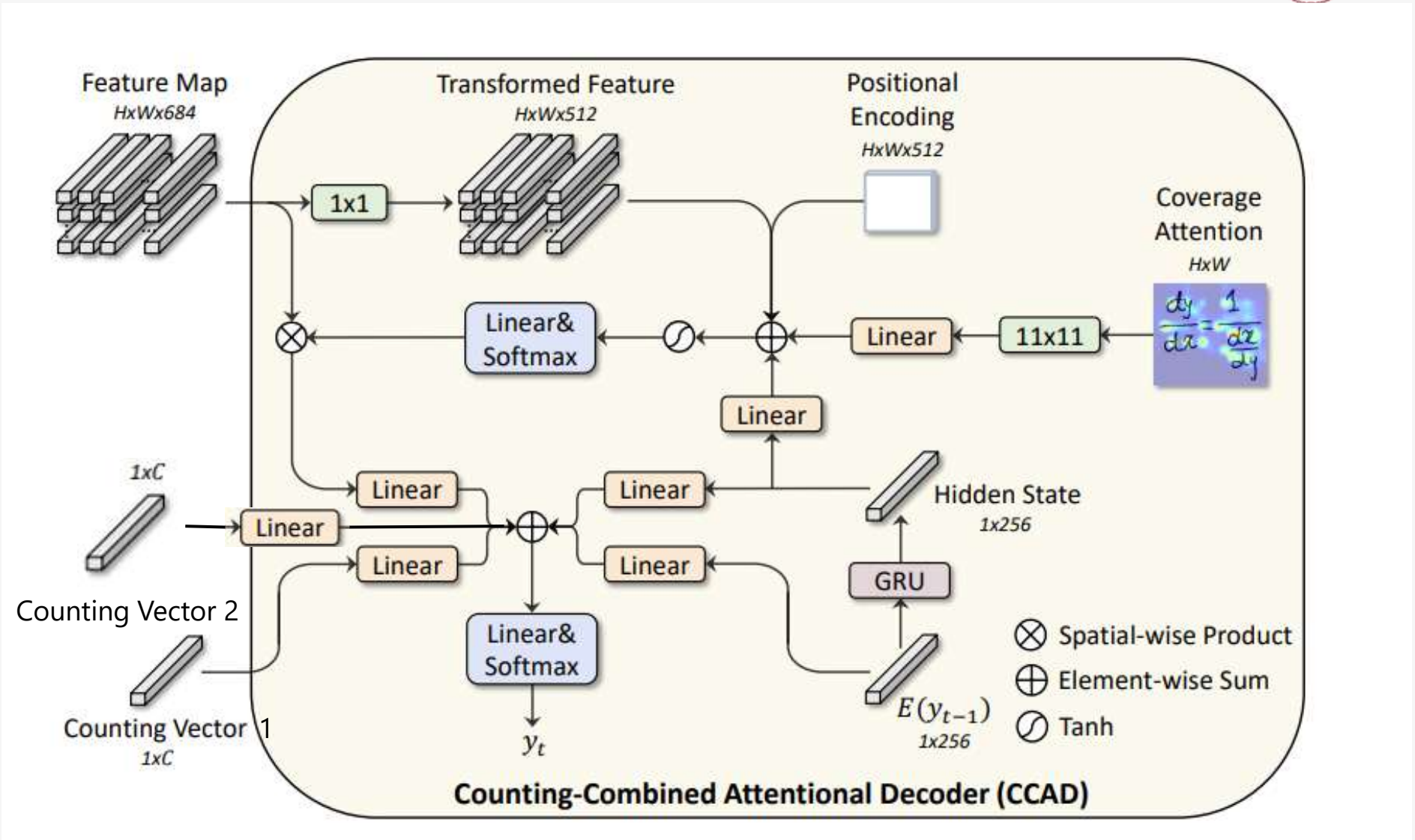




Further Improvement



Further Improvement

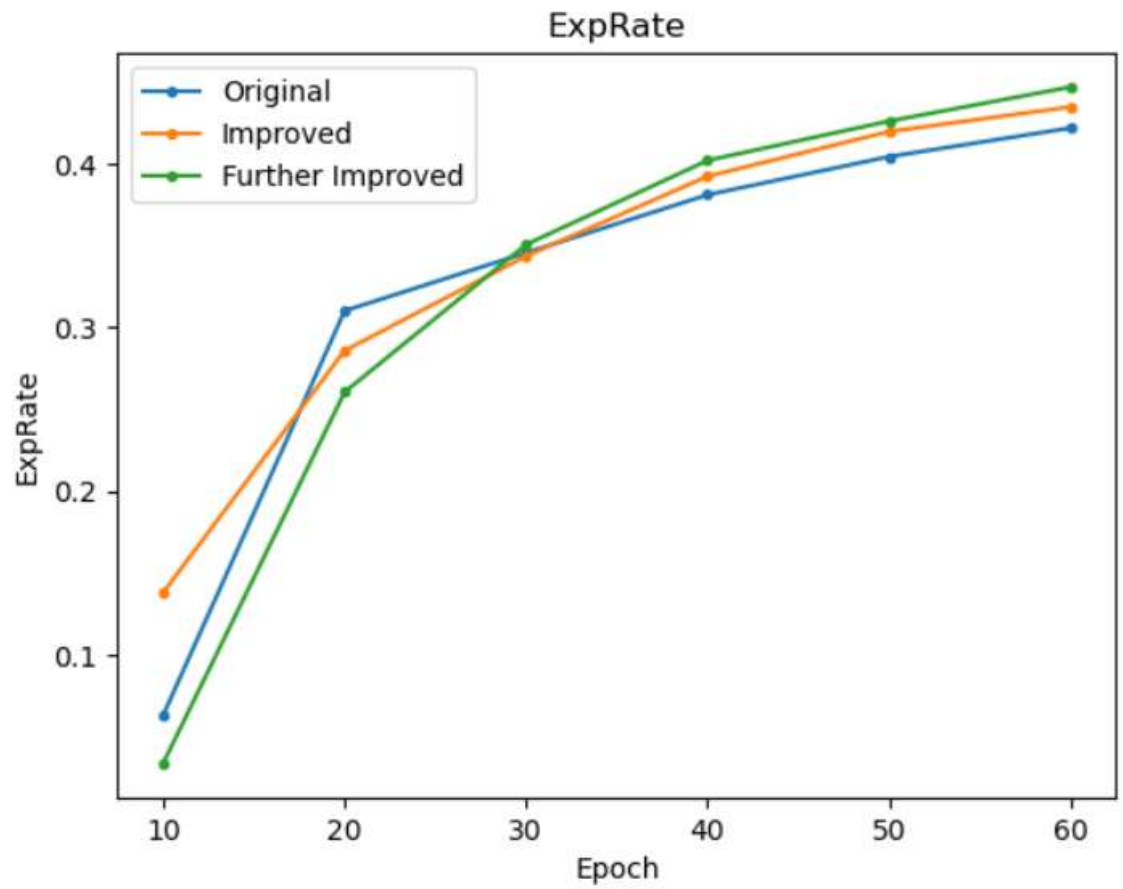


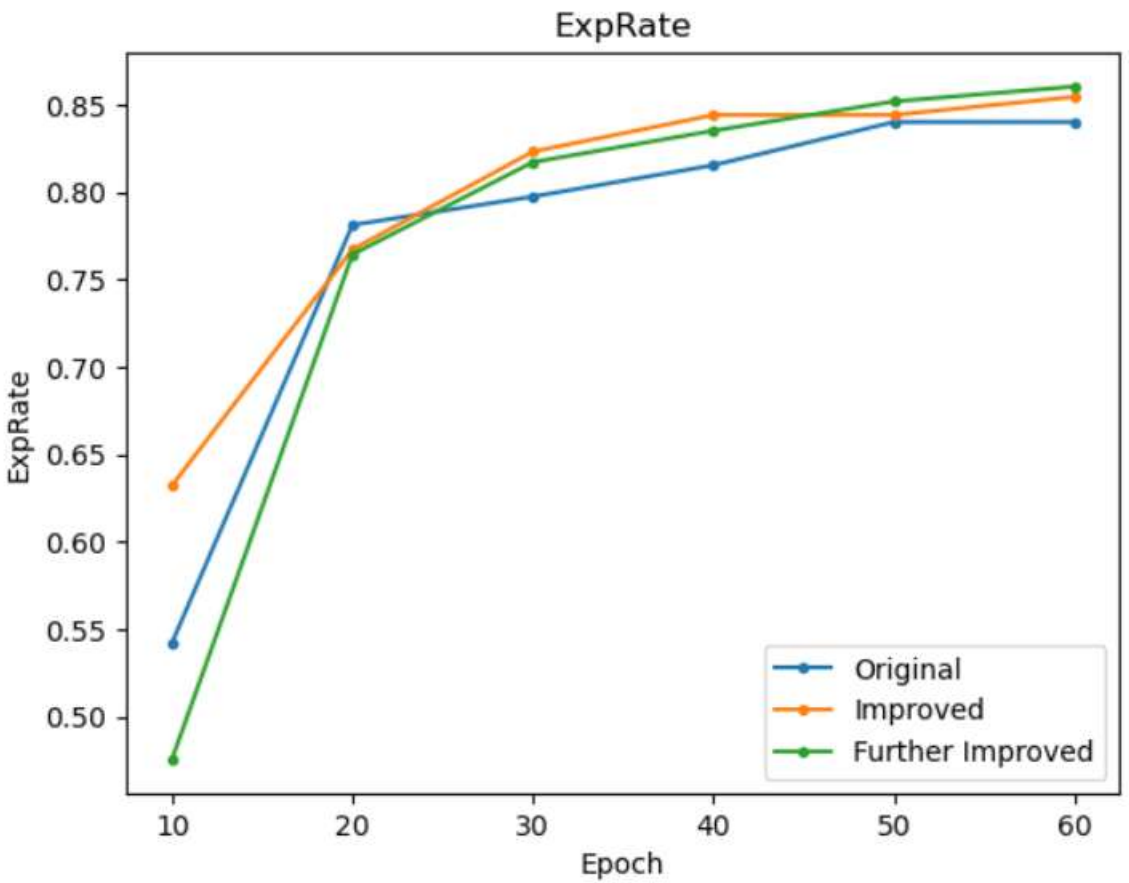
CCAD

Further Improvement

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Evaluating





上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

Thank you