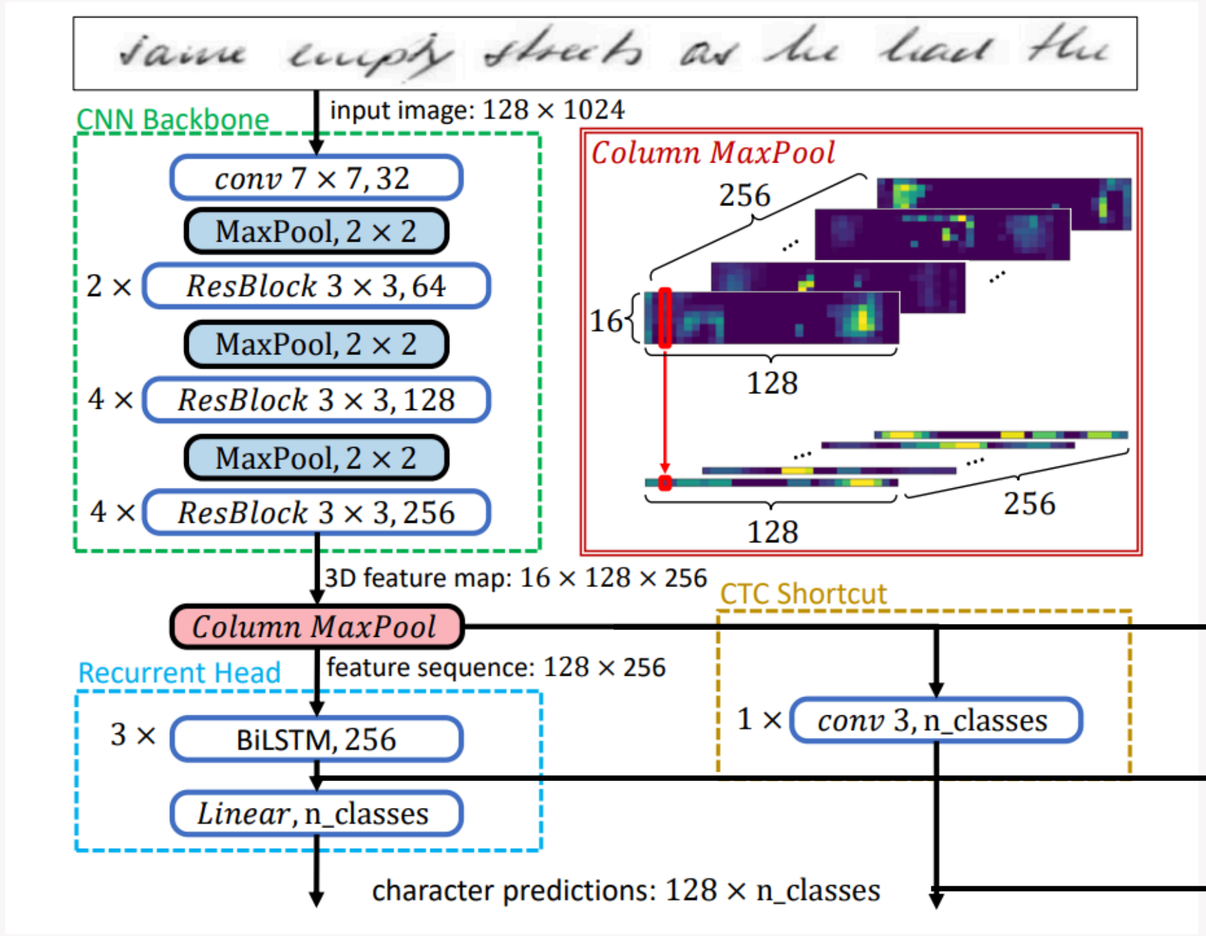
LLM条件付き確率最大化に基づく手書き文字認識の精度向上

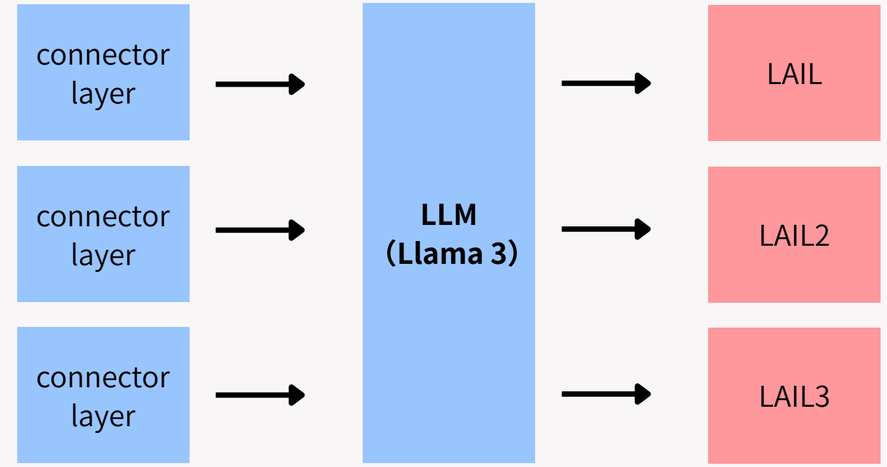
2025/10/08

前回のゼミでの疑問

1. 言語的文法的LLMタスクが解けるよう中間表現に働きかけるがその中間表現を獲得するCNNだけでは取れないのでは？？
2. もし,とれたとしてとれたものをRNNが具体的にどのようにくみ取ってもとの設計思想と喧嘩するのでは？？そうしないアルゴリズムに従った説明

提案手法





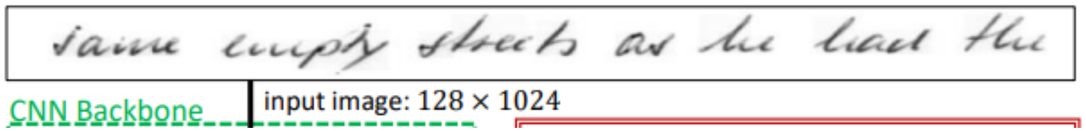
CTC Loss

CTC Loss2

[1]言語的文法的LLMタスクが解けるよう中間表現に働きかけるがその中間表現を獲得するCNNだけでは取れるのか

提案手法の組み合わせでCNNで長距離依存関係が取れるのかの調査

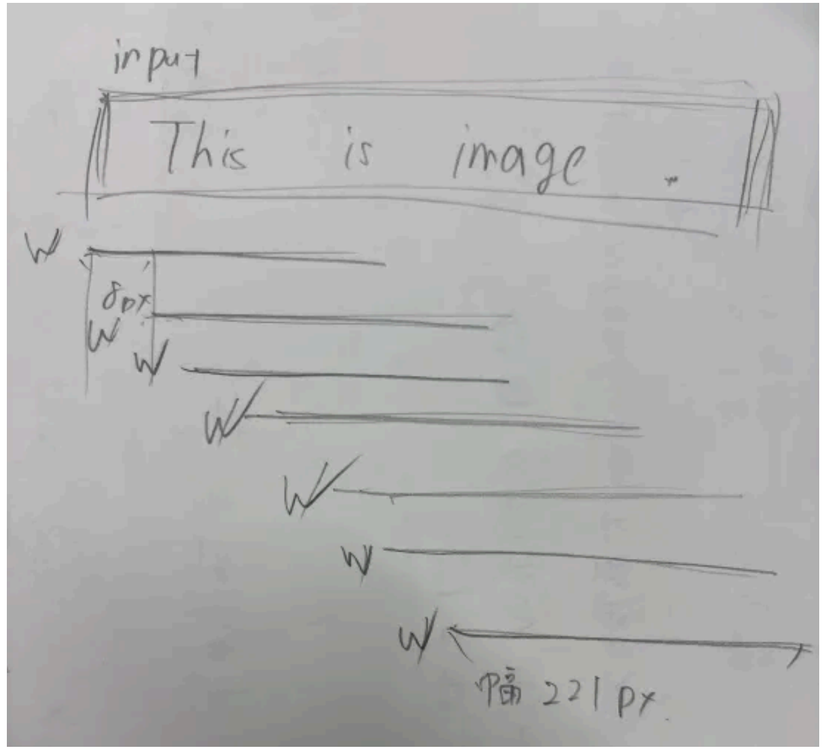
層	k	s	入力j	出力j	rの増分	r(累積)
7×7 conv	7	2	1	2	(7-1)*1=6	7
Res3×3 ×4 (Stage A)	3	1	2	2	4回×(3-1)*2=16	23
MaxPool 2×2	2	2	2	4	(2-1)*2=2	25
Res3×3 ×8 (Stage B)	3	1	4	4	8回×(3-1)*4=64	89
MaxPool 2×2	2	2	4	8	(2-1)*4=4	93
Res3×3 ×8 (Stage C)	3	1	8	8	8回×(3-1)*8=128	221
最終（幅方向）	–	–	–	**8**	–	**221**



output: feature sequence: 128 × 256

最終出力のある1セルは、元画像の幅221pxぶんの情報をまとめて見ていて、流石に、、、
意図したような言語的特徴をくみ取ってくれないかも

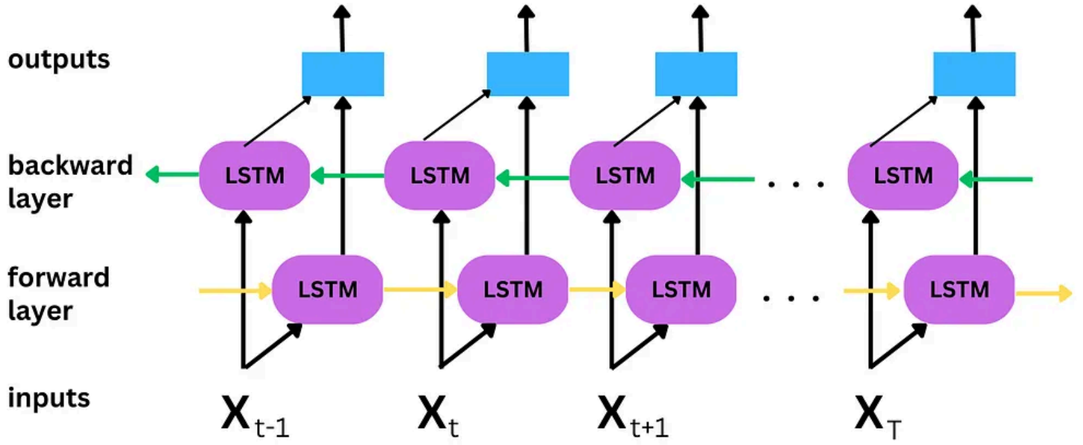
[1]言語的文法的LLMタスクが解けるよう中間表現に働きかけるがその中間表現を獲得するCNNだけでは取れるのか

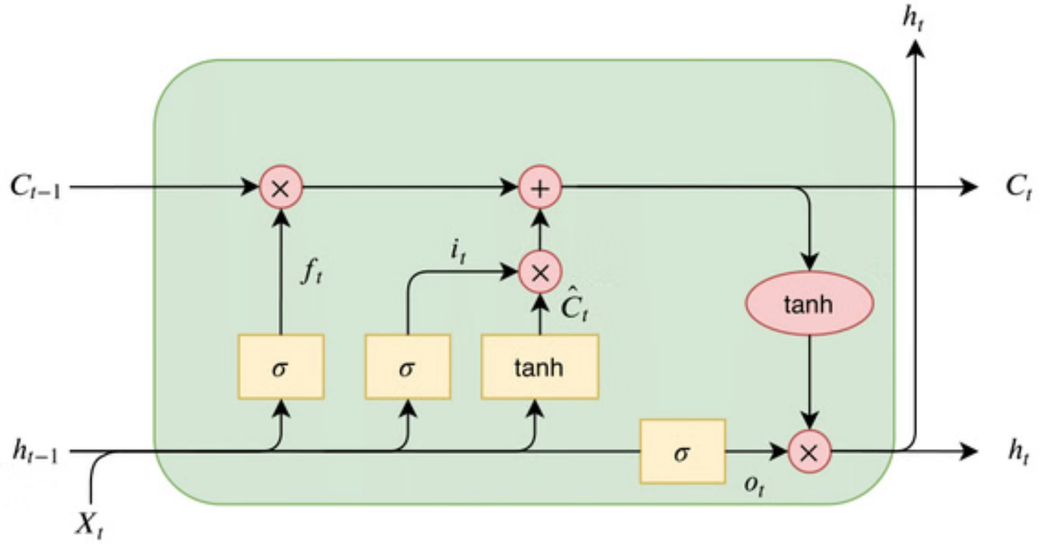


ただ、「BiLSTMの第1層出力に対しても補助的な言語モデリング損失を課しているので学習の進行に伴ってBiLSTMの層では、文脈表現を強化できそう

[2]RNNが有効活用できるかアルゴリズムから説明

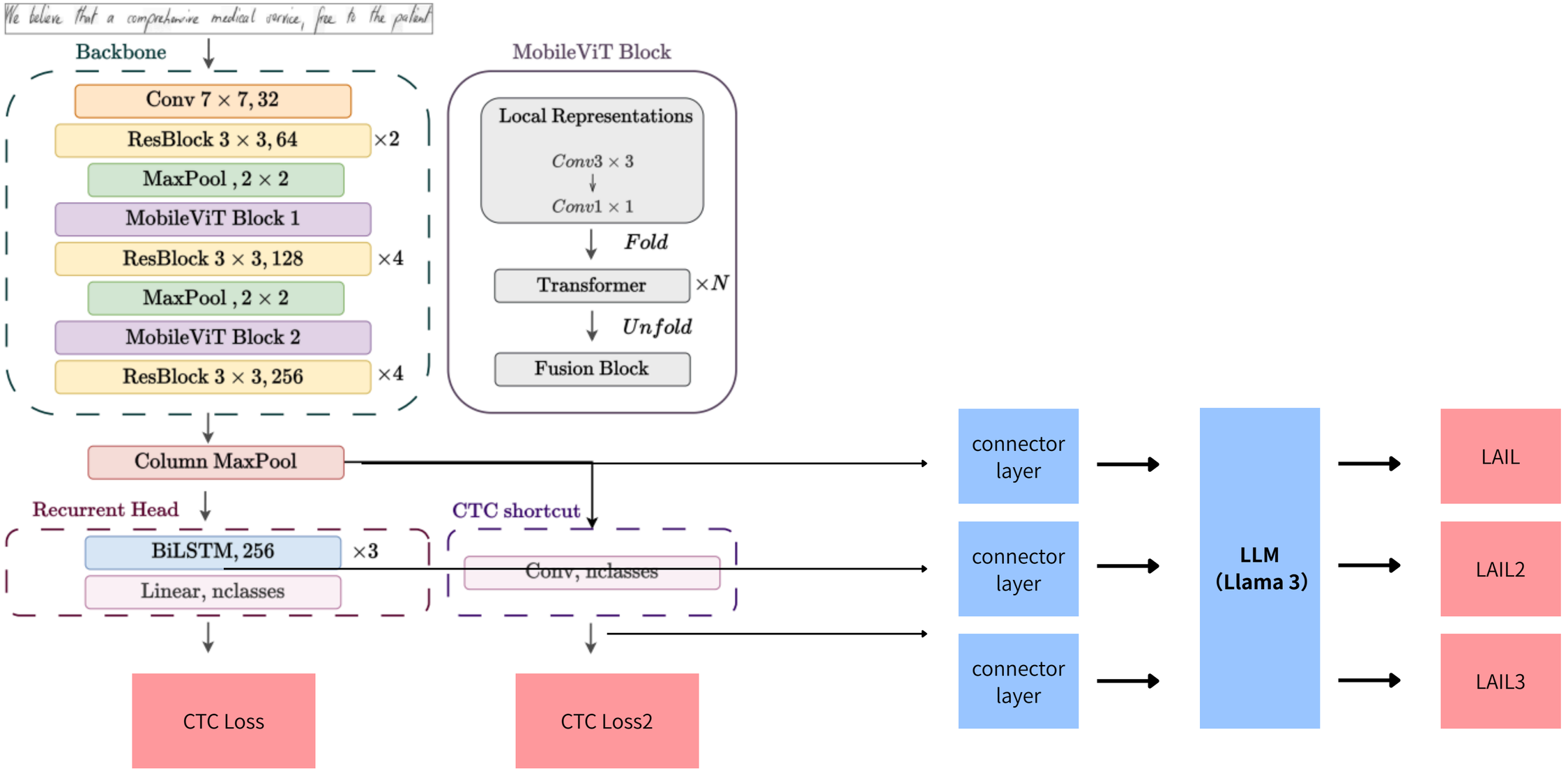
1. ゲートの分離性
2. “ブランク vs 文字”の後段分離が容易





$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad \text{（入力ゲート）}$$
$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad \text{（忘却ゲート）}$$
$$g_t = \tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \quad \text{（候補情報）}$$
$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad \text{（出力ゲート）}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad \text{（セル状態，足し算の記憶）}$$
$$h_t = o_t \odot \tanh(c_t) \quad \text{（隠れ状態）}$$

改善策


We believe that a comprehensive medical service, free to the patient
Backbone
Conv 7 × 7, 32
ResBlock 3 × 3, 64
×2
MaxPool , 2 × 2
MobileViT Block 1
ResBlock 3 × 3, 128
×4
MaxPool , 2 × 2
MobileViT Block 2
ResBlock 3 × 3, 256
×4
MobileViT Block
Local Representations
Conv3 × 3
Conv1 × 1
Fold
Transformer
×N
Unfold
Fusion Block
Column MaxPool
Recurrent Head
BiLSTM, 256
×3
Linear, nclasses
CTC shortcut
Conv, nclasses
CTC Loss
CTC Loss2
connector layer
connector layer
connector layer
LLM (Llama 3)
LAIL
LAIL2
LAIL3

mobile vit の精度

	ckpt	val_CER	val_WER	test_CER	test_WER
1	200	0.041	0.142	0.058	0.193
2	250	0.041	0.141	0.056	0.189
3	300	0.043	0.15	0.061	0.205
4	350	0.049	0.172	0.065	0.217
5	400	0.037	0.133	0.051	0.174
6	450	0.032	0.113	0.046	0.159
7	500	0.032	0.114	0.045	0.154
8	550	0.031	0.11	0.045	0.152
9	600	0.032	0.114	0.044	0.151
10	650	0.031	0.11	0.044	0.15
11	700	0.031	0.111	0.044	0.15
12	750	0.031	0.111	0.044	0.15
13	800	0.031	0.111	0.044	0.149