経過報告

2025年11月19日　工藤滉青

はじめに

やってきたこと

- コネクタ層比較
- 実験結果
- LLMのlossについて
- 今後の展望

実験

コネクタ層の良し悪し

実験設定1

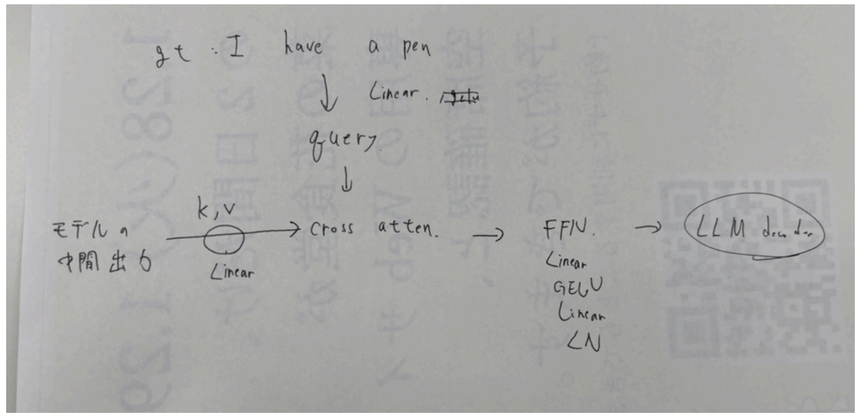
- コネクタ層:linear,gelu,ln

こいつめっちゃ時間かかる...

理由はあとで

実験設定2

- コネクタ層:Q-Former



線形層のみの差

Total samples: 2915

- Reference Loss (GT Text → Text): 5.7355
- Connector Loss (Image → Text, padding): 6.9723
- Downsampled Loss (Image → Text, resized): 8.3487

📈 Gaps:

Connector - Reference: 1.2368

Downsampled - Reference: 2.6132

Connector - Downsampled: 1.3764

Q-Formerのみの差

Total samples: 2915

- Reference Loss (GT Text → Text): 5.7300
- Connector Loss (Image → Text, padding): 6.9113

📈 Gaps:

Connector - Reference: 1.1813

実験

実験結果

コネクタ層は線形層本番

▼ 結果ログ

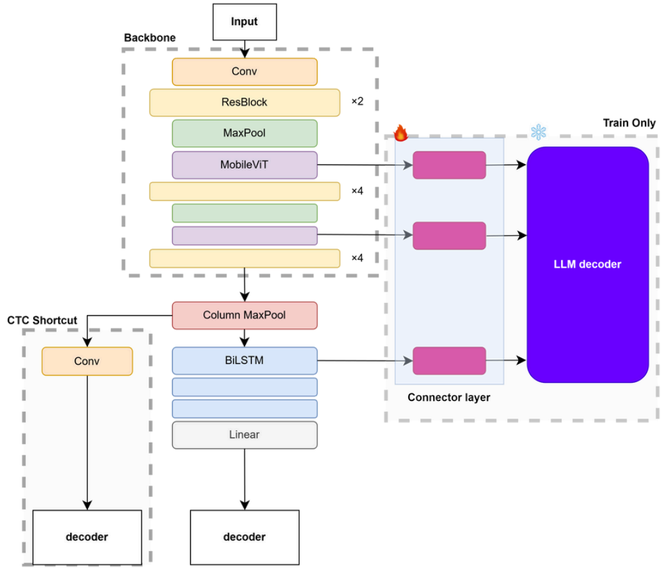
```
[val] CER at epoch 0: 0.0333
[val] WER at epoch 0: 0.1194
[test] CER at epoch 0: 0.0468
[test] WER at epoch 0: 0.1596
```

GPT3層+Q-Former

▼ 結果ログ

```
[val] CER at epoch 0: 0.0328
[val] WER at epoch 0: 0.1180

[test] CER at epoch 0: 0.0455
[test] WER at epoch 0: 0.1558
```



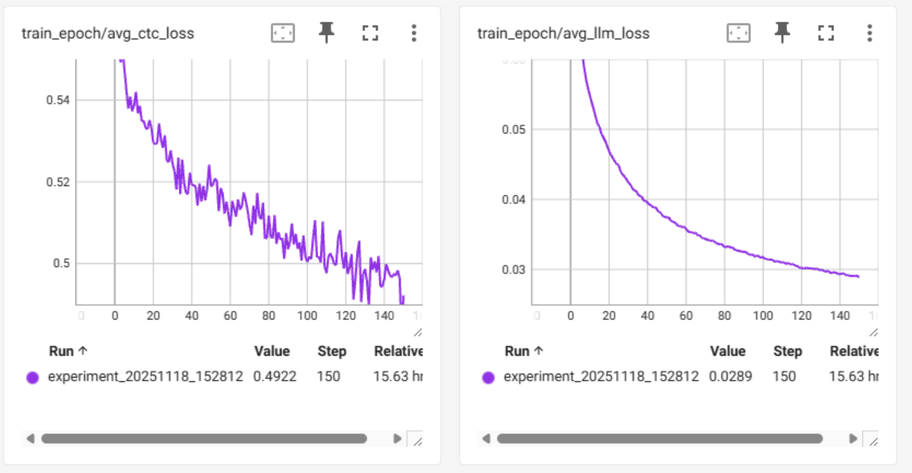
現状

LLMの損失の改善案模索結果

$$\log P_{b,t,v} = \log \mathrm{softmax}(\mathrm{connector\_logits}_{b,t,:})_v, \qquad Q_{b,t,v} = \mathrm{softmax}(\mathrm{reference\_logits}_{b,t,:})_v.$$

PyTorch の `F.kl_div(log_probs_connector, probs_reference)` の使い方に従うと、各バッチ・各時刻で計算される語彙上の和は次の式になります（Reference をターゲットにした KL）：

$$\mathrm{KL}_{b,t} = \sum_v Q_{b,t,v} \left(\log Q_{b,t,v} - \log P_{b,t,v} \right) = \sum_v Q_{b,t,v} \log Q_{b,t,v} \; - \; \sum_v Q_{b,t,v} \log P_{b,t,v}.$$



GPT3層＋Q-Former＋KDloss

▼ 結果ログ

PowerShell

```
[val] CER at epoch 0: 0.0326
[val] WER at epoch 0: 0.1163
[test] CER at epoch 0: 0.0451
[test] WER at epoch 0: 0.1532
```

現状

LLMの損失の改善案模索結果

$$\log P_{b,t,v} = \log \mathrm{softmax}(\mathrm{connector\_logits}_{b,t,:})_v, \qquad Q_{b,t,v} = \mathrm{softmax}(\mathrm{reference\_logits}_{b,t,:})_v.$$

PyTorch の `F.kl_div(log_probs_connector, probs_reference)` の使い方に従うと、各バッチ・各時刻で計算される語彙上の和は次の式になります（Reference をターゲットにした KL）：

$$\mathrm{KL}_{b,t} = \sum_v Q_{b,t,v} \left(\log Q_{b,t,v} - \log P_{b,t,v} \right) = \sum_v Q_{b,t,v} \log Q_{b,t,v} \; - \; \sum_v Q_{b,t,v} \log P_{b,t,v}.$$

現状

全体ゼミを終えてとか実験を終えて今後の展望

1. 線形層gelu 抜いて実験してみる
2. 1やと正直めっちゃ時間かかる→ダウンサンプリングすべきなのかいい塩梅でやるべきか
3. コネクタ層どこに引っ付けるべきか
4. ハイパラの比率何がいいのか