研究計画書

1 今週の進捗

- 手書き絵コンテの GPT OCR の結果比較
- 矢印 OCR のファインチューニング

2 手書き絵コンテの GPT OCR の結果比較

使用モデル: GPT-5.2 Thinking, Qwen3-VL-4B-Instruct

いただいた絵コンテの Cut, Action Memo, Dialogue に適当に考えたシナリオを作成し, それを手書きしてスキャンした画像を上記の 2 つのモデルで比較した. 表 1 に元の絵コンテにおける切り抜き画像と, スキャンした絵コンテの切り抜き画像を示す. いただいた絵コンテのデータをコピーして手書きしたあとにスキャンしているため, スキャン後の画像は少し線が薄くなっている. しかし, Piture においては今回重視する箇所は矢印であるため問題ないと考えている. また, Action Memo 等は手書きの文字がはっきりしている.

表 1: picture

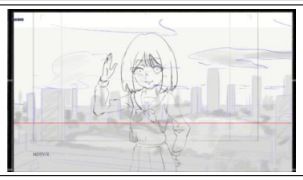
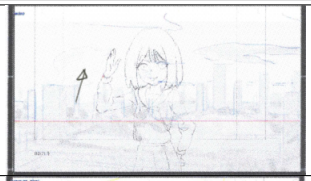
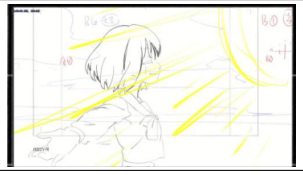
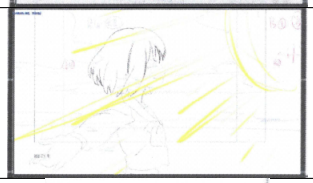
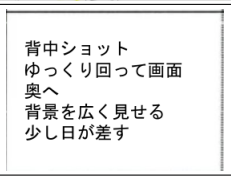
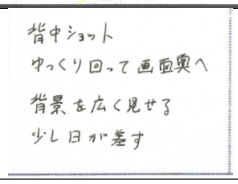
ID	元の画像	スキャン後の画像
page1 row3 picture		
page2 row3 picture		
page1 row5 action memo	 背中ショット ゆっくり回って画面奥へ 背景を広く見せる 少し日が差す	 背中ショット ゆっくり回って画面奥へ 背景を広く見せる 少し日が差す

表 2 に OCR の結果を示す. GPT と Qwen3-VL-4B の差はほとんどなく, “page1 row1 cut” でのみ差があった. また, “page2 row2 dialogue” ではどちらのモデルも “95+12” と出力しているが, 入力画像をみると正解は “05+12” であり, どちらも間違いとなった. 以下に載せている結果以外はすべてどちらも正解していた.

3 矢印 OCR のファインチューニング

3.1 矢印特化の評価指標

文字誤り率 Character Error Rate 出力文字列と正解文字列の編集距離を正解文字列長で割った値である。本研究では空白の差を吸収するために空白を正規化し、さらに中立記号である □ と △ を除去した上で計算する。値が小さいほど良い。なお、出力が正解より極端に長い場合には 1 を超えることがある。

完全一致率 Exact Match　正規化後の出力文字列と正規化後の正解文字列が完全に一致したサンプルの割合である。値が大きいほど良い。

矢印有無再現率　正解に矢印が含まれるサンプルに限定し、出力に矢印が一つ以上含まれた割合である。矢印の見逃しに対する強さを表す。値が大きいほど良い。

矢印個数完全一致率　正解に矢印が含まれるサンプルに限定し、出力の矢印個数が正解の矢印個数と一致した割合である。値が大きいほど良い。

矢印個数の平均絶対誤差　正解に矢印が含まれるサンプルに限定し、出力矢印数と正解矢印数の差の絶対値を平均したものである。値が小さいほど良い。

矢印方向の適合率、再現率、F1　矢印の方向を $\rightarrow, \leftarrow, \uparrow, \downarrow$ の四種類として数え上げ、マイクロ平均で適合率、再現率、F1 を計算する。過検出に弱い場合は適合率が低下し、見逃しに弱い場合は再現率が低下する。

矢印列完全一致率　正解に矢印が含まれるサンプルに限定し、出力に含まれる矢印の並び順が正解と完全一致した割合である。値が大きいほど良い。

矢印列の編集率、矢印列の類似度　矢印列に対する編集距離を用いて、編集率は正解矢印列長で正規化した値である。類似度は $1 - \Gamma\Gamma\Gamma$ として $[0, 1]$ に収まるように定義している。値が大きいほど良い。

4　質問

- 実験結果で見せる指標は “片田舎” を使用したもの, 修論と発表で見せる画像は Trigger データセット, GPT との比較で用いるのは 作成していただいた絵コンテ, という認識であっているのか.
- はたなかさん: 矢印等の切り抜き画像は見せてもいいのか.

5　今後の予定

- シーン説明文の生成および GPT との比較
- Trigger データセットでの実装

参考文献

表 2: cut, action memo, dialogue, time

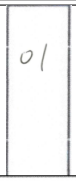
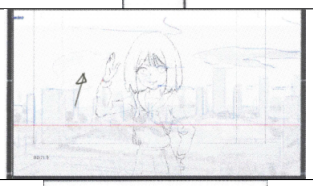
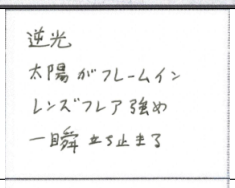
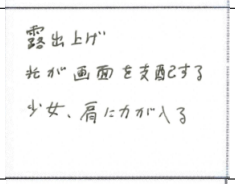
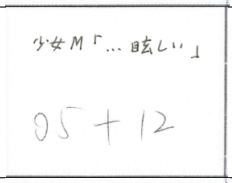
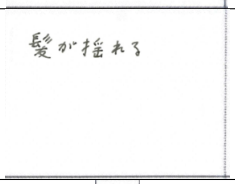
ID	入力画像	GPT-5.2 Thinking	Qwen3-VL-4B-Instruct
page1 row1 cut		01	01
page1 row3 picture		↑	□
page2 row1 cut		逆光 太陽がフレームイン レンズフレア強め 一瞬まぶしさ生まる	逆光 太陽がフレームイン レンズフレア強め 一目瞬立ち止まっ了
page2 row2 action memo		露出上げ 光が画面を支配する 少女、肩に力が入る	露出上が 光が画面を支配する 少女、肩にかが入る
page2 row2 dialogue		少女 M「…眩しい」 85 + 12	少女 M「…眩い」 05 + 12
page2 row4 dialogue		長い髪が揺れる	髪が揺れた
page1 row1 time		□	□

表 3: Base モデルと LoRA 適用モデルの評価結果

評価対象	指標	Base	LoRA
テキスト指標（全サンプル）			
テキスト	サンプル数 n	67	67
テキスト	文字誤り率 Character Error Rate	0.195	0.289
テキスト	完全一致率 Exact Match	0.343	0.313
矢印指標（正解に矢印が含まれるサンプルのみ）			
矢印ありのみ	サンプル数 n	5	5
矢印ありのみ	矢印有無再現率	0.200	0.200
矢印ありのみ	矢印個数完全一致率	0.200	0.200
矢印ありのみ	矢印個数の平均絶対誤差	0.800	0.800
矢印ありのみ	矢印方向の適合率 マイクロ平均	1.000	1.000
矢印ありのみ	矢印方向の再現率 マイクロ平均	0.200	0.200
矢印ありのみ	矢印方向の F1 マイクロ平均	0.333	0.333
矢印ありのみ	矢印列完全一致率	0.200	0.200
矢印ありのみ	矢印列の編集率 マイクロ平均	0.800	0.800
矢印ありのみ	矢印列の類似度平均 $[0, 1]$	0.200	0.200