

## 報告書

### 1 今週の進捗

- 種類ごとの絵コンテの OCR
- 種類ごとの OCR 結果の補正

### 2 研究の目的と全体構成

本研究の目的は、アニメーション制作に用いられる絵コンテ画像から、シーン内容・台詞・演出情報を構造化データとして抽出し、脚本との整合性を保ちながら意味的に理解可能な表現へ変換することである。

本研究では処理を以下の段階 (Stage) に分割して設計している。

- Stage0: 脚本の前処理 (知識化)
- Stage1: 絵コンテ画像からの OCR と CUT 単位への分割
- Stage2: 脚本情報を用いた OCR 結果の整形・補正
- Stage3: 絵コンテの意味理解とシーン説明生成

アニメ制作の用語の一覧についても作成した。以下に例を示す。

```
[
  {
    "id": "cam_pan",
    "word": ["PAN", "→", "←", "パン"],
    "category": "camera_movement",
    "meaning": "カメラを水平方向に移動させるカメラワーク。矢印は移動方向を示す。"
  },
  {
    "id": "cam_qpan",
    "word": ["QPAN", "クイックパン"],
    "category": "camera_movement",
    "meaning": "高速でカメラを水平方向に移動させるカメラワーク。通常の PAN より素早くカメラを動かす。"
  },
  {
    "id": "cam_follow",
    "word": ["Follow", "follow", "F"],
    "category": "camera_movement",
    "meaning": "被写体の動きに合わせてカメラが追従する。"
  },
  :
]
```

## 2.1 Stage0：脚本の知識化

脚本（docx 形式）を入力とし、以下の 2 種類のデータを生成する。

- 登場人物・用語辞書（`lexicon_entities.json`）
- 段落・台詞単位に分割された脚本コーパス（`script_chunks.jsonl`）

脚本中の「登場人物」セクションから正規表記を抽出し、人物名の正解集合として保持する。また、物語本文はシーン番号や台詞・地の文の構造を保持したまま分割し、後段での検索（RAG 的利用）を可能にしている。

## 2.2 Stage1：絵コンテ画像からの OCR

Stage1 では、絵コンテ画像を CUT 単位で処理する。各ページを行（row）に分割し、以下の列画像を入力とする。

『CUT 番号, PICTURE, アクションメモ, ダイアログ, TIM（時間）』

各列画像に対して OCR を行い、行単位の結果を保存する。その後、CUT 番号の推定結果を用いて行を統合し、CUT 単位の JSON データ（`cutXXXX.stage1.json`）を生成する。

### 2.2.1 使用モデル

Stage1 の OCR には視覚言語モデルを用いる。

- 使用モデル：Qwen3-VL-4B-Instruct
- 用途：画像入力に対するテキスト生成（OCR）

一般的な OCR エンジンではなく、視覚言語モデルを用いることで、日本語の手書き・装飾文字・制作現場特有の表記にも対応可能とした。

### 2.2.2 CUT 番号の系列補正

CUT 番号は OCR 誤認識が起こりやすいため、前後関係を用いた系列補正を導入している。

前の CUT 番号を  $c_{t-1}$ 、現在の OCR 推定値を  $\hat{c}_t$  とすると、期待値は

$$c_t^{exp} = c_{t-1} + 1$$

と定義できる。

$\hat{c}_t$  が  $c_t^{exp}$  から大きく乖離している場合、数字誤認（例：0 と 9）を考慮して補正する。これにより CUT 番号の飛びや誤結合を抑制している。

## 2.3 Stage2：脚本情報を用いた整形・補正

Stage2 では、Stage1 で得られた OCR 結果を、脚本由来の知識を用いて整形・補正する。

### 2.3.1 使用モデル

Stage1 と同様に、OCR 結果の補正には視覚言語モデルを用いる。

- 使用モデル：Qwen3-VL-4B-Instruct
- 用途：脚本やアニメ制作用語を用いた OCR 結果の補正

### 2.3.2 OCR 結果の正規化

OCR 結果には,

- 無意味なブレースホルダ (例: □, △)
- ノイズ行

が多く含まれるため, 連続するブレースホルダを 1 つにまとめるなどの正規化をする.

### 2.3.3 脚本との対応付け

絵コンテ中の台詞は脚本中にも存在するという前提のもと, OCR で得られた台詞をクエリとして, 脚本コーパスから一致・類似する台詞を検索する.

本研究では,

- 台詞の完全一致 (正規化後)
- 類似度スコア (しきい値付き)

を用いて候補を抽出している.

### 2.3.4 登場人物名の誤認識補正

OCR により誤認識された登場人物名 (例: 「ベルリ」→「ベリル」) に対し, 脚本由来の人物辞書を用いて補正する.

人物名  $x$  (OCR 結果) と正解候補  $y$  に対し, 編集距離  $d(x, y)$  を計算する. 隣接文字の転置も考慮した距離を用い, 距離が十分小さい場合に補正候補とする.

補正は元の OCR 結果を上書きせず, 正規化結果と補正ログを別フィールドとして保持する.

## 2.4 Stage3: 絵コンテの意味理解とシーン説明生成

ここでは, Stage2 において整形・補正された絵コンテ情報を用いて, 各カットおよび行 (row) に対応する意味的なシーン理解を実装する. 具体的には, 絵コンテ画像と補正済みテキスト, および脚本・制作知識を統合し, 構造化された意味表現と自然言語によるシーン説明を生成することを目的とする.

### 2.4.1 目的

Stage3 の目的は以下の 3 点である.

- 絵コンテ画像とテキスト情報を統合し, カットの内容を意味的に理解すること
- 脚本およびアニメ制作用語との整合性を保った解釈をすること
- 後続処理 (検索, 分析, 要約等) に利用可能な形でシーン情報を出力すること

Stage1 および Stage2 が主に文字情報の抽出・補正を目的としていたのに対し, Stage3 は視覚情報を含む意味理解を扱う点が本質的に異なる.

### 2.4.2 使用モデル

Stage3 でも、引き続き視覚言語モデルを用いる。

- 使用モデル：Qwen3-VL-4B-Instruct
- 用途：絵コンテ画像とテキスト・外部知識を統合した意味理解および文章生成

同一モデルを Stage1～3 で段階的に利用することで、OCR、補正、意味理解という処理の連続性を保つ設計としている。

### 2.4.3 入力

Stage3 への入力は、以下の情報から構成される。

**画像入力** 絵コンテ画像について、以下の 3 種類の入力形式を用いる。

- 行単位の picture 画像のみ
- 行単位で cut, picture, action\_memo, dialogue, time を結合した画像
- カット単位で複数行を結合した画像

これにより、画像構成の違いが意味理解およびシーン説明生成に与える影響を比較可能とする。

**テキスト入力** Stage2 において補正された以下のテキストを用いる。

- action\_memo（演出・動作メモ）
- dialogue（台詞）

行単位の処理では、対応する行の Stage1 OCR 結果を主情報とし、カット全体の補正済みテキスト（Stage2 出力）を補助的な文脈情報として利用する。

**外部知識（RAG）** Stage3 では、全脚本や辞書を一括で入力するのではなく、Stage2 において当該カットに対して抽出された情報のみを外部知識として用いる。具体的には、

- 脚本中の対応候補（script candidates）
- 登場人物辞書（該当カットに関連する人物のみ）
- 検出されたアニメ制作用語とその意味

を入力に含める。これにより、不要な知識による推測や誤解釈を抑制する。

### 2.4.4 出力

Stage3 の出力は、以下の 2 種類からなる。

**構造化 JSON** まず、絵コンテの内容を構造化した JSON 表現を生成する。この JSON には、

- 登場人物
- 行動・動作
- カメラワーク (PAN, FIX 等)
- 音響情報 (SE, BGM 等)
- 脚本との整合性判定
- 使用した脚本断片および制作記号

などが含まれる。この構造化表現は、後続の解析や検証を容易にするための中間表現として位置付けられる。

**シーン説明文** 次に、生成された構造化 JSON を主な根拠として、自然言語による短いシーン説明文を生成する。この際も画像を入力として与え、視覚情報と構造化情報の両方を参照することで、過度な推測を抑えた説明文生成をする。

#### 2.4.5 2 段階推論の採用

Stage3 では、

1. 構造化 JSON の生成
2. シーン説明文の生成

という 2 段階の推論を採用する。

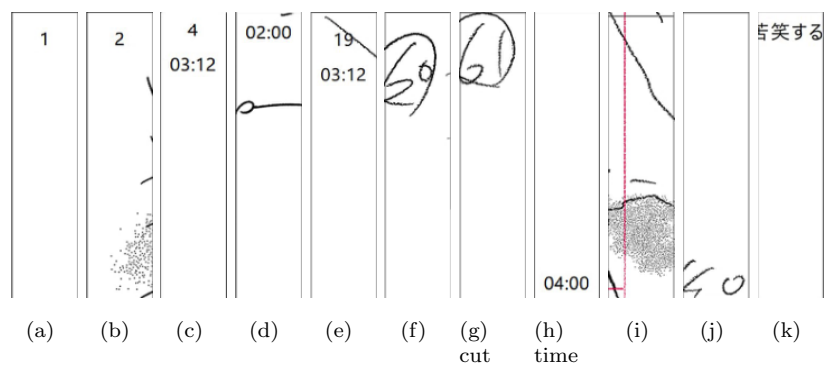
この分離により、

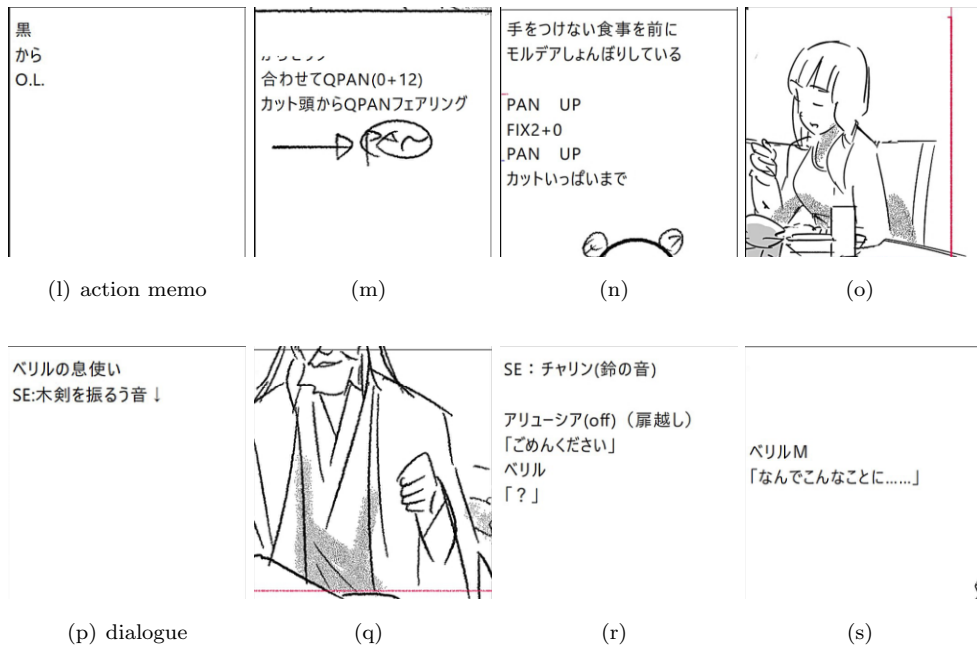
- 意味抽出と文章生成の責務を明確に分けられる
- 出力の一貫性および再現性が向上する
- 構造化情報を用いた分析・評価が可能になる

といった利点が得られる。

## 2.5 データ

### 2.5.1 CUT, TIM, アクションメモ, ダイアログ





## 2.6 結果

### 2.6.1 Stage0 (脚本の前処理)

脚本から登場人物の名前およびセリフやナレーションを取得.

```
{
  "episode_id": "episode01",
  "source_docx": "data/script_phase1/episode01.docx",
  "entities": [
    {
      "canonical": "ベリル・ガーデナント",
      "type": "character",
      "aliases": [],
      "evidence": [
        {
          "paragraph_index": 9,
          "text_snippet": "ベリル・ガーデナント"
        }
      ]
    },
    {
      "canonical": "ファナリー (セリフなし)",
      "type": "character",
      "aliases": [],
      "evidence": [
        {
          "paragraph_index": 22,
          "text_snippet": "ファナリー (セリフなし)"
        }
      ]
    }
  ]
}
```

```

    }
  ]
},
{
  "canonical": "モルデア・ガーデナント",
  "type": "character",
  "aliases": [],
  "evidence": [
    {
      "paragraph_index": 25,
      "text_snippet": "モルデア・ガーデナント"
    }
  ]
},
:

:
{"chunk_id": "c000029", "paragraph_index": 45, "text": "モルデアの視線、さらに鋭くなる。",
"kind": "narration", "scene_id": "①", "speaker": null, "speaker_note": null, "is_heading": false}

{"chunk_id": "c000030", "paragraph_index": 46, "text": "木剣、さらに2回ほどブン！と通過。",
"kind": "narration", "scene_id": "①", "speaker": null, "speaker_note": null, "is_heading": false}

{"chunk_id": "c000031", "paragraph_index": 47, "text": "振っているのは――ベリル。",
"kind": "narration", "scene_id": "①", "speaker": null, "speaker_note": null, "is_heading": false}

{"chunk_id": "c000032", "paragraph_index": 48, "text": "「…（チラッとモルデア見る）」",
"kind": "dialogue", "scene_id": "①", "speaker": "ベリルM", "speaker_note": null, "is_heading": false}

{"chunk_id": "c000033", "paragraph_index": 49, "text": "ベリル、素振り続けている。",
"kind": "narration", "scene_id": "①", "speaker": null, "speaker_note": null, "is_heading": false}
:

char_block =
ファナリー（セリフなし）
ランドリド・パトルロック
アリューシア・シトラス
ヘンブリッツ・ドラウト
モルデア・ガーデナント
クルニ・クルーシエル
フレン・ガーデナント
ベリル・ガーデナント
エヴァンス・ジーン
スレナ・リサンデラ
騎士団員A
騎士団員B

```

騎士団員C  
ジェイン

glossary =

ピン送り(奥から手前), グローフィルター, 残像エコー処理風, カット尻セリフ, フレームアウト, Follow follow, クイックパン, フェアリング, 作画リピート, 残像エコー風, Apart, Bpart, slide, × × ×  
つけ PAN, インサート, サイズ調整, フラッシュ, 話数内流用, 集中ブラー, A.C., A パート, B パート  
DOWN, H.V., QPAN, back, stop, wipe, ↓こぼし, カット尻, カット頭, スライド, ダブラシ, ピンボケ  
ピン送り, フレーム, ポンヨリ, 上→下へ, 回想明け, 湯気FX, FIX, L/O, O.L, OUT, PAN, S.L, T.B  
T.U, W ラシ, cut, m+n, off, wxp, こぼし, アオリ, エコー, カット, スロー, トメ絵, ニヤリ, マルチ  
入射光, 画面動, AC, FX, HV, IN, LO, SE, UP, on, じわ, ぼけ, ガヤ, トメ, ナメ, パン, ピン  
ボケ, ヨリ, 作画, 全景, 兼用, 回想, 引き, 欠番, 流用, 足し, F, M, S, ←, →, ↓, 仮, <, M

## 2.6.2 Stage1 (絵コンテ画像からの OCR) のプロンプト

○ Base Rules

Rules:

- Keep original line breaks.
- Preserve symbols exactly (e.g., arrows → ← ↑ ↓, brackets, punctuation).
- Do NOT correct typos.
- Do NOT translate.
- If unreadable, output '□'.
- Return ONLY the transcribed text. No explanations.

Notes:

- This storyboard contains BOTH handwritten text and computer-typed text.
- Handwritten text may include arrows and symbols.
- Arrows are important. Always transcribe arrows if present.

=== STORYBOARD TERMS (terms only) ===

:  
:

日本語訳 (Google 翻訳):

ルール:

- 元の改行を保持してください。
- 記号を正確に保持してください (例: 矢印 → ← ↑ ↓、括弧、句読点)。
- 誤字を直さないでください。
- 翻訳しないでください。
- 読めない場合は「□」と出力してください。
- 転写されたテキストのみを返してください。説明は不要です。

注意:

- この絵コンテには手書きのテキストとコンピュータで入力されたテキストの両方が含まれています。
- 手書きのテキストには矢印や記号が含まれる場合があります。
- 矢印は重要です。存在する場合は必ず転写してください。



## ○ cut

日本語訳 (Google 翻訳):

あなたはアニメの絵コンテ用の OCR エンジンです。

対象列: CUT。

タスク: CUT 番号があれば読み取ること。

mm:ss のような時間や「続」のような継続マークが含まれている場合もあります。

読めるものがなければ「□」と出力してください。

```
f"{BASE_RULES}n"
```

```
f"{HANDWRITING_NOTE}"
```

CUT 列の追加制約:

- 数字や短いマーカーを優先。
- カット番号と時間が両方ある場合は、見える通り別々の行に保持。
- 欠けている数字を推測してはいけません

## ○ picture

日本語訳 (Google 翻訳):

あなたはアニメの絵コンテ用 OCR エンジンです。

対象の列: PICTURE (主に絵)

タスク: 画像内にあるテキストを全て書き起こしてください。

ほとんどの領域は絵ですが、手書きの用語 (PAN、矢印など) が存在することがあります。

テキストがない場合は『□』と出力してください。

```
f"{BASE_RULES}"
```

```
f"{HANDWRITING_NOTE}"
```

PICTURE 列の追加制約:

- 見えるテキストや記号のみを書き起こしてください; 絵は無視します。
- 手書きの矢印や短い用語はよく出てきます; そのまま正確に残してください。
- 矢印が用語に付いている場合 (例: → PAN)、見える通りに一緒に残してください。

## ○ action memo

日本語訳 (Google 翻訳):

あなたはアニメ絵コンテ用の OCR エンジンです。

対象の列: ACTION MEMO。

タスク: アクション、カメラ、タイミング、エフェクトに関するすべての手書き/タイプされたメモを文字起こしすること。

この列には多くの制作用語 (PAN/FIX/ヨリ/トメ/etc.) が含まれることがあります。

```
f"{BASE_RULES}"
```

```
f"{HANDWRITING_NOTE}"
```

ACTION MEMO 列の追加制約：

- 制作用語は正確に保持すること。
- 括弧や矢印は正確に保持すること。
- 正規化や書き直しは行わないこと。

○ dialogue

日本語訳 (Google 翻訳)：

あなたはアニメ絵コンテ用の OCR エンジンです。

対象列：DIALOGUE (セリフ)

タスク：セリフと関連する音声メモを転写すること。

この列には、話者名、引用符「」、独白マーカ (M/M)、音声メモ (SE/MA/BGM/ギャ/ off/on) が含まれる場合があります。

```
f"{BASE_RULES}"
```

```
f"{HANDWRITING_NOTE}"
```

DIALOGUE 列に対する追加制約：

- 話者名や記号 (「」、()) をそのまま保持すること。
- SE:, MA:, BGM:, ギャ, off/on のような音声タグを正確に保持すること。
- 句読点を勝手に創作しないこと。

○ time

日本語訳 (Google 翻訳)：

あなたはアニメの絵コンテ用の OCR エンジンです。

対象列：TIME (時間)

タスク：時間情報があれば書き写してください。

一般的な形式：mm:ss、(m+n)、フレーム数、または短い数字のタイミングメモ。

読み取れるタイミングがない場合は『□』を出力してください。

```
f"{BASE_RULES}"
```

```
f"{HANDWRITING_NOTE}"
```

TIME 列に関する追加制約：

- 数字や時間表現を優先してください。
- 書かれている通りに正確に保持してください。

### 2.6.3 絵コンテ画像からの OCR と脚本情報を用いた OCR 結果の整形・補正の結果

前回: Qwen3-VL-4B-Instruct

今回: Qwen3-VL-8B-Instruct

#### 2.6.4 以前の GPT 系の結果

比較項目	gpt-4o-mini gpt-4.1-mini gpt-4.1-nano	gpt-4o	gpt-4.1
page002_row5	Se: 木剣を振る音↓	SE: 木剣を振るう音 ☒	SE: 木剣を振るう音↓
page007_row2	ハンセン@合わせて QPAN(0+12)@ カット頭から QPAN フェアリング@ →@PAC	・・・ンジ@合わせて QPAN(0+12) @ カット頭から QPAN フェアリング	パンニング@合わせて QPAN(0+12)@ カット頭から QPAN フェアリング
page053_row3	ベリル、地図を見て	ベルリ、地図を見て	ベリル、地図を見て
page053_row5	ベリル立ち止まって アリュージアへ向き@ follow stop	ベルル立ち止まって アリュージアへ@向き@ follow stop	ベル立ち止まって アリュージアへ@向き@ follow stop
page060_row4	木剣を差し出す。	木剣を差し出す。	木判を差し出す。
page061_row1	ナビカ PANUP!@ ピ送り（土）→（木）へ@ じっと木鉢を見る@スレ 64	トウカ PANUP!@ ピン送り—@ じっと木鉢を見つめる@ スレ 64	↑ ICH PANUP!@ ピ送り 上→下へ@ じっと木釘を見つめる スレイ 64

#### 2.6.5 Stage3 (絵コンテの意味理解) のプロンプト

You analyze an anime storyboard and output JSON only.

You are given an input image and text/context.

Rules:

- Do NOT invent details not supported by the image or provided text.
- If unknown, use null or empty list.
- Prefer ROW text as the primary evidence for row-based patterns.
- Use CUT-level text only as background context.

=== META ===

cut={cut\_num}

pattern={pattern\_id}

unit={json.dumps(unit\_obj, ensure\_ascii=False)}

=== ROW OCR (primary) ===

[action\_memo\_row]

{row\_action}

[dialogue\_row]

{row\_dialogue}

=== CUT OCR (secondary, Stage2) ===

[action\_memo\_cut]

```

{cut_action_norm}

[dialogue_cut]
{cut_dialogue_norm}

=== SCRIPT (retrieved) ===
{script_block}

=== TERMS (meaning attached) ===
{terms_block}

=== OUTPUT JSON SCHEMA ===
{
  "cut": <int>,
  "pattern": "<pattern_id>",
  "unit": { ... },
  "scene": {
    "setting": <string|null>,
    "characters": <array of strings>,
    "actions": <array of strings>,
    "camera": {"movement": <array of strings>, "framing": <string|null>, "focus": <string|null>},
    "sound": <array of {"type": "SE|BGM|GAYA|other", "text": string}>,
    "dialogue": <array of {"speaker": string|null, "text": string}>
  },
  "consistency": {"script_match": "match|partial|unknown|mismatch", "notes": <array of strings>},
  "evidence": {
    "used_script_chunk_ids": <array of strings>,
    "used_terms_with_meaning": <array of {"term": string, "meaning": string}>,
    "inputs": {
      "ocr_action_memo_row": string,
      "ocr_dialogue_row": string,
      "ocr_action_memo_cut": string,
      "ocr_dialogue_cut": string,
      "script_snippets": <array of strings>
    }
  }
}

Set cut={cut_num} and pattern='{pattern_id}'.
Return JSON only.

```

日本語訳 (Google 翻訳):

アニメの絵コンテを分析し、JSON のみを出力します。  
 入力として画像とテキスト/コンテキストが与えられます。

ルール:

- 画像や提供されたテキストで裏付けられていない詳細を作り出さないでください。
- 不明な場合は null または空のリストを使用してください。
- 行ベースのパターンの主要な証拠として ROW テキストを優先してください。
- CUT レベルのテキストは背景コンテキストとしてのみ使用してください。

### 2.6.6 Stage3 (シーン説明生成) のプロンプト

You write a short Japanese scene description for an anime storyboard.

You are given an input image, ROW OCR text, and a structured JSON analysis.

Rules:

- Use the structured JSON as the primary guide.
- Do NOT invent details.
- 2-4 sentences in Japanese.
- Mention camera movement if present.

=== META ===

cut={cut\_num}

pattern={pattern\_id}

unit={json.dumps(unit\_obj, ensure\_ascii=False)}

=== ROW OCR ===

[action\_memo\_row]

{row\_action}

[dialogue\_row]

{row\_dialogue}

=== STRUCTURED JSON ===

{json.dumps(structured\_json, ensure\_ascii=False)}

=== OUTPUT ===

Return ONLY the description text.

日本語訳 (Google 翻訳):

あなたはアニメの絵コンテ用に短い日本語のシーン説明を書きます。入力画像、ROW OCR テキスト、構造化された JSON 解析が与えられます。

ルール:

- 構造化 JSON を主なガイドとして使用します。
- 詳細を捏造しないでください。
- 日本語で 2~4 文にまとめてください。
- カメラの動きがあれば必ず記載してください。

### 2.6.7 Stage3 (絵コンテの意味理解とシーン説明生成) の結果


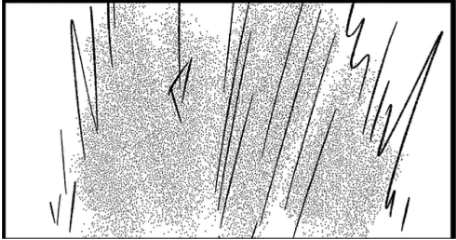
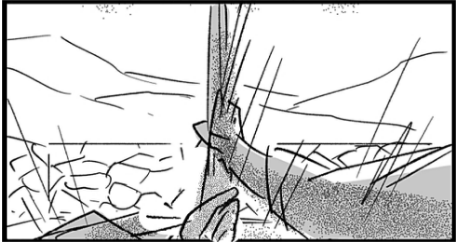
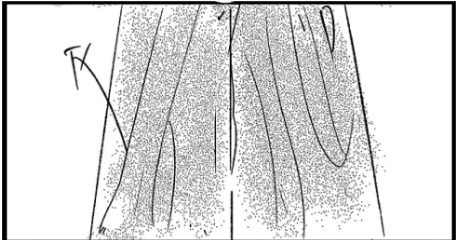
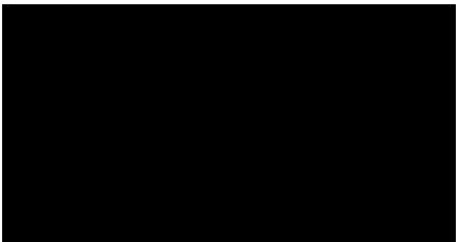
1		黒 から O.L.	ベリルの息使い SE:木剣を振るう音↓	
		木剣のヨリから		
		木剣を振り下ろしトメ 再び振り上げる 作画リピート		
		がカメラに近づく		
04:00		黒	SE:木剣を振るう音↓	04:00

図 1: 黒から O.L. で始まり、木剣のヨリから振り下ろす。カメラはトメで動きを止めて再び振り上げる。作画リピートで、被写体がカメラに近づく。ベリルの息使いが木剣を振る音を出す。

## 3 今後の予定

- OCR およびシーン説明モデルのファインチューニング
- 定量評価指標 (Accuracy, CER) の設計と比較実験
- Trigger データセットでの実装

11		子供たちが道場の入口付近で手を振る	子供たち 「先生おはようございまーす！」	
		ベリル(背)フレームINして手を上げる	ベリル(背) 「おはよう～」	
08:00		手をおろして立ち止まり	ベリル(背) 「道場継いで弟子も増えた、十分親孝行じゃないか」	08:00

図 2: 道場の入口付近で、子供たちは手を振って「先生おはようございます！」と挨拶。ベリルは背中を向けてフレームにインして手を上げ、「おはよう～」と返す。その後、手を下ろして立ち止まり、弟子が増えたことに満足した様子で「道場続いて弟子も増えた、十分親孝行じゃないか」と語る。

## 参考文献

表 1: 前回の OCR 結果と今回の OCR 結果

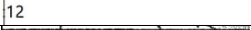

ID	入力画像	前回の OCR 結果	今回の OCR 結果
page4 row5 act	<p>モルディア ナメ ベリルが木剣を振っている 奥門の奥から子供達が見える</p> 	<p>モルデア ナメ ベルルが木剣を振っている 奥門の奥から子供達が見える</p>	<p>モルディア ナメ ベリルが木剣を振っている 奥門の奥から子供達が見える 12</p>
page7 row2 act	<p>カッテッ 合わせてQPAN(0+12) カット頭からQPANフェアリング</p> 	<p>合せて QPAN(0+12) カット頭から QPAN フェアリング</p>	<p>合わせて QPAN(0+12) カット頭から QPAN フェアリング</p>
page9 row3 act	<p>cut兼用(サイズ調整ヨリ) やれやれな表情のベリル</p> <p>セリフ終わりで 歩き出し(ピンonからoff) PANすると モルディア</p>	<p>cut 兼用 (サイズ調整ヨリ) やれやれな表情のベリル セリフ終わりで 歩き出し (ピン on から off) PAN すると モルディア</p>	<p>cut 兼用 (サイズ調整ヨリ) やれやれな表情のベリル セリフ終わりで 歩き出し (ピン on から off) PAN すると モルディア</p>
page13 row1 dia	<p>ベリルM 「若い頃は ひとかどの剣士になるという 望みも抱いたけど」 ↓こぼしあり</p>	<p>ベリル M 「若い頃は ひとかどりの剣士になるという 望みも抱いたけど」 ↓こぼしあり</p>	<p>ベリル M 「若い頃は ひとかどの剣士になるという 望みも抱いたけど」 ↓こぼしあり</p>
page16 row2 dia	<p>ベリルM 「剣の腕はほどほどで収まった俺 だけど、 人に教える才能はあったらしい」</p>	<p>ベリル M 「剣の腕はほどほどで収まった俺 だけど、 人に教える才能はあったらしい」</p>	<p>ベリル M 「剣の腕はほどほどで収まった俺 だけど、 人に教える才能はあったらしい」</p>



表 2: 前回の OCR 結果と今回の OCR 結果

ID	入力画像	前回の OCR 結果	今回の OCR 結果
page30 row1 cut		69 (60 に補正している)	69 (60 に補正している)
page30 row5 cut		<input type="checkbox"/>	<input type="checkbox"/>
page30 row5 act	 follow back ベリル窓外をぼんやり眺めながら	follow back ベリル窓外をぼんやり眺めながら	follow back ベリル窓外をぼんやり眺めながら
page30 row5 dia	ベリルM 「なんでこんなことに……」 	ベリルM 「なんでこんなことに……」	ベリルM 「なんでこんなことに……」
page42 row2 act	画面いっぱいまで迫る 	画面いっぱいまで迫る <input type="checkbox"/>	画面いっぱいまで迫る <input type="checkbox"/>