

## 概要

日本語のバナー広告画像ドメインでのタイポグラフィ属性解析 [1] を生成形式で提案し、合成データを用いたLoRAチューニングによる手法で実用性を検証する。

広告バナー風のテキストや広告の設定、それらに基づく装飾パラメタなどの擬似データをLLMで生成し、擬似データから日本語テキスト画像を合成することで学習データを作成した。合成画像を入力としてフォント種・太さ・揃え・配色・字間/行間などのタイポグラフィ属性を構造化テキストとして生成するタスクを提案した。Qwen2.5-VL [2]、Phi-4-Multimodal [3]、Gemma3 [4] といった7B以下の小規模VLMに対し、合成データでLoRAチューニングを行い、日本語広告特有のタイポグラフィ表現に対する小規模VLMの生成出力の傾向を調査した。

## データ合成

GPT-OSS 120Bを用いて広告デザインの説明と、説明に関連する広告テキストの装飾プロパティをzero-shotで合成する。合成したデータをFabric.jsで画像にレンダリングする。

### 広告デザイン指示

- Objective: 広告の目的
- TargetAudience: 広告ターゲット層
- ConceptualTheme: 広告コンセプトの説明
- ArtDirection: 広告デザインの方向性の指示
- FillColor: 背景色
- Text: 装飾対象の広告テキスト
- TextCategory: 対象テキストの広告内における意味分類 (main\_catchphrase, sub\_catchphrase, cta, noteのいずれか)

### テキスト装飾プロパティ

- FontFamily: SIL OPEN FONT LICENSEのフォント7種 (Noto Sans、Noto Sans JP、Noto Serif JP、Roboto、M PLUS 1p、M PLUS Rounded 1c、Murecho)
- FontSize: 10~88までの整数値 (px)
- FontWeight: 400, 500, 700, 900の4値
- Color: `rgba(255,255,255,0.2)` の形式の文字の色指定
- StrokeWidth: 0以上で指定する文字枠線の太さ (px: 0の時は枠なし)
- StrokeColor: `#FFFFFF` のような16進数形式の色指定
- ShadowOffsetX: 影のX方向のオフセット (px)
- ShadowOffsetY: 影のY方向のオフセット (px)
- ShadowBlur: 0~100までの整数値による影のぼかし指定 (px)
- ShadowColor: `rgba(255,255,255,0.2)` の形式の文字による影の色指定

## 実験

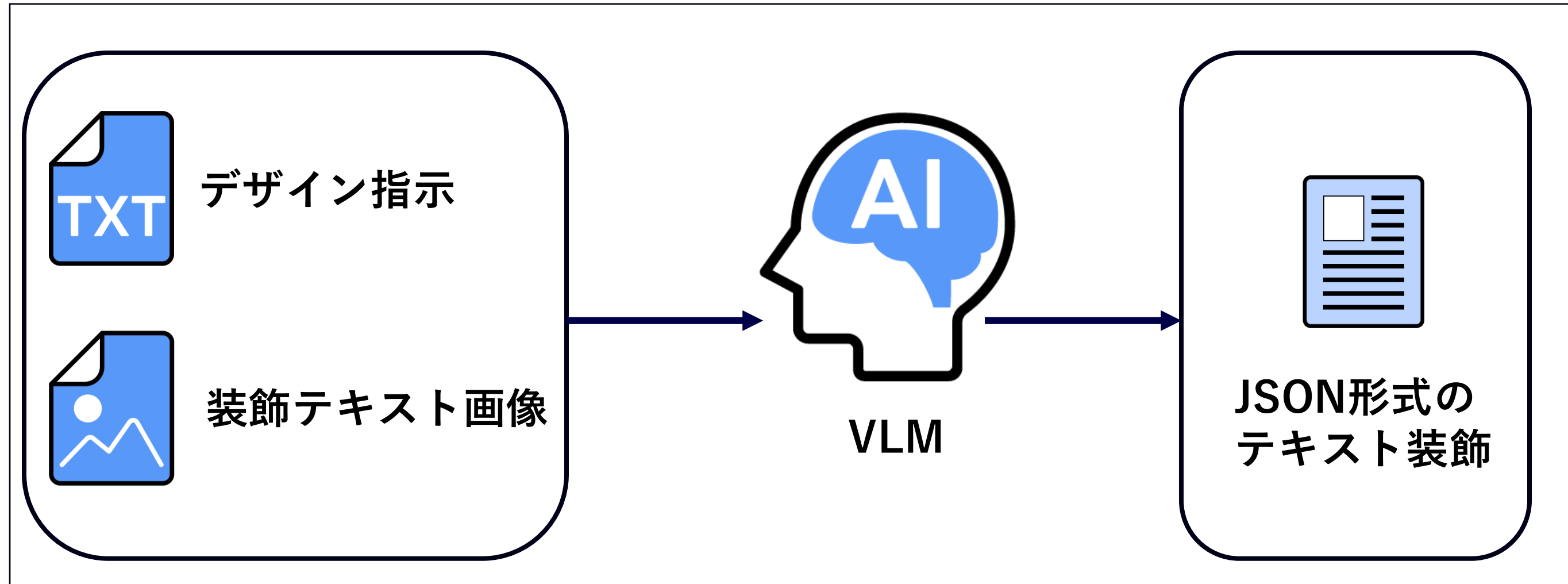
画像・広告デザイン指示を入力し、テキスト装飾プロパティのJSONを生成するように小規模VLMをチューニングする。

### モデル設定

- 訓練手法: Gemma, Qwen2.5-VLはLoRAチューニング。Phi-4-MMはVision LoRAレイヤーを追加訓練する。
- epoch: 1, 2, 5
- LoRA rank: 4, 8, 16
- Lr: 1e-5, 2e-5, 5e-5, 1e-4, 2e-4, 5e-4
- データ分割: 学習データ4,000件、評価データ1,000件。

### 評価指標

- Accuracy: FontFamily
- MAE: 単位がpxの項目
- CIEDE2000 [5]: 正解のrgb形式の色と生成結果の色差 (0~90)



## 実験結果

### 自動評価結果

モデル	Gemma3 (4B)	Qwen2.5-VL (3B)	Phi-4-MM (6B)	GPT-4o zeroshot
FontFamily (Acc) ↑	723	775	808	551
FontSize (MAE) ↓	1.199	0.528	0.396	8.217
FontWeight (MAE) ↓	17.28	13.97	11.59	56.84
Color (CIEDE) ↓	7.848	5.412	5.279	12.93
StrokeWidth (MAE) ↓	0.266	0.149	0.181	0.467
StrokeColor (CIEDE) ↓	14.87	11.84	11.80	28.77
ShadowOffsetX (MAE) ↓	0.053	0.051	0.044	0.424
ShadowOffsetY (MAE) ↓	0.062	0.071	0.048	0.237
ShadowBlur (MAE) ↓	0.641	0.710	0.442	4.152
ShadowColor (CIEDE) ↓	6.310	4.043	4.528	10.15

### サンプル入力

Objective	10月の三連休に向けた温泉旅館の早割プランを告知し、直前予約を促進する。
TargetAudience	30~50代の男女で、国内旅行や温泉、グルメ体験に関心があり、家族や夫婦で週末旅行を計画する層。
ConceptualTheme	秋の紅葉とぬくもりを感じる上質で落ち着いたトーン。温かみと安心感、少しの高級感。
ArtDirection	落ち着いた深い茜色を基調に、僅かに褐色を加えた温かい配色。余白を大きく取り、丸みのあるボタン風レイアウトを想定。文字は高コントラストの白で強めの太字、最小限のシャドウと細い縁取りで視認性を担保。アクセントは使用せず、CTAを画面中央に大きく配置。

### サンプル出力のレンダリング結果

Original	Gemma3 (4B)	Qwen2.5-VL (3B)	Phi-4-MM (6B)
今すぐ予約する	今すぐ予約する	今すぐ予約する	今すぐ予約する

## 参考文献

- [1] Takagi, et al. Analysis and Prediction of Attractive Fonts on Title-overlaid Food Images. MVA2025,
- [2] Bai, et al. Qwen2.5-VL Technical Report. arXiv:2409.12191.
- [3] Abouelenin, et al. Phi-4-Mini Technical Report: Compact yet Powerful Multimodal Language Models via Mixture-of-LoRAs. arXiv:2503.01743.
- [4] Kamath, et al. Gemma 3 Technical Report. arXiv:2503.19786.
- [5] Sharma, et al. The CIEDE2000 Color-Difference Formula: Implementation Notes, Supplementary Test Data, and Mathematical Observations. Color Research and Application, vol. 30. No. 1.