

# 広告画像生成のための自動生成テキストの有効性の検証

中田景子<sup>†</sup> 辻萌々華<sup>†</sup> 大野ひとみ<sup>†</sup> 新田直子<sup>†</sup>

<sup>†</sup> 武庫川女子大学 生活環境学部 情報メディア学科

## 1 はじめに

近年、テキストを用いた画像生成技術が飛躍的な進歩を遂げている。本研究では、特に広告画像の効率的な生成を目的とし、画像のキャプション生成技術に着目し、専門家が制作した広告画像から生成されるテキストを用いて、元の広告画像と同等の画像が生成可能か検証する。

## 2 検証方法

専門家が制作した広告画像と、テキストを用いて生成された広告画像に対し、以下の処理により、意味内容に基づく類似性を検証する。

- 1) 専門家が制作した広告画像に対し、画像からのキャプション生成手法である BLIP [2] を用いて画像内容を表すテキストを生成する。
- 2) 1) で生成したテキストを入力とし、テキストからの画像生成手法である Stable Diffusion [3] を用いて画像を生成する。
- 3) 3) で生成した画像に対し 1) と同様に BLIP を用いて画像内容を表すテキストを生成する。
- 4) 1) 及び 3) で生成したテキストをそれぞれ BERT [4] を用いた特徴量に変換し、コサイン類似度を算出する。

## 3 検証結果

異なるトピックに対する専門家が制作した広告画像を収集したデータセット [1] から9つのトピックを選択した。各トピックに対し、50枚の広告画像をランダムに選択し、BLIPにより生成したテキストを用いて50枚の画像を生成した。対となる広告画像と生成画像から生成されたテキストのトピックごとの類似度の分布を、図1の各トピックの左側の箱ひげ図として示す。全体的にテキストの類似度は高く、専門家が制作した広告画像と類似した内容の画像が生成されたと考えられる。しかし、図2に示すように、生成画像には人間が見ると広告らしくないものも多く見られた。そこで、広告であることを明示するため、'an advertisement of' を生成されたテキストの前に追加した上で画像を生成した。これにより、図2のように、画像中に文字が多く現れるといった違いが見られ、より広告らしい画像が生成された。しかし、図1の各トピックの右側の箱ひげ図に示すように、対となる生成画像と広告画像との類似度は、広告画像から生成したテキスト

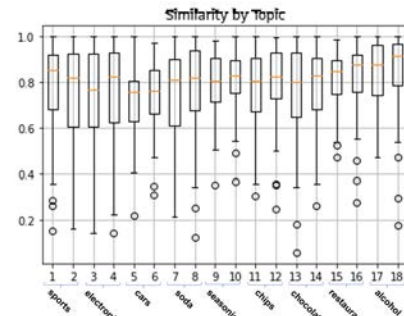


図1 キャプションの類似度分布



図2 広告画像と生成画像

のみを用いた場合と比べ、あまり変化は見られなかった。以上の結果より、テキストを少し変化させるだけでも生成される画像は大きく異なり得るが、専門家が制作したものと同等の画像を生成するには、広告画像から現状のキャプション生成技術により生成されるテキストよりも、より詳細な内容が必要と考えられる。

## 4 むすび

本稿は、広告画像を対象に、専門家が制作した画像から自動生成できるテキストにより、意味内容の類似した画像は生成できるものの、広告としての質は十分でないことを示した。今後は、より質の高い画像を生成するためのテキストの生成方法及び、より精度のよい、画像の類似性の評価手法について検討したい。

### 謝辞

本研究の一部は、JST CREST JPMJCR20D3, 科学研究費補助金基盤 (C) 22K12074 の助成による。

### 参考文献

- [1] Z. Hussain, et al., "Automatic Understanding of Image and Video Advertisements," CVPR, 2017.
- [2] J. Li, et al., "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation," ICML, 2022. AAAI, Vol.33, pp.726-733, 2019.
- [3] R. Rombach, et al., "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR, 2022.
- [4] N. Reimers, et al., "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," EMNLP, 2019.