

12 Text-to-Video モデルを用いて生成された動画に対する定量的評価指標の検討とその評価

Study and Evaluation of Quantitative Evaluation Metrics for Videos Generated by Text-to-Video Model.

本田涼大

指導教員 村上力

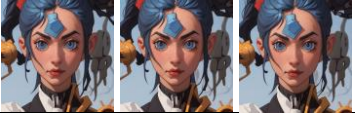


サムネイル			
プロンプト	In 3d game style, Salvador Dali with a robotic half face	unicorn sliding on a rainbow	Two white swans gracefully swam in the serene lake
キャプション	a painting of a woman wearing a clown mask	a horse is riding through the snow on a blue and white horse	two white swans swimming in a body of water
提案手法	0.3476	0.3817	0.8017
CLIPScore	0.2336	0.2314	0.2234
人間の評価	0.3333	0.5119	0.4048

表 1：提案手法による評価（抜粋）

1 はじめに

昨今、動画像生成の分野は大きな発展を遂げている。しかし、動画像生成モデルの評価手法は極端に少ない。本稿では、プロンプトと生成物の関係性を見る指標として、動画キャプションモデルを用いた評価手法を、既存手法である CLIPScore¹ と比較する。

2 提案手法

Text-to-Video モデルはプロンプトと呼ばれる指示文に従い動画を生成する。動画キャプションモデルは動画を元に説明文を作成する。此方には幾つか評価指標があり、高精度でキャプションが作成できることがわかっている。提案手法では動画に対し生成したキャプションがプロンプトと意味的にどれだけ近いのかを評価することでプロンプトへの動画の追従度を評価する。パラメータ θ を持つ文書埋込みモデルを $\text{emb}_{\theta}(\cdot)$ 、プロンプトの埋込みを p 、キャプションの埋込みを c とすると、プロンプトの埋込みは $\text{emb}_{\theta}(p)$ で表される。生成された動画のプロンプトへの追従度を以下のように定める。

$$\cos(p, c; \theta) = \frac{\text{emb}_{\theta}(p) \cdot \text{emb}_{\theta}(c)}{|\text{emb}_{\theta}(p)| |\text{emb}_{\theta}(c)|}$$

3 実験

動画キャプションモデルを用いて、動画に対してキャプション生成を行い、言語モデルを用いてキャプ

ションとプロンプトの埋込みを比較する。比較のため CLIPScore による評価も行う。提案手法の評価のため Animate-diff² を用いて生成した動画像とそのプロンプトに対して人間による評価を行った。31 組の動画像ペアに対し被験者 21 人に「動画像がプロンプトに従っているか 5 段階評価」「動画の説明」「品質の 5 段階評価」を聞いた。

4 結果と考察

結果を表 1 に示す。当初の予定ではアンケート結果を Ground-truth として提案手法の妥当性を評価する予定であったが、結果はほぼ無相関であった。高い相関が得られなかった理由として、プロンプトに稀に含まれる画風などの情報がキャプションには出力されず、スコアを下けていることが考えられる。

参考文献

- 1) Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, Yejin Choi. CLIPScore: A Reference-free Evaluation Metric for Image Captioning. arXiv preprint arXiv:2104.08718v3, 2022.
- 2) Yuwei Guo, Ceyuan Yang, Anyi Rao, Yaohui Wang, Yu Qiao, Dahua Lin, and Bo Dai. Animatediff: Animate your personalized text-to-image diffusion models without specific tuning. arXiv preprint arXiv:2307.04725, 2023.