

生成モデルによるドライビングシミュレータでの交通ルールの自動評価

Automatic Evaluation of Traffic Rules in Driving Simulators by Generative Models

山本 成騎[†] 今井優人[†] 梶原 祐輔[†]
 Yamamoto Seiki Imai Yuto Kajiwara Yusuke

1. はじめに

日本では、高齢ドライバーの事故が社会問題化している。警察庁によると、2011～21年の10年間で全体の死亡事故件数は4691件から2636件まで減少したが、75歳以上の高齢ドライバーによる死亡事故の割合は49.2%から57.7%に増加している[1]。これを受けて、令和2年改正道路交通法により、高齢者の運転免許証の更新等の手続において新たに運転技能検査が導入されることになった。75歳以上で一定の違反歴がある人は、運転技能検査に合格しなければ、運転免許証の更新を受けることができない。

運転技能検査の課題は「指示速度による走行」、「一時停止」、「右折・左折」、「信号通過」、「段差乗り上げ」等がある。これらの課題により、高齢ドライバーの運転技能を定量的に測定することが可能となる。しかし、運転技能検査は3年間隔で実施されるため、その間の運転技能は測定することはできない。そのため、この間に認知機能や反応速度、注意機構が低下した場合に起こり得る事故を防ぐことはできない。事故を未然に防ぐためには、簡便に運転技能を測定する方法を考える必要がある。

具体的な測定方法としてドライビングシミュレータ (DS) による運転技能の評価が挙げられるが、DSは運転コースの数に限りがあるため、慣れが生じてしまい正確な運転技能を測ることができない。また、DSは高価であるため、一般家庭に設置することはできない。

運転コースに限りがあるという問題は敵対的生成ネットワーク (Generative Adversarial Network (GAN)) [2]によって解決できると考えられる。2021年にKimらは自動運転の実証実験が可能でニューラルシミュレータであるDriveGANを開発した[3]。GANを用いることで異なるコースを自動生成できる。しかし、運転技能検査において評価できるようなDSを生成できるかどうかはまだ評価されていない。そこで、FIT2022のCI-005で発表した吉川は運転技能検査の課題に準じた3つの項目「鮮やかさ」、「形状」、「論理的整合性」と4つの評価対象「道路標識と規制表示」、「一時停止線」、「右折・左折のための道路」、「信号機」をもとに定性評価を行った[4]。その結果、「鮮やかさ」に関しては、オブジェクトごとに精度にばらつきがあり、右折・左折の精度が最も高かった。「形状」に関しては、道路と規制表示の精度は低かったが、それ以外のオブジェクトの精度は高かった。「論理的整合性」に関しては、どのオブジェクトも大部分で精度が高くなることがなかった。一方で、目視確認でDSの評価を行ったため、定性的でかつ確率的に生成されるDSを網羅的に評価できなかった。

そこで我々は、物体検出を用いて自動生成されたDSを定量評価可能なルールベース評価指標を考察し、提案した[5]。

残る問題はGANで生成したコースは確率的に生成されているため、出発地点に戻ることができないことや、任意のシーンが生成できない等の問題である。そこで本研究ではStyleGAN[6]の代わりに、テキストから画像を生成するDiffusion model[7]をDSに組み込み、任意の車載画像を生成する。

2. 関連研究

GANやVAEといった生成モデルは、画像生成や音声生成の分野で成功を収めてきたが、いくつかの制約や課題があった。近年においてDiffusion modelが注目され、使用されている背景には以下の要素がある。

1. 生成品質の向上

GANやVAEは高品質な生成を実現することができるが、モード崩壊やサンプルの多様性の欠如といった問題も存在する。一方、Diffusion modelは安定した生成品質を実現することが可能である。

2. 学習の安定性と効率性

GANは学習が困難であり、適切なハイパーパラメータの設定やトレーニングの安定性を確保するための工夫が必要である。一方、Diffusion modelは、データを生成するプロセス自体をモデル化し、単純な確率的反復プロセスによって学習を行うため、学習の安定性と効率性が向上する。

3. モデルの解釈可能性

GANやVAEはブラックボックスモデルとして知られており、内部の生成プロセスを理解することが難しい場合がある。一方、Diffusion modelは潜在変数の事後確率分布を解析的に求めることができるため、生成プロセスの解釈が容易になる。これにより、生成結果をより直感的に制御することができる。

Diffusion modelの活用例として、人間のリアルな声に近い発音を再現する[8]、コンピュータアニメーションにおける人間の動きを再現する[9]等が挙げられる。未知の動画を生成するという観点では、テキストによる条件付き動画生成を行うVideo diffusion model[10]が登場している。本研究ではDiffusion modelで既存の車載映像からヒヤリハットのシーンの生成を試みる。また生成した画像に対してimage to textを行う。生成されたテキストプロンプトをもとに、交通規則ルールの自動評価を試みる。

3. 検証

公立小松大学周辺の車載映像をDiffusion modelによって学習し、学習したものにテキストプロンプトを加えて任意のシーンを生成する。作業工程の流れは以下の通りである。

1. データ収集と前処理

交通事故のシーンや他の関連するシーンを含む多様な車載映像データを収集する。データセットは、事故の有無や異なる運転シナリオをカバーするように多様性を持たせる必要がある。

[†] 公立小松大学 サステイナブルシステム科学研究科,
 Komatsu University, Graduate School of Sustainable Systems
 Science

2. Diffusion model のトレーニング

収集した車載映像データおよびプロンプトを使用して、Diffusion model をトレーニングする。生成したいシーンの説明やテキストプロンプトを収集する。これには、交通事故のシーンや車両の位置、天候条件等のテキストプロンプトが含まれる。

3. シーン生成

学習済みの Diffusion model とテキストプロンプトを使用して、任意のシーンを生成する。テキストプロンプトをモデルに入力し、モデルが指示に基づいて画像を生成するようにする。生成された画像は指定されたテキストに応じて交通事故のシーンや他の要素を含むと予想される。

現状としては、公立小松大学周辺において任意のシーン画像とテキストプロンプトを生成することに成功した (図 1)。



図 1 高齢者と車が正面衝突するシーン (上：元画像, 下：生成画像)

4. 評価方法

これまでの研究[5]で使用していた yolov7 は指定した物体しか検出できない、データセットの収集やラベリングに時間がかかる等の課題があった。そのため、今回は Stable diffusion を用いて交通規則の評価を行う。以下に評価の手順を示す。

1. Stable diffusion で交通規則に基づいた画像をテキスト化する。
2. ファインチューニングを行い、任意のヒヤリハット画像を生成する。
3. 生成した画像をテキスト化する。
4. 元画像と生成画像のテキストを比較する。生成画像にヒヤリハットを示すようなテキストが表示されていれば、論理的整合性が取れているとする。

5. おわりに

Diffusion model を用いて任意のシーンを生成し、生成画像の論理的整合性をテキストプロンプトにより自動評価する方法を提案した。今後の研究においては、Diffusion model を利用して車載映像の生成を行い、より高品質で自由度のあ

る運転シーンを実現することが重要である。また、実際の運転状況や交通環境に近い運転シーンの生成を実現していく。

参考文献

- [1] 警察庁交通局令和 3 年における交通事故の発生状況等について (<https://www.npa.go.jp/bureau/traffic/bunseki/nenkan/040303R03nenkan.pdf> 閲覧日：2022 年 6 月 21 日)
- [2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [3] Kim, S. W., Phillion, J., Torralba, A., & Fidler, S. (2021). Drivegan: Towards a controllable high-quality neural simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5820-5829).
- [4] 吉川 晃太, 梶原 祐輔 (2022) 敵対的生成ネットワークによるドライビングシミュレータ生成時の評価指標に関する一考察
- [5] 山本 成騎, 梶原 祐輔 (2022) 物体検出を用いた自動生成されたドライビングシミュレータのルールベース評価
- [6] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN. *International Conference on Learning Representations*.
- [7] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, 33.
- [8] Zhang, C., Zhang, C., Zhang, M., & Kweon, I. S. (2021). Text-to-image Diffusion models in Generative AI: A Survey. *arXiv preprint arXiv:2107.01366*.
- [9] Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., & Bermano, A. H. (2021). Human Motion Diffusion model.
- [10] Ho, J., Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2021). Video Diffusion models.