先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	ソフトウェア設計学 (飯田　元 (教授))		
学籍番号	1911134	提出日	令和 3年 1月 25日
学生氏名	千田　将也		
論文題目	Towards improving the transferability of adversarial attacks against regression models ステアリング角度予測モデルに対する敵対的攻撃の転用可能性の向上		

要旨

近年，自動運転車のコア技術として機械学習が注目されている．例えば，カメラ画像を用いた機械学習モデルによる自律的ステアリング制御の研究開発が行われている．一方で，機械学習では敵対的攻撃に対する脆弱性が報告されており，さらに，自動車は今後外部との通信が行われるコネクテッドカーが一般的になることが予想され，セキュリティ上の脅威が課題となっている．機械学習に対する敵対的攻撃方法の研究として，敵対的生成ネットワークを応用して敵対的サンプルを生成する手法が提案されている．ここで，あるアーキテクチャの予測モデルを仮定して生成された敵対的サンプルが，異なるアーキテクチャの予測モデルに対しても有効である敵対的攻撃の性質を転用可能性という．先行研究において，回帰モデルでは転用可能性が低いことが報告されている一方で，分類モデルでは転用可能性が十分に高いことが示されている．本研究では，回帰モデルにおける敵対的攻撃の転用可能性を向上させることを目的に，敵対的生成ネットワークを用いた複数の回帰モデルに対する敵対的サンプルを生成する手法を提案する．