

GANを用いた反事実画像説明の生成

Generating Counterfactual Visual Explanation Using GAN

研究背景

- 機械学習モデルの予測、認識の判断根拠の研究は非常に重要
- 画像分類問題の反事実説明とは特定の属性がないことによって他のクラスに属していない理由をユーザに知らせること

研究目的

- 二つの画像間の自然な反事実画像説明を生成
- 人工画像分類に対して反事実画像説明の有用性を検討

提案手法

- クラス説明文書から特徴取得
- 特徴を表す文の置換、生成モデル画像生成
 - 画像分類モデルから二つのクラスの確率を取得
 - 確率によって反事実画像scoreを計算
 - Scoreが一番高い画像の特徴は最終的な反事実特徴

結果と考察

- クラウドソーシングの50人で下記のタスクを実行
 - ① 鳥の基本定義と真実画像を参考にした画像分類タスク
 - ② 反事実画像説明を参考に加えた画像分類タスク
- ①と②のタスク正解率で反事実画像説明の人による画像分類への有効性を検討する