

物体検出の画質改善ネットワークにおける RISE を用いた可視化と解析

荻野春紀・今村幸祐（金沢大学）

1. はじめに

近年、ニューラルネットワークを用いた AI の目覚ましい発展の一方で、ニューラルネットワークモデル自身の説明能力が低いという問題があり、判断根拠の可視化が求められている。一方、画像認識などでは学習やテストには鮮明な画像が用いられるが、実世界ではノイズや明暗等の劣化した画像の方が多く、ネットワークモデルに入力する前に画質を向上させる前処理が求められている。本研究では、ニューラルネットワークを用いた画質改善ネットワークによって物体検出モデルの判断根拠がどのように変化したかを可視化し、解析することを目的とする。

2. RISE による可視化

RISE[1][2]はネットワークモデルの内部構造によらないブラックボックス型の判断根拠可視化手法である。RISE では可視化したい画像と、画素値が 0 と 1 にサンプリングされたマスク画像を物体検出モデルの入力とする。マスク済み画像をネットワークモデルに入力し、その出力候補の中で元の画像に対するオブジェクトの検出結果に最も類似している出力値とマスク画像を掛け合わせる。これを任意のマスク枚数分行い、掛け合わせたマスク画像の総和をとったものが可視化ヒートマップ画像となる。

3. 判断根拠の比較における問題点

本研究では、前処理なしのモデルと複数の前処理ネットワークモデル間で解析を行う。比較の際の問題点として、あるネットワークモデルでは検出できたオブジェクトが別のネットワークモデルでは検出できない場合、その判断根拠が可視化できない。このような両モデル間で判断根拠の変化が大きいと予測されるオブジェクトの可視化を行うために、検出できたモデルでのオブジェクトの検出結果を未検出モデル側でも用いることとする。これにより、未検出のオブジェクトでも両方のモデル間で比較を行うことができるようになる。

4. シミュレーション実験

本研究では、画像の劣化としてガウスノイズとガンマ補正を用いて実験を行う。物体検出器に YOLOv3 を、前処理ネットワークに SR-CNN (3 層) と AR-CNN (4 層) 用いてモデルを学習する。

ガウスノイズにおける可視化の結果を図 1 に示す。AR-CNN・SR-CNN のモデルでは中央の飛行機は尾翼を含めた全体に注目しているのに対し、前処理なしではエンジン部分のみに注目している。ノイズにより尾翼あたりが検出できなかったと考えられる。SR-CNN・AR-CNN のモデルでは右端のオブジェクトが検出できているのに対し、前処理なしのモデルでは右端の飛行機の可視化が行えていないため、右端の飛行機はモデルの出力候補にすらなかったと考えられる。

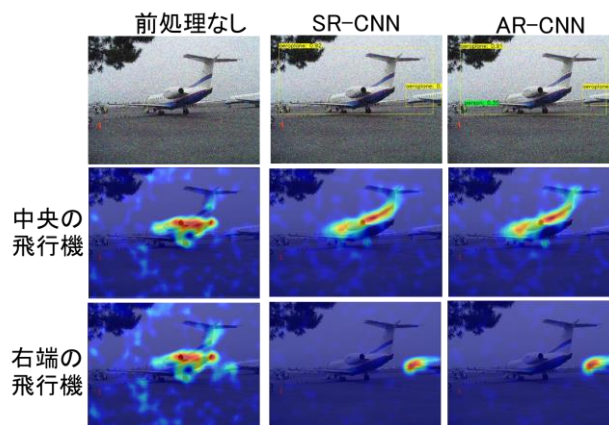


図 1 ガウスノイズでの可視化の結果

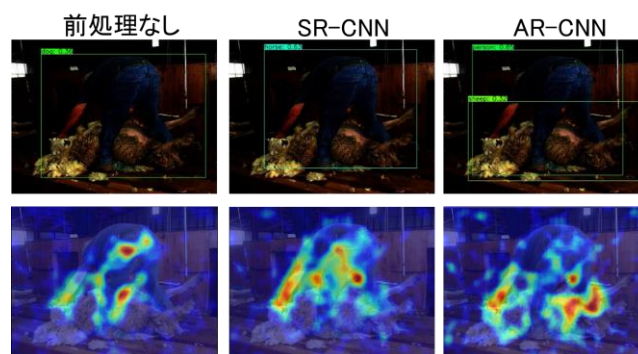


図 2 ガンマ補正での可視化の結果

図 2 は羊の検出結果に対しての可視化を行った結果である。AR-CNN のモデルでは羊の検出結果に対して、羊の顔や足の部分に注目しているのに対し、前処理なしのモデルと SR-CNN のモデルでは人の方に注目箇所がある。この結果から、前処理なしと SR-CNN のモデルでは 2 つのオブジェクトが 1 つのオブジェクトと認識されることが分かる。

5. まとめ

劣化により背景とオブジェクトやオブジェクト同士の境界が曖昧になり注目すべき箇所を注目していなかったものが、前処理によって画質が向上し境界がはっきりしたことで検出精度が上がったことが可視化によってわかった。今後の課題としてより複雑な前処理ネットワークに対して比較・解析を行う。

参考文献

- [1]V. Petsiuk, et al., “RISE: Randomized Input Sampling for Explanation of Black-box Models”, *BMVC 2018*
- [2]V. Petsiuk, et al., “Black-box Explanation of Object Detectors via Saliency Maps”, *CVPR 2020*