

pix2pix によるドライブレコーダーの画質改善手法

2017SC055 小畑憲吾

2017SC065 澤井秀斗

指導教員：河野浩之

1 はじめに

昨今の自動車社会において、あおり運転が問題視されていることや、運転支援システムが発達していくなかで、運転中の映像がより重要となることから、ドライブレコーダーの需要が増加している [7].

しかし、ドライブレコーダーの質や悪天候下などでは映像が不鮮明となり、実際の事故時等の状況確認が困難になるということがある.

これらの課題に対して、ドライブレコーダー画像の画質改善としてまだ使用されていない pix2pix の有効性を示すため、実際に使用しているドライブレコーダーを用い学習データを集め、pix2pix によって画質改善を行い、それを評価した.

2 先行研究

下記、表 1 にそれぞれの先行研究とその概要、表 2 に使用技術例を示す.

表 1 先行研究

著者	概要
石井ら [1]	画像解析で運転者の目の動きを観測
岡田ら [2]	道路環境の高精度認識アルゴリズム
戸部田ら [3]	動画に対する物体検出精度改善
石ら [4]	ビジョンベースの自動運転システム
柴田ら [5]	学習型超解像を用いた画像推定
原ら [6]	ドラレコ映像を用いた頭部検出

表 2 使用技術例

技術	概要
超解像技術	低解像画像をより高解像かつ鮮明な画像に復元
画質改善	圧縮された画像を入力とし非圧縮の画像を出力
ニューラルネットワークフレームワーク	ドライバーの注意視点を予測

3 提案手法

先行研究では、自動車に関連する認識技術の向上と運転支援システムへの応用に向け様々な手法によって研究されている. しかし、それぞれの手法で未だ改善の余地がある. 本研究では、これらと異なった手法によって課題に取り組

む. ここで、GAN(敵対的生成ネットワーク) に注目した.

近年、機械学習において GAN が使用されている技術は増加している. 2014 年に発表された GAN は、そこから派生研究され、CGAN, DCGAN や CoGAN などが開発されていった. 元はデータを増やす、質を上げる仕組みとして開発されたものが画像の分野に応用され始めた. GAN は入力されたデータや画像から新しいデータを生成することが可能であるため、画像の高画質化に利用できる.

画質改善において、GAN を使用するためのライブラリとして pix2pix を選択した. また、pix2pix が開発された 2016 年から現在に至るまで、ドライブレコーダーで録画された画像を pix2pix によって画質改善する研究は行われていない.

4 pix2pix の実行

4.1 節では Pix2Pix による画像変換、4.2 節では実行、4.3 節ではモデルのトレーニングについて示す.

4.1 Pix2Pix による画像変換

Pix2Pix は GAN を用いた画像生成アルゴリズムで、2 枚の対になった画像から相互間の関係を学習し、1 枚の元画像に対して学習結果に基づいて画像を生成することができる. 図 1 に pix2pix で実行される画像変換の例を示す. 図 1 の input がコンピュータに与えた学習用の画像、output がコンピュータが自動生成した画像、target がテスト用実画像となる. Pix2Pix は、「条件画像と実画像のペア」を学習データとしてその対応関係を学習する. 条件ベクトルの代わりに条件画像を使用し、画像から画像への変換問題を扱う.

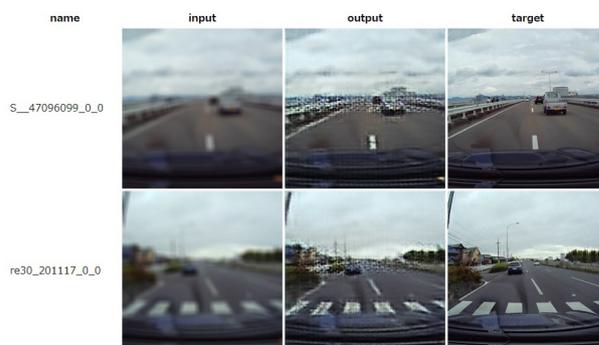


図 1 Pix2Pix で実行される画像変換の例

4.2 Pix2Pix の実行

Pix2Pix で条件画像と実画像のペアを用いて機械学習を行い、条件画像に対して本物に類似する画像を生成するた

めのトレーニングを行う。

4.3 モデルのトレーニング

CMP ファサードデータセットから 400 画像をダウンロードし、トレーニングを行う。Listing 1 に示すコードを入力し、BtoA トレーニングを実行する。

Listing 1 BtoA トレーニング実行コード

```
1 py pix2pix.py --mode train
2 --output_dir facades_train
3 --max_epochs 100
4 --input_dir facades\train
5 --which_direction BtoA
```

5 実験

5.1 節では学習データの作成について、5.2 節では世代数の比較について示す。

5.1 学習データの作成

画質改善が目的であるため、高画質化のための学習データを作成する必要がある。学習データに使用する画像の一例を図 2 に示す。



図 2 ドライブレコーダーの画像

pix2pix には予め学習用データセットが用意されているが、今回の目的はドライブレコーダー画像の高画質化なので、それに適した学習データを自作することになる。pix2pix は図 3 に示すような、2 枚の対になった画像を繋げて 1 枚にしたものを学習データとして用いる。(256px × 256px を 2 枚繋げた 256px × 512px の画像)



図 3 ドライブレコーダー画像の学習データ例

学習データの画像、図 3 を作成するには、以下の 3 つの

処理を行う必要がある。

- 手順 1. 元の画像を 256px × 256px に切り分ける。
- 手順 2. 切り分けたそれぞれの画像にぼかしを入れる。
- 手順 3. 切り分けた画像とぼかしを入れた画像を結合する。

この 3 つの処理を行う際に使用したプログラムを以下の Listing 2 に示す。

Listing 2 の学習データ生成の主要なプログラムについて述べる。まず、1 行目の splited 文で画像の切り分けを行う(手順 1)。これは 256px × 256px に画像を切り分ける。この際に、入力する画像は 256px × 256px 以上でなければならない。また、ファイル名に日本語が含まれていると、エラーが発生する。次に、3 行目の blurred 文で、切り分けたそれぞれの画像にぼかしを入れる。そして、5 行目の margined 文で、切り分けられた画像とぼかしを入れた画像を結合する。

Listing 2 学習データ生成のためのプログラムの一部

```
1 splited_image = img[256 * j : 256
2 * j + 256, 256 * k : 256 * k + 256]
3 blurred_image = cv2.blur(splited_
4 image,(10, 10))
5 margined_image = cv2.hconcat
6 ([splited_image, blurred_image])
```

5.2 世代数の比較

世代数とは、「1 つの訓練データ何回繰り返して学習させるか」の数のことである。学習回数を増やすほど、画質改善の精度が向上するとは限らず、「過学習」という状態に陥ることがある。過学習の状態になると画質改善の精度が落ちるので、「学習が進んで精度の向上がこれ以上見込めないとなったら、そこで学習を止める」、ということが必要となる。次に世代数 100 の画像 図 4 と、世代数 200 の画像 図 5 を示す。

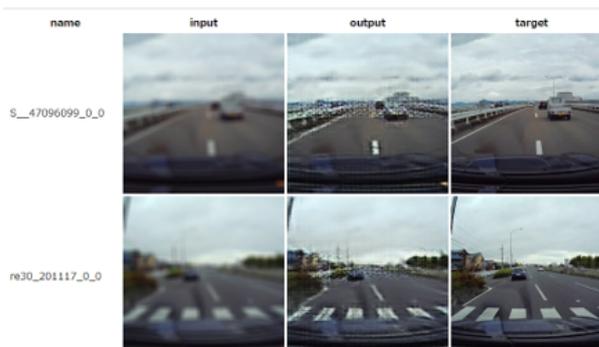


図 4 世代数 100 の画像例

世代数 100 の画像 図 4 と、世代数 200 の画像 図 5 の「output」(コンピュータが自動生成した画像)を見比べる。

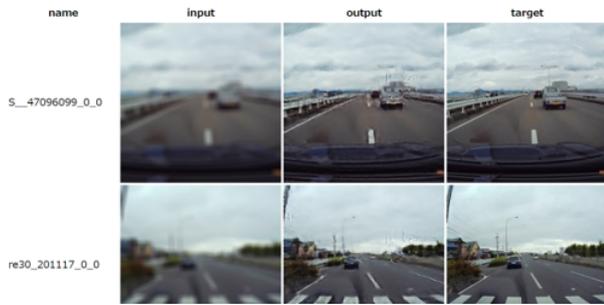


図 5 世代数 200 の画像例

世代数 100 の画像より世代数が 200 の画像の方が、車体部分を中心として画像の粗さが少なくなっている。

次に、世代数を変更する際の、プログラムの変更点について示す。Listing 1 を参照とする。Listing 1 のプログラムは、世代数が 100 である。例えばこの世代数を 200 にするには、プログラム 1 行目「epochs」の値を 100 から 200 に書き換え、実行する。

6 評価

6 章では、実験の評価について示す。まず 6.1 節では、PSNR について述べる。次に 6.2 節では、評価方法について述べる。最後に 6.3 節では、その結果について示す。

6.1 PSNR について

画質改善の評価方法は、PSNR(Peak signal-to-noise ratio) を使う。PSNR とは、2 枚の画像がどれだけ相違っているかを表す指標のことである。本研究では、ドライブレコーダー映像の画質改善を目的としており、学習データ数と世代数によって画質が改善される度合いが変化する。そのため、学習データ数と世代数による画像の相対的評価を算出すべく PSNR を採用した。ここで「2 枚の画像」という点の一つの大事な点であり、例えば 1 枚の画像だけをみて、それが画質が良い悪いを判断する指標ではないことを注意した。あくまで、相対的に画質を評価する際の指標として扱った。

6.2 評価方法

画質改善の評価方法として、まず比較対象の説明をする。3.1 節の図 3.2 の「コンピュータに与えた学習用の画像」と「コンピュータが自動生成した画像」を比較対象とした。これにより、PSNR の値が大きいほど、画質改善の度合いが大きいと判断した。

評価パターンとしては、次の 6 パターンの組み合わせとした。

1. 学習データ 100 枚 × 世代数 100 回
2. 学習データ 100 枚 × 世代数 200 回
3. 学習データ 200 枚 × 世代数 100 回
4. 学習データ 200 枚 × 世代数 200 回
5. 学習データ 300 枚 × 世代数 100 回

6. 学習データ 300 枚 × 世代数 200 回

またここでは、学習データ 100 枚 ~ 300 枚に共通する画像の中から 1 枚の学習データを図 6 に示す。



図 6 学習データ 1

6.3 結果

6.2 節で述べた評価方法から算出された評価結果を下記の表 3 に示す。縦軸は、学習データ 100 枚から 300 枚の中で共通して使用した画像の中から、無作為に選出した 10 枚をそれぞれ 1 から 10 とした。また、横軸は評価パターン番号とする。また、表 3 に示す PSNR の値の有効数字は小数点以下第三位とする。表 3 の見方について示す。例えば、3

表 3 評価結果

	1	2	3	4	5	6
1	22.243	22.462	22.576	22.000	21.949	22.803
2	22.983	24.386	25.399	24.128	23.806	24.988
3	22.592	23.334	23.164	22.917	22.563	23.407
4	21.235	21.070	20.437	20.657	20.733	21.382
5	22.320	21.278	20.742	21.227	21.126	21.858
6	20.860	20.054	20.090	20.203	20.432	20.840
7	21.378	21.772	21.019	21.395	21.245	22.164
8	21.553	21.130	21.113	20.916	20.752	21.495
9	21.934	21.822	21.991	21.821	21.553	21.553
10	21.267	21.630	21.015	21.410	21.129	21.826

列目の 2 行目の値 (25.399) は、学習データ 2 の評価パターン 3 の場合の PSNR の値を示す。

7 考察

まず、図 6 の評価パターン 1 から 6 の実行結果を図 7 から図 12 に示す。

次に、図 6 の入力画像と、図 7 に示す実行例の出力画像を見比べた時の観点から述べる。入力画像は画像全体において不鮮明で、建物や景色や車体が認識し辛い。比べて、出力画像は評価パターン 1 から 6 において差はあるものの、入力画像に比べ、建物や車体の輪郭、色覚、景色において鮮明に認識出来るようになっている。このことから、画質改善においては成功した。

また学習データの特徴も出力結果に関係していると考えられる。例えば、空や道路、全画像下部のボンネット部分

は、どの画像にも写っているのに、学習量が多く、画像の粗さは目立ちにくい。しかし木や固有建築や標識などの画像によって特徴が異なるものは粗さが目立った。

次に、表 3 から読み取ることのできる PSNR の値の差の観点から述べる。まず、同入力画像の評価パターンの違いについて述べる。例えば 1 行目 (学習データ 1 における評価パターン 1 から 6 の PSNR の値) においては、評価パターン 6 (学習データ 300 枚 × 世代数 100 回) の PSNR の値が 22.803 となり、最も値が大きい。しかし 9 行目 (学習データ 9 における評価パターン 1 から 6 の PSNR の値) においては、評価パターン 3 (学習データ 200 枚 × 世代数 100 枚) の場合が 21.991 となり、最も値が大きくなる。これは、評価パターン 1 から 6 の中で、評価パターン 3 の PSNR の値を最大値として過学習を起こしているためと考えられる。

また、表 3 の 6 列目 (評価パターン 6 における学習データ 1 から 6 の PSNR の値) について述べる。学習データ 1, 3 においては、評価パターン 1 から 6 において評価パターン 6 が最大値をとっていることを、読み取ることが出来る。しかし、学習データ 2, 4, 5, 6, 7, 8, 9 においてはこれに該当しない。これは学習データ、世代数共に値を大きくすることが、画質改善に繋がっているとは限らないことを示している。

以上 2 点から、PSNR 値が最大値をとる評価パターンは、画像によって異なることが考えられた。この理由としては、前述した、学習データの特徴量の偏りが出力画像に反映されており、各画像に適した評価パターンがあるためであると考察した。



図 7 評価パターン 1 図 8 評価パターン 2



図 9 評価パターン 3 図 10 評価パターン 4



図 11 評価パターン 5 図 12 評価パターン 6

8 終わりに

本研究で pix2pix を用いたドライブレコーダーの画質改善手法の提案をした。提案手法の有効性を確認することを目的とした実験の結果、実際にドライブレコーダー映像の高画質化をすることができ、認識精度の向上に有効性を認めた。学習データ、世代数に調整は必要なものの、実際のドライブレコーダーの画像を用いた状況検分時に本研究が役立つ。

参考文献

- [1] 石井 威望, 井口 雅一, 廣瀬 通孝, 藤岡 健彦, 岩田 洋夫, “画像解析を利用した自動車用ヒューマン・インターフェース,” 計測自動制御学会論文集, Vol. 22, No. 11, pp. 1183-1188, 1986 年.
- [2] 岡田 隆三, 田辺 淳, 伴野 守保, “自動車の運転支援・自動化のための画像センシング技術とその実践,” デジタルプラクティス, Vol. 8, No. 2, 2017 年, 4 月.
- [3] 戸部田 雅一, 高椋 佐和, 名取 直毅, 本田 芳寛, 平岩 直浩, 水野 貴史, “車載画像物体検出に向けた敵対的画像復元の検討,” 人工知能学会全国大会論文集, 1N3-GS-10-05, 2016 年, 6 月.
- [4] 石 晶, 李志豪, 本吉 俊之, 大西 直, 森 裕紀, 尾形 哲也, “End-to-End 自動運転モデル改善のための画像認識サブタスクの設計と評価,” 人工知能学会全国大会論文集, 1L2-J-11-01, 2019 年, 6 月.
- [5] 柴田 剛志, “学習型超解像を用いた自動車ナンバープレート画像推定,” 画像電子学会誌, vol. 44, No. 2, pp. 363-367, 2015 年.
- [6] 原 佑輔, 小島 颯平, 内山 彰, 梅津 高朗, 山口 弘純, 東野 輝夫, “ドライブレコーダー映像を用いた頭部検出に基づく人流推定法の提案,” マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, pp. 253-261, 2016 年, 7 月.
- [7] 一般社団法人日本自動車会議所, <https://www.aba-j.or.jp/info/industry/10071/>, 参照 May18, 2020.