

リカレント型ネットワークを用いた未来予測による自動車走行中の異常検知

正会員 ○中野雄大*1 同 山田悟史*2

11. 情報システム技術—6. ソフトウェア・アルゴリズム・システム化
異常検知, ラベリング処理, リカレント型ネットワーク, 類似度評価

1. はじめに

近年、我が国に限らず、モータリゼーションが進行し、車社会とも呼ばれるまでに我々人々と車は密接にかかわってきている。車とは我々人々を包み込む空間であり、ドアを閉めれば車内が自分達だけの居場所となる。これは玄関の扉を閉めれば自分達だけの居場所となる住宅と類似しており、車も住宅と同じプライベート空間と言える。しかし、車文化の普及とともに車を使用する年齢層も広がり、高齢ドライバーによる接触事故や、自動車運転者による故意的な危険運転や近年特に注目されている煽り運転、その他道路交通法を無視した無茶な運転による事故が多発しており、事故によるプライベート空間の侵害が近年問題となっている。

そういった事故を防止するべく、ドライブレコーダーとAIを用いた異常検知による様々なオートブレーキシステム（以下、従来の異常検知と表記）が開発され、実用化に向けて進んでいる。この従来の異常検知は、ドライブレコーダーに映るものをラベリングし、ラベリングされた物の急な飛び出しや、急接近などによって異常発生と認識するシステムになっている。この従来の異常検知は、ラベリングされている物に対しては高い確率で異常検知できる一方で、図1のように落下物やかかんでいる子供の後ろ姿などの想定が困難な障害物にはラベリング処理が弱いことが課題となっている。

この課題を解決する手法として、リカレント型ネットワークである ConvLSTM（Convolution Long Short-Term Memory）¹⁾とヒストグラム比較の2つを用いた従来の異常検知に代わり得る、ラベリングを要せず異常を検知できる新たなモデルの構築が考えられる。ConvLSTMを用いて過去の記憶から何事も無ければ来るはずの未来予測画像を生成し、ヒスト

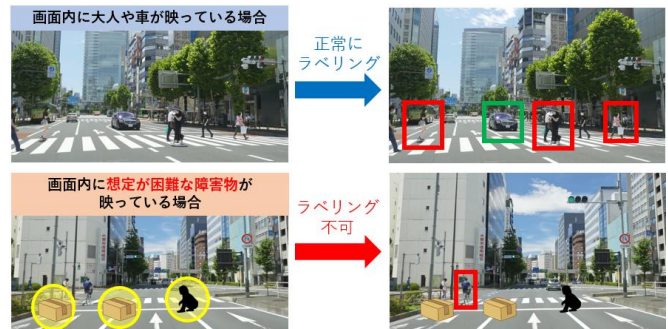


図1. ラベリング処理

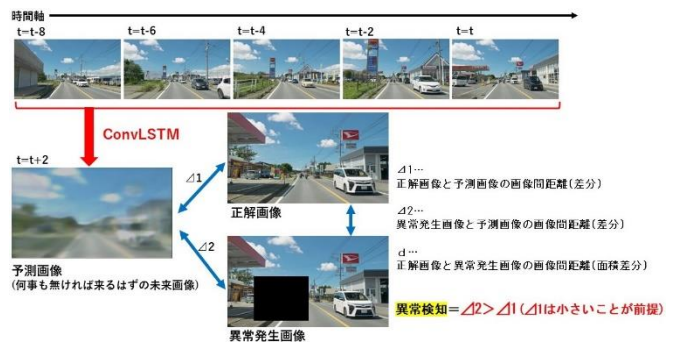


図2. 研究の流れ

グラム比較を用いて画像間の類似度を即座に測り、異常を検知するシステムである。

2. 概要

2-1. 提案する異常システムの流れ（図2）

システムでは車載カメラにより2秒間隔で生成された5枚の連続画像からConvLSTMを用いて2秒先の6枚目の何事も無ければ来るはずの未来画像（予測画像）を生成する。この予測画像と実際の2秒後の画像（正解画像）との差を $\Delta 1$ とする。異常が発生した場合の画像（異常発生画像）との差を $\Delta 2$ とする。 $\Delta 2 > \Delta 1$ となることは異常が無ければ訪れていた未来とは異なる状況になったことを意味している。本研究ではこの状況を異常発生と定義する。但し本研究では異常発生画像は正解画像の一部を任意面積（d）の正方形をランダムな単色で重ね

描きした画像とする。d を段階的に小さくした際の正しい異常検知率をシステムの検証値とする。検知率の高い△の計測方法の設定も研究課題である。

2-2. 対象道路

熊本県内の国道 57 号線から始まり、阿蘇西 IC を通過し、熊本市中心部までの約 70 km を対象道路とする。今回は都心部の様に交通量や建物量などの視覚的情報量が多い道路ではなく、ConvLSTM を用いて生成された予測画像の精度をより上げるために、建物量や交通量がそれほど多くない道路を選択した。

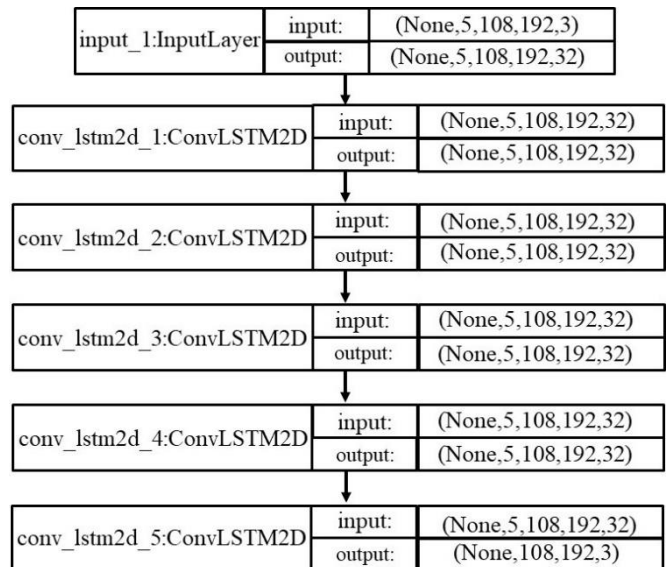
2-3. dataset の作成

データセットの元となる画像は、車載動画²⁾をフレームレート毎に静止画を切り出して生成した。フレームレートは 0.5fps とし、2327 枚の連番画像に分割したものを用意した。この画像群を訓練：検証：テスト用に 7.2:1.8:1 の比率で先頭から分割した。また、後述の学習モデルに合わせて画像に zero padding と resize を施し、(108×192×3) に形状を変更した。

2-4. ConvLSTM の構造

ConvLSTM の層構成は 5 層構成とした(図 3)。input の引数は、(サンプル数、フレーム数、画像の高さ、幅、RGB) となっており、最終的な output の引数は、(サンプル数、画像の高さ、幅、RGB) となっている。ランダムに選択された 5 フレームの連番画像を入力し、2 秒後の予測画像を 1 枚生成するモデルとなっている。今回使用したモデルでは各畳み込み層の間で batch normalization を行っている。これを François Chollet の運営するオープンソースの機械学習フレームワークである keras を利用して、Google 社が提供する Google Colaboratory 上で実行した。

学習時のハイパーパラメータは表 1 に示す。この各項目については以下で説明する。batch size は 1step 内で並列処理するデータ数を指し、本研究では 10 枚の画像を同時に計算させた。損失関数は正解画像と AI による予想画像の不一致度合いを示す loss を計算する関数である。評価関数は学習させたモデルの性能の良さをはかる関数である。最適化関数は loss が最小になるように学習時の重みを更新する関数である。また、momentum と kernal size に関しては、自ら調整して精度検証を行った。



入力(サンプル数、フレーム数、高さ、幅、RGB)
最終的な出力(サンプル数、高さ、幅、RGB)

図 3. ConvLSTM の構造

表 1. 学習時のハイパーパラメータ

項目	値
batch size	10
損失関数	平均絶対誤差(mae)
評価関数	平均二乗誤差(mse)
最適化関数	Adam
学習率	0.001
epoch数	10

表 2. 各方法の類似度表現

ピアソンの相関係数	-1から1の範囲で類似度を表す 1に近い方が類似度が高い
カイ二乗検定	0以上の値で類似度を表す 0に近い方が類似度が高い
KLダイバージェンス	0以上の値で類似度を表す 0に近い方が類似度が高い
交差法	0以上の値で類似度を表す 0から乖離しての方が類似度が高い
バタチャリア距離	0から1の範囲で類似度を表す 0に近い方が類似度が高い

2-5. 画像類似度評価の方法

画像間の RGB のヒストグラムの類似度とする。画像内の色の分布の様子を、X 軸を RGB (0 ≤ X ≤ 255)、Y 軸を分布度合いで示すヒストグラムで表現し、画像間でそのグラフを比較して類似度を測る。ヒストグラム作成には OpenCV³⁾ の calcHist 関数を用いる。また、表 2 に示したようにヒストグラム比較には 5 種の方法を検討する。

2-6. 異常発生素材の作成 (図 4)

前述の方法で作成し、d の割合の水準は 40%、35%、30%、25%、20%、17.5%、15%、12.5%、10%、7.5%、5%、2.5% の 12 段階とする。



図 4. 異常発生画像の一例

3. 検証結果

3-1. ConvLSTM による予測画像評価

3-1-1. 精度評価

kernel size を 5 パターン、momentum を 3 パターン設定し、計 15 パターンで精度評価を行った。(表 3) 今回は、一番精度が高いパターン O で生成された予測画像を使用する。

3-1-2. 予測画像と正解画像の比較

予測画像と正解画像の比較 12 セットを図 5 に示す。上段が正解 (cor) 画像で、下段が予測 (pre) 画像である。形状のエッジが不明瞭ではあるが、車や人、建物の配置、道路の中央線や横断歩道の位置はほぼ正確であることが読み取れる。また、直進だけでなく曲がり角についても同等の正確さが読み取れる。

3-2. 類似度結果

正解画像と予測画像、異常発生画像 (面積 d の正方形をランダムな単色で重ね描きした画像) の 3 枚から構成されるある 1 セットの RGB のヒストグラムを図 6 に例示する。グラフは R、G、B の 3 種類を重ねてプロットしたものである。また、図 6 で示した画像を用いて $\angle 1$ と $\angle 2$ の大小関係を画像類似度評価の方法で既述した 5 種の方法で測った結果を表 4 に示す。青色で塗られている箇所が $\angle 2 > \angle 1$ が成立しており、正しく異常を検知できたことを示す。反対に橙色で塗られた部分は、異常検知として成功していない部分である。

3-3. 検知率の検証結果

d は異常発生画像の作成で既述したように 12 水準、テストデータの 232 セット (1 セット正解画像、予測画像、異常発生画像の 3 枚の画像群)、異常発生パターン数 300 (単色と位置が一様分布でランダムに生成)、計 2,505,600 枚 (835,200 セット)

表 3. 各パターンの精度

パターン	kernel size	momentum	val loss
A	1*1	0.99	0.1009
B	1*1	0.1	0.1070
C	1*1	段階的※2	0.1048
D	3*3	0.99	0.0986
E	3*3	0.1	0.1005
F	3*3	段階的※2	0.1169
G	5*5	0.99	0.1049
H	5*5	0.1	0.1001
I	5*5	段階的※2	0.0998
J	7*7	0.99	0.1104
K	7*7	0.1	0.1051
L	7*7	段階的※2	0.1046
M	段階的※1	0.99	0.1034
N	段階的※1	0.1	0.1014
O	段階的※1	段階的※2	0.0983

※1 7*7、5*5、3*3、1*1と1層ずつkernel sizeを小さく設定

※2 0.4、0.6、0.8、0.99と段階的に大きく設定



図 5. 予測画像 (pre) と正解画像 (cor) の比較

の画像を用いて検知数を算出し、各 d の割合毎にテストデータのセット数で割って平均検知率を算出した結果を図 7 に示す。

表 4 と同様に、青色のグラフが検知成功、橙色のグラフが検知失敗である。KL ダイバージェンスに関しては、いずれの面積割合も検知率が 0 を示したためこの図から省略した。

3-4. 考察

ピアソンの相関係数による検知率が高かったのは、この手法は共分散や標準偏差を用いて類似度を測るため、視覚的類似度が高く読み取れた正解画像のヒストグラムの偏差よりも、ヒストグラムが一部乖離していた異常発生画像の方が予測画像との偏差が大きくなり、正解画像の方が予測画像との相関関係が強くなったためだと考えられる。また、カイ二乗検定は、ヒストグラムの差を二乗するため、グラフの乖離が大きい異常発生画像よりもグラフの一致率が高い正解画像の方が差分が小さく、検知率が高くなったと考えられる。一方で、交差法は、min 演算を適用しているため、グラフの乖離が類似度に影響せず、検知率が低くなったと考えられる。

4. まとめ

検知率に関しては、本研究で行った条件では、d の割合が 40% 以下であれば手法に関わらずほぼ確実に異常を検知することができることを把握した。また、7.5% 付近までであればピアソンの相関係数とカイ二乗検定は過半数以上の検知率を示すが、他の 3 つの手法については検知率が低くなり、ピアソンの相関係数とカイ二乗検定を使用する異常検知が有効であることを把握した。

検知時間に関しては、予測画像を出力するのに要する時間は GPU 環境下で 0.97482 秒、検知率を測るのに要する時間は CPU 環境下で 0.00478 秒で計 0.9796 秒であり、リアルタイムシステムに落とし込む場合、2 秒先の異常を検知するには約 1 秒の余裕があることも把握した。

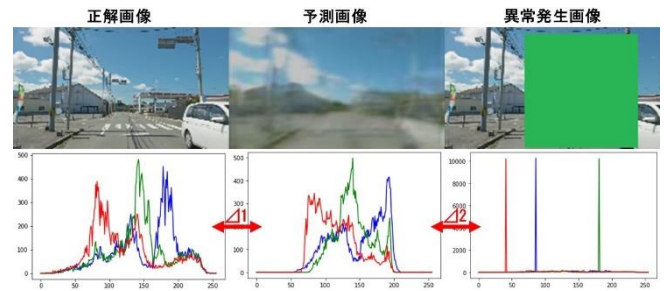


図 6. ヒストグラム比較 (X : RGB, Y : 分布度合い)

表 4. 類似度結果

比較手法	予測-正解 (∠1に關与)	予測-異常 (∠2に關与)
ピアソンの相関係数	0.755	0.422
カイ二乗検定	15190.357	6719156.668
KLダイバージェンス	73038.897	3261.883
交差法	14774.0	18049.0
パタチャリア距離	0.332	0.255

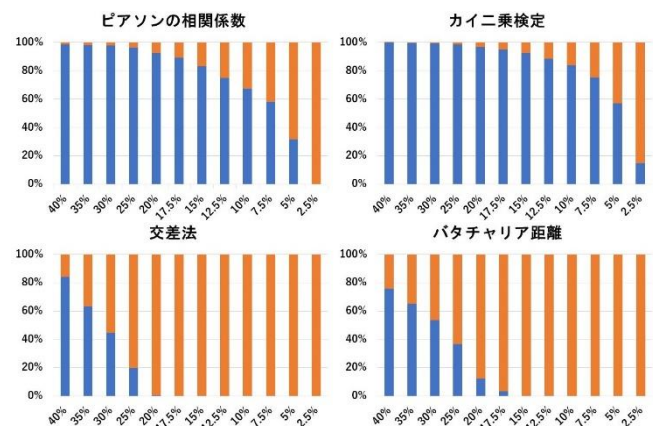


図 7. 検知率結果 (X : d の割合, Y : 検知率)

今後の方針としては、今回の様に異常発生箇所を単色の正方形で限定していたため色や形状のパターン数を増加させることや、色の分布に依存するヒストグラム比較以外の類似度評価を用いて検知調査を行うことが挙げられる。また、今回は突発的な異常発生について研究を行ったが、対向車が徐々に中央線をはみ出してくるといった突発的でない異常に対する検知の検証も挙げられる。

参考文献

- 1) Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, Wang-chun Woo, Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, arXiv:1506.04214v2, 2015
- 2) Dashcam Roadshow, “4K 阿蘇市大観峰→国道 57 号→熊本中心ドライブ”, [https://www.youtube.com/watch?v=fiSzmZjPeFM], (最終参照日 2022 年 12 月 4 日)
- 3) OpenCV, [https://opencv.org/], (最終参照日 2023 年 1 月 5 日)

*1 立命館大学 理工学部建築都市デザイン学科 Univ.

*2 立命館大学 理工学部建築都市デザイン学科 講師・工博 Ritsumeikan Univ., Dr.Eng.

Architecture and Urban Design, Ritsumeikan

Lecture of Architecture and Urban Design,