

GAN を活用した動画像異常検知手法の構築と 労働災害防止への応用の検討

橋本 慧志¹ 工藤 謙一² 高橋 孝幸³ 梅田 和昇⁴

1 中央大学大学院理工学研究科 〒112-8551 東京都文京区春日 1-13-27

2 中央大学研究開発機構 〒112-8551 東京都文京区春日 1-13-27

3 プリマハム株式会社開発本部 〒300-0841 茨城県土浦市中向原 635

4 中央大学理工学部 〒112-8551 東京都文京区春日 1-13-27

E-mail: 1 hashimoto@sensor.mech.chuo-u.ac.jp, 2 kkudo32h@tamajs.chuo-u.ac.jp

3 takayuki.takahashi@primaham.co.jp, 4 umeda@mech.chuo-u.ac.jp

あらまし 本稿では、GAN を用いた新たな動画像異常検知手法を提案し、労働災害防止のための活用可能性について検討する。時空間敵対的生成ネットワークに対して各種安定化手法を導入し、さらにマッチング項の追加により再構成品質を改善した。公開データセットと工場内の作業を模した独自のデータセットに対して定量的評価を行い既存手法に対しての有効性を確認した。

キーワード 異常検知, GAN, 労働災害

1. 序論

我が国の 2018 年度の製造業における、労働災害（労災）による死傷者は、127,329 人であった[1]。外国人労働者の増加や高齢者雇用の増加に伴い、製造業の現場は多様化の様相を呈している。これに起因し、労災の発生件数の減少割合は鈍化している。以前より労災を未然に防止するための様々な対策がなされてきた。しかし、リスクアセスメントなどの労災を防止する取り組みは、一般に人手によって実施され、その自動化に関する研究は国内外において乏しい。

一方、日常生活における異常をとらえる動画像異常検知に関する研究が行われている[2-6]。Luo ら[7]は、時空間ネットワーク（STN: spatio temporal networks）を用いて、encoder-decoder ベースの動画像の再構成による異常検知手法を提案している。さらに近年では、敵対的生成ネットワーク（GAN: generative adversarial networks）の活用が盛んであり[8-10]、Ravanbakhsh ら[11]は、pix2pix を用いて optical flow 画像と frame 画像間の関係性をモデル化して異常検知を行っている。しかし、GAN は学習が不安定になる問題がある。

本稿では、以上の背景を踏まえ、時空間敵対的ネットワークを用いた新たな動画像異常検知手法の構築を目的とする。さらに、労働災害防止のための現場での応用可能性について検討する。本稿の貢献は次のとおりである。

- 動画像異常検知技術を、労働災害防止を目的に活用する新しい取り組みである。
- GAN の課題である不安定性について、各種安定化手法の実装により改善した。

- マッチング項の導入により、モデルの再構成品質を改善した。

以下では、最初に時空間敵対的生成ネットワークと異常検知手法のフレームワークを示す。次に、検証実験に関して示し、最後に結論と今後の展望を述べる。本手法の概要図を図 1 に示す。

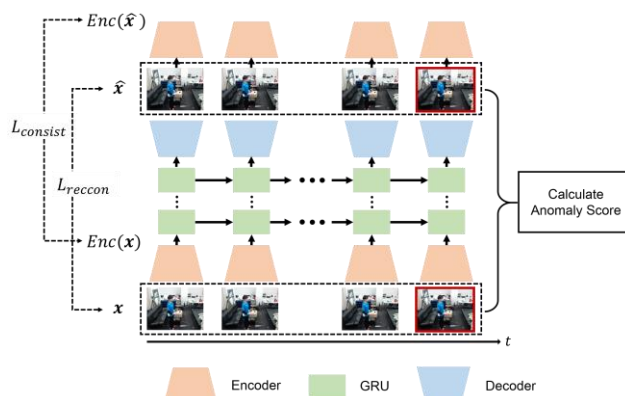


図 1 提案手法の概要

2. 提案手法

本手法では、動画像をモデル化するため、時空間敵対的生成ネットワークを用いる。また、各種安定化手法とマッチング項を導入することで、敵対的生成ネットワークの不安定性問題の解決と再構成品質の向上を図る。我々のモデルは、Encoder, Decoder, Discriminator の 3 つから構成されている。Encoder は入力を潜在空間へと写像し、Decoder は潜在変数を観測空間へと逆写像する。Discriminator は生のデータか Decoder による生成されたデータかを識別する。モデルが解くべき最適化問題は式(1)で表される。

$$\min_{Enc, Dec} \max_{Dis} L_{recon} + L_{gan} + L_{feat} + L_{consist} \quad (1)$$

ここで、 Enc は Encoder, Dec は Decoder, Dis は Discriminator を表す. また, L_{recon} は入出力間の再構成損失, L_{gan} は敵対的損失, L_{feat} は Feature matching 損失, $L_{consist}$ は, 潜在変数間の損失を表す.

さらに, 本手法の特徴について述べる. 本手法では, 入出力間の再構成誤差のみならず式(2)に示す潜在変数間のマッチング誤差項 $L_{consist}$ を導入する.

$$L_{consist} = \mathbb{E}_{\mathbf{x} \sim p_{data}} \|\text{Enc}(\mathbf{x}) - \text{Enc}(\text{Dec}(\mathbf{x}))\|_1 \quad (2)$$

ここで, $\mathbf{x}$ は入力である. これにより, 潜在空間上での誤マッチングを抑制し, 再構成品質の向上が見込まれる. また, 各種安定化手法として, [12]を参考に, TTUR, Self attention, Spectral normalization と, Feature matching[13]を導入することで, 敵対的学習の不安定性を改善する.

最後に, 異常度の算出手法について述べる. 入力 $\mathbf{x}$ に対する異常度 $a(\mathbf{x})$ を式(3)のように定義する.

$$a(\mathbf{x}) = \|\xi_T - \text{Dec}(\text{Enc}(\xi_T))\|_1 \quad (3)$$

ここで, 入力 $\mathbf{x}$ は固定長 T を有する部分時系列 $\xi_0, \xi_1, \dots, \xi_T$ から構成される. ξ は各フレームの画像を表す. 異常度は, 入力の最終フレームの再構成誤差の L1 ノルムから定義され, これを用いて異常検知が可能となる.

3. 検証実験

本稿では, 動画像異常検知において一般的な公開データセットである Avenue[14]と, 我々が独自に用意した労働災害検知データセットを用いて実験を行った. 図 2 にデータセットの例を示す. 前者は 16clip の訓練データ, 21clip のテストデータ, 後者は 5clip の訓練データ, 7clip のテストデータからなる. これら 2 つのデータセットに対して, Receiver Operating Characteristic (ROC) 曲線の AUROC によるモデルの定量的評価を行った. なお AUROC 算出時には正規化した異常度を用いた. 表 1 に Avenue に対する結果を, 表 2 に独自データセットに対する結果を示す.

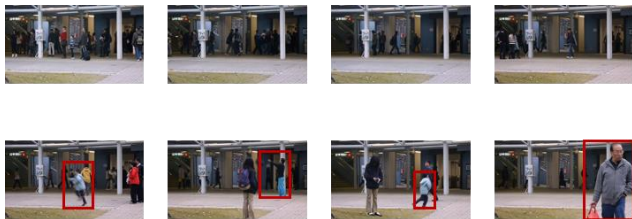


図 2 Avenue データセットの正常データ (上図) 及び異常データ (下図)

表 1 Avenue データセットに対する結果

	Ours	Conv-AE[5]	SRNN[7]
AUROC $\uparrow$	0.829	0.702	0.817

表 2 独自データセットに対する結果

	Ours	Conv-AE
AUROC $\uparrow$	0.728	0.571

2 つのデータセットにおいて既存手法を上回る結果を確認した. Avenue においては frame-level 検知で良好な結果となった. しかし, 独自データセットの結果を見るに, 数値的改善は現れたものの, 現場に適用可能な実用的性能には至っていない.

4. 結論

本稿では, 時空間敵対的生成ネットワークを用いた新たな動画像異常検知手法を構築した. Avenue データセットおよび独自の労災検知データセットにおいて, 既存手法を上回る結果を確認した. 今後の展望として, 異常の局所的な検知のためのフレームワークを導入するほか, 実環境における詳細な検証を行う予定である.

文 献

- [1] Ministry of Health, Labour and Welfare, “The status of occurrence of occupational accidents in 2018”, 2018.
- [2] Waqas Sultani, and et al., “Real-world Anomaly Detection in Surveillance Videos”, CVPR, 2018.
- [3] Yaxiang Fan, and et al., “Video Anomaly Detection and Localization via Gaussian Mixture Fully Convolutional Variational Autoencoder”, arXiv, 2018.
- [4] Guansong Pang, and et al., “Self-trained Deep Ordinal Regression for End-to-End Video Anomaly Detection”, CVPR, 2020.
- [5] Mahmudul Hasan, and et al., “Learning Temporal Regularity in Video Sequences”, CVPR, 2016.
- [6] Asim Munawar, and et al., “Spatio-Temporal Anomaly Detection for Industrial Robots through Prediction in Unsupervised Feature Space”, WACV, 2017.
- [7] Weixin Luo, and et al., “A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework”, ICCV, 2017.
- [8] Mahdyar Ravanbakhsh, and et al., “Training Adversarial Discriminators for Cross-channel Abnormal Event Detection in Crowds”, WACV, 2019.
- [9] Hung Vu, and et al., “Robust Anomaly Detection in Videos Using Multilevel Representations”, AAAI, 2019.
- [10] Mohammad Sabokrou, and et al., “AVID: Adversarial Visual Irregularity Detection”, arXiv, 2018.
- [11] Mahdyar Ravanbakhsh and et al., “Abnormal Event Detection in Videos using Generative Adversarial Nets”, ICIP, 2017.
- [12] Andrew Brock, and et al., “Large Scale GAN Training for High Fidelity Natural Image Synthesis”, ICLR, 2019.
- [13] Tim Salimans, and et al., “Improved Techniques for Training GANs”, NeurIPS, 2016.
- [14] Cewu Lu, and et al., “Abnormal Event Detection at 150 FPS in MATLAB”, ICCV, 2013.