

労働災害防止のための 動画像を用いた深層学習による異常検知手法

○橋本 慧志 (中央大学), 池 勇勳 (中央大学), 工藤 謙一 (中央大学),
高橋 孝幸 (プリマハム株式会社), 梅田 和昇 (中央大学)

Anomaly Detection Based on Deep Learning Using Video for Prevention of Industrial Accident

○Satoshi HASHIMOTO (Chuo Univ.), Yonghoon JI (Chuo Univ.), Kenichi KUDO (Chuo Univ.),
Takayuki TAKAHASHI (Prima Meat Packers, Ltd.), and Kazunori UMEDA (Chuo Univ.)

Abstract: This paper proposes an anomaly detection method for the prevention of industrial accidents using machine learning technology. Currently, a comprehensive system for preventing such accidents is basically done manually, and has not been automated. In order to carry out anomaly detection reliably, a sequential skeleton map of a person is used as training data from skeleton information extracted by OpenPose. Here, long short-term memory-based variational autoencoder (LSTM-VAE) is applied as a deep learning model. It is confirmed that this produces detection results with higher accuracy than conventional methods.

1. 序論

日本では、2018年度の製造業における、労働災害による死傷者は127,329人であった[1]。日本では、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。このため日本では減少傾向にあった、労働災害の発生件数の減少割合は鈍化している。Fig. 1に日本における労働災害の発生件数の推移を示す。

一度労働災害が発生すれば、企業に対して様々な責任が課せられることとなるので、以前より労働災害や事故を未然に防止するための、様々な対策がなされてきた。こうした取り組みの1つに、リスクアセスメントがある[2]。リスクアセスメントは、一般にハザードを特定してリスクを評価し、それに沿った改善行動を行う一連の流れのことを言う。リスクアセスメントの流れをFig. 2に示す。しかし、リスクアセスメントなどの包括的に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るものであり、その自動化に関する研究は、国内外において乏しいものとなっている。

一方、日常生活における様々な場面での異常をとらえる異常検知に関する研究が行われている[3-6]。一般的に異常検知とは、教師なし機械学習の手法を用いて正常な分布からの異常度を定義することで異常を発見するデータマイニング手法である。異常検知の流れをFig. 3に示す。異常検知は、防犯カメラから取得できる動画像を用いることで、犯罪行動の検出や、日常生活での転倒事故の検出を可能にするなど多方面での活用が期待されている。異常検知においては、主成分分析(PCA: principle component analysis)により正常データを表す部分空間を学習し、その分布の距離から異常度を算出する部

分空間法や、混合ガウス分布(GMM: gaussian mixture model)を用いて複数個の正規分布を最尤法により推定して異常検知を行う手法が存在している[7]。しかし、PCAやGMMなどの従来手法は、画像などの高次元のデータを扱う場合、高次元空間におけるクラス間距離とクラス内距離が近くなってしまう次元の呪いの問題が発生するため、高い精度での異常検知を行うことは困難である。そこで、近年深層学習技術の発展に伴い、異常検知分野でも生成モデルを活用した深層学習による手法が登場している。生成モデルの1つであるAutoEncoder(AE)[8]は、画像などの高次元のデータを潜在空間に符号化し、得た特徴から画像を生成するEncoder-Decoder型のニューラルネットワークである。また、Variational AutoEncoder(VAE)[9]は、上記のAutoEncoderにベイズ推論の枠組みを取り入れたものである。AEの潜在空間に正規分布を仮定することで、より高い精度の画像生成を可能にしている。高次元のデータ入力に対しては、AEやVAEなどの深層学習を用いた手法が用いられる。我々は、VAEを用いた労働災害防止のための異常検知手法を提案した[10]。しかし、これらのフレームワークは、その入力に対して、互いに独立同分布なデータである仮定を置く。その為、異常な行動を認識する場合の異常検知においては、時系列の相関を考慮できないので、これらのフレームワークは不適といえる。

そこで本研究では、以上の背景を踏まえ、目的を2つ定める。1つ目は、異常検知手法を用いて労災を防止する包括的な自動監視システムを構築することである。2つ目は、時系列を考慮可能な深層学習を用いることで、従来の異常検知手法と比べ高い精度での異常検知を可能にすることである。我々は、

[2]の資料および実際に存在する工場での労働災害の発生状況を根拠に、労働災害が発生する過程において、多くが工場内で定められているルールを逸脱した際に頻発すると考えている。本研究では、労働災害を未然に防止するために、こうしたルールを逸脱した行動を、工場内に設置されたカメラから取得できる動画像を用いて、異常行動として検知することに着目する。また、こうした行動を認識するために、時系列の入力を考慮可能な long short-term memory-based variational autoencoder (LSTM-VAE)[11]を用いて異常検知を行うこととする。具体的なタスクとして工場内での作業を模した複数の行動パターンの異常検知をあげ、検証実験により有効性を確認する。

最初に深層学習を用いた教師なし異常検知のフレームワークを紹介する。次に、検証実験に関して示し、最後に本手法の有効性に関して検証する。



Fig. 1 The number of industrial accidents in Japan



Fig. 2 The steps of risk assessment

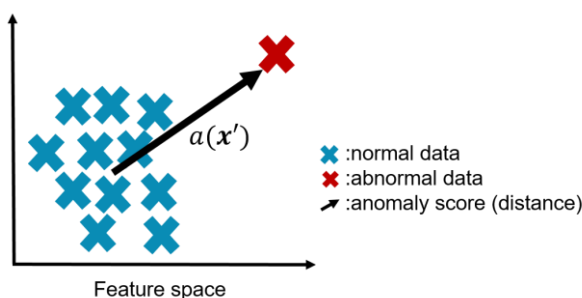


Fig. 3 Anomaly detection

2. 深層学習による異常検知

2.1 提案手法の概要

提案手法の流れを Fig. 4 に示す。本手法では、まずカメラを用いて人物の作業行動の様子を動画像に撮像し、OpenPoseを用いて人物の骨格情報を取得する。OpenPoseとは、Caoら[12]の骨格推定手法を実装したライブラリであり、Convolutional Neural Network (CNN) を組み合わせることで1枚の画像から25箇所の人物の骨格座標の検出を行うことが可能である。得られた骨格画像に前処理を施し、骨格動画像を作成する。そして、この動画像を入力としてLSTM-VAEによる学習を行う。学習を行う際には、事前に準備した正常な行動を収録したデータセットを用いて、モデルを学習する。最後に、上記の手法で学習されたモデルを用いて、入力 sequential skeleton map と、LSTM-VAEにより生成されたマップとの再構成誤差を取得し逐次的に異常度を算出することで異常検知を行う。

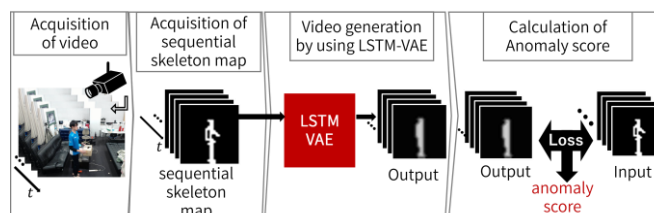


Fig. 4 The steps of the proposed method

2.2 OpenPose を用いた骨格情報の取得

撮像した入力動画像に対して、OpenPoseを用いて人物の骨格情報を取得する。次に、得られた骨格動画像に前処理を施す。処理の流れを Fig. 5 に示す。まず、OpenPoseを用いて得られた解像度 640×480 を有する人物の骨格画像を、首の座標を中心として 480×480 のサイズに切り取る。次に、28×28 のサイズに低解像度化を行う。ここで補間手法としてはバイリニア法を用いる。最後に、グレースケール化と2値化を行う。得られた 28×28 の骨格動画像に対して、sliding window 法を用いて、長さ T を有する部分時系列動画像を作成する。これを sequential skeleton map と呼ぶこととする。なお、sliding window における隣接する sequential skeleton map 間の幅は1とする。これを LSTM-VAE への入力とする。

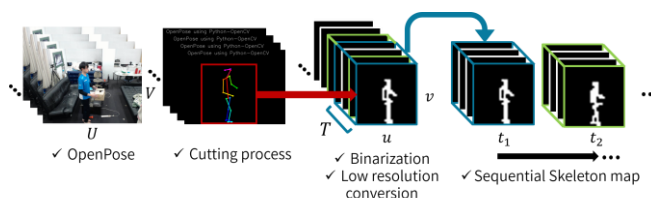


Fig. 5 Preprocessing flow

2.3 LSTM-VAE を用いた異常検知

本研究では、異常データの入手が困難であることを理由に、正常データのみでの学習が可能な教師なし学習を用いる。また、時系列を考慮可能な深層学習生成モデルである LSTM-VAE を用いることで、入力 *sequential skeleton map* と似た動画画像を生成させ、その入出力間の誤差から異常検知を行う。

本手法における LSTM-VAE のアーキテクチャを Fig. 6 に示す。ここで、 σ は分散ベクトル、 μ は平均ベクトルである。 ξ は長さ T の *sequential skeleton map* の構成要素である 1 フレームの画像であり、その下付き文字が各フレームに対応している。 z は潜在ベクトルである。また、FC は全結合層を表し、CNN は各フレームの特徴を学習する畳み込み層である。LSTM-VAE は、VAE に時系列情報の入力を可能とする LSTM 層を付加したものである。LSTM[13]は、RNN (Recurrent Neural Network) における勾配消失問題を、記憶メモリの概念を用いることで解決した時系列ニューラルネットワークの一つである。

本手法では、*sequential skeleton map* の各フレーム画像に対して、CNN を用いることで特徴を抽出する。CNN により得られた特徴量を LSTM 層にタイムステップ T の長さで入力し、時系列情報をモデリングする。LSTM からの出力は、正規分布を仮定した潜在空間 $N(0, I)$ へと次元圧縮される。この潜在空間を用いて *sequential skeleton map* を復号することで、入力と似た出力を生成することができる。学習時に用いる誤差関数を以下に示す。

$$L(x') = \frac{1}{2} \sum_{j=1}^J \left(1 + \ln(\sigma_j^2) - \mu_j^2 - (\sigma_j^2) \right) + \mathbb{E} \left[\sum_{i=1}^D (x'_i \ln y'_i + (1 - x'_i) \ln(1 - y'_i)) \right] \quad (1)$$

x' , y' は入力と出力を表し、その次元は D で表される。 J は潜在変数の次元であり、 $\mathbb{E}$ は期待値を表す。

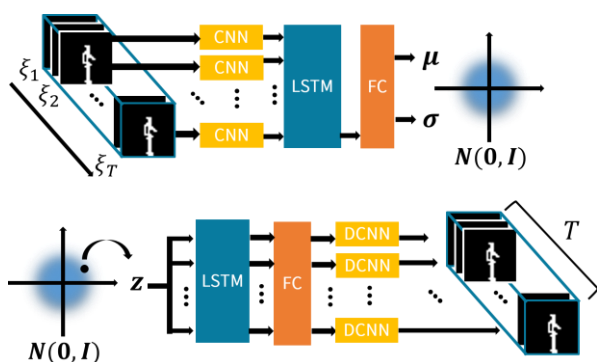


Fig. 6 The architecture of LSTM-VAE

3. 検証実験

3.1 実験概要

提案手法の有効性の確認および、AE, VAE を用いた従来手法との精度比較を目的として、検証実験を行った。まず、正

常データの学習に用いる *sequential skeleton map* を取得するために、工場内での作業を模した複数パターンの行動を行った。その行動一覧を Table 1 に示す。正常行動を収録した訓練データは LSTM-VAE の学習に用いる。そして、異常行動を含む検証データにより、モデルの性能を評価する。これらの行動は、実際の現場において実際に起こりうる逸脱行動を参考として作成している。その様子を、地上から 2.7m の高さに設置した単眼カメラを用いて、640×480 の解像度、30fps の撮像速度で収録した。収録した行動の一例を Fig. 7 に示す。このように収録した動画画像から、正常な行動のみを含む 4,618 フレームの骨格画像を取得した。そして、*sequential skeleton map* をタイムステップ数 $T=30$ として作成し、4,589 の *sequential skeleton map* を取得し、これを訓練データとした。同様に、Test に用いる異常な行動を含んだ 4,898 フレームの骨格画像を取得した。これら 2 つのデータの長さはおよそ 3 分である。上記と同様の方法で、4,758 の *sequential skeleton map* を取得した。最後に、正常な行動のみを含む *sequential skeleton map* を用いて、LSTM-VAE と比較対象である従来手法の AE, VAE の学習を行った。AE, VAE への入力は、4,618 フレームの骨格画像である。各手法のパラメータを Table 2 に示す。これらのハイパラメータは、後述する AUROC 値による定量的評価により、最適と思われる値を評価して用いている。なお、実装には、OS に Ubuntu16.04LTS、モデルの構築においては、ライブラリに Keras-2.1.4 を用い、GPU に Nvidia GTX1080Ti(11GB)を用いた。

評価には、モデルの精度を評価する指標である ROC (Receiver Operatorating Characteristic) 曲線が囲う面積の大きさである AUROC (Area Under ROC) を用いる。ROC 曲線は、異常検知手法の評価手法として一般に用いられ、横軸に誤報率 (FPR: false positive rate)、縦軸に真の検知率 (TPR: true positive rate) をとり、正常と異常を判別する異常度に対する閾値を変化させてプロットする。ROC 曲線が左上によるモデルほど、AUROC も大きくなるので、性能が良いことを示す。なお、本手法では、過学習を防止するために各 CNN 層に batch normalization (BN) [14]を行っている。

Table 1 The list of actions

Work	Normal action (for training and test dataset)	Abnormal action (for test dataset)
1	Carry the chair with both hands	Carry the chair with one hand
2	Cut the paper with both hands	Cut the paper with one hand
3	Discard trash in front of the box	Discard trash by throwing
4	Carry the box with both hands	Carry the box with one hand
5	-	Dig in the trash box
6	-	Work while running

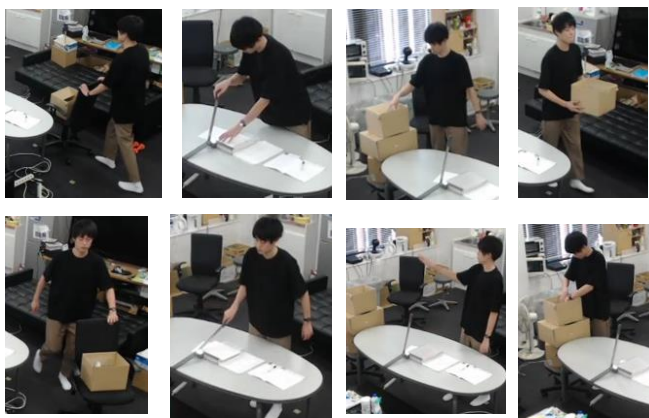


Fig. 7 Examples from our data set

Table 2 Parameters of the training phase

Parameters	LSTM-VAE	AE	VAE
Epochs	30	30	30
J	16	16	16
T	30	-	-
Latent dimension of LSTM	300	-	-
Batch size	16	32	32

3.2 実験結果

ROC 曲線を Fig. 8 に、縦軸に異常度、横軸にフレームをおいて描画した異常度の推移グラフを Fig. 9 に示す。異常度推移グラフにおいて、色がついている範囲は、異常行動が存在するフレームである。また、各モデルの AUROC 値を Table 3 に示す。提案手法では、異常な行動パターンをおおむね正しく異常と判別することができた。また、従来手法である AE, VAE と比較して高い精度での異常検知を達成した。LSTM-VAE により復号された sequential skeleton map の一例を Fig. 10 に示す。各画像上段が入力、下段が出力である。また、従来手法と比較して、ROC 曲線、AUC 値のどちらも提案手法のほうが良い精度を示す結果となった。LSTM-VAE の潜在空間の分布を Fig. 11 に示す。なお、青色、赤色の各点はそれぞれ正常データ、異常データからサンプリングされた z を表す。潜在空間の表示には、t-SNE[15]を用いて、潜在空間の次元(16 次元)を 2 次元に次元圧縮している。

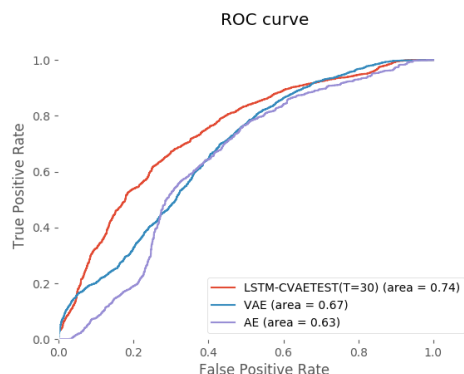


Fig. 8 ROC curve

Table 3 AUROC values

	AUROC
LSTM-VAE	0.74
AE	0.67
VAE	0.63

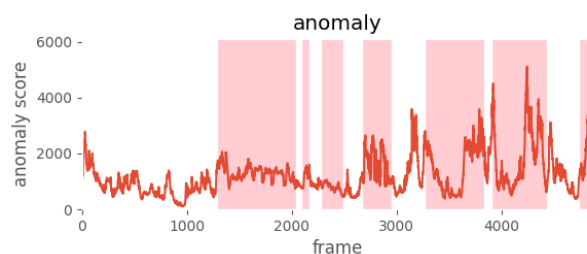


Fig. 9 The transition of the anomaly score

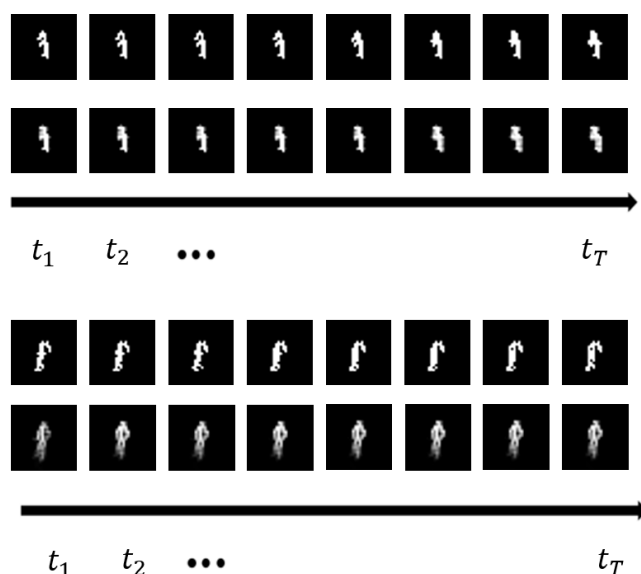


Fig. 10 Sequential skeleton map generated by LSTM-VAE

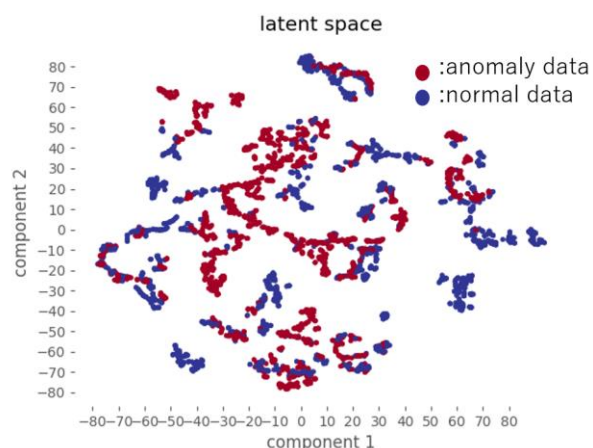


Fig. 11 The distribution of the latent space of LSTM-VAE

3.3 実験考察

提案手法は、3.2 節の結果を根拠に、労災防止のための自動異常検知手法として有効であると言える。また、LSTM-VAE による時系列を考慮可能な深層学習を用いた異常検知アプローチにより、従来の AE、VAE を用いた手法と比較しても高い精度を出すことができた。

しかし、本手法は OpenPose を用いた骨格情報が取得できる前提が必要であり、カメラから極端に遠距離に存在する人物の行動の異常検知は困難であると考えられる。また、LSTM-VAE の潜在空間に着目すると、次元圧縮による影響の存在も考えられるが、異常と正常が大きく分離できていないことがわかる。これは、sequential skeleton map の表す行動情報では、正常と異常とを大きく区別できないことを示している。すなわちこの sequential skeleton map の作成手法に関しても、低解像度化に伴い人物の姿勢を十分に表せていない可能性があるため、画像サイズの大きさなどで検討の余地がある。

4. 結論

本研究では、異常検知手法を用いて労災を防止する自動監視システムを構築した。また、時系列を考慮可能な深層学習を用いることで、従来の手法と比べ高い精度での異常検知を可能にした。

今後の展望として、骨格情報ではなく生の動画入力に対しての生成を行う手法の構築や、包括的な労災防止システムの実現に向けて周辺の物体や位置関係などの環境認識手法の検討がある。

参考文献

- [1] Ministry of Health, Labour and Welfare, “The status of occurrence of occupational accidents in 2018,” May 17, 2018.
- [2] Ministry of Health, Labour and Welfare, “To those working in the manufacturing industry For Work Safety and Health,” Jan. 17, p. 78, 2017.
- [3] W. Li, V. Mahadevan, and N. Vasconcelos, “Anomaly detection and localization in crowded scenes,” IEEE Trans. PAMI, 36 (1), pp.18-32, 2014.
- [4] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” CVPR, 2018.
- [5] D. Kimura, M. Narita, A. Munawar, and R. Tachibana, “Spatially-weighted Anomaly Detection with Regression Model,” arXiv, 2019.
- [6] V. Saligrama, and Z. Chen, “Video anomaly detection based on local statistical aggregates,” CVPR, pp. 2112–2119, 2012.
- [7] ファンチョンフィ, 岸礼子, 山本一真, 増田誠, “機械学習を用いた ATM 異常行動検知技術の検討”, 第 32 回人工知能学会全国大会論文集, 2018, p.10105.
- [8] G. E. Hinton and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks,” Science 313 (5786), pp. 504–507, 2006.
- [9] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” ICLR, 2014.
- [10] 橋本慧志, 池勇勲, 工藤謙一, 高橋孝幸, 梅田和昇, “骨格情報を利用した労働災害防止のための深層学習に基づく異常検知手法”, 日本機械学会ロボティクス・メカトロニクス講演会 2019 講演論文集, 2A1-G03, 2019.6.
- [11] D. Park, Y. Hoshi, and C. C. Kemp, “A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder,” IEEE Robot. Autom. Mag. Lett., 3 (3), pp. 1544–1551, July 2018.
- [12] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” CVPR, pp. 1302–1310, 2017.
- [13] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., 9 (8), pp. 1735–1780, 1997.
- [14] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” arXiv, 2015.
- [15] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., 9 (11), pp. 2579–2605, Nov. 2008.