

時空間敵対的生成ネットワークを用いた 動画画における異常行動検知

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

概要

近年、日常生活や製造業の現場等における深層学習を用いた動画画の異常を捉える試みが検討されている。特に、敵対的生成ネットワーク（GAN: generative adversarial networks）の活用が盛んであるが、非効率的であることや、不安定性が課題となっている。本稿では、時空間敵対的生成ネットワークを用いた教師なし学習による新たな動画画異常検知手法を提案する。提案手法はフレーム予測型のモデルであり、U-Net Discriminator の出力を活用するため、効率的かつ高精度な異常検知が可能である。

キーワード: 教師なし学習 敵対的生成ネットワーク 異常検知

1. 序論

近年の深層学習の発展に伴い、日常生活や製造業の現場等における監視カメラ画像の異常をとらえる動画画異常検知に関する研究が盛んに行われている[1-7]。動画画異常検知では主に時空間ネットワーク（STN: spatio temporal networks）を用いた手法[1-4]と、appearance 特徴と motion 特徴に分離してモデル化する手法[5-7]の 2 つに大別される。さらに、GAN（Generative Adversarial Networks）の活用が盛んであり[5-7]、異常検知の精度向上に貢献している。しかし、これら既存手法の多くに共通する課題として以下 3 点が挙げられる。1 つ目は非効率性である。STN を用いた手法は推論時に直近のフレームのみで異常度を計算する場合が多い。また、GAN を用いた手法の多くは推論時に Discriminator を無視する[8]。2 つ目は不安定性である。敵対的学習は一般的に不安定であり、手法の再現性に課題がある[9]。3 つ目はノイズの問題である。motion 特徴として用いる optical flow にはノイズがしばしば混入するので、性能に悪影響が及ぶ。

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画異常検知手法を提案する。本稿の貢献は次のとおりである。

- フレーム予測型の時空間敵対的生成ネットワークを構築し、効率的な手法を確立する。
- GAN を用いた手法の課題である不安定性について、安定化手法の実装により改善する。
- U-Net Discriminator の高レベル特徴の活用により、高精度な異常検知が可能となる。

以下では、最初に提案手法のフレームワークを示す。さらに、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。提案手法の概要を Fig. 1 に示す。

2. 提案手法

提案手法では、動画画を効率的にモデル化するために、フレーム予測型の時空間敵対的生成ネットワークを用いる。また、敵対的学習の不安定性を改善するために安定化手法を用いる。さらに、U-Net GAN[10]を参考に、Discriminator に U-Net を使い、その出力を融合することで、高精度な異常検知手法を確立する。

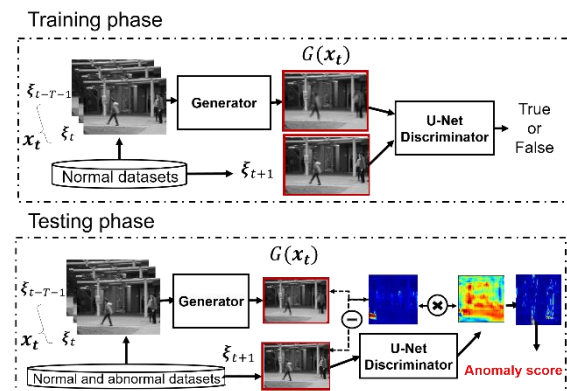


Fig. 1 提案手法の概要

2.1 時空間敵対的ネットワーク

モデルの詳細を Fig. 2 に示す。我々のモデルは、Generator, Discriminator の 2 つから構成されている。Generator は Encoder と Decoder からなる。Encoder は入力 of 動画画から畳み込み層と Convolutional-LSTM[11]を用いて特徴を抽出し、Decoder は Encoder により抽出された特徴を用いて逆畳み込み層により動画画の 1 フレーム先のフレーム画像を予測する。Generator の構造は、Lee ら[3]が採用したモデルを参考に、予測型のモデルへと拡張した。Discriminator には U-Net を採用し、真値か Decoder により予測されたフレーム画像かをピクセルレベルで識別する。

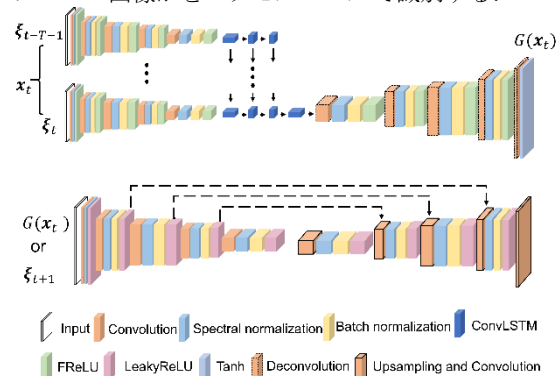


Fig. 2 Generator (上) と Discriminator (下) の詳細

2.2 敵対的学習

我々の時空間敵対的生成ネットワークは、以下の最小最大化問題を最適化する。

$$\min_G \max_D L_{D_{enc}} + L_{D_{dec}} + L_{consist} + L_{recon} \quad (1)$$

ここで、

$$L_{D_{enc}} = -\mathbb{E}_{\xi \sim p_\xi} [\log(D_{enc}(\xi_t))] - \mathbb{E}_{x \sim p_x} [\log(D_{enc}(1 - G(x_t)))] \quad (2)$$

$$L_{D_{dec}} = -\mathbb{E}_{\xi \sim p_\xi} \left[\sum_{i,j} \log [D_{dec}(\xi_t)]_{i,j} \right] - \mathbb{E}_{x \sim p_x} \left[\sum_{i,j} \log [D_{dec}(1 - G(x_t))]_{i,j} \right] \quad (3)$$

なお、 G はGenerator、 D, D_{enc}, D_{dec} はDiscriminatorとそのEncoder module, Decoder moduleである。 L_{recon} は真値と予測結果間の予測損失、 λ はそれに対する重みである。入力 x_t はある時刻 t における固定長 T を有する部分時系列 $\xi_{t-T-1}, \xi_{t-T-2}, \dots, \xi_t$ から構成され、 ξ は各フレーム画像である。 $[D_{dec}(\xi_t)]_{i,j}$ と $[D_{dec}(G(x_t))]_{i,j}$ は、ピクセル (i, j) におけるDiscriminatorの出力である。 $L_{consist}$ は[10]で導入されたCutmix[12]ベースのConsistency Regularizationである。さらに、提案手法では安定化手法として、TTUR[13], Self-Attention[14], Spectral normalization[15]を導入する。

2.3 異常検知

入力 x_t に対する異常度 $a(x_t)$ を以下に示す。

$$a(x_t) = \|\xi_{t+1} - G(x_t)\|_1 \quad (4)$$

$$map = (A - D(\xi_{t+1}))|F_2(\xi_{t+1}) - F_2(G(x_t))| \quad (5)$$

map はU-Net Discriminatorからの出力 $D(\xi_{t+1})$ と、成分がすべて1で $D(\xi_{t+1})$ と同じサイズの行列 A との差のマップ $A - D(\xi_{t+1})$ に、 D_{enc} の畳み込み第二層の中間表現 F_2 の差を融合したものである。我々がこのU-Net Discriminatorからの出力を融合するのは、異常な画像が入力されたときに、モデルがその領域を偽と判別することが期待できると考えるからである。異常算出に用いる最終的なスコア $S(x_t)$ は式(6)を用いて正規化することで求められる。

$$S(x_t) = \frac{a(x_t)}{\max(a(x_{1..m}))} \quad (6)$$

ここで、 m はテストデータの総数である。

3. 検証実験

本稿では、動画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[16]を用いて実験を行った。Fig. 3にAvenueの例を示す。Avenueは16clipの学習データ、21clipのテストデータからなる。図に示すように、定点の監視カメラ画像を収録しており、正常データは通常の速度で歩行する様子、異常データは走る、荷物を投げる等の異常な様子が収録されている。このAvenueに対して、Frame-levelのReceiver Operating Characteristic (ROC) 曲線に対するAUROCによる定量的評価を行った。表1にAvenueに対する結果を示す。なお、Ours only STNは map による重みづけをしない場合の結果である。AUROCを用いた評価により、提案手法の有効性を確認した。また、 map を融合することによる有効性も確認した。



Fig. 3 Avenue の例. 正常例 (上) と異常例 (下)

Table 1 Avenue の結果

	Ours	Ours only STN	STAN[3]	Future frame prediction[7]	SRNN[1]
AUROC↑	0.897	0.868	0.872	0.851	0.817

4. 結論

本稿では、動画画像を効率的にモデル化するため、フレーム予測型の時空間敵対的生成ネットワークを提案した。敵対的学習の不安定性を改善するため各種安定化手法を用い、U-Net Discriminatorの出力を融合することで効率的な異常検知手法を確立した。Avenue 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。

文献

- [1] Weixin Luo, et al., “A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework,” ICCV, 2017.
- [2] Weixin Luo, et al., “Remembering history with convolutional lstm for anomaly detection,” ICME, 2017.
- [3] Sangmin Lee et al., “STAN: Spatio-Temporal Adversarial Networks for Abnormal Event Detection,” ICASSP, 2018.
- [4] 橋本 慧志, 工藤 謙一, 高橋 孝幸, 梅田 和昇, “GAN を活用した動画画像異常検知手法の構築と労働災害防止へ向けた応用の検討”, 精密工学会画像応用技術専門委員会サマーセミナー2020, 2020.
- [5] Mahdyar Ravanbakhsh, et al., “Abnormal Event Detection in Videos using Generative Adversarial Nets,” ICIP, 2017.
- [6] Hung Vu, et al., “Robust Anomaly Detection in Videos Using Multilevel Representations,” AAAI, 2019.
- [7] Wen Liu, et al., “Future Frame Prediction for Anomaly Detection – A New Baseline,” CVPR, 2018.
- [8] Mohammad Sabokrou, et al., “AVID: Adversarial Visual Irregularity Detection,” arXiv, 2018.
- [9] Muhammad Zaigham Zaheer, et al., “Old is Gold: Redefining the Adversarially Learned One-Class Classifier Training Paradigm,” CVPR 2020.
- [10] Edgar Schonfeld, et al., “A U-Net Based Discriminator for Generative Adversarial Networks,” CVPR, 2020.
- [11] Xingjian Shi, et al., “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” NeurIPS, 2015.
- [12] Sangdoo Yun et al., “Cutmix: Regularization strategy to train strong classifiers with localizable features,” ICCV, 2019.
- [13] Martin Heusel, et al., “GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium,” NeurIPS, 2017.
- [14] Han Zhang, et al., “Self-Attention Generative Adversarial Networks,” ICML, 2019.
- [15] Takeru Miyato, et al., “Spectral Normalization for Generative Adversarial Networks,” ICLR, 2018.
- [16] Cewu Lu, et al., “Abnormal Event Detection at 150 FPS in MATLAB,” ICCV, 2013.