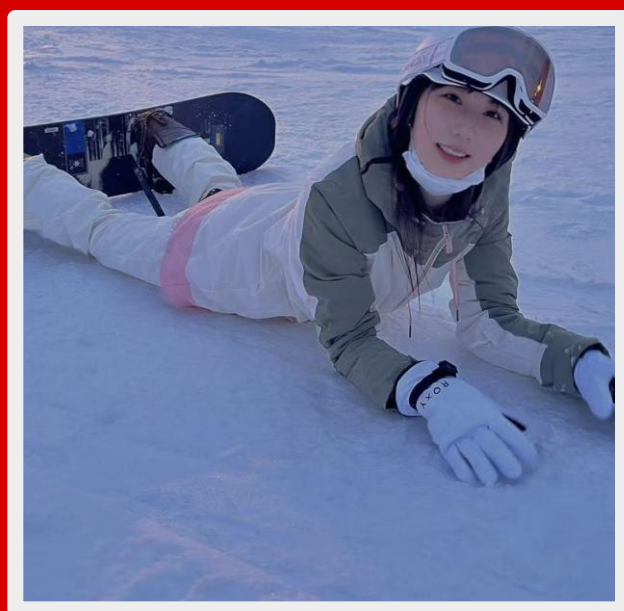


限定された教師情報下における 細粒度動画異常検知

Fine-Grained Video Anomaly Detection under Limited Supervision

王 汝沁
WANG Ruqin
オウ・ジョシン

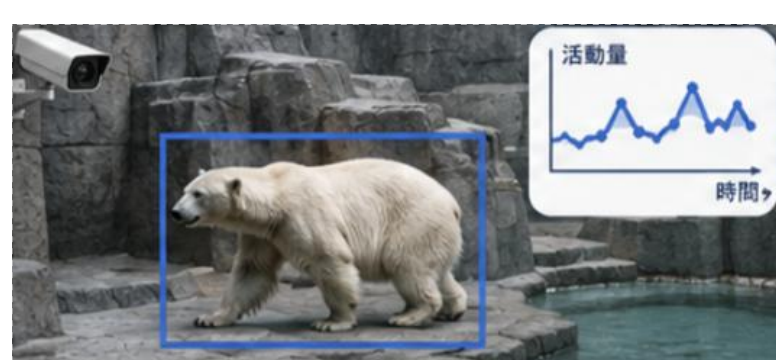
北海道大学 大学院情報科学院
自律系工学研究室



- 趣味：** ゲームやドライブ、写真撮影、新しい技術やデータ分析にも積極的に挑戦している。
- 強み：** 「深層学習の専門知識」＋「相手の気持ちに寄り添ったコミュニケーション力」
- 将来：** 深層学習モデルの研究開発に携わり、社会に役立つ技術を実装していきたい！

未来社会のあるべきかたち

- ◆ AIが長時間の監視映像から異常の兆候を早期に発見し、安全な都市・工場・交通を支援する。
- ◆ 「異常かどうか」だけでなく、「いつ・どこで」発生したのかを説明できる映像理解を実現する。
- ◆ 限られた教師情報から学習し、実環境で使える軽量・高効率な異常検知システムを目指す。



少ない学習データで実現する精密な映像異常検知

研究背景

異常検知（VAD）は製造、監視、防災、動物行動解析など幅広い分野で利用されている。しかし、実際の異常データは極めて少なく、収集やラベリングには多大なコストがかかるという課題がある。特に「いつ・どこで・なぜ」といった細かいレベルでの検出は困難であり、それを可能にする技術の確立が求められている。

研究目的

限られたラベル情報でも映像内の時空間的な異常をフレーム・ピクセルレベルで正確に特定できる手法を開発し、さらに異常データ生成の自動化を実現することを目指す。

研究結果

提案手法はベンチマークデータセットにおいて、既存の最先端（SOTA: State-of-the-Art）手法を上回り、精度向上と早期検出の両立を達成した。



図 1 VADタスクにおける典型的な異常イベントの例

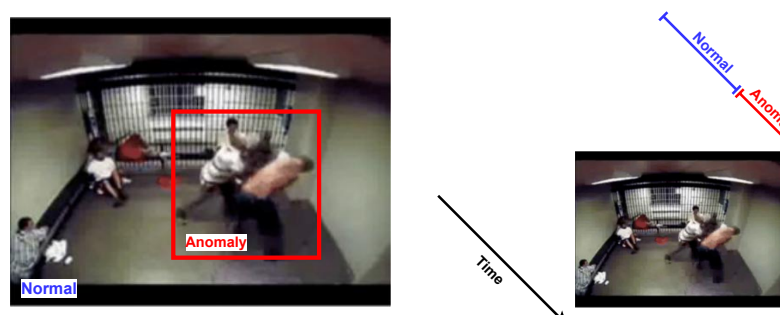


図 2 フレーム・ピクセル単位での時空間的なアノテーションの例

1. 弱ラベル環境での異常位置特定

研究概要：マルチインスタンス学習（MIL）

を用い、弱ラベル環境（映像全体にだけ「異常あり／なし」の情報があり、発生場所やタイミングは不明）で映像の細粒度（フレーム・ピクセルレベル）異常検知に挑戦した。

難点：細かい位置情報（細粒度ラベル）がなく、異常の

「いつ・どこ」を直接学習させることができなかった。

対策：映像を小さな領域に分け、それぞれの異常の可能性を推定することで、限られた情報から位置特定精度を高めた。

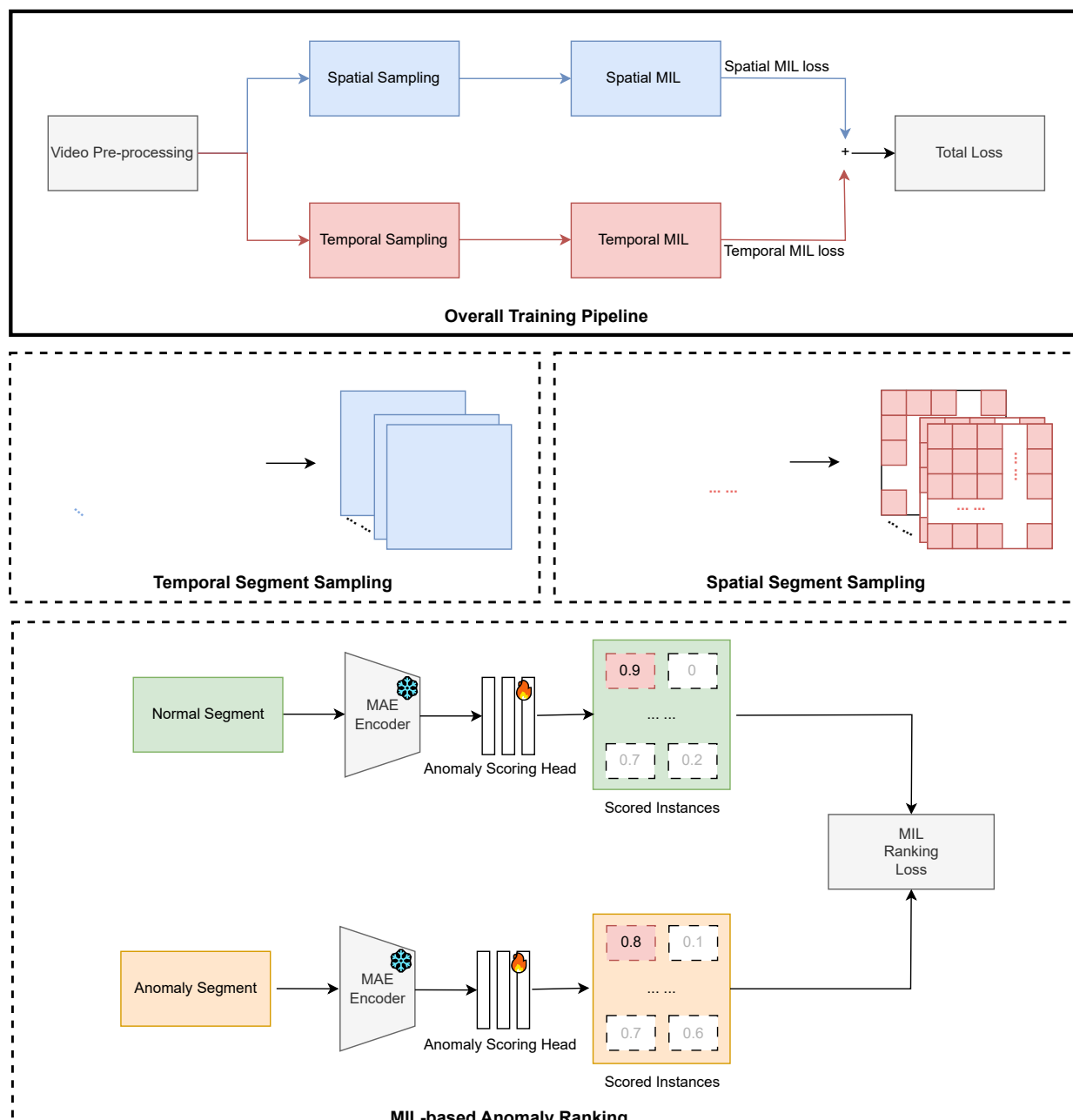


図 3 MILによる映像分割と異常スコア推定の流れ

2. 生成AIで自然な異常映像データセット作成

研究概要：生成AIを活用し、多様な異常映像とラベルを自動生成することで、データ不足を補った。

難点：既存の合成方法（カット＆ペースト型）は不自然さが残り、モデル学習に悪影響を与えていた。また、モデルごとに異なる特性に合わせたプロンプト設計や調整には、多くの試行が必要であった。

対策：大規模言語モデル（LLM）と視覚言語モデル（VLM）を組み合わせ、映像の文脈に自然に溶け込む異常シーンを生成しつつ、モデル特性に合わせたプロンプト最適化を行った。

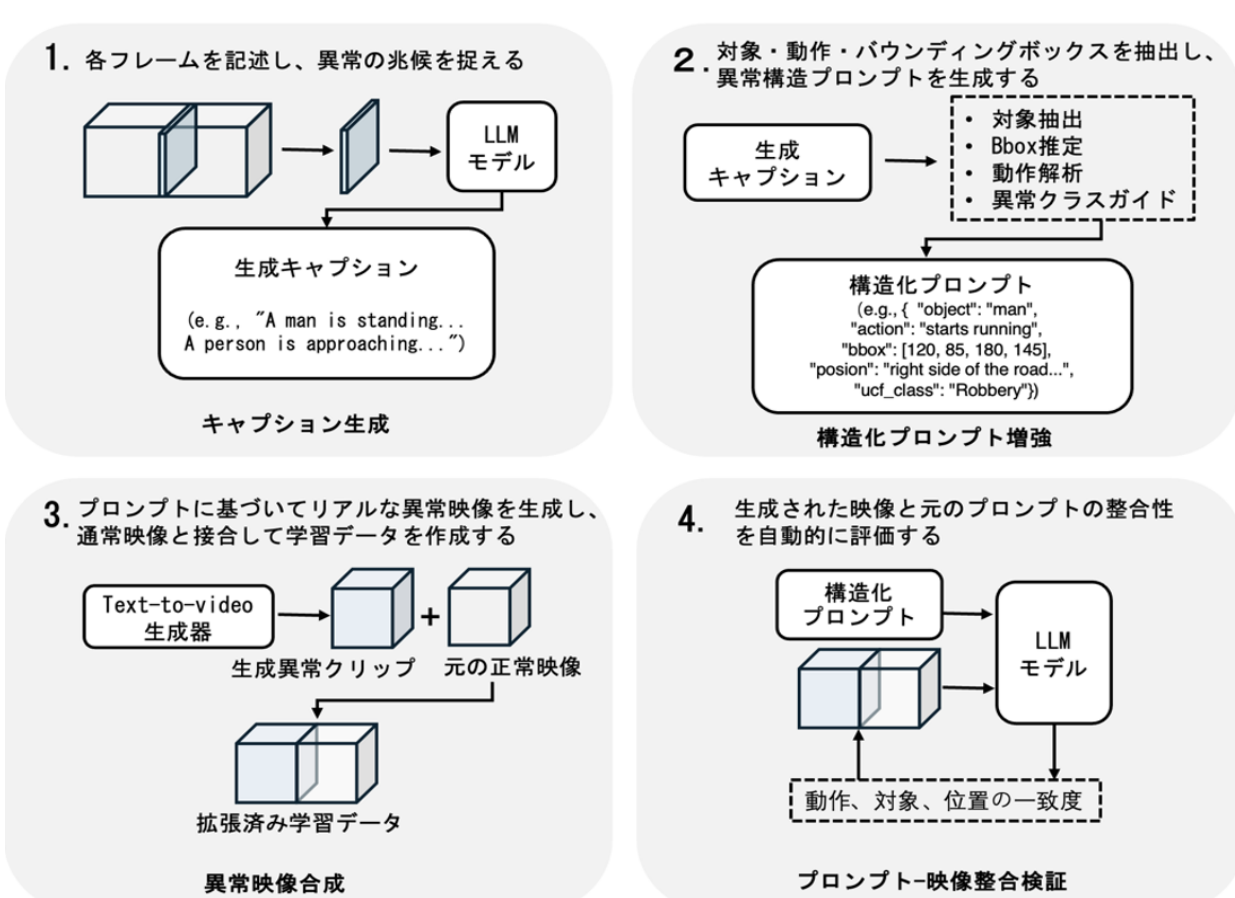


図 4 生成AIとプロンプト設計による異常映像生成プロセス

3. で高精度かつ効率的に検出

難点：高解像度映像での精密予測は計算コストが高く、リアルタイム処理が難しかった。

対策：空間・時間の二段階サンプリングを導入し、計算量を抑えつつ高精度を維持した。

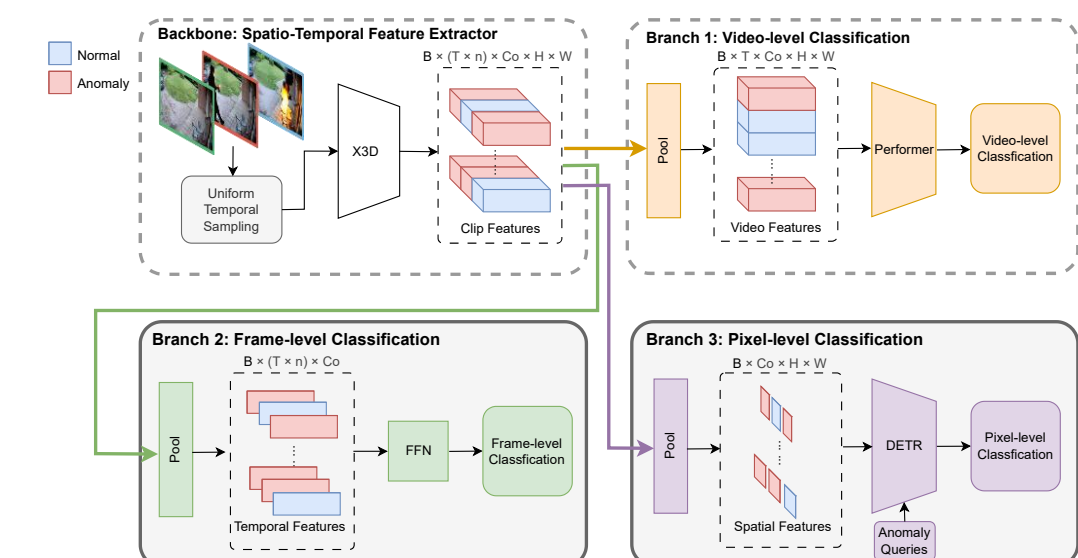


図 5 レーム・ピクセル・映像単位のマルチスケール予測モデル

可視化による有効性確認



結果1 MILベースの細粒度検出



結果2 マルチスケール予測

映像を領域ごとに分割し、異常範囲を特定。広域の異常検出に強み。

フレーム・ピクセル単位で異常を直接予測し、位置精度に優れる。