

異常検知モデルの出力データ活用によるクラウドワーカーの回答支援手法

Method for Assisting Crowdworkers' Answers Using Anomaly Detection Model Output Data

玉野 樹[†] 谷津 帆乃果[†] 板野 竜也[†] 小坂 隆浩[†]
Tatsuki Tamano Honoka Tanitsu Ryuya Itano Takahiro Koita

1. はじめに

近年、日本国内の犯罪検挙率は低い水準を保っている。令和4年度の犯罪検挙率は41.6%、令和5年度は38.3%となっており、直近の35年で一度も犯罪検挙率は50%以上を記録していない[1]。犯罪検挙率が低い水準を保っている原因として以下の2つが考えられる。1つ目は、監視カメラを有効的に活用できていない点である。近年、IoT技術の発展に伴い、監視カメラが低価格で手に入るようになり、公共の場だけでなく一般家庭にも設置されるようになった。しかし、人手不足のために監視カメラ映像を常時監視できず、犯罪を検知できていない。その結果、犯人を逮捕し損ねている。2つ目は、監視カメラ映像の事後確認では、犯罪の立証が困難である点である。逮捕には通常逮捕と現行犯逮捕が存在するが、通常逮捕の場合は逮捕状を発行するために十分な証拠を集める必要があり、検挙まで至らない可能性が高い[2]。検挙率の向上には、犯人を現行犯逮捕する必要がある。そのために監視カメラ映像を常時監視し、リアルタイムで犯罪を検知し通報する必要がある。本研究では、リアルタイムを数分以内と定義する。犯人を現行犯逮捕するには、犯行から30~40分以内に逮捕する必要がある[3]。また、犯人の捜索にある程度の時間を要することを想定すると、数分以内に犯罪を検知し通報することが望ましい。これら2つの原因から、犯罪検挙率を向上させるために、人手不足を解消し、リアルタイムで犯罪や事故などの異常を検知するシステムの構築が期待される。

Sultaniらは、人手不足を解消し、リアルタイムで犯罪や事故などの異常を検知するために、異常検知モデルを活用した手法を提案している[4, 5]。異常検知モデルは監視カメラ映像を入力として、0から1の実数値である異常スコアを出力とする。異常検知モデルの活用によって、監視カメラ映像を常時監視する人手なしで、リアルタイムで犯罪を検知することができる。しかし、モデルのみでの異常検知では十分な精度が得られていない。また、犯罪や事故などの異常が映っている学習用の映像データセットは少なく、モデルのみで信頼性のある異常検知システムを構築することは困難である。

板野らは、人手不足を解消し、リアルタイムで異常を

検知するために、異常検知モデルとクラウドソーシングを協調させた手法を提案している[6]。クラウドソーシングとはインターネットを介して不特定多数のクラウドワーカー（以下、「ワーカー」という）に仕事（タスク）を依頼する仕組みである[7]。クラウドソーシングによって、人手を雇用することなくインターネット上でワーカーに仕事を依頼することができ、人手不足を解消できる。板野らの手法では、異常検知モデル層とクラウドソーシング層の2層構造で犯罪検知を実現しており、まず、異常検知モデル層では深層学習を用いて異常スコアを算出する。そして、異常スコアがあらかじめ決められた閾値を超えた部分だけクラウドソーシングを使って異常検知を行う。板野らの手法は、モデルのみでの異常検知より精度が高く、ワーカーへの報酬やタスク設計を調整することで一定のリアルタイム性が得られることを示した[8, 9]。しかし、クラウドソーシング層での異常検知において、偽陽性率が高いという課題が存在する[10]。本研究で偽陽性とは、異常が発生していない映像に対して「異常が発生している」と判断してしまうことを指す。また、偽陰性は、異常が発生している映像に対して「異常が発生していない」と判断してしまうことを指す。この原因として、ワーカーのうち一定数スパムワーカー*が存在することや、タスク遂行能力の低いワーカー、ワーカーが持つバイアスなどが考えられる。ここで、タスク遂行能力が低いワーカーとは、集中力やモチベーションが欠如しているワーカーや、タスクの内容を正しく理解できないワーカーのことを指す。また、ワーカーの持つバイアスとは、異常が発生していない映像に対して、「異常が発生しているのではないか」と疑いの目を持ってしまう先入観のことである。偽陽性率が高いという課題に対して、ワーカーの偽陰性率を増加させずに、偽陽性率を抑えることで、誤検知の少ない信頼性の高いシステムの構築が期待される。

本研究では、クラウドソーシングによる異常検知において、偽陽性率が高いという課題を解決するために、異常検知モデルから出力されるデータを活用してワーカーのタスク回答を支援する手法を提案する。また、本研究の目的として、クラウドソーシングによる異常検知において、偽陰性率を上昇させずに、偽陽性率を低下させる

[†] 同志社大学大学院理工学研究科, Graduate School of Science and Engineering, Doshisha University

*効率良く報酬を稼ぐために、適当に回答するワーカー

ことを目的とする。偽陽性率が高い原因として、タスクを遂行する能力が低いワーカーや、ワーカーが持つバイアスが考えられるため、モデルから出力される異常スコアを参考に回答できるようタスクを設計する。タスクを遂行する能力が低いワーカーやバイアスを持つワーカーに対して、異常スコアの提示による回答支援を行うことで、ワーカーの負担軽減やバイアスの排除の効果が期待される。また、提案手法に対して評価実験を行い、偽陰性率を上昇させずに、偽陽性率を低下させられるか評価する。本研究では、具体的に以下の項目に貢献している。

- クラウドソーシングでの異常検知において、偽陽性率を低下させる新たな手法を提案した。
- クラウドソーシングでの異常検知において、異常検知モデルから出力されるデータの活用は偽陽性率低下に有効であることを示した。
- 出力データの活用によって、偽陽性率の低下だけでなく正解率、適合率の向上も示した。

2. 関連研究

本章では少ない人手かつリアルタイムで異常を検知するための手法として異常検知モデルによる手法とクラウドソーシングによる手法を紹介する。ここで、異常検知などの二値分類タスクの精度に関する一般的な評価指標である AUC について紹介する。AUC は百分率で 50 に近いほど低精度、100(もしくは 0) に近いほど高精度であることを示す。

本研究におけるセグメント動画とパート動画の定義について説明する。図 1 にセグメント動画とパート動画の例を示す。「セグメント動画」と「パート動画」はどちらも監視カメラ映像から分割した動画のことで、「セグメント動画」は映像の長さに関わらず k 等分 (k は 1 以上の整数) したもので、「パート動画」は、映像分割後の長さが t 秒 (t は 0 以上の整数) になるように分割したものである。本研究では、「セグメント動画」は、映像の長さに関わらず 32 等分したものの 1 つを表し、「パート動画」は、映像の長さを n 等分したものの 1 つを表す。ただし、 n は各パート動画の長さが 10 秒程度になるように調整した値である。 n は映像の長さに依存し、 f を監視カメラ映像の総フレーム数、 F を監視カメラ映像の FPS とし、式 (1) のように表される。

$$n = \frac{f}{10 \times F} + 0.5 \quad (1)$$

セグメント動画

(例) 全体の長さ: 100 秒, $k = 3$ の場合



パート動画

(例) 全体の長さ: 100 秒, $t = 20$ の場合



図 1: セグメント動画とパート動画の例

2.1 異常検知モデルによる異常検知手法

Sultani らは、監視カメラの映像から複数種類の異常を検知するモデルによる異常検知手法を提案している。監視カメラ映像を、長さに関わらず 32 本のセグメント動画に分割し、各セグメント動画に対して 4096 次元の C3D 特徴量を抽出する。C3D 特徴量とは、2 次元画像の幅、高さに加えて時間軸を追加した 3 次元畳み込みニューラルネットワークにより算出される特徴量であり、動きある行動の特徴を捉えることができる [11]。C3D 特徴量を抽出した後、3 層の全結合層ネットワークを介して、各セグメント動画に対して 0 から 1 の実数値をとる異常度が算出される。セグメント動画内の各フレームは同じ異常度として扱われる。テスト用データセットにはフレーム毎に 0 又は 1 の値をとる正解ラベルが与えられており、これを基にフレームレベルで異常度と正解ラベルが比較され、AUC が算出される。モデルの精度は UCF-Crime Dataset[†] という 13 種類の異常 (虐待, 逮捕, 放火, 暴行, 事故, 強盗, 爆発, 喧嘩, 泥棒, 銃撃, 窃盗, 万引き, 破壊行為) と正常な映像が含まれているデータセットを用いて評価している。Sultani らの手法を用いた場合、 $AUC = 75.41$ が得られる。Sultani らは、誤検知が生じる原因として、昆虫の飛翔によって動画の一部にノイズが乗る、あるいは覆い隠されてしまう点を挙げており、これにより偽陽性が生じる。このように、深層学習モデルによる手法は設置環境に左右されることがある。また、人混み等の集団行動を認識することができない点も誤検知の要因として挙げている。

Zhou らは、Sultani らの手法をベースとした異常検知モデルによる異常検知手法を提案している。Sultani らの手法と異なる点は C3D 特徴量ではなく、I3D 特徴量を採用している点である。I3D 特徴量とは、2 次元畳み込みニューラルネットワーク (2D-CNN) を時間軸に拡張し、3 次元畳み込みを用いることで算出される特徴量であり、空間情報と時間情報を同時に捉えることができる。

[†]<https://www.crcv.ucf.edu/projects/real-world/>

これにより、C3D よりも複雑な動作やシーンの変化をより詳細に捉えることができる [12]. Zhou らの手法を用いた場合、UCF-Crime Dataset に対して、 $AUC = 86.97$ が得られる。

異常検知モデルは監視カメラ映像を入力として、0 から 1 の実数値である異常スコアを出力とする。また、異常検知モデルの活用によって、監視カメラ映像を常時監視する人手なしでリアルタイムで犯罪を検知することができる。しかし、モデルのみでの異常検知では十分な精度が得られていない。また、犯罪や事故などの異常が映っている学習用の映像データセットは少なく、モデルのみで信頼性のある異常検知システムを構築することは困難である。

2.2 クラウドソーシングによる異常検知手法

クラウドソーシングによる異常検知は事前学習を必要とせず、計算機では抽出できない特徴量を捉え、高い精度が得られる。板野らは、人間の高度な認識能力を活用するクラウドソーシングによる異常検知手法を提案した [8, 9]. 板野らの研究では 10 秒程度のパート動画に対して、万引きが起こっているか否かを判別してもらうタスクを作成し、1パート動画当たり約 10 件の回答を収集して平均を取り、パート動画毎に異常スコアを算出している [9]. 異常スコアが得られたら、パート動画レベルで異常度と正解ラベルが比較され、 AUC が算出される。この手法では、UCF-Crime Dataset に対して、 $AUC = 99.70$ が得られている。クラウドソーシングによる手法では、パート動画レベルの AUC を算出しており、 AUC 算出方法の点では、フレームレベルで AUC を算出する深層学習モデルによる手法と異なるが、異常認識にクラウドソーシングを用いることで高い精度が得られることが期待できる。

また、板野らは異常検知モデルとクラウドソーシングを協調させることで複数種類の異常を検知するシステムを提案している [6]. ここで、異常検知モデルは Sultani らのモデルを使用しており、異常検知モデルから出力される異常スコアによってワーカーに回答を求めるか判断する方式を採用している。板野らの手法では、すべて異常検知モデルで異常検知を行うよりも高い精度が得られ、すべてクラウドソーシングで異常検知を行うよりも低コストで異常検知できる可能性を示した。また、リアルタイム性に関しても、ワーカーへの報酬やタスク設計を調整することで一定のリアルタイム性が得られることを示した [8, 9]. しかし、クラウドソーシングでの異常検知において、偽陽性率が高いという課題が存在する [10]. 偽陽性とは、ワーカーが異常が発生していない映像に対して「異常が発生している」と判断してしまうことである。この原因として、不正を行うスパムワーカーやタス

ク遂行能力の低いワーカー、ワーカーが持つバイアスなどが考えられる。ワーカーの偽陰性率を上昇させず、偽陽性率を抑えることで、誤検知の少ない信頼性の高いシステムの構築が期待される。

クラウドソーシングではワーカーの品質を確保するためにさまざまな研究が行われている。例えば、小林らは画像分類タスクにおいて、ワーカーが他のワーカーの回答を参考にして、タスクに回答する手法を提案している [13]. 小林らの手法では、第 1 段階で通常通りタスクに回答し、第 2 段階で他のワーカーの回答を参考にしながら、第 1 段階での自身の回答を修正する機会を設けた。結果として、分類の精度は向上し、長期的なワーカーの成長にも貢献できる可能性を示した。小林らの研究で、ワーカーが参考情報をもとにタスクに回答することで品質が向上する可能性を示した。

板野らの手法では、クラウドソーシングでの異常検知において偽陽性率を低下させるための対策が講じられていない。小林らの手法は、画像分類タスクにおいてのみ精度向上の有効性が示されている。また、ワーカーが参考しているデータが他のワーカーの回答である。本研究では、クラウドソーシングでの異常検知において、偽陽性率を低下させる対策を講じる。そして、ワーカーへの参考情報の提示は、画像分類タスクではなく、異常検知タスクでも精度向上に有効か、ワーカーが参考にするデータが異常検知モデルの出力データでも有効か検証する。

3. 提案手法

本研究では、クラウドソーシングでの異常検知において、偽陰性率を上昇させずに、偽陽性率を低下させるために、異常検知モデルから出力されるデータを活用してワーカーの回答を支援する手法を提案する。提案手法の概要を図 2 に示す。提案手法において、異常検知モデルから出力されるデータとして異常スコアを用いる。異常スコアは 0 から 1 の実数値で表され、1 に近いほど異常が発生している可能性が高いと推定される。また、異常スコアは Zhou らが提案した異常検知モデルから算出したものであり、このモデルは UCF-Crime Dataset で $AUC = 86.97$ が得られている。

提案手法での犯罪検知のフローを説明する。まず前処理として、監視カメラ映像を 10 秒程度に切り取ったパート動画に分割する。そして、パート動画を異常検知モデルに入力し、異常スコアを取得する。異常スコアは 16 フレームごとに 1 つモデルから算出するように設定されている。全ての異常スコアを対象のパート動画に割り当て、ワーカーは異常スコアが割り当てられたパート動画の異常の有無を判別する。ワーカーは異常が含まれてい

るか否かの二値で回答するため、「異常あり」と判断した場合は1、「異常なし」と判断した場合は0に変換し、複数ワーカーの回答の平均値を算出する。これを「クラウド異常スコア」とする。クラウド異常スコアが予め定められた閾値 α 以下の場合、異常なしと判別され、そのパート動画の認識は終了する。一方で、クラウド異常度が α より大きい場合、アラートを発生させる。

提案手法によってワーカーは異常スコアを参考に犯罪の有無を判断できる。小林らの研究 [13] では、参照するデータとして、他のワーカーの回答を使用している。提案手法では、参照するデータとして、異常スコアを使用するが、Zhou らが提案したモデルが $AUC = 86.97$ と、ある程度高い精度が示されている。そのため、ワーカーが異常スコアを参考に回答することで精度が向上するか検証する意義があると考えられる。板野らの手法において、タスク遂行能力の低いワーカーや、ワーカーが持つバイアスによって、偽陽性率が高くなっていると考えられる。提案手法では、タスクを遂行する能力が低いワーカーやバイアスを持つワーカーに対して、異常スコアの提示による回答支援を行うことで、ワーカーの負担軽減や、ワーカーが持つ「異常が発生しているのではないか」という先入観を排除できる可能性がある。その結果、ワーカーの誤認を防ぎ、偽陽性率を抑えられると考えられる。しかし、異常スコア (0~1) が正解ラベル (異常=1, 正常=0) と大きく異なっている場合、ワーカーの回答が異常スコアに依存してしまい、正解率が低下してしまう可能性がある。例えば、正常な映像に対して高い異常スコアが得られた場合、ワーカーの回答は異常スコアの値に依存してしまい正解率が低くなってしまう可能性がある。評価実験では、提案手法による偽陽性率の評価だけでなく、異常スコアが誤っている場合に、どの程度の正解率が保証されるか評価する。

4. 評価実験

4.1 実験方法

本研究では、実験 1, 2 で異常検知モデルの出力データの活用が、クラウドソーシングによる異常検知の精度にどのような影響を与えるか評価する。実験 1 では、板野らの手法 (以下、「既存手法」という) と提案手法の結果を比較して、提案手法が偽陰性率を上昇させずに、偽陽性率を低下させられるか評価する。実験 1 の評価指標として、偽陽性率、偽陰性率、正解率、適合率、再現率、 AUC を用いる。実験 2 では、パート動画に対する異常スコアが誤っていた場合でも、どの程度の正解率が保証されるか評価する。データセットは、既存手法に従い、UCF-Crime Dataset を使用して実験を行う。UCF-

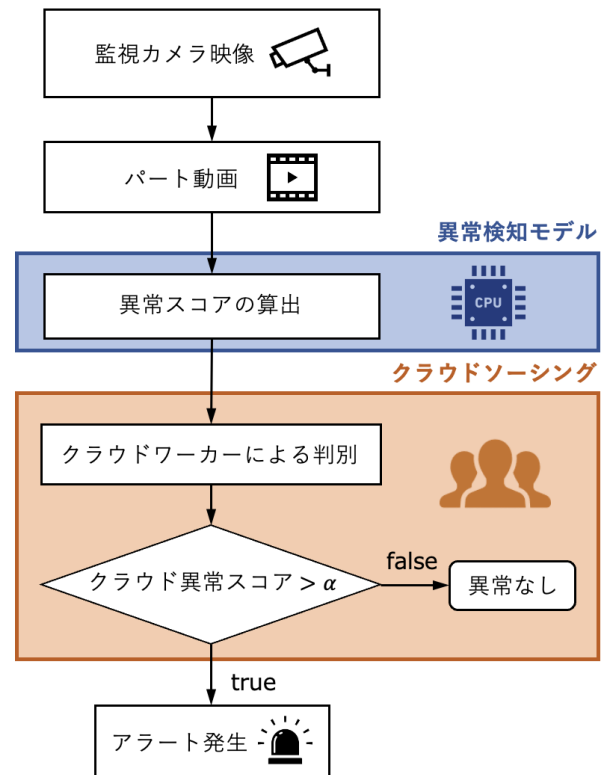


図 2: 提案手法の概要

Crime Dataset は、950 本の正常な映像、及び 950 本の異常な映像で構成されており、異常な映像には、公共安全に大きな影響を及ぼす主要な 13 種類の異常 (虐待、逮捕、放火、暴行、事故、強盗、爆発、喧嘩、泥棒、銃撃、窃盗、万引き、破壊行為) のいずれかが含まれている。既存手法では、UCF-Crime Dataset から正常な映像 24 本、異常な映像 26 本の計 50 本を試験用の映像としているが、実験 1 では、正常な映像 13 本、正常な映像 13 本を試験用の映像とした。各映像は長さが異なるため、ワーカーの負荷が偏らないように、既存手法と同様に全動画を 10 秒程度のパート動画に分割し、合計 86 本のパート動画が得られた。実験 2 では、正常な映像と異常な映像からそれぞれ 1 本のパート動画を選択し、それぞれのパート動画に人工的に誤った異常スコアを割り当てた。正常なパート動画には限りなく 1 に近い異常スコアを割り当て、異常なパート動画には限りなく 0 に近い異常スコアを人工的に割り当てた。

クラウドソーシングプラットフォームとして、Amazon Mechanical Turk (AMT)[§]を使用する。AMT はマーケットプレイス型のクラウドソーシングプラットフォームであり、ワーカーは公募されているタスクの中から自由に選択してタスクに回答する。AMT では常時 2,000 人以

[‡]正解率=正解したタスク数/全てのタスク数

[§]<https://www.mturk.com/>

上のワーカーが待機しており、他のクラウドソーシングプラットフォームと比較して一定レベルのリアルタイム性を確保することができる。ワーカーに与えるタスクは、文字や動画などを HTML や CSS で記述して自由に配置できる。

実験 1, 2 で異常スコアを算出するモデルは、Zhou らのモデル [12] を使用する。このモデルは、既存手法で用いられていた異常検知モデルよりも AUC が高く、映像の複雑な動作やシーンの変化をより詳細に捉えることができる。

4.2 実験 1

実験 1 では、既存手法と提案手法の結果を比較し、提案手法が偽陰性率を上昇させずに、偽陽性率を低下させられるか評価する。また、提案手法によって正解率、適合率、再現率にどのような影響を与えるのかを評価し、ROC 分析を用いて既存手法と提案手法の AUC を比較する。

ここで、既存手法と提案手法のタスク構成について説明する。既存手法は、画面上部にパート動画が表示されており、「映像内で異常が発生しているか回答せよ」という指示を記載している。ワーカーはそのパート動画に対して YES か NO で回答する構成になっている。スパムワーカーに対しては、「早すぎる回答は受け付けられない」といった旨の注意書きを行うことで最小限の対策を行った。また、異常の定義については不明瞭である。例えば、爆発といっても爆発の瞬間のみを異常と捉えるのか、爆発後の煙が出ている間も異常と捉えるのかで結果が大きく変化すると考えられる。既存手法では、異常がはっきりとは映っておらず、疑わしいのみ場合は NO と回答するように説明して、異常に対する曖昧さを取り除いている。

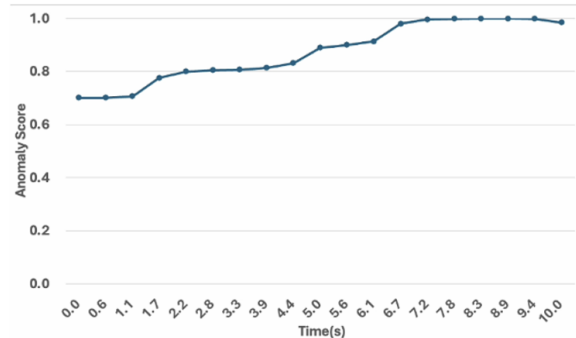
提案手法のタスク構成は、「説明フェーズ」、「回答フェーズ」の 2 つに分かれている。説明フェーズでは、タスクの内容を理解してもらうために設定しており、モデルについての簡単な説明、異常スコアについての説明、回答サンプルを提示している。モデルについての簡単な説明として、モデルの精度やモデルは必ず正しい異常スコアを算出するとは限らないことを示す。異常スコアについての説明では、異常スコアが 1 に近ければ異常が発生している可能性が高いことを示す。図 3 は回答サンプル例である。回答サンプルは、2 種類用意しており、1 つ目は、異常な映像に対して高い異常スコアを示しているケース、2 つ目は、正常な映像に対して低い異常スコアを示しているケースである。異常スコアは、16 フレームごとに 1 つの異常スコアを算出しており、すべての異常スコアをグラフにプロットしている。また、パート動画全体の異常スコアの平均値も算出している。回答フェーズでは、ワーカーに異常検知モデルから出力された異常スコアを

Sample Answer 1

Surveillance camera video



Anomaly scores(0~1)



Average anomaly score: **0.873625549**
Correct answer: **YES**

図 3: 回答サンプル例

参考にパート動画内の異常の有無を判別してもらおう。図 4 は回答フェーズの内容である。回答フェーズでは、「映像内で異常が発生しているか回答せよ」という指示とともに、異常スコアを参考にすることを指示している。説明フェーズの回答サンプル時と同様に、パート動画、16 フレームごとの異常スコア、異常スコアの平均値を提示し、異常の有無を YES か NO で回答してもらおう。提案手法でも既存手法と同様に、スパムワーカーへの最小限の対策と、異常の定義を行っている。また、異常がはっきりとは映っておらず、疑わしいのみ場合は NO と回答するように説明して、異常に対する曖昧さを取り除いている。

実験 1 では、延べ 1720 名のワーカーにタスクを依頼し、全 86 本の各パート動画につきワーカーから 10 件の回答を集め、各回答を 0 又は 1 で表した際の平均値をクラウド異常スコアとした。ここで、パート動画内のフレームは同じ異常度とする。また、報酬の増加に伴って精度が向上する可能性がある [14] ため、本研究では、報酬は 1 タスクあたり 0.01[\$] に固定した。そして、同様

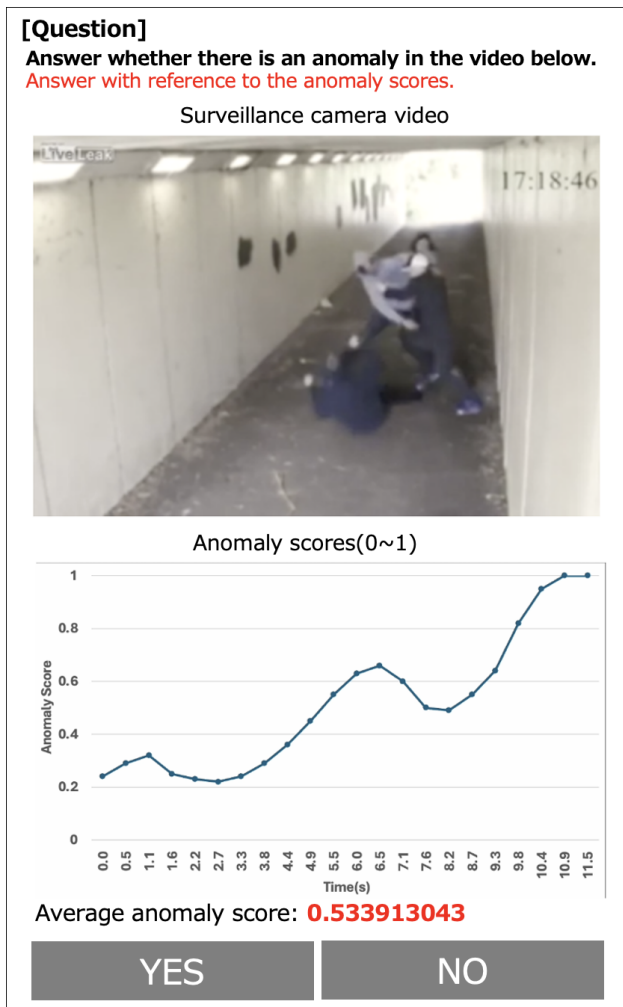


図 4: 回答フェーズ

の 86 本のパート動画に対して、既存手法で実験を行い、提案手法の結果を比較する。

4.3 実験 2

実験 2 では、パート動画に対する異常スコアが誤っていた場合、どの程度の正解率を保証されるか調査する。また、評価の基準として、パート動画に対して異常スコアが正しく付与された場合の正解率を同時に示す。依頼するタスクは、実験 1 で使用したタスクをベースとした、「異常_高」、「異常_低」、「正常_低」、「正常_高」の 4 種類である。例えば、「異常_高」のタスクの場合、異常なパート動画に対して限りなく 1 に近い異常スコアを人工的に割り当てている。また、「異常_低」のタスクの場合、異常なパート動画に対して限りなく 0 に近い異常スコアを人工的に割り当てている。「正常」なパート動画に対しても、同じように異常スコアを割り当てる。「異常_低」と「正常_高」のタスクは、パート動画に対して誤った異常スコアが割り当てられ、「異常_高」と「正常_低」のタス

クは、パート動画に対して正しい異常スコアが割り当てられる。「異常_高」と「正常_低」のタスクに対する結果は、「異常_低」と「正常_高」のタスクに対する結果を評価するための基準である。実験 2 では、「正常」と「異常」の 2 種類のパート動画が使用されており、UCF-Crime Dataset の正常な映像と異常な映像からそれぞれ 1 本のパート動画を選択し、それぞれのパート動画に人工的に異常スコアを割り当てた。異常スコアは、16 フレームごとに 1 つ算出しており、全ての異常スコアをパート動画に付与している。実験 2 では評価指標として、正解率を用いる。本実験で異常スコアが誤っていた場合、どの程度の正解率を得ることができるか評価する。例えば、クラウドソーシングで複数の回答を集計する際によく用いられる多数決の場合を考える。2 値分類タスクにおいて多数決は、回答の 50% 以上が正しい回答の場合システム全体として精度が低下することはない。そのため、異常検知モデルから出力される異常スコアが間違っていた場合でも、正解率が 50% 以上であることが望まれる。

実験 2 では、延べ 1320 名のワーカーにタスクを依頼し、「異常_高」、「異常_低」、「正常_低」、「正常_高」のそれぞれのパターンで 330 人のワーカーにタスクを依頼した。信頼水準 95%、許容誤差 5% の場合で考え、依頼するワーカーの数を 330 人に設定している。また、報酬は実験 1 と同様に、1 タスクあたり 0.01[\$] に固定した。

4.4 実験結果

表 1 に実験 1 における、既存手法と提案手法の正解率、適合率、再現率、偽陽性率、偽陰性率を示す。表 1 では、閾値 α が 0.5 と 0.8 の場合の結果である。 $\alpha = 0.5$ の場合、既存手法の偽陽性率は **82.56%**、提案手法の偽陽性率は **1.16%** となった。偽陰性率は、既存手法、提案手法ともに **0.00%** となった。正解率、適合率、再現率に関しては、提案手法は既存手法より高い数値が得られた。 $\alpha = 0.8$ の場合、既存手法の偽陽性率は **13.95%**、提案手法の偽陽性率は **0.00%** となった。偽陰性率は、既存手法で **6.98%**、提案手法で **8.14%** となった。再現率に関しては、提案手法は既存手法よりわずかに低い数値となったが、正解率と適合率に関しては、 $\alpha = 0.5$ の場合と同様、提案手法は既存手法より高い数値が得られた。

図 5 に実験 1 における、既存手法と提案手法の ROC 曲線を示す。既存手法では、 $AUC = 88.70$ が得られた。一方、提案手法では、 $AUC = 99.97$ が得られた。

図 6 に実験 2 における各パターンの正解率を示す。正しい異常スコアが割り当てられている「異常_高」の正解率は、**91.21%**、「正常_低」の正解率は、**71.48%** となった。そして、誤った異常スコアを割り当てている「異常_低」の正解率は、**68.18%**、「正常_高」の正解率は、**57.58%** となった。

表 1: 評価指標ごとの比較

	$\alpha = 0.5$		$\alpha = 0.8$	
	既存手法	提案手法	既存手法	提案手法
正解率 (%)	58.72	99.42	89.54	95.93
適合率 (%)	54.78	98.85	86.96	100.00
再現率 (%)	100.00	100.00	93.02	91.86
偽陽性率 (%)	82.56	1.16	13.95	0.00
偽陰性率 (%)	0.00	0.00	6.98	8.14

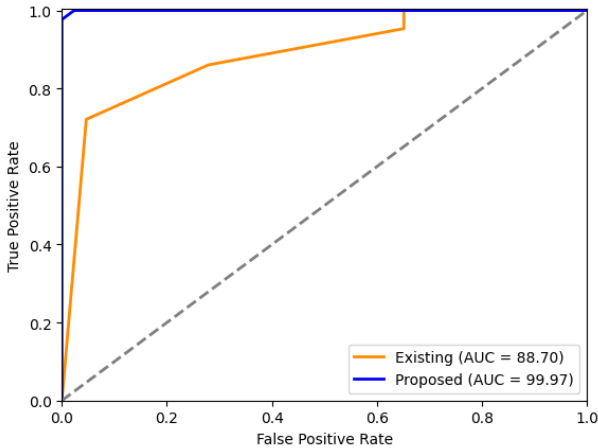


図 5: ROC 曲線 (橙線: 既存手法, 青色: 提案手法)

4.5 考察

まず, 表 1 に関して, 提案手法によって閾値 α が 0.5 の場合に, 偽陰性率を上昇させずに, 偽陽性率を低下させられていることが分かる. 閾値 α が 0.8 の場合でも, わずかに偽陰性率が上昇してしまっただが, 偽陽性率が低下させられている. この結果から提案手法は, クラウドソーシングでの異常検知において, 偽陰性率を上昇させずに, 偽陽性率を低下させるのに有効であることが明らかになった. また, 正解率と適合率に関しても, 閾値 α が 0.5, 0.8 の場合ともに, 既存手法より高い結果が得られた. よって, 提案手法によって偽陽性率を低下だけでなく, 正解率や適合率を向上させられることが明らかになった. また, $\alpha = 0.5$ の場合, 既存手法と提案手法で同じ再現率となったが, $\alpha = 0.8$ の場合, 提案手法は既存手法よりもわずかに低い値となった. この原因として, ワーカーに提示する異常スコア (0~1) が正解ラベル (異常=1, 正常=0) と大きく異なっている場合, ワーカーは異常スコアを参考にタスクに回答するため再現率が低くなってしまふことが考えられる. また, $\alpha = 0.5$ の場合, 既存手法での偽陽性率が 82.56% と比較的高い結果が得られた. この原因として, スパムワーカーやタスク遂行能力の低いワーカーの割合が高いことやワーカーが持つバイアスなどが考えられる.

図 5 に関して, 提案手法は既存手法の AUC を大きく

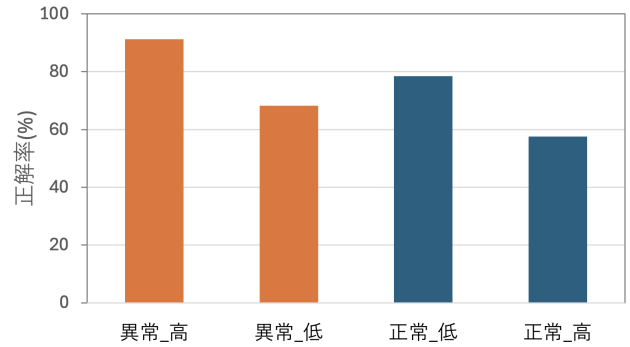


図 6: 正解率 (要素名は正解ラベル_異常スコアとする. 例えば, 要素名「異常_高」は, 異常な映像に対して高い異常スコアを付与したケース)

上回る結果となった. この結果から, ワーカーのタスク回答時に, 異常スコアを提示することで, タスク遂行能力の低いワーカーの回答支援や, ワーカーがバイアスを持っていたとしても異常スコアを参考にすることによりバイアスを排除できる可能性があることが分かる..

図 6 について, パート動画に対して誤った異常スコアを割り当てると高い正解率が得られないことが示された. また, 異常なパート動画よりも正常なパート動画の場合により低い正解率になってしまうことが分かった. しかし, 異常, 正常どちらのパート動画に誤った異常スコアを割り当てたとしても 50%以上の正解率が得られたため, 閾値の設定次第でシステム全体の精度を保つことができる.

提案手法によって, クラウドソーシングによる異常検知において偽陽性率を低下させ, 高い正解率と適合率が得られることが示された. しかし, 異常スコアが正解ラベルと大きく異なる値をとっていた場合に, 再現率が低下してしまう可能性がある. 異常検知モデルがパート動画に対して誤った異常スコアを割り当てた場合でも, 正解率は 50%以上得られるので, システム全体の精度を低下させない閾値 α の設定が重要となる.

5. まとめ

本研究では, クラウドソーシングでの異常検知において, 偽陽性率を低下させる新たな手法を提案し, 異常映像を含んだデータセットを用いて評価実験を行った. 実験の結果として, 提案手法は偽陽性率低下に有効であり正解率, 適合率の向上にも有効であることが明らかとなった. しかし, 異常スコアが正解ラベルと大きく異なる値をとっていた場合に, 再現率が低下してしまう可能性がある. 今後は, システム全体の精度を低下させない閾値の設定について検討していく必要がある.

参考文献

- [1] 警察庁, 令和 5 年の犯罪情勢,
https://www.npa.go.jp/publications/statistics/crime/situation/r5_report.pdf
- [2] 全国万引犯罪防止機構, 第 12 回全国小売業不明ロス・店舗セキュリティ実態調査分析報告書,
[https://www.manboukikou.jp/01/wp-content/uploads/2020/08/第 12 回全国小売業調査.pdf](https://www.manboukikou.jp/01/wp-content/uploads/2020/08/第12回全国小売業調査.pdf)
- [3] 裁判所, 最高裁判例,
https://www.courts.go.jp/app/hanrei_jp/detail2?id=51534
- [4] Waqas Sultani, Chen Chen, and Mubarak Shah : Real-world anomaly detection in surveillance videos, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.6479–6488 (2018)
- [5] Zhou Hang, Junqing Yu, and Wei Yang : Dual memory units with uncertainty regulation for weakly supervised video anomaly detection, Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 37, No. 3, pp.4020-4024 (2023)
- [6] Itano Ryuya, Tomoya Nohara, and Takahiro Koita : Video Anomaly Detection Method using Deep Learning Models and Crowd Workers, IEICE Technical Report, pp.1-6 (2022)
- [7] 鹿島久嗣, 梶野洸 : クラウドソーシングと機械学習, 人工知能学会誌, Vol.27, No. 4, pp.381-388 (2012)
- [8] 板野竜也, 目良侃太郎, 小坂隆浩 : クラウドソーシングにおける回答のリアルタイム性向上手法の検討, 信学技報. Vol.121, pp.46-51 (2021)
- [9] Ryuya Itano and Takahiro Koita. : Evaluation of the effectiveness of a crowdsourcing-based crime detection system, IEICE Communications Express, pp.607-611 (2022)
- [10] 玉野樹, 板野竜也, 小坂隆浩 : クラウドワーカーによる異常検知に向けた動画提供方式の評価, 電子情報通信学会学生会研究発表講演会, p.18, (2023)
- [11] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri : Learning spatiotemporal features with 3d convolutional networks, In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.4489–4497 (2015)
- [12] D. Koshti, S. Kamoji, N. Kalnad, S. Sreekumar and S. Bhujbal : Video Anomaly Detection using Inflated 3D Convolution Network, 2020 International Conference on Inventive Computation Technologies (ICICT), pp.729-733 (2020)
- [13] Kobayashi, M., Morita, H., Matsubara, M., Shimizu, N. and Morishima : An Empirical Study on Short- and Long-Term Effects of Self-Correction in Crowdsourced Microtasks, Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, pp.79-87 (2018)
- [14] Mason, Winter and Watts, Duncan J. : Financial Incentives and the Performance of Crowds, ACM SigKDD Explorations Newsletter, pp.100-108 (2009)