感情分析は"物語の中身"を見ているのか、"言葉の見た目"に左右されているのか

なぜ小説は難しいのか

口コミ

対応が最悪だった。二度と行かない。

言葉どおりの否定 → 判定しやすい

小説

それは実に見事な失敗だった。

皮肉・反語 → 言葉と意味がずれる
→ 判定が難しい

感情分析の結果を決めているのは"物語の内容"か、それとも"表現の形"か

切り分けができれば、
口コミ分析への応用可能性も判断できる

確かめる手順

パブリックドメインの小説を感情分析にかける

複数の手法を使い、手法ごとの偏りを避ける

↓

文脈は変えず、語尾と一人称だけを機械的に書き換える

↓

改変前と改変後の結果を比べる

出来事も筋も同じ。
変えるのは語尾と一人称だけ

改変前	改変後
私は門の下で雨やみを待っていた	俺は門の下で雨やみを待ってたんだ

結果が変わる → 表層の表現に左右されている

結果が変わらない → 物語の内容を見ている

いまは"書き換えなし"の測定まで完了。この結果が、あとで比べるための基準になる

今回やったこと

モデル：日本語BERTの感情分析モデル

対象：青空文庫の作品（芥川龍之介『羅生門』）

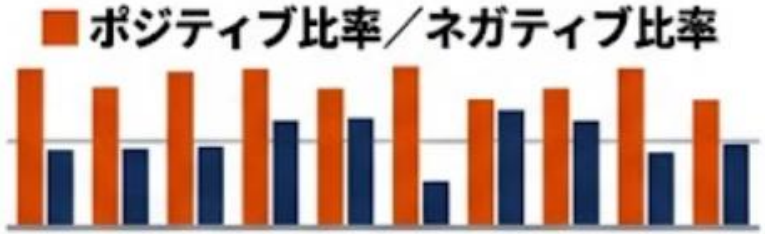
分類：ポジティブ／ニュートラル／ネガティブの3つ

文章の書き換えは、まだ行っていない

得られた2種類の結果

作品の進行順（冒頭→結末）

どの場面でどちらの感情が強まるかの移り変わりが見える



作品を10等分した区間

作品全体を数値に要約したもの。比較のときの物差しになる

これから比べる条件

	規則的に統一	ランダムに
一部だけ書き換え		
全体を書き換え		

書き換える範囲と規則性を変え、条件どうしの差を見る

比べる相手は、いつも“書き換えなし”の結果

この計画にまだ書かれていないこと

語尾と一人称を選んだ理由　結果がどちら向きに変わるかの予想

小説の感情分析に関する先行研究の整理（参考文献はBERTの原論文1件のみ）

顔の映像から目の開き具合を EARという数値にし、しきい値を下回った瞬間を『瞬き1回』と数える

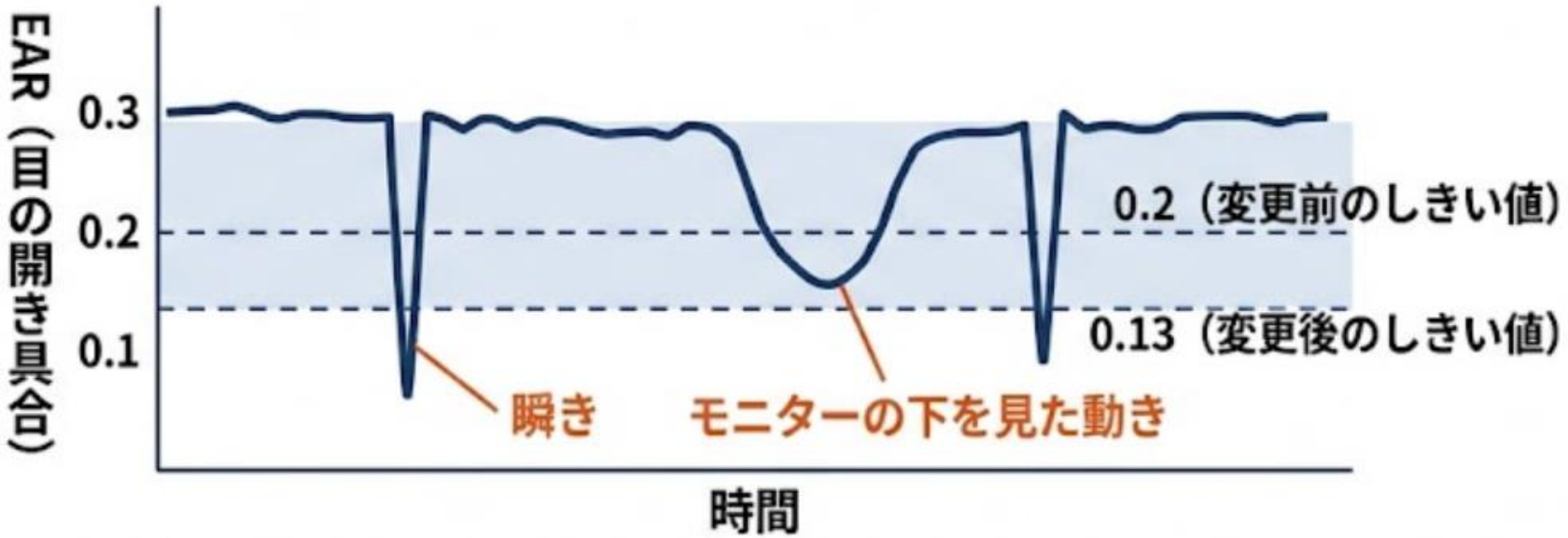
カメラで
顔を撮影

MediaPipeで目の
周囲の特徴点を検出

EARを計算
目の開き具合を数値化

開いた目＝EAR大／
閉じた目＝EAR小

EARが
しきい値を下回る
＝瞬き1回

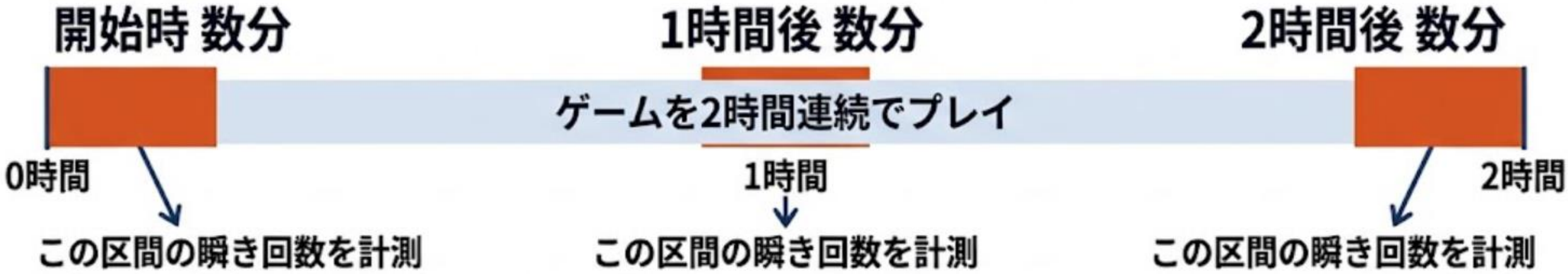


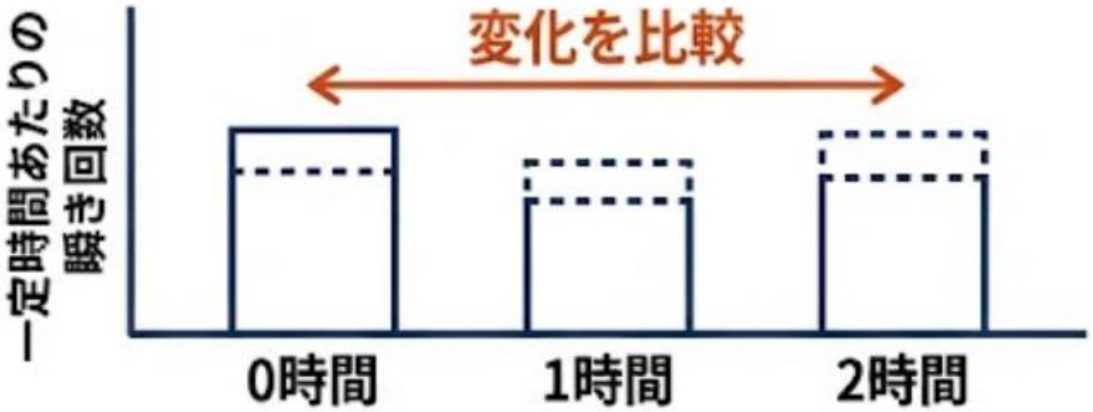
下方向を見ただけで瞬きと誤判定されたため、しきい値を0.2から0.13へ下げた

テキストファイルに自動保存

処理フレーム数
瞬き回数
経過時間
タイムスタンプ
左右それぞれのEAR値

2時間続けてプレイし、そのうち離れた3つの区間だけを数分ずつ計測して、瞬きの回数を比較する





経過時間そのものを条件として扱い、
同じ人の中での増減を見る

本資料に書かれていないこと

集中している／していないを分ける具体的な回数の基準
瞬きが多い側と少ない側のどちらが集中かという向き
2時間を通しで計測しない理由

そのため絶対的な基準ではなく、
同一の被験者の時間変化を評価する

照明やモニターの明るさを変え、
EAR値の変動と誤検出を比較する

検出精度を高め、集中を保てる
適切なプレイ時間を明らかにする

深層学習のフェイク判定は根拠が見えない。計算式の分かる数値と対応づけて説明できるかを調べる

フェイクとされる人物動画 10本（動画共有サイトから収集）

一定間隔でフレームを抜き出す

顔検出 → 顔だけを切り出す（顔クロップ）

A 深層学習モデルの判定

入力＝顔クロップ画像1枚だけ → 学習済み分類モデル（Vision Transformer / SigLIP、real か fake の2択） → 出力：Realの確率＋Fakeの確率＝100%

なぜそう判定したかは外から見えない

B 画像処理で計算する数値（画像特徴量）

動きの大きさ：連続する2枚で画素がどれだけ動いたかの平均（オプティカルフロー）

顔の揺れ：目・鼻・口などの特徴点のフレーム間の平均移動量

明るさの偏り：画面を左右・上下に分けたときの平均輝度の差

計算式が決まっているので、何を測っているか人が説明できる

同じフレームでAとBを並べ、相関係数（−1〜+1）を求める

連動が強い → モデルの判定を、説明できる数値の言葉に置き換えられる

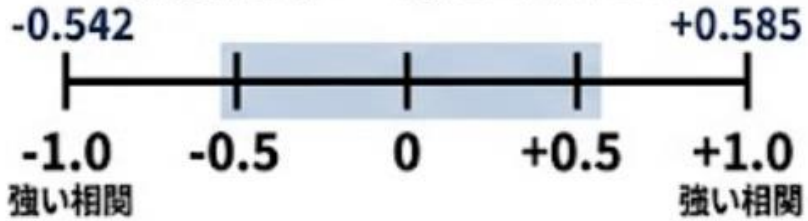
連動が弱い → モデルは別のものを見ている

相関係数は連動の度合いを示すだけで、原因と結果の関係を示すものではない

相関は弱かった。モデルは静止画1枚しか見ておらず、測った数値は時間の変化量だった

動画10本の相関係数

判定確率 × 動きの大きさ



判定確率 × 顔の揺れ

符号は動画ごとに正にも負にも入れ替わり、一貫した傾向なし
＝ 強い相関なし

理由の対比図

モデルが見ているもの		特徴量が測っているもの
顔クロップ画像1枚のみ	≠	隣り合うフレームの差
時間の情報は入力されない		時間方向のの変化量
1枚の中の質感・輪郭・不自然さで判定		動きの滑らかさ・揺れ

測っている対象がそもそも違う。
相関が弱いのは当然の結果と整合する

これから確かめること

モデルが1枚の顔画像の中で何に反応しているかは、まだ分かっていない

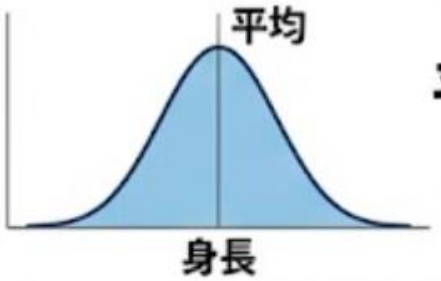
10本中4本はRealの確率が50%以上だった。これをこれを『失敗』と呼べるのは、10本すべてが偽物だという前提が正しい場合だけ

→ 動画はタグ検索で集めたもの。本当に生成AI由来かを確認してから結果を読み直す

異常の見本は学習しない。正常な見え方だけを覚え、そこからの外れ具合を数値にする。覚え方は二通り。

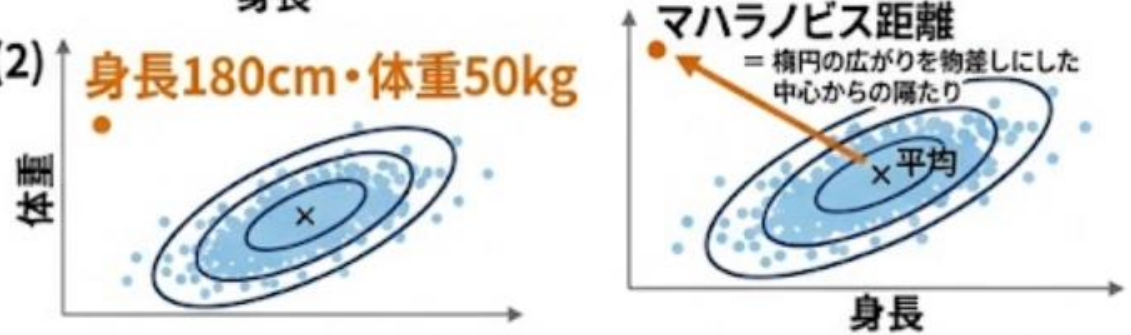
PaDiM ― 区画ごとに正常の分布をつくる

(1)



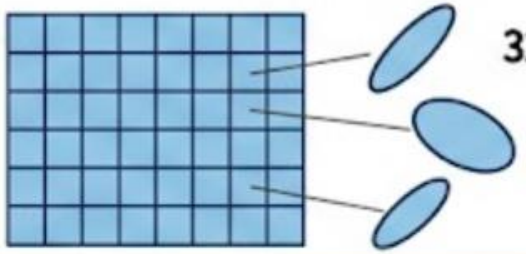
平均の付近に集中する山型
＝ガウス分布

(2)



量どうしの連動（共分散）まで表す＝多変量ガウス分布
一つずつ見れば普通でも、組合せとしては珍しい、が判定できる

(3)



32×32の小領域（パッチ）に区切る

場所ごとに写るものが違うので、
分布は画像全体で一つではなく
区画ごとに別々

検知時：その区画の見え方が、正常時のばらつき
何個分ずれているか＝異常度

PatchCore ― 正常の見本を代表だけ残す

(1)

正常画像 20枚

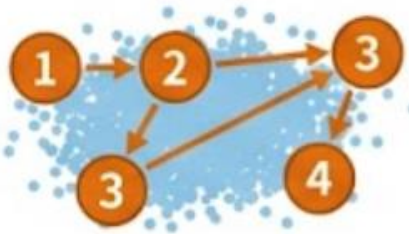
見本 20480個 → 多すぎるので一部だけ残す

(2)

無作為に
間引く

めったに現れない正常
な見え方が抜け落ちる
→後で異常と誤判定

貪欲法
（コアセット
選択）



すでに選んだどの
見本からも最も遠い
ものを次に選ぶ

その時点で最善に見える選択を確定し、後戻りしない＝貪欲法
少数でも正常な見え方の範囲を偏りなく覆える

検知時：最も似た見本との違いが大きいほど、
正常時に見たことのない見え方＝異常度が高い

色の見た目では判定しない。数値で判定し、しきい値を動かして精度を測る。

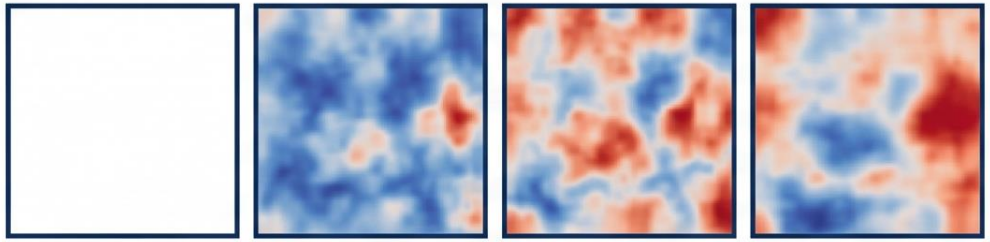
異常度が指す『異常』

正常画像と比べた、見え方の外れ具合だけを数値化

見つけたい外観の変化	同じく異常度が上がるもの（外乱）
•異物が写り込む •あるべき物が無い •位置や個数や大きさが変わる •表面に傷や汚れ	•照明が変わる •カメラがずれる

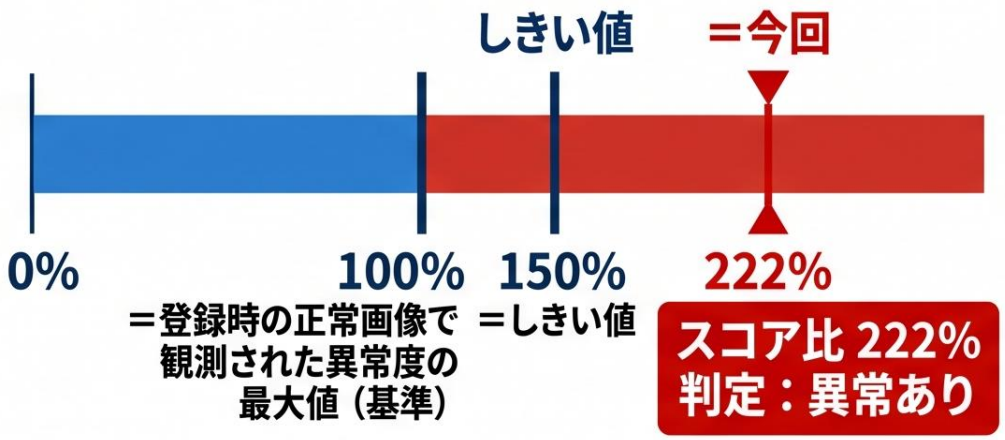
この両者を原理的に区別しない＝本手法の性質

結果の見方



ライブ映像　手法1　手法2　手法3

色はその画像の中での相対表示。最も低い所が青、最も高い所が赤
正常な画像でもどこかは赤くなる → 色だけでは判断できない



異常度の測り方と特徴を取り出すモデルが手法ごとに違うため。基準値も別々なので、手法どうしで数値の大小は比べない

検出精度の求め方

被写体・カメラ・照明を固定し、正常登録は一度だけ（やり直すと基準値が変わり比較不能）

↓

正常な状態と、異常を作った状態を撮影し、一枚ずつ正解を記録

↓

	しきい値を超えた	超えない
実際は異常	検出	見逃し
実際は正常	誤検知	正しい

↓

再現率＝見逃さなかった割合／適合率＝異常と判定したうち正しかった割合

しきい値を100%～300%まで動かして描く

3手法の比較、調整前後の比較に使う

実際に使うしきい値は、誤検知が許容範囲に収まる最小の値

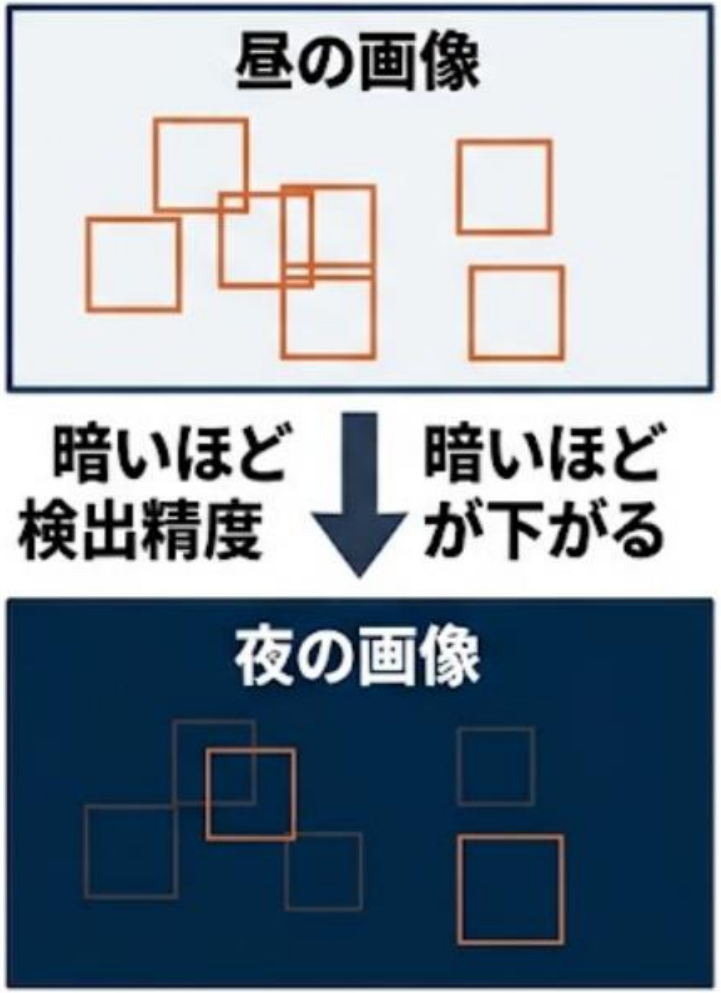
想定用途：定位置の部品の欠品、異物の混入、床の落下物や汚れ、表面の傷 ／ 対象外：人物や動いている物 ／ 明るさの変化は異常ではなく外乱として扱い、誤検知の測定に使う

YOLO系の物体検出AIで、映像の明るさを下げながら検出結果を比べた実験

実験の条件

- 使用モデル：YOLO系 (YOLOv12/YOLO26)
- 検出対象：COCO80クラスのみ、他は検出されない
- 信頼度閾値：**0.25** (下回る候補は検出に数えない)
- 前処理：ガンマ補正+CLAHE
- 入力：カメラまたは動画なのの各フレーム

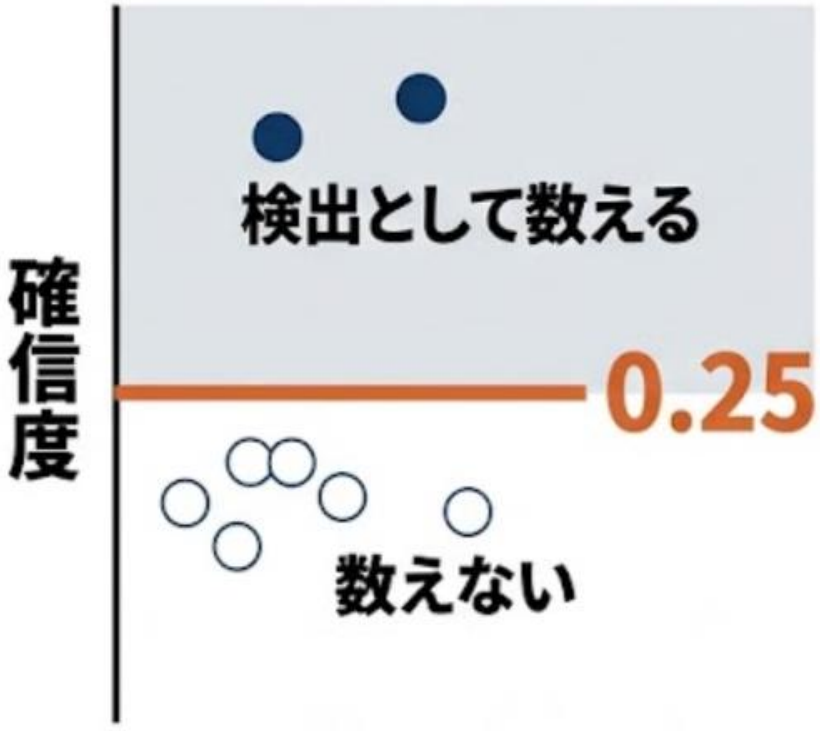
段階1：昼と夜の静止画(YOLOv12)



段階2：車の走行動画(YOLO26)

変えたもの：明るさ(ガンマ補正)とコントラスト(CLAHE)　見たもの：検出数と確信度

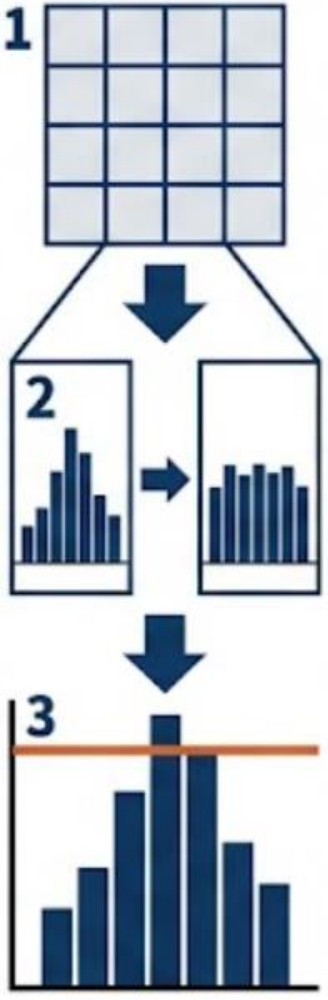
「『暗すぎる』は数値で決めた線ではなく、検出が0になった結果から後から決まる」



0.25以上の検出が一つも出ない
明るさ＝本資料が言う『暗すぎる』

前処理でコントラストを強調しても
検出数が戻らない領域がある

CLAHE（コントラスト制限付き適応ヒストグラム均等化）

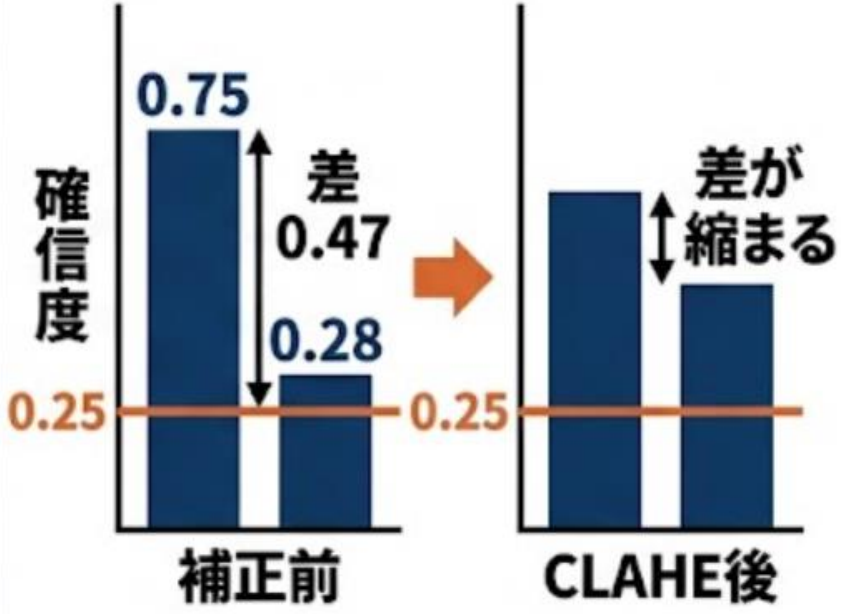


1 画像を小さな領域に分ける

2 領域ごとに明るさの分布をならす

3 上限を設け、強調しすぎとノイズ増幅を抑える

明るさを変えずにCLAHEだけ適用



検出数は2台のまま変わらない
（どちらも0.25を下回らない）

確信度のばらつきが均され、
安定して認識できる

本資料に記載がないこと：元より明るくした場合 / 輝度ヒストグラムによる解析 / 人間の見え方との比較