AI画像理解と3次元体験



資料中の図などは Google Nano Banana 2 を用いて作成

全体像と画面構成

教材動画（街・人物）をAIで解析し、モデルの出力値と計算の途中の値を見ながら、操作して確かめる





操作キー	機能
0：ホーム	1〜4：各章
Space：一時停止／再生	F：全画面
H：ヘルプの表示／非表示	Esc・右クリック：ヘルプを閉じる
Q：終了	

ヘルプを開くと動画は止まり、閉じると止めた所から続く

動画はコマを飛ばさずに1コマずつ解析する。
処理が追いつかない環境では、動画は本来より遅く再生される

第1章 3×3フィルタの積和

画像は輝度の数値の行列。周囲3×3画素の輝度と重み9個の積和が、出力の1画素になる

街の映像 → 320×180画素に縮小 → 輝度に変換

$Y = 0.299R + 0.587G + 0.114B$

周囲3×3の輝度

50	50	200
50	50	200
50	50	200

左が暗く右が明るい＝縦の輪郭

×

重み（縦エッジ）

-1	0	1
-2	0	2
-1	0	1

＝

輝度×重み

-50	0	200
-100	0	400
-50	0	200

→ 積和 600

出力の1画素

この積和を全画素で行うと、出力画像になる

重みの合計が0

縦エッジ　横エッジ　輪郭

出力：赤＝正、青＝負

重みの合計が1

ぼかし　シャープ

出力：0未満は0、255超は255として輝度で表示

入力画像をクリック：調べる画素を選ぶ

［最も強く反応した画素へ］：
出力の絶対値が最大の画素へ移動

CNNの畳み込み層も同じ積和を行う。重みは学習で決まる

第2章 物体検出としきい値

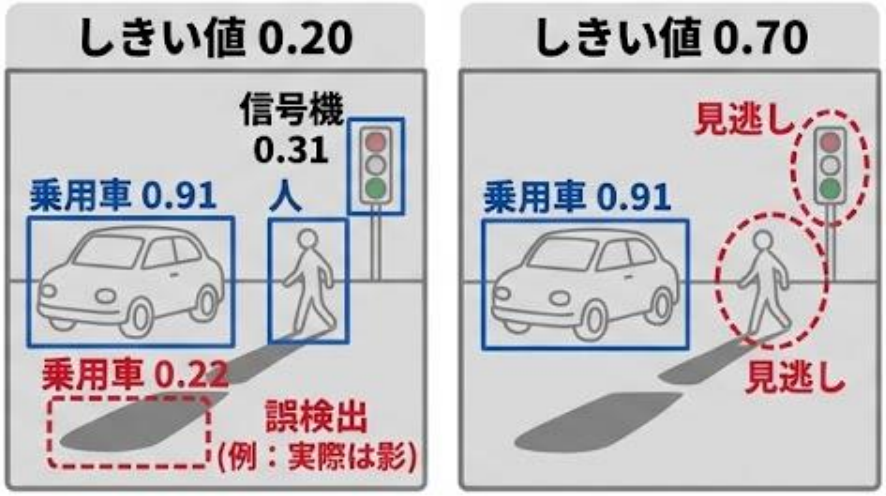
物体検出AIは、物体ごとに枠・クラス・スコアを出力する。
映像に表示するのは、スコアがしきい値以上の検出だけ

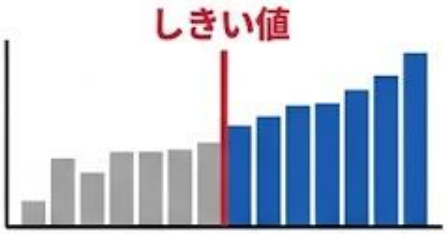
【1つの検出の中身】



領域分割：物体の領域も画素単位で出力する

しきい値による比較





スコア 0.05～1.0

赤線より右の検出だけを映像に表示

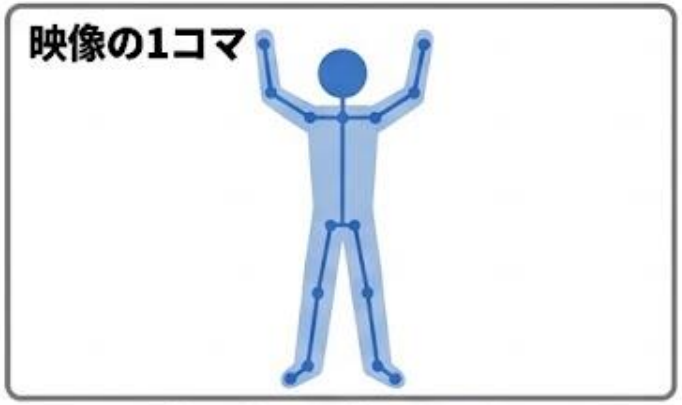
スライダー：しきい値 0.05～0.95
枠をクリック：スコア・枠の　位置と大きさ・領域の画素数

しきい値を下げる：見逃しが減り、誤検出が増えやすい ⟷ **しきい値を上げる：誤検出が減り、見逃しが増えやすい**

スコアは確信度の目安。正解である確率とは限らない
YOLO26はCOCOの80クラスで学習済み。80クラスに含まれない物体は検出できない

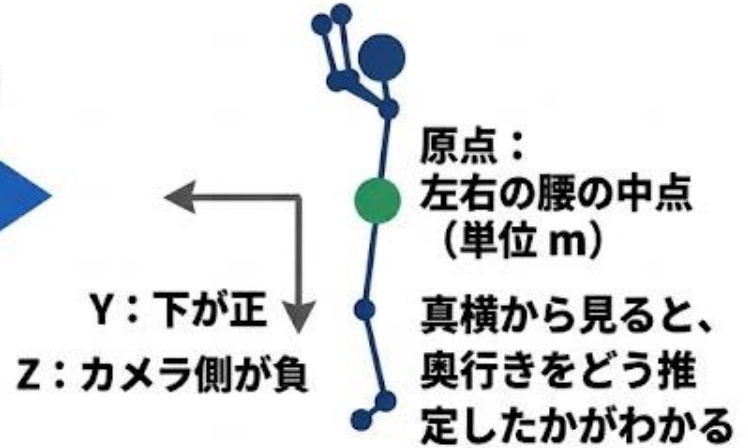
第3章 1台のカメラから人の3D姿勢

1台のカメラの映像から、1人の人物の33点の3D座標を推定する。
3D座標は、学習した人体の形にもとづく推定値

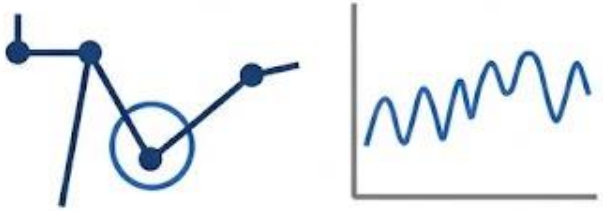


MediaPipe Pose Landmarker

33点を推定



関節の角度



ひじ・ひざの角度：3D座標から計算

直近150コマの推移（0〜180°）

visibility



visibility: その点が隠れずに写っているとモデルが推定する度合い（0〜1）

visibilityが0.5未満の点を含む角度は『参考値』と表示

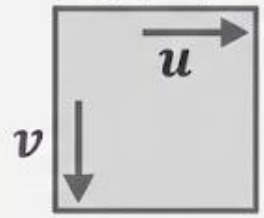
操作

3D骨格をドラッグ：回転
（4秒操作しないと自動で回転）
映像上の点をクリック：
点の名前・3D座標・visibility
スライダー：visibilityがこの値未満の点を表示しない
左右は、写っている人物自身から見た左右

第4章① 奥行き Z から3D座標へ

AIが画素ごとの奥行き Z を推定し、仮定した水平画角から3D座標 X・Y を計算する

画素の位置
(u, v)



→ **AIが奥行き Z [m] を推定**

Depth Anything V2

→ **3D座標を計算**

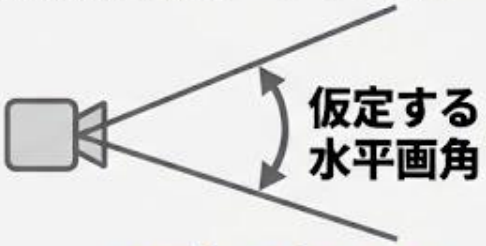
$$X = (u - c_x)\,Z / f$$
$$Y = (v - c_y)\,Z / f$$

c_x, c_y：画像の中心
X：右　Y：下　Z：前方

→ **3D点群**



焦点距離 f の求め方



$$f = \frac{(\text{画像の幅} / 2)}{\tan(\text{水平画角} / 2)}$$

水平画角を変える
↓
焦点距離 f が変わる
↓
X・Y・2点間の3D距離・路面からの高さが変わる

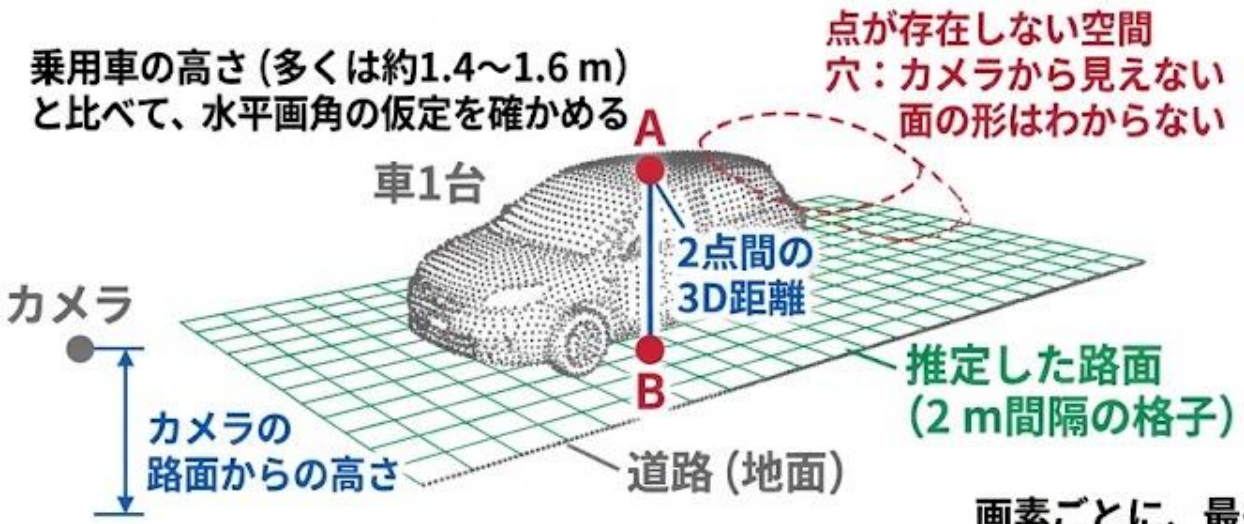
奥行き Z は変わらない
(AIが画像だけから推定する)

奥行き Z が推定上限の97%以上の点は除く

主に空。奥行き Z が不確かなため

第4章② 回す・測るで確かめる

3D点群を別の方向から見たり、2点間の3D距離を計算したりして、AIの推定と水平画角の仮定を確かめる



RANSACによる路面推定

画像の下半分・40 m以内の点を使う

↓

3点から平面の候補を作る
（路面の傾きは約45°以内と仮定）

↓

距離0.15 m未満の点が
最も多い候補を選ぶ

↓

その点群に主成分分析で
平面を当てはめ直す

画素ごとに、最も手前の点だけを描く（Zバッファ法）

ドラッグ：視点を回転 3点目をクリック：最も古い点と入れ替わる	2点をクリック：2点間の3D距離を計算 計測は一時停止してから行う
選んだ点は画像上の同じ位置に固定され、 動画が進むと別の物体を指す	色：実画像／路面からの高さ／奥行き Z ドリー：中央付近の大きさを保ったまま、 遠近感を変える
表示の焦点距離：表示だけを変える（3D座標は変わらない）	
[元の視点]：回転と表示の焦点距離を初期状態に戻す	

起動から動作までの流れ

インターネット接続が必要なのは初回だけ。取得したファイルは models フォルダに保存され、2回目以降はオフラインで動作する

起動前：コードの CONTACT に連絡先（メールアドレスまたは Webサイト）を書く → **初回起動**（インターネット接続が必要） → **8項目を取得**（AIモデル4種（奥行き推定モデルは3項目）と教材動画2本）

項目ごとに進捗バーで表示。
奥行き推定モデルの重みは約100 MB

→ **models フォルダに保存**（プログラムと同じ場所） → **[体験をはじめる]**

取得が終わると押せる

2回目以降はオフラインで動作

表示の解像度：通常（縦768画素）／4Kディスプレイ向け（縦2160画素）
取得中も選べる。画面の縦が2000画素以上なら、4K向けが初めから選ばれている
[models フォルダを開く]：いつでも押せる

数分待ってから［再試行］を押す
サーバーの利用制限（HTTP 429/503）のときは、自動で待って再試行する
教材動画を取得できないとき：ブラウザで保存し、person.webm・street.webm の名前で models フォルダに置く

第2章・第4章：NVIDIA GPUがあれば自動で使う

第3章：常にCPU

GPUがない環境：第2章・第4章の表示が遅くなる場合がある

処理時間は [AIの中身] 欄に表示

exe にした場合も、models フォルダは exe と同じ場所に作られる