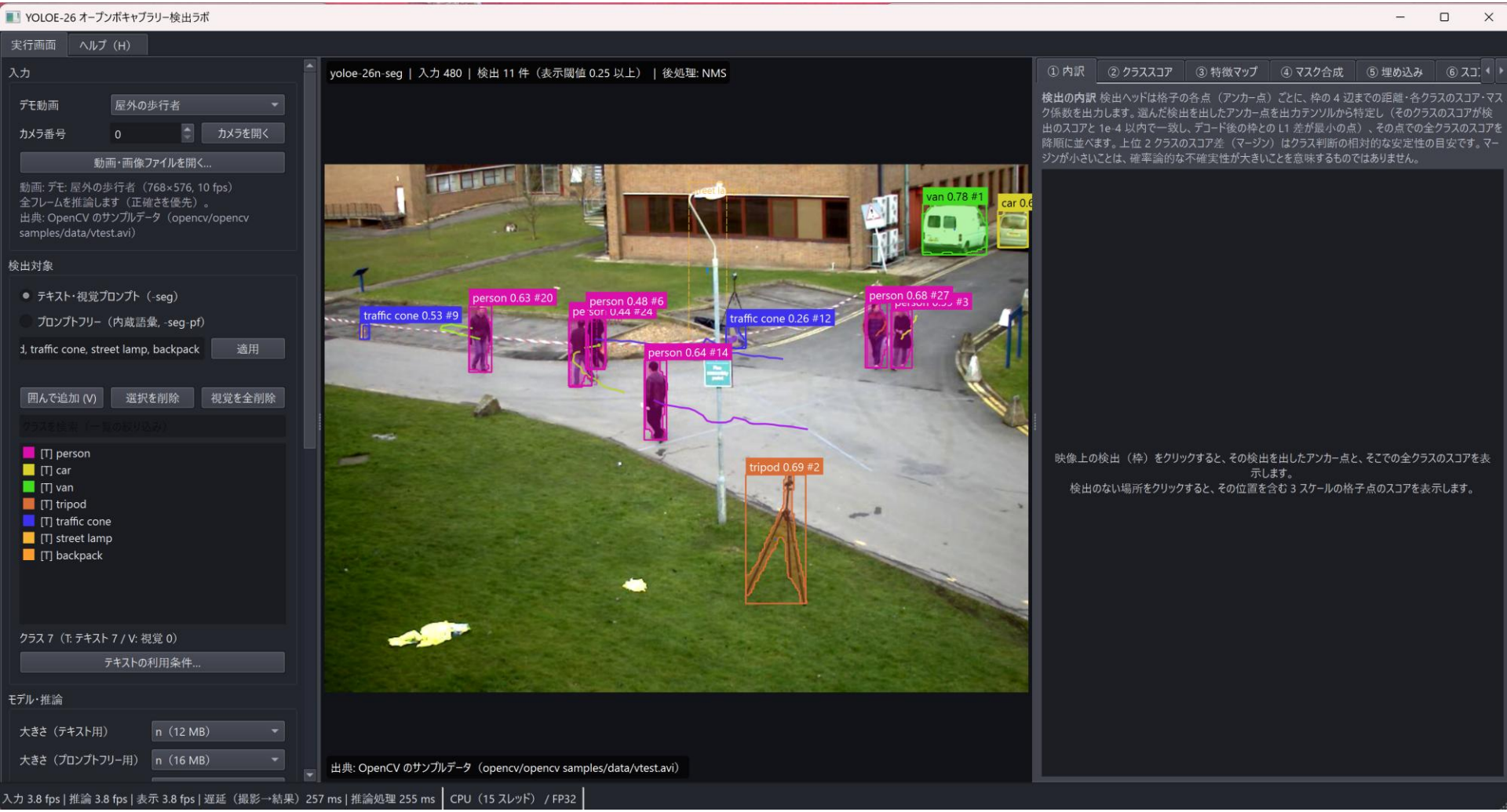
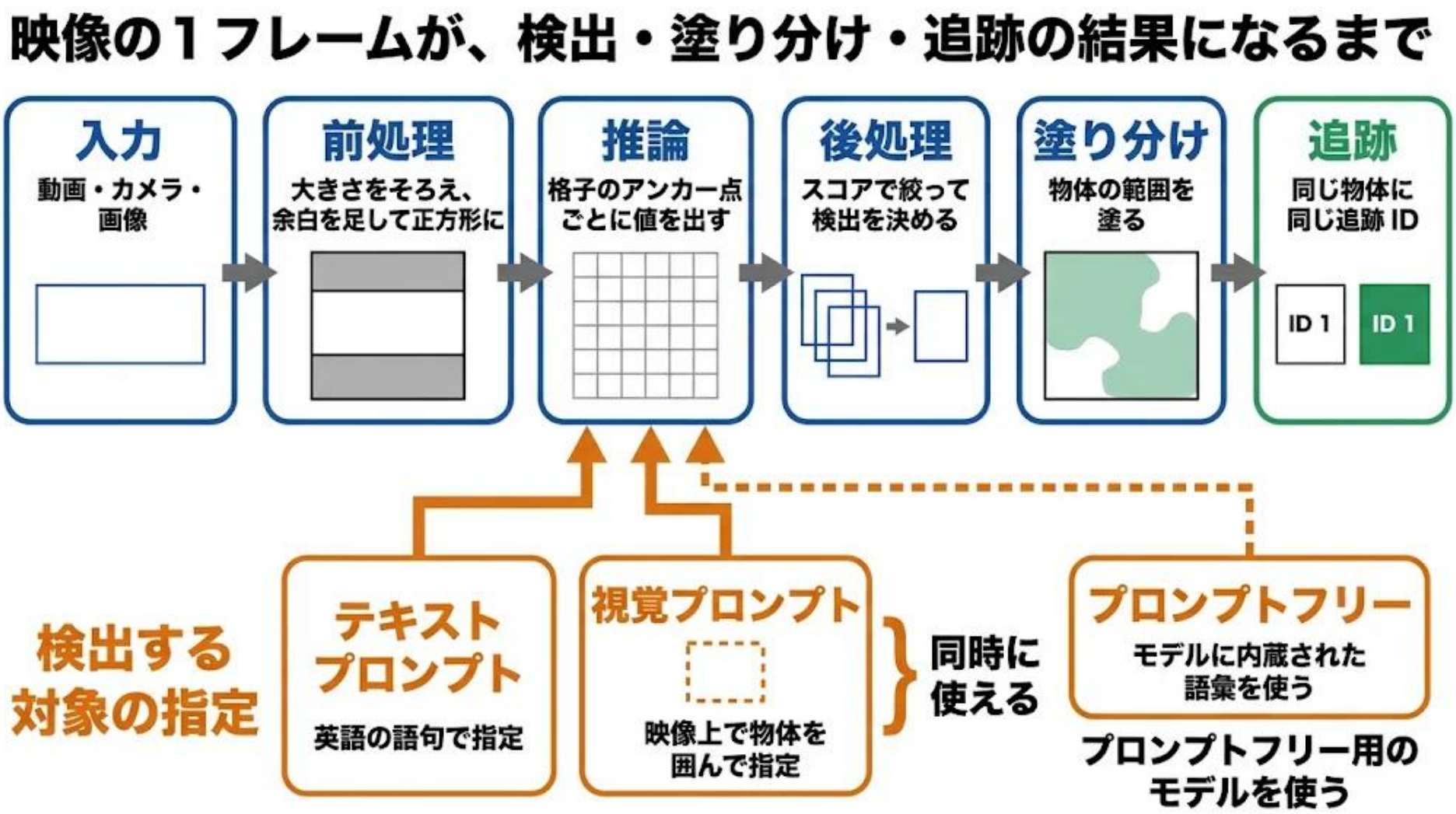
YOLOE-26 オープンボキャブラリー検出

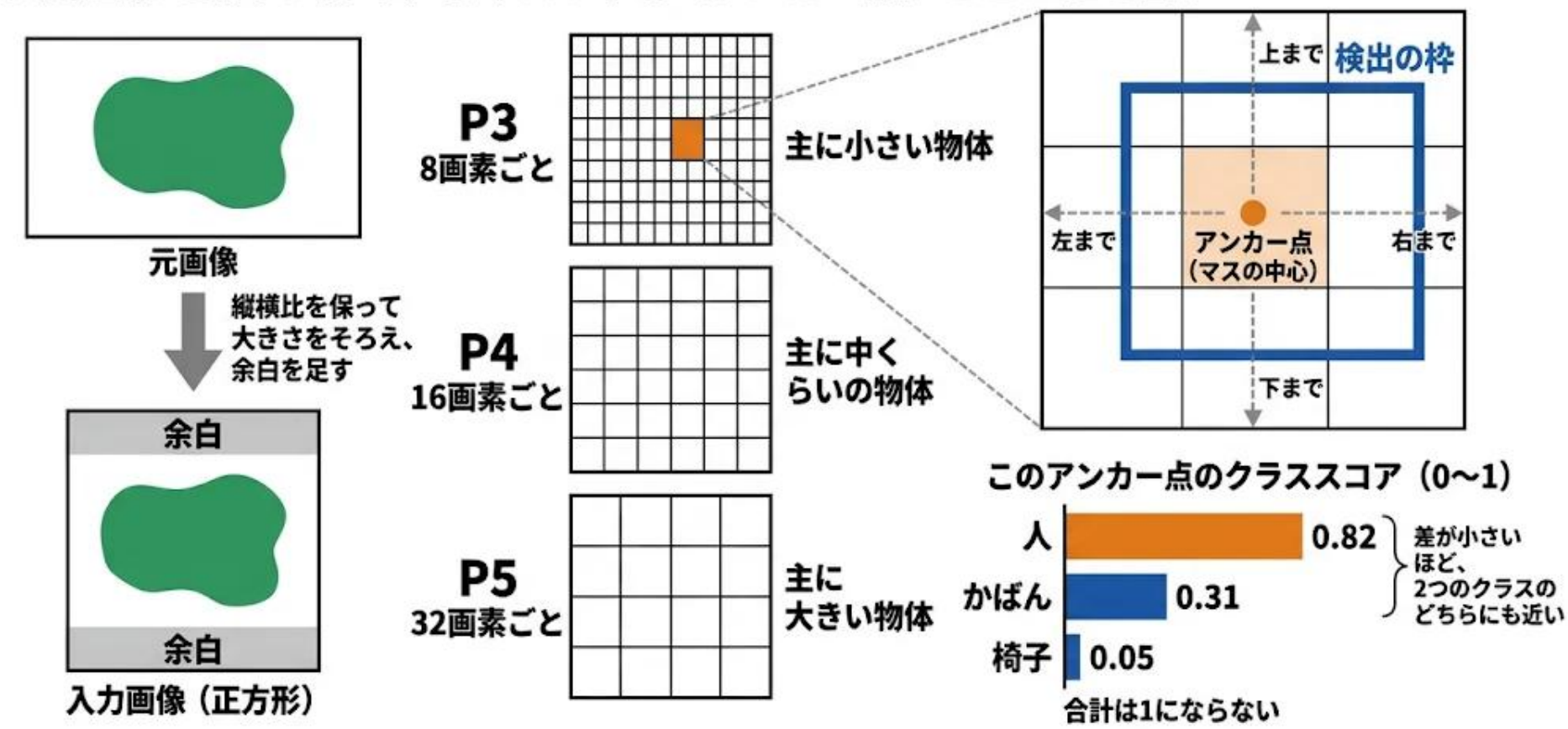


資料中の図などは Google Nano Banana 2 を用いて作成

処理の全体の流れ


映像の１フレームが、検出・塗り分け・追跡の結果になるまで
入力
動画・カメラ・画像
前処理
大きさをそろえ、余白を足して正方形に
推論
格子のアンカー点ごとに値を出す
後処理
スコアで絞って検出を決める
塗り分け
物体の範囲を塗る
追跡
同じ物体に同じ追跡 ID
ID 1
ID 1
検出する対象の指定
テキストプロンプト
英語の語句で指定
視覚プロンプト
映像上で物体を囲んで指定
同時に使える
プロンプトフリー
モデルに内蔵された語彙を使う
プロンプトフリー用のモデルを使う

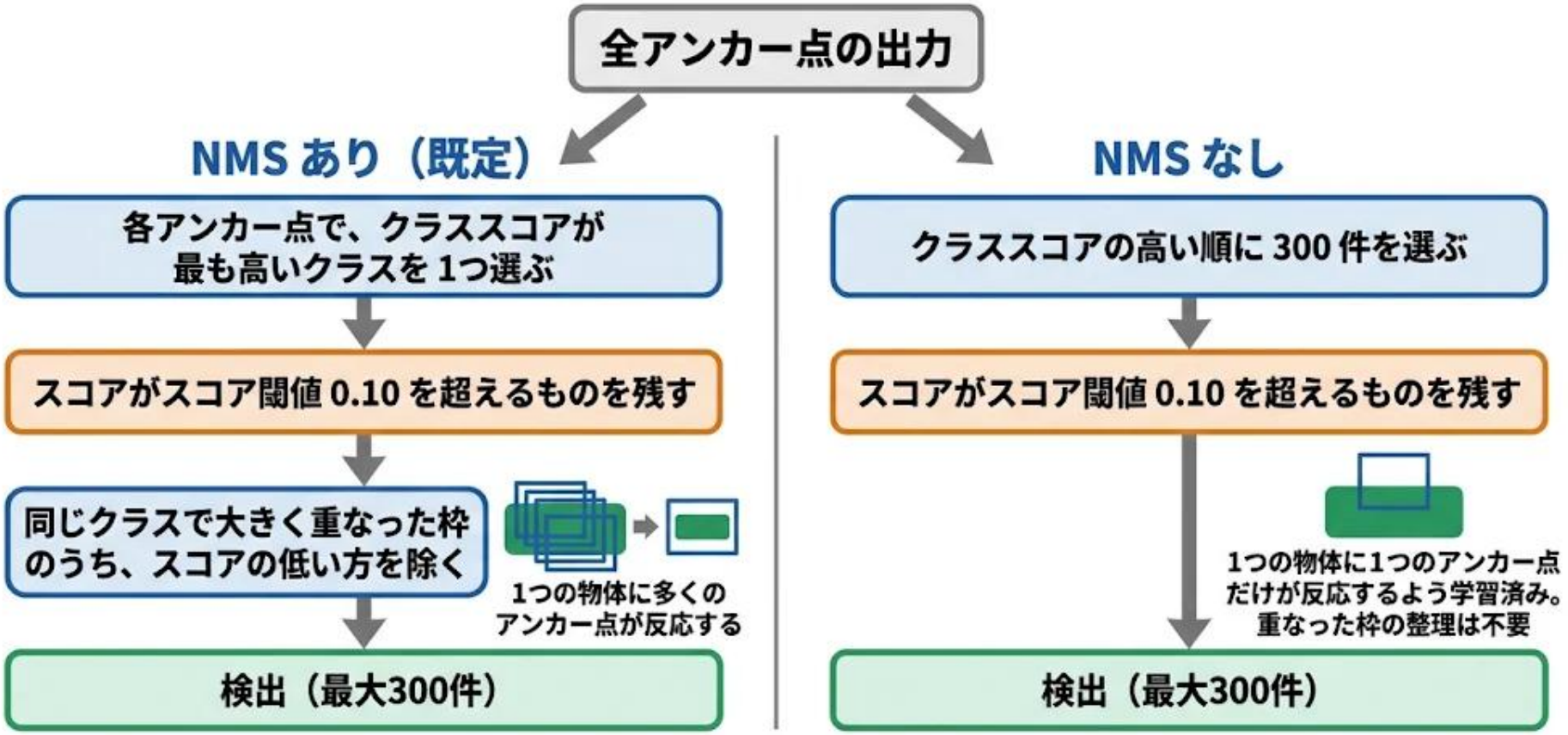
座標・格子・アンカー点と、各点の出力



各アンカー点の出力：枠の4辺までの距離/クラススコア/マスク係数32個

後処理（スコア閾値と重なりの整理）

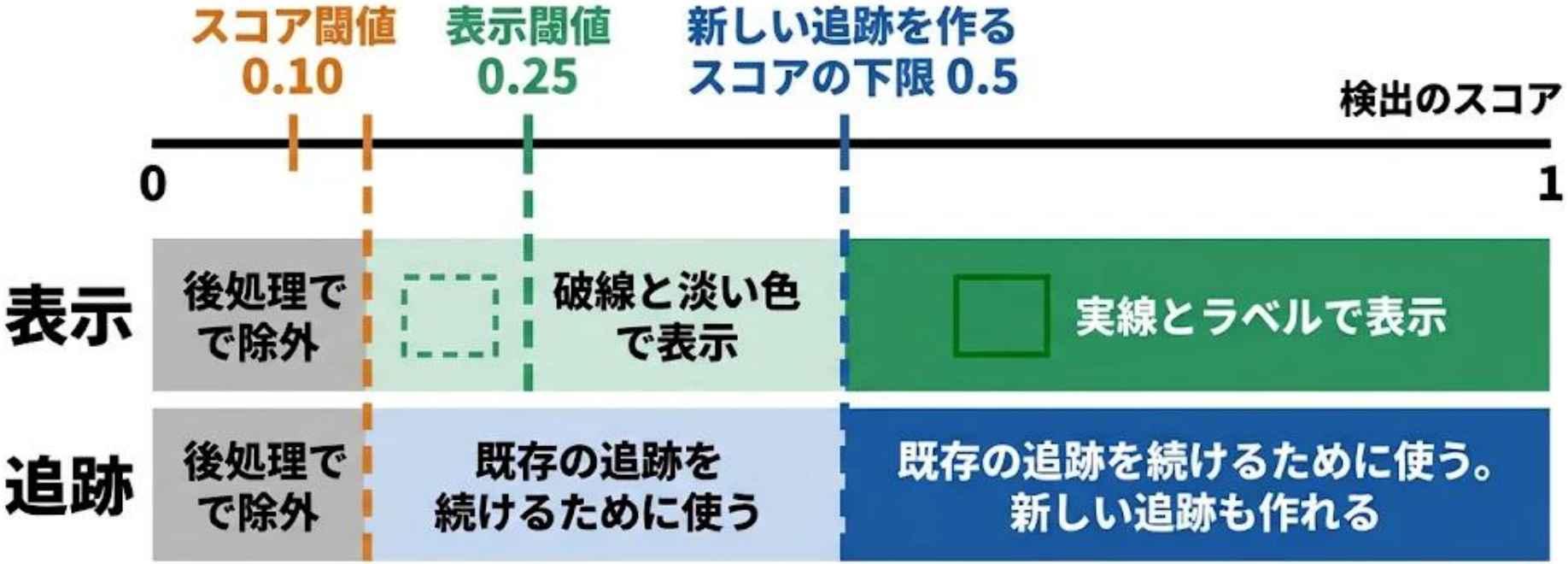
後処理の方式は２つ。どちらも、全アンカー点の出力から検出を最大 300 件決める



方式を切り替えると推論し直す。一時停止中なら、同じフレームで比べられる

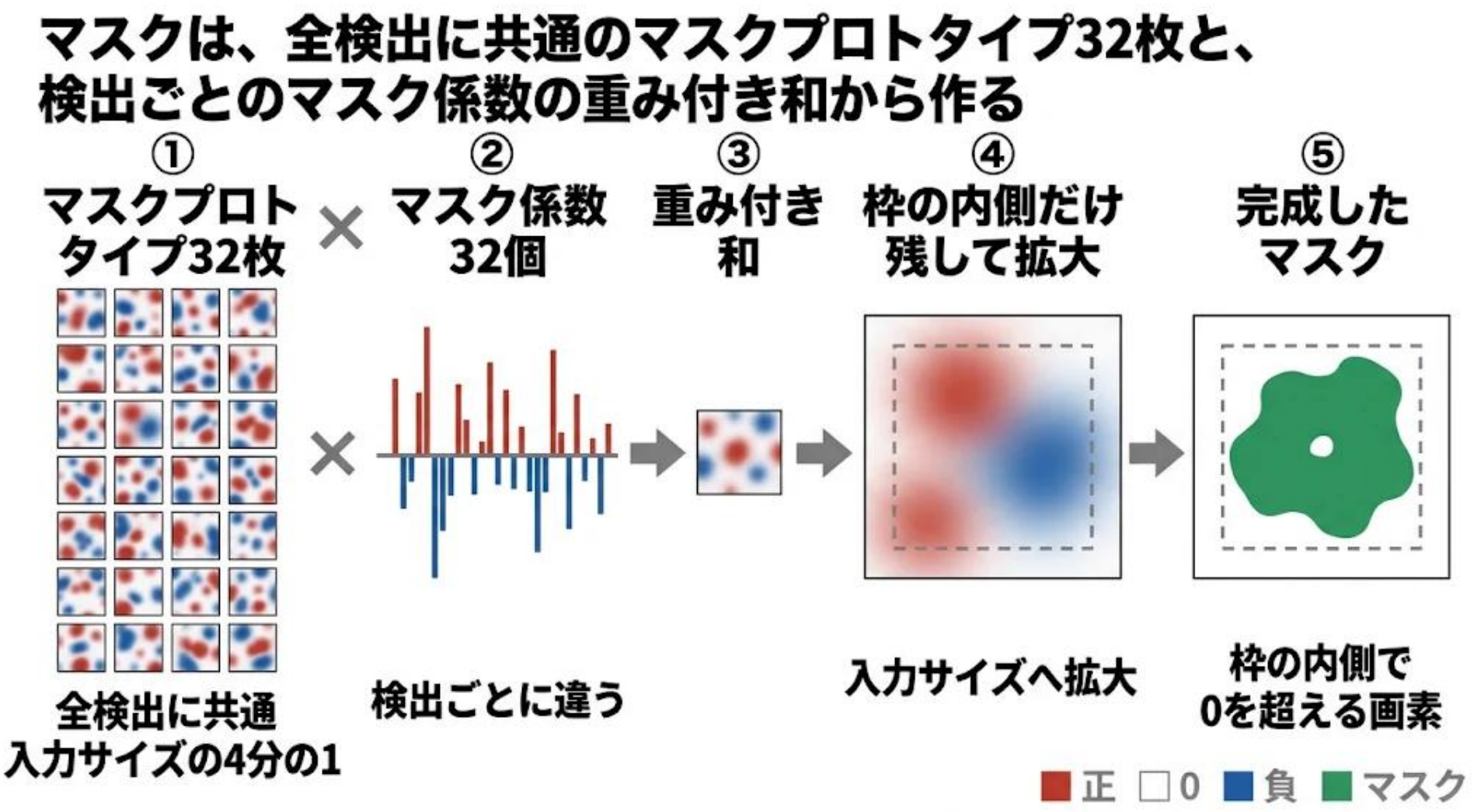
表示と追跡（スコアによる扱いの違い）

後処理を通った検出は、スコアによって表示のされ方と追跡での使われ方が決まる



表示閾値は追跡に影響しない
スコア閾値 0.10 と表示閾値 0.25 は既定値で、画面で変更できる。0.5 は固定値
BoostTrack++ (既定) では、既存の追跡と重なり、見た目が似た検出のスコアを引き上げてから 0.5 と比べる
TrackTrack では、後処理で除外された検出の一部も、既存の追跡を続けるために使う

マスクの合成


マスクは、全検出に共通のマスクプロトタイプ32枚と、
検出ごとのマスク係数の重み付き和から作る
①
マスクプロトタイプ32枚
×
②
マスク係数32個
③
重み付き和
④
枠の内側だけ残して拡大
⑤
完成したマスク
×
全検出に共通
入力サイズの4分の1
検出ごとに違う
入力サイズへ拡大
枠の内側で
0を超える画素
正
0
負
マスク