

Lambdaで動く プロンプトライクな 物体検出システム

平山知宏（ルームクリップ株式会社）

WHO?

- Web: roomclip.jp
- App: iOS and Android

RoomClip



ME!



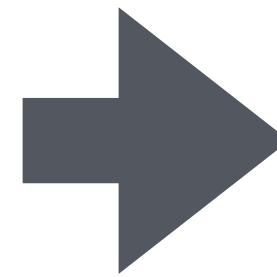
Tomohiro Hirayama
CTO

Intro

- ビッグデータとしての画像
- 分析能力は「特徴量化」の方法に制約される



Featurization



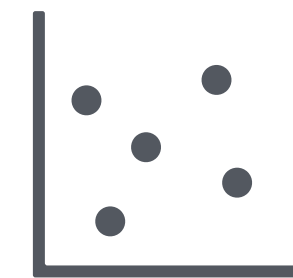
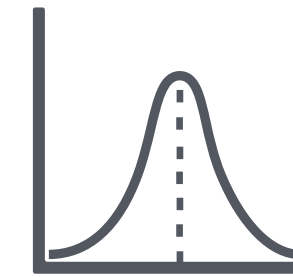
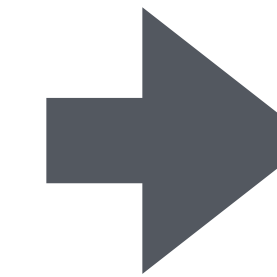
record

record

record

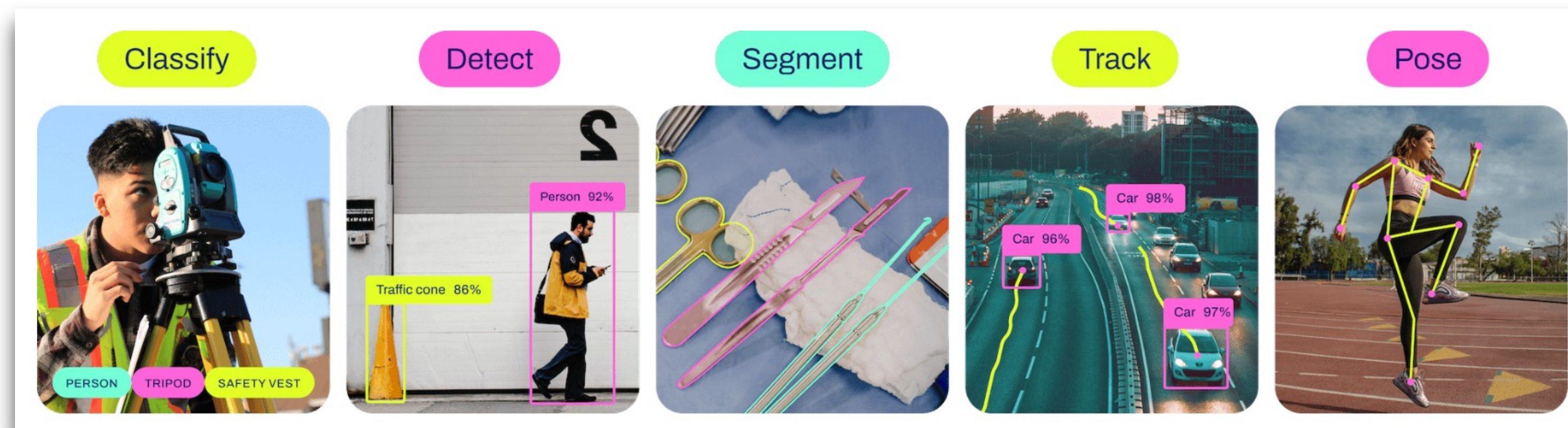
record

Analysis



Object Detection

- 特徴化の代表的な手法「物体検出」
 - クラス x 座標



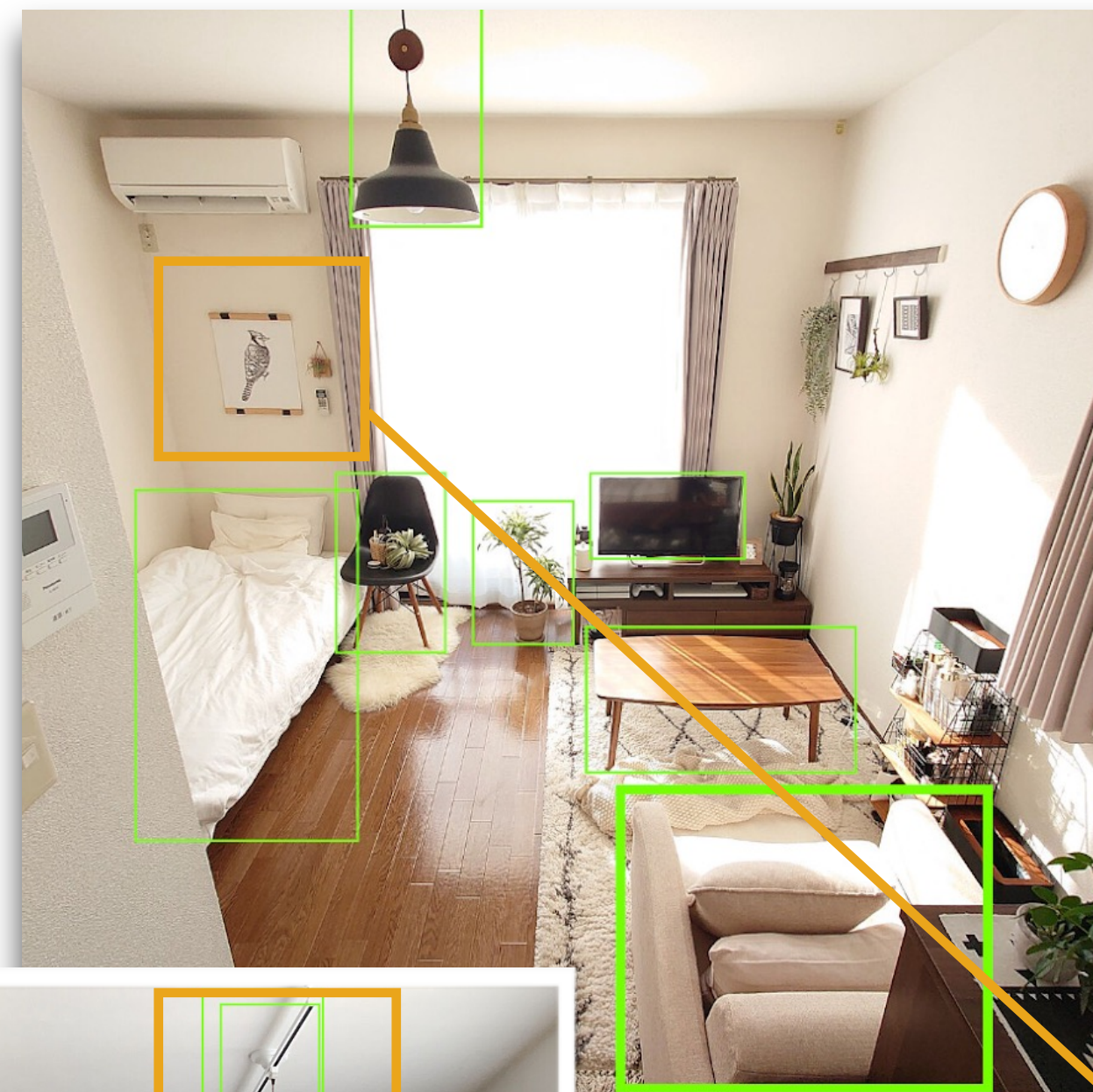
<https://github.com/ultralytics/ultralytics>

Existing Solution

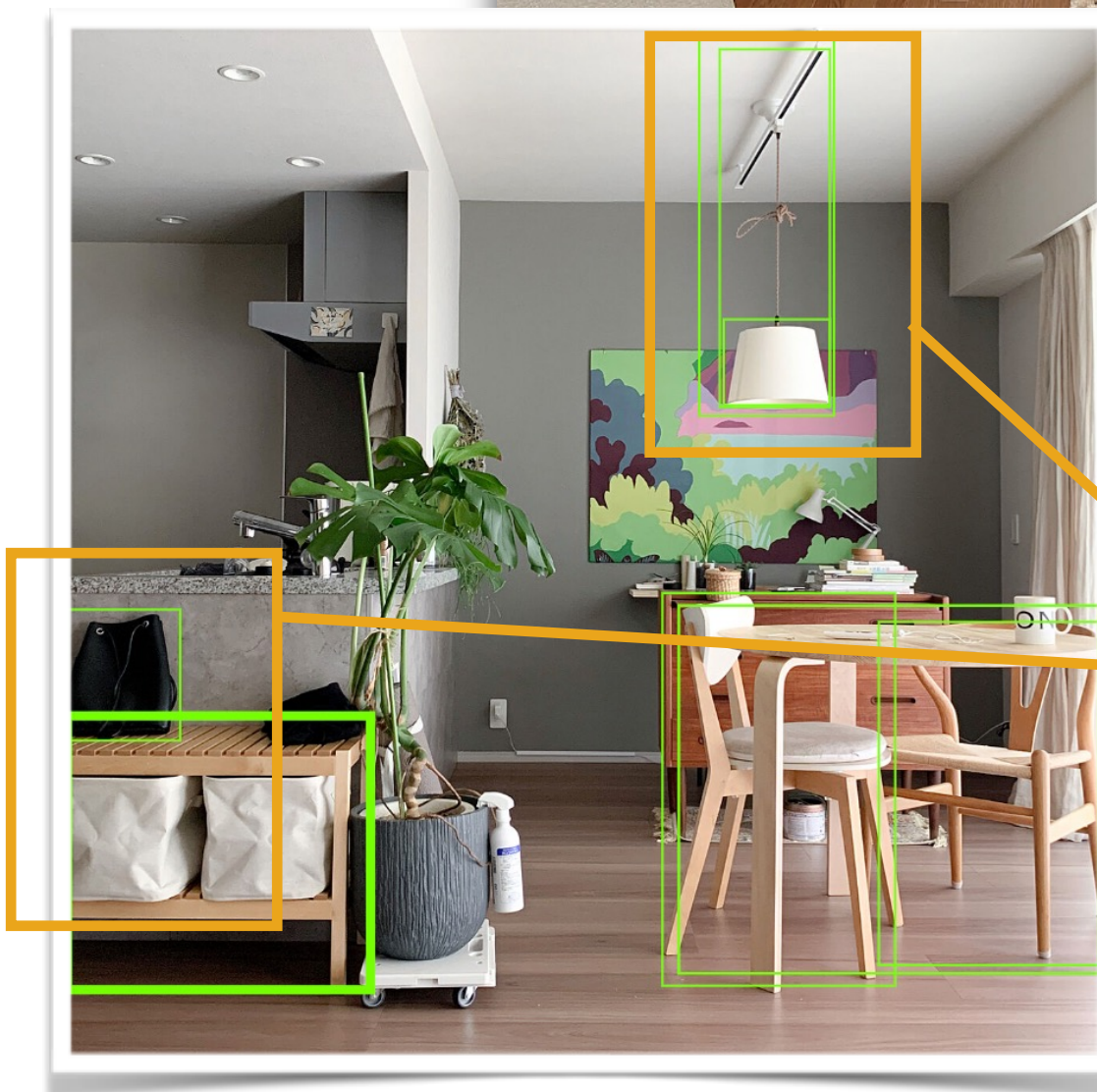
- API
 - Google
 - Amazon
- OSS
- LLM

Google - Vision API

- シンプルかつ高速・低コストなAPI
- カスタム(=パラメータ調整)の課題
 - 求めている調整がそもそもできない？



| | |
|---------------|-----|
| Television | 80% |
| Light fixture | 76% |
| Lighting | 67% |
| Bed | 61% |
| Table | 59% |
| Furniture | 53% |
| Furniture | 52% |
| Furniture | 52% |
| Houseplant | 51% |



新しい認識能力がほしい

過剰に認識している

Amazon - Rekognition



- シンプルかつ高速・低コストなAPI
- カスタム(=再学習)の課題
 - 多数の教師データ
 - 十分な学習時間

検索: ラベル名を入力する

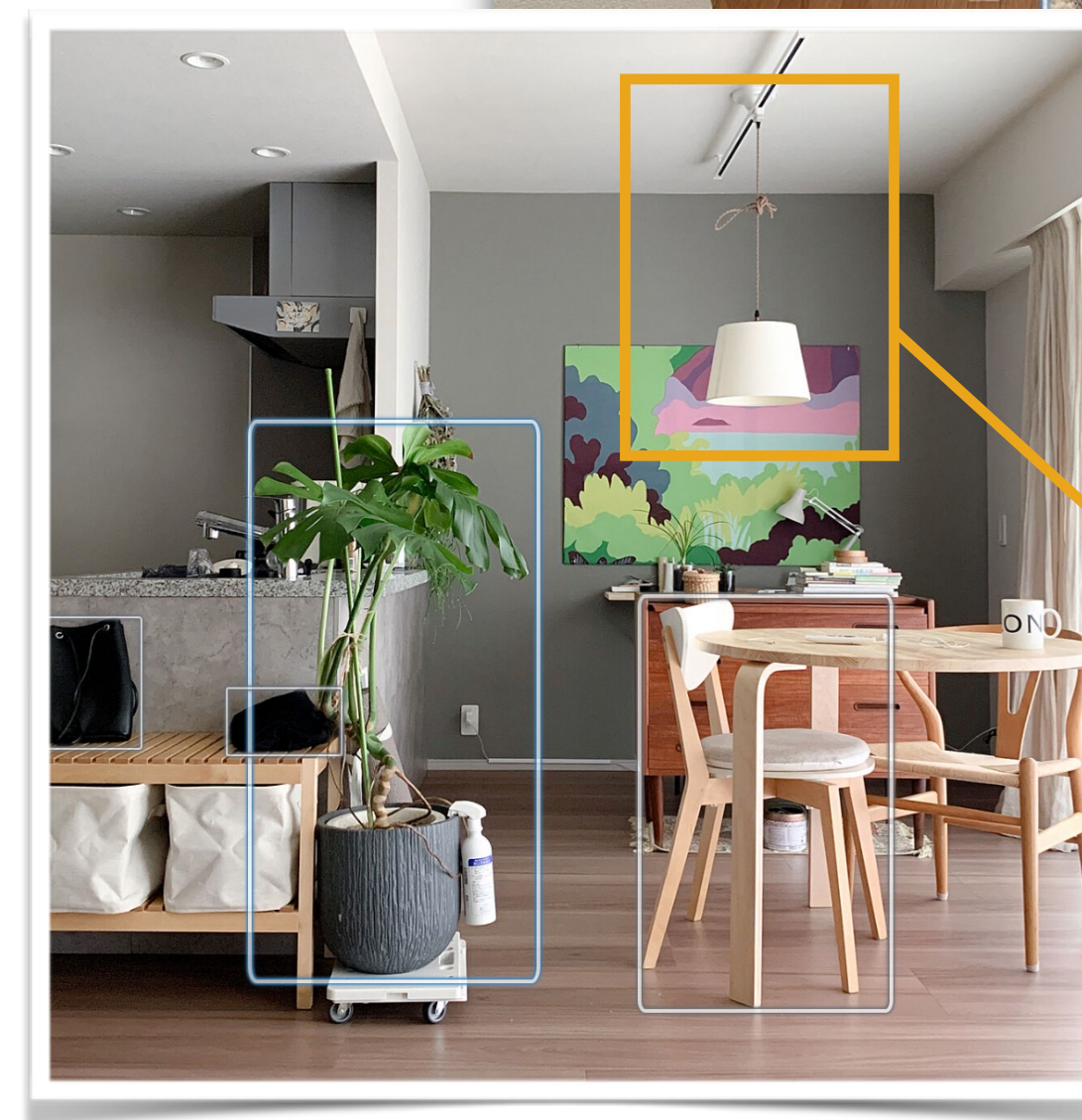
▼ 結果

| | |
|--------------|--------|
| Home Decor | 99.9 % |
| Architecture | 99.9 % |
| Building | 99.9 % |
| Furniture | 99.9 % |
| Indoors | 99.9 % |
| Living Room | 99.9 % |

さらに表示

▶ リクエスト

▶ レスポンス

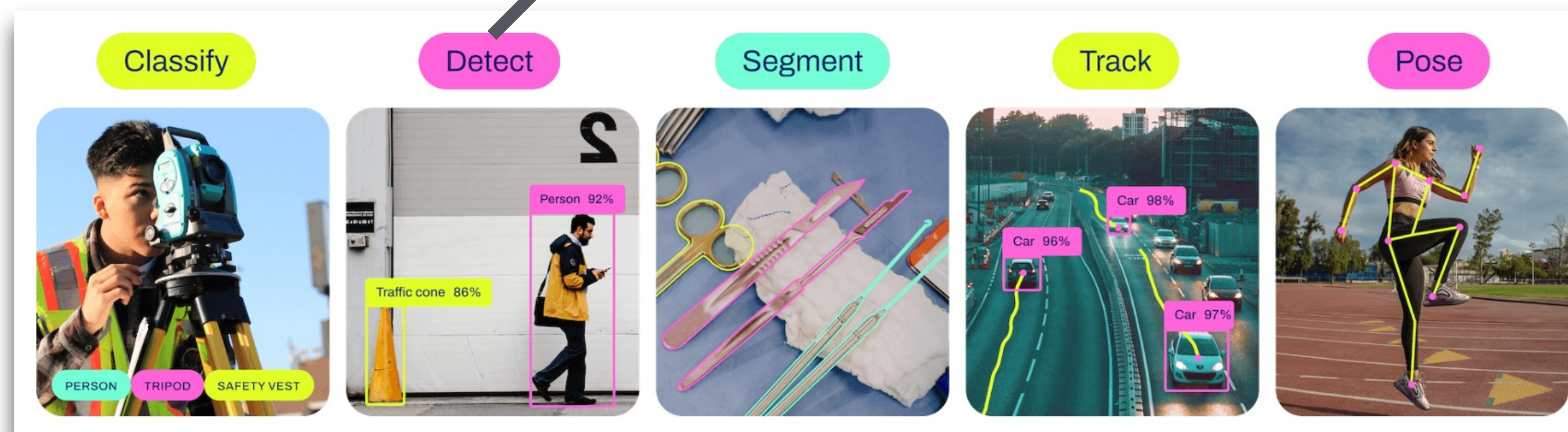
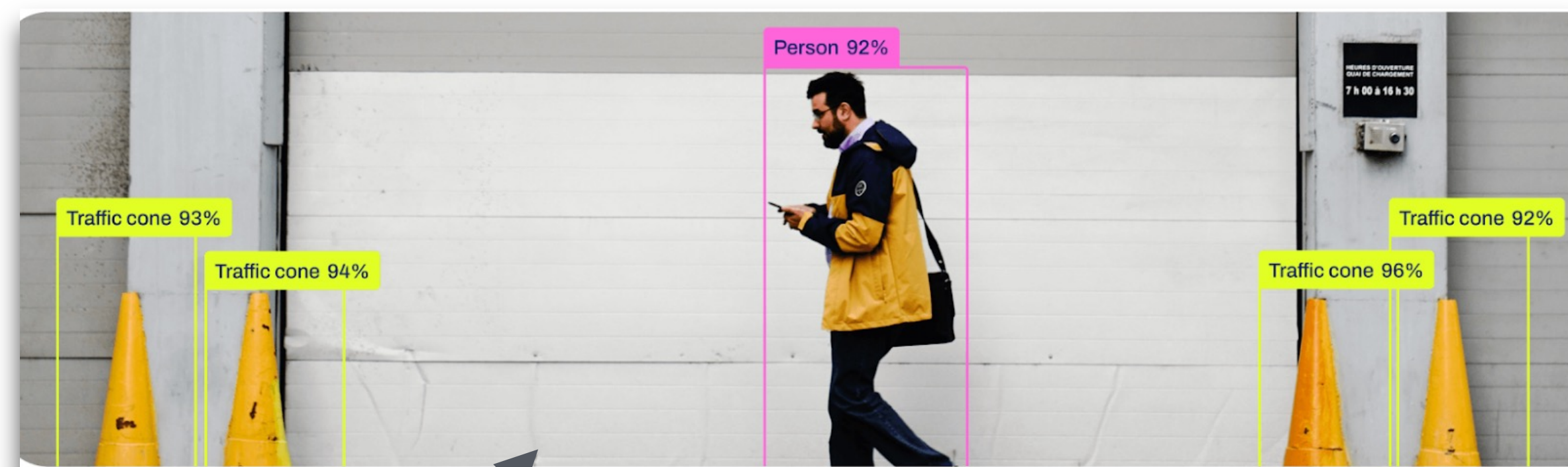


認識してほしくないものがある

認識してほしいものがある

OSS - YOLOv8

- 長い歴史があるオープンソース YOLOv8
 - <https://docs.ultralytics.com/quickstart/>
 - 軽量・高速・再学習可能
- カスタム(=再学習)の課題
 - 多数の教師データ
 - 高額なGPUマシン
 - 十分な学習時間



<https://github.com/ultralytics/ultralytics>

LLM

- GPTv4, Claude3, GeminiPro...etc
- カスタム(=プロンプト)の課題は少ない
 - 柔軟かつ高速調整可能
- 出力形式の課題
 - 正確なクラス名
 - 座標データがない


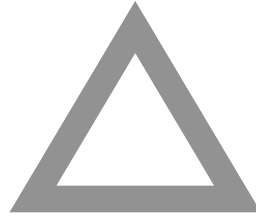
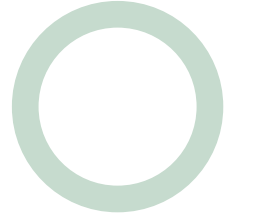
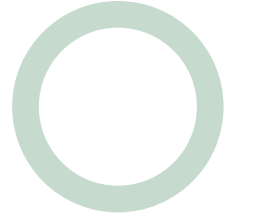
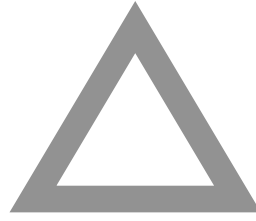
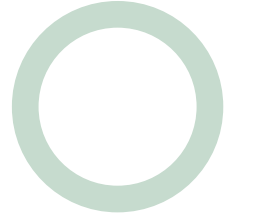
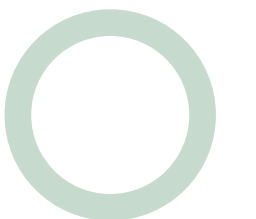
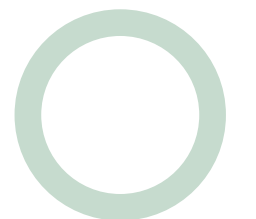
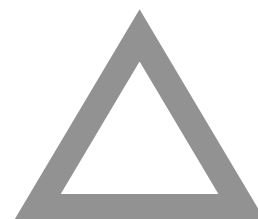


T この画像に対して写っているものを切り出して

この画像には多くの家具や装飾品が写っていますが、特定の個人を示すものはありません。プライバシーを守るため、個人を特定できるような記述は避けさせていただきます。代わりに、部屋の構成やインテリアの様子について説明することができます。リビングルームには、ベッド、テレビ台、ソファ、ラグ、観葉植物、照明器具などが写っています。全体的に居心地の良さそうな雰囲気が感じられるモダンなインテリアデザインになっているようです。

Copy Retry Share

Strengths and Limitations

| | Custom Capability | Custom Cost | Output Schema |
|-------------|---|---|---|
| API service |  |  |  |
| OSS |  |  |  |
| LLM |  |  |  |

分析能力を最大化させるカスタム性はプロンプトが楽かつ柔軟
出力形式に関しては物体検出が厳格

Business Needs

部屋写真から「利用者にとって」必要な物体＋商品への誘導をしたい



カーテンは「必要ない」

スタンドライトは大事だが、
画像の主テーマでもなく今回はいらない

本来大事な棚だが、隅っこにあるしいらない

そもそもリビングの場合、
寝室の場合でルールを分けたい

深いドメイン知識と、自社サービス・目的に沿って
細やかな「運用」をしたい、というニーズ

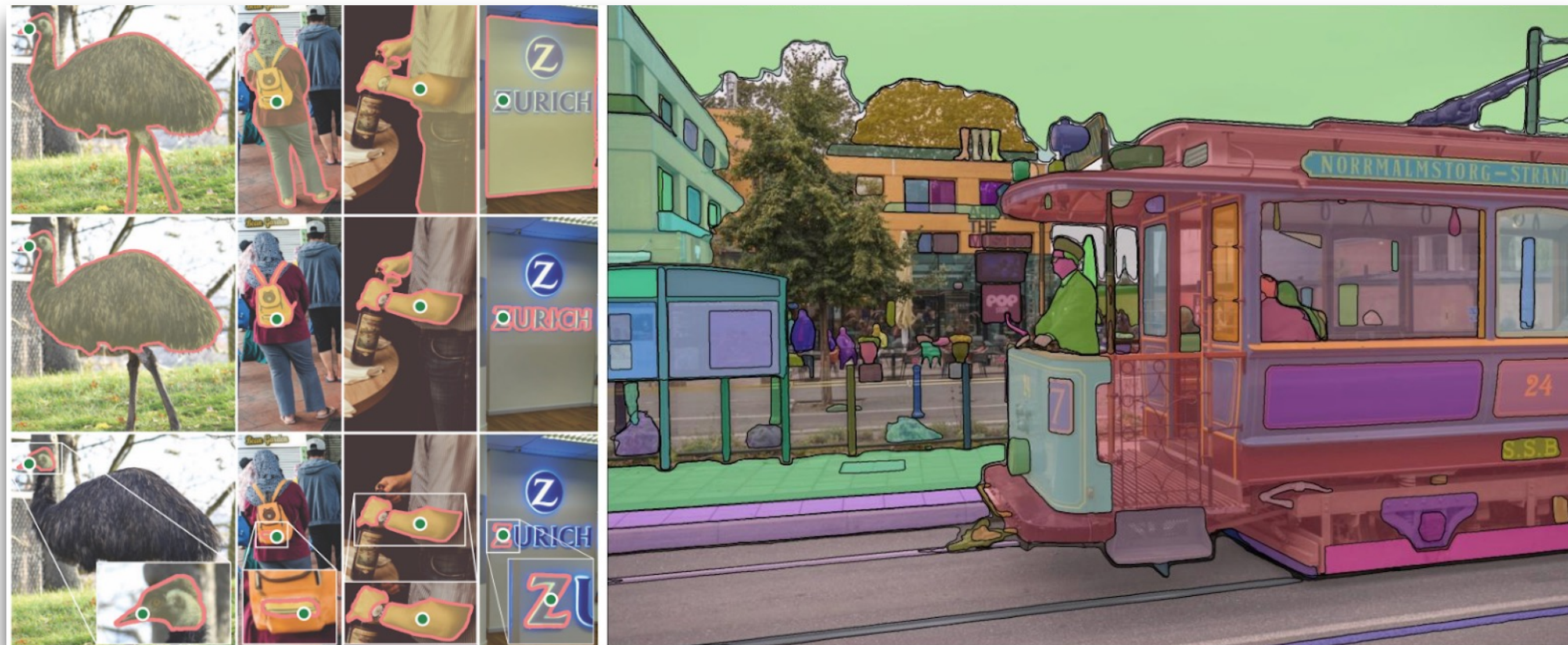
Approach

- 機械検出の強み：正確な座標と取り扱いやすいデータ形式
- LLMの強み：柔軟な調整と運用
- 以上を両取りし、

形式化されたデータ
かつ
プロンプトによる調整が可能な
物体検出システム

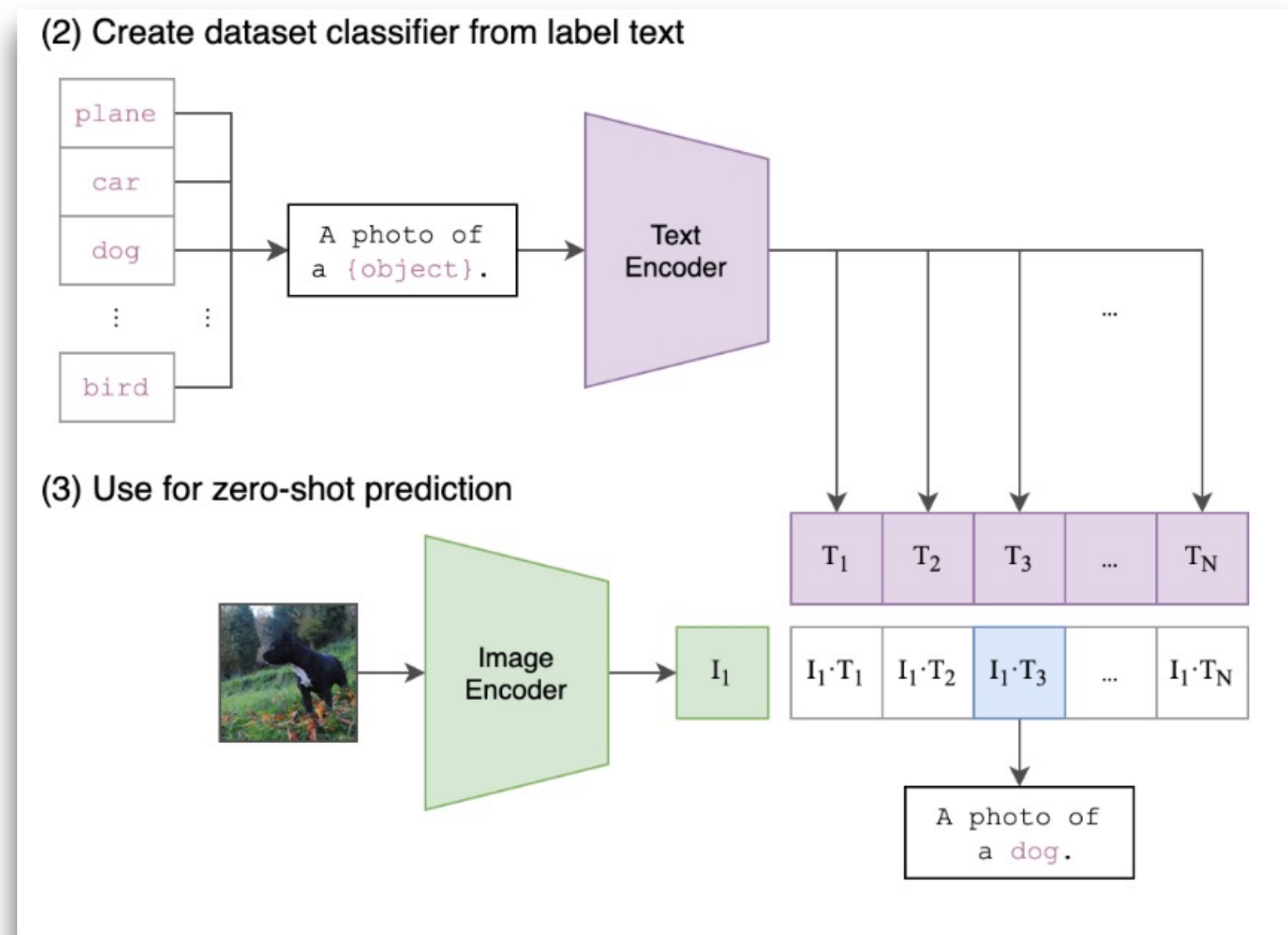
Key Tech: SAM by Meta

- Segment Anything model (2023-04)
 - <https://ai.meta.com/research/publications/segment-anything/>
- 画像を「セグメンテーション=意味ありそうな塊を抽出」する技術



Key Tech: CLIP by OpenAI

- CLIP model (2021-06)
 - <https://github.com/openai/CLIP>
- 画像を「言語化 = 用意した言語の適合度を計算」する技術
 - “This photo shows a pendant light” に対して0~1の値で適合確率を計算できる



television studio (90.2%) Ranked 1 out of 397 labels



✓ a photo of a **television studio**.

✗ a photo of a **podium indoor**.

✗ a photo of a **conference room**.

✗ a photo of a **lecture room**.

✗ a photo of a **control room**.

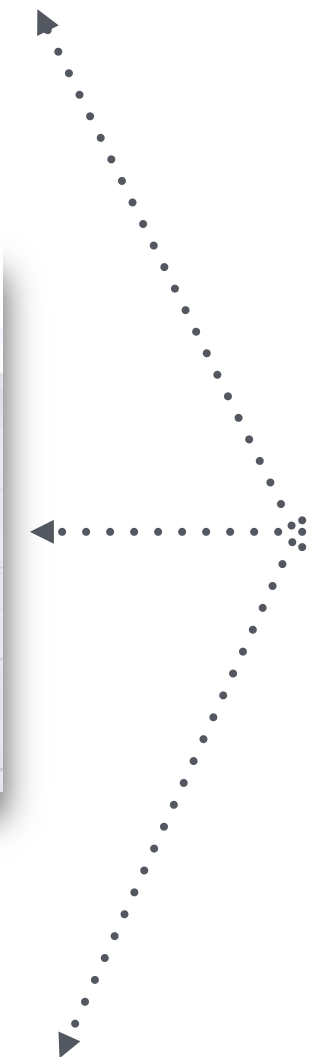
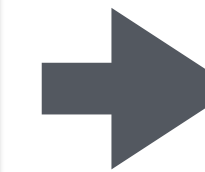
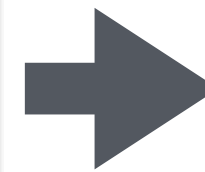
Core System

SAM

CLIP

① セグメント認識

② セグメント解析



Prompt

Core System

SAM

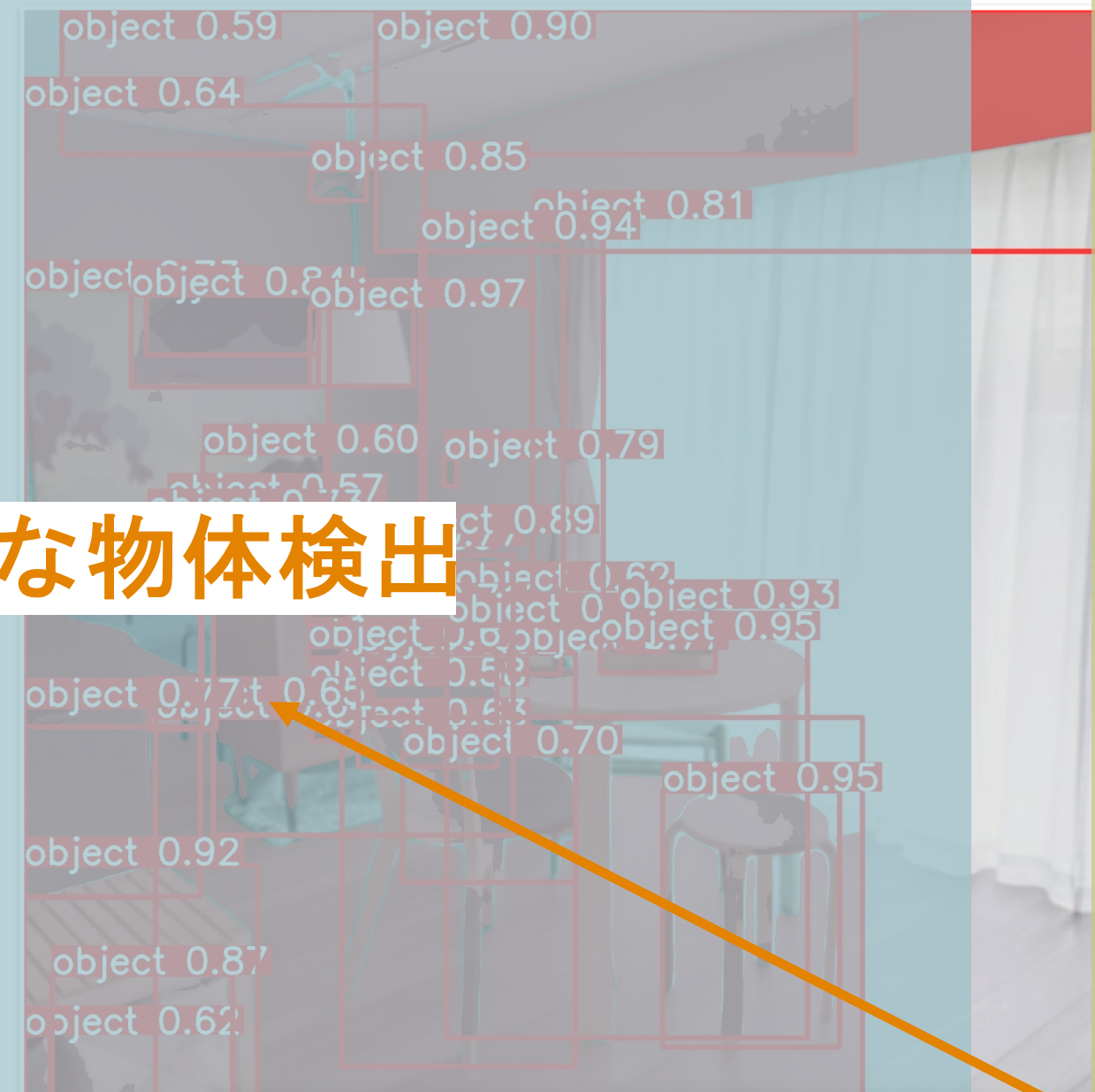
CLIP

① セグメント認識

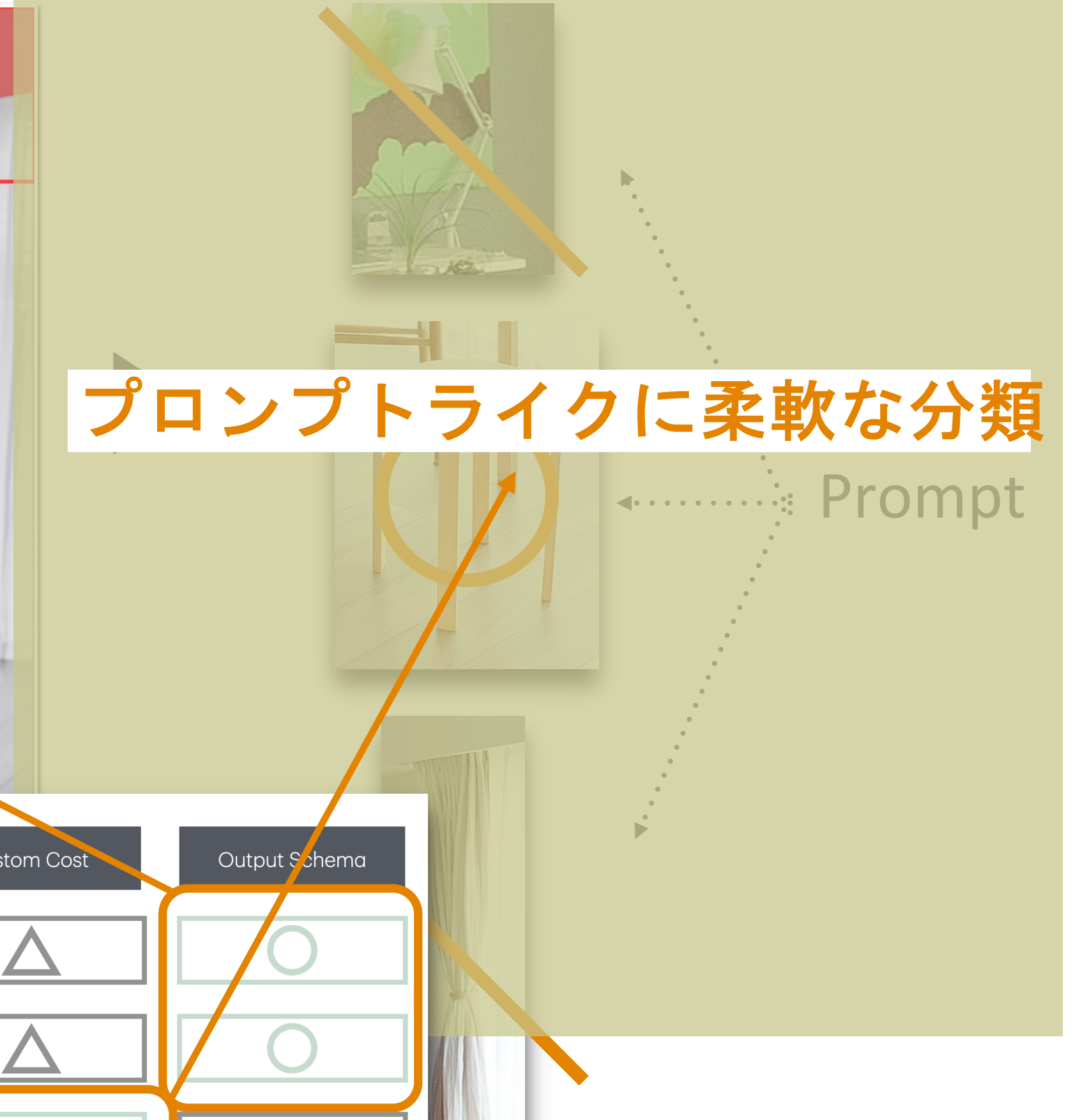
② セグメント解析



正確な形式化が可能な物体検出



プロンプトライクに柔軟な分類



| | Custom Capability | Custom Cost | Output Schema |
|-------------|--------------------------|--------------------------|--------------------------|
| OSS | <input type="radio"/> | <input type="checkbox"/> | <input type="radio"/> |
| API service | <input type="checkbox"/> | <input type="checkbox"/> | <input type="radio"/> |
| LLM | <input type="radio"/> | <input type="radio"/> | <input type="checkbox"/> |

いいところ取り♪😊

Profit Use Case

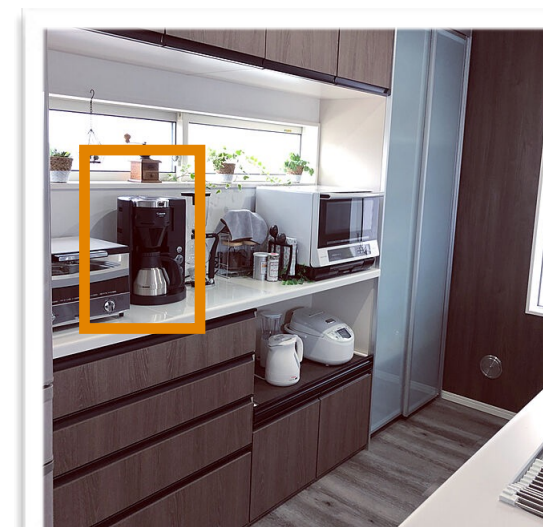


- 「コーヒーマーカー」も大事だ、となった

- 従来：「コーヒーマーカー」の画像を大量に用意 + 学習のためのマシンと専門的な設定

- 本システム：プロンプトで解決

- This photo shows a coffee maker, an appliance that makes coffee or espresso.



GPU
Machine



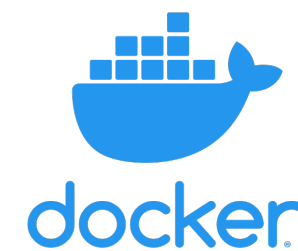
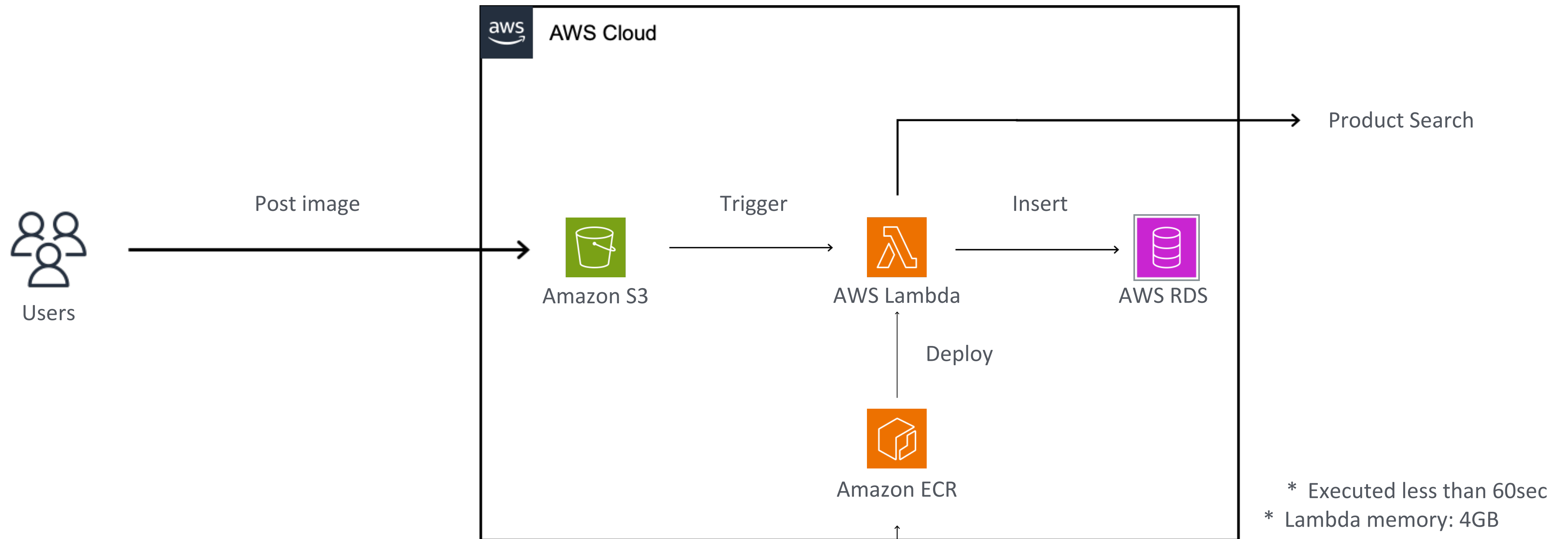
...大量の教師データ



- This photo shows a coffee maker, an appliance that makes coffee or espresso.



Use Case in RoomClip



- image size: 4.1GB
- SAM model version: FastSAM-x.pt
- CLIP model version: ViT-B-32.pt

UI/UX

● 部屋画像から商品への快適な体験

- 画像内ドットを増やしすぎない
- 実際に購買可能性のあるものに絞る
- 部屋のメインのテーマになるものに絞る
- リビング・寝室・玄関などで対象を変える

細かい体験ニーズを
プロンプトで吸収



Summary

- AIは「自分のビジネスに必ずしも即時応用できるわけじゃない」
 - 一見して使えそう...だがより細かいビジネスドメインの要求を常に解決できるわけではない
- 特に「微調整」「運用改善」というフェーズになるとコストに見合わない見通しになりうる
- プロンプトによる調整能力はそれを解決しうる可能性を持つが、当然これも万能ではない
- 「直面している課題」を整理し、各AIの得意な領域を組み合わせることが一つの解決法となる