Convergence Lab. Whitepaper

画像認識 AI、導入前に何を決めるべきか

用途選定・実証実験・本番運用の進め方

何に使うか。何で評価するか。どう業務につなげるか。
導入を検討する担当者のための、判断の手引き。

Convergence Lab. 株式会社
2026 年 11 月 6 日

01 / 導入の全体像

最初に決めるのは、AI を使って変えたい業務

「画像認識 AI で検査を効率化したい」「カメラ映像から状況を把握したい」。構想があっても、用途、必要な画像、精度の目標が曖昧なままでは、何を試し、どこで導入を判断するかが定まりません。

本資料は、画像認識 AI の導入を検討する業務責任者、開発担当者、情報システム担当者に向けた手引きです。用途選定から運用までを、各工程で決めることと、次へ進むために確かめることに分けて説明します。

技術の選択肢と、業務で使う条件を整理する

画像認識には、画像全体の分類、物体の位置の検出、領域の切り分け、文字の読み取りなどがあります。近年は、テキストで指定した対象を画像や動画から検出・領域分割するモデルも公開されています。例えば Meta の SAM 3 は、テキストによる対象の指定に対応しています。[1]

こうした技術を導入候補に加えたうえで、自社の画像、必要な処理速度、誤りが起きた際の対応を確認します。続くページで主要タスクと VLM を整理し、その後に用途選定から運用までの判断方法を説明します。

工程	決めること	次へ進むための確認
用途選定	対象業務と改善したい指標	画像から判断でき、効果を測れるか
実証準備	収集条件、正解、評価方法	現場を代表する画像を用意できるか
開発・評価	比較する方式と合格条件	精度・速度・確認負荷が条件を満たすか
本番導入	業務連携、例外処理、担当	現場で無理なく対応できるか
運用・改善	監視、再評価、更新の手順	変化を発見し、復旧できるか

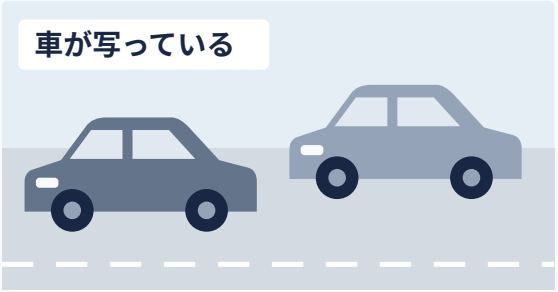
各工程で残す判断材料を決めると、実証実験の結果を導入判断につなげられます。

01 / 主要タスク / 図で見る出力の違い

同じ画像から、どんな答えを取り出すか

二台の車が写った同じ場面を使い、四つのタスクの出力を比べます。分類は画像全体のラベル、検出は各対象の矩形、領域分割は画素の範囲を返します。

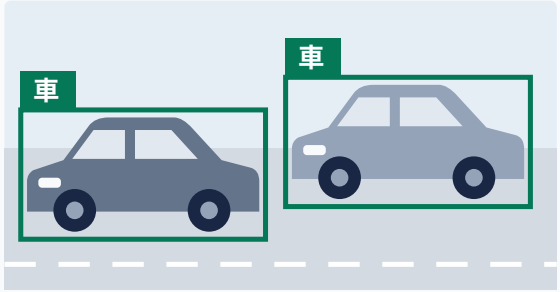
01 画像分類



出力：クラス

画像全体に「車が写っている」というラベル

02 物体検出



出力：クラス ＋ 矩形

車ごとの位置を矩形で示す

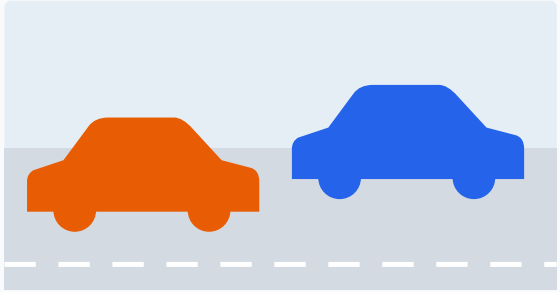
03 セマンティック領域分割



車　道路　背景

同じクラスの車を同じ色で表す

04 インスタンス領域分割



車 1　車 2

車ごとに別の色で輪郭を表す

図 1　認識結果の模式図。色はクラスや個体の区別を示す表示用の色です。実際のモデルの推論結果ではありません。

領域分割の二つの方式は、ここが違う

セマンティック：二台とも「車」という同じクラスの色になります。背景や道路もクラスごとに分けます。
インスタンス：「車 1」と「車 2」を別の個体として区別します。色が違っても、どちらもクラスは「車」です。

「何があるか」「どこにあるか」「どの画素か」「どの個体か」で、必要な出力を選びます。

01 / 主要タスク / 種類・位置・領域

何を知りたいかで、認識タスクを選ぶ

画像認識 AI のタスクは、入力画像から何を出力するかで区別できます。例えば同じ部品の画像でも、「種類を知る」「個数を数える」「傷の面積を測る」では必要な出力が異なります。代表的なタスクとして、分類・物体検出・領域分割などがあります。[6]

タスク	何を出力するか	用途例・代表的な指標
画像分類 Image Classification	画像全体や切り出した領域に、種類や状態のラベルを付ける。単一ラベルか複数ラベルかも定義する。対象の位置は通常出力しない。	部品の種類、良否の判定。 Accuracy、クラス別の Precision・Recall・F1。
物体検出 Object Detection	画像内の各対象に、クラスと位置を示す矩形（バウンディングボックス）を付ける。複数の対象を個別に検出できる。	部品や車両の位置把握、画像内の個数集計。 AP・mAP、運用設定での Precision・Recall。
セマンティック・セグメンテーション	各画素に「道路」「人」「傷」などのクラスを割り当てる。同じクラスの物体を、一つずつの個体としては区別しない。	道路領域や傷の範囲の抽出。 クラス別 IoU、クラス平均の mIoU、Dice。
インスタンス・セグメンテーション	各物体を個体ごとに区別し、それぞれの輪郭に沿った画素領域（マスク）を出力する。同じ種類が隣接する画像にも使う。	複数部品の輪郭抽出、個体ごとの面積計測。 マスクの IoU に基づく AP・mAP。

位置だけでよいか、輪郭まで必要か

対象の位置と個数が分かればよい場合は、物体検出が候補です。面積や形状を使う場合は、領域分割を検討します。用途に合う出力を先に決めると、画像への正解付けに必要な作業も具体化できます。領域分割では、一般に矩形より細かな輪郭の指定が必要になります。

領域の一致度には、IoU（共通部分 ÷ 合併部分）や、Dice（共通部分の面積の 2 倍 ÷ 予測と正解それぞれの面積の合計）を使います。面積を実寸へ換算する場合は、カメラの校正や撮影距離なども確認します。

同じ画像でも、欲しい答えは異なる

分類：「この画像は不良品」
検出：「この位置に不良箇所が二つある」
領域分割：「この画素の範囲が不良箇所」
個体ごとに区別する必要があるかで、領域分割の方式も選びます。

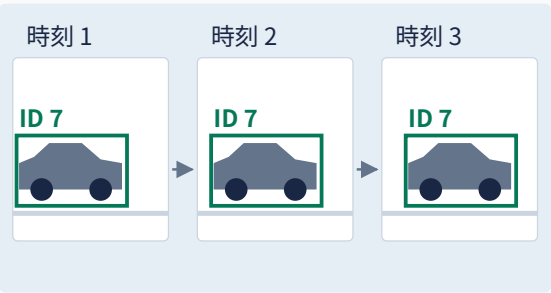
出力の粒度を決めると、タスク・正解データ・評価指標の選択がつながります。

01 / 主要タスク / 図で見る追跡・姿勢・文字・異常

動きや文字、異常はどのように表すか

追跡は時間をまたぐ対象の対応付け、姿勢推定はキーポイント、OCR は文字列、異常検知は通常との違いを出力します。それぞれの結果の見え方を、模式図で示します。

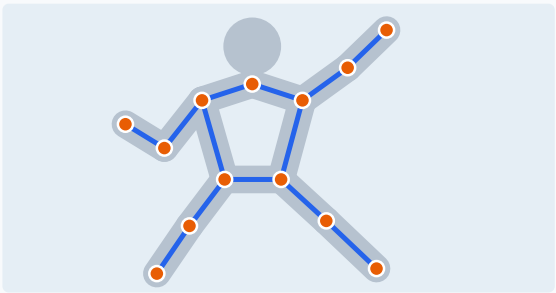
01 物体追跡



出力：対象ID ＋ 位置の時系列

フレームが変わっても同じ対象に同じID

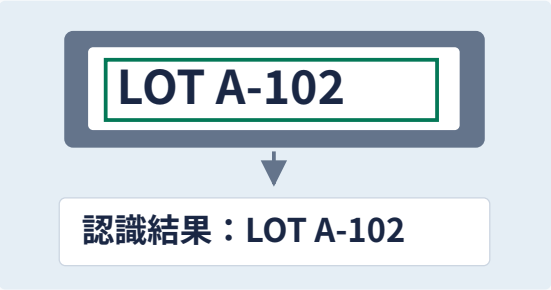
02 姿勢推定



出力：キーポイントの座標

肩・肘・膝などの位置を点と線で示す

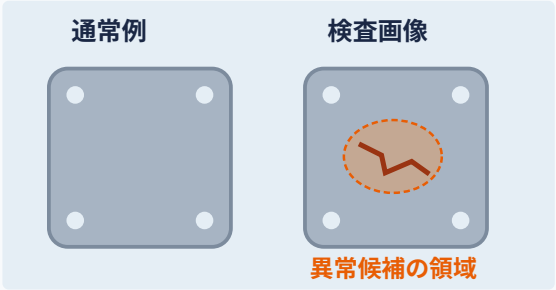
03 文字認識（OCR）



出力：文字領域 ＋ 文字列

文字の場所を見つけ、文字列へ変換する

04 異常検知



出力：異常スコアや候補領域

通常と違う箇所を異常候補として示す

図 2　各タスクの出力例を表した模式図。実際の推論結果ではありません。異常候補領域の表示に対応するかどうかは方式によって異なります。

図を読むときのポイント

追跡：ID は映像内で同じ対象を区別するためのものです。
姿勢推定：図の点と線は、関節などの位置と接続関係を表します。
OCR：文字の読み取り後に、製造番号などの項目への対応付けを行う場合もあります。
異常検知：異常候補が業務上の不良に該当するかは、別途基準に照らして判断します。

必要な出力を確認し、時系列・動作・記録・検査といった業務へつなげます。

01 / 主要タスク / 時間・動作・文字・異常

動きや文字、通常との違いを捉える

業務によっては、一枚の画像での種類や位置に加え、時間をまたぐ対応付け、関節の位置、文字列、異常の兆候が必要です。複数のタスクを組み合わせる場合は、それぞれの出力と最終的な業務成果を評価します。

タスク	何を出力するか	用途例・評価の観点
物体追跡 Object Tracking	動画のフレーム間で同じ対象を対応付け、識別 ID と位置の推移を出す。追跡 ID は、実在の人物の氏名などを特定するものではない。[7]	動線、滞在時間、重複を避けた通過数集計。 HOTA・IDF1、ID の入れ替わりや追跡の途切れ。
姿勢推定 Pose Estimation	人や物体のキーポイントの位置を推定する。人体なら肩・肘・膝などを使う。2D 座標か 3D 座標かを区別する。[6]	作業姿勢の把握、動作分析の入力。 キーポイントの位置誤差や OKS に基づく AP など。
文字認識 OCR	画像内の文字領域を検出し、文字列へ変換する。項目名と値の対応付けや帳票の構造化は、追加の処理として設計する。[8]	ラベル、製造番号、書類の読み取り。 文字誤り率（CER）、文字列・項目単位の完全一致率。
異常検知 Anomaly Detection	通常の状態と異なる度合いをスコアとして出し、方式によっては異常候補の領域も示す。正常画像を中心に学習する方式もある。[9]	不良例を集めにくい外観検査など。 画像単位と領域単位を分け、運用しきい値での見逃し・誤検出も測る。

追跡の HOTA・IDF1 は、検出の正しさや対象の対応付けを評価する指標です。姿勢推定の OKS は、対象の大きさやキーポイントごとの許容差を考慮した位置の一致度を表します。OCR の CER は、正解に合わせるための文字の置換・削除・挿入の合計を、正解の文字数で割った値です。

単体の認識と、業務上の意味付けをつなぐ

例えば作業動作の判定には、姿勢推定の結果を時系列で分析する工程が必要になることがあります。異常検知の高スコアも、出荷不可の不良とは限りません。照明や位置のずれなどと、業務上の異常を区別して評価します。

組み合わせるタスクごとの性能と、集計・判定・記録までの最終結果を確かめます。

01 / VLM による画像認識 / 特徴

ラベル付きの学習データを用意せずに試す選択肢

VLM（Vision-Language Model：視覚言語モデル）は、画像と言葉を関連付けて扱うモデルです。画像と指示文を入力し、内容の説明や質問への回答を得る生成型のほか、画像と言葉の類似度を使う方式があります。[11・12] 画像を分類する、文字を読むといったタスクに対して、VLM という実現方式を選ぶ、と整理できます。

最大の特徴は、追加学習なしで試せること

事前学習済みモデルを利用すれば、自社用の大量のラベル付き学習データを作る前に、文章で対象や判断内容を指定して認識を試せます。対象タスクの例を与えずに試す方法をゼロショット、少数の例を入力して判断を補助する方法をフューショットと呼びます。例をプロンプトに添えることと、モデルを追加学習することは区別します。

方式	入力と出力	導入時の特徴
生成型 VLM	画像と質問・指示から、説明文、回答、指定形式のテキストなどを生成する。[11]	「何が写っているか」「この条件を満たすか」などを指示できる。対応する出力や精度はモデルで異なる。
画像・テキストの照合型	画像と候補の説明文の類似度を求める。CLIP 系などを用いたゼロショット分類が例。[12]	候補文を変更して分類を試せる。スコアは候補や表現に依存し、そのまま正解確率とは扱わない。
言語で対象を指定する検出	画像と対象名・説明から矩形などを出す。Grounding DINO などが例。[13]	固定クラスの専用検出器を学習する前に、言葉で対象を指定して検出を試せる。

最後の方式は、画像と言語を扱う関連技術であり、対話型の生成 VLM と同じ機能を持つとは限りません。文章で画像の内容を答えられるモデルが、正確な矩形・マスクも出せるとは限らないため、必要な出力に合う方式を選びます。

「ラベルデータがいらない」の範囲を整理する

試行開始：用途に合う学習済みモデルなら、自社のラベル付き学習データなしで始められる場合があります。
導入判断：自社画像で正解を確認し、性能を評価するためのデータは必要です。
改善：独自の対象や基準への対応に、追加の例示やラベル付きデータによる学習が必要になる場合もあります。

VLM は、データ整備に先立って用途の実現可能性を確かめる入口にもなります。

01 / VLM による画像認識 / 選定と評価

指示の柔軟さを活かし、自社の条件で確かめる

VLM では、認識したい内容を文章で変えられます。用途の探索、画像の説明、確認作業の補助などを試す候補になります。一方、細かな欠陥、厳密な個数、独自の良否基準への適合性は、実際の画像と要件で確認します。

専用モデルと、同じ業務条件で比較する

観点	VLM を使う場合	専用の分類・検出モデルを使う場合
立ち上げ	追加学習なしの試行が可能。指示文や候補文、出力形式を設計する。	既存モデルの適用や、自社データでの追加学習を検討する。
対象の変更	文章の変更で試せる範囲がある。変更後の性能は再評価する。	クラス追加や基準変更に、学習データやモデルの更新が必要な場合がある。
精度と出力	言葉による柔軟な指定が特徴。表現や入力画像の条件で結果が変わりうる。	対象と出力を限定し、その条件に合わせて性能を調整しやすい。
速度と費用	モデル規模、画像解像度、生成量、実行環境で変わる。実測が必要。	処理を絞った小型モデルも候補。自社環境で速度・維持費を比較する。

同じ画像を何度も処理する場合や、多数のカメラで連続稼働させる場合は、一件の試行から総量を見積もります。VLM と専用モデルのどちらが適するかは、認識品質、遅延、処理量、運用費用をそろえて判断します。

自然な説明と、認識の正しさを分けて評価する

生成型 VLM には、画像と一致しない対象や内容を回答する誤りがあり、物体の幻覚として研究されています。[14] 「判断不能と返す」と指示した場合も、その指示が守られるかを検証します。回答の自然さや、モデル自身が述べる確信の強さだけでは採点しません。

評価では、自社の通常画像、対象がない画像、難しい条件を集め、正解と採点基準を定めます。分類なら Precision・Recall・F1、検出なら AP など、出力に合う指標を使います。自由文なら必須情報・誤った記述・判断不能の扱いを項目ごとに採点します。モデル版、指示文、画像の前処理、生成設定も記録します。

VLM から専用モデルへつなぐ進め方もある

VLM で用途を試す → 人が正解と誤りを確認する → 必要なラベルを整える → 専用モデルや追加学習と比較する。VLM を候補ラベルの作成に使う場合も、付け漏れと誤りを人が確認します。

学習データを作る前に試し、評価結果から VLM の継続利用・追加学習・専用モデルを選びます。

02 / 用途選定

「何を認識するか」を、「どう仕事が変わるか」へ

最初に、現在の作業を一連の流れとして書き出します。誰が、何を見て、どのように判断し、その後に何をしているでしょうか。AI を組み込む場所と、人が担当する場所を具体的にします。

例えば外観検査なら、傷を見つけることに加えて、製品を止める、再検査する、結果を記録するといった処理があります。AI が候補を示しても、確認作業が増えれば、期待した省力化につながらない場合があります。

最初の用途は、この四つで比較する

比較する観点	確認する問い
業務上の効果	作業時間、再検査、見落とし、記録のどれを改善したいか。現状値を把握できるか。
画像での判断可能性	必要な特徴が画像に写るか。人も画像だけで判断できるか。別の情報が必要か。
試しやすさ	対象を限定できるか。画像収集や現場での検証に協力を得られるか。
誤りへの対応	見逃しや誤検出が起きたとき、確認・やり直し・停止が可能か。

自動化する範囲を段階的に設定する

初期の用途には、人が確認する対象を絞る、候補箇所を表示する、件数を集計するといった形もあります。全件を自動処理する場合と、人の判断を補助する場合では、必要な精度と許容できる処理時間が異なります。

対象の範囲も限定します。「全製品の全不良」より、「特定ラインの特定製品における、定義済みの欠陥候補」の方が、実証で確かめる内容を明確にできます。対象外の製品や、画像から判断できないケースの扱いも記録します。

用途を一文で書く

「誰の」「どの作業で」「何を認識し」「結果をどう使い」「何を改善するか」を埋めます。
記入例（架空）：検査担当者の再確認作業で、部品表面の欠陥候補を検出し、候補画像を優先表示することで、確認時間を減らす。

用途選定の成果物は、対象業務・期待効果・自動化の範囲をまとめた一枚の計画です。

03 / 実証準備

件数を決める前に、集める画像の条件を決める

必要な画像数は、対象の種類、ばらつき、失敗の頻度、求める性能、利用する方式によって変わります。まず少量の画像を確認して課題を整理し、不足する条件を追加収集する計画を立てます。

日常の条件と、困る条件の両方を集める

通常の対象だけでなく、似た別物、対象がない画像、隠れや重なり、反射、汚れなども検討します。カメラ、距離、角度、時間帯、製品型番といった条件を記録すると、後から失敗の原因を調べやすくなります。

準備するもの	整理する内容
画像と撮影条件	取得日時、場所、カメラ、対象、撮影距離など。収集できていない条件も記録する。
正解の基準	何を検出・分類するか。境界が曖昧な例、対象外、判断不能の扱い。
データの分割	学習用、調整用、最終評価用の分け方。撮影日や対象物の重複を確認する。
取り扱いのルール	収集・利用の範囲、提供可能な相手、保存場所、アクセス権、保存期間。

同じ動画の連続フレームや、同じ製品を繰り返し撮影した画像を無作為に分けると、学習側と評価側が似すぎることがあります。実際に確かめたい条件に合わせ、撮影日、対象物、拠点などの単位で分割を検討します。

低照度は、条件を分けて評価するための一例

低照度画像の公開データセット ExDARK は、可視光で撮影した画像を、影や薄暮など 10 種類の照明条件に分けて収録しています。[2] 自社の評価でも、全体の平均に加え、利用する時間帯や照明条件ごとの結果を確認する考え方が使えます。

暗い現場への対策では、可視光 AI の改善、照明の追加、カメラやセンサーの変更を比較対象にします。何を見分けたいか、設備を変更できるか、設置や維持にどれだけ費用がかかるかを含めて検証します。

実証準備の成果物は、画像一覧・正解の基準・収集条件表・評価用データの分割方針です。

03 / アノテーション / 正解の設計

何を「正解」とするかを、データに書き込む

アノテーションとは、画像や動画に、種類・位置・領域・文字列などの情報を付ける作業です。教師あり学習の材料になるだけでなく、モデルの出力を採点する基準にもなります。VLM を追加学習せずに使う場合も、評価に必要な正解を整理します。

タスクに合わせて、ラベルの形を選ぶ

対象タスク	付ける情報	先に決めること
画像分類	画像・切り出し領域ごとのクラス。複数の状態が共存する場合は複数ラベルも検討する。	良否・種類の定義、複数該当や判断不能の扱い。
物体検出	対象ごとの矩形とクラス。	小さな対象、隠れた対象、画像端で切れた対象をどこまで含めるか。
領域分割	クラスごとの画素領域。個体を区別する方式では個体 ID も付ける。	輪郭の境界、穴、重なり、背景の定義。
追跡・姿勢推定	追跡はフレーム間で共通の対象 ID。姿勢推定はキーポイント座標と可視性など。	一度画面外に出た対象の ID、見えない点の扱い。
OCR	文字領域と、そこに書かれた文字列。	読み順、記号、空白、読めない文字の表記。
異常検知	正常／異常の区分。領域評価には異常箇所のマスクなど。	業務上の不良と、見た目の変化の違い。

正常画像を中心に学習する異常検知でも、性能を評価するには、正常と異常を区別した評価データが必要です。領域を評価する場合は、その正解領域も用意します。学習に必要なラベルと、評価に必要なラベルを分けて計画します。

境界例を含むガイドラインを作る

クラス名の一覧に加え、ラベルを付ける例・付けない例・判断が分かれる例を画像付きで整理します。物体検出なら、見えている部分だけを矩形で囲むか、隠れた部分まで推定して囲むかを統一します。低照度で対象を特定できない画像を、推測で特定クラスへ割り当てないことも重要です。

「対象がない」と「未確認」を区別する

ラベルがない画像が、本当に対象なしなのか、作業前・付け忘れなのかを管理します。判別不能な領域を評価から除く場合も、その条件と件数を記録します。後から評価結果に合わせて除外基準を変えないようにします。

アノテーションの最初の成果物は、作業者が同じ判断をできる正解のガイドラインです。

03 / アノテーション / 品質と工数

少量で試し、判断のずれを直してから増やす

大量の作業を始める前に、通常例と難しい例を含む少量の画像で試行します。作業者が迷った点を記録し、現場の判断に責任を持つ担当者と基準を修正します。委託する場合も、判断基準の承認と例外の裁定を担う担当を決めておきます。

作業・確認・修正の流れを決める

工程	具体的に行うこと
1. 試行	一部の画像を複数人で独立に付け、クラス、矩形、輪郭、付け漏れなどの不一致を比較する。
2. 基準の修正	不一致の原因を確認し、説明と具体例をガイドラインへ追加する。判断できない場合の相談先も決める。
3. 本作業と検査	作業済み画像を抽出して確認し、誤りの傾向を調べる。重要クラスや難しい条件は重点的に見る。
4. 是正と版管理	基準の変更が既存ラベルに及ぼす範囲を確認し、修正する。画像・ラベル・基準の版を対応付ける。

品質確認には、複数人の結果の比較や、確認済みの参照ラベルとの照合などがあります。CVAT などのツールにも、参照ラベルを用いた検証やレビューの機能があります。[10] 一致率が高くても基準自体が誤っている可能性はあるため、業務上の正解を確認する工程は残します。

自動アノテーションは、修正作業まで含めて評価する

既存モデルで矩形やマスクの候補を作り、人が修正する方法もあります。採用前に、手作業との所要時間と修正後の品質を比較します。候補に出なかった対象の付け漏れや、同じ種類の誤りをそのまま採用していないかも確認します。

評価用の正解を作る際は、評価対象モデルの出力を無検証で流用しません。元画像とガイドラインに照らして正解を確認し、そのモデルに有利な採点基準にならないようにします。

画像枚数だけで工数を見積もらない

一枚あたりの対象数、輪郭の複雑さ、動画の長さ、専門知識の必要性で作業量は変わります。試行で計測した作業・レビュー・修正の時間を基に、対象条件別に見積もります。ツール設定、基準の整備、形式変換、変更時の再作業も含めます。

見積もり時に共有するもの

サンプル画像、クラス数、必要な形状、画像・対象の数量、品質の検査方法、納品形式、判断基準の承認者をそろえます。まず試行を行い、その結果で本作業の範囲と費用を決める進め方が考えられます。

件数に加え、ラベルの一貫性・付け漏れ・修正履歴を管理し、学習と評価の信頼性を支えます。

04 / 開発・評価 / 指標の基礎

「精度」を分解し、見逃しと誤検出を測る

方式を比較する前に、何を合格とするかを決めます。「精度 95%」だけでは、何を測った数字かが分かりません。分類や物体検出では、適合率・再現率などを使い、誤りの種類を分けて確認します。[4]

適合率・再現率・F1 を使い分ける

TP（真陽性）は正しく検出した対象、FP（偽陽性）は誤った検出、FN（偽陰性）は見逃した対象です。物体検出では、クラスと位置の一致条件を決め、予測と正解を対応付けて数えます。

指標	定義	読み取れること
適合率 Precision	TP ÷ (TP + FP)	検出した候補のうち、正しいものの割合。低いと確認対象に誤検出が多く含まれる。
再現率 Recall	TP ÷ (TP + FN)	検出すべき対象のうち、拾えた割合。見逃し率は同じ条件で 1 − Recall。
F1 スコア	2 × Precision × Recall ÷ (Precision + Recall)	適合率と再現率の調和平均。一つの値で比較できるが、業務ごとの失敗の重みは反映しない。

計算例：候補の正しさと、対象の拾い上げは別の数字

説明用の架空例：正解の対象 100 個に対し、120 個の候補を出し、そのうち 90 個が正解だった場合、TP＝90、FP＝30、FN＝10 です。
適合率は 75%、再現率は 90%、F1 は約 0.82。見逃しは 10 個、誤検出は 30 個と、件数でも示します。

実運用では、候補を採用する自信度のしきい値を決め、その設定で適合率と再現率を測ります。しきい値を変えた結果を適合率–再現率曲線（PR 曲線）などで確認し、見逃しの許容範囲と人の確認能力から運用する設定を選びます。

正解率（Accuracy）だけでは足りない場合がある

二値分類の正解率は、正しく陽性・陰性と判定した件数を全件数で割った値です。例えば 1,000 件中、不良が 10 件しかなければ、全件を良品と判定しても正解率は 99%ですが、不良の再現率は 0%です。分類ではクラス別の結果や混同行列も確認します。物体検出では背景の「陰性」を一意に数えにくいため、通常は適合率・再現率や AP を用います。

指標名・対象クラス・しきい値と、TP・FP・FN の件数を合わせて報告します。

04 / 開発・評価 / 物体検出と導入判断

位置の正しさを測り、業務の合格条件につなげる

物体検出では、対象の種類に加え、位置が正しいかを評価します。領域の重なりを表す IoU と、検出性能をまとめる AP・mAP を押さえると、比較表の数字を読み解きやすくなります。 [5]

指標	意味と、比較する際の注意
IoU Intersection over Union	予測領域と正解領域の「共通部分の面積 ÷ 合併部分の面積」。完全一致で 1。物体検出では通常、矩形の重なりを測る。領域分割ではマスクにも使える。
AP Average Precision	自信度の順に候補を評価して得る PR 曲線を、所定の補間・集計方法で要約した値。単一の運用しきい値での適合率とは異なる。
mAP mean AP	各クラスの AP の平均。mAP@0.5 は IoU しきい値 0.5 で評価。COCO 方式の mAP@0.5:0.95 は、0.50 から 0.95 まで 0.05 刻みの 10 条件でも平均する。

例えば IoU のしきい値を 0.5 とした評価では、同じクラスで IoU が基準を満たす予測を正解へ対応付けます。通常の対象では一つの正解に対する重複検出は FP となります。COCO などの評価仕様には例外規則もあるため、評価実装と設定を固定します。

mAP@0.5 と mAP@0.5:0.95 は、同じ数字として比較できません。 データの分割、クラス、入力解像度、最大検出数、評価方式を併記し、クラス別・撮影条件別の結果も確認します。用途や評価条件、制約の記録には Model Cards の考え方も参考になります。 [3]

技術指標を、現場の処理能力と組み合わせる

観点	測るもの	導入判断との関係
見逃し・誤検出	運用しきい値での Recall・Precision、単位時間あたりの FP	重大な見逃しと、人の確認件数を許容範囲にできるか
応答時間	画像取得から通知までの遅延。中央値と 95 パーセンタイルなど	遅いケースも現場の処理周期に間に合うか
業務の効果	確認時間、処理量、再作業などの変化	導入・運用の費用に見合うか

対象が写っていない画像も含めて測定します。少数例で見逃しがゼロでも、本番での保証にはなりません。評価件数を併記し、最終評価用データを繰り返し調整に使わないようにします。

開発・評価の成果物は、比較結果・条件別の失敗例・本番へ進む条件と残る課題です。

05 / 本番導入

AI の出力を、現場の次の行動につなげる

実証実験で性能を確認した後は、画像取得、結果表示、設備・業務システムとの連携を含めて検証します。映像を受け取り、認識し、人へ知らせ、対応を記録するまでを一つの処理として設計します。

正常時と例外時の流れを決める

業務フローの基本形

画像取得 → 認識 → 結果の振り分け → 自動処理または人の確認 → 対応の記録
各段階で、担当者、処理期限、入力不足や通信断が起きた際の対応を決めます。

AI が判断しにくい場合、人が確認する画面には、対象画像、検出位置、関連情報を表示します。「要確認」になった対象が溜まる場合や、担当者が不在の場合の処理も必要です。カメラ映像が途絶えたときに、対象がない状態と区別できることも確かめます。

配置と費用は、利用条件から比較する

処理を現場側で行うか、クラウド側で行うかは、許容遅延、通信量、画像の持ち出し条件、保守体制で検討します。推論だけの時間に加え、画像の転送、前処理、結果の保存・通知を含む時間を測ります。

費用はモデル開発だけではありません。カメラや照明、計算機、通信、保存、画像への正解付け、人の確認、再学習・再評価の工数を含めて見積もります。期待する効果も、削減できる作業と残る作業を分けて算出します。

実際の現場で、段階的に範囲を広げる

まず既存業務を続けながら AI の結果を記録し、実際の判断と比較する方法があります。その後、人が確認する範囲や自動処理の対象を限定して導入します。各段階で評価期間、担当、停止・継続の条件を設定します。

本番導入の成果物は、業務フロー・例外対応・費用見積もり・段階導入と復旧の計画です。

06 / 運用・改善

変化を見つけ、改善できる状態を保つ

導入時に決めた条件は、運用中も確認します。製品、カメラ位置、照明、背景などが変わる場合は、変更の記録と再評価を結び付けます。出力の件数を監視するだけでなく、正解と照合する機会を設けることが重要です。

成功例と失敗例を、次の評価へ戻す

人が訂正した結果、問い合わせ、見逃しの報告を集めます。AI が候補にした画像だけでは見逃しを発見できないため、候補にならなかった画像も抽出して確認します。対象や撮影条件ごとの偏りにも注意します。

確認するタイミング	対応する内容
日常の監視	画像取得の停止、遅延、処理件数、要確認件数、未対応の蓄積を確認する。
定期的な点検	抽出画像を正解と照合し、見逃し・誤検出・条件別の傾向を確認する。
現場の変更時	カメラ、照明、製品などの変更を記録し、影響する条件で再評価する。
モデルの更新時	従来版と同じ基準で比較し、悪化した条件を確認する。切り戻す手順も用意する。

更新を判断できる記録を残す

画像、正解、モデル、設定、評価結果を版ごとに対応付けます。いつ、何を変更し、どの条件が改善・悪化したかを追えるようにします。失敗の原因によっては、照明や運用手順の修正が対応策になるため、モデル更新以外の選択肢も検討します。

運用開始前に、担当を決める三つの仕事

現場の判断：訂正や例外に対応し、使いにくさを報告する。
システムの保守：障害を検知し、停止・復旧・問い合わせに対応する。
性能の評価：正解を確認し、再評価と更新の要否を判断する。

運用・改善の成果物は、点検項目・担当体制・評価記録・更新と切り戻しの手順です。

07 / 検討を具体化する

まず、この七つを一枚に整理してください

すべてが決まっている必要はありません。分かっていることと、確かめたいことを分けるだけでも、次の検討を具体化できます。

	整理する項目	記入する内容
01	解決したい業務課題	現在の作業、困りごと、改善したい指標
02	対象と利用範囲	何を認識するか、対象外は何か、人が担う範囲
03	手元の画像	件数、撮影条件、正解の有無、追加収集の可否
04	評価と合格条件	見逃し、誤検出、速度、確認負荷の目標
05	導入環境	カメラ、ネットワーク、処理場所、連携するシステム
06	費用と体制	予算の範囲、現場の協力、確認・保守・評価の担当
07	次に確かめたいこと	実証実験で解消する不確実性、継続・見直しの条件

構想段階から、導入後の改善までご相談ください

Convergence Lab.は、用途の整理、データ収集・評価設計、モデル開発、システム実装、本番運用と継続的な改善まで、画像認識 AI の導入に必要な各工程を支援します。

「用途の候補を比較したい」「手元の画像で実現可能性を確かめたい」「実証実験から本番へ進む条件を整理したい」など、現在の検討段階をお知らせください。上の七項目で分かる範囲や、画像の状況を共有いただくと、最初に確認する課題を整理しやすくなります。

画像認識 AI の導入・評価・改善に関するお問い合わせ
Convergence Lab. 株式会社
kimura@convergence-lab.com

参考文献

［1］ Meta. “New Segment Anything Models Make it Easier to Detect Objects and Create 3D Reconstructions.” 2025. Meta 公式発表.
［2］ Loh, Y. P. and Chan, C. S. “Getting to Know Low-light Images with The Exclusively Dark Dataset.” CVIU, 2019. arXiv:1805.11227.
［3］ Mitchell et al. “Model Cards for Model Reporting.” FAT*, 2019. Google Research.
［4］ Google. “Classification: Accuracy, recall, precision, and related metrics.” Machine Learning Crash Course.
［5］ COCO. “Detection Evaluation.” COCO 公式評価仕様.
［6］ PyTorch. “Models and pre-trained weights.” Torchvision 公式ドキュメント.
［7］ Luiten et al. “TrackEval.” 追跡評価の公式実装.
［8］ PaddlePaddle. “General OCR Pipeline.” PaddleOCR 公式ドキュメント.
［9］ Bergmann et al. “MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection.” CVPR, 2019. CVF 論文公開ページ.
［10］ CVAT. “Quality control.” CVAT 公式ドキュメント.
［11］ Hugging Face. “Image-text-to-text.” Transformers 公式ドキュメント.
［12］ Hugging Face. “Zero-shot image classification.” Transformers 公式ドキュメント.
［13］ Liu et al. “Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection.” arXiv:2303.05499.
［14］ Li et al. “Evaluating Object Hallucination in Large Vision-Language Models.” EMNLP, 2023. arXiv:2305.10355.

本資料の業務例は説明用であり、導入実績や性能保証を示すものではありません。未公開の研究結果および ExDARK の画像は掲載していません。参照情報の確認日：2026 年 11 月 6 日。