Introducing Computer Vision

COMPUTER VISION · IMAGE & VIDEO ANALYSIS · AMAZON REKOGNITION · FACES · CUSTOM LABELS · GROUND TRUTH · METRICS

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

01 **Computer vision, defined:** the **automated extraction of information from digital images** — machines identifying people, places, and things at or above human accuracy, faster and at scale. It splits into two areas, **image analysis** and **video analysis**, and is usually built on **deep learning** models.

02 **The image-analysis ladder: object classification** names the category (binary = two classes, multi-class = more); **object detection** adds *where* — a **bounding box** (top, left, width, height) plus a per-object **confidence**; **object segmentation** goes further to per-pixel boundaries (not covered in this course).

03 **Amazon Rekognition:** a **managed, deep-learning** service for image and video analysis that needs **no ML expertise** — faces, sentiment and demographics, text, unsafe content, and searchable libraries. It hosts and scales the models and integrates with **Amazon S3**, **IAM**, **Lambda**, and **Kinesis**; images are JPEG or PNG, and results return as **JSON**.

04 **Confidence rules every result:** each label, face, and detection carries a **confidence score** — choose a threshold that fits the stakes. Face **gender and emotion are inferred from the image, not identity**, and **facial recognition should only narrow the field for human review** — never decide autonomously or violate privacy.

05 **Custom domains need Custom Labels:** a pre-built model detects only what it was trained on. **Amazon Rekognition Custom Labels** trains a domain model on **a few hundred** images (**at least 2 labels, at least 10 images per label**) through a **6-step** workflow with automated ML, and you evaluate it with **precision, recall, and F1**.

02 KEY TERMS

TERM	DEFINITION
Computer vision	The automated extraction of information from digital images; often built with deep learning models
Object classification	Deciding which category or categories a picture or object belongs to (two classes = binary, more = multi-class)
Object detection	Returns the object categories plus the location of each object and a per-object confidence
Bounding box	The box surrounding a detected object, given as top, left, width, and height (as a ratio of image size)
Object segmentation	Semantic segmentation — fine-grained, per-pixel boundaries for each object; not covered in this course
Confidence score	A percentage indicating how probable it is that a detection or label is correct
Amazon Rekognition	AWS managed deep-learning service that adds image and video analysis to your applications
Facial landmarks	An array of face feature points (left eye, right eye, mouth), each with x and y coordinates
Rekognition Custom Labels	Rekognition feature that trains a model on your domain-specific images using automated ML
SageMaker Ground Truth	Service that builds high-quality labeled training datasets using a workforce and active learning
Precision	The proportion of positive predictions that were correct (raise the threshold to improve it)
Recall / F1	Recall = the fraction of test-set labels correctly classified; F1 combines precision and recall into one score

03 COMPUTER VISION TASKS

two areas, six tasks

AREA	TASK	WHAT IT DOES
Image analysis	Object classification	Names the category or categories of the image or object
Image analysis	Object detection	Categories **plus** each object's bounding box and confidence
Image analysis	Object segmentation	Per-pixel fine boundaries; not covered in this course
Video analysis	Instance tracking	Captures the path each person follows through a scene (pathing)
Video analysis	Action recognition	Studies behavior over time — e.g., shopper density and demographics
Video analysis	Motion estimation	Uses motion to recognize complex activities and thousands of objects

04 AMAZON REKOGNITION AT A GLANCE

managed, deep-learning, no ML expertise

CAPABILITY	WHAT IT DOES
Searchable libraries	Discover objects and scenes in stored images and videos
Face-based verification	Confirm identity by comparing a live image with a reference image
Sentiment & demographics	Interpret expressions (happy, sad, surprised) and details such as gender
Unsafe content detection	Flag inappropriate content in images and stored videos
Text detection	Recognize and extract text from images (Rekognition Text in Image)
Integrations	Works with S3, IAM, Lambda, Kinesis; JPEG/PNG in, JSON out

05 REKOGNITION CUSTOM LABELS WORKFLOW

train a model on your domain

01 Collect images → 02 Create training dataset → 03 Create test dataset → 04 Train the model → 05 Evaluate (precision · recall · F1) → 06 Use the model

06 TUNING A CUSTOM MODEL

threshold trades one error for the other

REDUCE FALSE POSITIVES → PRECISION

the model over-predicts your label

- **Raise** the confidence threshold (diminishing returns)
- Add the confusing class as its own label
- Fix any mislabeled training or test images

REDUCE FALSE NEGATIVES → RECALL

the model misses your label

- **Lower** the confidence threshold
- Use better, more representative examples
- Split one label into easier-to-learn classes

07 EXAM TRAPS

where the knowledge check gets you

- × “Classification and detection are the same task.” — Classification only names the category; **detection** also returns each object's **bounding box** and confidence, and **segmentation** goes to per-pixel boundaries.
- × “Rekognition can identify any object out of the box.” — It detects only what it was **trained on** (tens of millions of images, but not your niche domain); for domain objects you must train with **Custom Labels**.
- × “Rekognition's gender and emotion outputs reflect the person's true identity or feelings.” — Both are **inferred from the image only** — gender is not identity, and emotion may not match the actual emotional state.
- × “Facial recognition can make the decision automatically.” — It should only **narrow the field** of matches for a human to review; never use it autonomously where human analysis is required, or in ways that violate privacy.
- × “Raise the confidence threshold to improve recall.” — Backwards. **Raising** the threshold cuts false positives (better *precision*); **lowering** it cuts false negatives (better *recall*).
- × “You need thousands of labeled images to use Custom Labels.” — It builds on Rekognition's pretraining, so **a few hundred** domain images work — but you still need **at least 2 labels and at least 10 images per label**.
- × “Stored and streaming video use the same setup.” — Stored: async **Start/Get** operations, completion via **SNS** → **SQS**, video in **S3**. Streaming: **Kinesis Video Streams** in, a Rekognition Video **stream processor**, **Kinesis data stream** out.
- × “Ground Truth automated labeling works on any dataset size.” — It needs **large** datasets (minimum **1,250** objects, **5,000+** recommended) so the active-learning network can reach the accuracy threshold.

08 SELF-QUIZ

answers are upside-down below

Q1 What is the one-sentence definition of computer vision?

Q2 Which image-analysis task returns both an object's category and its location in the image?

Q3 A bounding box is expressed with which four values?

Q4 Name the three video-analysis tasks in computer vision.

Q5 Which AWS managed service adds deep-learning image and video analysis without requiring ML expertise?

Q6 Which two image file formats does Amazon Rekognition accept, and in what format does it return results?

Q7 Which Rekognition feature trains a model to detect objects and scenes specific to your own business domain?

Q8 To train a Custom Labels model, what is the minimum number of labels and of images per label?

Q9 Which single evaluation metric combines precision and recall into one number?

Q10 Sample exam: to detect a known face in a live camera feed, which two AWS streaming services form the pipeline with Rekognition Video?

ANSWER KEY (rotate the page)

Q1 the automated extraction of information from digital images **Q2** object detection (with a bounding box and confidence) **Q3** top, left, width, and height **Q4** instance tracking, action recognition, and motion estimation **Q5** Amazon Rekognition **Q6** JPEG or PNG images; results as JSON **Q7** Amazon Rekognition Custom Labels **Q8** at least 2 labels and at least 10 images per label **Q9** the F1 score **Q10** Kinesis Video Streams (input) and a Kinesis data stream (output)