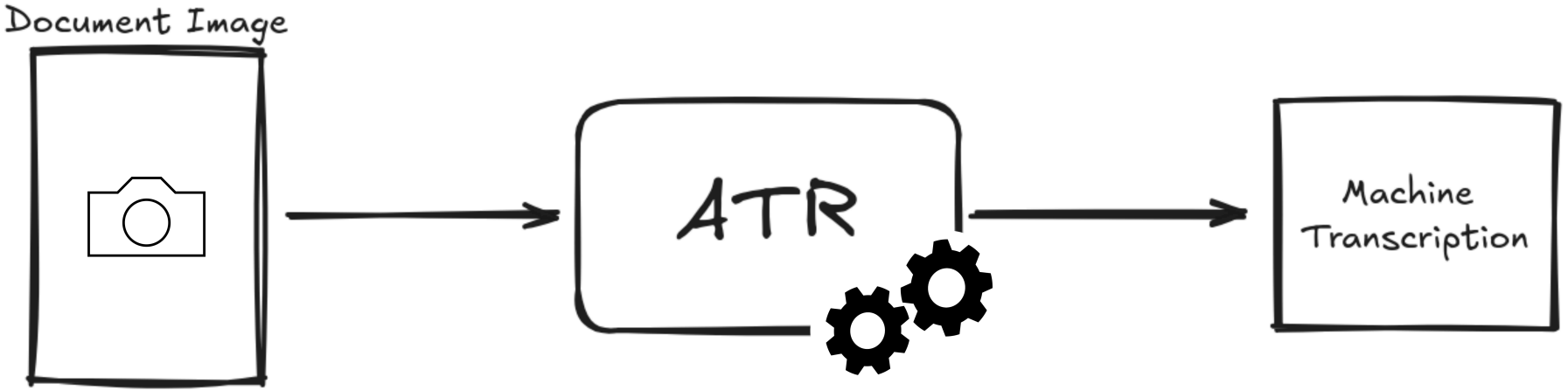
1 - Introduction to Automatic Text Recognition

What is Automatic Text Recognition (ATR)?



What about HTR and OCR?

What about HTR and OCR?

OCR

Optical Character Recognition

typically printed/typewritten material

HTR

Handwritten Text Recognition

What about HTR and OCR?

OCR

Optical Character Recognition

typically printed/typewritten material

HTR

Handwritten Text Recognition

both are forms of ATR

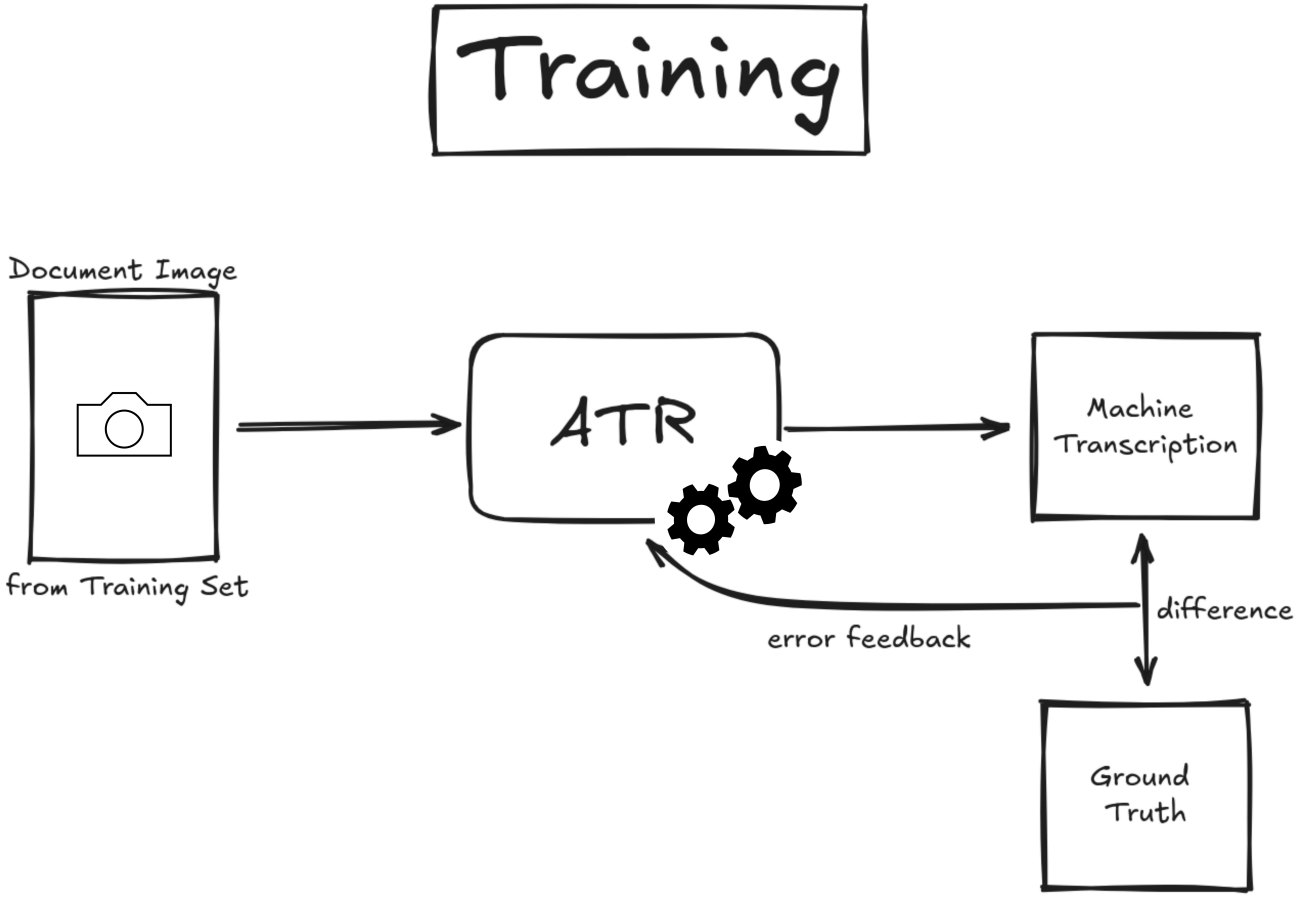
How does ATR work?

- nowadays: deep learning
- repeatedly show data to algorithm
- correct algorithm's "prediction"
- **goal**: "learn" to recognise specific patterns
  - e.g.: all of the different ways a character can be written

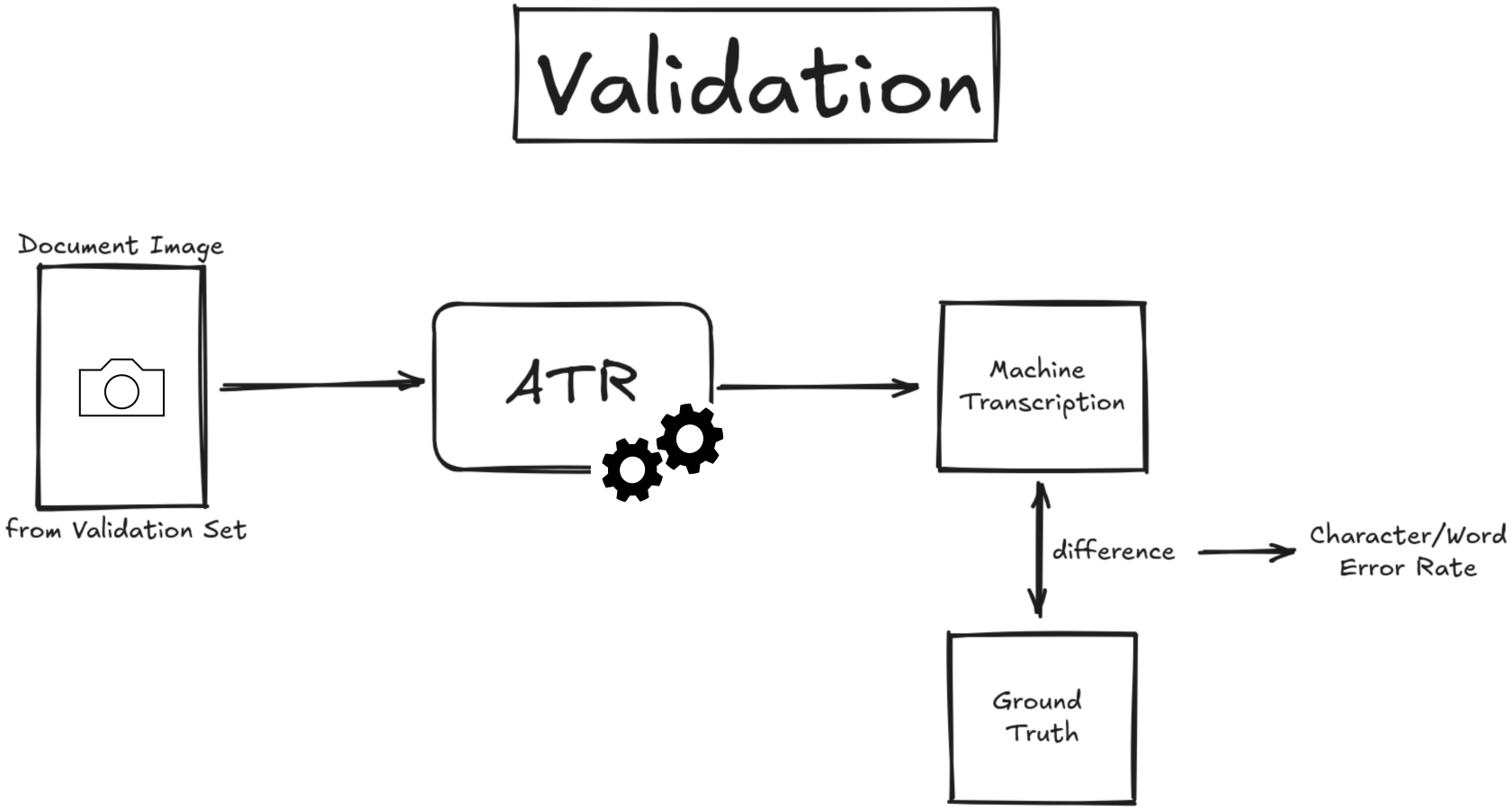
The General Deep Learning Pipeline

1. Training a model
2. Testing a model
3. Inference - i.e. using a model “in production”

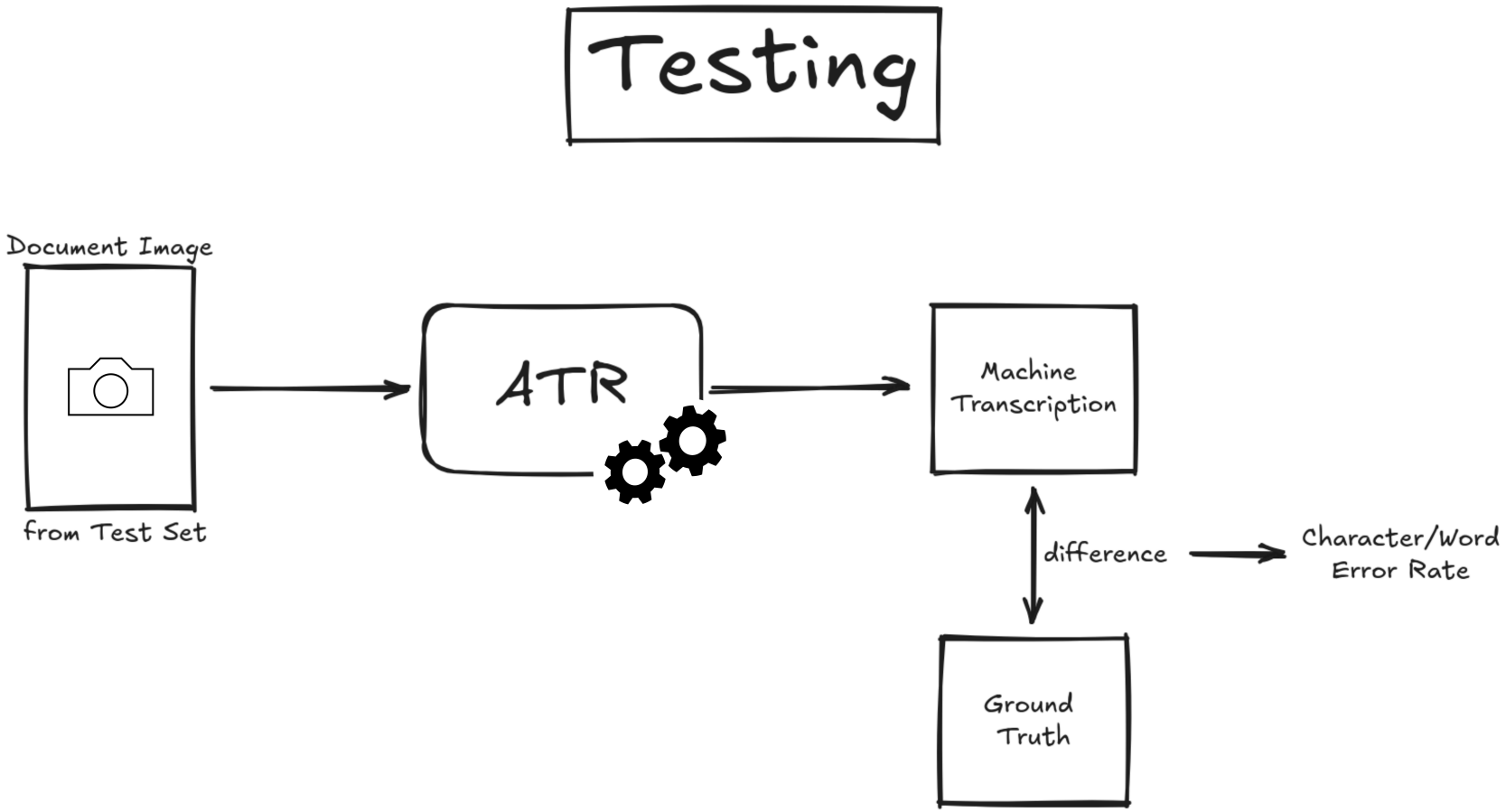
Training a Deep Learning Model



Training a Deep Learning Model -- Validation



Testing a Deep Learning Model



How to Evaluate ATR Results?

- most common metrics:
  - Character Error Rate (CER)
  - Word Error Rate (WER)
- same underlying idea

$$ER = \frac{\text{Insertions} + \text{Deletions} + \text{Substitutions}}{\text{GT Char/Word Count}}$$

Error Rates - Worked Example

Ground Truth

optisk tegngjenkjenning

Error Rates - Worked Example

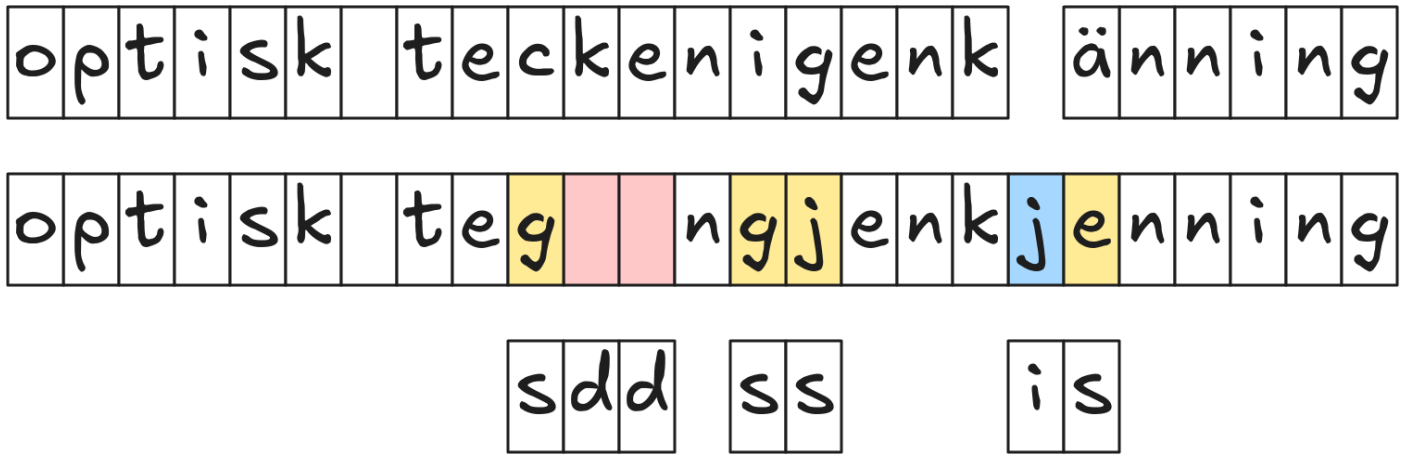
Ground Truth

optisk tegngjenkjenning

Comparison (e.g. ATR result)

optisk teckenigenkänning

Error Rates - Worked Example: CER



$$CER = \frac{1\ \text{Insertion} + 2\ \text{Deletions} + 4\ \text{Substitutions}}{23\ \text{GT Characters}} = \frac{7}{23} \approx 30.4\%$$

Error Rates - Worked Example: WER

optisk	teckenigenkänning
optisk	tegngjenkjenning

$$WER = \frac{0 \text{ Insertions} + 0 \text{ Deletions} + 1 \text{ Substitution}}{2 \text{ GT Words}} = \frac{1}{2} = 50\%$$

Data for ATR

- digital document images
- corresponding transcriptions
  - training + evaluation

Transcription Formats

- standards:
  - PAGE XML
  - ALTO XML
  - TEI XML (less frequent)

- "homebrew":
  - CSV, json
  - custom format text files

Detour: Image Segmentation

- automatically cut image into meaningful chunks
  - individual lines, words, characters
  - layout elements, e.g. paragraphs, headers, marginalia

- ATR state-of-the-art: line segmentation

- tool `Laypa` ⇒ hidden away within `Loghi`