Image Segmentation in Refinery Visual Inspection

Teacher Notes and Student Handout | YESREF Computer Vision Training

Purpose of this handout

This handout extends the refinery equipment visual inspection scenario from image classification to image segmentation. The goal is to help trainees understand why pixel-level localisation is valuable in refinery inspection tasks such as leakage detection, corrosion monitoring, physical damage assessment, and maintenance prioritisation.

1. Core Idea: From Image-Level Labels to Pixel-Level Labels

In the earlier classification setting, the model receives a refinery equipment image and predicts one overall class, such as Normal, Corrosion, Leakage sign, or Physical damage. This is useful, but it does not show the exact location, size, or boundary of the problem. Image segmentation provides a more detailed output by assigning a class label to each pixel or image region.

Task	YESREF example	Typical output
Image classification	Is this pump normal, corroded, leaking, or damaged?	One label, e.g., Leakage sign
Object detection	Where is the leaking valve or damaged component?	Bounding box around the relevant object or fault region
Semantic segmentation	Which pixels correspond to equipment, corrosion, leakage, damage, and background?	A coloured pixel-level mask over the image
Instance segmentation	If there are multiple valves or damaged regions, which mask belongs to each individual instance?	Separate masks for separate objects or fault instances
Panoptic segmentation	What is every pixel in the image, including equipment parts, defects, floor, wall, and background?	Complete scene-level pixel map

Classroom framing

A classification model behaves like an inspector who gives one final comment: “this image shows leakage.” A segmentation model behaves more like an inspector who marks the exact leakage region on the image, so Maintenance, HSSE, and Engineering teams can see where the problem is and judge its severity.

2. YESREF Example 1: Leakage Sign on a Pump

This is the main running example for introducing segmentation. A pump image may be classified as Leakage sign, but the engineer still needs to know where the leakage appears and whether it is close to a seal, flange, valve, or pipe connection.

Version	Input	Output	Limitation / benefit
Classification	Image of a pump	Leakage sign	Indicates the fault type but does not localise the leakage region.
Segmentation	Same pump image	Coloured mask showing pump body, wet/oily region, rusted bolts, and background	Localises the fault region and supports severity estimation, prioritisation, and maintenance planning.

Pixel region	Suggested label
Pump body	Equipment
Wet or oily region under the seal	Leakage sign
Rusted bolts or rusty metal patch	Corrosion
Floor, wall, or surrounding area	Background

Discussion question

If the model only predicts "Leakage sign", is that enough for Maintenance? What additional information does the coloured mask provide?

Expected direction: the mask helps locate the issue, estimate severity, support repair planning, and reduce ambiguity in the inspection report.

3. YESREF Example 2: Corrosion on a Pipe

Corrosion may cover only a small portion of a pipe, flange, bolt, or tank surface. A whole-image label may be too broad when the affected area is small. Segmentation makes the output more useful by identifying the exact surface patches that appear corroded.

Classification output	Segmentation output	Operational value
Corrosion	These exact surface patches on the pipe are corrosion.	The affected area can be inspected, measured, compared over time, and prioritised for maintenance.

Discussion question

If corrosion covers only 5% of the pipe surface, should the whole image simply be labelled "Corrosion", or should the affected region be marked?

Expected direction: marking the affected region is more informative because severity, area, boundary, and location matter in engineering decisions.

4. YESREF Example 3: Wet Surface After Cleaning vs Real Leakage

This is a strong example for showing why context matters. Classification may confuse cleaning residue and leakage because both can appear as wet or dark regions. Segmentation can help trainees reason about where the wet region starts, how it spreads, and whether it is connected to a relevant equipment component.

Case	Segmentation clue	Reasoning focus
Real leakage	Wet region starts near a joint, seal, flange, pipe connection, or valve.	The location and origin of the wet trail support a leakage interpretation.
Cleaning residue	Wider irregular wet area that may not originate from an equipment connection.	The pattern may reflect cleaning activity rather than equipment failure.

Discussion question

Look at the wet region. Is it coming from a valve, seal, flange, or pipe connection, or is it spread randomly like cleaning residue?

Expected direction: the model should use both local texture and surrounding equipment context rather than relying only on colour or darkness.

5. Why a Pixel Needs Context

A single pixel may only appear dark brown, grey, black, or shiny. By itself, that pixel may not reveal whether it belongs to corrosion, shadow, oil, background, or normal metal. The model therefore needs neighbouring pixels and wider equipment context to make a reliable prediction.

Local pixel appearance	Possible meanings without context	Context needed
Dark brown patch	Corrosion, paint stain, shadow, dirt	Is it on a pipe surface? Does it have rust texture? Is it near damaged coating?
Black or shiny patch	Leakage, shadow, wet floor, oil stain	Is it connected to a seal, flange, valve, or joint?
Grey metallic region	Normal equipment, worn metal, damaged surface	Are there cracks, dents, sharp edges, or abnormal texture?

Key teaching point

Pixel-level classification is not simply about colour. The model should learn from local texture, nearby boundaries, object structure, and the broader refinery equipment scene.

6. Sliding Window View of Segmentation

One intuitive way to introduce segmentation is to imagine a small inspection window moving across the refinery image. At each position, the model looks at the patch and predicts the class of the centre pixel.

Patch location	Centre pixel prediction
Patch around rusty pipe surface	Corrosion
Patch around oily stain under a flange	Leakage sign
Patch around clean metal	Normal equipment
Patch around floor or wall	Background

Discussion question

Why would this become slow for every pixel in a high-resolution refinery image?

Expected direction: many overlapping patches repeat similar computation. This motivates a fully convolutional approach that processes the whole image more efficiently.

7. Fully Convolutional Segmentation

A more efficient approach is to pass the whole refinery image through a convolutional network once and produce a full pixel-level prediction map. The network first extracts visual features from the whole image, then outputs a mask with the same spatial size as the input image.

Input image	Network processing	Output mask
Pump, pipe, valve, flange, tank surface, or compressor area	Convolutional layers learn equipment structure, defects, texture, and context	Each pixel receives a label such as equipment, corrosion, leakage sign, physical damage, or background

Teacher note

The key transition is from repeated patch-by-patch classification to a single image-level forward pass that produces dense predictions for all pixels.

8. Downsampling and Upsampling in Segmentation Networks

Segmentation networks commonly use an encoder-decoder structure. The encoder compresses the image to understand the overall scene, while the decoder upsamples the learned features back to image resolution so that a pixel-level mask can be produced.

Network stage	Refinery inspection analogy	Purpose
Downsampling / encoder	The inspector steps back to understand the whole equipment scene.	Captures global context such as pump area, pipe route, valve location, and surrounding environment.
Bottleneck	The inspector forms a high-level interpretation of the scene.	Represents abstract evidence about equipment type, defect possibility, and fault context.
Upsampling / decoder	The inspector zooms back in and marks the exact leakage, corrosion, or damage pixels.	Recovers spatial resolution and produces the final segmentation mask.

Important caution

Downsampling helps the model understand the big picture, but it may lose fine spatial details. Thin cracks, small rust patches, and narrow leakage trails are especially vulnerable to this loss.

9. Residual / Skip Connections and U-Net

Skip connections help the decoder recover fine spatial details by bringing earlier high-resolution features from the encoder directly into later upsampling layers. This is valuable when the target defects are small, thin, or boundary-sensitive.

Issue caused by downsampling	How skip connections help	YESREF example
Small features may disappear.	Earlier high-resolution features are reused during decoding.	Small corrosion patches on pipe surface.
Edges and boundaries may become blurry.	Spatial detail from shallow layers improves boundary recovery.	Leakage boundary under pump seal.
Thin structures may be missed.	Fine-grained information supports precise localisation.	Hairline crack or narrow wet trail.

Classroom summary line

Deep layers help the model understand what is present; early layers help it remember where exactly edges and small details are.

U-Net combines an encoder, a decoder, and skip connections. In the YESREF setting, the encoder learns equipment context, the decoder reconstructs the pixel-level defect map, and the skip connections carry sharp spatial details such as corrosion boundaries, crack edges, and leakage outlines.

10. Trainee Activity: Manual Pixel-Level Reasoning

Use a refinery image or describe the following scene: a pipe section has a dark stain near a flange, rust-coloured marks on the lower side, and background floor. Trainees decide suitable labels for different regions before discussing how a segmentation model would learn the same task.

Region in image	Suggested student label
Pipe metal	Normal equipment
Rust-coloured area	Corrosion
Dark wet trail near flange	Leakage sign
Crack or dent	Physical damage
Floor, wall, or background	Background

Discussion question

Would classification, object detection, or segmentation be most useful for this image?

Expected direction: segmentation is most useful when the exact fault region, size, boundary, and location matter. Detection may be enough when an approximate region is sufficient. Classification is useful for broad triage but gives the least localisation detail.

11. Consolidated Summary

In refinery equipment inspection, classification predicts the overall condition of an image, while segmentation identifies where different conditions appear inside the image. For a pump image, segmentation can label the pump body as equipment, a wet stain near the seal as leakage, a rusty bolt as corrosion, and the floor as background. This makes the model output closer to real inspection practice because engineers need the fault type, location, size, and boundary. A single pixel is often ambiguous, so segmentation models use convolutional context from surrounding pixels and the full equipment scene. Encoder-decoder networks compress the image to understand the scene and then upsample features to recover the mask. Skip connections, as used in U-Net, help preserve fine spatial details such as cracks, rust patches, and leakage boundaries.

12. Quick Reference for Delivery

Concept	One-line refinery interpretation
Classification	One decision for the whole image.
Segmentation	A decision for every pixel or region.
Context	A pixel needs surrounding evidence to be interpreted correctly.
Sliding window	A simple but inefficient way to classify centre pixels patch by patch.

Fully convolutional network	Processes the full image once and predicts dense pixel labels.
Downsampling	Captures the big picture but may lose detail.
Upsampling	Restores spatial resolution to produce a mask.
Skip connection	Carries high-resolution detail from encoder to decoder.
U-Net	Encoder-decoder segmentation model with skip connections.