1. Model Setup

- **Model**: U-Net
- **Loss function**: `BCEWithLogitsLoss()` — applies binary cross-entropy **with logits**, making it suitable for **multi-label segmentation** (each pixel can belong to more than one class).
- **Optimizer**: `Adam`
- **AMP**: **automatic mixed precision (AMP)**, which speeds up training and reduces memory usage by using float16 operations where safe.

2. Loss Function

`BCEWithLogitsLoss` is used, which is:

$$\mathcal{L}(x, y) = -[y \cdot \log(\sigma(x)) + (1 - y) \cdot \log(1 - \sigma(x))]$$

Where,

x is the raw model output (logits), and y is the target mask.
The sigmoid function $\sigma(x)$ maps logits to probabilities.
This loss is computed **per pixel per class** — ideal for your six-channel output.

3. Metrics

Metric	Use
IoU (Intersection over Union)	Measures overlap between predicted and actual mask
Dice Coefficient	Another overlap metric, more sensitive to small regions
Precision / Recall	Helpful when analyzing specific classes (e.g. defect types)