UNet Model

Core Concepts of the U-Net Model

- **Semantic Segmentation**: The model classifies each pixel in an image into one of several classes (6 classes for defects).

- **Encoder-Decoder Architecture**: Compresses and then reconstructs the image to perform dense prediction.

- **Skip Connections**: Help preserve spatial resolution by combining encoder feature maps with decoder feature maps.

Model Components and Architecture

- **DoubleConv:** This block performs two convolutional layers back-to-back, each followed by Batch Normalization and ReLU.
- **Parameters**: `in_channels=1`: Grayscale image input.
  `out_channels=6`: Output has 6 segmentation masks, one per defect class.
- **Encoder (Downsampling Path):**
  Each level: Applies `DoubleConv`.
  Saves the output (for skip connections).
  Reduces spatial resolution using `MaxPool2d`.
- **Bottleneck:** This layer acts as the "bridge" between the encoder and decoder — capturing the most abstracted features.
- **Decoder (Upsampling Path):** This helps restore spatial resolution while retaining detailed features from the encoder.
  Each decoder step: Applies a `ConvTranspose2d` (upsampling by a factor of 2).
  Concatenates with the corresponding encoder output
  Applies `DoubleConv`.

- **Final Layer:** Outputs a 6-channel image (one for each class).

Statistics and Model Size

Layer	Channels	Params (approx)
DoubleConv-1	1 → 64	~38K
DoubleConv-2	64 → 128	~221K
DoubleConv-3	128 → 256	~885K
DoubleConv-4	256 → 512	~3.5M
Bottleneck	512 → 1024	~14.2M
Decoder total	Multiple blocks	~18M
Final conv	64 → 6	~390
Total		~36M params