NUST

EFFICIENT TIMBER SURFACE DEFECT DETECTION USING A COMPACT YOLO ARCHITECTURE FOR RESOURCE-CONSTRAINED DEVICES

Team Members: Sara Adnan Ghori (411228) & Umar Farooq (406481)
Supervisor: Dr. Hashir Moeed Kiani

PROBLEM STATEMENT

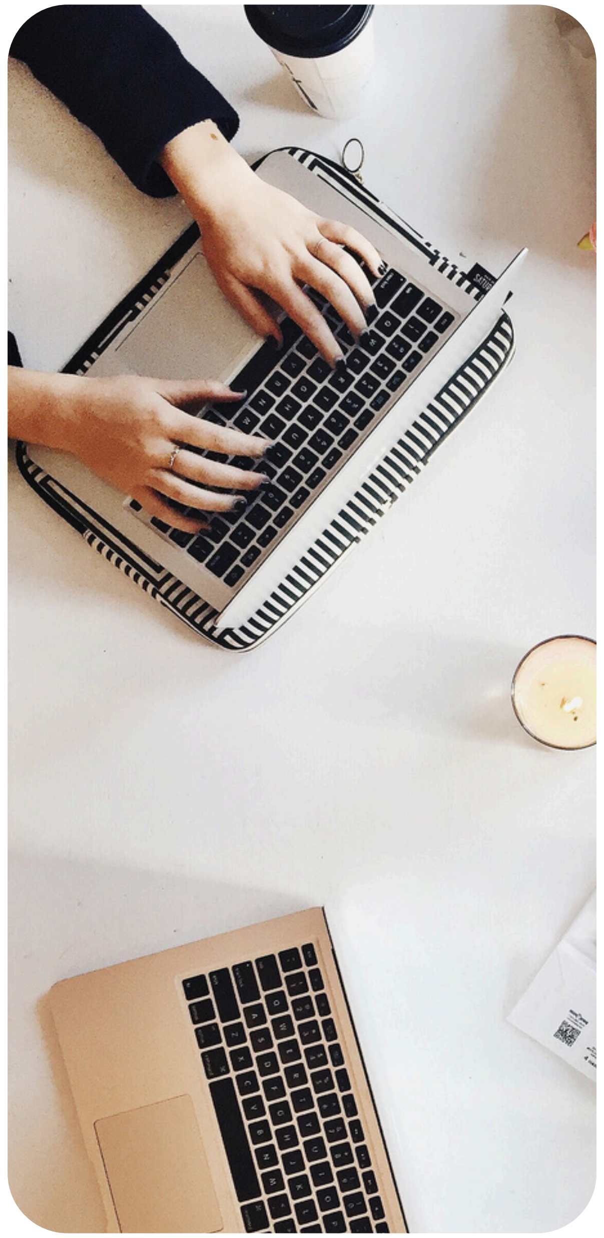
01

Standard timber inspection is largely **manual**, making it **prone to human error** and dependent on operator stamina. Although deep learning-based models offer potential for automated inspection, most **existing approaches are trained on generic wood types** and therefore **perform poorly on timber-**specific tasks. Furthermore, high-performance detectors such as SGN-YOLO and WDNet-YOLO are **computationally intensive** and unsuitable for real-time deployment on edge devices, while **lighter models fail to deliver acceptable accuracy.**

GOAL

02

Develop a lightweight, timber-specialized YOLO architecture optimized for resource-constrained edge deployment without sacrificing accuracy on small or irregular defects.



LITERATURE REVIEW & IDENTIFIED GAPS

Current State:

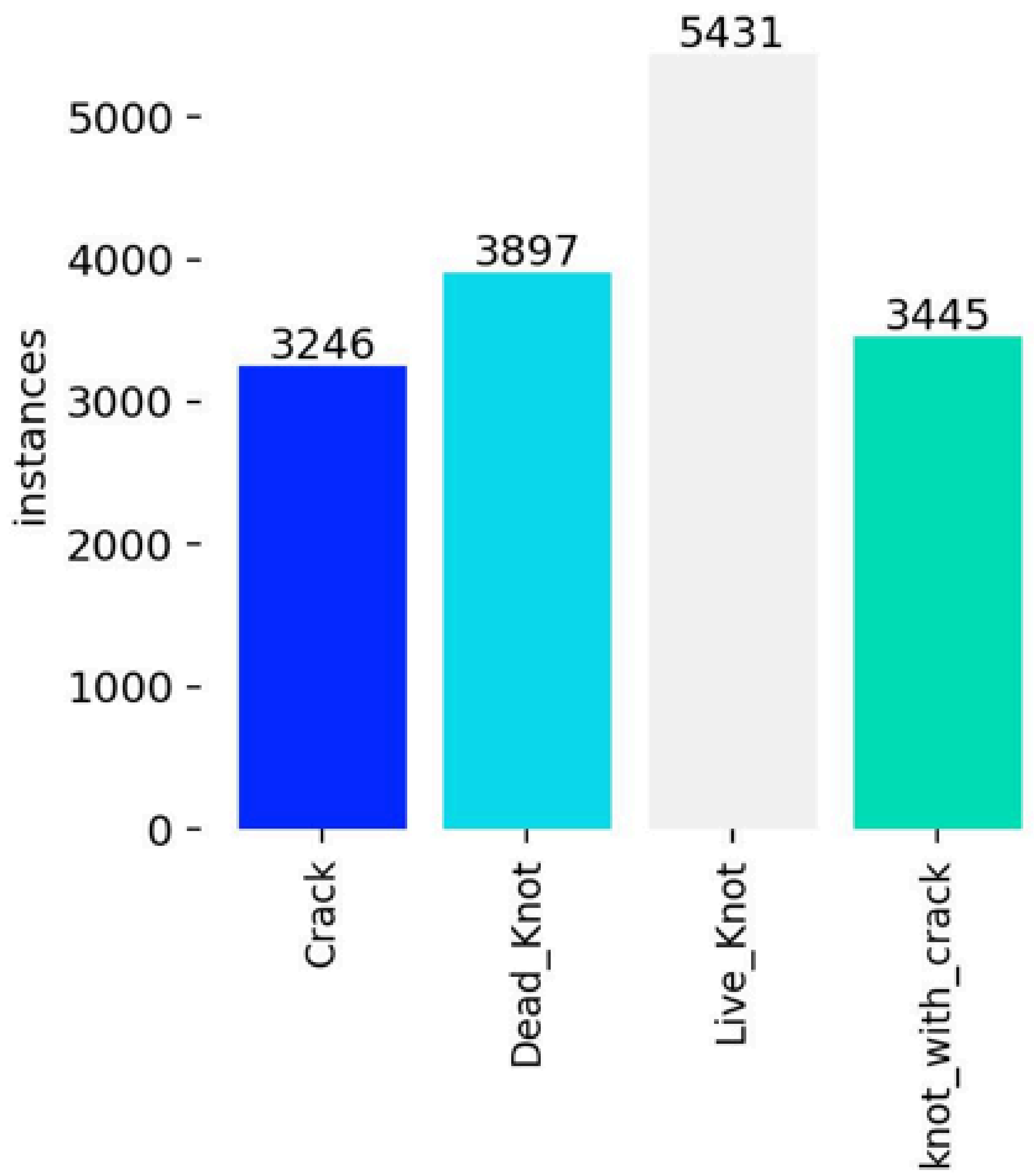
- Advanced models like SGN-YOLO and WDNET-YOLO use heavy attention and large backbones, leading to high inference latency. Lightweight methods usually trade accuracy over computations.

Gaps:

- Excessive computational requirements fot SOTA Models.
- Weak performance of ultra-light models on fine defects.
- Scarcity of timber-specific detection models.
- Lack of real-time solutions for edge cards in industrial pipelines.

DATASET OVERVIEW

- **Base Dataset:** Roboflow "defects-in-timber".
- **Classes:** Crack, Dead_Knot, Live_Knot, and Knot_with_Crack.
- **Augmentation:** Albumentations (Flip, Rotate, Shift Scale)
- **Final Statistics:** The training set reached 5,431 instances for Live Knots and 3,246 for Cracks



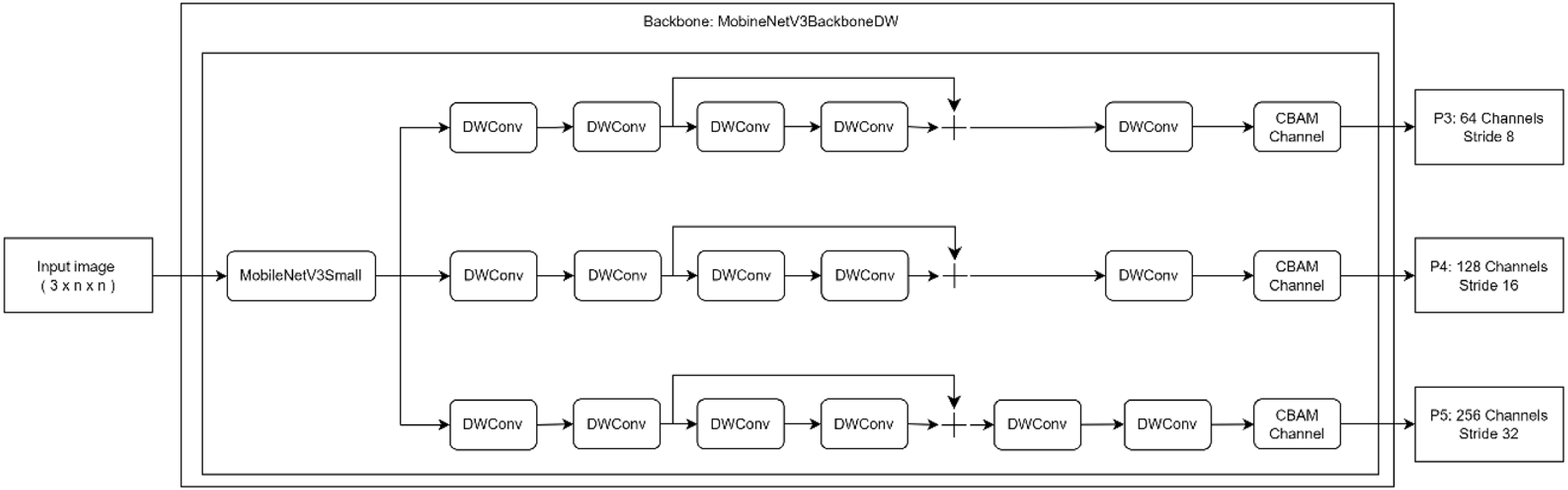
HIGH-LEVEL PROPOSED ARCHITECTURE

- **Core Philosophy:** A three-component system combining efficiency with multi-scale feature extraction.

Three Pillars:

1. **Backbone**: Enhanced MobileNetV3 Small with multi-scale refinement.
2. **Neck:** UltraLite FPN-PAN style network with attention and a P5 Transformer.
3. **Detection Head:** YOLOv8-style anchor-free predictor.

THE BACKBONE: MOBILENETV3BACKBONEDW



THE BACKBONE: MOBILENETV3BACKBONEDW

Base: MobileNetV3 Small for its edge-friendliness.

Hierarchical Features: Extracts maps at three stride levels:

- P3 (Stride 8): Fine details for small cracks.
- P4 (Stride 16): Mid-level texture and shape.
- P5 (Stride 32): High-level semantic context for large knots.

Multi-Stage Refinement: Each scale includes 4-6 sequential refinement blocks (e.g., 5 stacks for P3, 6 for P5).

BACKBONE INNOVATION – DWCONVCUSTOM & CBAM

DWConvCustom

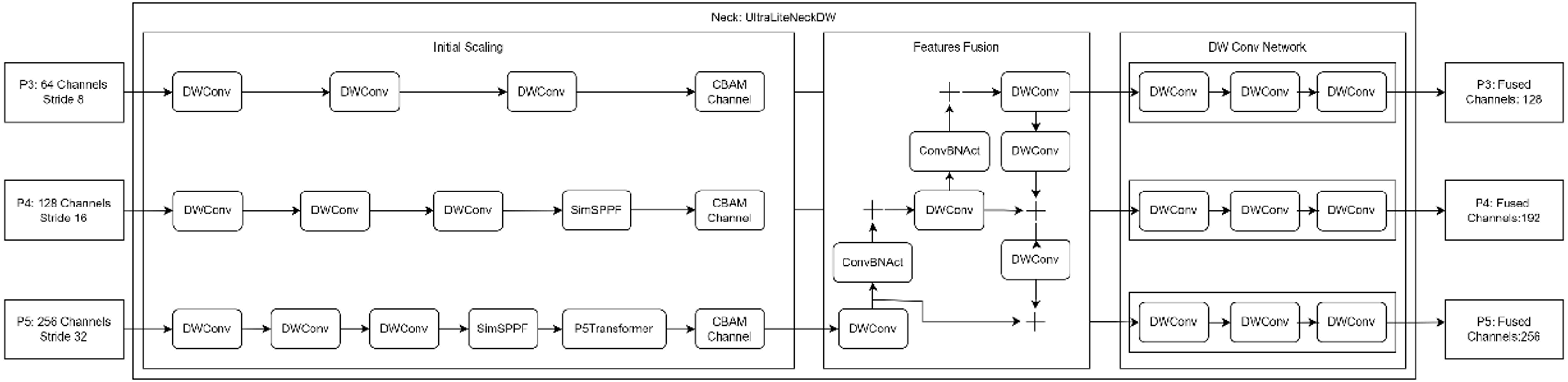
Uses Depthwise Separable Convolutions to factorize 3X3 (filtering) and 1X1 (mixing) operations, reducing FLOPs while allowing deeper stacks

CBAM Attention

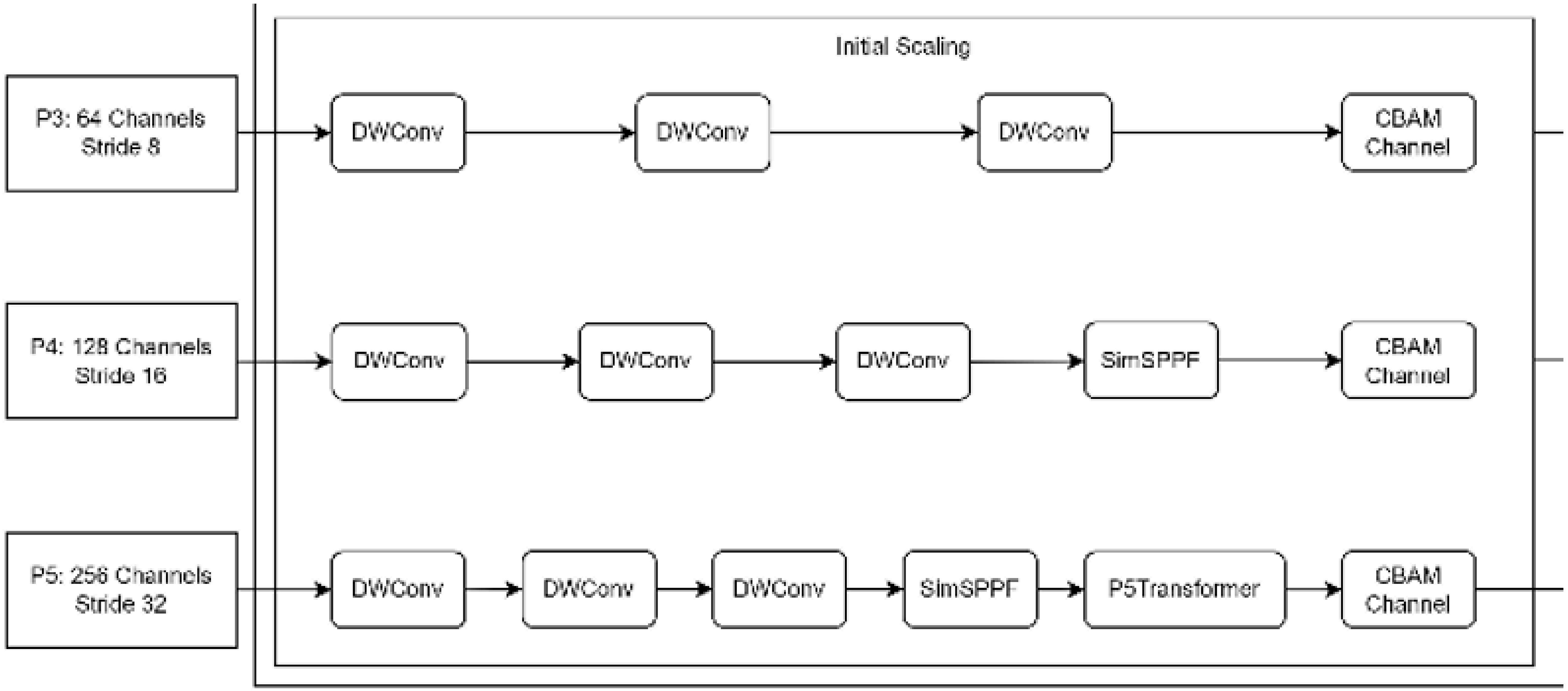
Every scale output (P3-P5) is passed through a CBAM module:

- **Channel Attention:** Highlights discriminative defect channels.
- **Spatial Attention**: Suppresses background wood grain to focus on defect-prone regions

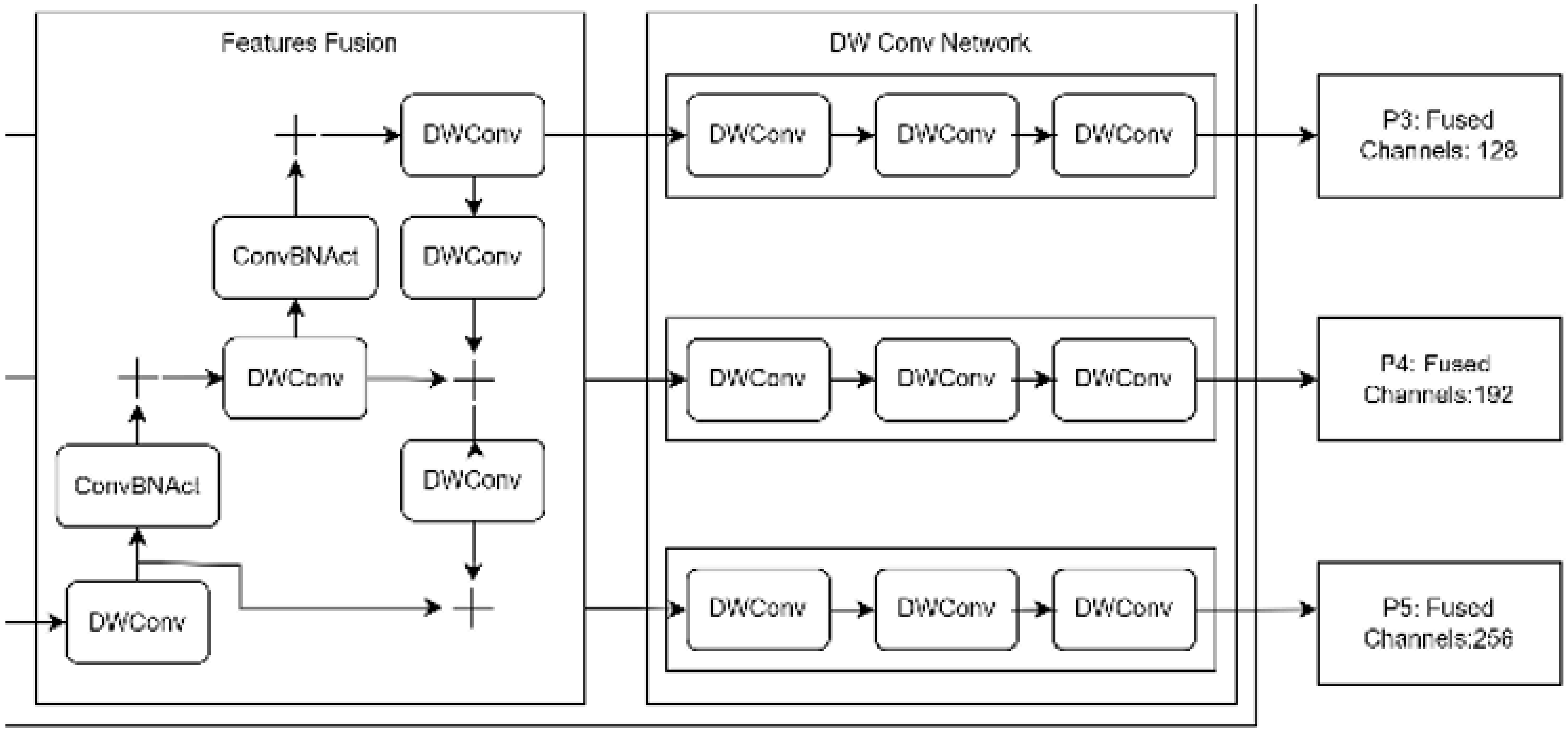
THE NECK: ULTRALITENECKDW



THE NECK: INITIAL SCALING



THE NECK: FEATURE FUSION



THE NECK: ULTRALITENECKDW

- **Fusion Strategy:** Combines Top-Down (FPN) and Bottom-Up (PAN) paths.
- **Top-Down (FPN)**: Up-samples P5 to fuse with P4 and P3, injecting context into fine-resolution maps to detect tiny cracks.
- **Bottom-Up (PAN):** Re-samples maps upward to preserve spatial flow and multi-scale richness.
- **SimSPPF:** A simplified Spatial Pyramid Pooling block provides global context with minimal overhead.

THE NECK: P5 TRANSFORMER

Why a Transformer? Added specifically to the P5 layer to improve long-range context modeling for large, irregular knots.

Lightweight Design: Includes 1X1 projection, LayerNorm, and 2 encoder layers to maintain low computational cost.

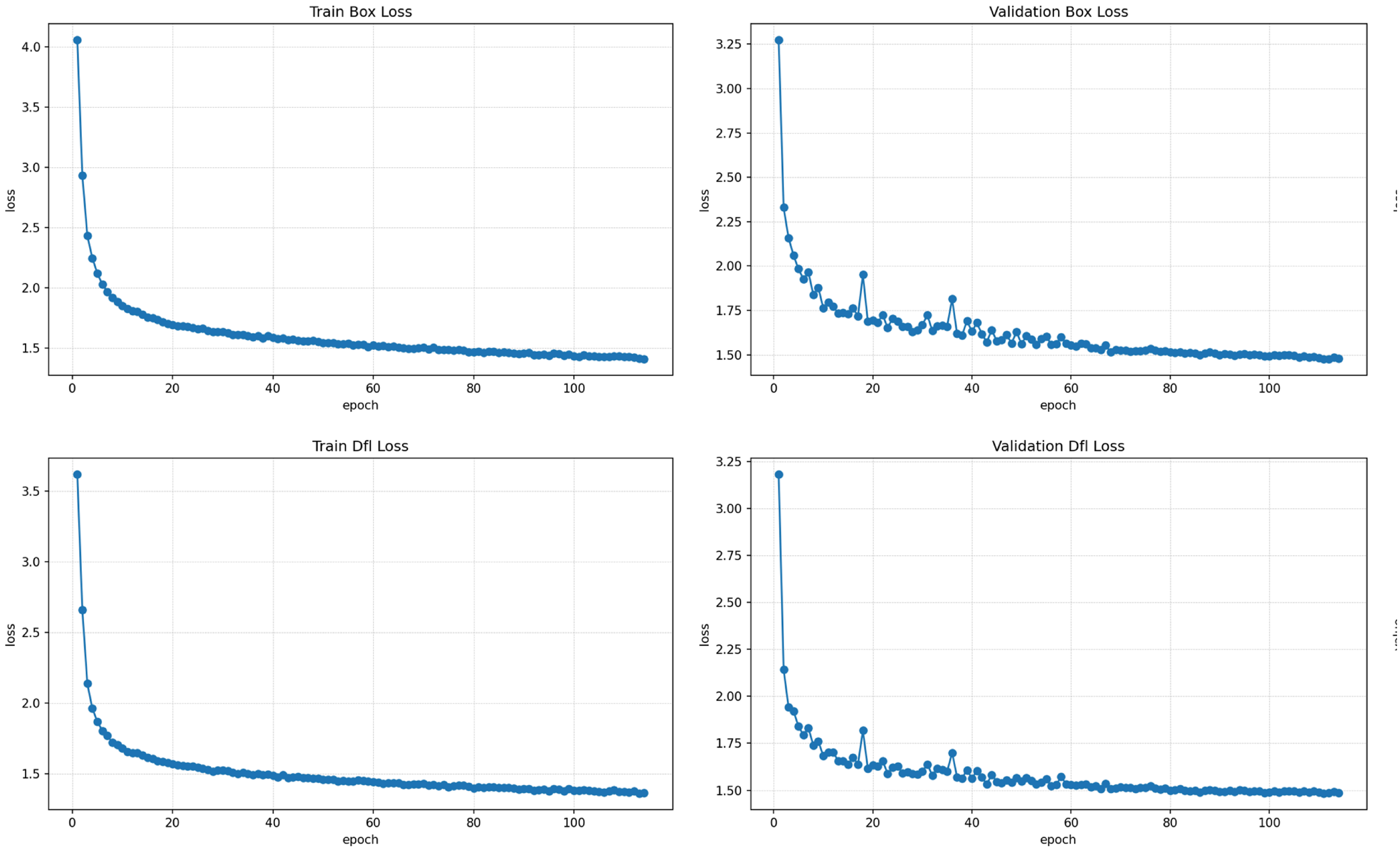
Result: Better detection of spatially complex defects that standard convolutions might miss.

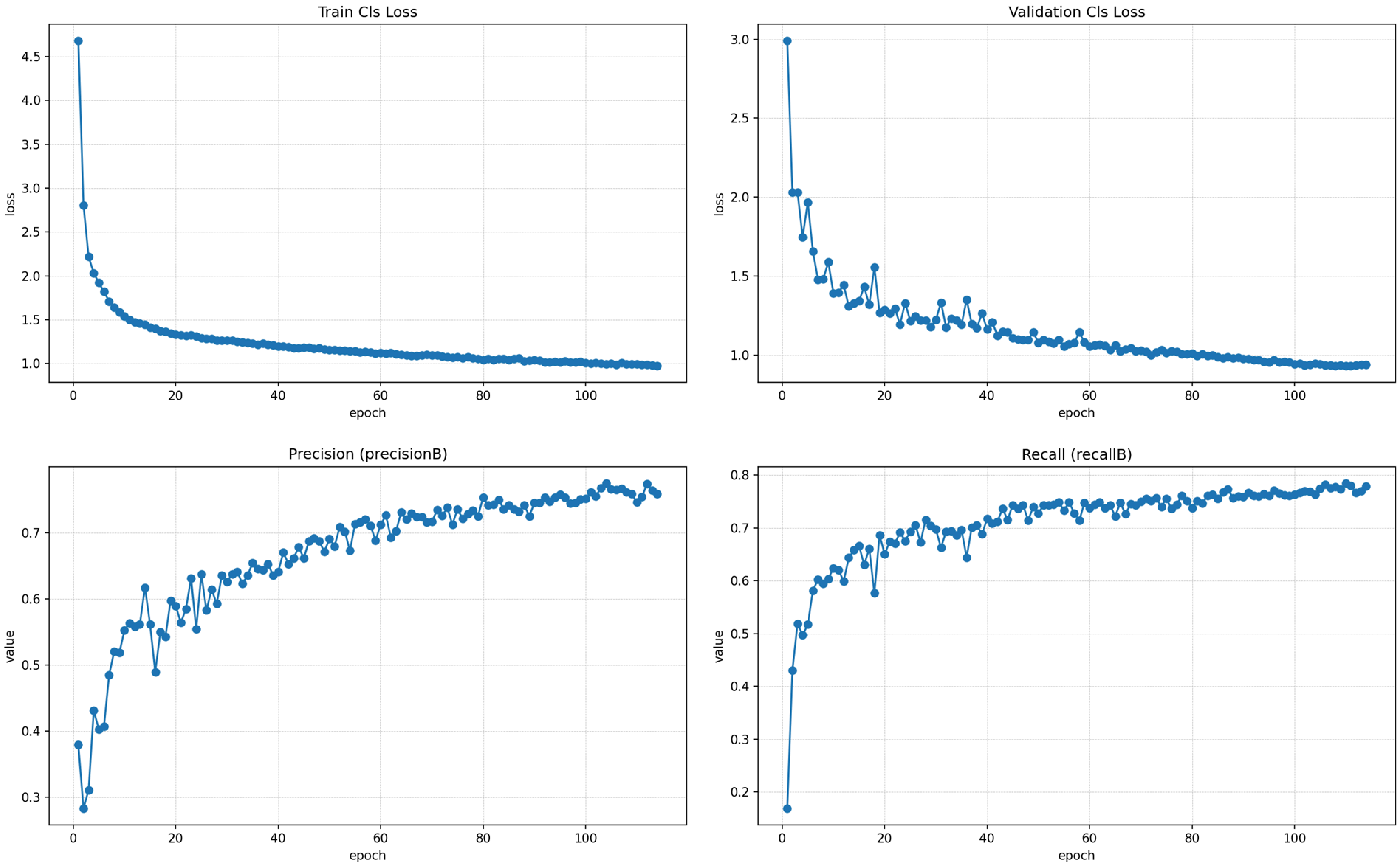
THE DETECTION HEAD: STANDARD YOLO-V8 DETECTION HEAD

- **Standard_YOLOv8:** Uses YOLOv8-style decoding to produce bounding box offsets and class probabilities.
- **Significance:** Simplifies computation and significantly improves performance on irregular-shaped defects like knots.
- **Multi-Scale Predictions:** Produces three maps corresponding to stride 8 (small), 16 (medium), and 32 (large defects/clusters).

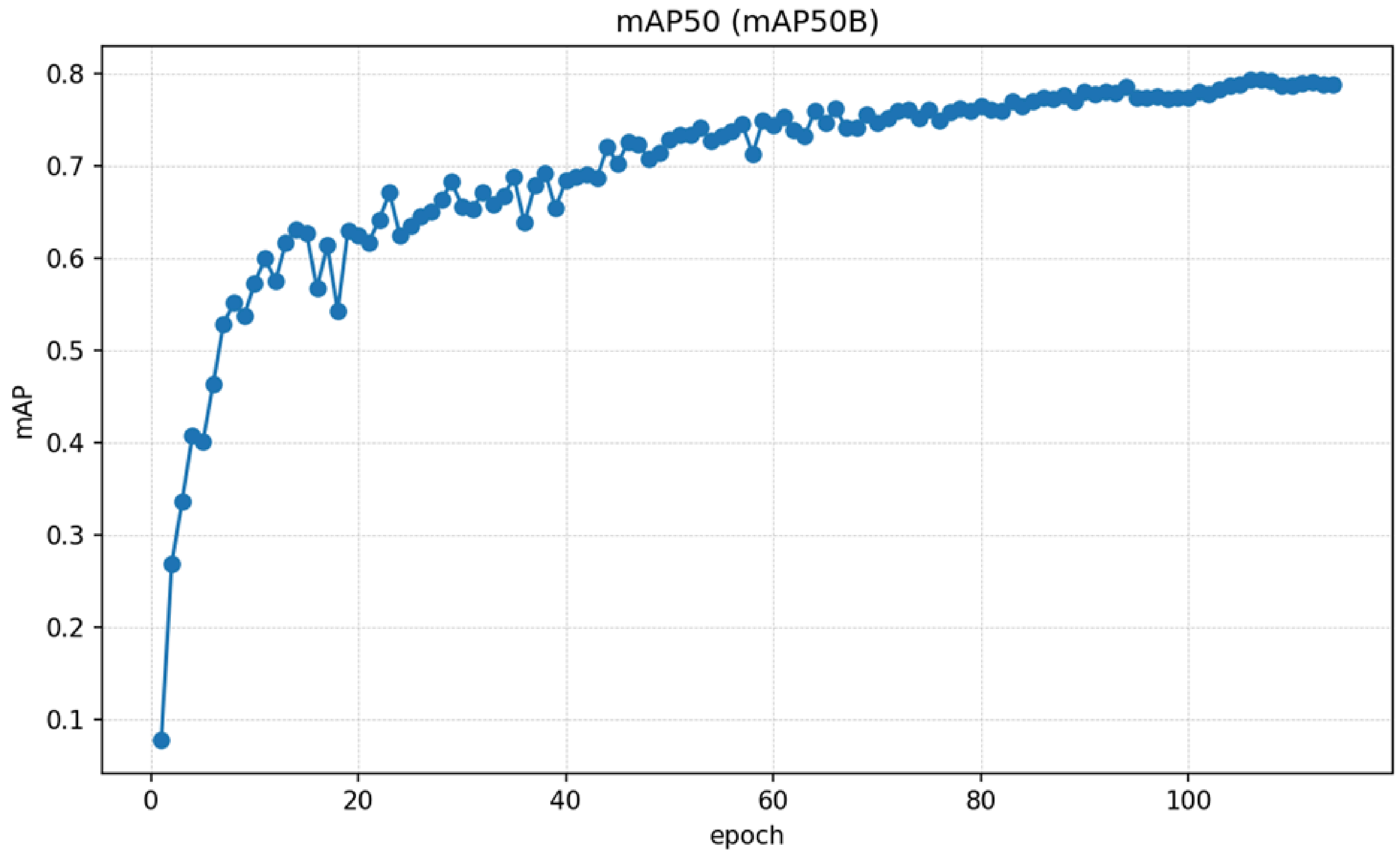
TRAINING SETUP & METHODOLOGY

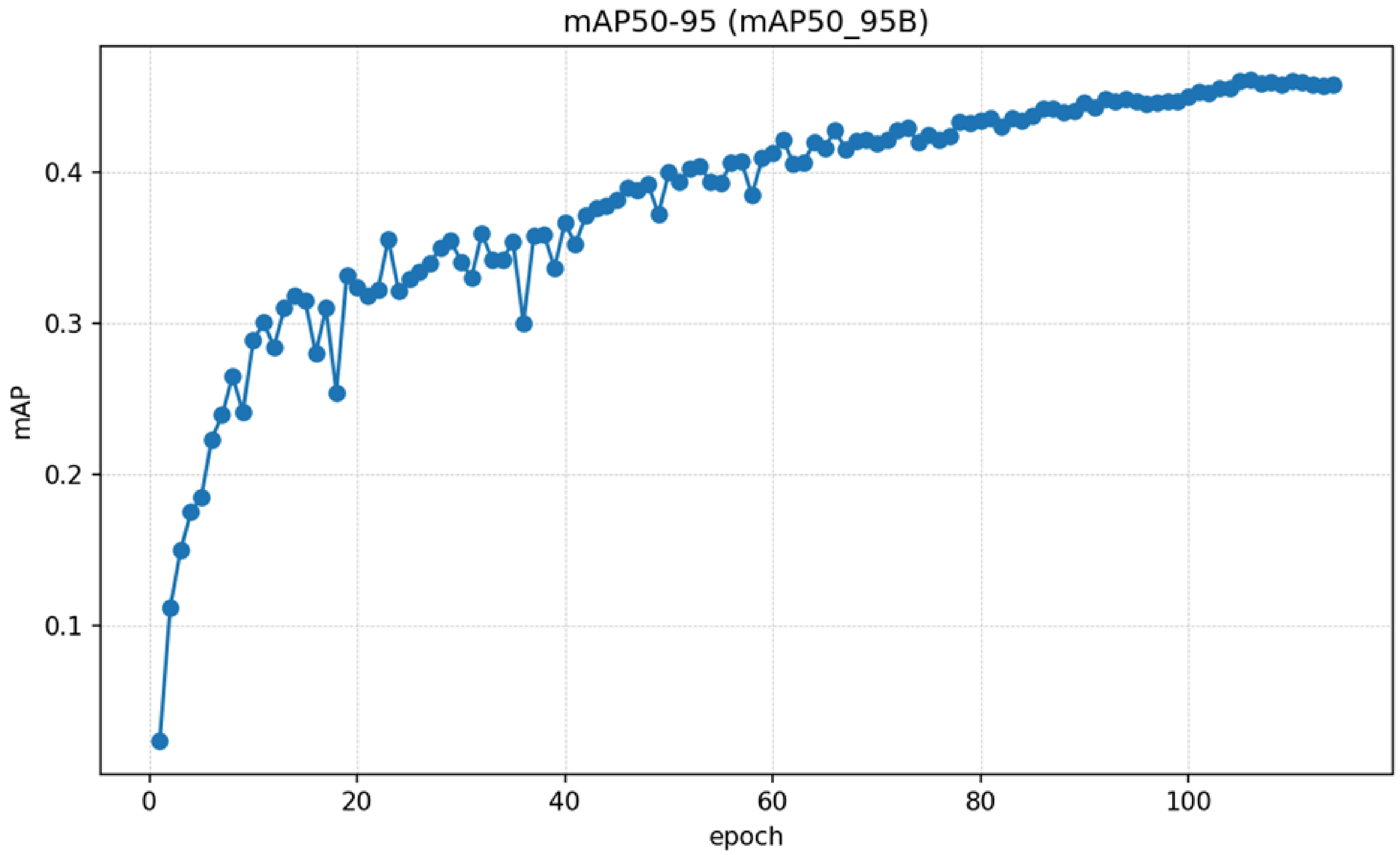
- **Optimization:** SGD optimizer with cosine learning rate decay and a 5-epoch warmup.
- **Loss Functions:** A composite of CIoU (Box Loss), BCE (Classification Loss), and Distribution Focal Loss (DFL) for fine-grained localization.
- **Metrics:** mAP@50 and mAP@50-95 used to measure overall accuracy and localization precision.


Train Box Loss
loss
epoch
Validation Box Loss
loss
epoch
Train Dfl Loss
loss
epoch
Validation Dfl Loss
loss
epoch



Train Cls Loss
Validation Cls Loss
Precision (precisionB)
Recall (recallB)
loss
value
epoch


mAP50 (mAP50B)
mAP
epoch
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
20
40
60
80
100


mAP50-95 (mAP50_95B)
mAP
epoch
0.1
0.2
0.3
0.4
0
20
40
60
80
100

RESULTS – EFFICIENCY & COMPLEXITY

- **Parameters**: 4.24 Million.
- **Computational Cost**: 4.46 GFLOPs (2.23 GMACS).
- **Inference Speed**: Average of 14.9 ms per image, confirming real-time readiness for industrial pipelines.
- **Summary**: Much less complex than traditional YOLO detectors while remaining robust

Performance metrics over 114 epochs

- **Precision (P): 0.38 → 0.76**
- **Recall (R): 0.16 → 0.77**
- **mAP@50 (B): 0.07 → 0.78**
- **mAP@50-95 (B): 0.02 → 0.46**

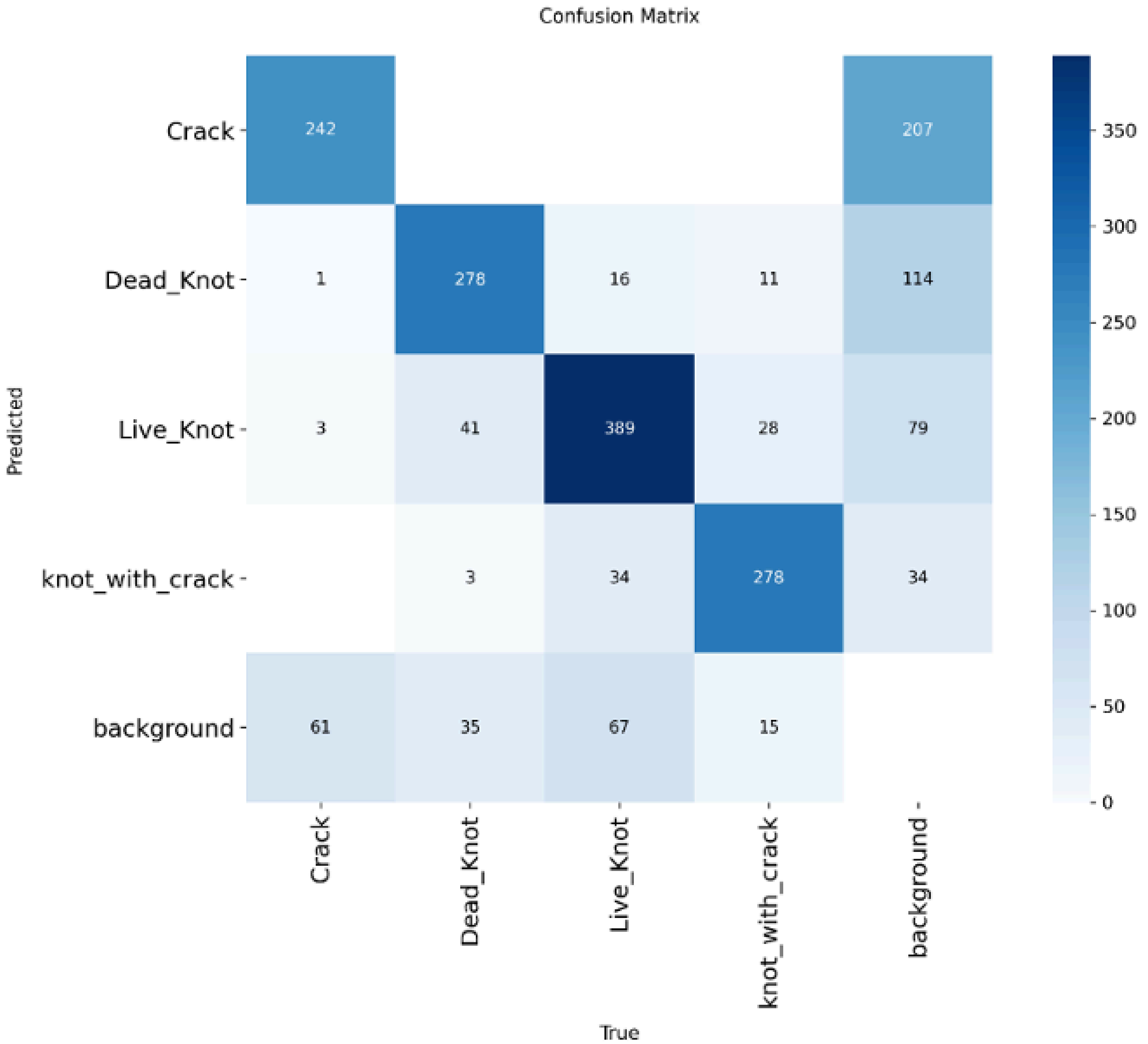
TEST METRICS

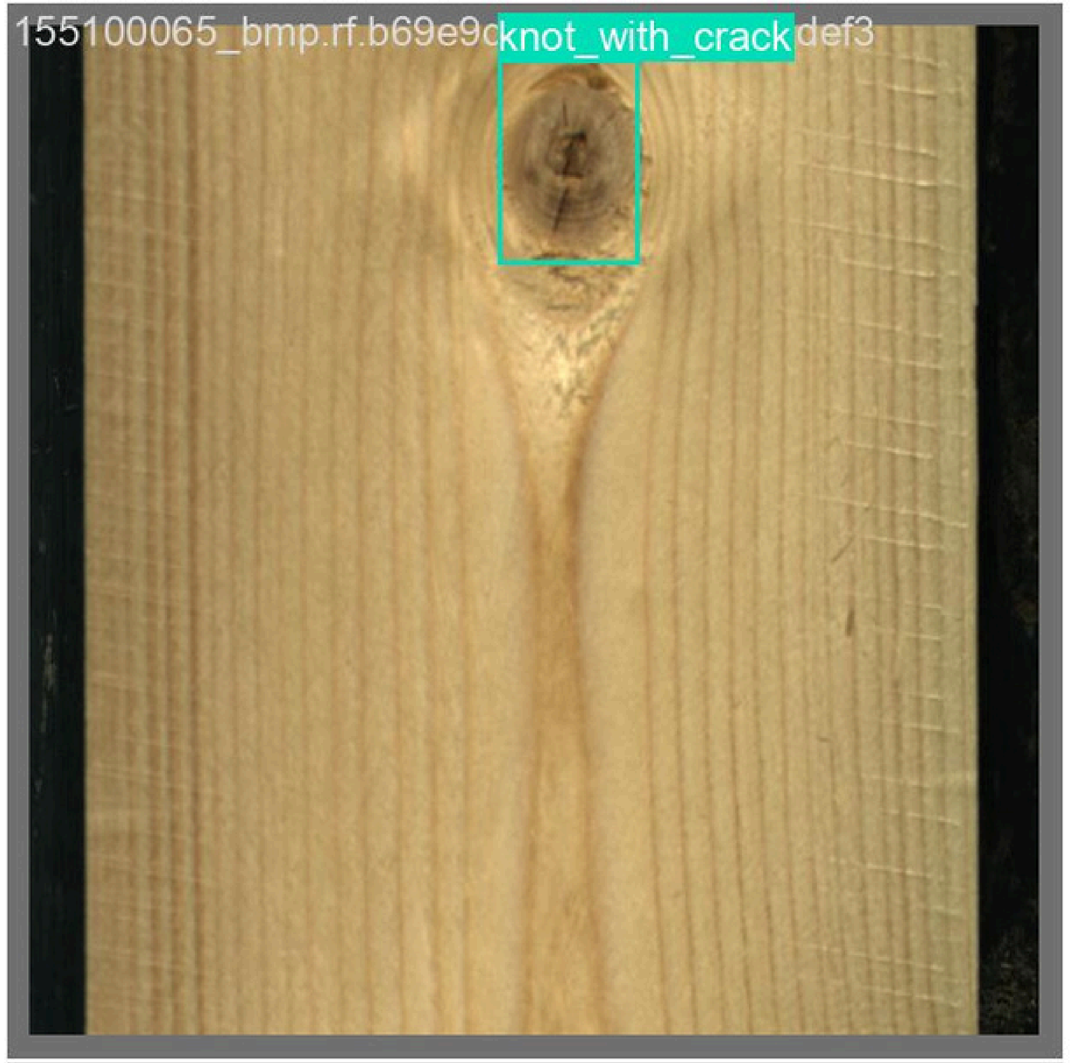
Class	Images	Instances	Precision	Recall	mAP@50	mAP@50–95
Crack	254	307	0.709	0.707	0.711	0.337
Dead_Knot	263	357	0.738	0.767	0.758	0.426
Live_Knot	305	506	0.800	0.737	0.806	0.416
knot_with_crack	269	332	0.849	0.898	0.923	0.672
All Classes	627	1502	0.774	0.777	0.800	0.463

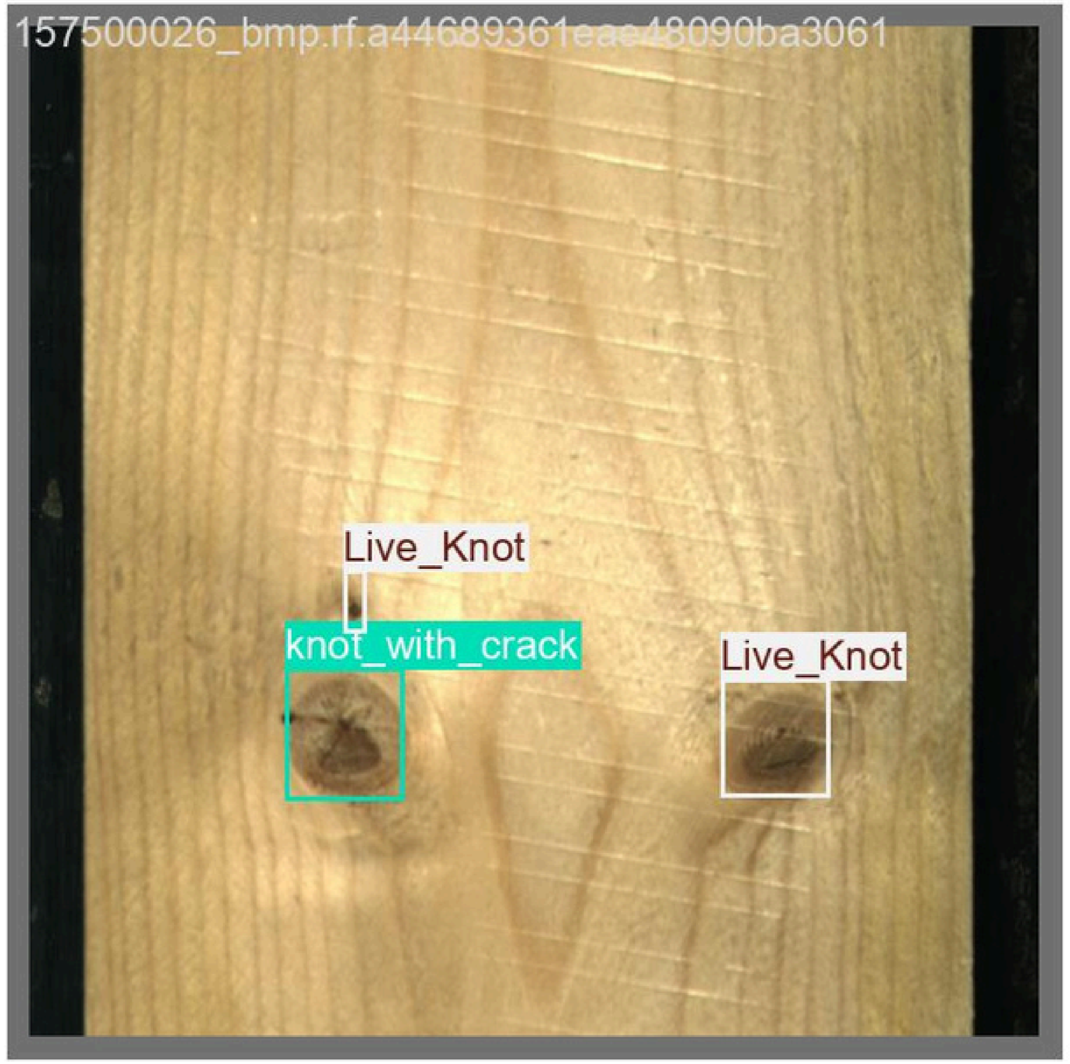
RESULTS – PERFORMANCE FINDINGS

- **Validation mAP@50:** 0.794.
- **Test mAP@50:** 0.800.
- **Key Observations:**
- **Best Performance:** On "Knot with Crack" (0.923 mAP@50) due to effective contextual modeling.
- **Challenges:** Small "Cracks" showed lower recall, reflecting the difficulty of fine-grained detection on textured wood.
- **Confusion Matrix:** Some misclassification between Dead Knots and Live Knots due to similar visual appearances.

CONFUSION MATRIX – TEST

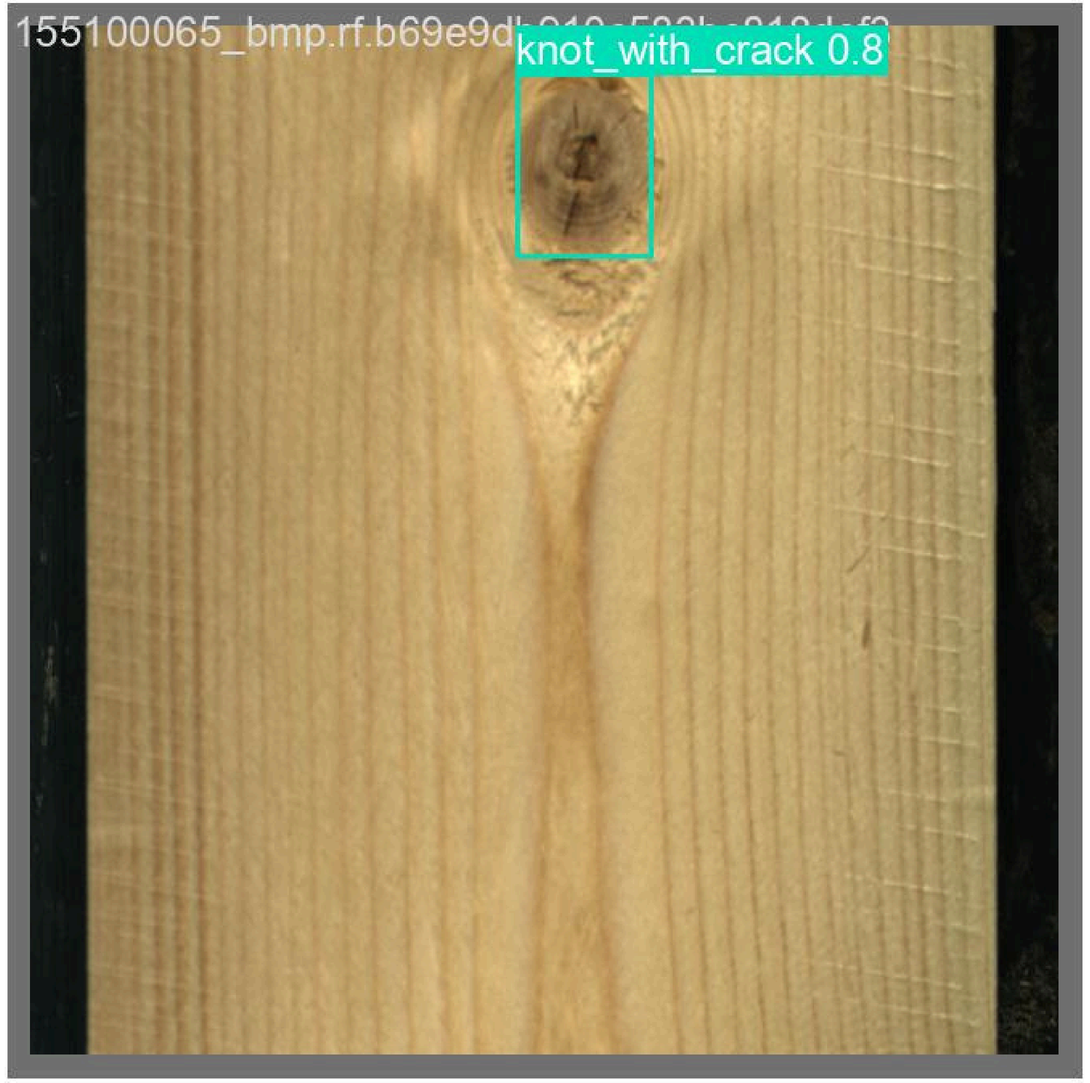


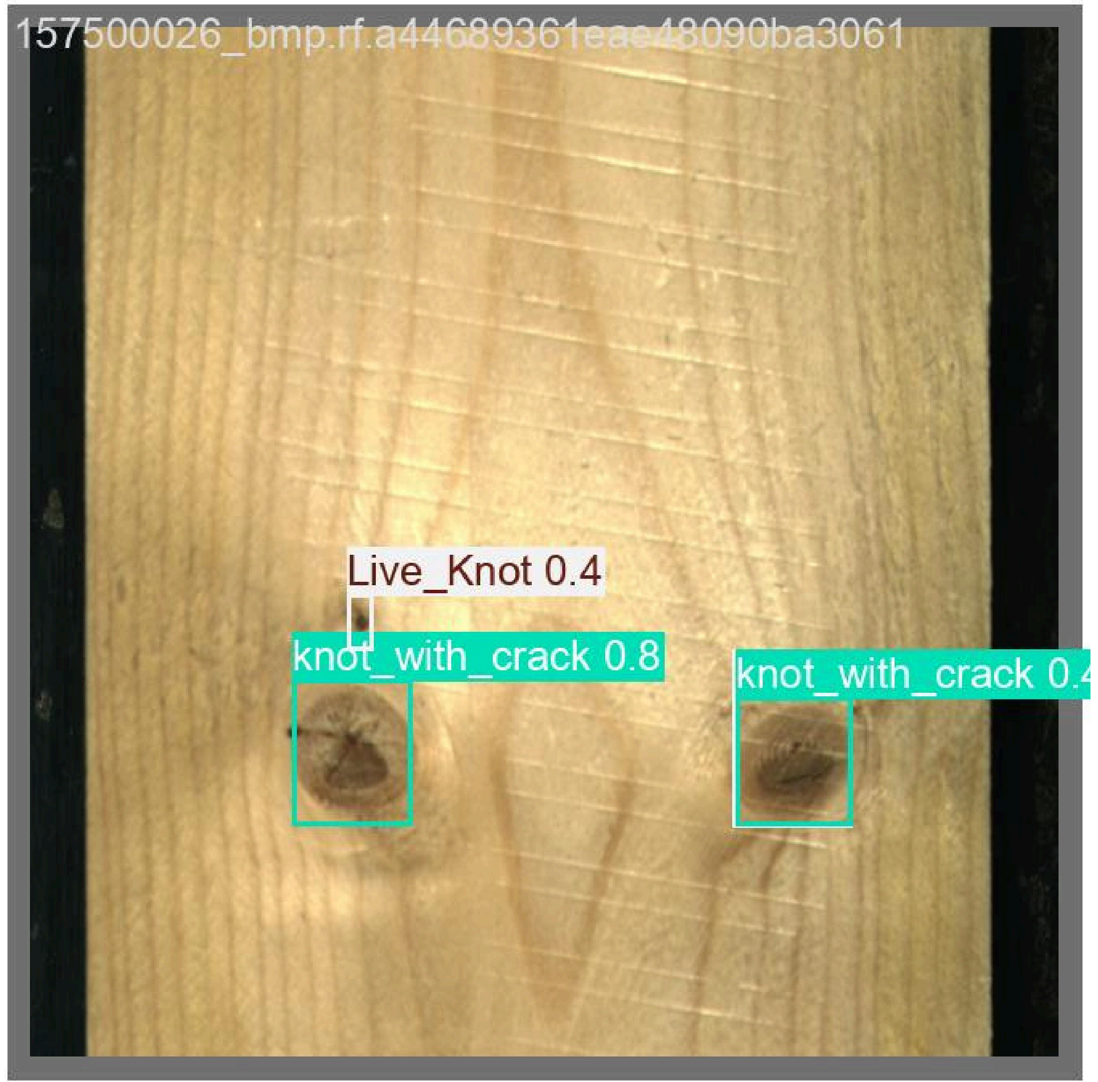


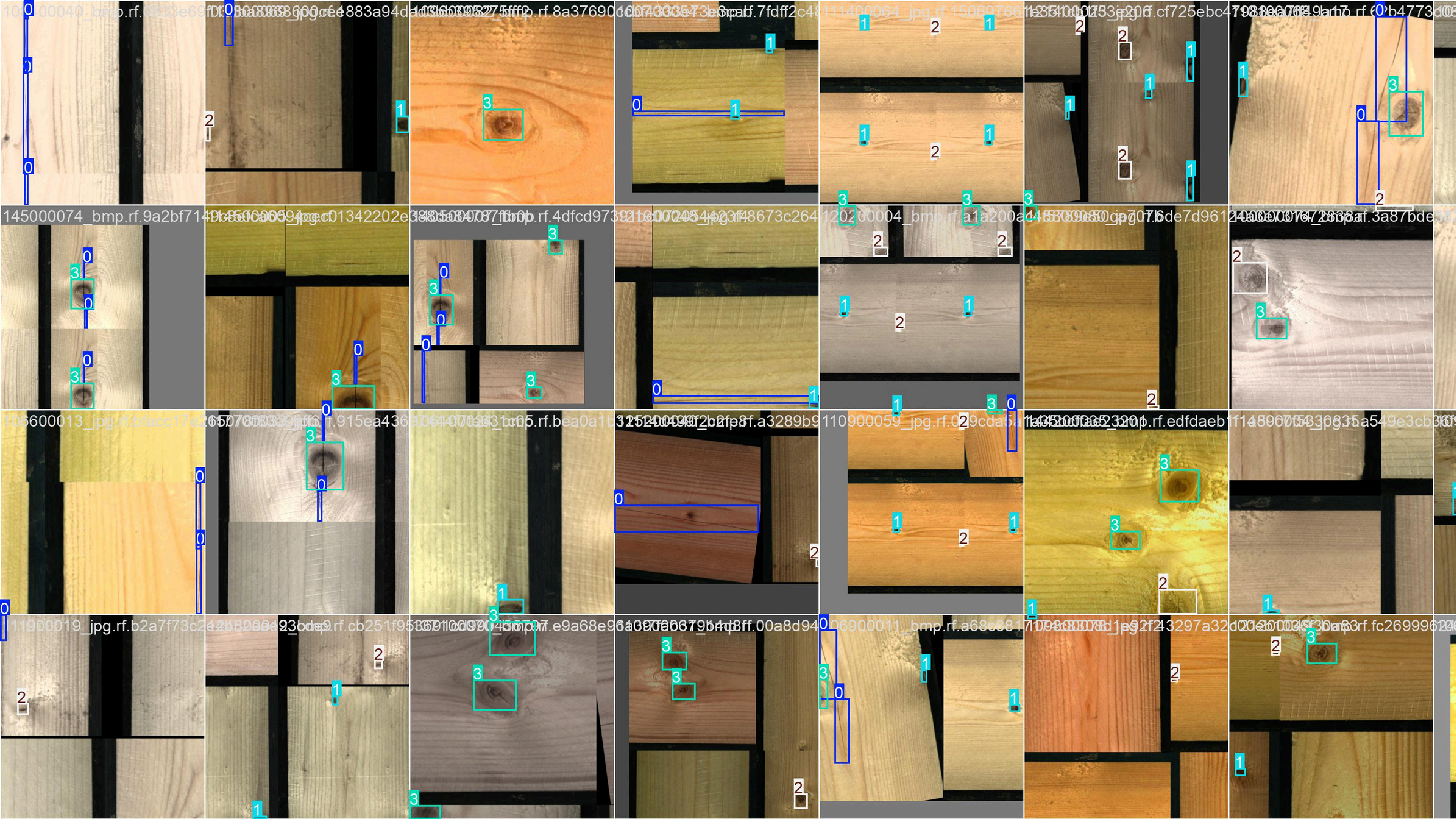


← ACTUAL

PREDICTED →







CONCLUSION

- **Contributions:** Proved that a compact, 4.46 GFLOP model can achieve high accuracy (80% mAP) for real-time timber grading.
- **Industrial Impact:** Allows low-latency quality control on edge-level hardware in timber-processing facilities.
- **Future Work:** Incorporating multi-modal inputs (depth or hyperspectral) and exploring self-supervised learning to handle unseen timber species

THANK YOU