The Nineteenth International Conference on Advances in Computer-Human Interactions

ACHI 2026

May 24, 2026 to May 28, 2026 - Venice, Italy

Semantic Segmentation of Extremely Small Defects in Sliced Apples

Yueying Shi and Oky Dicky Ardiansyah Prima

Yueying Shi (s236w002@s.iwate-pu.ac.jp)
Graduate School of Software and Information Science,
Iwate Prefectural University

SELF INTRODUCTION

- **Name:**
  - Yueying Shi
- **Institution:**
  - PhD student,
  - Graduate School of Software and Information Science,
  - Iwate Prefectural University, Japan.
- **Research field:**
  - Computer vision for food inspection,
  - Semantic segmentation and anomaly detection.

AGENDA

1. Introduction
2. Related Work
3. Materials and Method
4. Experimental Setup
5. Results and Discussion
6. Conclusion and Future Work

I. INTRODUCTION - BACKGROUND

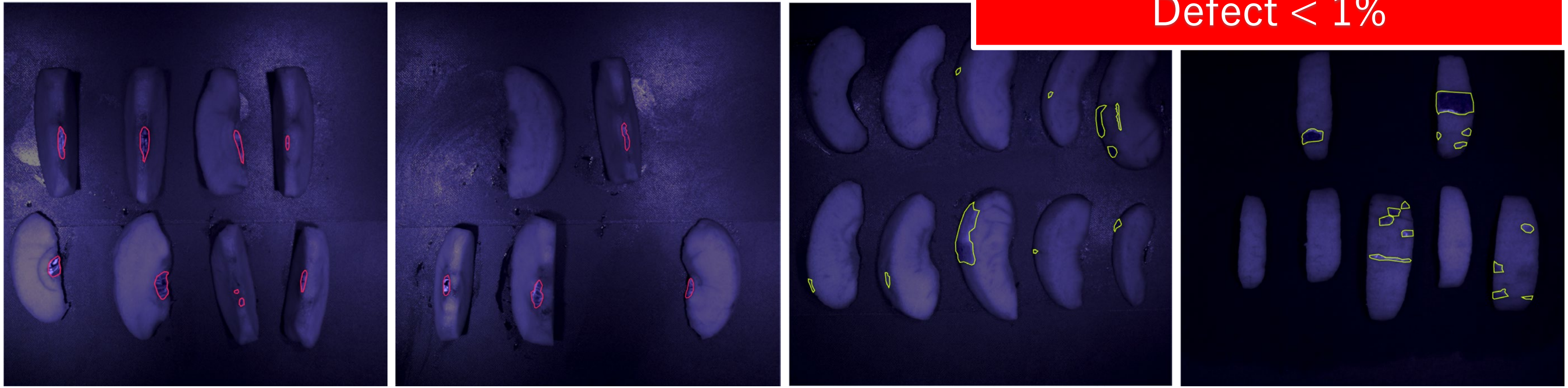
- **Automated visual inspection plays an important role in food processing.**
  - **Sliced apple** inspection still depends on human operators.
  - Residual skin and core fragments affect appearance and safety.
  - Manual inspection requires high labor and time costs.

I. INTRODUCTION – CHALLENGE

Challenge: Extremely small and sparse defects

- Defect pixels occupy less than 1% of the image area.
- High accuracy can still miss critical defects.
- Recall-oriented evaluation is required.

Background > 99%
Defect < 1%

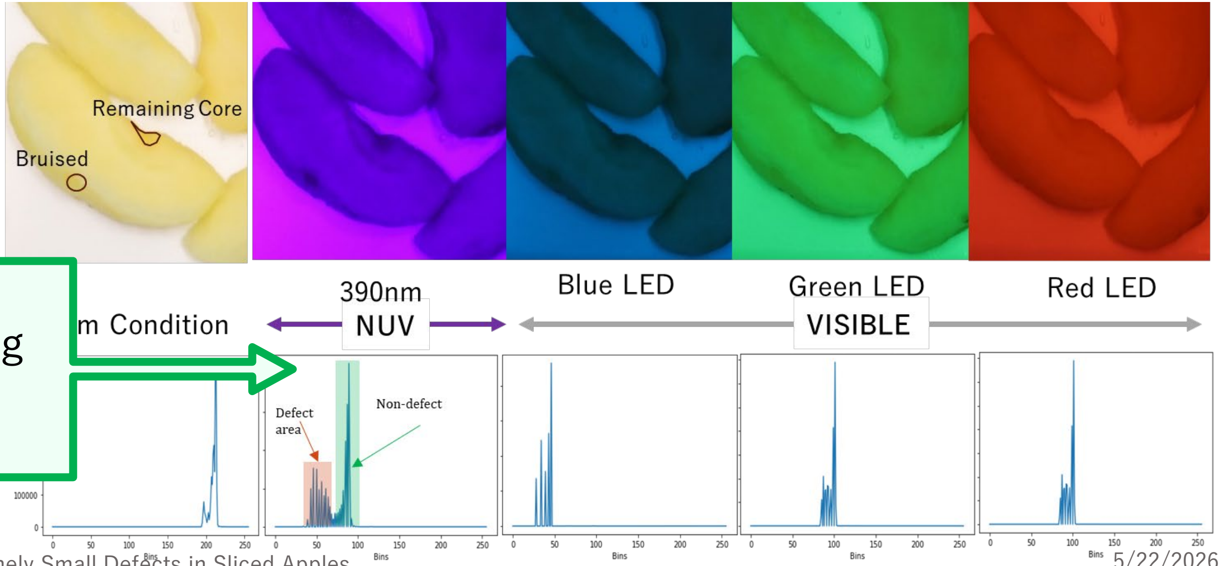


Red circles mark the remaining core, green circles mark remaining skin area.

I. INTRODUCTION – LIGHT MODALITY

■ Why UV (ultraviolet) imaging:

- ➢ UV illumination enhances defect contrast.
- ➢ Skin and core fragments become more distinguishable.
- ➢ Better visibility supports pixel-level segmentation.



Only UV imaging showed two distinct peaks, enabling separation of defect and non-defect regions.

I. INTRODUCTION – INSPECTION MODEL

- **Why semantic segmentation:**
  - Predicts pixel-level defect location and shape.
  - Suitable for small and irregular defects.
- **Aim of the semantic segmentation model:**
  - Minimize missed detections under extreme imbalance.

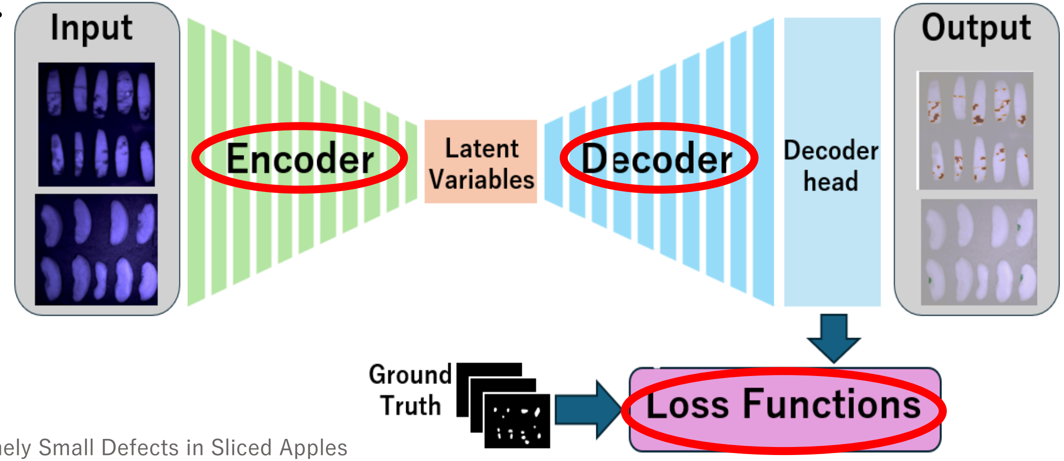
II. RELATED WORK – PRIOR STUDIES AND CHALLENGES

- **Simultaneous detection of remaining skin and core remains challenging:**
  - RGB, SWIR, and X-ray have been explored for apple inspection.
  - However, tiny residual skin and core fragments remain challenging.
  - This study focuses on UV-based pixel-level segmentation.

Imaging Type	Remaining Peel	Core Detection
RGB	✅ Good	❌ Not visible
SWIR	⚠ Possible	✅ Good
X-ray	❌ Not used	✅ Excellent
※ General characteristics from previous studies		

Overall architecture of the segmentation framework.

- **Encoder:** feature extraction
- **Decoder:** spatial reconstruction
- **Loss function:** optimization signal
- **Stepwise evaluation:** Loss, encoder, and decoder were evaluated independently.

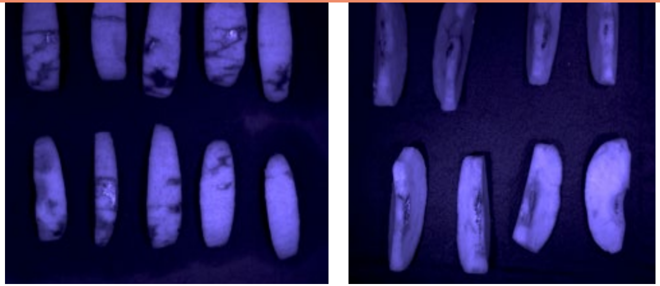


III. MATERIALS AND METHOD - PREPARATION

Image acquisition:

- 339 UV images under multi-view acquisition
- 3 classes: background, core, skin
- Annotation: pixel-level, manual
- Split: 310 training, 19 validation, 10 test images
- Augmentation: rotation and flipping

Each defect class occupies less than 1% of the image area



UV images: (L)remaining skin, (R) remaining core

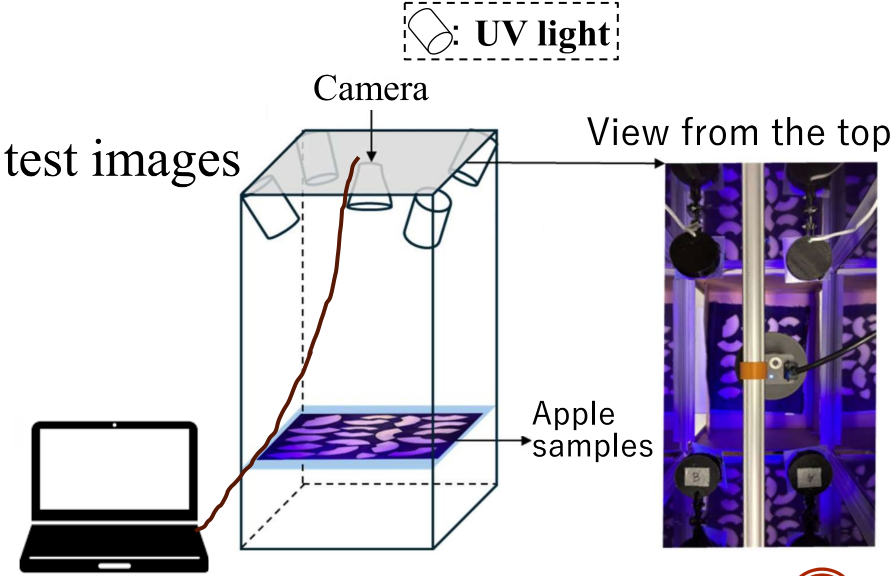
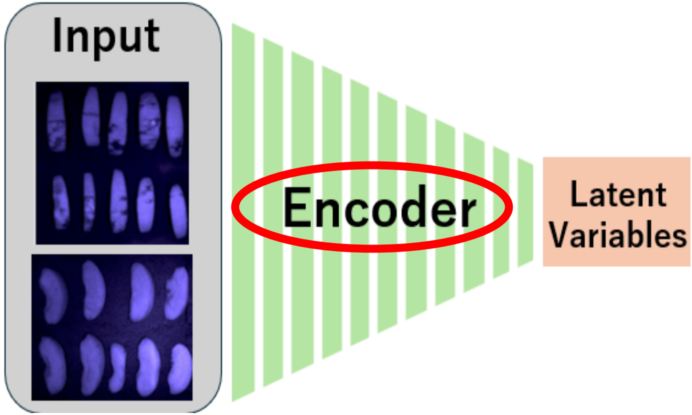


Image acquisition devices and the view from the top.

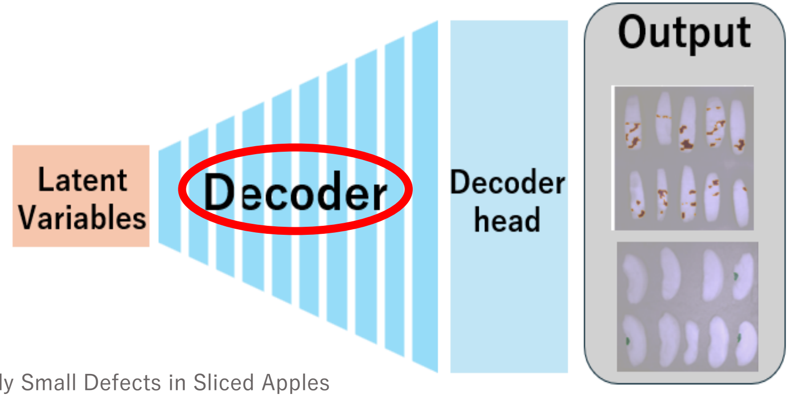
III. MATERIALS AND METHOD – ENCODER

- **Encoder candidates:** Lightweight vs deeper architectures
  - **MobileNetV2:** lightweight encoder
  - **ResNet-34 / ResNet-50:** deeper residual networks
  - **EfficientNet-B0:** scaled efficient architecture



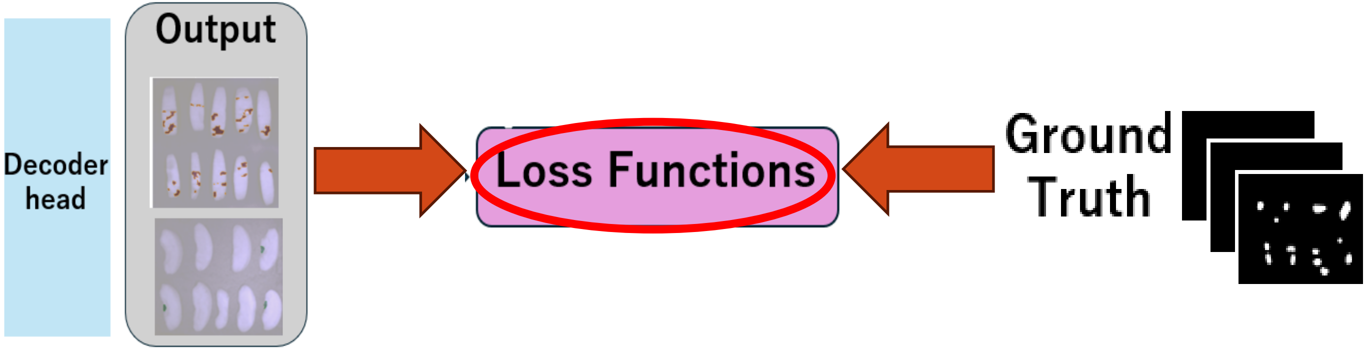
III. MATERIALS AND METHOD – DECODER

- **Decoder candidates:** Complementary designs to enhance feature fusion and localization
  - **UNet:** skip connections
  - **UNet++:** dense skip pathways
  - **DeepLabV3+:** multi-scale context
  - **Pyramid Attention Network (PAN):** attention-based refinement



III. MATERIALS AND METHOD – LOSS FUNCTION

- **Loss function:** designed for extreme class imbalance.
  - **Weighted Cross-Entropy Loss:** Pixel-wise class weighting
  - **Dice Loss:** Overlap maximization
  - **Tversky Loss:** FP–FN balance
  - **Focal Tversky Loss:** focuses on hard-to-segment regions



III. MATERIALS AND METHOD - EVALUATION

Focused metrics:

- **F2-score:** recall-oriented balanced metric
- Recall: ability to detect defect pixels
- IoU: spatial overlap quality

$$F_2 = \frac{(1+2^2)\cdot(\text{Precision}\cdot\text{Recall})}{(2^2\cdot\text{Precision})+\text{Recall}}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

IV. EXPERIMENTAL SETUP

- **Input size: 224 × 224**
- **Training setup: 200 epochs, batch size 8**
  - Same dataset split across all experiments
  - Same optimizer and learning rate schedule
- **Stepwise evaluation: one variable at a time**
  - Other variables fixed for fair comparison
- **Device: NVIDIA GeForce RTX 4080 Laptop GPU**

V. RESULTS AND DISCUSSION – LOSS FUNCTION

- **Region-based losses outperform pixel-wise weighting.**
  - **Tversky:** best for core defects
  - **Focal Tversky:** best for skin defects
  - Weighted CE: achieved high recall but lower spatial consistency.

Loss Function	Core F2	Skin F2
Tversky	★ **0.828**	0.469
Focal Tversky	0.708	★ **0.578**
Dice	0.801	0.535
Weighted CE	0.651	0.523

V. RESULTS AND DISCUSSION - ENCODER

- **MobileNetV2 achieved the strongest overall performance.**
  - Lightweight encoder preserves fine spatial details.
  - Deep encoders may suppress tiny defect cues.

UNDER FIXED **TVERSKY** LOSS AND DECODER		
Encoder	Core F2	Skin F2
MobileNetV2	★ **0.868**	★ 0.711
ResNet-34	0.828	0.469
ResNet-50	0.795	0.593
EfficientNet-B0	0.759	0.441

UNDER FIXED **FOCAL TVERSKY** AND DECODER		
Encoder	**Core F2**	**Skin F2**
MobileNetV2	0.829	★ **0.707**
ResNet-50	0.892	0.572
ResNet-34	0.708	0.578
EfficientNet-B0	0.637	0.477

V. RESULTS AND DISCUSSION - DECODER

Importance of skip connections

- UNet: stronger core segmentation.
- UNet++: stronger skin segmentation.
- Skip connections preserve high-resolution spatial features.
- This is critical for tiny and fragmented defects.

Decoder	Core F2	Skin F2
UNet	★ **0.879**	0.678
UNet++	0.868	★ **0.711**
PAN	0.862	0.677
DeepLabV3+	0.8	0.649

V. RESULTS AND DISCUSSION - ROBUSTNESS

Robust and stable performance:

- Reproducibility analysis across random seeds.
- UNet and MobileNetV2 achieved stable performance.
- UNet++ and MobileNetV2 showed higher variance.
- Lightweight encoders with skip connections improved robustness.

Decoder	Core F2 Score (mean ± std.)	Skin F2 Score (mean ± std.)
UNet	★ **0.870 ± 0.022**	★ **0.734 ± 0.035**
UNet++	0.833 ± 0.060	0.678 ± 0.102

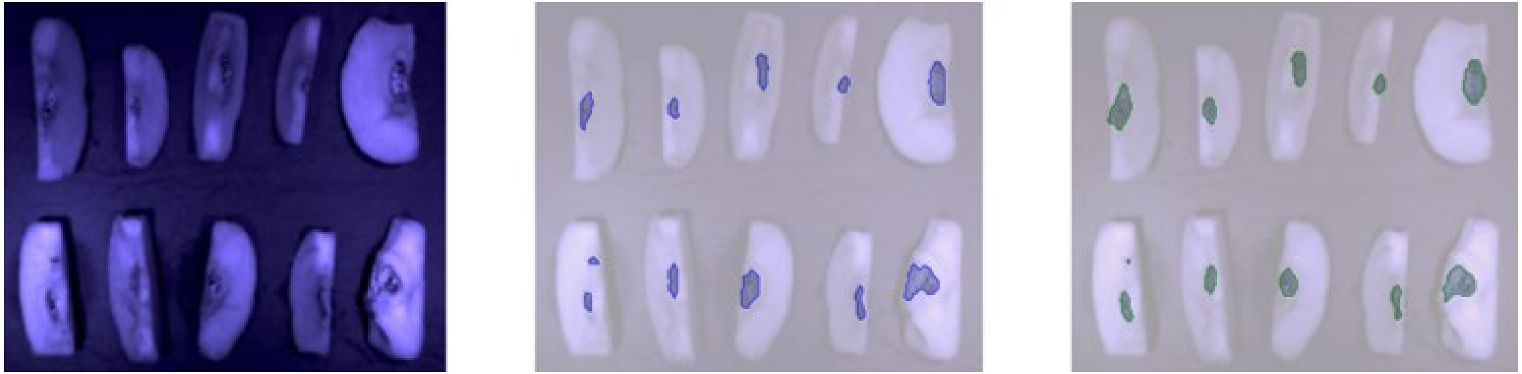
V. RESULTS AND DISCUSSION - QUALITATIVE RESULTS

Visual prediction results

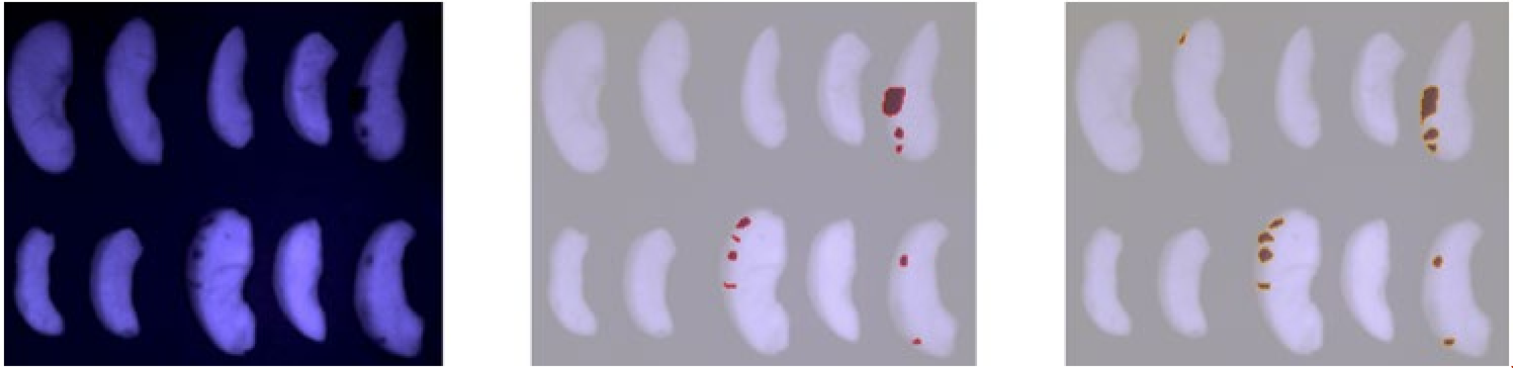
- Most defects were successfully localized.
- Minor false positives were observed.
- Minor false positives are acceptable because missed defects are more critical.

V. RESULTS AND DISCUSSION
- QUALITATIVE RESULTS (IMAGE)

Core
GT: 11
TP: 11
Missed: 0



Skin
GT: 9
TP: 9
FP: 1
Missed: 0



VI. CONCLUSION AND FUTURE WORK - FINDINGS

- **Main findings:**
  - Semantic segmentation of extremely small defects in sliced apples under severe class imbalance.
- **Loss function:** Region-based losses improve F2-score.
- **Encoder:** MobileNetV2 preserves tiny defect features.
- **Decoder:** UNet-based decoders reconstruct small fragmented regions.

VI. FUTURE WORK

Future work:

- Higher resolution training
- Boundary-aware or post-processing methods
- Real-time inspection pipeline