Why Does Aggressive Resizing Preserve Malware Image Classification Performance?

Evaluating the Impact of Interpolation and Spatial Detail on Family-Discriminative Signals

Cristina Mitu
BSc Computer Science and Engineering
Delft University of Technology
Responsible Professor: Tom Viering
Supervisor: Akash Amalan

1 INTRODUCTION

Malware binaries can be represented as grayscale **byteplot images** and classified using **convolutional neural networks (CNNs)**. Although classification accuracy can remain surprisingly high after **aggressive resizing,** it is unclear what information survives once most spatial detail is removed. This study investigates whether low-resolution byteplots retain meaningful **family-discriminative structure** or rely on coarse **dataset-specific regularities**.

2 RESEARCH QUESTIONS & OBJECTIVE

Main question:
How can malware classifiers remain accurate after aggressive byteplot resizing?

Subquestions:

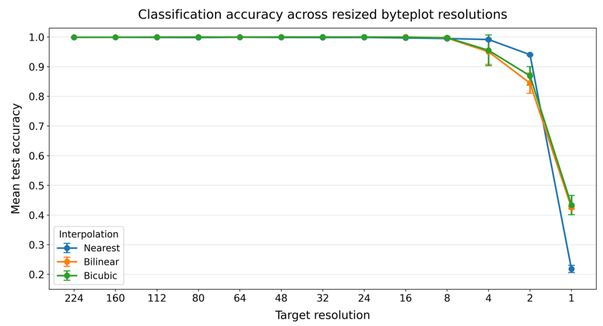
1. How do downsampling and interpolation affect accuracy?
2. Which malware families remain robust at low resolution?
3. Do high-resolution models transfer to resized inputs?
4. What information remains in the final retained pixels?

3 DATASET & METHODOLOGY

The experiments use **10,010 unpacked malware and benign byteplots** from **14 balanced families**, including PE and ELF binaries. ResNet-18 models were trained from scratch at resolutions from **224×224 to 1×1** using **nearest-neighbour, bilinear, and bicubic** interpolation, evaluated with three-fold cross-validation. Additional analyses test sensitivity to Gaussian blur, cross-resolution transfer, family-level robustness, and the predictive value and binary origins of pixels retained at ultra-low resolutions.

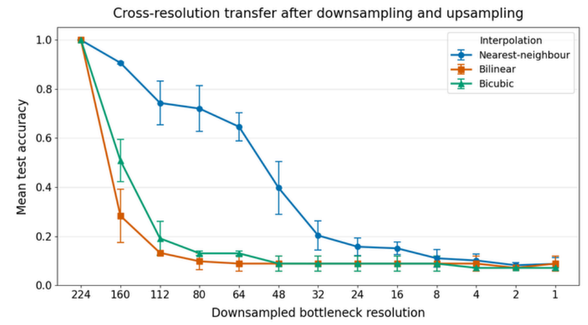
4 RESULTS

A Same-resolution performance



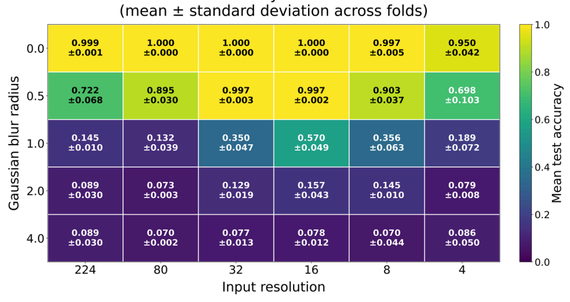
All interpolation methods remain above 0.99 mean accuracy through 8x8. At 4x4 and 2x2, nearest-neighbour performs best; at 1x1 bilinear and bicubic perform better.

B Cross-resolution transfer fails



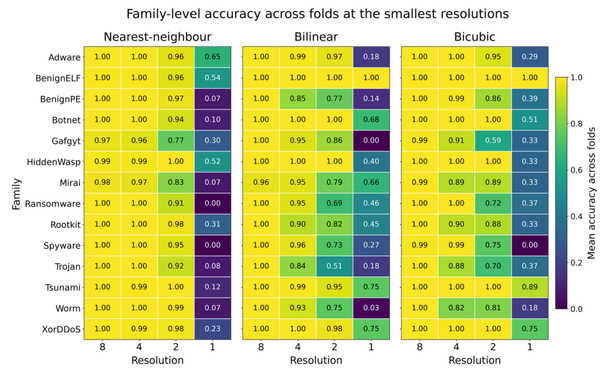
A model trained at 224x224 does not reliably classify aggressively downsampled and re-upsmapled inputs. This shows that low-resolution models learn resolution-specific cues.

C Blur reduces performance



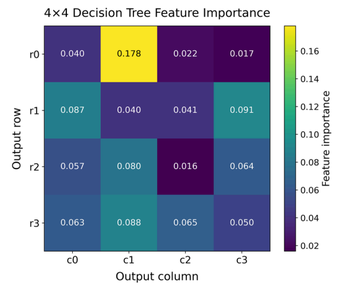
Gaussian blur sharply reduces accuracy, showing that broad texture alone is not sufficient for classification; preserved local contrast matters.

D Family robustness differs

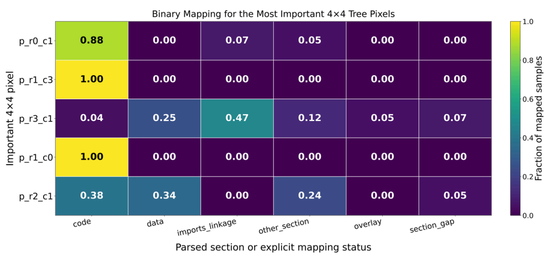


Some malware families remain robust at the smallest resolutions, while others degrade rapidly. Low-resolution signal is not equally strong across families.

E Interpretability of Retained Pixels



A decision tree trained at 4x4 resolution reaches 0.832 accuracy. However, only a subset of retained pixels remain representative for the family-signal. Backmapping of these pixels shows that these often overlap with executable code, but also data and imports/linkage. This suggests coarse family-level structure and sampled byte values remain informative, but does not prove semantice code understanding



5 CONCLUSION & NEXT STEPS

Aggressive resizing preserves useful malware-family information through coarse layout, sampled byte values, and local contrast. Nearest-neighbour performs best when a few spatial positions remain, while smoothing reduces accuracy and becomes more useful only at 1×1. Robustness differs across families, and high-resolution models cannot directly reuse their learned representation after aggressive compression. Retained pixels often overlap meaningful binary regions, especially executable code, but this does not demonstrate semantic code understanding.
Future work should validate these findings on other datasets and compare them with simple baselines such as file size, byte histograms, and entropy.