Approximating Classical Feature Detectors using CNN Trained on a Single Image

1 Problem Definition

This project explores the feasibility of training Convolutional Neural Networks (CNNS) to replicate the functionality of classical computer vision algorithms, specifically Harris Corner Detection and Canny Edge Detection, under the constraint of using only a single image for training. Given patches extracted from one image of the John Curtin School of Medical Research, the goal was to train a CNN to predict corresponding Harris corner likelihood heat-maps and Canny edge probability maps. This investigation tests the limits of deep learning with minimal data, examining the model's ability to approximate complex functions versus achieving generalizable feature understanding.

2 Method

- **Ground Truth Generation**

The Canny edge map was generated using OpenCV's 'cv2.Canny' function with thresholds selected to capture prominent edges. For Harris corners the code from lab 3 was used which consisted a manual implementation of Sobel filters which computed image gradients (Ix, Iy), the structure tensor components (Ix^2, Iy^2, Ixy) were calculated and smoothed using Gaussian filtering (sigma=1.0), and the Harris response (R) was computed using $R = \mathrm{det(M)} - \mathrm{k} \times (\mathrm{trace(M)})^2$ with k$= 0.04$. Non-maximum suppression (NMS) was applied to the response map to identify keypoint locations (shown in Figure 1). For CNN training, the raw Harris response map was clipped and normalized to [0,1] to serve as the target heat-map, while the Canny map was converted to a binary float map 0.0, 1.0. Figure 2 shows the input image and generated ground truth maps.

- **Data Preparation**

The single grayscale input image and its corresponding Harris and Canny ground truth maps were divided into overlapping patches. To artificially increase the dataset size, geometric augmentations (random rotations and flips) were applied, ensuring the same transformation was applied to an input patch and its corresponding ground truth patches (see Figure 3). The resulting pool of patches was randomly split into training (80%) and validation (20%) sets. Custom PyTorch 'Dataset' and 'DataLoader' classes were implemented to serve batches of (input, Harris ground truth, Canny ground truth) tensors during training.

- **Model Architecture and Training**

Two separate U-net models, a standard encoder-decoder architecture with skip connections suitable for image-to-image tasks, were employed - one for Harris prediction and one for Canny prediction. Both models took single channel grayscale patches as input and outputted single channel maps via a Sigmoid activation. The models were trained using the Adam optimizer for 100 epochs. Mean Squared Error loss was used for the Harris heat-map regression task, and Binary Cross-Entropy loss was used for the Canny edge probability prediction. Model weights yielding the lowest validation loss for each task were saved to mitigate overfitting.

3 Results

The training and validation loss curves for both model are shown in Figure 4. Both models exhibited rapid initial learning, with losses decreasing significantly in early epochs. However, clear signs of overfitting started to emerge. The Harris model's validation loss plateaued after around 40 epochs, while the Canny model's validation loss began to consistently increase after approximately 30 epochs, indicating that continued training was degrading performance on unseen patches from the source image. The best recorded validation loss for Harris was approximately 0.0001 and for Canny was approximately 0.015.

Evaluation on the original ANU building image (Figure 5) revealed mixed results. The Canny CNN prediction (thresholded at 0.5) provided a remarkably clean edge map, closely matching the classical Canny output visually and achieving high quantitative scores (Table 1), suggesting successful function approximation for this specific image. However, the CNN's predicted Harris heatmap (Figure 5, Figure 6) differed noticeably from the classical ground truth, activating broadly along edges rather than producing sharp corner peaks. Consequently, applying Non-Maximum Suppression (NMS) to extract discrete keypoints from the CNN Harris map was ineffective.

To explicitly test generalization despite the single image training, the models were applied to three entirely different test images of buildings. Counter-intuitively, both CNNs demonstrated surprising generalization. For each test image, the CNN Harris heatmap closely resembled the classical Harris output, and the CNN Canny prediction was nearly identical to the classical Canny edges for that respective image. This indicates that the models learned more than just the specific structure of the training building; they successfully approximated the fundamental, local operations inherent in the classical algorithms.

4 Reflection

This project yielded intriguing insights into learning with extreme data limitations. While the validation loss curves indicated overfitting to the specific features of the single training image, the subsequent tests on unseen images revealed unexpected generalization capabilities for both Harris and Canny approximations. This suggests the CNNs, particularly the U-Net architecture, were able to learn the underlying local computations of these classical detectors from the varied patch examples. The Harris CNN however, learned a representation emphasizing edges more broadly than the classical corner response, making direct keypoint extraction via NMS problematic.

The initial expectation of generalization failure was hence challenged. The ethical consideration remains to accurately represent these findings, i.e., the models are approximations that generalized surprisingly well for these specific tasks, but are not guaranteed replacements for classical methods across all domains and were still trained on biased, minimal data. The key takeaway is nuanced, in that while overfitting is a risk with limited data, CNNs can sometimes learn fundamental, low-level image operations robustly if those operations are sufficiently represented locally within the training examples (edges / corners). This highlights the difference between overfitting on specific high-level features versus learning generalizable low-level functions.

5 Conclusion

Convolutional Neural Networks trained on patches from a single image successfully approximated the Canny edge detection algorithm, demonstrating unexpected generalization to novel test images. The attempt to replicate Harris corners resulted in a CNN that learned relevant structural features, also generalizing visually to test images, but produced an edge-like heatmap unsuitable for standard keypoint extraction via NMS. While validation loss curves indicated overfitting relative to the source image, the practical tests revealed the models learned the fundamental local operations of the classical algorithms surprisingly well.

To improve corner localization from the Harris CNN, post-processing techniques could be applied to its output heatmap before NMS, such as applying different sharpening filters or morphological operations to enhance peaks relative to edges. Experimenting with alternative loss functions designed to encourage sparsity or penalize non-peak activations might also yield heatmaps more suitable for keypoint extraction. Furthermore, investigating different network architectures or analyzing the learned filters within the current Harris CNN could provide insights into why it favoured edge representation.

6 Bibliography

[1] Philipp Fischer Olaf Ronneberger and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany.*

[2] Jiao Ding, Jie Chang, Renrui Han, Li Yang. CDSE-UNet: Enhancing COVID-19 CT Image Segmentation with Canny Edge Detection and Dual-Path SENet Feature Fusion. In *School of Electrical and Electronic Engineering, Anhui Institute of Information Technology, Wuhu 241199, China*

[3] Kanija Muntarina, Rafid Mostafiz, Fahmida Khanom, Sumaita Binte Shorif, Mohammad Shorif Uddin. MultiResEdge: A deep learning-based edge detection approach. In *https://doi.org/10.1016/j.iswa.2023.200274*

[4] Javier Sánchez, Nelson Monzón, Agustín Salgado. An Analysis and Implementation of the Harris Corner Detector. In *10.5201/ipol.2018.229*

[5] Erik Cuevas, Margarita-Arimatea Díaz-Cortes, Efrén Mezura-Montes. Corner detection of intensity images with cellular neural networks (CNN) and evolutionary techniques. In *https://doi.org/10.1016/j.neucom.2019.03.014*

[6] Mastering U-Net: A Step-by-Step Guide to Segmentation from Scratch with PyTorch. In *medium.com*

[7] Bharath K. U-Net Architecture For Image Segmentation. In *digitalocean.com*

Table 1: Canny quantitative metrics.

Method	Pixel Acc	Edge IoU	Precision	Recall	F1-score
Canny	**0.9982**	**0.9577**	**0.9795**	**0.9772**	**0.9784**

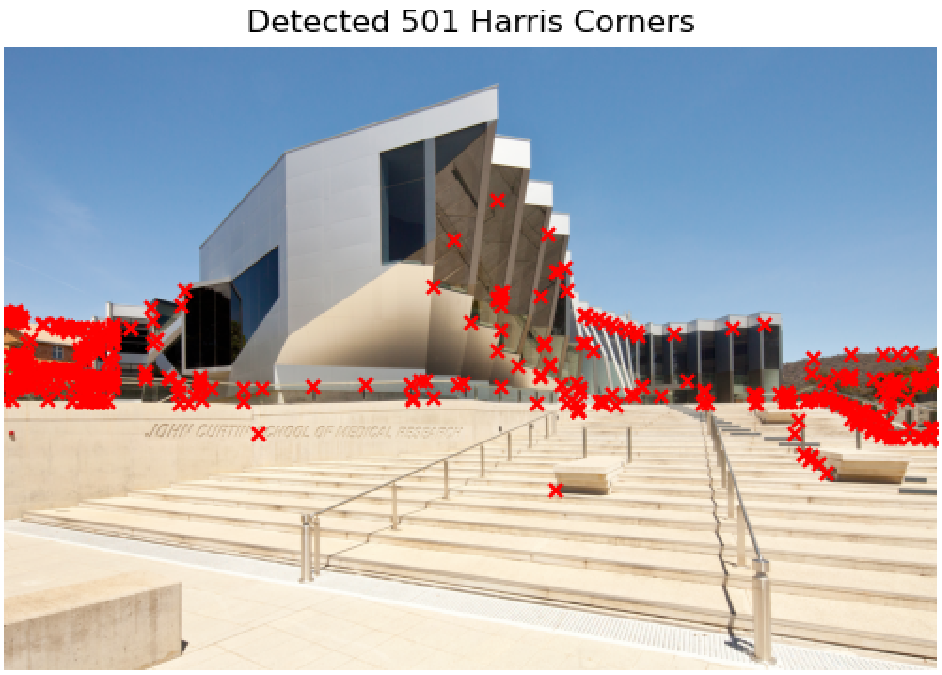

Figure 1: Harris corners detected.

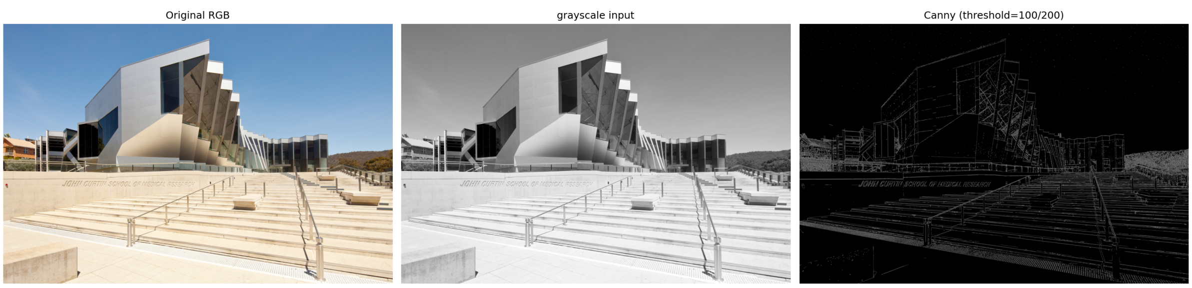

Figure 2: Canny edge ground truth.

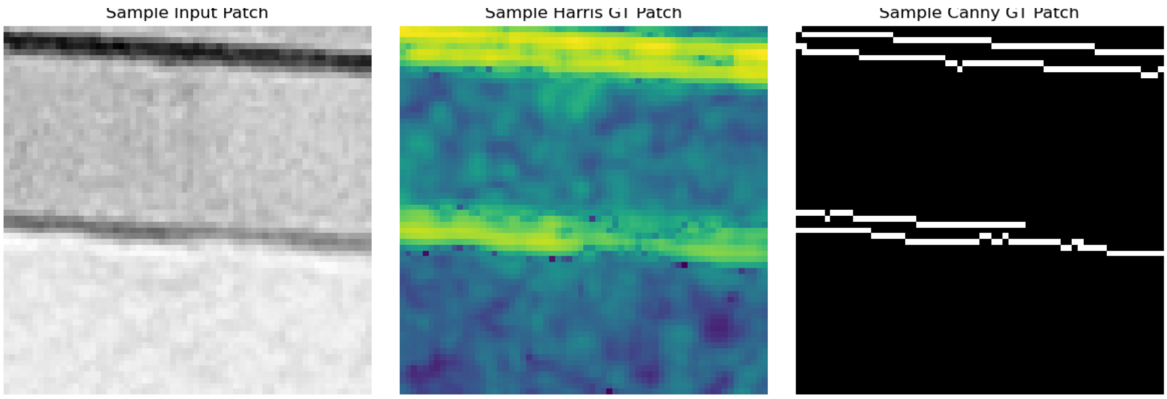

Figure 3: Sample batches.

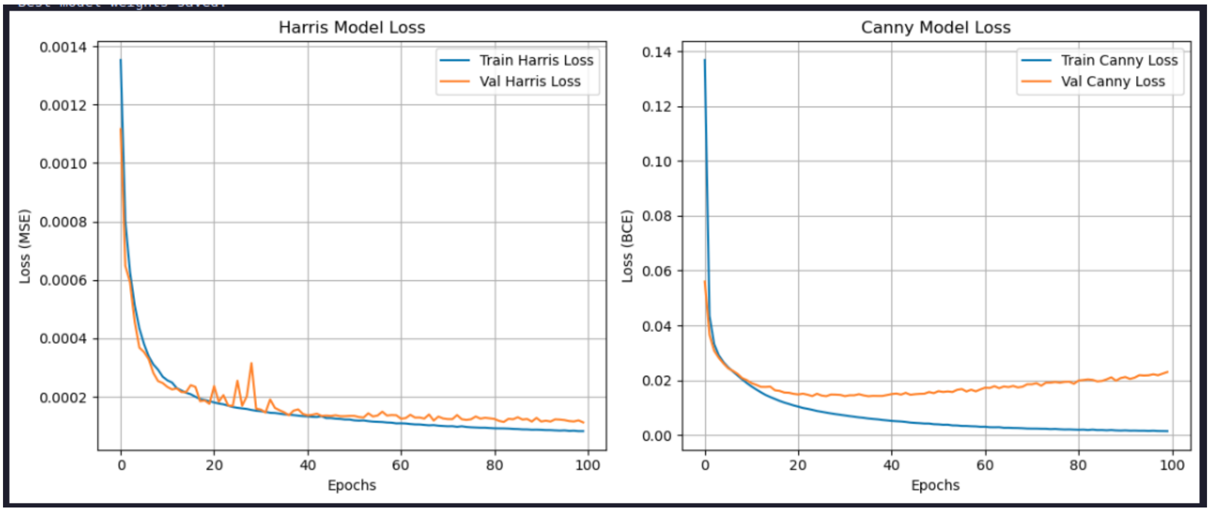

Figure 4: training and validation loss curves.

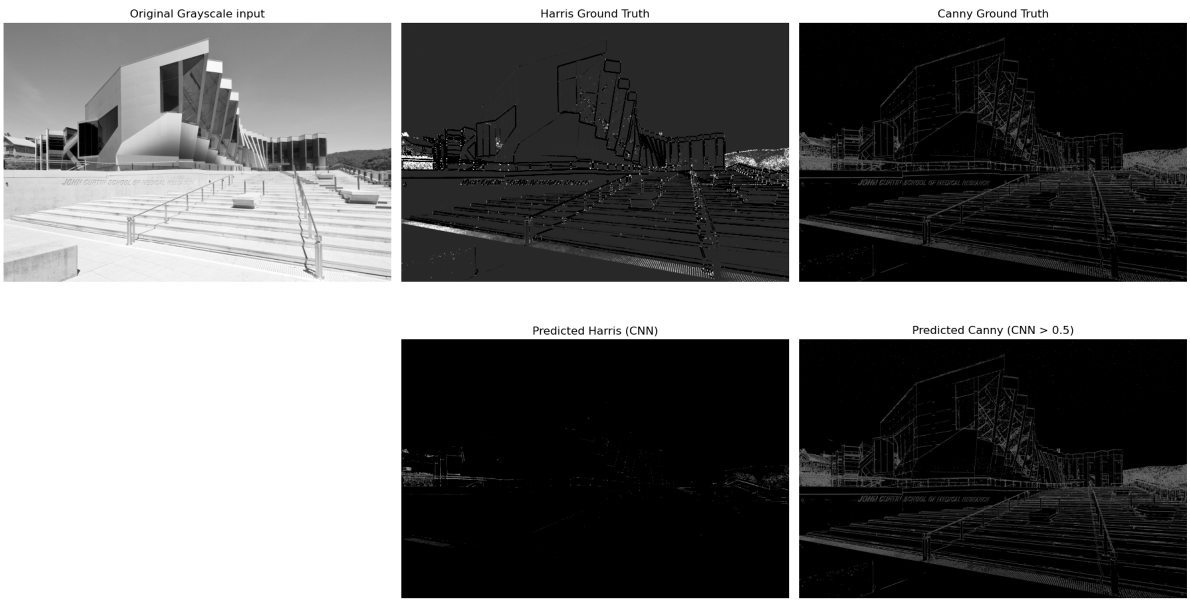

Figure 5: Final predictions

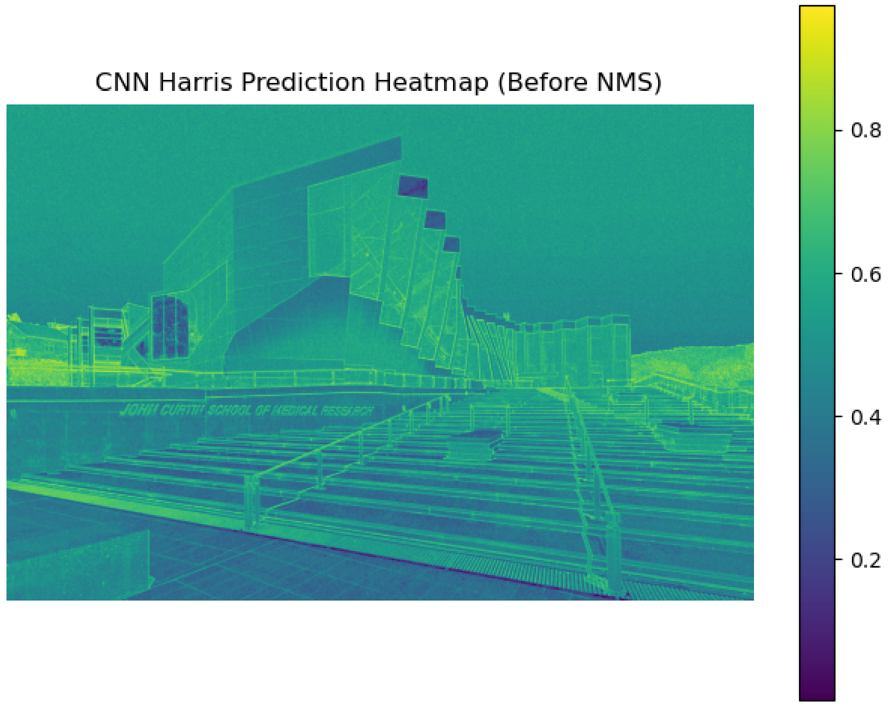

Figure 6: CNN Harris Prediction Heatmap (before nms)

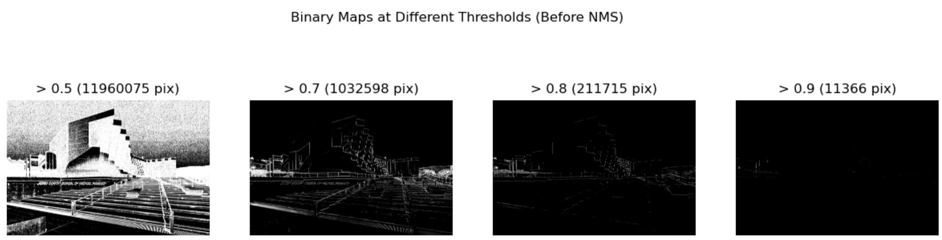

Figure 7: Binary maps at different thresholds (before nms)

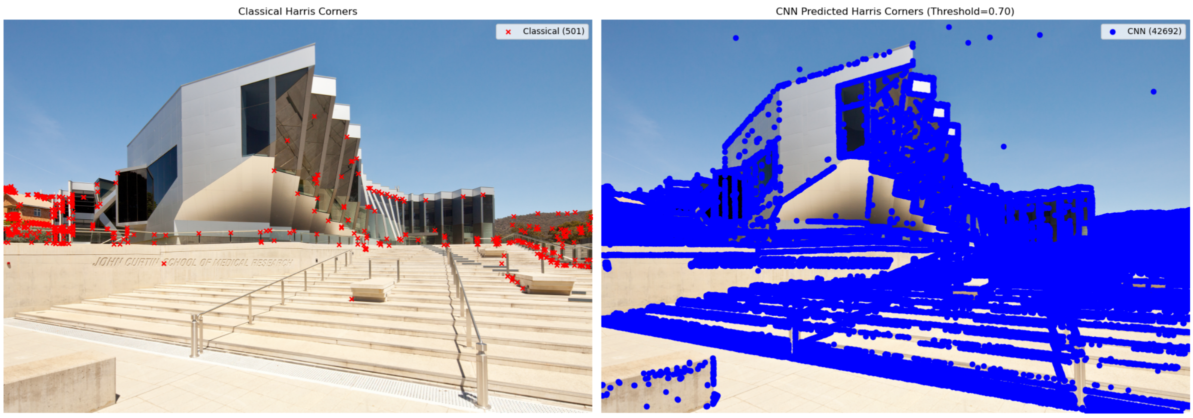

Figure 8: Harris corner comparison