UMD Building Classification

UNIVERSITY OF MARYLAND

Jerry Li, Jeffery Tian, Can Li, Christopher Lim, Justin Zhao

Code available at:
github.com/BlueTurtle123/CMSC472_Final_Project

Motivation and Data Collection

We compared various CNN classification techniques

- **Collected**: 4000+ images of 20 different UMD buildings
- **Augmented**: using SyRA synthesized rain

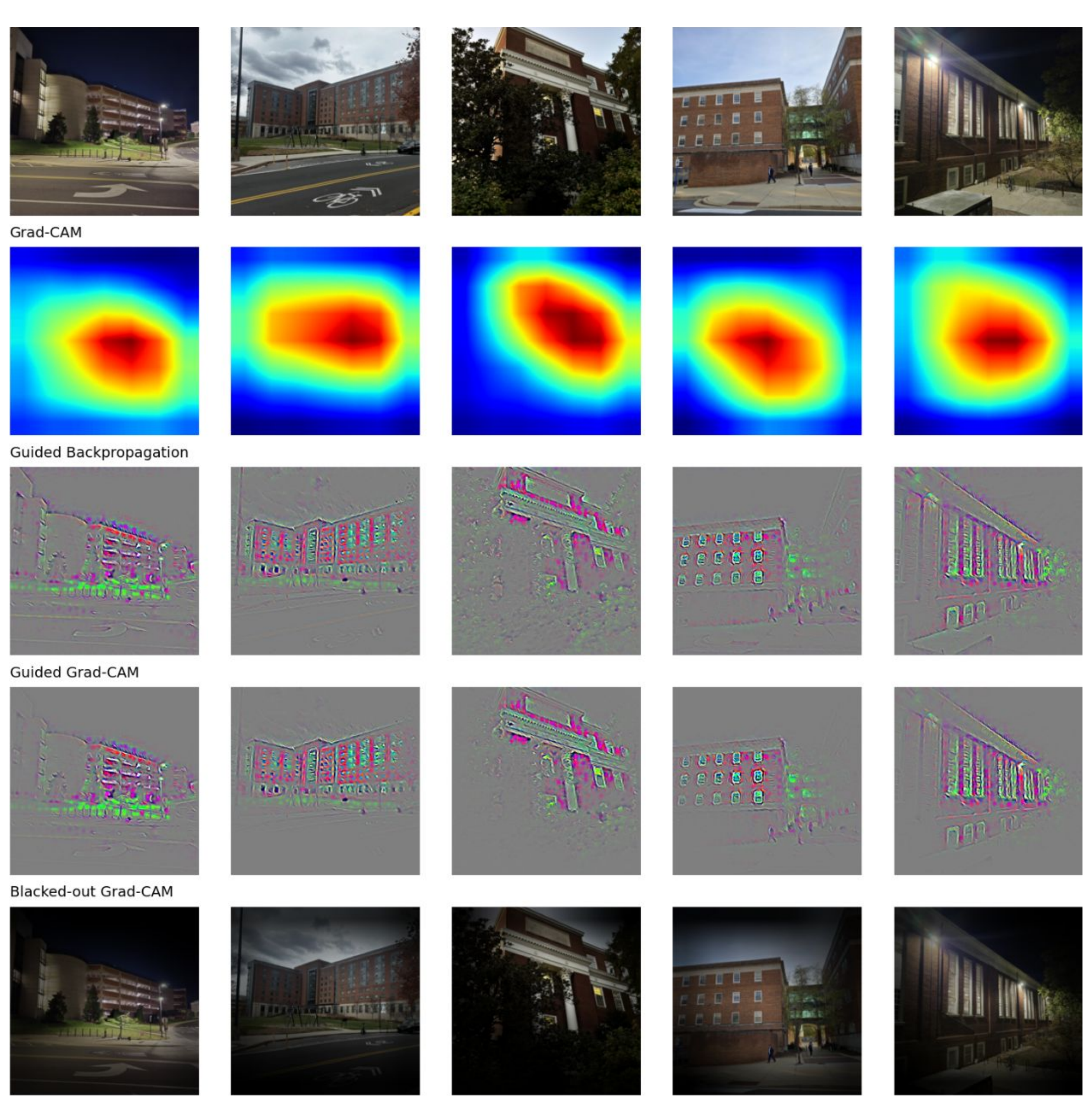
Fully Supervised Baseline

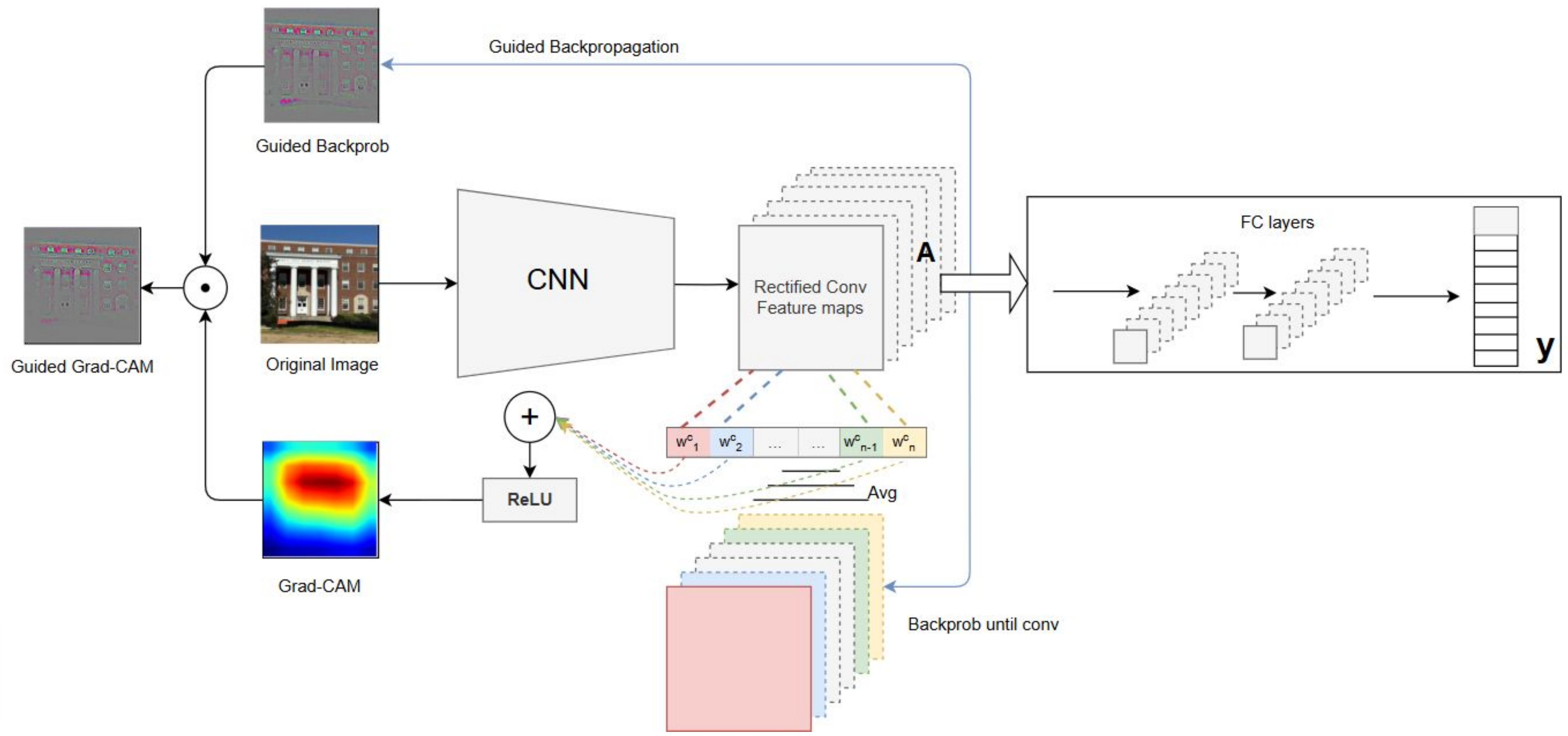
Architecture	Val Accuracy	Test Accuracy
DenseNet121	97.47%	87.77%
DenseNet201	95.26%	87.93%
ResNet18	97.63%	88.54%
ResNet34	98.89%	86.38%
ResNet50	96.05%	80.13%

A CNN is a **Black Box**. We need a way to see what is going on in there. The answer is to use Grad-CAM.

Grad-CAM generates visual explanations for CNN decisions by highlighting important regions in an image.

Where is ResNet-18 looking at?





Grad-CAM computes class-specific visualizations by forwarding an image through a CNN, followed by task-specific calculations to obtain a raw class score.

The gradients for all classes except the target class are set to zero, and the signal is back propagated to the convolutional feature maps to generate a coarse heatmap.

This heatmap is then multiplied with guided backpropagation to produce high-resolution, concept-specific visualizations, known as **Guided Grad-CAM**.

Semi-supervised Training Methods

MeanTeacher with ResNet-18

Method: Leverages both labeled and unlabeled data to improve model performance.
Key Feature: Uses consistency regularization through teacher-student networks.
Results: Student Model Test Acc: 75%, Teacher Model Test Acc: 73%

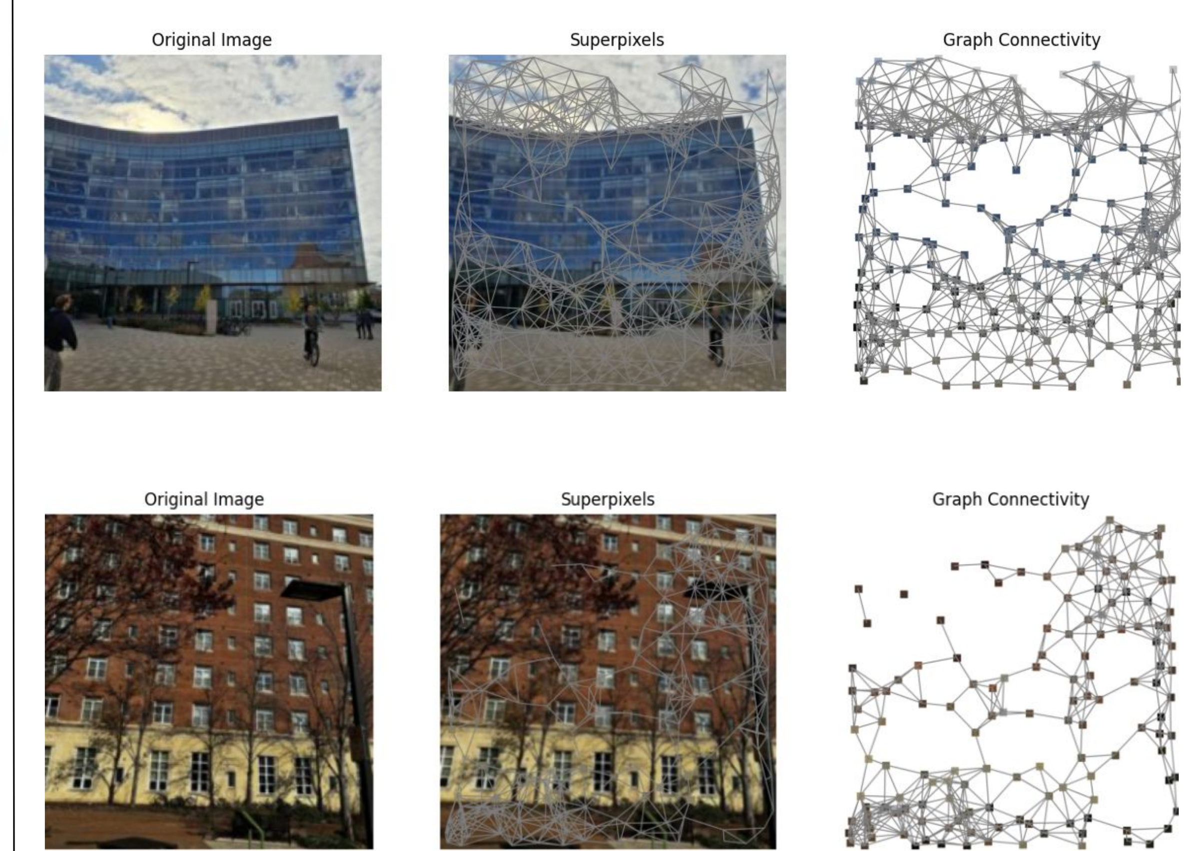
FixMatch with WideResNet

FixMatch is a semi-supervised learning algorithm that improves model performance by combining pseudo-labeling with consistency regularization, using both labeled and unlabeled data.

- 20 labeled images per class
- Top1 accuracy: 40.82%
- Top5 accuracy: 76.63% (within the top 5 predictions that the model makes)

Graph Neural Network (GNN)

The model did not converge in the time constraints, but we completed the SLIC transform for certain examples.



Human-in-the-loop (HITL)

Definition: Integrates human expertise to guide model learning and improve decision-making.

Benefit: Enhances model accuracy by refining predictions with minimal labeled data.

Application: Perfect for tasks where expert input is valuable but labeled data is limited.

Result: achieved 67% test accuracy training ResNet-18 from scratch, compared to 31% accuracy without doing HITL.