Deep Learning-based Scene Recognition — Report

Model: EfficientNet-B0 (timm) | Checkpoint: best_model.pth

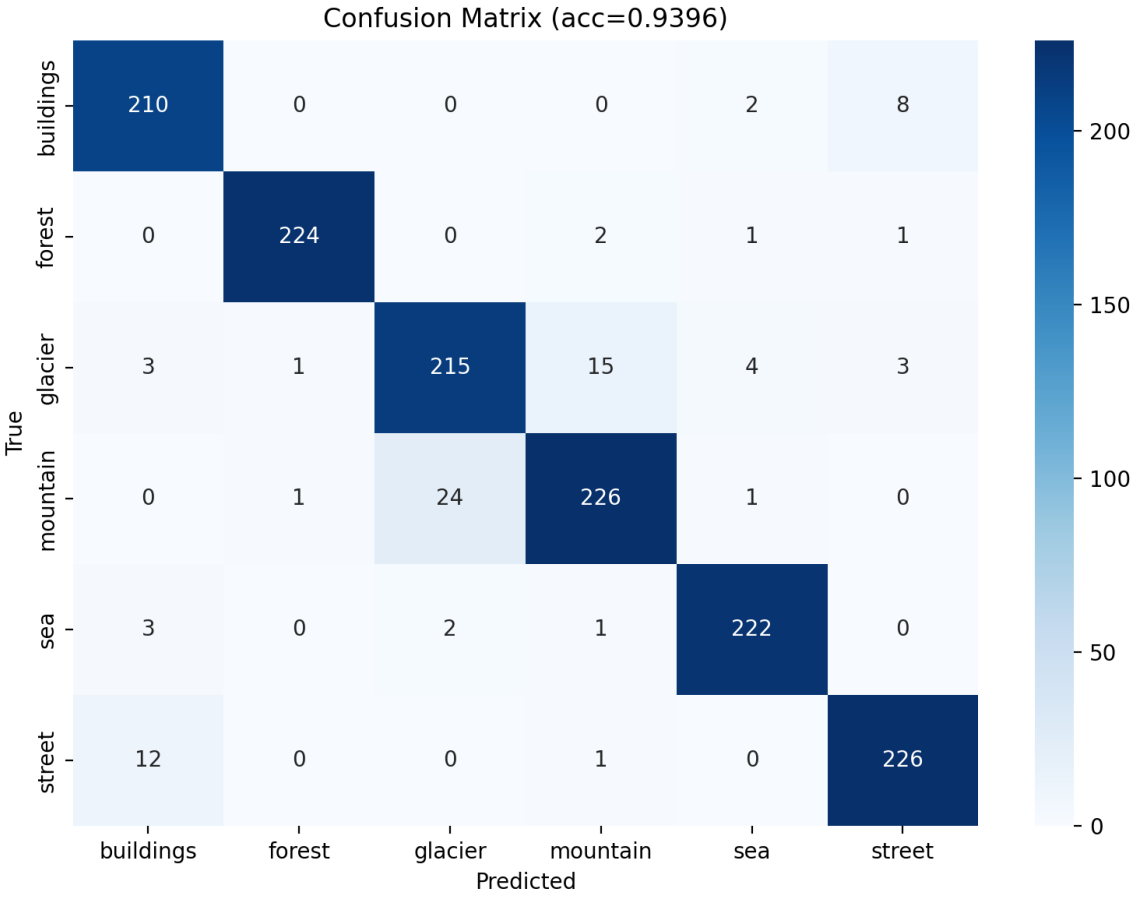
Dataset split: /content/dataset_split | Test size: 1408 images

Overall test accuracy: 0.9396

Classification Report (precision / recall / f1):

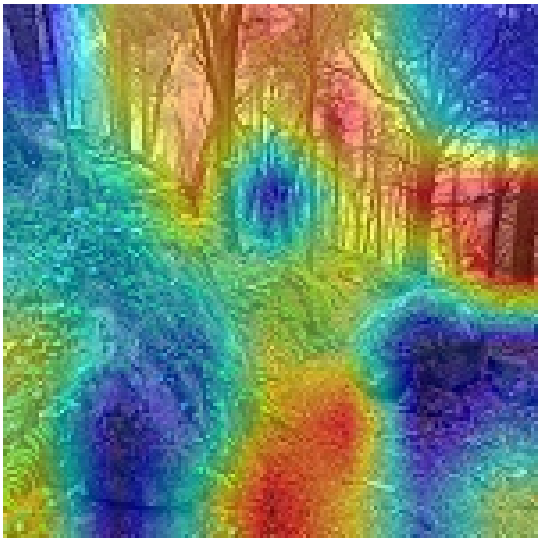
	precision	recall	f1-score	support
buildings	0.9211	0.9545	0.9375	220
forest	0.9912	0.9825	0.9868	228
glacier	0.8921	0.8921	0.8921	241
mountain	0.9224	0.8968	0.9095	252
sea	0.9652	0.9737	0.9694	228
street	0.9496	0.9456	0.9476	239
accuracy			0.9396	1408
macro avg	0.9403	0.9409	0.9405	1408
weighted avg	0.9397	0.9396	0.9396	1408

Confusion Matrix

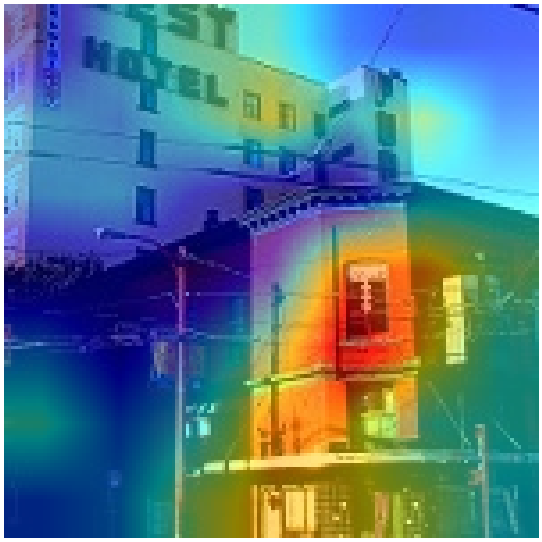


Grad-CAM Visualizations (sample images)

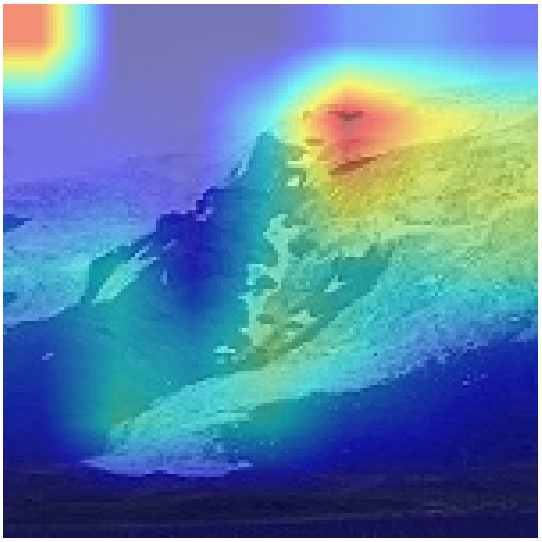
gradcam_0_street.png



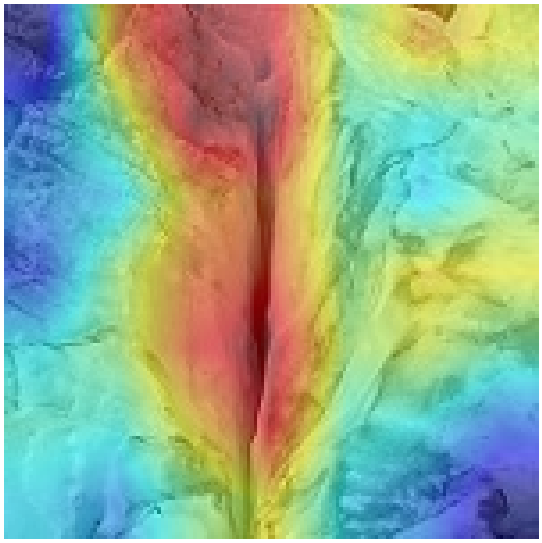
gradcam_1_forest.png



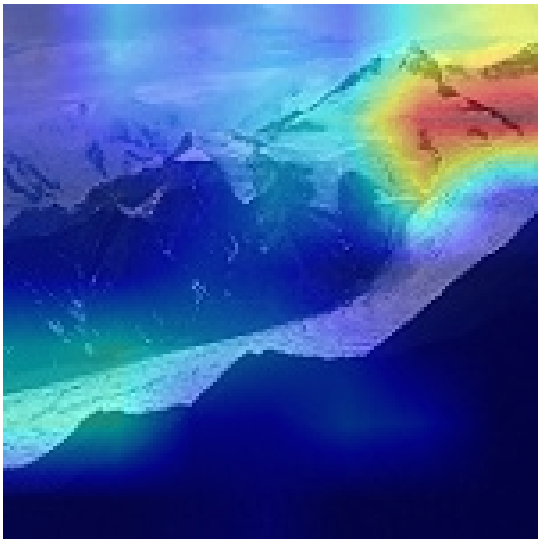
gradcam_2_buildings.png



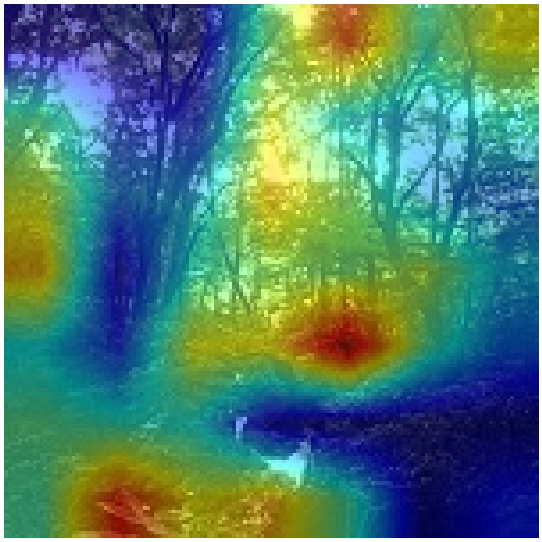
gradcam_3_glacier.png



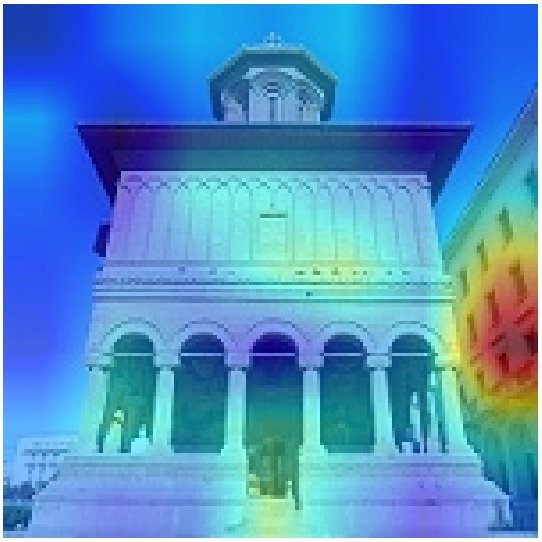
gradcam_4_glacier.png



gradcam_5_glacier.png



gradcam_6_forest.png



gradcam_7_buildings.png

Notes & Next Steps

- Consider TTA and small ensemble for +1~2% boost if needed.
- For deployment: quantize with torch.quantization or export to ONNX/TorchScript.
- Create Streamlit demo to show predictions + Grad-CAM (app provided).