When Auxiliary Cues Fail: A Data-Centric Study of Clean-to-Stressed Generalization in Battlefield Infrared Imagery

Noyon Dey, Nasid Habib Barna, Suchendra M. Bhandarkar
School of Computing, University of Georgia, Athens, GA 30602, USA.

Visual and Parallel Computing Laboratory

School of Computing
UNIVERSITY OF GEORGIA

Motivation

- IR images contain limited visual details, making classification difficult, especially under stressed battlefield conditions where data are scarce.
- Imaging stress can alter IR-specific thermal characteristics, highlighting the need to evaluate IR recognition under stressed conditions.
- Auxiliary information improves RGB recognition, but its effectiveness for IR, particularly stressed IR remains largely unexplored.

Fig.1: Thermal characteristics changes due to applied stress.

Contributions

- Study of battlefield vehicle recognition in data-scarce, synthetically stressed IR imagery.
- Evaluation of semantic attributes and shape-context (SC) fusion across multiple state-of-the-art backbones.
- Demonstration of the fact that auxiliary fusion provides only marginal improvement (+0.2%) in stressed IR image classification.
- Identification of key failure modes: low visibility, poor thermal separability, incomplete structural cues, and modality bias.
- Suggestion of future directions: thermal-consistent modeling, improved shape representations, and Mixture-of-Experts (MoE) architectures.

Auxiliary Data: Attributes and Shape

- **Attributes:** We use discriminative vehicle cues, including tracks, wheel count, turret, barrel, viewpoint, and turret-to-body ratio.
- **Shape Context (SC)** and **Generalized Shape Context (GSC)**: SC and GSC encode global shape and boundary structure using relative orientation histograms of sampled contour points, with GSC additionally incorporating local edge-orientation information.

Methodology

- **Auxiliary Information Generation:** Fine-tuned YOLOv8 extracts vehicle attributes, viewpoints, and segmentation masks; turret-to-body ratio is computed from the masks.
- **Visual Feature Extraction:** *ConvNeXt V2* processes replicated 3-channel IR images and produces a 768-D visual feature vector.
- **Auxiliary Feature Encoding:** The 6-D attribute vector is projected to 128-D, while the GSC shape descriptor is represented as a 512-D vector.
- **Late Feature Fusion:** Visual (768-D), attribute (128-D), and GSC (512-D) features are concatenated into a unified representation for robust vehicle classification in the presence of human-induced and environmental stressors.

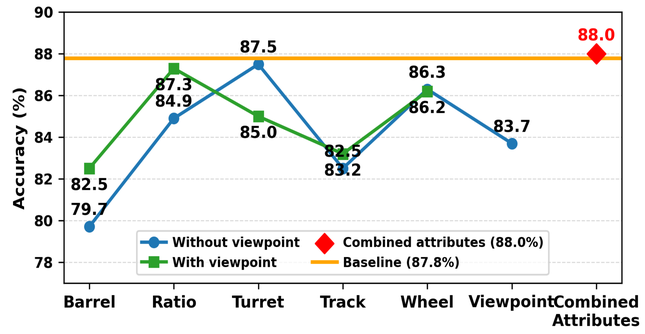
Results and Discussion

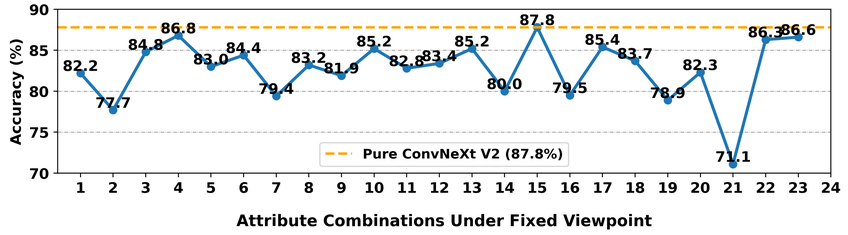
Dataset:

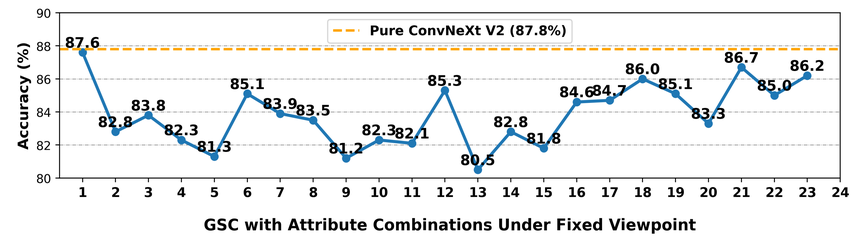
- **Army:** Train (12.5K), Test Stressed (1.5K). [Classification]
- **FLIR:** Train (8.9K), Test 1.4K. [Synthetic data generation]

Employed stressors:

- Commonly used battlefield stressors.
- Camouflage, Distortion, Defilade, Vertical and Horizontal Coloration.







Late fusion of auxiliary information:

- **Attributes:** both individual and with constant viewpoint combination do not help improve classification performance.
  - Various attribute combinations with constant viewpoint do not improve accuracy .
- **GSC with attributes:** Do not improve classification performance.

Reasons for reduced performance:

Underperforming classes and stressors:

- **Classes:** HMMWV and Stryker.
- **Stressors:** Defilade and Camouflage.

Modality bias:

- Attributes from IR image modality improves performance for corresponding RGB images, but not for the IR image modality.

Table 1: Experimental results on modality bias.

Model	RGB Input			IR Input		
	Pure	+ IR Attr.	Δ	Pure	+ IR Attr.	Δ
ViT	86.1	**86.5**	+0.4	**82.5**	81	-1.5
CvT	77.5	**83.6**	+6.1	**75.1**	74.6	-0.5
ConViT	86.2	**89.3**	+3.1	**80.6**	76.4	-4.2
ConvNeXt V2	89.0	**89.3**	+0.3	87.8	**88.0**	+0.2

Foreground-background separability:

- Important vehicle parts are often blended in the background, due to the nature of IR imagery.

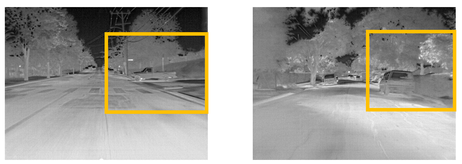


Fig.3: Lack of foreground-background color separability in real-world IR scenes

Reduced visual details and distortion of auxiliary cues:

- Critical parts are missing due to reduced visual details.
- Auxiliary cues prove disruptive rather than helpful, in such situations.

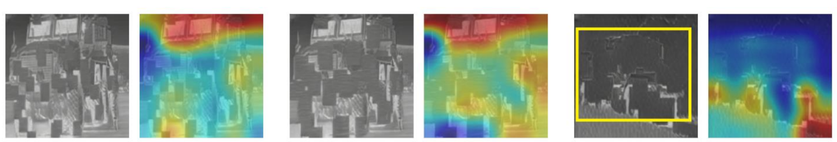


Fig.4: Late-fused GSC descriptor and attributes: disrupts attention as stress increases.

Conclusions and future recommendations:

- Auxiliary fusion provides only marginal gains (+0.2%) in stressed IR imagery, unlike its stronger effectiveness in RGB imagery.
- Key limitations include low visual detail, weak foreground-background thermal separability, and modality bias.
- Performance drops notably for HMMWV and Stryker, particularly under camouflage and defilade stressors.
- Future directions: physics-guided boundary learning, improved shape representations, and Mixture-of-Experts (MoE) architectures.

Acknowledgement:
This research was sponsored by the DEVCOM Analysis Center and was accomplished under Cooperative Agreement Number W911NF-22-2-0001.

Contact: {noyon.dey,nasidhabib.barna, suchi}@uga.edu