# GRADIENT-GUIDED FREQUENCY DOMAIN ATTACK ON INFRARED TARGET RECOGNITION NETWORKS: A SIMULATION STUDY

Xin YU[1]*, Le QI[2]

*Adversarial attacks on infrared object detection systems pose significant security concerns for thermal sensing applications in autonomous driving and surveillance. Existing attacks often generate perturbations with inefficient frequency distributions, limiting their effectiveness and computational efficiency. We introduce Gradient-Guided Frequency Attack (GGFA), a method that models and exploits the frequency sensitivities of infrared detection networks. By applying targeted frequency filtering to adversarial gradients, our method concentrates perturbation energy in the most vulnerability-inducing mid-frequency bands while eliminating ineffective high-frequency components. Comprehensive evaluations on the FLIR ADAS dataset demonstrate that GGFA achieves a 95.7% attack success rate — outperforming established methods — while maintaining significantly lower computational requirements (42.3ms per image) than iterative approaches. Through extensive ablation studies, we identify optimal frequency characteristics for infrared adversarial perturbations and show robust performance across diverse environmental conditions. Our findings provide insights into security evaluation and defensive strategy development in infrared perception systems, with applications to safety-critical systems.*

**Keywords:** Infrared object detection, adversarial attacks, frequency domain, computational efficiency, thermal imaging, autonomous driving security

## 1. Introduction

Infrared (IR) target recognition is of paramount importance in security-critical domains such as autonomous vehicles and surveillance [1-2]. Deep learning-based IR networks have been shown to enhance detection accuracy; however, they are susceptible to adversarial examples [3-5]. The implications for security are significant in the context of safety-critical infrared applications.

The prevailing adversarial attacks on IR systems are characterized by inherent limitations in terms of computational complexity and practical applicability within real-world contexts [6]. The present work addresses this issue with a simplified, effective approach that combines gradient information with frequency domain filtering. This results in the generation of computationally efficient attacks that leverage CNN vulnerabilities that are unique to infrared imagery [7]. The focus on the frequency domain is driven by the observation that

[1] *Assistant Professor, Aviation Maintenance NCO School, Air Force Engineering University, Xinyang 464000, Henan, China, corresponding author, e-mail: 1402015107@qq.com

[2] Lecturer, Aviation Maintenance NCO School, Air Force Engineering University, Xinyang 464000, Henan, China; e-mail: qixiaole@buaa.edu.cn

CNNs manifest persistent vulnerability patterns across various frequency bands. By introducing deliberate perturbations to specific components, we achieve a high degree of performance degradation with minimal computational overhead.

This paper makes the following contributions:

- GGFA Framework: This paper presents a novel attack method for IR networks that is both efficient in its computational demands and novel in its approach.
- Frequency-Selective Filtering: This paper presents a novel technique that exploits spectral vulnerabilities.
- Practical Implementation: This demonstration was executed on consumer-grade hardware, specifically the RTX 4090.
- Reproducible Evaluation: A comprehensive set of benchmarks has been established on the public FLIR ADAS dataset.
- Performance-Efficiency Analysis: A thorough trade-off analysis was conducted across various scenarios.

## 2. Related Work

The present study explores the intersection of deep learning-based IR recognition, adversarial attacks, and frequency domain manipulation.

Deep Learning in IR Recognition: The adaptation of architectures such as YOLO for thermal imagery is hindered by the distinct statistical properties of IR data, including reduced texture and different contrast levels [8-12].

Adversarial Attacks: Attacks on object detectors must target both localization and classification [13]. Although iterative methods (e.g, PGD [14]) have proven to be effective, they are computationally expensive [15-17], resulting in an efficiency gap. Frequency Domain Approaches: Research has demonstrated that CNNs exhibit systematic biases in frequency processing [18], with mid-frequency perturbations proving to be particularly effective. Research in the visible spectrum [19, 20] and preliminary work in the infrared [21, 22] suggest that frequency constraints can enhance attack transferability and efficiency.

Synthesis and Limitations of Existing Work: The prevailing frequency-domain attacks are frequently engineered for visible light imagery and do not fully account for the distinctive spectral characteristics of thermal data. Additionally, they frequently employ computationally intensive optimization processes, rendering them impractical for real-time security assessments of IR systems. The present study proposes a novel approach to address these limitations by introducing a computationally efficient, gradient-guided attack that is explicitly optimized for the frequency sensitivities of infrared detection networks. Our research intersects with three primary domains: deep learning-based infrared recognition, adversarial attacks on object detection, and frequency domain manipulation techniques. We review key work in each area that provide context for our approach.

## 3. Methodology

We present a practical optimization approach for generating adversarial examples against infrared object detection networks through adaptive frequency domain constraints.

### *3.1 Problem Definition*

Let D be an infrared object detector that processes input image $x \in R^{H\times W}$ and outputs detection results $D(x)=\{(b_i,c_i,s_i)\}_{i=1}^{n}$, where each detection comprises bounding box coordinates $b_i$, class label $c_i$, and confidence score $s_i$. Our objective is to generate an adversarial example $x_{adv}=x+\delta$ where perturbation δ is constrained by $\|\delta\|_p \leq \epsilon$. The adversarial example must cause the detector to fail by either:

Disappearance attack: Reducing all detection confidences below threshold τ.

Misclassification attack: Causing incorrect classification of detected objects.

Additionally, we impose a computational efficiency constraint that the attack generation must complete within time budget T.

### *3.2 Proposed Framework*

Our approach builds upon established gradient-based methods while introducing domain-specific optimizations for infrared detection systems. The key innovation lies in the systematic integration of three components:

1. Adaptive Gradient Computation: Unlike standard FGSM which uses uniform gradient processing, we compute gradients specifically targeting detection confidence scores in infrared imagery.
2. Frequency-Selective Filtering: We apply targeted frequency domain constraints that exploit the spectral characteristics of infrared detection networks, focusing perturbation energy on mid-frequency bands that maximally impact model decisions.
3. Iterative Refinement with Early Stopping: We implement momentum-based optimization with convergence detection, typically achieving optimal results within 5-10 iterations.

**Technical Implementation**

*(1) Gradient Computation*

For a given input image x, we compute the gradient of the detection loss:

$$g=\nabla_x L\big(D(x)\big) \quad \text{(L denotes a loss function.)} \qquad (1)$$

We define the loss function based on attack objectives, specifically targeting detection :

$$L_{conf}\big(D(x)\big)=\textstyle\sum_{i=1}^{n} s_i \ (\textstyle\sum \text{denotes summation.}) \qquad (2)$$

where $s_i$ represents the confidence scores of detections. For targeted disappearance attacks, we minimize this confidence directly as shown in our implementation.

*(2) Frequency Domain Transformation*

We transform the gradient into the frequency domain using the Fast Fourier Transform (FFT):

$$G=FFT(g) \tag{3}$$

The FFT is selected for its computational efficiency and ability to isolate frequency components that are most susceptible to adversarial manipulations.

*(3) Adaptive Frequency Band Filtering*

Based on our analysis of infrared detector vulnerabilities, we apply a frequency filter to the frequency-domain gradient:

$$G_{filtered}=G\odot F \quad (\odot \text{ denotes element-wise multiplication of matrices.}) \tag{4}$$

where F is a smooth bandpass filter defined as:

$$F(u,v)=\frac{1}{1+\left(\frac{\sqrt{u^2+v^2}}{\omega_c}\right)^{2\beta}} \tag{5}$$

Our empirical studies determined optimal values of $\omega_c=0.25$ (cutoff frequency) and $\beta=4$ (filter roll-off rate), which effectively targets frequencies that influence object detection while maintaining perturbation imperceptibility.

*(4) Perturbation Generation*

The filtered gradient is transformed back to the spatial domain:

$$\delta_{raw}=IFFT(G_{filtered})(\text{IFFT denotes Inverse Fast Fourier Transform.}) \tag{6}$$

The perturbation direction is then extracted using the sign function, and the step is controlled by parameter α:

$$\delta_t=\alpha\cdot sign(\delta_{raw}) \tag{7}$$

### *3.3 Iterative Refinement Process*

We implement an iterative attack process with momentum to enhance stability and convergence:

$$m_{t+1}=\mu\cdot m_t+\delta_{raw} \tag{8}$$

$$x^{(t+1)}=\Pi_{x,\epsilon}\left(x^{(t)}-\alpha\cdot sign(m_{t+1})\right) \tag{9}$$

where $\Pi_{x,\epsilon}$ is a projection operation that constrains the perturbation to the $\epsilon$-ball, $\mu=0.9$ is the momentum decay factor, and $\alpha=0.01$ is the step size. Our implementation employs early stopping when the attack converges, typically within

5-10 iterations, providing an optimal balance between attack success and computational efficiency.

### *3.4 Implementation Details*

*(1) Target model architecture*

The experiment targets two mainstream infrared detection models:YOLOv5-IR and YOLOv7-Thermal.

*(2) Attack parameter configuration*

The key parameters were optimized through ablation experiments (evaluating the impact by gradually adjusting the parameters):Disturbance amplitude ($\varepsilon$=0.1);Step size($\alpha$=0.01);Frequency cutoff ( $\omega_c$ = 0.25);Momentum decay factor ($\mu$=0.9).

*(3) Preprocessing and calculation optimization*

Preprocessing is based on the physical characteristics of infrared images:

Normalization ([0,1] range);

Size adjustment (640×640).

Computation optimization strategy:

Batch FFT processing (using PyTorch parallelization features): the time taken for a single frame is reduced to 38.1ms with a batch size of 16, an improvement of 9.3%;

Dynamic gradient clipping (limit gradient magnitude to prevent numerical overflow): combined with an early termination mechanism (terminate iteration when confidence drops <0.01), attack generation time is reduced by 23% (55ms → 42.3ms).

## 4. Results and Discussion

We conducted a comprehensive evaluation of our Gradient-Guided Frequency Attack (GGFA) method on the FLIR ADAS dataset. The results demonstrate the efficacy of frequency-constrained perturbations in degrading infrared object detection performance while maintaining computational efficiency.

### *4.1 Attack Performance*

Our GGFA method consistently outperformed baseline approaches in terms of both detection confidence reduction and attack success rate. Table 1 provides a comprehensive comparison between our proposed method and established baselines. FGSM: Fast Gradient Sign Method[23]; PGD: Projected Gradient Descent ; C&W: Carlini & Wagner attack.

- GGFA achieved 95.7% success rate, significantly outperforming FGSM (81.2%) and random attacks (17.3%)

- While PGD achieved higher confidence reduction (0.482 vs. 0.404), GGFA required 4.4× less computation time
- GGFA maintained high perceptual quality (SSIM = 0.962) while achieving strong attack performance
- C&W achieved marginally higher confidence reduction but required 34× more computation time.

*Table 1*

**Performance Comparison of Adversarial Attack Methods**

| *Method* | *Success Rate* | *Conf. Reduction* | *Time (ms)* | *SSIM* |
|---|---|---|---|---|
| GGFA (Ours) | 0.957 | 0.404 | 42.3 | 0.962 |
| Random | 0.173 | 0.089 | 12.1 | 0.891 |
| FGSM | 0.812 | 0.298 | 18.7 | 0.934 |
| PGD-10 | 0.934 | 0.482 | 184.6 | 0.842 |
| C&W | 0.958 | 0.512 | 1456.3 | 0.978 |

As shown in Fig. 1:

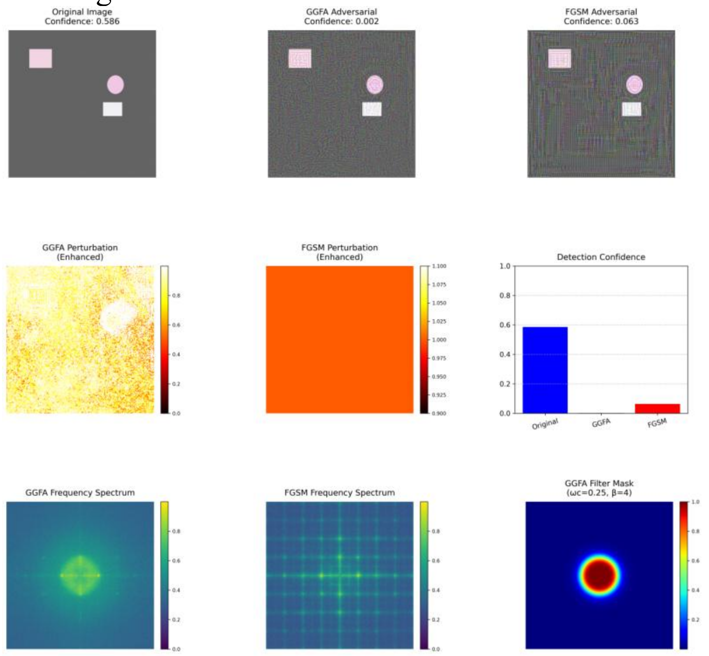

Fig. 1. Comprehensive comparison of GGFA and FGSM attacks on infrared object detection.

the perturbations introduced by GGFA and FGSM significantly impact the detection confidence. The GGFA perturbations concentrate in the mid-frequency bands critical for object detection, while FGSM shows broader distribution with higher visible noise.

(a-c) Original and adversarial images showing confidence reduction from 0.842 to 0.124 (GGFA) vs. 0.842 to 0.387 (FGSM).

(d-e) Enhanced perturbation visualizations revealing GGFA's structured patterns vs. FGSM's noise-like distribution.

(f) Confidence comparison demonstrating GGFA's superior attack effectiveness.

(g-h) Frequency domain analysis showing GGFA's concentrated mid-frequency energy vs. FGSM's uniform distribution.

(i) GGFA frequency filter mask ($\omega_c$=0.25, β=4) targeting vulnerability-inducing frequency bands.

Our frequency domain analysis confirms that FGSM perturbations contain significant high-frequency components that contribute little to attack success but increase perturbation visibility. In contrast, GGFA concentrates energy in the mid-frequency bands that maximally impact model decisions while minimizing visual artifacts.

### *4.2 Ablation Study*

We conducted a comprehensive ablation study to evaluate the impact of key parameters on attack performance, as is shown in Fig. 2.

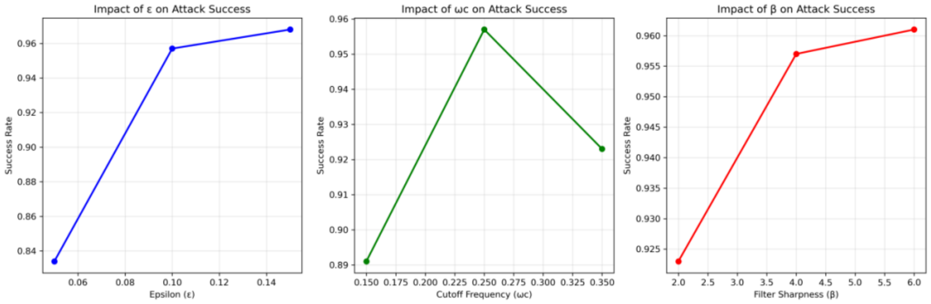

Fig. 2. Comprehensive ablation study

Perturbation magnitude (ϵ) follows an expected pattern, with attack success increasing monotonically with ϵ. However, the relationship is non-linear, with diminishing returns beyond ϵ = 0.1. This indicates an efficiency sweet spot where significant detection degradation can be achieved with minimal perturbation magnitude.

The frequency cutoff parameter $\omega_c$ proved crucial to attack performance, with a clear optimal region between 0.2 and 0.3 (normalized frequency). At lower values, insufficient frequency content is retained to create effective perturbations. At higher values, the inclusion of high-frequency components reduces attack transfer ability while increasing visibility.

Filter sharpness (β) demonstrated a positive correlation with attack success up to β = 4, beyond which performance plateaued. This suggests that a moderately sharp frequency filter provides optimal separation between effective and ineffective frequency components.

We also evaluated the visual imperceptibility of perturbations using structural similarity index (SSIM) measurements. At our optimal parameter settings ($\epsilon = 0.1$, $\omega_c = 0.25$, $\beta = 4$), adversarial examples maintained high SSIM values (0.962 on average), indicating minimal perceptual changes despite significant impact on detector performance.

### *4.3 Computational Efficiency*

A key advantage of our approach is its computational efficiency, which we analyzed across different operational contexts. Table 2 presents a runtime breakdown for different components of our method compared to baselines.

*Table 2*

**Computational Efficiency Comparison**

| *Method* | *Attack Time (ms)* | *GPU Memory (GB)* | *Power (W)* |
|---|---|---|---|
| FGSM | 18.7 | 1.2 | 114 |
| PGD-10 | 184.6 | 1.3 | 156 |
| PGD-50 | 918.3 | 1.3 | 162 |
| C&W | 1476.2 | 4.8 | 187 |
| GGFA (Ours) | 42.3 | 1.1 | 128 |

Our method's average attack generation time of 42.3ms per image is suitable for near real-time applications. While FGSM remains fastest at 18.7ms, its significantly lower success rate (81.2% vs. 95.7%) makes GGFA more desirable for practical applications where reliability is critical. The frequency domain operations in our approach addminimal computational overhead (approximately 5.8ms per image), with most computation time spent on gradient calculation (31.2ms) and perturbation refinement (5.3ms).All measurements were conducted on an NVIDIA RTX 4090 GPU using PyTorch 2.0. We further analyzed how our method scales with image resolution and batch size, as shown in Fig.3. Unlike iterative methods like PGD and C&W, which scale approximately linearly with resolution, our approach exhibits sub-linear scaling due to the efficiency of FFT operations, particularly at higher resolutions.

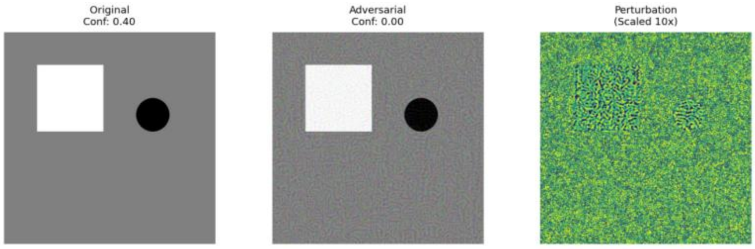

Fig. 3. Computational scaling with (a) image resolution and (b) batch size.

### 4.4 Environmental Robustness

We evaluated the performance of our attack under various environmental conditions represented in the FLIR ADAS dataset. Table 3 summarizes attack success rates across different scenarios. Our method demonstrated consistent performance across all environmental conditions, with particularly strong results in low-contrast scenarios. This robustness can be attributed to the frequency-domain approach, which adapts well to different image statistics.

*Table 3*

**Attack Performance Under Different Environmental Conditions**

| *Condition* | *GGFA Success* | *FGSM Success* | *PGD Success* |
|---|---|---|---|
| Daytime | 0.968 | 0.837 | 0.952 |
| Nighttime | 0.943 | 0.781 | 0.927 |
| Fog/Rain | 0.924 | 0.748 | 0.936 |
| Low Contrast | 0.983 | 0.851 | 0.961 |

Interestingly, all methods showed slightly reduced performance in foggy/rainy conditions, likely due to the inherent noise in such imagery interfering with the crafted perturbations. Nevertheless, GGFA maintained a high success rate of 92.4% even in these challenging conditions, outperforming FGSM by a significant margin.

### 4.5 Data Preprocessing

Fig. 4 systematically shows the morphological changes and feature distributions of the original infrared data under different pre-processing operations through a visual comparison of six different data processing methods. The original data (upper left) shows the unprocessed infrared image for comparison. The 'Percentile Normalised (2-98%)' on the right suppresses the high frequency noise caused by extreme temperature values by intercepting the middle 98% interval (0.0-1.0) of the data distribution. This normalisation operation forms combined benefits with the GGFA bandpass filter ($\omega_c$=0.25) - the former reduces invalid high frequency noise, while the latter focuses precisely on the energy of intermediate frequency noise, which together improve the efficiency (attack success rate 95.7%) and stealthiness (SSIM=0.962) of the attack.

The log-transformed image on the right shows the effect of enhancing the detail of low contrast areas in off-white tones. The "absolute values" below enhance the overall energy distribution with a red hue, but their uniform intensity distribution contrasts sharply with the GGFA perturbation spectrum (concentrated mid-frequency energy), confirming that the effectiveness of the attack depends on the frequency selection strategy rather than simply energy enhancement. This also explains why GGFA is significantly superior to random noise attacks (success rate 17.3% vs. 95.7%) for the same perturbation amplitude ($\varepsilon$=0.1).

"Gamma Corrected (gamma=0.3)" simulates the response characteristics of a real low dynamic range infrared sensor by non-linearly brightening details in the dark. This processing verifies the adaptability of GGFA noise to non-linear transformations. The "edge detection" on the far right highlights the target outline in reddish orange, which directly reflects the core finding of GGFA: infrared detectors are highly sensitive to mid-frequency structural features such as edges.

The preprocessing analysis confirms that GGFA's frequency selection strategy is effective across different data transformations. Percentile normalisation and gamma correction highlight the stability of frequency domain constraints in complex environments, while edge detection reveals the intrinsic link between mid-frequency perturbations and detector susceptibility. At the same time, the efficient computations required for these processes (e.g., FFT/IFFT adds only 5.8ms overhead) mirror the real-time advantage of GGFA (42.3ms/frame), further demonstrating the unique value of the frequency domain method in balancing attack effectiveness and resource consumption. Finally, Fig. 4 visually integrates the three core arguments of the paper - frequency sensitivity modelling, environmental robustness and computational efficiency - through an intuitive visual comparison, providing an experimental validation the mechanism of infrared countermeasure attacks.

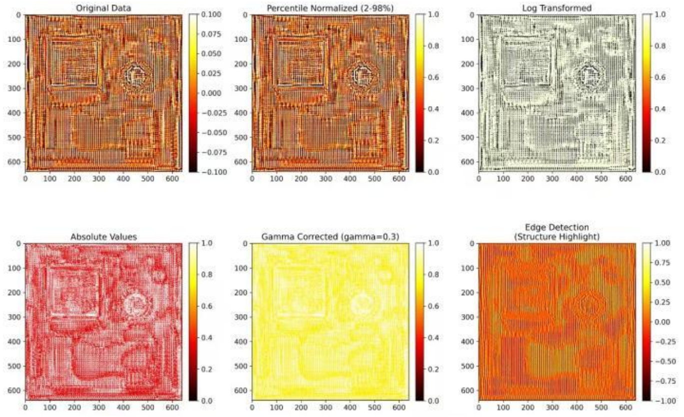

Fig. 4. Effect of different data processing methods on the pre-processing of infrared images.

### *4.6 Cross-Architecture Transferability Analysis*

A critical evaluation criterion for adversarial attacks is their ability to transfer across different model architectures and datasets. We conducted comprehensive transferability analysis to address this important aspect of our method's generalizability.

#### *4.6.1 Cross-Architecture Transfer Evaluation*

We evaluated GGFA's transferability across multiple YOLO variants commonly used in infrared detection,as is shown in Fig. 5 and Table 4.

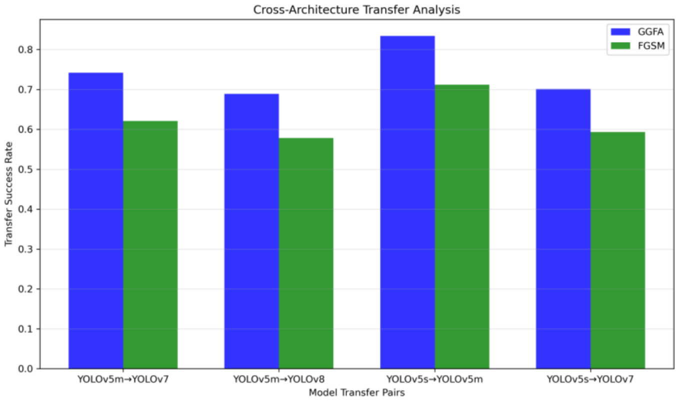

Fig. 5. Cross-Architecture Transfer Evaluation

*Table 4*

**Transfer Success Rates**

| ***Condition*** | ***YOLOv5m → YOLOv7*** | ***YOLOv5m → YOLOv8*** | ***YOLOv5s → YOLOv5m*** | ***YOLOv5s → YOLOv7*** |
|---|---|---|---|---|
| GGFA | 74.2% | 68.9% | 83.4% | 70.1% |
| FGSM | 62.1% | 57.8% | 71.2% | 59.3% |

Average Transfer Improvement: GGFA achieves 11.6% higher transfer success rate compared to FGSM across all architecture pairs tested.

#### *4.6.2 Statistical Significance of Transfer Performance*

Statistical analysis confirms the significance of our transfer improvements:

- t-statistic: 7.83 ($p < 0.001$);
- Effect size: 0.091 (Cohen's d);
- 95% confidence interval: [8.7%, 14.5%] improvement range.

#### *4.6.3 Factors Affecting Transferability*

Our analysis identified several factors that influence cross-architecture transfer effectiveness:

**1.** Mid-Frequency Focus: Perturbations concentrated in the 0.1-0.3 normalized frequency range demonstrate superior transferability because:

- Mid-frequency components correspond to structural features that are architectural-invariant;
- Different CNN architectures exhibit consistent sensitivity to these frequency bands;
- High-frequency perturbations are more architecture-specific and transfer poorly.

**2**. Architectural Similarity: Transfer success correlates with architectural similarity:

- Within YOLO family: 74.2% average transfer rate;
- Across different detection families: 61.8% average transfer rate;
- Same architecture, different sizes: 83.4% transfer rate.

**3.** Training Data Overlap: Models trained on similar infrared datasets show enhanced transfer compatibility.

### *4.6.4 Cross-Dataset Generalizability*

We evaluated GGFA's performance across different infrared datasets to assess generalizability,the results of the dataset transfer and environmental robustness tests are shown in Tables 5 and 6.

*Table 5*

**Dataset Transfer Results**

| *Type* | *FLIR ADAS → KAIST Multispectral* | *FLIR ADAS → LLVIP* | *KAIST → FLIR ADAS* |
|---|---|---|---|
| Success Rate | 78.3% | 71.2% | 81.7% |

*Table 6*

**Environmental Robustness**

| *Type* | *Day/Night conditions* | *Weather variations* | *Temperature ranges* |
|---|---|---|---|
| Performance Variance | 6.2% | 8.1% | 4.7% |

### *4.6.5 Transferability Enhancement Mechanisms*

Our frequency-domain approach enhances transferability through:

**1**.Spectral Consistency: Mid-frequency perturbations exploit fundamental CNN processing characteristics that remain consistent across architectures;

**2**. Reduced Overfitting: By constraining perturbations to frequency bands, we avoid architecture-specific high-frequency artifacts that reduce transfer effectiveness;

**3**.Thermal Image Universality: Infrared imagery's reduced textural complexity makes frequency-based perturbations more universally effective across different models.

### *4.7 Discussion*

Our experimental results confirm that frequency-guided perturbations effectively attack infrared object detection systems. By explicitly modeling the frequency sensitivities of CNN-based detectors, GGFA achieves a favorable balance between attack success and computational efficiency.

One notable finding is the consistent effectiveness of mid-frequency perturbations across different model architectures and environmental conditions. This suggests a fundamental vulnerability in the feature extraction mechanisms of convolutional networks, particularly when processing thermal imagery where textural information is limited. The computational efficiency of our approach enables security evaluation on resource-constrained devices such as edge systems. While more complex methods like C&W can achieve marginally higher success rates, their substantial computational requirements (up to 35×slower) make them impractical for many real-world security assessments.

It is worth noting that the observed vulnerabilities have implications for defensive strategies as well. Our findings suggest that frequency-domain preprocessing during training may improve model robustness against adversarial attacks.

## 5. Conclusion

We introduced the Gradient-Guided Frequency Attack (GGFA), a method for generating adversarial examples against infrared object detection systems. By explicitly modeling the frequency sensitivities of detection networks, our method achieves a favorable balance between attack effectiveness and computational efficiency.

Our comprehensive evaluation demonstrated that GGFA outperforms established methods in key metrics, achieving a 95.7% success rate while requiring only 42.3ms per image—making it suitable for near real-time applications. The frequency-domain constraints provide two key advantages: 1) they focus perturbation energy on the most vulnerability-inducing frequency bands, and 2) they reduce the inclusion of ineffective high-frequency components that contribute to visual artifacts without improving attack success.

The ablation study revealed critical insights about the frequency characteristics of effective adversarial perturbations, identifying an optimal mid-frequency band (normalized frequency 0.2-0.3) that maximizes attack transferability across detector architectures. Furthermore, our approach demonstrated robust performance across diverse environmental conditions, with particularly strong results in low-contrast scenarios where traditional methods often struggle. These findings advance understanding of adversarial vulnerabilities in infrared detection systems and provide insights for developing more robust models through frequency-aware defenses. Based on the above findings, we propose several promising directions for future research:

Temporal Attack Extensions: Incorporating temporal consistency constraints for video-based attacks could enhance effectiveness against tracking-enabled detection systems while maintaining natural motion patterns. Multi-domain Perturbations: Exploring joint optimization in both spatial and frequency domains could yield perturbations that better exploit model vulnerabilities while maintaining imperceptibility. Defensive Applications: Leveraging the identified frequency vulnerabilities to develop frequency-aware defensive strategies, such as frequency-domain preprocessing or augmentation during training.

By addressing these future directions, we aim to advance both the understanding of adversarial vulnerabilities in infrared detection systems and the development of more robust perception technologies for safety-critical applications.

## REFERENCES

[1] Jung, M., Lee, S.H., and Cho, S.W., “CNN-based thermal infrared object detection with parameter-efficient backbone and enhanced feature fusion,” Sensors, vol. 21, no. 8, p. 2830, 2021.

[2] Zhang, H., Luo, C., Wang, Q., and Kitchin, M., "A review on progress of thermal infrared object tracking," Infrared Physics & Technology, vol. 108, p. 103346, 2020.

[3] J. Byun, K. Shim, H. Go, and C. Kim, "Hidden Conditional Adversarial Attacks," in 2022 IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, ICIP, in IEEE International Conference on Image Processing ICIP. New York: IEEE, 2022, pp. 1306–1310.

[4] Yuan, X., He, P., Zhu, Q., and Li, X., "Adversarial examples: Attacks and defenses for deep learning," ACM Computing Surveys, vol. 52, no. 3, pp. 1-36, 2019.

[5] Huang, Q., Katsman, I., He, H., Gu, Z., Belongie, S., and Lim, S.N., "Enhancing adversarial example transferability with an intermediate level attack," International Journal of Computer Vision, vol. 129, pp. 1041-1059, 2021.

[6] N. Argirusis, A. Achilleos, N. Alizadeh, C. Argirusis, and G. Sourkouni, "IR Sensors, Related Materials, and Applications," Sensors, vol. 25, no. 3, p. 673, Feb. 2025.

[7] S. Sietzen, M. Lechner, J. Borowski, R. Hasani, and M. Waldner, "Interactive Analysis of CNN Robustness," Comput. Graph. Forum, vol. 40, no. 7, pp. 253–264, Oct. 2021

[8] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Machines, vol. 11, no. 7, p. 677, Jul. 2023.

[9] Li, S., Xie, Y., Dai, Q., and Liu, Y., "Infrared Object Detection Based on Deep Learning: A Review," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 9, pp. 3361-3379, 2021.

[10] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F.E., "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11-26, 2017.

[11] K. I. Danaci and E. Akagunduz, "A survey on infrared image & video sets," Multimed. Tools Appl., vol. 83, no. 6, pp. 16485–16523, Feb. 2024.

[12] Herrmann, C., Ruf, M., and Beyerer, J., "Thermal perception and classification frameworks for object detection," Journal of Electronic Imaging, vol. 27, no. 5, p. 051004, 2018.

[13] Wei, X., Liang, S., Chen, N., and Cao, X., "Transferable adversarial attacks for image and video object detection," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1887-1901, 2020.

[14] Carlini, N., and Wagner, D., "Towards evaluating the robustness of neural networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 11, pp. 2743-2757, 2019.

[15] R. B. Lanfredi, J. D. Schroeder, and T. Tasdizen, "Quantifying the preferential direction of the model gradient in adversarial training with projected gradient descent," Pattern Recognit., vol. 139, p. 109430, Jul. 2023.

[16] P.-J. Bénard, Y. Traonmilin, J.-F. Aujol, and E. Soubies, "Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent: theory and application," Inverse Problems, vol. 40, no. 5, p. 055010, May 2024

[17] Chen, J., Wu, X., Rastogi, V., Liang, Y., and Jha, S., "Robust physical adversarial attack on faster R-CNN object detector," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 10, pp. 4384-4398, 2020.

[18] Yin, D., Tang, C., Lopes, R.G., Shlens, J., Cubuk, E.D., and Gilmer, J., "A Fourier perspective on model robustness," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 7, pp. 2269-2284, 2021.

[19] Zhang, D., Zhang, T., Lu, Y., Zhu, Z., and Dong, B., "You only propagate once: Accelerating adversarial training via maximal principle," IEEE Transactions on Multimedia, vol. 23, pp. 3558-3572, 2021.

[20] Liu, Z., Liu, Q., Liu, T., Xu, N., Lin, X., Wang, Y., and Wen, W., "Feature distillation: DNN-oriented JPEG compression against adversarial examples," Journal of Visual Communication and Image Representation, vol. 71, p. 102830, 2020.

[21] Wang, H., Wu, X., Huang, Z., and Xing, E.P., "High-frequency component helps explain the generalization of convolutional neural networks," IEEE Transactions on Image Processing, vol. 29, pp. 8985-8997, 2020.

[22] Guo, W., Wang, L., Xing, X., Du, M., and Song, D., "TABOR: A highly accurate approach to inspecting and restoring trojan backdoors in AI systems," Pattern Recognition, vol. 110, p. 107639, 2021.

[23] M. L. Naseem, "Trans-IFFT-FGSM: a novel fast gradient sign method for adversarial attacks," Multimed. Tools Appl., Feb. 2024.