Evaluation of Neural Object Detection Models for Human Detection in Infrared Images

PROJECT REPORT T1000

from the course of studies Computer Science - Artificial Intelligence

at the Cooperative State University Baden-Württemberg
Ravensburg Campus Friedrichshafen

by

Lukas Florian Richter

13.08.2025

Completion Time:	16 Wochen
Student ID, Course:	None, TIK24
Company:	Airbus Defence & Space, Taufkirchen
Supervisor in the Company:	René Loeneke

Declaration of Authorship

In accordance with clause 1.1.13 of Annex 1 to §§ 3, 4 and 5 of the Cooperative State University Baden-Württemberg's Study and Examination Regulations for Bachelor's degree programs in the field of Technology, dated 29.09.2017. I hereby declare that I have written my thesis on the topic:

Evaluation of Neural Object Detection Models for Human Detection in Infrared Images

independently and have used no other sources or aids than those specified. I further declare that all submitted versions are identical.

Taufkirchen 13.08.2025

Lukas Florian Richter

Abstract

This project report evaluates the performance of neural object detection models for detecting humans in infrared images. The study focuses on comparing different variations of the SSD (Single Shot Mutlibox Detector) model architecture, assessing their accuracy and inference speed, and identifying the most suitable model for the given task. Additionally, different preprocessing techniques are evaluated to improve the detection performance.

More specifically, the main contributions of this project are:

- Conceptualization of a simple and cost-efficient hardware setup for the purpose of on-premise human detection in infrared images
- Evaluation of different SSD model architectures
- Comparison between different preprocessing techniques
- Identification of the most suitable model for the given task
- A theoretical pipeline for the secure transmission of the detection results to a remote server

Table of Contents

List of Figures

List of Tables

Code Snippets

List of Acronyms

AI	Artificial Intelligence
AP	Average Precision
API	Application Programming Interface
AdaBoost	Adaptive Boosting
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
DL	Deep Learning
FC	Fully Connected
FN	False Negative
FP	False Positive
FP16	16-bit floating point
FP32	32-bit floating point
FPS	Frames per Second
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HNM	Hard-Negative Mining
HOG	Histogram of Oriented Gradients
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
INT8	8-bit integer

IR	Infrared
IoU	Intersection over Union
ML	Machine Learning
MPS	Metal Performance Shaders
NMS	Non-Maximum Suppression
NN	Neural Network
OSI	Open Systems Interconnection
R-CNN	Region-based Convolutional Neural Network
RAM	Random Access Memory
REST	Representational State Transfer
RGB	Red, Green, Blue
RNN	Recurrent Neural Network
ROI	Region of Interest
RPN	Region Proposal Network
ReLU	Rectified Linear Unit
ResNet	Residual Network
SGD	Stochastic Gradient Descent
SSD	Single Shot MultiBox Detector
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TN	True Negative
TP	True Positive
TPU	Tensor Processing Unit
VGG	Visual Geometry Group
VOC	Visual Object Classes

ViT	Vision Transformer
YOLO	You Only Look Once
mAP	mean Average Precision

Glossary

Batch A batch is a group of data processed together as a unit.

Batch Gradient Descent Batch Gradient Descent is an optimization algorithm used to minimize the loss function in machine learning models by iteratively updating the model parameters based on their gradients with respect to the

Exploit An exploit is a method or piece of code that takes advantage of vulnerabilities in software, applications, networks, operating systems, or hardware, typically for malicious purposes.

Patch A patch is data that is intended to be used to modify an existing software resource such as a program or a file, often to fix bugs and security vulnerabilities.

Stochastic Gradient Descent Stochastic Gradient Descent (SGD) is an optimization algorithm used to minimize the loss function in machine learning models by iteratively updating the model parameters based on their partial derivatives with

Vulnerability A Vulnerability is a flaw in a computer system that weakens the overall security of the system.

1 Introduction

With the increase in security threats to critical infrastructure, automated surveillance systems have become essential for ensuring the safety and security of people, infrastructure and property at scale. The ability to detect individuals on critical infrastructure premises is crucial to preventing unauthorized access and potential damage to assets. While conventional RGB-based surveillance systems remain prevalent in many application, they face inherent limitations in challenging scenarios such as low-light conditions, adverse weather, fog, smoke and complete darkness during nighttime [1].

A compelling alternative to systems operating in the visible light domain are such capturing wavelengths in the infrared spectrum and thus offering consistent detection capabilities that are fundamentally independent of ambient lighting conditions. Unlike regular RGB cameras, thermal sensors detect light with longer wavelengths that correspond to the heat signatures of emitted directly by objects. This characteristic provides unique advantages for human detection, as the human body maintains a relatively constant temperature of approximately 37°C, creating distinct thermal signatures that remain visible regardless of environmental illumination [2].

The integration of deep learning architectures with thermal imaging thus opens new possibilities for automated systems that can reliably detect humans in the aforementioned scenarios with adverse conditions for conventional RGB-based concepts. However, most state-of-the-art object detection models have been primarily developed for and trained on RGB imagery. Given that the spectral, tectural and contrast characteristics of infrared images differ substantially from visible-light imagery, both due to the properties of those wavelengths themselves and of the sensors, those existing models might need to be adapted to achieve optimal performance.

This research addresses the critical need for systematic evaluation of neural object detection models specifically tailored for thermal human detection applications. The study focuses on the Single Shot MultiBox Detector (SSD) architecture, a prominent one-stage detection framework known for its balance between accuracy and computational efficiency. By examining multiple model variants with different backbone networks (VGG16 and ResNet152), initialization strategies (pretrained versus scratch training), and thermal-specific preprocessing techniques (image inversion and edge enhancement), this work provides comprehensive insights into optimal configurations for infrared surveillance systems.

This project addresses the need for systematic evaluation of neural object detection models and preprocessing techniques tailored for thermal human detection applications. For the purpose of developing and edge-deployable network, the focus of this study lies on the SSD architecture, a prominent one-stage detection framework with relatively low complexity compared to newer architectures like the Vision Transformer (ViT).

This work provides comprehensive insights into optimal configurations for infrared surveillance networks by examining multiple model variants with different:

1. **backbone networks** (Visual Geometry Group (VGG) and Residual Network (ResNet))
2. **initialization strategies** (parameters pretrained on the RGB-Dataset IMAGENET1K_V2 versus randomly sampled)
3. **preprocessing techniques** (inversion of the image or enhancement of its edges)

The practical significance of this research extends beyond academic interest, addressing real-world challenges faced by the security and defense industry. In partnership with Airbus Defence & Space, this project explores the development of cost-efficient, edge-deployable thermal surveillance solutions that can operate reliably in challenging environments where traditional RGB systems fail.

1.1 Research Objectives and Contributions

This thesis makes several key contributions to the field of thermal image processing and computer vision:

1. **Preprocessing Technique Analysis**: Quantitative evaluation of thermal-specific image enhancement methods, including polarity inversion and edge enhancement, and their impact on detection accuracy.

2. **Backbone Network Comparison**: Detailed comparison between VGG16 and ResNet152 architectures in the context of thermal imagery, addressing the trade-offs between model complexity and performance.

3. **Practical Implementation Guidelines**: Development of actionable recommendations for deploying thermal surveillance systems in real-world environments, considering computational constraints and accuracy requirements.

4. **Dataset Integration Framework**: Unified evaluation approach across five diverse thermal datasets (FLIR ADAS v2, AAU-PD-T, OSU-T, M3FD, KAIST-CVPR15), enabling robust performance assessment.

1.2 Thesis Organization

The remainder of this thesis is structured to provide a comprehensive examination of thermal human detection using neural networks. Section 2 presents a thorough review of object detection fundamentals, SSD architecture principles, and thermal image processing techniques, establishing the theoretical foundation for the experimental work. Section 3 details the systematic approach employed for model evaluation, including dataset preparation, experimental design, and evaluation metrics. Section 4 presents comprehensive performance analysis across all model configurations and preprocessing techniques. Section 5 interprets the findings within the context of practical deployment scenarios and industrial requirements. Finally, Section 6 synthesizes the key contributions and outlines directions for future research in thermal surveillance technologies.

2 Literature Review and Theoretical Background

The field of object detection has undergone significant evolution from traditional computer vision techniques to sophisticated deep learning architectures. Understanding this progression is essential for contextualizing the current work's contribution to thermal image analysis. This section examines the theoretical foundations of object detection, with particular emphasis on the Single Shot MultiBox Detector (SSD) architecture and its applicability to thermal imagery processing challenges.

Key areas to develop:

- Evolution from traditional methods (HOG, SIFT) to deep learning
- Comparison of one-stage vs. two-stage detection models
- SSD architecture fundamentals and anchor box mechanisms
- Backbone network analysis (VGG vs. ResNet trade-offs)
- Thermal imaging characteristics and preprocessing challenges
- Existing work on infrared human detection
- Gap analysis: Limited research on SSD for thermal surveillance

2.1 Object Detection Fundamentals

Most object detection methods can be broadly categorized into two main approaches: traditional methods and deep learning-based methods. Traditional methods mainly rely on handcrafted features and sliding window techniques, while deep learning-based methods in this field leverage convolutional neural networks (CNNs) or vision transformer (ViT) architectures to automatically learn features from data.

2.1.1 Traditional Object Detection Methods

Simple approaches to object detection entail applying manually constructed feature detector kernels in a sliding window fashion to images.

An example of this is the **Viola-Jones-Algorithm** [3]:

1. Compute the integral image of the input image, that is, the sum of pixel intensities from the top-left corner of the image to each pixel. This allows for quick computation of the sum of pixel intensities in any rectangle in the image by subtracting the value of the upper left pixel of the rectangle from that of the lower right pixel.

2. Apply a series of Haar-like features to detect potential objects. Haar-like features are computed by subtracting the sum of pixels in one rectangle from the sum of pixels in an adjacent rectangle. These features capture various simple patterns, such as edges and lines.

3. Use the Adaptive Boosting (AdaBoost) technique to build a cascaded strong classifier consisting of several weak classifiers that can detect simple patterns consisting of Haar-like features.

4. Split the image into subwindows and classify each subwindow using the cascaded classifier as either containing the object or not.

Other approaches employ Histogram of Oriented Gradients (HOG) descriptors. The HOG is attained by dividing the image into a grid of cells, contrast-normalizing them and then computing the vertical as well as horizontal gradients of their pixels. The gradients for each cell are accumulated in a one-dimensional histogram which serves as that cell's feature vector. After labeling the cells in the training data, a Support Vector Machine (SVM) can be trained to find an optimal hyperplane separating the feature vectors corresponding to the object that should be detected from those that do not contain the object.

2.1.2 Deep Learning-Based Object Detection

Discusses the evolution of deep learning models, including R-CNN, Fast R-CNN, Faster R-CNN, and YOLO, highlighting their strengths and limitations.

However, those methods are highly dependent on engineering the correct priors, such as the Haar-like features, and

2.2 Single Shot MultiBox Detector (SSD) Architecture

Detailed explanation of SSD model architecture, including backbone networks (VGG, ResNet) and detection mechanisms.

2.2.1 Backbone Networks for Feature Extraction

Explores the role of backbone networks (VGG, ResNet) in feature extraction and their impact on SSD performance.

2.2.2 Feature Maps and Anchor Boxes

Describes the multi-scale feature maps and anchor boxes used in SSD for object detection.

2.2.3 MultiBox Loss Function

Explains the MultiBox loss function that combines localization loss and confidence loss for training SSD models.

Non-Maximum Suppression (NMS):

Examines the NMS technique used to filter duplicate detections and improve detection accuracy.

2.3 Thermal Image Processing

Discusses characteristics of thermal images, preprocessing techniques (inversion, edge enhancement), and challenges specific to infrared imagery.

3 Methodology

This study employs a systematic experimental approach to evaluate the effectiveness of SSD-based neural networks for human detection in thermal imagery. The methodology encompasses dataset selection and preparation, implementation of multiple model variants with different backbone architectures, application of thermal-specific preprocessing techniques, and comprehensive evaluation metrics. The experimental design ensures reproducible results while addressing the unique challenges posed by infrared image characteristics.

Key areas to develop:

- Dataset description: FLIR ADAS v2, AAU-PD-T, OSU-T, M3FD, KAIST-CVPR15
- Model configurations: SSD300-VGG16 vs. SSD300-ResNet152
- Training setup: Pretrained vs. scratch initialization strategies
- Preprocessing techniques: Image inversion and edge enhancement
- Data augmentation and split strategies (train/validation/test)
- Evaluation metrics: mAP, precision, recall, inference speed
- Hardware setup and computational requirements
- Statistical significance testing approach

3.1 Dataset Description

Details the thermal image datasets (FLIR ADAS v2, AAU-PD-T, OSU-T, M3FD, KAIST-CVPR15) and their characteristics.

3.2 Model Implementation

Explains the implementation of SSD models with different backbones and preprocessing configurations.

3.3 Experimental Design

Outlines the systematic approach to comparing model variants and the evaluation framework.

4 Results and Analysis

The experimental evaluation reveals significant performance variations across different model configurations and preprocessing approaches when applied to thermal human detection tasks. This section presents comprehensive results from training 16 distinct model variants, combining backbone architectures (VGG16 vs. ResNet152), initialization strategies (pretrained vs. scratch), and preprocessing techniques (none, inversion, edge enhancement, combined). The analysis demonstrates clear patterns in model behavior and identifies optimal configurations for thermal surveillance applications.

Key areas to develop:

- Training convergence analysis: Loss curves and stability patterns
- Detection accuracy results: mAP scores across all model variants
- Preprocessing impact: Quantitative comparison of enhancement techniques
- Backbone architecture comparison: VGG16 vs. ResNet152 performance
- Initialization strategy effects: Pretrained vs. scratch training outcomes
- Computational efficiency: Inference speed and memory requirements
- Dataset-specific performance: Results breakdown by thermal dataset
- Error analysis: Common failure cases and detection limitations

4.1 Training Performance

Reports training loss curves, converegence behavior, and computational requirements for different mdoel variants.

4.2 Detection Accuracy Analysis

Provides detailed mAP scores and detection performance metrics for each model configuraiton and preprocessing technique.

4.3 Preprocessing Impact Evaluation

Analzyes the effects of image inversion and edge enhancement on detection performance.

5 Discussion

The experimental results provide valuable insights into the practical applicability of SSD architectures for thermal human detection systems. While certain configurations demonstrate superior performance, the choice of optimal model depends on specific deployment requirements, including accuracy thresholds, computational constraints, and operational environments. This section interprets the findings within the context of real-world surveillance applications and addresses the broader implications for thermal imaging-based security systems.

Key areas to develop:

- Performance trade-offs: Accuracy vs. computational efficiency analysis
- Preprocessing effectiveness: When and why certain techniques work better
- Backbone selection criteria: Situational advantages of VGG16 vs. ResNet152
- Real-world deployment implications: Edge computing considerations
- Limitations and constraints: Environmental factors affecting performance
- Comparison with existing thermal detection systems
- Cost-benefit analysis for industrial implementation
- Future optimization potential and research directions

5.1 Model Performance Comparison

Compares SSD-VGG and SSD-ResNer performance and discusses trade-offs between accuracy and computational efficiency.

5.2 Practical Deployment Considerations

Discusses real-world application scenarios and system requirements for thermal surveillance.

6 Conclusion and Future Work

This thesis has systematically evaluated the application of Single Shot MultiBox Detector architectures for human detection in thermal imagery, providing empirical evidence for optimal model configurations in surveillance applications. The comprehensive analysis of 16 model variants across multiple thermal datasets has yielded practical insights for deploying neural networks in infrared-based security systems. The findings contribute to both academic understanding and industrial implementation of thermal computer vision technologies.

Key areas to develop:

- Key findings summary: Best-performing model configurations identified
- Methodological contributions: Systematic evaluation framework for thermal detection
- Practical implications: Guidelines for industrial thermal surveillance deployment
- Technical achievements: Successful adaptation of RGB models to thermal domain
- Research limitations: Dataset constraints and environmental factors
- Future research directions: Advanced architectures and multi-modal approaches
- Industry impact: Potential applications beyond security surveillance
- Recommendations: Implementation guidelines for practitioners

7 Examples

Just a couple of examples to demonstrate proper use of the typst template and its functions.

7.1 Acronyms

Use the `acr` function to insert acronyms, which looks like this Hypertext Transfer Protocol (HTTP).

Application Programming Interfaces are used to define the interaction between different software systems.

REST is an architectural style for networked applications.

7.2 Glossary

Use the `gls` function to insert glossary terms, which looks like this:

The Stochastic Gradient Descent is an optimization algorithm used in Machine Learning.

7.3 Lists

Create bullet lists or numbered lists.

- This
- is a
- bullet list

1. It also
2. works with
3. numbered lists!

7.4 Figures and Tables

Create figures or tables like this:

7.4.1 Figures

Figure 1 — Image Example

7.4.2 Tables

	Area	Parameters
cylinder.svg	$\pi h \frac{D^2 - d^2}{4}$ (1)	h: height D: outer radius d: inner radius
tetrahedron.svg	$\frac{\sqrt{2}}{12} a^3$ (2)	a: edge length

Table 1 — Table Example

7.5 Code Snippets

Insert code snippets like this:

```
const ReactComponent = () => {
  return (
    <div>
      <h1>Hello World</h1>
    </div>
  );
};

export default ReactComponent;
```

Listing 1 — Codeblock Example

7.6 References

Cite like this K. R. Akshatha, A. K. Karunakar, S. B. Shenoy, A. K. Pai, N. H. Nagaraj, and S. S. Rohatgi [2]. Or like this [1].

You can also reference by adding `<ref>` with the desired name after figures or headings.

For example this Table 1 references the table on the previous page.

References

[1] M. A. Farooq, P. Corcoran, C. Rotariu, and W. Shariff, “Object Detection in Thermal Spectrum for Advanced Driver-Assistance Systems (ADAS),” no. arXiv:2109.09854. arXiv, Oct. 2021. doi: 10.48550/arXiv.2109.09854.

[2] K. R. Akshatha, A. K. Karunakar, S. B. Shenoy, A. K. Pai, N. H. Nagaraj, and S. S. Rohatgi, “Human Detection in Aerial Thermal Images Using Faster R-CNN and SSD Algorithms,” *Electronics*, vol. 11, no. 7, p. 1151, Jan. 2022, doi: 10.3390/electronics11071151.

[3] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA: IEEE Comput. Soc, 2001, p. I–511–I–518. doi: 10.1109/CVPR.2001.990517.