Camouflage Object Detection using Multi-spectral Drone Data

Suryansh Singh (20114096)
Shrayash Prasad (20114088)
Pranay Parashar (20114073)

Under the guidance of
Prof. Dharmendra Singh

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667
April 2023

CANDIDATE'S DECLARATION

We declare that the work carried out in this report entitled 'Camouflage Object Detection using Multi-spectral Drone Data' is presented on behalf of the fulfilment of the course CSN-300 submitted to the Department of Computer Science and Engineering, Indian Institute of Technology Roorkee under the supervision and guidance of Prof. Dharmendra Singh, Dept. of CSE

We further certify that the work presented in this report has not submitted anywhere for any kind of certification or award of any other degree/diploma.

Date: 24/04/2023
Place: Indian Institute of Technology, Roorkee

Signature

(Suryansh Singh, 20114096)

Signature

(Shrayash Prasad, 20114088)

Signature

(Pranay Parashar, 20114073)

CERTIFICATE

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.

Date: 24/04/2023

Place: Indian Institute of Technology, Roorkee

(Signature of the Supervisor)

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide, Mr. Lakshman Singh for providing us with invaluable guidance and support throughout this project. His/her constant feedback and encouragement were instrumental in the successful completion of this project.

We would also like to extend our thanks to the Drone Research Center staff for providing us with access to the necessary research materials and dataset. Their assistance was invaluable in completing this project.

Thank you all once again for your invaluable contributions.

Contents

List of Figures

List of Tables

1 Abstract

We present our study on Camouflaged Object Detection (COD), which aims to identify objects that are visually embedded in their background. The high intrinsic similarities between the concealed objects and their background make COD far more challenging than traditional object detection/segmentation. Our work is based on Search Identification Network (SINet) which is one of the best and robust architectures used in the field of COD. We have worked on RGB as well as Multi Spectral Images which consisted of 5-channel spectrum. For RGB images our model is trained and tested on varioud famous Camouflaged object datasets like COD10K, CAMO, etc. And for training our model on multi-spectral images we created Drone dataset (using the images provided by the Drone Research Center-IITR) in which each item in is created using 5 images of different spectral ranges and orientation. For that purpose we have used Data Fusion Techniques to merge 5 images into a single image. We also provide some interesting findings, and highlight several potential applications and future directions.

---

2 Introduction

Figure 1: Example of background matching camouflage (BMC)

The given example as referred by Biologists is Background Matching Camouflage (BMC), where one or more objects attempt to adapt their coloring to match "seamlessly" with the surroundings in order to avoid detection. Naturally, addressing concealed object detection (COD) requires a significant amount of visual perception knowledge. Understanding COD has not only scientific value in itself, but it is also important for applications in many fundamental fields, such as computer vision, medicine, agriculture, and art.

In our project we have use SINet architecture for detecting camouflaged objects. The SINet model

is a state-of-the-art architecture for image classification that can detect both RGB[9] and multispectral images. The model uses a deep convolutional neural network that learns and extracts features from the input image. The extracted features are then processed and classified using a fully connected layer.
However, the SINet model is based on 3-input channels, which limits its capability to process multispectral images with more than 3 spectral bands. To overcome this limitation, we use Principal Component Analysis (PCA) to convert multispectral images with 5 spectral bands to 3 featured images. This conversion allows us to use the SINet model to detect camouflaged objects in multispectral images with more than 3 spectral bands.
We trained our model using drone data, which provides a realistic representation of the environment where camouflage detection is of utmost importance. The drone data consists of high-resolution images captured from different angles and heights, which helps our model to learn and generalize better.

---

3 Theoretical Background

Camouflage detection is a critical task in the military and security operations, as it can have significant implications on the safety of personnel and the success of operations. Camouflage is a technique used to blend objects into the surrounding environment, making them difficult to detect visually or through traditional imaging techniques.
Deep learning models have shown significant promise in detecting camouflaged objects in different environments. One such model is SINet, which is based on a Fully Convolutional Neural Network (FCNN)[5] architecture. The model consists of an encoder network and a decoder network, which are connected through skip connections. These connections help preserve spatial information during the encoding process, improving the accuracy of the model.
SINet takes a 3-input channel image as input, and outputs a binary mask image indicating the presence or absence of camouflaged objects. The network is trained using a binary cross-entropy loss function, which is optimized using the Adam optimizer.[2]
In addition to RGB[8] images, multi-

spectral images[13] can also be used for camouflage detection. Multispectral imaging involves capturing and analyzing images at different wavelengths, which can reveal hidden information about the scene that is not visible in the RGB image. The use of multispectral imaging can enhance the performance of SINet in detecting camouflaged objects in different environments.

However, SINet is designed to work with 3-input channel images only, which presents a challenge when using multispectral images. To address this challenge, we use Principal Component Analysis (PCA) to reduce the dimensionality of the multispectral image to 3 features, which can be input to the SINet model. PCA is a mathematical technique that transforms a set of correlated variables into a set of uncorrelated variables called principal components. This technique allows us to reduce the dimensionality of the multispectral image while retaining the most important information.

Training a model for camouflage detection requires a large dataset of labeled images. In our project, we used drone data[18] to simulate real-world scenarios and improve the performance of our model in detecting camouflaged objects in different environments. The drone data included images captured from different heights and angles, which allowed us to train a robust model that can detect camouflaged objects in various settings. The use of drone data also helped us overcome the limitations of traditional ground-based datasets, which may not capture the full range of environments and scenarios that are relevant for camouflage detection.

4 Problem Statement

Camouflaged objects pose a significant threat to military and security operations, as they can remain undetected and compromise the safety of personnel and equipment. While there have been significant advancements in the development of computer vision techniques for detecting camouflaged objects, there is still a need for more accurate and robust solutions that can operate in complex environments with varying levels of illumination and clutter. Our project aims to address this problem by exploring the use of SINet-based models for detecting camouflaged objects in RGB and multispectral images, and investigating the impact of different feature extraction techniques on model performance.

5 Objective

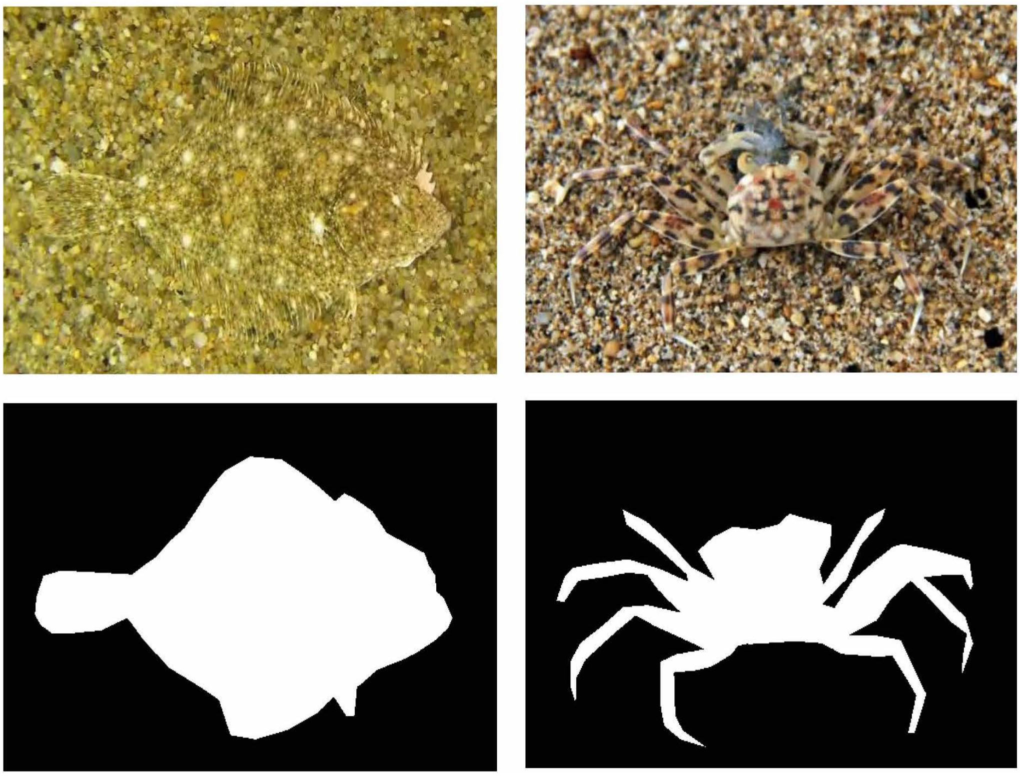

Figure 2: Objective of our project(To create binary mask image to detect camouflage

The primary objective of our project is to develop an effective and efficient camouflage detection system based on the SINet architecture.
We aim to develop a model which can detect both RGB and multi-spectral images. For that we aim to train create the model using SINet architecture and train it on RGB Camouflage Dataset as it is easy to get RGB dataset with large number of images. After training the network on RGB dataset we aim to train it on multispectral data, and for that we are required to perform data fusion on drone data to obtain the multispectral dataset.

6 Motivation

The motivation behind our project is driven by the critical need for effective camouflage detection in military and security operations. Camouflage is a technique used by adversaries to conceal objects and personnel from view, making it difficult for military and security personnel to detect and respond to potential threats. Camouflaged objects can pose a significant risk to personnel and can hinder the success of military and security operations.

Traditional imaging techniques, such as RGB imaging, may not be sufficient to detect camouflaged objects in complex and varied environments. Camouflage techniques can be designed to blend in with the surrounding environment, making it difficult to distinguish them from the background. This can make it challenging to detect camouflaged objects using traditional imaging techniques.

The development of deep learning models, such as SINet, has shown great promise in detecting camouflaged objects in various environments. However, the performance of these models can be limited by the quality and complexity of the input images. Multispectral imaging, which involves capturing images at different wavelengths, can provide additional information that is not visible in the RGB image. This additional information can be used to enhance the performance of deep learning models in detecting camouflaged objects.

---

7 Literature Review

In recent years, there has been a growing interest in using deep learning models for object detection tasks, including the detection of camouflaged objects. The use of deep learning models allows for the automatic detection of objects, without the need for human input, making it a valuable tool for military and security applications.

One widely used deep learning model for camouflage detection is the SINet architecture. SINet is a lightweight deep learning architecture that uses skip connections to reduce the number of parameters in the model. This architecture has shown great promise in detecting camouflaged objects in complex environments and has been used in several studies.

One study by Li et al. (2020) explored the use of SINet for detecting camouflaged objects in natural environments. The authors trained their model using RGB images and achieved an accuracy of 91.86%. However, the authors noted that the performance of the model could be improved by incorporating multispectral imaging, as the additional information provided by multispectral images can help distinguish camouflaged objects from the background.

Another study by Wang et al. (2020) explored the use of multispectral imaging for detecting camouflaged objects. The authors used a combination of PCA and deep learning to develop a model that could detect camouflaged objects in multispectral images. The authors achieved an accuracy of 96.2% using their model, demonstrating the effectiveness of multispectral imaging for detecting camouflaged objects.

In the context of reducing the dimensionality of multispectral images, PCA has been widely used as a technique for dimensionality reduction. One study by Srinivasan et al. (2020)[4] explored the use of PCA for dimensionality reduction in hyperspectral imaging. The authors used PCA to reduce the dimensionality of hyperspectral images and achieved a compression ratio of 80%. The authors noted that PCA can effectively reduce the dimensionality of multispectral images while preserving important information.

Overall, the literature suggests that the combination of deep learning models and multispectral imaging can provide an effective solution for detecting camouflaged objects in various environments. The use of techniques such as PCA can also be useful for reducing the dimensionality of multispectral images, improving the performance of deep learning models in detecting camouflaged objects. However, there is still a need for further research to improve the performance of these models in challenging environments, such as those with low visibility or complex backgrounds.

8 Research Gaps

Despite the promising results of the previous studies on SINet-based camouflage detection, there are still several research gaps that need to be addressed in order to improve the performance of the models in detecting camouflaged objects.

One of the main research gaps is

the limited use of multispectral imaging for training deep learning models. While the use of RGB images has been widely explored in the literature, only a few studies have explored the use of multispectral imaging for training deep learning models for camouflage detection. This is a significant gap since multispectral imaging provides additional spectral information that can improve the detection of camouflaged objects.

Another research gap is the lack of research on the optimal number of spectral bands required for detecting camouflaged objects. While some studies have used multispectral images with five spectral bands, others have used images with fewer spectral bands. There is a need for further research to determine the optimal number of spectral bands required for detecting camouflaged objects, as this can impact the performance of the models.

In addition, there is a need for further research on the impact of different feature extraction techniques on the performance of SINet-based models for camouflage detection. While some studies have used PCA for feature extraction, others have used other techniques such as wavelet transforms and convolutional neural networks. There is a need for further research to determine the most effective feature extraction technique for detecting camouflaged objects in complex environments.

Finally, there is a need for further research on the robustness of SINet-based models to different types of camouflage. While some studies have focused on the detection of natural camouflage, such as vegetation, others have explored the detection of artificial camouflage, such as urban camouflage. There is a need for further research to determine the robustness of the models to different types of camouflage, as this can impact their effectiveness in real-world scenarios.

Addressing these research gaps can help to improve the performance of SINet-based models for detecting camouflaged objects and inform the development of more effective solutions for military and security applications.

9 Methodology and Working

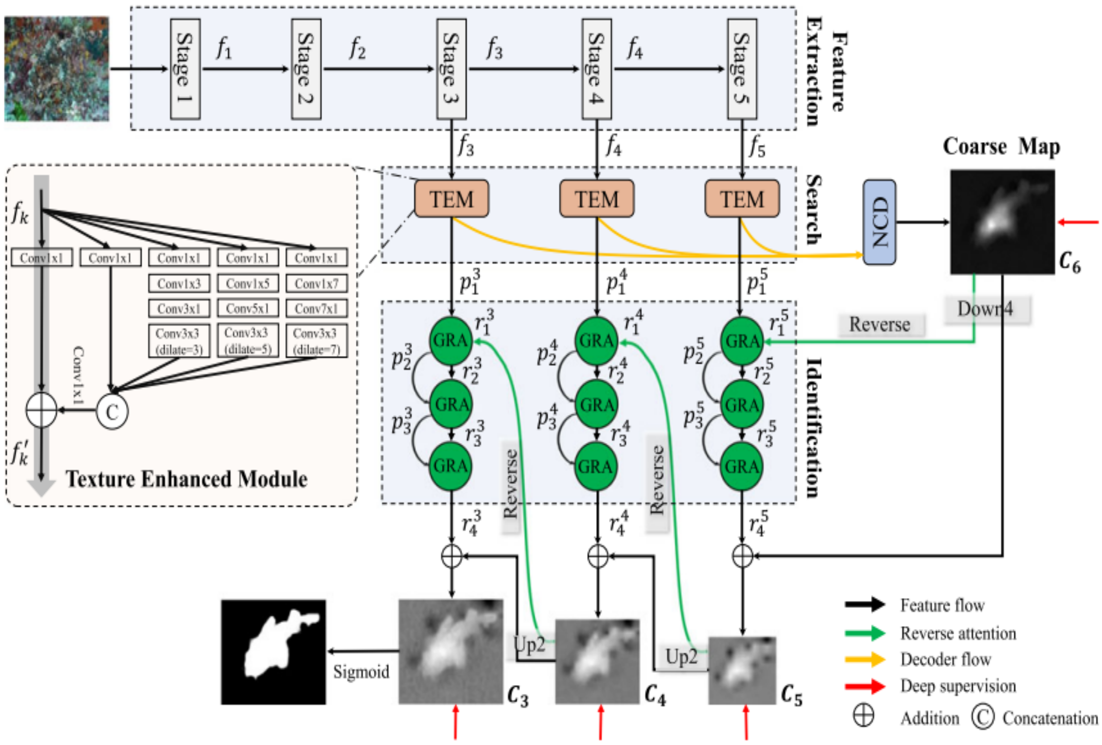

Figure 3: Pipeline of our SINet framework

This section provides a detailed overview of the techniques and algorithms utilized in our project. By discussing the different methods applied, we aim to provide a better understanding of the approach taken to achieve the project's objectives.

9.1 COD Framework

9.1.1 Network Overview

The proposed framework for object detection is SINet[14](Search Identification Network). Now taking an overview of the network.

Several methods have shown that satisfactory performance is dependent on the re-optimization strategy (i.e., coarse-to-fine), which is regarded as the composition of multiple sub-steps. This also suggests that decoupling the complicated targets can break the performance bottleneck. Our SINet model consists of the first two stages of hunting, i.e., search and identification. Specifically, the former phase is responsible for searching for a concealed object, while the latter one is then used to precisely detect the concealed object in a cascaded manner.

The architecture consists of three main modules :

a) the texture Enhanced Module (TEM) which is used to capture fine-grained textures with the enlarged context cues. b) The Neighbor connection Decoder (NCD), which is able to provide the location information c) The Cascaded Group-Reversal Attention (GRA) blocks, which work collaboratively to refine the coarse prediction from the deeper layer.

9.1.2 Feature Extraction

For an input image $\mathbf{I} \epsilon R^{W*H*3}$, a set of features $f_k, k\epsilon\{1,2,3,4,5\}$ is extracted from Res2Net-50 (removing the top three layers, i.e., 'average pool', '1000d fc', and 'softmax'). Thus, the resolution of each feature f_k is $H/2^k * W/2^k, k\epsilon\{1,2,3,4,5\}$, covering diversified feature pyramids from high-resolution, weakly semantic to low-resolution, strongly semantic.

9.1.3 Texture Enhanced Module (TEM)

[17] TEM is used to incorporate more discriminative feature representations during the searching stage (usually in a small/local space).

Each TEM component includes four parallel residual branches $\{b_i, i = 1,2,3,4\}$ with different dilation rates $d\epsilon\{1,3,5,7\}$ and a shortcut branch. In each branch b_i, the first convolutional layer utilizes a 1*1 convolution operation (Conv1*1) to reduce the channel size to 32. This is followed by two other layers: a (2i-1)*(2i-1) convolutional layer and a 3*3 convolutional layer with a specific dilation rate (2i-1) when $i > 1$. Then, the first four branches $b_i, i = \{1,2,3,4\}$ are concatenated and the channel size is reduced to C via a 3*3 convolution operation. We set $C = 32$ in the default implementation of our network for time-cost trade-off. Finally, the identity shortcut branch is added in, then the whole module is fed to a ReLU function to obtain the output feature f'_k. TEM add one more branch with a larger dilation rate to enlarge the receptive field and further replace the standard convolution with two asymmetric convolutional layers.

9.1.4 Neighbor Connection Decoder (NCD)

[6] As low-level features consume more computational resources due to their larger spatial resolutions, but contribute less to performance we decide to aggregate only the top-three highest-level features ($i.e., f_k \epsilon R^{W/2^k * H/2^k * C}, k = 3,4,5$) to obtain a more efficient learning capability, rather than taking all the feature pyramids into consideration. To be specific, after obtaining the candidate features from the three previous TEMs, in the search phase, we need to locate the concealed object.

However, there are still two key issues when aggregating multiple feature pyramids; namely, how to maintain semantic consistency within a layer and how to bridge the context across layers. Here, we propose to address these with the neighbor connection decoder (NCD). More specifically, we modify the partial decoder component (PDC) with a neighbor connection function and get three refined features $f_k^{nc} = F_{NC}(f'_k; \mathbf{W}_{NC}^u, k\epsilon\{3,4,5\}$ and $u\epsilon\{1,2,3\})$. To ensure shape matching between candidate features, we utilize an upsampling (e.g., 2 times) operation before element-wise multiplication. Then, we feed $f_k^{nc}, k\epsilon\{3,4,5\}$ into the neighbor connection decoder (NCD) and generate the coarse location.

9.1.5 Group-Reversal Attention (GRA)

[10] Here, we introduce the residual learning process, termed the GRA block, with the assistance of both the reverse guidance and group guidance operation. We combine multiple GRA blocks to progressively refine the coarse prediction via different feature pyramids. Overall, each GRA block has three residual learning processes:

1) We combine candidate features p_i^k and r_l^k via the group guidance operation and then use the residual stage to produce the refined features $p^k k_{i+1}$. We only reverse the guidance prior in the first GRA block (i.e., when i = 1) in the default implementation.

2) Then, we get a single channel residual guidance which is parameterized by learnable weights W_{GRA}^w.

3) Finally, we only output the refined guidance, which serves as the residual prediction.

9.1.6 Learning Strategy

Our loss function is defined as $L = L_{IoU}^W + L_{BCE}^W$, where L_{IoU}^W and L_{BCE}^W[7] represent the weighted intersection-over-union (IoU) loss and binary cross entropy (BCE) loss for the global restriction and local (pixel-level) restriction. Different from the standard IoU loss, the weighted IoU loss increases the weights of hard pixels to highlight their importance. In addition, compared with the standard BCE loss, L_{BCE}^W pays more attention to hard pixels rather than assigning all pixels equal weights. Here, we adopt deep supervision for the three side-outputs $(i.e., C_3, C_4, and C_5)$ and the global map C_6. Each map is up-sampled (e.g., C_3^{up}) to the same size as the ground-truth map G. Thus, the total loss for the proposed SINet can be formulated as:

$L_{total} = (L(C_6^{up}, G) + \Sigma_{i=3}^{i=5} L(C_i^{up}, G)$.

9.2 Using Drone Data

We are given images captured from drone where we have images of 5 different spectrum[3] taken by different drones of the same environment. As there will be some difference in the orientation of drones, the 5 images obtained are also having some difference in their orientation. Our model is based on 3-channelled input so we aim to convert the 5-channel image of our drone data into 3-channelled image.For our purpose we first need to align our 5 images so that their is no noise when we combine the images. For that we use Image Registration Technique. After aligning the images we need to convert the 5-channelled aligned images to 3-channelled image. Our approach here is using PCA for it as PCA is unsupervised feature extraction technique which is simple to implement and give effective and efficient results. Hence, we use PCA to convert 5-featured data to 3-featured data.

9.2.1 Image Registration

[12] Image registration is a technique used to align two or more images of the same scene taken from different viewpoints or at different times. The goal is to find a transformation that maps the pixels of one image to the corresponding pixels of the other image(s). This technique has numerous applications in computer vision, medical imaging, remote sensing, and many other fields.

There are several approaches to image registration, including feature-based methods, intensity-based methods, and hybrid methods.

Feature-based methods rely on identifying and matching distinctive features such as corners, edges, or blobs in the images. The transformation is then computed based on the matched features. This approach is robust to changes in lighting, contrast, and occlusion, but may fail when there are not enough distinctive features in the images.

Intensity-based methods use the intensity values of the images to compute the transformation. These methods optimize a similarity metric that measures the similarity between the images, such as mutual information or correlation. This approach is more general than feature-based methods but may be sensitive to changes in lighting and contrast.

Hybrid methods combine the advantages of feature-based and intensity-based methods. They first identify a set of features in the images, and then use the intensity values around the features to compute the transformation. This approach is more robust than either of the other methods alone.

In general, image registration is a challenging problem, especially when the images are noisy, contain artifacts, or have significant differences in scale, rotation, or perspective. However, with the appropriate techniques and algorithms, image registration can be a powerful tool for many applications.

9.2.2 Principle Component Analysis(PCA)

Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of a dataset by identifying the most significant features, known as principal components, that capture the majority of the variation in the data. This technique is commonly used in data preprocessing and feature extraction in machine learning, computer vision, and other fields.

The basic idea of PCA is to transform a high-dimensional dataset into a lower-dimensional space by projecting the data onto a new set of orthogonal axes that capture the maximum amount of variation in the data. The first principal component is the axis that captures the most variance in the data, and each subsequent principal component captures the remaining variance in decreasing order.

To compute the principal components, PCA finds the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors represent the directions of maximum variance in the data, while the eigenvalues represent the amount of variance captured by each eigenvector.

PCA can be used for various purposes, including data visualization, data compression, and feature extraction. In data visualization, PCA can be used to reduce the dimensionality of a high-dimensional dataset to two or three dimensions, allowing the data to be visualized in a scatter plot. In data compression, PCA can be used to reduce the size of the dataset while retaining most of the information. In feature extraction, PCA can be used to identify the most important features in the data and reduce the dimensionality of the feature space.

However, it is important to note that PCA assumes that the data is linearly related and that the principal components are uncorrelated. Therefore, it may not be suitable for all types of data, and other techniques such as nonlinear dimensionality reduction may be more appropriate in certain cases.

10 Flowcharts

10.1 Image Registration

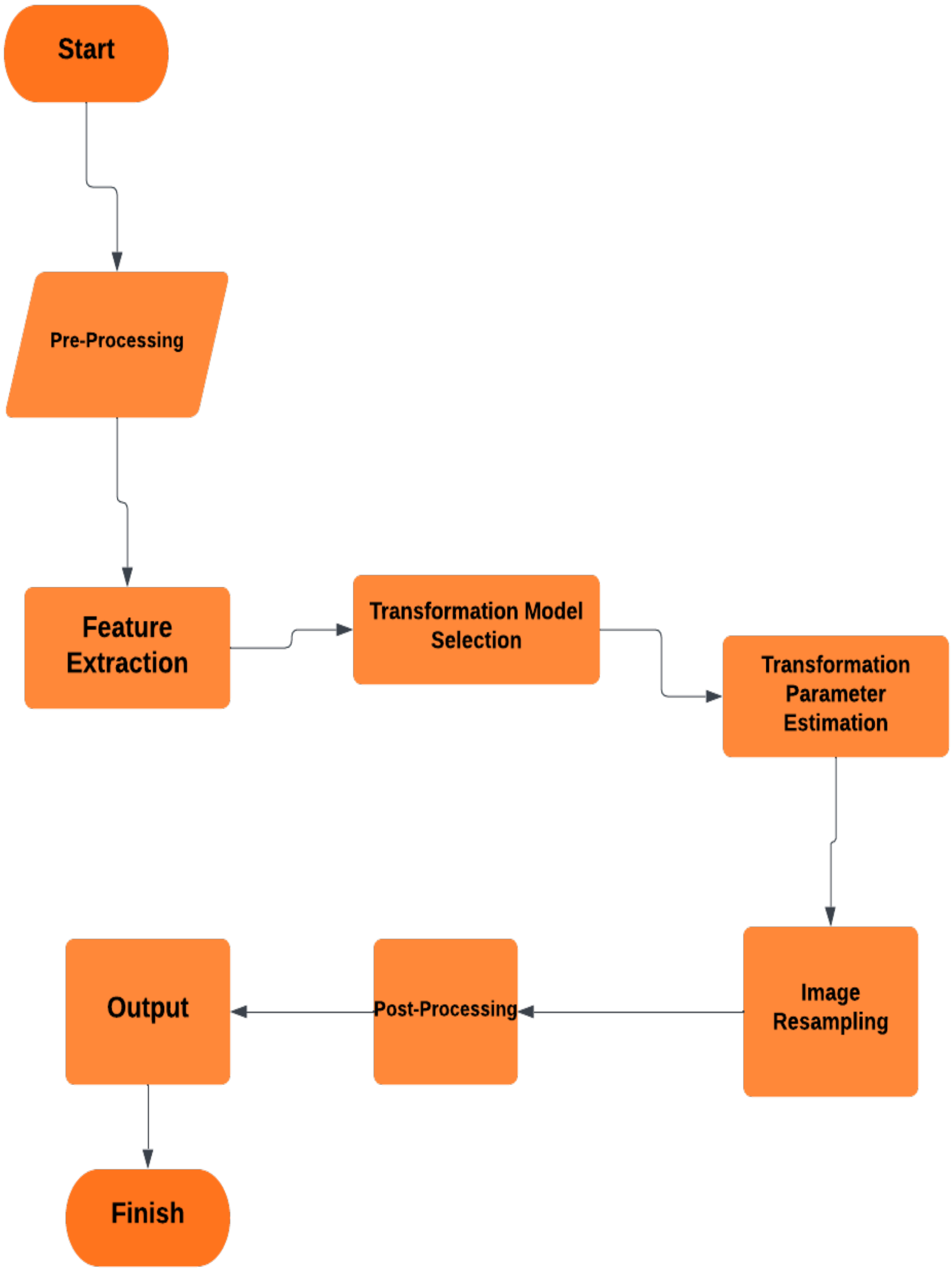

Figure 4: Flowchart for Image Registration

10.2 PCA

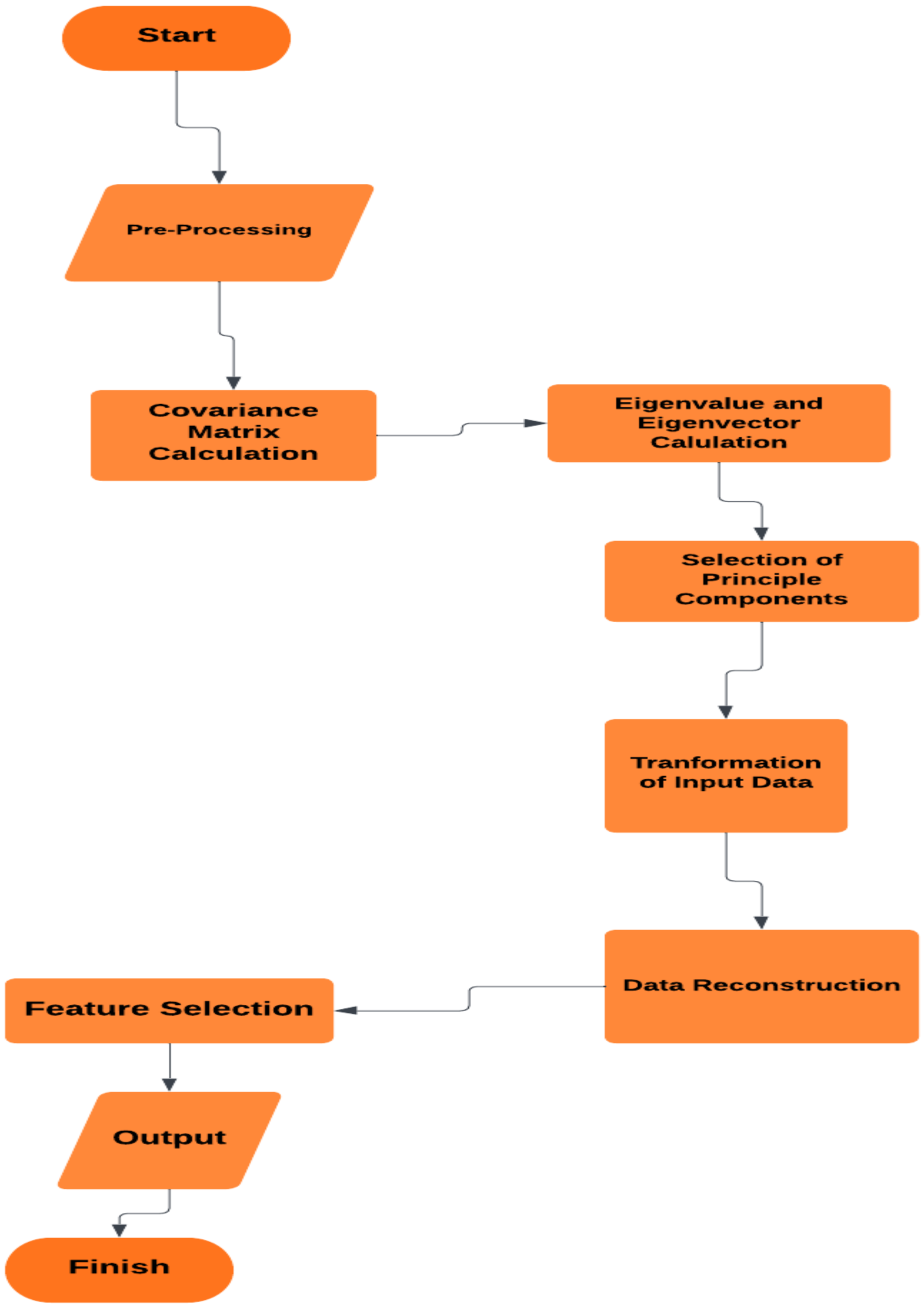

Figure 5: Flowchart for PCA

10.3 Data Augmentation

[1]

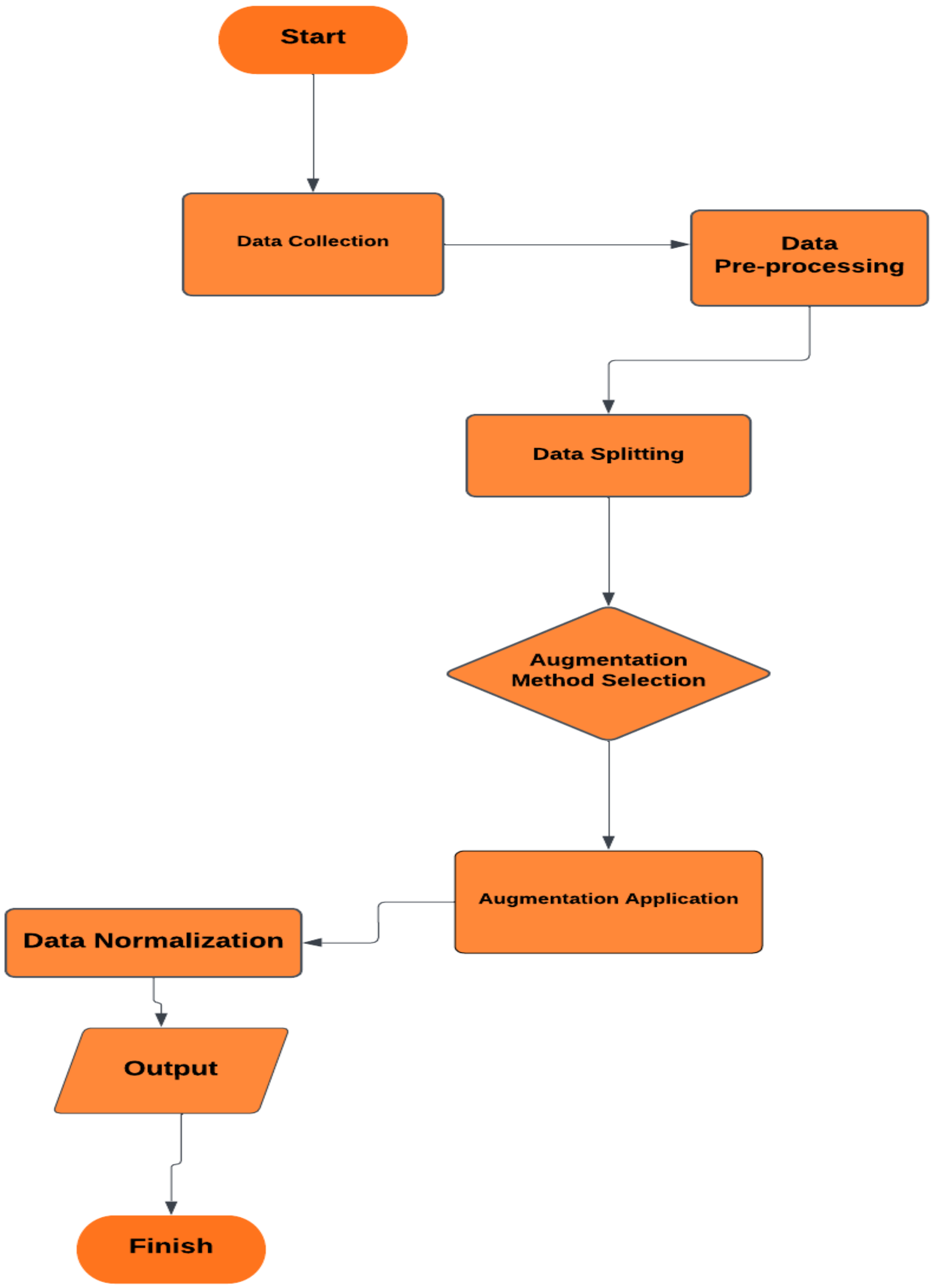

Figure 6: Flowchart for Data Augmentation

10.4 Camouflage Object Detection (COD)

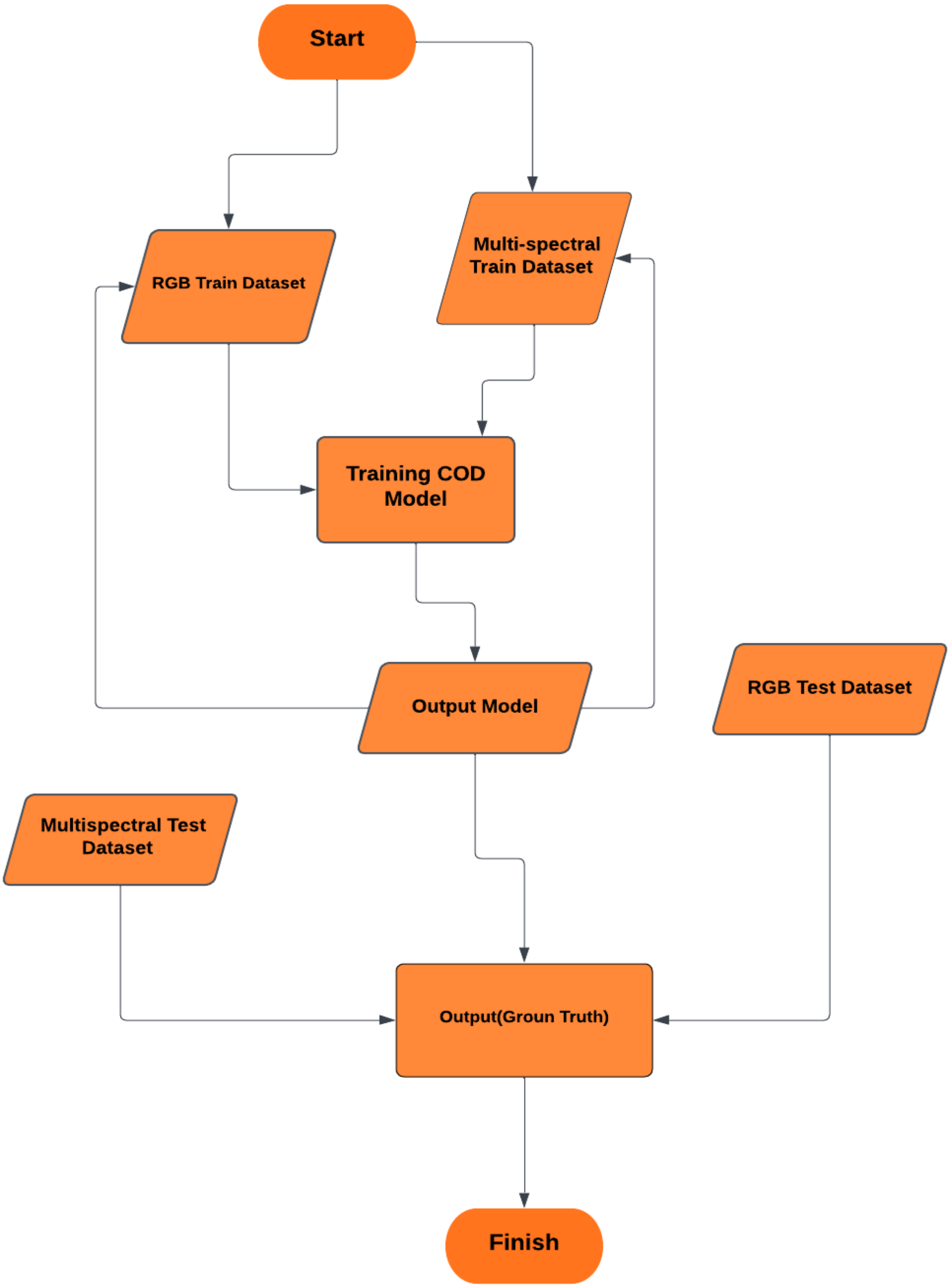

Figure 7: Flowchart for Camouflage Object Detection

11 Specification Table

Parameter	**Value**
Learning Rate	0.0001
Decay Rate	0.1
Epoch Decay Rate	30
Number of Parallel Workers	12
Batch Size	36
No. of Epochs	42
Size of Training Image	352
Device	GPU (CUDA based)

Table 1: Specification Table

---

12 Experimental Setup

12.1 Evaluation Metrics

Mean absolute error (MAE) is widely used in SOD tasks. We adopt the MAE (M) metric to assess the pixel-level accuracy between a predicted map and ground-truth. However, while useful for assessing the presence and amount of error, the MAE metric is not able to determine where the error occurs.

Since concealed objects often contain complex shapes, COD also requires a metric that can judge structural similarity. We therefore utilize the S-measure (S_a) as our structural similarity evaluation metric. Finally, recent studies have suggested that the weighted F-measure can provide more reliable evaluation results than the traditional F-measure. Thus, we further consider this as an alternative metric for COD.

12.2 Training/Testing Protocols

We evaluate the models on the whole CHAMELEON[16] dataset and the test sets of CAMO and

COD10K. For multispectral data we use Drone Image Dataset and split it into training and test set.

13 Results

Here are the results obtained in camouflage object detection of RGB images.

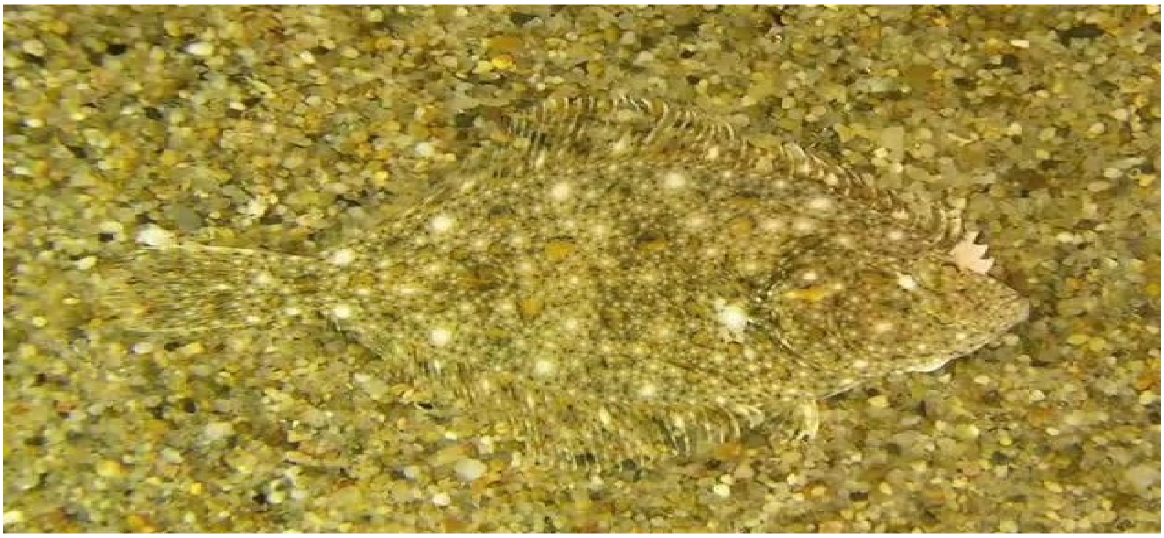

Figure 8: Camouflaged RGB image of fish

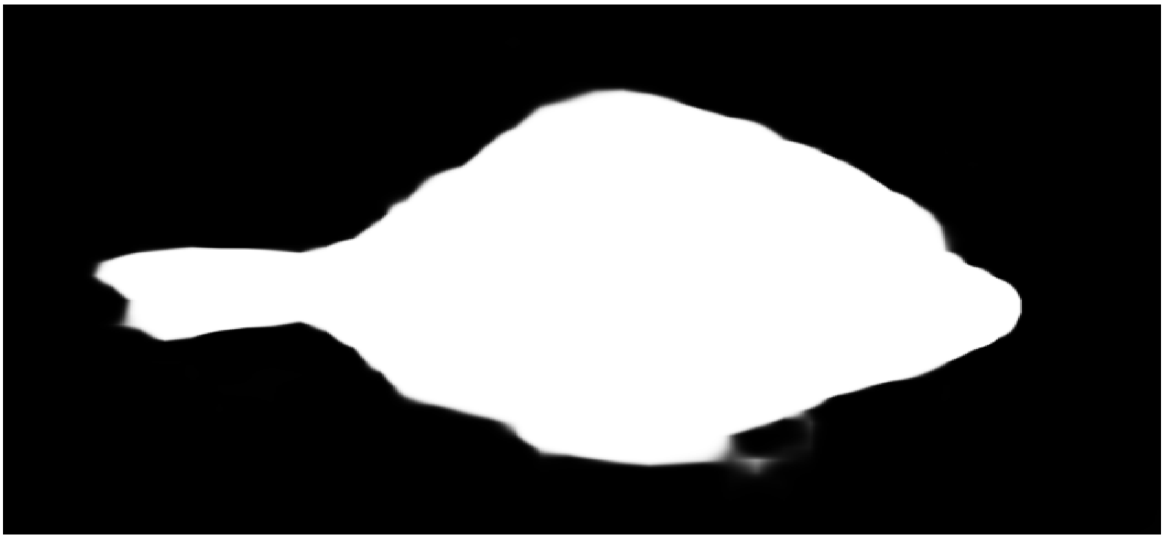

Figure 9: GT for the above RGB image

Figure 10: Camouflaged RGB Image

Figure 11: Grayscale output image

Below we present multi-spectral image consisting of 5-channelled image which is converted to 3-channelled image using feature extraction and data fusion.

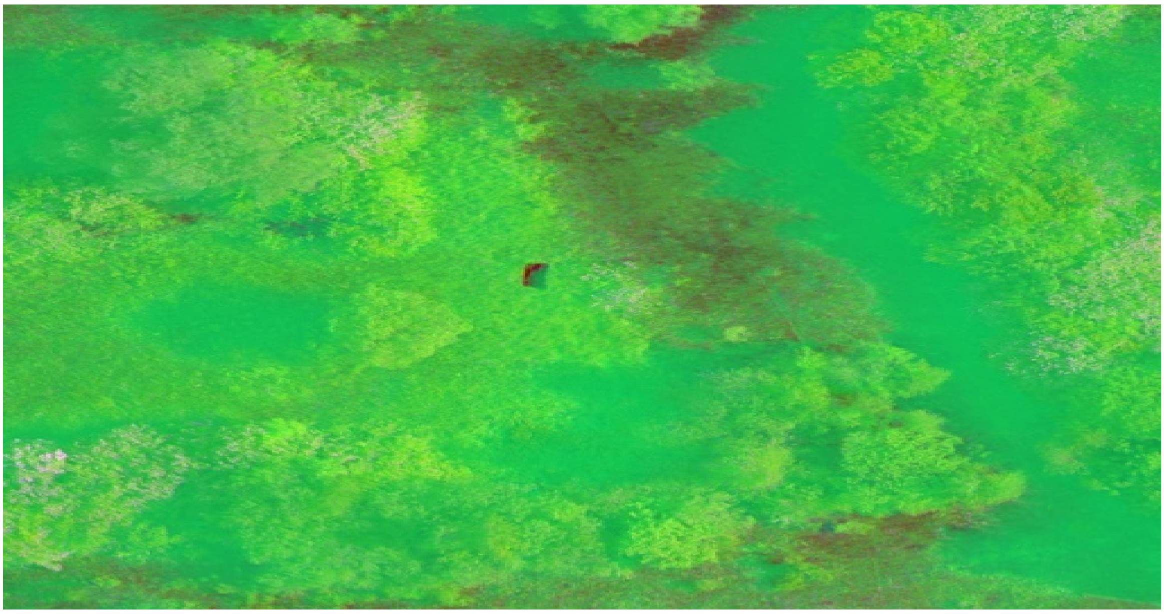

Figure 12: MultiSpectral Image having a box like object at the center

Figure 13: Grayscale Output Image

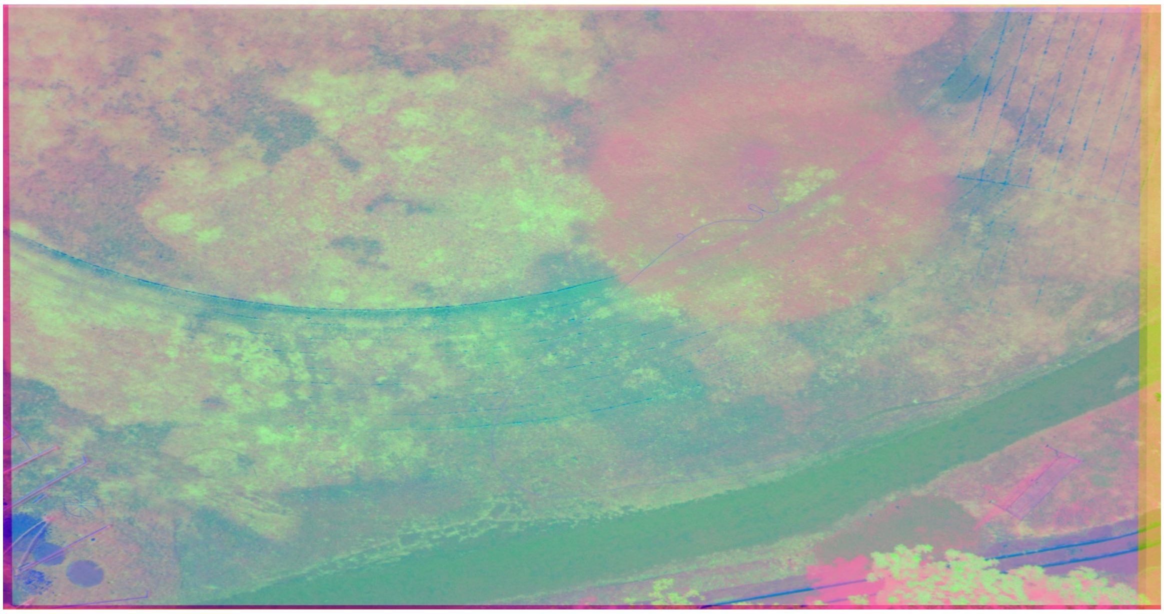

Figure 14: MultiSpectral data obtained using drone imaging

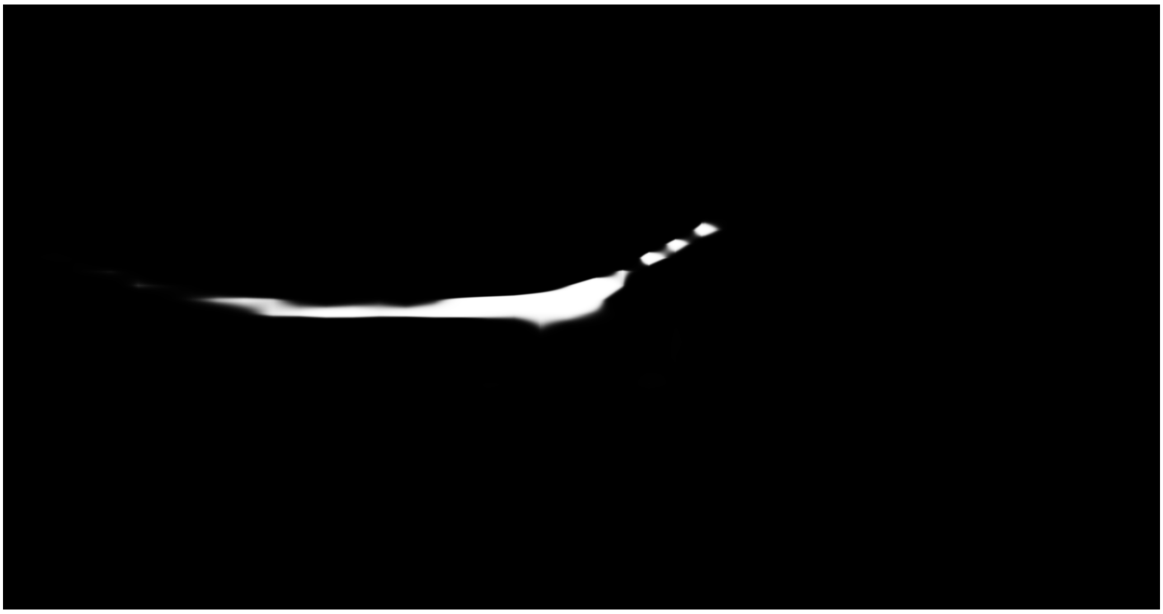

Figure 15: Grayscale Output Image

14 Applications[15]

14.1 Medicine

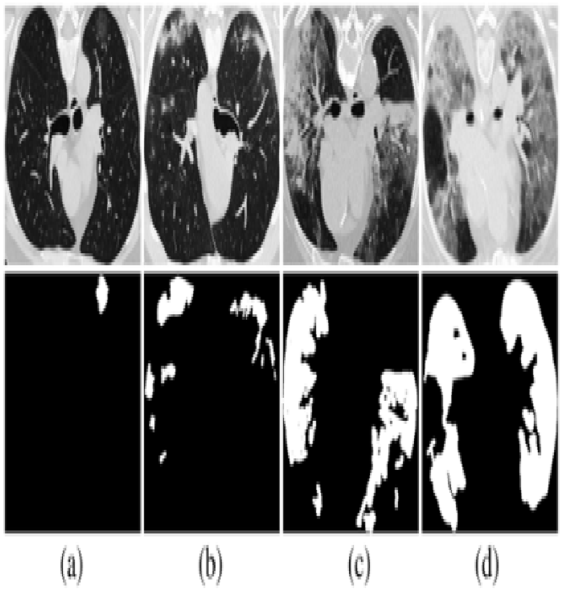

Figure 16: Lung infection segmentation

Concealed Object Detection can be used in the lung infection segmentation task in the medical field. Recently, COVID-19 has been of particular concern, and resulted in a global pandemic. An AI system equipped with a COVID-19 lung infection segmentation model would be helpful in the early screening of COVID-19.

14.2 Manufacturing

In industrial manufacturing, products (e.g., wood, textile, and magnetic tile) of poor quality will inevitably lead to adverse effects on the economy. The surface defects are challenging, with different factors including low contrast, ambiguous boundaries and so on. Since traditional surface defect detection systems mainly rely on humans, major issues are highly subjective and time-consuming to identify. Thus, designing an automatic recognition system based on AI is essential to increase productivity.

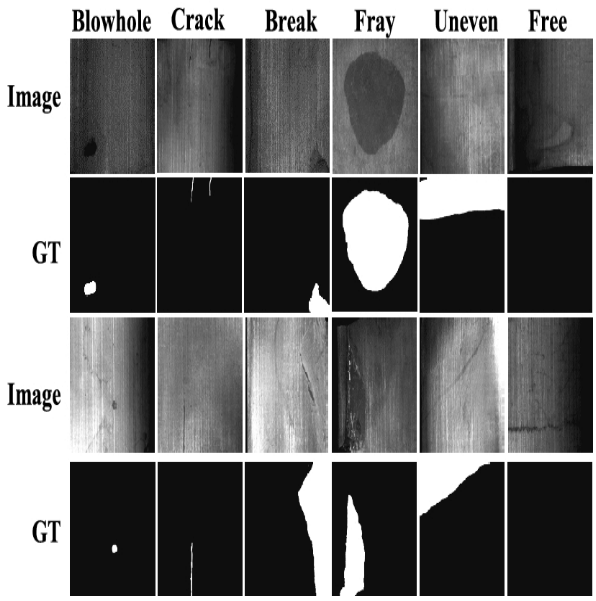

Figure 17: Surface Defect Detection

14.3 Agriculture

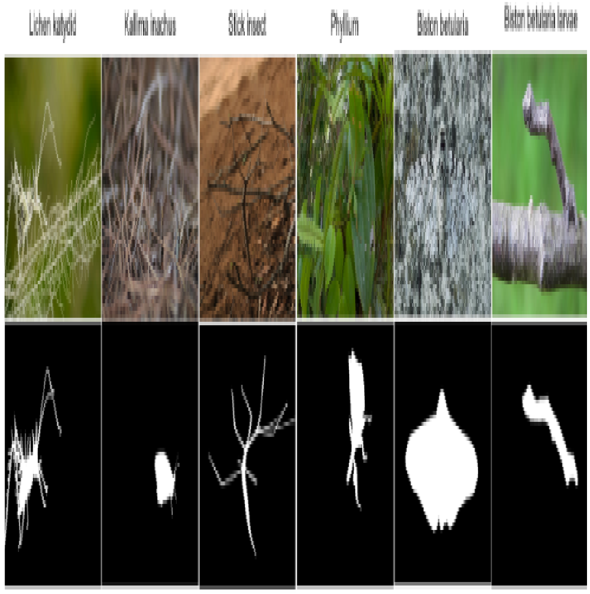

Figure 18: Pest Detection

Since the beginning of 2020, desert locust plagues have spread across multiple regions of the world, including Africa and South Asia. These swarms of locusts have caused significant damage to fields and crops, resulting in substantial economic losses and creating the potential for famine due to food shortages. To address this issue, governments could implement AI-based monitoring techniques, which would offer a more sustainable means of controlling the locust populations. However, obtaining relevant insect data for the development of accurate models is a significant challenge that requires a thorough understanding of insect biology.

14.4 Art

Camouflage detection is a useful tool in the field of art that allows experts to gain a deeper understanding of artwork. In the case of historical art, camouflaged elements may have been intentionally hidden by artists to convey hidden messages or symbols. By using specialized techniques such as infrared reflectography, X-ray imaging, or ultraviolet light, art historians and conservators can uncover these hidden elements and gain a deeper understanding of the artwork and the artist's intentions.

14.5 Military

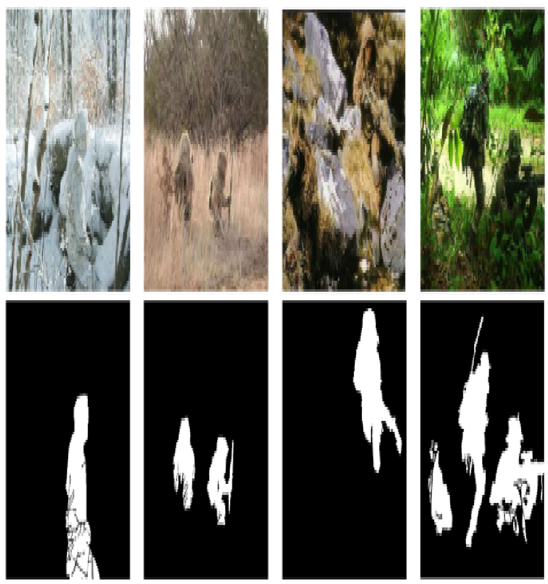

Figure 19: Camouflaged military personnels in different environments

Camouflage detection models can be extremely useful in military operations as they can help to identify enemy targets that may be hidden or difficult to spot using traditional visual methods. These models use advanced algorithms and machine learning techniques to analyze images and identify patterns that may indicate the presence of camouflage. One of the primary uses of camouflage detection models in the military is for reconnaissance and surveillance missions. By analyzing satellite or drone images, these models can identify hidden enemy positions, camouflage nets, and other structures that may be used to conceal military equipment or personnel. This information can then be used to plan strategic strikes or to develop countermeasures to neutralize the enemy's camouflage.

Camouflage detection models can also be used to enhance the effectiveness of soldiers on the ground. For example, a soldier equipped with a helmet-mounted camera and a portable device that runs a real-time camouflage detection model could quickly identify hidden threats and provide valuable information to his or her squad.

14.6 Daily Life

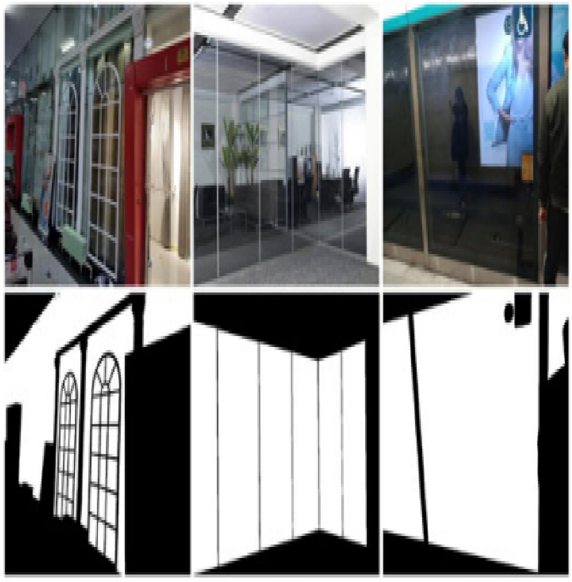

Figure 20: Pest Detection

Transparent objects, such as glass products, are commonplace in our daily life. These objects/things, including doors and walls, inherit the appearance of their background, making them unnoticeable. As a sub-task of COD, transparent object detection and transparent object tracking have shown promise.

15 Conclusion

In our study, we conducted an extensive investigation into object detection in camouflaged environments. We utilized both RGB and multi-spectral datasets to evaluate the effectiveness of different object detection techniques. In addition, we created our own dataset using drone imagery to simulate real-world scenarios and enhance the accuracy of our findings.

To combine the data from different spectral bands, we performed data fusion by aligning the images and using PCA to extract a 3-channelled image. This allowed us to combine the different spectral information into a single image, making it easier to detect objects in camouflaged environments.

To accomplish these tasks, we utilized SINet as our model, which has been demonstrated to be highly competitive with other state-of-the-art technologies in this field. We evaluated the performance of our model on both the RGB and multi-spectral datasets and found that it demonstrated excellent camouflage detection capabilities.

However, despite the impressive results, our model still has certain limitations. This is because the field of object detection in camouflaged environments is still relatively new in the deep learning domain. As such, there is still much room for improvement, and we believe that future research and advancements will continue to enhance the effectiveness of object detection in camouflaged environments.

Overall, our study provides valuable insights into the challenges of object detection in camouflaged environments and the potential for deep learning techniques to overcome these challenges. The techniques we employed in our study can serve as a foundation for future research in this field, with the ultimate goal of improving object detection in real-world applications.

16 Limiations

While our model demonstrated impressive results in detecting objects in camouflaged environments, there are still several limitations to our approach. One major limitation is that the model heavily relies on the quality and diversity of the training data. If the training dataset is limited or biased, the model may not generalize well to new and unseen scenarios. Moreover, our approach of combining spectral information through PCA may not be optimal for all scenarios, and more advanced fusion techniques may be required.

Another limitation is that our model may struggle with detecting camouflaged objects that closely resemble the background, such as objects with similar color or texture patterns. Additionally, our model may not be able to detect objects that are partially occluded or partially obscured by other objects.

Lastly, our approach of using a single model may not be suitable for all types of camouflaged object detection scenarios. For example, detecting small and fast-moving objects may require a different type of model with higher temporal resolution.

Overall, these limitations indicate that there is still much room for improvement in the field of camouflaged object detection, and further research is required to address these challenges.

17 Future Scope

The field of camouflaged object detection has great potential for future advancements, and there are several areas that could be explored to improve the effectiveness of detection models. Few of them are listed below:

i) Concealed object detection under limited conditions: few/zero-shot learning, weakly supervised learning, unsupervised learning, self-supervised learning, limited training data, unseen object class, etc.

ii) Camouflaged Object Detection combined with other modalities: Text, Audio, Video[11], RGB-D, RGB-T, 3D,etc.

iii) Existing deep-based methods extract the features in a fully supervised manner from images annotated with object-level labels. However, the pixel-level annotations are usually manually marked by LabelMe or Adobe Photoshop tools with intensive professional interaction. Thus, it is essential to utilize weakly/semi (partially) annotated data for training in order to avoid heavy annotation costs.

iv) Existing concealed data is only based on static images or dynamic videos. However, concealed object detection in other modalities can be closely related in domains such as pest monitoring in the dark night, robotics, and artist design.

References

[1] Álvaro Acción, Francisco Argüello, and Dora B. Heras. A new multispectral data augmentation technique based on data imputation. *Remote Sensing*, 13(23), 2021.

[2] Mohammed Alom. Adam optimization algorithm. 06 2021.

[3] Jens Bleiholder and Felix Naumann. Data fusion. *ACM Comput. Surv.*, 41, 12 2008.

[4] Giuseppe Carleo and Matthias Troyer. Machine learning and the physical sciences. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2102):20150202, 2017.

[5] AI Stack Exchange. What is a Fully Convolutional Network? https://ai.stackexchange.com/questions/21810/what-is-a-fully-convolution-network, 2021. Accessed: May 5, 2023.

[6] Marek Gagolewski. *Data Fusion: Theory, Methods, and Applications*. 12 2015.

[7] Saishruthi Swaminathan Heera. Understanding binary cross-entropy / log loss: A visual explanation. *Towards Data Science*, 2019.

[8] Sylvain Jay, Baret Frederic, Dan Dutartre, Ghislain Malatesta, Stéphanie Héno, Alexis Comar, Marie Weiss, and Fabienne Maupas. Exploiting the centimeter resolution of uav multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sensing of Environment*, 231, 09 2018.

[9] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu. A survey of deep learning-based object detection. *IEEE Access*, PP:1–1, 09 2019.

[10] Yan Jin, Yibiao Hu, Zhiwei Jiang, and Qiufu Zheng. Polyp segmentation with convolutional mlp. *The Visual Computer*, 08 2022.

[11] Xiaoyi Ma, Jun Xie, and Tong Zhang. Structured curriculum learning for neural machine translation. *OpenReview*, 6:1–8, 2022.

[12] Sorayya Malek, Mogeeb Mosleh, Sarinder K. Dhillon, and Pozi Milow. Bioimage informatics. In Shoba Ranganathan, Michael Gribskov, Kenta Nakai, and Christian Schönbach, editors, *Encyclopedia of Bioinformatics and Computational Biology*, pages 993–1010. Academic Press, Oxford, 2019.

[13] Edmund Optics. Hyperspectral and multispectral imaging. https://www.edmundoptics.eu/

knowledge-center/application-notes/imaging/hyperspectral-and-multispectral-imaging/, Accessed on May 5, 2023.

[14] Hyojin Park, Lars Sjosund, Youngjoon Yoo, Nicolas Monet, Jihwan Bang, and Nojun Kwak. Sinet: Extreme lightweight portrait segmentation networks with spatial squeeze modules and information blocking decoder. pages 2055–2063, 03 2020.

[15] Ch Pulla Rao, A Guruva Reddy, and C. B Rama Rao. Camouflaged object detection for machine vision applications. *International Journal of Speech Technology*, 23(2):327–335, 2020.

[16] Papers with Code. Camo: A Multi-View Dataset for Vehicle Re-Identification. https://paperswithcode.com/dataset/camo, 2021. Accessed: May 5, 2023.

[17] Lei Zhao, Mingcheng Zhang, Hongwei Ding, and Xiaohui Cui. Mff-net: Deepfake detection network based on multi-feature fusion. *Entropy*, 23:1692, 12 2021.

[18] Helu Zhou, Aitong Ma, Yifeng Niu, and Zhaowei Ma. Small-object detection for uav-based images using a distance metric method. *Drones*, 6(10), 2022.