El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

Deep Learning Models Integrating Attention Mechanisms For Military Camouflaged Object Detection

Nilgün Şengöz[1a], Gül Karaman[1b], Mert Samet Çeliker[1c], Nazmi Yücel Çan[1d]
[1]Department of Information Systems and Technologies, Mehmet Akif Ersoy University, Gölhisar School of Applied Sciences, Burdur, 15030, Türkiye

nilgunsengoz@mehmetakif.edu.tr

DOI : 10.31202/ecjse.1747013
Received: 20.07.2025 Accepted: 10.03.2026
How to cite this article:
Şengöz N., Karaman G., Çeliker M. S., Çan N. Y., "Deep Learning Models Integrating Attention Mechanisms for Military Camouflaged Object Detection", El-Cezeri Journal of Science and Engineering, Vol: 13, Iss: 2, (2026), pp.(146-160).
ORCID: [a]0000-0001-5651-8173, [b]0009-0004-0443-2895, [c]0009-0005-3327-5064, [d]0009-0009-5885-7406.

Abstract: The Adaptive Camouflage Dataset (ACD1K) comprises meticulously curated, high-quality images designed to facilitate research in camouflage detection and segmentation. This dataset is divided into training, validation, and testing subsets, enabling comprehensive evaluation of deep learning models. We employed Attention-ResUNet and Attention-ResUNet++ models, built upon the ResNet-50 architecture and enhanced with attention mechanisms, to ensure robust feature extraction. Performance evaluation was conducted using common metrics such as accuracy, precision, recall, F1-score, and intersection over union. The Attention-ResUNet, with CLAHE preprocessing, Adamax optimizer, 1e-5 learning rate, and 0.2 dropout rate, is recommended for segmentation-focused scenarios with an accuracy 96.88% and an intersection over union of 92.01%. Under similar experimental conditions, the Attention-ResUNet++ model, using the Adam optimization algorithm, is recommended for classification-focused scenarios with 98.32% accuracy and 82.09% intersection over union (IoU). These results demonstrate that model selection should be determined by application requirements. These findings underscore the efficacy and versatility of CNN-based architectures for detecting camouflaged objects in visually complex environments.

Keywords: Attention Mechanisms, Deep Learning, Image Segmentation, Object Detection, Adaptive Camouflage Dataset

1 Introduction

Camouflaged object detection is a computer vision problem that aims to identify objects integrated with natural environments by distinguishing their images from the background, which is of paramount importance, particularly in defense applications. Achieving high-accuracy detection has become critical in environments where targets are disguised or indistinguishable, rendering traditional methods limited. Today's military operations are becoming increasingly technology-driven, and the need for advanced analysis methods, such as automatic detection of concealed objects, is growing. Accurate detection of camouflaged objects minimizes manpower losses by enabling early intervention for surveillance systems, unmanned aerial vehicles, and autonomous defense systems. In this context, camouflaged object detection in modern warfare technologies is not only a tool, but also a technological requirement at the heart of the defense industry [1-2].

However, camouflaged object detection in a military context poses significant challenges due to the use of camouflage. Camouflage reduces the probability of target detection and is a concealment mechanism that allows objects to blend into their environment [3]. This limits the detection capabilities of traditional object detection methods. Especially in complex terrain environments (forest, desert, urban area) and variable weather/light environments, the effectiveness of camouflage increases, making it difficult to detect camouflaged objects [4].

At this point, advanced technologies such as artificial intelligence, machine learning, and deep learning come into play to solve the problem of camouflaged object detection. These technologies, integrated with image processing, aim to achieve high accuracy in pattern recognition and detection in cases where classical methods are insufficient thanks to their data-driven learning capabilities [5]. Given that access to military-domain imagery is often restricted due to security and privacy concerns, this can directly affect the performance of CNN-based models [6]. Therefore, segmentation-oriented architectures such as U-Net and U-Net++, which can offer high performance even under limited data conditions, have gained importance [7]. Thanks to their ability to capture even fine details in images, these models have provided effective solutions, especially in challenging conditions such as camouflaged object detection [3-6]. In addition, the integration of attention mechanisms is predicted to significantly improve segmentation performance by focusing not only on the general features of the image but also on the most critical regions of the image [1-3]. Similarly, recent deep learning-based detection and segmentation studies on aerial imagery have demonstrated the strong potential of CNN architectures for complex UAV-based visual detection tasks [8].

Despite recent advances, most existing studies focus either on improving segmentation accuracy or boundary precision, without explicitly analyzing the trade-off between these objectives (segmentation accuracy vs. intersection over union (IoU)).

In addition, the combined effects of attention mechanisms, ResNet-50 backbones, and CLAHE-based contrast enhancement remain insufficiently investigated in military camouflage detection scenarios.

In this study, this trade-off is systematically analyzed through a comparative evaluation of Attention-ResUNet and Attention-ResUNet++ architectures with a ResNet-50 backbone under limited-data military conditions. The proposed approach distinguishes model configurations optimized for high accuracy and high IoU performance, providing a clear contribution beyond existing works.

Finally, this study recommends Attention-ResUNet++ for applications requiring high classification accuracy, while Attention-ResUNet is recommended for scenarios where precise segmentation boundaries and high IoU are critical.

1.1 Research Questions

This study aims to address the following research questions:

RQ1: Can Attention-ResUNet and Attention-ResUNet++ architectures improve camouflaged object segmentation performance in complex military environments?

RQ2: How does the integration of a ResNet-50 backbone affect feature extraction robustness under limited data conditions?

RQ3: What is the trade-off between segmentation accuracy and boundary delineation (IoU) in the proposed models?

RQ4: Does CLAHE-based preprocessing contribute to performance improvement in camouflage detection tasks?

1.2 Contributions

The main contributions of this study are as follows:

We propose two attention-enhanced ResNet50-based U-Net architectures for camouflaged object detection and segmentation.

We present a systematic analysis of the trade-off between segmentation accuracy and boundary delineation (IoU), offering a clear model-selection guideline (High-IoU vs. High-Accuracy).

We evaluate the combined effect of attention mechanisms, CLAHE preprocessing, and optimizer selection under limited-data military conditions on the ACD1K dataset.

This study was conducted without external funding. The limitations of the proposed approach and future research directions are discussed in the concluding section.

2. LITERATURE SUMMARY

In military domains, camouflage plays a critical role in concealing military targets. In this context, the fields of object detection and image segmentation are among the most active research topics in the computer vision literature. Modern approaches are largely based on deep learning models, and researchers have experimented with classification and prediction approaches for camouflaged object detection using different methods.

In one of the studies on camouflaged object detection and segmentation, a novel model called CoFiNet was proposed. By focusing on multi-scale feature fusion and extraction, this model aims to improve detailed segmentation performance even in challenging situations where object boundaries are ambiguous. The model has improved its capacity to capture both contextual and detailed information by using a combination of coarse and fine segmentation strategies, and the learning process is optimized with selective kernel structures and a double mask approach. CoFiNet was tested on four different datasets to demonstrate its effectiveness [9].

In another study, U-NET and VGG16 architectures were integrated with CBAM attention mechanism to detect camouflaged objects. In the experiments performed on CAMO and CAMO-COCO datasets, the learning process of the model was enhanced by using data augmentation techniques. The results show that the proposed method performs well on different datasets [10].

In another segmentation study using the U-Net architecture, a different dataset was used. The encoder-decoder structure of the model increased its ability to distinguish hidden objects in images by focusing on preserving fine details. The model achieved an average accuracy rate of 70.14% in the training and testing processes using different parts of the same dataset. It has shown that U-Net can work in challenging scenarios such as camouflage [11].

DMNet, one of the recent works, uses a diffusion-based method. This model learns to generate target object boxes by a process of noise reduction over random boxes.The performance of the model has been extensively tested with six different evaluation metrics and reported to be superior to competing models. In particular, a significant improvement in segmentation accuracy was achieved thanks to the position enhancement module that improves edge recall and the adaptive weighting approach between layers [12].

Finally, K. Karthiga et al. developed a new model based on YOLOv7, which provides a more specific approach for the detection of camouflaged targets. This model integrates GAM-ELAN (Global Attention Mechanism + Efficient Layer Aggregation Network) into the network structure for feature extraction. The features are integrated into the detection mechanism of YOLOv7 by combining them with FF-FPN (Feature Fusion-Feature Pyramid Network). The model was tested on the ACD1K and COD10K datasets and achieved 98.2% accuracy, 92.37% IOU and 95.4% accuracy, 90.4% IOU,

respectively. Metrics such as F-Measure, Precision, Recall, MAPE, and IoU were also used in the study [1].
The current work contributes to the field of image segmentation by focusing specifically on the accurate pixel-level discrimination of camouflaged objects. Inspired by segmentation and image processing papers in the literature, the camouflaged object detection problem is treated as a segmentation problem. Unlike the YOLO-based work of K. Karthiga et al., our current work aims to identify the boundaries of objects more clearly by adopting a mask-based segmentation approach [1-9-10-11-12].

ACD1K, the Adaptive Camouflage Dataset used in the current study, provides comprehensive coverage of objects hidden in various environments [13]. This dataset of high-quality and carefully curated images is designed to support research in the field of camouflage. The diversity of the dataset provides an ideal basis for testing the ability of the developed deep learning models to accurately detect different camouflage patterns. Using sophisticated deep learning models developed for image segmentation, it is aimed to accurately discriminate camouflage patterns in the ACD1K dataset on a pixel-by-pixel basis [14].

Table 1: Comparison of metric values of existing studies.

Publication	Methods	Dataset	Accuracy	IOU	Mape
K. Karthiga [1]	Proposed Yolov7	ACD1K	0.982	0.9237	0.018
K. Karthiga [1]	Proposed Yolov7	COD10K	0.954	0.904	0.023
Deep Learning Models Integrating Attention Mechanisms for Military Camouflaged Object Detection	Attention-ResUNet (proposed)	ACD1K	0.9704	0.9401	-
Deep Learning Models Integrating Attention Mechanisms for Military Camouflaged Object Detection	Attention-ResUNet++ (proposed)	ACD1K	0.9832	0.8209	-

Accuracy and Intersection over Union (IOU) metrics were used to compare the metric results of our current study and the study of K. Karthiga et al. in Table 1. Accuracy indicates the rate at which all pixels are correctly classified, while IOU is the ratio of the intersection area between the predicted mask and the actual mask to the area of the union and is an important metric for evaluating segmentation performance. While the MAPE (Mean Absolute Percentage Error) metric used in the YOLOv7-based CAMOUFLAGE-Net study by K. Karthiga et al. produces meaningful results in scenarios with predicted values, such as object detection, this metric gives meaningless and unreliable results since pixel-based classification is performed in our mask-based segmentation model [1-15-16].

In the current study, using the ResNet50 model and Attention-ResUNet++ architecture, 98.32% accuracy and 82.09% IOU values are obtained for the ACD1K dataset. When the ResNet50 model and Attention-ResUNet are used, 97.04% accuracy and 94.01% IOU values are obtained. It is recommended to use the ResNet50 model and Attention-ResUNet++ if the accuracy value is aimed to be higher, and the ResNet50 model and Attention-ResUNet method if the IOU value is aimed to be higher.
One of the strengths of our work is that we are able to identify the boundaries of objects more clearly thanks to our mask-based segmentation approach. In addition, the different hyperparameter optimizations we applied to the data with different hyperparameter optimizations increased the generalization capability of our model by preventing memorization in raw and masked data. The times in the training and testing phases are also given in seconds. These factors reveal the efficiency of the model. In addition, low prediction times are important for real-time applications. These factors make our work more powerful and unique compared to other approaches in the literature.

3. MATERIAL METHOD

In this study, deep learning-based semantic segmentation approaches are used to address the problem of camouflaged object detection, with the aim of accurately detecting objects hidden in complex natural environments at the pixel level. The methods are tested on a special dataset consisting of high-resolution RGB images and their corresponding accurately labeled mask (ground truth) maps. The dataset is composed of images taken in various terrain types (forest areas, desert surfaces, urban environments) that are diversified to represent different environmental conditions and include camouflaged soldiers. The mask files corresponding to each image show the pixel-level regions where camouflaged objects are located, providing a strong basis for training and evaluating segmentation algorithms.

The images were resized to 256×256 pixels to conform to the model requirements during the processing process, and then the pixel intensities were normalized to the range [0, 1]. To enhance overall model performance and mitigate overfitting, various data augmentation techniques were applied, various data augmentation techniques were applied to the training data. These techniques include random horizontal and vertical translation, rotation, zooming, contrast/brightness adjustments, and low-level geometric distortions. The aim was to make the model more robust to different positions, scales, and lighting conditions.

U-Net and U-Net++ architectures were preferred for the segmentation task, and the pre-trained ResNet-50 network was integrated as a backbone in order to provide powerful feature extraction in the encoder layers of these models. Furthermore, by incorporating attention mechanisms into the model architectures, it is aimed to focus on more critical regions in the images and increase segmentation performance. All models were built using the Python programming language with TensorFlow and Keras libraries. Auxiliary libraries such as OpenCV, NumPy and scikit-image were used for visual processing. During the training process, different optimizers (Adam, Adamax, and Adadelta), learning rate, number of epochs and dropout values were tested to determine the optimal hyperparameter combinations and the results were comprehensively evaluated with performance metrics such as accuracy, precision, recall, F1 score, and Intersection over Union (IoU).

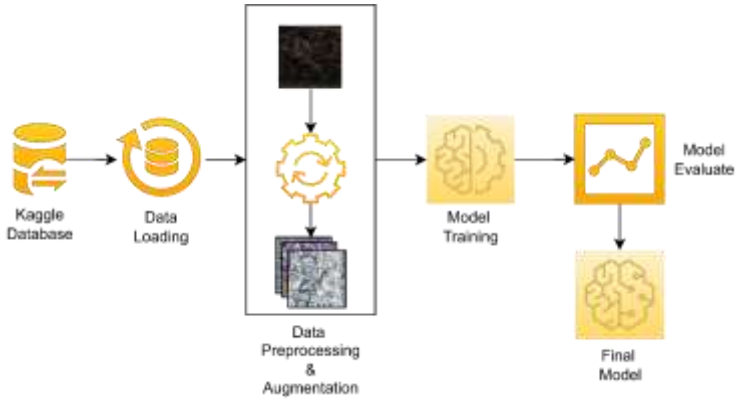


Figure 1: Segmentation model augmented with transfer learning model.

The model in Figure 1 schematically summarizes the end-to-end operation of a deep learning-based system developed for camouflaged object detection. The process starts with the collection of camouflaged soldier images obtained from the field and continues by ensuring data security by backing up these images. Subsequently, a preprocessing phase—including resizing, normalization, and noise reduction—is executed to prepare the images for model training. After the preprocessing, the data augmentation step aims to make the model more robust to different variations by subjecting the images to various transformation techniques. This process is critical to improve the generalization performance of the model. Then, the dataset is trained with deep learning architectures (U-Net and U-Net++) and the learning performance of the model is evaluated with metrics such as accuracy, precision, specificity, F1 score, AUC, and IOU.

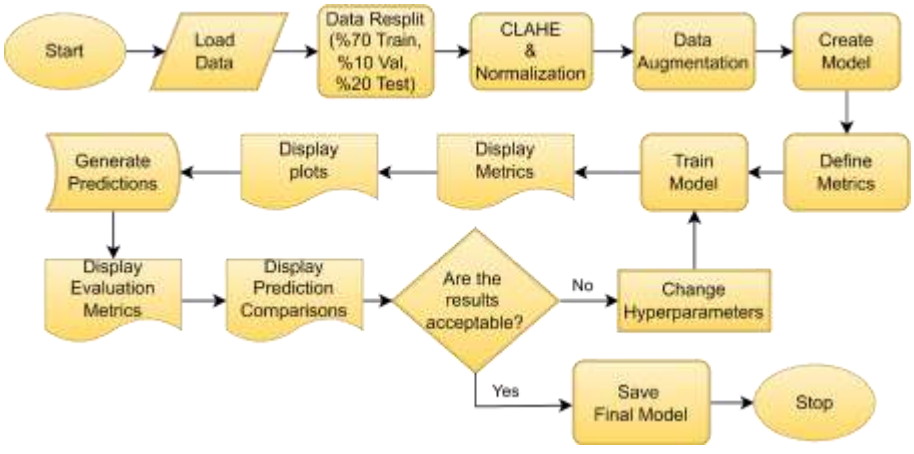


Figure 2: Flow diagram of the model used for segmentation.

The flow diagram in Figure 2 summarizes the flowchart of the deep learning-based classification system for camouflaged object detection. The process starts with uploading the data and then splitting it into 70% training, 10% validation and 20% testing, preprocessing with CLAHE and normalization, and data augmentation. Then the model is created, success metrics are defined and the model is trained. After training, metrics and graphs are displayed, predictions are generated on the test data, evaluation metrics and comparisons are presented based on these predictions. If the results are not acceptable, the process

returns to training by changing the hyperparameters, and if they are acceptable, the final model is saved and the process is terminated.

In this study, training and testing of the developed deep learning-based segmentation models were performed on the Kaggle platform. The working environment used is running on the Linux 6.6.56+ operating system and configured with Python version 3.11.11. The system is equipped with an x86_64 architecture processor with a total RAM capacity of 31.35 GB and 4 logical and 2 physical cores. NVIDIA Tesla P100-PCIE-16GB GPU was used as the graphics processing unit and configured with CUDA 12.4 and cuDNN 9.0.3 versions active. These hardware features enabled fast and efficient training of deep learning models. The codes were run on the Jupyter Notebook environment and the GPU acceleration was fully utilized throughout the training process.

3.1. Dataset

In this study, the dataset used for camouflaged object segmentation is called ACD1K and was used in the training and testing processes of deep learning-based image processing methods [13]. The dataset consists of high-resolution images in RGB format and corresponding pixel-level labeled mask images. In the training phase, a total of 748 camouflaged images and 748 masks corresponding to these images were used. In the testing phase, 330 images and 330 masks were used. In the images, the camouflaged objects are integrated into the natural environment and the masks represent the boundaries of these objects at the pixel level. All images and masks are in three-channel (RGB) format and resized to fit the model input. The dataset provides an ideal reference to test the effectiveness of object segmentation algorithms in challenging visual conditions (complex backgrounds, low contrast, similar color distributions).

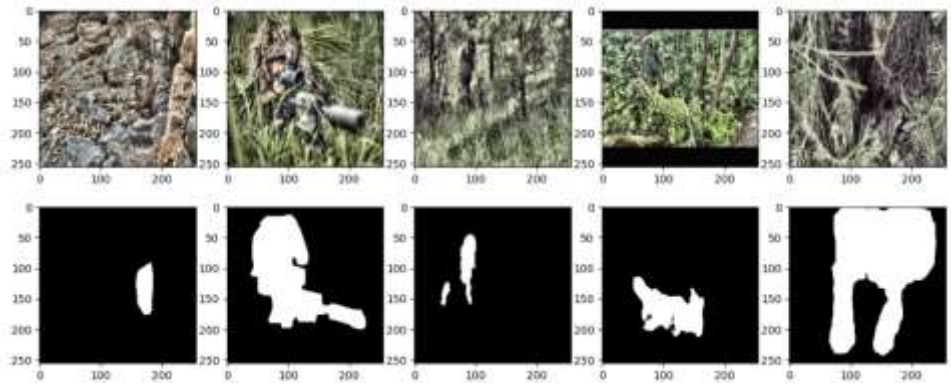


Figure 3: Sample images and their masks in the dataset.

Figure 3 shows sample images of the dataset we used in the current study.

Table 2: Number of images and masks in the dataset.

Training	Ground Truth	754
	Images	754
Validation	Ground Truth	108
	Images	108
Testing	Ground Truth	216
	Images	216

Table 2 shows the images in the dataset and the number of corresponding masks.

3.2. Data Augmentation

Data augmentation is the transformations and manipulations applied to the original dataset to improve the generalization ability of deep learning models. In this study, various data augmentation techniques were applied to the image and mask data used for camouflaged object detection. These operations aim to prevent overfitting by training the model on a wider range of data.

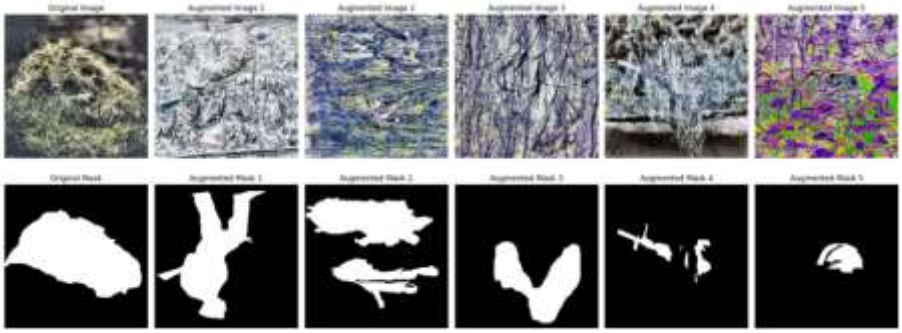


Figure 4: Example images and masks with data augmentation techniques applied.

The existing dataset used in Figure 4 has been augmented using some pre-processing steps. The augmented version of the existing dataset is also indicated. Data augmentation was performed with the Albumentations library [17]. The transformations applied are as follows:

The images in Figure 4 show the original images and masks before data augmentation techniques were applied, and the augmented samples obtained as a result of various transformations applied to these images. As can be seen from the image, a richer training data set was created by diversifying the limited amount of original data with certain transformations. This was done in order to increase the generalizability of the deep learning model to wider variations and reduce the risk of overfitting. Data augmentation was implemented using the Albumentations library. This library was chosen because it offers both flexible and high performance transformation functions. The first transformation applied is the RandomRotate90 transformation. This operation ensures that each image is randomly rotated 90 degrees with a probability of 80%. This transformation especially allows objects to be modeled from different angles. Thus, the model is able to recognize objects in different orientations, not only in a fixed perspective. This makes the model more robust to angular variations [17].

The second transformation, HorizontalFlip, involves inverting the images on the horizontal axis with a probability of 70%. This process makes it possible to recognize symmetrical structures from different directions. The left-right transformation of objects in the image helps the model to recognize the same object in different positions and orientations. Similarly, the VerticalFlip transformation involves inverting the images on the vertical axis with a 70% probability. This transformation both facilitates the model's learning of up-down variations of objects and introduces the model to different perspectives that can be encountered in real-world environments.

All these transformations were applied simultaneously to both the original images and the corresponding segmentation masks, as shown in Figure 4. Thus, the alignment between the data and the label is preserved and the model learns correctly. Such data augmentation steps significantly improve the performance of the model, especially in areas such as medical imaging, where there is a limited number of annotations.

The CLAHE method was also applied to the data as an extra. Contrast Limited Adaptive Histogram Equalization (CLAHE) is a contrast enhancement method that is frequently used in image processing, especially in medical images. Unlike traditional histogram equalization methods, CLAHE divides the image into small tiles and applies local histogram equalization to each region. This both enhances local contrast and improves the overall image quality. Furthermore, by limiting the contrast enhancement, sudden intensity changes and noise spikes are prevented, which is an important advantage, especially for low-contrast and noise-sensitive medical images [18]. The CLAHE method has been widely used in the healthcare field, especially in low-contrast medical images to enhance details. In this study, CLAHE is applied to overcome the low contrast problems caused by the similarity of camouflaged objects to the environmental background and to make object boundaries more distinct. In the experiments, it was observed that the application of CLAHE positively affects the performance of camouflaged object detection and improves the segmentation results. Recent studies have shown that CLAHE significantly improves model performance when used as a preprocessing step in deep learning-based classification and segmentation tasks. Especially in medical images, it has been observed that preprocessing with CLAHE enables artificial intelligence models to discriminate small details more successfully [19]. In another study, deep learning-based segmentation models were reported to provide significant improvements in accuracy and recall in MRI images with CLAHE [20]. Moreover, the proper tuning of the parameters of CLAHE (e.g., frame size and contrast boundary) is critical for the image quality. Data-driven methods such as Bayesian optimization are recommended for optimizing these parameters [21]. For all these reasons, CLAHE has become an effective image enhancement method that is frequently used in deep learning supported medical image analysis [22].

The masks were also processed in harmony with the images and updated correctly during these transformations. Thanks to the parameter that allows the mask to undergo the same processing as the raw data, the masks are processed in the same way as the images.

The data augmentation function enriches the original dataset by applying a certain number of augmentation operations. The function applies the defined transformations to each image and mask, creating an augmented dataset. In order to avoid possible mismatches between the images and the corresponding ground truth masks, the number of samples in the dataset is equalized.

The augmented data set was added to the training data set to increase the number of data. A total of 1508 images were obtained by augmenting the training data set, which initially contained 754 images.
Augmented dataset visualizations and masks allow the model to work with more diverse data during the training process, improving its overall performance. Using the preprocessed data, segmentation was performed with the Attention-ResUNet and Attention-ResUNet++ models, both incorporating a ResNet-50 backbone and attention mechanisms.

3.3. ResNet50

In the current study, a deep learning-based approach is developed on the camouflaged object detection problem, and ResNet-50 architecture is preferred as the backbone. ResNet-50 is a powerful network structure consisting of 50 layers that solves the vanishing gradient problem encountered in deep convolutional neural networks through residual connections [23]. The model consists of an input layer, four groups of residual blocks, and an output layer. The residual blocks consist of bottleneck blocks with a triple convolutional structure (1x1, 3x3, 1x1). This structure makes it possible to reduce computational cost and maintain learning efficiency while increasing network depth [23-24].

Residual connections allow the model to have deeper structures and allow gradients to propagate backwards more effectively during learning. Especially in models with a high number of layers, these connections are critical for stable and effective learning [23]. This structure of the ResNet architecture has inspired later models such as SE-ResNet [25] and ResNeXt [26], where residual connections are further optimized by channel weighting or grouped convolutions.



Figure 5: Architecture of the ResNet50 model used as backbone.

Figure 5 shows the architecture of the ResNet50 model used as backbone. The ResNet-50 architecture starts with an initial 7x7 convolution (64 filters, stride=2) followed by 3x3 maximum pooling. This is followed by four basic convolutional block groups Conv2_x (3 bottleneck blocks), Conv3_x (4 blocks), Conv4_x (6 blocks), and Conv5_x (3 blocks). In these blocks, the filter sizes are gradually increased so that the model can learn more complex features. The output layer includes global average pooling and fully connected layers, and finally softmax activation is used. This structure is designed to provide high performance for both classification and segmentation tasks [23].

In terms of transfer learning, ResNet-50 is a very common choice, and successful results have been obtained in different tasks using weights previously trained on ImageNet dataset [23-27]. It exhibits superior generalization capability especially in applications such as medical imaging, object recognition, and face recognition [25-28].

In the current study, the ACD1K dataset containing masking information is used for the camouflaged object detection problem. This dataset is suitable for challenging segmentation tasks because it requires detecting camouflaged objects. The ResNet-50 architecture is used as the backbone, with both U-Net++ and the attention mechanisms integrated into this model. This architecture is powerful in terms of depth and learning capacity and is supported by hyperparameter optimization to improve segmentation performance.

3.4. Attention-ResUNet

U-Net is a convolutional neural network (CNN) architecture developed by Olaf Ronneberger et al. in 2015 for biomedical image segmentation. This architecture stands out with its advantages such as the ability to work with limited labeled data, high accuracy segmentation and low computational cost. U-Net consists of two main parts: a narrowing path (encoder) that processes the input image and an expanding path (decoder) that produces the segmentation mask. Between both paths are skip connections that connect feature maps of the same resolution level. These connections improve segmentation accuracy by allowing low-level details to be combined with high-level semantic information.

Figure 6 visualizes the overall structure of the Attention-ResUNet architecture with ResNet-50 as the backbone. Nowadays, various enhancements have been made on the U-Net architecture. For example, U-Net++ aims to reduce the semantic gap by densifying jump connections and adding deep supervision [29]. U-Net v2 enriches multi-level feature maps using the Hadamard product and improves segmentation accuracy [30]. TransUNet utilizes both local and global context information by integrating Transformer modules into the U-Net architecture, which improves performance, especially in multi-organ segmentation tasks [31]. Similarly, the U-Netmer model provides high accuracy in segmentation tasks at different scales by combining U-Net and Transformer structures [32].

In the present paper, the classical U-Net algorithm is enhanced with the Attention gate mechanism.

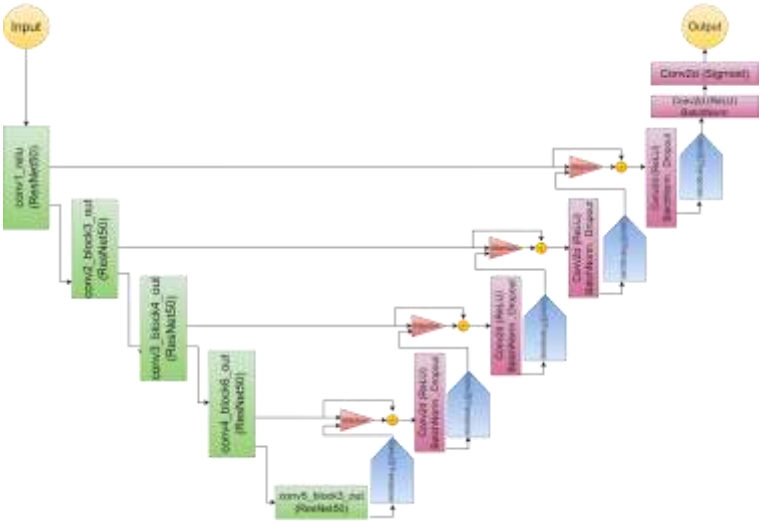


Figure 6: Attention-ResUNet model architecture.

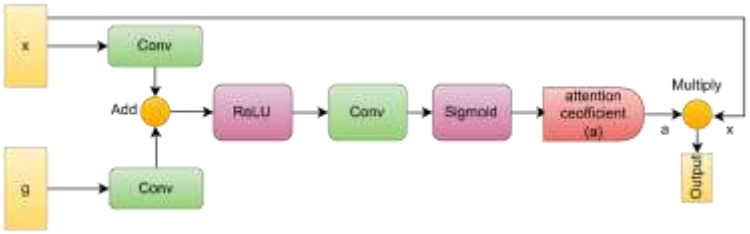


Figure 7: Architecture of the Attention Gate mechanism.

Figure 7 shows the Attention Gate mechanism. The Attention Gate (AG) mechanism shown in Figure 7 is a structure developed to direct attention-based information flow, especially in medical image segmentation. This structure allows the network to focus only on relevant regions, suppressing information from unimportant or irrelevant regions. This allows the model to perform more precise and accurate segmentation, especially in border regions.

There are two basic inputs to the AR mechanism: low-level feature map (x) from the encoder path and high-level semantic information (g) from the decoder path. First, these two inputs are dimensionally aligned by passing them through convolutional layers. These two components are then summed and combined and processed with a non-linear ReLU activation function. This intermediate result is then normalized between 0 and 1 using a second convolutional layer and a sigmoid activation function. These values are called attention coefficients (a). These coefficients are dot multiplied by the input feature (x) to give more weight to regions of interest while suppressing irrelevant regions.

The AG mechanism enables better capture of small lesions or boundaries, which are especially important in complex medical images. It has been shown in many studies that when integrated into encoder-decoder architectures such as U-Net, both precision and segmentation accuracy are improved [33]. Moreover, this mechanism contributes to the generalization capability of the model as it creates learnable attention masks without additional parameter overhead [34]. Especially with the widespread use of attention-based models, the use of structures such as AG in medical image processing is increasing.

U-Net, Attention Gate, and ResNet50 are advanced image segmentation architectures designed to both preserve low-level details and efficiently model high-level semantic information. It is based on the U-Net's encoder-decoder architecture, while the encoder part uses ResNet50 as a powerful feature extractor. Thus, highly representative features are extracted from images through deep convolutional layers. The residual connections of ResNet50 reduce gradient loss, enabling deep architectures to learn more stably and efficiently.

The Attention Gate (AG) mechanism in architecture, especially integrated into skip connections, ensures that only meaningful and target-oriented features are transferred to the decoder layer during the segmentation process. Each AG module compares the low-level details from the encoder with the more abstract information from the decoder and combines these two sources of information through attention coefficients, allowing only the activations of important regions to stand out. This allows the model to better discriminate target objects from environmental clutter (similar textures for the current study).

This architectural combination offers a great advantage in visual scenes with low contrast and complex backgrounds, where conventional methods such as camouflaged object detection struggle. Since camouflaged objects often have similar

colors, textures or edge structures to their surroundings, purely convolutional filters may be insufficient for such situations. The combination of ResNet50's strong semantic representation capability, U-Net's multilayer feature transfer and AG's localized recall results in significant performance gains, especially in the clear delineation of object boundaries and the capture of small details.

In the present work, this architecture is supported by various hyperparameter optimization techniques. Critical parameters such as learning rate, batch size, dropout rate and optimizer type were systematically screened to determine the best performing combinations. At the same time, by increasing the diversity of the training set with dataset-specific augmentation techniques, the risk of overfitting was reduced and the generalization capability of the model was strengthened. In this way, the obtained results showed high success in terms of metrics such as accuracy, F1 score, recall, specificity and Intersection over Union (IoU). The Attention-ResUNet architecture has been optimized to successfully perform object segmentation even in scenes with high camouflage density.

3.5. Attention-ResUNet++

In the remainder of the present paper, hyper-parameter optimizations are also applied to the U-Net++ model augmented with the Attention mechanism.

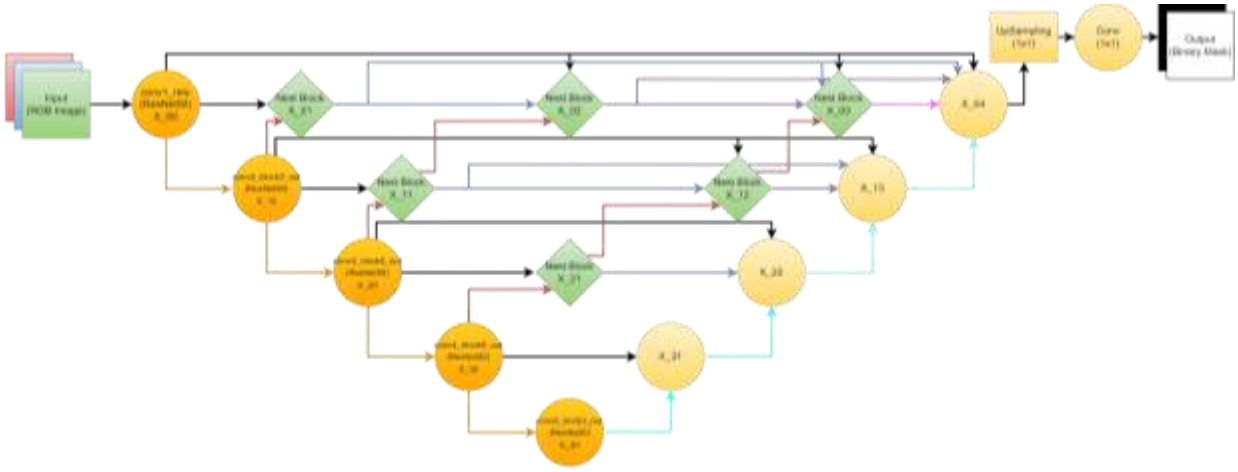


Figure 8: Architecture of the Attention-ResUNet++.

Figure 8 shows the architecture of the Attention-ResUNet++. The architecture in Figure 8 is an advanced derivative of U-Net++ and aims to achieve high accuracy and segmentation performance, especially in complex segmentation tasks. The model uses the ResNet50 architecture in the encoder part to perform deep feature extraction on the three-channel RGB image given as input. With this backbone, the model generates feature maps at different levels such as conv1_relu, conv2_block3_out, conv3_block4_out, conv4_block6_out and conv5_block3_out. Each block carries more abstract semantic information than the previous level, while learning increasingly more general concepts by decreasing the resolution. Each of the outputs from these layers is stored for use in the decoder side of the nested structure.

The U-Net++ architecture underlying the model differs from the classical U-Net structure in that it includes dense skip connections. These connections provide not only direct transfers from encoder to decoder, but also a multi-layered, hierarchical structure created by combining the outputs of previous decoder blocks. For example, blocks such as X_01, X_02, X_03 are created by combining the encoder outputs with information from lower level blocks located horizontally. Similarly, deep nested blocks like X_11, X_12, X_13, X_21, X_22 work with multiple inputs from both the encoder and the previous decoder intermediate outputs. These structures increase the feature reuse capability of the model and contribute to more efficient propagation of gradients. Thus, the training process becomes more stable, and the learning capacity of the model expands.
Another important element of this architecture is the integrated attention mechanisms. In particular, attention modules are integrated into nested decoder blocks to learn the contextual importance of each pixel. Thanks to this mechanism, the model can more clearly distinguish between object and non-object regions during segmentation. Especially in cases where the background has similar features to the object, such as camouflaged object detection, attention mechanisms allow the model to focus on the relevant regions, reducing misclassifications. Furthermore, this structure facilitates the filtering of patterns according to their regional density, producing segmentation masks with high visual integrity.

On the decoder side, the feature maps from different levels are first upsampled and then the channel size is reduced with 1x1 convolution layers. The information coming from intermediate blocks such as X_04, X_13, X_22, X_31 are combined by

upsampling to a common resolution. The 1x1 convolution process applied in the last layer generates only one output value for each pixel, creating the segmentation mask. This mask classifies pixels belonging to the target class as 1 and all other pixels as 0. The overall structure of the model produces strong outputs in terms of both structural integrity and classification accuracy thanks to multi-scale feature fusion, dense connections, attention modules and deep feature extraction. This allows the model to be stable. In addition, the conv blocks in the architecture are shown in Figure 9.

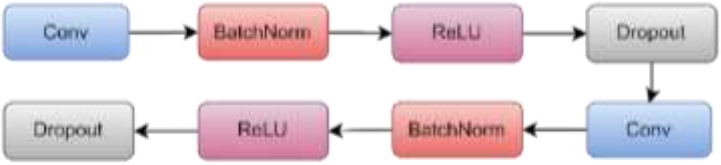


Figure 9: Conv blocks of the Attention-ResUNet++.

The Conv blocks in Figure 9 have been carefully designed to increase learning capacity and improve organizational capability in segmentation tasks. Each Conv block consists of a two-stage symmetric structure. The first stage consists of convolution, batch normalization, ReLU activation, and dropout, while the second stage repeats the same components in reverse order: dropout, ReLU, batch normalization, and convolution. This structure both facilitates the extraction of deep features and makes the model robust to overlearning. In particular, the bidirectional placement of regularization layers such as batch normalization and dropout allows the network to learn more stably. Moreover, thanks to this structure, each Conv block provides a balanced and optimized flow of information from input to output. This improves the segmentation performance of the model and minimizes fluctuations in the training process.

The feature maps weighted by the Attention mechanism are then merged with other features of the same resolution (the merging process indicated by the cross symbol). This enriched representation is then passed through the Conv Block in the structure presented in Figure 10 for higher level feature extraction.

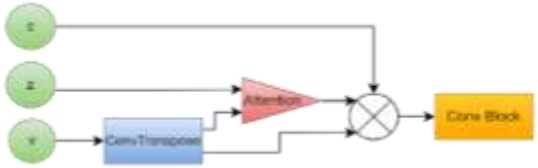


Figure 10: Nest blocks of the Attention-ResUNet++ model architecture.

Figure 10 visually presents the overall architecture of the nested blocks and explains how they work with careful information integration. First, multiple inputs from the encoder or earlier decoder levels are upsampled with the help of the ConvTranspose layer. This information is then filtered through the attention module so that only contextually meaningful features are highlighted. The output of the attention module is combined with other inputs of the same resolution to create a rich feature representation, and then the two-stage Conv Block structure described in Figure 9 comes into play. This block both reinforces feature fusion and balances the learning process with regularization components.

The architecture in Figure 10 presents the information fusion and processing steps of nested decoder blocks in a systematic way, especially in the U-Net++ structure. In this way, the model can effectively blend information from different scales to produce segmentation outputs that are both semantically and spatially robust. In this context, hyperparameter optimization was performed to improve model performance in segmentation problems using the Attention-ResUNet++ architecture. The findings indicate that both models demonstrate potential in the segmentation of camouflaged objects.

Performance of the Attention-ResUNet Model:

The Attention-ResUNet model demonstrated superior performance, particularly concerning the IoU metric. As mentioned in the summary section, when trained with CLAHE preprocessing, an Adamax optimizer, hyperparameters used include learning rate = 1e-5, dropout rates = 0.2 and 0.5. Three different optimization algorithms (Adam, Adamax, and Adadelta) were used in the optimization process and the success metrics of the model were calculated for each of them.

4. CONCLUSION AND RECOMMENDATIONS

The present study investigates the performance of Attention-ResUNet and Attention-ResUNet++ architectures, which are deep learning models augmented with attention mechanisms, for the detection of military camouflaged objects. The models

were trained and tested on the ACD1K dataset, and their performance was evaluated using metrics such as accuracy, precision, recall, F1 score, and intersection over union (IoU). The effect of the Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing technique on model performance was also analyzed.

4.1. Conclusions

With a learning rate of 1e-5, and a dropout rate of 0.2, the Attention-ResUNet model achieved 96.88% accuracy, 92.42% precision, 93.92% recall, 90.13% F1 score, and 92.01% IoU. This configuration was particularly successful in accurately identifying object boundaries (high IoU).

As illustrated in Table 3, the Attention-ResUNet model's performance metrics are presented, excluding the CLAHE method, and incorporating various optimizer and dropout values.

Table 3: Attention-ResUNet model metric results without CLAHE.

Optimizer	Adam		Adamax		Adadelta	
Dropout Rate	**0.2**	**0.5**	**0.2**	**0.5**	**0.2**	**0.5**
Accuracy	0.9670	0.9721	0.9681	0.9650	0.4852	0.5709
AUC	0.9806	0.9748	0.9739	0.9698	0.5231	0.5535
Dice Coefficient	0.8366	0.8369	0.8175	0.8035	0.2679	0.2645
F1 score	0.8366	0.8369	0.8175	0.8035	0.2679	0.2645
IoU	0.9432	0.9097	0.9233	0.8969	0.8194	0.8194
MCC	0.8118	0.8116	0.7887	0.7749	0.0240	0.0424
Mean IoU	0.8373	0.7318	0.7761	0.6897	0.4097	0.4097
Precision	0.8748	0.9206	0.9091	0.9343	0.1897	0.2014
Recall	0.9535	0.9251	0.9146	0.8673	0.5660	0.4644

The findings presented in Table 3 demonstrate that, in experiments employing various optimizers and dropout values, the Attention-ResUNet model, utilizing the Adam optimizer and a dropout value of 0.2, attains an even higher IoU value of 94.32% (wherein the accuracy was measured at 96.70%). As illustrated in Table 3, the performance of the Adamax and Adadelta optimizers is examined at varying dropout rates. The findings indicate that the Adam optimizer consistently yields superior outcomes in terms of IoU and accuracy metrics.

Table 4: Attention-ResUNet model metric results applied to CLAHE.

Optimizer	*Adam*		*Adamax*		Adadelta	
Dropout Rate	**0.2**	**0.5**	**0.2**	**0.5**	**0.2**	**0.5**
Accuracy	0.9734	0.9746	0.9688	0.9634	0.4554	0.3214
AUC	0.9788	0.9753	0.9724	0.9749	0.5319	0.5433
Dice Coefficient	0.8497	0.8445	0.8214	0.8178	0.2775	0.3016
F1 score	0.8497	0.8445	0.8214	0.8178	0.2775	0.3016
IoU	0.9387	0.9261	0.9201	0.9132	0.8194	0.8194
MCC	0.8262	0.8217	0.7941	0.7883	0.0339	0.0791
Mean IoU	0.8224	0.7831	0.7656	0.7446	0.4097	0.4097
Precision	0.9154	0.9442	0.9242	0.8760	0.1924	0.1957
Recall	0.9392	0.9136	0.9013	0.9284	0.6303	0.4644

The results in Table 4 demonstrate that CLAHE preprocessing enhances the IoU value to 93.87% (dropout 0.2) and 92.61% (dropout 0.5) when employed with the Adam optimizer. Nevertheless, the maximum IoU value (94.32%) is still obtained in the case without CLAHE (see Table 3, Adam, D=0.2). This finding indicates that the Attention-ResUNet model is a promising candidate, particularly in scenarios where optimal segmentation quality is paramount.

Performance of the Attention-ResUNet++ Model:

The Attention-ResUNet++ model, augmented with attentional mechanisms, exhibited superior performance, particularly with regard to the Accuracy metric. As indicated in the summary section of the Word document, the Attention-ResUNet++ model, which was trained with CLAHE preprocessing, an Adam optimizer, a learning rate of 1e-5, and a dropout rate of 0.2, achieved 98.32% accuracy, 96.75% precision, 93.87% recall, 90.58% F1 score, and 82.09% intersection over union (IoU). The findings

indicate that the Attention-ResUNet++ model demonstrates notable efficacy in achieving overall pixel classification accuracy.

Table 5: Metric results of Attention-ResUNet++ without CLAHE.

Optimizer	Adam		Adamax		Adadelta	
Dropout Rate	**0.2**	**0.5**	**0.2**	**0.5**	**0.2**	**0.5**
Accuracy	0.9807	0.9761	0.9763	0.9738	0.7784	0.7347
AUC	0.9895	0.9897	0.9843	0.9870	0.5741	0.5831
Dice Coefficient	0.8920	0.8846	0.8737	0.8605	0.1248	0.2198
F1 score	0.8920	0.8846	0.8737	0.8605	0.1248	0.2198
IoU	0.8199	0.8195	0.8202	0.8194	0.8194	0.8194
MCC	0.8753	0.8663	0.8554	0.8403	0.0314	0.0688
Mean IoU	0.4114	0.4099	0.4127	0.4097	0.4097	0.4097
Precision	0.9479	0.9111	0.9436	0.9247	0.2272	0.2411
Recall	0.9453	0.9612	0.9238	0.9307	0.0945	0.2185

Table 6: Metric results of Attention-ResUNet++ applied to CLAHE

Optimizer	Adam		Adamax		Adadelta	
Dropout Rate	**0.2**	**0.5**	**0.2**	**0.5**	**0.2**	**0.5**
Accuracy	0.9832	0.9826	0.9766	0.9709	0.2871	0.9448
AUC	0.9869	0.9808	0.9830	0.9862	0.5431	0.9875
Dice Coefficient	0.9058	0.9056	0.8748	0.8683	0.2949	0.8102
F1 score	0.9058	0.9056	0.8748	0.8683	0.2949	0.8102
IoU	0.8209	0.8677	0.8249	0.8194	0.8194	0.8194
MCC	0.8927	0.8915	0.8563	0.8469	0.0525	0.7828
Mean IoU	0.4153	0.5882	0.4307	0.4097	0.4097	0.4097
Precision	0.9675	0.9655	0.9464	0.8962	0.1888	0.7803
Recall	0.9387	0.9373	0.9227	0.9489	0.8946	0.9667

As indicated by the findings presented in Table 6, the (CLAHE, Adam, Dropout 0.2) configuration achieved the maximum accuracy value (98.32%) for Attention-ResUNet++. An analysis of Table 6 reveals that the 98.32% accuracy value achieved with the Adam optimizer employing a 0.2 dropout rate surpasses the 98.26% accuracy attained by the same optimizer with a 0.5 dropout rate and the outcomes yielded by the Adamax optimizer. In terms of intersection over union (IoU), as illustrated in Table 6, the Adam optimizer with CLAHE and a dropout value of 0.5 can attain an IoU value of 86.77%, yielding an accuracy of 98.26%. In comparison, the Attention-ResUNet++ results without CLAHE in Table 5 recorded the highest IoU value of 81.99% with the Adam optimizer and 0.2 dropout, indicating that CLAHE improves the IoU metric in this model.

A comparison of models and the CLAHE effect:

The analysis confirms that the Attention-ResUNet model is more successful in achieving the highest IoU values (Table 3: 94.32%, Adam, D0.2, without CLAHE), while the Attention-ResUNet++ model offers the highest Accuracy values (Table 6: 98.32%, Adam, D0.2, with CLAHE). The model with the highest IoU value is the Attention-ResUNet without CLAHE, while the model with the highest Accuracy value is the Attention-ResUNet++ with CLAHE.

An analysis of the effect of CLAHE pre-processing is conducted:

- In the Attention-ResUNet++ model, the highest accuracy (98.32%, Table 6) and the largest intersection over union (IoU) (86.77%, Table 6) are achieved in scenarios with CLAHE. A comparison with Table 5 (without CLAHE) demonstrates that CLAHE is generally beneficial for this model. For instance, when employing the Adam optimizer with a 0.2 dropout rate, the accuracy metric exhibited an enhancement from 98.07% (Table 5) to 98.32% (Table 6), while the intersection over union (IoU) metric demonstrated an improvement from 81.99% (Table 5) to 82.09% (Table 6).
- In the Attention-ResUNet model, the optimal configuration (Adamax, D0.2, with CLAHE) documented in the summary section attained 92.01% IoU (see Table 4). However, the highest IoU value achieved by the model (94.32%, Adam, D0.2) was obtained without CLAHE, as demonstrated in Table 3. This finding indicates that the efficacy of CLAHE for Attention-ResUNet may be contingent on the utilized optimizer and the designated metric. For instance, a comparison of Tables 3 and 4 reveals that CLAHE reduces the IoU from 94.32% to 93.87% (mean_iou from 83.73%

to 82.24%) when utilizing a 0.2% dropout with the Adam optimizer. However, this reduction does not occur when employing, as demonstrated in Figure 2, the Adamax optimizer exhibits a 2% dropout rate. It is evident that this phenomenon impacts the IoU from 92.33% to 92.01% (dice coefficient from 81.75% to 82.14%). It is noteworthy that the IoU values directly compared are 92.01% with CLAHE and 92.33% without CLAHE for Adamax D0.2. It is imperative to maintain consistency between metrics.

4.2. Prediction Times

Prediction times were analyzed as part of model performance evaluations. The Attention-ResUNet model's prediction time was documented as 6.3198 seconds. The prediction time of the Attention-ResUNet++ was found to be 8.5995 seconds.

4.3. Suggestions

In light of the findings, the following suggestions can be made for military applications involving the detection of camouflaged objects:

Model Selection and Metrics to Focus on:

- In scenarios where the primary objective of the application is to delineate the boundaries of camouflaged objects with the utmost precision and to optimize the segmentation quality (i.e., scenarios where attaining a high Intersection Over Union [IoU] value is imperative), the Attention-ResUNet model may prove to be a more suitable option. Notably, the IoU result of 94.32% with no CLAHE, Adam optimizer, and low dropout (i.e., 0.2) detailed in Table 3 is particularly salient in this regard.
- The primary objective of the application is to maximize the overall rate of correct classification of pixels (i.e., to achieve high accuracy). In such cases, it is recommended to utilize the Attention-ResUNet++ model with CLAHE preprocessing, an Adam optimizer, and low dropout (e.g., 0.2). The model demonstrates a high degree of accuracy, with an overall detection accuracy of 98.32%, as illustrated in Table 6.
- The F1 Score is a crucial metric for both models as it strikes a balance between precision and recall. As illustrated in Table 6, the Attention-ResUNet++ model achieved an F1 Score of 90.58% with the Adam optimizer and 0.2 dropout rate. For the Attention-ResUNet model, the optimal F1 scores were found to be 84.97% with CLAHE (Table 4) and 83.66% without CLAHE (Table 3), both utilizing the Adam optimizer with 0.2 dropout. These results indicate that while both models provide a robust balance between metrics, Attention-ResUNet++ demonstrates a higher overall F1 performance in military camouflage scenarios.

Use of CLAHE Preprocessing:

- A comparison of Tables 5 and 6 reveals that the application of CLAHE preprocessing for Attention-ResUNet++ tends to enhance performance metrics. Consequently, it is recommended to consider CLAHE as a standard preprocessing step when utilizing Attention-ResUNet++.
- In the Attention-ResUNet model, the effect of CLAHE is more variable. While the highest IoU value was obtained without CLAHE in Table 3, configurations with CLAHE also yielded competitive results, as demonstrated in Table 4. Consequently, the employment of CLAHE for Attention-ResUNet warrants further evaluation, contingent upon the selected optimizer and the performance objectives in question.

Hyperparameter Optimization:

- Tables 3, 4, 5, and 6 demonstrate that the selection of optimizer (Adam or Adamax) and hyperparameters, such as the dropout rate, exerts a substantial influence on the outcomes of both models. The Adadelta optimizer demonstrated substandard performance across all scenarios when compared to the other algorithms. In subsequent studies, a comprehensive search with a broader range of hyperparameters and various optimization algorithms (e.g., SGD with momentum, RMSprop) may further enhance the capabilities of the models.

4.4. Limitations

Although the proposed models achieve high performance, the computational load of the ResNet-50 backbone is relatively high for edge devices. Additionally, the models were tested mainly on the ACD1K dataset; performance on real-time video streams requires further validation.

Future Studies:

- A rigorous investigation into the various attentional mechanisms and forms of integration among these mechanisms may prove beneficial.
- The diversification of data augmentation techniques, along with the incorporation of more advanced methods (e.g., GAN-based data augmentation), has been demonstrated to enhance the generalizability of models.
- It has been demonstrated that experimenting with different backbone networks (e.g., EfficientNet, ResNeSt) can affect

the quality of feature extraction and, consequently, the segmentation performance.

- A thorough examination of model compression and quantization techniques is imperative to ascertain the viability of models for execution in real-time or on platforms with limited resources.
- The findings of this study demonstrate that U-Net models, augmented with attention mechanisms, can effectively address the intricate challenge of detecting camouflaged objects. The selection and configuration of models must be executed with meticulous care, taking into account the particular demands of the application and the performance metrics that have been prioritized.

Authors' Contributions

All authors contributed equally to all stages of the study, including conceptualization, code development, literature review, manuscript preparation, and final approval. Each author was actively involved in the research process and shared responsibility for the integrity and accuracy of the work. All authors have read and approved the final version of the manuscript.

Competing Interests

The authors declare that they have no competing interests.

References

[1] K. Karthiga and A. Asuntha, "CAMOUFLAGE-Net: comprehensive advanced model for optimal camouflaged target detection and analysis using groundbreaking elements," Signal, Image and Video Processing, 2025. doi: 10.1007/s11760-025-02928-5.

[2] B. Janakiramaiah et al., "Military object detection in defence using multi-level capsule networks," Research Square Preprint RS-3210306, 2023. [Online]. Available: https://www.researchsquare.com/article/rs-3210306/v1.

[3] C. Guo and H. Huang, "Enhancing camouflaged object detection through contrastive learning and data augmentation techniques," Engineering Applications of Artificial Intelligence, vol. 141, p. 109703, 2025. doi: 10.1016/j.engappai.2025.109703.

[4] X. Jiang et al., "MAGNet: a camouflage object detection network simulating the observation effect of magnifier," Research Square Preprint RS-976369, 2021. [Online]. Available: https://www.researchsquare.com/article/rs-976369/v1.

[5] B. Li, R. Zhou, L. Yang, Q. Wang and H. Chen, "MilDetr: detection transformer for military camouflaged target detection," IEEE Access, vol. 12, art. no. 10450905, 2024. doi: 10.1109/ACCESS.2024.3366974.

[6] O. Ronneberger, P. Fischer and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," arXiv preprint arXiv:1505.04597, 2015.

[7] Z. Ahmed, S. A. Tanim, F. S. Prity, H. Rahman, and T. B. M. Maisha, "Improving biomedical image segmentation: an extensive analysis of U-Net for enhanced performance," in Proc. ICETITE, Vellore, India, Feb. 2024.

[8] K. Türkarslan ve F. Hardalac, "Derin Öğrenme Yöntemleri Kullanılarak Havadan Elde Edilen Görüntüler Üzerinde Nesne Tespiti", ECJSE, c. 9, sy. 4, ss. 1398–1410, 2022, doi: 10.31202/ecjse.1135509.

[9] T. Yuan et al., "CoFiNet: unveiling camouflaged objects with multi-scale feature integration," SciSpace/Typeset.io, [Online]. Available: https://typeset.io/papers/cofinet-unveiling-camouflaged-objects-with-multi-scale-feature-oov8w5r71u. [Accessed: May 14, 2025].

[10] D. P. Naik and S. C. Tripathi, "Improving camouflage object detection using U-Net and VGG16," SciSpace/Typeset.io, [Online]. Available: https://typeset.io/papers/improving-camouflage-object-detection-using-u-net-and-vgg16-x4vj1d3zrv. [Accessed: May 14, 2025].

[11] H. Anandan, V. Karthikeyan and R. Jaisankar, "U-Net based camouflaged object detection," SciSpace/Typeset.io, [Online]. Available: https://typeset.io/papers/u-net-based-camouflaged-object-detection-28p6o029t7. [Accessed: May 14, 2025].

[12] Y. Zhang, J. Zhang, Y. Cao and H. Liu, "DMNet: diffusion model network for camouflaged object detection," Electronics, vol. 13, no. 18, p. 3690, 2024. doi: 10.3390/electronics13183690.

[13] "Adaptive camouflaged dataset (ACD1K)", Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/5f06c6964edcaa5f421efa39144d101532e19410e1323d2a75757d9f34baaf87. [Accessed: 04-May-2025].

[14] R. Kamble and P. Rajarajeswari, "Revealing hidden patterns: a deep learning approach to camouflage detection," International Journal of Computational Methods and Experimental Measurements, vol. 12, no. 1, pp. 97–105, Mar. 2024. doi: 10.18280/ijcmem.120111.

[15] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," International Journal of Forecasting, vol. 22, no. 4, pp. 679–688, Oct. 2006. doi: 10.1016/j.ijforecast.2006.03.001.

[16] X. Ji and S. Jiang, "Research on natural image semantic segmentation method based on improved U-Net model," in Third International Conference on Electronics, Electrical and Information Engineering (ICEEIE 2023), SPIE, p. 87, Oct. 2023. doi: 10.1117/12.3009019.

[17] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," Information, vol. 11, no. 2, p. 125, 2020. doi: 10.3390/info11020125.

[18] A. Mishra, "Contrast limited adaptive histogram equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints," arXiv preprint arXiv:2109.00886, 2021.

[19] S. Henna and A. Reji, "A data augmented approach to transfer learning for Covid-19 detection," arXiv preprint arXiv:2108.02870, 2021.

[20] A. Almansoori et al., "Image preprocessing with contrast-limited adaptive histogram equalization improves the segmentation performance of deep learning for the articular disk of the temporomandibular joint on MRI," Computers in Biology and Medicine, vol. 152, p. 106334, 2023. doi: 10.1016/j.compbiomed.2022.106334.

[21] A. Zafar et al., "Bayesian optimization for CLAHE parameter tuning in medical image analysis," Scientific Reports, vol. 14, no. 1, p. 5581, 2024. doi: 10.1038/s41598-024-55460-w.

[22] S. Sahu and A. Singh, "A comparative study on CLAHE and deep learning based enhancement techniques for low-contrast medical images," Biomedical Signal Processing and Control, vol. 75, p. 103591, 2022. doi: 10.1016/j.bspc.2022.103591.

[23] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[24] S. Zagoruyko and N. Komodakis, “Wide residual networks,” arXiv preprint arXiv:1605.07146, 2016.
[25] J. Hu, L. Shen and G. Sun, “Squeeze-and-excitation networks,” in 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00744.
[26] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, “Aggregated residual transformations for deep neural networks,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 1492–1500. doi: 10.1109/CVPR.2017.165.
[27] G. Huang, Z. Liu, L. van der Maaten and K. Q. Weinberger, “Densely connected convolutional networks,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 4700–4708. doi: 10.1109/CVPR.2017.243.
[28] Y. Sun, X. Wang and X. Tang, “Deeply learned face representations are sparse, selective, and robust,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, Jun. 2015, pp. 2892–2900. doi: 10.1109/CVPR.2015.7298881.
[29] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, “UNet++: a nested U-Net architecture for medical image segmentation,” in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11, Springer, 2018. doi: 10.1007/978-3-030-00928-1_1.
[30] Y. Peng, M. Sonka and D. Z. Chen, “U-Net v2: rethinking the skip connections of U-Net for medical image segmentation,” arXiv preprint arXiv:2311.17791, 2023.
[31] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang and Y. Zhou, “TransUNet: transformers make strong encoders for medical image segmentation,” arXiv preprint arXiv:2102.04306, 2021.
[32] S. He, R. Bao, P. E. Grant and Y. Ou, “U-Netmer: U-Net meets transformer for medical image segmentation,” arXiv preprint arXiv:2304.01401, 2023.
[33] O. Oktay et al., “Attention U-Net: learning where to look for the pancreas,” arXiv preprint arXiv:1804.03999, 2018.
[34] Y. Zhang, C. Wang and W. Ji, “Attention-gated networks for medical image segmentation: a survey of methods and applications,” Computers in Biology and Medicine, vol. 149, p. 105939, 2022. doi: 10.1016/j.compbiomed.2022.105939