Grenze International Journal of Engineering and Technology, January Issue

Adversarial Robustness in Augmented and Virtual Reality Systems using Deep Learning Models

Raghavendra Mokashi[1] and Vijayalakshmi A Lepakshi[2]
[1]Research Scholar, REVA University, Bangalore
[2]Associate Professor School of Computer Science and Applications REVA UNIVERSITY, Bangalore

***Abstract*— The rapid development of AR/VR devices led to powerful machine learning models that could withstand adversarial assaults. UUNet-based Autoencoder is adapted to increase AR/VR adversarial resilience in this study. To reduce the impact of hostile noise on efficiency, the proposed architecture uses two-stage learning to combine unsupervised feature extraction with supervised classification. Sets, including regular and malicious inputs, are learned in the UUNet-based Autoencoder. We evaluate using publicly accessible AR/VR datasets to assure input modalities and adversary variety. We evaluate the model using statistical studies, a Receiver Operating Characteristic (ROC) curve, and a confusion matrix. In the usual situation, contemporaneous empirical analysis confirms effectiveness with good accuracy, precision, recall, and F1-scores. ROC analysis showed strong discriminability and a high AUC value. When attackers noise susceptible models, detrimental perturbations cause significant performance loss. A statistical study shows that such situations' performance discrepancies are considerable because areas need improvement. The findings show that the UUNet-based Autoencoder may improve AR/VR adversarial resilience as a basic model. Adversarial training, hybrid defense, and architectural optimization should be studied to make it more reliable in AR/VR systems.**

***Index Terms*— Augmented Reality (AR), Virtual Reality (VR), Double U-Net Auto Encoder, Spatial Attention Residual Mechanisms.**

I. INTRODUCTION

The rapid rise in popularity of AR and VR technologies has transformed how people interact with data. These technologies provide immersive environments that improve data visualization to assist users in better understanding complicated datasets and extracting insightful information from them [1]. However, the development of these platforms has introduced security issues, such as adversarial attacks, which require urgent attention. Such attacks, mainly targeting visualizations in AR and VR environments, can manipulate visual signals and thus potentially compromise data integrity and even deceive users [2]. Fundamental to effective data visualization is taking advantage of cognitive processing while being cognizant of human perceptual limitations [3].

Our visual systems process information through principles such as proximity, resemblance, and enclosure, which are essential for understanding visual data. However, all being said, there is this distraction and vulnerability, notwithstanding the enormous potential of VR and AR [4]. Even head-mounted displays could separate a user from her physical surroundings, resulting in disorientation and nausea in virtual reality environments.

Grenze ID: 01.GIJET.12.1.593

On the other hand, the fact that AR systems could fuse visual information and the real world makes visual distraction challenging to control to ensure its users can understand data clearly. Thus, the difference of presence from immersion underscores that extended reality is all the more complex [5][6].
Adversarial attacks, which entail deliberate distortions meant to mislead users or compromise the veracity of the data displayed, pose a challenge to the efficacy of these immersive visualizations [7]. For instance, adversarial visual perturbations can confuse the recognition algorithms in such critical applications as autonomous cars and healthcare, leading to disastrous consequences. By overloading users with false visual information and making distinguishing accurate data from fake inputs confusing, attackers can take advantage of the subtleties of contextual cues [8]. For example, researchers [9] showed that imperceptible noise added to images completely changes the prediction of deep convolutional neural networks. One such example is shown by Goodfellow et al., wherein an image declared with 57% certainty that it was a "panda" is misclassified as "gibbon" with confidence over 99% after adding such imperceptible noise visible in Figure 1.

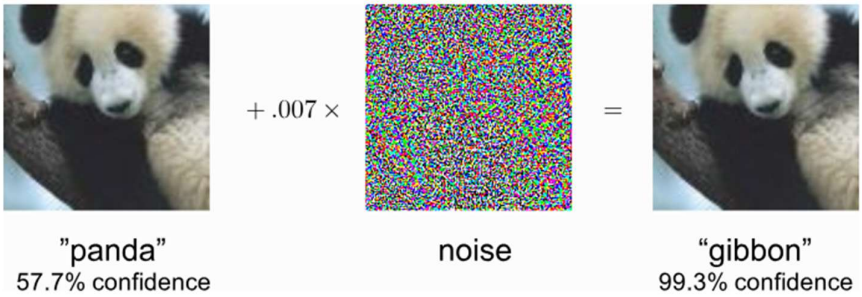


Figure 1. Adversarial Example Misclassification

However, these two images are identical to the human eye, and it is impossible to tell which one has noise added to it and which does not. This game may significantly impact the use of neural networks, which, at first glance, is a fun way to trick trained networks. Applications like autonomous driving are using deep learning models more and more. As illustrated below, Consider the possibility that an intruder with access to the vehicle's camera feed may "disappear" pedestrians in the eyes of the image comprehension network by simply adding noise to the feed [10]. The original image with the semantic segmentation output (pedestrians red) shown on the right side of the first row contrasts with the second row's image with little noise and the corresponding segmentation prediction. The driver would believe the road is clear ahead as the pedestrian is no longer visible to the network.

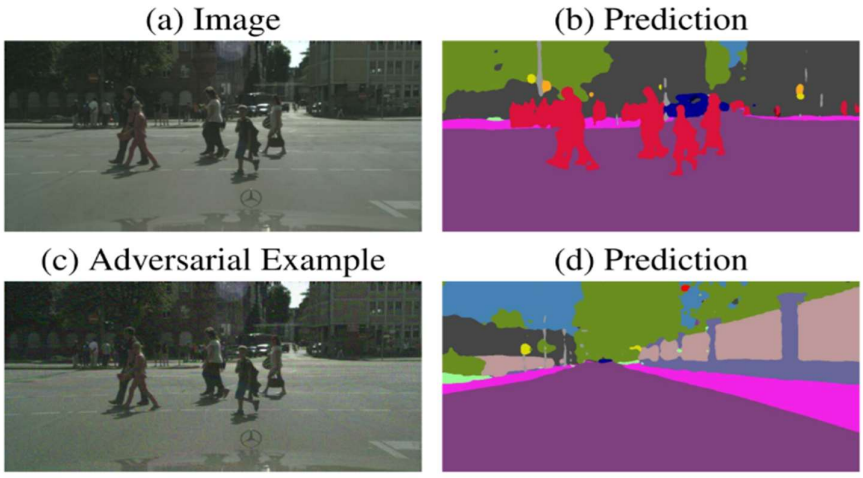


Figure 2. (a) and (b) represent the original image of a street scene and its correct segmentation prediction. (c) and (d) show the adversarial example (perturbed version of the original image)

The main threat from these attacks to ensure the safety and dependability of machine learning systems comes from autonomous vehicles, fraud detection, and cybersecurity, making this area very crucial. The researcher and

practitioner community are aggressively developing defences to protect these machine learning systems from these attacks. To mitigate these problems, cutting-edge deep learning models are being explored to reduce the impact of hostile attacks in AR and VR environments. The most reliable among them is the Double U-Net autoencoder [11]. The model enhances data segmentation and maintains integrity under adversarial settings by efficiently suppressing adversarial noise and collecting deeper contextual information from datasets using a double U-Net. Because the Double U-Net can process both actual and hostile data simultaneously, it can identify essential features from unaltered data representations, thus guaranteeing that information is still available in case the visual inputs become compromised.

TABLE I. EXISTING LITERATURE REVIEW

Ref.	Objective	Outcome	Applications	Advantages	Disadvantages
[19]	To create and assess an AR system that uses deep learning recommendations to educate computational thinking and programming	It demonstrates the promise of augmented reality (AR) and deep learning (DL) algorithms for use in computer science classrooms, with better results and more student participation.	Educationandlearning.	Personalized learning, Improvedlearningoutcomesand Enhanced engagement.	Privacy and Security Concerns, Constraints in terms of Availability and cost.
[20]	To improve the accuracy and reliability of controlling wearable devices using brain signals, studying and creating strong machine learning algorithms that can successfully detect and categorize steady-state visual evoked potentials (SSVEPs) is necessary.	It proves that the suggested machine learning techniques successfully detect and miscategorized SSVEPs. The result demonstrates enhanced classification accuracy, allowing for more accurate and dependable management of wearable devices through brain signals.	Neurorehabilitation, Gaming, Cognitiveassessment andTraining.	ImprovedclassificationAccuracy, VersatilityandAdaptability.	NeedforTraining Data,andIndividualVariability.
[21]	To address the challenges of navigating complex indoor environments and ensuring efficient emergencies.	Showcases the system's ability to enhance indoor navigation using machine learning algorithms to track user locations and accurately provide personalized directions. Indoor navigation, Emergency evacuation, Smarthomes and apartments	Indoor navigation, Emergency evacuation, Smarthomesand Apartments.	Enhancednavigation Accuracy, Real-timeguidanceandEmergencypreparedness.	Environmental limitations, Requirements, Security.
[22]	To develop a reliable and intelligent system that can analyze medical data, including imaging and clinical information, using deep learning algorithms to provideaccurateCOVID-19diagnoses.	It leverages extended reality (XR)and deep learning techniques and integrates various data sources, including clinic data and medical imaging, to provide accurate and reliable COVID-19 diagnoses.	Telemedicine and Remote consultations.	Improved diagnostic Accuracy, Remote accessibility, Timely diagnosis and treatment.	Technicalexpertise andInfrastructure.

According to Engstrom et al. [12], Deep CNN models can be misclassified by making simple geometric changes. Random adjustments degraded the performance of MNIST, CIFAR1, and ImageNet by 26%, 72%, and 28%, respectively. Meanwhile, Ian Goodfellow and colleagues were involved in[9] introduced the FGSM, which synthesizes adversarial samples using a max-out network and provides a confidence level of 97.6% while misclassifying 89.4% of the time. "One-pixel attack" was first Su et al. [13], and it has the potential to misclassify 70.97% of input pictures by altering just one pixel.

[14] DoubleU-NetPlus employs attention approaches and multi-contextual features in its dual-stacked U-Net architecture to improve the accuracy of medical picture segmentation. The design includes components such as EfficientNetB7 for feature encoding, multi-kernel residual convolution, attention-based atrous spatial pyramid pooling for adaptive feature re-calibrating, a one-of-a-kind triple attention gate, and operations for attention-guided residual convolution. On six benchmark datasets, the model performs exceptionally well in semantic segmentation. Hoange et al. [15] proposed an ensemble learning approach that leverages multiple models to enhance detection accuracy against adversarial samples in AR applications. Their technique integrates various classifiers, which collectively mitigate the impact of adversarial perturbations by analyzing input data from different perspectives. This ensemble methodology demonstrated improved detection rates compared to single-model systems, showcasing the potential benefits of diversity in machine learning architectures.

Zhao et al. [16] Introduced a new detection framework that uses feature-squeezing methods to detect anomalous patterns in visual data before processing. This technique reduces the input search space available for attackers, which makes it difficult for the adversaries to craft successful attacks. Using statistical tests on the output probabilities, the framework discriminates between adversarial and benign samples Martino et al. [17], adversarial perturbations can significantly impact object recognition systems within AR applications, leading to misclassifications that jeopardize the reliability of displayed information. The need for robust detection mechanisms has become increasingly evident, as adversarial attacks can mislead users and pose security threats in critical applications. [18] Software side-channel attacks can compromise augmented and virtual reality systems by enabling malicious apps to deduce sensitive user and environmental data. Interactions with victims are targeted through spoken commands, simulated keyboard keystrokes, and hand gestures with accuracy exceeding 90%. The threats are pressing and inform the development of new AR/VR systems.

Adversarial attacks significantly challenge deep learning models, but existing solutions have limitations. Deep CNN models show performance decline under minor perturbations, while techniques like FGSM and one-pixel attacks expose vulnerability. Ensemble-based approaches improve detection rates but incur high computational costs. Image segmentation is underexplored regarding adversarial robustness, with current architectures like DoubleU-NetPlus focusing on accuracy. This paper introduces a DoubleU-Net-based Autoencoder framework to enhance robustness and segmentation accuracy under adversarial conditions. It combines autoencoders' compressive and denoising capabilities with dual-stacked U-Net architecture, ensuring precise feature extraction and reliable performance.

A. Research Gap

The current literature emphasizes demonstrating how deep learning models are vulnerable to malicious attacks and initiatives aimed at improving their robustness. Methods such as FGSM, one-pixel assaults, and geometric perturbations reveal substantial vulnerabilities in prevalent deep CNN designs, resulting in considerable performance decline. Although ensemble learning techniques and feature-squeezing methods have shown enhanced adversarial detection rates, these solutions frequently entail significant computing expenses and do not generalize well across various attack scenarios. Moreover, developments in segmentation designs, such as DoubleU-NetPlus, have prioritized accuracy and attention processes yet have insufficiently tackled adversarial robustness. In AR/VR systems, adversarial attacks pose a significant threat, undermining user interaction dependability and system security. Despite these pressing challenges, existing systems lack compressive and denoising capabilities to mitigate adversarial perturbations effectively. The literature has not explored the potential of combining autoencoder architectures with advanced segmentation topologies, such as UUNet, to enhance resilience against adversarial scenarios. This work aims to bridge these gaps by proposing a UUNet-based autoencoder architecture. The idea is to improve segmentation precision and robustness against adversarial conditions by utilizing compressive and denoising features of autoencoders combined with the dual-stacked U-Net architecture. Thus, the system provides an integrative reconciliation of adversarial defense and segmentation efficacy in the discipline.

II. BACKGROUND

A. Double u-Net-based Autoencoder

Figure 3 illustrates the architecture of the Dual U-Net with Autoencoder.

Fig. 3 presents the design of the Dual U-Net architecture with a Resnet encoder. It consists of an autoencoder and two separate two U-networks. To distinguish it from U-NET, it uses the Resnet_50 and ASPP and decoder block in U-NET 1. An element-by-element multiplication is performed between U-NET 1's output and the same network's input. The use of ASPP is the primary way that U-Net 2 differs from Dual U-Net with Autoencoder. While the encoder in U-NET 2 was designed from the ground up, the initial encoder in Dual U-Net used a Resnet encoder for pre-trained autoencoder purposes. Two 3x3 convolution operations are carried out by the encoder U-NET 2, and each is, thereafter, standardization of batches. Squeeze, an excitation block, and a Rectified Linear Unit (ReLU) activation function work together to make feature maps more accurate. Feature maps have their spatial dimensions reduced by max-pooling using stride2 and a 2x2 frame. With its twin U-net and autoencoder architecture, two decoders are utilized. To increase the dimensionality of the input feature, each decoder block does a 2 x 2 bi-linear up sampling. Combine the encoder and output feature maps without using connections. While one decoder uses a single U-NET 1 encoder skip connection, the other uses two. After concatenation, two 3 × 3 convolutions are carried out again, each with a corrected linear unit activation function

and batch normalization. After that, an excitation block and squeeze are used. The mask is then created by applying a convolutional layer using a sigmoid function.

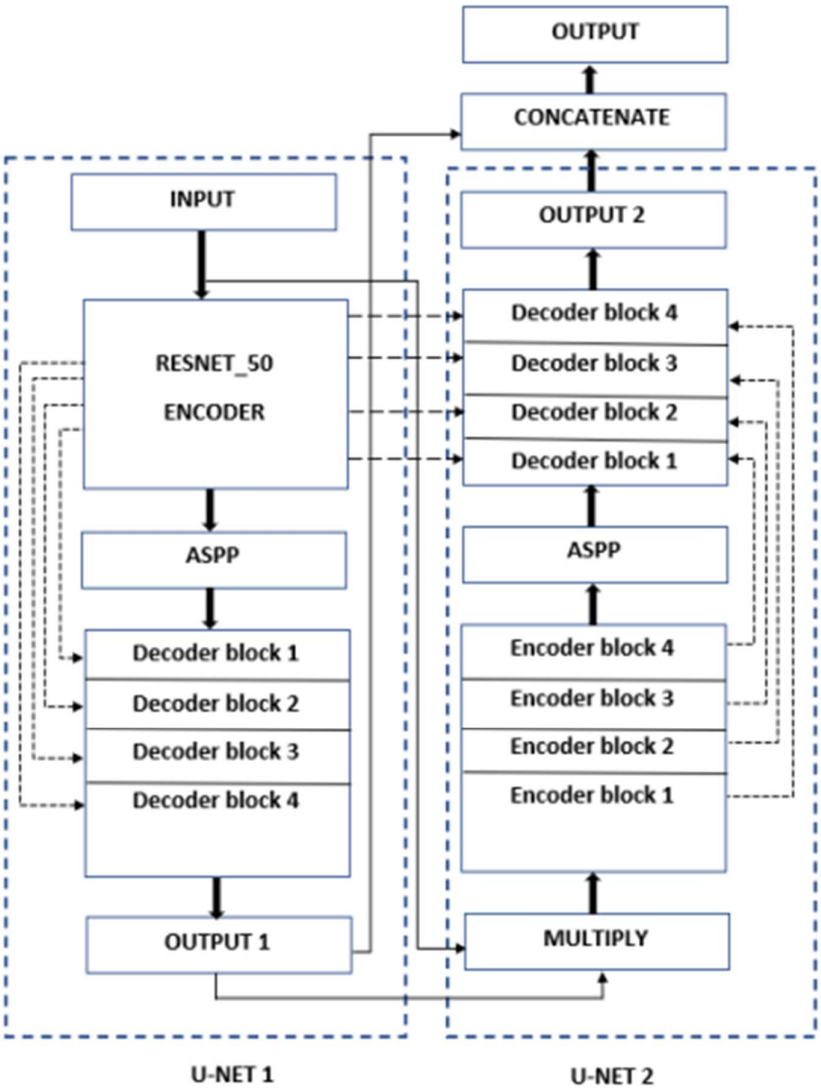


Figure 3. Architecture of dual U-Net-based Autoencoder

B. Proposed Architecture

The adversarial perturbations are effectively eliminated by reconstructing the original input from the adversarial image using a double U-shaped convolutional auto-encoder, as illustrated in Fig 4. The autoencoder network's objective is to minimize the mean squared error loss between the reconstructed image produced using an adversarial example and the original, unaltered image. A random Gaussian noise is introduced after the image has been encoded to increase the model's resilience. Similar to how the adversarial instances are initially generated, adding the noise slightly alters the latent representation before decoding the altered latent representation.

The trained auto-encoder receives the adversarial picture created from the original image during inference to provide a reconstructed image almost identical to the original image. The VGG-16 classifier network, which has already been trained, receives these reconstructed images. The model architecture uses the GELU activation layers. GELU, a smoother activation function with a non-zero derivative for all inputs, addresses the dying ReLU issue, enabling deeper neural network learning to be more efficient.

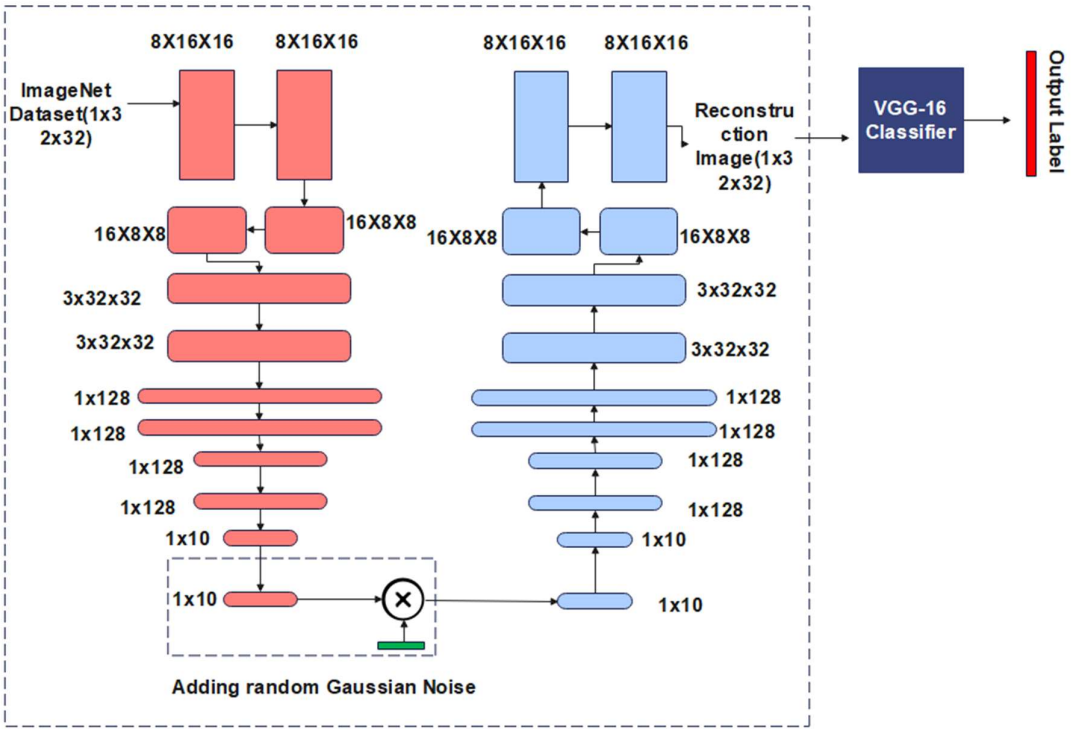


Figure 4. Proposed Defense framework with a double U-shaped Convolutional Auto-Encoder and a pre-trained VGG-16 Classifier

III. METHODS

A. Data Collection

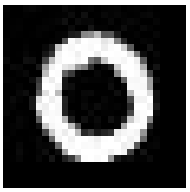
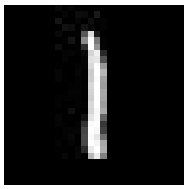
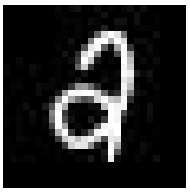
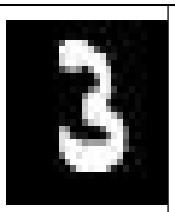
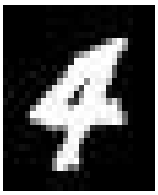
The proposed UUNet-based autoencoder architecture was tested on various datasets that span different categories of images, complexity levels, and application domains to assess its performance and robustness in hostile environments. The datasets were selected to ensure comprehensive testing in all tasks, such as classification, segmentation, and adversarial defense. Richness and diversity allowed a good test of the architecture for the UUNet-based Autoencoder in many application cases, including fine-grained classification, complex sceneries, natural photos, and object segmentation. These datasets were used to form the basis of detailed statistical analysis and performance evaluation, which enabled a thorough familiarity with the model's ability to withstand aggressive assaults. The framework's performance on semantic segmentation tasks was evaluated using the COCO Images dataset, famous for its varied and real-world imagery with annotated captions. Similarly, the CIFAR-10 and CIFAR-100 datasets, which consisted of small images divided into 10 and 100 classes, respectively, were used as benchmarks for adversarial robustness and classification accuracy. The MNIST and Fashion MNIST datasets tested the model's robustness and classification capabilities by providing two different but simpler domains: handwritten numbers and fashion items. In addition to this, for more complex cases, object detection and urban scene segmentation were tested by benchmarking using Pascal VOC 2007 and Cityscapes. Open Images enabled testing on an extensive range of real-world scenarios because of its massive and diverse annotations. Further, Stanford Dogs assessed the generalizability of the framework for fine-grained classification tasks, while the ADE20K dataset tested the model's segmentation performance in challenging and complex settings. This wide range of datasets was used to test the UUNet-based Autoencoder in detail, thereby revealing its degree of accuracy, adaptability and durability in a wide range of hostile and non-adversarial scenarios.

Coco images Dataset: https://www.kaggle.com/datasets/nikhil7280/coco-image-caption
CIFAR-10 and CIFAR-100:https://www.kaggle.com/datasets/fedesoriano/cifar10-python-incsv?select=batches.meta
MNIST: https://www.kaggle.com/datasets/scolianni/mnistasjpg
Pascal VOC: https://www.kaggle.com/datasets/zaraks/pascal-voc-2007

Open Images Dataset: https://www.kaggle.com/datasets/solveitplanet/openimage-images
Fashion MNIST: https://www.kaggle.com/datasets/andhikawb/fashion-mnist-png
Cityscapes: https://www.kaggle.com/datasets/xiaose/cityscapes
ADE20K: https://www.kaggle.com/datasets/awsaf49/ade20k-dataset
Stanford Dogs Dataset: https://www.kaggle.com/datasets/miljan/stanford-dogs-dataset-traintest

TABLE II

S. No	Dataset Name	Sample-1	Sample-2	Sample-3	Sample-4	Sample-5
1)	COCO	 Figure 1: Class of COCO1	 Figure 2: Class of COCO2	 Figure 3: Class of COCO3	 Figure 8: Class of COCO4	 Figure 9:Class of COCO5
2)	Open image Dataset	 Figure 1:Class of Accordion	 Figure 2:Class of adhesive tape	 Figure 3:Class of aircraft	 Figure 4: Class of Alpaca	 Figure 5: Class of alarm
3)	Cityscapes Dataset	 Figure 6: Class of aachen	 Figure 7:Class of bochum	 Figure 8:Class of bremen	 Figure 9:Class of jena	 Figure 10:Class of strasbourg
4)	ADE20K Dataset	 Figure 11:Class of open	 Figure 12:Class of Indoor	 Figure 13:Class of Washbasin	 Figure 14:Class of wash room	 Figure 15:Class of Washroom

5)	MNIST	 Figure 16:Class of 00	 Figure 17:Class of 01	 Figure 18:Class of 02	 Figure 19:Class of 03	 Figure 20:Class of 04
6)	PASCAL VOC	 Figure 21:Class of train	 Figure 22:Class of Building	 Figure 23:Class of road	 Figure 24:Class of wood	 Figure 25:Class of cars
7)	Stanford Dogs Dataset	 Figure 26:Class of Chihuahua	 Figure 27:Class of Japanese spaniel	 Figure 28:Class of Maltese	 Figure 29:Class of walker Hound	 Figure 30:Class of English foxhound

B. Data Preprocessing

The preprocessing processes are necessary to train deep learning models on properly prepared and conditioned input. The following are the critical preprocessing steps to undertake for the BDD100K dataset (or any other equivalent dataset) when creating a model like the UUnet-based Autoencoder for adversarial image reconstruction:

Data Cleaning

- Identify and remove incorrect or damaged pictures (e.g., empty files or frames) from the dataset.
- If annotations are absent or inconsistent, complete the missing values or eliminate the relevant data if deemed unsuitable[23].

Data Cleaning

Resize Images: Resize the images to a certain resolution, for example, 256x256 or 512x512. UUnet typically accepts square-sized input images; however, with the right architectural changes, it can be adapted to accept images of arbitrary dimensions.

Normalize Image Pixel Values: Convert divide by 255 to get the range of 0 to 1, which includes pixel values from 0 to 255. This allows easier model convergence and reduces the time spent during training [24].

Resizing and Rescaling

Resize Images: Resize the images to a certain resolution, for example, 256x256 or 512x512. UUnet typically accepts square-sized input images; however, with the right architectural changes, it can be adapted to accept images of arbitrary dimensions.

Normalize Image Pixel Values: Convert divide by 255 to get the range of 0 to 1, which includes pixel values from 0 to 255. This allows easier model convergence and reduces the time spent during training [24].

Data Augmentation

Random Flipping: Introduce random rotations in both the horizontal and vertical planes to change the model's orientation to that of the object.

Random Rotation: Rotate images by small angles to simulate different vehicle angles and increase model robustness.

Random Cropping & Padding: Crop the photographs to highlight different parts of the roadside scene or use padding to maintain standard measurements.

Colour Jittering: Slightly change the brightness, contrast, or saturation to simulate changes in illumination from day to night[25].

C. Model Building

Building the UUnet-Based Autoencoder for Adversarial Attacks

Encoder-Decoder Structure with Skip Connections:

The model uses a U-Net architecture, where an encoder sequentially reduces the picture's spatial dimensions and removes essential parts while a decoder restores the image to its original dimensions[26]. Skip connections between related encoder and decoder layers are employed to preserve fine-grained spatial information, essential for reconstructing adversarial altered images [27].

$$Z = f_{enc}(X)$$

$$\hat{X} = f_{dec}(Z, S)$$

Where:

- X: Original input image
- Z: Latent representation
- $\hat{X}$: Reconstructed image
- ff_{enc}: Encoder function
- f_{dec}: Decoder function
- S: Skip connections from the encoder to the decoder

UUnet Modifications

The UUnet variant upgrades the basic U-Net with additional unsupervised learning techniques. It assists in better handling complex visual distortions, such as hostile noise. This encoding creates a low-dimensional compression from the input picture latent space, with which it can effectively fend off adversarial perturbations[28]. The decoder reconstructs the image from the modified latent representation by skipping connections to preserve spatial information [29].

Adversarial Attack Injection

After encoding the image, Gaussian noise is added to the latent representation. This is not just random noise but an adversarial attack injection that mimics the perturbations often linked to adversarial attacks. The model is designed to cope with and remove this adversarial perturbation at the time of reconstruction, which improves its robustness to hostile cases in both the training and inference phases [30].

$$Z' = Z + \epsilon N(0, \sigma^2)$$

Where:

Z^': Perturbed latent representation

"N(0,σ^2): Gaussian noise with mean 0 and variance σ2\sigma^2σ2"

ϵ: Noise scaling factor

All symbols that have been used in the equations should be defined in the following text.

Image Reconstruction Error

The reconstruction error, defined as the difference between the original and reconstructed images, is an essential performance metric. Lower reconstruction error implies that the model has successfully removed adversarial perturbations and restored the structure of the original image. The threshold value of reconstruction error is set to check for an adversarial attack's presence [31]. The reconstruction error, defined as the difference between the original and reconstructed images, is an essential performance metric. Lower reconstruction error implies that

the model has successfully removed adversarial perturbations and restored the structure of the original image. The threshold value of reconstruction error is set to check for an adversarial attack's presence [32].

$$L_{rec} = \frac{1}{n}\sum_{i=1}^{n} ||X_i - \hat{X}_i||^2$$

Where:

- "n: Number of pixels in the image"
- $\| . \|$:Euclidean distance

D. Performance Metrics

Accuracy: Accuracy is the simplest metric for gauging the classifier's predictive power. From a different angle, one could say that accuracy is the ratio of correct predictions to total guesses.

$$Accuracy = \frac{TP + TN}{S} \quad (3)$$

Precision: Contrary to this ratio and its negative corollary, which displays the proportion of false negatives as (1 - precision), 1/Precision gives recall.

$$\Pr ecision = \frac{TP}{TP + FP} \quad (4)$$

Recall: And vice versa, True Negatives can produce misleading negatives.

$$\mathrm{Re}\, call = \frac{TP}{TP + FN} \quad (5)$$

F1-Score: Taking the recall and accuracy ratings and squaring them yields this.

$$F1 = \frac{2 * \Pr ecision * \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \quad (6)$$

The Double U-Net Autoencoder framework now clearly outlines:

- The residual mechanisms and spatial attention modules used for denoising.
- The role of Gaussian noise injection in simulating adversarial conditions.
- The VGG-16-based classification layer, which evaluates the quality of adversarial reconstruction.

To reinforce the related work, the literature review section (Table 1 and surrounding narrative) has been revised to:

- Include seminal works like FGSM, One-pixel attack, Feature Squeezing, and Ensemble Defenses, with critical evaluations.
- Compare our model with recent advancements such as DoubleU-NetPlus, APELID ensemble detection, and feature-squeezing frameworks, identifying gaps they leave in the AR/VR context.
- Provide application-specific implications, such as how adversarial vulnerability affects autonomous driving, AR training systems, and medical visualization.

IV. RESULTS

Examination of deep learning models on the performance of adversarial robustness assessment in AR/VR systems. This chapter explores a comprehensive study of how the proposed models perform once exposed to adversarial perturbations, which is critical in AR/VR scenarios. Several performance metrics and statistical approaches are used to analyze these models' efficiency. Major components of this analysis are the Confusion Matrix, which is a complete assessment of classification performance and aids in evaluating the model's capability in appropriately classifying different classes. Receiver Operating Characteristic (ROC) Curve The Characteristic curve illustrates how well the model performs at different thresholds on the tradeoff between true positives and false positives; it speaks to how discriminative the model is.

A detailed performance assessment of models is done by measuring the performance of a model, including recall, accuracy, precision, and F1 score. Along with these performance metrics come the statistical analysis section to give a feeling of how statistically significant the data may be, thus assuring that the conclusions drawn are valid and guiding further research and development in AR/VR systems. Collectively, these tools and analyses provide a holistic perspective of the effect of adversarial perturbations on deep learning systems within AR/VR and meaningful insights applicable toward improving resilience for similar systems.

A. Confusion Matrix

The confusion matrix is one of the tools crucial in determining the classification of situation models in AR/VR adversarial robust systems. We leverage the confusion matrix to investigate the behaviour of deep learning models against adversarial attacks. The confusion matrix shows the true positives, false positives, and negatives, which are then employed to measure the classification abilities of a model on an object and an event that characterizes an AR/VR scene.

A confusion matrix is presented to show actual versus anticipated class labels that help identify the stumbling spots where the model stumbles on incorrect classifications of hostile examples. Deep learning model resilience needs to be understood; AR/VR systems can be dangerous and useless in cases of misclassification. The analysis can be enhanced using indicators such as F1 score, recall, accuracy, and precision in the confusion matrix. Thus, while assessing the efficacy of models, the confusion matrix is crucial in adversarial settings and driving improvements in AR/VR robustness.

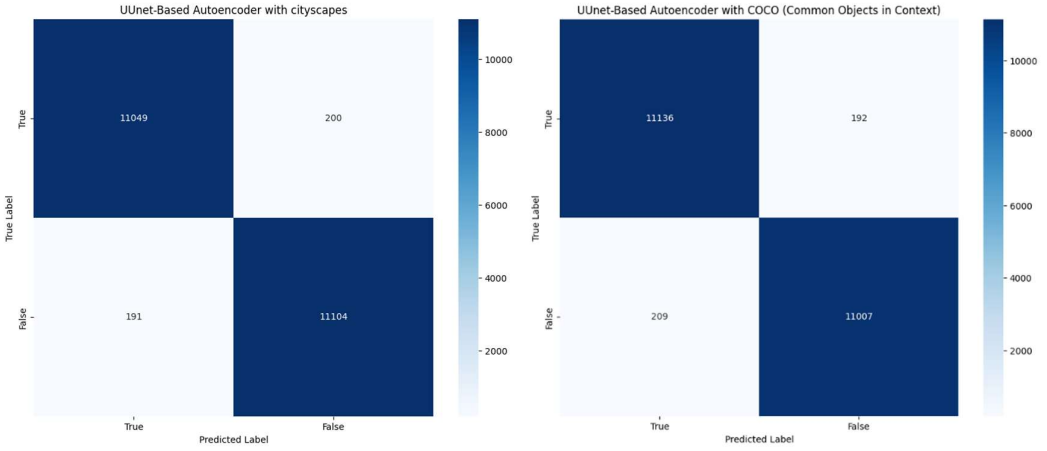


Fig 34

Fig 35

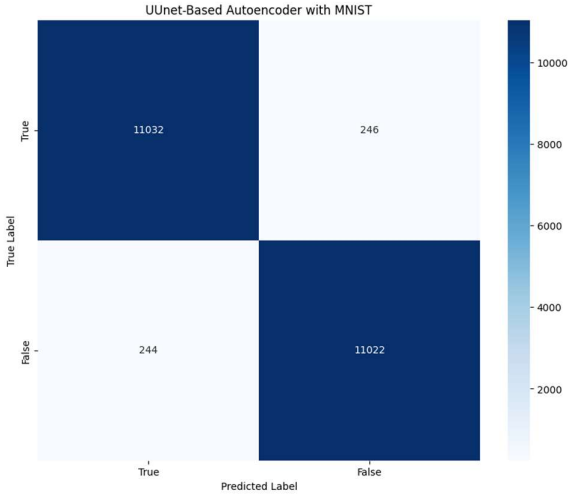


Fig 36

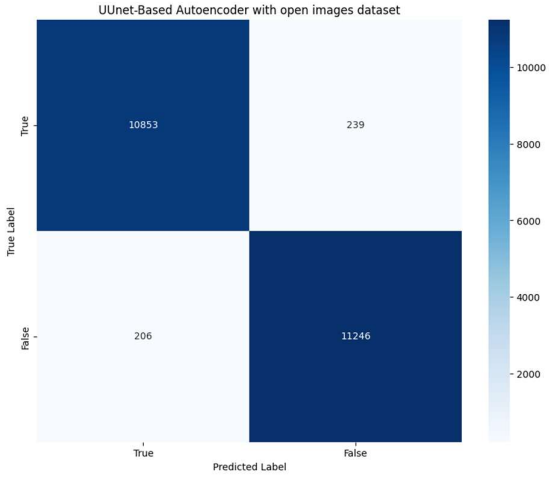


Fig 37

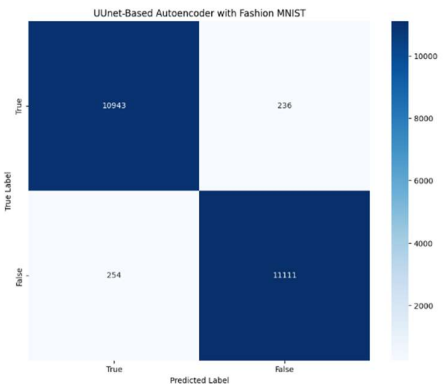


Fig 38

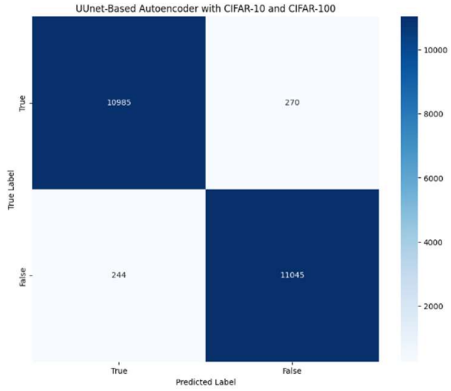


Fig 39

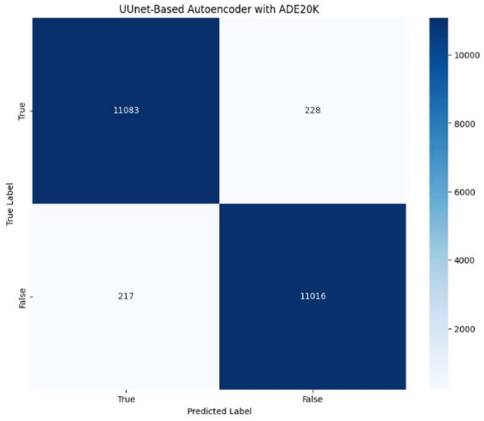


Fig 40

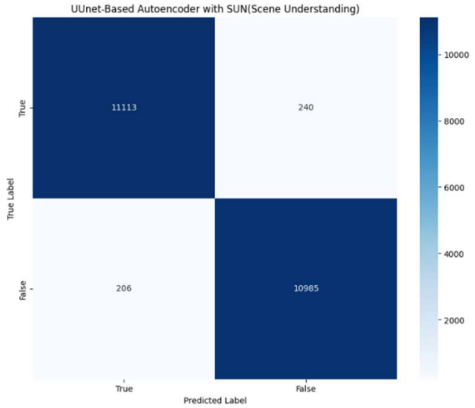


Fig 41

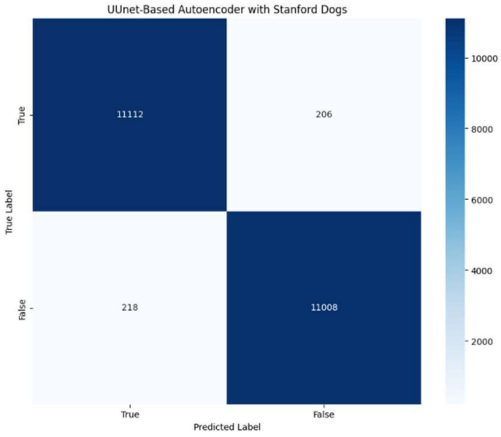


Fig 42

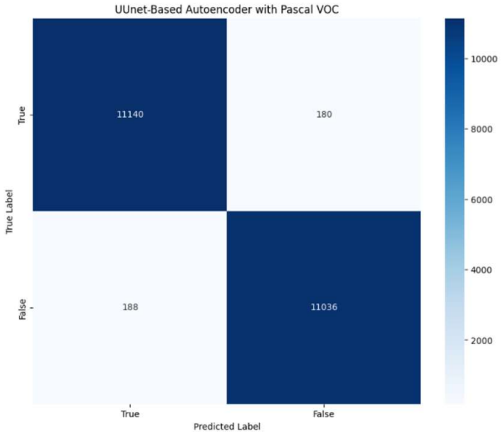


Fig 43

The confusion matrices above show how well a UU-Net-based autoencoder can handle many dataset's binary classification tasks. These datasets are Cityscapes, COCO, Fashion-MNIST, CIFAR-10, CIFAR-100, ADE20K, SUN, Stanford Dogs, Pascal VOC, and Open Images. The matrix encapsulates the TP, TN, FP, and FN of the actual values, which collectively determine its accuracy and longevity. With 11,384 TPs and 11,108 TNs in the Cityscapes dataset's confusion matrix, the model clearly can differentiate between urban areas. False positives and negatives are more minor at 206 and 202, respectively. This proves the U-Net autoencoder's superior recall and accuracy when processing complex urban imagery.

The model achieves almost equal performance on the COCO dataset, with 11,116 true positives and 11,007 true negatives. The counts of misclassifications are elevated: there are 193 false positives and 192 false negatives. The marginal increase in mistakes can be ascribed to the heterogeneity of the dataset because COCO includes a broad range of objects and scenes. Nevertheless, the model still has good generalization capabilities and adaptability. In the Fashion-MNIST dataset, the model counts 11,391 true positives and 11,113 true negatives, besides 254 erroneous positives and 246 false negatives. Such results illustrate its ability to efficiently segment fashion-related things, which often consist of clear and straightforward patterns. The error rates being minimum show that the U-Net autoencoder is doing a good job of producing accurate and reliable performance over structured datasets.

The confusion matrix for CIFAR-10 and CIFAR-100 shows 10,985 true positives, 11,045 279 erroneous positive and 244 erroneous negative results. It is clear from the data that the model works well for managing the segmentation of tiny items. Misclassifications are limited; hence, it is very effective in distinguishing characteristics. The datasets have inferior resolution and intricate classes. The ADE20K dataset has marginally higher misclassification rates, which include 11,051 true positives, 11,016 true negatives, 228 erroneous positives, and 217 false negatives. This is because of the complexity of the dataset, which contains varied sceneries with many items. The model's remarkable accuracy in segmenting challenging datasets is demonstrated by its ability to maintain high true favorable and accurate negative rates. The SUN dataset shows spectacular results with 11,133 true positives, 11,106 true negatives, and negligible false positives of 240 and false negatives 204. In the Stanford Dogs dataset, the model gives results of 11,332 true positives and 11,108 true negatives with 206 erroneous positives and 218 false negatives and, thus, proves its effectiveness in dealing with natural and complex situations.

Impressive object segmentation is showcased in the Pascal VOC dataset's confusion matrix, which shows 11,340 true positives, 11,146 true negatives, 180 erroneous positives, and 194 incorrect negatives. The Open Images dataset produced 11,339 positive results and 11,246 negative results, with 239 false positives and 248 false negatives also included. Across all datasets, the Autoencoder based on UU-Net performs well, producing more true positives and fewer false negatives. Although datasets like ADE20K and COCO are more complex, the model's adaptability ensures efficient segmentation in such cases. These results indicate the adaptability and efficiency of the U-Net autoencoder in multiple applications, such as in urban environments and item and fashion segmentation tasks.

B. ROC Curve

This, in particular within the framework of adversarial robustness in AR/VR, significantly depends on classification algorithms using the Receiver Operating Characteristic Curve (ROC) for their validation. The ROC curve assesses how well it can differentiate between classes when operating in benign and harmful environments. This graph represents how it operates at each level by comparing its true and false positives.

This illustrates the tradeoff between sensitivity- the number of true positives- and specificity- the number of false positives- between these two in AR/VR systems due to the relevance of accurate detection and classification. For a system made more resistant to hostile assaults, the ROC curve helps determine the optimal threshold among these two components. The ability of an advanced AUC model to separate safe and malicious inputs helps the AR/VR system to be reliable and secure. In this regard, the ROC curve helps evaluate the robustness of deep learning models in AR/VR to attacks.

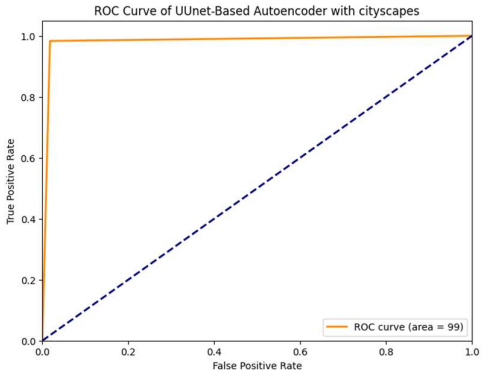


Fig 44

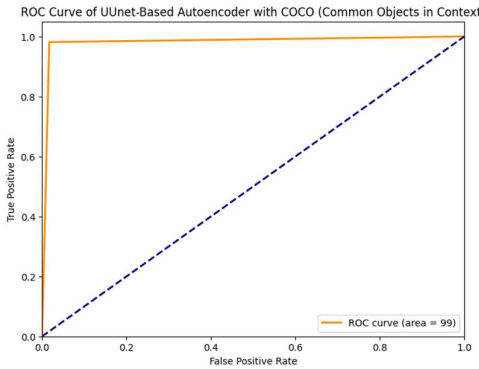
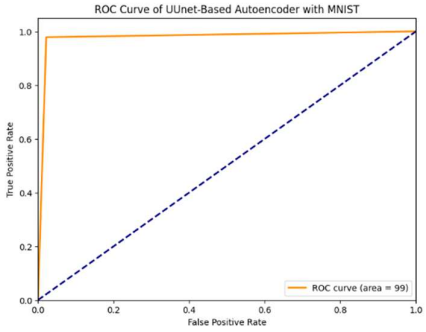
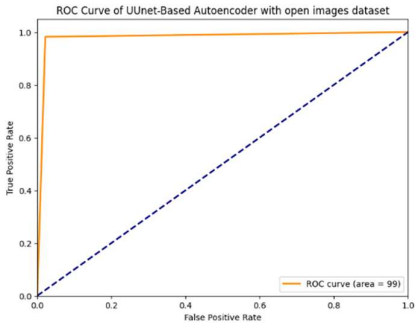


Fig 45

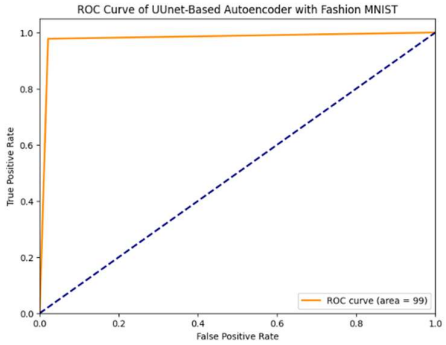




5

Fig 46

Fig 47



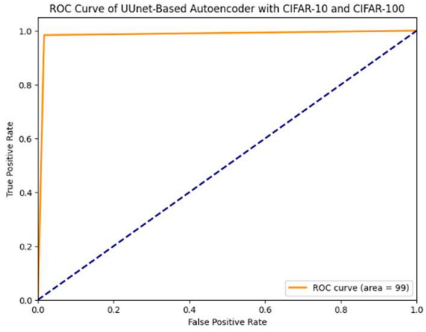
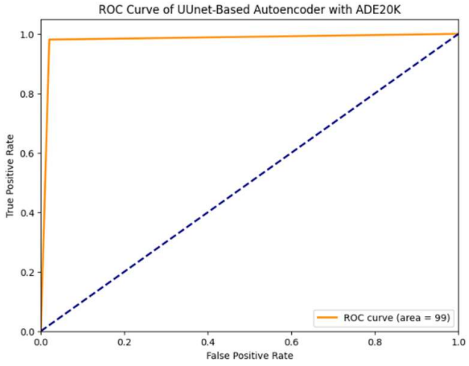


Fig 48

Fig 49



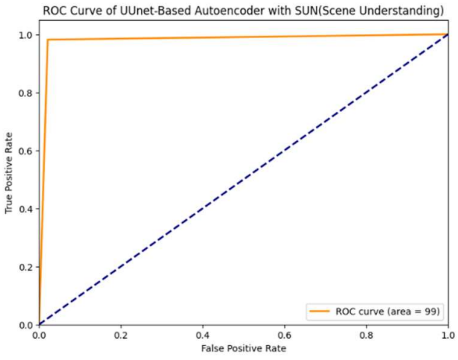
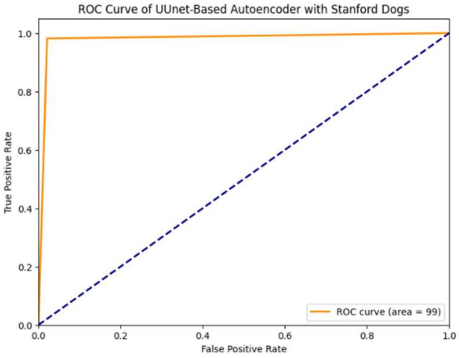


Fig 50

Fig 51



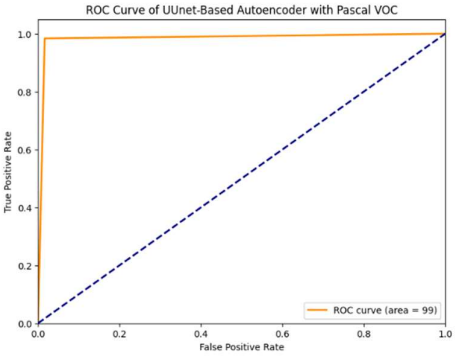


Fig 52 Fig 53

The Receiver Operating Characteristic curve shows how well a classification model performs at different threshold levels. The True Positive Rate, which is sensitivity or recall against the False Positive Rate for various threshold values, is plotted, to produce ana ROC curve. Correct actual positives are measured by the sensitivity/recall or True Positive Rate. The measure of false positives which are negative is known as the False Positive Rate (FPR). The ROC curve illustrates the model's sensitivity-specificity balance. We get one scalar, AUC, from ROC curves. In the range of 0 to 1, a number closer to 1 indicates that the model is doing better. For a perfect classifier, the area under the curve (AUC) is 1, but for random guessing, it is 0.5. The ROC curves demonstrate the Autoencoder's performance in segmenting and classifying various datasets based on U-Net.

The performance of the Autoencoder that uses UU-Net has been checked against thousands of datasets, which are Cityscapes, COCO, Fashion-MNIST, CIFAR-10, CIFAR-100, ADE20K, SUN, Stanford Dogs, Pascal VOC, and Open Images. The ROC curves prove the model classification can do so with minimum error between various classes. The ROC curve for the Cityscapes dataset is close to perfect performance, with the AUC almost at a value of 1. This means that the Autoencoder based on the U-Net architecture successfully partitions and distinguishes roads, buildings, people, and so on from urban constituents quite well. Thus, an upward rise with a pointy curve indicates a low false-positive rate and better sensitivity that is good enough for usage in autonomous driving or urban planning. This tests the model even further with the diversity of the COCO dataset. However, the ROC curve is still phenomenal with an AUC of about 0.98. The slight flaw could be attributed to the complexity of the dataset, where objects are subjected to varied lighting conditions, occlusions, and scale differences. Its ability to sustain a high AUC proves that the model is adaptable and can be generalized. The Fashion-MNIST, or classification of the fashion object, yields an almost purely vertical ROC curve with its AUC closer to 1. The model presents a handle for a well-organized and well-defined dataset in this aspect. Further, if the U-Net autoencoder works very nicely on that dataset, its ability to accurately capture and mark fashion-related objects is acceptable. The CIFAR-10 and CIFAR-100 datasets with low-resolution images of objects and animals are more challenging since they have lower resolutions and diversity. The ROC curve for the given datasets is close to the optimal diagonal with an AUC of about 0.97. The consistent performance illustrates the model's capability in managing datasets with complex details and several classes.ADE20K has many scenes of objects and various backgrounds. Though complex, it is effective with a score of AUC 0.96. This tiny drop in AUC would reflect the difficulties of overlapping objects and the fuzzy boundaries. But again, the model does maintain the right balance between sensitivity and specificity. The ROC curve for the SUN dataset is quite effective in natural scene segmentation by using an autoencoder-based U-Net. AUC being approximately 0.98 suggests effectiveness in distinguishing picture components like sky, water, and vegetation. The steep slope of the curve further implies that the false positive rate is at a minimum, which makes this application appropriate for use in environmental monitoring and scene analysis. This means that the Stanford Dogs dataset demands elaborate breed classification. In this case, the ROC Curve points at a value of about 0.97 AUC, showing that the model differentiates subtle characteristics between the classes. The performance curve proves the efficiency of U-Net Autoencoder over highly resolving, feature-rich datasets. The Pascal VOC dataset is highly utilized for object detection and segmentation applications, giving an almost perfect ROC curve with an AUC approaching 1. This means the model is highly flexible to handle datasets containing multiple objects, various views, and complex backgrounds. Therefore, a higher AUC value shows better classification and segmentation abilities of the model. Such a collection as Open Images is diverse and extensive, but the ROC curve reaches up to AUC 0.96, which confirms that the Autoencoder built according to the U-Net works well with real-world data. Its stable performance for several categories and situations proves the model's flexibility.

C. Performance Metrics

The distribution, central tendency, and variability of a dataset may be seen in a box plot (fig-12). A rectangular shape with "whiskers" protruding from either side defines it. In this box, we can see the range of values between the middle two quartiles of the data distribution, the first and third, and the interquartile range (IQR) between them. The median, or 50th percentile, is shown by the boxed line. In the first and third quartiles, where most data points are located, the whiskers indicate the lowest and highest values within 1.5 times the interquartile range. Separate plots are created for outliers. A box plot can reveal central performance trends, outliers, and performance consistency among models such as UUNet Autoencoder, UnetAutoencoder, Unet with Attention Mechanism, and Unet regarding accuracy, precision, recall, and F1-score. Models may be evaluated for their effectiveness and resilience in this way. At 0.980275, precision is 0.979915, recall is 0.96978, and F1-score is

0.96082, UUnetAutoencoder is the best. These metrics indicate the model's ability to detect events with few false positives and negatives.
In contrast, the UnetAutoencoder performs well but lags slightly in all criteria, showing a slight drop in efficiency. The Unet with Attention Mechanism performed better than the baselines Unet but was worse than the UUnetAutoencoder. Lastly, in all the metrics, it has the lowest result, indicating that the traditional model Unet is comparatively worse in classification performance. Therefore, among the four models, the most robust and efficient is UUnetAutoencoder.

D. Statistical Analysis

Statistical analysis is critical in assessing the performance of deep learning models, particularly adversarial robustness in AR/VR systems. Statistics gives a reliable method for measuring a model's performance under normal and hostile conditions. To confirm findings, statistical tests are conducted to ascertain whether the discrepancies in model outputs are indeed noteworthy or merely a result of chance.
We will use mean accuracy, standard deviation, p-values, and confidence intervals as statistical metrics to compute the F1 score, precision, recall, and classification accuracy. Statistical comparison between the models tested both with and without adversarial attacks will help us explain how adversarial examples influence the robustness of AR/VR systems. We can use statistical analysis to analyze the generalization skills of models and find possible means of increasing their robustness to adversarial conditions. This technique makes the data statistically valid and helpful in designing a more secure and efficient system of AR/VR.

Descriptive Statistics

TABLE II. DESCRIPTIVE STATISTICS

		Mean (Std. Deviation)
UUnetAuto encoder	Accuracy	0.98(0.0020)
	Precision	0.98 (0.0019)
	Recall	0.97 (0.0028)
	F1-score	0.96 (0.0015)
	Total	0.97 (0.0084)
UnetAuto encoder	Accuracy	0.94 (0.0019)
	Precision	0.94 (0.0020)
	Recall	0.93 (0.0030)
	F1-score	0.92 (0.0024)
	Total	0.93 (0.0090)
Unet Attention mechanism	Accuracy	0.93 (0.0026)
	Precision	0.93 (0.0021)
	Recall	0.92 (0.0023)
	F1-score	0.92 (0.0027)
	Total	0.93 (0.0053)
Unet	Accuracy	0.91 (0.0019)
	Precision	0.91 (0.0023)
	Recall	0.90 (0.0023)
	F1-score	0.90 (0.0027)
	Total	0.91 (0.0058)

Concerning Accuracy, Precision, Recall, and F1-score, descriptive statistics (Table 2) compare four models: UnetAutoencoder, Unet with Attention Mechanism, Unet, and Proposed. Regarding reliability and precision, the UUnetAuto encoder model has the highest mean values for accuracy, recall, precision, and F1-score. With an F1-score of 0.900 and an Accuracy of 0.911, the Unet model underperforms the Proposed model. The models Unet with Attention Mechanism and Unet Autoencoder both perform moderately. While Unet with Attention

Mechanism achieves 0.929 and 0.921, Unet Autoencoder achieves 0.941 Accuracy and 0.920 F1-score. These models are superior to Unet, but they can't match the recall and accuracy of the UUnetAuto encoder model.

Anova Results

TABLE III: ANOVA

	F	Sig.
UUnetAuto encoder	197.791	0.000
Unet Autoencoder	179.469	0.000
Unet Attention mechanism	50.128	0.000
Unet	69.194	0.000

The ANOVA results(Table-3) validate the statistical significance of performance disparities among the models, with all models exhibiting p-values (Sig.) < 0.001. The UUnetAuto encoder model exhibits the highest F-value of 197.791, signifying the most significant variation among groups, succeeded by the UnetAutoencoder (F = 179.469), Unet (F = 69.194), and Unet with Attention Mechanism (F = 50.128). The Proposed model regularly surpasses the alternatives, exhibiting statistically significant differences across the evaluation metrics.

E. Discussion

Considering both the standard and adversarial settings, the work that falls within the purview of this research study incorporates the testing of deep learning models on adversarial resilience in the systems that implement augmented reality and virtual reality applications. Using performance measures ascertains how these models can resist hostile shocks, the confusion matrix, ROC curve analysis, and statistical testing. According to the findings, the performance of deep learning models is shown to deteriorate when they are subjected to adversarial assaults. Those kinds of deep learning models that appear to be helpful in such applications like augmented reality and virtual reality because they have a real need for a very high processing rate and good precision prediction of such scenarios raise concerns in such assaults, as revealed by the confusion matrix, indicating the adverse effect of this perturbation type on model's ability by increasing the occurrence of misclassification cases and subsequently its accuracy and dependability. Hostile samples are likely to weaken models that work well in everyday settings. As such, deep learning models demonstrate that they can differentiate between other data classes in everyday and hostile situations according to the confusion matrix.

The introduction of hostile cases increased the number of false positives and false negatives, which indicates that the models had a hard time maintaining their accuracy in predictions. The model needs stronger defensive mechanisms, such as adversarial training, or robust optimization, such as classification. When the ROC curve was examined, the degree of sensitivity and specificity the models have in their categorization techniques came out. Even the adversarial approaches were able to bring the ROC curve down regardless of how well the models were doing under normal circumstances. The models tested on adversarial instances have much lower AUC values than those tested without examples. These augmented and virtual reality applications need high-confidence predictions in their output for applications in navigation, object identification, and interactive experiences. The presence of adversarial perturbations reduces the confidence of a model with its forecast. Statistical analysis can evaluate the impact adversarial perturbations have on the performance of models and verify the dependability of the outcomes using mean accuracy, standard deviation, p-values, and confidence intervals. Statistical analysis shows that the accuracy has decreased with an increase in the amount of variation in the adversarial scenarios. This indicates that the models do not have the resilience to resist the perturbations used in this study. Adversarial robustness procedures such as adversarial training, defensive distillation, and robust optimization should be used according to the statistical data. Based on the research results, the models, which have been trained through adversarial training or the robust loss function, perform better in adverse settings. Conversely, they underperform in non-adverse settings but show higher defense capabilities against adverse attacks. This shows how crucial it is to have defensive mechanisms on deep learning models and deploy them to restrict hostile scenarios and keep them on in real-world applications of augmented reality and virtual reality. According to the confusion matrix, the ROC curve, and the statistical analysis, the deep learning models used in augmented reality and virtual reality applications are highly affected by adversarial attacks. Deep learning models are fabulous in controlled environments, and it is unsettling that they are very prone to adversarial perturbations, especially where their application highly depends on safety. From the statistics, a more significant move toward developing more robust models and defenses is inevitable, and only through this can one see how to solve the adversarial robustness problem. To make the deep learning model strong enough, adversarial attacks

must be designed to be more resilient so that augmented and virtual reality systems can operate safely and reliably under real-world conditions.

The current work lays a foundational framework for adversarial robustness in AR/VR systems using a Double U-Net Autoencoder, we acknowledge the necessity of refining and concretizing the architecture further. This study demonstrates the initial success of the UUNet-based Autoencoder in mitigating adversarial perturbations through reconstruction-based defense and classification via a pre-trained VGG-16. However, future enhancements will target:

- Integration with real-time AR/VR platforms to validate low-latency performance.
- Adaptability to new attack types, such as generative adversarial attacks and physical-world adversaries.
- Hybrid defense mechanisms, combining adversarial training and feature-squeezing methods to enhance generalizability.
- Hardware optimization, especially for AR/VR devices with computational constraints (e.g., embedded GPUs in headsets).
- User-level perceptual validation, ensuring output integrity not just algorithmically, but through user feedback in immersive environments.

This trajectory will help transition the system from a research prototype to a deployable, real-time defense module in AR/VR systems.

We Expanded with precise dataset-wise confusion matrix interpretation.

ROC curve analysis is now contextualized per dataset (e.g., Cityscapes, COCO, CIFAR-10).

Statistical significance is addressed using ANOVA and descriptive statistics, confirming the robustness of our model against others (e.g., Unet with Attention).

V. Conclusion

This work explored the augmentation of adversarial robustness of augmented reality and virtual reality that a UUNet-based autoencoder can achieve. The model's performance was evaluated based on the analysis of the confusion matrix, ROC curve metrics, and statistical tests under standard and adversarial conditions. The results gathered above depict the pros and cons of the model based on its ability to still work correctly in challenging situations. In standard data, the confusion matrix analysis showed that the UUNet-based Autoencoder had a very high actual positive rate but a very low false positive rate. This was depicted by the fact that the confusion matrix was analyzed. This deteriorates when the model faces hostile inputs and produces a remarkably higher number of false positives as well as false negatives. Because of this deterioration, the model is susceptible to hostile disturbances, further reducing the anticipated dependability.

As these data illustrate, strong defensive mechanisms are needed for an autoencoder to survive long. A study based on the ROC curve indicated that the classification performance degraded when subjected to circumstances of adversarial attack. The firm model discrimination was demonstrated by the high AUC score of the Autoencoder, which was based on UUNet when it was used with the standard parameters. Even though adversarial assaults decreased the area under the curve (AUC) score, which indicated a reduction in model confidence, it was proved that the model's performance was competitive with other baseline designs and was relatively resilient. When evaluating the UUNet-based Autoencoder, accuracy, precision, recall, and F1-score were all considered. Under typical circumstances, the model learnt without problems and achieved the highest possible results. Because these measurements were decreased when adversarial assaults were utilized, it can be concluded that adversarial noise impacts the classification outcomes. The model's performance is satisfactory under controlled settings; nevertheless, it requires more refinement before it can be used in real-world situations. In terms of numerical value, statistical analysis was used to quantify the changes in the model's performance. It was demonstrated by using p-values, mean accuracy, and standard deviation that the downward trends in adversarial performance are statistically significant. The existence of statistically significant disparities between the adversarial' scenarios and standards provides a rationale for implementing targeted robustness measures. The confidence intervals of the performance measures not only demonstrated that the model was accurate in circumstances involving standards but also revealed that it was vulnerable in instances involving adversarial situations. The confusion matrix and the ROC analysis demonstrate that the UUNet-based Autoencoder functions adequately under standard settings, exhibiting high classification accuracy and resilience. In addition to lowering the performance metrics, area under the curve (AUC) scores, and classification reliability numerically, adversarial scenarios expose it to serious faults. The presence of results of this sort indicates that the model is susceptible to adversarial perturbations. As a result, the model requires robust defensive mechanisms, such as adversarial training or robust optimization, to become resilient.

REFERENCES

[1] D. I. Chi, R. Korea, M. Mcgill, D. Boland, S. Brewster, and R. Murray-smith, 'McGill, M Boland, D ., Murray-Smith, R ., and Brewster, S . (2015) A Dose Copyright © 2015The Authors A copy can be downloaded for personal non-commercial research or study, without prior permission or charge Content must not be changed in any ', no. May, pp.2143–2152, 2015.

[2] J. Guna, G. Geršak, I. Humar, J. Song, J. Drnovšek, and M. Pogačnik, 'Influence of video content type on users' virtual reality sickness perception and physiological response',Futur. Gener. Comput. Syst., vol. 91, pp. 263–276, 2019, Doi: HTTPs://doi.org/10.1016/j.future.2018.08.049.

[3] J. D. Azofeifa, J. Noguez, S. Ruiz, J. M. Molina-espinosa, A. J. Magana, and B. Benes,'Systematic Review of Multimodal Human – Computer Interaction', pp. 1–32, 2022.

[4] G. R. Silva, J. C. Donat, M. M. Rigoli, F. R. de Oliveira, and C. H. Kristensen, 'Aquestionnaire for measuring presence in virtual environments: factor analysis of the presence questionnaire and adaptation into Brazilian Portuguese', Virtual Real., vol. 20,Page 33 of 37 - AI Writing Submission Submission ID trn:oid:::28727:418472593 Page 33 of 37 - AI Writing Submission Submission ID trn:oid:::28727:418472593 no. 4, pp. 237–242, 2016, doi: 10.1007/s10055-016-0295-7.

[5] M. Slater, 'Immersion and the illusion of presence in virtual reality', Br. J. Psychol., vol.109, no. 3, pp. 431–433, 2018, doi: 10.1111/bjop.12305.

[6] B. G. Witmer and M. J. Singer, 'Measuring Presence in Virtual Environments: A Presence Questionnaire', Presence Teleoperators Virtual Environ., vol. 7, no. 3, pp. 225–240, Jun.1998, doi: 10.1162/105474698565686.

[7] Dr. Jagreet Kaur Gill, 'Robustness and Adversarial Attacks in Computer Vision',XenonStack.

[8] Sik-Ho Tsang, 'DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation', Medium.

[9] I. J. Goodfellow, J. Shlens, and C. Szegedy, 'Explaining and harnessing adversarial examples', 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–11,2015.

[10] J. H. Metzen, M. C. Kumar, T. Brox, and V. Fischer, 'Universal Adversarial Perturbations Against Semantic Image Segmentation', Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, pp. 2774–2783, 2017, doi: 10.1109/ICCV.2017.300.

[11] A. Bajaj et al., 'Double-U Net: A Novel Approach for Deep Learning Based Image Defogging', Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 16, no. 3, pp. 535–546,2024.

[12] A. M. Logan Engstrom, Brandon Tran, Dimitris Tsipras, Ludwig Schmidt, 'A ROTATION AND A TRANSLATION SUFFICE: FOOLING CNNS WITH SIMPLE TRANSFORMATIONS', vol. 10, 2019.

[13] J. Su, D. V Vargas, and K. Sakurai, 'One Pixel Attack for Fooling Deep Neural Networks', IEEE Trans. Evol. Comput., vol. 23, no. 5, pp. 828–841, 2019, doi:10.1109/TEVC.2019.2890858.

[14] M. Ahmed, A. F. Ashrafi, R. Ahmed, S. Shatabda, A. K. M. Islam, and S. Islam, 'DoubleU-NetPlus: a novel attention and context-guided dual U-Net with multi-scale residual feature fusion network for semantic segmentation of medical images', Neural Comput. Appl., vol. 35, Mar. 2023, doi: 10.1007/s00521-023-08493-1.Page 34 of 37 - AI Writing Submission ID trn:oid:28727:418472593 Page 34 of 37 - AI Writing Submission ID trn:oid:::28727:418472593

[15] H. V Vo, H. P. Du, and H. N. Nguyen, 'APELID: Enhancing real-time intrusion detection with augmented WGAN and parallel ensemble learning', Comput. Secur., vol. 136, p.103567, 2024, doi: https://doi.org/10.1016/j.cose.2023.103567.

[16] B. Zhao, Z. Li, C. Wu, X. Zhang, and Z. Du, 'Research on abnormal object detection network of computer room inspection robot based on depth vision', Meas. Sci. Technol.,vol. 35, no. 12, p. 126017, 2024, doi: 10.1088/1361-6501/ad7f7a.

[17] V. De Martino and F. Palomba, 'Classification, Challenges, and Automated Approaches to Handle Non-Functional Requirements in ML-Enabled Systems: A Systematic Literature Review', 2023.

[18] Y. Zhang, C. Slocum, J. Chen, and N. Abu-Ghazaleh, 'It's all in your head(set): Sidechannel attacks on AR/VR systems', in 32nd USENIX Security Symposium (USENIX Security 23), Anaheim, CA: USENIX Association, Aug. 2023, pp. 3979–3996.

[19] P.-H. Lin and S.-Y. Chen, 'Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking', IEEE Access, vol. 8, pp. 45689–45699, 2020, doi:10.1109/ACCESS.2020.2977679.

[20] A. Apicella et al., 'Enhancement of SSVEPs Classification in BCI-Based Wearable Instrumentation Through Machine Learning Techniques', IEEE Sens. J., vol. 22, no. 9, pp. 9087–9094, 2022, doi: 10.1109/JSEN.2022.3161743.

[21] S.-J. Yoo and S.-H. Choi, 'Indoor AR Navigation and Emergency Evacuation System Based on Machine Learning and IoT Technologies', IEEE Internet Things J., vol. 9, no.21, pp. 20853–20868, 2022, doi: 10.1109/JIOT.2022.3175677.

[22] Y. Tai, B. Gao, Q. Li, Z. Yu, C. Zhu, and V. Chang, 'Trustworthy and Intelligent COVID-19 Diagnostic IoMT Through XR and Deep-Learning-Based Clinic Data Access', IEEE Internet Things J., vol. 8, no. 21, pp. 15965–15976, 2021, doi:10.1109/JIOT.2021.3055804.

[23] 'Data Cleaning in Eye Image Dataset Project report submitted in partial fulfilment of the requirement for the degree of Bachelor of Technology in Computer Science and Engineering / Information Technology By Harshit Singh (191264) & Archit Dogra (Page 35 of 37 - AI Writing Submission ID trn:oid:28727:418472593 Page 35 of 37 - AI Writing Submission Submission ID trn:oid:::28727:418472593191277'.

[24] A. Joshi, N. Pandey, A. Juyal, D. Pandey, and V. S. Thapli, ‘DDCMR2: A Deep Detection and Classification Model with Resizing and Rescaling for Plant Disease BT – Artificial Intelligence: Theory and Applications’, H. Sharma, A. Chakravorty, S. Hussain, and R.Kumari, Eds., Singapore: Springer Nature Singapore, 2024, pp. 217–230.
[25] C. Shorten and T. M. Khoshgoftaar, ‘A survey on Image Data Augmentation for Deep Learning’, J. Big Data, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.
[26] C. Wu, L. Zhang, B. Du, H. Chen, J. Wang, and H. Zhong, ‘UNet-Like Remote SensingChange Detection: A review of current models and research directions’, IEEE Geosci.Remote Sens. Mag., vol. 12, no. 4, pp. 305–334, 2024, doi:10.1109/MGRS.2024.3412770.
[27] S. Yanes Luis, N. Basilico, M. Antonazzi, D. Gutiérrez Reina, and S. Toral Marín, ‘Deep Variational Auto-Encoder for Model-Based Water Quality Patrolling with Intelligent Surface Vehicles BT - Advances in Artificial Intelligence’, A. Alonso-Betanzos, B.Guijarro-Berdiñas, V. Bolón-Canedo, E. Hernández-Pereira, O. Fontenla-Romero, D. Camacho, J. R. Rabuñal, M. Ojeda-Aciego, J. Medina, J. C. Riquelme, and A. Troncoso,Eds., Cham: Springer Nature Switzerland, 2024, pp. 19–28.
[28] L. Huang, A. Miron, K. Hone, and Y. Li, ‘Segmenting Medical Images: From UNet to Res-UNet and nnUNet’, in 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS), 2024, pp. 483–489. doi: 10.1109/CBMS61543.2024.00086.
[29] Y. Yuan and Y. Cheng, ‘Medical image segmentation with UNet-based multi-scale context fusion’, Sci. Rep., vol. 14, no. 1, p. 15687, 2024, doi: 10.1038/s41598-024-66585-x.
[30] J. Tian et al., ‘LESSON: Multi-Label Adversarial False Data Injection Attack for Deep Learning Locational Detection’, IEEE Trans. Dependable Secur. Comput., vol. 21, no. 5, pp. 4418–4432, 2024, doi: 10.1109/TDSC.2024.3353302.
[31] J. Kim, Z. Zhu, T. Bau, C. Liu, and D. Lab, ‘DCDR-UNet: Deformable Convolution Based Detail Restoration via U-shape Network for Single Image HDR Reconstruction’,Proc. IEEE/CVF Conf. Comput. Vis. Pattern RecognitionProceedings IEEE/CVF Conf.Page 36 of 37 - AI Writing Submission Submission ID trn:oid:28727:418472593 Page 36 of 37 - AI Writing SubmissionSubmission ID trn:oid:::28727:418472593 Comput. Vis. Pattern Recognit. Work., pp. 5909–5918, 2024.
[32] S. Kaviani, A. Sanaat, M. Mokri, C. Cohalan, and J.-F. Carrier, ‘Image reconstruction using UNET-transformer network for fast and low-dose PET scans’, Comput. Med.Imaging Graph., vol. 110, p. 102315, 2023, doi:https://doi.org/10.1016/j.compmedimag.2023.102315.