www.itim.unige.it

Genuense Athenaeum

Università degli Studi di Genova

Scuola Politecnica & Facoltà di Ingegneria

www.itim.unige.it/strategos

Laurea Magistrale in Ingegneria Strategica

International MSc in Engineering Technology for Strategy & Security

of Genoa University

Study and Design of Generative Learning tools for Threat Assessment in Defence

Advisor: **Prof. Agostino G. Bruzzone, Simulation Team**

Co-Advisors: **Eng. Umberto Battista, Dr. Federico Taddei Santoni**

Candidate: **Dr. Carolina Badano**

Academic Year 2023/24

STRATEGOS 4th Edition

Acknowledgments

To Professor Agostino G. Bruzzone, who gave me the opportunity to undertake the STRATEGOS master degree two years ago and today, as my supervisor, provided me the necessary support to successfully complete this journey.

To Federico and Umberto, my corporate supervisors, for supporting and advising me with patience and dedication since my first steps in the professional world.

To my Family, especially my mother, who, through her example, encouraged me to believe in myself and achieve my goals, no matter how ambitious.

To Gianluca, always by my side, in moments of joy as well as in times of distress.

To Dr. G. B. Crosa di Vergagni, for welcoming me like a granddaughter and offering me the opportunity to stay in Genoa at the facility he manages.

To my grandparents, my dearest memories; specifically to Nonno Franco, Master of Work and Master of Life. I will always have you by my side, wherever you are, just as you have been with me all these years, teaching me, through your life, the values that are most important to me: education, determination, and willpower. I hope I have made you proud on this day, we both so desired and pursued.

Ringraziamenti

Al Professor Agostino G. Bruzzone, il quale due anni fa mi diede la possibilità di intraprendere il corso di studi STRATEGOS ed oggi, in qualità di mio relatore, il sostegno necessario per concludere con successo tale percorso.

A Federico ed Umberto, miei relatori aziendali, per avermi sostenuta e consigliata, con pazienza e dedizione, sin dai miei primi passi nel mondo del lavoro.

Alla mia Famiglia, in particolare a mia madre, che, con l'esempio, mi ha spronata a credere in me stessa e raggiungere i miei obiettivi, pur ambiziosi.

A Gianluca, sempre vicino, nei momenti di gioia come di sconforto.

Al Dott. G. B. Crosa di Vergagni, per avermi accolta come una nipote, offrendomi la possibilità di soggiornare a Genova presso la struttura da lui diretta.

Ai miei nonni, i miei ricordi più cari; nello specifico al Nonno Franco, Maestro del Lavoro e Maestro di Vita. Ti avrò sempre accanto, ovunque tu sia, proprio come mi sei stato in questi anni, insegnandomi, attraverso la tua vita, i valori per me più importanti: l'educazione, la determinazione e la volontà. Spero di averti reso fiero in questo giorno, da noi due così desiderato e perseguito.

…ho scalato le montagne,
immagino che hai sorriso…

Abstract

Artificial Intelligence (AI) is one of the most important and rapidly developing emerging trends, offering unique benefits through the development of computer systems capable of performing tasks that typically require human intelligence. AI applications span many sectors, enabling machines to analyse vast amounts of data, identify trends and make predictions, thereby improving decision-making, productivity and user experience. However, a significant challenge remains: the quality and scarcity of data, particularly in the military sector, due to confidentiality issues, technological limitations and the operational environment.

This Thesis aims to analyse and address the concerns of frugal and robust AI in the military domain. Emphasising the growing importance of data-centric AI, this research evaluates innovative approaches and explores techniques for generating high-resolution synthetic imagery and applies them to AI-based threat assessment, decision support and soldier training.

The dissertation proposes two main approaches to address image scarcity and quality issues for effective threat assessment using synthetic data. The first approach is to increase the resolution of low-quality images using generative AI, and the second is to generate realistic synthetic scenarios using stable diffusion methods. The research was carried out during an internship at Stam S.r.l., a company active in this field.

The Thesis highlights the centrality of high-quality data in the development of efficient AI-based methods. Moreover, the results demonstrate the impressive contributions that AI can bring to the military, particularly in supporting decision making through autonomous threat assessment methods.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence	NF	Normalizing Flow
ANN	Artificial Neural Network	NN	Neural Network
C2	Command and Control	NVAE	Nouveau VAE
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance	OE	Operational Environment
CBRN	Chemical, Biological, Radiological, and Nuclear	OSINT	Open-Source Intelligence
CLIP	Contrastive Language-Image Pre-training	PSNR	Peak Signal-to-Noise Ratio
CNN	Convolutional Neural Network	RCL	Recurrent Convolutional Layer
CVAE	Conditional Variational Autoencoder	RCNN	Recurrent Convolutional Neural Network
DGL	Deep Generative Learning	ROS	Random OverSampling
DGP	Deep Gaussian Process	RRDB	Residual-in-Residual Dense Block
DNN	Deep Neural Network	RSTB	Residual Swin Transformer Blocks
ERSGAN	Enhanced Super-Resolution Generative Adversarial Network	RaGAN	Relativistic Average GAN
GA	Genetic Algorithm	RoI	Region of Interest
GAN	Generative Adversarial Network	SDXL	Stable Diffusion XL
GFPGAN	Generative Facial Prior-Generative Adversarial Network	SISR	Single Image Super-Resolution
GL	Generative Learning	SMOTE	Synthetic Minority Oversampling Technique
GMM	Generative Moment Matching	SRGAN	Super-Resolution GAN
GMMN	Generative Moment Matching Network	SUPIR	Super-Resolution Perceptual Image Restoration
GN	Group Normalization	SVM	Support Vector Machine
HCP	Hypotheses-CNNPooling	SWINIR	Swin Transformer for Image Restoration
HFT	Hypotheses-fine-tuning	UAN	Universal Adaptation Network
ISR	Intelligence, Surveillance, and Reconnaissance	UDA	Unsupervised Domain Adaptation
KL	Kullback-Leibler	VAE	Variational Autoencoder
LIIF	Local Implicit Image Function	WGAN	Wasserstein GAN
ML	Machine Learning	ZPD	Zone of Proximal Development
NAP	Neural Abstraction Pyramid		

Table 1: List of Acronyms

1. Introduction

In today's world, one of the most important subjects is Artificial Intelligence. It is an increasingly important topic that has developed exponentially in recent years and offers unique advantages. AI refers to the development of computer systems capable of performing tasks that typically require human intelligence. These tasks include understanding natural language, recognising patterns, learning from experience, and making decisions. Its applications extent many sectors. With AI, machines can analyse vast amounts of data, identify trends, and make predictions, leading to improved decision-making, increased productivity and enhanced user experiences. Thus, one of the main requirements of AI are a multitude of real datasets. Despite technological advantages lead to a large and growing number of datasets, especially in the military sector, one of the main challenges is the low quality and scarcity of data.

The aim of this Thesis, which focuses on the military domain, is first to analyse and then to address frugal and robust AI and its consequences; indeed, very innovative approaches have been tested and evaluated since decades thanks to extensive use of Simulation [1]. The Thesis examines various methods and show promise for both increasing the amount of data available through generation methods and improving its quality. Today, the concept of data-centricity is growing exponentially, as AI-based models only get better if they are trained on high quality datasets. This Thesis attempts to give meaning to data by first focusing on methods to increase the quality and quantity of data, and then performing threat assessment on these synthetic images to support decision making and soldier training. In fact, A-based threat assessment relies on the analysis of images, both real and synthetic, to identify and classify potential risks, and so it requires a lot of good data on which to train and then perform the analysis.

The Thesis proposes two possible approaches to address the problems of image scarcity and low quality and then to perform threat assessment based on the synthetic data. One approach focuses on increasing the resolution of low-quality images using generative AI. Segmentation and labelling methodologies are then performed to improve the performance of the generative AI and to detect threats based on the increased resolution data. The second approach focuses on generating synthetic realistic scenarios through the stable diffusion method. Then different

threat levels are attributed to the generated scenarios by experienced soldiers and then a classifier is trained to perform risk assessment.

The work reported in this Thesis has been conducted during the internship period within the company Stam S.r.l., which is currently involved in several activities in this domain.

The Thesis demonstrates the impressive contributions that AI can bring to the military, particularly in supporting decision-making through autonomous methods of threat assessment. These methods can assist soldiers during training and in real time. This is only possible by overcoming the challenges of low and scarce data through frugal and robust AI, as the proposed approaches in this Thesis have shown, focusing on the central aspect of data centricity, thus giving importance to data in order to have efficient AI-based methods.

1.1 Artificial intelligence overview

Artificial Intelligence corresponds to the engineering and scientific methodologies that result in a system which poses intelligent reasoning and behaviour. Recent technological advances have established AI as a technology with impressive impact on almost every aspect of the modern socio-economic environment, thus significantly enhancing the capabilities of, among others, decision-making and support systems and autonomous processes based on various types of data. Modern Machine Learning (ML) and Deep Learning (DL) techniques rely on mathematical models that seek to extract and exploit information from a huge dataset that is relevant to a given application. By identifying correlations between complicated data, AI systems are able to automatically understand patterns and structures. [2]

Specifically, AI is the simulation of human intelligence processes by machines, especially computer systems. AI requires a foundation of specialized hardware and software for writing and training ML algorithms [3]. AI systems work by ingesting large amounts of labelled training data, analysing the data for correlations and patterns, and using these patterns to make predictions about future states. In this way, a chatbot that is fed examples of text can learn to generate lifelike exchanges with people, or an image recognition tool can learn to identify and describe objects in images by reviewing millions of examples. AI programming focuses on cognitive skills that include:

- Learning, AI programming focuses on acquiring data and creating rules for how to turn it into actionable information. The rules, which are called algorithms, provide computing devices with step-by-step instructions for how to complete a specific task.
- Reasoning, AI programming aims on choosing the right algorithm to reach a desired outcome.
- Self-correction, AI programming is designed to continually fine-tune algorithms and ensure they provide the most accurate results possible.
- Creativity, AI uses Neural Networks (NN), rules-based systems, statistical methods and other AI techniques to generate new images, new text, new music and new ideas. [2]

When the tasks to be performed comes to be repetitive and detail-oriented, AI can perform them better than humans only if it can access to large available data and information easily than humans. [2]

AI should not be viewed in isolation; rather, it should be seen as a supporting technology for existing functional applications. It relies on carefully crafted algorithms designed to address specific challenges, encompassing tasks such as data collection, organization, processing, analysis, transmission, and responsive actions, especially when dealing with extensive datasets. These AI algorithms can align with and approximating the cognitive capabilities of the human mind, effectively enhancing and complementing various operational aspects. [4]

AI brings several advantages. It excels in detail-oriented tasks, significantly reduces time for data-heavy processes, saves labour, and ensures consistent results. AI's ability to personalize customer experiences and the availability of AI-powered virtual agents contribute to increased satisfaction and improved service accessibility. However, challenges exist. Developing AI is expensive and demands technical expertise, creating a shortage of qualified professionals. AI systems may also perpetuate biases present in training data, limiting their fairness. Generalization across tasks remains a hurdle, and the widespread adoption of AI raises concerns about job displacement, necessitating proactive measures for workforce adaptation. [2]

In navigating these challenges, responsible and ethical approaches are crucial to harness the benefits of AI while addressing its drawbacks. Achieving this balance ensures that AI contributes positively to society and aligns with human values.

Over the past seven decades, dating back to its inception, the field of AI has evolved through three distinct developmental phases. In the initial stage, the primary focus was on rule-based methods, employing tools such as decision trees, Boolean logic, and fuzzy logic, collectively known as expert systems. Subsequently, the second phase witnessed a pivotal shift towards the development and application of statistical techniques, giving rise to the concept and methodology of ML. During this stage, notable achievements were made, including the development of e-mail spam filters and internet search engines. The third and current developmental stage introduces the adoption of human-like learning techniques, prominently featuring NNs and defining the concept and technology of DL. [4]

AI can be subdivided in four main categories based on their functions and capabilities: reactive machines, limited memory, theory of mind, and self-aware. In addition to the four main types, AI can be further classified into three broader categories: narrow intelligence, general intelligence, and superintelligence. Reactive machines respond to prompts without the use of memory or a broader understanding of context. They are often used in game design to create opponents and are commonly seen in marketing tools, such as chatbots. These AI systems can analyse customer behaviour and market trends to optimize marketing campaigns in real time. Limited memory AI can learn from a limited amount of data or feedback but does not retain long-term memories. For example, ChatGPT has a token limit and cannot recall earlier parts of a conversation beyond that limit. In marketing, limited memory AI is used to analyse data, make predictions, and offer recommendations. However, the accuracy of its output depends on the quality of the input data. Theory of mind AI is a concept that represents an advanced class of technology. It aims to create AI systems that understand the mental states of humans. While it has potential for enhancing human-AI interactions, it is still in its early stages of development and implementation. Self-aware AI represents the next evolutionary step beyond theory of mind, where machines not only understand human emotions and mental states but also have their own emotions, needs, and beliefs. However, self-aware AI only

exists as a hypothetical concept in science fiction and is not a reality at this time. These categories can be further simplified as follows[4]:

- Narrow Intelligence: Includes reactive machines and limited memory AI. These systems are specialized and can't perform beyond their programmed capabilities.
- General Intelligence: Represents the concept of AI with human-like understanding, adaptability, and reasoning. Theory of mind and self-aware AI fall under this category, although they are still under development or purely hypothetical.
- Superintelligence: This is a theoretical concept where AI surpasses human intelligence by a significant margin. It is a subject of speculation and debate in the field of AI. [5]

1.2 AI in military applications

Modern military systems are becoming increasingly complex because of the variety of challenges they must contend with. To balance this complexity and support operations in a demanding and dynamic Operational Environment (OE) new skills, capacities, and novel technologies are needed. It is crucial to exploit multimodal data fusion techniques to extract relations between the different data source in the systems that incorporate frameworks for improved situation awareness and intelligence operations [6], multiple sources and types of information need to be employed and processed when frugality in data is present. Especially in military operations raw data has relatively limited utility, so it’s important to relate them with other information about the OE and consider the past experiences [7]. Mission planning and operational Command&Control (C2) in complicated scenarios with little data, when decision-making in real time or a limited response timeframe are a mandatory request, are another key concern during military operations. [21] Thus, it is evident the importance of using AI in this context to evolve towards new C2 paradigms and achieve new capabilities in this field such as agility for challenging scenarios [8].

From the military point of view, the AI is defined as the capability of computer systems to perform tasks that normally require human intelligence like perception, conversation and decision-making. Moreover, AI is able to incorporate abstraction and interpretation into

information processing and make decisions at a level of sophistication that would be considered intelligent in humans. [21]

Therea have been multiple successful uses of AI in Defence to address different issues such as Training and Scenario Development in combination with several Digital Solutions [9]. Indeed, nowadays the advances in AI covers many new opportunities addressing major challenges [10]; for instance Autonomous Systems are a crucial source of data for AI as well as an asset that need intelligence to operate remotely under supervision with large challenges due to latency, communication, etc. [11]. It should be outlined that moving to multi domain and multi dimension scenarios, especially touching new emerging aspects such integration of space and cyber space it turns crucial the use of AI [12].

Indeed, AI can be defined as the simulation of human intelligence on a machine, enabling it to efficiently utilize knowledge to solve problems. It serves as an comprehensive term encompassing various disciplines such as ML, computer vision, natural language processing, DL, and cognitive computing. Each of these disciplines has distinct characteristics that set them apart. [22]

AI holds a significant position within the military domain, as it empowers computer systems to emulate human intelligence, effectively tackling tasks that encompass perception, decision-making, and conversational capabilities. By harnessing AI's capabilities, the military can enhance the efficiency of its operations, bolstering everything from autonomous drone missions to advanced data analysis, ultimately leading to more informed and timely decisions in complex military scenarios. Furthermore, AI serves as an ever-evolving, multifaceted field encompassing specialized disciplines which collectively contribute to bolstering the military's technological prowess and strategic advantage. [21]

It should be outlined, that AI has a crucial role in addressing Human Behaviours in specific operations such as CIMIC and Psyops [13].

Asymmetric threats and defence against terrorism are other fields that propose challenges requiring use of AI and several researches have been successfully conducted in this field [14] Obviously, the great potential of AI have been demonstrated also in Logistics to redesign and

optimize Logistics [15]. Indeed, modern operations relies on strong interoperability and this capability requires the necessity to create complex scenario where AI has a crucial role to play different parties, entities and units that are necessary to test these capabilities at Strategic, Operational and Tactical Levels [16].

1.2.1 Today's situation

In recent years, AI has gained significant attention due to its civilian and military applications. Modern warfare is increasingly relying on Artificial Intelligence. Military systems involving AI can manage greater amounts of data more effectively than conventional systems. Additionally, due to its intrinsic computation and decision-making powers, AI improves the self-control, self-regulation, and self-actuation of combat systems. AI is deployed in almost every military application, and increased research and development funding from defence research agencies to develop new and advanced applications of AI is projected to drive the increased adoption of AI-driven systems in the defence sector. AI is expected to contribute to different military application: autonomous warfare platforms, cybersecurity, logistic and transportation, target acquisition, battlefield information processing, predictive maintenance.

NATO member states have recognized this potential and have started investing in AI, incorporating it into their defence strategies. The strategic importance of AI has led nations to develop it for various military functions, including Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR), cyber operations, and the deployment of autonomous and semi-autonomous machines. This technology can significantly contribute to the acquisition and processing of Open-Source Intelligence (OSINT) by acting as the initial filter for vast amounts of internet-available data, which can later be evaluated by personnel. By incorporating AI into the decision-making process at all levels of warfare (tactical, operational, and strategic), leaders can benefit from calculated input that is unbiased by emotions and other factors that can affect human judgment. This can provide valuable support in making well-informed decisions. [23]

AI holds significant applications across a spectrum of domains within the military field. AI can enhance the coordination, situational awareness, and performance of land, naval, air, and

space weapon systems. It can also reduce maintenance and operational costs of them. Moreover, AI can autonomously protect networks, computers, software, and data against unauthorized access. AI-based systems can detect patterns in cyber-attacks, learn from past events, and develop appropriate countermeasures with higher speed and effectiveness than humans. Additionally, AI can accurately identify targets for unmanned aerial vehicles, cruise missiles, and other systems. It contributes to intelligence, surveillance, and reconnaissance (ISR) operations, providing improved threat monitoring and real-time analysis of unstructured data. AI enhances simulation and training through systems engineering, software, and computer engineering. It facilitates multidisciplinary training by providing realistic scenarios, intelligent adversaries, and personalized feedback. Talking about logistic aspects, AI can improve efficiency and timeliness in transport activities, reduce costs, and enhance maintenance effectiveness. It enables just-in-time delivery, rapid identification of potential malfunctions, and anticipation of failures or component replacements. Also, the medical sector can benefit from AI, indeed robotic surgical systems and platforms can be integrated with AI to enable remote medical assistance, complex diagnosis, correlation with medical histories, and recommendation of treatments. Furthermore, at the political and strategic level, AI can be used to destabilise an adversary by producing and publishing massive amounts of false information. In this case, AI is also likely to be the best candidate to defend against such attacks. [23]

In the military, decisions often have life-or-death consequences. Massive volumes of data may be collected, stored, and analysed using Big Data technologies from a variety of sources, such as sensors, satellites, drones, and intelligence reports. This data is processed by AI systems to produce useful insights, enabling military leaders to make choices in real time. Moreover, nowadays cybersecurity is paramount for military operations. AI assists in identifying and mitigating cyber threats, while Big Data analytics detect patterns of cyberattacks. Big Data technology optimizes logistics and supply chains in the military, ensuring that troops have the necessary equipment and resources at the right time and place. AI-driven algorithms improve resource allocation and distribution. Additionally, AI and Big Data are used to create realistic training simulations for military personnel. These simulations help troops prepare for real-world

scenarios and improve their decision-making skills. AI and Big Data are employed to enhance communication networks, making them more resilient and secure. So, the fusion of AI and Big Data is enhancing military capabilities and effectiveness, because their integration drives transformative changes in the way armed forces operate, plan, and make decisions. [4] In military operations, reliable and secure communication is vital for coordination and information sharing.

A human defence planner is limited by their cultural background, education, and personal experiences, which are specific to a particular society and historical context. Additionally, predictors are influenced by the social context in which they operate and may even shape that context. In contrast, AI has the potential to process vast amounts of data in an unbiased manner, identify alternative scenarios that human planners may overlook, and effectively anticipate future trends in technology, economy, demographics, and military affairs. AI is less influenced by contextual factors and cognitive biases that often affect human planners. In uncertain and complex environments, human planners tend to rely on familiar mindsets and can be swayed by social pressures from peers and superiors. AI can help overcome these challenges by continuously reviewing past planning assumptions, comparing them to current events without biases or hubris. [21]

In the upcoming decade, it is anticipated that AI systems would prioritize intelligence and integrate knowledge-focused analytical skills. The AI solutions will then be connected to the network of physical and virtual domains, including sensors, businesses, people, and autonomous agents, while also utilizing blockchain technology's advantages for data integrity. To make use of large-scale, decentralized sensor networks, storage, and processing, they will be dispersed. In order to facilitate new disruptive impacts, they will digitally integrate the human, physical, and information worlds. [4]

In addition the new discipline defined Strategic Engineering is fast developing in Defence and strongly relies on the use of AI in combination with Simulation and data analytics [17]. This fact requires more and more to AI solutions considering the new dimensions and paradigms from Hybrid to Cognitive Warfare [18]. Combining AI with Simulation could allow also to cover

transformation of strategic aspects such as Defence Acquisitions to promote the renewed Smart Defence concept [19].

Moreover, recently major challenges such as Chemical, Biological, Radiological, and Nuclear (CBRN), due to the complexity of contamination and diffusion as well due to current crisis situation, resulted as strategic application field requiring use of advance in AI to properly evaluate the situation and future developments and very interesting achievements have demonstrated [20].

1.2.2 Challenges

Incorporating AI into military operations presents significant challenges, primarily centred on frugality, robustness, and AI explainability. Implementing AI in military operations necessitates frugal data use for accuracy, robust systems that withstand various challenges, and explainable AI to ensure operators can trust and comprehend AI-driven decisions. User interfaces must be simple, ensuring effective use of AI systems on the battlefield, where the time pressure is high. Addressing these issues requires a collaborative effort between AI experts, military strategists, and policymakers to balance innovation and security.

A key challenge that can slow or limit the use of modern AI in military applications is that the main ingredient in any ML application is data from which the machines can learn and ultimately gain insight. [23]This is why today there is a significant shift from model-centricity to data-centricity. In fact, in the past the focus was on developing and refining models, often neglecting the quality of data. As the limitations of this approach became clear, the emphasis has shifted towards the importance of high-quality, well-curated data. Even the most sophisticated models require robust data to perform optimally. Consequently, there is a growing recognition of the need to invest in better data collection, management, and processing techniques. This transition prioritizes data as the key driver of success in AI development.

Military organisations are often good at collecting data for debriefing or reconstruction purposes. However, there is no guarantee that the same data can be successfully used for ML. As a result, military organisations may need to adapt their data collection processes to take full advantage of modern AI techniques such as DL. [23]

The military domain is also characterised by a lack of available datasets. Much of the data collected and used in military operations is highly classified and sensitive; therefore, most data are not available. This is mainly because military operations involve personal data of soldiers, intelligence agents or other personnel, and information sharing within the military community is limited due to the dynamics of cooperation and different branches. In addition, the military operates in different environments, from land to sea to air to space, and data is generated in a variety of formats and standards, leading to a lack of standardisation. Finally, in some cases, historical data may not have been systematically collected or preserved, particularly in previous military operations. This can result in a lack of historical data for retrospective analysis.

Another issue related to data in military domain is that this field operates within a fragile ecosystem that is susceptible to biases and inequalities present in our society. Additionally, ethics and risk should be taken into consideration to prevent any misunderstanding and treated. It is essential to stimulate critical reflection on the protection of individuals and groups at the most fundamental level. This involves fostering innovation and technology that adhere to ethical values, promote individual well-being, and contribute to the betterment of society. There is the potential for significant unintended or malicious consequences for individuals, organizations, and society at large. The inherent risks associated with AI deployment must be carefully assessed and mitigated to prevent adverse outcomes. This includes considerations for privacy, security, algorithmic biases, and potential misuse of AI technologies. [23]

Further research is needed in the field of Human-Machine Teaming, as human involvement remains essential in military operations regardless of the advancement of AI techniques. ML, DL, adversarial AI, and emerging computing methods like neuromorphic and probabilistic computing require ongoing research to develop advanced algorithms that are effective in military applications. Given that AI has the potential to enhance predictive and cognitive data analytics, the transformation of both structured and unstructured data into actionable insights for decision-makers is crucial. Structured data can be processed using ML techniques and NNs, while unstructured data can be effectively processed using DL and natural language techniques. These tools can be applied to achieve more realistic analysis and personalized

training, particularly in areas such as modelling and simulation, adaptive techniques in electronic warfare, and the development of AI agents for defensive and offensive cyber operations across the entire information space. The information acquired through AI can be leveraged for military purposes, including the development of countermeasures for hybrid warfare. [4]

Concluding, the implementation of AI in military applications needs to address specific challenges, in particular transparency and consistency, vulnerability mitigation, and data challenges. Researchers have already made significant progress in addressing these challenges within AI, but a nuanced approach is required to translate them to the military domain, which requires risk tolerances, data quality criteria and legal constraints that require tailored solutions. In addition, exploring visualisation techniques from the field of visual analytics may prove invaluable in understanding and effectively presenting complex AI-derived information in military contexts. Future studies should prioritise these considerations to ensure that AI technologies meet the unique needs and challenges of military applications. [23]

1.2.3 Possible solutions

One of the primary goals of integrating AI into the military is to improve decision-making, both in training and in real-time operations. One of the most effective ways AI can achieve this is through threat assessment. However, training AI algorithms for threat assessment requires large amounts of high-quality data, which is a significant challenge for military applications due to data scarcity and variability in data quality. Indeed, in military operations data collection is often secondary to the primary objectives of the mission, making in-mission data collection challenging. In addition, privacy and security concerns associated with military data further complicate the collection and use of real-world data.

To address these challenges, one of the most convenient and practical solutions is to generate synthetic data. By creating realistic synthetic data, it is possible to simulate a wide range of scenarios and threats for the AI to learn from, without the need for extensive real-world data collection. This approach not only mitigates the risks associated with handling sensitive information, but also allows for the generation of large data sets that can be tailored to specific

training needs. Another way to increase the amount of data available is to improve the resolution and quality of existing data. Advanced algorithms can be used to enhance lower quality data, making it more useful for training AI models. Techniques such as data augmentation, where existing data is transformed in various ways to create new data samples, can also be used to enrich the dataset.

Once the available data is increased and privacy concerns are adequately addressed, it becomes feasible to employ methods that require training on large and high-quality datasets to achieve optimal efficiency. This improvement in data quality and quantity allows for more accurate and robust models Specifically, this work will focus on the area of AI-based threat assessment, fed with synthetic data obtained through generative learning approaches.

2. Literature review

2.1 Artificial Intelligence applications

The field of AI has significantly advanced in recent years, offering powerful tools and methodologies for solving complex problems across various domains. Among these, Genetic Algorithms (GAs) and Artificial Neural Networks (ANNs) stand out as prominent techniques with unique strengths and applications. The integration of GAs and ANNs can bring several advantages to different domains. ML, a subset of ML involving Deep Neural Networks (DNN), further enhances the capabilities of AI by enabling the modeling of complex patterns and relationships within large datasets. DL techniques, including Convolutional Neural Networks (CNNs) and Generative Learning (GL) models, have shown remarkable success in various applications, from image recognition to data generation.

2.1.1 Genetic Algorithm

A GA is a metaheuristic optimization algorithm inspired by biological evolution, commonly used as a search method and for modelling evolutionary systems. In this approach, binary strings are stored in a computer's memory and undergo modifications over time, resembling the evolution of populations under natural selection. GAs have the capability to evolve complex and interesting structures known as individuals, which represent potential solutions to problems. The process begins by randomly creating a population of individuals. In a simple case, each individual is represented as a bit string, representing a candidate solution for a specific problem. Variation among individuals in the population leads to varying degrees of fitness. These differences are utilized to bias the selection of a new set of candidate solutions in the next iteration, known as selection. During selection, a new population is formed by replicating more successful individuals while removing less successful ones. However, the replicas are not exact copies. There is a chance for mutation (random bit flips), crossover (exchange of substrings between two individuals), or other modifications to the bit strings during replication. By transforming the previous set of individuals through mutation and crossover, a new set of individuals is generated, ideally with an improved chance of being

successful. The GA is intriguing from a computational perspective due to claims of its effectiveness as a biased sampling algorithm. The argument for GA performance comprises three components: large populations initialized randomly provide independent sampling; selection preserves high-fitness individuals, biasing the sampling process towards regions of high fitness; crossover combines partial solutions from different strings onto the same string, exploiting parallelism within the population. While GAs are primarily recognized as problem-solving methods, they can also be employed to study and model evolution in various domains, including biological, social, and cognitive systems. GAs excel at manipulating multiple parameters simultaneously, which contributes to their strength as an optimization technique. [24]

GA, despite its inherent simplicity, has served as a foundational framework with numerous applications across scientific and engineering domains. GAs have found extensive utility in the realm of optimization. They have been effectively employed in tasks spanning numerical optimization, as well as complex combinatorial optimization problems, such as circuit layout and job-shop scheduling. GAs have been harnessed to evolve computer programs tailored for specific tasks. Furthermore, they have been instrumental in the design of various computational structures, including cellular automata and sorting networks. In the field of ML, GAs have been applied to a multitude of tasks, encompassing classification and prediction. These applications extend to tasks like weather prediction and protein structure forecasting. GAs have also played a role in evolving components of machine-learning systems, such as optimizing weights for NNs, refining rules for learning classifier systems, symbolic production systems, and even enhancing sensor configurations for robots. GAs have been leveraged to model intricate economic processes, shedding light on innovation dynamics, the development of bidding strategies, and the emergence of economic markets. GAs have been utilized to model various aspects of the natural immune system. This includes simulating somatic mutations occurring during an individual's lifetime and unravelling the discovery of multi-gene families over evolutionary time. GAs have provided valuable insights into ecological phenomena, ranging from biological arms races and host-parasite co-evolution to the study of symbiosis and the analysis of resource flow within ecological systems. [40]

The adaptability and problem-solving capabilities of GAs have led to their widespread use as a versatile tool in addressing complex challenges across multiple disciplines, illuminating new pathways for scientific and engineering advancements.

2.1.2 Artificial Neural Network

Inspired by biological neural networks, ANNs are highly parallel computing systems comprising numerous interconnected simple processors. ANN models attempt to incorporate organizational principles believed to be used in the human brain. One type of ANN considers nodes as artificial neurons, forming the basis of ANNs. Artificial neurons are computational models inspired by natural neurons. ANNs process information, making them useful in fields related to information processing. The connections between neurons in ANNs are associated with weights, which represent the flow of information. These weights are computed using a mathematical function that determines the activation level of the neuron. Input neurons have a single input, and their output is the input multiplied by a weight. There are algorithms that can adjust the weights of ANNs to achieve the desired network output. The training process starts with random weights, and the objective is to refine them to minimize errors. ANNs are employed for various challenging tasks such as pattern recognition (assigning labels to input), categorization and clustering, and prediction and forecasting. Network architectures can be classified based on their connection patterns into two categories: feed-forward networks (no loops in the graph) and recurrent networks (loops exist due to feedback connections). The connection weights in the network are typically learned from available training data, and performance improves iteratively by updating the weights. There are three major learning paradigms in ANNs. In supervised learning, the network is provided with correct answers for every input pattern, and the weights are adjusted to produce answers as close as possible to the known correct ones. Unsupervised learning, on the other hand, explores the underlying structure and correlations in the data to organize patterns into categories. Hybrid learning combines aspects of both supervised and unsupervised learning, where some weights are determined through supervised learning and others through unsupervised learning. Reinforcement learning involves providing the network with critiques on the correctness of

outputs rather than the correct answers themselves. The perceptron is an algorithm used for supervised classification, mapping inputs to non-binary outputs. It is an online learning algorithm, processing elements in the training set one at a time. In the context of ANNs, a perceptron is similar to a linear neuron but focuses on classification rather than regression. While a perceptron adjusts its weights to separate training examples into respective classes, a linear neuron adjusts its weights to minimize real-valued prediction errors. The perceptron algorithm is often referred to as the single-layer perceptron, distinguishing it from more complex neural networks. Boltzmann learning is a variation of error-correction learning, measuring error based on differences in correlations among outputs of neurons under specific operating conditions. The Hebbian rule increases the weight between two neurons when they activate simultaneously and decreases it when they activate separately. Competitive learning involves output units competing for activation, resulting in only one unit being active at a time, known as winner-take-all. Competitive learning can cluster or categorize input data based on correlations, automatically grouping similar patterns. The backpropagation algorithm is used in layered feed forward ANNs. Neurons are organized in layers, with signals propagating forward, and errors being propagated backward. Inputs are received by neurons in the input layer, and the network's output is determined by neurons in the output layer, with one or more intermediate hidden layers. Backpropagation employs supervised learning, where errors are computed, and weights are adjusted using the gradient descent method. [25]

2.1.3 Genetic Neural Network

Genetic Neural Networks integrate the principles of both GAs and NNs to solve complex problems and optimize the performance of neural networks. In a GNN, the GA is used to evolve and optimize the architecture and parameters of the NN. The GA operates on a population of NN structures, treating them as individuals in the population. Each individual represents a particular NN architecture, including the arrangement of layers, the number of neurons in each layer, and the connection weights. The GA applies genetic operators, such as selection, crossover, and mutation, to iteratively evolve the population of NNs. After each iteration of the GA, the resulting NN architectures are evaluated on the problem or task. The best-performing

individuals are selected to form the next generation, while less fit individuals may be eliminated. This process continues until a satisfactory NN architecture or solution is obtained. The advantage of using GAs with NNs is that it allows for automatic optimization of the network's structure and parameters, avoiding the need for manual tuning.

Genetic neural networks have found applications in various domains, leveraging their strong mapping and learning abilities, as well as their global search capability and robustness. These networks offer advantages compared to other methods by simplifying the consideration of complex relationships among threat factors and eliminating the need for tedious calculations. As a result, genetic neural networks are well-suited for threat assessment, which holds significant importance in military applications. Accurate threat assessment forms the foundation for commanders to make informed decisions regarding their actions. [26]

2.1.4 Deep Learning

Artificial Deep Neural Networks are computational models inspired by the brain, allowing computers to excel at cognitive tasks. These models consist of interconnected layers of simple processing units, similar to neurons, working in parallel. While basic neural networks have an input and output layer, the addition of more layers creates deep networks. During training, DNNs learn to perform specific tasks by adjusting the connections between units. Once trained, these DNNs can apply the learned task to new inputs. [27]

DL empowers computational models with multiple layers to learn data representations that have various levels of abstraction. By using the backpropagation algorithm, DL uncovers intricate structures within large datasets. This algorithm guides the machine to adjust its internal parameters, allowing each layer to compute representations based on the previous layer's representation. DL is a representation-learning approach that achieves multiple levels of representation by composing simple, non-linear modules. Each module transforms the representation from one level to a slightly more abstract level. In classification tasks, higher layers of representation amplify essential aspects of the input for discrimination while suppressing irrelevant variations. The progress of DL has enabled significant advancements in solving problems that have long challenged the artificial intelligence community. [29]

2.1.4.1 Convolutional Neural Network

In DL, a CNN is a class of feedforward artificial neural specifically designed to extract features from data using convolutional structures. Its architecture is inspired by visual perception, where biological neurons correspond to artificial neurons, CNN kernels represent different receptors responsive to various features, and activation functions simulate neural electric signals transmitting only when exceeding a certain threshold. CNN offers several advantages: local connections, weight sharing, and down-sampling dimension reduction. Local connections reduce parameters and expedite convergence by connecting each neuron to only a small number of neurons in the previous layer. Weight sharing allows a group of connections to share the same weights, further reducing parameters. Down-sampling, achieved through pooling layers, leverages image local correlation principles to reduce data while retaining crucial information. [28]

CNNs are specifically designed to process data in the form of multiple arrays. They capitalize on four key ideas that leverage the properties of natural signals: local connections, shared weights, pooling, and the use of multiple layers. The architecture of a typical ConvNet consists of a series of stages. The initial stages comprise convolutional and pooling layers. In a convolutional layer, units are organized in feature maps, and each unit connects to local patches in the previous layer's feature maps through a set of weights known as a filter bank. The result undergoes a non-linear activation function, such as a ReLU. All units in a feature map share the same filter bank, while different feature maps utilize distinct filter banks. The convolutional layer detects local conjunctions of features, whereas the pooling layer merges semantically similar features. Backpropagation through a ConvNet is similar to a regular deep network, enabling the training of all weights in all filter banks. The convolutional and pooling layers draw direct inspiration from the concepts of simple and complex cells in visual neuroscience. The application of deep convolutional networks to datasets has achieved remarkable results, significantly reducing error rates compared to other approaches. This success is attributed to efficient GPU utilization, the use of ReLUs, the regularization technique called dropout, and data augmentation methods involving deformations of existing examples. [29]

2.1.5 Generative Learning

GL is a theory that involves the active integration of new ideas with the learner's existing schemes. At its core, GL emphasizes that genuine comprehension arises when learners proactively construct meaning. This approach asserts that learners forge connections between stimuli and their stored information, encompassing knowledge and personal experiences. This process, referred to as generation, entails the creation of unique and meaningful associations between stimuli and stored information. Hence, GL represents the active process of constructing meaning through the generation of connections and associations between stimuli and the learner's existing knowledge, beliefs, and experiences. [30]

Focusing on computer science, GL algorithms are ML algorithms that aim to model the joint probability distribution of the input features and the corresponding labels or classes. They learn the underlying data distribution and generate new samples that are similar to the training data. Specifically, generative approaches try to build a model of the Positives and a model of the Negatives. A decision boundary is formed where one model becomes more likely and these create models of each class. Generative models can be used for tasks such as data generation, missing data imputation, and estimating the likelihood of data. On the other hand, discriminative models focus on learning the decision boundary between different classes or labels directly. They aim to model the conditional probability distribution of the labels given the input features. Discriminative models learn to classify or predict the labels based on the observed input features. They are often used for tasks such as classification, regression, and anomaly detection. [31]

The distinction between complex "adaptive" systems and complex "generative" systems is significant. Complex adaptive systems are associated with self-organization, while complex generative systems are linked to self-transcendence, driving agents toward the implicate order. Adaptive learning refines existing competencies without challenging underlying beliefs, while GL questions operating norms and transcends to develop a new order. Self-organization and self-transcendence occur under specific conditions across individual, social, and impersonal dimensions. Adaptive learning involves self-organization within the explicate order,

while GL involves self-transcendence and approaching the implicate order. Logic and deductive reasoning are predominant in adaptive learning, while intuition, attention, and dialogue play a key role in GL. As result, adaptive AI can modify its own code in response to real-world changes that were not anticipated at the time of its creation; generative AI focuses on developing unsupervised and semi-supervised algorithms capable of producing new content, such as text, audio, video, images, and code, by utilizing existing data. [33] The choice between complex adaptive and complex generative systems hinges on the nature of the data environment and the desired outcomes. GL emerges as a favourable option when faced with situations characterized by scarcity and low quality of data. Its capacity for self-transcendence, coupled with the ability to generate new content, positions it as a promising approach to navigate through data limitations. By leveraging existing data creatively, generative AI can pave the way for innovation and adaptation in scenarios where traditional adaptive learning might fall short. Therefore, embracing GL methodologies holds significant potential in addressing the challenges posed by data scarcity and quality concerns, propelling advancements in AI towards more resilient and versatile solutions.

DL models, which consist of deep, hidden layers in ANNs, can capture complex patterns within data. Deep Generative Learning (DGL) models combine probabilistic generative models with DL architectures, allowing for efficient modeling. DGL models, such as recurrent neural networks with a generative subsystem, can capture complex probabilistic distributions even in high-dimensional or noisy data. They can generate unseen data efficiently, even with limited labeled or unsupervised data. DGL has the potential to enhance existing features, improve human-AI interaction, differentiate software-based offerings, and break innovation barriers. [32]

2.2 Frugal AI

Having high-quality data is relevant in developing successful algorithms. However, the processes of data collection, processing, and storage are time-consuming and expensive. As ML systems become more complex, it becomes increasingly challenging to characterize their objectives. Therefore, the goal is to design AI systems that can achieve a certain level of

robustness within their intended field of use while requiring less data. This is where the concept of frugal AI comes into play. In certain domains like the military, complete databases may not be readily available due to security reasons. In such cases, frugal AI becomes essential. [36]

Frugal AI is a technique that aims to achieve robustness in AI models while using fewer data and computational resources. It involves training AI systems with limited resources, focusing on input frugality and learning frugality. Input frugality refers to the emphasis on data and the use of fewer training datasets or features compared to what is typically required in a non-frugal setting. This approach can be motivated by resource constraints or privacy concerns. The goal is to achieve prediction quality while using a smaller amount of data. Learning frugality, on the other hand, focuses on minimizing the costs associated with the learning process, such as computational and memory resources. This aspect is primarily driven by constraints like limited computational power or battery capacity. By optimizing the learning process, Frugal AI aims to achieve efficiency in terms of resource usage. Traditionally, the conventional approach in AI involved acquiring large datasets and investing significant effort in perfecting the code. However, with the availability of vast amounts of data, the code required for processing and gaining insights has matured. Achieving Frugal AI goes beyond data efficiency, but focusing on the data is a logical starting point. Training AI systems on small, well-engineered datasets may seem counterintuitive to established AI and ML practices. However, ensuring robustness within the intended field of use is a crucial aspect of building trust in AI systems. By achieving robustness despite limited resources, frugal AI offers the potential to address resource constraints while maintaining reliable and trustworthy AI models. [35]

The implementation of Frugal AI represents an innovative approach for addressing the challenges posed by the military domain, which is often characterized by a scarcity of relevant and robust datasets, where data availability may be limited due to the sensitive and classified nature of information, traditional ML approaches can encounter significant barriers in terms of training and deploying models. Moreover, the development and deployment of specialized ML models and architectures that are designed to function effectively with minimal sample sizes for training and inference is essential, given that military applications often require a high level of accuracy, reliability, and real-time decision-making capabilities, making it imperative to find

ways to leverage the available data efficiently. Frugal AI can bring about several advantages in the military field. By adapting ML techniques to work with smaller datasets, the military can unlock the potential of AI in areas like target recognition, autonomous systems, and threat detection, even in scenarios where data scarcity would typically hinder progress. Moreover, the concept of Frugal AI is not only relevant to the military but also holds potential for broader applications in sectors where data availability is a concern. Its development highlights the adaptability and resilience of AI technologies to perform effectively in resource-constrained environments, opening the door to new possibilities in various fields beyond defence.

Researchers and engineers are exploring diverse methods to address the scarcity of data, aiming for efficiency and frugality. One approach is transfer learning, which involves leveraging an existing AI system that has already learned from a sufficient dataset. The idea is to start with an AI that has learned from data that is somewhat similar to the missing data, thereby benefiting from the prior knowledge. Another technique is data generation, where a virtual environment is utilized to generate data that closely resembles the conditions of a real environment. The objective is to create missing data for training directly from this simulated environment. Data augmentation is another method that involves generating new data by applying transformations or modifications to existing data. Combining data generation and augmentation can lead to a more comprehensive and diverse dataset. By employing these techniques, frugal AI aims to address the challenge of limited data by making the most out of the available resources and generating additional data to enhance the training process. [37]

2.3 Synthetic data generation

Data is ever-present and holds significant value, but to unlock that value, data quality is paramount. When dealing with sensitive information like medical records, credit datasets or restricted military information, safeguarding data privacy is crucial, without compromising quality. The need for high-quality data and privacy preservation has gained prominence as both businesses and researchers increasingly rely on data. Synthetic data, comprising artificially generated information, emerges as a powerful solution to address these challenges. Synthetic data, being generated rather than collected or measured, often exhibits higher

quality compared to real data. Furthermore, privacy safeguards can be enforced to prevent the disclosure of critical information, such as restricted data in the military.

The military sector, with its unique operational challenges and stringent security and privacy requirements, often faces with a scarcity of available datasets. This scarcity can be attributed to a combination of factors, including the inherent difficulties in collecting accurate and error-free data in the field and the pressing concerns surrounding data privacy. In the military domain, collecting data is challenging due to the dynamic and high-stakes nature of operations. Field conditions, which are often hostile and unpredictable, can introduce errors into data collection processes. This is particularly problematic because military operations demand precision and reliability. Furthermore, privacy concerns are dominant in the military sector, given the sensitive nature of the data involved. Military operations involve classified and confidential information, and protecting the privacy of individuals involved in these operations is not only a legal requirement but also a moral imperative. This makes it challenging to share, access, or utilize real military data for research, development, or analysis. Considering these challenges, synthetic data generation emerges as a valuable and innovative solution. It offers the capability to create data that closely mimics the characteristics of real data without being constrained by the difficulties of collecting authentic data. Synthetic data can be used for a wide range of applications within the military sector, including training simulations, equipment testing, and algorithm development. It provides a way to generate data that is representative of real-world scenarios while maintaining data privacy and security. This allows researchers, developers, and analysts to work with realistic data without compromising the confidentiality of sensitive information. Synthetic data generation, when executed properly, becomes an essential tool in overcoming the scarcity of real military datasets. It not only aids in ensuring the accuracy and reliability of data but also safeguards the privacy and security of sensitive military information, making it an asset in the modern military landscape.

While synthetic data is a compelling concept, its generation demands precision. It must be plausible and adhere to the underlying distribution of the original data. Consequently, the algorithms responsible for generating synthetic data must exhibit robustness and effectively capture the patterns inherent in real data. SMOTE (Synthetic Minority Oversampling

Technique), introduced as early as 2002, is one of the pioneering algorithms aiming to replicate data distributions, alongside techniques like Random OverSampling (ROS) and traditional approaches like rotation or scaling. Over the years, this idea has matured, leading to the proposal of several variants. Notably, the advent of DL has spawned more promising ideas, including Variational AutoEncoders (VAEs) in 2013 and, most significantly, Generative Adversarial Networks (GANs) in 2014. [37][37]

There are various strategies and techniques available to develop robust ML applications, especially when faced with limited training data. In this context, transfer learning, GL, and modelling and simulation are tools that can significantly enhance the capabilities of ML systems. Moreover, they can be used combined for developing ML applications in the military. These techniques not only mitigate the constraints imposed by limited training data but also improve the adaptability and accuracy of models while reducing risks and costs associated with data collection. In this way, they enhance the military's ability to leverage the benefits of cutting-edge ML technologies, while also ensuring the security and precision demanded by such critical applications.

2.3.1 Transfer learning

The idea of transfer learning finds its roots in the field of educational psychology. Drawing on the generalization theory of transfer, it postulates that the ability to transfer learning is a consequence of experiences being generalized. Essentially, individuals can apply their knowledge to new situations by extrapolating from their prior experiences. [38]

An excellent technique to be used when the availability of dataset is scares is transfer learning. Transfer learning in DL applications offers two distinctive approaches, each tailored to optimize model performance. The first approach involves "Relearning the Output Layer." Here, the final layer of a pre-trained model is replaced with a new output layer designed to suit the expected output of the new task. During training, only the weights associated with this new output layer are updated, while all other model parameters remain unaltered. This method is particularly effective when the source and target tasks share a significant degree of similarity. The second approach is "Fine Tuning the Entire Model." This method resembles the first approach, but

with a key difference: it permits updates to the weights of the entire DNN. Fine-tuning the entire model can be advantageous when the source and target tasks exhibit a moderate to high degree of relatedness, although it typically requires a more substantial amount of training data to achieve optimal results. [23]

Several methods are commonly used to increase the efficiency of transfer learning. Pre-trained models, trained on large datasets for specific tasks, are a common and efficient way to transfer knowledge. These pre-trained models act as a starting point, providing learned representations that can be adapted to new tasks through fine-tuning or feature extraction. [23] Fine-tuning involves taking a pre-trained model and training it on a new dataset or task. This process allows the model to adapt its learned representations to the specific characteristics of the new data, often leading to improved performance on the target task. Fine-tuning is particularly useful when the source and target tasks are similar, or when labelled data in the target domain is limited. Instead of fine-tuning the entire pre-trained model, feature extraction uses only the learned representations (features) from the pre-trained model. These features can then be used as input to a new model trained specifically for the target task. Feature extraction is useful when computational resources are limited or when the target task differs significantly from the source task. Data augmentation techniques, such as flipping, rotating, scaling, or adding noise to the input data, can be used to increase the diversity of the training data and improve model generalisation. By applying data augmentation techniques learned from one task or domain to another, transfer learning can be made more efficient, especially when labelled data in the target domain is limited. In cases where the source and target domains have different data distributions, domain adaptation techniques can be used to align the data distributions between the domains. This can involve techniques such as adversarial training or domain-specific regularization to make the transfer of knowledge more effective. Ensemble methods, such as model stacking or model averaging, can be used to combine multiple pre-trained models or fine-tuned models to improve performance. By leveraging the complementary strengths of different models, ensemble methods can often achieve better performance than any single model alone.

Transfer learning's major benefit lies in its potential to enhance a model's generalization capabilities, allowing it to perform effectively across a wide spectrum of tasks. By transferring knowledge from pre-trained models, even in cases of limited training data, models can achieve better performance. However, it's important to acknowledge that the efficacy of transfer learning diminishes as the dissimilarity between the source and target tasks increases. In essence, transfer learning is a versatile tool in the DL toolkit, enabling practitioners to build models that excel in various domains, leveraging prior knowledge and experiences to adapt to new challenges.

2.3.1.1 Categorization

Transfer learning can be subdivided in 2 main categories: problem and solution categorization. [38][38] The first focuses on understanding the nature of the tasks being transferred between the source and target domains. It helps to define the setting in which transfer learning is applied and provides insights into the challenges and strategies involved. The second focuses on the methods and approaches used to transfer knowledge between domains. It helps to understand how the transfer is carried out and what techniques are used to exploit the knowledge from the source domain.

Problem categorisation is divided into label-setting-based categorisation, which focuses on the availability of labelled data in the target domain and the relationship between the tasks in the source and target domains, and space-setting-based categorisation, which instead efforts on the similarity or dissimilarity between the source and target domains in terms of their feature spaces or distributions. Label-setting-based categorization encompasses three distinct sub settings, namely inductive transfer learning, transudative transfer learning, and unsupervised transfer learning. [38] In the inductive transfer learning setting, the target task differs from the source task, regardless of whether the source and target domains are the same or not. This scenario necessitates some labelled data in the target domain to induce a predictive model for use in that domain. When an abundance of labelled data exists in the source domain, it resembles multitask learning. However, in this case, the focus is solely on achieving high performance in the target task by transferring knowledge from the source task, while multitask

learning concurrently learns both source and target tasks. When no labelled data is available in the source domain, it mirrors self-taught learning. In self-taught learning, differences in label spaces between the source and target domains imply that the side information from the source domain cannot be directly employed. This situation aligns with the inductive transfer learning setting where labelled data in the source domain is inaccessible. In the transudative transfer learning setting, the source and target tasks are the same, but the source and target domains differ. This scenario lacks labelled data in the target domain, although it benefits from an abundance of labelled data in the source domain. In the unsupervised transfer learning setting, similar to inductive transfer learning, the target task differs from the source task but remains related to it. Unsupervised transfer learning focuses on addressing unsupervised learning tasks in the target domain, such as clustering, dimensionality reduction, and density estimation. In this case, there are no labelled data available in either the source or target domains during training. [39] Space-setting-based categorization comprises homogeneous transfer learning, where the source and target data are in the same feature space, and heterogeneous transfer learning, where the source and target data are represented in different feature spaces. [38]

Solution categorization comprises four groups: instance-based, feature-based, parameter-based, and relational-based approaches. Instance-based transfer learning is based on selectively weighting instances from the source domain to support learning in the target domain. The goal is to identify which instances from the source domain are most relevant to the target task. This is often achieved through techniques such as instance reweighting or instance selection, where instances with higher similarity or relevance to the target task are given greater weight during training. This approach adapts the model's learning process to prioritise knowledge that is most useful for the target domain. Feature-based transfer learning involves the transformation of feature representations to bridge the gap between the source and target domains. This approach recognises that the original features may not be directly applicable or optimal for the target task. Asymmetric feature-based transfer learning modifies the source features to match the characteristics of the target domain. On the other hand, symmetric feature-based transfer learning seeks to establish a shared latent feature space

where both source and target features can be mapped. By reshaping the feature space, this approach facilitates the transfer of relevant knowledge while accounting for domain differences. Parameter-based transfer learning focuses on the transfer of knowledge encoded in model parameters from the source to the target domain. This involves using pre-trained models or learned parameters from the source task to initialise the model for the target task. Subsequent fine-tuning of these parameters allows the model to adapt to the nuances of the target domain. Parameter-based transfer learning enables the efficient use of knowledge from previous tasks, providing a head start for learning in the target domain while allowing for task-specific adjustments. Relational transfer learning addresses tasks where relational structures play a central role, such as in knowledge graphs, social networks or recommender systems. These tasks involve understanding complex relationships between entities or objects. Relational transfer learning focuses on transferring learned relational knowledge, such as logical rules or graph structures, from the source domain to the target domain. By encoding and transferring these relational patterns, the model can effectively capture and exploit the underlying structures present in both domains, thereby improving performance in relational domains. [38]

2.3.1.2 Style transfer learning

Style transfer learning is a technique in ML and computer vision that aims to transfer the style of one image to another. Specifically, the technique consists of specifying an input image as the base image, the content image, and at the same time specifying another image as the desired image style. The image style transfer algorithm then transforms the image style while preserving the structure of the content image. In this way, the final output composite image represents a perfect combination of the input image content and the desired style. A crucial step is to analyse an image of a particular style and developing a mathematical or statistical model to represent that style. By understanding the underlying patterns and characteristics of the style image, a model can be constructed to guide the transformation process. This model serves as a reference for adjusting the target image to better align with the established style, resulting in visually pleasing outcomes. However, a notable drawback of traditional style

transfer approaches is their limited versatility. These methods typically excel at replicating a specific style or scene but struggle to adapt to diverse styles or accommodate multiple stylistic elements within a single image. As a result, the practical applicability of conventional style transfer research remains constrained. This challenge has only been more effectively solved and its performance has greatly improved in recent years, driven by the DNN's remarkable performance in large-scale picture classification and its potent multi-level image feature extraction and representation capabilities. [66]

Before the application of NNs to style transfer, several methods were employed to model and extract style from images: texture synthesis, image analogy and image filtering. Texture synthesis techniques aimed to generate new images with similar texture characteristics as a given style image. These methods typically relied on statistical analysis of pixel neighbourhoods to capture texture patterns and then reproduce them in the synthesized image. Image analogy techniques were used to transfer the style of one image onto another by establishing correspondences between similar image regions. By analysing the structural and visual similarities between corresponding regions in the style and content images, these methods attempted to transfer the stylistic elements while preserving the content of the target image. Image filtering techniques involved applying various image processing filters to manipulate the appearance of an image. [66] While these techniques could achieve some degree of style transfer, they often struggled to produce results that were both visually appealing and faithful to the original content. Moreover, these techniques relied heavily on handcrafted features and heuristics, which limited their flexibility and adaptability to different styles and content.

NNs have the significant advantage to automatically extract the most useful features from images after training. Before their introduction, features were extracted by cutting the original object into smaller pieces. Style transfer methods integrated with NNs can be classified as Slow Transfer, based on image optimization, and Fast Transfer, based on model optimization. The initial focus is on optimising the size and format of images transferred between devices or servers. To achieve this, methods such as image compression, resizing, and format conversion are commonly used. By reducing the size of the image data, transfer times are

decreased, resulting in faster overall processing. However, this approach may result in some loss of image quality, depending on the level of compression applied. Fast Transfer based on model optimization focuses on optimizing the ML model itself to reduce its size and complexity. Techniques such as model pruning, quantization, and distillation are employed to create smaller and faster models. This approach reduces the computational resources required during inference, resulting in faster processing and transfer times. However, it is important to note that there may be a trade-off between reducing the size of the model and maintaining its accuracy. Extreme optimization techniques may sacrifice some level of model performance. [66] The primary distinction between these methods is that Slow Transfer, which is based on image optimization, aims to reduce transfer times by optimizing the images themselves. In contrast, Fast Transfer, which is based on model optimization, aims to achieve faster processing by optimizing the ML model used for inference.

2.3.1.3 Domain adaptation

Domain adaptation is a specific type of transfer learning; these two closely linked problem settings are subdisciplines of ML that use information from another related area with sufficient labeled data to improve the performance of a target model with inadequate or no annotated data. In contrast to transfer learning, where tasks may change between source and destination, in domain adaptations the tasks remain static and only domains vary. [68]

The goal of domain adaptation is to minimize the domain gap in order to successfully transfer the model trained on the source domain to the target domain. The domain gap corresponds to the differences between the data collected in the target domain and the source domain, captured by different sensors, from different perspectives or under different illumination conditions. [69]

Current domain adaption techniques assume that label sets are consistent between domains. They address the domain gap in one of three ways: by producing features or samples for target domains, by learning domain invariant feature representation, or by using generative models to convert samples between domains. Two methods of domain adaptation are recommended by recent research: open set domain adaptation and partial domain adaptation. Partial domain

adaptation requires that the source labels include the target labels. On the other hand, open set domain adaptation introduces 'unknown' classes in both domains and assumes knowledge of common classes during training. Moreover, modified open set domain adaptation removes data associated with source unknown classes to ensure the source label set is a subset of the target label set. Some approaches allow for partially shared label sets and require labeled data in the target domain. These advancements address practical challenges in domain adaptation. In real-world situations, these presumptions are frequently broken. Partial domain adaptation is appropriate in the situations when the source label set adequately encompasses the target labels. Conversely, when the source label set shares classes or is a subset of the target label set, open set domain adaptation is the preferred method. [69]

The main problem in the general scenario is that we have no prior knowledge about the label set of the target domain, and therefore it is not possible to choose the appropriate domain adaptation method. Unsupervised Domain Adaptation (UDA) offers a promising solution to this problem by using labelled source data and unlabelled target data to train a classification model. The goal is to ensure that the learned model can effectively generalise and perform well in the target domain despite the lack of labelled examples. By learning from the labelled data in the source domain and exploiting the unlabelled data in the target domain, UDA techniques aim to bridge the domain gap and improve the adaptability of the model to new, unseen data distributions. [70] There are two key technological obstacles when creating domain adaptation models with UDA. The first one has to do with the inability to determine which portion of the source domain should match which portion of the target domain in the absence of knowledge about the target label set. If the entire source domain and the complete target domain match, the model will decline. The second problem pertains to the model's ability to classify target samples as "unknown" if they are not associated with any class within the source label set. The classifier is unable to determine the precise category of these classes since there are no labelled training data available for them. To address these challenges can be used Universal Adaptation Network (UAN), which incorporates a novel criterion aimed at quantifying the transferability of individual samples. This criterion combines measures of domain similarity with prediction uncertainty for each sample, creating a sample-level

weighting mechanism. With the enhanced UAN model, samples belonging to the common label set shared between the source and target domains are automatically identified and matched, while target samples associated with private label sets can be effectively flagged as "unknown" using a rejection pipeline. [69] In real-world scenarios, the applicability of UAN can lead to improved adaptability and performance of models in different domains. By effectively quantifying the transferability of samples and facilitating the automatic identification of common labels while accurately labelling unknown ones, UAN increases the robustness of models in universal domain adaptation tasks.

2.3.2 Modelling and Simulation

The military has leveraged modelling and simulation extensively in various capacities, including training, decision support, and research. The factor of time that the commander and his staff need for decision-making process is increasingly important. The possibilities of modelling and simulation can significantly shorten and specify the planning process of military engineering tasks. Consequently, there exists a substantial repository of well-established and validated models that have evolved over extended durations. These models also hold the potential for generating synthetic data to support ML applications. [23]

Modelling and simulation techniques are important for generating synthetic data that closely mimics real-world data, and they have a wide range of applications in various fields. At its core, the process involves creating artificial data that captures the statistical and structural characteristics of real data, making it a valuable resource when genuine data is scarce, sensitive, or expensive to acquire. Synthetic data will be constructed defining as foundation a Data Generating Process (DGP), a conceptual framework that outlines the essential elements and relationships within the data. It includes specifications on data distributions, correlations, and other attributes that need to be preserved in the synthetic data. Choosing the appropriate modelling approach is the next crucial step. This choice depends on the nature of the real data and the intricacies of the DGP. Whether it's a statistical model like a linear regression or a ML model such as a DNN, the modelling approach needs to align with the complexity and characteristics of the real data to effectively capture its essence. Parameter estimation is vital

when using a model-based approach. It involves estimating the model parameters using real data. The accuracy of these estimates is pivotal because it directly influences how closely the synthetic data will resemble the real data. Properly estimated parameters ensure that the synthetic data generation process accurately reflects the underlying statistical properties of the original data. With the DGP and model parameters in place, the process moves to simulating data generation. This is where the chosen modelling approach comes into play, allowing for the generation of synthetic data points that follow the patterns and relationships specified in the DGP. The goal is to create synthetic data that mirrors the structure and statistical properties of the real data, thus making it a reliable substitute for various purposes. Data validation and adjustment are critical steps to ensure the quality and accuracy of the synthetic data. Comparing the synthetic data to the real data helps identify any discrepancies. If disparities exist, adjustments may be necessary, whether it involves fine-tuning the model, revisiting parameter estimates, or refining the data generation process to enhance fidelity. [41]

In cases where privacy and data security are paramount, synthetic data proves to be an essential tool. Privacy-preserving techniques can be applied to synthetic data generation to protect sensitive information, making it impossible to re-identify individuals from the synthetic data. This is especially important in situations where compliance with data protection regulations is mandatory. The performance assessment of models or algorithms trained on synthetic data is a necessary step to gauge the utility of the synthetic data. By comparing the results obtained with synthetic data to those from real data, we can evaluate how well the synthetic data replicates the behavior of the real data. This step is fundamental in ensuring the reliability and applicability of the synthetic data. The process of generating synthetic data is iterative, making it an ever more reliable resource. Modelling and simulation for synthetic data generation is a systematic and adaptable process, offering solutions to various data-related challenges, from data privacy concerns to research and development needs. The overarching aim is to create synthetic data that faithfully reflects the intricacies of the real data, enabling accurate and meaningful analysis and modelling.

2.3.3 Generative learning

GL is an important theory that delves into the active cognitive process of integrating new knowledge and ideas with a learner's existing mental framework. At its core, this approach places a strong emphasis on the concept that genuine comprehension occurs when learners proactively construct meaning rather than passively absorbing information. This cognitive process, known as "generation," involves the creation of unique and meaningful associations between the stimuli encountered and the pre-existing information stored in an individual's mind. GL represents the dynamic process of constructing meaning through the active generation of connections and associations between external stimuli and an individual's pre-existing knowledge, beliefs, and experiences. Rather than simply memorizing facts or following a prescribed set of instructions, GL encourages learners to question, explore, and develop a deeper understanding of the subject matter. This approach often leads to better retention and application of knowledge, as learners are actively engaged in the process of connecting new information to what they already know. [30]

In practical terms, GL can involve various strategies and techniques, such as problem-solving, critical thinking, and active discussion. It is not limited to any specific educational domain but can be applied to a wide range of subjects and disciplines. GL, in the context of synthetic data generation, is a cutting-edge technique that leverages ML models to create artificial data that mimics the statistical patterns and characteristics of real-world data. This approach is particularly valuable in scenarios where obtaining sufficient and diverse real data may be challenging, expensive, or privacy sensitive. Generative models, such as GANs and VAEs are commonly used in GL for synthetic data generation. These models learn the underlying data distribution by analysing the patterns, relationships, and structures present in authentic data. Once trained, they can produce synthetic data samples that closely resemble the original data in terms of statistical properties and features. However, it's important to note that the quality and effectiveness of synthetic data depend on the accuracy and representativeness of the generative model. Careful evaluation and validation are critical to ensure that the synthetic data aligns with the requirements of the intended applications. Moreover, GL is extending its reach into the domain of generating high-resolution images from low-resolution inputs, marking

a remarkable expansion of its applicability. In this quest, researchers have explored a wide range of techniques. These approaches harness the power of generative models to improve image resolution, providing a compelling demonstration of the versatility inherent in GL methods. Using DL architectures and sophisticated algorithms, these techniques excel at the task of generating high-quality images with finer detail and improved resolution. [37]

2.3.3.1 Generative Adversarial Network

GANs were introduced in 2014 as a powerful class of generative models. They are inspired by a two-player zero-sum game, where the gains of the two players balance each other exactly. GANs typically consist of a generator and a discriminator that learn simultaneously. The generator aims to capture the potential distribution of real samples and generate new data samples, while the discriminator acts as a binary classifier, accurately distinguishing between real and generated samples. [42]

More specifically, one network, the generator, defines data implicitly and is capable of drawing samples from the distribution of this data. The generator is defined by a prior distribution over a vector, which serves as input to the generator function. In the generator, noise is intentionally added to create synthetic data resembling the real training data but not identical. This introduces variability and prevents the generator from reproducing the exact training set, fostering diversity in generated samples. The other player in this game is the discriminator. The discriminator examines samples and estimates whether they are real or fake. GANs provide a way to learn deep representations without extensive annotated training data, making them valuable for both semi-supervised and unsupervised learning. In the training process of GANs, the loss functions for both the generator and discriminator are present and fundamental. The generator's loss function guides it to produce samples that are more likely to be classified as real by the discriminator. This incentivizes the generator to create increasingly realistic samples as training progresses. The discriminator's loss function encourages it to correctly classify real samples as real and fake samples as fake. This guides the discriminator to improve its ability to discern real from fake samples. [34]

The elements and the structure of GAN is showed in **Figure 1**.

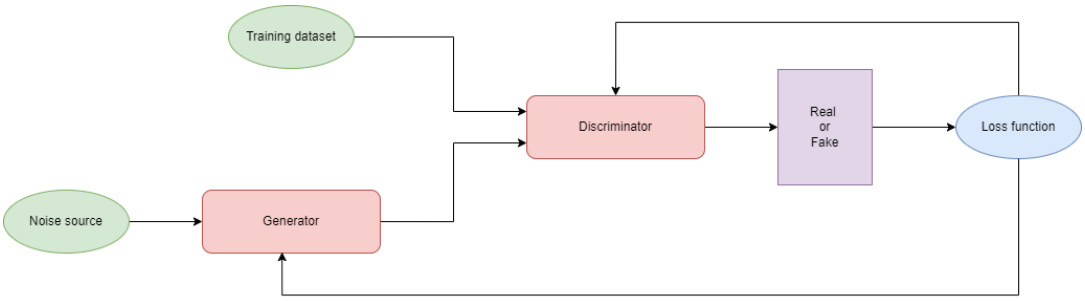

***Figure 1**: Generative Adversarial Network*

A relevant issue related to the parallel training of these two players is that the generator can continue to be trained even when the discriminator is optimal, aiming to reduce the accuracy of the discriminator. If the generator distribution perfectly matches the distribution of real data, the discriminator will be maximally confused, predicting 0.5 for all inputs. However, in practice, the discriminator may not be trained to optimality. GANs learn by implicitly computing the similarity between the distribution of a candidate model and the distribution corresponding to real data. [42]

NNs have contributed to the success of GANs due to their training process using the backpropagation algorithm, flexible structural design, and wide range of applications. However, the convergence and existence of equilibrium points in GANs have not been formally proven. Ensuring balance and synchronization between the two adversarial networks during training can be challenging, potentially leading to instability in the training process. [34]

In the military field, GANs offer numerous benefits and applications. They can be used to develop advanced camouflage techniques, enabling soldiers to blend seamlessly into their surroundings, enhancing stealth, and reducing the risk of detection. GANs are also instrumental in generating realistic 3D models of terrains and urban environments, crucial for military simulations and mission planning. Furthermore, GANs can play a role in intelligence and reconnaissance efforts by creating deceptive images and decoys to confound enemy surveillance. In image recognition tasks essential for military operations, GANs augment limited datasets with synthetic data, improving the accuracy of ML models, particularly when collecting real-world data is costly or risky. GANs outclass in anomaly detection, helping identify deviations from normal patterns in data streams, crucial for early threat detection and cybersecurity. In the aerospace and naval domains, GANs optimize stealth capabilities by

virtually generating and evaluating designs for radar-absorbing materials and shapes, enhancing the ability of military assets to evade detection. GANs have the potential to significantly enhance military operations, from improving camouflage to optimizing resource allocation and enhancing stealth technology.

An advanced and improved version of GANs is the Wasserstein GAN, also known as WGAN. WGAN introduces a different loss function, the Wasserstein distance, which leads to significant enhancements in training stability and the quality of generated samples. [52] Instead of the traditional binary output indicating whether an image is detected as real or fake, the discriminator in WGAN outputs a continuous value representing the Wasserstein distance, which quantifies how far the generated image is from a real image in terms of their underlying data distributions. [51] Using the Wasserstein distance as the loss function gives the generator a more meaningful and informative gradient during training. As a result, the generator receives input that is clearer and easier to understand regarding how much improvement is required for it to match the distribution of real images. This helps prevent typical problems like mode collapse, where the generator only produces a small range of samples, and also speeds up and improves convergence during training.

Moreover, one key aspect of WGAN is its focus on the Lipschitz continuity constraint. In traditional GANs, there is no explicit constraint on the discriminator, which can lead to instability during training. The Wasserstein distance introduces the concept of Lipschitz continuity, which means that the discriminator's function should be locally Lipschitz. In practical terms, this constraint prevents the discriminator from having extreme gradients that could cause the training process to become unstable. To enforce the Lipschitz continuity constraint, some early versions of WGAN used weight clipping, where the weights of the discriminator are restricted to a small range, effectively constraining the Lipschitz constant. While weight clipping can stabilize training, it has some drawbacks, such as potential gradient vanishing or exploding. [51]

Despite its advantages, WGAN and its variations can be more computationally intensive to implement than regular GANs and necessitate careful hyperparameter tweaking.

Nevertheless, they have achieved significant advances in the field of generative modelling because of their capacity to offer a more reliable and explanatory training process, and continuing research is working to improve and broaden their capabilities.

A particular type of GAN designed for image-to-image translation tasks, eliminating the necessity for paired training data is CycleGAN. It is a robust deep-learning framework that leverage two GANs, to learn the correspondence between disparate image domains. Through training, it adeptly captures the defining features of the target domain, facilitating the generation of novel images from the source domain that encapsulate these attributes. The training process initializes by independently training the two GANs. Initially, one GAN's generator learns to produce target domain images from the source domain, while its discriminator distinguishes real target domain images from generated ones. Concurrently, the second GAN's generator is trained to generate a source domain image from a target domain image. Subsequently, the two GANs are combined to form the CycleGAN. Their generators constitute the CycleGAN's generators, and their discriminators serve as its discriminators. The model is then trained to generate images from one domain that closely resemble those from the other domain, facilitated by a cycle-consistent loss function. This loss function encourages the model to generate images that are perceptually indistinguishable from genuine target domain images. Additionally, an identity mapping component is integrated to preserve the original characteristics of source domain images during translation to the target domain, further enhancing the model's performance. [71]

By employing two GANs, CycleGANs enhances accuracy, robustness, and efficiency compared to other GAN methods while maintaining lower complexity. However, this architecture's susceptibility to overfitting, along with its tendency to slow down training and occasionally pose challenges during training, remains a concern. [71] Despite these limitations, CycleGANs remain a valuable tool for various image translation tasks, demonstrating their potential to effectively bridge the gap between disparate domains, especially when aiming to emulate the characteristics of a target domain.

2.3.3.2 Real ERSGAN

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) builds on the foundation laid by Super-Resolution Generative Adversarial Networks (SRGAN), which is a pioneering approach in the field of Single Image Super-Resolution (SISR) and it introduced the concept of using GANs to improve the visual quality of super-resolved images. SRGAN aims to overcome the limitations of traditional Peak Signal-to-Noise Ratio (PSNR) oriented methods, techniques aimed at optimizing the PSNR value minimizing the distortion introduced during compression or reconstruction, by optimising images in a perceptual space rather than a pixel space. SRGAN achieves this by incorporating adversarial loss and perceptual loss functions into the training process. The adversarial loss encourages the generator to produce images that are indistinguishable from high-resolution images, while the perceptual loss ensures that the generated images are visually similar to the ground truth images in terms of perceptual features. ESRGAN addresses some of the limitations of SRGAN and further improves the visual quality of super-resolved images. ESRGAN introduces novel techniques such as Residual-in-Residual Dense Block (RRDB) for improved network architecture, Relativistic Average GAN (RaGAN) for improved discriminator performance, and adjustments to the perceptual loss function to overcome drawbacks observed in SRGAN. The RRDB is an improved version of the Dense Block, which is commonly used in CNNs. It comprises multiple residual blocks, each containing densely connected layers. Within each residual block, a residual-in-residual structure is used to enable more effective feature reuse and propagation. This facilitates the learning of complex features across multiple scales, resulting in improved super-resolution performance. RaGAN instead of solely trying to increase the probability of the discriminator correctly classifying real samples as real and fake samples as fake, it also aims to minimize the difference between the probability of the discriminator classifying real samples as real and the probability of classifying generated samples as real. This modification encourages the generator to produce images that are not only visually realistic but also have high fidelity compared to real images. In addition, ESRGAN proposes network interpolation as a strategy to effectively balance perceptual quality and PSNR. These improvements result in superior super-resolution performance, making ESRGAN a significant advance in the field of

SISR. [75][74] Nonetheless, ERSGAN may struggle to preserve fine details, resulting in images that appear overly smooth or lacking sharpness, particularly in textures or high-frequency regions. Moreover, GAN-based models like ESRGAN are prone to training instability and training and deploying ESRGAN require substantial computational resources, especially for high-resolution images or large datasets.

Real-ESRGAN (Real-Enhanced Super-Resolution Generative Adversarial Network) is an enhanced version of ESRGAN that addresses its main issues. Real-ERSGAN adopts the same generator architecture as ESRGAN, featuring RRDB that capture intricate details across different scales. Furthermore, the architecture has been expanded to accommodate not only ×4 scale super-resolution but also ×2 and ×1 scales. To reduce computational burden and GPU memory consumption, a preprocessing step using pixel-unshuffle is employed to decrease spatial dimensions and increase channel sizes before inputting them into the main network. The discriminator architecture is redesigned to address a wider range of degradation. The original discriminator is replaced by a U-Net architecture with skip connections, which provides realness values for each pixel, enabling detailed per-pixel feedback to the generator. Spectral normalization regularization is used to enhance training stability and counteract the complexity of the U-Net structure and degradation effects. This regularization technique helps to reduce the artifacts that may be introduced during GAN training. Real-ESRGAN training combines L1 loss, perceptual loss, and GAN loss to enhance local details and suppress artifacts. Perceptual loss captures high-level image features, while GAN loss encourages the generator to produce visually realistic images. Real-ESRGAN achieves significant improvements in addressing various image degradation challenges through its refined architecture and training process. [76]

2.3.3.3 GFPGAN

GFPGAN (Generative Face Priority GAN) is a cutting-edge technique for creating high-resolution images. GFPGAN uses flow-based algorithms within a GAN framework to generate realistic images while preserving key photographic characteristics. To achieve this, GFPGAN uses an adversarial loss function alongside its flow-based approach. This approach learns to

map a simple distribution to a more complex target distribution. normalization (AdaIN) and an adversarial loss function. GFPGAN can produce impressive results with fewer parameters and computational resources compared to other GAN-based methods, making it suitable for real-time applications. Additionally, GFPGAN represents a significant advancement in face image generation. It utilizes a unique generator architecture that combines elements of a U-Net-like network with a GAN architecture. GFPGAN can generate realistic face images from semantic maps that encode detailed facial attributes such as hair, eyes, and nose. The generator network takes a semantic map as input and produces a corresponding feature map. This feature map is then processed through a series of convolutional layers, batch normalization layers, and activation functions within the GAN architecture to produce the final high-quality image output. In contrast, the discriminator network in GFPGAN is a CNN trained to discern between real and generated images using a specific loss function. This setup ensures that the generator produces images that are indistinguishable from real ones, enhancing the overall realism and quality of the generated images. [77]

2.3.3.4 Autoregressive models

Autoregressive statistical models are widely used for predicting future values based on past data. They assume that past values significantly influence current values and are applied in various fields to analyse time-varying phenomena. In autoregressive models, white noise is used to represent random fluctuations in the data. However, these models have limitations when the underlying dynamics change over time, potentially leading to inaccurate predictions. To overcome this, advanced modelling techniques incorporate additional components like moving averages, seasonal patterns, and trend analysis to capture evolving dynamics and improve accuracy. These extended autoregressive models offer better insights into complex systems. [43]

Autoregressive models have several applications in the military field, particularly in areas where time series data analysis, prediction, and decision-making are crucial. These models are designed to capture temporal dependencies within data, making them valuable in military scenarios that involve time-sensitive information and planning. Autoregressive models can be

used to predict when maintenance is required by analysing historical data on equipment performance. This helps prevent unexpected failures and ensures operational readiness. Moreover, Autoregressive models can be used for long-term planning and strategy development. By analysing historical data and trends, these models can provide insights into potential future challenges and opportunities, aiding in strategic decision-making. Additionally, Autoregressive models can analyse and predict patterns in data from various sensors and surveillance systems. They can be used to detect unusual activity or anomalies, enhancing situational awareness and aiding in threat detection.

However, Autoregressive models assume that the statistical properties of the data remain constant over time. In military contexts, this assumption may not hold if conditions change due to evolving threats, technology, or geopolitical factors. Furthermore, the accuracy and reliability of autoregressive models are highly dependent on the quality and completeness of the historical data. Inaccurate or incomplete data can lead to unreliable predictions. Autoregressive models tend to be better suited for short-term predictions, as the accuracy of predictions tends to degrade as the forecast horizon extends further into the future. [43]

2.3.3.5 Generative moment matching networks

Generative Moment Matching Network (GMMN) is a type of generative models that offers a way to assess generalization in the data space and allow for qualitative evaluation. GMMNs utilize a simple prior distribution for easy sampling. This prior is then deterministically propagated through the hidden layers of the NN, and the output represents a sample from the model. This enables GMMNs to quickly generate independent random samples. The core idea behind GMMNs is to use a NN to learn a deterministic mapping from samples of an easily sampleable distribution to model from the data distribution. The architecture of the generative network in GMMNs is similar to that of a GAN. However, instead of using the challenging minimax objective function used in GAN training, it trains the network by simply minimizing the MMD criterion. The generative network consists of a stochastic hidden layer with independent prior uniform distributions for each hidden unit. [66]

In general, Generative Moment Matching (GMM) is a technique used in ML and generative modelling to align the moments (statistical properties like mean and variance) of a generated distribution with those of a target distribution. GMM can be useful in generating synthetic data that closely matches the statistical properties of real-world military data. This can be valuable for augmenting limited datasets and improving ML model performance, such as in image recognition or anomaly detection. Moreover, GMM can assist in creating realistic simulations of military scenarios. By matching the statistical moments of generated scenarios with real historical data, it helps in creating training environments and simulations that closely resemble actual situations, enhancing the effectiveness of military training.

However, real-world military data is often complex and dynamic, with many variables and dependencies. GMM assumes that the data can be described by simple statistical moments. This oversimplification may not capture the full complexity of military scenarios. For this reason, GMM is not typically used for threat assessment in the military field because it usually involves more complex and dynamic considerations beyond statistical moments.

2.3.3.6 Variational autoencoder

A VAE is a NN architecture commonly used in unsupervised ML. Its primary objective is twofold: first, to acquire a concise and continuous representation of data, and second, to generate new data samples that closely resemble the input. An Autoencoder represents a straightforward NN architecture comprising two fundamental components: the encoder and the decoder. The encoder's role is to condense the original input into a compact representation within a significantly smaller vector space. Conversely, the decoder aims to reconstruct the compressed data back into its original form, albeit with some degree of loss. In contrast, a VAE, **Figure 2**, extends this concept by not only compressing the data but also learning its underlying distribution. By leveraging this distribution, the VAE can decode and generate entirely new data points. In the context of VAE, the encoder endeavours to learn parameters that facilitate the compression of input data into a latent vector. The resulting encoding is sampled from a Gaussian density governed by these learned parameters. On the other hand, the decoder takes the encoded representation as input, which is generated by the encoder. It

then parametrizes the reconstruction over a set of parameters. The output is drawn from the distribution representing the original data. The metric utilized for evaluating and refining the model is the VAE loss function, which comprises two essential terms:

- Negative Log-Likelihood of the Decoder: This term assesses how effectively the latent vector is reconstructed into the original data point for each data point. It is essentially the reconstruction error. In the implementation, the built-in binary cross-entropy loss function is commonly employed, where the input is used to approximate the reconstructed output.
- Kullback-Leibler (KL) Divergence: This term acts as a regularizer and measures the information loss incurred when the encoder produces the latent vector. In the VAE framework, the prior distribution is typically chosen as a standard Gaussian distribution. The KL divergence encourages the encoder to generate latent vectors that are close to this Gaussian distribution. This regularization ensures that the latent representations of different types of data are sufficiently diverse yet close to each other in the latent space. By minimizing the KL divergence, the encoder learns to produce meaningful and structured latent representations. [63]

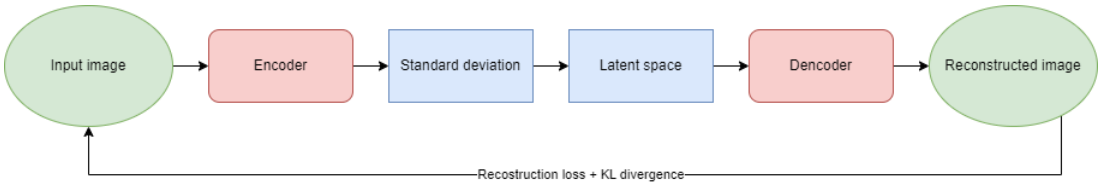

Figure 2: *Variational Autoencoder*

Furthermore, as models undergo the process of learning to generate realistic output, they inherently grasp significant features embedded within the data. These learned features hold valuable information that can potentially be leveraged for classification tasks. This notion is particularly evident in the context of Conditional Variational Autoencoders (CVAE) and semi-supervised learning models. [65]

However, the conventional approach that relies on pixel-level measurement loss often falls short in capturing perceptual differences and spatial correlations between images. To enhance the quality of generated images, an alternative approach prioritizes the preservation of

consistency in hidden representations and spatial correlations between the input and output images. This results in significantly improved image generation that aligns more effectively with human perception. [44]

Integrating high-level feature perceptual loss, derived from pretrained deep CNNs, into VAE frameworks involves optimizing the VAE's loss function to prioritize the preservation of semantically meaningful features extracted by the CNN. [44]Traditionally, VAEs use a loss function that combines a reconstruction loss, typically measured as the pixel-wise difference between the input and reconstructed images, with a regularization term, such as KL divergence, to encourage the latent space to follow a specific distribution. By contrast, high-level feature perceptual loss replaces the pixel-wise reconstruction loss with a loss term that measures the difference between the high-level features extracted from the input and reconstructed images by a pretrained deep CNN. These high-level features represent abstract semantic information about the content and style of the images, allowing the VAE to focus on preserving important visual attributes during image generation. The use of high-level feature perceptual loss enables the VAE to generate images with enhanced perceptual quality. Furthermore, integrating feature perceptual loss facilitates the training of feed-forward networks for real-time style transfer and super-resolution. These networks leverage the learned high-level features to efficiently transform input images, enabling rapid image transformations without sacrificing quality.

Nouveau VAE (NVAE) is a deep hierarchical VAE designed to produce high-quality images through architectural enhancements. NVAE achieves state-of-the-art results among non-autoregressive likelihood-based generative models. The key innovation of NVAE lies in its carefully designed network architecture, with depth wise convolutions serving as the main building block. Depth wise convolutions rapidly increase the receptive field of the network without significantly increasing the parameter count, thereby enhancing its ability to capture complex features in the input images. Additionally, batch normalization is a critical component for the success of deep VAEs, contrary to previous VAE methods. However, training instability issues are present when increasing the number of hierarchical groups, irrespective of batch normalization’s presence. To address this challenge, it is possible to aid a novel residual

parameterization technique for approximate posterior parameters, improving the optimization of the KL divergence term. Furthermore, it's been demonstrating the importance of spectral regularization in stabilizing VAE training. Notably, NVAE represents the first successful application of VAEs to images as large as 256x256 pixels. [64]

VAEs have the potential for various applications in the military field, given their capabilities in generative modelling and representation learning. VAEs can be employed for anomaly detection in military systems. By learning the normal distribution of data, they can detect deviations from the norm, which might indicate security breaches, equipment malfunctions, or unusual activities. Moreover, VAEs can generate synthetic data that closely resembles real-world military scenarios and synthetic training environments. This is useful for creating realistic simulations for training purposes, such as flight simulations, battlefield scenarios, or medical training simulations. These simulations help in skill development, tactical training, and decision-making exercises, providing a safe and controlled learning environment. Additionally, VAEs can assist in identifying and classifying targets, whether they are enemy combatants, vehicles, or objects of interest. They can enhance the accuracy of target recognition systems.

Nevertheless, VAEs tend to produce slightly blurry images. This is because the model generates samples from the latent space, which can result in a loss of fine details and sharpness in the generated images. This blurriness can be a drawback when generating high-fidelity images. [65] Moreover, Training VAEs can be computationally intensive and require careful tuning of hyperparameters. The training process involves optimization over a complex loss function, and achieving a good balance between the reconstruction loss and the regularization term can be challenging.

2.3.3.7 Normalizing flow

Normalizing flows (NFs) are a sequence of simple functions that can be inverted or have an analytical inverse. These flows transform complex data points into simple Gaussian Distributions, and vice versa. Unlike GANs, where the generator is trained to produce images from random vectors, flow-based models transform data points into simple distributions during training. These models are trained using the negative log-likelihood loss function as the

probability function. The loss function is derived using the change of variables formula from basic statistics. Flow-based models do not require noise on the output, allowing for powerful local variance models. The training process of flow-based models is more stable compared to GANs, which necessitate careful tuning of hyperparameters for both generators and discriminators. NFs also converge more easily. However, the quality of samples generated by flow-based models is not as good as those produced by GANs and VAEs. [45]

The density of a sample can be determined by transforming it back to the original simple distribution and calculating the product of the density of the sample after the inverse transformation under this distribution and the change in volume induced by the sequence of inverse transformations. This methodology enables the creation of new distribution families by selecting an initial density and subsequently linking together several parameterized, invertible, and differentiable transformations. For NFs to be practical, they must fulfil several criteria: they should be invertible, possess enough expressiveness to adequately model the desired distribution, and maintain computational efficiency. [46]

NFs can generate synthetic data that closely matches real-world military scenarios. This is particularly valuable for creating realistic training simulations and scenarios. It allows military personnel to practice in a controlled and lifelike environment, improving their readiness and preparedness for various situations. NFs can be used for anomaly detection in military systems. By learning the normal distribution of data, they can identify deviations or anomalies in real-time, which can be indicative of security breaches or equipment malfunctions. Moreover, in military operations, accurately identifying targets, such as enemy combatants, vehicles, or objects of interest, is crucial. NFs can assist in target recognition systems by providing a generative model that can separate normal environmental conditions from potential threats. However, the complexity of NF models can make them harder to maintain and adapt in the field, where the availability of expertise and resources for model management may be limited. Moreover, validating the quality and accuracy of data generated by NF can be challenging, particularly when ground truth data is limited, and this can reduce the stability of the model.

2.3.3.8 Stable diffusion

Diffusion models are a class of generative models that aim to generate images by modelling the process of diffusion. They provide a powerful framework for text-to-image generation tasks, yielding cutting-edge results. Diffusion models work by iteratively adding noise to an initial image until it becomes a sample from the target distribution. This noise addition occurs in a latent space, where the model manipulates representations of the image to gradually transform it into the desired output. Stable diffusion models use compressed latent space to optimise memory usage and computation time. Instead of working directly in pixel space, latent diffusion trains the model to generate compressed representations of images, improving efficiency. The process involves progressively adding noise to an image, transforming it into a noisy representation until it approximates pure noise. [73] Diffusion models are built from a hierarchy of denoising autoencoders. Additionally, they are likelihood-based models, so they do not exhibit mode-collapse and training instabilities as GANs and, by heavily exploiting parameter sharing, they can model highly complex distributions of natural images without involving billions of parameters as in autoregressive models. [72]

Stable Diffusion uses latent images derived from training data as input, with the algorithm dynamically adding noise to create noisy images. By incorporating inputs such as time steps and text prompts, image diffusion algorithms can effectively learn to predict the noise added at each step, resulting in refined image generation. In latent diffusion, several key components play a crucial role in the image generation process. These include the auto-encoder, the U-net and the text-encoder. The auto-encoder model consists of an encoder and a decoder. During training, the encoder reduces a image to a lower dimensional latent representation for forward diffusion. At each training step, increasing levels of noise are applied to these latents, which serve as input to the U-net model. This process significantly reduces memory requirements, with a 48-fold reduction compared to pixel-space diffusion models. The decoder then reconstructs the latent representation back into an image. During inference, the VAE decoder transforms the denoised image into its true form. Meanwhile, the U-Net predicts the denoised image representation from noisy latent. A conditional model, informed by time step and text embedding, guides the process. The text encoder converts input prompts into

embeddings that facilitate U-net guidance. Typically, a simple transformer-based encoder such as CLIP is used for this purpose. This comprehensive architecture ensures effective noise reduction and accurate image generation in latent diffusion models. [73]

2.3.3.9 SWIN IR

CNNs have become one of the most used approaches to image restoration tasks. CNN-based methods have made significant advances by exploiting techniques such as residual learning and dense connections; however they still face inherent limitations due to their basic building block, the convolutional layer. First, the interactions between images and convolution kernels are content-independent, meaning that the same convolution kernel is applied to all image regions regardless of their content. This lack of content awareness can lead to suboptimal results. Secondly, due to their local processing nature, convolutions struggle to effectively capture long-range dependencies within images.

In contrast, Transformer architectures, with their self-attention mechanism, offer a promising solution to image restoring by capturing global interactions between contexts. However, traditional vision transformers for image restoration typically process input images by dividing them into fixed-size patches and processing each patch independently, limiting their ability to effectively capture spatial dependencies. Swin Transformer is a novel architecture that combines the strengths of both CNNs and transformers. In particular, it exploits the ability of CNNs to handle large images with local attention mechanisms, and the ability of transformers to model long-range dependencies through a shifted windowing scheme. Swin Image Resolution (SwinIR), an image restoration application of Swin Transformer, consists of three main modules. Shallow Feature Extraction extracts shallow features using a convolutional layer, preserving low frequency information that is critical for image reconstruction. Deep Feature Extraction consists of Residual Swin Transformer Blocks (RSTB), each of which contains multiple Swin Transformer layers for local attention and cross-window interaction. Additional convolutional layers enhance features, with residual connections facilitating feature aggregation. High quality image reconstruction, where both shallow and deep features are fused to facilitate high quality image reconstruction. SwinIR offers several advantages,

including content-aware interactions between image content and attention weights, effective long-range dependency modelling via the shifted window mechanism, and achieving superior performance with fewer parameters. [78]

2.3.3.10 LIIF

The Local Implicit Image Function (LIIF) is a novel method proposed to represent natural and complex images in a continuous manner. Traditionally, images are represented as 2D arrays of pixels, which limits the resolution and fidelity of the representation. LIIF aims to overcome this limitation by modelling images as functions defined in a continuous domain, allowing for arbitrary resolution and high fidelity. Inspired by recent advances in implicit neural representation for 3D shape reconstruction, LIIF represents an image as a function mapping coordinate to RGB values, where the function is parameterised by a DNN. Unlike previous encoder-based methods that have struggled to represent complex images, LIIF introduces the concept of local implicit image functions. In LIIF, an image is represented as a set of latent codes distributed in spatial dimensions. Given a coordinate, the decoding function queries the local latent codes around the coordinate and predicts the RGB value at that coordinate. This continuous representation allows LIIF to generate images of arbitrary resolution. To train the LIIF framework, an encoder is trained via a self-supervised super-resolution task. During training, a single image is down sampled to generate an input, and a ground truth representation is obtained by representing the image as pixel samples. The encoder maps the input image onto a 2D feature map as its LIIF representation. The coordinates of the ground-truth pixels are then used to query the LIIF representation, and the decoding function predicts the RGB values for each coordinate. A training loss, such as the L1 loss, is calculated between the predicted and ground truth RGB values. [79]

LIIF bridges the gap between discrete and continuous image representations, providing a novel approach to image representation that can generalise to higher resolutions while maintaining high fidelity. Its applications extend to image generation, super-resolution and other image-to-image translation tasks.

2.3.3.11 SUPIR

Scaling-Up Image Restoration (SUPIR) is at the forefront of cutting-edge image restoration technology, promising a leap forward in visual effects and intelligence. At its core, SUPIR relies on the power of generative priors, in particular StableDiffusion-XL (SDXL), which boasts a staggering 2.6 billion parameters. This is a deliberate choice, as SDXL provides a direct path to generating high-resolution images without the constraints of hierarchical design, which fits perfectly with SUPIR's goals of improving image quality. However, the use of such a powerful generative prior requires an equally sophisticated adaptor capable of exploiting SDXL's potential to recover images from low-quality inputs. SUPIR is trained on over 20 million high-quality, high-resolution images, each accompanied by detailed descriptive text. However, SUPIR takes an unconventional approach by including low-quality, negative samples in its training dataset, a strategy that surprisingly improves visual effects. This method, combined with restoration-guided sampling, ensures fidelity to the input image - a crucial aspect in maintaining the authenticity of the restoration process. Scaling image restoration models presents its own set of challenges, from fine-tuning image encoders to designing efficient SDXL-compatible adapters. These obstacles are met with innovative solutions to ensure that SUPIR not only scales effectively, but also maintains stability and fidelity in the restoration process. [80]

2.4 Scenario generation

In general, the notion “scenario” could be defined as a synthetic description of an event or series of actions and events about the future. The “plausible future” determined within a set of scenarios can encompass different areas of defence capabilities: security policy development, operations, training. The scenario generation is based on initial featured experts’ opinions and believes usage as “information input”. As far as the information of that kind could be considered as rather subjective, techniques like: brainstorming, back casting, workshop method, roundtables, discussions and questionnaires fill-up are used for the initial information gathering supported with tools for group work like: flipcharts, whiteboards, multimedia, etc. [47]

Scenario generation process is mainly composed by 5 steps:

- Preparation: this step includes definition of the time horizon, experts' team formation, goals definition, database creation, methodological preparation, scenarios' security level and time schedule definition.
- Strategic Base Analysis: it can be conducted over the whole spectrum of national security and in accordance with the scope of the strategic base a concrete focus for the analysis should be determined.
- Analysis of the Characteristics of the Future: analysis and selection of the most important characteristics, which are significant for the decision-making process in the planned "plausible future", are performed. This step aims at narrowing the scenario development field in a reasonable way and, at the same time, producing a scenario explanation of the future projection.
- Definition of Zones of Security Interests: it enables the establishment of a clear geopolitical foundation for the development of the scenarios.
- Development and Analysis of the Scenarios: it is a complex task and it takes into consideration the definitions of Step 4 and is implemented in thirteen sub-steps: selection of main dimensions (key factors), definition and selection of alternatives for each scenario dimension, linking alternatives, scenario entitling, scenario system evaluation, scenario logic selection, scenario wild-cards analysis, scenario text elaboration, development of scenario portfolio, scenario validation, scenario approval, scenario presentation, implementation of the scenario. [47]

A scalable scenario generation system is a crucial component in training and skill development. Such a system should encompass several critical capabilities, each contributing to the overall effectiveness of the training process. First and foremost, the concept of "replayability" is essential. Mastery of skills often hinges on the ability to apply those skills across a broad spectrum of realistic scenarios. Frequent training is crucial, but there is a limit to the benefits of repeated exposure to the same scenario. When trainees encounter the same situations repeatedly, they may inadvertently memorize the scenarios themselves, rather than truly grasping the underlying concepts. Therefore, to mitigate the problem of diminishing returns from repetitive scenarios, the scenario generation system must possess the capability

to create a multitude of distinct variations from a given set of input parameters. This ensures that trainees experience a wide range of situations, preventing rote memorization and fostering the development of adaptable skills. Furthermore, the system should have the ability to tailor scenarios to individual needs. Training scenarios are often designed with an "average" trainee in mind, which can lead to suboptimal experiences for both above-average and below-average individuals. The Zone of Proximal Development (ZPD) theory emphasizes that learners excel when challenged at an appropriate level. By customizing scenarios to individual or small group requirements, the system can effectively address the unique needs and abilities of each trainee. It may involve assessing prior performance, adjusting difficulty levels, or even customizing scenarios to match different learning styles, thereby ensuring that each trainee operates within their ZPD and experiences more effective training. Finally, the adaptability of the scenario generation system to changing conditions in the world is crucial. The world is in a constant state of flux, with new tactics, techniques, and procedures emerging to respond to evolving challenges. The scenarios used for training can quickly become outdated and less effective. Therefore, the scenario generation system should be designed to be easily reconfigurable to accommodate changes in training objectives, missions, and challenges. This adaptability relies on having knowledge structures that can be updated swiftly to incorporate the latest information, tactics, and procedures. It ensures that the training remains up-to-date and can accurately simulate real-world conditions. These capabilities collectively facilitate skill development, ensure the relevance of training, and enhance the overall effectiveness of the training process. It is a versatile tool that can cater to the ever-evolving demands of the training and development landscape. [48]

Focusing on military field, the overarching objective of the scenario generation system is to meticulously craft a sequence of events that serves three paramount purposes. Firstly, the narrative generated should authentically simulates a realistic mission. Secondly, the system should be designed to meticulously construct scenarios that align with a set of training objectives, properly tailored to the specific needs and skill levels of each individual trainee. Lastly, the system should aspire to diversify the training landscape by generating a multitude of distinct scenarios, ensuring that training can be conducted repeatedly without tedious

repetition. Overall, the scenario generation system endeavours to strike a harmonious balance between realism, individualized learning objectives, and variety, thereby fostering a highly effective and engaging training environment.

2.4.1 Methodologies

Modelling and simulation are sophisticated techniques used to generate scenarios in various fields, including military operations. These methods involve creating virtual representations of complex systems and running simulations to predict their behaviour under different conditions. In the context of military planning, modelling and simulation play a crucial role in training, decision support and research efforts.

At the core of modelling and simulation is the ability to construct mathematical or computational models that capture the dynamics and interactions within a given system. These models can range from simple representations to highly detailed, intricate simulations, depending on the complexity of the system and the level of fidelity required. Once a model has been developed, it can be used to simulate different scenarios by adjusting input parameters or initial conditions. [41] In military planning, for example, scenarios might include changes in enemy tactics, changes in terrain characteristics, or changes in logistical constraints. By running simulations under these different scenarios, planners can explore potential outcomes and assess the effectiveness of different strategies and courses of action.

Modelling and simulation techniques offer several advantages for scenario generation in a military context. First, they provide a controlled environment in which planners can test hypotheses and evaluate alternative courses of action without the risks associated with real-world experimentation. In addition, simulations can be conducted quickly and at relatively low cost compared to field exercises or live training events. Modelling and simulation also allow planners to consider a wide range of factors simultaneously, including both deterministic variables and stochastic elements. On the other hand, modelling and simulation are resource and time consuming and can be inaccurate if model simplifications are made. They can also be imprecise due to poor data quality and unaccounted for human factors. Over-reliance and

high development costs are concerns, as are ethical and security issues and potential unintended consequences.

Text and image prompts can be used to generate scenarios through advanced ML models, such as GL, to quickly produce diverse and creative scenarios. This approach has several key advantages. First, generative models excel at creativity and diversity, allowing a wide range of scenarios to be explored from simple prompts. This ability to generate multiple possibilities is particularly valuable in areas such as military training, game design and strategic planning, where a rich variety of scenarios can provide deeper insights and more engaging experiences. Speed and efficiency are also key benefits. These models can generate scenarios quickly, making them ideal for applications that require rapid prototyping or iterative development. For example, in disaster preparedness or strategic military planning, the ability to quickly generate multiple potential scenarios helps planners and decision makers prepare for a wide range of possibilities. Customisation is another key benefit. By adjusting the input prompts, users can tailor the generated scenarios to meet specific needs or conditions. This flexibility makes generative models highly adaptable to different contexts and requirements. Generative models are also able to handle complex relationships and nuanced details, producing realistic and intricate scenarios that capture multiple facets of real-world situations. Despite these advantages, there are some challenges to consider. The quality and accuracy of generated scenarios can vary, the process is resource-intensive, and models can perpetuate biases present in their training data. In addition, careful validation is required to ensure the validity and reliability of generated scenarios.

2.5 Threat assessment

A threat is a potential harm that can occur due to a vulnerability that exists in various scenarios. The key feature of a threat is that it is uncertain. There is no certainty of what will happen, when it will happen, how bad the consequences will be, how long it will last, and of the probability which it will happen. Therefore, all possible threats to the system must be identified and their consequences neutralized or mitigated, if fundamental. To do this, especially in military field where the occurrence of threat can be catastrophic, threat assessment techniques

are used. Assessment is the process of gathering information for the use in making decisions. Hence, threat assessment can be defined as the practice of determining the credibility and severity of a potential threat, as well as the likelihood that the threat will be a reality, using the information available on the environment. [49]

Risk assessment plays a vital role in today's military operational planning, and threat assessment is central to this process. However, threat assessment is not a process without complications and problems. The military context is one of uncertainty, which means that in analysis cannot be limited to mathematical aspects.

Identifying potential threats is a complex task influenced by various factors. Institutions, particularly in the military field, play a significant role in shaping threat assessments by providing guidelines for organized human interaction across different domains like families, governments, businesses, and religions. These institutions are essentially shared concepts existing in participants' minds, not always explicitly known. Cognitive influence adds another layer of complexity, making prediction and assessment challenging. Risk, being dynamic and immeasurable, requires continuous evaluation. Although not constant, a reasonable representation of risk can be derived through a thorough risk analysis based on accurate threat assessments, probability assessments, and objective descriptions of consequences. This analysis guides risk management decisions, especially in military operations where decision options and potential negative outcomes are identified and evaluated. Choosing the alternative with the lowest expected risk is often rational when probabilities are comparable. However, challenges arise in assessing probabilities and comparing consequences, particularly in analyses focusing on conceptual and psychological outcomes. [50]

To make informed decisions about potential risks, having access to a wealth of data is crucial. It's like having a detailed map to navigate through uncertainties. The more information available, the better equipped we are to assess the likelihood of threats and understand the potential consequences. This data serves as a foundation for risk assessments, allowing us to analyse probabilities with more accuracy and to objectively compare the potential outcomes. It's akin to having a comprehensive toolkit that enables a thorough evaluation of various factors

influencing risk. Without ample data, the risk assessment process becomes challenging, and decisions may be less informed and more susceptible to uncertainties.

Conducting risk assessment in situations where there is a lack of data, particularly in sensitive areas such as the military, can be challenging. In such scenarios, it is necessary to rely on alternative methods and strategies to gather relevant information and make informed assessments. Traditional methods such as relying on expert judgement, conducting red team exercises, and analysing historical data form the basis of military threat assessment. Subject matter experts from the military or related fields provide invaluable insights based on their experience and contribute to a qualitative understanding of potential risks. Red team exercises simulate adversarial roles and systematically identify weaknesses and vulnerabilities, while historical analysis, even when data is limited, can provide critical insights from declassified information or historical accounts. Scenario planning is another traditional method that allows military strategists to develop hypothetical situations based on available information and expert insights. This forward-thinking approach facilitates the exploration of various potential risks and their potential impacts on military operations. Simulation and modelling techniques further contribute by allowing the testing of different scenarios, providing a nuanced understanding of potential consequences and assessing the resilience of military systems. In data-scarce situations, GL techniques such as AI, natural language processing, and anomaly detection greatly improve military risk assessment. These techniques make it easier to create artificial data for training, spot irregular patterns that indicate potential threats, create a variety of threat scenarios for training and readiness, improve current data through variation, apply knowledge from related tasks to improve threat recognition, summarize complex information for effective threat communication, work with generative models and human analysts to create scenarios, update threat models continuously based on new information, and visualize threat landscapes through immersive simulations. An approach to military threat analysis that is more thorough and flexible is offered by integrating these techniques.

Threat detection is the crucial first step in threat assessment, where risks in images are identified through segmentation and labelling. However, transitioning from threat detection to threat assessment involves moving beyond simply identifying potential risks in images to

evaluating their context, severity, and potential impact. This process includes analysing extracted features, understanding the surrounding environment, assessing risk levels, and making informed decisions based on the findings. It's a structured approach that combines technical analysis with contextual understanding to effectively manage and respond to identified threats. Achieving effective threat assessment often necessitates the use of sophisticated algorithms and methodologies, underscoring the importance of high-quality data.

2.5.1 Segmentation Methodologies

Image segmentation plays a pivotal role in breaking down these intricate synthetic images into meaningful segments or regions. This process involves dividing the image into distinct and semantically meaningful parts, such as objects, backgrounds, or specific features. This task can be completed by two procedures. One method is segmentation comparison, which ranks how well various approaches perform when segmenting the same kind of pictures. It is an inter-technique procedure. An additional method is segmentation characterisation, which is an intra-technique procedure for identifying how the technique under consideration behaves when segmenting different types of images. [53] By segmenting synthetic images, we gain a more granular understanding of the content within them, which is crucial for accurate threat assessment and object identification.

Three distinct groups of evaluation methods exist: analytical methods, goodness methods, and discrepancy methods. Analytical methods directly address segmentation algorithms by examining their principles, requirements, utilities, and complexity. While this approach seems straightforward, analytical studies alone may not capture all the properties of segmentation algorithms. Both analysis and practical experience indicate that analytical methods can only augment information provided by other methods and are rarely employed in isolation. Goodness methods assess algorithm performance indirectly by evaluating segmented images using predefined quality measures based on human intuition. Numerous goodness measures have been proposed to encompass various aspects of an "ideal" or "good" segmentation. These measures provide a means to judge the effectiveness of segmentation algorithms by comparing the segmented images to qualitative standards. Discrepancy methods quantify the

disparity between an segmented image and an ideally segmented image, often referred to as the reference image, gold standard, or ground truth. These methods aim to gauge how closely the actually segmented image aligns with the reference image. In essence, discrepancy methods seek to determine the extent of deviation between the actual and ideal segmentation outcomes. [53]

2.5.1.1 R-CNN

Object detection finds widespread application in intelligent surveillance, automatic driving, surgical instrument positioning, and various other domains. Its primary objective is to discern both the classification and location details of a specified object within complex scenes. This information proves invaluable for intricate tasks like subsequent object tracking. Notably, object detection goes beyond merely identifying object classification and positioning; it also involves determining the quantity and size of objects. Within the realm of object detection models employing DL, two prominent classes emerge: regression/classification-based methods and region proposal-based methods. These models play a pivotal role in addressing the multifaceted challenges associated with object detection tasks. The integration of DL into object detection marked a significant advancement, notably with the introduction of the Region-Convolution Neural Network (R-CNN). As a region proposal-based method, R-CNN pioneered object adaptive detection, showcasing the capacity of DL to enhance the precision and efficiency of object detection in diverse applications. [54]

The region-based CNN framework has revolutionised object detection and semantic segmentation in computer vision. R-CNN proposes a multi-stage approach to the challenge of accurately localising and classifying objects within images. In the first stage, region proposals are generated using a method such as selective search, which efficiently identifies potential object locations within the image. These proposals serve as candidate regions for further analysis. In the second stage, each proposed region is independently processed by a pre-trained CNN to extract rich feature representations. This CNN acts as a feature extractor and transforms the raw image data into a high-dimensional feature vector for each region. These feature vectors are then fed into a set of Support Vector Machines (SVMs), which are trained

to classify the content of each region into predefined classes. In addition, bounding box regressors are trained to refine the proposed regions, thereby improving localisation accuracy. Finally, a post-processing step called non-maximum suppression is applied to merge highly overlapping bounding boxes and produce the final set of detected objects with their corresponding class labels. This multi-stage approach of region proposal, feature extraction, classification and bounding box refinement enables R-CNN to achieve state-of-the-art performance in object detection and semantic segmentation tasks, marking a significant advancement in the field of computer vision. [74]

Mask R-CNN showed in **Figure 3**, signifying a conceptual extension of Faster R-CNN, enriches its capabilities by introducing a third branch dedicated to predicting segmentation masks for each candidate object. Essentially, Faster R-CNN initially provides class labels and bounding-box offsets for each candidate object, and Mask R-CNN seamlessly incorporates an additional branch to furnish object masks. This intuitive extension facilitates a finer spatial layout extraction, a pivotal requirement for accurate segmentation. The foundational structure of Mask R-CNN derives from the two-stage framework of Faster R-CNN. The initial stage involves the Region Proposal Network (RPN), proposing bounding boxes for candidate objects. Subsequently, akin to Fast R-CNN, the second stage extracts features using Region of Interest (RoI) Pooling from each candidate box, facilitating classification and bounding-box regression. A distinctive feature of Mask R-CNN's second stage is its concurrent output of binary masks for each RoI, deviating from recent systems where classification depends solely on mask predictions. During the training phase, Mask R-CNN employs a multi-task loss on each sampled RoI. The classification loss and bounding-box loss echo those in Faster R-CNN. The mask branch produces a Km^2-dimensional output for each RoI, encoding K binary masks of resolution m × m for each of the K classes. The per-pixel sigmoid, applied to the output, contributes to the average binary cross-entropy loss. The distinctive aspect of Mask R-CNN lies in its approach to mask representation, preserving the explicit m × m spatial layout without losing spatial dimensions. The fully convolutional representation, using an FCN, demands fewer parameters compared to methods relying on fully connected layers for mask prediction.

The pixel-to-pixel behavior, crucial for spatial correspondence, is facilitated by the RoIAlign layer. [55]

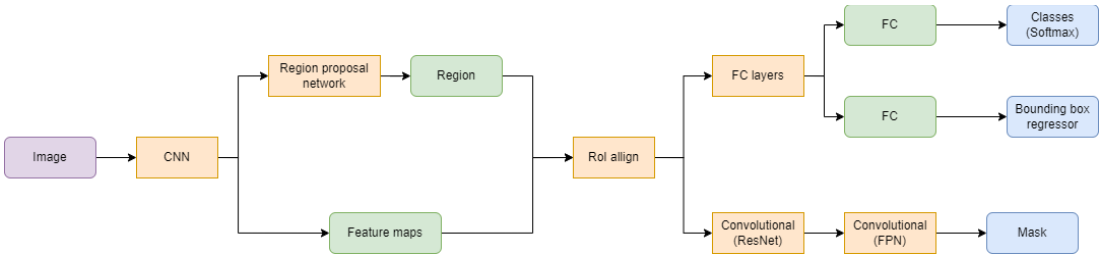

Figure 3: *Mask R-CNN*

Here, the integration of additional information comes into play. The use of ResNet as the backbone is a strategic move to address concerns related to network depth, calculation, and parameter quantity. Additionally, Group Normalization (GN) is introduced to enhance detection accuracy, and cascade training with IoU thresholds is employed to further refine detector performance. This proposed algorithm strategically tackles the limitations of Mask R-CNN, ensuring accurate bounding box and mask information for subsequent classification and regression tasks. The combination of Mask R-CNN's innate capabilities with these enhancements underscores the adaptability and effectiveness of this comprehensive approach. [54]

2.5.2 Labelling Methodologies

The process of image labelling stands as a critical phase in the analysis of synthetic images. It's important to note that the labelling phase occurs after the segmentation process, focusing on the segmented images and various objects within them. This sequential approach allows for a more targeted and refined analysis of individual objects within the synthetic environment. Labelling synthetic images offers a comprehensive understanding of their content, a crucial factor for accurate threat assessment and object identification. The segmentation of these synthetic images through labelling addresses the challenge of intricate details by isolating individual objects or elements. This not only facilitates a more precise analysis of each labelled segment but also enables a focused evaluation of potential threats within the synthetic environment. Labelling in image analysis offers unique advantages. It enhances the interpretation of image content by associating meaningful categories, aids in precise

identification and classification of individual objects, adds context to segmented elements for better scenario analysis, serves as valuable training data for ML models, generates easily understandable results for collaboration with experts, supports informed decision-making across various domains, extracts actionable information tailored to specific use cases, and contributes to visually informative representations for improved analysis.

In the realm of real-world image recognition systems, the triumvirate of success revolves around three pivotal factors: firstly, the creation of representative image features; secondly, the design and implementation of potent learning methods; and thirdly, the acquisition of vast amounts of training data. The crafting of representative image features holds paramount importance as it is instrumental in effectively describing and differentiating various entities within the visual data. One of the primary challenges addressed in this context involves the discernment of relevant and informative data from the noise present in the dataset. The intricate task of separating signal from noise is vital to ensure that the image features extracted truly encapsulate the distinguishing characteristics of the entities under consideration. This involves the implementation of sophisticated algorithms and methodologies capable of filtering out extraneous information, thereby enhancing the quality of the image features. Simultaneously, the effective utilization of the massive amount of available data is critical to the success of the image recognition system. The development of an efficient end-to-end pipeline is essential to streamline the process of feature extraction, learning, and recognition. This pipeline ensures a seamless integration of various stages, optimizing the utilization of computational resources and facilitating a coherent flow of information from raw data to actionable insights. In essence, the journey towards a robust real-world image recognition system involves navigating the challenges of extracting representative image features, developing advanced learning methods, and orchestrating the handling of extensive training data. It is the synergy of these elements that propels image recognition systems towards accuracy, efficiency, and applicability in diverse practical scenarios. [56]

The primary computational challenge encountered in object recognition lies in addressing the inherent problem of variability. A robust vision system must effectively generalize across extensive variations in the appearance of an object, such as changes in viewpoint, illumination,

or occlusions. Simultaneously, it must uphold specificity to accurately identify and categorize objects. Within the realm of recognition, two fundamental tasks emerge: identification and categorization. Typically, computer vision techniques can easily achieve identification, while categorization poses a more formidable challenge. Categorization demands a broader generalization, encompassing not only diverse viewing conditions but also different exemplars within a class. The complexity of the task is not inherently tied to the nature of identification versus categorization but hinges on parameters like the size and composition of the training set. The extent to which the training examples cover the required variability for effective generalization significantly influences the difficulty of the recognition task. [57]

2.5.2.1 CNN

In the realm of labelling within computer vision, CNNs play a fundamental role in extracting relevant features and making predictions based on visual data. A CNN is a specialized type of NN designed to process grid-like data, such as images. Its architecture includes convolutional layers that apply filters to input data, enabling the network to automatically learn hierarchical representations of features. For labelling tasks, a CNN typically consists of multiple layers, including convolutional layers for feature extraction and pooling layers for down sampling. The network learns to recognize patterns, shapes, and textures at different levels of abstraction. In the context of multi-label classification, where each image can be associated with multiple labels, a common approach involves transforming the problem into several single-label classification tasks. During training, the CNN learns to map input images to a set of labels, optimizing its parameters using techniques like ranking loss or cross-entropy loss. However, it's crucial to address the challenge of capturing dependencies between multiple labels effectively. To overcome this, joint image/label embedding techniques are employed. These methods enable the network to learn a cohesive representation of images and their associated labels, enhancing its ability to recognize and predict multiple labels in a coherent manner. Overall, CNNs serve as powerful tools for automating labelling tasks by learning intricate patterns and relationships within visual data. [58]

An innovative solution for multi-label classification is the Hypotheses-CNNPooling (HCP) structure, a flexible deep CNN architecture. HCP efficiently processes an arbitrary number of object segment hypotheses, potentially generated by advanced objectiveness detection techniques like Binarized Normed Gradients or Edge Boxes. Each hypothesis is seamlessly connected to a shared CNN, and a novel pooling layer is introduced for aggregating single-label CNN predictions into multi-label results. The HCP infrastructure brings several advantages. Primary, HCP does not demand ground-truth bounding box information during training on multi-label image datasets. This departure from traditional methods reduces the annotation burden, enhancing generalization across diverse multi-label image datasets. Additionally, to address potentially noisy hypotheses, HCP employs a cross-hypothesis max-pooling operation, effectively suppressing noise and discarding redundant hypotheses. This robustness ensures reliable performance even in the presence of imperfect input hypotheses. Moreover, the shared CNN within HCP is flexible, allowing pre-training with large-scale single-label image datasets like ImageNet. Fine-tuning on the target multi-label dataset is facilitated. The architecture accommodates various advanced CNNs, providing adaptability to different network structures. Lastly, HCP's outputs, processed through the softmax layer, result in normalized probability distributions over labels. The predicted probability values inherently serve as the final classification confidence for corresponding categories, ensuring a robust multi-label prediction outcome. The HCP deep network is designed to handle multi-label image classification, addressing challenges such as the absence of ground-truth bounding box information and the need for robustness to noisy or redundant hypotheses. In the first phase, known as Hypotheses Extraction, objectless detection techniques like Binarized Normed Gradients or Edge Boxes generate a set of candidate object windows. To manage computational resources effectively, a subset of these candidates is selected as hypotheses using a dedicated method. The selected hypotheses are then input into a shared CNN. The Training phase involves pre-training the shared CNN on a large-scale single-label dataset, such as ImageNet. Fine-tuning is subsequently performed on the target multi-label dataset (e.g., Pascal VOC), utilizing the entire image for this purpose. Hypotheses-fine-tuning (HFT) is crucial and involves feeding all hypotheses of a given image into the shared CNN. A cross-

hypothesis max-pooling operation is applied to suppress noisy hypotheses and produce an integrative prediction. For Multi-label Classification on a test image, the generated hypotheses undergo the shared CNN to obtain c-dimensional predictive results. The final prediction is then derived through cross-hypothesis max-pooling, accompanied by softmax. This step ensures that high responses, indicative of predicted labels, are preserved. The overall architecture of HCP demonstrates a flexible and efficient approach, particularly notable for its ability to handle scenarios where ground-truth bounding box information is unavailable and to mitigate the impact of noisy or redundant hypotheses. The training strategy, including pre-training and fine-tuning, contributes to the adaptability of the model across different datasets. [59]

Another innovative solution proposes the Neural Abstraction Pyramid (NAP), a hierarchical Recurrent Neural Network designed for image processing. NAP draws inspiration from biology, incorporating both vertical and lateral recurrent connectivity. This architecture gradually refines image interpretation to address visual ambiguities. RCNN, or Recurrent CNN, stands out due to its unique relationship with certain sparse coding models. The use of fixed-point updates in inference in these models implicitly defines recurrent neural networks. The Recurrent Convolutional Layer (RCL) serves as the cornerstone of RCNN, comprising a stack of RCLs, optionally interspersed with max-pooling layers. The computational advantages of the recurrent connections in RCNN are notable. They enable each unit to integrate context information from an arbitrarily large region in the current layer. With increasing time steps, the influence of each unit expands, encompassing a larger neighborhood in the current layer. This results in an enhanced ability to capture spatial relationships in the input space. It's worth mentioning that simply increasing the depth of a CNN by sharing weights between layers may yield a model with the same depth and number of parameters as RCNN. However, empirical evidence suggests that such a model might not achieve comparable performance to RCNN, as demonstrated in experiments conducted. [60]

2.5.2.2 CLIP

Large-scale visual-language pre-training models, exemplified by CLIP, excel at capturing rich and expressive features in both visual and language domains. The utilization of raw CLIP

features for zero-shot image classification proves to be a robust and competitive strategy, demonstrating performance comparable to fully-supervised counterparts. What sets CLIP apart is its departure from traditional pre-training tasks that focus on iconic images. Instead, CLIP learns from images of complex scenes and their accompanying natural language descriptions. This unique learning paradigm encourages the embedding of local image semantics in its features. Moreover, it empowers CLIP to learn concepts in an open vocabulary, accommodating a wide range of objects and capturing rich contextual information. Notably, CLIP's ability to grasp the co-occurrence and relations of certain objects, along with spatial priors, contributes to its versatility in diverse tasks. CLIP functions as a visual-language pre-training method, leveraging a vast dataset of raw web-curated image-text pairs. Its architecture consists of an image encoder and a text encoder, both jointly trained to map input images and text into a unified representation space. The training objective employs contrastive learning, treating ground-truth image-text pairs as positives and creating negatives from mismatched image-text combinations. CLIP offers two alternative implementations: a Transformer and a ResNet with a global attention pooling layer. In the realm of segmentation network training, the prevalent approach involves a three-step process. Firstly, the backbone network is initialized with pre-trained weights from ImageNet. Subsequently, segmentation-specific network modules with randomly initialized weights are added. Finally, the joint fine-tuning of the backbone and newly added modules takes place, resulting in a cohesive and effective segmentation model. [61]

MaskCLIP simplifies the process of extracting dense patch-level features from CLIP's image encoder while preserving the integrity of visual-language associations. Leveraging the value features of the last attention layer, MaskCLIP acquires classification weights for dense prediction directly from CLIP's text encoder embeddings, utilizing 1×1 convolutions without the need for explicit mapping. Compatibility extends to all CLIP variants, encompassing ResNets and ViTs. To enhance MaskCLIP's performance, two non-training-dependent mask refinement techniques are introduced: key smoothing and prompt denoising. Key smoothing involves computing the similarity between the key features of different patches, contributing to the smoothing of predictions. Prompt denoising strategically removes prompts associated with

classes unlikely to exist in the image, reducing distractors and enhancing prediction accuracy. To overcome architectural constraints and accommodate advanced structures like PSPNet and DeepLab, a novel approach is proposed. Instead of deploying MaskCLIP at inference time, it is integrated into the training process, functioning as a versatile and robust annotator that furnishes high-quality pseudo labels. Harnessing the generality and robustness of CLIP features, MaskCLIP+ readily adapts to various semantic segmentation scenarios, including fine-grained class segmentation, novel concept segmentation, and the segmentation of moderately corrupted inputs. Functioning as a segmentation annotator, MaskCLIP provides rich and novel supervision signals, particularly beneficial for segmentation methods dealing with limited labels. Retaining the visual-language association of CLIP endows MaskCLIP with the innate ability to segment open vocabulary classes and fine-grained classes described by free-form phrases. The robustness exhibited by CLIP against natural distribution shifts and input corruptions is preserved in MaskCLIP, contributing to its reliability in various applications. Expanding its versatility, MaskCLIP+ transcends its role in annotation-free and open-vocabulary segmentation. It demonstrates efficacy in the transductive zero-shot semantic segmentation task, generating pseudo labels exclusively for unseen classes. Leveraging the generality and robustness of CLIP features, MaskCLIP+ seamlessly adapts to diverse semantic segmentation scenarios, encompassing fine-grained class segmentation, novel concept segmentation, and segmentation in the presence of moderate corruption. While MaskCLIP excels without the need for training, its rigidity in adopting CLIP's image encoder limits its adaptability to advanced segmentation architectures like DeepLab and PSPNet. To address this, it is introduced MaskCLIP+, liberating it from the architectural constraints and facilitating integration with more sophisticated segmentation models. In MaskCLIP+, predictions are treated as training-time pseudo ground-truth labels, allowing for the incorporation of diverse segmentation architectures tailored to the task at hand. By utilizing MaskCLIP's predictions as pseudo ground-truth labels during target network training, we replace the classifier of the target network with that of MaskCLIP, preserving the network's ability for open-vocabulary prediction. MaskCLIP+ guided learning extends to the transductive zero-shot segmentation setting, where MaskCLIP generates pseudo labels exclusively for

unlabeled pixels. Distinguishing itself from related attempts targeting object detection, which employ knowledge distillation between CLIP's image-level visual features and a target model's features, our approach uses pseudo labels due to the structural dissimilarity between our target network and CLIP's image encoder. This choice proves advantageous, as it avoids conflicts in performance between seen and unseen classes, a challenge encountered with feature-level guidance. MaskCLIP+, relying on pseudo labels, maintains consistent performance across seen classes. [62]

3. Evaluation of methodologies

3.1 Generative learning methods

3.1.1 SWOT analysis

SWOT analysis is usually used in companies for identifying and analysing an organization's strengths, weaknesses, opportunities and threats. These words make up the SWOT acronym. A SWOT analysis draws information from internal sources (strengths and weaknesses) as well as external forces that may have uncontrollable impacts on decisions (opportunities and threats), so it considers the internal and external environment together.

Below a SWOT analysis of the different GL methods is presented in order to get an overview and evaluate their possible use. The analysis is based on the literature review performed in the precedent section of the Thesis.

Table 2: *SWOT Generative Adversarial Network*

STRENGHTS	WEAKNESSES
Hight-quality outputs Variety of applications Unsupervised learning Generative diversity	Training instability Mode collapse Evaluation metrics
OPPORTUNITIES	THREATS
Hybrid models Improved training algorithms	Bias amplification Resources intensive

Table 3: *SWOT Wasserstein Generative Adversarial Network*

STRENGHTS	WEAKNESSES
Hight-quality outputs Variety of applications Stable training Mode collapse mitigation Meaningful gradients	Hyperparameter sensitivity Computational effort Complexity
OPPORTUNITIES	THREATS
Improved GAN training Interdisciplinary applications	Resources intensive Hyperparameter complexity

Table 4: *SWOT Cycle Generative Adversarial Network*

STRENGHTS	WEAKNESSES
Hight-quality outputs Variety of applications Robust domain shifts in data Small dataset required	Slow training Prone to overfitting Difficult interpretation
OPPORTUNITIES	THREATS
Increased creativity in image generation	Unexpected/undesirable data

Table 5: *SWOT Autoregressive Generative Model*

STRENGHTS	WEAKNESSES
Sequential generation Interpretability Training stability Flexibility	Step-by-step generation Limitations in dynamic context Limited global view Large memory footprint
OPPORTUNITIES	THREATS
Hybrid models Conditional generation	Long-range dependencies Data efficiency demand large datasets

Table 6: *SWOT Generative Moment Matching Networks*

STRENGHTS	WEAKNESSES
Moment matching Simplicity No adversarial training	Limited complexity Optimization challenges Evaluation metrics
OPPORTUNITIES	THREATS
Hybrid models Transfer learning	Practical applicability Competing approaches

Table 7*: SWOT Variational Autoencoder*

STRENGHTS	WEAKNESSES
Probabilistic modelling Latent space Regularization Applicability	Mode collapse Blurry samples Limited expressiveness
OPPORTUNITIES	THREATS
Hybrid models Conditional generation	Model complexity Approximation quality

Table 8: *SWOT Normalizing Flows*

STRENGHTS	WEAKNESSES
Exact likelihood estimation Invertibility Interpretable latent space	Limited flexibility for complex data Training complexity
OPPORTUNITIES	THREATS
Hybrid models	Computational demand

3.1.2 Benchmarking analysis

Benchmarking algorithms is a critical process in evaluating their performance across various tasks. It involves defining the problem at hand, selecting appropriate metrics to measure success, and carefully choosing representative datasets. Comparisons is done with baselines, such as existing algorithms or heuristic approaches, offer additional context.

The benchmarking analysis showed in **Table 9** compares GL methodologies on a 1 to 5 scale. The characteristics used to evaluate the choice of the algorithm include:

- Output Quality: it assesses the visual appeal and realism of generated images.
- Easy Algorithm: it evaluates the ease of use and implementation of an algorithm, which is crucial for practical applications.
- Stability: it measures the consistency of image generation across different datasets or runs.
- Adaptability: it assesses how well an algorithm can be adjusted to generate images with specific characteristics or adapt to different datasets.
- Entirety: it evaluates how comprehensively an algorithm captures the entirety of the data distribution or generates images that cover a wide range of features.
- Applicability: it assesses the versatility of an algorithm in various image generation tasks and datasets.
- Feasibility: Feasibility considers practical constraints such as computational resources, time, and data requirements.
- Suitability for threat assessment: it evaluates how the use of the algorithm will benefit the threat assessment process.

The suitability of threat assessment can be considered the problem at hand, and so the primary factor influencing the choice of the algorithm.

***Table 9**: Benchmarking analysis of Generative Learning methods*

	Output quality	Easy algorithm	Stability	Adaptability	Entirety	Applicability	Feasibility	Suitability for threat assessment	TOT
GAN	5	4	3	4	4	5	4	4	**33**
WGAN	5	3	4	4	5	4	4	5	**34**
CGAN	5	4	4	5	4	5	4	5	**36**
AM	4	4	3	3	4	4	3	5	**30**
GMMN	3	4	4	3	3	3	3	2	**25**
VAE	4	4	4	5	4	5	4	3	**33**
NF	5	3	3	5	4	4	3	4	**31**

Based on the comprehensive benchmarking analysis performed on the results from the literature review, it becomes evident that GANs and VAEs stand out as the most efficient solutions. These sophisticated models excel in their respective domains, offering distinct advantages depending on the primary objectives of the task at hand. In scenarios where the paramount goal revolves around generating high-quality data, GANs reign supreme. Leveraging their adversarial training framework, GANs exhibit a remarkable capability to produce synthetic samples that closely resemble real data, boasting visually appealing characteristics that mirror those found in the original dataset. The ability of GANs to generate realistic samples with fidelity makes them an ideal choice for tasks where data quality is of utmost importance. Conversely, when stability takes precedence over other considerations, VAEs emerge as the preferred option. VAEs, characterized by their probabilistic framework and latent variable representations, offer a more consistent and robust performance in generating synthetic data. Their inherent ability to capture underlying data distributions while providing principled uncertainty estimates renders them particularly well-suited for tasks where maintaining stability and reliability in data generation is critical. Thus, the selection between GANs and VAEs hinges upon the specific requirements and priorities of the task at hand. For tasks where optimizing data quality is paramount, GANs stand as the recommended choice due to their prowess in producing visually appealing and realistic samples. On the other hand, if stability and robustness in data generation are the primary concerns, VAEs offer a more suitable solution, ensuring consistent performance across varying conditions. By carefully aligning the choice of model with the objectives of the task, practitioners can effectively

harness the capabilities of GANs and VAEs to achieve optimal results in synthetic data generation.

3.1.3 Diffusion methods

Stable diffusion is emerging as a leading-edge method of synthetic image generation, characterised by its ability to produce images of exceptional realism and precision. Unlike conventional techniques, stable diffusion works within a compressed latent space, where noise is incrementally added to create highly detailed image representations. This unique process ensures that the synthetic images produced are remarkably lifelike, capturing even the most intricate details and subtle nuances.

A key advantage of stable diffusion is its reliance on a likelihood-based framework. This approach effectively addresses and mitigates common problems found in other generative models, such as mode collapse, where the model produces a limited variety of images, and training instabilities, which can affect the quality and consistency of the generated images. By overcoming these challenges, stable diffusion provides a more robust and reliable method for image generation.

The rationale for choosing stable diffusion for synthetic image generation lies in its ability to preserve the core characteristics of the original images while introducing variations that closely mimic real-world scenarios. This is particularly important for applications that require a high level of detail and authenticity, such as medical imaging, virtual reality and digital art.

Stable diffusion also provides a high degree of flexibility in controlling the level of noise within the latent representations. Users can fine-tune this parameter to produce images that meet specific requirements, whether they need more realistic and detailed images or those with greater variability and abstraction. This control makes Stable Diffusion an invaluable tool for creating a wide range of synthetic images tailored to different use cases, from enhancing training datasets for ML to developing realistic simulations in games and films.

3.2 Transfer learning methods

Transfer learning can reduce the time and resources required for the training phase of ML models because it can leverage the knowledge of pre-trained models on large datasets. However, there are several challenges associated with pre-trained models in transfer learning. First, there is the issue of availability of pre-trained models. In addition, the integration of different models can be difficult. Furthermore, pre-trained models can lead to overfitting, as they tend to capture overly specific features as the dataset size increases.

On the other hand, the use of pre-trained models can lead to higher performance compared to training from scratch, especially when dealing with small output datasets. It's important to note that transfer learning requires an output dataset to train the model on the new task, which distinguishes it from GL. GL aims to produce images like the input, whereas transfer learning produces images in a desired style.

Transfer learning has the advantage of better generalisation, especially when the input data set is large. In addition, pre-trained models are more adaptable to different tasks with less difficulty in training.

Style transfer is a technique in image processing and computer vision in which the artistic style of one image is transferred to another while preserving its content. It allows the creation of visually appealing and artistic images by blending the style of one image with the content of another. On the other hand, while style transfer can produce visually appealing results, it can also lead to a loss of originality in the content. The final images may be so similar to the style of the reference image that the original content becomes less recognisable or distinctive. In addition, style transfer techniques are limited by the styles and reference images available.

Style transfer techniques can be customised to suit individual preferences or branding requirements. In addition, many style transfer algorithms and tools are readily available and easy to use, requiring minimal technical expertise. It is important to note, however, that these algorithms can be computationally intensive, especially when processing high-resolution images or complex artistic styles, and can introduce imperfections into the final output, such as blur, distortion or mismatched textures.

Style transfer is a powerful tool for artistic expression and visual enhancement and is most often used when the goal is to achieve an image with a desired style.

Domain transfer is a technique used in ML and AI to transfer knowledge from a source domain to a target domain, enabling models trained on one domain to generalize well to a different but related domain. However, one of the main challenges in domain transfer is the presence of a domain gap between the source and target domains. If the domains are too dissimilar, it may be difficult for the model to effectively transfer knowledge, leading to decreased performance in the target domain.

Domain transfer allows for the transfer of specific features or knowledge from the source domain that are relevant to the target domain. This enables models to focus on learning domain-specific characteristics while leveraging general knowledge from the source domain. Nevertheless, differences in labels or annotations associated with the data between the source and target domains can pose challenges for domain transfer algorithms. The model needs to adapt not only to differences in data distribution but also to changes in the labeling scheme.

The effectiveness of domain transfer techniques heavily depends on the quality and relevance of the source domain data. Noisy, biased, or insufficiently representative source domain data may hinder the performance of the transferred model in the target domain. However, leveraging pre-existing knowledge from a source domain can mitigate this issue, thereby reducing the time and resources required for training models in the target domain.

3.3 Segmentation and Labelling techniques

The evaluation of segmentation and labelling methods will build upon the state-of-the-art review previously conducted.

RCNN, primarily developed for object detection tasks, operates by proposing regions of interest within an image and subsequently classifying these regions into different object categories. While effective for object detection, RCNN does not inherently offer segmentation or labeling capabilities. However, the introduction of Mask RCNN extends the capabilities of RCNN by incorporating a segmentation component. This enhancement allows Mask RCNN to

not only identify objects within images but also generate pixel-level segmentation masks, precisely delineating object boundaries. Through this extension, Mask RCNN enhances the utility of RCNN, particularly in scenarios where detailed segmentation is crucial for accurate analysis and understanding of image content.

In contrast, HCP specializes in human-centric labeling tasks, focusing specifically on recognizing and labeling various parts of the human body in images. Unlike RCNN and Mask RCNN, which are more generalized in their applications, HCP is tailored to meet the requirements of human-related labeling tasks, offering a specialized solution for tasks such as pose estimation and anatomical analysis.

Meanwhile, CLIP represents a paradigm shift in image understanding, leveraging natural language descriptions to label images. By associating textual descriptions with visual content, CLIP enables tasks like image labeling based on semantic understanding. However, unlike Mask RCNN, CLIP does not provide segmentation capabilities inherently. To address this limitation, MaskCLIP emerges as an extension of CLIP, enhancing its capabilities by providing segmentation masks for objects within images. This integration of segmentation capabilities with CLIP's contextual understanding enables more detailed analysis and interpretation of visual content, facilitating tasks such as object localization and scene understanding. Moreover, the subsequent improvement offered by MaskCLIP+ further enhances the quality and accuracy of segmentation masks, ensuring more precise delineation of object boundaries compared to its predecessor.

3.3.1 Benchmarking analysis

A comprehensive benchmarking analysis based on the literature review assesses and rates the effectiveness of segmentation and labelling methods, assigning scores between 1 and 5 based on several key evaluation criteria. These criteria encompass the following aspects:

- Accuracy: it focuses on the model's precision and its ability to provide accurate results.
- Robustness: it refers to the stability and performance consistency across various conditions and scenarios.

- Ease of Use: it encompasses the user experience aspects associated with the model, including installation, integration, and workflow usability.
- Flexibility: it assesses the model's adaptability and versatility across a range of tasks and applications.
- Feasibility: it evaluates the practical use of the model in terms of computational resources, including memory, processing power, and energy consumption.

Table 10: *Benchmarking analysis of Segmentation and Labeling methods*

	Accuracy	Robustness	Ease of use	Flexibility	Feasibility	**TOT**
RCNN	4	3	3	3	3	**16**
Mask RCNN	5	4	4	4	4	**21**
HCP	4	4	4	2	4	**18**
CLIP	3	3	3	4	3	**16**
Mask CLIP	4	4	4	5	4	**21**
Mask CLIP+	5	5	4	5	4	**23**

Based on the benchmarking analysis showed in **Table 10** the standout method emerges as Mask CLIP+. It represents a cutting-edge solution that seamlessly integrates state-of-the-art segmentation capabilities with advanced contextual understanding. This fusion of capabilities empowers Mask CLIP+ to provide a comprehensive and nuanced approach to threat assessment tasks, effectively addressing the multifaceted requirements inherent in such scenarios. By leveraging advanced segmentation techniques, Mask CLIP+ excels in precisely delineating and isolating objects or regions of interest within complex visual data, facilitating accurate and granular labeling—a critical aspect of threat assessment tasks. Furthermore, its contextual understanding capabilities, epitomized by the CLIP framework, enable Mask CLIP+ to comprehend and interpret the broader context surrounding identified threats, thereby enhancing the accuracy and relevance of its assessments. Moreover, Mask CLIP+ exhibits commendable robustness, demonstrating consistent performance across diverse environmental conditions and scenarios. Its ease of use and flexibility further contribute to its appeal, facilitating seamless integration into existing workflows and adaptability to evolving threat landscapes.

4. Proposed approaches

In threat assessment within military contexts, two distinct yet interrelated approaches are currently under evaluation. These approaches aim to tackle the core challenges encountered with military datasets: poor data quality and restricted availability.

The first approach concentrates on improving data quality by enhancing resolution. Through the application of advanced generative AI techniques, this method endeavours to elevate the clarity and fidelity of existing imagery, thereby rendering it more suitable for in-depth analysis. This process entails a multi-step procedure: initially, employing generative AI to augment low-resolution images, followed by subjecting both the original and enhanced images to segmentation and labelling algorithms for object identification and classification. A critical phase involves comparing the accuracy of segmentation and labelling between the original and enhanced images, facilitating the refinement of the AI model. Finally, the upgraded images, enriched with finer details, are utilized for enhanced threat detection and assessment.

On the other hand, the second approach revolves around the generation of realistic synthetic images to supplement existing datasets. Employing stable diffusion techniques across two distinct pipelines – the Text2Img and Img2Img pathways – this method creates synthetic scenarios based on textual descriptions or video game images, respectively. These synthetic scenarios are then endowed with threat levels and organized into training, validation, and test sets, primed for the education of AI models tasked with threat assessment.

The results obtained from these approaches are not included in this Thesis, as they were developed during the internship conducted within an European Defence Project, which is affected by confidentiality issues. For this reason, all data produced during the project are confidential. So, to demonstrate the proposed approaches in the Thesis, it was decided to generate and use images through an open-source site called Leonardo.ai.

While these approaches are developed and evaluated autonomously, their fusion holds significant promise. By harmonizing the enhanced resolution of existing data with the generation of realistic synthetic imagery, it becomes possible to address both the quality and

availability challenges inherent in military datasets comprehensively. This synergistic integration not only can fortify the capabilities of AI-driven threat assessment but also amplify decision support mechanisms and augments soldier training programs.

4.1 Super resolution of low-quality images

The “super resolution of low-quality images” approach to perform threat assessment via GL is based on enhancing the quality of existing low-resolution images rather than generating new synthetic ones. **Figure 4** illustrates the methodology selected. Firstly, a generative AI model is employed to increase the quality of low-resolution images. This enhancement process involves using advanced algorithms to add details and improve clarity, resulting in high-resolution versions of the original images. Following this, both the low-resolution and the newly generated high-resolution images undergo segmentation and labelling. This step involves identifying and classifying the objects present in the images. The purpose of conducting segmentation and labelling on both sets of images is to allow for a comparison of the accuracy between the two sets. The expectation is that the high-resolution images, being clearer and more detailed, will enable more precise segmentation and labelling. By analysing the differences in accuracy, the performance of the generative AI can be fine-tuned and improved, ensuring it produces the most effective image enhancements possible. After refining the generative AI model based on these comparisons, the next step focuses on using the segmentation and labelling results from the high-resolution images to perform threat detection and assessment. This involves analysing the classified objects to determine potential threats. The enhanced clarity and detail in the high-resolution images should facilitate a more accurate and reliable threat assessment process. The proposed methodology leverages generative AI to enhance image quality, improving the accuracy of subsequent image analysis tasks. This enhancement not only aids in better segmentation and labelling but also significantly contributes to more precise threat detection and assessment.

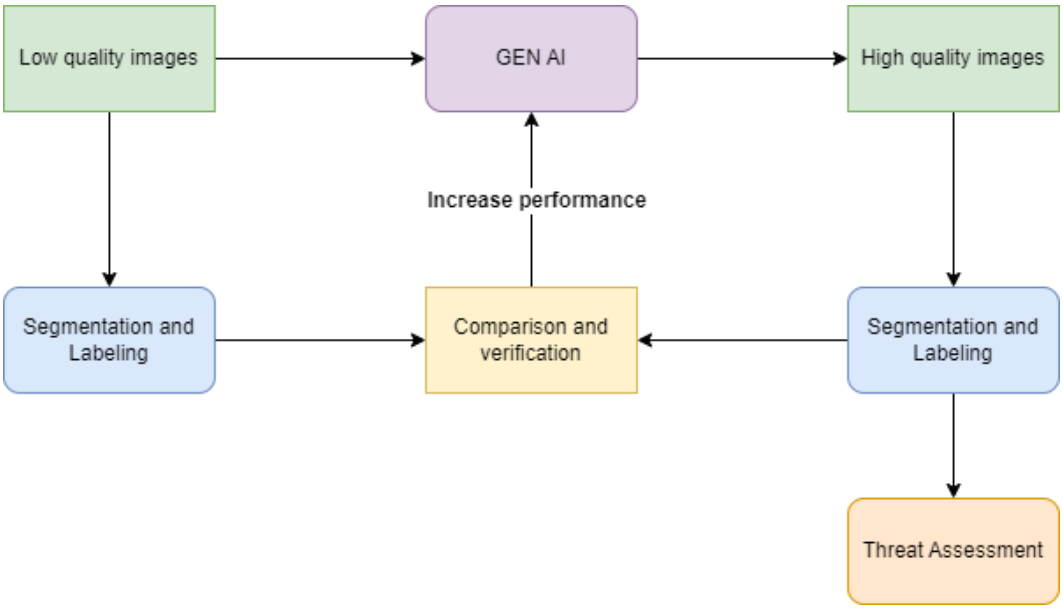

Figure 4: *Super resolution of low-quality images scheme*

4.1.1 GEN AI to increase images quality

Low resolution in image datasets presents a significant challenge across various domains, but it becomes particularly pronounced in military applications. The complexities of military operations often entail data collection in challenging environments, ranging from adverse weather conditions to high-risk combat zones. In such scenarios, obtaining high-quality images can be arduous due to factors like limited visibility, rapid movement, and the need for covert surveillance. Additionally, the equipment used to capture these images may be subject to constraints in terms of size, weight, and power consumption, further compromising image quality. Military-grade sensors, while rugged and adaptable, often compromise resolution for durability and versatility, especially when operating in low-light or long-range scenarios. Another significant factor exacerbating the issue of low resolution is the data security. Military imagery often contains sensitive information that must be protected from unauthorized access or interception by adversaries. Data security measures can introduce additional compression and encryption steps in the images that degrade image quality, further diminishing the resolution of the transmitted data.

A promising method to increase the resolution of images is to leverage generative AI models. Real-ERSGAN is one of the most effective algorithms to be used on order to increase the

resolution of images. The results of the increased image resolution can be seen in **Figure 5**, where the first image is characterized by low resolution and the second image is the one characterized by an increased resolution.

Figure 5: *Image from low to high resolution*

The comparison between the performance of classification and segmentation tasks using low-resolution and high-resolution images is used as evaluation metric to increase the ability of generating high quality images of the generative AI. For classification the measures considered include F1 score, precision, recall, and accuracy, while segmentation metrics comprise mean IoU, F1 score, precision, recall, and accuracy. The evaluation will be

conducted on datasets of military objects and benchmarked against state-of-the-art methods for super-resolution and object detection.

4.1.2 Segmentation and labelling

Segmentation and labelling are essential steps in efficiently comparing low- and high-resolution images to improve the performance of generative AI. Accurate segmentation and labelling allow for accurate evaluation of synthetic high-resolution images, ensuring that the generative algorithm effectively increases resolution. If these high-resolution images show easier segmentation and identification of elements, it indicates the success of the generative algorithm; otherwise, its performance needs to be improved.

In addition, segmentation and labelling are fundamental to subsequent threat detection and assessment. Segmentation isolates individual objects or elements within images, allowing more precise analysis of each segment. This enables a focused assessment of potential threats within the synthetic environment. In addition, segmented objects provide valuable input for further computer vision tasks such as object recognition, classification and tracking. These isolated segments provide a basis for training ML models to identify and categorise objects within synthetic scenes. This contributes to a more complete understanding of the content and context of the generated images, supporting robust threat assessment and scenario analysis. In comprehensive image analysis, the labelling phase contributes significantly to the understanding of the content. This process takes place after segmentation and focuses on the segmented images and the various objects within them. This sequential methodology allows for targeted exploration of individual objects. Labelling is critical for accurate threat assessment and object identification. It enhances the interpretation of image content by associating meaningful categories, assists in the accurate identification and classification of individual objects, provides context to segmented elements for improved scenario analysis, provides valuable training data for ML models, and produces results that are easy to understand for collaboration with experts. Labelling also facilitates informed decision-making across multiple domains, extracts actionable information tailored to specific use cases, and

contributes to visually informative representations that enhance the overall analytical capabilities of generated synthetic images.

4.1.3 Threat assessment

The objective of the threat assessment is to comprehensively identify and evaluate potential threats within the synthetic high-resolution images. This process is intricately linked to the segmentation and labelling, which collectively contribute crucial information for a nuanced understanding of the elements present in the synthetic environment. The segmentation module isolates and delineates the elements, providing a foundational understanding of the synthetic environment's composition. This segmented information, enriched by the subsequent labelling module, equips the threat assessment module with detailed insights into the nature and characteristics of each identified object.

4.2 Realist scenarios generation

The "realist scenarios generation" approach utilizes stable diffusion to create synthetic scenarios through two distinct pipelines: Text2Img and Img2Img. As illustrated in ***Figure 6***, the input to the stable diffusion model can be either text or images, depending on the desired pipeline. The first pipeline generates desired scenarios based on textual descriptions; by inputting specific text, the model produces corresponding synthetic images that depict the described scenes. In the other one, images from video games are used as conditioning inputs to generate similar synthetic scenarios; this approach leverages the visual characteristics of the provided images to produce new, yet comparable, synthetic scenes. Once the synthetic scenarios are generated, threat levels are assigned to each scenario. These scenarios are then divided into training, validation, and test sets to facilitate the learning process of the classifier. This subdivision ensures that the classifier can effectively learn from the training set, validate its performance, and be tested for accuracy in threat assessment.

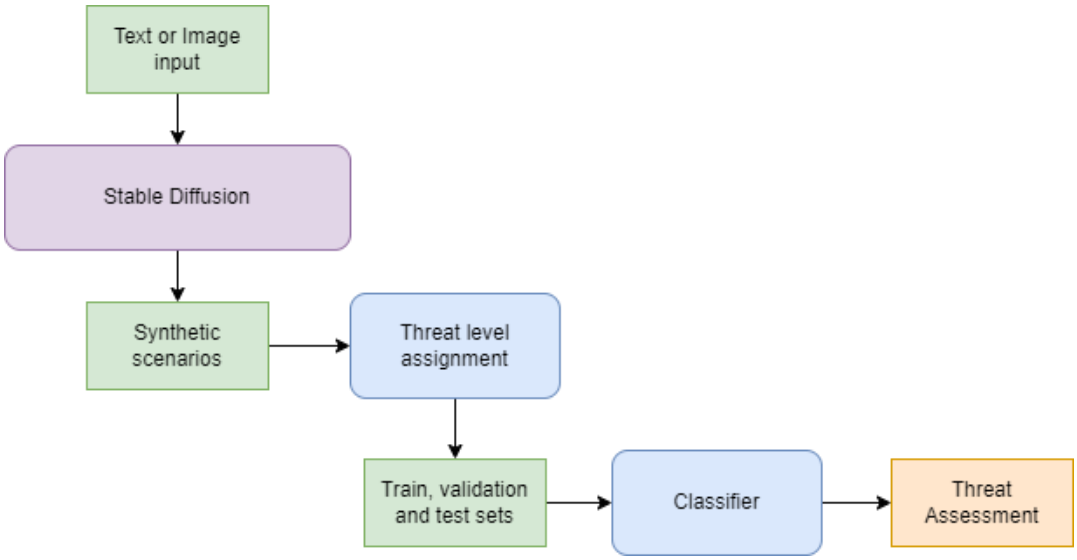

***Figure 6**: Realist scenarios generation scheme*

4.2.1 Stable diffusion to generate scenarios

Stable diffusion is an advanced technique used to generate synthetic realistic scenarios. It operates within a compressed latent space, incrementally adding noise to create detailed and nuanced images. In the proposed approach this powerful method uses two pipelines: text2img and img2img, each with different capabilities and applications.

The text2img pipeline focuses on generating images based on a textual prompt used as conditioning input. This process allows the user to describe the desired scene or scenario in words, and the stable diffusion model translates this description into a highly detailed and realistic image. In addition, a negative prompt can be incorporated to explicitly exclude certain elements or effects from the generated scenario, ensuring that the final image closely matches specific requirements.

Figure 7 shows the results inputting as text “Desert field with several tanks”. Instead, ***Figure 8*** show the results obtained by inputting as text “Tank in desert field under attack”.

Figure 7: *Desert scenario with several tanks*

Figure 8: *Tank under attack in desert scenario*

In contrast, the img2img pipeline requires both a prompt and an existing image as conditioning inputs. This method can be further divided into two different setups: basic img2img and instruct pix2pix. The basic img2img setup uses an image as the primary conditioning element along with the prompt. This approach allows significant changes to be made to the original image, transforming it extensively based on the given prompt. It is particularly useful for creating entirely new scenarios by exploiting the visual information contained in the conditioning image. This technique is effective for generating scenarios that need to retain some visual coherence with the input image while undergoing substantial changes. The instruct pix2pix setup refines the img2img pipeline by using the prompt as a set of instructions for modifying the input image. This method preserves the core structure and detail of the original image, resulting in more controlled and less drastic changes. It is ideal for applications where maintaining the integrity of the original image is critical, such as when subtle changes or enhancements are required without significantly altering the overall composition.

Indeed, the ***Figure 9*** shows the results obtained.

Figure 9: *Urban scenario during battle*

The text2img pipeline excels in scenarios with fewer objects, where the text description can succinctly capture the essential details. In contrast, the img2img pipeline can generate entirely new and complex scenarios, especially when using images from video games or other rich visual sources. By exploiting the information embedded in the conditioning images, this approach can produce highly varied yet coherent scenes, facilitating the creation of diverse and complex threat scenarios. The ability to generate realistic and varied synthetic scenarios is helpful in threat assessment. The text2img pipeline enables the rapid generation of specific scenes based on detailed textual descriptions, useful for rapid prototyping of potential threats. Meanwhile, the img2img pipeline, with its ability to drastically transform existing images, enables the creation of novel scenarios that can simulate a wide range of potential threats. This diversity is critical to comprehensive threat analysis and preparedness, ensuring that the scenarios generated cover a wide range of possibilities.

4.2.2 Domain transfer to increase realism

Domain transfer, coupled with stable diffusion, is a powerful synergy that aims to increase the realism of the scenarios generated, while at the same time adding a greater degree of variation and complexity. Domain transfer involves the transfer of knowledge learned in one domain to another, often from a source domain where data is abundant to a target domain where data may be scarce or difficult to obtain. In the context of scenario generation, domain transfer allows the integration of knowledge from different sources to enrich the synthetic scenarios. To further enhance the authenticity and diversity of these scenarios generated by the two pipelines based on stable diffusion, domain transfer becomes instrumental. Domain transfer uses information from different domains, such as images, text or even real-world data, to enrich the scenario generation process. In addition, domain transfer enables the adaptation of stable diffusion models trained on one dataset or domain to perform effectively in a different domain or scenario. This adaptability allows the use of pre-trained models, reducing the need for extensive training data and speeding up the scenario generation process. Furthermore, domain transfer is used to introduce novel elements or changes to the generated scenarios, thereby increasing their variability and richness. By transferring knowledge from domains with

different characteristics or introducing transformations learned from real-world data, the stable diffusion model can generate scenarios with unexpected or innovative elements, enhancing their realism and relevance for threat assessment.

4.2.3 Threat level assignment

The methodology used to assign different threat levels to the synthetic scenarios generated relies heavily on the insights of experienced soldiers. Drawing on the expertise and first-hand experience of military personnel, this approach provides a nuanced understanding of threats, enabling not only the assignment of generic threat levels, but also the differentiation of threat levels in different contexts. By using the knowledge and judgement honed by practical experience, analysts can effectively identify subtle nuances in threat scenarios and thereby delineate the level of risk associated with each situation. In essence, this methodology goes beyond a simplistic categorisation of threats into broad levels and instead delves into the intricacies of each scenario. Furthermore, by incorporating the perspectives of experienced soldiers, the threat assessment process becomes more contextual and attuned to real-world scenarios. Rather than relying solely on theoretical models or abstract classifications, analysts can draw on the practical wisdom and tacit knowledge accumulated by those with frontline experience. Ultimately, incorporating the insights of experienced soldiers into the threat assessment methodology enriches the analysis and decision-making process, enabling a more nuanced understanding of threats and their implications. By drawing on the expertise of those who have faced similar scenarios in the real world, analysts can better anticipate, assess and mitigate potential risks, thereby increasing the effectiveness and readiness of military operations.

4.2.4 Classifier to risk assessment

A ML classifier is used to identify and classify elements within synthetic realistic scenarios based on data collected from experienced soldiers. This process begins with the collection of detailed and accurate data from soldiers who have direct experience in the field, as explained in the previous section. The collected data serves as a robust basis for training the ML classifier.

Once trained, the classifier can analyse synthetic images and scenarios and effectively identify elements present in these environments. The identification process involves recognising different objects, people and other significant components within the synthetic scene. This recognition is critical to understanding the context and specific details of the scenario being evaluated. Once the elements have been identified, the classifier proceeds to classify the potential threats within the synthetic environment. This classification is not limited to a binary presence or absence of risk. Instead, it is a comprehensive risk assessment that takes into account the intricate details of the environment in which the threat is located. Factors such as surrounding infrastructure, spatial relationships between objects, environmental conditions and other contextual information play an important role in this assessment. The classifier's ability to make nuanced risk assessments enhances its effectiveness in real-world applications. For example, in a synthetic battlefield scenario, the classifier can distinguish between different types of threat and also assess the threat level based on the environment. In addition, the integration of ML classifiers with soldier-collected data ensures that the system remains adaptive and continuously improves over time. As soldiers encounter new threats and environmental conditions, they provide updated data that can be used to retrain and refine the classifier. The comprehensive risk assessment provided by the ML classifier supports informed decision making. Commanders and decision-makers can use the detailed analysis provided by the classifier to strategies and respond effectively. In addition, this technology helps in pre-emptive threat detection and mitigation. By continuously analysing synthetic scenarios, the classifier can identify potential threats before they manifest themselves in real-world situations.

5. Evaluation of the proposed approaches

The two proposed approaches address the two main critical problems of military datasets: the scarcity and low quality of images. Indeed, the super-resolution of low-quality images approach addresses the everyday problem of low-quality images, due to the fact that in the military field, the systems used to capture images are often of low resolution and the task of capturing images is generally considered to be of low importance. On the other hand, the realistic scenario generation approach addresses the problem of low availability of images, mainly due to privacy and confidentiality concerns. Given the different nature and specific problems addressed by these two approaches, their combined application offers a significant advantage, leading to the creation of growing datasets and thus increasing the possibilities for use in decision-making processes, particularly those promoted by threat assessment operations. The super-resolution approach to low quality imagery offers several advantages. First, it maximises the use of existing data by ensuring that it is both authentic and directly from the battlefield. This approach allows thorough threat assessments to be made of past real-world situations, thereby preparing for future scenarios. However, this method does not have the potential to generate new scenarios and can involve significant computational effort without providing proportional results. Conversely, the realistic scenario generation approach has the potential to generate new images from sources such as video games or textual descriptions, without the need for real images. This allows users to specify the types of elements to be generated, giving them control over the generated content and ensuring that it meets their requirements. Another advantage is the involvement of experts in risk classification, which allows the algorithm to learn from real-life experience. However, this method requires more evaluation effort to ensure the accuracy and reliability of the generated scenarios. It is clear that these proposed approaches need further refinement, particularly in the area of threat assessment. Nevertheless, their combined use could represent a significant breakthrough in the military field by effectively addressing the primary problems of low quality and scarce data. It also opens up the possibility of applying similar methods to other areas, such as the medical sector, where data quality and availability are also critical issues.

6. Conclusions

Artificial Intelligence has become a critical component of modern technology, evolving rapidly and finding applications in various sectors, including the military. AI involves the development of computer systems capable of performing tasks that require human intelligence, such as natural language understanding, pattern recognition, experience-based learning and decision-making. These systems can analyse vast amounts of data, identify trends and make predictions, leading to improved decision-making, increased productivity and enhanced user experiences.

AI plays an important role in various military applications, including autonomous warfare platforms, cybersecurity, logistics, target acquisition, battlefield intelligence processing and predictive maintenance. NATO member states and other military organisations are investing in AI to enhance their defence capabilities. AI can make a significant contribution to command, control, communications, computers, intelligence, surveillance and reconnaissance operations, cyber operations and the use of autonomous and semi-autonomous machines. In particular, in this Thesis, AI-based threat assessment is analysed as it can improve soldiers' decision-making by providing insight into risks both in training scenarios and in real time.

Despite its potential, the integration of AI into military operations poses significant challenges, primarily related to data frugality, robustness, and the explainability of AI. Military personnel, require user-friendly interfaces to effectively use AI systems due the scares time available to make decisions. Furthermore, ethical considerations need to be addressed to prevent bias and ensure fair and responsible use of AI. Finally, despite technological advances that have led to a proliferation of data sets, particularly in the military sector, challenges remain due to the low quality and scarcity of data.

Generating synthetic data and improving the resolution and quality of existing data are practical solutions to the challenges. In addition, these approaches improve frugal and robust AI, enabling AI-based systems to work efficiently with scarce real-world data. Synthetic data enables the simulation of a wide range of scenarios and threats, providing a rich dataset for training AI models without the need for extensive real-world data collection and avoiding

privacy issues. Enhancing existing data with advanced algorithms also helps create high-quality datasets for effective AI training.

Various techniques are employed for generating synthetic images and enhancing their resolution. In this Thesis have been studied and evaluated:

- Transfer learning: it involves applying knowledge acquired source domain to another target domain. In the context of style transfer and domain transfer, this entails utilizing pre-trained models to adapt styles or features from one dataset to another, thereby not only generating new images but also imbuing them with the desired style.
- Generative learning: it focuses on constructing models capable of generating new data, often through methods like Generative Adversarial Networks or Variational Autoencoders. These models learn to produce data resembling a specified distribution, facilitating the creation of synthetic data. Generative learning methods can also be employed to enhance image quality.
- Stable diffusion: it describes a process in which a system gradually attains equilibrium without significant fluctuations. In ML, it is frequently utilized in generative models to regulate the generation process and ensure that the produced samples are coherent and stable.
- Modelling and simulation: they entail the development of mathematical or computational models that simulate real-world phenomena. In synthetic data generation, these techniques are utilized to simulate data closely resembling real-world data, furnishing valuable training data for ML models without necessitating large, labelled datasets.

AI-based threat assessment can be performed though segmentation and labelling, which are employed to identify and classify threats. Segmentation involves dividing an image into distinct regions or segments to isolate and highlight objects or areas of interest. This process is critical for identifying potential threats within an image by focusing on specific segments rather than analysing the entire image at once. Labelling assigns pre-defined categories or classes to these segmented regions. In the context of threat identification and classification, this means

that segmented areas are categorised as specific types of risk. By combining segmentation and labelling, security systems can more accurately and efficiently identify and classify threats, improving overall security and response times.

In this Thesis two main approaches have been proposed and explained to firstly address the issues of image scarcity and low resolution and then to perform threat assessment in order to enhance decision making process.

The first approach increases the resolution of low-quality imagery through generative AI, through Real-ERSGAN. This algorithm improves the clarity and detail of images, making them more useful for training AI models. Segmentation and labelling are used to improve the performance of the generative AI system comparing the results obtained from low- and high-resolution images. Then segmentation and labelling on the increased resolution images are used to detects threats.

The second approach focuses on the generation of realistic synthetic scenarios using stable diffusion techniques. Two different pipelines have been used: text to image pipeline using as input a text describing the desired scenario; image to image using as input videogames images. Different threat levels are then assigned to the generated scenarios based on soldiers' experience. Training classifiers on these generated scenarios with related threat levels enables the AI to make accurate risk assessments based on real life knowledge.

The next step to achieve is to combine the two approaches in order to integrate high-resolution image enhancement with realistic synthetic scenario generation. This integration will enable a comprehensive AI-based threat assessment framework that leverages both enhanced visual data clarity and diverse, simulated scenarios.

This Thesis emphasizes the essential role of data centrality in advancing AI applications, with a focus on the military field. By prioritizing the enhancement of data quality through techniques such as high-resolution image generation and realistic scenario simulation, the research highlights how robust and diverse datasets are essential for improving decision-making capabilities. This approach not only tackles current challenges in integrating AI within military operations but also lays a foundation for future innovations in defence technology that rely on

comprehensive and reliable data-driven insights. By harnessing state-of-the-art generative AI techniques to enhance image resolution and simulate realistic scenarios, alongside precise segmentation and labelling methodologies, significant advancements are demonstrated in AI-driven threat assessment within military contexts. These advancements directly contribute to enhancing decision-making processes critical for optimizing defence operations and ensuring superior situational awareness and response capabilities.

Bibliography

[1]. Hodson, D.D., Hill, R.R., Bruzzone A. (2014) Special issue: Art and science of live, virtual and constructive simulation and intelligent agents to support defense and homeland security testing and analysis, Journal of Defense Modeling and Simulation, DOI 10.1177/1548512914528095

[2]. Laskowski, N., & Tucci, L. (2024) What is artificial intelligence (AI)? Everything you need to know. CIO/IT Strategy. Retrieved from https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

[3]. Bruzzone, A.G., Longo, F., Massei, M., Nicoletti, L. (2014) Human modeling for multi coalition joint operations, Simulation Series, DOI

[4]. Szabadföldi, I. (2021). Artificial intelligence in military application–opportunities and challenges. Land Forces Academy Review, 26(2), 157-165.

[5]. Rodrigue, E., & Needle, F. (2023). 4 Types of Artificial Intelligence & What Marketers Are Using Most (Research). Retrieved from https://blog.hubspot.com/marketing/types-of-ai

[6]. Bruzzone A.G., Remondino, M., Battista, U., Tardito, G., Santoni, F.T. (2021) Modelling & Data Fusion to support Acquisition in Defence, 11th International Defense and Homeland Security Simulation Workshop, DHSS 2021, DOI 10.46354/i3m.2021.dhss.012

[7]. Bruzzone, A.G., Caussanel J, Giambiasi N., Frydman C., (2007) "From Abstract Representation to Formal Modelling of Tactical Military Operations", Proceedings of Summer Computer Simulation Conference 2007, San Diego, July

[8]. Bruzzone, A.G., Bocca, E., Tarone, F. (2011) Simulating urban environment for assessing impact of alternative command & control netcentric maturity models within asymmetric scenarios, International Defense and Homeland Security Simulation Workshop, DHSS 2011, Held at the International Mediterranean and Latin American Modeling Multiconference, I3M 2011, DOI

[9]. Bruzzone, A. G., & Massei, M. (2017). Simulation-based military training. Guide to Simulation-Based Disciplines: Advancing Our Computational Future, 315-361.

[10]. Mazal, J., Fagiolini, A., Vasik, P., Storto, S.L., Bruzzone, A.G., Pickl, S., Neumann, V., Stodola, P. (2023) Preface, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), DOI

[11]. Carrera, A., Tremori, A., Caamaño, P., Been, R., Pereira, D.C., Bruzzone, A.G. (2016) HLA interoperability for ROS-based autonomous systems, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) DOI 10.1007/978-3-319-47605-6_10

[12]. Bruzzone, A.G., Fontane, J.-G., Berni, A., Brizzolara, S., Longo, F., Dato, L., Poggi, S., Dallorto, M. (2013) Simulating the marine domain as an extended framework for joint collaboration and competition among Autonomous Systems, 3rd International Defense and Homeland Security Simulation Workshop, DHSS 2013, Held at the International Multidisciplinary Modeling and Simulation Multiconference, I3M 2013, DOI

[13]. Bruzzone, A.G., Tremori, A., Longo, F., Turi, M., Franzinetti, G. (2012) Models & interactive simulation for civil military interoperability in humanitarian aid and civil protection, 11th International Conference on Modeling and Applied Simulation, MAS 2012, Held at the International Multidisciplinary Modeling and Simulation Multiconference, I3M 2012, DOI

[14]. Bruzzone, A.G., Massei, M., Tremori, A., Camponeschi, M., Nicoletti, L., Di Matteo, R., Franzinetti, G. (2015) Distributed virtual simulation supporting defense against terrorism, 5th International Defense and Homeland Security Simulation Workshop, DHSS 2015, DOI

[15]. Bruzzone A.G., Orsoni, A. (2003) AI and simulation-based techniques for the assessment of supply chain logistic performance, Proceedings - Simulation Symposium, DOI 10.1109/SIMSYM.2003.1192809

[16]. Bruzzone, A.G., Massei, M., Bartolucci, C., Poggi, S., Martella, A., Franzinetti, G. (2014) Interoperability requirements for developing simulation solutions for innovative integrated systems, 4th International Defense and Homeland Security Simulation Workshop, DHSS 2014, DOI

[17]. Bruzzone A.G., Massei, M., Sinelshchikov, K., Giovannetti, A., Gadupuri, B.K. (2021) Strategic Engineering Applied to Complex Systems within Marine Environment, Proceedings of the 2021 Annual Modeling and Simulation Conference, ANNSIM 2021, Simulation Series DOI 10.23919/ANNSIM52504.2021.9552035

[18]. Cayirci, E., Bruzzone, A.G., Longo, F., Gunneriusson, H. (2016) A model to describe hybrid conflict environments, 6th International Defense and Homeland Security Simulation Workshop, DHSS 2016, DOI

[19]. Bruzzone, A.G., Massei, M., Maglione, G.L., Sinelshchikov, K., Di Matteo, R. (2017) A strategic serious game addressing system of systems engineering, 16th International Conference on Modeling and Applied Simulation, MAS 2017, Held at the International Multidisciplinary Modeling and Simulation Multiconference, I3M 2017, DOI

[20]. Bruzzone, A.G., Gadupuri, B., Schmidt, W., Nikolov, O., Massei, M., Di Bella, P., Pedemonte, M. (2021) AI & Interoperable Simulation for Pandemics and Crisis Management, 11th International Defense and Homeland Security Simulation Workshop, DHSS 2021, DOI 10.46354/i3m.2021.dhss.010

[21]. Constantinescu, M., & Dumitrache, V. I. (2022). Artificial intelligence and the future of defence planning and resources management. In International conference KNOWLEDGE-BASED ORGANIZATION (Vol. 28, No. 1, pp. 180-186).

[22]. Skalický, P., Palasiewicz, T., Kyjovský, J., & Zelený, J. (2017). Possibilities of Modelling and Simulation in Military Engineering. In International Workshop on Modelling and Simulation for Autonomous Systems (pp. 402-409). Cham: Springer International Publishing.

[23]. Svenmarck, P., Luotsinen, L., Nilsson, M., & Schubert, J. (2018, May). Possibilities and challenges for artificial intelligence in military applications. In Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting (pp. 1-16).

[24]. Forrest, S. (1996). Genetic algorithms. ACM computing surveys (CSUR), 28(1), 77-80.

[25]. Gupta, N., VAYA, N., & MISHRA, B. (1997). Artificial Neural Network. Eastern pharmacist, 40(475), 39-41.

[26]. Chen, H., & Zhang, K. (2012). Target threat assessment based on genetic neural network. In 2012 International Conference on Industrial Control and Electronics Engineering (pp. 1789-1792). IEEE.

[27]. Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. Trends in cognitive sciences, 23(4), 305-317.

[28]. Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019.

[29]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[30]. Seel, N. M. (Ed.). (2011). Encyclopedia of the Sciences of Learning. Springer Science & Business Media.

[31]. Natsu. (2018). What are generative learning algorithms? Deep Learning. Retrieved from https://mohitjain.me/2018/03/12/generative-learning-algorithms/

[32]. Hofmann, P., Rückel, T., & Urbach, N. (2021). Innovating with artificial intelligence: capturing the constructive functional capabilities of deep generative learning.

[33]. Chiva, R., Grandío, A., & Alegre, J. (2010). Adaptive and generative learning: Implications from complexity theories. International Journal of Management Reviews, 12(2), 114-129.

[34]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[35]. Larsson, S. (2022, April 22). Frugal AI: Value at Scale Without Breaking the Bank. Scaling AI. Retrieved from https://blog.dataiku.com/frugal-ai-value-at-scale-without-breaking-the-bank

[36]. Ioualalen, A. (2021, July). How can frugal AI overcome the lack of data? Retrieved from https://www.linkedin.com/pulse/how-can-frugal-ai-overcome-lack-data-arnault-ioualalen

[37]. Figueira, A., & Vaz, B. (2022). Survey on synthetic data generation, evaluation methods and GANs. Mathematics, 10(15), 2733.

[38]. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1), 43-76.

[39]. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[40]. Mitchell, M. (1995, September). Genetic algorithms: An overview. In Complex. (Vol. 1, No. 1, pp. 31-39).

[41]. Lawrence, A. R., Kaiser, M., Sampaio, R., & Sipos, M. (2021). Data generating process to evaluate causal discovery techniques for time series data. arXiv preprint arXiv:2104.08043.

[42]. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE signal processing magazine, 35(1), 53-65.

[43]. Fernando, J. (2022, April 06). What Are Autoregressive Models? How They Work and Example. Retrieved from https://www.investopedia.com/terms/a/autoregressive.asp

[44]. Krasser, M. (2018). Deep feature consistent variational auto-encoder. Retrieved June, 12, 2019.

[45]. Omray, A. (2021, July 16). Introduction to Normalizing Flows: Why and how to implement normalizing flows over GANs and VAEs. Retrieved from https://towardsdatascience.com/introduction-to-normalizing-flows-d002af262a4b

[46]. Kobyzev, I., Prince, S. J., & Brubaker, M. A. (2020). Normalizing flows: An introduction and review of current methods. IEEE transactions on pattern analysis and machine intelligence, 43(11), 3964-3979.

[47]. Minchev, Z., & Shalamanov, V. (2010, April). Scenario Generation and Assessment Framework Solution in Support of the Comprehensive Approach. In Proceedings of SAS-

081 Symposium on “Analytical Support to Defence Transformation”, RTO-MP-SAS-081, Sofia, Boyana, April (pp. 26-28).

[48]. Zook, A., Lee-Urban, S., Riedl, M. O., Holden, H. K., Sottilare, R. A., & Brawner, K. W. (2012). Automated scenario generation: toward tailored and optimized military training in virtual environments. In Proceedings of the international conference on the foundations of digital games (pp. 164-171).

[49]. Hussain, S., Kamal, A., Ahmad, S., Rasool, G., & Iqbal, S. (2014). Threat modelling methodologies: a survey. Sci. Int.(Lahore), 26(4), 1607-1609.

[50]. Bang, M., & Liwång, H. (2016). Influences on threat assessment in a military context. Defense & Security Analysis, 32(3), 264-277.

[51]. Weng, L. (2019). From gan to wgan. arXiv preprint arXiv:1904.08994.

[52]. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

[53]. Zhang, Y. J. (2001, August). A review of recent evaluation methods for image segmentation. In Proceedings of the sixth international symposium on signal processing and its applications (Cat. No. 01EX467) (Vol. 1, pp. 148-151). IEEE.

[54]. Wu, M., Yue, H., Wang, J., Huang, Y., Liu, M., Jiang, Y., ... & Zeng, C. (2020). Object detection based on RGC mask R-CNN. IET Image Processing, 14(8), 1502-1508.

[55]. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[56]. Hua, X. S., & Li, J. (2015, February). Prajna: Towards recognizing whatever you want from images without image labeling. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 29, No. 1).

[57]. Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. Nature neuroscience, 3(11), 1199-1204.

[58]. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2285-2294).

[59]. Wei, Y., Xia, W., Lin, M., Huang, J., Ni, B., Dong, J., ... & Yan, S. (2015). HCP: A flexible CNN framework for multi-label image classification. IEEE transactions on pattern analysis and machine intelligence, 38(9), 1901-1907.

[60]. Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3367-3375).

[61]. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PMLR.

[62]. Zhou, C., Loy, C. C., & Dai, B. (2022, October). Extract free dense labels from clip. In European Conference on Computer Vision (pp. 696-712). Cham: Springer Nature Switzerland.

[63]. Hou, X., Sun, K., Shen, L., & Qiu, G. (2019). Improving variational autoencoder with deep feature consistent and generative adversarial training. Neurocomputing, 341, 183-194.

[64]. Vahdat, A., & Kautz, J. (2020). NVAE: A deep hierarchical variational autoencoder. Advances in neural information processing systems, 33, 19667-19679.

[65]. El-Kaddoury, M., Mahmoudi, A., & Himmi, M. M. (2019). Deep generative models for image generation: A practical comparison between variational autoencoders and generative adversarial networks. In Mobile, Secure, and Programmable Networking: 5th International Conference, MSPN 2019, Mohammedia, Morocco, April 23–24, 2019, Revised Selected Papers 5 (pp. 1-8). Springer International Publishing.

[66]. Li, Y., Swersky, K., & Zemel, R. (2015, June). Generative moment matching networks. In *International conference on machine learning* (pp. 1718-1727). PMLR.

[67]. Liu, L., Xi, Z., Ji, R., & Ma, W. (2019). Advanced deep learning techniques for image style transfer: A survey. Signal Processing: Image Communication, 78, 465-470.

[68]. Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. (2021). A brief review of domain adaptation. Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020, 877-894.

[69]. You, K., Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2019). Universal domain adaptation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2720-2729).

[70]. Ma, X., Gao, J., & Xu, C. (2021). Active universal domain adaptation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8968-8977).

[71]. Noema, Y. (2022). What is CycleGAN and how to use it? Medium. Retrieved from https://medium.com/imagescv/what-is-cyclegan-and-how-to-use-it-2bfc772e6195

[72]. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).

[73]. Mishra, O. (2023). Stable Diffusion Explained: How does Stable diffusion work? Explaining the tech behind text to image generation. Medium. Retrieved from https://medium.com/@onkarmishra/stable-diffusion-explained-1f101284484d

[74]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[75]. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

[76]. Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1905-1914).

[77]. Dubey, R., & Gajjar, R. (2023, December). Real-Time Image Super-Resolution using Drone through GFPGAN and Nvidia Jetson Nano. In 2023 9th International Conference on Signal Processing and Communication (ICSC) (pp. 440-444). IEEE.

[78]. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).

[79]. Chen, Y., Liu, S., & Wang, X. (2021). Learning continuous image representation with local implicit image function. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8628-8638).

[80]. Yu, F., Gu, J., Li, Z., Hu, J., Kong, X., Wang, X., ... & Dong, C. (2024). Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild. arXiv preprint arXiv:2401.13627.