DEEPBLOCK

EBOOK

Geospatial MLOps in Defense

Trends & Challenges in 2023

*Geospatial data and machine learning are the two powerful technologies that are transforming the way the military operates. The combination of these technologies has given rise to a new field called **Geospatial MLOps** (Geospatial Machine Learning Operations). Geospatial MLOps enables military organizations, intelligence agencies, and key industrial players to develop state-of-the-art AI models. These models enable analysts to process large volumes of remote sensing imagery and commanders to shorten the path to critical decision-making.*

However, the adoption of Geospatial MLOps in defense comes with its own set of challenges. This ebook sums up the research and industry know-how of Deep Block experts and explores the current trends and challenges associated with the adoption of Geospatial MLOps in the defense sector.

At Omnis Labs, the company behind Deep Block, our steadfast vision is to actualize a future where artificial intelligence is not a far-fetched idea, but a ubiquitous reality. We fervently believe that everyone should have the opportunity to delve into the realm of AI and construct their own models. To that end, we are pioneering one of the foremost AI collaborative ecosystems. Our mission is to empower individuals and businesses alike to seamlessly harness the power of AI and revolutionize the way we live, work, and interact.

I hope you have a great read and that this ebook helps you shed a light on new and fascinating AI technologies.

Gwihwan Moon, CEO of Omnis Labs, Creator of Deep Block

SUMMARY

Geospatial MLOps in Defense
Trends & Challenges in 2023

Chapter 1: Introduction to Geospatial MLOps in Defense.

The surge in geospatial data accessibility from both state-owned and commercial Earth Observation (EO) satellites has paved the way for an increasing number of governmental actors and international organizations to monitor and scrutinize different aspects of the battlefield. In fact, the landscape of satellite ownership and operation has **transformed drastically since the 1960s**, when only the U.S, France, the USSR, UK, and Canada held the ability to launch or possess satellites. As of 2023, approximately 40 countries are equipped with EO satellites.

Moreover, the commercial satellite imagery industry has experienced an exponential upsurge over the last decade, escalating **from a mere 11 satellites in 2012 to a staggering 441 in 2022**. In times of active conflicts, the use of commercial imagery has proven to be crucial, as evidenced by the war in Ukraine, where Ukrainian troops gained access to data from U.S. companies within minutes of its collection. This accessibility to readily available data has indeed made a critical difference, not only from a military planning perspective but also with regard to public information dissemination.

In 2020, **commercial entities had surpassed governmental and military organizations** in terms of the quantity of satellites in operation. Concurrently, the United States launched its new First Space Force Doctrine, aimed at supporting the increasing number of privately-held ventures beyond the Earth's atmosphere. Now, **over 50% of privately-owned satellites are held by companies based in the United States.**

Today, **close to 1000 EO satellites are orbiting in space and capturing petabytes of data every day**, a massive amount of information begging to be analyzed.

Growing number of Earth Observation satellites operated worldwide over the last 30 years, by activity.

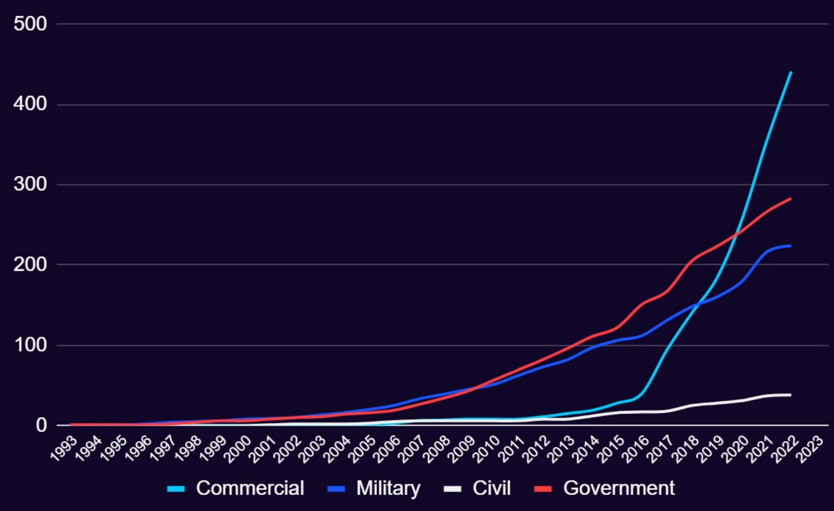


Chart: Deep Block - Source: Union of Concerned Scientists

The sheer volume and complexity of geospatial data make it **challenging to analyze and extract meaningful insights**. This is where machine learning comes in. Machine learning algorithms can be trained to recognize patterns in geospatial data and provide real-time predictions based on those patterns. The combination of geospatial data and machine learning enables the military to **make informed decisions in real-time**.

Geospatial technology is the use of location data to solve problems and answer questions. It encompasses everything from GPS tracking and satellite imagery to geocoding and spatial analysis. **Geospatial data** can come from a variety of sources, including satellites, drones, ground-based sensors, and other devices. **Machine Learning Operations (MLOps)** is the practice of automating machine learning workflows, from data preparation to deployment.

Geospatial MLOps is the intersection of these two fields, combining geospatial technology with MLOps to **improve the accuracy and efficiency of location-based models**, transforming the way that military organizations approach geospatial data analysis. By automating the process of analyzing, processing, and visualizing location data, analysts can streamline the identification of patterns, trends, and anomalies in geospatial data, providing a level of accuracy and insight that was previously impossible.

The integration of geospatial data analysis and machine learning operations is not a new concept. However, the term Geospatial MLOps has gained popularity in recent years due to its relevance in defense. Its **flexibility and customizability make it a powerful tool** that enables the military to process and analyze large volumes of geospatial data, which ultimately helps provide situational awareness and support decision-making.

Geospatial MLOps consists of different components, including **data acquisition, data management, model training, and model deployment.** The data acquisition component involves the collection and processing of geospatial data from different sources. The data management component focuses on the storage and processing of the data. The model training component involves the development and training of machine learning algorithms, while the model deployment component focuses on the integration of the trained models into operational systems.

Deep Block specializes in the model training phase of the Geospatial MLOps workflow (*see chapter 5*) and helps governmental agencies, military organizations, key industry players develop and deploy cutting-edge computer vision models.

Chapter 2: History and modern trends in Geospatial MLOps in Defense.

The defense industry has always been was one of the primary drivers of technological innovation, especially in the realm of Artificial Intelligence. The military's need for advanced technologies to support their operations and improve their strategic capabilities has always been a driving force behind the development of new technologies, including Machine Learning. Geospatial MLOps is a relatively new field in the defense industry. However, the use of geospatial data and machine learning in defense has a longer history.

- Geospatial data has been used for military purposes for decades. In the 1960s, satellite imagery was used in the Cuban Missile Crisis to monitor missile sites and confirm their destruction. Since then, satellite imagery has become an **important source of geospatial data for the military**. However, the volume and complexity of geospatial data have increased significantly in recent years with the advent of drones and other advanced sensors. This has led to the need for advanced machine learning algorithms to process and analyze the data.

- The use of machine learning in defense has also been around for decades. In the 1980s, the US Department of Defense (DoD) created the Defense Advanced Research Projects Agency (DARPA) to **fund research in emerging technologies, including machine learning**. One of the earliest successful applications of machine learning in defense was the DARPA-funded Autonomous Land Vehicle project in the 1990s, which used neural networks to recognize obstacles and plan routes for unmanned ground vehicles.

- In the 2000s, machine learning was **increasingly used in defense for tasks such as target recognition and classification**. For example, the US Air Force used machine learning algorithms to classify objects in airborne radar imagery. However, the deployment of machine learning models in the defense industry was often limited by technical and logistical challenges.

- It was not until the mid-2010s that Geospatial MLOps started to emerge as a new field in the defense industry. This was driven in part by the increasing volume and complexity of geospatial data and the **need for advanced analytics to extract insights from the data**. Additionally, advances in cloud computing and other technologies have made it easier to deploy and manage machine learning models at scale.

Today, Geospatial MLOps is a rapidly growing field in the defense industry. Defense agencies around the world are **heavily investing in developing new machine learning algorithms** for geospatial analysis, as well as new tools and platforms for managing and deploying machine learning models in operational settings. Its adoption is pushed by several trends:

1. **Increased use of cloud-based solutions for Geospatial MLOps**: Cloud-based solutions provide scalability and flexibility, allowing organizations to quickly and easily deploy ML models for geospatial analysis. Cloud-based solutions also enable real-time monitoring of geospatial data, allowing organizations to make quick and informed decisions.

2. **AI-driven automation for geospatial data analysis**: AI-driven automation can help to reduce the amount of time and effort required to analyze geospatial data, enabling defense organizations to focus on other critical tasks. Automation can also help to reduce the risk of human error, improving the accuracy and reliability of geospatial analysis.

3. **Edge computing for Geospatial MLOps:** Edge computing involves processing data at the edge of the network, rather than in the cloud or a central data center. Edge computing can help to reduce latency and improve the performance of ML models, enabling real-time analysis of geospatial data.

4. **Increased adoption of explainable AI for Geospatial MLOp:** Explainable AI involves making ML models more transparent and interpretable, enabling humans to understand how the model arrived at a particular decision. This is particularly important in defense applications, where decisions made based on ML insights can have significant consequences.

5. **Use of synthetic data for Geospatial MLOps**: Synthetic data can help to address the challenge of representative training data, enabling organizations to train ML models on a wider range of scenarios. This can help to improve the accuracy and reliability of geospatial analysis, particularly in scenarios where real-world data is limited.

Chapter 3: The benefits of using Geospatial MLOps in Defense.

Geospatial MLOps is rapidly becoming a game-changing technology, providing unparalleled opportunities and benefits for the defense industry:

1. **Improved Situational Awareness**: Analyze data from a variety of sources, including satellites, UAVs, planes, and ground-based sensors, to gain a real-time understanding of the battlefield. This enables commanders to make informed decisions and quickly adjust to changing conditions. For example, during the Iraq War, the U.S. military used geospatial data to monitor insurgent movements and identify potential threats. This allowed them to adjust their tactics and ultimately reduce the number of attacks.

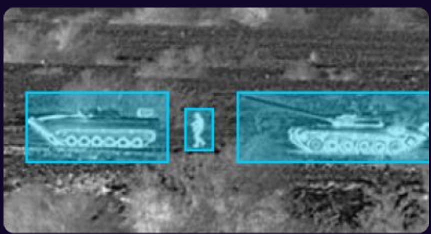
2. **Reduced Human Error**: Automate time-consuming and error-prone tasks, such as image analysis and target identification, freeing up human analysts to focus on more complex tasks. This can greatly reduce the risk of human error and improve overall accuracy. For instance, automated target recognition systems use machine learning algorithms to analyze video feeds and identify potential targets. This reduces the workload on human analysts and improves the accuracy of target identification.

3. **Enhanced Predictive Capabilities**: Analyze vast amounts of data to identify patterns and make predictions about future events. This enables defense agencies to anticipate potential threats and take proactive measures to prevent them. Indeed, predictive analytics can forecast the likelihood of an attack in a particular region, allowing commanders to adjust their strategy accordingly.

4. **Improved Efficiency**: Automate many of the tasks associated with data analysis, including data cleaning, feature extraction, and model selection. This can greatly improve the efficiency of data processing and reduce the time required to generate actionable insights. For example, **Deep Block** has developed an automated image analysis system that uses machine learning algorithms to detect changes in satellite imagery. This greatly reduces the time required for manual analysis and improves the overall efficiency of the process.

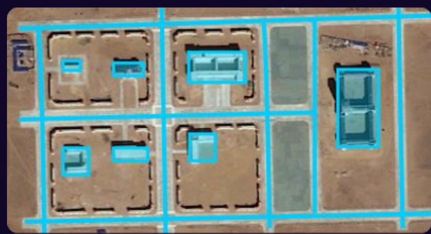
5. **Cost Savings**: Optimize resource allocation and reduce operational costs. By automating many of the tasks associated with data analysis, defense agencies can reduce the need for human analysts and improve overall efficiency. This can result in significant cost savings over time. For example, the U.S. Navy has developed an automated system for detecting underwater mines using sonar data. This reduces the need for expensive human divers and improves the accuracy of mine detection.

USE CASES



Target Recognition
ML algorithms are used to analyze satellite and drone imagery to identify targets on the ground, such as vehicles or weapons caches. This enables faster and more accurate targeting by military forces.

Maritime Domain Awareness
With thousands of vessels on the high seas at any given time, monitoring maritime activity can be a daunting task. Geospatial MLOps can help identify patterns in vessel activity. IMINT analysts can more easily detect and deter threats, such as piracy and smuggling.



Mission Planning
Planning military operations is a complex and time-consuming process. Geospatial MLOps can provide real-time insights into the location and movements of targets. By doing so, mission planners can plan and execute operations more quickly and with greater accuracy.

Border Security
Geospatial models can also be used to monitor borders and detect intrusions in real-time, providing faster response times to potential security threats. This is particularly important in border areas that are difficult to monitor using traditional surveillance methods.

Terrain Analysis
ML algorithms can analyze satellite imagery and geospatial data to identify terrain features such as hills, valleys, and waterways. This information is crucial for military planners to develop effective operational plans.

Air Situation Monitoring
Collect information related to the enemy's air force strength, movements, and activity on land, in the air, or at sea. Train your AI to detect, classify, and inventorize any type of aircraft as well as map airfield infrastructures and capabilities.

Chapter 4: The challenges of Geospatial MLOps adoption in Defense

Despite the many benefits of machine learning in the defense industry, there are also **significant challenges that must be overcome**. However, recent Geospatial MLOps solutions development like Deep Block can help defense agencies and industrial players surmount every hurdle.

CHALLENGES

- **Data Access and Quality**: Access to high-quality data is essential for any AI model to work effectively. In the defense industry, this can be a significant challenge. According to a survey by the Defense Advanced Research Projects Agency (DARPA), data access and quality were identified as the top challenge in deploying AI models in the defense industry, with **60% of respondents citing it as a significant obstacle.**

- **Lack of Expertise**: AI is a complex field, and deploying AI models in the defense industry requires a high level of expertise. In addition, there is a well-known scarcity of professionals in this particular domain. According to a survey by the Center for a New American Security, **68% of defense industry professionals identified a lack of AI talent as a significant challenge.** This shortage of expertise can make it difficult to develop and deploy effective AI models.

- **Deployment Success Rate**: Even when an AI model has been developed, deploying it in the defense industry can be challenging. A survey by the Government Business Council found that **only 21% of federal government employees believed that their agency had been successful in deploying AI models.** This low success rate can be attributed to the challenges outlined above, as well as other factors such as funding constraints and cultural resistance to new technologies.

- **Integration with Existing Systems**: The defense industry relies on a complex network of systems, and integrating AI models into this network can be challenging. For example, an AI model may need to integrate with existing sensor networks or command and control systems. According to a survey by Accenture, **35% of defense industry professionals identified integration with existing systems as a significant challenge in deploying AI models.**

- **Security Concerns**: The defense industry is particularly concerned with security, and this includes the use of AI models. While AI can help with identifying and mitigating security threats, it can also pose its own security risks. For example, adversaries may attempt to hack into an AI system to manipulate the data or cause the model to fail. A survey by Deloitte found that **44% of defense industry professionals believe that security concerns are a significant challenge in deploying AI models.**

SOLUTIONS

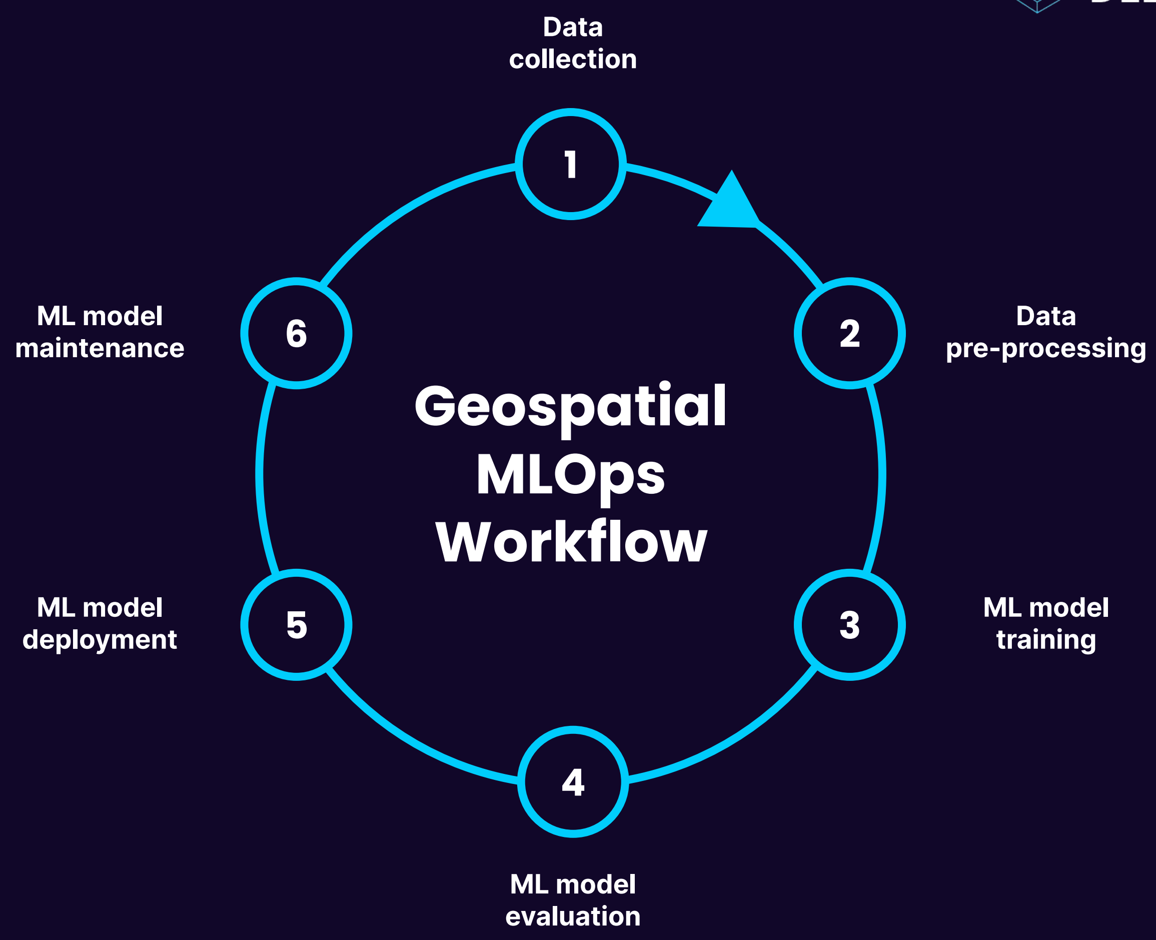
- **Data Access and Quality:** One solution to address the challenge of data access and quality is to invest in **data management systems**. These systems can help standardize data and ensure its quality, making it easier for AI models to process and analyze. Additionally, the defense industry can **collaborate with data providers** to increase the availability of high-quality data.

- **Lack of Expertise**: To address the shortage of AI talent, defense agencies can **invest in training programs** to develop in-house expertise. This includes providing education and training for existing employees and recruiting new talent with specialized skills. Additionally, developers can **rely on AI model training platform** such as Deep Block to bring their projects to fruition without investing in an enterprise-grade AI pipeline.

- **Deployment Success Rate**: One solution to improve deployment success rates is to **prioritize collaboration and communication** among all stakeholders. This includes involving end-users in the development process to ensure that the AI models meet their needs. Furthermore, by conducting thorough testing and evaluation before deployment, the defense industry can increase the likelihood of successful implementation.

- **Integration with Existing Systems**: To overcome the challenge of integrating AI models into existing systems, defense actors can invest in **application programming interfaces (APIs)**, which Deep Block also provides, that enable seamless integration between systems. Furthermore, by involving system developers in the AI model development process, it is possible to ensure that the AI models can integrate into existing systems with minimal disruption.

- **Security Concerns**: Defense organizations can address security concerns by **developing robust security protocols** and investing in cybersecurity. This includes implementing multi-factor authentication and encrypting data to prevent unauthorized access. Additionally, by conducting regular security audits, the defense industry can identify and address vulnerabilities before they are exploited.



Chapter 5: Geospatial MLOps Workflow

Geospatial MLOps involves a series of processes that enable the training and deployment of machine learning (ML) models for geospatial analysis in defense applications. The Geospatial MLOps workflow can be divided into several key steps.

1. **Data collection**: Geospatial data can come from various sources, including satellites, drones, and other sensors. The data needs to be in a format that can be used for ML training. Data collection can involve a range of tools and technologies, including data processing pipelines and cloud-based storage solutions.

2. **Data pre-processing**: The data needs to be cleaned and pre-processed to ensure that it is in a format that can be used for ML training. This step can involve the use of data cleaning tools, data normalization techniques, and data augmentation methods.

3. **ML model training**: This step involves selecting the appropriate ML algorithms and training the model on the pre-processed data. The use of synthetic data can help to address the challenge of representative training data. Model development can involve a range of ML algorithms, including deep learning, decision trees, and support vector machines.

4. **Model evaluation**: The performance of the ML model needs to be evaluated to ensure that it is accurate and reliable. This step can involve the use of performance metrics, such as precision, recall, and F1 score.

5. **Model deployment**: This step involves deploying the ML model into a production environment. This can involve the use of containerization and microservices to enable scalability and flexibility. The deployment of the model needs to be closely monitored to ensure that it is performing as expected.

6. **Model maintenance**: The ML model needs to be updated and maintained to ensure that it continues to perform effectively. This step can involve the use of monitoring and feedback loops to ensure that the system is performing as expected.

One of the main challenges of the Geospatial MLOps workflow is the need to balance accuracy with interpretability. ML models can be complex and difficult to interpret, which can make it challenging to explain how the model arrived at a particular decision. This is particularly important in defense, where decisions made based on ML insights can have significant consequences.

Chapter 6: Geospatial Data Acquisition

Defense, Military and Intelligence agencies around the world use geospatial data for a variety of purposes, including intelligence gathering, planning and executing military operations, and monitoring global events. Intelligence gathering primarily falls under two categories - **Geospatial Intelligence (GEOINT)** and its subset, **Imagery Intelligence (IMINT)**. These disciplines have the most to gain from the adoption of Geospatial MLOps. GEOINT uses geospatial data in combination with human intelligence and signals intelligence to gain a holistic understanding of an area of interest, including terrain, infrastructure, and other relevant features. IMINT focuses on analyzing imagery to provide more detailed information on the geographic features of an area, as well as the activities of individuals, organizations, and foreign military forces.

As such, security organizations collect a wide range of geospatial data types that provide critical information for their operations.

Imagery data
helps identify potential targets, monitor military activity in a particular area, and assess the impact of military operations

Topographic data
provide information on the terrain and elevation of an area, which is critical for planning military operations.

Weather data
can provide information on current and future weather conditions, which can impact tactical execution.

Demographic data
provide insights into the population of an area, which can inform military decision-making.

Geospatial data acquisition is a critical component of Geospatial MLOps. One of the main challenges in geospatial data acquisition is the large volume of data that needs to be processed. The quality and quantity of geospatial data directly affect the accuracy and reliability of the insights that can be obtained from it.

This data needs to be collected, stored, and managed in a way that makes it accessible and useful for analysis. The process of data acquisition involves several steps, including **data collection, pre-processing, and integration**.

In defense, **Data collection** involves the use of various sensors and platforms to capture geospatial data.

Satellites provide a global coverage of Earth's surface and can capture high-resolution imagery. They can be classified into two main types: optical and radar. Optical satellites use visible and near-infrared light to create images of the Earth's surface, while radar satellites use radio waves to create images, which are particularly useful for imaging through clouds and at night.

UAVs, which include drones are a growing source of information. They, can be used to capture aerial imagery and perform other sensing tasks. They can operate at lower altitudes than satellites, providing higher resolution imagery, and can be equipped with a variety of sensors, including cameras, LiDAR, and thermal sensors.

Planes can be used to collect geospatial data through aerial photography, LiDAR, and other sensors. Aerial photography involves capturing high-resolution images of the Earth's surface from an airplane, while LiDAR uses lasers to measure distances and create 3D maps of the terrain.

Ground-based sensors are used to collect weather or seismic data but also to capture footage and monitor critical infrastructure such as bridges, dams, and power plants for any signs of damage or sabotage. They can be deployed in remote or hazardous areas where it may not be safe or practical to send human observers, such as in conflict zones or disaster areas.

The ongoing conflict between Ukraine and Russia has ushered in the era of the first full-scale drone war. The utilization of **hundreds of reconnaissance and attack drones on a daily basis** has become increasingly integrated with every aspect of combat and a growing source of intelligence. These drones have demonstrated their ability to rapidly bridge the gap with enemy positions, mark the coordinates of enemy command posts, artillery, or ammunition depots, coordinate attacks with other units, and confirm the destruction of the enemy's infrastructure.

Once the data has been collected, it needs to be pre-processed to remove any errors or inconsistencies. **Pre-processing** may involve tasks such as image correction, filtering, and calibration. The pre-processing step is critical to ensure that the data is accurate and ready for analysis.

After pre-processing, the data needs to be integrated with other sources of data. This can involve tasks such as **data fusion**, where data from different sources are combined to provide a more comprehensive picture. Integration can be challenging, as the data may come in different formats and resolutions.

The **integration** of geospatial data from different sources is critical for defense applications. It enables the military to have a more comprehensive understanding of the environment and potential threats. For example, geospatial data can be used to monitor troop movements, detect changes in infrastructure, and identify potential risks.

Chapter 7: Model development in Geospatial MLOps

Model development is a crucial step in the process of implementing Geospatial MLOps in the defense industry. It involves selecting the appropriate ML algorithms and training the model on pre-processed data.

Even if algorithms and models are used interchangeably, they actually refer to different things:

- An **AI algorithm** is a set of rules or instructions that a computer program follows to perform a specific task. The algorithm may involve mathematical calculations, statistical analysis, or other computational techniques.

- An **AI model** is a specific implementation of an AI algorithm that has been trained on data to perform a specific task. The model is essentially the result of the algorithm's application to a particular problem, and it reflects the learned patterns in the data.

In other words, an **AI algorithm is like a blueprint, while an AI model is the actual implementation of that blueprint**. An AI algorithm defines the rules and procedures for a specific problem, while an AI model is the end result of applying those rules and procedures to a specific dataset.

Model development can involve a range of ML algorithms, such as:

- **Convolutional Neural Networks (CNNs)** are one of the most widely used deep learning algorithms used in Geospatial MLOps. These networks are particularly useful for image recognition tasks, such as detecting vehicles, buildings, and other objects. CNNs use a technique called convolution, where a small filter is passed over the image to extract features. These features are then used to classify the object in the image. CNNs have proven to be very effective for object detection in satellite imagery.

- **Recurrent Neural Networks (RNNs)** are another type of neural network which, unlike CNNs, are designed for sequential data, such as time-series data or text. RNNs can be used for tasks such as forecasting weather patterns, predicting enemy movements, and more. In addition, RNNs can be combined with CNNs to analyze both image and sequential data, making them a powerful tool for defense agencies.

- **Decision Trees** are a type of supervised learning algorithm used for classification and regression tasks. They are particularly useful in Geospatial MLOps for tasks such as land cover classification, where the goal is to classify different types of terrain or vegetation. Decision Trees work by recursively splitting the data into subsets based on the most significant feature. The result is a tree-like model where each leaf node represents a class label.

- **Support Vector Machines (SVMs)** are a type of supervised learning algorithm used for classification and regression tasks. SVMs work by finding the hyperplane that maximally separates the data into different classes. They are particularly useful in Geospatial MLOps regarding tasks such as landmine detection, where the goal is to classify a particular area as safe or dangerous. SVMs have proven to be very effective in detecting subtle patterns in the data.

- **Generative Adversarial Networks (GANs)** are neural network variations used for generating new data based on a given set of training data. GANs consist of two networks: a generator network that creates new data, and a discriminator network that tries to distinguish between the generated data and real data. GANs are particularly useful in Geospatial MLOps for generating synthetic data for training CNNs. Synthetic data can help address the challenge of representative training data, which is often a problem in defense applications.

COMPUTER VISION MODELS

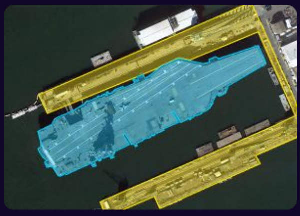


Image segmentation
Image segmentation works by dividing an image into several segments, each with a unique label, to make it easier to analyze and extract information from the image. It is used for target identification, terrain analysis, and building mapping.

Object detection
Object detection can identify and localize objects in an image or video stream. This model is used in various defense applications, such as aerial reconnaissance, target tracking, and vehicle identification.

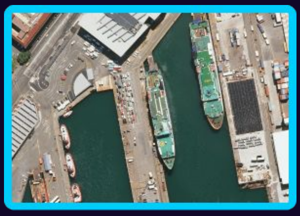


Image classification
Image classification works by classifying an image into specific categories based on its features. It is used for terrain analysis, target identification, and mission planning.

Change Detection
Multi-temporal change detection can analyze changes that occur over time in an area of interest. It is used for environmental monitoring, land cover mapping, and infrastructure analysis.

Chapter 8: Model training in Geospatial MLOps

To build efficient and accurate models, the data used for training must be appropriately labeled. Data labeling involves tagging or annotating data to provide context and meaning to the model, allowing it to learn patterns and make accurate predictions.

Data labeling, also known as annotation, is the process of adding metadata or labels to data to make it more informative and easier to analyze. In defense, geospatial data, such as satellite imagery, is essential for a wide range of tasks, including surveillance, reconnaissance, and mapping. However, the data is usually massive and unstructured, making it challenging to analyze and extract meaningful insights. Data labeling **helps to organize and structure the data, making it easier to use for machine learning models**. For example, in object detection tasks, such as identifying military equipment or personnel in satellite imagery, data labeling involves marking the location, size, and shape of each object of interest. Similarly, in image segmentation tasks, data labeling involves assigning a label to each pixel in the image to indicate the class it belongs to, such as land, water, or buildings.

There are different types of data labeling methods, including manual and automatic labeling. **Manual labeling** involves the use of human annotators to manually add labels to data. This process is time-consuming and can be prone to errors, but it provides more accurate and reliable results. **Automatic labeling**, on the other hand, uses AI algorithms to label data automatically, based on predefined rules or machine learning models. This process is faster and more scalable, but it may not always produce accurate results.

Data labeling remains a labor-intensive task that requires skilled human annotators to go through large volumes of data and accurately label each item. The accuracy and quality of the labels are essential for the success of the machine learning models. Hence, it is crucial to have a well-defined data labeling process and quality control mechanisms in place:

1. **Define clear labeling guidelines:** To ensure consistency in data labeling, it is important to define clear guidelines for annotators. These guidelines should cover the types of labels to be used, the format of the labels, and any specific rules or criteria for labeling.

2. **Use multiple annotators:** To ensure the accuracy and reliability of data labeling, it is recommended to use multiple annotators to label the same data. This helps to identify any inconsistencies or errors in the labeling and improve the quality of the data.

3. **Validate the labeling:** After the data has been labeled, it is important to validate the labeling to ensure its accuracy and reliability. This can be done by comparing the labeled data with ground truth data or by using metrics such as precision, recall, and F1 score.

4. **Use tools for automatic labeling:** To improve the efficiency and scalability of data labeling, it is recommended to use software tools for automatic labeling. These tools can help to speed up the labeling process and reduce the risk of errors.

5. **Monitor the labeling process:** It is important to monitor the data labeling process to ensure that it is progressing according to the defined guidelines and quality standards. This can be done by regularly reviewing the labeled data and providing feedback to annotators.

KEY PRINCIPLES

Supervised learning involves providing the model with labeled data, where each data point has a corresponding label or target. The model then learns to identify patterns in the data that relate to the labels, making it possible to predict the label of new, unlabeled data. In Geospatial MLOps, this technique is often used to classify satellite imagery into different categories such as vegetation, water bodies, and man-made structures.

Unlike supervised learning, **unsupervised learning** involves training the model on unlabeled data. The model then learns to identify patterns and structures within the data, without any pre-existing knowledge of the target labels. Unsupervised learning is used for Geospatial MLOps applications such as clustering and anomaly detection.

Semi-supervised learning is a hybrid approach that involves training the model on both labeled and unlabeled data. The model uses the labeled data to learn the patterns associated with the labels, and then applies this knowledge to the unlabeled data to make predictions. Semi-supervised learning can be useful in Geospatial MLOps for tasks where labeled data is scarce, but unlabeled data is abundant.

Active learning is a technique that involves the model actively selecting data points for labeling. The model selects the data points that it is most uncertain about, and then requests human input to label those data points. This approach can be useful in Geospatial MLOps, where human expertise is often required to identify and label complex features in the data.

Chapter 9: Model deployment in Geospatial MLOps

In Geospatial MLOps, computer vision models are trained on large datasets to detect and classify objects, perform image segmentation, and identify changes in the environment over time. However, the real value of these models is realized when they are deployed in operational environments. Here are the different steps to achieve this:

1. **Preparing the Model for Deployment**: Before deploying a computer vision model, it is important to ensure that it is compatible with the hardware and software infrastructure of the operational environment. This includes optimizing the model for performance and memory usage on the target hardware, such as a remote sensing platform or a ground-based system. The model must also be packaged into a deployable format, such as a Docker container, that can be easily deployed and managed in the operational environment. The container should include all the necessary dependencies, such as the model's architecture, weights, and pre-processing code, as well as any required libraries and tools.

2. **Deploying the Model:** Once the model is prepared, it can be deployed to the operational environment. The deployment process involves installing the container on the target system and configuring it to receive input data and generate output results. In Geospatial MLOps, computer vision models may be deployed in a variety of settings, including ground-based sensors, unmanned aerial vehicles (UAVs), and satellites. Each environment presents unique challenges, such as limited bandwidth, processing power, and memory, that must be taken into account during deployment.

3. **Input Data Processing:** Once the model is deployed, it must be configured to receive input data in the appropriate format. This may involve pre-processing the data to ensure that it is in a format that can be understood by the model. For example, in geospatial applications, input data may be in the form of satellite imagery or LiDAR data. The data may need to be pre-processed to remove noise, enhance features, and adjust for atmospheric conditions. The pre-processing steps will depend on the specific model and the nature of the input data.

4. **Model Execution:** After the input data is processed, it is fed into the deployed model for execution. The model generates output results, such as object detection, image segmentation, or change detection, based on the input data and the model's training. The output results may be in the form of raw data, such as pixel values or feature vectors, or they may be in a more user-friendly format, such as a visualization or a report. The output results may also be used as input to other systems or models for further analysis and decision-making.

5. **Result Analysis and Feedback:** Finally, the output results must be analyzed and interpreted in the context of the operational environment. This may involve visualizing the output results to identify patterns and trends, or it may involve statistical analysis to identify significant changes over time. The results may also be compared to ground truth data, such as data collected by human operators, to validate the model's performance and identify areas for improvement. The feedback generated from this analysis can be used to refine the model and improve its performance in future deployments.

KEY PRINCIPLES

The deployment of ML models is a critical step in the Geospatial MLOps process, as it determines how well the model performs in real-world scenarios. Here are the key principles and practices of model deployment:

- **Scalability** — Geospatial ML models often require high-performance computing resources to process large geospatial datasets. It is essential to have scalable deployment architectures, such as containerization and cloud computing, to ensure that the models can handle large workloads.

- **Performance monitoring** — It is important to monitor the performance of ML models in real-time, to ensure that they are meeting their performance goals. This requires monitoring the key performance indicators of the model, such as accuracy, precision, recall, and F1 score.

- **Interpretability** — Geospatial ML models often make decisions that have significant impacts on the mission outcomes. It is essential to have interpretability measures in place to understand how the model is making decisions and to ensure that these decisions are aligned with mission objectives.

- **Model retraining** — Geospatial data is often dynamic, with new data being generated on a continuous basis. It is essential to have effective retraining procedures in place to ensure that the model remains up-to-date and continues to perform effectively over time.

- **Versioning** — Geospatial ML models often undergo iterative development, with multiple versions of the model being deployed in different scenarios. It is essential to have effective versioning procedures in place to ensure that the correct version of the model is deployed in each scenario.

- **Security** — Geospatial ML models often deal with sensitive and classified data, and it is important to have appropriate security measures in place to protect the model and the data it processes. This includes measures such as access control, encryption, and secure communication protocols.

Chapter 10: Model monitoring and maintenance in Geospatial MLOps

The deployment of these models is not a one-time process; rather, it is an ongoing effort that requires continuous monitoring and maintenance to ensure their optimal performance.

- A study by Accenture found that companies that regularly update and maintain their machine learning models achieve an **average improvement of 25% in model performance.**
- According to a survey by Algorithmia, the most common reason for model failure is data drift, which occurs when the data used to train the model no longer accurately reflects the real-world data. In the same survey, it was found that **60% of data scientists spend at least 20% of their time on model maintenance.**
- In a survey by Databricks, **72% of data professionals reported that they had experienced at least one production issue** with their machine learning models in the previous 12 months.
- A report by Gartner estimates that **75% of machine learning models require human intervention** to ensure their accuracy and reliability.
- A survey by NewVantage Partners found that **77% of executives believe that their company is experiencing significant challenges** in operationalizing machine learning, with the majority of challenges related to data management, data quality, and data governance.

In defense, computer vision models are often used in critical applications such as surveillance and target recognition. As such, any error or delay in these models can have serious consequences. Model monitoring and maintenance are therefore essential practices to ensure that these models are functioning optimally at all times.

Model monitoring involves tracking the performance of a model over time, with the goal of identifying any issues that may arise. This can involve monitoring metrics such as accuracy, precision, and recall, as well as tracking the frequency and severity of false positives and false negatives. By keeping an eye on these metrics, teams can quickly identify when a model's performance begins to degrade and take corrective action.

Model maintenance, on the other hand, involves making updates and improvements to a model over time to ensure it continues to perform at its best. This can include retraining the model on new data, updating the model architecture to improve performance, or fine-tuning hyperparameters for optimal performance. By regularly maintaining models, teams can ensure that they continue to deliver the desired results and stay up-to-date with the latest advances in computer vision technology.

EXAMPLES

- **Object Detection for Surveillance**: Imagine a computer vision model that is used to detect suspicious activity in a high-security area. The model was initially trained on a dataset of images taken during the day, but now it is being used to monitor the area 24/7. Over time, the model begins to perform poorly at night, with a high rate of false positives and false negatives. By monitoring the model's performance, the team is able to quickly identify this issue and take corrective action - for example, by retraining the model on a dataset of night-time images.
- **Image Segmentation for Target Recognition**: Another example involves a computer vision model that is used to segment images of military targets, enabling more accurate targeting and decision-making. Over time, the model begins to produce inconsistent results, with segments that are incorrectly classified or misaligned with the target. By monitoring the model's performance, the team is able to identify the issue and take corrective action - for example, by fine-tuning the model's hyperparameters or updating its architecture.
- **Image Classification for Intelligence Analysis**: Finally, let's consider a computer vision model that is used to classify images of potential military targets, enabling intelligence analysts to quickly identify key information. Over time, the model begins to struggle with certain types of images, such as those with low contrast or unusual lighting conditions. By monitoring the model's performance, the team is able to identify the issue and take corrective action - for example, by retraining the model on a more diverse set of images.

QUALITY • INTEGRATION • STORAGE • GOVERNANCE • LABELING • VERSIONING

Chapter 11: Data management in Geospatial MLOps

Data is the lifeblood of Geospatial MLOps, and effective data management is essential for the success of Geospatial MLOps applications. In this chapter, we will explore the key principles and practices of data management in Geospatial MLOps.

1. **Data quality** — Geospatial data is often complex, heterogeneous, and subject to errors and biases. It is essential to ensure that the data used in Geospatial MLOps applications is accurate, complete, and free from errors and biases. This requires careful data cleaning, preprocessing, and quality control procedures.

2. **Data integration** — Geospatial data comes from a variety of sources, such as sensors, satellites, and other data providers. It is essential to integrate these data sources into a single, unified data repository, to enable effective geospatial analysis and ML model training.

3. **Data storage and retrieval** — Geospatial data is often large, complex, and computationally intensive. It is therefore important to store and retrieve data efficiently, using appropriate storage technologies such as cloud storage, object stores, and data lakes. This enables fast and efficient access to geospatial data for analysis and model training.

4. **Data governance** — Geospatial data is often sensitive and classified, and it is important to have appropriate governance and security measures in place to protect it. This includes measures such as access control, data encryption, and data masking.

5. **Data labeling and annotation** — Geospatial ML models often require labeled and annotated data for training. It is essential to have efficient and accurate data labeling and annotation procedures in place, to enable effective ML model training.

6. **Data versioning and lineage** — Geospatial ML models often require iterative development and training, and it is essential to track the versioning and lineage of data used in each training iteration. This enables effective tracking of data changes, and helps ensure that ML models are based on the most accurate and up-to-date geospatial data.

Chapter 12: Securing data labeling practices

As defense agencies continue to integrate machine learning and computer vision technologies into their operations, ensuring the security and integrity of their data and models becomes increasingly critical. This includes developing and labeling data sets that are both accurate and secure.

CHALLENGE

- **Model Security**: Defense agencies must ensure that the computer vision models they develop are secure and protected from unauthorized access or manipulation.
- **Model Accuracy**: The accuracy of the models is crucial for successful implementation in defense operations. Models should be developed and tested rigorously to ensure that they meet the desired accuracy levels.
- **Model Robustness**: Defense organizations must ensure that their computer vision models are robust enough to handle real-world scenarios and diverse environments.
- **Model Explainability**: Explainable AI (XAI) techniques should be used to develop models that can be easily understood and validated by humans, including decision-makers and operators.

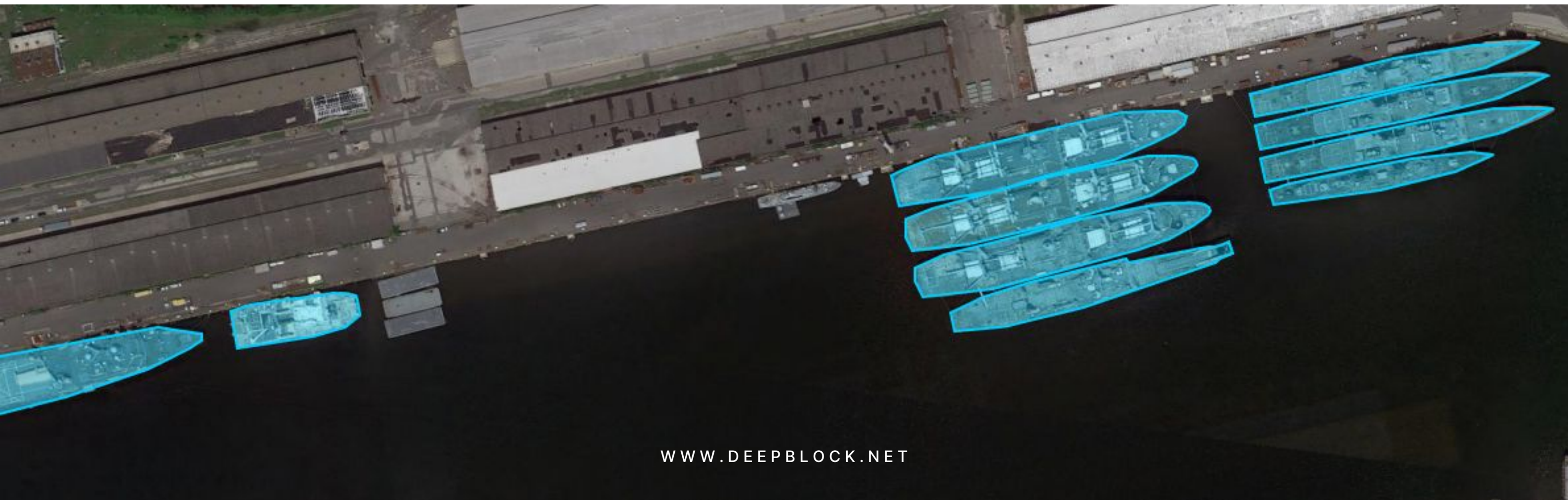
SOLUTION

- **Threat Modeling**: Threat modeling can be used to identify potential vulnerabilities in the models and develop appropriate security measures to mitigate them.
- **Code Reviews**: Conducting regular code reviews can help identify potential security vulnerabilities and ensure that the models are developed securely.
- **Testing and Validation**: Rigorous testing and validation of the models should be performed to ensure they are accurate, robust, and meet the desired performance levels.
- **Explainability Techniques**: Using XAI techniques such as LIME, SHAP, or Local Interpretable Model-Agnostic Explanations (LIME) can help ensure that models are explainable and can be validated by human operators.

In addition to the technical measures, it's important for defense agencies to implement strong policies and procedures for data handling and labeling. These should include clear guidelines for who has access to the data, how it's used, and how it's stored and secured. It's also important to have a **well-defined process for data labeling**, including training for those performing the labeling and quality control checks to ensure accuracy and consistency.

One example of a defense agency successfully implementing these principles is the United States Department of Defense (DoD). The DoD has implemented **strict guidelines for data handling and labeling**, including the use of secure environments and multi-factor authentication for access to sensitive data. The DoD also employs trained personnel for data labeling and has established quality control measures to ensure accuracy.

Statistics show that investing in secure development and labeling practices can have significant benefits for defense agencies. A study by the RAND Corporation found that **improved data quality and management can lead to increased effectiveness in military operations**, while a report by Accenture found that companies that prioritize cybersecurity and data protection see an **average return on investment of 30%**.



Chapter 13: The future of Geospatial MLOps

The field of Geospatial MLOps in the defense industry is rapidly evolving, with new developments and innovations emerging at a fast pace. In the coming years, we will see exponential improvement in computing power, which will make possible exciting new developments in Geospatial MLOps, with a powerful impact on the defense industry:

- **Edge computing in Geospatial MLOps** — Edge computing refers to the processing of data at the edge of the network, closer to the source of the data. This can reduce latency and improve the speed of data processing. In the context of Geospatial MLOps, edge computing can allow for real-time processing of geospatial data and improve the accuracy and responsiveness of Geospatial MLOps applications.

- **Integration of Geospatial MLOps with augmented reality (AR) and virtual reality (VR) technologies** — AR and VR technologies can provide a more immersive and interactive experience for the human operator, allowing them to visualize and interact with geospatial data in new and innovative ways. This can improve situational awareness and enhance the decision-making process.

- **Use of blockchain technology in Geospatial MLOps** — Blockchain technology can provide a secure and transparent way to manage and share geospatial data, which is critical for the defense industry. By using blockchain technology, organizations can ensure the integrity and authenticity of geospatial data, which can improve the accuracy and effectiveness of Geospatial MLOps applications.

- **Use of explainable AI (XAI) in Geospatial MLOps** — XAI refers to the ability of AI algorithms to provide explanations for their decision-making processes. This is particularly important in the defense industry, where the human operator needs to understand the reasoning behind the recommendations provided by Geospatial MLOps applications. By using XAI, organizations can improve the transparency and accountability of Geospatial MLOps applications and build trust in the technology.

- **Use of advanced machine learning algorithms, such as deep learning and reinforcement learning, in Geospatial MLOps** — These algorithms are where Deep Block is focusing its R&D efforts. They can provide more accurate and sophisticated predictions and recommendations, which can improve the effectiveness of military applications.

In summary, the future of Geospatial MLOps in the defense industry is bright, with many potential developments and innovations on the horizon. These include edge computing, AR and VR technologies, blockchain technology, XAI, and advanced machine learning algorithms. By staying abreast of these developments and incorporating them into their military applications, organizations can continue to improve their decision-making capabilities and maintain their competitive edge.

ABOUT US

Deep Block is the world's fastest Geospatial MLOps platform. Using cutting-edge AI algorithms, geospatial specialists can analyze remote sensing imagery at an unprecedented scale by easily training and deploying machine learning models, without any coding required.

Deep Block delivers a production-grade AI pipeline that facilitates the deployment of high-performing Machine Learning models (MLOps) in days instead of months. On top of a library of pre-trained models, AI experts and amateurs alike can train and use their own computer vision algorithms, without any coding knowledge required.

Deep Block is seeing a growing variety of applications for geospatial analysis ranging from Intelligence, Surveillance, and Reconnaissance (ISR), environmental monitoring, disaster management, urban planning, and agriculture.

Deep Block’s algorithms can detect ground objects, and analyze their shape, size, location, and classification (purpose) over a defined period of time. This proves to be particularly useful in surveillance operations for air situation monitoring, maritime awareness, and strategic site monitoring. For example, the rapid expansion of an industrial complex, or the conversion of a civilian airport into a military base can be picked up by the algorithm and automatically flagged.

- **Would you like to see more stories like this?** Visit The Deep Dive, Deep Block's Blog. It regroups all our successes and failures. It is an honest take on the challenges of a Computer Vision startup, an open opinion column for our experts, and the place to get our latest news.

- **Would you like to try Deep Block?** Unlock all our core features for free, for 30 days by registering on app.deepblock.net