

Semantic Image Encoding and Communication for Earth Observation with LEO Satellites

Van-Phuc Bui*, Thinh Q. Dinh^{†§}, Israel Leyva-Mayorga*, Shashi Raj Pandey*, Eva Lagunas[‡], Petar Popovski*,

*Department of Electronic Systems, Aalborg University, Denmark ({vpb, ilm, srp, petarp}@es.aau.dk)

[†]University of Information Technology, Ho Chi Minh City, Vietnam (thinhdq@uit.edu.vn)

[§]Vietnam National University, Ho Chi Minh City, Vietnam

[‡]University of Luxembourg, Luxembourg (eva.lagunas@uni.lu)

Abstract—The substantial volume of data generated by Earth observation (EO) satellites poses a significant challenge to the limited-rate satellite-to-ground links. This paper addresses the downlink communication problem of change detection in multi-spectral satellite images for EO purposes. The proposed method is based on a cohesive strategy capable of eliminating clouds and performing semantic encoding during image processing, as it encodes vital information for the target application, in the form of changed multi-spectral pixels (MPs) to minimize energy consumption. The proposed method is based on a three-stage end-to-end scoring mechanism, which quantifies the significance of each MP before determining its transmission. Specifically, the sensing image is (1) normalized and passed through a high-performance cloud filtering via the Cloud-SLR model, (2) passed to the proposed scoring algorithm that uses Change-Net to identify MPs that have a high likelihood of being changed, compress them, and forward to the ground station, and (3) reconstructed at ground gateway based on the reference image and received data. The numerical results show the effectiveness of the proposed framework in achieving energy savings of up to 58% while upholding the transmission of high-quality data for satellite-based EO applications.

Index Terms—Low Earth Orbit (LEO) satellite communications, Semantic communication, Earth observation, Change detection, Image processing.

I. INTRODUCTION

Remote sensing satellites are vital for environmental monitoring as they swiftly provide comprehensive coverage of targeted regions, allowing for land use surveys, urban studies, and hazard management through image acquisition [2]. However, a major problem is that the use of high-resolution sensors results in large volumes of data, which necessitates significant communication resources and on-board data storage capacity for transmitting data to ground-based end-users. For instance, the Sentinel-2 system acquires an extensive amount of data (2.4 Terabits per day) for transmitting to the terrestrial gateway, with each surface location being captured at periodic intervals of five days [3]. The grow of Low Earth Orbit (LEO) satellite deployments in Earth Observation (EO) applications, coupled with the constrained communication capacity of LEO

This work was supported by the Villum Investigator Grant “WATER” from the Velux Foundation, Denmark. The work of Eva Lagunas has received funding from the Luxembourg National Research Fund (FNR) under the project SmartSpace (C21/IS/16193290). An earlier version of this paper was presented in part at the IEEE Globecom 2023 [1]. (Corresponding Author: Thinh Dinh)

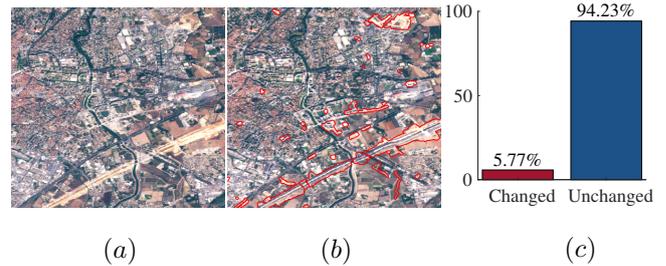


Fig. 1: An example: (a), (b) a pair of images in different times where changed areas are bounded with red; (c) a bar chart on the co-responding contribution of changed and unchanged pixels.

satellites, poses a limitation on the handling of daily generated EO data. While forwarding the complete dataset can be advantageous for accurately detecting any changes or anomalies, it is not regarded as efficient in terms of data storage and transmission capabilities. Therefore, the conventional approach of transmitting captured images to the ground for analysis and distribution may be inefficient. Thus, the vast amount of data generated by EO satellites requires novel techniques for high-spectral image processing and transmission.

Numerous forthcoming satellite missions aim to address memory and bandwidth limitations by implementing on-board processing operations that allow for data processing transfer from the ground segment to the space segment. The adoption of the new processing workflow is projected to significantly improve the efficiency of downlink data transmission, leading to a decrease in the required transmission resources. In order to improve bandwidth utilization, compression should be semantically related to change detection and the transmission of differentially encoded data. This holds the potential to diminish bandwidth requirements and energy consumption during communication but will increase energy expenditure associated with computational and compression processes. As illustrated in Fig. 1, a pair of images captured at different times from a single location has been randomly selected from the OSCD dataset [4]. Areas differing in Fig. 1(b) from the corresponding reference image shown in Fig. 1(a) are outlined in red. Fig. 1(c) shows the ratio of changed to unchanged areas in Fig. 1(b). By accurately detecting only the changed regions down to the pixel level and encoding them, we can save up to 94% of communication overhead. Our primary focus is the

identification of terrestrial changes. However, clouds present a significant source of variability in image data, covering over 50% of the Earth's surface, including approximately 55% of terrestrial landmasses and 72% of the oceanic expanse [5]. Consequently, any methodology devised for change detection necessitates the inclusion of a mechanism to discriminate against or eliminate clouds from the image. In this study, we develop a three-stage method specifically designed to reduce the volume of data transmitted by identifying and encoding only those non-cloud multi-spectral pixels that exhibit significant changes. This approach not only preserves the quality and relevance of the transmitted data but also significantly cuts down on the energy required for data processing and transmission.

A. Related Works

1) *Semantic communication*: Semantic information represents the meaning and veracity of source information [6], which has attracted extensive research interest very recently. Popular approaches for reducing the transmitted information payload while simultaneously preserving the quality of the original data source encompass strategies such as (1) optimizing compression parameters [7], (2) employing a machine learning-based technique referred to as joint source-channel coding (JSCC) [8], and (3) removing redundant or irrelevant data. In the first task, authors in [7] introduced a novel method by optimizing the compression ratio, segmentation, and distributed processing of EO images, employing an iterative optimization approach within the framework of satellite mobile edge computing. In the second, the JSCC framework has garnered significant attention due to its remarkable performance in operating effectively without reliance on precise channel state information [8]. Utilizing an end-to-end approach, JSCC is trained to attain high Peak signal-to-noise ratio (PSNR) values when transmitting a range of data types, including images [8], [9], audio [10], text [11], or other specific tasks [12]. Nonetheless, we emphasize that the aforementioned techniques tend to excessively compress data, whereas our objective lies on the last point, focusing on transmitting changed information to the terrestrial station. By scoring and encoding data from observed images that exhibit a high probability of change, our strategy seeks to save communication resources by directly reducing the volume of information that must be transmitted prior to the compression or encryption stages. This approach is anticipated to yield groundbreaking improvements in energy efficiency for data processing and transmission within the domain of EO.

2) *Change detection*: The primary aim of change detection in the field of remote sensing is to identify pixels exhibiting "semantic change" within multi-spectral sensing images obtained at distinct time points within a particular region [13]. Various factors can contribute to this change, including object deformation, relative motion, alterations in appearance, or object disappearance. Traditional approaches to change detection, including basic image difference measurement, Change Vector Analysis (CVA) [14], Principal Component Analysis (PCA) with K-means [15], and Gramm-Schmidt transformation [16]

have demonstrated effective results in certain straightforward scenarios. Enhanced variants have also been developed to achieve superior performance, such as Parcel CVA [17] and Robust CVA [18]. Additionally, the fuzzy local information clustering method with decomposition is applied for land-cover change detection [19], while the fuzzy C-means (FCM) algorithm is combined with a Bayesian network to detect and classify structural changes between two sensing images [20].

In recent years, the utilization of various techniques and components derived from neural networks has been increasingly employed in the task of scene segmentation for change detection. The goal is to extract more profound representations from the data. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), possess the ability to automatically learn image features during the training process, leading to notable advancements in performance. Many of these approaches utilize an encoder-decoder structure, with the U-Net architecture being a prominent example [21]–[24]. Typically, CNN-based change detection methods encompass two primary approaches: change detection based on post-classification, or direct detection at the patch-level or pixel-level [25]. The first approach involves training CNNs to classify remote sensing images captured during two distinct periods for deriving change information [26], [27]. The latter approach entails designing a CNN utilizing patch or pixel-level samples for network training, which necessitates a meticulously designed structure to attain good performance [28]–[31]. However, because of the complex nature of sensing images from LEO satellites, change detection approaches have not made real breakthroughs in very high-precision localization of changes to the pixel-level. That leads to the fact that it is highly plausible that information-carrying data would be lost if these methods were applied directly to satellite processing where only pixels determined to be changed are sent. As a result, conventional change detectors are only meaningful in assisting ground-based data analysis when we have all the data received from the satellite. Our methodology advances this by scoring each pixel and selectively transmitting those potentially to be changed, i.e., maintaining a 95% probability of selecting and encoding the changed pixels.

3) *Cloud removal*: The majority of processing tasks involving optical remote sensing imagery necessitate cloud detection at the pixel scale, making it a fundamental step in the analysis and preprocessing of such imagery. Researchers have dedicated their efforts to investigating and developing cloud detection techniques centered around spectral thresholds [32], [33]. These methods compute thresholds using diverse attributes like cloud reflectivity and brightness. Furthermore, the application of machine learning is explored for cloud detection [34], encompassing the extraction of features such as cloud texture, color, and geometric characteristics, followed by the design and training of classifiers. However, these approaches necessitate the manual configuration of appropriate thresholds or the creation of distinct features tailored to different instances, demanding specialized expertise and often exhibiting limited robustness. CNN is capable of autonomously learning suitable features, thereby eliminating the need for manual feature selection. Consequently, various

approaches employing CNN have been introduced for cloud detection, predominantly relying on fully convolutional networks (FCNs) [35]–[39]. The U-Net [40], a well-established encoder–decoder architecture, incorporates skip-connections to fuse corresponding features from both the encoding and decoding stages, thereby recovering spatial information. The effectiveness of U-Net in delineating clouds in Landsat 8 imagery, aided by automatically generated Ground Truths (GTs), has been firmly established [36]. Authors in [41] proposed a global–local fusion architecture based on convolutional neural networks CNNs to detect clouds. Besides, [42] utilized a feature pyramid network to reconstruct missing information. In [43], temporal-information-based methods were employed to use corresponding areas in ancillary images from different acquisition times to recover cloud-covered regions. These methods aim to restore parts obscured by clouds but often fail when unexpected events occur. In these works, the authors concentrated on enhancing or proposing models designed to augment the precision in discerning cloud pixels or generating image pixels to fulfill the missing parts. Here, instead, the focus is on the preservation of non-cloud meteorological parameters for subsequent procedural stages.

Previous research efforts have primarily concentrated on enhancing noise reduction techniques [32]–[35], particularly for cloud-related data, or refining the precision of ground-based change detection methods upon receiving comprehensive data from satellites. However, these existing solutions prove inadequate for complex end-to-end systems due to the lack of comprehensive consideration given to the vital transmission of critical information contained within noisy sensing images to the gateway. In addition, the factor of processing power for intelligent processing and communications has not been investigated.

B. Contributions

This paper analyzes the energy required to support all operations for an end-to-end system that utilizes deep learning-based architecture to process and transmit valuable information from noisy sensing images captured by satellites. Our contributions are listed as follows:

- This work extends and improves our preliminary end-to-end semantic communication framework [1] for satellite-based EO by including the energy consumption for the execution of the models to extract changed MPs at the satellites. This allows us to evaluate the total energy consumption for processing and transmission of satellite images to the ground stations.
- We present a 3-stage architecture for semantic communication of EO images. The architecture includes a learning-based transmitter that extracts changed MPs from the captured images, which drastically reduces the data to be transmitted by the satellites and, hence, the required communication resources, without performance degradation.
- We present two models that were developed based on the U-Net encoder-decoder architecture [40]: Cloud-SLR (Cloud removal model enhanced by Smooth Labeling

TABLE I: Key parameters defined in the paper.

Symbol	Description
Scenario	
\mathbf{X}^{t_0}	Reference image taken at time t_0
\mathbf{X}^{t_1}	Observed image taken at time t_1
H	Height of each image [pixels]
W	Width of each image [pixels]
D	Depth of each image [pixels]
\mathcal{X}^{t_i}	Set of available MPs in image $\mathbf{X}^{t_i}, i \in \{0, 1\}$
\mathcal{P}	Set of selected MPs for transmission
Communication	
f_c	Carrier frequency [Hz]
γ_t	SNR for downlink satellite-to-ground link at a time t
G_{tx}	Transmitter antenna gain
G_{rx}	Receiver antenna gain
d_t	Distance between that satellite and gateway at time t
P_{tx}	Transmission power at the satellite [W]
μ_{RF}^{amp}	Inefficiency of the power amplifier
P_{RF}^{sta}	Static power consumption by communication [W]
$R_{\mathcal{P}}$	Selected communication rate at the time t
Processing	
f_{CPU}	Processor frequency [Hz]
N^{MAC}	Number of MAC operation
ε	Number of consumes energy units per MAC operation [fJ]
ρ	Compression ratio
κ	Parameter that determines the complexity of the compression algorithm
$C(\rho_k, \kappa)$	Number of CPU cycles to compress one bit of data
$\mathcal{D}_{cloud} = \{(\mathbf{a}_i, z_i)\}$	Cloud learning set over input \mathbf{a}_i and output z_i for SLR-Cloud model
\varkappa	Degree of smoothing in training SLR-Cloud model
$\mathcal{D}_{change} = \{(\mathbf{b}_i, y_i)\}$	Change learning set over input \mathbf{b}_i and output y_i for Change-Net model

Regularization technique) and Change-Net. Cloud-SLR was designed for a highly accurate cloud detection through increased depth, width, and skip connections with varying receptive fields and by applying the label smoothing technique. Change-Net was designed to accurately detect changed multi-spectral pixels through a dual-branch structure to process pre- and post-change images, subsequently generating change maps with pixel scoring.

- The proposed framework is evaluated through numerical simulations that utilize available standard datasets (38-Cloud and OSCD datasets) and real-world data obtained from Sentinel-2 about the North Jutland Region, Denmark. The results clearly show that our approach reduces both energy consumption at the satellite and the amount of data for encoding and transmission while preserving critical information, leading to truly intelligent edge processing.

II. SYSTEM MODEL

In this section, we introduce the semantic encoding system architecture, which differs from existing literature primarily

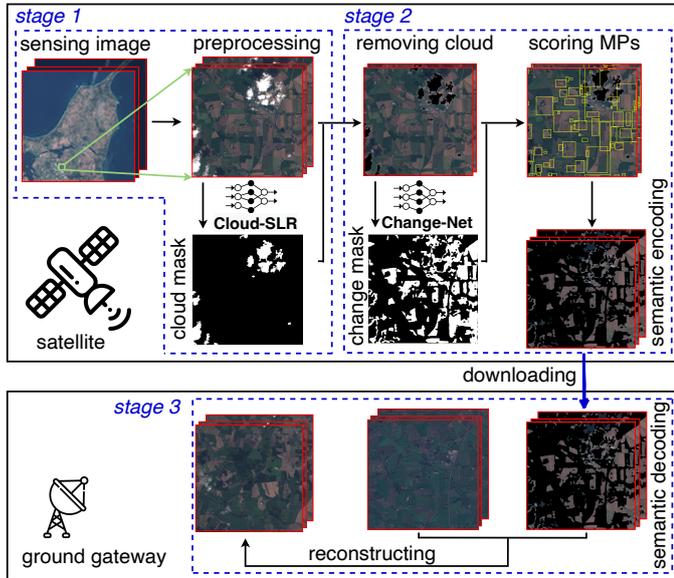


Fig. 2: Proposed end-to-end scoring architecture.

focusing on optimizing compression ratios [7] or establishing efficient JCSF frameworks for data transmission [8]. Our emphasis, in contrast, is on enhancing energy efficiency by optimizing the extraction and encoding of new sensing data from satellites before transmission. This is done while adhering to established communication protocols, which feature built-in mechanisms for error detection and packet retransmission, ensuring the accurate delivery of information. This strategic approach allows our system to remain adaptable for future integration of optimization techniques, including compression techniques or the incorporation of the JSCC framework. Next, motivated by the shortcomings of previous works in handling the energy constraints, we formulate the problem of minimizing energy consumption, both in processing and communication, under the constraint of a given data downloading performance.

A. Satellite-Based Image Transmission System

We consider a framework for satellite communication, depicted in Fig. 2 involving a LEO satellite engaging in communication with a ground-level gateway directly connected to a server. The satellite assumes responsibility for capturing multispectral images and is equipped with an AI module designed for image feature extraction. These features, such as potential pixel changes and their associated coordinates, must be transmitted within the limited time period when the satellite and gateway are connected. In this context, the proposed end-to-end framework comprises three sequential stages: (1) the acquired images undergo preprocessing and cloud removal using the Cloud-SLR model. This step is imperative, as its omission may result in the inadvertent misclassification of cloud pixels, a common occurrence in satellite imagery, as change pixels. (2) The resultant set of cloud-free image pairs is then subjected to the Change-Net model, alongside a scoring mechanism, which serves the purpose of identifying semantic

map including scheduled MPs. These selected MPs are subsequently encoded and transmitted to a designated gateway. (3) The ultimate reconstruction of multi-spectral images is accomplished by leveraging reference images in conjunction with received changed data.

Consider two coregistered multi-spectral satellite images taken at different time instances

$$\mathbf{X}^{t_0} = \{x^{t_0}(i, j, k) | 1 \leq i \leq H, 1 \leq j \leq W, 1 \leq k \leq D\}, (1)$$

$$\mathbf{X}^{t_1} = \{x^{t_1}(i, j, k) | 1 \leq i \leq H, 1 \leq j \leq W, 1 \leq k \leq D\} (2)$$

of size $H \times W \times D$, where H , W , and D are the size of height, width, and number of spectral bands, respectively. \mathbf{X}^{t_0} is the reference image taken at the reference time t_0 and \mathbf{X}^{t_1} is the newly observed image taken at time t_1 .

Definition 1. Pixel change is defined as the temporal variation occurring of coregistered images at a specific location, including changes in land use, urban coverage, deforestation, and other similar types of deviation [44], [45].

Our objective is to communicate the changed multi-spectral pixels (MPs) from the sensing image \mathbf{X}^{t_1} captured by the satellite to Earth while minimizing the data transmitted to ground and the energy consumption of the entire process. Specifically, given $\mathcal{X}^{t_1} = \{\mathbf{x}_{ij}^{t_1}\}$ as a set of available MPs in image \mathbf{X}^{t_1} , e.g., $\mathbf{x}_{ij}^{t_1}$ could be represented by the MP (i, j) including all spectral bands, we need to select a subset of MPs $\mathcal{P} \in \mathcal{X}^{t_1}$. The reference pixel set $\mathcal{X}^{t_0} = \{\mathbf{x}_{ij}^{t_0}\}$ of the reference image \mathbf{X}^{t_0} is processed similarly. We define $\mathcal{S} = \{s_{ij}\}$ as the accurate change map where $s_{ij} \in \{0, 1\}$ represents the changed/unchanged MP. In particular, $s_{ij} = 1$ if there is a pixel change at location (i, j) , and vice versa.

Remark 1. The considered change scenarios bring superior improvements in information encoding transmitting and storing. However, the complexity will be problematic when we simply classify change MPs and semantic data. In fact, the sensing images typically include noise (i.e., clouds [5]), then noise MPs will be classified as changed, leading to unnecessary propagation down the gateway and waste of valuable communication resources.

Our work addresses this raising issue defined in Remark 1 by using high-performance Cloud-SLR to remove cloud MPs before detecting change MPs, which is performed by the Change-Net module. Our focus is on analyzing the energy efficiency of our architecture compared to conventional schemes, rather than on the training process of CNNs.

B. Energy consumption

We evaluate the energy consumption by calculating the combined energy consumption associated with image processing, compression, and communication. We denote $\alpha_{ij} \in \{0, 1\}$ is a binary decision variable with $\alpha \triangleq \{\alpha_{ij} | 1 \leq i \leq H, 1 \leq j \leq W\}$, where $\alpha_{ij} = 1$ if we select the MP $\mathbf{x}_{ij}^{t_1}$ as a candidate of \mathcal{P} . As a consequence, the transmission set \mathcal{P} could be constructed as

$$\mathcal{P} \triangleq \{\mathbf{x}_{ij}^{t_1} | \alpha_{ij} = 1\}, \quad \forall \alpha_{ij}, \mathbf{x}_{ij}^{t_1}. (3)$$

Then, the total amount of data $C(\alpha)$ to transmit the set \mathcal{P} is computed by

$$C(\alpha) = \sum_{\alpha_{ij} \in \alpha} D\alpha_{ij}. \quad (4)$$

The total energy for processing and transmitting the selected data \mathcal{P} is derived as

$$E^{\text{tot}}(\alpha) = E^{\text{proc}} + E^{\text{comp}}(\alpha) + E^{\text{trans}}(\alpha), \quad (5)$$

where $E^{\text{proc}}(\alpha)$ is the required energy for processing raw images and selecting the transmitted set \mathcal{P} , $E^{\text{comp}}(\alpha)$ stands for compress energy and $E^{\text{trans}}(\alpha)$ indicates the consumed communication energy.

1) *Processing energy*: The energy consumption $E^{\text{proc}}(\alpha)$ for image processing is mainly consumed by well-trained CNN-based models, formulated by [46]

$$E^{\text{proc}} = E^{\text{MAC}} + E^{\text{mov}}, \quad (6)$$

where E^{MAC} and E^{mov} are energy for CPU computation for Multiply-And-Accumulate (MAC) operations and memory access for data movement between CPU and memory, respectively. Herein, we focus on the convolution layers for energy consumption because these layers dominate the overall computation and energy consumption [46]. Considering one convolution layer with the input feature map, assume there are M filters with height, width, and number of channels are R , S , and D , respectively. As a result, we have M output feature map with size as $E \times F$. Considering the stride with size $\bar{S} \times \bar{S}$, the number of MAC operations in this convolution layer is $N^{\text{MAC}} = RSDEFM/\bar{S}^2$. Then, given each MAC operation consumes energy of ε units [47], the total energy consumed during MAC operations is calculated as

$$E^{\text{MAC}} = \varepsilon N^{\text{MAC}} = \varepsilon RSDEFM/\bar{S}^2. \quad (7)$$

The data movement energy E^{mov} is consumed when CPU reads input image pixels, filter weights, partial sums, and then writes updated partial sums to the cache or memory. In the context of computational energy estimation, the initial step involves the computation of the quantity of MACS operations, which is subsequently adjusted by the energy expenditure associated with the hardware execution of a MAC. Consequently, the total computational energy outlay is contingent solely upon the number of MACs [48].

2) *Compression energy*: The model to calculate the satellite's energy spending for processing tasks captures the most relevant CPU parameters [49]. The model establishes a direct relationship between the energy consumed per clock cycle and the square of the CPU clock frequency f_{CPU} multiplied by the effective capacitance coefficient, which is specific to the processor under consideration [49]. Thereby, we have

$$E_{\text{cycle}}^{\text{proc}} = \frac{P^{\text{proc}}(f_{\text{CPU}})}{f_{\text{CPU}}}, \quad (8)$$

where $P^{\text{proc}}(f_{\text{CPU}})$ is the power consumption during processing at the maximum CPU frequency. Noting that the supplied power is linear with the number of processor cores N_{CPU} , the

energy consumption to process data $E^{\text{proc}}(\alpha)$ is modeled as

$$E^{\text{comp}}(\alpha) = C(\alpha) \hat{R}^{\text{comp}} E_{\text{cycle}}^{\text{proc}}, \quad (9)$$

where

$$\hat{R}^{\text{comp}} = e^{\kappa\rho} - e^{\kappa} \quad (10)$$

represents the compressing complexity, which is defined as the number of CPU cycles to compress one bit of data by a compression ratio ρ and the positive constant κ depending on used compression algorithm [50].

3) *Communication energy*: We consider an interference-free communication channel affected by additive-white Gaussian noise (AWGN) with zero-mean and variance ϖ^2 [7]. Consequently, the signal-to-noise ratio (SNR) for the downlink satellite-to-ground link at a particular time t , represented by γ_t , is calculated as

$$\gamma_t = G_{\text{tx}} G_{\text{rx}} P_{\text{tx}} \left(\frac{c}{4\pi d_t f_c \varpi} \right)^2, \quad (11)$$

where G_{tx} and G_{rx} stand for the transmitter and receiver antenna gains. P_{tx} is the transmission power at the satellite, f_c is the carrier frequency, and $c = 2.998 \times 10^8$ is the speed of light. The distance between the satellite and gateway d_t can be calculated due to satellite's predefined connection time to the gateway, orbit time, and satellite's altitude. Once we know the SNR, a proper modulation and coding scheme is selected to achieve reliable communication, i.e., following the DVB-S2 system [51]. We denote a available throughput \bar{R} [bps] and $\gamma_{\min}(\bar{R})$ as the minimum SNR to achieve a reliable rate \bar{R} to achieve a block error rate $< 10^{-5}$ with rate \bar{R} . Then the set of ordered pairs is defined as $\mathcal{Q}_{\text{DVB-S2}} = \{\bar{R}, \gamma_{\min}(\bar{R})\}$. The selection of transmission rate for the selected set \mathcal{P} is determined by adapting the modulation and coding scheme in order to attain the maximum data rate for reliable communication. Let γ_t be minimum SNR experienced at the gateway at time t , when the transmission of the data is initiated. At this time instant, the rate is selected as

$$R_{\mathcal{P}} = \max \{ \bar{R} \in \mathcal{Q}_{\text{DVB-S2}} : \gamma_t \geq \gamma_{\min}(\bar{R}) \}. \quad (12)$$

The energy consumption associated with the downlink transmission of data encompasses the data's size \mathcal{P} , the inefficiency of power amplifiers, and the fixed power consumption originating from communication modules [52]. As a result, the power expenditure attributed to the downlink phase in the communication process, denoted as P_{RF} , can be computed as

$$P_{\text{RF}} = \mu_{\text{RF}}^{\text{amp}} P_{\text{tx}} + P_{\text{RF}}^{\text{sta}}, \quad (13)$$

where $\mu_{\text{RF}}^{\text{amp}}$ represents the inefficiency of the power amplifier, $P_{\text{RF}}^{\text{sta}}$ is the static power consumption by communication. Under the assumption that the RF link of the recipient satellite consistently consumes a fixed power of $P_{\text{RF}}^{\text{sta}}$ without variation, this aspect is exempt from the energy consumption assessment of the considered framework [7]. Consequently, the downlink energy consumption for download data \mathcal{P} from the satellite to the gateway is calculated by

$$E^{\text{trans}}(\alpha) = \frac{P_{\text{tx}} C(\alpha)}{R_{\mathcal{P}}}. \quad (14)$$

C. Energy Efficient Data Downloading With Change-Detection Constraint

In contrast to conventional approaches aimed at improving the efficiency of the change detection algorithm that is applied to the sensed images at the satellite [53], our focus is centered on the reduction of the energy required to transmit all change to the nearby gateway, which shifts the focus towards the communication aspects of the system. That is, all x_{ij} that have change map $s_{ij} = 1$ will be scheduled to the transmission set \mathcal{P} . This is achieved by setting the values of the decision variable α as

$$\alpha \triangleq \{\alpha_{ij} | \alpha_{ij} \geq s_{ij}\}, \forall \alpha_{ij}, s_{ij}, \quad (15)$$

which ensures that all the changed MPs are transmitted. Accordingly, the optimization problem is to minimize the communication energy subject to change detection constraints, which is formulated as

$$P1 : \min_{\alpha} E^{\text{tot}}(\alpha) \quad (16a)$$

$$\text{s.t. } \alpha_{ij} \geq s_{ij}. \quad (16b)$$

Proposition 1. *The problem P1 achieves optimality when the constraints (16b) satisfy $\alpha_{ij} = s_{ij}, \forall i = \{1, 2, \dots, H\}, j = \{1, 2, \dots, W\}$. This is straightforward to prove due to the fact that the energy decreases monotonously as the number of transmitted pixels, identified by the number of non-zero entries in α , decreases.*

III. SEMANTIC ENCODING SYSTEM

This section introduces a practical solution for MP scheduling by proposing a system designed to identify most changed MPs in real-world observation images. To accomplish this, problem (1) is reformulated in a practical manner for resolution. We then outline a proposed architecture for detecting changed MPs in EO applications. This architecture includes a preprocessing step using Cloud-SLR to eliminate cloud cover, followed by the deployment of a deep learning model, Change-Net, which aims to detect as changed MPs in captured images before they are transmitted to the gateway.

A. Approximate Formulation of Data Downloading with Change Detectors

We underscore that the primary difficulty in addressing problem (16) resides in managing the binary constraint (16b). Under realistic conditions, where the satellite cannot precisely determine the indices of all the changed MPs in captured sensing images, fully adhering to constraint (16b) proves infeasible. Instead, it is necessary to implement an AI model at the satellite to estimate s_{ij} and, hence, perform inference on the changed MPs, which in turn requires a reformulation of P1. In this paper, we would like to build a practical predictor that can recognize most of the change MPs. We define $s_{ij}^p \in \{0, 1\}$ as the predictor for s_{ij} given by the selected change detection algorithm, which achieves an error

level ϵ . Then, P1 is reformulated as

$$P2 : \min_{\{\alpha_{ij}\}} \sum_{i=1}^H \sum_{j=1}^W \alpha_{ij} \quad (17a)$$

$$\text{s.t. } \alpha_{ij} \geq s_{ij}^p, \quad (17b)$$

$$\Pr(s_{ij}^p = 0 | s_{ij} = 1) \leq \epsilon, \quad (17c)$$

where $\epsilon \in [0, 1]$ is the threshold that ensures a sufficient number of change MPs are selected. It is evident that when the value of epsilon is decreased, the criteria for identifying change MPs become more rigorous, leading to the selection of additional pixels for transmission. As an illustration, when setting $\epsilon = 0.05$, the problem P2 is devised to guarantee the selection and labeling of no less than 95% of change MPs. Accordingly, we propose an algorithm to obtain an efficient suboptimal solution to the problem (17).

In the following, we describe the steps to detect change MPs in EO applications, which include preprocessing to remove cloud cover and the use of a deep learning model to determine the likelihood that each specific pixel is changed from reference versions of the image. The aim is to ensure that the reconstructed images at the gateway are as precise as possible, thereby facilitating their accurate interpretation by human/domain experts in EO.

B. Preprocessing and Cloud Removal

1) *Preprocessing:* As part of the preprocessing stage, the following steps are undertaken: (1) the selection of bands representative of the complete set of MP images; (2) the adjustment of illumination levels to facilitate the detection (cloud and change) in the subsequent stage; and (3) the elimination of cloud cover from the observed image. First, N ($N < D$) useful bands are selected for processing based on the principle that a group of appropriate bands (e.g., visible and near-infrared bands) could reduce the processing requirements for both training and execution phases. This is consistent with realizing cloud detection on the satellite with limited hardware and processing capabilities. Before removing clouds, a radiometric correction procedure is implemented to rectify radiometric disparities among multi-spectral images stemming from diverse imaging conditions such as sun angle, light intensity, and atmospheric circumstances. The chosen approach is relative radiometric normalization, which is founded on the z-score method, which normalizes images to zero mean and a unit standard deviation. Given a MP pair, denoted by $\mathbf{x}_{ij}^{t_0} = [x_1^{t_0}, x_2^{t_0}, \dots, x_N^{t_0}]$ and $\mathbf{x}_{ij}^{t_1} = [x_1^{t_1}, x_2^{t_1}, \dots, x_N^{t_1}]$. The radiometric relative normalization can be expressed as [54]

$$\hat{x}_n^{t_0} = \frac{x_n^{t_0} - \mu_{\mathbf{x}_n^{t_0}}}{\sigma_{\mathbf{x}_n^{t_0}}^2}; \quad \hat{x}_n^{t_1} = \frac{x_n^{t_1} - \mu_{\mathbf{x}_n^{t_1}}}{\sigma_{\mathbf{x}_n^{t_1}}^2}, \quad (18)$$

where $\mu_{\mathbf{x}_n^{t_k}}$ is the mean and $\sigma_{\mathbf{x}_n^{t_k}}^2$ is the variance for band n of image \mathbf{X}^{t_k} , ($k \in \{0, 1\}$). The radiometric correction would suppress the radiometric difference between multi-spectral images caused by different conditions.

2) *Cloud removal with Cloud-SLR:* In this work, the Cloud-SLR model for identifying cloud regions in multi-spectral Sentinel-2 images based on U-Net using a combination of

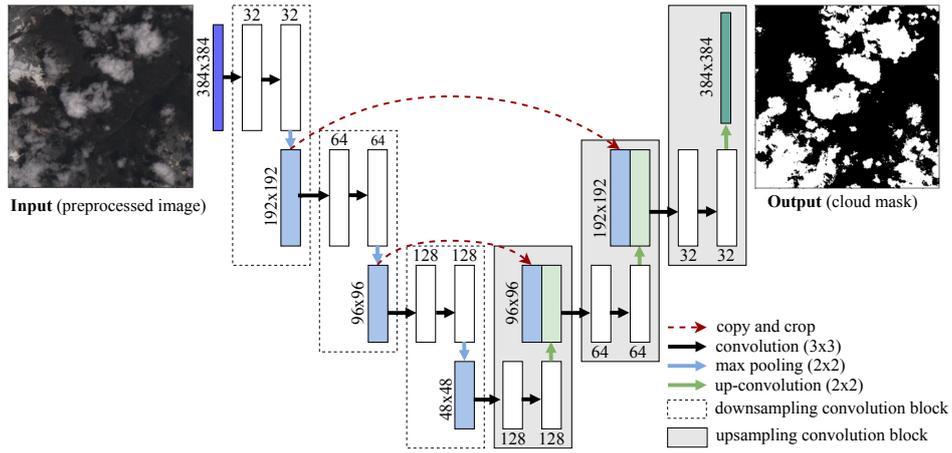


Fig. 3: Cloud-SLR for cloud detection based on U-net architecture.

thresholding and deep learning techniques is deployed. This approach utilizes four spectral bands - Red, Green, Blue, and Near Infrared (RGBNir) - for both training and prediction. The detailed design of the Cloud-SLR architecture, which encompasses downsampling and upsampling convolutional blocks along with their corresponding structures, is visually depicted in Fig. 3. As each spectral band in Sentinel-2 encompasses a substantial number of pixels (approximately 10980×10980), it is necessary to segment them into smaller image patches using a cropping technique. This is achieved by dividing the selected spectral band image into $384 \times 384 \times 4$ non-overlapping patches, with each of the four patches corresponding to the RGBNir bands being combined to create a 4D input. To enhance the resilience towards patterns of clouds or similar, the input patches underwent geometric transformations, such as horizontal flipping, rotation, and zooming. The output probability map is obtained via the use of a sigmoid activation function in the final convolution layer of the network. The cloud learning set consists of tuples

$$\mathcal{D}_{\text{cloud}} = \{(\mathbf{a}_i, z_i) \in (\mathcal{A} \times \mathcal{Z}) | i = 1, \dots, M_{\text{cloud}}\}, \quad (19)$$

obtained from an unknown joint distribution $P_{\mathcal{D}_{\text{cloud}}}$ over $\mathcal{A} \times \mathcal{Z}$, where M_{cloud} is the total number of instances, and z_i represents the corresponding label of the input \mathbf{a}_i . Our aim is to estimate a function $g(\mathbf{a}_i | \psi)$ that maps inputs \mathbf{a}_i to outputs z_i , where ψ is a set of parameters that are optimized using the training set. To estimate ψ , one can utilize the Adam gradient descent [55] method to implement the Cross-Entropy (CE) loss function as

$$\mathcal{L}(\hat{z}_{\psi,i}, z_i) = -(z_i \log(\hat{z}_{\psi,i}) + (1 - z_i) \log(1 - \hat{z}_{\psi,i})), \quad (20)$$

with $\hat{z}_{\psi,i} = g(\mathbf{a}_i | \psi)$ is the output of the model.

Remark 2. In order to enhance the training efficiency of Cloud-SLR, we utilize the CE loss function empowered by the label smoothing regularization technique [56]. Specifically, let \bar{z}_i as a label introduced to facilitate label smoothing. The smoothed label z_i^{SL} is formulated as

$$z_i^{\text{SL}} = (1 - \varkappa)z_i + \varkappa\bar{z}_i, \quad (21)$$

where the parameter $\varkappa \in (0, 1)$ denotes the degree of smoothing. The label \bar{z}_i is sampled from the $P_{\bar{z}}$ distribution.

Algorithm 1 Proposed algorithm for problem P2

Input: Geometrical multi-spectral image pair \mathbf{X}^{t_0} and \mathbf{X}^{t_1}

Output: The transmission set \mathcal{P} , Reconstructed MP image $\hat{\mathbf{X}}^{t_1}$, and Energy requirement for processing and transmission

Preprocessing and Cloud Removal

- 1: Select process bands (R, G, B, Nir) for cloud removal and change detection
- 2: Normalize \mathbf{X}^{t_0} and \mathbf{X}^{t_1}

- 3: Cloud detection and removal using Cloud-SLR

Change Scoring and Semantic Encoding

- 4: Scoring MPs using Change-Net and update $\{s_{ij}^p\}$ as in (25)
 - 5: Perform threshold segmentation to get the transmission set $\mathcal{P} = \{\mathbf{x}_{ij}^{t_1} | s_{ij}^p = 1, i = 1, 2, \dots, H, j = 1, 2, \dots, W\}$
 - 6: Update $\alpha_{ij} = s_{ij}^p \forall i, j$ and calculate required energy $E^{\text{tot}}(\alpha)$ as in (5)
-

Herein, we adopt a uniform distribution $\bar{z}_i = 1/2$ for all labels. Consequently, the CE loss (20) is replaced by

$$\mathcal{L}(\hat{z}_{\psi,i}, z_i^{\text{SL}}) = -(z_i^{\text{SL}} \log(\hat{z}_{\psi,i}) + (1 - z_i^{\text{SL}}) \log(1 - \hat{z}_{\psi,i})). \quad (22)$$

The concept of label smoothing entails substituting the conventional one-hot vector labels with smoothed labels, which are the result of averaging between the distinct target labels and a uniformly distributed array of alternative labels. This technique has demonstrated significant efficacy in addressing the challenge of overfitting without requiring interventions within the latent activations or parameters of the neural network induced by other methodologies, i.e., such as the application of penalties like ℓ_1 -norm or ℓ_2 -norm on model parameters, the utilization of Dropout to stochastically nullify neuron outputs, batch normalization, and data augmentation techniques. Through the introduction of noise into the target labels during the training process, the label smoothing technique prompts the model to exhibit reduced levels of confidence in

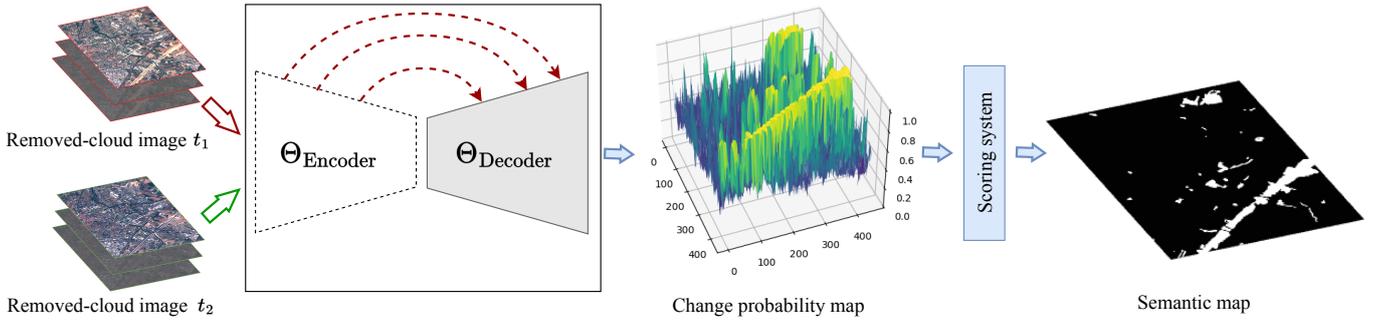


Fig. 4: Change-Net for semantic encoding (Θ indicates the network branch's learnable parameters).

its predictions, thus resulting in a more equitable dispersion of the probability mass across all classes. The employment of the label smoothing methodology in Cloud-SLR training yields remarkable outcomes, as extensively showcased and examined in the section IV. The well-trained weights of the Cloud-SLR model are then applied to the target image to generate a predicted cloud probability map. This map is binarized using a global threshold of γ and multiplied with the image \mathbf{X}^{t_1} to produce the cloud-removed image $\tilde{\mathbf{X}}^{t_1}$ with $\tilde{\mathcal{X}}^{t_1} = \{\tilde{\mathbf{x}}_{ij}^{t_1}\}$ as a set of MPs in $\tilde{\mathbf{X}}^{t_1}$.

C. Change Scoring and Semantic Encoding

In this stage, we have two multi-spectral images that have been geometrically aligned, have consistent lighting conditions, and have had cloud cover removed. Our goal is to propose a computationally efficient automatic change detection method for these two images that are practical for satellite applications. Motivated by [40], [57], the Change-Net based on U-Net architecture is developed to perform semantic segmentation of images. We use a negative log-likelihood loss function to classify every MP in the observed images as either changed or not. Specially, we are considering a change learning set consisting of tuples $\mathcal{D}_{\text{change}}$ as

$$\mathcal{D}_{\text{change}} = \{(\mathbf{b}_i, y_i) \in (\mathcal{B} \times \mathcal{Y}) | i = 1, \dots, M_{\text{change}}\} \quad (23)$$

with inputs \mathbf{b}_i , and labels y_i . With θ as a set of learnable parameters to be optimized, the likelihood $p(\mathbf{b}_i, \theta)$ is defined as the joint density of the observed data, which can be viewed as a function of θ that maps any given input \mathbf{b}_i to outputs y_i . By denoting m as the mini-batch size, the parameter set θ_{model} is then achieved through training process as

$$\theta_{\text{model}} = \arg \max_{\theta} \prod_{i=1}^m p(y_i | \mathbf{b}_i, \theta). \quad (24)$$

In our case, the system output $\bar{s}_{ij} \in [0, 1]$ is defined as the change score of the corresponding cloud-removed MP $\tilde{\mathbf{x}}_{ij}^{t_1}$, which is the probability of $p(s_{ij} = 1 | \tilde{\mathbf{x}}_{ij}^{t_1}, \theta_{\text{model}}) = \xi(\theta_{\text{model}}^T \tilde{\mathbf{x}}_{ij}^{t_1})$ with $\xi(\cdot)$ is the Log Softmax function. As \bar{s}_{ij} is close to 1, it is likely that there is a change in location (i, j) . Thus, to make a good prediction, we would like to learn a scoring system such that

$$s_{ij}^p = \mathbf{1}(\bar{s}_{ij} \geq \tau), \quad (25)$$

where τ is the predefined threshold, derived experimentally to satisfy the constraint (17c). The configuration of our change

scoring and semantic encoding framework is depicted in Fig. 4. Herein, the FC-EF-Res structure [57] is employed in the encoder-decoder block for training over a dataset. In particular, a 4-band image pair (RGBNir bands) rendered cloud-free is the outcome of the Cloud-SLR process, serving as input for Change-Net. In distinction from typical classification models that endeavor to categorize MPs within an image as either changed or unchanged, Change-Net generates a change probability map, subsequently subjected to a scoring phase, culminating in the generation of a semantic map comprising compacted and transmitted MPs. This procedural stage is devised to address the inefficiencies inherent in prevailing change detection methodologies. Further analysis of this process is expounded upon in Section IV.

It is important to note that the computation of energy consumption in the Cloud-SLR and Change-Net frameworks is executed via (6), wherein the determination of MACs is contingent upon the dimensions of the input image and the intricacy inherent in both architectural models. Specifically, enabling, for processing an input, Cloud-SLR requires 6.99×10^9 MACs, as the architecture depicted in Figure 3, while Change-Net, utilizes an encoder-decoder architecture similar to [57], consumes 3.66×10^9 MACs. For the adaptability of our algorithm to scenarios, this framework is specifically designed for EO satellites. This three-stage process ensures that our framework can handle a variety of environmental conditions and image qualities, i.e., those from Sentinel-2 systems. Our evaluations on diverse datasets and real-world Sentinel-2 images, demonstrate the robustness and adaptability of our framework, which is detailed analysis in the next section. Algorithm 1 captures steps for solving problem P2.

IV. NUMERICAL RESULTS

A. Implementation Details

We evaluate our proposed algorithm in the Sentinel-2 system when the transmission process initiates upon the satellite's entry into the gateway's coverage area. This is a worst-case scenario for the energy consumption for communication. Other important parameters for performance analysis are included in Table II. We use the 38-Cloud dataset for training and testing Cloud-SLR, which covers 38 Landsat-8 scene images and their manually extracted pixel-level ground truths for cloud detection. The 38-Cloud dataset, originally introduced in [58], comprises 8400 non-overlapping (NOL) patches, each measuring 384 x 384 pixels. These patches are extracted from

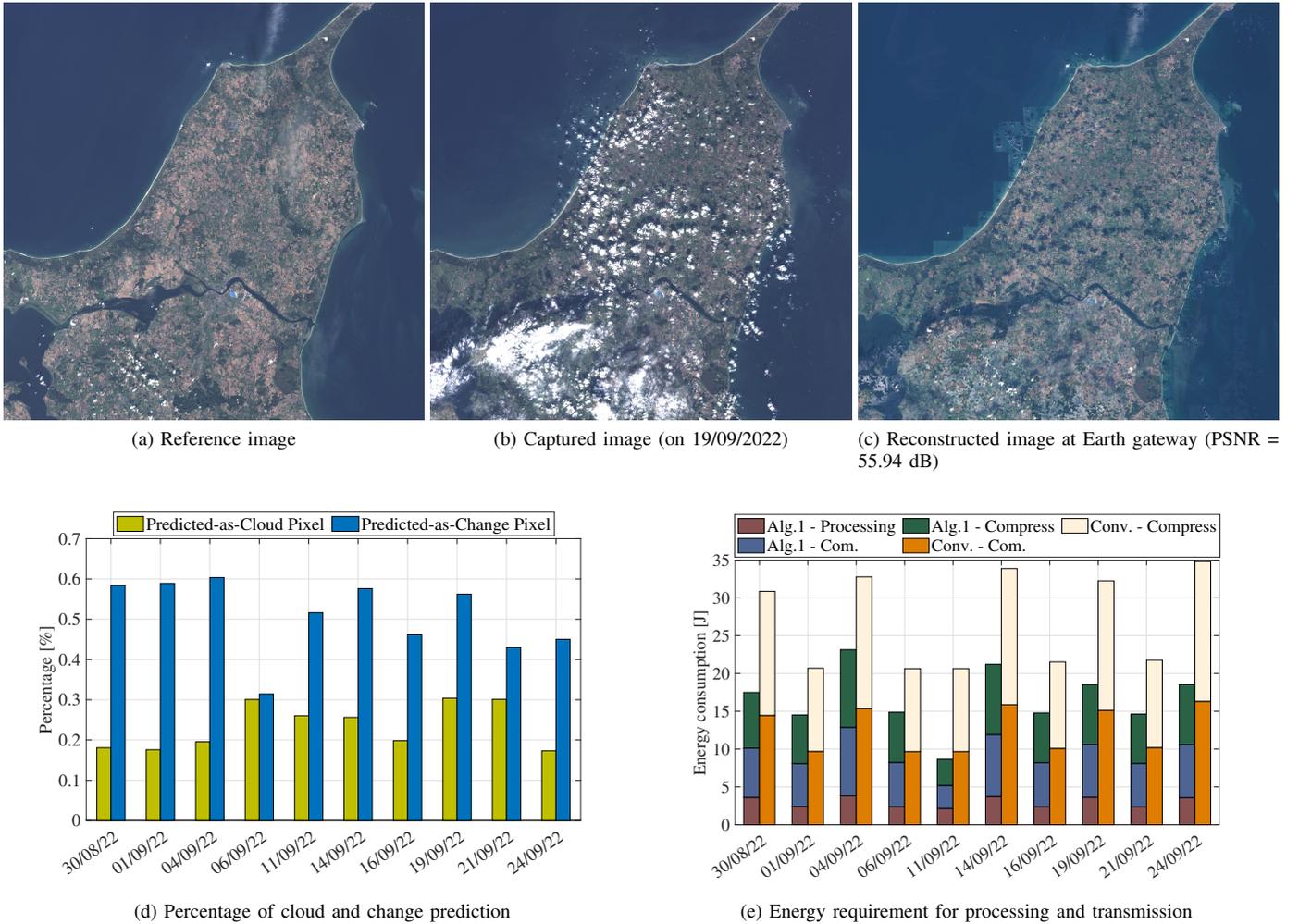


Fig. 5: The performance of the proposed algorithm on the real Sentinel-2 data (the multi-spectral images were acquired in the North Jutland Region of Denmark during the period from 23/08/2022 to 24/09/2022), with the reference image captured on 23/08/2022 (figures (a), (b), and (c) are display using RGB channels).

18 Landsat 8 Collection 1 Level-1 scenes, serving as the training dataset. Additionally, the test dataset is constituted by 9201 patches of equivalent spatial dimensions, sourced from 20 Landsat 8 scenes. The scenes primarily originate from North American locations, with their corresponding ground truths (GTs) manually derived. For Change-Net model, we conduct the Onera Satellite Change Detection (OSCD) dataset, publicly available on the IEEE-DataPort repository [4], widely used in the literature to detect changes in multi-spectral images of Sentinel-2 satellite [59]. The use of Sentinel-2 datasets, which provide multispectral imagery, helps us mitigate cold start issues with trained models during practical deployment. Although acquiring high-quality annotated data is complex and crucial for improving model performance, this paper primarily focuses on optimizing the use of existing data systems to enhance energy efficiency in satellite communications. Enhancing annotated datasets further is left for future works. The weights were assigned inversely proportional to the number of examples in each class in order to address class imbalances between the two categories (change versus no change). The features are preprocessed classically to facilitate training. The dataset was split into a shuffled and stratified training set and

a test set, comprising 60% and 40% of the data, respectively. Both Cloud-SLR and Change-Net are trained on the ground and applied to satellites for further processing of sensing data. Following training and evaluation on the dataset with annotated ground truth, we apply the trained models to real-world data collected from a Sentinel-2 image to identify changes at the pixel level relative to reference images. These identified semantic map \mathcal{S} are subsequently labeled and encoded prior to being transmitted to the gateway.

We conduct our experiments on a 28-core Intel(R) Xeon(R) Gold @ 2.8 GHz server (52 virtual cores) with 256 GB memory and a Nvidia Tesla V100 GPU. Our implementation for the proposed models and algorithm is Pytorch-based.

B. Overall energy efficiency performance

In order to evaluate the effectiveness of our architectural framework using real-world data, we integrated data sourced from the Sentinel-2 system, encompassing considerations of data reduction volume as well as an examination of the computational energy requisites for processing and transmitting the data, operating under the performance from both the Cloud-SLR and Change-Net components. Furthermore, in our

TABLE II: Simulation parameters.

Parameter	Symbol	Value
Carrier frequency	f_c	20 GHz
Altitude of satellites [3]	h	786 km
Processor frequency	f_{CPU}	1.8 GHz
Compression factor	ρ	5
Power consumption for processing	$P^{proc}(f_{CPU})$	10 W
Complexity of the image processing algorithm	κ	0.1
Inefficiency of the downlink		
RF power amplifier	μ_{RF}	1
Bandwidth	B	500 MHz
Noise power	ϖ^2	-115 dB
Satellite antenna gain	G_{tx}	32.13 dBi
Gateway antenna gain	G_{rx}	34.2 dBi
Performance in change detection	ϵ	0.05
Transmission power	P_{tx}	10 W
Communications duration	T^{pass}	15 min
Orbital period		100 min
Initial learning rate of Cloud-SLR	η^{cloud}	10^{-4}
Degree of smoothing	\varkappa	0.1
Initial learning rate of Change-Net	η^{change}	10^{-4}
Energy cost of each MAC operation [47]	ε	51 fJ

assessment of the collective energy consumption by the two networks, we postulated that the energy attributed to E^{MAC} , constitutes approximately 10% of the total processing energy, E^{proc} , a proportion akin to the one observed in the GoogleNet architecture [60]. We opted to employ acquired Sentinel-2 data not only to assess the proficiency of our proposed framework but also due to the unavailability of a dataset encompassing both cloud patterns and contemporaneous alterations. This limitation is one of the underlying factors hindering the direct practical applicability of machine learning applications for advanced intelligent satellite systems. The specific utility of images is exemplified through Fig. 5(a)-(c). The outcomes reveal that our proposed architecture confers advantages, particularly when the Peak Signal-to-Noise ratio (PSNR) between the reconstructed image (Fig. 5(c)) and the original captured image (Fig. 5(b)) attains a value of 55.94 dB, post cloud removal and absent compression. To achieve this, Change-Net strategically selected up to 60% of MPs (as demonstrated in Fig. 5(d) on the date 04/09/2023) in order to ensure a satisfactory selection of changed MPs.

Figure 5(d) presents the count of MPs identified as changed and those associated with cloud cover. We can see that our schematic works effectively through its capability to achieve efficient data reduction by up to 55% for transmission to Earth (as of September 24, 2022). As seen in the pair of illustrations in Fig. 5(a)-(b), subsequent to cloud area removal, a substantial proportion of terrestrial regions are earmarked and dispatched for monitoring (constituting approximately 40%-60% of the total), while marine regions (approximating 20%) are excluded unless experiencing noteworthy alterations. It becomes apparent that in contrast to the conventional OSCD

dataset comprised of cleared pairs resulting in a few number of changed MPs are selected, Changed-Net excels in the selection of the almost land areas for forwarding. This observation underscores the importance of employing advanced image pre-processing methods, particularly in real-world datasets affected by noise and cloud interference. Such techniques enable accurate and targeted localization of changed regions, facilitating more streamlined data transmission, which is left for our future works. Fig. 5(e) shows the energy requisites for the processing (both in Cloud-SLR and Change-Net), compression, and transmission of satellite-derived data. We note that the required energy for conventional algorithm (transmit all capture data) changes over time despite the same area because captured data from Sentinel-2 does not always capture the entire region and output the same amount of data. In a comparative analysis to the conventional paradigm of transmitting the entire sensing image, our architectural framework results in a minimum conservation of 28% in energy requirements, which saves up to $2.39\times$ the energy consumption compared to conventional one. This conservation translates directly into not only diminished energy consumption for data processing and satellite transmission, but also economical storage capacity and processing duration at terrestrial gateways.

In pursuit of the attainment of such communication efficiency, we have engineered the Cloud-SLR and Change-Net components to ensure their substantial reliability and efficiency. A detailed evaluation of their performance and capabilities is delineated in the following.

C. Cloud-SLR Performance

It is imperative to underscore that our principal objective centers on proposing an approach with the capacity to detect clouds with high performance while retaining a significant portion of non-cloud MPs. Consequently, we place special emphasis on three evaluation metrics, Specificity, Precision, and Overall Accuracy which are computed as

$$\begin{aligned} \text{Specificity} &= \frac{TN}{TN + FP}, & \text{Precision} &= \frac{TP}{TP + FP}, \\ \text{Accuracy} &= \frac{TP}{TP + FN}, \end{aligned} \quad (26)$$

where TP, FN, TN, FP are respectively True Positive, False Negative, True Negative, and False Positive. As seen in (26), the metric that we focus on most is specificity as it reflects the amount of non-cloud information retained after using Cloud-SLR. A higher specificity value is evident in its direct correlation to the quantity of non-cloud information preserved, which corresponds to valuable information.

We examine the convergence behavior of Cloud-SLR, as depicted in Fig. 6. Notably, our proposed model demonstrates fast convergence coupled with robust stability. In addition, Table III presents a comparative analysis between our Cloud-SLR and other contemporary approaches, conducted on the 38-Cloud dataset. Building upon the aforementioned discourse, by employing the label smoothing technique in tandem with an appropriately set classification threshold, Cloud-SLR effectively preserves 99.56% of non-cloud MPs-exhibiting superiority over existing methodologies. Moreover, our model ensures

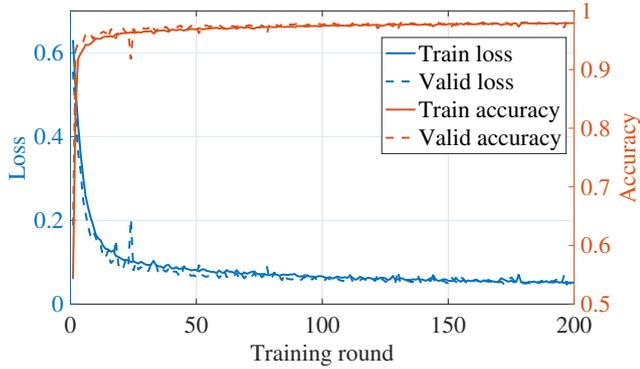


Fig. 6: Evolution of loss and accuracy versus training round.

TABLE III: Numerical performance over 38-Cloud dataset (%)

Method	Precision	Specificity	Overall Accuracy
FCN [36]	96.15	98.34	95.05
Fmask [61]	88.65	94.20	94.94
Cloud-Net [58]	97.60	98.97	95.86
Cloud-Net+ [62]	97.33	98.83	96.36
Cloud-SLR (Ours)	97.94	99.56	95.12

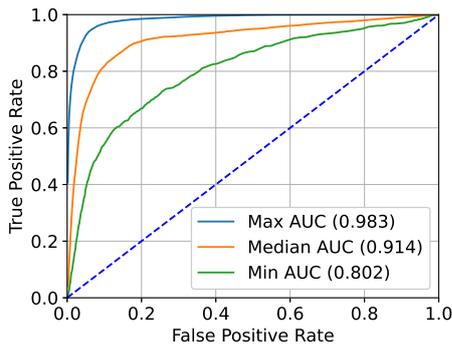


Fig. 7: ROC curves over OSCD dataset; the AUC scores are shown in the legends.

an overall accuracy exceeding 95%, while also attaining the highest precision in comparison to extant techniques. The heightened dependability and accuracy of Cloud-SLR ensure the retention of changed MPs without compromise, which constitutes the focal content intended for transmission through the gateways.

D. Semantic Encoding Performance

We evaluate the Receiver Operating Characteristics (ROC) of the change detection model with different thresholds to show trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) ratio. Herein, the TPR and FPR are respectively defined as $TPR = TP / (TP + FN)$, and $FPR = FP / (FP + TN)$. Based on ROC, the Area Under the ROC Curve (AUC) could be calculated, which measures how well predictions are ranked.

The efficacy of the Change-Net model is validated in Fig. 7 by analyzing Receiver Operating Characteristic (ROC) curves

across the OSCD dataset under varying thresholds with the highest, median, and lowest AUC levels, respectively. The notable performance of our Change-Net is readily observed as it achieves a high level of performance with a maximum AUC of 0.98 and a minimum AUC of 0.80. Therefore, selecting the appropriate score threshold to guarantee the fulfillment of constraints in (17) while simultaneously ensuring the selection of an adequate number of modified points can be considered a dependable approach. The demonstrated success of Change-Net with the OSCD dataset is significant as it sets the stage for subsequent endeavors when enabling the proficient selection of changed MPs as a foundational step toward the efficient identification of semantic map pixels in real-world scenarios. This strategic approach not only holds the potential to enhance data selection but also contributes to saving transmission resources.

Figure 8 shows the efficacy of Change-Net through four illustrative scenarios drawn from the OSCD dataset. More specifically, images in Fig. 8(a) and Fig. 8(b) correspond to pairs of images captured at the same geographical location but at different times. Fig. 8(c) manifests as a cartographic representation of change probabilities at the pixel level, intricately portraying the probability of alterations within the visual domain. Subsequently, Fig. 8(d) enhances this discourse by integrating two tiers of information: firstly, the ground truth depicted in blue, capturing the change map between images in Fig. 8(a) and Fig. 8(b); and secondly, the composite representation of the semantic encoding map rendered in red. The ensuing interplay between the probabilistic rendering of figure in Fig. 8(c) and a well-structured scoring mechanism facilitates the extraction of pixels endowed with semantic depth. It is noteworthy that the semantic encoding map, covering a substantial portion of the ground truth, achieves a coverage rate of approximately 95%, contingent upon the specific geographical locations. This map then plays as a semantic mask to extract information for subsequent transmission to the gateway. Based on the probabilistic alteration representations depicted in Fig. 8(c), the extent of semantic information chosen exhibits variability. For example, in scenarios involving image pairs captured in locales such as Hong Kong or Montpellier, where the probability change map distinctly delineates the unchanged-changed MP pairs, the discernment of semantic MP candidates closely approximates ground truth. Consequently, the data transmission burden upon the encoder is notably diminished. Conversely, instances like that of Las Vegas may manifest a profusion of MPs bearing elevated probabilities of being designated as changed MPs. Hence, ensuring a 95% assurance in the selection of altered MPs entails the concurrent selection of numerous unchanged MPs. In contrast to prevailing methodologies which merely convey changed pixels, our proposed approach represents a logical progression, preserving crucial insights into the sensing images and thus enabling practical applications within the domain of satellite systems.

In Figure 9, we present a graphical representation illustrating the cumulative percentage of downloaded content, as well as the probabilities associated with changed and unchanged MPs, juxtaposed against the score attributed to multi-spectral

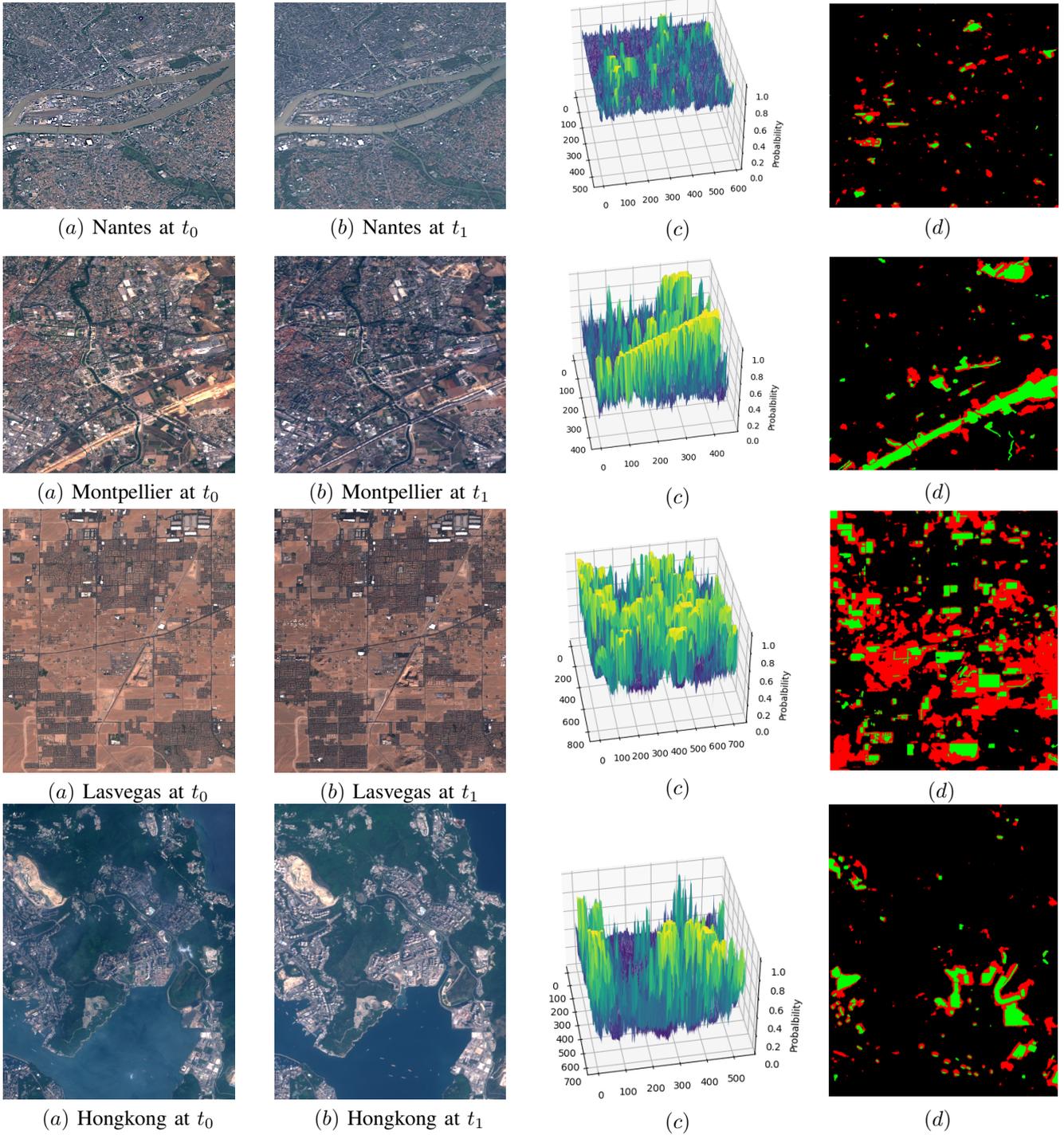


Fig. 8: (a), (b) Example pair of images in different times (displayed in RGB bands). (c) Probability map of change. (d) Ground truth and Semantic encoding map, respectively (Ground Truth is displayed in blue and Semantic map is display in red).

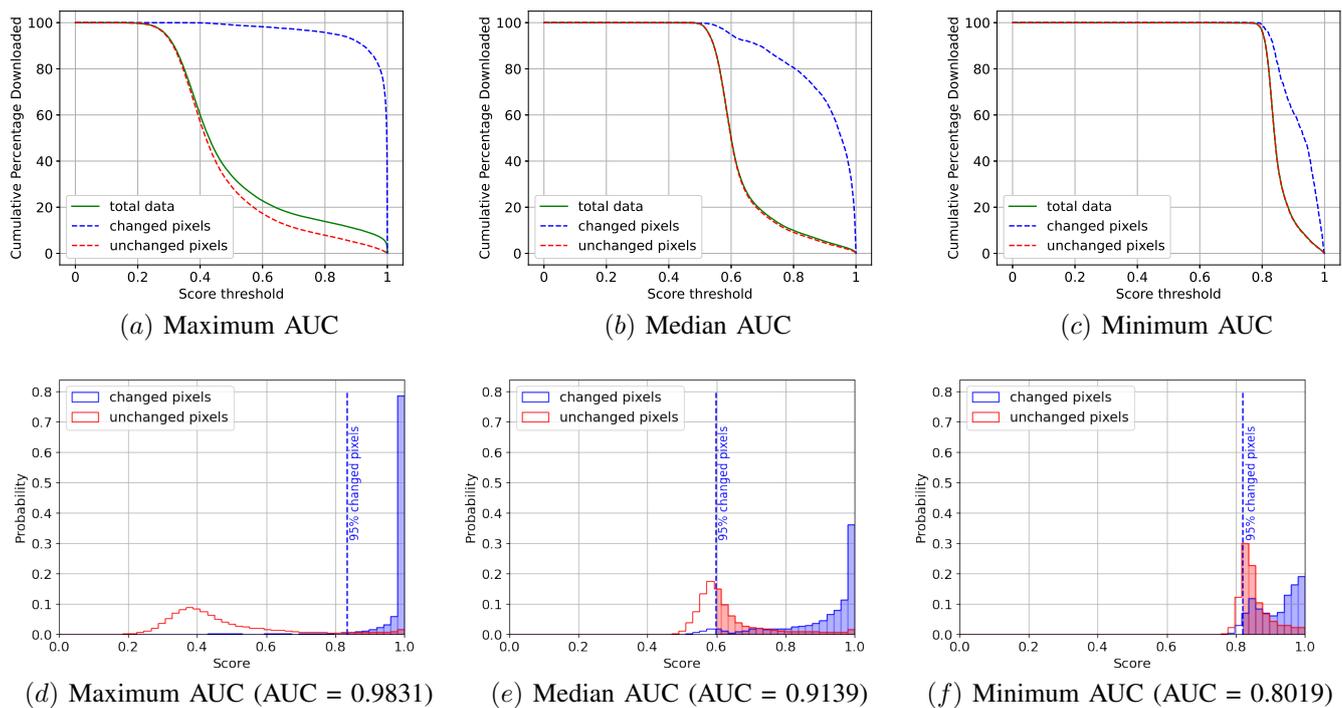


Fig. 9: (a-c) Cumulative Percentage Downloaded versus score of multi-spectral images under different AUC; (d-f) Probability of changed and unchanged MPs versus score of multi-spectral images.

images, across distinct AUC values. It is easy to see that the utilization of varied AUC values results in a considerable disparity in the distribution pattern of changed and unchanged MPs, as exemplified in Fig. 9(a)-(c). Consequently, the attainment of a 95% ratio of changed MPs necessitates an adjustment of the score threshold. Specifically, MPs annotated as 0 and 1 exhibit a degree of overlap, as evidenced in Fig. 9(d)-(f), engendering the situation wherein the requisite count of changed MPs selection also entails the selection of a number of unchanged MPs (denoted by the red regions in Fig. 8(d)). Specifically, in order to ensure the identification of a satisfactory quantity of changed MPs at a 95% frequency, approximately 50% of the unchanged MPs are classified with a value of 1. We emphasize that our approach places primacy on the establishment of a threshold for designating a specific fraction of changed MPs post-training, rather than solely evaluating the model’s performance metrics on the test dataset.

Summing up, we have evaluated our proposed algorithm in the Sentinel-2 system, demonstrating significant improvements in energy efficiency. Our results indicate that the proposed architecture reduces data transmission volume by up to 55% and conserves at least 28% of energy compared to traditional methods. The reliability and efficiency of the Cloud-SLR and Change-Net models have been validated through experimental results, ensuring their high practical applicability. Specially, we used the 38-Cloud dataset to train and test Cloud-SLR, achieving an overall accuracy exceeding 95% and a specificity of 99.56%. The Change-Net model’s performance was confirmed through OSCD dataset, achieving a maximum AUC of 0.98, demonstrating high performance in change detection.

V. CONCLUSION AND DISCUSSION

This study investigated the context of EO with the primary goal of augmenting the efficiency of processing operations and mitigating energy utilization within edge devices throughout the phases of sensing data transmission and storage. To realize this objective, we introduce a comprehensive end-to-end learning-based scoring framework, which is designed to autonomously encode semantic MPs from multi-spectral sensing images. In accordance with this approach, initial preprocessing and the elimination of cloud interference from captured images are executed via the implementation of the Cloud-SLR technique. Subsequently, the establishment of the semantic map is facilitated through the utilization of the Change-Net model, coupled with a scoring system, thereby enabling streamlined encoding and effective data transmission to ground-based stations. The restoration of captured images is facilitated through the integration of received data and reference images. The numerical results demonstrate the enhanced data detection capabilities of our proposed solution, underscoring its potential to achieve a remarkable reduction of up to 58% in energy consumption. The proposed architecture serves as a valuable foundation for future research endeavors, providing inspiration for conducting more comprehensive case studies that incorporate cutting-edge techniques for mitigating various forms of noise, thus paving the way for the realization of truly intelligent satellite systems.

REFERENCES

- [1] V.-P. Bui, T. Q. Dinh, I. Leyva-Mayorga, S. R. Pandey, E. Lagunas, and P. Popovski, “On-board change detection for resource-efficient earth observation with leo satellites,” in *Proc. IEEE GLOBECOM*, 2023.
- [2] D. Poursanidis and N. Chrysoulakis, “Remote sensing, natural hazards and the contribution of esa sentinels missions,” *Remote Sensing Applications: Society and Environment*, vol. 6, pp. 25–38, 2017.

- [3] ESA. Sentinel-2 operations. Accessed: 17 Feb. 2023. [Online]. Available: https://www.esa.int/Enabling_Support/Operations/Sentinel-2_operations
- [4] R. Caye Daudt *et al.*, "OSCD - onera satellite change detection," 2019. [Online]. Available: <https://dx.doi.org/10.21227/asqe-7s69>
- [5] M. D. King *et al.*, "Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, 2013.
- [6] R. Carnap, Y. Bar-Hillel *et al.*, "An outline of a theory of semantic information," 1952.
- [7] I. Leyva-Mayorga *et al.*, "Satellite edge computing for real-time and very-high resolution Earth observation," *IEEE Trans. Commun.*, pp. 1–1, 2023.
- [8] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 3, pp. 567–579, 2019.
- [9] J. Xu *et al.*, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 4, pp. 2315–2328, 2022.
- [10] Z. Weng *et al.*, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 9, pp. 6227–6240, 2023.
- [11] X. Peng *et al.*, "A robust deep learning enabled semantic communication system for text," in *Proc. IEEE GLOBECOM, 2022*, pp. 2704–2709.
- [12] Y. Duan *et al.*, "Multimedia semantic communications: Representation, encoding and transmission," *IEEE Netw.*, vol. 37, no. 1, pp. 44–50, 2023.
- [13] S. Fang, K. Li, J. Shao, and Z. Li, "Snunet-cd: A densely connected siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [14] W. A. Malila, "Change vector analysis: An approach for detecting forest changes with landsat," in *LARS symposia*, 1980, p. 385.
- [15] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k -means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, 2009.
- [16] J. B. Collins and C. E. Woodcock, "An assessment of several linear change detection techniques for mapping forest mortality using multi-temporal landsat tm data," *Remote Sens. Environ.*, vol. 56, no. 1, pp. 66–77, 1996.
- [17] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.
- [18] F. Thonfeld *et al.*, "Robust change vector analysis (RCVA) for multi-sensor very high resolution optical satellite data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 50, pp. 131–140, 2016.
- [19] W. Fang and C. Xi, "Land-cover change detection for sar images based on biobjective fuzzy local information clustering method with decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [20] Y. Li, X. Li, J. Song, Z. Wang, Y. He, and S. Yang, "Remote-sensing-based change detection using change vector analysis in posterior probability space: A context-sensitive bayesian network approach," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 3198–3217, 2023.
- [21] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- [22] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 2, pp. 22–40, 2016.
- [23] H. Chen, Y. He, L. Zhang, W. Yang, Y. Liu, B. Gao, Q. Zhang, and J. Lu, "A multi-input channel u-net landslide detection method fusing sar multi-source remote sensing data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2023.
- [24] K. Bhuyan, S. R. Meena, L. Nava, C. van Westen, M. Floris, and F. Catani, "Mapping landslides through a temporal lens: an insight toward multi-temporal landslide mapping using the u-net deep learning model," *GIScience & Remote Sensing*, vol. 60, no. 1, p. 2182057, 2023.
- [25] M. Zhang and W. Shi, "A feature difference convolutional neural network-based change detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7232–7246, 2020.
- [26] J. Sherrah, "Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery," *arXiv preprint arXiv:1606.02585*, 2016.
- [27] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
- [28] C. Cao, S. Dragičević, and S. Li, "Land-use change detection with convolutional neural network methods," *Environments*, vol. 6, no. 2, p. 25, 2019.
- [29] T. Yan, Z. Wan, and P. Zhang, "Fully transformer network for change detection of remote sensing images," in *Proceedings of the Asian Conference on Computer Vision, 2022*, pp. 1691–1708.
- [30] A. Codegoni, G. Lombardi, and A. Ferrari, "Tinydc: A (not so) deep learning model for change detection," *Neural Comput. Appl.*, vol. 35, no. 11, pp. 8471–8486, 2023.
- [31] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, 2018.
- [32] R. R. Irish, "Landsat 7 automatic cloud cover assessment," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049. SPIE, 2000, pp. 348–355.
- [33] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, 2012.
- [34] Z. An and Z. Shi, "Scene learning for cloud detection on remote-sensing images," *IEEE J. Sel. Top. Appl. Earth Obs.*, vol. 8, no. 8, pp. 4206–4222, 2015.
- [35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE CVPR*, 2015, pp. 3431–3440.
- [36] S. Mohajerani, T. A. Krammer, and P. Saedi, "Cloud detection algorithm for remote sensing images using fully convolutional neural networks," *arXiv preprint arXiv:1810.05782*, 2018.
- [37] J. Dröner, N. Korfhage, S. Egli, M. Mühlhng, B. Thies, J. Bendix, B. Freisleben, and B. Seeger, "Fast cloud segmentation using convolutional neural networks," *Remote Sensing*, vol. 10, no. 11, p. 1782, 2018.
- [38] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftgaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote Sens. Environ.*, vol. 229, pp. 247–259, 2019.
- [39] A. Francis, P. Sidiropoulos, and J.-P. Muller, "Cloudfcn: Accurate and robust cloud detection for satellite imagery with deep learning," *Remote Sensing*, vol. 11, no. 19, p. 2312, 2019.
- [40] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*. Springer, 2015, pp. 234–241.
- [41] F. Xu, Y. Shi, P. Ebel, L. Yu, G.-S. Xia, W. Yang, and X. X. Zhu, "Glf-cr: Sar-enhanced cloud removal with global-local fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 192, pp. 268–278, 2022.
- [42] C. Duan, M. Belgiu, and A. Stein, "Efficient cloud removal network for satellite images using sar-optical image fusion," *IEEE Geosci. Remote Sens. Lett.*, 2024.
- [43] C. Duan, J. Pan, and R. Li, "Thick cloud removal of remote sensing images using temporal smoothness and sparsity regularized tensor optimization," *Remote Sensing*, vol. 12, no. 20, p. 3446, 2020.
- [44] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, 2019.
- [45] Q. Wang, Z. Yuan, Q. Du, and X. Li, "Getnet: A general end-to-end 2-d cnn framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, 2019.
- [46] T.-J. Yang, Y.-H. Chen, J. Emer, and V. Sze, "A method to estimate the energy consumption of deep neural networks," in *Proc. IEEE ACSSC*, 2017, pp. 1916–1920.
- [47] Z. Wang and N. Verma, "A low-energy machine-learning classifier based on clocked comparators for direct inference on analog sensors," *IEEE Trans. Circuits Syst. I: Regul. Pap.*, vol. 64, no. 11, pp. 2954–2965, 2017.
- [48] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
- [49] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wirel. Commun.*, vol. 12, no. 9, pp. 4569–4581, 2013.
- [50] X. Li *et al.*, "Wirelessly powered crowd sensing: Joint power transfer, sensing, compression, and transmission," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 2, pp. 391–406, 2018.
- [51] A. Morello and V. Mignone, "Dvb-s2: The second generation standard for satellite broad-band services," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 210–227, 2006.
- [52] B. Matthiesen *et al.*, "A globally optimal energy-efficient power control framework and its efficient implementation in wireless interference networks," *IEEE Trans. Signal Process.*, vol. 68, pp. 3887–3902, 2020.
- [53] K. S. Basavaraju *et al.*, "Ucdnet: A deep learning model for urban change detection from bi-temporal multispectral sentinel-2 satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–10, 2022.
- [54] C. Wu *et al.*, "Unsupervised change detection in multitemporal VHR images based on deep kernel PCA convolutional mapping network," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 12 084–12 098, 2021.
- [55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [56] C. Szegedy *et al.*, "Rethinking the inception architecture for computer vision," in *Proc. IEEE CVPR*, 2016, pp. 2818–2826.
- [57] R. C. Daudt *et al.*, "Multitask learning for large-scale semantic change detection," *Comput. Vis. Image Underst.*, vol. 187, p. 102783, 2019.
- [58] S. Mohajerani and P. Saedi, "Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery," in *Proc. IEEE IGARSS*, 2019, pp. 1029–1032.
- [59] R. C. Daudt *et al.*, "Urban change detection for multispectral earth observation using convolutional neural networks," in *Proc. IEEE IGARSS*, 2018, pp. 2115–2118.
- [60] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE CVPR*, 2015, pp. 1–9.
- [61] Z. Zhu *et al.*, "Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images," *Remote Sens. Environ.*, vol. 159, pp. 269–277, 2015.
- [62] S. Mohajerani and P. Saedi, "Cloud-net+: A cloud segmentation cnn for landsat 8 remote sensing imagery optimized with filtered jaccard loss function," *arXiv e-prints*, pp. arXiv–2001, 2020.



Van-Phuc Bui (Student Member, IEEE) received the B.Sc. degree (Hons.) in electronics and telecommunications from the Ho Chi Minh University of Technology, Vietnam, in 2018 and the M.Sc. degree in telecommunication from Soongsil University, South Korea, in 2020. He was a Research Assistant at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, in 2021. He is currently pursuing a Ph.D. at Aalborg University, Denmark. His research interests include optimization algorithms and artificial intelligence for wireless communications, semantic communications and networking, and satellite communications.



Eva Lagunas (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications engineering from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 2010 and 2014, respectively. She has held positions at UPC, Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), University of Pisa, Italy; and the Center for Advanced Communications (CAC), Villanova University, PA, USA. In 2014, she joined the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, where she currently holds a Research Scientist position. Her research interests include terrestrial and satellite system optimization, spectrum sharing, resource management and machine learning.



Thinh Quang Dinh (Member, IEEE) received the B.Eng. degree (Hons.) in electrical and electronic engineering (specializing in telecommunications) from the Ho Chi Minh City University of Technology (HCMUT), Vietnam, in 2013, and the Ph.D. degree from the Singapore University of Technology and Design (SUTD) in 2019 under the President's Graduate Fellowship. In 2018, he was a Visiting Student with the Wireless Computing Laboratory, University of Toronto. In 2019, he was a Data Scientist with Trusting Social. He was honored with the 2020 IEEE Stephen O. Rice Prize. His research interests include both quantum and classical communication systems and networking.



Israel Leyva-Mayorga (Member, IEEE) received the B.Sc. degree in telematics engineering and the M.Sc. degree (Hons.) in mobile computing systems from the Instituto Politécnico Nacional (IPN), Mexico, in 2012 and 2014, respectively, and the Ph.D. degree (cum laude and extraordinary prize) in telecommunications from the Universitat Politècnica de València (UPV), Spain, in 2018. He was a Visiting Researcher at the Department of Communications, UPV, in 2014, and at the Deutsche Telekom Chair of Communication Networks, Technische Universität Dresden, Germany, in 2018. He is currently

an Assistant Professor at the Connectivity Section (CNT) of the Department of Electronic Systems, Aalborg University (AAU), Denmark, where he served as a Postdoctoral Researcher from January 2019 to July 2021. He is an Associate Editor for IEEE WIRELESS COMMUNICATIONS LETTERS and a Board Member for one6G. His research interests include beyond-5G, and 6G networks, satellite communications, and random and multiple access protocols.



Petar Popovski (Fellow, IEEE) is a Professor at Aalborg University, where he heads the section on Connectivity and a Visiting Excellence Chair at the University of Bremen. He received his Dipl.-Ing and M. Sc. degrees in communication engineering from the University of Sts. Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005. He received an ERC Consolidator Grant (2015), the Danish Elite Researcher Award (2016), the IEEE Fred W. Ellersick Prize (2016), the IEEE Stephen O. Rice Prize (2018), the Technical Achievement Award from the IEEE Technical Committee on Smart Grid Communications (2019), the Danish Telecommunication Prize (2020) and Villum Investigator Grant (2021). He was a Member at Large of the Board of Governors of IEEE Communication Society 2019-2021. He is currently an Editor-in-Chief of the IEEE Journal on Selected Areas in Communications and a Chair of the IEEE Communication Theory Technical Committee. His research interests are in the area of wireless communication and communication theory. He authored the book "Wireless Connectivity: An Intuitive and Fundamental Guide".

mittee on Smart Grid Communications (2019), the Danish Telecommunication Prize (2020) and Villum Investigator Grant (2021). He was a Member at Large of the Board of Governors of IEEE Communication Society 2019-2021. He is currently an Editor-in-Chief of the IEEE Journal on Selected Areas in Communications and a Chair of the IEEE Communication Theory Technical Committee. His research interests are in the area of wireless communication and communication theory. He authored the book "Wireless Connectivity: An Intuitive and Fundamental Guide".



Shashi Raj Pandey (Member, IEEE) received his B.E. degree in Electrical and Electronics with a specialization in communication engineering from Kathmandu University, Nepal, and the Ph.D. degree in Computer Science and Engineering from Kyung Hee University, Seoul, South Korea. He is currently an Assistant Professor at the Department of Electronic Systems, Aalborg university. Prior, he was a Postdoctoral Researcher at the Connectivity Section, Aalborg University, from 2021 to 2023. He served as a Network Engineer at Huawei Technologies Nepal Co. Pvt. Ltd, Nepal, from 2013 to 2016. His research

interests include network economics, game theory, wireless communications, data markets and distributed machine learning. He was the recipient of the Best Paper Award at several conferences, including IEICE APNOMS 2019. He was a Member at Large at the IEEE Communication Society Young Professionals 2020 – 2021. He currently serves as a Member at Large in the IEEE Communication Society (ComSoc) On-Line Content Board, IT Standing Committee, and is on the editorial advisory board of IEEE's Spectrum The Institute 2022 – 2024. He is an affiliated member of the Pioneer Center for AI, Denmark.