HIGH-RESOLUTION SEMANTICALLY CONSISTENT IMAGE-TO-IMAGE TRANSLATION

Neha Fathima Begum
CSE Department,
Shadan women's College of Engineering and Technology,
Hyderabad, India
nehacse30@gmail.com

Dr. Mohd Umar Farooq
CSE Department,
Shadan women's College of Engineering and Technology,
Hyderabad, India
umarfarooq.mohd@gmail.com

Dr. K. Palani
CSE Department,
Shadan women's College of Engineering and Technology,
Hyderabad, India
principalswcet2020@gmail.com

ABSTRACT

In recent years, deep learning has emerged as one of the most effective computer vision methods available to remote sensing experts. To reduce the disparity across satellite image datasets, scientists must address the domain adaptation (DA) problem because the remote sensing datasets lack training labels. Consequently, after being trained, image segmentation models could be more broad and make use of an already-existing set of labels rather than needing the creation of new ones. This study suggests an unsupervised DA model that maintains the images' per-pixel quality and semantic consistency during the style-transferring process. The main contribution of this paper is the revised SemI2I model architecture, which greatly improves the model's performance and allows it to compete with the most advanced Cy CADA model. The future development of the proposed method could include ecological domain transfer, a priori evaluation of dataset quality in terms of data distribution, or exploration of the inner architecture of the DA model. A second contribution is to test the Cy CADA model using multiband remote sensing datasets such as SPOT-6 and WorldView-2. The proposed model preserves the semantic coherence and per-pixel quality of the images during the style-transfer procedure. Consequently, the semantic segmentation model outperforms the SemI2I model and obtains results comparable to the state-of-the-art Cy CADA model after being trained on the changed photos. The proposed method could be extended to investigate the internal architecture of the DA model, transfer of ecological domains, and a priori dataset quality assessment based on data distribution.

Keywords: Domain Adaptation (DA), Remote Sensing (RS), Unsupervised Domain Adaptation (UDA), Data Flow Diagram (DFD).

I. INTRODUCTION

The statement highlights the considerable difficulties posed by the heavy reliance on large-scale supervised land cover labels for the production of useful map products in the era of big Earth data, often known as remote sensing (RS) big data [1]. To overcome this difficulty, RS image representations from various satellites and geographical locations can be aligned using unsupervised domain adaptation (DA) techniques [16]. Examples are circumstances in which a new area lacks land cover labels or their availability is questionable, in which the sensor attributes in the two domains differ, or in which both of these scenarios happen simultaneously [16]. Furthermore, existing processes for producing maps can effectively include these adaptation techniques [12]. This DA strategy lessens the difference between the target domain (pictures without labels) and the source domain (images with associated semantic labels) so that the segmentation model that is trained later can function well with both datasets [15]. It is crucial to make sure the modifications are carried out in a way that maintains the significant land content and structure of the picture because the application domain is RS-based satellite photography [13]. Thus, the key feature of the presented work is that the adapted images are highly accurate in terms of semantic consistency, i.e., the objects in the adapted images preserve their original logical meaning [6]. Semantically consistent adaptation is crucial in RS because each pixel brings certain information which must be preserved [14]. Focusing on semantically consistent approaches means that the solution presented here is increasingly important for government and private industries since existing map products (i.e., training labels) in RS are limited, not publicly available, or expensive to obtain [11].

OBJECTIVE

A large number of skilled professionals are involved in the creation of Land Use and Land Cover (LULC) maps. The creation of LULC maps was done in a semi-automated manner prior to the introduction of deep learning (DL) for RS data [4]. It used pre-existing tools and algorithms in addition to requiring a human in the loop. An expert has to visually evaluate each pixel in the satellite image and provide a label to it in order to label one [2]. The hunt for fully automated solutions was fuelled by the need for higher temporal frequency of LULC maps and the labor-intensive and costly nature of the labor involved [3].

II. PROBLEM STATEMENT

In the realm of big Earth data and remote sensing (RS) big data, the reliance on extensive supervised land cover labels for producing accurate map products poses significant challenges [1]. The need for large-scale labeled data hinders effective map generation, especially in areas lacking such labels or where sensor characteristics vary across domains [5]. Unsupervised domain adaptation (DA) methods emerge as a solution by aligning RS images from diverse satellites and regions, facilitating knowledge transfer from labeled to unlabeled domains [16]. This adaptation becomes crucial when land cover labels are scarce or inaccessible, and when sensor variations exist between domains [8]. Integrating DA methods into map production workflows offers a viable strategy, bridging the gap between labeled and unlabeled datasets for segmentation model training [6]. The key focus lies in preserving semantic consistency during adaptation, ensuring that objects in adapted images retain their logical meaning [14]. This semantic fidelity is paramount in RS applications, where pixel-level information integrity is crucial [10]. Thus, the central challenge addressed here is enabling accurate, semantically consistent adaptation of RS imagery to overcome limitations in existing map products, benefiting government and private sectors reliant on RS data for decision-making and analysis [11].

EXISTING SYSTEM

We present a framework for Ancient Hieroglyphic Language translation based on the Histogram of Oriented Gradients (HOG) technique [9]. The performance metric utilized in this context is the handover failure rate [12]. The framework assesses mobility issues, encompassing both too-early and too-late handovers, to translate images to text [4]. It allows for quick training and prediction, making it suitable for scenarios where computational efficiency is crucial [8]. It can provide rapid insights into the likelihood of translations from images to text. However, it may not capture intricate relationships between features as effectively as more complex algorithms [7].

Disadvantage of Existing System

While our framework based on the Histogram of Oriented Gradients (HOG) technique offers quick training and prediction for Ancient Hieroglyphic Language translation, it comes with certain disadvantages. One notable limitation is its potential inability to capture intricate relationships between features as effectively as more complex algorithms. This could result in a loss of nuanced semantic understanding, leading to inaccuracies or ambiguities in the translated text.

PROPOSED SYSTEM

The proposed system has the ability to automatically learn relevant features and patterns from scanned images without the need for manual feature engineering [13]. It is especially beneficial in complex tasks like image-to-text translation, where there is subtle and hard-to-define information using traditional handcrafted features [15]. The system is capable of handling variations in image-to-text translation, lighting conditions, and complex images [16]. It can learn to generalize across different images, predict language, and translations [6]. This capability enhances the accuracy and efficiency of the translation process, making it a valuable tool for applications where precise and reliable translations are critical [14].

Advantages of Proposed System

Well-suited for tasks such as image classification and object detection. Robustly identify ancient language from image.

III. RELATED WORKS

Several related works contribute to the field of Ancient Hieroglyphic Language translation, each leveraging different techniques and methodologies. One notable approach involves the use of deep learning models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants like Long Short-Term Memory (LSTM) networks. These models excel in capturing complex patterns and dependencies within hieroglyphic images, leading to more accurate and contextually relevant translations. Another line of research focuses on utilizing advanced feature extraction techniques, such as Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF), to improve the robustness and reliability of translation systems by enhancing the recognition of hieroglyphic symbols and their spatial relationships. Additionally, some studies explore the integration of domain-specific knowledge and linguistic rules into translation algorithms, aiming to enhance semantic understanding and improve the coherence of translated texts. Furthermore, advancements in natural language processing (NLP) techniques, such as word embeddings and attention mechanisms, have been applied to hieroglyphic translation tasks, facilitating better alignment between image features and textual representations. Overall, these related works contribute diverse perspectives and methodologies to the challenge of Ancient Hieroglyphic Language translation, driving innovation and progress in this specialized field.

IV. METHODLOGY OF PROJECT

The methodology for the Ancient Hieroglyphic Language translation project employing the Histogram of Oriented Gradients (HOG) technique involves a systematic approach to handling hieroglyphic images. Initially, a diverse dataset of hieroglyphic images is collected, encompassing various symbols, characters, and textual contexts to ensure comprehensive coverage. These images undergo preprocessing steps, including noise reduction, contrast enhancement, and standardization of sizes and formats, to improve their quality and uniformity. Subsequently, the HOG technique is applied to extract relevant features from the pre-processed images, capturing shape and texture information critical for symbol recognition.

Following feature extraction, a machine learning model, such as a Support Vector Machine (SVM) or

Random Forest classifier, is trained using the extracted HOG features and labeled data associating features with hieroglyphic symbols. This training phase enables the model to learn patterns and associations necessary for accurately recognizing and classifying hieroglyphic symbols.

MODULES:

Image Dataset: A dataset is a collection of data treated as a single unit by a computer. It contains separate pieces of data but can be used to train an algorithm with the goal of finding predictable patterns within the entire dataset. High-quality datasets are essential for AI advances, often more critical than the algorithms themselves. In fact, having high-quality datasets can lead to breakthroughs in AI six times faster than improvements in algorithms.

Data Analysis: Data analysis is the process of cleaning, analyzing, interpreting, and visualizing data using various techniques and business intelligence tools. Data analysis tools help you discover relevant insights that lead to smarter and more effective decision-making.

Preprocessing: Preprocessing is the process of preparing data for analysis. It can involve Data preprocessing is a key step in data mining, machine learning, and other data science tasks. It can help improve the accuracy of new models and reduce the amount of compute required.

CNN Model / Feature Extraction: A Convolutional Neural Network (CNN) is a type of deep learning algorithm particularly well-suited for image recognition and processing tasks. It is composed of multiple layers, including: Convolutional Layers: These are the key components of a CNN. Filters are applied to the input image to extract features such as edges, textures, and shapes. Pooling Layers: These down-sample the feature maps, reducing spatial dimensions while retaining essential information. Fully Connected Layers: These make predictions or classify the image.

GAN Model Apply: A Convolutional Neural Network (CNN) is a type of deep learning algorithm specifically designed for image processing and recognition tasks. Compared to alternative classification models, CNNs require less preprocessing as they can automatically learn hierarchical feature representations from raw input images.

Train Model: In machine learning, training a model is the process of teaching a machine learning algorithm to make predictions or decisions based on data. This process is similar to teaching a child to recognize patterns from examples.

Test Model: Definition. In machine learning, model testing is referred to as the process where the performance of a fully trained model is evaluated on a testing set. Unit test. Check the correctness of individual model components. Regression test. Check whether your model breaks and test for previously encountered bugs. Integration test. Check whether the different components work with each other within your machine learning pipeline.

Evaluation Model: Model evaluation in machine learning is the process of determining a model's effectiveness and quality. It involves using a variety of metrics and approaches to evaluate whether the model achieves the required goals and how well it generalizes to fresh, untested data.

Deployment: Model deployment in machine learning is the process of integrating your model into an existing production environment where it can take in an input and return an output. The goal is to make the predictions from your trained machine learning model available to others.

V. ALGORITHM USED IN PROJECT

An Unsupervised Domain Adaptation (UDA) model for images typically integrates Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) to address domain shift between datasets. CNNs are used for feature extraction and learning representations from images, capturing hierarchical features like edges, textures, and object shapes. In the context of UDA, a CNN pre-trained on the source domain with labeled data is leveraged as a feature extractor. It learns to encode useful information from the source domain images, such as identifying objects or patterns.

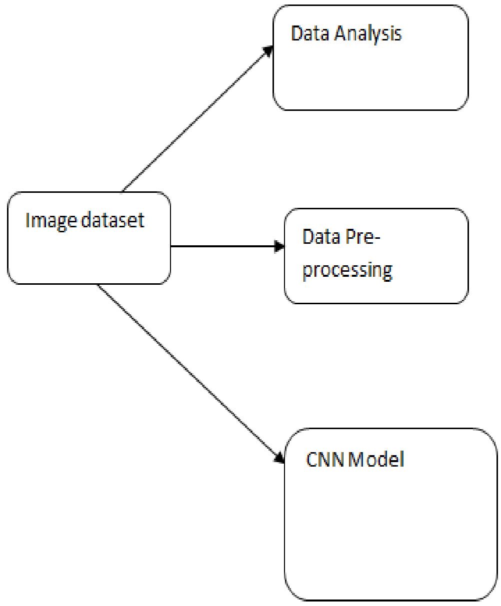
On the other hand, GANs play a crucial role in domain adaptation by generating realistic images in the target domain. The GAN consists of two components: a generator and a discriminator. The generator creates synthetic images resembling the target domain, while the discriminator learns to distinguish between real target domain images and generated ones. This adversarial training process encourages the generator to produce images that are indistinguishable from real target domain images.

Benefits

- Leveraging a pre-trained CNN for feature extraction allows the model to capture relevant and discriminative features from the source domain, enhancing the learning process in the target domain.
- The combination of CNNs and GANs helps address domain shift by generating synthetic images in the target domain, bridging the gap between the distributions of the source and target domains.
- Fine-tuning the CNN using target domain data, augmented with GAN-generated images, improves the

model's ability to generalize across domains, leading to better performance on unseen target domain samples.

DATA FLOW DIAGRAM



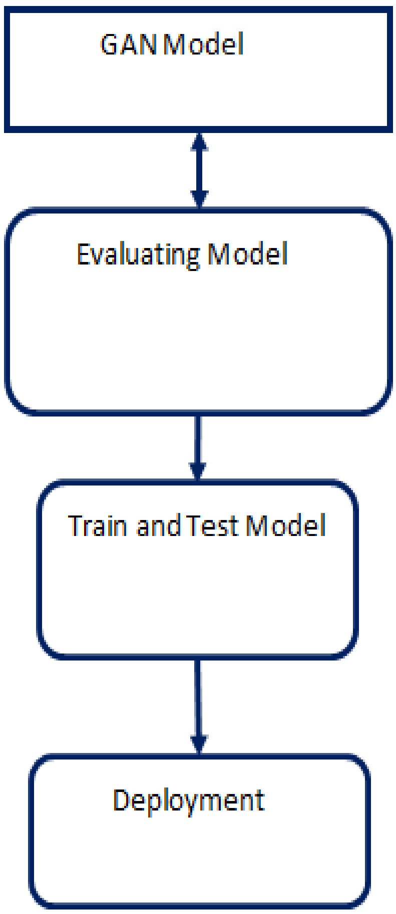


Fig: 5 Flow Diagrams of Modules

Impediments of DL

GANs, while powerful for generating realistic data, face several challenges. These include mode collapse (limited output diversity), training instability, hyperparameter sensitivity, mode dropping (missing data variations), difficulty in evaluation metrics, limited interpretability, data quality/quantity impact, potential for overfitting, and a lack of transparency in generated data's relationship to the underlying distribution. Overcoming these requires algorithmic enhancements, architectural adjustments, hyperparameter tuning, and domain-specific expertise for effective training and usage.

VI. SYSTEM ARCHITECTURE

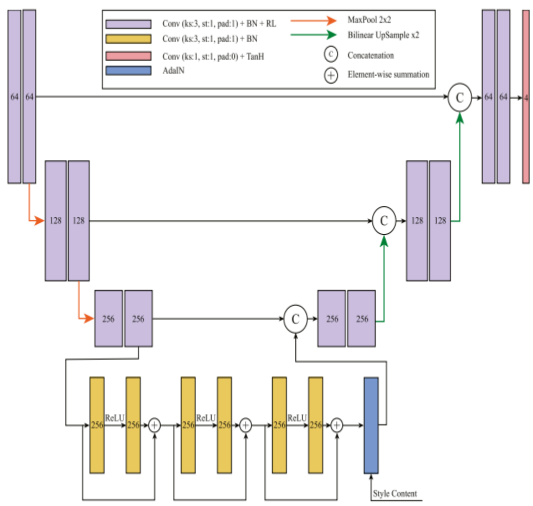
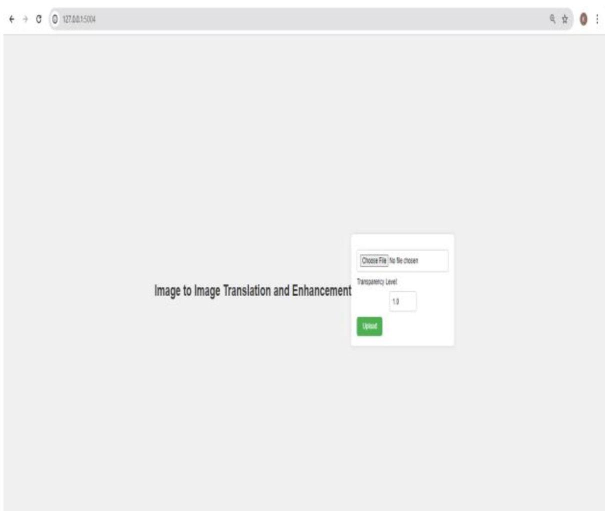
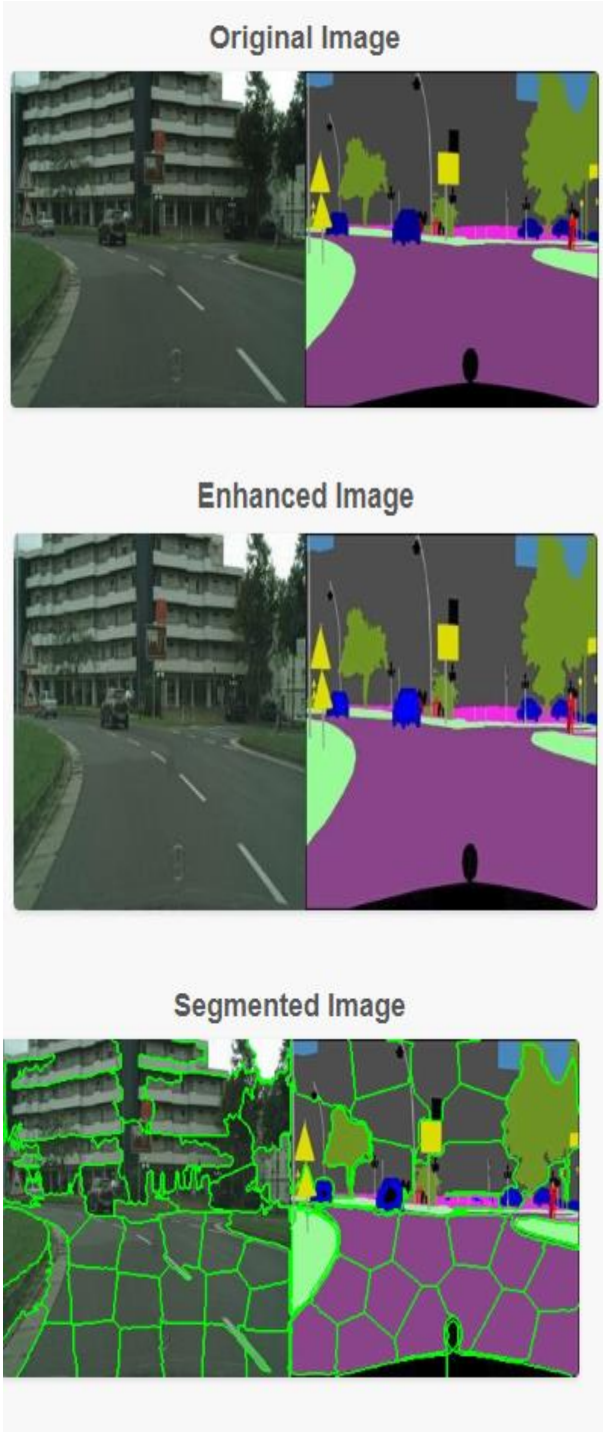


Fig: 6 SYSTEM ARCHITECTURE OF PROJECT

VII. RESULTS AND DISCUSSION





VIII. FUTURE ENHANCEMENT

The future development of the proposed method could include ecological domain transfer, a priori evaluation of dataset quality in terms of data distribution, or exploration of residual blocks in the encoder. First, both datasets (the source and the target) are composed of satellite images taken in different ecological regions, and, thus, represent different style characteristics. Potentially, there could be a situation when the target domain is represented by so many samples taken from different regions that the task of acquiring its common style becomes meaningless because of the increased complexity of the style-transferring task.

IX. REFERENCES

[1] Y. Li, J. Ma, and Y. Zhang, “Image retrieval from remote sensing Big Data: A survey,” Inf. Fusion, vol. 67, pp. 94–115, 2021.

[2] Natural Resources Canada, Ottawa, ON, Canada, “Natural resources Canada,” Jul. 2016. Accessed: 2022. [Online]. Available: https://www. nrcan.gc.ca/home

[3] D. Dobrijevic, “Space calendar 2022: Rocket launches, sky events, missions & more! Space,” 2022. [Online]. Available: https://www.space.com/ 32286-space-calendar.html

[4] D. Lu and Q. Weng, “A survey of image classification methods and techniques for improving classification performance,” Int. J. Remote Sens., vol. 28, no. 5, pp. 823–870, Mar. 2007.

[5] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deep Lab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018.

[7] G. Lin, C. Shen, A. van den Hengel, and I. Reid, “Efficient piecewise training of deep structured models for semantic segmentation,” in Proc. IEEE Conf. Comput.Vis. Pattern Recognit., 2016, pp. 3194–3203.

[8] Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang, “Semantic image segmentation via deep parsing network,” in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 1377–1385.

[9] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 3354–3361.

[10] Y.-H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, X. Lu, and M.-H. Yang, “Deep image harmonization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 3789–3797.

[11] V. Alhassan, C. Henry, S. Ramanna, and C. Storie, “A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery,” Neural Comput. Appl., vol. 32, no. 12, pp. 8529–8544, Jun. 2020.

[12] C. D. Stories and C. J. Henry, “Deep learning neural networks for land use land cover mapping,” in Proc. IEEE Int. Geo sci. Remote Sens. Symp., 2018, pp. 3445–3448.

[13] C. J. Henry et al., “Automated LULC map production using deep neural networks,” Int. J. Remote Sens., vol. 40, pp. 4416–4440, 2019.

[14] I. Good fellow, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., vol. 27, 2014, pp. 1–9.

[15] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, and M.-H. Yang, and M. Chandraker, “Learning to adapt

structured output space for semantic segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.,
2018, pp. 7472–7481.

[16] M. Liu, P. Zhang, Q. Shi, and M. Liu, "An adversarial domain adaptation framework with KL-constraint for remote sensing land cover classification," IEEE Geo sci. Remote Sens. Lett. vol. 19, 2022, Art. No. 3002305.

[17] W. Liu and F. Su, "Unsupervised adversarial domain adaptation network for semantic segmentation," IEEE Geo sci. Remote Sens. Lett. vol. 17, no. 11, pp. 1978–1982, Nov. 2020.

[18] O. Tasar, S. L. Happy, Y. Tarabalka, and P. Alliez, "Color Map GAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks," IEEE Trans. Geo sci. Remote Sens., vol. 58, no. 10, pp. 7178–7193, Oct. 2020.

[19] O. Tasar, S. L. Happy, Y. Tarabalka, and P. Alliez, "SEMI2I: Semantically consistent image-to-image translation for domain adaptation of remote sensing data," in Proc. IEEE Int. Geo sci. Remote Sens. Symp., 2020, pp. 1837–1840.

[20] O. Tasar, A. Giros, Y. Tarabalka, P. Alliez, and S. Clerc, "Daug Net: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images," IEEE Trans. Geo sci. Remote Sens., vol. 59, no. 2, pp. 1067–1081, Feb. 2021