DCV^{2}I: A Practical Approach for Supporting Geographers' Visual Interpretation in Dune Segmentation with Deep Vision Models

Anqi Lu[1*], Zifeng Wu[2*], Zheng Jiang[1], Wei Wang[1], Eerdun Hasi[2], Yi Wang[3]

[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, China
[2]School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, China
[3]School of Computer Science, Beijing University of Posts and Telecommunications, China
{laq2023, jiangzheng, weiwang, yiwang}@bupt.edu.cn, wuzifeng@mail.bnu.edu.cn, hasi@bnu.edu.cn

Abstract

Visual interpretation is extremely important in human geography as the primary technique for geographers to use photograph data in identifying, classifying, and quantifying geographic and topological objects or regions. However, it is also time-consuming and requires overwhelming manual effort from professional geographers. This paper describes our interdisciplinary team's efforts in integrating computer vision models with geographers' visual image interpretation process to reduce their workload in interpreting images. Focusing on the dune segmentation task, we proposed an approach featuring a deep dune segmentation model to identify dunes and label their ranges in an automated way. By developing a tool to connect our model with ArcGIS, one of the most popular workbenches for visual interpretation, geographers can further refine the automatically-generated dune segmentation on images without learning any CV or deep learning techniques. Our approach thus realized a non-invasive change to geographers' visual interpretation routines, reducing their manual efforts while incurring minimal interruptions to their work routines and tools they are familiar with. Deployment with a leading Chinese geography research institution demonstrated the potential of our approach in supporting geographers in researching and solving drylands desertification.

Introduction

Human geography is always a visual discipline (Rose 2003). An image, whether aerial photograph or other digital sensor image, must be interpreted to construct geographic knowledge by geographers (Tolia-Kelly 2012). Visual interpretation is usually the beginning but essential stage of various analysis operations (Lillesand, Kiefer, and Chipman 2015). During visual interpretation, geographic objects or regions are identified, classified, measured, and assessed within geographic information systems (GIS) (Trotter 1991). Imagery data and its derivatives produced in visual interpretation may be integrated with other data, such as climatic and socioeconomic data, to conduct spatial reasoning. While visual interpretation is a powerful tool in geographic knowledge production, it is also a demanding task for geographers (Lillesand, Kiefer, and Chipman 2015). It does not only time-consuming but also labor-intensive (Hua et al. 2022). For example, a professional geographer may take a few days to perform a comprehensive visual interpretation for an image containing hundreds of objects manually. Meanwhile, geographers' cognitive perception and real-world experience are still inevitable in manual visual interpretation, introducing human biases and subjectivity into the analysis (Manley, Filomena, and Mavros 2021).

With the recent progress of computer vision technologies, traditional visual interpretation has been increasingly supported by deep learning for computer vision (Wiley and Lucas 2018). Compared with traditional 'per-pixel' Geographic Object-based Image Analysis (GOBIA) (Blaschke et al. 2014), deep learning could better deal with image semantics and knowledge integration without explicitly extracting image features (Fang et al. 2019). Among many visual interpretation tasks, dune segmentation is particularly suitable for incorporating deep vision models. First, unlike large geographical landscapes, a small-scale aerial photograph of the desert might contain hundreds of dunes, costing overwhelming human efforts even for creating training data. Second, dunes, even in the same area, may exhibit diverse morphological traits, limiting the applicability of standard tools such as Ersi's GeoAI. Third, as a specific class of problem across multiple domains (e.g., different deserts), segmenting dunes has to deal with multiple nuances in imagery data, restricting the capability of general zero-shot segmentation models such as SAM (Kirillov et al. 2023) or SEEM (Zou et al. 2023b).

Focusing on the dune segmentation task, we developed an approach called **DCV^{2}I** to integrate **d**eep **c**omputer **v**ision models with geographers' **v**isual **i**nterpretation workflow. Our approach features a deep dune segmentation model which integrates the multiscale characteristics of Atrous Spatial Pyramid Pooling (ASPP) module (Chen et al. 2018) into the Attention U-Net architecture (Oktay et al. 2018), owning the advantage of segmenting dunes with multiple nuances. A simple plug-in tool for ArcGIS was built to encapsulate the model and manage the workflow automatically. It first transfers the to-be-segmented image data to the model from ArcGIS, and then pushes the editable segmentation results back to ArcGIS. During this process, geographers do not need to leave ArcGIS, and can revise the automatically-segmented images following the same visual interpretation routines of ArcGIS.

*These authors contributed equally.

Our interdisciplinary team has deployed **DCV^{2}I** with field surveys of geographers from a leading Chinese human geography research institution (Faculty of Geographical Science, Beijing Normal University). The deployment demonstrated: (1) **DCV^{2}I**'s deep vision model could deliver state-of-the-art performances in identifying dunes and labeling their ranges; (2) **DCV^{2}I**'s realized minimal interruptions to geographers' work routines and allowed them to use the tool they familiar with without touching any deep vision models, thus being highly appreciated by professional geographers. The deployment shows **DCV^{2}I**'s promising potential in supporting geographers' work. Moreover, it accelerated geographers' research related to drylands desertification which is among the most urgent economic, social and environmental challenges of sustainable development, thus having potential to support the growth of sustainable societies.

So far, **DCV^{2}I** has supported a group of eight geographers and played an essential role in one Ph.D. dissertation and one master thesis work, at least. Multiple research papers using it have been published or are in progress with prestigious geography journals such as *Geography and Sustainability* and *Land Degradation & Development*. We will further push its adoption in the geography community.

Related Work

Object Segmentation in Geography

Object segmentation is a classic problem in Geography (Castilla and Hay 2008). It groups neighboring pixels to form real-world geographic objects to reduce image complexity, making image content understandable and producing meaningful geographic objects for further analysis (Kucharczyk et al. 2020). Literature often falls into edge-based, region-based, and hybrid methods (Hossain and Chen 2019). Edge-based segmentation first detects edges (boundaries between objects) by capturing objects' geometrical and physical characteristics of objects (Ikonomopoulos 1982; Wang, Sun, and Chen 2015). For instance, hard-coded operators (Canny 1986) and fuzzy-based approaches (Trivedi and Bezdek 1986) are employed for edge detection. Then, these discontinuous edges are connected to form continuous segment edges, dealing with false edges or missing edges. Region-based segmentation starts within an object and expands towards the boundaries (Gaetano, Scarpa, and Poggi 2009; Wuest and Zhang 2009; Zhang et al. 2013; Chen et al. 2015), which contains two parts: merging and splitting. The merging process increases region size step by step based on the homogeneity criterion, and the splitting process splits the regions into sub-regions using the in-homogeneity criterion. Hybrid methods combine the strengths of both categories. Most studies in this line first employ the edge-based strategy, resulting in an image with over-segmentation. Subsequently, the region-based technique is used to merge similar segments (Akçay and Aksoy 2008; Zarrinpanjeh, Samadzadegan, and Schenk 2013; Zhang, Xiao, and Feng 2017). For instance, the tree Markov random field model was proposed in (Zhang, Xiao, and Feng 2017) to combine hierarchical segmentations, resulting in more homogeneous regions and precise edge localization.

Segmentation in Computer Vision

Meanwhile, image segmentation is arguably the most important yet challenging problem in computer vision due to its central role in visual perception. Segmentation tasks could be divided into three categories: instance, semantic, and panoptic segmentation (Kirillov et al. 2019). Extensive literature has focused on these three classes of image segmentation tasks and made considerable progress (Ghosh et al. 2019; Grady 2006; Hao, Zhou, and Guo 2020).

There are also fast-growing research interests in the computer vision community to build unified vision models capable of dealing with multiple tasks, including image segmentation. Such unified vision models are often trained in two ways. First, multiple tasks could be trained together to produce one model to deal with all training tasks without fine-tuning on each specific task (Lu et al. 2023; Yan et al. 2023; Zou et al. 2023a). Second, generalist training strategies have been proposed to enable models to handle new tasks in a zero-shot manner (Abdollahzadeh et al. 2023; Wang et al. 2023a). For example, SegGPT utilized the generalist Painter framework (Wang et al. 2023b) to realize in-context learning for image segmentation (Wang et al. 2023c).

Besides, interactive segmentation has attracted much attention, partially inspired by the success of large language models. It is the task of segmenting objects by interactively taking user inputs, e.g., clicks, boxes, polygons, and scribbles, as guidance for model refinements. There has been considerable progress since interactive segmentation is increasingly integrated with promptable design to enable better model performances and flexible integration with other systems. For example, SAM allows users to use points, boxes, and text, and encodes them to perform prompt-based learning to gain improved performances (Kirillov et al. 2023).

Summary

Segmentation has been a key research area in geography and computer vision but has different emphases. Research in geography focuses on applying segmentation algorithms in geographers' workflow, while CV research puts more effort into designing novel algorithms aiming at better performances. Unfortunately, there are often some significant lags for the geography community to apply the progress in computer vision, partially due to geographers' willingness to keep algorithms out of their work and their limited expertise to deal with algorithms by themselves.

DCV^{2}I in A Nutshell

Design Rationale

Our interdisciplinary team consists of AI researchers and geographers, giving us a unique opportunity to design a practical approach to integrate deep vision models into geographic tasks. Being *practical* means that the approach should be easy to use for professional geographers and bring utilities to them. While developing the **DCV^{2}I**, We had extensive discussions and exchanges and formulated the following three principles as the design rationale for **DCV^{2}I**.

1. *Interoperability.* **DCV^{2}I** should be able to interoperate with the tools frequently used by professional geographers, e.g., ArcGIS. We thus chose a data-centric mechanism in implementing the communication between deep vision models and geographers' tools using standard geographic data formats.
2. *Fully-encapsulated.* **DCV^{2}I**'s deep vision model should be fully encapsulated. We first attempted to design an interactive segmentation model but later gave up. We made such decisions because we learned that interactivity means interruptions for geographers. They have little motivation to learn and manage a prompt language, and are more willing to fix the errors in the automatic segmentation output by themselves rather than interacting with the model to improve the accuracy.
3. *Stability.* **DCV^{2}I** should deliver stable performances over dunes from multiple domains. Therefore, some state-of-the-art deep vision models may not be the best choices for practical use.

Guided by the above design rationale, **DCV^{2}I** was designed to consist of two key elements since only the model cannot fulfill the above principles. The first is the workflow for professional geographers to utilize the deep vision model in their dune segmentation task, while the second is the deep vision model itself. How they support the design principles will be introduced accordingly.

Workflow of DCV^{2}I

Fig. 1 describes the overall workflow of **DCV^{2}I**. From a geographer's view, it contains three stages. First, geographers may load the image into their work environment, e.g., ArcGIS, and choose to use **DCV^{2}I**. Then, there would be a few minutes for them to wait, depending on the complexity of the image to be segmented. Usually, after no more than a couple of minutes, the automatically segmented image will appear in the environment. They will check the segmentation of dunes in the image and make revisions if necessary.

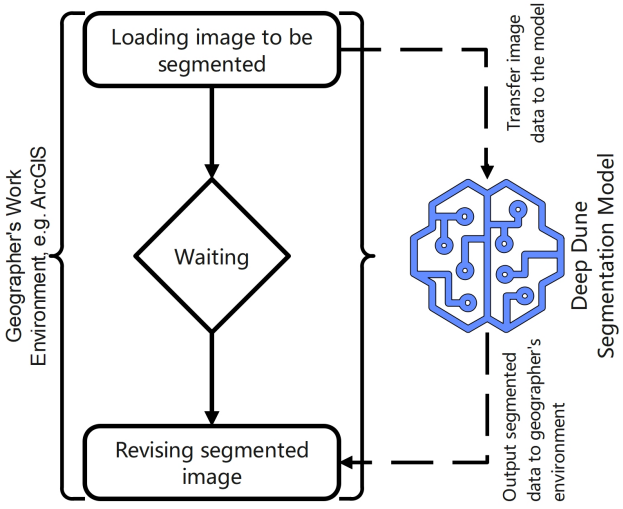

Figure 1: The workflow of **DCV^{2}I**.

In this workflow, the deep vision model is totally invisible to geographers. They do not need to know anything about the segmentation performed by the model. They only need to check the model's output in their ordinary work environments and make necessary fixes for any segmentation errors. The workflow is identical to the in-house GOBIA tools they are familiar with. Therefore, no cognitive overhead is required in adapting to the new model. Also, no interruption occurs. The second design rationale–*Fully-encapsulated.* is thus guaranteed. Meanwhile, using the standard data format facilities the bi-directional data communication between geographers' work environment and the model, which achieves high *Interoperability*.

Deep Dune Segmentation Model

We chose the U-Net (Ronneberger, Fischer, and Brox 2015) as the basic building block for deep dune segmentation. U-Net employs skip connections to merge high-level semantic features from the decoder with low-level semantic features from the encoder, forming a robust combination of features. However, relying solely on these skip connections was difficult to reduce the semantic gap caused by the absence of multi-scale features. This gap limited the ability of the network to extract the edge of the large object and the small object itself. Therefore, applying U-Net directly to dune morphology extraction could result in suboptimal performance. The dunes also often exhibited diverse morphological traits with slight differences. Furthermore, there are both large dunes and tiny dunes that have formed recently.

To address this problem, we proposed a dune semantic segmentation model which embeds the ASPP module (Chen et al. 2018) into the Attention U-Net (Oktay et al. 2018) architecture. The model mainly comprised an encoder module, an ASPP module, and a decoder module. Our specific network architecture is shown in Fig. 2, with detailed explanation in the following section. Compared with the emerging models in computer vision, our model delivered relatively stable and effective performance (please see the evaluations).

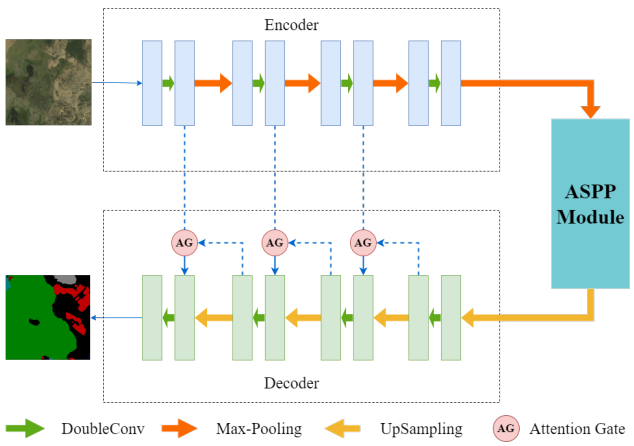

Figure 2: The deep dune segmentation model's architecture.

Encoder We employed VGG16 (Simonyan and Zisserman 2014) as our model's encoder module. To facilitate the implementation of parameter transfer, we removed the fully connected layer of VGG16 and retain part of the convolutional layers. During training, we loaded the pre-trained parameters of VGG16 (pre-trained on the VOC (Everingham et al. 2010) dataset) to improve the generalization ability and reduced the training cost.

ASPP In **DCV^{2}I**, the ASPP module was used in the bottleneck part to extract multi-scale features from the high-level feature maps. It connected four atrous convolution layers and a global average pooling layer in parallel, followed by a 1×1 convolution operation. Each atrous convolution layer consisted of a convolution operation, a Batch Normalization (BN) layer, and a ReLU activation function. The ASPP module overcomed the drawbacks of local information loss caused by the grid effect and the lack of correlation in distant information. In this way, it obtained features of different scales without pooling layers.

Decoder While the ASPP module effectively captured features at multiple scales and the fine details of dune images, it also introduced some irrelevant information, such as noise during upsampling. To mitigate the impact of irrelevant image features on target features and enhance dune segmentation accuracy, we introduced the attention mechanism (AG) (Oktay et al. 2018) within the decoding phase, similar to the Attention U-Net.

Loss Function The imbalance of the dune type distributions often led to easily distinguishable negative samples accounting for the majority of the overall loss when using the traditional standard cross-entropy loss function for model training. So we used Focal Loss (Lin et al. 2017) to mitigate such a problem of low model accuracy caused by the afore-mentioned imbalance.

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{1}$$

where α is the category weight parameter, $(1-p_t)^{\gamma}$ is the moderator. The value of α depended on the proportionality of positive and negative samples in the dataset, and the shared weights of positive and negative samples on the total loss were controlled by the introduction of α. A smaller α reduces the weight of the category accounting for a high proportion.

Field Surveys

Background and Problem

Background As we mentioned, our team iwas formed by the joint efforts of AI researchers and geographers. The geographers are with the Faculty of Geographical Science, Beijing Normal University, which ranked as A+[1] in the fourth round of national discipline assessment. As a leading research institution, one of its main directions is combating desertification which requires a deep understanding of the evolution of dunes in deserts. Every year, they organize multiple field surveys of the deserts in China to collect dune data, often by performing field investigations in areas of remote sensing images or taking aerial photographs using UAVs designed for geographical surveys.

Problem Each field survey resulted in a large amount of longitudinal image data which was impossible to be analyzed manually. After multiple years, the scale of data became even larger, containing images of millions of dunes. Such data provided opportunities for geographers to understand the life cycle of dunes by analyzing how dunes form, grow, and survive over time. Conventionally, they only focused on a handful of dunes or the rough overall estimations and statistics of the target area. By performing visual interpretation manually, they derived insights about the evolution of these dunes, e.g., Dong et al. (2023); Guan et al. (2022); Wang et al. (2007). However, such a process only used a very small part of the collected data, wasting most data and the efforts for collecting them. It also restricted the ability to develop a precise view of dune systems without losing details and to discover novel patterns in dune evolution.

Geographers' problems could be solved by automated dune segmentation tools that free them from manual efforts. The native GOBIA tools in their work environment often fail to deal with such data since their algorithms often could only deal with regular objects at a large scale. Most deep vision models were neither integrated nor tested in the problem domain. Therefore, approaches such as **DCV^{2}I** could be a potential solution. In the rest of this section, we would take a field survey as an example, while segmenting data collected from this survey serves **DCV^{2}I**'s first use case.

Target Area and Data Sources

Location and Natural Environment of the Target Area As one of China's four major sandy lands, the Mu Us Sandy Land is located in the agricultural and pastoral transitional zone in northwest China, spanning the geographic coordinates of 37°27'–40°22'N and 106°20'–111°30'E. Situated on the Ordos Plateau, it serves as a transitional area between the Ordos Plateau and the Loess Plateau in northern Shaanxi, without a clearly defined geographical boundary. The region spans across Inner Mongolia Autonomous Region, Shaanxi Province, and Ningxia Hui Autonomous Region.

The Mu Us Sandy Land is characterized by a fragile and sensitive ecological environment (Wang et al. 2022). The predominant land surface cover in the entire region is grassland and sandy areas, making it an important ecological barrier in northern China. Over thousands of years of human activities, such as reclamation, logging, excessive grazing, and other forms of disturbance, the natural vegetation of the sandy land has nearly disappeared. The three major vegetation communities in this region include: grassland and shrub vegetation on elevated ridges and terraces; woody shrubs and sand-fixing vegetation on semi-fixed and fixed dunes and sandy land; grasslands, salt-tolerant, and swamp vegetation on floodplains and marshlands.

The Mu Us Sandy Land's unique geographical location and natural environment make it a typical region for study-

[1] A+ is the highest rank, only two institutions nationwide were awarded this rank in every discipline.

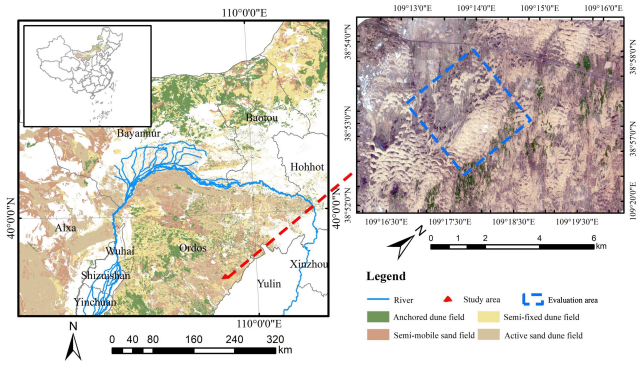

Figure 3: The target area of the field survey.

ing regional aeolian geomorphology and desertification in northern China. There have been multiple field surveys happened in this area. The geographers in our team conducted their most recent field survey of this area in 2020 due to the travel complexities caused by the COVID-19 pandemic. We selected this area as the study experimental area. The geographical coordinates of the field survey are approximately between 109°13' to 109°20'E and 38°52' to 38°58'N. The location of the target area is depicted in the following map in Fig. 3.

Data Sources and Field Survey Our remote sensing image came from Pleiades satellites with a resolution of 0.5 meters. The satellite image within our study area was captured in April 2020. Orthorectification and geometric correction are further performed. In the process, the selected control points were evenly distributed throughout the entire area (the geometric centers of special vegetation, typical trees, etc.), with a particular increase in the number of control points around sand dunes.

In 2020, field surveys were also conducted to assess the actual types and distribution of dunes accurately. During the field survey process, the geographical coordinated of ground sample points were recorded using GPS. The actual sand dune types at each sample point and specific information about the surrounding environment were documented. The data collected during the survey helped gain a preliminary understanding of the spatial distribution characteristics of dunes and to train our deep semantic segmentation model.

Figure 4: The field explorations conducted by geographers.

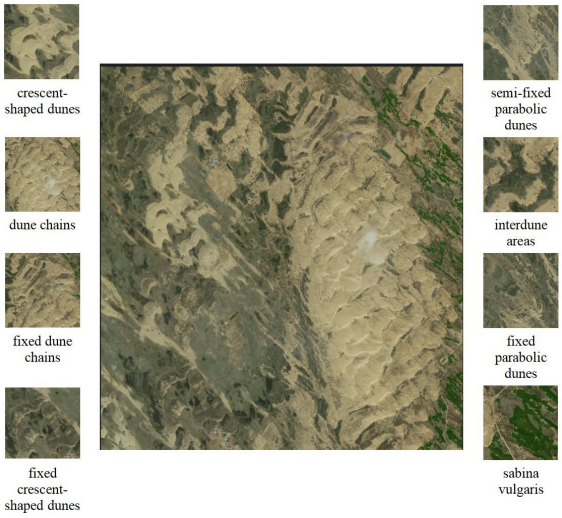

Figure 5: Dune types in the target area.

The on-site field exploration is depicted in Fig. 4.

In this area, there are eight main dune morphological types, namely: 1) crescent-shaped dunes; 2) dune chains; 3) fixed dune chains; 4) fixed crescent-shaped dunes; 5) semi-fixed parabolic dunes; 6) fixed parabolic dunes; 7) interdune areas; 8) Sabina vulgaris. The dune types of the target area are shown in Fig. 5.

Evaluating DCV^{2}I with Deployment

The evaluations to **DCV^{2}I** were conducted from two perspectives. The first focused on the model performances, which followed the routines of evaluating deep learning systems. The second was user-centric, focusing on geographers' experiences in using it to deliver dune segmentation. Code and data used in evaluations are available at: https://doi.org/10.6084/m9.figshare.24564874.v1.

Model Performance Evaluation

Dataset We constructed a new sand image dataset after our geographers concluded the aforementioned field survey of the Mu Us Sandy Land in a four-step process. First, we obtained the target area's satellite images using ArcGIS by clipping the original satellite images. Then, our geographers used ArcGIS's ArcMap Editor to draw boundaries of dunes in the images, and also labeled the categories of dunes through manual visual interpretation. The third step cropped the large-size images into small ones.Due to the substantial size of the original images, some of which exceeded dimensions of 10,000 $\times$ 10,000 pixels, directly inputting them into the deep learning model was impractical. Therefore, we sliced the original images into 512 $\times$ 512 pieces for more efficient training and testing. We employed the sliding window to achieve precise cropping of the image. In the last step, data augmentation techniques (sliding window with different repetitions, image flipping and random rotation) were used to enhance the datasets. Doing so reduced our geographers' manual burdens in constructing the dataset. The

above process resulted in a dataset containing 1,537 dune images where all dune boundaries were marked. Concerning the distribution of dunes, the original images featured 428 dunes of various sizes and types, including 33 crescent-shaped dunes, 30 dune chains, 14 fixed dune chains, 31 fixed crescent-shaped dunes, 34 semi-fixed parabolic dunes, 74 interdune areas, 58 fixed parabolic dunes, and 154 sabina vulgaris.

Experiment Settings We randomly divided evaluation area data into three subsets: the training set (1,211 images), the validation set (135 images), and the testing set (191 images). Due to the limited training data, the training process did not start from scratch but adopted the official pre-trained weights on VOC data. In the first 100 training epochs, the backbone network was frozen, and the other network was fine-tuned using training data, where the batch size was 8 with an SGD optimizer. In the following 100 training epochs, the whole network was trained together, and the batch size was 4. The above process' learning rate was $1e$-4, and weight decay was $1e$-4.

Comparison Methods We systematically compared our model with the following state-of-the-art image segmentation methods, including FCN (Long, Shelhamer, and Darrell 2015), U-Net (Ronneberger, Fischer, and Brox 2015), SegNet (Badrinarayanan, Kendall, and Cipolla 2017) and DeepLabV3+ (Chen et al. 2018).

- **FCN**. FCN is a convolutional neural network architecture for the semantic segmentation of images. It enables end-to-end pixel-level prediction by converting a fully connected layer into a convolutional layer.
- **U-Net**. U-Net is a CNN-based image segmentation network mainly used for medical image segmentation. It performs well on small datasets, partially due to its effective architectural design and skip-connect application.
- **SegNet**. SegNet is an encoding-decoding architecture based on CNNs, which has been used in areas such as medical image segmentation and autonomous driving.
- **DeepLabV3+**. DeepLab V3+ introduces spatial pyramid pooling (ASPP) and multiscale prediction mechanisms to capture detailed information and process multiscale features. This has led to good results in tasks such as medical image and natural scene segmentation.

Note that the native GOBIA tools were not included in comparisons since their practical performances were very limited. For our dataset, they often failed to provide any meaningful results. Four standard metrics were used in the evaluations: accuracy, F1 Score, mPA, and MIoU, at the pixel-level. We also compared the segmentation results produced by the different models qualitatively.

Model Performances & Segmentation Results Table 1 shows the comparison results on our dataset of all the models. Our deep dune segmentation model exhibited significant improvements across various metrics. It achieves accuracy, F1 Score, mPA, and MIoU values of 96.72%, 72.96%, 80.67%, and 75.09%, respectively. Compared to the traditional U-Net algorithm and DeepLab V3+ algorithm, the proposed model outperformed them by 3.75%, 5.1%, 2.73%, 2.18%, and 3.94%, 5.54% 3.02%, 2.2% in accuracy, F1 Score, mPA, and MIoU, respectively.

Models	Accuracy	F1 Score	mPA	MIoU
FCN	88.72	45.64	70.27	62.48
SegNet	91.81	59.48	75.38	68.37
DeepLab V3+	92.78	67.42	77.65	72.89
U-Net	92.97	67.86	77.94	72.91
Ours	**96.72**	**72.96**	**80.67**	**75.09**

Table 1: Accuracy (%) Results in Evaluation Area.

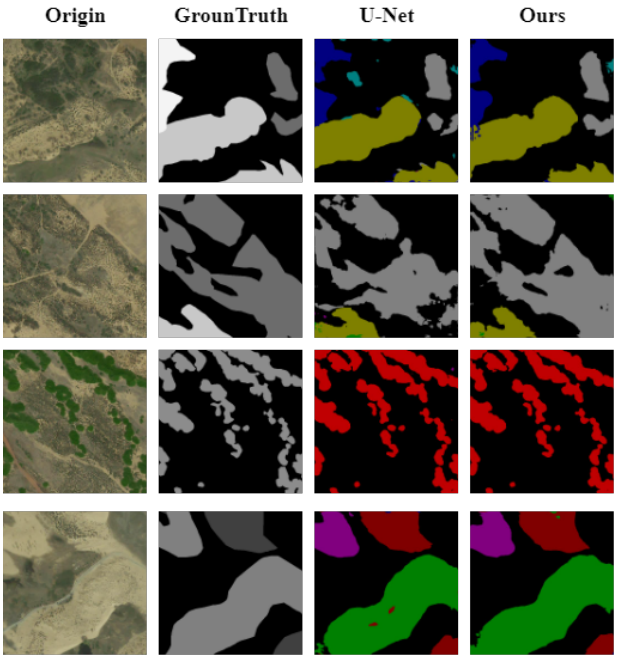

Figure 6: Actual segmentation of different models. We randomly selected four images for comparison. The first column shows the original dune image, the second shows the labeled image, the third shows the segmentation result using U-Net, and the last shows the segmentation result using our method, where different colors mark different dune classes, e.g., red for sabina vulgaris, etc.

In addition to these metrics, the effectiveness of image segmentation was also important. The comparisons are presented in Fig. 6 intuitively. Our model exhibited much better segmentation performances, highlighting the dune regions and minimizing misclassifications. In contrast, U-Net incurs significant misclassifications. For instance, in the first and fourth rows, U-Net inaccurately labeled some background portions as dunes. The edge detailing in the U-Net model also fell short compared with our approach, as evident from the segmentation results in the second row.

Models	Accuracy	F1 Score	mPA	MIoU
a	92.97	67.86	77.94	72.91
b	95.41	70.99	80.21	74.62
c	94.78	69.92	79.65	74.89
Ours	**96.72**	**72.96**	**80.67**	**75.09**

Table 2: Ablation (%) Results in Evaluation Area.

Ablation We used the same dataset to validate the model's enhancements in relative to the baseline U-Net model through comparative and ablation experiments. The investigation primarily involved the incorporation of the ASPP and Attention modules. The network architecture proposed in this study is assessed against three configurations: a) the basic U-Net network, b) the U-Net network integrated with the Attention mechanism, and c) the U-Net network enhanced with the ASPP module. The outcomes were juxtaposed against the algorithmic outcomes presented in this paper, as summarized in Tab. 2. The integration of distinct modules contributed positively to the overall performances. Notably, the augmentation achieved by combining the baseline with both the Attention and ASPP modules surpassed the enhancements seen in cases where only the baseline + ASPP or baseline + Attention modules were employed. This observation underscores the efficacy of our model.

Formative Evaluation with Users

While the proposed deep dune segmentation model achieved good performances, **DCV^{2}I**'s practical value was still determined by the acceptance from its users, i.e., professional geographers. We further qualitatively evaluated **DCV^{2}I** by observing its usage in two professional geographers' work and collecting their feedback. We had the following insights.

First, we found that there was minimal learning barrier for professional geographers to use **DCV^{2}I**. Both geographers could independently use **DCV^{2}I** after a 5-minute tutorial video without any difficulty. Second, the geographers were generally satisfied with the segmentation results; they told us our model's results were much better than the native automated segmentation tools in ArcGIS.The percentage of segments that were used as-is was consistently over 95%. We found geographers were more tolerant of minor imperfections. They viewed such imperfections as a natural part of visual interpretation. For them, these imperfections may be not errors but just ambiguities. Third, **DCV^{2}I**'s workflow provided great conveniences to geographers' work by allowing them to perform all their jobs in ArcGIS. The feature of pushing editable segmentation results back to ArcGIS for further revising (see Fig. 7) was highly acclaimed. Both geographers exhibited strong interest to continue using **DCV^{2}I** and introduce it to their colleagues. By 08/2023, images (including remote-sensing and UAV's) of over 13,000 square km (>30%) of the entire Mu Us Sandy Land had been segmented by **DCV^{2}I**, much more than ten times of all the areas manually segmented by the entire geographer team done in the last 5 years. The benefits are thus at the magnitude level.

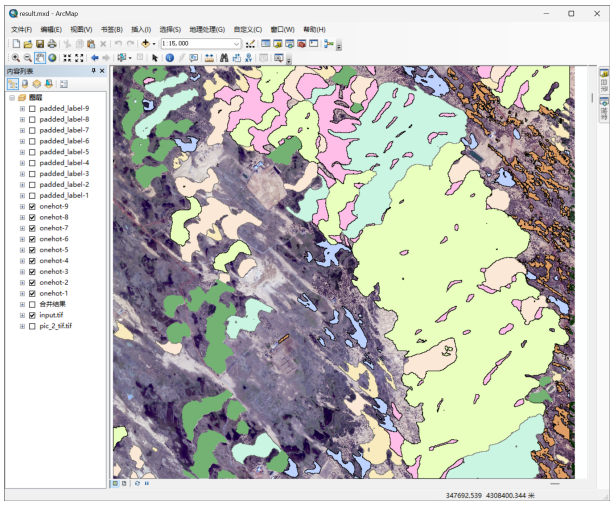

Figure 7: Editable segmentation results in ArcGIS.

Besides, after the SAM (Kirillov et al. 2023) was introduced in early 2023, the same geographers were asked to perform dune segmentation tasks using SAM again. Unfortunately, none of them could deliver satisfying results with SAM after several dozen interactions. For them, learning how to provide effective prompts could be very challenging, particularly when their tasks were in the specific dune segmentation domain. Meanwhile, they felt frustrated for not allowing them directly revise segmentation results. This indicated that the generalist's interactive models might still be premature for the tasks.

For the time taken for without **DCV^{2}I**, we found that a geographer with moderate experience usually spent a couple of days to finish the task, while using **DCV^{2}I** could reduce the time to less than an hour. Note that manually segmenting dunes was boring and tiring, so a geographer could only work 3-4 hours effectively per day. However, the waiting period (see Fig. 1) for **DCV^{2}I**, which is a limitation, is still too long for personal computers, even though it runs quite fast on workstations. For example, an image of 200M pixels usually takes 10-15 minutes to run before outputting the segmentation to ArcGIS on a laptop with RTX 4060 GPU and 8G memory. While a progress bar could partially reduce geographers' anxiety in waiting, accelerating the model's speed in personal computing devices shall be considered to further improve user experiences. Anyway, the waiting time about 10-15 minutes is still much faster than the native segmentation tool in ArcGIS, which usually takes forever to run on our geographer colleagues' machines.

Summary & Lesson Learned

The above evaluations confirmed that **DCV^{2}I** fulfilled its goals of being a practical approach to support geographers' visual interpretation in dune segmentation. The model recorded outstanding performances, and geographers highly appreciated the workflow design. The design and deployment of **DCV^{2}I** offer valuable lessons to artificial intelligence researchers interested in building interdisciplinary ap-

plications of deep vision models. The most important lesson is that *users' work routines must be honored*. I.e., a successful AI application should minimize the interruptions to their workflow, and empower their familiar work environments and tools rather than replace them.

Concluding Remarks

This paper presents **DCV^{2}I**, a practical approach to using a deep vision model to support geographers in performing visual interpretation in dune segmentation tasks efficiently. **DCV^{2}I** was designed to achieve high *interoperability* with *fully-encapsulated* dune segmentation model to deliver *stable* performances. These design principles guided our interdisciplinary teams to make a series of critical design decisions of **DCV^{2}I**, which features a workflow design and a deep dune segmentation model. The deep dune segmentation model was hidden in the workflow, allowing geographers to work in their routine work environments, such as ArcGIS. Instead of interactive segmentation, **DCV^{2}I** used the deep dune segmentation model to produce a reasonably good segmentation and guarantee the freedom of revising it to geographers with tools they are familiar with. Therefore, **DCV^{2}I** realized a non-invasive change to geographers' visual interpretation routines, reducing their manual efforts while incurring minimal interruptions to their work routines. Our extensive evaluations demonstrated its high utility and usability. As an outcome of interdisciplinary efforts, **DCV^{2}I** have exhibited significant potential and helped generate a number of research outputs such as dissertations and research papers. We plan to further push its adoption in the geography community.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China under grants 62076232 and 62172049. Corresponding author: Wei Wang.

References

Abdollahzadeh, M.; Malekzadeh, T.; Teo, C. T. H.; Chandrasegaran, K.; Liu, G.; and Cheung, N.-M. 2023. A Survey on Generative Modeling with Limited Data, Few Shots, and Zero Shot. arXiv:2307.14397.

Akçay, H. G.; and Aksoy, S. 2008. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE transactions on Geoscience and Remote Sensing*, 46(7): 2097–2111.

Badrinarayanan, V.; Kendall, A.; and Cipolla, R. 2017. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481–2495.

Blaschke, T.; Hay, G. J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R. Q.; Van der Meer, F.; Van der Werff, H.; Van Coillie, F.; et al. 2014. Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87: 180–191.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679–698.

Castilla, G.; and Hay, G. 2008. Image objects and geographic objects. In *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*, 91–110. Springer.

Chen, B.; Qiu, F.; Wu, B.; and Du, H. 2015. Image segmentation based on constrained spectral variance difference and edge penalty. *Remote Sensing*, 7(5): 5980–6004.

Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; and Adam, H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, 801–818.

Dong, Y.; Fu, S.; Zhang, S.; and Hasi, E. 2023. Type, Distribution, Formation and Evolution of Coastal Aeolian Dunes. In *Sand Dunes of the Northern Hemisphere*, 163–178. CRC.

Everingham, M.; Gool, L. V.; Williams, C. K. I.; Winn, J.; and Zisserman, A. 2010. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2): 303–338.

Fang, W.; Wang, C.; Chen, X.; Wan, W.; Li, H.; Zhu, S.; Fang, Y.; Liu, B.; and Hong, Y. 2019. Recognizing global reservoirs from Landsat 8 images: A deep learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9): 3168–3177.

Gaetano, R.; Scarpa, G.; and Poggi, G. 2009. Hierarchical texture-based segmentation of multiresolution remote-sensing images. *IEEE Transactions on geoscience and remote sensing*, 47(7): 2129–2141.

Ghosh, S.; Das, N.; Das, I.; and Maulik, U. 2019. Understanding deep learning techniques for image segmentation. *ACM computing surveys (CSUR)*, 52(4): 1–35.

Grady, L. 2006. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11): 1768–1783.

Guan, C.; Hasi, E.; Yang, Y.; Sun, Y.; and Du, H. 2022. Determinants and dynamics of blowouts in Hulun Buir sandy grassland, Inner Mongolia, China from 1959 to 2018. *Earth Surface Processes and Landforms*, 47(11): 2676–2694.

Hao, S.; Zhou, Y.; and Guo, Y. 2020. A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406: 302–321.

Hossain, M. D.; and Chen, D. 2019. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150: 115–134.

Hua, Y.; Marcos, D.; Mou, L.; Zhu, X. X.; and Tuia, D. 2022. Semantic Segmentation of Remote Sensing Images With Sparse Annotations. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.

Ikonomopoulos, A. 1982. An approach to edge detection based on the direction of edge elements. *Computer Graphics and Image Processing*, 19(2): 179–195.

Kirillov, A.; He, K.; Girshick, R.; Rother, C.; and Dollár, P. 2019. Panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '19)*, 9404–9413.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643*.

Kucharczyk, M.; Hay, G. J.; Ghaffarian, S.; and Hugenholtz, C. H. 2020. Geographic object-based image analysis: a primer and future directions. *Remote Sensing*, 12(12): 2012.

Lillesand, T.; Kiefer, R. W.; and Chipman, J. 2015. *Remote Sensing and Image Interpretation*. John Wiley & Sons.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

Lu, J.; Clark, C.; Zellers, R.; Mottaghi, R.; and Kembhavi, A. 2023. Unified-IO: A unified model for vision, language, and multi-modal tasks. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR '23)*.

Manley, E.; Filomena, G.; and Mavros, P. 2021. A spatial model of cognitive distance in cities. *International Journal of Geographical Information Science*, 35(11): 2316–2338.

Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; et al. 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Rose, G. 2003. On the need to ask how, exactly, is geography "visual"? *Antipode*, 35(2): 212–221.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tolia-Kelly, D. P. 2012. The geographies of cultural geography II: Visual culture. *Progress in Human Geography*, 36(1): 135–142.

Trivedi, M. M.; and Bezdek, J. C. 1986. Low-level segmentation of aerial images with fuzzy clustering. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(4): 589–598.

Trotter, C. M. 1991. Remotely-sensed data as an information source for geographical information systems in natural resource management a review. *International Journal of Geographical Information System*, 5(2): 225–239.

Wang, M.; Sun, Y.; and Chen, G. 2015. Refining high spatial resolution remote sensing image segmentation for man-made objects through a collinear and ipsilateral neighborhood model. *Photogrammetric engineering & Remote sensing*, 81(5): 397–406.

Wang, W.; Zhang, G.; Han, H.; and Zhang, C. 2023a. Correntropy-Induced Wasserstein GCN: Learning Graph Embedding via Domain Adaptation. *IEEE Transactions on Image Processing*, 32: 3980–3993.

Wang, X.; Eerdun, H.; Zhou, Z.; and Liu, X. 2007. Significance of variations in the wind energy environment over the past 50 years with respect to dune activity and desertification in arid and semiarid northern China. *Geomorphology*, 86(3-4): 252–266.

Wang, X.; Song, J.; Xiao, Z.; Wang, J.; and Hu, F. 2022. Desertification in the Mu Us Sandy Land in China: Response to climate change and human activity from 2000 to 2020. *Geography and Sustainability*, 3(2): 177–189.

Wang, X.; Wang, W.; Cao, Y.; Shen, C.; and Huang, T. 2023b. Images speak in images: A generalist painter for in-context visual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '23)*, 6830–6839.

Wang, X.; Zhang, X.; Cao, Y.; Wang, W.; Shen, C.; and Huang, T. 2023c. Seggpt: Segmenting everything in context. *arXiv preprint arXiv:2304.03284*.

Wiley, V.; and Lucas, T. 2018. Computer vision and image processing: a paper review. *International Journal of Artificial Intelligence Research*, 2(1): 29–36.

Wuest, B.; and Zhang, Y. 2009. Region based segmentation of QuickBird multispectral imagery through band ratios and fuzzy comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1): 55–64.

Yan, B.; Jiang, Y.; Wu, J.; Wang, D.; Luo, P.; Yuan, Z.; and Lu, H. 2023. Universal instance perception as object discovery and retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '23)*, 15325–15336.

Zarrinpanjeh, N.; Samadzadegan, F.; and Schenk, T. 2013. A new ant based distributed framework for urban road map updating from high resolution satellite imagery. *Computers & Geosciences*, 54: 337–350.

Zhang, X.; Xiao, P.; and Feng, X. 2017. Toward combining thematic information with hierarchical multiscale segmentations using tree Markov random field model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131: 134–146.

Zhang, X.; Xiao, P.; Song, X.; and She, J. 2013. Boundary-constrained multi-scale segmentation method for remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 78: 15–25.

Zou, X.; Dou, Z.-Y.; Yang, J.; Gan, Z.; Li, L.; Li, C.; Dai, X.; Behl, H.; Wang, J.; Yuan, L.; et al. 2023a. Generalized decoding for pixel, image, and language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '23)*, 15116–15127.

Zou, X.; Yang, J.; Zhang, H.; Li, F.; Li, L.; Gao, J.; and Lee, Y. J. 2023b. Segment everything everywhere all at once. *arXiv preprint arXiv:2304.06718*.