Literature Review

Project title: Integration of Deep Learning (DL) in Geospatial Object-Based Image Analysis (GEOBIA) for Image Segmentation in Remote Sensing Tasks

Your name: Vivek Sahukar ASU ID: 1230067360

Step 1: Survey relevant work [10 points]

1. Du, S., Du, S., Liu, B., & Zhang, X. (2021). Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high-resolution remote sensing images. International Journal of Digital Earth, 14(3), 357-378.
2. Herlawati, H., Handayanto, R. T., Atika, P. D., Sugiyatno, S., Rasim, R., Mugiarso, M., ... & Purwanti, S. (2022, December). Semantic Segmentation of Landsat Satellite Imagery. In 2022 Seventh International Conference on Informatics and Computing (ICIC) (pp. 1-6). IEEE.
3. Song, A., Kim, Y., & Han, Y. (2020). Uncertainty analysis for object-based change detection in very high-resolution satellite images using deep learning network. Remote Sensing, 12(15), 2345.
4. Luo, C., Li, H., Zhang, J., & Wang, Y. (2023, July). OBViT: A high-resolution remote sensing crop classification model combining OBIA and Vision Transformer. In 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) (pp. 1-6). IEEE.
5. Wang, J., Zheng, Y., Wang, M., Shen, Q., & Huang, J. (2020). Object-scale adaptive convolutional neural networks for high-spatial resolution remote sensing image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 283-299.
6. Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2018). An object-based convolutional neural network (OCNN) for urban land use classification. Remote sensing of environment, 216, 57-70.
7. Zaabar, N., Niculescu, S., & Kamel, M. M. (2022). Application of convolutional neural networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in Ain Témouchent, Algeria. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 5177-5189.
8. Hénaff, O. J., Koppula, S., Alayrac, J. B., Van den Oord, A., Vinyals, O., & Carreira, J. (2021). Efficient visual pretraining with contrastive detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10086-10096).
9. Ibrahim, A., & El-kenawy, E. S. M. (2020). Image segmentation methods based on superpixel techniques: A survey. Journal of Computer Science and Information Systems, 15(3), 1-11.

Step 2: Summary of relevant work [20 points]

Zaabar, N., Niculescu, S., & Kamel, M. M. (2022). Application of convolutional neural networks with object-based image analysis for land cover and land use mapping in coastal areas: A case study in Ain Témouchent, Algeria. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 5177-5189.

Brief summary:

- The paper uses a method that combines CNN with OBIA for the Land Use and Land Cover Classification (LULC) task for two datasets: Sentinel-2 and Pléiades Imagery.
- Custom CNN models are used to extract the heatmaps from the images, which were later utilized as input features to perform the OBIA.
- The results for the proposed method were then compared with two ML methods viz. RF and SVM.

Strengths:

- The paper uses a novel approach to integrate CNN into OBIA methods to avoid doing pixel-level classification by traditional CNN models.
- The paper does a fair evaluation of the proposed method by comparing it to other pixel-based and OBIA methods across different land cover type classes (in the dataset used). The paper explains the data preprocessing steps in a detailed manner.
- OBIA methods are better suited for high-resolution (<10 m) datasets. The paper used high-resolution datasets (2m and 10 m), which made the comparison of different existing methods with the proposed method more appropriate.

Limitations:

- The proposed method still uses the different scale parameter values obtained through trial-and-error in the OBIA step, which defeats the purpose of integrating DL into OBIA to obviate the need for manual parameter selection in OBIA methods.
- It would have been more beneficial to use existing DL models that are better suited for segmentation tasks such as Unet and DeepLab rather than custom designing the CNN.
- The paper uses two different CNN models for each of the two datasets. This raises questions about the proposed method's adaptability to other similar datasets: Do the CNN models need to be designed each time a different dataset is used?
- The paper could have benefited from implementing and using the CNN model in Python for better reproducibility and open-source support.

Luo, C., Li, H., Zhang, J., & Wang, Y. (2023, July). OBViT: A high-resolution remote sensing crop classification model combining OBIA and Vision Transformer. In 2023 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) (pp. 1-6). IEEE.

Brief summary:

- The paper uses SLICO (Zero Parameter Version of Simple Linear Iterative Clustering) segmentation methods to produce superpixel objects from remote-sensing images with similar shapes and close areas.
- Then, ViT (Vision Transformer) is used for superpixel classification.
- Finally, an object-based K-nearest neighbor filtering algorithm is used as a post-processing method to reduce the pretzel phenomenon ("prediction of incorrect superpixel objects leads to many spots in the predicted image").

Strengths:

- The paper uses SLICO instead of SLIC for image segmentation, obviating the need for selecting tightness measure 'm' subjectively.
- The paper employs the innovative use of Vision Transformer (ViT) instead of CNN, which is better than CNN in working with different sizes of superpixels and adapting to small-size images.
- Post-processing the model's prediction results using an object-based K-nearest neighbor filtering algorithm reduces the "pretzel phenomenon."

Limitations:

- The main aim of integrating DL methods in OBIA is to obviate the need for manual hyperparameter selection in the segmentation stage. However, the user still has to choose the parameters for SLICO (or any other superpixel algorithm) and object-based K-nearest neighbor algorithms.
- The paper does not mention if the same parameters are used by the SLICO algorithm for superpixel segmentation.
- The paper does not discuss what methods are used to stitch predicted superpixels to obtain final segmentation results.
- Results are reported using metrics viz. Precision, Recall, and F1-score: the proposed model's performance needs to be compared with existing OBIA methods by reporting metrics appropriate for image semantic segmentation tasks viz. Dice Score and IoU (Intersection over Union).

Herlawati, H., Handayanto, R. T., Atika, P. D., Sugiyatno, S., Rasim, R., Mugiarso, M., ... & Purwanti, S. (2022, December). Semantic Segmentation of Landsat Satellite Imagery. In 2022 Seventh International Conference on Informatics and Computing (ICIC) (pp. 1-6). IEEE.

Brief summary:

- The paper uses DL-based model viz. DeepLabV3+ for semantic segmentation for the land cover classification task.
- The paper compares the performance of DeepLabV3+ (by reporting the accuracy metric) against Iterative Self-Organizing Clustering (ISOCLUST) and OBIA. The accuracy of DeepLabV3+ was 95%, which is higher than OBIA (accuracy of 80%) but lower than ISOCLUST (accuracy of 100%). However, the paper still recommends using DeepLabV3+ due to its much less computation time compared to ISOCLUST.
- The paper uses MATLAB to run the DeepLabV3+ model.

Strengths:

- The paper clearly mentions splitting the dataset into training, validation, and testing to avoid class imbalance and data leakage.
- To make a fair comparison, the paper reports both computational time and performance metric (accuracy) for all three methods.
- The paper uses the MATLAB App Designer to facilitate the run of the code for the DL model for users who lack coding expertise.

Limitations:

- The paper boldly claims that, unlike ISOCLUST and OBIA, DL methods do not need experts in remote sensing and GIS (geographic information systems), which is not entirely true. We still need experts to evaluate the results of these black-box models.
- Only accuracy is used as a metric for comparing semantic segmentation performance. Other metrics more relevant for such tasks viz. Precision, Recall, F1 score, Dice Score, and IoU need also to be reported to do a fair evaluation.
- The paper could have benefited from implementing and using DeepLabv3+ in Python for better reproducibility and open-source support.

Step 2: Organization of relevant work [20 points]

What are you trying to do? Articulate your objectives using no jargon. [6 points]

My goal is to use an end-to-end learning model that involves both optimizing SLIC hyperparameters (where gradients cannot be calculated) and training a CNN for superpixel classification (where gradients can be calculated). For this I will use a hybrid approach that combines gradient-free optimization for SLIC parameters and gradient-based optimization for the CNN. The hybrid optimization strategy is as follows:

1. SLIC Hyperparameter Optimization (Gradient Free): Use Nevergrad (Facebook's library) or any other gradient-free optimization library to optimize the SLIC hyperparameters. This process is iterative and involves evaluating the performance of the entire pipeline (SLIC + CNN) for each set of hyperparameters proposed by Nevergrad.
2. CNN Optimization (Gradient-Based): Use standard gradient-based optimization techniques, such as SGD or Adam, to adjust the weights of the network based on the loss calculated from the classification performance on the superpixel segments.

The rough step-by-step end-to-end training process is as follows:

- Iteration Cycle: The training process alternates between optimizing the SLIC hyperparameters using Nevergrad and updating the CNN weights using gradient descent.
- Outer Loop (Nevergrad optimization):
    - Nevergrad proposes a new set of SLIC hyperparameters.
    - With these parameters, perform SLIC segmentation on your training images.
    - Inner Loop (Gradient Descent for CNN):
        - Feed resulting superpixels into CNN
        - Compute the loss based on the CNN's classification performance (e.g., cross-entropy loss for classification tasks). Use the modified version of pixel-wise loss to adapt it for superpixel objects.
        - Update the CNN weights using backpropagation and gradient descent
    - Use the performance metric (Dice score/IoU) as the objective for Nevergrad to propose a new set of parameters in the next iteration.

Within each iteration of the Nevergrad optimization loop, the CNN is trained for several epochs or until convergence. After each full cycle of CNN training, the overall pipeline performance (segmentation + classification) is evaluated to guide the next set of hyperparameters proposed by Nevergrad.

How is it done today, and what are the limits of current practice? [7 points]

Today, DL and OBIA methods are used separately, mostly for semantic segmentation tasks. The reason is that OBIA is an object-based method, and DL is pixel-based. However, both methods have their own pros and cons. OBIA consists of two steps: segmentation (grouping of pixels into super-pixels to generate “objects”) and classification (classifying the objects obtained from the previous steps into different required classes). OBIA offers advantages by incorporating spatial and spectral context, reducing noise, and better reflecting real-world structures. However, OBIA's segmentation stage often involves manual feature selection and parameter tuning, making it subjective and time-consuming. While deep learning (DL) excels at automatic feature extraction, its integration within OBIA workflows remains limited. This is due to challenges in adapting pixel-centric DL segmentation models to object-based frameworks, concerns about computational overhead, and the lack of clear advantages of DL over traditional machine learning techniques in terms of accuracy for OBIA classification.

There has been some limited work in integrating the DL methods into the OBIA workflow. Most of the methods still require manual parameter selection for the segmentation algorithm (such as SLIC) used for the segmentation stage of the OBIA, which makes integrating DL into OBIA less useful. The current work does not use more useful performance metrics for image segmentation tasks viz. Dice Score and IoU (Intersection over Union) make the fair evaluation of their proposed methods difficult. Moreover, custom CNN models are used with the OBIA analysis, which limits the adaptability of the proposed method for the general task, as using a different DL model for different datasets is not a very efficient and scalable approach.

What is new in your approach, and why do you think it will be successful? [7 points]

I would use segmentation algorithms apart from SLIC to evaluate which one gives the best results for superpixel segmentation. I would use a modified version of the loss function to calculate the object-level loss for superpixel. In my approach, the DL model automatically selects parameters for superpixel segmentation. Furthermore, instead of custom designing the DL model from scratch, I will use the existing architecture (which has already been tested and proven useful for a variety of tasks), such as Resnet, Unet, and Transformer, for feature extraction and superpixel parameter selection. This would allow me to just one model for different datasets for image segmentation tasks. Finally, I would use Dice Score and IoU as evaluation metrics, which would be a fair comparison against the baseline methods of using DL and OBIA methods separately. These refinements would allow me to train a single hybrid model that would select both SLIC and superpixel segmentation parameters while optimizing the performance metrics for semantic segmentation tasks. In this manner, my proposed method will integrate DL into the OBIA workflow, utilizing both approaches' benefits.