Improved Automated Building Extraction from High Resolution Remote Sensing Imagery using Time-optimized Deep Learning Techniques

Chintan Maniyar[1] and Minakshi Kumar[1]

[1]Indian Institute of Remote Sensing

November 23, 2022

Abstract

Deep learning techniques are being increasingly used in earth science applications - from climate change modelling to feature extraction from remote sensing imagery, given their advantage of increased contextual and hierarchical feature representation. However, deep learning comes at an expense of extensive computational resources and long training time to achieve benchmark results. This study suggests time-optimized deep learning techniques for training deep convolutional networks for one of the most sought after feature extraction subsets – building extraction from satellite/aerial imagery. Building extraction is one of the most important tasks in the dynamic pipeline of urban applications such as urban planning and management, disaster management, urban mapping etc. among other geospatial applications. Automatically extracting buildings from remotely sensed imagery has always been a challenging task, given the spectral homogeneity of buildings with the non-building features as well as the complex structural diversity within the image. With the availability of high resolution open-source satellite and UAV data, deep learning techniques have greatly improved building extraction. However, training on such high resolution data requires the networks to be significantly deeper, resulting in long model training and inference times. This study proposes a combination of two time efficient methods to train a Dynamic Res-U-Net for building extraction in less time without decreasing the training parameters: 1) Using Cyclical Learning and SuperConvergence concepts by dynamically changing the learning rate while training the network to achieve very high accuracy in very less time and 2) Using a specific order to train the layers of the network(s) to specially have the last layers of the networks perform better, leading to an overall improved network performance in lesser time. Building extraction results are gauged using the metrics of Accuracy, Dice Score and Intersection over Union (IoU) and F1-Score. The metrics comparison of training the Res-U-Net in the conventional way vs the proposed techniques shows an evident optimisation in terms of time. Better results are achieved in lesser training epochs using the proposed time-optimised training techniques.

Improved Automated Building Extraction from High Resolution Remotely-sensed Imagery using Time-optimised Deep Learning Techniques

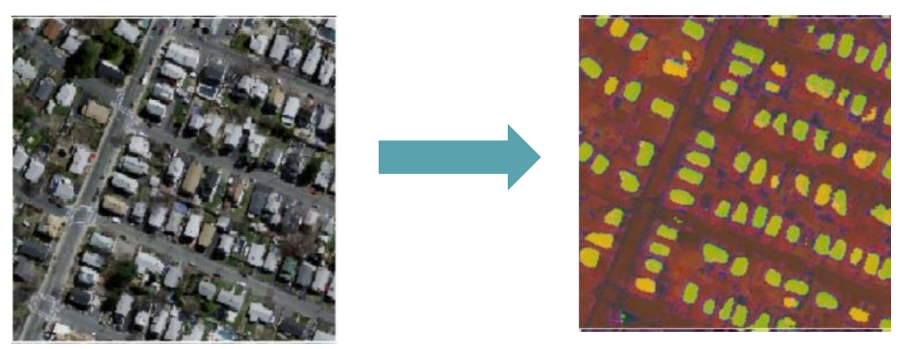
Chintan B. Maniyar* and Minakshi Kumar

Correspondence: chintanmaniyar@gmail.com

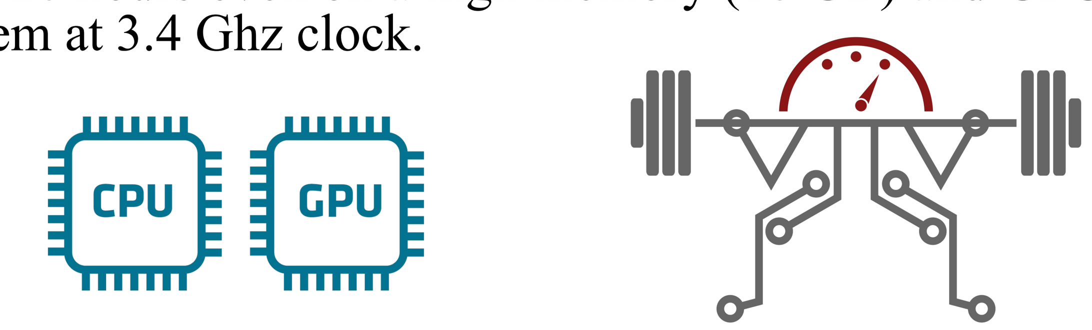
Indian Institute of Remote Sensing, Indian Space Research Organization, Dehradun - 248001, Uttarakhand, INDIA

INTRODUCTION

- Automating building extraction has been one of the hottest research topics given its applications in urban planning, urban monitoring, urban change detection, urban heat islands, damage assessment, disaster management, building index growth, cadastral mapping etc.



- Being a highly cognitive problem, there is a hardcore requirement of high spectral-spatial contextual information, which also poses as a significant challenge as there is always a trade-off between spectral and spatial resolution in remote sensing.
- Since the past decade, Deep Learning algorithms, with their representational ability, have stepped up to address this issue by producing excellent results with high and very high resolution spatial data hence reducing the requirement for a densely spectral information plot. Moreover, there are always performance challenges in case of blurred or irregular boundaries.
- However, deep learning algorithms come with a downside of high time and resource requirement to train the model. And on a large dataset of aerial/satellite imagery, the training takes more than 20 hours even on a high memory (16 GB) and GPU (4 GB) system at 3.4 Ghz clock.



OBJECTIVE

This study has the following objectives:

- To develop a single deep neural network for multiscale building extractions.
- To propose a cost function that can handle irregular shapes and blurred boundaries for building detection
- To optimize the training process of neural network without reducing the trainable parameters and meeting the benchmarks for building extraction

METHODOLOGY

- U-Net with dynamic encoder-decoder architecture is built on top of a pretrained ResNet34
- A new loss function, "Combo Loss", which is a combination of two fundamental cost functions based on edge loss and region loss, is suggested. This way a single loss function accounts for irregular shapes (region loss) and blurred boundaries (edge loss)

- Loss Functions – Unique Combo Loss

$$ComboLoss = BCE_{Loss} + DiceLoss$$

Region + Edge

Where g = ground truth image and p = predicted image

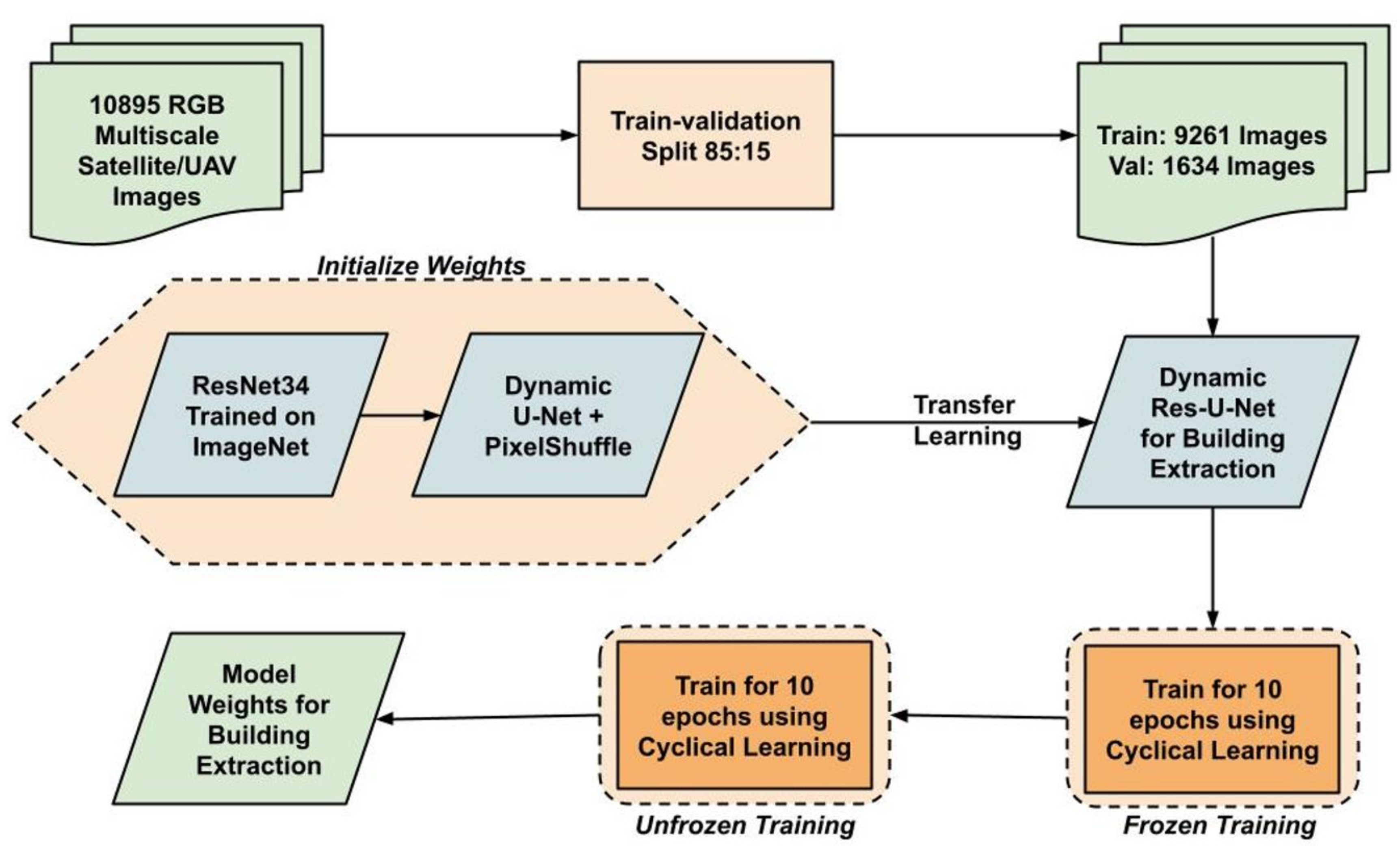
$$BCE_{Loss} = -\frac{1}{patchsize}\sum_{i=1}^{patchsize} g_i \times \log p_i + (1-g_i) \times \log(1-p_i)$$

Edge Based

$$DiceLoss = \frac{2\times\sum_{i=0}^{patchsize} p_i g_i}{\sum_{i=0}^{patchsize} p_i^2 + \sum_{i=0}^{patchsize} g_i^2}$$

Region Based

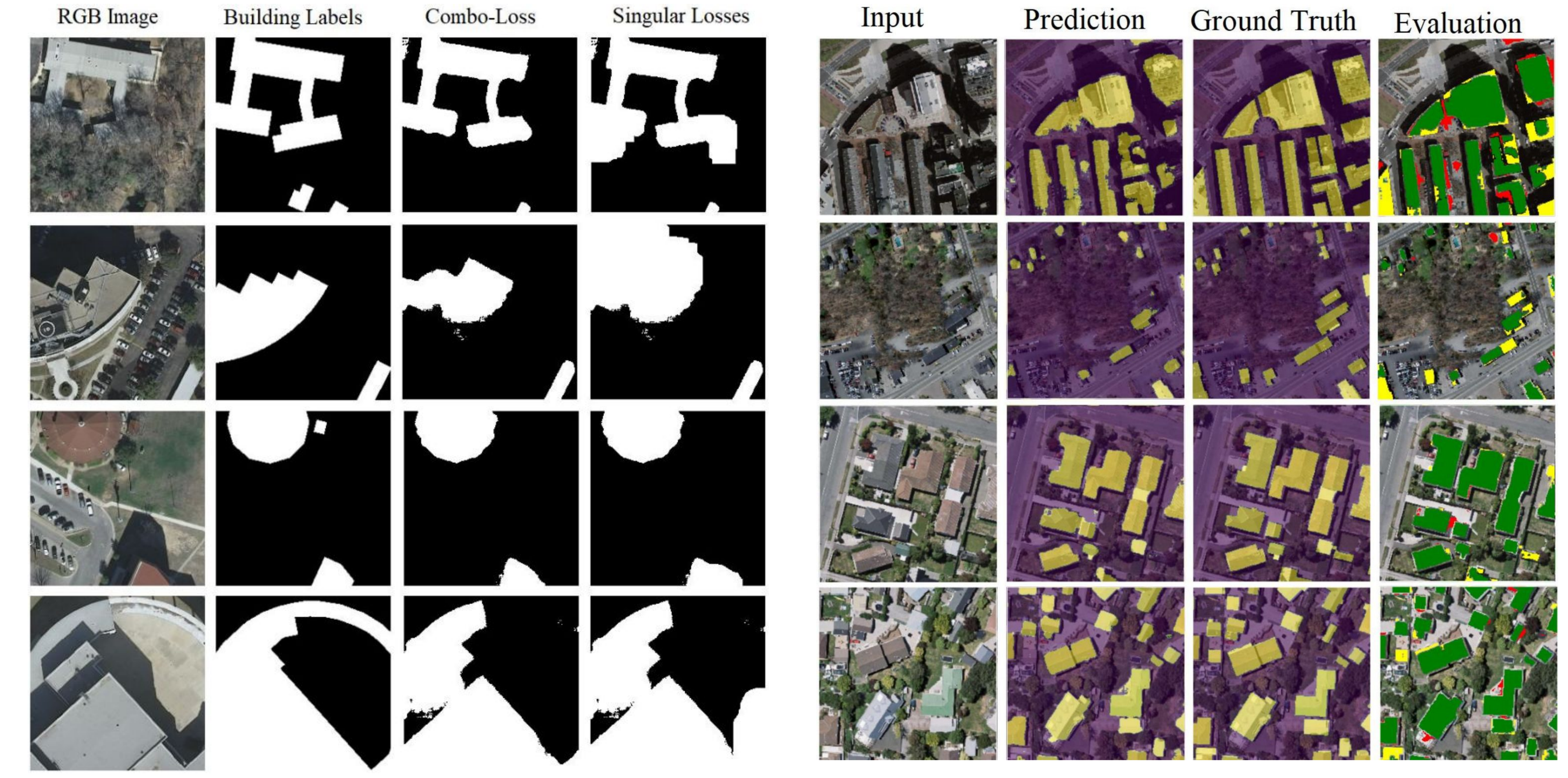
- To reduce the training time and consequently the resource consumption for the deep learning model, a unique order to train the network layers is suggested. Since the last few layers are most difficult to train, we train them first, until they start to converge.
- This is done by freezing the model, training the last layers until earliest convergence and then unfreezing the model, training the entire model from the input layer to the output layer.



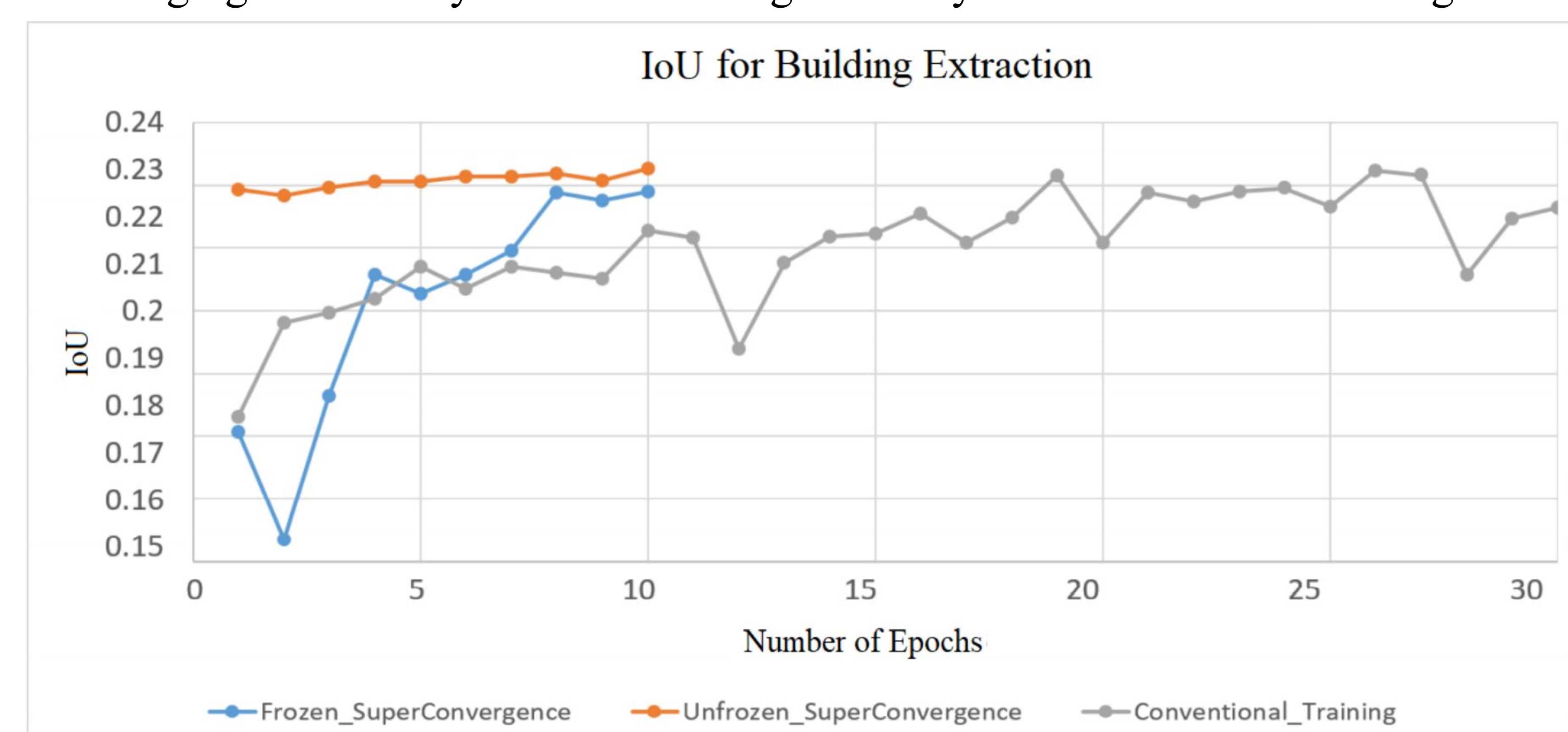
- In addition to changing the order of training the layers, we also implement the policy of cyclical learning, which dynamically changes the learning rate while training the model, depending on the gradient of loss.

RESULTS

- Combo Loss improves building extraction for irregular shapes and blurred boundaries. One-cycle fit policy identifies optimum learning rate dynam



- Changing order of layers while training evidently makes the model converge faster



	Conventional Training Method	Proposed Training Method
Accuracy	95.6%	**96.55%**
Dice Score	0.772	**0.785**
Mean IoU	0.80	**0.84**
Mean F1-Score	0.88	**0.91**
Training Epochs	30	10
Training Time	~11.2 hours	**~4.5 hours**