Improving land cover segmentation across satellites using domain adaptation

Nadir Bengana, Janne Heikkilä

Abstract—Image segmentation for Land Use and Land Cover (LULC) mapping is a valuable asset that saves a lot of time and effort as opposed to manual annotation. However the lack of datasets to train a model good enough to cover a variety of locations on the planet does not exist. Domain adaptation which proved to be quite useful in segmenting street view images can be used to solve the problem of scarcely labeled land cover datasets. In this paper we build a few labeled datasets based on multispectral imagery from Sentinel-2, Worldview-2 and Pleiades-1 satellites and test domain adaptation between those datasets. Experiments show that domain adaptation manages to make it possible to semantically segment images from different areas on the planet with a limited amount of labeled data.

Index Terms—Domain adaptation, Land cover segmentation, image segmentation

I. INTRODUCTION

O

VER the last few years Remote Sensing (RS) data became easily obtainable thanks to the surge of open data from some Earth Observation (EO) satellites. Data such as the Sentinel-2 data from the European Copernicus program [1] and Landsat data from the U.S. Geological Survey (USGS) [2] are available for free. These satellites provide multispectral high resolution imagery (up to 10m). This gave the opportunity to use big data in RS applications which in turn facilitated the usage of Deep Learning (DL) tools in the processing of that data. Land Use and Land Cover (LULC) classification and segmentation is an example of such applications. It is indubitably an important tool in RS providing a way to monitor forests, agriculture, oceans, etc.

Lately classical Machine Learning (ML) tools have fallen out of favor when it comes to many machine vision notably LULC segmentation. Otávio et al. [3] showed that Convolutional Neural Networks (CNN) vastly outperform the classical ML methods when it comes to land cover classification. In the land cover segmentation section of the DeepGlobe challenge [4] the leaderboards are completely dominated by Deep Neural Networks (DNN). In land cover segmentation, [5], [6], and [7] are leading examples in the land cover segmentation where they rely on DNNs such as ResNet and DenseNet.

Deep Neural Networks (DNN) perform well, but, they require a lot of data to show their true potential. The availability of free data from a few satellites does not translate well to ground truth LULC annotation. Corine Land Cover (CLC) is an example of annotation over all of Europe, but the pixel resolution is quite low at 100m/px. Better resolution counterparts exist such as the 20m/px version covering Finland provided by the Finnish environment institute [8] or the 10m/px version covering Germany by Geodatenzentrum [9]. Albeit these versions are limited and not fool proof they can be used to train a DNN to adapt from one location to another or from one satellite to another.

Removing the domain shift between different datasets due to different distribution, lighting, and other factors is called Domain Adaptation. In the context of semantic segmentation, domain adaptation is used to segment images from a target dataset using a source dataset. There exist two types of domain adaptation approaches, supervised where some or all of the target data is labeled, and unsupervised where the target data is unlabeled. Domain adaptation for semantic segmentation is usually applied on street-view images or basic character recognition images. It has the potential to improve upon results achieved with simple transfer learning.

Our contributions are:

1) To the best of our knowledge this is the first paper that goes into applying domain adaptation to LULC mapping.
2) Building various labeled datasets using CLC.
3) Adapting an existing domain adaptation method for satellite imagery.

II. RELATED WORK

1) LULC segmentation: Land use and land cover image segmentation is a field that has not been exploited much especially using deep learning. Usually classical machine learning methods are applied such as support vector machines and decision trees [10] [11]. This is due to the lack of labeled data to train the required DNNs for the task. A small amount of research has been done to tackle the task land cover segmentation using deep learning. As an example the state of the art semantic segmentation DNN based methods were translated to satellite imagery, however, the results were not as good. This is caused by the difference between regular objects images and satellite images where unlike the regular shapes of objects the land cover types have random shapes like forests stands and bodies of water. Kuo et al. [7] presents a method that is one of the leading results in the Deepglobe challenge, in which improving the results relied on a variation of Deeplabv3+ [12] where the fully connected layers of the ResNet backbone are replaced with ASPP (Atrous Spacial Pyramid Pooling). In addition an encoder-decoder architecture is deployed to reduce the effect of resolution loss due to pooling and strided convolution. The encoder-decoder method is a repeating approach on the leading methods in the field of
land cover segmentation. This is due to the necessity of high resolution to land cover since not only structure affects the segmentation but texture and color play a big role resulting from the irregularities of land cover; for the most part. Hasan et al. [13] used state of the art semantic segmentation method, and added LIDAR\footnote{LIDAR uses light in the form of a pulsed laser to measure ranges} data that improved upon the results a bit compared to other methods.

2) Domain adaptation: Domain adaptation takes two different sets (source and target) with different distributions and aims at reducing the domain shift between them by changing and aligning the distribution of one of the sets to match the other. This can also be done by aligning the sets to a common space different from both sets. There are mainly three forms of domain adaptation approaches [14], [15]. The first method is a classical form of minimizing a distance between the source and the target data. Maximum Mean Discrepancy (MMD) is an example of such measures to minimize to achieve domain-invariant feature representation that performs well in both source and target domains. The second method is adversarial based domain adaptation that relies on using Generative Adversarial Networks (GANs) [16] to make one of the sets appear similar to the other one. Eric et al. [17] is an example of the adversarial methods where the target data is encoded into the source data with a discriminator to recognize between the two. The third method creates a shared representation for each domains so to not only translate one domain to another but both domain to a common space, or have a transfer function capable of translating both models to each other and back to the original one. CycleGAN [18] is a example of the third method where two discriminators are used to map images from the source data to the target data and vice-versa.

3) Domain adaptation for semantic segmentation: Domain adaptation can be used for multiple applications including semantic segmentation. All methods mentioned previously have been tested in the field of semantic segmentation with varying level of success. In fact domain adaptation is very useful for semantic segmentation. The reason behind it is that the pixel annotation of images is expensive and require a lot of resources. Generally domain adaptation methods used for classification are not well translated for semantic segmentation [19]. Therefore adversarial and reconstruction methods have been preferred. Architectures such as [20], [21] are examples of adversarial domain adaptation that aims at using a GAN to generate source like images and then segment them using a network trained on the source data. The reconstruction approach has been tested by many methods with some having different variations [22], [23], [24], [25]. The datasets that they use are almost exclusively the street view images datasets, including Cityscape [26], GTA5 [27], and Synthia [28]. Chang et al. [29] is an example where a Domain Invariant Structure Extraction (DISE) framework was proposed to disentangle images into domain-invariant structure and domain-specific texture representations. Another new method that adds an extra step was proposed by Li et al. [30] called bidirectional learning (BDL). The principle behind it is to add a bidirectional learning system and alternate between learning the segmentation adaptation and the image translation model and adding a loss that supervises the translation using the segmentation adaptation model. This prevents the translation model from converging to a point where the discriminator sees the images as being from the same distribution while not aligning the classes correctly which causes the segmentation to fail.

III. MATERIALS

A. Satellite data

The satellite data used in the study comprised of four EO satellites. The first one containing the bigger bulk of the data is from the Sentinel-2 constellation of satellites. The second satellite is the Worldview-2 satellite containing a much fewer amount of data. The third satellite is Worldview-3 through an already available dataset with images taken from Worldview-3 Vivid data [31]. Finally the Pleiades satellite data with the least amount of images.

Sentinel-2 satellites is composed of two polar-orbiting satellites placed in the same sun-synchronous orbit, phased at 180° to each other. The satellites are named Sentinel-2A and Sentinel-2B. They each contain a multispectral sensor with 12 bands ranging from 60m/pix resolution up to 10m/pix. Of which the four 10m/pix bands were chosen to be used. These bands shown in Table I contain the Near Infra Red (NIR) band which is extremely useful in detecting vegetation.

Through the European Space Agency’s (ESA) third party mission data from Worldview-2 satellite is available for free as WorldView-2 European Cities [32]. Worldview-2 is a Very High Resolution (VHR) satellite that contains an 8-bands multispectral sensor with 1.8m/pix resolution shown in Table II and a 0.46cm/pix pan-chromatic sensor.

Pleiades-1 satellite is part of the Pleiades constellation of satellites whose data is not available for the public for free. It is similar in its properties to Worldview-2 (Table I), so it would be interesting to see the results of using Worldview-2 to train a model that can do land mapping on the Pleiades data.

The data obtained from satellite imagery comes in the form of raster images in floating point format. Both Sentinel-2 and Worldview have a 12bit radiometric resolution encoded in 32bit floating point. These had to be encoded in 8bit unsigned integer format to be able to handle them as image files using Python libraries. This process required the use of QGIS which is a Geographical Information System (GIS) allowing the handling of high resolution multi-channel rasters. The main steps consisted of merging the rasters, translating the format to 8bit images and extracting the RGB channels before normalizing them to have a unified illumination.

The bands mentioned in III-A were extracted to form a 3 channel 8bit RGB raster. This in turn got divided into patches of PNG images with 224 × 224 resolution for Sentinel-2, 512 × 512 for Worldview-2, and 448 × 448 for Pleiades-1. The resulting amount of images from Sentinel-2’s is 37706 RGB images. The amount of Worldview-2 images is much lower resulting in 3570 RGB images. The Pleiades-1 dataset is the one with the least amount of images at only 500 RGB images.

1LIDAR uses light in the form of a pulsed laser to measure ranges.
TABLE I
SENTEL2-B PROPERTIES

| Bands | Bandwidth (nm) | Central wavelength (nm) | Spacial resolution (m) |
|-------|---------------|-------------------------|------------------------|
| B2: Blue | 66 | 492.1 | 10 |
| B3: Green | 56 | 559 | 10 |
| B4: Red | 51 | 665 | 10 |
| B5: NIR | 106 | 833 | 10 |

TABLE II
WORLDVIEW-2 PROPERTIES

| Bands | Bandwidth (nm) | Central wavelength (nm) | Spacial resolution (m) |
|-------|---------------|-------------------------|------------------------|
| B2: Blue | 60 | 480 | 1.84 |
| B3: Green | 70 | 545 | 1.84 |
| B5: Red | 60 | 660 | 1.84 |
| B7: NIR | 125 | 833 | 1.84 |

TABLE III
PLEIADES-1B PROPERTIES

| Bands | Bandwidth (nm) | Central wavelength (nm) | Spacial resolution (m) |
|-------|---------------|-------------------------|------------------------|
| B2: Blue | 120 | 490 | 2 |
| B3: Green | 120 | 550 | 2 |
| B4: Red | 120 | 660 | 2 |
| B5: NIR | 200 | 850 | 2 |

The DeepGlobe land cover segmentation dataset [4] is used for comparison with available methods. Containing 1146 images of which only 803 are labeled. The dataset is built from Worldview-3’s vivid images [31]. It covers India, Indonesia, and Thailand. This dataset is readily available without the need for preparation. It contains 12847 images with a size of 612 × 612 pixels, which I cropped from the full size of 2000×2000 pixels since using the full resolution images would require more GPU memory. The format of the images is 8bit with a PNG RGB labeling.

For all the datasets data augmentation has been introduced during training as a mixture of random rotation and cropping.

B. Label data

The label data chosen is from Corine Land Cover by Copernicus (CLC) [33]. CLC is a manually annotated map of Europe based on 44 classes ranging from natural covers such as forests and water surfaces to man made covers such as buildings and crops. Roughly in every 6 years a new version of CLC is available since the year 2000. The pixel resolution of it is 100m/px, however, there exists a version with 20m/px covering the whole area of Finland [8] and a 10m/px version covering Germany. This modified version of CLC2018 is the one used in the study with further modifications involving the classes. These modifications consist of merging some classes together to get them down to 7 classes instead of the original 44 classes. Table IV shows the details of the classes used. The lack of label data in the field of LULC is a serious problem that feeds itself. Without good data it is difficult to build reliable models and thus limiting the ability to use them to generate global label data. While the data covering Finland is accurate, the one covering Germany is not. Several errors can be spotted that ignores small objects such as individual houses in farms which defies the purpose of having a high resolution raster. The label data was aligned to the same Coordinate Reference System (CRS) as the corresponding satellite rasters. Furthermore, the CLC raster was upsampled to match the pixel resolution of the satellite raster it is covering, then divided into patches. It was then converted into a single channel 8bit PNG with the values of each pixel ranging from 0 to 6 representing the corresponding class at that pixel.

DeepGlobe dataset comes with its own labels which the Corine labels were made to match for easier comparison. The labels are the same ones shown in Table IV. They are made with human annotators.

C. Study area

The study area varies between Sentinel-2 and Worldview-2 both covering parts of Finland and Germany with none of the areas between the satellites overlapping. Worldview-2 covers 1520.17km² in Finland and 1310.18km² in Germany (Fig 2). The rasters were carefully chosen to avoid any cloud coverage that might compromise the efficiency of the training. Sentinel-2 covers a far larger area in Finland at around 128320.21km² from all over the country. The area covered in Germany by Sentinel-2 is also larger at around 74361.98km² (Fig 1). The availability of more data from Sentinel-2 is due to the fact that all of its data is available for free whereas a limited amount of the Worldview-2 data is freely available. Finally the Pleiades-1 data which covers a small area of Finland at around 519.67km² with no overlapping data with the Worldview-2 data. Some of the area overlaps with Sentinel-2.

IV. Method

The method applied on this study relies on using domain adaptation to be able to semantically segment satellite images from a different dataset than the one with the ground truth images already available.

A. Network architecture

The network used are based on the ones used by [30]. We also explore the effect of newer semantic segmentation architectures on the results.

The architecture proposed by Li et al. [30] is shown in Fig 3 containing two subnetworks. The first network (F) is an
image to image translation network based on GANs that takes images from the source dataset and learns to translate them into the target dataset’s distribution. The architecture of F is CycleGAN with a 9 block residual network as generator and a 3 layer discriminator. The second network (M) is the segmentation network that learns to segment images using ground truth labels. The segmentation network is based on Deeplabv2 which is a network specialized for semantic segmentation with ResNet as its backbone, and in this case ResNet101 pretrained with ImageNet dataset. M is connected to a discriminator that learns to distinguish between segments generated from target images and translated source images.

B. Training

The training was performed on a different set of Nvidia GPUs (Tesla V100, Tesla P100, Tesla T4) with 16GB of video memory. For about 250,000 iteration with a batch size of 4. The batches were randomized for every iteration in the epoch.

The training of the BDL network uses the loss functions $l_M$ to train the segmentation network shown below:

\[
l_M = \lambda_{adv} l_{adv}(M(S'), M(T)) + l_{seg}(M(S'), Y_s) \tag{1}\]

as for the translation network $l_F$ to train the translation network shown below:

\[
l_F = \lambda_{GAN}[l_{GAN}(S', T) + l_{GAN}(S, T')] + \\
\lambda_{recon}[l_{recon}(S, F^{-1}(S')) + l_{recon}(T, F^{-1}(T'))] + \\
l_{per}(M(S), M(S')) + l_{per}(M(T), M(T')) \tag{2}\]

where $S$ is the source data, $T$ is the target data, $S'$ is the translated data from source to target, $T'$ is the translated data from target to source, $l_{adv}$ is the adversarial loss by the discriminator added to the segmentation network, $l_{seg}$ is the cross entropy loss between images and labels, $F^{-1}$ is the reverse function of $F$, $l_{GAN}$ is the GAN adversarial loss, $l_{recon} = E[||F^{-1}(S') - S||_1]$ is the reconstruction loss, and $l_{per}(M(S), M(S')) = \lambda_{per}E[||M(S) - M(S')||_1] + \lambda_{per-recon}E[||F^{-1}(S') - M(S)||_1]$ is the perceptual loss which is the loss that connects the segmentation network back the translation network. The coefficients $\lambda_{adv}$, $\lambda_{GAN}$, $\lambda_{recon}$, $\lambda_{per}$, and $\lambda_{per-recon}$ are used to emphasize on some losses more than others. Those coefficients are the ones that will help guide the translation network using the segmentation network. And Fig 5 show the difference between setting $\lambda_{per} = 0.1$ and $\lambda_{per} = 1$ respectively. The result with setting $\lambda_{per} = 0.1$ which works with streetview images causes forestry to be replaced with desert since both of them have a random shape and since these classes both exist it doesn’t trigger the discriminator to think it is wrong. Therefore it is important to find the best coefficients for each task. A multitude of combinations have been tested based on previous results and intuition. In addition the loss function has been modified to the following one:

\[
l_F = \lambda_{GAN}(E[\lambda_B D(T) + E[1 - \lambda_B D(S')]) + \\
E[\lambda_B D(T')] + E[1 - \lambda_B D(S')]) + \\
\lambda_{recon}[l_{recon}(S, F^{-1}(S')) + l_{recon}(T, F^{-1}(T'))] + \\
\lambda_{per-A}E[||M(S) - M(S')||_1] + \\
\lambda_{per-recon}E[||F^{-1}(S') - M(S)||_1] + \\
\lambda_{per-B}E[||F^{-1}(T') - M(T)||_1] + \\
\lambda_{per-recon}E[||F^{-1}(T') - M(T)||_1] \tag{3}\]

$2 F^{-1}$ is its own trained network and not just an inverse of F. However, its goal is to do the reverse of F.
The main changes are the separation of $\lambda_{per}$ to $\lambda_{perA}$ and $\lambda_{perB}$ which were modified from the default value of 0.1 to a higher one to avoid an issue where the classes are not matched correctly. Fig 4 shows an example of such occurrence where trees from Worldview-2 are being replaced by barren class in DeepGlobe domain. Another change is the addition of $\lambda_D$ as a result of the generator not focusing on making good enough images with the inclusion of a higher value of $\lambda_{perA}$ and $\lambda_{perB}$. This causes the generated image to look a lot like the source dataset therefore $\lambda_D$ would force the discriminator to put the generator back in track while preserving the classes alignments. Fig 5 shows the previous error fixed with by changing the values of $\lambda_D$, $\lambda_{perA}$, and $\lambda_{perB}$. The values change depending on the experiment ran. While translating Worldview-2 to DeepGlobe requires $\lambda_D = 1.5$, $\lambda_{perA} = 0.5$, and $\lambda_{perB} = 0.1$ translating Sentinel-2 to DeepGlobe requires $\lambda_D = 100$, $\lambda_{perA} = 2$, and $\lambda_{perB} = 0.5$. The reason for that is the difference of resolution where in the Worldview-2 to DeepGlobe translation the resolution is similar so texture information can be used by the generator to match classes from source to target. In the Sentinel to DeepGlobe translation the resolution difference is much higher and there is not much information shared between the classes from both sets apart from color where it could also be different.

C. Metrics

Semantic segmentation uses a varying set of metrics to measure how accurate it is compared to the GT data. Those include Mean Intersection over Union (MIoU), Average Precision (AP), Pixel Accuracy (PA), and Boundary F1 (BF) score. In this experimentation MIoU has been used.

MIoU is a widely used metric for the evaluation of semantic segmentation. It computes the mean of the rate of overlap between the GT segments and the resulting segmentation. MIoU is obtained using the following equation:

$$MIoU = \frac{1}{n} \sum_{i=1}^{n} \frac{GT_i \cap Output_i}{GT_i \cup Output_i}$$

(4)

where $n$ is the number of classes. Another way to write the formula is:

$$MIoU = \frac{1}{n} \sum_{i=1}^{n} \frac{TP}{TP + FP + FN}$$

(5)

V. RESULTS AND DISCUSSION

DeepGlobe dataset is the only available standard dataset for land cover segmentation so trying to fit results to it is a good way to compare with other methods. The Worldview-2 dataset contains somewhat similar images to the ones in the DeepGlobe dataset since they are both from the Worldview constellation albeit DeepGlobe uses images from Worldview-3 and with different post processing methods. The test of using domain adaptation from Worldview-2 (Finland and Germany) to DeepGlobe is going to be referred to as "WV2 to DG". Similarly a test that also adapts to DeepGlobe is the Sentinel-2 dataset to DeepGlobe referred to as "Sen to DG". Unlike WV2 to DG where Worldview-2 is a satellite of somewhat similar properties with the big difference being the location, Sen to DG is trying to have domain adaptation between two very different satellites. In addition to DeepGlobe images being captured from Asian countries close to the equator, the Sentinel-2 dataset covers Finland and Germany only. Having a test between two similar satellites such as Worldview-2 and
Pleiades-1 satellites from the same location is helpful to know whether or not it is necessary to use domain adaptation and how useful it is. The test will be referred to as "WV2FI to PLFI". To test how well domain adaptation performs between satellites when the location is similar we implemented a test between Sentinel-2 and Worldview-2 referred to as "Sen to WV2". Both satellites cover non-overlapping areas from Finland and Germany. The final test aims at seeing how well domain adaptation works for different locations when the sensor used is the same. Therefore, we implemented Worldview-2 Germany to Worldview-2 Finland referred to as "WV2GR to WV2FI". Finland and Germany do not share much of the land cover distribution. The tree species for once are quite different and Finland has much more lakes and forests compared to Germany with more urban areas and agriculture.

A. No domain adaptation results

To test the improvements over no domain adaptation, we implemented a separate run with only the segmentation network. Then we ran a test on the model with the target dataset’s validation subset. The results of WV2 to DG test without domain adaptation can be seen in Table V. The results are quite bad and that is caused by the fact that images from Worldview-2 vary a lot from images from DeepGlobe in both sensor and location on the planet with many classes being different as seen in Fig 8. Fig 7 shows an example of a few test images with the ground truth and the model output without domain adaptation. Everything is considered agriculture by the network which makes very unreliable. Similarly Sen to DG results are not reliable albeit better than the WV2 to DG since Sentinel-2 has way more data which results in a slightly more variety. This however is not enough to produce acceptable results as seen in Fig 9.

The Sen to WV2 results without domain adaptation although not being too bad they have quite a lot of inaccuracy especially mixing up similarly looking classes when seen in different resolutions such as forestry and agriculture. Examples can be seen in Fig 10 highlighting the errors mentioned.
| Network                      | Unknown | Urban | Agriculture | Rangeland | Forest | Water | Barren | MIoU  |
|-----------------------------|---------|-------|-------------|-----------|--------|-------|--------|-------|
| WV2 to DG                   | 0.0     | 0.02  | 57.55       | 0.0       | 0.01   | 0.56  | 0.0    | 8.31  |
| Sen to DG                   | 0.0     | 9.29  | 42.19       | 8.02      | 19.66  | 28.74 | 0.11   | 16.86 |
| WV2/FL to PLFI              | 14.24   | 11.33 | 41.58       | 2.74      | 30.28  | 58.21 | 1.2    | 25.7  |
| WV2GR to WV2FL              | 8.90    | 40.15 | 52.73       | 2.74      | 39.85  | 9.27  | 0.0    | 24.8  |

Fig. 11. Results of training without domain adaptation from Worldview-2 Finland to Pleiades-1. Right: output of model. middle: test images from Pleiades-1 dataset. Left: ground truth images from Pleiades-1 dataset.

Fig. 12. Results of training without domain adaptation from Worldview-2 (Germany) to Worldview-2 (Finland). Right: output of model. middle: test images from Worldview-2 (Finland) dataset. Left: ground truth images from Worldview-2 (Finland) dataset.

The results of WV2FI to PLFI without using domain adaptation is not as good as expected. Even though the differences of the sensors between both satellites are not big the results do not translate well. Fig 11 shows an example of results of WV2FI to PLFI without domain adaptation.

Finally the results of WV2GR to WV2FI are surprisingly low considering that the satellite is the same and the labels were both based from CLC. Although the last point is not very relevant since the accuracy of the labels are not similar. Table V shows that the results are quite low with just 9.25 MIoU which is the second lowest score in the list of tests ran. Fig 11 shows an example of the results from the run.

B. BDL results

The results of using domain adaptation for the WV2 to DG can be seen in Table VI. Although the result is not very impressive numerically, there is a big difference between that and the result without domain adaptation from 8.31 to 29.8. Fig 13 shows an example of a few test images with the ground truth and the model output after domain adaptation showing very similar labeling to the ground truth. In addition to that the results in some cases could be better than the ground truth and since the logic behind the labeling of the classes is not clear the errors could be debatable. As an example it is unclear what is considered a forest and what is considered rangeland, in addition to that some small villages have been completely ignored in the ground truth while parts of them are not as seen in Fig 14.

Using domain adaptation between Sentinel-2 and Worldview-2 is another successful example. The difference of results between using domain adaptation and not using it is quite big, almost double as seen in Table VI. Samples from the test set can be seen in Fig 15. As mentioned in Section III-B the ground truth labels for Germany are lacking in precision which would limit the results even when training on that specific dataset. However having good data covering Finland would help correcting the errors in it and end up getting results that are better than human annotation in some...
TABLE VI
RESULTS WITH BDL AND DEEPLABV2

| Network            | Unknown | Urban | Agriculture | Rangeland | Forest | Water | Barren | MIoU |
|--------------------|---------|-------|-------------|-----------|--------|-------|--------|------|
| WV2 to DG          | 0.0     | 43.99 | 69.81       | 4.99      | 39.76  | 55.79 | 0.0    | 29.76|
| Sen to DG          | 0.1     | 29.78 | 40.42       | 9.73      | 23.3   | 62.67 | 0.58   | 23.8 |
| Sen to WV2         | 67.71   | 34.65 | 79.87       | 6.16      | 76.27  | 77.08 | 0.0    | 48.82|
| WV2FI to PLFI      | 0.0     | 49.64 | 80.85       | 21.14     | 76.71  | 6.81  | 0.04   | 33.6 |
| WV2GR to WV2FI     | 0.0     | 21.9  | 39.91       | 2.68      | 71.07  | 51.84 | 0.0    | 27.06|

Fig. 14. Improving the ground truth using domain adaptation. Right: output of model. Middle: test images from DeepGlobe dataset. Left: ground truth images from DeepGlobe dataset.

Fig. 15. Results of training with domain adaptation from Sentinel-2 to Worldview-2. Right: output of model. Middle: test images from Worldview-2 dataset. Left: ground truth images from Worldview-2 dataset.

Fig. 16. Results of training with domain adaptation from Sentinel-2 to Worldview-2. Right: output of model. Middle: test images from Worldview-2 dataset. Left: ground truth images from Worldview-2 dataset.

Sen to DG test was successful since it improved the results from a MIoU of 16.86 to a 23.8 which might not sound like much especially since the results with Deeplabv2 on Deepglobe dataset is 52.24 [35], but, visually it translates to good results as seen in Fig 17. The results are still not as good as with WV2 to DG and that could be explained with the huge difference of pixel resolution between Sentinel-2 satellite and Worldview-3 where the DeepGlobe dataset comes from. This is still a good step as the Sentinel data is free whereas Worldview-3 data is very expensive.

With a limited amount of data like the case with the Pleiades-1 dataset, training a DNN on this data is quite difficult if not impossible. Therefore, applying domain adaptation is a good way to test the efficiency of such method to be able to have land mapping on a very limited dataset. WV2FI to PLFI show good results with almost an extra 10% MIoU as seen in Fig 18 which is quite good considering that it would be costly to train on Pleiades-1 data.

The results obtained from WV2GR to WV2FI using domain adaptation have improved significantly compared to the previous results from not using domain adaptation. As seen in Table VI, the MIoU score rose up 27 to from 9 without domain adaptation. Fig 19 shows example from test images using this domain adaptation. Although the results are not very precise especially on small details it is much better compared to Fig 12.
C. Other tests

In addition to the previous experiments Deeplabv3+ has been tested as the segmentation network. Replacing the backbone with Deeplabv3+ did not bring much bigger improvements compared to Deeplabv2 which is shown in Table VII. The reason behind that is the imperfection of the dataset. In fact the results start to diverge after a while due to the network being too good at mimicking the errors in the ground truth images. So in order to make use of the best potential of deep learning methods it is important to have a very good and precise dataset in land cover mapping.

![Fig. 17. Results of training with domain adaptation from Sentinel-2 to DeepGlobe. Right: output of model. Middle: test images from DeepGlobe dataset. Left: ground truth images from DeepGlobe dataset.](image1)

![Fig. 18. Results of training with domain adaptation from Worldview-2 (Finland) to Pleiades-1 (Finland). Right: output of model. Middle: test images from Pleiades-1 (Finland) dataset. Left: ground truth images from Pleiades-1 (Finland) dataset.](image2)

![Fig. 19. Results of training with domain adaptation from Worldview-2 (Germany) to Worldview-2 (Finland). Right: output of model. Middle: test images from Worldview-2 (Finland) dataset. Left: ground truth images from Worldview-2 (Finland) dataset.](image3)
VI. CONCLUSION

In this paper, we experimented on applying domain adaptation to semantically segment satellite images for LULC. We chose to test on the state of the art method available for street view images as they are the most popular in the field of semantic segmentation. The experimental results show that the use of domain adaptation could highly benefit the ability to label LULC without the need for various labeled datasets, which is something that it lacks. The results suggest that it is even possible to segment satellite images from areas completely different than the ones used to train the model.

ACKNOWLEDGMENT

This research was financially supported by Business Finland (Grant no. 1259/31/2018). Authors would also like to thank the project partners from VTT Technical Research Centre of Finland for providing their datasets and technical support.

REFERENCES

[1] “Copernicus open access hub.” [Online]. Available: https://scihub.copernicus.eu/dhus/
[2] “Ugs.” [Online]. Available: https://earthexplorer.usgs.gov/
[3] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, “Do deep features generalize from everyday objects to remote sensing and aerial scenes?” in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2015, pp. 44–51.
[4] I. Demir, K. Koperski, D. Lindenbaum, G. Fang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, “Deepglobe 2018: A challenge to parse the earth through satellite images,” CoRR, vol. abs/1805.06561, 2018. [Online]. Available: http://arxiv.org/abs/1805.06561
[5] C. Tian, C. Li, and J. Shi, “Dense fusion classmate network for land cover classification,” 06 2018, pp. 262–2624.
[6] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” CoRR, vol. abs/1608.06993, 2016. [Online]. Available: http://arxiv.org/abs/1608.06993
[7] T.-S. Kuo, K.-S. Tseng, J. Yan, Y.-C. Liu, and Y.-C. F. Wang, “Deep aggregation net for land cover classification,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 247–2474, 2018.
[8] “Metatietopavalu.” [Online]. Available: http://metatieto.ymapisto.fi:8800/geoparial/catalog/search/resource/details?resource=%7B6833C06E-BF77-4F0B-A066-B94AE939292E%7D
[9] “Bkg - corine land cover.” [Online]. Available: https://www.bkg.bund.de/DE/UBer-das-BKG/Geoinformation/Fernerkundung/Landbedeckungsmodell/CorinelandCover/clc.html
[10] M. S. Tehrany, B. Pradhan, and M. N. Jebu, “A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using spot 5 imagery,” Geocarto International, vol. 29, no. 4, pp. 351–369, 2014. [Online]. Available: https://doi.org/10.1080/10106049.2013.768300
[11] R. Khatami, G. Mountrakis, and S. V. Stelman, “A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research,” Remote Sensing of Environment, vol. 177, pp. 89 – 100, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425716300578
[12] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” CoRR, vol. abs/1802.02611, 2018. [Online]. Available: http://arxiv.org/abs/1802.02611
[13] H. A. Arief, G.-H. Strand, H. Tveite, and U. G. Indahl, “Land cover segmentation of airborne lidar data using stochastic atrous network,” Remote Sensing, vol. 10, no. 6, 2018. [Online]. Available: http://www.mdpi.com/2072-4292/10/6/973
[14] M. Wang and W. Deng, “Deep visual domain adaptation: A survey,” CoRR, vol. abs/1802.03601, 2018. [Online]. Available: http://arxiv.org/abs/1802.03601
[15] G. Wilson and D. J. Cook, “Adversarial transfer learning,” CoRR, vol. abs/1812.02849, 2018. [Online]. Available: http://arxiv.org/abs/1812.02849
[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in neural information processing systems, 2014, pp. 2672–2680.
[17] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” CoRR, vol. abs/1702.05464, 2017. [Online]. Available: http://arxiv.org/abs/1702.05464
[18] J. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” CoRR, vol. abs/1703.10593, 2017. [Online]. Available: http://arxiv.org/abs/1703.10593
[19] Y. Zhang, P. David, and B. Gong, “Curriculum domain adaptation for semantic segmentation of urban scenes,” 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2039–2049, 2017.
[20] J. Hoffman, D. Wang, F. Yu, and T. Darrell, “F cn s in the wild: Pixel-level adversarial and constraint-based adaptation,” ArXiv, vol. abs/1612.02649, 2016.
[21] Y.-H. Chen, W.-Y. Chen, Y.-T. Chen, B.-C. Tsai, Y.-C. F. Wang, and M. Sun, “No more discrimination: Cross city adaptation of road scene segmenters,” 10 2017, pp. 2011–2020.
[22] W. Hong, Z. Wang, M. Yang, and J. Yuan, “Conditional generative adversarial network for structured domain adaptation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.
[23] W. Chang, H. Wang, W. Peng, and W. Chiu, “All about structure: Adapting structural information across domains for boosting semantic segmentation,” CoRR, vol. abs/1903.12212, 2019. [Online]. Available: http://arxiv.org/abs/1903.12212
[24] Y. Tsai, W. Hung, S. Schulter, K. Sohn, M. Yang, and M. Chandraker, “Learning to adapt structured output space for semantic segmentation,” CoRR, vol. abs/1802.10349, 2018. [Online]. Available: http://arxiv.org/abs/1802.10349
[25] T.-H. Vu, H. Jain, M. Bucher, M. Cord, and P. Perez, “Advnet: Adversarial entropy minimization for domain adaptation in semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
[26] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[27] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, “Playing for data: synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.
[29] W.-L. Chang, H.-P. Wang, W.-H. Peng, and W.-C. Chiu, “All about structure: Adapting structural information across domains for boosting semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[30] Y. Li, L. Yuan, and N. Vasconcelos, “Bidirectional learning for domain adaptation of semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[31] Digitalglobe basemap +vivid. [Online]. Available: https://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/2/DF_Basemap_Vivid_DS_1.pdf

[32] “Worldview-2 european cities - view data product - earth online - esa.” [Online]. Available: https://earth.esa.int/web/guest/ worldview-2-european-cities-dataset

[33] B. Kosztra, G. Büttner, G. Hazeu, and S. Arnold, “Updated clc illustrated nomenclature guidelines.” [Online]. Available: https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20170930.pdf

[34] “Nvidia tesla v100 data center gpu.” [Online]. Available: https://www.nvidia.com/en-us/data-center/tesla-v100/ 

[35] C. Tian, C. Li, and J. Shi, “Dense fusion classmate network for land cover classification,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.