

Fully Convolutional Open Set Segmentation

Hugo Oliveira · Caio Silva · Gabriel L. S. Machado · Keiller Nogueira · Jefersson A. dos Santos

Received: 15/06/2020 / Accepted:

Abstract In semantic segmentation knowing about all existing classes is essential to yield effective results with the majority of existing approaches. However, these methods trained in a Closed Set of classes fail when new classes are found in the test phase. It means that they are not suitable for Open Set scenarios, which are very common in real-world computer vision and remote sensing applications. In this paper, we discuss the limitations of Closed Set segmentation and propose two fully convolutional approaches to effectively address Open Set semantic segmentation: OpenFCN and OpenPCS. OpenFCN is based on the well-known OpenMax algorithm, configuring a new application of this approach in segmentation settings. OpenPCS is a fully novel approach based on feature-space from DNN activations that serve as features for computing PCA and multi-variate gaussian likelihood in a lower dimensional space. Experiments were conducted on the well-known Vaihingen and Potsdam segmentation datasets. OpenFCN showed little-to-no improvement when compared to the simpler and much more time efficient SoftMax thresholding, while being between some orders of magnitude slower. OpenPCS achieved promising results in almost all experiments by overcoming both OpenFCN and SoftMax thresholding. OpenPCS is also a reasonable compromise between the runtime performances of the extremely fast SoftMax thresholding and the extremely slow OpenFCN, being close able to run close to real-time. Experiments also indicate that OpenPCS is effective, robust and suitable for Open Set segmentation, being able to improve the recognition of unknown class pixels without reducing the accuracy on the known class pixels.

Keywords Open Set Recognition · Semantic Segmentation · Generative Modeling · Deep Learning · Remote Sensing

H. Oliveira · Caio Silva · Gabriel L. S. Machado · Keiller Nogueira · Jefersson A. dos Santos
Av. Antônio Carlos, 6627 – ICEx Building,
Pampulha, Belo Horizonte, Brazil
E-mail: oliveirahugo@dcc.ufmg.br, caiosilva@dcc.ufmg.br, gabriel.lucas@dcc.ufmg.br,
keiller.nogueira@dcc.ufmg.br, jefersson@dcc.ufmg.br

1 Introduction

The development of new technologies for the acquisition of aerial images onboard satellites or aerial vehicles has made it possible to observe and study various phenomena on the Earth’s surface, both on a small and large scale. A highly requested task, in this sense, is automated geographic mapping, which gives an easier and faster approach of monitoring cities, regions, countries, or entire continents. Automatic mapping of remote sensing images is typically modeled as a supervised classification task, commonly known as semantic segmentation, in which a model is first trained using labeled pixels and then used to classify other pixels in a new region. Commonly, this process is based on the Closed Set (or Closed World) assumption: it assumes that all training and testing pixels come from the same label space, i.e., train and test sets have the same set of classes. It is easy to notice that this assumption does not hold in real-world scenarios, mainly for Earth Observation applications, such as geographic mapping, given the huge size of the images and the (possible) elevated number of distinct objects (classes). In these scenarios, the model is likely to observe, during the prediction phase, samples of classes not seen during the training. In these cases, Closed Set semantic segmentation methods are error-prone to unknown classes given that they will wrongly recognize it as one of the known classes during the inference. This limits the use of such approaches to real-world Earth Observation applications, such as automated geographic mapping.

Towards solving this, Open Set Recognition (OSR) can be described as the set of algorithms that address the problem of identifying, during the inference phase, samples of unknown classes, i.e., instances of classes not seen during the training. Using the same definitions of Geng *et al.* [11] and Scheirer *et al.* [32], during the inference phase, an OSR system, such as an Open Set semantic segmentation approach, should be able to correctly classify the instances/pixels of classes employed during the training (Known Known Classes (KKCs)) whereas recognizing the samples/pixels of classes **not** seen during training (Unknown Unknown Classes (UUCs)).

By this definition, it is possible to state that the main difference between the closed and Open Set scenarios is related to the knowledge of the world, i.e., the knowledge of all possible classes. Specifically, while in the Closed Set scenario the methods should have full knowledge of the world, Open Set approaches must assume that they do not know all the possible classes during the training. Obviously, different approaches may have a distinct knowledge of the world depending on the problem. A visual example of this difference, in terms of knowledge of the world, is depicted in Figure 1.

Technically, OSR is usually achieved via transductive learning [22] by trying to infer samples from UUCs using only the test data distribution. Another approach is to adapt Anomaly Detection techniques for OSR by using auxiliary data to try to learn a generative model via supervision, such as the Outlier Exposure (OE) [15]. However, neither of these approaches is ideal for Open Set semantic segmentation. This is because transduction requires continuous updates of a model in order to cope with new data, presenting an overhead that might be too expensive in most real-world applications, and the OE [15] depends on the existence and availability of auxiliary Out-of-Distribution (OOD) data for training, which may not be possible or even useful in semantic segmentation. Therefore, an Open Set

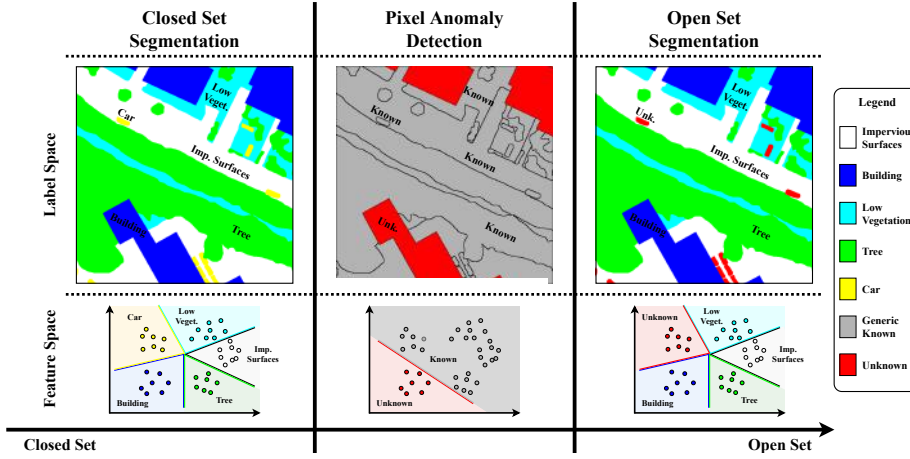


Fig. 1: Graphical depiction of the problem settings of Closed Set, Anomaly Detection and Open Set Recognition in dense labeling scenarios. The x -axis ranges from fully Closed Set (that is, assuming full knowledge of the world) on the left side to Open Set on the rightmost example. In the middle, there is the binary task of Pixel Anomaly Detection, wherein pixels are segmented either into KKC's or UUC's without discerning between distinct KKC's. We also depict the Label Space (Semantic Segmentation map) and a sample of some pixels' representation in a 2D manifold of the Feature Space separated by labels and decision boundaries for each class.

semantic segmentation algorithm must rely solely on inductive learning and use only the available training data and KKC's, while also addressing performance concerns, given the higher output dimensions of segmentation.

In this paper, we proposed two novel approaches for Open Set semantic segmentation of remote sensing image. We evaluate our methods and compare it with baselines by simulating Open Set situations in well known urban scenes such as Vaihingen and Potsdam datasets. It is important to mention that the concept of Open Set semantic segmentation is still very little explored in the literature. The first and unique work that introduces this problem is [6], which proposes a method based on OpenMax [3] for pixelwise classification, which have many limitations concerning both effectiveness and efficiency.

The main contributions of this work are:

1. The first fully convolutional methodology for semantic segmentation in RS imagery or otherwise adapted from OpenMax [3], explained in Section 3.1;
2. Proposal of a completely novel fully convolutional methodology for identifying UUC's in dense labeling scenarios using Principal Components from the internal feature space of the DNNs, further described in Section 3.2;
3. Definition of a benchmark evaluation protocol with standard threshold-dependent and threshold-independent metrics for testing OSR Semantic Segmentation tasks;
4. Extensive evaluations of architectures and thresholding for the proposed approaches and both Open and Closed Set baselines.

This paper is organized in sections, as follow: Section 2 presents the related work, other papers that try to solve the Open Set semantic segmentation problem; Section 3 describes the proposed methods, explaining in details how it works; Section 4 shows the setup used for this paper, the datasets evaluated and the metrics; Section 5 presents the obtained results from the experimental setup; Finally, section 6 contains the conclusion over the obtained results and the proposed method.

2 Related Work and Background

Convolutional Neural Networks (CNNs) have become the backbone of visual recognition for the last decade. AlexNet [20] reintroduced image feature learning, allowing for better scalability than the first CNNs (i.e. LeNet [21]) in order to perform inference over harder tasks (i.e. ImageNet [7] and CIFAR [19]). AlexNet took advantage of larger convolutional kernels in the earlier layers and contained a total of eight layers, between convolutional and fully-connected ones. VGG [37] simplified CNN architectures by using the same 3×3 kernels in all convolutions and max-poolings for downscaling. In contrast to VGG, the GoogleNet architecture [40] – also known as Inception – studied a diverse set of kernel sizes to enforce disentanglement in activations. Inception modules mix combinations of multiple kernel sizes and poolings. Both VGG and Inception allow for deeper networks with smaller convolutional kernels in each module, which proved to be more efficient than shallower networks with larger convolutions, at least up to around 20 layers.

It was observed that adding layers beyond a total of 20 was detrimental to the training of CNNs, as the gradients did not reach the earlier layers, effectively preventing their training. Residual Networks (ResNets) [14] based on residual identity functions that allow for shortcuts in the backpropagation were then introduced. ResNets with between 18 and 151 convolutional blocks were investigated by He *et al.* [14], with little benefit being observed beyond that. With time, some tweaks were proposed to the standard ResNet architecture, with the more noteworthy ones being Wide ResNets (WRNs) [46] and ResNeXt [43], which yielded considerable improvements to the traditional ResNets. However, ResNets were observed to be highly inefficient, as the activations of most convolutions throughout the network could be dropped with little-to-no effect on classification performance [38]. Densely Connected Convolutional Networks (DenseNets) [16] improved on the parameter efficiency of ResNets by replacing the identity function by concatenation and adding bottleneck and transition layers, which lowered the parameter requirements of the architecture. Huang *et al.* [16] tested DenseNet with between 121 and 264 layers, observing them to be more efficient than ResNets in both parameter and flops, when similar error values were compared in the validation set.

The remainder of this section presents the main concepts to the understand of this work and the recent literature about semantic segmentation (Section 2.1) and Open Set recognition (Section 2.2).

2.1 Deep Semantic Segmentation

CNNs [20] are considered the state-of-the-art for sparse labeling tasks as object/scene classification due to their feature learning capabilities. The literature quickly learned to adapt CNNs for dense labeling tasks by patchwise training, using the label for the central pixel [9, 31]. However, Fully Convolutional Networks (FCNs) [25] were shown to be considerably more efficient than patchwise training, providing the first end-to-end framework for semantic segmentation. Besides the accuracy and efficiency benefits of fully convolutional training, any traditional CNN architecture could be converted into an FCN by adding a bilinear interpolation to the activations and replacing the dense layers by convolutional ones, as shown in Figure 2. This simple scheme also allowed for transfer learning from large labeled datasets as ImageNet [7] to relatively smaller semantic segmentation datasets as Pascal VOC [8] and COCO [24]. FCNs are also shown to benefit from skip connections that merge the high semantic level activations at the end of the network with the high spatial resolution information from earlier layers [25].

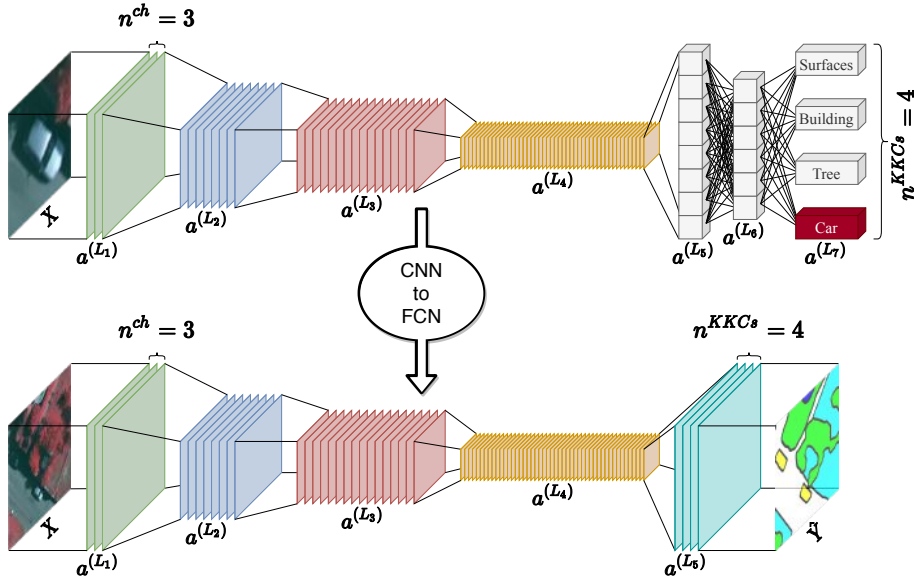


Fig. 2: Architecture example of a CNN for image classification and its equivalent FCN architecture with the same backbone for semantic segmentation. Activations from layer l are depicted as $a^{(l)}$ for all layers in the network (L_1 through L_7 for the CNN and L_1 through L_6 for the FCN). One should notice that in both architectures the input layer $a^{(L_1)}$ has the number of channels n^{ch} depending on the input data's number of channels (in the case of RGB images, $n^{ch} = 3$). In the CNN, the number of neurons in the output layer must match the number of KKCs (n^{KKCs}) in the data. Equivalently, in the FCN, the number of channels in the output layer n^{KKCs} , as suggested by the notation, depends on the number of KKCs of the dataset. In this example, $n^{KKCs} = 4$ as there are 4 known classes in this example.

More recently, several semantic segmentation methods have been proposed specifically to deal with specific aspects of remote sensing images such as spatial constraints [28, 26, 27, 42, 2, 29] or non-RGB data [17, 13]. Nogueira et al. [28] use patchwise semantic segmentation in RS imaging for both urban and agricultural scenarios. Maggiori et al. [26] proposed a multi-context method based on upsampled and concatenated features extracted from distinct layers of a fully convolutional network. In [27], the authors proposed multi-context methods that combine boundary detection with deconvolution networks. In [2], the authors fine-tuned a deconvolutional network using 256×256 fixed size patches. To incorporate multi-context knowledge into the learning process, they proposed a multi-kernel technique at the last convolutional layer. Wang et al. [42] proposed to extract features from distinct layers of the network to capture low- and high-level spatial information. In [17], the authors adapt state-of-the-art semantic segmentation approaches to work with multi-spectral images. Guiote et al. [13] proposed an approach for semantic segmentation from LiDAR point clouds.

2.2 Open Set Recognition

The Open Set recognition problem was first introduced by Scheirer et al. [33]. They discussed about the notion of “openness”, that occurs when we do not have knowledge of the entire set of possible classes during supervised training, and must account for unknowns during predicting phase. The first studies and applications involving Open Set recognition were adaptations of “shallow” methods that acted in the feature space of visual samples and consisted mainly of threshold-based or support-vector-based methods [33]. More recent work has extended the concept to deep neural networks [3, 5].

A recent survey by Geng *et al.* [11] splits Open Set methods mainly between discriminative and generative approaches. Discriminative approaches usually use 1-vs-All Support Vector Machines (SVMs) [33] in order to delineate the space between valid samples from the training classes and outliers – which ideally would identify UUCs. Meta-recognition can also be used to predict failures in visual learning tasks [33, 32]. Extreme Value Theory (EVT) is one of the most common modelings for meta-recognition using DNNs [3, 10, 30] for classification.

Earlier deep OSR methods [3, 10, 23] aimed to incorporate the prediction of UUCs directly onto the prediction of the DNN output layer. Bendale *et al.* [3] and Ge *et al.* [10] perform this by reweighting the output probabilities of the SoftMax activation to accommodate a UUC into the prediction during test time. This approach is known as OpenMax [3], and further developments to it have been proposed; for instance, aiding the computation of OpenMax with synthetic images from a Generative Adversarial Network (GAN) in G-OpenMax [10]. OpenMax [3] will be further detailed in the methodology (Section 3.1), as it is the basis for one of the proposed methods in this paper.

Inspired by adversarial attacks [12], Out-of-Distribution Detector for Neural Networks (ODIN) [23] insert small perturbations in the input image x in order to increase the separability in the SoftMax predictions between in- and out-of distribution data (\mathcal{D}^{in} and \mathcal{D}^{out} , respectively). This separability allows ODIN to work similarly to OpenMax [3] and operate close to the label space, using a threshold over class probabilities to discern between KKC and UUCs. The manuscript re-

ports Area Under ROC curve metrics between 0.90 and 0.99 for CIFAR-10 [19] as \mathcal{D}^{in} and between 0.70 and 0.85 for CIFAR-100 [19] as \mathcal{D}^{in} , depending on the \mathcal{D}^{out} (i.e. TinyImageNet¹, LSUN [45], random noise, etc). Extensive hyperparameter tuning experiments are reported in the paper as well. As will be further discussed in Section 3.2, restricting the information used for OSR to the activations in the last layers has severe limitations. Thus, modern methods have employed different strategies than simply thresholding the output probabilities to split samples into KKC and UUCs.

A recent trend in both Anomaly Detection and OSR for deep image classification has been to incorporate input reconstruction error in supervised DNN training as a way to identify OOD samples [44, 30, 39]. These approaches fall under the branch of generative OSR, according to the taxonomy by Geng *et al.* [11]. Classification-Reconstruction learning for Open-Set Recognition (CROSR) [44] trains conjointly a supervised DNN for classification model ($x \rightarrow y$) and an AutoEncoder (AE) to encode the input (x) into an bottleneck embedding (z) and then decode it to reconstruct \tilde{x} . Conjoint training allows the DNN to optimize a compound loss function that minimizes both the classification and reconstruction errors. During the test phase, the reconstruction error magnitude between $err(x, \tilde{x}) = ||x - \tilde{x}||$ dictates if the input x is indeed from the predicted class \hat{y} or an OOD sample.

Class Conditional AutoEncoder (C2AE) [30], similarly to CROSR, uses the reconstruction error of the input ($||x - \tilde{x}||$) from an AE and EVT modeling to determine a threshold in order to discern between KKC and UUC samples. Following the same trend of thresholding a certain point in the density function of the reconstruction error from the inputs, Conditional Gaussian Distribution Learning (CGDL) [39] uses a Variational AutoEncoder (VAE) to model the bottleneck representation of the input images according to a vector of gaussian means μ and standard deviations σ in a lower-dimensional high semantic-level space. This modeling allows CGDL to unsupervisedly discriminate between KKC and UUCs by thresholding the likelihood of the embedding z_i generated from a novel sample x_i pertaining to the multivariate gaussians $\mathcal{N}(z_i, \mu_k, \sigma_k^2)$, where k represents the predicted class for sample x_i .

2.2.1 Open Set Semantic Segmentation

OpenPixel [6] is based on patchwise training of classification DNNs for image classification [28] of RS images. The method builds on top of OpenMax [3] in order to recognize out-of-distribution pixels in urban scenarios. However, OpenPixel is highly inefficient during both training and test times due to the patchwise training using a customly built CNN.

As we have discussed in the introduction, to the best of our knowledge, the unique work in the literature that address the Open Set segmentation problem was proposed by Silva *et al* [6]. The authors introduce the concept and proposes two methods based on OpenMax [3] for pixelwise classification. Although promising, the approaches proposed in [6] have several limitations both in effectiveness and efficiency aspects. In this work we have extended the OpenPixel method proposed

¹ <https://tiny-imagenet.herokuapp.com/>

in [6] to be feasible in practical situations. We better explain the improvements and adaptations in Section 3.1.

As far as the authors are aware, there are no fully convolutional architectures for deep Open Set semantic segmentation in neither the remote sensing nor computer vision communities. Section 3 bridges this gap with the proposal of two approaches based on OpenMax [3] (Section 3.1) and Principal Component likelihood scoring [41] (Section 3.2) in the domain of urban scene segmentation.

3 Proposed Methods

This section details the two proposed methods presented in this work: 1) Open Fully Convolutional Network (OpenFCN), a fully convolutional extension of OpenMax [3, 6] for dense labeling tasks (Section 3.1); and 2) Open Principal Component Scoring (OpenPCS), a novel approach that uses feature-level information to fit multivariate gaussian distributions to a low-dimensional manifold of the data in order to obtain a score based on the data likelihood for identifying failures in recognition (Section 3.2).

3.1 OpenFCN

OpenFCN relies on traditional FCN-based architectures, which are normally composed of traditional CNN backbones with the inference layers replaced by bilinear interpolation and more convolutions. As the dense prediction is treated at training time as a classification task, the distinction between OpenFCN and FCN can be seen more clearly during validation and predicting. A meta-recognition module based on OpenMax [3] is added to the prediction procedure of traditional FCNs, as can be seen in Figure 3.

Let $\{X, Y\}$ be a paired set of image pixels and semantic labels from a dataset containing C KKC. A deep model \mathcal{M} can be trained in a stochastic manner by feeding samples X_i to a gradient descent optimizer as Adam [18] with a loss function as Cross Entropy, given by:

$$\mathcal{L}_{CE}(Y, \hat{Y}^{pri}) = -Y \log(\hat{Y}^{pri}) - (1 - Y) \log(1 - \hat{Y}^{pri}) \quad (1)$$

This strategy ultimately yields an activation $a_i \in \mathbb{R}^C$ for each pixel i after the last layer. Thus, \mathcal{M} can be seen as a function $\mathbb{R}^3 \rightarrow \mathbb{R}^C$ that converts the input space X_i into the prior SoftMax prediction σ_i for sample i . Obtaining the class prediction \hat{Y}_i^{pri} can be easily done by finding the class with the larger probability in σ_i across all KKC. Thus, OpenFCNs are trained using the exact same procedure as traditional FCNs for Closed Set semantic segmentation. Posteriori predictions \hat{Y}^{pos} are only computed on validation and testing, as described in the following paragraphs.

OpenMax (\mathcal{O}) relaxes the traditional SoftMax requirement that prediction probabilities for KKC must add to 1, introducing an extra class to the posterior prediction \hat{Y}^{pos} to the prediction set for X . The function $\mathcal{O}(X, Y, \mathcal{M})$ is, therefore, able to reweight the SoftMax predictions, aggregating the probability of misclassifications due to UUCs. Following the protocol of OpenMax [3], during OpenFCN's

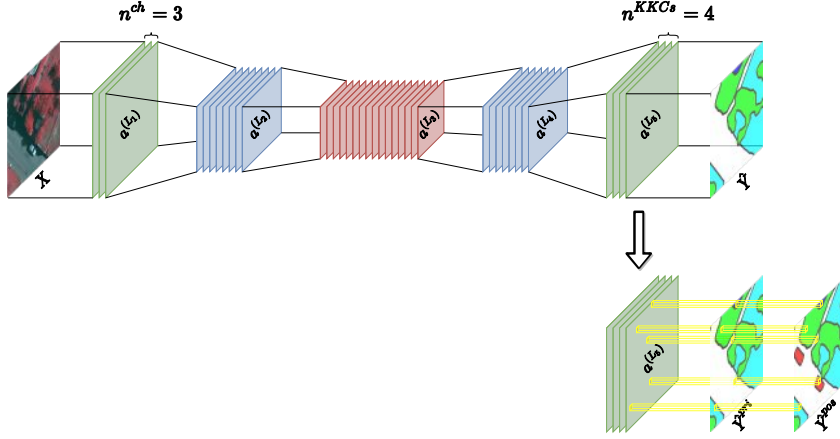


Fig. 3: OpenFCN scheme for Open Set Semantic Segmentation. During training, OpenFCN behaves like the traditional Closed Set FCN, with only Known Known Classes (KKCs) being fed to a supervised loss, such as Cross Entropy. OpenFCN differs from FCN only during validation and testing, when OpenMax is computed and the probabilities are thresholded in order to predict Unknown Unknown Classes (UUCs).

validation procedure each KKC $c_k, k \in \{0, 1, \dots, C - 1\}$ yields one Weibull distribution \mathcal{W}_k . \mathcal{W}_k is fit to the deviations from the mean μ_k of $a^{(L_5)}$ according to some distance (i.e. euclidean, cosine, hybrid distances). Averages μ_k are computed in the validation set according only to the correctly classified pixels of class c_k .

Finally, in order to identify Out-of-Distribution (OOD) samples, quantiles from the Cumulative Distribution Function (CDF) for \mathcal{W}_k are computed, with all pixels in the set \hat{Y}^{pri} predicted to be from c_k and with less confidence than \mathcal{T}_k being attributed to be from a UUC. Thus, the posterior OpenMax prediction \hat{Y}_i^{pos} for a specific pixel X_i is given by:

$$\hat{Y}_i^{pos} = \begin{cases} c_k, & \text{if } \max(a_i^{(l)}) \geq \mathcal{T}_k \\ c_{unk}, & \text{if } \max(a_i^{(l)}) < \mathcal{T}_k, \end{cases} \quad (2)$$

where l is the output layer in a DNN and c_{unk} is the identifier for the UUC in the Open Set scenario. This scheme is shown in Figure 4.

Dense labeling tasks (i.e. instance/semantic segmentation or detection) inherently have higher dimensional outputs than sparse labeling tasks (i.e. classification or single-target regression). Dense predictions also present their particular set of difficulties, including how to handle boundaries between adjacent objects. Early in our experiments, we have seen a large number of border artifacts in OpenFCN predictions. As depicted in Figure 5, adjacent areas between objects with distinct classes naturally yield class predictions with smaller certainties than the central pixels of these objects, resulting in warped Weibull distributions. This happens because the last layer in the network tries to model directly the label space

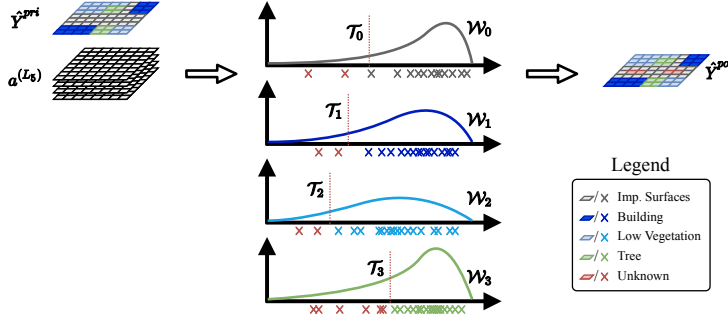


Fig. 4: OpenFCN’s generative modeling for detecting UUC samples. A Weibull model \mathcal{W}_c for KKC c is fit using the last layer’s activations ($a^{(L_5)}$) from correctly classified samples of this class in the training set. According to the CDF for each class’ Weibull distribution, a threshold τ_c is set and samples that do not reach this threshold are classified as pertaining from a UUC.

distribution, retaining little-to-no information about the original input. Hence, activations from later layers in the network are more affected by object boundary uncertainties.

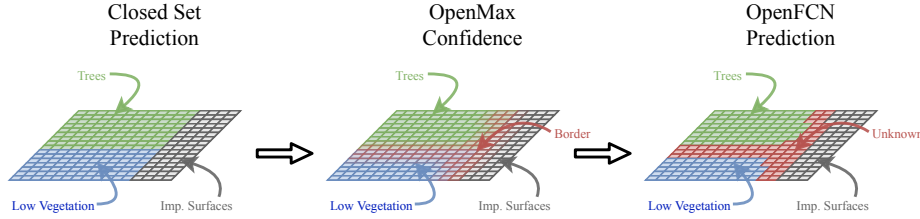


Fig. 5: Depiction of OpenFCN’s prediction confidence degradation on object boundaries due to the use of information close to the label space. SoftMax and OpenMax probabilities on dense labeling tasks are naturally lower on boundary regions between objects from distinct classes.

In order to mitigate this limitation of OpenFCNs in dense labeling tasks, we propose a completely novel method that is able to fuse information from low and high semantic level information into one single model. This method will be presented in Section 3.2.

3.2 OpenPCS

OpenPCS works similarly to CGDL [39], but with three important differences: 1) we fit gaussian priors using not only the input images X , but also the intermediary activations (i.e. $a^{(L_3)}$, $a^{(L_4)}$, $a^{(L_5)}$, etc); 2) we use a PCA instead of a VAE; and 3) the training is purely supervised and the low dimensional gaussians are fitted

only during the validation phase, and not conjointly with the training, as CGDL. This was all done for simplicity and aiming to ease computational complexity, as open set semantic segmentation is a very recent field of research.

It is well known that the deeper a certain layer l is placed in a DNN, the closer to the label space the activation features $a^{(l)}$ are [36]. In fact, Shwartz and Tishby [36] argue that any supervised DNN can be seen as a Markov chain of sequential tensorial representations that gradually morph the information processed by the network from the input space (in the input layer) to the label space (in the output layer). Thus, by using only the last layer’s activations to fit Weibull distributions to each KKC, OpenFCNs – and, by extension, OpenMax [3] – limit themselves to work with information close to the label space.

Unlike OpenFCN, OpenPCS takes into account feature maps from earlier layers, which encode information closer to the input space, and combine them with activations from the last layers, fusing low and high semantic level information. This can be seen in Figures 6 and 7 in the form of the yellow columns shown in the lower part of the image. Each output pixel in \hat{Y}^{pri} gets a correspondent activation vector (a^*) made by the concatenation of earlier layer activations for the corresponding prediction map region in the channel axis. As earlier layers ($a^{(L_4)}$ and $a^{(L_3)}$) have lower spatial resolution due to the network’s bottleneck, the activations from these layers are upsampled using the \uparrow^t function, where t indicates the number scaling factor. In the example shown in Figure 6, $a^* = (a^{(L_5)}, \uparrow^2 a^{(L_4)}, \uparrow^4 a^{(L_3)})$, so, for instance, if $a^{(L_5)} \in \mathbb{R}^{4 \times MN}$, $\uparrow^2 a^{(L_4)} \in \mathbb{R}^{8 \times MN}$ and $\uparrow^4 a^{(L_3)} \in \mathbb{R}^{16 \times MN}$, then a^* has dimensionality $28 \times MN$. The concatenated feature vector for each input/output pixel of index i in this example would be, therefore, $a_i^* \in \mathbb{R}^{(28)}$.

Concatenating activations from multiple layers into a^* yields high dimensionality feature vectors for each pixel, as modern CNNs/FCNs easily output hundreds or thousands of activation channels from each layer. The large redundancy found in activation maps from convolutional layers [38, 16] should also render a^* to be highly redundant. OpenPCS mitigates both problems by computing a lower dimension manifold a^{low} of each pixel’s activation with Principal Component Analysis (PCA) previously to fitting a generative model (\mathcal{G}), as shown in Figure 7. This approach grants two desirable properties to OpenPCS: 1) faster inference time during testing, as PCA implementations can be highly parallelized via vectorial operations and low-dimensional gaussian likelihood scoring can be computed in a fast manner; and 2) PCA feature selection guarantees that only the most important activation channels are used to compute a scoring function to detect OOD samples and, consequently, UUCs.

3.2.1 Open Set Scoring with Principal Components

Besides being purely a tensorial operation used for dimensionality reduction, PCA can be seen as a probability density estimator with gaussian priors. As described by Tipping and Bishop [41], this allows PCA to be used as a generative model for novelty detection. In other words, the low dimensional Principal Components generated by PCA using a multivariate gaussian prior can yield likelihoods that allow for OOD recognition in new data.

PCA reduces dimensionality by finding latent variables composed of combinations of features in the original input space such that the reconstruction error when returning to this original space is minimized. This operation works by computing

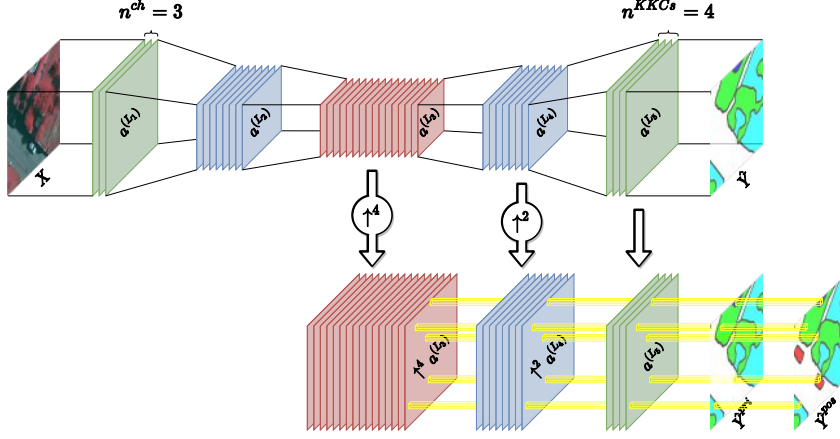


Fig. 6: OpenPCS general schematics. Subsequent activations ($a^{(L_1)}$, $a^{(L_2)}$, ..., $a^{(L_4)}$) from an FCN with a certain CNN backbone (as in Figure 2) is shown. Activations from the last layers (i.e. $a^{(L_5)}$, $a^{(L_4)}$ and $a^{(L_3)}$, in this case) are concatenated to form column vectors for each predicted output pixel in \hat{Y}^{pri} . The prior prediction \hat{Y}^{pri} is then processed according to the scheme shown in Figure 7 using a generative model \mathcal{G} in order to detect OOD pixels and, thus, classify them as unknowns.

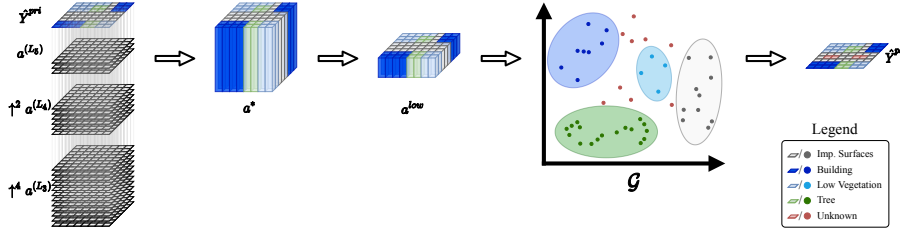


Fig. 7: Graphical scheme for the generative modeling in OpenPCS. In contrast to OpenFCN's Weibull fitting, OpenPCS uses a gaussian modeling in a low-dimensional representation a^{low} of the activations a^* . The Principal Components for each pixel are computed according to the concatenation of activations a^* from multiple layers (i.e. $a^{(L_3)}$, $a^{(L_4)}$, $a^{(L_5)}$, etc) for the corresponding region of each specific pixel. One multivariate gaussian is fit for each KKC and thresholds defined according to likelihoods from these models are used in order to identify OOD pixels.

eigenvalues (λ) and eigenvectors (v) of a covariance matrix (A). Those values can be calculated using the equation $Av = \lambda v$. In the PCA procedure, the eigenvalues represent how impacting their correspondent eigenvector direction is to the data variability. Since reconstructions provided by PCA emphasize variation, in order to perform dimensionality reduction, the standard procedure is to project your

data into an orthogonal basis composed by the eigenvectors that have the biggest correspondent eigenvalues, i.o.w., a subspace with the highest variance possible.

Exemplifying the application of PCA in OpenPCS, first we compute a covariance matrix using feature vector a^* – that is, a concatenation of activations from different layers in a DNN for a specific set of pixels. After that, we calculate the eigenvalues and eigenvectors of this matrix and select the set of *low* eigenvectors (v^{low}) that have the correspondent *low* largest eigenvalues (λ^{low}). Finally, we use those eigenvectors as a basis and project our input vectors (a^*) on it, resulting in $a^{low} = a^* \cdot v^{low}$.

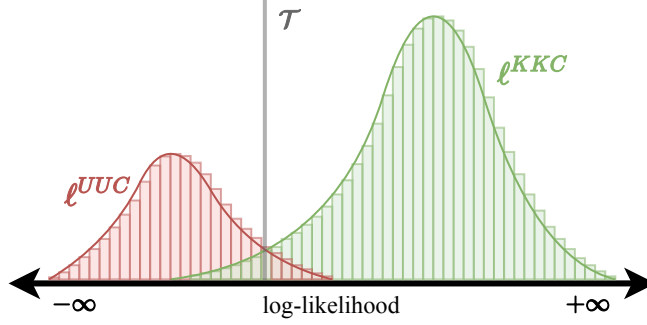


Fig. 8: Depiction of log-likelihood probability density functions. This toy example’s bimodal distribution is composed of two distinct peaks from the log-likelihoods of KKC’s (ℓ^{KKC}), which tend to have higher values, as the low dimensional gaussians were fit to them; and log-likelihoods from UUC samples (ℓ^{UUC}), which assume lower values, as they were not fed to the gaussian density estimator.

3.2.2 OpenIPCS

OpenPCS is highly memory intensive, as fitting n^{KKC} PCA models using millions of pixels with feature vectors in the scale of hundreds or thousands of bins requires all this data to be stored in RAM. In practice, for training the traditional OpenPCS we used a subsample of randomly selected patches to fit the models using a reasonable amount of memory. Empirically, we found that 150 patches with 224×224 pixels each resulted in an acceptable trade-off between memory and enough training data for the computation of Principal Components for each class. Even with this approach, OpenPCS required between 20 and 30 GB of memory, depending on the dimensionality of the feature vectors a^* . In addition, PCA models trained using subrepresented classes in the imbalanced datasets (i.e. Cars) did not fit correctly due to a low number of correctly classified samples used for training the generative model. Ultimately, OpenPCS for large and highly imbalanced datasets can result in underperformance on the task of UUC identification due to the subsampling required by the large memory usage of the method.

In an effort to minimize this problem, in addition to the traditional OpenPCS we also evaluate an Incremental PCA (IPCA) for generative modeling and

likelihood scoring. The full OSR segmentation methodology will be henceforth referred to as OpenIPCA. In contrast to the traditional offline training of the standard PCA, IPCA allows for mini-batch online training, which is highly useful for correctly computing the Principal Components in large datasets. OpenIPCA also allows for the model to be further updated in an efficient manner whenever new data might arise, not requiring a full retraining of the standard PCAs from scratch, but instead updating the IPCA model by feeding newly acquired patches. We emphasize that, apart from the incremental/online training that allows all the training dataset to be used in the computation of the Principal Components, all other aspects of the implementation of OpenIPCS were identical to the standard OpenPCS.

4 Experimental Setup

This section describes the experimental setup for the evaluation of OpenFCN against open and Closed Set baselines. In order to encourage reproducibility, we provide details regarding the datasets and evaluation protocol (Section 4.1), fully convolutional architectures and baselines (Section 4.2). In addition, we are publicizing the code for OpenFCN, OpenPCS and OpenIPCS in this project’s webpage² in a conscious effort to encourage reproducibility of our results and follow-up research on OSR segmentation. OpenFCN’s Weibull fitting and distance computation was based on libMR³ and OpenPCS’ Principal Components and likelihood scoring were computed using the scikit-learn⁴ library.

4.1 Datasets and Evaluation Protocol

In order to validate the effectiveness of OpenFCN and OpenPCS on RS image segmentation, we used two urban scene 2D semantic labeling datasets from the International Society for Photogrammetry and Remote Sensing (ISPRS) with pixel-level labeling: Vaihingen⁵ and Potsdam⁶. Vaihingen presents a spatial resolution of 5cm/pixel with patches ranging from 2,000~2,500 pixels in each axis and Potsdam has 9cm/pixel samples with 6,000 × 6,000 patches. Both datasets contain IR-R-G-B spectral channels paired with semantic maps divided into 6 KKC: impervious surfaces, buildings, low vegetation, high vegetation, cars and miscellaneous; and 1 KUC: segmentation boundaries between objects.

The data also allows for 3D information to be incorporated into the models via Digital Surface Model (DSM) images. However, in order to follow standard procedures in the RS literature [35, 2, 6], we ignored the DSM and blue channel in our evaluation, limiting the experiments to the IR-R-G channels. Aiming to ease the computation of our experiments, we also ignored the miscellaneous class, as it contains a rather small number of samples in Vaihingen. For both datasets we used the standard procedure in the literature of training with most patches, while

² <http://patreo.dcc.ufmg.br/2020/03/20/openfcn/>

³ <https://github.com/Vastlab/libMR>

⁴ <https://scikit-learn.org/stable/>

⁵ <http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-vaihingen.html>

⁶ <http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html>

separating some specific patches for testing: 11, 15, 28, 30 and 34 for Vaihingen; and 2_11, 2_12, 4_10, 5_11, 6_7, 7_8 and 7_10 in the case of Potsdam.

In order to simulate the Open Set environments on the Closed Set RS datasets, we employed a Leave-One-Class-Out (LOCO) protocol. LOCO works by selecting one class as UUC and ignoring its samples during training as shown in Figure 9. This protocol allows for verifying both the overall performances of the segmentation architectures and class-by-class metrics, as there is a large class imbalance mainly when comparing classes as impervious surfaces and cars in the datasets.

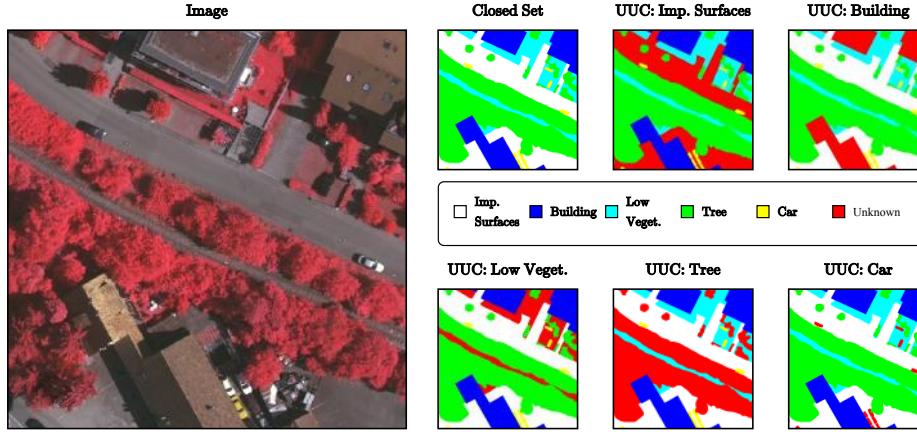


Fig. 9: LOCO procedure for Open Set evaluation. On the left one can see a mini-patch chosen from one of the larger patches that compose sample 11 of Vaihingen. On the right we present the Closed Set labels and different KKC's marked as Unknown using the LOCO protocol.

One should notice that all samples in the data used in our experiments pertain to Closed Set Semantic Segmentation datasets, which allows for an objective evaluation of the effect of OSR segmentation on well-known classification metrics. Aiming to capture all nuances of training and testing Open Set scenarios, we propose a new standard set of test metrics that accounts for KKC's and UUC's, while also capturing the overall performance over all known and unknown classes. In order to evaluate the impact of adding the meta-recognition step to the inference procedure of KKC's, we computed the accuracy for known classes (Acc^K). The main metric adopted by Silva *et al.* [6] for evaluating the performance of OSR segmentation on UUC's is the binary accuracy between KKC's and UUC's (Acc^U), as in the case of Anomaly Detection. However, Acc^U is artificially inflated due to the large imbalance between the number of pixels from KKC's and UUC in most cases. We solved this by adding the precision of unknown detection (Pre^U) to the evaluation in order to observe the trade-off in performance in the binary meta-recognition task. The Kappa (κ) metric was used to evaluate the overall performance in both KKC's and UUC of the algorithms, as it is a common metric for the evaluation of semantic segmentation in remote sensing. All of the metrics above are, however, threshold dependent, making it hard to compare methods

with distinct threshold ranges. As the thresholds do not represent equal statistical entities in SoftMax^T , OpenFCN and OpenPCS, we define the thresholds for the methods based on preset values of True Positive Rate (TPR). This evaluation protocol also has the benefit of yielding information about the recall of the methods, as TPR and recall are equivalent in this case. Also, in order to evaluate the methods in a threshold-independent fashion, we complement our analysis with the Receiver Operating Characteristic (ROC) curve – which evaluates the whole range of thresholds for the different methods on a TPR vs. False Positive Rate (FPR) 2D plot – and its corresponding Area Under Curve (AUC).

Due to the large spatial resolution of Vaihingen and Potsdam patches, we sliced them into 224×224 mini-patches for training, in order to fit ImageNet’s [7] traditional patch size, which is the most commonly expected input size for the CNNs used as backbones in our experiments. During training, we used data augmentation based on random cropping for patch selection, random flips in the horizontal and vertical dimensions and random rotations by multiples of 90 degrees. In order to solve inconsistencies due to lack of context and to compensate for patch border uncertainties, we employed overlapping patches of 224×224 pixels during testing. We explicit that this procedure considerably increased the computational budget of our experiments mainly in the Potsdam dataset for OpenFCN due to inefficiencies and non-vectorized operations in libMR. A more thorough evaluation of time complexity can be found in Section 5.2.5.

4.2 Fully Convolutional Architectures and Baselines

We compared several distinct fully convolutional architectures for OSR semantic segmentation, including five distinct CNN backbones: VGG [37], ResNet-50 [14], Wide ResNet-50-2 [46] (WRN), ResNeXt-50-32x4d [43] (ResNeXt) and DenseNet-121 [16].

Based on the methodology of OpenMax [3], we also compare OpenFCN and OpenPCS with traditional SoftMax followed by thresholding. SoftMax Thresholding (SoftMax^T) follows the premise that least certain network predictions may be motivated by outlier classes seen in test time. In addition to these Open Set approaches, we evaluate the proposed methods in comparison with traditional Closed Set fully convolutional architectures for dense labeling using the LOCO protocol. Closed Set FCNs wrongly classify UUCs by design, forcing the pixel to be segmented according to the higher probability prediction among KKC’s. One should notice that OpenPCS, OpenFCN, SoftMax^T and Closed Set FCNs were evaluated according to the same pretrained DNNs, differing only on test time. This protocol allows for direct comparisons in objective metrics disregarding performance variability due to the random nature of gradient descent optimization.

5 Results and Discussion

In this section we present and discuss the obtained results. First, an analysis is performed in Section 5.1 in order to define the most suitable architectures for the proposed approaches. This section contains both overall and per-class threshold-independent analysis. Section 5.2 compares the proposed methods with the base-

line in terms of quantitative metrics (Sections 5.2.2 and 5.2.3), qualitative segmentation maps (Section 5.2.4), and runtime performance (Section 5.2.5).

Note that, for the sake of simplicity and clarity, only the most relevant results were reported in this Section. For a full report of the results, please, check the supplementary material on this project’s webpage linked in Section 4.

5.1 Architecture Analysis

In this Section, we analyze the networks in order to define the most suitable architectures for the proposed techniques. To perform this evaluation, we employed the AUC, a threshold-independent performance measurement that allows comparisons between methods without resorting to (potentially) arbitrary thresholds. Furthermore, it is important to emphasize that we performed this analysis by using only the Vaihingen dataset. This is due to the fact that the Potsdam dataset is very similar to Vaihingen one and, therefore, analysis and decisions made over the latter dataset are also applicable to the former one.

Table 1 presents the average AUC results over all runs of the LOCO protocol. For SoftMax^T and OpenFCN, the ResNet-50 architecture [14] produced the best outcomes, followed closely by WRN-50 [46] and DenseNet-121 [16]. OpenPCS showed better results with WRN-50 [46], followed closely by DenseNet-121 [16] and VGG [37]. Since WRN [46] and DenseNet [16] were the most stable networks, producing good results in all approaches, they were selected and used in further experiments shown in the following sections.

Table 1: Average AUC values in the LOCO protocol for each evaluated FCN backbone in the **Vaihingen** dataset. Bold values indicate the best overall results including all methods.

Backbone	SoftMax ^T	OpenFCN	OpenPCS
WRN-50 [46]	0.68 ± 0.09	0.69 ± 0.10	0.82 ± 0.12
DenseNet-121 [16]	0.68 ± 0.08	0.68 ± 0.09	0.79 ± 0.12
ResNeXt-50 [43]	0.64 ± 0.06	0.65 ± 0.04	0.72 ± 0.12
ResNet-50 [14]	0.69 ± 0.12	0.71 ± 0.12	0.69 ± 0.16
VGG-19 [37]	0.67 ± 0.06	0.68 ± 0.07	0.77 ± 0.12

Based on the best results for Vaihingen, we also report the threshold independent average AUC analysis over all UUCs on Potsdam. Table 2 shows the results for DenseNet-121 and WRN-50 on the considerably larger Potsdam dataset, wherein extensive architectural comparisons were not feasible.

Table 2: Average AUC values in the LOCO protocol for **DenseNet-121** and **WRN-50** backbones in the **Potsdam** dataset. Bold values indicate the best overall results including all methods.

Backbone	SoftMax ^T	OpenFCN	OpenPCS	OpenIPCS
WRN	0.62 ± 0.06	0.62 ± 0.07	0.73 ± 0.13	0.73 ± 0.13
DenseNet	0.61 ± 0.11	0.62 ± 0.12	0.66 ± 0.25	0.71 ± 0.23

One can easily see that in both tables OpenPCS and OpenIPCS present better AUC results than OpenFCN and SoftMax^T, while the distinction between OpenFCN and SoftMax^T is often rather small or nonexistent. These evaluations are averages and standard deviations computed across all classes in the LOCO protocol, therefore, apart from a raw evaluation of the overall performance of each method, the previously mentioned.

5.2 Baseline Comparison

Based on the analysis performed in previous sections, we have conducted several experiments to investigate the effectiveness of the proposed methods using the most promising architectures: WRN-50 and DenseNet-121. We further investigate the performance on both KKC and UUCs of the proposed methods (OpenFCN and OpenPCS) and baseline (SoftMax^T) using the threshold-dependent metrics described in Section 4.1. This section also shows qualitative segmentation predictions from our experiments.

5.2.1 Per-Class Analysis

Table 3 presents the results, in terms of AUC per UUC, for the Vaihingen dataset. As can be seen, the OpenPCS obtained most of the best results when evaluating the UUCs and networks independently (bold values). When considering all the networks together and only the UUCs independently (values with †), OpenPCS produced the best outcomes for four UUCs while OpenFCN yielded the best result for one UUC.

Table 3: AUC values for each UUC of the **Vaihingen** dataset. Bold values indicate the best results when comparing SoftMax^T, OpenFCN and OpenPCS for each network and UUC. Results followed by † represent the best overall results for a certain UUC.

Backbone	Methods	UUC: Imp. Surf.	UUC: Building	UUC: Low Veg.	UUC: Tree	UUC: Car
DenseNet	SoftMax ^T	.78 ± .03	.63 ± .05	.73 ± .05	.67 ± .06	.58 ± .06
	OpenFCN	.81 ± .03	.64 ± .05	.72 ± .05	.66 ± .06	.58 ± .05
	OpenPCS	.84 ± .03	.94 ± .01†	.70 ± .08	.81 ± .07†	.65 ± .06
WRN	SoftMax ^T	.80 ± .03	.64 ± .04	.75 ± .05	.63 ± .06	.58 ± .05
	OpenFCN	.83 ± .03	.67 ± .03	.74 ± .05	.62 ± .06	.58 ± .05
	OpenPCS	.87 ± .03	.94 ± .02†	.77 ± .04	.63 ± .09	.87 ± .03†
ResNeXt	SoftMax ^T	.66 ± .06	.55 ± .07	.71 ± .07	.66 ± .04	.65 ± .04
	OpenFCN	.67 ± .06	.59 ± .05	.71 ± .07	.64 ± .03	.63 ± .04
	OpenPCS	.88 ± .03†	.81 ± .07	.58 ± .08	.69 ± .09	.65 ± .05
ResNet	SoftMax ^T	.83 ± .03	.62 ± .06	.81 ± .05†	.60 ± .05	.60 ± .05
	OpenFCN	.86 ± .02	.64 ± .05	.80 ± .05	.61 ± .05	.62 ± .05
	OpenPCS	.75 ± .03	.92 ± .02	.65 ± .07	.50 ± .07	.63 ± .05
VGG	SoftMax ^T	.72 ± .02	.59 ± .04	.74 ± .04	.66 ± .05	.66 ± .05
	OpenFCN	.74 ± .03	.59 ± .04	.74 ± .03	.65 ± .05	.65 ± .04
	OpenPCS	.82 ± .04	.90 ± .02	.71 ± .06	.60 ± .07	.81 ± .04

Analysis of Table 3 reveal that OpenPCS excels in almost all architectures and UUCs in the LOCO protocol. The overall best results for four out of five UUCs

Table 4: AUC values for each UUC of the **Postdam** dataset. Bold values indicate the best results when comparing SoftMax^T, OpenFCN, OpenPCS and OpenIPCS for each network and UUC. Results followed by † represent the best overall results for a certain UUC.

Backbone	Methods	UUC: Imp. Surf.	UUC: Building	UUC: Low Veg.	UUC: Tree	UUC: Car
DenseNet	SoftMax ^T	.66 ± .09	.49 ± .07	.53 ± .13	.64 ± .08†	.76 ± .02
	OpenFCN	.66 ± .09	.50 ± .06	.51 ± .13	.62 ± .07	.80 ± .03
	OpenPCS	.83 ± .06	.87 ± .13	.28 ± .03	.54 ± .11	.77 ± .04
	OpenIPCS	.84 ± .07†	.88 ± .14†	.41 ± .07	.54 ± .09	.91 ± .04†
WRN	SoftMax ^T	.68 ± .06	.53 ± .06	.60 ± .12	.62 ± .07	.66 ± .10
	OpenFCN	.69 ± .05	.54 ± .05	.58 ± .12	.60 ± .07	.69 ± .09
	OpenPCS	.66 ± .08	.86 ± .10	.71 ± .08	.57 ± .10	.86 ± .04
	OpenIPCS	.64 ± .09	.86 ± .10	.76 ± .07†	.55 ± .10	.83 ± .04

in the LOCO protocol (*Impervious Surfaces*, *Building*, *Tree* and *Car*) pertain to OpenPCS. Most AUCs for OpenPCS in these UUCs achieve values larger than 0.80, peaking at 0.94 in the class *Building*. Not coincidentally, these classes are also the most semantically and visually distinct from the other ones. More detailed insights regarding the nature of the superiority of OpenPCS over the other methods will be discussed further in Section 5.2.

The outlier in this pattern is the class *Low Vegetation*, as its samples present a much larger intraclass variability than the others, encompassing sidewalks, bushes, grass fields in the interior and exterior of buildings, gardens, backyards, some sections of parking lots and even train tracks partially covered with vegetation. Due to this higher variation in samples, the lower-dimensional multivariate gaussian fitting of OpenPCS was not able to properly map the whole variation in the data, as was the simple Weibull fitting of OpenFCN. Thus, SoftMax^T overcame both proposed methods and achieved a peak AUC of 0.81 on ResNet-50. Even though the best result in the UUC *Tree* was achieved by OpenPCS, the peak AUC for this class was also 0.81, this time using a DenseNet-121 backbone. The main visual distinction between *Trees* and *Low Vegetation* is not in the visible spectrum, though. While textures for most patches of both classes are rather similar when taking into account grass fields and bushes for *Low Vegetation*, the DSM data is more visually distinct than the IRRG data, as the peak altitude for tree tops are above the mostly plain areas of *Low Vegetation*. This implies that using the inputs channels (IRRG and DSM) coupled with the middle and later activations of the DNNs on the computation of PCAs for OpenPCS could have improved the performance of both classes. However, more research on this must be done to either confirm or deny this hypothesis.

To better demonstrate the differences in performance between the techniques, Figures 10 and 11 present the ROC curves for samples from the Vaihingen and Potsdam datasets, respectively. It is important to emphasize that each figure was created using a binary mask that separates KKC from the evaluated UUC. Through Figures 10, it is possible to observe that, in general, OpenPCS produces better results with lower FPRs for the same TPRs. It is also possible to visually assess that the detection capabilities of the methods for UUC *Impervious Surfaces* were considerably superior to other classes. The ROCs for *Buildings*, *Tree* and *Car* in Vaihingen also highlight the superiority of OpenPCS when compared to the other approaches.

Figure 11 provides distinct insights, as Potsdam is a much harder dataset than Vaihingen. Figures 11b and 11e show that OpenPCS and its online/incremental implementation (OpenIPCS) behaves dramatically better than both OpenFCN and SoftMax^T. OpenFCN and SoftMax^T stay mostly close to randomly selecting pixels in the images, which is delineated in the figures by the 45 degree dashed line in the plot. For *Building* and *Car*, the ROCs and AUCs show that SoftMax^T and OpenFCN achieved worse than random results, while OpenPCS and OpenIPCS reached close to or above 0.90 in AUC.

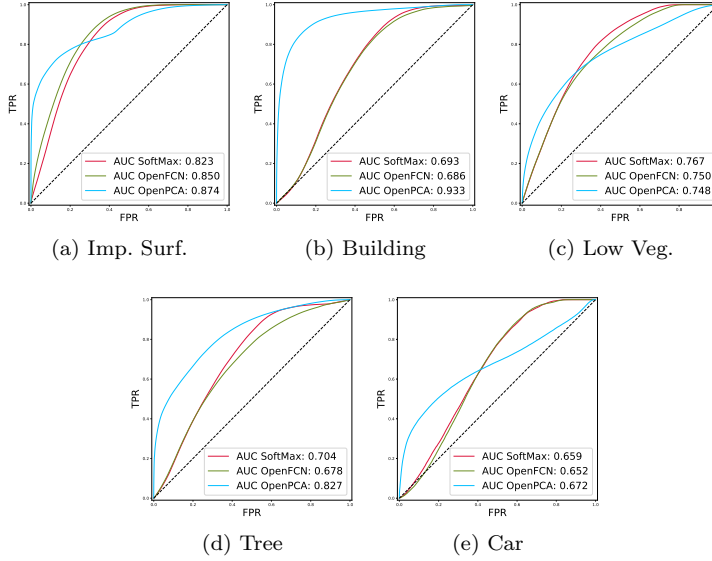


Fig. 10: ROC curves for SoftMax^T, OpenFCN and OpenPCS on sample 11 from **Vaihingen** using a **DenseNet-121** backbone.

Analysis of the per-class ROCs and Tables reveal that OpenPCS (and its variant OpenIPCS) behave considerably better than both SoftMax^T and OpenFCN in the binary task of Pixel Anomaly Detection for most UUCs. One can clearly see the distinction of the distinct methods in Pixel Anomaly Detection performance between UUCs in Figures 10 and 11. OpenPCS excels by a large margin on some UUCs, especially *Building* and *Tree* on Vaihingen, presenting lower FPR values than OpenFCN and SoftMax^T by almost any TPR threshold in The ROC. The evaluation on the other classes is more nuanced, as the plots cross at some points. OpenPCS tends to perform better on the lower end of FPR detections, providing the first evidence that OSR using Principal Components lessens the effect of UUC segmentation on the KKC classification performance, as smaller FPRs indicate that a smaller number of KKC pixels were misclassified as UUCs. This relation between UUCs and their distinction in performance on both KKC and UUCs will further explored in Sections 5.2.2 and 5.2.3.

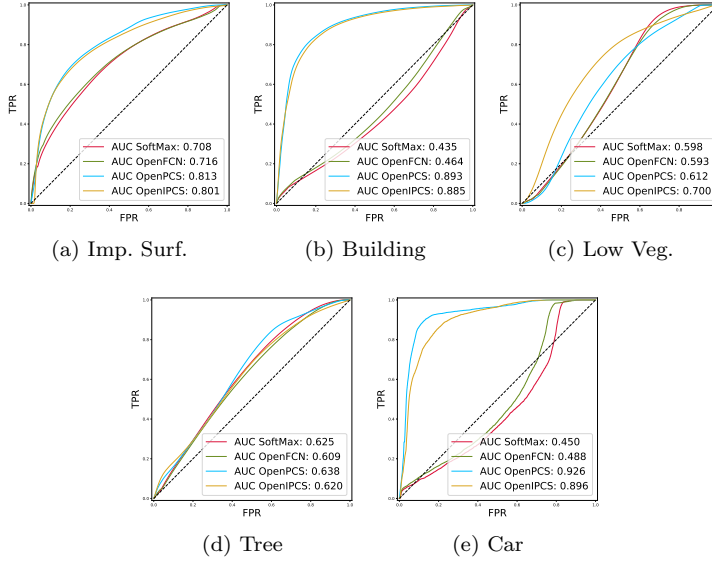


Fig. 11: ROC curves for SoftMax^T , OpenFCN, OpenPCS and OpenIPCS on sample 6_7 from **Potsdam** using a **WRN-50** backbone.

5.2.2 Performance on Vaihingen for KKC's and UUC's

OSR tasks are inherently multi-objective, as Open Set algorithms must be able to successfully discern UUCs from KKC's, while still being able to correctly classify samples from KKC's. Tables 5 and 6 show results from Acc^K , Pre^U and κ in the Vaihingen dataset using FCNs with WRN-50 and DenseNet-121 backbones, respectively.

Tables 5, and 6 show the performance of the OSR segmentation approaches on both KKC's and UUC's. The comparison of TPRs larger than one – that is, thresholds that allow for OSR – with their Closed Set counterparts ($\text{TPR} = 0$) using κ reveals that, in many cases, assuming openness improves object recognition in scenarios where full knowledge of the world is not possible. The most noticeable improvements in κ happen in UUC's that are visually and semantically more distinct from the other classes in the dataset: *Impervious Surfaces*, *Building*, *Tree* and *Car*. In the following paragraphs we will discuss the performance of the methods for each UUC in LOCO, starting from *Impervious Surfaces*.

Impervious Surfaces pixels in both Vaihingen and Potsdam are mainly composed of streets and driveways in residential/commercial buildings. These surfaces are considerably distinct from vegetation areas, as the “gray” signature in IRRG bands of asphalt is considerably distinct from the redness of plants due to their large reflection of near-infrared radiation. On Vaihingen, κ using fully closed DNNs achieve 0.53 with WRN-50 and 0.52 with DenseNet-121. By adding openness to these DNNs via OpenPCS, we were able to increase these κ metrics to 0.68 and 0.65, respectively, representing gains in κ between 0.13 and 0.15. The OSR methods allowed the methods to correctly delineate most streets and driveways in the

Table 5: Results for different unknown TPR thresholds applied to SoftMax^T, OpenFCN and OpenPCS from an FCN with **WRN-50** backbone in **Vaihingen**. Bold values indicate the best overall Acc^K , Pre^U and κ results for a certain UUC. “Closed” rows mean unknown TPRs of 0.0, which is equivalent to Closed Set segmentation.

UUC: Impervious Surfaces								
TPR	SoftMax ^T			OpenFCN			OpenPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.86±.01	.00±.00	.53±.04	.86±.01	.00±.00	.53±.04	.86±.01	.00±.00
0.10	.85±.01	.61±.11	.55±.03	.85±.01	.77±.07	.56±.04	.86±.01	1.00±.00
0.30	.81±.02	.52±.10	.57±.02	.83±.02	.63±.10	.59±.02	.86±.01	.95±.03
0.50	.76±.02	.50±.09	.59±.02	.78±.02	.55±.10	.61±.02	.84±.02	.85±.08
0.70	.70±.03	.47±.10	.59±.03	.72±.03	.50±.10	.62±.02	.78±.04	.64±.13
0.90	.56±.05	.42±.10	.52±.06	.59±.04	.44±.09	.56±.05	.53±.07	.41±.10
UUC: Building								
TPR	SoftMax ^T			OpenFCN			OpenPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.82±.02	.00±.00	.49±.05	.82±.02	.00±.00	.49±.05	.82±.02	.00±.00
0.10	.79±.02	.34±.07	.49±.05	.79±.02	.34±.08	.49±.05	.82±.02	1.00±.00
0.30	.73±.03	.36±.06	.49±.04	.73±.03	.36±.07	.49±.04	.82±.02	.99±.01
0.50	.65±.03	.36±.05	.48±.04	.65±.03	.36±.06	.49±.04	.82±.02	.93±.04
0.70	.52±.05	.33±.04	.42±.04	.54±.03	.35±.05	.44±.03	.80±.03	.85±.06
0.90	.23±.12	.28±.04	.20±.11	.34±.05	.31±.05	.30±.04	.71±.04	.64±.09
UUC: Low Vegetation								
TPR	SoftMax ^T			OpenFCN			OpenPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.92±.01	.00±.00	.61±.04	.92±.01	.00±.00	.61±.04	.92±.01	.00±.00
0.10	.91±.01	.47±.05	.62±.04	.90±.01	.46±.05	.62±.04	.90±.02	.46±.12
0.30	.86±.02	.45±.05	.64±.04	.86±.02	.44±.06	.63±.04	.84±.04	.47±.12
0.50	.79±.03	.43±.05	.63±.04	.77±.04	.41±.06	.61±.05	.77±.05	.44±.08
0.70	.66±.07	.39±.05	.56±.07	.63±.07	.36±.06	.53±.08	.67±.05	.41±.06
0.90	.44±.09	.32±.05	.40±.09	.43±.08	.32±.04	.39±.08	.49±.04	.35±.05
UUC: Tree								
TPR	SoftMax ^T			OpenFCN			OpenPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.88±.02	.00±.00	.54±.07	.88±.02	.00±.00	.54±.07	.88±.02	.00±.00
0.10	.81±.04	.22±.05	.50±.08	.81±.04	.23±.04	.51±.08	.88±.02	.88±.08
0.30	.70±.05	.28±.06	.46±.08	.69±.06	.27±.05	.45±.08	.81±.07	.62±.17
0.50	.62±.06	.32±.06	.45±.07	.60±.06	.31±.06	.43±.07	.61±.14	.37±.08
0.70	.53±.06	.35±.07	.43±.06	.51±.06	.34±.07	.41±.06	.38±.07	.29±.05
0.90	.40±.05	.34±.06	.37±.05	.37±.05	.33±.06	.33±.05	.17±.06	.28±.05
UUC: Car								
TPR	SoftMax ^T			OpenFCN			OpenPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.84±.02	.00±.00	.77±.02	.84±.02	.00±.00	.77±.02	.84±.02	.00±.00
0.10	.77±.05	.01±.01	.69±.05	.77±.05	.01±.01	.70±.04	.84±.02	.27±.06
0.30	.64±.06	.01±.01	.56±.06	.64±.06	.01±.01	.56±.06	.83±.02	.25±.05
0.50	.54±.05	.01±.01	.46±.05	.55±.06	.01±.01	.47±.05	.81±.02	.13±.04
0.70	.45±.04	.02±.01	.38±.04	.46±.05	.02±.01	.39±.04	.72±.04	.05±.02
0.90	.35±.04	.02±.01	.28±.03	.34±.04	.02±.01	.28±.03	.56±.05	.03±.01

samples – as can also be seen in a qualitative manner in Section 5.2.4. SoftMax^T and OpenFCN also improved the performance of the closed DNNs for *Impervious Surfaces*, albeit in a smaller magnitude when compared to OpenPCS, achieving gains of between 0.05 and 0.09.

Closed Set WRN-50 achieved an Acc^K of 0.86, followed closely by DenseNet-121 with 0.85 when *Impervious Surface* pixels were set as UUC in LOCO. For the same TPR, Acc^K was only barely affected by the openness of OpenPCS, maintaining their original levels of 0.86 and 0.85 up until a TPR of 0.30. In contrast, Acc^K was considerably sacrificed in SoftMax^T and OpenFCN for TPRs larger than 0.10, meaning that pixels that were correctly classified by the original DNNs were cast as UUCs by the OSR post-processing. OpenPCS maintained large Acc^K values

Table 6: Results for different unknown TPR thresholds applied to SoftMax^T, OpenFCN and OpenPCS from an FCN with **DenseNet-121** backbone in **Vaihingen**. Bold values indicate the best overall Acc^K , Pre^U and κ results for a certain UUC. “Closed” rows mean unknown TPRs of 0.0, which is equivalent to Closed Set segmentation.

UUC: Impervious Surfaces									
TPR	SoftMax ^T			OpenFCN			OpenPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.85±.02	.00±.00	.52±.04	.85±.02	.00±.00	.52±.04	.85±.02	.00±.00	.52±.04
0.10	.83±.03	.47±.11	.53±.04	.85±.02	.67±.12	.55±.04	.85±.02	.97±.02	.55±.04
0.30	.79±.03	.48±.10	.56±.02	.81±.03	.56±.11	.58±.02	.85±.03	.91±.07	.61±.03
0.50	.74±.03	.47±.11	.57±.02	.76±.03	.52±.11	.59±.02	.82±.04	.77±.14	.65±.03
0.70	.67±.04	.45±.10	.57±.04	.70±.04	.49±.11	.59±.03	.71±.09	.57±.15	.61±.09
0.90	.54±.05	.41±.09	.51±.06	.58±.04	.43±.10	.54±.05	.38±.06	.35±.08	.35±.06
UUC: Building									
TPR	SoftMax ^T			OpenFCN			OpenPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.83±.03	.00±.00	.50±.06	.83±.03	.00±.00	.50±.06	.83±.03	.00±.00	.50±.06
0.10	.77±.03	.26±.06	.48±.05	.78±.03	.28±.07	.48±.05	.82±.03	.98±.02	.53±.05
0.30	.70±.03	.32±.06	.47±.04	.71±.03	.32±.06	.47±.04	.82±.03	.96±.02	.59±.05
0.50	.62±.03	.33±.06	.45±.04	.62±.03	.34±.06	.46±.04	.82±.03	.94±.03	.65±.04
0.70	.50±.06	.33±.06	.40±.06	.51±.06	.33±.06	.41±.06	.80±.03	.87±.06	.70±.03
0.90	.30±.13	.30±.06	.26±.13	.29±.15	.30±.06	.25±.14	.71±.04	.64±.07	.68±.03
UUC: Low Vegetation									
TPR	SoftMax ^T			OpenFCN			OpenPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.92±.01	.00±.00	.61±.04	.92±.01	.00±.00	.61±.04	.92±.01	.00±.00	.61±.04
0.10	.90±.01	.45±.08	.62±.04	.90±.01	.46±.09	.62±.04	.92±.01	.70±.06	.64±.04
0.30	.85±.02	.43±.07	.63±.04	.85±.03	.43±.08	.63±.04	.88±.02	.56±.05	.66±.04
0.50	.77±.04	.41±.06	.60±.05	.76±.05	.40±.06	.60±.06	.77±.06	.44±.07	.61±.07
0.70	.64±.06	.37±.04	.54±.06	.61±.07	.35±.05	.51±.07	.55±.14	.33±.07	.45±.14
0.90	.42±.07	.31±.04	.38±.07	.39±.06	.30±.04	.35±.06	.23±.11	.26±.04	.19±.11
UUC: Tree									
TPR	SoftMax ^T			OpenFCN			OpenPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.87±.02	.00±.00	.53±.05	.87±.02	.00±.00	.53±.05	.87±.02	.00±.00	.53±.05
0.10	.84±.02	.33±.09	.53±.05	.84±.02	.34±.09	.53±.05	.87±.02	.97±.04	.56±.05
0.30	.77±.04	.37±.09	.53±.05	.77±.04	.37±.09	.53±.06	.86±.02	.82±.11	.62±.05
0.50	.68±.05	.37±.09	.51±.06	.68±.06	.37±.08	.50±.06	.80±.04	.64±.10	.63±.06
0.70	.57±.06	.37±.08	.46±.06	.54±.07	.35±.08	.44±.07	.70±.07	.50±.04	.60±.08
0.90	.35±.10	.33±.09	.31±.11	.30±.08	.31±.08	.27±.08	.47±.10	.38±.04	.43±.09
UUC: Car									
TPR	SoftMax ^T			OpenFCN			OpenPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.84±.02	.00±.00	.77±.03	.84±.02	.00±.00	.77±.03	.84±.02	.00±.00	.77±.03
0.10	.77±.04	.01±.01	.69±.04	.76±.04	.01±.00	.68±.04	.83±.02	.17±.08	.76±.02
0.30	.66±.06	.01±.01	.58±.06	.65±.05	.01±.01	.57±.05	.80±.04	.07±.03	.72±.04
0.50	.55±.07	.01±.01	.47±.06	.56±.05	.01±.01	.48±.05	.68±.07	.03±.01	.60±.08
0.70	.45±.06	.02±.01	.38±.06	.47±.06	.02±.01	.39±.05	.40±.10	.02±.01	.33±.09
0.90	.32±.05	.02±.01	.26±.04	.34±.05	.02±.01	.27±.04	.13±.05	.01±.00	.10±.04

up until TPRs of 0.50, with a large drop only being seen in a TPR of 0.70. These results imply that OpenPCS is more accuracy efficient when identifying OOD *Impervious Surface* pixels, barely sacrificing KKC segmentation performances. The precision of unknowns (Pre^U) is another evidence for the superiority of OpenPCS compared to OpenFCN and SoftMax^T, as it reaches a perfect precision of 1.00 in WRN-50 for a TPR (recall) of 0.10, followed closely by DenseNet’s Pre^U of 0.97 for the same TPR. Pre^U remains larger than 0.90 for both WRN-50 and DenseNet-121 up to a TPR of 0.30, meaning that UUC predictions for *Impervious Surfaces* from OpenPCS were highly reliable, even allowing for real-world applications of the method. Similarly to Acc^K , the precision of unknowns Pre^U was significantly

lower from the start on SoftMax^T and OpenFCN, with the best results for these methods peaking (with 0.10 TPR using WRN-50) on 0.61 and 0.77, respectively.

The UUC *Building* yielded the most drastic gains in κ when OpenPCS was introduced, as showed by the increase from 0.50 to 0.70 with the DenseNet-121 backbone (Table 6). *Building* pixels are certainly among the most visually distinct features in the IRRG images in both texture and shape. In addition to that, these structures are rather distinct in the DSM data, as they can be easily seen on depth maps as sudden variations in altitude in comparison with the surrounding terrain. Finally, context also plays a larger role in *Building* detection, as both commercial and residential structures reside close to driveways or parkways (both *Street* samples) and are often surrounded by vegetation in suburban areas as Vaihingen and Potsdam. Predictions for UUC *Building* using a TPR of 0.10 were also highly precise, with Pre^U values of 1.0 (perfect scoring) and 0.98 (near perfect) for WRN-50 and DenseNet-121, respectively.

Overall results from UUC *Tree* followed similar patterns to *Street* and *Building*, but in a smaller scale. κ gains using OpenPCS with WRN-50 and DenseNet-121 backbones were in the scale of 0.03 and 0.10, respectively, showing a noticeable boost in performance mainly in DenseNet-121. In contrast, the best κ values for OpenFCN and SoftMax^T were the Closed Set ones, with κ quickly dropping even since TPR 0.10 on WRN-50. The best Pre^U was also achieved by OpenPCS on both backbones, reaching a maximum of 0.97 for a TPR of 0.10 on DenseNet-121, while no noticeable drop in Acc^K could be seen in this setting. Acc^K remained relatively close to the Closed Set KKC classification performance up to 0.30 of TPR with DenseNet-121 and up to 0.10 on WRN-50.

Car pixels represent a tiny proportion of samples in both Potsdam and Vaihingen. Thus, analysis of Open vs. Closed Set methods by only using κ would yield basically no gain in adding OSR to the pipeline, as can be seen in Tables 5 and 6. This is due to the benefits of detecting UUCs not outweighing the added errors in KKC classification in metrics that measure general multiclass performance as Balanced Accuracy or κ . Instead, a more accurate picture of *Car* segmentation can be seen when taking into account Pre^U and Acc^K , together with the TPR, which is automatically given because it served to compute the thresholds for the methods. One can easily see that OpenPCS' Acc^K performance was barely degraded for TPR values up to 0.50 in WRN-50, while also remaining considerably close to the Closed Set accuracy in DenseNet-121 until a TPR of 0.30. In other words, there was virtually no harm to KKC classification in WRN-50 in exchange for the correct identification of 50% of *Car* pixels, while DenseNet-121 also kept high Acc^K values for 30% of the *Car* pixels being correctly segmented.

Pre^U values were not particularly high in either method, even for small TPRs, reaching a peak of 0.27 in WRN-50 with a TPR of 0.10 – that is, only approximately one in four pixels predicted to be pertaining to a vehicle in this setting was really from a vehicle. The low Pre^U of UUC *Car* can be explained by the relatively high intra-class variability in class *Building*, as most of the pixels misclassified as *Car* were, in fact, from parts of *Building* structures that were not common in the rest of the houses and warehouses in the data. We reached this conclusion after the qualitative evaluation shown in Section 5.2.4 that can be fully appreciated in the project's webpage. It is natural that houses have distinct shapes, sizes and rooftop textures, resulting. OpenPCS' Pre^U results, though indeed small, are still more than one order of magnitude larger than both OpenFCN and SoftMax^T.

Lastly, one important distinction between segmentation and detection must be highlighted in the case for *Car*: while many vehicles were not perfectly segmented, almost all automobiles were at least partially identified by OpenPCS. Qualitative results available at the project’s webpage show that in both datasets a large proportion of the automobiles were correctly identified by having a large subset of their pixels correctly identified as pertaining to a UUC. The same cannot be said for OpenFCN and SoftMax^T, which correctly detected (even if only partially) a much lower proportion of vehicles. In addition to that, one can see that even for small TPRs of 0.10 OpenFCN and SoftMax^T presented large drops to both Acc^K and κ performances.

5.2.3 Performance on Potsdam for KKC’s and UUC’s

Similarly to the presentation of the results from Vaihingen, Tables 7 and 8 show the same per-class threshold-dependent metrics (κ , Acc^K and Pre^U) in Potsdam for WRN-50 and DenseNet-121, respectively. As Potsdam is a considerably harder dataset when compared to Vaihingen – mainly due to the large imbalance and much larger spatial resolution – almost all metrics achieve lower values than the ones from Tables 5 and 6. We highlight that the standard OpenPCS with random patch selection for the training of PCA was also computed for Potsdam (as shown in Figure 11), but this section only reports OpenIPCS values. The incremental training of PCA showed to be much more stable in the larger Potsdam dataset, while we choose to not show the original OpenPCS results mostly due to spatial constraints and organization

Discussion regarding Potsdam results closely resembles the ones presented from Vaihingen on Section 5.2.2, albeit with one major distinction: relatively poor performance in comparison with Vaihingen due to the hardships encountered in processing Potsdam samples. These difficulties include the massive size of the dataset, restricting tuning of the networks’ hyperparameters due to computational constraints; the larger spatial resolution (6000×6000 per image) resulting from smaller physical areas covered by each pixel in comparison with Vaihingen, which also accentuate the downsides of introducing a larger correlation between adjacent pixels and providing less context on a 224×224 patch; and an extremely imbalanced set of labels.

One can indeed observe significant κ improvements by introducing OpenIPCS to the segmentation procedure in UUCs *Impervious Surfaces*, *Building*, *Tree* and *Car*, mainly with a DenseNet-121 backbone (Table 8). Closed Set κ values for these classes were 0.42, 0.41, 0.51 and 0.56, while the best OSR results (assuming optimal TPR choice) increased these values to 0.55, 0.57, 0.52 and 0.57. Analogously to Vaihingen, the increases to κ were larger for *Impervious Surfaces* and *Building* (between 0.13 and 0.16), while *Tree* and *Car* achieve much more modest gains of only 0.01, which are not statistically significant. *Building* again achieved a close to perfect Pre^U (0.96) using DenseNet-121 and TPR 0.10, followed by the also high *Impervious Surfaces* Pre^U of 0.87 also with 0.10 of TPR.

Low Vegetation again was an outlier in the LOCO protocol using DenseNet-121, with SoftMax^T and OpenFCN achieving less degradation in Acc^K and a larger Pre^U for the same TPR values. Since *Low Vegetation* fills a much larger proportion of pixels in Potsdam than in Vaihingen, it would be expected that adding OSR would considerably improve κ values in comparison with Closed Set

Table 7: Results for different unknown TPR thresholds applied to SoftMax^T, OpenFCN and OpenIPCS from an FCN with **WRN-50** backbone in **Potsdam**. Bold values indicate the best overall Acc^K , Pre^U and κ results for a certain UUC. “Closed” rows mean unknown TPRs of 0.0, which is equivalent to Closed Set segmentation.

UUC: Impervious Surfaces								
TPR	SoftMax ^T			OpenFCN			OpenIPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.73±.09	.00±.00	.40±.09	.73±.09	.00±.00	.40±.09	.73±.09	.00±.00
0.10	.71±.09	.58±.14	.41±.08	.72±.09	.70±.13	.42±.08	.71±.08	.54±.18
0.30	.66±.09	.46±.11	.43±.08	.68±.09	.51±.10	.44±.08	.64±.08	.47±.09
0.50	.59±.10	.40±.11	.41±.09	.60±.10	.42±.11	.42±.09	.54±.08	.40±.09
0.70	.47±.11	.35±.11	.37±.10	.48±.11	.36±.11	.37±.10	.39±.08	.33±.09
0.90	.29±.11	.31±.11	.25±.10	.28±.12	.31±.11	.25±.11	.16±.04	.28±.08
UUC: Building								
TPR	SoftMax ^T			OpenFCN			OpenIPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.77±.04	.00±.00	.44±.08	.77±.04	.00±.00	.44±.08	.77±.04	.00±.00
0.10	.74±.04	.34±.15	.43±.07	.75±.04	.38±.16	.44±.07	.77±.04	.90±.06
0.30	.64±.07	.29±.15	.40±.06	.65±.07	.30±.15	.40±.06	.76±.04	.82±.05
0.50	.52±.10	.28±.15	.35±.08	.54±.09	.29±.15	.36±.07	.74±.04	.77±.08
0.70	.37±.10	.27±.14	.26±.08	.39±.09	.28±.14	.28±.07	.71±.05	.69±.14
0.90	.16±.07	.26±.13	.13±.06	.18±.08	.26±.13	.14±.07	.56±.13	.48±.20
UUC: Low Vegetation								
TPR	SoftMax ^T			OpenFCN			OpenIPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.86±.04	.00±.00	.46±.14	.86±.04	.00±.00	.46±.14	.86±.04	.00±.00
0.10	.81±.04	.37±.11	.45±.14	.81±.04	.35±.11	.44±.15	.83±.05	.55±.14
0.30	.71±.09	.39±.12	.43±.16	.69±.11	.38±.12	.41±.18	.78±.07	.56±.15
0.50	.60±.14	.40±.12	.40±.18	.58±.15	.39±.12	.38±.19	.72±.08	.55±.15
0.70	.49±.15	.40±.13	.37±.16	.47±.15	.39±.14	.35±.16	.64±.10	.53±.15
0.90	.34±.13	.39±.16	.30±.11	.31±.11	.38±.16	.27±.10	.50±.13	.47±.15
UUC: Tree								
TPR	SoftMax ^T			OpenFCN			OpenIPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.89±.05	.00±.00	.60±.07	.89±.05	.00±.00	.60±.07	.89±.05	.00±.00
0.10	.85±.05	.26±.04	.58±.07	.85±.05	.25±.04	.58±.07	.80±.06	.28±.17
0.30	.77±.05	.28±.05	.55±.08	.76±.05	.27±.05	.54±.08	.67±.09	.23±.09
0.50	.65±.08	.26±.05	.48±.11	.63±.10	.25±.06	.46±.12	.55±.11	.23±.07
0.70	.50±.12	.25±.05	.39±.13	.47±.13	.24±.05	.35±.14	.39±.13	.21±.06
0.90	.29±.11	.22±.04	.23±.11	.24±.11	.21±.04	.18±.10	.16±.09	.19±.04
UUC: Car								
TPR	SoftMax ^T			OpenFCN			OpenIPCS	
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U
Closed	.75±.07	.00±.00	.64±.09	.75±.07	.00±.00	.64±.09	.75±.07	.00±.00
0.10	.74±.08	.22±.13	.63±.10	.75±.08	.28±.16	.64±.09	.75±.07	.29±.15
0.30	.69±.12	.04±.03	.58±.12	.70±.12	.06±.03	.60±.12	.74±.07	.14±.08
0.50	.59±.13	.02±.02	.49±.12	.62±.13	.03±.02	.52±.12	.71±.06	.07±.05
0.70	.47±.13	.02±.01	.38±.11	.51±.12	.02±.01	.41±.10	.62±.07	.04±.03
0.90	.31±.13	.02±.01	.25±.10	.34±.12	.02±.01	.27±.09	.45±.09	.02±.02

when this class was set to UUC in the LOCO protocol. However, the best κ results were the Closed Set ones for all OSR methods on the DenseNet-121 backbone, implying that neither of the methods converged well enough on this architecture to compensate for the KKC prediction losses. *Low Vegetation* results were opposite on the WRN-50 backbone in Potsdam: κ increased by 0.06 (for TPR 0.70); the best Pre^U results were indeed from OpenIPCS, peaking at 0.56 with TPR 0.30 and Acc^K degradation was much slower than OpenFCN and SoftMax^T as TPRs were allowed to grow. These discrepancies between architectures highlights that, even though OpenPCS and OpenIPCS were found to be the current state-of-the-art for OSR segmentation, they can be highly dependent on the DNN architecture, hyperparameters and proved to be rather unstable, being still a work in progress.

Table 8: Results for different unknown TPR thresholds applied to SoftMax^T, OpenFCN and OpenIPCS from an FCN with **DenseNet-121** backbone in **Potsdam**. Bold values indicate the best overall Acc^K , Pre^U and κ results for a certain UUC. “Closed” rows mean unknown TPRs of 0.0, which is equivalent to Closed Set segmentation.

UUC: Impervious Surfaces									
TPR	SoftMax ^T			OpenFCN			OpenIPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.77±.08	.00±.00	.42±.10	.77±.08	.00±.00	.42±.10	.77±.08	.00±.00	.42±.10
0.10	.74±.08	.44±.11	.43±.09	.74±.08	.45±.11	.43±.09	.76±.08	.87±.11	.45±.10
0.30	.68±.09	.40±.14	.43±.09	.69±.09	.42±.14	.44±.09	.75±.09	.84±.10	.51±.09
0.50	.59±.11	.37±.14	.41±.10	.60±.10	.37±.14	.42±.10	.73±.10	.76±.12	.55±.09
0.70	.47±.14	.35±.14	.36±.13	.47±.14	.35±.14	.36±.13	.66±.10	.63±.13	.55±.09
0.90	.29±.18	.31±.13	.25±.17	.28±.18	.31±.13	.25±.16	.46±.12	.42±.11	.40±.09
UUC: Building									
TPR	SoftMax ^T			OpenFCN			OpenIPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.74±.08	.00±.00	.41±.07	.74±.08	.00±.00	.41±.07	.74±.08	.00±.00	.41±.07
0.10	.71±.08	.37±.17	.41±.06	.72±.08	.40±.18	.42±.06	.74±.08	.96±.05	.44±.07
0.30	.61±.10	.30±.17	.37±.06	.62±.09	.32±.17	.38±.06	.74±.08	.91±.11	.49±.06
0.50	.46±.12	.27±.15	.29±.09	.47±.11	.28±.15	.30±.08	.73±.08	.85±.13	.55±.06
0.70	.25±.14	.25±.14	.16±.11	.26±.12	.25±.14	.17±.10	.69±.09	.73±.17	.57±.09
0.90	.09±.07	.24±.13	.06±.06	.07±.05	.24±.12	.05±.04	.55±.22	.53±.21	.50±.22
UUC: Low Vegetation									
TPR	SoftMax ^T			OpenFCN			OpenIPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.83±.03	.00±.00	.44±.14	.83±.03	.00±.00	.44±.14	.83±.03	.00±.00	.44±.14
0.10	.78±.04	.35±.13	.43±.15	.78±.05	.34±.13	.42±.15	.74±.06	.24±.06	.38±.17
0.30	.65±.13	.36±.14	.37±.19	.63±.15	.35±.14	.36±.21	.57±.09	.28±.11	.30±.17
0.50	.53±.16	.36±.14	.33±.19	.51±.16	.35±.14	.31±.19	.40±.08	.30±.14	.20±.13
0.70	.39±.16	.36±.15	.27±.17	.36±.15	.35±.15	.24±.16	.21±.06	.31±.15	.10±.08
0.90	.20±.12	.35±.16	.16±.12	.18±.10	.35±.16	.14±.09	.06±.02	.32±.16	.02±.03
UUC: Tree									
TPR	SoftMax ^T			OpenFCN			OpenIPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.80±.10	.00±.00	.51±.11	.80±.10	.00±.00	.51±.11	.80±.10	.00±.00	.51±.11
0.10	.77±.10	.25±.05	.50±.11	.77±.09	.26±.07	.50±.10	.79±.09	.34±.12	.52±.10
0.30	.69±.10	.25±.05	.47±.11	.69±.09	.25±.06	.48±.10	.69±.09	.26±.08	.47±.13
0.50	.59±.11	.25±.05	.42±.11	.59±.09	.25±.05	.43±.10	.54±.13	.23±.07	.36±.16
0.70	.49±.13	.25±.05	.38±.12	.47±.11	.24±.05	.36±.10	.31±.13	.19±.05	.19±.12
0.90	.39±.12	.25±.04	.34±.10	.32±.10	.23±.04	.27±.09	.10±.06	.18±.04	.06±.04
UUC: Car									
TPR	SoftMax ^T			OpenFCN			OpenIPCS		
	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ	Acc^K	Pre^U	κ
Closed	.68±.11	.00±.00	.56±.13	.68±.11	.00±.00	.56±.13	.68±.11	.00±.00	.56±.13
0.10	.68±.12	.15±.10	.56±.14	.68±.11	.16±.09	.56±.13	.68±.11	.58±.29	.57±.13
0.30	.65±.12	.05±.03	.53±.13	.65±.11	.06±.03	.54±.13	.68±.11	.35±.19	.57±.14
0.50	.60±.12	.03±.02	.49±.13	.62±.10	.04±.02	.51±.12	.67±.11	.20±.11	.56±.14
0.70	.52±.10	.03±.01	.42±.10	.56±.09	.03±.02	.45±.10	.65±.11	.10±.06	.54±.13
0.90	.38±.08	.02±.01	.30±.07	.44±.08	.02±.01	.34±.08	.58±.10	.05±.03	.47±.11

For almost all UUCs, Acc^K degraded much faster on Potsdam than on Vaihingen, as larger TPRs were evaluated. One important outlier in almost all aspects presented previously was *Car* on Potsdam. While on Vaihingen the κ did not increase at all in comparison to the Closed Set DNNs with the introduction of OpenPCS, the κ value for *Car* had a slight, but noticeable, increase from 0.56 to 0.57. Pre^U values for *Car* segmentation were also considerably larger on Potsdam, peaking at 0.58 for TPR 0.10 using DenseNet-121 and 0.29 using WRN-50. At the same time, Acc^K remained close to the Closed Set counterpart with TPRs up to 0.70, only observing a noticeable decrease in KKC classification performance on TPR 0.90. As for OpenFCN and SoftMax^T, Pre^U results for *Car* were also higher and Acc^K also degraded slowly with the increase of TPR (comparing with results

from Vaihingen), even if these metrics did not achieve the same gains as the ones from OpenPCS.

5.2.4 Qualitative Analysis

Figures 12 and 13 present some visual examples of the results generated by the FCN with DenseNet-121 backbone in the Vaihingen and Potsdam datasets respectively. Qualitatively, the effectiveness of OpenPCS to distinguish KKC from UUCs is even clearer. As can be observed, OpenPCS is capable of producing more accurate UUC identification for both KKC and UUCs, when compared to the ground-truth label. On the other hand, outcomes generated by the SoftMax^T and OpenFCN are very similar to each other, but not very alike to the ground-truth label, mainly for the UUCs. This corroborates with previous analysis and conclusions about the effectiveness of the OpenPCS mainly for discriminating UUCs.

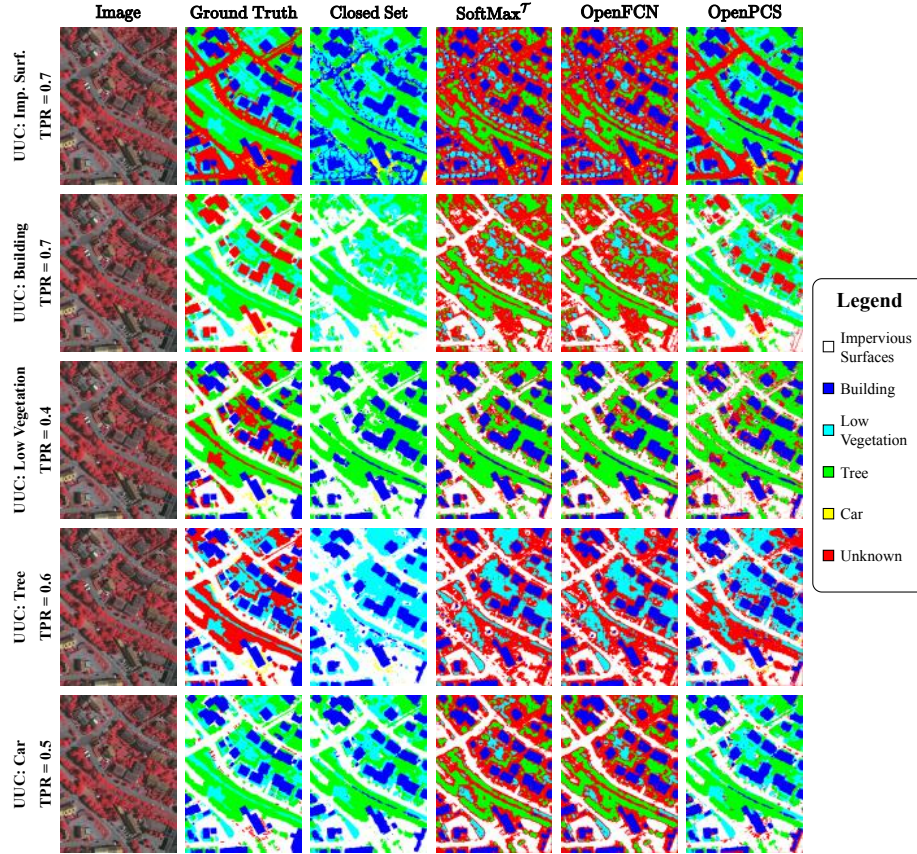


Fig. 12: Some visual result samples obtained for the **Vaihingen** dataset according to distinct UUCs in the LOCO protocol and distinct TPR thresholds for SoftMax^T , OpenFCN and OpenPCS.

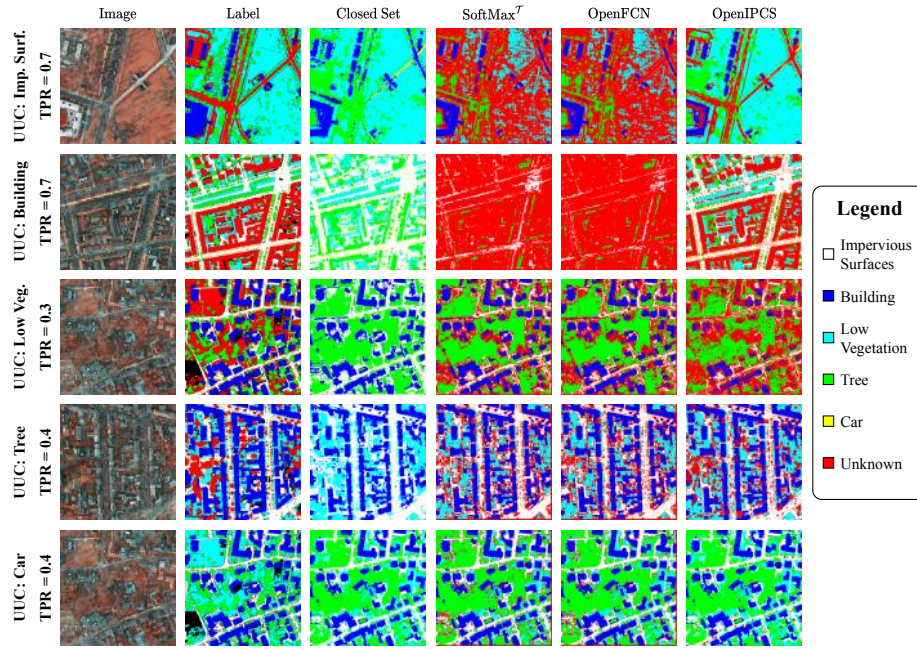


Fig. 13: Some visual result samples obtained for the **Potsdam** dataset according to distinct UUCs in the LOCO protocol and distinct TPR thresholds for SoftMax^T , OpenFCN and OpenIPCS.

Performing a deep analysis of the qualitative results, we can see that the proposed OpenPCS identified the UUC *Building* almost perfectly, while the SoftMax^T and OpenFCN techniques quite confused this UUC with other KCCs using the same threshold TPR. This same outcome can be seen for the UUC *Car*. Although this UUC comprises only a tiny percentage of the total amount of pixels and does not contribute a lot in terms of κ , it was much better identified by the OpenPCS than by other approaches. In fact, the above outcomes are repeated for all UUCs, except for the *Low-Vegetation* one, which has enormous intra-class variability (with pixels of grass fields, sidewalk-like areas and other structures) and, consequently, natural erratic behaviour. For a more detailed qualitative analysis, please, check the project’s webpage.

5.2.5 Runtime Performance Analysis

One of the most important motivations for proposing OpenPCS was the observation that OpenMax did not scale well to dense labeling, being prohibitively expensive when operating in a pixelwise fashion, while the simple SoftMax^T performed exponentially faster. In order to properly quantify the time differences between SoftMax^T , OpenFCN and OpenPCS, we computed the per-patch run-times of each method for 224×224 patch resolution. These results are shown in Figure 14 for both Vaihingen and Potsdam (Figures 14a and Figures 14b, respectively) and these same results are also presented in \log_{10} scale (Figures 14c and

Figures 14d, respectively), as the linear time comparisons severely hampered the visualization of SoftMax^T 's performance when compared to OpenFCN.

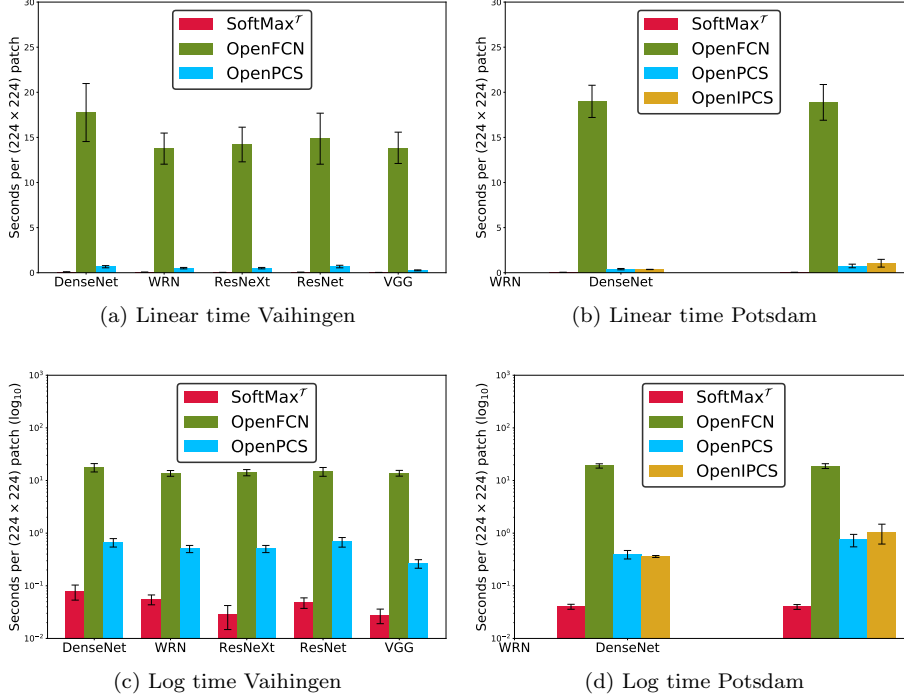


Fig. 14: Per-patch time comparison between the proposed approaches. The time presented in the y -axis is shown in seconds for each (224×224) patch across all test patches. The left part of the figure (a, c) shows times for Vaihingen, while the rightmost figures (b, d) depict times for Potsdam. Results are also shown in linear (a, b) and \log_{10} scale (c, d) in order to show the exponential distinction between execution times across SoftMax^T, OpenFCN and OpenPCS. Confidence intervals for each plot are shown as error bars computed with the t-Student distribution on the average execution runtimes for patches over 5 runs of the LOCO protocol (one for each class set as UUC) for each backbone.

Visual analysis of Figure 14 reveals the discrepancies between OSR methods, with SoftMax^T inference being the fastest, usually taking between 0.03 and 0.1 second per 224×224 patch. On the opposite side, OpenFCN was observed to be by far the slowest method, with runtimes for one single patch in the range between 10 and 20 seconds. OpenFCN's runtime makes it highly unsuitable for real-time computer vision applications.

On the faster end of the spectrum between SoftMax^T and OpenFCN, there were OpenPCS and OpenIPCS, with execution times between 0.25 and 0.7 seconds per patch. Both the online and offline PCAs used for the inference of UUCs from the proposed methods are highly vectorized operations, which allows them

to be parallelized into several processing cores and be faster even in single-core architectures. While we did not tune the algorithms for this purpose, OpenPCS' and OpenIPCS' runtimes allow for near real-time inference on applications as self-driving cars or autonomous drone control. In contrast, the original implementation of OpenMax from libMR is not naturally vectorized, requiring inferences to be performed linearly on pixels and severely hampering OpenFCN's performance.

6 Conclusion

In this manuscript we introduced, formally defined and explored the problem of Open Set Segmentation and proposed two approaches for solving OSR segmentation tasks: 1) OpenFCN, which is based on the well-known OpenMax [3] approach; and 2) OpenPCS, a completely novel method for Open Set Semantic Segmentation. A comprehensive evaluation protocol based on a set of standard objective metrics to be used on this novel field that takes into account KKC's and UUC's was proposed, executed and discussed for the proposed methods and the main baseline: SoftMax^T.

OpenPCS – and its more scalable variant referred to as OpenIPCS – yielded significantly better results than OpenFCN and SoftMax^T, suggesting that using middle-level feature information is highly useful for OSR segmentation. OpenFCN showed relatively poor performance in most experiments, as it is limited to using the activations of the last layer to detect OODs. This scheme works well for sparse labeling scenarios [3] (i.e. image classification), however, dense labeling inherently consists of highly correlated samples – in this case, pixels – and labels – in this case, semantic maps – with border regions naturally presenting lower confidence than the interior pixels of objects. This distinction between sparse and dense labeling severely hampered the capabilities of OpenFCN and SoftMax^T of detecting OOD samples.

As expected, UUCs with larger semantic and visual distinctions in relation to the remaining KKC's also showed considerably better classification performance in the LOCO protocol. Specifically, *Low Vegetation* samples have a high intra-class variability, representing grass fields, sidewalk-like areas and other structures in the Vaihingen and Potsdam datasets, rendering them similar to samples from the classes *Street* and *Tree*. This resulted in lower performances in almost all metrics for all methods assessed in our experimental procedure. Closed Set predictions by the DNN architectures evaluated in Section 5.2 generally resulted in lower overall κ results than when OSR was added in the form of OpenPCS mainly, with gains up to 0.20 in this metric for the *Building* UUC. The analysis conducted in Section 5.2.5 also revealed the discrepancies between OSR methods in inference runtime, with SoftMax^T being the fastest method, OpenFCN being the exponentially slower and OpenPCS being a reasonable compromise, being able to run close to real-time, depending on GPU memory for larger patches and/or image size.

As OpenFCN is simply a fully convolutional version of OpenPixel [6], also based on OpenMax [3], we can confidently infer that even OpenFCN performed better than OpenPixel. That is because OpenPixel was based on a traditional CNNs trained in a patchwise fashion for the classification of the central pixel and fully convolutional training was already shown to improve both segmentation performance and dramatically improve runtime efficiency of patchwise approaches in

the pattern recognition literature [25]. While OpenFCN differs in this aspect, it even uses the same meta-recognition lib and some of the code of OpenPixel for the computation of OpenMax. One important variation of OpenPixel’s approach that still needs to be verified in the context of fully convolutional training is the post processing using morphology, called Morph-OpenPixel by Silva *et al.* [6]. This post-processing was shown to considerably improve the performance of OpenPixel. At last, even though time comparisons were not made available by Silva *et al.* [6], it is also possible to infer that OSR segmentation using even OpenFCN is exponentially faster than patchwise training also due to the experiments performed on the original FCN paper [25].

Approaches as CGDL [39] and C2AE [30] couple the training of the generative model (usually a VAE) with the supervised training for KKC classification. These approaches allow for and end-to-end training of DNNs for OSR image classification tasks, and are currently considered the state-of-the-art for this task. Future works for OpenPCS include the adaptation of these methods for Open Set Semantic Segmentation, which would incorporate low-, middle- and high-level semantic features into the recognition of OOD samples, possibly allowing for an end-to-end training of Pixel Anomaly Detection and OSR segmentation. We also aim to investigate the capabilities of other lighter and/or more robust generative models in the OpenPCS pipeline, such as One-Class SVM (OCSVM) [34,33] or Gaussian Mixture Models (GMMs) [4,1] instead of the simpler Principal Components used for computing OpenPCS likelihoods.

Authors also aim to perform future analysis by varying the Openness of the problems – that is, setting more than one class to UUC in a variation of the LOCO protocol. Early experiments indicate that the superiority of OpenPCS might be even more pronounced in settings with larger Openness and/or larger number of known classes.

Acknowledgments

The authors would like to thank NVIDIA for the donation of the GPUs that allowed the execution of all experiments in this paper.

References

1. Attias, H.: A variational bayesian framework for graphical models. In: Advances in neural information processing systems, pp. 209–215 (2000) [32](#)
2. Audebert, N., Le Saux, B., Lefèvre, S.: Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In: Asian conference on computer vision, pp. 180–196. Springer (2016) [6](#), [14](#)
3. Bendale, A., Boulton, T.E.: Towards open set deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1563–1572 (2016) [3](#), [6](#), [7](#), [8](#), [11](#), [16](#), [31](#)
4. Bishop, C.M.: Pattern recognition and machine learning. springer (2006) [32](#)
5. Cardoso, D.O., Gama, J., França, F.M.: Weightless neural networks for open set recognition. Machine Learning **106**(9-10), 1547–1567 (2017) [6](#)
6. da Silva, C.C.V., Nogueira, K., Oliveira, H.N., dos Santos, J.A.: Towards open-set semantic segmentation of aerial images. arXiv preprint arXiv:2001.10063 (2020) [3](#), [7](#), [8](#), [14](#), [15](#), [31](#), [32](#)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255. Ieee (2009) [4](#), [5](#), [16](#)
8. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision **111**(1), 98–136 (2015) [5](#)
9. Farabet, C., Couprie, C., Najman, L., LeCun, Y.: Learning hierarchical features for scene labeling. IEEE transactions on pattern analysis and machine intelligence **35**(8), 1915–1929 (2012) [5](#)
10. Ge, Z., Demyanov, S., Chen, Z., Garnavi, R.: Generative openmax for multi-class open set classification. arXiv preprint arXiv:1707.07418 (2017) [6](#)
11. Geng, C., Huang, S.J., Chen, S.: Recent advances in open set recognition: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020) [2](#), [6](#), [7](#)
12. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014) [6](#)
13. Guiolette, F., Pham, M., Dambreville, R., Corpetti, T., Lefvre, S.: Semantic segmentation of d points clouds: Rasterization beyond digital elevation models. IEEE Geoscience and Remote Sensing Letters pp. 1–4 (2020) [6](#)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016) [4](#), [16](#), [17](#)
15. Hendrycks, D., Mazeika, M., Dietterich, T.: Deep anomaly detection with outlier exposure. arXiv preprint arXiv:1812.04606 (2018) [2](#)
16. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708 (2017) [4](#), [11](#), [16](#), [17](#)
17. Kemker, R., Salvaggio, C., Kanan, C.: Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing **145**, 60 – 77 (2018). Deep Learning RS Data [6](#)
18. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) [8](#)
19. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009) [4](#), [7](#)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012) [4](#), [5](#)
21. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998) [4](#)

22. Li, F., Wechsler, H.: Open set face recognition using transduction. *IEEE transactions on pattern analysis and machine intelligence* **27**(11), 1686–1697 (2005) [2](#)
23. Liang, S., Li, Y., Srikant, R.: Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690* (2017) [6](#)
24. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: *European conference on computer vision*, pp. 740–755. Springer (2014) [5](#)
25. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440 (2015) [5](#), [32](#)
26. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: High-resolution aerial image labeling with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* **55**(12), 7092–7103 (2017) [6](#)
27. Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U.: Classification with an edge: improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing* **135**, 158–172 (2018) [6](#)
28. Nogueira, K., Dalla Mura, M., Chanussot, J., Schwartz, W.R., dos Santos, J.A.: Learning to semantically segment high-resolution remote sensing images. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 3566–3571. IEEE (2016) [6](#), [7](#)
29. Nogueira, K., Dalla Mura, M., Chanussot, J., Schwartz, W.R., dos Santos, J.A.: Dynamic multicontext segmentation of remote sensing images based on convolutional networks. *IEEE Transactions on Geoscience and Remote Sensing* **57**(10), 7503–7520 (2019) [6](#)
30. Oza, P., Patel, V.M.: C2ae: Class conditioned auto-encoder for open-set recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2307–2316 (2019) [6](#), [7](#), [32](#)
31. Pinheiro, P.H., Collobert, R.: Recurrent convolutional neural networks for scene labeling. In: *31st International Conference on Machine Learning (ICML), CONF* (2014) [5](#)
32. Scheirer, W.J., Jain, L.P., Boulton, T.E.: Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence* **36**(11), 2317–2324 (2014) [2](#), [6](#)
33. Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., Boulton, T.E.: Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence* **35**(7), 1757–1772 (2012) [6](#), [32](#)
34. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. *Neural computation* **13**(7), 1443–1471 (2001) [32](#)
35. Sherrah, J.: Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery. *arXiv preprint arXiv:1606.02585* (2016) [14](#)
36. Shwartz-Ziv, R., Tishby, N.: Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810* (2017) [11](#)
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014) [4](#), [16](#), [17](#)
38. Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. *arXiv preprint arXiv:1505.00387* (2015) [4](#), [11](#)
39. Sun, X., Yang, Z., Zhang, C., Peng, G., Ling, K.V.: Conditional gaussian distribution learning for open set recognition. *arXiv preprint arXiv:2003.08823* (2020) [7](#), [10](#), [32](#)
40. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9 (2015) [4](#)
41. Tipping, M.E., Bishop, C.M.: Mixtures of probabilistic principal component analyzers. *Neural computation* **11**(2), 443–482 (1999) [8](#), [11](#)
42. Wang, H., Wang, Y., Zhang, Q., Xiang, S., Pan, C.: Gated convolutional neural network for semantic segmentation in high-resolution images. *Remote Sensing* **9**(5), 446 (2017) [6](#)
43. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500 (2017) [4](#), [16](#), [17](#)
44. Yoshihashi, R., Shao, W., Kawakami, R., You, S., Iida, M., Naemura, T.: Classification-reconstruction learning for open-set recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4016–4025 (2019) [7](#)
45. Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., Xiao, J.: Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365* (2015) [7](#)
46. Zagoruyko, S., Komodakis, N.: Wide residual networks. *arXiv preprint arXiv:1605.07146* (2016) [4](#), [16](#), [17](#)