Improving Deep Learning based Point Cloud Classification using Markov Random Fields with Quadratic Pseudo-Boolean Optimization

Qipeng Mei[1] , Kevin Qiu[2], Dimitri Bulatov[2], Dorota Iwaszczuk[1]

[1] Technical University of Darmstadt, Germany - (qipeng.mei,dorota.iwaszczuk)@tu-darmstadt.de
[2] Fraunhofer IOSB, Ettlingen, Germany - (kevin.qiu,dimitri.bulatov)@iosb.fraunhofer.de

Keywords: 3D, Semantic Segmentation, Urban, Outdoor Point Clouds, Class Imbalance.

Abstract

3D point clouds are a relevant source of information for multiple applications, including digital twins, building modeling, disaster and risk management, forestry, autonomous driving, and many others. Assigning points to the semantic classes is one of the essential data interpretation steps to effectively use them for further analysis. Deep learning models for semantic segmentation, such as RandLA-Net, are state-of-the-art methods for this task. Although the overall accuracy of classification is usually satisfactory,there are still several shortcomings not allowing assigning correct labels across all the classes. For instance, the receptive field of these networks is often too small to correctly classify point clouds in all cases. These networks suffer also from class imbalance, typical in real-world data sets, and tend to oversmooth small classes. Post-processing approaches help to overcome these problems and achieve better classification accuracy. In this work, we investigate the feasibility of improving the deep-learning outputs by introducing prior knowledge. To do this, the output probabilities of point classes obtained using RandLA-Net are post-processed with a workflow based on Markov Random Fields, in which the unary potentials are adjusted to preserve smaller classes while the pairwise potentials take into account. a hand-tailored inter-class reliability matrix. To validate our method, we apply it to the Hessigheim benchmark. Our MRF-based approach further optimizes these prediction results, effectively and efficiently improving the overall accuracy by approximately 1 to 2 percentage points.

1. Introduction

Multiple applications nowadays require a three-dimensional representation of the environment and its changes over time. Therefore, direct 3D measurements became the state-of-the-art technique to collect information about the world, like mobile sensing using different platforms such as drones, vehicles, or human-carried devices. Using Light Detection and Ranging (LiDAR) acquisition technique, point clouds with varying point density can be generated. These point clouds representing surface geometry can be enriched with radiometric information, such as LiDAR intensity and color information from additional cameras. This data representation can be used for visualization and manual interpretation. Considering, however, the large amounts of data, particularly for time series and outdoor measurements, automatic data processing and interpretation are necessary to make use of this data. Semantic segmentation is one of the techniques giving a sense to the data and allowing for automatic analysis in information models or digital twins. Here, each point is assigned to one class, which allows for identifying objects in 3D space. For this task, multiple methods have been developed in the last decades, whereas even the most advanced machine learning methods, like (Weinmann et al., 2015), relying on hand-crafted features, were gradually replaced by deep-learning-based methods. Semantic segmentation models based on deep learning became particularly efficient with the advancement ofcomputer processing power and related algorithms. Benefiting from their intricate neural network structures, they are capable of acquiring abilities similar to humans in perceiving the world. More specially, based on the provided data, they are able to form an understanding of specific tasks by continuously updating intrinsic parameters.

In the human learning process besides continuous practice, summarizing experiences to form knowledge is also a crucial aspect. Common-place knowledge enables humans to quickly familiarize themselves with unfamiliar environments. Similarly, we can utilize the common-place knowledge to rectify the predictions of the deep learning models to make them more in line with objective reality.

Therefore, in this work, we explore the feasibility of optimizing deep learning (DL) prediction results by introducing the prior knowledge with the assistance of Markov Random Fields (MRFs). We will see that advanced move-making methods, such as Quadratic Pseudo-Boolean Optimization (QPBO), have to be applied to deal with supermodular priors derived from common-place knowledge and formulated as a prior tailered to preserve smaller classes and an interclass reliability matrix.

1.1 Previous Work

DL has been state of the art for semantic segmentation of image data for almost one decade. Due to the inherent structure of point clouds, however, CNN-based methods from the image domain cannot be directly applied. Point clouds are sparse, unstructured, unsorted, and have non-uniform densities and occlusion effects (Ye et al., 2018), whereas image data are simply in a uniform two-dimensional pixel grid. Point clouds also tend to be quite large, especially in the remote sensing field. Their irregular shape makes it difficult to choose a sensible value for the receptive field. The first machine learning approaches on point clouds were projection-based, where normal convolutional networks and back-projection are used, or voxel-based, where 3D CNN layers are employed. However, projection and voxelization always result in loss of information. The breakthrough came in 2017 when PointNet (Qi et al., 2017a), the first network that directly operates on individual points, was introduced. Since then, state-of-the-art pipelines for 3D semantic segmentation, such as PointNet++ (Qi et al., 2017b), KP Conv

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-229-2024 | © Author(s) 2024. CC BY 4.0 License.

(Thomas et al., 2019), RandLA-Net (Hu et al., 2020), Bilateral Context-based segmentation (Qiu et al., 2021), and PointNeXt (Qian et al., 2022) and many others, mostly operate directly on the points. The receptive field of these networks is still frequently insufficient to consider large 3D objects, such as large flat roofs, which may be confused with the ground. This is especially the case in dense, high-resolution point clouds, which is why we consider further features, inspired by remote sensing and image processing, as additional inputs into the DL network. Recently, the research into transformer-based networks for point cloud classification has taken off, with networks such as Point Transformer (Zhao et al., 2021) or Point-MAE (Pang et al., 2022). While these approaches are, certainly, interesting, especially in combination with self-supervised learning, transformer-based networks are simply too resource-intensive. This is a problem for remote sensing, where point clouds have easily more than a hundred million points, so computational efficiency has to be accounted for.

Imbalanced data is also problematic in the sense that small classes are subject to oversmoothing. The DL methods incorporate inherent tools for increasing accuracy for smaller classes. One could perform data augmentation, improve data balancing, or employ alternative performance metrics, such as focal loss (Lehner et al., 2022, Lin and Nguyen, 2020, Sander, 2020). Data augmentation is a very wide scientific field, starting with manipulating (rotating, re-scaling) point clouds around objects belonging to small classes and ending at using generative techniques as well as synthetic examples, e. g. for cars (Lehner et al., 2022). This strategy is, however, less attractive if one wishes to keep the pipeline more generic for not very evident rare classes. Regarding better data balancing, (Sander, 2020) argues in favor of upsampling the smallest class to equal its points cardinality to be the same as in the largest class against decimating all point clouds in a batch to the number of points of the smallest point class as it would help to preserve as much information as possible. Post-processing routines, such as relabeling non-confident pixels, may bring some improvements as well, as we can see in the image-based task (Li et al., 2017, Jia et al., 2021). These relabelings, targeting the softmax outputs of the DL-based approaches, may particularly help the small classes. For example, (Qiu et al., 2022) leveraged the elevation above ground feature to improve segmentation for shapes of cars. Additionally, to consider spacial neighborhoods, Markov or Conditional Random Fields represent well-known tools frequently applied to post-processing classification results (Li et al., 2017). They suppress noise by encouraging same-class-neighborhoods, but decrease the class-weighted score by favoring large classes. Several improvements have been proposed: sometimes, higher-order potentials encourage a whole group of pixels to belong to the same class (Niemeyer et al., 2016). An alpha shape (Edelsbrunner and Mücke, 1994) for building pixels is an example of such a constraint (Montoya-Zegarra et al., 2015). Alternatively, (Pham et al., 2019) applied Multi-Value Conditional Random Fields: the imbalanced classes are aggregated to instances via embeddings. These embeddings are incorporated into the energy minimization workflow designed for indoor point clouds. However, configurations concerning labels of neighboring nodes of different classes and/or higher order priors cannot be easily considered without causing massive numerical issues during binarization (Bulatov et al., 2016). It must be noted that the MRF and CRF routines do not necessarily have to constitute the post-processing step of the classification procedure but can be devised as additional layers within the DL network while pack-propagating the error of MRF inference (Schwing and Urtasun, 2015). However, such simplifications of MRFs are often unable to capture certain fine-grained details or correlations present in the original system (Liu et al., 2017). Additional challenges may be given by irregular graphs induced by 3D points neighborhoods and by a high number of classes, as (Liu et al., 2017) have concluded.

1.2 Contribution

In this study, we exploit the potential of MRFs to include context information for point cloud classification to reach more reliable results. We hypothesize that assignments of DL-based pipelines, such as the state-of-the-art RandLA-Net approach, can be improved by introducing prior knowledge. The priors are formulated in a module allowing to increase a-priori probabilities of smaller classes, on the one hand, and in an inter-class reliability matrix and used to penalize improbable neighborhoods, on the other hand. Both priors can be easily integrated into an MRF-based workflow, allowing obtaining the a-posteriori probabilities using the QPBO strategy.

Our contributions include:

- Application of RandLA-Net to the classification of urban point cloud data.
- Improvement of the classification results through the approach based on MRFs and incorporating a small-class-preserving module and an inter-class reliability matrix as components of unary and pairwise potentials, respectively, to rectify unreliable results. These alterations are based on common-place knowledge.
- Application of the QPBO optimization to minimize the MRF-energy function with super-modular potentials.

2. Methodology

2.1 Deep-learning-based Classification

The network used for classification of the points is RandLA-Net (Hu et al., 2020). This network has been specifically designed to handle large point clouds efficiently. One of the biggest bottlenecks in point cloud classification is given by the downsampling steps in the encoder, which are necessary to get hierarchical features. PointNet++ (Qi et al., 2017b), the first network in point cloud segmentation to do so, uses farthest point sampling (or FPS) as the downsampling technique. While FPS tends to select quite meaningful points, improving the network performance, it is also very computationally expensive and inefficient to calculate in very large point clouds. RandLA-Net however employs random sampling, the fastest possible sampling, making it ideal for large point clouds, as is common in remote sensing.

Random sampling may discard meaningful points. To counteract this, the authors built sophisticated so-called local feature aggregation modules that encode the relative positions and features of the neighboring points of each point using computationally efficient MLPs and attentive pooling. The overall structure of RandLA-Net exhibits an encoder and decoder with skip connections, while the local feature aggregation modules in each layer are reminiscent of residual blocks from ResNet. To improve the classification of smaller classes, the data sampler prioritizes regions with rare classes during training.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-229-2024 | © Author(s) 2024. CC BY 4.0 License.

The loss function is also weighted by class occurrence. Other than the coordinates of the points, RandLA-Net accepts arbitrary additional inputs for each point. To further improve the classification results, we found it helpful to add also features that are typical in remote sensing. The features we added are: RGB, relative elevation, number of points in a fixed neighborhood, variation of these points in the z-axis, planarity, and normal vector. The relative elevation above ground is calculated using the method in (Bulatov et al., 2014).

During inference, RandLA-Net processes 40,000 points at once with a batch size of 16 on an NVIDIA V100 GPU with a RAM of 16 GB, resulting in a throughput of about 200,000 points per second. Higher throughput may be possible, we did not optimize the parameters or the implementation for speed. The output of the network is saved as softmax probabilities of each class, which are then given to the MRF-based workflow for post-processing.

2.2 Application of Markov Random Fields

The MRF is an undirected probabilistic graphical model denoted as $G = \langle V, E \rangle$. Here, V signifies a set of nodes, each corresponding to a random variable. The set of edges $E = \{(i,j) : i \neq j; i,j \in V\}$ indicates the dependency of the node i and j. In our case, each point within the point cloud dataset is treated as a node, while edges establish connections between nodes and their neighbors. The interactions among the nodes are characterized by an energy function, comprised of unary and pairwise potentials. Minimizing this energy function yields the optimal node classification, satisfying the prescribed constraints.

2.2.1 Unary Potentials The unary potentials depict the costs associated with assigning nodes to various classes without considering the state of other nodes. We convert the output probabilities of RandLA-Net into the unary potentials for each node using the following equation:

$$E_u\left(s(i)\right) = \min\left(-\log_2 P\left(s(i)\right), 2048\right), \tag{1}$$

with $E_u(s(i))$ and $P_u(s(i))$ denoting the unary potentials and the output probability of node i in class s, respectively. Basically, by using the standard negative logarithm trick, we establish a mechanism ensuring that a low probability results in a substantial cost for the MRF. The normalization factor 2048 is the maximum value among all unary potentials. Normalizing $E_u(s(i))$ to a range of 0 to 2048 allows processing memory-expensive graphs as integer values at a sufficiently high resolution and without numerical issues (overflow).

Because the subsequent application of MRFs may oversmooth smaller classes, we decided to modify the unary potentials to encourage assignments towards these less frequent classes. We formalize and extend the procedure of (Li et al., 2017) (who just performed relabeling of classes road and car according to the distribution of $P(s)$) in the following way. Let s_1 and s_2 be the two classes and $s_1 \notin \mathcal{S}_\circ$, where $\mathcal{S}_\circ$ is the set of small classes, Let i denote a pixel for which $P(s(i))$ takes on its highest (p_1) and second-highest (p_2) value for s_1 and s_2, respectively. Finally, let ϕ be a feature and $\theta_p, \theta_s \in [0,1], \theta_u = \pm 1$, θ_ϕ four thresholds. If $\theta_p p_1 < p_2$ and $\theta_u \phi(i) < \theta_\phi$, then we increase the second-highest probability before applying (1):

$$P(s_2(i)) = p_2(1 + \theta_s) \tag{2}$$

Doing so, points from the bigger class s_1 change the labels towards s_2 if 1) s_2 is the second-most-probable, with probability close enough to the winner and 2) the key feature ϕ is larger or smaller (depending on θ_u) than θ_ϕ.

Choice of s_1, s_2, ϕ, and $\theta_{\{p,u,\phi,s\}}$ is data-specific, however, very intuitive. For outdoor point clouds, one can already assume which pairs of classes can be confused with each other (i.e. roof and chimney, road and car, etc.) or can extract such pairs s_1 and s_2 from the referenced data, for example, as large off-diagonal elements of the confusion matrix. Our intuition can also help to retrieve the characteristic feature and the separation thresholds $\theta_{\{u,\phi\}}$ (chimney tends to have a higher linearity than roof, and a car usually has a higher elevation than road, etc.) while analysis of Kullback-Leibler divergences for points with reference offers a way towards automatized determination of these parameters. Thresholds $\theta_{\{p,s\}}$ are the least critical. They can either be set to the standard values (0.5 and 0.15, respectively) or determined using the reference data: it must not *degrade too much* the a-priori likelihoods. The advantage of the concept lies in the a-posteriori estimation. If the probability of only a few car points is increased, then, during MRF inference, also their neighbors can change their labels to be correctly assigned to the car class while an isolated point spuriously assigned to car class, will not cause much harm during the MRF inference and will most likely be oversmoothed.

2.2.2 Pairwise Potentials The pairwise potentials take into account the cost caused by the adjacency relationship, which depend on the classes of the endpoint nodes of the edge. In this study, the k-nearest neighbor (kNN) search is employed to locate the nearest N neighbors for each point in 3D space, establishing edges between them. The distances d_{ij} between two neighboring points i, j are converted into weights

$$W_{ij} \;=\; \lambda \cdot \exp\left(-\frac{d_{ij}}{\omega}\right) \tag{3}$$

to regulate their mutual influence. Hereby, ω adjusts the sensitivity to d, and λ controls the gain of the weights. Hence, the pairwise potentials are defined as:

$$E_p(s(i), s(j)) = W_{ij} \cdot \begin{cases} 0, & \text{if } s(i) = s(j) \\ 1, & \text{if } s(i) \neq s(j), \text{reliable} \\ R, & \text{if } s(i) \neq s(j), \text{unreliable} \end{cases} \tag{4}$$

with W_{ij} from (3). We do not penalize edges formed by nodes of the same class to encourage smoothing. In other cases, depending on the reliability of the edge, we impose two types of penalty, 1 and $R\,(\gg 1)$ to suppress the appearance of unreliable adjacency. The inter-class reliability matrix is as in Figure 1. It is grounded in the perception of objective reality. For instance, a chimney ($s = 10$) is only likely to be connected to itself or the roof ($s = 4$). Consequently, the reliability of the edges formed between chimneys and other classes is extremely low.

2.2.3 Energy Minimization Based on the previous description and Equations (1) and (4), the energy function for the MRF is as follows:

$$E(s) = \sum_i \left[E_u\left(s(i)\right) + \sum_{j \in N(i)} E_p\left(s(i), s(j)\right) \right] \tag{5}$$

Achieving energy minimization on MRFs is recognized as an

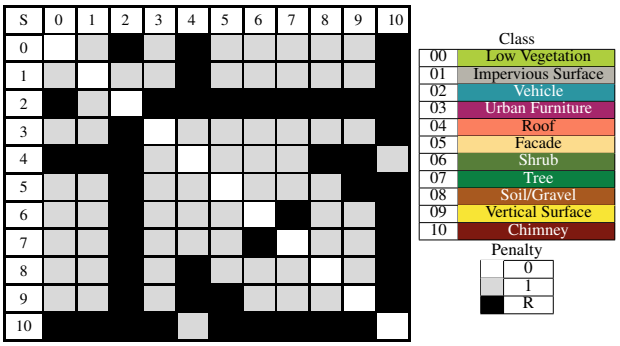


Figure 1. Matrix of the inter-class reliability.

NP-hard problem, prompting the innovation of diverse graph cut algorithms to provide approximate solutions (Boykov et al., 2001, Rother et al., 2007). In our MRF, R breaks the submodularity. For example, in clique consisting of low vegetation ($s = 0$), roof ($s = 4$) and facade ($s = 5$), the reliability of the edges between the low vegetation and the facade, as well as between the roof and the facade, is 1. However, the reliability between the low vegetation and the roof is R. This non-submodular term ($1 + 1 \ll R$) leads to the inability to apply the classical Alpha-Expansion algorithm (Boykov et al., 2001). Therefore, we chose the QPBO algorithm (Rother et al., 2007), whose unique ability to handle non-submodular energy terms fits seamlessly with the characteristics of our designed MRF.

Algorithm 1 Energy minimization approach with QPBO algorithm as black box.

INPUT: $E_u, E_p, I_{\max}$ (max. number of iterations), δ (tolerance)
OUTPUT: Improved label map s

Initialize s as local result on E_u, $\delta_s = +\infty$
while iter < $I_{\max}$ & $\delta_s > \delta$ do
  iter++
  Set order of expansions S^*
  for $\alpha \in S^*$ do
    Compute binary graph G_b:
    $E_u(0), E_u(1), E_p(0,0), ...E_p(1,1)$
    Perform QPBO
    Update s to be α where G_b is 1
  end for
  Compute deviation δ_s
end while
Return s

Algorithm 1 demonstrates the approach of energy minimization. The reassignment problem of eleven classes (in our dataset, see next section) is transformed into eleven binary classification problems, similarly to α-expansion. Through iterations, we approximately obtain class assignments that minimize the energy of MRF. In particular, during each outer iteration, we first initialize an update sequence S^* that includes all classes and is randomly ordered. Randomization allows the algorithm to process classes in a different order each time in the inner iteration, thereby increasing the diversity in the exploration of the solution space. This contributes to a more comprehensive exploration of possible solutions and increases the robustness of the algorithm.

Within an inner iteration, we perform a graph cut operation on current class s assignments with a fixed potentially updated class $\alpha \in S^*$ (inner iteration). In order to achieve this, for each point i, we retrieve the potentials $E_u(0) = E_u(0(i))$ representing the current unary cost, and $E_u(1)$, representing the unary cost of the class α. Additionally, we use Equation (4) with α to compute the pairwise costs $E_p(0,0)$, $E_p(0,1)$, $E_p(1,0)$, and $E_p(1,1)$ to adapt to different cases. Here, 0 implies the current class of the point, while 1 signifies the conversion of the class into α. Thus, $E_p(0,0)$ denotes the smoothness penalties between the current labels in s and $E_p(1,1)$ means that both labels are overwritten with α (and so it is 0 because of Eq.(4)). Potentials $E_p(0,1)$ are $E_p(1,0)$ are defined analogously.

The QPBO algorithm processes the data in polynomial time. The algorithm design does not differ much from Alpha expansion, which in the case of sparse graphs, is intrinsically linear in number of edges and labels. In an inner iteration of the Algorithm 1, a label update typically takes about 4 seconds, under the main computer specifications: CPU (16 cores, 4.20 GHz), RAM (128 GB) and GPU (NVIDIA RTX4090, 24 GB).

3. Results

3.1 Dataset

The dataset used in this paper is the Hessigheim dataset from the IFP institute of the University of Stuttgart, Germany (Kölle et al., 2021). This H3D dataset was captured using LiDAR with a resolution of 800 points/m^2 that are enriched by colors from RGB cameras. The points are labelled into eleven classes. In total, four measurements at different times are available, and we opted to use the most popular March 2018 capture. For each capture, predefined training, validation, and testing sets exist. The neural network was trained on the training and validation set. For the test set, the labels are not released to the public, instead, researchers are encouraged to submit their classification results via their website. Since the data providers of the dataset discourage multiple submissions with only marginal differences, we opted to do all evaluations on the validation set, which contains about 14.5 million points.

To assess the accuracy, we applied metrics most often used in remote sensing, namely the overall accuracy (abbreviated as OA) and the average F1 score, needed to track the algorithm's performance on smaller classes. Furthermore, we investigated the performance of the optimization algorithm by comparing energies at the beginning and at the end of the energy minimization procedure with that of the ground truth, bearing in mind that a few model violations, such as grass-covered roofs can be present in the reference data.

3.2 Findings

3.2.1 Quantitative Evaluation We examine first the parameters relevant for the pairwise potentials, which we established in Section 2.2.2: λ, ω, R, and N. Hence, we conducted several sets of experiments with the unmodified unary potentials to explore the effect of different parameter options on the energy optimization.

Figure 2 shows the impact of different parameters on the overall accuracy (OA). The result of the deep-learning-based procedure, RandLA-Net is already relatively accurate with more than 85.17% of overall accuracy, given that the number of classes is eleven. Still, the MRF-based approach effectively further optimizes this prediction results, improving the overall accuracy by approximately 1 to 2 $p.p.$ (percentage points).

Overall, different N in combination with λ have different effects on OA, when $N = 10$, a large λ constantly improves OA, whereas when $N = 40$, at a low λ there is already a satisfactory performance. Furthermore, with the increase of λ there is a tendency of decreasing OA, and the imposition of penalties on neighboring elements diminishes the optimization. This phenomenon could be attributed to an excessive correction induced by the consideration of an extensive number of neighbors. Moreover, in the context of identical parameters elsewhere, the performance when ω equals 1 surpasses that when it equals 0.4. The parameter ω, which in some related work is being attributed to differentiate between MRFs and CRFs, has not shown an exuberant influence on the results and will be therefore set to 1 in further experiments.

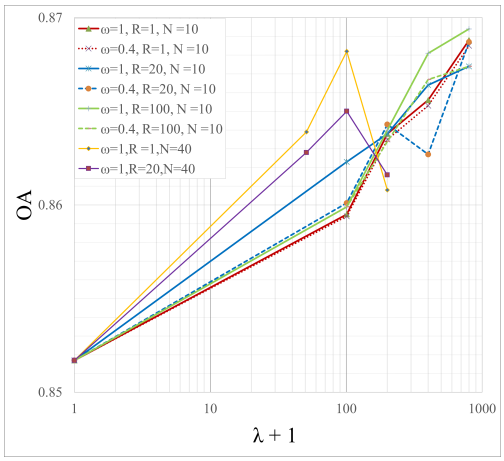
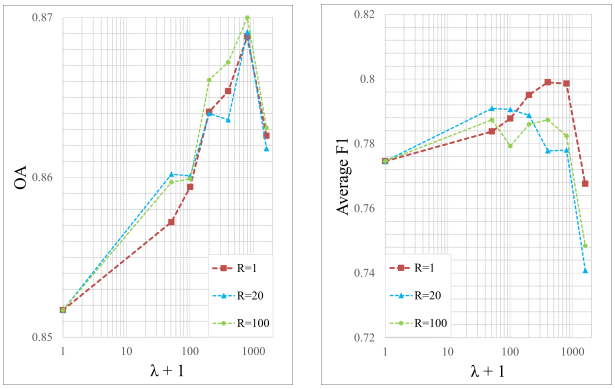


Figure 2. Impact of different parameters on the overall accuracy.



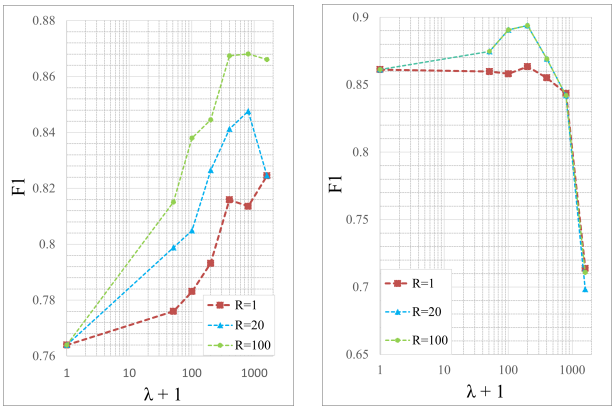
(a) Overall Accuracy (OA) (b) Average F1 Score

Figure 3. Impact of R and λ on the metrics ($\omega = 1$, $N = 10$).

To more intuitively observe the effect of R and λ on the performance, we repeated our experiments using an identical optimization sequence S^* for all configurations. This helps to avoid fluctuations that may be introduced by different optimization sequences. As mentioned before, the improvement is positively correlated with λ. It comes from increasing the distance weight, making each node more inclined to refer to neighboring nodes that are closer in reassignment. Nevertheless, it is worth noting that when the λ is too large, excessive smoothing leads to a reduction in OA.

A larger R further improves the accuracy, verifying the effectiveness of our pairwise potentials. The reliability matrix makes the result closer to the real situation by limiting the generation of unreliable edges, which are ignored by the deep learning model. Notably, when the R is not sufficiently large, increased λ renders its impact less pronounced, even below the effect of Alpha-Expansion ($R = 1$).

After our approach, there is a certain increase in the macro F1 score as well (see Figure 3(b)). In some cases, a large λ would weaken the effect and this turning point surfaces earlier with higher R.



(a) Vehicle (S2) (b) Chimney (S10)

Figure 4. F1 Score of small classes.

In focusing the observation on some of the small classes where strict neighbors' penalties have been imposed, such as vehicle ($s = 2$) only exists on the road (Impervious surface, $s = 1$), and chimney ($s = 10$) can only exist on roof ($s = 4$), the enhancement effect of R on individual F1 score is significant (see Figure 4). This further demonstrates that our approach effectively improves the prediction results for several small classes. It is worth noting that larger values of R enhance the F1 score of these classes quite well in comparison to the standard Alpha-Expansion. However, with an excessively large λ, a discernible weakening is observed.

Through Table 1, it can be observed that, with $N = 10$, the effective constraint of R prevents the occurrence of unreliable edges, and as the λ increases, this penalty becomes more stringent. When $R = 1$, the smoothing effect also suppresses unreliable edges under a larger λ. Moreover, when applying a complete penalty to all 40 neighbors, unreliable edges are almost nonexistent.

N	ω	R	λ	$S10, \neg S4$	$S2, \neg S1$	$S6, S7$	$S5, S9$
10	1	100	100	160	1484	9606	3620
			800	66	696	1238	1070
		20	100	319	3554	12711	4419
			800	141	1071	1386	316
		1	100	4784	22731	89082	22156
			800	644	7369	11008	4347
40	1	20	50	0	14	49	0
			200	0	2	36	0
		1	50	538	5480	20667	6013
			200	191	2371	2561	3779
Ground Truth				630	2525	35750	132
RandLA-Net				12020	137445	497614	116454

Table 1. Statistics for unreliable neighbors using 40 neighboring instances as observations (excerpt).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-229-2024 | © Author(s) 2024. CC BY 4.0 License.

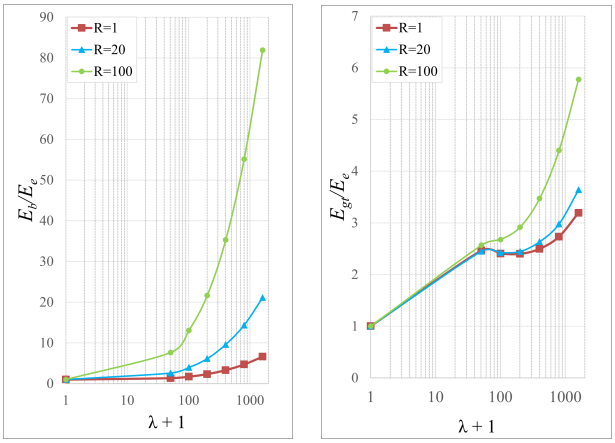


Figure 5. Energy ratio of MRF (left: E_b/E_e, right: E_{gt}/E_e) under configurations $\omega = 1$ and $N = 10$.

In order to track the mathematical performance of the MRF-based optimization, we plotted in Figure 5 the ratios E_b/E_e and E_{gt}/E_e for different values of λ and R. Hereby E_b, E_e, E_{gt} denote the energy at the beginning, at the end and that of the ground truth configuration, respectively; For all configurations, ω was set to 1 and N to 10.

The plots in Figure 5(a) show that, despite the higher energy associated with large λ and R, our approach consistently achieves a significant (1-2 orders of magnitude) reduction in total energy through optimized class assignments during the energy minimization. The plots in Figure 5(b), denoting E_{gt}/E_e and measuring the impact of violations of model assumptions, indicate that the energy of the configuration optimized by our approach is lower than that of the ground truth. Fortunately, with growing R, the curves remain quite close to each other. This phenomenon can be explained in Table 1, since there are, albeit very few, forbidden choices of neighbors within the ground truth. Nevertheless, in the results of the RandLA-Net, it seems that there are significantly more unreliable neighbors. Thus, our design of pairwise potentials is meaningful and capable of substantially optimizing the results. In addition, a reasonable choices of R and λ are necessary to avoid deviating from our assumptions and overfitting from excessively harsh penalties.

Since the manipulation of unary potentials requires determination of quite many parameters $(\theta_p, \theta_s, \theta_u, \theta_\phi)$, we have considered one such set of parameters and run the QPBO-based optimization with the previously selected parameters ($R = 1, N = 10, \lambda = 400, \omega = 1$). The accuracies increased, achieving 86.54%. It is notable that the a-priori result (that is, point-wise calculation of altered by (2) unaries merely yielded 85.24%. In our modification, we fine-tuned the unary potential of pixels on soil ($s = 8$). Consequently, when comparing the results optimized through MRF under the same configuration with unmodified unary potentials, there was a slight improvement in the F1 score for soil ($s = 8$), around 0.1 *p.p.*. This confirms our conjecture that increasing probabilities of small classes is especially helpful while attracting neighbors in clusters.

3.2.2 Qualitative Evaluation Figure 6 illustrates the visualization of point cloud labels, with the output of MRF obtained from one of the configurations that exhibit superior OA and F1 score. The colors of the labels are consistent with those in Figure 1.

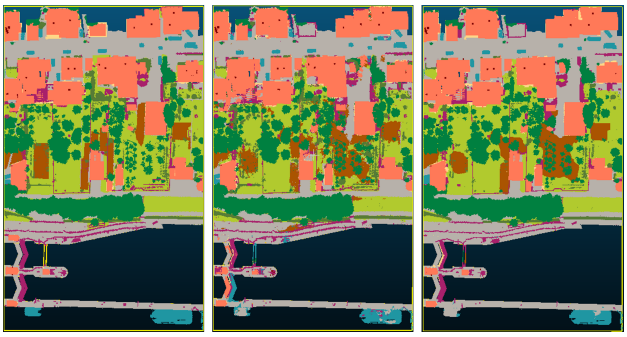


Figure 6. Overview of ground truth (left) and outputs (middle: RandLA-Net, right: MRF) with classes colored according to Figure 1.

In general, compared to the output of RandLA-Net, the results of MRF are smoother and closely resemble the ground truth. It tends to achieve excellent results for objects with clear structures, but faces challenges with complex contours.

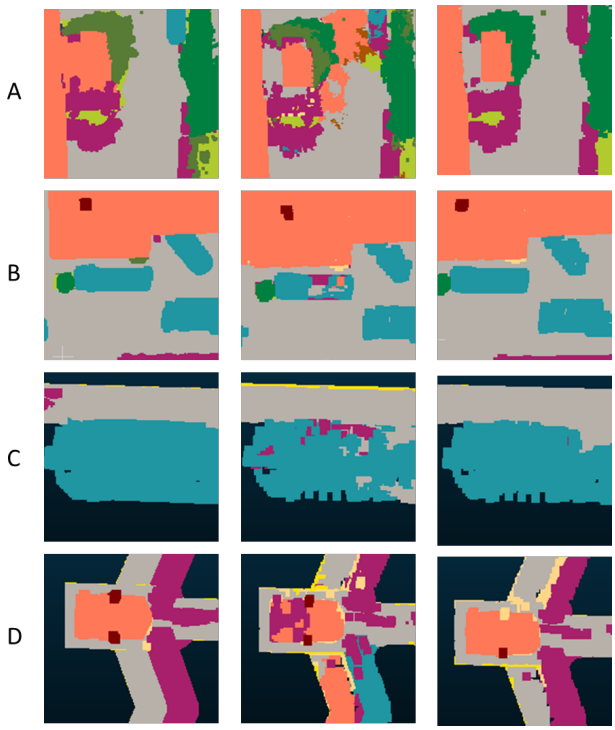


Figure 7. Nadir-views of slices of positive results: A for false roof on the road surface, B for noise on car, C for noise on ship, and D for noise on roof. Classes are colored according to Figure 1 (left: Ground truth, middle: RandLA-Net, right: MRF).

Figure 8. Slice of results related to neighbours, frontal view. Classes are colored according to Figure 1 (left: Ground truth, middle: RandLA-Net, right: MRF).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W5-2024-229-2024 | © Author(s) 2024. CC BY 4.0 License.

In detail, MRF can well fix some small prediction errors of RandLA-Net on objects (see Figure 7). For example, in case A the road surface (Impervious surface, $s = 1$) is predicted to have a false roof ($s = 4$). In case B and C, there is noise on a vehicle ($s = 2$). And in case D the roof ($s = 4$) is polluted by urban furniture ($s = 3$). In some regions, we observe a facilitating effect of neighbors on MRF optimization. Since we utilize KNN search in 3D space for each node, the neighbors at nearby elevations can sometimes bring contributions. In the example of Figure 8, although the output of RandLA-Net completely misclassifies this channel (Urban funiture, $s = 3$) as a vehicle ($s = 2$), our approach successfully restores it. This may be due to the combined effect of R and other neighbors at nearby elevations.

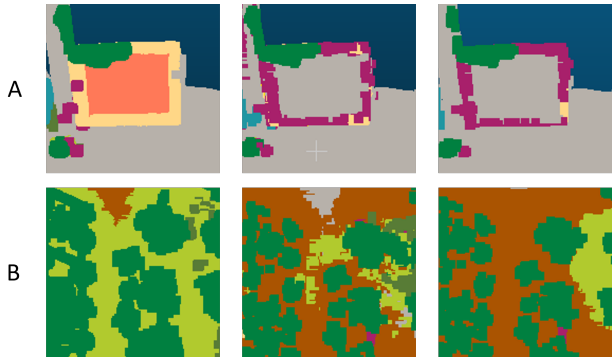


Figure 9. Slices of negative results: A for completely incorrect roof, B for misclassification of grass as soil. Classes are colored according to Figure 1 (left: Ground truth, middle: RandLA-Net, right: MRF).

Conversely, when dominant incorrect predictions prevail, MRF remains indifferent to rectifying the results and may even amplify such errors (see Figure 9). As in case A, the MRF is unable to restore the roof ($s = 4$), because in Hessigheim, there are only few such roofs. In case B, the misclassification of grass (Low vegetation, $s = 0$) as soil ($s = 8$) are further amplified.

4. Conclusion

We presented a deep-learning-based workflow for outdoor point classification. In our work, RandLA-Net has once again proven to be a computationally efficient method for analyzing large-scale point clouds. Attention mechanisms and relative class weighting make the correct assignment of even the smallest classes possible.

In order to deal with the unconventional neighborhoods (grass-chimney, etc.), which occasionally occur despite the absence or scarcity of such neighborhoods in the training data, we design a post-processing step based on Markov Random Fields with semantic priors given by unary potentials encouraging preservation of certain classes and the so-called inter-class reliability matrix applied to pairwise potentials. This matrix indicates how plausible a certain neighborhood of 3D points is, which is justified by common-place knowledge. Even though neighborhood relations may depend on application (military applications may mean vehicles parked on grass, for example), the context is very intuitive and can be easily adjusted according to the situation and passed to a non-local optimization approach. In our future work, we will explore the potential for adaptive adjustment of interclass reliability matrices based on the characteristics of the datasets to enhance the generality of our approach.

The resulting MRF problem is no longer sub-modular for $R > 1$. As a consequence, advanced minimization methods have to be applied, such as the QPBO method. Since this method only works on binary problems, we implemented a workaround allowing binarizing our graph given a fixed label. While already with the standard Alpha-Expansion approach ($R = 1$), we could achieve a considerable improvement in OA against the RandLA-Net result, the application of context-based neighborhood priors allows further smaller progress. However, it seems that large classes grow at the cost of the small classes because the MRFs oversmooth towards dominant classes. As a consequence, we observe a decay in the F1 score while the OA is still growing. Then, in the future work, imposing further losses that allow the preservation of smaller classes, including those discouraging certain neighborhoods, will be considered.

The validation set of the Hessigheim dataset is relatively small, and it has a different characteristic than the actual test set, for example, regarding point density. Evaluation of the test set takes place by the data provider only. Due to this, it was not yet possible to assess so many configurations as in the validation set. While the first tests of the proposed methodology have been successful, more configurations are to be submitted.

Acknowledgements

The authors thank the China Scholarship Council (CSC) for supporting this research, Grant/Award Number: 202308080109.

References

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11), 1222–1239.

Bulatov, D., Häufel, G., Meidow, J., Pohl, M., Solbrig, P., Wernerus, P., 2014. Context-based automatic reconstruction and texturing of 3D urban terrain for quick-response tasks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, 157–170.

Bulatov, D., Kottler, B., Rottensteiner, F., 2016. Energy minimization of discrete functions with higher-order potentials for depth map generation. *23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2344–2349.

Edelsbrunner, H., Mücke, E. P., 1994. Three-dimensional alpha shapes. *ACM Transactions On Graphics (TOG)*, 13(1), 43–72.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11108–11117.

Jia, M., Reiter, A., Lim, S.-N., Artzi, Y., Cardie, C., 2021. When in doubt: Improving classification performance with alternating normalization. *arXiv preprint arXiv:2109.13449*.

Kölle, M., Laupheimer, D., Schmohl, S., Haala, N., Rottensteiner, F., Wegner, J. D., Ledoux, H., 2021. The Hessigheim 3D (H3D) benchmark on semantic segmentation of high-resolution 3D point clouds and textured meshes from UAV LiDAR and Multi-View-Stereo. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 1, 11.

Lehner, A., Gasperini, S., Marcos-Ramiro, A., Schmidt, M., Mahani, M.-A. N., Navab, N., Busam, B., Tombari, F., 2022. 3D-VField: Adversarial augmentation of point clouds for domain generalization in 3D object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17295–17304.

Li, J., Ding, W., Li, H., Liu, C., 2017. Semantic segmentation for high-resolution aerial imagery using multi-skip network and Markov Random Fields. *2017 IEEE International Conference on Unmanned Systems (ICUS)*, IEEE, 12–17.

Lin, H.-I., Nguyen, M. C., 2020. Boosting minority class prediction on imbalanced point cloud data. *Applied Sciences*, 10(3), 973.

Liu, Z., Li, X., Luo, P., Loy, C. C., Tang, X., 2017. Deep learning Markov Random Field for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8), 1814–1828.

Montoya-Zegarra, J. A., Wegner, J. D., Ladickỳ, L., Schindler, K., 2015. Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, 127–133.

Niemeyer, J., Rottensteiner, F., Sörgel, U., Heipke, C., 2016. Hierarchical higher order CRF for the classification of airborne lidar point clouds in urban areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 655–662.

Pang, Y., Wang, W., Tay, F. E., Liu, W., Tian, Y., Yuan, L., 2022. Masked autoencoders for point cloud self-supervised learning. *Proceedings of the European Conference on Computer Vision*, Springer, 604–621.

Pham, Q.-H., Nguyen, T., Hua, B.-S., Roig, G., Yeung, S.-K., 2019. Jsis3d: Joint semantic-instance segmentation of 3d point clouds with multi-task pointwise networks and multi-value conditional random fields. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8827–8836.

Qi, C. R., Su, H., Mo, K., Guibas, L. J., 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.

Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30.

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., Ghanem, B., 2022. PointNeXt: Revisiting pointNet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems*, 35, 23192–23204.

Qiu, K., Bulatov, D., Lucks, L., 2022. Improving car detection from aerial footage with elevation information and markov random fields. *SIGMAP*, 112–119.

Qiu, S., Anwar, S., Barnes, N., 2021. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1757–1767.

Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M., 2007. Optimizing binary MRFs via extended roof duality. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 1–8.

Sander, R., 2020. Sparse data fusion and class imbalance correction techniques for efficient multi-class point cloud semantic segmentation. *Preprint*, 10.

Schwing, A. G., Urtasun, R., 2015. Fully connected deep structured networks. *arXiv preprint arXiv:1503.02351*.

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., Guibas, L. J., 2019. KPConv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6411–6420.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286-304.

Ye, X., Li, J., Huang, H., Du, L., Zhang, X., 2018. 3D recurrent neural networks with context fusion for point cloud semantic segmentation. *Proceedings of the European Conference on Computer Vision*, 403–417.

Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., 2021. Point transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16259–16268.