A. Implementation details

All results presented in this paper were computed on a GPU Nvidia RTX 3090.

A.1. State-of-the-art architectures

To compute the inference and for re-training, the following open-source codes were used (all found on github), alongside these associated parameters:

- KPConv: HuguesTHOMAS/KPConv-PyTorch. Model: KPFCNN; learning rate 1e-2; number of iterations: 500 000; maximum number of points per batch: 13400. Same parameters for nuScenes and SemanticKITTI. The choice of accumulating the 10 past frames for the accumulated KPConv comes from the recommendation inside the aforementioned repository.
- Cylinder3D: xinge008/Cylinder3D. Model: cylinder asym; learning rate: 1e-3; number of epochs: 40; batch size: 2; grid size: 480x360x32. Same parameters for nuScenes and SemanticKITTI.
- SRUNet and SPVCNN: mit-han-lab/spvnas. Model: MinkUnet and SPVCNN; learning rate: 2.4e-1; number of epochs: 15 for SemanticKITTI, 20 for nuScenes; batch size: 2 for SemanticKITTI, 8 for nuScenes; voxel size: 0.05m.
- CENet: huixiancheng/CENet . Model: senet-512; learning rate: 1e-2; number of epochs: 100; batch size: 6; Image resolution: 512x64. Only trained on SemanticKITTI.
- Helix4D: romainloiseau/Helix4D. Model: Helix4D; learning rate: 2e-3; number of epochs: 50; batch size: 2. Only trained on SemanticKITTI.

A.2. 3DLabelProp

As mentioned in the main paper, the KPConv model used inside 3DLabelProp is the standard layout, namely the KPFCNN. It consists of 4 downsampling blocks and 4 upsampling blocks.

In every case, the model architecture does not change. For the training, we use a Lovasz loss and weighted cross-entropy loss. We use a SGD optimizer with a weight decay of 1e-4 and a momentum of 0.98. The learning rate scheduler is a cosine annealing scheduler. It is trained with cluster extracted by the 3DLabelProp pipeline when assuming that past inferences are the ground truth. Training hyperparameters are:

- SemanticKITTI - learning rate: 0.005, batch size: 12, number of iterations: 400 000.
- nuScenes - learning rate: 0.001, batch size: 16, number of iterations: 350 000.

Due to the annotation process of nuScenes, only 2 frame per seconds are annotated, contrary to other dataset where 10 frames per seconds are annotated. As we are only using the annotated frames for the training of the nuScenes model, we reproduce the same accumulation pattern at inference time.

In practice, we accumulate only one out of 5 frames when processing other datasets to emulate a 2Hz frame rate. It means that to predict semantic of points in frame n, we use accumulation of frames n, $n-5$, $n-10$; $n-15$; $n-20$.

A.3. Data augmentation

All data augmentations applied are the same for the various methods. They are also the one used for 3DLabelProp, but it is important to note that in the 3DLabelProp case, these augmentations are applied at the cluster level and not at the scan level.

Here is the list:

- Centering.
- Random rotation around z.
- Random scaling.
- Gaussian noise.
- Flipping around x and around y.

B. Label mappings

The choice of labels for the different mIoU computations can have a significant impact on the results. We share in Table 9 to Table 13 all the details of the mappings between datasets so that all the results can be easily reproduced and that in the future new generalization methods can be compared with our results.

Some mapping can look unnecessarily coarse. For instance, SemanticKITTI define trucks as ”Trucks, vans with a body that is separate from the driver cabin, pickup trucks, as well as their attached trailers.”, whereas nuScenes define trucks as ”Vehicles primarily designed to haul cargo including pick-ups, lorrys, trucks and semi-tractors.”, which means that the sub element ”van” can belong either to truck or not depending on the dataset. In order to not be sensitive to annotation choices, we have to group car and truck together into a single coarse category, otherwise we could be measuring annotation discrepancies.

This need for coarseness is exacerbated for $\mathcal{L}_{NS \cap SP}$ which is extremely coarse, and as a result can be considered an easy label set.

SemanticKITTI	$\mathcal{L}_{SK \cap NS}$	nuScenes
motorcycle motorcyclist	motorcycle	motorcycle
bicycle bicyclist	bicycle	bicycle
person	person	pedestrian
road parking	d. ground	driveable surface
sidewalk	sidewalk	sidewalk
other ground	o. ground	other flat
building fence pole traffic sign	manmade	barrier traffic cone manmade
vegetation trunk	vegetation	vegetation
car truck other vehicle	vehicle	bus car construction vehicle trailer truck
terrain	terrain	terrain

Table 9: Details of the mapping from the original label sets of SemanticKITTI and nuScenes to $\mathcal{L}_{SK \cap NS}$.

SemanticKITTI	$\mathcal{L}_{SK \cap SP}$	SemanticPOSS
person	person	person
bicyclist motorcyclist	rider	rider
bicycle motorcycle	bike	bike
other vehicle car truck	car	car
road terrain parking sidewalk other ground	ground	ground
trunk	trunk	trunk
vegetation	vegetation	plants
traffic sign	traffic sign	traffic sign
pole	pole	pole
building	building	building garbage can cone/stone
fence	fence	fence

Table 10: Details of the mapping from the original label sets of SemanticKITTI and SemanticPOSS to $\mathcal{L}_{SK \cap SP}$.

nuScenes	$\mathcal{L}_{NS \cap SP}$	SemanticPOSS
pedestrian	person	person
bicycle motorcycle	bike	rider bike
car bus construction vehicle trailer truck	car	car
driveable surface other flat sidewalk terrain	ground	ground
vegetation	vegetation	vegetation plant
barrier manmade traffic cone	manmade	traffic sign pole garbage can building cone/stone fence

Table 11: Details of the mapping from the original label sets of nuScenes and SemanticPOSS to $\mathcal{L}_{NS \cap SP}$.

SemanticKITTI	$\mathcal{L}_{SK \cap PS}$	PandaSet
bicycle motorcycle bicyclist motorcyclist	2-wheeled	pedicab personal mobility device motorized scooter bicycle motorcycle
person	pedestrian	pedestrian pedestrian w/ objects
road parking	d. ground	road road marking driveway
sidewalk	sidewalk	sidewalk
other ground terrain	o. ground	ground
building fence pole traffic sign	manmade	pylons road barriers signs cones construction signs construction barriers rolling containers building other static object
vegetation trunk	vegetation	vegetation
car truck other vehicle	4-wheeled	pickup truck medium sized truck semi-truck towed object construction vehicle uncommon vehicle emergency vehicle bus car

Table 12: Details of the mapping from the original label sets of SemanticKITTI and PandaSet to $\mathcal{L}_{SK \cap PS}$.

nuScenes	$\mathcal{L}_{NS \cap PS}$	PandaSet
bicycle motorcycle	2-wheeled	pedicab personal mobility device motorized scooter bicycle motorcycle
pedestrian	pedestrian	pedestrian pedestrian w/ objects
driveable ground	d. ground	road road marking driveway
sidewlak	sidewalk	sidewalk
other flat terrain	o. ground	ground
barrier manmade traffic cone	manmade	pylons road barriers signs cones construction signs construction barriers rolling containers building other static object
vegetation	vegetation	vegetation
car bus construction vehicle trailer truck	4-wheeled	pickup truck medium sized truck semi-truck towed object construction vehicle uncommon vehicle emergency vehicle bus car

Table 13: Details of the mapping from the original label sets of nuScenes and PandaSet to $\mathcal{L}_{NS \cap PS}$.

C. Quantitative & Qualitative results

In Table 14 to Table 25, we present the details by class of all semantic segmentation results presented in the main article. They are complementary of Table 4 and Table 5

As mentioned in the main article, Table 19 we see the ability of KPConv to perform well on the Person class (+12% over the second best), which explains KPConv outranking every other methods. We could believe that 3DLabelProp has trouble reproducing such a high quality of dynamic objects due to the accumulation process, which leaves trail inside the scan. Nonetheless, we did try a variant of 3DLabelProp by discarding past dynamic object (to remove trails), and we notice a decrease in general performance (especially ”mobile” objects like parked cars).

We see the impact of the coarseness of $\mathcal{L}_{NS \cap SP}$ in Table 25 as all methods are well performing.

In the various result tables, we did not include the number of parameters of the various models, even if it is an important variable in real time semantic segmentation. This choice was made because, from one segmentation paradigm to another, the parameters don’t have the same impact.

At the end of this part in Figure 8, we illustrate the 3DlabelProp process.

	Car	Bicycle	Motorcycle	Truck	Other-vehicle	Person	Bicyclist	Motrocyclist	Road	Parking	sidewalk	Other-ground	building	Fence	Vegetation	Trunk	Terrain	Pole	Traffic-sign	mIoU
CENet	91.3	24.8	60.6	81.8	57.1	56.9	76.5	0.0	93.2	51.6	78.7	0.1	84.6	54.9	82.9	60.1	68.6	54.6	38.1	58.8
Helix4D	96.0	21.2	63.5	68.2	61.4	64.9	73.9	0.5	93.3	39.4	79.2	1.4	88.6	54.1	88.5	60.7	77.4	61.2	46.0	60.0
KPConv	94.7	36.8	58.6	45.1	47.7	62.7	76.8	1.0	90.1	32.3	75.5	3.8	88.4	59.8	87.6	67.2	74.3	61.7	44.4	58.3
SRUNet	96.2	7.9	51.6	65.5	51.5	64.7	64.5	0.0	93.2	48.3	80.1	0.1	91.0	62.3	88.4	67.7	75.4	62.8	43.3	58.6
SPVCNN	96.5	22.6	59.4	79.4	61.8	67.1	81.0	0.0	93.3	47.7	80.5	0.2	91.1	63.8	87.7	66.9	73.2	63.5	47.7	62.3
Cylinder3D	96.4	21.0	53.2	79.1	57.4	69.2	80.5	0.0	93.2	43.8	79.4	2.6	90.4	55.7	86.0	64.4	71.6	63.5	45.8	60.7
3DLabelProp (Ours)	96.4	38.7	80.3	63.5	62.6	76.5	92.6	0.0	85.9	45.6	66.5	4.2	87.4	49.0	86.4	65.6	70.0	48.5	36.4	60.8

Table 14: Source-to-source segmentation results for SemanticKITTI on $\mathcal{L}_{SK}$.

	Car	Bicycle	Motorcycle	Truck	Other-vehicle	Person	Bicyclist	Motrocyclist	Road	Parki-ng	sidewalk	Other-ground	building	Fence	Vegetation	Trunk	Terrain	Pole	Traffic-sign	mIoU
CENet	82.7	1.8	20.8	38.2	16.4	13.1	51.2	0.0	86.7	18.5	67.4	0.2	69.7	32.6	76.3	38.4	69.9	46.5	11.7	39.1
Helix4D	93.6	20.3	42.1	35.0	52.7	56.1	63.4	2.7	89.6	24.9	73.4	3.3	86.3	48.4	86.8	56.3	73.7	58.2	43.5	53.2
KPConv	92.4	30.6	52.6	27.4	30.1	55.5	70.5	2.0	84.4	21.6	67.2	8.4	83.7	56.9	87.3	60.1	74.1	58.6	37.8	52.7
SRUNet	94.0	7.3	45.1	57.0	43.9	54.8	53.6	0.0	90.7	37.1	75.5	1.4	88.9	54.9	87.6	62.8	74.1	60.1	37.2	54.0
SPVCNN	94.0	21.4	46.3	72.7	52.7	60.4	67.6	0.0	90.7	37.3	75.4	1.9	89.1	57.9	87.3	62.8	72.5	59.2	41.2	57.4
Cylinder3D	92.7	20.8	38.4	65.2	48.7	56.5	64.5	0.0	90.4	18.6	73.1	1.4	88.0	40.6	83.4	58.1	70.2	59.3	38.8	53.1
3DLabelProp (Ours)	96.2	35.4	86.4	63.8	66.7	71.7	87.8	0.0	77.8	38.7	59.3	0.7	83.5	42.6	82.6	53.6	56.3	43.3	31.0	56.7

Table 15: Validation set results on generalization from SemanticKITTI to SemanticKITTI32 on $\mathcal{L}_{SK}$.

	2-wheeled	Pedestrian	D. Ground	Seidewalk	O.Ground	Manmade	Vegetation	4-Wheeled	mIoU
CENet	0.1	0.7	5.2	7.2	7.0	51.0	28.2	7.4	13.3
Helix4D	07.6	13.9	41.7	23.7	6.4	54.6	32.2	42.5	27.7
KPConv	6.4	24.6	44.9	25.7	15.2	51.3	40.5	53.0	32.7
SRUNet	17.8	34.7	54.8	31.8	16.9	77.7	57.9	61.6	44.2
SPVCNN	11.0	23.8	51.2	33.9	17.5	77.1	53.4	53.4	40.2
Cylinder3D	2.7	11.9	15.5	18.4	10.4	52.2	25.1	10.9	18.4
3DLabelProp (Ours)	33.0	69.7	52.3	37.3	16.3	83.5	68.2	87.3	56.0

Table 16: Validation set results on generalization from SemanticKITTI to Panda64 on the label set $\mathcal{L}_{SK \cap PS}$.

	2-wheeled	Pedestrian	D. Ground	Seidewalk	O.Ground	Manmade	Vegetation	4-Wheeled	mIoU
CENet	0.0	0.0	4.9	1.4	2.0	11.1	21.4	3.1	4.9
Helix4D	3.5	4.0	24.0	5.4	4.7	28.4	20.4	23.5	14.2
KPConv	10.1	14.6	13.7	7.0	13.1	20.1	42.2	48.0	21.1
SRUNet	5.6	12.1	29.4	7.2	7.1	41.9	42.2	32.2	22.2
SPVCNN	2.0	6.7	28.7	7.6	6.1	49.9	37.5	16.9	19.4
Cylinder3D	0.0	1.4	1.9	4.8	8.4	20.9	13.2	1.4	6.5
3DLabelProp (Ours)	41.9	61.8	65.2	27.0	10.1	73.9	72.6	89.5	55.2

Table 17: Validation set results on generalization from SemanticKITTI to PandaFF on the label set $\mathcal{L}_{SK \cap PS}$.

	Person	Rider	Bike	Car	Ground	Trunk	Vegetation	Traffic-sign	Pole	Building	Fence	mIoU
CENet	2.9	40.2	0.0	31.9	73.9	33.3	65.6	22.2	36.5	77.6	37.0	27.9
Helix4D	27.8	18.4	4.0	59.7	64.4	27.7	58.0	24.3	29.2	66.5	16.4	36.0
KPConv	43.9	32.7	8.6	63.4	74.8	31.4	59.7	7.6	32.6	57.6	17.6	39.1
SRUNet	38.9	26.1	2.9	82.9	75.1	35.5	65.6	15.5	41.6	77.1	37.3	45.3
SPVCNN	44.6	18.1	4.6	83.2	75.0	36.0	65.8	18.5	41.6	75.8	36.0	45.4
Cylinder3D	44.6	13.6	2.8	81.5	73.6	33.2	61.6	13.1	36.0	71.3	30.1	41.9
3DLabelProp (Ours)	71.6	43.9	3.6	89.3	73.9	33.3	65.6	22.2	36.5	77.6	37.0	50.4

Table 18: Validation set results on generalization from SemanticKITTI to SemanticPOSS on the label set $\mathcal{L}_{SK \cap SP}$.

	Motorcycle	Bicycle	Person	Road	Sidewalk	O. Ground	Manmade	Vegetation	Car	Terrain	mIoU
CENet	0.0	0.0	0.0	0.8	0.2	1.3	24.6	16.8	6.0	0.2	5.0
Helix4D	7.5	2.8	23.2	75.7	31.7	3.8	57.2	58.5	53.7	28.5	34.3
KPConv	33.7	7.8	52.8	80.3	34.9	4.0	70.7	70.5	74.3	37.8	46.7
SRUNet	17.1	2.4	24.8	87.8	39.4	5.1	74.7	72.6	78.5	25.5	42.8
SPVCNN	35.9	3.0	27.3	88.3	39.6	6.6	75.0	72.4	78.2	24.7	45.1
Cylinder3D	3.3	2.2	1.0	77.3	32.0	7.6	67.3	60.6	56.2	19.8	32.7
3DLabelProp (Ours)	27.4	8.8	40.9	79.3	38.3	5.0	69.8	69.4	72.8	32.7	44.4

Table 19: Validation set results on generalization from SemanticKITTI to nuScenes on the label set $\mathcal{L}_{SK \cap NS}$.

	barrier	bicycle	bus	car	cnstrctn-vhcl	motorcycle	pedestrian	traffic-cone	trailer	truck	drvbl-grnd	other-flat	sidewalk	terrain	manmade	vegetation	mIoU
KPConv	66.2	17.8	72.4	87.8	26.3	67.8	69.8	51.5	25.6	73.5	93.8	53.6	66.3	71.2	83.0	82.8	63.1
SPVCNN	72.6	14.0	82.9	88.7	32.2	73.4	69.9	47.8	46.1	78.3	95.0	64.4	69.4	71.8	84.6	83.4	67.2
Cylinder3D	71.5	29.4	84.3	86.4	40.5	70.5	72.9	54.3	57.2	79.7	96.1	65.8	71.9	71.6	86.4	85.0	70.2
3DLabelProp (Ours)	71.8	46.9	84.1	84.5	39.7	83.1	77.0	50.4	47.9	79.7	93.1	62.5	68.5	72.9	87.7	86.3	71.0

Table 20: Source-to-source segmentation results for nuScenes on $\mathcal{L}_{NS}$.

	Motorcycle	Bicycle	Person	Road	Sidewalk	O. Ground	Manmade	Vegetation	Car	Terrain	mIoU
KPConv	18.9	3.9	49.8	67.2	28.8	0.4	73.4	81.8	79.0	45.8	44.9
SPVCNN	25.7	11.6	47.7	69.8	46.3	0.0	77.1	81.0	90.0	45.2	49.4
Cylinder3D	3.7	0.0	21.0	54.3	21.5	0.0	65.0	70.9	46.1	34.4	31.7
3DLabelProp (Ours)	36.8	37.2	66.1	77.9	61.1	0.4	81.6	84.3	93.2	66.7	60.5

Table 21: Validation set results on generalization from nuScenes to SemanticKITTI on the label set $\mathcal{L}_{NS \cap SK}$.

	Motorcycle	Bicycle	Person	Road	Sidewalk	O. Ground	Manmade	Vegetation	Car	Terrain	mIoU
KPConv	23.1	3.5	46.3	78.9	48.6	1.9	75.9	83.7	81.1	62.7	50.6
SPVCNN	28.6	8.3	45.0	81.4	50.9	0.2	81.6	84.5	91.5	60.0	53.2
Cylinder3D	14.0	5.6	37.5	75.8	48.0	0.0	74.5	77.9	78.8	52.0	46.4
3DLabelProp (Ours)	45.9	39.1	63.7	83.3	63.7	0.3	81.8	85.2	93.6	68.1	62.5

Table 22: Validation set results on generalization from nuScenes to SemanticKITTI32 on the label set $\mathcal{L}_{NS \cap SK}$.

	2-wheeled	Pedestrian	D. Ground	Seidewalk	O.Ground	Manmade	Vegetation	4-Wheeled	mIoU
KPConv	3.9	6.2	20.3	17.1	10.8	64.5	53.2	24.2	25.0
SPVCNN	19.2	43.1	38.0	30.6	15.3	71.9	62.1	69.4	43.6
Cylinder3D	0.3	0.8	13.3	6.9	7.8	58.9	30.5	8.3	15.8
3DLabelProp (Ours)	48.9	71.5	71.9	49.8	28.6	89.2	76.9	86.1	65.4

Table 23: Validation set results on generalization from nuScenes to Panda64 on the label set $\mathcal{L}_{NS \cap PS}$.

	2-wheeled	Pedestrian	D. Ground	Seidewalk	O.Ground	Manmade	Vegetation	4-Wheeled	mIoU
KPConv	5.0	6.8	7.0	7.5	4.6	38.2	48.7	17.8	16.9
SPVCNN	1.4	7.4	3.0	3.2	8.0	28.9	28.5	8.5	11.1
Cylinder3D	0.0	0.0	1.6	0.6	2.4	24.3	7.8	0.9	4.7
3DLabelProp (Ours)	51.6	68.2	90.0	40.8	28.5	85.6	81.2	87.6	66.7

Table 24: Validation set results on generalization from nuScenes to PandaFF on the label set $\mathcal{L}_{NS \cap PS}$.

	Person	Bike	Car	Ground	Vegetation	Manmade	mIoU
KPConv	56.0	17.2	72.4	73.8	70.0	74.5	60.7
SPVCNN	64.5	15.5	85.6	74.2	71.8	78.1	64.8
Cylinder3D	22.9	0.6	29.5	74.1	63.0	66.9	42.8
3DLabelProp (Ours)	72.1	0.9	79.4	73.3	70.8	80.9	64.3

Table 25: Validation set results on generalization from nuScenes to SemanticPOSS on the label set $\mathcal{L}_{NS \cap SP}$.

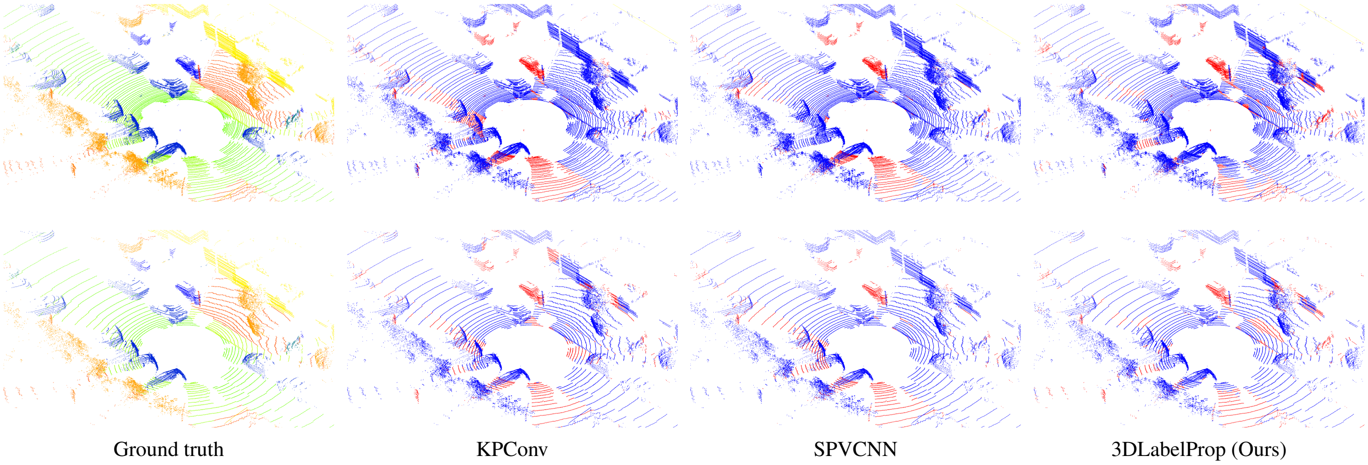


Figure 4: Qualitative results when trained on SemanticKITTI and tested on SemanticKITTI (top row) and SemanticKITTI32 (bottom row). For illustration, the same frame was selected for SemanticKITTI and SemanticKITTI32. From left to right: Ground truth labels, results from KPConv, results from SPVCNN, results from 3DLabelProp. In blue, points with correct semantic segmentations. In red, errors.
We can observe the decrease in performance from SemanticKITTI to SemanticKITTI32 of SPVCNN on the left sidewalk which is badly labeled for SemanticKITTI32. For KPConv, the buildings on the right are errors appearing only with SemanticKITTI32. For 3DLabelProp, while the segmentation is slightly worse than SPVCNN on the full resolution, because the decrease of performance is small between SemanticKITTI and SemanticKITTI32, the overall quality for SemanticKITTI32 is on par with SPVCNN.

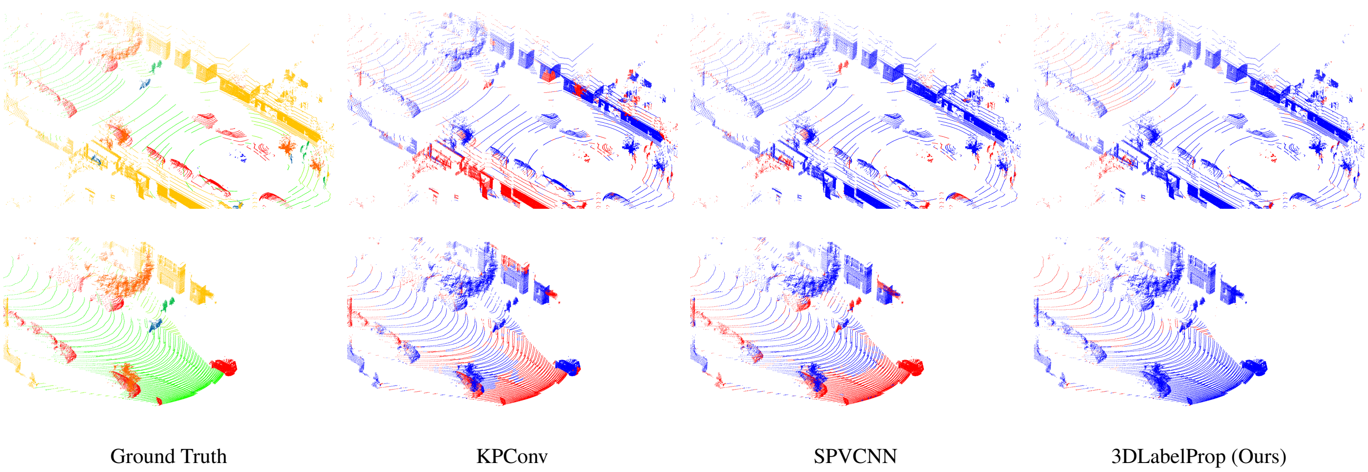
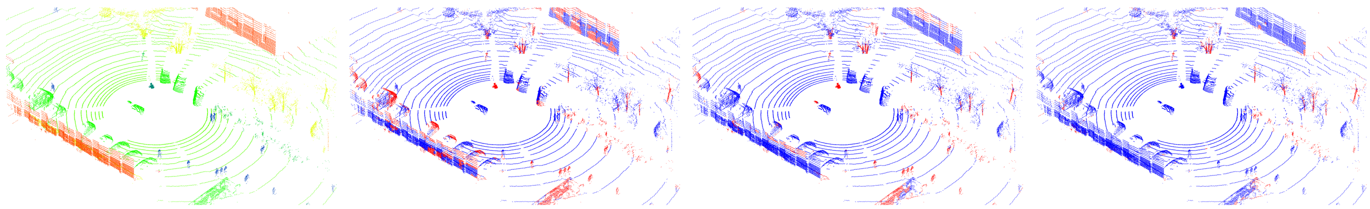
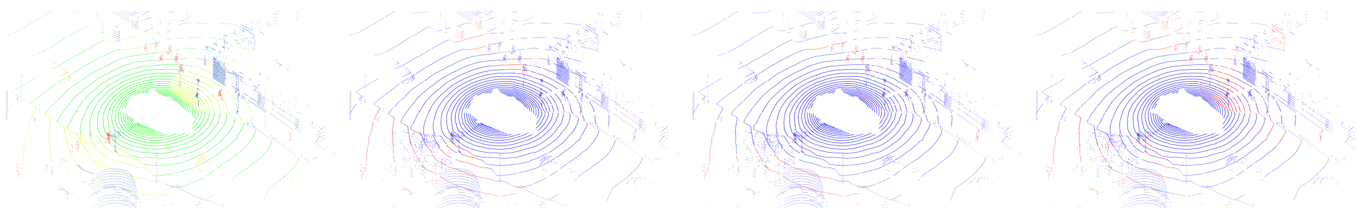


Figure 5: Qualitative results when trained on SemanticKITTI and tested on Panda64 (top row) and PandaFF (bottom row). For comparison in the same environment, two frames acquired at the same time were chosen for Panda64 and PandaFF. From left to right: Ground truth labels, results from KPConv, results from SPVCNN, results from 3DLabelProp. In blue, points with correct semantic segmentation. In red, errors.
For Panda64 we see the ability of 3DLabelProp to properly segment point close to the sensor contrary to SPVCNN where most of its errors lie. The same pattern can be seen for PandaFF where SPVCNN and KPConv have a lot of trouble understanding points with limited neighborhoods, i.e points close to the sensor, whereas 3DLabelProp display the same pattern of errors for Panda64 and PandaFF and has no issue with points with limited neighborhoods.



Ground truth KPConv SPVCNN 3DLabelProp (Ours)

Figure 6: Qualitative results when trained on SemanticKITTI and tested on SemanticPOSS. From left to right: Ground truth labels, results from KPConv, results from SPVCNN, results from 3DLabelProp. In blue, points with correct semantic segmentation. In red, errors.
In these images, we see the ability of 3DLabelProp to properly recognize all the parked cars, in contrary to KPConv. Buildings are also more accurately detected for 3DLabelProp compared to SPVCNN. We see that trunks are systematically badly segmented.



Ground truth KPConv SPVCNN 3DLabelProp (Ours)

Figure 7: Qualitative results when trained on SemanticKITTI and tested on nuScenes. From left to right: Ground truth labels, results from KPConv, results from SPVCNN, results from 3DLabelProp. In blue, points with correct semantic segmentation. In red, errors.
We see the higher quality of KPConv on pedestrians compared to SPVCNN and 3DLabelProp.

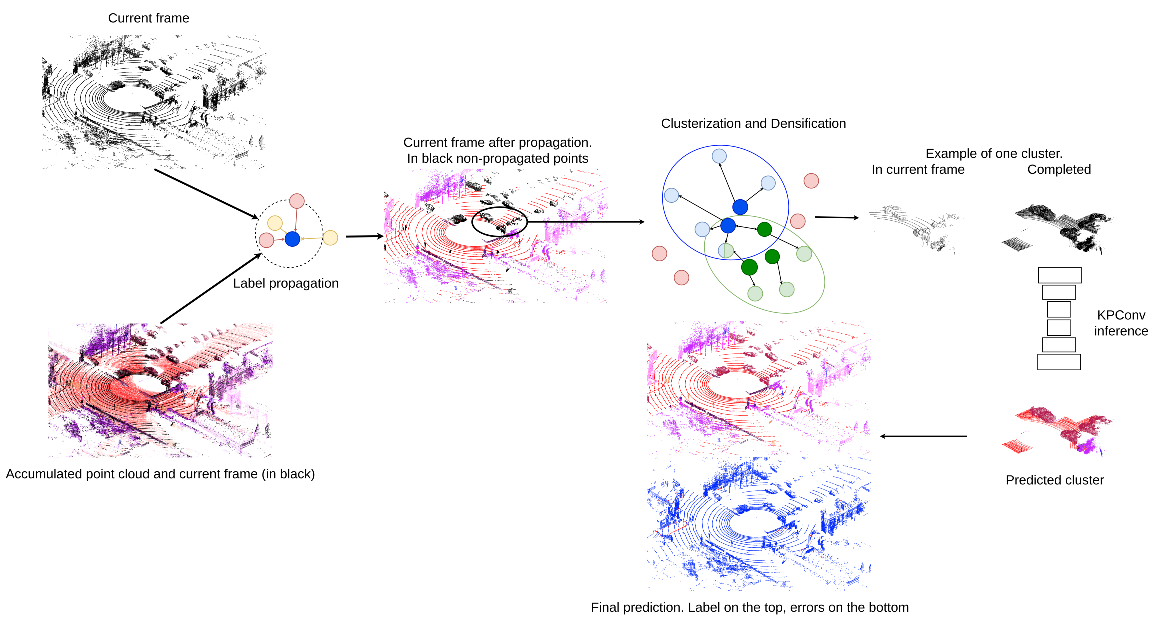


Figure 8: Example of the 3DLabelProp pipeline. Trained on SemanticKITTI and tested on SemanticPOSS.