Supplementary Material for PointFlowNet: Learning Representations for Rigid Motion Estimation from Point Clouds

Aseem Behl Despoina Paschalidou Simon Donné Andreas Geiger
Autonomous Vision Group, MPI for Intelligent Systems and University of Tübingen
{aseem.behl,despoina.paschalidou,simon.donne,andreas.geiger}@tue.mpg.de

Abstract

The supplementary material provides additional qualitative comparisons of our method with the baselines methods along with examples of failure cases for our method. We also describe our workflow for generating the Augmented KITTI dataset and provide the hyperparameter values for balancing weights in our multi-task loss function. In addition, we provide a video of our qualitative results on sequences from KITTI test dataset.

1. Augmented KITTI

Figure 1 describes our workflow for generating the Augmented KITTI. We start by fitting the ground plane using RANSAC 3D plane fitting; this allows us to detect obstacles and hence the drivable region. In a second step, we randomly place virtual cars in the drivable region, and simulate a new LIDAR scan that includes these virtual cars. Our simulator uses a noise model learned from the real KITTI scanner by fitting a Gaussian distribution conditioned on the horizontal and vertical angle of the rays, based on KITTI LIDAR scans. Our simulator also produces missing estimates at transparent surfaces by ignoring them with a probability equal to their transparency value provided by the CAD models. Additionally, we remove points in the original scan which become occluded by the augmented car by tracing a ray between each point and the LIDAR, and removing those points whose ray intersects with the car mesh. Finally, we sample the augmented car's rigid motion using a simple approximation of the Ackermann steering geometry, place the car at the corresponding location in the next frame, and repeat the LIDAR simulation.

2. Hyperparameters

The loss functions used by our approach comprises four parts:

$$\mathcal{L} = \alpha\mathcal{L}_{flow} + \beta\mathcal{L}_{rigmo} + \gamma\mathcal{L}_{ego} + \mathcal{L}_{det} \qquad (1)$$

Here, α, β, γ are positive constants for balancing the relative importance of the task specific loss functions. In order to fix the balancing weights in our multi-task loss function, we performed a hyperparameter search over the loss function weights, maximizing the performance over the scene flow and object rigid motion. We set $\alpha = 4.0$, $\beta = 1.0$, $\gamma = 1.0$ in Eqn. 1.

3. Qualitative Comparison to Baseline Methods

Figures 2 - 23 show qualitative comparison of our method with the best performing baseline methods on examples from the test set of the Augmented KITTI dataset. The qualitative results show that our method predicts motion for both background and foreground parts of the scene with higher accuracy than all the baselines on a diverse range of scenes and motions.

In Figures 21 - 23, we provide challenging examples where our method fails to predict the correct scene flow. We observe here that in case of scenes with two or more cars in very close proximity, our method may predict wrong scene flow for points on one car in the reference point cloud at frame t by matching them with points on the other car in close proximity at frame $t + 1$. However, we note that, even for these failure cases our method performs better than the baseline methods.

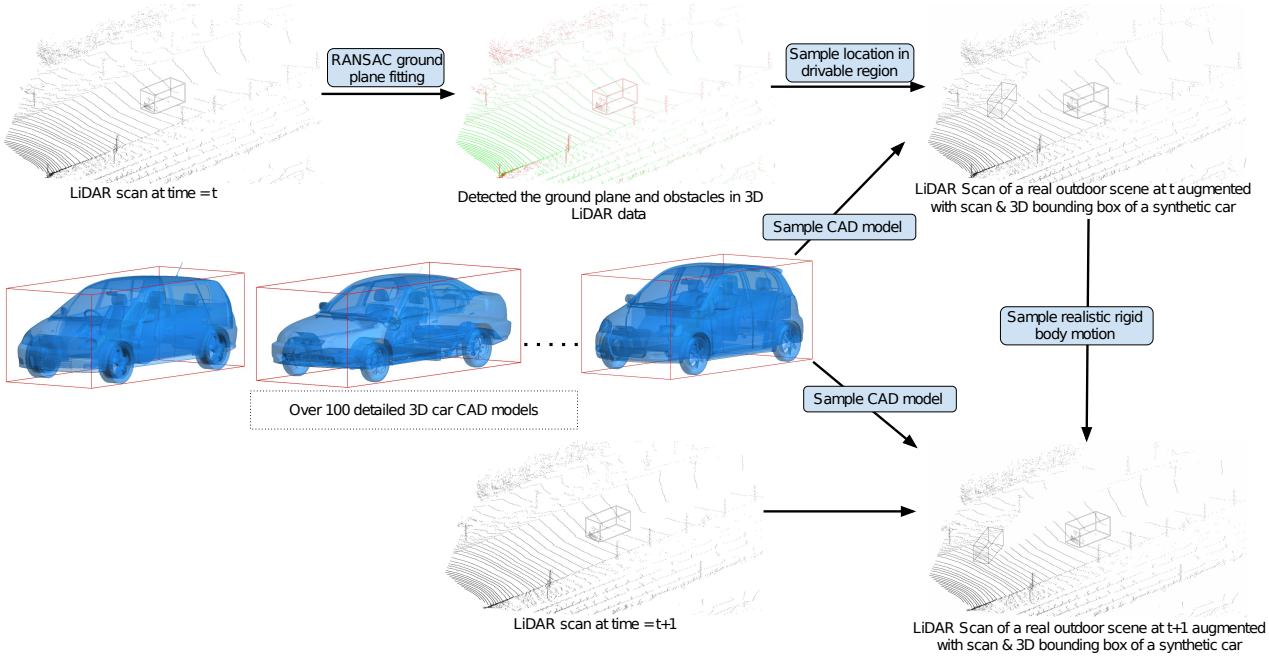

Figure 1: **Augmented KITTI.** Workflow for generating the Augmented KITTI dataset.

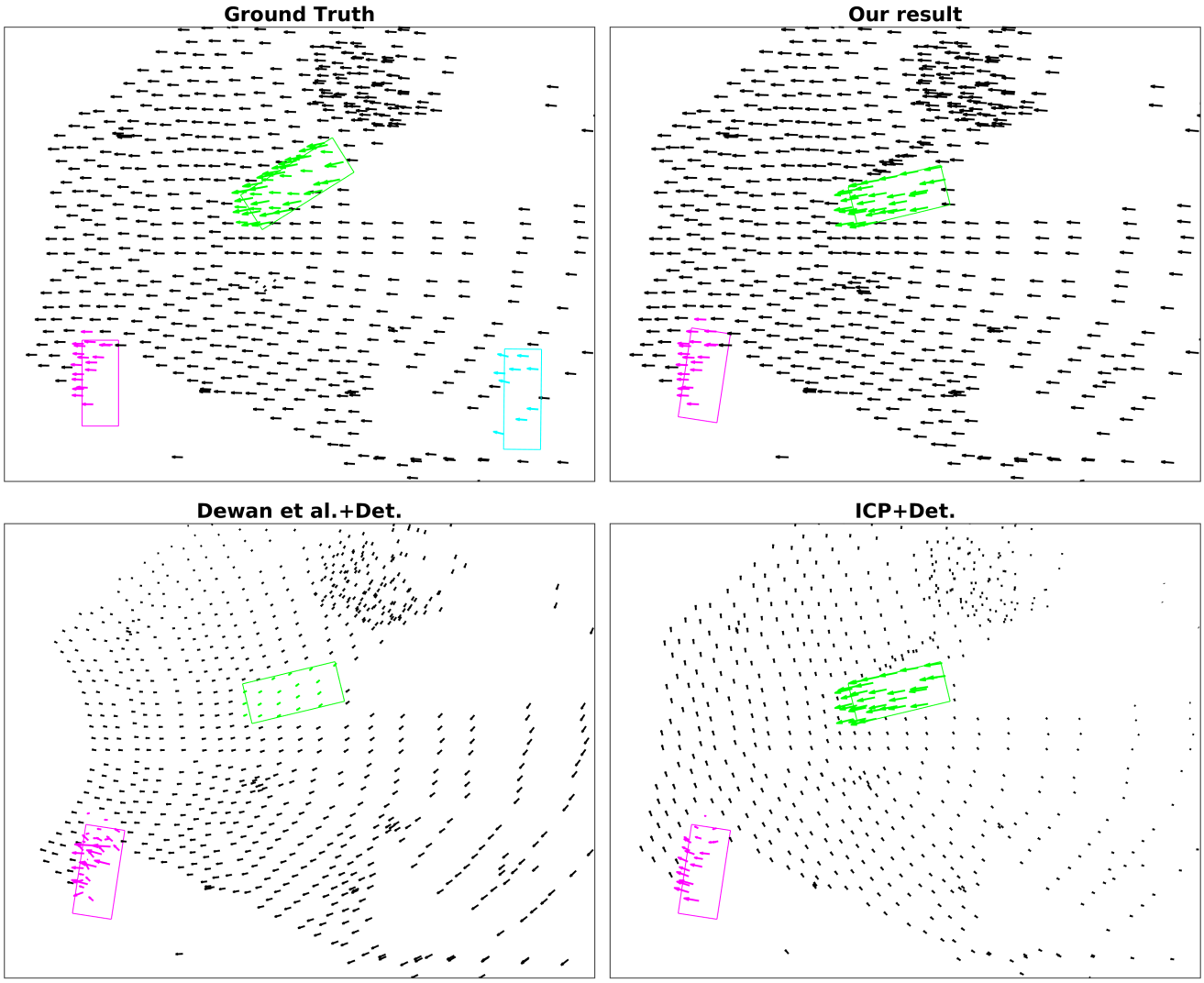

Figure 2: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

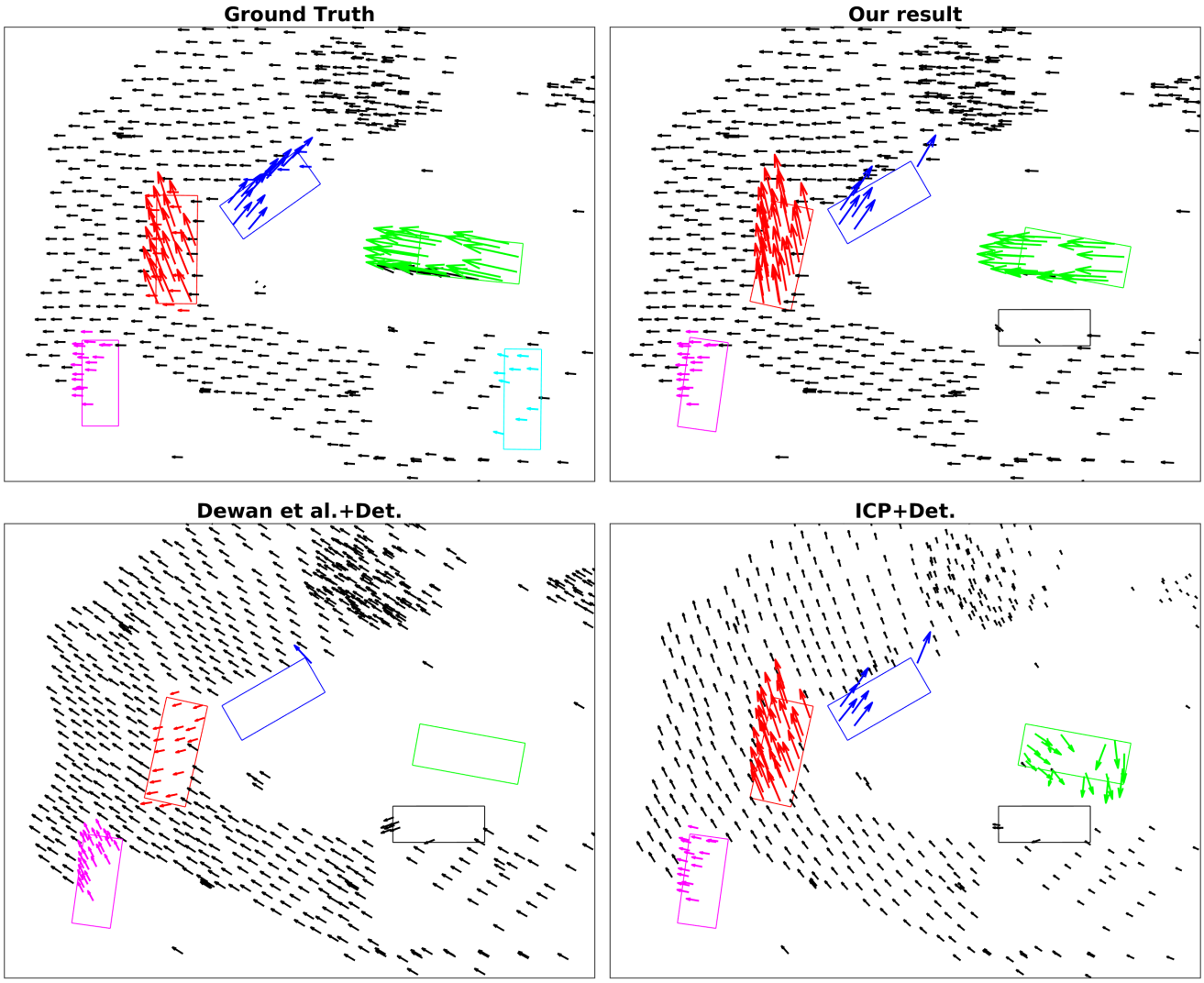

Figure 3: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

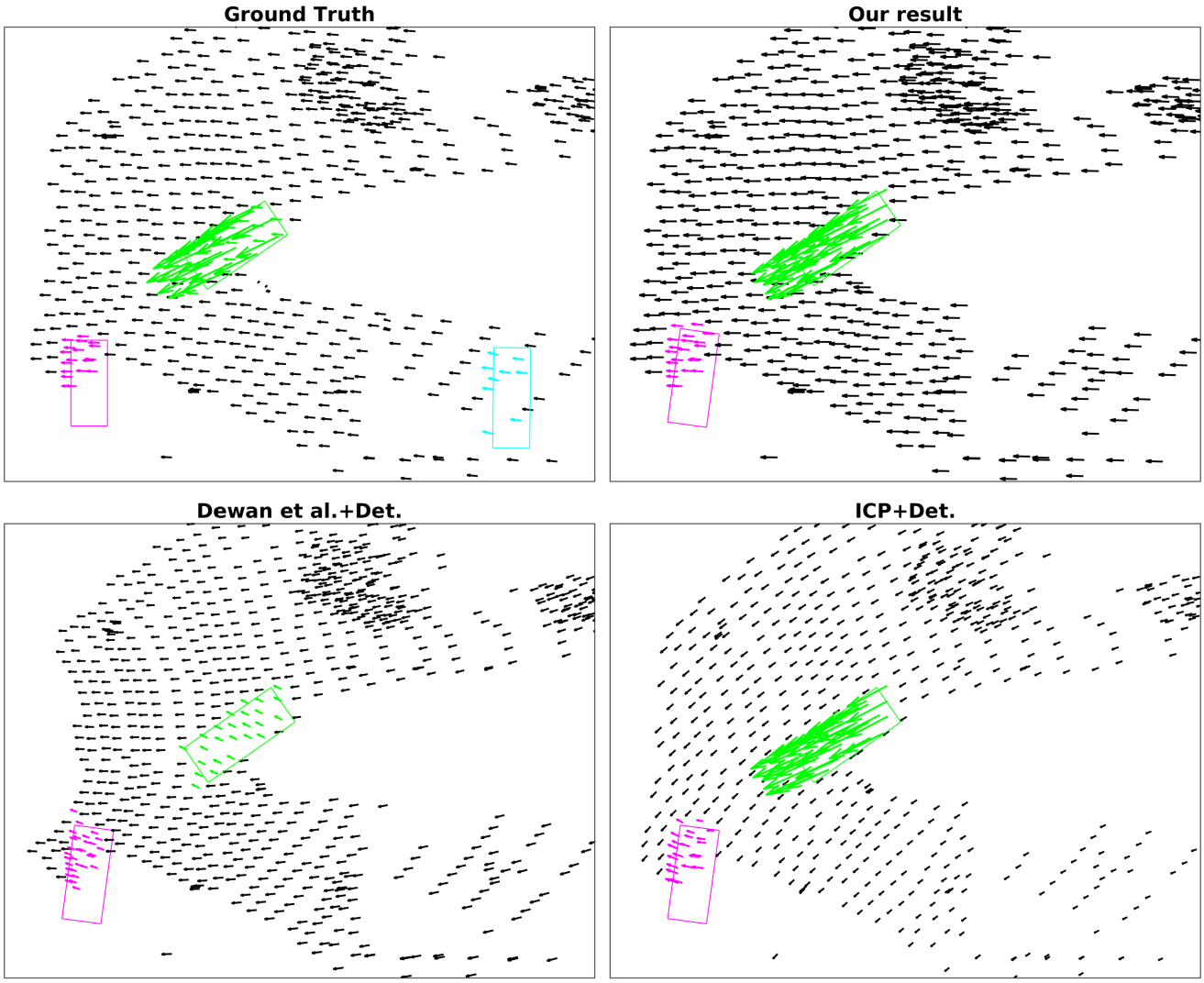

Figure 4: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

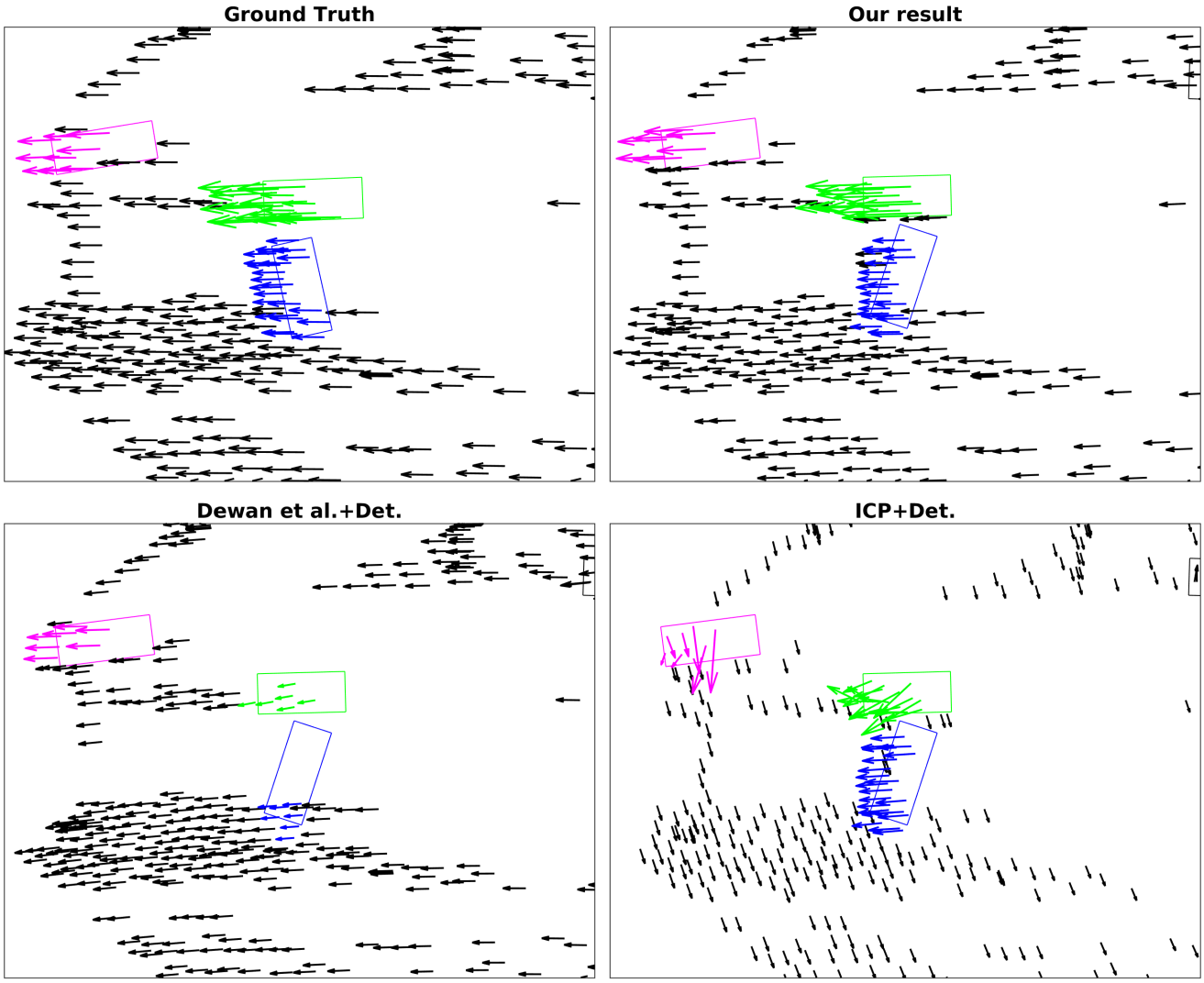

Figure 5: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

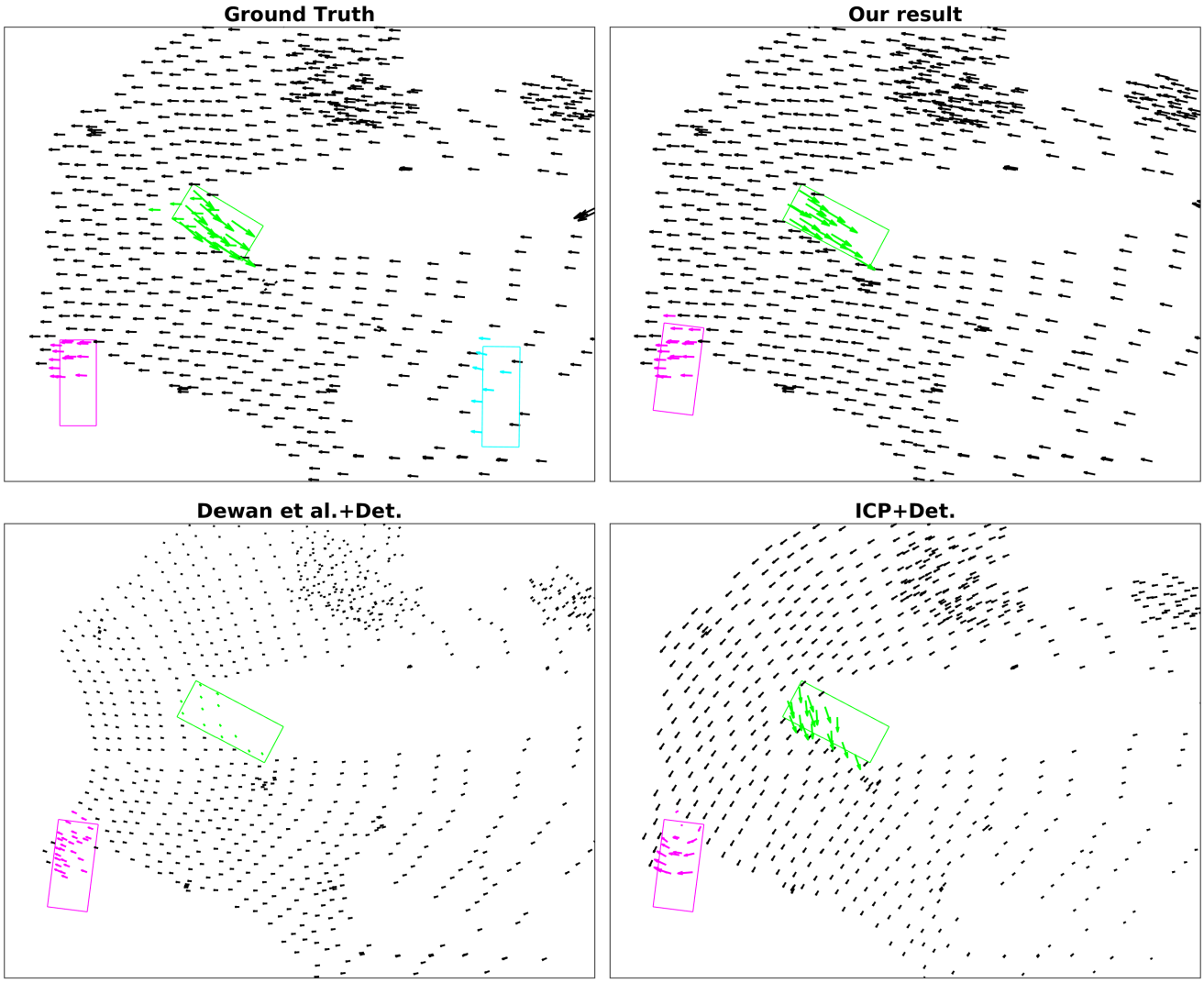

Figure 6: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

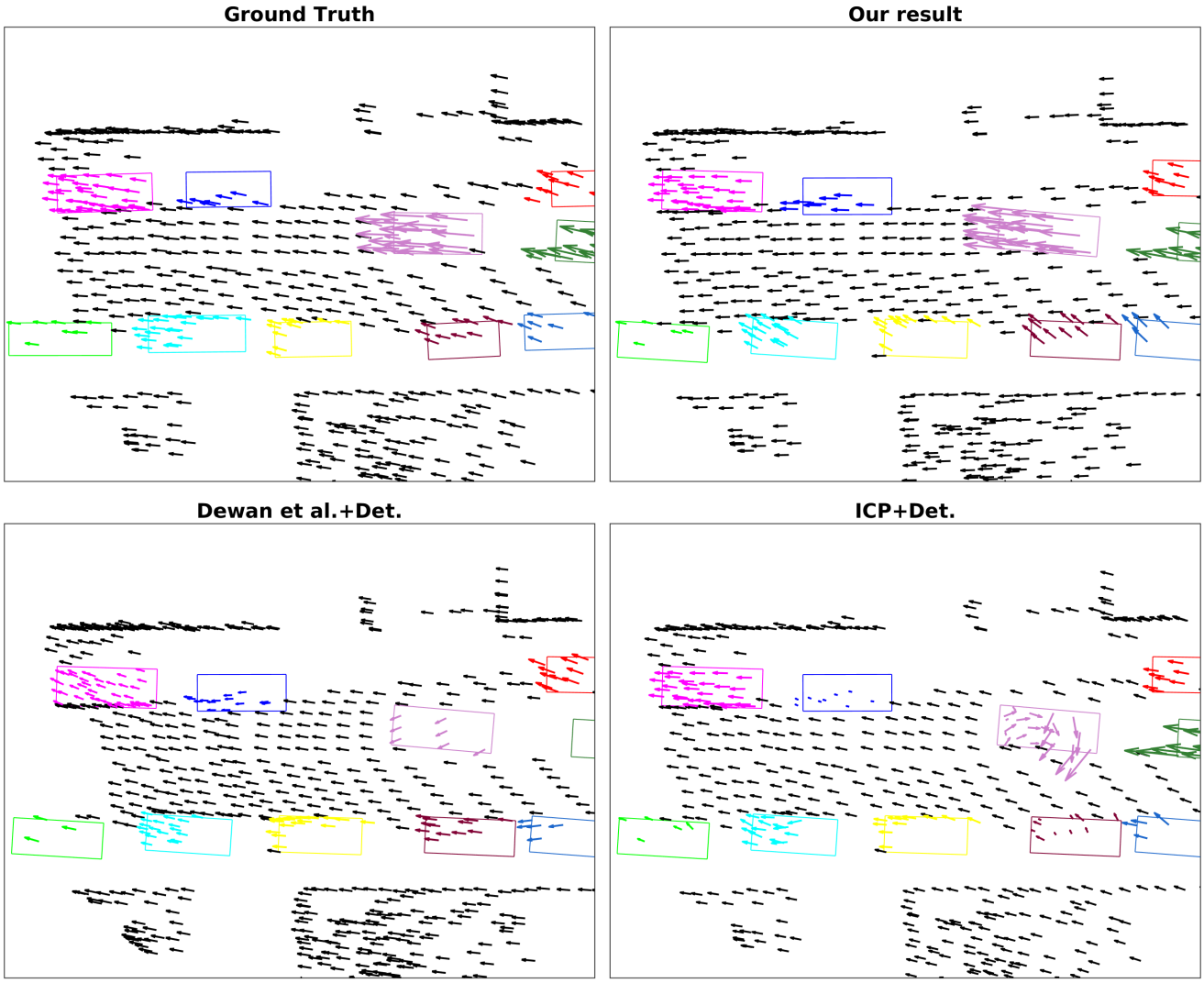

Figure 7: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

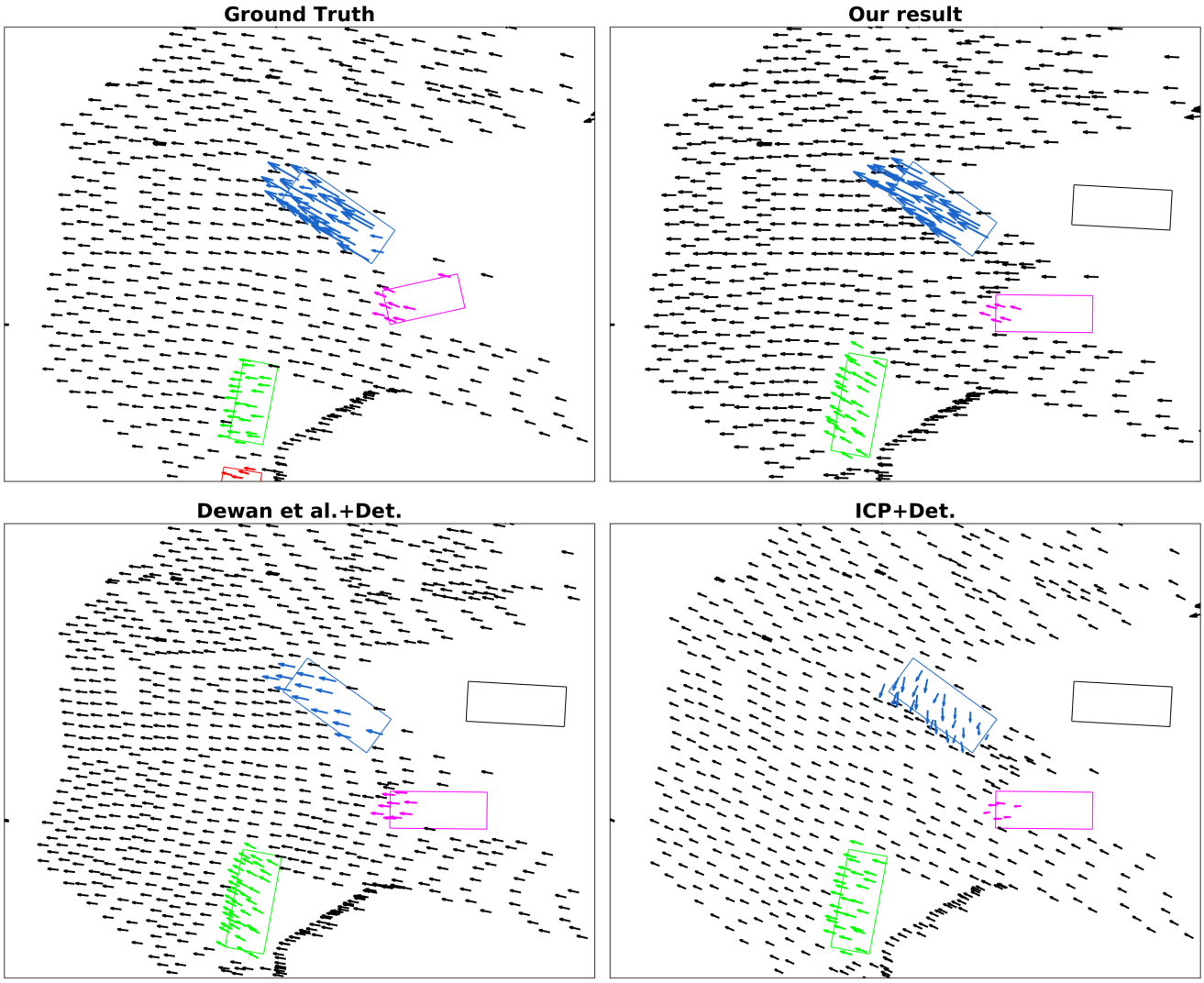

Figure 8: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

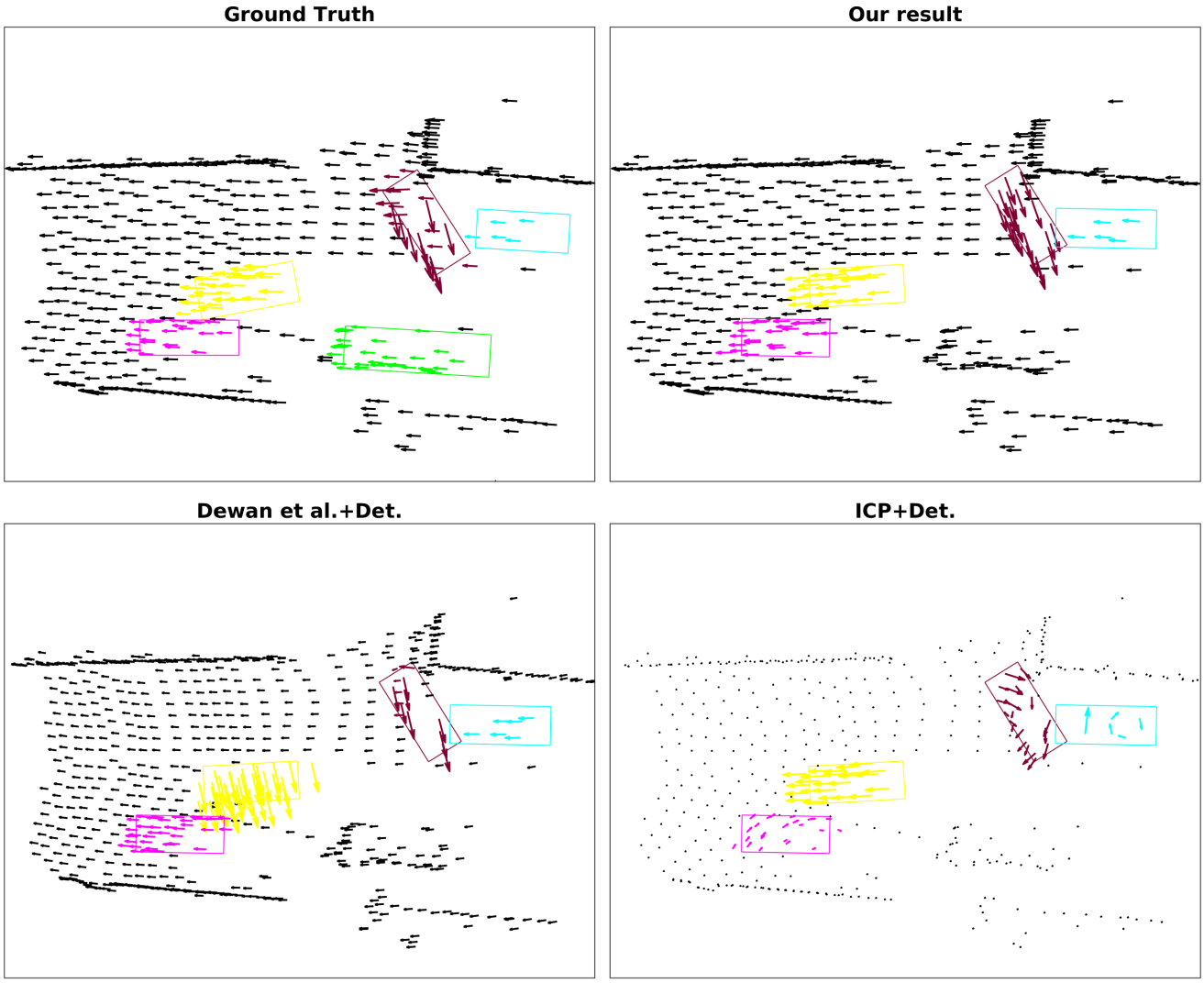

Figure 9: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

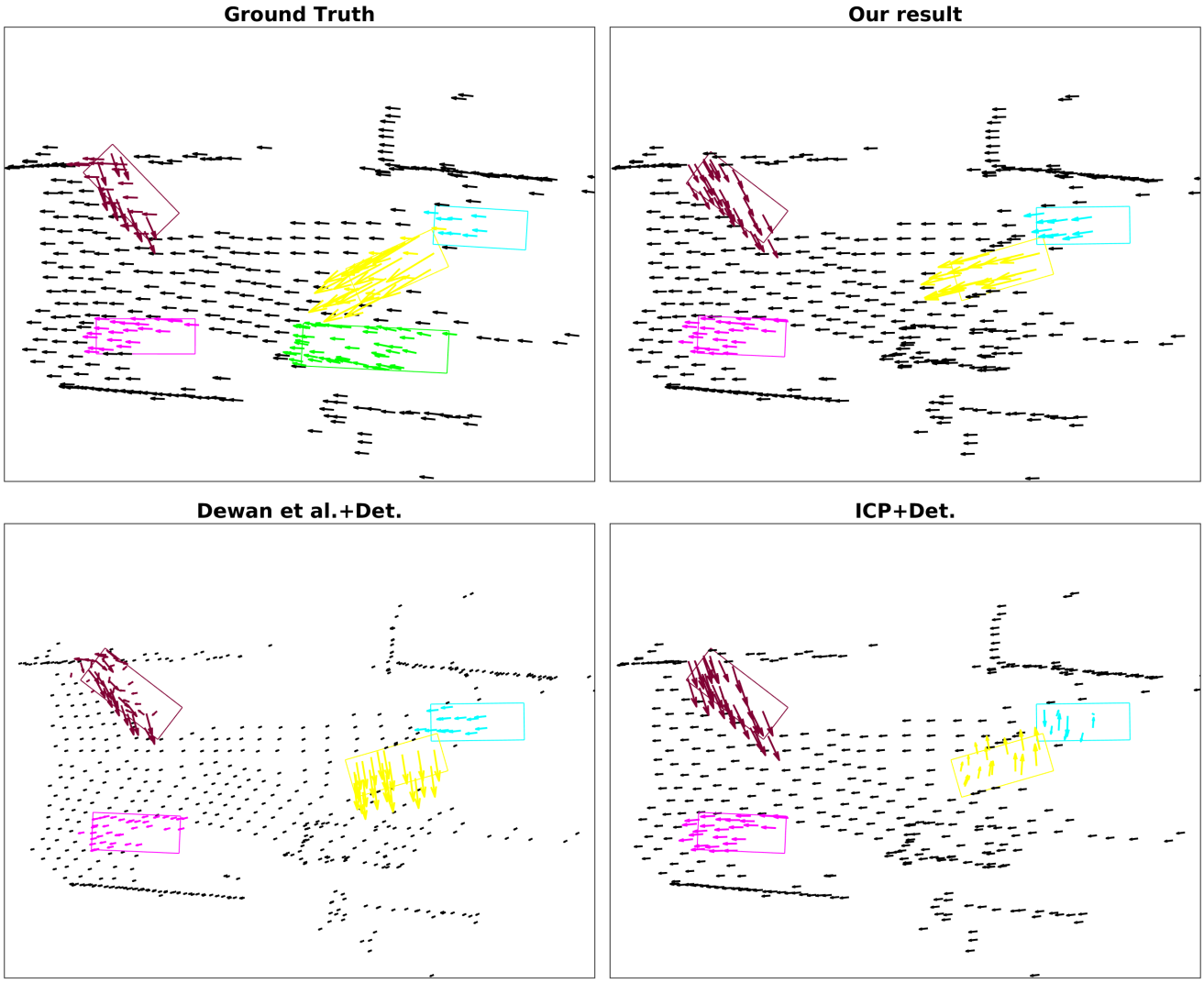

Figure 10: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

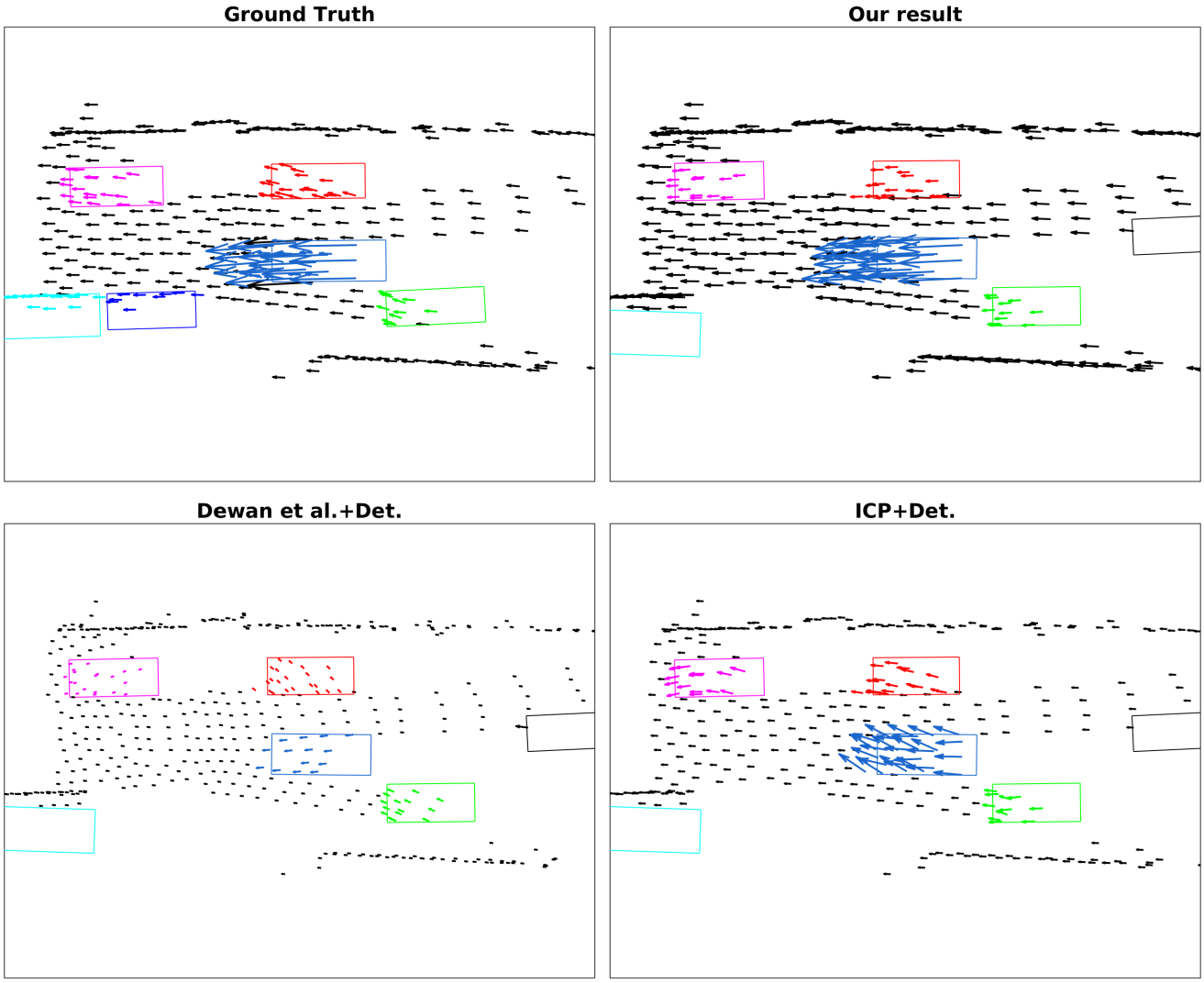

Figure 11: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

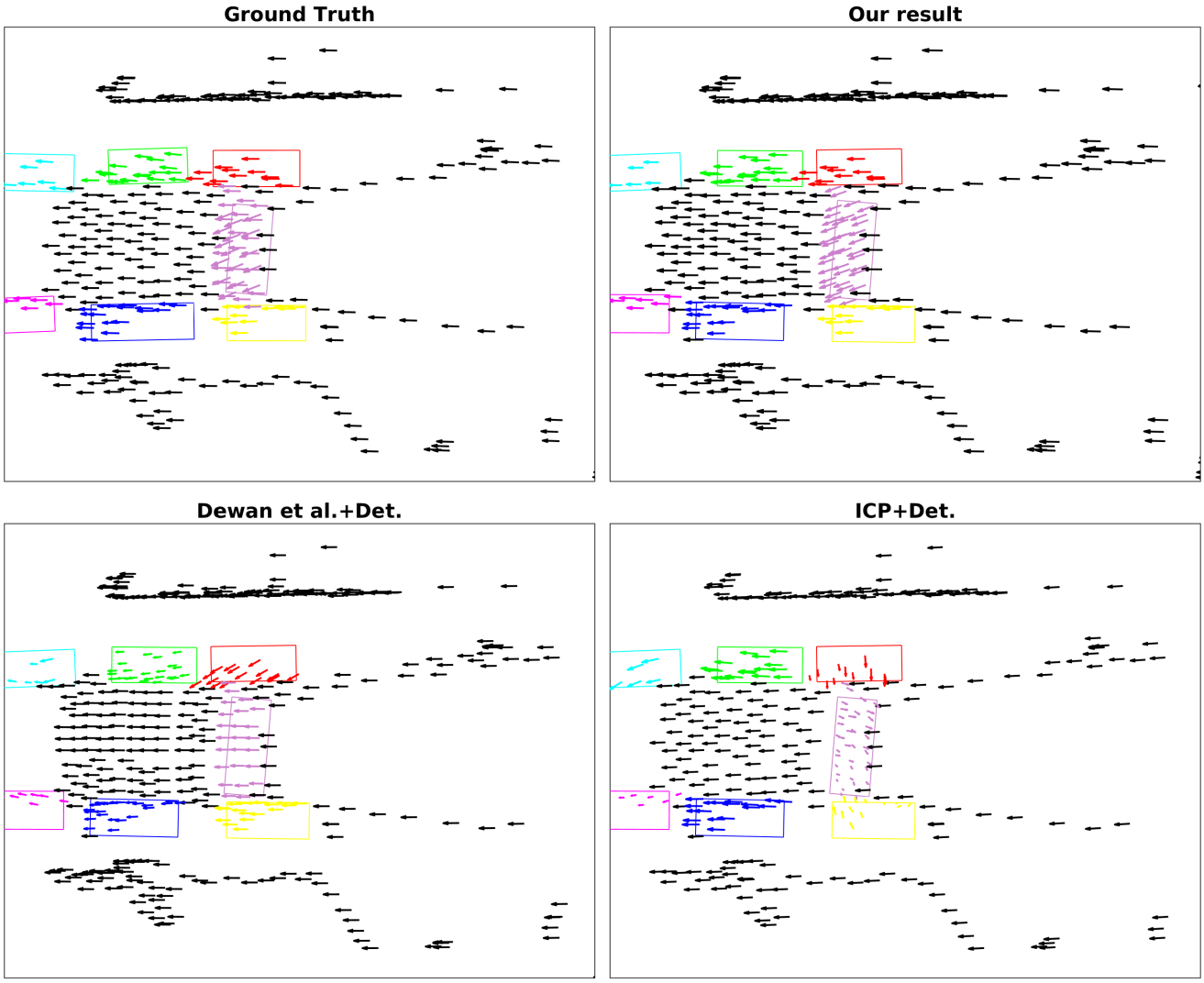

Figure 12: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

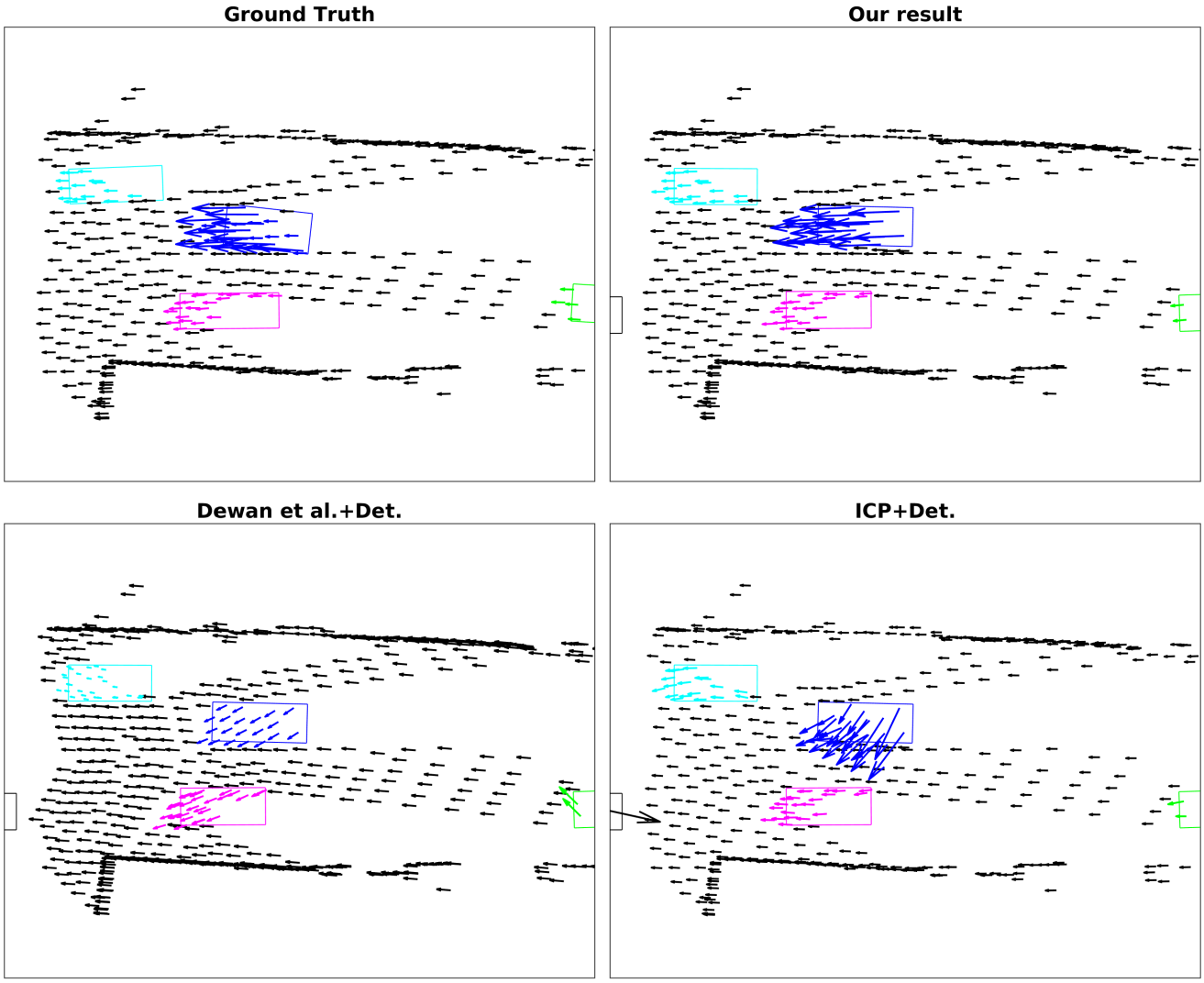

Figure 13: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

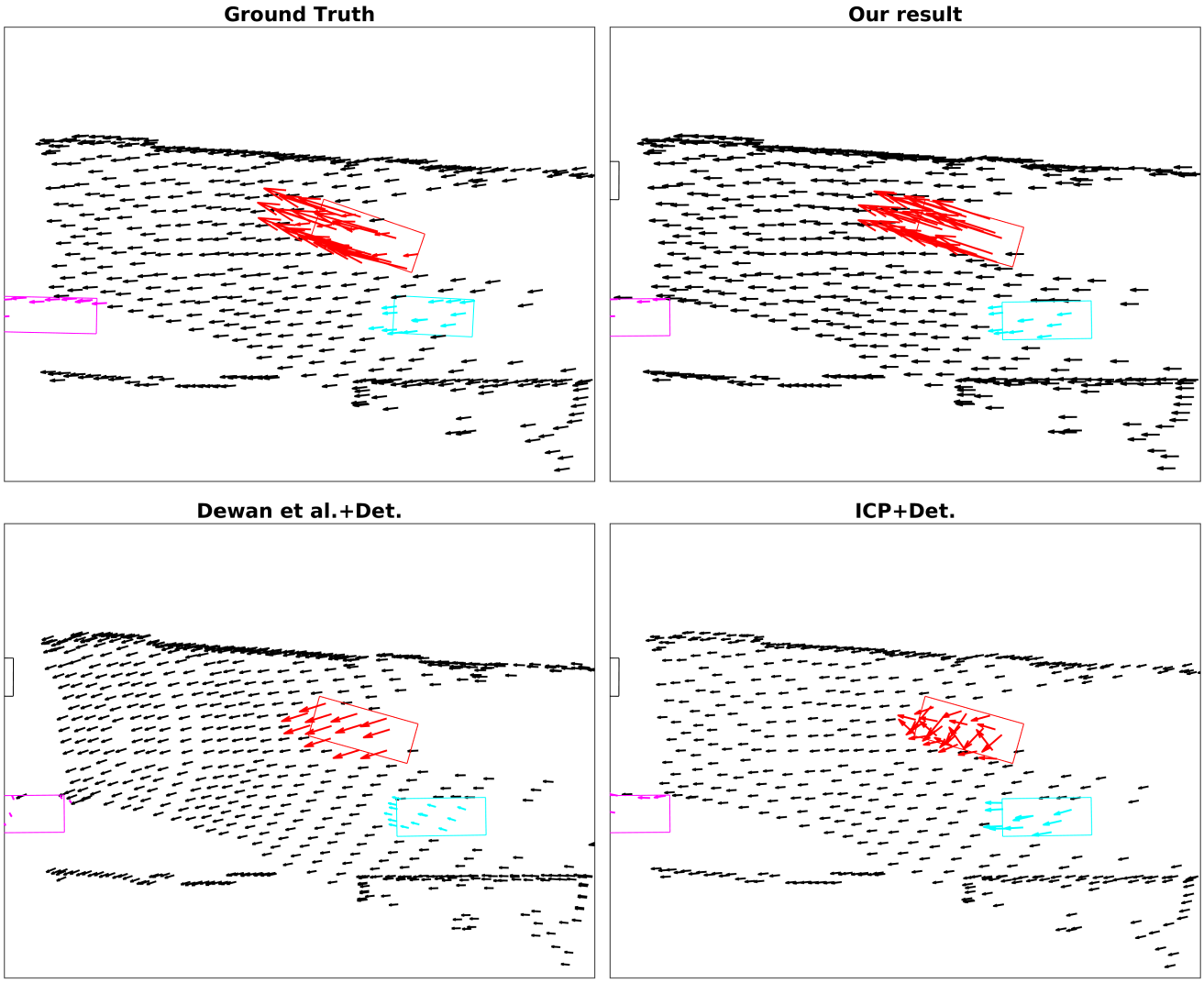

Figure 14: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

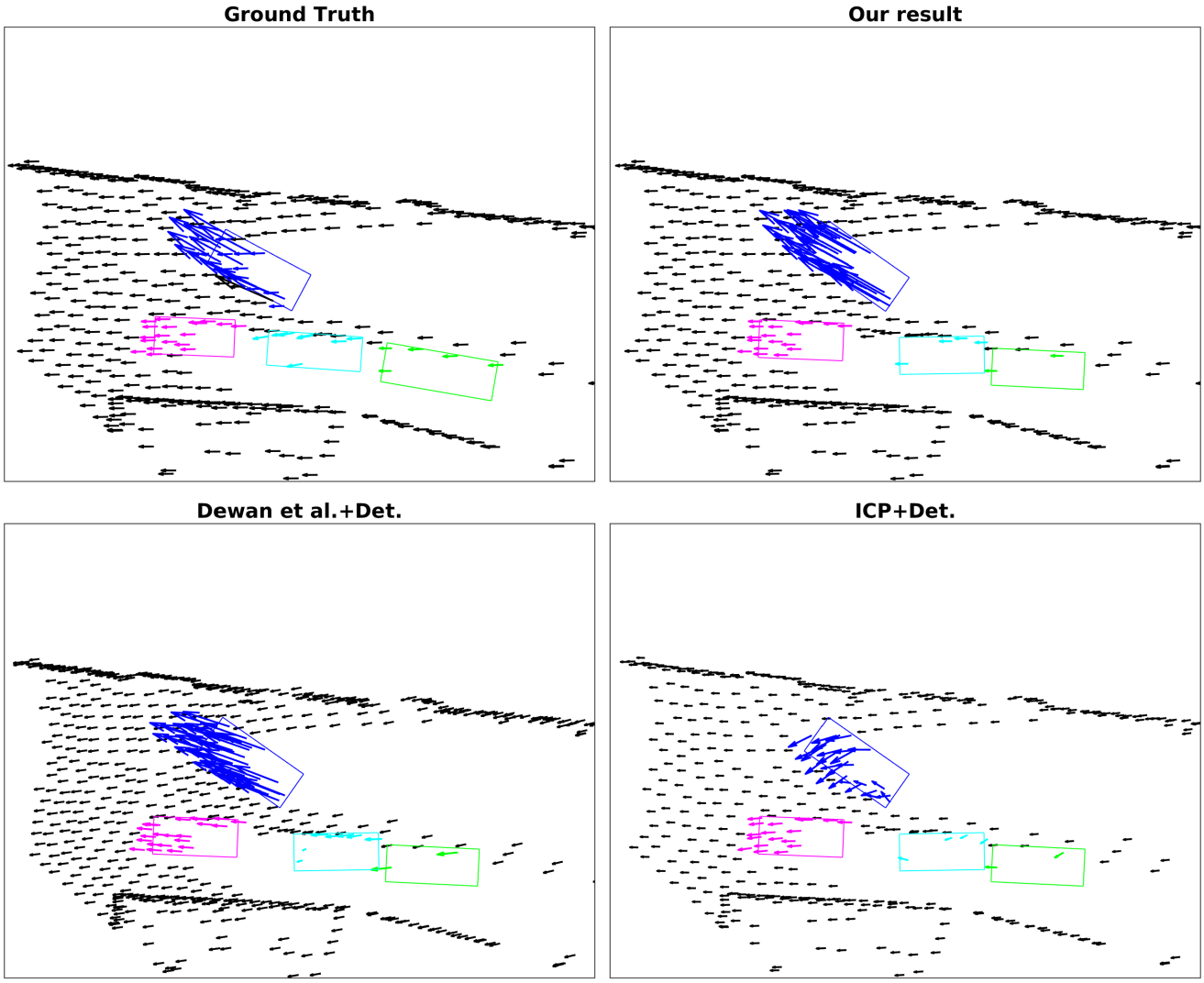

Figure 15: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

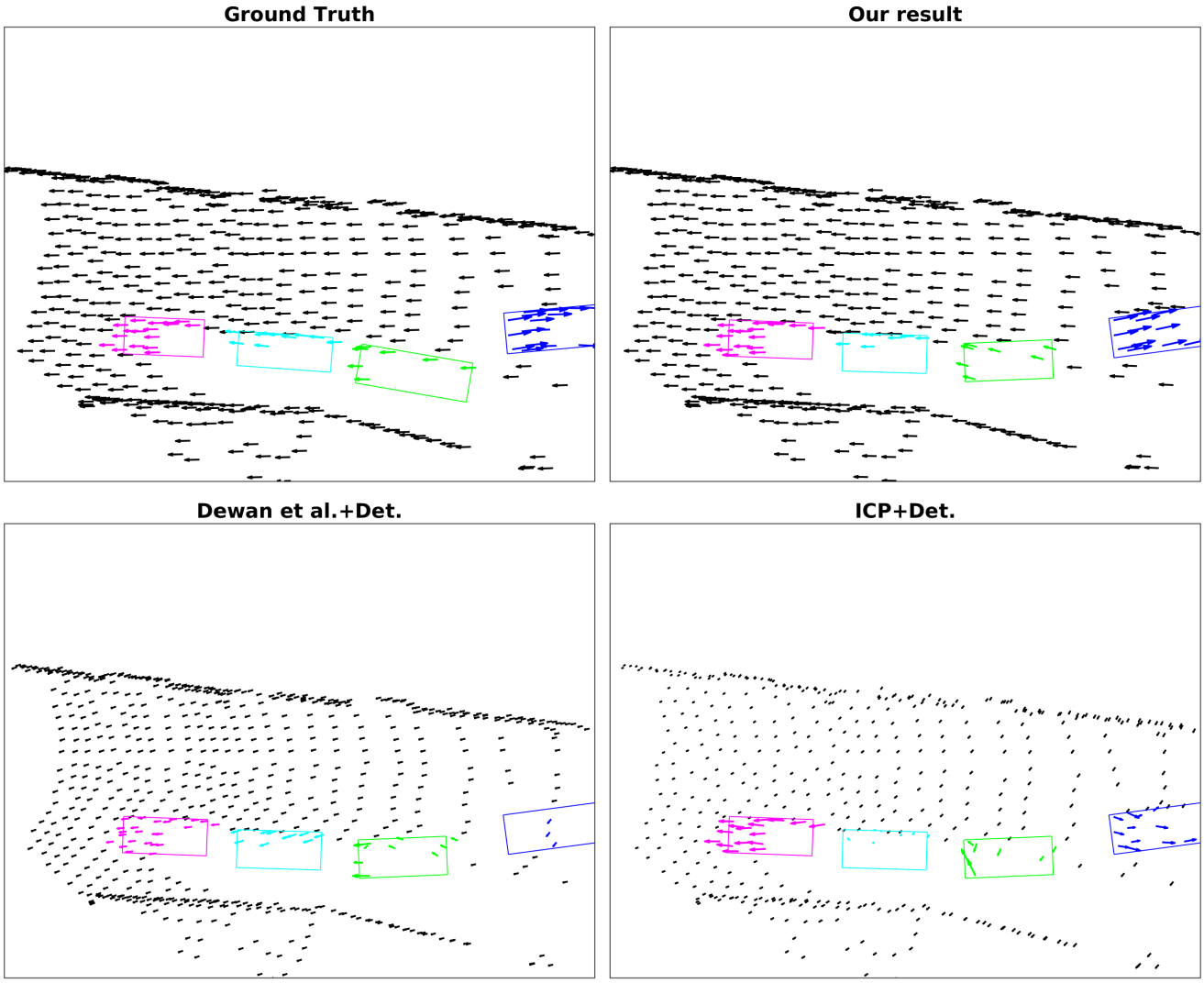

Figure 16: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

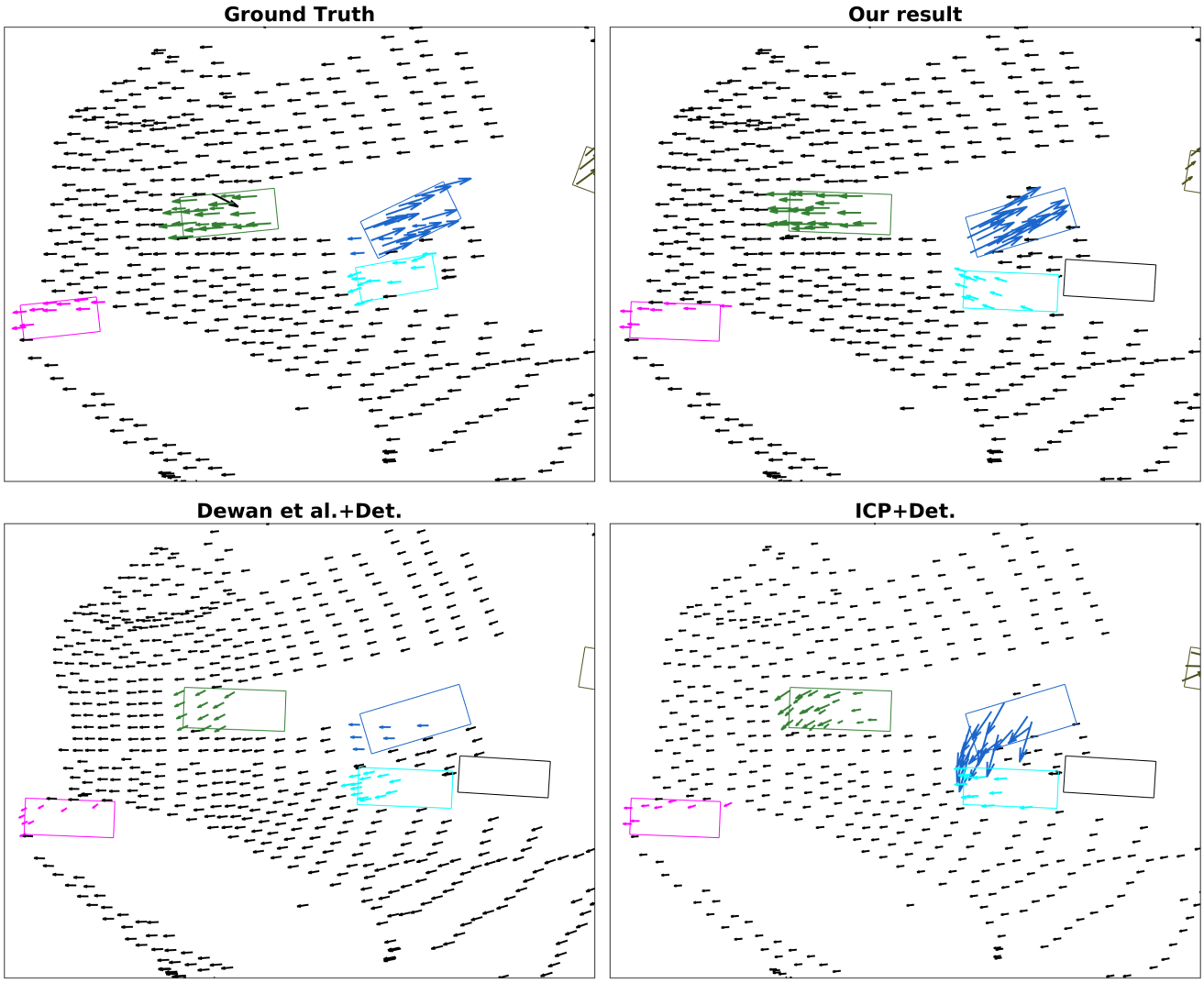

Figure 17: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

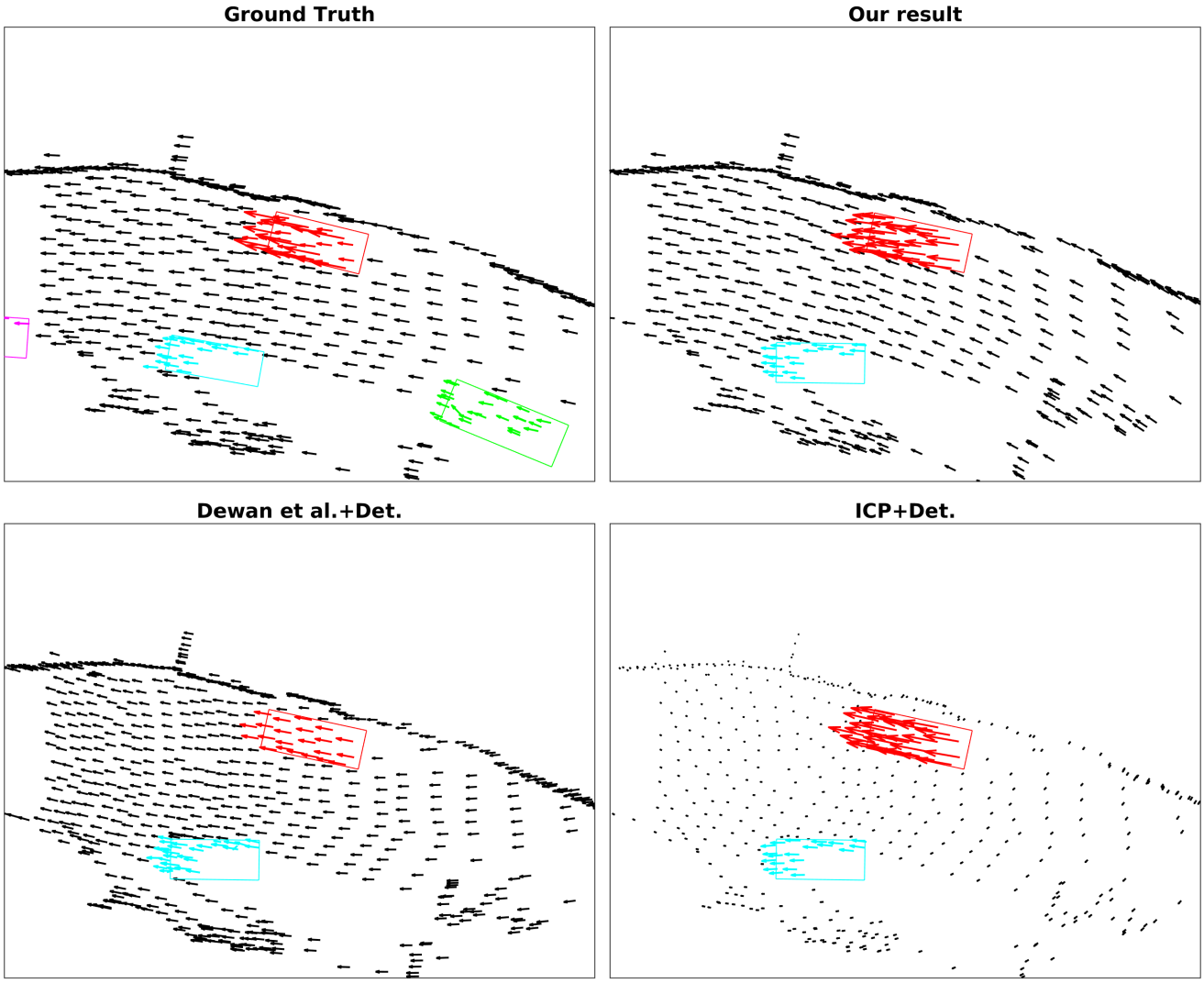

Figure 18: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

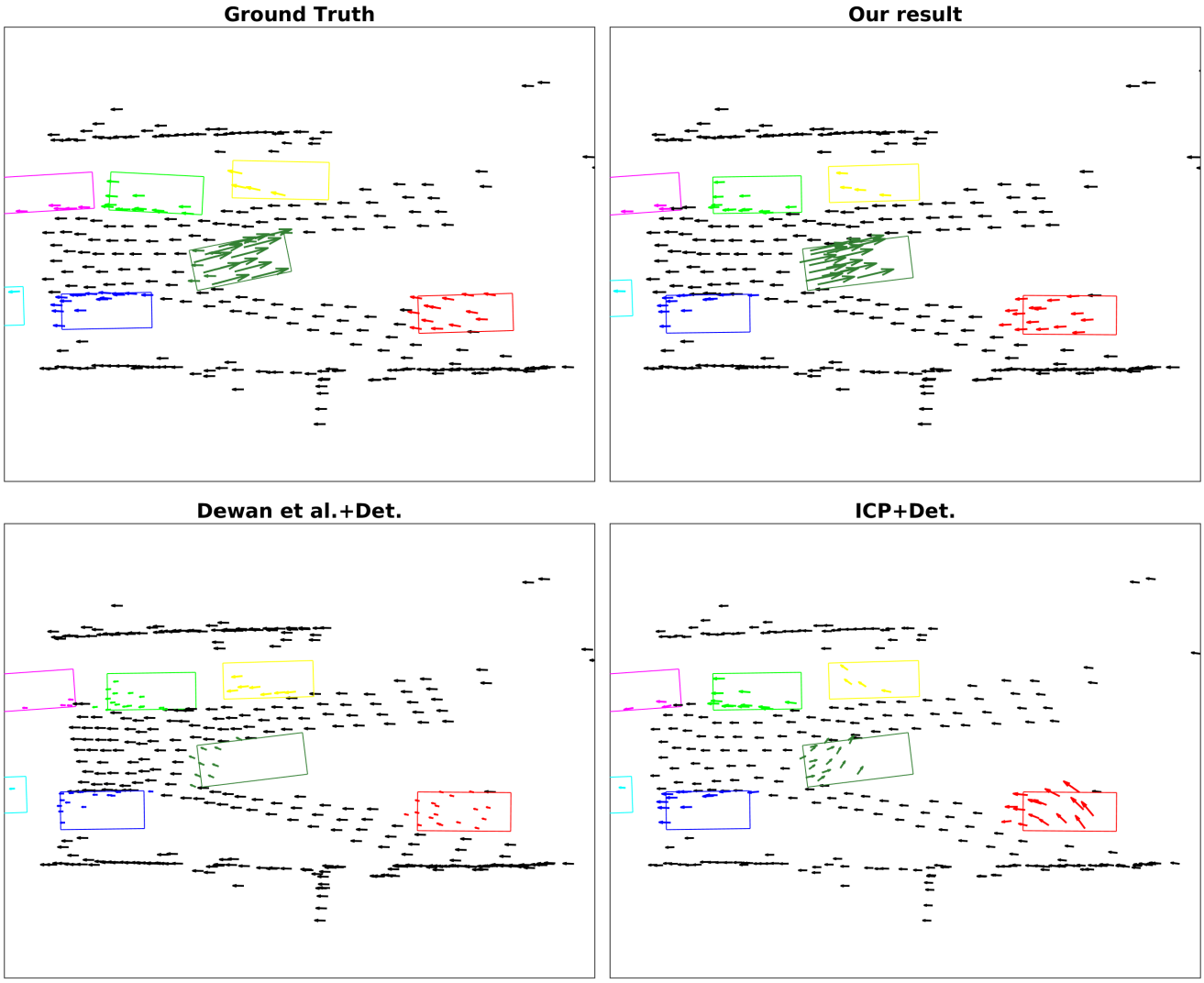

Figure 19: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

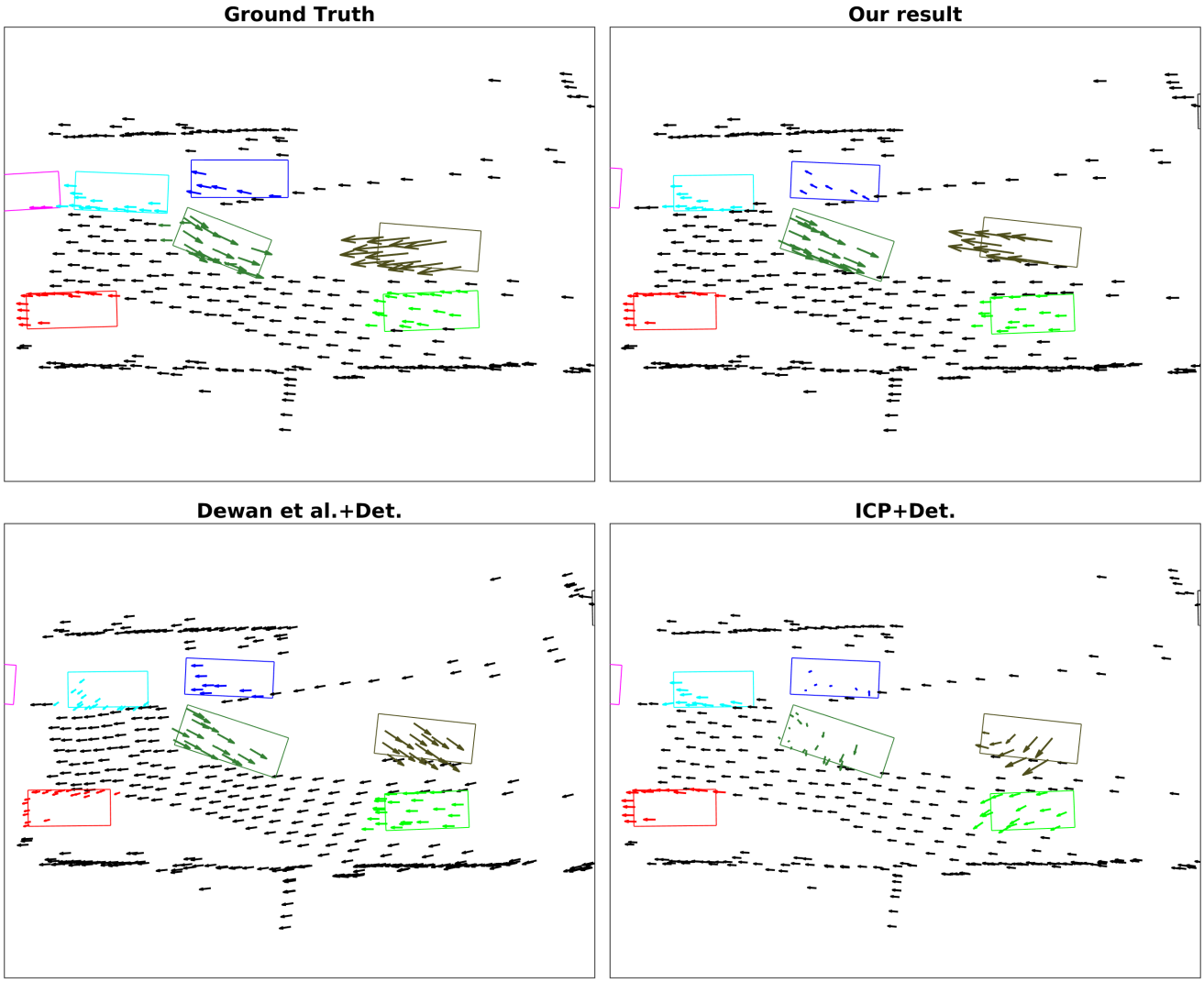

Figure 20: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. For clarity, we visualize only a subset of the points.

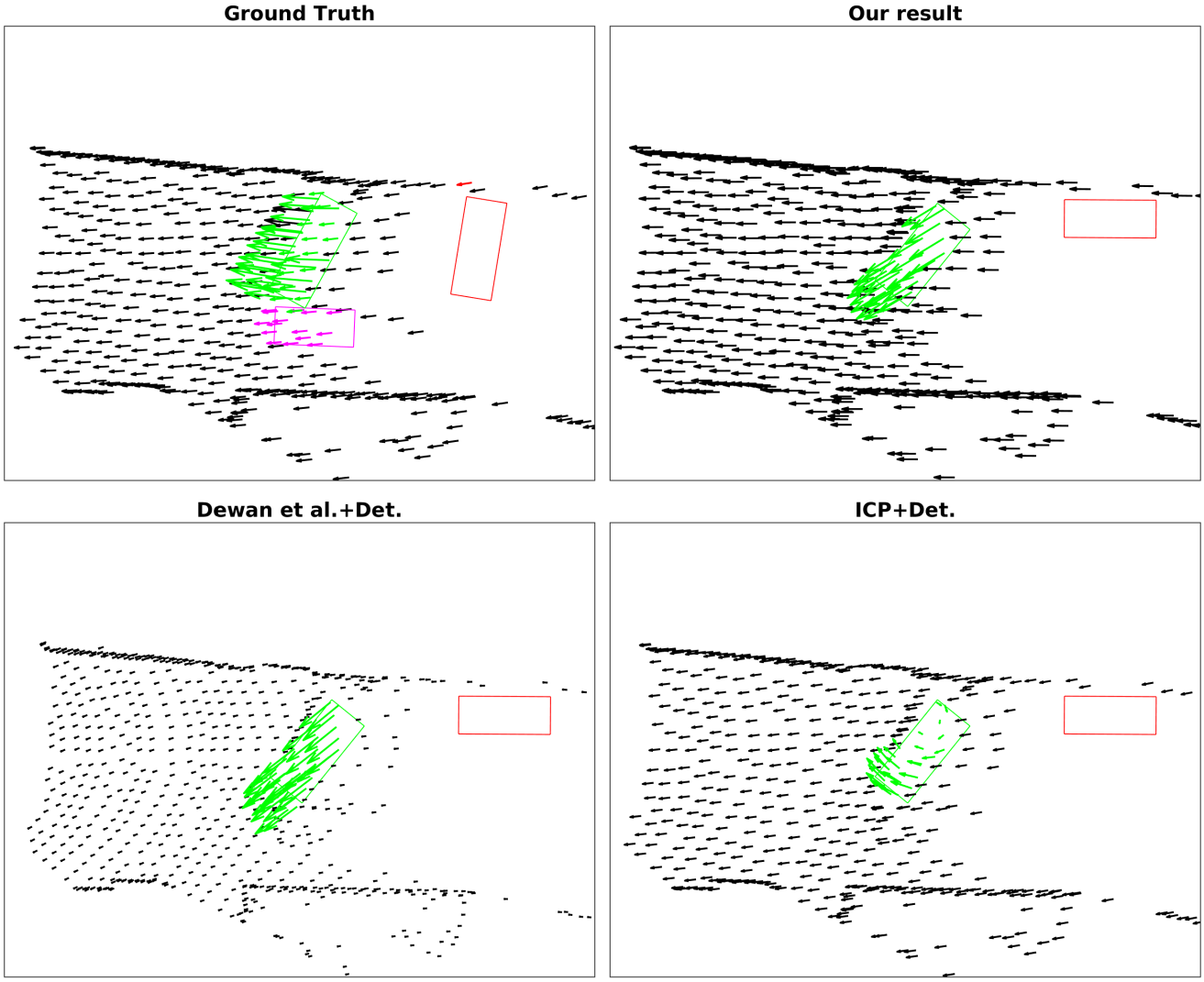

Figure 21: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. We observe here that our method predicts wrong scene flow for points on the green car in the reference point cloud at frame t by matching them with points on the other car (pink) in close proximity at frame $t + 1$. For clarity, we visualize only a subset of the points.

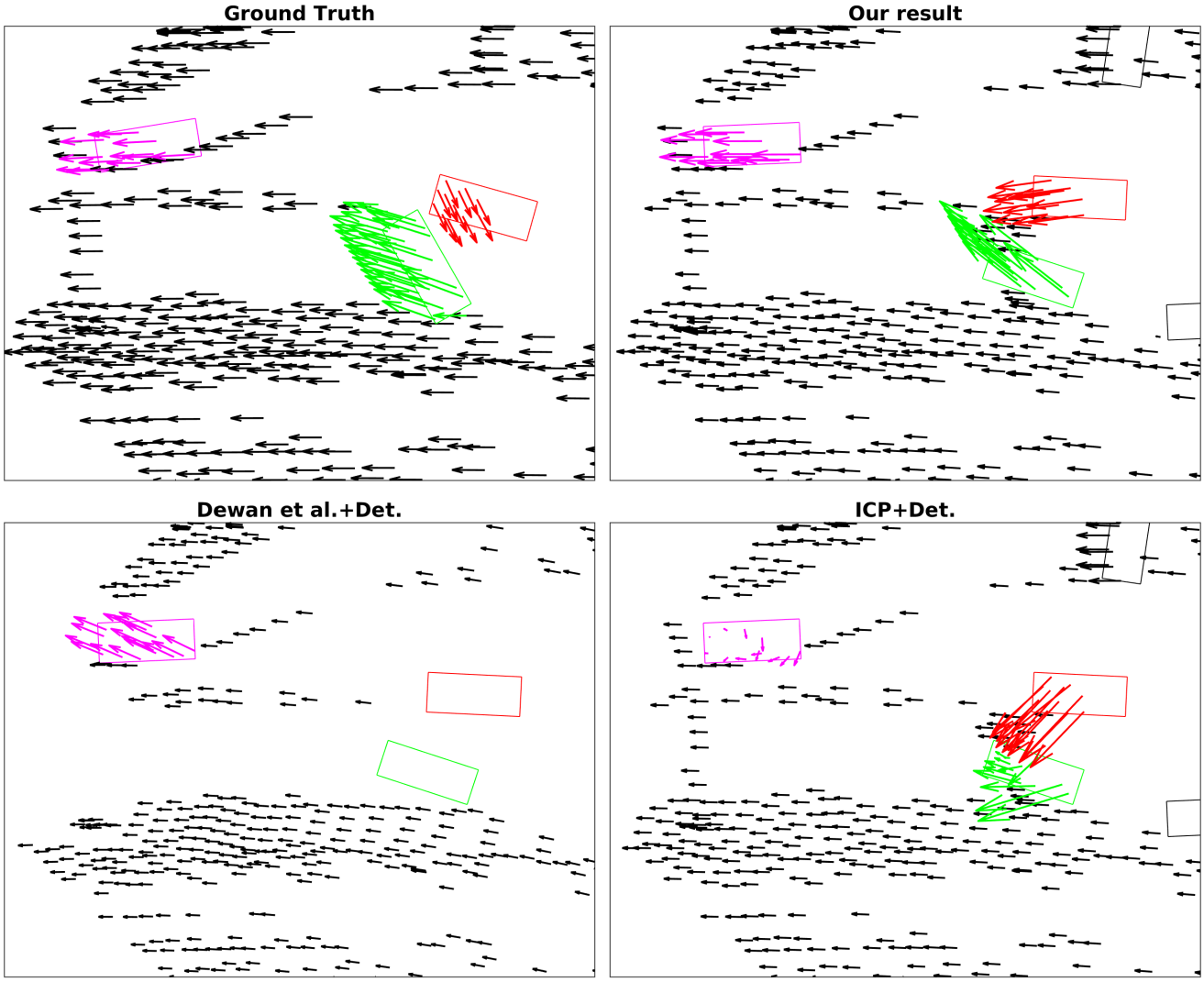

Figure 22: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. We observe here that our method predicts wrong scene flow for points on the red car in the reference point cloud at frame t by matching them with points on the other car (green) in close proximity at frame $t + 1$. For clarity, we visualize only a subset of the points.

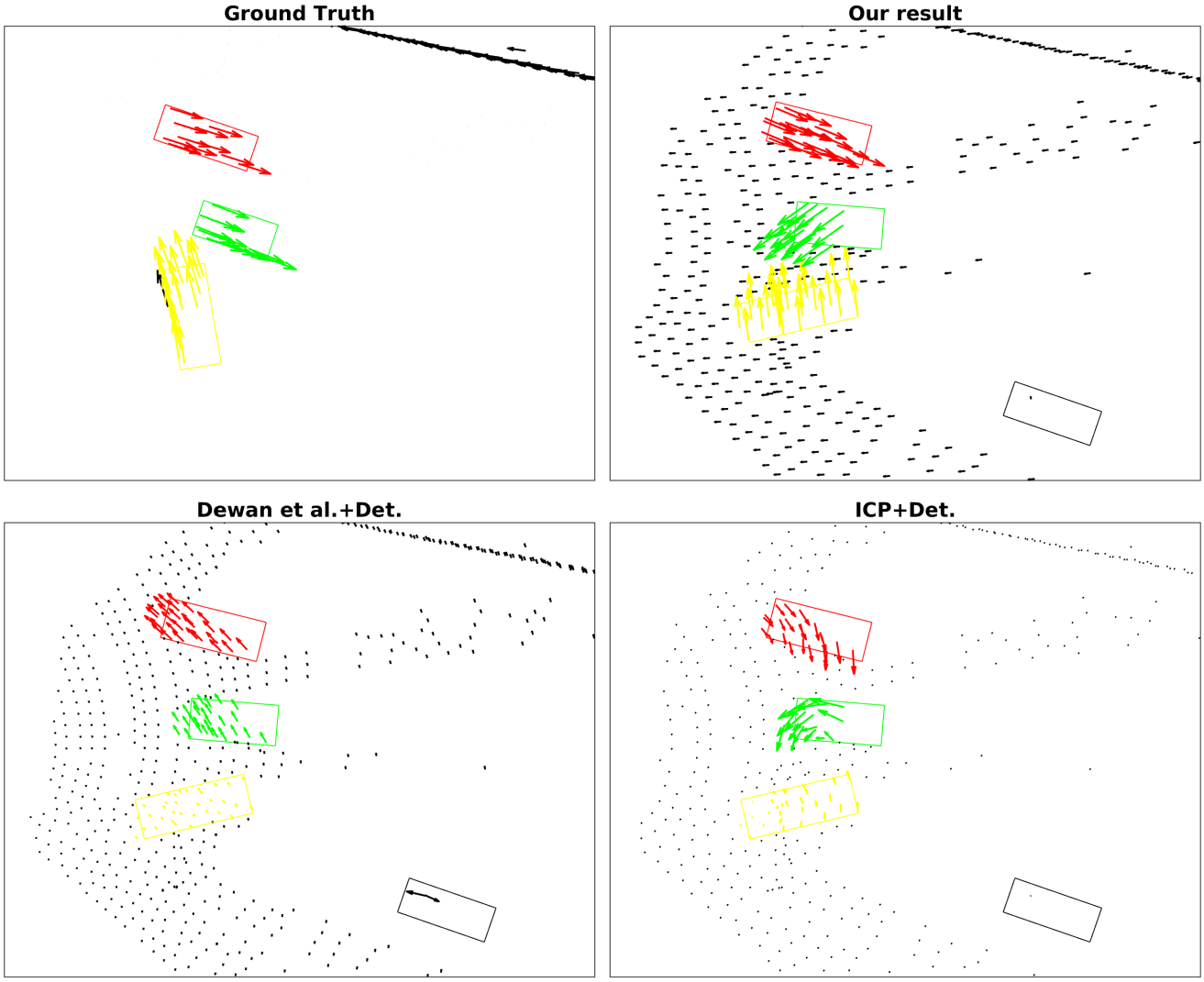

Figure 23: **Qualitative Comparison** of our method with the best performing baseline methods on an example from the test set of the Augmented KITTI dataset. We observe here that our method predicts wrong scene flow for points on the green car in the reference point cloud at frame t by matching them with points on the other car (yellow) in close proximity at frame $t + 1$. For clarity, we visualize only a subset of the points.