法線情報を追加した点群からの３次元物体検出の高精度化

○繆 継樹†,　平川 翼†, 山下 隆義†, 藤吉 弘亘†
†: 中部大学
miaoj1229@mprg.cs.chubu.ac.jp

概要：自動運転では，昼夜を問わず道路上の物体の３次元形状を検出する必要があるため，LiDAR によって撮影された点群データを入力として利用することが多い．また，精度と速度を両立させるために，点群データから生成した擬似画像を用いる３次元物体検出手法が注目されている．しかしながら，点群データとして取得した反射強度や距離情報だけでは，３次元物体の角度に対する精度が不十分である．そこで，本研究では，物体の点群だけでなく，点群から推定した法線情報を擬似画像に変換し，他の擬似画像と組み合わせることで，より高精度な３次元物体検出手法を提案する．評価実験の結果，法線情報を利用しない場合と比べて，KITTI ベンチマークデータセットにおいて検出精度の向上を実現した．
＜キーワード＞ 物体検出，点群処理，自動運転支援

1. はじめに

自動運転シーンにおいて，物体検出は重要なタスクの１つである．このタスクは，物体検出の精度と速度の両方が重要であり，検出速度を高速化しつつ，精度を向上するための様々な手法が研究されている[1][2][3]．自動運転には，一般的なカメラから撮影された RGB 画像を用いる場合もあるが，昼夜を問わず物体を検出するために，カメラに比べて環境光の影響を受けない LiDAR によって撮影された点群データを利用することが多い．点群データを利用することで３次元空間での物体の位置を検出できる．これにより，物体の移動方向の予測や物体形状を捉えることができるため，自動運転での経路計画をより正確に行うことが可能となる．

一方，点群データは RGB 画像とは異なり，点群の並びが不規則なため，２次元畳み込みを利用することが困難である．そのため，点群データを Bird’s-eye view (BEV)マップで表現する方法がある[4]．BEV マップは，点群データを俯瞰視点からの２次元の擬似画像で表現する．これにより点群データを並びが規則的な画像に変換できる．しかしながら，異なる点群が同じ位置に配置されることがあり，点群データよりも表現力が低下する．また，３次元情報を２次元に削減するため，物体の形状情報が損なわれてしまう．

そこで本研究では，点群データから３次元法線ベクトル推定し，「Normal-map」として２次元の擬似画像に追加する．これにより，物体の形状情報を含む入力となり，検出精度の向上を図ることができる．また，従来の BEV マップを利用する手法に用いられるベースネットワークを You Only Look Once v4 (YOLOv4)[5]にすることで，より精度の高い物体検出を可能とする．

2. 関連研究

物体検出手法は，RGB 画像を入力とする手法と LiDAR により撮影した点群データを入力する手法がある．RGB 画像を入力とする手法は，速度を重視した１ステージ検出手法と精度を重視した２ステージ検出手法がある．１ステージ検出手法は，SSD [2]や YOLO [3]，YOLOv4[5]のように物体のクラススコアとバウンディングボックスを同時に出力することができる．一方，２ステージ検出手法は，Faster R-CNN[1]のように物体候補領域を抽出後，詳細に判定を行い，物体のクラススコアとバウンディングボックスを出力する．RGB 画像を入力とする手法は，高速かつ高精度であるが，検出した物体の３次元形状を捉えることができない．

点群データは並びが不規則なため，Deep Convolutional Neural Network(DCNN)に直接入力することができない．そのため，３次元空間を一定のグリッドに分割して，点群データをサンプリングする 3D voxels[6]，Range View[7]，BEV[4]，Point Set[8]などが提案されている．これにより，３次元空間に対して畳み込み処理を行うことができる．

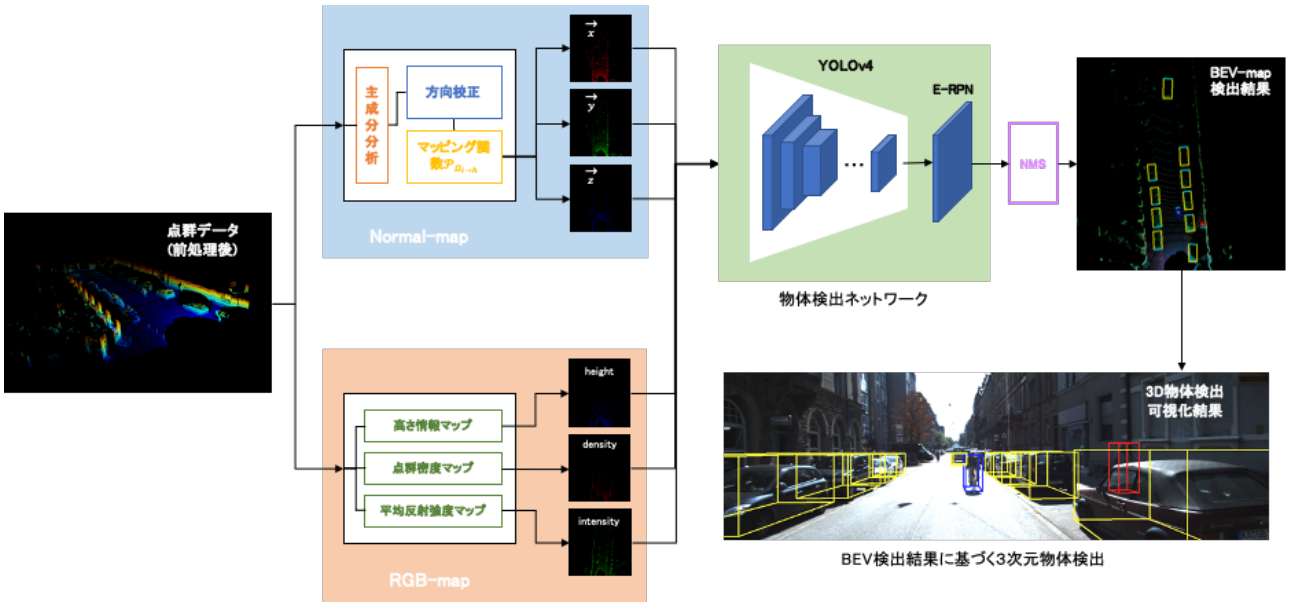


図 1 提案手法の概要

また，点群データを２次元の擬似画像に変換する MV3D[9], Complex-YOLO[4], BirdNet[10]が提案されている． Complex-YOLO では，点群データから高さ情報を表すチャンネル，点群密度を表すチャンネル，反射強度を表すチャンネルの３つの BEV マップを生成して RGB-map とする．これにより，RGB 画像と同じ３チャンネルとなるため，２次元畳み込みを使用することができ，YOLO のような高速なネットワークを用いて高速化が可能である．

3. 提案手法

本研究では，法線情報を追加した点群データからの３次元物体検出手法を提案する．図１に提案手法の概要を示す．

3.1. 点群の前処理

点群データは，LiDAR から照射したレーザーが物体に反射して戻ってきた点の集合であり，点の飛行時間から距離情報も取得できる．点群データは，LiDAR に近く物体は多くの点を取得でき，遠くの物体は疎となる．このように点の密度が距離により異なるため，点の並びが不規則となる．そこで，点群データの前処理が必要となる．

まず，点群データから点の並びが規則的となる BEV マップを作成する．BEV マップの画像サイズを幅 608 ピクセル，高さ 608 ピクセルと定義し，LiDAR の位置を原点とする．そして，奥行き 50m，水平 50m の領域内（$x \in [-25,25], y \in [0,50]$）の点群データを利用することとして，それより遠方の点群データを削除する．LiDAR は地上 1.73m の高

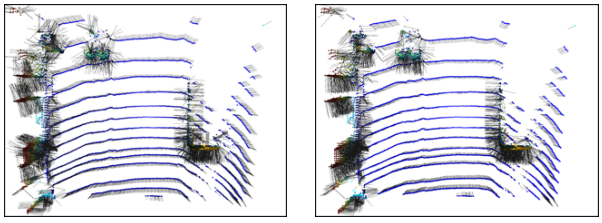


図 2 方向校正による法線方向の比較

さに設置しているとして，z 軸の範囲を$z \in [-2.73,1.27]$に設定する．各点群は，式(1)の$\mathcal{P}_\Omega$となる．

$$\mathcal{P}_\Omega = [x, y, z]^T \tag{1}$$

z 軸の範囲は，地上約 4m 以内の物体の点群を網羅できる．これは，トラックやバンなどの高さのある大きな物体を検出可能な範囲としている．

3.2. 物体法線の推定

法線は，主成分分析(PCA)[11]により前処理した点群データから探索半径 30cm，探索点数kとして，法線ベクトルを推定する．ここで，探索点数の最大を 50 点とする．式(2)の$\mathcal{C}$は，ある注目点の近傍における，点p_iと中央点$\bar{p}$の距離が要素である共分散行列である．PCA により，共分散行列から固有ベクトル$\vec{v}_j$を求める．第 1 主成分と第２主成分が点の多い方向であるため，平面と見なすことができる．λ_iは共分散行列の j 番目の固有値であり，その平面の垂線が法線であると推定できる．

$$\mathcal{C} = \frac{1}{k}\sum_{i=1}^{k}(p_i - \bar{p}) \cdot (p_i - \bar{p})^T \tag{2}$$

表 1 KITTI ベンチマークでの BEV 性能の評価結果

Method	FPS	Car			Pedestrian			Cyclist			Average
		Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	
BirdNet[10]	9.1	**84.17**	59.83	57.35	**28.20**	**23.06**	**21.65**	**58.64**	**41.56**	**36.94**	**45.93**
Complexer-YOLO[13]	16.7	77.24	68.96	64.95	21.42	18.26	17.06	32.00	25.40	22.88	38.68
提案手法	5.5	72.84	**71.52**	**67.50**	26.71	21.19	20.17	42.50	36.06	31.18	43.30

しかし，このままでは，どちらが内側か外側か判断できない．

そこで方向校正により，平面上における各法線が，同じ方向になるように LiDAR に対して，反対側に向いている法線のみを反転させる．図２は校正した法線の可視化例である．図２の左側と比較して，右側の図では法線が自動車の外側に向くように修正されていることが分かる．

3.3. Normal-map の生成

前節で求めた各点に対する法線を２次元での畳み込み処理を可能とするために, Normal -map を作成する．Normal-map の作成には，式(3)に示すマッピング関数を用いる．これにより，各点群を２次元空間にマッピングする際，最も高い位置にある点$\mathcal{P}_{\Omega_{j\to h}}$の法線ベクトル$\vec{x}, \vec{y}, \vec{z}$を利用することができる．

$$\mathcal{P}_{\Omega_{i\to h}} = \left\{ \mathcal{P}_{\Omega_i} = [x, y, z]^T \middle| \mathcal{S}_j = f_{PS}\left(\mathcal{P}_{\Omega_i}, g\right) \right\} \quad (3)$$

各点の法線ベクトルを式(4)のように，軸ごとに$\mathcal{P}_{\Omega_{i\to h}}$から$normal_{\vec{x}}$, $normal_{\vec{y}}$, $normal_{\vec{z}}$として抽出する．法線ベクトルを軸ごとに２次元擬似画像で表現する．

$$\begin{aligned} normal_{\vec{x}}\left(\mathcal{S}_j\right) &= \vec{x}\left(\mathcal{P}_{\Omega_{j\to h}}\right), \\ normal_{\vec{y}}\left(\mathcal{S}_j\right) &= \vec{y}\left(\mathcal{P}_{\Omega_{j\to h}}\right), \\ normal_{\vec{z}}\left(\mathcal{S}_j\right) &= \vec{z}\left(\mathcal{P}_{\Omega_{j\to h}}\right). \end{aligned} \quad (4)$$

法線推定の探索範囲（半径 30cm）は BEV マップのピクセルの表現範囲$50m\ /680\ pixels = \ 8cm/pixel$より広いため，より大域の情報を含むことができる．また，法線ベクトルから算出した Normal-map も BEV マップであるため，他の BEV マップと自由に組み合わせて，入力データにすることができる．

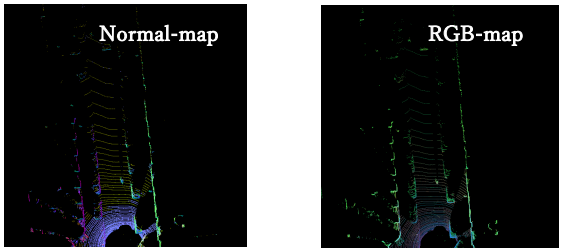


図 3 入力画像

3.4. 入力情報

提案手法のネットワークの入力は，図３に示す RGB-map と Normal map を用いる．RGB-map は，Complex-YOLO と同様であり，高さ情報の height マップ，点群密度を表す density マップ，平均反射強度を表す intensity マップから構成される．Normal map は，x, y, z の各軸おける法線ベクトルである．入力には，これらのマップをチャネル方向に連結した６チャンネルの BEV マップを用いる．

3.5. 物体検出ネットワーク

前節で述べた BEV マップを入力とし，Euler-Region Proposal Network(E-RPN)[4]で物体のクラスやサイズを予測し，３次元物体検出を行う．3 次元物体検出のベースネットワークには YOLOv4 を用いる．E-RPN は，物体位置b_xとb_y，サイズb_wとb_l，ヨー角度 b_ϕ及び BEV マップの物体のクラススコアと分類の信頼度を出力する．b_ϕは直接予測できないため，$b_\phi = tan^{-1}(t_{Im}, t_{Re})$により求める．３次元のバウンディングボックスのうち，幅および奥行きをネットワークで推定する．高さは，計算コストを節約するために事前定義した値を使用する．

ここで，本研究では，自動車，歩行者，自転車の３クラス検出対象とする．損失関数には，Mean-square error(MSE) を用いる．

4. 評価実験

提案手法の有効性を調査するため，KITTI データセット[12]を用いて，従来手法との比較実験を

表 2 入力に用いる BEV マップによる精度比較

	BEV マップ		mAP	クラス		
	RGB-map	Normal-map	(IoU-50)	自動車	歩行者	自転車
従来手法	○	-	87.73	96.26	76.83	90.10
提案手法	○	○	91.94(4.21↑)	97.03	84.54	94.24

行う．Normal-map の有無による検出精度の比較も行う．また，きょう角算出関数を追加することにより，物体ヨー角の精度を評価する．

4.1. 実験概要

本実験では，KITTI 訓練データを 4 対 1 の割合で分割し，6000 枚を提案手法の学習，1481 枚を検証に用いた．学習時，水平反転，スケーリング，random crop，モザイクと random padding のデータ拡張を RBG-map に適用した．本研究では，点群データのみを用いており，画像データは検出結果の可視化にのみ利用する．

3.5 章で述べたように，提案手法では高さ情報を出力しない．よって，出力および正解情報に対して，幅および奥行き情報から算出した 2 次元矩形について mean Intersection over Union (mIoU)を式(5)のように算出して評価する．

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (5)$$

4.2. KITTI ベンチマーク評価結果

KITTI ベンチマーク[12]での評価結果を表 1 に示す．従来手法と比べて，提案手法は低難易度（Easy）において，自動車の検出精度が 72.84%と他の手法よりも低いが，高難度（Hard）の場合に 67.50%と最も精度が高い．また，同じ BEV ベースの BirdNet[10]と，ほぼ同程度の精度を達成している．同一クラスの物体であっても，検出の難しさが異なる場合には，法線情報の追加により物体検出に頑健性が向上している．

また，提案手法は同じ YOLO ベースのネットワークを用いた Complexer-YOLO [13]よりも高い Average Precision(AP)を達成している．入力は BEV マップのみであるが，歩行者や自転車などの非平面形状の物体の精度は，難易度ごとに優れている．

4.3. Normal-map の有効性評価

Normal-map の有効性を確認するため，RGB-map と Normal-map の組み合わせによる比較を行った．表 2 に各組み合わせによる検出精度を示す．従来手法の RGB-map のみを入力した場合では自動車，

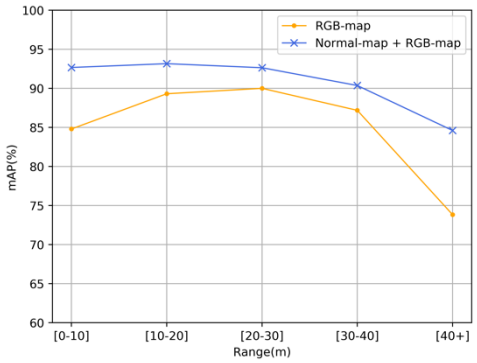


図 4 距離値による物体検出 mAP

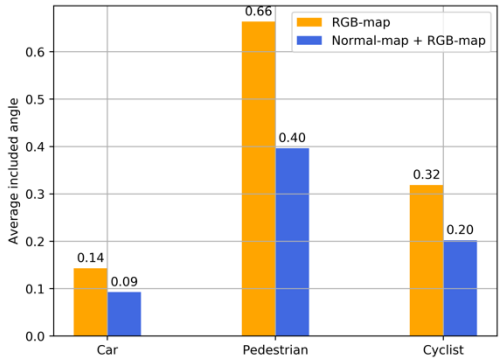


図 5 各クラスの平均きょう角

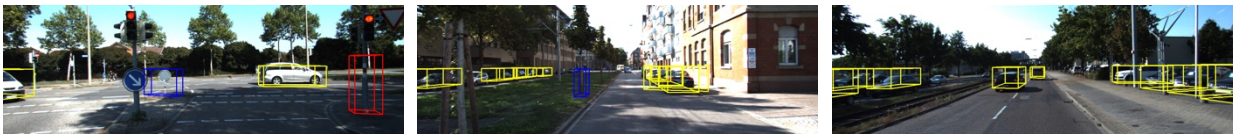
歩行者，自転車の 3 つのクラスの mAP は 87.73 である．Normal-map と RGB-map の両方を入力した場合，mAP は 91.94 となった．これより Normal-map を追加することで全体の検出精度が向上していることがわかる．一方，クラスごとの検出精度では，自動車クラスで若干精度が向上しているものの，歩行者クラスで大幅に検出精度が向上していることが分かる．

4.4. 距離値による評価

点群データは距離値により，密度が変わるため，距離ごとの検出精度を評価する．評価結果を図 4 に示す．法線情報を追加した検出精度は従来手法と比較して，30m 程度まで検出精度が低下しない．また，40m 以上でも従来手法よりも精度低下が小さいことが分かる．

4.5. 角度予測の評価

法線は物体の形状情報であるため，Normal-map を追加することでヨー角の精度を向上させること



RGB-map のみ

Normal-map のみ

Normal-map + RGB-map

図 6 物体検出結果の可視化例

ができる．そこで，物体角度予測の評価実験を行う．空間座標系において，πと$-\pi$の両方が同じ方向を向いていることを考慮すると，予測したヨー角と真値の誤差を直接計算することはできない．従って，予測したヨー角は(Im_p, Re_p)ベクトルであり，真値は(Im_{gt}, Re_{gt})であると想定する．きょう角θは式(7) に示すように計算できる．

$$\overrightarrow{a_p} \cdot \overrightarrow{b_{gt}} = \left|\overrightarrow{a_p}\right|\left|\overrightarrow{b_{gt}}\right| \cos\theta \tag{6}$$

$$\cos\theta = \frac{\sqrt{{Im_p}^2 + {Re_p}^2} \cdot \sqrt{{Im_{gt}}^2 + {Re_{gt}}^2}}{Im_p Im_{gt} \cdot Re_p Re_{gt}} \tag{7}$$

推定した物体角度値と真値の平均きょう角$angle_{cls}$を算出する．

$$angle_{cls} = \frac{1}{n}\sum_{k=1}^{n} \arccos\theta_k \tag{8}$$

図５に示すように，従来手法と比較して，提案手法はきょう角の推定精度が向上していることが分かる．特に，自動車のような大きく，平面な形状を持つ物体に対して角度予測の精度がより向上している．また，歩行者のような円柱な物体に対しても，角度予測の精度が向上することが分かる．

4.6. 可視化結果

３次元物体検出の可視化結果を図 6 および図 7 に示す．従来手法は，信号機や交通標識などの物体を人間やサイクリストに誤検出しやすい．これら物体は，円柱状の形状をしており，高さが人間に類似しているため，誤って予測している．一方，法線情報を追加した提案手法は Normal-map のみでも，誤検出が減少している．このことから，法線情報を利用することで，人の特徴と似ている物体の誤検出を改善できることがわかった．

5. おわりに

本研究では，法線情報を追加した点群からの３次元物体検出手法を提案した．従来手法と比較して高い検出精度を獲得することを確認した．Normal-map はシミュレータのデータセットにも使用できるため，実環境でないデータセットを用いることも可能である．一方，法線を推定するため，全体の処理時間が増加する．今後は，精度向上を実現するために，セマンティック画像と点群データを融合した手法を検討する．また，より多くクラスの物体を高速で検出させる手法を検討する．

参考文献

[1] S. Ren, *et al.*, “Faster r-cnn: Towards real-time object detection with region proposal networks”, NeurIPS, 91-99, 2015.

[2] W. Liu, *et al.*, “Ssd: Single shot multibox detector”, ECCV, 21-37, 2016.

[3] J. Redmon, *et al.*, “You only look once: Unified, real-time object detection”, CVPR, 779-788, 2016.

[4] M. Simony, *et al.*, “Complex-yolo: An euler-region-proposal for real-time 3d object detection on point clouds”, ECCV, 0-0, 2018.

[5] A. Bochkovskiy, *et al.*, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, arXiv, 2004.10934, 2020.

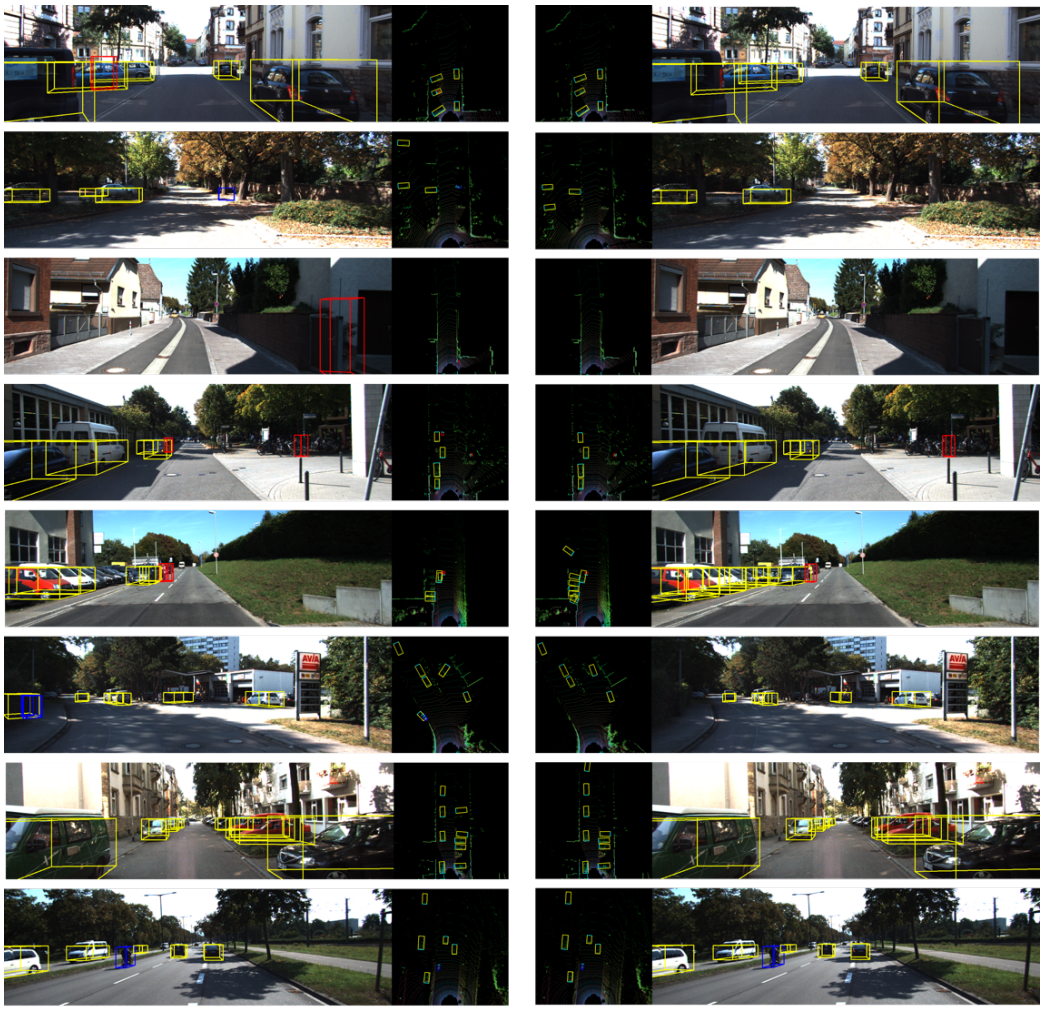


図 7 物体検出結果の可視化例

[6] Y. Zhou, *et al.*, “Voxelnet: End-to-end learning for point cloud based 3d object detection”, CVPR, pp. 4490-4499, 2018.

[7] G. P. Meyer, *et al.*, “Lasernet: An efficient probabilistic 3d object detector for autonomous driving”, CVPR, pp. 12677-12686, 2019.

[8] C. Qi, *et al.*, “Pointnet++: Deep hierarchical feature learning on point sets in a metric space”, NeurIPS, pp. 5099-5108, 2017.

[9] X. Chen, *et al.*, “Multi-view 3d object detection network for autonomous driving”, CVPR, 1907-1915, 2017.

[10] J. Beltran, *et al.*, “Birdnet: a 3d object detection framework from lidar information”, ITSC, pp. 3517-3523, 2018.

[11] S. Wold, *et al.*, “Principal component analysis”, Chemome--trics and intelligent laboratory systems, 37-52, 1987.

[12] A. Geiger, *et al.*, “Are we ready for autonomous driving? the kitti vision benchmark suite”, CVPR, 3354-3361, 2012.

[13] M. Simon, *et al.*, “Complexer-YOLO: Real-time 3D object detection and tracking on semantic point clouds”, CVPR, 0-0, 2019.

繆継樹：2018 年紹興文理学院工学部機械工学科卒業．現在中部大学大学院修士課程在学中，自動運転において点群を用いた物体検出の研究に従事．

平川翼：2013 年広島大学大学院博士課程前期修了，2014 年広島大学大学院博士課程後期入学，2017 年中部大学研究員(~2019)，2017 年広島大学大学院博士後期課程修了．2019 年中部大学特任助教．コンピュータビジョン，パターン認識，医用画像処理の研究に従事．

山下隆義：2002 年奈良先端科学技術大学大学院大学博士前期課程修了．2002 年オムロン株式会社入社，2009 年中部大学大学院博士後期課程修了(社会人ドクター)，2014 年中部大学講師，2017 年より同大学准教授．人の理解に向けた動画像処理，パターン認識・機械学習の研究に従事．

藤吉弘亘：1997 年中部大学大学院博士後期課程修了．1997～2000 年米カーネギーメロン大学ロボット工学研究所 – Postdoctoral Fellow．2000 年中部大学講師．2004 年より同大学教授．2005～2006 年米カーネギーメロン大学ロボット工学研究所客員研究員，計算機視覚，動画像処理，パターン認識・理解の研究に従事．